

	A	B	C	D	E	F	G
1	Date ▾	Weekday ▾	Amount ▾	AcctType ▾	OpenedBy ▾	Branch ▾	Customer ▾
2	1-Nov	Friday	5,000	IRA	New Accts	Central	Existing
3	1-Nov	Friday	14,571	CD	Teller	Central	New
4	1-Nov	Friday	500	Checking	New Accts	Central	Existing
5	1-Nov	Friday	15,000	CD	New Accts	Central	Existing
6	1-Nov	Friday	4,623	Savings	New Accts	North Cou	Existing
7	1-Nov	Friday	8,721	Savings	New Accts	Westside	New
8	1-Nov	Friday	15,276	Savings	New Accts	North Cou	Existing
9	1-Nov	Friday	5,000	Savings	New Accts	Westside	Existing

Figure 1.1 The bank data.

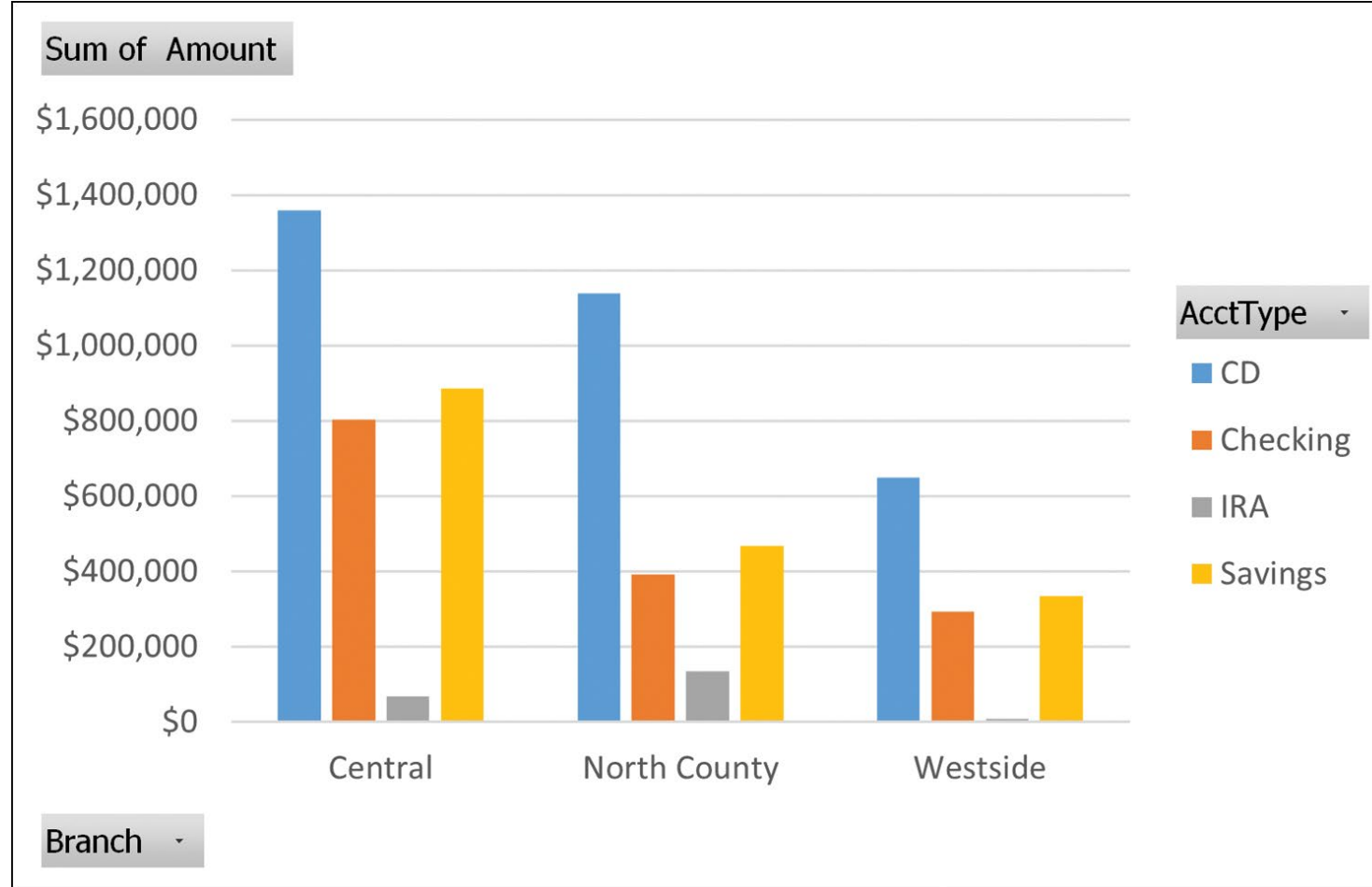


Figure 1.2 Bank deposits by branch and account type.

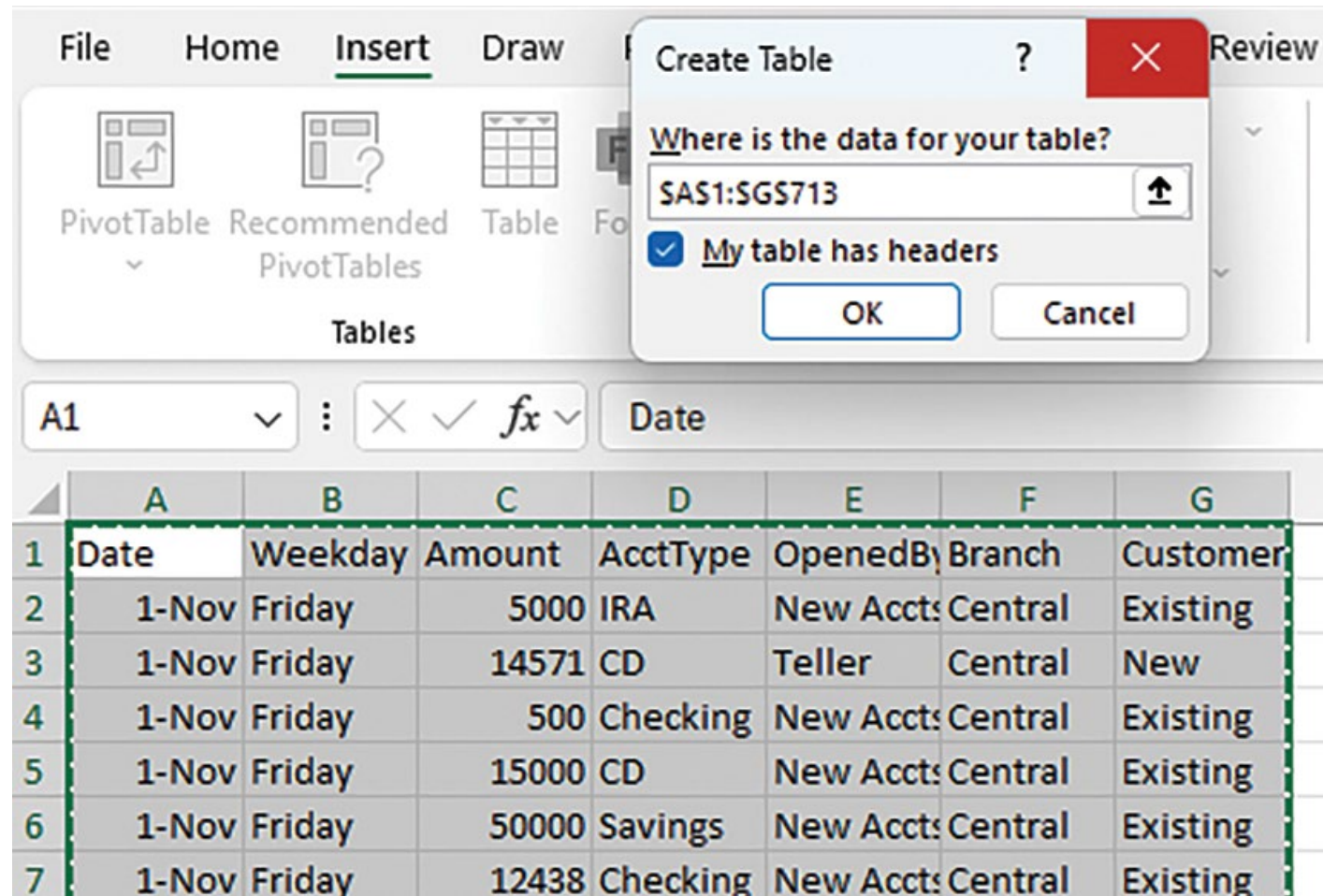


Figure 1.3 The Create Table dialog.

	A	B	C	D	E	F	G
1	Date	Weekday	Amount	AcctType	OpenedBy	Branch	Customer
2	1-Nov	Friday					Existing
3	1-Nov	Friday					New
4	1-Nov	Friday					Existing
5	1-Nov	Friday					Existing
6	1-Nov	Friday					Existing
7	1-Nov	Friday					Existing
8	1-Nov	Friday					Existing
9	1-Nov	Friday					New
10	1-Nov	Friday					Existing
11	1-Nov	Friday					Existing
12	1-Nov	Friday					Existing
13	1-Nov	Friday					New
14	1-Nov	Friday					New
15	1-Nov	Friday					New
16	1-Nov	Friday					New
17	1-Nov	Friday					Existing
18	1-Nov	Friday					Existing
19	1-Nov	Friday					Existing
20	1-Nov	Friday					New
21	1-Nov	Friday					Existing
22	1-Nov	Friday					New
23	1-Nov	Friday					Existing
24	1-Nov	Friday					Existing
25	1-Nov	Friday	7427	Checking	New Accts	North Cou	Existing
26	1-Nov	Friday	4500	Checking	New Accts	North Cou	New
27	1-Nov	Friday	45000	CD	Teller	North Cou	Existing

Figure 1.4 Bank data showing the Branch sorting and filtering option.

	A	B	C	D	E	F	G
1	Date	Weekday	Amount	AcctType	OpenedBy	Branch	Customer
2	1-Nov	Friday	5000	IRA	New Accts	Central	Existing
3	1-Nov	Friday	14571	CD	Teller	Central	New
4	1-Nov	Friday	500	Checking	New Accts	Central	Existing
5	1-Nov	Friday	15000	CD	New Accts	Central	Existing
6	1-Nov	Friday	4623	Savings	New Accts	North Cou	Existing
7	1-Nov	Friday	8721	Savings	New Accts	Westside	New
8	1-Nov	Friday	15276	Savings	New Accts	North Cou	Existing

Figure 1.5 The CSV version of the file has lost the table architecture.

Date	Weekday	Amount	AcctType	OpenedBy	Branch	Customer
1-Nov	Friday	5,000	IRA	New Accts	Central	Existing
1-Nov	Friday	14,571	CD	Teller	Central	New
1-Nov	Friday	500	Checking	New Accts	Central	Existing
1-Nov	Friday	15,000	CD	New Accts	Central	Existing
1-Nov	Friday	4,623	Savings	New Accts	North County	Existing
1-Nov	Friday	8,721	Savings	New Accts	Westside	New
1-Nov	Friday	15,276	Savings	New Accts	North County	Existing
1-Nov	Friday	5,000	Savings	New Accts	Westside	Existing
1-Nov	Friday	15,759	CD	Teller	Westside	Existing
1-Nov	Friday	12,000	CD	New Accts	Westside	Existing
1-Nov	Friday	7,177	Savings	Teller	North County	Existing
1-Nov	Friday	6,837	Savings	New Accts	Westside	Existing
1-Nov	Friday	3,171	Checking	New Accts	Westside	Existing

Figure 1.6 The data in a Notepad text file.

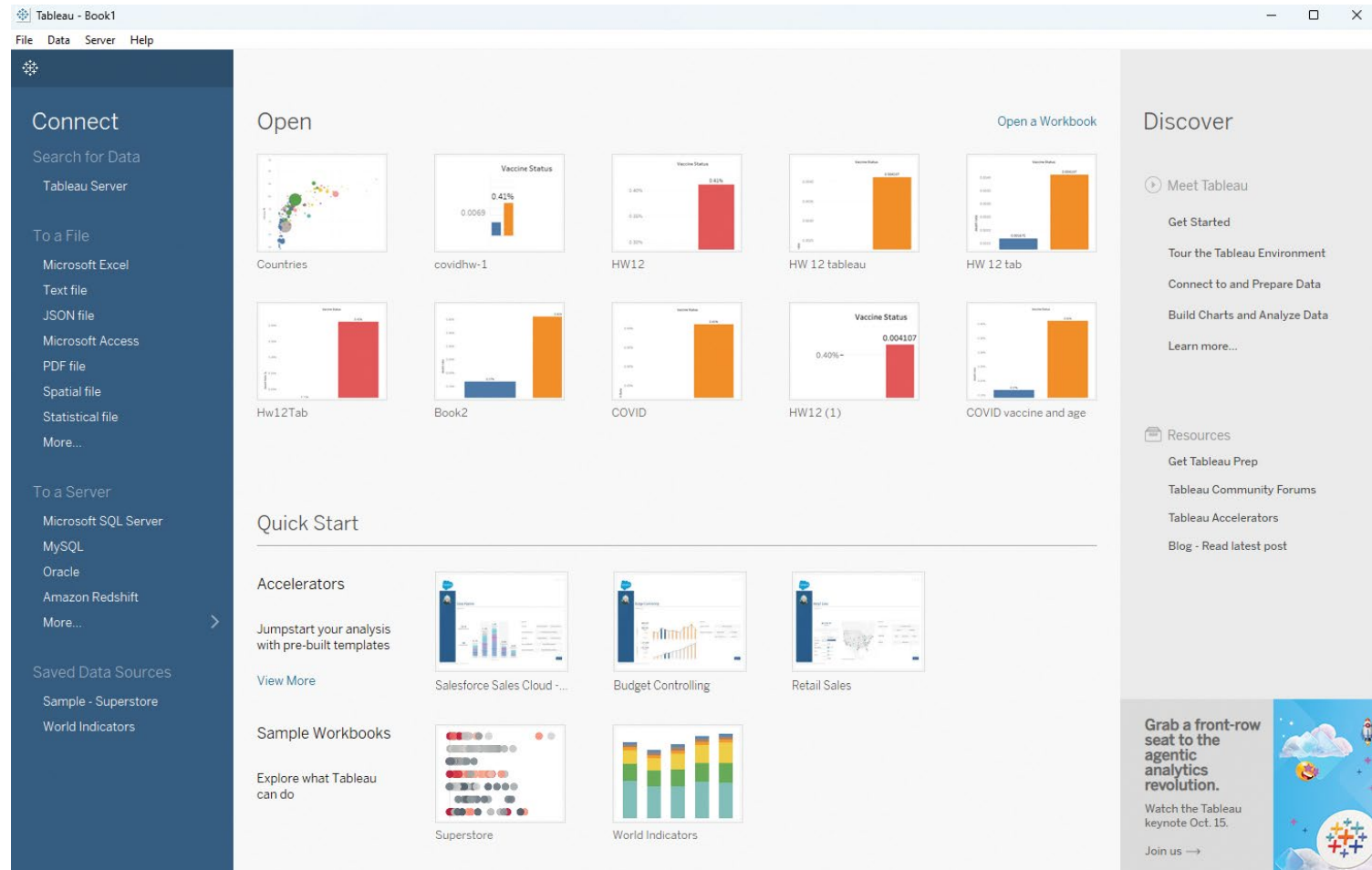


Figure 1.7 The opening screen for Tableau.

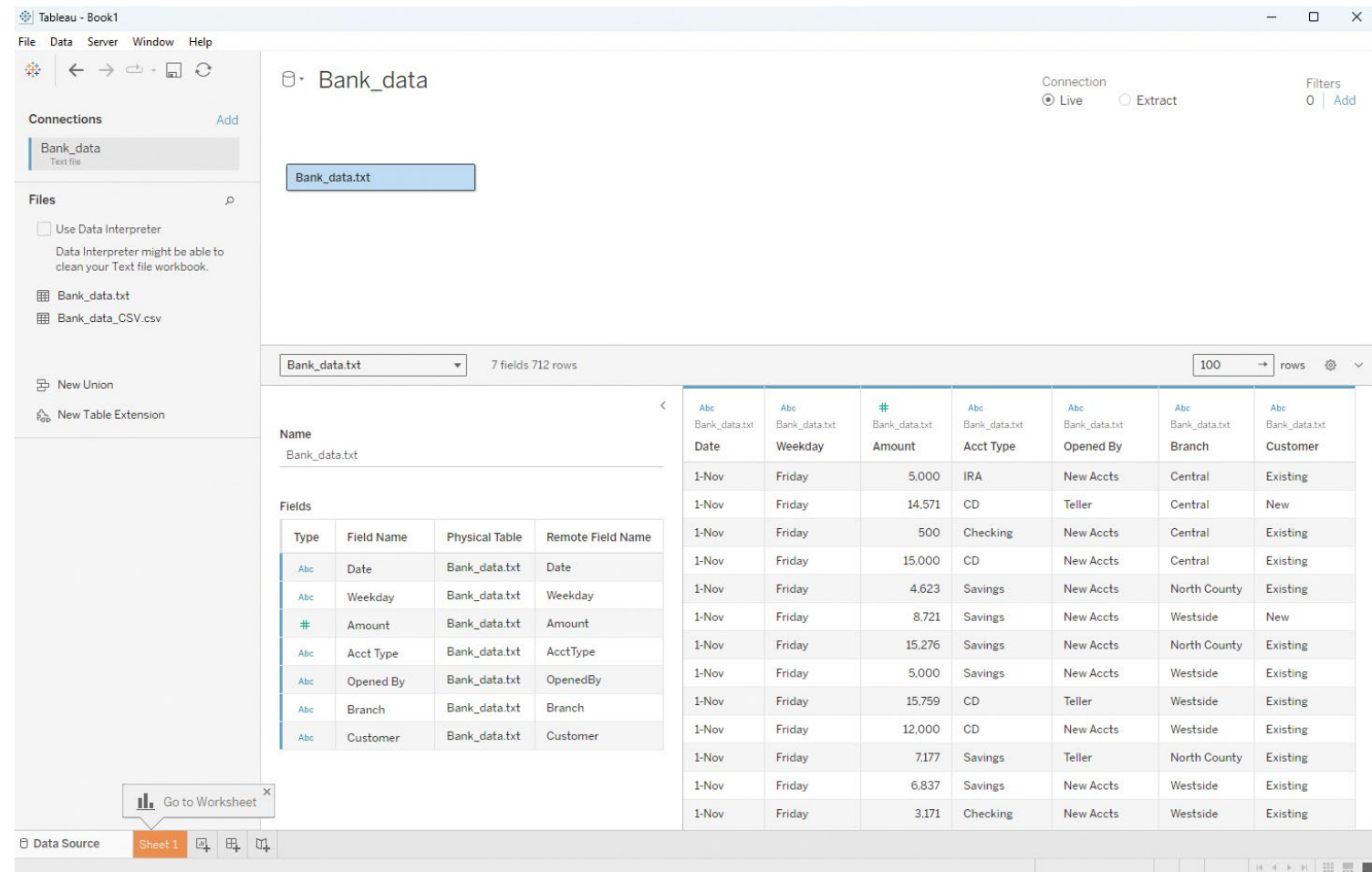


Figure 1.8 Bank data in Tableau.

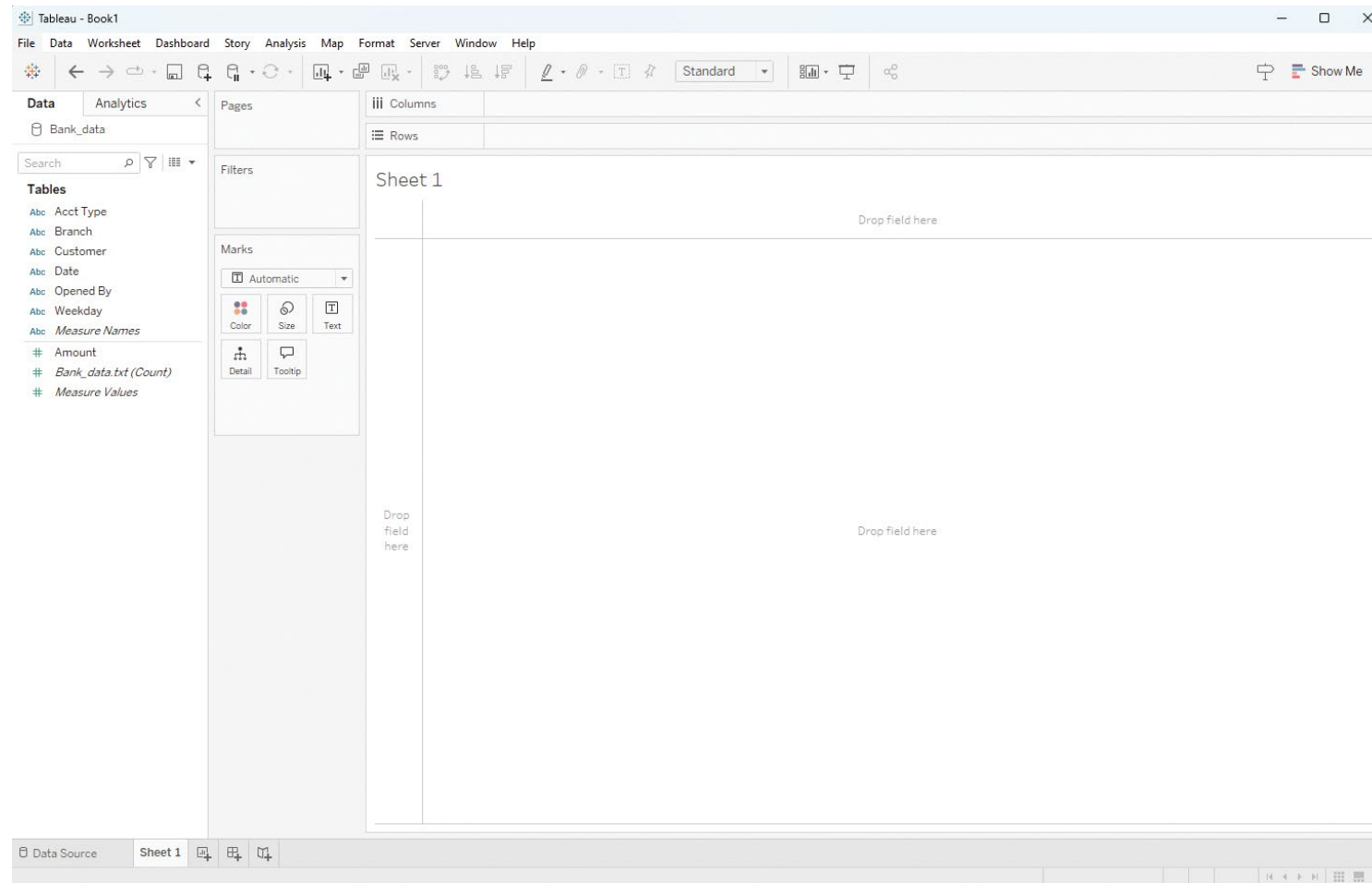


Figure 1.9 Bank data as a worksheet in Tableau.

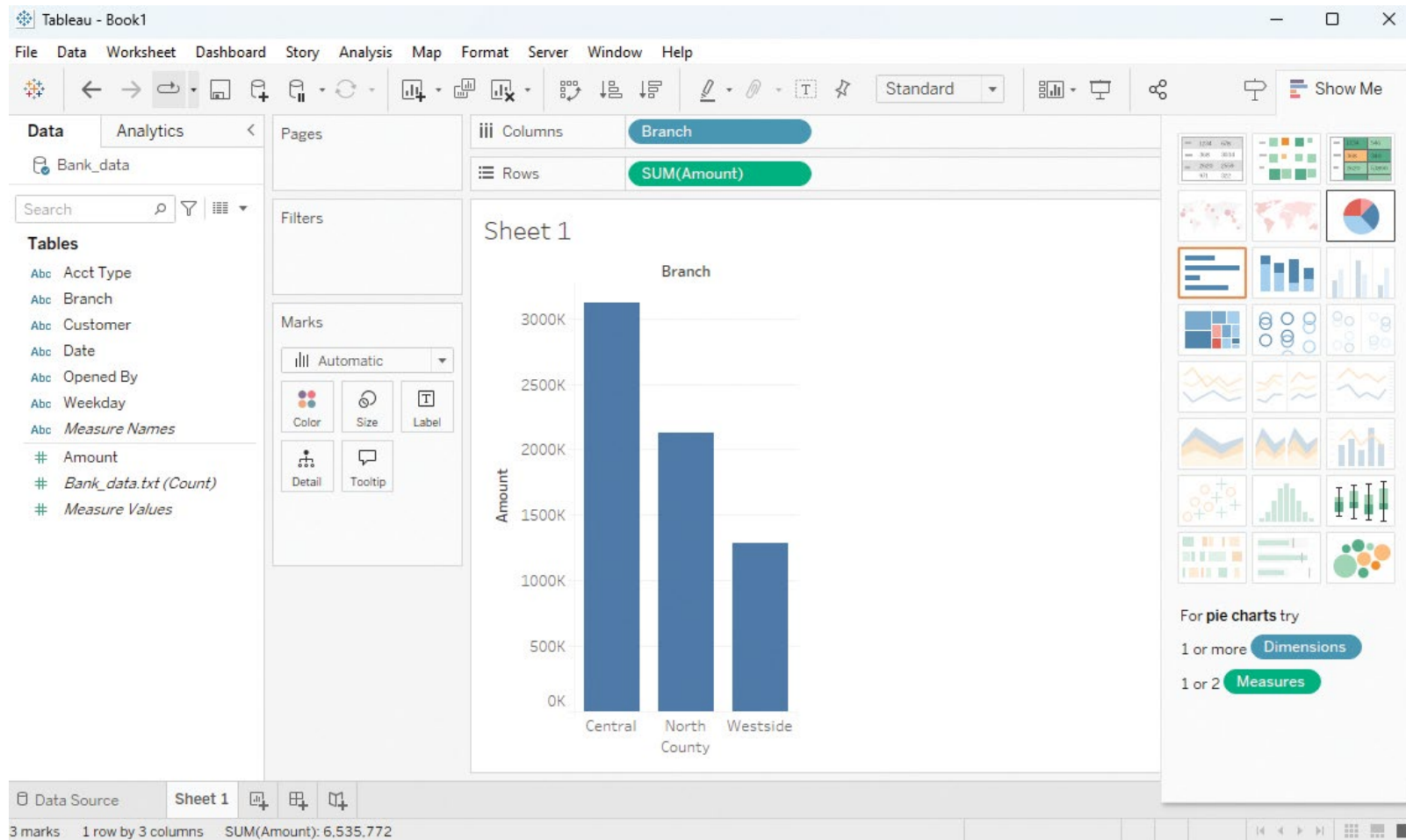


Figure 1.10 The result of dragging the Amount into the Rows shelf and the Branch into the Columns shelf.

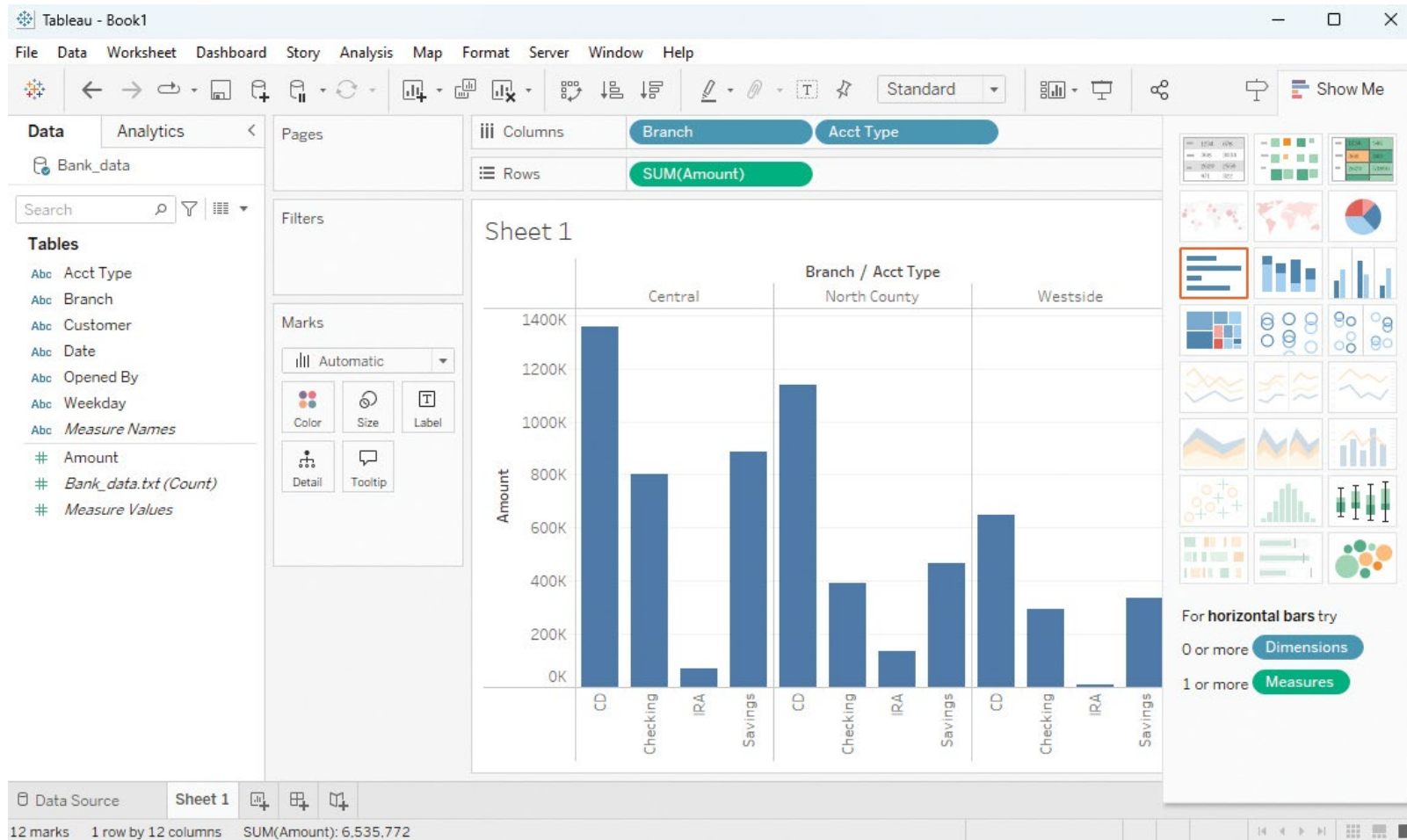


Figure 1.11 The result of adding the Acct Type.

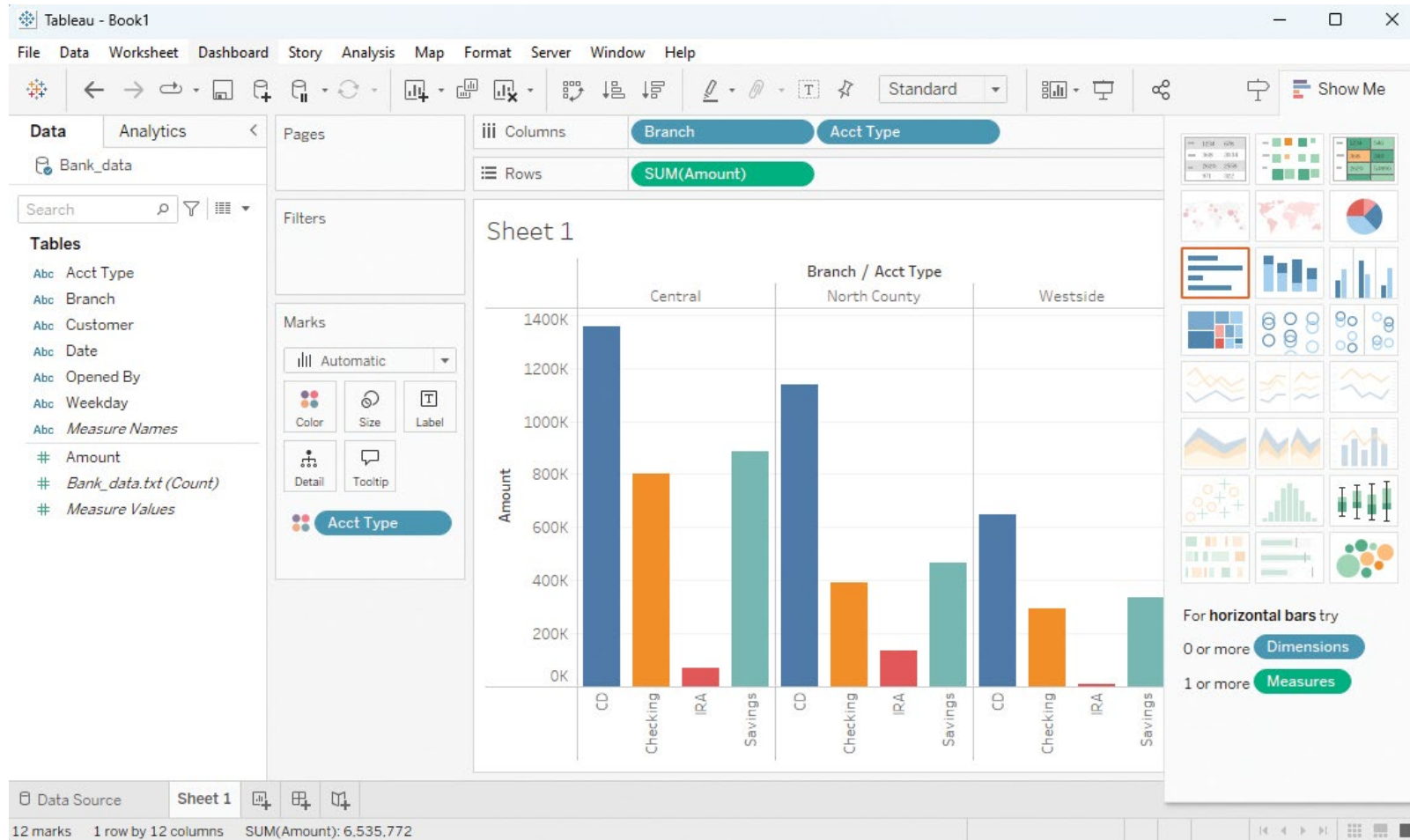


Figure 1.12 Coloring by Acct Type by dragging this dimension onto the Color Mark.

AutoSave On Bank_data2 No Label Saved

File Home **Insert** Draw Page Layout Formulas Data Review View

PivotTable Recommended PivotTables Table Forms

Pictures 3D Models Shapes SmartArt Icons Screenshot

Checkbox

Tables Illustrations Controls

A2 : fx Date

	Date	Weekday	Amount	AcctType	OpenedBy	Branch	Customer
2	1-Nov	Friday	5,000	IRA	New Accts	Central	Existing
3	1-Nov	Friday	14,571	CD	Teller	Central	New
4	1-Nov	Friday	500	Checking	New Accts	Central	Existing
5	1-Nov	Friday	15,000	CD	New Accts	Central	Existing
6	1-Nov	Friday	4,623	Savings	New Accts	North Cou	Existing
7	1-Nov	Friday	8,721	Savings	New Accts	Westside	New

Figure 1.13 To insert a PivotTable, highlight all of the data and click PivotTable.

PivotTable from table or range ? X

Select a table or range

Table/Range: Table2 ↗

Choose where you want the PivotTable to be placed

☒ New Worksheet

☐ Existing Worksheet

Location: ↗

Choose whether you want to analyze multiple tables

☐ Add this data to the Data Model

OK Cancel

Figure 1.14 The Create PivotTable dialog.

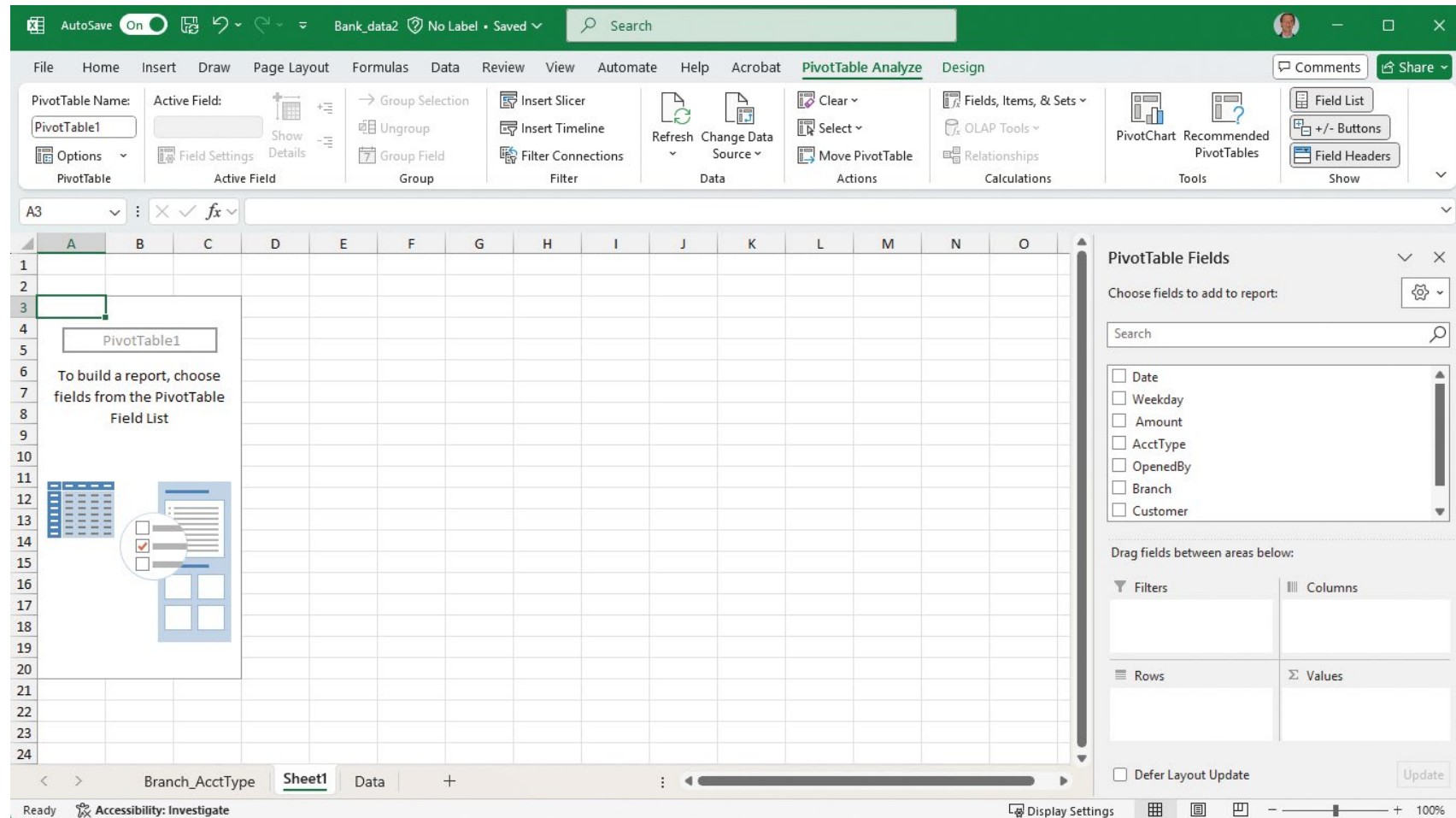


Figure 1.15 Inserting a PivotTable into a new sheet.

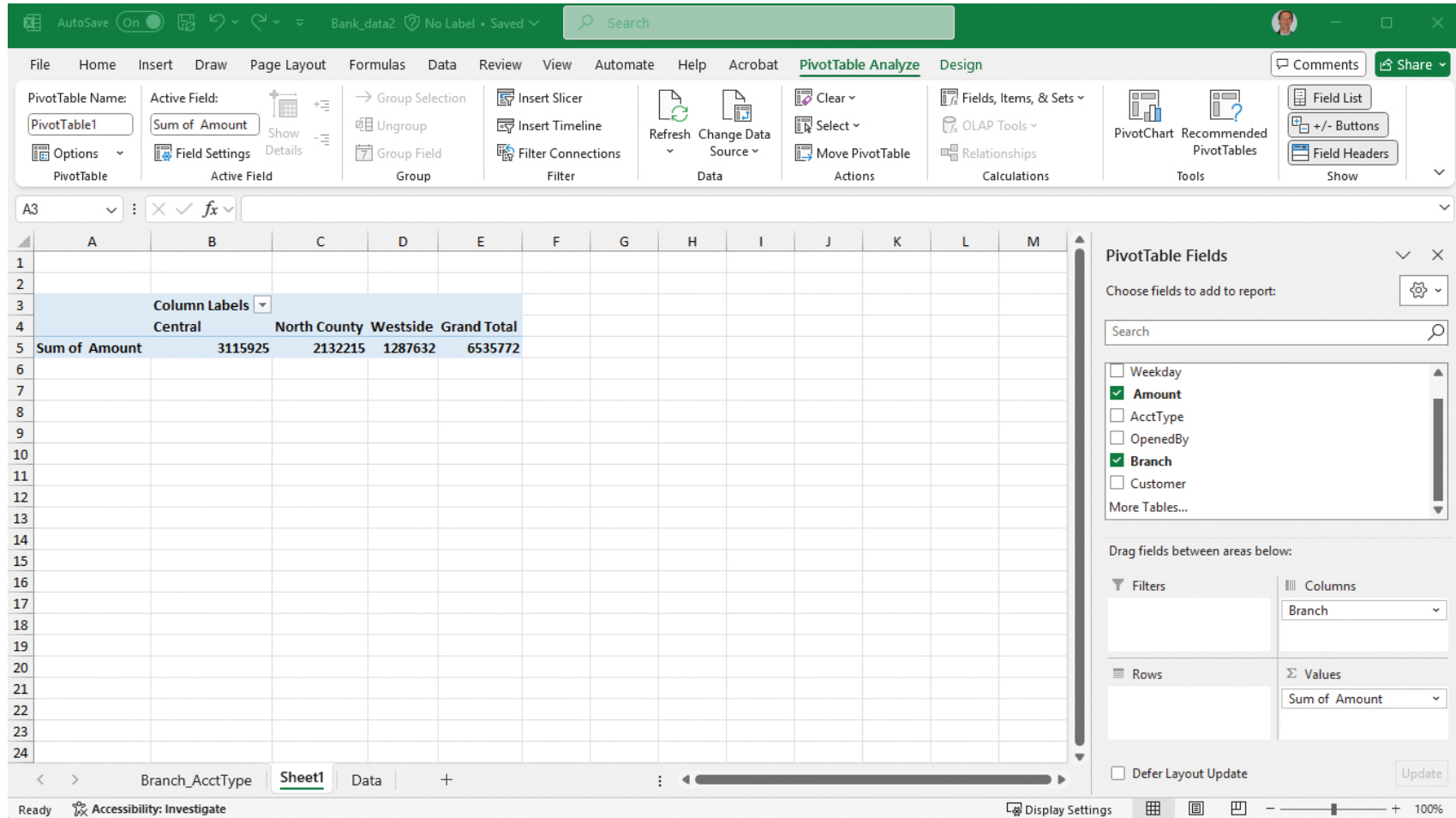


Figure 1.16 PivotTable of Branch by Amount with aggregation being done using SUM, and a drop-down arrow providing more options.

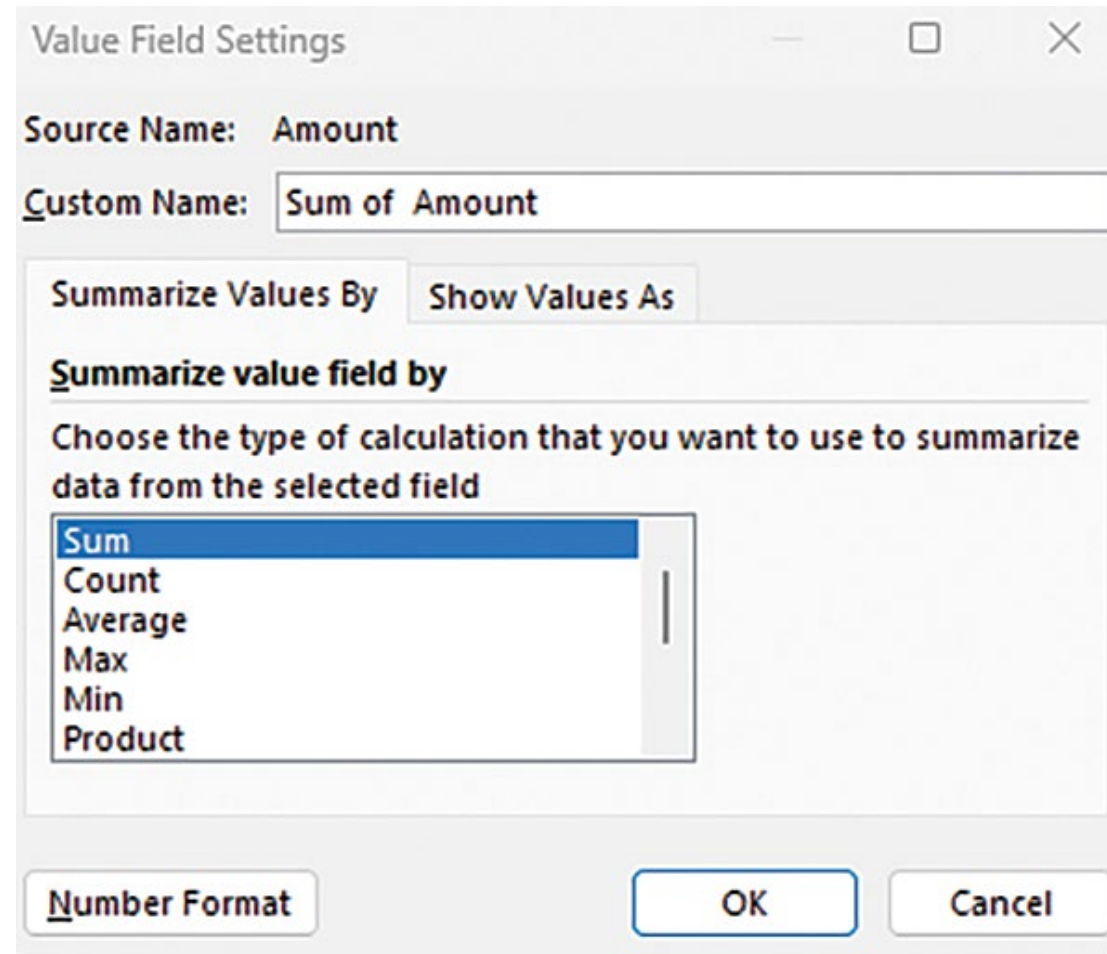


Figure 1.17 The Value Field Settings dialog box.

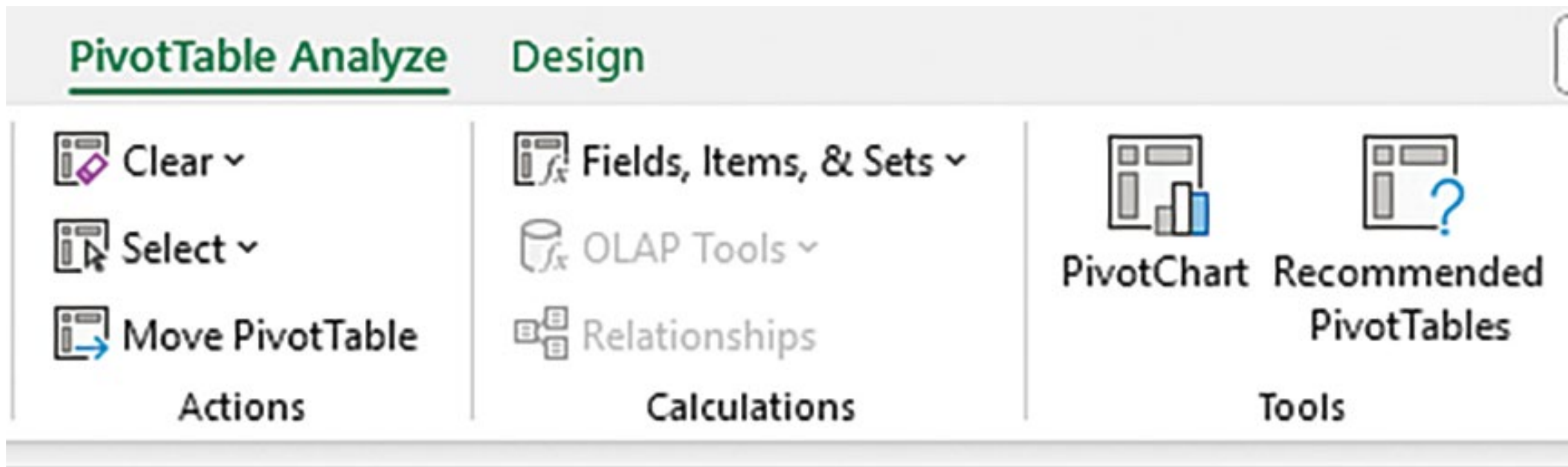


Figure 1.18 PivotChart option in the PivotTable Tools menu.

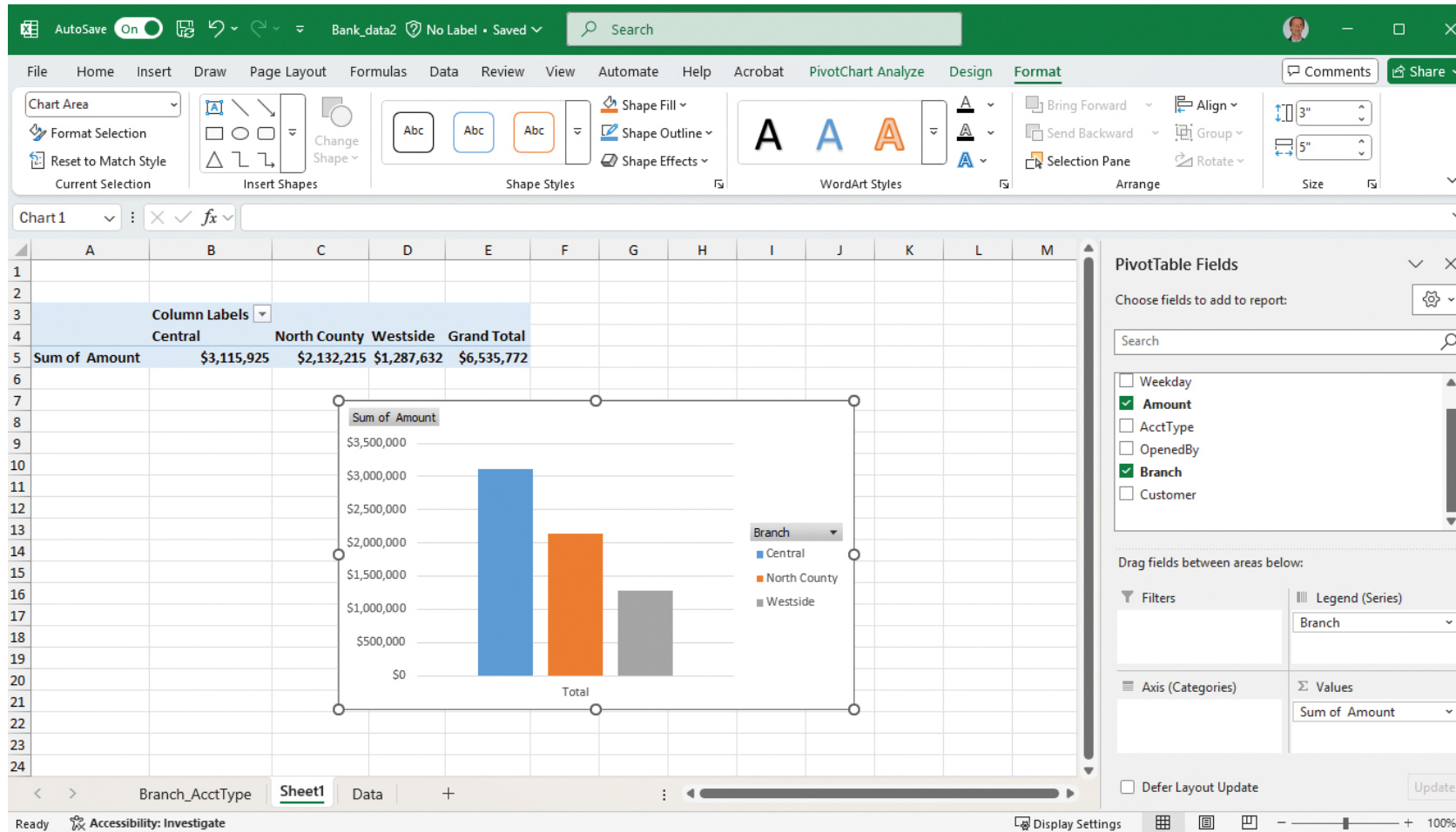


Figure 1.19 Column chart of Amount by Branch.

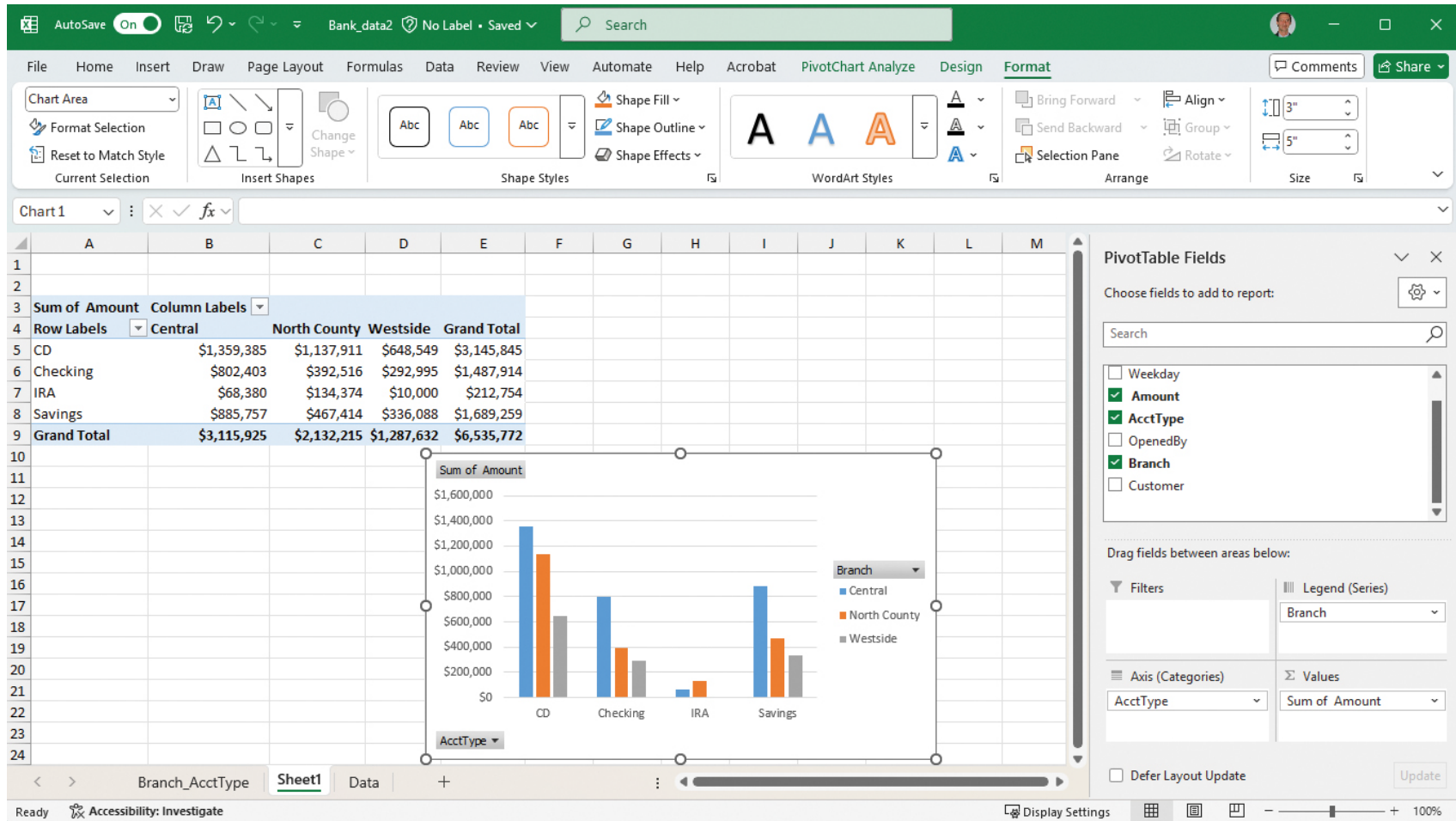


Figure 1.20 Column chart of Amount by Branch and AcctType.

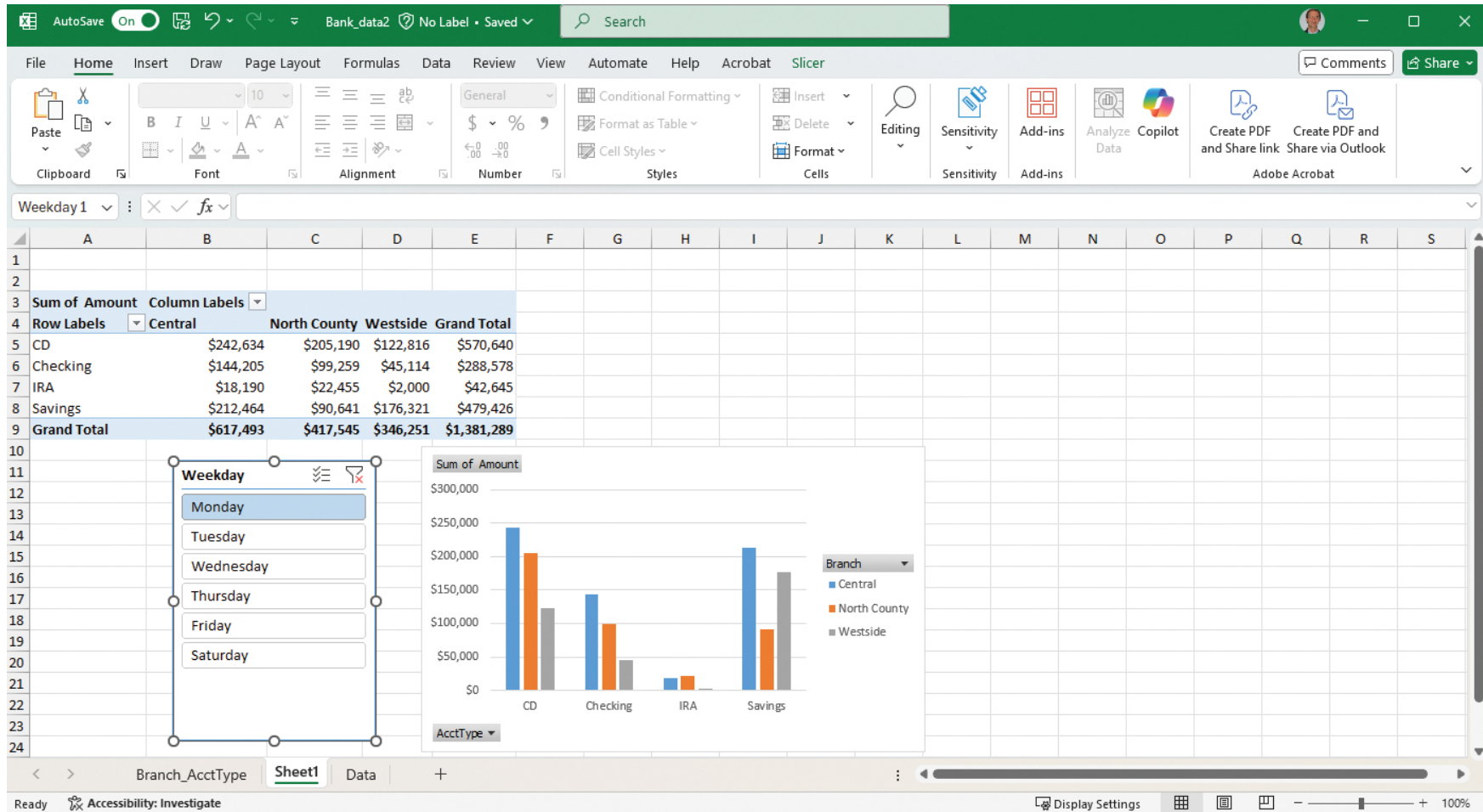


Figure 1.21 Total deposits by branch and account types for only Monday using a slicer.

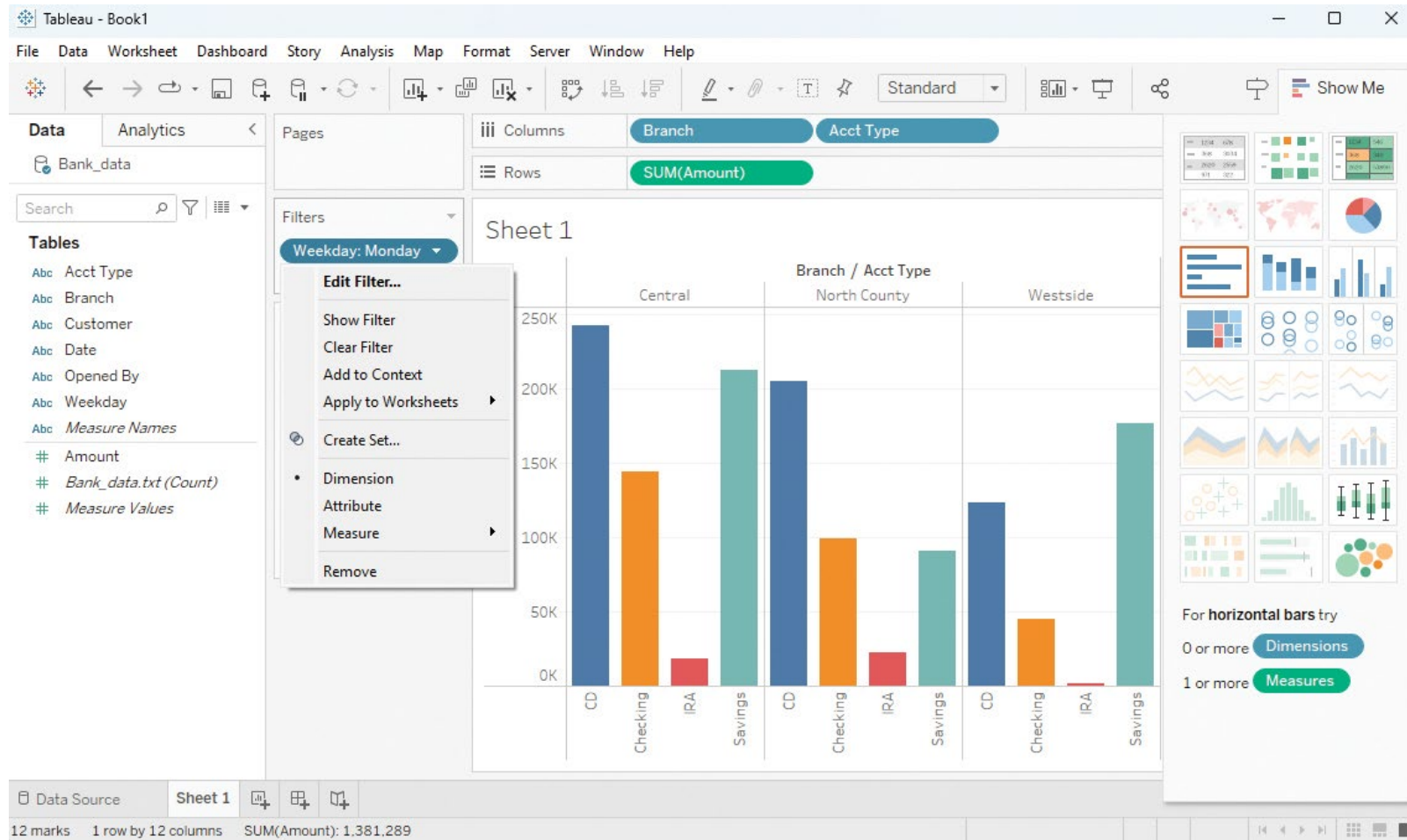


Figure 1.22 Choose a different day or set of days by editing the filter.

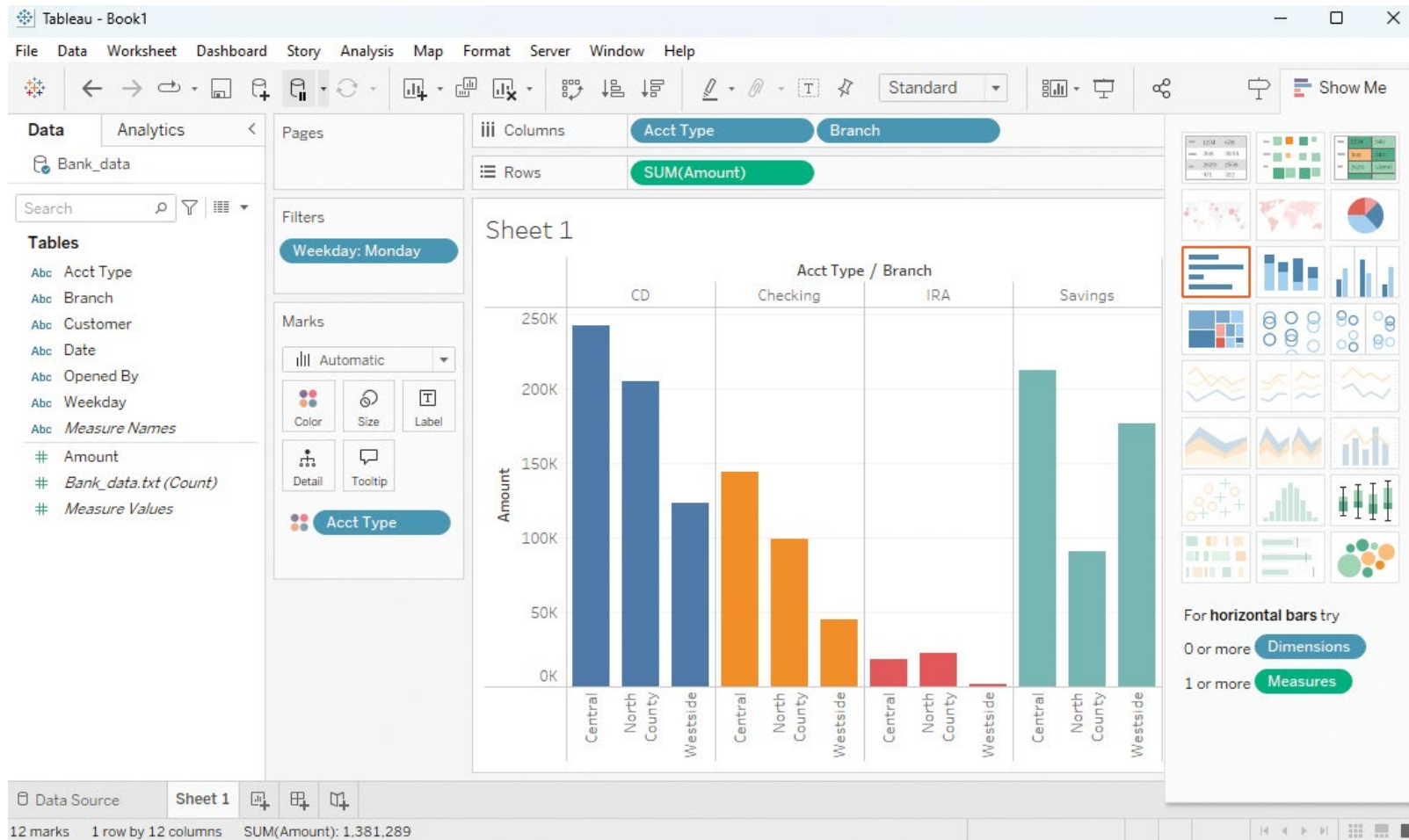


Figure 1.23 Coloring by account type.

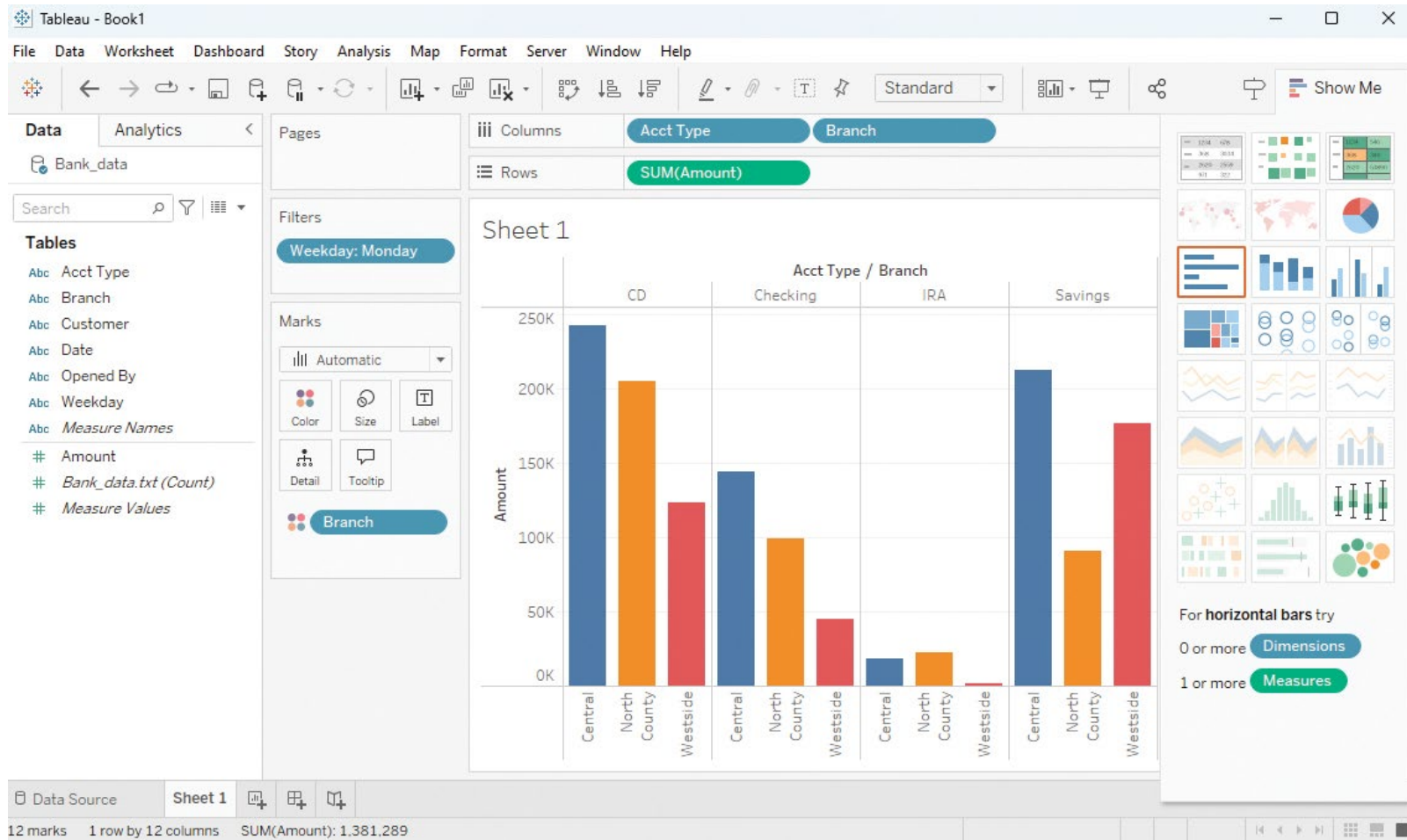


Figure 1.24 Coloring by Branch.

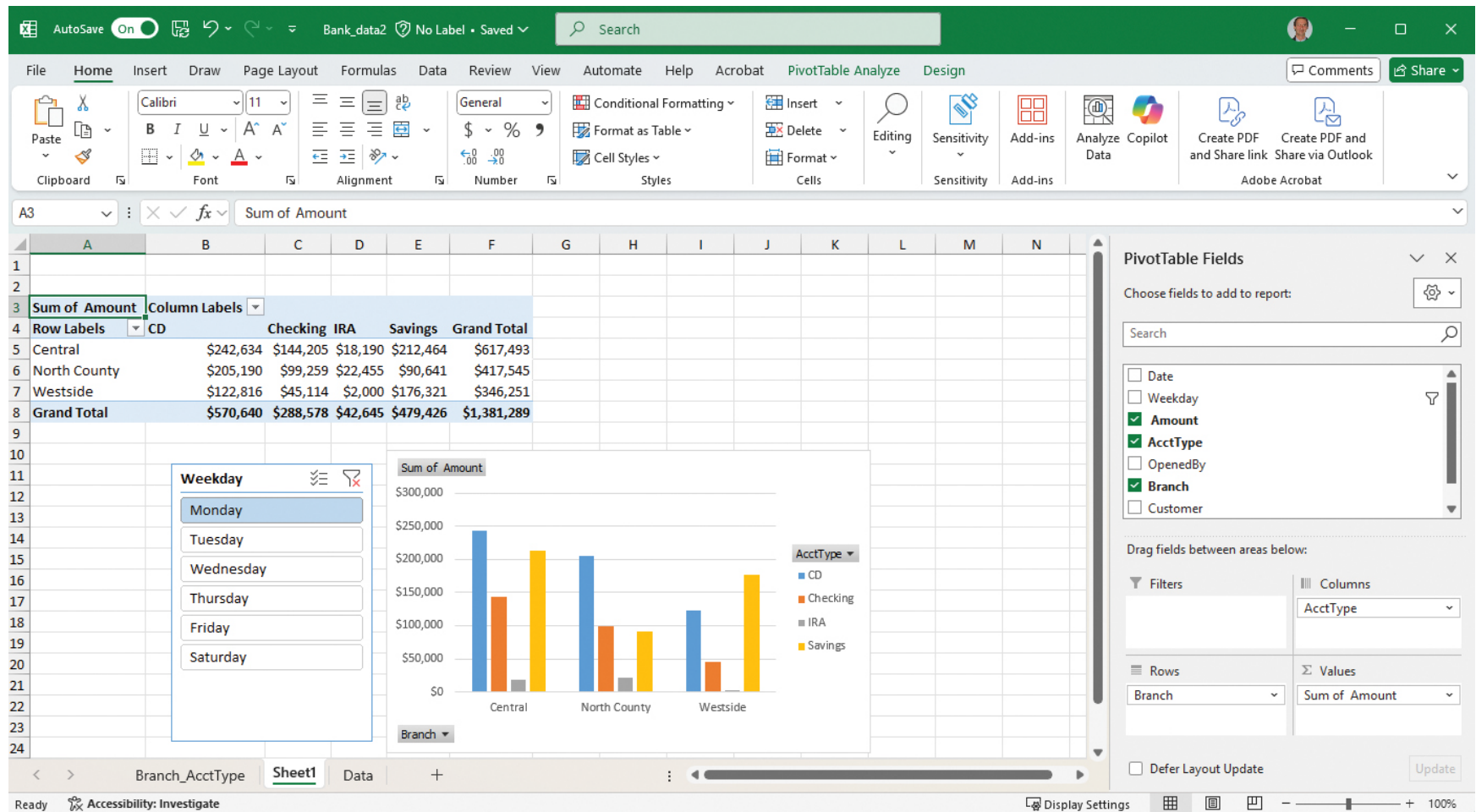


Figure 1.25 Column chart showing Amount by Branch, then Account Type.

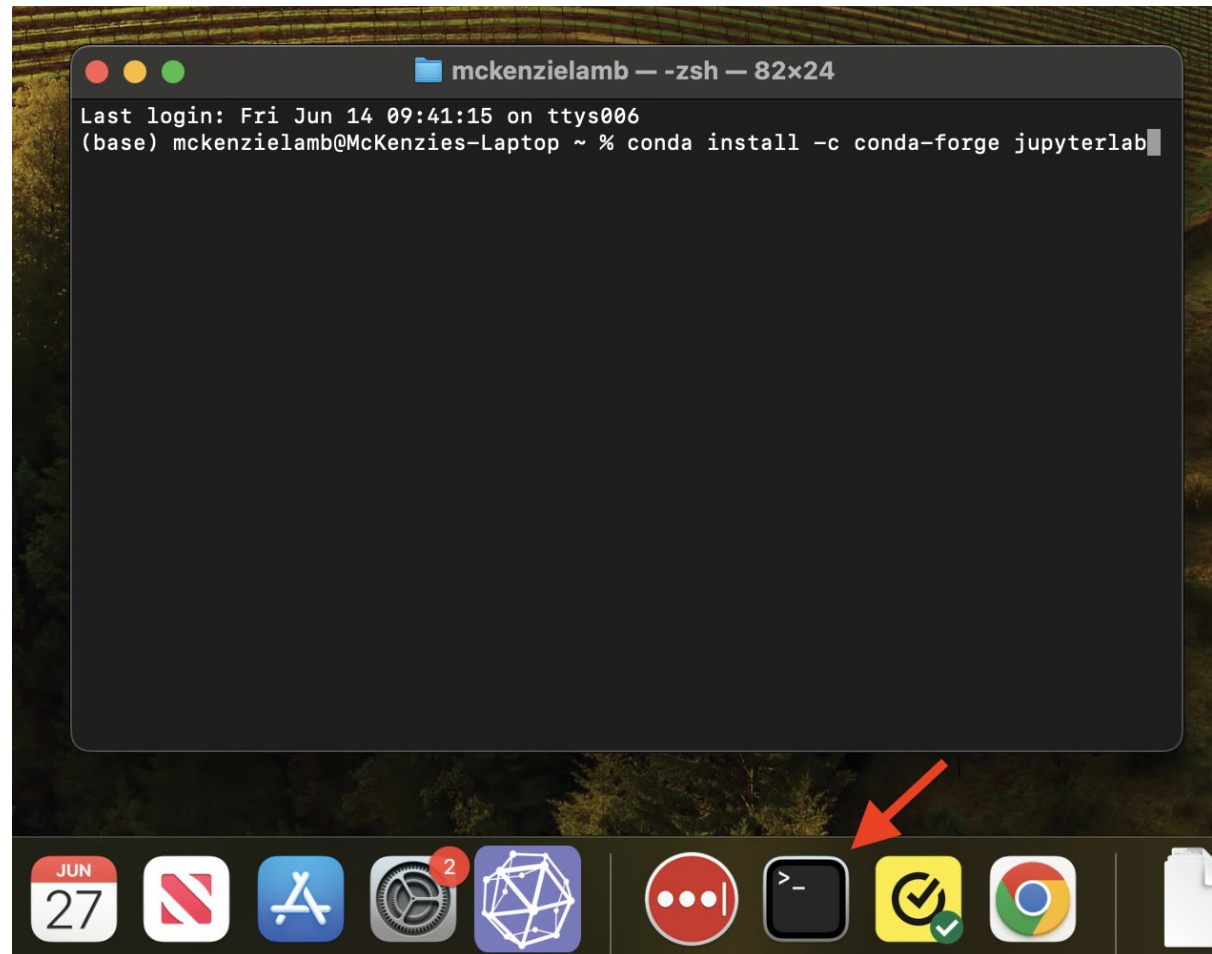


Figure 1.26 The terminal window.

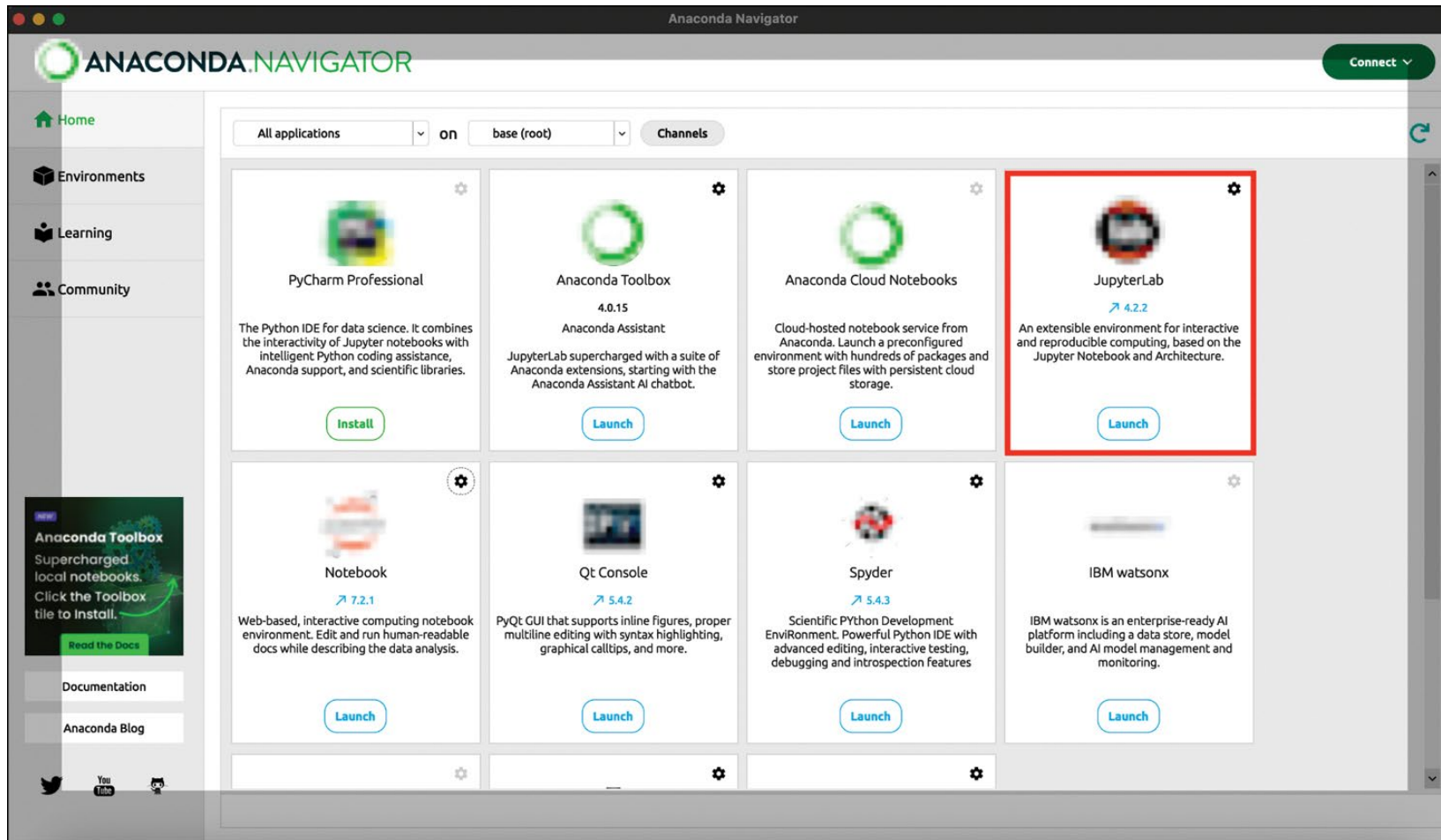


Figure 1.27 Click the Jupyter Lab button.

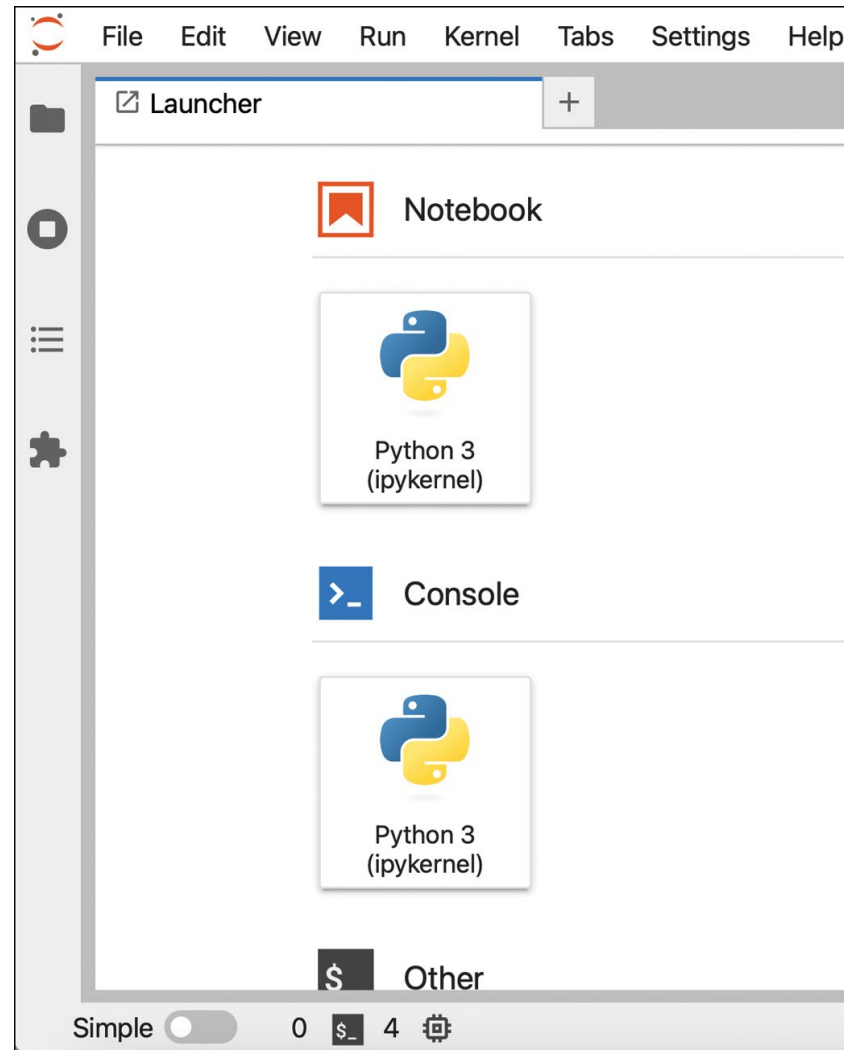


Figure 1.28 The Jupyter Lab interface.

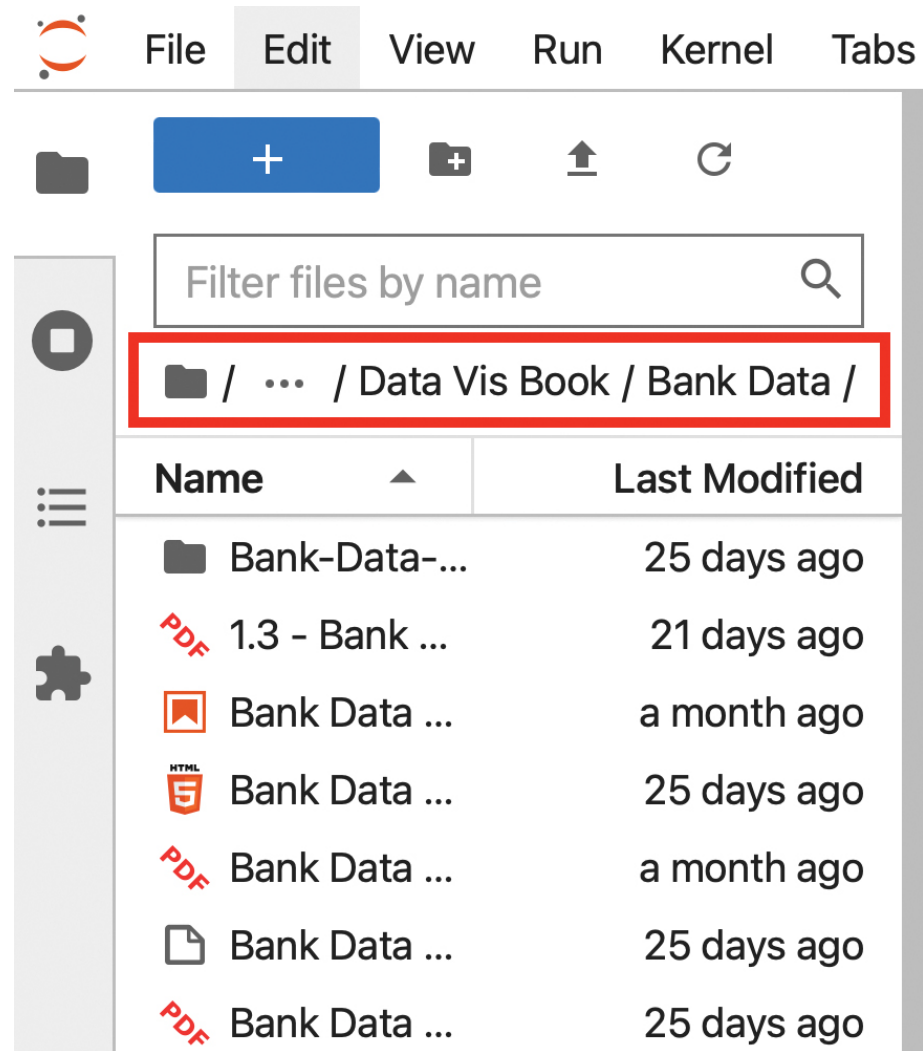


Figure 1.29 Navigating folders.

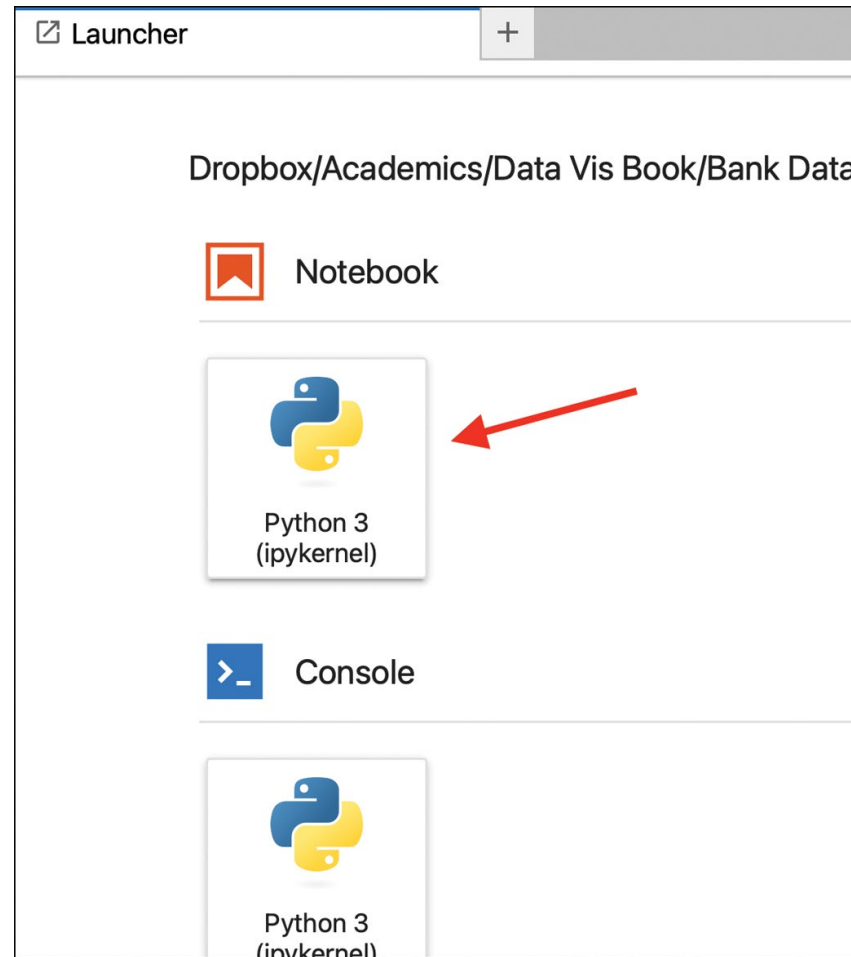


Figure 1.30 Choosing a kernel.

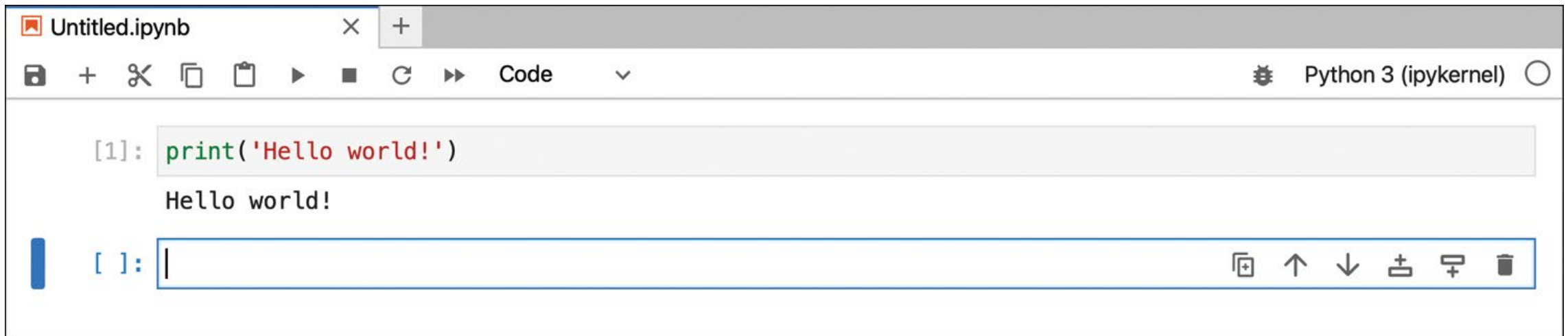


Figure 1.31 Output is displayed below the code cell.

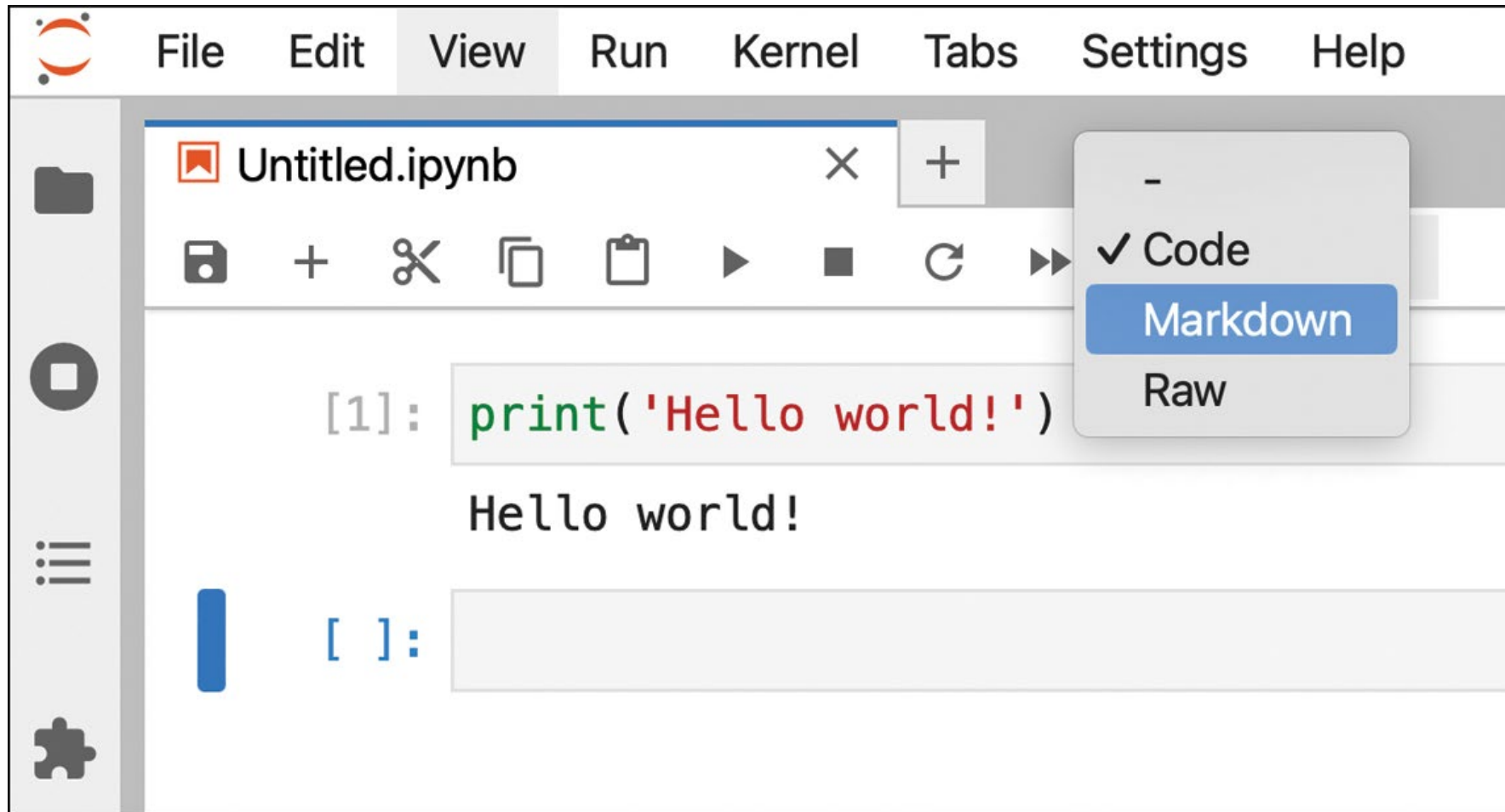


Figure 1.32 A Markdown cell.

Integration

All you really need to know about integration is that $\int_0^1 x^2 dx = \frac{1}{3}$.

Figure 1.33 Markdown cell with mathematical symbols.

Integration

All you really need to know about integration is that $\int_0^1 x^2 dx = \frac{1}{3}$.

Figure 1.34 The typeset output of the code.

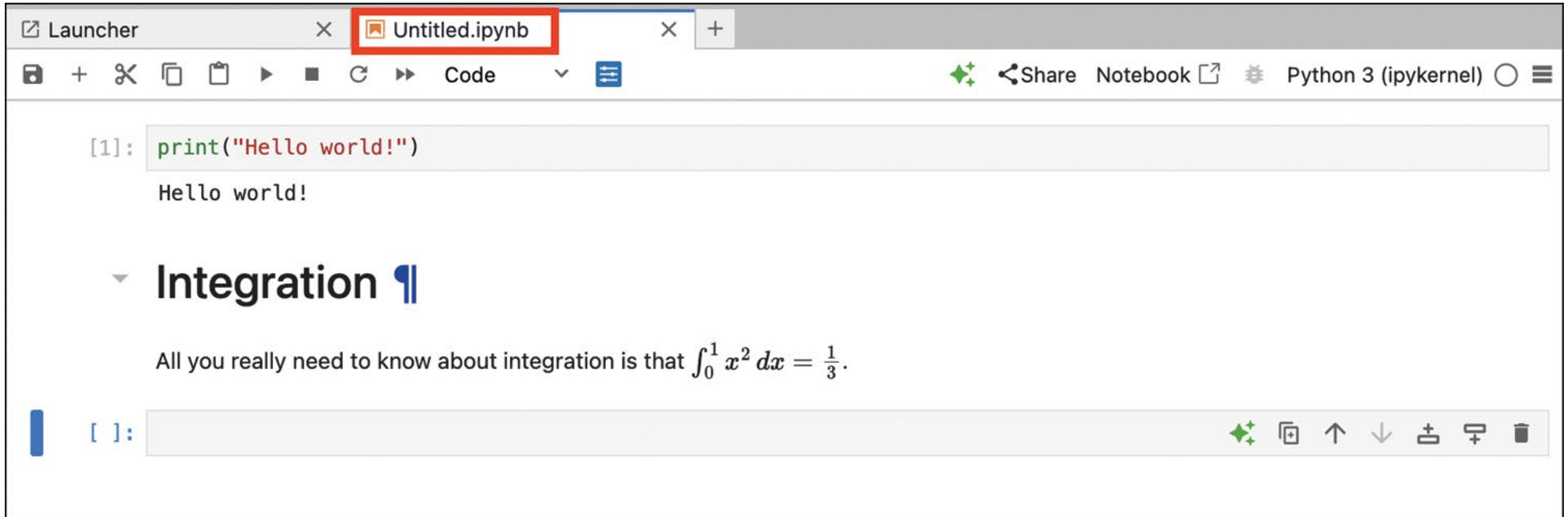


Figure 1.35 The notebook file name is shown at the top of the tab.

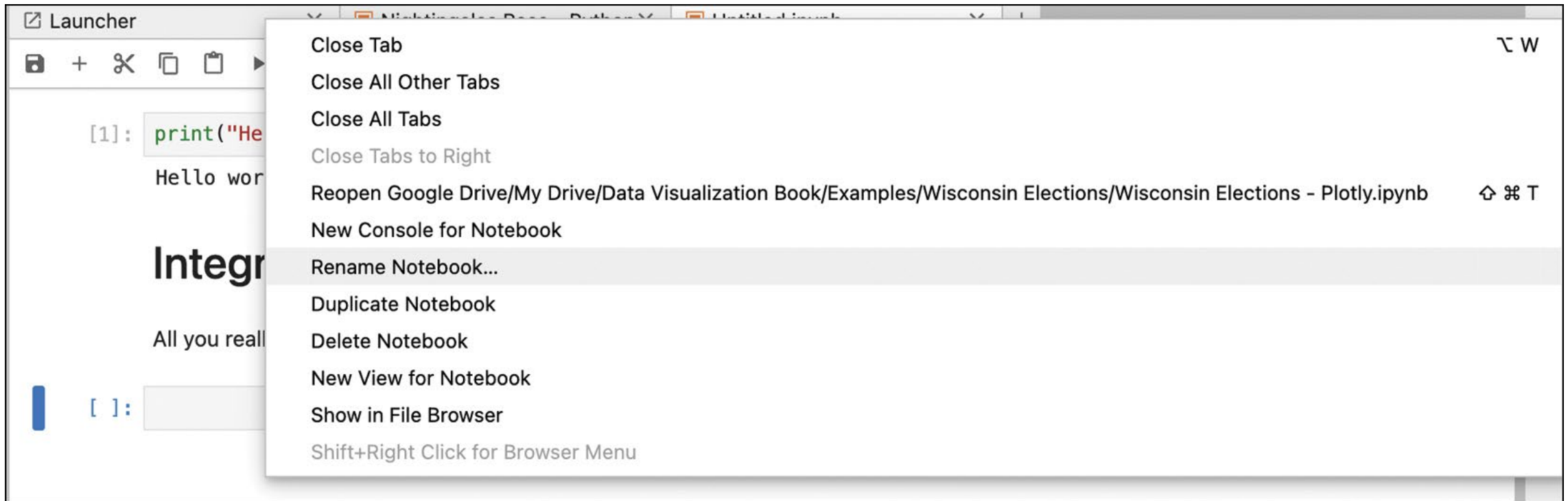
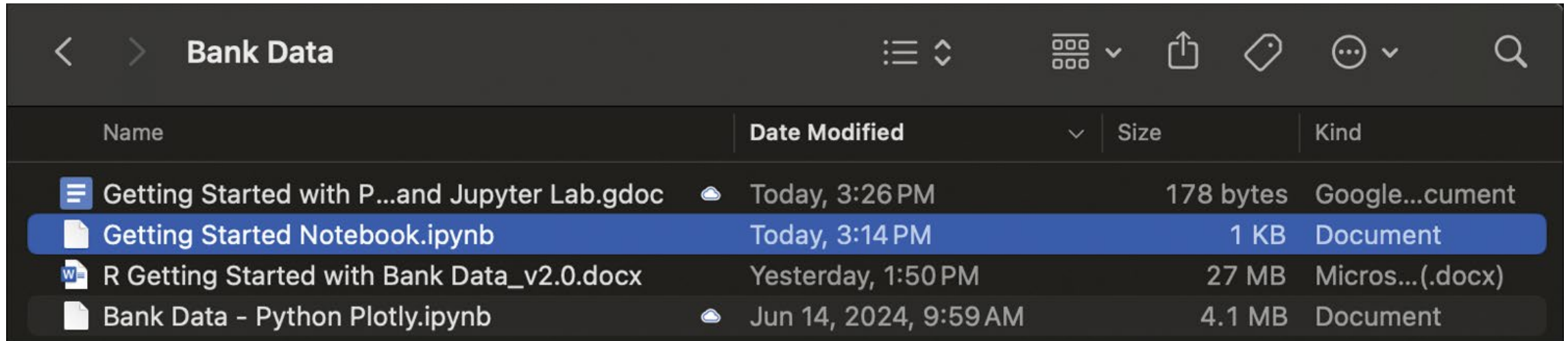


Figure 1.36 The Rename Notebook option.



Bank Data				
Name	Date Modified	Size	Kind	
 Getting Started with P...and Jupyter Lab.gdoc	Today, 3:26 PM	178 bytes	Google...cument	
 Getting Started Notebook.ipynb	Today, 3:14 PM	1 KB	Document	
 R Getting Started with Bank Data_v2.0.docx	Yesterday, 1:50 PM	27 MB	Micros...(.docx)	
 Bank Data - Python Plotly.ipynb	Jun 14, 2024, 9:59 AM	4.1 MB	Document	

Figure 1.37 The file in the Bank Data folder.

```
[7]: import os
os.getcwd()

[7]: '/Users/mckenzielamb/Library/CloudStorage/GoogleDrive-mckenzielamb@depauw.edu/My Drive/Data Visualization Book/Examples/Bank Data'
```

A horizontal toolbar with icons for running code (green play button), saving (floppy disk), undo (curved arrow left), redo (curved arrow right), copy (two overlapping squares), and delete (trash can).

Figure 1.38 The `os.getcwd()` command result.


```
[9]: import pandas as pd
df = pd.read_csv('Bank_data.csv')
df
```

```
[9]:
```

	Date	Weekday	Amount	AcctType	OpenedBy	Branch	Customer
0	1-Nov	Friday	5,000	IRA	New Accts	Central	Existing
1	1-Nov	Friday	14,571	CD	Teller	Central	New
2	1-Nov	Friday	500	Checking	New Accts	Central	Existing
3	1-Nov	Friday	15,000	CD	New Accts	Central	Existing
4	1-Nov	Friday	4,623	Savings	New Accts	North County	Existing
...	...	...	...	...	...	...	...
707	30-Nov	Saturday	7,000	IRA	New Accts	North County	Existing
708	30-Nov	Saturday	4,257	Savings	New Accts	Westside	Existing
709	30-Nov	Saturday	400	Checking	Teller	Central	Existing
710	30-Nov	Saturday	12,673	Checking	New Accts	Westside	New
711	30-Nov	Saturday	4,000	Checking	Teller	North County	Existing

712 rows x 7 columns

Figure 1.39 The command to read the Bank_data.csv file.

```
mckenzielamb — python3.11 < python3.11 /opt/homebrew/anaconda3/bin/ju...
s/__init__.py", line 189, in ensure_async
    result = await obj
            ^^^^^^^^^
File "/opt/homebrew/anaconda3/lib/python3.11/site-packages/jupyter_server/se
rvices/contents/manager.py", line 902, in update
    await self.rename(path, new_path)
File "/opt/homebrew/anaconda3/lib/python3.11/site-packages/jupyter_server/se
rvices/contents/manager.py", line 889, in rename
    await self.rename_file(old_path, new_path)
File "/opt/homebrew/anaconda3/lib/python3.11/site-packages/jupyter_server/se
rvices/contents/filemanager.py", line 1015, in rename_file
    raise web.HTTPError(409, "File already exists: %s" % new_path)
tornado.web.HTTPError: HTTP 409: Conflict (File already exists: Google Drive/M
y Drive/Data Visualization Book/Examples/Bank Data/Untitled.ipynb)
[W 2024-06-27 15:19:07.835 ServerApp] 409 PATCH /api/contents/Google%20Drive/My%20
Drive/Data%20Visualization%20Book/Examples/Bank%20Data/Getting%20Started%20Notebo
ok.ipynb?1719515947828 (b555e8f814724f209ded39d234e59d5e@::1) 4.16ms referer=http:/
/localhost:8889/lab/tree/Google%20Drive/My%20Drive/Data%20Visualization%20Book/Exa
mples/Bank%20Data/Getting%20Started%20Notebook.ipynb?
[I 2024-06-27 15:19:14.830 ServerApp] Kernel shutdown: 9a26161f-f901-48b9-b6a1-81e
37e88e728
[W 2024-06-27 15:19:15.066 ServerApp] delete /Google Drive/My Drive/Data Visualiza
tion Book/Examples/Bank Data/Untitled.ipynb
```

Figure 1.40 The Jupyter terminal window remains open.

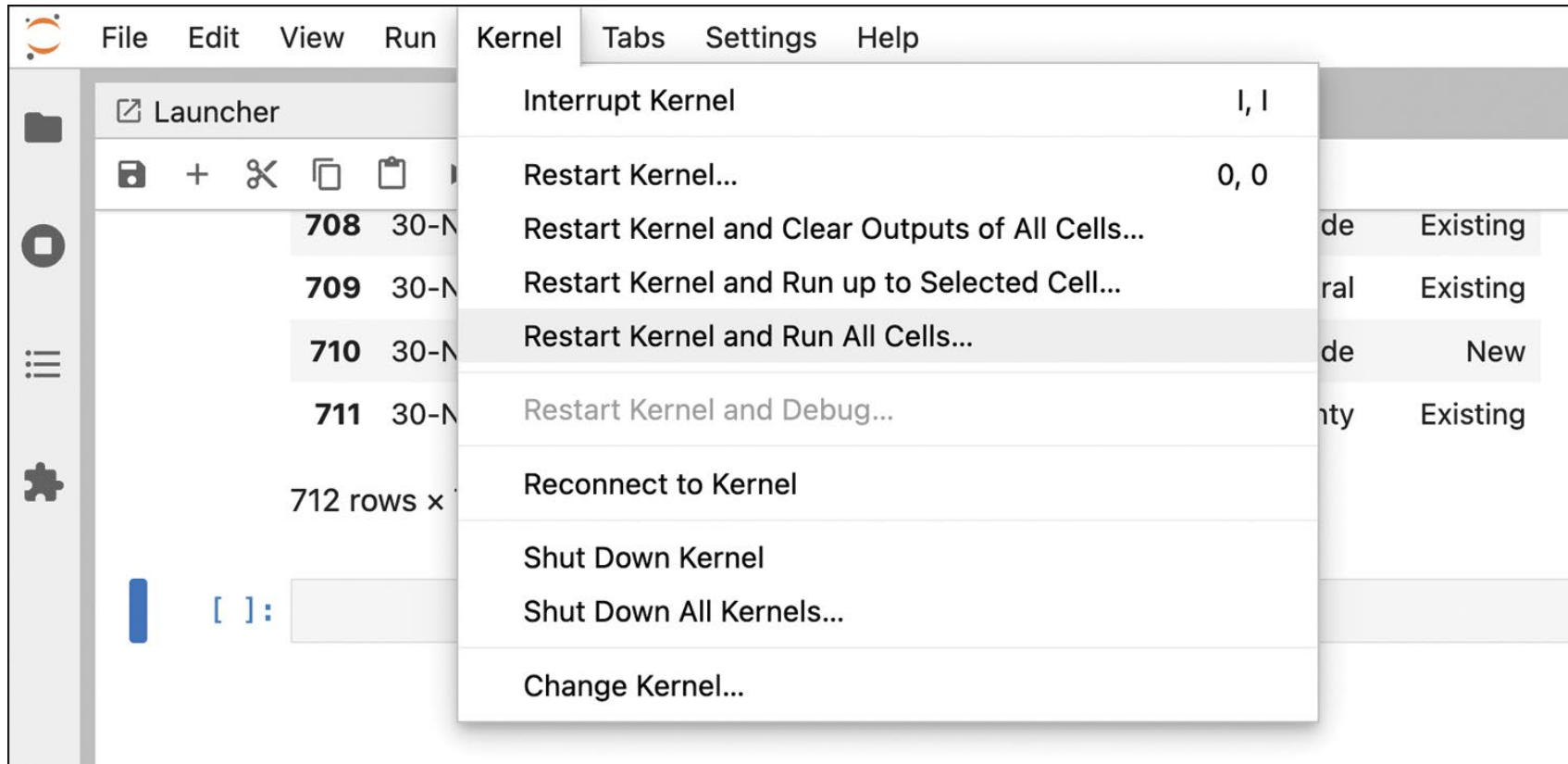


Figure 1.41 The Restart Kernel and Run All Cells option.

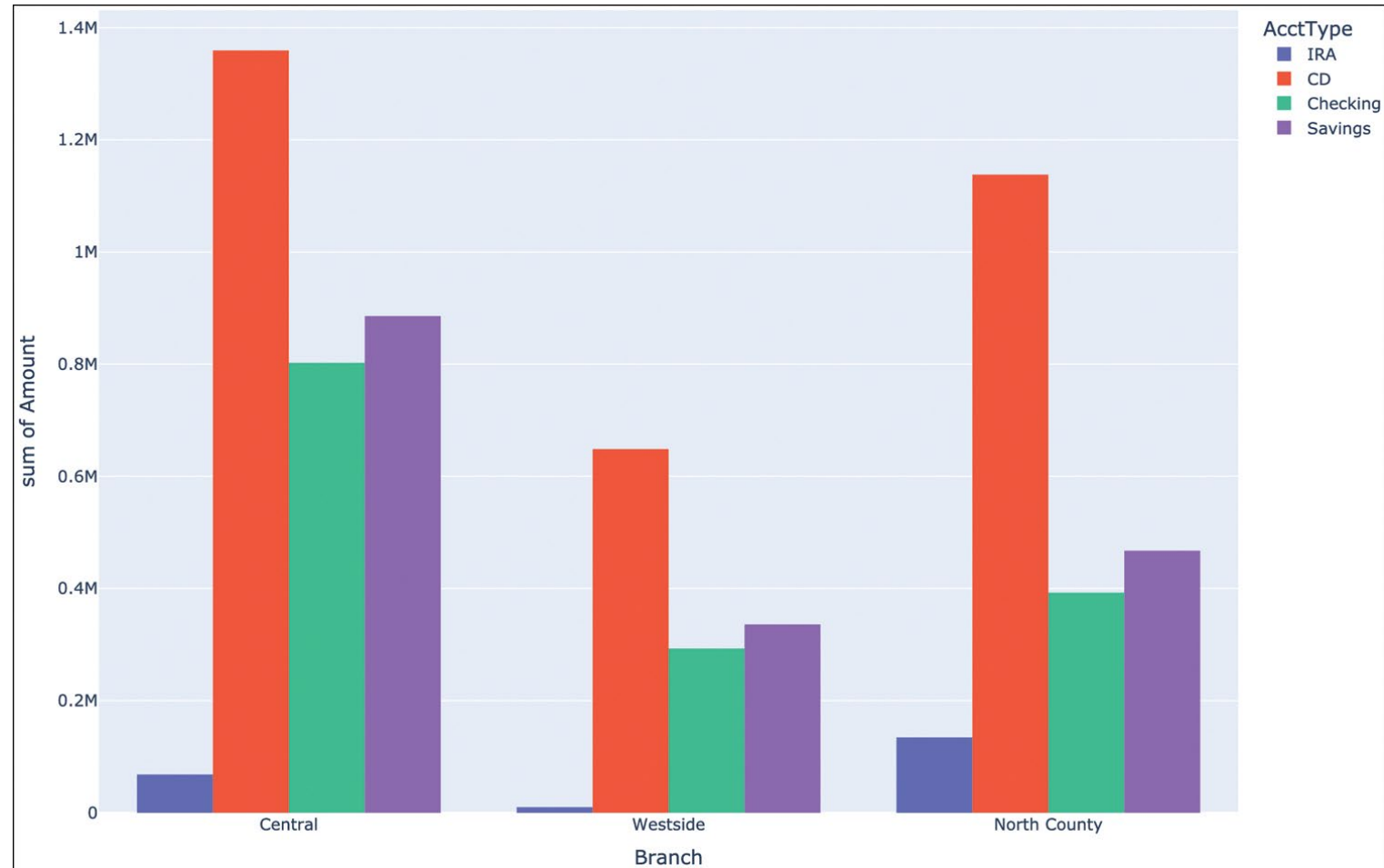


Figure 1.42 The plot result.

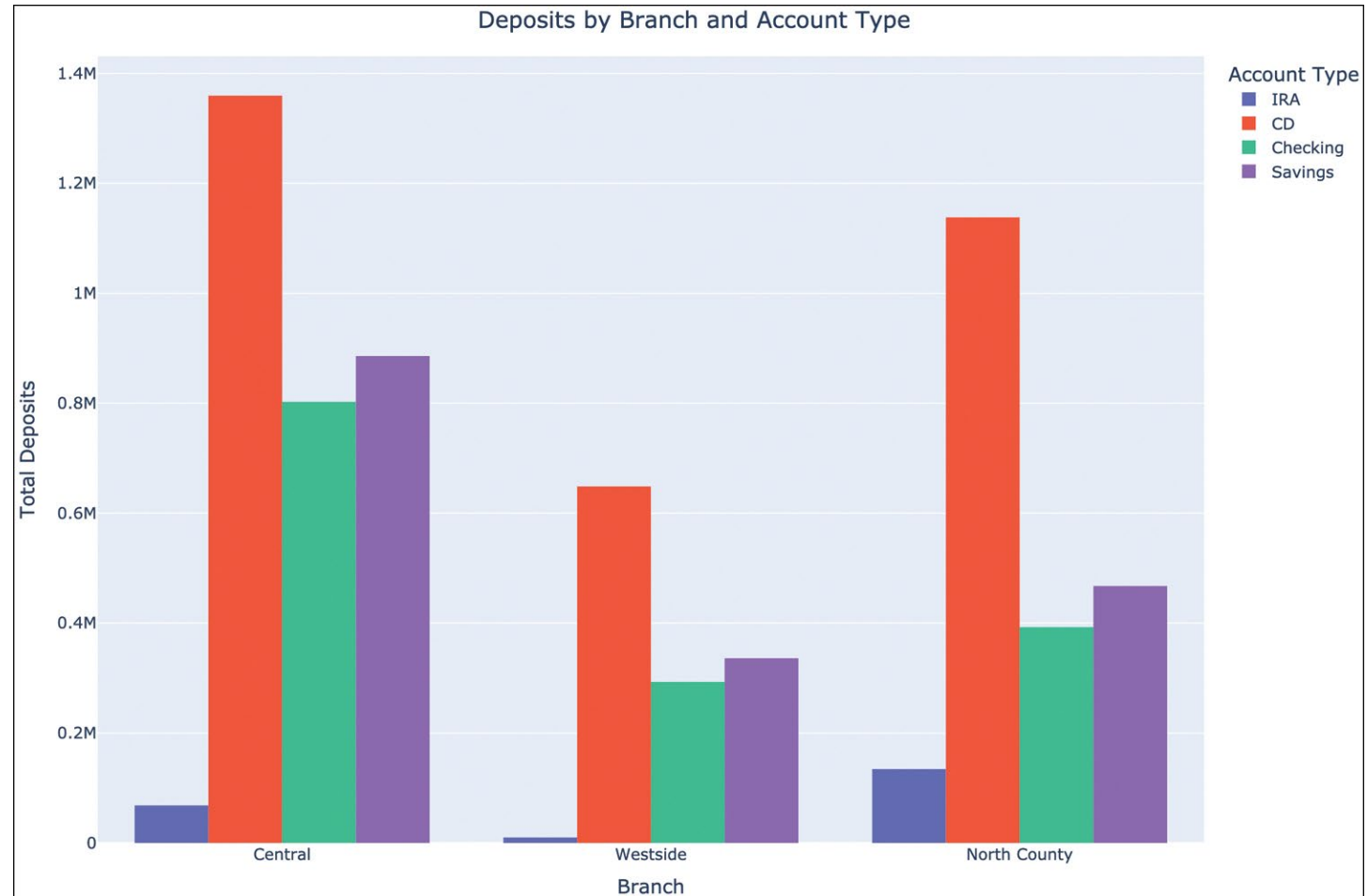


Figure 1.43 The fine-tuned plot.

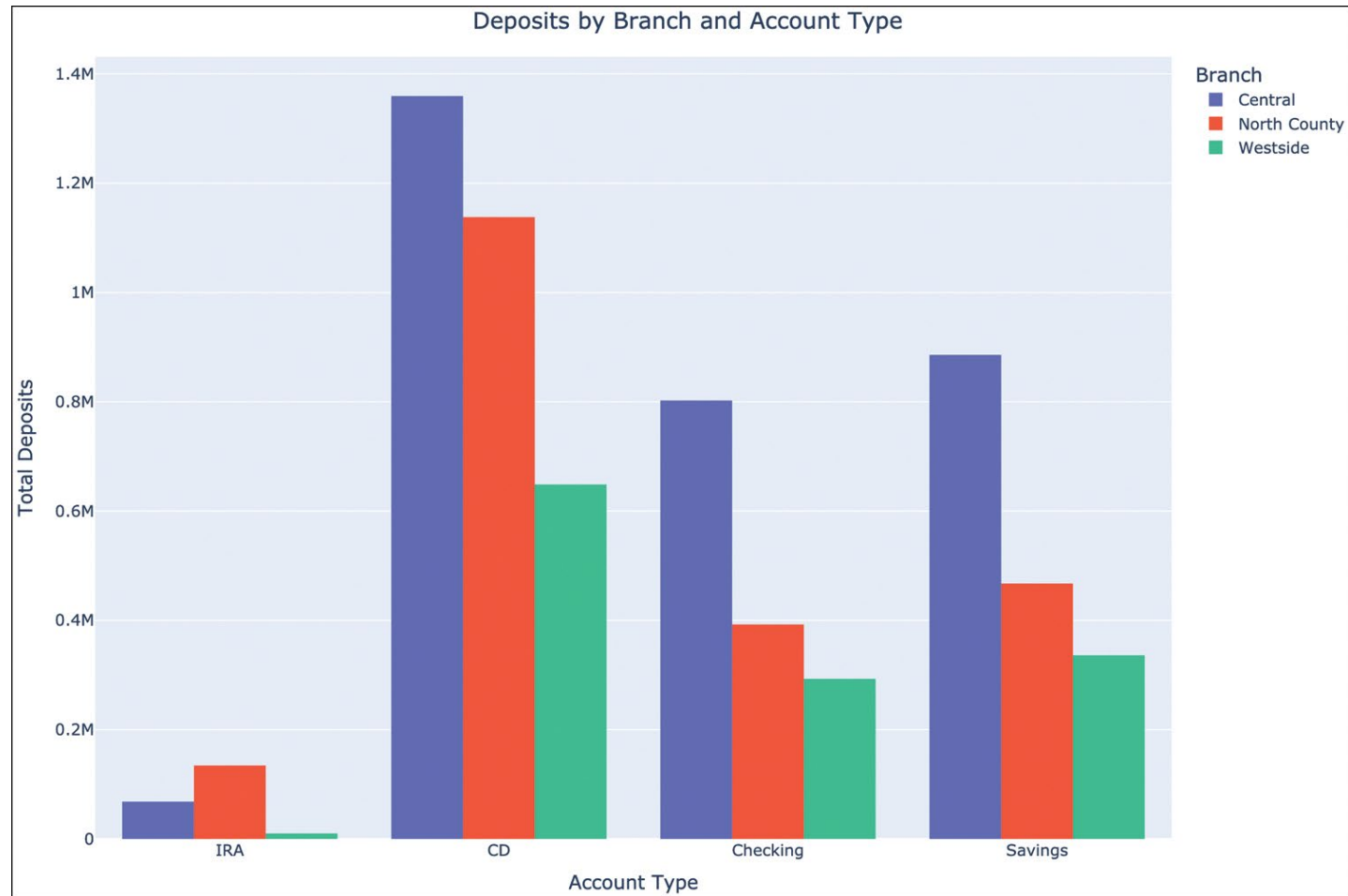


Figure 1.44 A plot organized by account type and branch.

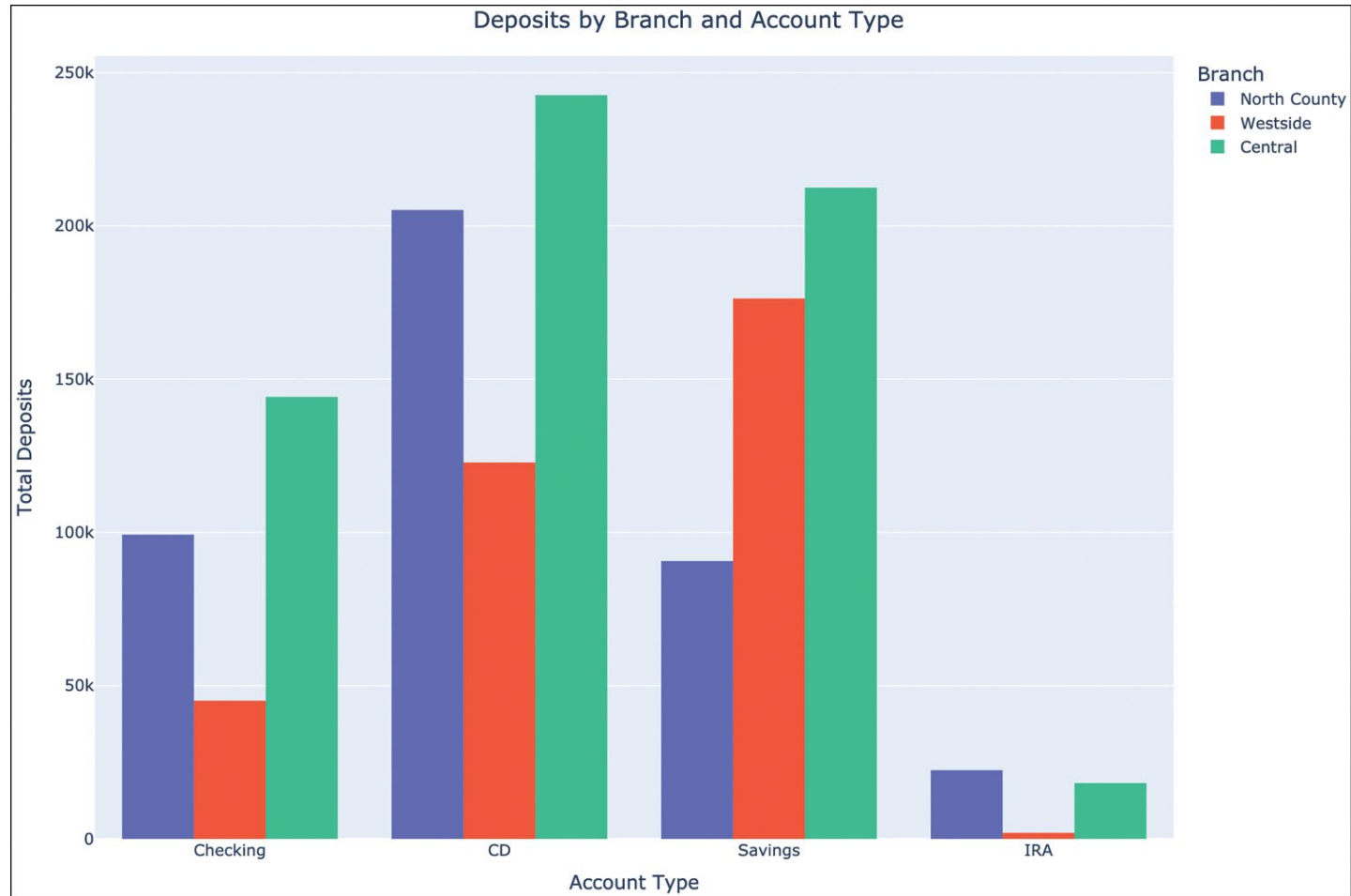


Figure 1.45 Only the Monday deposits.

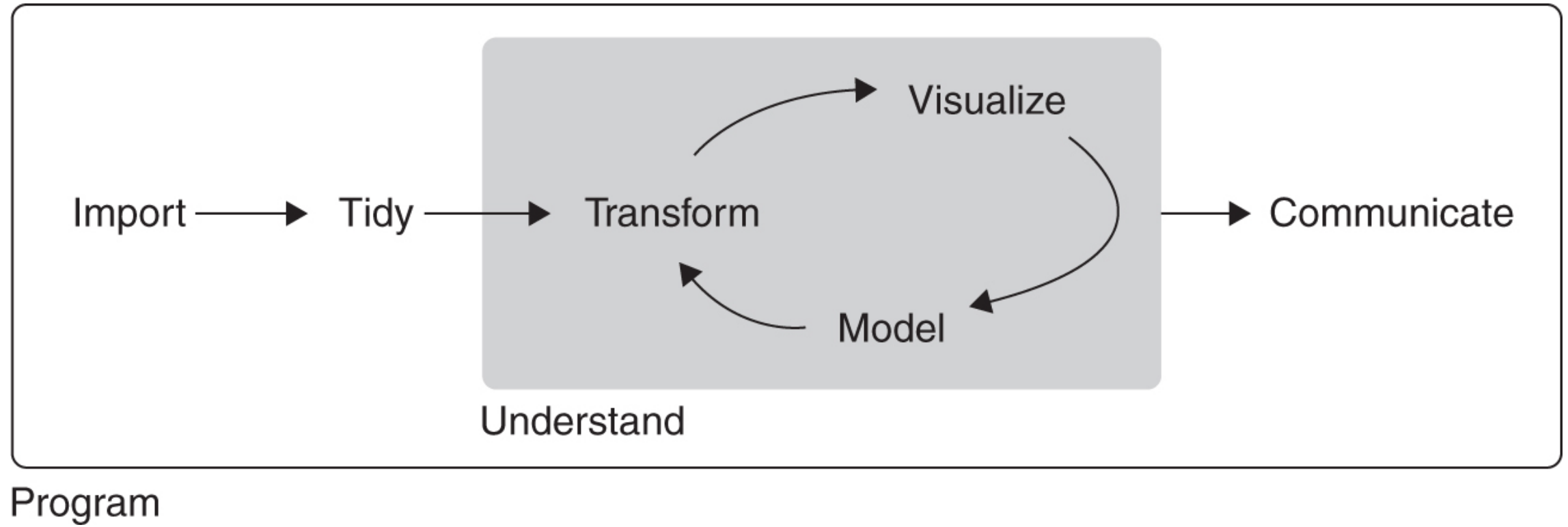


Figure 1.46 The data science process.



CRAN
[Mirrors](#)
[What's new?](#)
[Search](#)
[CRAN Team](#)

About R
[R Homepage](#)
[The R Journal](#)

Software
[R Sources](#)
[R Binaries](#)
[Packages](#)
[Task Views](#)
[Other](#)

Documentation
[Manuals](#)
[FAQs](#)

Donations
[Donate](#)

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux \(Debian, Fedora/Redhat, Ubuntu\)](#)
- [Download R for macOS](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2025-06-13, Great Square Root) [R-4.5.1.tar.gz](#), read [what's new](#) in the latest version.
- The CRAN directory [src/base-prerelease](#) contains R alpha, beta, and rc releases as daily snapshots in time periods before a planned release.
- Between releases, the same directory [src/base-prerelease](#) contains snapshots of current patched and development versions. Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Alternatively, daily snapshots are [available here](#).
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#).

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

Copyright © 2025 CRAN

Figure 1.47 The CRAN home page.

R for Windows

Subdirectories:

base	Binaries for base distribution. This is what you want to install R for the first time .
contrib	Binaries of contributed CRAN packages (for R $\geq$ 4.0.x).
old contrib	Binaries of contributed CRAN packages for outdated versions of R (for R $<$ 4.0.x).
Rtools	Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

Figure 1.48 Click the download for your computer/operating system type.

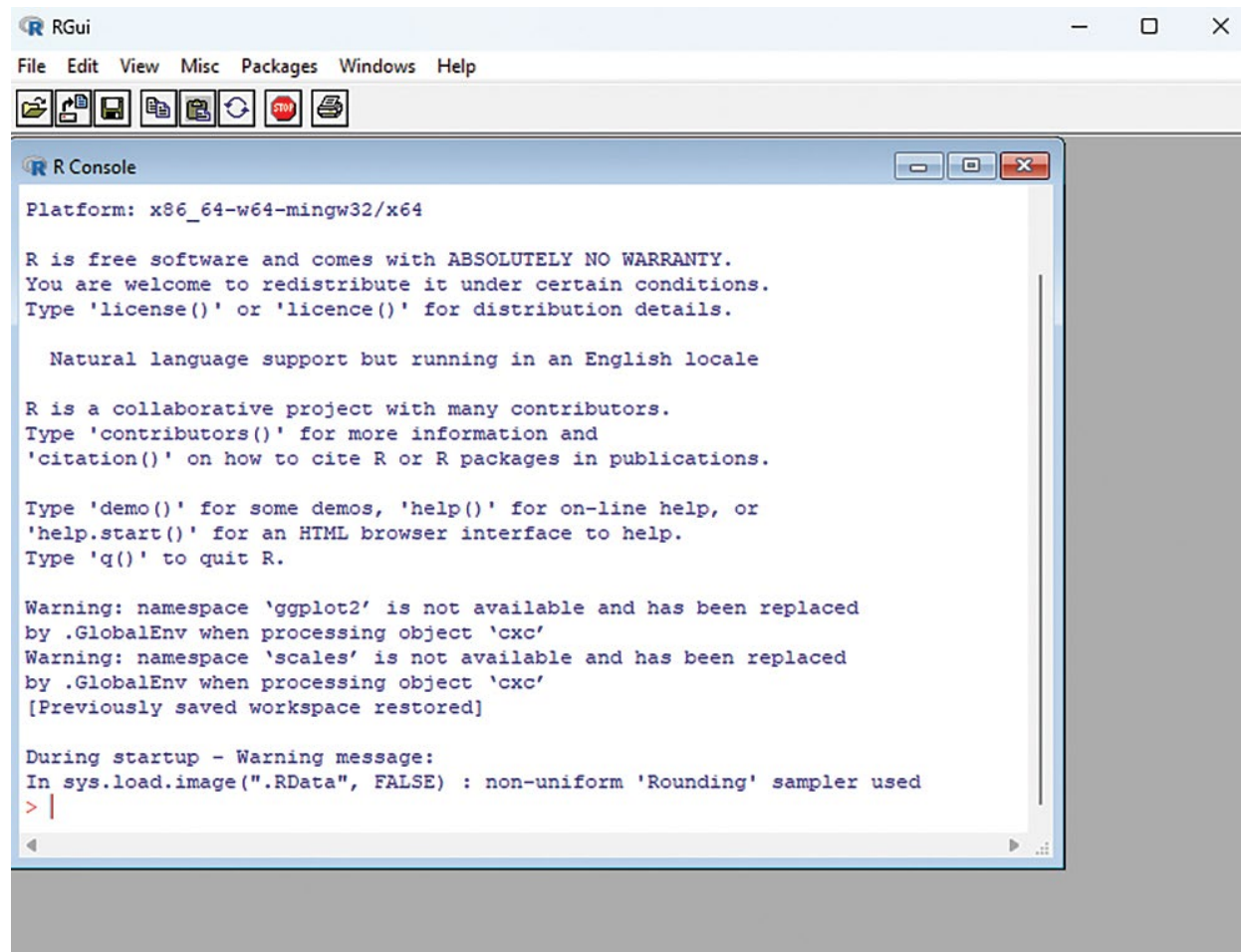


Figure 1.49 The R console.

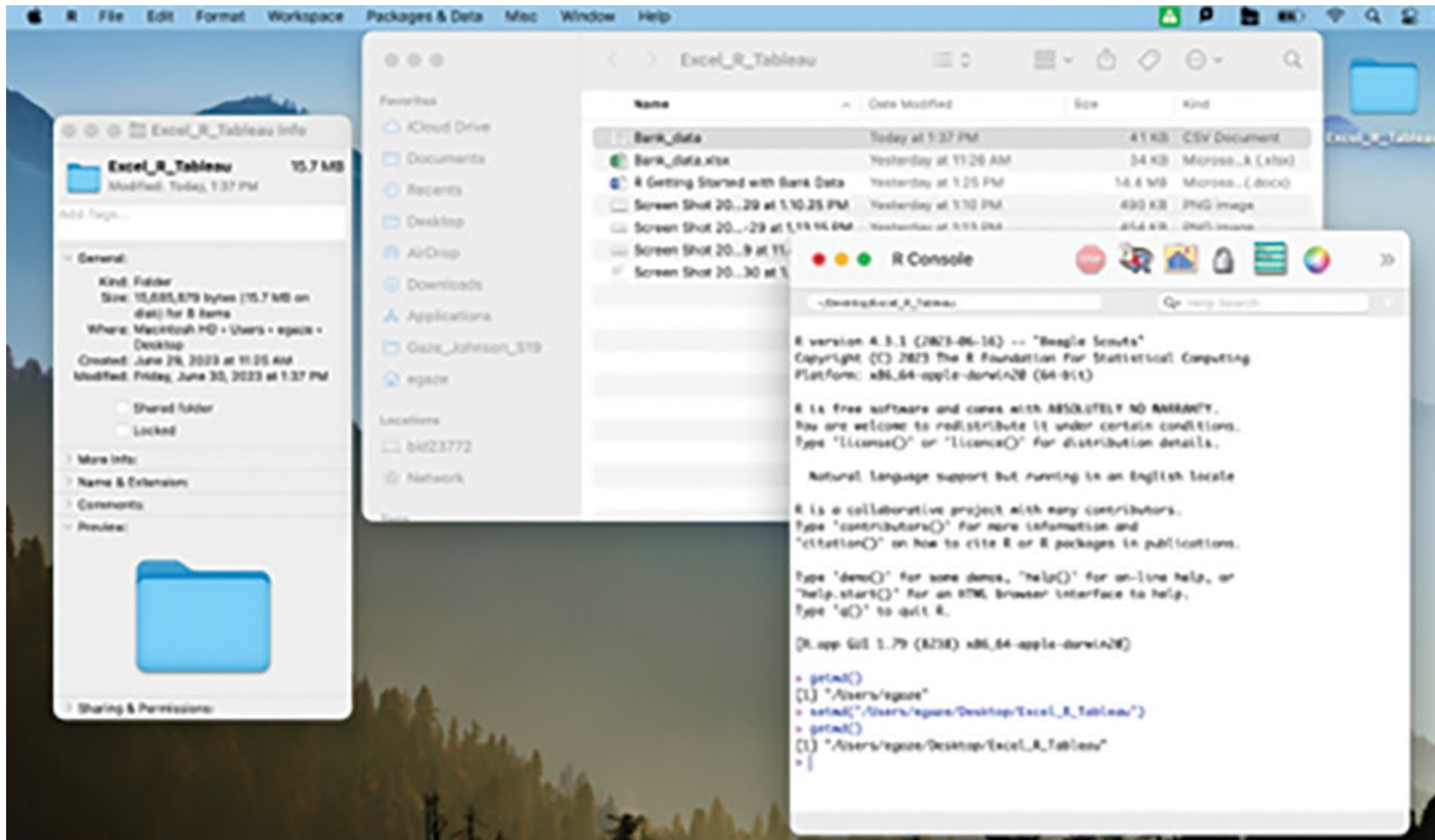
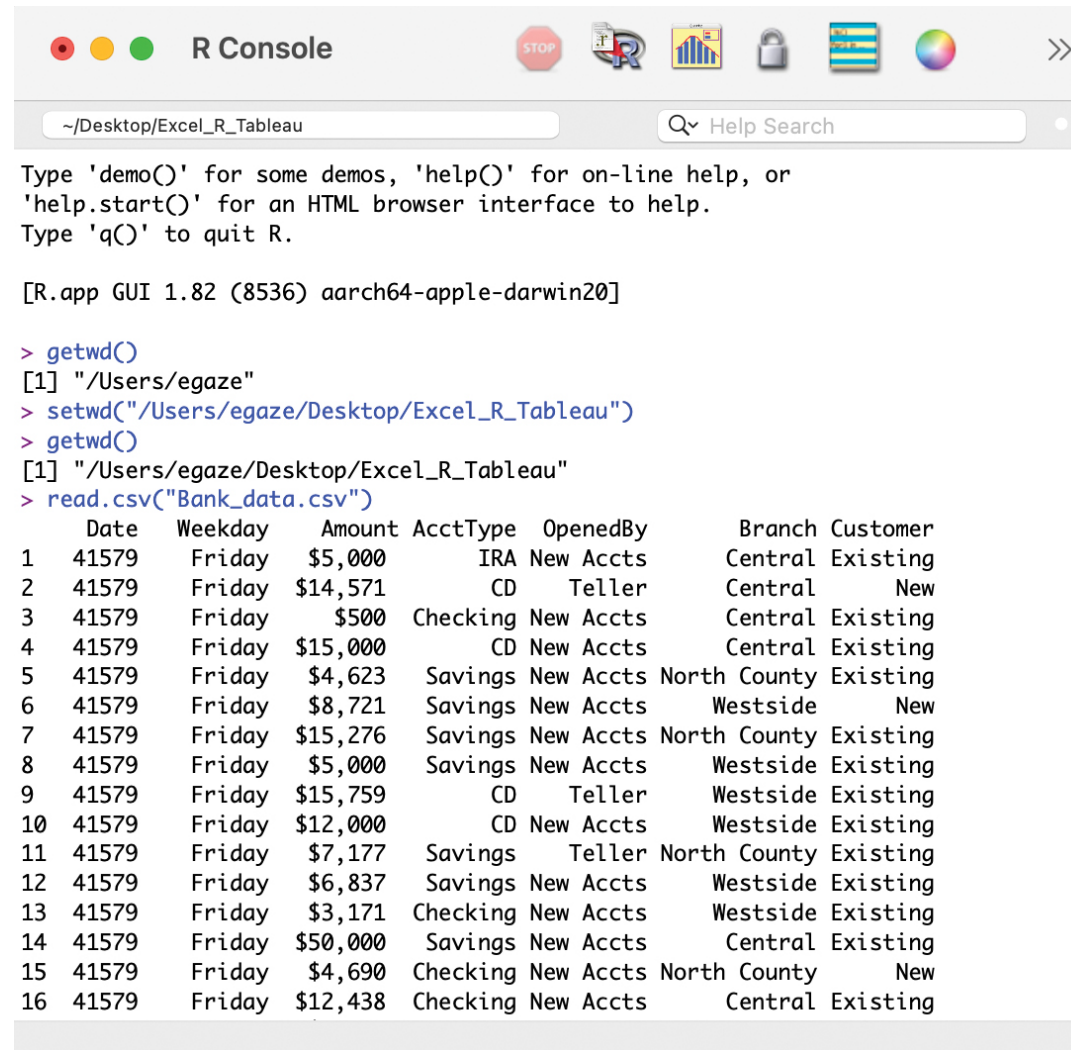


Figure 1.50 Using the setwd() function to change the working directory.



The screenshot shows the R Console window with a title bar containing window control buttons and the text 'R Console'. Below the title bar is a search bar with the text 'Help Search'. The main area of the console displays the following text:

```
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.  
  
[R.app GUI 1.82 (8536) aarch64-apple-darwin20]  
  
> getwd()  
[1] "/Users/egaze"  
> setwd("/Users/egaze/Desktop/Excel_R_Tableau")  
> getwd()  
[1] "/Users/egaze/Desktop/Excel_R_Tableau"  
> read.csv("Bank_data.csv")
```

	Date	Weekday	Amount	AcctType	OpenedBy	Branch	Customer
1	41579	Friday	\$5,000	IRA	New Accts	Central	Existing
2	41579	Friday	\$14,571	CD	Teller	Central	New
3	41579	Friday	\$500	Checking	New Accts	Central	Existing
4	41579	Friday	\$15,000	CD	New Accts	Central	Existing
5	41579	Friday	\$4,623	Savings	New Accts	North County	Existing
6	41579	Friday	\$8,721	Savings	New Accts	Westside	New
7	41579	Friday	\$15,276	Savings	New Accts	North County	Existing
8	41579	Friday	\$5,000	Savings	New Accts	Westside	Existing
9	41579	Friday	\$15,759	CD	Teller	Westside	Existing
10	41579	Friday	\$12,000	CD	New Accts	Westside	Existing
11	41579	Friday	\$7,177	Savings	Teller	North County	Existing
12	41579	Friday	\$6,837	Savings	New Accts	Westside	Existing
13	41579	Friday	\$3,171	Checking	New Accts	Westside	Existing
14	41579	Friday	\$50,000	Savings	New Accts	Central	Existing
15	41579	Friday	\$4,690	Checking	New Accts	North County	New
16	41579	Friday	\$12,438	Checking	New Accts	Central	Existing

Figure 1.51 File output in the R console.

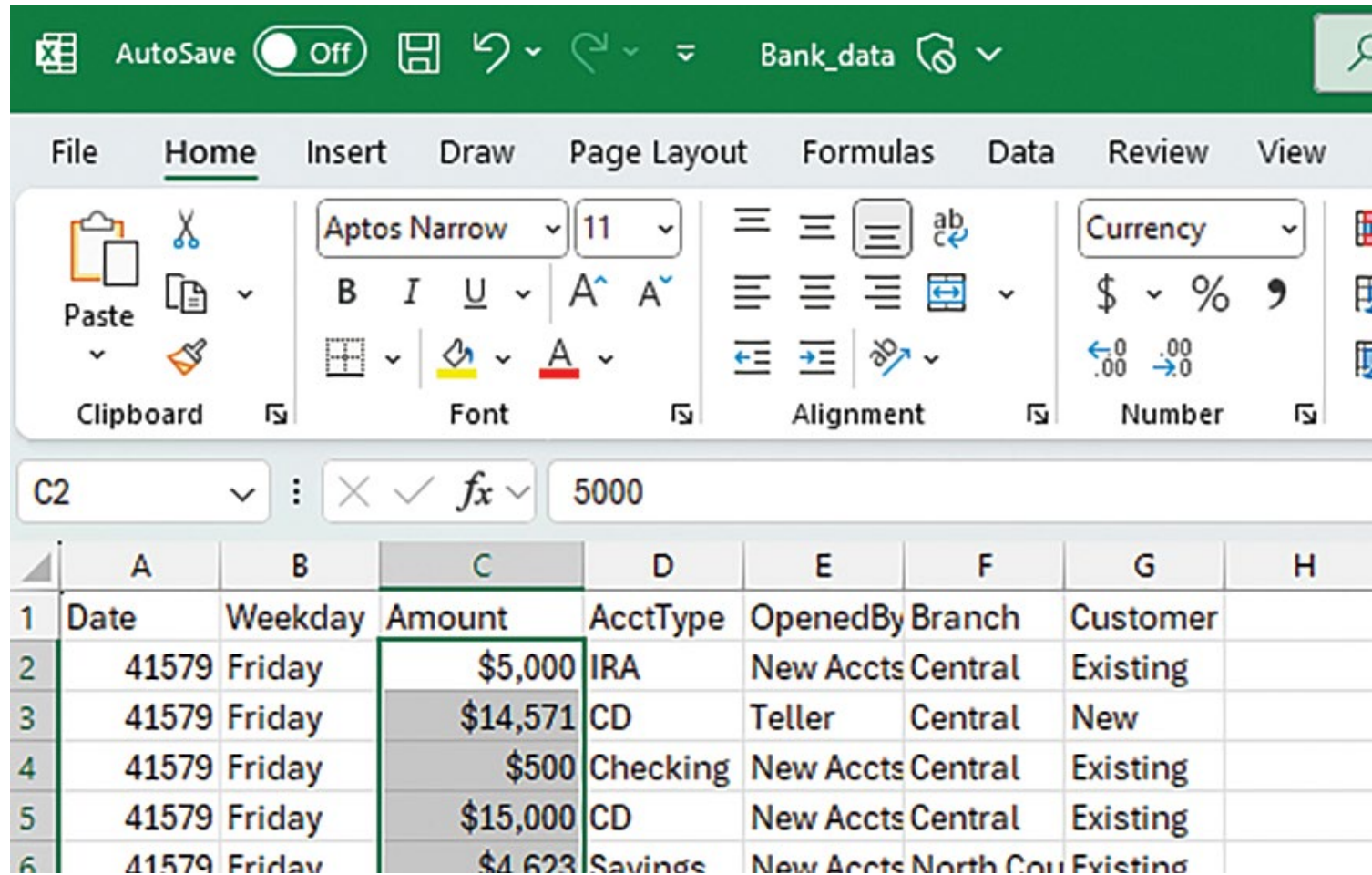


Figure 1.52 The Amount variable formatted as currency in the Excel file.

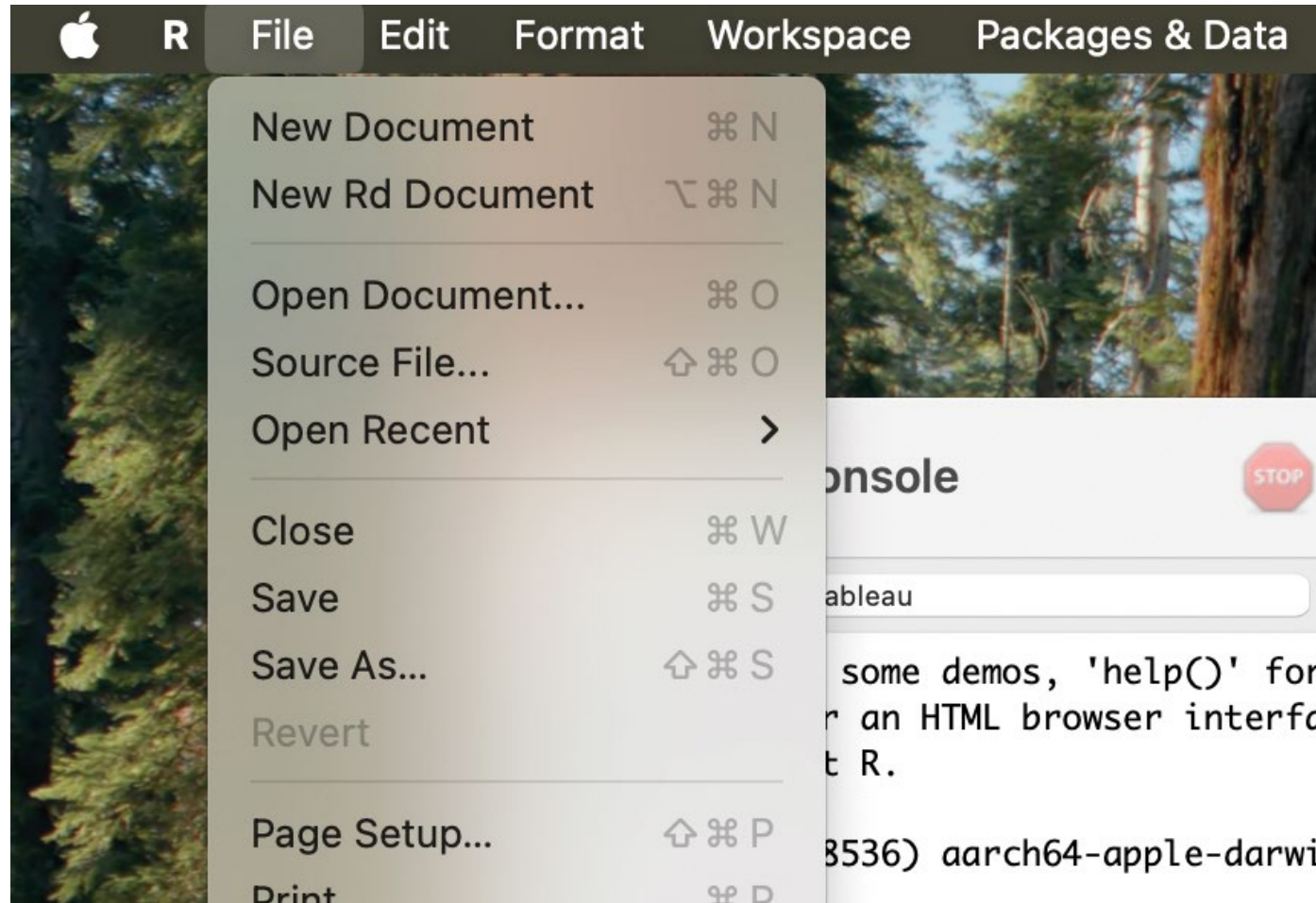
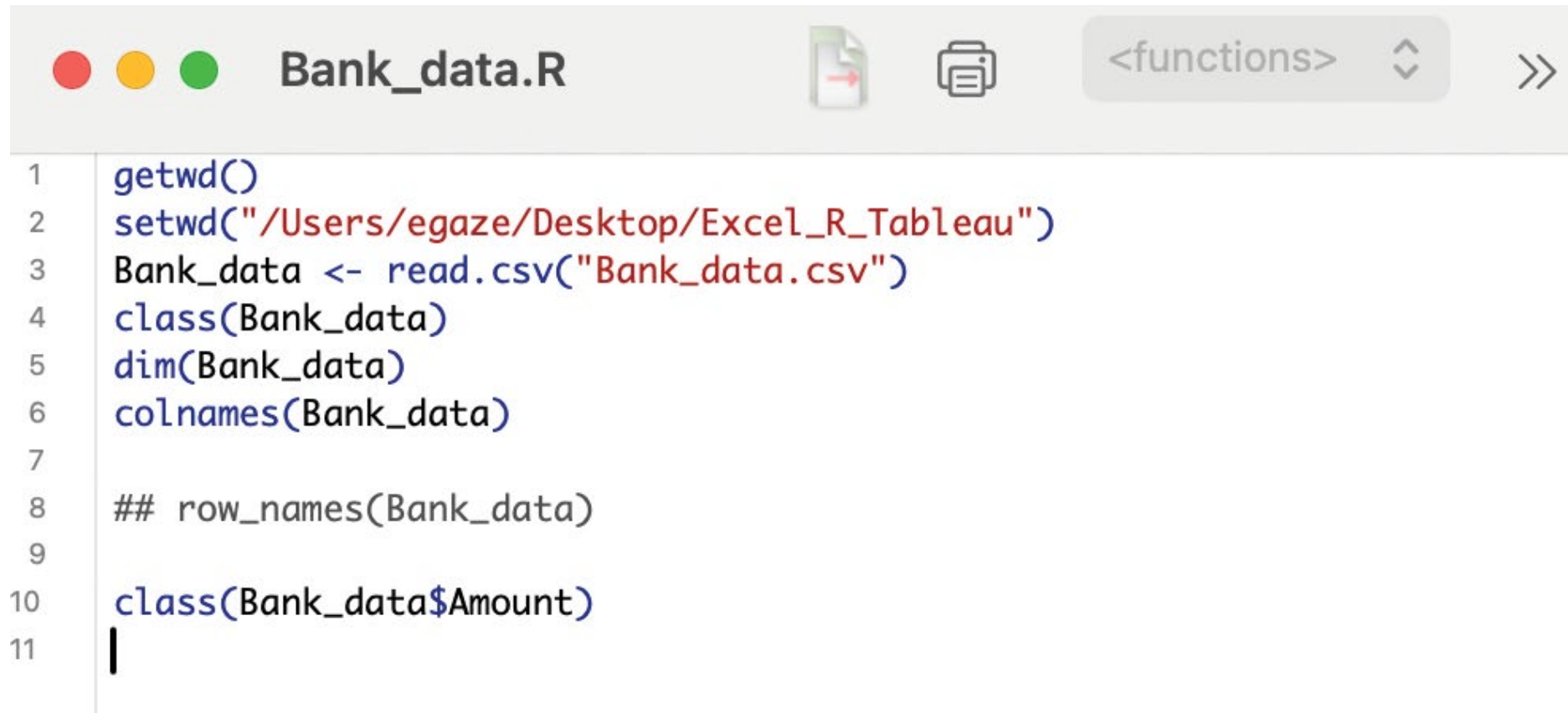


Figure 1.53 Creating a script in R using File → New Document.



```
1 getwd()
2 setwd("/Users/egaze/Desktop/Excel_R_Tableau")
3 Bank_data <- read.csv("Bank_data.csv")
4 class(Bank_data)
5 dim(Bank_data)
6 colnames(Bank_data)
7
8 ## row_names(Bank_data)
9
10 class(Bank_data$Amount)
11 |
```

Figure 1.54 Pasting R code from the console into a new R script document.

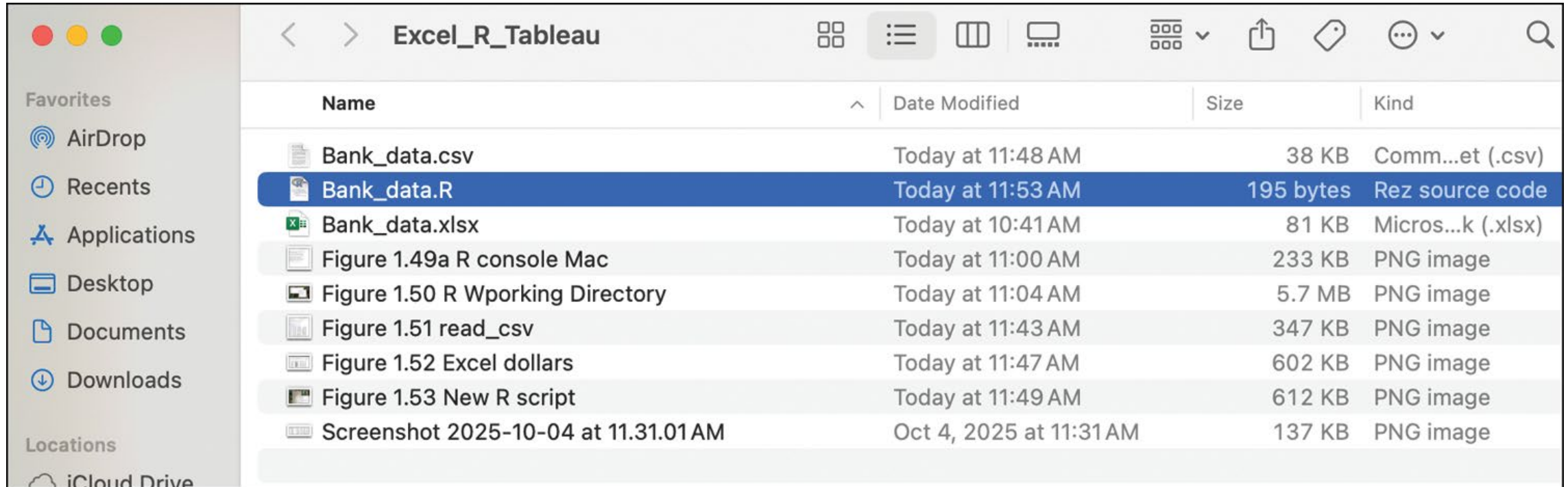


Figure 1.55 The R source file saved in a folder and listed below the CSV document.

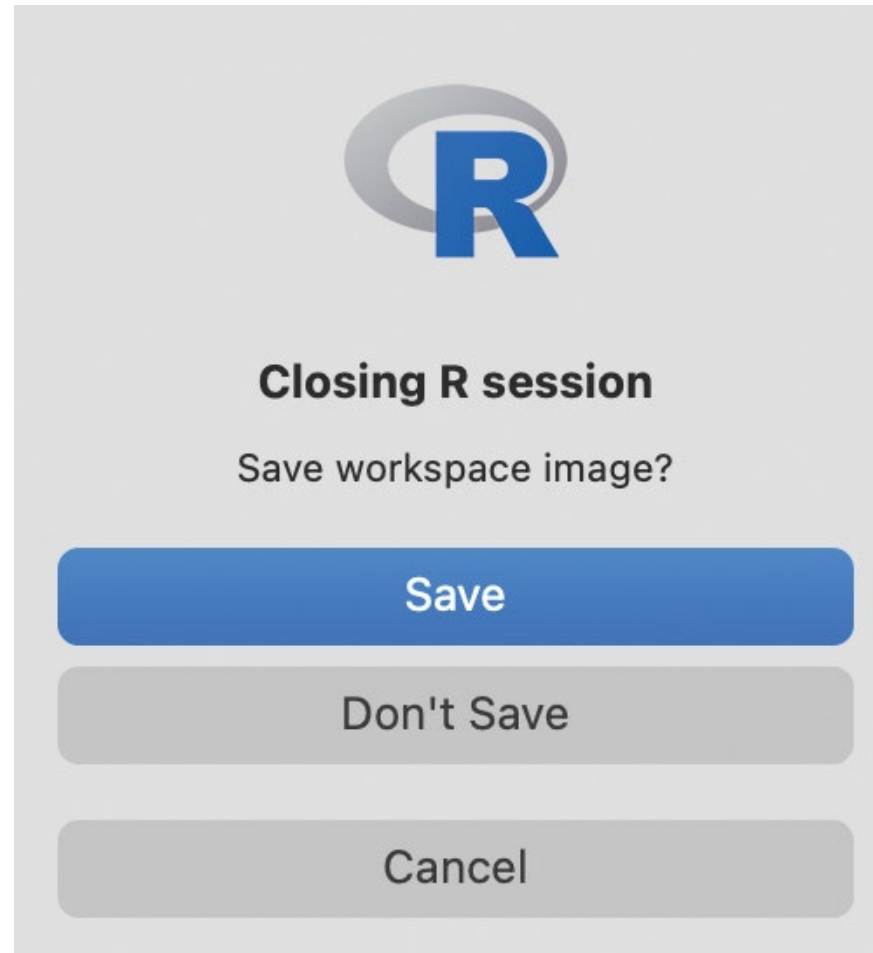


Figure 1.56 R asks if you want to save your workspace image.

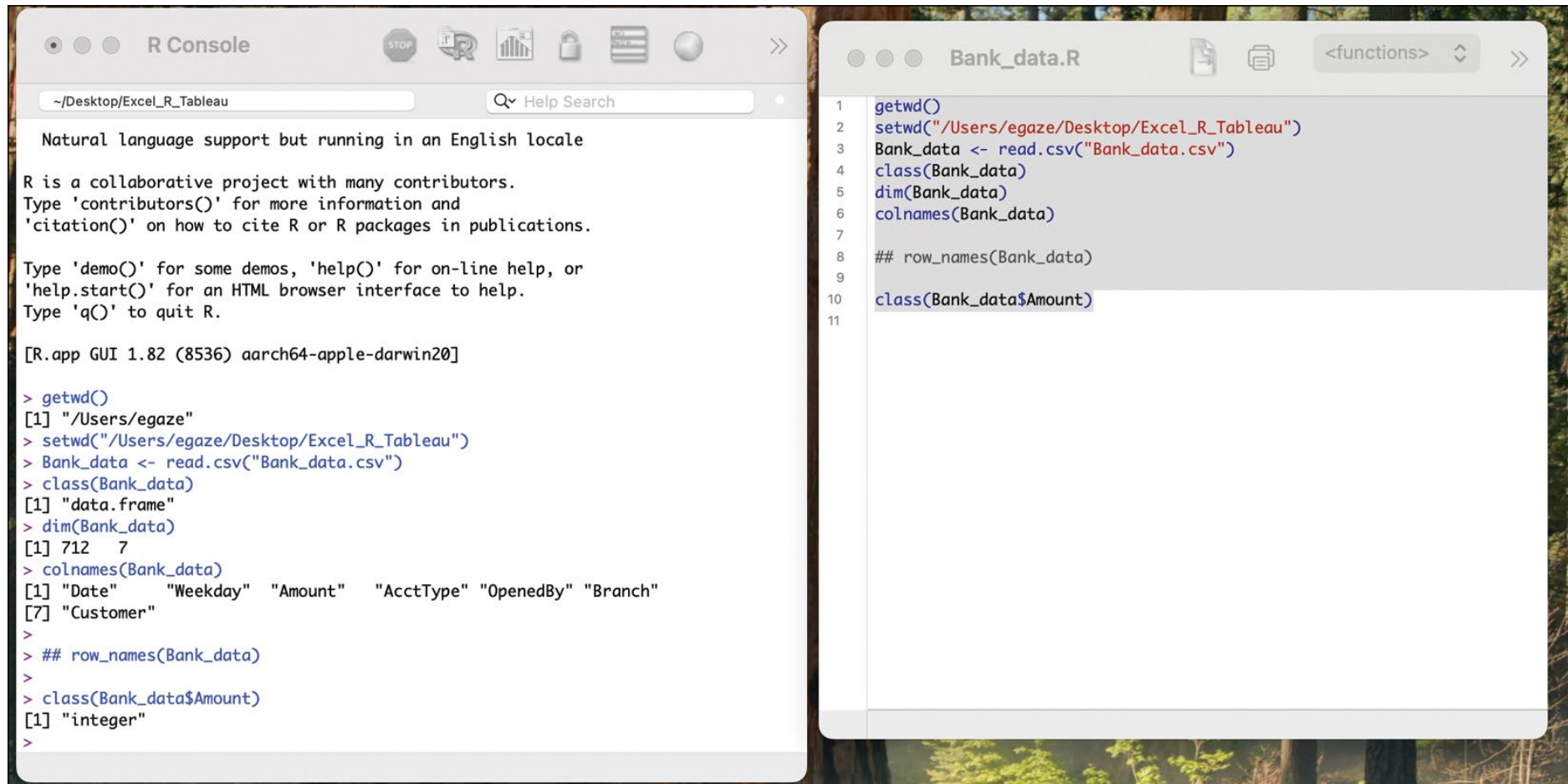


Figure 1.57 Run the script to get back to your previous workspace.

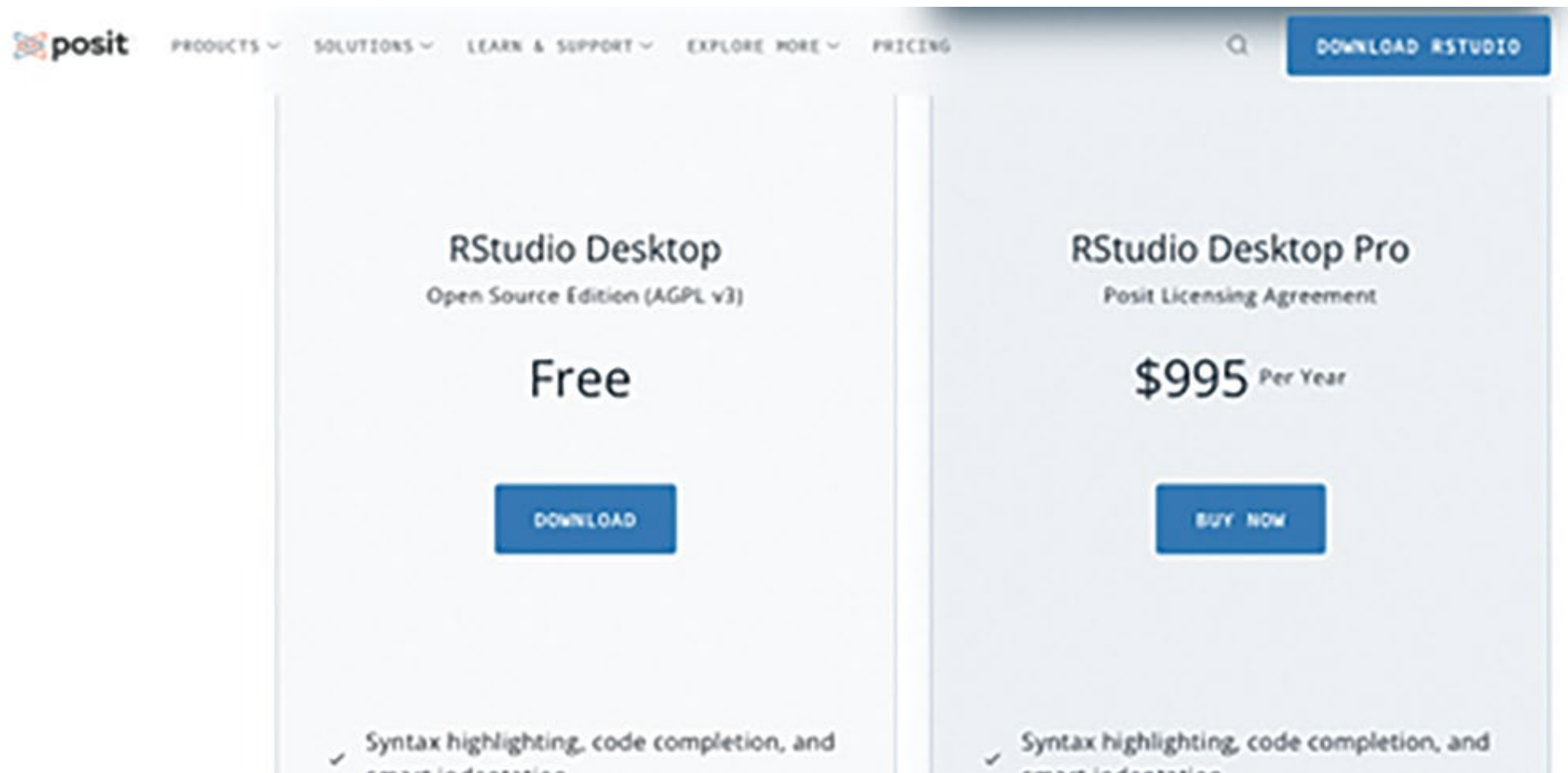


Figure 1.58 Choose the version of RStudio you want to download.

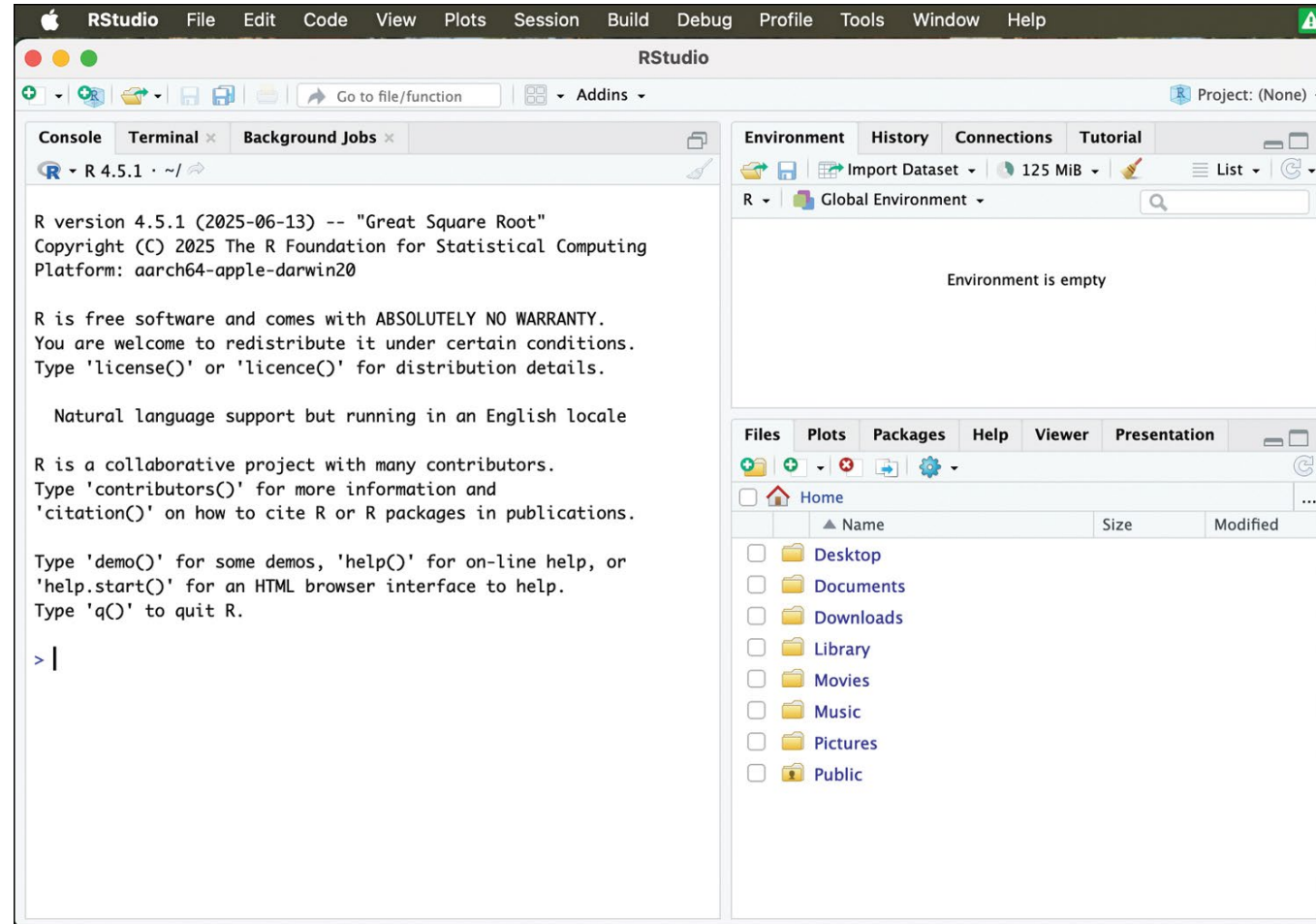


Figure 1.59 RStudio environment with only the console pane on the left; notice that there is no script.

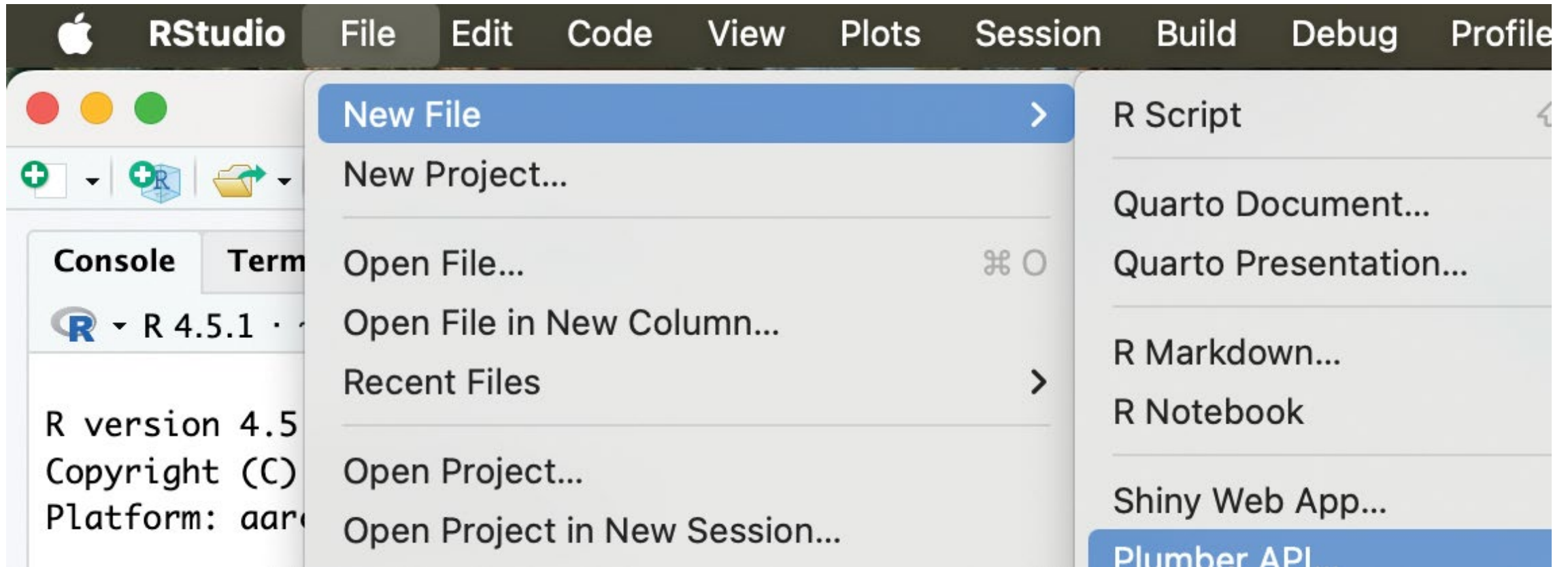


Figure 1.60 Choosing the R Script option.

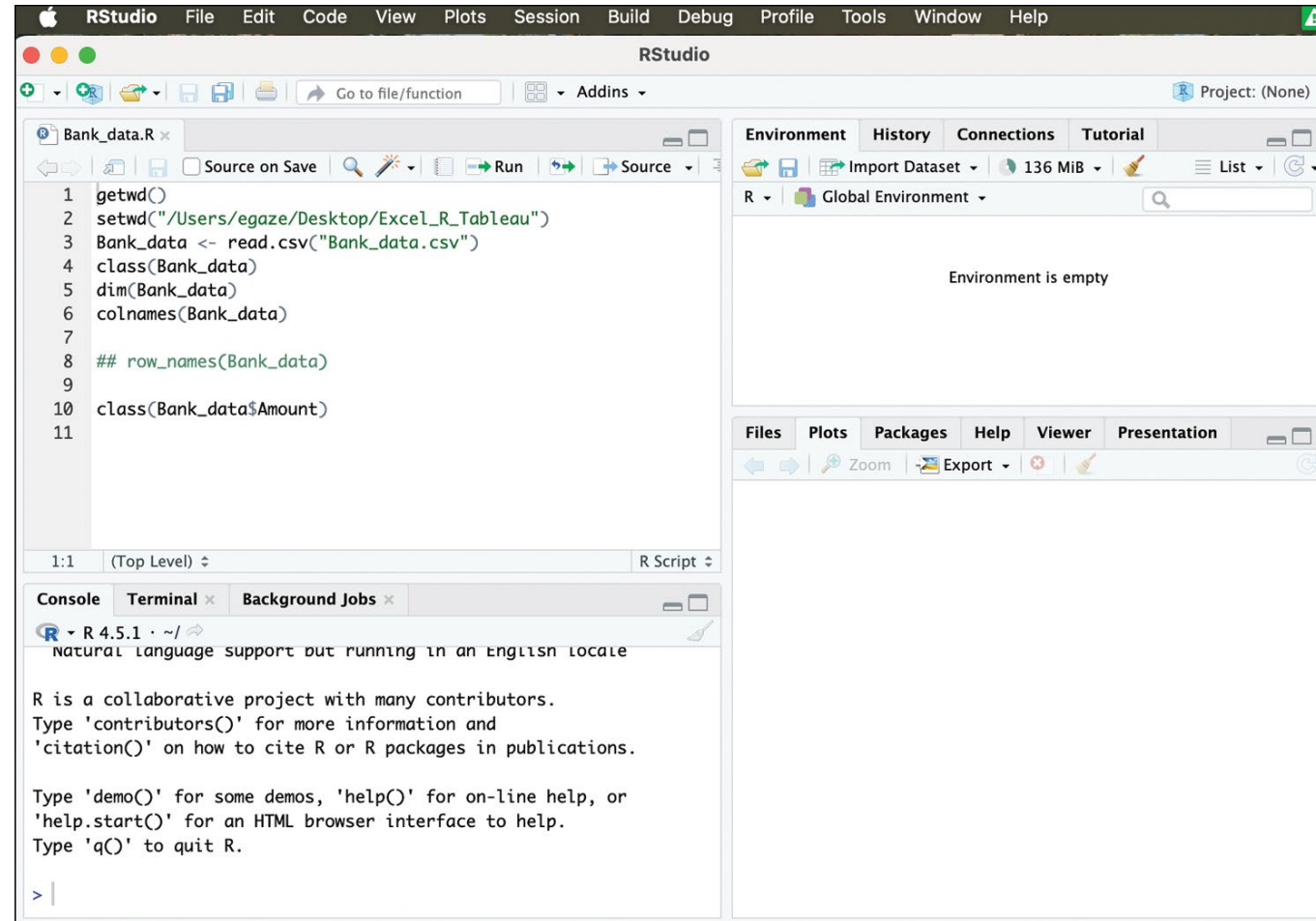


Figure 1.61 The Bank_data.R script opens in a new pane in RStudio.

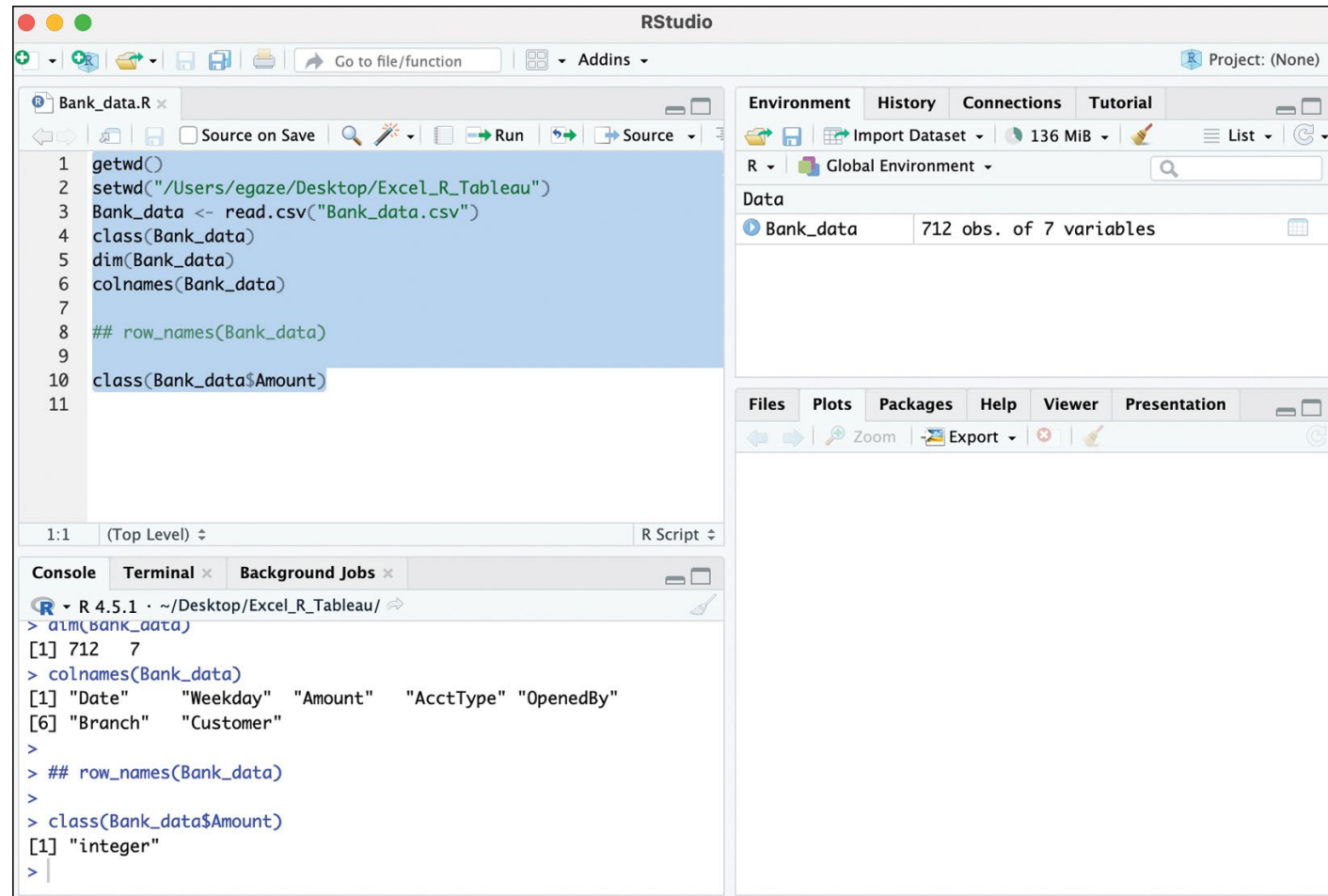


Figure 1.62 Running the highlighted code and its output in the console.

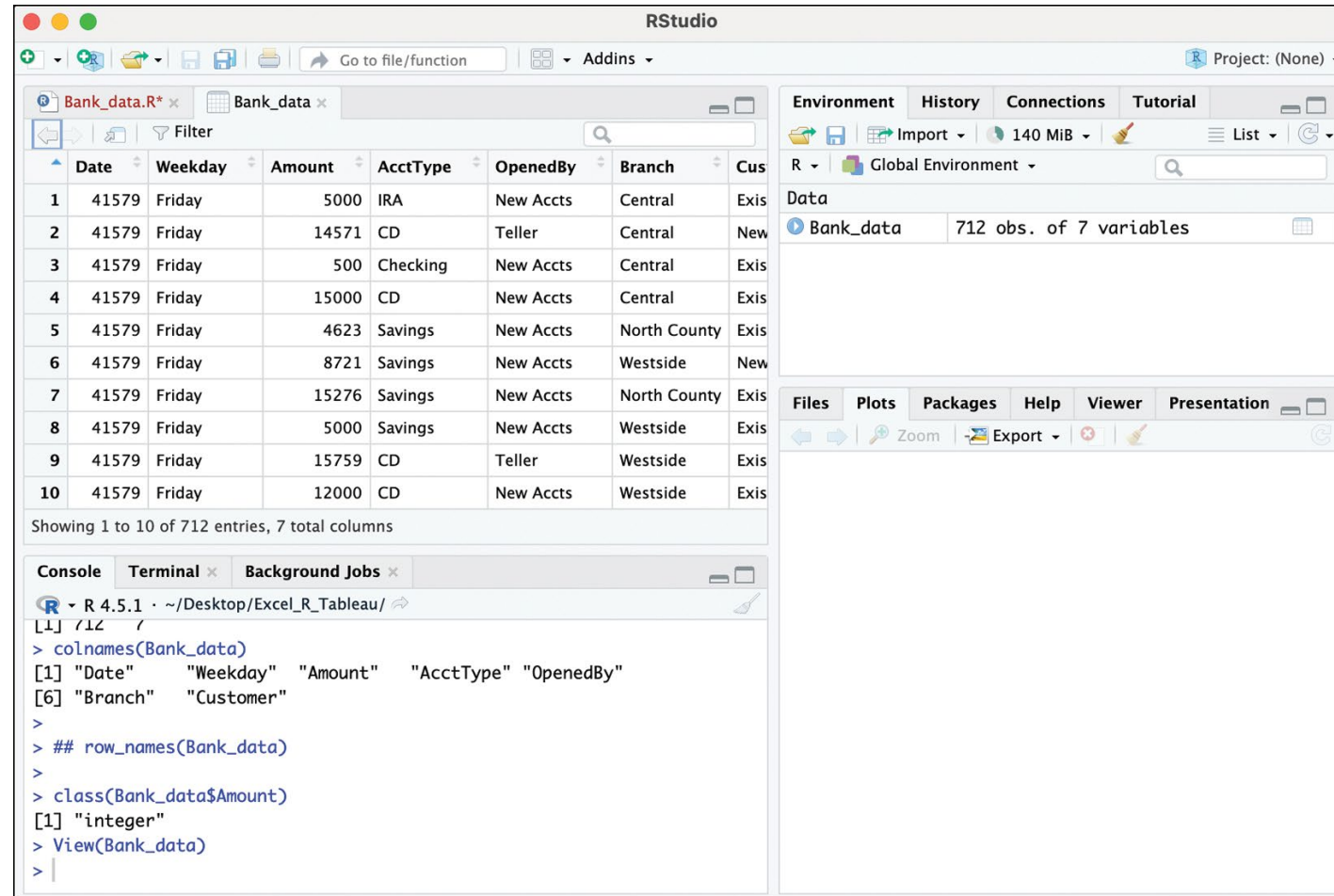


Figure 1.63 Output from the `View(Bank_data)` function appears in a tab next to the R script.

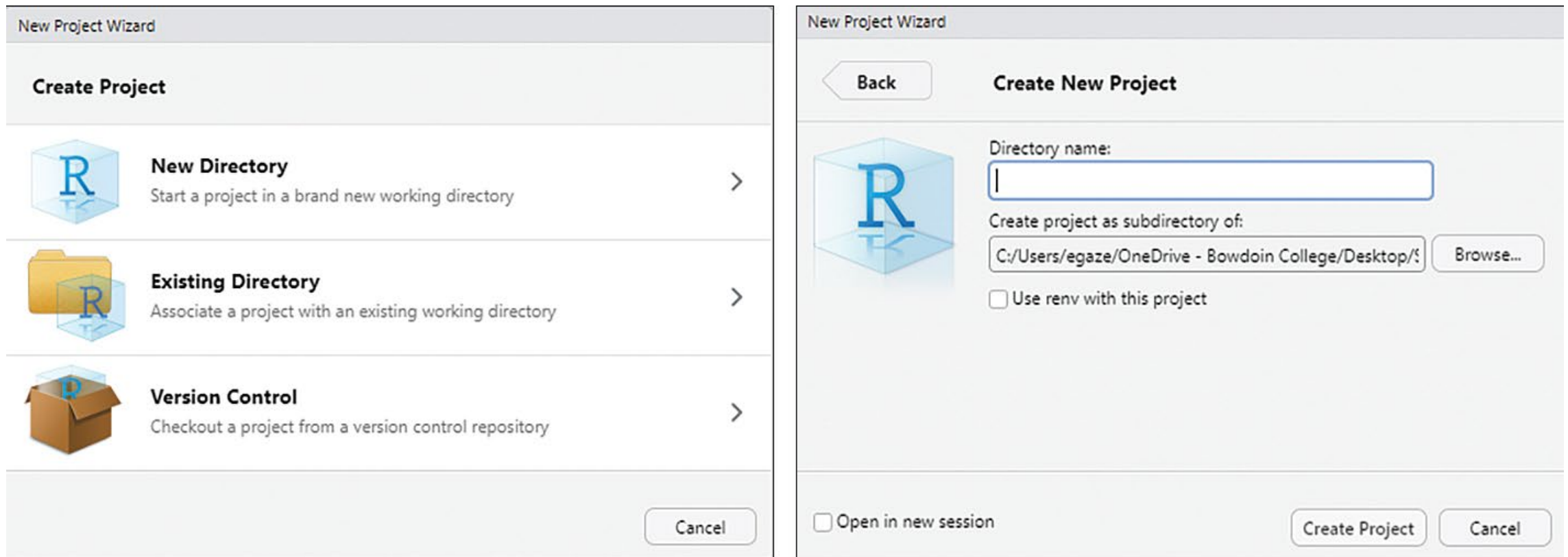


Figure 1.64 The Create Project and Create New Project dialogs in the New Project Wizard.

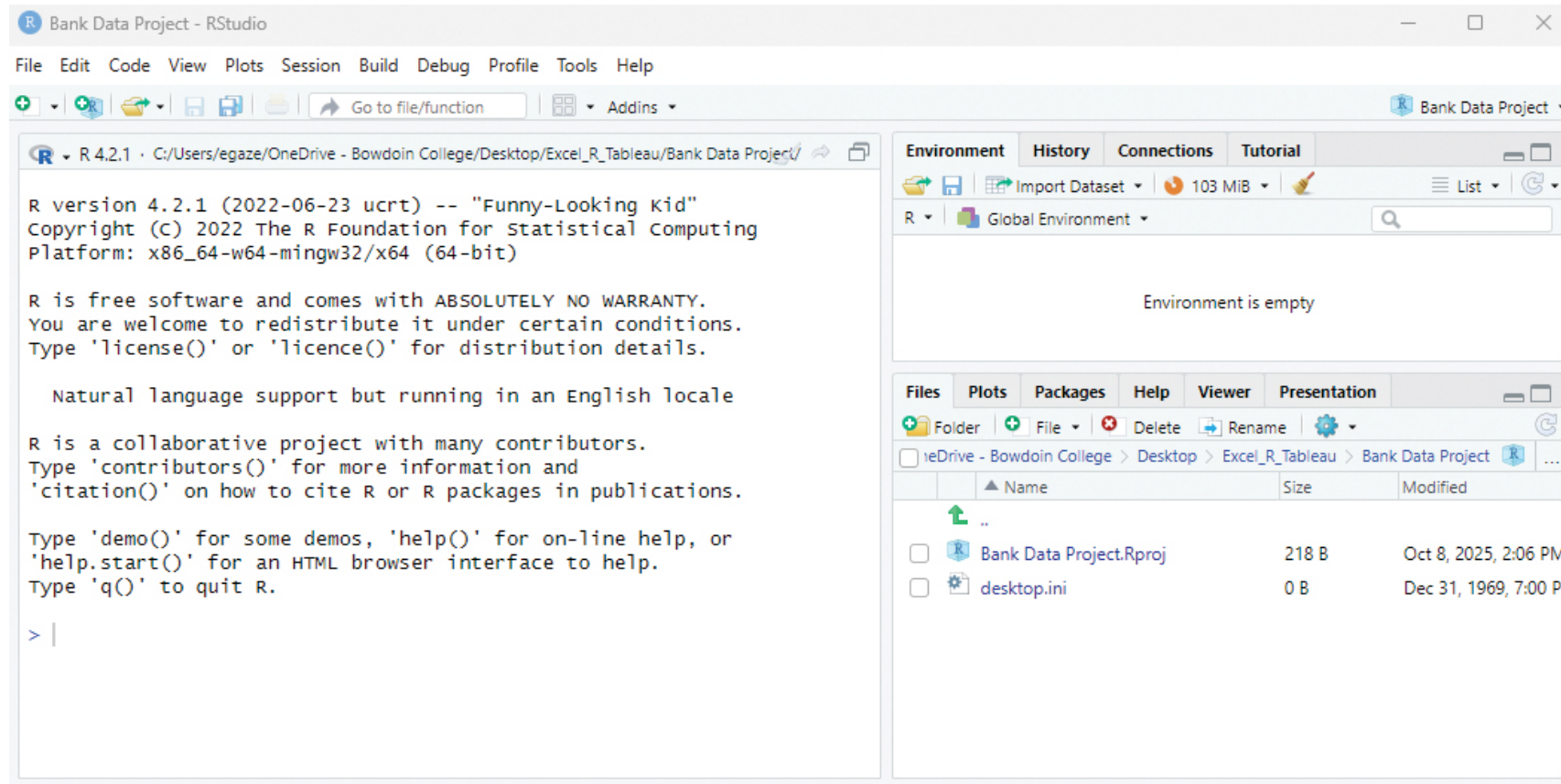


Figure 1.65 New project session in RStudio named Bank Data Project.

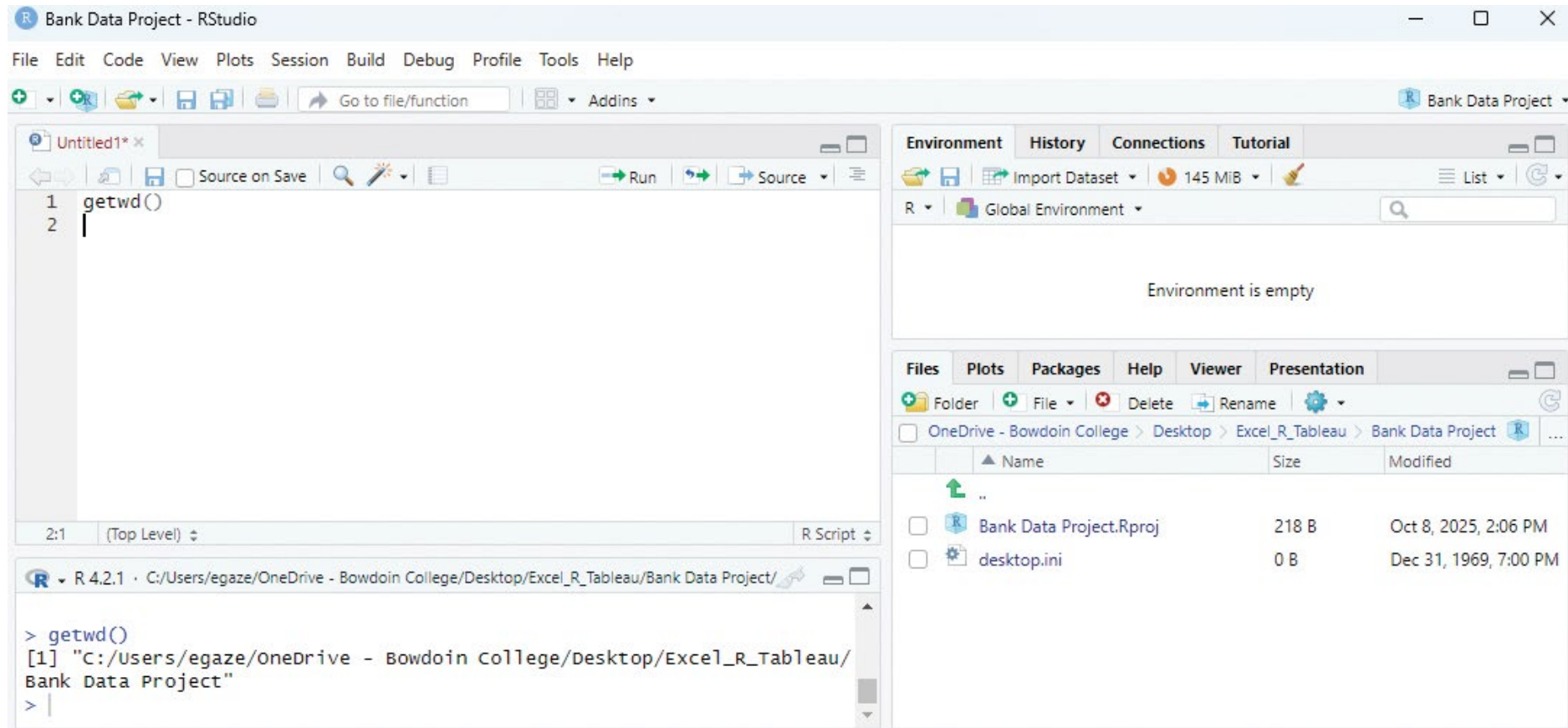


Figure 1.66 The working directory is set to the Bank Data Project folder.

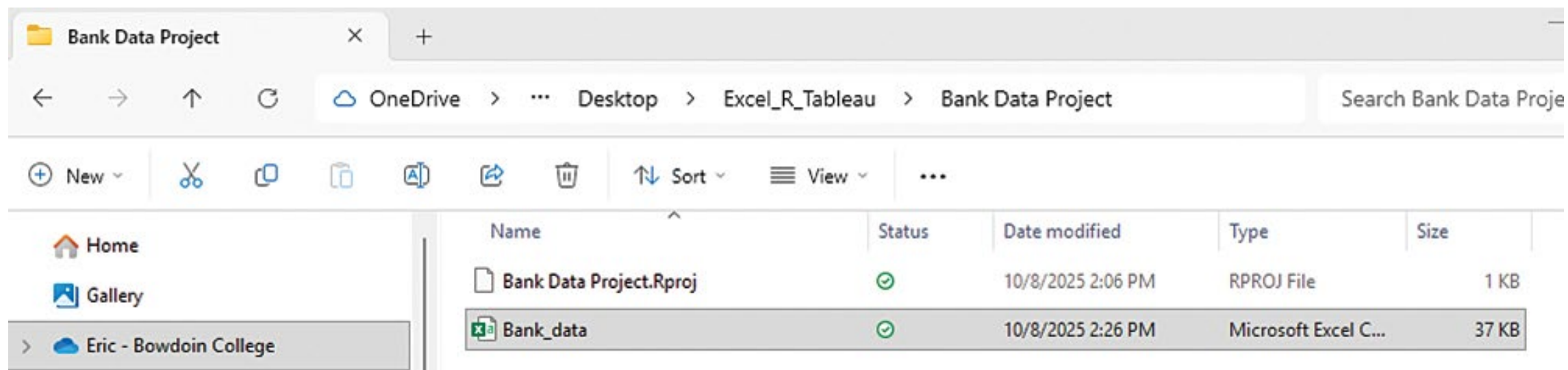


Figure 1.67 Import Dataset in RStudio opens a search in the Bank Data Project folder, which shows the Bank_data CSV file and the R project file.

Import Dataset

Name:

Input File:

```
Date, weekday, Amount, AcctType, OpenedBy, Branch, Customer
41579, Friday, 5000, IRA, New Accts, Central, Existing
41579, Friday, 14571, CD, Teller, Central, New
41579, Friday, 500, Checking, New Accts, Central, Existing
41579, Friday, 15000, CD, New Accts, Central, Existing
41579, Friday, 4623, Savings, New Accts, North County, Existing
41579, Friday, 8721, Savings, New Accts, Westside, New
41579, Friday, 15276, Savings, New Accts, North County, Existing
41579, Friday, 5000, Savings, New Accts, Westside, Existing
41579, Friday, 15759, CD, Teller, Westside, Existing
41579, Friday, 12000, CD, New Accts, Westside, Existing
41579, Friday, 7177, Savings, Teller, North County, Existing
41579, Friday, 6837, Savings, New Accts, Westside, Existing
41579, Friday, 3171, Checking, New Accts, Westside, Existing
41579, Friday, 50000, Savings, New Accts, Central, Existing
41579, Friday, 4690, Checking, New Accts, North County, New
41579, Friday, 12438, Checking, New Accts, Central, Existing
```

Encoding:

Heading: ☒ Yes ☐ No

Row names:

Separator:

Decimal:

Quote:

Comment:

na.strings:

☐ Strings as factors

Data Frame

Date	weekday	Amount	AcctType	OpenedBy	Branch
41579	Friday	5000	IRA	New Accts	Central
41579	Friday	14571	CD	Teller	Central
41579	Friday	500	Checking	New Accts	Central
41579	Friday	15000	CD	New Accts	Central
41579	Friday	4623	Savings	New Accts	North Count
41579	Friday	8721	Savings	New Accts	Westside
41579	Friday	15276	Savings	New Accts	North Count
41579	Friday	5000	Savings	New Accts	Westside
41579	Friday	15759	CD	Teller	Westside
41579	Friday	12000	CD	New Accts	Westside
41579	Friday	7177	Savings	Teller	North Count
41579	Friday	6837	Savings	New Accts	Westside
41579	Friday	3171	Checking	New Accts	Westside
41579	Friday	50000	Savings	New Accts	Central
41579	Friday	4690	Checking	New Accts	North Count
41579	Friday	12438	Checking	New Accts	Central

Figure 1.68 The data in CSV format.

Bank Data Project - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Bank_data x

	Date	Weekday	Amount	AcctType	OpenedBy	Branch	Customer
1	41579	Friday	5000	IRA	New Accts	Central	Existing
2	41579	Friday	14571	CD	Teller	Central	New
3	41579	Friday	500	Checking	New Accts	Central	Existing
4	41579	Friday	15000	CD	New Accts	Central	Existing
5	41579	Friday	4623	Savings	New Accts	North County	Existing
6	41579	Friday	8721	Savings	New Accts	Westside	New
7	41579	Friday	15276	Savings	New Accts	North County	Existing
8	41579	Friday	5000	Savings	New Accts	Westside	Existing

Showing 1 to 9 of 712 entries, 7 total columns

R 4.2.1 · C:/Users/egaze/OneDrive - Bowdoin College/Desktop/Excel_R_Tableau/Bank Data Project/

```

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> Bank_data <- read.csv("C:/Users/egaze/OneDrive - Bowdoin College/Desktop/Excel_R_Tableau/Bank Data Project/Bank_data.csv")
> view(Bank_data)
>

```

Environment History Connections Tutorial

Import Dataset 144 MiB

R Global Environment

Data

Bank_data 712 obs. of 7 variables

Files Plots Packages Help Viewer Presentation

Folder File Delete Rename

OneDrive - Bowdoin College > Desktop > Excel_R_Tableau > Bank Data Project

	Name	Size	Modified
	..		
☐	.Rhistory	325 B	Oct 8, 2025, 2:32 PM
☐	Bank Data Project.Rproj	218 B	Oct 8, 2025, 2:33 PM
☐	Bank_data.csv	36.9 KB	Sep 14, 2023, 11:47 AM

Figure 1.69 The data file is read into RStudio.

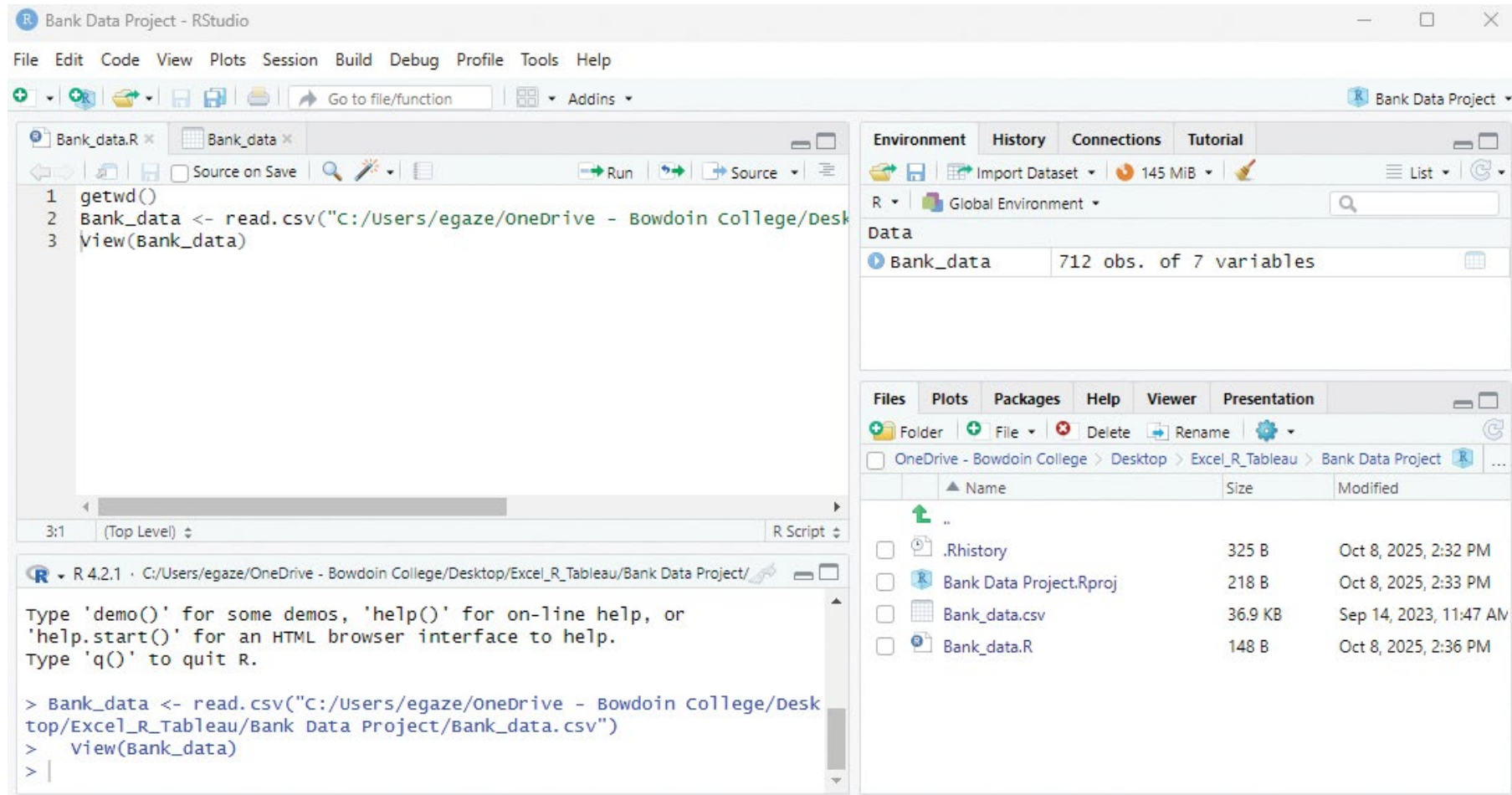


Figure 1.70 The various RStudio panes: Console, Script, Data Environment, and Files.

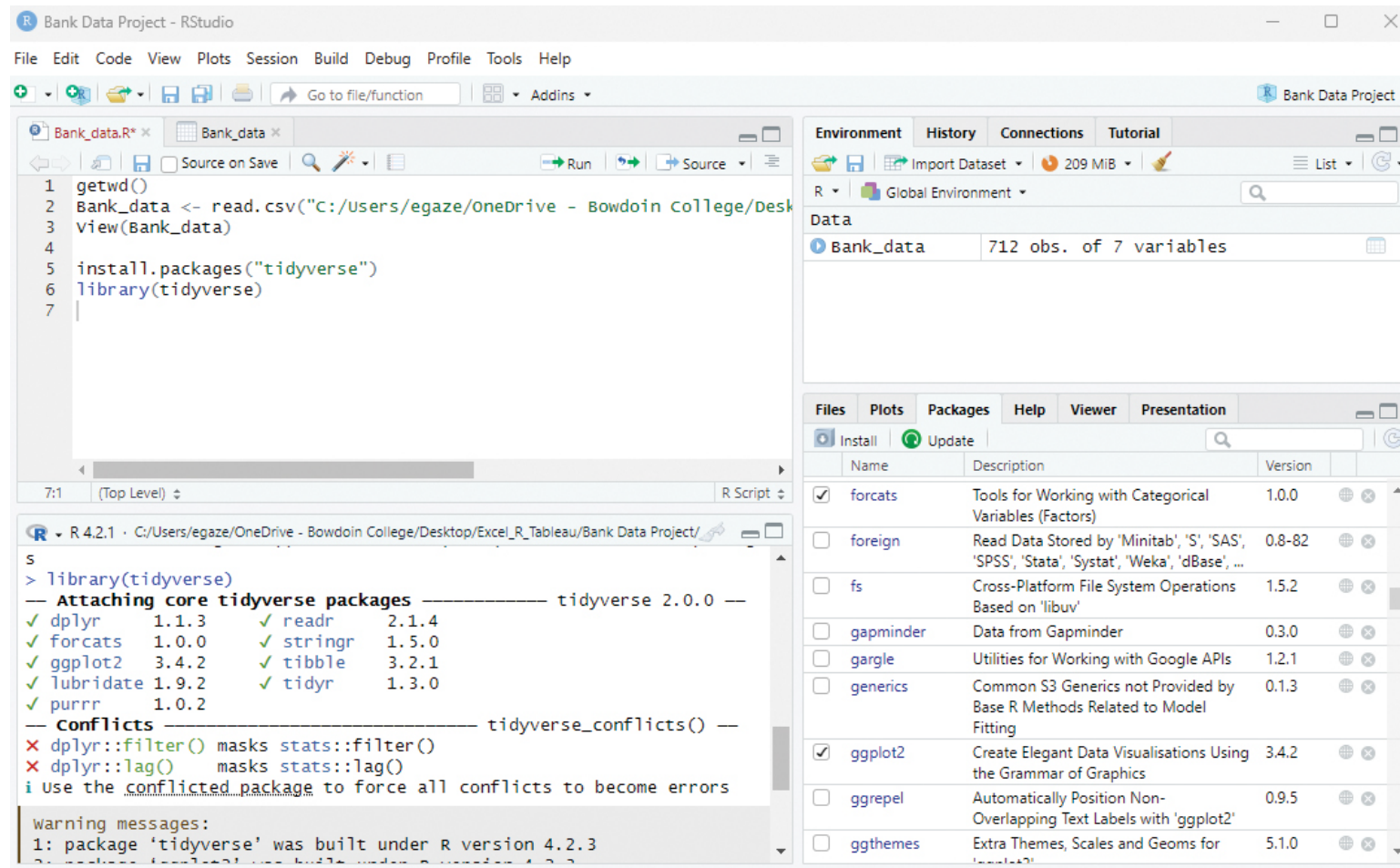


Figure 1.71 Installing the tidyverse package.

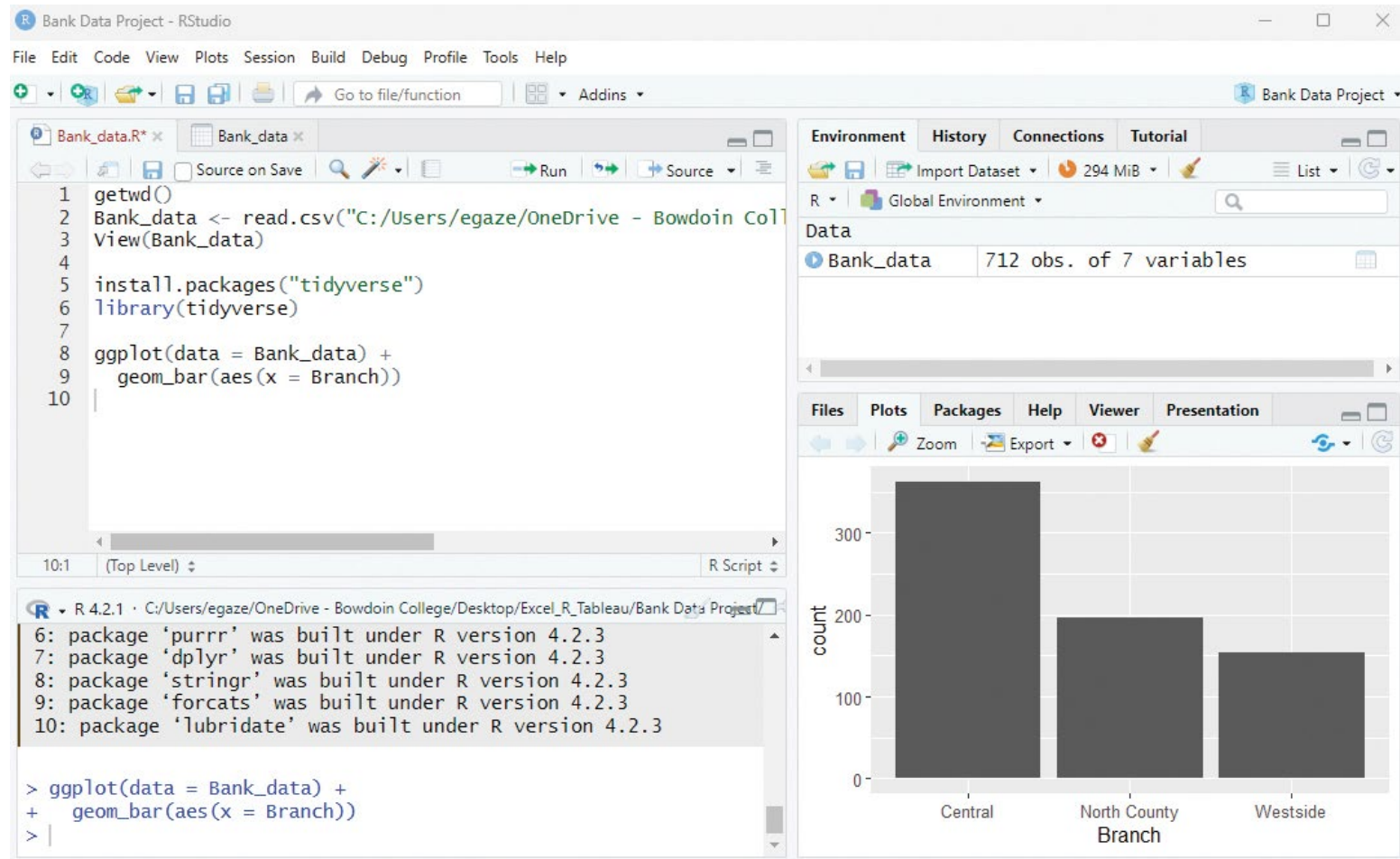


Figure 1.72 Bar plot showing counts of branches.

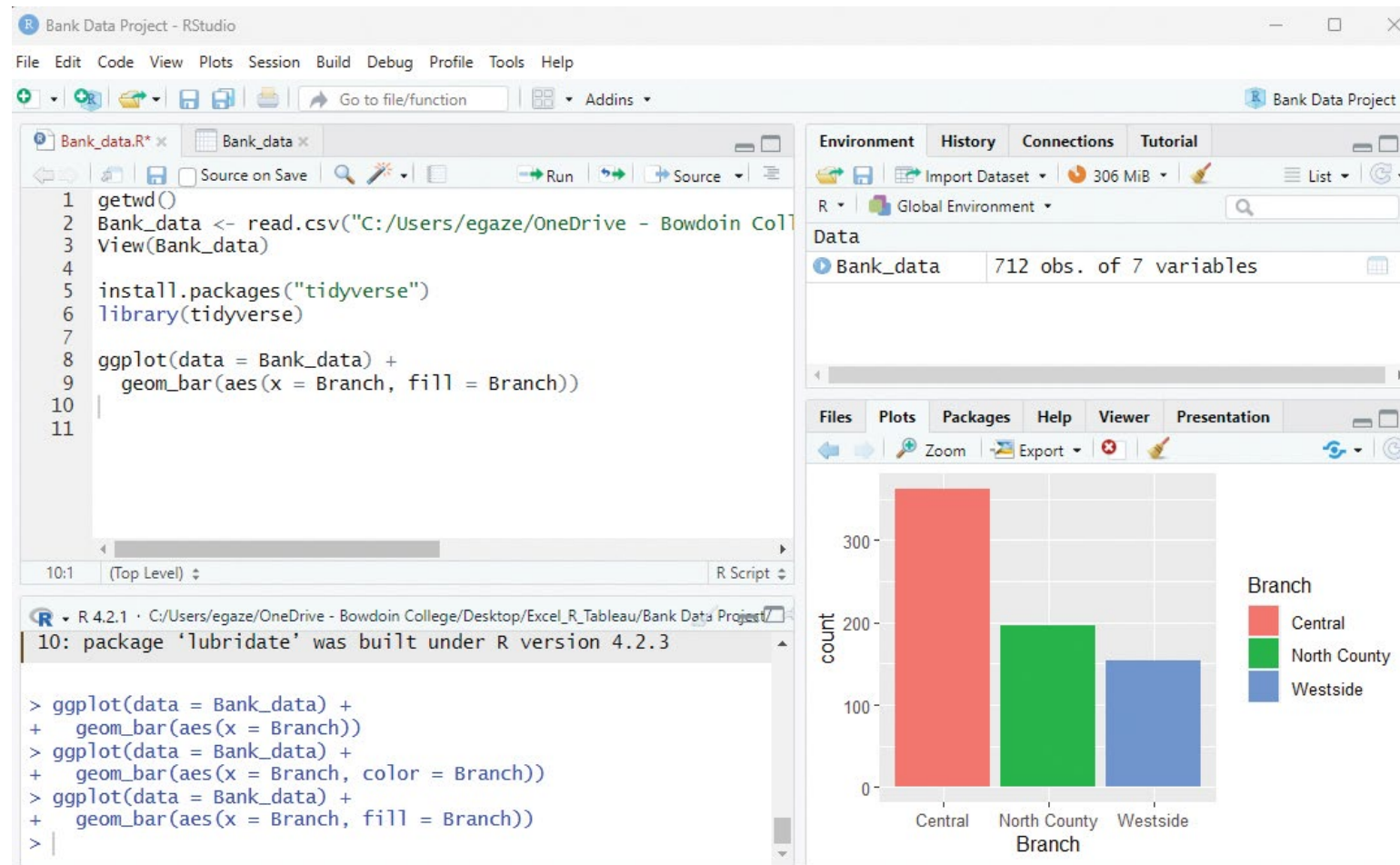


Figure 1.73 Color added by mapping the Branch variable to the fill aesthetic.

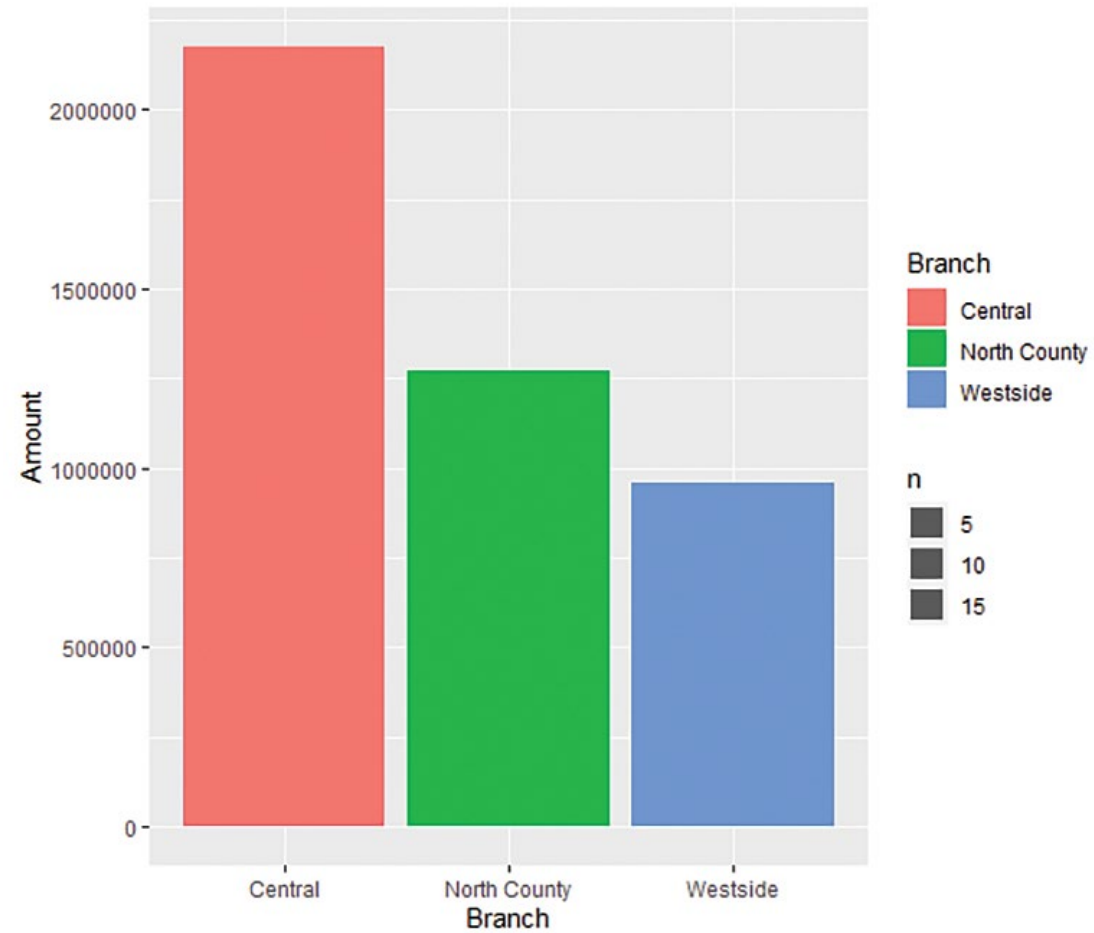


Figure 1.74 First attempt at plotting the sum of Amount by Branch using the statistical transformation set equal to sum. The sums are incorrect!

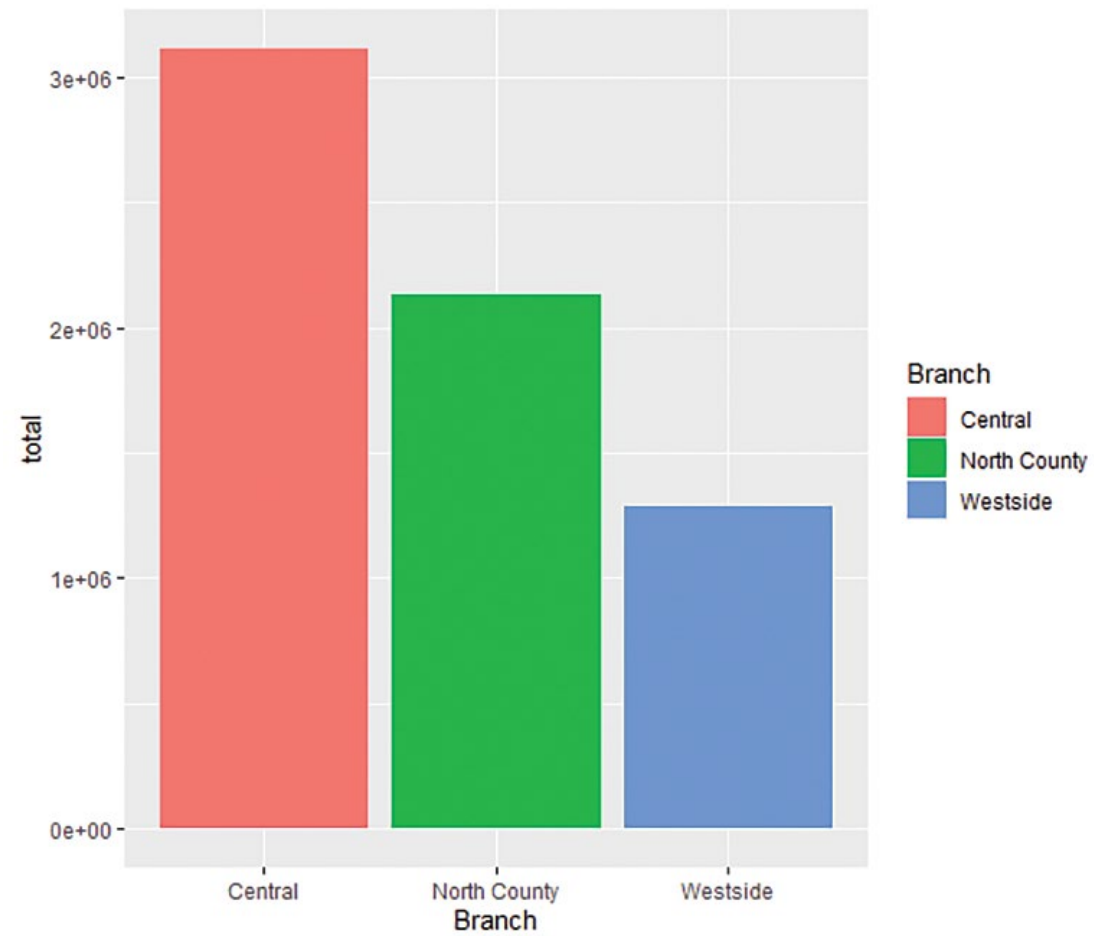
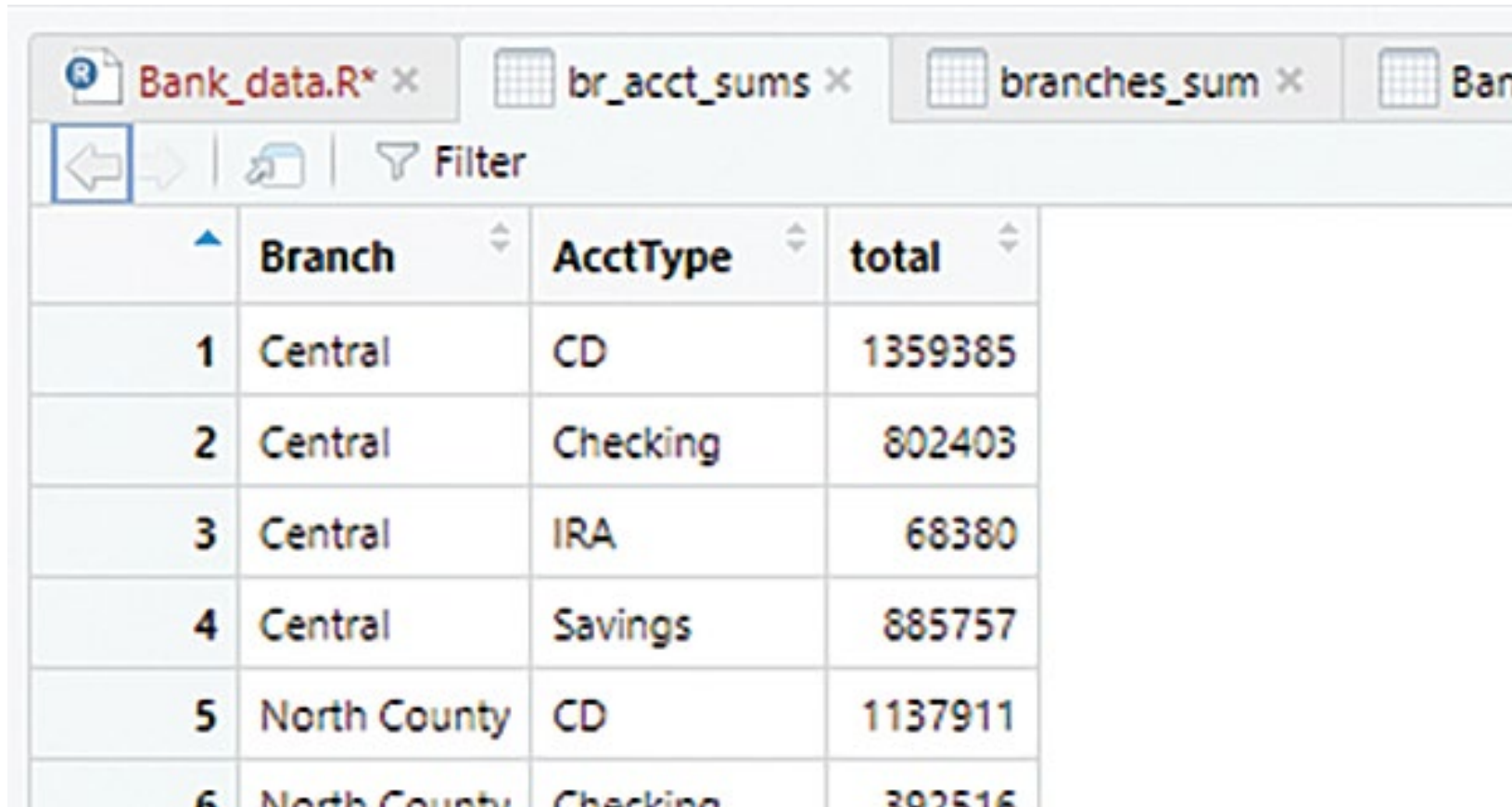


Figure 1.75 Actual values used rather than counts.



The screenshot shows an RStudio window with three tabs: 'Bank_data.R*', 'br_acct_sums', and 'branches_sum'. The 'br_acct_sums' tab is active, displaying a data table. The table has four columns: an index column, 'Branch', 'AcctType', and 'total'. The data is as follows:

	Branch	AcctType	total
1	Central	CD	1359385
2	Central	Checking	802403
3	Central	IRA	68380
4	Central	Savings	885757
5	North County	CD	1137911
6	North County	Checking	302516

Figure 1.76 Sum of Amount for each Branch and account type.

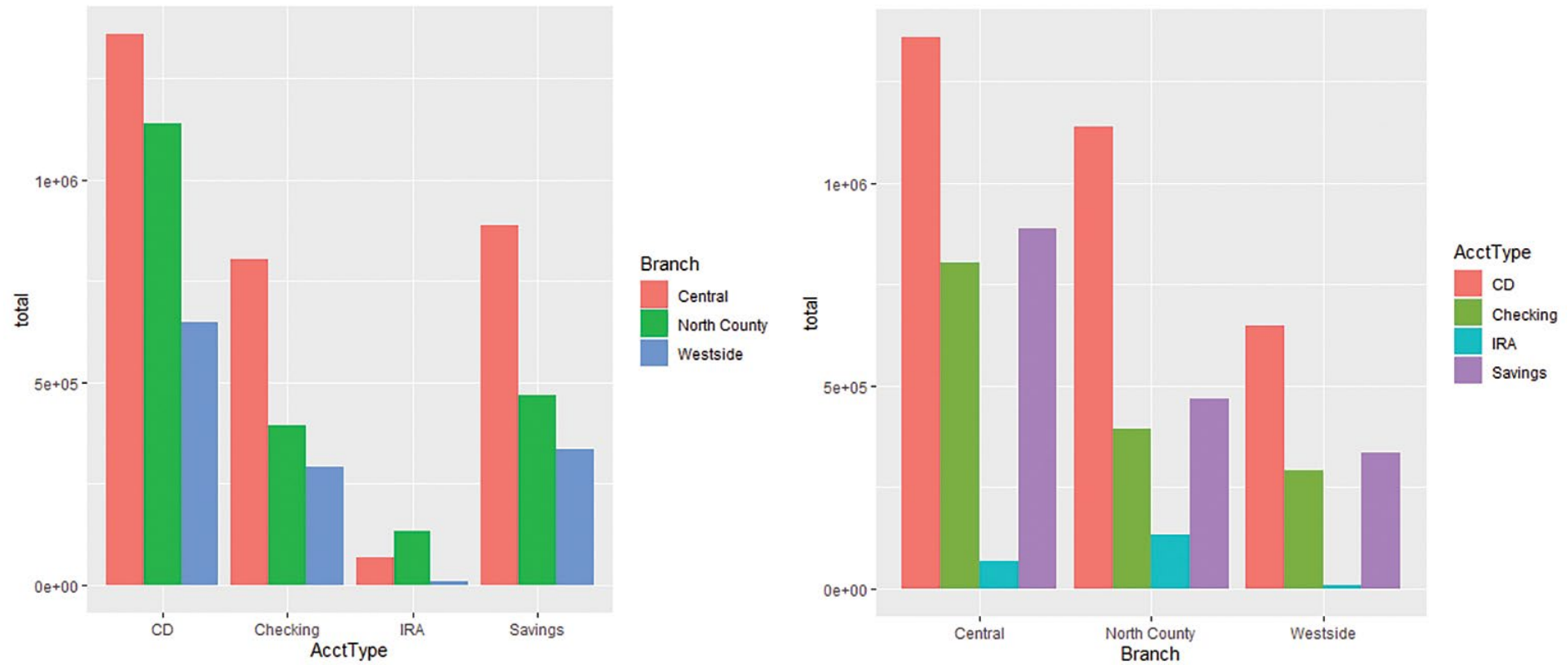


Figure 1.77 Switching Branch and Account Type order to get both target graphics.

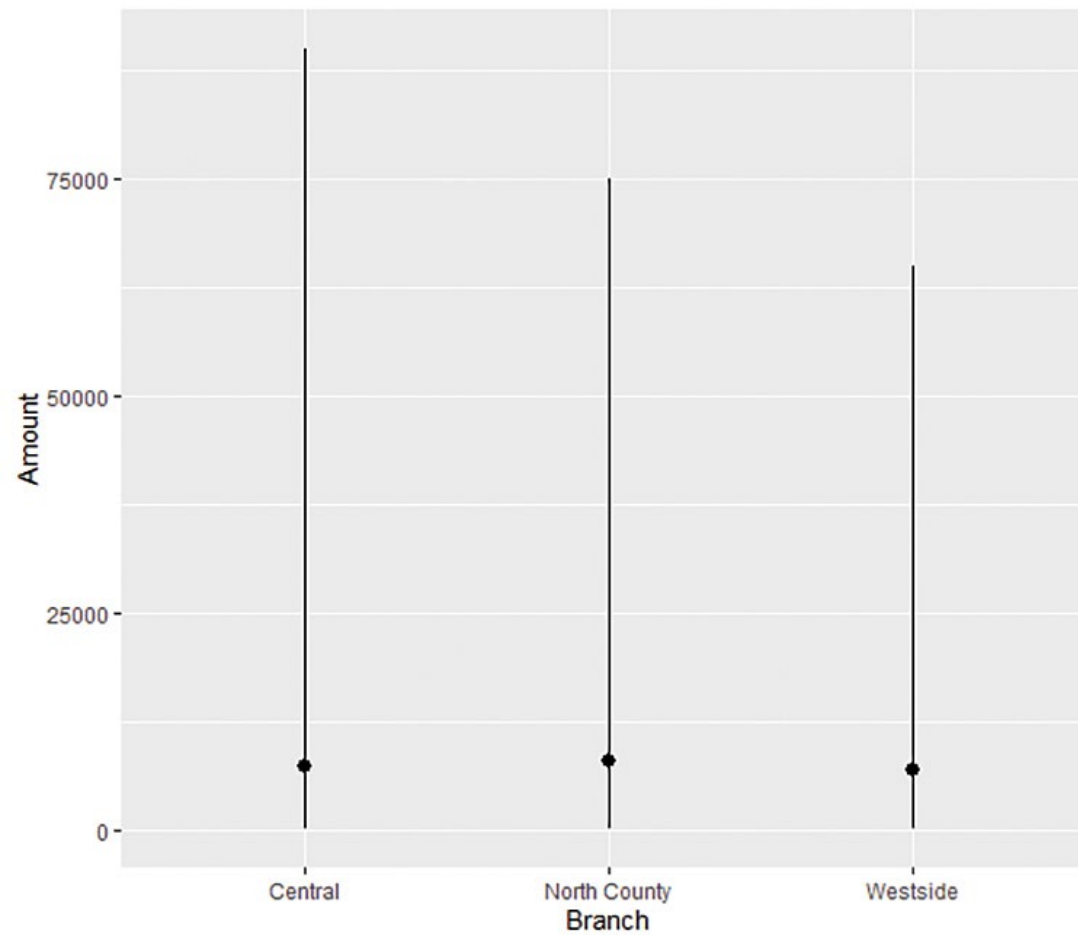


Figure 1.78 Using `stat_summary` to show the minimum, maximum, and median values in the distribution of Amount by Branch.

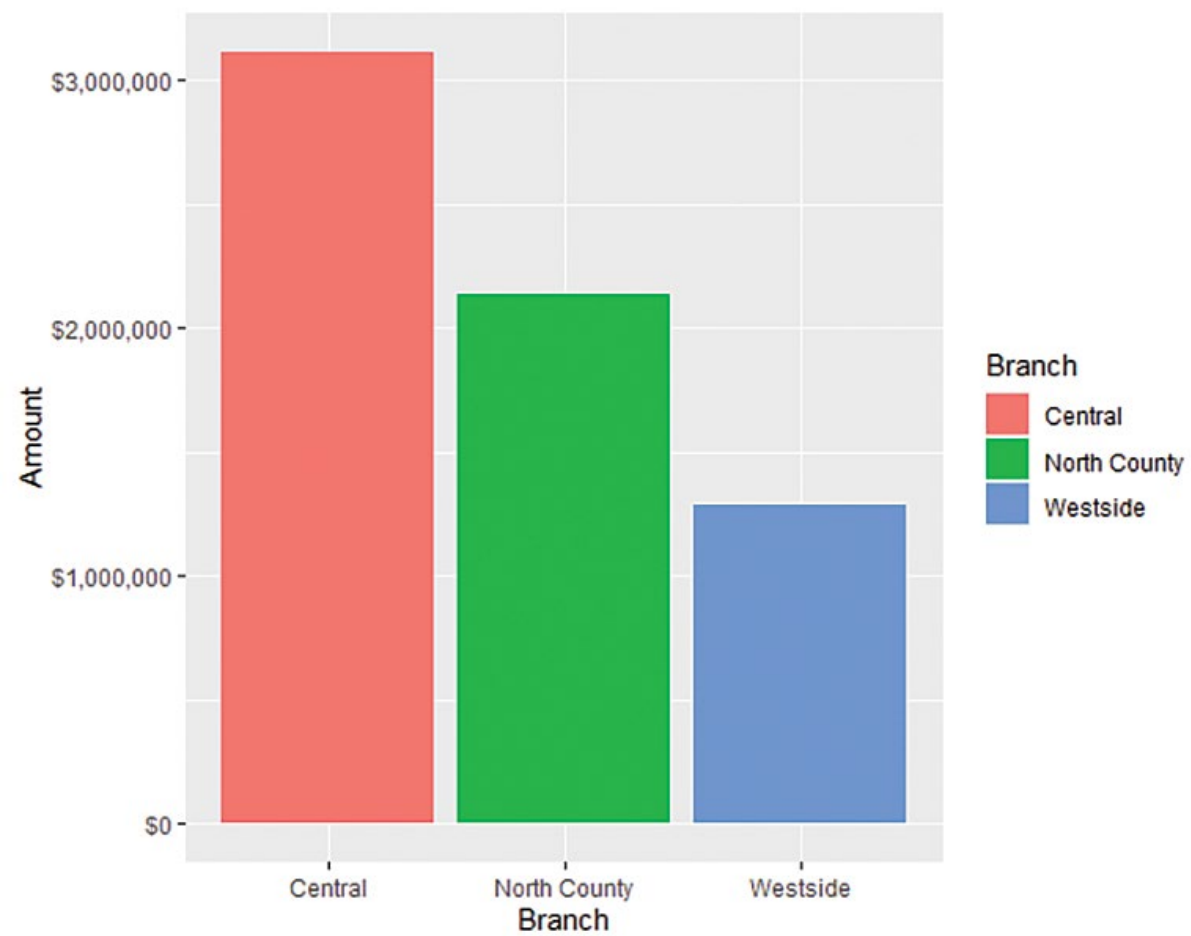


Figure 1.79 The original plot with the correct sums and dollar formatting.

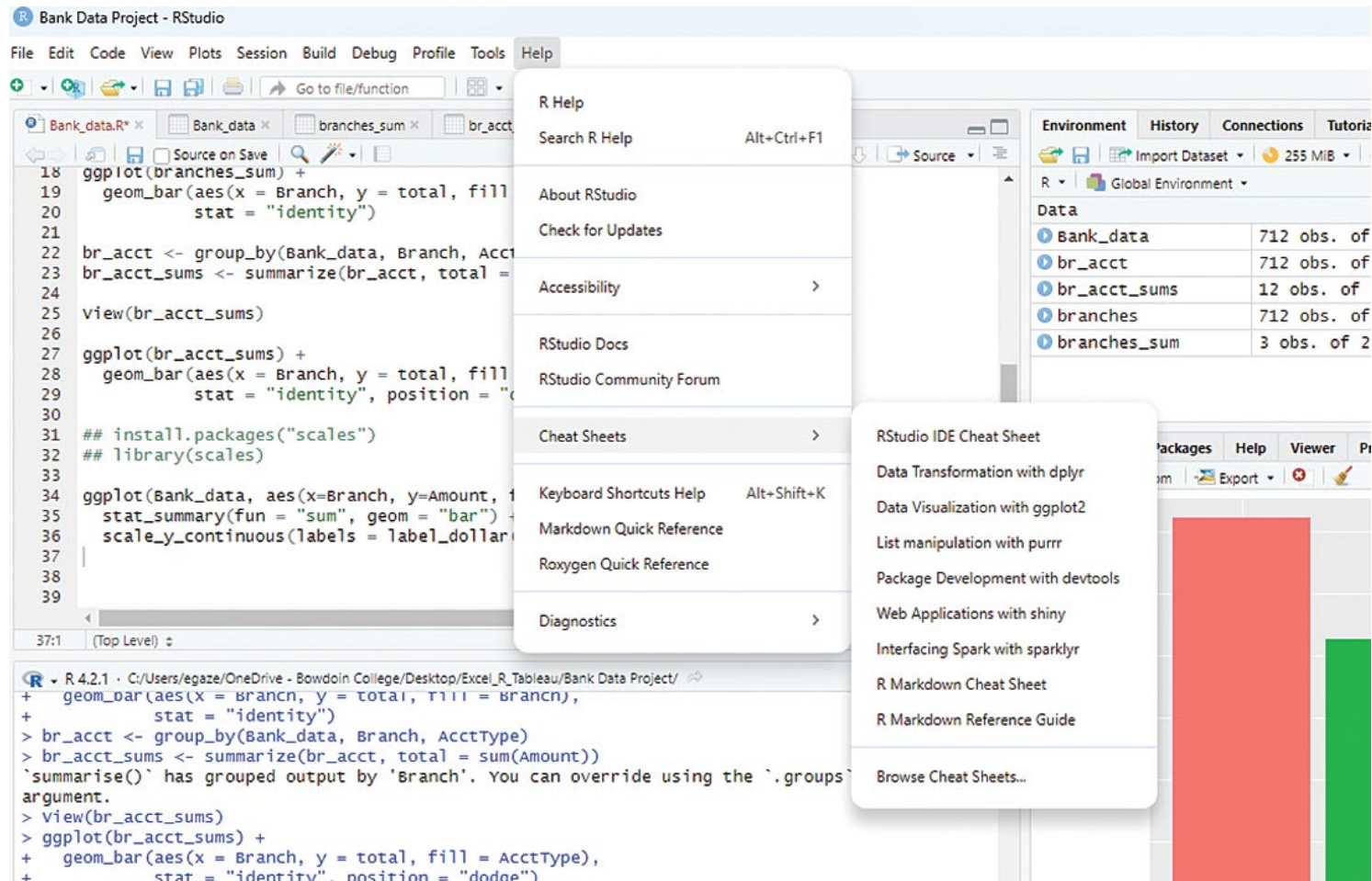


Figure 1.80 Cheat sheets!

The screenshot shows the RStudio interface. The left pane displays an R script with the following code:

```

6
7
8 class(Bank_data)
9 class(Amount)
10 class(Bank_data$Amount)
11
12 install.packages("tidyverse")
13 library(tidyverse)
14
15 # install.packages("scales")
16 library(scales)
17
18 ?ggplot
19 ?mpg
20 mpg
21
22
23 # Note that geom_bar defaults to count
24 # Check out ggplot2 cheat sheet
25 # Reminder we don't need to specify data and mapping arguments
26 # And we can move the aesthetic from global to local in geom_bar

```

The bottom pane shows the console output for the `mpg` command:

```

# A tibble: 234 × 11
  manufacturer model      displ  year   cyl trans      drv      cty   hwy fl      class
  <chr>         <chr>    <dbl> <int> <int> <chr>    <chr> <int> <chr> <chr>
1 audi         a4         1.8   1999    4 auto(l5) f        18    29 p    compact
2 audi         a4         1.8   1999    4 manual(m5) f        21    29 p    compact
3 audi         a4         2     2008    4 manual(m6) f        20    31 p    compact
4 audi         a4         2     2008    4 auto(av) f        21    30 p    compact
5 audi         a4         2.8   1999    6 auto(l5) f        16    26 p    compact
6 audi         a4         2.8   1999    6 manual(m5) f        18    26 p    compact
7 audi         a4         3.1   2008    6 auto(av) f        18    27 p    compact
8 audi         a4 quattro 1.8   1999    4 manual(m5) 4        18    26 p    compact
9 audi         a4 quattro 1.8   1999    4 auto(l5) 4        16    25 p    compact
10 audi         a4 quattro 2     2008    4 manual(m6) 4        20    28 p    compact
# i 224 more rows
# i Use `print(n = ...)` to see more rows
>

```

The right pane shows the Environment pane with a list of objects:

Object	Size
account_types	712 obs. of 7 variables
AcctType	712 obs. of 7 variables
AcctType_sum	4 obs. of 2 variables
AT_Sum	4 obs. of 2 variables
Bank_data	712 obs. of 7 variables
bank_df	712 obs. of 7 variables
bank_df2	712 obs. of 11 variables
BankData	712 obs. of 8 variables

The bottom right pane shows the documentation for the `mpg` dataset:

Fuel economy data from 1999 to 2008 for 38 popular models of cars

Description

This dataset contains a subset of the fuel economy data that the EPA makes available on <https://fuelconomy.gov/>. It contains only models which had a new release every year between 1999 and 2008 - this was used as a proxy for the popularity of the car.

Usage

```
mpg
```

Format

A data frame with 234 rows and 11 variables:

```
manufacturer
```

Figure 1.81 The mpg data that comes with the ggplot2 package.

```

22 e <- ggplot(mpg, aes(cty,hwy))
23 e + geom_point()
24
25
26
27 # Note that geom_bar defaults to count
28 # Check out ggplot2 cheat sheet
29 # Reminder we don't need to specify data and mapping arguments
30 # And we can move the aesthetic from global to local in geom_bar
27:1 (Top Level)

```

R 4.2.1 · C:/Users/egaze/OneDrive - Bowdoin College/Desktop/Spring 2026/MATH 1100 Data Viz Sp26/Case Studies/Bank Data/Bank Data R/

	manufacturer	model	displ	year	cty	trans	drv	cty	hwy	fl	class
	<chr>	<chr>	<dbl>	<int>	<int>	<chr>	<chr>	<int>	<int>	<chr>	<chr>
1	audi	a4	1.8	1999	4	auto(l5)	f	18	29	p	compact
2	audi	a4	1.8	1999	4	manual(m5)	f	21	29	p	compact
3	audi	a4	2	2008	4	manual(m6)	f	20	31	p	compact
4	audi	a4	2	2008	4	auto(av)	f	21	30	p	compact
5	audi	a4	2.8	1999	6	auto(l5)	f	16	26	p	compact
6	audi	a4	2.8	1999	6	manual(m5)	f	18	26	p	compact
7	audi	a4	3.1	2008	6	auto(av)	f	18	27	p	compact
8	audi	a4 quattro	1.8	1999	4	manual(m5)	4	18	26	p	compact
9	audi	a4 quattro	1.8	1999	4	auto(l5)	4	16	25	p	compact
10	audi	a4 quattro	2	2008	4	manual(m6)	4	20	28	p	compact

```

# i 224 more rows
# i use `print(n = ...)` to see more rows
> e <- ggplot(mpg, aes(cty,hwy))
> e + geom_point()
> |

```

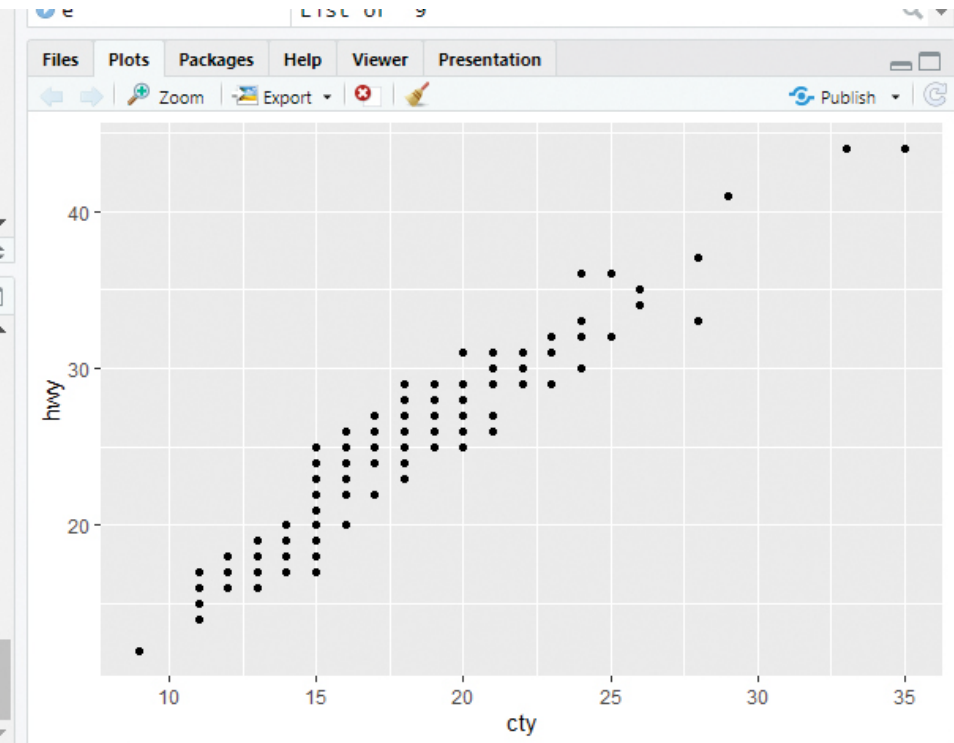


Figure 1.82 A scatterplot showing the relationship between city and highway mpg.


```
22 e <- ggplot(mpg, aes(cty,hwy))
23 e + geom_point()
24 ?ggplot|
25
26
27 # Note that geom_bar defaults to count
24:8 (Top Level) ↕
```

R 4.2.1 · C:/Users/egaze/OneDrive - Bowdoin College/Desktop/Spring 2026/MATH 1100 Data Viz Sp26/Case Studies/Bank Data/Bank Data R/

```
> mpg
# A tibble: 234 × 11
  manufacturer model    displ  year   cyl trans      drv    cty   hwy fl   class
  <chr>         <chr>    <dbl> <int> <int> <chr>   <chr> <int> <int> <chr> <chr>
1 audi         a4         1.8  1999     4 auto(l5) f      18    29 p    compact
2 audi         a4         1.8  1999     4 manual(m5) f      21    29 p    compact
3 audi         a4         2    2008     4 manual(m6) f      20    31 p    compact
4 audi         a4         2    2008     4 auto(av) f      21    30 p    compact
5 audi         a4         2.8  1999     6 auto(l5) f      16    26 p    compact
6 audi         a4         2.8  1999     6 manual(m5) f      18    26 p    compact
7 audi         a4         3.1  2008     6 auto(av) f      18    27 p    compact
8 audi         a4 quattro  1.8  1999     4 manual(m5) 4      18    26 p    compact
9 audi         a4 quattro  1.8  1999     4 auto(l5) 4      16    25 p    compact
10 audi         a4 quattro  2    2008     4 manual(m6) 4      20    28 p    compact
# i 224 more rows
# i Use `print(n = ...)` to see more rows
> e <- ggplot(mpg, aes(cty,hwy))
> e + geom_point()
> ?ggplot
> |
```

Files Plots Packages Help Viewer Presentation

R: Create a new ggplot Find in Topic

ggplot {ggplot2}

R Documentation

Create a new ggplot

Description

`ggplot()` initializes a ggplot object. It can be used to declare the input data frame for a graphic and to specify the set of plot aesthetics intended to be common throughout all subsequent layers unless specifically overridden.

Usage

```
ggplot(data = NULL, mapping = aes(), ..., environment = parent.frame())
```

Arguments

data	Default dataset to use for plot. If not already a data.frame, will be converted to one by fortify(.) . If not specified, must be supplied in each layer added to the plot.
mapping	Default list of aesthetic mappings to use for plot. If not specified, must be supplied in each layer added to the plot.

Figure 1.83 Information on ggplot.

Country Comparisons —

Real GDP per capita

Compares GDP on a purchasing power parity basis divided by population, as of 1 July for the same year.

219 Results Clear Filters

Regions

All

DOWNLOAD DATA

Rank	Country	Date of Information
1	Monaco	\$270,100 2024 est.

 **CENTRAL INTELLIGENCE AGENCY**
THE WORK OF A NATION.
THE CENTER OF INTELLIGENCE.

Report Information Contact

Search CIA.gov SEARCH

العربية 中文 Français Русский Español More

HOME ABOUT CIA CAREERS & INTERSHIPS OFFICES OF CIA NEWS & INFORMATION LIBRARY SPY KIDS

Library

Library

Publications

• The World Factbook

World Leaders

CIA Maps

Center for the Study of Intelligence

Freedom of Information Act

Electronic Reading Room

Kent Center

Occasional Papers

Intelligence Literature

Reports

Related Links

My Way Reader

Home » Library » Publications » The World Factbook

THE WORLD FACTBOOK

Please select a country to view

ABOUT REFERENCES APPENDICES FAQs CONTACT THE WORLD FACTBOOK ARCHIVE

COUNTRY COMPARISON :: GDP - PER CAPITA (PPP)

GDP - per capita (PPP) compares GDP on a purchasing power parity basis divided by population as of 1 July for the same year.

Filter by the Region: All

RANK	COUNTRY	GDP - PER CAPITA (PPP)	DATE OF INFORMATION
1	LIECHTENSTEIN	\$139,100	2009 est.
2	QATAR	\$124,100	2017 est.
3	MACAU	\$122,900	2018
4	MONACO	\$115,700	2015 est.

Figure 2.1 The CIA World Factbook home page (2019 access).

From *Data Analytics Toolkit* by Eric Gaze and McKenzie Lamb (ISBN 978-0-13-539786-2). Copyright © 2027 Pearson Education, Inc. All rights reserved.

	A	B	C	D
1	RANK	COUNTRY	GDP - PER CAPITA (PPP)	DATE OF INFORMATION
2	1	LIECHTENSTEIN	\$139,100	2009 est.
3	2	QATAR	\$124,100	2017 est.
4	3	MACAU	\$122,000	2018
5	4	MONACO	\$115,700	2015 est.
6	5	LUXEMBOURG	\$105,100	2017 est.

	A	B	C	D
1	RANK	COUNTRY	GDP - PER	DATE OF IN
2	1	LIECHTEN	139100	2009 est.
3	2	QATAR	124100	2017 est.
4	3	MACAU	122000	2018
5	4	MONACO	115700	2015 est.
6	5	LUXEMBO	105100	2017 est.

	A	B	C	D	E	F	G	H
1	name	slug	value	date_of_inf	ranking	region		
2	Monaco	monaco	\$270,100	2024	1	Europe		
3	Liechtenst	liechtenste	\$210,600	2024	2	Europe		
4	Singapore	singapore	\$132,600	2024	3	East and Southeast Asia		
5	Luxembou	luxembour	\$128,200	2024	4	Europe		
6	Ireland	ireland	\$115,300	2024	5	Europe		
7	Macau	macau	\$112,800	2024	6	East and Southeast Asia		
8	Qatar	qatar	\$110,900	2024	7	Middle East		
9	Bermuda	bermuda	\$105,300	2024	8	North America		
10	Norway	norway	\$91,100	2024	9	Europe		
11	Switzerland	switzerland	\$82,000	2024	10	Europe		

Figure 2.2 Formatting comes with the paste into Excel.

Country Comparison :: GDP - per capita (PPP)

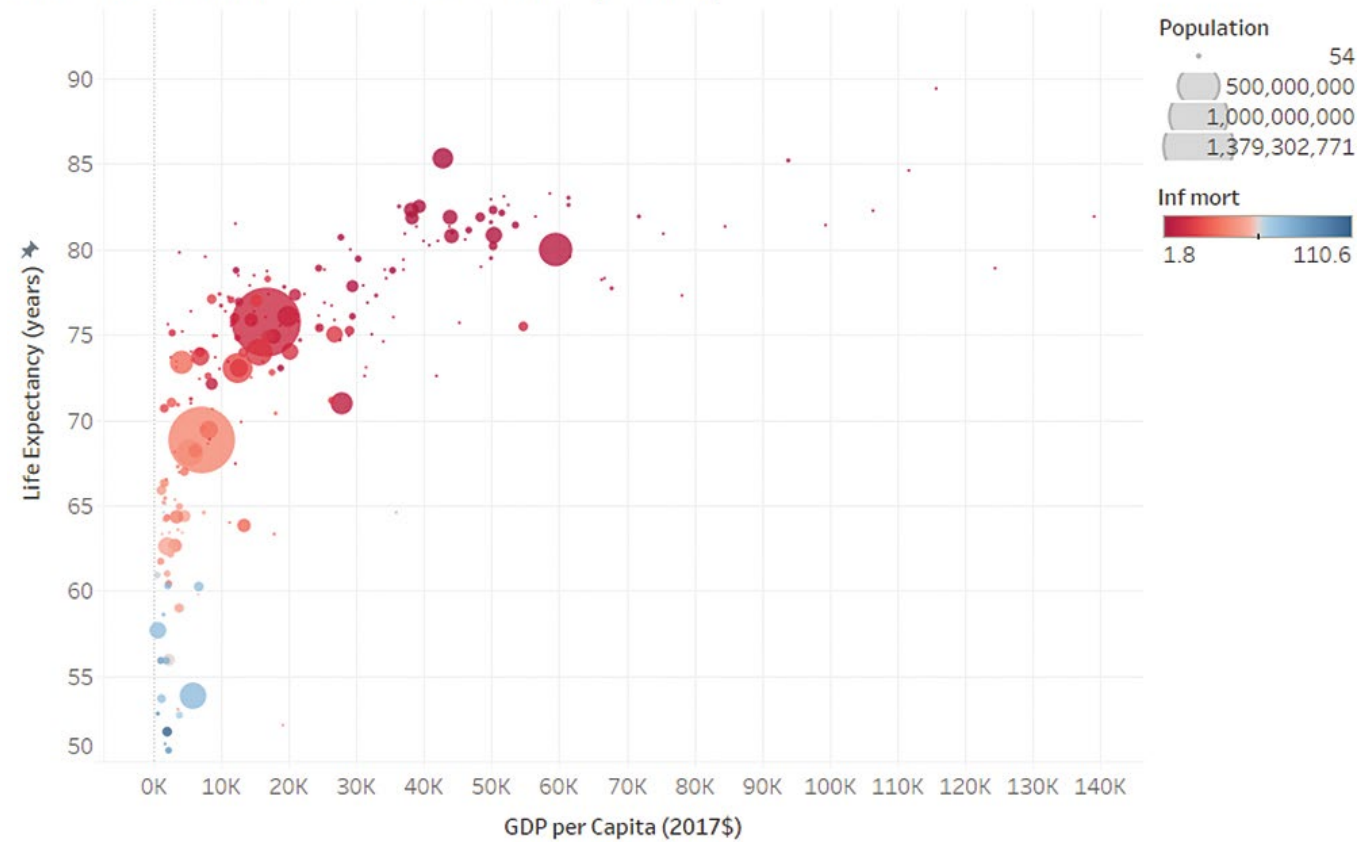
1	Liechtenstein	\$139,100	2009 est.
2	Qatar	\$124,100	2017 est.
3	Macau	\$122,000	2018
4	Monaco	\$115,700	2015 est.
5	Luxembourg	\$105,100	2017 est.
6	Bermuda	\$99,400	2016 est.
7	Singapore	\$94,100	2017 est.
8	Isle of Man	\$84,600	2014 est.
9	Brunei	\$78,900	2017 est.
10	Ireland	\$73,200	2017 est.
11	Norway	\$72,100	2017 est.
12	Falkland Islands (Islas Malvinas)	\$70,800	2015 est.
13	United Arab Emirates	\$68,600	2017 est.
14	Sint Maarten	\$66,800	2014 est.

Figure 2.3 The data in the webpage format.

	A	B	C	D	E	F
1	GDP_rank	Country	GDP_per_cap	IM_rank	Country	Inf_mort
2	207	Afghanistan	2000	1	Afghanistan	110.6
3	125	Albania	12500	124	Albania	11.9
4	111	Algeria	15200	81	Algeria	19.6
5	136	American Samoa	11200	128	American Samoa	11.3
6	32	Andorra	49900	199	Andorra	3.6
7	161	Angola	6800	12	Angola	67.6
8	130	Anguilla	12200	211	Anguilla	3.3
9	79	Antigua and Barbuda	26300	118	Antigua and Barbuda	12.1
10	89	Argentina	20900	139	Argentina	9.8

Figure 2.4 Pasting datasets into one file.

Life Expectancy vs GDP per Capita (2017\$)



GDP per cap vs. Life exp. Color shows Inf mort. Size shows Population. Details are shown for Country.

Figure 2.5 Life expectancy versus GDP per capita with population (size) and infant mortality rates (color).

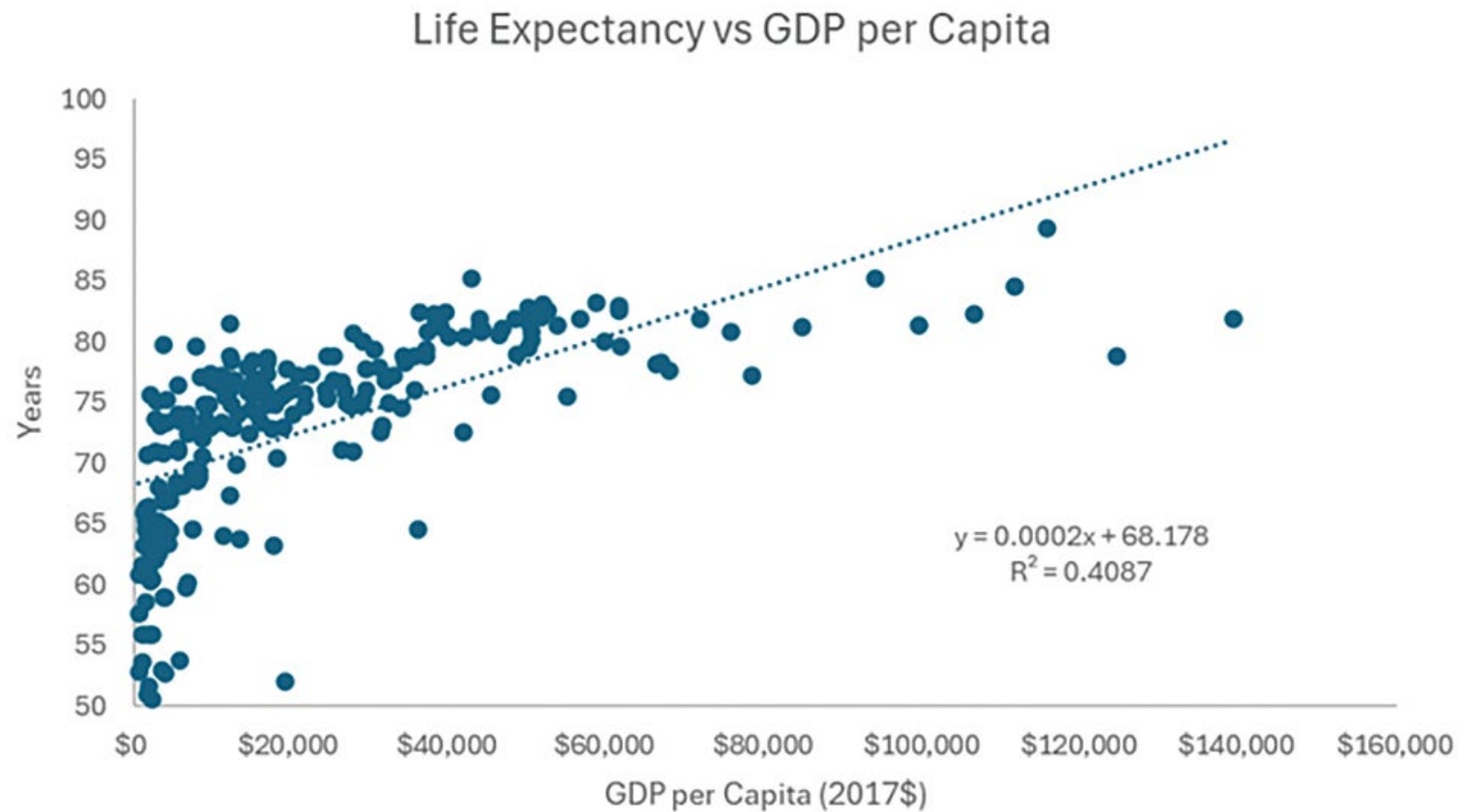


Figure 2.6 Bivariate scatterplot showing life expectancy against GDP per capita.

Life Expectancy versus GDP per Capita



Figure 2.7 Using a logarithmic scale on the x-axis.

Map of Life Expectancy

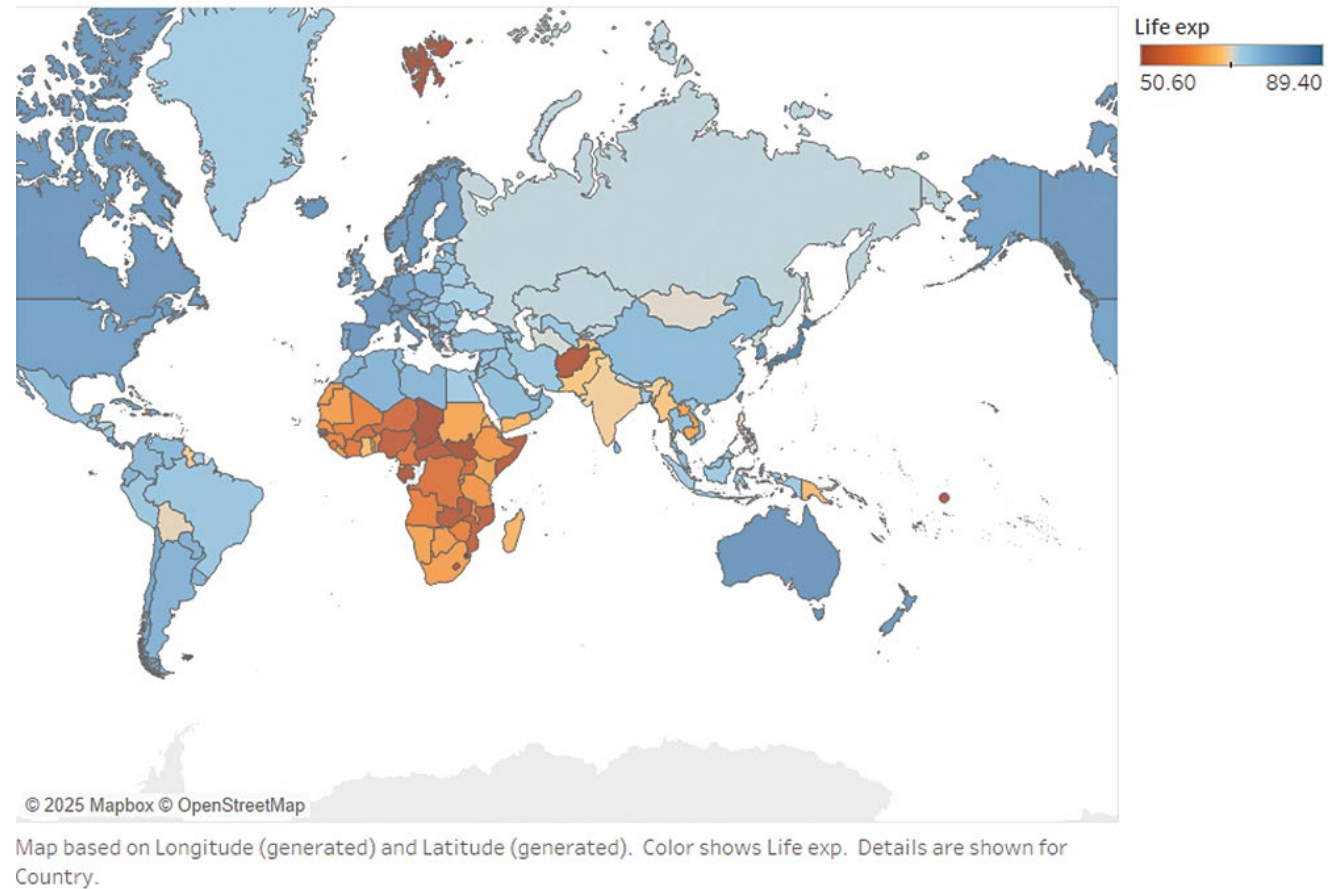


Figure 2.8 A choropleth map showing life expectancy by country.

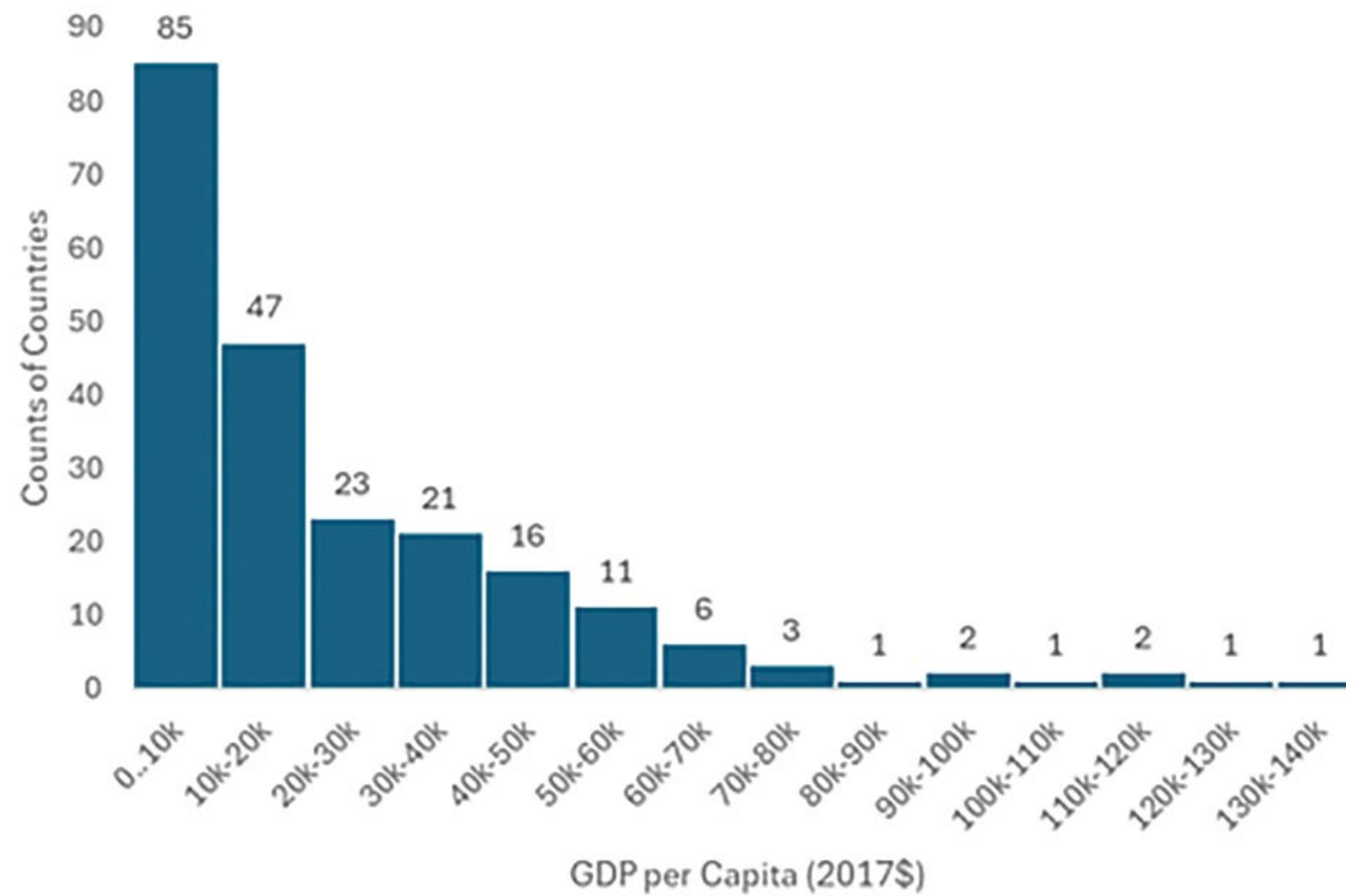


Figure 2.9 Distribution of GDP per capita values displayed in a histogram.

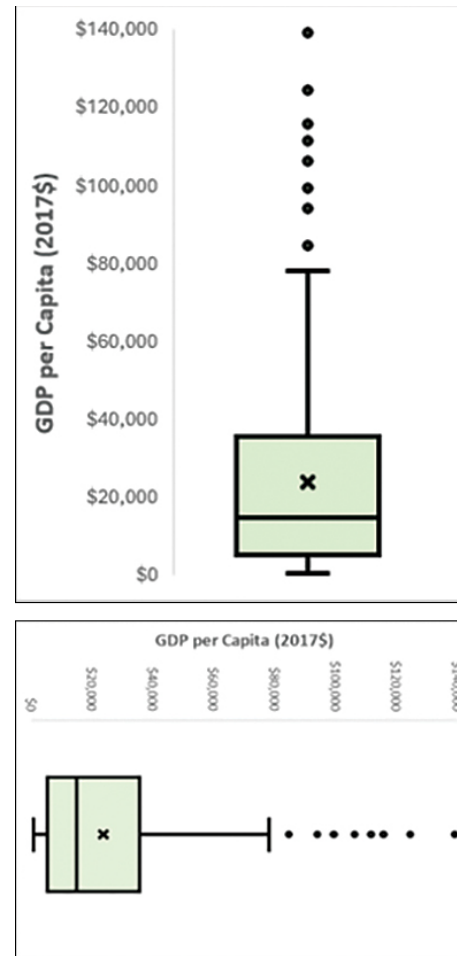


Figure 2.10 A boxplot of GDP per capita: vertical and horizontal options.

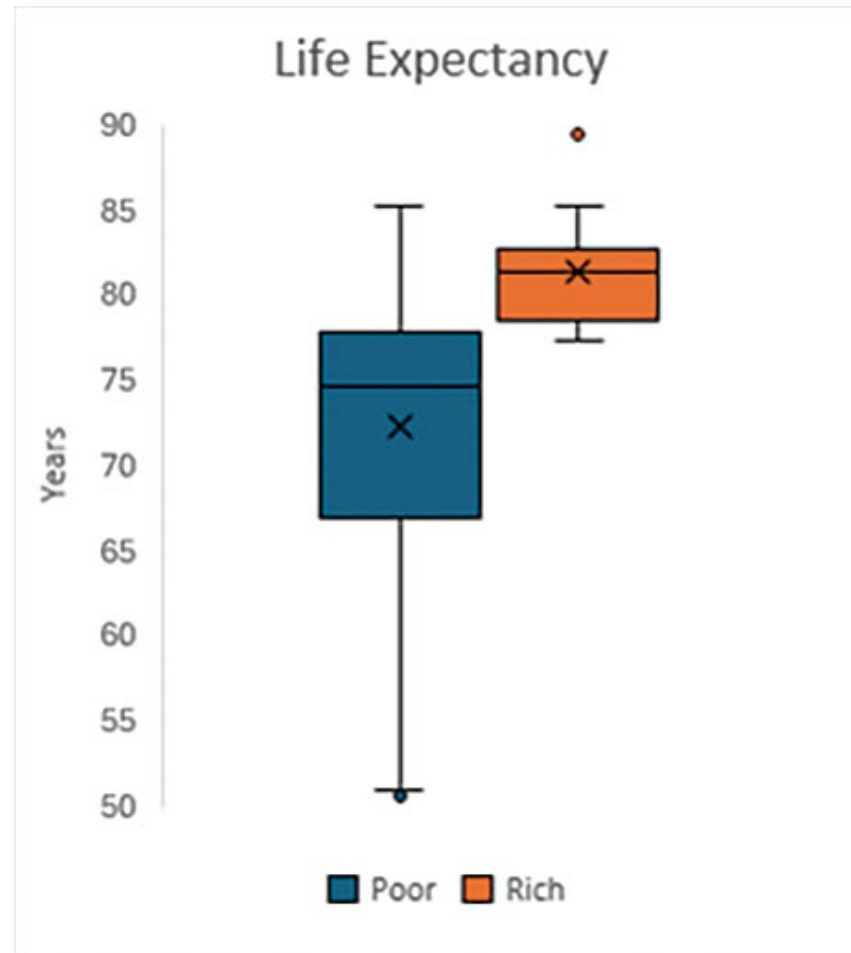


Figure 2.11 Life expectancy by wealth, using \$60,000 GDP per capita as the cutoff between rich and poor countries.

AutoSave On Countries_joined_2 No Label • Last Modified: Just now Search

File Home Insert Draw Page Layout Formulas Data Review View Automate Help Acrobat

Clipboard Font Alignment Number Styles Cells Editing Sensitivity Add

Z24

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	GDP_rank	Country	GDP_per_c	LE_rank	Life_exp	IM_rank	Inf_mort	Pop_rank	Population			GDP_per_cap	Life_exp	
2	1	Liechtenst	139100	19	81.9	190	4.2	213	38244			\$139,100	81.9	
3	2	Qatar	124500	52	78.9	167	6.2	143	2314307			\$124,500	78.9	
4	3	Monaco	115700	1	89.4	225	1.8	217	30645			\$115,700	89.4	
5	4	Macau	111600	4	84.6	214	3.1	170	601969			\$111,600	84.6	
6	5	Luxembou	106300	13	82.3	209	3.4	171	594130			\$106,300	82.3	
7	6	Bermuda	99400	26	81.4	219	2.5	203	70864			\$99,400	81.4	
8	7	Singapore	93900	3	85.2	222	2.4	113	5888926			\$93,900	85.2	
9	8	Isle of Man	84600	29	81.3	193	4	199	88815			\$84,600	81.3	
10	9	Barbados	78800	73	77.2	140	8.0	174	442502			\$78,800	77.2	

Figure 2.13 The cleaned data file with two columns ready for inserting a scatterplot.

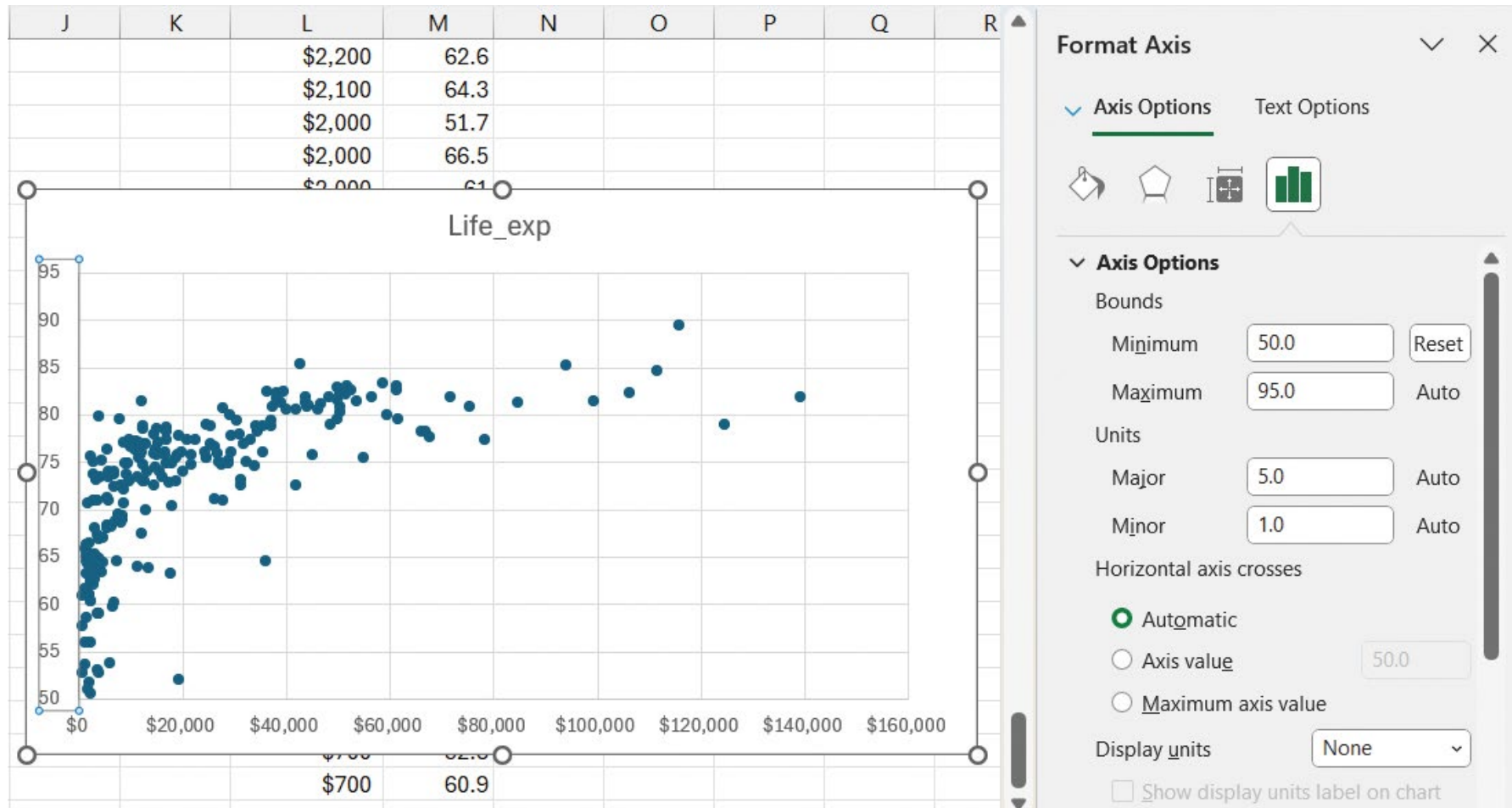


Figure 2.14 The scatterplot in Excel.

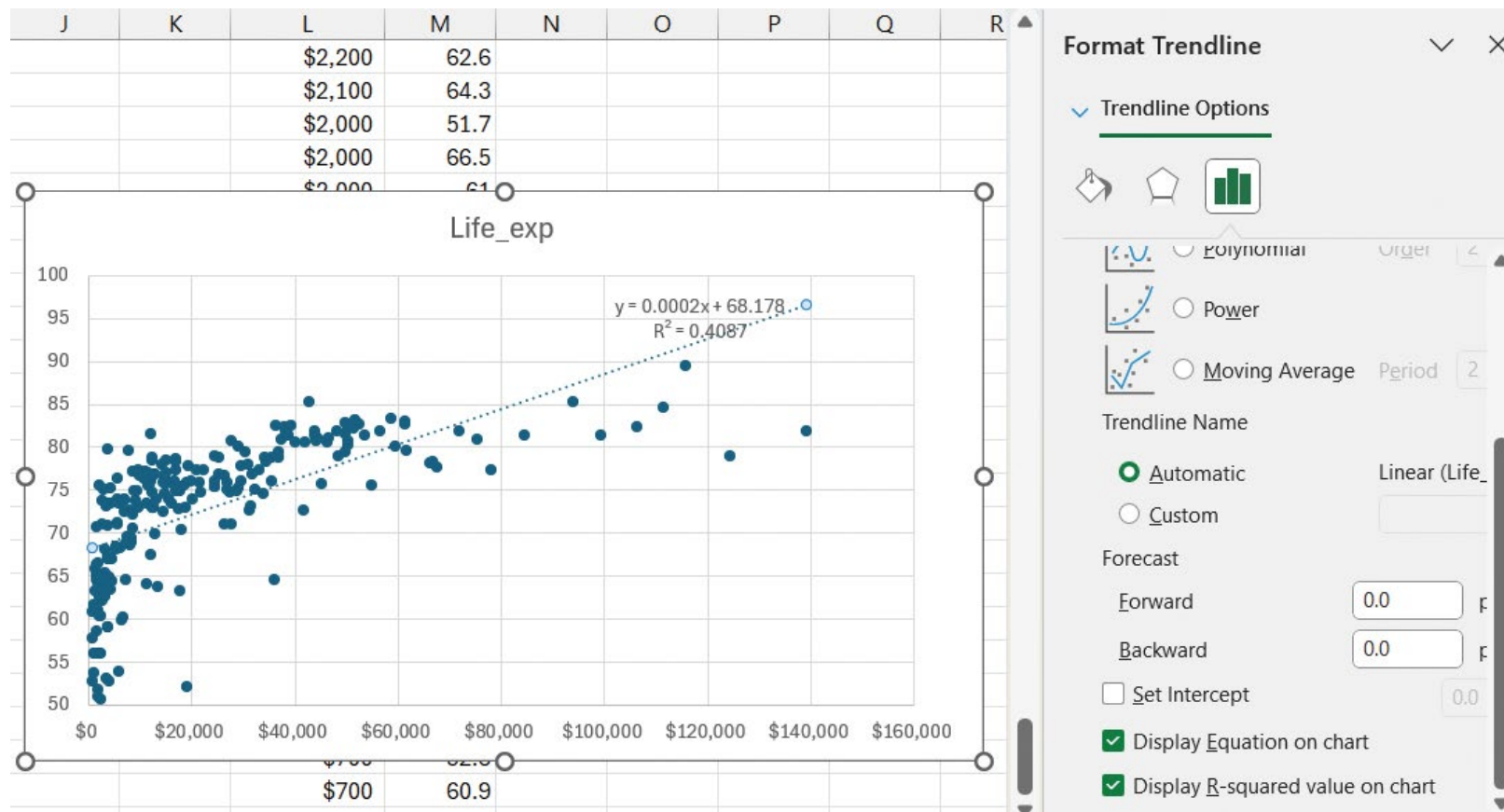


Figure 2.15 Adding the trendline and displaying the equation and R^2 value.

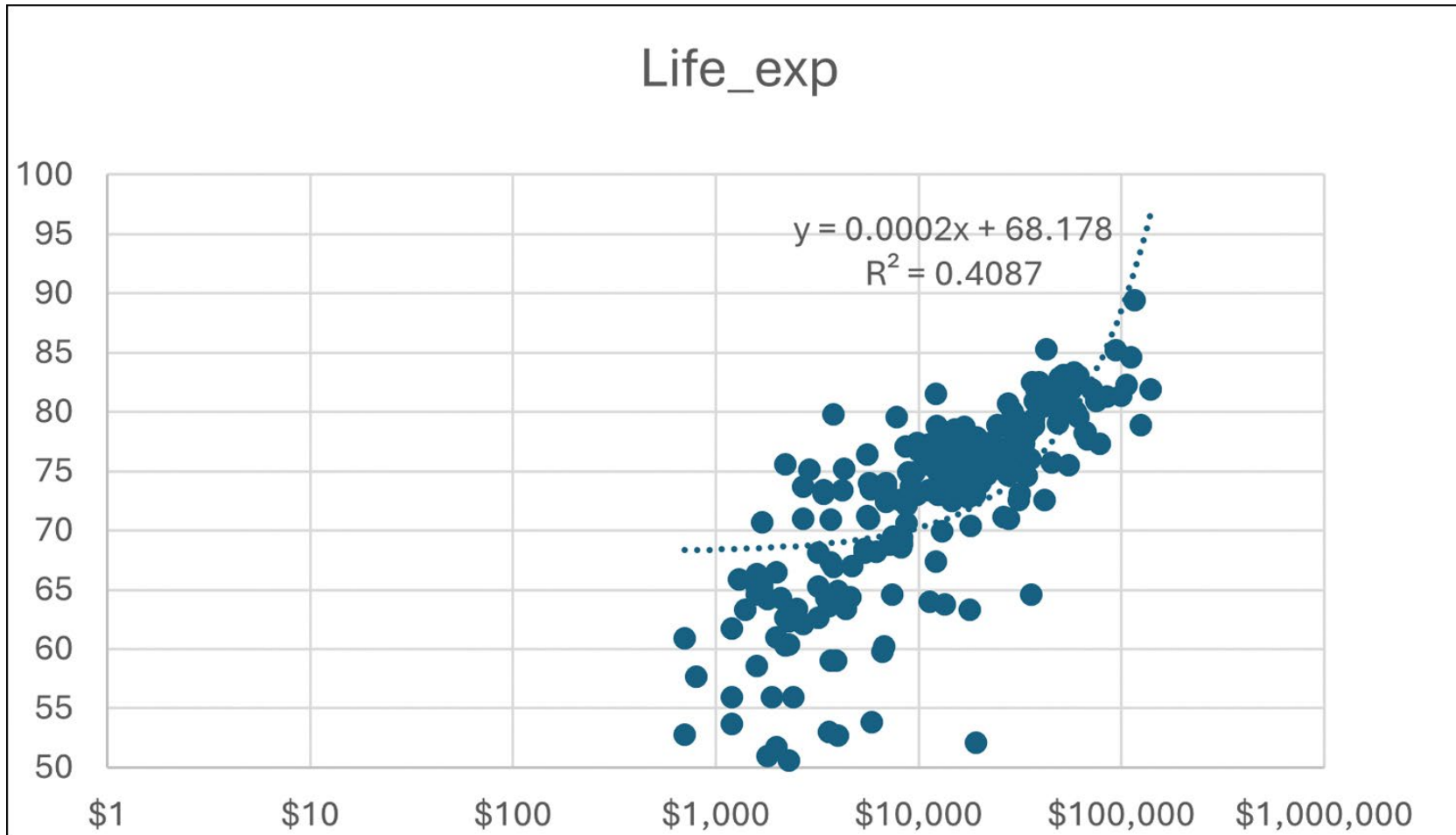


Figure 2.16 Logarithmic scale added to the x-axis.

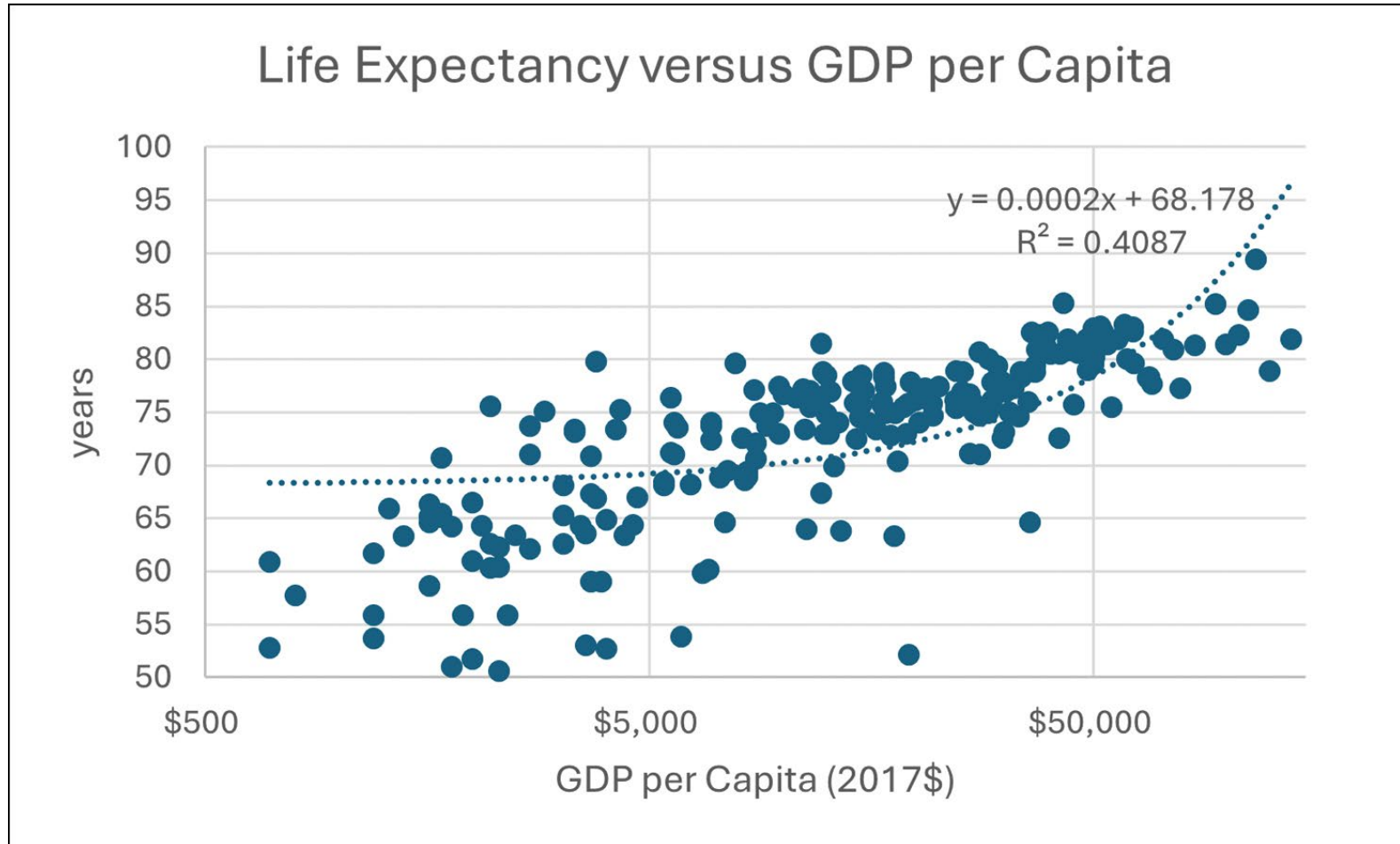


Figure 2.17 Changing the scale of the x-axis.

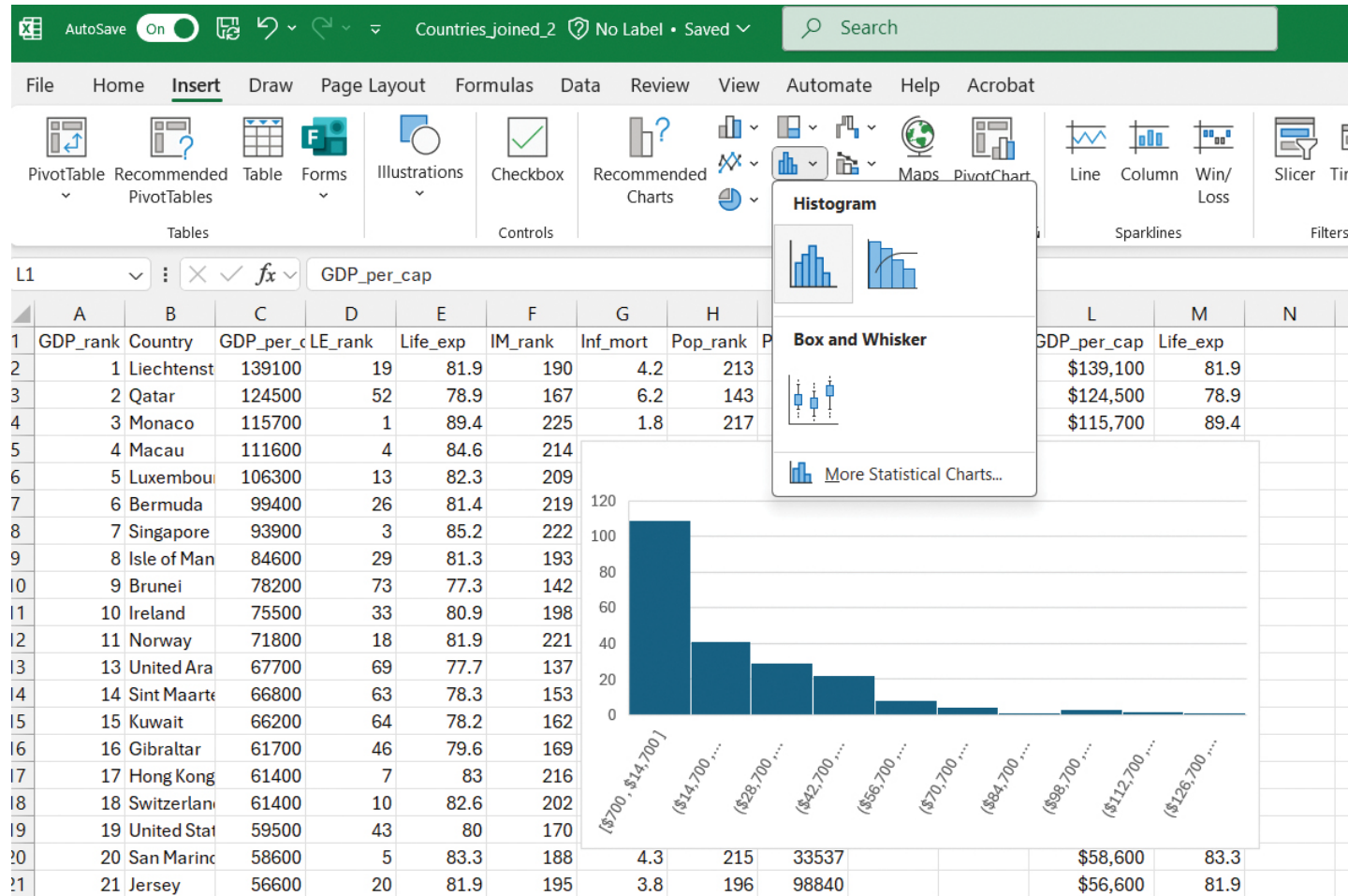


Figure 2.18 The automatic histogram using the built-in statistical charts.

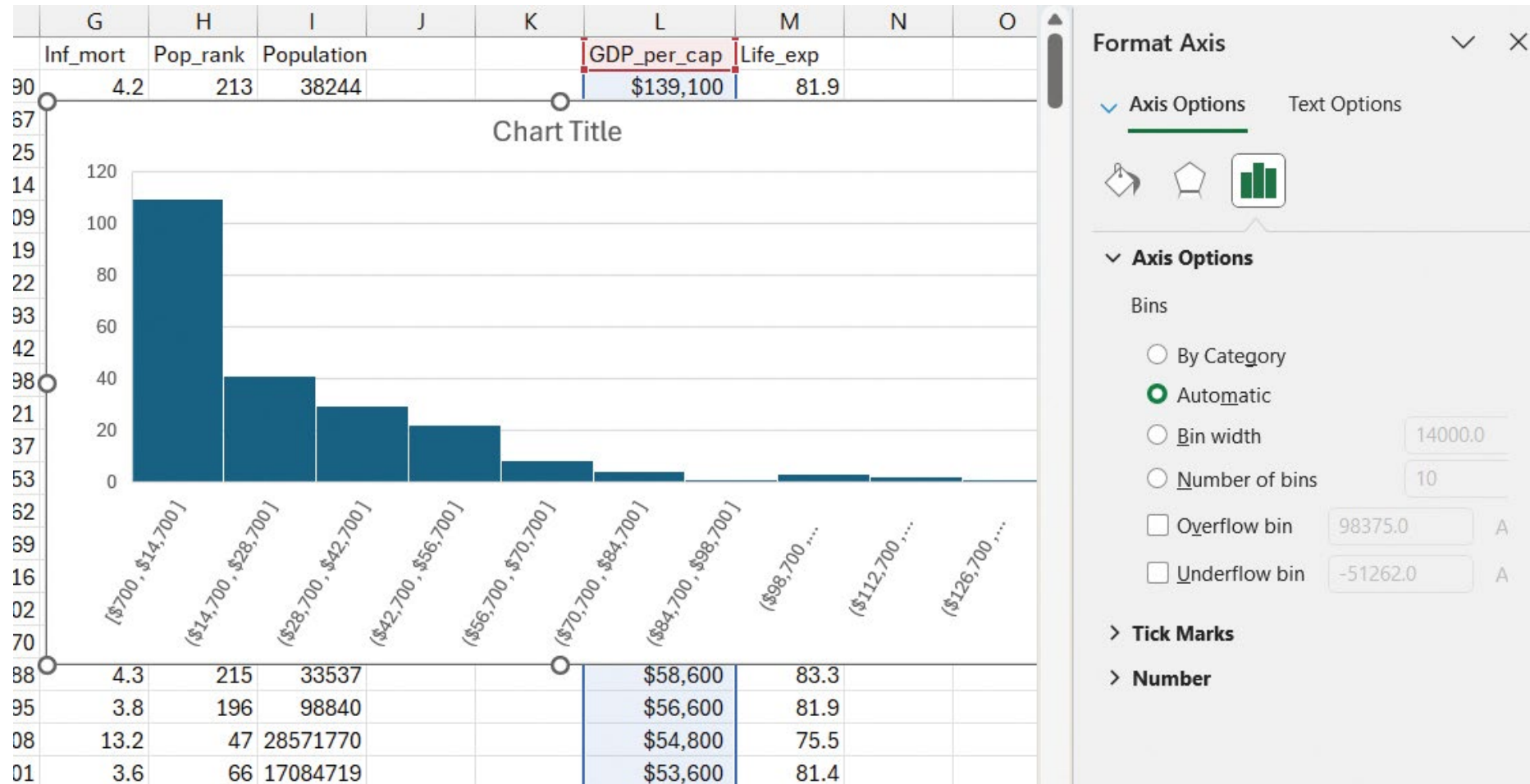


Figure 2.19 Histogram axis options include bin width and number of bins.

R	S	T	U
Cutoffs	COUNTIFS	Bins	Counts
0	85	0..10k	85
10000	132	10k-20k	=S4-S3
20000	155	20k-30k	23
30000	176	30k-40k	21

L	M	N	O	P	Q	R	S	T	U
GDP_per_cap	Life_exp								
\$139,100	81.9					Cutoffs	COUNTIFS	Bins	Counts
\$124,500	78.9					0	=COUNTIF(\$L\$2:\$L\$221,"<"&R4)		
\$115,700	89.4					10000	COUNTIF(range, criteria)		47
\$111,600	84.6					20000	155	20k-30k	23
\$106,300	82.3					30000	176	30k-40k	21
\$99,400	81.4					40000	192	40k-50k	16
\$93,900	85.2					50000	203	50k-60k	11

Figure 2.20 Counting the number of data values in column L below the cutoffs in column R.

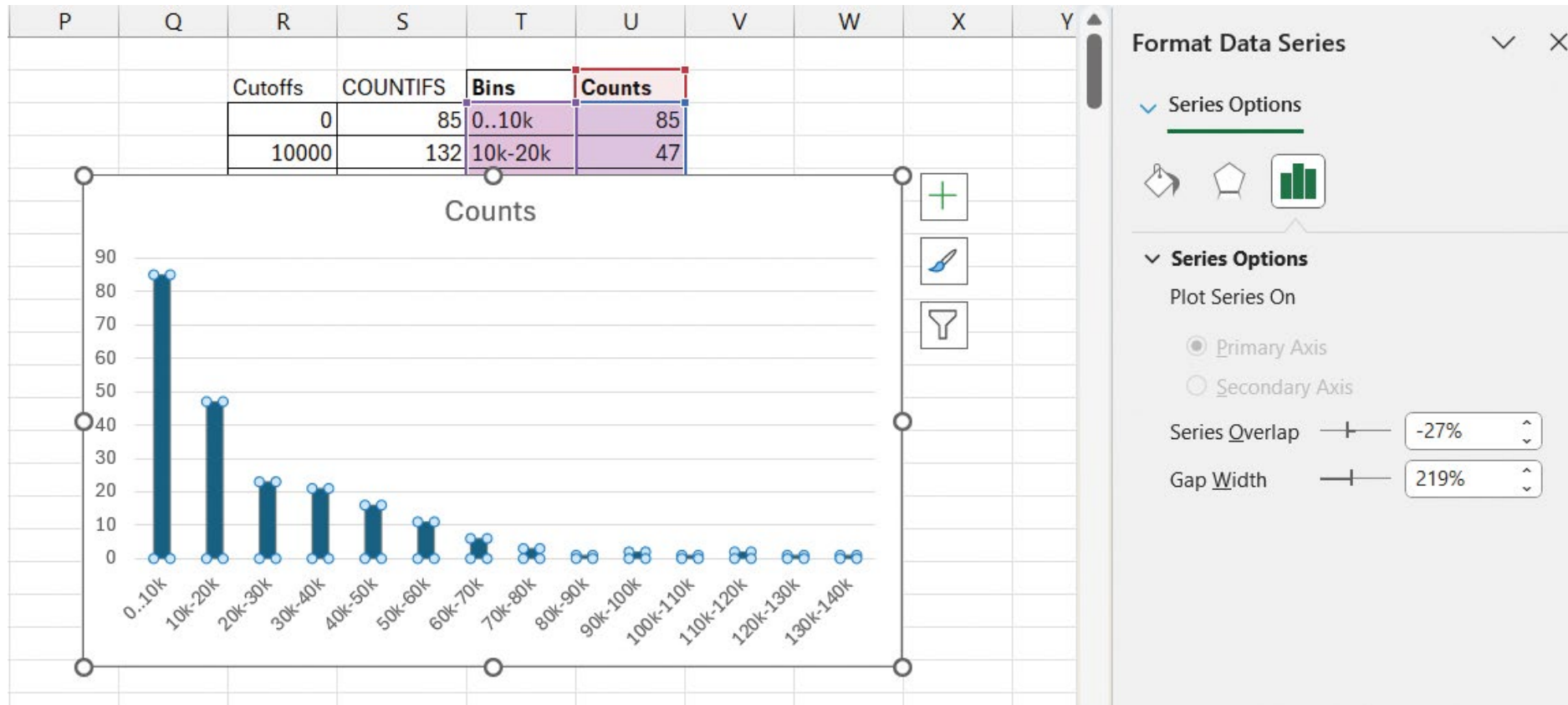


Figure 2.21 Insert a column chart if you are manually creating bins and counting values for the histogram.

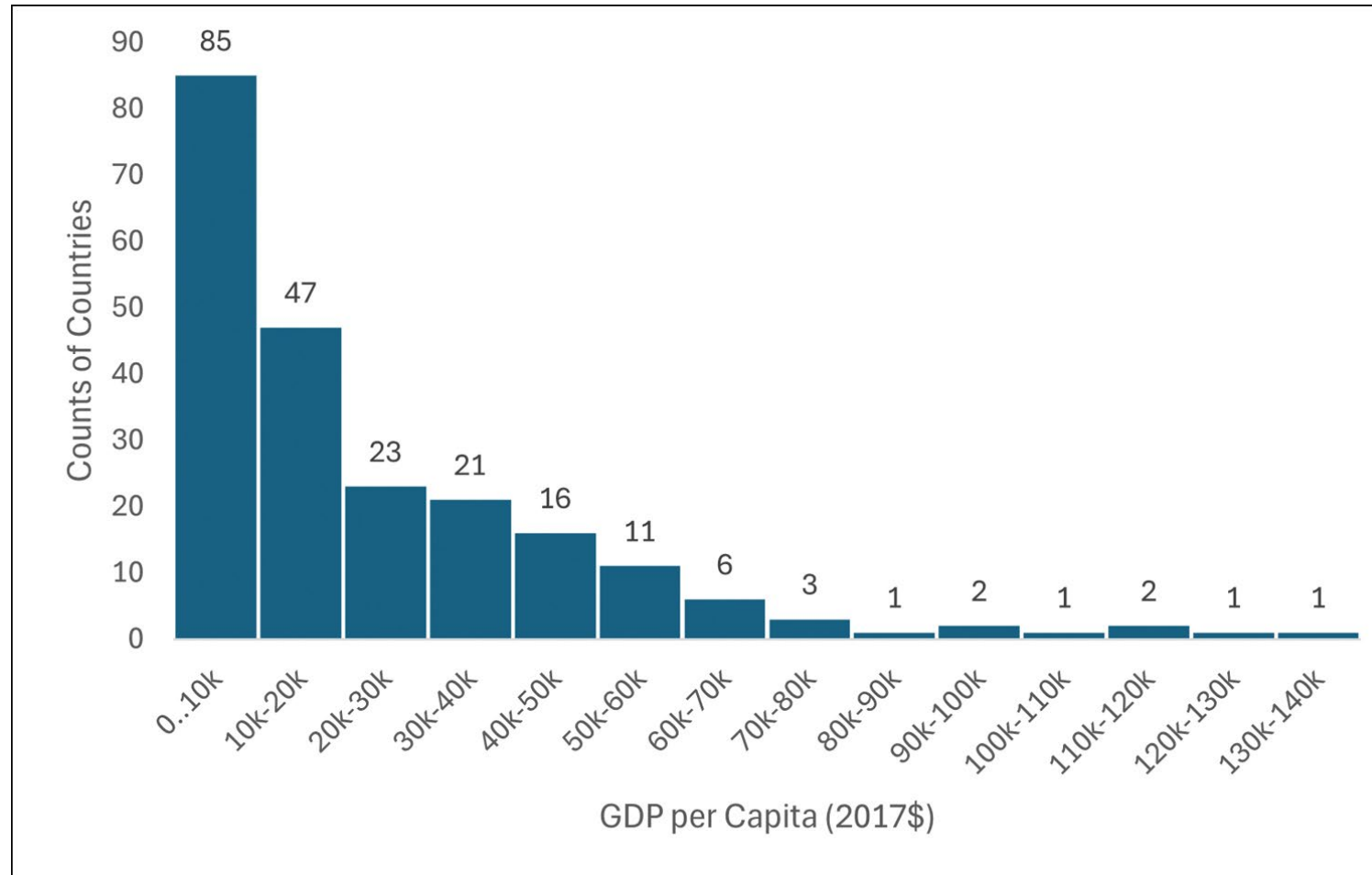


Figure 2.22 The completed histogram.

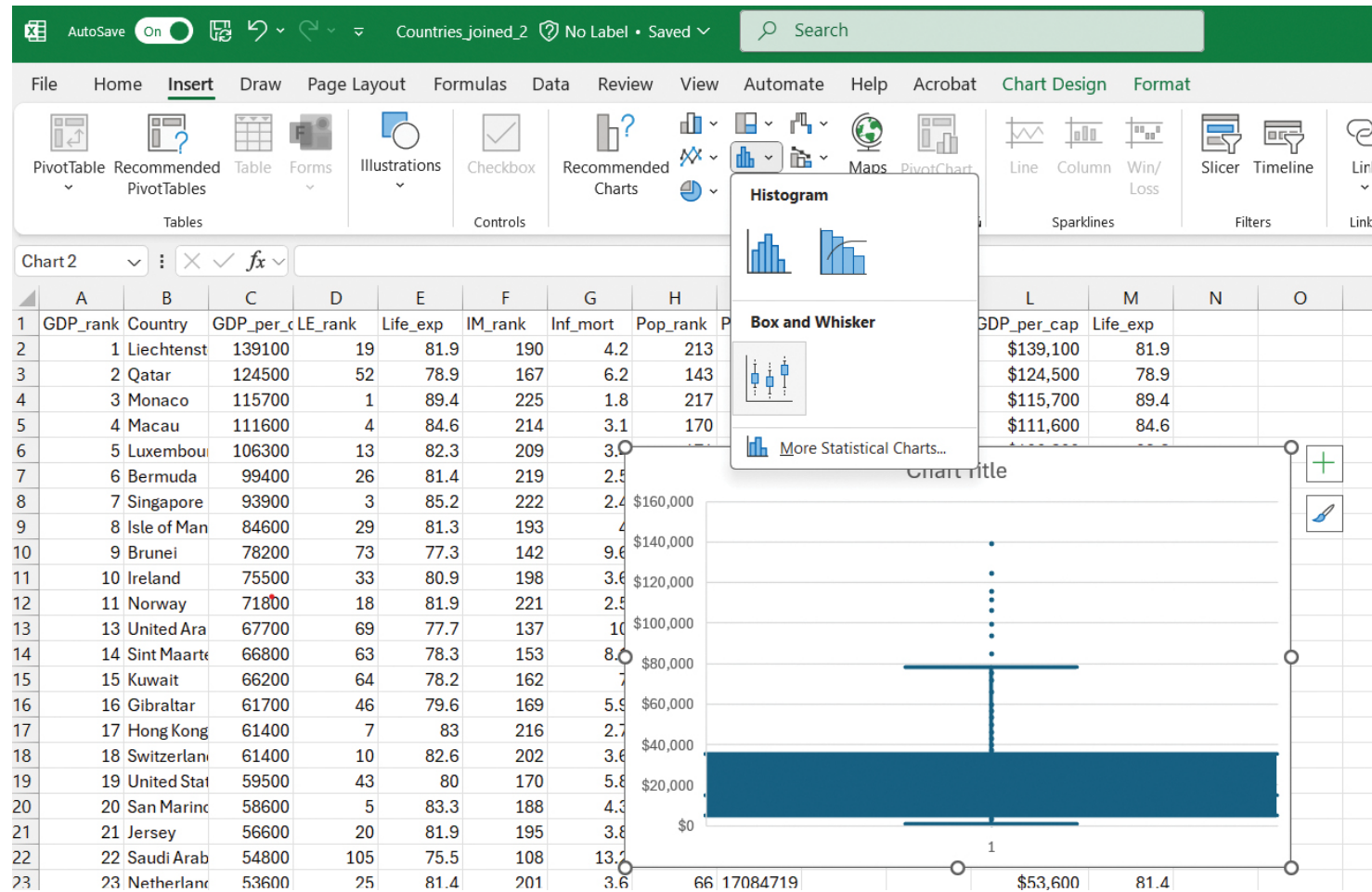


Figure 2.23 The box-and-whisker plot.

LOGEST ⌵ ⋮ ✗ ✓ <i>fx</i> ⌵ =IF(B2>60000, "Rich", "Poor")								
	A	B	C	D	E	F	G	
1	Country	GDP_per_c	Life_exp	Wealth				
2	Liechtenst	139100	81.9	=IF(B2>60000, "Rich", "Poor")				
3	Qatar	124500	78.9	R	IF(logical_test, [value_if_true], [value_if_false])			
4	Monaco	115700	89.4	Rich				
5	Macau	111600	84.6	Rich				
6	Luxembou	106300	82.3	Rich				
7	Bermuda	99400	81.4	Rich				

Figure 2.24 The Wealth column using the IF function.

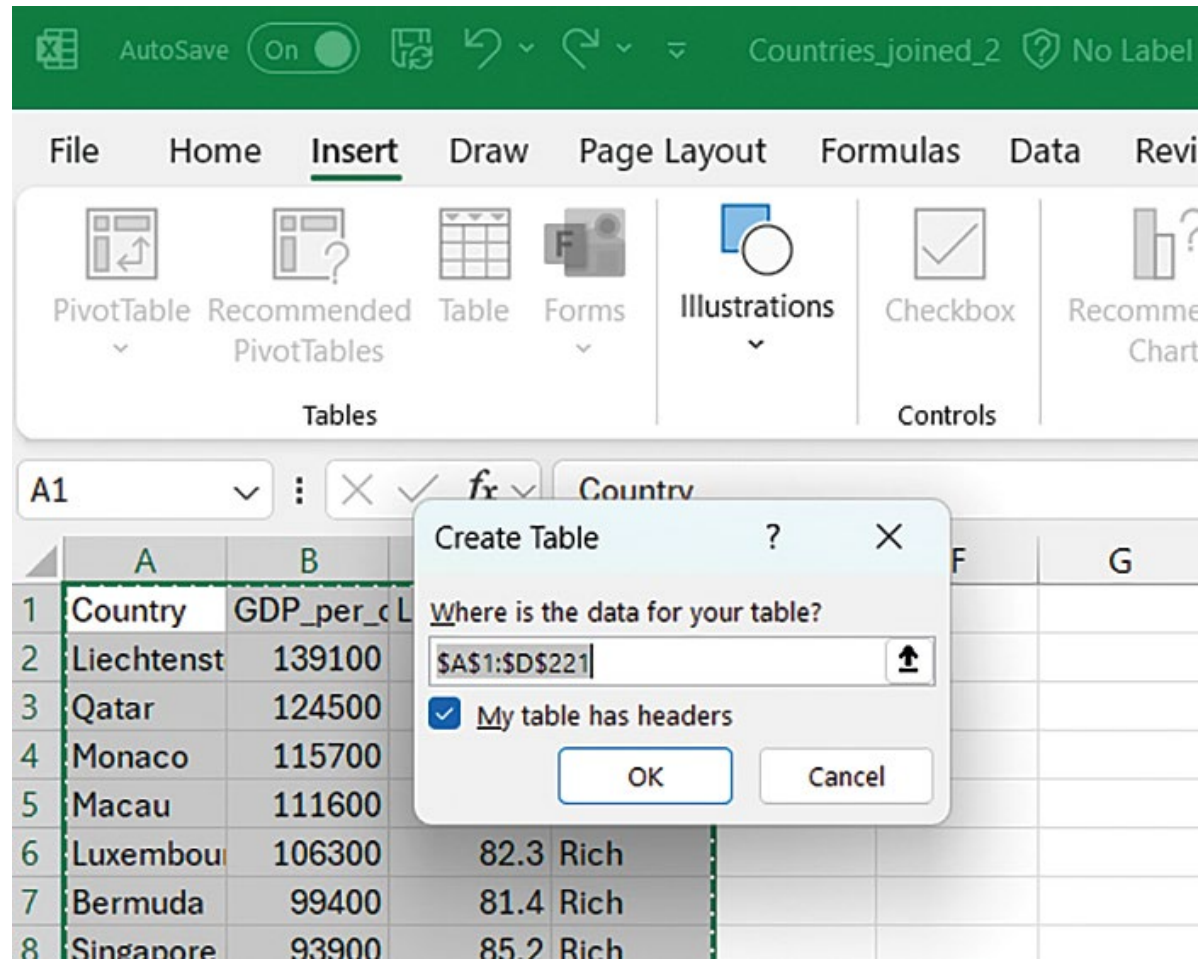


Figure 2.25 Making your data a table in Excel.

A2

</

Figure 2.26 The data as a table allowing for sorting by country.

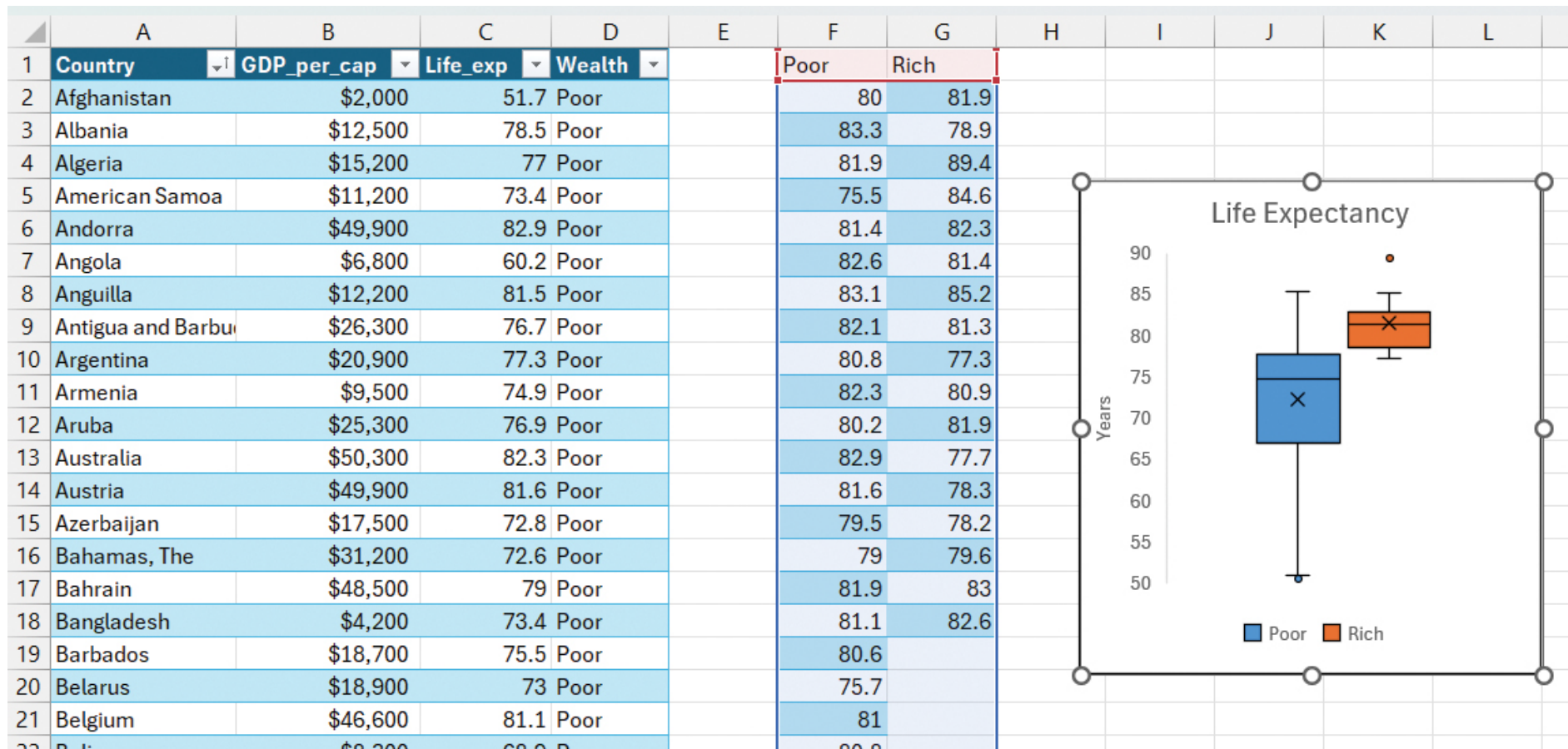


Figure 2.27 The boxplot showing life expectancy by wealth status of countries.

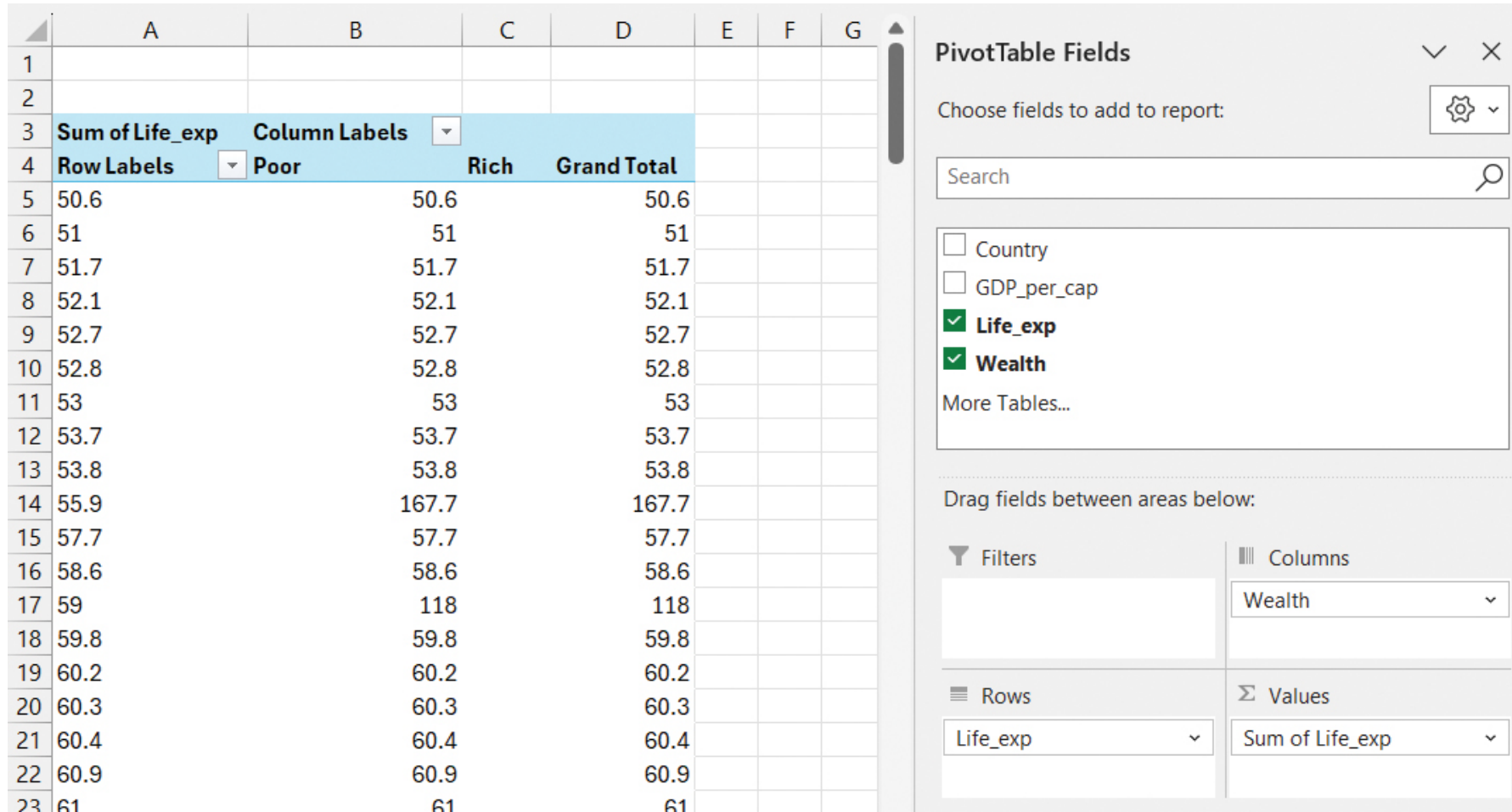


Figure 2.28 Sorting by wealth using a PivotTable.

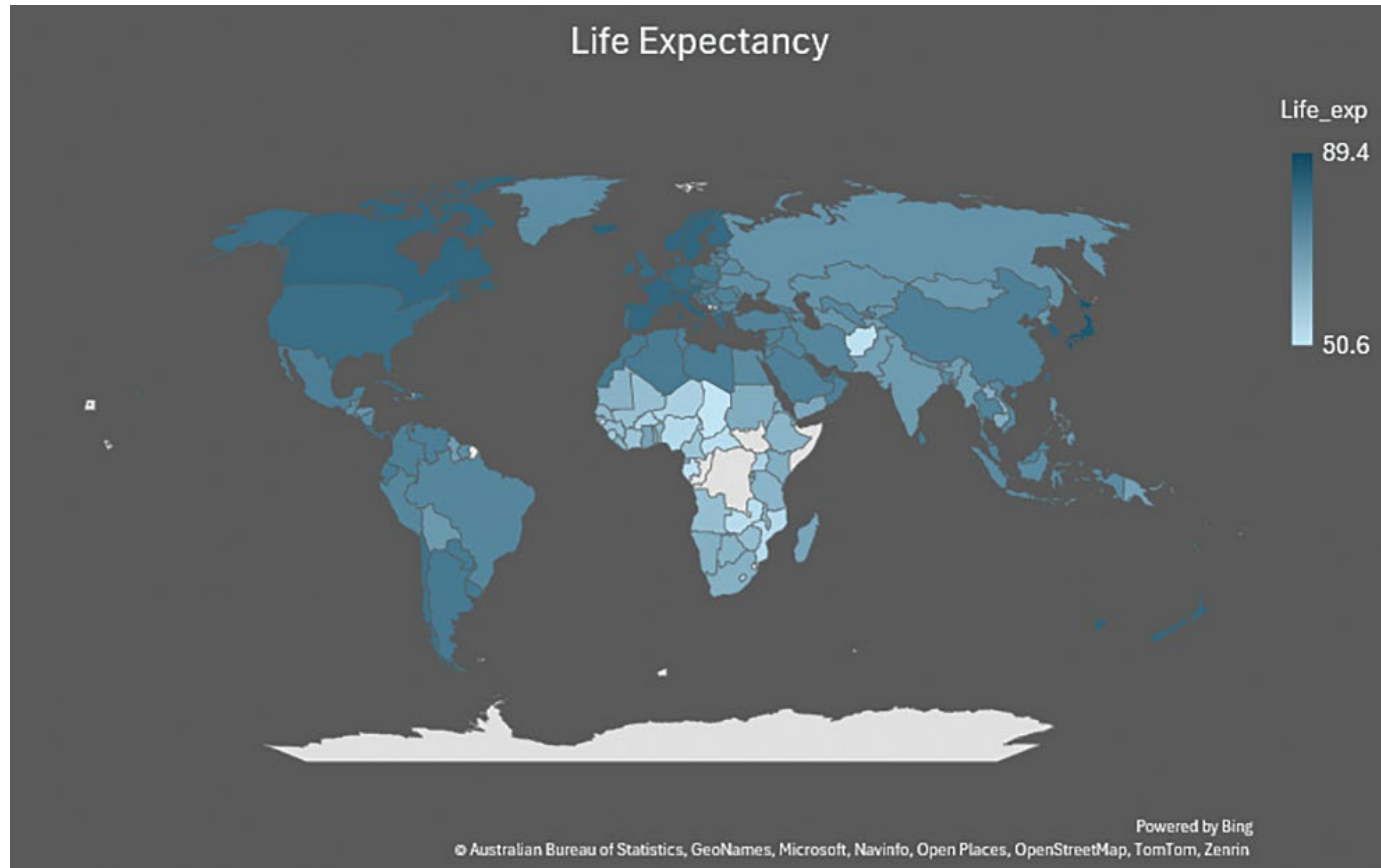


Figure 2.29 The choropleth or heat map showing life expectancy by country.

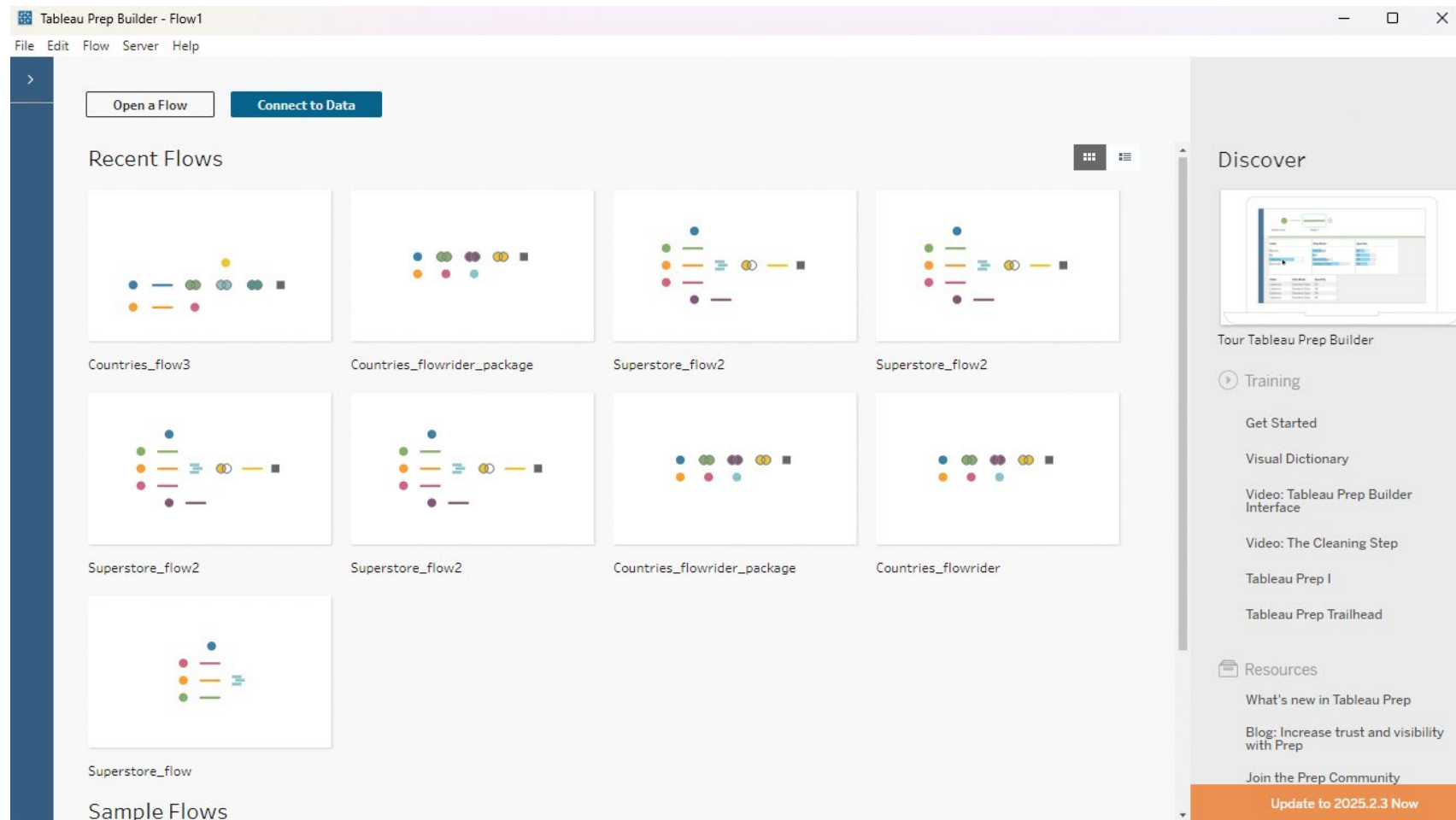


Figure 2.30 Tableau Prep Builder.

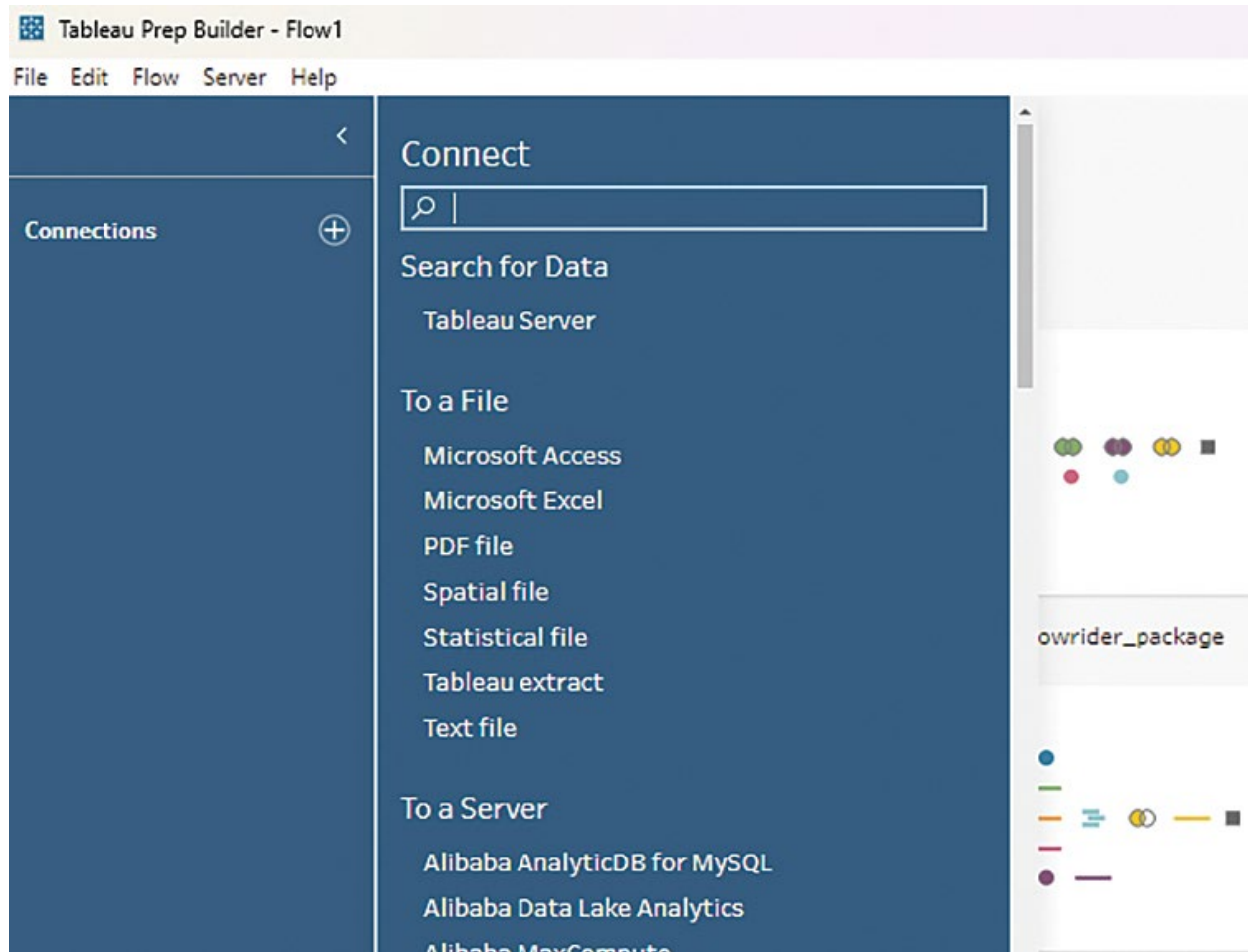


Figure 2.31 The Connections panel.

Tableau Prep Builder - Flow1*

File Edit Flow Server Help

Connections +

GDP.csv
Text file

Search

Tables

GDP

Input 4 fields Filter Values...

Settings Tables Data Sample Changes (1)

Connection Details

Text file

File

C:\Users\egaze\OneDrive - Bowdoin
College\Desktop\Excel_R_Tableau\Images\Countries\GDP.csv

Browse

Header Options

☐ Generate header (ex. F1, F2, F3)

☒ Set header

Header and Data Start Row

Header row

1

Data start row

2

3 Recommendations Data Preview Field List Search

Preview GDP 4 fields

	#	#	Abc	#
	Source Row Number	GDP_rank	Country	GDP_per_cap
1	1	1	Liechtenstein	139,100
2	2	2	Qatar	124,500
3	3	3	Monaco	115,700
4	4	4	Macau	111,600
5	5	5	Luxembourg	106,300
6	6	6	Bermuda	99,400
7	7	7	Singapore	93,900
8	8	8	Isle of Man	84,600
9	9	9	Brunei	78,200
10	10	10	Ireland	75,500
11	11	11	Norway	71,800
12	12	12	Falkland Islands (Islas...)	70,800
13	13	13	United Arab Emirates	67,700

Figure 2.32 The GDP data from the .csv file.

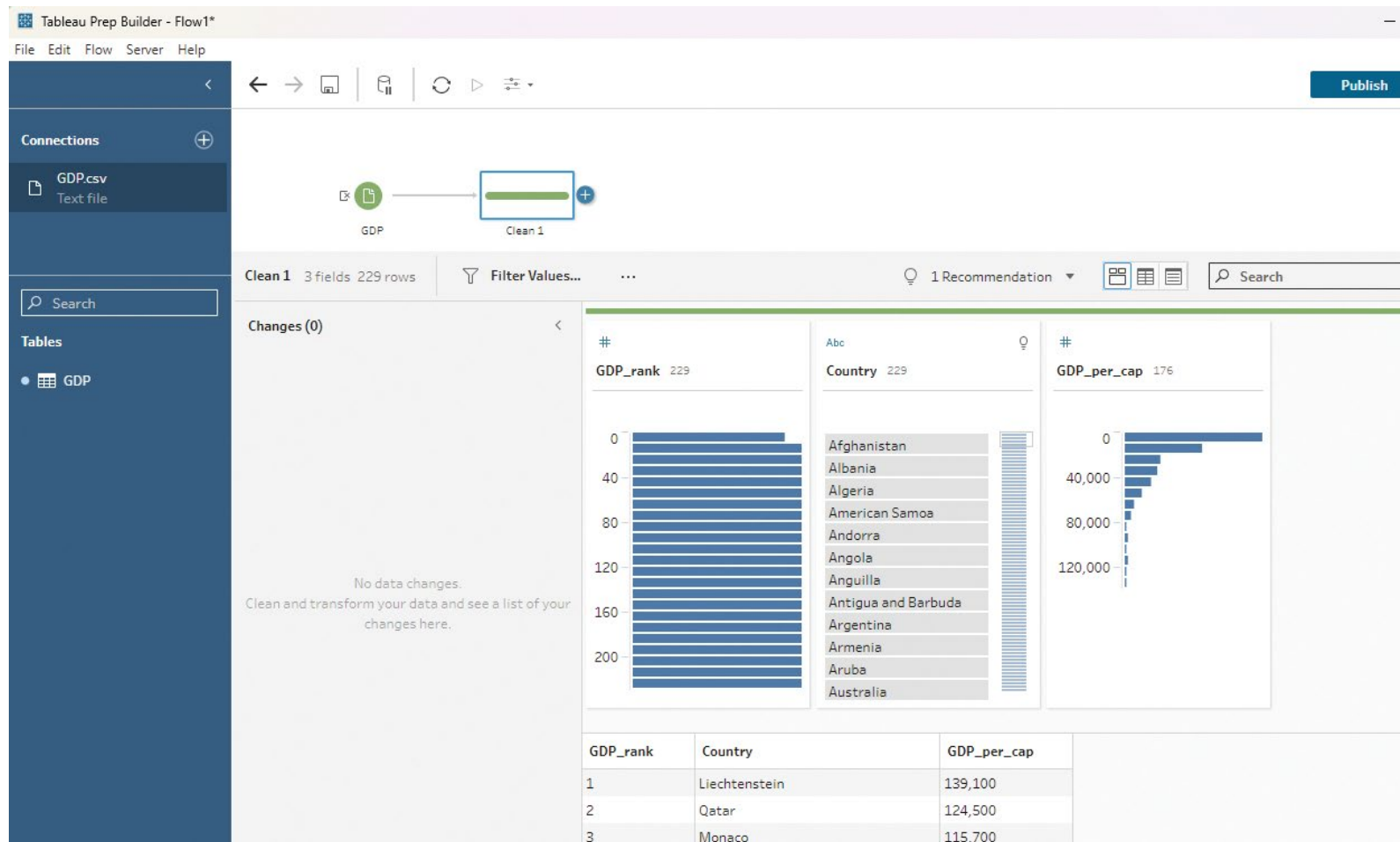


Figure 2.33 Click the + in the flow pane to clean the data.

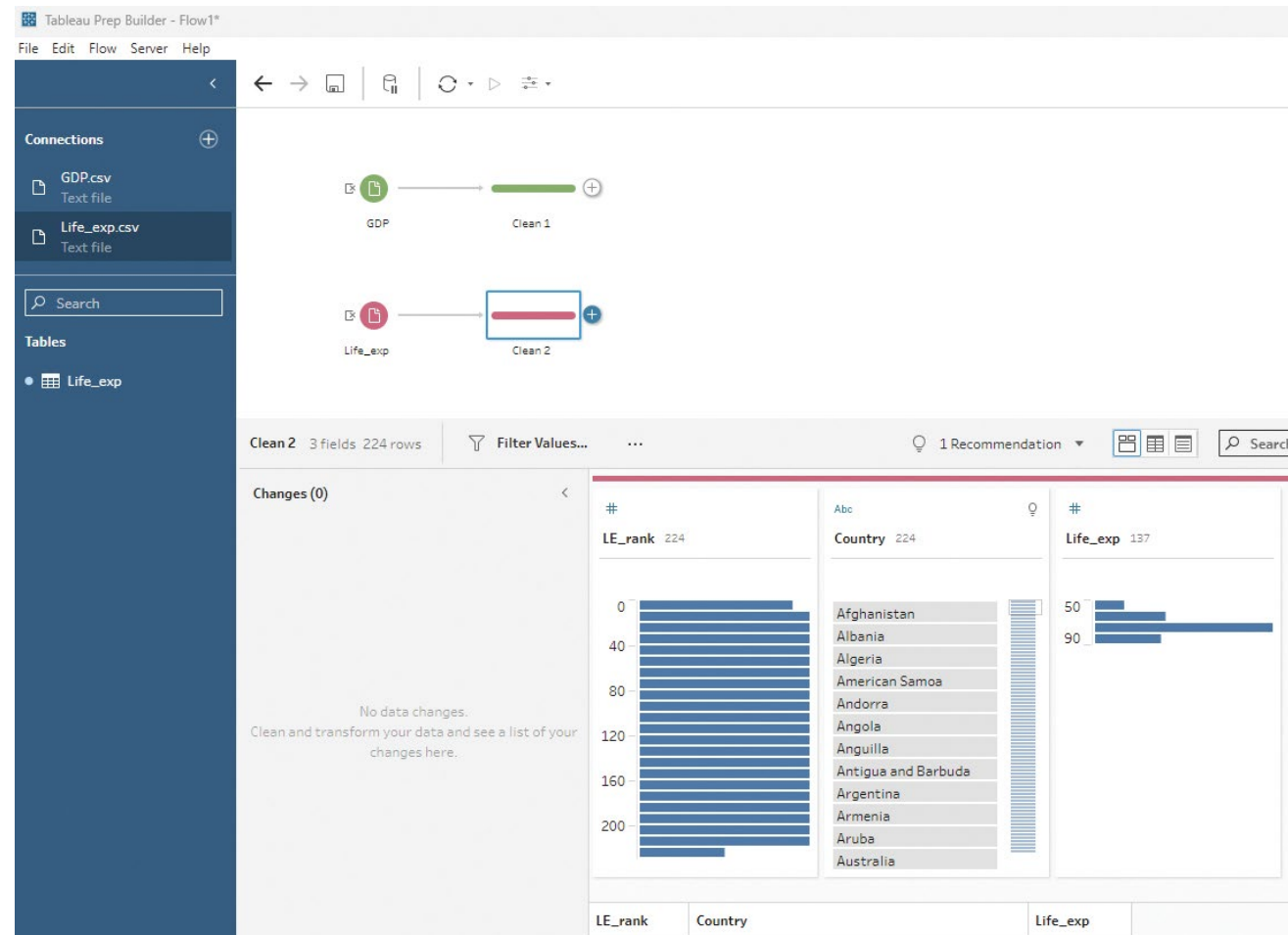


Figure 2.34 There are 224 unique countries in the life expectancy data.

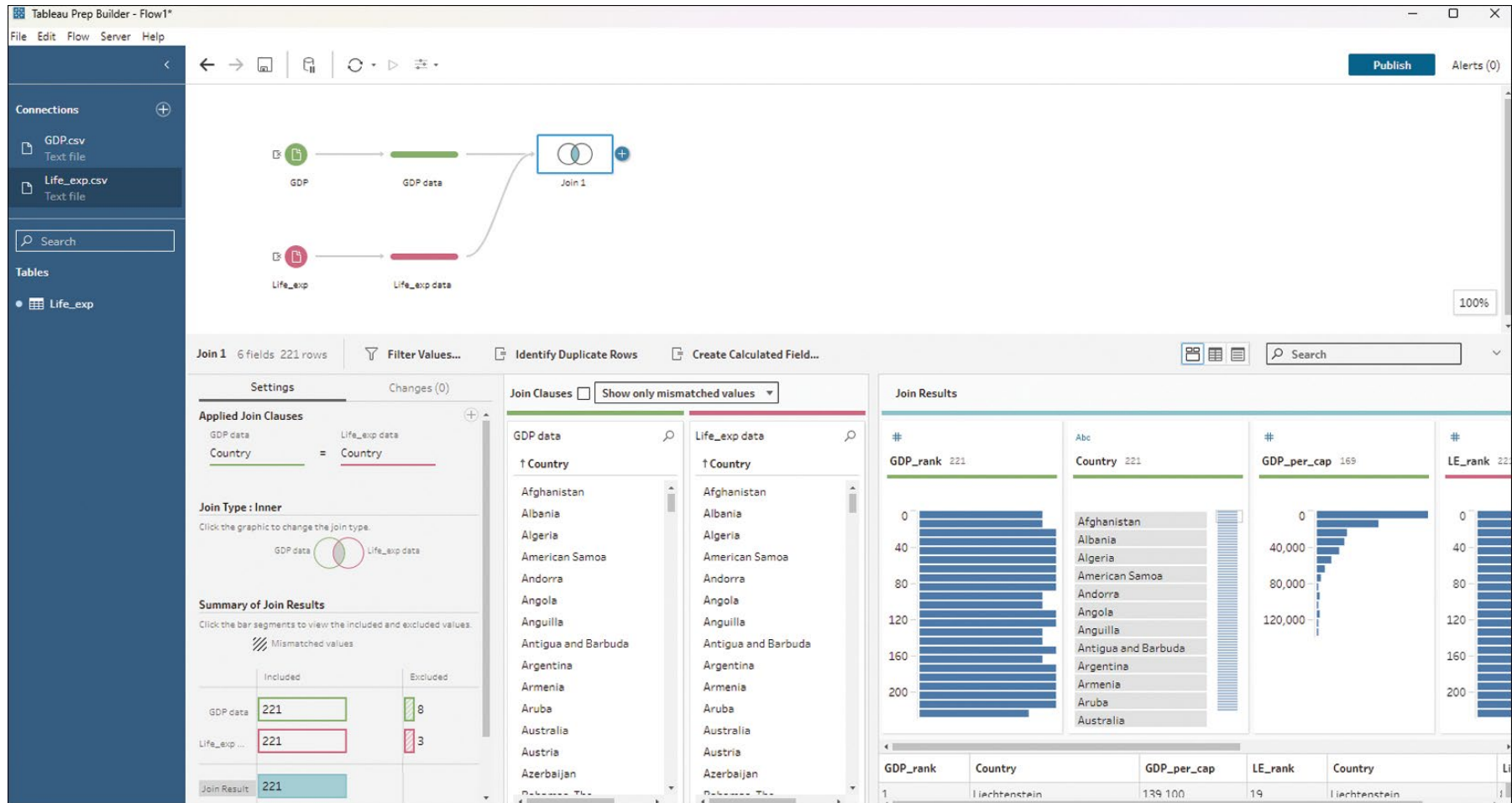


Figure 2.35 Choose join to merge the GDP and life expectancy data.

Join Clauses ☒ Show only mismatched values ▼	
GDP data	Life_exp data
↑ Country	↑ Country
Eswatini	Gaza Strip
Falkland Islands (Islas Malvina	Somalia
Kosovo	Swaziland
Montenegro	
Niue	
Saint Martin	
South Sudan	
Tokelau	

Figure 2.36 Checking “Show only mismatched values” to see the countries excluded from each data file in the merge.

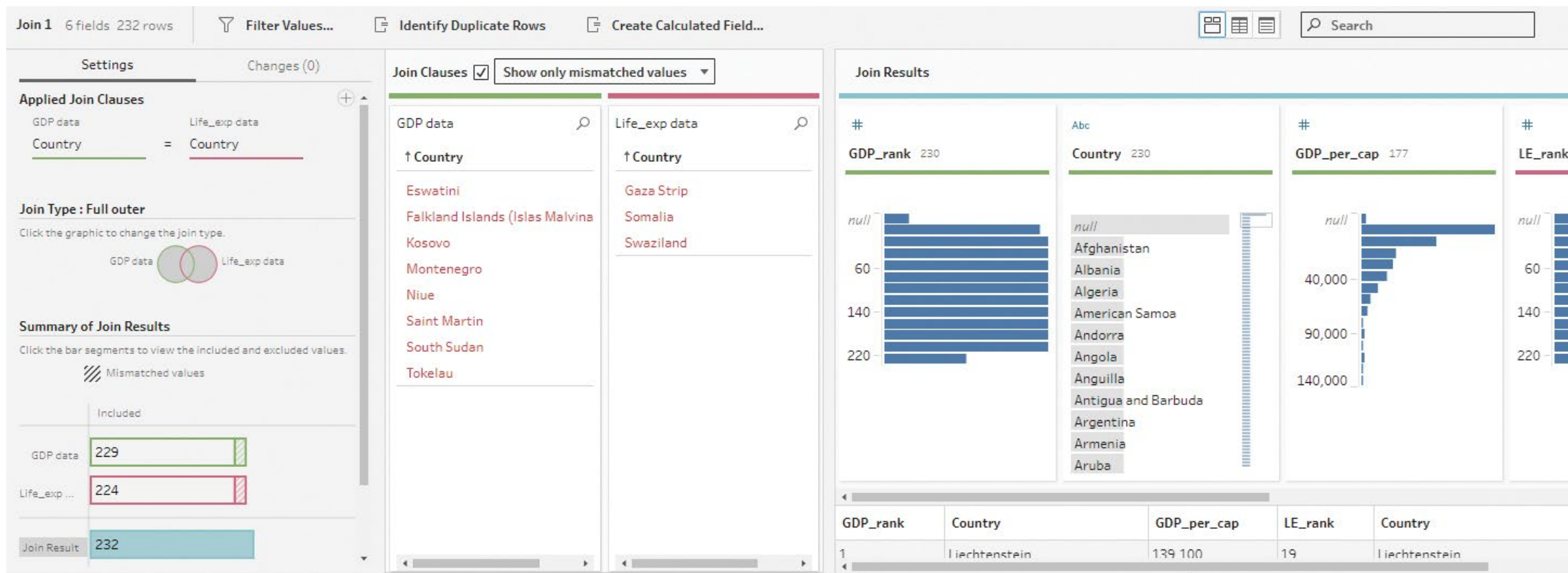


Figure 2.37 The result of a full join that keeps all the countries from both datasets.

		Aruba		
GDP_rank	Country	GDP_per_cap	LE_rank	Country
229	Burundi	700	202	Burundi
<i>null</i>	<i>null</i>	<i>null</i>	218	Somalia
<i>null</i>	<i>null</i>	<i>null</i>	124	Gaza Strip
<i>null</i>	<i>null</i>	<i>null</i>	220	Swaziland

Figure 2.38 Null values for the GDP data.

Add Field

Field Name

Country_join

IFNULL([Country],[Country-1])

Reference

All

< Search

ABS

ACOS

AND

IFNULL(expr1, expr2)

Returns <expr1> if it is not null,
otherwise returns <expr2>.

Example: IFNULL([Profit], 0)

Calculation is valid ^

Create with Tableau Agent

Apply

Apply and Close

Figure 2.39 Creating new calculated field named Country-Join using IFNULL([Country],[Country-1]).



Figure 2.40 Clicking the Changes tab displays the computation.

Tableau Prep Builder - Flow1*

File Edit Flow Server Help

Connections

- GDP.csv Text file
- Life_exp.csv Text file

Search

Tables

- Life_exp

Join 1 5 fields 232 rows

Filter Values... Identify Duplicate Rows Create Calculated Field...

Settings Changes (4)

- Join
 - [Country] == [Country]
- Calculated Field
 - Country_join
 - IFNULL([Country],[Country-1])
- Remove Field
 - Country
- Remove Field
 - Country-1
- Rename Field
 - Country
 - From [Country_join] to [Country]

Join Clauses ☒ Show only mismatched values

Join Results

Country 232 GDP_rank 230

Afghanistan null

Albania 60

Algeria 140

American Samoa 220

Andorra

Angola

Anguilla

Antigua and Barbuda

Argentina

Armenia

Aruba

Australia

Country	GDP_rank	GDP
Liechtenstein	1	139,1
...	...	...

Figure 2.41 The changes made to the data.

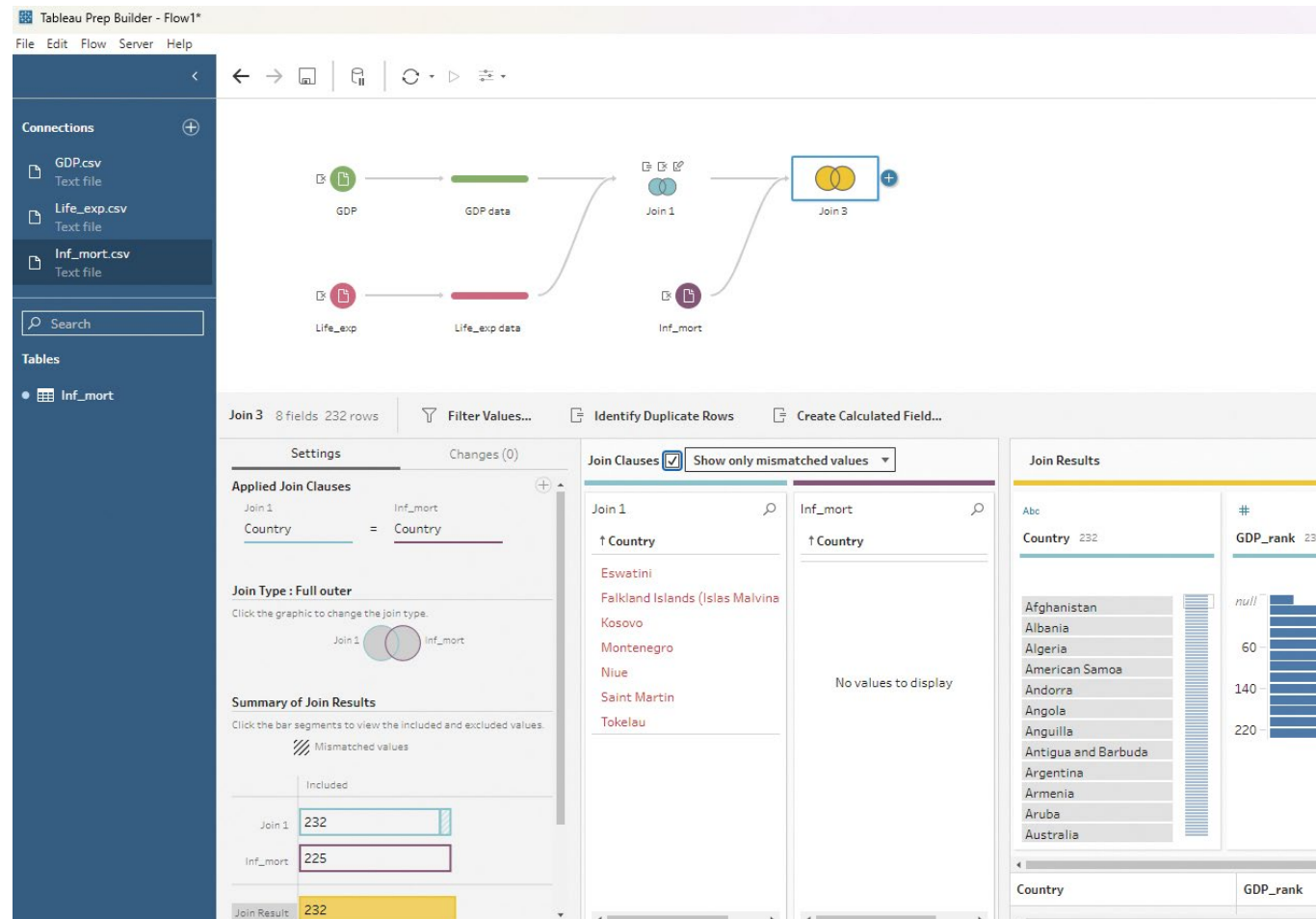


Figure 2.42 Repeat the process and add the Inf_mort data.

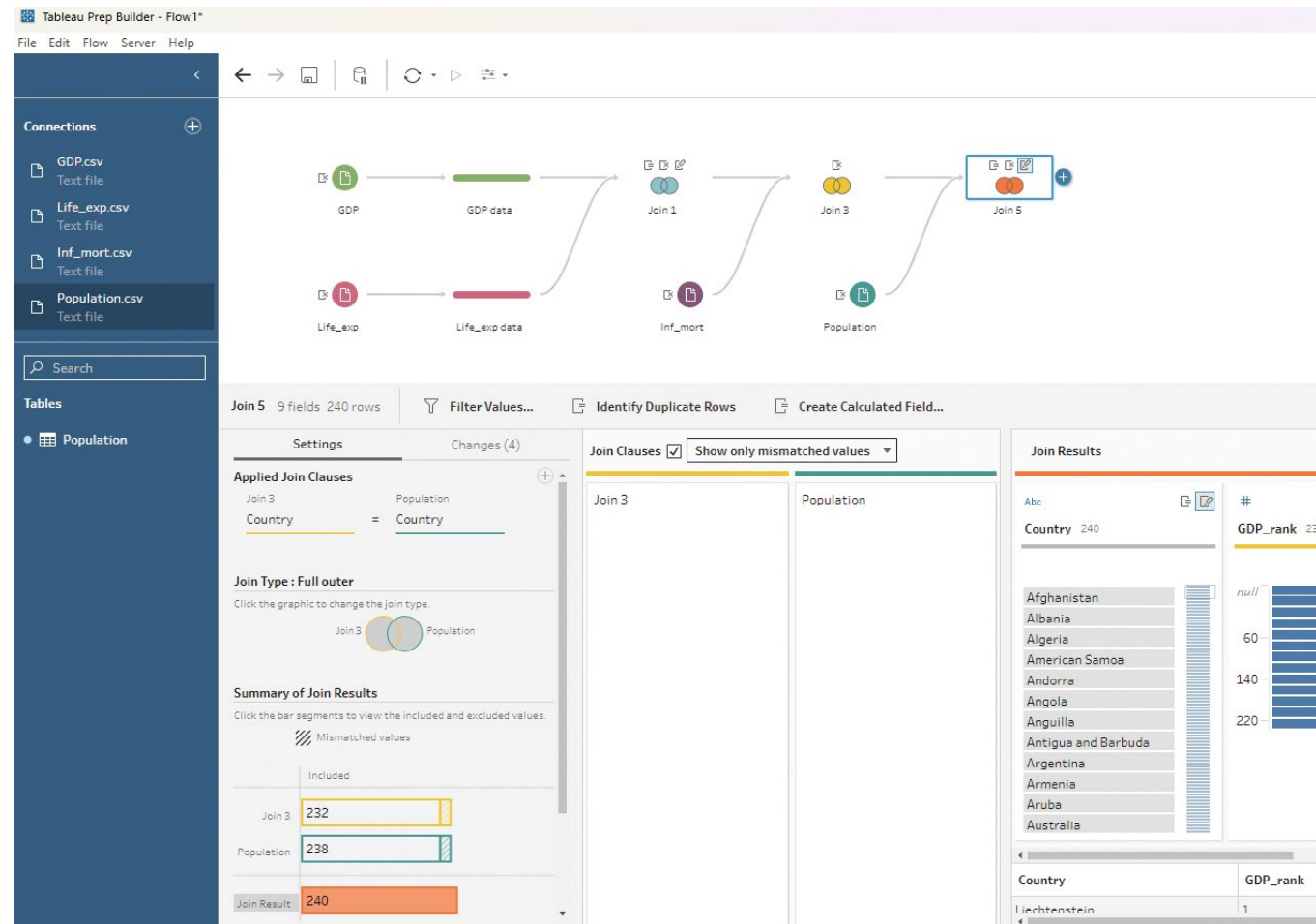


Figure 2.43 Adding the population data.

Tableau Prep Builder - Flow1*

File Edit Flow Server Help

Connections

- GDP.csv Text file
- Life_exp.csv Text file
- Inf_mort.csv Text file
- Population.csv Text file

Search

Tables

- Population

Output 9 fields

Save output to

File

Browse

Name

Countries_Joined

Location

C:\Users\egaze\OneDrive - Bowdoin College\Desktop\Excel_R_Tableau\images\ Countries

Output type

Comma Separated Values (.csv)

Write Options

Select an option to create or update your output table.

Full refresh

Create table

Run Flow

Flow Diagram:

```

graph LR
    GDP[GDP] --> GDP_data[GDP data]
    Life_exp[Life_exp] --> Life_exp_data[Life_exp data]
    GDP_data --> Join1[Join 1]
    Life_exp_data --> Join1
    Inf_mort[Inf_mort] --> Join3[Join 3]
    Population[Population] --> Join3
    Join1 --> Join3
    Join3 --> Countries_Joined[Countries_Join...]
    Countries_Joined --> Output[Output]
  
```

Save to Countries_Joined.csv

Country	GDP_rank	GDP_per_cap	IE_rank	Life_exp	IM_rank	Inf_mor
Liechtenstein	1	139,100	19	81.9	190	4.2
Qatar	2	124,500	52	78.9	167	6.2
Monaco	3	115,700	1	89.4	225	1.8
Macau	4	111,600	4	84.6	214	3.1
Luxembourg	5	106,300	13	82.3	209	3.4
Bermuda	6	99,400	26	81.4	219	2.5
Singapore	7	93,900	3	85.2	222	2.4
Isle of Man	8	84,600	29	81.3	193	4
Brunei	9	78,200	73	77.3	142	9.6
Ireland	10	75,500	33	80.9	198	3.6
Norway	11	71,800	18	81.9	221	2.5
Falkland Islands (Islas Malvinas)	12	70,800	null	null	null	null
United Arab Emirates	13	67,700	69	77.7	137	10
Sint Maarten	14	66,800	63	78.3	153	8.1
Kuwait	15	66,200	64	78.2	162	7
Gibraltar	16	61,700	46	79.6	169	5.9

Figure 2.44 The output step to generate the joined data file.

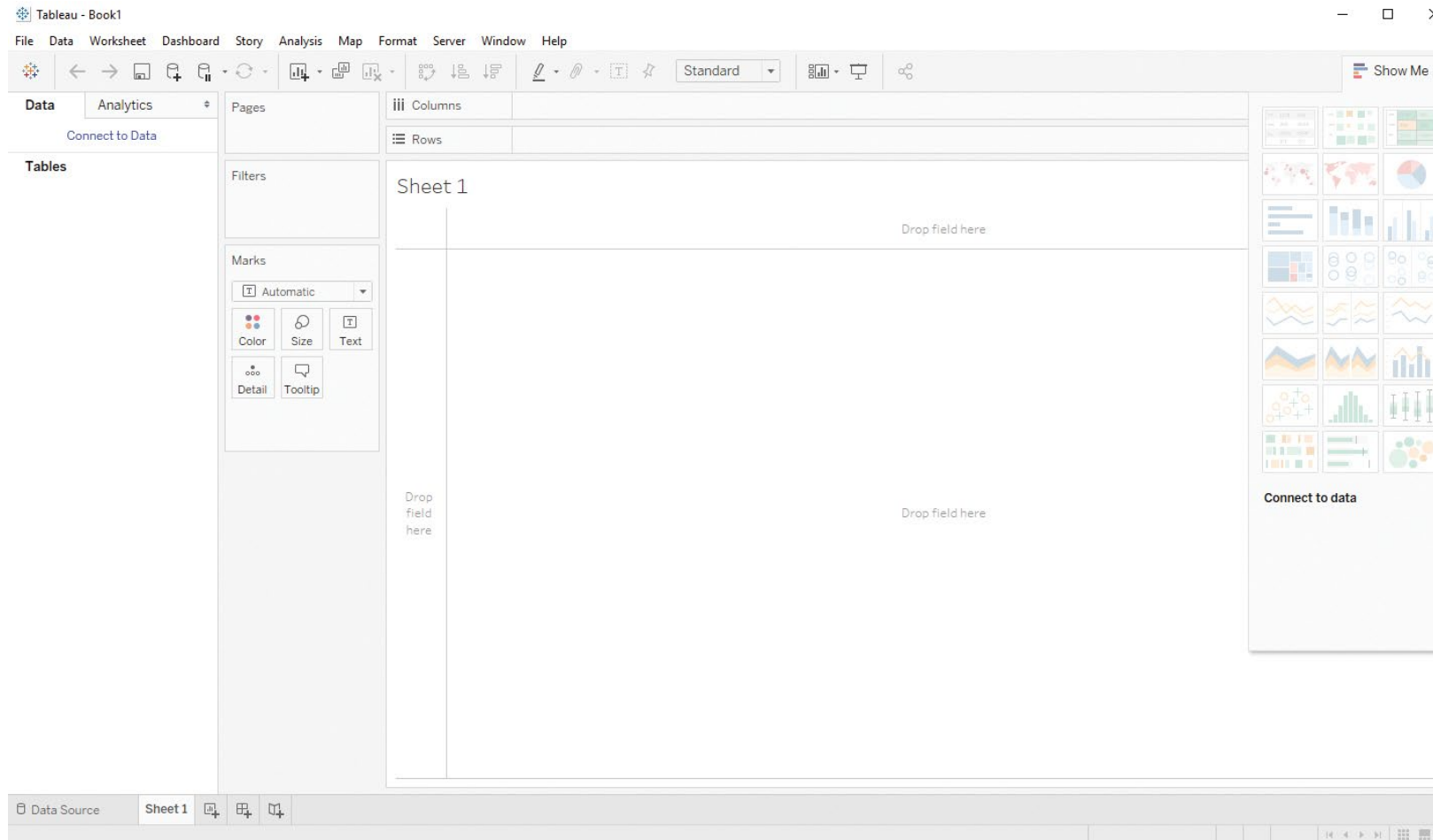


Figure 2.45 The Connect to data button.

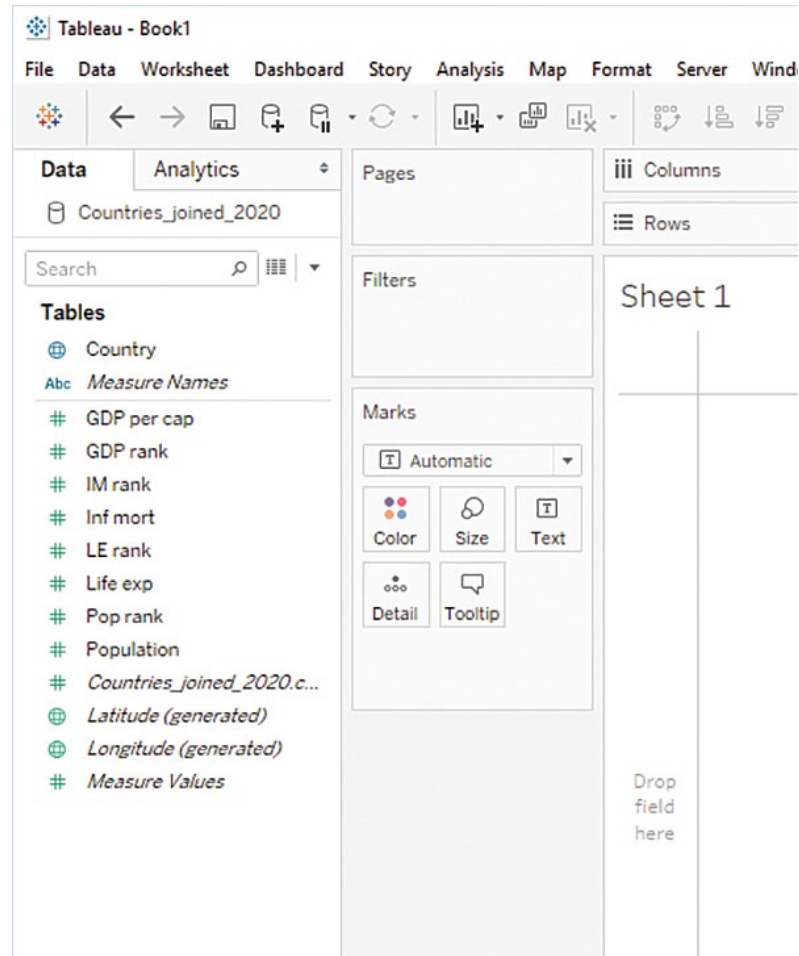


Figure 2.46 The Countries_joined.csv data file in Tableau.

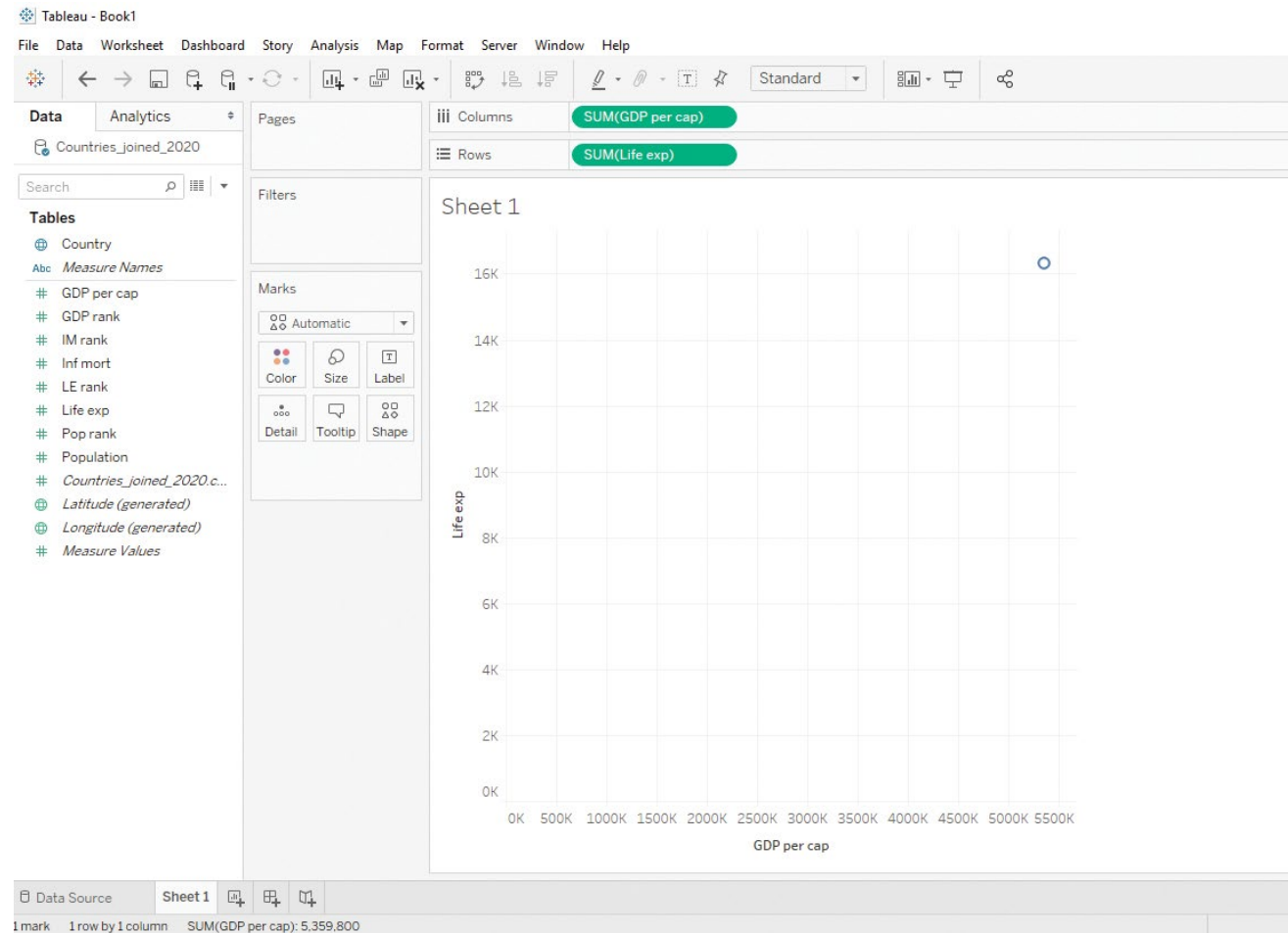


Figure 2.47 Tableau scatterplots default to summing each variable and plotting a single point.

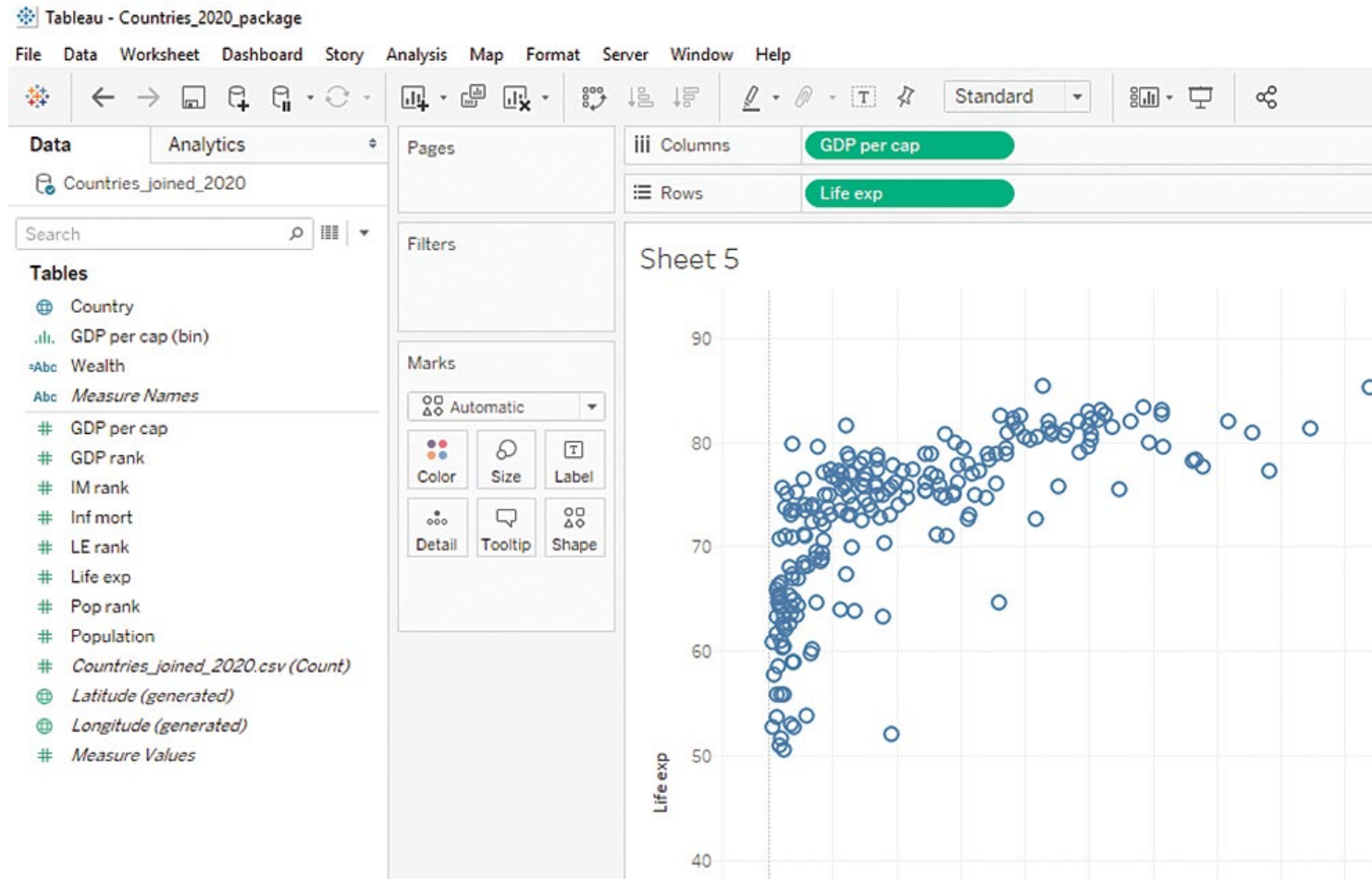


Figure 2.48 The disaggregated points.

Edit Colors Infant Mortality Rate

×

Palette

 Orange-Blue-White Diverging



1.8

163.8

Start

Custom

▼

1.8

Center

Custom

▼

30

End

Custom

▼

163.8

☐ Stepped Color

5

▲▼

steps

☒ Reverse

☒ Use full color range

☐ Include totals

↺ Reset

Figure 2.49 The Edit Colors dialog box.

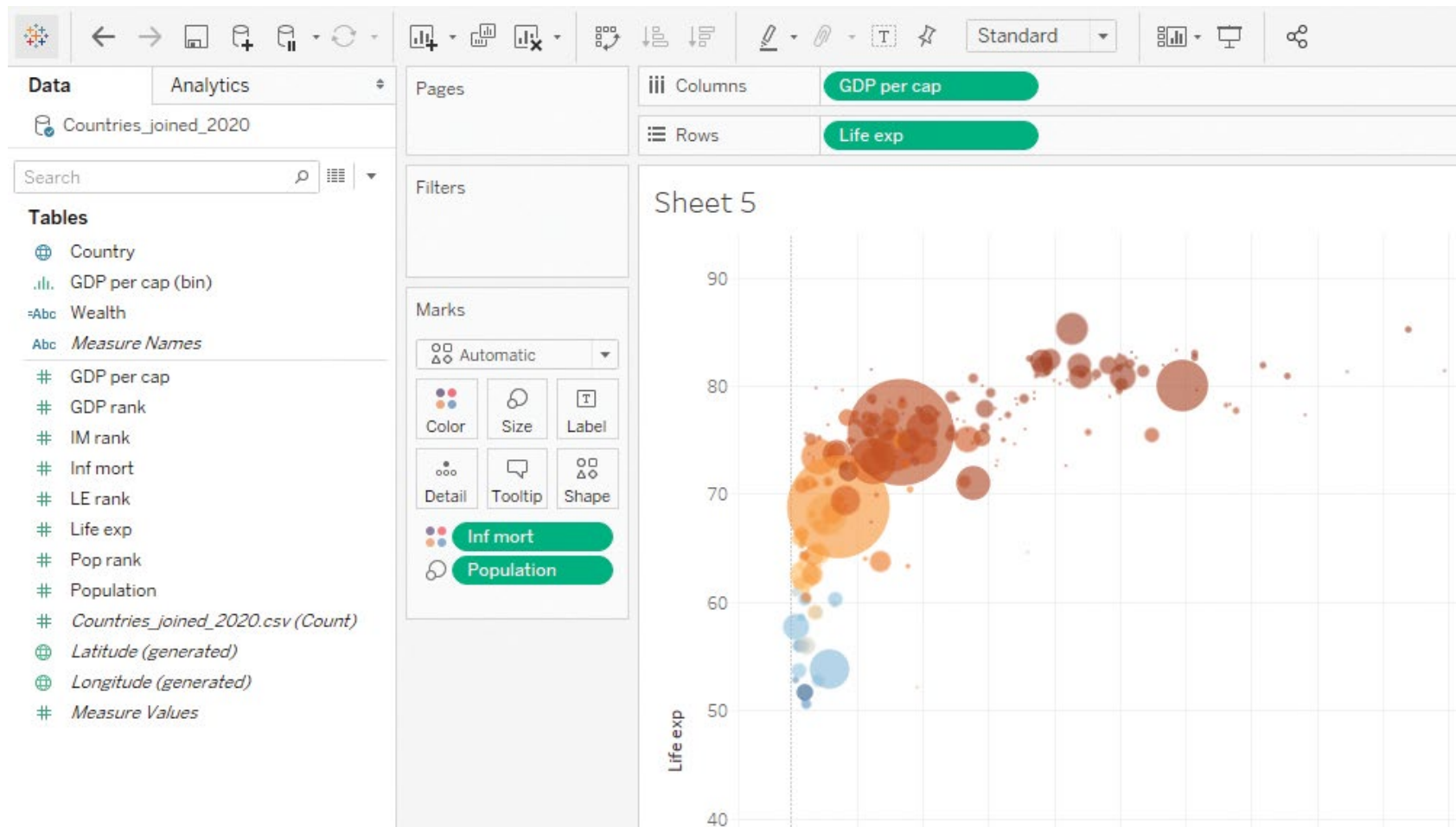


Figure 2.50 Sizing the dots by population.

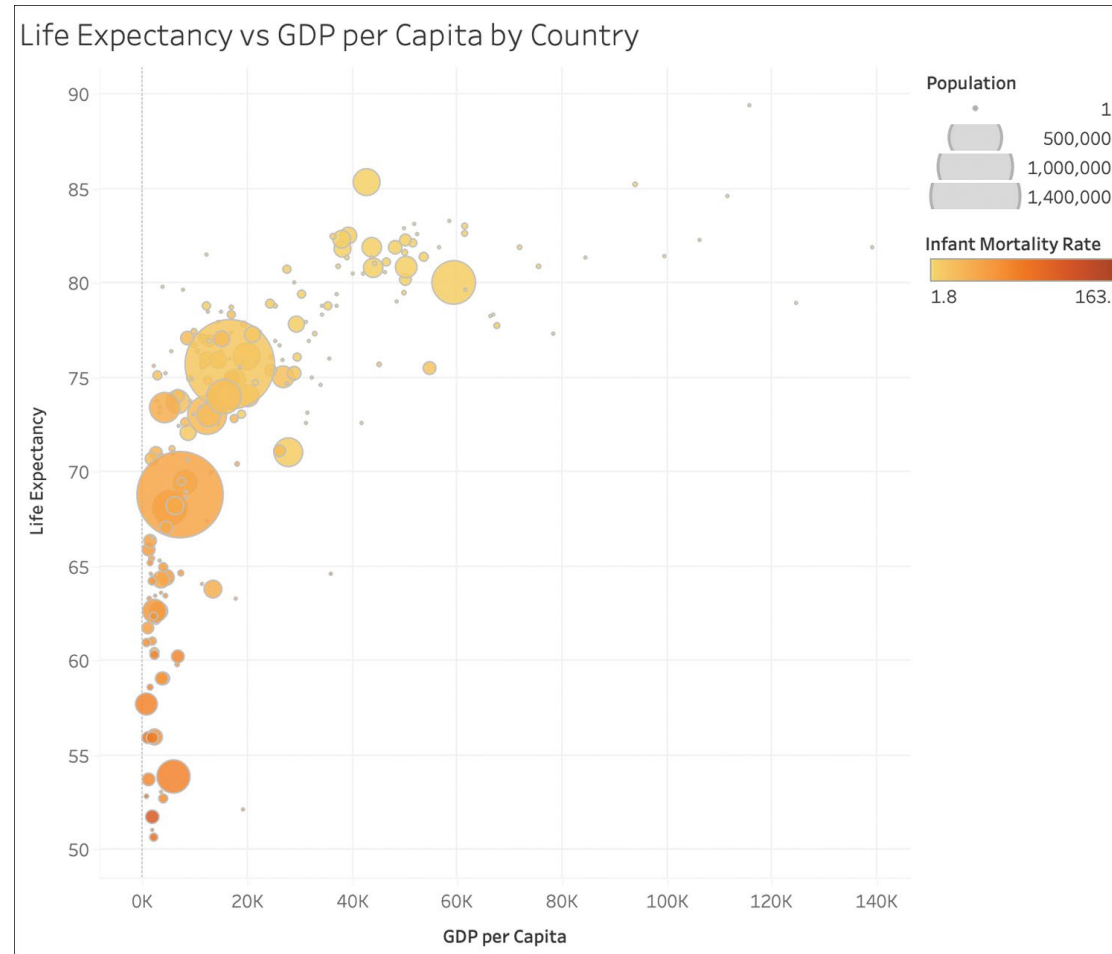


Figure 2.51 The final scatterplot of life expectancy versus GDP per capita.

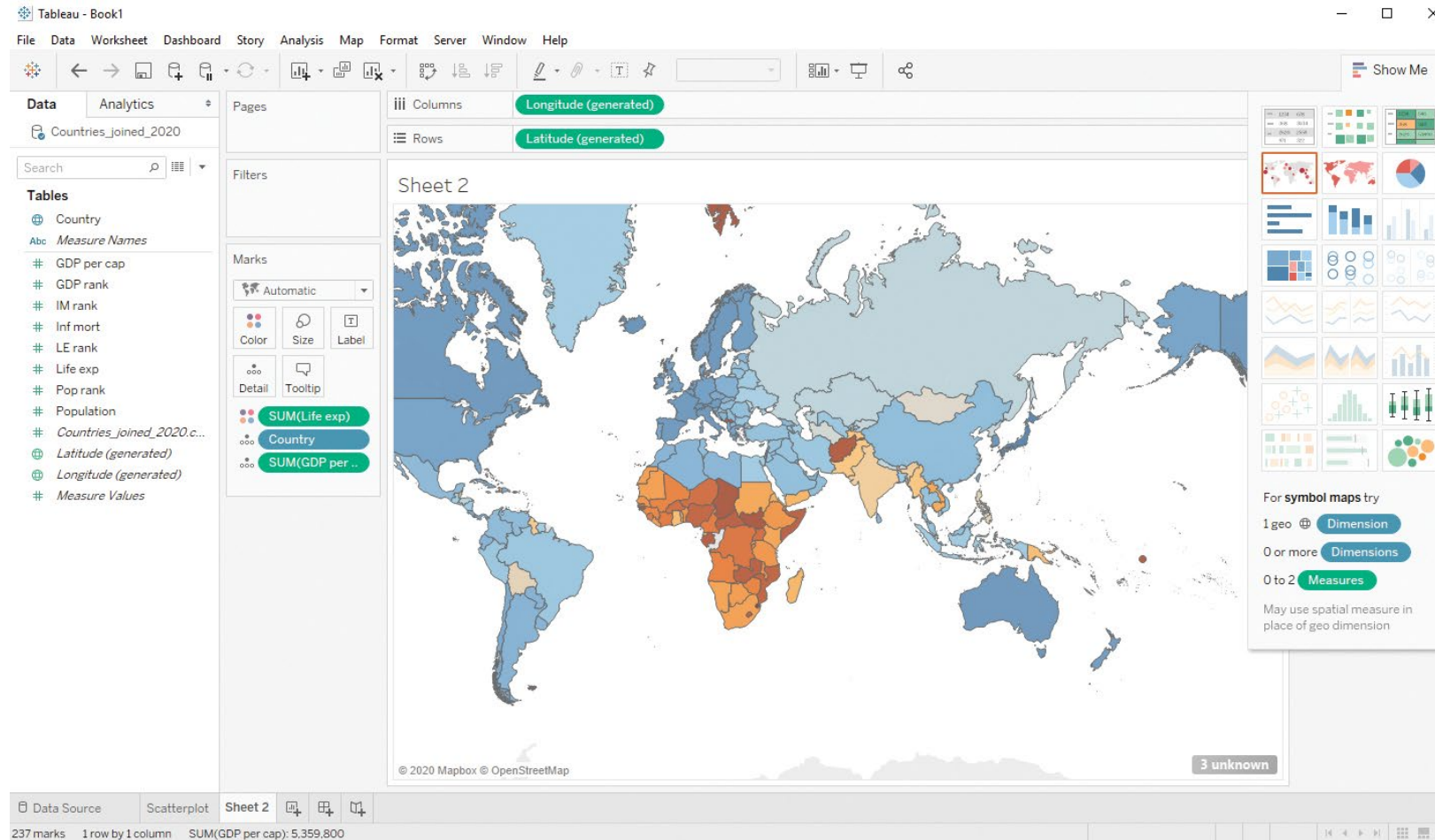


Figure 2.52 Making a map by dragging the Country field to the Detail mark.

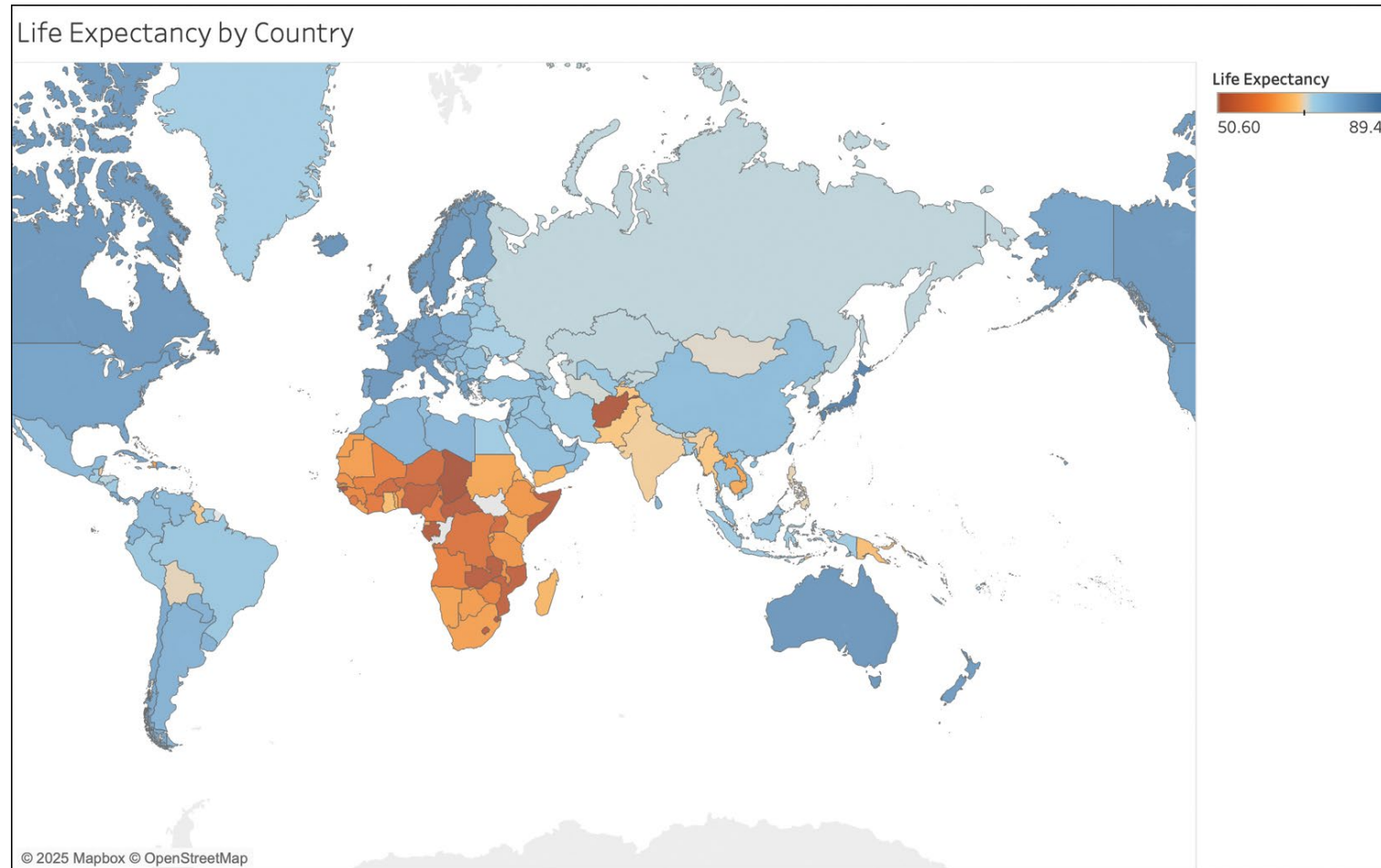


Figure 2.53 The choropleth or heat map showing life expectancy by country.

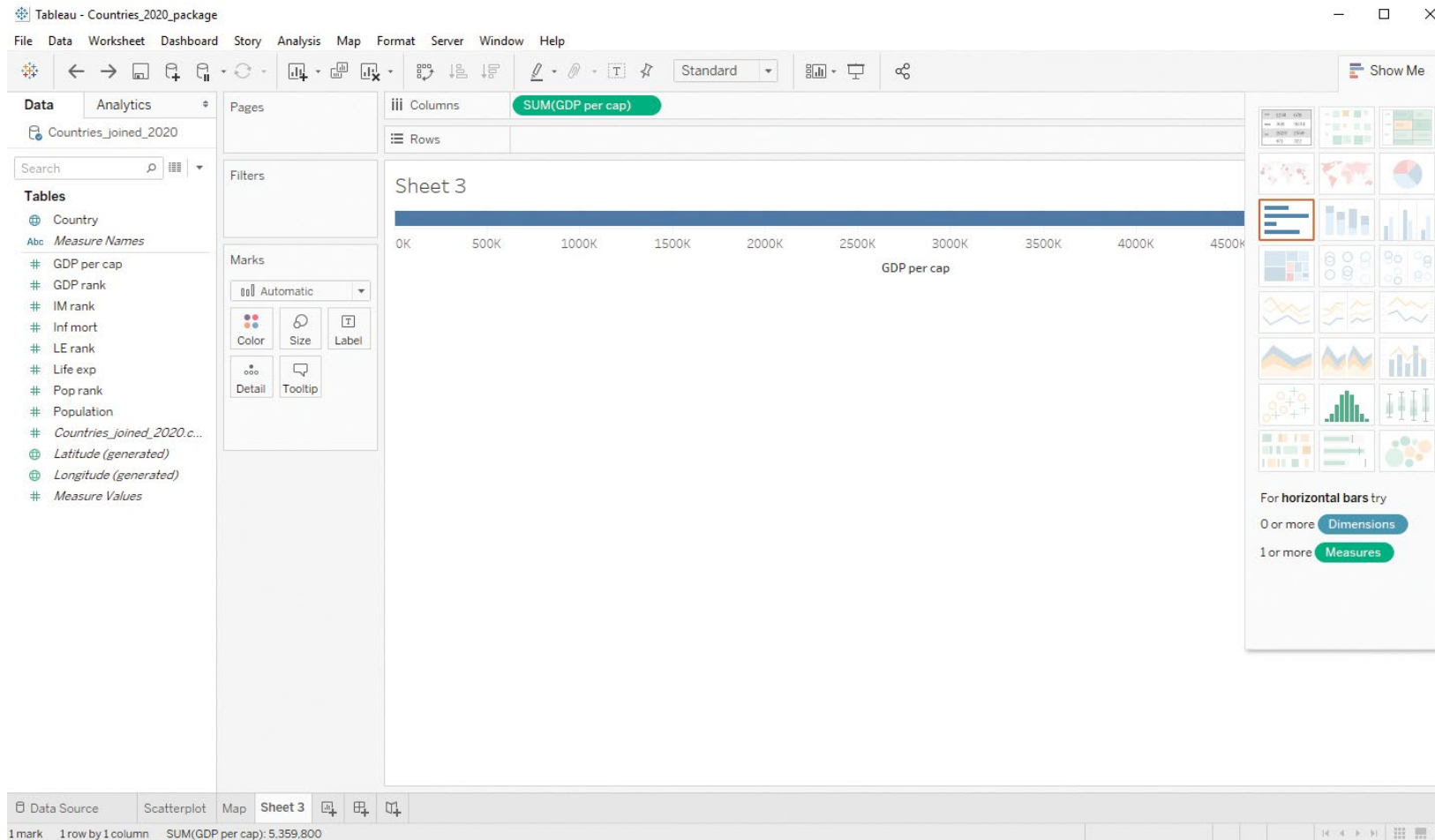


Figure 2.54 Clicking the histogram icon in the Show me field.

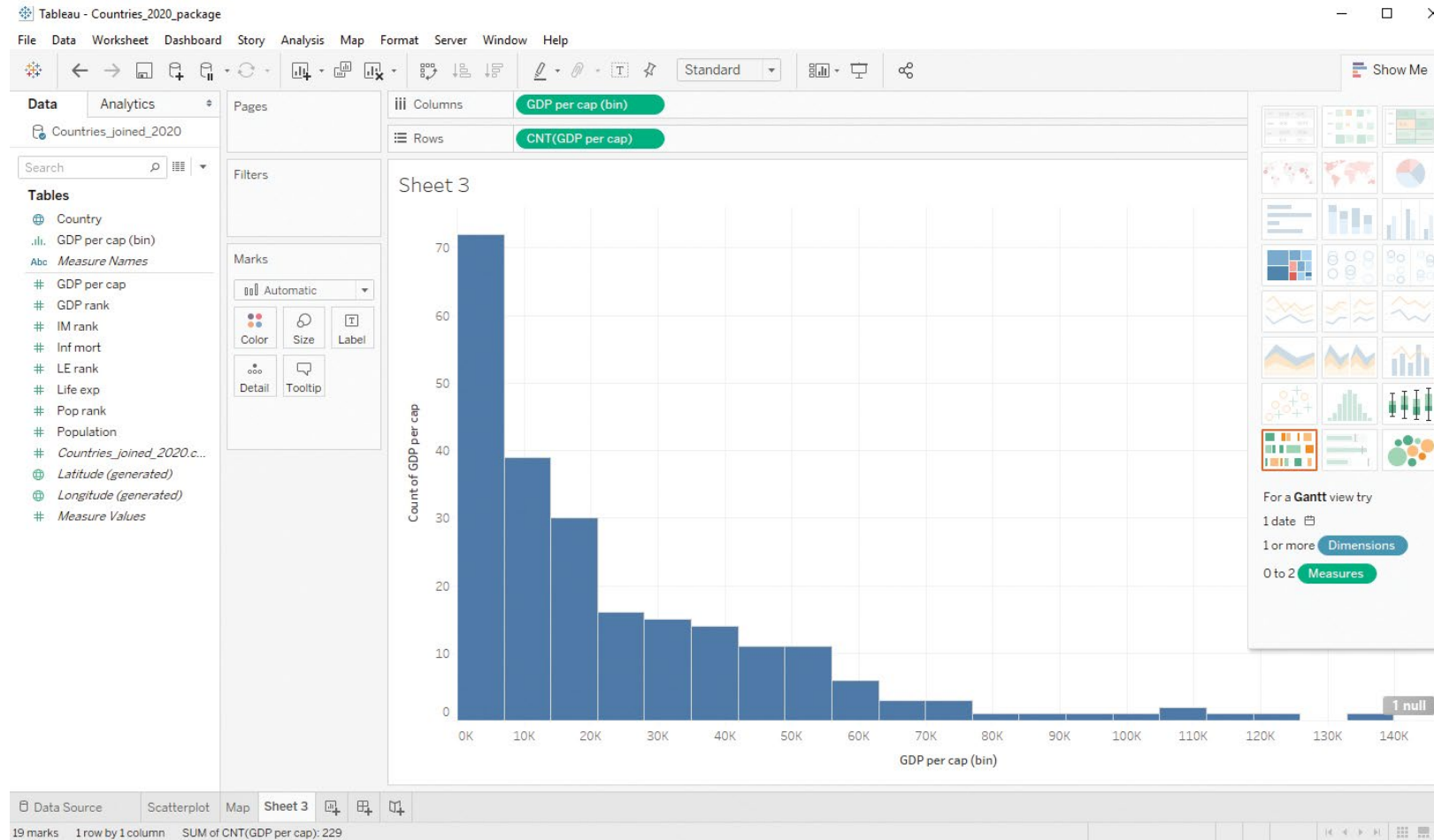


Figure 2.55 Automatic histogram and bin creation.

J2

✕ ✓ *fx*

=IF(C2 > 60000, "Rich", "Poor")

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Country	GDP_rank	GDP_per	LE_rank	Life_exp	IM_rank	Inf_mort	Pop_rank	Populatio	Wealth				
2	Liechtenst	1	139100	19	81.9	190	4.2	213	38244	=IF(C2 > 60000, "Rich", "Poor")				
3	Qatar	2	124500	52	78.9	167	6.2	143	2314307	IF(logical_test, [value_if_true], [value_if_false])				
4	Monaco	3	115700	1	89.4	225	1.8	217	30645					

Figure 2.56 Entering the IF statement in Excel.

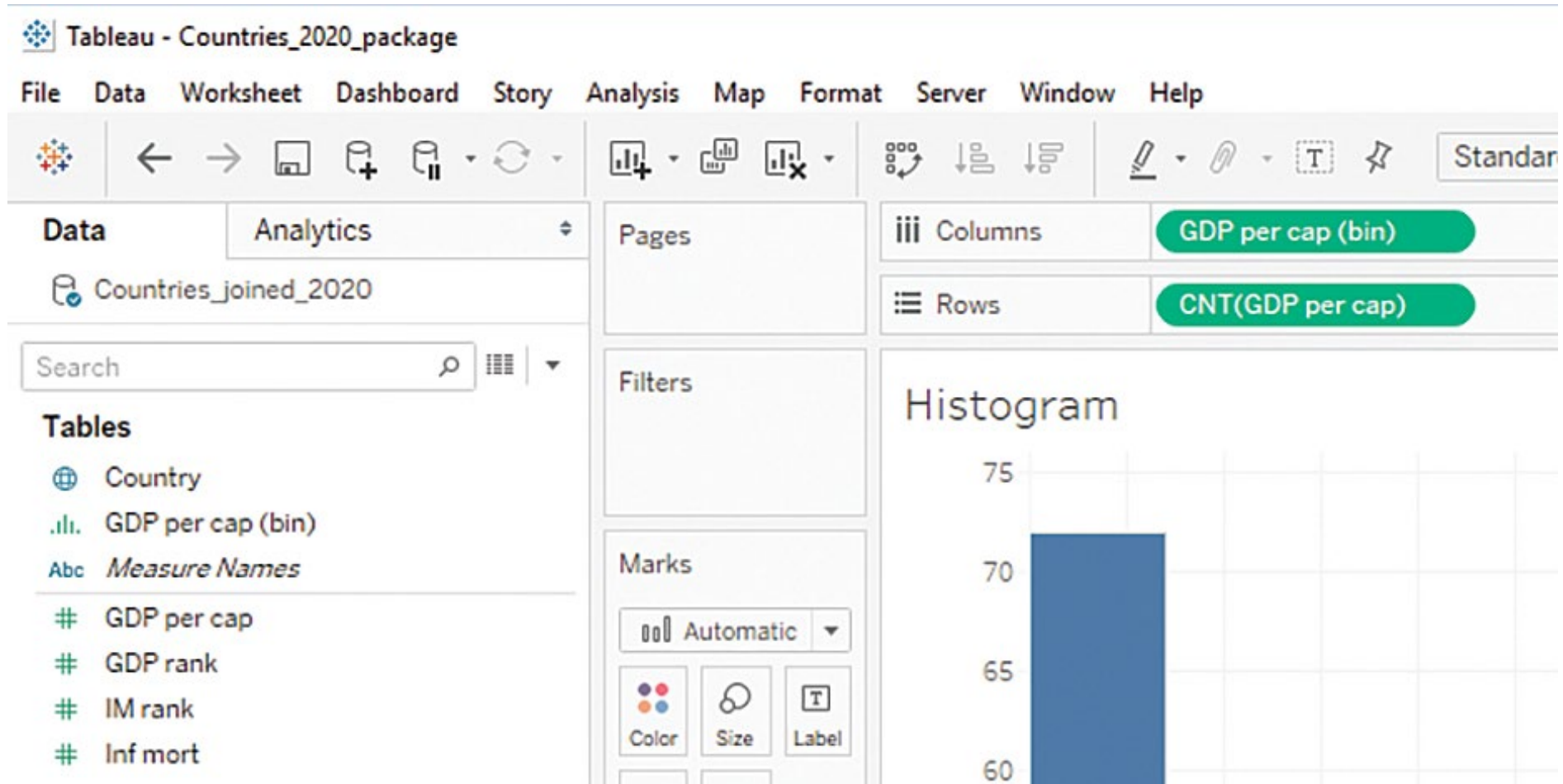


Figure 2.57 The drop-down arrow next to the Search box in the data pane.

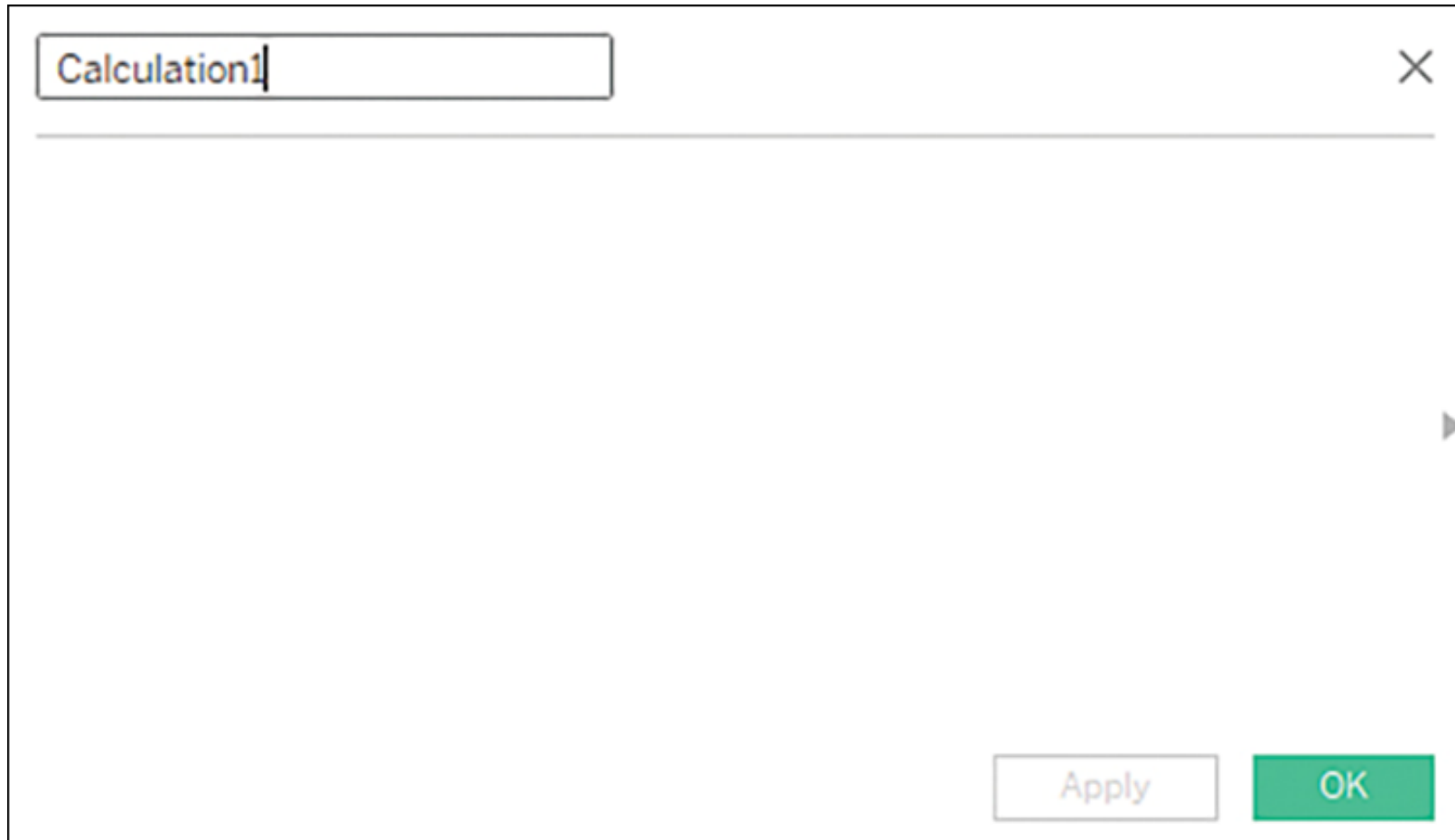


Figure 2.58 The code editor for creating a calculated field.

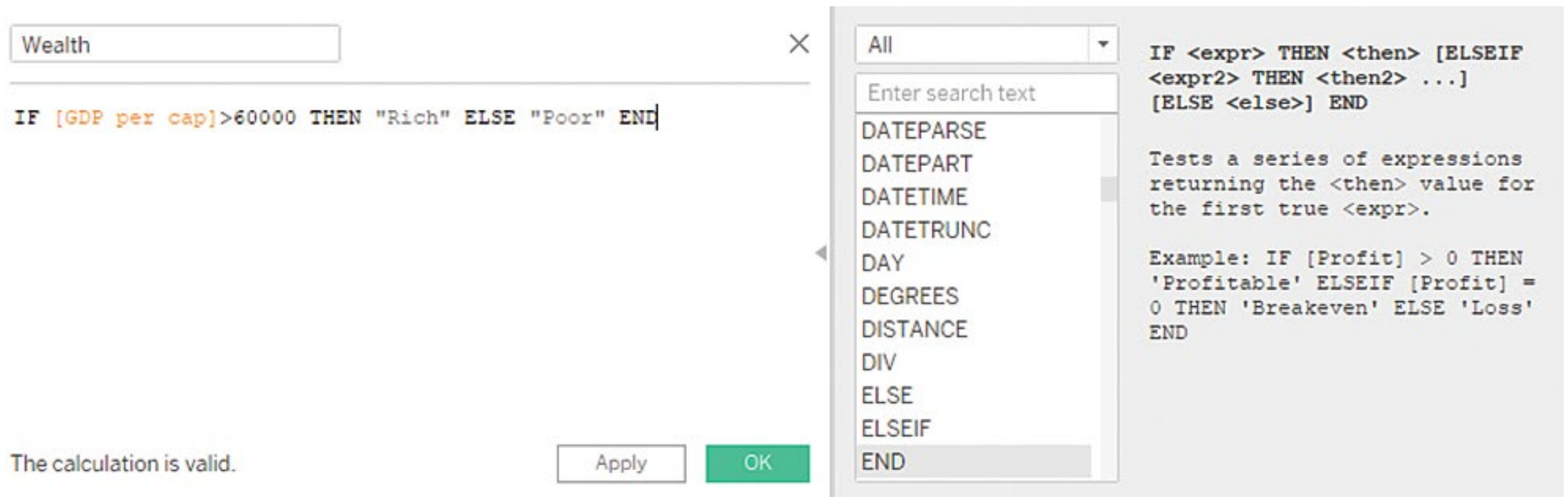


Figure 2.59 The Wealth field calculated with the IF function in Tableau.

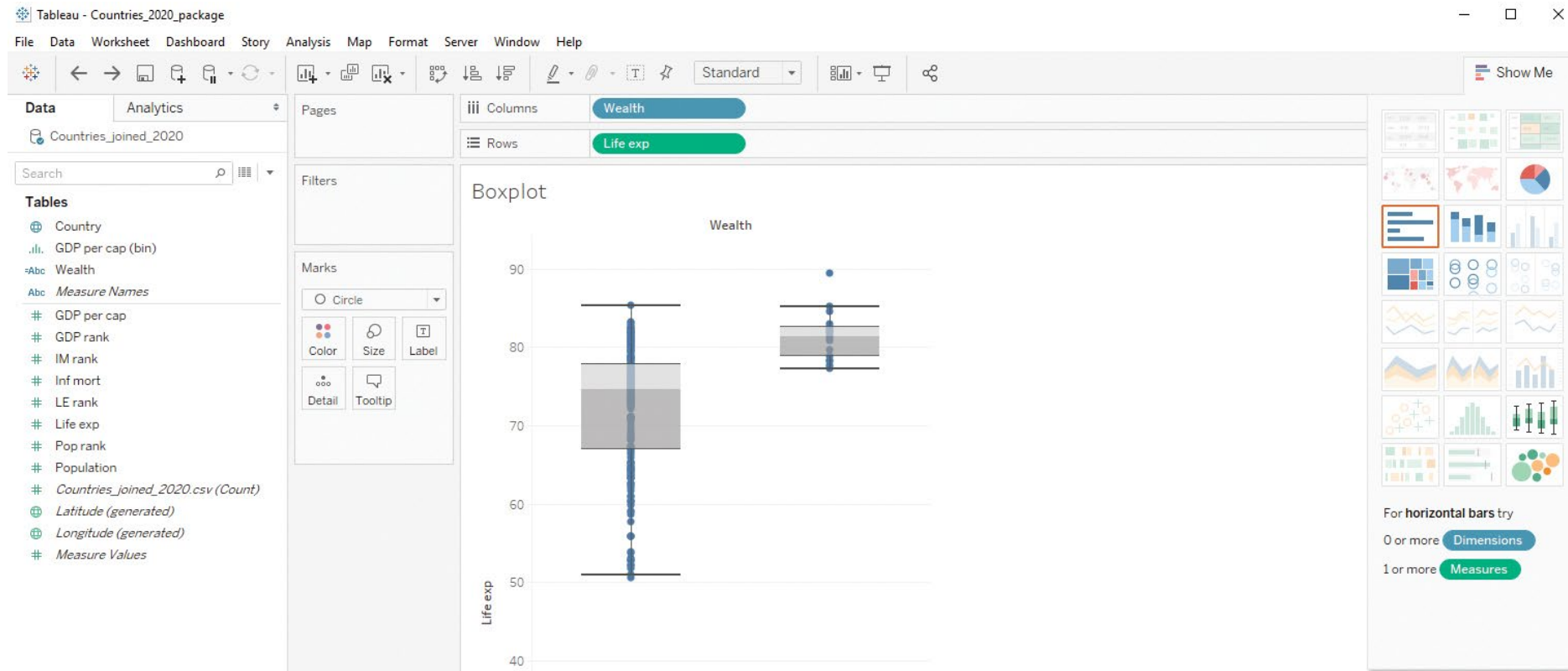


Figure 2.60 The boxplot of life expectancy by wealth with the data disaggregated.

Box Plots

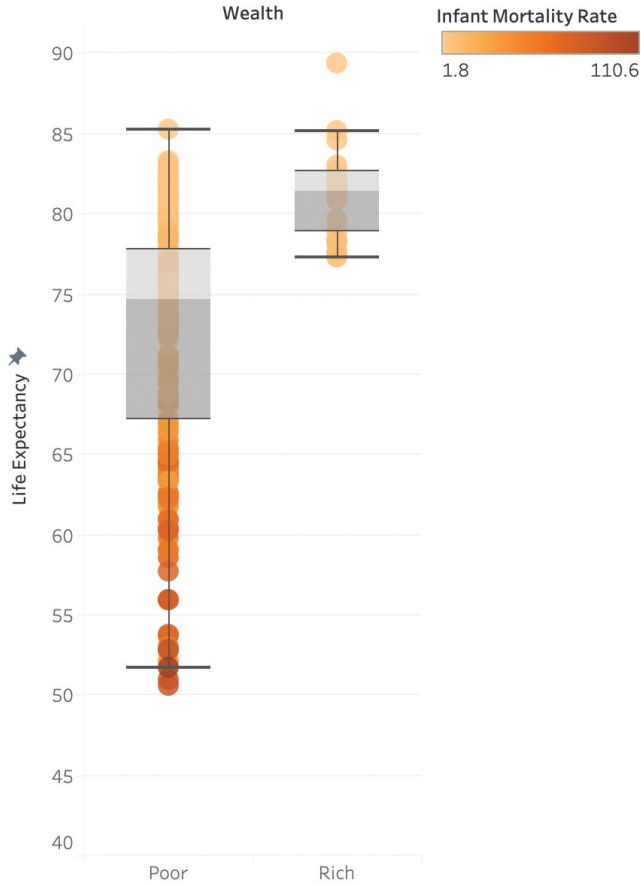


Figure 2.61 The finished boxplot.

Connections

GDP

Text file

Add

Files

☐ Use Data Interpreter

Data Interpreter might be able to clean your Text file workbook.

GDP.csv

Inf_mort.csv

Life_exp.csv

Population.csv

New Union

New Table Extension

GDP.csv

GDP.csv

3 fields 229 rows

Name

GDP.csv

Fields

Type	Field Name	Physical Table	Rem...
#	GDP rank	GDP.csv	GDP_r...
🌐	Country	GDP.csv	Country
#	GDP per cap	GDP.csv	GDP_p...

#	GDP.csv	🌐	GDP.csv	#	GDP.csv
	GDP rank		Country		GDP per cap
	1		Liechtenstein		139,100
	2		Qatar		124,500
	3		Monaco		115,700
	4		Macau		111,600
	5		Luxembourg		106,300
	6		Bermuda		99,400
	7		Singapore		93,900

Go to Worksheet

Data Source

Sheet 1

Figure 2.62 The files in the Files list.

ConnectionsAdd

GDP
Text file

Files

☐ Use Data Interpreter
Data Interpreter might be able to clean your Text file workbook.

GDP.csv

Inf_mort.csv

Life_exp.csv

Population.csv

New Union

New Table Extension

GDP.csv — Inf_mort.csv

How do relationships differ from joins? Learn more

GDP.csv

Operator

Inf_mort.csv

Abc Country

=

Abc Country (Inf mor

+ Add more fields

> Performance Options

Remove Relationship

Inf_mort.csv
IM rank

Inf_mort.csv

Country (Inf mort.csv)

Inf_mort.csv
Inf mort

1

Afghanistan

110.6000

2

Somalia

94.8000

3

Central African Republic

86.3000

4

Guinea-Bissau

85.7000

5

Chad

85.4000

6

Niger

81.1000

7

Burkina Faso

72.2000

Data Source

Sheet 1

Figure 2.63 Inf_mort.csv to the right of GDP.csv.

From *Data Analytics Toolkit* by Eric Gaze and McKenzie Lamb (ISBN 978-0-13-539786-2). Copyright © 2027 Pearson Education, Inc. All rights reserved.

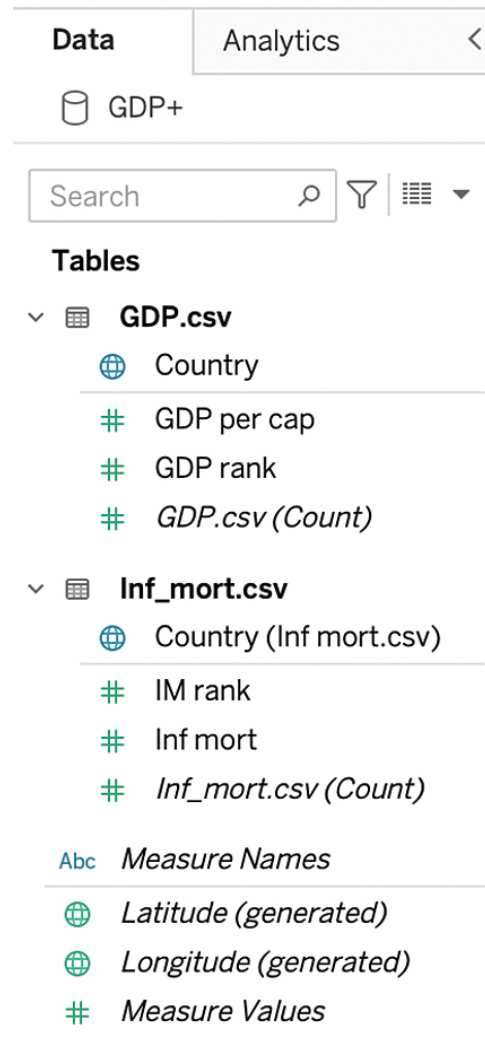


Figure 2.64 Field for both datasets.

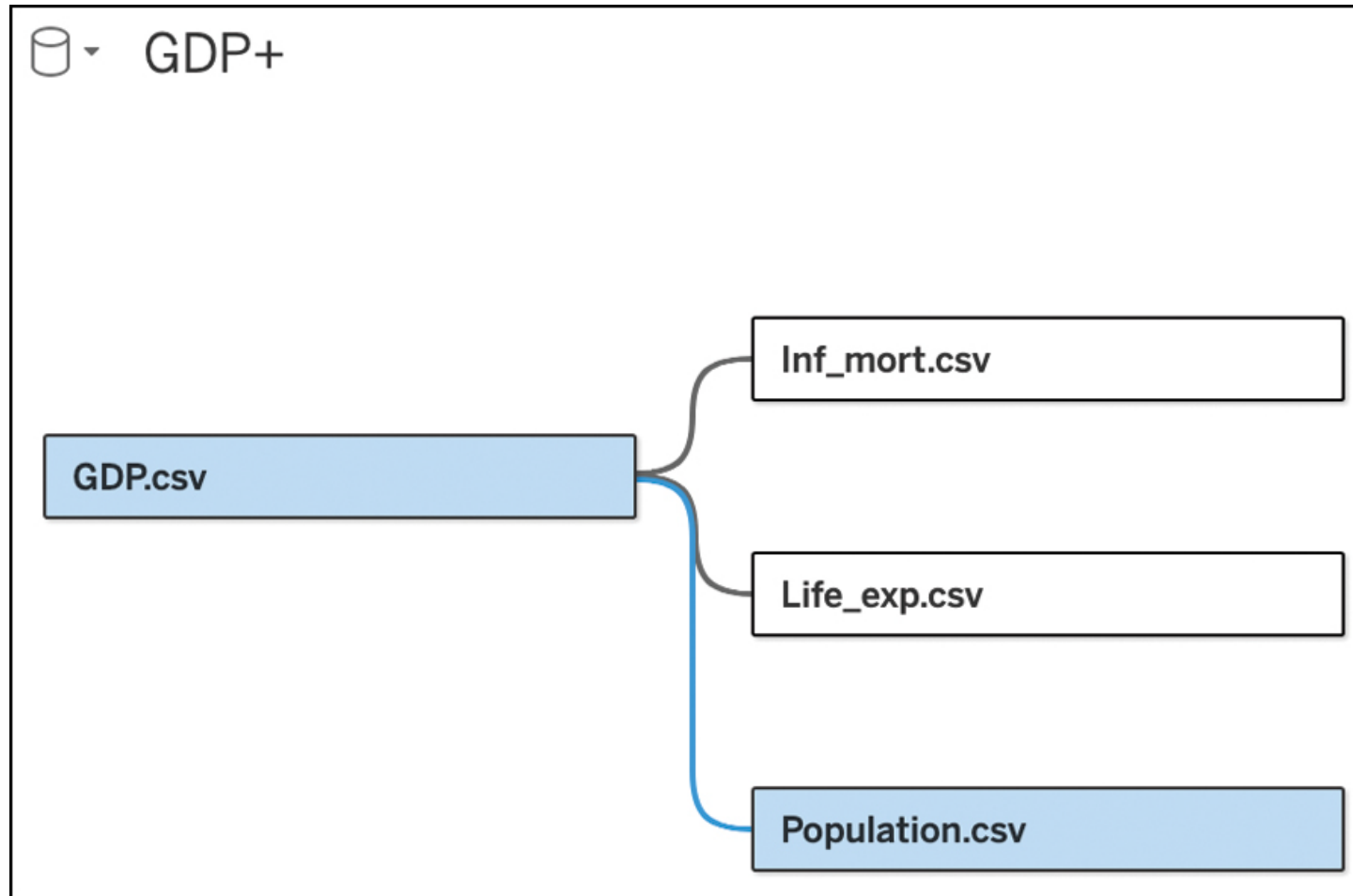


Figure 2.65 All of the other datasets connect to GDP.csv.

The screenshot shows a data tool interface. At the top, a diagram illustrates a relationship between **GDP.csv** and three other datasets: **Inf_mort.csv**, **Life_exp.csv**, and **Population.csv**. Below this, a dropdown menu shows the selected relationship: **GDP.csv — Life_exp.csv**.

The main workspace is divided into two sections. The left section, titled "How do relationships differ from joins? [Learn more](#)", contains a query builder. It shows the relationship between **GDP.csv** and **Life_exp.csv** with the operator **=**. The fields **Abc Country** and **Abc Country (Life exp)** are selected. Below the query builder is a button labeled **+ Add more fields**.

The right section displays a table of results. The table has three columns: **LE rank**, **Country (Life exp.csv)**, and **Life exp**. The data is as follows:

LE rank	Country (Life exp.csv)	Life exp
1	Monaco	89.4000
2	Japan	85.3000
3	Singapore	85.2000
4	Macau	84.6000
5	San Marino	83.3000
6	Iceland	83.1000
7	Hong Kong	83.0000

At the bottom left of the workspace, there is a red button labeled **Remove Relationship**.

Figure 2.66 The fields that match Life exp.

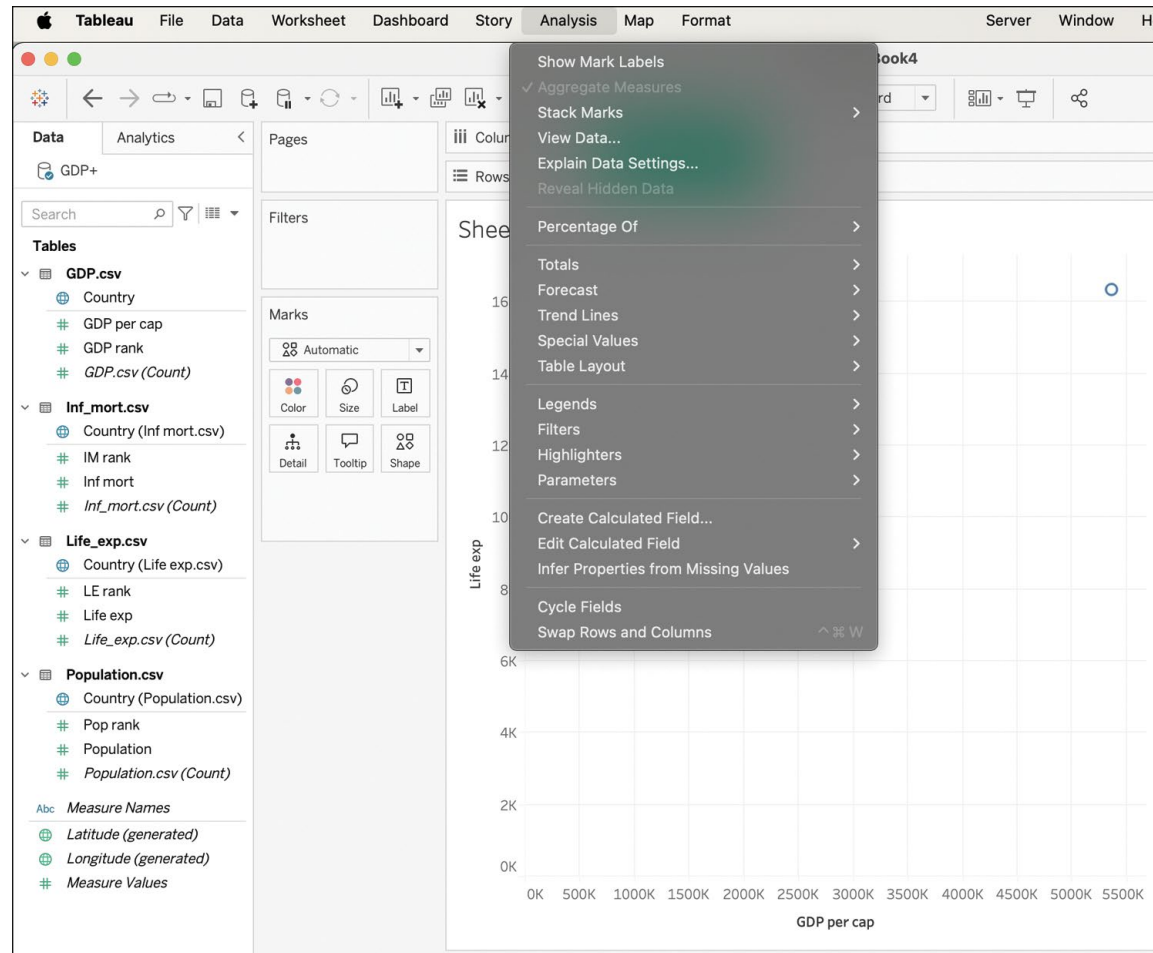


Figure 2.67 The Aggregate Measures option in the Analysis menu is checked and cannot be modified.

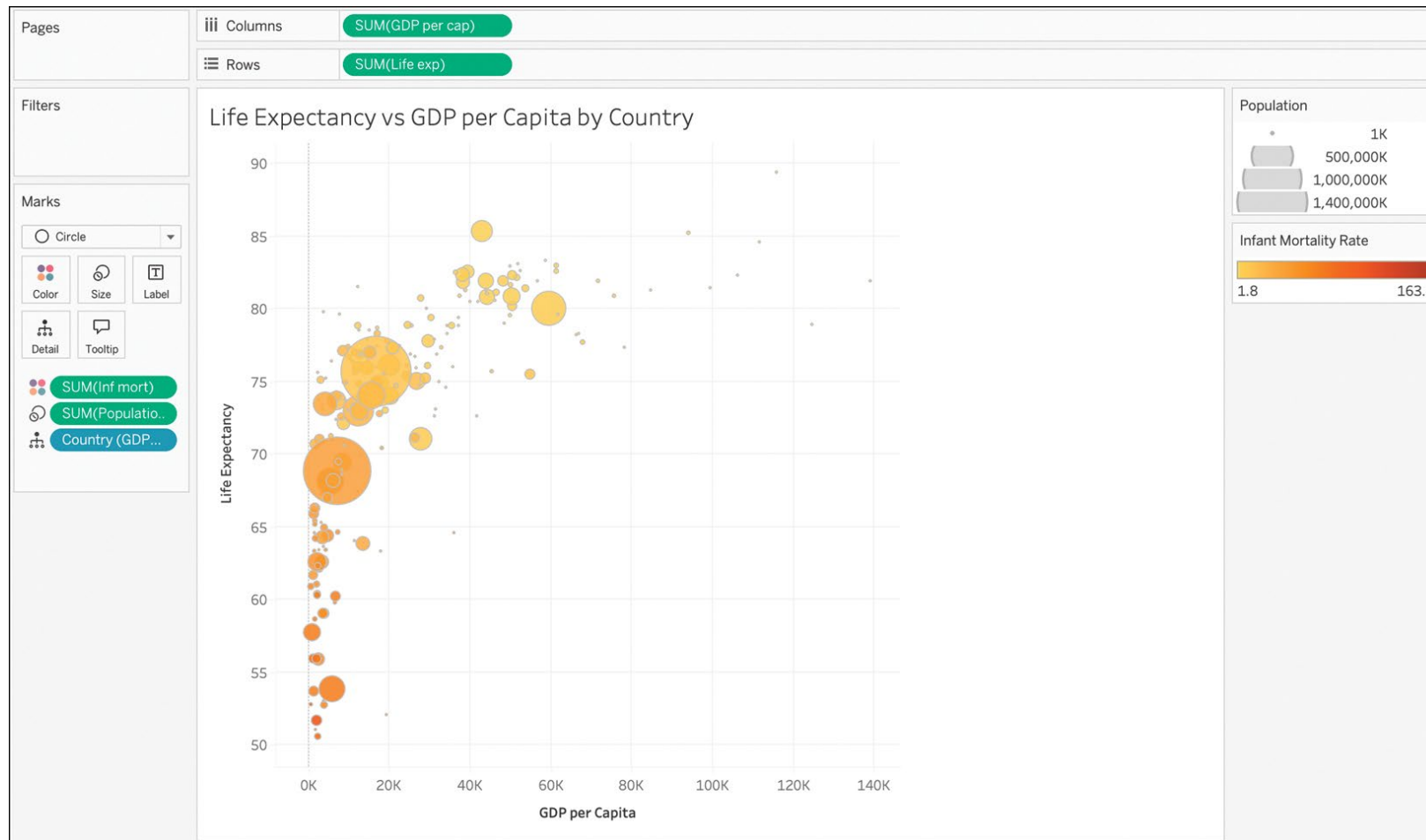


Figure 2.68 The scatterplot.

Columns	
Rows	Country (Life exp.csv) Country (Population.c..) Country (Inf mort.csv) Country (GDP.csv)

Sheet 2

Country (Life exp.csv)	Country (Population.csv)	Country (Inf mort.csv)	Country (GDP.csv)	
Afghanistan	Afghanistan	Afghanistan	Afghanistan	Abc
Albania	Albania	Albania	Albania	Abc
Algeria	Algeria	Algeria	Algeria	Abc
American Samoa	American Samoa	American Samoa	American Samoa	Abc
Andorra	Andorra	Andorra	Andorra	Abc
Angola	Angola	Angola	Angola	Abc
Anguilla	Anguilla	Anguilla	Anguilla	Abc
Antigua and Barbuda	Antigua and Barbuda	Antigua and Barbuda	Antigua and Barbuda	Abc
Argentina	Argentina	Argentina	Argentina	Abc
Armenia	Armenia	Armenia	Armenia	Abc
Aruba	Aruba	Aruba	Aruba	Abc
Australia	Australia	Australia	Australia	Abc
Austria	Austria	Austria	Austria	Abc
Azerbaijan	Azerbaijan	Azerbaijan	Azerbaijan	Abc
Bahamas, The	Bahamas, The	Bahamas, The	Bahamas, The	Abc
Bahrain	Bahrain	Bahrain	Bahrain	Abc
Bangladesh	Bangladesh	Bangladesh	Bangladesh	Abc
Barbados	Barbados	Barbados	Barbados	Abc
Belarus	Belarus	Belarus	Belarus	Abc
Belgium	Belgium	Belgium	Belgium	Abc
Belize	Belize	Belize	Belize	Abc
Benin	Benin	Benin	Benin	Abc
Bermuda	Bermuda	Bermuda	Bermuda	Abc
Bhutan	Bhutan	Bhutan	Bhutan	Abc
Bolivia	Bolivia	Bolivia	Bolivia	Abc
Bosnia and Herzegovina	Bosnia and Herzegovina	Bosnia and Herzegovina	Bosnia and Herzegovina	Abc
Botswana	Botswana	Botswana	Botswana	Abc
Brazil	Brazil	Brazil	Brazil	Abc
British Virgin Islands	British Virgin Islands	British Virgin Islands	British Virgin Islands	Abc

Figure 2.69 The Country fields appear as rows.

Columns				
Rows	Country (GDP.csv)	Country (Inf mort.csv)	Country (Life exp.csv)	Country (Population.c..)

Sheet 2

Country (GDP.csv)	Country (Inf mort.csv)	Country (Life exp.csv)	Country (Population.csv)	
Null	Gaza Strip	Null	Null	16.6
	Somalia	Null	Null	94.8
	Swaziland	Null	Null	48.4
Afghanistan	Afghanistan	Afghanistan	Afghanistan	110.6
Albania	Albania	Albania	Albania	11.9
Algeria	Algeria	Algeria	Algeria	19.6
American Samoa	American Samoa	American Samoa	American Samoa	11.3
Andorra	Andorra	Andorra	Andorra	3.6
Angola	Angola	Angola	Angola	67.6
Anguilla	Anguilla	Anguilla	Anguilla	3.3
Antigua and Barbuda	Antigua and Barbuda	Antigua and Barbuda	Antigua and Barbuda	12.1
Argentina	Argentina	Argentina	Argentina	9.8
Armenia	Armenia	Armenia	Armenia	12.7
Aruba	Aruba	Aruba	Aruba	10.7
Australia	Australia	Australia	Australia	4.3
Austria	Austria	Austria	Austria	3.4
Azerbaijan	Azerbaijan	Azerbaijan	Azerbaijan	23.8
Bahamas, The	Bahamas, The	Bahamas, The	Bahamas, The	11.3
Bahrain	Bahrain	Bahrain	Bahrain	8.9
Bangladesh	Bangladesh	Bangladesh	Bangladesh	31.7
Barbados	Barbados	Barbados	Barbados	10.2
Belarus	Belarus	Belarus	Belarus	3.6
Belgium	Belgium	Belgium	Belgium	3.4
Belize	Belize	Belize	Belize	18.9
Benin	Benin	Benin	Benin	52.8
Bermuda	Bermuda	Bermuda	Bermuda	2.5
Bhutan	Bhutan	Bhutan	Bhutan	32.1
Bolivia	Bolivia	Bolivia	Bolivia	35.3
Bosnia and Herzegovina	Bosnia and Herzegovina	Bosnia and Herzegovina	Bosnia and Herzegovina	5.5

Figure 2.70 Null values appear when adding the Inf mort field to the table.

NULLS

×

ISNULL([Country (GDP.csv)]) OR
ISNULL([Country (Inf mort.csv)]) OR
ISNULL([Country (Life exp.csv)]) OR
ISNULL([Country (Population.csv)])

The calculation is valid.

2 Dependencies ▼

Apply

OK

Figure 2.71 The formula for calculating a new field named NULLS.

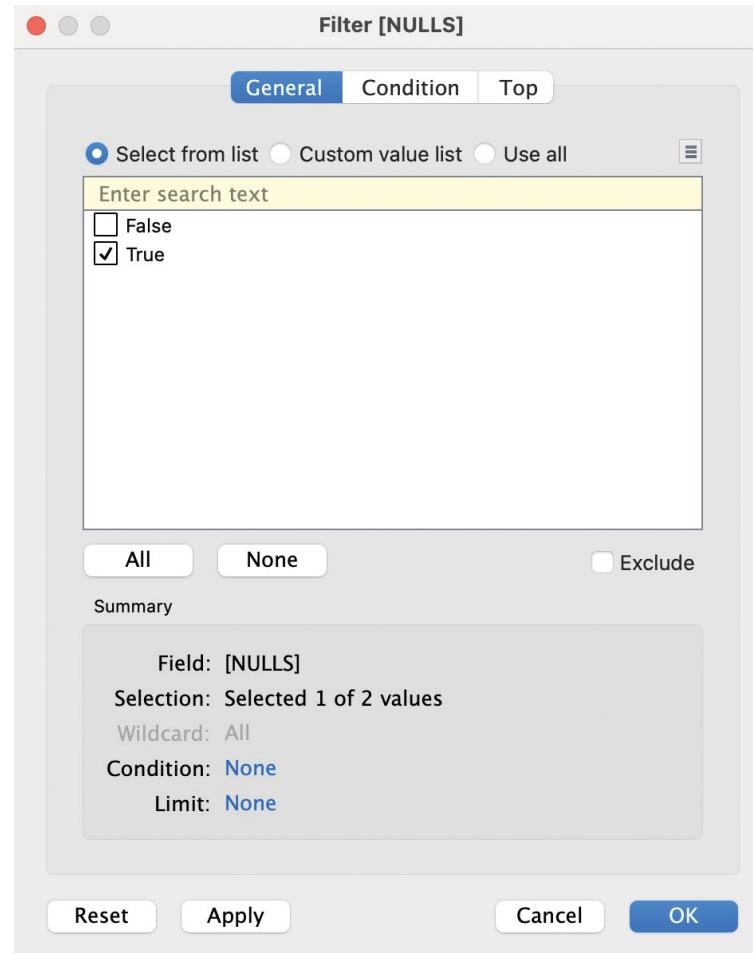


Figure 2.72 Filtering the NULLS field.

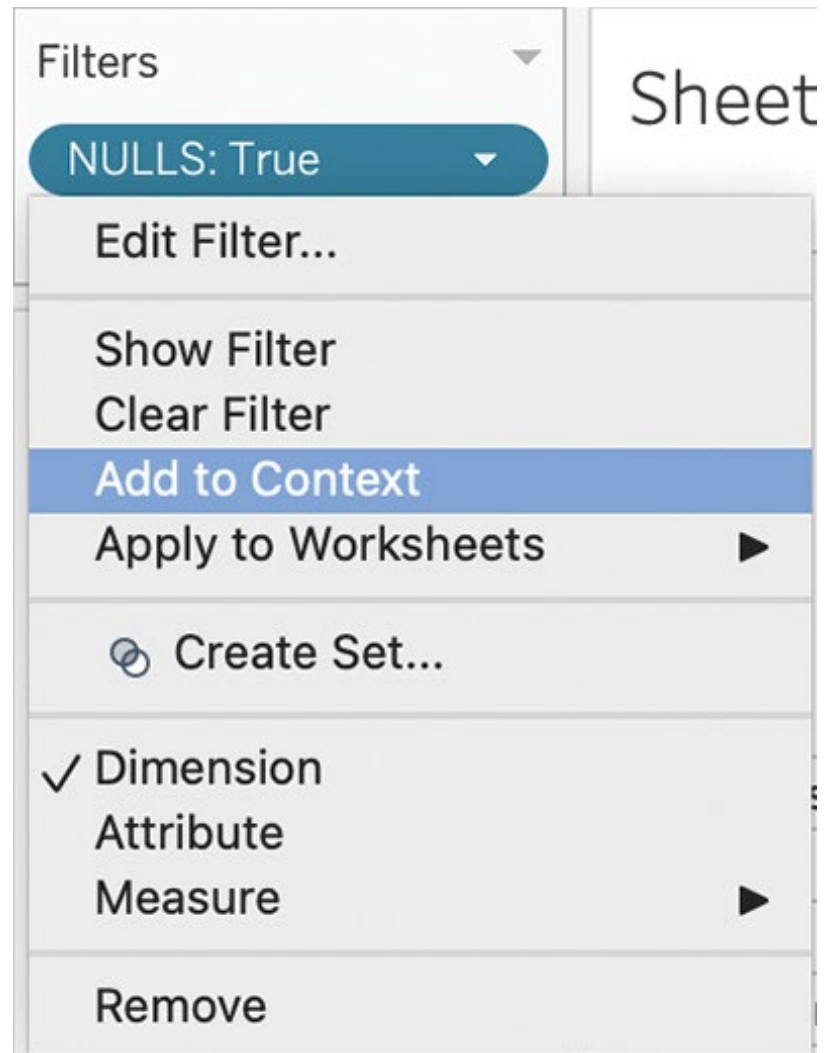


Figure 2.73 Choose Add to Context.

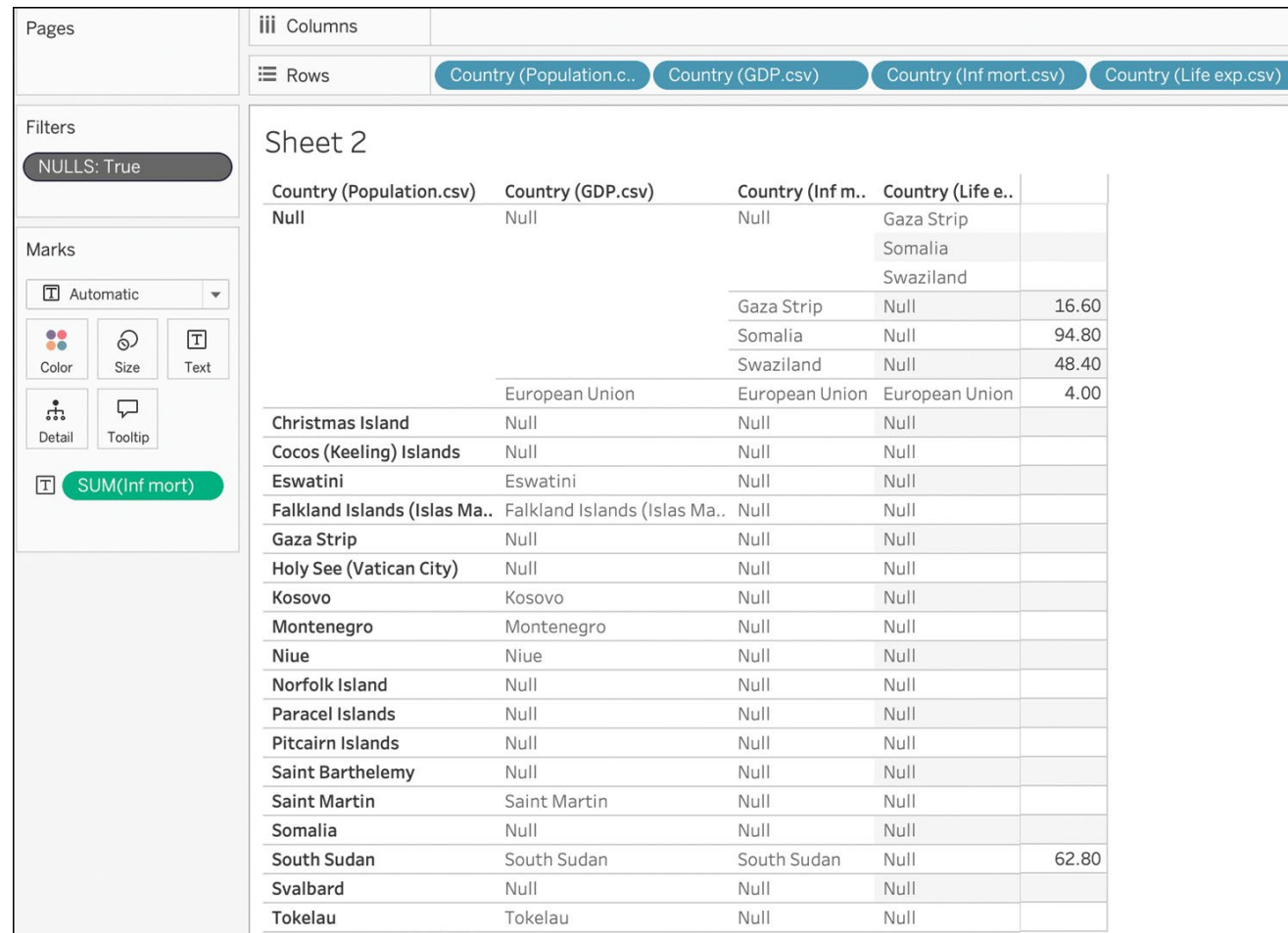


Figure 2.74 The mismatches between the four data files using the NULLS field.



Figure 2.75 A basic scatterplot of life expectancy by GDP per capita.



Figure 2.76 Adding in infant mortality.



Figure 2.77 Customizing the color scale.

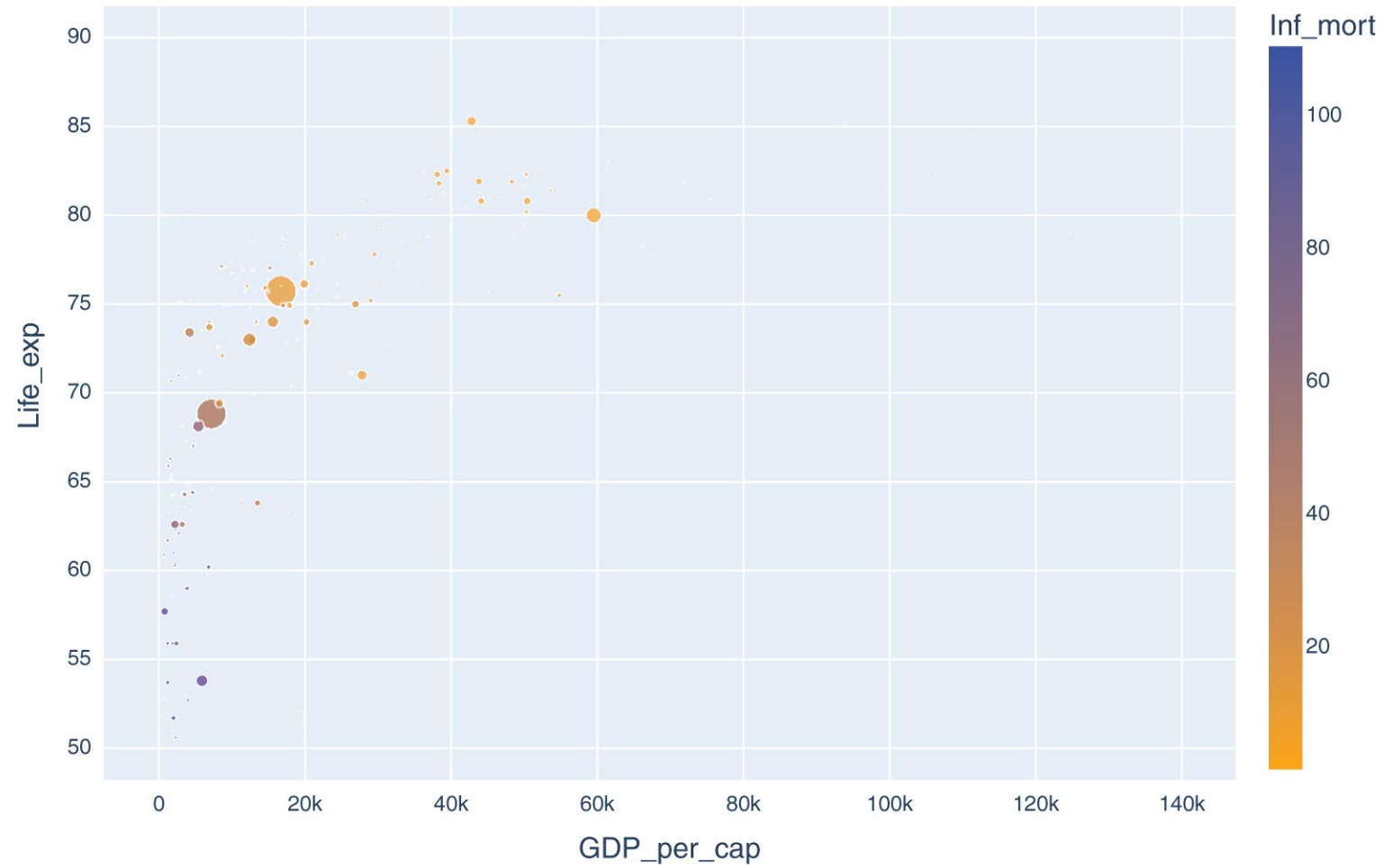


Figure 2.78 Adding the population data.

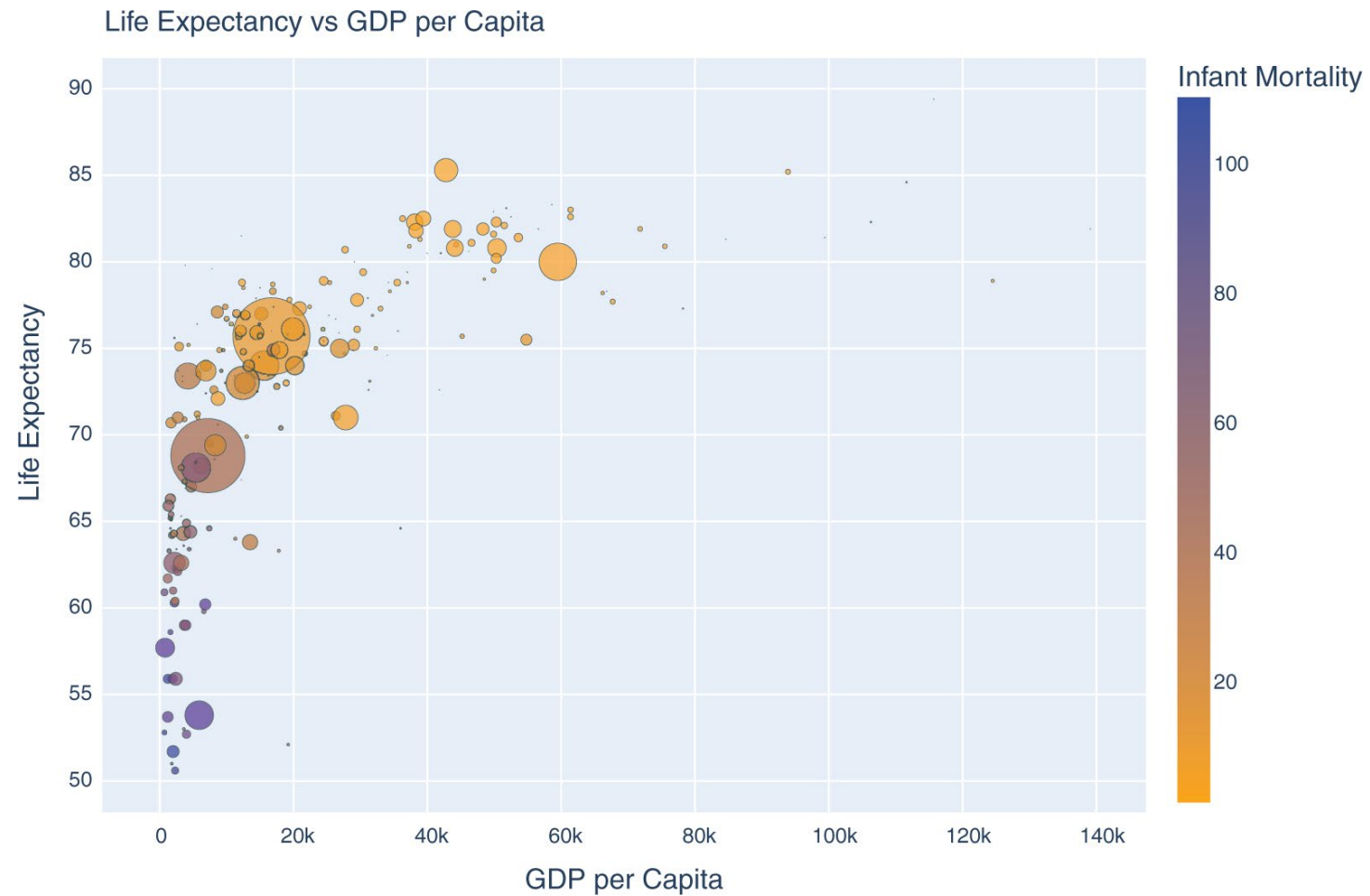


Figure 2.79 Changing the graph title and axis titles.

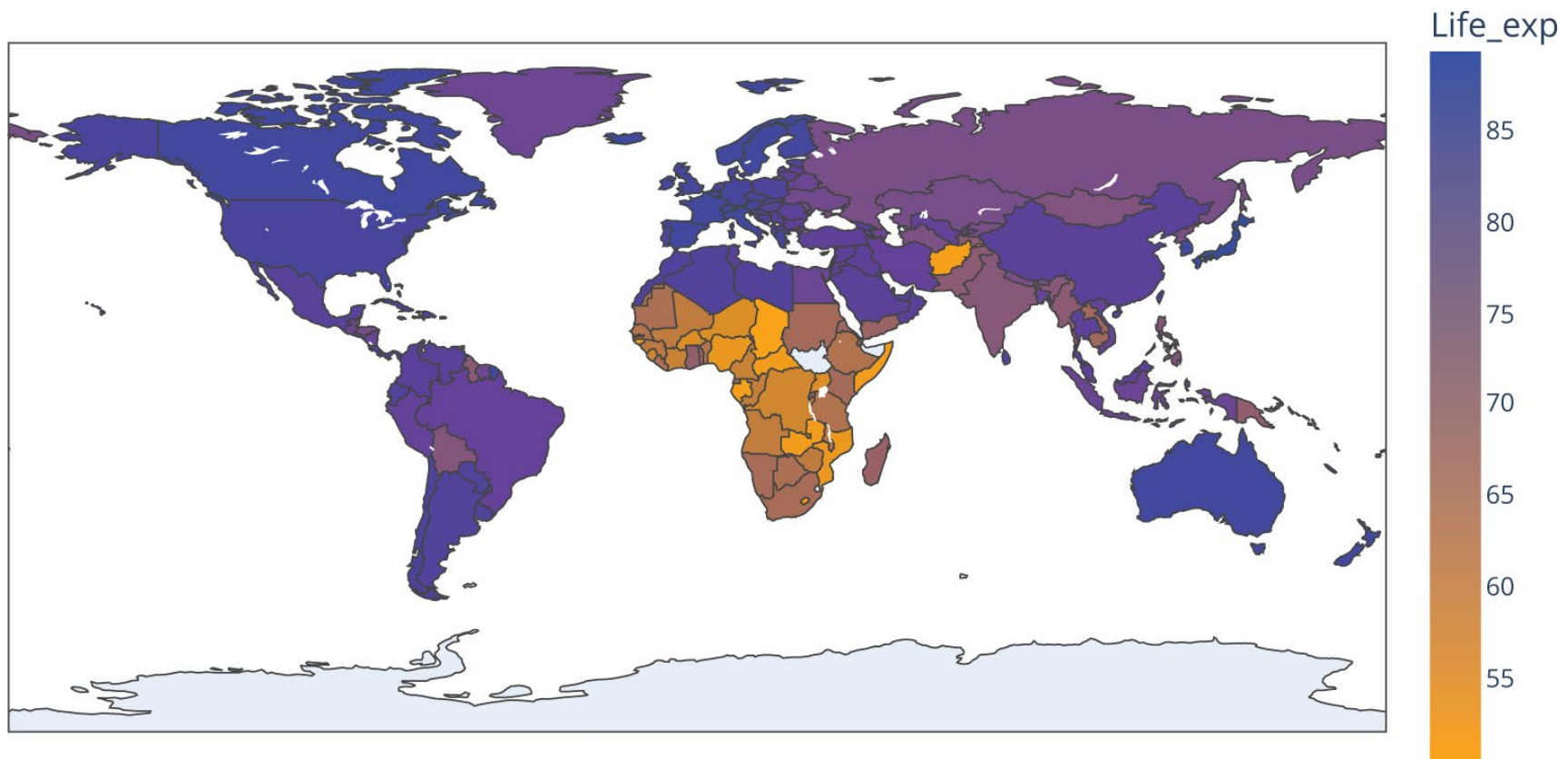


Figure 2.80 Using the choropleth function.

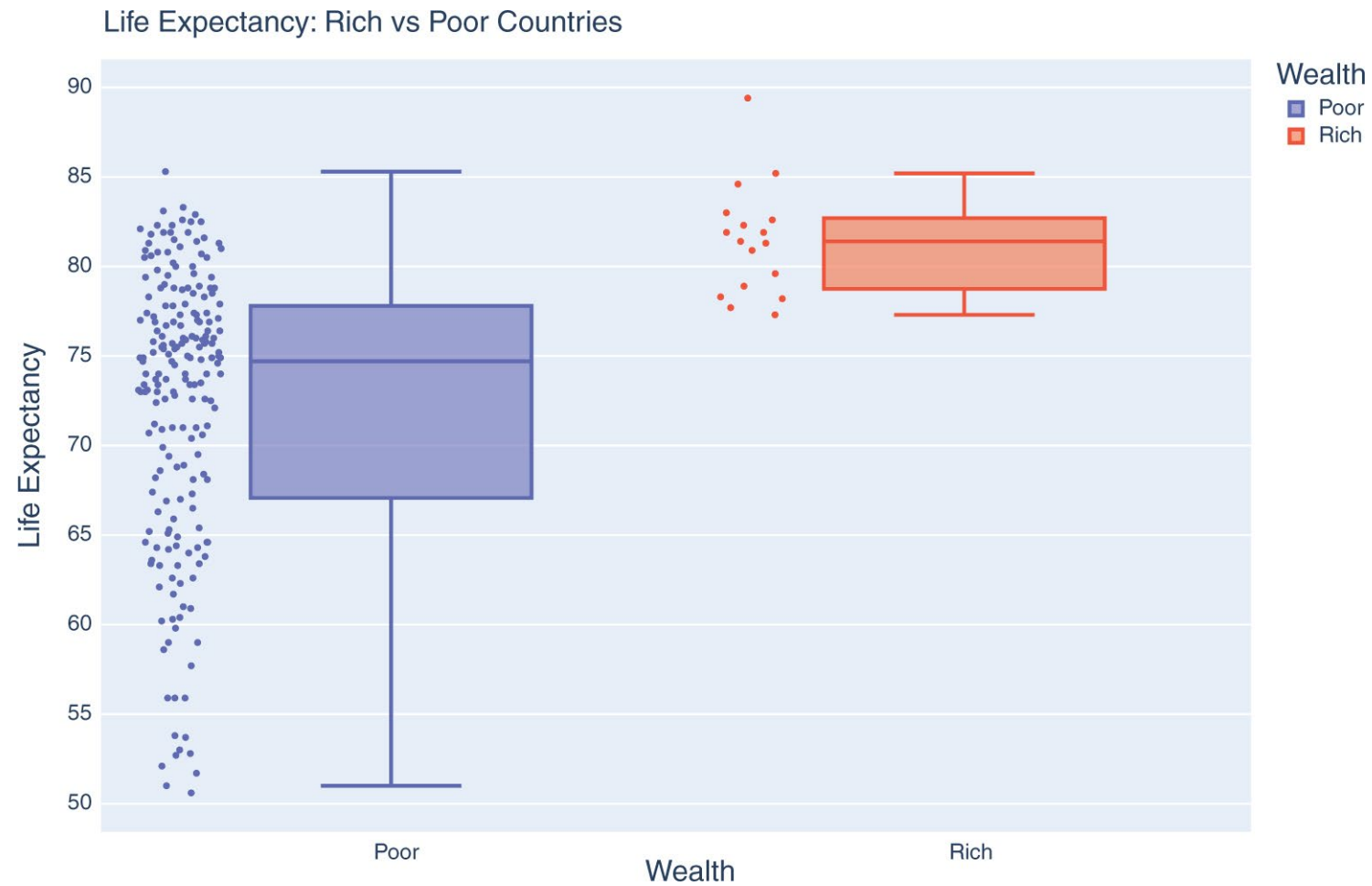


Figure 2.81 The box command creates a boxplot.

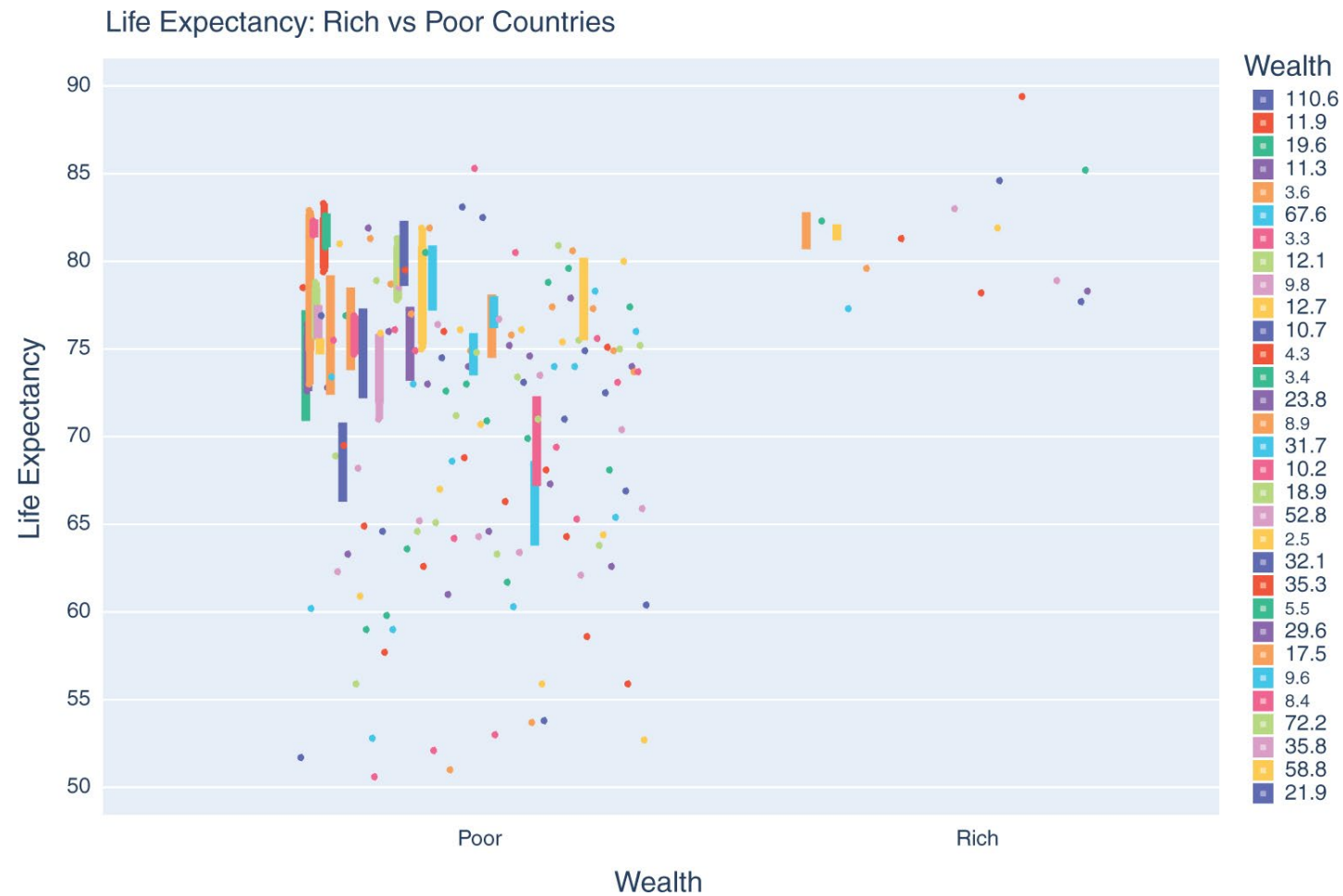


Figure 2.82 Not what we expected when trying to color by the Inf mort field.

Histograms for Country Measures



Figure 2.83 Histograms showing various country measures.

Distribution of GDP per Capita

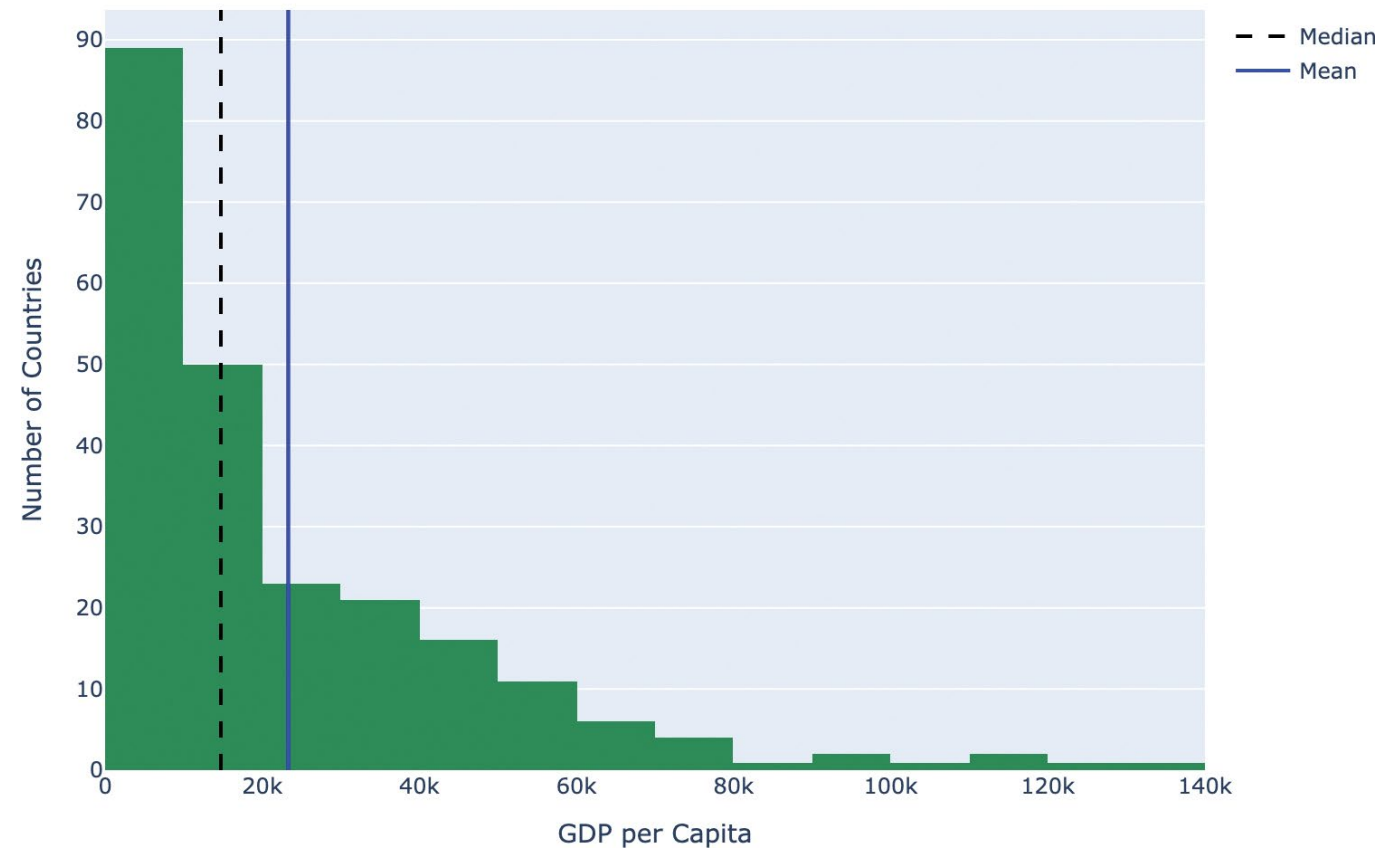


Figure 2.84 A single histogram for GDP per capita created using Plotly Express.

R Countries - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

+ +R + Go to file/function Addins

Countries_script.R x gdp x countries x world_map x pop x

← → ↗ Filter

	GDP_rank	Country	GDP_per_cap
1	1	Liechtenstein	139100
2	2	Qatar	124500
3	3	Monaco	115700
4	4	Macau	111600
5	5	Luxembourg	106300
6	6	Bermuda	99400

Figure 2.85 Viewing the gdp data.frame in RStudio.

R Countries - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

+ - +R - - - - - Go to file/function - Addins -

Countries_script.R x gdp x countries x world_map x pop x

← → - Filter

	Country	GDP_rank	GDP_per_cap	Pop_rank	Population	IM_rank	Inf_I
1	Afghanistan	207	2000	39	34124811	1	
2	Albania	125	12500	136	3047987	124	
3	Algeria	111	15200	33	40969443	81	
4	American Samoa	136	11200	210	51504	128	
5	Andorra	32	49900	200	76965	199	
6	Angola	161	6800	46	29310273	12	
7	Anguilla	130	12200	220	17087	211	

Figure 2.86 The merged countries data.frame.

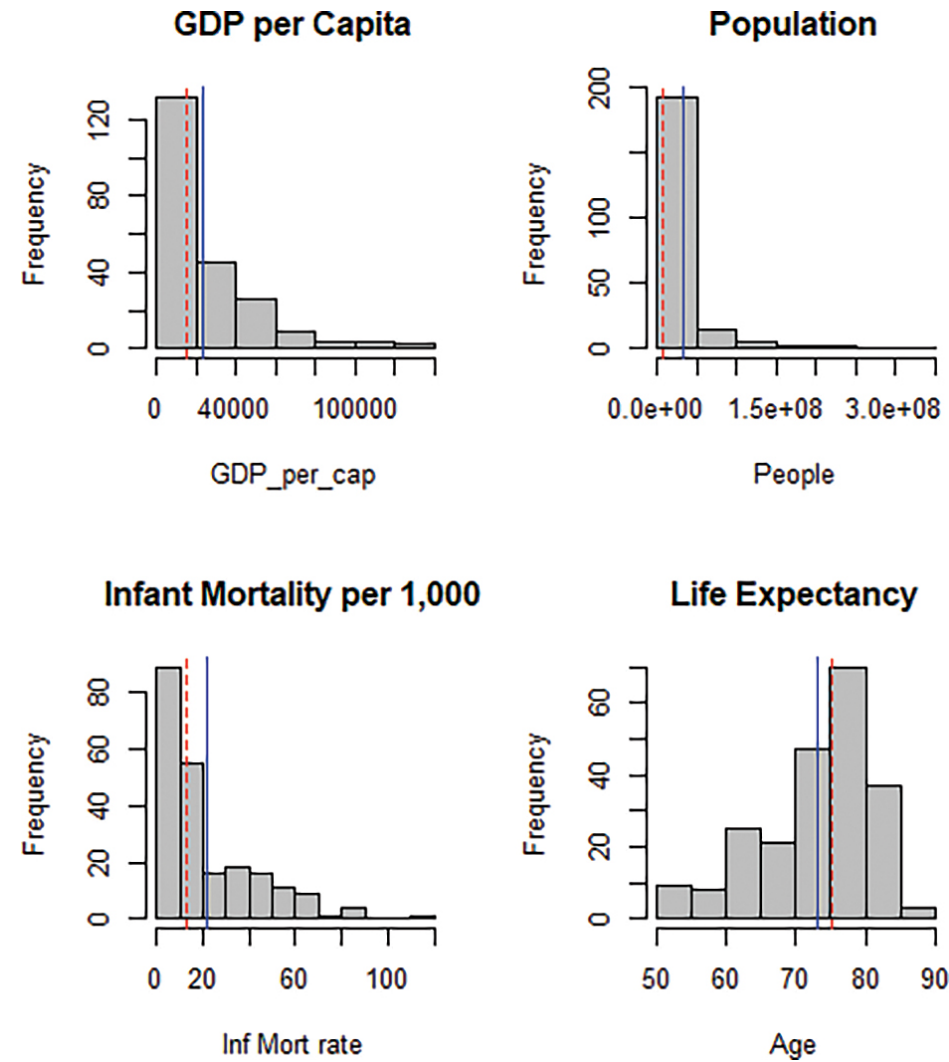


Figure 2.87 The histograms.

lty

The line type. Line types can either be specified as an integer (0=blank, 1=solid (default), 2=dashed, 3=dotted, 4=dotdash, 5=longdash, 6=twodash) or as one of the character strings "blank", "solid", "dashed", "dotted", "dotdash", "longdash", or "twodash", where "blank" uses 'invisible lines' (i.e., does not draw them).

Alternatively, a string of up to 8 characters (from `c(1:9, "A":"F")`) may be given, giving the length of line segments which are alternatively drawn and skipped. See section 'Line Type Specification'.

Functions such as `lines` and `segments` accept a vector of values which are recycled.

lwd

The line width, a *positive* number, defaulting to 1. The interpretation is device-specific, and some devices do not implement line widths less than one. (See the help on the device for details of the interpretation.)

Functions such as `lines` and `segments` accept a vector of values which are recycled: in such uses lines corresponding to values NA or NaN are omitted. The interpretation of 0 is device-specific.

Figure 2.88 The help results for the par function.

```

> bins = quantile(countries$Population, prob = (0:10)/10)
> bins
      0%      10%      20%      30%      40%      50%      60%
5292.0  65695.9  282065.2  1220546.7  3238393.4  5903922.5  9960850.6
      70%      80%      90%     100%
16529054.8  29457209.4  62488831.8 1379302771.0
> vals = cut(countries$Population, bins, labels = F, include.lowest = T)
> vals
 [1] 9  4  9  1  2  8  1  2  9  4  2  8  6  7  3  4 10  3  6  7  3  7  2  3  7  5  4 10
[29] 1  3  6  8  9  7  3  7  8  9  1  5  7  8 10  9  3 10  5  1  5  8  5  7  2  4  7  5
[57] 3  2  7  7 10  6  3  6  4 10  1  3  5 10  3  4  4  5 10  8  1  7  1  2  2  7  2  7
[85] 4  3  7  6  6  6  3 10 10 10  9  5  2  6  9  4 10  2  7  8  9  2  8  9  4  5  6  4
[113] 6  4  5  6  1  4  3  3  4  8  8  9  3  8  3  2  5  4 10  2  5  1  4  1  9  8  4  1
[141] 8  8  2  5  6  8 10  1  5  5 10  1  5  6  6  9 10  9  7  5  4  8 10  7  1  1  2  1
[169] 2  2  1  2  8  7  6  2  6  5  1  5  4  3  9  9  8  9  3  6  6  8  8  6  9 10  4  6
[197] 2  3  7 10  5  1  1  9  9  6 10 10  5  9  3  9 10  2  1  4  3  8  7  7

```

Figure 2.89 The decile values.

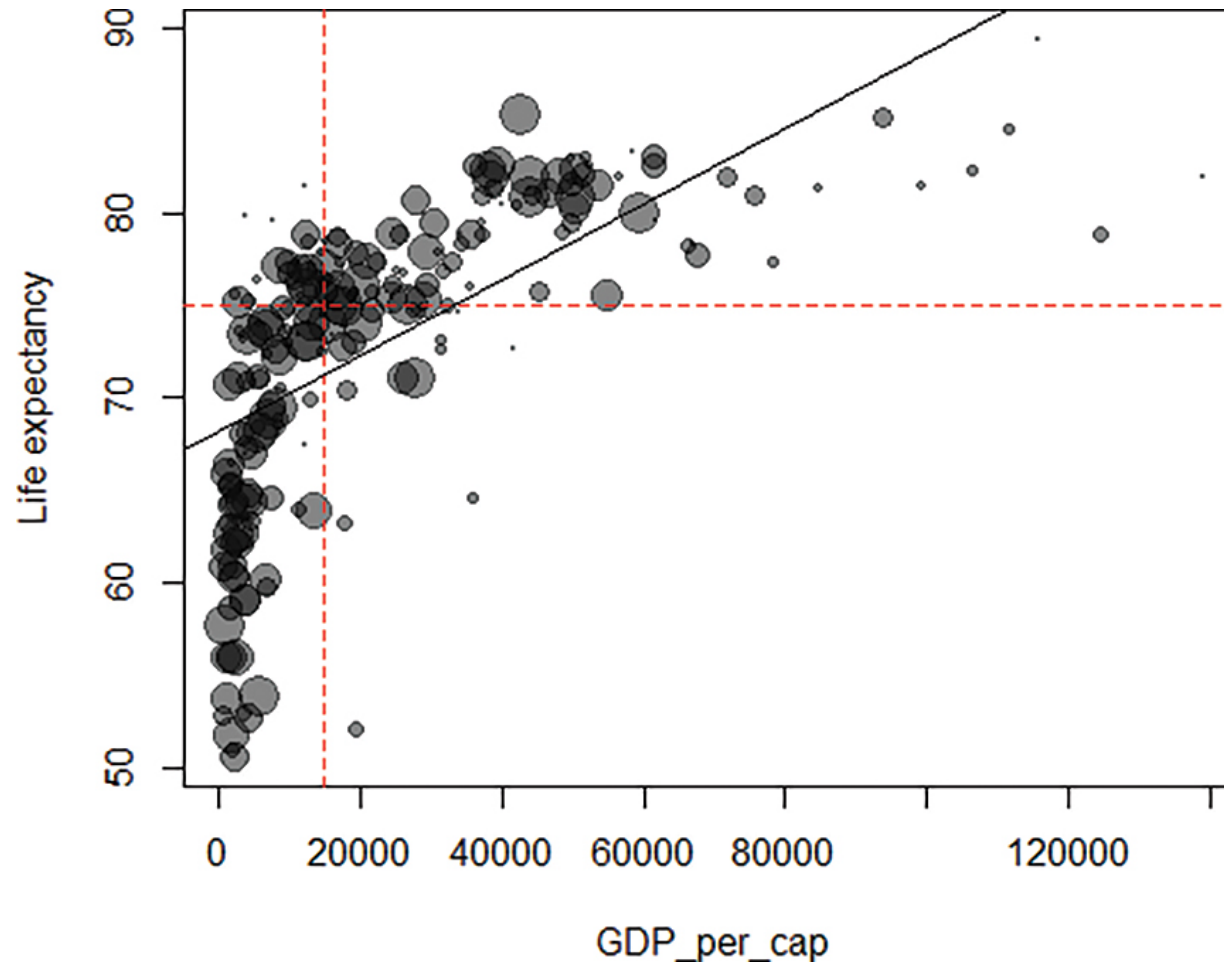


Figure 2.90 Base R scatterplot with lines for the means and the regression line.

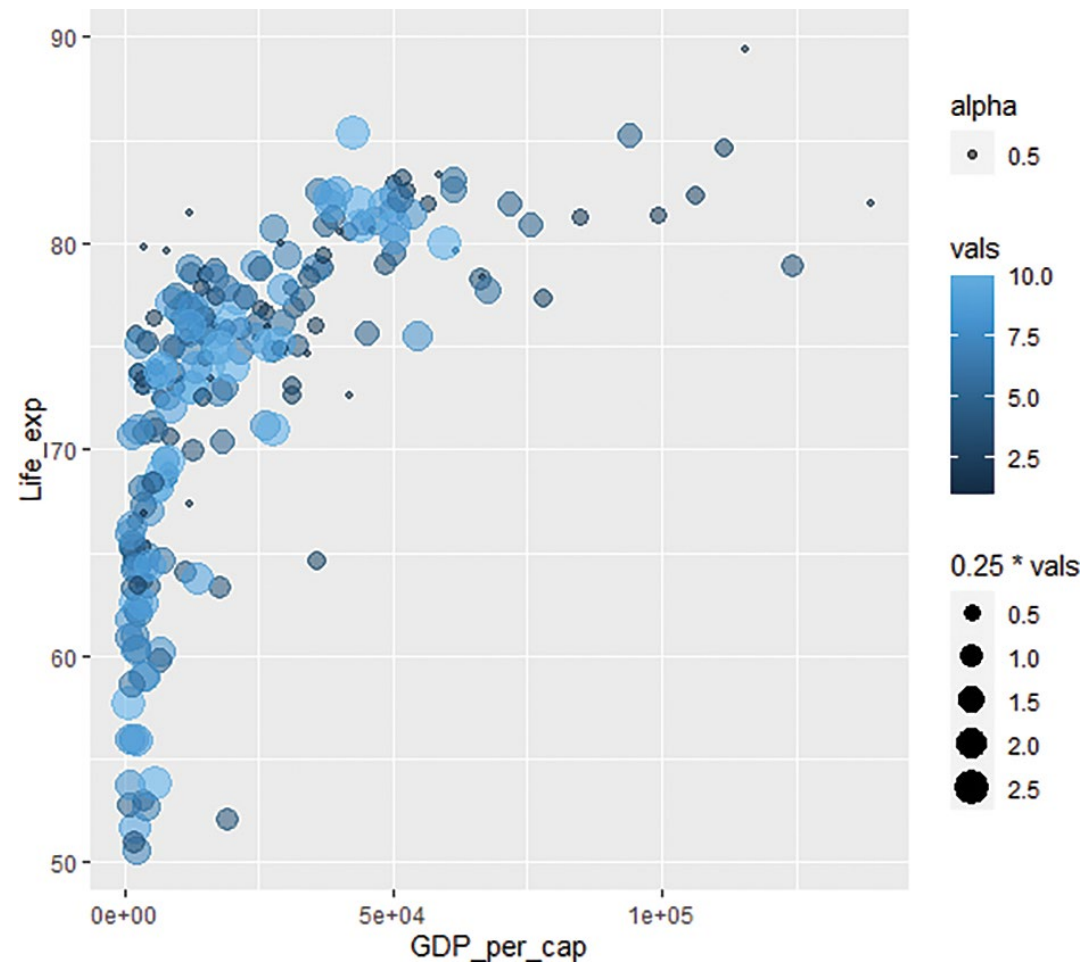


Figure 2.91 Another version of the scatterplot using `qplot`.



Figure 2.92 Scatterplot with correctly sized bubbles: The area (not radius) is proportional to the population.

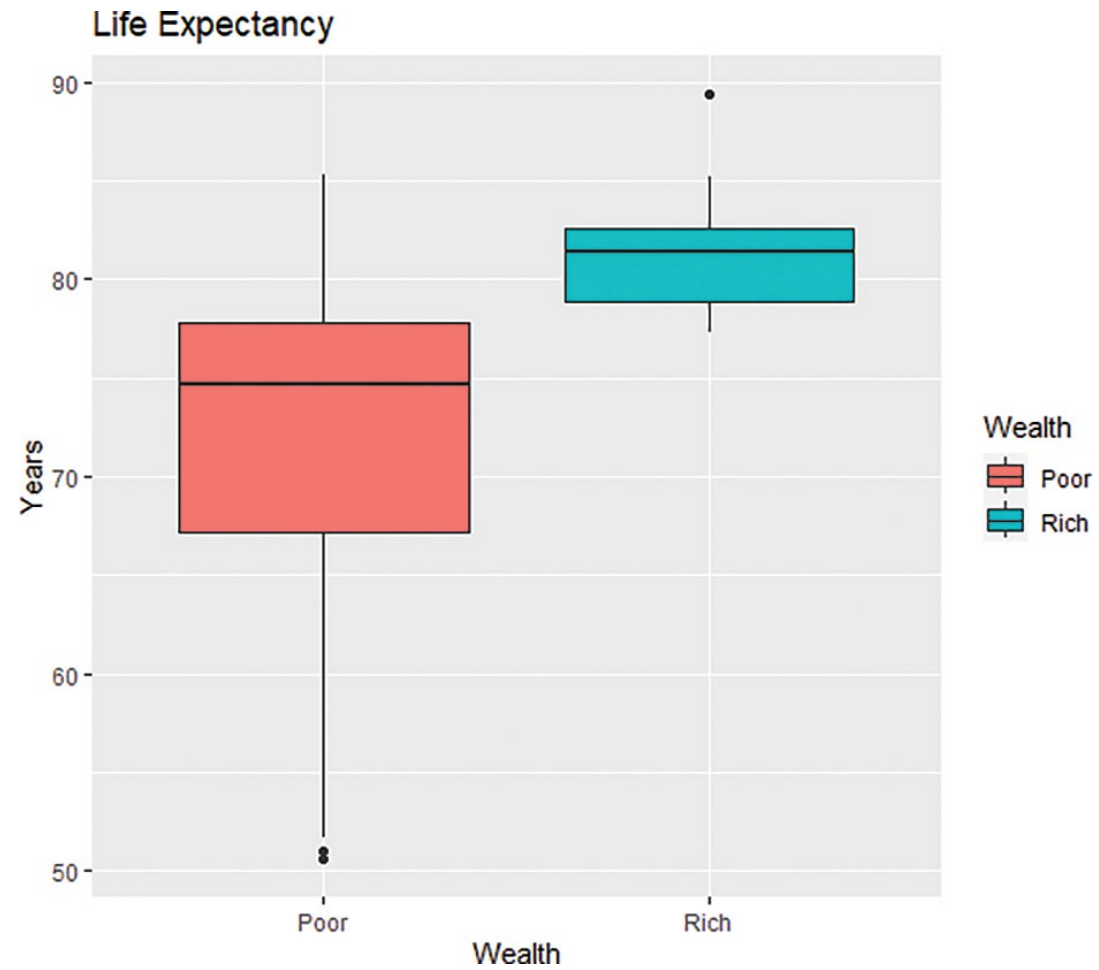


Figure 2.93 The side-by-side boxplots.

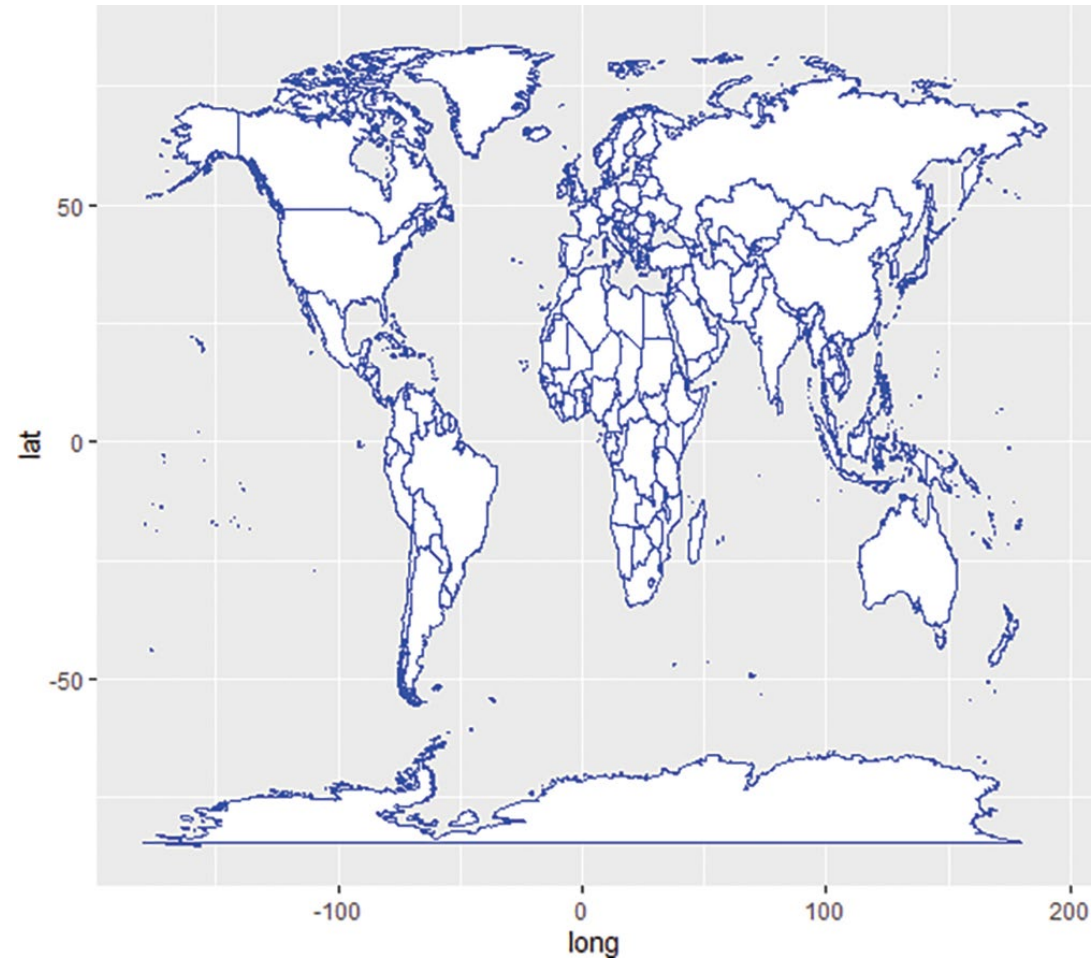


Figure 2.94 The boundaries of each country.

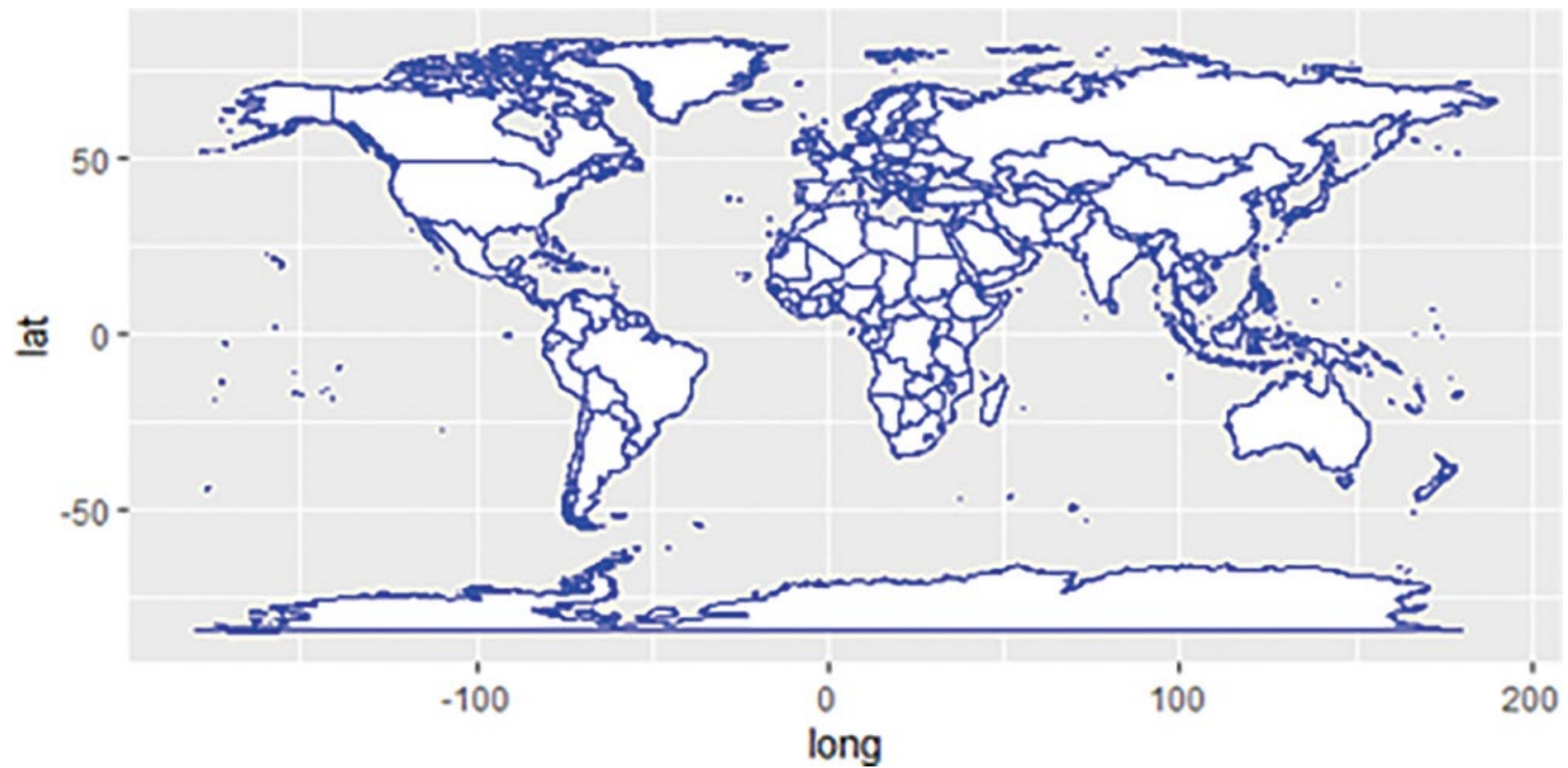


Figure 2.95 The corrected map.

```

> countries$region <- countries$Country
> head(countries)
  Country GDP_rank GDP_per_cap Pop_rank Population IM_rank Inf_mort LE_rank
1  Afghanistan    207      2000      39   34124811      1    110.6    222
2    Albania     125     12500     136    3047987    124     11.9     59
3    Algeria     111     15200      33   40969443     81     19.6     78
4 American Samoa   136     11200     210      51504    128     11.3    134
5    Andorra      32     49900     200      76965    199      3.6      8
6    Angola     161      6800      46   29310273     12     67.6    205

  Life_exp wealth      region
1    51.7   Poor  Afghanistan
2    78.5   Poor    Albania
3    77.0   Poor    Algeria
4    73.4   Poor American Samoa
5    82.9   Poor    Andorra
6    60.2   Poor    Angola

> head(world_map)
  long      lat group order region subregion
1 -69.89912 12.45200     1     1  Aruba      <NA>
2 -69.89571 12.42300     1     2  Aruba      <NA>
3 -69.94219 12.43853     1     3  Aruba      <NA>
4 -70.00415 12.50049     1     4  Aruba      <NA>
5 -70.06612 12.54697     1     5  Aruba      <NA>
6 -70.05088 12.59707     1     6  Aruba      <NA>

```

Figure 2.96 The new column in the countries data.frame named region.


```

> choro_df = merge(world_map, countries, by = "region", all.x = TRUE)
> head(choro_df)

```

	region	long	lat	group	order	subregion	Country	GDP_rank	GDP_per_cap
1	Afghanistan	73.73183	36.88779	2	30	<NA>	Afghanistan	207	2000
2	Afghanistan	74.00185	36.82310	2	27	<NA>	Afghanistan	207	2000
3	Afghanistan	62.61055	35.23315	2	263	<NA>	Afghanistan	207	2000
4	Afghanistan	66.59580	31.01997	2	143	<NA>	Afghanistan	207	2000
5	Afghanistan	74.03887	36.82573	2	26	<NA>	Afghanistan	207	2000
6	Afghanistan	71.53085	36.84512	2	385	<NA>	Afghanistan	207	2000

	Pop_rank	Population	IM_rank	Inf_mort	LE_rank	Life_exp	Wealth
1	39	34124811	1	110.6	222	51.7	Poor
2	39	34124811	1	110.6	222	51.7	Poor
3	39	34124811	1	110.6	222	51.7	Poor
4	39	34124811	1	110.6	222	51.7	Poor
5	39	34124811	1	110.6	222	51.7	Poor
6	39	34124811	1	110.6	222	51.7	Poor

```

> length(unique(choro_df$region))
[1] 252

```

Figure 2.97 Merging to keep all the regions from world_map.

```

> choro <- choro_df[order(choro_df$order), ]
> head(choro)

```

	region	long	lat	group	order	subregion	Country	GDP_rank	GDP_per_cap	Pop_rank
7142	Aruba	-69.89912	12.45200	1	1	<NA>	Aruba	81	25300	189
7140	Aruba	-69.89571	12.42300	1	2	<NA>	Aruba	81	25300	189
7134	Aruba	-69.94219	12.43853	1	3	<NA>	Aruba	81	25300	189
7135	Aruba	-70.00415	12.50049	1	4	<NA>	Aruba	81	25300	189
7139	Aruba	-70.06612	12.54697	1	5	<NA>	Aruba	81	25300	189
7136	Aruba	-70.05088	12.59707	1	6	<NA>	Aruba	81	25300	189

	Population	IM_rank	Inf_mort	LE_rank	Life_exp	wealth
7142	115120	131	10.7	80	76.9	Poor
7140	115120	131	10.7	80	76.9	Poor
7134	115120	131	10.7	80	76.9	Poor
7135	115120	131	10.7	80	76.9	Poor
7139	115120	131	10.7	80	76.9	Poor
7136	115120	131	10.7	80	76.9	Poor

Figure 2.98 The blank space in the square brackets indicates we want all the columns of choro_df.

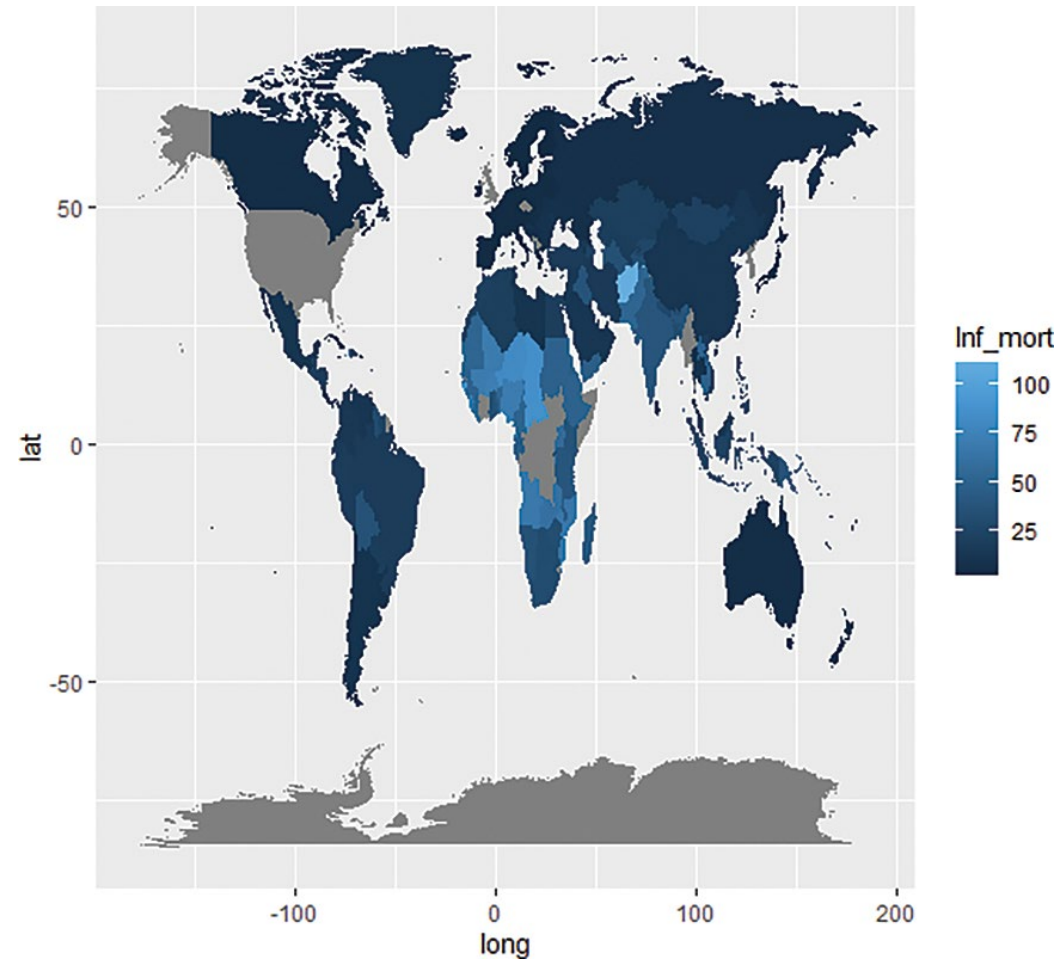


Figure 2.99 Our map from the quick plot function.

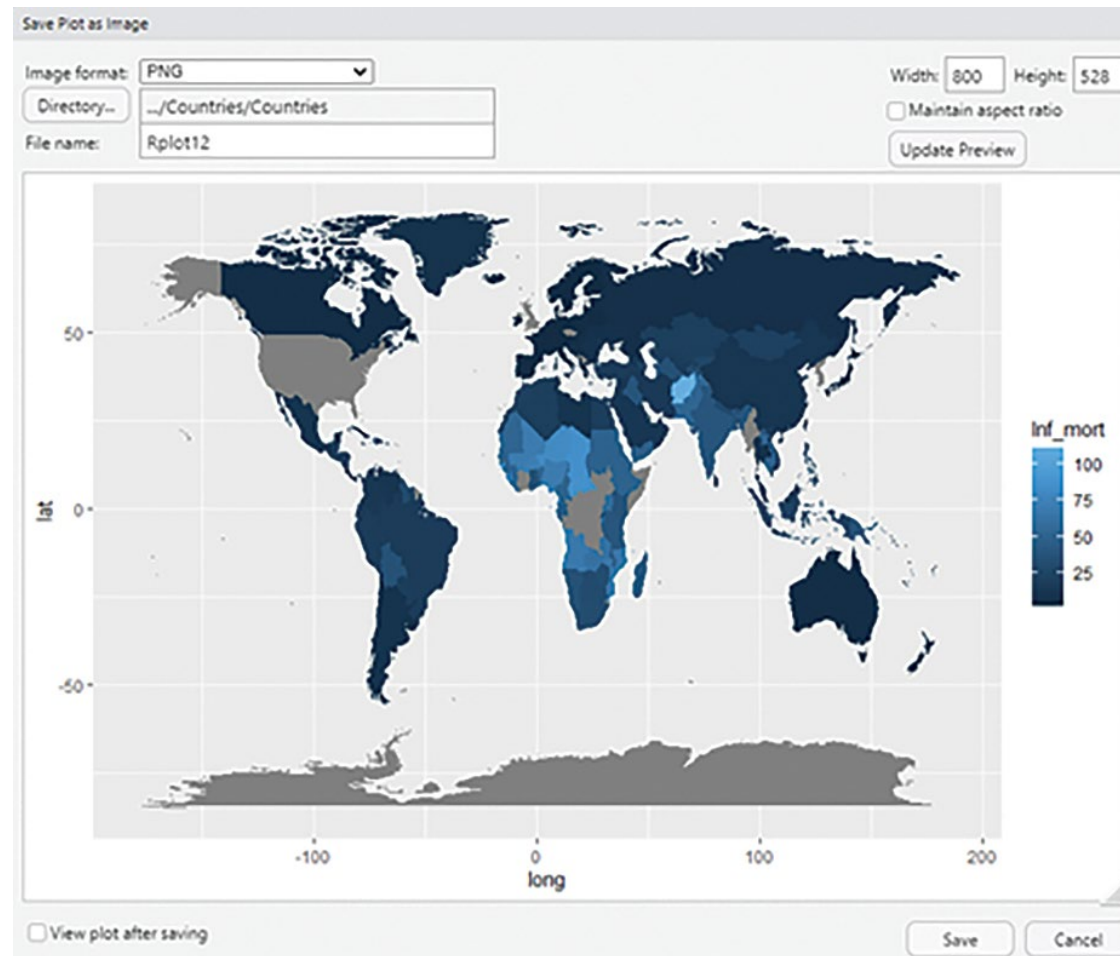


Figure 2.100 The resized map.

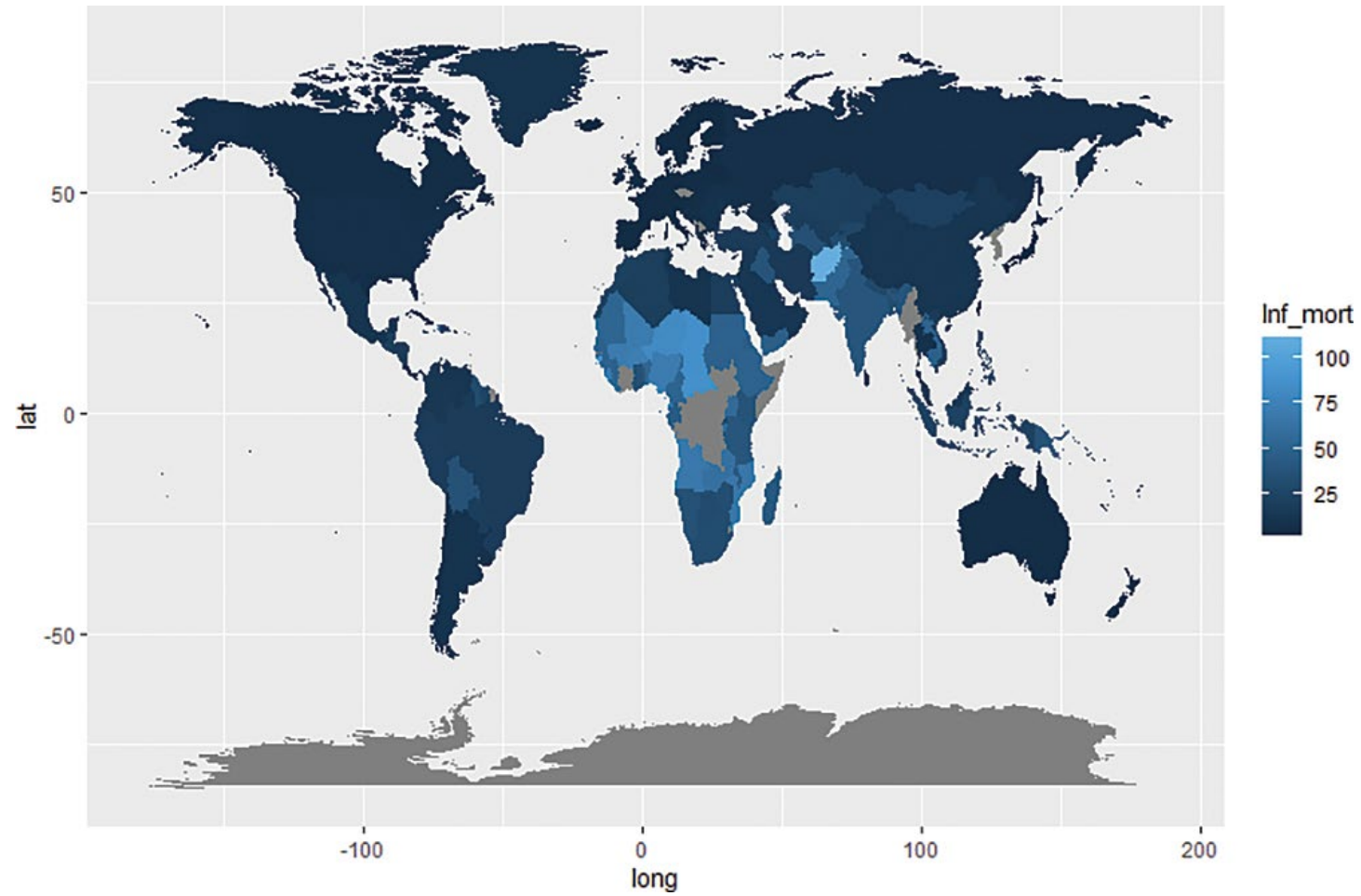


Figure 2.101 Heat map showing infant mortality by country.

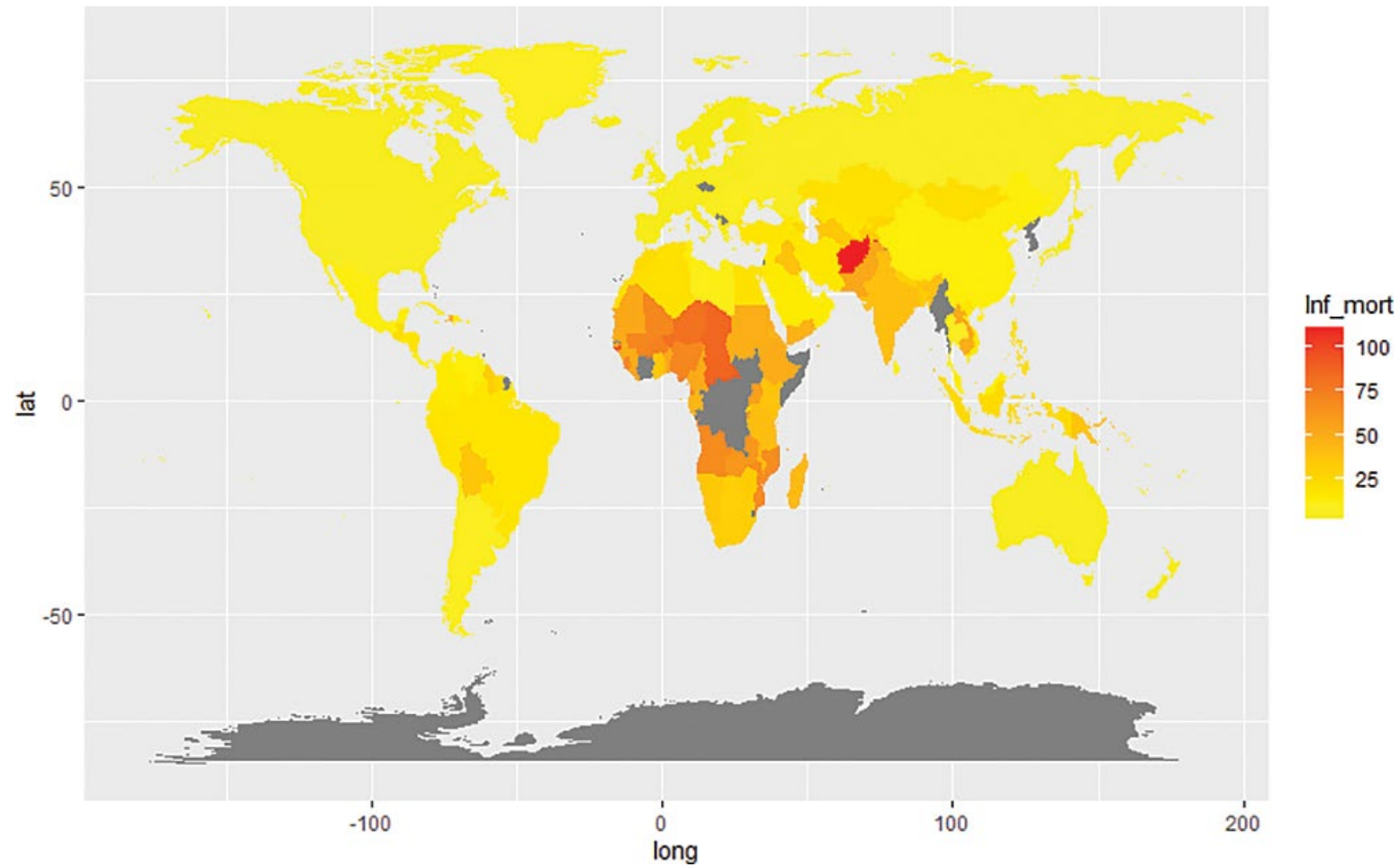


Figure 2.102 The yellow-to-red color scheme for infant mortality is too yellow.

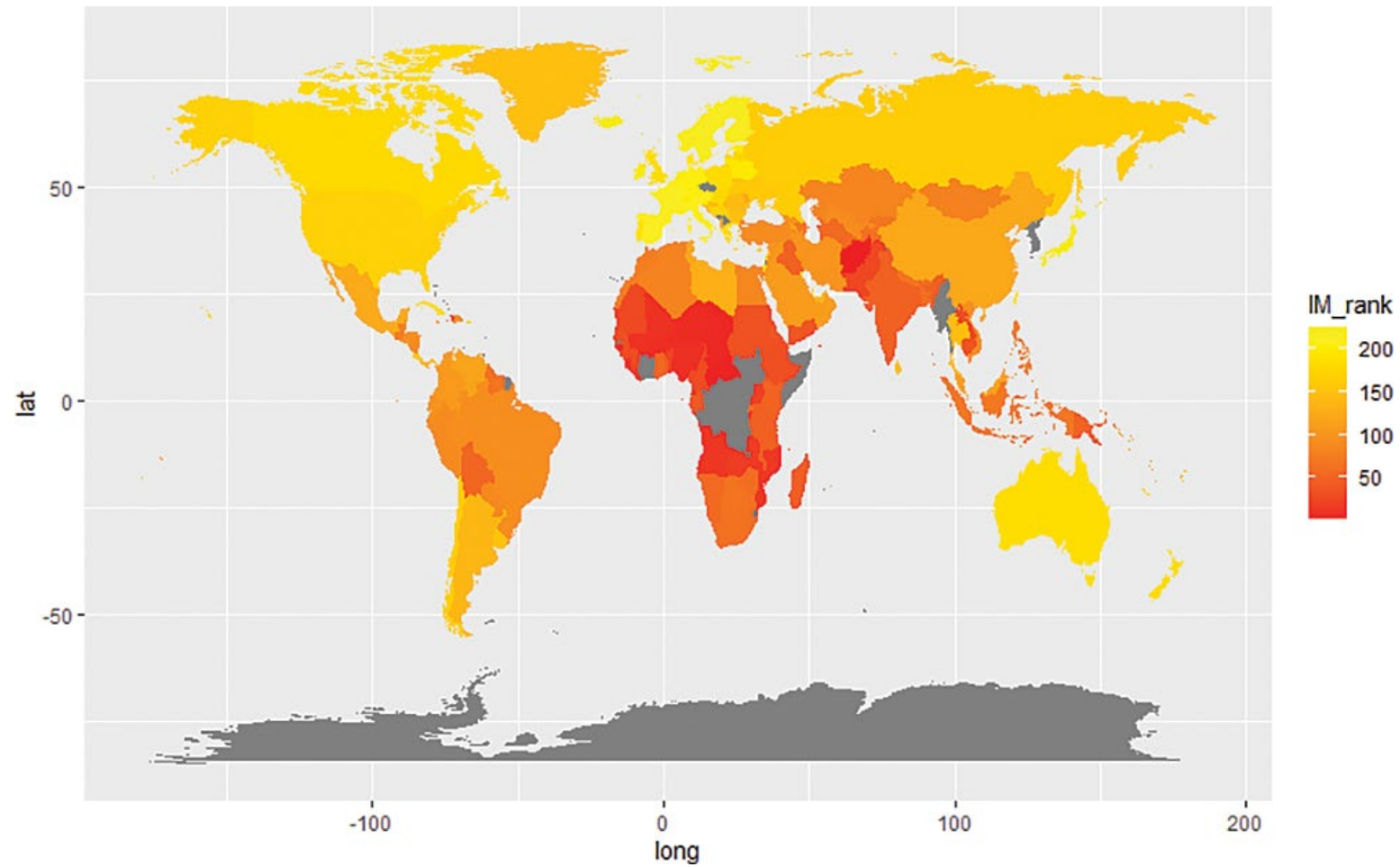


Figure 2.103 Filling by infant mortality rank.

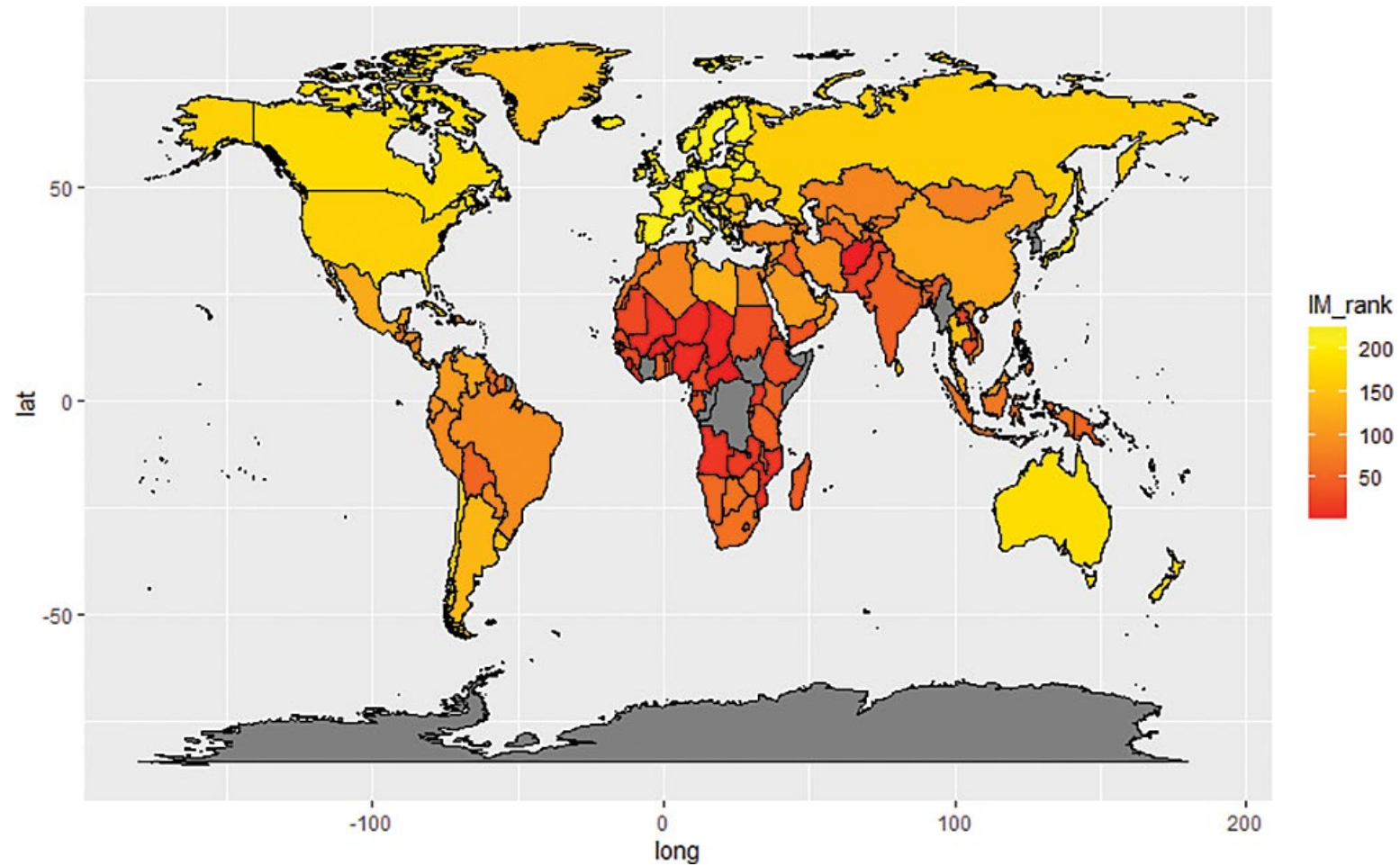


Figure 2.104 Completed choropleth map by infant mortality rank with borders.

Comparing Elections

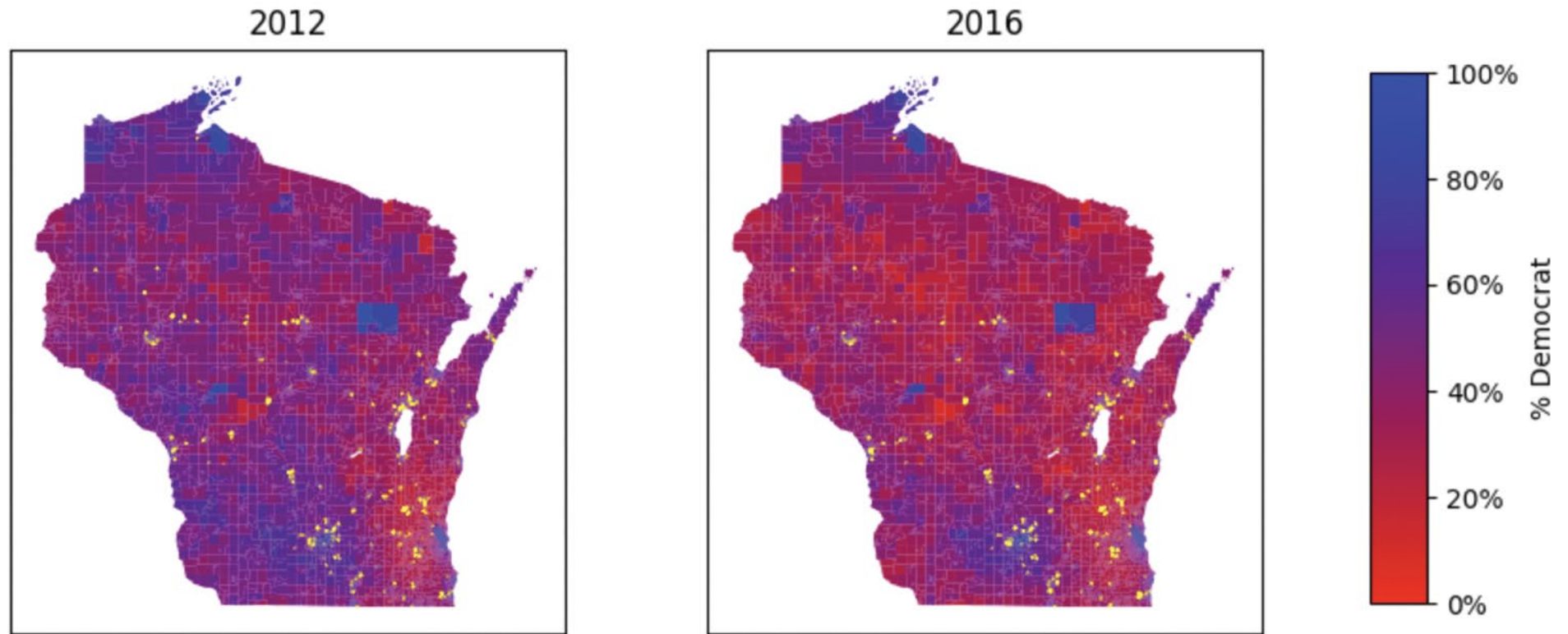


Figure 3.1 Side-by-side maps of Democratic vote share.

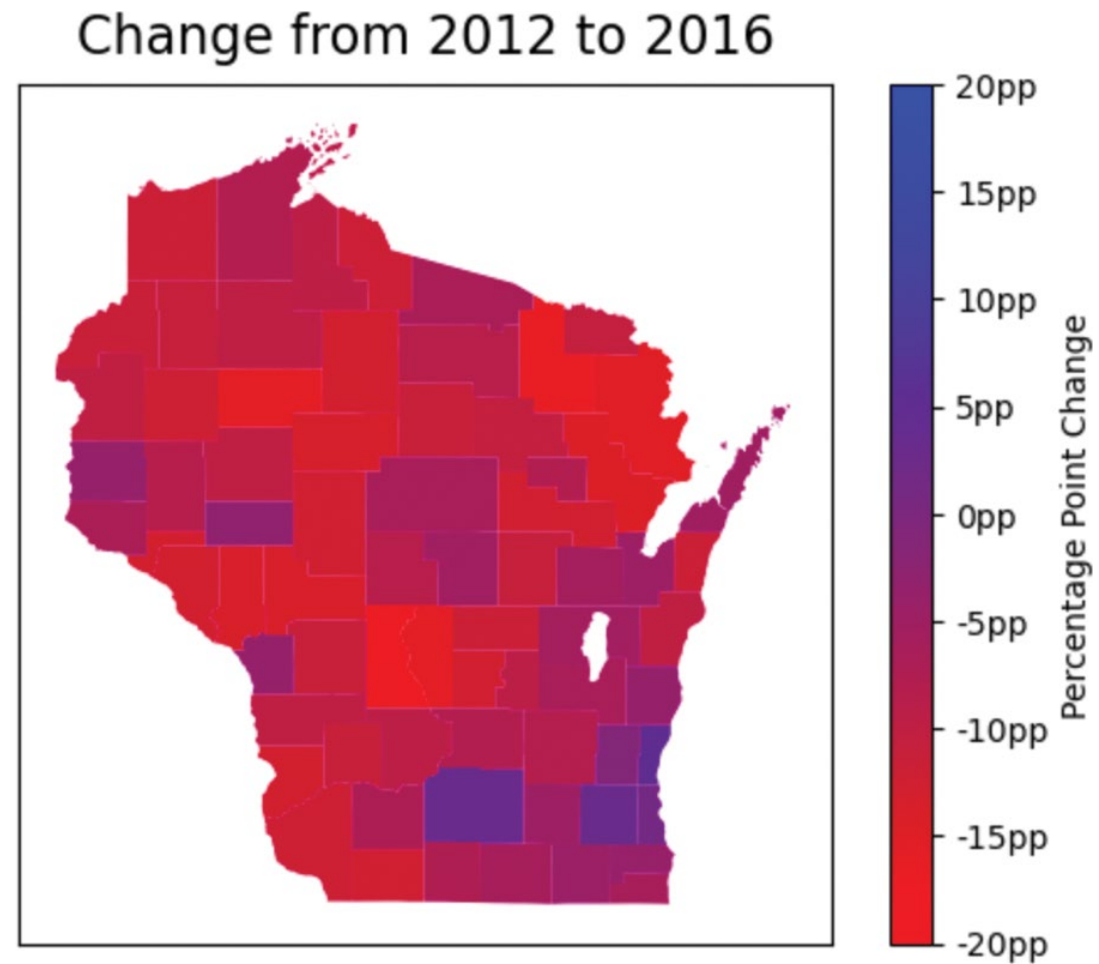


Figure 3.2 County change in the Democratic vote share.

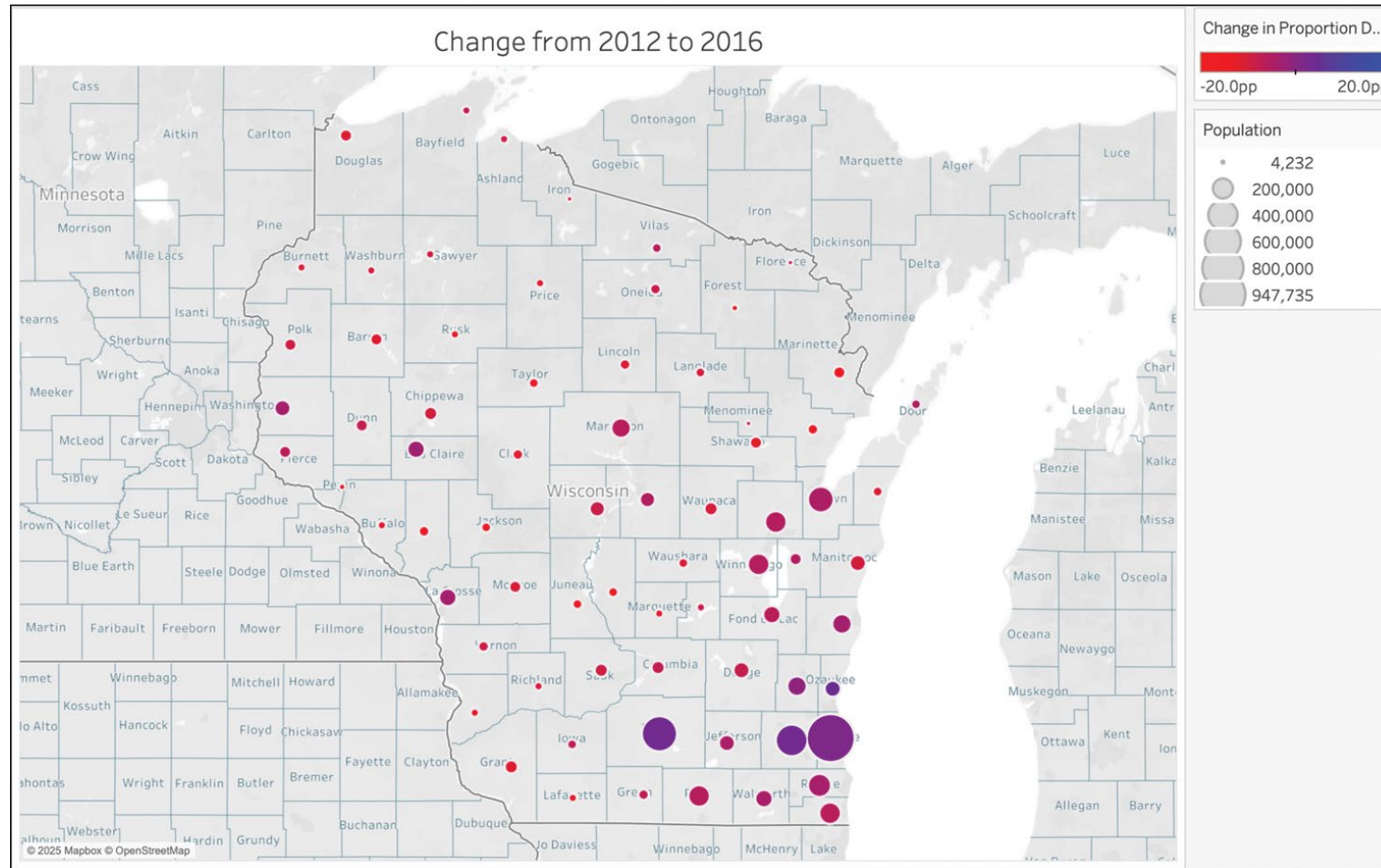


Figure 3.3 Bubble chart showing the change in Democratic vote share sized by county population.

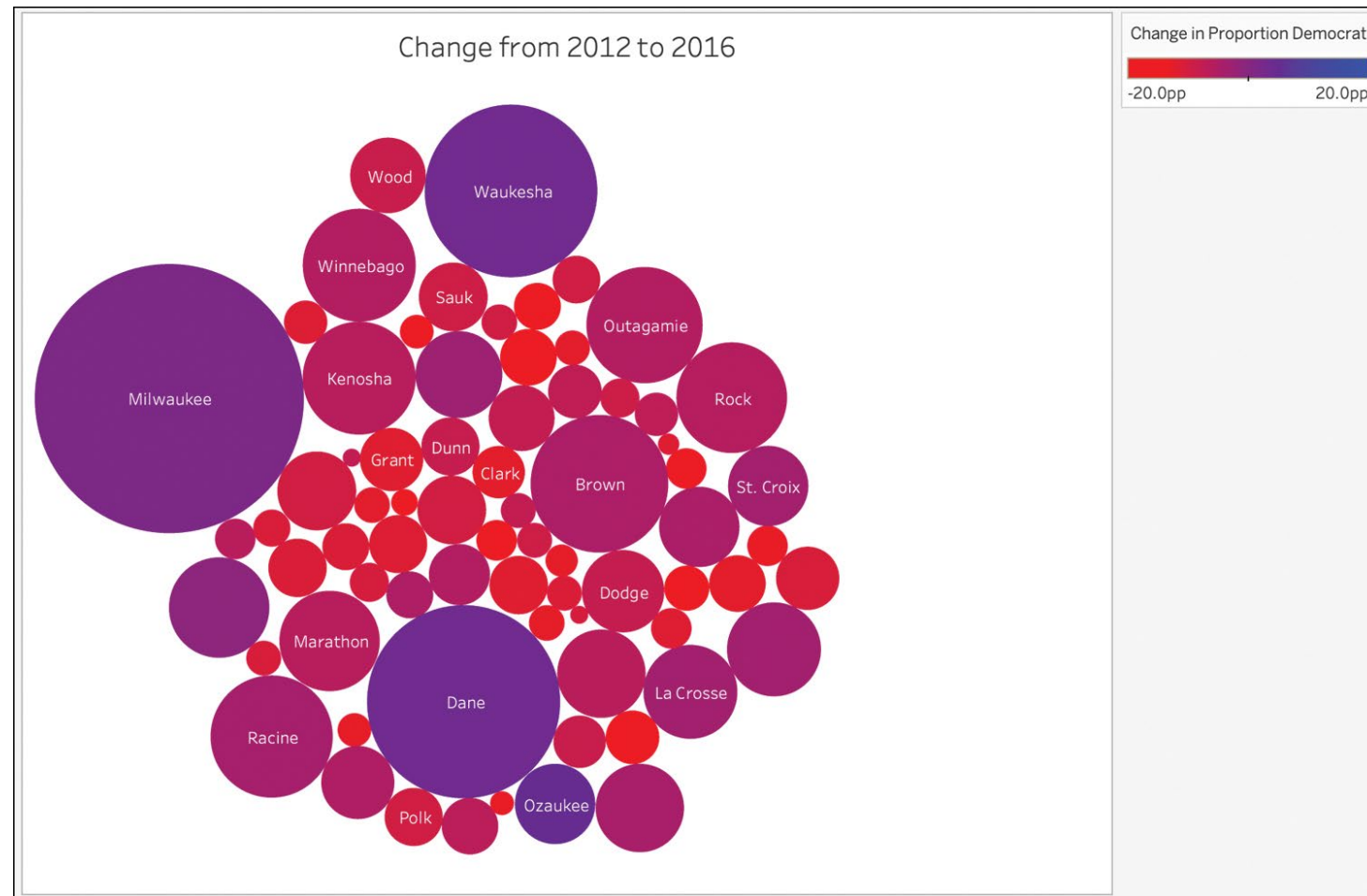


Figure 3.4 Packed circle chart showing the change in Democratic vote share sized by county population.

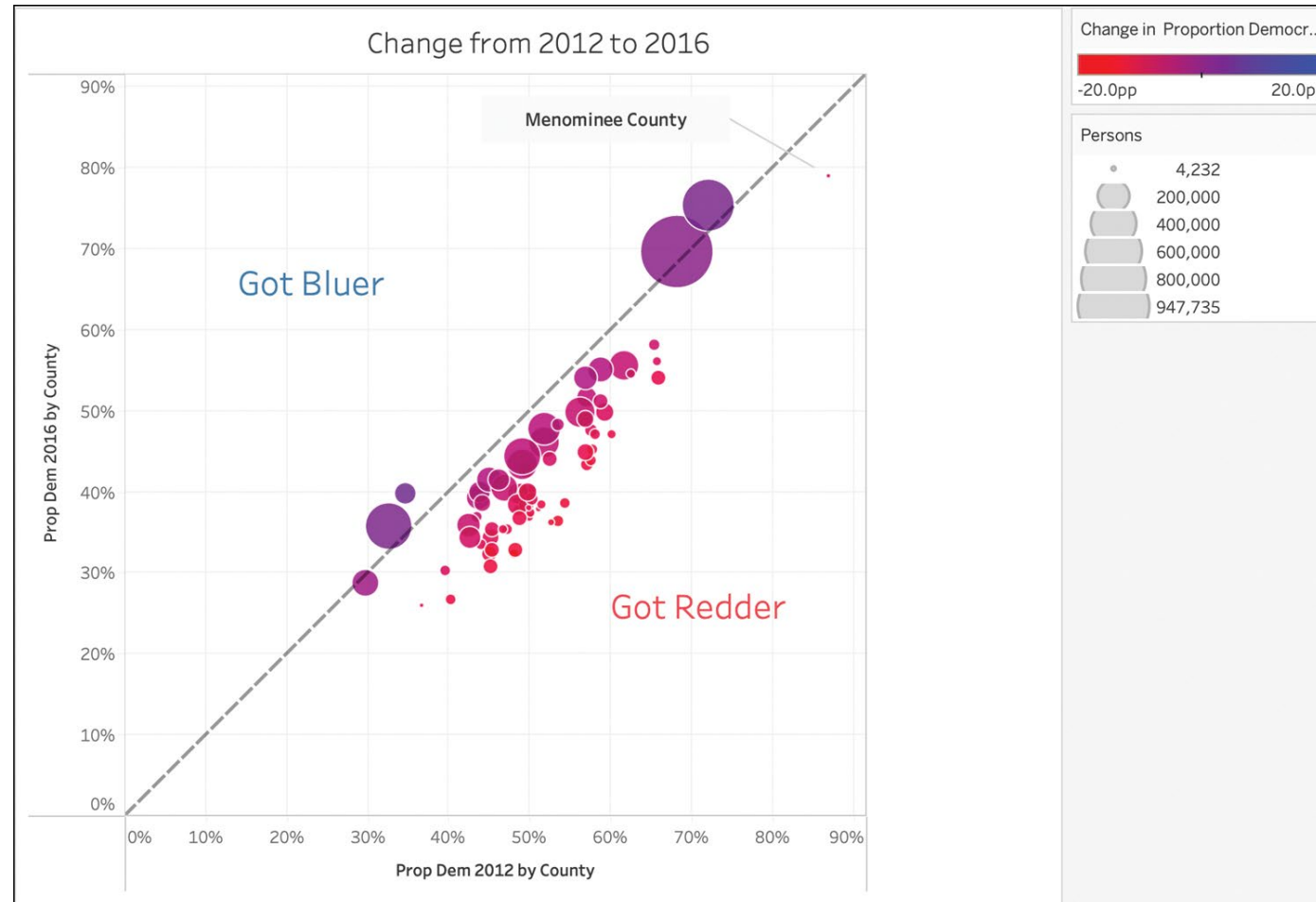


Figure 3.5 Scatterplot showing the change in Democratic vote share sized by county population.

Text Import Wizard - Step 2 of 3

This screen lets you set the delimiters your data contains.

Delimiters

☐ Tab
 ☐ Treat consecutive delimiters as one

☐ Semicolon
 Text qualifier: " ⌵

☒ Comma

☐ Space

☐ Other:

Preview of selected data:

OBJECTID	GEOID	CNTY_FIPS	CNTY_NAME	COUSUBFP	MCD_FIPS	MCD_NAME	CTV	LABEL	LSAD	NAME
1	55001002750001	55001	Adams	00275	5500100275	Adams	C	Adams - C 0001	25	Adams
2	55001002750002	55001	Adams	00275	5500100275	Adams	C	Adams - C 0002	25	Adams
3	55001002750003	55001	Adams	00275	5500100275	Adams	C	Adams - C 0003	25	Adams
4	55001002750004	55001	Adams	00275	5500100275	Adams	C	Adams - C 0004	25	Adams
5	55001003000001	55001	Adams	00300	5500100300	Adams	T	Adams - T 0001	43	Adams
6	55001003000002	55001	Adams	00300	5500100300	Adams	T	Adams - T 0002	43	Adams
7	55001003000003	55001	Adams	00300	5500100300	Adams	T	Adams - T 0003	43	Adams
8	55001073000001	55001	Adams	07300	5500107300	Big Flats	T	Big Flats - T 0001	43	Big F

Figure 3.6 Text Import Wizard.

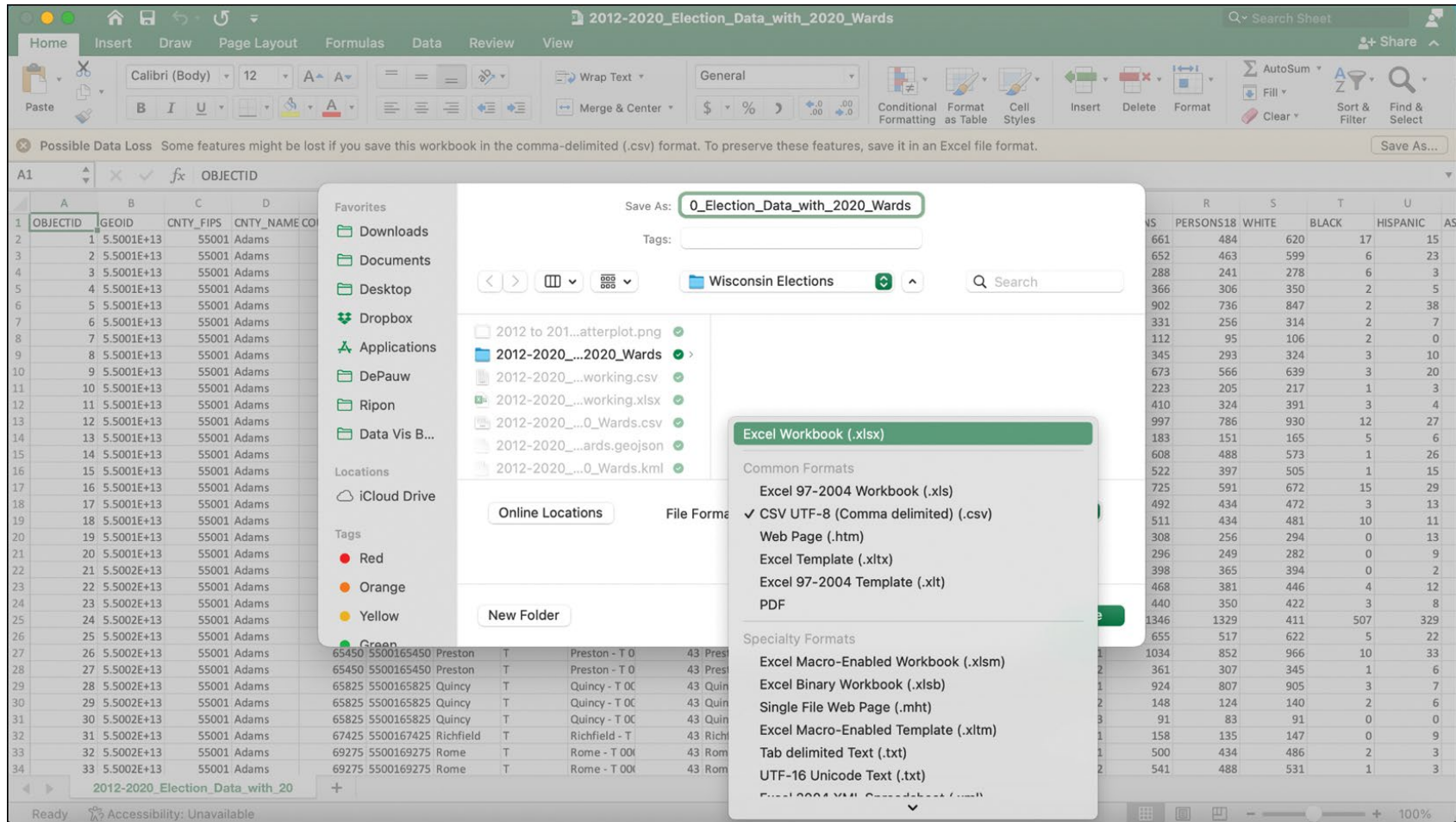


Figure 3.7 Saving a copy of the .csv file as an Excel .xlsx file.

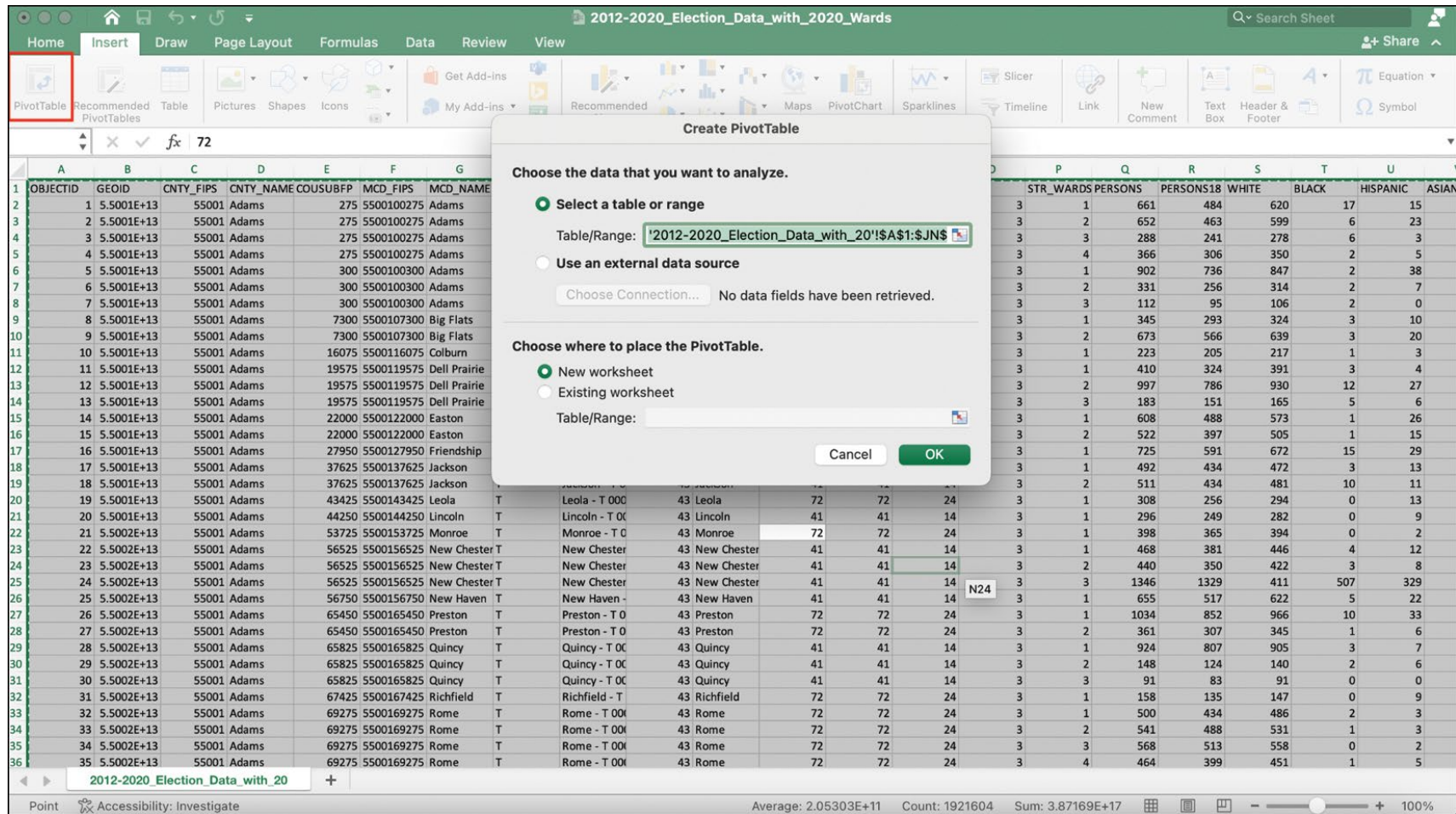


Figure 3.8 Creating a pivot table.

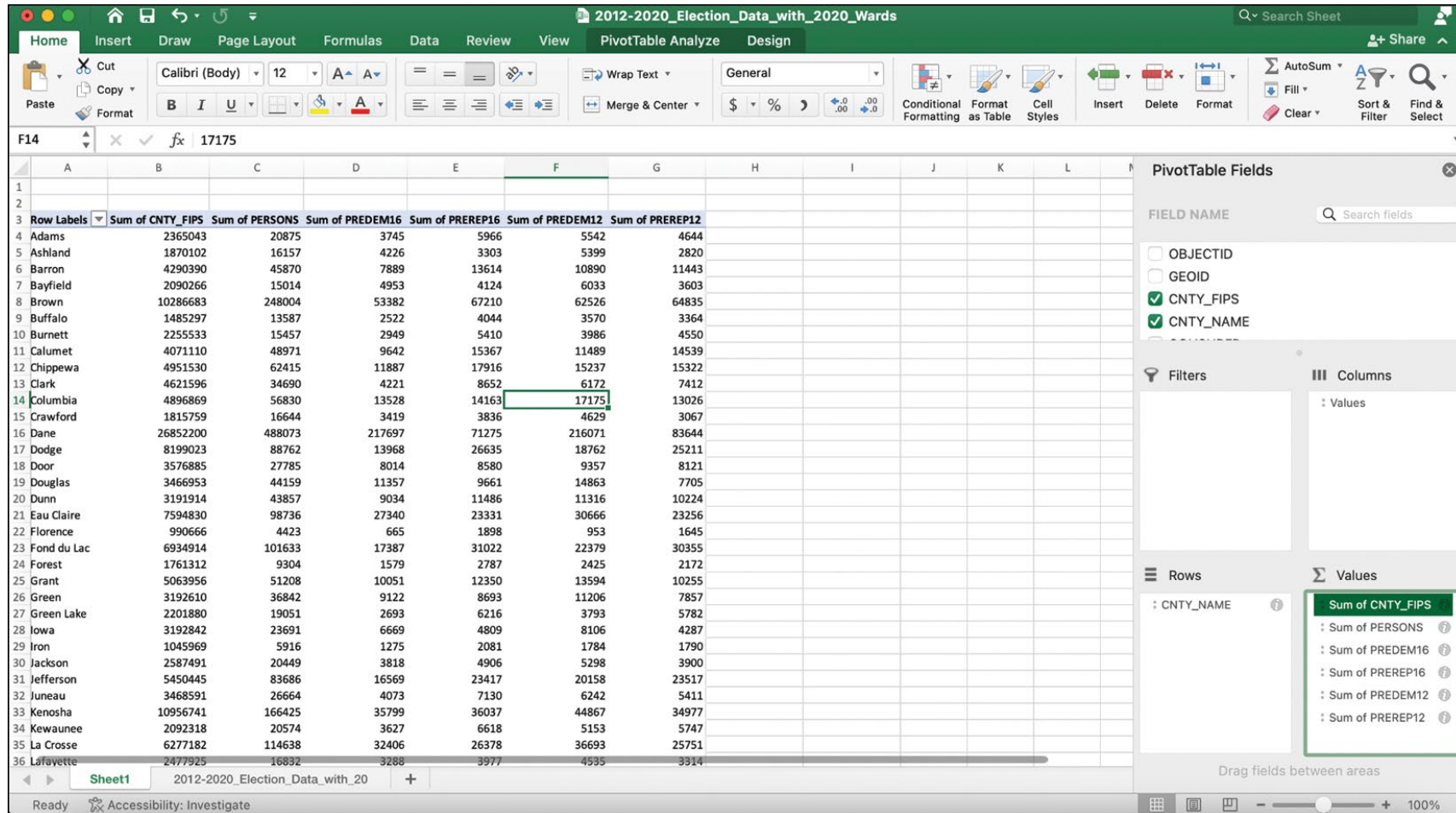


Figure 3.9 Editing the pivot table by county.

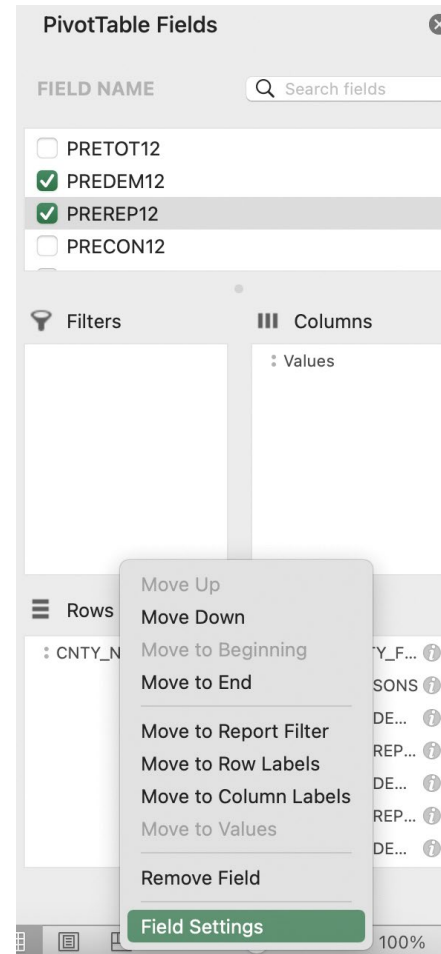


Figure 3.10 Changing the aggregation method.

PivotTable Field

Source field: CNTY_FIPS

Field name:

Summarize by Show data as

- Sum
- Count
- Average
- Max
- Min**
- Product
- Count Numbers
- StdDev

Figure 3.11 Summarize by Min.

	A	B	C	D	E	F	G
1							
2							
3	Row Labels	Sum of CNTY_FIPS	Sum of PERSONS	Sum of PREDEM16	Sum of PREREP16	Sum of PREDEM12	Sum of PREREP12
4	Adams	2365043	20875	3745	5966	5542	4644
5	Ashland	1870102	16157	4226	3303	5399	2820
6	Barron	4290390	45870	7889	13614	10890	11443
7	Bayfield	2090266	15014	4953	4124	6033	3603
8	Brown	10286683	248004	53382	67210	62526	64835
9	Buffalo	1485297	13587	2522	4044	3570	3364
10	Burnett	2255533	15457	2949	5410	3986	4550
11	Calumet	4071110	48971	9642	15367	11489	14539
12	Chippewa	4951530	62415	11887	17916	15237	15322
13	Clark	4621596	34690	4221	8652	6172	7412
14	Columbia	4896869	56830	13528	14163	17175	13026
15	Crawford	1815759	16644	3419	3836	4629	3067
16	Dane	26852200	488073	217697	71275	216071	83644
17	Dodge	8199023	88762	13968	26635	18762	25211
18	Door	3576885	27785	8014	8580	9357	8121
19	Douglas	3466953	44159	11357	9661	14863	7705
20	Dunn	3191914	43857	9034	11486	11316	10224
21	Eau Claire	7594830	98736	27340	23331	30666	23256
22	Florence	990666	4423	665	1898	953	1645
23	Fond du Lac	6934914	101633	17387	31022	22379	30355
24	Forest	1761312	9304	1579	2787	2425	2172
25	Grant	5063956	51208	10051	12350	13594	10255
26	Green	3192610	36842	9122	8693	11206	7857
27	Green Lake	2201880	19051	2693	6216	3793	5782
28	Iowa	3192842	23691	6669	4809	8106	4287
29	Iron	1045969	5916	1275	2081	1784	1790
30	Jackson	2587491	20449	3818	4906	5298	3900
31	Jefferson	5450445	83686	16569	23417	20158	23517
32	Juneau	3468591	26664	4073	7130	6242	5411
33	Kenosha	10956741	166425	35799	36037	44867	34977
34	Kewaunee	2092318	20574	3627	6618	5153	5747
35	La Crosse	6277182	114638	32406	26378	36693	25751

Figure 3.12 Modified pivot table.

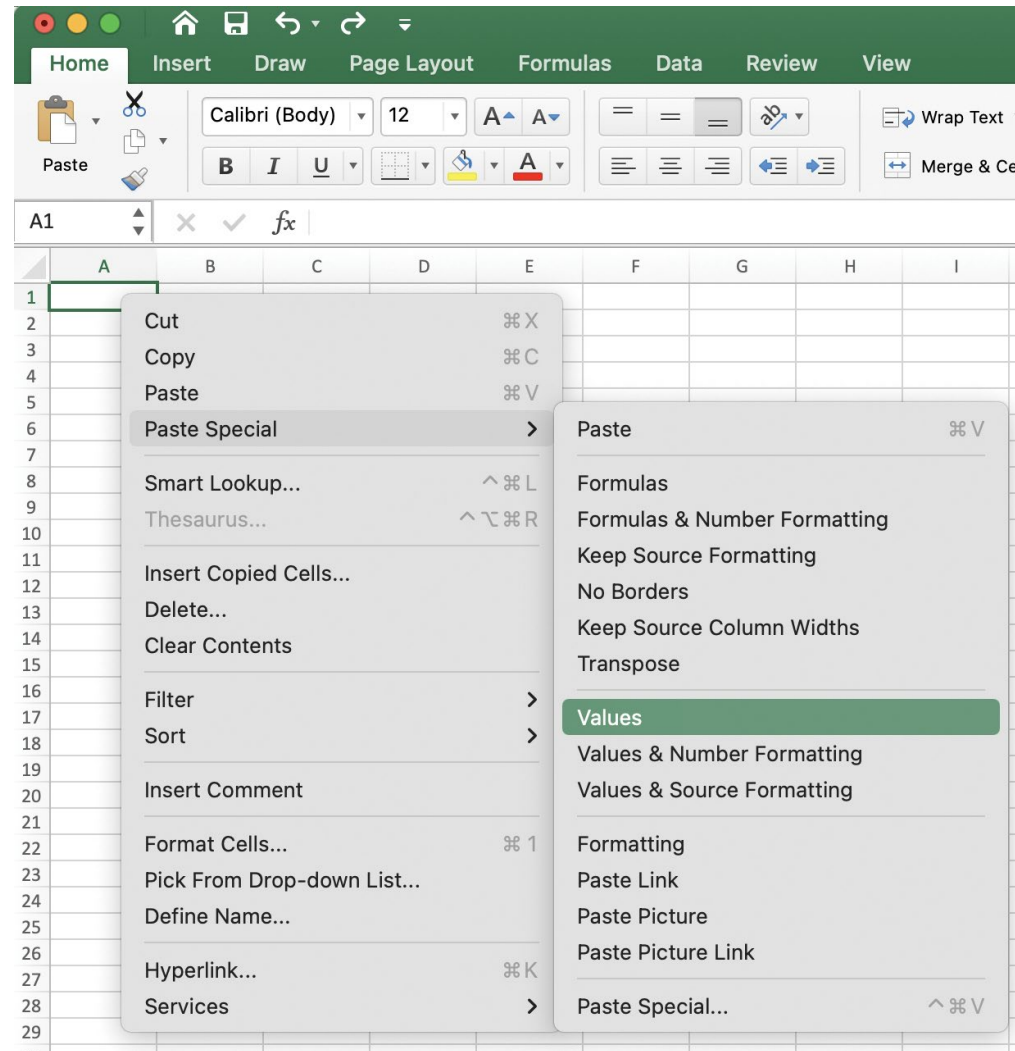
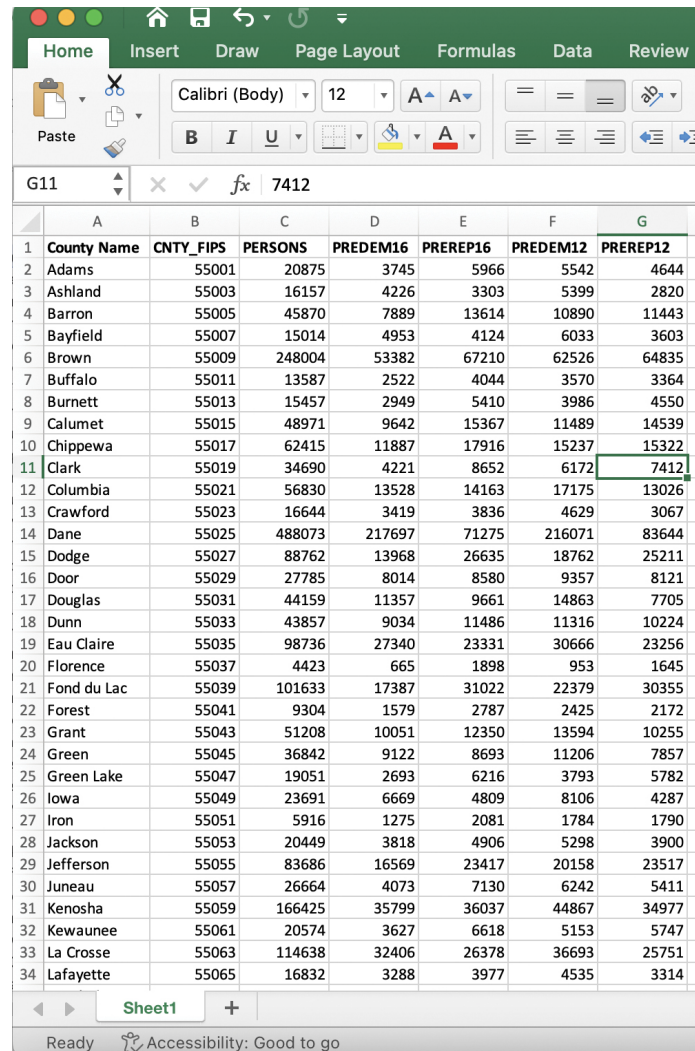


Figure 3.13 Pasting the pivot table values using Paste Values.



The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G
	County Name	CNTY_FIPS	PERSONS	PREDEM16	PREREP16	PREDEM12	PREREP12
1	Adams	55001	20875	3745	5966	5542	4644
2	Ashland	55003	16157	4226	3303	5399	2820
3	Barron	55005	45870	7889	13614	10890	11443
4	Bayfield	55007	15014	4953	4124	6033	3603
5	Brown	55009	248004	53382	67210	62526	64835
6	Buffalo	55011	13587	2522	4044	3570	3364
7	Burnett	55013	15457	2949	5410	3986	4550
8	Calumet	55015	48971	9642	15367	11489	14539
9	Chippewa	55017	62415	11887	17916	15237	15322
10	Clark	55019	34690	4221	8652	6172	7412
11	Columbia	55021	56830	13528	14163	17175	13026
12	Crawford	55023	16644	3419	3836	4629	3067
13	Dane	55025	488073	217697	71275	216071	83644
14	Dodge	55027	88762	13968	26635	18762	25211
15	Door	55029	27785	8014	8580	9357	8121
16	Douglas	55031	44159	11357	9661	14863	7705
17	Dunn	55033	43857	9034	11486	11316	10224
18	Eau Claire	55035	98736	27340	23331	30666	23256
19	Florence	55037	4423	665	1898	953	1645
20	Fond du Lac	55039	101633	17387	31022	22379	30355
21	Forest	55041	9304	1579	2787	2425	2172
22	Grant	55043	51208	10051	12350	13594	10255
23	Green	55045	36842	9122	8693	11206	7857
24	Green Lake	55047	19051	2693	6216	3793	5782
25	Iowa	55049	23691	6669	4809	8106	4287
26	Iron	55051	5916	1275	2081	1784	1790
27	Jackson	55053	20449	3818	4906	5298	3900
28	Jefferson	55055	83686	16569	23417	20158	23517
29	Juneau	55057	26664	4073	7130	6242	5411
30	Kenosha	55059	166425	35799	36037	44867	34977
31	Kewaunee	55061	20574	3627	6618	5153	5747
32	La Crosse	55063	114638	32406	26378	36693	25751
33	Lafayette	55065	16832	3288	3977	4535	3314

Figure 3.14 The new data file aggregated by county with population size and number of votes for each party's presidential candidate in 2012 and 2016.

SUM ✖ ✔ <i>fx</i> =F2+G2												
	A	B	C	D	E	F	G	H	I	J	K	L
	County Name	CNTY_FIPS	PERSONS	PREDEM16	PREREP16	PREDEM12	PREREP12	Major Party Votes 2012	Major Party Votes 2016	Prop Dem 2012	Prop Dem 2016	Prop Dem Change
1												
2	Adams	55001	20875	3745	5966	5542	4644	=F2+G2				
3	Ashland	55003	16157	4226	3303	5399	2820					
4	Barron	55005	45870	7889	13614	10890	11443					

Figure 3.16 Calculating total major party votes.

H2												
	A	B	C	D	E	F	G	H	I	J	K	L
1	County Name	CNTY_FIPS	PERSONS	PREDEM16	PREREP16	PREDEM12	PREREP12	Major Party Votes 2012	Major Party Votes 2016	Prop Dem 2012	Prop Dem 2016	Prop Dem Change
2	Adams	55001	20875	3745	5966	5542	144	10186				
3	Ashland	55003	16157	4226	3303	5399	2820	8219				
4	Barron	55005	45870	7889	13614	10890	11443	22333				
5	Bayfield	55007	15014	4953	4124	6033	3603	9636				
6	Brown	55009	248004	53382	67210	62526	64835	127361				
7	Buffalo	55011	13587	2522	4044	3570	3364	6934				
8	Burnett	55013	15457	2949	5410	3986	4550	8536				
9	Calumet	55015	48971	9642	15367	11489	14539	26028				
10	Chippewa	55017	62415	11887	17916	15237	15322	30559				
11	Clark	55019	34690	4221	8652	6172	7412	13584				
12	Columbia	55021	56830	13528	14163	17175	13026	30201				
13	Crawford	55023	16644	3419	3836	4629	3067	7696				
14	Dane	55025	488073	217697	71275	216071	83644	299715				
15	Dodge	55027	88762	13968	26635	18762	25211	43973				
16	Door	55029	27785	8014	8580	9357	8121	17478				
17	Douglas	55031	44159	11357	9661	14863	7705	22568				
18	Dunn	55033	43857	9034	11486	11316	10224	21540				
19	Eau Claire	55035	98736	27340	23331	30666	23256	53922				
20	Florence	55037	4423	665	1898	953	1645	2598				
21	Fond du Lac	55039	101633	17387	31022	22379	30355	52734				
22	Forest	55041	9304	1579	2787	2425	2172	4597				
23	Grant	55043	51208	10051	12350	13594	10255	23849				
24	Green	55045	36842	9122	8693	11206	7857	19063				
25	Green Lake	55047	19051	2693	6216	3793	5782	9575				
26	Iowa	55049	23691	6669	4809	8106	4287	12393				
27	Iron	55051	5916	1275	2081	1784	1790	3574				
28	Jackson	55053	20449	3818	4906	5298	3900	9198				
29	Jefferson	55055	83686	16569	23417	20158	23517	43675				
30	Juneau	55057	26664	4073	7130	6242	5411	11653				
31	Kenosha	55059	166425	35799	36037	44867	34977	79844				
32	Kewaunee	55061	20574	3627	6618	5153	5747	10900				
33	La Crosse	55063	114638	32406	26378	36693	25751	62444				

Sheet1

Ready
Accessibility: Good to go
Average: 829

Figure 3.17 Filling the formula down the column.

I2												
	A	B	C	D	E	F	G	H	I	J	K	L
1	County Name	CNTY_FIPS	PERSONS	PREDEM16	PREREP16	PREDEM12	PREREP12	Major Party Votes 2012	Major Party Votes 2016	Prop Dem 2012	Prop Dem 2016	Prop Dem Change
2	Adams	55001	20875	3745	5966	5542	4644	10186	9711	0.54408011	0.38564514	-0.158435
3	Ashland	55003	16157	4226	3303	5399	2820	8219				

Figure 3.18 Adding remaining formulas.

L1	Prop Dem Change											
	A	B	C	D	E	F	G	H	I	J	K	L
								Major Party Votes 2012	Major Party Votes 2016	Prop Dem 2012	Prop Dem 2016	Prop Dem Change
1	County Name	CNTY_FIPS	PERSONS	PREDEM16	PREREP16	PREDEM12	PREREP12					
2	Adams	55001	20875	3745	5966	5542	4644	36	9711	0.54408011	0.38564514	-0.158435
3	Ashland	55003	16157	4226	3303	5399	2820	8219	7529	0.65689257	0.56129632	-0.0955962
4	Barron	55005	45870	7889	13614	10890	11443	22333	21503	0.48761922	0.36687904	-0.1207402
5	Bayfield	55007	15014	4953	4124	6033	3603	9636	9077	0.62608966	0.54566487	-0.0804248
6	Brown	55009	248004	53382	67210	62526	64835	127361	120592	0.49093522	0.44266618	-0.048269
7	Buffalo	55011	13587	2522	4044	3570	3364	6934	6566	0.51485434	0.38409991	-0.1307544
8	Burnett	55013	15457	2949	5410	3986	4550	8536	8359	0.46696345	0.3527934	-0.1141701
9	Calumet	55015	48971	9642	15367	11489	14539	26028	25009	0.44140925	0.38554121	-0.055868
10	Chippewa	55017	62415	11887	17916	15237	15322	30559	29803	0.49860925	0.39885246	-0.0997568
11	Clark	55019	34690	4221	8652	6172	7412	13584	12873	0.45435807	0.3278956	-0.1264625
12	Columbia	55021	56830	13528	14163	17175	13026	30201	27691	0.56868978	0.48853418	-0.0801556
13	Crawford	55023	16644	3419	3836	4629	3067	7696	7255	0.60148129	0.4712612	-0.1302201
14	Dane	55025	488073	217697	71275	216071	83644	299715	288972	0.72092154	0.75334981	0.03242826
15	Dodge	55027	88762	13968	26635	18762	25211	43973	40603	0.42667091	0.34401399	-0.0826569
16	Door	55029	27785	8014	8580	9357	8121	17478	16594	0.53535874	0.48294564	-0.0524131
17	Douglas	55031	44159	11357	9661	14863	7705	22568	21018	0.65858738	0.54034637	-0.118241
18	Dunn	55033	43857	9034	11486	11316	10224	21540	20520	0.52534819	0.44025341	-0.0850948
19	Eau Claire	55035	98736	27340	23331	30666	23256	53922	50671	0.56871036	0.53955912	-0.0291512
20	Florence	55037	4423	665	1898	953	1645	2598	2563	0.36682063	0.25946157	-0.1073591
21	Fond du Lac	55039	101633	17387	31022	22379	30355	52734	48409	0.42437517	0.35916875	-0.0652064
22	Forest	55041	9304	1579	2787	2425	2172	4597	4366	0.52751795	0.36165827	-0.1658597
23	Grant	55043	51208	10051	12350	13594	10255	23849	22401	0.57000294	0.44868533	-0.1213176
24	Green	55045	36842	9122	8693	11206	7857	19063	17815	0.58784032	0.51204042	-0.0757999
25	Green Lake	55047	19051	2693	6216	3793	5782	9575	8909	0.39613577	0.30227859	-0.0938572
26	Iowa	55049	23691	6669	4809	8106	4287	12393	11478	0.65407892	0.58102457	-0.0730543
27	Iron	55051	5916	1275	2081	1784	1790	3574	3356	0.4991606	0.37991657	-0.119244
28	Jackson	55053	20449	3818	4906	5298	3900	9198	8724	0.57599478	0.43764328	-0.1383515
29	Jefferson	55055	83686	16569	23417	20158	23517	43675	39986	0.46154551	0.41437003	-0.0471755
30	Juneau	55057	26664	4073	7130	6242	5411	11653	11203	0.53565605	0.36356333	-0.1720927
31	Kenosha	55059	166425	35799	36037	44867	34977	79844	71836	0.56193327	0.49834345	-0.0635898
32	Kewaunee	55061	20574	3627	6618	5153	5747	10900	10245	0.47275229	0.35402635	-0.1187259
33	La Crosse	55063	114638	32406	26378	36693	25751	62444	58784	0.5876145	0.55127246	-0.036342

Sheet1

+

Ready Accessibility: Good to go

Figure 3.19 Completed data file with new calculations.

B2 ✕ ✓ fx =A2 & " County, Wisconsin"				
	A	B	C	D
1	County Name	County	CNTY_FIPS	PERSONS
2	Adams	Adams County, Wisconsin	55001	20875
3	Ashland	Ashland County, Wisconsin	55003	16157
4	Barron	Barron County, Wisconsin	55005	45870
5	Bayfield	Bayfield County, Wisconsin	55007	15014
6	Brown	Brown County, Wisconsin	55009	248004
7	Buffalo	Buffalo County, Wisconsin	55011	13587
8	Burnett	Burnett County, Wisconsin	55013	15457
9	Calumet	Calumet County, Wisconsin	55015	48971
10	Chippewa	Chippewa County, Wisconsin	55017	62415
11	Clark	Clark County, Wisconsin	55019	34690

Figure 3.20 Pasting “ County, Wisconsin” onto county names.

Home

Insert

Draw

Page Layout

Formulas

Data

Review

Get External Data

Refresh All

Queries & Connections

Properties

Edit Links

Currencies

Geography

Convert

"Lond

popula

Power

B2

×

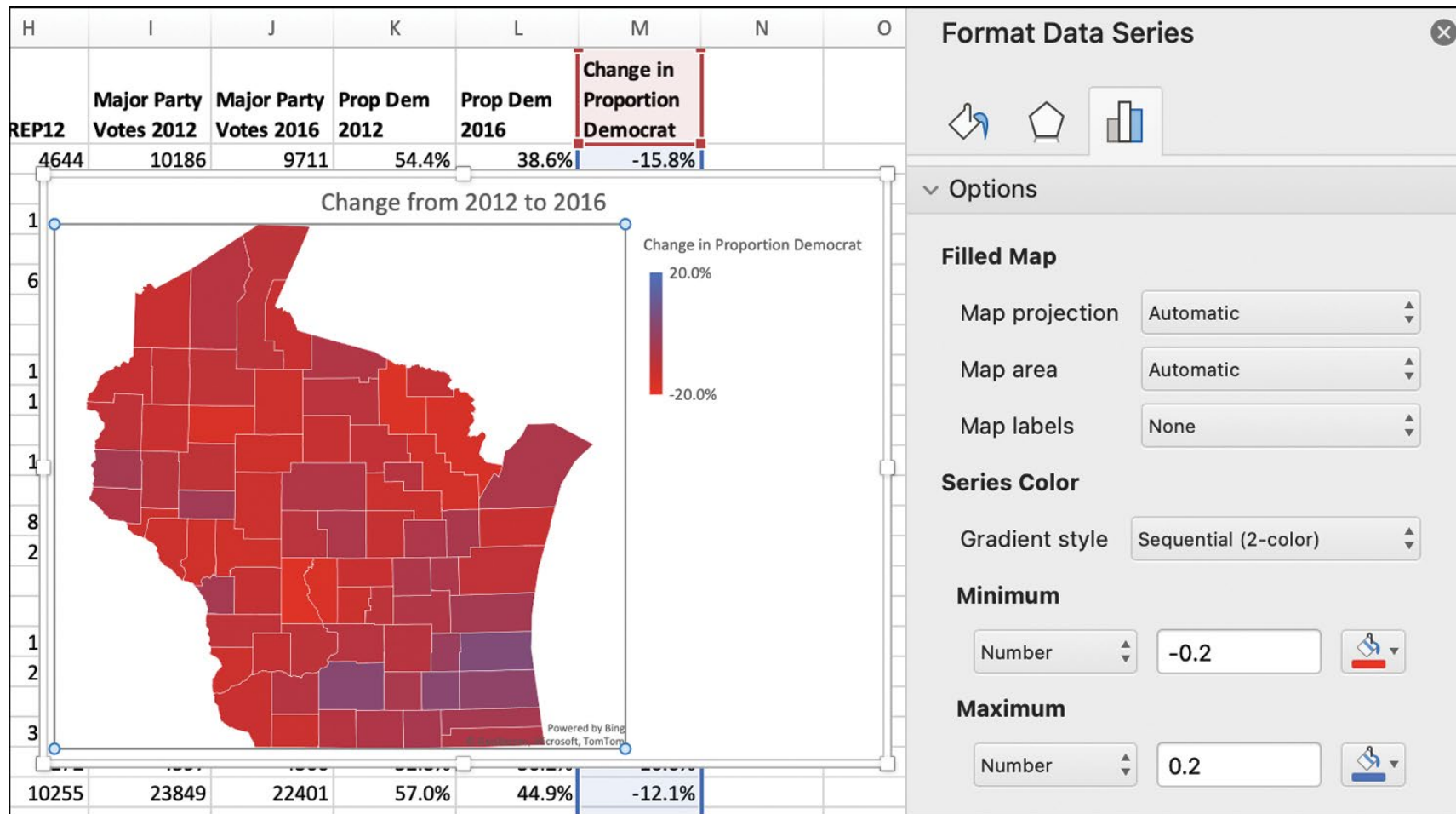
✓

fx

Adams County, Wisconsin

	A	B	C	D	E	
1	County Name	County	FIPS	PERSONS	PREDEM16	PREREP
2	Adams	Adams County, Wisconsin	55001	20875	3745	5
3	Ashland	Ashland County, Wisconsin	55003	16157	4226	3
4	Barron	Barron County, Wisconsin	55005	45870	7889	13
5	Bayfield	Bayfield County, Wisconsin	55007	15014	4953	4
6	Brown	Brown County, Wisconsin	55009	248004	53382	67
7	Buffalo	Buffalo County, Wisconsin	55011	13587	2522	4
8	Burnett	Burnett County, Wisconsin	55013	15457	2949	5
9	Calumet	Calumet County, Wisconsin	55015	48971	9642	15

Figure 3.21 Column converted to Geography type after pasting these names as values.



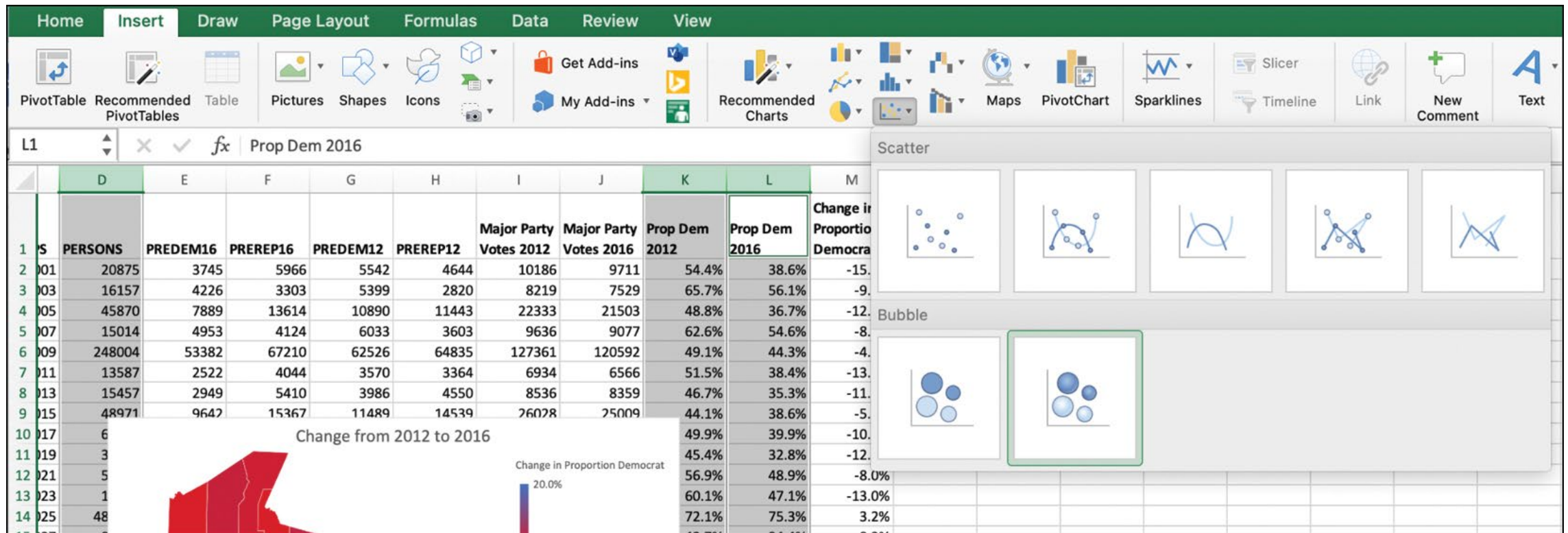


Figure 3.23 Creating a bubble chart.

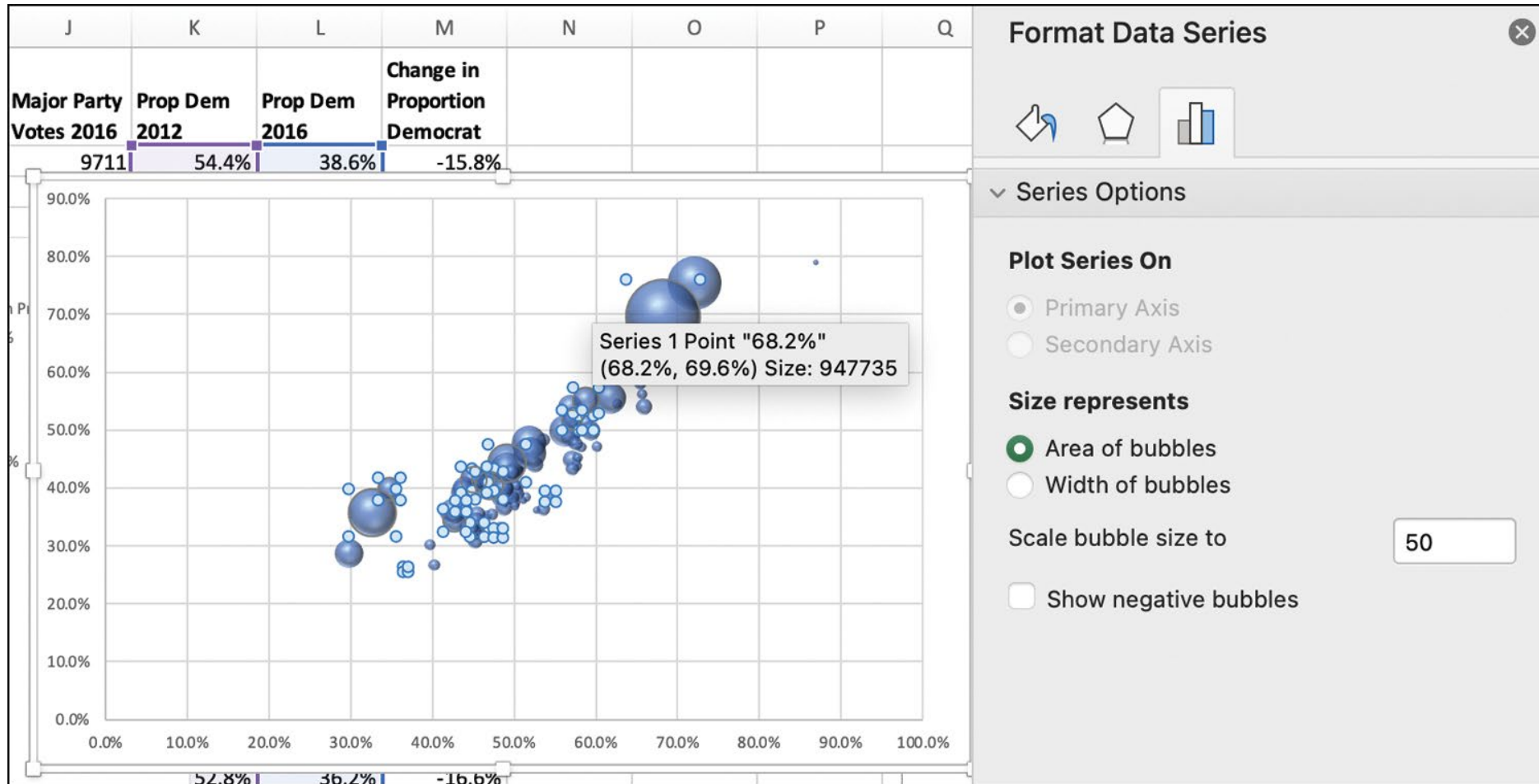


Figure 3.25 Sizing by area.

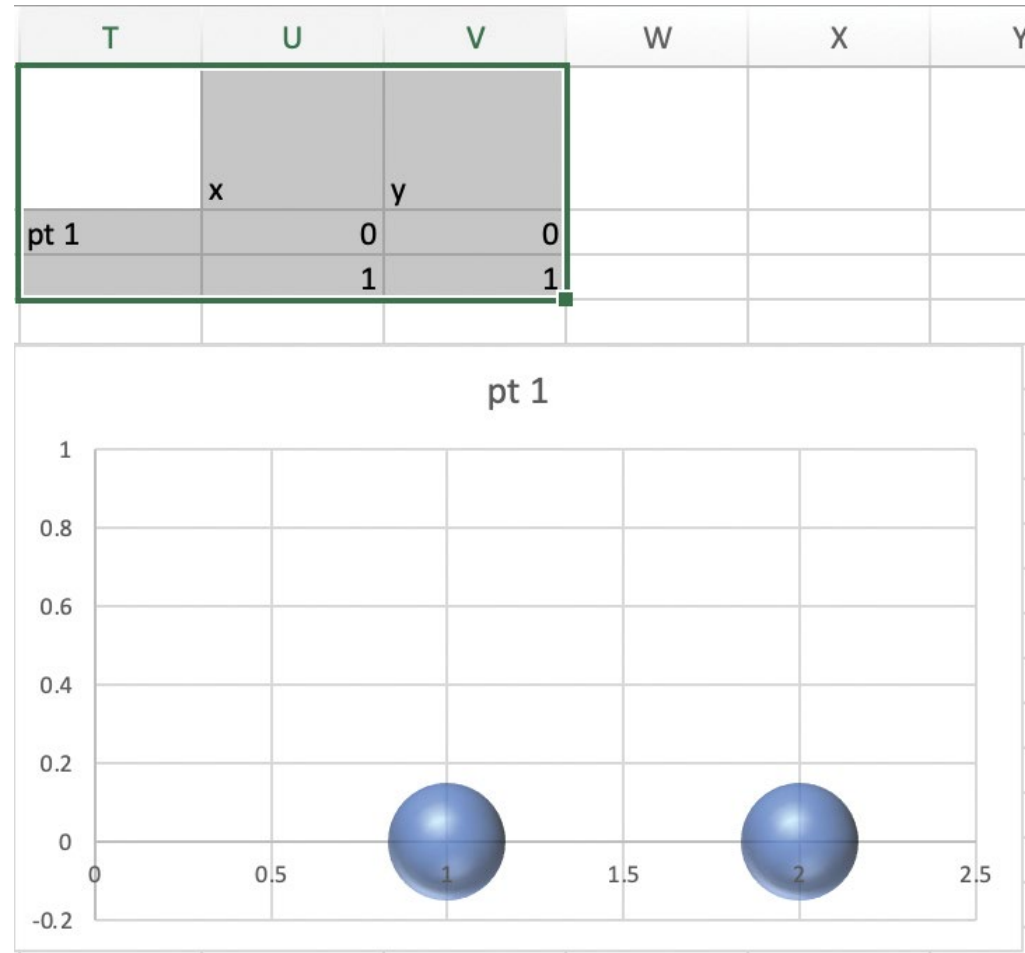


Figure 3.26 Initial auxiliary bubble chart.

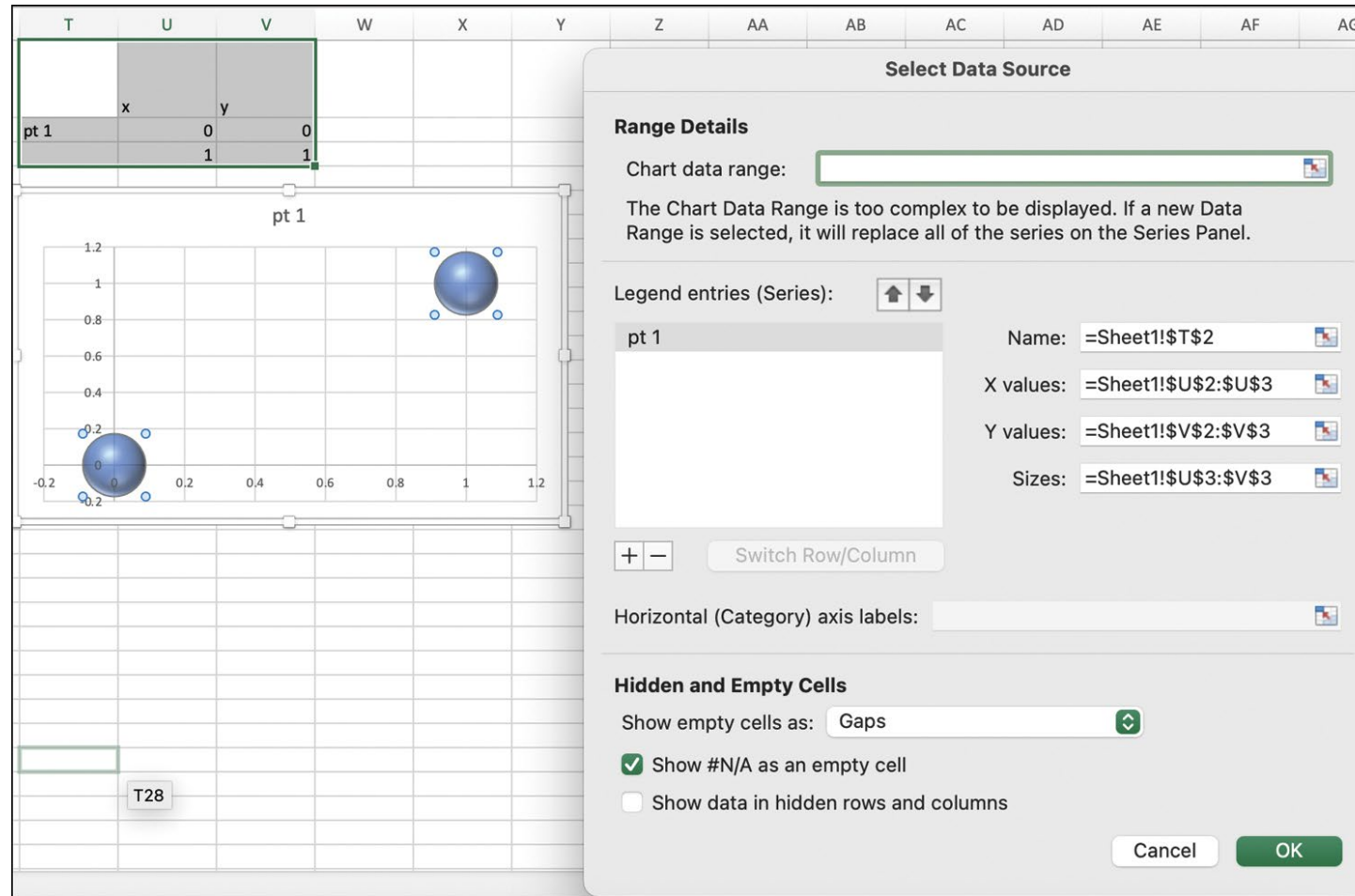


Figure 3.27 Assigning columns to bubble attributes in the auxiliary chart.

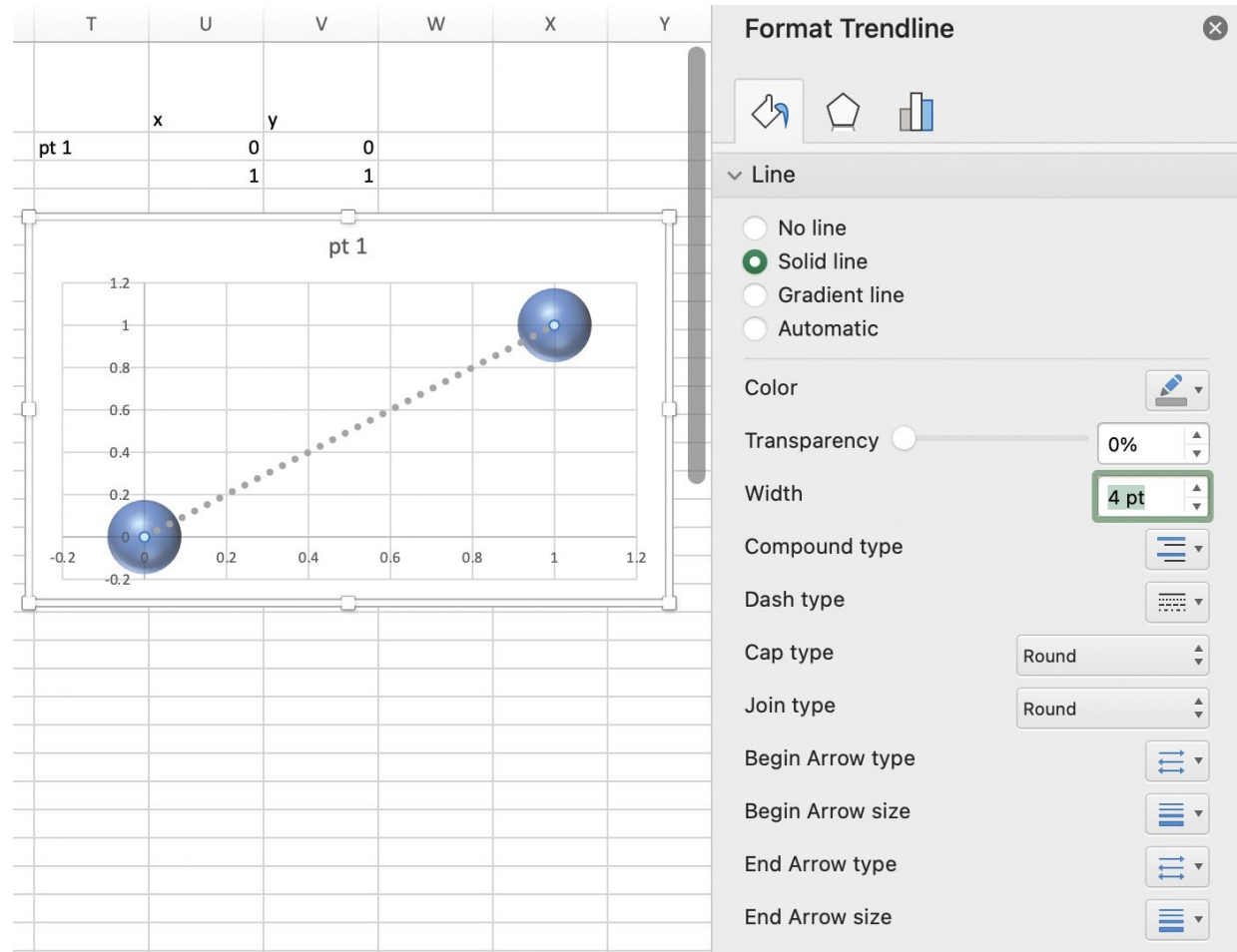


Figure 3.28 Formatting the trendline.

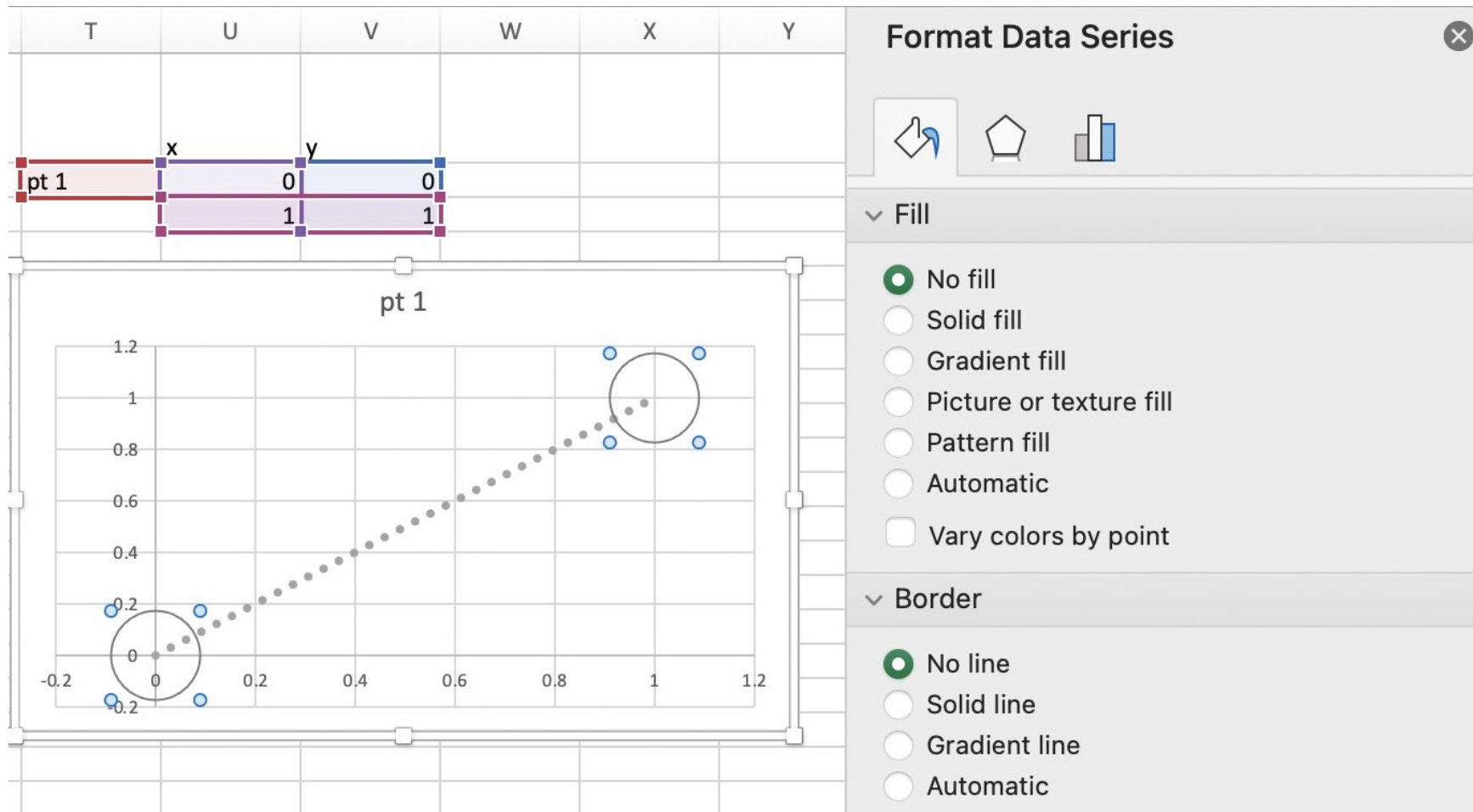


Figure 3.29 Hiding points.

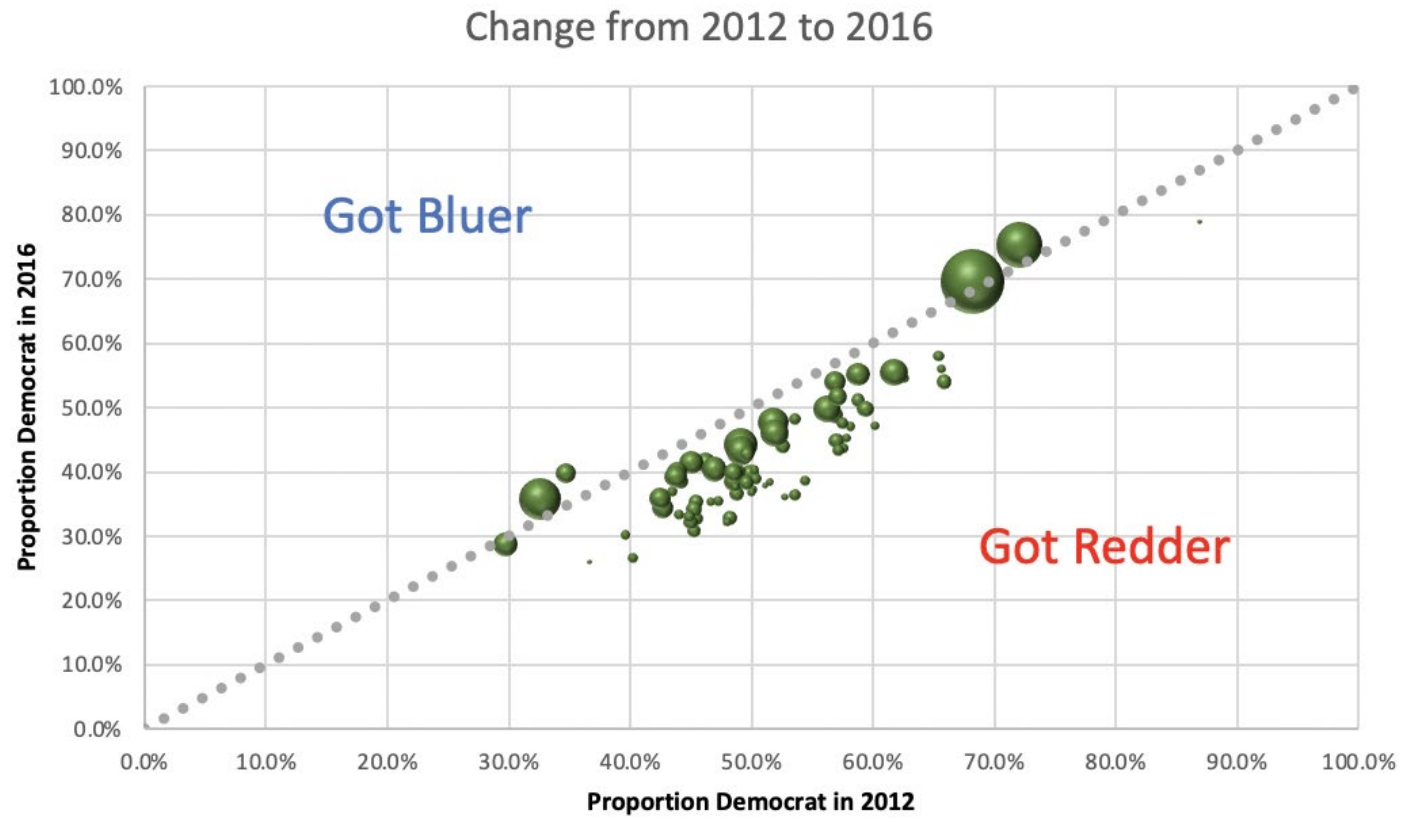


Figure 3.30 Final bubble chart.

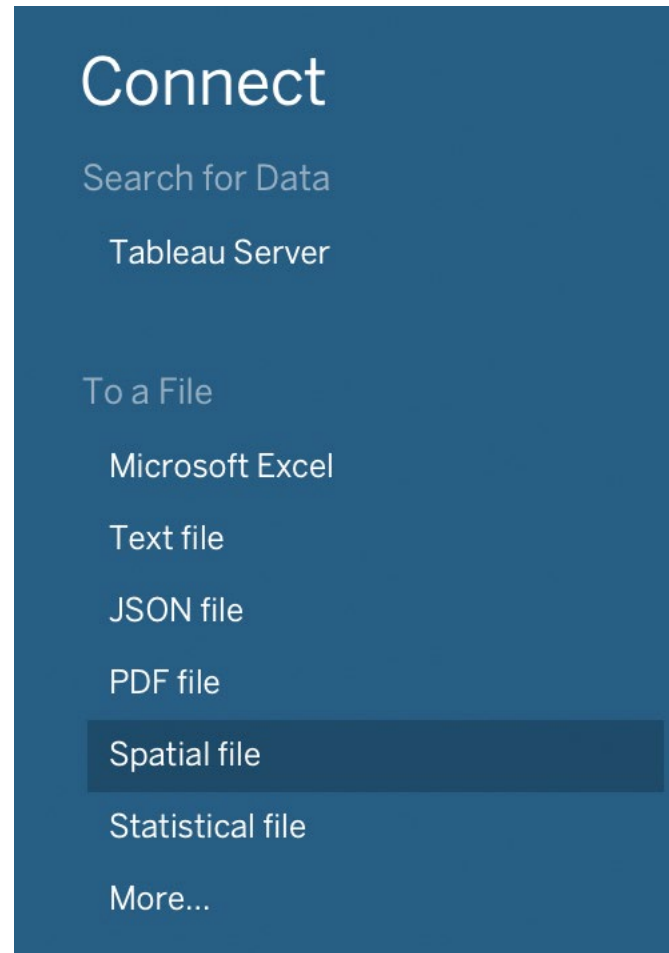


Figure 3.31 Connecting to a spatial file in Tableau.

< > ☰ ▼ ☰ ▼		📁 2012-2020_Election_Dat... ⬆ ⬇ ⬆		🔍 Search	
Name		Date Modified	Size	Kind	
📄	2012-2020_Electio...h_2020_Wards.dbf	✓ Jun 28, 2023 at 3:30 PM	5.9 MB	Document	
📄	2012-2020_Electio...h_2020_Wards.shp	✓ Jun 28, 2023 at 3:30 PM	36.1 MB	Document	
📄	2012-2020_Electio...h_2020_Wards.shx	✓ Jun 28, 2023 at 3:30 PM	57 KB	Document	
📄	2012-2020_Electio...h_2020_Wards.cpg	✓ Jun 28, 2023 at 3:29 PM	5 bytes	Document	
📄	2012-2020_Electio...ith_2020_Wards.prj	✓ Jun 28, 2023 at 3:29 PM	475 bytes	Document	

Figure 3.32 Choosing the .shp file.

Summary

2012-2020 Wisconsin Election Data with July 2020 Wards

Attribute Field Definitions

Figure 3.33 Finding the field definition key.

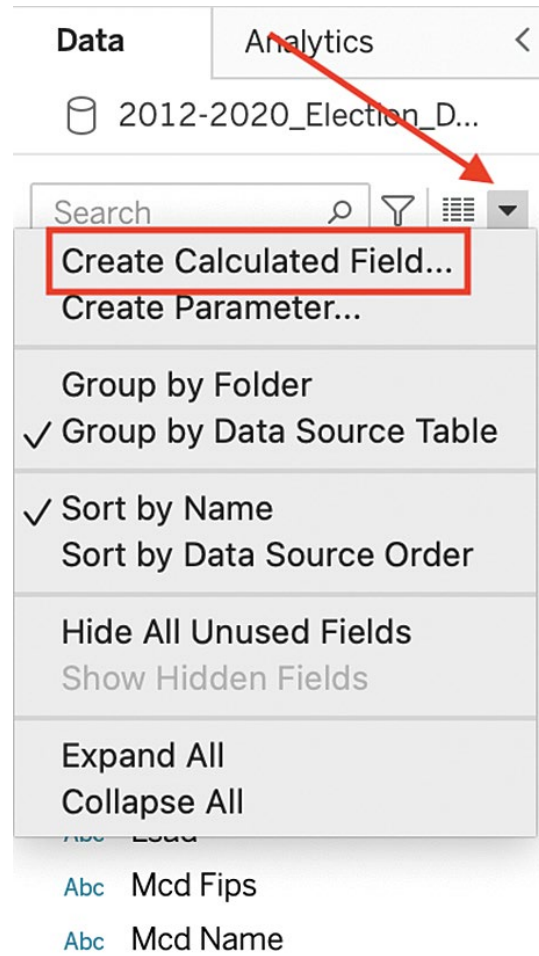


Figure 3.34 Creating a calculated field.



Figure 3.35 Total major party votes calculation.

Prop Dem 2012

×

$$[\text{Predem12}] / [\text{Major Party Votes 2012}]$$

Figure 3.36 Calculating the proportion of votes that went to the Democratic candidate.

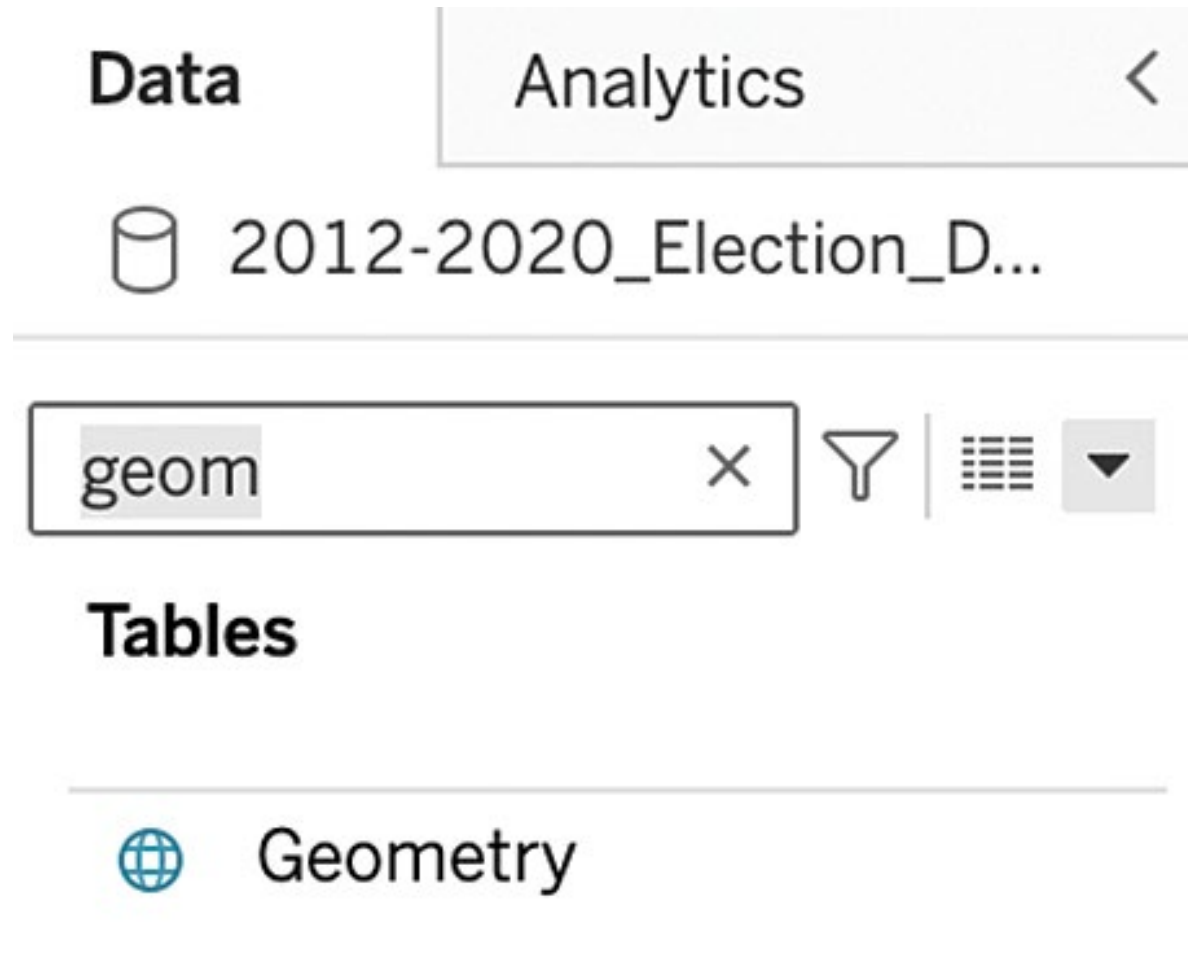


Figure 3.37 Finding the field with spatial information.

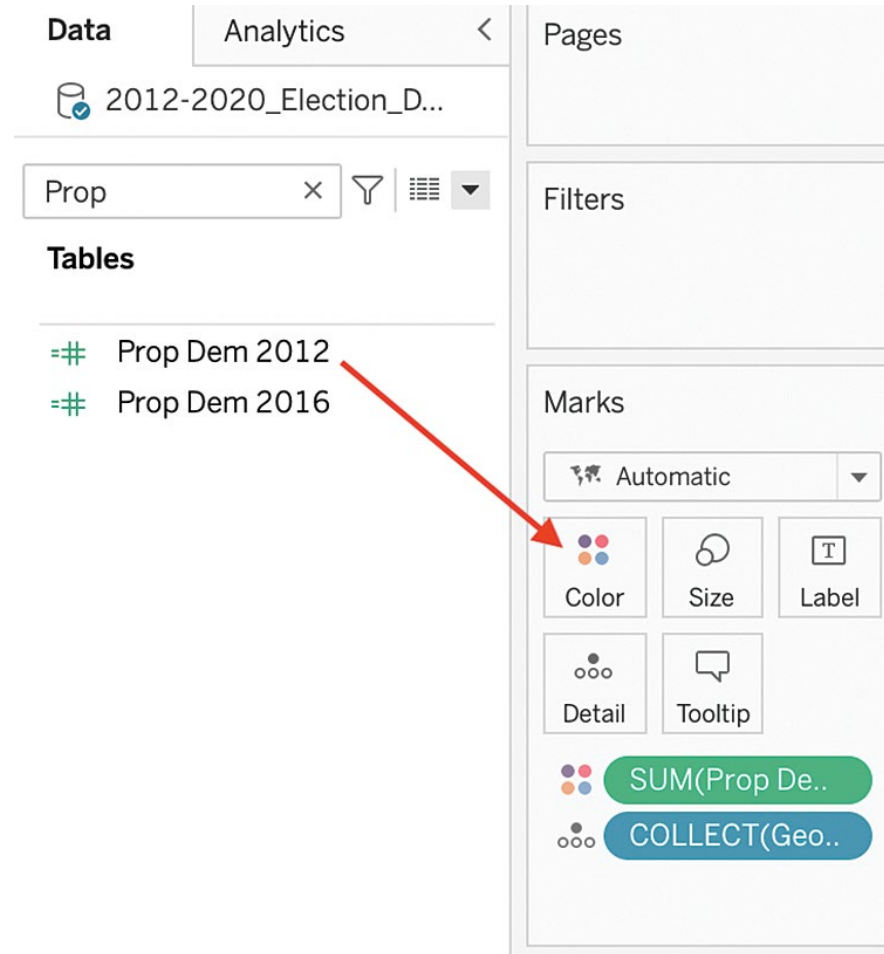


Figure 3.38 Coloring by the proportion of Democratic votes.

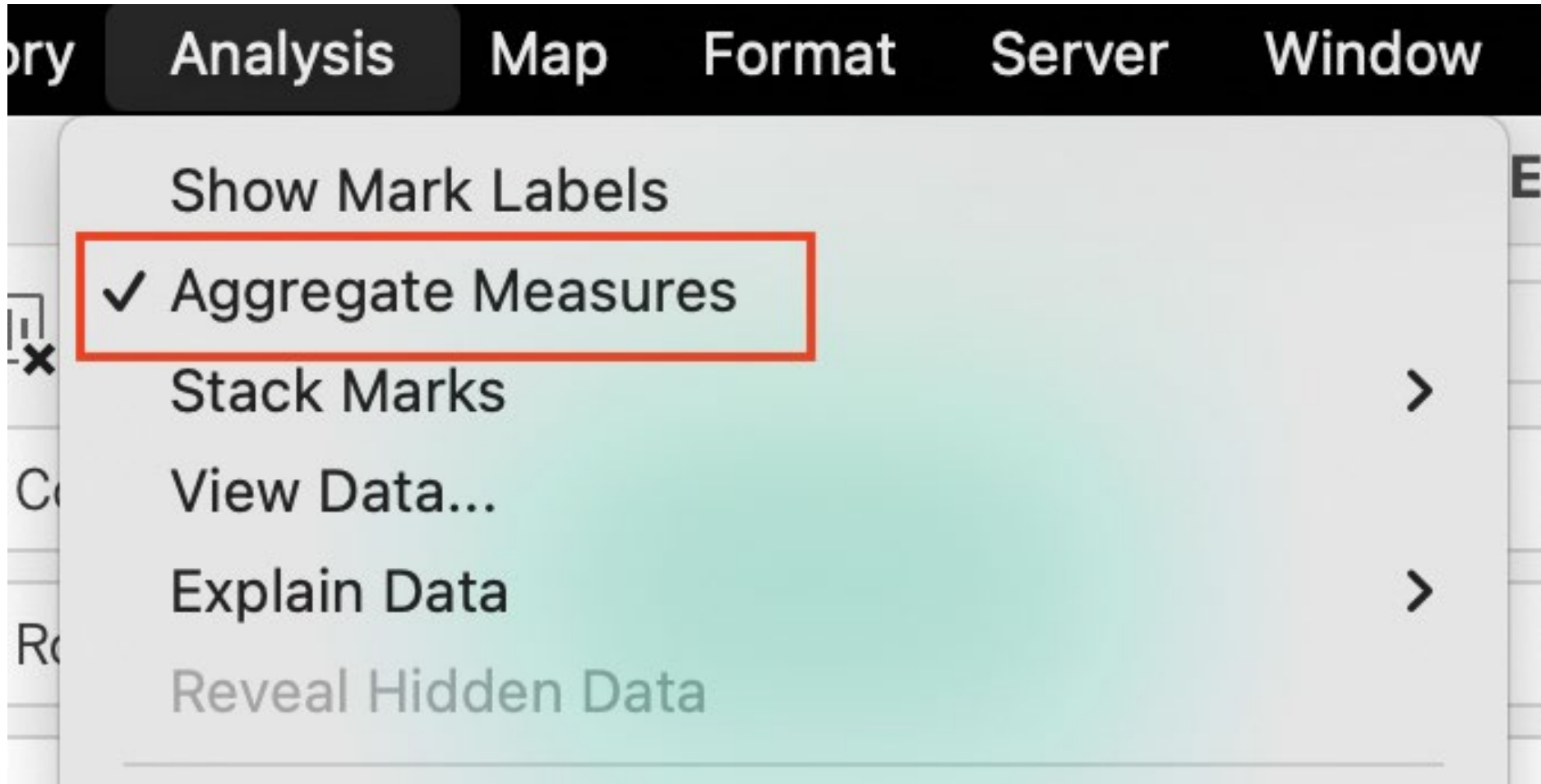


Figure 3.39 Disaggregating the data to the row level.

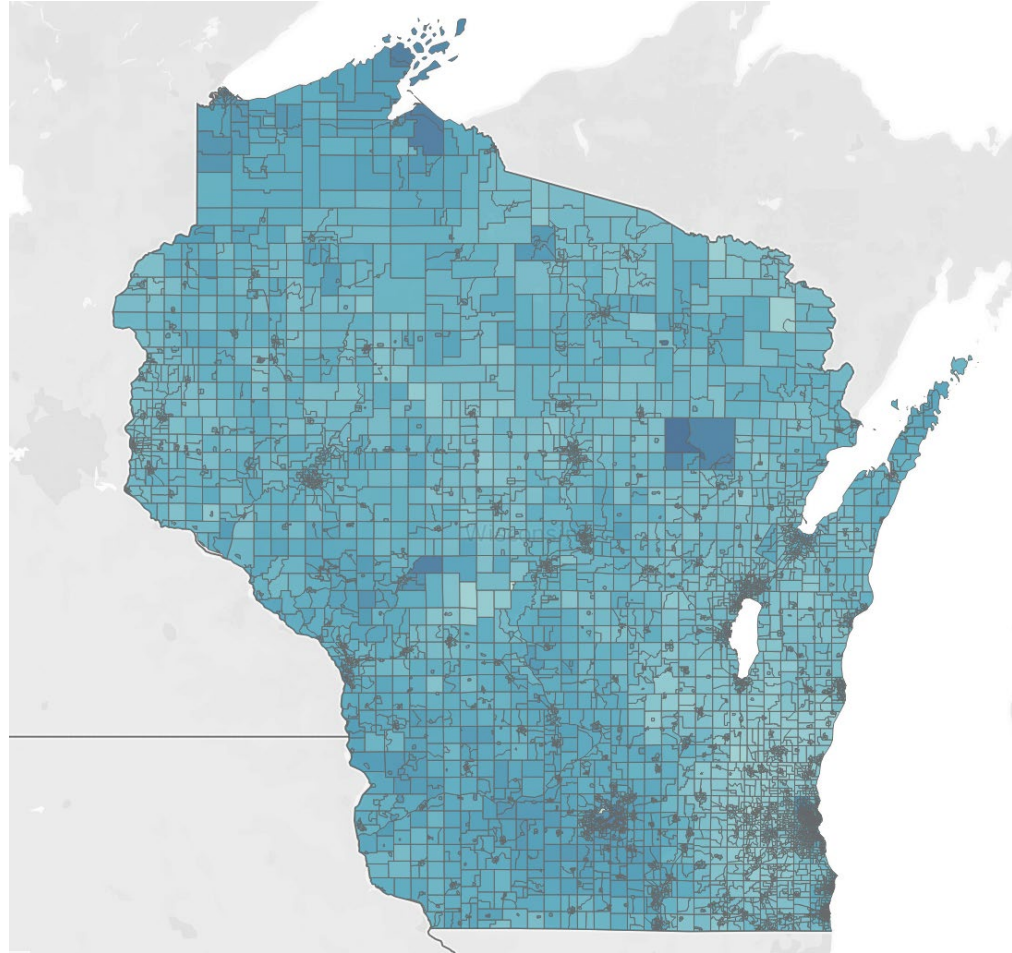


Figure 3.40 Initial map of Democratic vote share shaded by ward.

Edit Colors Prop Dem 12

Palette

Red-Blue Diverging

0%100%

Start

Automatic0

Center

Automatic0.5

End

Automatic1

☐ Stepped Color 5 steps

☐ Reverse

☐ Use full color range

☐ Include totals

Reset

Figure 3.41 Choosing color scale options.

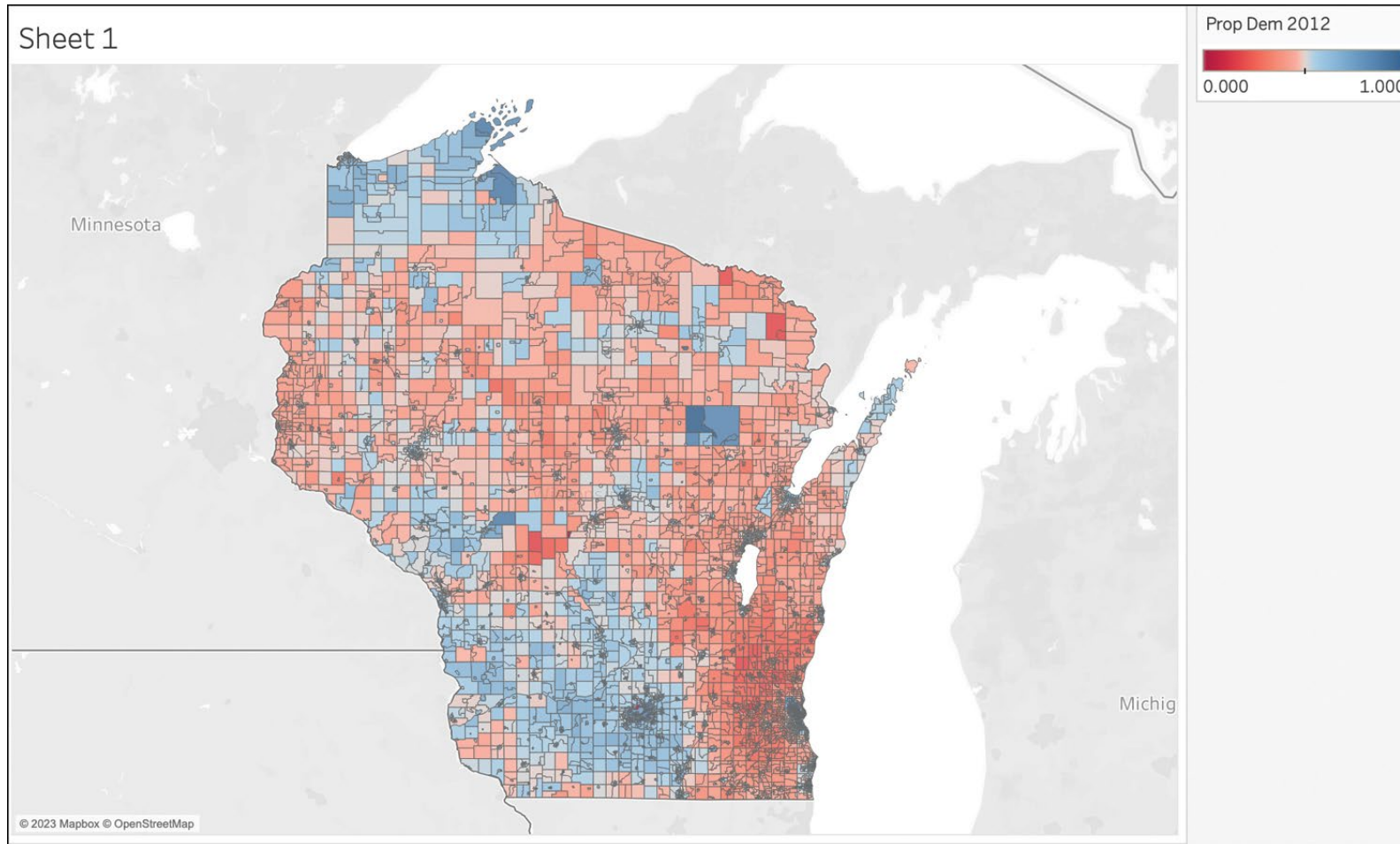


Figure 3.42 Democratic vote share shaded by ward.

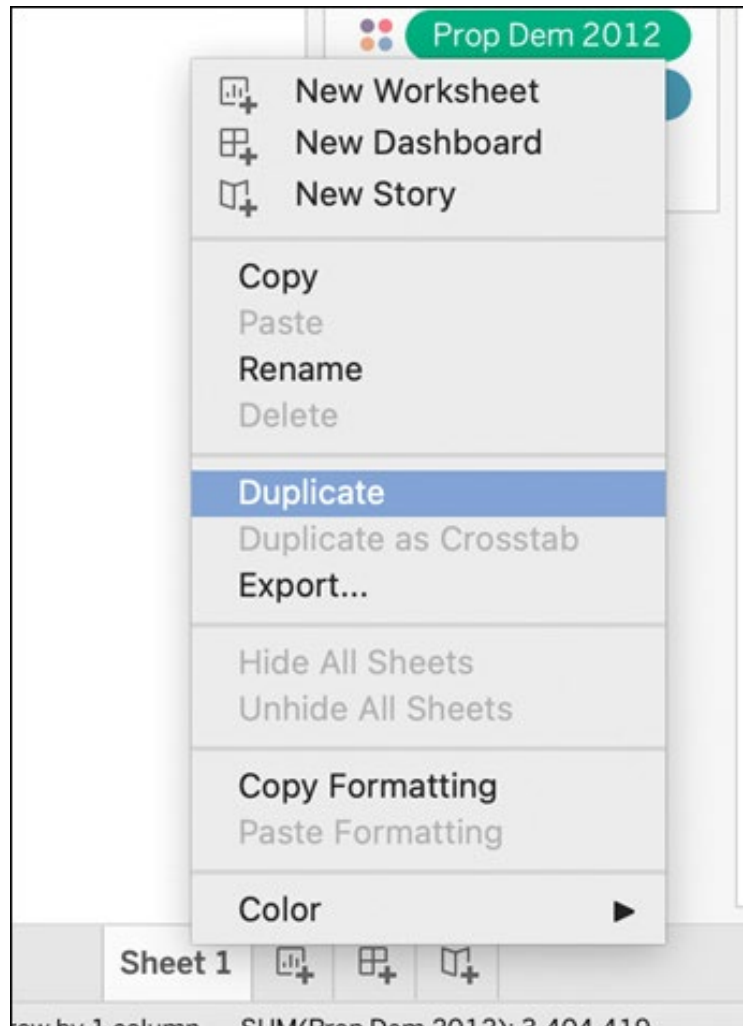


Figure 3.43 Duplicating a sheet.

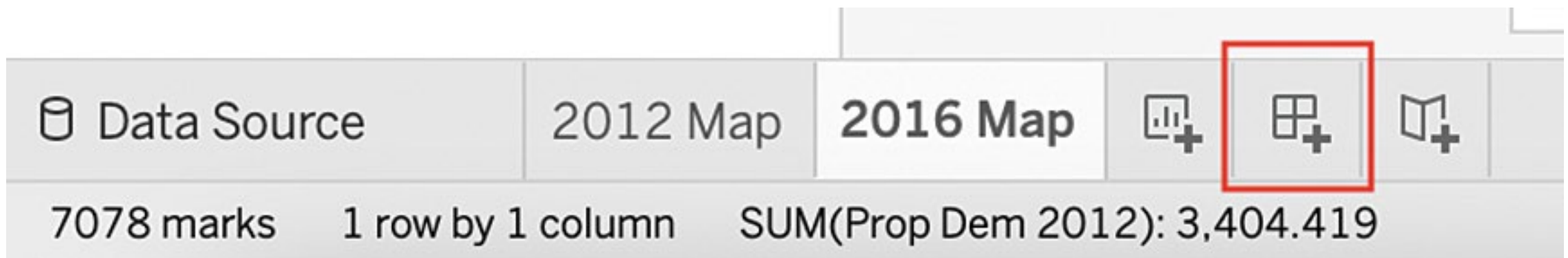


Figure 3.44 Creating a new dashboard.

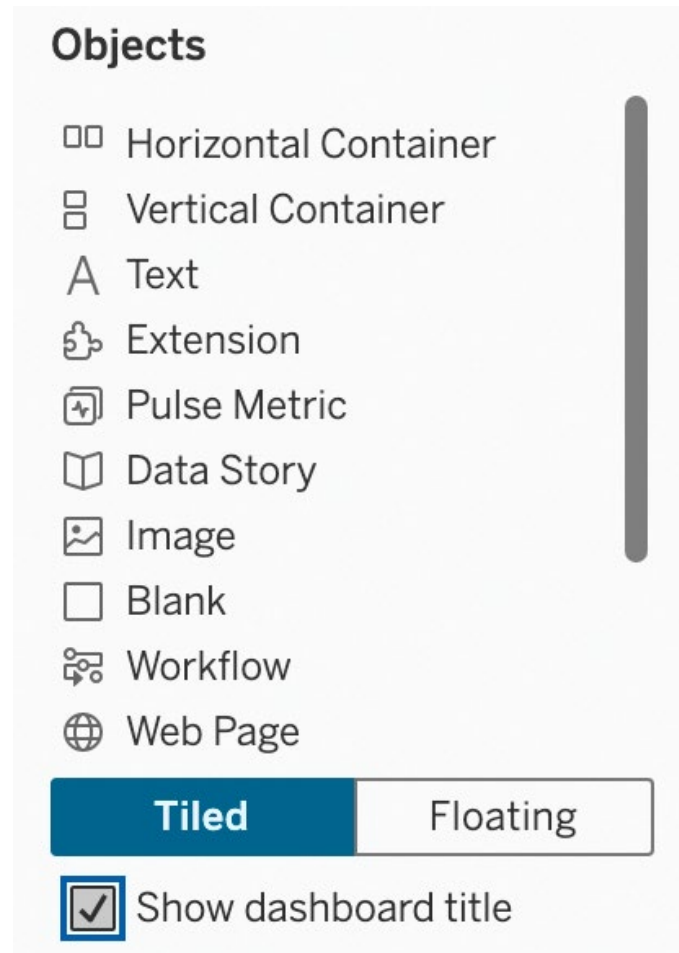


Figure 3.45 Showing a dashboard title.

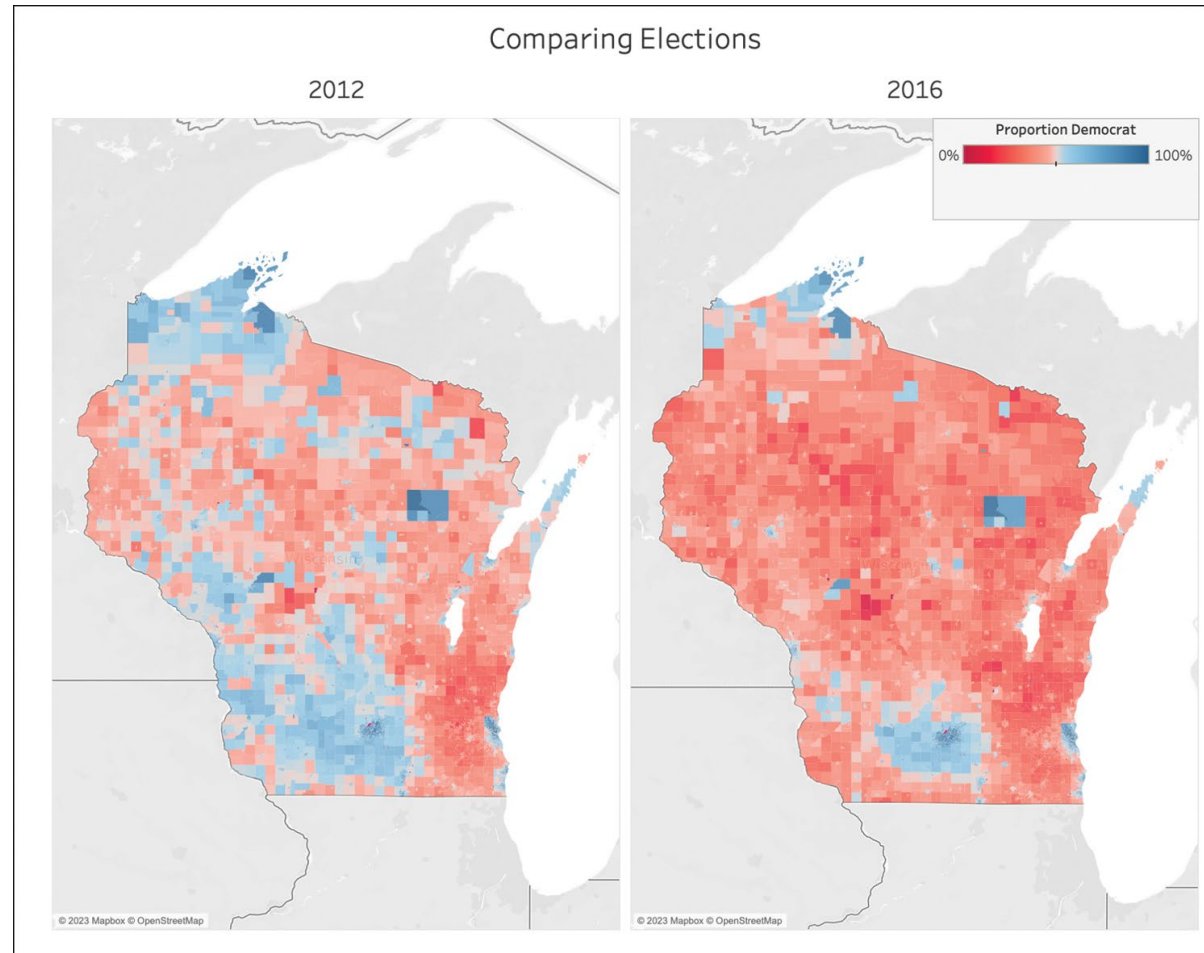


Figure 3.46 Side-by-side maps comparing the 2012 and 2016 Democratic election results.

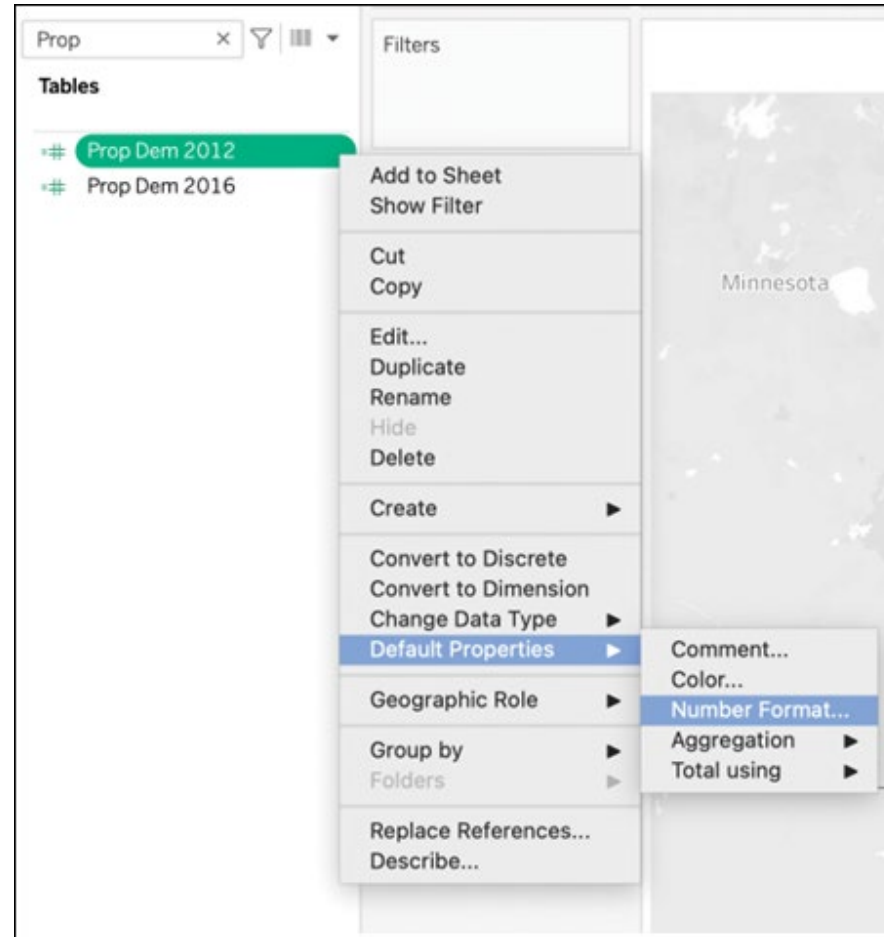


Figure 3.47 Changing the number format.

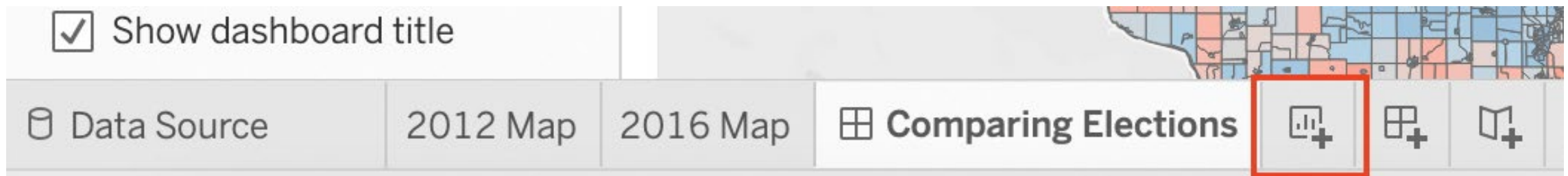


Figure 3.48 New worksheet icon.

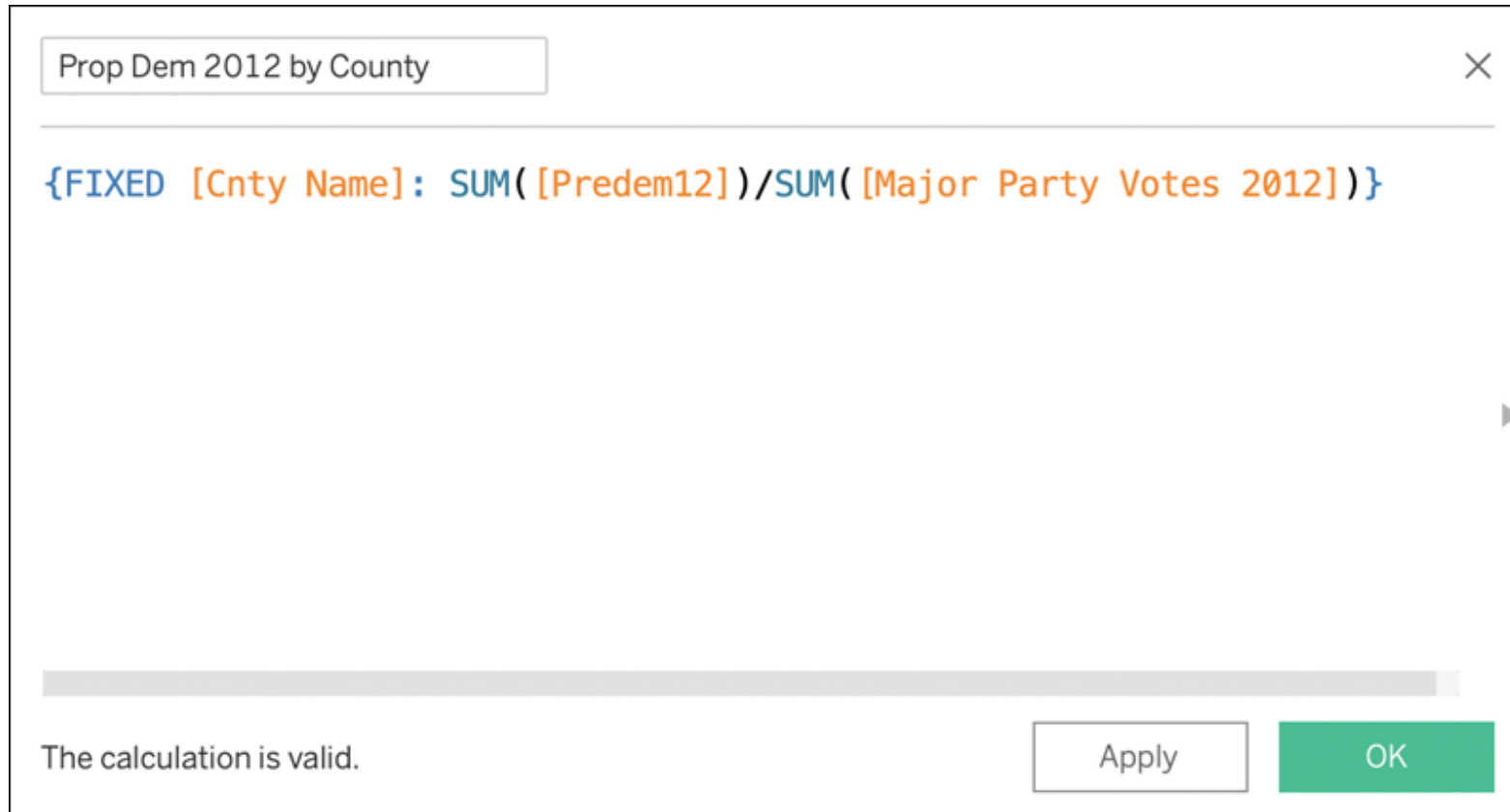


Figure 3.49 Level-of-detail calculation for the Democratic vote share by county.

Prop Dem Change

$$100 * ([\text{Prop Dem 2016 by County}] - [\text{Prop Dem 2012 by County}])$$

Figure 3.50 Calculating the change in the Democratic vote share from 2012 to 2016.

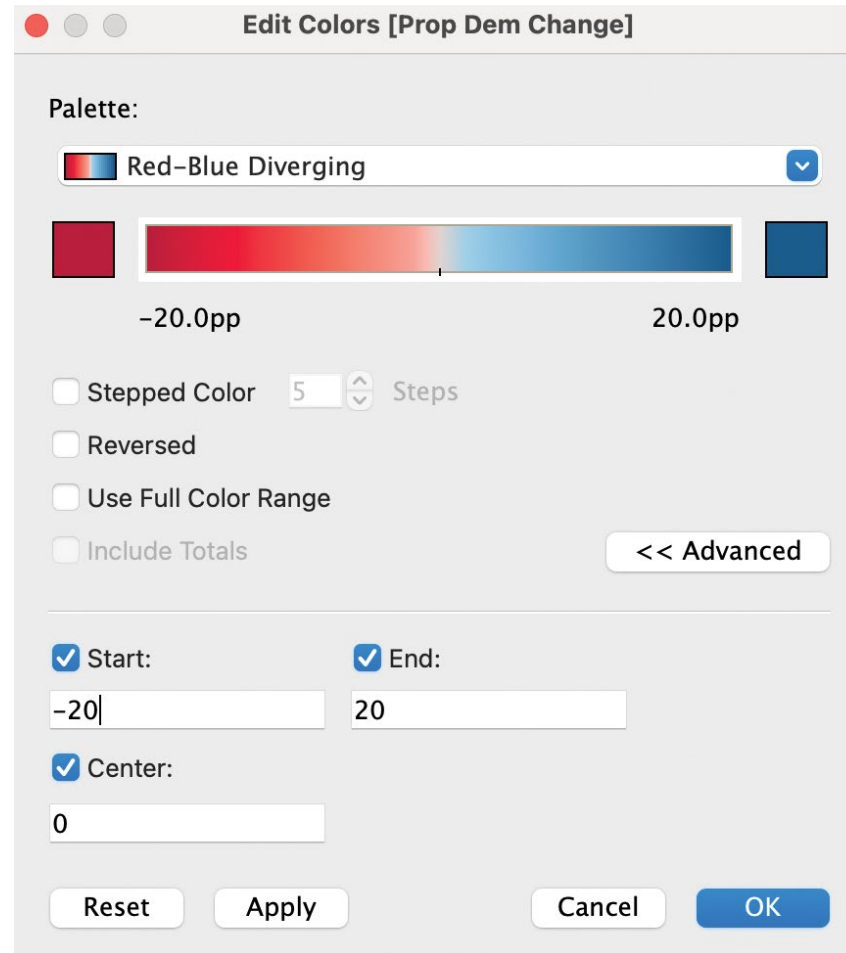
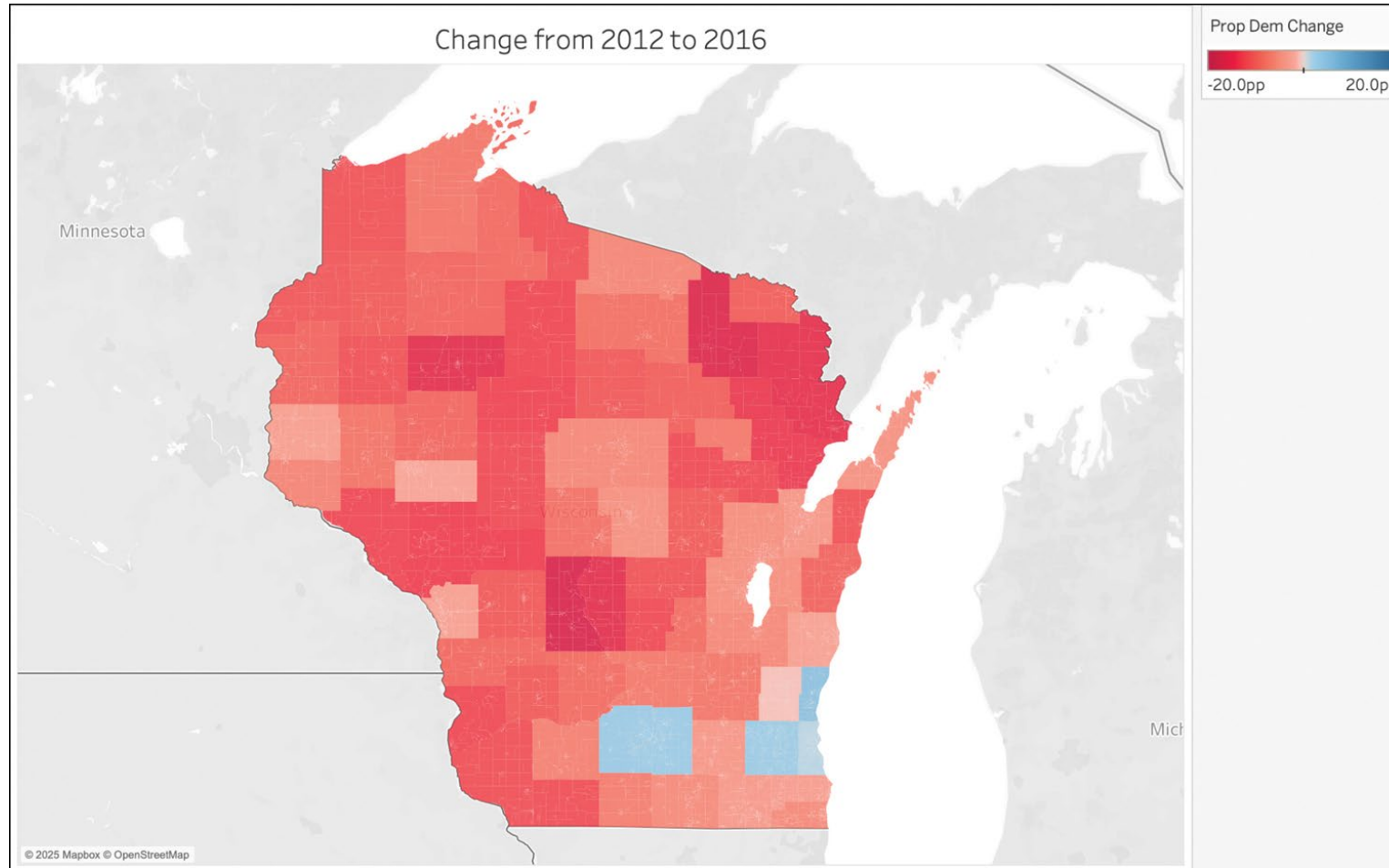


Figure 3.51 Choosing a meaningful color scale.



03fig52

Figure 3.52 Map shaded by the county change in Democratic vote share.

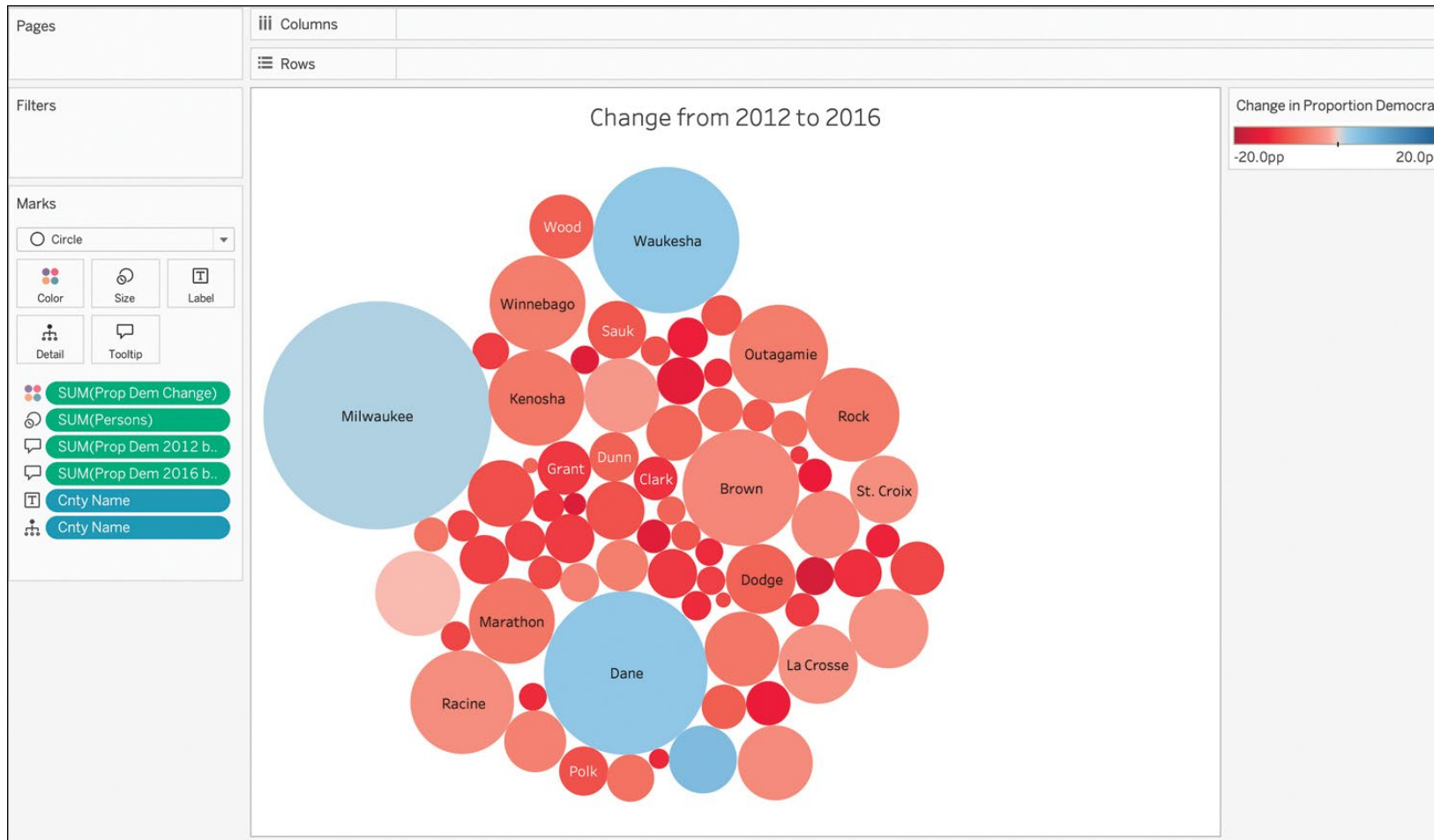


Figure 3.53 Bubble chart of the total change in Democratic vote share by county.

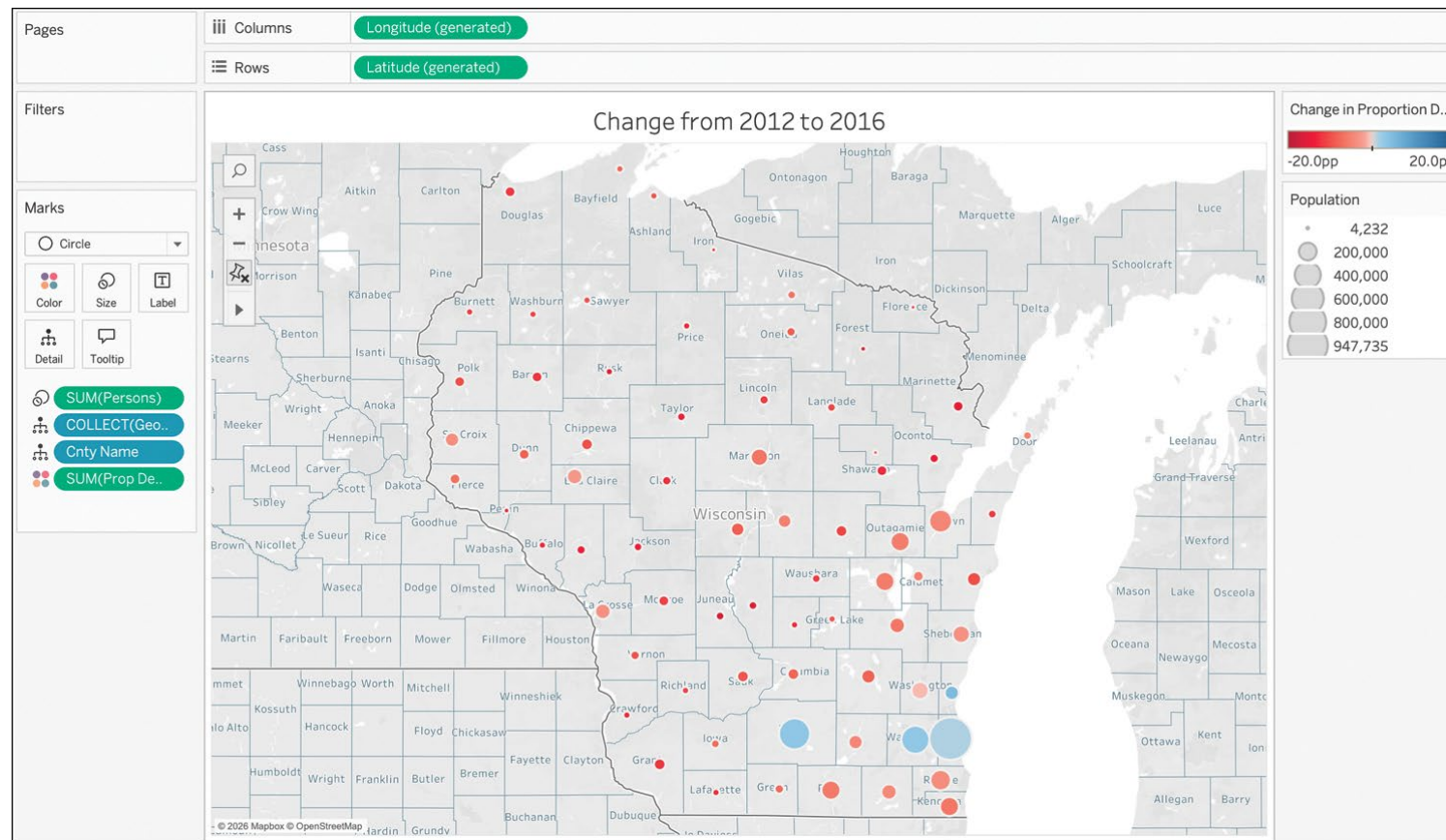


Figure 3.54 Geographic bubble chart.

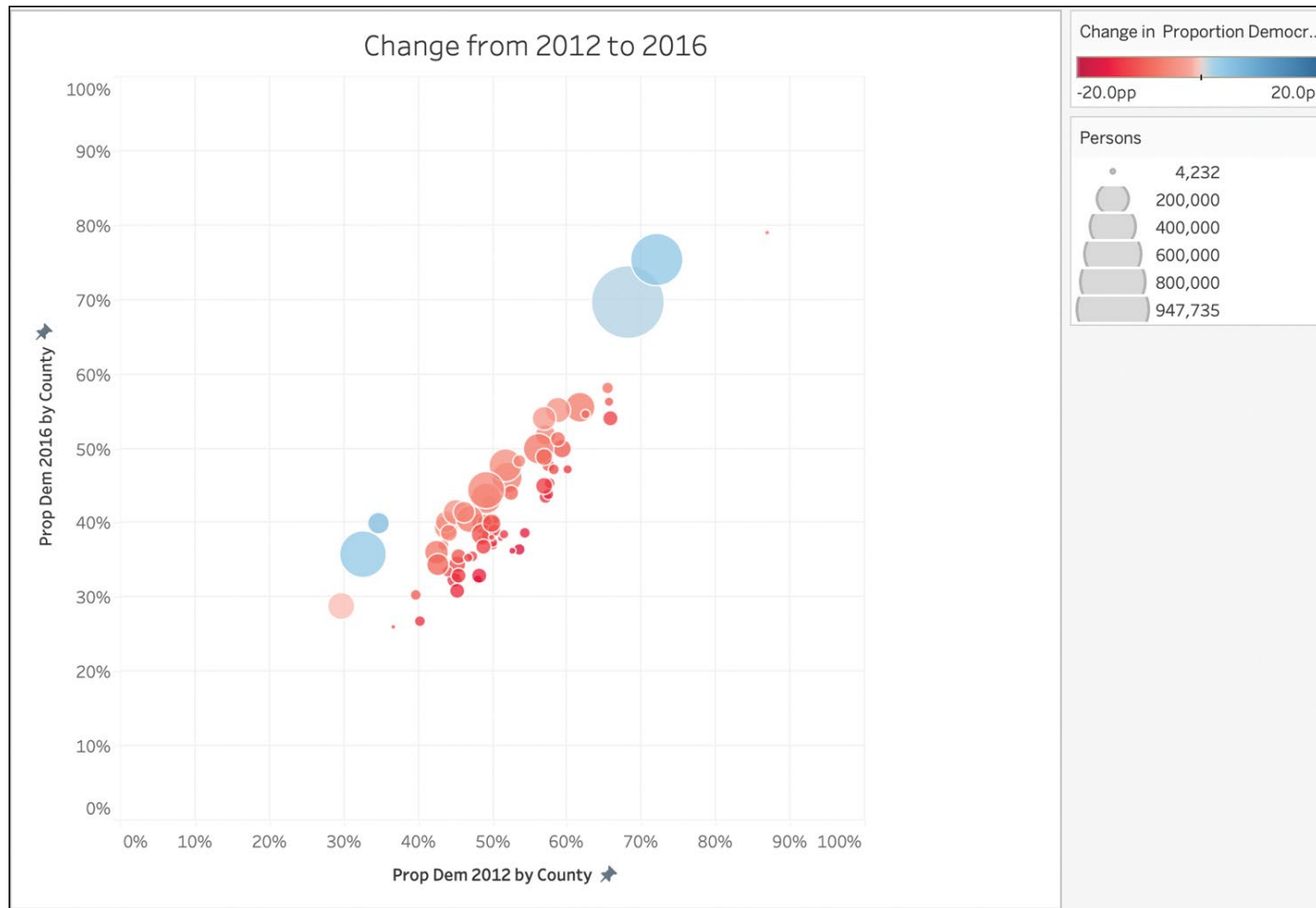


Figure 3.55 Initial scatterplot for the Democratic county vote shares in 2012 and 2016, colored by change and sized by population.

Edit Axis [Prop Dem 2012 by County]

General

Tick Marks

Range

☐ Automatic

☐ Uniform axis range for all rows or columns

☐ Independent axis ranges for each row or column

☒ Fixed

Fixed start

▼

Fixed end

▼

0

1

Scale

☐ Reversed

☐ Logarithmic

☒ Positive

☐ Symmetric

Axis Titles

Title

Prop Dem 2012 by County

Subtitle

Subtitle

☒ Automatic

↶ Reset

Figure 3.56 Editing an axis range.

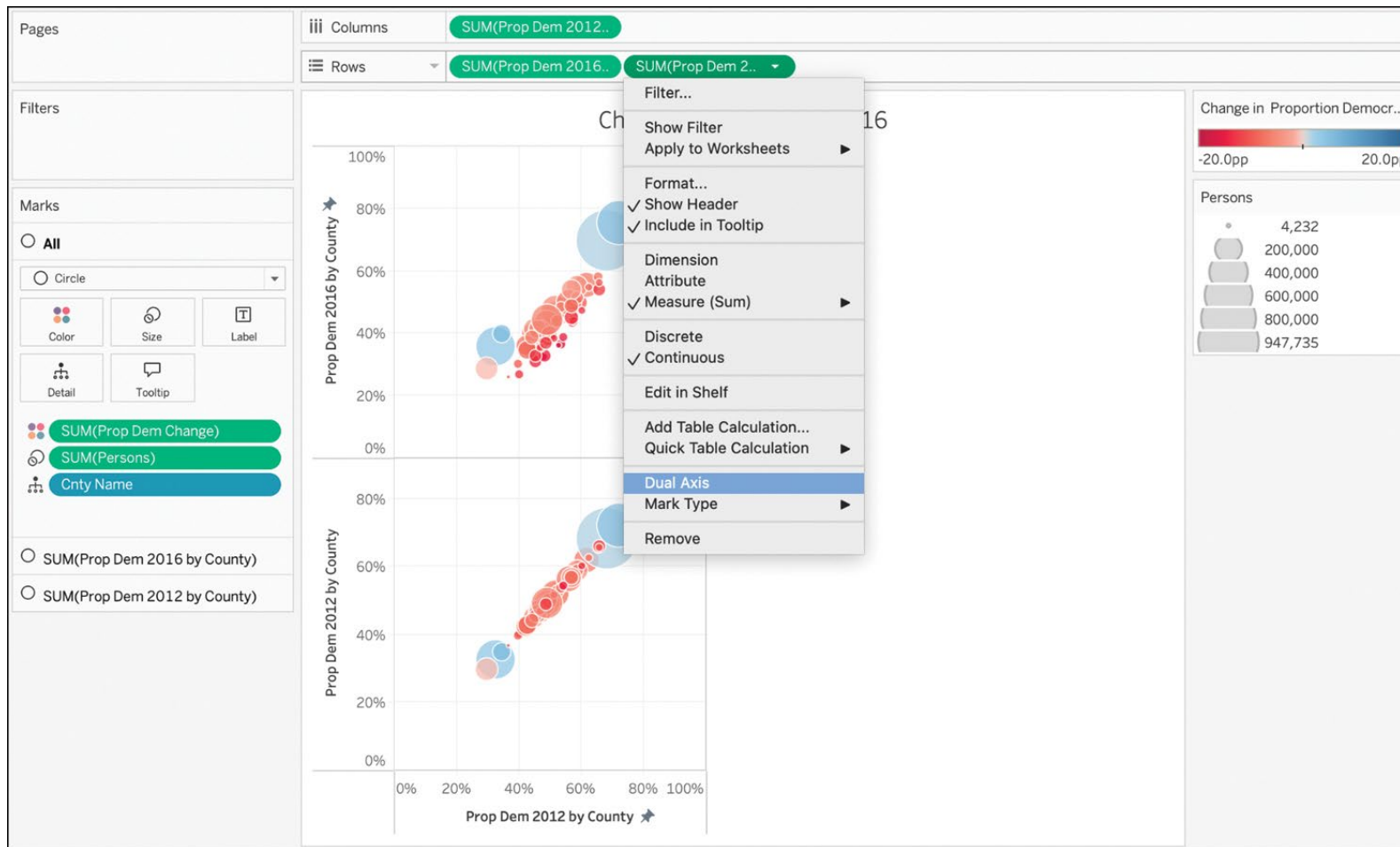


Figure 3.57 Creating a dual-axis graphic.

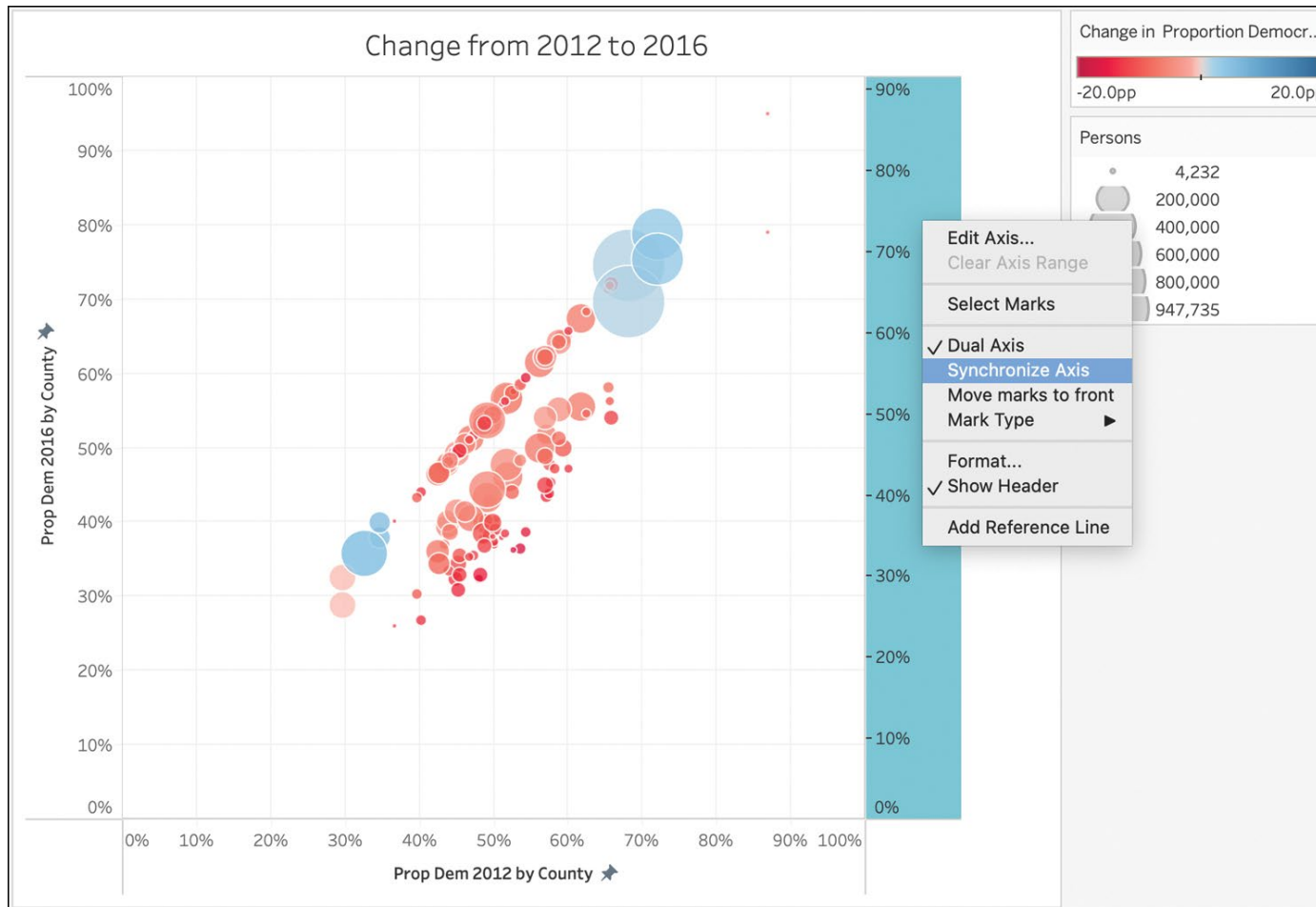


Figure 3.58 Synchronizing the vertical axes.

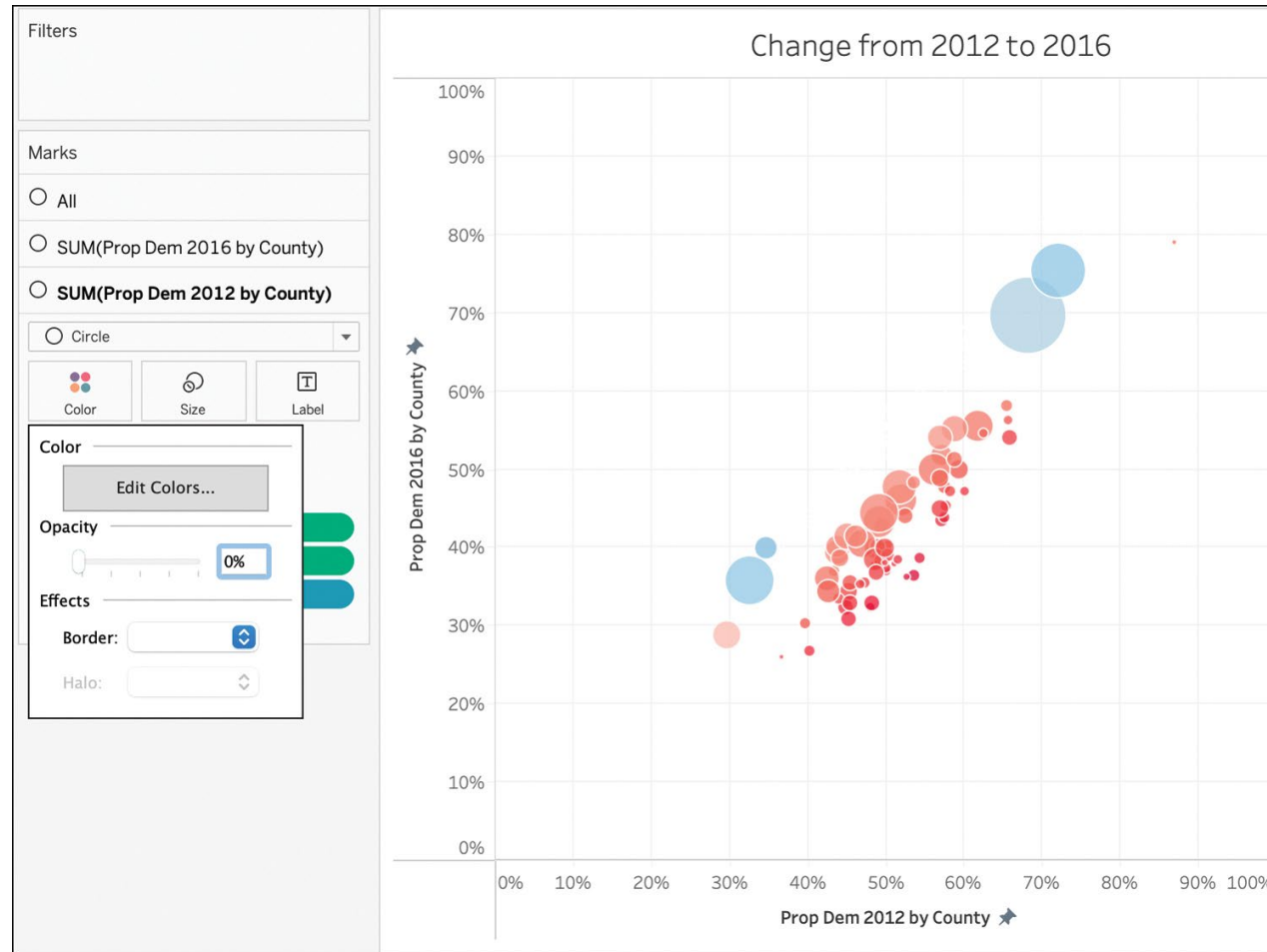


Figure 3.59 Hiding auxiliary bubbles.

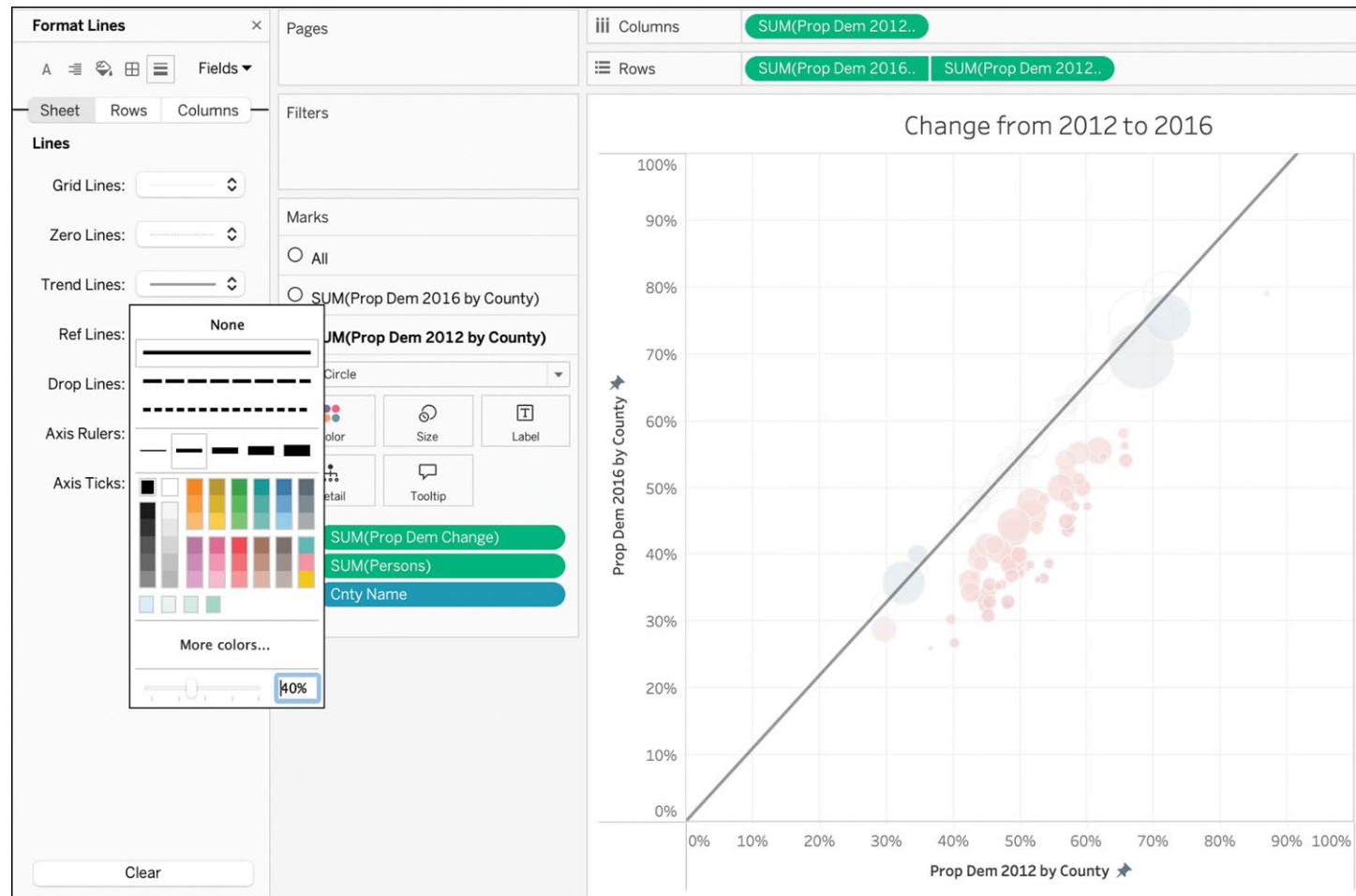


Figure 3.60 Adding a trendline.

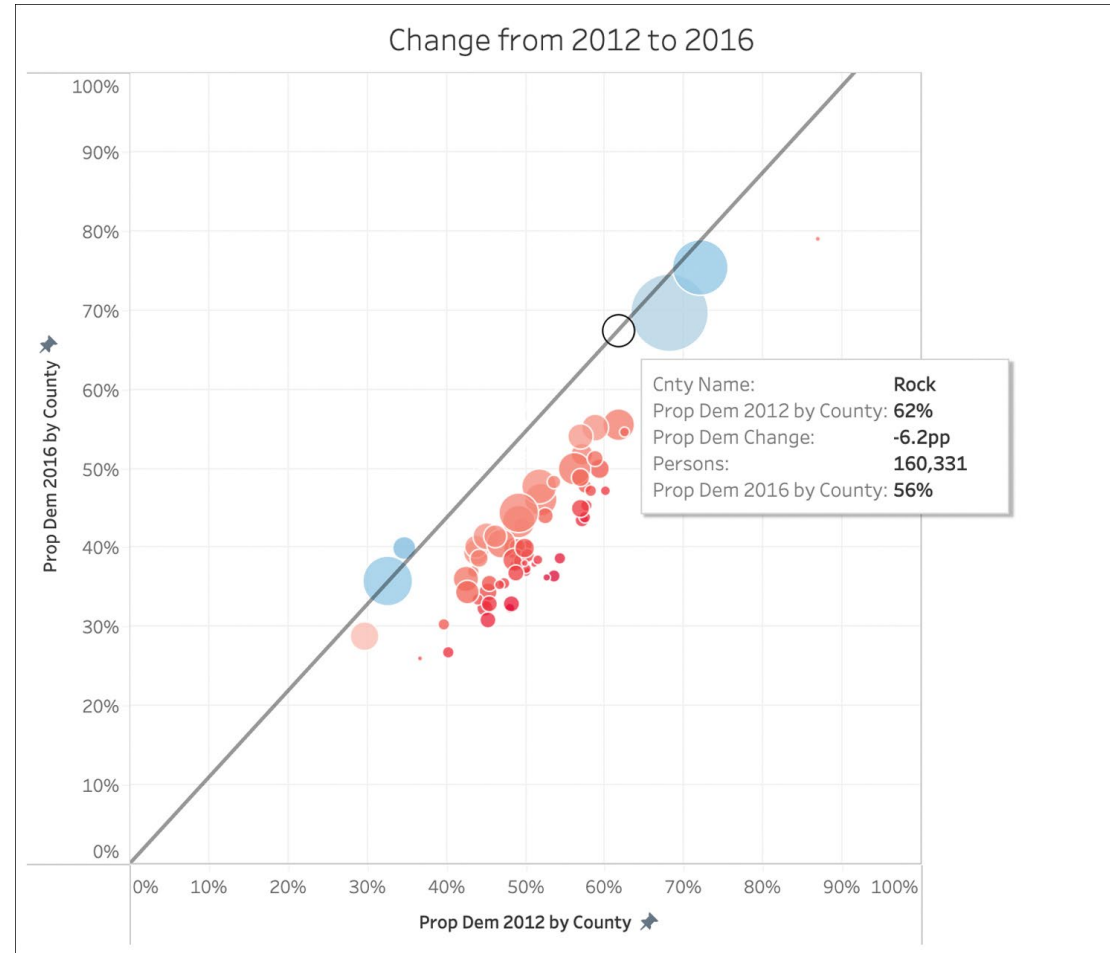


Figure 3.61 Hover effect to remove.

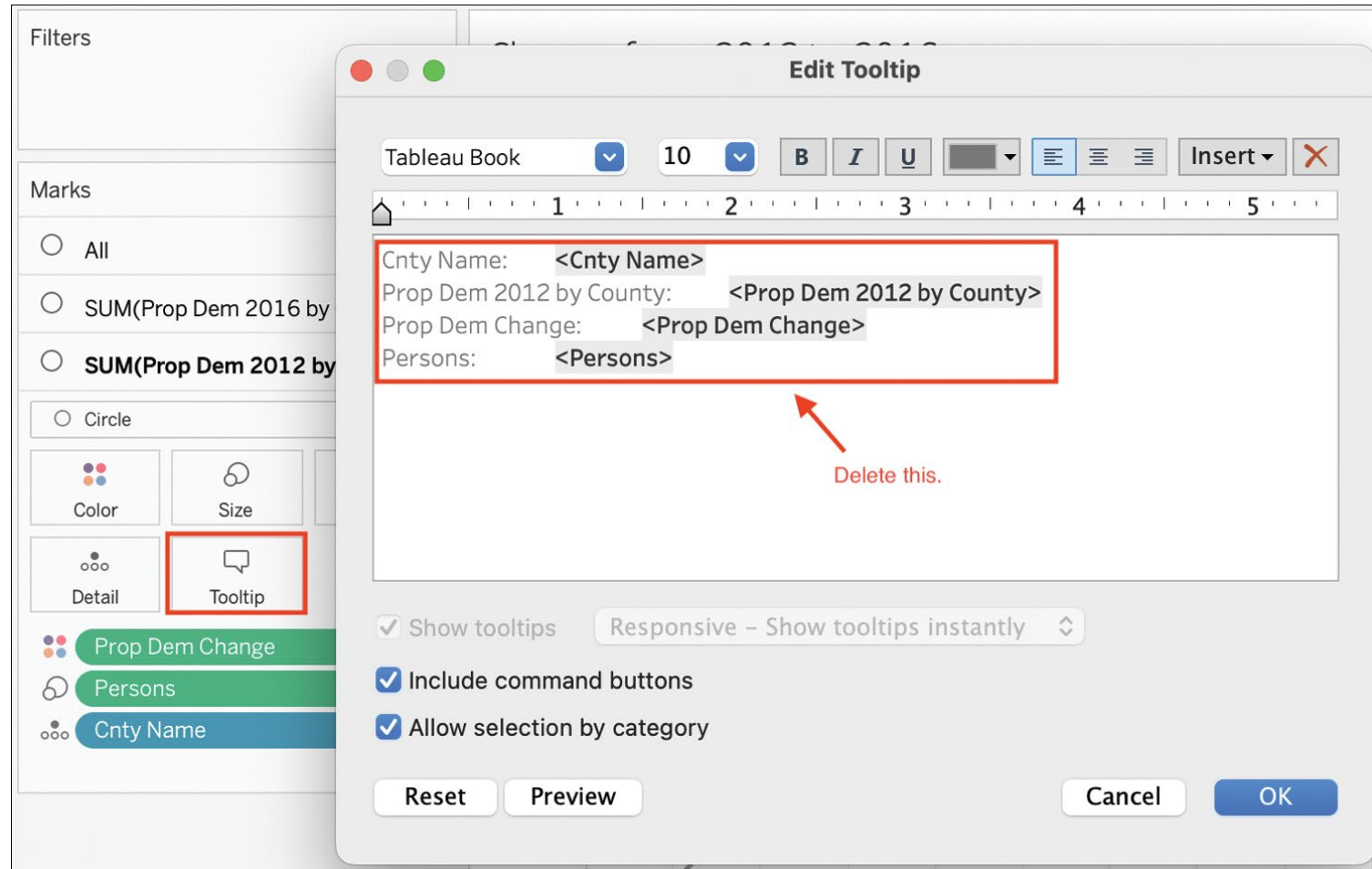


Figure 3.62 Removing hover information.

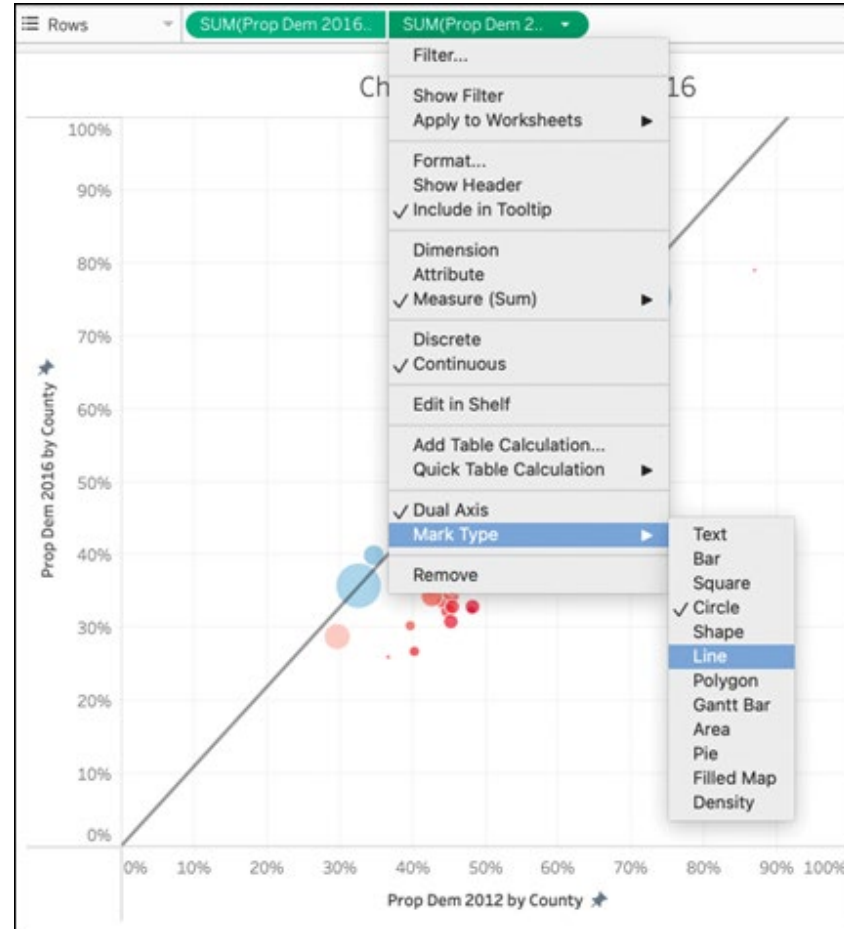


Figure 3.63 Changing mark type.

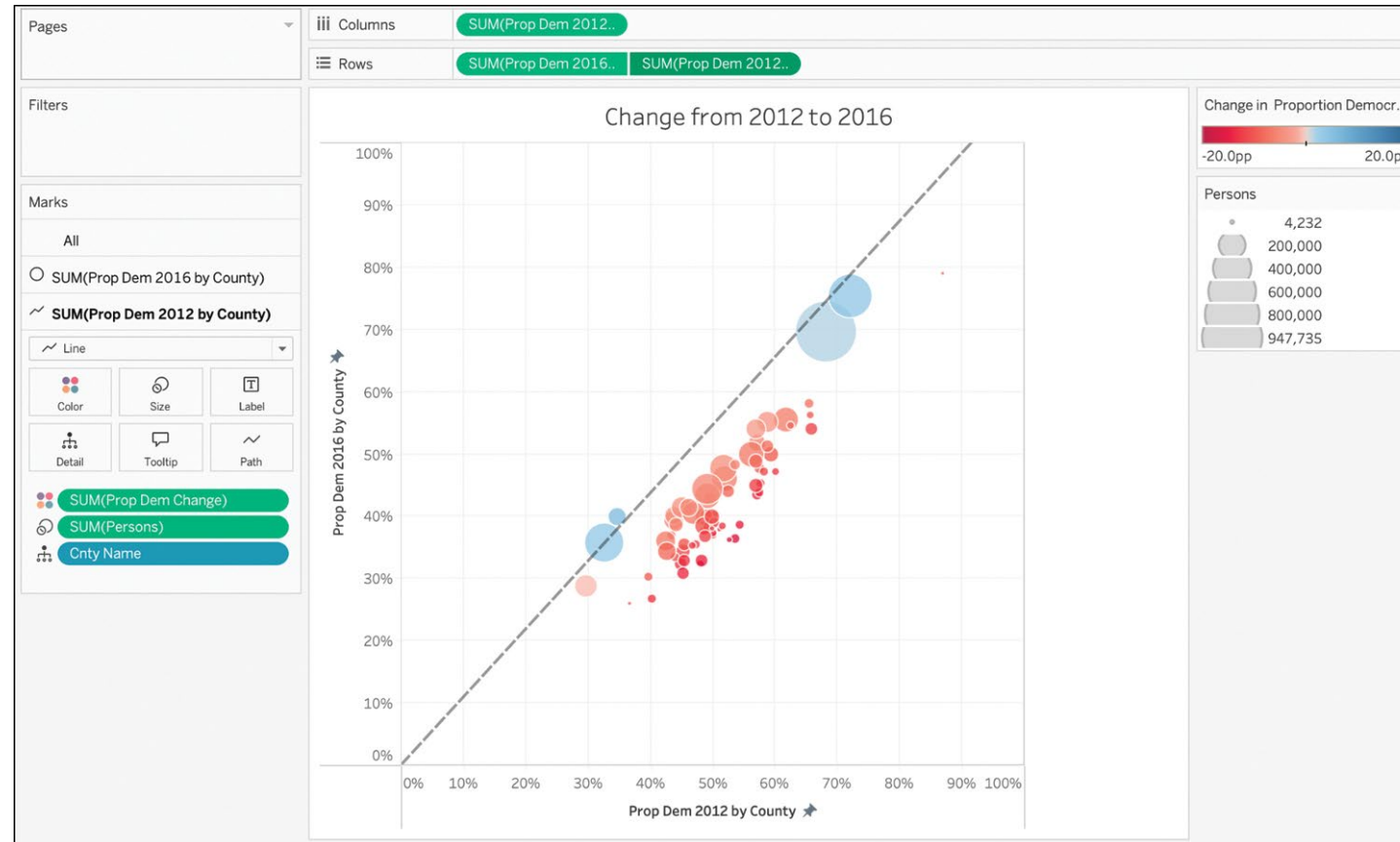


Figure 3.64 Completed scatterplot.

Figure 3.65 XML code to create a custom color scale in Tableau.

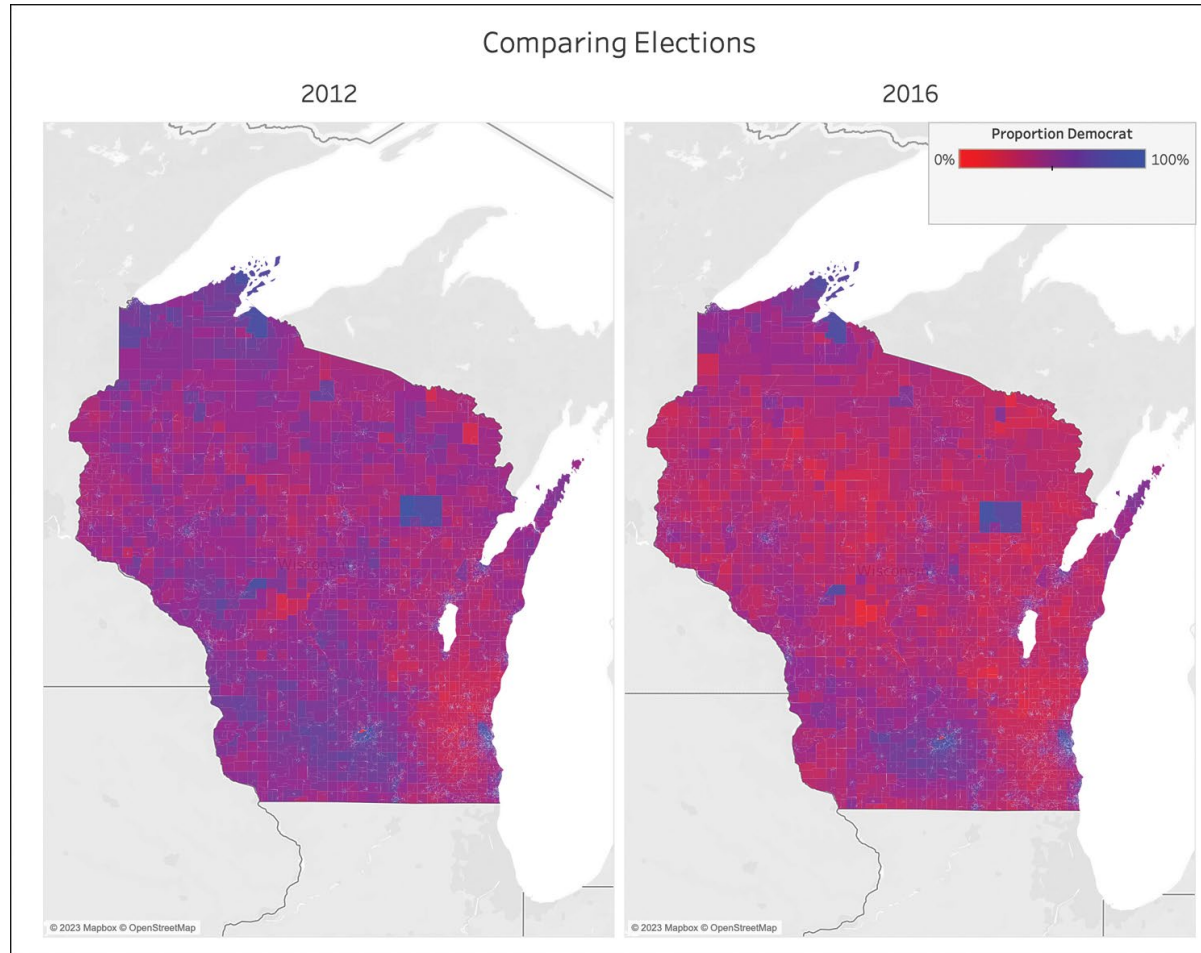


Figure 3.66 Side-by-side maps.

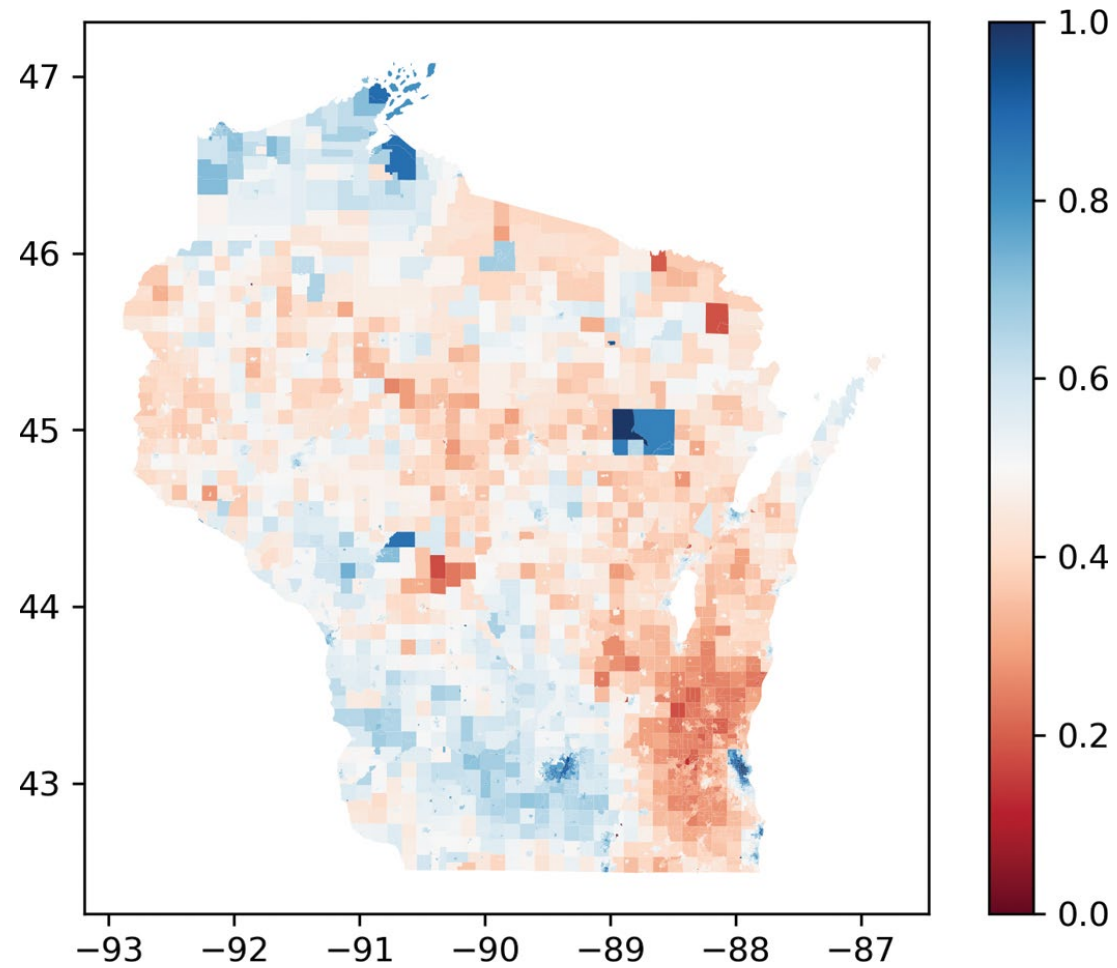


Figure 3.67 Initial matplotlib choropleth map of the Democratic vote share in 2012.

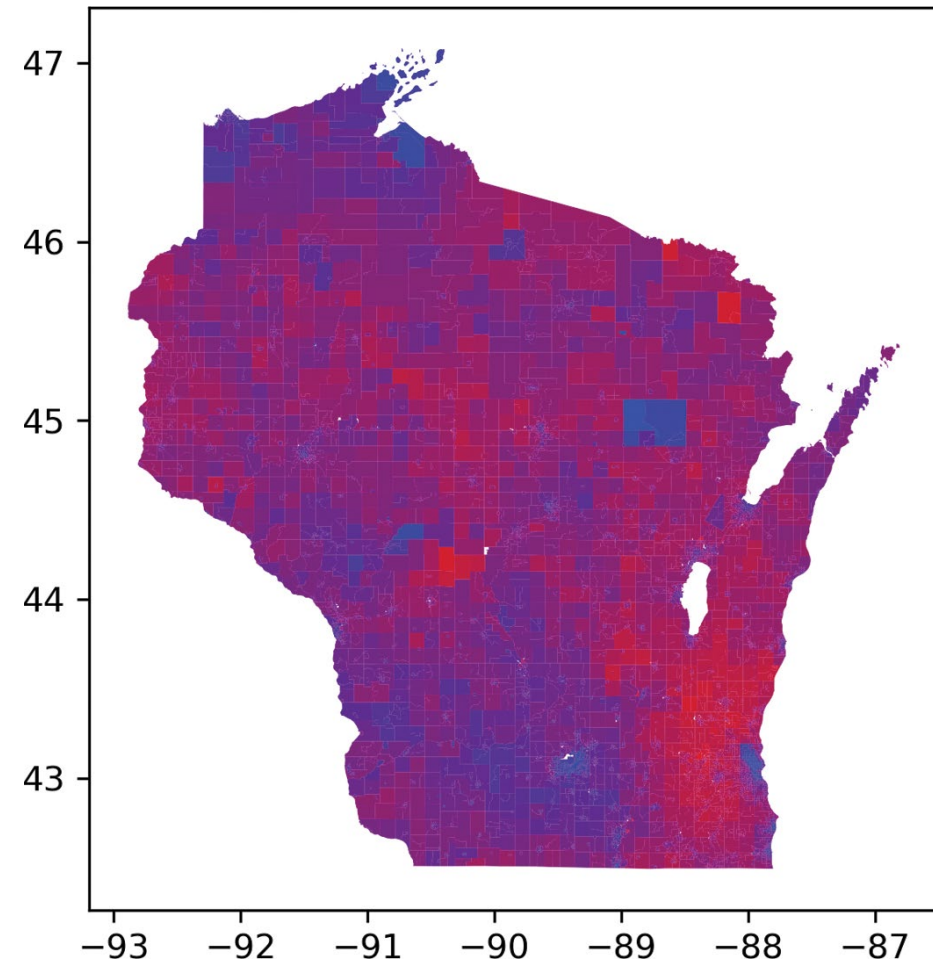


Figure 3.68 A better color scale of the Democratic vote share.

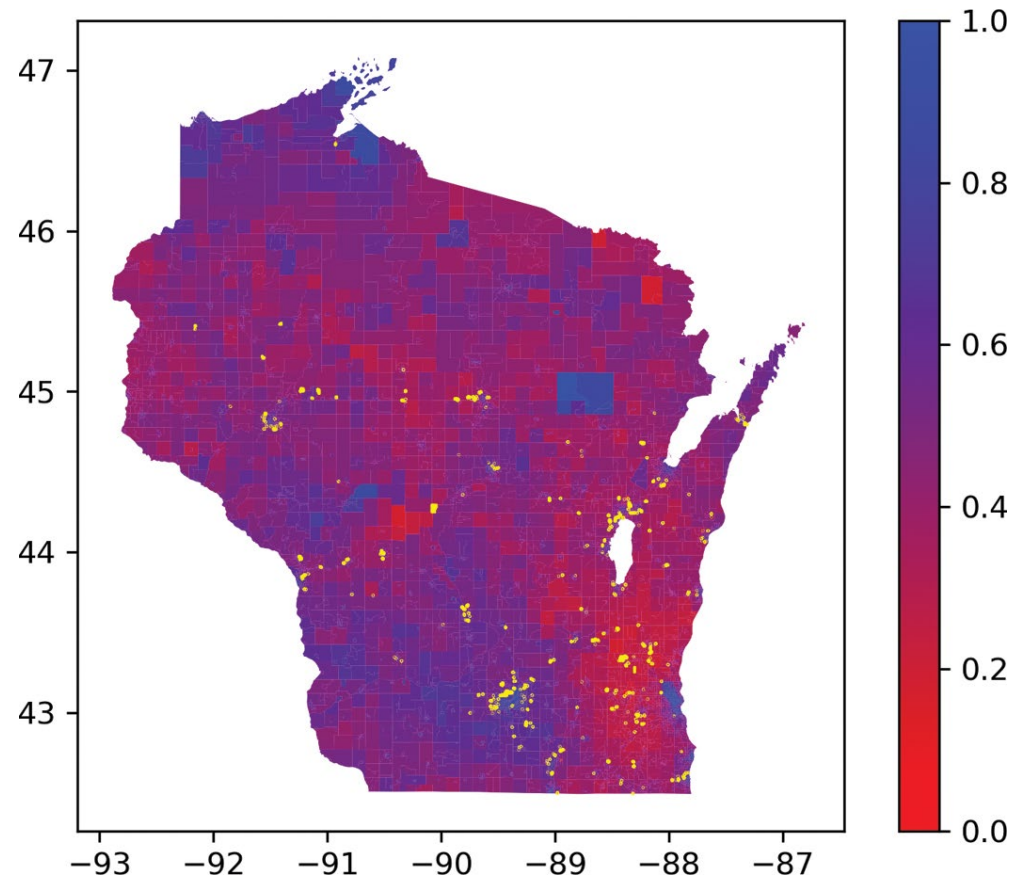


Figure 3.69 Showing missing values in yellow.

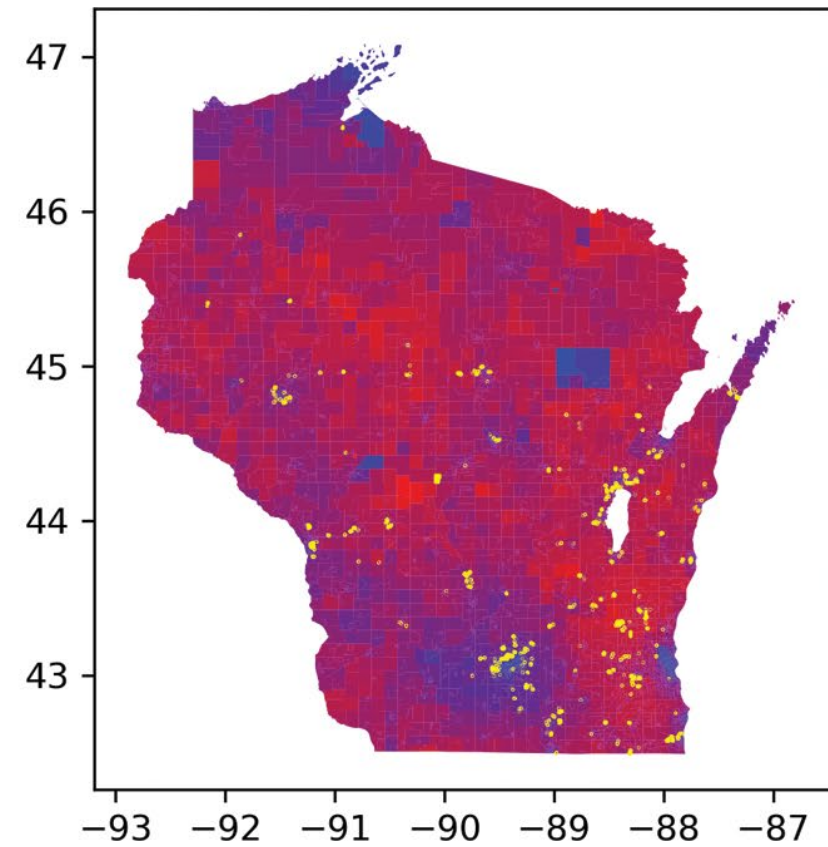
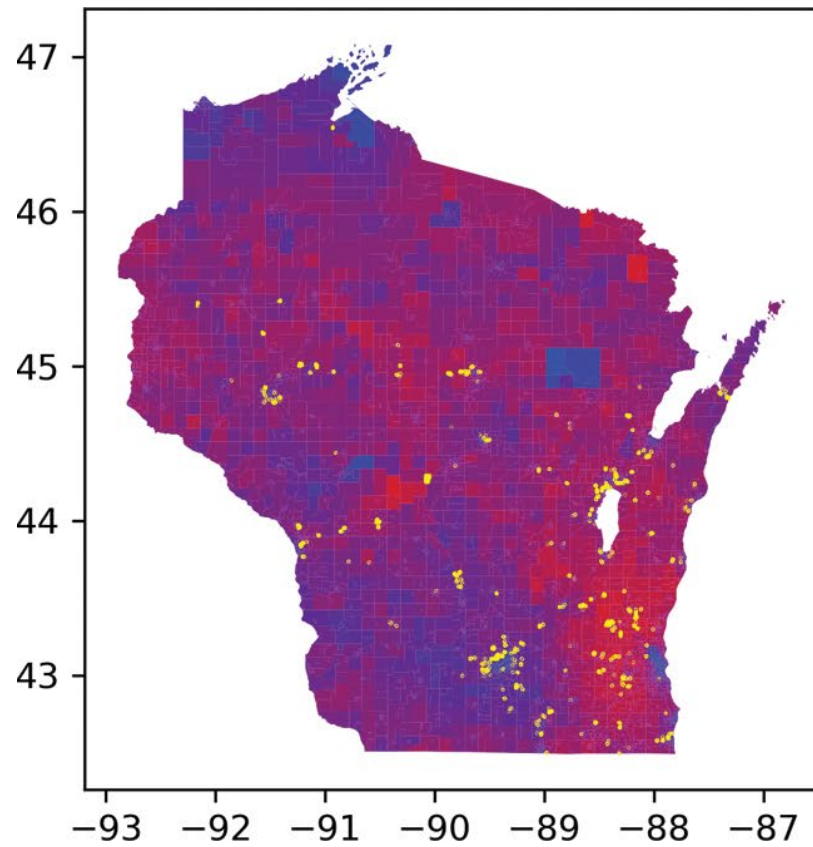


Figure 3.70 Generating side-by-side maps.

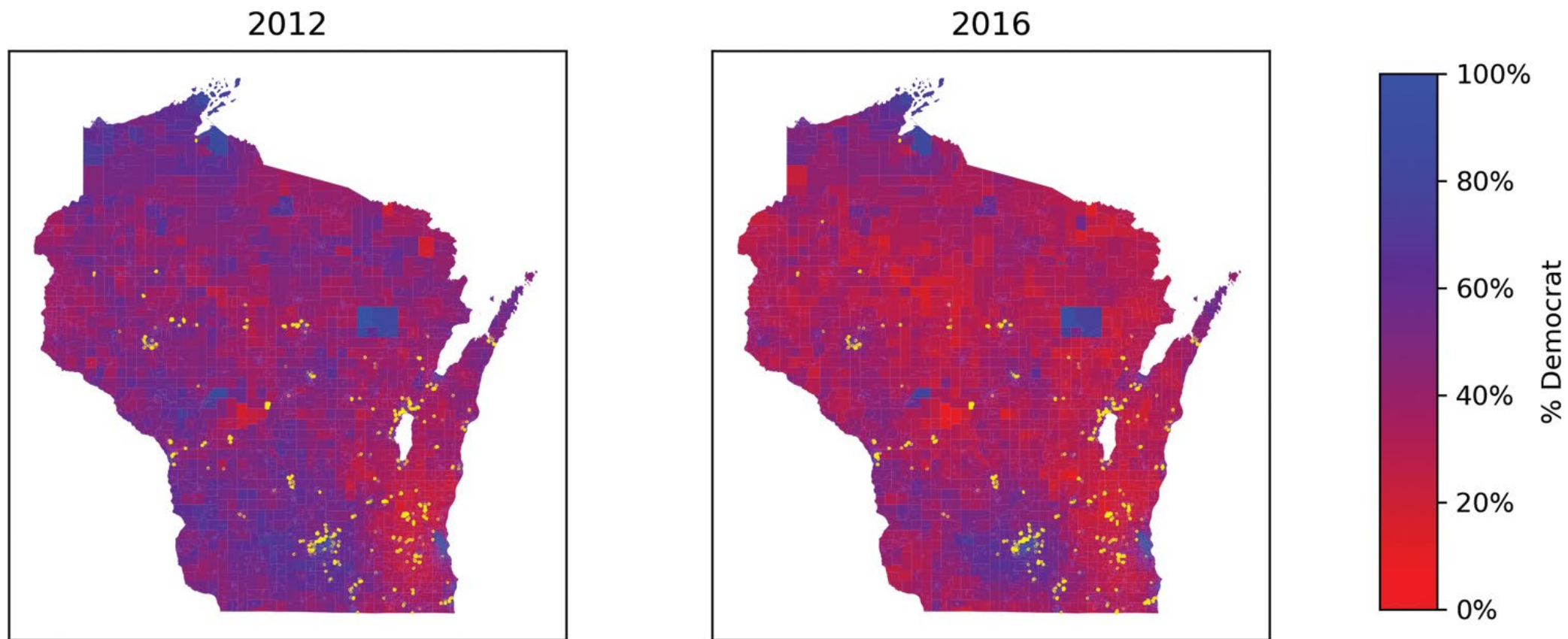


Figure 3.71 Side-by-side maps with a shared color legend.

Change from 2012 to 2016

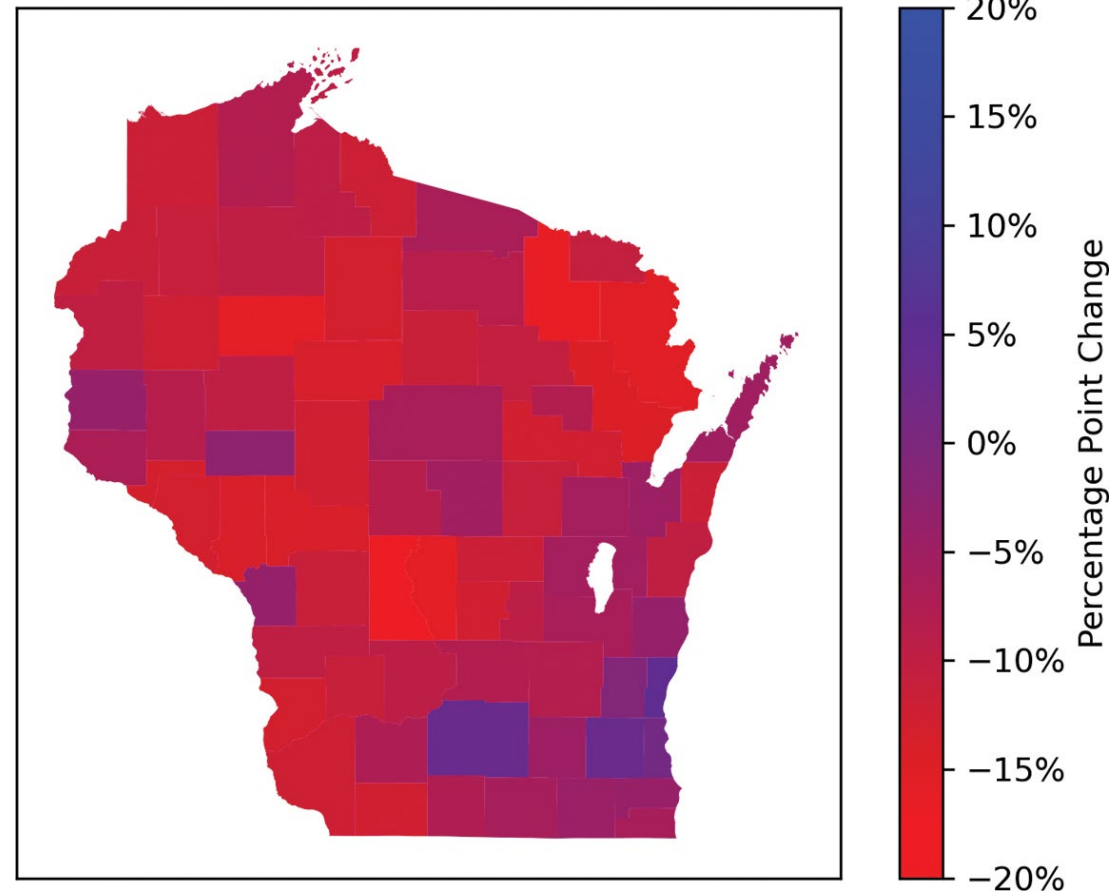


Figure 3.72 Choropleth shaded by county change.

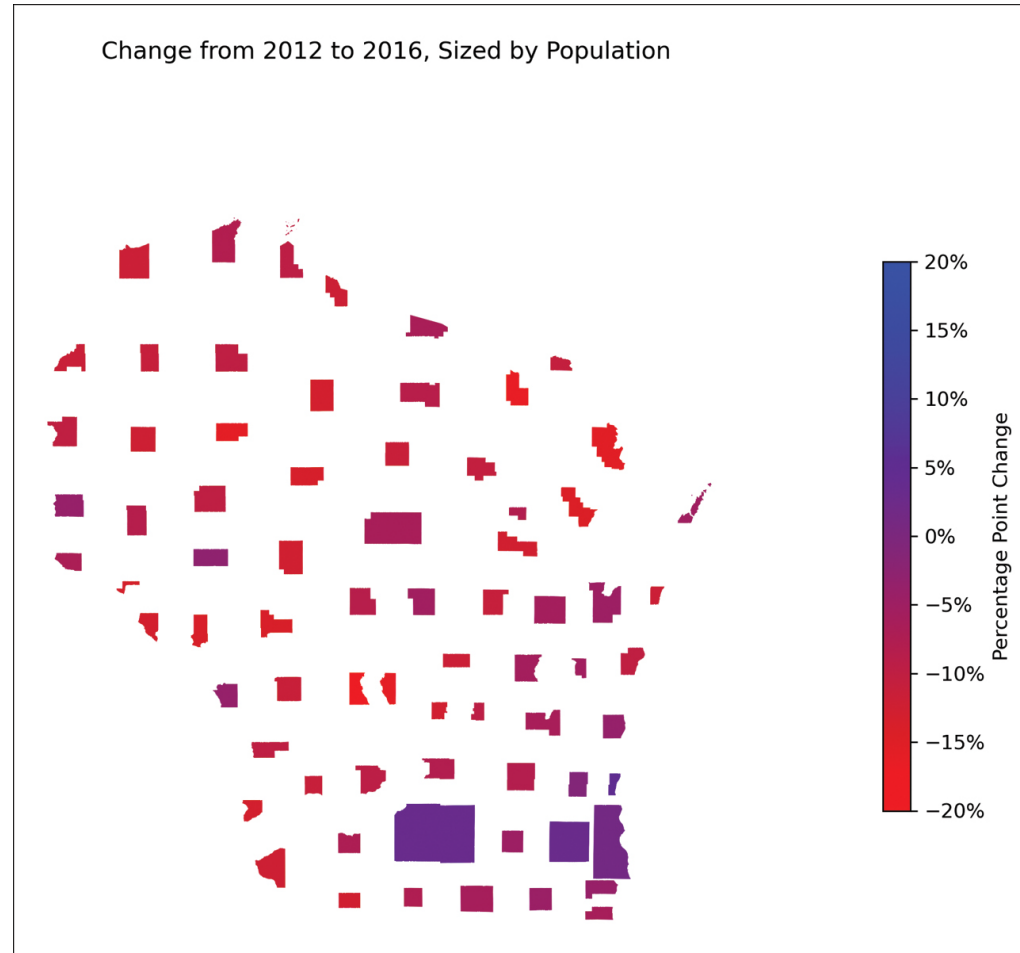


Figure 3.73 Cartogram showing the total change in Democratic vote share by county.

Change from 2012 to 2016

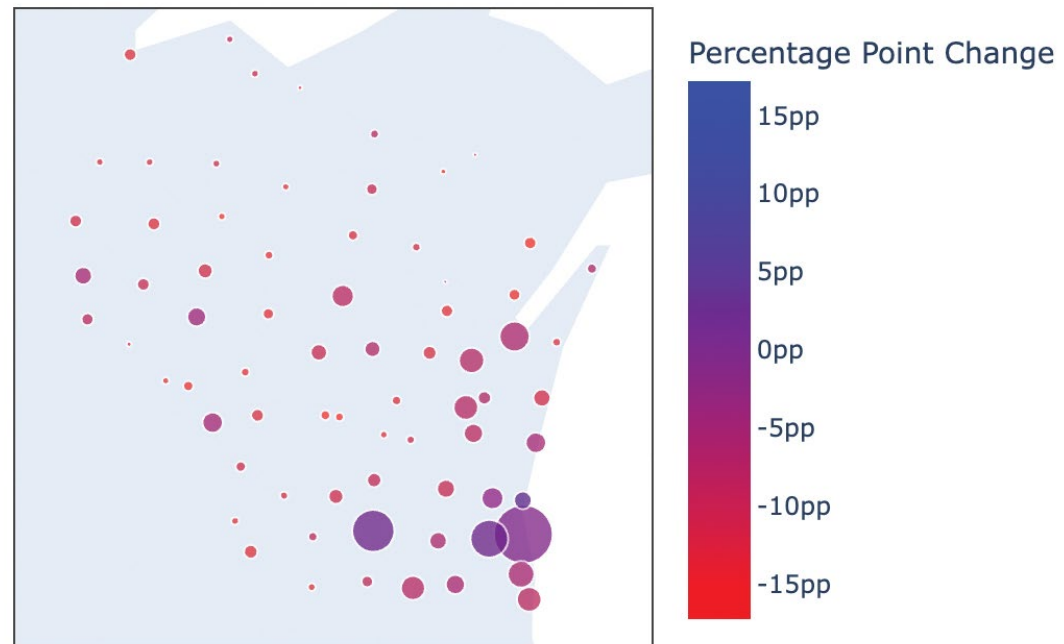


Figure 3.74 Geographic bubble chart.

Change from 2012 to 2016

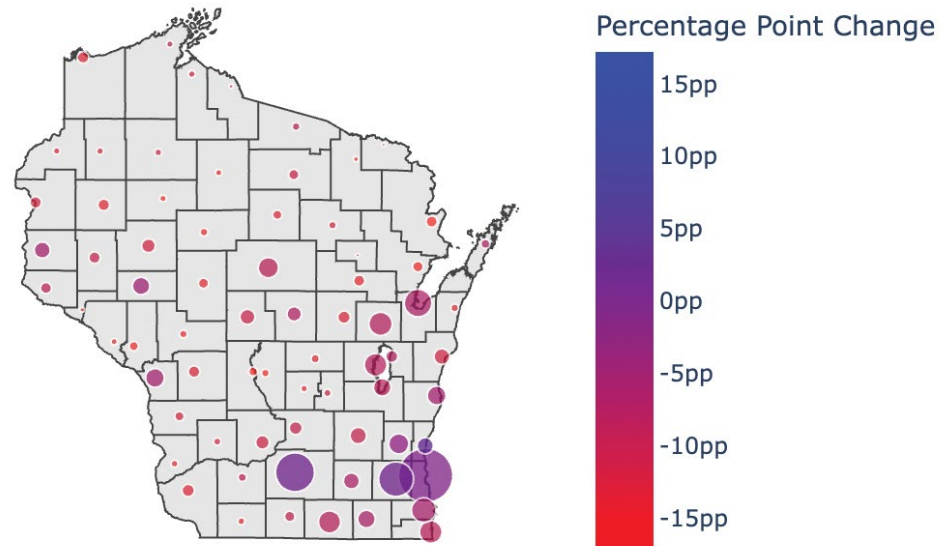


Figure 3.75 Geographic bubble chart with improved background map.

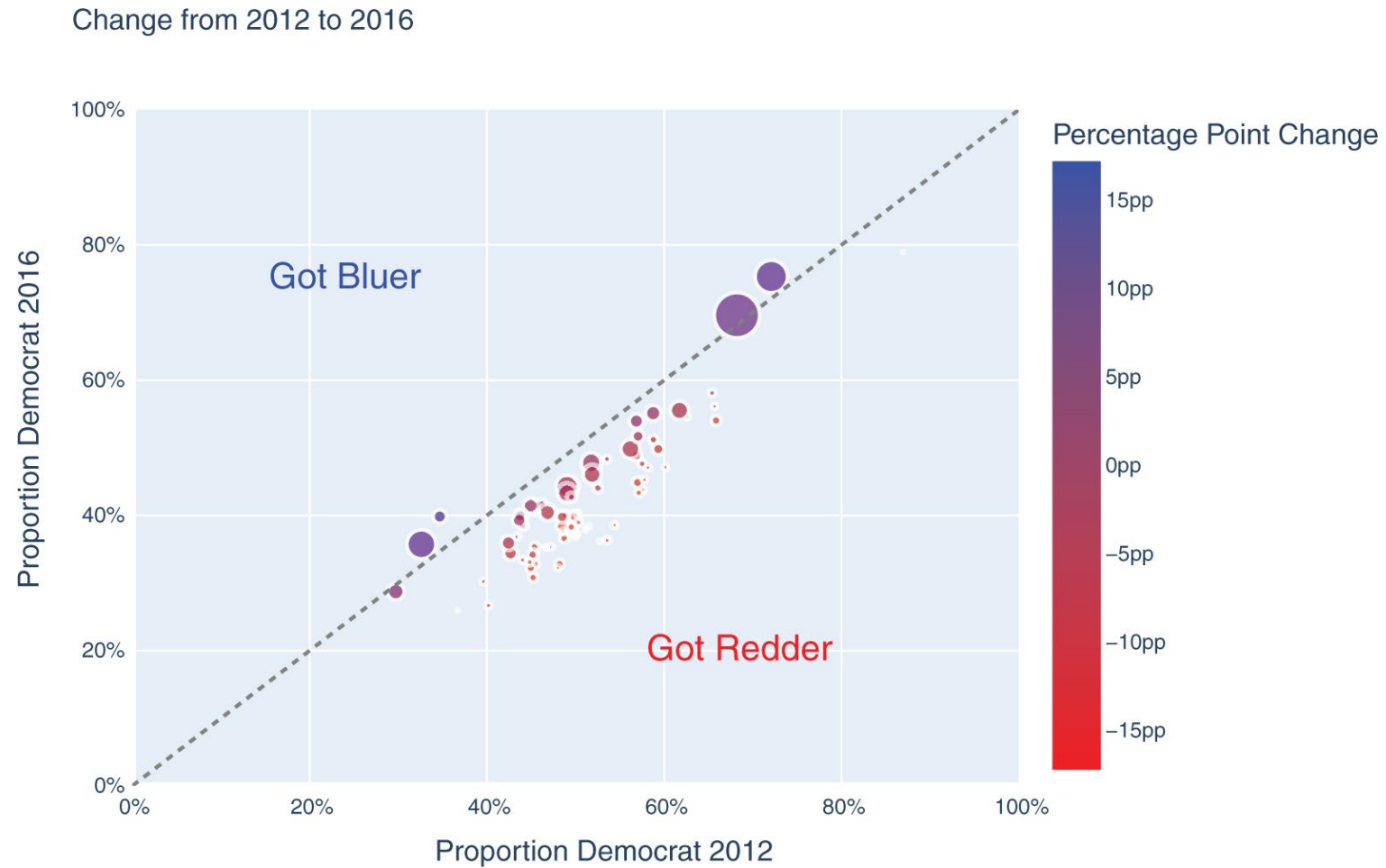


Figure 3.76 Scatterplot by county.

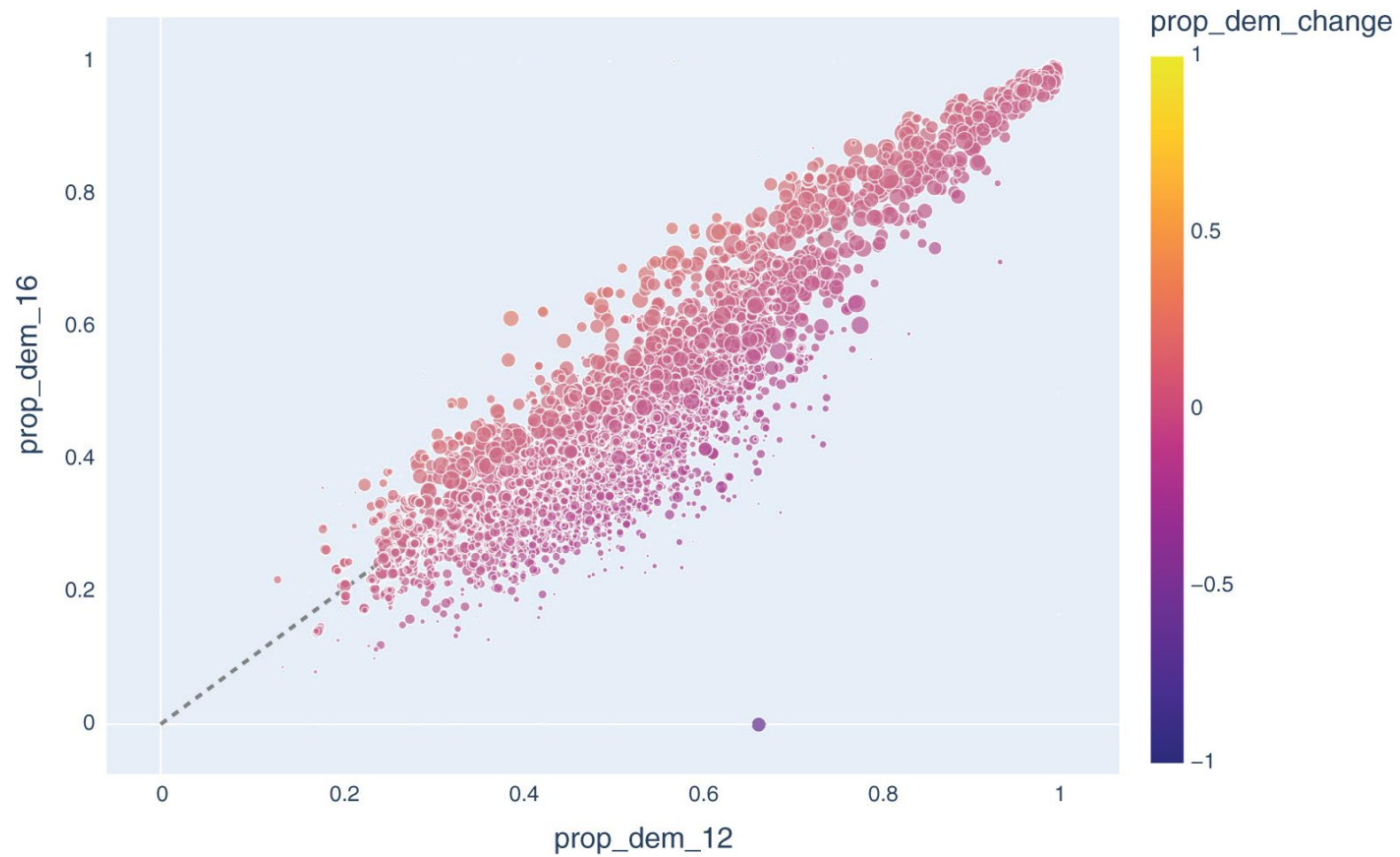


Figure 3.77 Scatterplot by ward.

Wisconsin_R - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

+ R + Go to file/function Addins

Wisconsin.R x state x

Filter Cols: << < 1 - 50 > >>

	OBJECTID	GEOID	CNTY_FIPS	CNTY_NAME	COUSUBFP	MCD_FIPS	MCD_NAME	CTV
1	1	55001002750001	55001	Adams	00275	5500100275	Adams	C
2	2	55001002750002	55001	Adams	00275	5500100275	Adams	C
3	3	55001002750003	55001	Adams	00275	5500100275	Adams	C
4	4	55001002750004	55001	Adams	00275	5500100275	Adams	C
5	5	55001003000001	55001	Adams	00300	5500100300	Adams	T

Figure 3.78 The state sf data.frame.

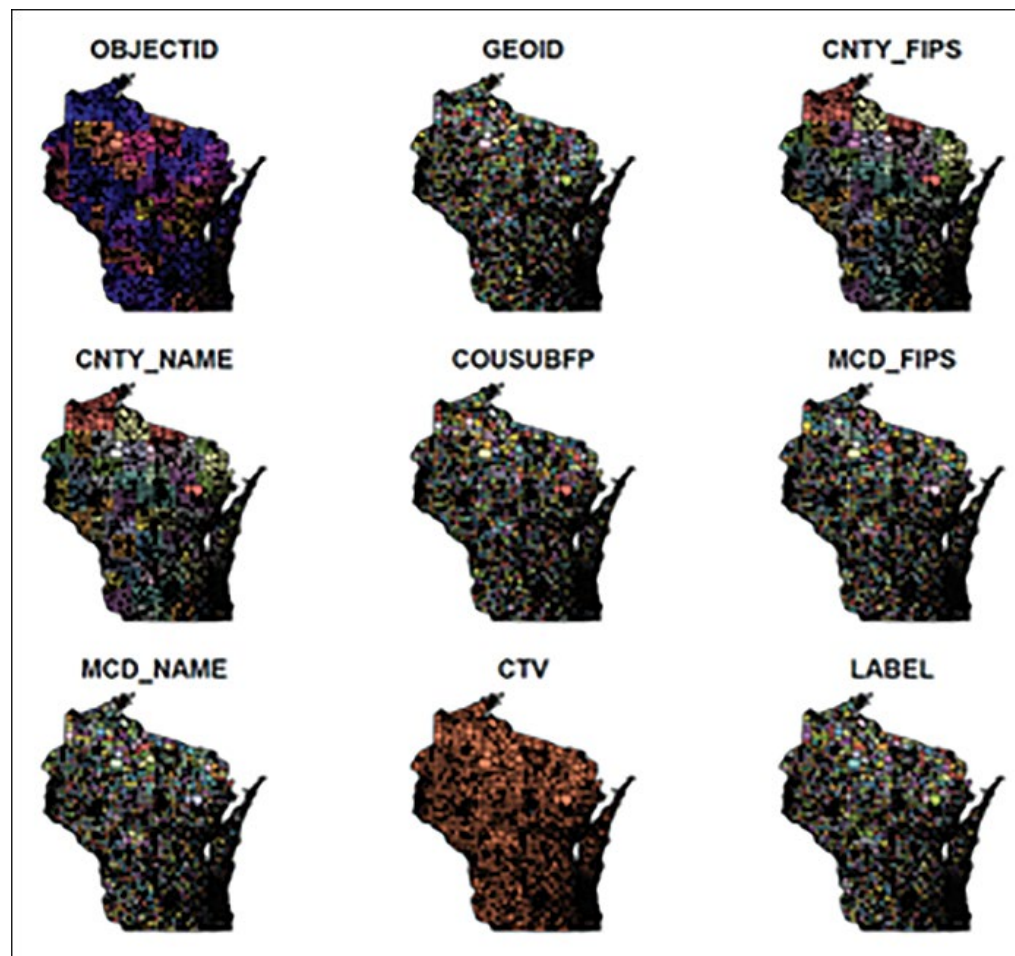


Figure 3.79 Using `plot(state)`, R generates a map for each variable!

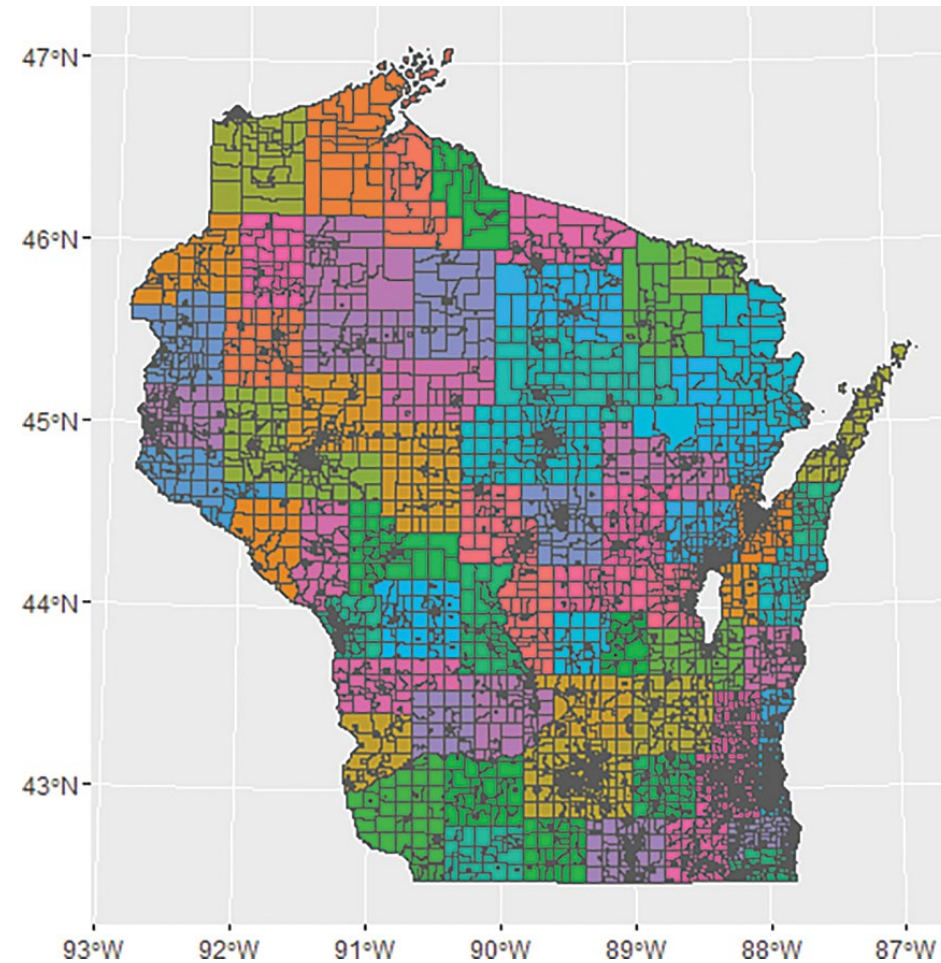


Figure 3.80 Using ggplot and geom_sf we are able to make a map showing all the wards and colored by county.

Wisconsin.R x state x smist x							
← → 📄 🔍 Filter							
	CNTY_FIPS	CNTY_NAME	PREREP12	PREDEM12	PREREP16	PREDEM16	PERS
1	55001	Adams	101	168	127	110	
2	55001	Adams	98	164	124	107	
3	55001	Adams	44	73	56	48	
4	55001	Adams	55	92	70	60	
5	55001	Adams	181	240	253	139	
6	55001	Adams	67	88	93	51	

Figure 3.81 Wisconsin elections dataset after selecting and mutating.

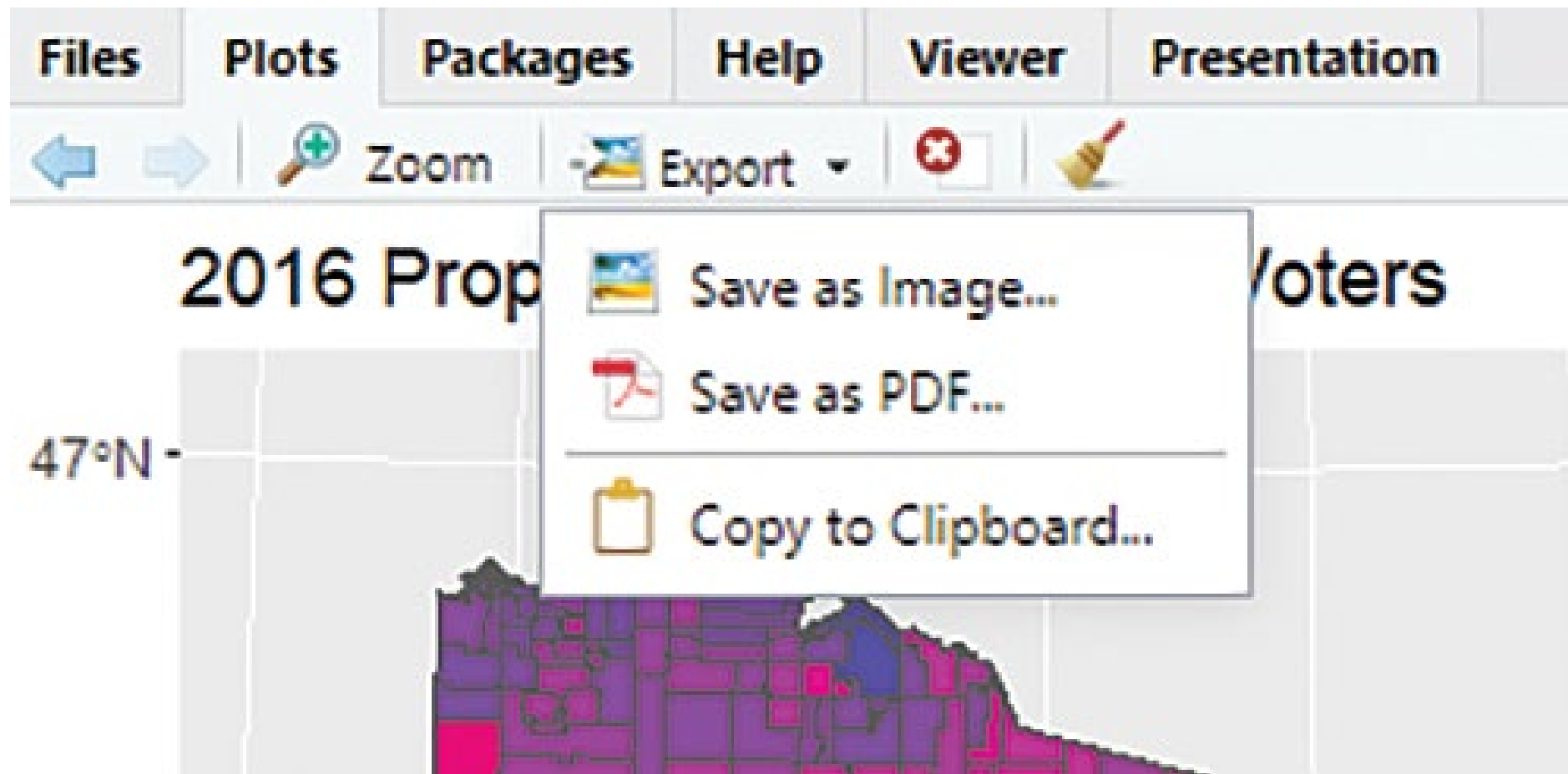


Figure 3.82 Exporting the map as an image file.

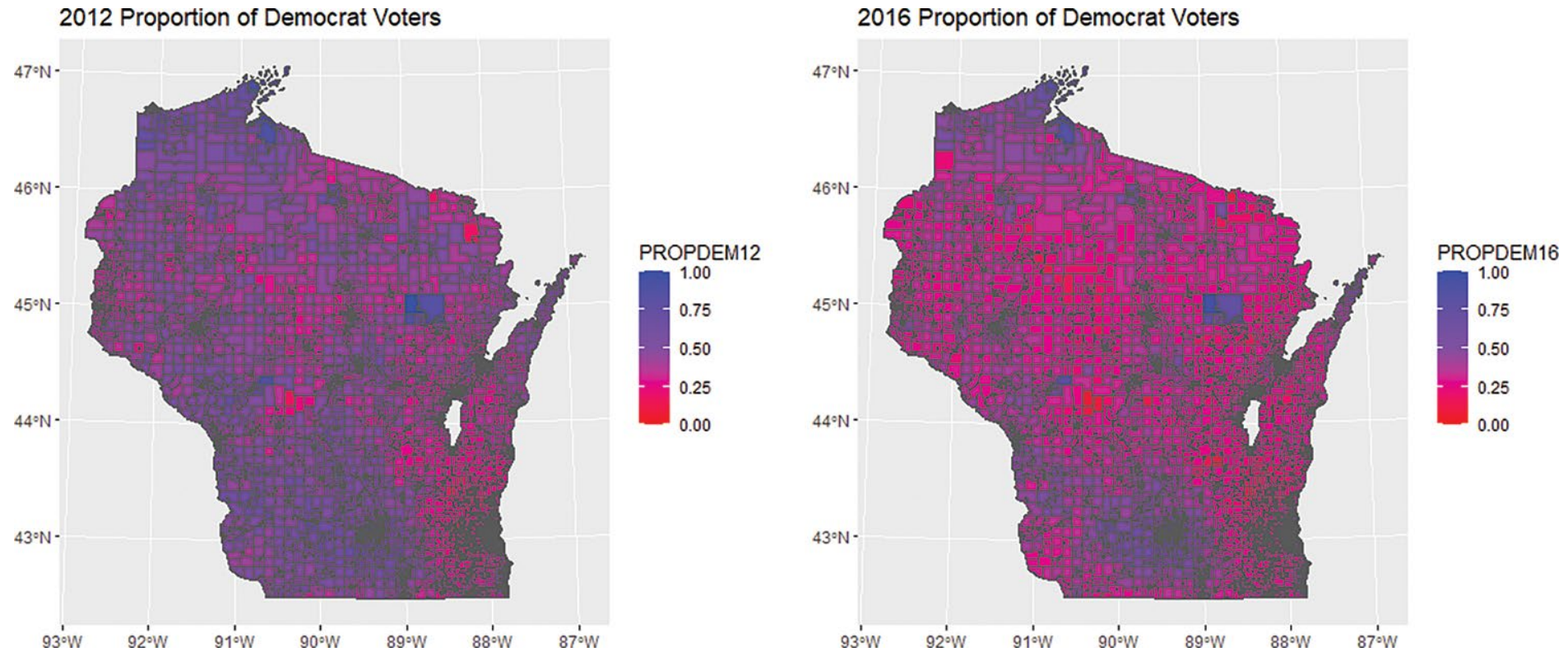


Figure 3.83 Maps showing political leanings in 2012 and 2016 by ward.

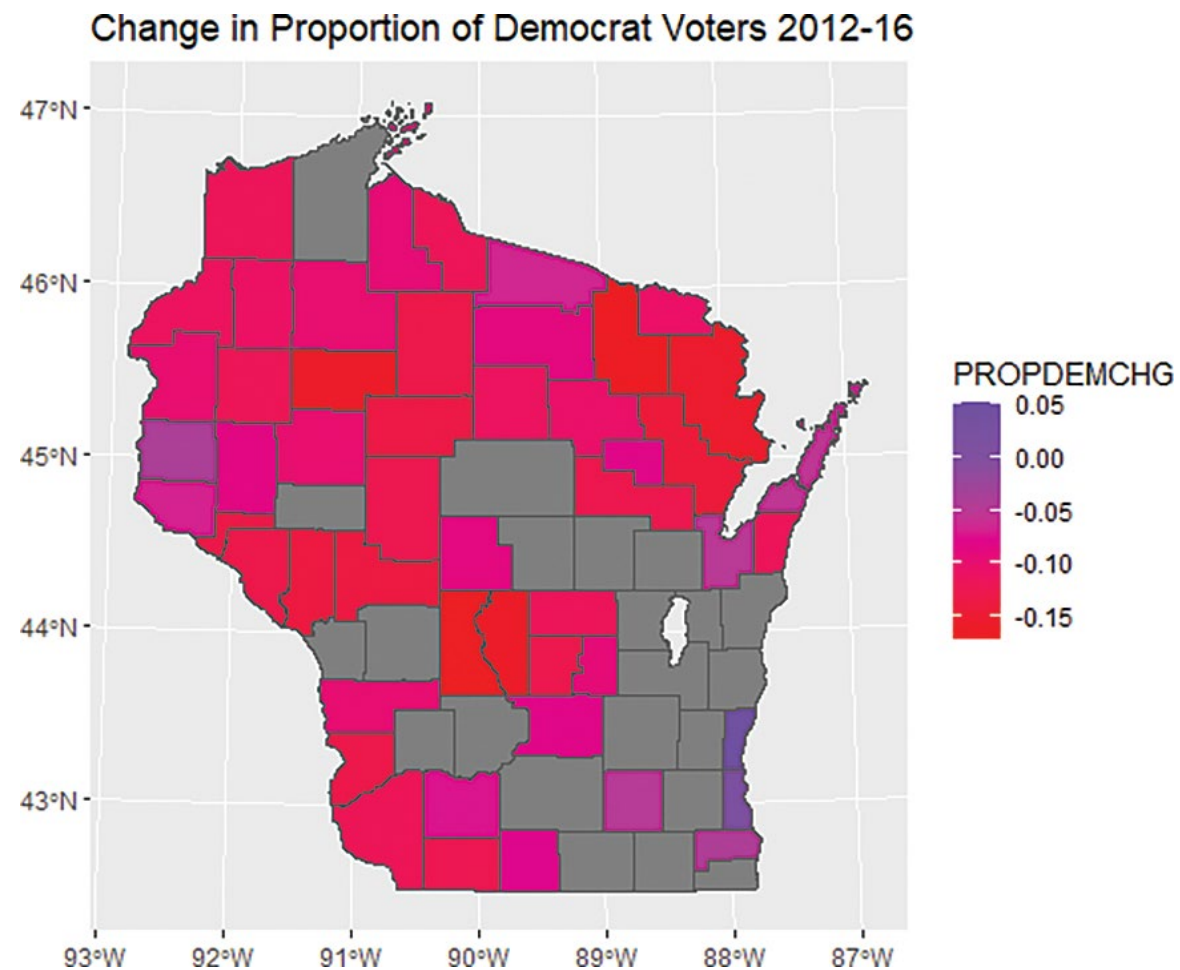


Figure 3.84 Map with NA values.

Wisconsin.R x county2 x county x state x smist x								
Filter								
	predem16	prerep16	persons	geometry	PROPDEM12	PROPDEM16	PptDEMCHG	
644	3745	5966	20875	POLYGON ((1762636 1159366, ...	0.5440801	0.3856451	-15.84349	
320	4226	3303	16157	MULTIPOLYGON (((1581926 200...	0.6568926	0.5612963	-9.55962	
443	7889	13614	45870	POLYGON ((1300275 1729516, ...	0.4876192	0.3668790	-12.07401	
603	4953	4124	15014	MULTIPOLYGON (((1472111 209...	0.6260897	0.5456649	-8.04248	
335	53382	67210	248004	MULTIPOLYGON (((2259109 138...	0.4909352	0.4426662	-4.82690	
364	2522	4044	13587	POLYGON ((1266358 1323800, ...	0.5148543	0.3840999	-13.07544	
550	2949	5410	15457	POLYGON ((1060611 1924297, ...	0.4669634	0.3527934	-11.41700	
539	9642	15367	48971	MULTIPOLYGON (((2195527 125...	0.4414093	0.3855412	-5.58680	

Figure 3.85 Calculating the change in the proportion of Democratic votes by county.

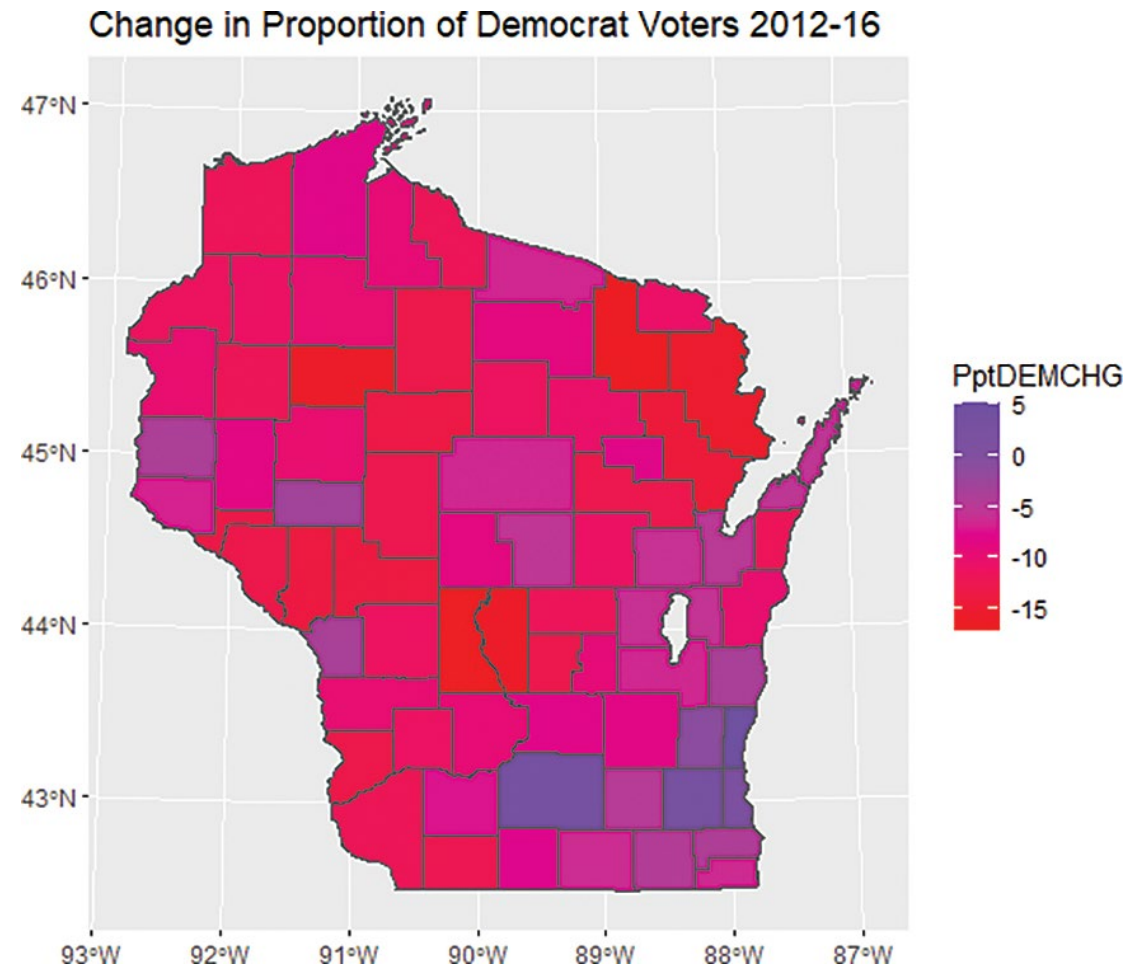


Figure 3.86 Map shaded by change in the Democratic vote share from 2012 to 2016.

Wisconsin.R x county2 x county x state x smlst x				
Filter				
eometry	PROPDEM12	PROPDEM16	PptDEMCHG	centroid
POLYGON ((1762636 1159366, ...	0.5440801	0.3856451	-15.843497	POINT (1766425 1274558)
MULTIPOLYGON (((1581926 200...	0.6568926	0.5612963	-9.559625	POINT (1534845 2129580)
POLYGON ((1300275 1729516, ...	0.4876192	0.3668790	-12.074018	POINT (1231626 1809939)
MULTIPOLYGON (((1472111 209...	0.6260897	0.5456649	-8.042480	POINT (1403745 2207477)
MULTIPOLYGON (((2259109 138...	0.4909352	0.4426662	-4.826904	POINT (2227142 1456772)
POLYGON ((1266358 1323800, ...	0.5148543	0.3840999	-13.075443	POINT (1247563 1428869)
POLYGON ((1060611 1924297, ...	0.4669634	0.3527934	-11.417005	POINT (1102959 1973459)
MULTIPOLYGON (((2195527 125...	0.4414093	0.3855412	-5.586805	POINT (2183202 1321281)

Figure 3.87 Data with centroid column added.

Wisconsin.R x county2 x county x state x smlst x							
Filter							
eometry	PROPDEM12	PROPDEM16	PptDEMCHG	centroid	long	lat	
OLYGON ((1762636 1159366, ...	0.5440801	0.3856451	-15.843497	POINT (1766425 1274558)	1766425	1274557.9	
MULTIPOLYGON (((1581926 200...	0.6568926	0.5612963	-9.559625	POINT (1534845 2129580)	1534845	2129580.5	
OLYGON ((1300275 1729516, ...	0.4876192	0.3668790	-12.074018	POINT (1231626 1809939)	1231626	1809938.5	
MULTIPOLYGON (((1472111 209...	0.6260897	0.5456649	-8.042480	POINT (1403745 2207477)	1403745	2207476.6	
MULTIPOLYGON (((2259109 138...	0.4909352	0.4426662	-4.826904	POINT (2227142 1456772)	2227142	1456771.6	
OLYGON ((1266358 1323800, ...	0.5148543	0.3840999	-13.075443	POINT (1247563 1428869)	1247563	1428868.7	
OLYGON ((1060611 1924297, ...	0.4669634	0.3527934	-11.417005	POINT (1102959 1973459)	1102959	1973459.0	
MULTIPOLYGON (((2195527 125...	0.4414093	0.3855412	-5.586805	POINT (2183202 1321281)	2183202	1321281.3	

Showing 1 to 8 of 73 entries. 14 total columns

Figure 3.88 Latitude and longitude columns.

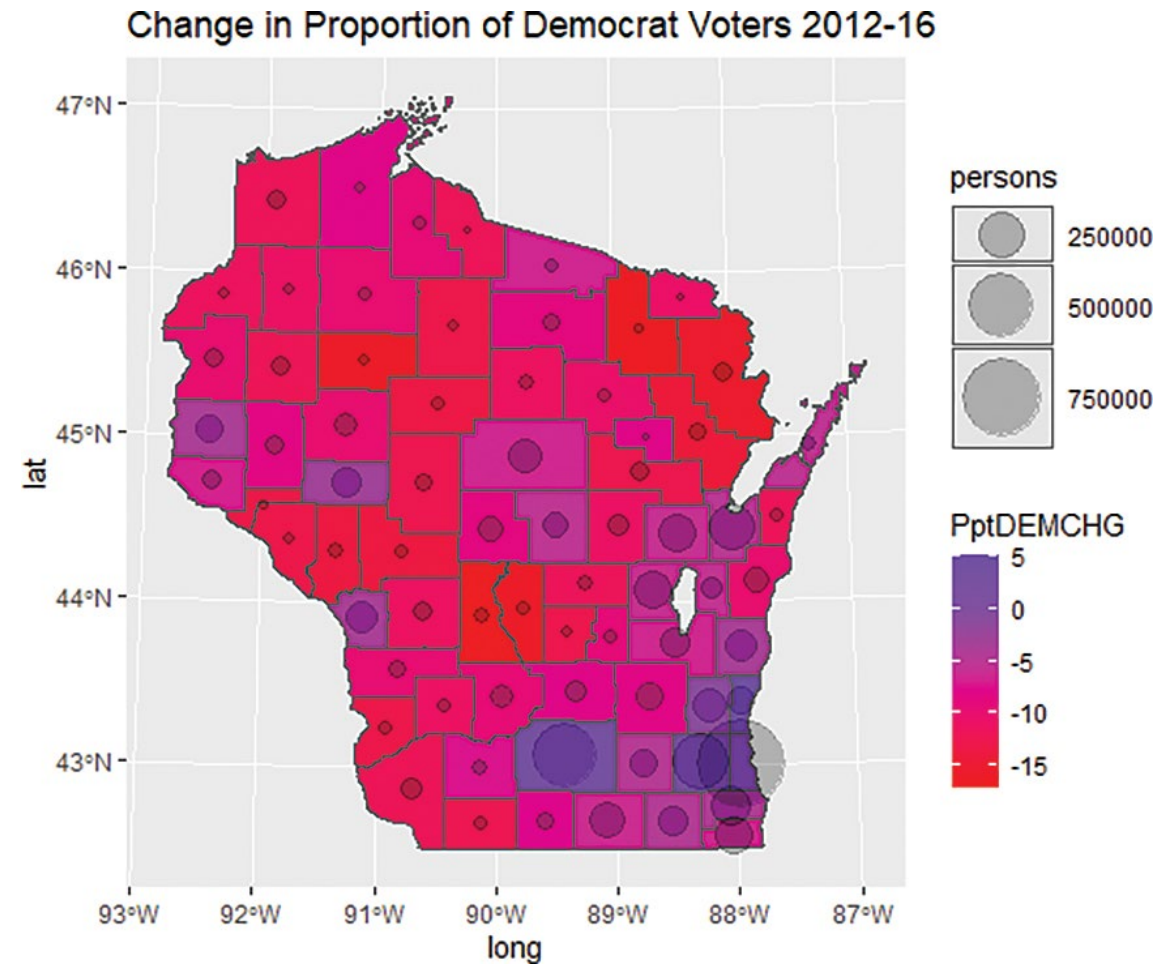


Figure 3.89 Map with bubbles showing county population size.

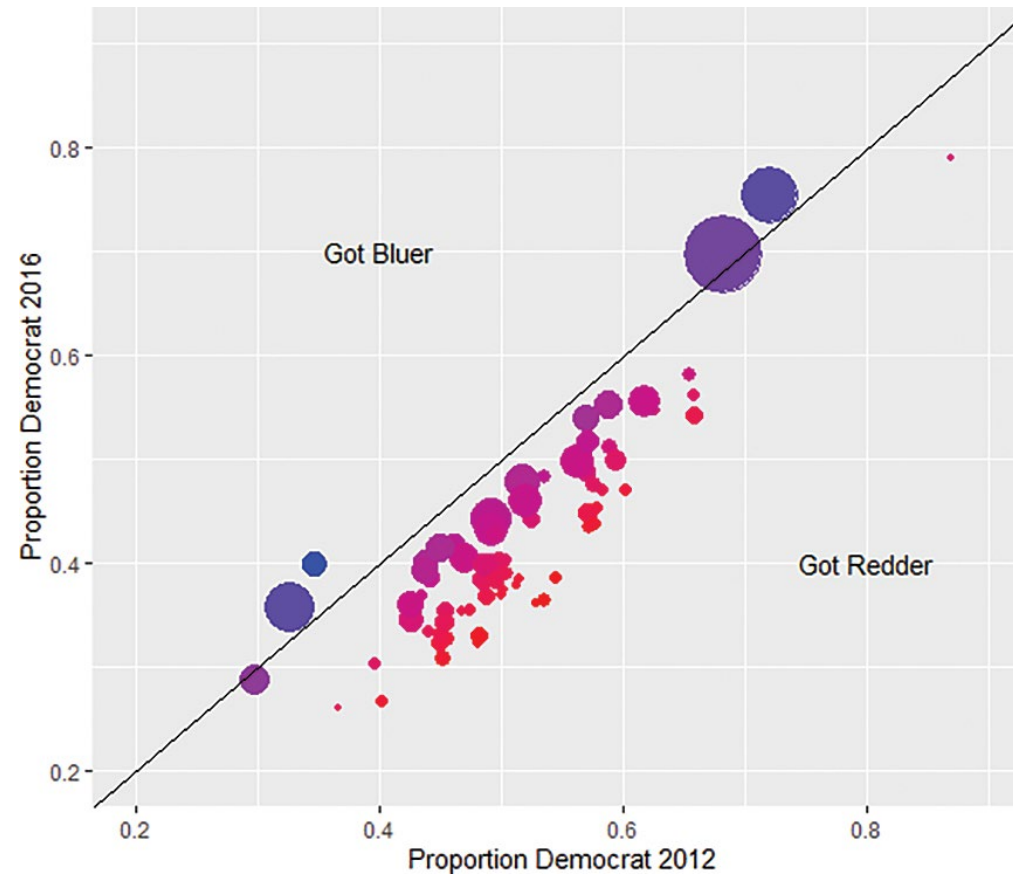


Figure 3.90 R scatterplot showing the proportion of Democratic votes by county in 2012 and 2016.

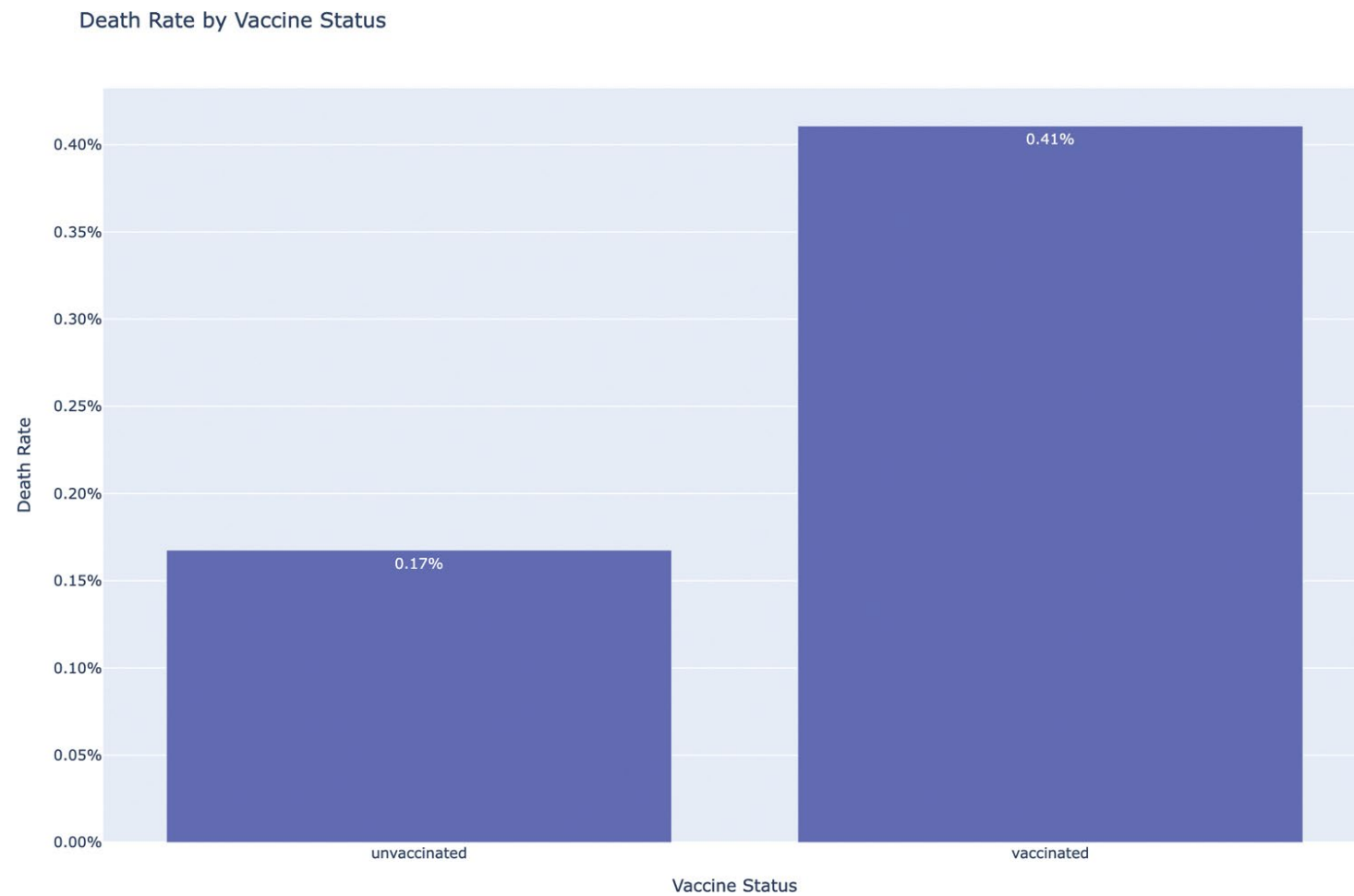


Figure 4.1 COVID-19 death rates by vaccination status.

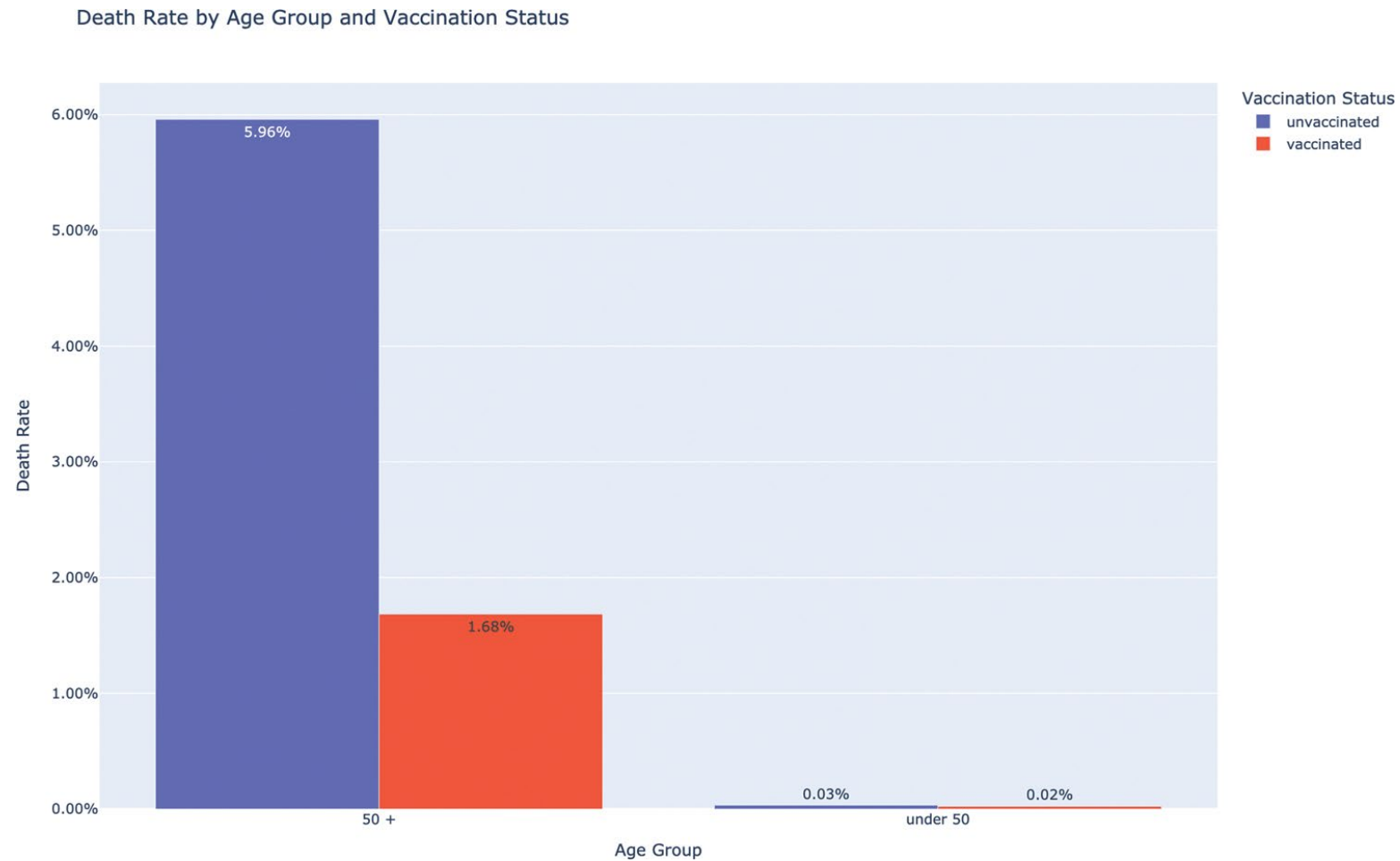


Figure 4.2 COVID-19 death rates by vaccination status and age group.

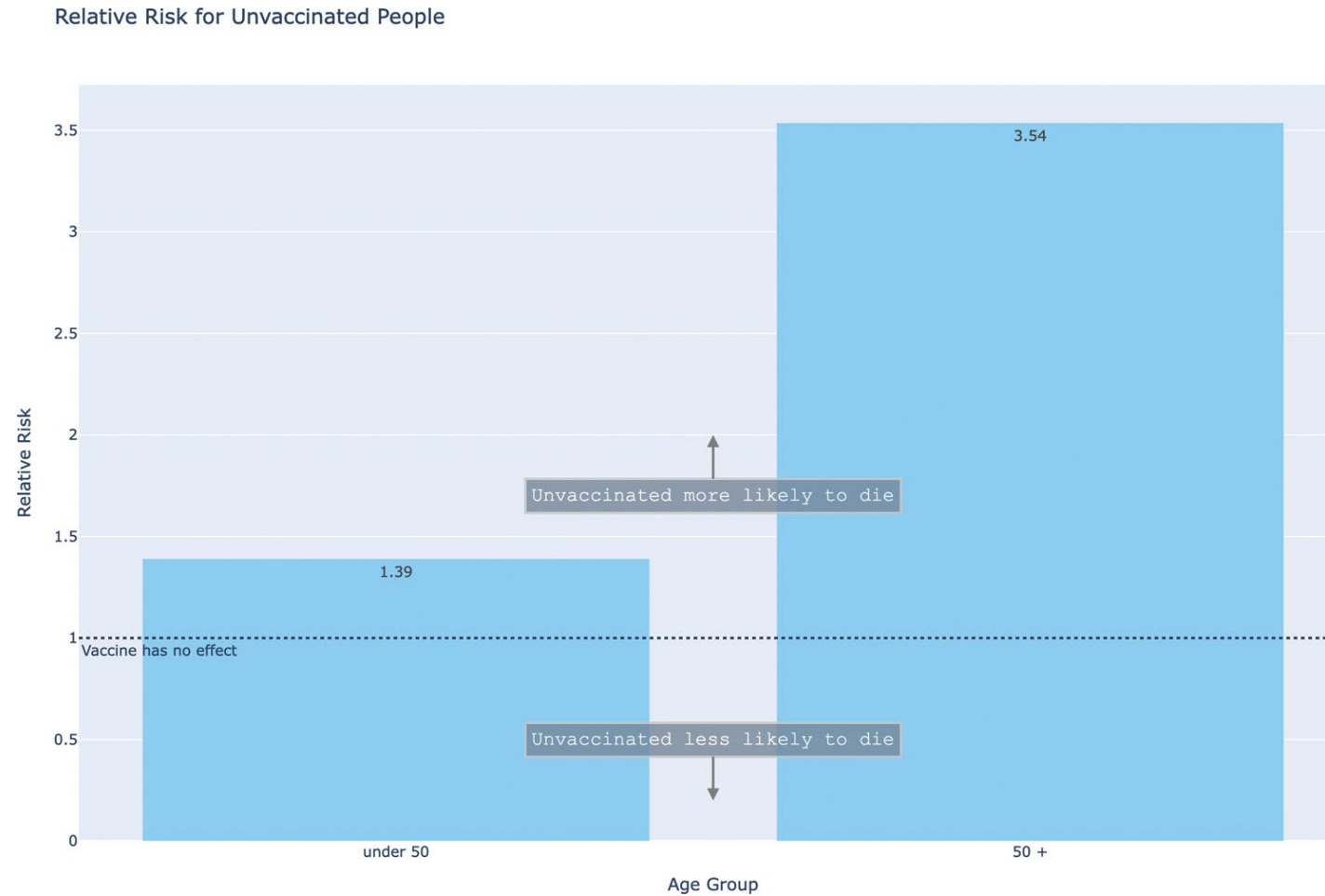


Figure 4.3 Relative risk of dying for unvaccinated people from COVID-19.

	A	B	C	D
1	age_group	vaccine_status	outcome	
2	under 50	vaccinated	death	
3	under 50	vaccinated	death	
4	under 50	vaccinated	death	
5	under 50	vaccinated	death	
6	under 50	vaccinated	death	
7	under 50	vaccinated	death	
8	under 50	vaccinated	death	
9	under 50	vaccinated	death	
10	under 50	vaccinated	death	

Figure 4.4 COVID-19 data showing three categorical or qualitative variables.

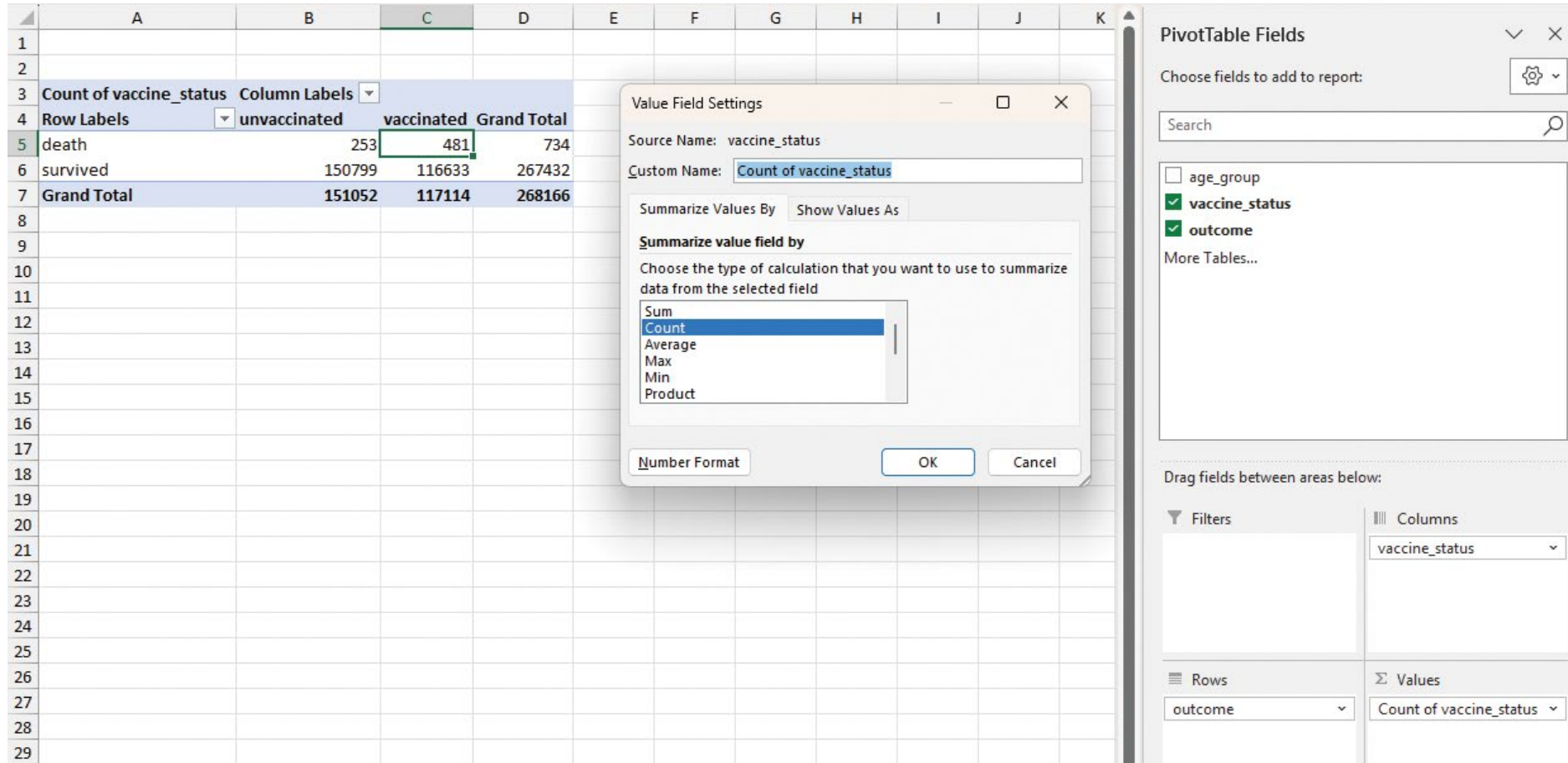


Figure 4.5 PivotTable counting outcomes by vaccine_status.

Value Field Settings

Source Name: vaccine_status

Custom Name: Count of vaccine_status

Summarize Values By Show Values As

Show values as

% of Column Total

Base field:

age_group
vaccine_status
outcome

Base item:

Number Format OK Cancel

Count of vaccine_status			
Column Labels			
Row Labels	unvaccinated	vaccinated	Grand Total
death	0.17%	0.41%	0.27%
survived	99.83%	99.59%	99.73%
Grand Total	100.00%	100.00%	100.00%

Figure 4.6 Formatting PivotTable counts of outcomes by vaccine_status as percentages of the column total.

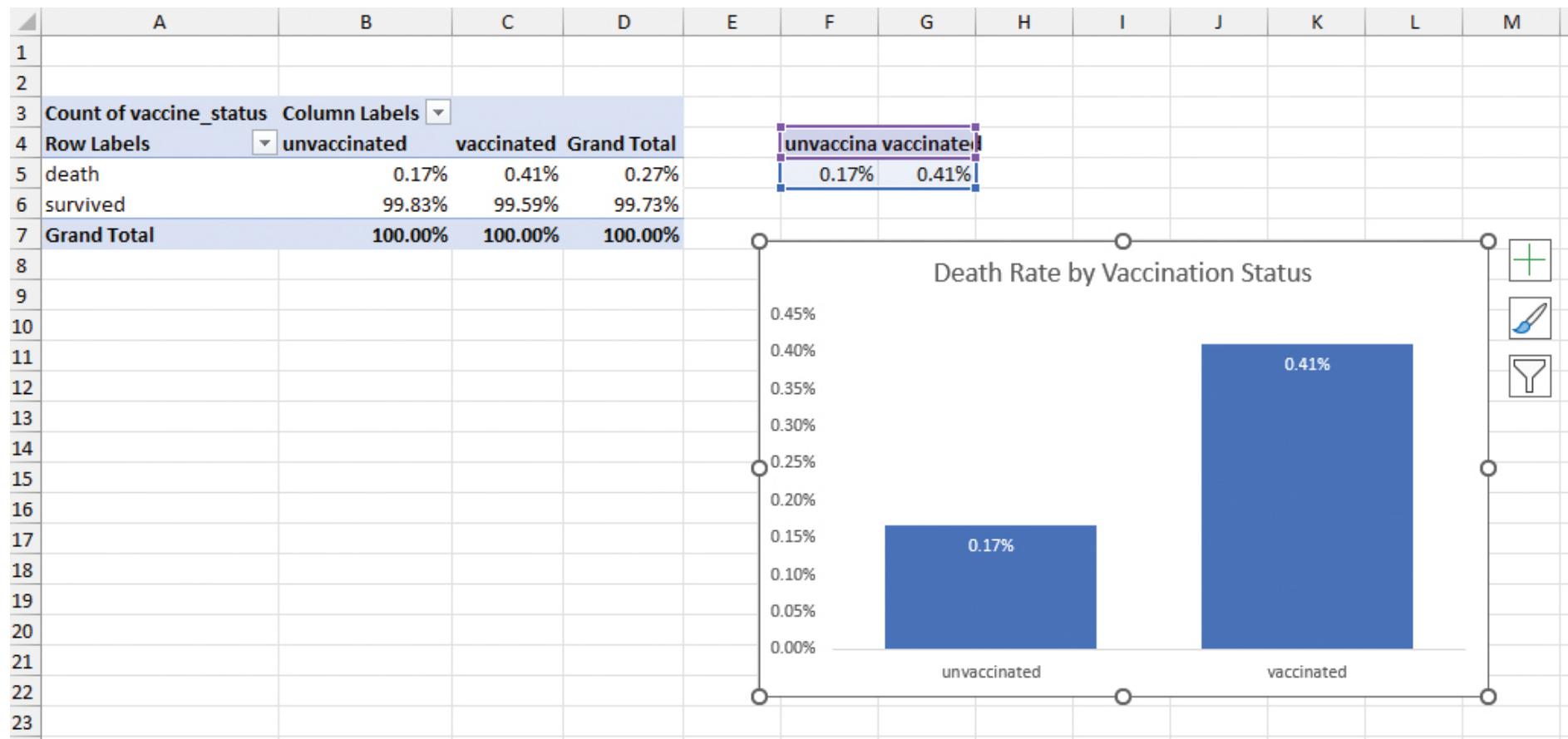


Figure 4.7 Creating the first target graphic of death rates by vaccination status.

Count of outcome		Column Labels	
Row Labels		unvaccinated	vaccinated
50 +		3440	27307
death		205	460
survived		3235	26847
under 50		147612	89807
death		48	21
survived		147564	89786
Grand Total		151052	117114

Count of outcome		Column Labels	
Row Labels		unvaccinated	vaccinated
50 +		2.28%	23.32%
death		5.96%	1.68%
survived		94.04%	98.32%
under 50		97.72%	76.68%
death		0.03%	0.02%
survived		99.97%	99.98%
Grand Total		100.00%	100.00%

Figure 4.8 Formatting PivotTable counts of outcomes by age and vaccine_status as percentages of the parent row total.

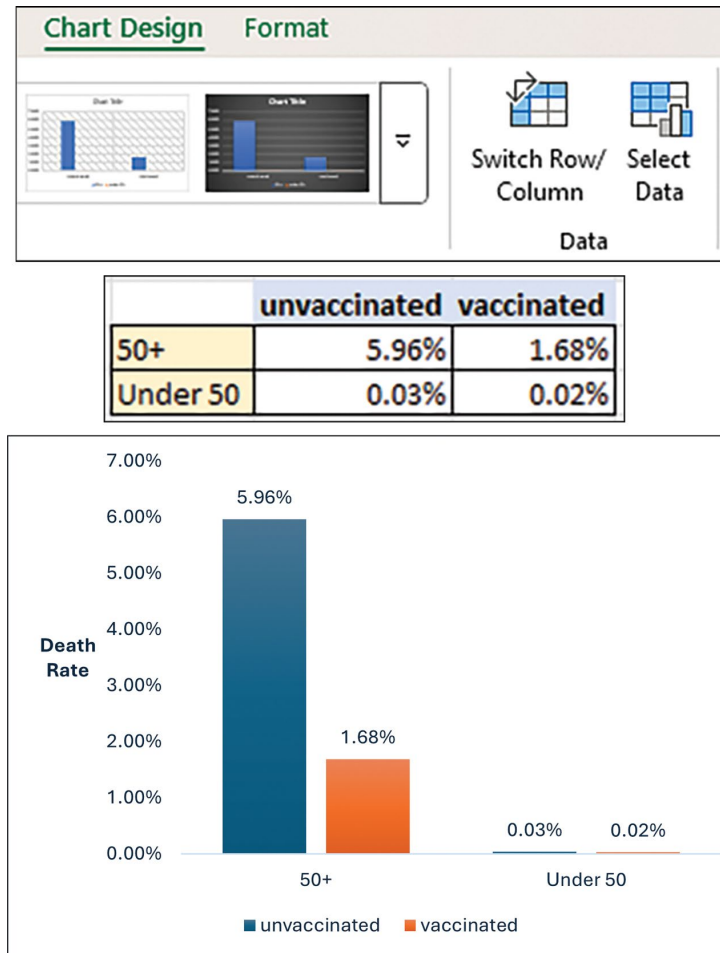


Figure 4.9 Creating the second target graphic showing the death rates by vaccination status for each age group.

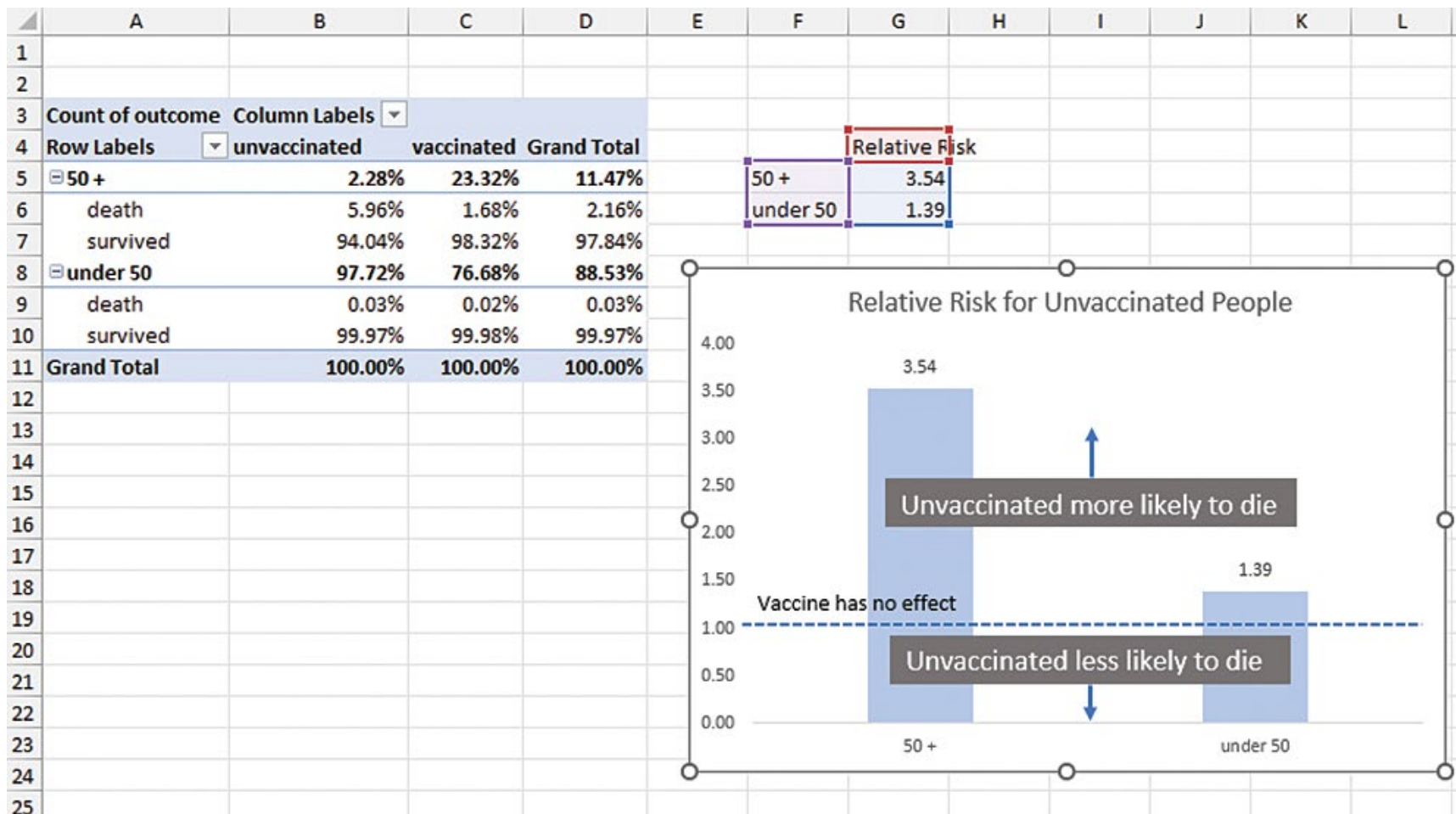


Figure 4.10 Creating the third target graphic showing the relative risk for unvaccinated people by age group.

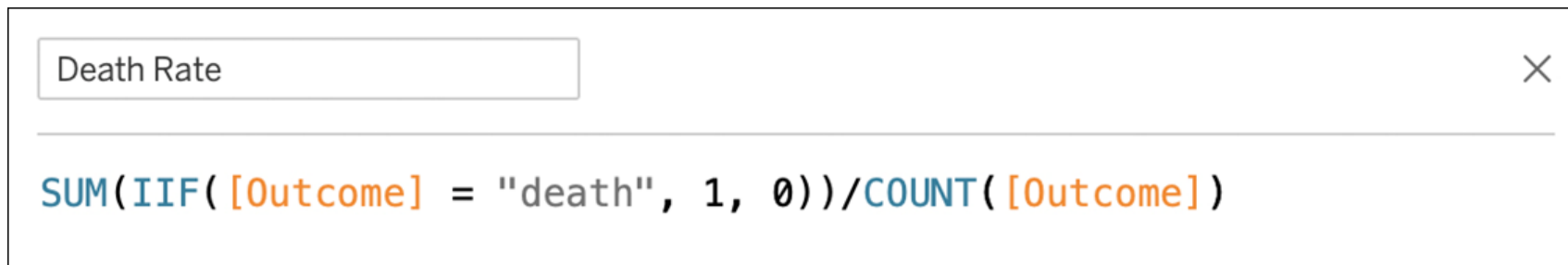


Figure 4.11 Death rate formula.

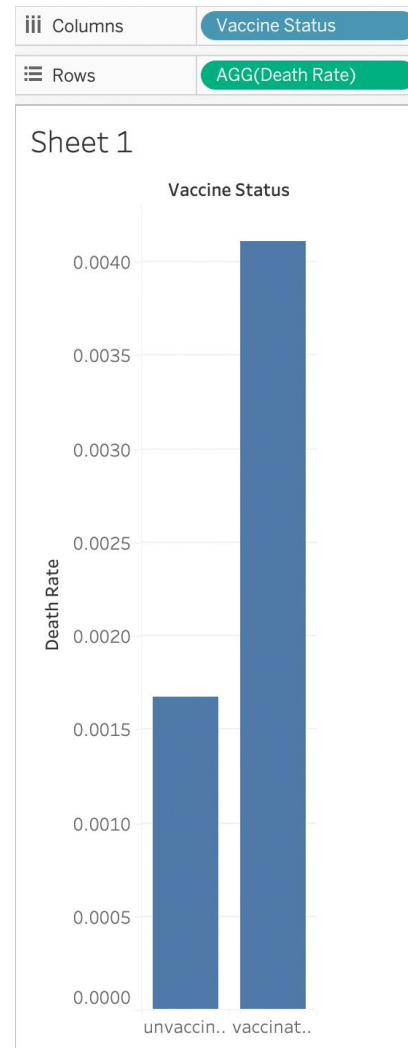


Figure 4.12 Initial bar chart.

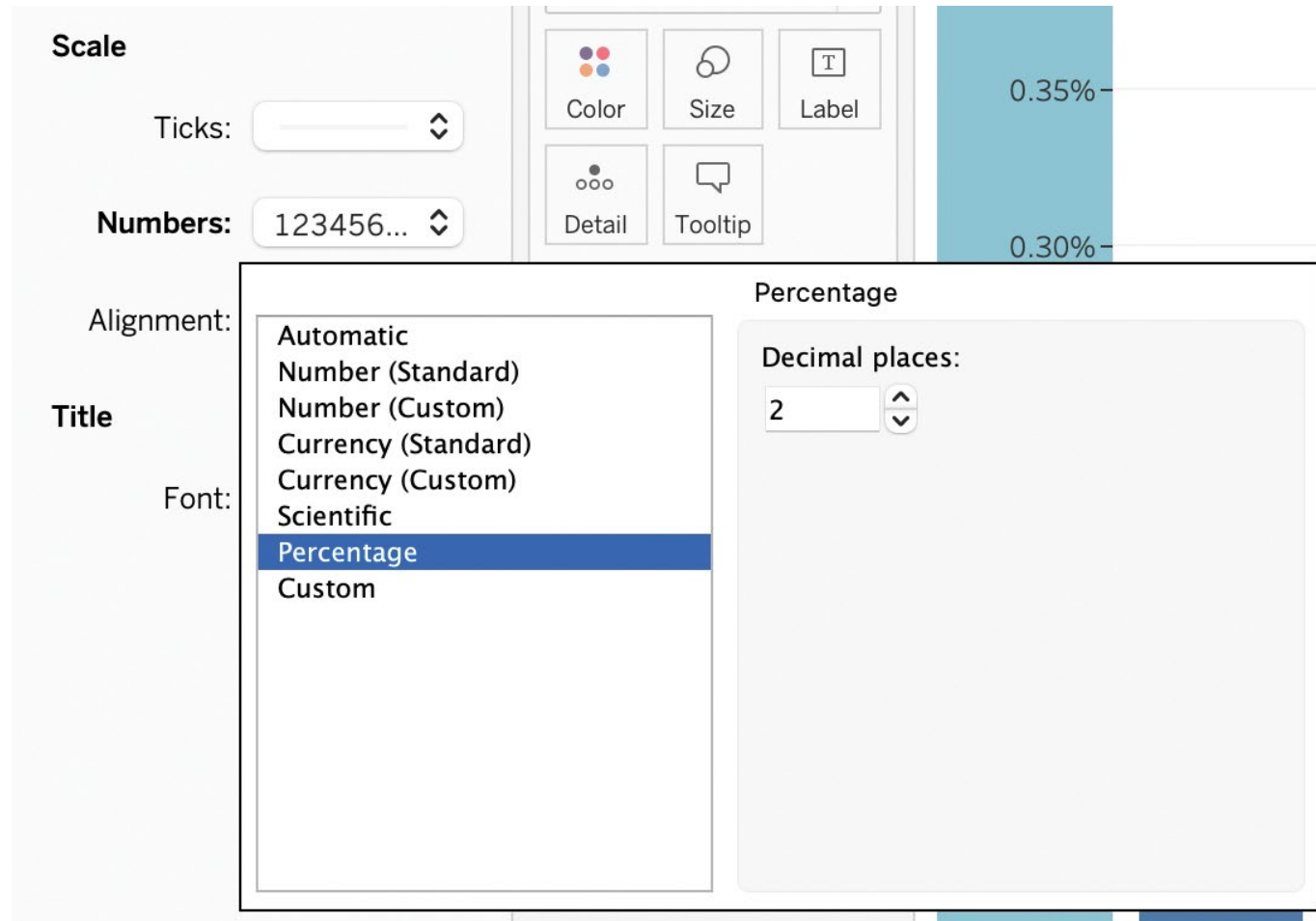


Figure 4.13 Formatting the vertical axis.

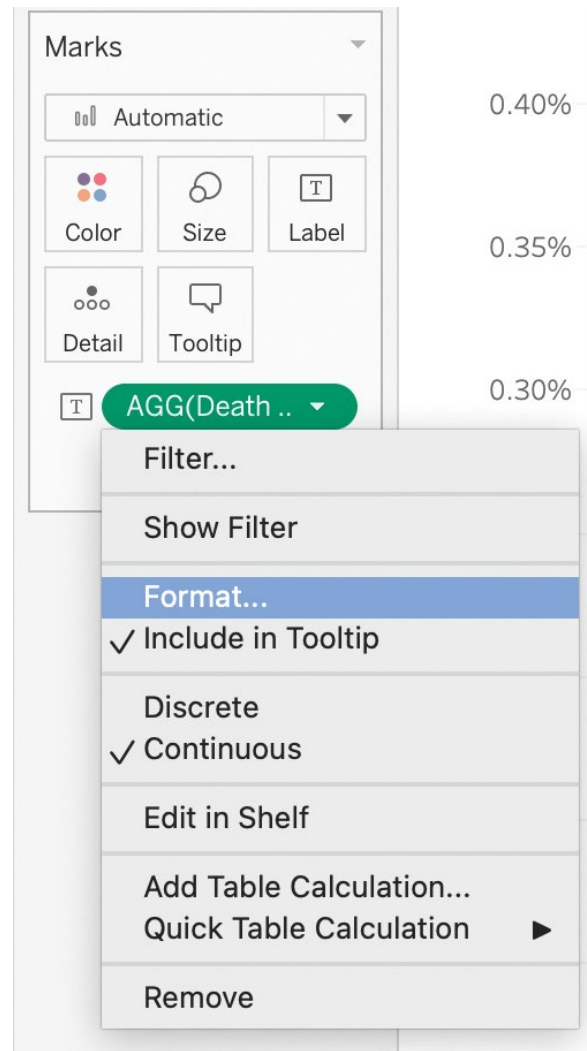


Figure 4.14 Formatting the bar labels.

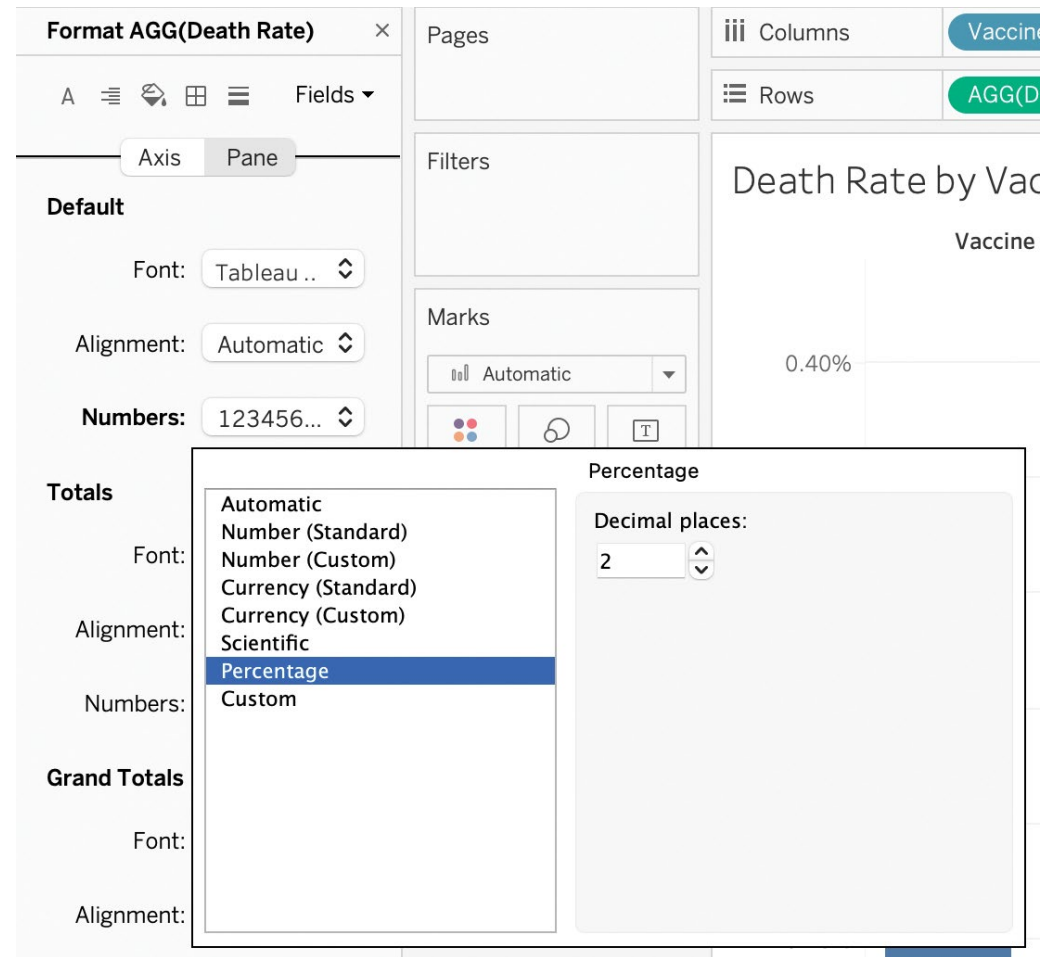


Figure 4.15 Specifying the number format.

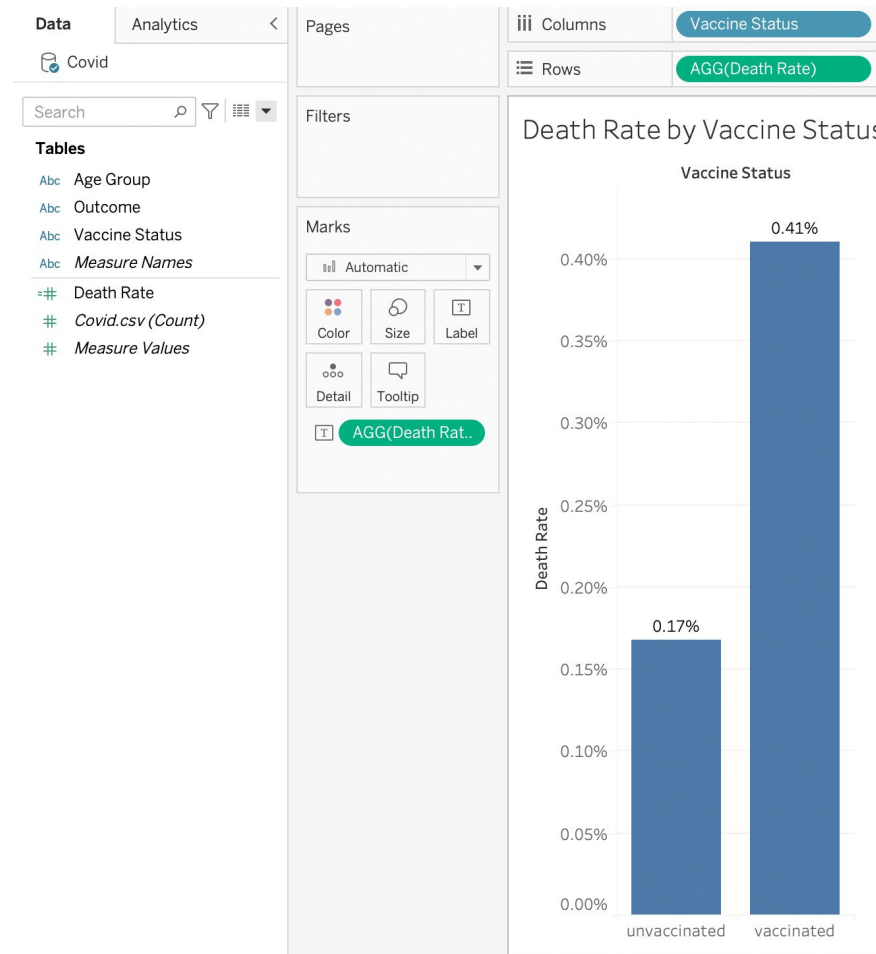


Figure 4.16 Bar chart by vaccine status.

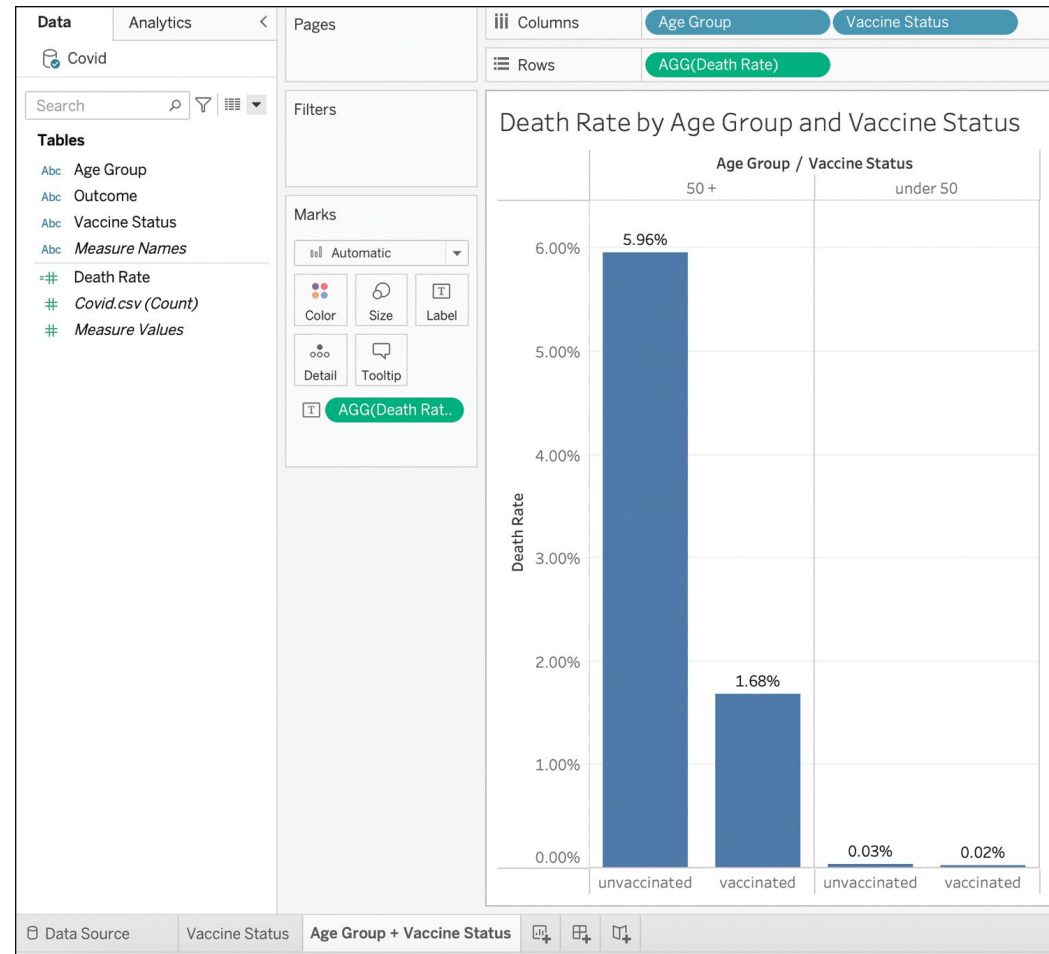


Figure 4.17 Bar chart by age group and vaccine status.

Risk if Vaccinated

×

SUM(IIF([Outcome] = "death" and [Vaccine Status] = "vaccinated", 1, 0))/
SUM(IIF([Vaccine Status] = "vaccinated", 1, 0))

▶

The calculation is valid.

2 Dependencies ▼

Apply

OK

Figure 4.18 Risk If Vaccinated calculation.

Relative Risk

×

[Risk if Unvaccinated]/[Risk if Vaccinated]

The calculation is valid.

1 Dependency ▾

Apply

OK

Figure 4.19 Relative risk calculation.

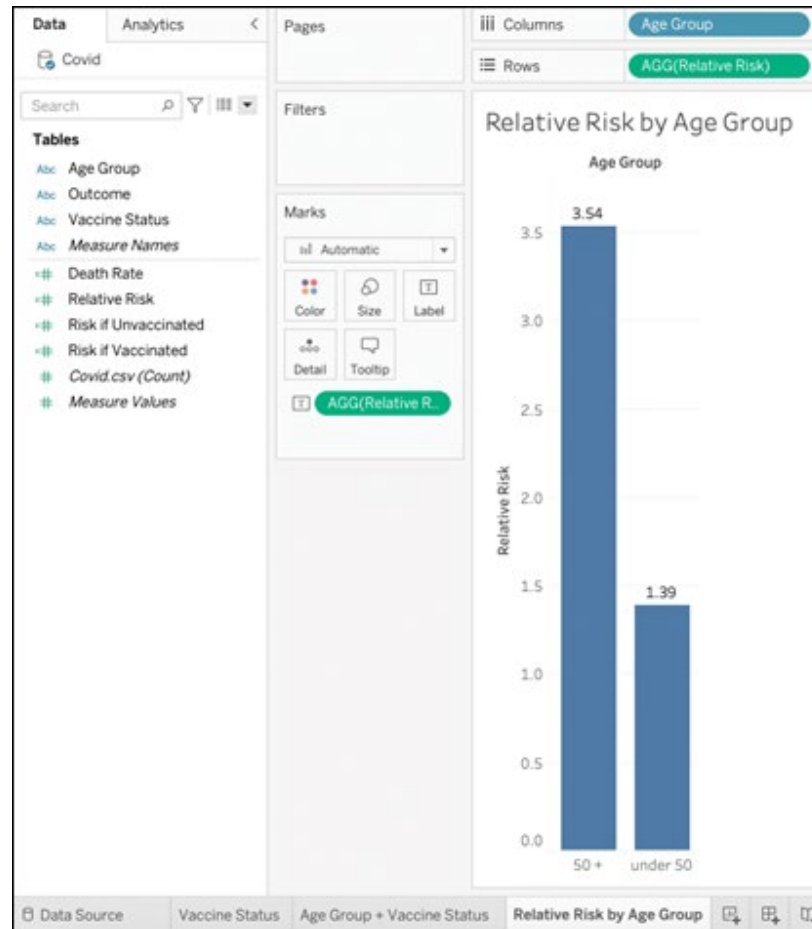


Figure 4.20 Relative risk bar chart.

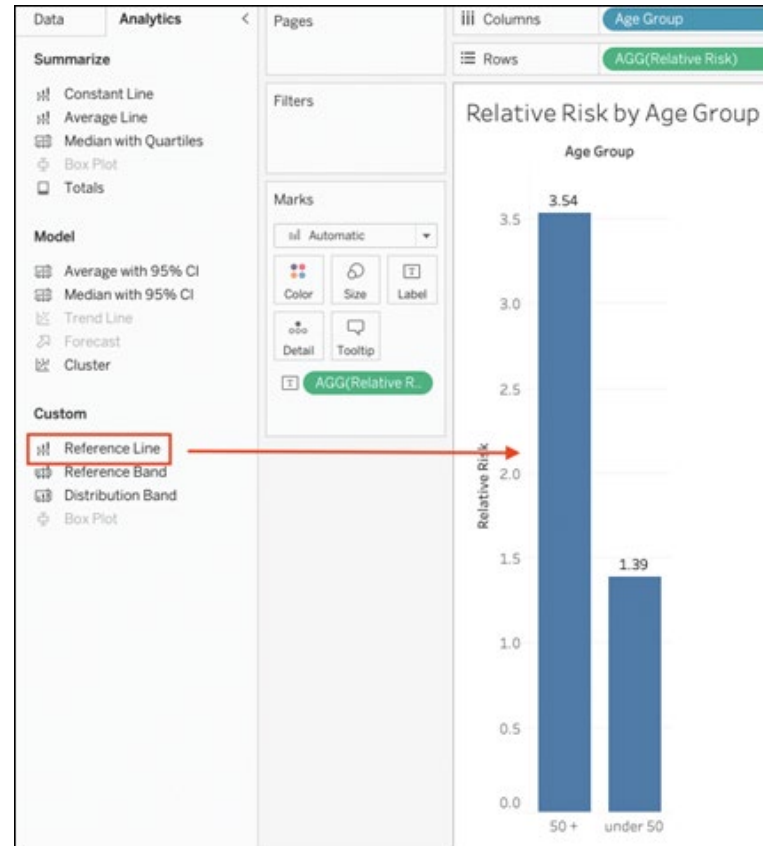


Figure 4.21 Adding a reference line.

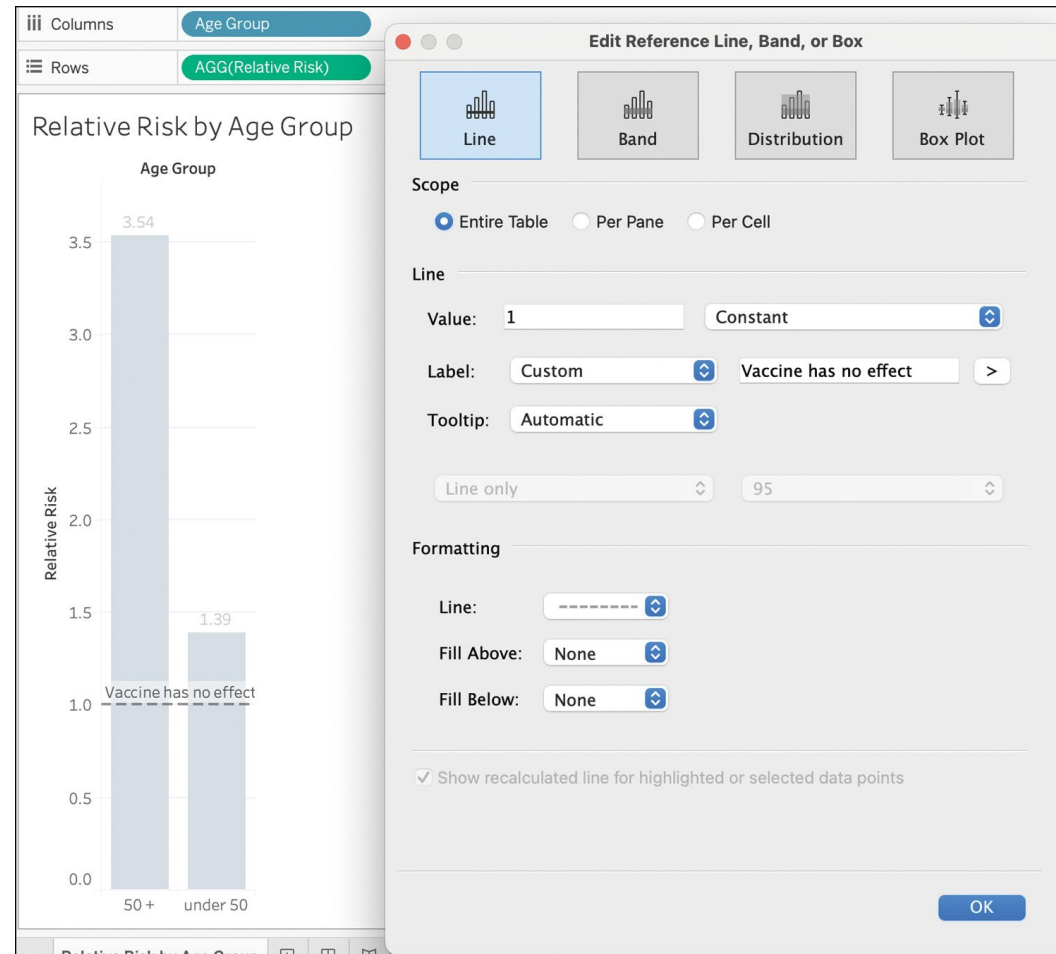


Figure 4.22 Reference line formatting.

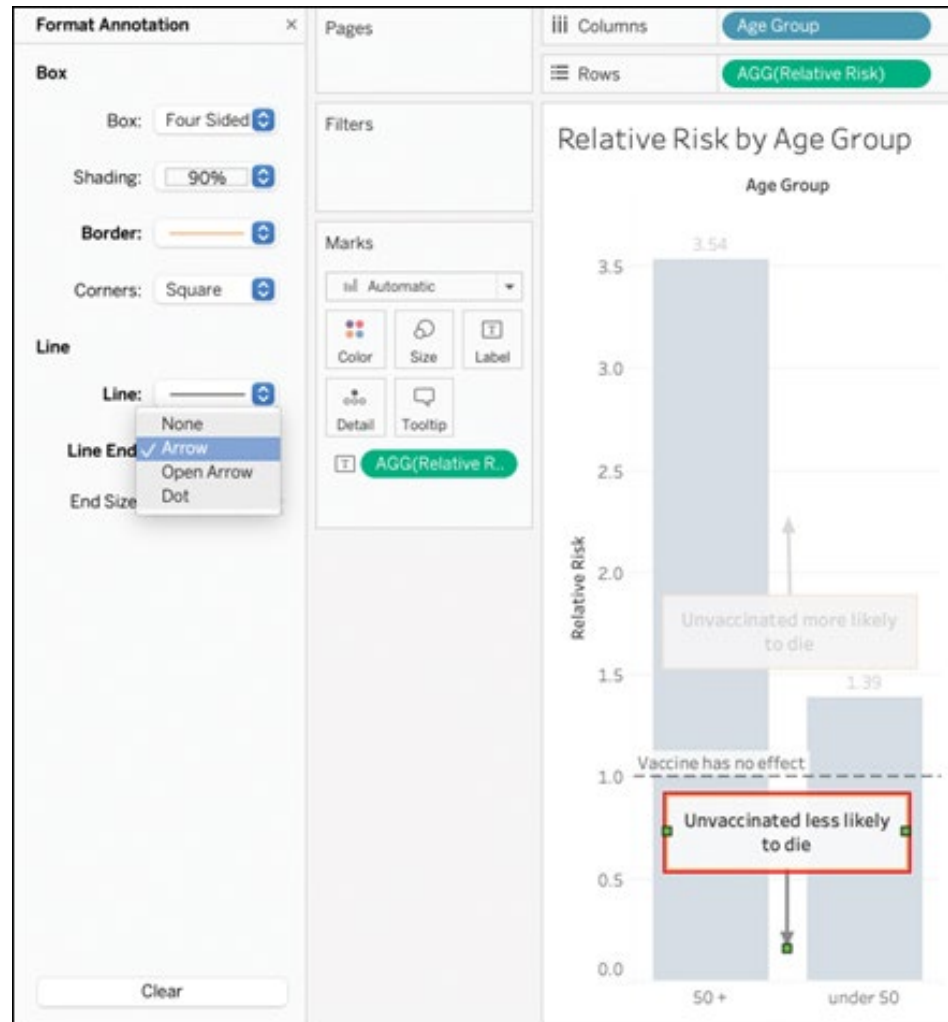


Figure 4.23 Adding text annotations with arrows.

Relative Risk by Age Group

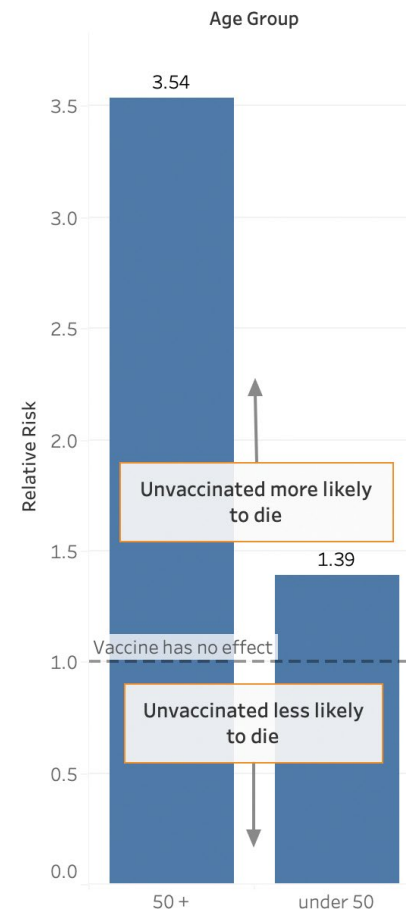


Figure 4.24 Final relative risk bar chart.

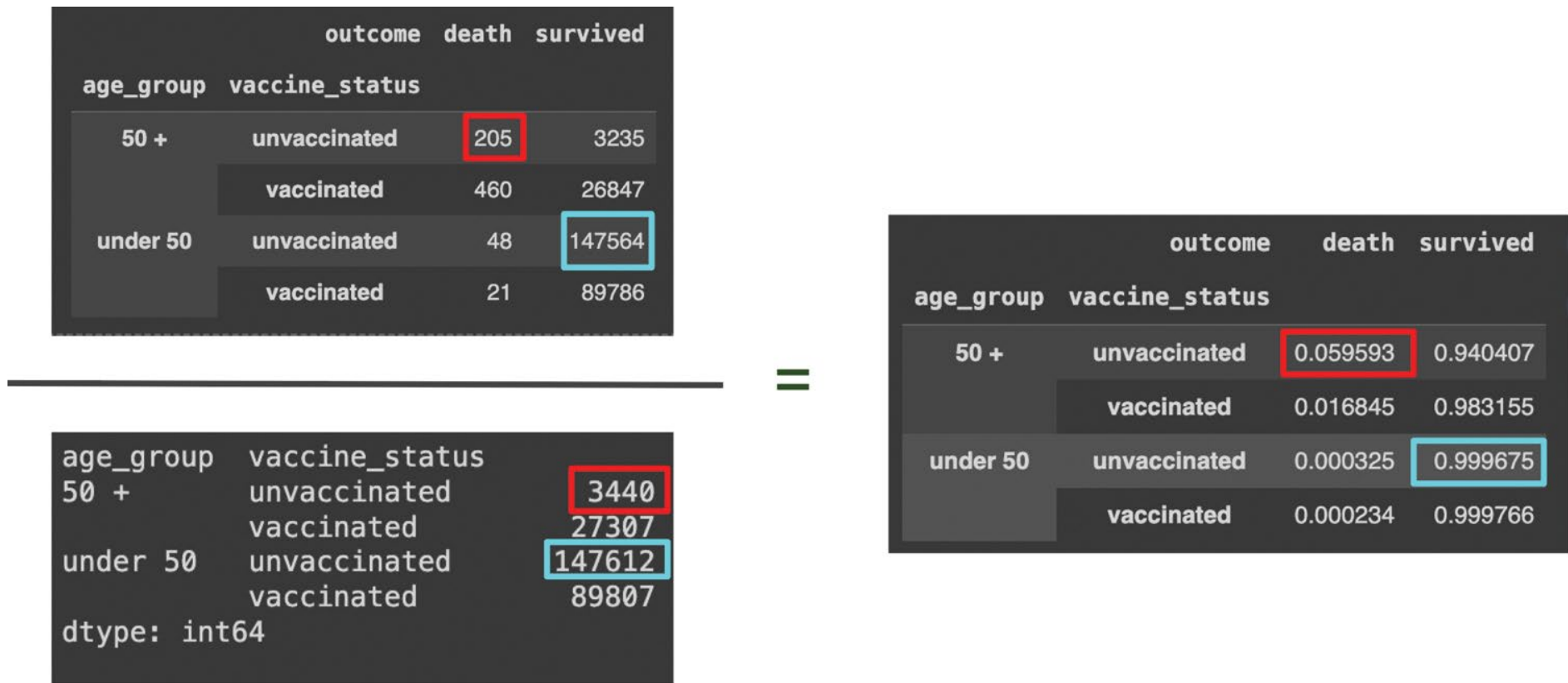


Figure 4.25 COVID-19 division diagram.

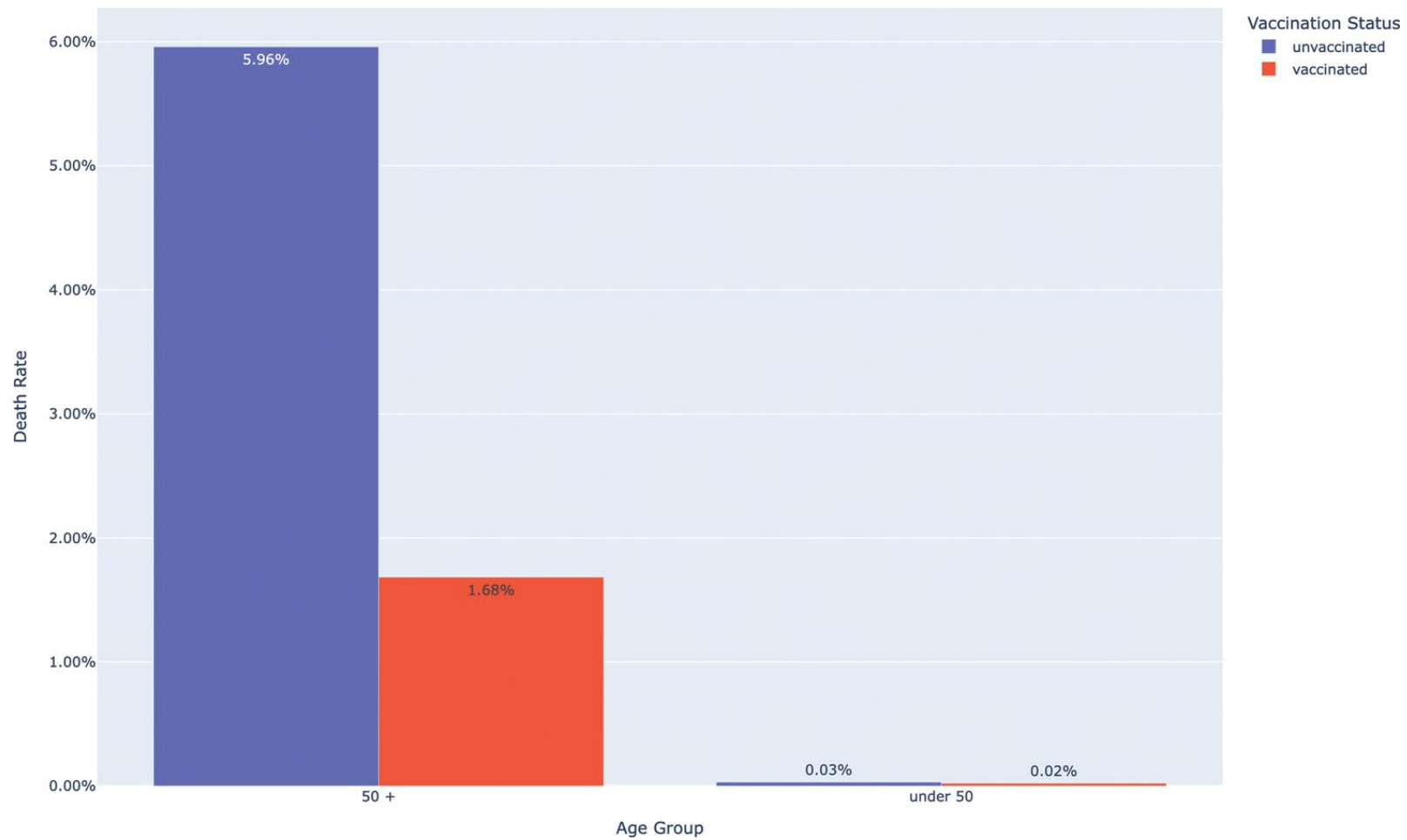


Figure 4.26 Absolute risk by age and vaccine status.

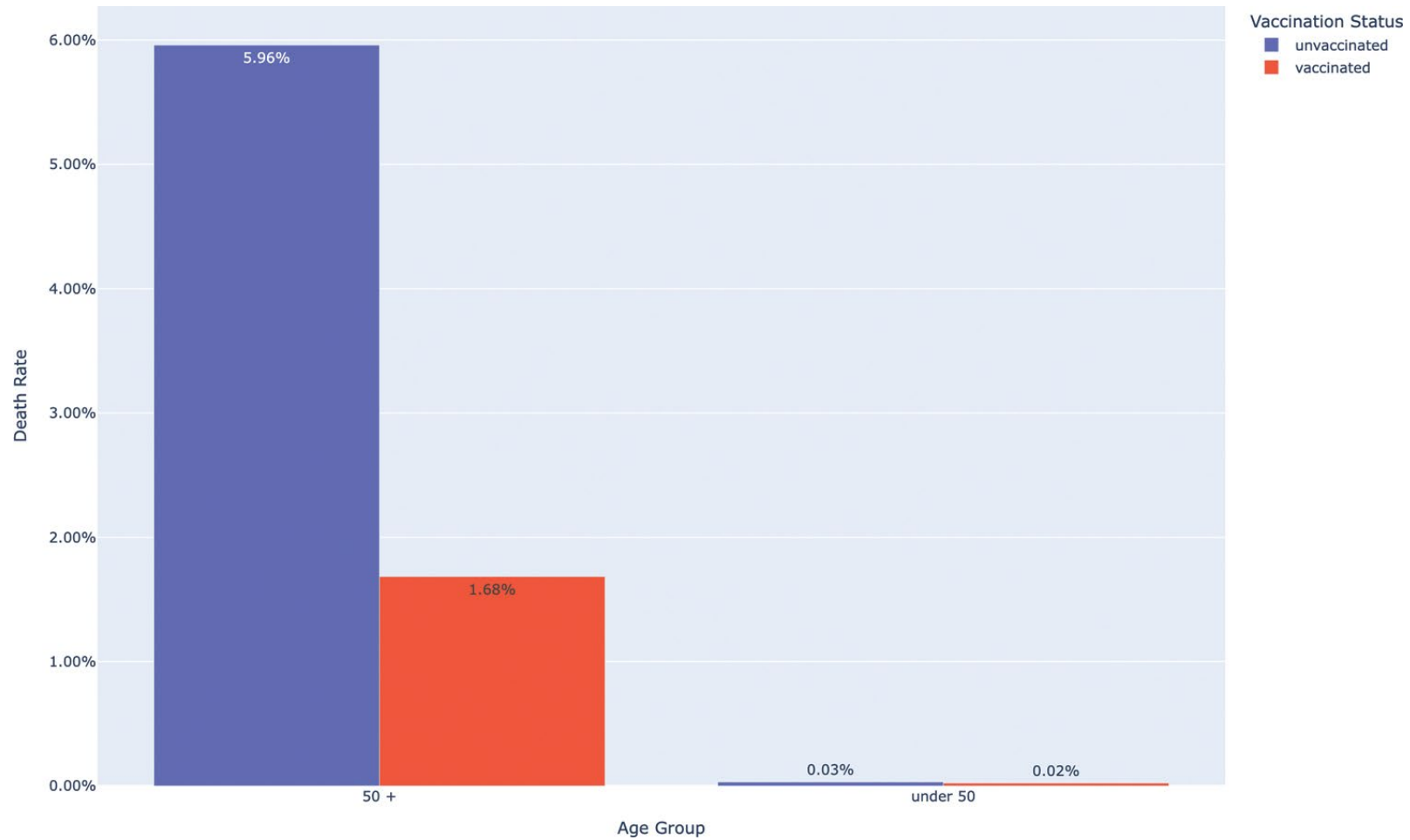


Figure 4.27 Absolute risk by age and vaccine status—long version.

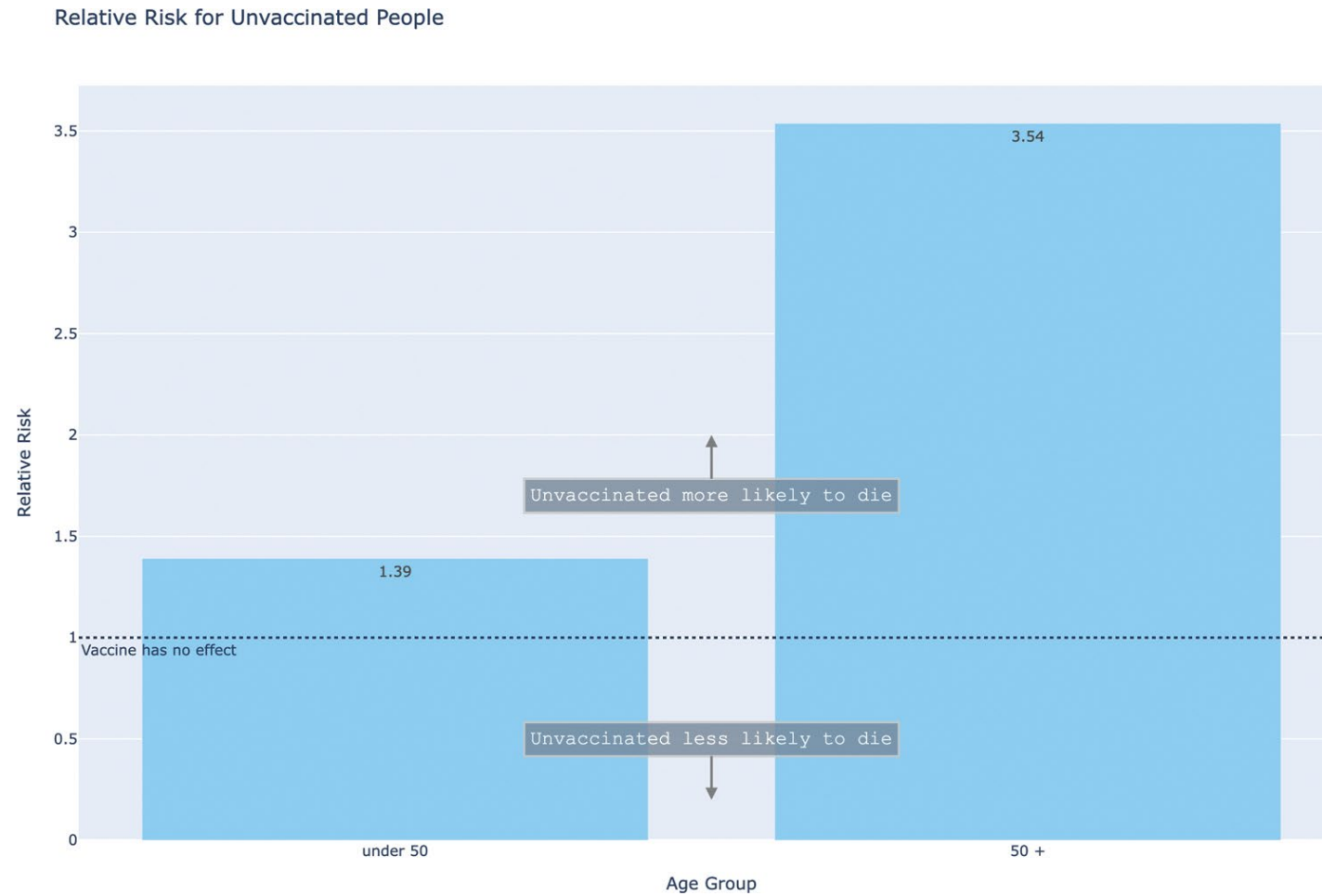


Figure 4.28 Relative risk.

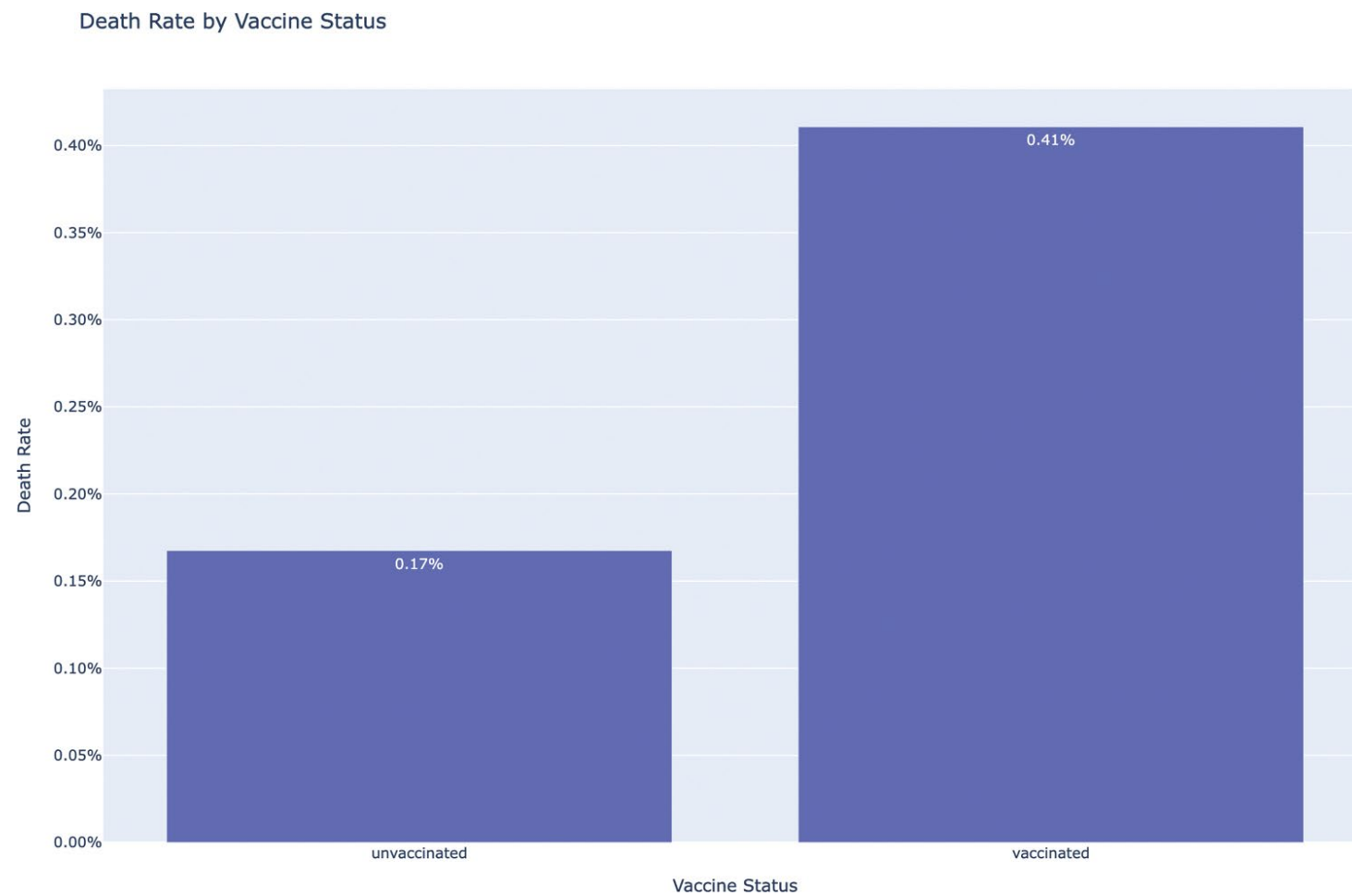


Figure 4.29 COVID-19 death rate by vaccine status.

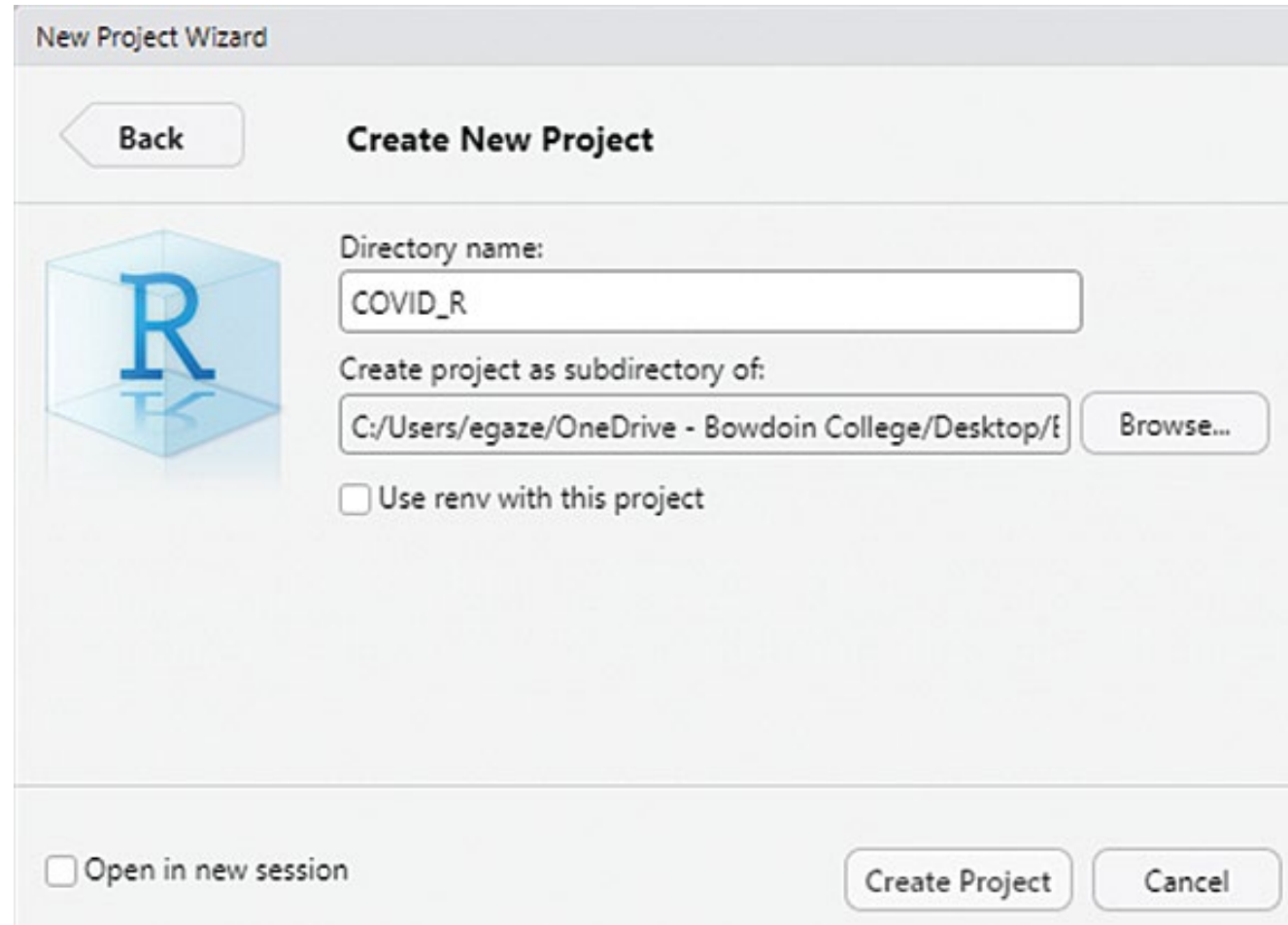


Figure 4.30 Creating a new project in RStudio.

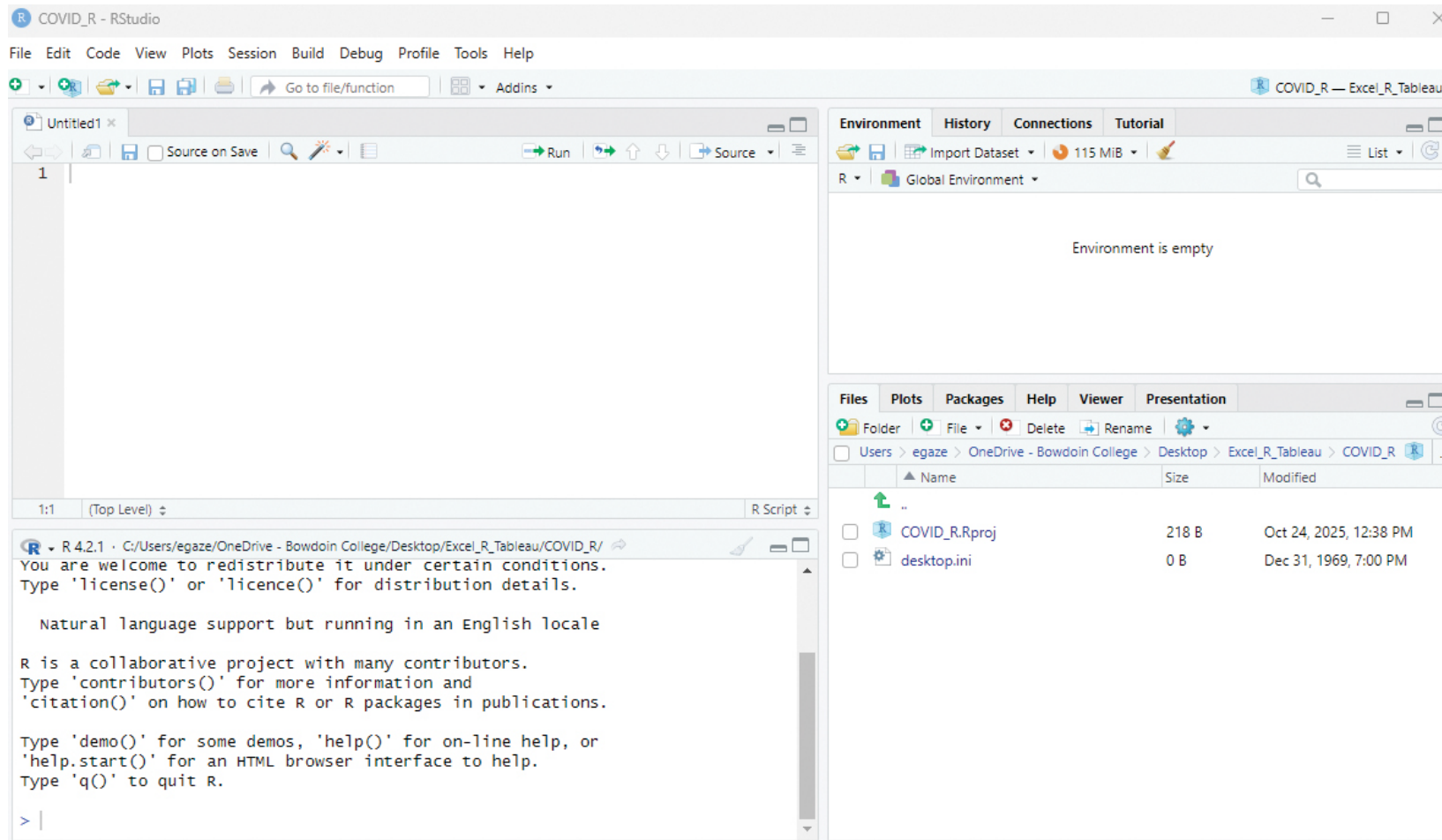


Figure 4.31 RStudio environment.

 covid_deaths_by_age_and vaccination		8/24/2023 3:30 PM	Microsoft Excel C...	7,768 KB
 COVID_project.Rproj		5/13/2025 10:22 AM	RPROJ File	1 KB

Figure 4.32 RStudio project folder with files.

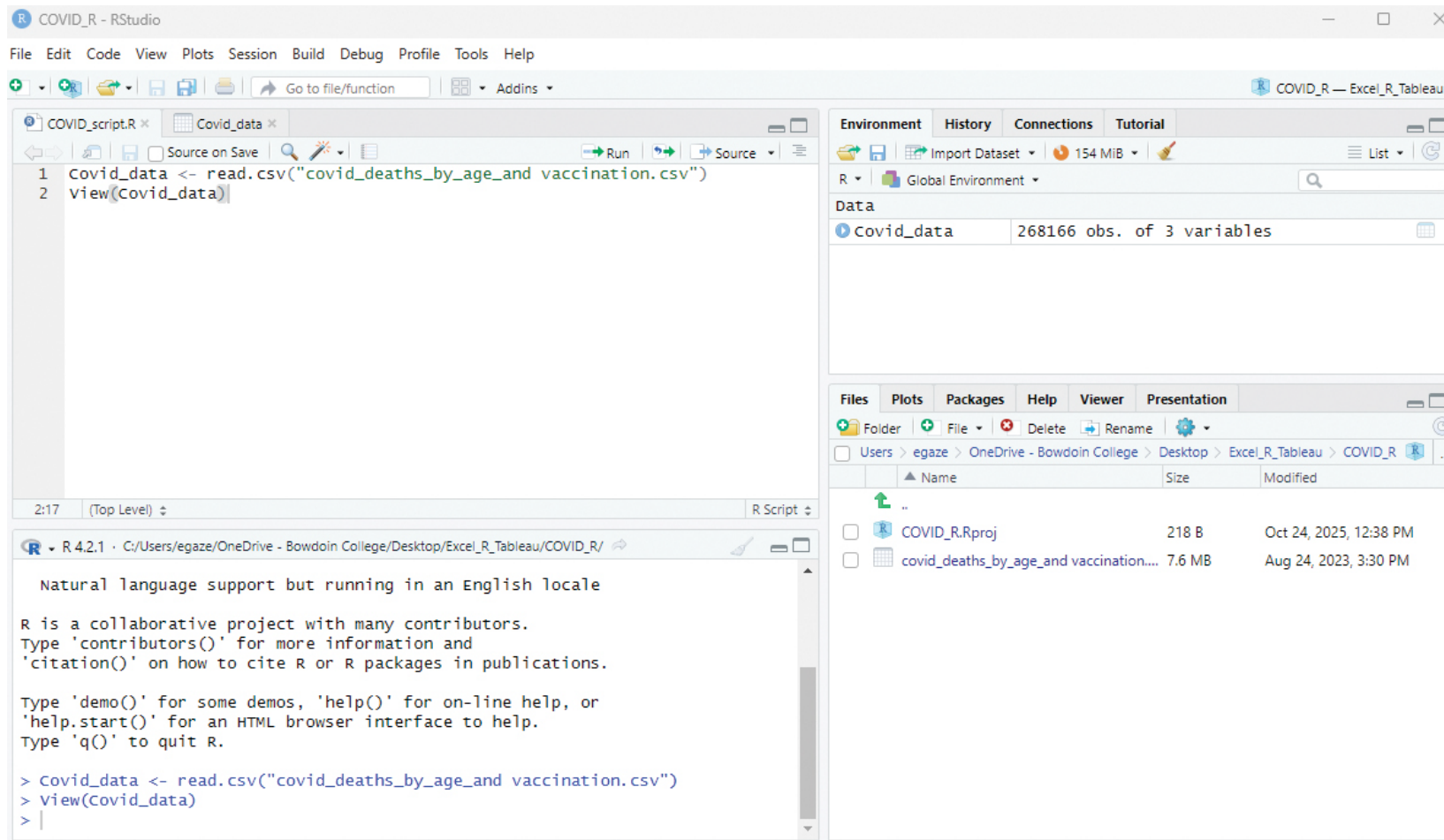


Figure 4.33 RStudio environment with the script started, code run in the console, files in the project folder, and the first data object created.

Covid_script.R ×		Covid_data ×	
← →		Filter	
	age_group	vaccine_status	outcome
1	under 50	vaccinated	death
2	under 50	vaccinated	death
3	under 50	vaccinated	death
4	under 50	vaccinated	death
...	...	...	...

Figure 4.34 COVID-19 data.

```
> unique(Covid_data)
```

	age_group	vaccine_status	outcome
1	under 50	vaccinated	death
22	under 50	unvaccinated	death
70	under 50	vaccinated	survived
89856	under 50	unvaccinated	survived
237420	50 +	vaccinated	death
237880	50 +	unvaccinated	death
238085	50 +	vaccinated	survived
264932	50 +	unvaccinated	survived

Figure 4.35 Results of the unique function.

Environment

History

Connections

Tutorial

Import Dataset

238 MiB

List

R

Global Environment

Data

Covid_data

268166 obs. of 3 variables

dfperc

2 obs. of 2 variables

nvax_d

253 obs. of 3 variables

pdf

2 obs. of 1 variable

perc_tab

num [1:2, 1:2] 0.17 99.83 0.41 99.59

vax_d

481 obs. of 3 variables

Figure 4.36 RStudio data environment with three data objects.

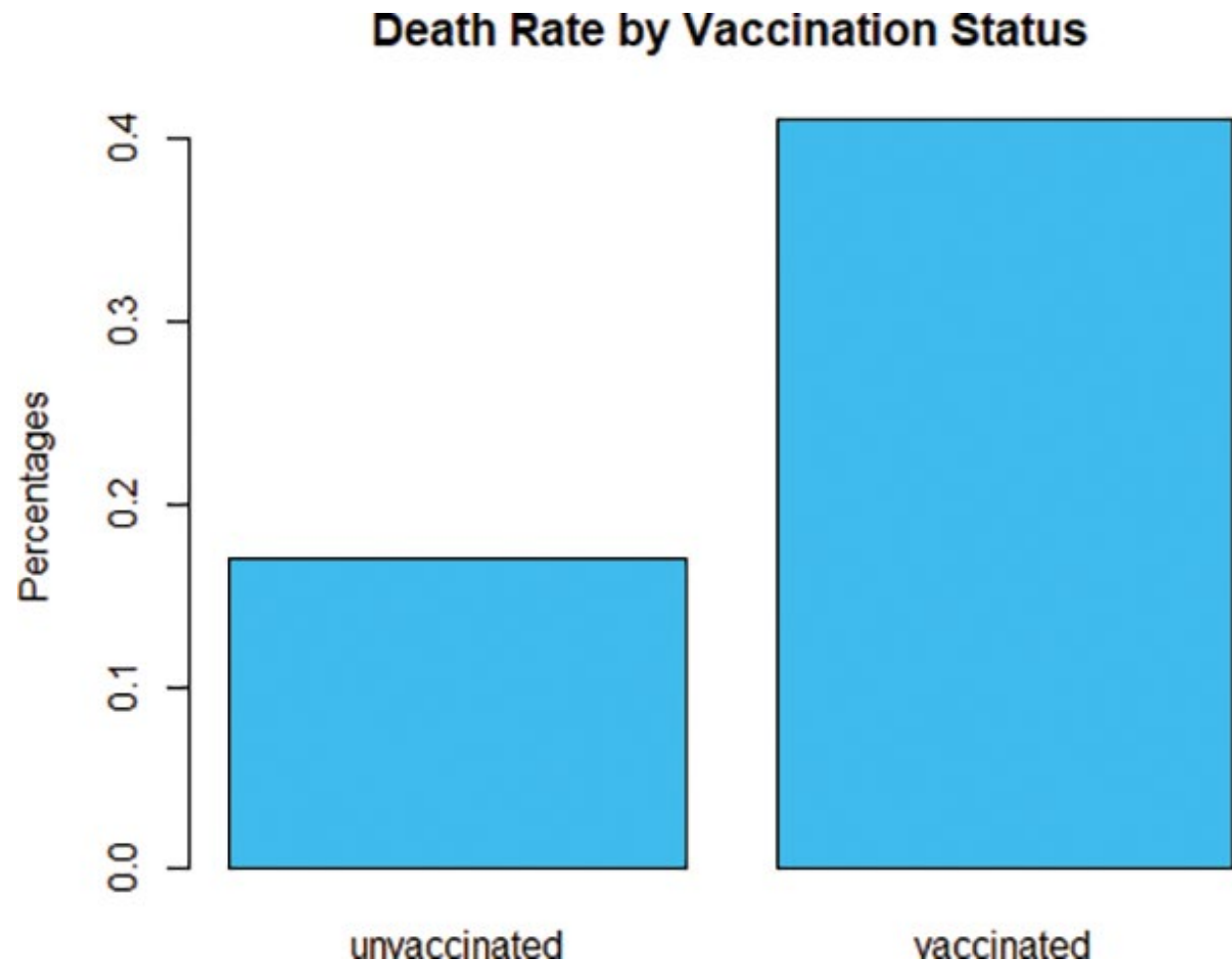


Figure 4.37 First target graphic without labels.

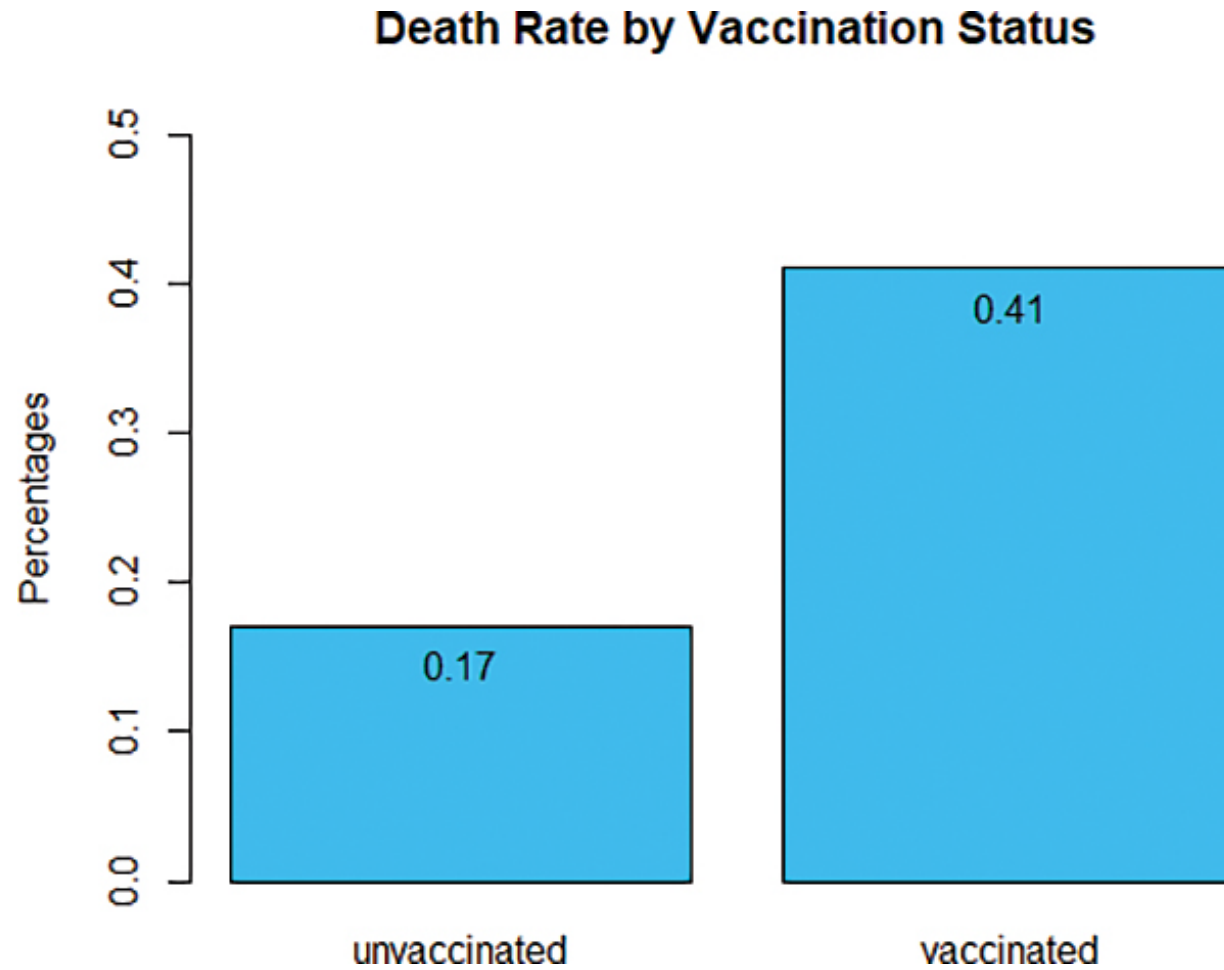


Figure 4.38 Completed first target graphic using the barplot function with labels created using the text function.


```
ggplot(data = dfperc, mapping = aes(x = rownames(dfperc), y = death.rate)) +  
  geom_bar(stat = "identity", fill = "blue") +  
  ylim(0, max(dfperc$death.rate)+0.1) +  
  xlab(NULL) + ylab("Percentage") + ggtitle("Death Rate by Vaccination Status") +  
  geom_text(aes(label = death.rate), vjust = 1.5, color = "white")
```

Figure 4.39 ggplot.

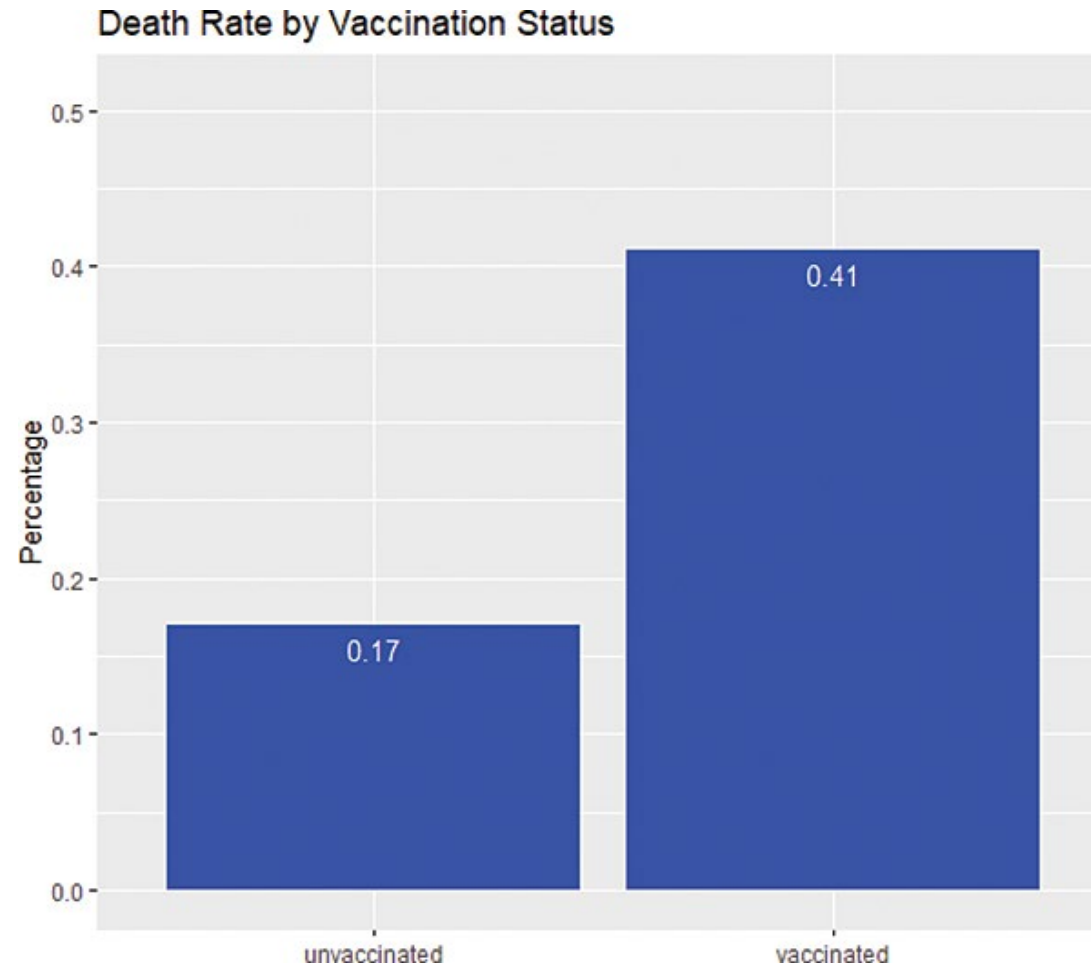


Figure 4.40 Completed first target graphic using the ggplot function with labels using the geom_text function.

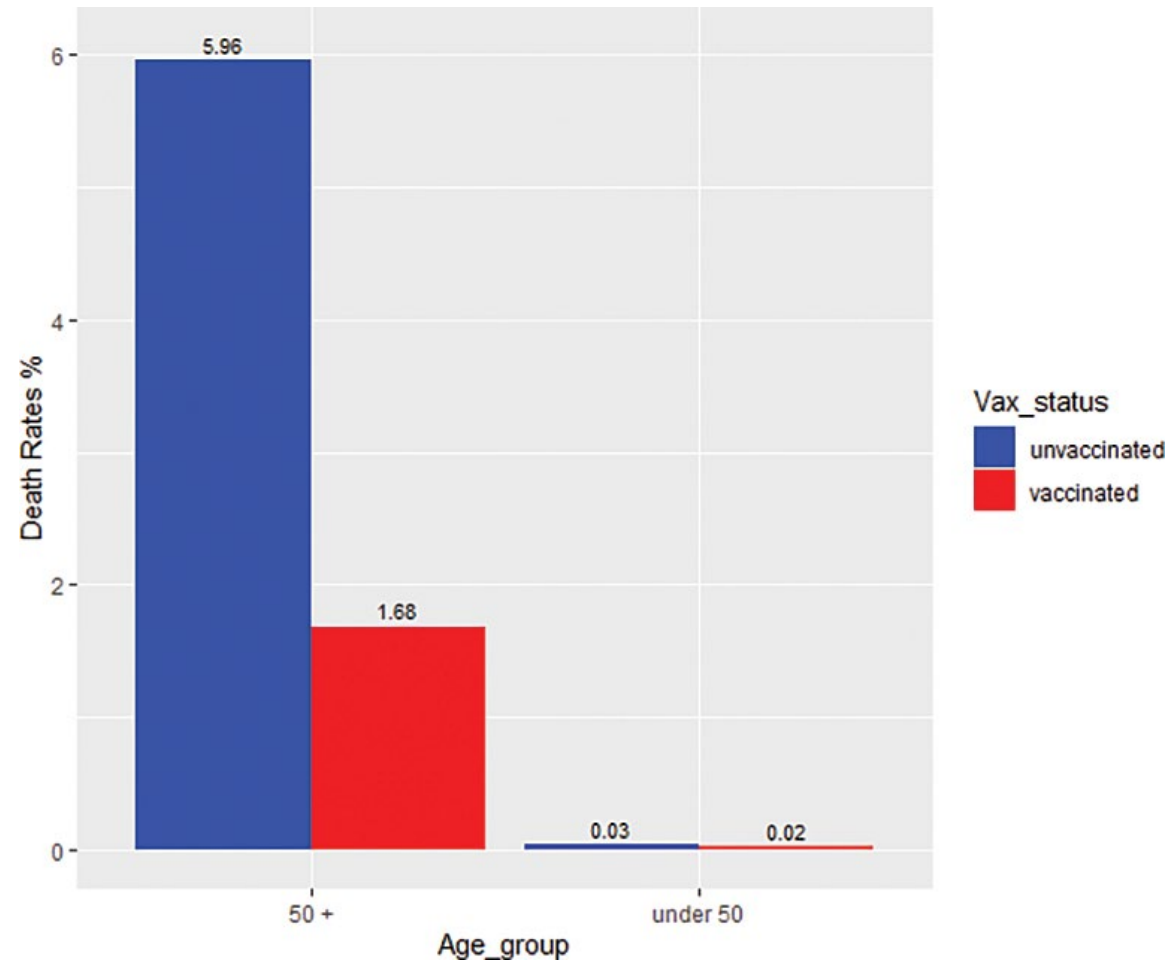


Figure 4.42 Completed second target graphic.

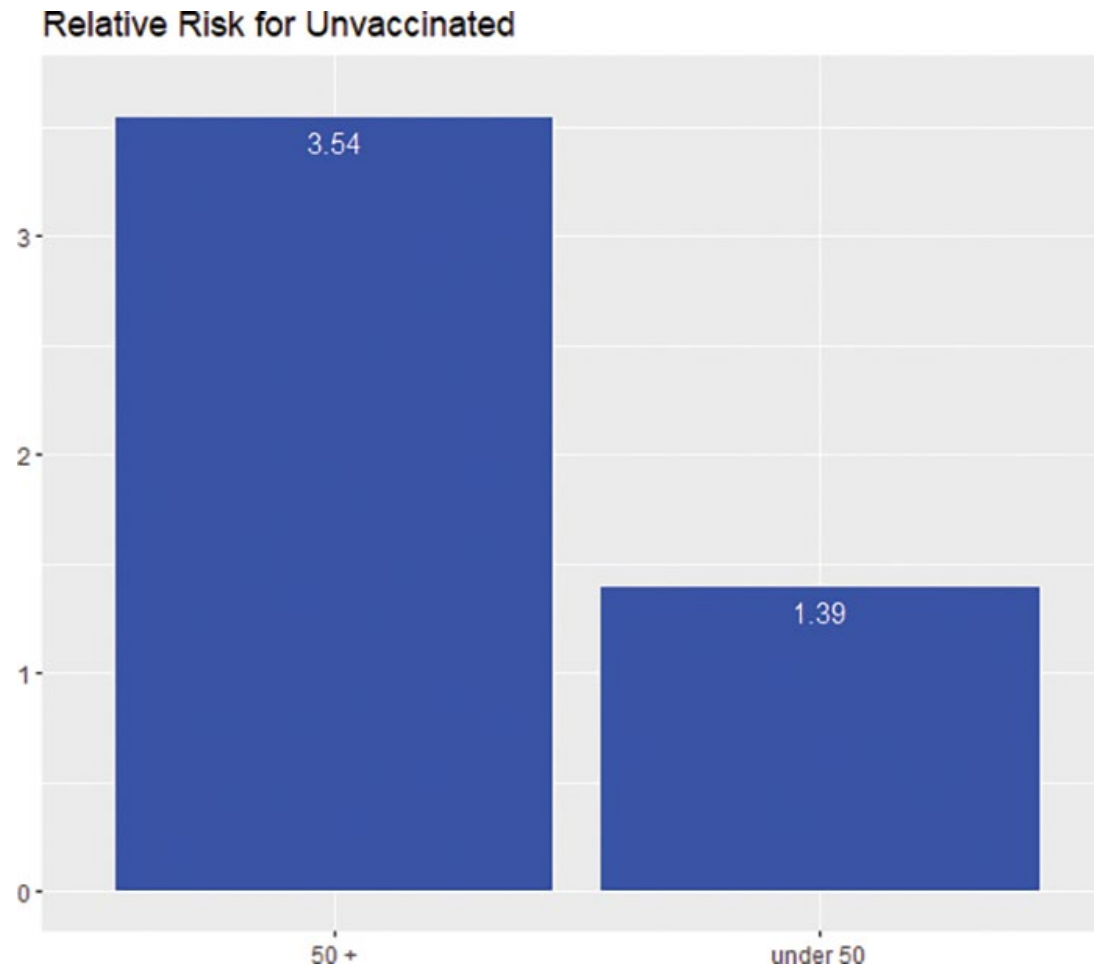


Figure 4.43 Third target graphic without text boxes.

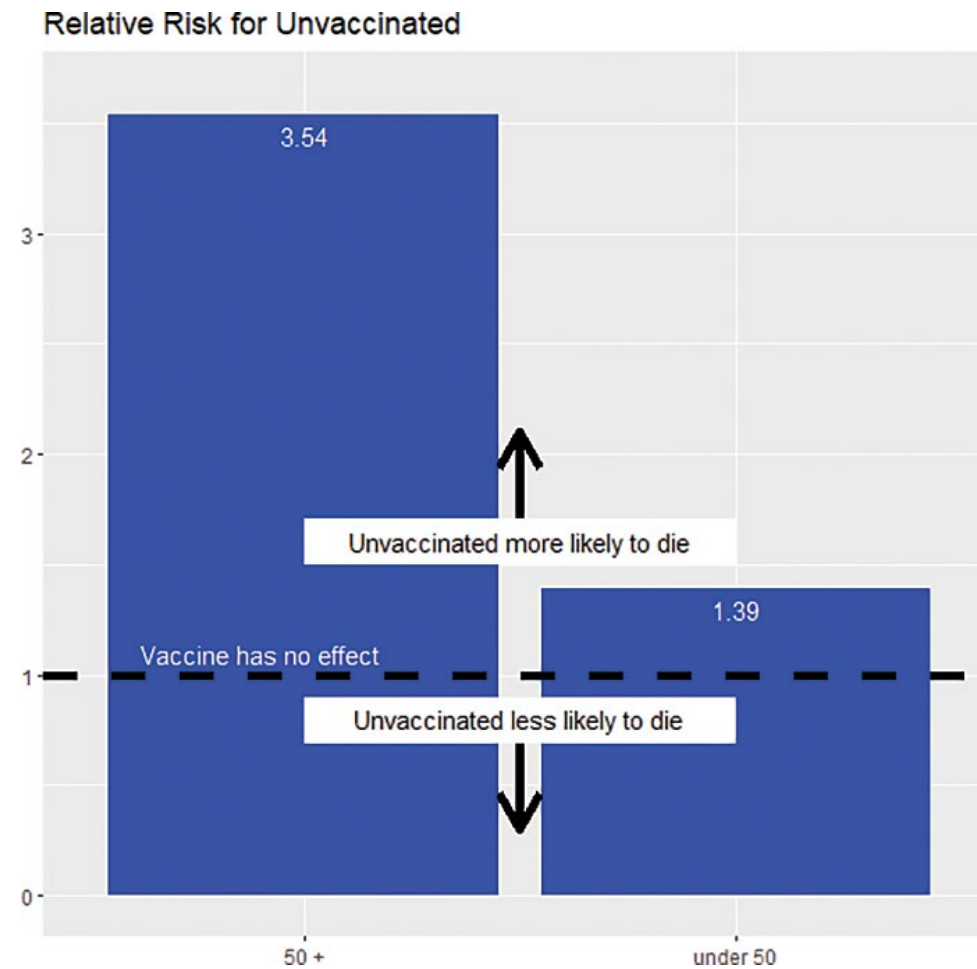


Figure 4.44 Completed third target graphic with text boxes and horizontal line.

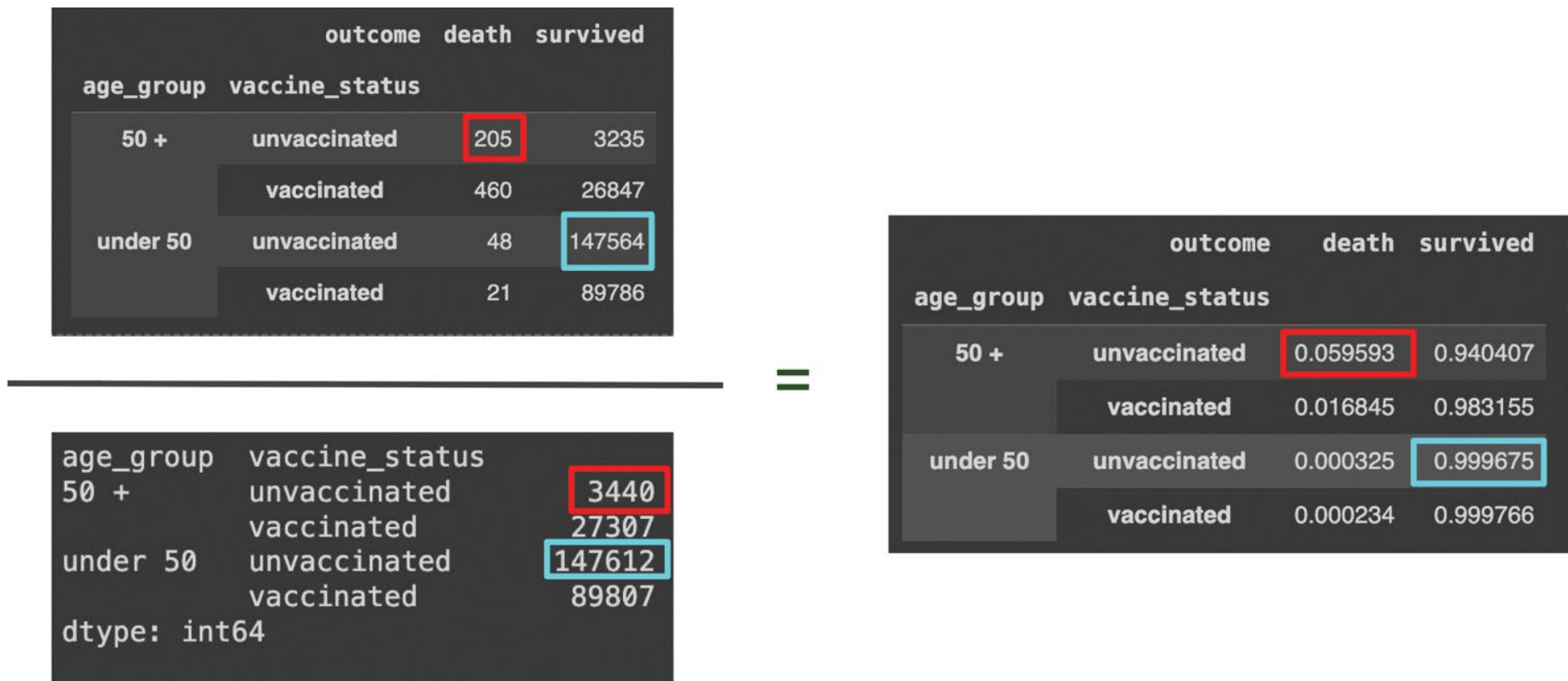


Figure 4.45 COVID-19 data division diagram.

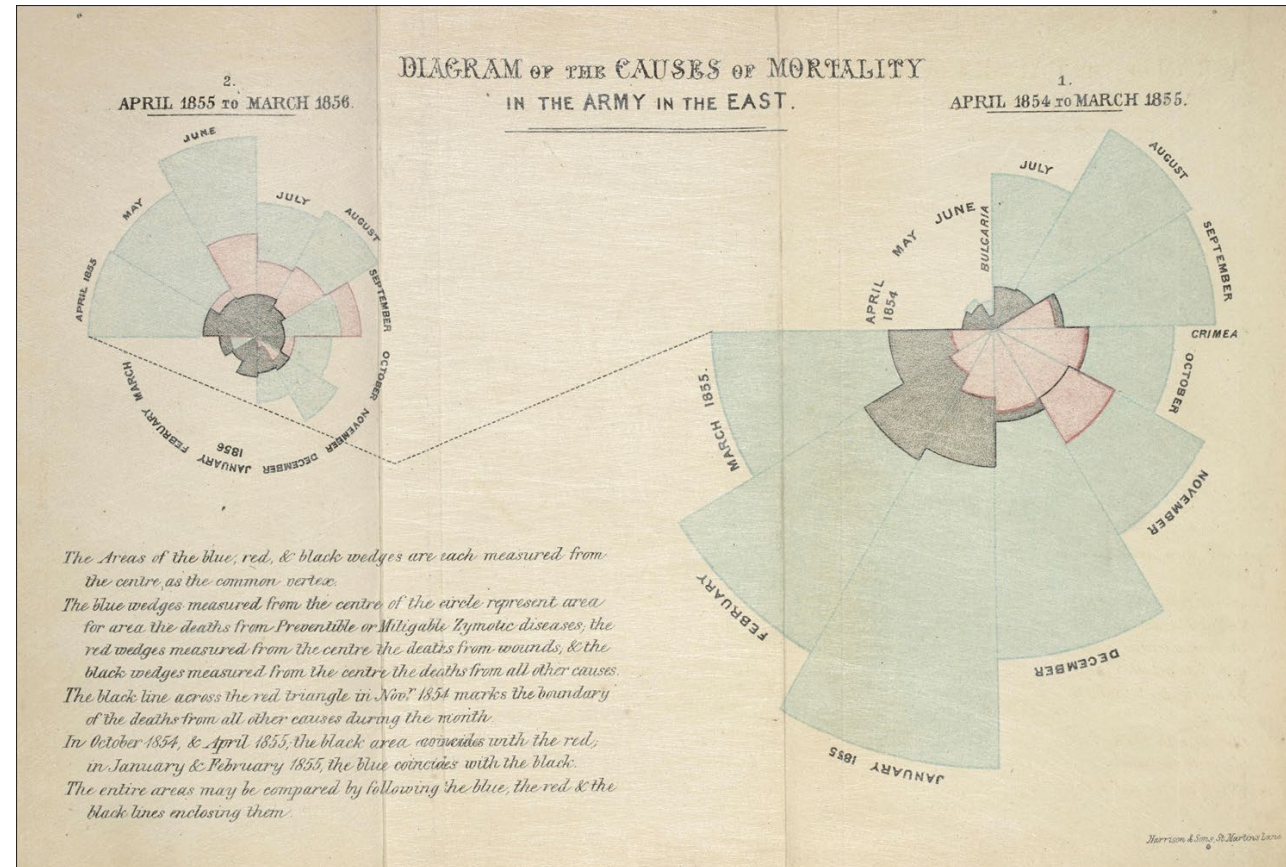


Figure 5.1 Nightingale's Rose. (Credit: Album / British Library / Alamy stock photo)

	Date	Month	Year	Army	Disease	Wounds	Other	Disease.rate	Wounds.rate	Other.rate
1	1854-04-01	Apr	1854	8571	1	0	5	1.4	0.0	7.0
2	1854-05-01	May	1854	23333	12	0	9	6.2	0.0	4.6
3	1854-06-01	Jun	1854	28333	11	0	6	4.7	0.0	2.5
4	1854-07-01	Jul	1854	28722	359	0	23	150.0	0.0	9.6
5	1854-08-01	Aug	1854	30246	828	1	30	328.5	0.4	11.9
6	1854-09-01	Sep	1854	30290	788	81	70	312.2	32.1	27.7

Figure 5.2 Nightingale's data.

Name	Status	Date modified	Type	Size
 Images	✓	11/4/2025 4:29 PM	File folder	
 .RData	✓	7/31/2024 6:02 PM	RDATA File	168 KB
 .Rhistory	✓	8/19/2024 11:12 AM	RHISTORY File	18 KB
 Nightingale	✓	5/5/2025 9:24 AM	Microsoft Excel C...	2 KB
 Nightingale.R	✓	7/31/2024 6:02 PM	R File	6 KB
 Nightingale.Rproj	✓	11/4/2025 4:30 PM	RPROJ File	1 KB

Figure 5.3 Exporting the Nightingale data.frame as a .csv file using the write.csv function.

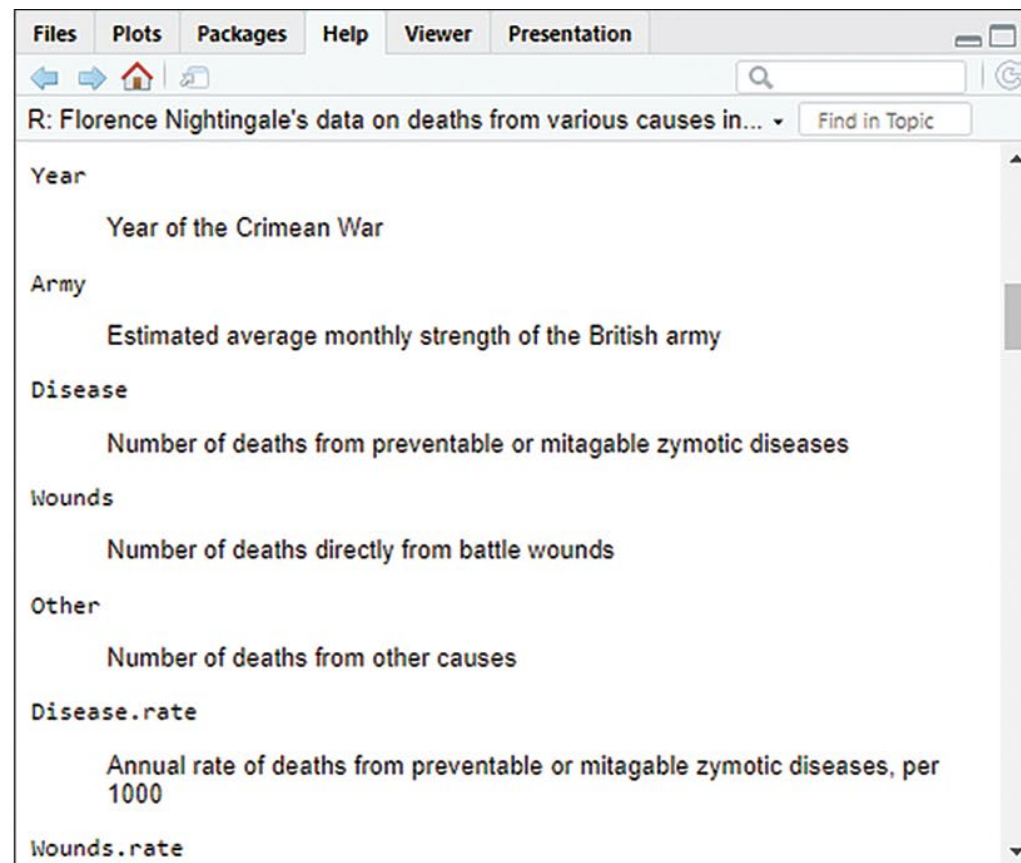
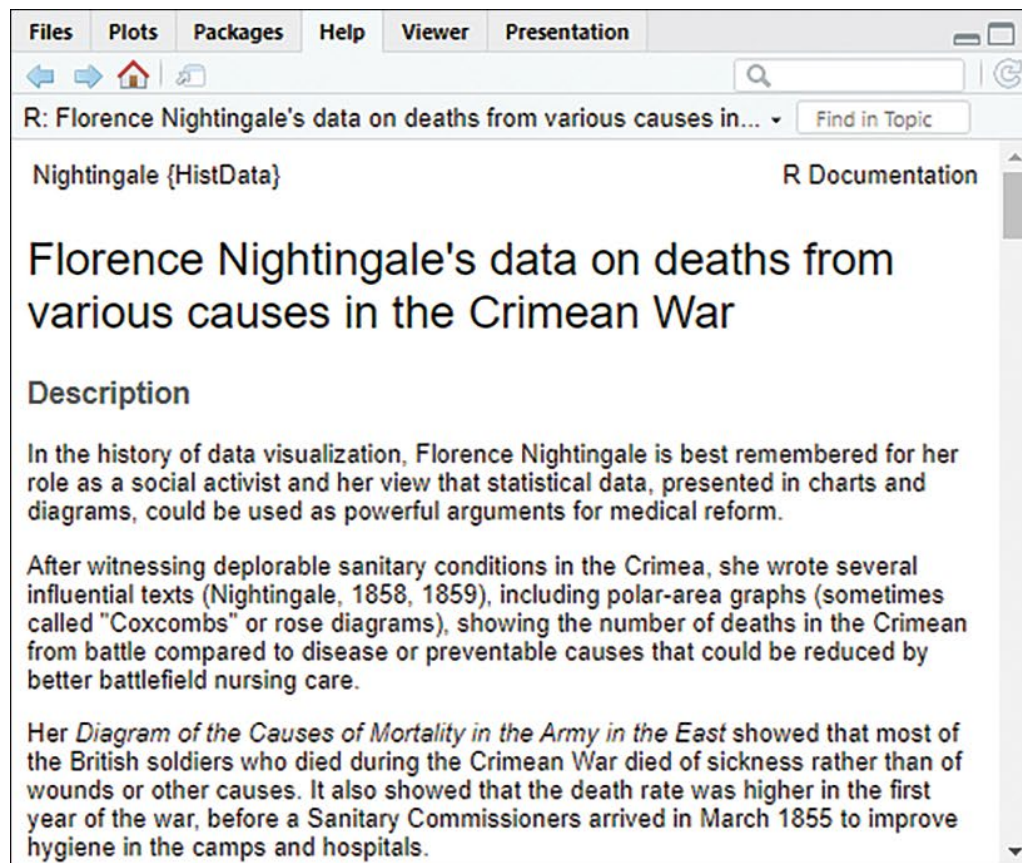


Figure 5.4 Florence Nightingale data description using the help option in R.

LOGEST ⌵ : ✖ ✔ *fx* ⌵ =(F2/\$E2)*1000*12

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Date	Month	Year	Army	Disease	Wounds	Other	Disease.r	Wounds.r	Other.rate	Rates Checked		
2	1	1854-04-0	Apr	1854	8571	1	0	5	1.4	0	7	=(F2/\$E2)*1000*12		7.00035
3	2	1854-05-0	May	1854	23333	12	0	9	6.2	0	4.6	6.171517	0	4.628638
4	3	1854-06-0	Jun	1854	28333	11	0	6	4.7	0	2.5	4.658878	0	2.541206
5	4	1854-07-0	Jul	1854	28722	359	0	23	150	0	9.6	149.9896	0	9.609359
6	5	1854-08-0	Aug	1854	30246	828	1	30	328.5	0.4	11.9	328.5062	0.396747	11.9024

Figure 5.5 Verifying the death rates with a formula in Excel.

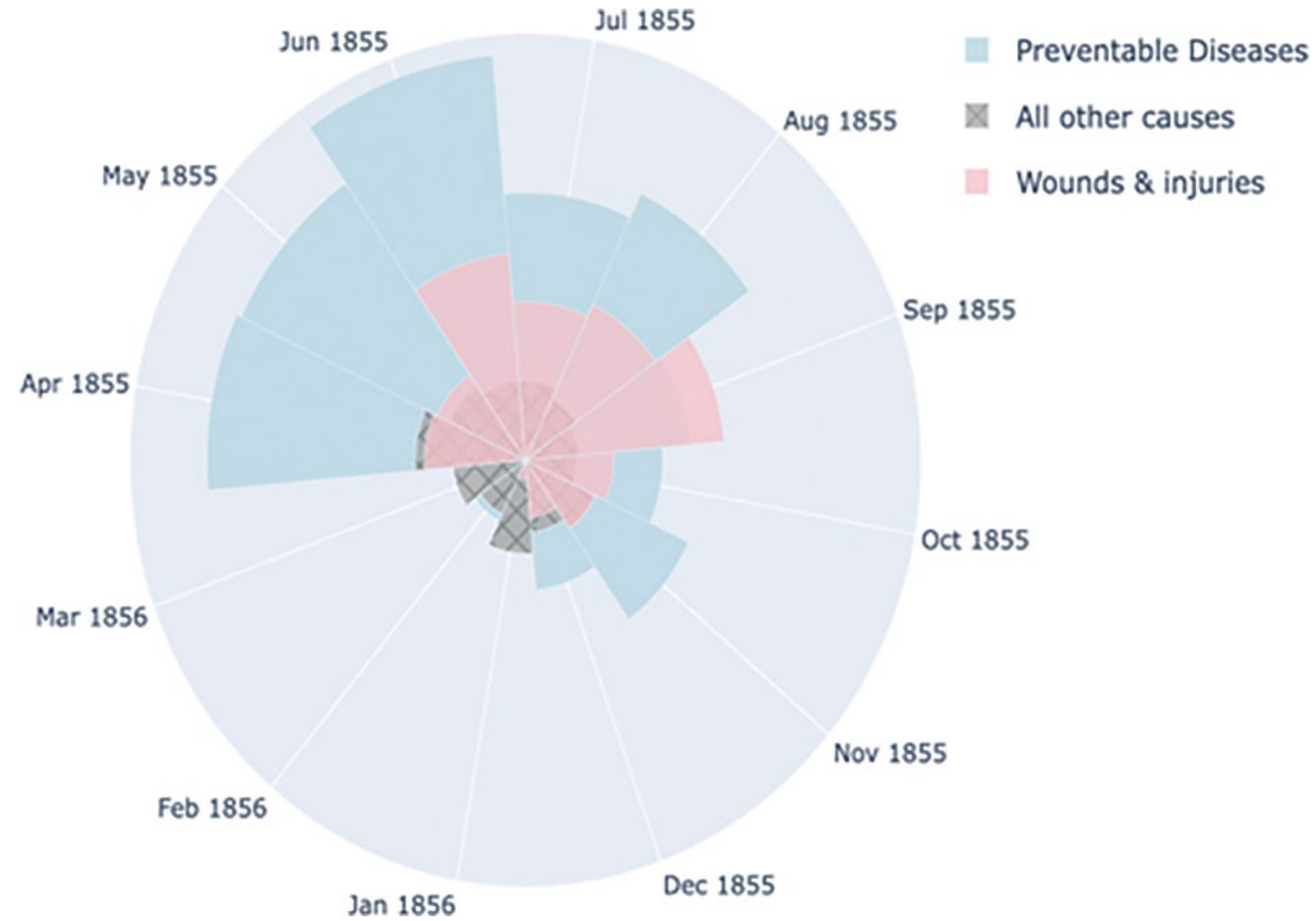


Figure 5.6 Nightingale's Rose after sanitation measures created using Python.

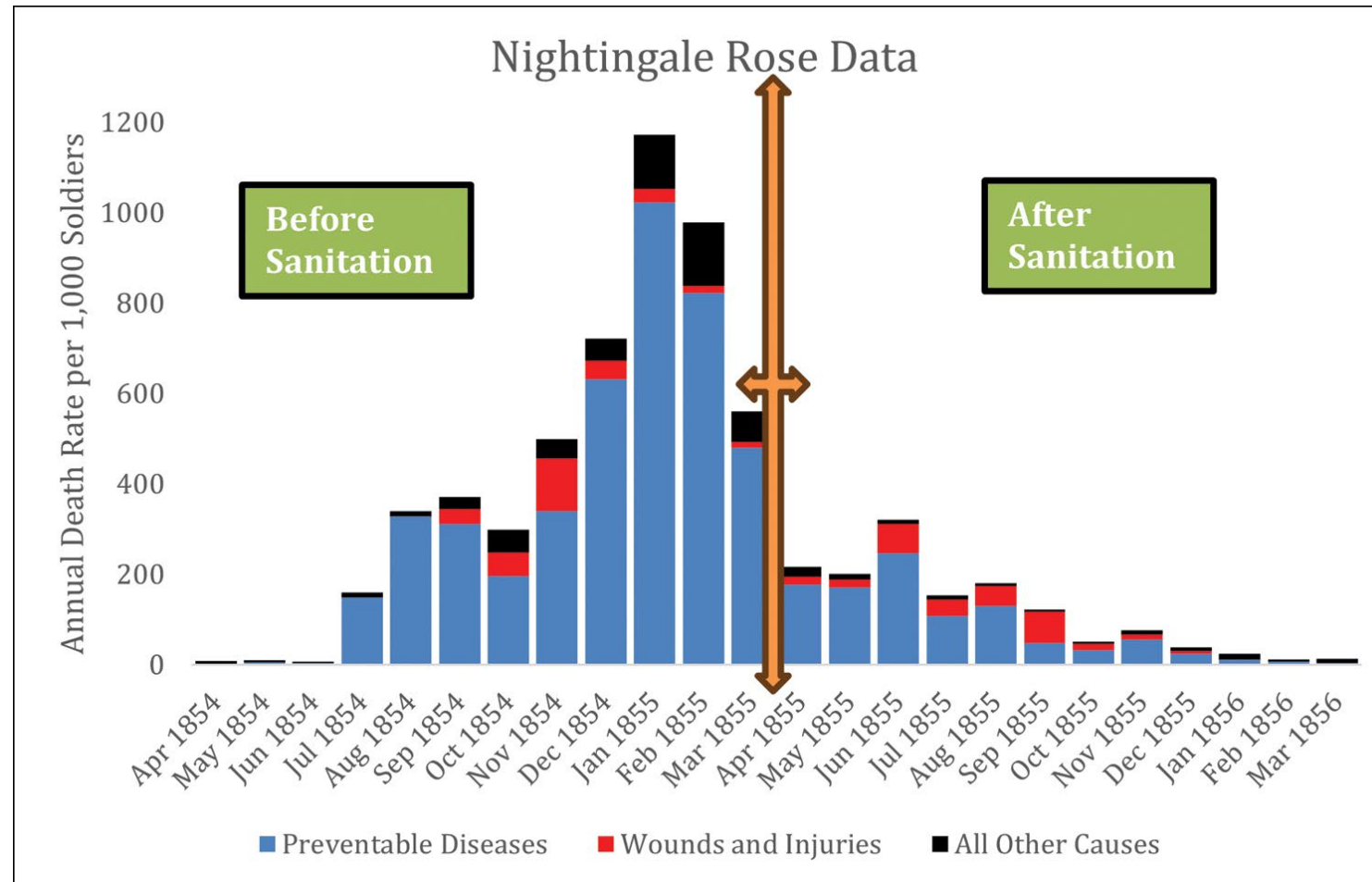


Figure 5.7 Stacked column chart version of Nightingale's Rose created using Excel.

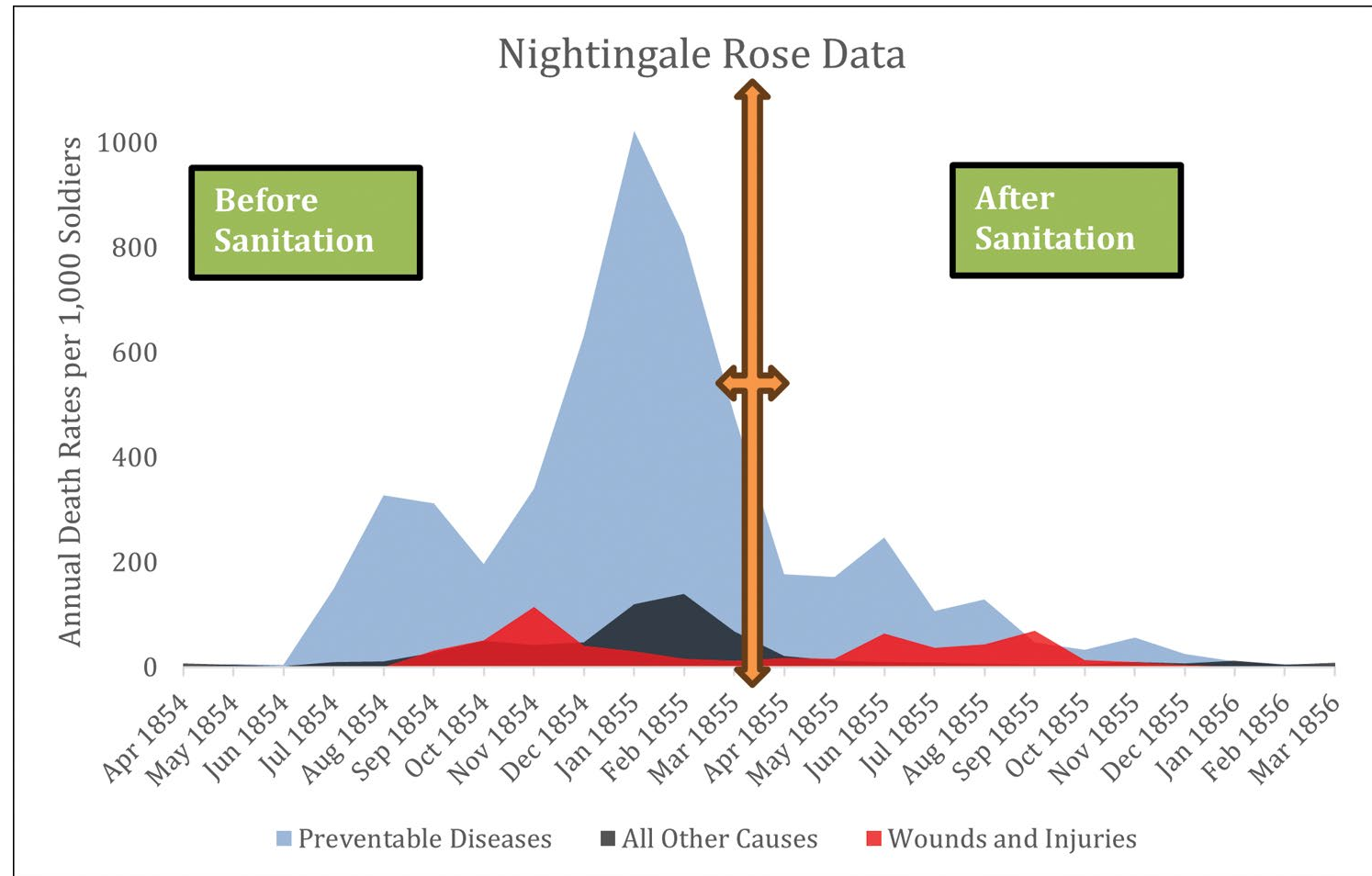


Figure 5.8 Area chart version of Nightingale's Rose created using Excel.

I2 ✖ ✓ fx =C2&" "&D2												
	A	B	C	D	E	F	G	H	I	J	K	L
1		Date	Month	Year	Army	Disease	Wounds	Other	Month	Preventable Diseases	Wounds and Injuries	All Other Causes
2	1	1854-04-01	Apr	1854	8571	1	0	5	=C2&" "&D2		0	7
3	2	1854-05-01	May	1854	23333	12	0	9	May 1854	6.2	0	4.6
4	3	1854-06-01	Jun	1854	28333	11	0	6	Jun 1854	4.7	0	2.5
5	4	1854-07-01	Jul	1854	28722	359	0	23	Jul 1854	150	0	9.6
6	5	1854-08-01	Aug	1854	20246	828	1	20	Aug 1854	228.5	0.4	11.0

Figure 5.9 Concatenating (combining) the Month and Year columns in Excel.



Figure 5.10 Cock's comb. (Credit: Potapov Alexander / Shutterstock)

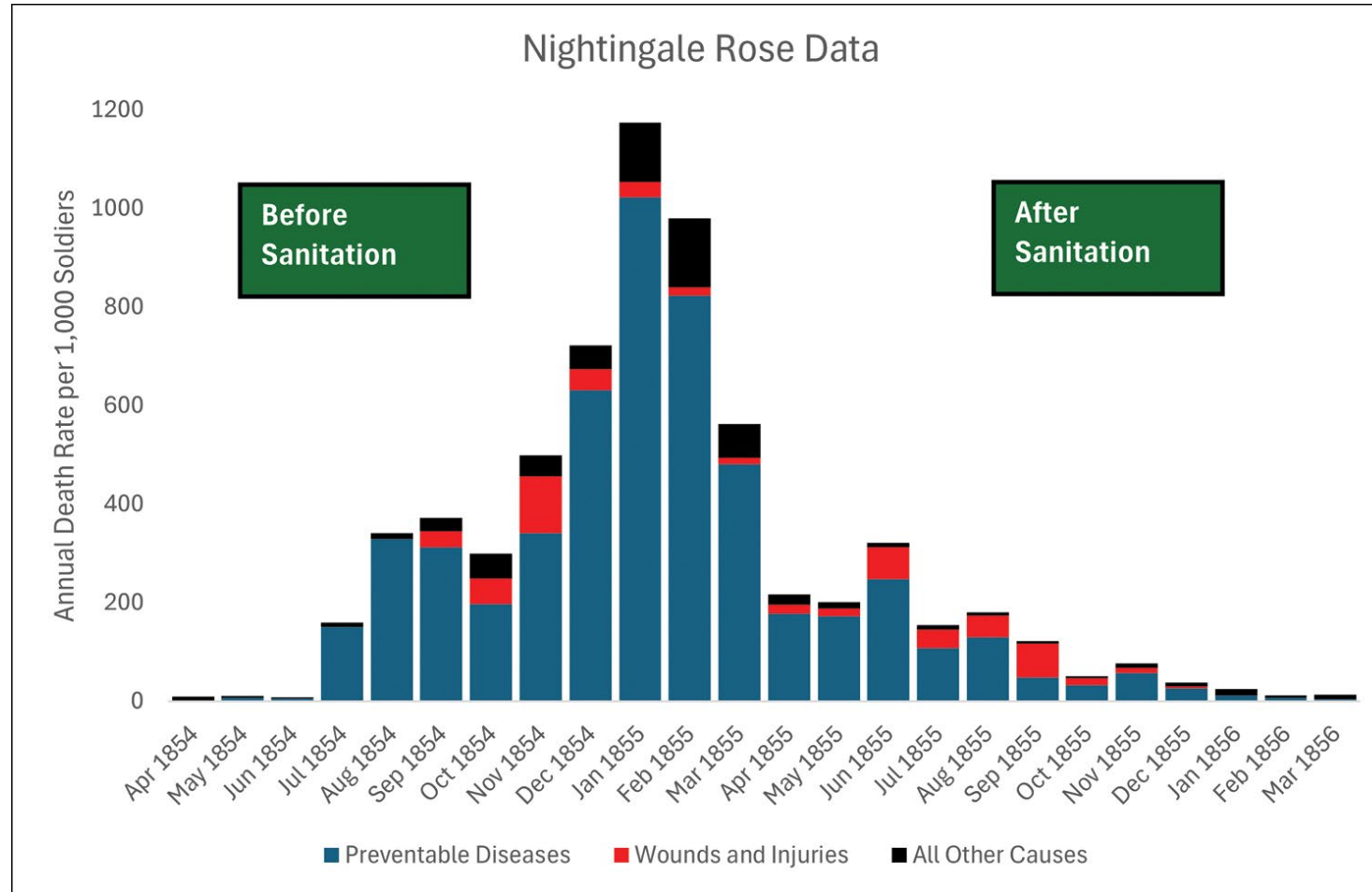


Figure 5.11 Stacked column chart version of Nightingale's Rose created using Excel.

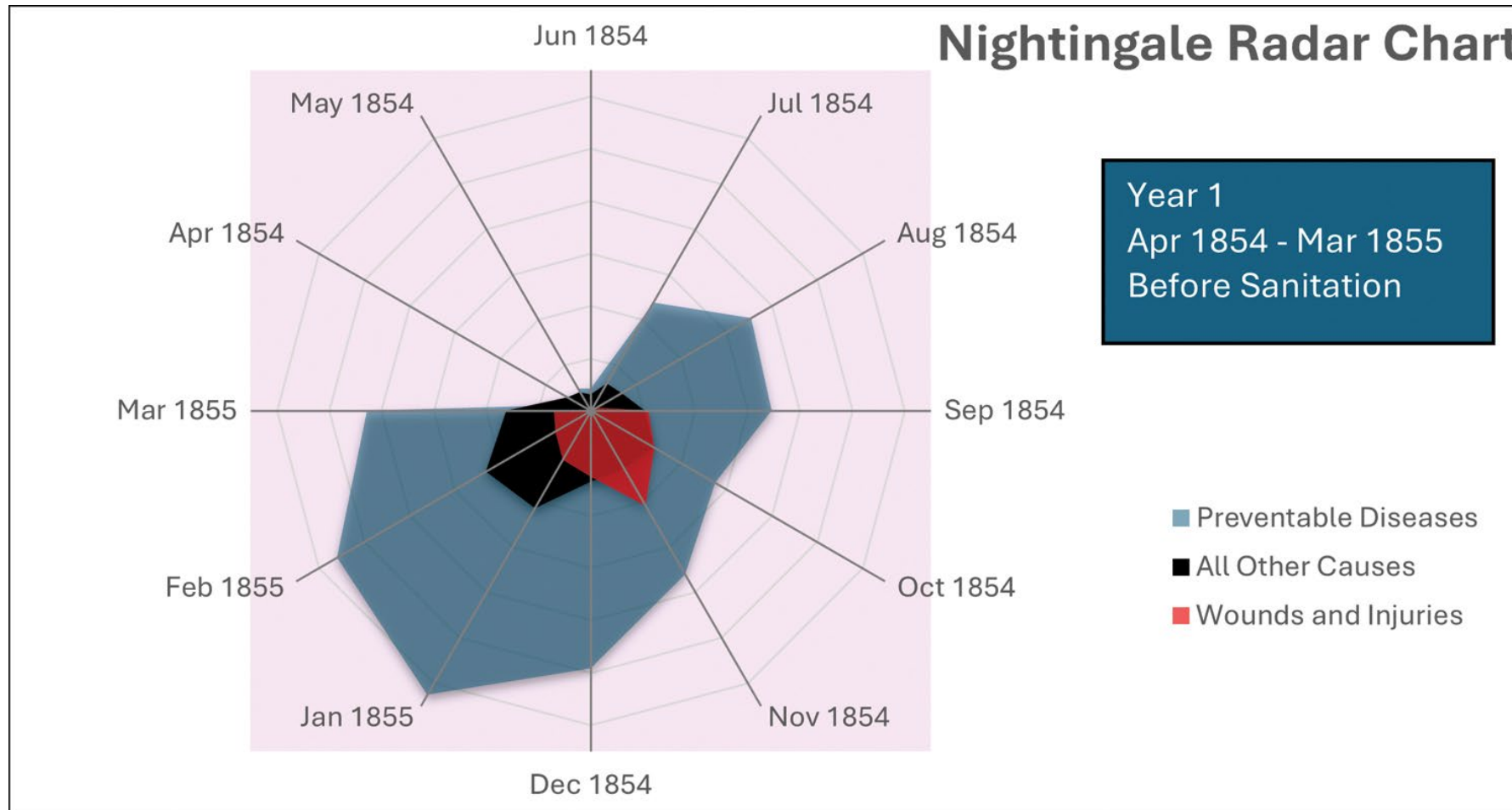


Figure 5.12 Radar chart version of Nightingale's Rose created using Excel.



Figure 5.13 The Chart Design menu with Change Chart Type option in Excel.

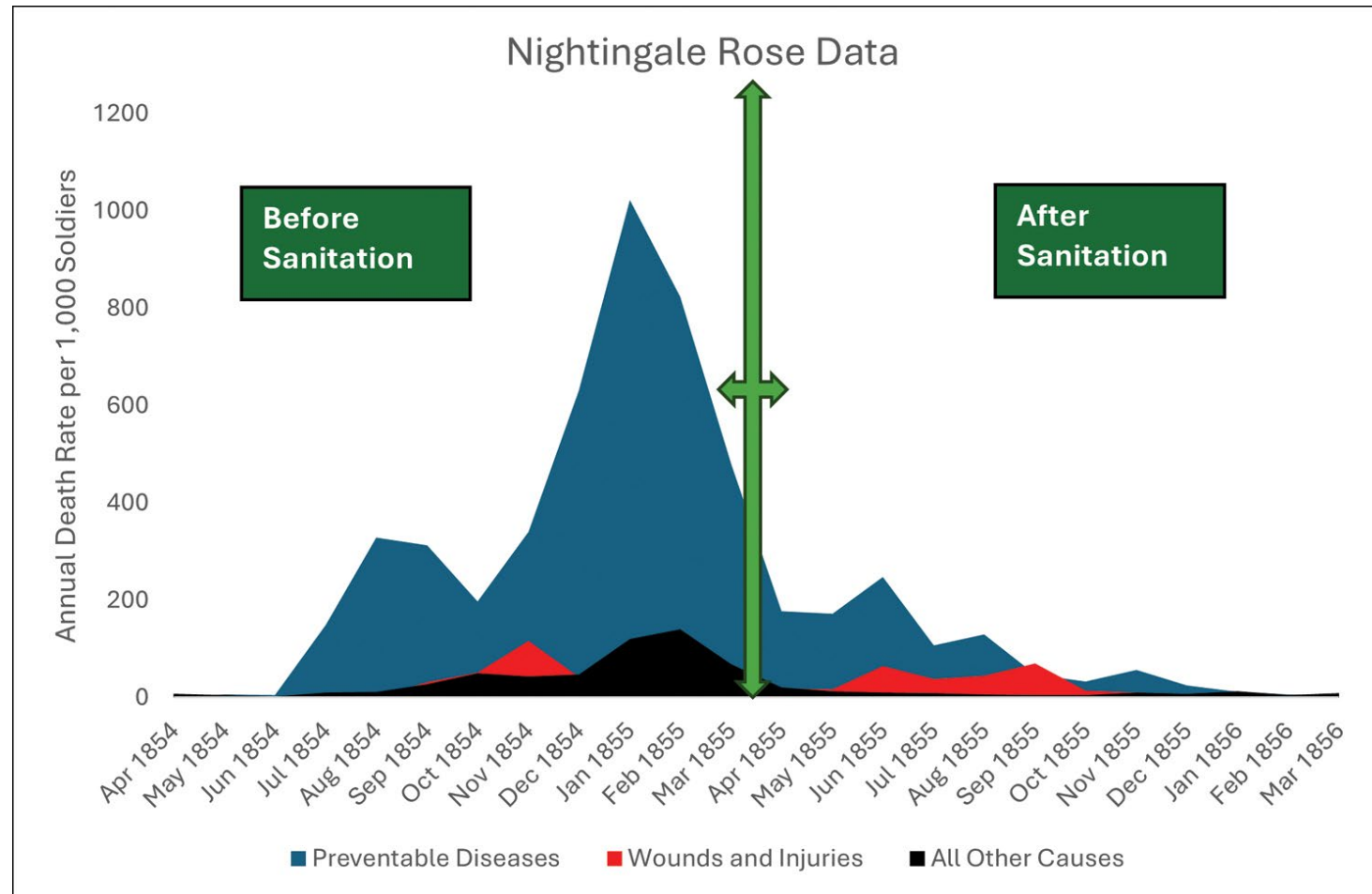


Figure 5.14 Area chart version of Nightingale's Rose created using Excel.

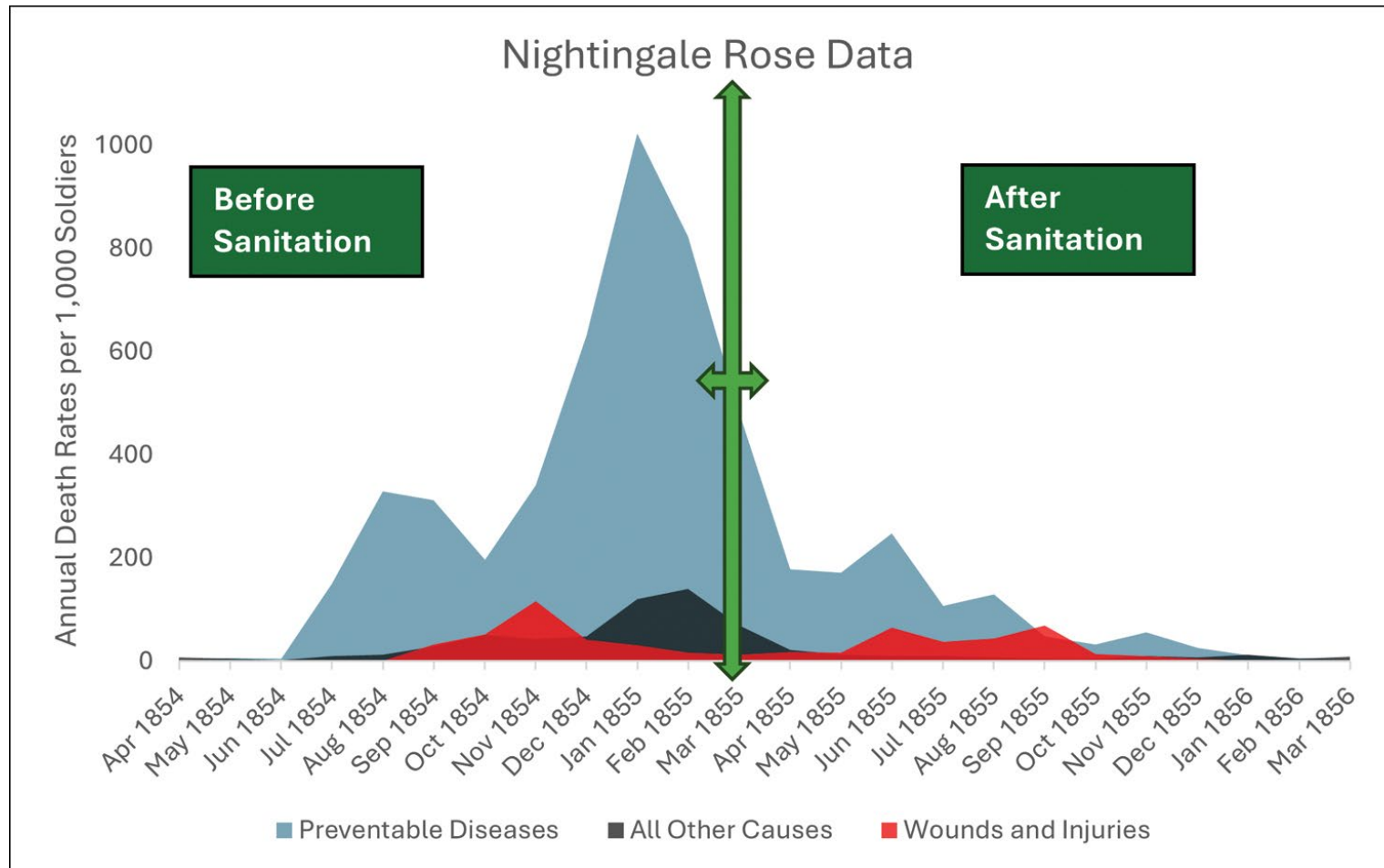


Figure 5.15 Area chart version of Nightingale's Rose with wounds in front and transparency added using Excel.

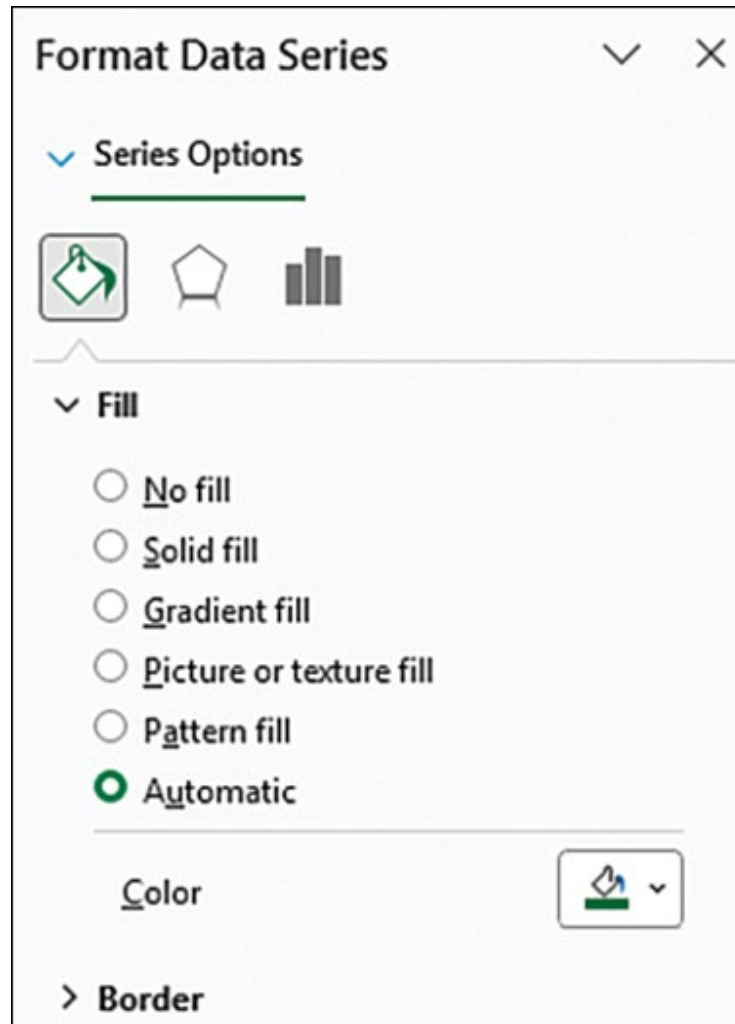


Figure 15.16 Format Data Series dialog in Excel.

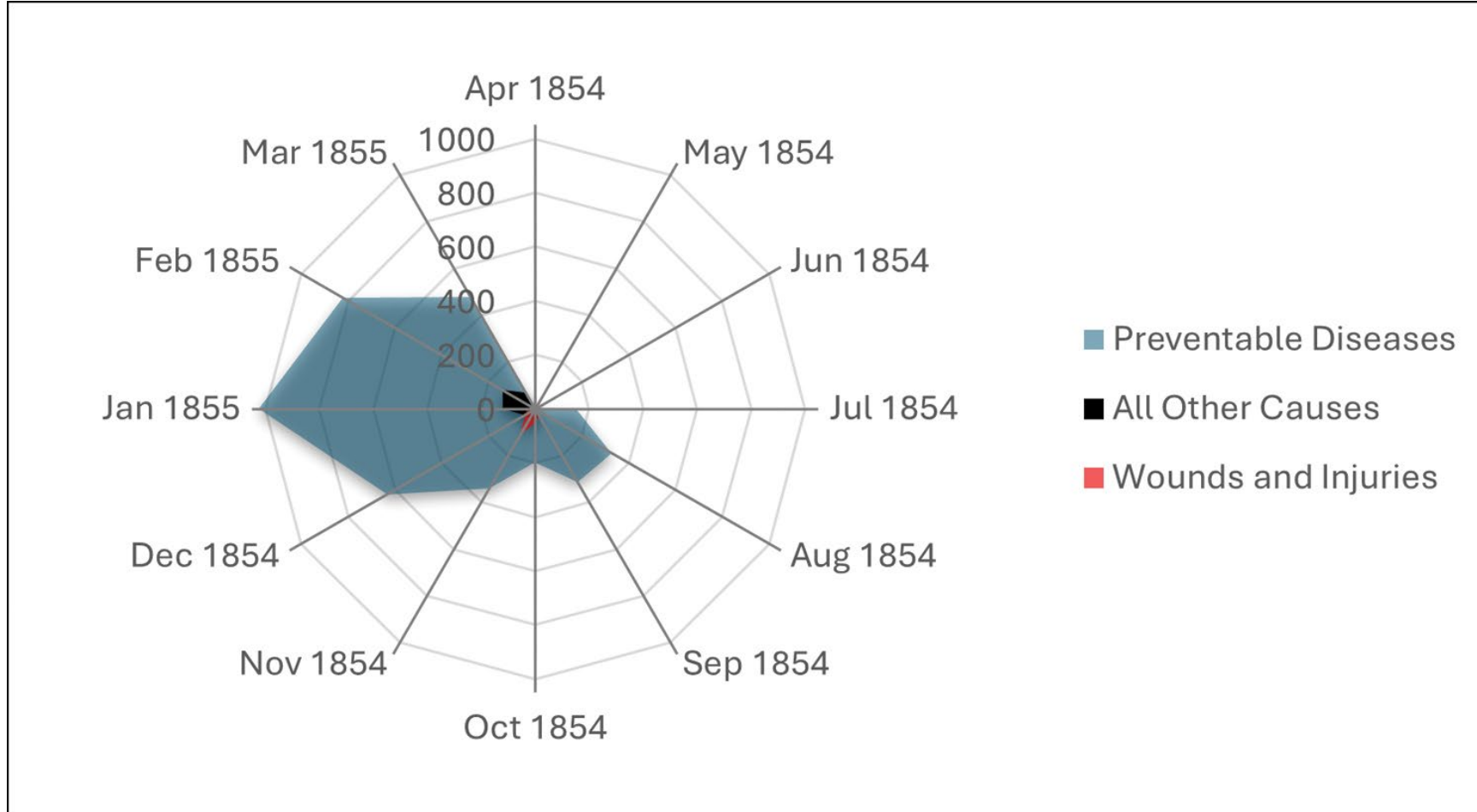


Figure 5.17 Radar chart version of Nightingale's Rose created using Excel.

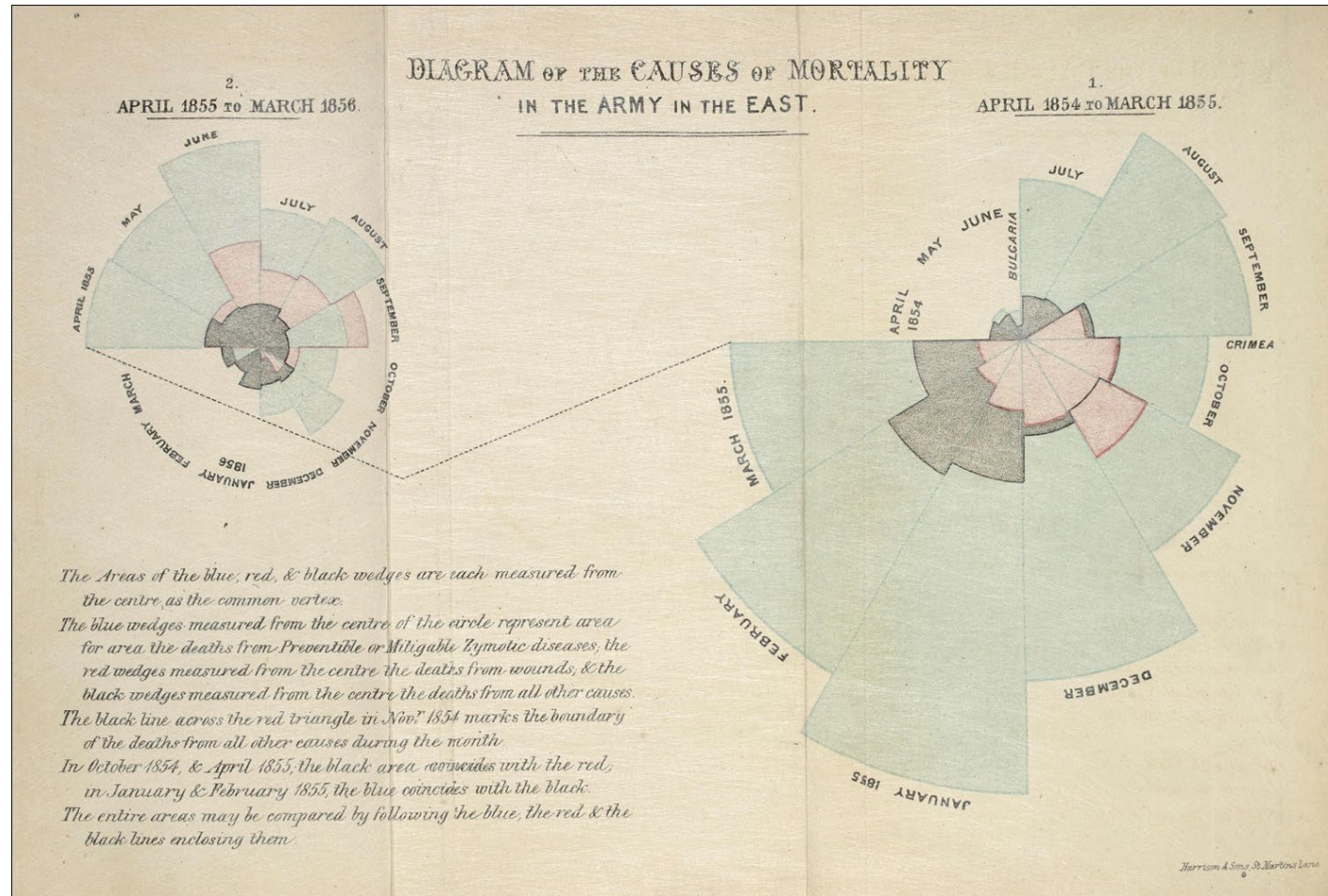


Figure 5.18 Nightingale's Rose Diagram has data set equal to area, not radius. (Credit: Album / British Library / Alamy stock photo)

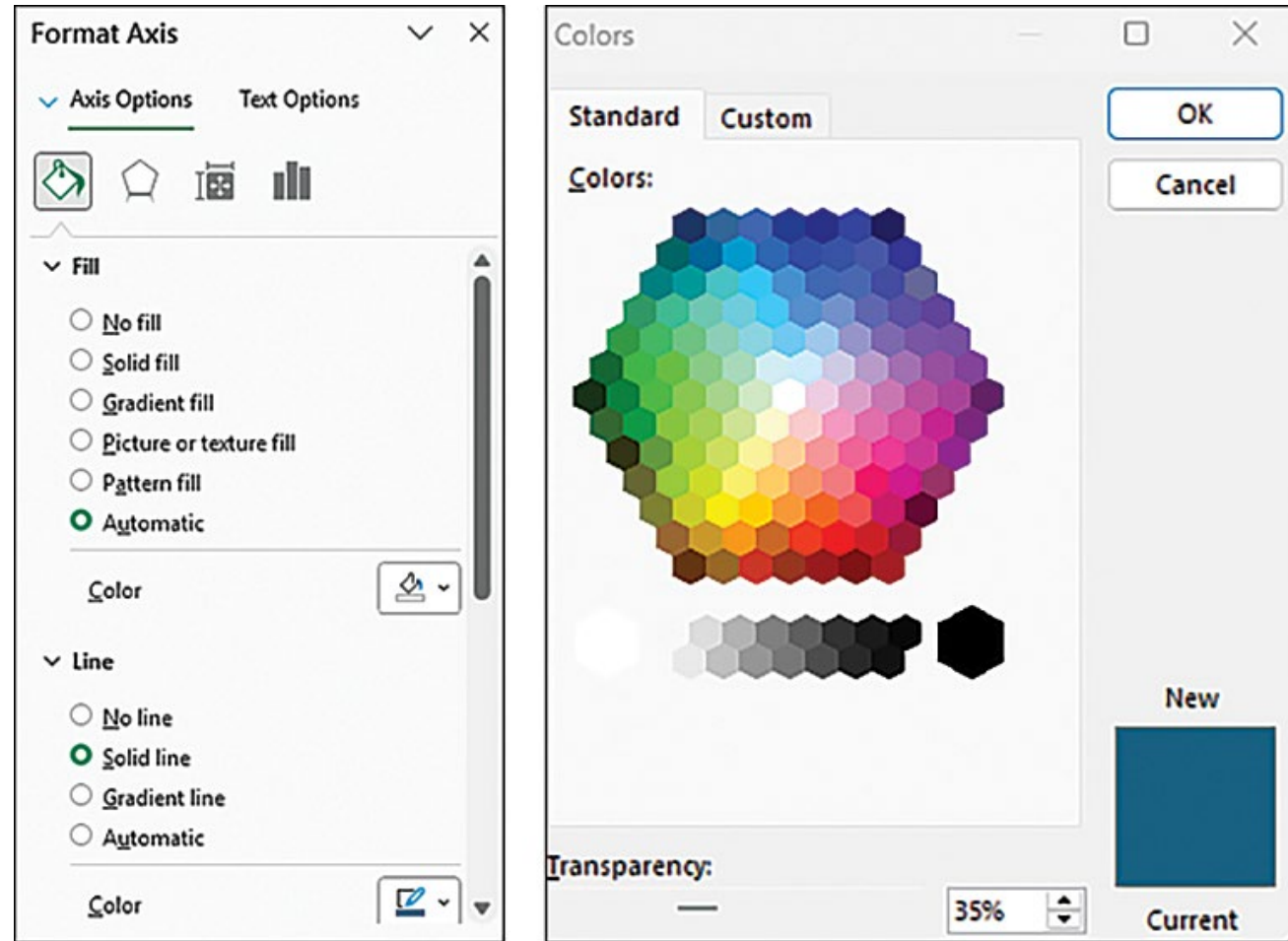


Figure 5.19 Formatting options for radar charts in Excel.

	J	K	L	M	N	O
	Preventable Diseases	All Other Causes	Wounds and Injuries	Preventable Diseases	All Other Causes	Wounds and Injuries
	1.4	7	0	$-\text{SQRT}(J2 \cdot 12 / \text{PI}())$		0.00
4	6.2	4.6	0	$\text{SQRT}(\text{number})$	4.19	0.00
	4.7	2.5	0	4.24	3.09	0.00
	150	9.6	0	23.94	6.06	0.00
	328.5	11.9	0.4	35.42	6.74	1.24
	312.2	27.7	32.1	34.53	10.29	11.07
	107	50.1	51.7	27.12	12.82	14.05

Figure 5.20 Computing the radius of a circular wedge with the area set equal to the death rate.

M2																
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
		Date	Month	Year	Army	Disease	Wounds	Other	Month	Preventable Diseases	All Other Causes	Wounds and Injuries	Preventable Diseases	All Other Causes	Wounds and Injuries	
1																
2	1	1854-06-0	Jun	1854	28333	11	0	6	Jun 1854	4.7	2.5	0	4.24	3.09	0.00	
3	2	1854-07-0	Jul	1854	28722	359	0	23	Jul 1854	150	9.6	0	23.94	6.06	0.00	
4	3	1854-08-0	Aug	1854	30246	828	1	30	Aug 1854	328.5	11.9	0.4	35.42	6.74	1.24	
5	4	1854-09-0	Sep	1854	30290	788	81	70	Sep 1854	312.2	27.7	32.1	34.53	10.29	11.07	
6	5	1854-10-0	Oct	1854	30643	503	132	128	Oct 1854	197	50.1	51.7	27.43	13.83	14.05	
7	6	1854-11-0	Nov	1854	29736	844	287	106	Nov 1854	340.6	42.8	115.8	36.07	12.79	21.03	
8	7	1854-12-0	Dec	1854	32779	1725	114	131	Dec 1854	631.5	48	41.7	49.11	13.54	12.62	
9	8	1855-01-0	Jan	1855	32393	2761	83	324	Jan 1855	1022.8	120	30.7	62.50	21.41	10.83	
10	9	1855-02-0	Feb	1855	30919	2120	42	361	Feb 1855	822.8	140.1	16.3	56.06	23.13	7.89	
11	10	1855-03-0	Mar	1855	30107	1205	32	172	Mar 1855	480.3	68.6	12.8	42.83	16.19	6.99	
12	11	1854-04-0	Apr	1854	8571	1	0	5	Apr 1854	1.4	7	0	2.31	5.17	0.00	
13	12	1854-05-0	May	1854	23333	12	0	9	May 1854	6.2	4.6	0	4.87	4.19	0.00	
14	13	1855-06-0	Jun	1855	38863	802	209	31	Jun 1855	247.6	9.6	64.5	30.75	6.06	15.70	
15	14	1855-07-0	Jul	1855	42647	382	134	33	Jul 1855	107.5	9.3	37.7	20.26	5.96	12.00	
16	15	1855-08-0	Aug	1855	44614	483	164	25	Aug 1855	129.9	6.7	44.1	22.28	5.06	12.98	
17	16	1855-09-0	Sep	1855	47751	189	276	20	Sep 1855	47.5	5	69.4	13.47	4.37	16.28	
18	17	1855-10-0	Oct	1855	46852	128	53	18	Oct 1855	32.8	4.6	13.6	11.19	4.19	7.21	
19	18	1855-11-0	Nov	1855	37853	178	33	32	Nov 1855	56.4	10.1	10.5	14.68	6.21	6.33	
20	19	1855-12-0	Dec	1855	43217	91	18	28	Dec 1855	25.3	7.8	5	9.83	5.46	4.37	
21	20	1856-01-0	Jan	1856	44212	42	2	48	Jan 1856	11.4	13	0.5	6.60	7.05	1.38	
22	21	1856-02-0	Feb	1856	43485	24	0	19	Feb 1856	6.6	5.2	0	5.02	4.46	0.00	
23	22	1856-03-0	Mar	1856	46140	15	0	35	Mar 1856	3.9	9.1	0	3.86	5.90	0.00	
24	23	1855-04-0	Apr	1855	32252	477	48	57	Apr 1855	177.5	21.2	17.9	26.04	9.00	8.27	
25	24	1855-05-0	May	1855	35473	508	49	37	May 1855	171.8	12.5	16.6	25.62	6.91	7.96	

Figure 5.21 Rearranging the data so the dates align properly in the radar chart.

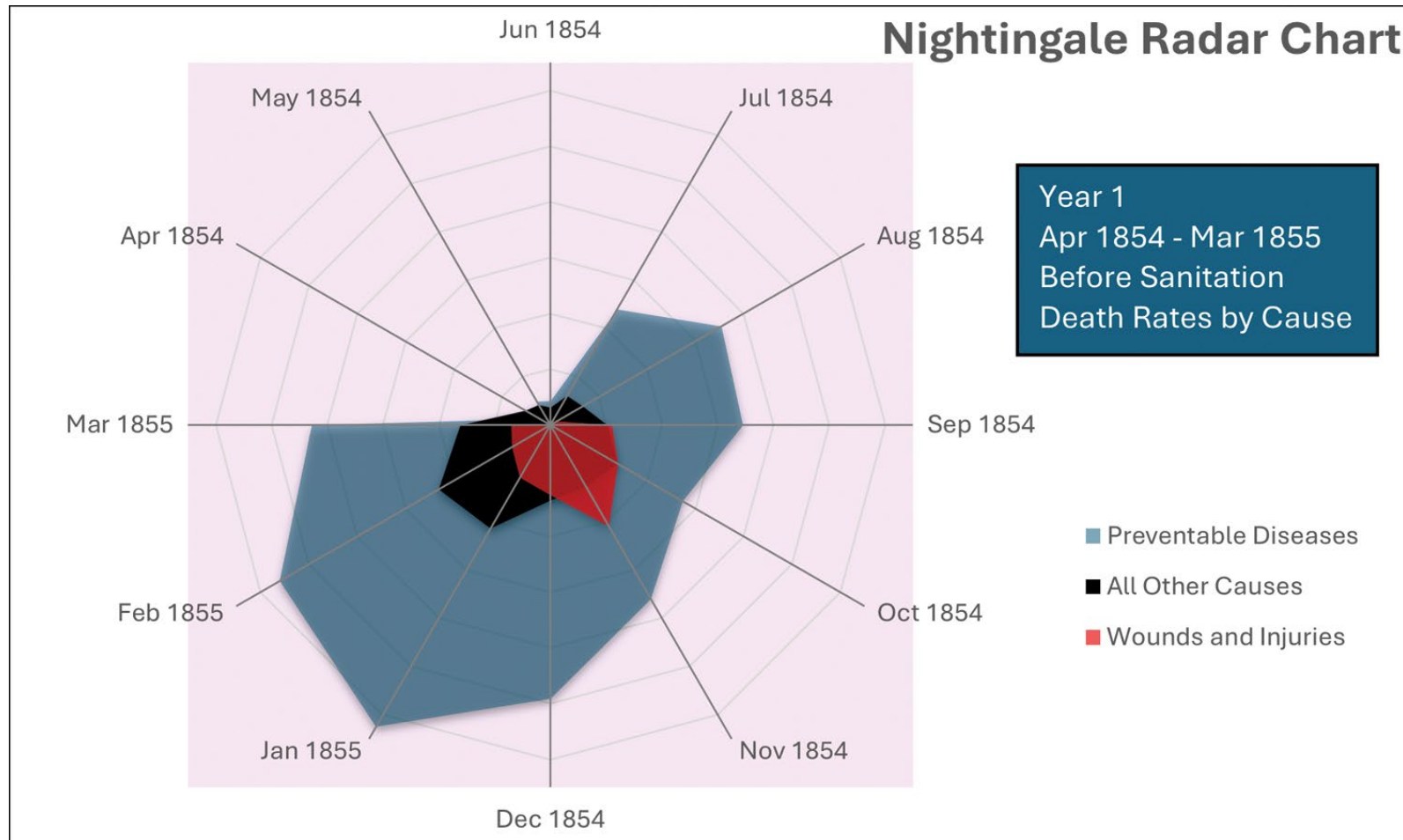


Figure 5.22 Radar chart version of Nightingale's Rose before sanitation created using Excel.

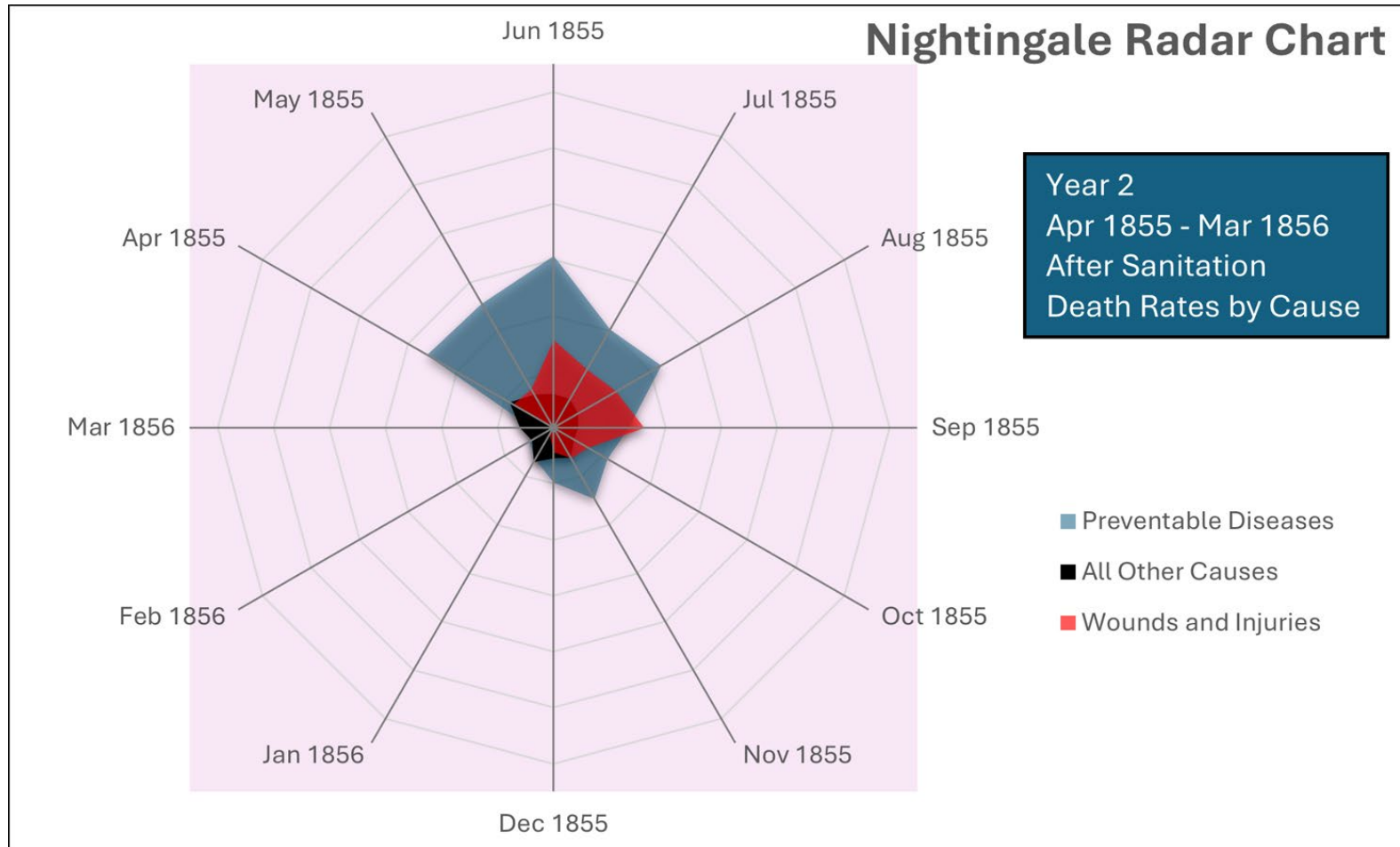


Figure 5.23 Radar chart version of Nightingale's Rose after sanitation created using Excel.

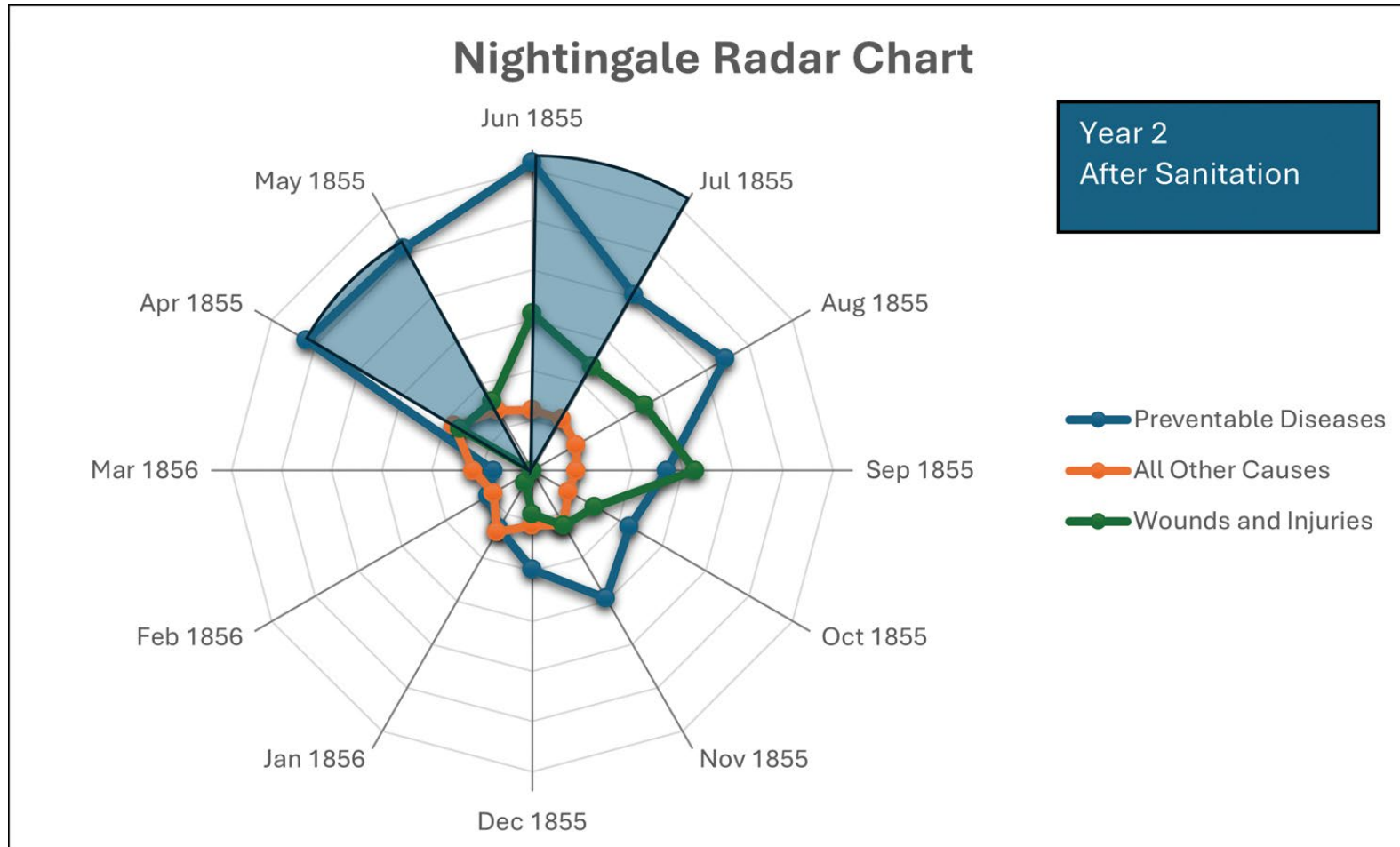


Figure 5.24 Radar chart version of Nightingale's Rose with circular wedges created using Excel.

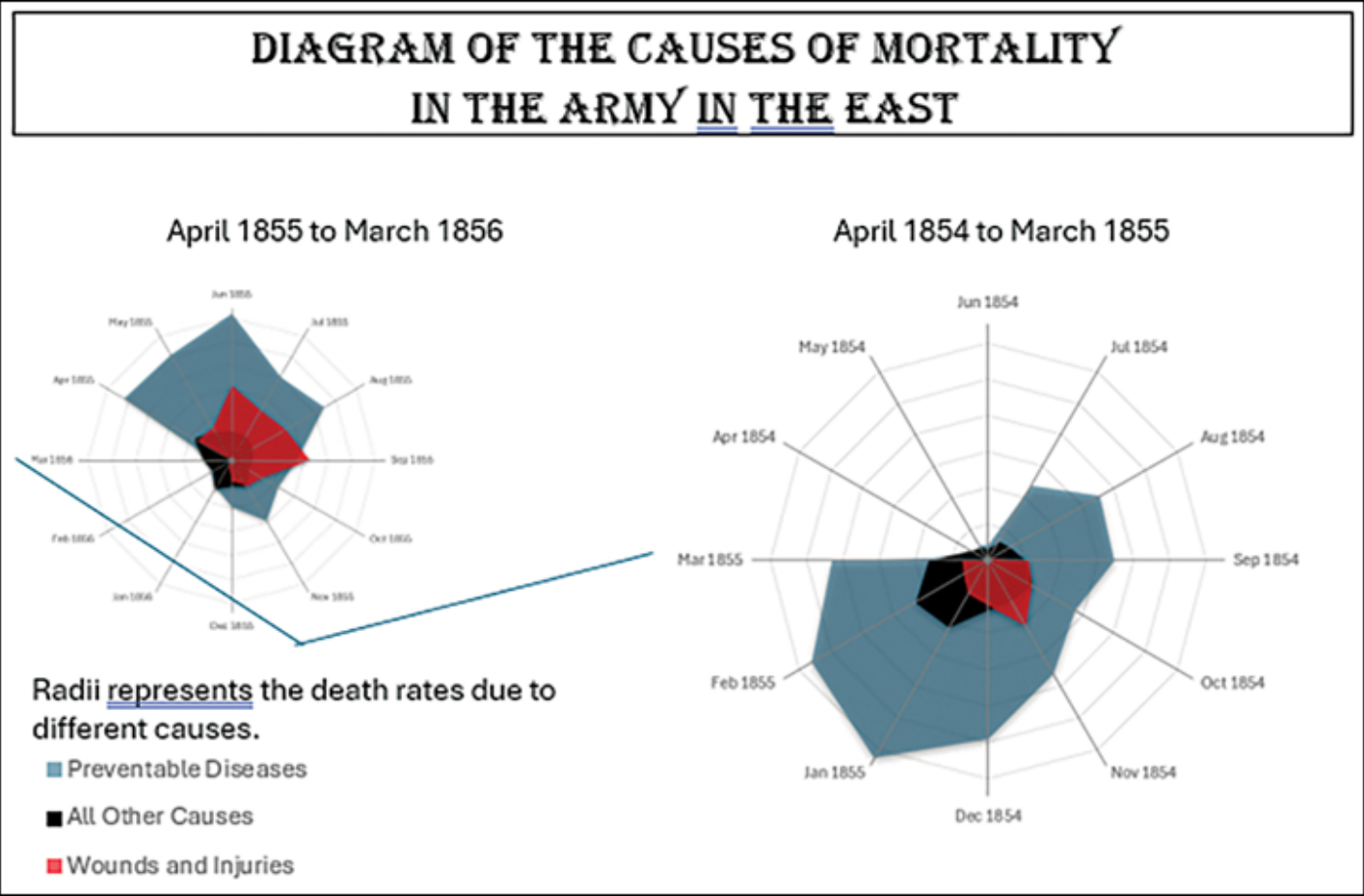


Figure 5.25 Double radar chart version of Nightingale’s Rose created using Excel.

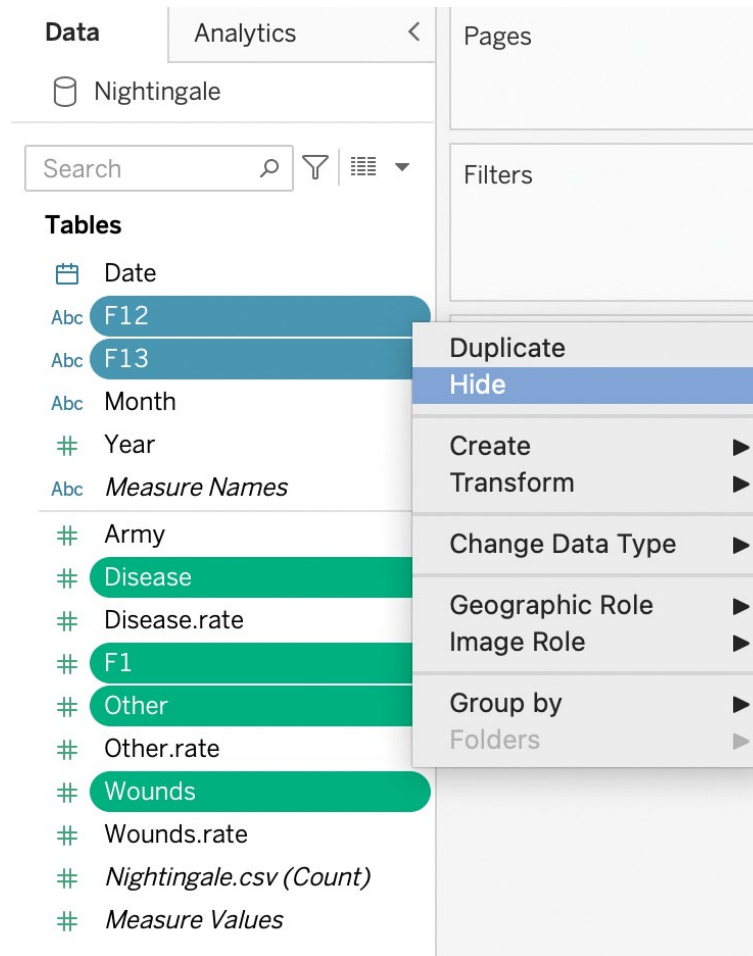


Figure 5.26 Hiding data variables in Tableau.

 Nightingale.csv	 Nightingale.csv	 Nightingale.csv	 Nightingale.csv	 Nightingale.csv		
Date	Month	Year	Army	Disease.rate		
4/1/1854	Apr	1854	8,571	1.40		7.000
5/1/1854	May	1854	23,333	6.20		4.600
6/1/1854	Jun	1854	28,333	4.70		2.500
7/1/1854	Jul	1854	28,722	150.00		9.600
8/1/1854	Aug	1854	30,246	328.50	0.400	11.900
9/1/1854	Sep	1854	30,290	312.20	32.100	27.700
10/1/1854	Oct	1854	30,643	197.00	51.700	50.100
11/1/1854	Nov	1854	29,736	340.60	115.800	42.800
12/1/1854	Dec	1854	32,779	631.50	41.700	48.000
1/1/1855	Jan	1855	32,393	1,022.80	30.700	120.000
2/1/1855	Feb	1855	30,919	822.80	16.300	140.100
3/1/1855	Mar	1855	30,107	480.30	12.800	68.600

▼

Rename

Copy Values

Hide

Create Calculated Field...

Pivot

Merge Mismatched Fields

Figure 5.27 Pivoting in Tableau.

 Nightingale.csv	 Nightingale.csv	 Nightingale.csv	 Nightingale.csv	 Pivot	 Pivot
Date	Month	Year	Army	Pivot Field Names	Pivot Field Values
4/1/1854	Apr	1854	8,571	Disease.rate	1.40
4/1/1854	Apr	1854	8,571	Other.rate	7.00
4/1/1854	Apr	1854	8,571	Wounds.rate	0.00
5/1/1854	May	1854	23,333	Disease.rate	6.20
5/1/1854	May	1854	23,333	Other.rate	4.60
5/1/1854	May	1854	23,333	Wounds.rate	0.00
6/1/1854	Jun	1854	28,333	Disease.rate	4.70
6/1/1854	Jun	1854	28,333	Other.rate	2.50
6/1/1854	Jun	1854	28,333	Wounds.rate	0.00
7/1/1854	Jul	1854	28,722	Disease.rate	150.00
7/1/1854	Jul	1854	28,722	Other.rate	9.60
7/1/1854	Jul	1854	28,722	Wounds.rate	0.00

Figure 5.28 The data after pivoting.

Abc Nightingale.csv	# Nightingale.csv	# Nightingale.csv	Abc Pivot	# Pivot	=Abc Calculation
Month	Year	Army	Pivot Field Names	Pivot Field Values	Pivot Field Names - Split 1
Apr	1854	8,571	Disease.rate	1.40	Disease
Apr	1854	8,571	Other.rate	7.00	Other
Apr	1854	8,571	Wounds.rate	0.00	Wounds
May	1854	23,333	Disease.rate	6.20	Disease
May	1854	23,333	Other.rate	4.60	Other
May	1854	23,333	Wounds.rate	0.00	Wounds
Jun	1854	28,333	Disease.rate	4.70	Disease
Jun	1854	28,333	Other.rate	2.50	Other
Jun	1854	28,333	Wounds.rate	0.00	Wounds
Jul	1854	28,722	Disease.rate	150.00	Disease
Jul	1854	28,722	Other.rate	9.60	Other
Jul	1854	28,722	Wounds.rate	0.00	Wounds

Figure 5.29 Using a split to clean up the pivot field names.

 Nightingale.csv	 Nightingale.csv	 Nightingale.csv	 Nightingale.csv	 Pivot	 Calculation
Date	Month	Year	Army	Mortality Rate	Cause of Death
4/1/1854	Apr	1854	8,571	1.40	Disease
4/1/1854	Apr	1854	8,571	7.00	Other
4/1/1854	Apr	1854	8,571	0.00	Wounds
5/1/1854	May	1854	23,333	6.20	Disease
5/1/1854	May	1854	23,333	4.60	Other
5/1/1854	May	1854	23,333	0.00	Wounds
6/1/1854	Jun	1854	28,333	4.70	Disease
6/1/1854	Jun	1854	28,333	2.50	Other
6/1/1854	Jun	1854	28,333	0.00	Wounds
7/1/1854	Jul	1854	28,722	150.00	Disease
7/1/1854	Jul	1854	28,722	9.60	Other
7/1/1854	Jul	1854	28,722	0.00	Wounds

Figure 5.30 Pivoted and cleaned data.

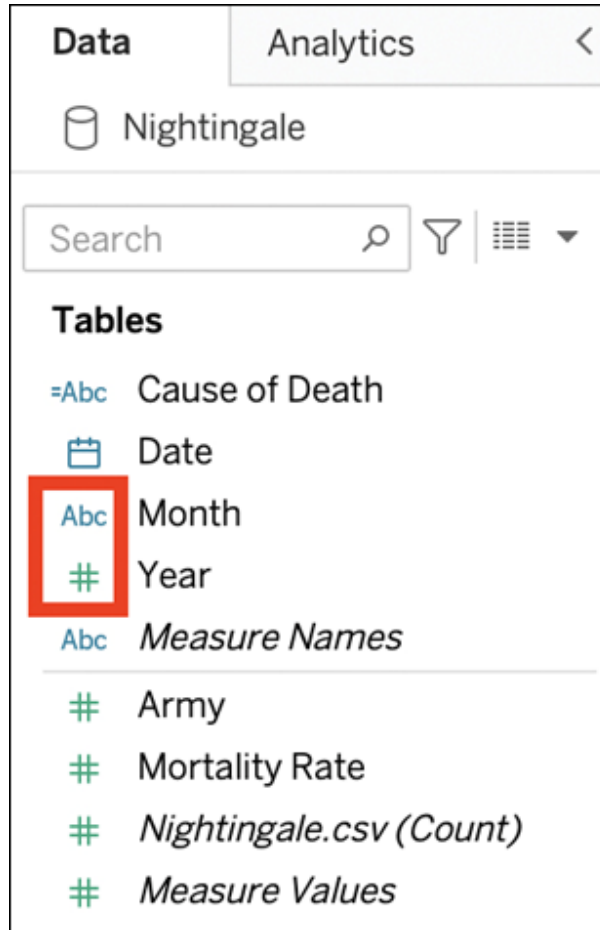


Figure 5.31 Tableau showing month as a dimension and year as a measure.

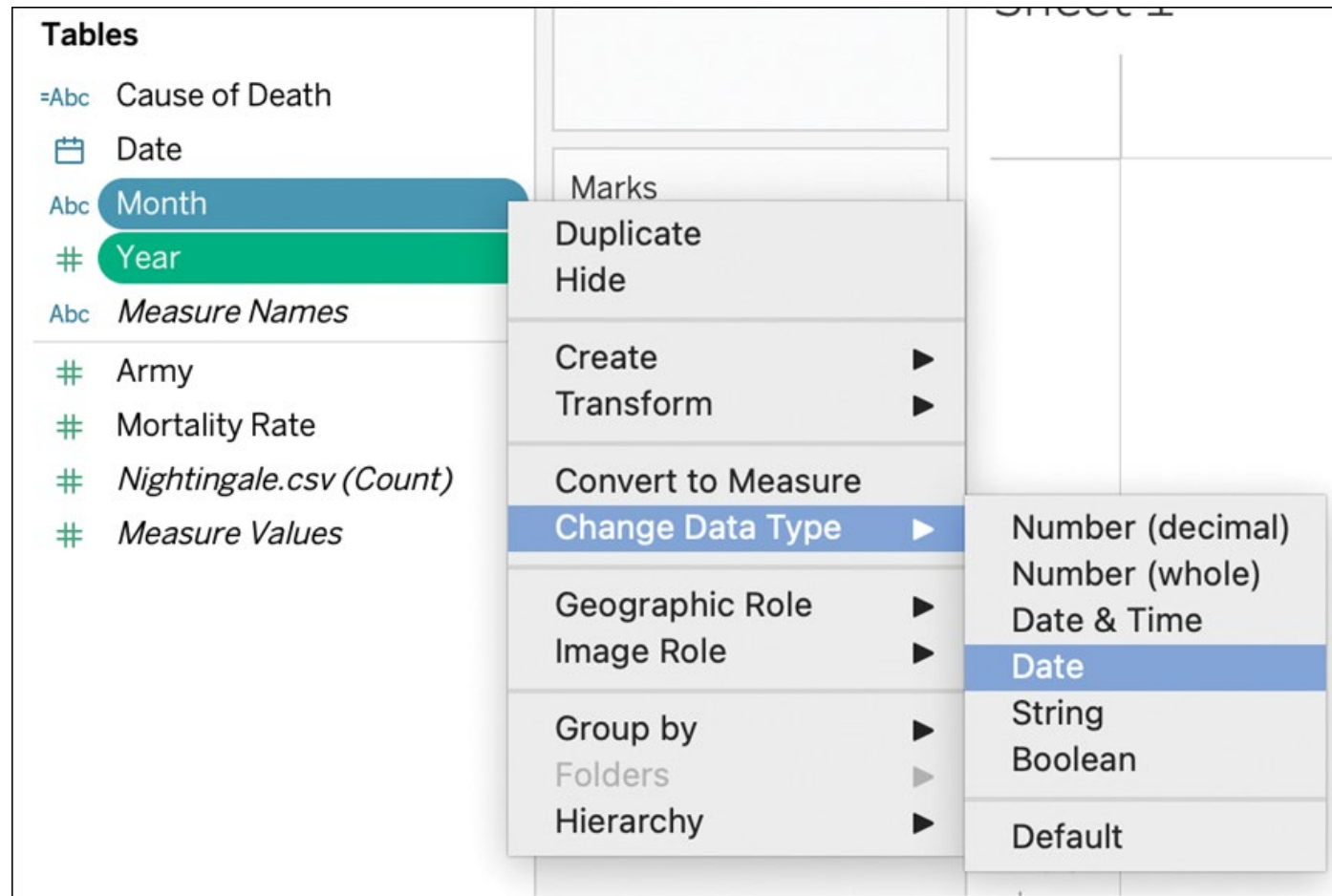


Figure 5.32 Converting to the Date data type.

Timeframe

×

IF [Date]< DATE("1855-04-01") THEN "Before" ELSE "After" END|

The calculation is valid.

Apply

OK

Figure 5.33 A calculation to distinguish between the “Before” and “After” time frames.

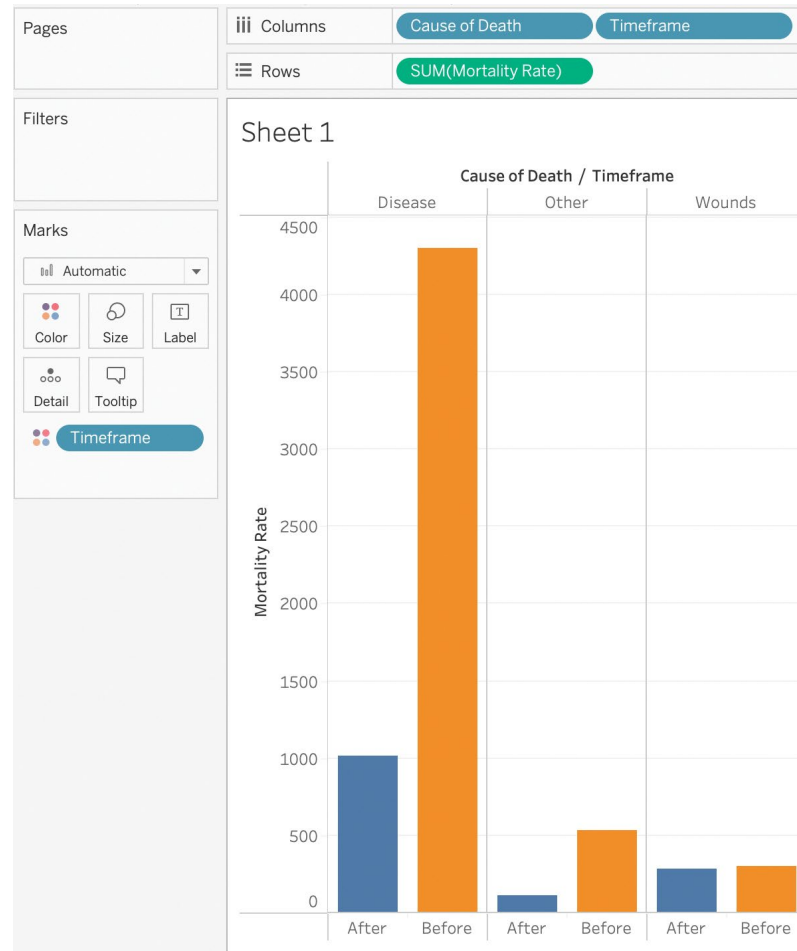


Figure 5.34 A simple, aggregate bar chart.

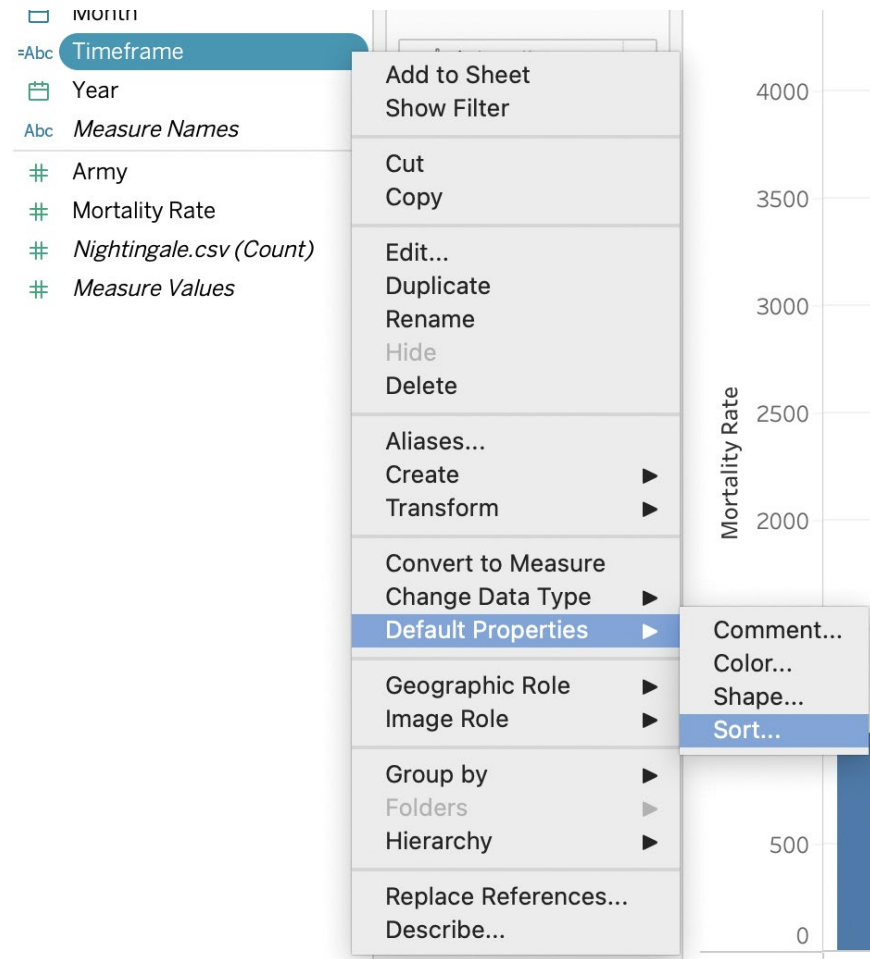


Figure 5.35 Changing the order of the horizontal axis entries.

Before vs After Sanitation by Cause of Death

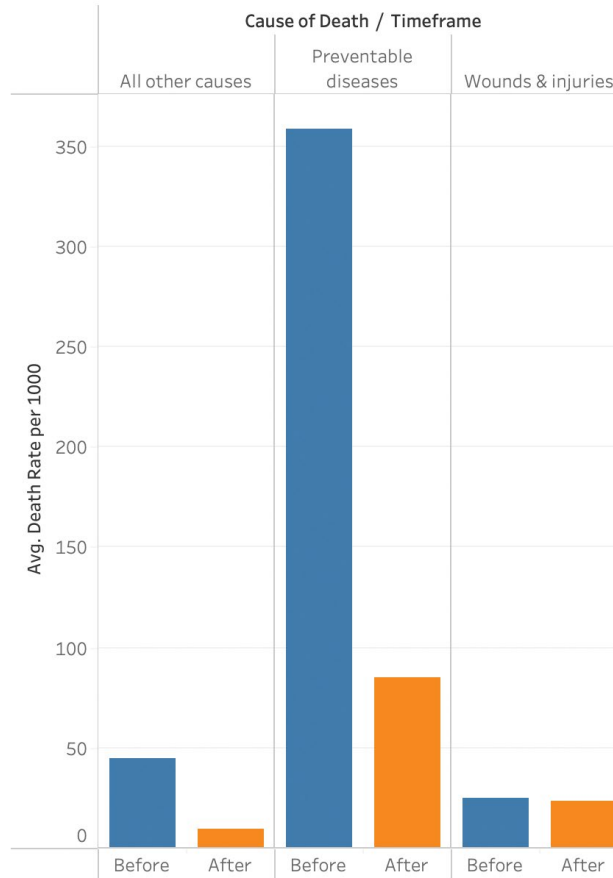


Figure 5.36 Bar chart using average to aggregate the data.

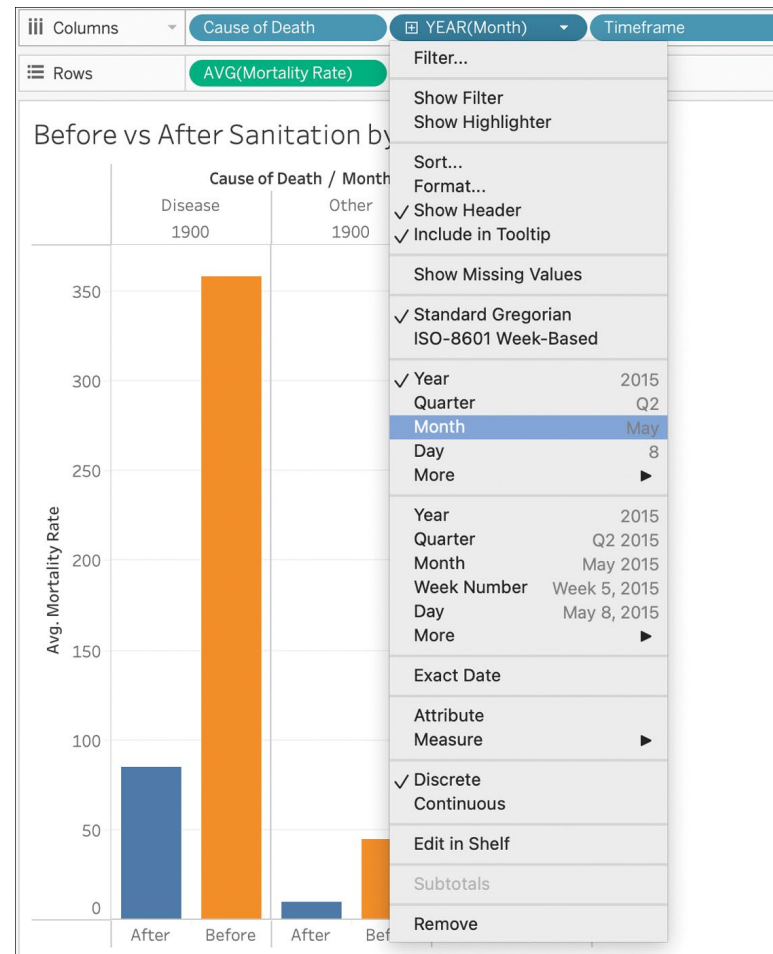


Figure 5.37 Changing how the Month field is displayed.

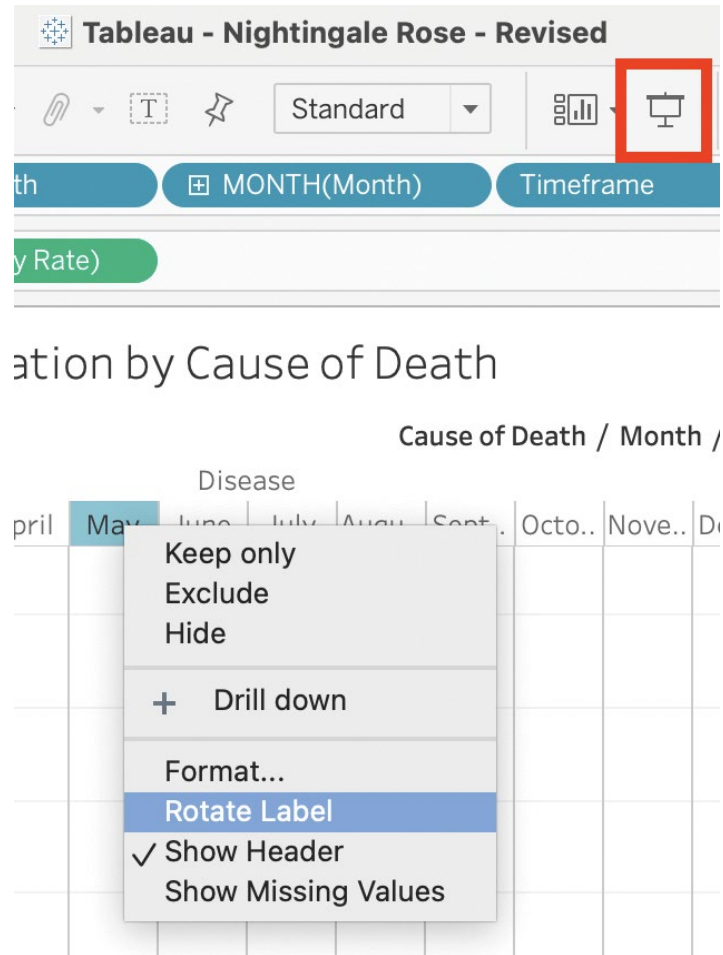


Figure 5.38 Switching to presentation mode and rotating the data labels.

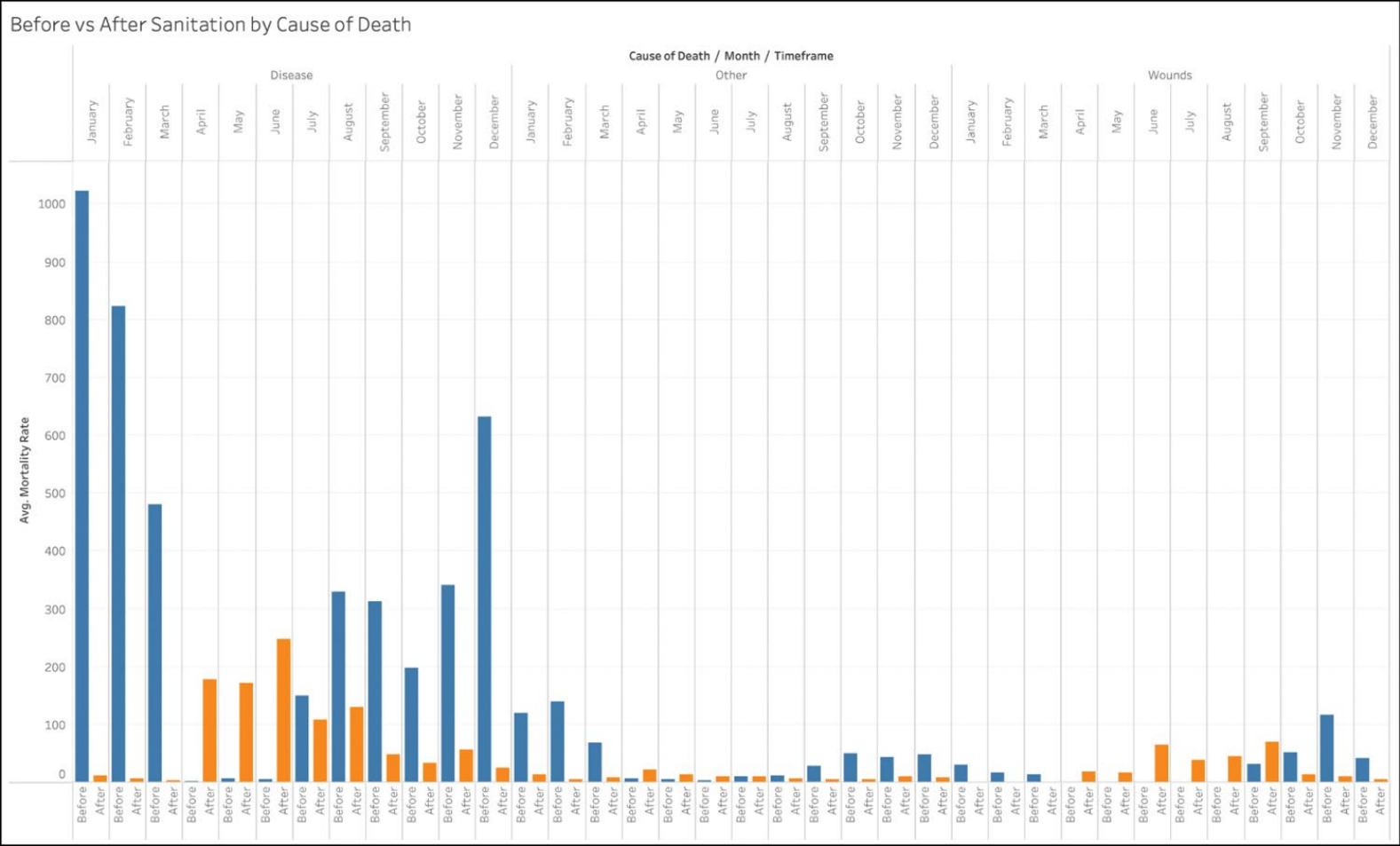


Figure 5.39 Bar graph comparing the before-sanitation to after-sanitation death rates by month.

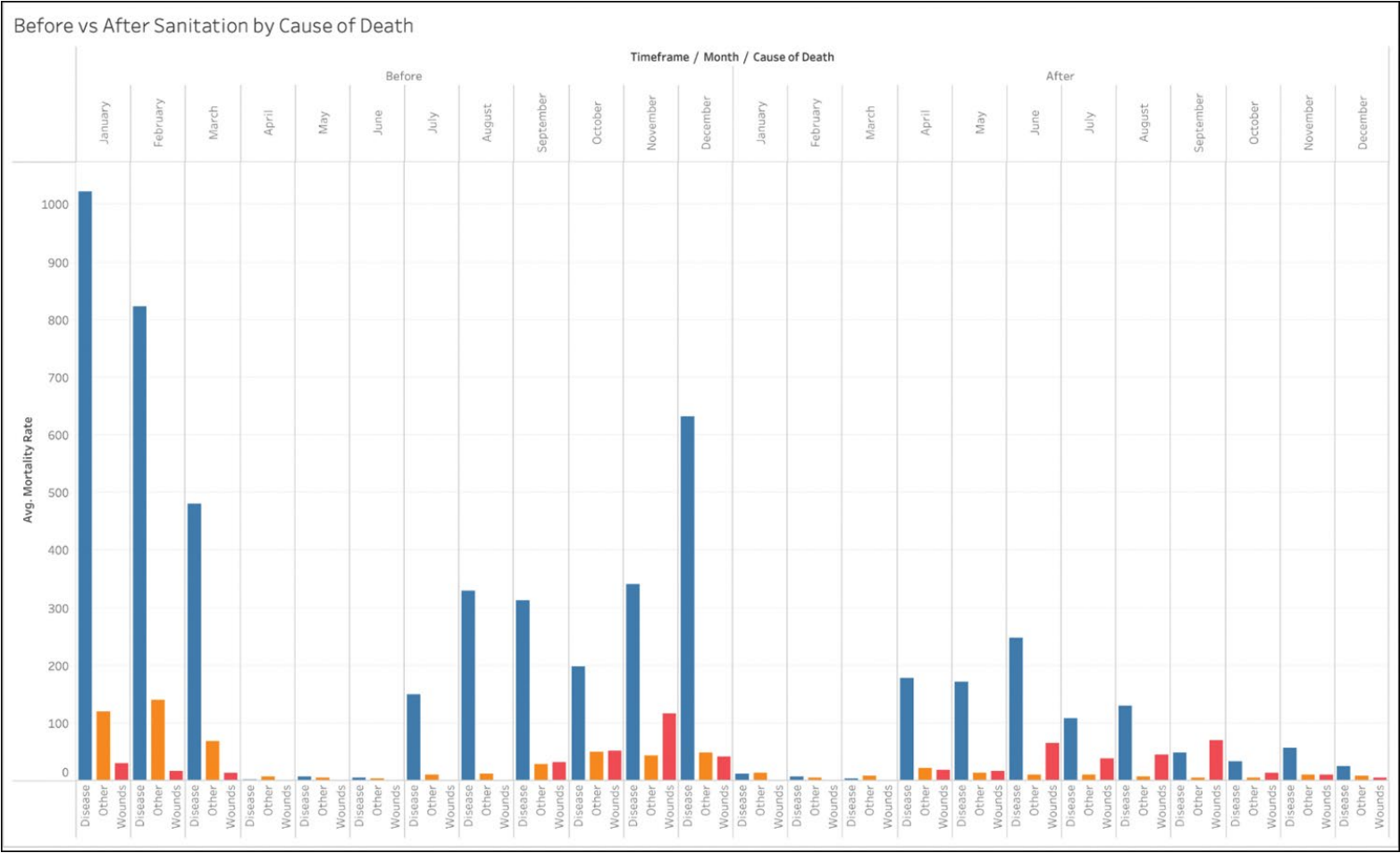


Figure 5.40 Bar chart organized by month, then cause of death.



Figure 5.41 A quick stacked area chart.

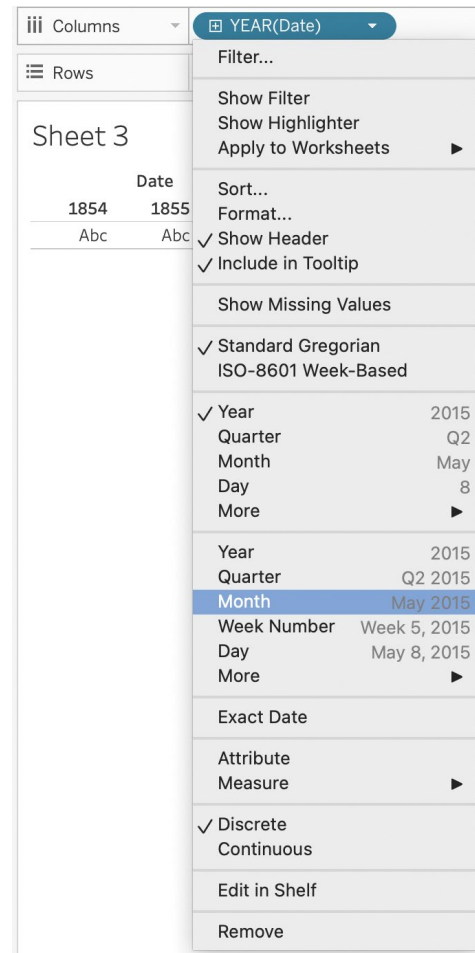


Figure 5.42 Choosing display date options for the Date field.

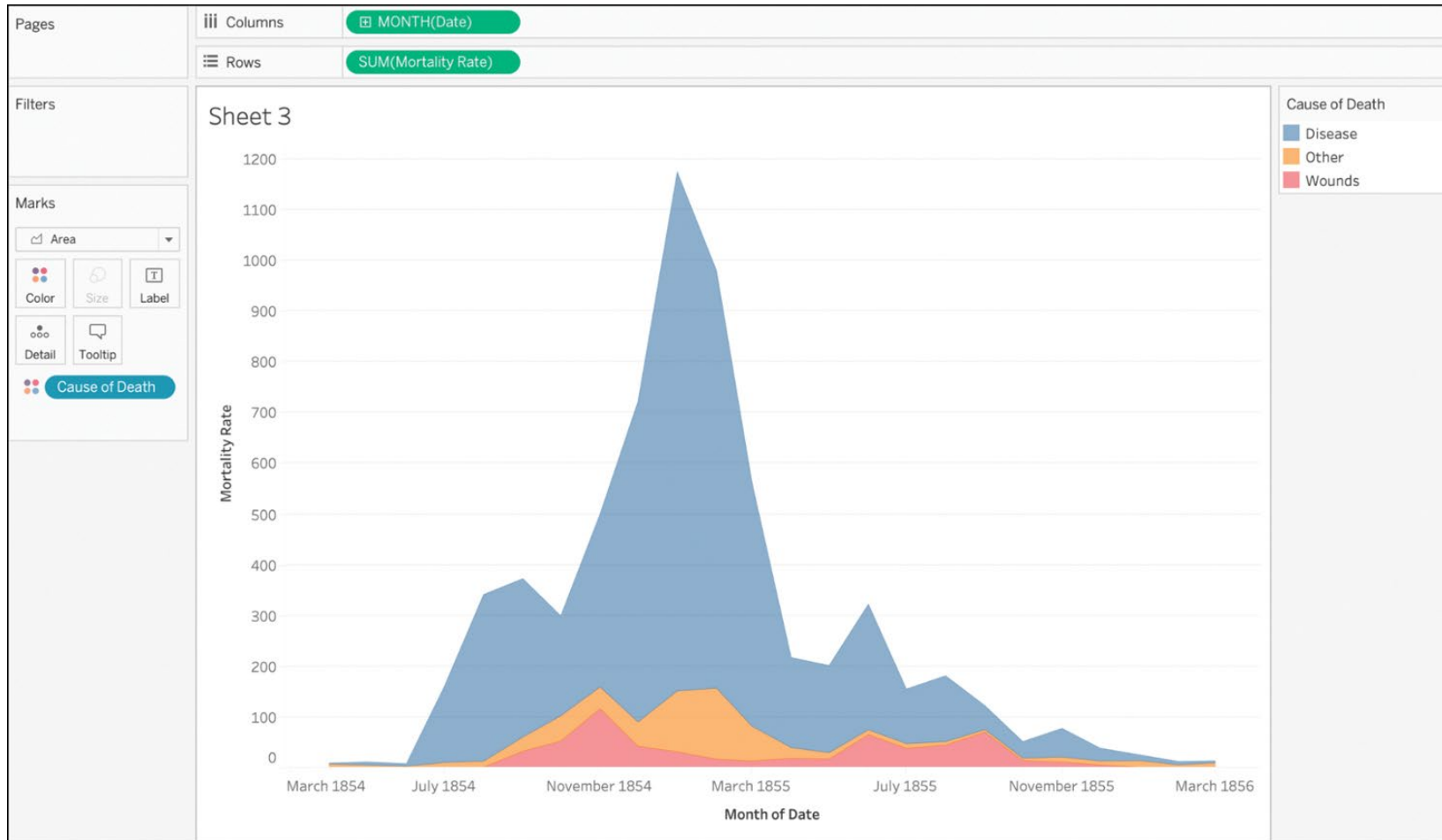


Figure 5.43 A stacked area chart.

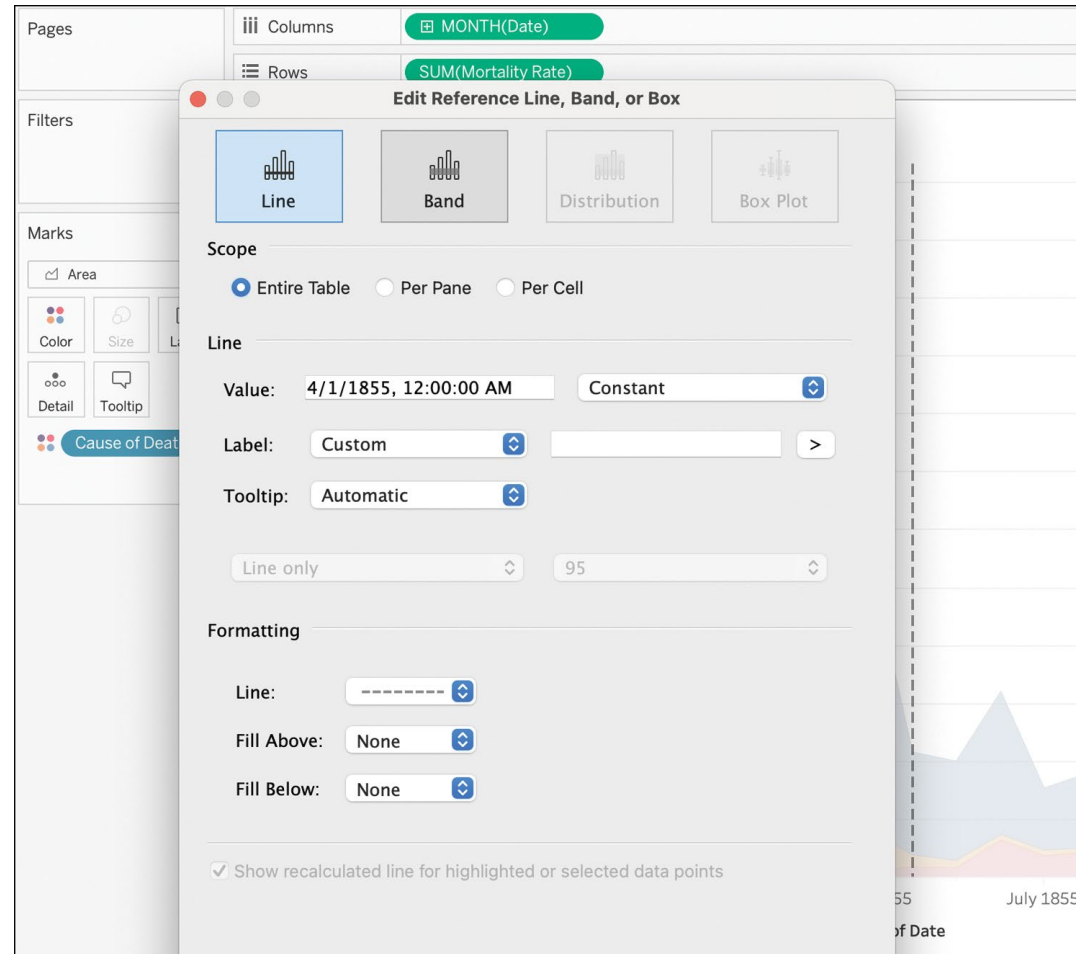


Figure 5.44 Adding a divider.

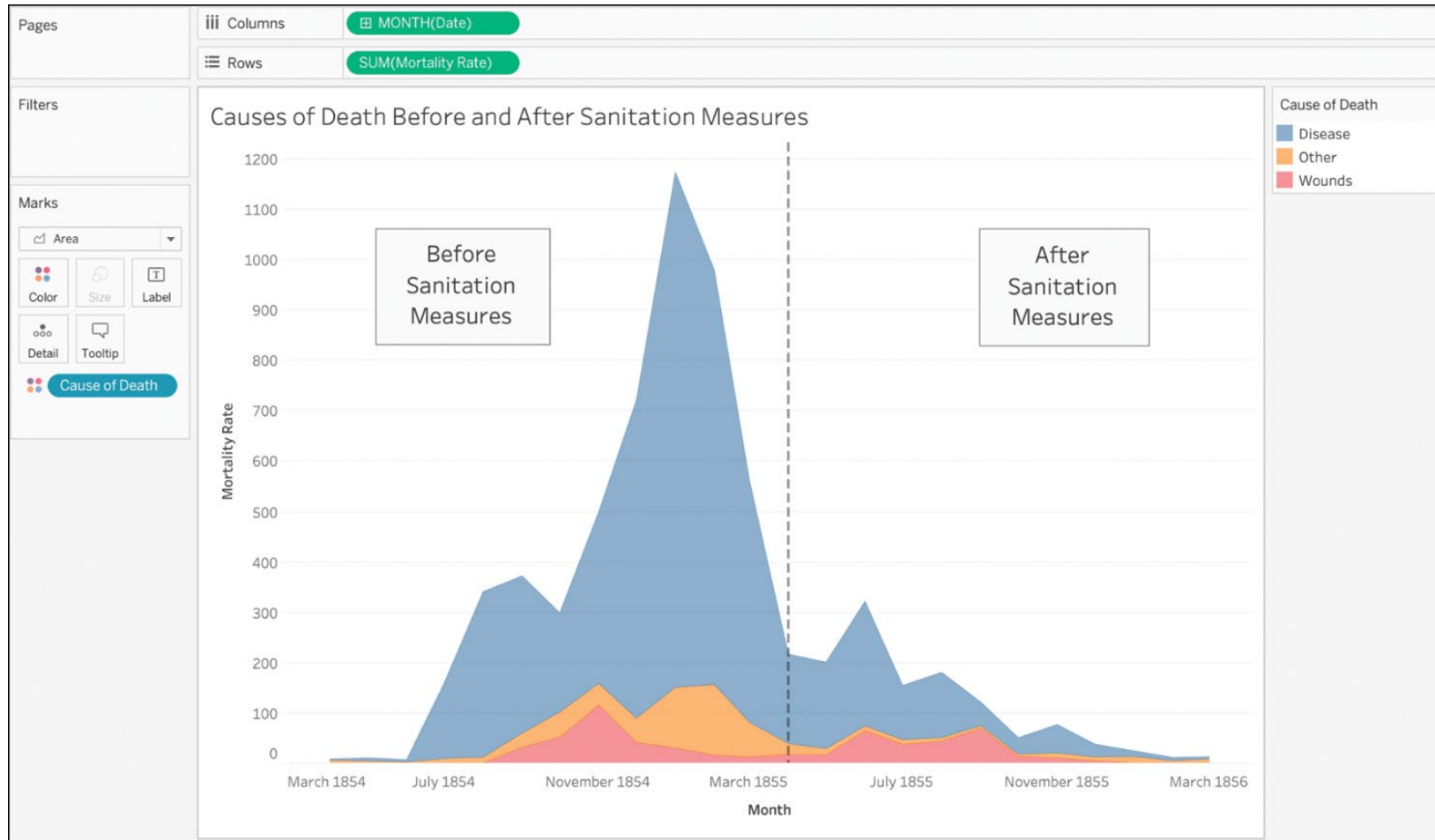


Figure 5.45 The final stacked area chart with annotations and a reference line.

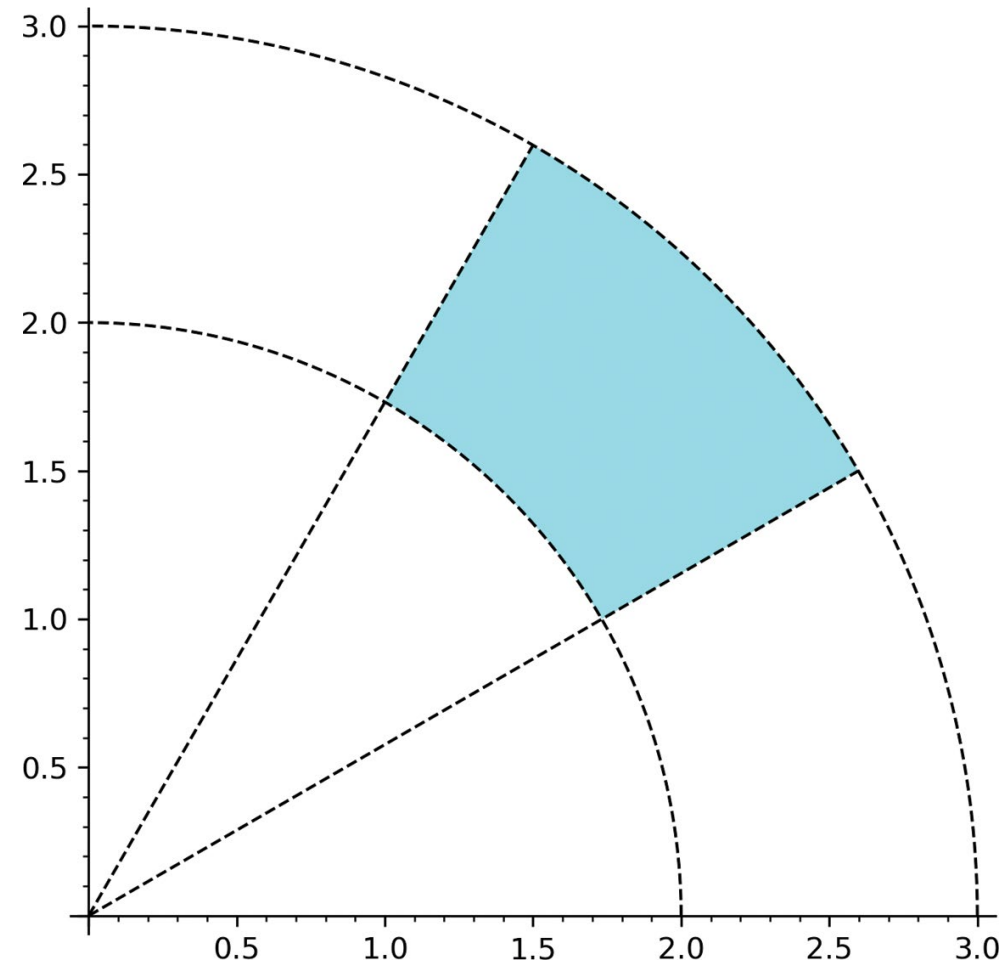


Figure 5.46 Polar rectangles.

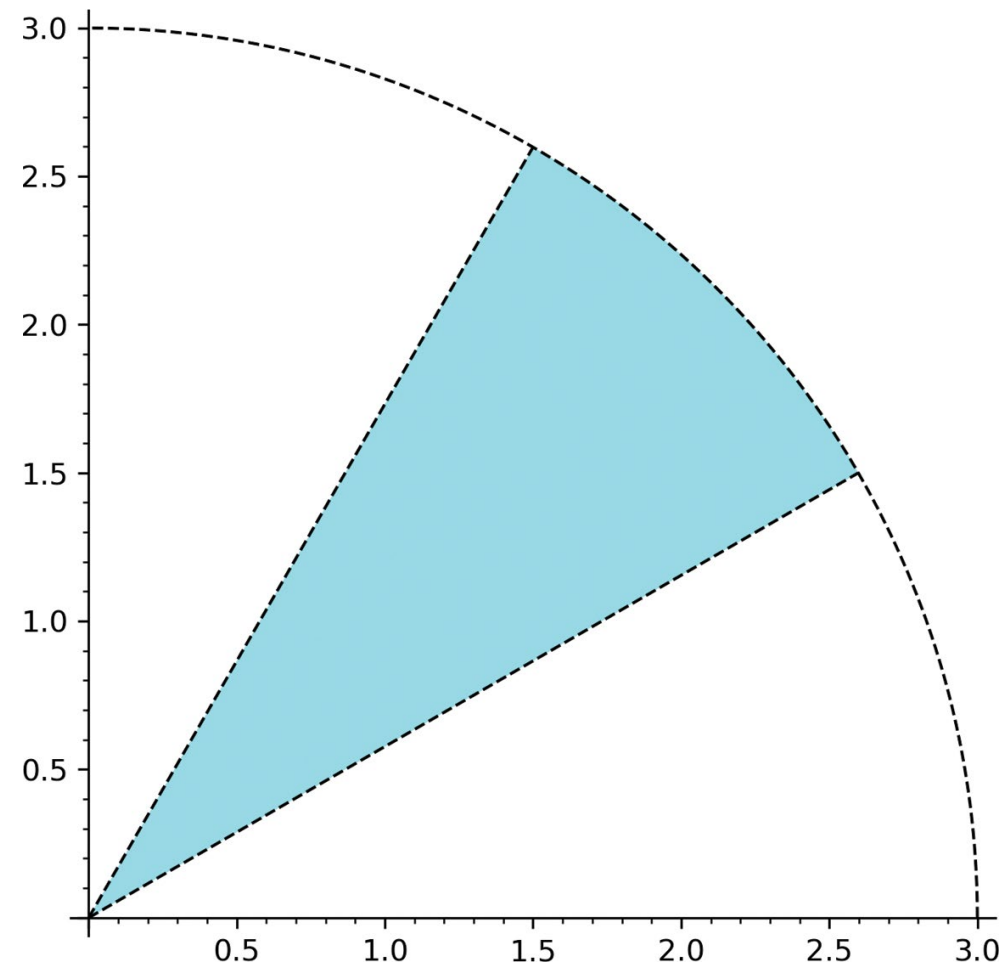


Figure 5.47 A sector.

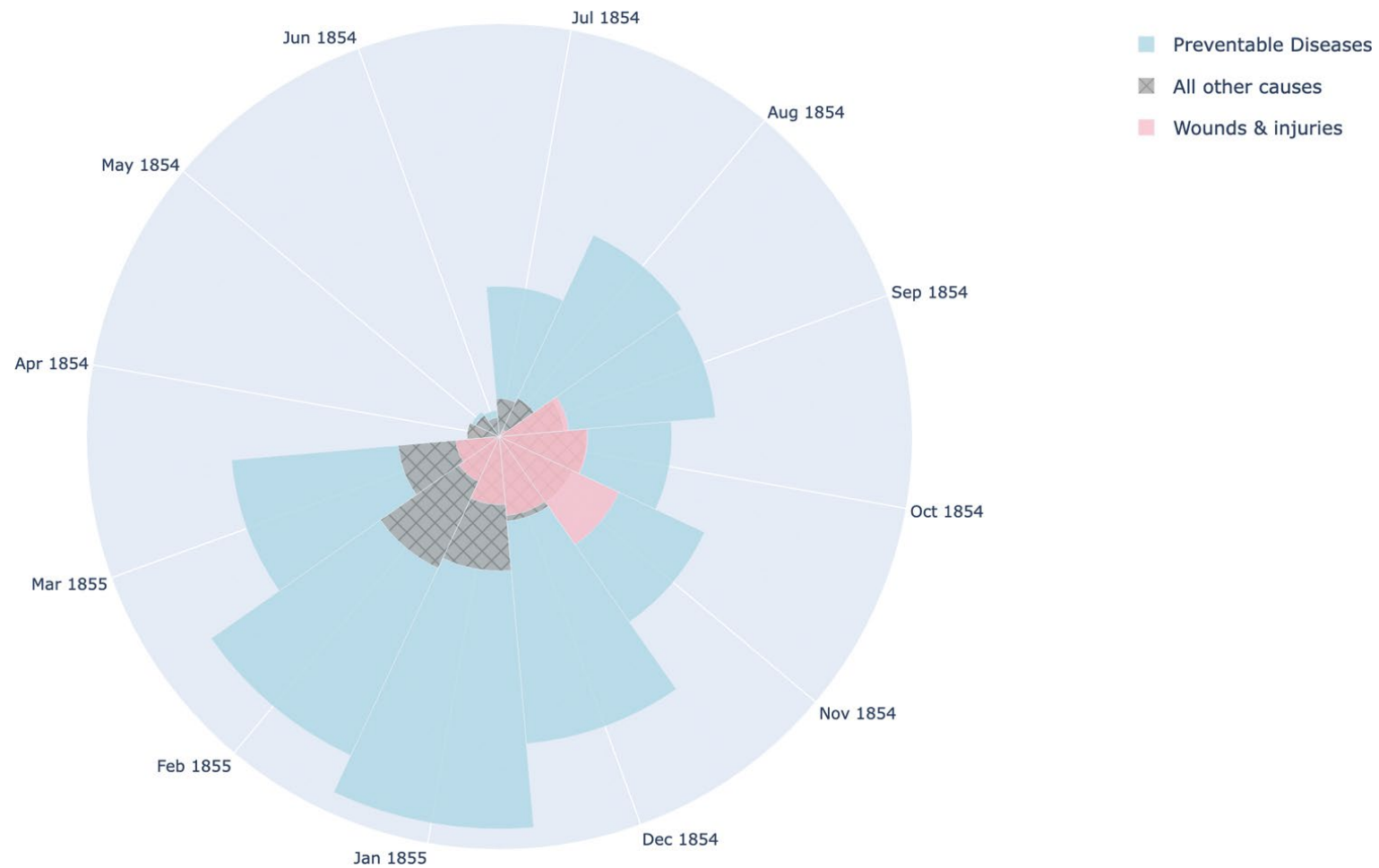


Figure 5.48 Before sanitation measures, area from center version.

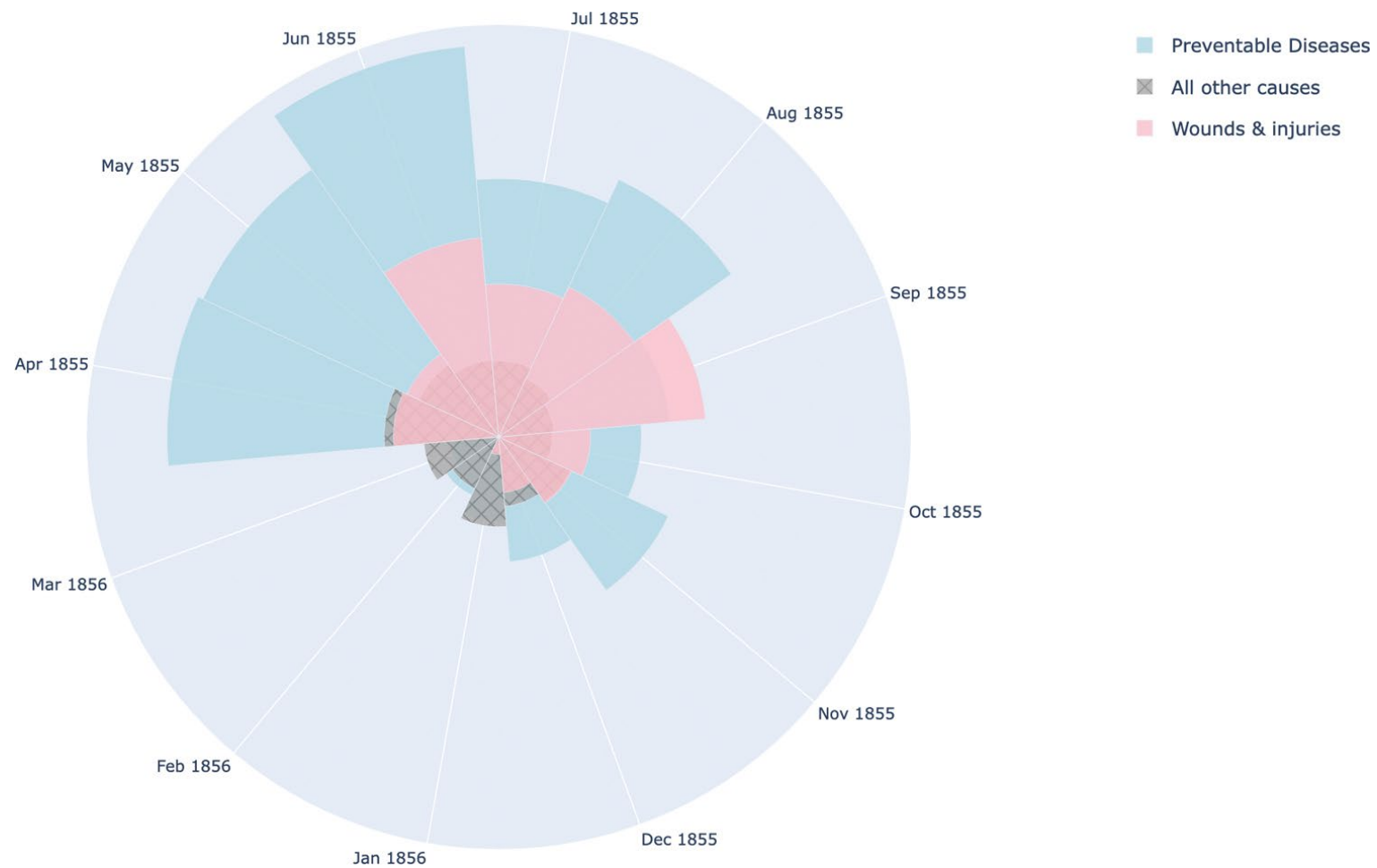


Figure 5.49 After sanitation measures, area from center version.

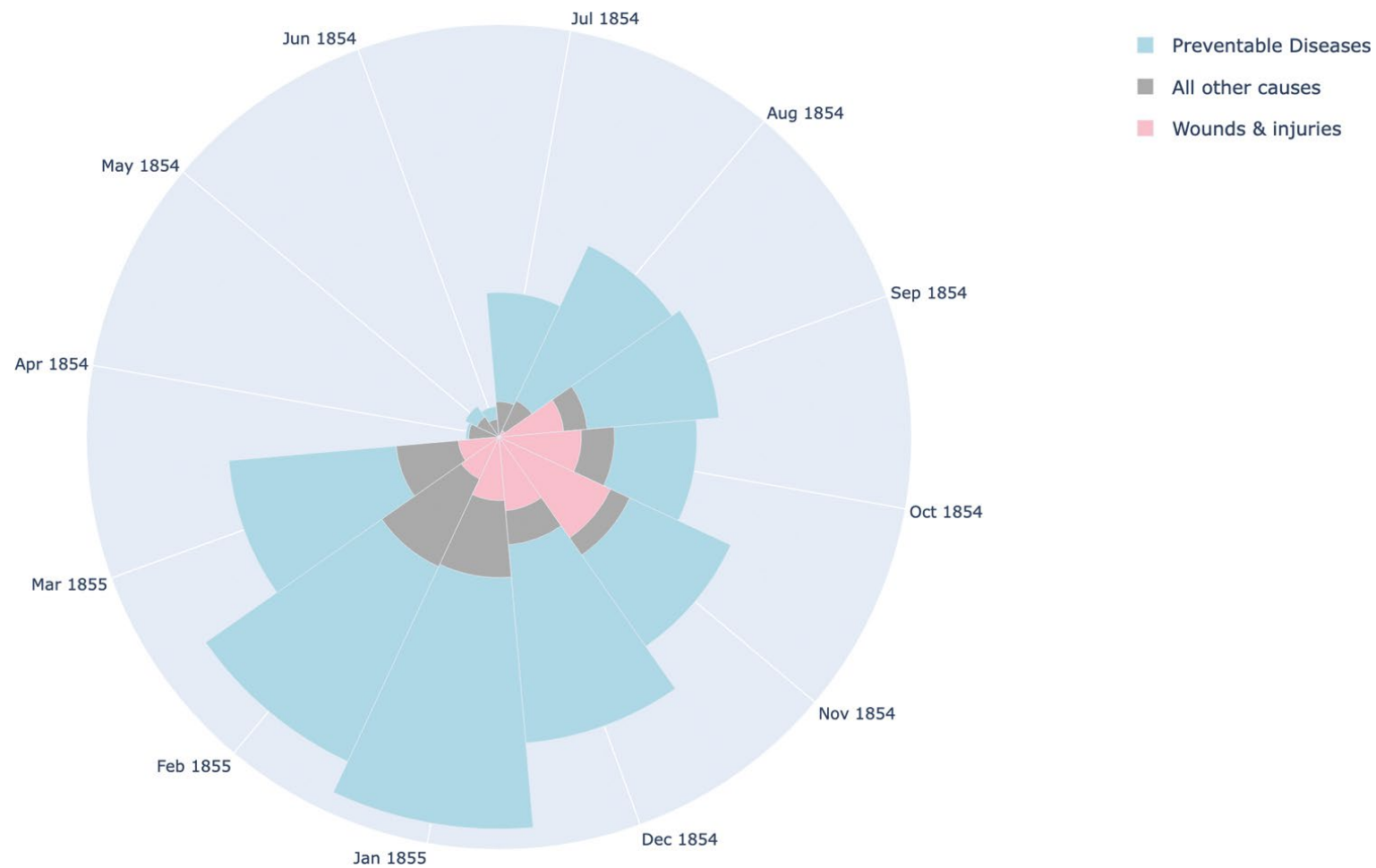


Figure 5.50 Before sanitation measures, stacked version.

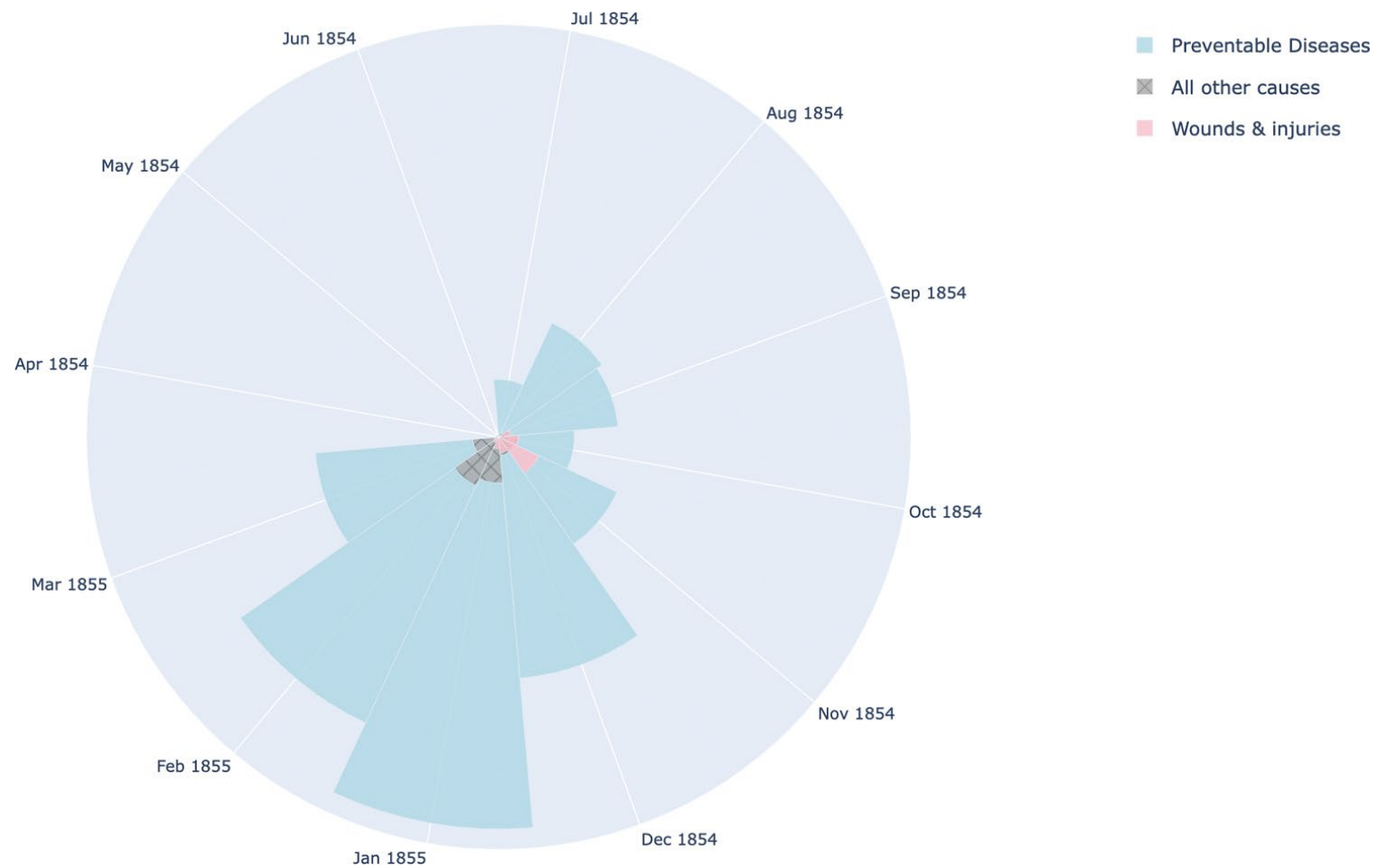


Figure 5.51 Before sanitation measures, radius from center version.

Causes of Mortality (Stacked Histogram Version)

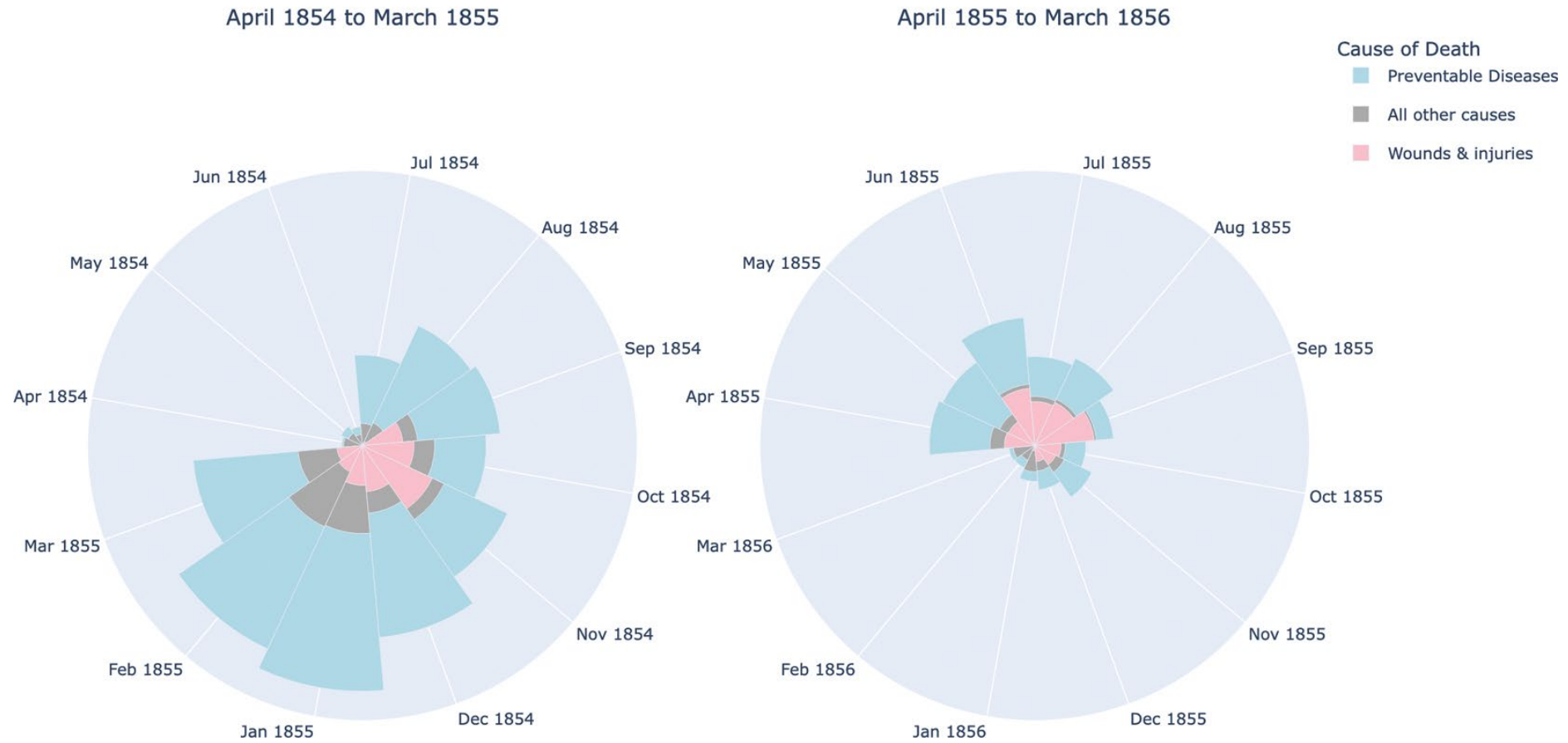


Figure 5.52 Before and after sanitation measures, stacked version.

April 1854 to March 1855

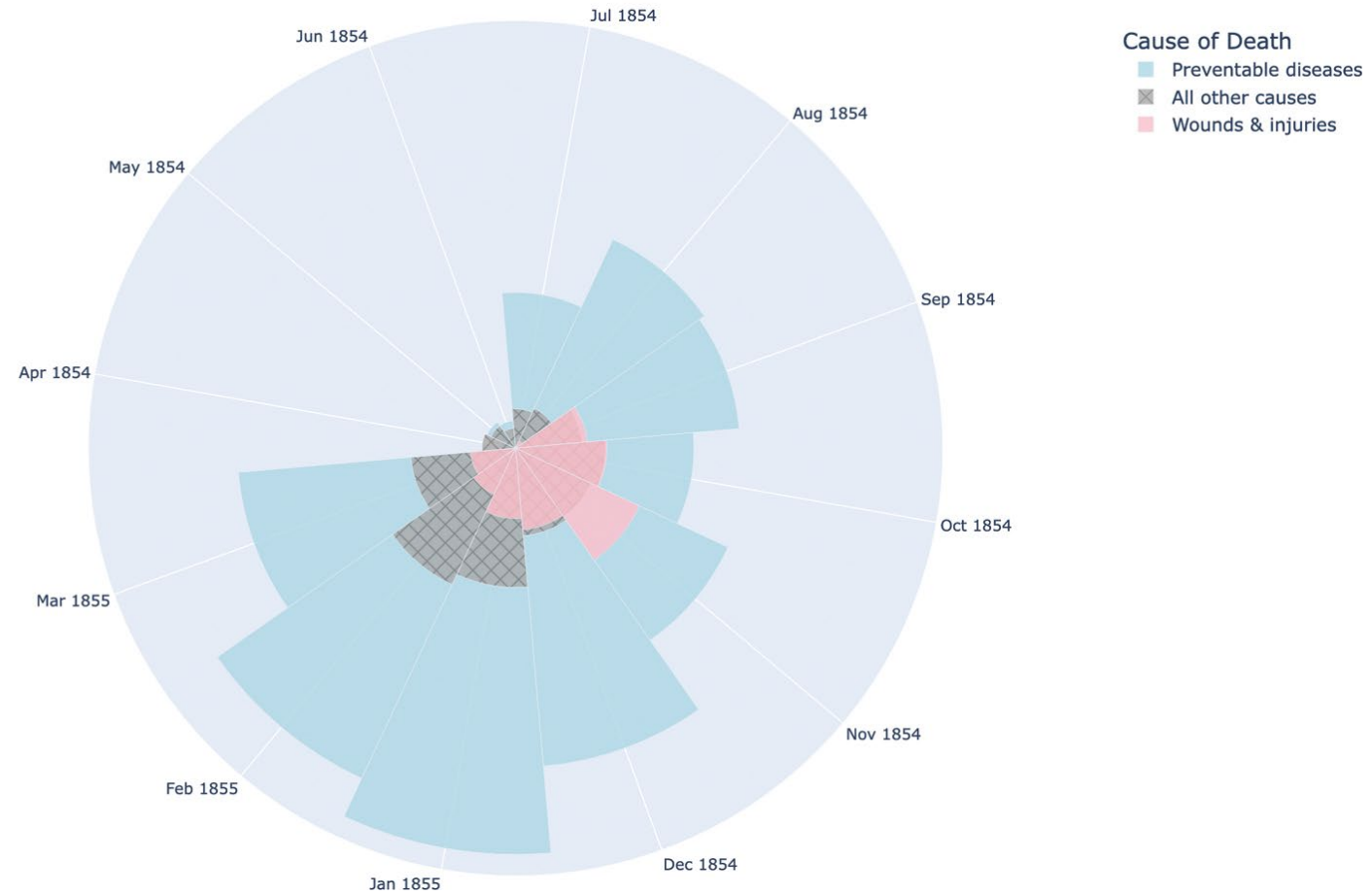


Figure 5.53 Plotly Express: before sanitation measures, area from center version.

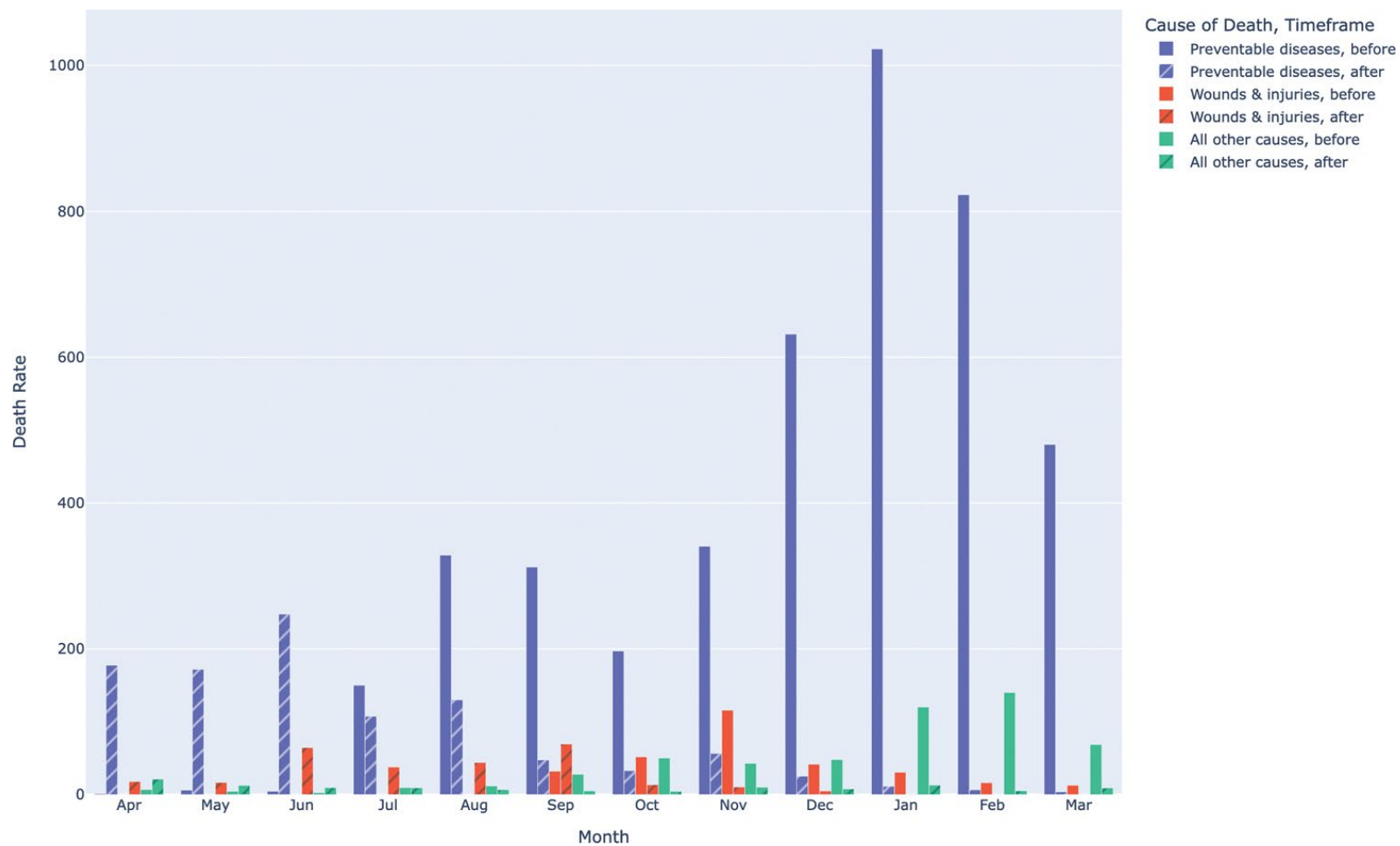


Figure 5.54 First bar chart with patterns.

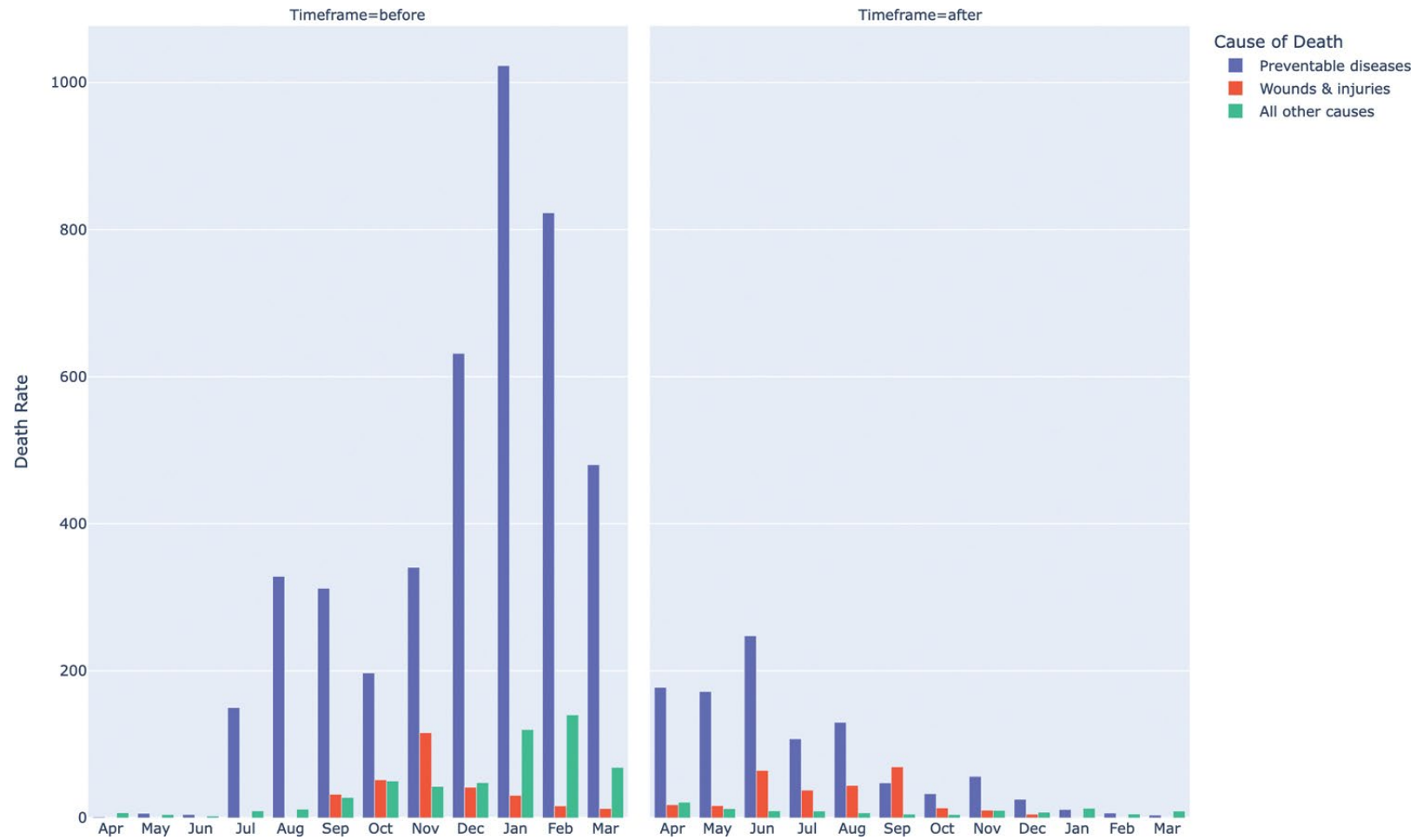


Figure 5.55 Faceted bar chart.

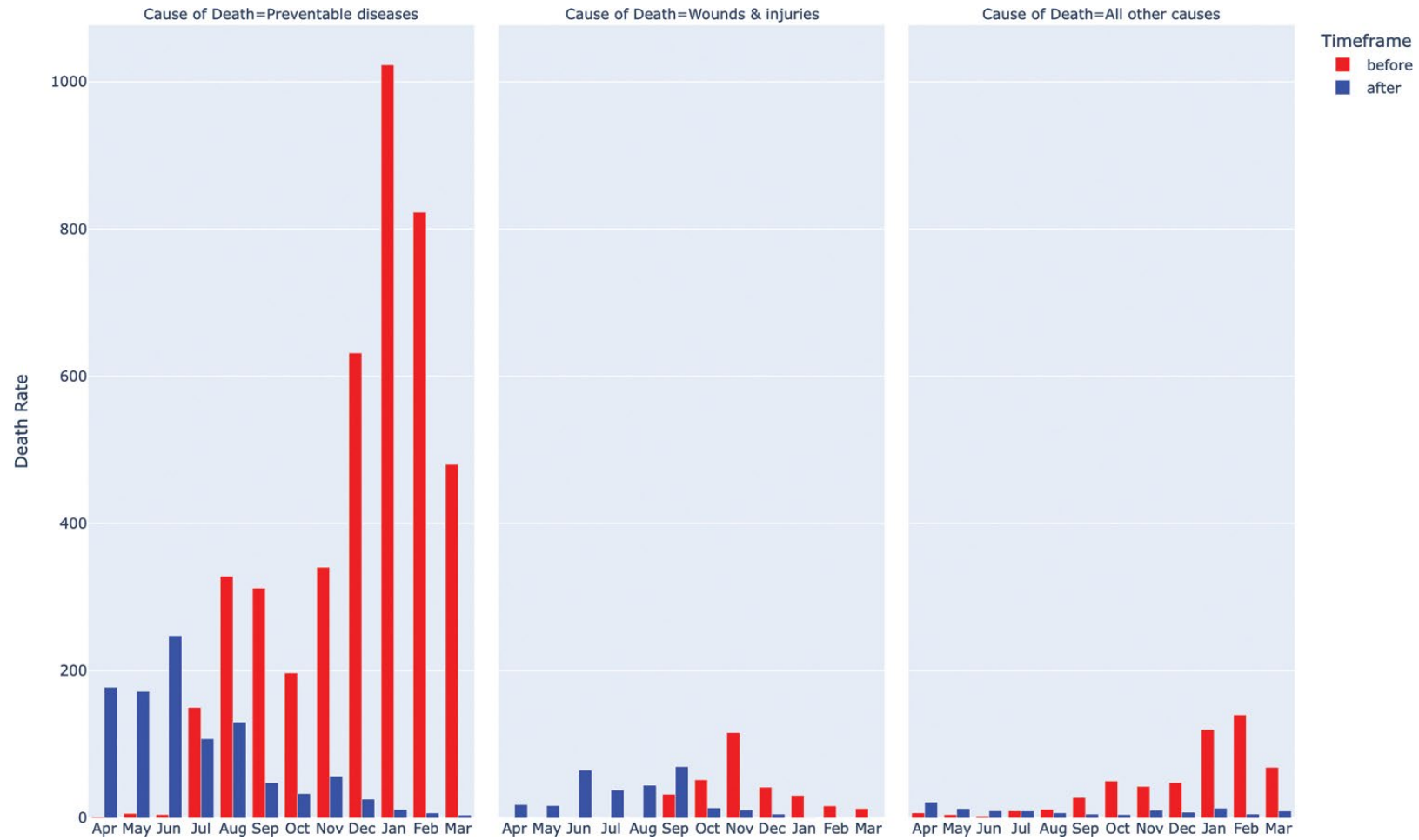


Figure 5.56 Bar chart faceted by cause of death.

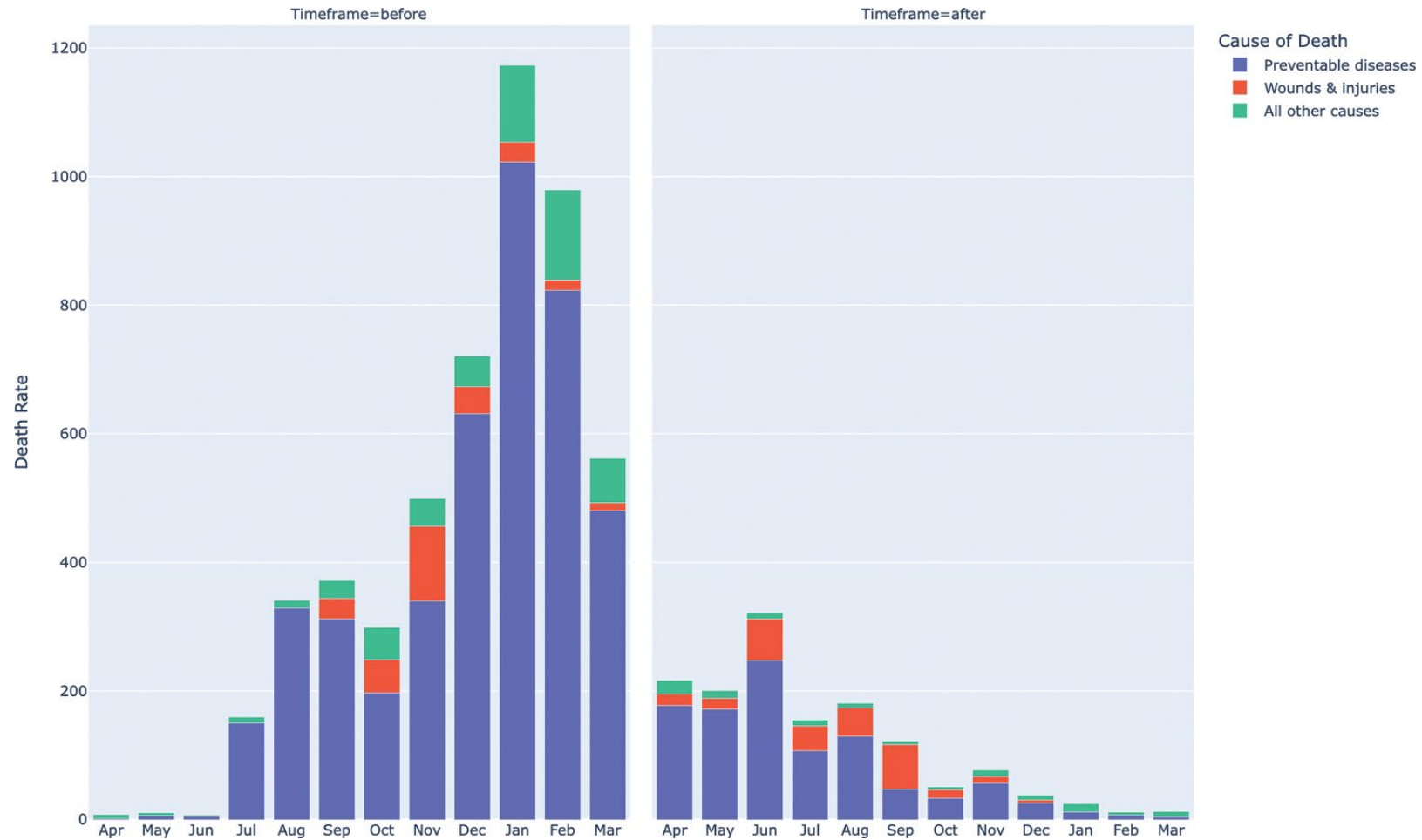


Figure 5.57 Stacked histogram.

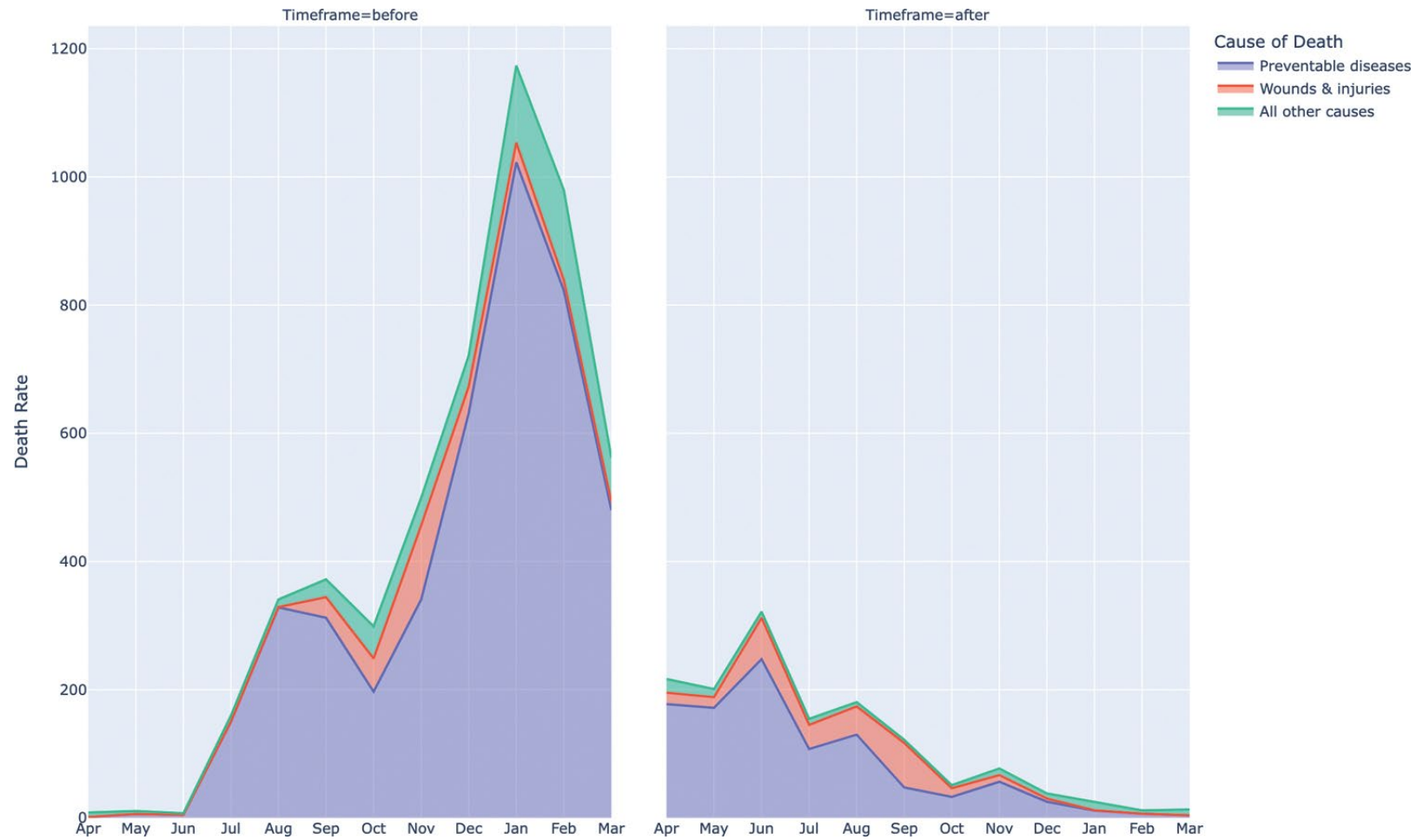


Figure 5.58 Stacked area chart.

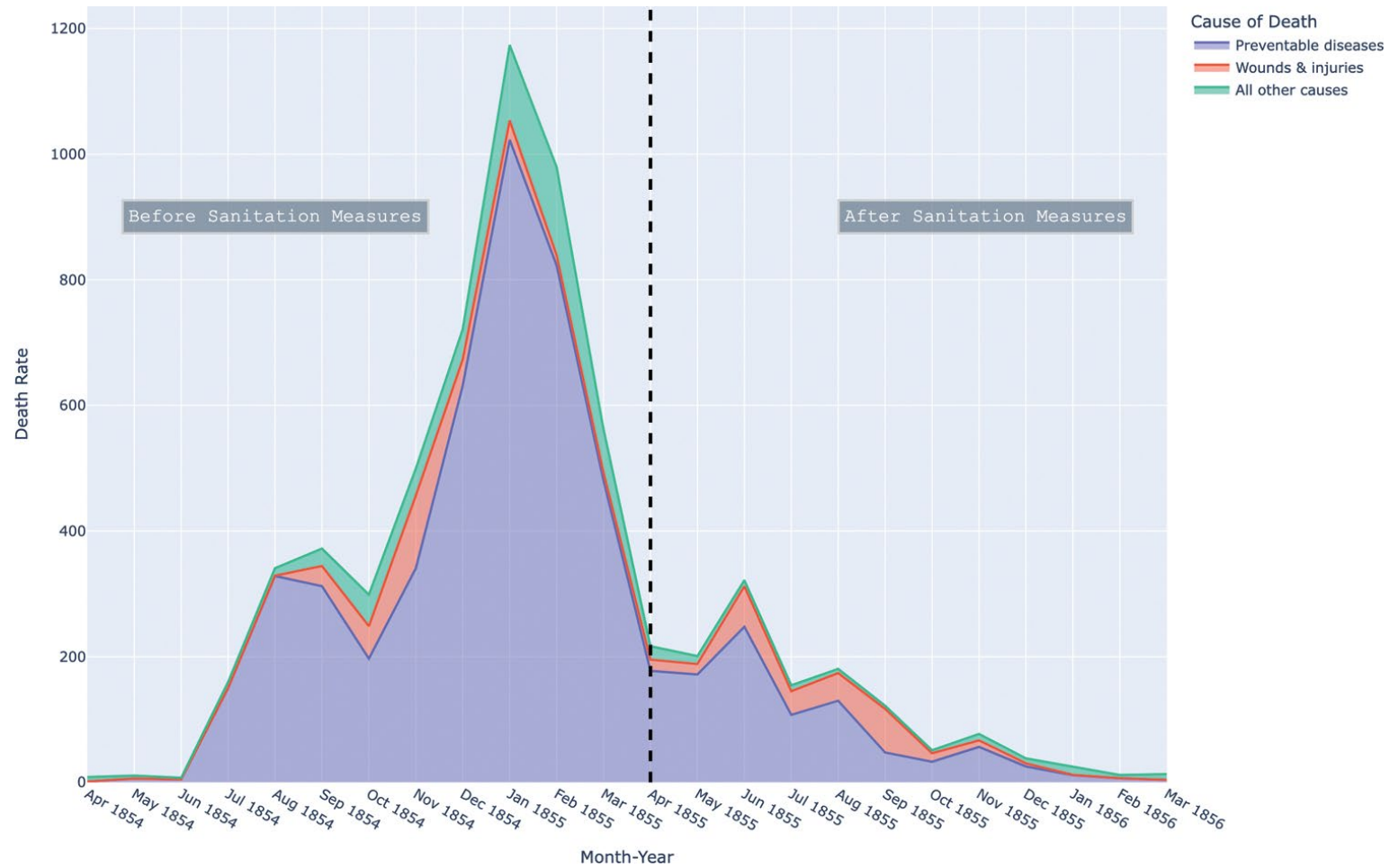


Figure 5.59 A stacked area chart with a dividing line.

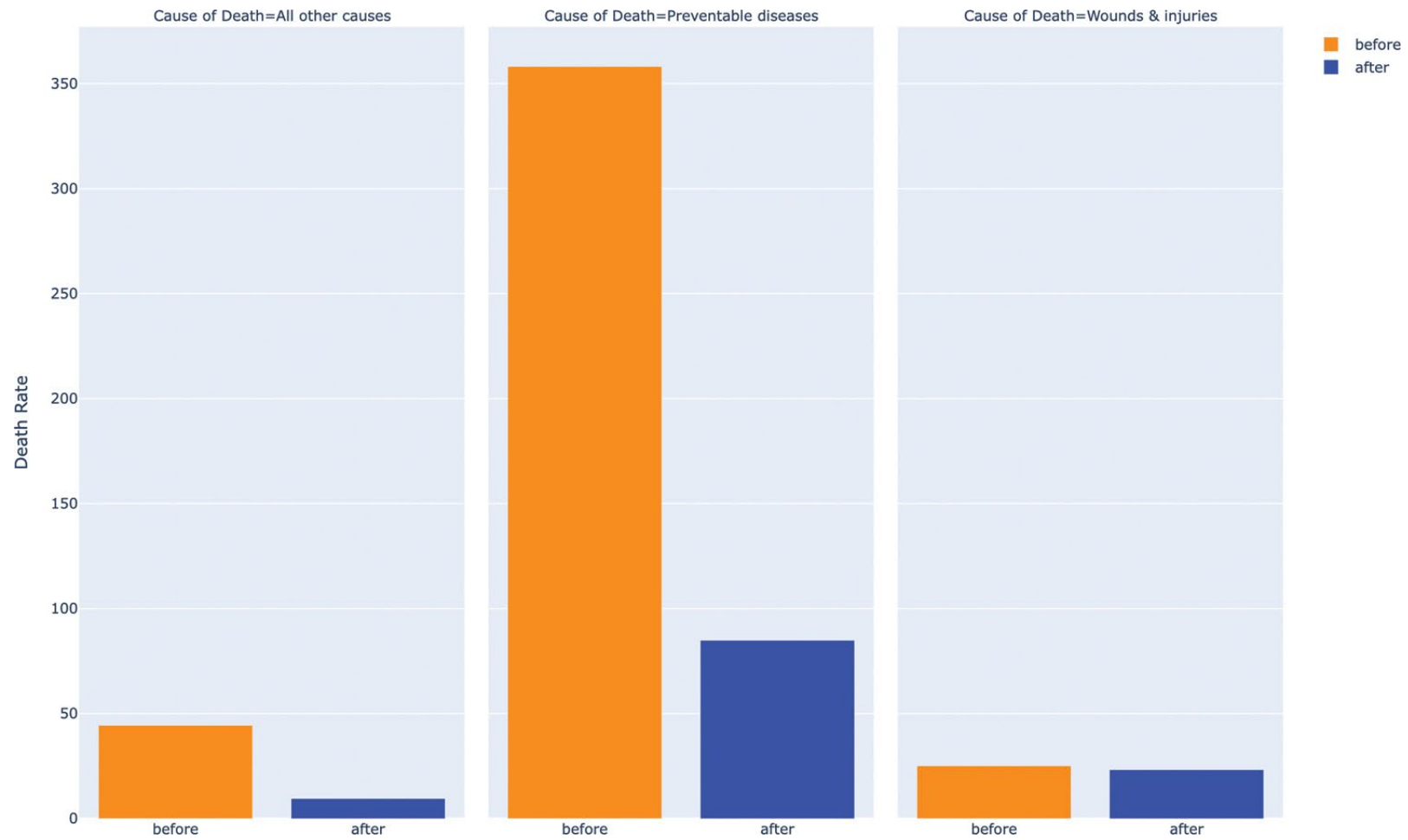


Figure 5.60 A bar chart showing aggregate data.

▲	Date	Month	Year	Army	Disease	Wounds	Other	Disease.rate	Wounds.rate	Other.rate
1	1854-04-01	Apr	1854	8571	1	0	5	1.4	0.0	7.0
2	1854-05-01	May	1854	23333	12	0	9	6.2	0.0	4.6
3	1854-06-01	Jun	1854	28333	11	0	6	4.7	0.0	2.5
4	1854-07-01	Jul	1854	28722	359	0	23	150.0	0.0	9.6
5	1854-08-01	Aug	1854	30246	828	1	30	328.5	0.4	11.9
6	1854-09-01	Sen	1854	30290	788	81	70	312.2	32.1	27.7

Figure 5.61 Florence Nightingale's data from the HistData R package.

Nightingale.R × Nightingale × Night ×				
← → 📄 🔍 Filter				
	Date	Disease.rate	Wounds.rate	Other.rate
1	1854-04-01	1.4	0.0	7.0
2	1854-05-01	6.2	0.0	4.6
3	1854-06-01	4.7	0.0	2.5
4	1854-07-01	150.0	0.0	9.6
5	1854-08-01	328.5	0.4	11.9
6	1854-09-01	312.2	32.1	27.7

Figure 5.62 Subset of Nightingale data.frame choosing columns 1, 8, 9, and 10.

Nightingale.R x melted x Nightingale x Night x			
Filter			
	Date	variable	value
20	1855-11-01	Disease.rate	56.4
21	1855-12-01	Disease.rate	25.3
22	1856-01-01	Disease.rate	11.4
23	1856-02-01	Disease.rate	6.6
24	1856-03-01	Disease.rate	3.9
25	1854-04-01	Wounds.rate	0.0
26	1854-05-01	Wounds.rate	0.0
27	1854-06-01	Wounds.rate	0.0
28	1854-07-01	Wounds.rate	0.0
29	1854-08-01	Wounds.rate	0.4
30	1854-09-01	Wounds.rate	32.1
31	1854-10-01	Wounds.rate	51.7

Figure 5.63 Melt the data by Date to create tidy data for R.

Nightingale.R × melted × Nightir			
← → 📄 🔍 Filter			
	Date	Cause	Deaths
1	1854-04-01	Disease	1.4
2	1854-05-01	Disease	6.2
3	1854-06-01	Disease	4.7
4	1854-07-01	Disease	150.0
5	1854-08-01	Disease	328.5
6	1854-09-01	Disease	312.2
7	1854-10-01	Disease	197.0
8	1854-11-01	Disease	340.6

Figure 5.64 Renaming variables and removing “.rate”.

```
data(Nightingale)
```

Format

A data frame with 24 observations on the following 10 variables.

Date

a Date, composed as `as.Date(paste(Year, Month, 1, sep='-'), "%Y-%b-%d")`

Figure 5.65 The help page on Nightingale describes each variable, including Date.

Nightingale.R × melted × Nightingale × Night ×					
←→ 📄 🔍 Filter					
	Date	Cause	Deaths	Regime	Month
1	1854-04-01	Disease	1.4	Before	Apr 1854
2	1854-05-01	Disease	6.2	Before	May 1854
3	1854-06-01	Disease	4.7	Before	Jun 1854
4	1854-07-01	Disease	150.0	Before	Jul 1854
5	1854-08-01	Disease	328.5	Before	Aug 1854
6	1854-09-01	Disease	312.2	Before	Sen 1854

Figure 5.66 The Night data.frame with new variables Regime and Month.

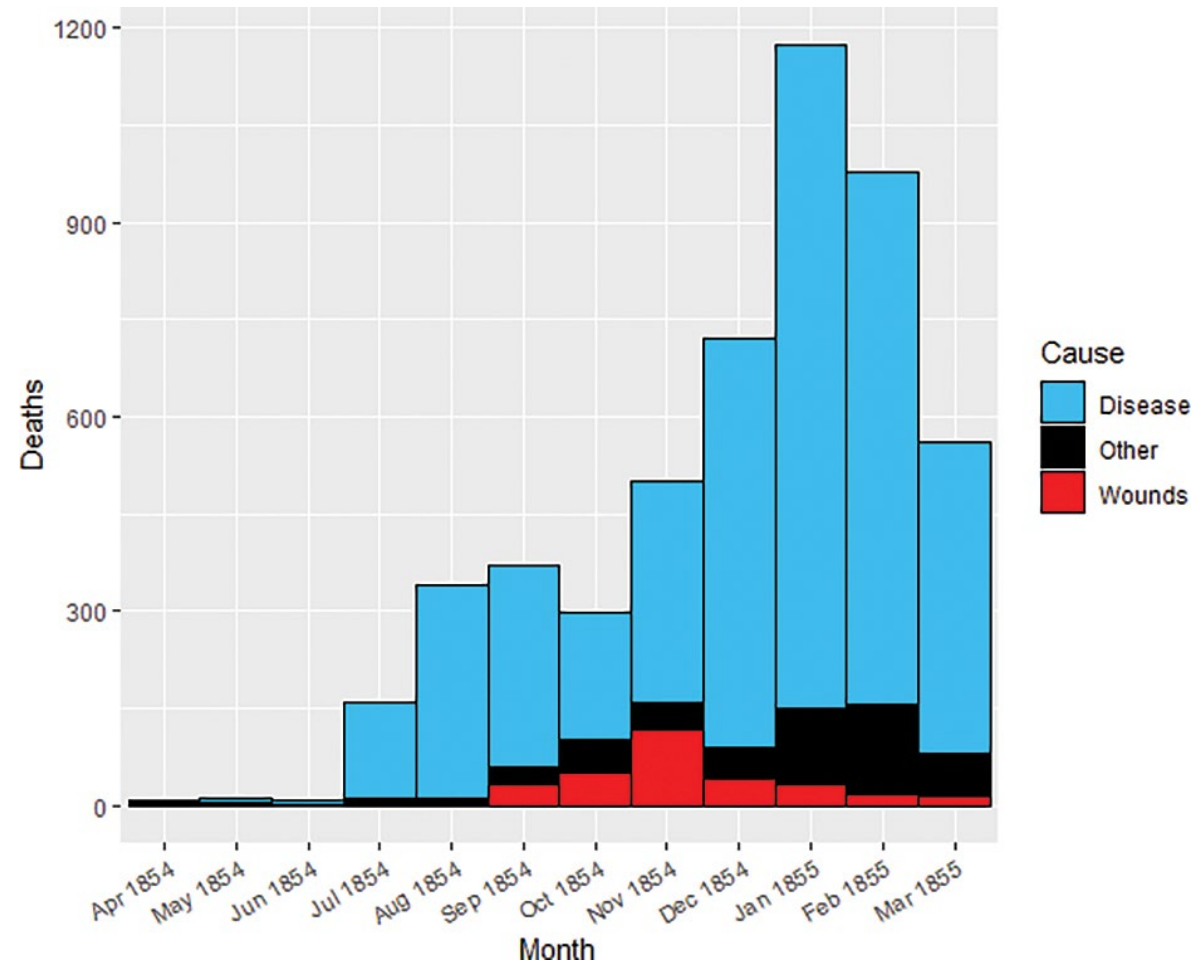


Figure 5.67 Stacked column chart version of Nightingale's Rose before sanitation created using R.

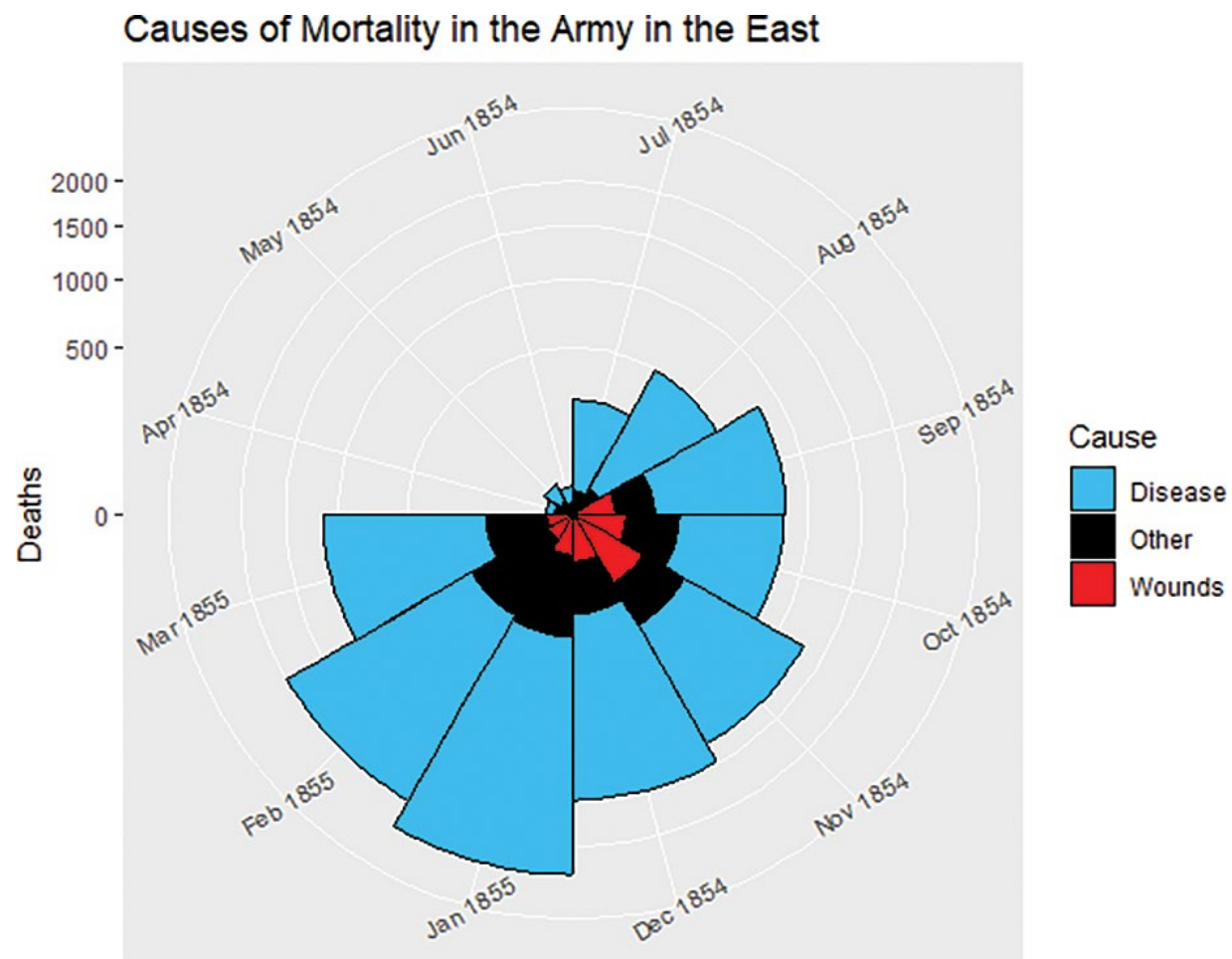


Figure 5.68 Stacked polar chart version of Nightingale's Rose before sanitation created using R.

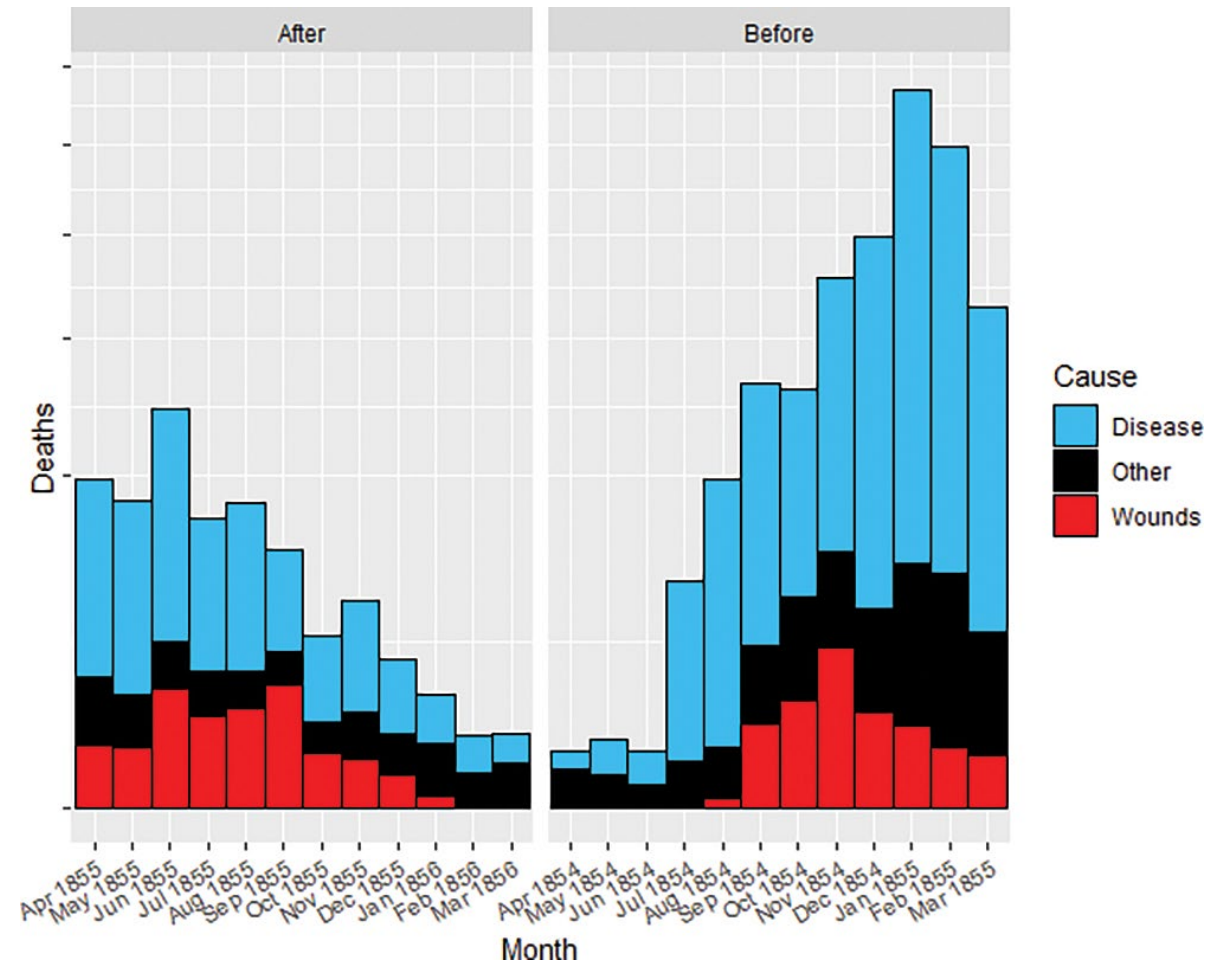


Figure 5.69 Stacked column chart version of Nightingale's Rose created using facet_grid.

Causes of Mortality in the Army in the East

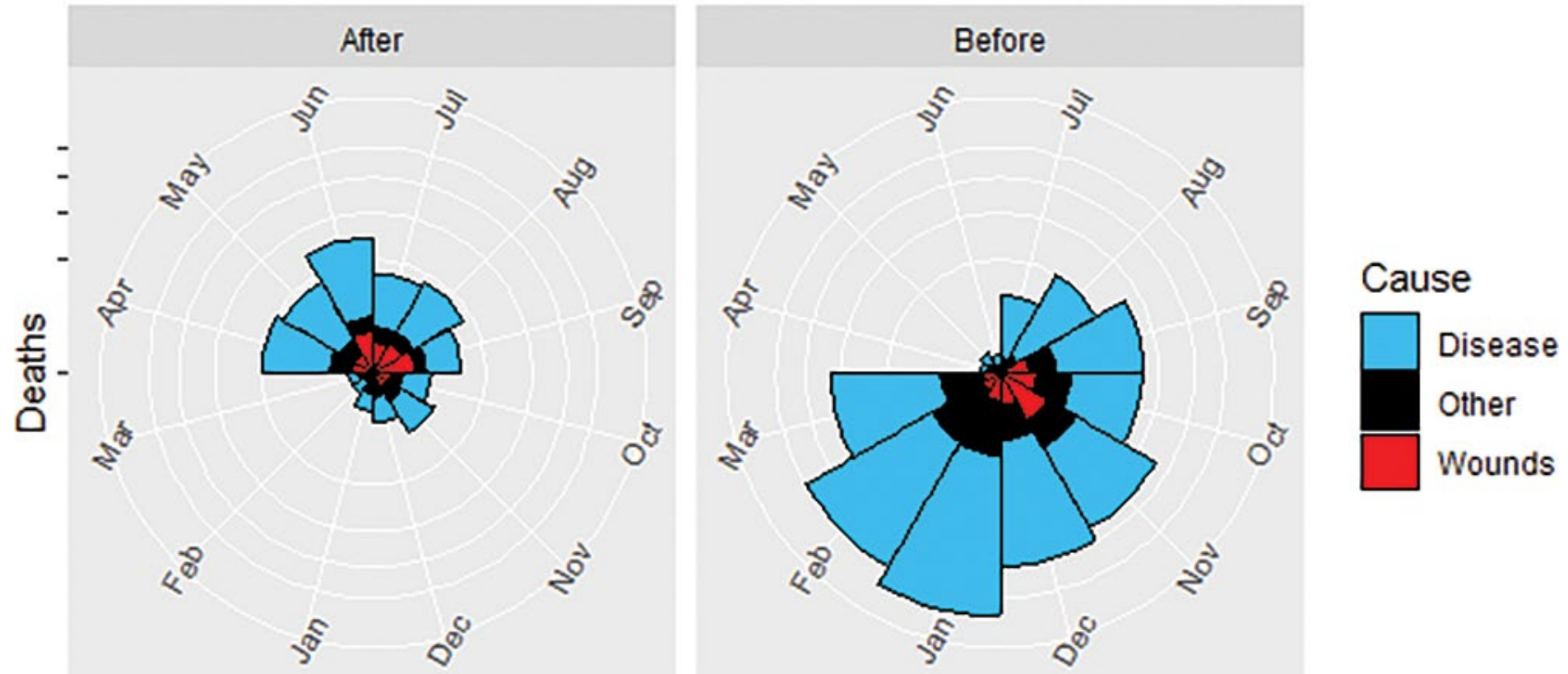


Figure 5.70 R version of Nightingale's Rose created using `facet_grid`.

	Date	Month	Year	Army	Disease	Wounds	Other	Disease.rate	Wounds.rate	Other.rate
1	1854-04-01	Apr	1854	8571	1	0	5	1.4	0.0	7.0
2	1854-05-01	May	1854	23333	12	0	9	6.2	0.0	4.6
3	1854-06-01	Jun	1854	28333	11	0	6	4.7	0.0	2.5
4	1854-07-01	Jul	1854	28722	359	0	23	150.0	0.0	9.6
5	1854-08-01	Aug	1854	30246	828	1	30	328.5	0.4	11.9

Figure 5.71 The Nightingale data as a two-way table.

	Date	variable	value
21	1855-12-01	Disease.rate	25.3
22	1856-01-01	Disease.rate	11.4
23	1856-02-01	Disease.rate	6.6
24	1856-03-01	Disease.rate	3.9
25	1854-04-01	Wounds.rate	0.0
26	1854-05-01	Wounds.rate	0.0
27	1854-06-01	Wounds.rate	0.0
28	1854-07-01	Wounds.rate	0.0
29	1854-08-01	Wounds.rate	0.4
30	1854-09-01	Wounds.rate	32.1
31	1854-10-01	Wounds.rate	51.7
32	1854-11-01	Wounds.rate	115.8
33	1854-12-01	Wounds.rate	41.7

Figure 5.72 The Nightingale data as tidy data in R after being melted.



Figure 5.73 Tableau calculation to distinguish time frames.

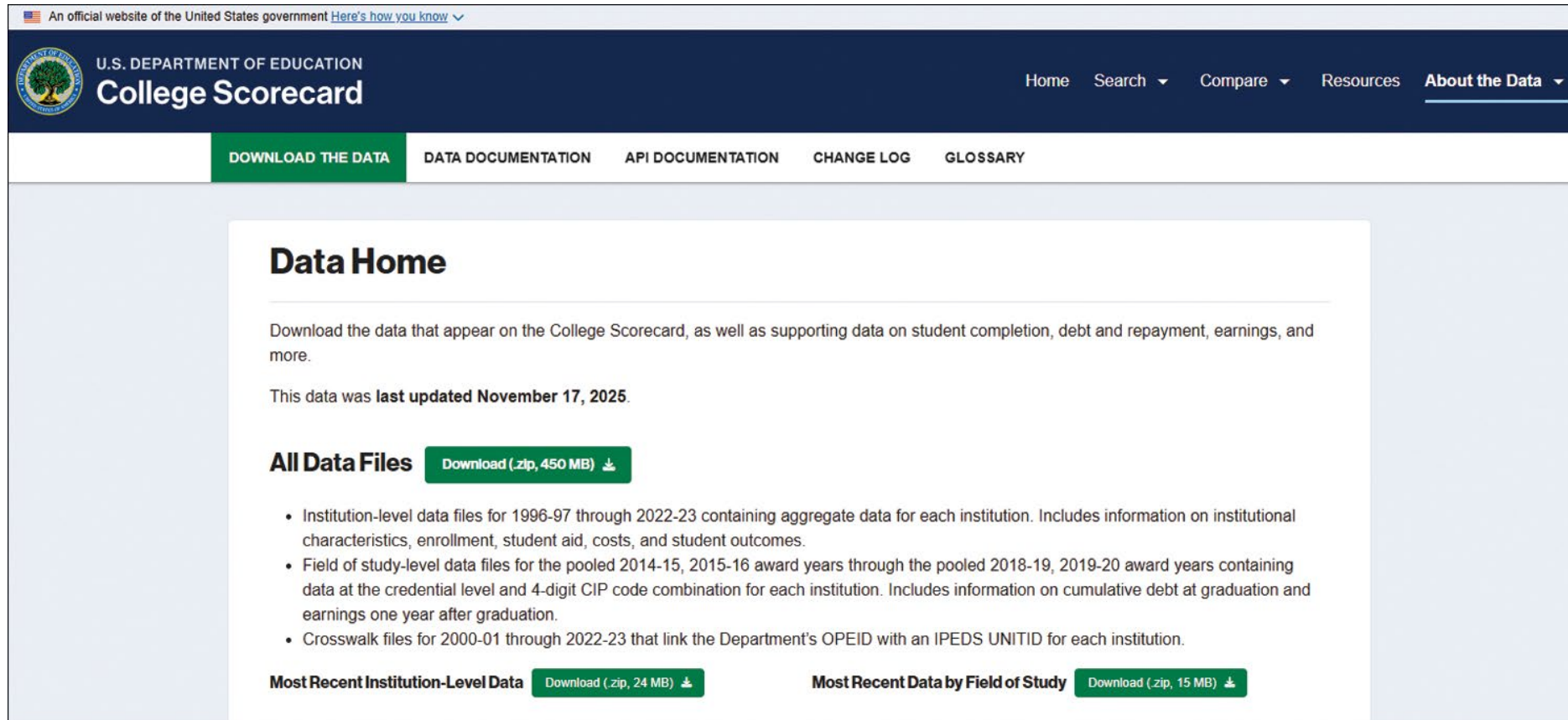


Figure 6.1 College Scorecard webpage interface for downloading the data.

AutoSaveOff

Most-Recent-Cohorts-Institution_202...

Saved to this PC

Search

FileHomeInsertDrawPage LayoutFormulasDataReviewViewAutomateHelpAcrobat

Paste

Clipboard

Aptos Narrow11A

B

I

U

Font

General

Conditional Formatting

Format as Table

Cell Styles

Alignment

Number

Styles

Insert

Delete

Format

Cells

Editing

Sensitivity

Add-ins

Copilot

Create PDF and Share link

Create PDF and Share via Outlook

Comments

Share

A1

UNITID

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	
1	UNITID	OPEID	OPEID6	INSTNM	CITY	STABBR	ZIP	ACCRED	INSTURL	NPCURL	SCH_DEG	HCM2	MAIN	NUMBRAN	PREDDEG	HIGHDEG	CONTROL	ST_FIPS	REGION	LOCALE	LOCALE2	LATITUDE	LONGITUDE	CCBASIC	CCU
2	100654	100200	1002	Alabama	Normal	AL	35762	Southern	/www.aam	www.aam	3	0	1	1	3	4	1	1	5	12	NA	34.78337	-86.5685	18	
3	100663	105200	1052	University	Birmingham	AL	35294-011	Southern	/https://ww	https://tcc	3	0	1	1	3	4	1	1	5	12	NA	33.5057	-86.7993	15	
4	100690	2503400	25034	Amridge U	Montgomery	AL	36117-355	Southern	/https://ww	https://ww	3	0	1	1	3	4	2	1	5	12	NA	32.36261	-86.174	20	
5	100706	105500	1055	University	Huntsville	AL	35899	Southern	/www.uah	finalid.uah	3	0	1	1	3	4	1	1	5	12	NA	34.72456	-86.6404	15	
6	100724	100500	1005	Alabama	Montgomery	AL	36104-027	Southern	/www.alas	www.alas	3	0	1	1	3	4	1	1	5	12	NA	32.36432	-86.2957	17	
7	100751	105100	1051	The Univer	Tuscaloos	AL	35487-010	Southern	/www.ua.e	financiala	3	0	1	1	3	4	1	1	5	13	NA	33.21188	-87.546	15	
8	100760	100700	1007	Central Al	Alexander	AL	35010	Southern	/www.cacc	https://ww	2	0	1	1	1	2	1	1	5	32	NA	32.92478	-85.9453	5	
9	100812	100800	1008	Athens St	Athens	AL	35611	Southern	/www.athe	www.athe	3	0	1	1	3	4	1	1	5	31	NA	34.80679	-86.9647	22	
10	100830	831000	8310	Auburn Un	Montgomery	AL	36117-359	Southern	/www.aum	www.aum	3	0	1	1	3	4	1	1	5	12	NA	32.36736	-86.1775	18	
11	100858	100900	1009	Auburn Un	Auburn	AL	36849	Southern	/www.aub	https://ww	3	0	1	1	3	4	1	1	5	13	NA	32.59938	-85.4883	15	
12	100937	101200	1012	Birmingham	Birmingham	AL	35254	Southern	/www.bsc	www.bsc	3	0	1	1	3	3	2	1	5	12	NA	33.51377	-86.8506	21	
13	101028	1218200	12182	Chattahoo	Chattahoochee	AL	36869	Southern	/https://ww	https://ext	2	0	1	1	1	2	1	1	5	41	NA	32.42391	-85.0315	4	
14	101116	1303906	13039	South Uni	Montgomery	AL	36116	Southern	/www.sout	www.sout	3	0	0	9	2	4	3	1	5	12	NA	32.34268	-86.2165	22	
15	101143	101500	1015	Enterprise	Enterprise	AL	36330-130	Southern	/www.escc	https://ww	2	0	1	1	2	2	1	1	5	32	NA	31.2975	-85.837	2	
16	101161	106000	1060	Coastal Al	Bay Minetti	AL	36507-269	Southern	/www.coas	www.coas	2	0	1	1	2	2	1	1	5	32	NA	30.85134	-87.7782	4	
17	101189	100300	1003	Faulkner U	Montgomery	AL	36109-339	Southern	/www.faulk	www.faulk	3	0	1	1	3	4	2	1	5	12	NA	32.38418	-86.2164	18	
18	101240	101700	1017	Gadsden U	Gadsden	AL	35903	Southern	/www.gads	www.gads	1	0	1	1	1	2	1	1	5	13	NA	33.994	-85.9914	7	
19	101277	4187200	41872	New Begin	Albertville	AL	35951	National	/www.nbcc	nbccosme	1	0	1	1	1	1	3	1	5	32	NA	34.27871	-86.1969	-2	
20	101286	101800	1018	George C	Dothan	AL	36303-923	Southern	/www.wall	www.wall	2	0	1	1	1	2	1	1	5	13	NA	31.31527	-85.4658	7	
21	101295	787100	7871	George C	Hanceville	AL	35077-200	Southern	/https://ww	https://ww	2	0	1	1	1	2	1	1	5	32	NA	34.07244	-86.7819	7	
22	101301	569900	5699	George C	Selma	AL	36702-280	Southern	/https://ww	https://ww	2	0	1	1	2	2	1	1	5	32	NA	32.44592	-87.0133	5	
23	101365	962107	9621	Herzing U	Birmingham	AL	35209	Higher Lee	https://ww	https://ww	2	0	0	10	1	4	2	1	5	21	NA	33.46847	-86.8325	26	
24	101435	101900	1019	Huntingdon	Montgomery	AL	36106-214	Southern	/www.hunt	www.hunt	3	0	1	1	3	3	2	1	5	12	NA	32.35103	-86.2844	22	
25	101453	2199700	21997	Heritage C	Florence	AL	35630-997	Associatio	www.hcu	www.hcu	3	0	1	1	3	4	2	1	5	13	NA	34.85615	-87.6631	24	
26	101462	526000	5260	J. F. Drake	Huntsville	AL	35811	Southern	/https://ww	https://ww	1	0	1	1	1	2	1	1	5	12	NA	34.77186	-86.5729	8	
27	101471	3002500	30025	J F Ingram	Deatsville	AL	36022	Council o	www.istc	istc.edu/N	1	1	1	1	1	1	1	1	5	21	NA	32.58231	-86.3924	-2	
28	101480	102000	1020	Jacksonvi	Jacksonvi	AL	36265	Southern	/www.jsu	www.jsu	3	0	1	1	3	4	1	1	5	23	NA	33.82294	-85.7653	18	
29	101505	102200	1022	Jefferson U	Birmingham	AL	35215-309	Southern	/www.jeffe	www.jeffe	2	0	1	1	1	2	2	1	5	12	NA	33.65254	-86.7072	8	

Most-Recent-Cohorts-Institution

ReadyAccessibility: Unavailable

Figure 6.2 College Scorecard most recent cohorts by institution data.

NAME OF DATA ELEMENT				
Unit ID for institution	programs	cip_4_digit.unit_id	integer	UNITID
6-digit OPE ID for institution	programs	cip_4_digit.ope6_id	string	OPEID6
Institution name	programs	cip_4_digit.school.name	autocomplete	INSTNM
Control of institution	programs	cip_4_digit.school.type	integer	CONTROL
Flag for main campus	programs	cip_4_digit.school.main_campus	integer	MAIN
Classification of Instructional Programs (CIP) code for the field of study	programs	cip_4_digit.code	string	CIPCODE
Text description of the field of study CIP Code	programs	cip_4_digit.title	string	CIPDESC
Level of credential	programs	cip_4_digit.credential.level	integer	CREDLEV
Text description of the level of credential	programs	cip_4_digit.credential.title	string	CREDDESC
Number of awards to all students in year 1 of the pooled debt cohort	programs	cip_4_digit.counts.ipeds_awards1	integer	IPEDSCOUNT1
Number of awards to all students in year 2 of the pooled debt cohort	programs	cip_4_digit.counts.ipeds_awards2	integer	IPEDSCOUNT2
Student recipient count for average/median Parent PLUS loan debt disbursed at all institutions	programs	cip_4_digit.debt.parent_plus.all.all_inst.count	integer	DEBT_ALL_PP_ANY_N
Average Parent PLUS loan debt disbursed at all institutions	programs	cip_4_digit.debt.parent_plus.all.all_inst.average	integer	DEBT_ALL_PP_ANY_MEAN
Median Parent PLUS loan debt disbursed at all institutions	programs	cip_4_digit.debt.parent_plus.all.all_inst.median	integer	DEBT_ALL_PP_ANY_MDN
Student recipient count for average/median Parent PLUS loan debt disbursed to males at all institutions	programs	cip_4_digit.debt.parent_plus.male.all_inst.count	integer	DEBT_MALE_PP_ANY_N
Average Parent PLUS loan debt disbursed to males at all institutions	programs	cip_4_digit.debt.parent_plus.male.all_inst.average	integer	DEBT_MALE_PP_ANY_MEAN
Median Parent PLUS loan debt disbursed to males at all institutions	programs	cip_4_digit.debt.parent_plus.male.all_inst.median	integer	DEBT_MALE_PP_ANY_MDN
Student recipient count for average/median Parent PLUS loan debt disbursed to non-males at all institutions	programs	cip_4_digit.debt.parent_plus.not_male.all_inst.count	integer	DEBT_NOTMALE_PP_ANY_N
Average Parent PLUS loan debt disbursed to non-males at all institutions	programs	cip_4_digit.debt.parent_plus.not_male.all_inst.average	integer	DEBT_NOTMALE_PP_ANY_MEAN
Median Parent PLUS loan debt disbursed to non-males at all institutions	programs	cip_4_digit.debt.parent_plus.not_male.all_inst.median	integer	DEBT_NOTMALE_PP_ANY_MDN
Student recipient count for average/median Parent PLUS loan debt disbursed to Pell recipients at all institutions	programs	cip_4_digit.debt.parent_plus.pell.all_inst.count	integer	DEBT_PELL_PP_ANY_N
Average Parent PLUS loan debt disbursed to Pell recipients at all institutions	programs	cip_4_digit.debt.parent_plus.pell.all_inst.average	integer	DEBT_PELL_PP_ANY_MEAN
Median Parent PLUS loan debt disbursed to Pell recipients at all institutions	programs	cip_4_digit.debt.parent_plus.pell.all_inst.median	integer	DEBT_PELL_PP_ANY_MDN
Student recipient count for average/median Parent PLUS loan debt disbursed to non-Pell-recipients at all institutions	programs	cip_4_digit.debt.parent_plus.not_pell.all_inst.count	integer	DEBT_NOPELL_PP_ANY_N
Average Parent PLUS loan debt disbursed to non-Pell-recipients at all institutions	programs	cip_4_digit.debt.parent_plus.not_pell.all_inst.average	integer	DEBT_NOPELL_PP_ANY_MEAN

Figure 6.3 College Scorecard data dictionary.

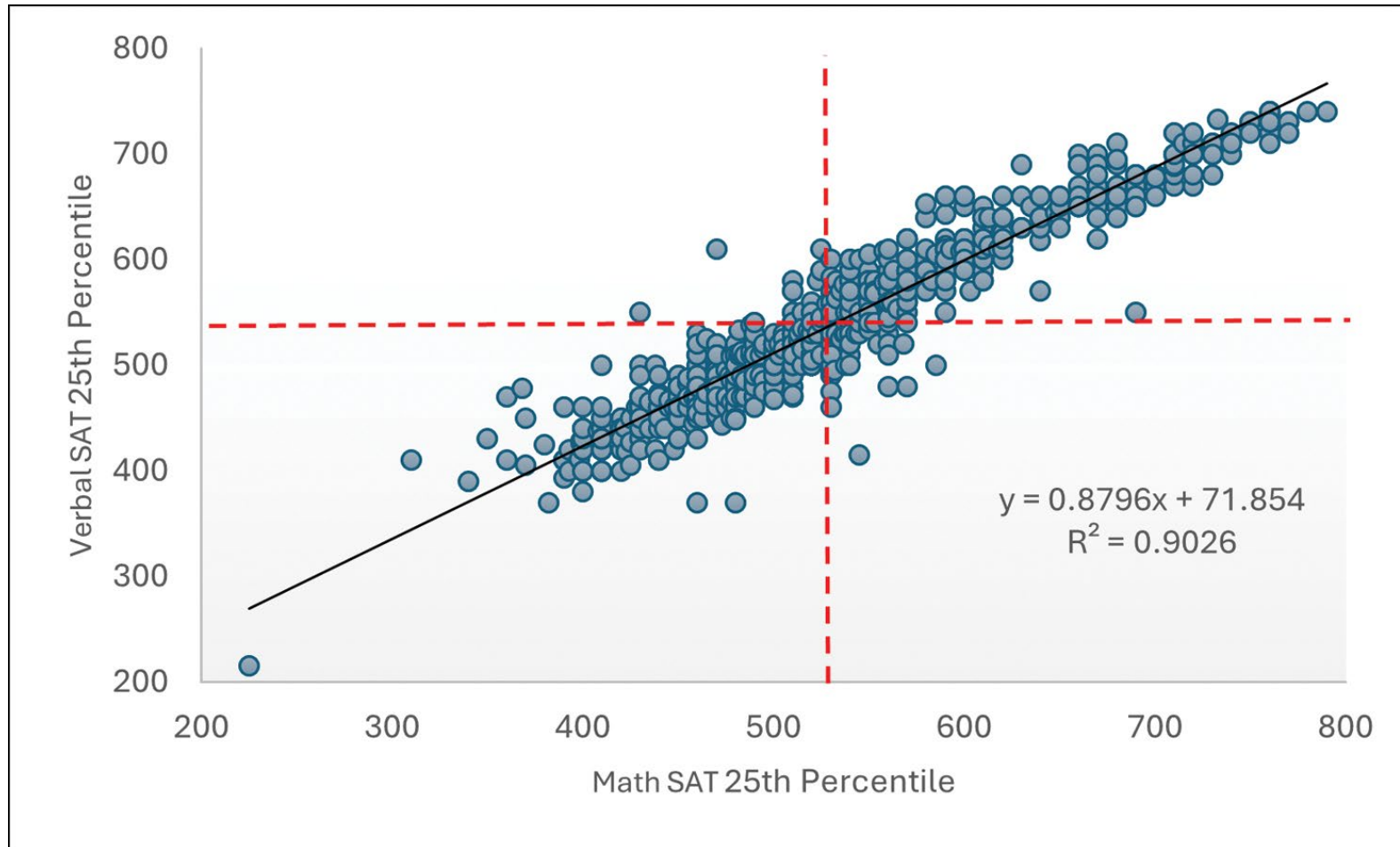


Figure 6.4 A scatterplot showing verbal versus math SAT scores.



Figure 6.5 A scatterplot showing total cost (sticker price) versus admission rate with Maine schools highlighted in a different color.

	A	B	C	D	E
1	INSTNM	SATMT25	SATVR25	ADM_RATE_ALL	COSTT4_A
2	Alabama A & M University	400	430	0.683956439	23167
3	University of Alabama at Birmingham	590	610	0.866794008	26257
4	Amridge University	NA	NA	NA	NA
5	University of Alabama in Huntsville	613	613	0.781042824	25777
6	Alabama State University	399	429	0.965978309	21900
7	The University of Alabama	560	570	0.800599201	31024
8	Central Alabama Community College	NA	NA	NA	14041
9	Athens State University	NA	NA	NA	NA
10	Auburn University at Montgomery	500	510	0.922270742	19771
11	Auburn University	610	610	0.437441184	33650
12	Birmingham-Southern College	560	590	0.571712488	35495
13	Chattahoochee Valley Community College	NA	NA	NA	9320
14	South University-Montgomery	NA	NA	NA	32195
15	Enterprise State Community College	NA	NA	NA	9452
16	Coastal Alabama Community College	NA	NA	NA	14446
17	Faulkner University	480	470	0.824074074	36169

Figure 6.6 New Excel file containing the five variables of interest.

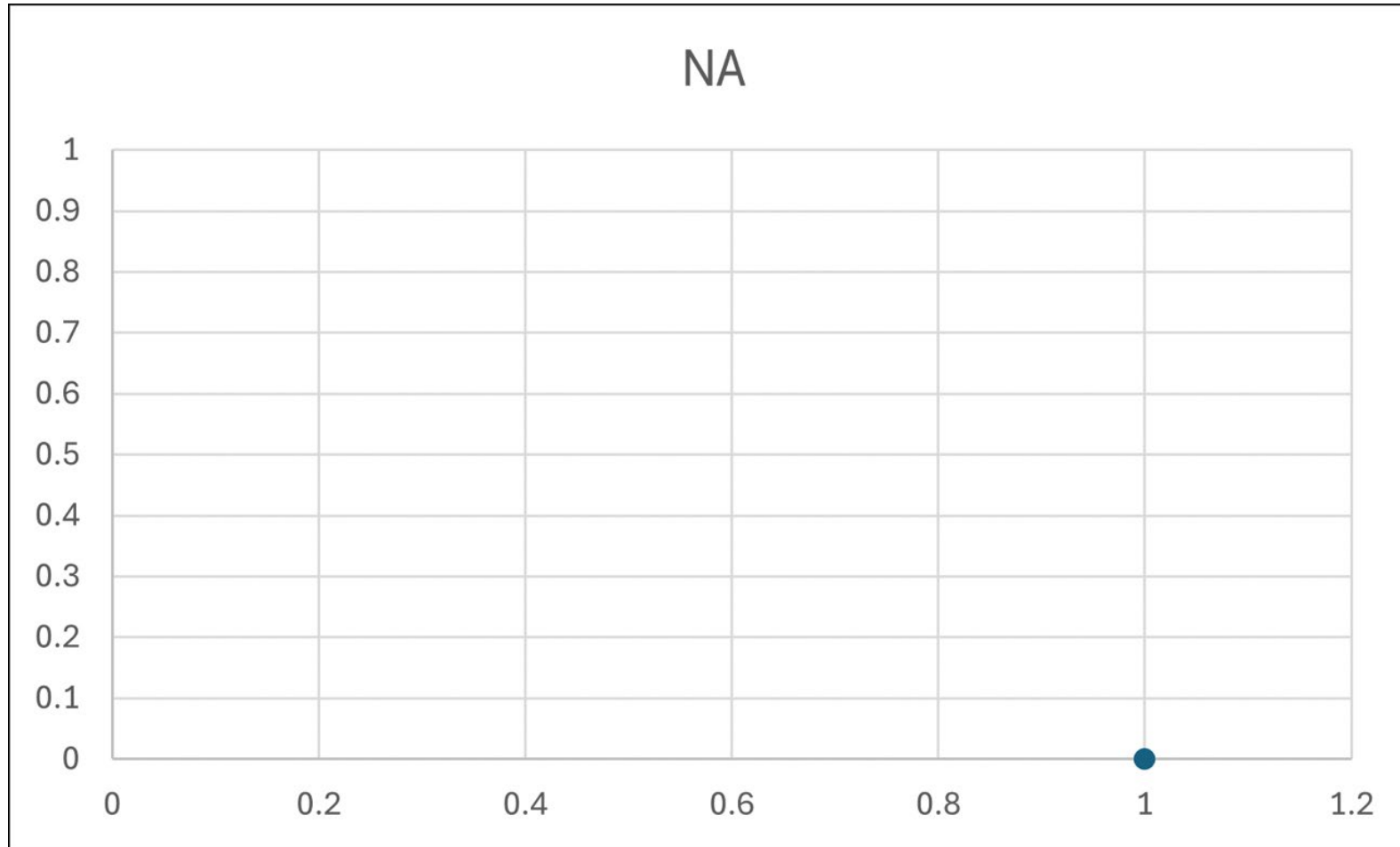


Figure 6.7 Underwhelming scatterplot in Excel due to missing (NA) values.

	A	B	C	D	E
1	INSTNM	SATMT25	SATVR25	ADM_RATE_ALL	COSTT4_A
2	Alabama A & M University	400	430	0.683956439	23167
3	University of Alabama at Birmingham	590	610	0.866794008	26257
5	University of Alabama in Huntsville	613	613	0.781042824	25777
6	Alabama State University	399	429	0.965978309	21900
7	The University of Alabama	560	570	0.800599201	31024
10	Auburn University at Montgomery	500	510	0.922270742	19771
11	Auburn University	610	610	0.437441184	33650
12	Birmingham-Southern College	560	590	0.571712488	35495
17	Faulkner University	480	470	0.824074074	36169
24	Huntingdon College	500	500	0.659541391	39717
28	Jacksonville State University	460	480	0.784645982	22153
32	University of West Alabama	450	430	0.688597892	22456
44	Samford University	530	550	0.832603182	52514
48	University of South Alabama	493	510	0.653625498	22667
49	Spring Hill College	560	550	0.726341463	35488
53	Troy University	460	490	0.948400853	22005
54	Tuskegee University	475	495	0.296852944	41924
74	University of Arizona	570	570	0.86750700	27680

Figure 6.8 Making your data a table in Excel and filtering out the NA values.

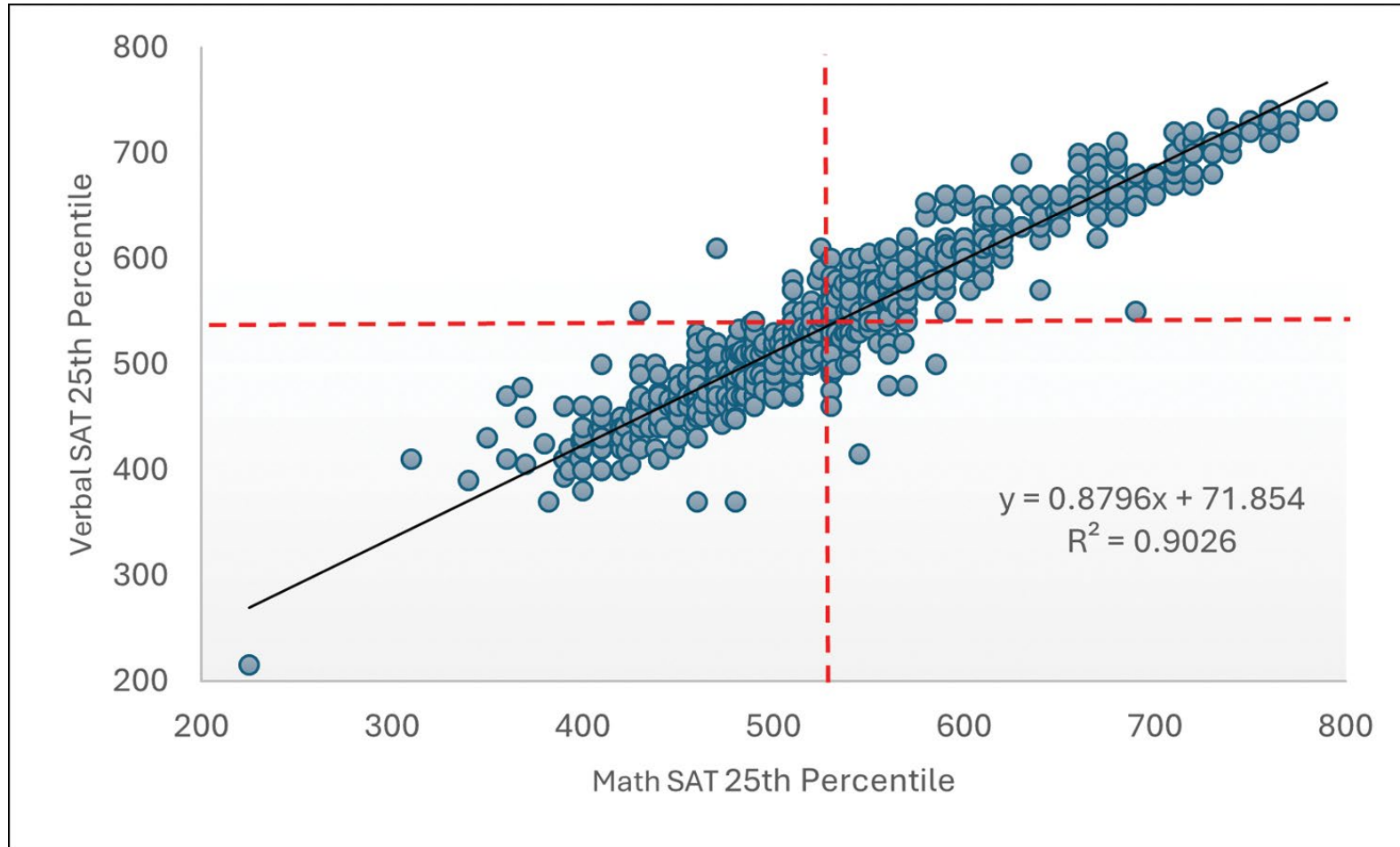


Figure 6.9 Scatterplot of verbal versus math SAT 25th percentiles by institution.

	A	B	C
1	INSTNM ▼	SATMT25 ▼	SATVR25 ▼
2	Rust College	225	215
3	University of Providence	310	410
4	Villa Maria College	340	390
5	Mercy College of Ohio	350	430
6	Bloomfield College	360	410
7	The University of Texas Rio Grande Valle	360	470
8	Escuela de Artes Plasticas y Diseno de P	368	478
9	Fairmont State University	370	450
10	University of North Texas at Dallas	370	405
11	University of Arkansas at Pine Bluff	380	425

Figure 6.10 Sorting the math scores in ascending order to see the minimum values.

	SATMT25	SATVR25
MIN	225	215
25th	480	480
MEDIAN	520	523
Mean	529.5	537.6
75th	570	580
MAX	790	740

Figure 6.11 Five-number summary of verbal versus math SAT 25th percentiles by institution.

	A	B	C
1	INSTNM ▼	ADM_RATE_ALL ▼	COSTT4_A ▼
2	Alabama A & M University	68.4%	\$ 23,167
3	University of Alabama at Birmingham	86.7%	\$ 26,257
4	University of Alabama in Huntsville	78.1%	\$ 25,777
5	Alabama State University	96.6%	\$ 21,900
6	The University of Alabama	80.1%	\$ 31,024
7	Auburn University at Montgomery	92.2%	\$ 19,771
8	Auburn University	43.7%	\$ 33,650
9	Birmingham-Southern College	57.2%	\$ 35,495
10	Faulkner University	82.4%	\$ 36,169
11	Huntingdon College	66.0%	\$ 39,717

Figure 6.12 A table showing admission rates and costs.

	A	B	C
1	INSTNM	ADM_RATE_ALL	COSTT4_A
2	United States Air Force Acad	16.2%	NA
3	United States Coast Guard /	23.9%	NA
4	Principia College	47.8%	NA
5	The Southern Baptist Theolo	87.2%	NA
6	United States Naval Acader	10.8%	NA
7	Hillsdale College	20.6%	NA
8	United States Military Acad	12.0%	NA
9	Grove City College	71.0%	NA
10	New Saint Andrews College	83.0%	NA
11	Patrick Henry College	75.8%	NA
12	Jersey College	66.0%	NA
13	Jewish Theological Semina	50.8%	\$ 86,964
14	University of Chicago	5.4%	\$ 84,126
15	Northwestern University	7.2%	\$ 83,803
16	University of Pennsylvania	6.5%	\$ 83,298
17	Columbia University in the C	3.9%	\$ 82,617
18	Brown University	5.1%	\$ 82,570
19	Wesleyan University	14.4%	\$ 82,245
20	Harvey Mudd College	13.4%	\$ 82,236
21	Yale University	4.6%	\$ 82,170
22	Washington University in St	11.8%	\$ 81,620

Figure 6.13 Sorting by total cost to remove NAs and see the maximum values.

	A	B	C	D	E	F
992	University of Chicago	IL	760	740	0.054341453	84126
993	Northwestern University	IL	760	730	0.07208209	83803
994	Harvard University	MA	760	730	0.032407181	81292
995	Dartmouth College	NH	760	740	0.063805759	81501
996	Princeton University	NJ	760	730	0.056997817	78490
997	Columbia University in the City of New York	NY	760	730	0.039488165	82617
998	Cornell University	NY	760	710	0.074687477	80278
999	Duke University	NC	760	730	0.063459693	81138
1000	Brown University	RI	760	730	0.050583427	82570
1001	Vanderbilt University	TN	760	730	0.066692542	79538
1002	Rice University	TX	760	730	0.086823776	71745
1003	Washington University in St Louis	MO	770	730	0.117573562	81620
1004	Carnegie Mellon University	PA	770	720	0.113043986	77465
1005	University of Pennsylvania	PA	770	730	0.065014289	83298
1006	Stanford University	CA	770	730	0.036805137	78898
1007	Franklin W Olin College of Engineering	MA	770	720	0.191415313	79024
1008	Johns Hopkins University	MD	780	740	0.072534616	78415
1009	Massachusetts Institute of Technology	MA	790	740	0.039594871	77020
1010	Amridge University	AL	NA	NA	NA	NA
1011	Central Alabama Community College	AL	NA	NA	NA	14041
1012	Athens State University	AL	NA	NA	NA	NA
1013	Chattahoochee Valley Community College	AL	NA	NA	NA	9320
1014	South University-Montgomery	AL	NA	NA	NA	32195

Figure 6.14 Sorting by SATMT25 to remove NAs and see the maximum values after adding the state abbreviation column.

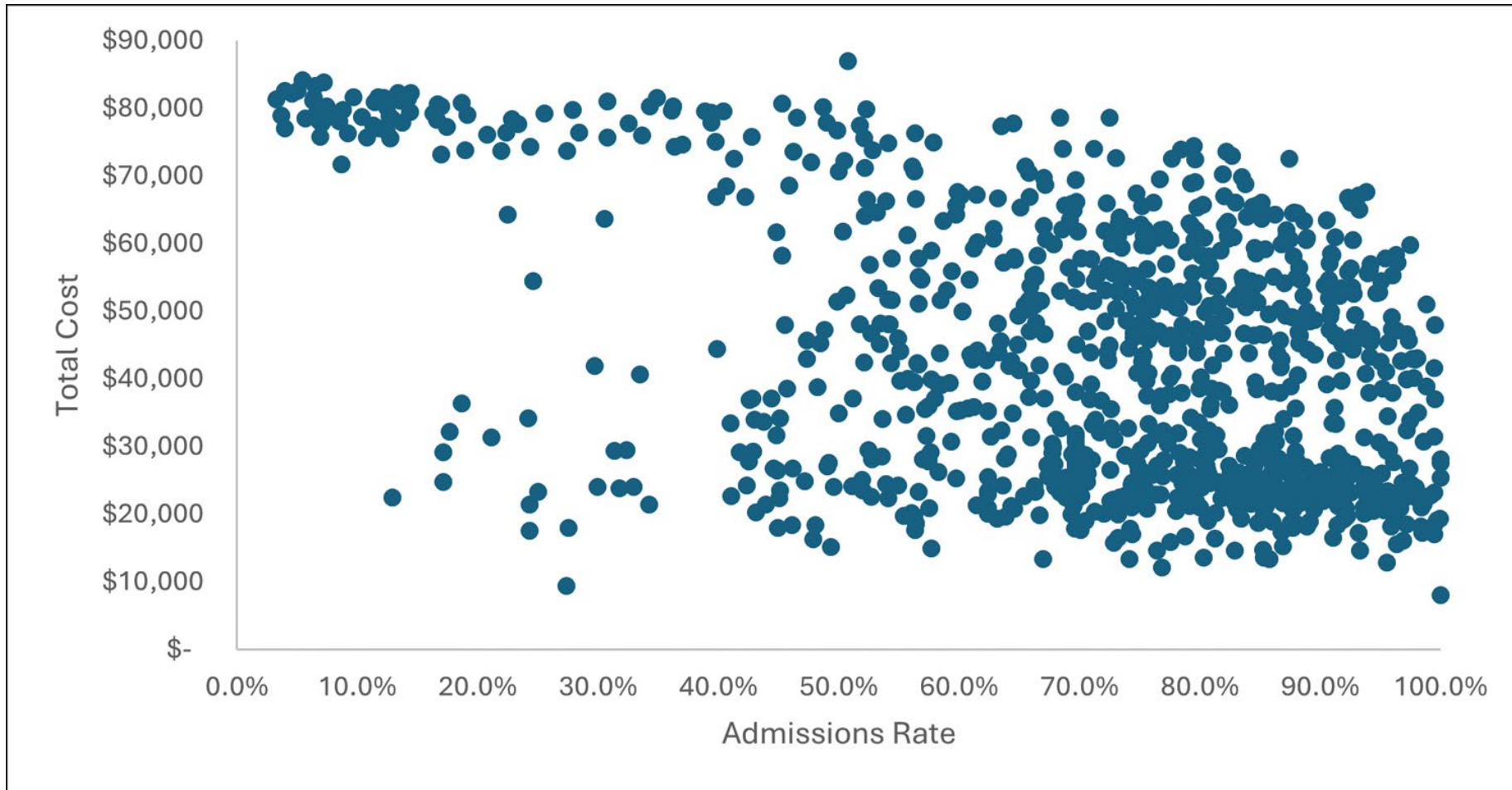


Figure 6.15 Scatterplot of total cost versus admissions rate.

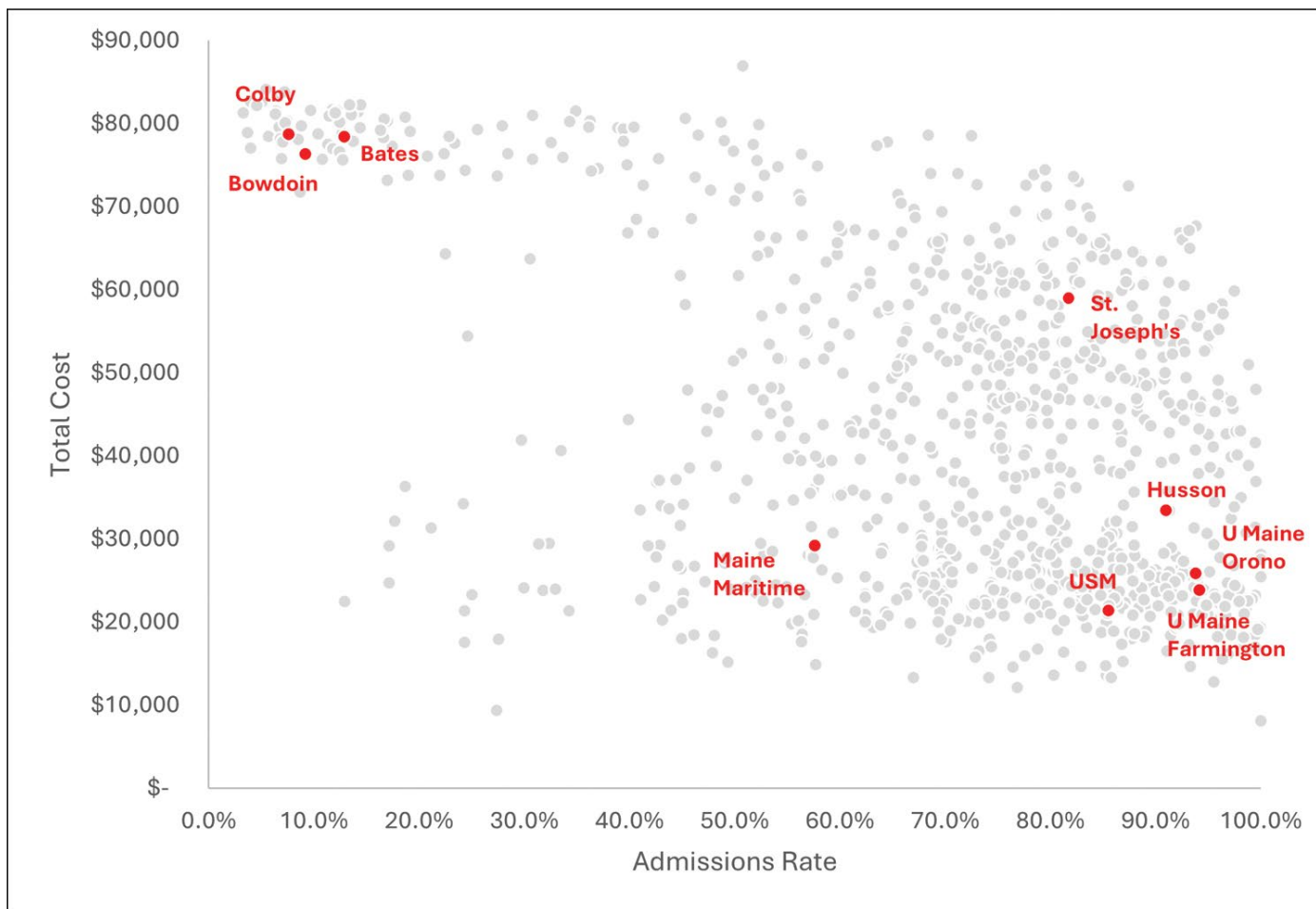


Figure 6.16 Finished scatterplot with formatting choices and Maine schools highlighted.

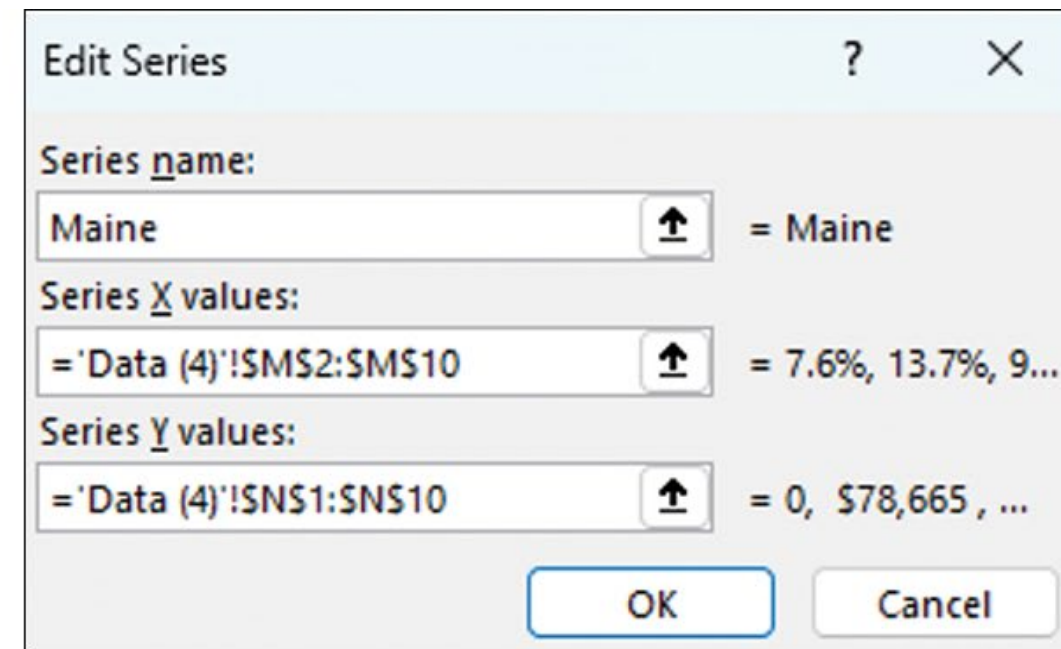
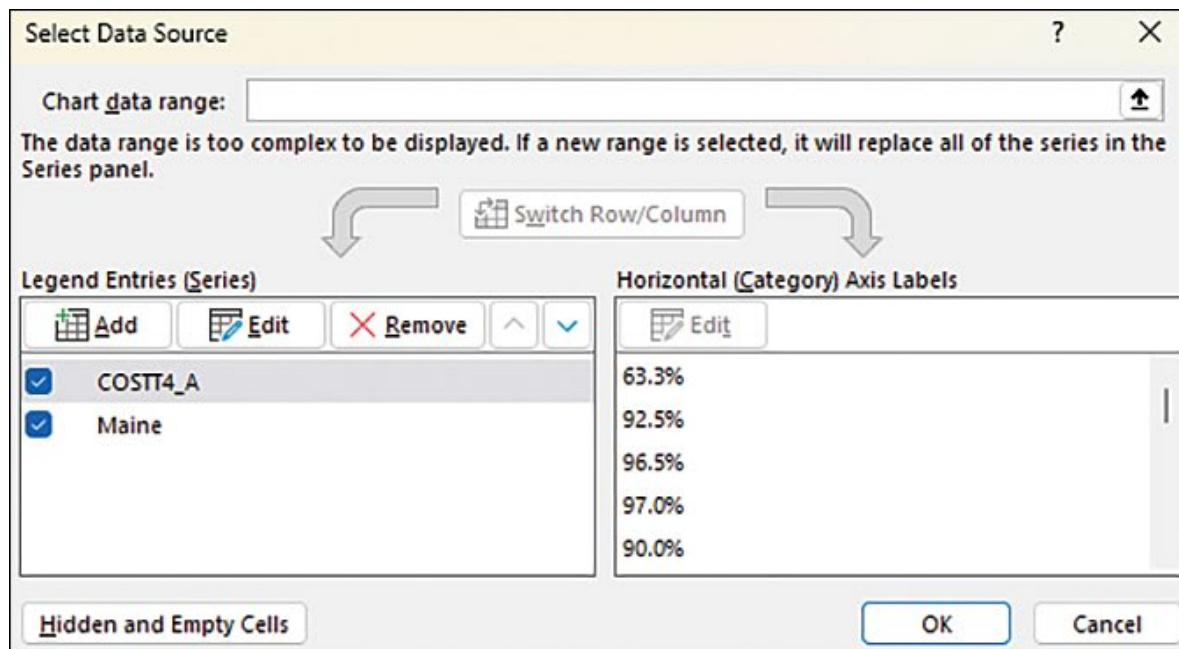


Figure 6.17 The Select Data menu option to add a new data series for just the Maine schools.

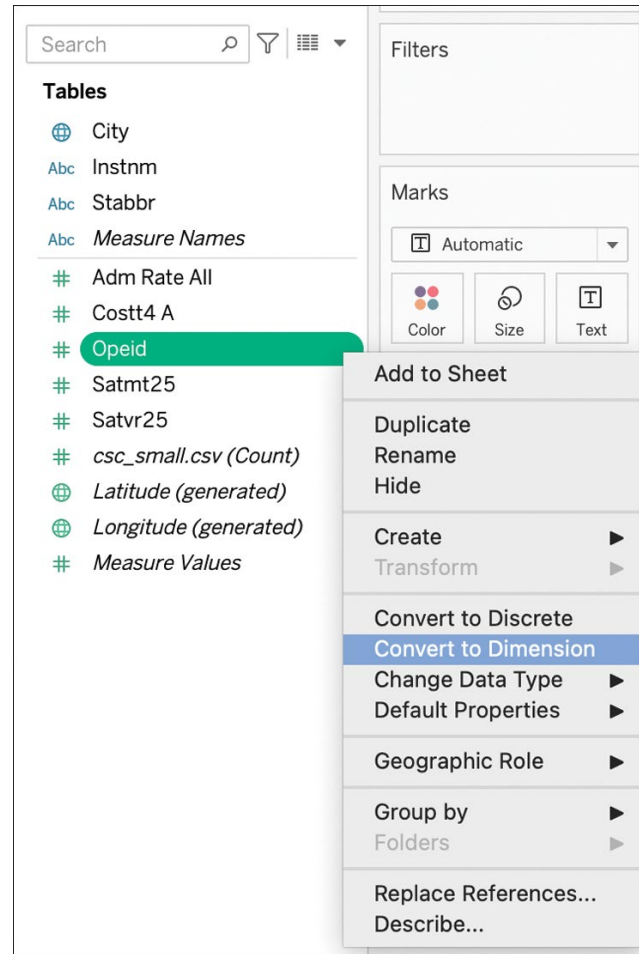


Figure 6.18 Converting a table from a measure to a dimension.

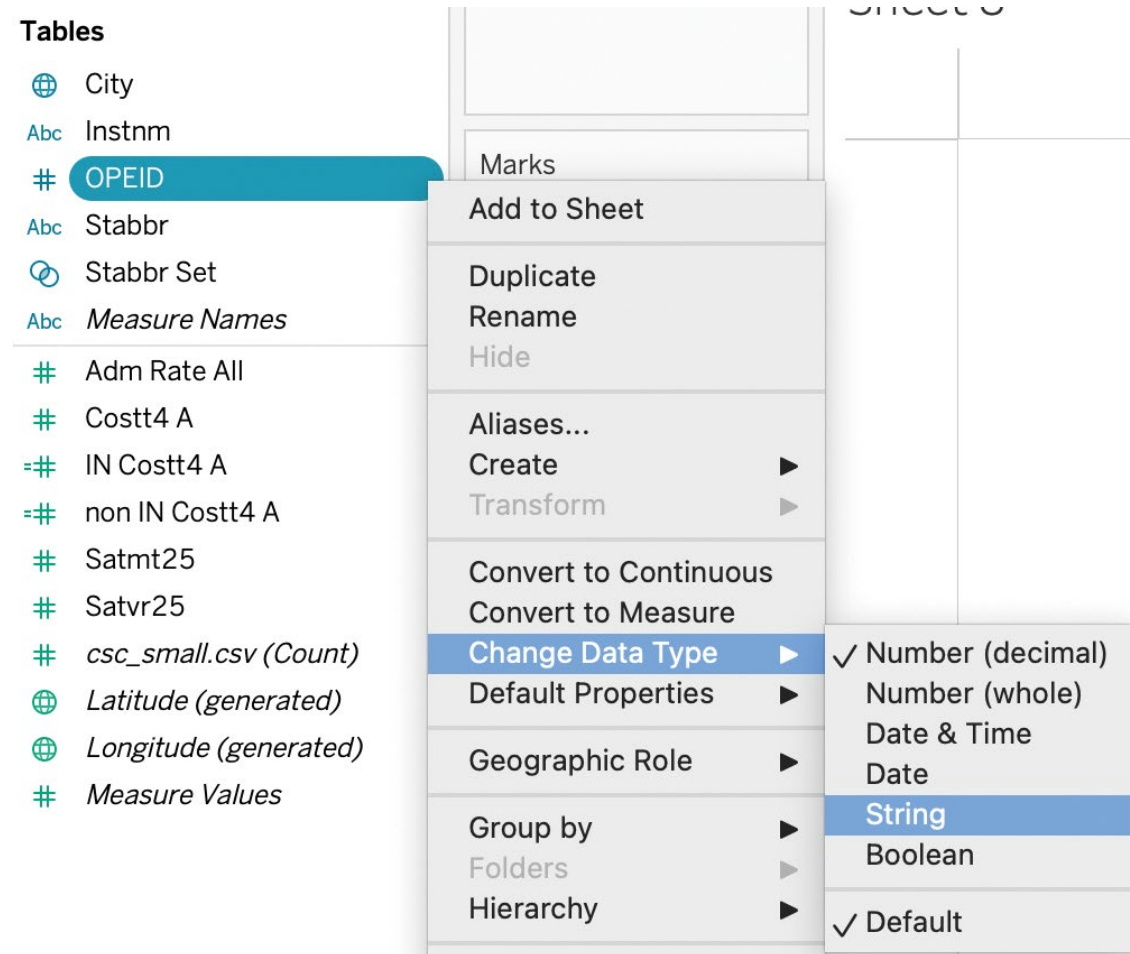


Figure 6.19 Converting from a numeric data type to a string data type.

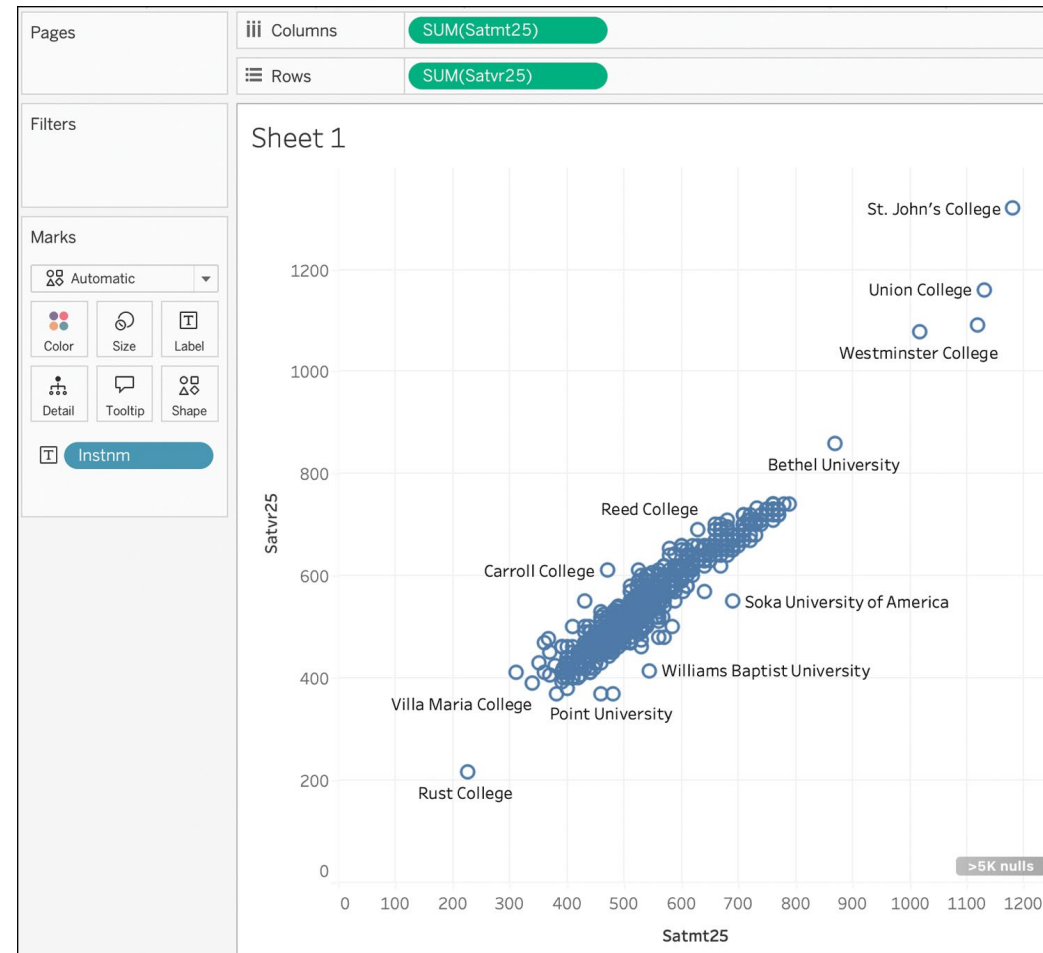


Figure 6.20 Incorrect SAT scores scatterplot.



Figure 6.21 Scatterplot comparing math and verbal SAT scores.

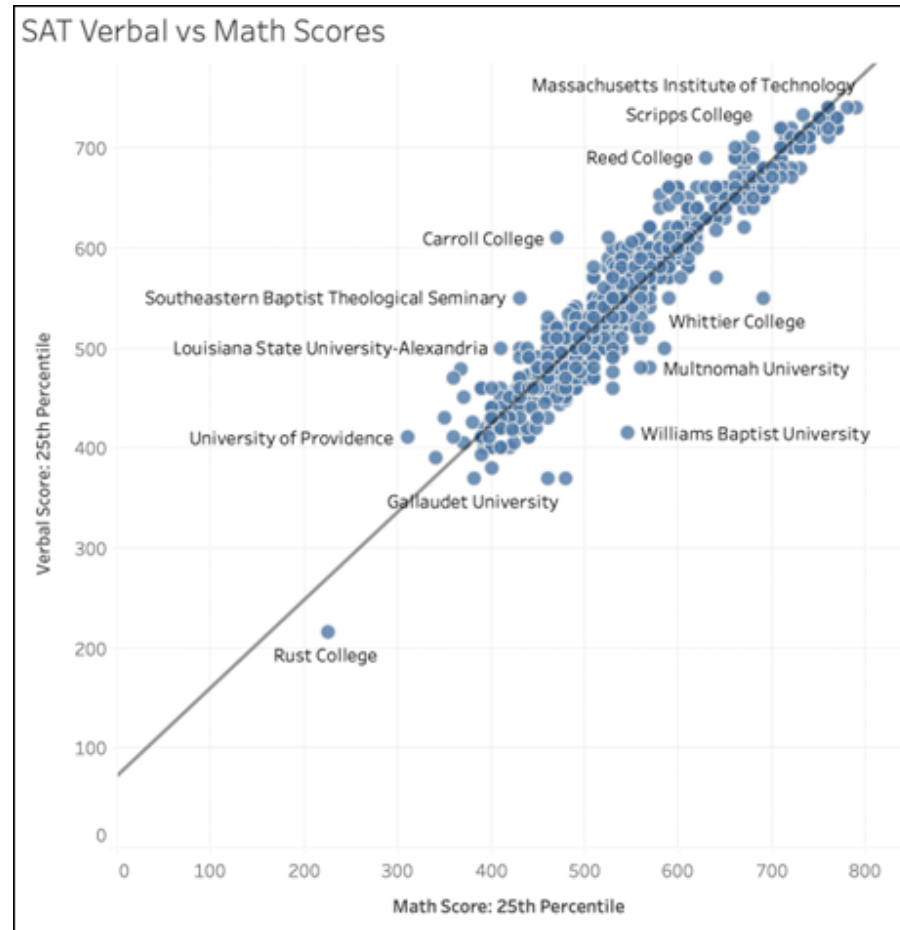


Figure 6.22 SAT scores scatterplot with trendline.



Figure 6.23 Filtering the scatterplot by institution name for “University.”

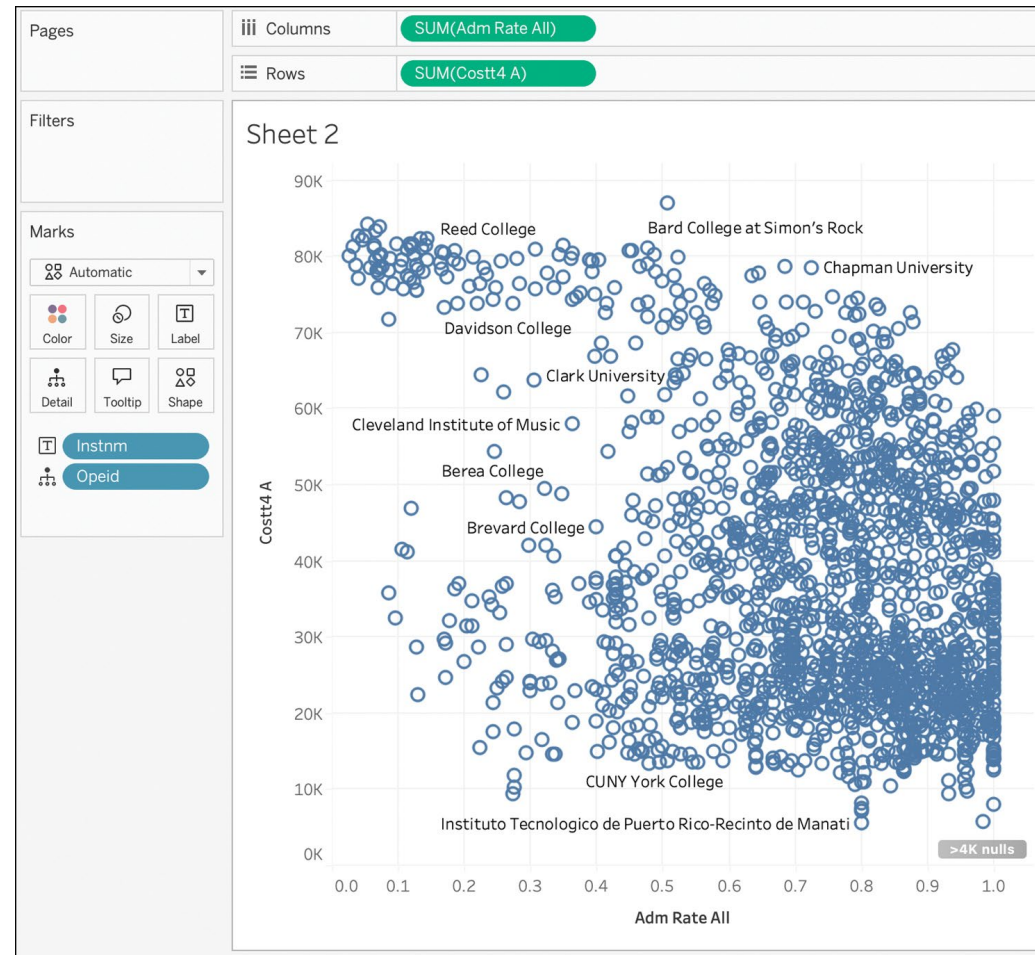


Figure 6.24 A basic graphic comparing sticker price to admission rate.

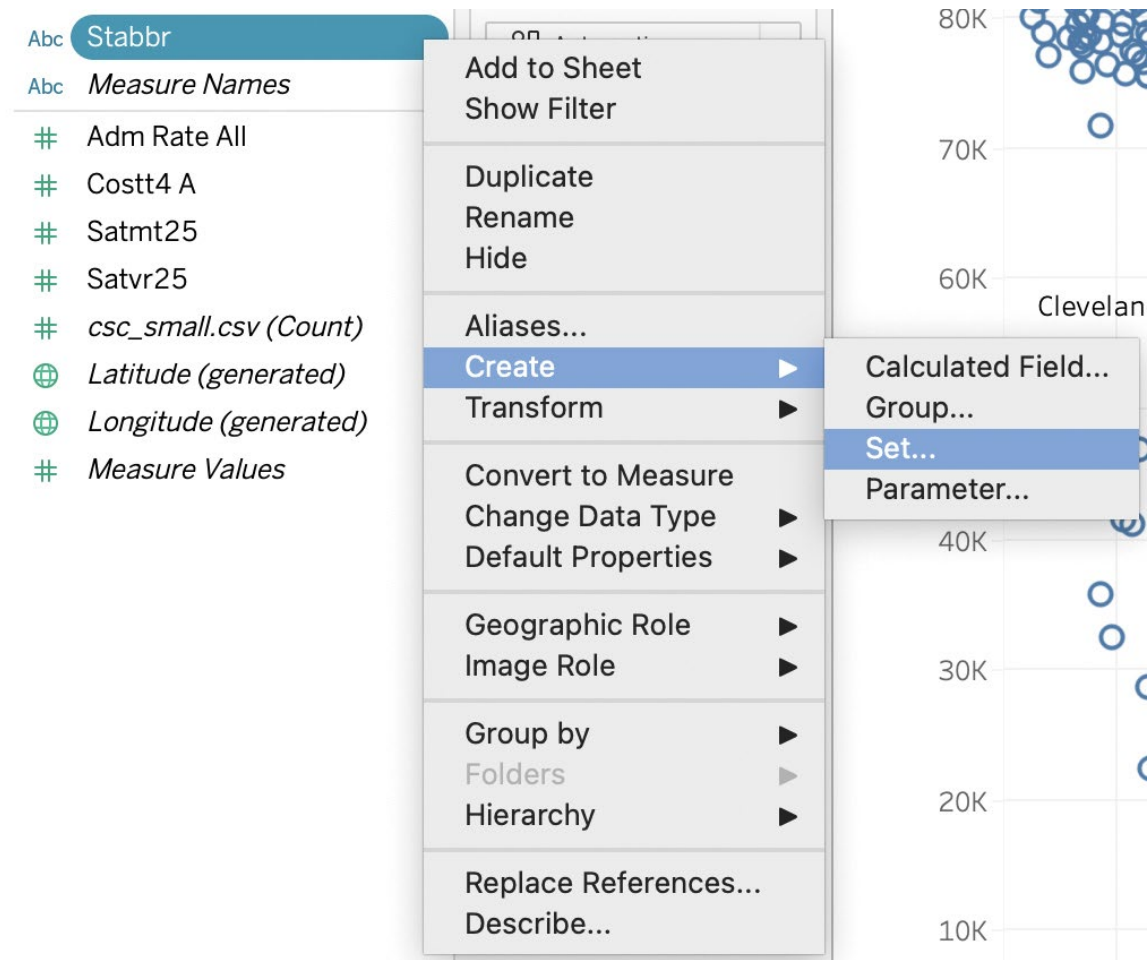


Figure 6.24 A basic graphic comparing sticker price to admission rate.

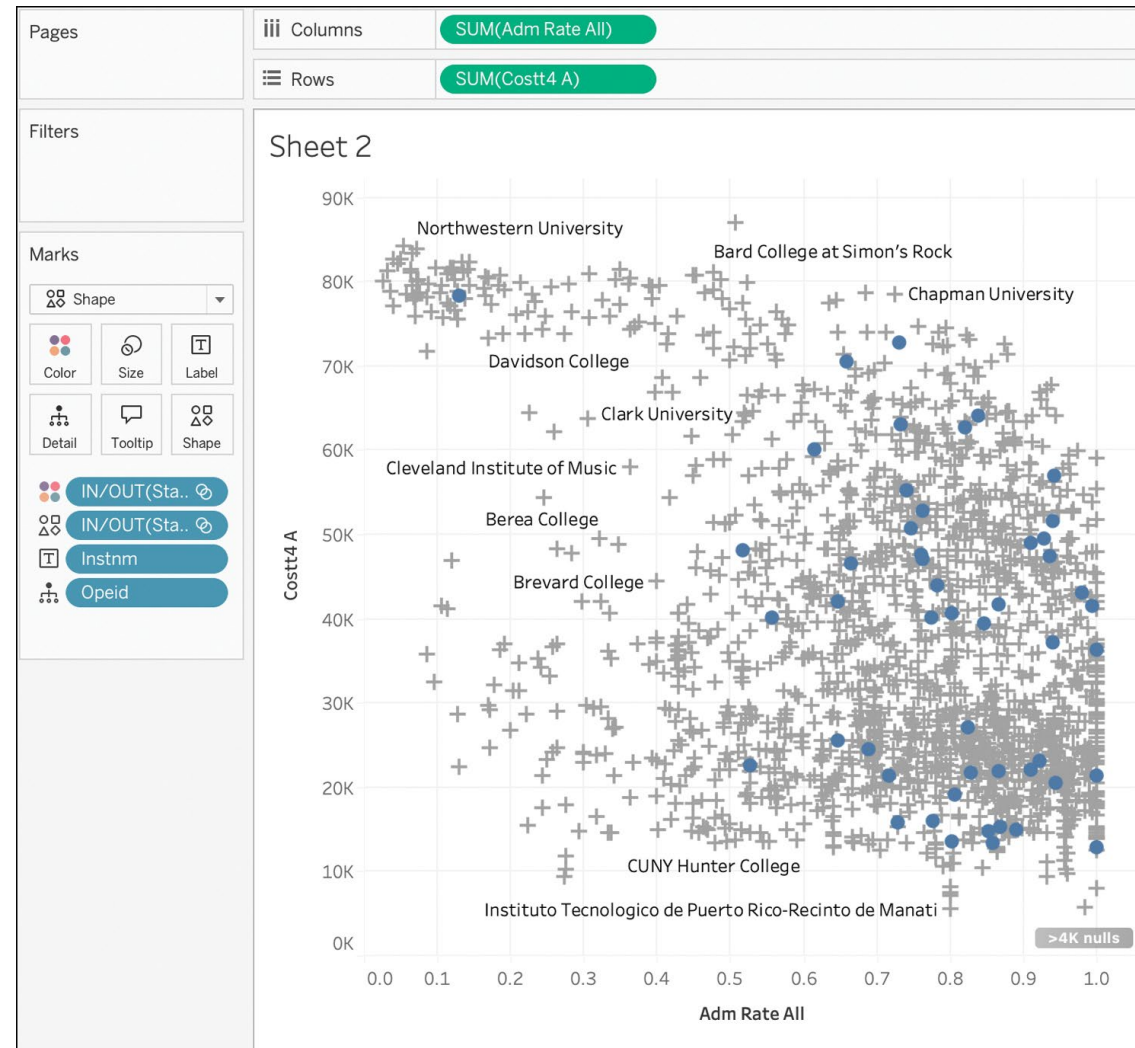


Figure 6.26 A quick graphic highlighting Indiana schools.

non IN Costt4 A

×

IF NOT [Stabbr Set] THEN [Costt4 A] ELSE NULL END

The calculation is valid.

ApplyOK

IN Costt4 A

×

IF [Stabbr Set] THEN [Costt4 A] ELSE NULL END

The calculation is valid.

1 Dependency ▼ApplyOK

Figure 6.27 Formulas used to independently control formatting that distinguishes Indiana schools from non-Indiana schools.

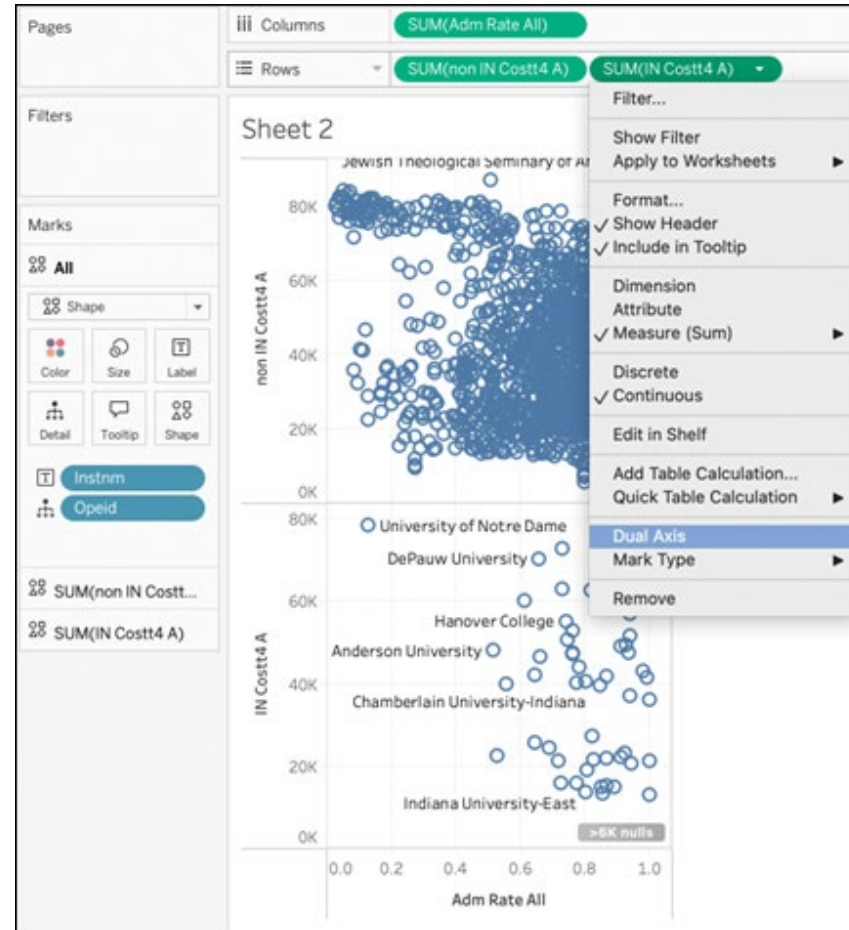


Figure 6.28 Creating a two-layer plot using the dual-axis option.

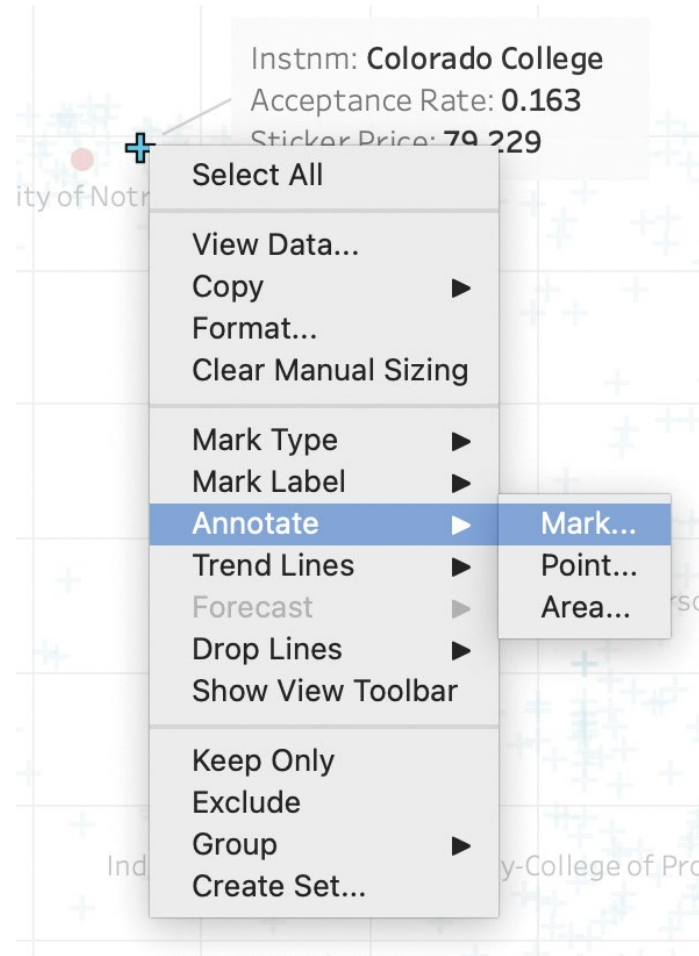


Figure 6.29 Annotating a point.

Sticker Price vs Acceptance Rate

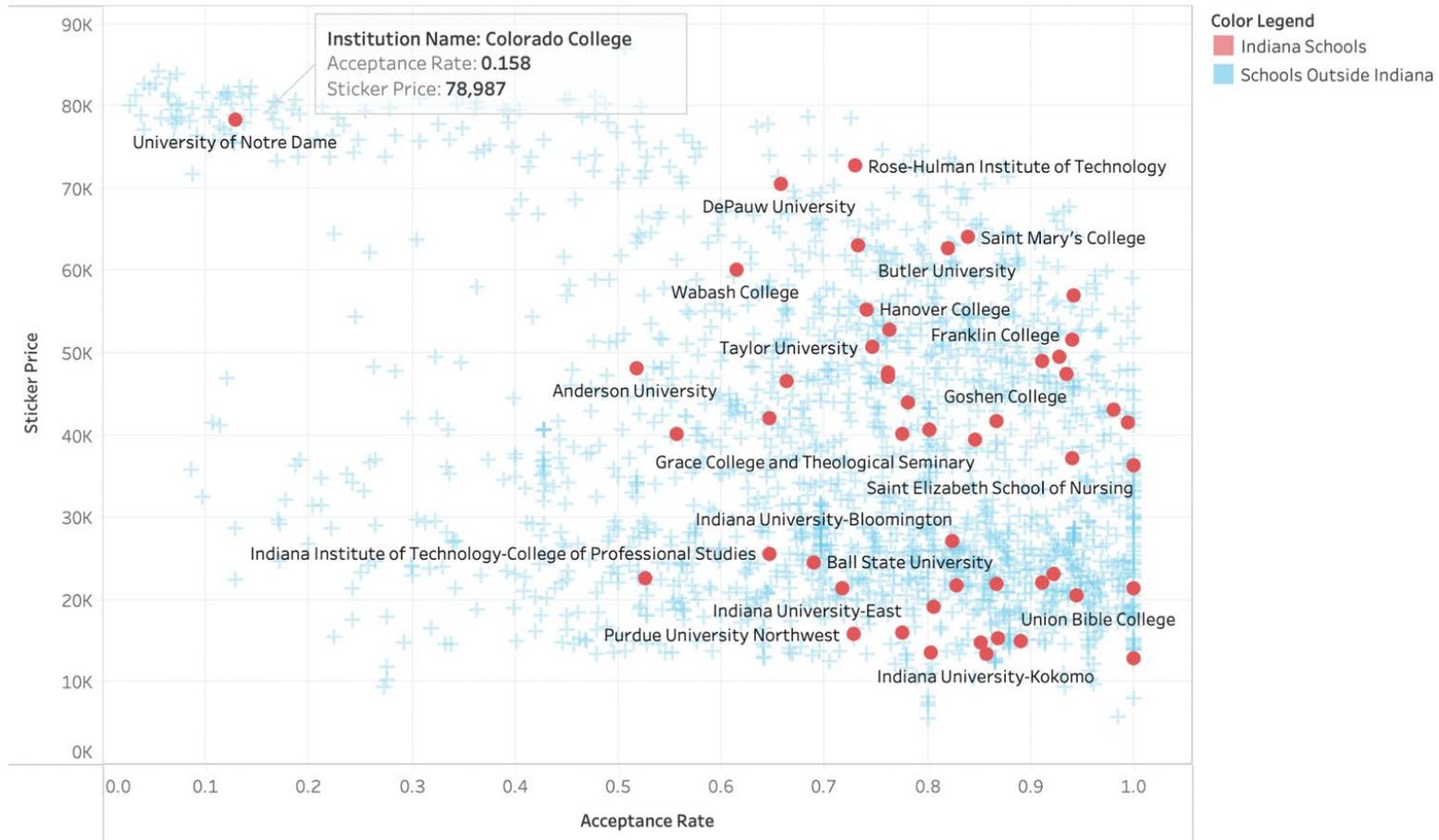


Figure 6.30 Sticker price versus acceptance rate scatterplot with Colorado College highlighted.

Sticker Price vs Acceptance Rate

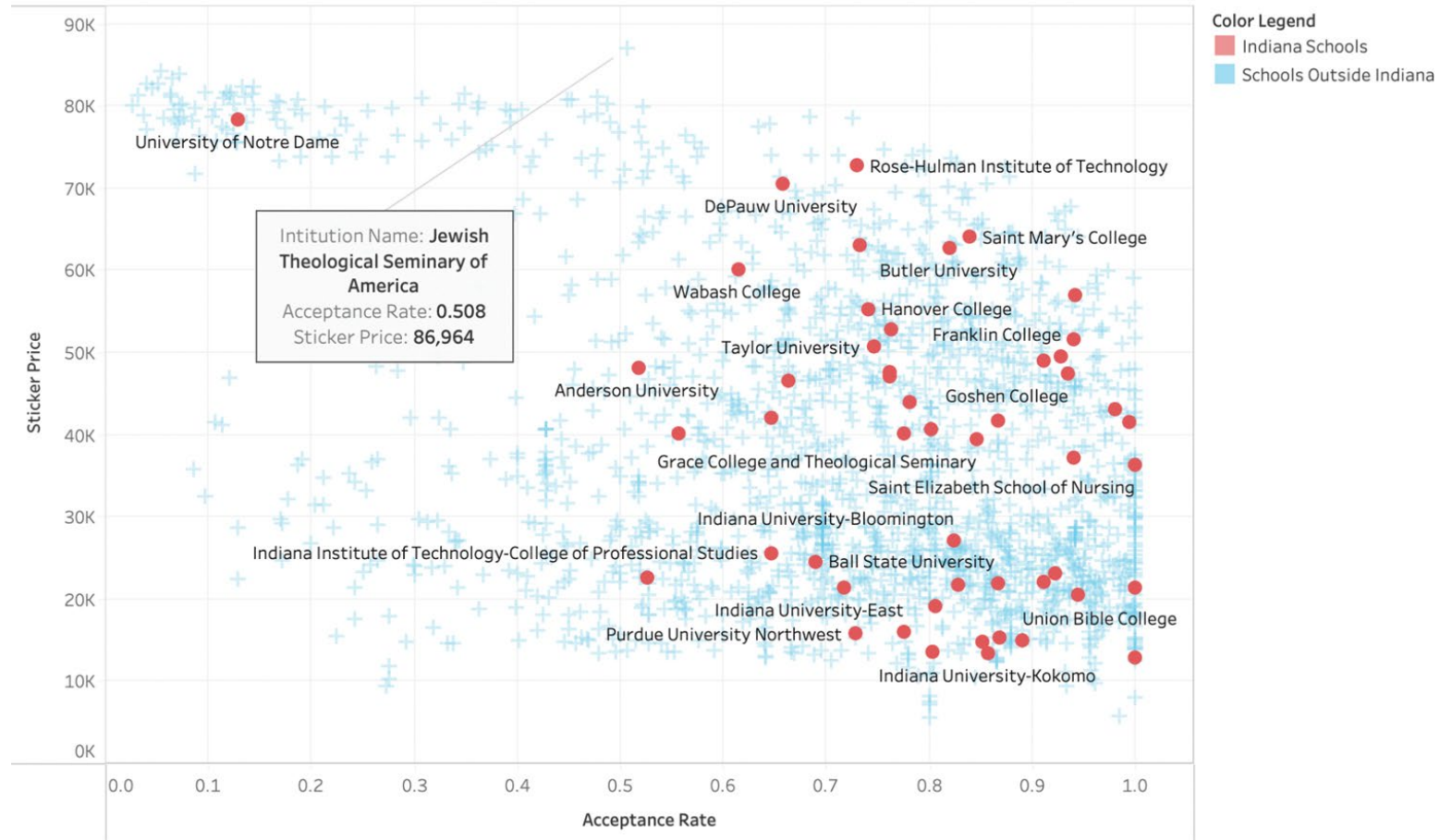


Figure 6.31 Scatterplot highlighting the most expensive (by sticker price) school in the dataset.

Filter [Costt4 A]

Range of values
Range of values

At least

At most

Special

5,500 86,964

5500 86964

Show: Only Relevant Values ☐ Include Null Values

Reset Apply Cancel OK

Figure 6.32 Filtering out schools without a cost value.

Instnm	OPFID	
Aaniih Nakoda College		
Abilene Christian Universi		
Abilene Christian Universi		
Abraham Baldwin Agricul.		
Academia Medical Institute	4303000	26,460
Academy College	2050300	34,100
Academy of Art University	753100	44,484
Academy of Interactive	4248400	35,103
Entertainment	4223600	34,602
Acupuncture and Massag..	3414500	37,935
Adams State University	134500	22,772
Adelphi University	266600	55,566

Data source order
Alphabetic

Field ▶

Nested ▶

MAX(Costt4 A)

Figure 6.33 Sorting by sticker price.

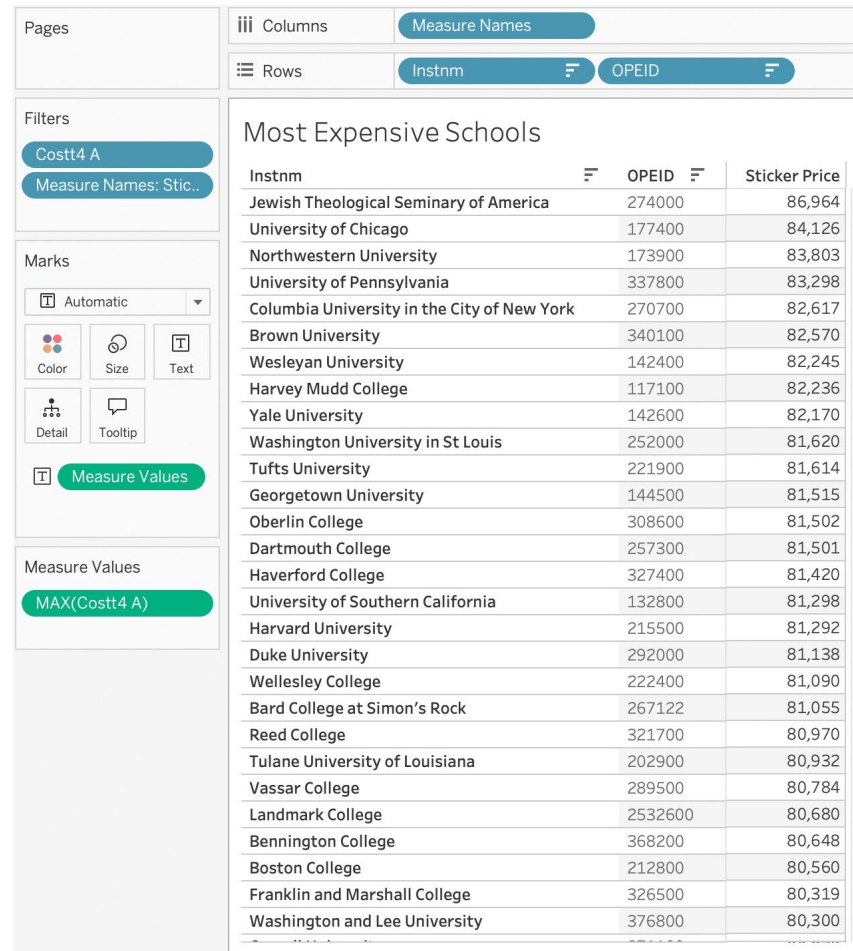


Figure 6.34 Table sorted by sticker price with a column label added.

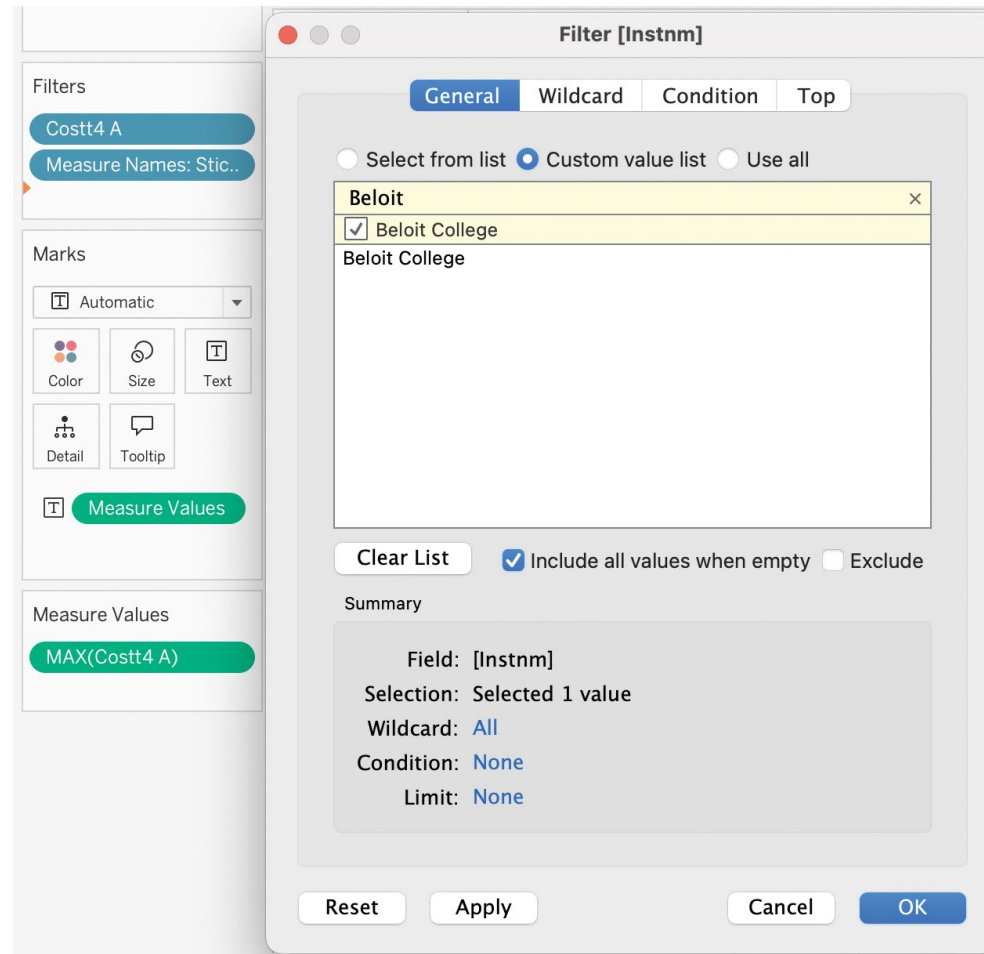


Figure 6.35 Using the search bar to find the sticker price for an individual school.

Sticker Price		
Instnm	OPEID	Sticker Price
Beloit College	383500	66,936

Figure 6.36 Table showing the sticker price for an individual school.

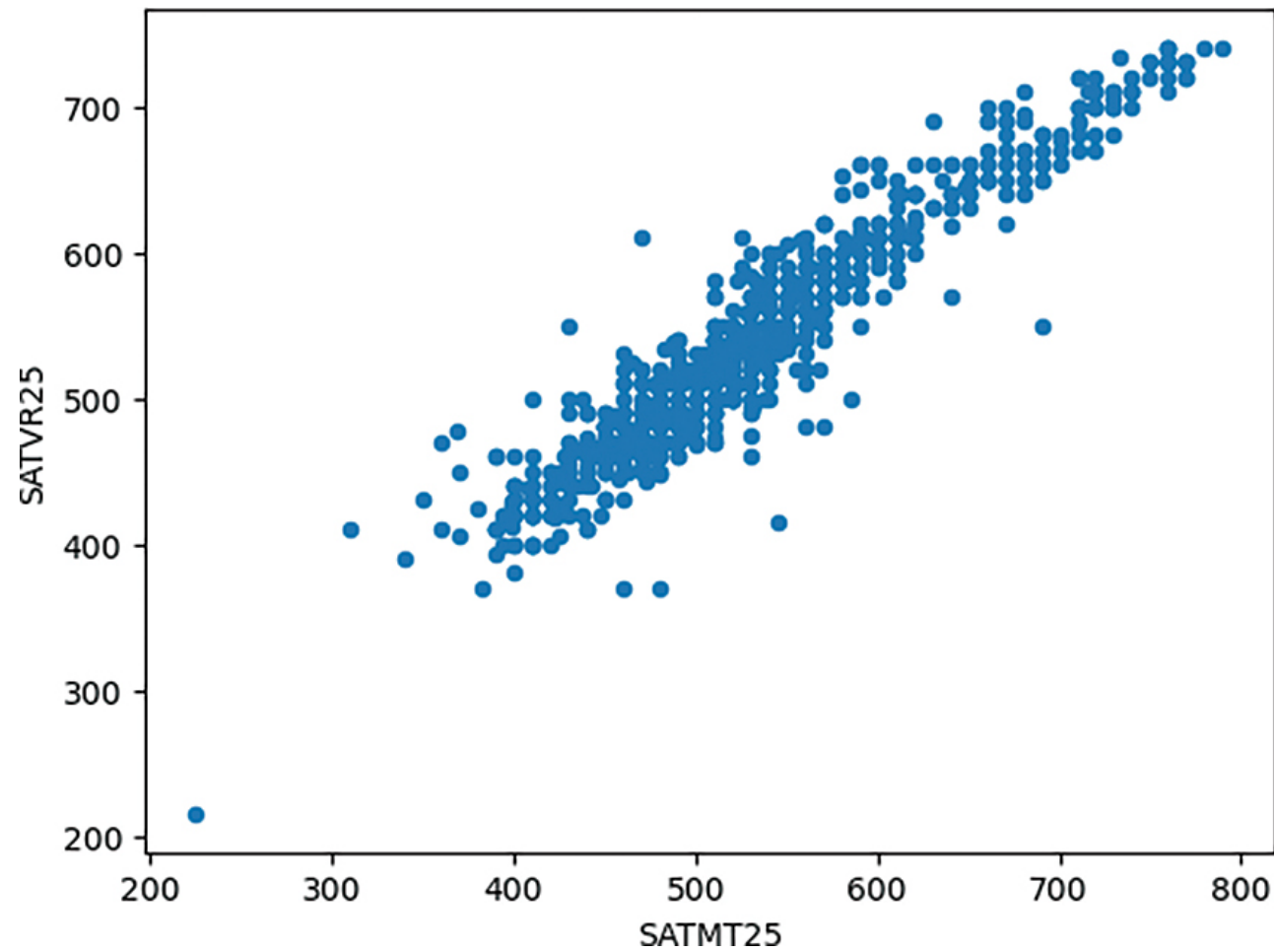


Figure 6.37 Quick Pandas plot.



Figure 6.38 Plotly scatterplot.

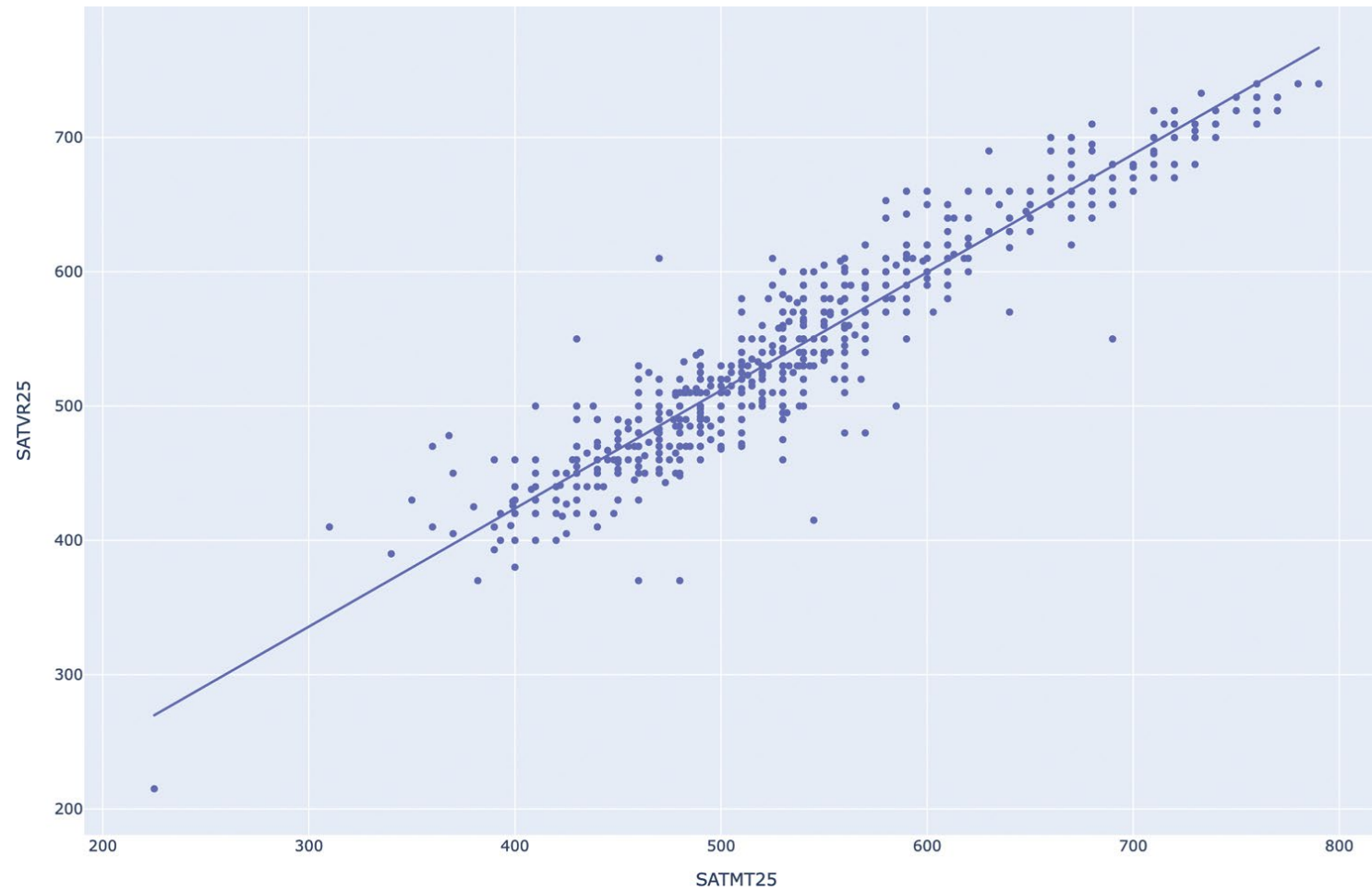


Figure 6.39 Scatterplot with trendline.

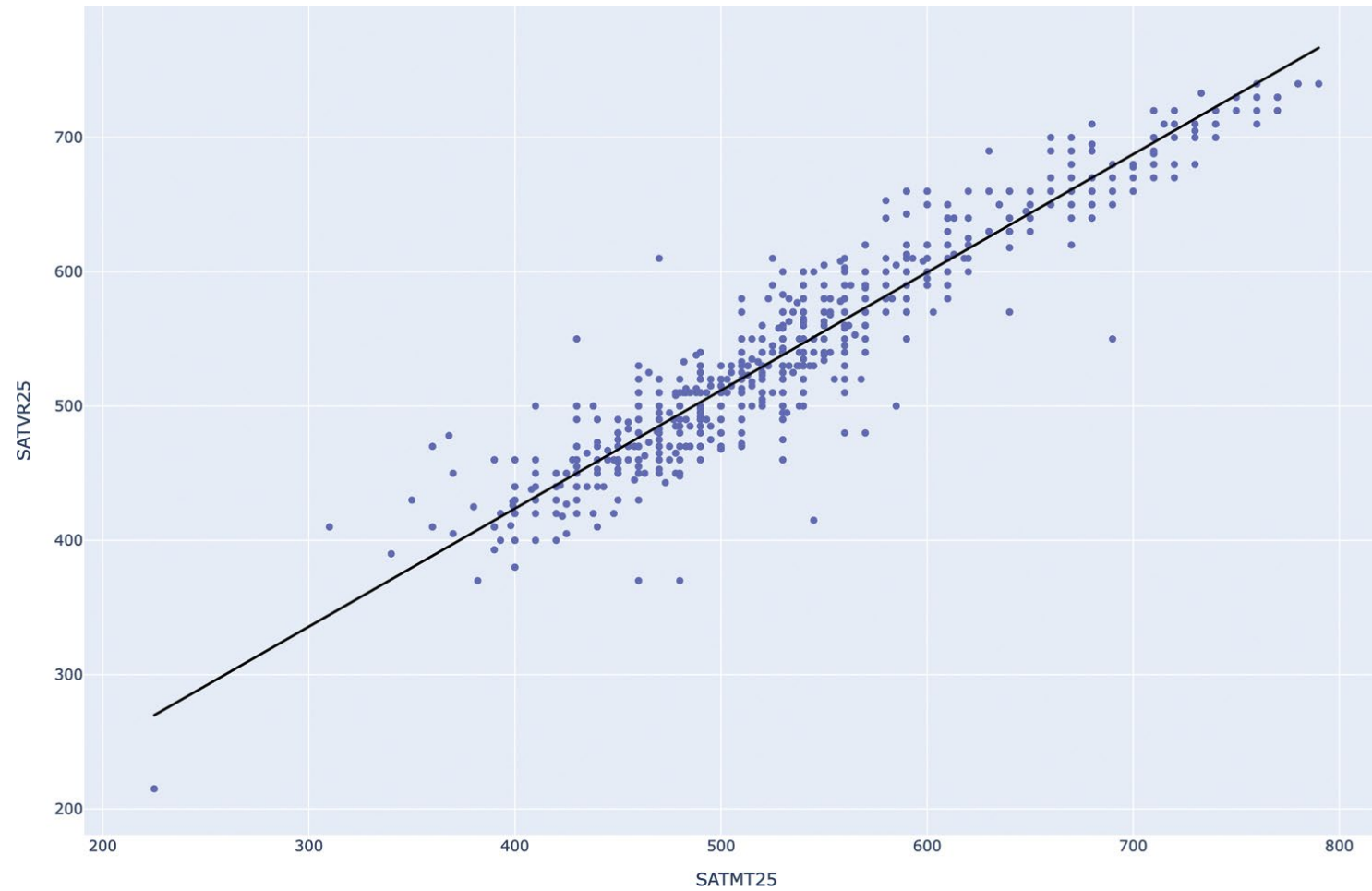


Figure 6.40 Black trendline.

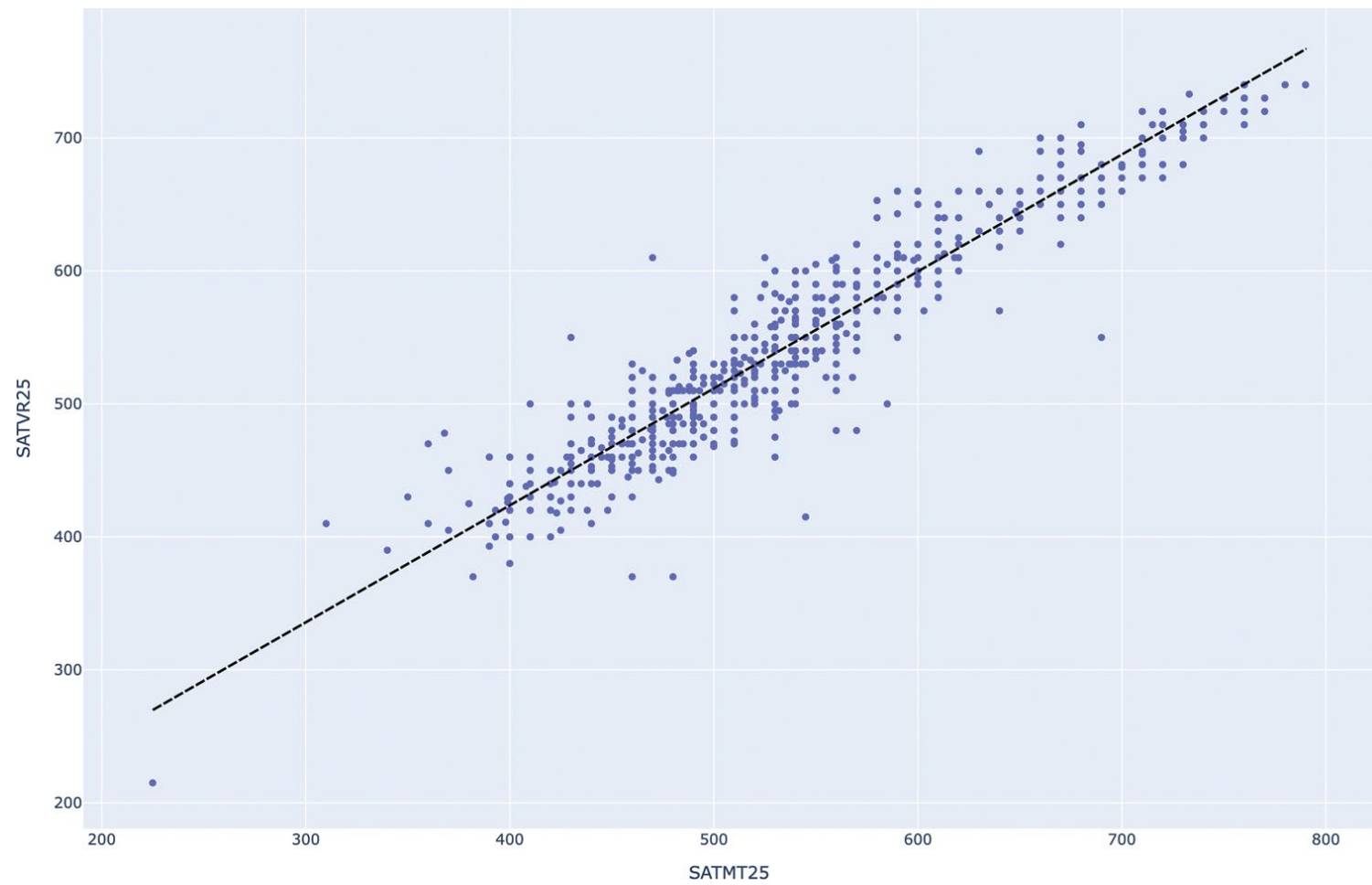


Figure 6.41 Dashed trendline.

Dep. Variable:	y	R-squared:	0.903
Model:	OLS	Adj. R-squared:	0.902
Method:	Least Squares	F-statistic:	9322.
Date:	Mon, 07 Jul 2025	Prob (F-statistic):	0.00
Time:	13:45:25	Log-Likelihood:	-4636.6
No. Observations:	1008	AIC:	9277.
Df Residuals:	1006	BIC:	9287.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	71.8545	4.883	14.714	0.000	62.272	81.437
x1	0.8796	0.009	96.550	0.000	0.862	0.898

Omnibus:	108.921	Durbin-Watson:	1.955
Prob(Omnibus):	0.000	Jarque-Bera (JB):	704.797
Skew:	-0.229	Prob(JB):	9.02e-154
Kurtosis:	7.071	Cond. No.	3.45e+03

Figure 6.42 Acceptance rate versus cost.

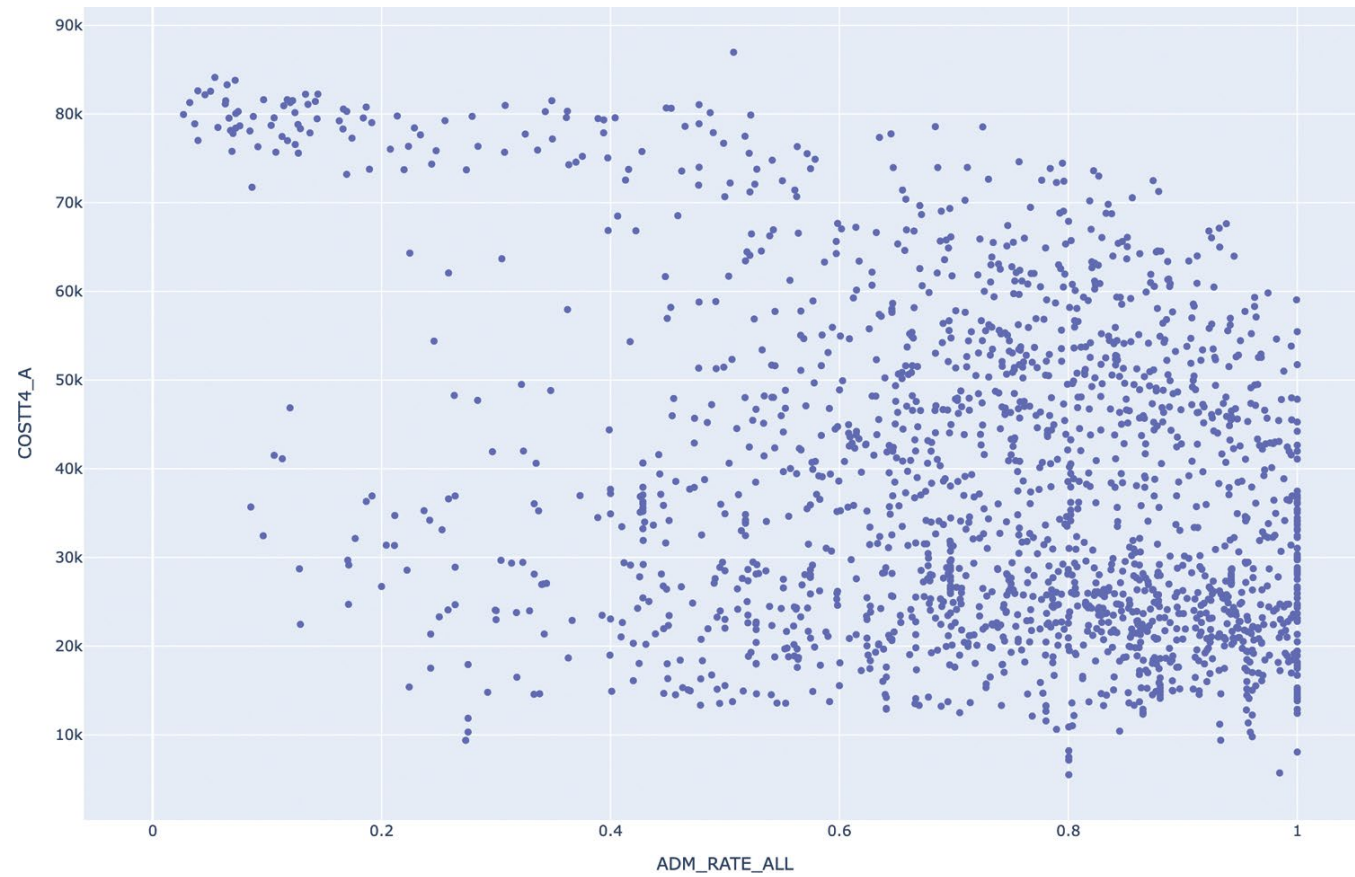


Figure 6.43 Acceptance rate by costs.

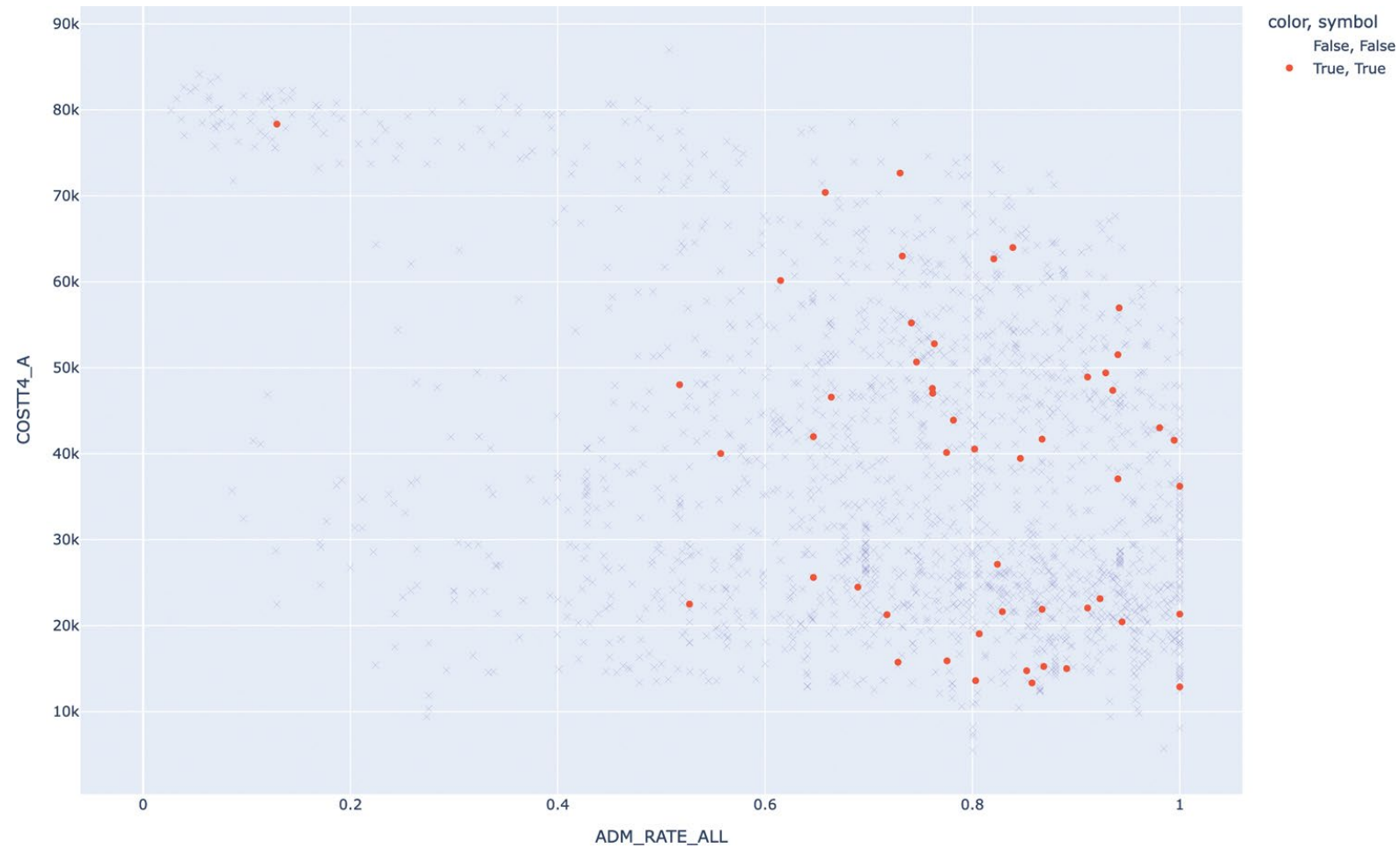


Figure 6.44 Acceptance rate versus cost—Indiana.

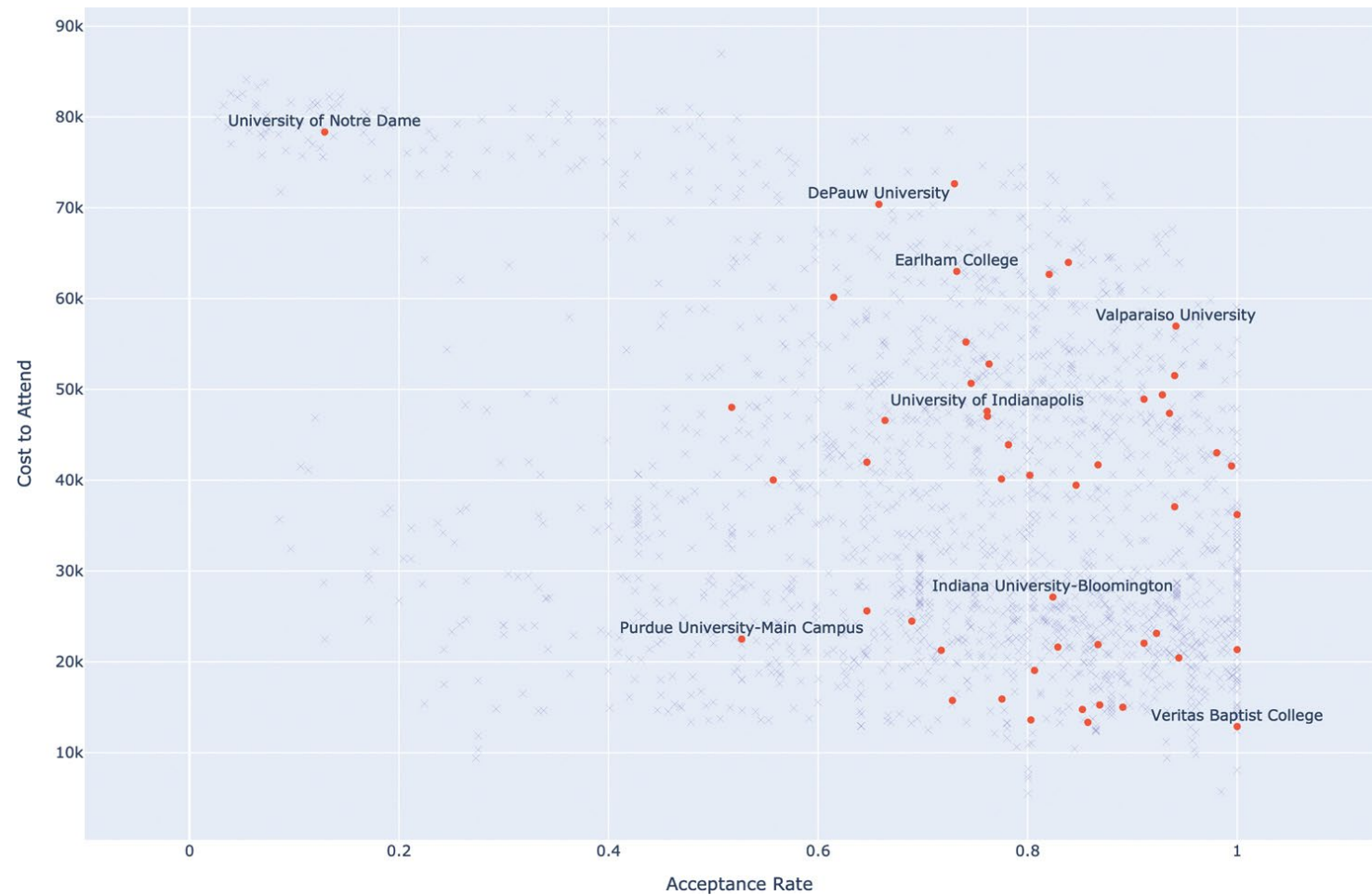


Figure 6.45 Acceptance rate versus cost, labeled in Plotly Express.

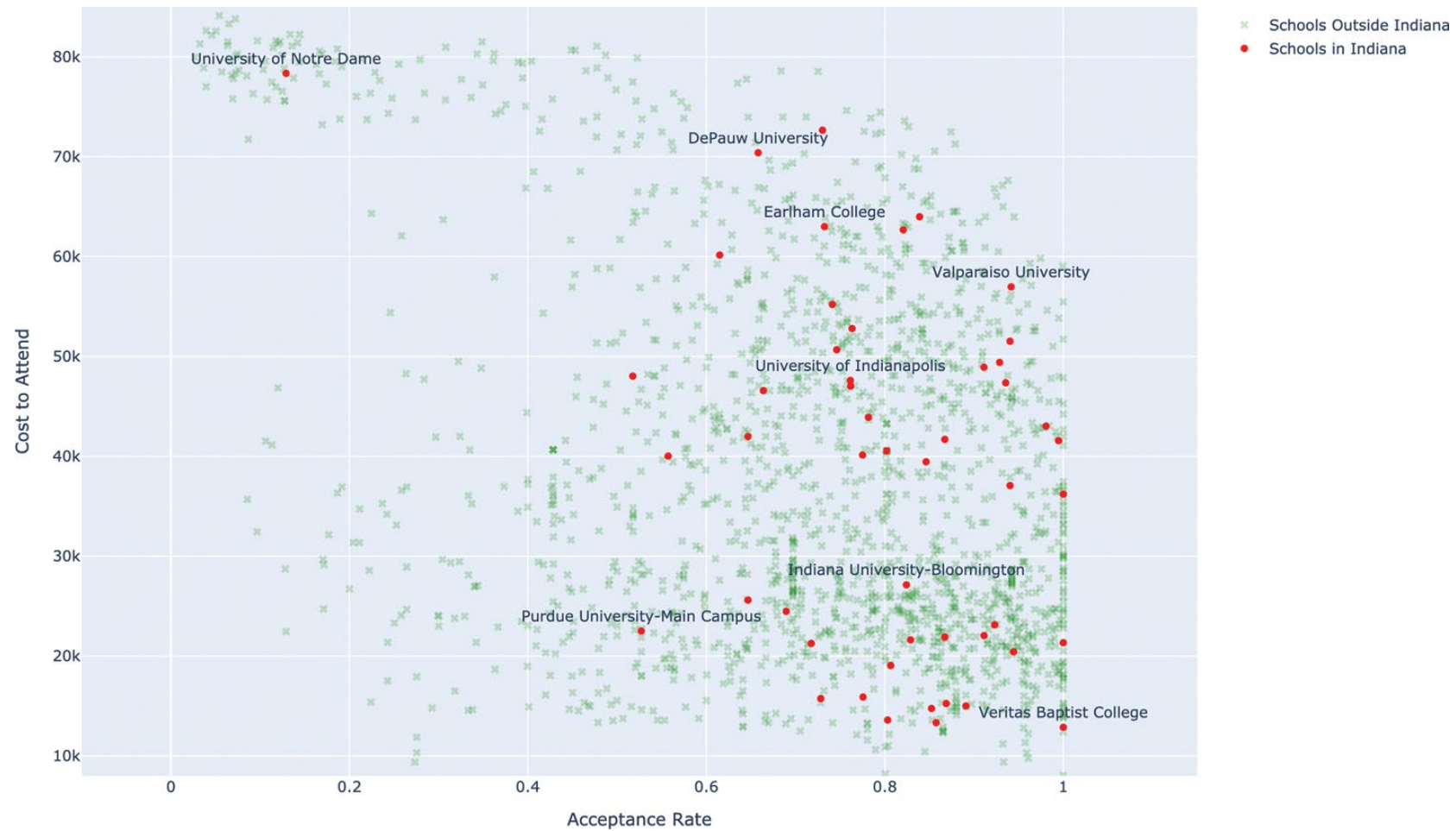


Figure 6.46 Graph Objects scatterplot.

Colleges.R × csc × bach × Univ_df ×								
Filter Cols: « < 1 - 50 > »								
	UNITID	OPEID	OPEID6	INSTNM	CITY	STABBR	ZIP	
1	100654	100200	1002	Alabama A & M University	Normal	AL	35762	
2	100663	105200	1052	University of Alabama at Birmingham	Birmingham	AL	35294-01	
3	100690	2503400	25034	Amridge University	Montgomery	AL	36117-35	
4	100706	105500	1055	University of Alabama in Huntsville	Huntsville	AL	35899	
5	100724	100500	1005	Alabama State University	Montgomery	AL	36104-02	
6	100751	105100	1051	The University of Alabama	Tuscaloosa	AL	35487-01	
7	100760	100700	1007	Central Alabama Community College	Alexander City	AL	35010	
8	100812	100800	1008	Athens State University	Athens	AL	35611	
9	100830	831000	8310	Auburn University at Montgomery	Montgomery	AL	36117-35	
10	100858	100900	1009	Auburn University	Auburn	AL	36849	

Figure 6.47 Viewing the College Scorecard data.frame in R.

	A	B	C	D	E	
1	NAME OF DATA ELEMENT	dev-category	developer-friendly name	API data type	INDEX	VARIABLE NAME
277						
278	Admission rate	admissions	admission_rate.overall	float	float	ADM_RATE
279	Admission rate for all campuses rolled up to the 6-digit OPE ID	admissions	admission_rate.by_ope_id	float		ADM_RATE_ALL
280	25th percentile of SAT scores at the institution (critical reading)	admissions	sat_scores.25th_percentile.critical_reading	float	float	SATVR25
281	75th percentile of SAT scores at the institution (critical reading)	admissions	sat_scores.75th_percentile.critical_reading	float	float	SATVR75
282	25th percentile of SAT scores at the institution (math)	admissions	sat_scores.25th_percentile.math	float	float	SATMT25
283	75th percentile of SAT scores at the institution (math)	admissions	sat_scores.75th_percentile.math	float	float	SATMT75
284	25th percentile of SAT scores at the institution (writing)	admissions	sat_scores.25th_percentile.writing	float	float	SATWR25
285	75th percentile of SAT scores at the institution (writing)	admissions	sat_scores.75th_percentile.writing	float	float	SATWR75
286	Midpoint of SAT scores at the institution (critical reading)	admissions	sat_scores.midpoint.critical_reading	float	float	SATVRMID
287	Midpoint of SAT scores at the institution (math)	admissions	sat_scores.midpoint.math	float	float	SATMTMID
288	Midpoint of SAT scores at the institution (writing)	admissions	sat_scores.midpoint.writing	float	float	SATWRMID
289	25th percentile of the ACT cumulative score	admissions	act_scores.25th_percentile.cumulative	float		ACTCM25
290	75th percentile of the ACT cumulative score	admissions	act_scores.75th_percentile.cumulative	float		ACTCM75
291	25th percentile of the ACT English score	admissions	act_scores.25th_percentile.english	float		ACTEN25
292	75th percentile of the ACT English score	admissions	act_scores.75th_percentile.english	float		ACTEN75
293	25th percentile of the ACT math score	admissions	act_scores.25th_percentile.math	float		ACTMT25
294	75th percentile of the ACT math score	admissions	act_scores.75th_percentile.math	float		ACTMT75
295	25th percentile of the ACT writing score	admissions	act_scores.25th_percentile.writing	float		ACTWR25
296	75th percentile of the ACT writing score	admissions	act_scores.75th_percentile.writing	float		ACTWR75
297	Midpoint of the ACT cumulative score	admissions	act_scores.midpoint.cumulative	float	float	ACTCMMID
298	Midpoint of the ACT English score	admissions	act_scores.midpoint.english	float		ACTENMID
299	Midpoint of the ACT math score	admissions	act_scores.midpoint.math	float		ACTMTMID
300	Midpoint of the ACT writing score	admissions	act_scores.midpoint.writing	float		ACTWRMID
301	Average SAT equivalent score of students admitted	admissions	sat_scores.average.overall	float	float	SAT_AVG
302	Average SAT equivalent score of students admitted for all campuses rolled up to the 6-digit OPE ID	admissions	sat_scores.average.by_ope_id	float		SAT_AVG_ALL
303	Percentage of degrees awarded in Agriculture, Agriculture Operations, And Related Sciences	academics	program_percentage_agriculture	float		PC1001

Figure 6.48 The data dictionary for the College Scorecard data.

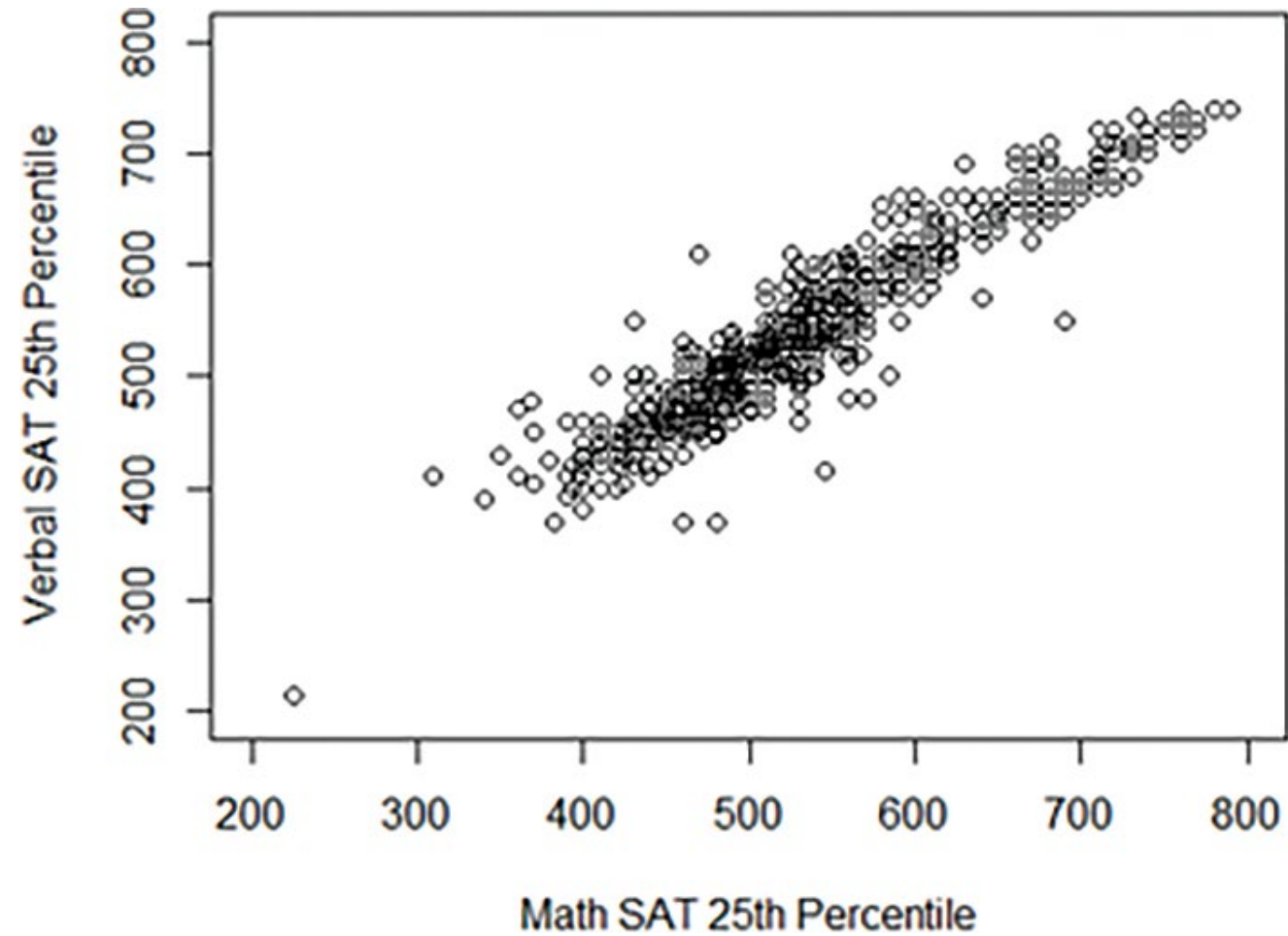


Figure 6.49 Scatterplot of verbal versus math 25th percentiles by institution using the base R plot function.

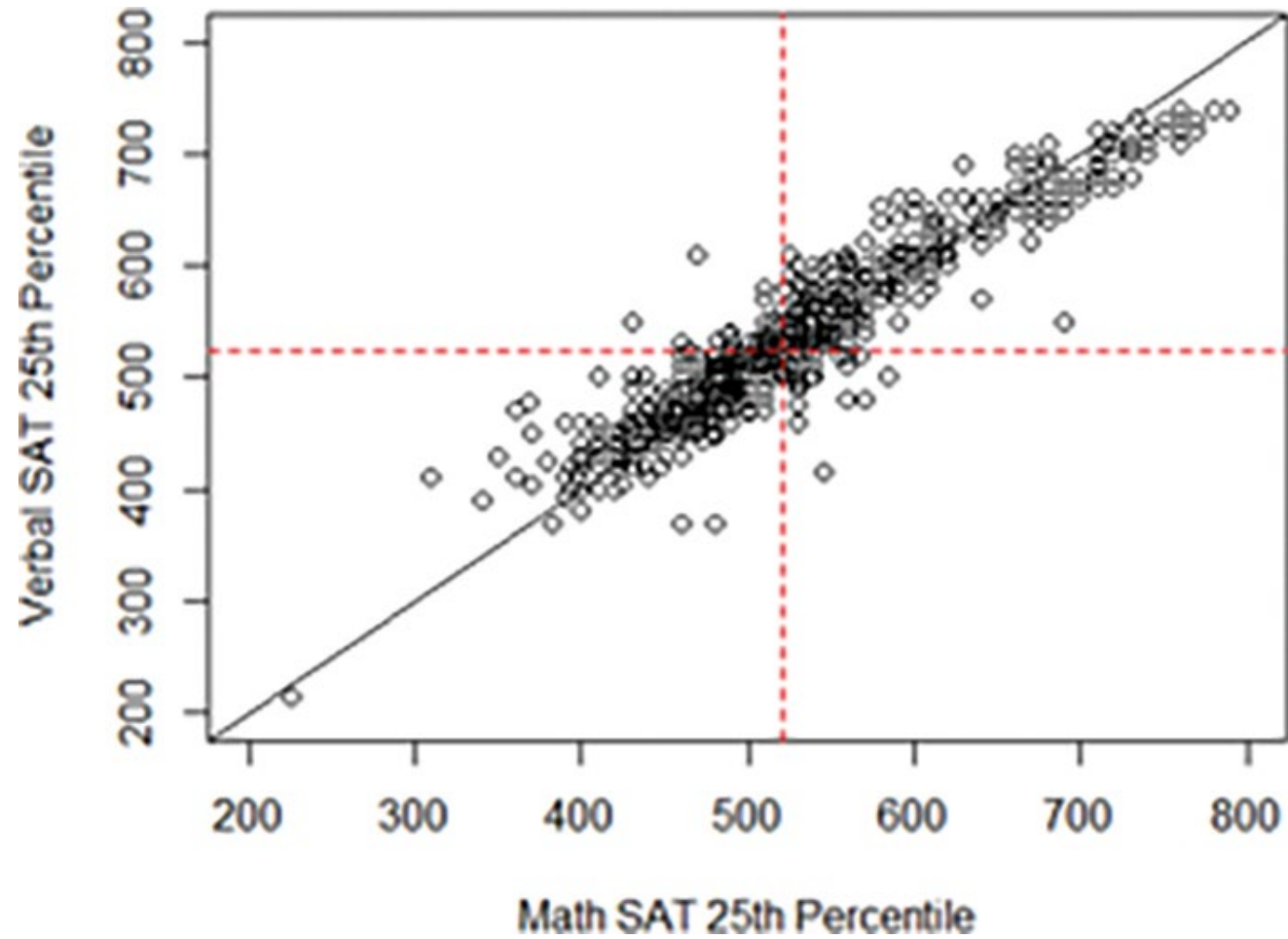


Figure 6.50 Scatterplot of verbal versus math 25th percentiles by institution with median and $y = x$ lines.

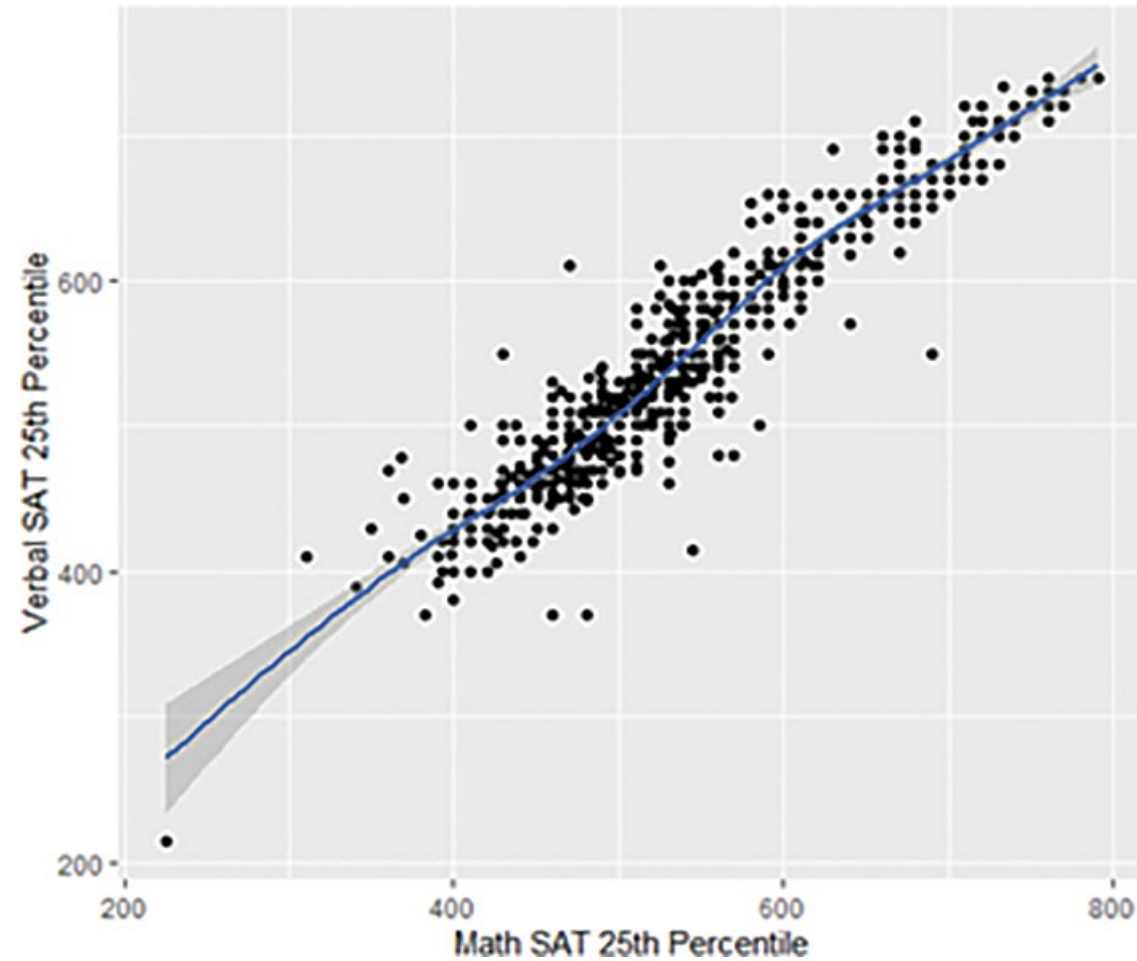


Figure 6.51 Scatterplot of verbal versus math 25th percentiles by institution, created with a smoother using ggplot in R.

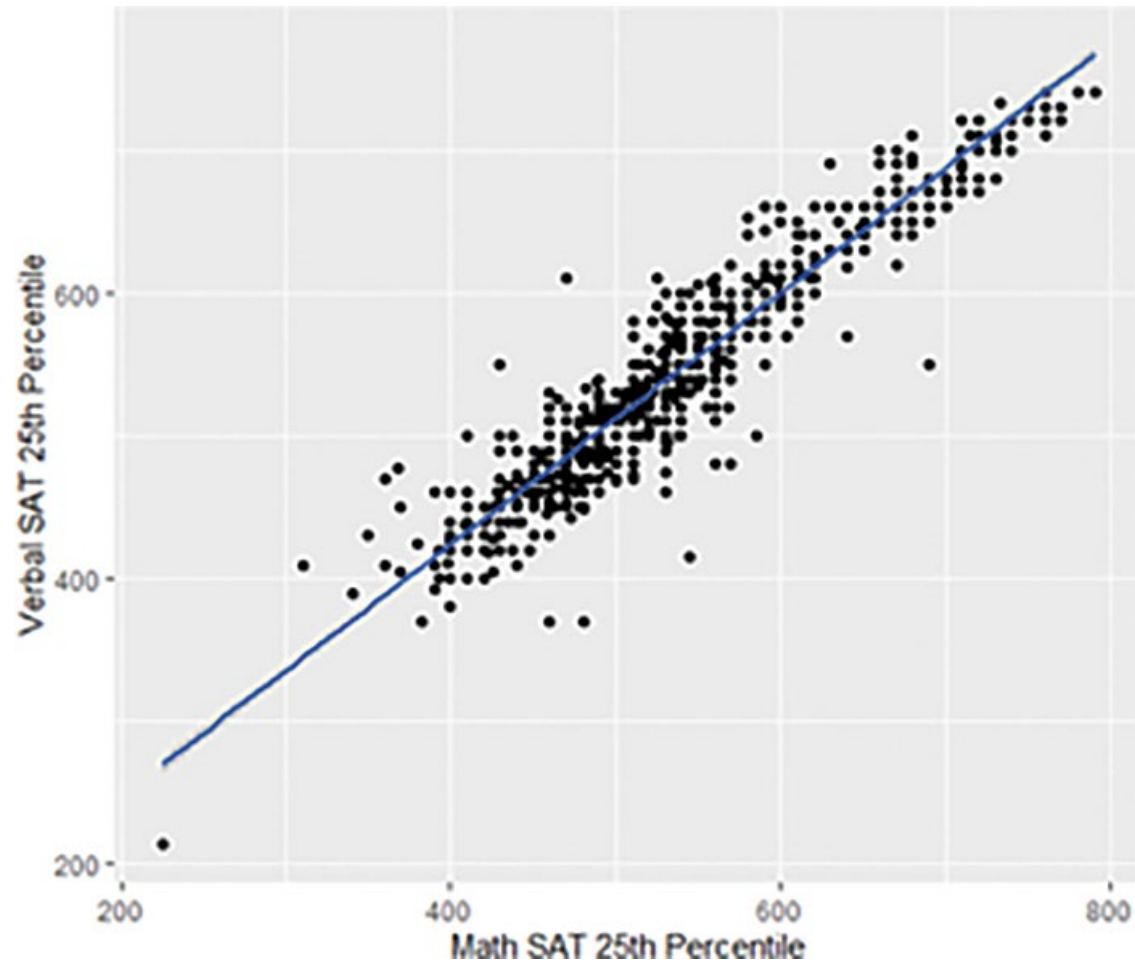
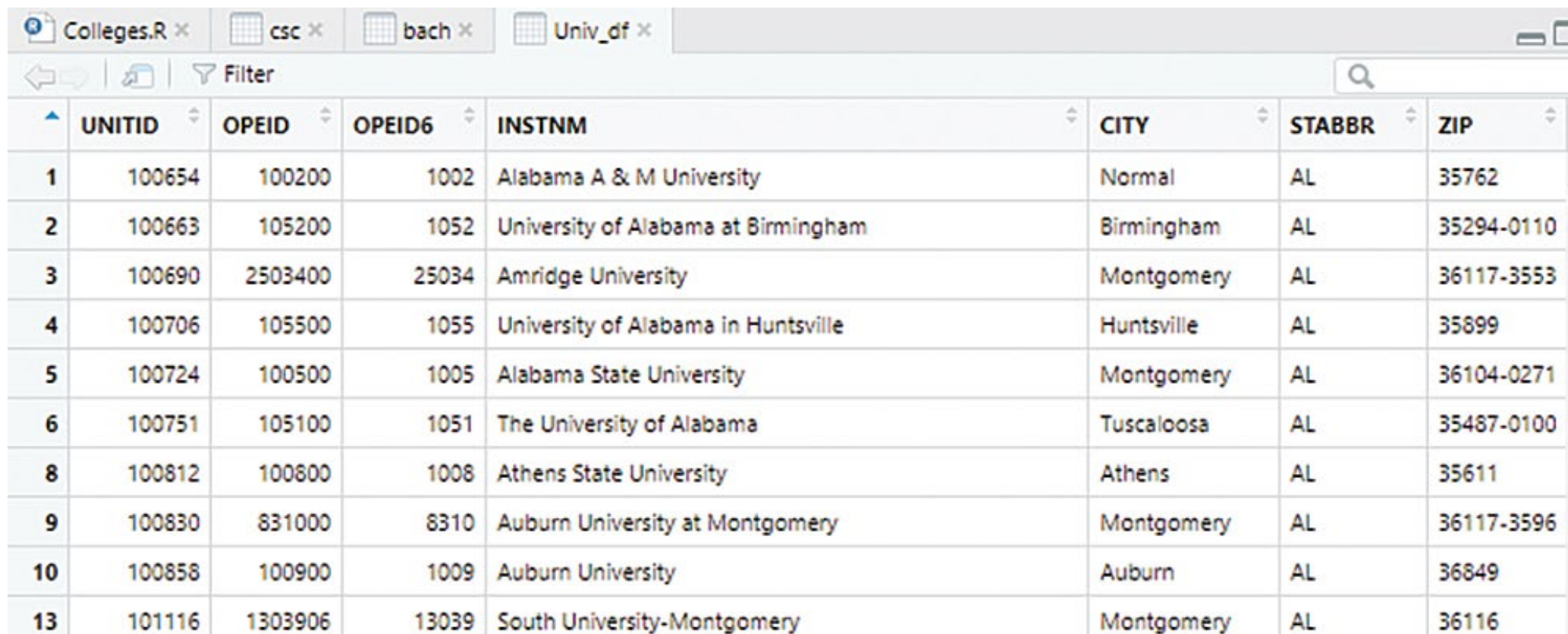


Figure 6.52 Scatterplot of verbal versus math 25th percentiles by institution with linear model, created using ggplot in R.



	UNITID	OPEID	OPEID6	INSTNM	CITY	STABBR	ZIP
1	100654	100200	1002	Alabama A & M University	Normal	AL	35762
2	100663	105200	1052	University of Alabama at Birmingham	Birmingham	AL	35294-0110
3	100690	2503400	25034	Amridge University	Montgomery	AL	36117-3553
4	100706	105500	1055	University of Alabama in Huntsville	Huntsville	AL	35899
5	100724	100500	1005	Alabama State University	Montgomery	AL	36104-0271
6	100751	105100	1051	The University of Alabama	Tuscaloosa	AL	35487-0100
8	100812	100800	1008	Athens State University	Athens	AL	35611
9	100830	831000	8310	Auburn University at Montgomery	Montgomery	AL	36117-3596
10	100858	100900	1009	Auburn University	Auburn	AL	36849
13	101116	1303906	13039	South University-Montgomery	Montgomery	AL	36116

Figure 6.53 The data.frame of universities using the grep function to identify this subset.

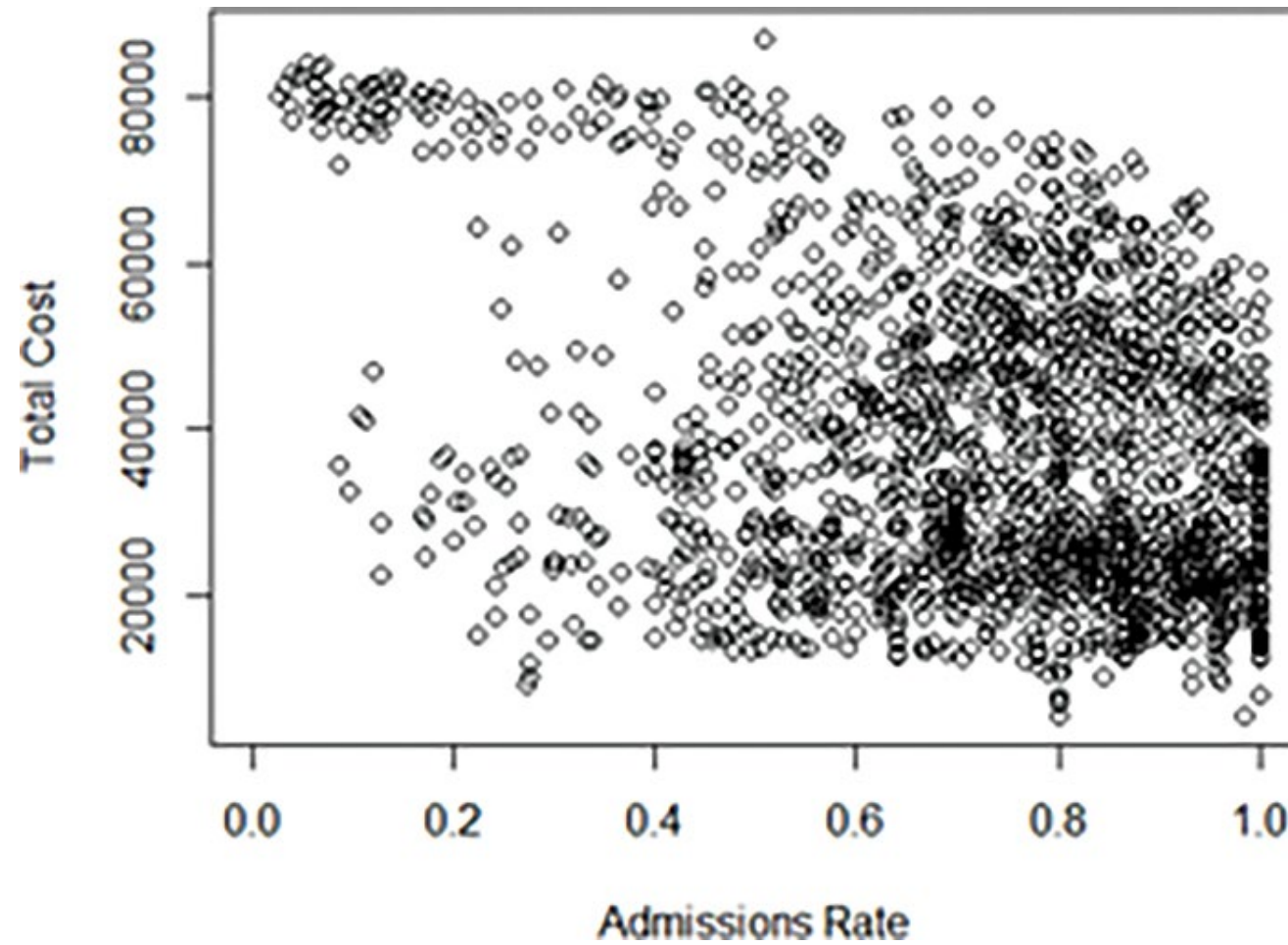


Figure 6.54 Scatterplot of total cost versus admissions rate by institution, created using the base R plot function.

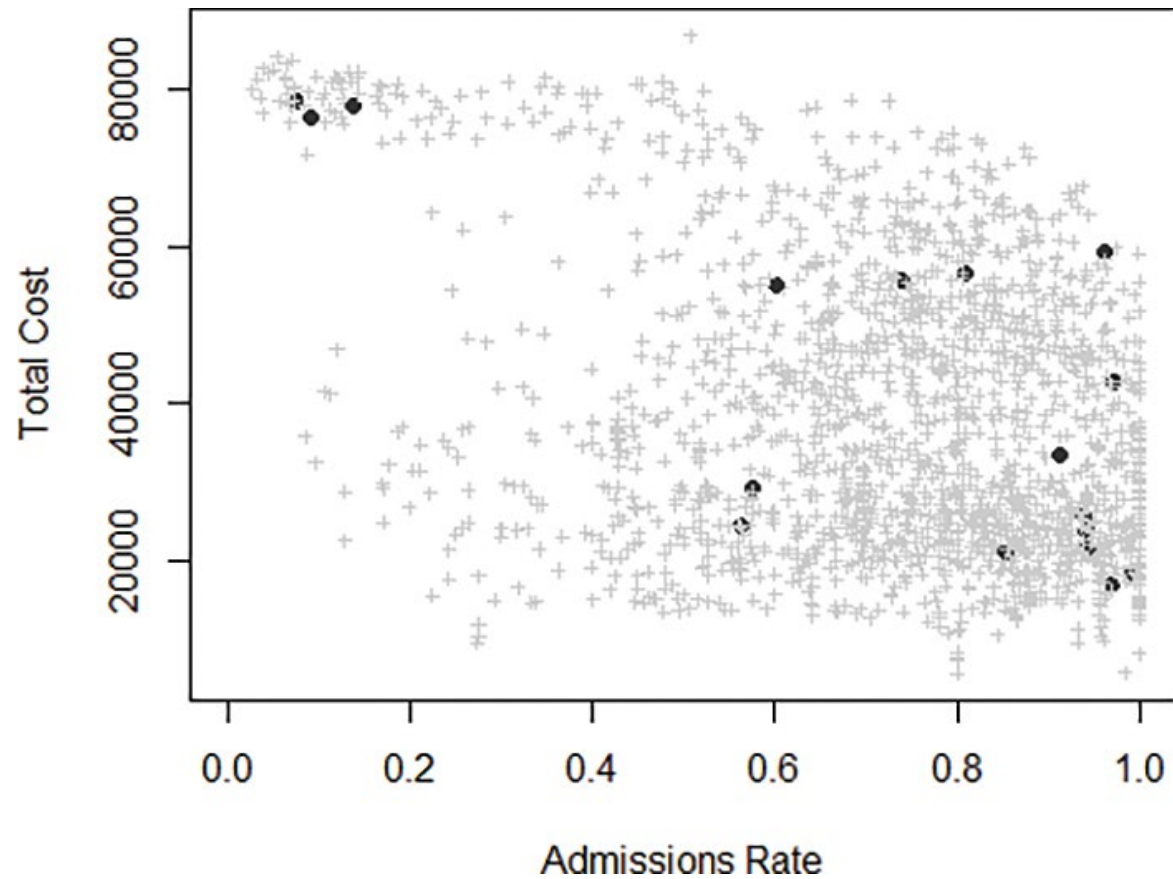


Figure 6.55 Scatterplot of total cost versus admissions rate by institution, created using the base R plot function with formatting and Maine schools highlighted.

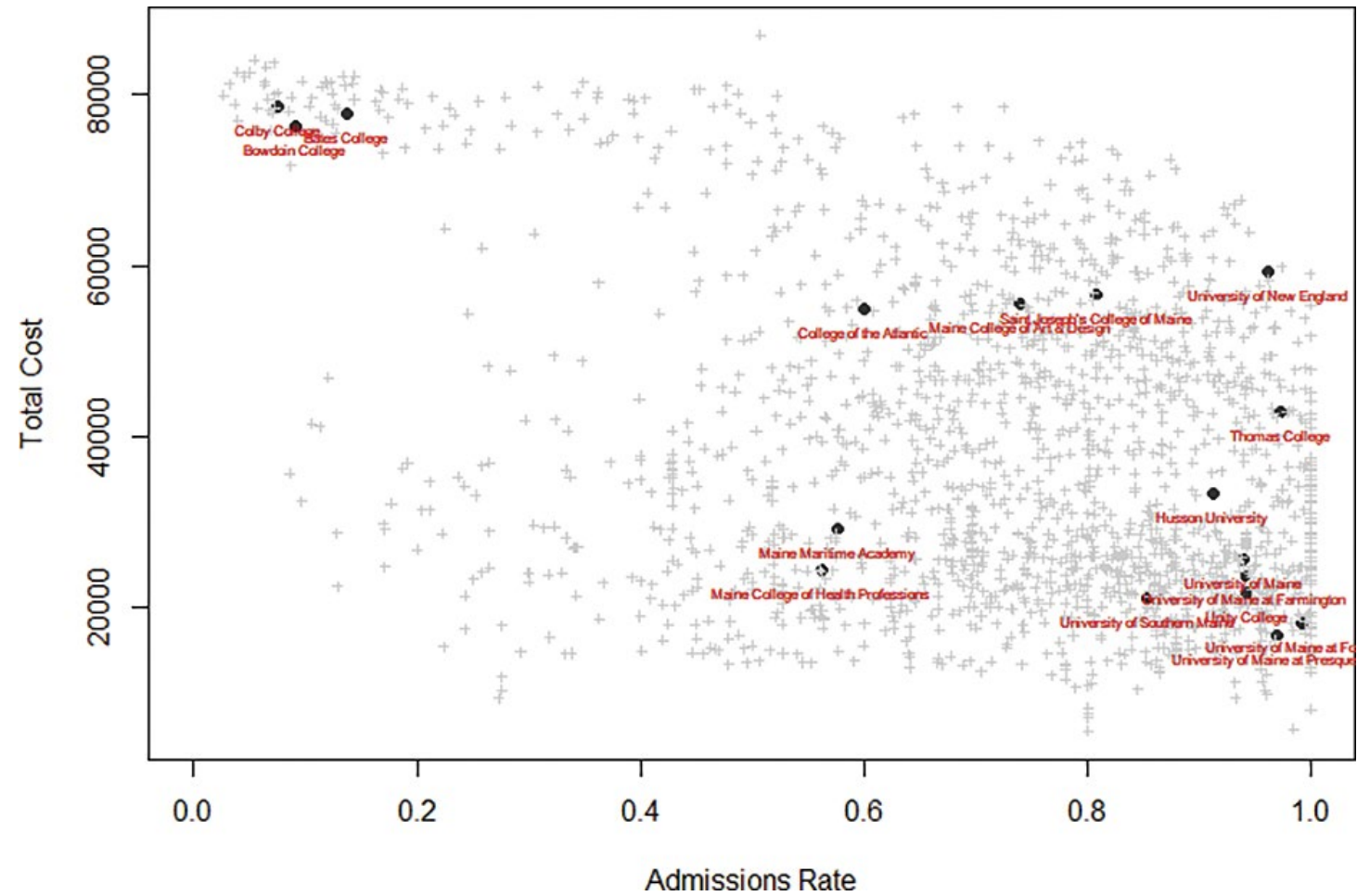


Figure 6.56 Scatterplot of total cost versus admissions rate by institution, created using the base R plot function with Maine schools highlighted.

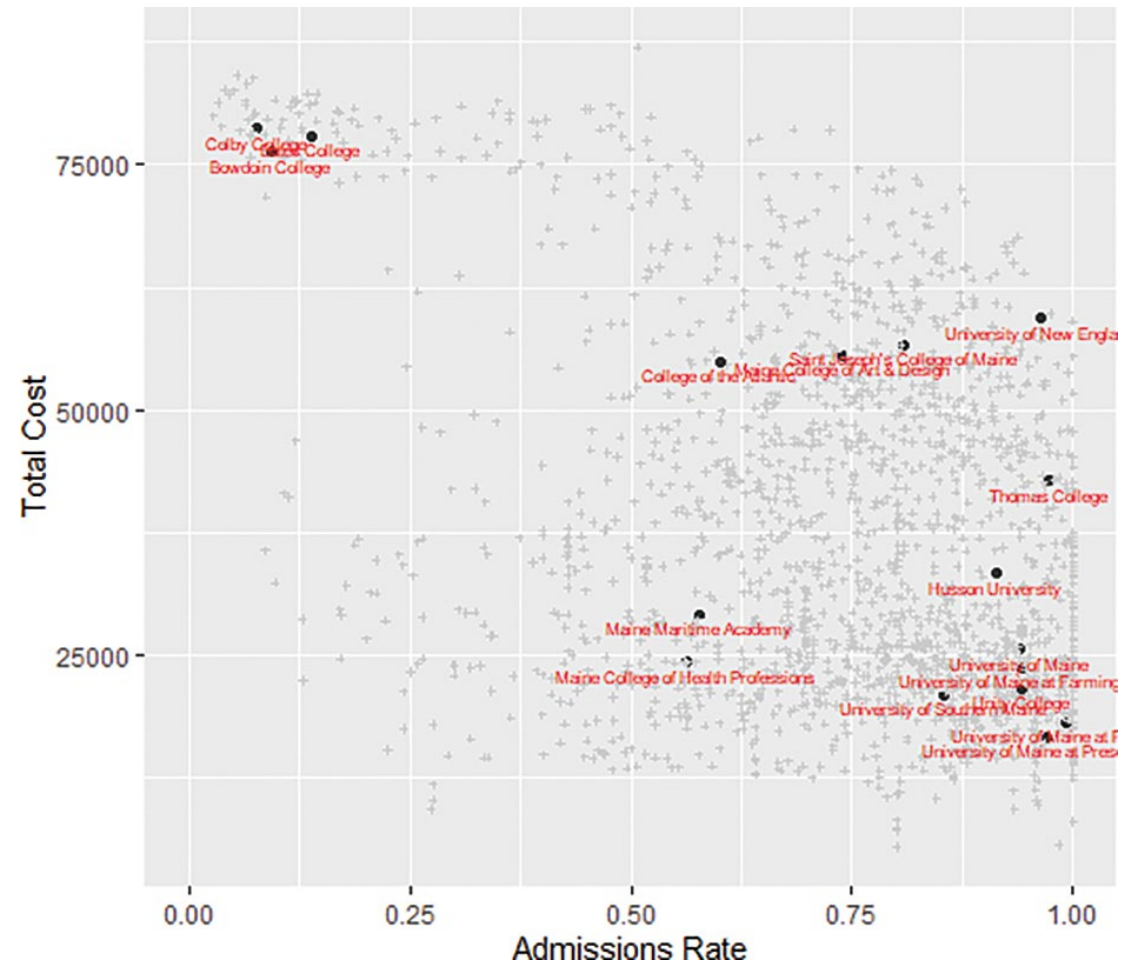


Figure 6.57 Scatterplot of total cost versus admissions rate by institution, created using the ggplot function with Maine schools highlighted.

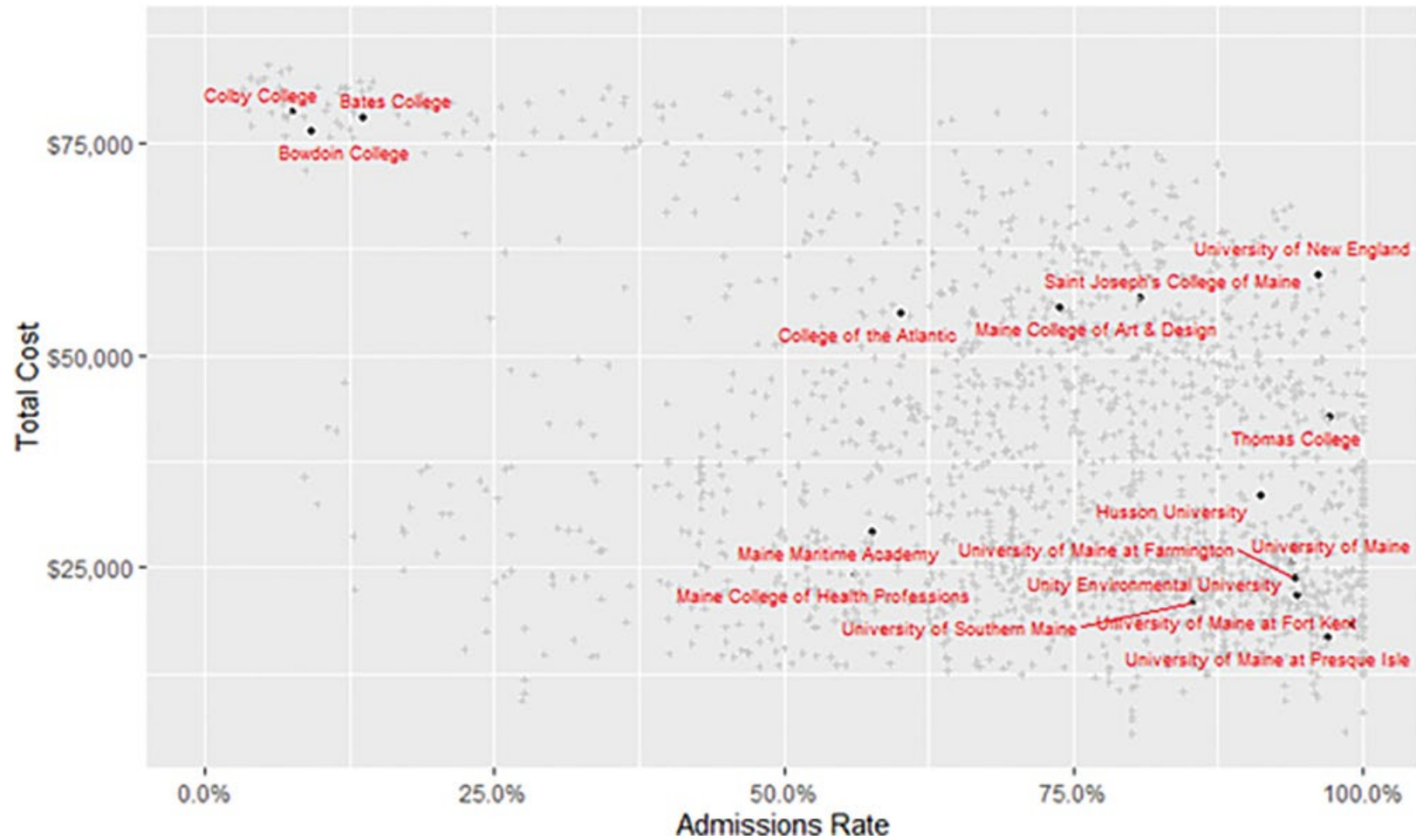


Figure 6.58 Scatterplot of total cost versus admissions rate by institution, created using the ggplot function with Maine schools highlighted and the ggrepel package for the labels.



 [Home](#)

[Research Topics](#)

[Research & Training](#)

[Clinical Resources](#)

[Grants & Funding](#)

[Home](#) / [Research Topics](#) / [Trends & Statistics](#) / [Drug Overdose Deaths: Facts and Figures](#)

Drug Overdose Deaths: Facts and Figures

Highlights

- The [National Center for Health Statistics](#) (NCHS) at the U.S. Centers for Disease Control and Prevention (CDC) reports data on fatalities involving drugs commonly associated with fatal overdose. [NIDA analyzes NCHS data](#) to create the figures and descriptions on this page. For more information or to contact CDC, please visit the [CDC's Drug Overdose Deaths](#) web page.
- Timely data related to unintentional and undetermined intent drug overdose deaths by participating jurisdiction is also available from the CDC's [State Unintentional Drug Overdose Reporting System](#) (SUDORS) system.
- Data related to nonfatal drug overdoses are reported through the [CDC's Drug Overdose Surveillance and Epidemiology](#) (DOSE) system.

Figure 7.1 National Institute on Drug Abuse webpage.

Download Links

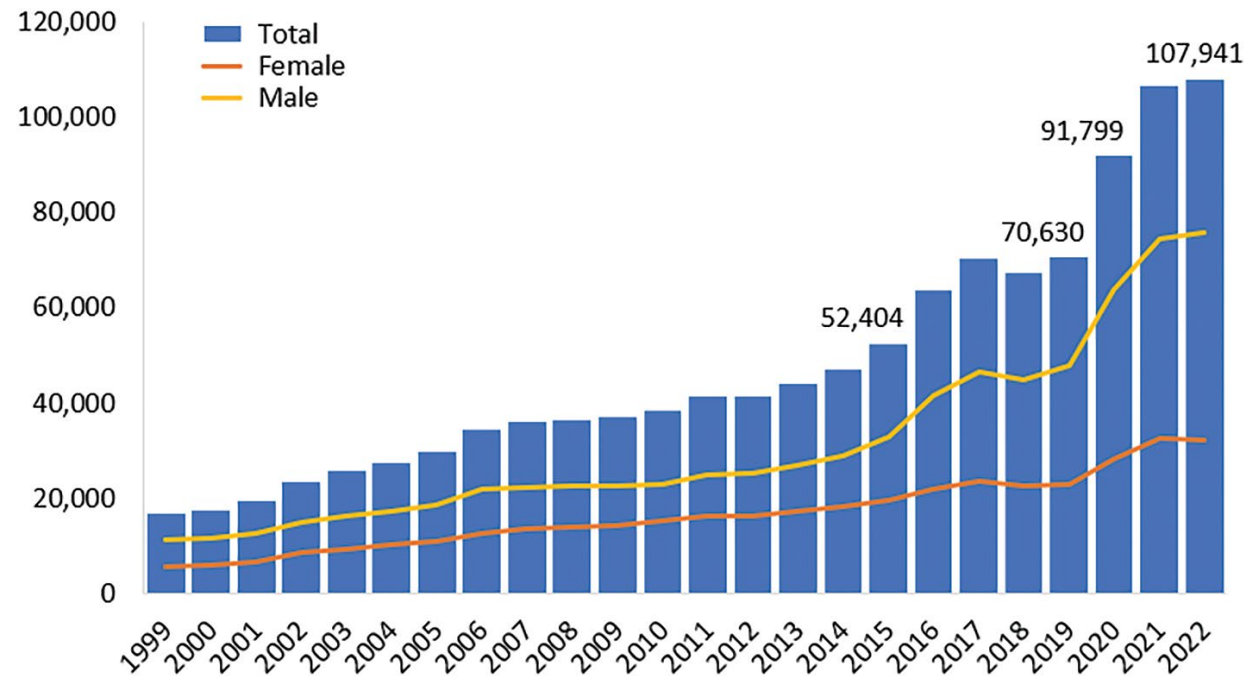
- Download graphs in PowerPoint - [National Drug Involved Overdose Deaths 1999-2022.pptx](#)
- Download the data used to create these figures - [Overdose_data_1999-2022 5.7.24.xlsx](#)

Figure 7.2 National Institute on Drug Abuse webpage interface for downloading the data.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	
1		 National Institute on Drug Abuse																										
2		Number of National Drug Overdose Deaths*	Involving Select Prescription and Illicit Drugs																									
3		Source: National Center on Health Statistics, CDC WONDER																										
4																												
5																												
6																												
7			1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2015-2022 Fold Change	
8		Total Overdose Deaths	16,849	17,415	19,394	23,518	25,785	27,424	29,813	34,425	36,010	36,450	37,004	38,329	41,340	41,502	43,982	47,055	52,404	63,632	70,237	67,367	70,630	91,799	106,699	107,941	2.1	
9		Female	5,591	5,852	6,736	8,490	9,386	10,304	11,089	12,532	13,712	13,982	14,411	15,323	16,352	16,390	17,183	18,243	19,447	22,074	23,685	22,426	22,749	28,071	32,398	32,127	1.7	
10		Male	11,258	11,563	12,658	15,028	16,399	17,120	18,724	21,893	22,298	22,468	22,593	23,006	24,988	25,112	26,799	28,812	32,957	41,558	46,552	44,941	47,881	63,728	74,301	75,814	2.3	
11		Any Opioid¹	8,050	8,407	9,496	11,920	12,940	13,756	14,918	17,545	18,516	19,582	20,422	21,089	22,784	23,166	25,052	28,647	33,091	42,249	47,600	46,802	49,860	68,630	80,411	81,806	2.5	
12		Female	2,057	2,264	2,767	3,760	4,138	4,643	5,161	5,945	6,581	6,819	7,287	7,734	8,325	8,432	9,055	10,227	11,420	13,751	15,263	14,724	15,225	19,970	23,654	23,421	2.1	
13		Male	5,993	6,143	6,729	8,160	8,802	9,113	9,757	11,600	11,935	12,763	13,135	13,355	14,459	14,734	15,997	18,420	21,671	28,498	32,337	32,078	34,635	48,660	56,757	58,385	2.7	
14		Select Opioids^{1a} AND Synthetic Opioids other than Methadone	167	205	225	358	383	428	485	781	642	703	923	1,002	961	944	1,219	2,493	4,806	9,299	12,556	13,491	13,596	16,465	15,444	20,413	4.2	
15		Prescription Opioids²	3,442	3,785	4,770	6,483	7,461	8,577	9,612	11,589	12,796	13,149	13,523	14,583	15,140	14,240	14,145	14,838	15,281	17,087	17,029	14,975	14,139	16,416	16,706	14,716	1.0	
16		Female	1,022	1,236	1,608	2,304	2,681	3,144	3,572	4,274	4,863	4,959	5,212	5,644	6,082	5,995	6,049	6,506	6,664	7,109	7,156	6,252	5,755	6,441	6,623	5,713	0.9	
17		Male	2,420	2,549	3,162	4,179	4,780	5,433	6,040	7,315	7,933	8,190	8,311	8,939	9,058	8,245	8,096	8,332	8,617	9,978	9,873	8,723	8,384	9,975	10,083	9,003	1.0	
18		Prescription Opioids AND Synthetic Opioids other than Methadone	142	167	199	322	344	384	426	573	601	655	872	939	889	861	1,015	1,489	2,263	4,055	5,444	5,417	5,876	8,626	9,644	8,310	3.7	
19		Female	65	76	86	157	151	184	207	246	286	309	444	453	426	445	488	661	898	1,394	1,859	1,872	1,949	2,798	6,393	2,635	2.9	
20		Male	77	91	113	165	193	200	219	327	315	346	428	486	463	416	527	828	1,365	2,661	3,585	3,545	3,927	5,828	3,251	5,675	4.2	
21		Prescription Opioids WITHOUT Synthetic Opioids other than Methadone	3,300	3,618	4,571	6,161	7,117	8,193	9,186	11,016	12,195	12,494	12,651	13,644	14,251	13,379	13,130	13,349	13,018	13,032	11,585	9,558	8,263	7,790	7,062	6,406	0.5	
22		Female	957	1,160	1,522	2,147	2,530	2,960	3,365	4,028	4,577	4,650	4,768	5,191	5,656	5,550	5,561	5,845	5,766	5,715	5,297	4,380	3,806	3,643	230	3,078	0.5	
23		Male	2,343	2,458	3,049	4,014	4,587	5,233	5,821	6,988	7,618	7,844	7,883	8,453	8,595	7,829	7,569	7,504	7,252	7,317	6,288	5,178	4,457	4,147	6,832	3,328	0.5	
24		Synthetic Opioids other than Methadone (primarily fentanyl)³	730	782	957	1,295	1,400	1,664	1,742	2,707	2,213	2,306	2,946	3,007	2,666	2,628	3,105	5,544	9,580	19,413	28,466	31,335	36,359	56,516	70,601	73,838	7.7	
25		Female	330	374	447	614	643	798	823	1,030	1,053	1,083	1,445	1,440	1,247	1,195	1,431	2,079	3,020	5,578	7,942	8,807	10,076	15,250	19,571	19,880	6.6	
26		Male	400	408	510	681	757	866	919	1,677	1,160	1,223	1,501	1,567	1,419	1,433	1,674	3,465	6,560	13,835	20,524	22,528	26,283	41,266	51,030	53,958	8.2	
27		Heroin⁴	1,960	1,842	1,779	2,089	2,080	1,878	2,009	2,088	2,399	3,041	3,278	3,036	4,397	5,925	8,257	10,574	12,989	15,469	15,482	14,996	14,019	13,165	9,173	5,871	0.5	
28		Female	306	279	313	359	358	341	389	344	399	551	577	584	878	1,213	1,732	2,414	3,108	3,717	3,886	3,705	3,520	3,284	2,372	1,504	0.5	

Figure 7.3 National Institute on Drug Abuse overdose death data.

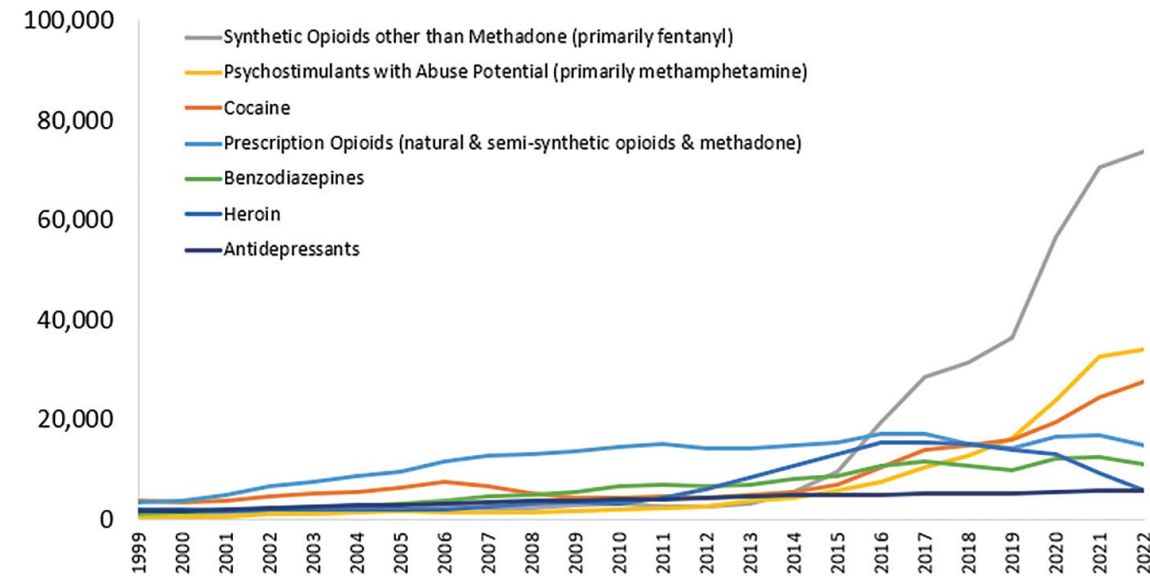
Figure 1. U.S. Overdose Deaths* by Sex, 1999-2022



*Includes deaths with underlying causes of unintentional drug poisoning (X40–X44), suicide drug poisoning (X60–X64), homicide drug poisoning (X85), or drug poisoning of undetermined intent (Y10–Y14), as coded in the International Classification of Diseases, 10th Revision. Source: Centers for Disease Control and Prevention, National Center for Health Statistics. Multiple Cause of Death 1999-2022 on CDC WONDER Online Database, released 4/2024.

Figure 7.4 Total U.S. overdose deaths by sex, 1999–2023, from the NIDA website.

Figure 2. U.S. Overdose Deaths*, Select Drugs or Drug Categories, 1999-2022



*Includes deaths with underlying causes of unintentional drug poisoning (X40–X44), suicide drug poisoning (X60–X64), homicide drug poisoning (X85), or drug poisoning of undetermined intent (Y10–Y14), as coded in the International Classification of Diseases, 10th Revision. Source: Centers for Disease Control and Prevention, National Center for Health Statistics. Multiple Cause of Death 1999-2022 on CDC WONDER Online Database, released 4/2024.

Figure 7.5 U.S. overdose deaths by drug category, 1999–2023, from the NIDA website.

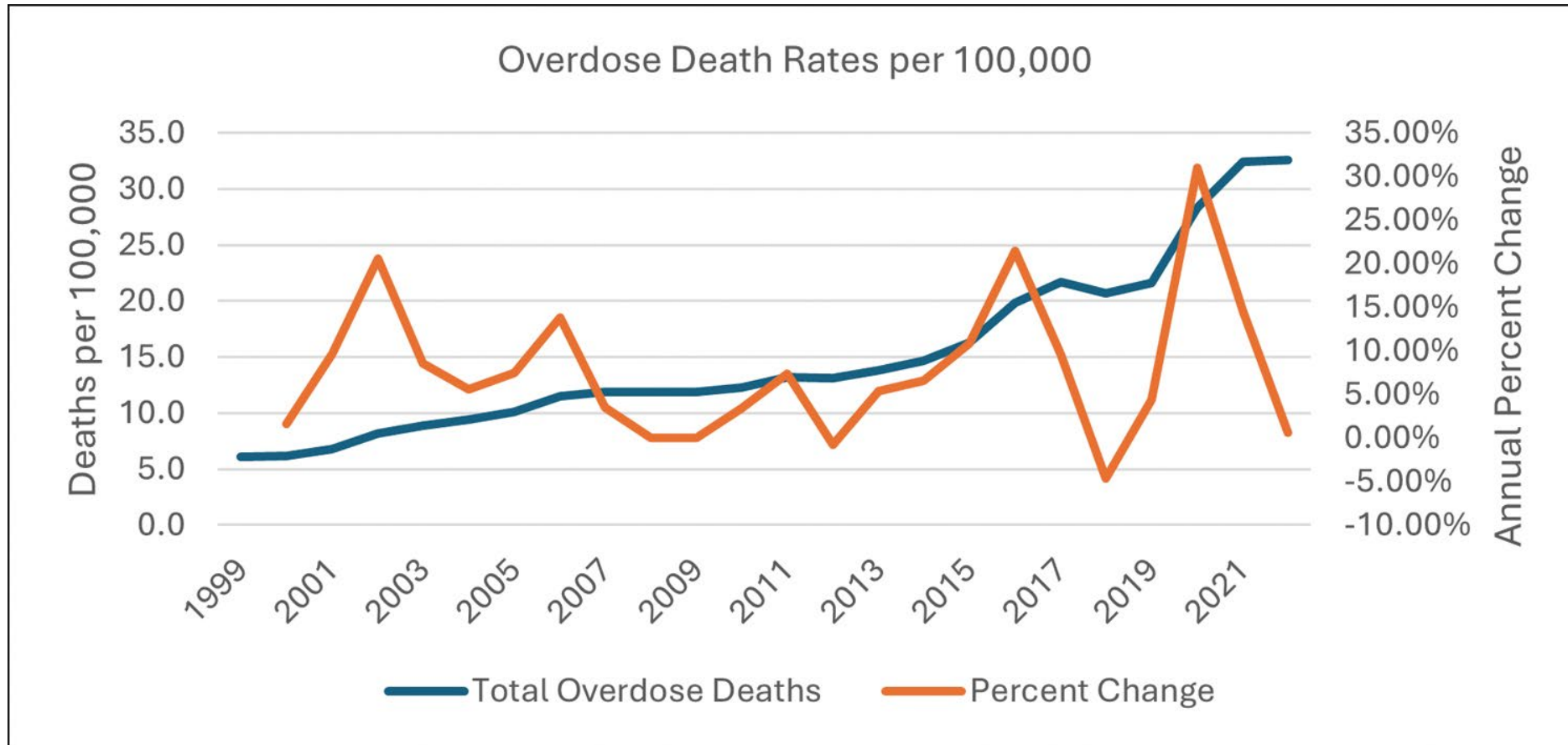


Figure 7.6 Total U.S. overdose death rates, 1999–2022, with annual percent change.

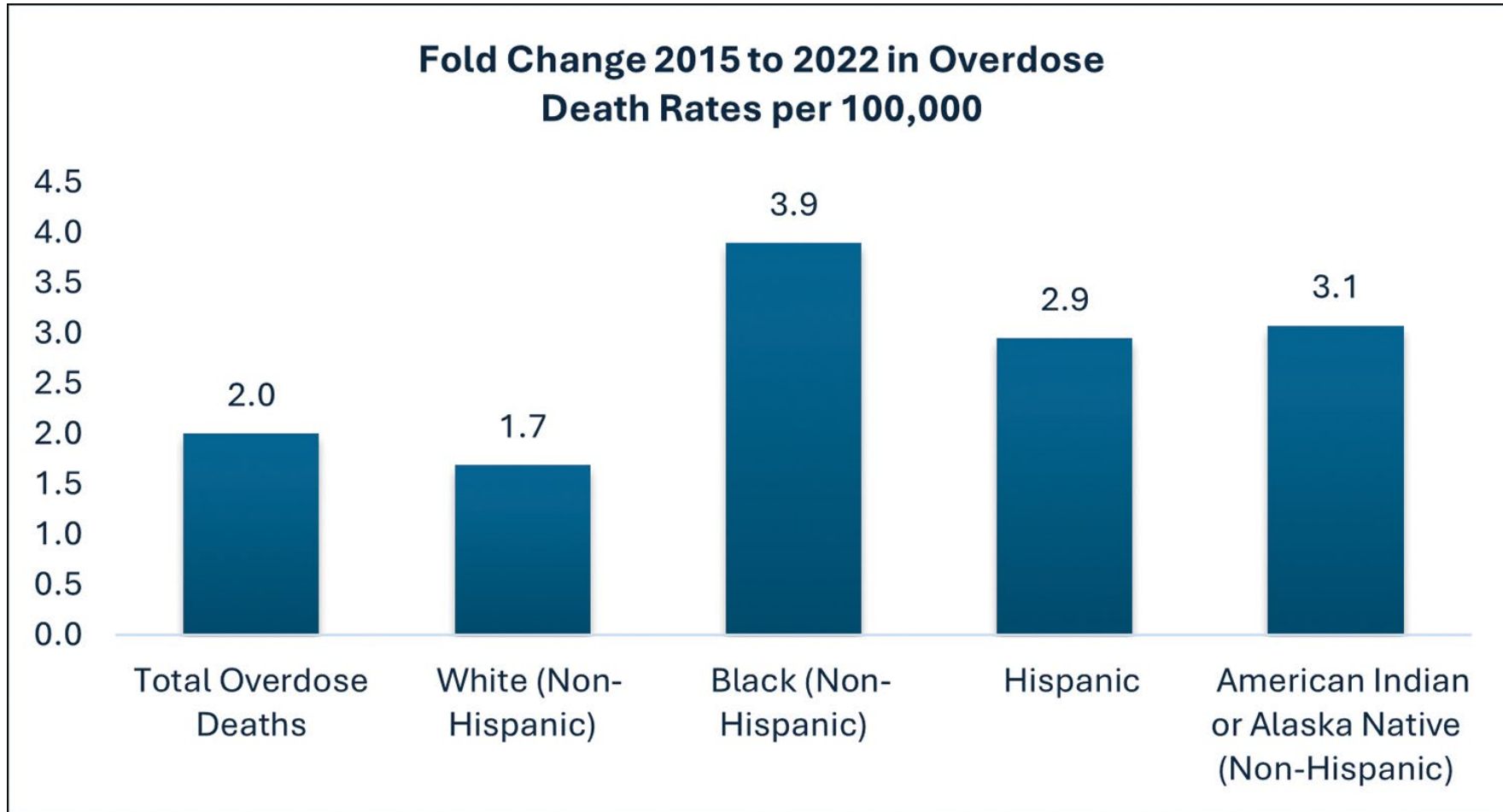


Figure 7.7 Change (as a multiple) in U.S. overdose death rates, 2015–2022, by demographic group.

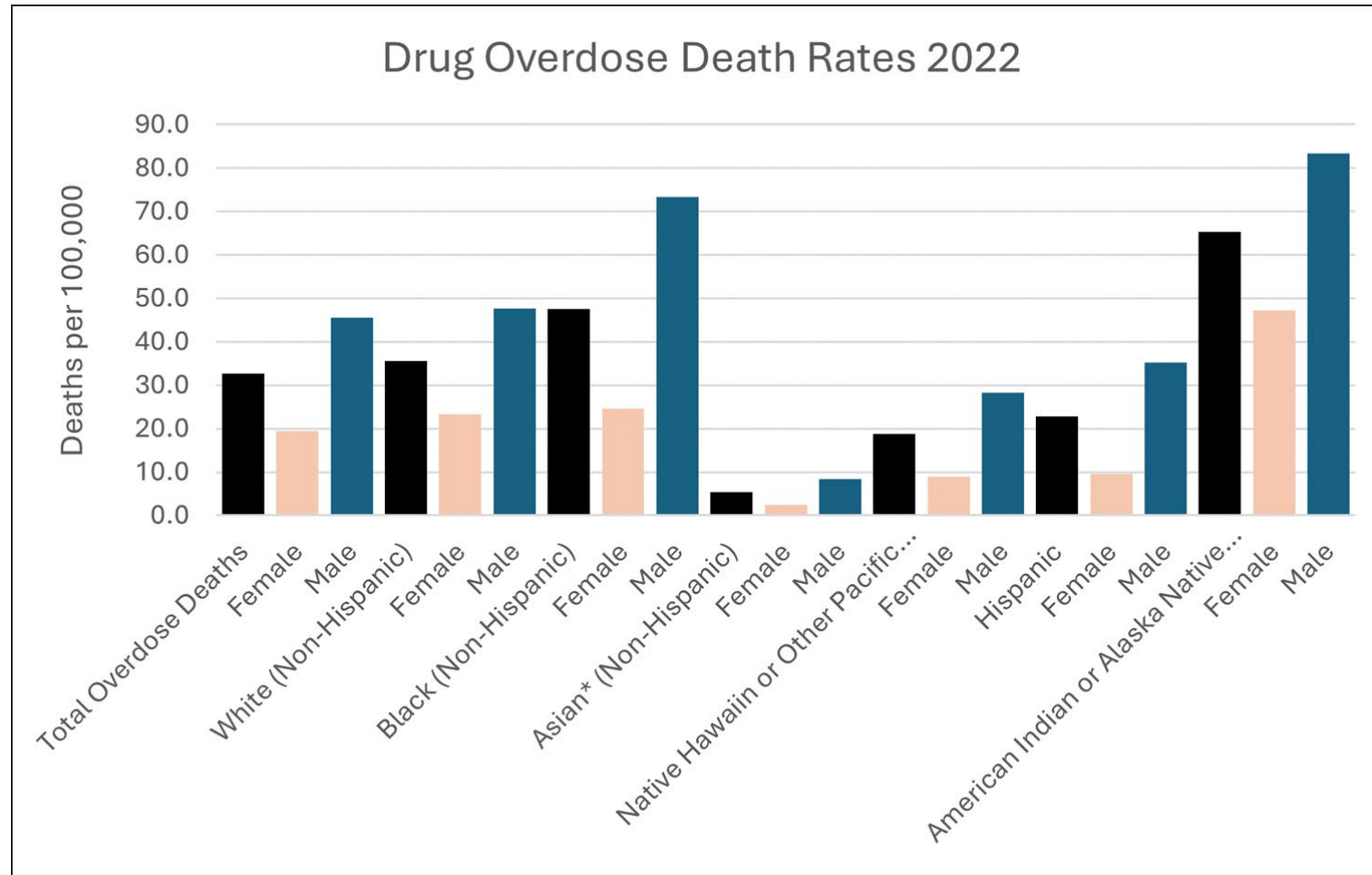
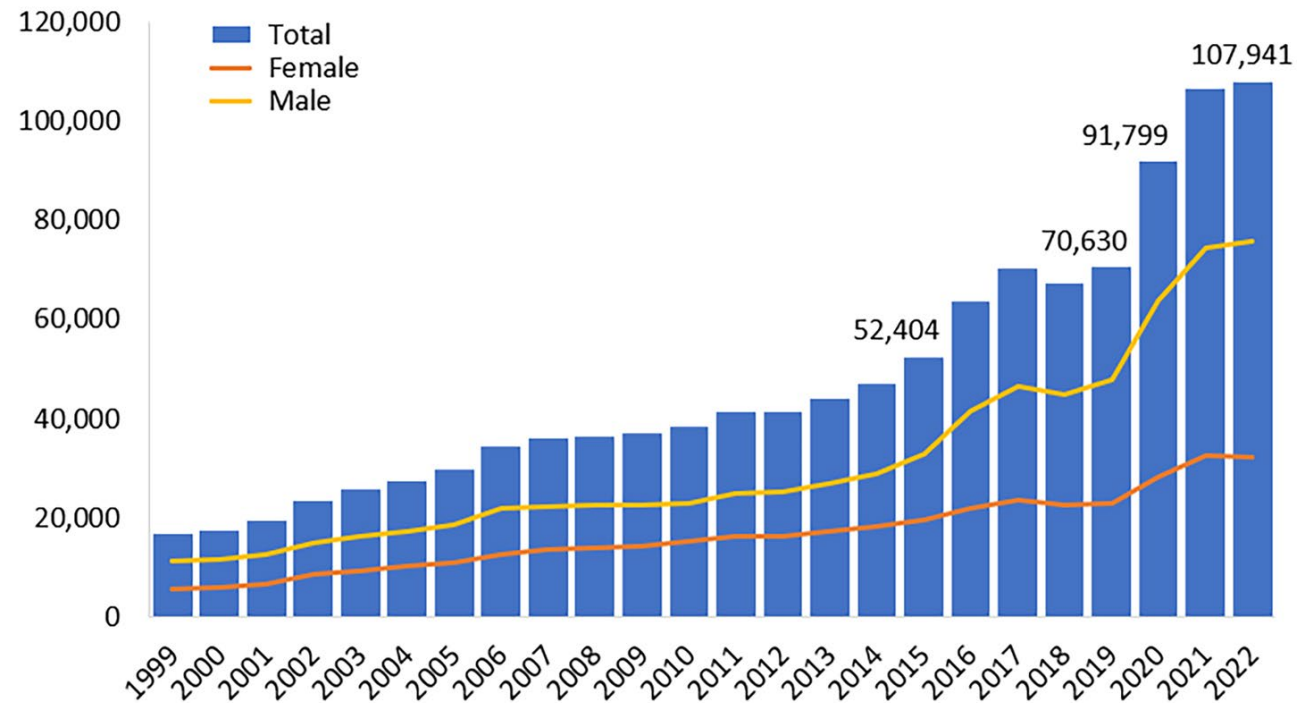


Figure 7.8 U.S. overdose death rates, 2022, by demographic group and sex.

Figure 1. U.S. Overdose Deaths* by Sex, 1999-2022



*Includes deaths with underlying causes of unintentional drug poisoning (X40–X44), suicide drug poisoning (X60–X64), homicide drug poisoning (X85), or drug poisoning of undetermined intent (Y10–Y14), as coded in the International Classification of Diseases, 10th Revision. Source: Centers for Disease Control and Prevention, National Center for Health Statistics. Multiple Cause of Death 1999-2022 on CDC WONDER Online Database, released 4/2024.

Figure 7.9 Total U.S. overdose deaths by sex, 1999–2023, from the NIDA website.

	A	B	C	D
1		Total Over	Female	Male
2	1999	16,849	5,591	11,258
3	2000	17,415	5,852	11,563
4	2001	19,394	6,736	12,658
5	2002	23,518	8,490	15,028
6	2003	25,785	9,386	16,399
7	2004	27,424	10,304	17,120
8	2005	29,813	11,089	18,724
9	2006	34,425	12,532	21,893
10	2007	36,010	13,712	22,298
11	2008	36,450	13,982	22,468
12	2009	37,004	14,411	22,593
13	2010	38,329	15,323	23,006
14	2011	41,340	16,352	24,988

Figure 7.10 Total U.S. overdose deaths by sex, 1999–2022, after copying and transposing the data.

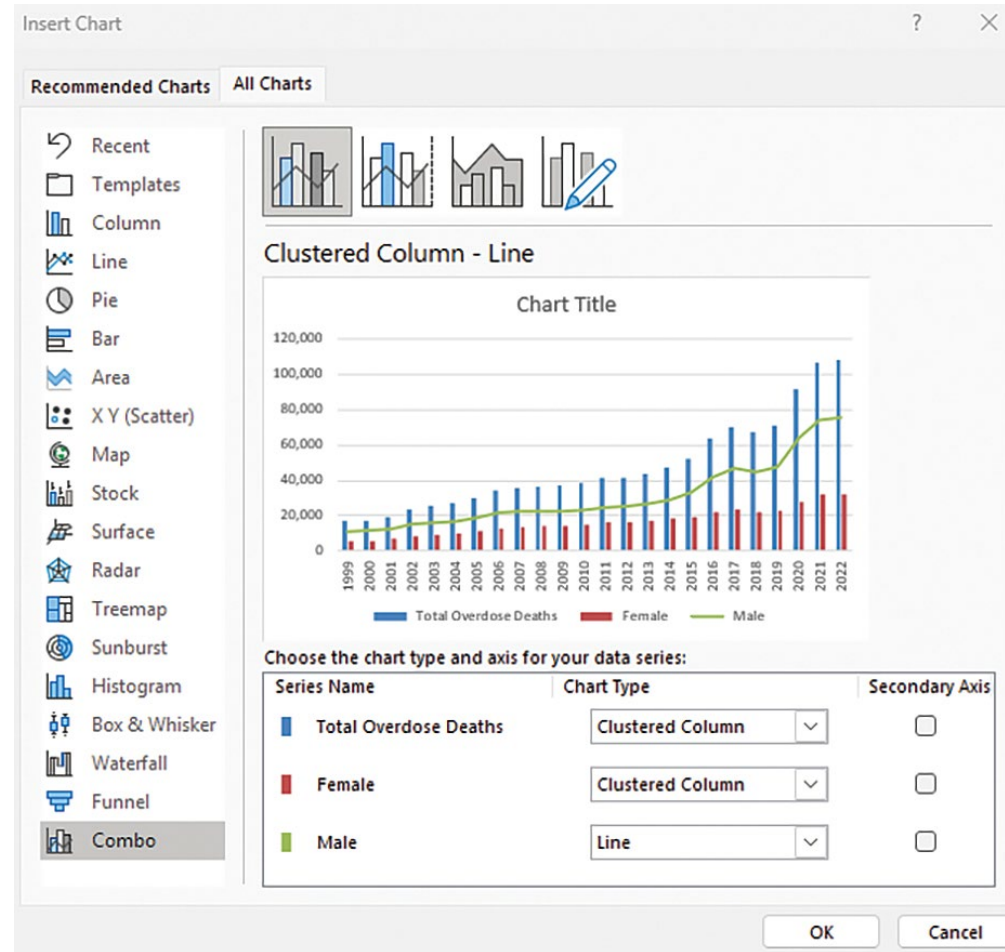


Figure 7.11 The All Charts menu with Combo chart option selected.

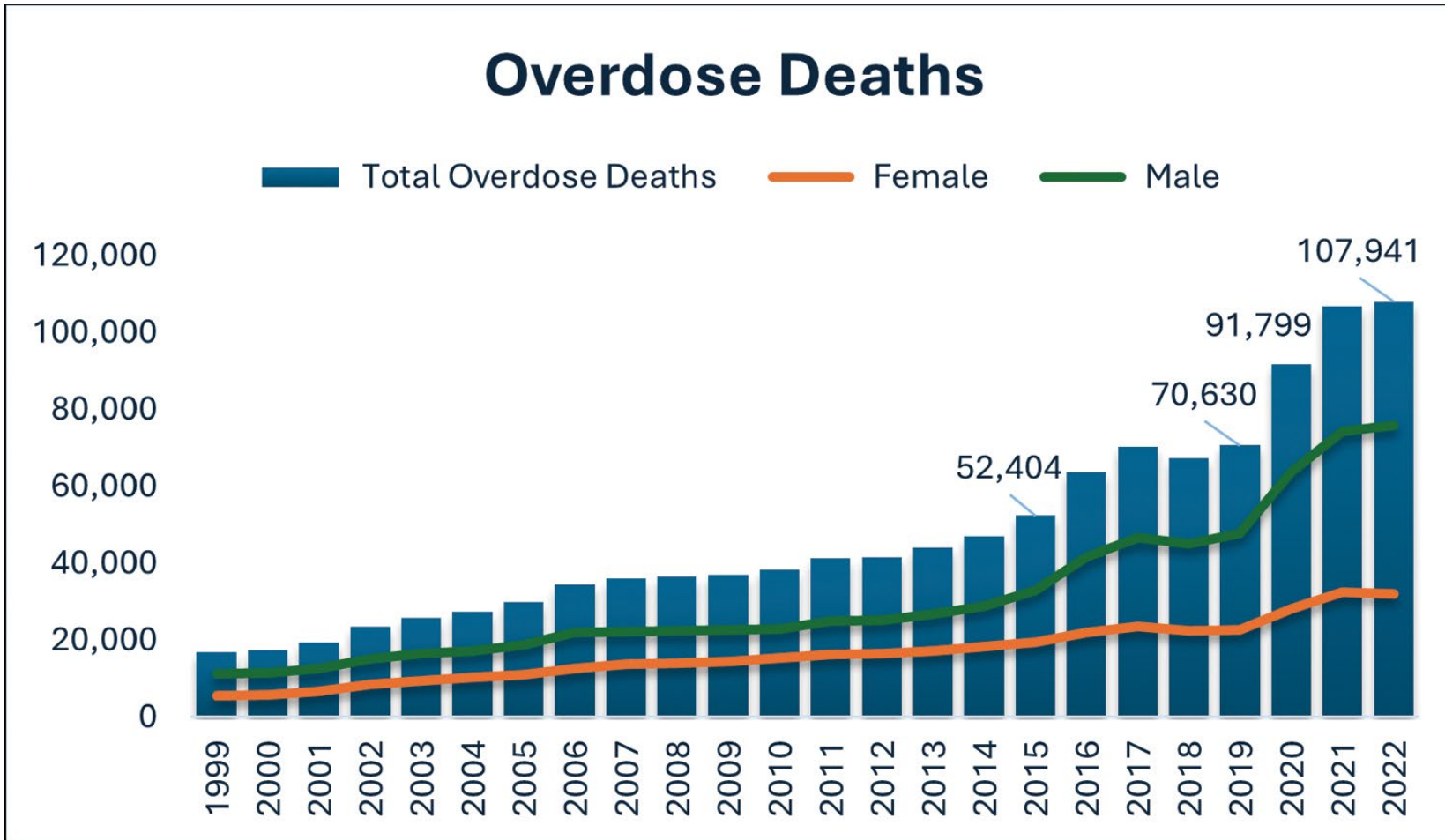
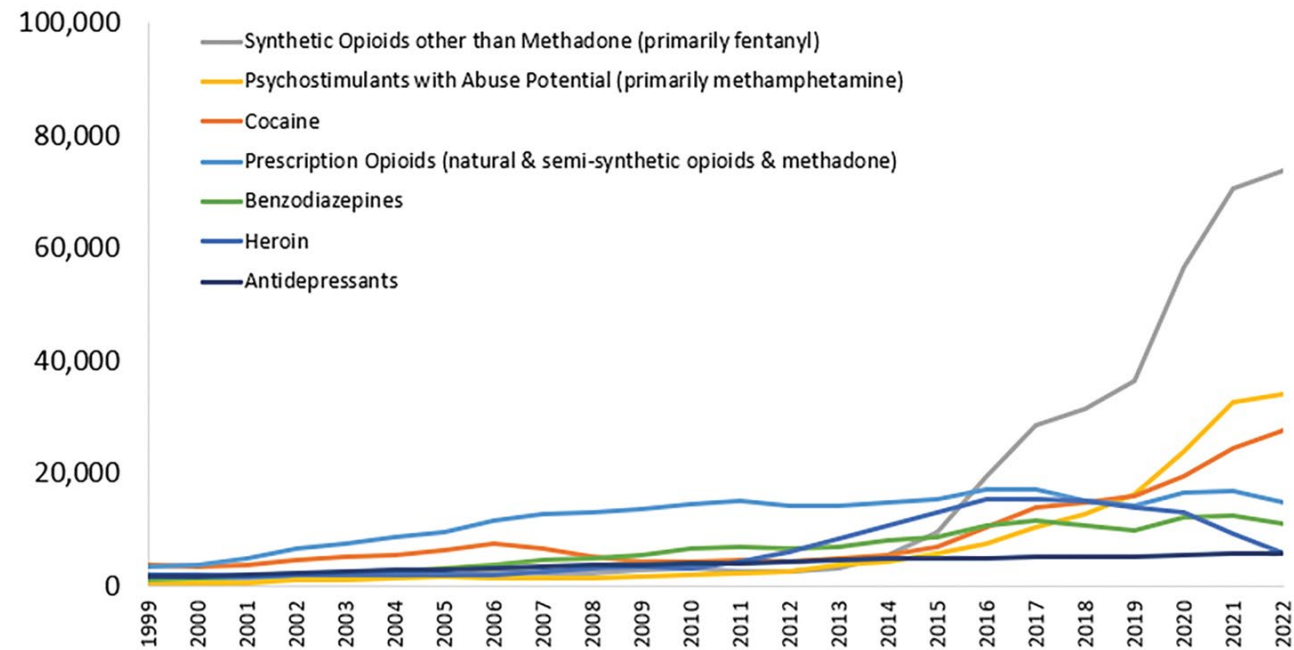


Figure 7.12 Total U.S. overdose deaths by sex, 1999–2022, in Excel.

Figure 2. U.S. Overdose Deaths*, Select Drugs or Drug Categories, 1999-2022



*Includes deaths with underlying causes of unintentional drug poisoning (X40–X44), suicide drug poisoning (X60–X64), homicide drug poisoning (X85), or drug poisoning of undetermined intent (Y10–Y14), as coded in the International Classification of Diseases, 10th Revision. Source: Centers for Disease Control and Prevention, National Center for Health Statistics. Multiple Cause of Death 1999–2022 on CDC WONDER Online Database, released 4/2024.

Figure 7.13 U.S. overdose deaths by drug category, 1999–2023, from the NIDA website.

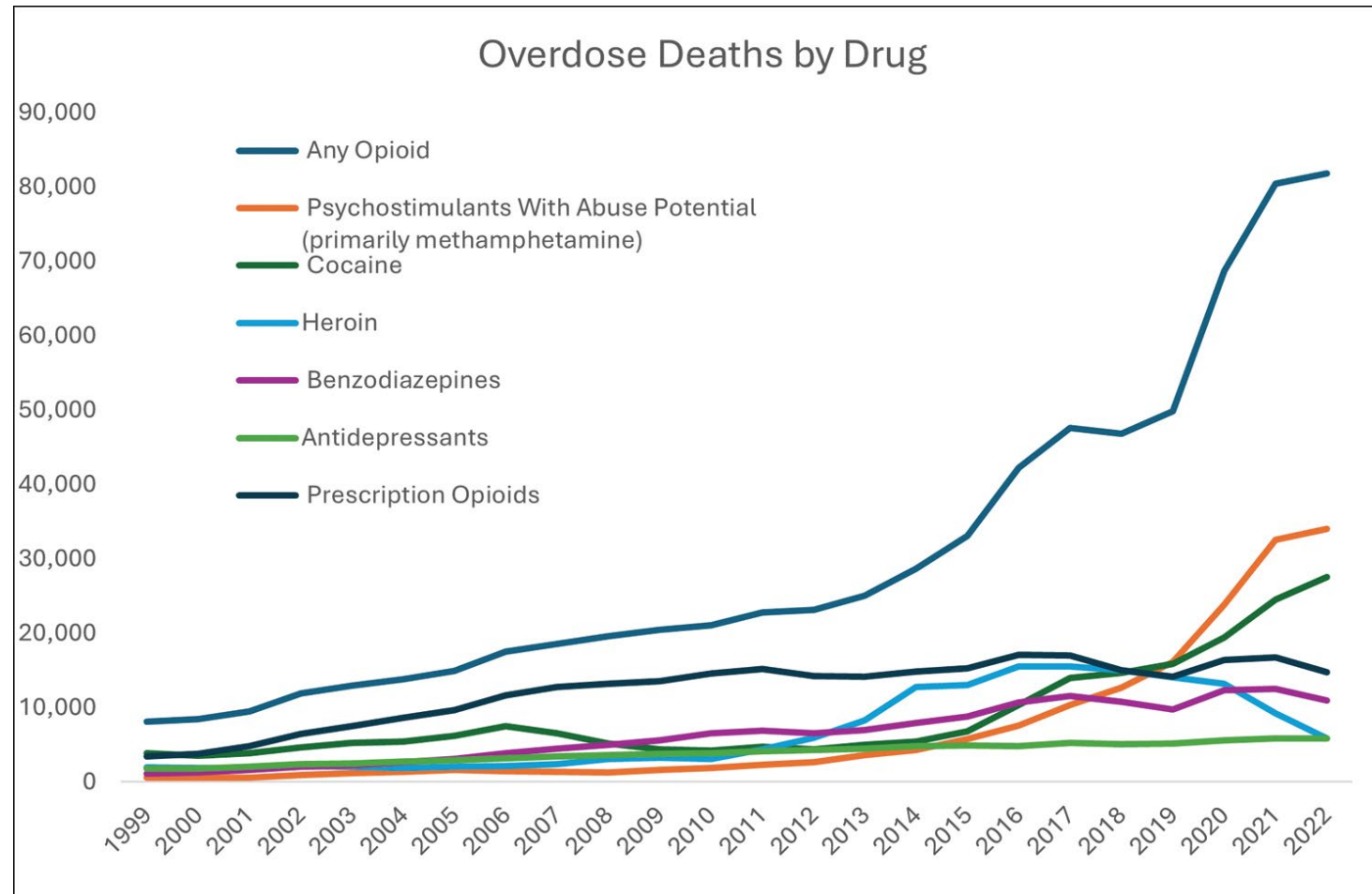


Figure 7.14 U.S. overdose deaths by drug category, 1999–2022, in Excel.

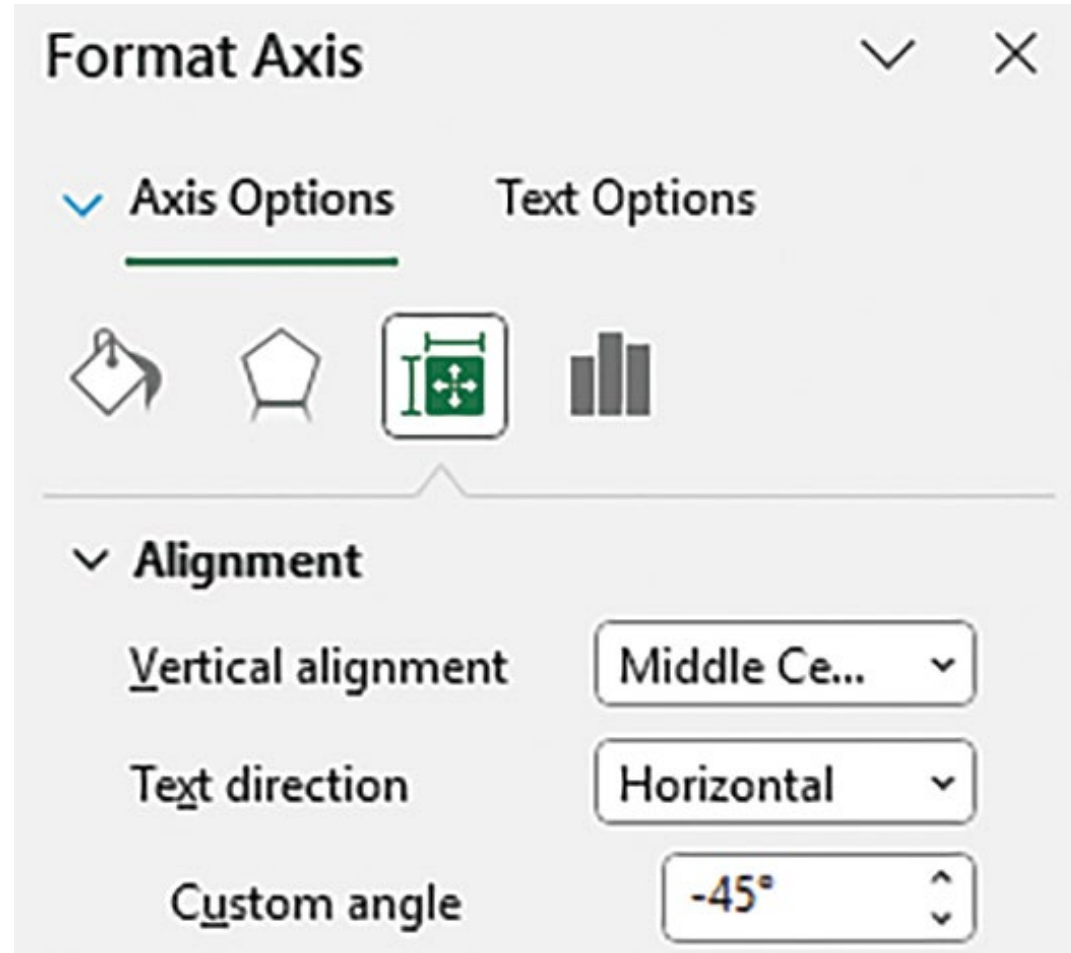


Figure 7.15 Format Axis options for angling axis labels.

C3

⌵

⋮

✕ ✓ *fx* ⌵

</

Figure 7.16 Overdose death rates per 100,000 and annual percent change.

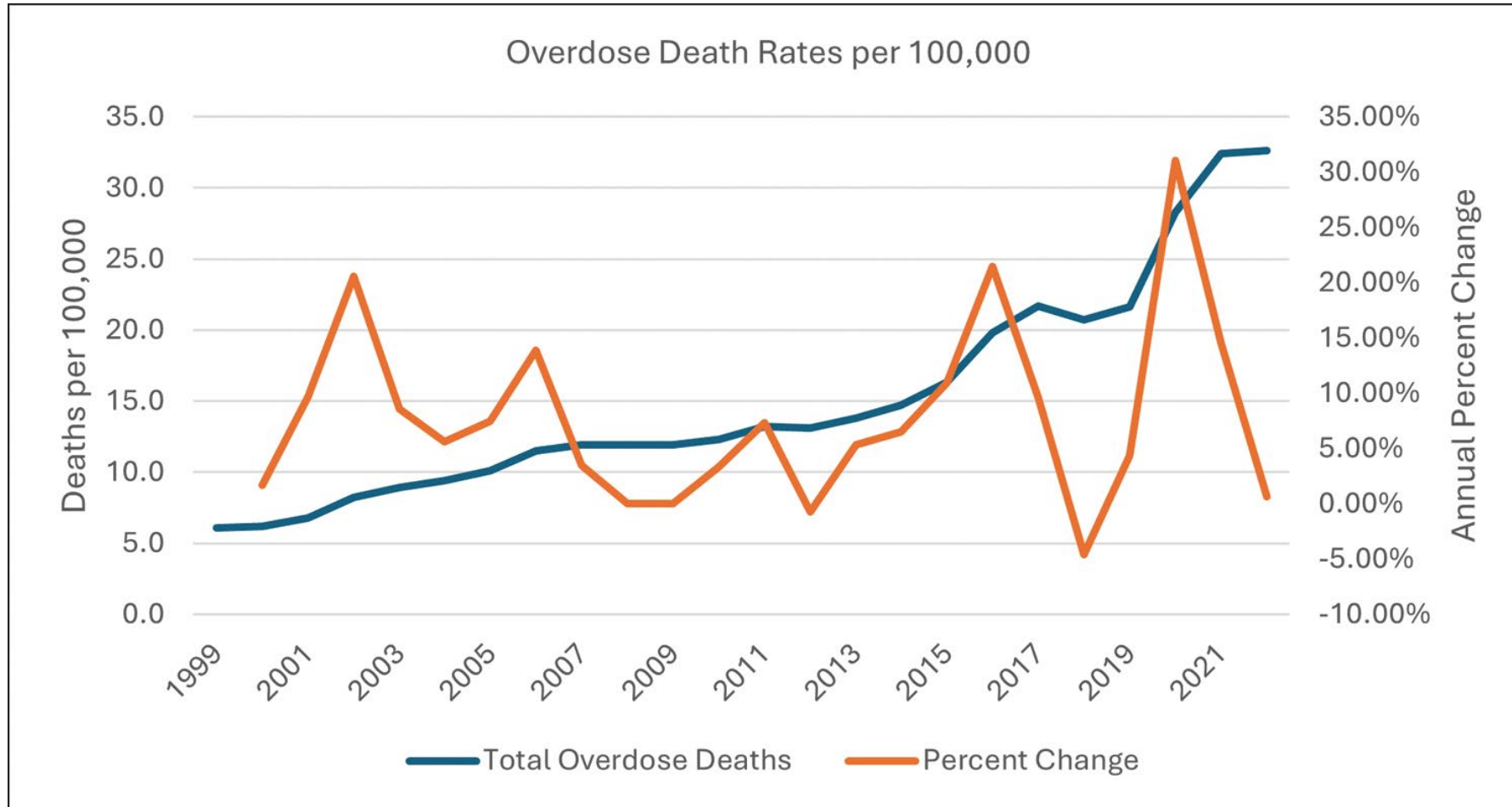


Figure 7.17 Total U.S. overdose death rates, 1999–2022, with annual percent change.

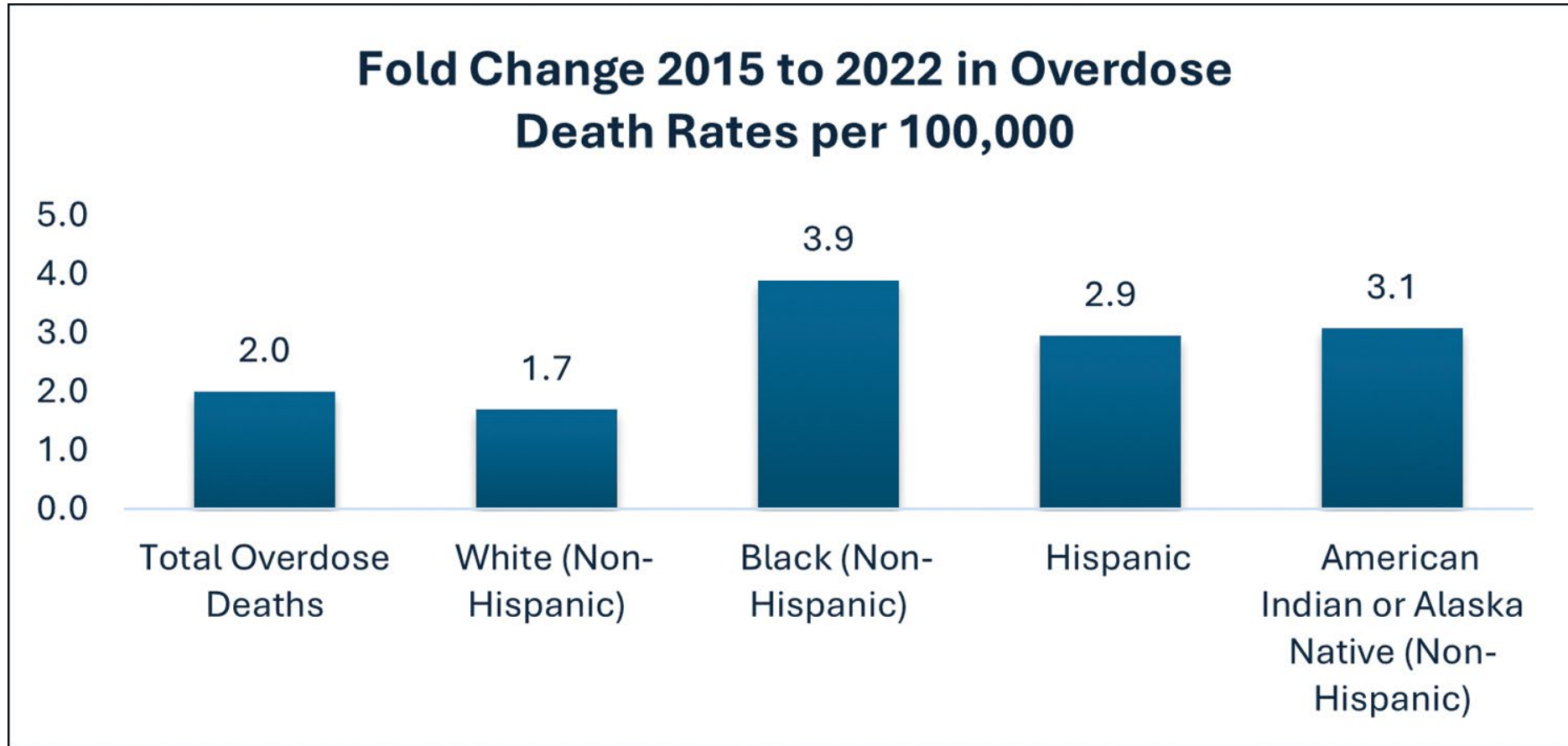


Figure 7.18 Change (as a multiple) in U.S. overdose death rates, 2015–2022, by demographic group.

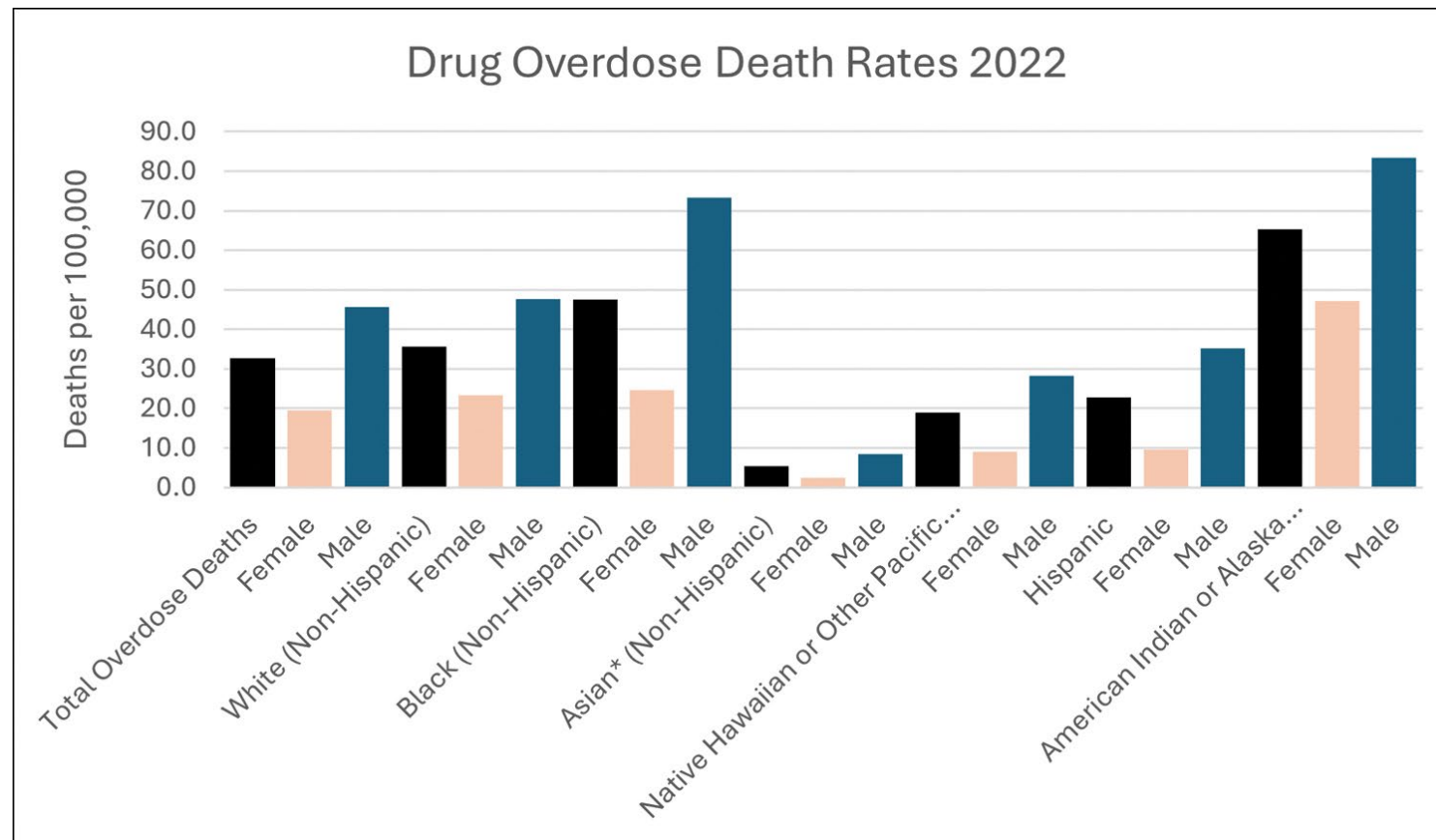


Figure 7.19 U.S. overdose death rates, 2022, by demographic group and sex.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	
1		 National Institute on Drug Abuse																										
2		Number of National Drug Overdose Deaths* Involving Select Prescription and Illicit Drugs																										
3		Source: National Center on Health Statistics, CDC WONDER																										
4																												
5																												
6																												
7			1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2015-2022 Fold Change	
8		Total Overdose Deaths	16,849	17,415	19,394	23,518	25,785	27,424	29,813	34,425	36,010	36,450	37,004	38,329	41,340	41,502	43,982	47,055	52,404	63,632	70,237	67,367	70,630	91,799	106,699	107,941	2.1	
9		Female	5,591	5,852	6,736	8,490	9,386	10,304	11,089	12,532	13,712	13,982	14,411	15,323	16,352	16,390	17,183	18,243	19,447	22,074	23,685	22,426	22,749	28,071	32,398	32,127	1.7	
10		Male	11,258	11,563	12,658	15,028	16,399	17,120	18,724	21,893	22,298	22,468	22,593	23,006	24,988	25,112	26,799	28,812	32,957	41,558	46,552	44,941	47,881	63,728	74,301	75,814	2.3	
11		Any Opioid¹	8,050	8,407	9,496	11,920	12,940	13,756	14,918	17,545	18,516	19,582	20,422	21,089	22,784	23,166	25,052	28,647	33,091	42,249	47,600	46,802	49,860	68,630	80,411	81,806	2.5	
12		Female	2,057	2,264	2,767	3,760	4,138	4,643	5,161	5,945	6,581	6,819	7,287	7,734	8,325	8,432	9,055	10,227	11,420	13,751	15,263	14,724	15,225	19,970	23,654	23,421	2.1	
13		Male	5,993	6,143	6,729	8,160	8,802	9,113	9,757	11,600	11,935	12,763	13,135	13,355	14,459	14,734	15,997	18,420	21,671	28,498	32,337	32,078	34,635	48,660	56,757	58,385	2.7	
14		Select Opioids^{1,2} AND Synthetic Opioids other than Methadone	167	205	225	358	383	428	485	781	642	703	923	1,002	961	944	1,219	2,493	4,806	9,299	12,556	13,491	13,596	16,465	15,444	20,413	4.2	
15		Prescription Opioids²	3,442	3,785	4,770	6,483	7,461	8,577	9,612	11,589	12,796	13,149	13,523	14,583	15,140	14,240	14,145	14,838	15,281	17,087	17,029	14,975	14,139	16,416	16,706	14,716	1.0	
16		Female	1,022	1,236	1,608	2,304	2,681	3,144	3,572	4,274	4,863	4,959	5,212	5,644	6,082	5,995	6,049	6,506	6,664	7,109	7,156	6,252	5,755	6,441	6,623	5,713	0.9	
17		Male	2,420	2,549	3,162	4,179	4,780	5,433	6,040	7,315	7,933	8,190	8,311	8,939	9,058	8,245	8,096	8,332	8,617	9,978	9,873	8,723	8,384	9,975	10,083	9,003	1.0	
18		Prescription Opioids AND Synthetic Opioids other than Methadone	142	167	199	322	344	384	426	573	601	655	872	939	889	861	1,015	1,489	2,263	4,055	5,444	5,417	5,876	8,626	9,644	8,310	3.7	
19		Female	65	76	86	157	151	184	207	246	286	309	444	453	426	445	488	661	898	1,394	1,859	1,872	1,949	2,798	6,393	2,635	2.9	
20		Male	77	91	113	165	193	200	219	327	315	346	428	486	463	416	527	828	1,365	2,661	3,585	3,545	3,927	5,828	3,251	5,675	4.2	
21		Prescription Opioids WITHOUT Synthetic Opioids other than Methadone	3,300	3,618	4,571	6,161	7,117	8,193	9,186	11,016	12,195	12,494	12,651	13,644	14,251	13,379	13,130	13,349	13,018	13,032	11,585	9,558	8,263	7,790	7,062	6,406	0.5	
22		Female	957	1,160	1,522	2,147	2,530	2,960	3,365	4,028	4,577	4,650	4,768	5,191	5,656	5,550	5,561	5,845	5,766	5,715	5,297	4,380	3,806	3,643	230	3,078	0.5	
23		Male	2,343	2,458	3,049	4,014	4,587	5,233	5,821	6,988	7,618	7,844	7,883	8,453	8,595	7,829	7,569	7,504	7,252	7,317	6,288	5,178	4,457	4,147	6,832	3,328	0.5	
24		Synthetic Opioids other than Methadone (primarily fentanyl)³	730	782	957	1,295	1,400	1,664	1,742	2,707	2,213	2,306	2,946	3,007	2,666	2,628	3,105	5,544	9,580	19,413	28,466	31,335	36,359	56,516	70,601	73,838	7.7	
25		Female	330	374	447	614	643	798	823	1,030	1,053	1,083	1,445	1,440	1,247	1,195	1,431	2,079	3,020	5,578	7,942	8,807	10,076	15,250	19,571	19,880	6.6	
26		Male	400	408	510	681	757	866	919	1,677	1,160	1,223	1,501	1,567	1,419	1,433	1,674	3,465	6,560	13,835	20,524	22,528	26,283	41,266	51,030	53,958	8.2	
27		Heroin⁴	1,960	1,842	1,779	2,089	2,080	1,878	2,009	2,088	2,399	3,041	3,278	3,036	4,397	5,925	8,257	10,574	12,989	15,469	15,482	14,996	14,019	13,165	9,173	5,871	0.5	
28		Female	306	279	313	359	358	341	389	344	399	551	577	584	878	1,213	1,732	2,414	3,108	3,717	3,886	3,705	3,520	3,284	2,372	1,504	0.5	
29		Male	1,654	1,563	1,466	1,730	1,722	1,537	1,620	1,744	2,000	2,490	2,701	2,452	3,519	4,712	6,525	8,160	9,881	11,752	11,596	11,291	10,499	9,881	6,801	4,367	0.4	
30		Heroin AND Synthetic Opioids other than Methadone	15	18	15	15	16	13	34	113	13	28	29	45	44	69	209	1,027	2,685	5,781	8,091	9,068	8,746	8,990	6,783	4,675	1.7	
31		Female	4	7	4	5	3	6	9	25	3	13	10	8	11	19	58	275	670	1,430	2,035	2,267	2,256	2,294	1,791	1,186	1.8	
32		Male	11	11	11	10	13	7	25	88	10	15	19	37	33	50	151	752	2,015	4,351	6,056	6,801	6,490	6,696	4,992	3,489	1.7	
33		Heroin WITHOUT Synthetic Opioids other than Methadone	1,945	1,824	1,764	2,074	2,064	1,865	1,975	1,975	2,386	3,013	3,249	2,991	4,353	5,856	8,048	9,547	10,304	9,688	7,391	5,928	5,273	4,175	2,390	1,196	0.1	
34		Female	302	272	309	354	355	335	380	319	396	538	567	576	867	1,194	1,674	2,139	2,438	2,287	1,851	1,438	1,264	990	581	318	0.1	
35		Male	1,643	1,552	1,455	1,720	1,709	1,530	1,595	1,656	1,990	2,475	2,682	2,415	3,486	4,662	6,374	7,408	7,866	7,401	5,540	4,490	4,009	3,185	1,809	878	0.1	
36		Stimulants^{5,6}	4,271	4,017	4,308	5,423	6,215	6,591	7,606	8,668	7,697	6,320	5,824	5,914	6,765	6,879	8,338	9,935	12,122	17,258	23,139	25,877	30,231	40,643	53,495	57,497	4.7	
37		Female	980	980	1,083	1,400	1,626	1,767	2,001	2,214	2,028	1,667	1,586	1,683	1,955	2,026	2,412	2,720	3,527	4,895	6,645	7,529	8,532	11,338	15,087	15,763	4.5	
38		Male	3,291	3,037	3,225	4,023	4,589	4,824	5,605	6,454	5,669	4,653	4,238	4,231	4,810	4,853	5,926	6,675	8,595	12,363	16,494	18,348	21,699	29,305	38,408	41,743	4.9	
39		Stimulants AND Any Opioid	2,101	1,977	1,996	2,528	2,732	2,850	3,215	3,764	3,394	3,085	2,766	2,662	3,255	3,340	4,037	4,999	6,594	10,222	14,455	16,165	19,192	27,966	37,682	41,361	6.3	

Figure 7.20 Overdose death data in its original form (as provided by NIDA).

	A	B	C	D	E	F	G	H	I	J
1	Category	Gender	1999	2000	2001	2002	2003	2004	2005	2006
2	Total Overdose Deaths	All	16,849	17,415	19,394	23,518	25,785	27,424	29,813	34,425
3	Total Overdose Deaths	Female	5,591	5,852	6,736	8,490	9,386	10,304	11,089	12,532
4	Total Overdose Deaths	Male	11,258	11,563	12,658	15,028	16,399	17,120	18,724	21,893
5	Any Opioid	All	8,050	8,407	9,496	11,920	12,940	13,756	14,918	17,545
6	Any Opioid	Female	2,057	2,264	2,767	3,760	4,138	4,643	5,161	5,945
7	Any Opioid	Male	5,993	6,143	6,729	8,160	8,802	9,113	9,757	11,600
8	Prescription Opioids	All	3,442	3,785	4,770	6,483	7,461	8,577	9,612	11,589
9	Prescription Opioids	Female	1,022	1,236	1,608	2,304	2,681	3,144	3,572	4,274
10	Prescription Opioids	Male	2,420	2,549	3,162	4,179	4,780	5,433	6,040	7,315
11	Synthetic Opioids other than Methadone (primarily fentanyl)	All	730	782	957	1,295	1,400	1,664	1,742	2,707
12	Synthetic Opioids other than Methadone (primarily fentanyl)	Female	330	374	447	614	643	798	823	1,030
13	Synthetic Opioids other than Methadone (primarily fentanyl)	Male	400	408	510	681	757	866	919	1,677
14	Heroin	All	1,960	1,842	1,779	2,089	2,080	1,878	2,009	2,088
15	Heroin	Female	306	279	313	359	358	341	389	344
16	Heroin	Male	1,654	1,563	1,466	1,730	1,722	1,537	1,620	1,744
17	Stimulants	All	4,271	4,017	4,308	5,423	6,215	6,591	7,606	8,668
18	Stimulants	Female	980	980	1,083	1,400	1,626	1,767	2,001	2,214
19	Stimulants	Male	3,291	3,037	3,225	4,023	4,589	4,824	5,605	6,454
20	Cocaine	All	3,822	3,544	3,833	4,599	5,199	5,443	6,208	7,448
21	Cocaine	Female	850	843	957	1,143	1,322	1,405	1,620	1,860
22	Cocaine	Male	2,972	2,701	2,876	3,456	3,877	4,038	4,588	5,588
23	Psychostimulants With Abuse Potential (primarily methamphetamine)	All	547	578	563	941	1,179	1,305	1,608	1,462
24	Psychostimulants With Abuse Potential (primarily methamphetamine)	Female	158	164	152	285	353	393	438	411
25	Psychostimulants With Abuse Potential (primarily methamphetamine)	Male	389	414	411	656	826	912	1,170	1,051
26	Benzodiazepines	All	1,135	1,298	1,594	2,022	2,248	2,627	3,084	3,835
27	Benzodiazepines	Female	420	480	614	763	885	1,079	1,209	1,472
28	Benzodiazepines	Male	715	818	980	1,259	1,363	1,548	1,875	2,363

Figure 7.21 Overdose death data reorganized in Excel for use in Tableau.

	A	B	C	D	E	F	G	H
1	Race	Gender	1999	2000	2001	2002	2003	2004
2	All	All	6.1	6.2	6.8	8.2	8.9	9.4
3	All	Female	3.9	4.1	4.6	5.8	6.4	6.9
4	All	Male	8.2	8.3	9.0	10.6	11.5	11.8
5	White	All	6.2	6.6	7.4	9.2	10.2	11.0
6	White	Female	4.3	4.5	5.3	6.8	7.5	8.3
7	White	Male	8.0	8.6	9.6	11.6	12.9	13.7
8	Black	All	7.5	7.3	7.6	8.2	8.2	8.3
9	Black	Female	4.0	4.2	4.4	5.1	5.4	5.6
10	Black	Male	11.5	10.9	11.2	11.7	11.6	11.4
11	Asian	All						
12	Asian	Female						
13	Asian	Male						
14	Native Hawaiian or Other Pacific Islander	All						
15	Native Hawaiian or Other Pacific Islander	Female						
16	Native Hawaiian or Other Pacific Islander	Male						
17	Hispanic		5.4	4.6	4.5	5.4	5.6	5.2
18	Hispanic	Female	2.2	2.0	2.2	2.7	2.9	2.9
19	Hispanic	Male	8.6	7.1	6.7	8.0	8.3	7.5
20	American Indian or Alaska Native	All	6.0	5.5	6.9	8.5	10.8	12.5
21	American Indian or Alaska Native	Female	5.2	4.3	6.5	7.1	9.4	10.4
22	American Indian or Alaska Native	Male	6.7	6.7	7.5	10.0	12.2	14.7

Figure 7.22 Overdose rate data reorganized in Excel for use in Tableau (partial view).

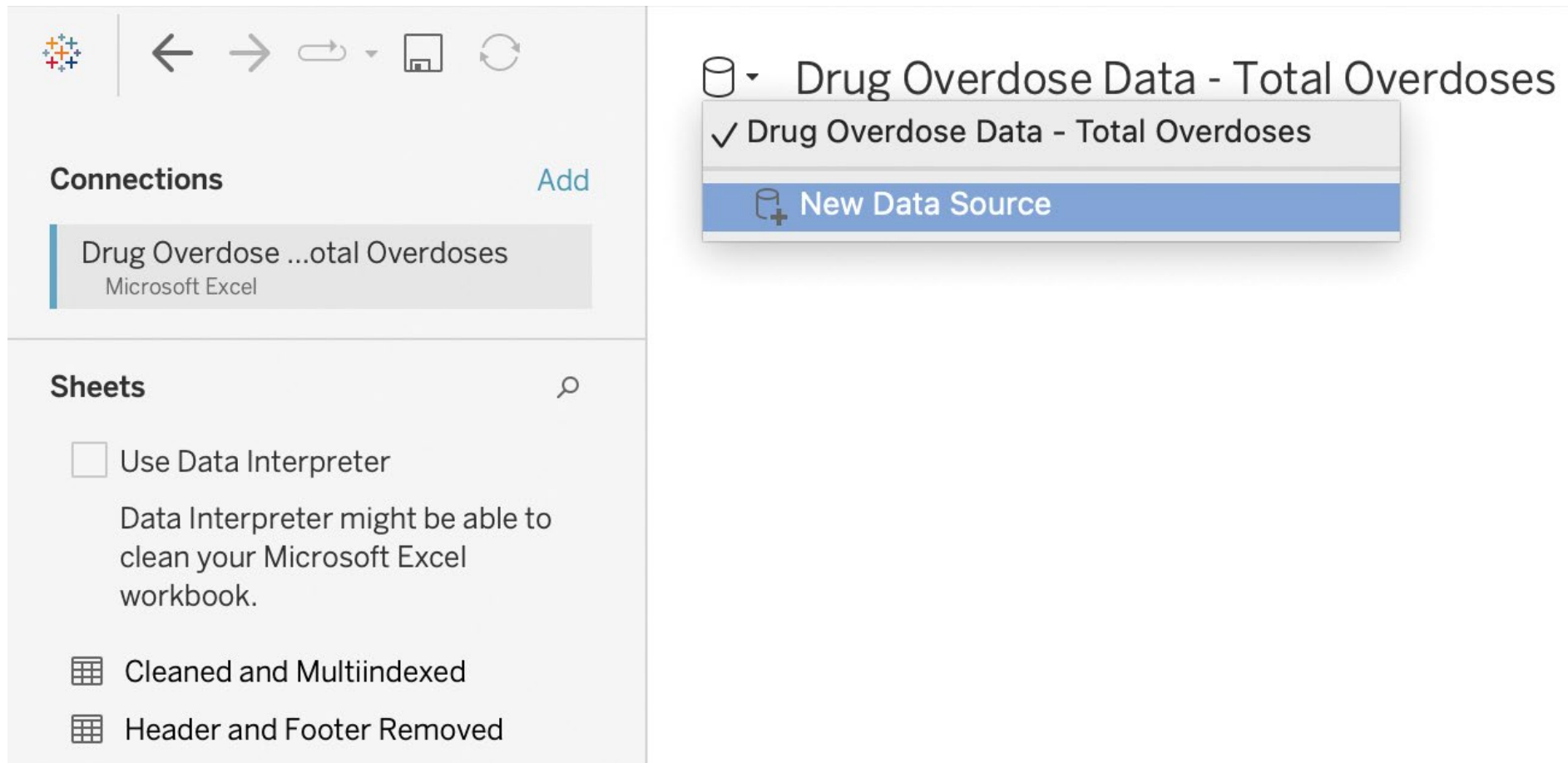


Figure 7.23 Connecting to a second data source.

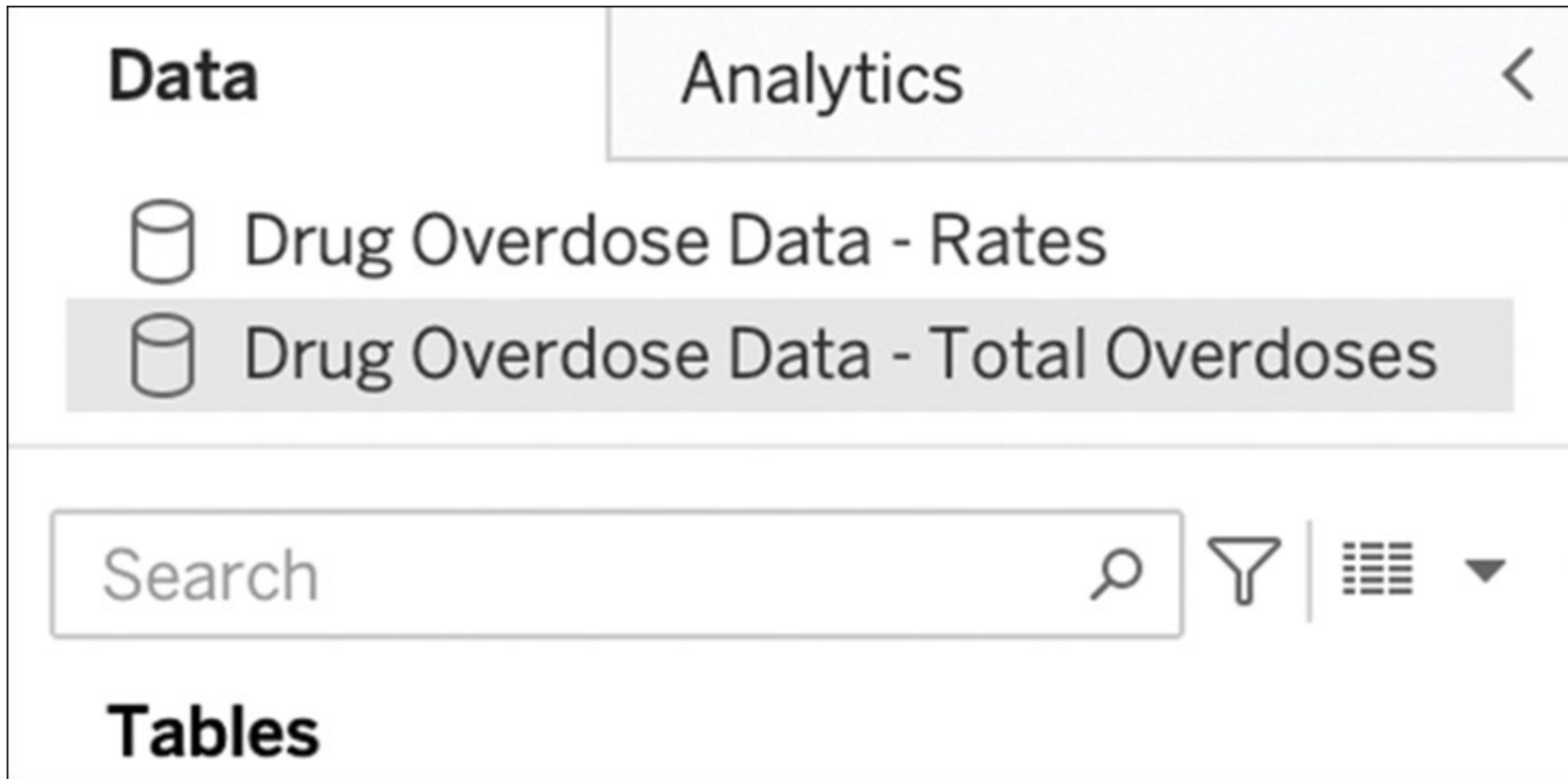


Figure 7.24 Selecting a data source in a sheet.

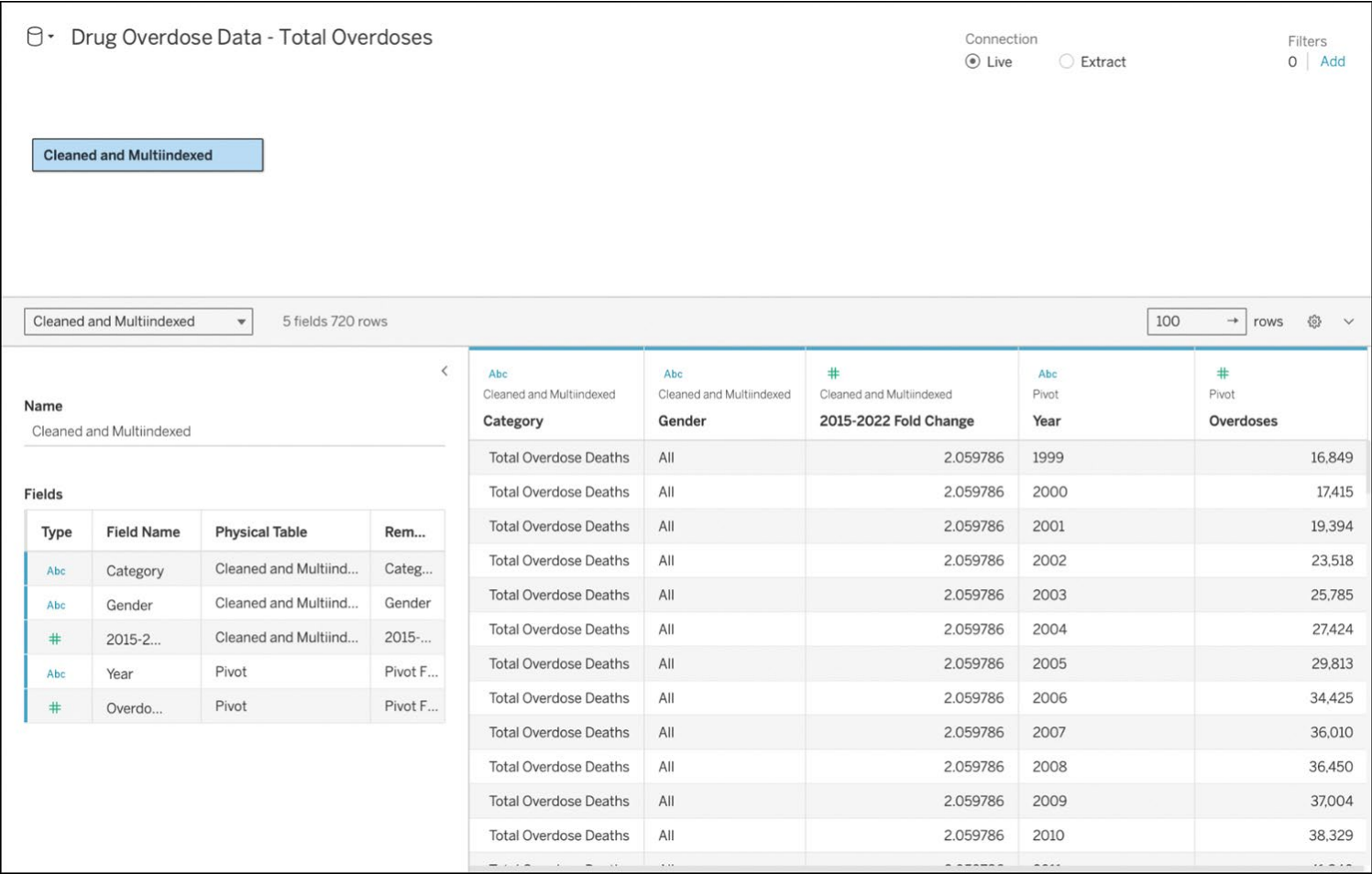


Figure 7.25 Pivoted overdose death data.

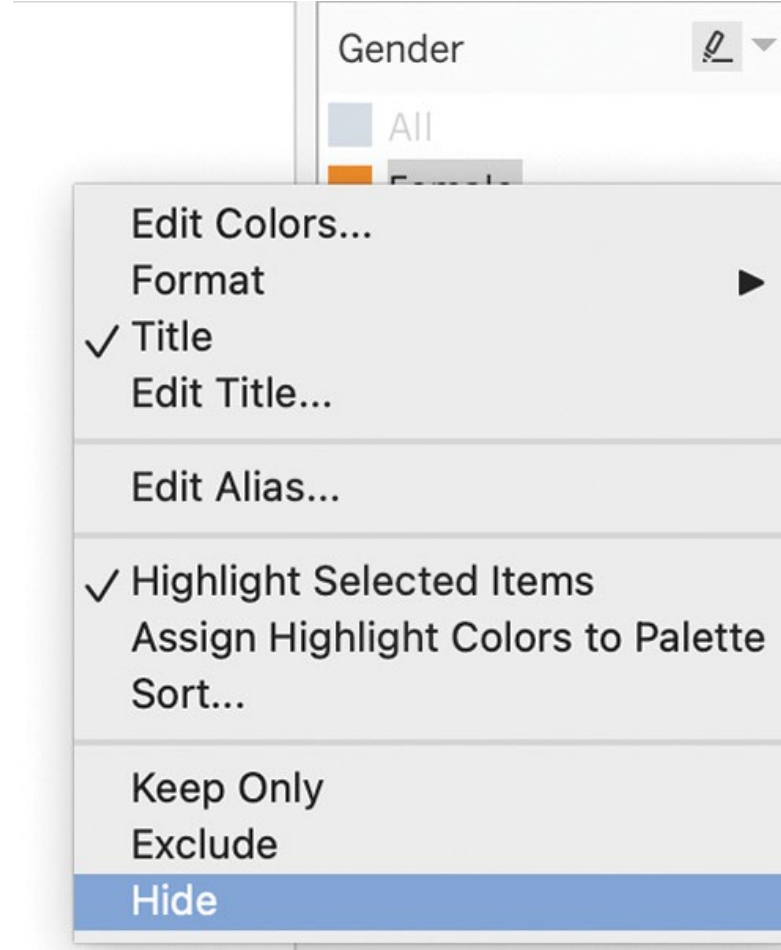
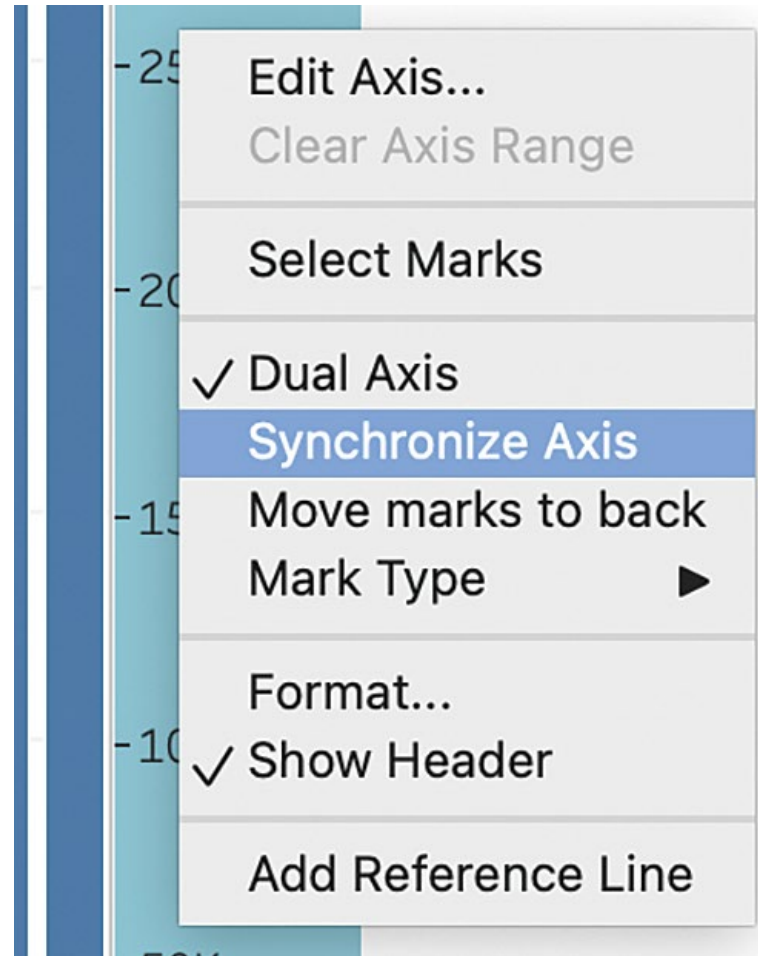


Figure 7.26 Hiding values from the color legend.



07fig27

Figure 7.27 Synchronizing dual axes.



Figure 7.28 Creating a duplicate of a dimension using a calculated field.

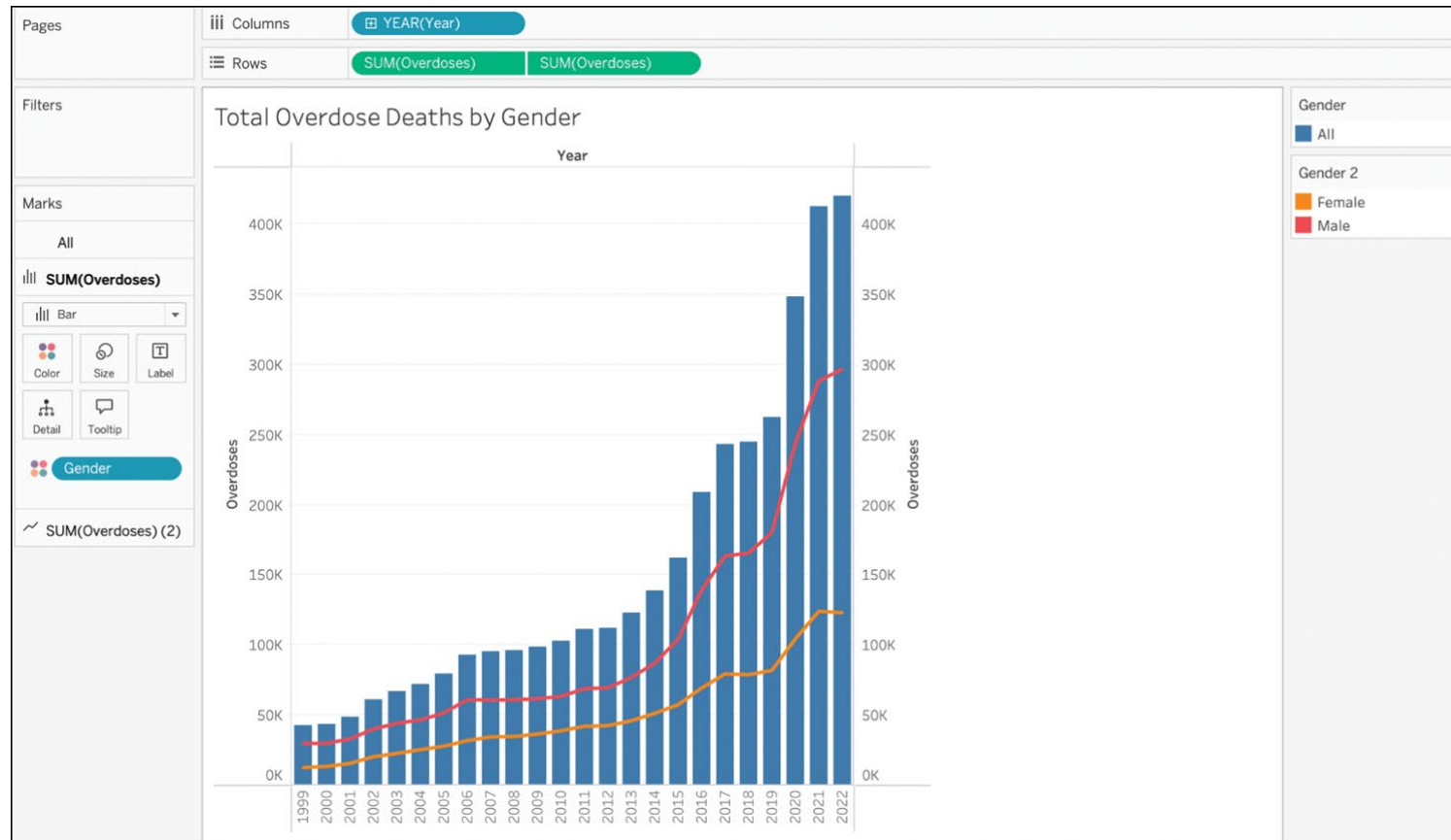


Figure 7.29 Incorrect combined chart.

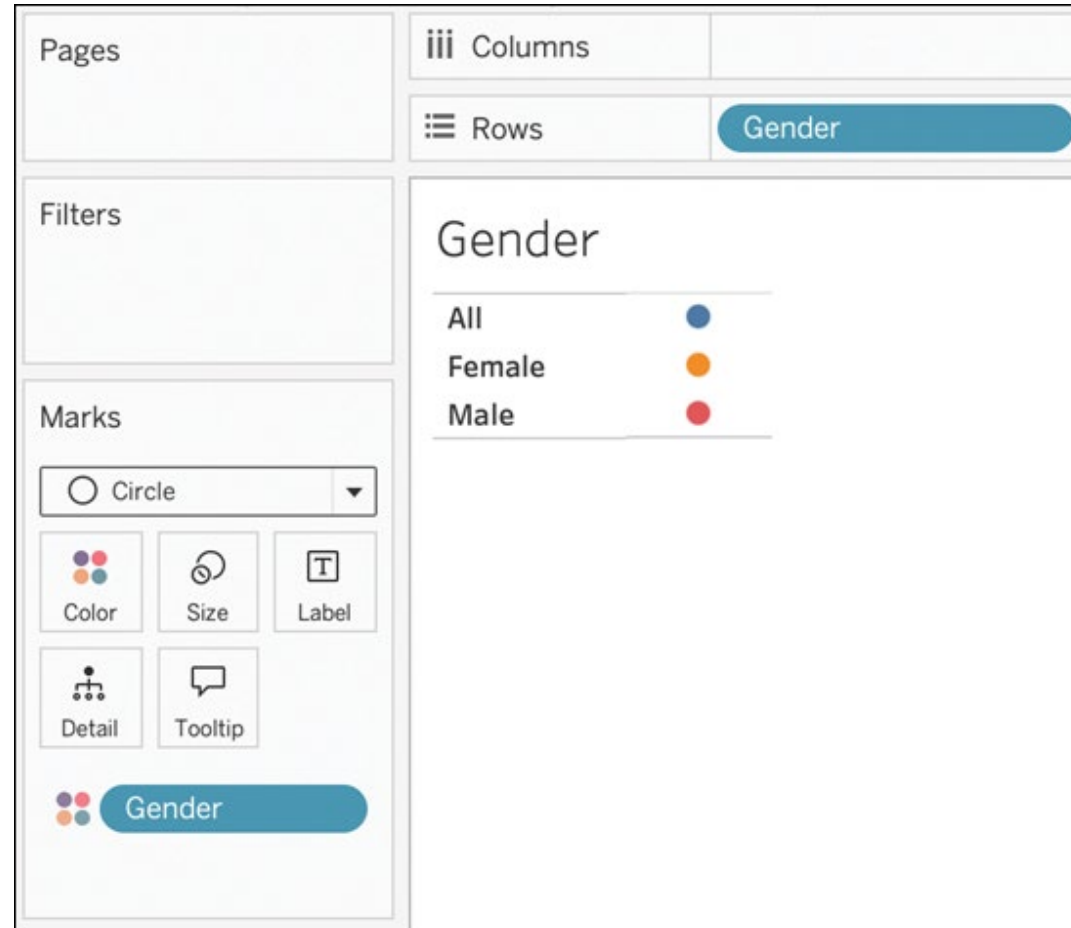


Figure 7.30 Worksheet designed to function as a color legend.

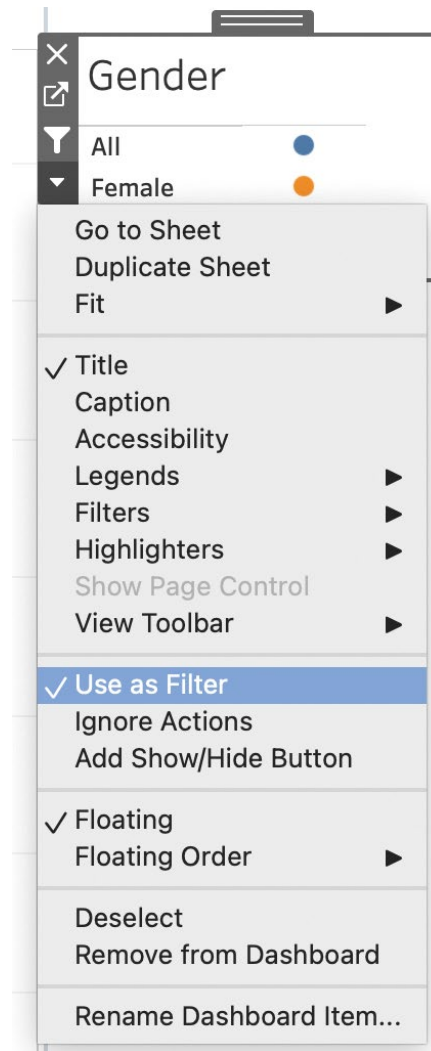


Figure 7.31 Using a sheet as a filter.

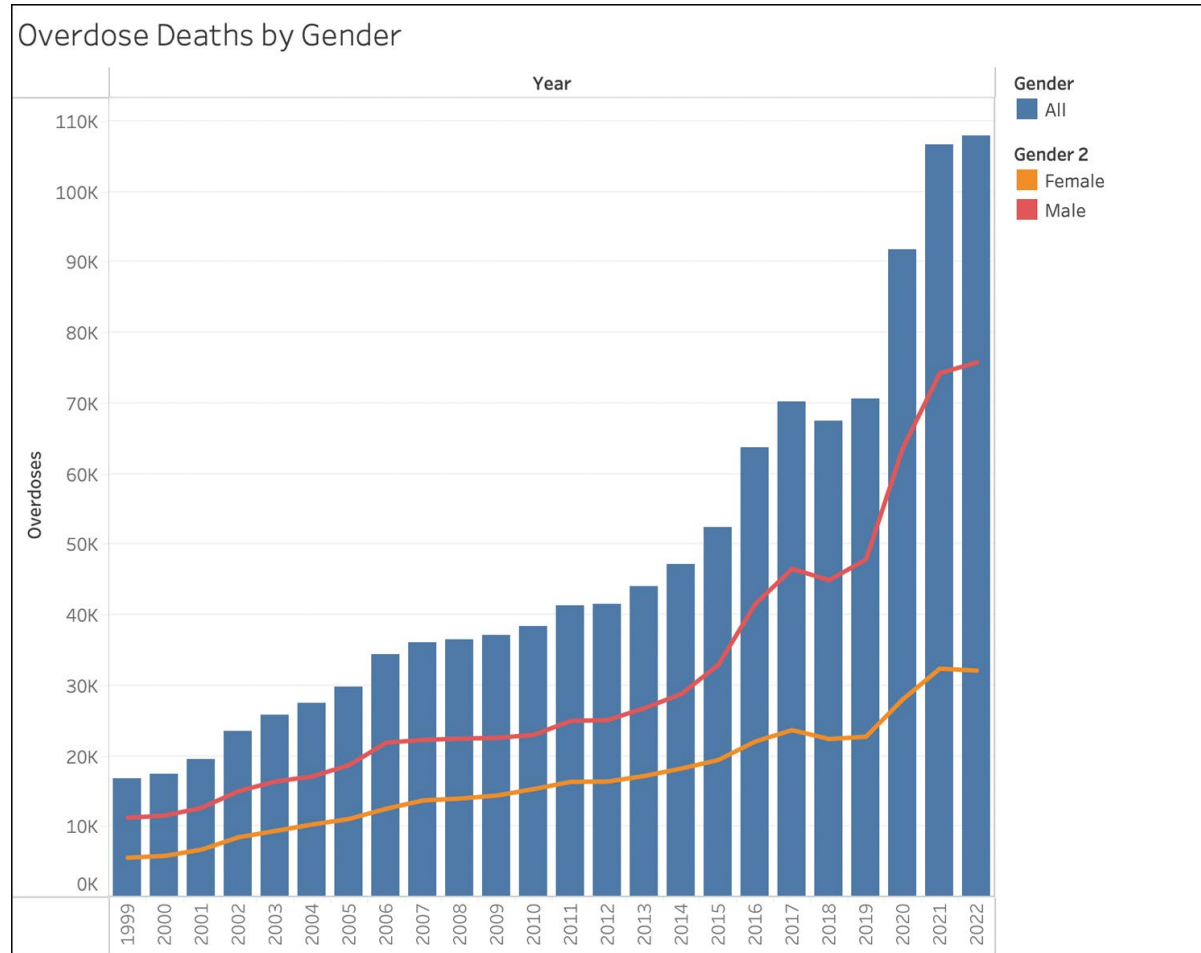


Figure 7.32 Combined bar/line chart showing overdose deaths by gender.

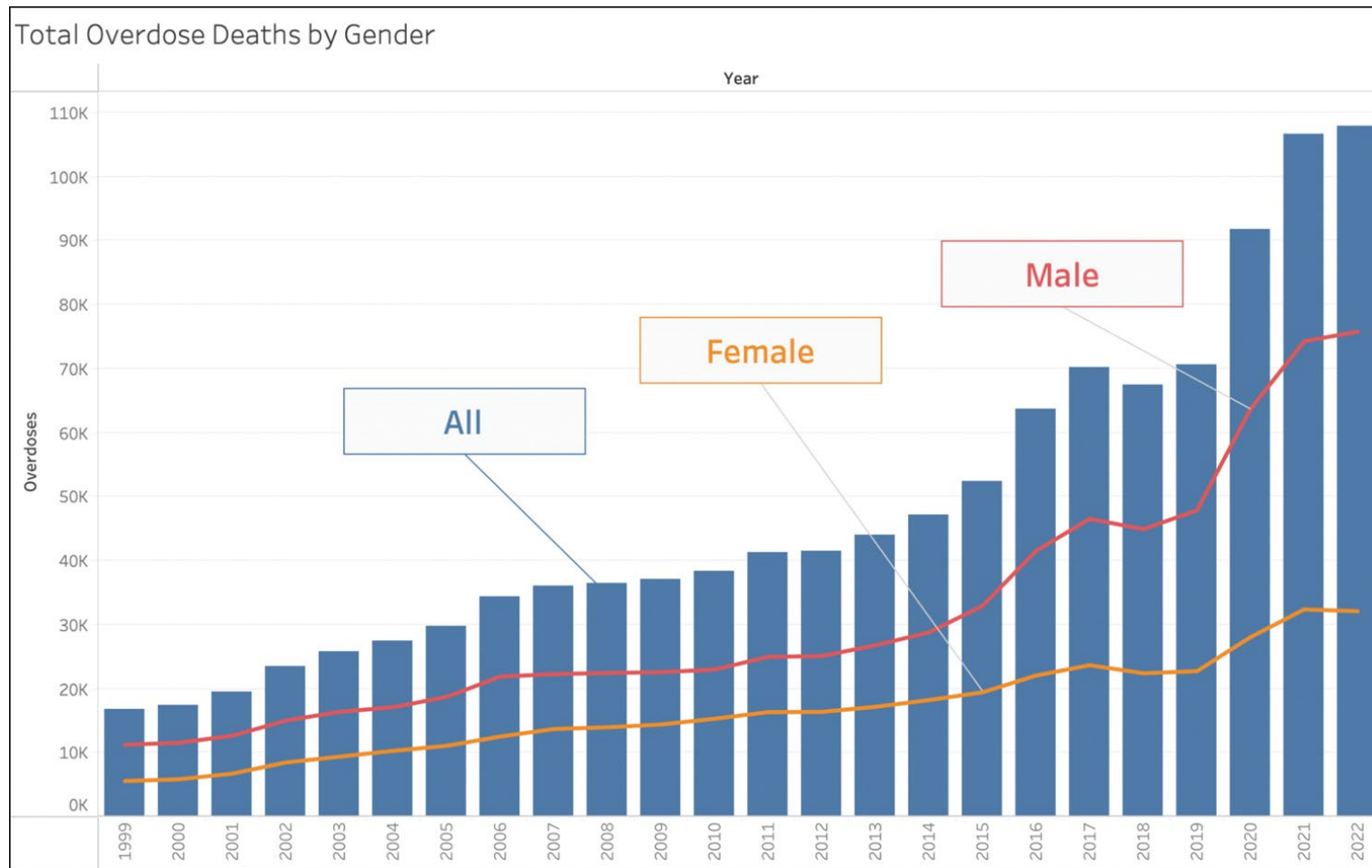


Figure 7.33 Combined bar/line chart showing overdose deaths by gender—annotated version.

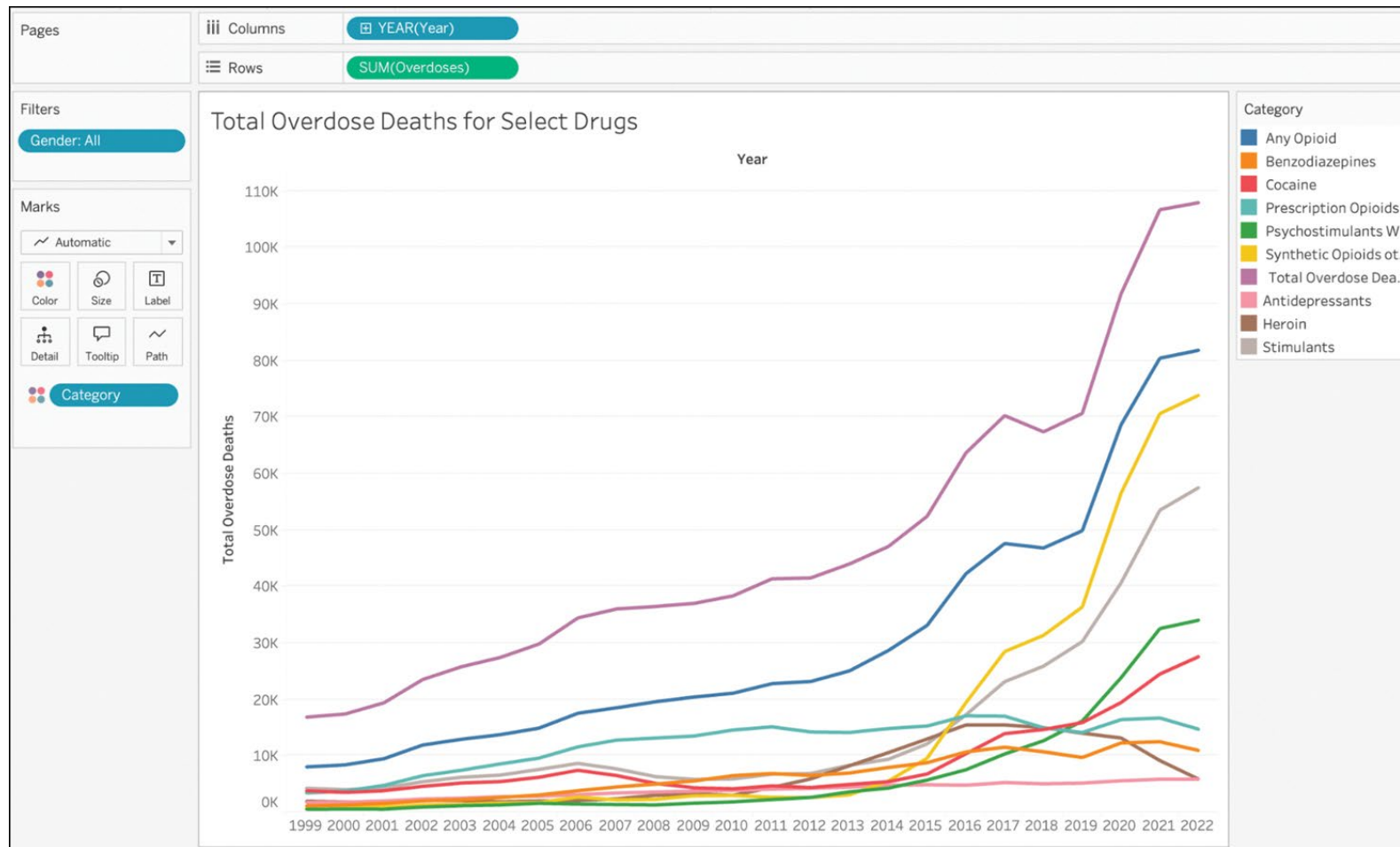


Figure 7.34 Overdose deaths for selected drugs.

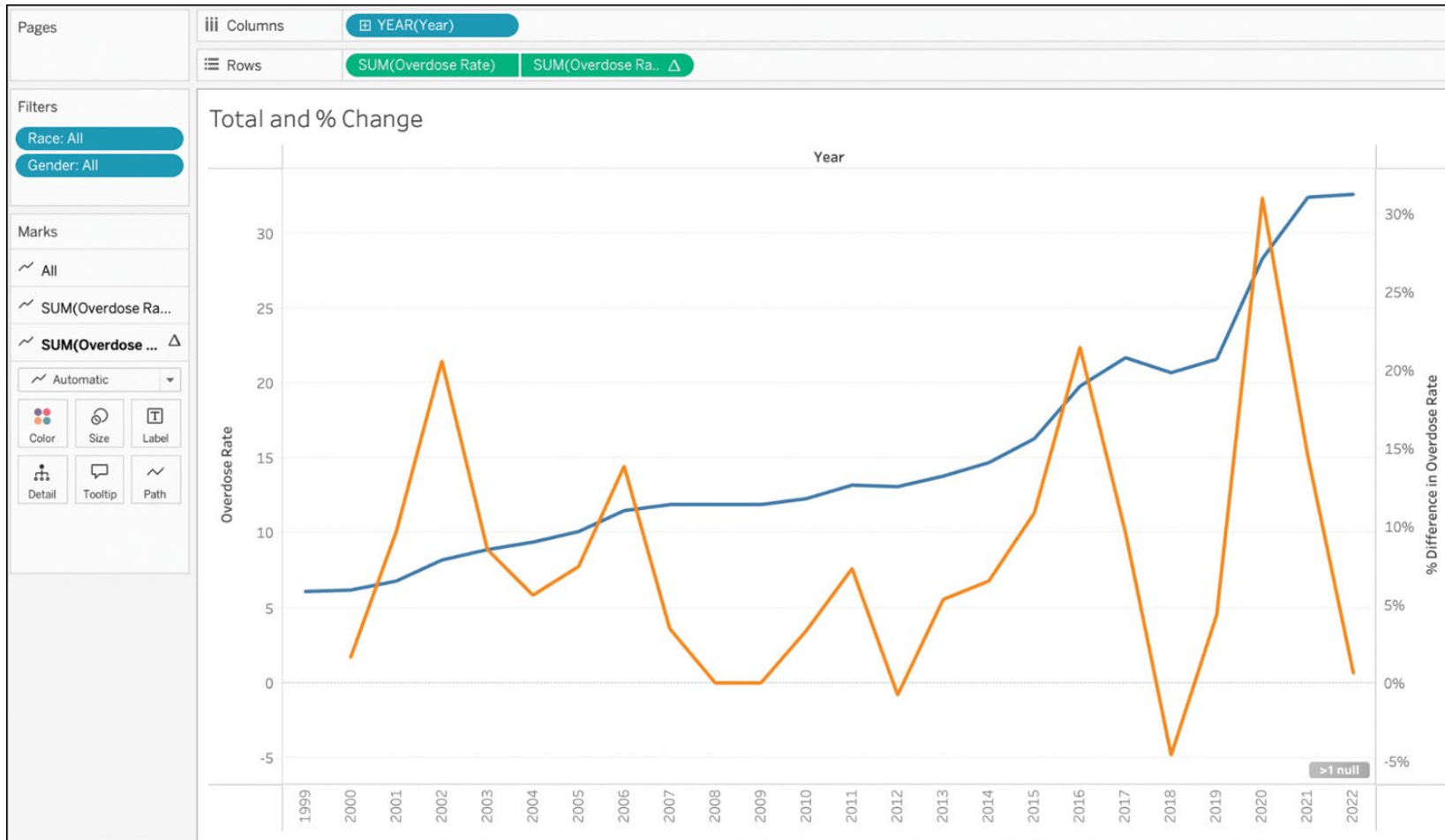


Figure 7.35 Dual-axis chart showing total overdose deaths and percent change from year to year.

Table Calculation

% Difference in Overdose Rate

Calculation Type

Percent Difference From

Compute Using

Table (across)

Cell

Specific Dimensions

☒ Year of Year

At the level

Relative to

Previous

☒ Show calculation assistance

Figure 7.36 Table calculation for percent change.

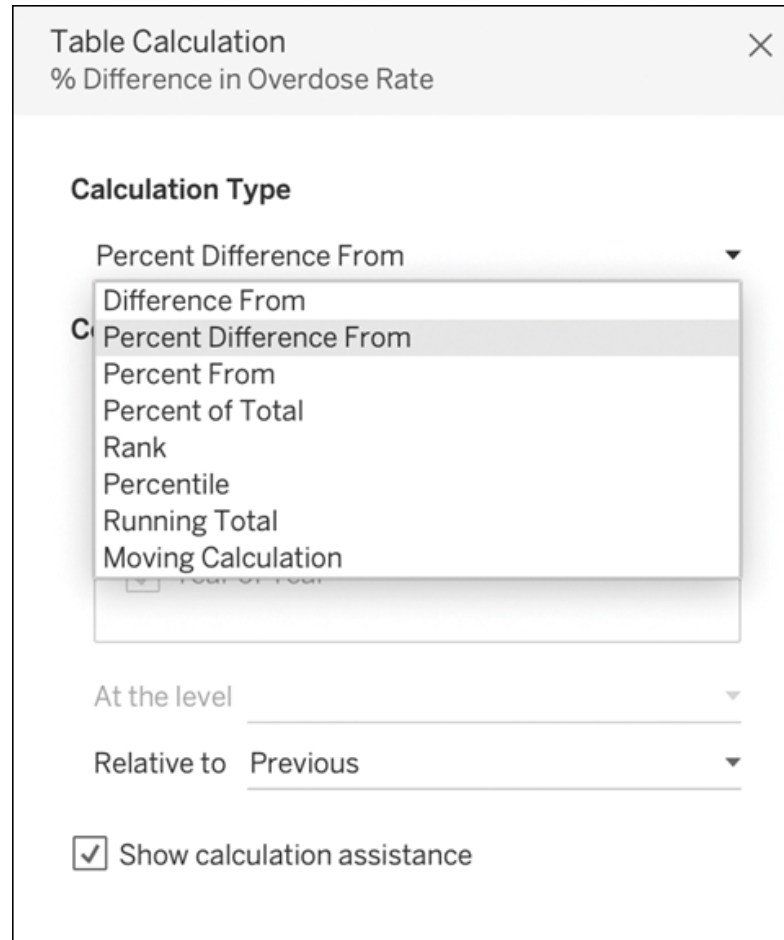


Figure 7.37 Table calculation menu options.

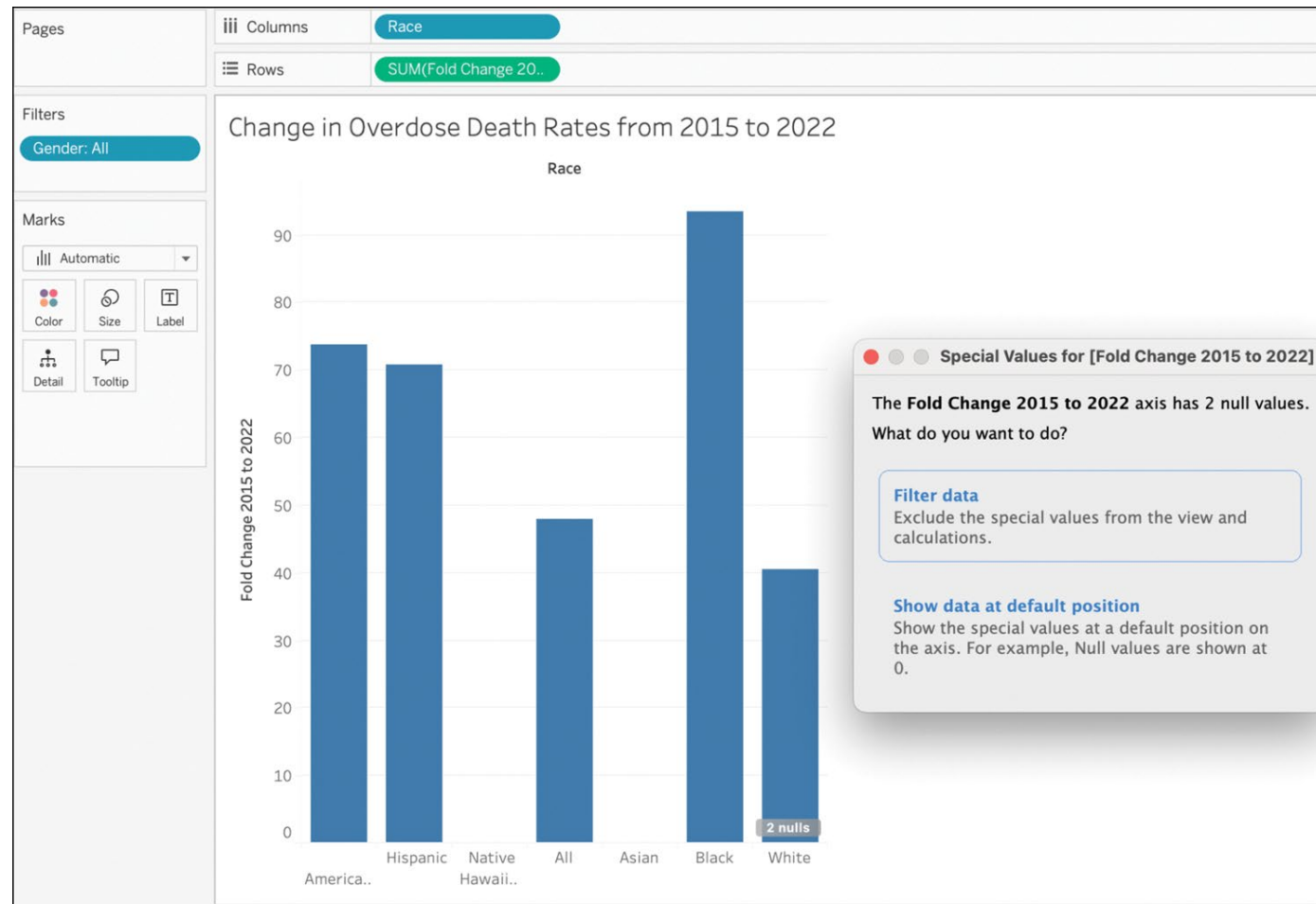


Figure 7.38 Filtering null columns.

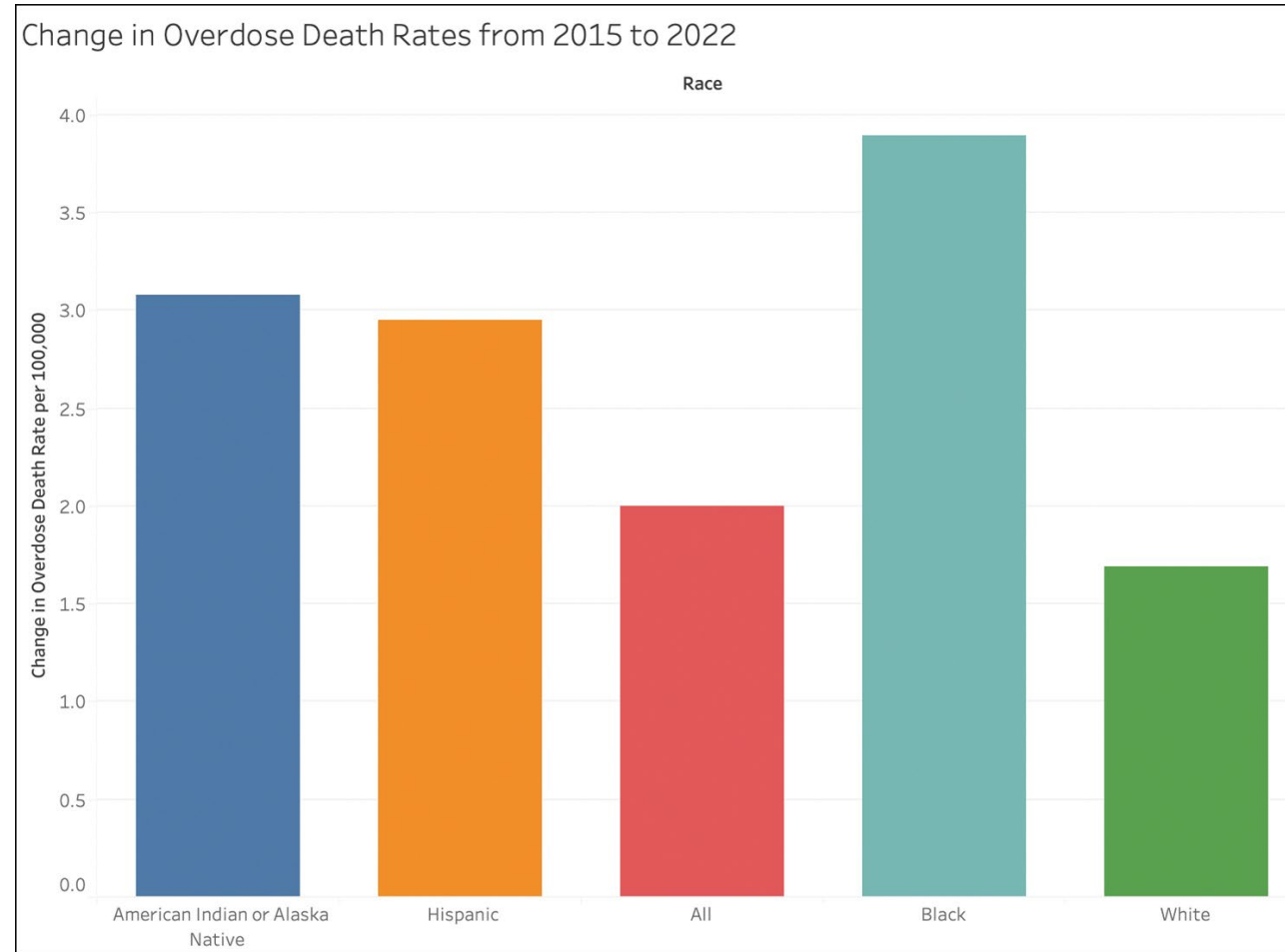


Figure 7.39 Bar chart showing the fold change in overdose death rates by race between 2015 and 2022.



Figure 7.40 Overdose death rates by race and gender.

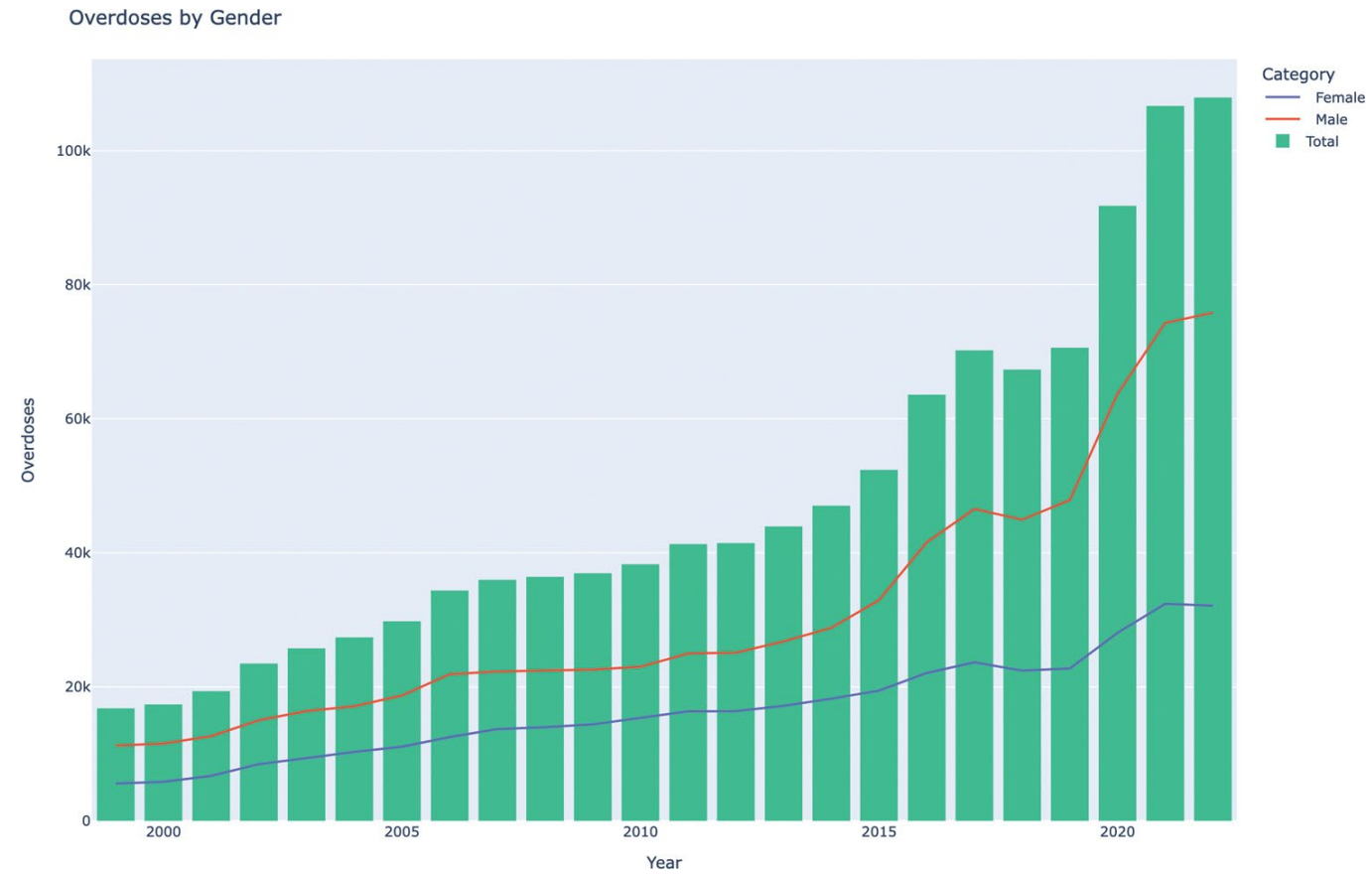


Figure 7.41 Bar/line chart of overdoses by gender, created in Plotly Express.

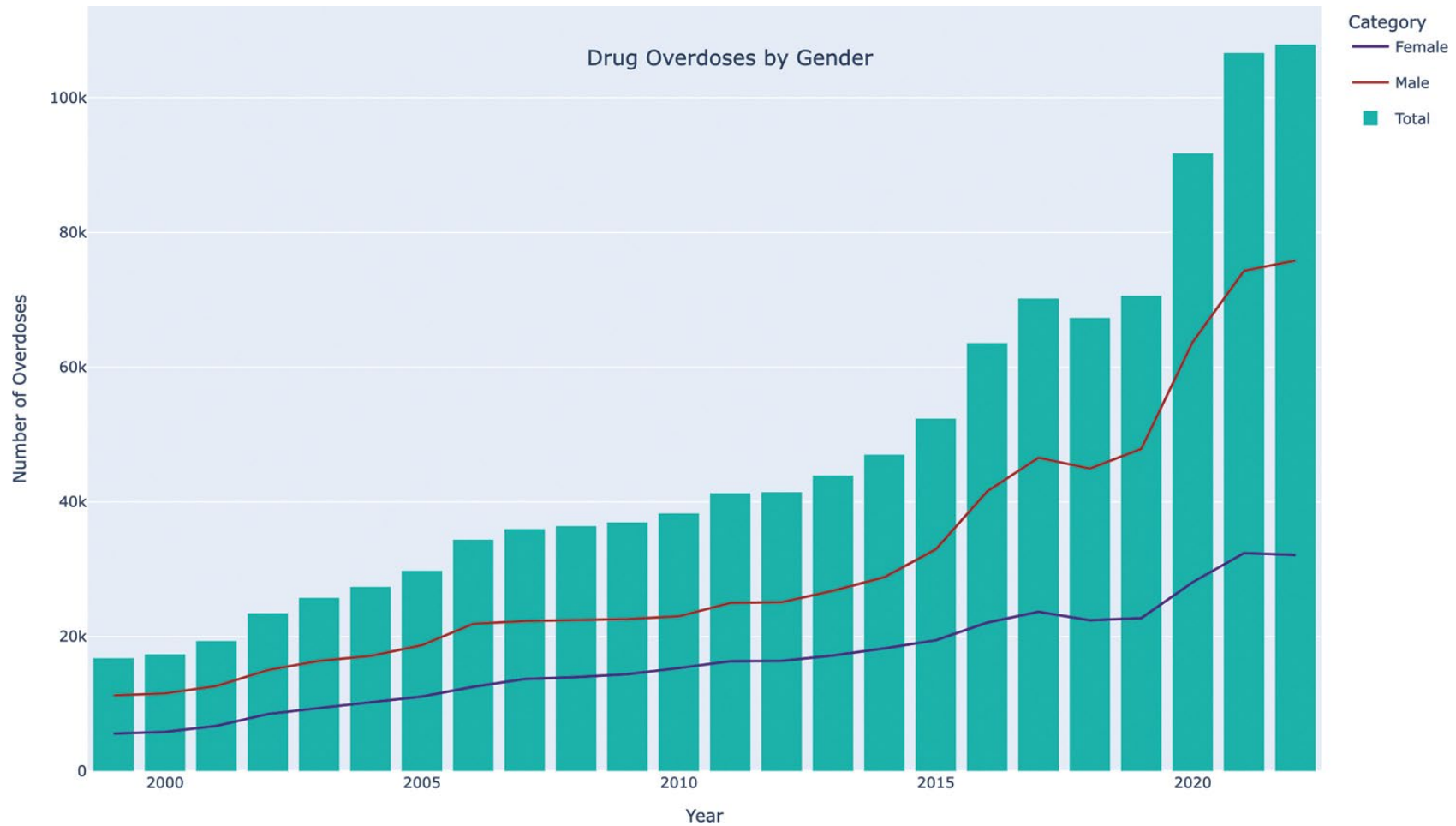


Figure 7.42 Bar/line chart of overdoses by gender, created using Graph Objects.

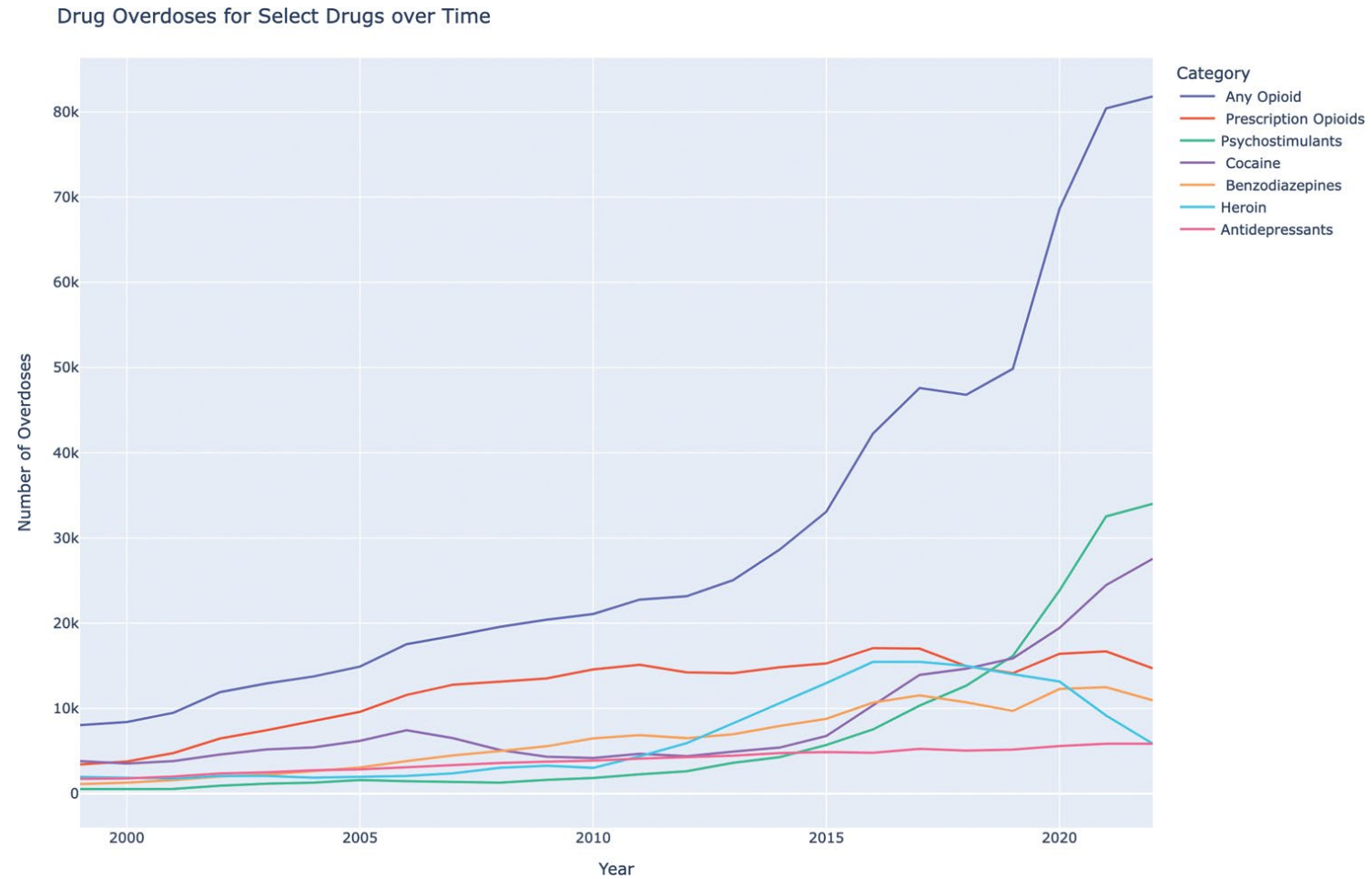


Figure 7.43 Line chart of overdoses on selected drugs over time.

Overdose Rate and Annual Percent Change

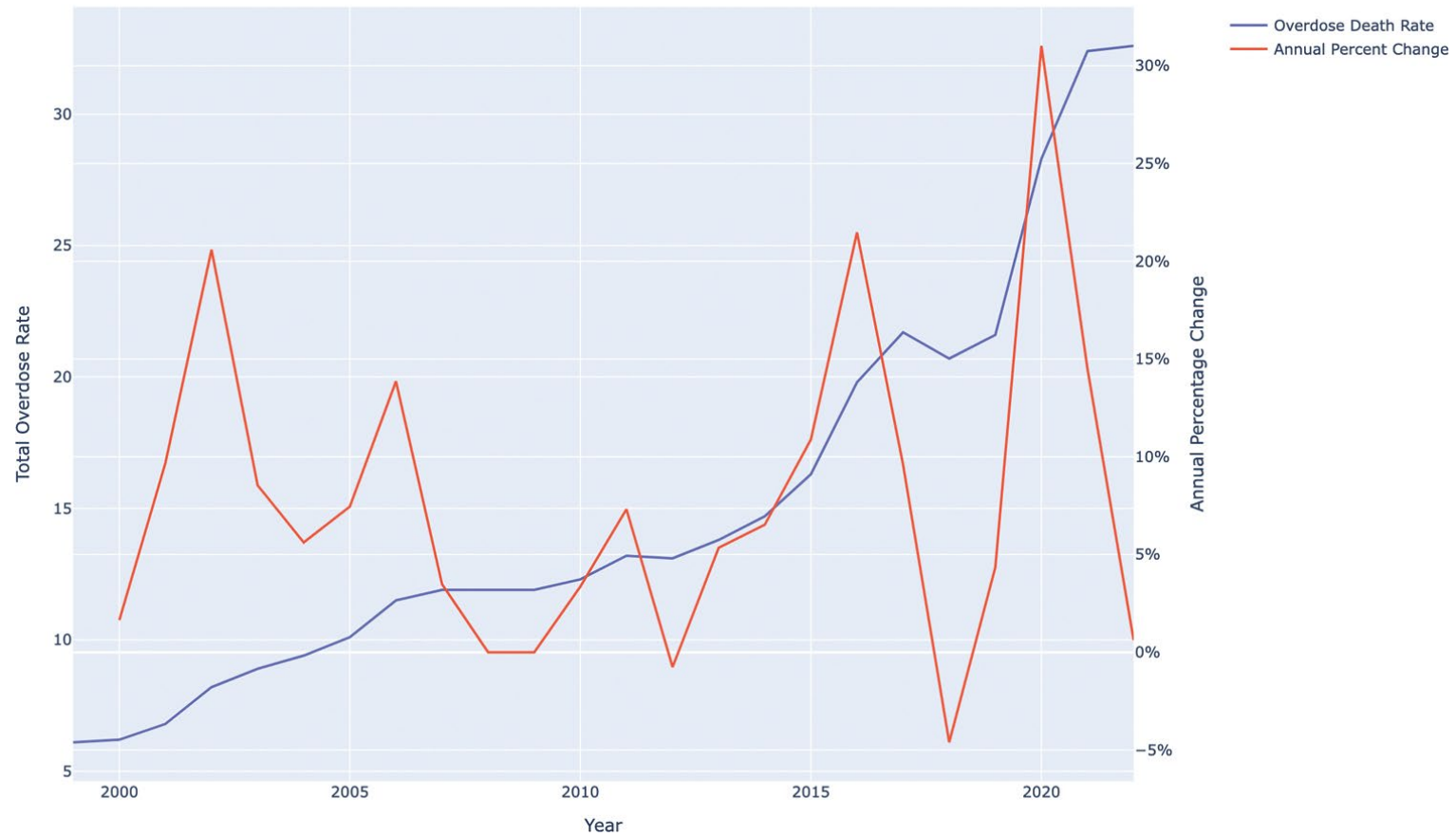


Figure 7.44 Overdose rate and annual percentage change.

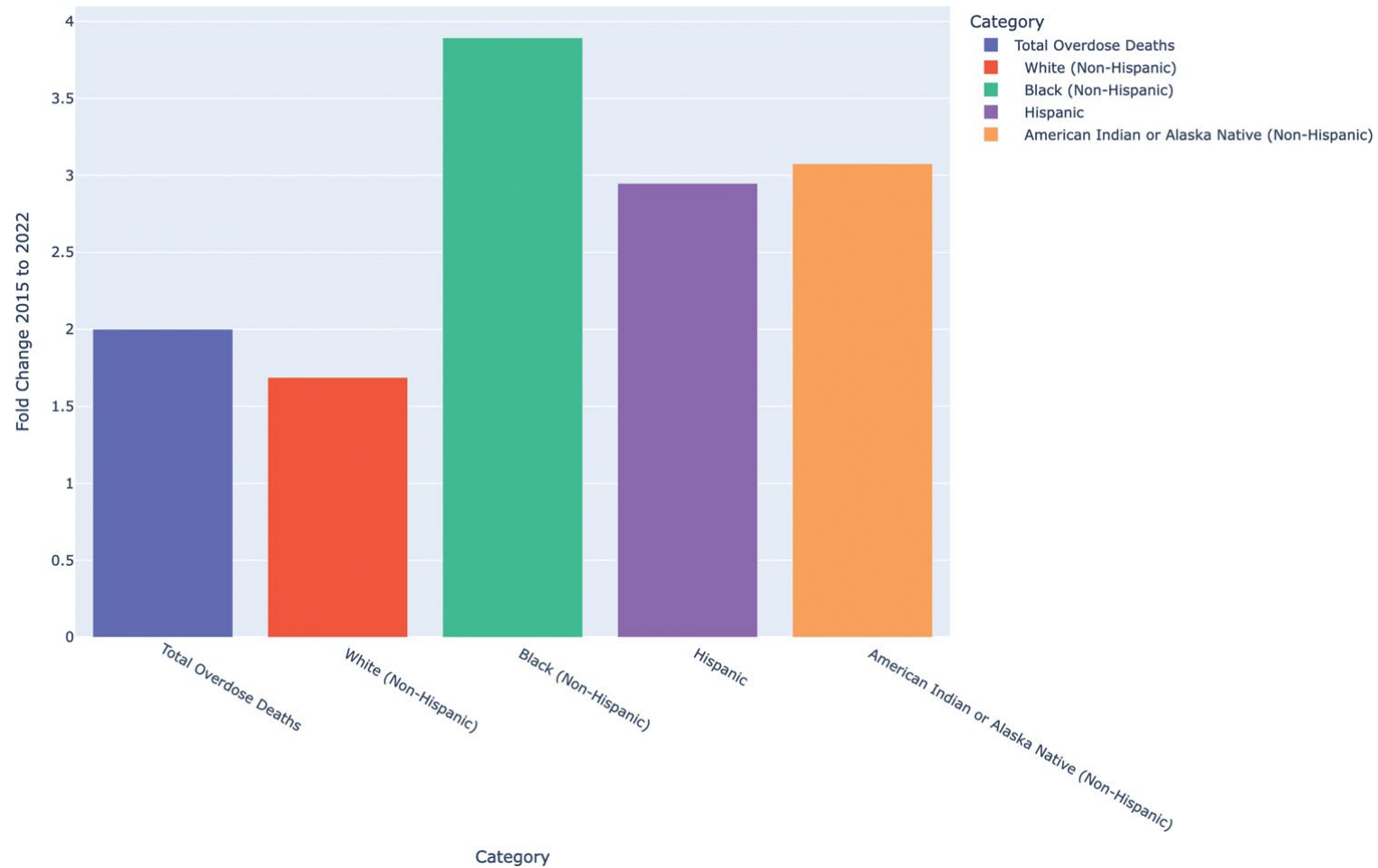


Figure 7.45 Overdoses by race.

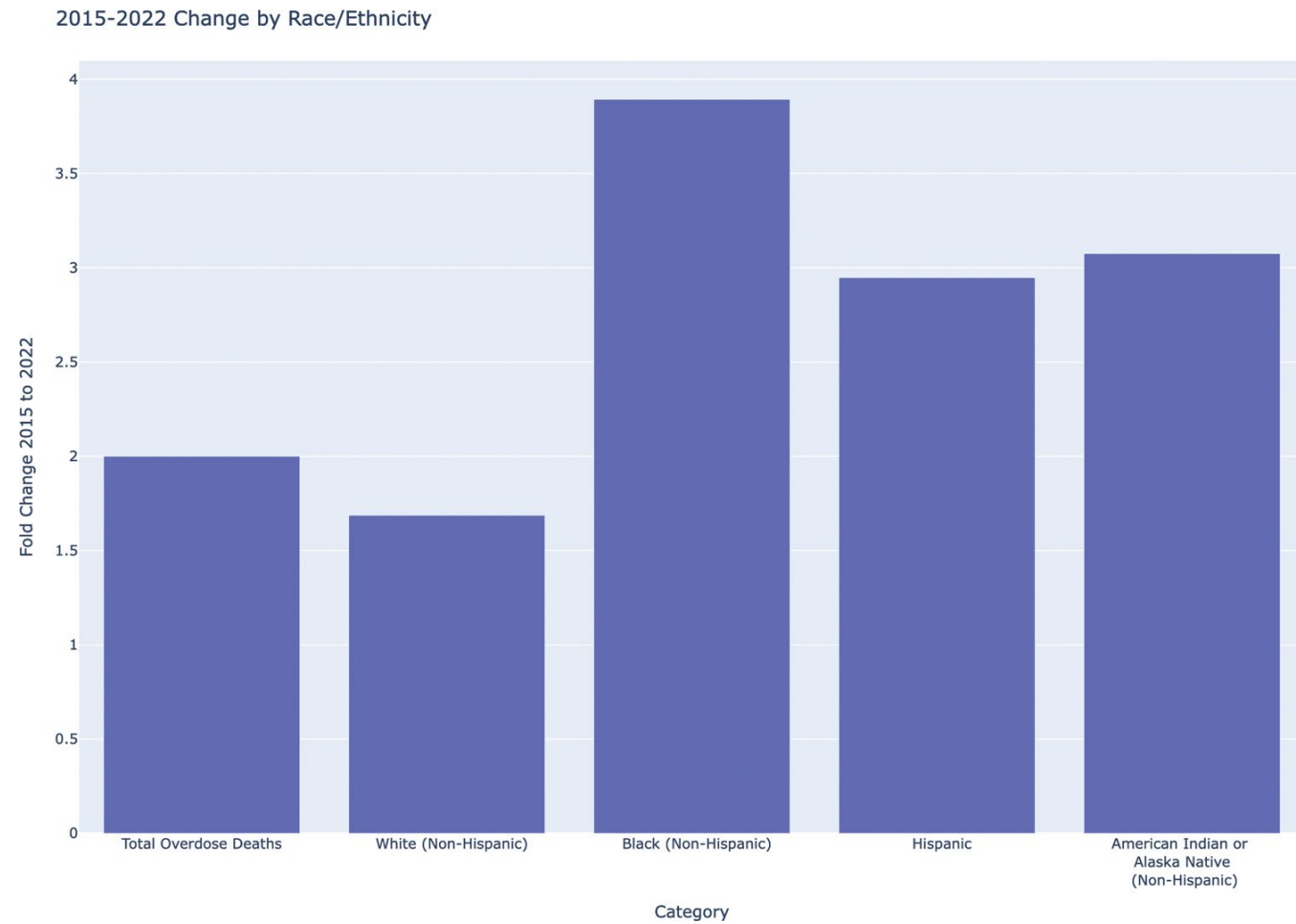
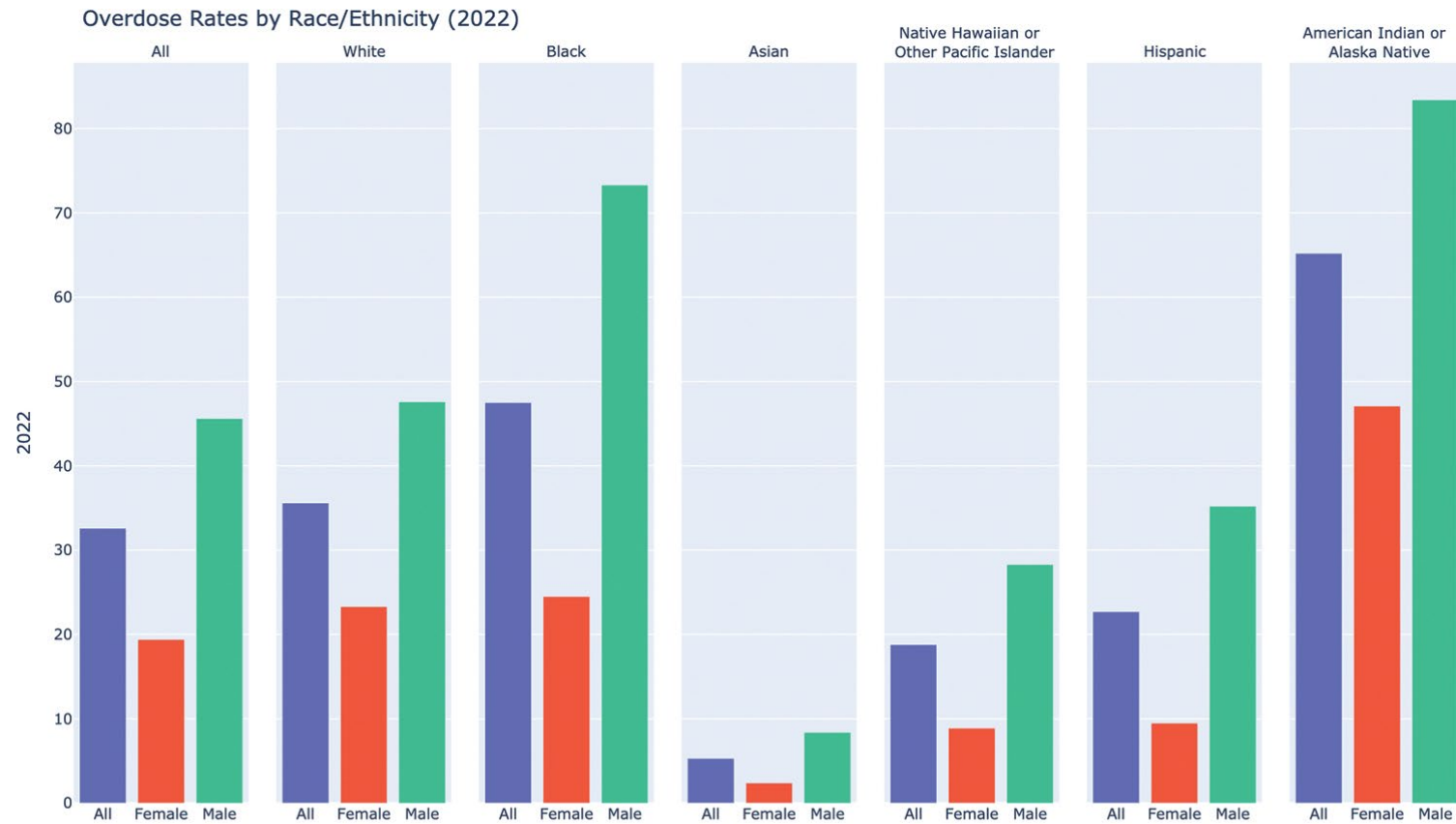


Figure 7.46 Overdoses by race (monochrome).



Note: Categories other than Hispanic and All do not include Hispanics.

Figure 7.47 Overdose rates faceted by race and gender.

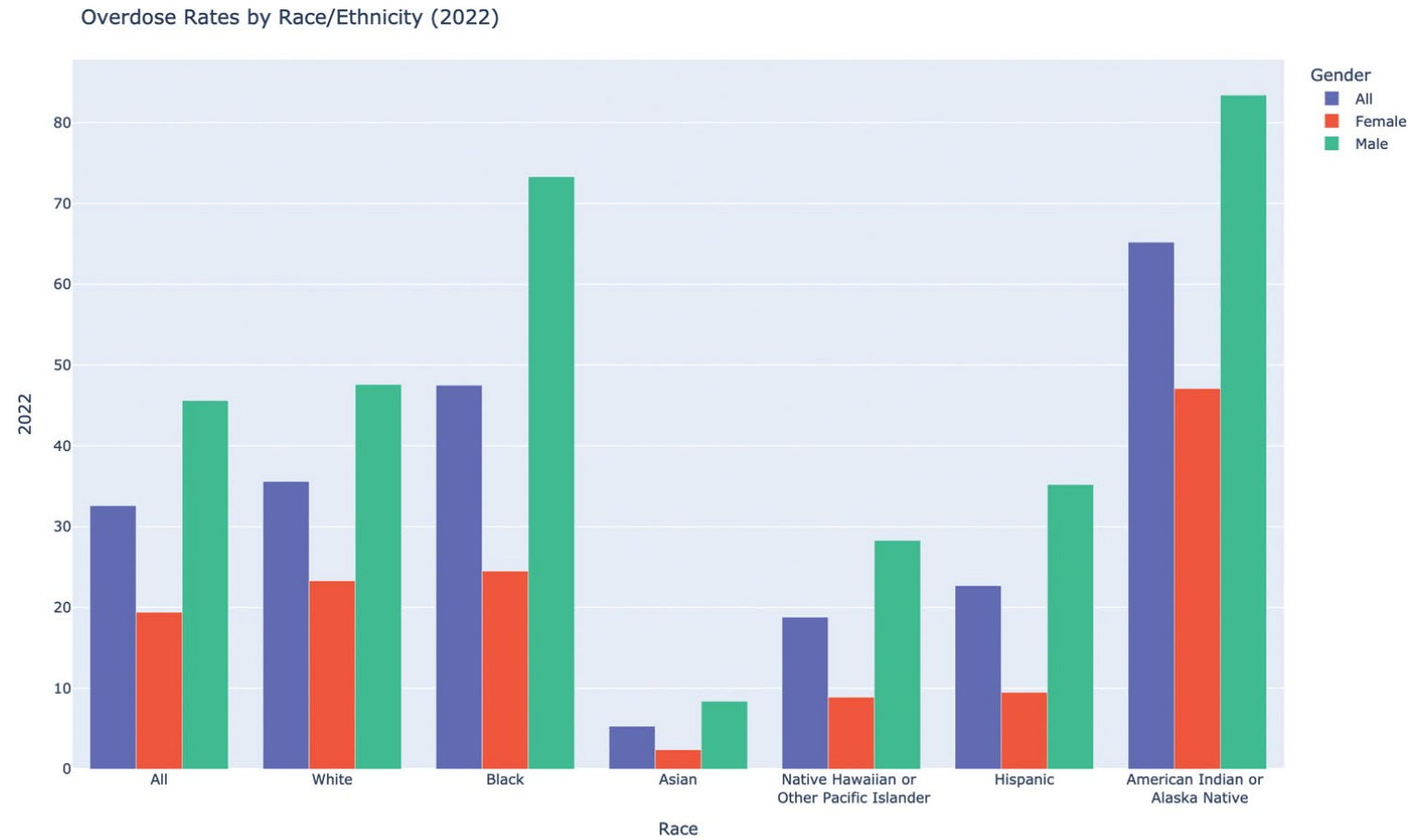
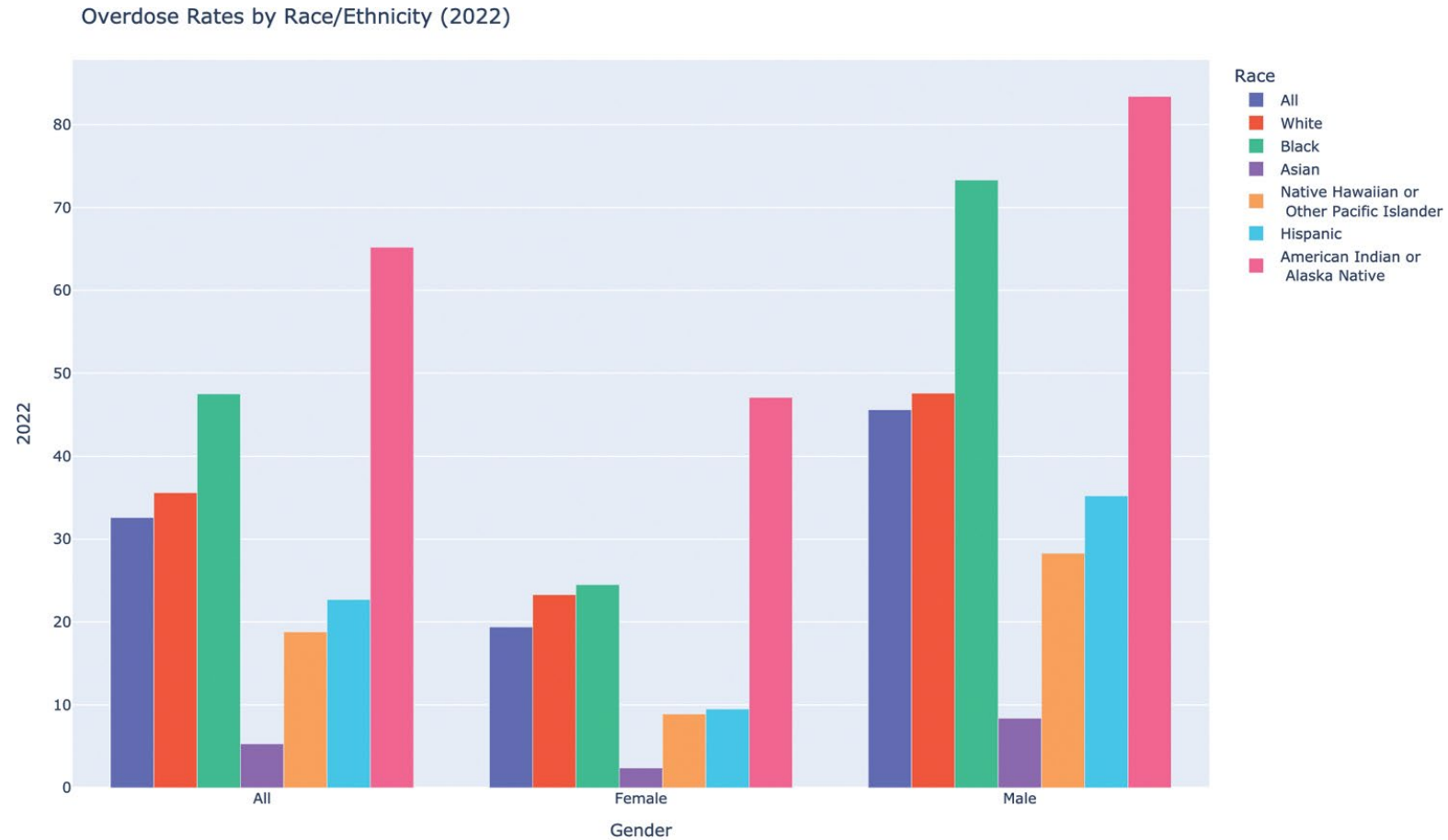
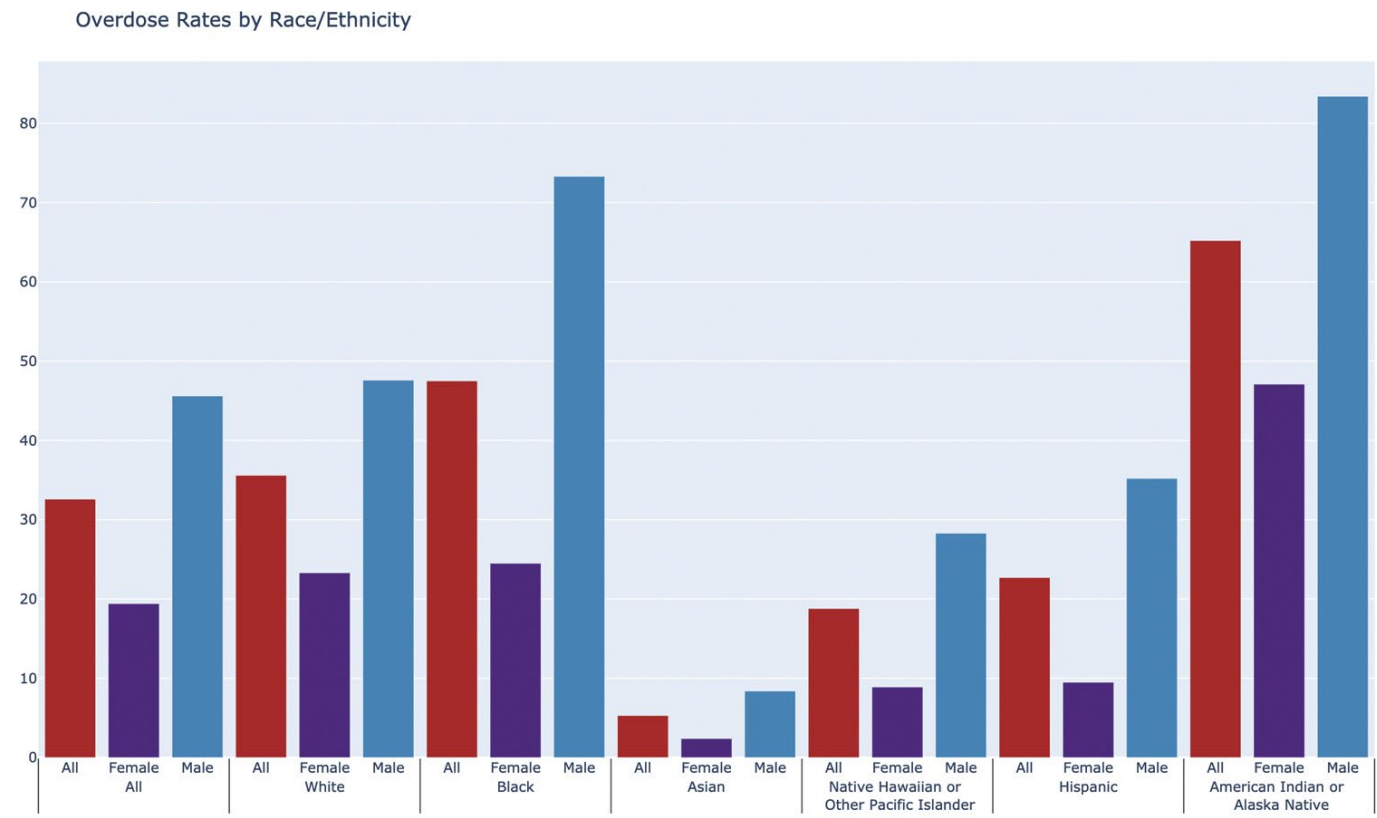


Figure 7.48 Overdose rates organized by race and gender.



Note: Categories other than Hispanic and All do not include Hispanics.

Figure 7.49 Overdose rates organized first by gender, then by race.



Note: Categories other than Hispanic and All do not include Hispanics.

Figure 7.50 Overdose rates multi-indexed with column dividers.

OD_script.R x

odmf x

←→



 Filter

	Year	Total.Overdose.Deaths	Female	Male
1	1999	16849	5591	11258
2	2000	17415	5852	11563
3	2001	19394	6736	12658
4	2002	23518	8490	15028
5	2003	25785	9386	16399
6	2004	27424	10304	17120

Figure 7.51 Total U.S. overdose deaths by sex, 1999–2022, in R.

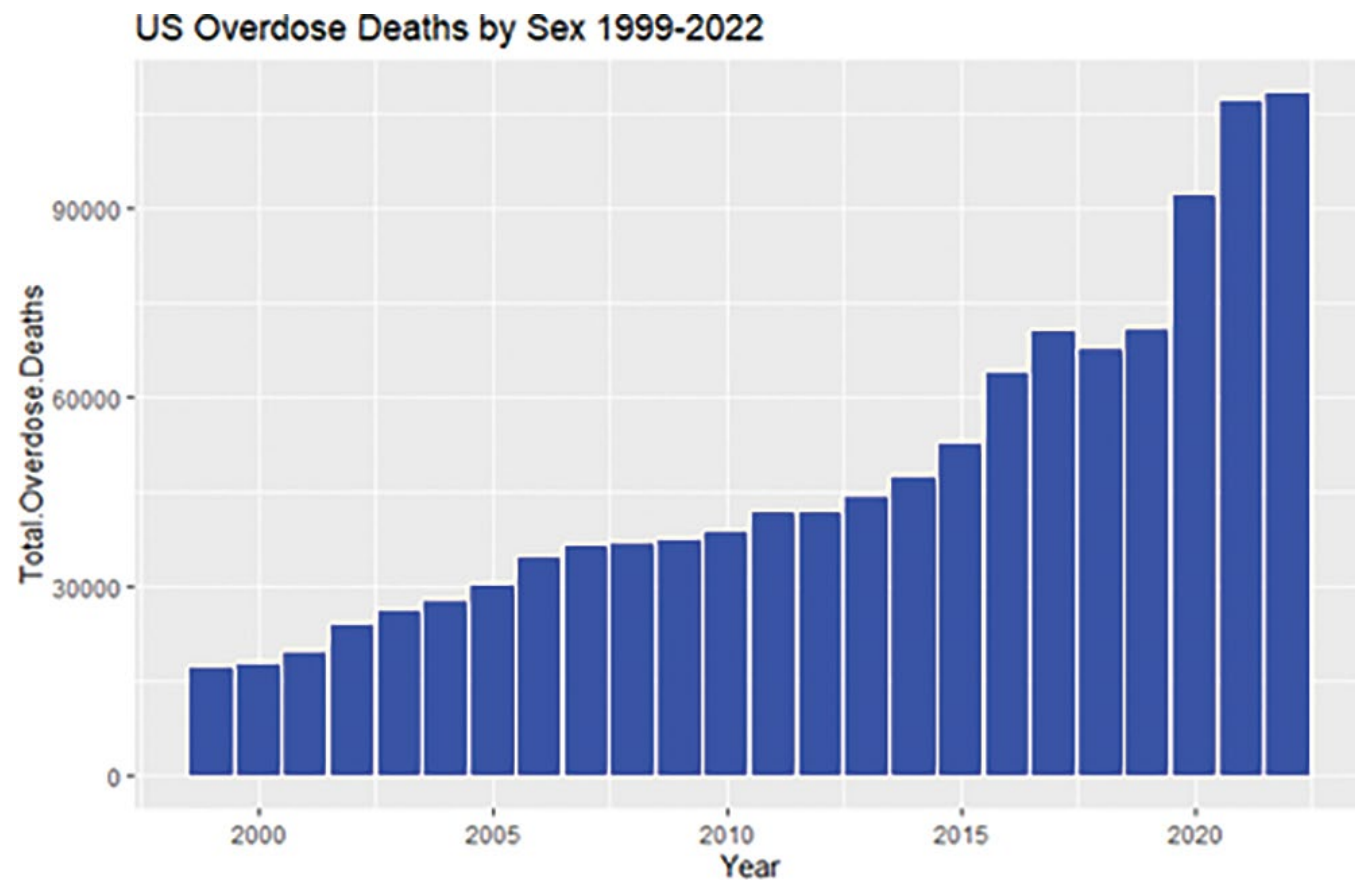


Figure 7.52 Total U.S. overdose deaths, 1999–2022, created using ggplot in R.

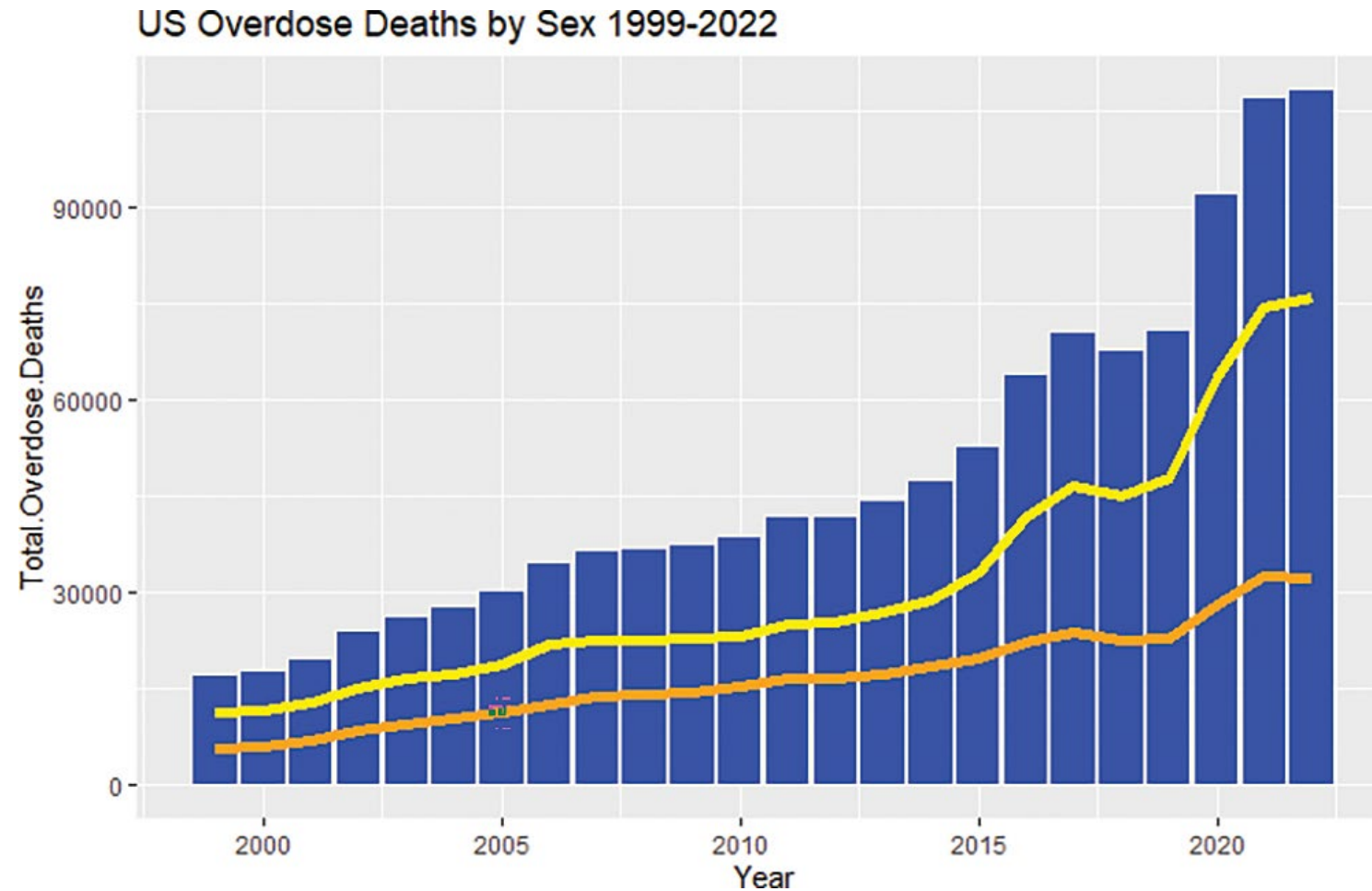


Figure 7.53 Total U.S. overdose deaths, 1999–2022, with lines for sex but no legend, created using ggplot in R.

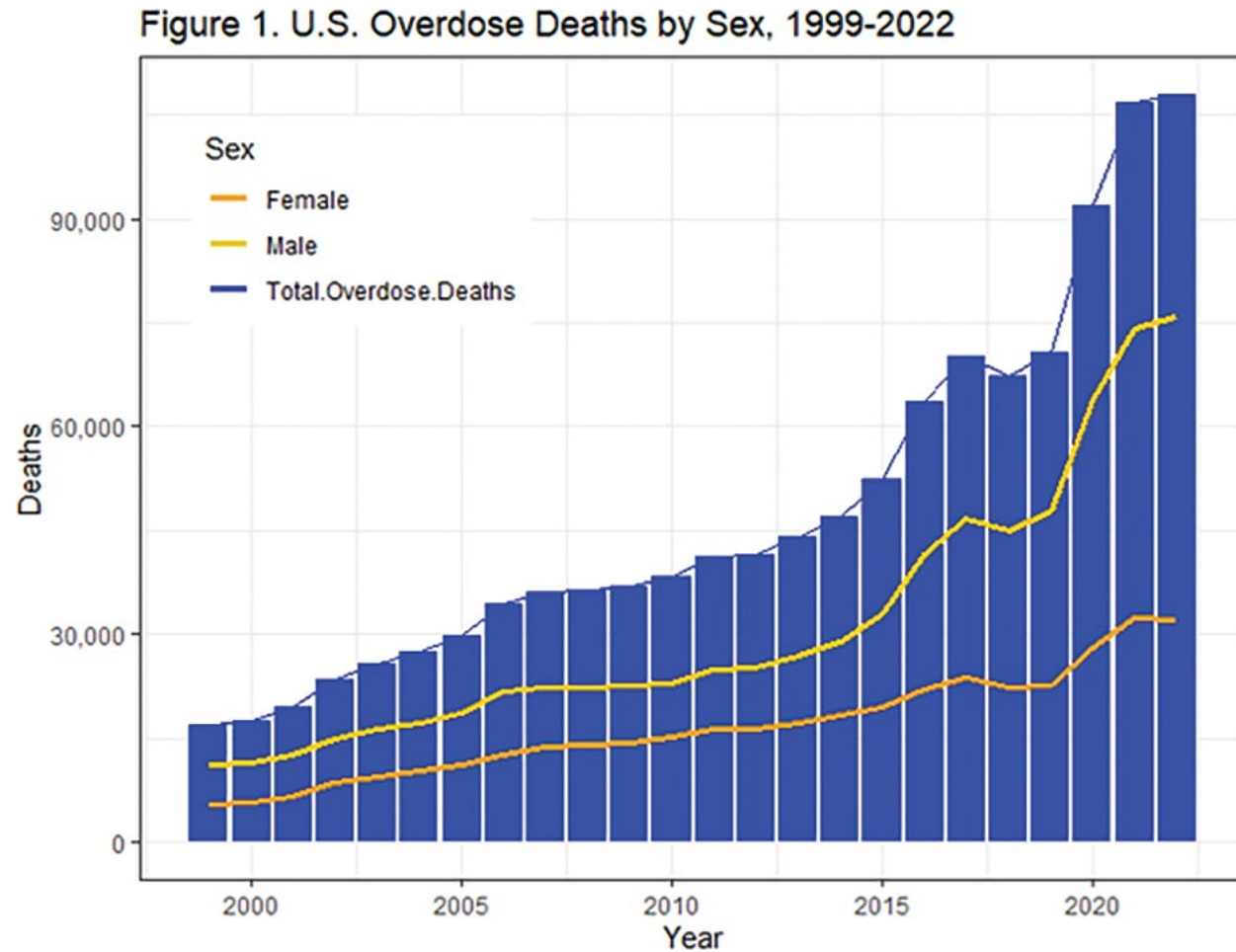


Figure 7.54 Total U.S. overdose deaths, 1999–2022, with lines for sex and a legend, created using ggplot and the scales package in R.

US Overdose Deaths by Sex 1999 - 2022

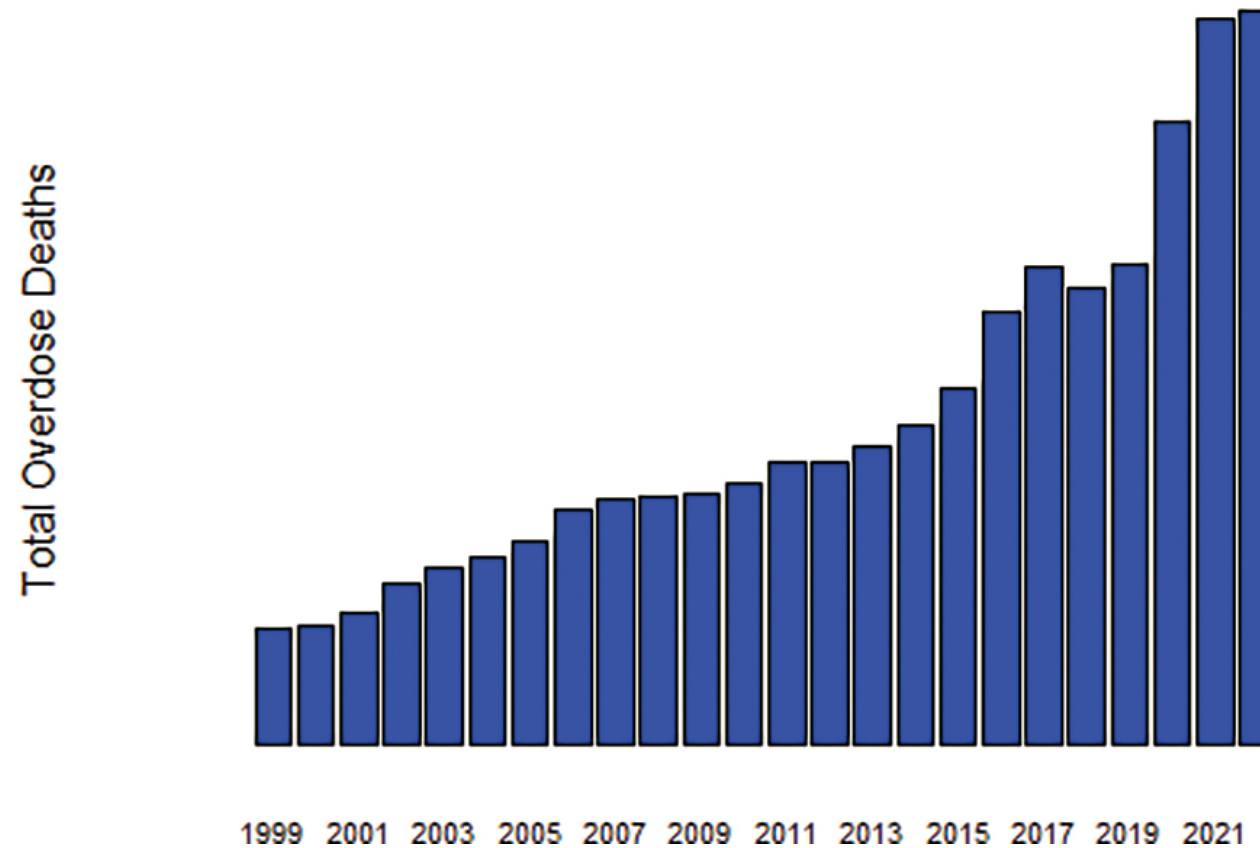


Figure 7.55 Total U.S. overdose deaths, 1999–2022, using the base bar plot in R.

US Overdose Deaths by Sex 1999 - 2022

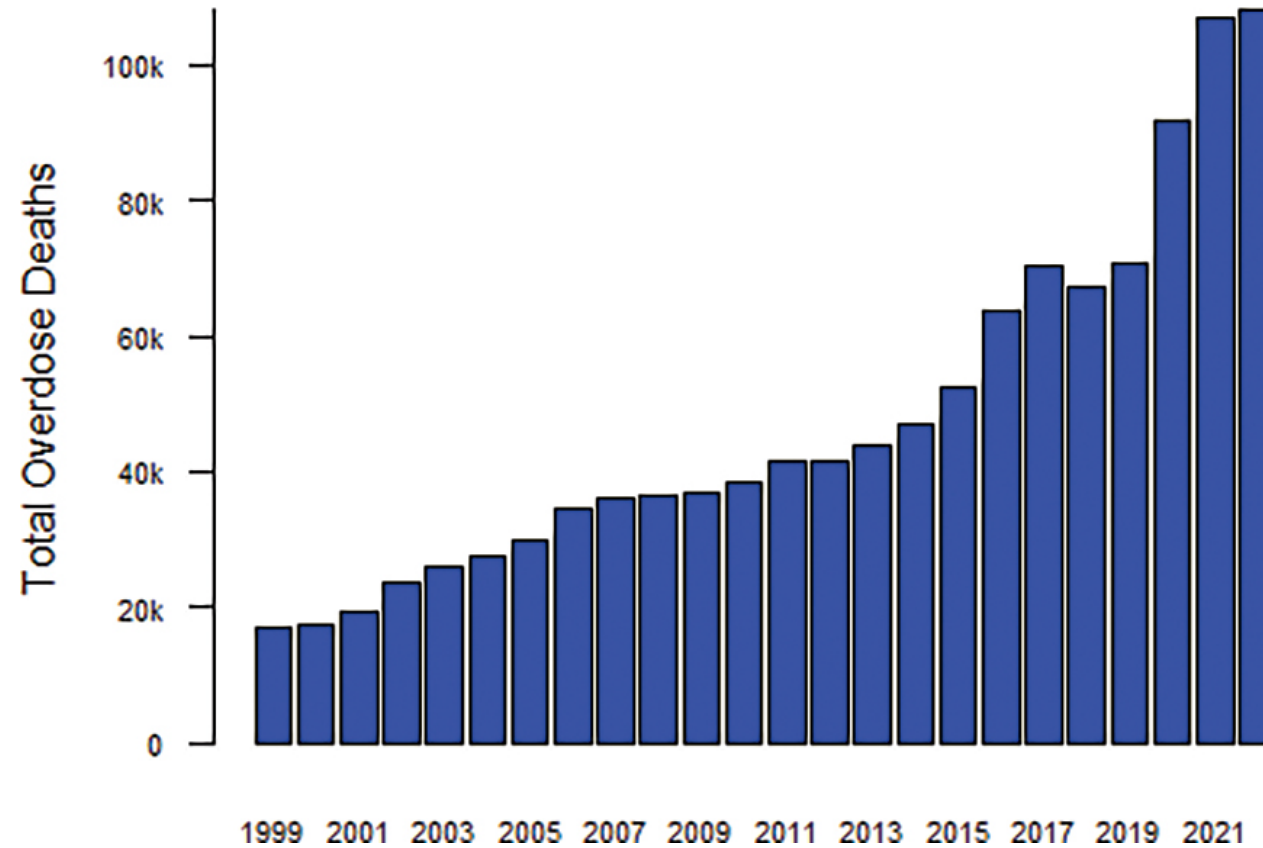


Figure 7.56 Total U.S. overdose deaths, 1999–2022, with custom axis labels using the base bar plot in R.

US Overdose Deaths by Sex 1999 - 2022

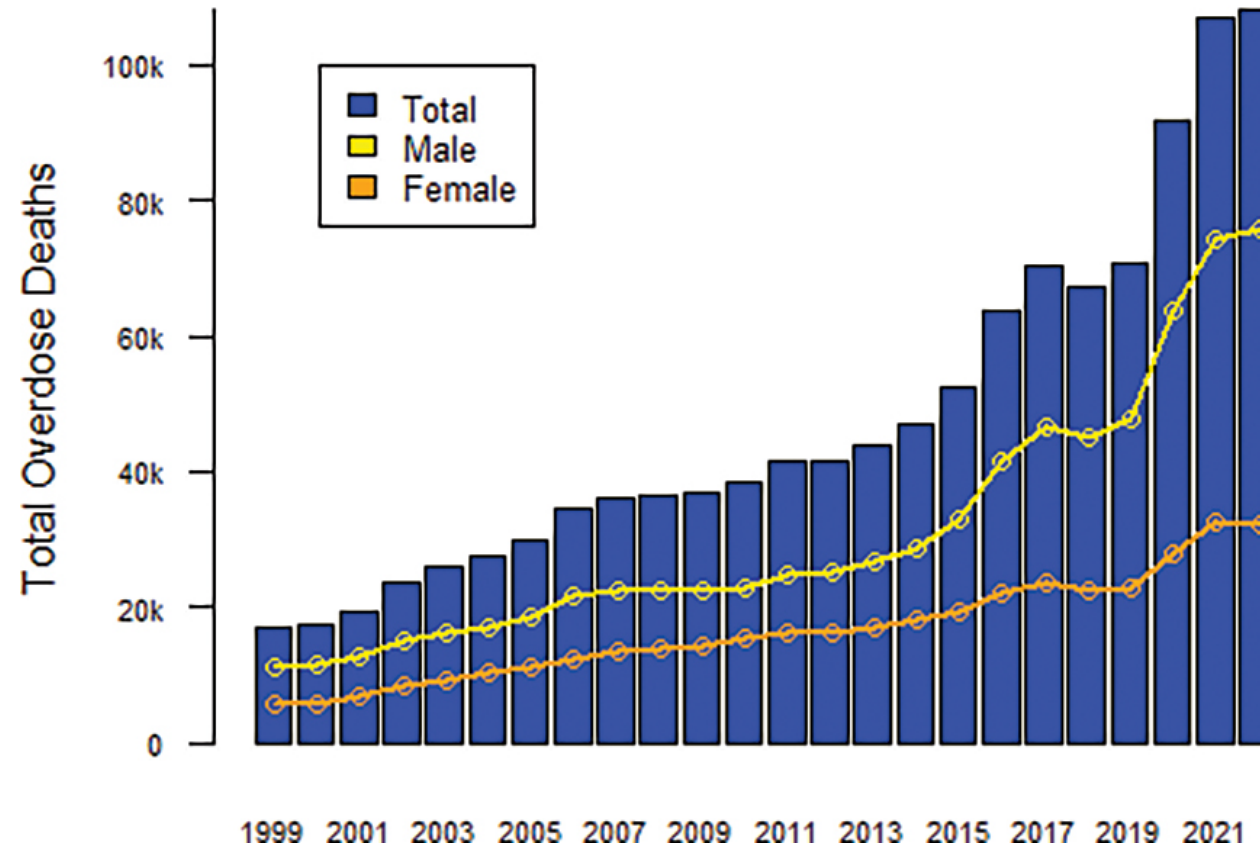


Figure 7.57 Total U.S. overdose deaths by sex, 1999–2022, using the base bar plot in R.

OD_script.R × oddrug × odmelt × odmf ×							
← → 📄 🔍 Filter							
	Year	Any.Opioid	Meth	Cocaine	Heroin	Benzodiazepines	Antid
1	1999	8050	547	3822	1960	1135	
2	2000	8407	578	3544	1842	1298	
3	2001	9496	563	3833	1779	1594	
4	2002	11920	941	4599	2089	2022	
5	2003	12940	1179	5199	2080	2248	
6	2004	13756	1305	5443	1878	2627	
7	2005	14918	1608	6208	2009	3084	
8	2006	15515	1600	7110	2000	3035	

Figure 7.58 Total U.S. overdose deaths by drug type, 1999–2022, in R.

The screenshot shows an RStudio window with three tabs: 'OD_script.R', 'odmelt2', and 'oddrug'. Below the tabs is a toolbar with navigation icons and a 'Filter' button. The main area displays a data table with the following structure:

	Year	Drug	Deaths
1	1999	Any.Opioid	8050
2	2000	Any.Opioid	8407
3	2001	Any.Opioid	9496
4	2002	Any.Opioid	11920
5	2003	Any.Opioid	12940
6	2004	Any.Opioid	13756

Figure 7.59 Total U.S. overdose deaths by drug type, 1999–2022, as tidy data in R.

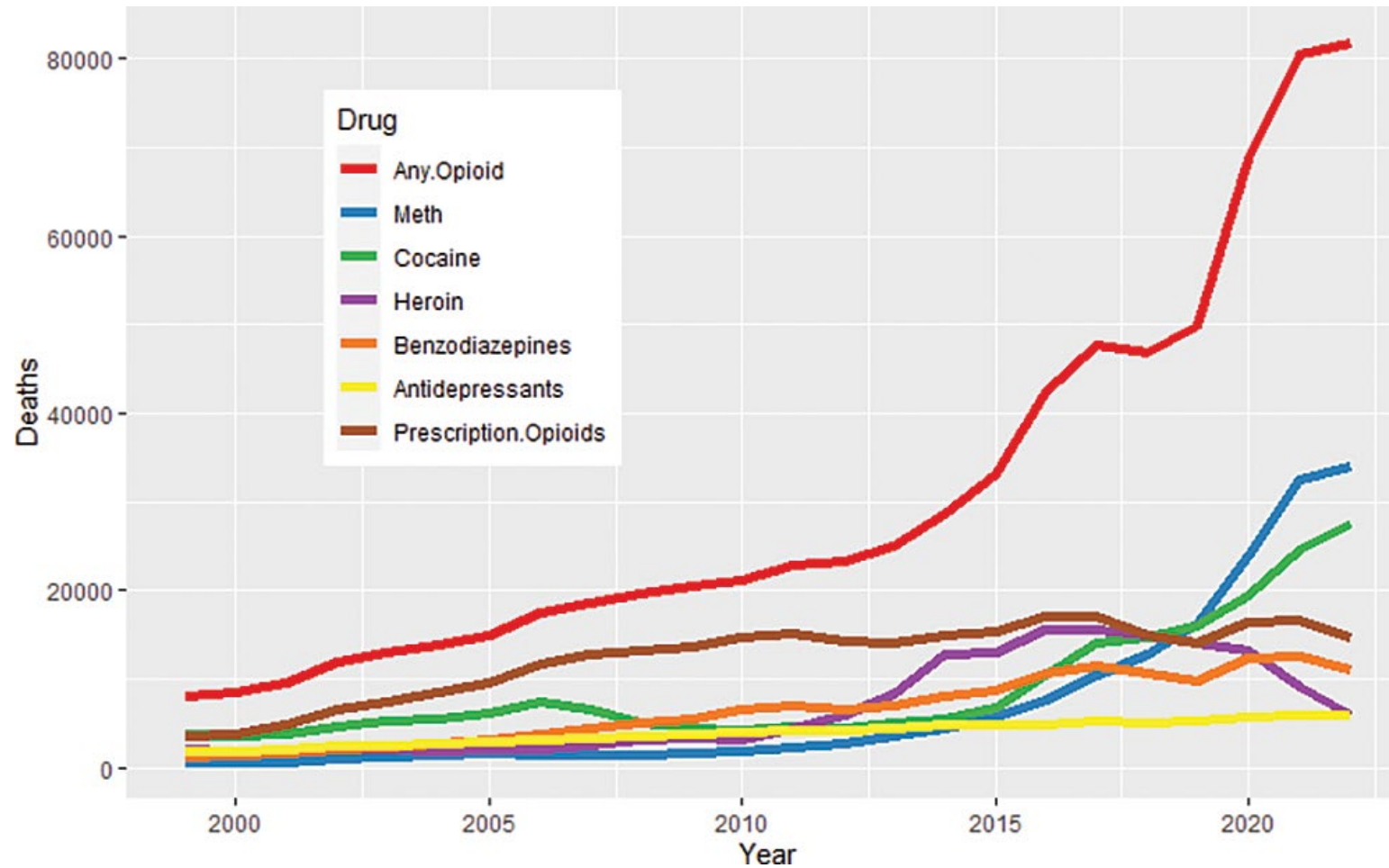


Figure 7.60 Total U.S. overdose deaths by drug type, 1999–2022, created using tidy data and ggplot in R.

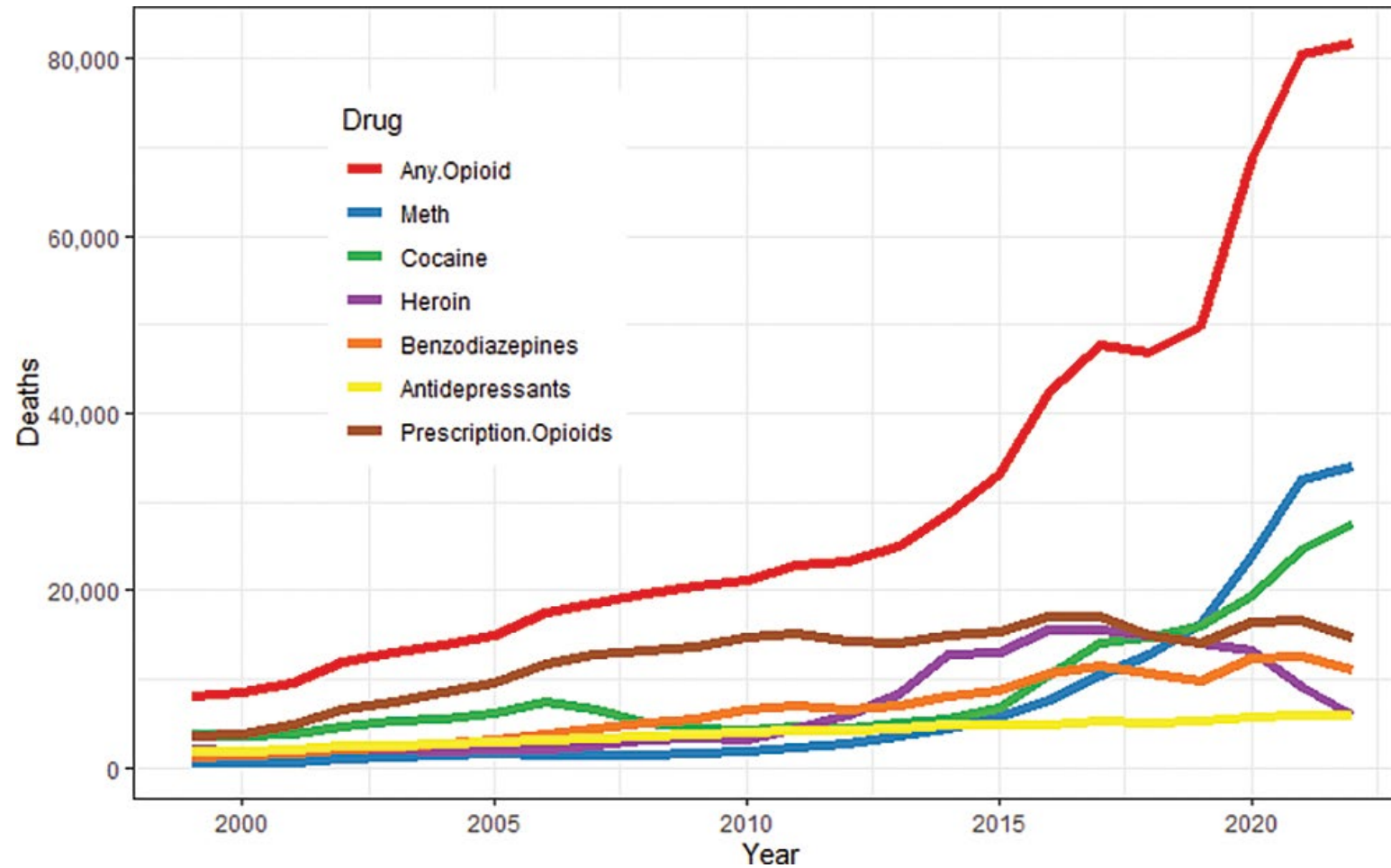


Figure 7.61 Total U.S. overdose deaths by drug type, 1999–2022, with formatting choices in R.

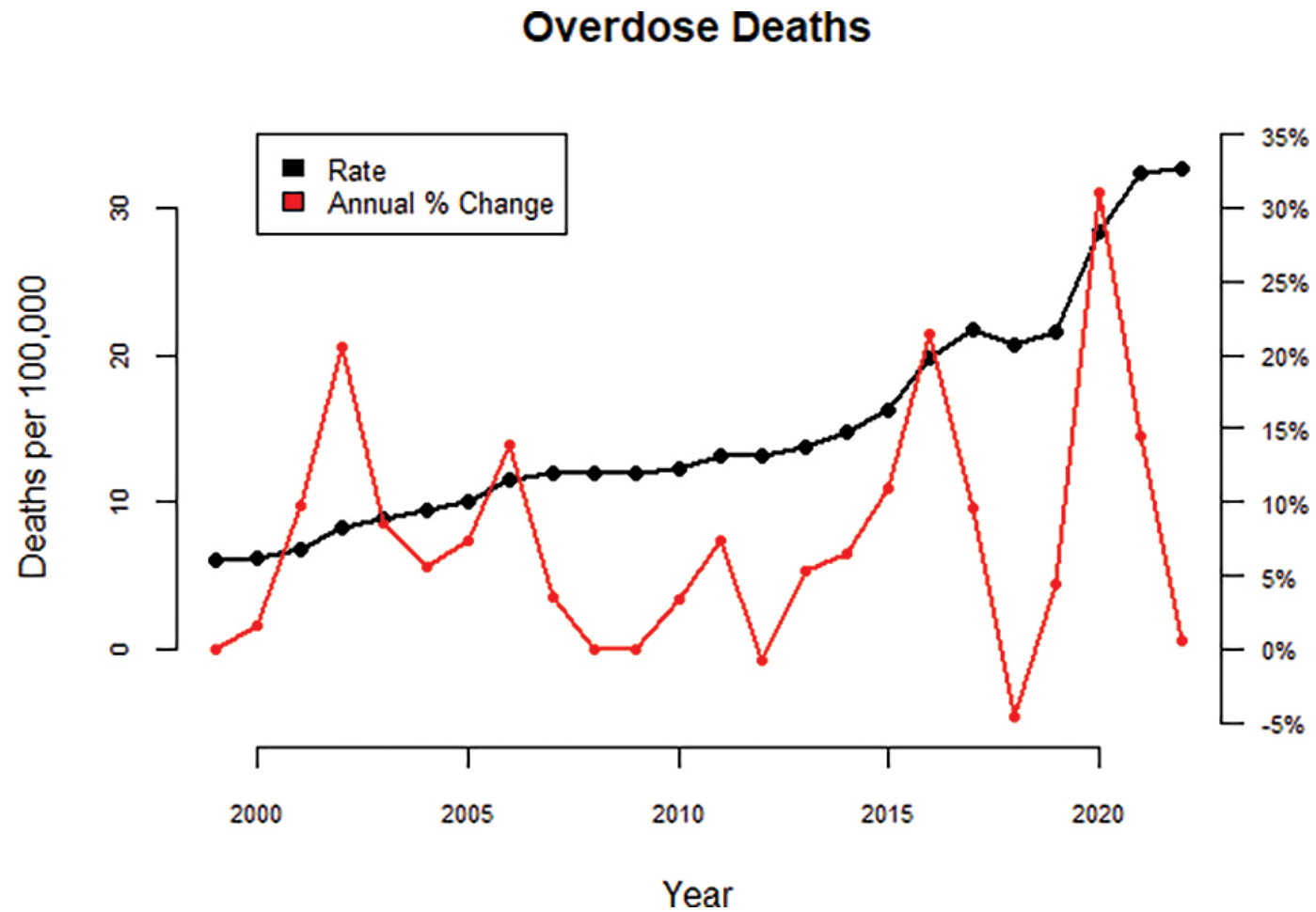


Figure 7.62 Total U.S. overdose death rates, 1999–2022, with annual percent change.

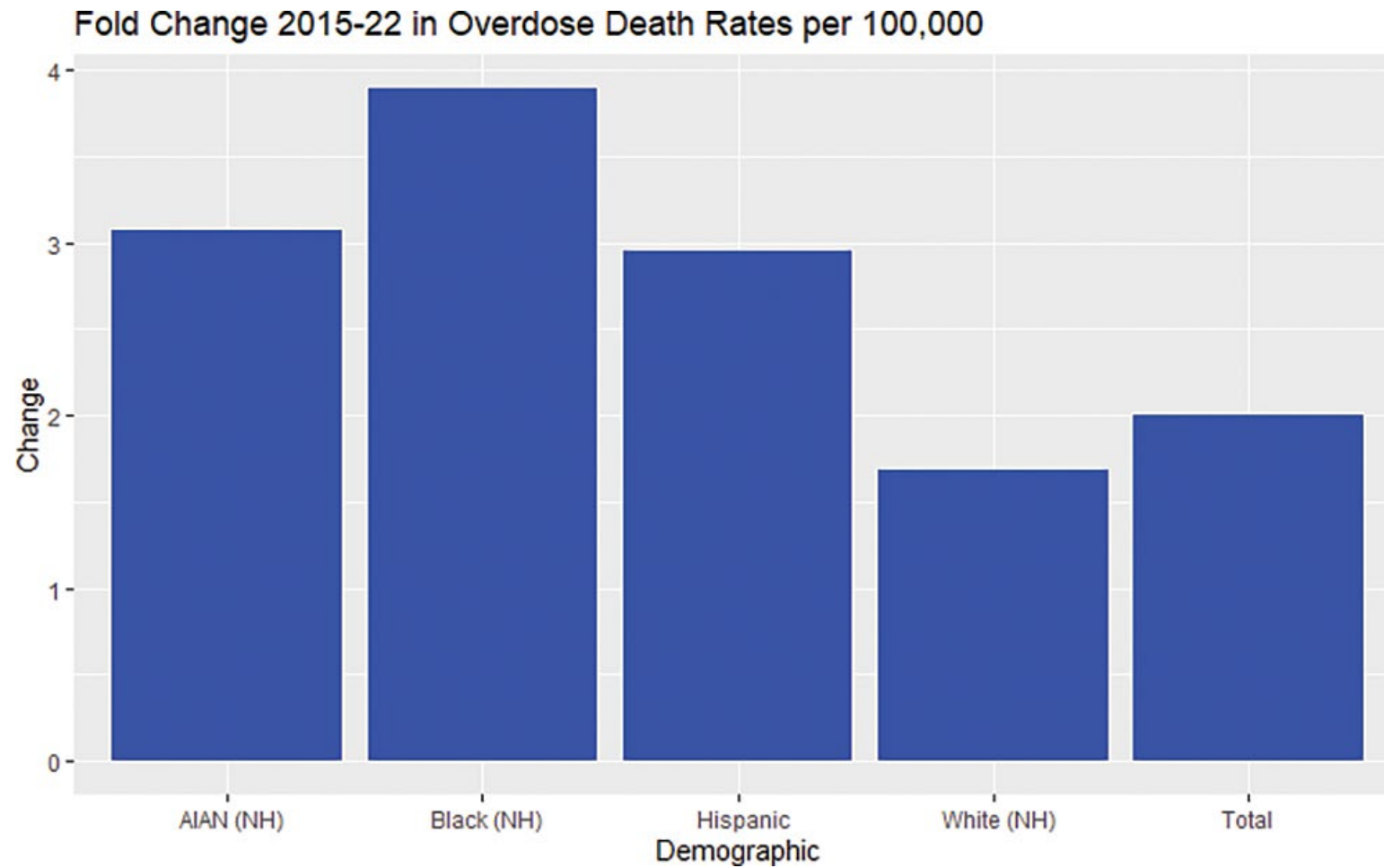
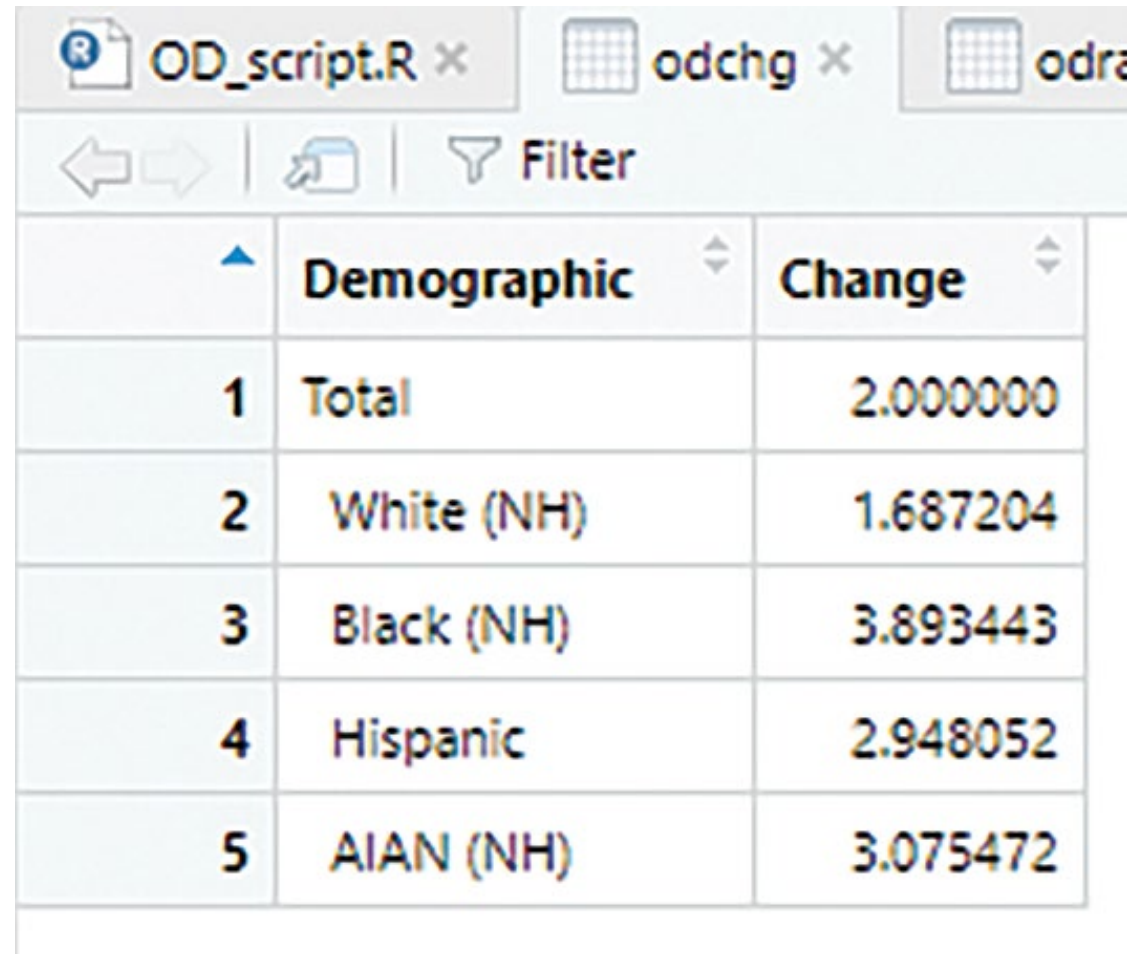


Figure 7.63 Change (as a multiple) in U.S. overdose death rates, 2015–2022, by demographic group.



	Demographic	Change
1	Total	2.000000
2	White (NH)	1.687204
3	Black (NH)	3.893443
4	Hispanic	2.948052
5	AIAN (NH)	3.075472

Figure 7.64 Change (as a multiple) in U.S. overdose death rates, 2015–2022, with spacing.

	Demographic	Rate
1	Total	32.6
2	Female	19.4
3	Male	45.6
4	White (NH)	35.6
5	Female	23.3
6	Male	47.6
7	Black (NH)	47.5
8	Female	24.5
9	Male	73.3
10	Asian (NH)	5.3

Figure 7.65 U.S. overdose death rates, 2022, by demographic group, with spacing issues.

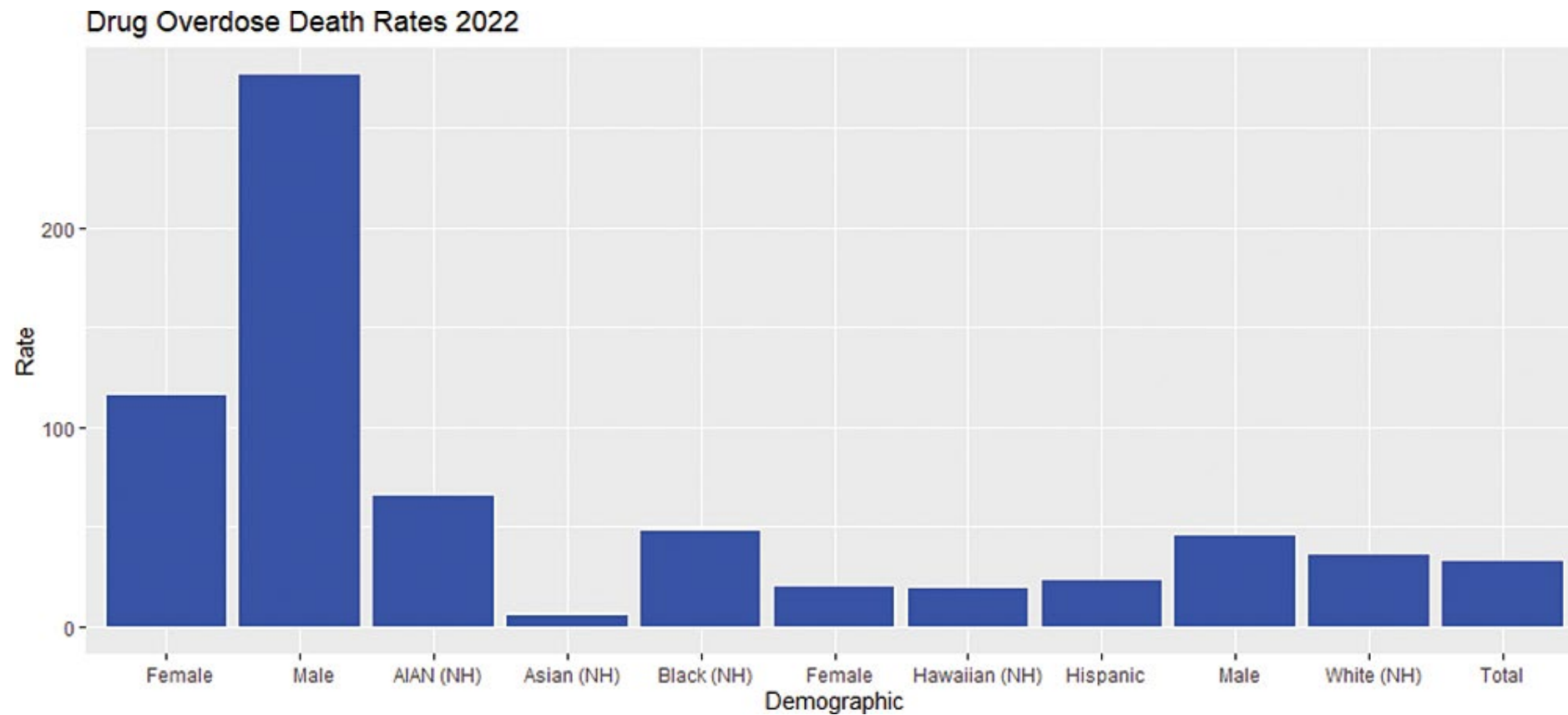
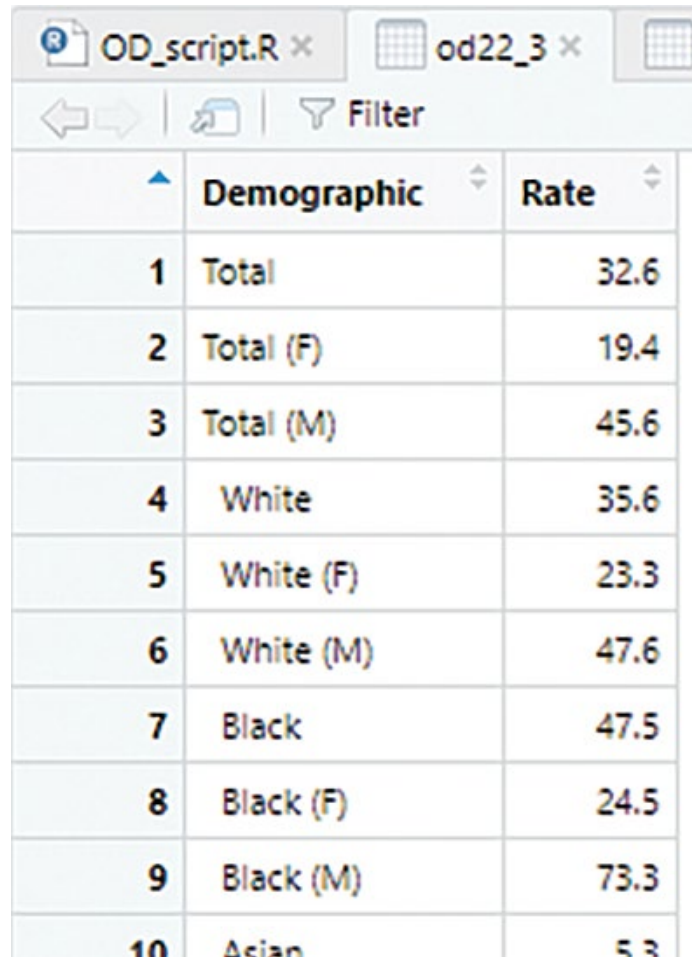


Figure 7.66 U.S. overdose death rates, 2022, by demographic group, with spacing issues.



	Demographic	Rate
1	Total	32.6
2	Total (F)	19.4
3	Total (M)	45.6
4	White	35.6
5	White (F)	23.3
6	White (M)	47.6
7	Black	47.5
8	Black (F)	24.5
9	Black (M)	73.3
10	Asian	5.3

Figure 7.67 U.S. overdose death rates, 2022, by demographic group, with spacing issues resolved.

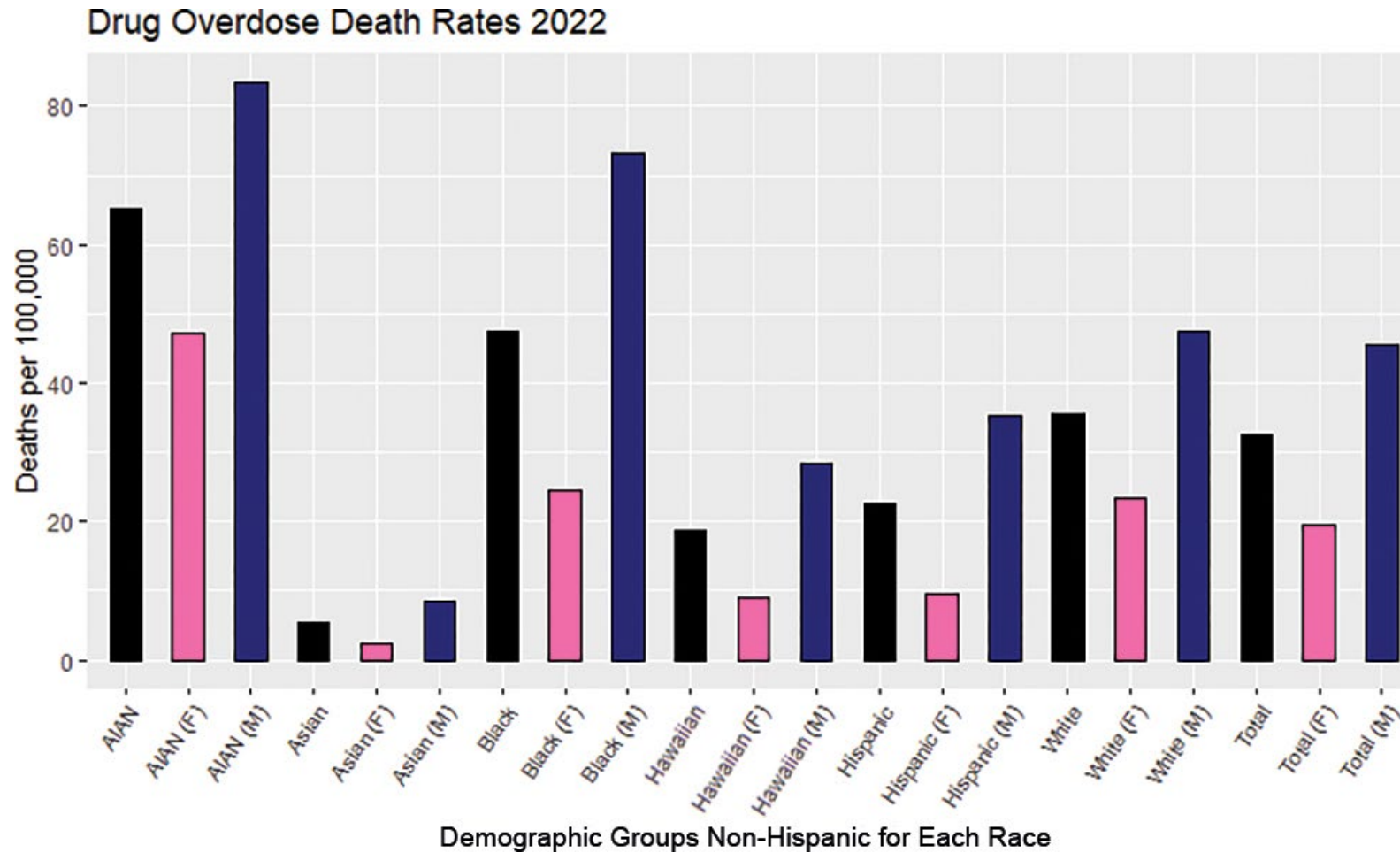


Figure 7.68 U.S. overdose death rates, 2022, by demographic group.

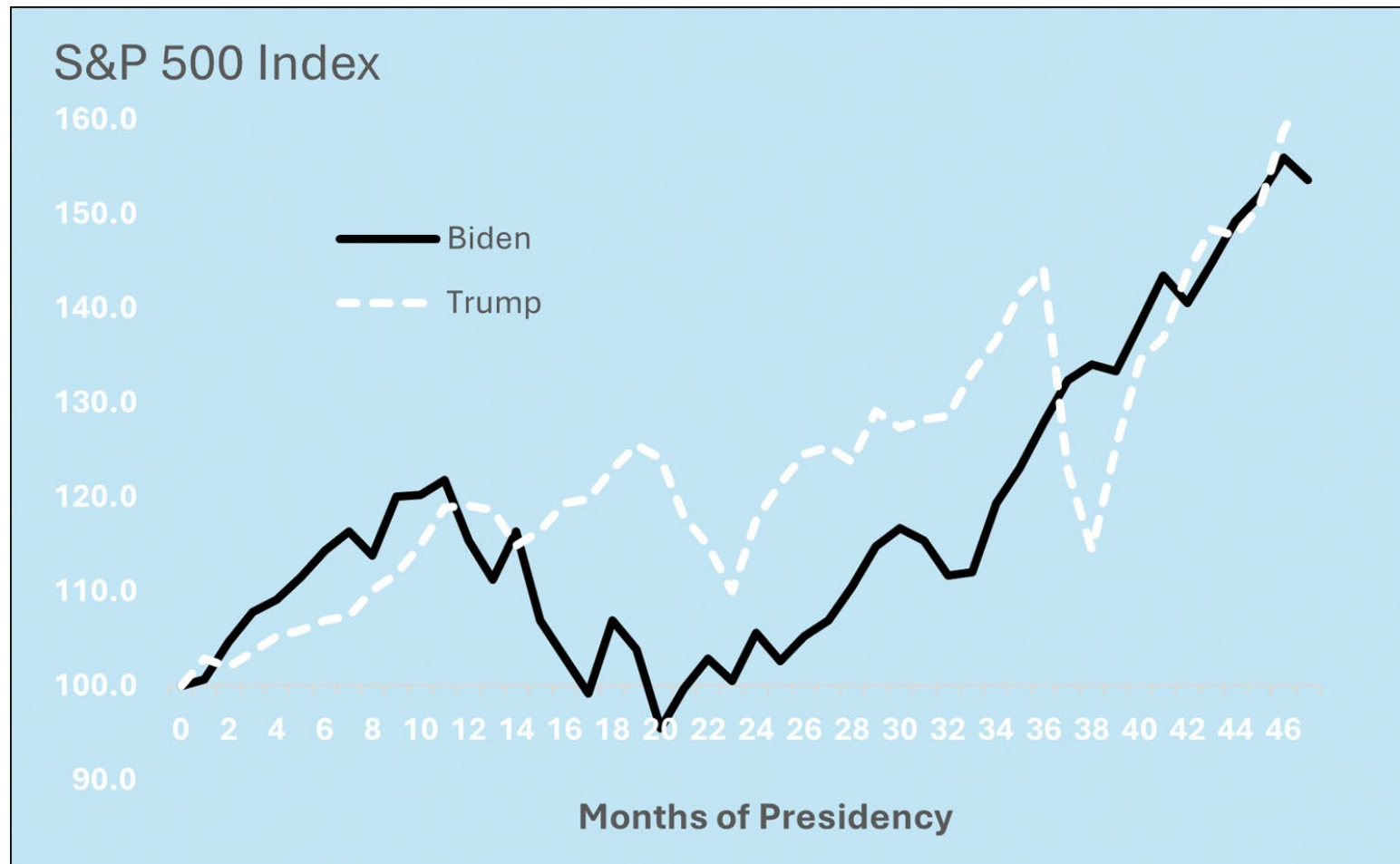


Figure 8.1 Comparing the change in the S&P 500 Index over the first Trump presidency and the Biden presidency.

SPX / Historical Data

S&P 500 (SPX) Historical Data

1M 6M YTD 1Y 5Y **MAX** [Download historical data](#)

Date	Close/Last	Open	High	Low
02/28/2025	5,954.50	5,856.74	5,959.40	5,837.66
02/27/2025	5,861.57	5,981.88	5,993.69	5,858.78
02/26/2025	5,859.00	5,870.07	5,880.00	5,800.00

Figure 8.2 S&P 500 Index historical data.

H2 × ✓ <i>fx</i> =DATEDIF(G2,A2,"m")								
	A	B	C	D	E	F	G	H
1	Date	Close/Last	Open	High	Low	President	Start	Months
2	1/17/2025	5996.66	5995.4	6014.96	5978.44	Biden	1/21/2021	47
3	1/16/2025	5937.34	5963.61	5964.69	5930.72	Biden	1/21/2021	47
4	1/15/2025	5949.91	5905.21	5960.61	5905.21	Biden	1/21/2021	47
5	1/14/2025	5842.91	5859.27	5871.92	5805.42	Biden	1/21/2021	47
6	1/13/2025	5836.22	5782.02	5838.61	5773.31	Biden	1/21/2021	47
7	1/10/2025	5827.04	5890.35	5890.35	5807.78	Biden	1/21/2021	47
8	1/8/2025	5918.25	5910.66	5927.89	5874.78	Biden	1/21/2021	47

Figure 8.3 S&P 500 Index historical data with columns for each president, their term start date, and the number of months into each term.

Average of Open Column Labels ▼		
Row Labels ▼	Biden	Trump
0	3862.5	2305.2
1	3887.8	2372.2
2	4042.1	2350.8
3	4163.5	2388.2
4	4215.8	2427.2
5	4307.8	2440.9
6	4414.7	2465.3
7	4492.9	2474.2

Drag fields between areas below:

Filters 	Columns President ▼
Rows Months ▼	Values Average of Open ▼

Figure 8.4 S&P 500 Index historical data as a PivotTable.

	A	B	C	D	E	F	G	H	I
1									
2									
3	Average of Open Column Labels ▼								
4	Row Labels ▼	Biden	Trump	Grand Total				Biden	Trump
5	0	3862.5	2305.2	3065.7			0	100.0	100.0
6	1	3887.8	2372.2	3130.0			1	=B6/B\$5*100	
7	2	4042.1	2350.8	3196.5			2	104.7	102.0
8	3	4163.5	2388.2	3296.5			3	107.8	103.6
9	4	4215.8	2427.2	3278.9			4	109.1	105.3

Figure 8.5 S&P 500 Index historical data as a PivotTable with start values set to 100 and other values scaled proportionally.

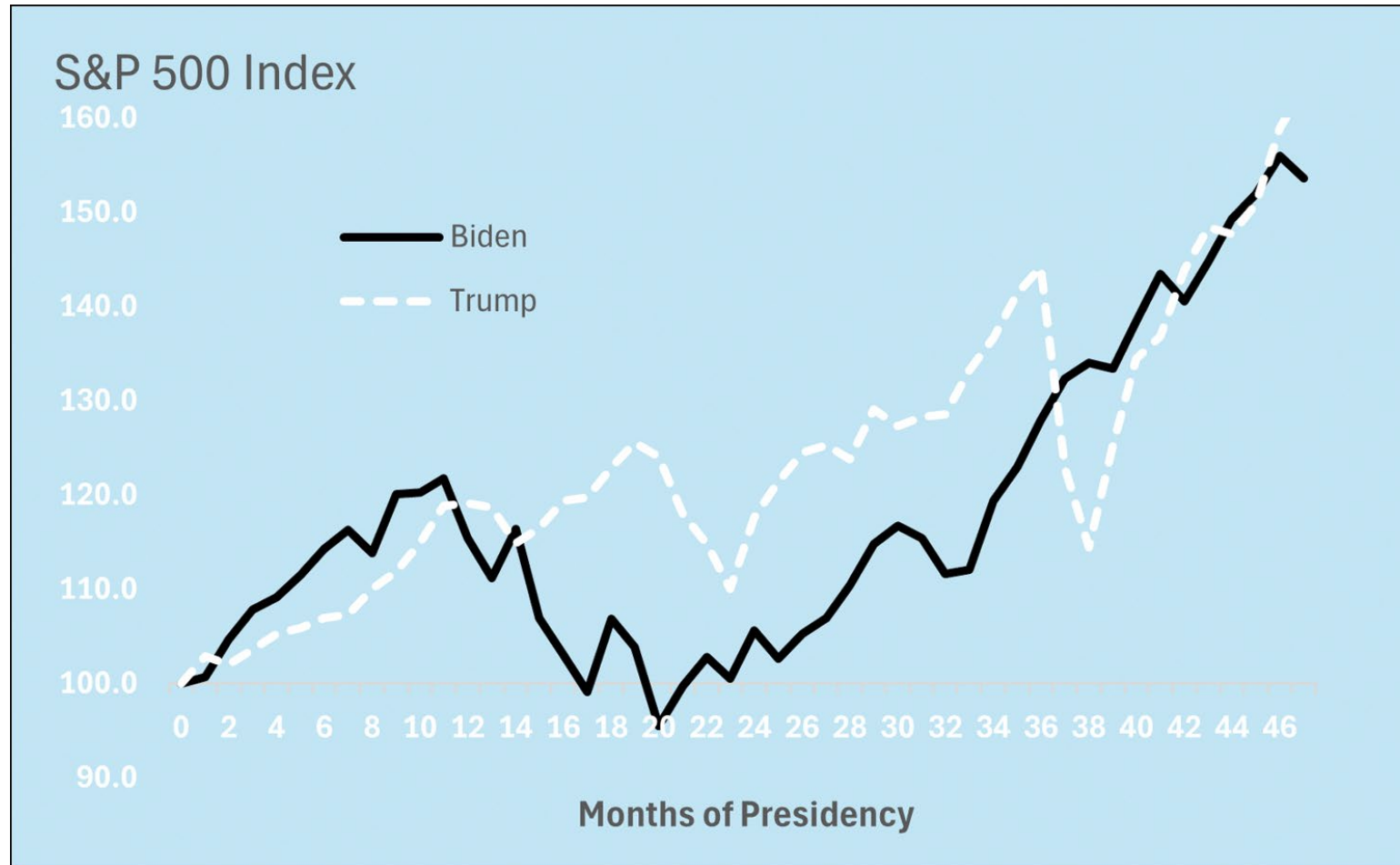


Figure 8.6 Comparing the change in the S&P 500 Index over the first Trump presidency and the Biden presidency.

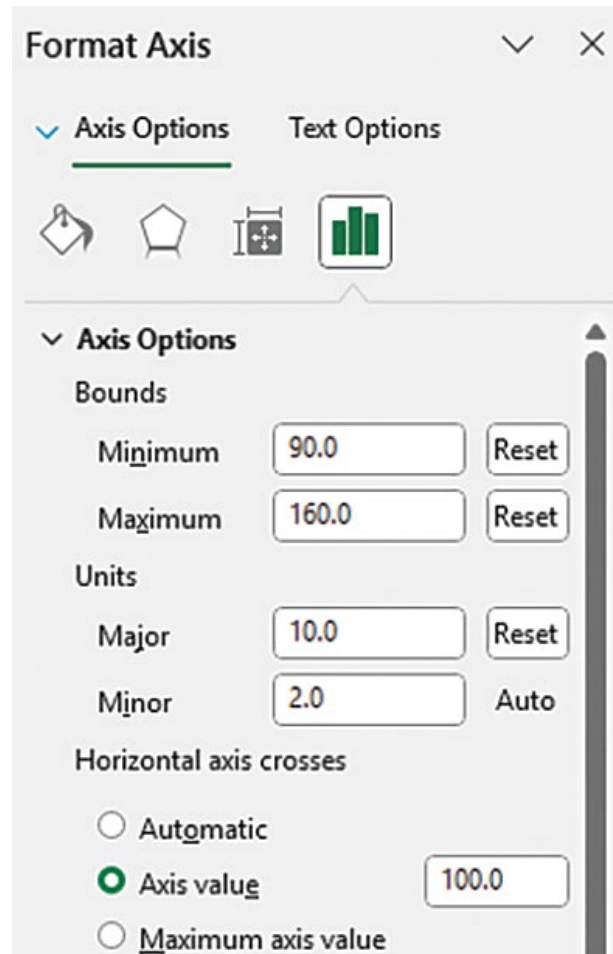


Figure 8.7 Formatting the horizontal axis to cross at 100 requires formatting the vertical axis.

Start Date

```
CASE [President]
WHEN "Biden" THEN DATE("1/21/2021")
WHEN "Trump" THEN DATE("1/23/2017")
END
```

Figure 8.8 A CASE statement.

Months Since Start

```
DATEDIFF("month", [Start Date], [Date])
```

Figure 8.9 Using the DATEDIFF command.

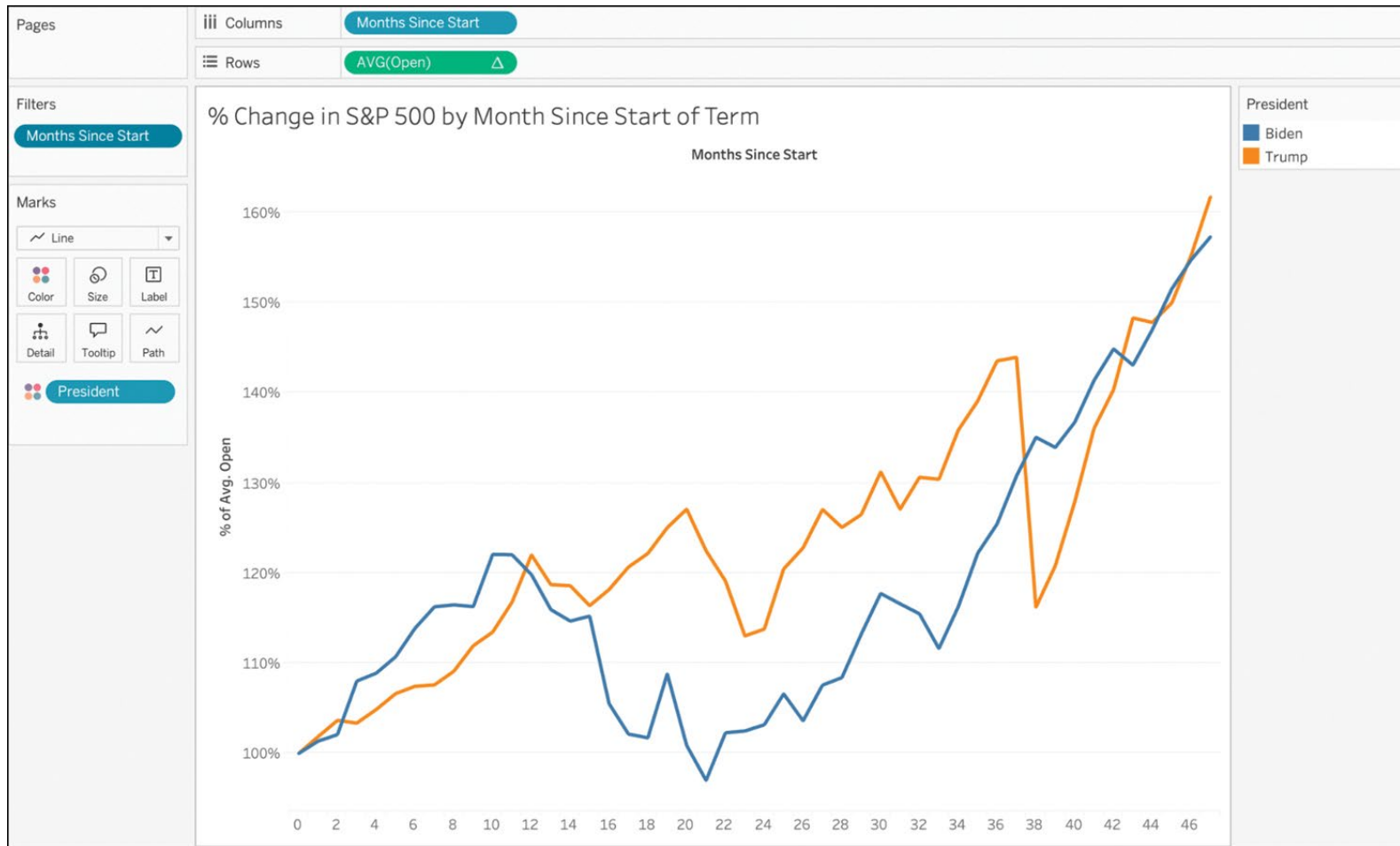


Figure 8.10 A quick version created using a table calculation.

Table Calculation

% of Avg. Open

×

Calculation Type

Percent From

Compute Using

Table (across)

Cell

Specific Dimensions

☒ Months Since Start

☐ President

At the level

Relative to

First

☒ Show calculation assistance

Figure 8.11 Table calculation menu options.

Indexed Open

Results are computed along Table (across).

$100 * (\text{AVG}([Open]) / \text{LOOKUP}(\text{AVG}([Open]), \text{FIRST()}))$

Figure 8.12 Index calculation.



Figure 8.13 The final, indexed version of the target graphic.

S&P 500 Performance - Biden vs Trump (1st Term)

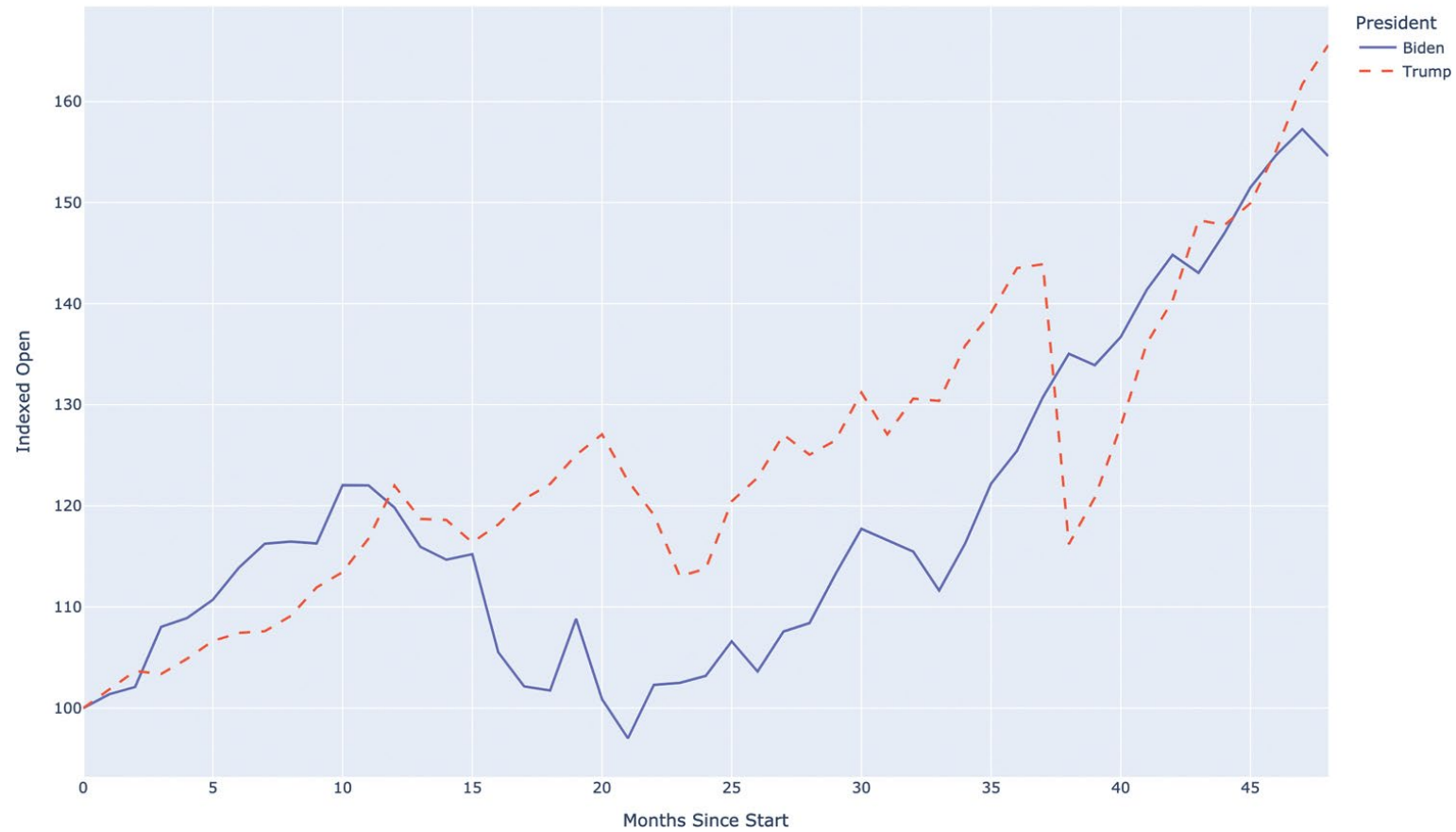
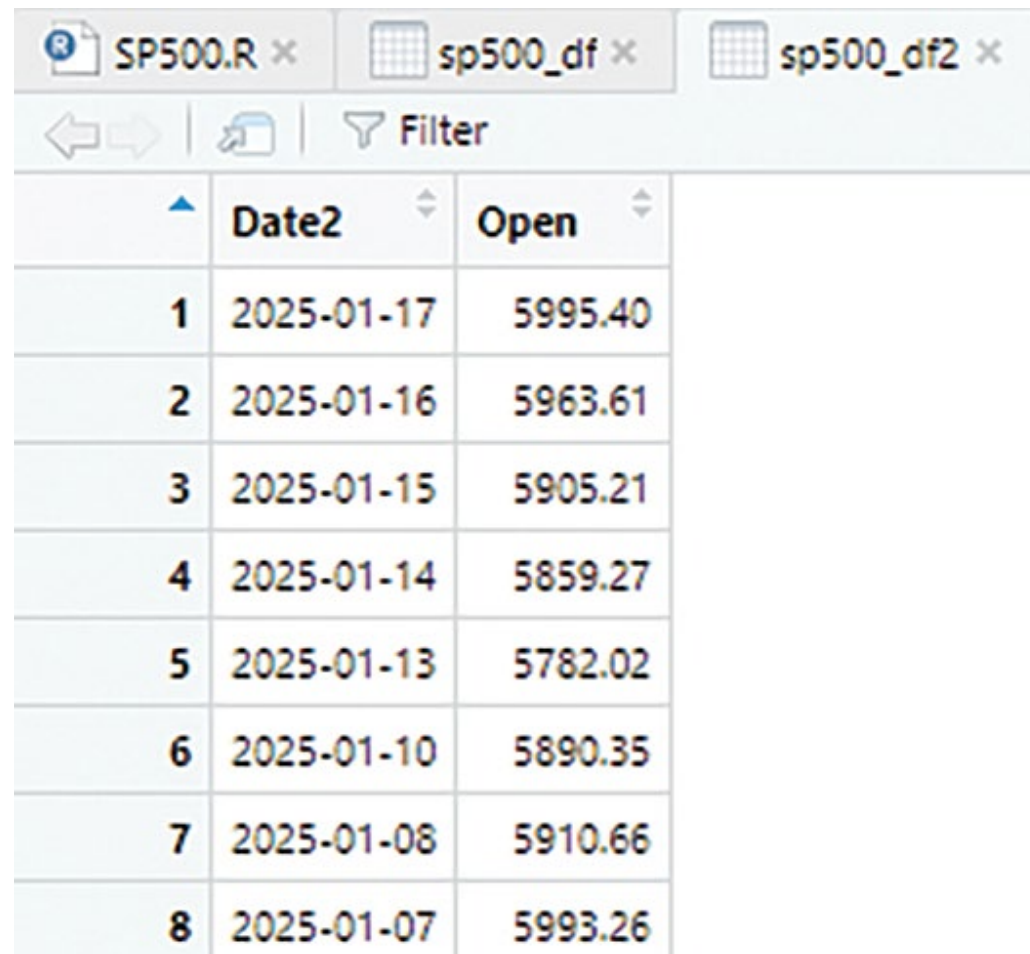


Figure 8.14 S&P 500 Index performance for the first Trump presidency and the Biden presidency.

SP500.R x sp500_df x sp500_df2 x sp500 x						
← → 📄 🔍 Filter						
	Date	Close.Last	Open	High	Low	
1	2/28/2025	5954.50	5856.74	5959.40	5837.66	
2	2/27/2025	5861.57	5981.88	5993.69	5858.78	
3	2/26/2025	5956.06	5970.87	6009.82	5932.69	
4	2/25/2025	5955.25	5982.73	5992.65	5908.49	
5	2/24/2025	5982.25	6026.60	6042.65	5977.82	

Figure 8.15 S&P 500 Index historical data.



	Date2	Open
1	2025-01-17	5995.40
2	2025-01-16	5963.61
3	2025-01-15	5905.21
4	2025-01-14	5859.27
5	2025-01-13	5782.02
6	2025-01-10	5890.35
7	2025-01-08	5910.66
8	2025-01-07	5993.26

Figure 8.16 S&P 500 Index historical data after wrangling it a bit.

SP500.R × sp500_df × sp500_df2 × sp500 >				
← → 📄 🔍 Filter				
	Date2	Open	President	Start
1	2025-01-17	5995.40	Biden	2021-01-21
2	2025-01-16	5963.61	Biden	2021-01-21
3	2025-01-15	5905.21	Biden	2021-01-21
4	2025-01-14	5859.27	Biden	2021-01-21
5	2025-01-13	5782.02	Biden	2021-01-21
6	2025-01-10	5890.35	Biden	2021-01-21
7	2025-01-08	5910.66	Biden	2021-01-21

Figure 8.17 S&P 500 Index historical data after adding in President and Start variables.

The figure consists of two screenshots of an RStudio interface, showing the S&P 500 Index historical data. The top screenshot displays the 'sp500' data frame with columns: Date2, Open, President, Start, and Months. The bottom screenshot displays the same data frame with an additional 'Months' column added, containing the value 47 for all rows.

	Date2	Open	President	Start	Months
1	2025-01-17	5995.40	Biden	2021-01-21	2021-01-21 UTC--2025-01-17 UTC
2	2025-01-16	5963.61	Biden	2021-01-21	2021-01-21 UTC--2025-01-16 UTC
3	2025-01-15	5905.21	Biden	2021-01-21	2021-01-21 UTC--2025-01-15 UTC
4	2025-01-14	5859.27	Biden	2021-01-21	2021-01-21 UTC--2025-01-14 UTC
5	2025-01-13	5782.02	Biden	2021-01-21	2021-01-21 UTC--2025-01-13 UTC
6	2025-01-10	5800.35	Biden	2021-01-21	2021-01-21 UTC--2025-01-10 UTC

	Date2	Open	President	Start	Months
1	2025-01-17	5995.40	Biden	2021-01-21	47
2	2025-01-16	5963.61	Biden	2021-01-21	47
3	2025-01-15	5905.21	Biden	2021-01-21	47
4	2025-01-14	5859.27	Biden	2021-01-21	47
5	2025-01-13	5782.02	Biden	2021-01-21	47
6	2025-01-10	5800.35	Biden	2021-01-21	47

Figure 8.18 S&P 500 Index historical data after adding in the Months variable.

SP500.R × sp500_df × sp500_df2 × sp500 ×				
← → 📄 🔍 Filter				
	Months	President	sp_avgs	
1	0	Biden	3862.488	
2	0	Trump	2305.173	
3	1	Biden	3887.758	
4	1	Trump	2372.228	
5	2	Biden	4042.117	
6	2	Trump	2350.791	
7	3	Biden	4163.541	
8	3	Trump	2388.226	

Figure 8.19 Average opening prices of S&P 500 Index historical data after grouping by month and president.

SP500.R × sp500_df × sp500_df2 × sp500 ×					
← → 📄 🔍 Filter					
	Months	President	sp_avgs	Index	
1	0	Biden	3862.488	100.00000	
2	1	Biden	3887.758	100.65425	
3	2	Biden	4042.117	104.65059	
4	3	Biden	4163.541	107.79428	
5	4	Biden	4215.752	109.14602	
6	5	Biden	4307.822	111.52971	
7	6	Biden	4414.742	114.29788	
8	7	Biden	4492.851	116.32013	
9	8	Biden	4305.300	113.70708	

Figure 8.20 Index of average opening prices of S&P 500 Index historical data.

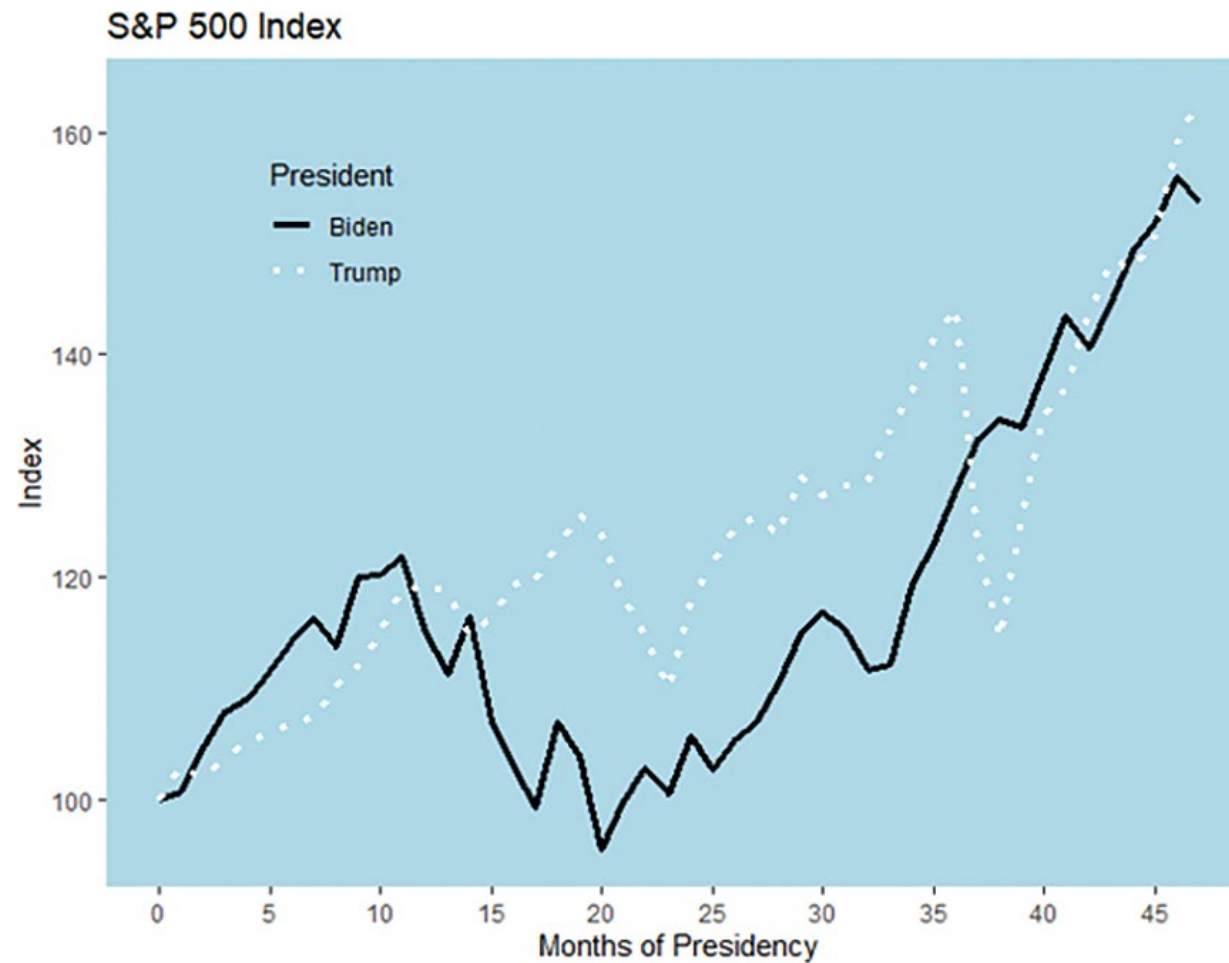


Figure 8.21 The finished graphic showing the change in the S&P 500 Index over each president's term.

	A	B	C	D	E	F	G	H	I
1									
2									
3	Average of Open Column Labels ▾								
4	Row Labels ▾	Biden	Trump	Grand Total				Biden	Trump
5	0	3862.5	2305.2	3065.7			0	100.0	100.0
6	1	3887.8	2372.2	3130.0			1	=B6/B\$5*100	
7	2	4042.1	2350.8	3196.5			2	104.7	102.0
8	3	4163.5	2388.2	3296.5			3	107.8	103.6
9	4	4215.8	2427.2	3278.0			4	109.1	105.3

Figure 8.22 The data in wide form.

SP500.R* x sp500_df x sp500_df2 x sp500 x					
← → 📄 🔍 Filter					
	▲	Months	President	sp_avgs	Index
1		0	Biden	3862.488	100.00000
2		1	Biden	3887.758	100.65425
3		2	Biden	4042.117	104.65059
4		3	Biden	4163.541	107.79428
5		4	Biden	4215.753	109.14603

Figure 8.23 The data in long form.