

Bank_R_script

2025-06-10

To start we read in the data and put the tidyverse package in the library. Remember you must install this package once and only once on your machine.

```
Bank_data <- read.csv("Bank_data_CSV.csv")
View(Bank_data)
```

```
# install.packages("tidyverse")
library(tidyverse)
```

To make the first chart we use basic ggplot syntax, noting that geom_bar defaults to count. To get color we can map the Branch variable to fill.

```
ggplot(data = Bank_data) +
  geom_bar(aes(x = Branch, fill = Branch))
```

Trying to change the aggregation from count to sum using a statistical transformation unfortunately gave incorrect sums.

```
ggplot(data = Bank_data) +
  geom_bar(aes(x = Branch, y = Amount, fill=Branch), stat = "sum")
```

So we manually create the correct sums using the group_by and summarize functions. These can also be coded using pipe, %>%, notation.

```
branches <- group_by(Bank_data, Branch)
branches_sum <- summarize(branches, total = sum(Amount))

View(branches_sum)

#We could use a PIPE %>% to avoid the interim steps
#Read %>% as "then"
branches_sum <- Bank_data %>%
  group_by(Branch) %>%
  summarize(total = sum(Amount))
```

Now plot the bars using stat = identity because we want to plot the exact sums we just computed.

```
ggplot(branches_sum) +
  geom_bar(aes(x=Branch, y=total, fill = Branch), stat = "identity")
```

Grouping by branch and account type is similar, and plotting requires us to dodge the position, else we get stacked bars.

```
br_acct <- group_by(Bank_data, Branch, AcctType)
br_acct_sums <- summarize(br_acct, total = sum(Amount))
View(br_acct_sums)

ggplot(br_acct_sums) +
  geom_bar(aes(x=AcctType, y=total, fill = Branch),
    stat = "identity", position = "dodge")
```

The `stat_summary` function is another way to do this.

```
ggplot(Bank_data) +
  stat_summary(aes(x = Branch, y = Amount), fun.min = min,
    fun.max = max, fun = median)

ggplot(Bank_data, aes(x=Branch, y=Amount, fill=AcctType)) +
  stat_summary(fun = "sum", geom = "bar", position = "dodge")

ggplot(Bank_data, aes(x=AcctType, y=Amount, fill=Branch)) +
  stat_summary(fun = "sum", geom = "bar", position = "dodge")
```

We can recreate the original barplot of amount by branch and also format the values as dollars using the `scales` package. Note that `geom_col` defaults to `stat = identity` and also works.

```
# install.packages("scales")
library(scales)

ggplot(Bank_data, aes(x=Branch, y=Amount, fill=Branch)) +
  stat_summary(fun = "sum", geom = "bar") +
  scale_y_continuous(labels = label_dollar())

ggplot(Bank_data, aes(x = Branch, y = Amount, fill=Branch)) +
  geom_col() +
  scale_y_continuous(labels = label_dollar())
```