

SP500_R_script

2025-07-07

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

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

library(tidyverse)
```

First read in the data and get a date variable in correct format. Check it works with logical operators.

```
sp500_df <- read.csv("S_P 500_clean.csv")
View(sp500_df)

#OK need to deal with the dates
class(sp500_df$Date)

sp500_df$Date2 <- as.Date(sp500_df$Date, tryFormats = c("%m/%d/%Y"))
class(sp500_df$Date2)

# Test to see how we can use dates with logical operators
sp500_df$Date2[1] == "2025-02-28"
```

Next filter out the rows we don't want, choose just two columns to start, and then create the president and start columns.

```
# filter rows so we have only rows in the two terms
sp500_df2 <- filter(sp500_df, Date2 <= "2025-01-17" & Date2 >= "2017-01-23")
View(sp500_df2)

# select just two columns needed
sp500_df2 <- select(sp500_df2, Date2, Open)

# Create new column with names and new with start dates just like Rich/Poor in Countries
sp500_df2$President <- rep("Biden", length = nrow(sp500_df2))
sp500_df2$President[sp500_df2$Date2 < "2021-01-21"] = "Trump"

sp500_df2$Start <- rep("2021-01-21", length = nrow(sp500_df2))
sp500_df2$Start[sp500_df2$Date2 < "2021-01-21"] = "2017-01-23"
```

Next create the months variable and then use group_by and summarize.

```
# Create new column with months
?difftime
?interval
```

```

# Sometimes you just have to google some code...
sp500_df2$Months <- interval(ymd(sp500_df2$Start), ymd(sp500_df2$Date2))
sp500_df2$Months = sp500_df2$Months %/% months(1)

# Now average the open for each month and President
# Using pipe notation just like in Banksy
# This will recreate the two way PivotTable from Excel and melt it so it's tidy
# Rows are the months 0-47 and columns are Presidents
# Values are averages of the Open by month

sp500 <- sp500_df2 %>%
  group_by(Months, President) %>%
  summarize(sp_avgs = mean(Open))
View(sp500)

```

Almost there! Just have to create the new index column with start values set to 100.

```

# Arrange rows by President
sp500 <- arrange(sp500, President)

# Create new variable for the Index set to 100
sp500$Index <- rep(100, nrow(sp500))

# Now compute the Index values for each President by dividing by the first value for each term

sp500$Index[sp500$President == "Biden"] =
  sp500$sp_avgs[sp500$President == "Biden"]/sp500$sp_avgs[1]*100
sp500$Index[sp500$President == "Trump"] =
  sp500$sp_avgs[sp500$President == "Trump"]/sp500$sp_avgs[49]*100

```

Perfect! Now we can make our graphic.

```

# Phew! Now let's try to get the graphic....
ggplot(sp500, aes(x = Months, y = Index, group = President)) +
  geom_line(linewidth = 1.2, aes(linetype = President, color = President)) +
  scale_color_manual(values = c("black", "white")) +
  scale_linetype_manual(values = c(1,3)) +
  scale_x_continuous(breaks = seq(0,45, by = 5)) +
  scale_y_continuous(breaks = c(100, 120, 140, 160)) +
  ggtitle("S&P 500 Index") +
  xlab("Months of Presidency") +
  theme(panel.background = element_rect(fill = "lightblue"),
        panel.grid = element_line(color = "lightblue"),
        legend.position = c(0.2,0.8),
        legend.background = element_rect(fill = "lightblue"),
        legend.key = element_rect(fill = "lightblue"))

```

The End!