Maxim Dokukin

017025164

MATH 167R Sec 01

Prof. Gao

05/05/2024

Final Project Report

Introduction and background

The debate over whether government officials should participate in stock trading persists. Investment is a key wealth-building strategy, but it could lead to conflicts of interest. Officials are expected to prioritize the public good, which may not always align with profit-maximizing motives. There's also the risk that politicians may exploit non-public information, leading to insider trading.

This project's goal is to examine the trading activities of politicians and determine if their transactions outperform the general market. The S&P 500 index serves as the benchmark for average market returns. The analysis includes several subquestions:

1. Is there a correlation between the reporting gap and trade profitability?
2. Is there a difference in the frequency or profitability of trades made by Democrats versus Republicans?
3. Who is the most successful investor in the government?
4. Are there potential cases of insider trading?
5. Are certain states more active in terms of the number of transactions?
6. Which ticker symbols are most frequently involved in politicians' transactions?

Data for this study was scraped from https://www.capitoltrades.com/trades, yielding a dataset of 39,114 transactions spanning from 2021-03-14 to 2024-05-01. Scraping was the chosen method due to the lack of existing comprehensive datasets. Further processing of this data set was performed to clean it and update the transaction prices, as some of the reported prices were different from the market prices on that date.

Additional processing used the data from the transactions dataset to generate trades dataset. Trades were formed using a cycle algorithm that grouped all transactions by politician names and tickers. For each politician and ticker, the algorithm generated trade cycles. A trade cycle starts with a series of stock purchases and ends with a series of sales. The average purchase and sell prices are calculated based on the mean of all buy and sell prices in the cycle. The average days between transactions are computed as the difference between the mean sell and purchase dates. From these two factors, the trade's annualized percentage profit is calculated.

Additionally, more data was acquired, to estimate the stock performance before and after transaction dates. For each transaction, prices were fetched before and after the transaction date using day offsets of [-90, -60, -40, -20, -10, 10, 20, 40, 60, 90] from the transaction date. Based on this data, the average performance was estimated before and after each purchase.

As of today, there's been a lot of discussion about how well certain politicians do on the stock market, often beating the average investor by a wide margin. Studies, like one featured in the Journal of Financial and Quantitative Analysis, suggest that politicians' investment portfolios often outperform market. This raises concerns on whether they have efficient investment strategies or whether they are exercising insider trading. Because of these suspicions, there's a growing demand for tougher rules and more transparency around how politicians trade stocks.

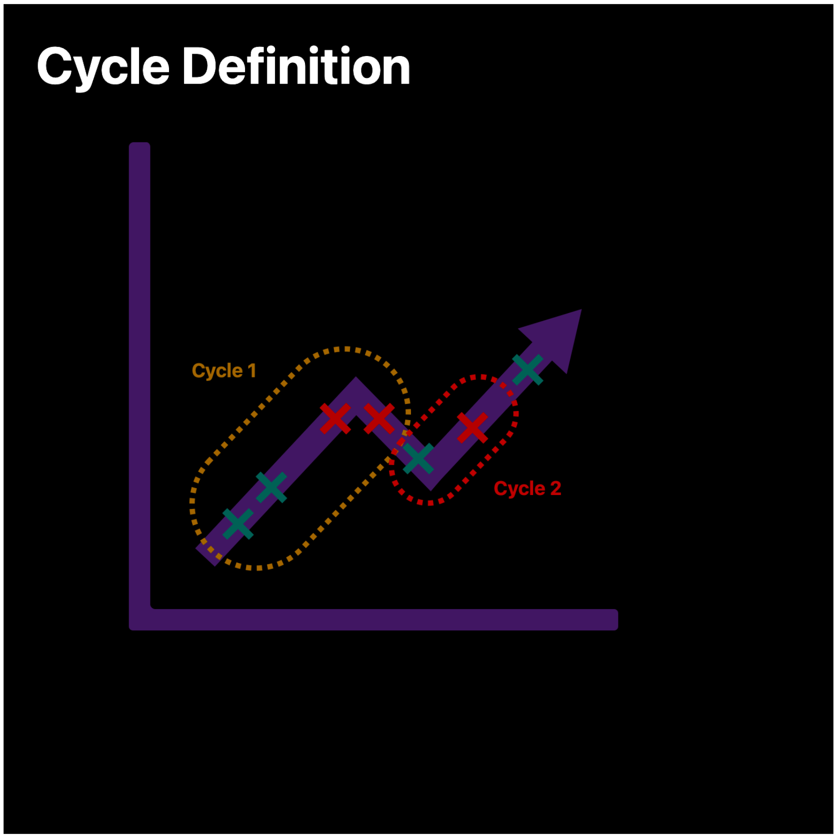
Methodology 1: Transaction Cycle Analysis

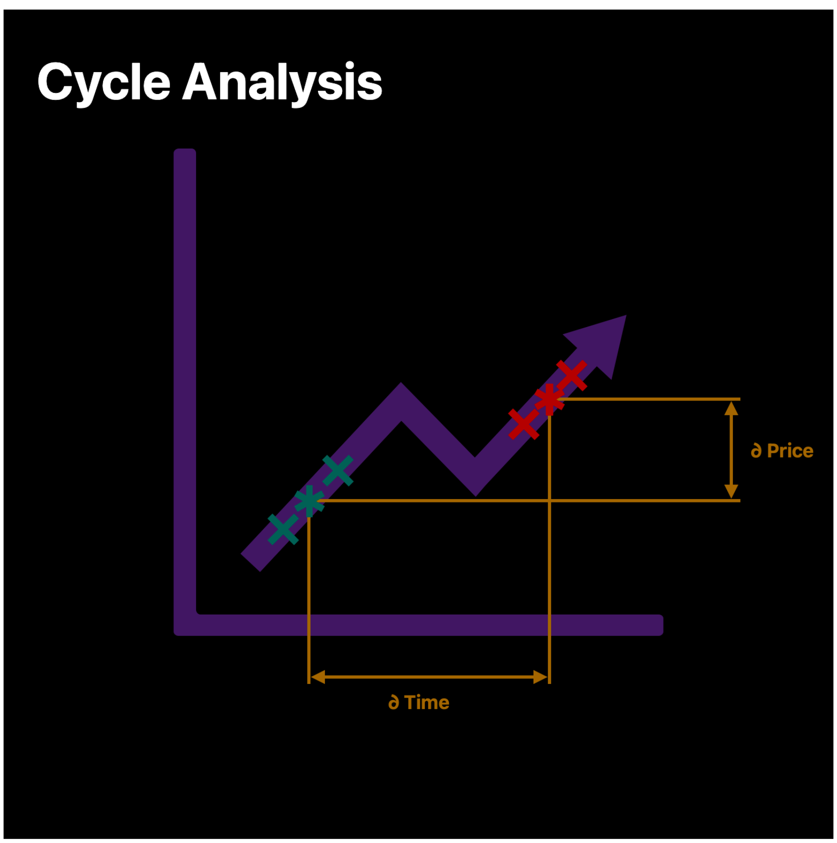
This methodology, implemented as a Python function evaluates the trading behavior of politicians by analyzing patterns and performance across their stock transactions. The function begins by grouping the data by 'Politician Name' and 'Ticker', which isolates the transactions for each individual politician and stock combination. These grouped transactions are then sorted by date to ensure correct processing order.

As the function goes through each sorted group, it identifies and processes cycles of buying followed by selling. Each transaction type ('BUY' or 'SELL') is processed according to its sequence within these cycles, with the function recording relevant details such as transaction prices, IDs, amounts, and dates. The transition from buying to selling phases marks the start and end of investment cycles.

Following the completion of a sell phase, the function computes several key performance metrics. These include average buy and sell prices, the total percentage profit, and the annualized profit, calculated based on the duration between the first purchase and the last sale within the cycle. Additionally, the function calculates the duration of the investment as the number of days between the initial buy and final sell.

The results from each processed cycle are compiled into a structured format, incorporating data points such as the names of the politicians, transaction IDs, total amounts involved in the transactions, and the calculated financial metrics. The final output is a DataFrame that consolidates all cycle results, providing a detailed and quantifiable overview of the trading performance for each politician-stock pair.


Cycle Definition
Cycle 1
Cycle 2


Cycle Analysis
∂ Price
∂ Time

Methodology 2: Evaluating Stock Performance Pre- and Post-Transaction

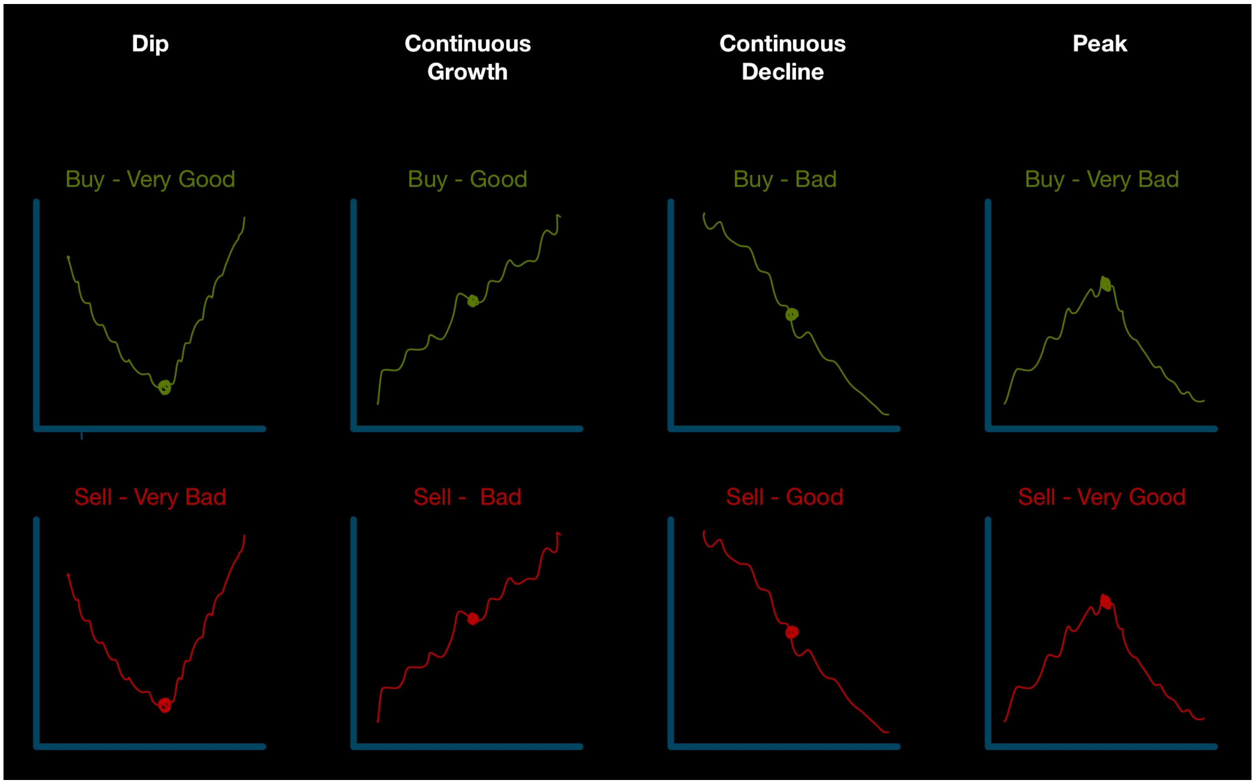
This methodology uses R to assess the stock performance around transaction dates, using the average annualized rates of change before and after purchases. The analysis starts by calculating percentage changes for stock prices at specific intervals: 90, 60, 40, 20, and 10 days, before a transaction and 10, 20, 40, 60, and 90 days after. These percentage changes are then annualized to express them as yearly rates, providing a clearer picture of short-term stock performance in an annualized context.

Further, the methodology uses the mutate function to create columns for each interval's annualized performance and then calculates the average performance across all intervals before and after transactions (pd_avg and fd_avg).

The data is further processed by adjusting transaction types, where 'RECEIVE' is reclassified as 'BUY', and excluding 'EXCHANGE' transactions, and transactions by parties labeled as 'Other'.

Additionally, the market_figure function categorizes stock performance trends based on these averages, identifying whether stocks were consistently up, down, peaked, or dipped. The eval_success function assesses the effectiveness of 'BUY' or 'SELL' decisions based on these trends, categorizing them as 'Good Decision', 'Bad Decision', 'Very Good Decision', or 'Very Bad Decision' depending on the market context at the time of the transaction.

This methodology not only measures stock performance dynamics related to specific transactions but also evaluates the decision-making effectiveness in a financial context, offering a comprehensive view of investment strategy outcomes.


Dip
Continuous Growth
Continuous Decline
Peak
Buy - Very Good
Buy - Good
Buy - Bad
Buy - Very Bad
Sell - Very Bad
Sell - Bad
Sell - Good
Sell - Very Good

Data Dictionary

transactions_cleaned.csv

Variable Name	Data Type	Description
Politician Name	String	Name of the politician involved in the transaction
Party	String	Political party affiliation of the politician
Chamber	String	Legislative chamber to which the politician belongs
State	String	State from which the politician hails or represents
Issuer Name	String	Name of the company involved in the transaction
Ticker	String	Stock ticker symbol of the company
Publication Date	Date	Date when the transaction was made public
Transaction Date	Date	Date when the transaction occurred
Reporting Gap	Integer	Number of days between the transaction date and the publication date
Owner	String	The holder of the stock
Transaction Type	String	Type of transaction
Value Range	String	Range of the value of the transaction
Price	Float	Price of the stock at the time of transaction
Transaction ID	String	Unique identifier for the transaction
Price_0_days	Float	Price of the stock at the market open on the transaction day
Approx Transaction Amount	Integer	Approximate dollar amount of the transaction
Approx Share Count	Integer	Approximate number of shares transacted

trades.csv

Variable Name	Data Type	Description
Politician Name	String	Name of the politician involved in the transaction
Party	String	Political party affiliation of the politician
Chamber	String	Legislative chamber to which the politician belongs
State	String	State from which the politician hails or represents
Ticker	String	Stock ticker symbol of the company involved in the transaction
Report Gap	Integer	Number of days between the last transaction date and the reporting date
Annualized Percentage Profit	Float	Annualized profit percentage from the trade

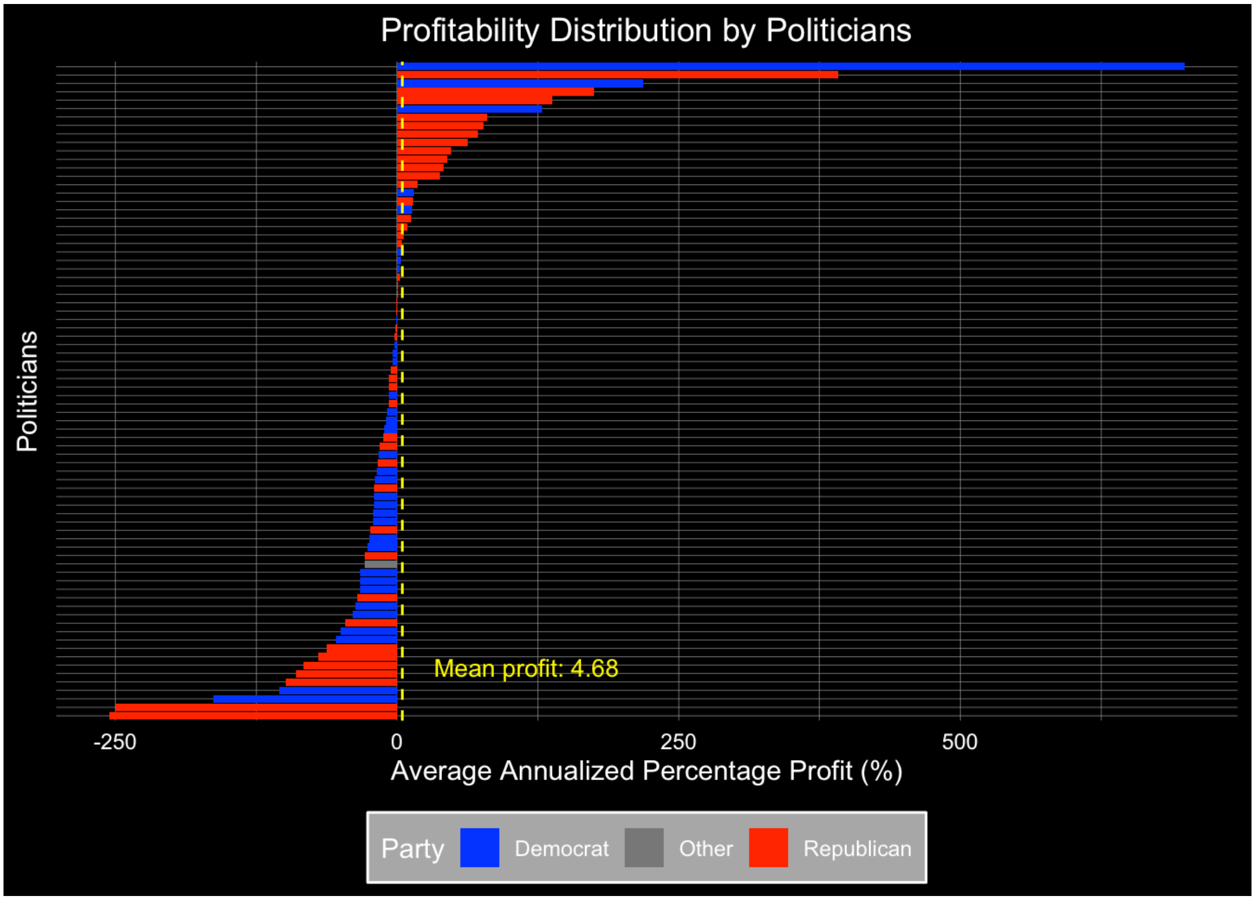
Variable Name	Data Type	Description
Purchase Transaction IDs	String	Comma-separated list of transaction IDs for purchases
Sell Transaction IDs	String	Comma-separated list of transaction IDs for sales
Total Purchase Amount	Float	Total dollar amount of all purchase transactions
Total Sell Amount	Float	Total dollar amount of all sell transactions
Days Between	Integer	Number of days between midpoints of purchase and sell transactions

more_prices.csv

Variable Name	Data Type	Description
Politician Name	String	Name of the politician involved in the transaction
Party	String	Political party affiliation of the politician
Chamber	String	Legislative chamber to which the politician belongs
State	String	State from which the politician hails or represents
Issuer Name	String	Name of the company involved in the transaction
Ticker	String	Stock ticker symbol of the company
Publication Date	Date	Date when the transaction was made public
Transaction Date	Date	Date when the transaction occurred
Reporting Gap	Integer	Number of days between the transaction date and the publication date
Owner	String	The holder of the stock
Transaction Type	String	Type of transaction
Value Range	String	Range of the value of the transaction
Price	Float	Price of the stock at the time of transaction
Transaction ID	String	Unique identifier for the transaction
Approx Transaction Amount	Float	Approximate dollar amount of the transaction
Approx Share Count	Float	Approximate number of shares involved in the transaction
Price_-90_days	Float	Price of the stock 90 days before the transaction
Price_-60_days	Float	Price of the stock 60 days before the transaction
Price_-40_days	Float	Price of the stock 40 days before the transaction
Price_-20_days	Float	Price of the stock 20 days before the transaction
Price_-10_days	Float	Price of the stock 10 days before the transaction

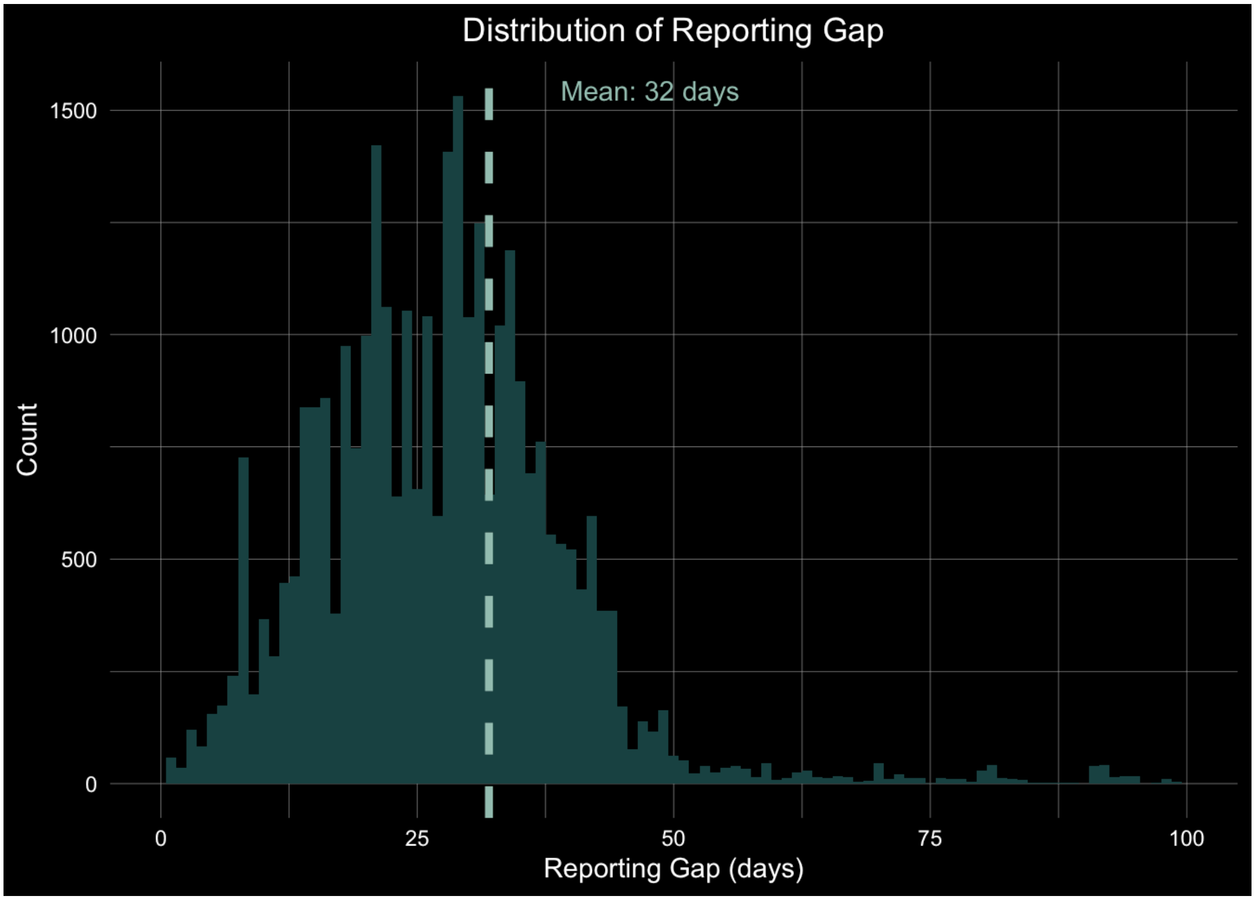
Variable Name	Data Type	Description
Price_0_days	Float	Price of the stock at market close on the transaction day
Price_10_days	Float	Price of the stock 10 days after the transaction
Price_20_days	Float	Price of the stock 20 days after the transaction
Price_40_days	Float	Price of the stock 40 days after the transaction
Price_60_days	Float	Price of the stock 60 days after the transaction
Price_90_days	Float	Price of the stock 90 days after the transaction

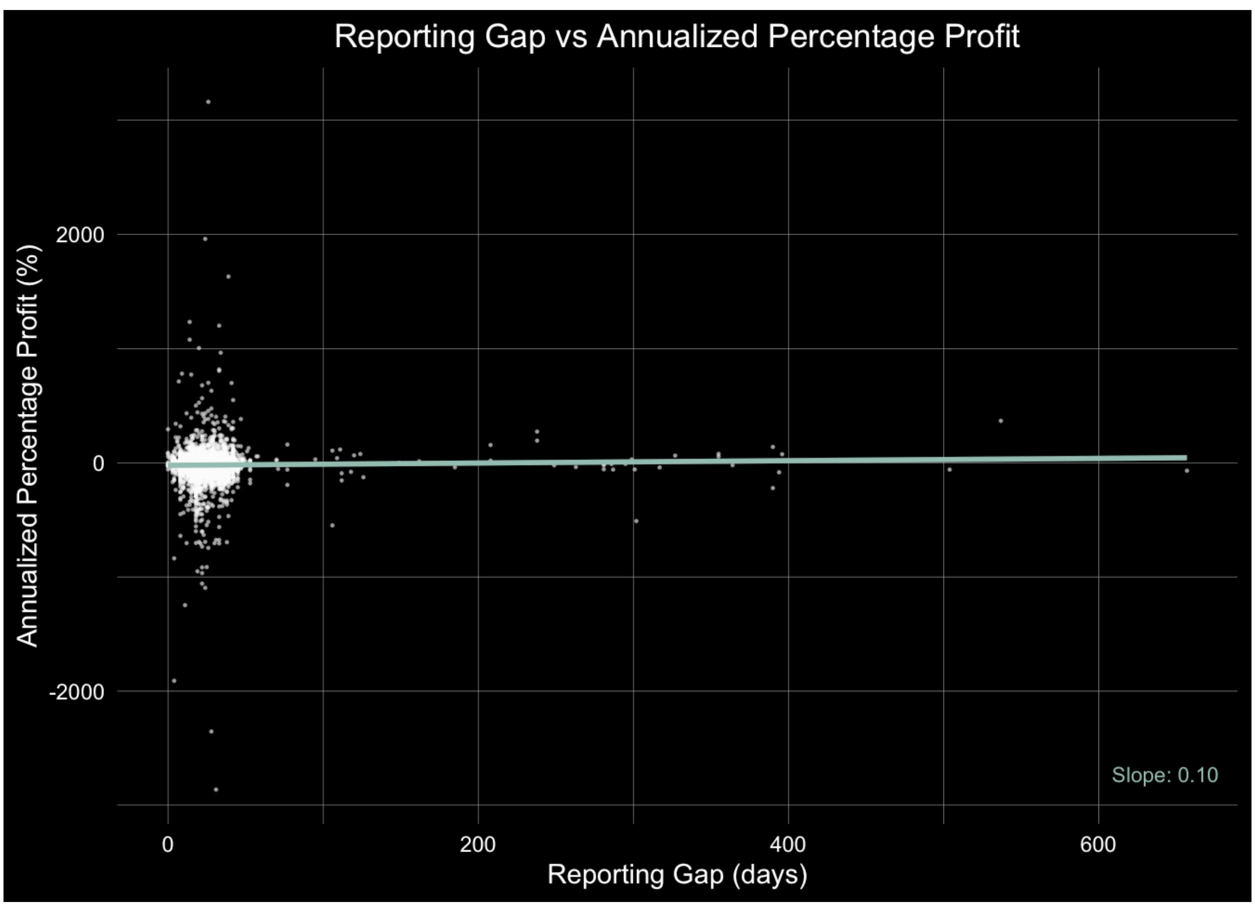
Q1: Does average politician outperform stock market?



The average annualized percentage profit yield from S&P500 was 6.8% in the 2021-2024 period. While the average annualized percentage profit for an average politician was 4.7% in the same time period. Therefore, the average politician did not outperform the stock market.

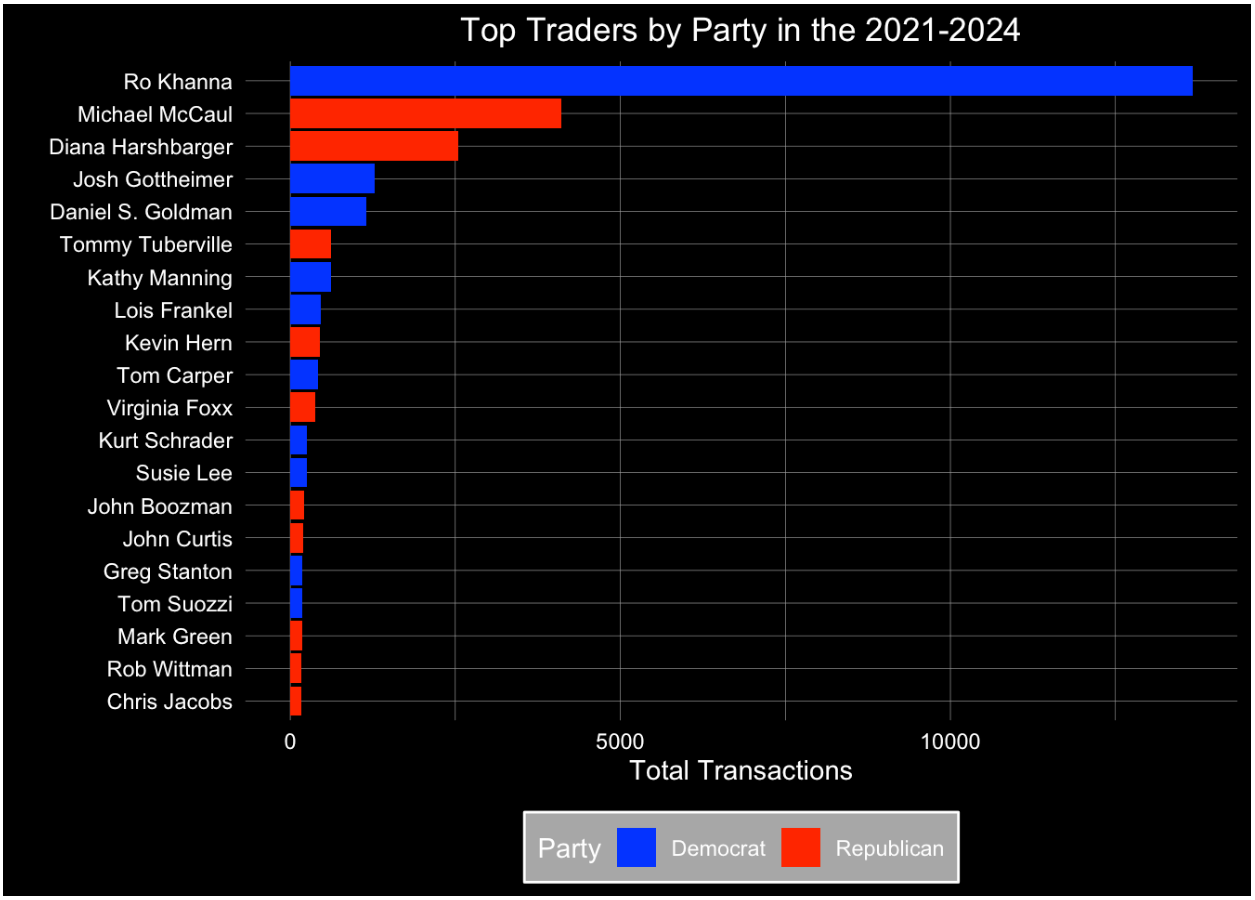
Is there a correlation between the Reporting Gap and trade profitability?

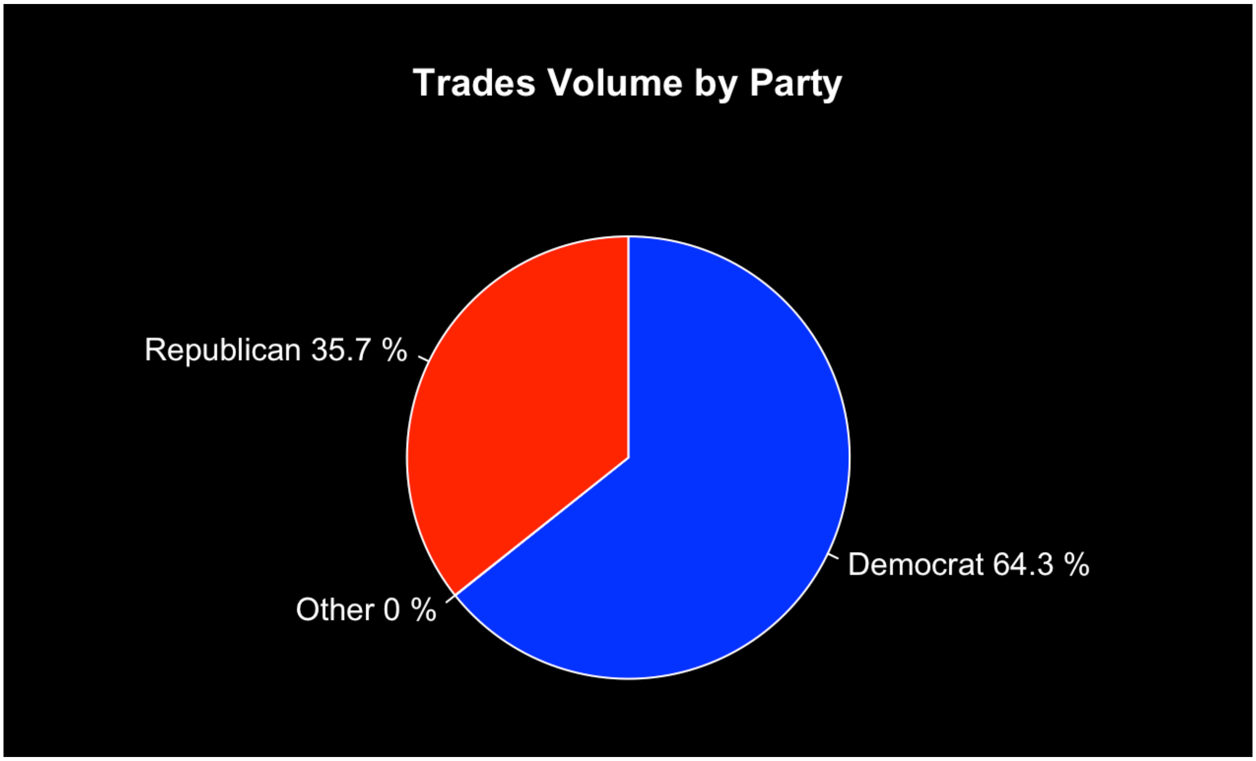


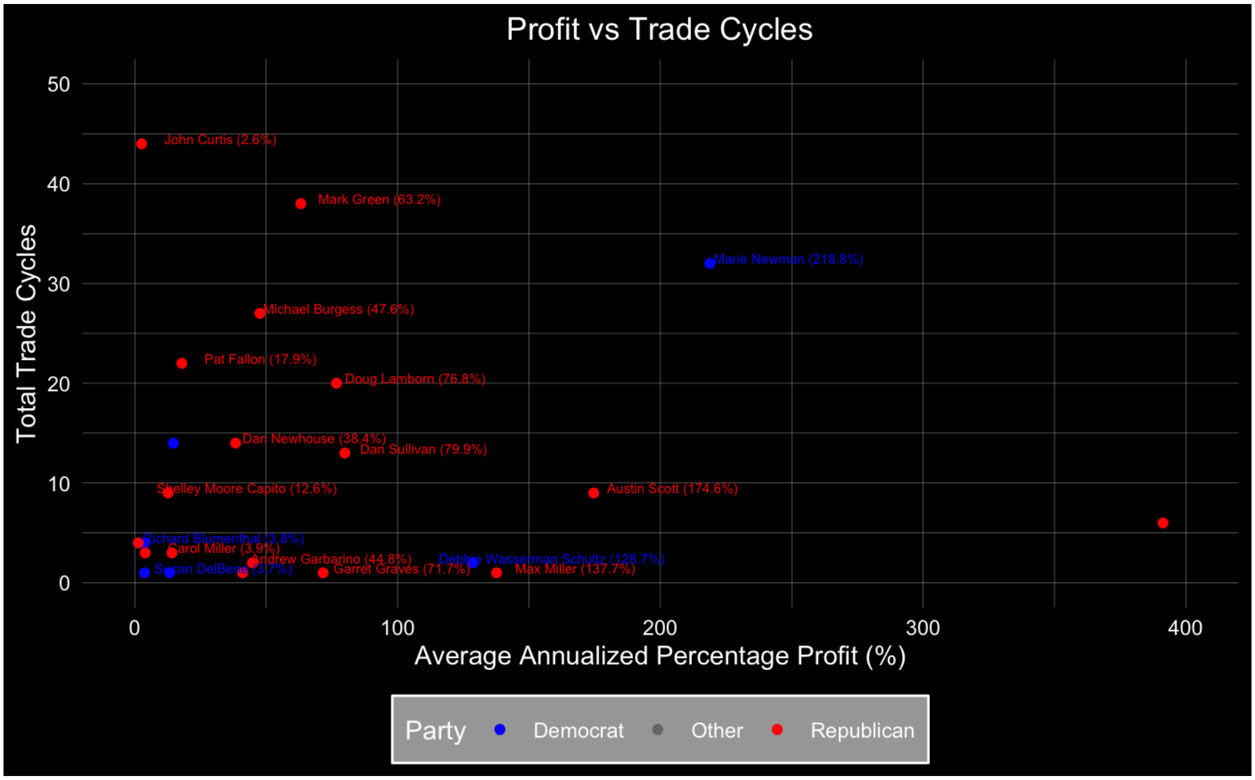


Linear regression model was used to calculate the slope coefficient between trades profitability and reporting gap. Data suggests that there is no correlation between trades profitability and reporting gap.

Q2: Is there a difference in the frequency or profitability of trades made by Democrats versus Republicans?

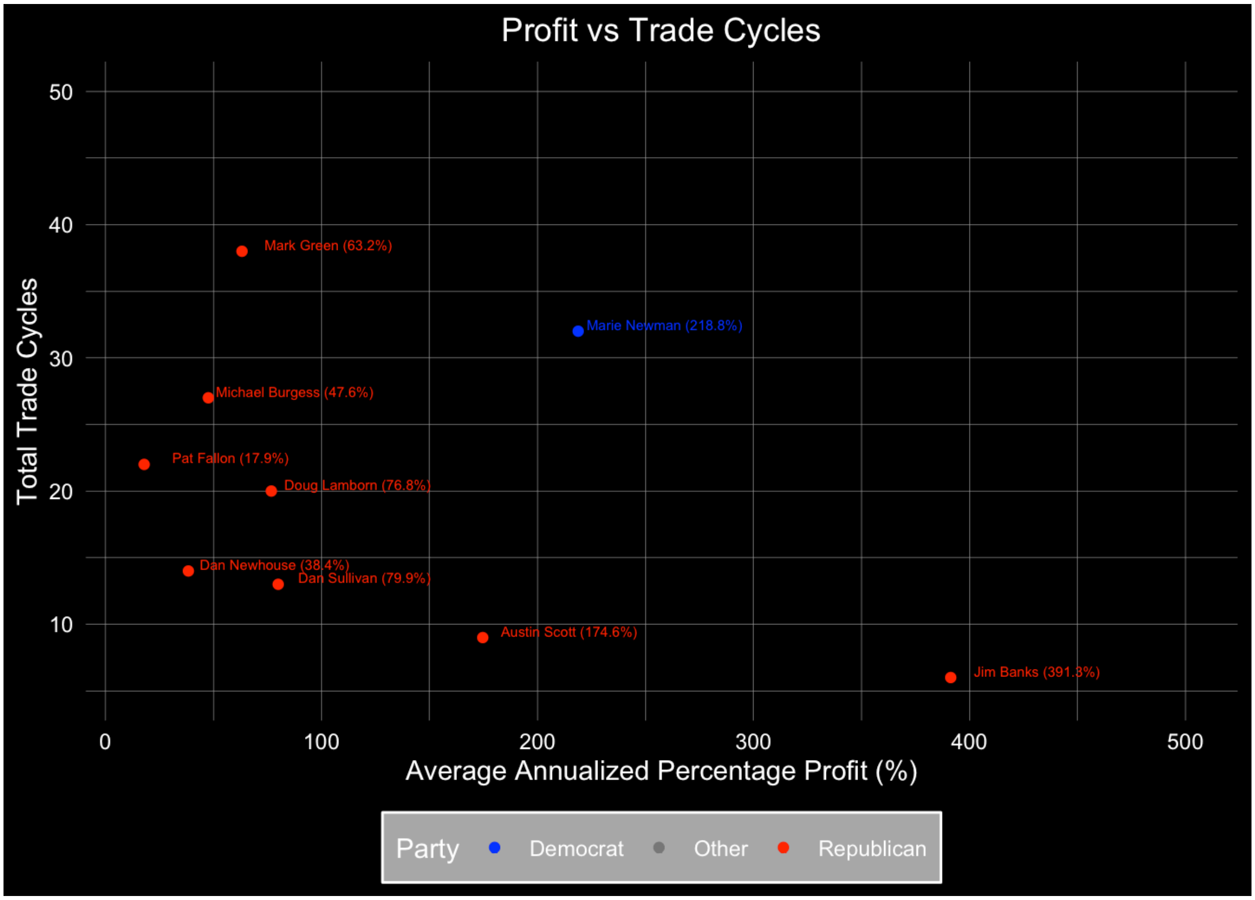






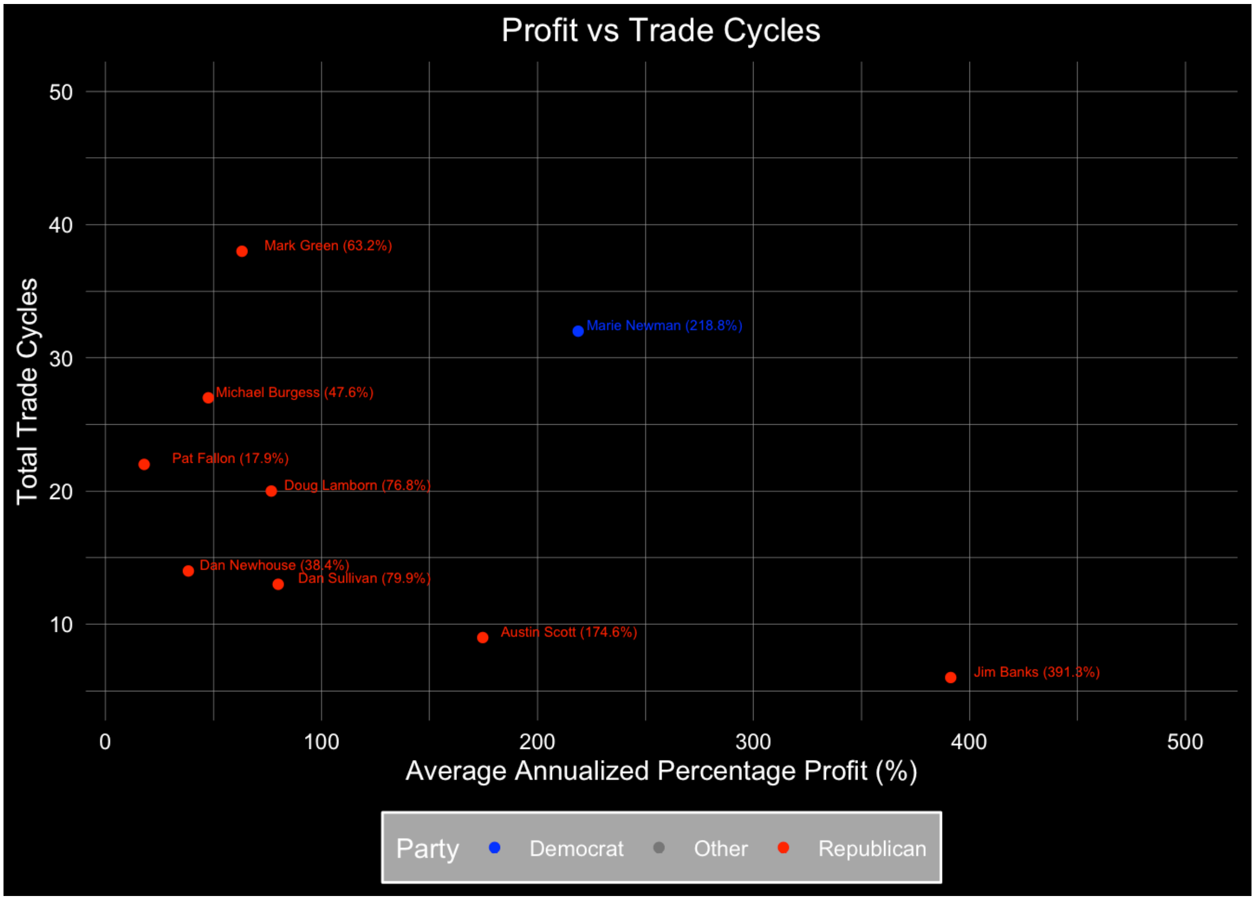
The data suggests that the majority (2/3) of trades are made by Democrats. If accounted for Democrat outlier Ro Khanna with 136668 transactions, the proportion flips. Yet, Republicans have more profitable trades.

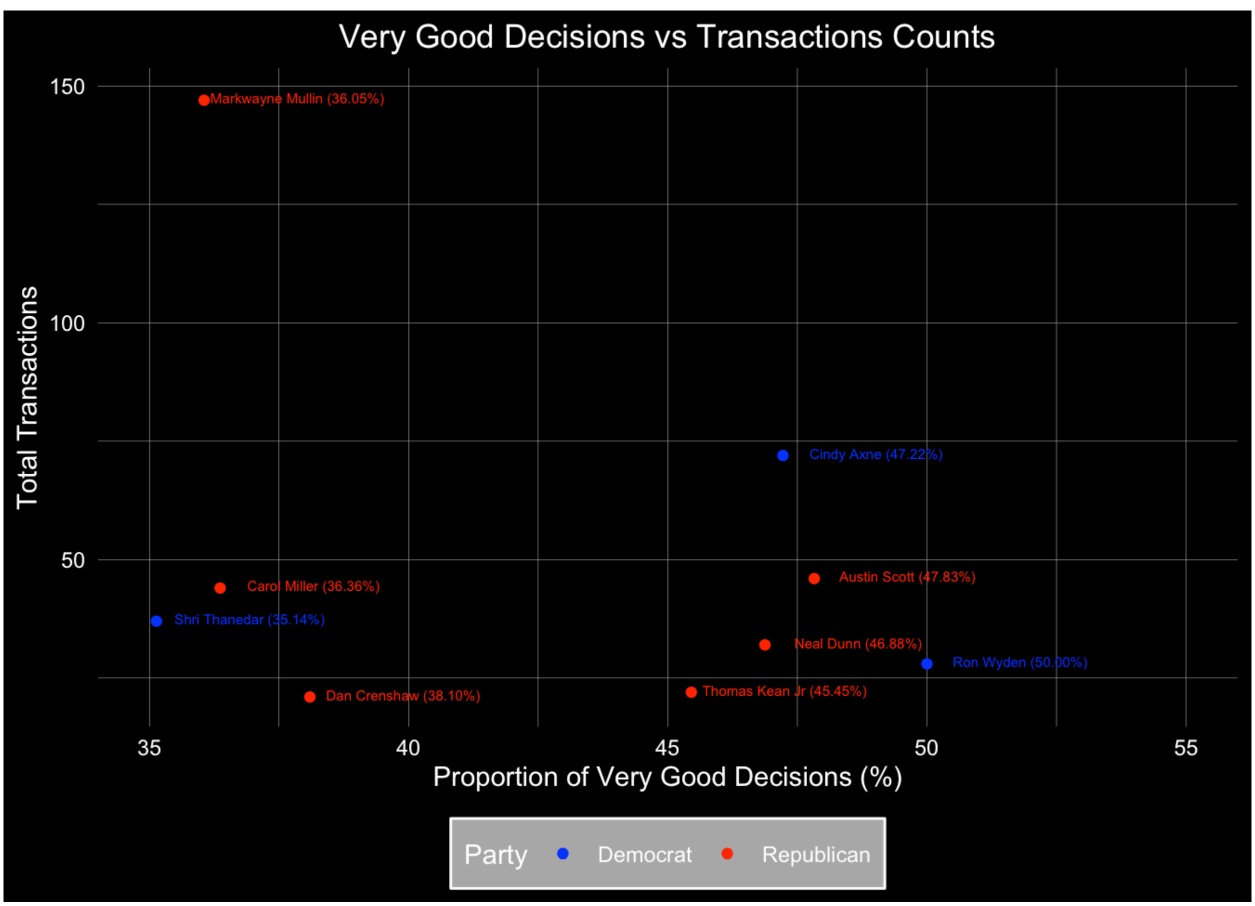
Q3: Who is the most successful investor in the government?



Marie Newman reached 219% annualized average percent profit, with 32 Trade Cycles and 120 transactions in total over the past 3 years. It is 32 times more profitable than the S&P500 index over the same period of time.

Q4: Are there potential cases of insider trading?





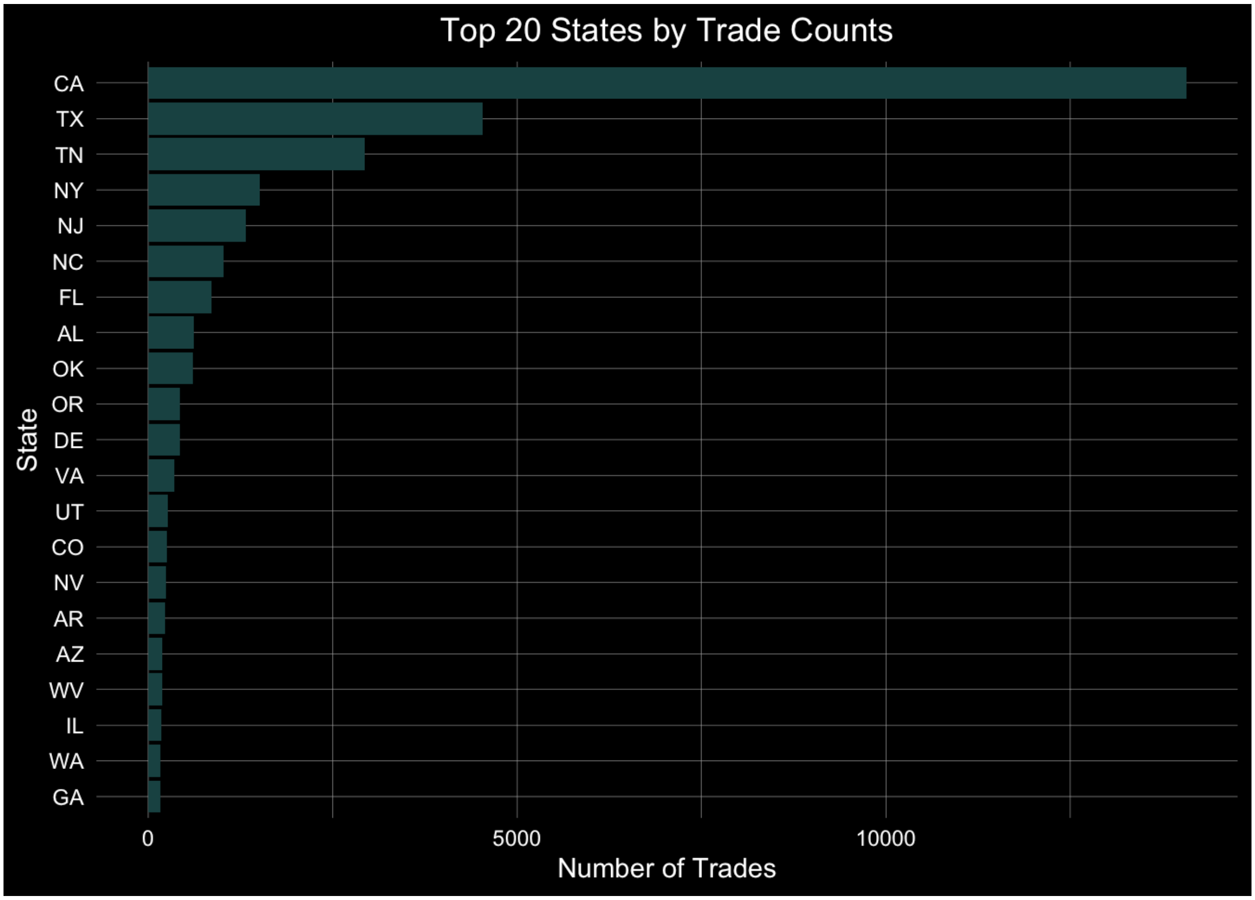
Politicians who made over 15% annualized average percent profit with more than 5 trades, or whose proportion of very good decisions exceeds 35% with more than 10 transactions, can be subjects for further investigation.

For statistical analysis, an assumption was made that an average trader would make about 25% of very good, 25% of good, 25% of bad, and 25% of very bad decisions. Furthermore, the distribution of the decisions assumed to follow binomial, with p=0.25.

binom.test() was used to classify unusually high proportions of very good decisions among the politicians.

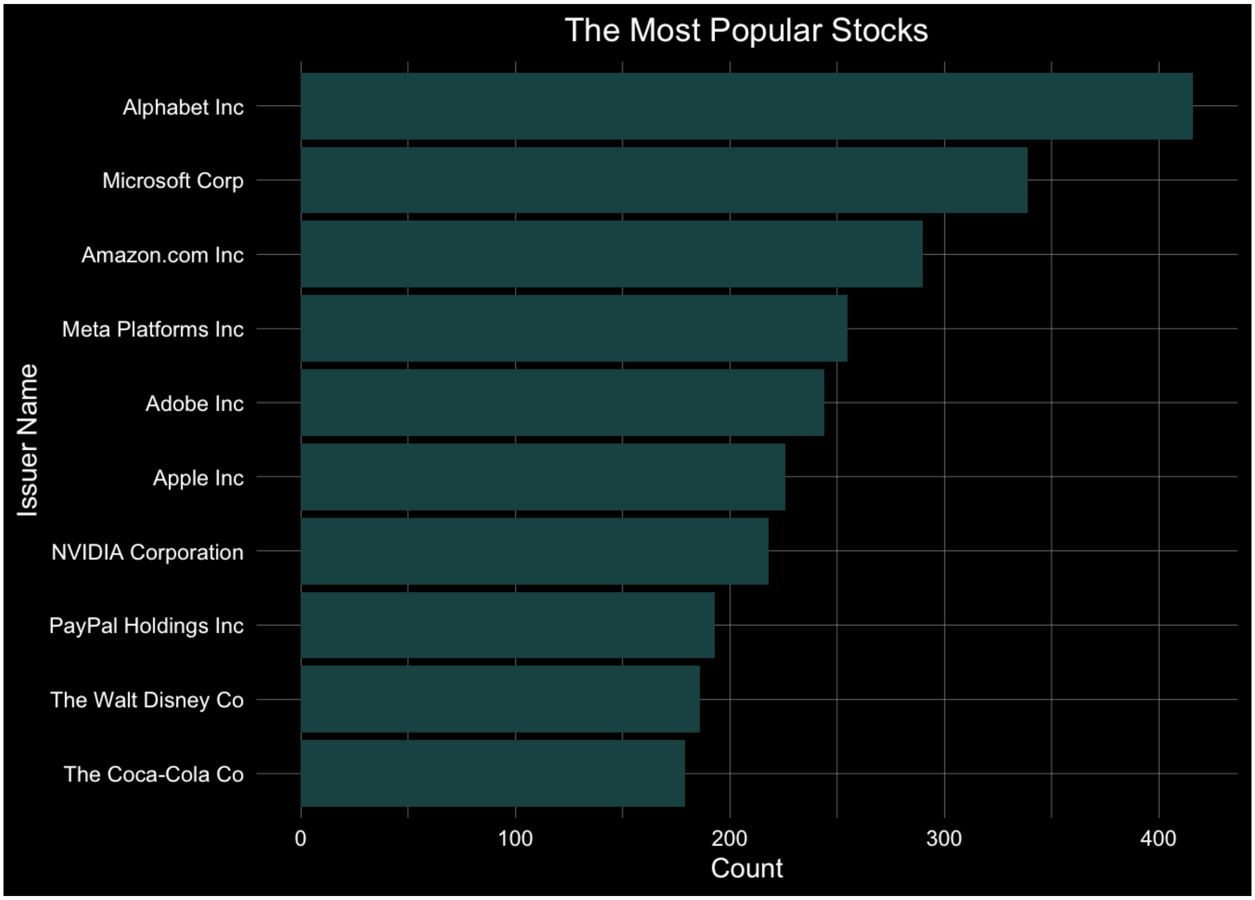
Politician Name	Total Transactions	Proportion of Very Good Decisions	p value
Eric Burlison	5	100.0%	0.0010
Trey Hollingsworth	9	88.9%	0.0001
Frank Pallone	5	80.0%	0.0156
Bobby Scott	9	66.7%	0.0100
Van Taylor	9	55.6%	0.0489
Ron Wyden	28	50.0%	0.0041
Austin Scott	46	47.8%	0.0009
Cindy Axne	72	47.2%	0.0001
Neal Dunn	32	46.9%	0.0071
Thomas Kean Jr	22	45.5%	0.0444
Markwayne Mullin	147	36.1%	0.0030
Chuck Fleischmann	154	33.8%	0.0152
Diana Harshbarger	2541	31.2%	0.0000
Michael McCaul	4113	26.3%	0.0497

Q5: Are certain states more active by the number of transactions?



Yes. The most trades happen in California, Texas, Tennessee, New York, and New Jersey.

Q6: Which ticker symbols are most frequently involved in politicians' transactions?



80% of the top 10 traded companies are tech companies.

Conclusions and Future Questions:

This report explored the trading activities of politicians and analyzed whether their transactions outperform the general market, along with several sub-questions. The findings indicate that on average, politicians do not outperform the market; the S&P 500 yielded an average annualized profit of 6.8% over the studied period, compared to 4.7% for politicians.

The investigation revealed no significant correlation between the Reporting Gap and trade profitability, suggesting that the timing of trade disclosures may not influence the financial outcomes of these trades directly. The analysis also showed that while Democrats engage in a higher number of transactions, Republicans appear to realize higher profitability on their trades.

Marie Newman appeared as the most successful investor within the government, achieving an annualized profit vastly exceeding that of the S&P 500. The data also identified potential cases of insider trading, signaling the need for further scrutiny into trades that substantially outperform the market under suspicious circumstances.

Certain states, such as California, Texas, Tennessee, New York, and New Jersey, showed higher trading activities, which may reflect the socio-economic and political dynamics within these regions. Additionally, the predominance of tech stocks among the most traded symbols aligns with broader market trends but also points to potential sector biases in trading decisions.

Moving forward, alternative methodologies can be applied to the existing data to expand the analysis. Additionally, there could be an impact of the political events on the trading activity.

This study contributes to our understanding of the financial activities of politicians, providing a foundation for future research and potentially informing policy discussions on transparency and regulation in political stock trading. The need for more detailed and accessible data remains a significant challenge, underscoring the importance of transparency for fair market conditions and public trust in political figures.

Limitations

- The three-year data duration limits evaluation of long-term investment strategies, potentially missing broader market cycles and politicians' investment skills.
- Methodology 1 assumes that the amount of stock bought and sold within a cycle is equal, which might not reflect reality. Additionally, it assumes that all stocks are sold by the end of the cycle, an assumption that often does not align with typical investing practices where holdings may persist across multiple cycles or remain unsold for strategic reasons. These assumptions can lead to inaccuracies in analyzing the financial performance and decision-making strategies of politicians, whose trading patterns may exhibit complex behaviors not fully captured by these assumptions.
- Politicians are required to report only transaction volume ranges, not exact amounts, which restricts the ability to precisely assess the impact of their trades. This limitation makes it difficult to determine the exact quantity of stock bought or sold. Such constraints lead to significant gaps in analysis, particularly when trying to evaluate the financial outcomes or potential market influence of these trades. This lack of precision can also hinder efforts to detect patterns that might suggest unethical behavior, like insider trading.
- Some trading activities by politicians might not be disclosed, particularly those conducted through intermediaries or associates, creating blind spots.

Code Appendix

Please review the pages below.

Final Report

Maxim Dokukin

2024-04-19

Data Loading

```
transactions_df <- read.csv(url('https://raw.githubusercontent.com/maxdokukin/Politic
ian-Trades/main/Data/data/transactions_cleaned.csv'))
trades_df <- read.csv(url('https://raw.githubusercontent.com/maxdokukin/Politician-Tr
ades/main/Data/data/trades.csv'))
more_prices_df <- read.csv(url('https://raw.githubusercontent.com/maxdokukin/Politici
an-Trades/main/Data/data/more_prices.csv'))
```

Libraries

```
library(dplyr)
library(ggplot2)
library(tidyr)
library(broom)
```

Data Processing

```
# Calculating stock performance before and after transaction dates

# before and after purchase average annualized rates of change
more_prices_df <- more_prices_df |>
  mutate(
    pd1 = ((Price_.60_days / Price_.90_days - 1) / 30) * 365,
    pd2 = ((Price_.40_days / Price_.60_days - 1) / 20) * 365,
    pd3 = ((Price_.20_days / Price_.40_days - 1) / 20) * 365,
    pd4 = ((Price_.10_days / Price_.20_days - 1) / 10) * 365,
    pd5 = ((Price_0_days / Price_.10_days - 1) / 10) * 365,
    pd_avg = rowMeans(cbind(pd1, pd2, pd3, pd4, pd5), na.rm = TRUE),

    fd1 = ((Price_10_days / Price_0_days - 1) / 10) * 365,
    fd2 = ((Price_20_days / Price_10_days - 1) / 10) * 365,
    fd3 = ((Price_40_days / Price_20_days - 1) / 20) * 365,
    fd4 = ((Price_60_days / Price_40_days - 1) / 20) * 365,
```

```
    fd5 = ((Price_90_days / Price_60_days - 1) / 30) * 365,
    fd_avg = rowMeans(cbind(fd1, fd2, fd3, fd4, fd5), na.rm = TRUE),

    perf_delta = fd_avg - pd_avg
  )

more_prices_df$Transaction.Type[more_prices_df$Transaction.Type == 'RECEIVE'] <- 'BU
Y'
more_prices_df <- more_prices_df[!more_prices_df$Transaction.Type == 'EXCHANGE', ]
more_prices_df <- more_prices_df[!more_prices_df$Party == 'Other', ]

# calculate market figures
market_figure <- function(row) {
  pd_avg <- as.numeric(row[33])
  fd_avg <- as.numeric(row[39])

  if (is.na(pd_avg) || is.na(fd_avg)) {
    return(NA)
  } else if (pd_avg <= 0 && fd_avg <= 0) {
    return('CTS_DOWN')
  } else if (pd_avg >= 0 && fd_avg >= 0) {
    return('CTS_UP')
  } else if (pd_avg > 0 && fd_avg < 0) {
    return('PEAK')
  } else if (pd_avg < 0 && fd_avg > 0) {
    return('DIP')
  }
}

more_prices_df$market_figure <- apply(more_prices_df, 1, market_figure)
more_prices_df <- na.omit(more_prices_df, cols = "market_figure")


# calculate success of transactions
eval_success <- function(row) {
  transaction_type <- row[11]
  market_figure <- row[41]

  if (market_figure == "CTS_DOWN" && transaction_type == "SELL") {
    return("Good Decision")
  } else if (market_figure == "CTS_DOWN" && transaction_type == "BUY") {
    return("Bad Decision")
  } else if (market_figure == "CTS_UP" && transaction_type == "SELL") {
    return("Bad Decision")
  } else if (market_figure == "CTS_UP" && transaction_type == "BUY") {
    return("Good Decision")
  } else if (market_figure == "PEAK" && transaction_type == "SELL") {
```

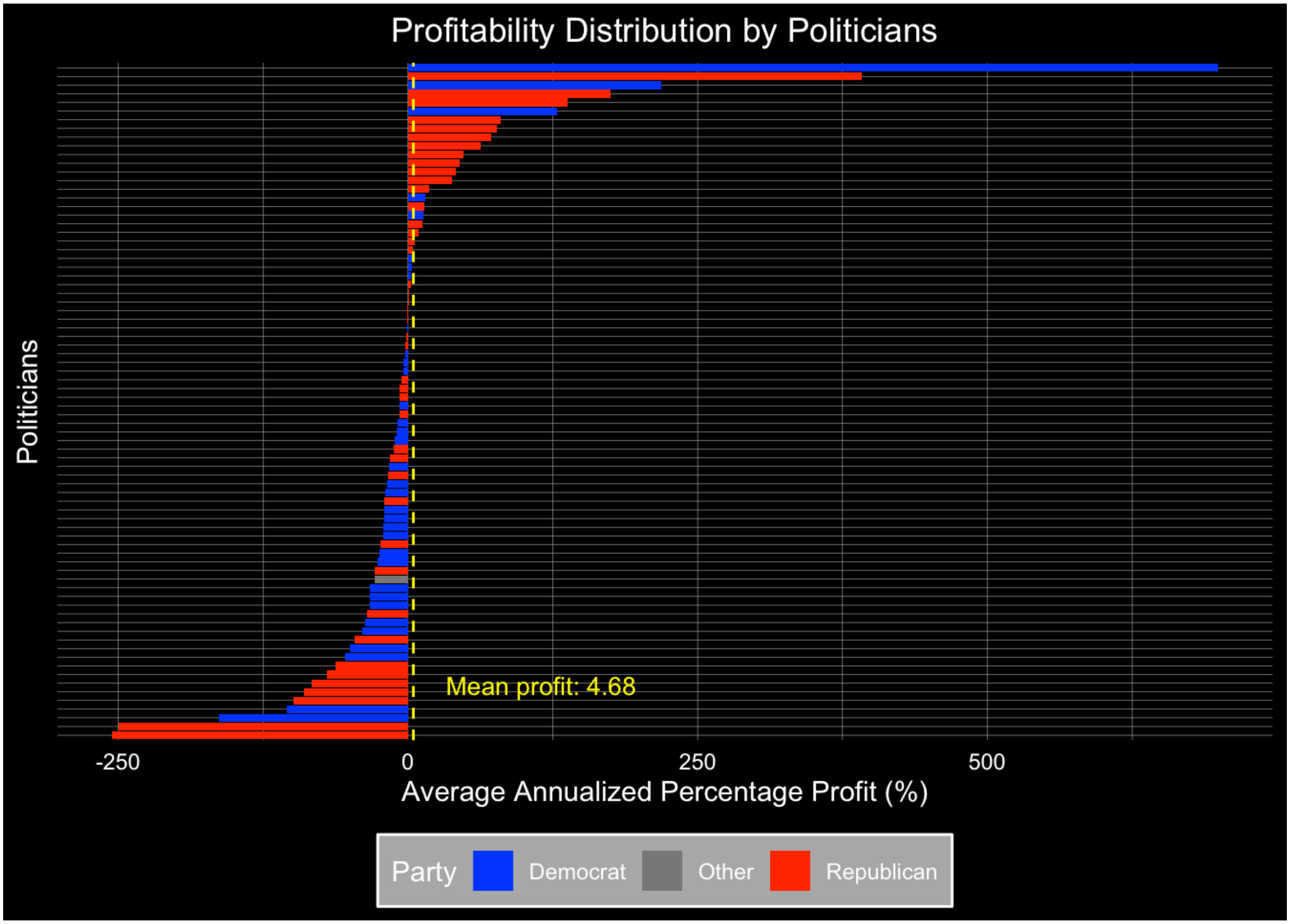
```
    return("Very Good Decision")
  } else if (market_figure == "PEAK" && transaction_type == "BUY") {
    return("Very Bad Decision")
  } else if (market_figure == "DIP" && transaction_type == "SELL") {
    return("Very Bad Decision")
  } else if (market_figure == "DIP" && transaction_type == "BUY") {
    return("Very Good Decision")
  }
}

more_prices_df$transaction_eval <- apply(more_prices_df, 1, eval_success)
```

```
# create df that contains info by politician names

general_info <- transactions_df |>
  group_by(Politician.Name) |> distinct(Politician.Name, Party, Chamber)

avg_apy_trades <- trades_df |>
  group_by(Politician.Name) |>
  summarise(
    Average_Annualized_Percentage_Profit = mean(Annualized.Percentage.Profit),
    Total_Trades = n(),
    .groups = 'drop'
  )

total_transactions <- transactions_df |>
  group_by(Politician.Name) |>
  summarise(Total_Transactions = n(), .groups = 'drop')

decisions <- more_prices_df |>
  group_by(Politician.Name, transaction_eval) |>
  summarise(Decision_Counts = n(), .groups = 'drop') |>
  pivot_wider(
    names_from = transaction_eval,
    values_from = Decision_Counts,
    values_fill = list(Decision_Counts = 0)
  ) |>
  full_join(total_transactions, by="Politician.Name") |>
  mutate(
    Proportion_of_Very_Bad_Decisions = (`Very Bad Decision` / Total_Transactions) * 1
00,
    Proportion_of_Bad_Decisions = (`Bad Decision` / Total_Transactions) * 100,
    Proportion_of_Good_Decisions = (`Good Decision` / Total_Transactions) * 100,
    Proportion_of_Very_Good_Decisions = (`Very Good Decision` / Total_Transactions) *
100
  )

# very useful for plotting
politician_info_df <- as.data.frame(
  general_info |>
  full_join(avg_apy_trades, by = "Politician.Name") |>
  full_join(decisions, by = "Politician.Name"))

rm(avg_apy_trades, decisions, general_info, total_transactions)
```

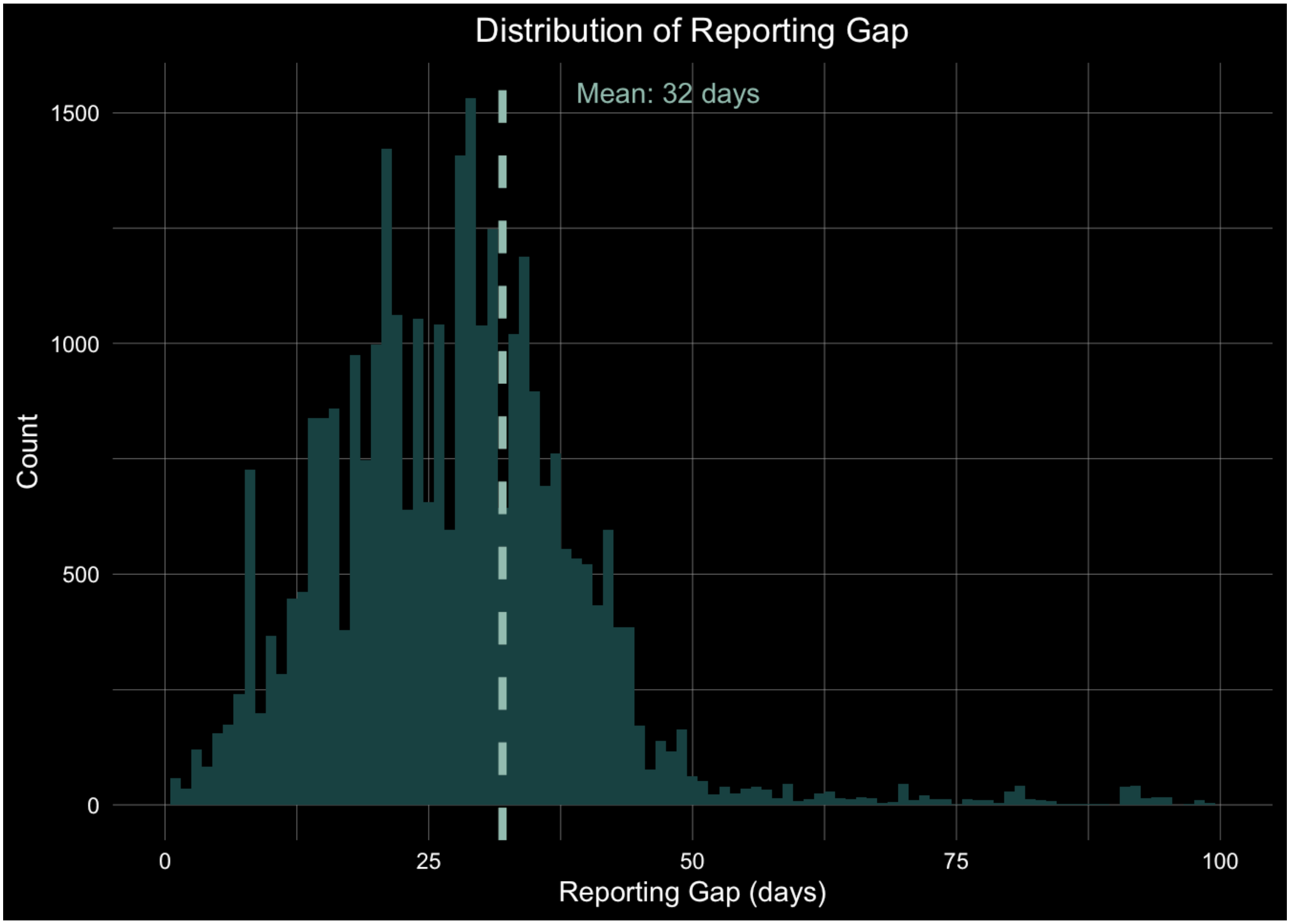
```r
# global graphing vars
party_colors <- c(Democrat = "blue", Republican = "red", Other = "#777777")

dark_theme <- function() {
  theme_minimal() +
  theme(plot.background = element_rect(fill = "black"),
        panel.background = element_rect(fill = "black"),
        text = element_text(color = "white"),
        axis.title = element_text(color = "white"),
        axis.text = element_text(color = "white"),
        legend.background = element_rect(fill = "darkgrey", color = "white"),
        legend.position = "bottom",
        panel.grid.major = element_line(color = "darkgray", size = 0.1),
        panel.grid.minor = element_line(color = "darkgray", size = 0.1),
        plot.title = element_text(hjust = 0.5))
}
```

Plots

```
# prepare data
plot_data <- politician_info_df[!is.na(politician_info_df$Average_Annualized_Percenta
ge_Profit), ]
mean_profit <- mean(politician_info_df$Average_Annualized_Percentage_Profit, na.rm =
TRUE)

# plot
ggplot(plot_data,
       aes(y = reorder(Politician.Name, Average_Annualized_Percentage_Profit),
           x = Average_Annualized_Percentage_Profit, fill = Party)) +
  # main data
  geom_bar(stat = "identity", position = "dodge") +
  scale_fill_manual(values = party_colors) +
  # mean line
  geom_vline(xintercept = mean_profit, linetype = "dashed", color = "yellow", size =
0.5) +
  annotate("text", y = mean_profit, x = 0, label = sprintf("Mean profit: %.2f", mean_
profit),
            vjust = -0.5, hjust = -0.2, color = "yellow", size = 3.5) +
  # theme
  dark_theme() +
  theme(axis.text.y = element_blank())+
  # labels
  labs(title = "Profitability Distribution by Politicians",
       x = "Average Annualized Percentage Profit (%)",
       y = "Politicians")
```


Profitability Distribution by Politicians
Politicians
Mean profit: 4.68
-250
0
250
500
Average Annualized Percentage Profit (%)
Party
Democrat
Other
Republican

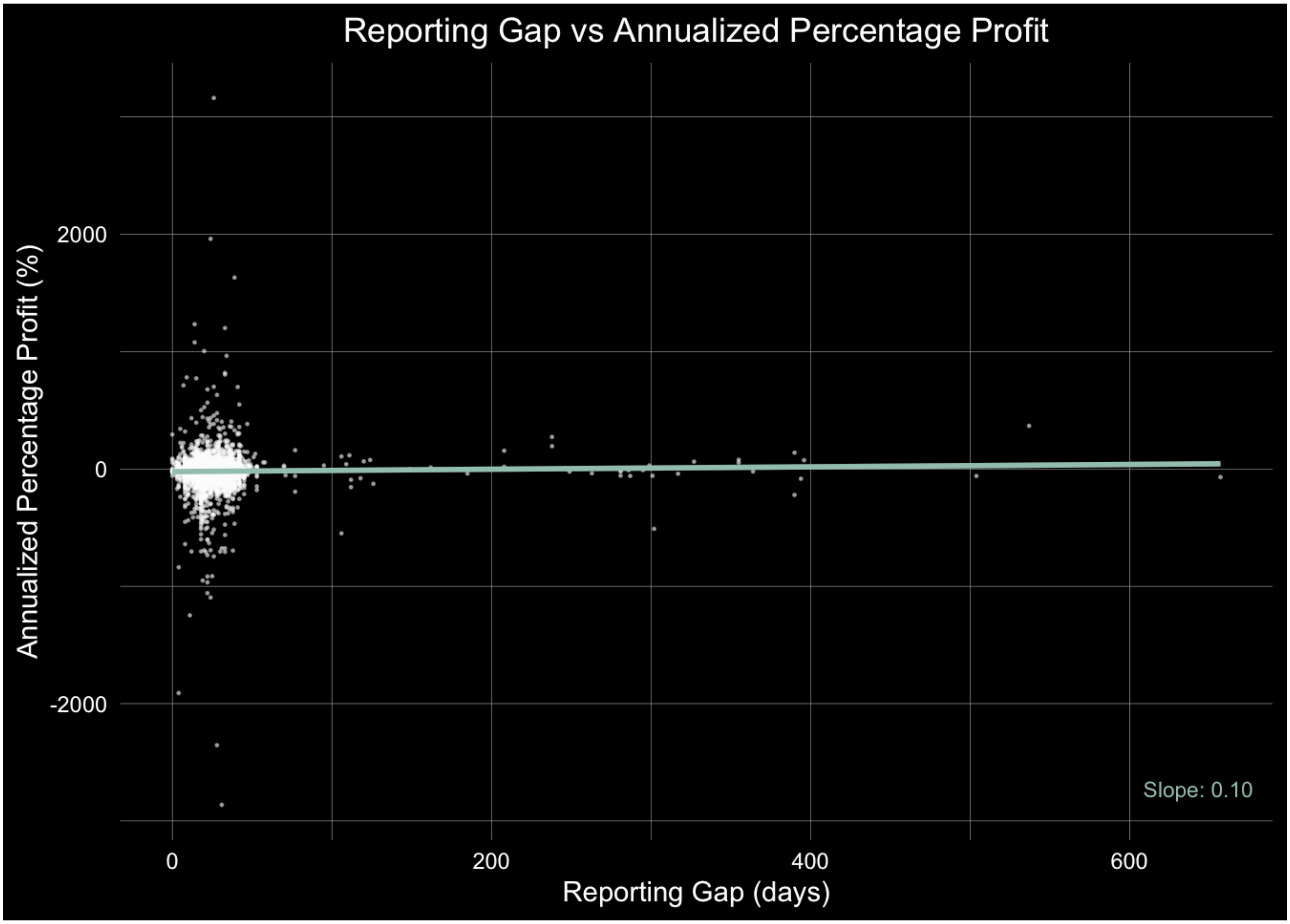
```
# prepare data
plot_data <- transactions_df
mean_gap <- mean(transactions_df$Reporting.Gap, na.rm = TRUE)

# plot
ggplot(plot_data, aes(x = Reporting.Gap)) +
  # main data
  geom_histogram(binwidth = 1, fill = "#1B4242") +
  # mean line
  geom_vline(aes(xintercept = mean_gap), color = "#96BDB0", linetype = "dashed", size
= 1.5) +
  annotate("text", x = mean_gap, y = Inf, label = sprintf("Mean: %.0f days", mean_ga
p),
           vjust = 2, hjust=-0.4, color = "#96BDB0") +
  # theme
  dark_theme() +
  # labels
  labs(title = "Distribution of Reporting Gap",
       x = "Reporting Gap (days)",
       y = "Count") +
  # axes
  xlim(0, 100)
```



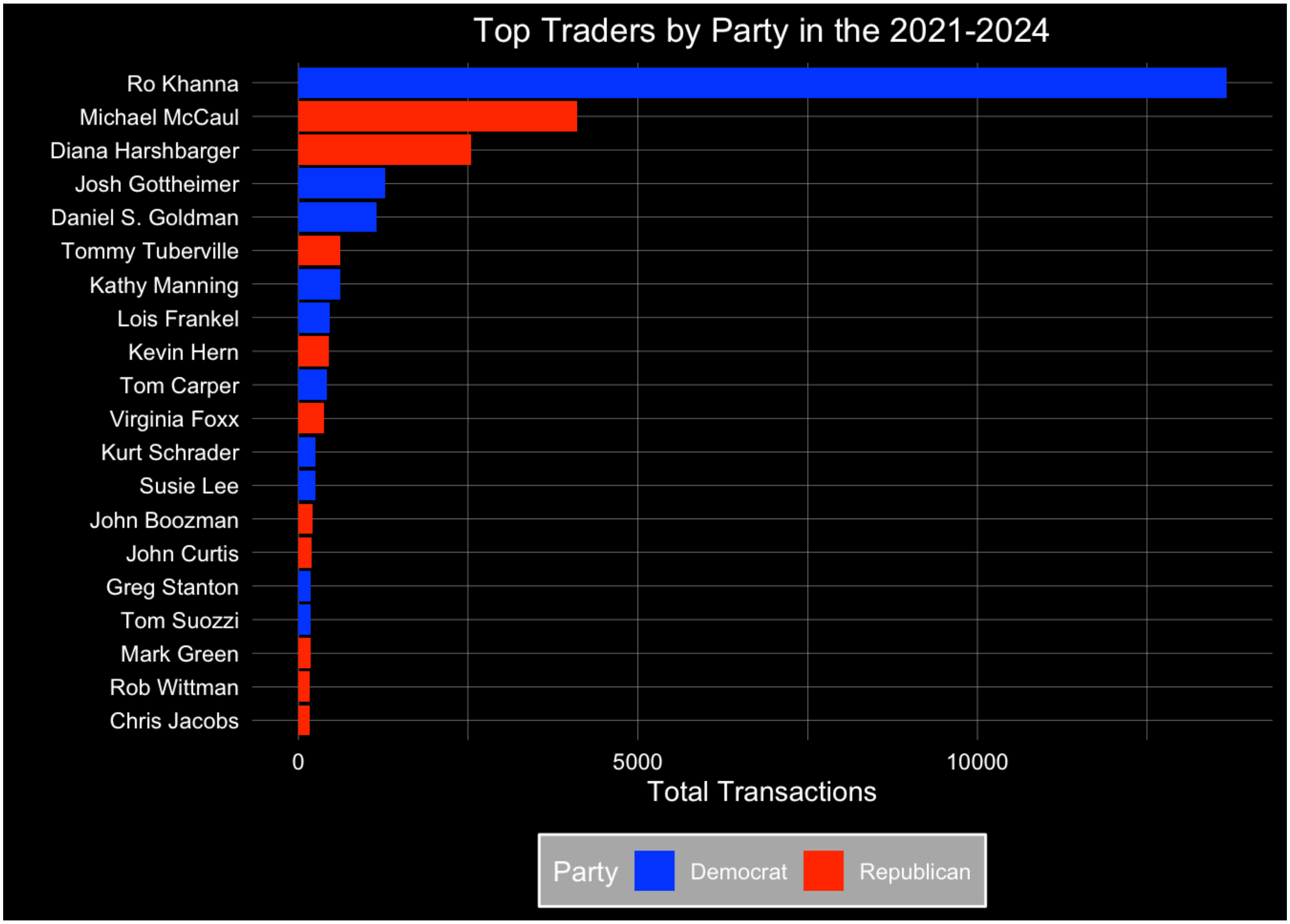

```
# prepare data
plot_data <- trades_df

# model
model <- lm(Annualized.Percentage.Profit ~ Report.Gap, data = plot_data)
summary(model)
```

```
##
## Call:
## lm(formula = Annualized.Percentage.Profit ~ Report.Gap, data = plot_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2843.4   -24.9     9.9    32.7  3181.6
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -22.76493    2.70443  -8.418   <2e-16 ***
## Report.Gap    0.10136    0.07266   1.395    0.163
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 137.4 on 5319 degrees of freedom
## Multiple R-squared:  0.0003658,  Adjusted R-squared:  0.0001778
## F-statistic: 1.946 on 1 and 5319 DF,  p-value: 0.163
```

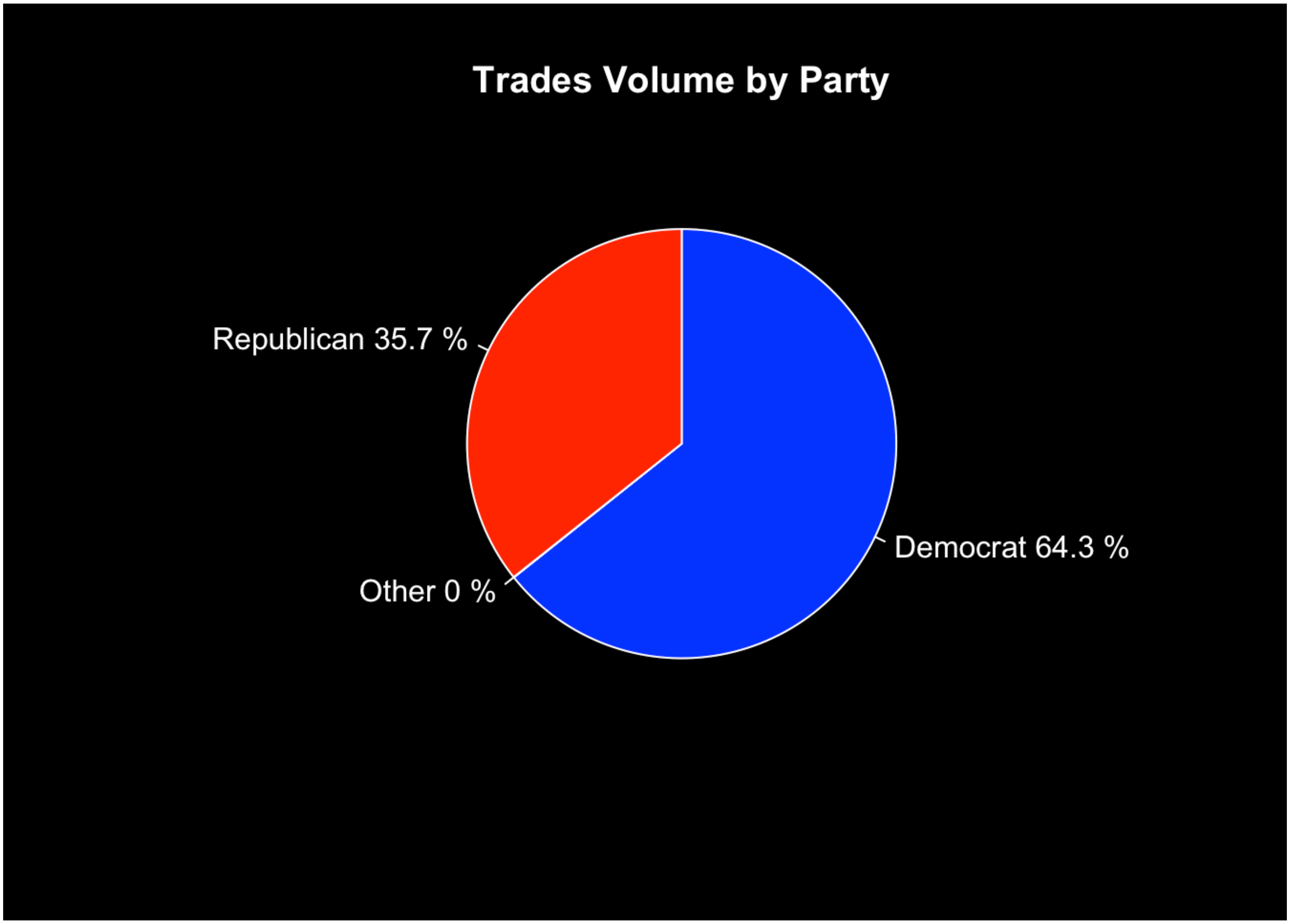
```
slope <- coef(model)["Report.Gap"]

# plot
ggplot(plot_data, aes(x = Report.Gap, y = Annualized.Percentage.Profit)) +
  # main data
  geom_point(alpha = 0.4, size = 0.2, color = "white") +
  # reg line
  geom_smooth(method = "lm", se = FALSE, color = "#96BDB0") +
  annotate("text",
           x = max(trades_df$Report.Gap, na.rm = TRUE),
           y = min(trades_df$Annualized.Percentage.Profit,
           na.rm = TRUE),
           label = sprintf("Slope: %.2f", slope),
           hjust = 0.7,
           vjust = -0.5,
           size = 3,
           color = "#96BDB0") +
  # theme
  dark_theme() +
  # labels
  labs(title = "Reporting Gap vs Annualized Percentage Profit",
       x = "Reporting Gap (days)",
       y = "Annualized Percentage Profit (%)")
```



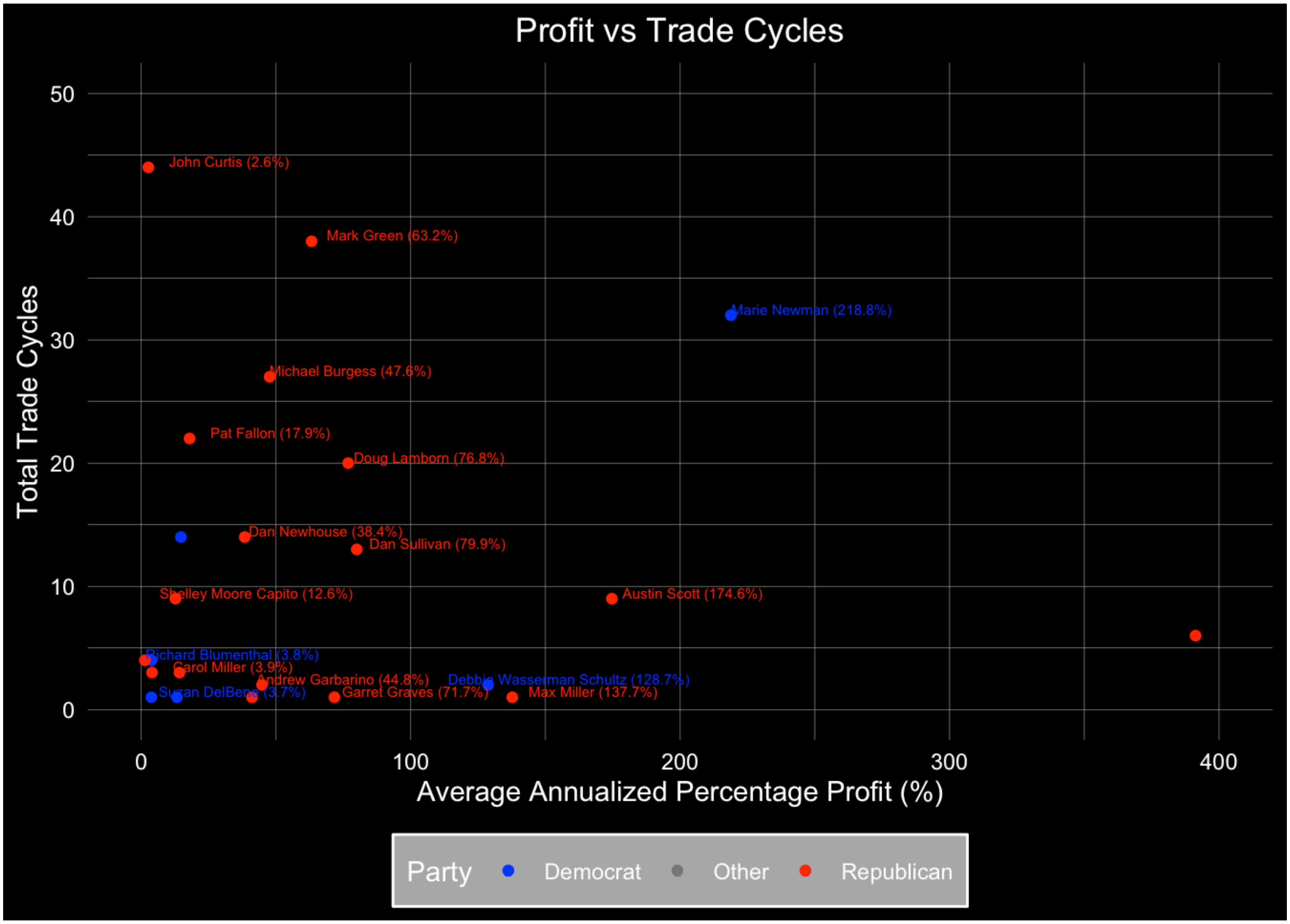

```r
# prepare data
plot_data <- politician_info_df |>
  top_n(20, Total_Transactions)

# plot
ggplot(plot_data, aes(y = reorder(Politician.Name, Total_Transactions),
                      x = Total_Transactions,
                      fill = Party)) +
  # main data
  geom_bar(stat = "identity") +
  scale_fill_manual(values = party_colors) +
  # theme
  dark_theme() +
  # labels
  labs(title = "Top Traders by Party in the 2021-2024",
       x = "Total Transactions",
       y = "")
```



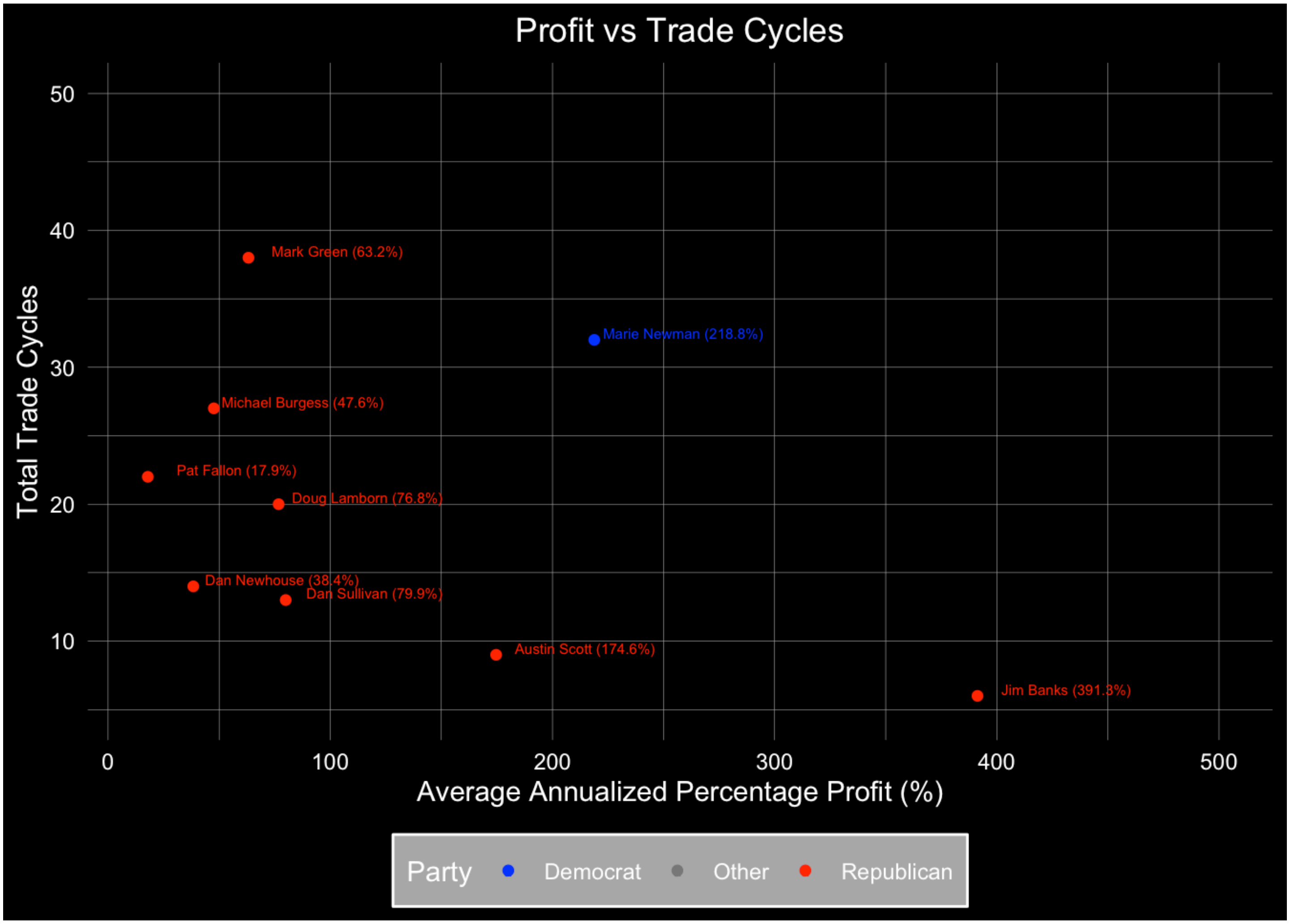

```
# prepare data
plot_data <- transactions_df |>
  count(Party)
percentages <- round(100 * plot_data$n / sum(plot_data$n), 1)
labels <- paste(plot_data$Party, percentages, "%", sep=" ")

# plot
par(bg = "black", col.main = "white", col.lab = "white", col.axis = "white", fg = "wh
ite")
pie(plot_data$n, labels = labels,
    col = c("blue", "gray", "red"),
    main = "Trades Volume by Party",
    init.angle = 90,
    clockwise = TRUE)
```


Trades Volume by Party
Republican 35.7 %
Democrat 64.3 %
Other 0 %

```r
# prepare data
plot_data <- politician_info_df |>
  mutate(label = paste(Politician.Name,
                       sprintf("(%.1f%%)",
                       Average_Annualized_Percentage_Profit),
                       sep = " "))

# plot
ggplot(plot_data, aes(x = Average_Annualized_Percentage_Profit,
                           y = Total_Trades, color = Party)) +
  # main data
  geom_point() +
  geom_text(aes(label = label),
            nudge_x = 30,
            nudge_y = 0.5,
            size = 2,
            check_overlap = TRUE) +
  scale_color_manual(values = party_colors) +
  # theme
  dark_theme() +
  # labels
  labs(title = "Profit vs Trade Cycles",
       x = "Average Annualized Percentage Profit (%)",
       y = "Total Trade Cycles") +
  # axes
  xlim(0, 400) +
  ylim(0, 50)
```


Profit vs Trade Cycles
Total Trade Cycles
Average Annualized Percentage Profit (%)
0
10
20
30
40
50
0
100
200
300
400
John Curtis (2.6%)
Mark Green (63.2%)
Marie Newman (218.8%)
Michael Burgess (47.6%)
Pat Fallon (17.9%)
Doug Lamborn (76.8%)
Dan Newhouse (38.4%)
Dan Sullivan (79.9%)
Shelley Moore Capito (12.6%)
Austin Scott (174.6%)
Richard Blumenthal (3.8%)
Carol Miller (3.9%)
Andrew Garbarino (44.8%)
Debbie Wasserman Schultz (128.7%)
Suzan DelBene (3.7%)
Garret Graves (71.7%)
Max Miller (137.7%)
Party
Democrat
Other
Republican

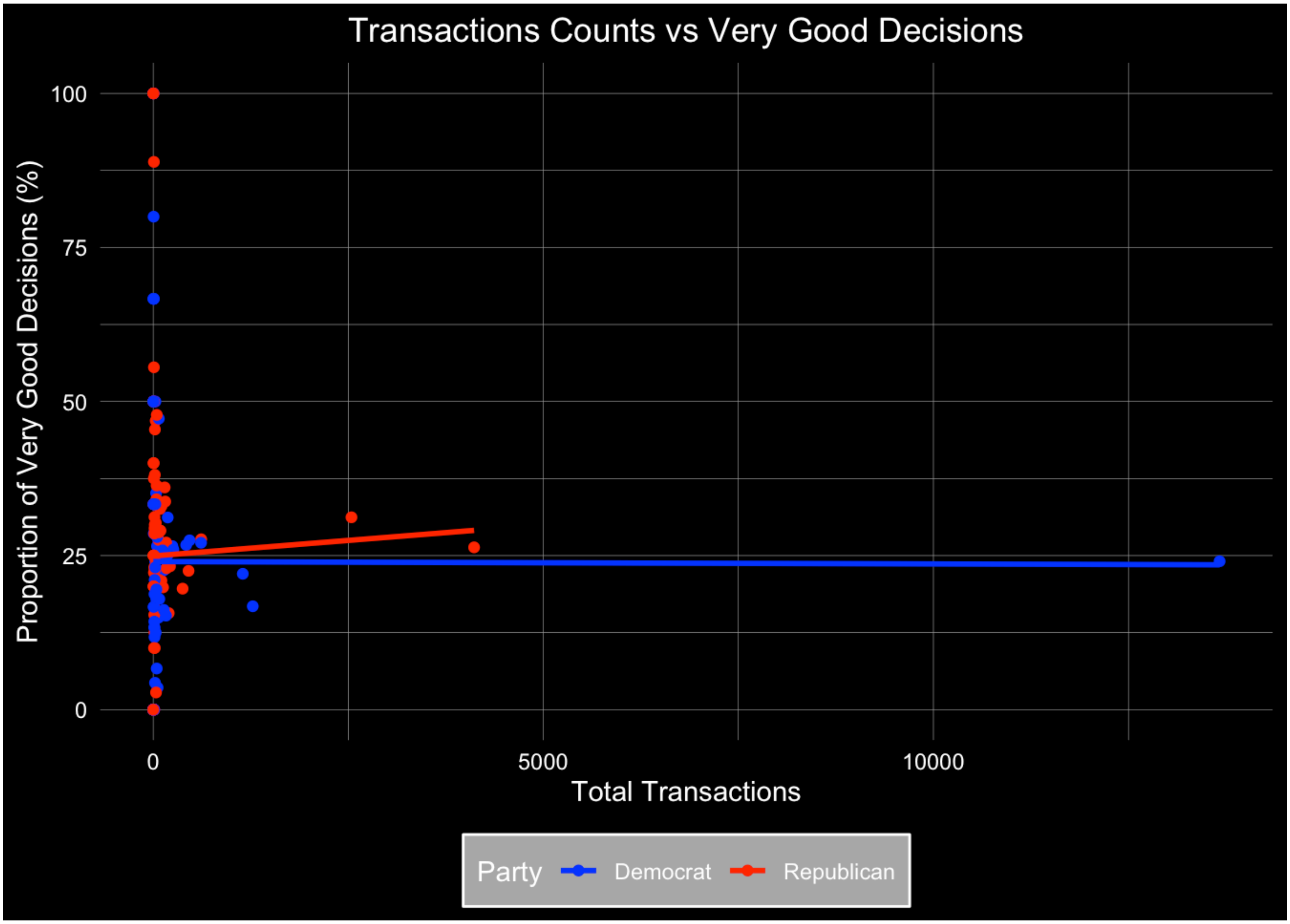
```r
# plot
ggplot(plot_data, aes(x = Average_Annualized_Percentage_Profit,
                           y = Total_Trades, color = Party)) +
  # main data
  geom_point() +
  geom_text(aes(label = label),
            nudge_x = 40,
            nudge_y = 0.5,
            size = 2,
            check_overlap = TRUE) +
  scale_color_manual(values = party_colors) +
  # theme
 dark_theme() +
  # labels
  labs(title = "Profit vs Trade Cycles",
       x = "Average Annualized Percentage Profit (%)",
       y = "Total Trade Cycles") +
  # axes
  xlim(15, 500) +
  ylim(5, 50)
```



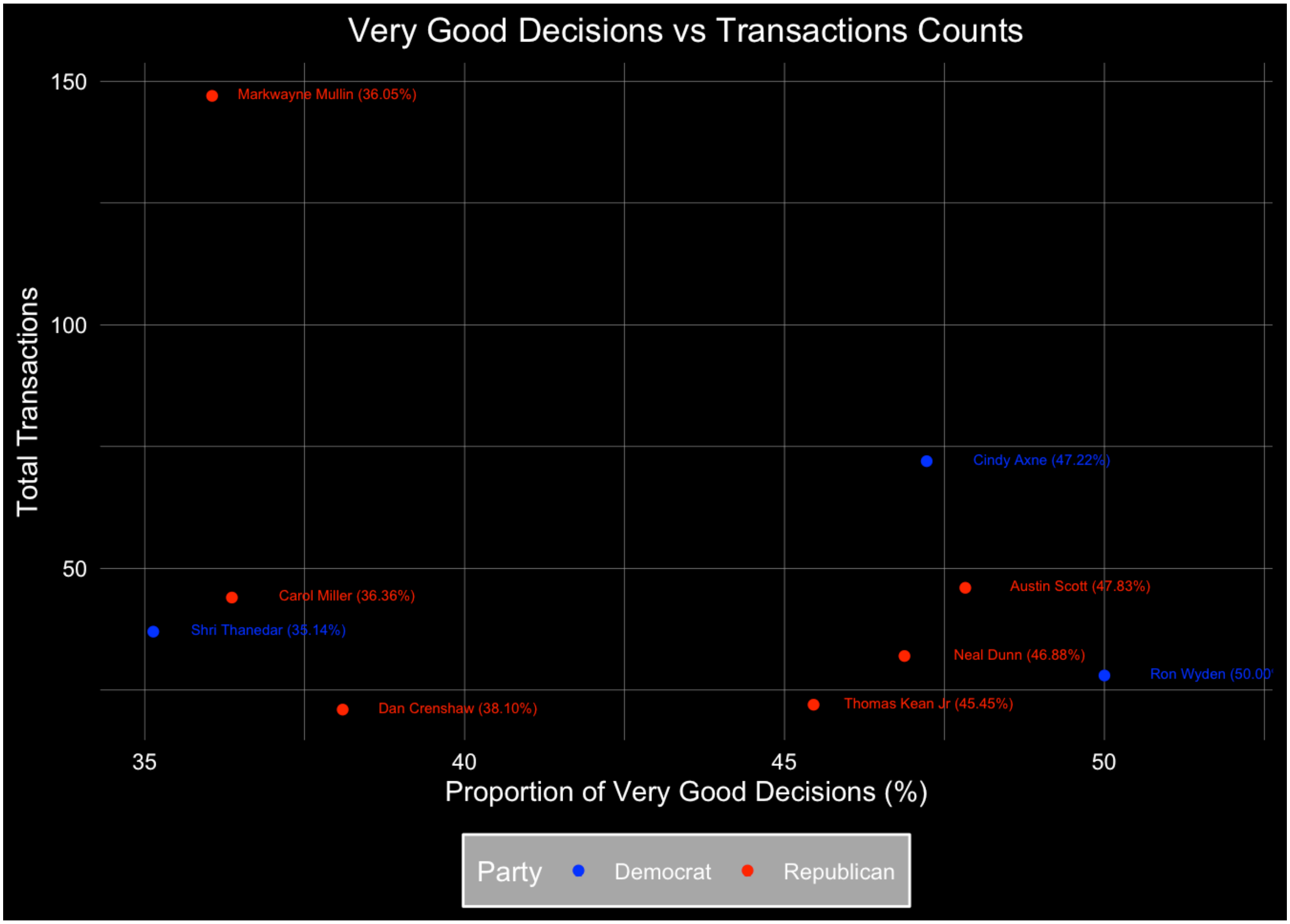

```
# prepare data
plot_data <- politician_info_df[!is.na(politician_info_df$Proportion_of_Very_Good_Dec
isions), ]

# model
models <- plot_data |>
  group_by(Party) |>
  do(tidy(lm(Proportion_of_Very_Good_Decisions ~ Total_Transactions, data = .)))

intercepts <- models |>
  filter(term == "(Intercept)") |>
  select(Party, intercept = estimate)

print(intercepts)
```

```
## # A tibble: 2 × 2
## # Groups:   Party [2]
##   Party      intercept
##   <chr>          <dbl>
## 1 Democrat        24.0
## 2 Republican      24.9
```

```
# plot
ggplot(plot_data, aes(x = Total_Transactions,
                      y = Proportion_of_Very_Good_Decisions, color = Party)) +
  # main data
  geom_point() +
  geom_smooth(method = "lm", formula = y ~ x, se = FALSE) +
  scale_color_manual(values = party_colors) +
  # theme
  dark_theme() +
  # labels
  labs(title = "Transactions Counts vs Very Good Decisions",
       x = "Total Transactions",
       y = "Proportion of Very Good Decisions (%)")
```


Transactions Counts vs Very Good Decisions
Proportion of Very Good Decisions (%)
100
75
50
25
0
0
5000
10000
Total Transactions
Party
Democrat
Republican

```
# prepare data
plot_data <- politician_info_df |>
  filter(Total_Transactions < 500) |>
  filter(Total_Transactions > 10 ) |>
  filter(Proportion_of_Very_Good_Decisions > 35) |>
  mutate(label = sprintf("%s (%.2f%%)", Politician.Name, Proportion_of_Very_Good_Deci
sions))

ggplot(plot_data, aes(y = Total_Transactions,
                      x = Proportion_of_Very_Good_Decisions,
                      color = Party)) +
  # main data
  geom_point() +
  geom_text(aes(label = label),
            nudge_x = 1.8,
            nudge_y = 0.5,
            size = 2,
            check_overlap = TRUE) +
  scale_color_manual(values = party_colors) +  # Define custom colors for the groups
  # theme
  dark_theme() +
  # labels
  labs(title = "Very Good Decisions vs Transactions Counts",
       x = "Proportion of Very Good Decisions (%)",
       y = "Total Transactions")
```




```
xlim(35, 55)
```

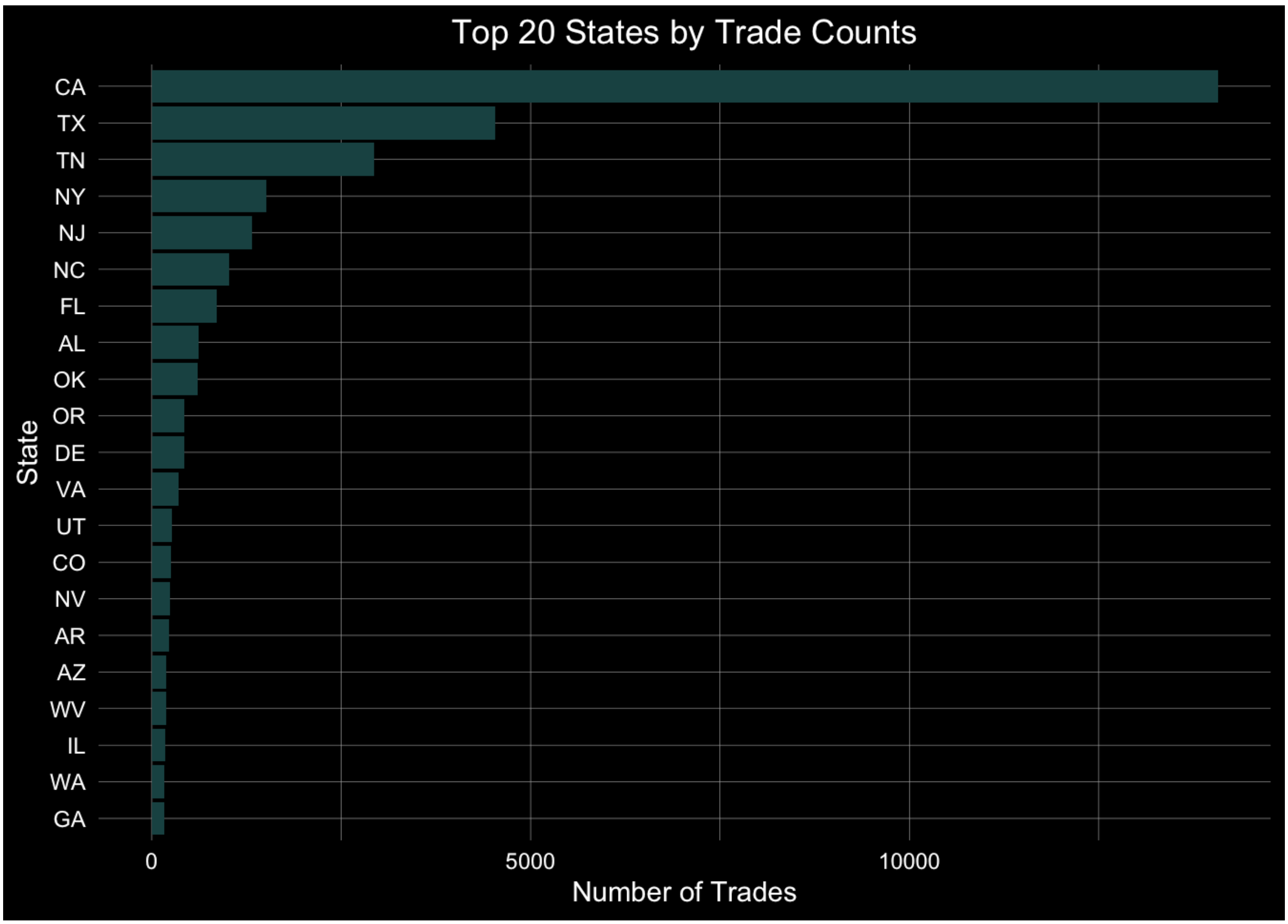
```
## <ScaleContinuousPosition>
##  Range:
##  Limits:   35 --   55
```

```r
# binom test on the proportion of very good decisions
tst_data <- politician_info_df[!is.na(politician_info_df$`Very Good Decision`), ]

tst_results <- tst_data |>
  rowwise() |>
  mutate(
    test_result = list(binom.test(x = `Very Good Decision`, n = Total_Transactions, p
= 0.25)),
    p_value = test_result$p.value,
    statistic = test_result$statistic
  ) |>
  ungroup()

sig_results <- tst_results %>%
  filter(p_value < 0.05, Proportion_of_Very_Good_Decisions > 25) |>
  arrange(desc(Proportion_of_Very_Good_Decisions)) |>
  select(`Politician.Name`, Total_Transactions, Proportion_of_Very_Good_Decisions, p_
value)

print(sig_results)
```

```
## # A tibble: 14 × 4
##    Politician.Name    Total_Transactions Proportion_of_Very_Good_Deci…¹  p_value
##    <chr>                           <int>                          <dbl>    <dbl>
##  1 Eric Burlison                       5                          100   9.77e- 4
##  2 Trey Hollingsworth                  9                           88.9 1.07e- 4
##  3 Frank Pallone                       5                           80   1.56e- 2
##  4 Bobby Scott                         9                           66.7 9.99e- 3
##  5 Van Taylor                          9                           55.6 4.89e- 2
##  6 Ron Wyden                          28                           50   4.10e- 3
##  7 Austin Scott                       46                           47.8 9.12e- 4
##  8 Cindy Axne                         72                           47.2 5.24e- 5
##  9 Neal Dunn                          32                           46.9 7.14e- 3
## 10 Thomas Kean Jr                     22                           45.5 4.44e- 2
## 11 Markwayne Mullin                  147                           36.1 2.99e- 3
## 12 Chuck Fleischmann                 154                           33.8 1.52e- 2
## 13 Diana Harshbarger                2541                           31.2 1.84e-12
## 14 Michael McCaul                   4113                           26.3 4.97e- 2
## # ℹ abbreviated name: ¹Proportion_of_Very_Good_Decisions
```

```r
write.csv(sig_results, "sus_poltics.csv")
```

```r
plot_data <- transactions_df |>
  count(State) |>
  arrange(desc(n)) |>
  slice_max(n, n = 20)

ggplot(plot_data, aes(x = reorder(State, n), y = n)) +
  # main data
  geom_bar(stat = "identity", fill = "#1B4242") +
  coord_flip() +
  # theme
  dark_theme() +
  # labels
  labs(x = "State",
       y = "Number of Trades",
       title = "Top 20 States by Trade Counts")
```



```
plot_data <- transactions_df |>
  drop_na(Issuer.Name) |>
  count(Issuer.Name) |>
  top_n(10, n) |>
  arrange(desc(n))

ggplot(plot_data, aes(x = reorder(Issuer.Name, n), y = n)) +
  # main data
  geom_bar(stat = "identity", fill = "#1B4242") +
  coord_flip() +
  # theme
  dark_theme() +
  # labels
  labs(title = "The Most Popular Stocks",
       x = "Issuer Name",
       y = "Count")
```