Analysis of the Triboro line

Requirements

Libraries

```
library(tidyverse)
```

Files

```
subway_lines <- readr::read_csv('data/nta-subway-lines.csv')
subway_times <- readr::read_csv('data/nta-subway-times.csv')
driving_times <- readr::read_csv('data/nta-driving-times.csv')
walking_times <- readr::read_csv('data/nta-walking-times.csv')
```

Data

```
subway_lines <- subway_lines %>%
  dplyr::mutate(
    log_S000 = log(S000)
  )
```

Exploratory data analysis

Distribution of job counts
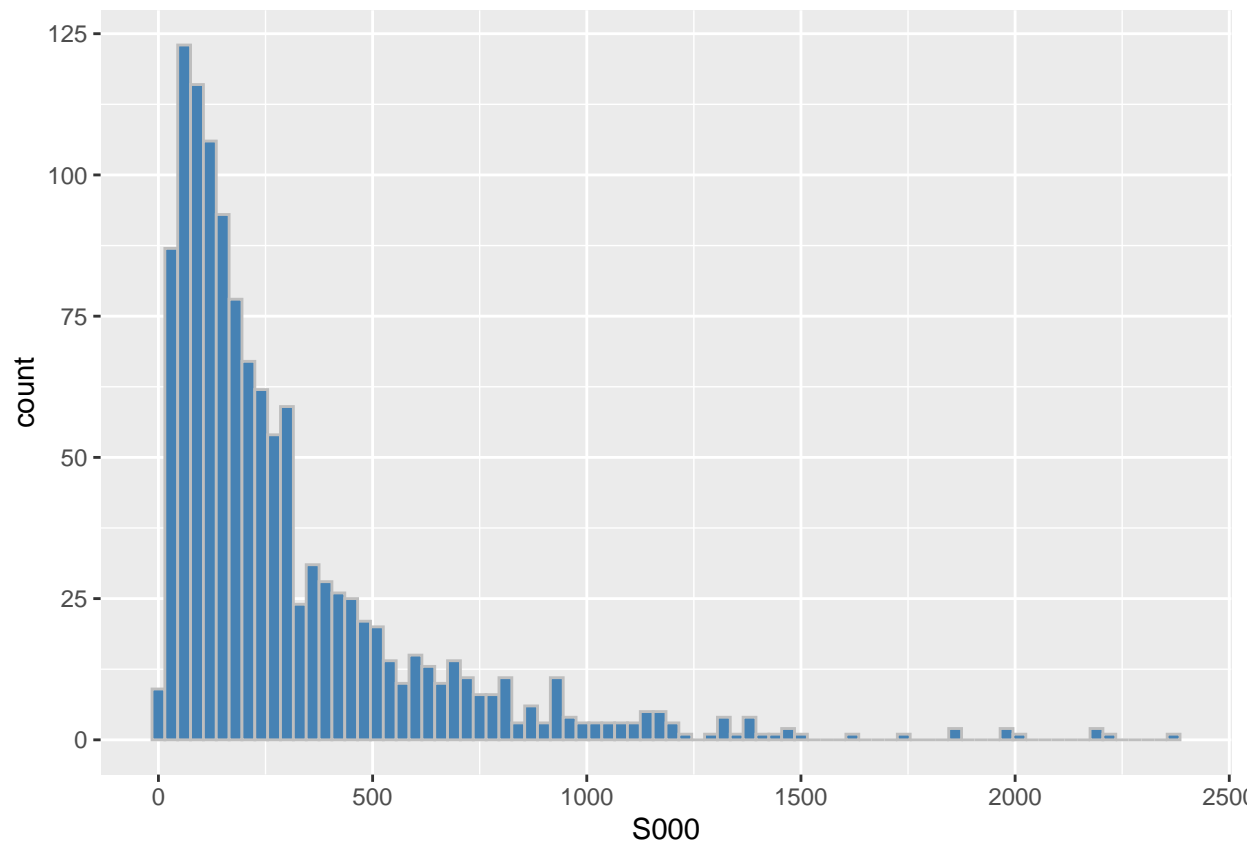
Original

Natural log

Distribution of commute counts
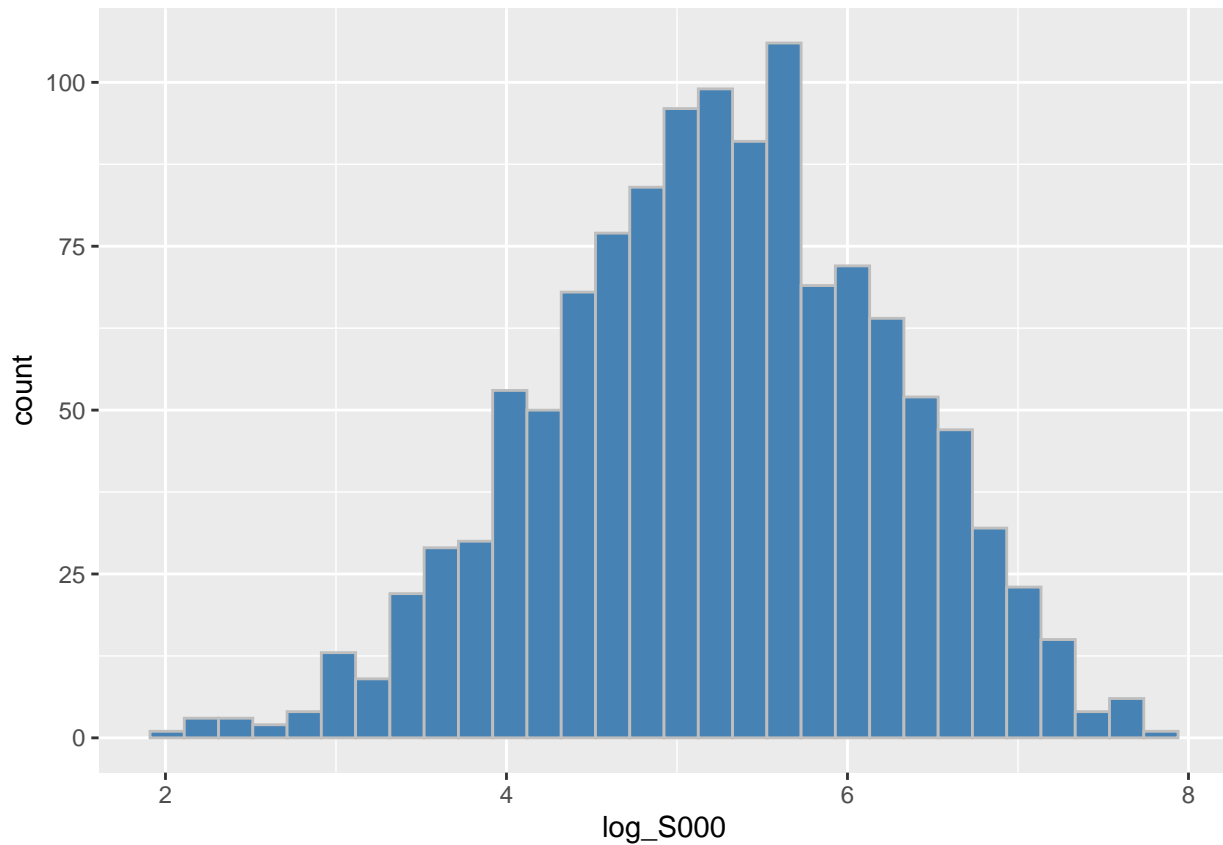
Original

```
commute_counts <- subway_lines %>%
  dplyr::select(trip, S000) %>%
  dplyr::mutate(
    log_S000 = log(S000)
  )
```

```
ggplot(commute_counts) +
  geom_histogram(aes(x = S000), fill = "steelblue", color = "grey", binwidth = 30)
```

Natural log

```
ggplot(data = commute_counts) +
  geom_histogram(aes(x = log_S000), bins = 30, fill = "steelblue", color = "grey")
```

Number of subway lines and commute count

Original

Transformed

Subway Transit time and commute count

Original

Transformed

Driving in traffic time and commute count

Original

Transformed

Walking time and commute count

Original

Transformed

Regression of Subway, Driving, and walking

Subway model

Driving

Walking

Multiple linear regression for all three factors

Equations plotted for all factors

Along all axis

Cut to most pivotal times (10 to 50 minutes) Table of values at 10, 25, 50

Auto Correlation of Subway, Driving, and Walking

Global Moran's I

Subway   Statistic Plot

Driving   Statisic Plot

Walking   Statustic Plot

LISA

Driving   Statistic Plot with major roadways overlayed

Walking   Statistic Plot of local area

Subway   Statistic Plot with subway lines

Network autocorrelation

Visualization of network

Visualization of network's complement

Global Moran's I

LISA   Plot by coloring desire lines