

# Project 1

Ekemini MbekAnako

2023-10-31

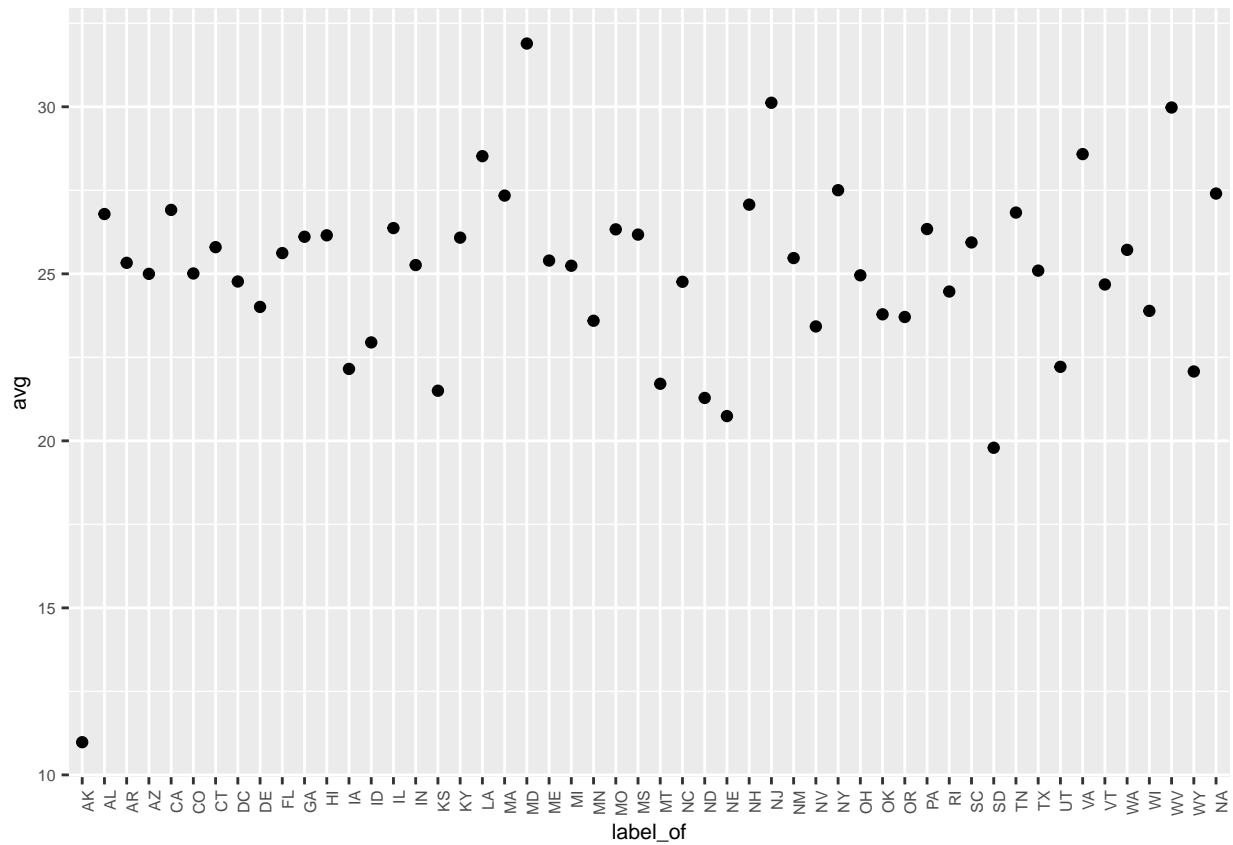
## #Introduction

The goal of this project is to find the correlation of commute times and population density of each state in the United States of America. It is hypothesized that states with high population density have increased commute times due to their condensed nature.

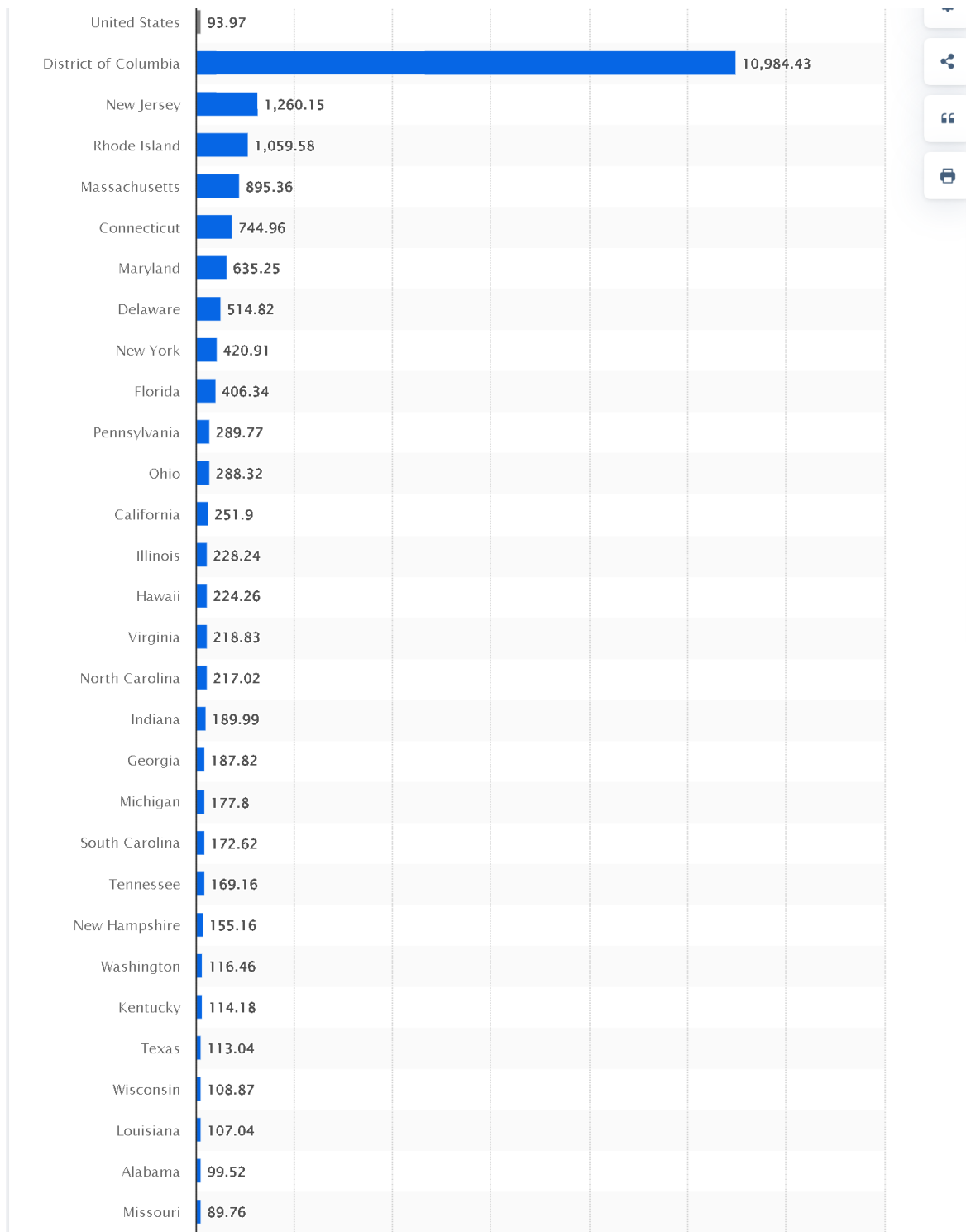
```
setwd("C:/Users/ekemi/Documents/R STUFF")
library(readr)
data <- read_csv('avg-commute-time-by-state.csv')

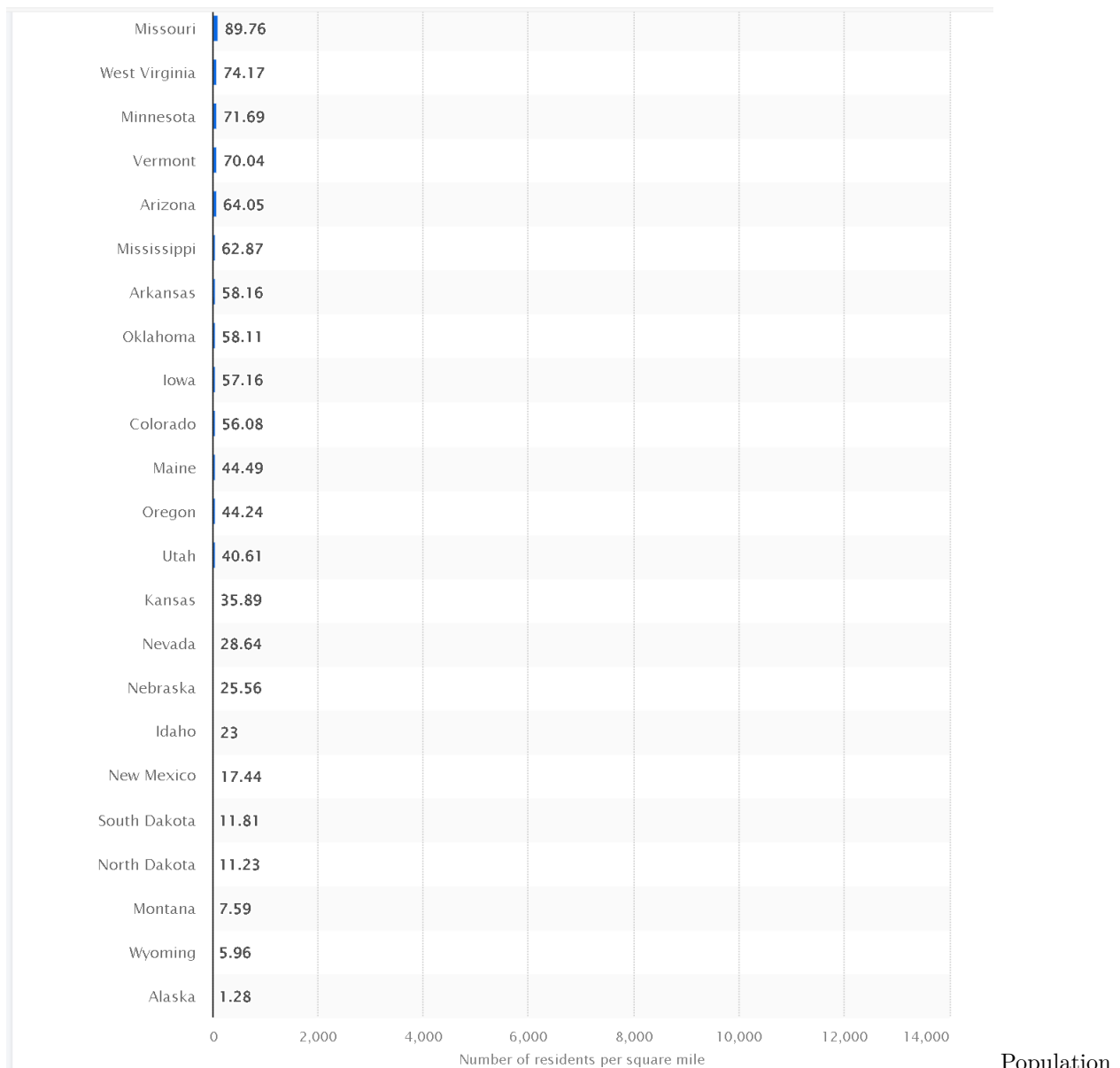
## Rows: 52 Columns: 2
## -- Column specification -----
## Delimiter: ","
## chr (1): label_of
## dbl (1): avg
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

library(ggplot2)
ggplot(data, aes(x = label_of, y = avg)) + geom_point() + theme(text = element_text(size=8),
  axis.text.x = element_text(angle=90, hjust=1))
```



This graph shows the average commute time of each state. The state abbreviation are the USPS standard. As shown, majority of the commute times are between 20 and 30 minutes.



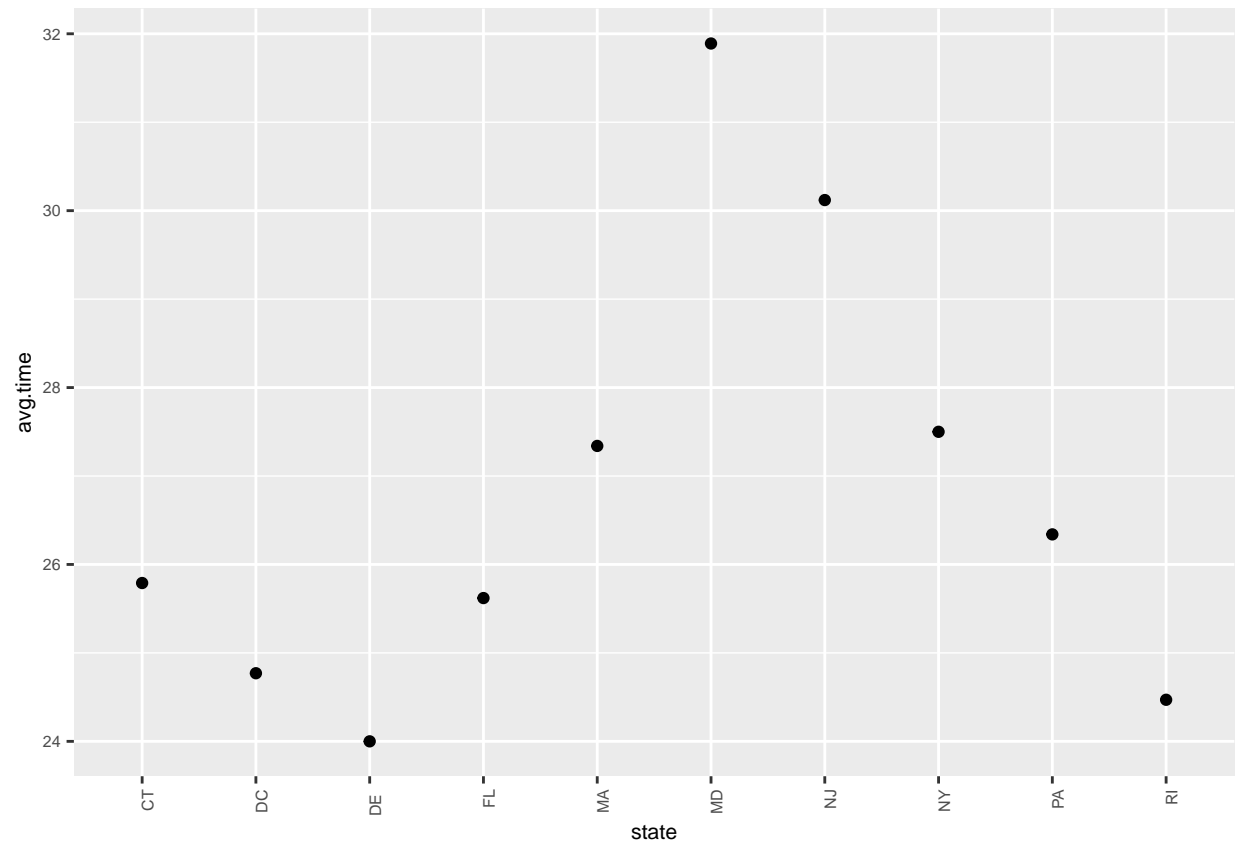


density for all states. Couldn't find an actual dataset with it (At least, one that was free).

For sake of time, I decided to take the 10 densest states and compared their commute times.

```
setwd("C:/Users/ekemi/Documents/R STUFF")
library(readr)
data2 <- read_csv('commute-times-top-ten.csv')

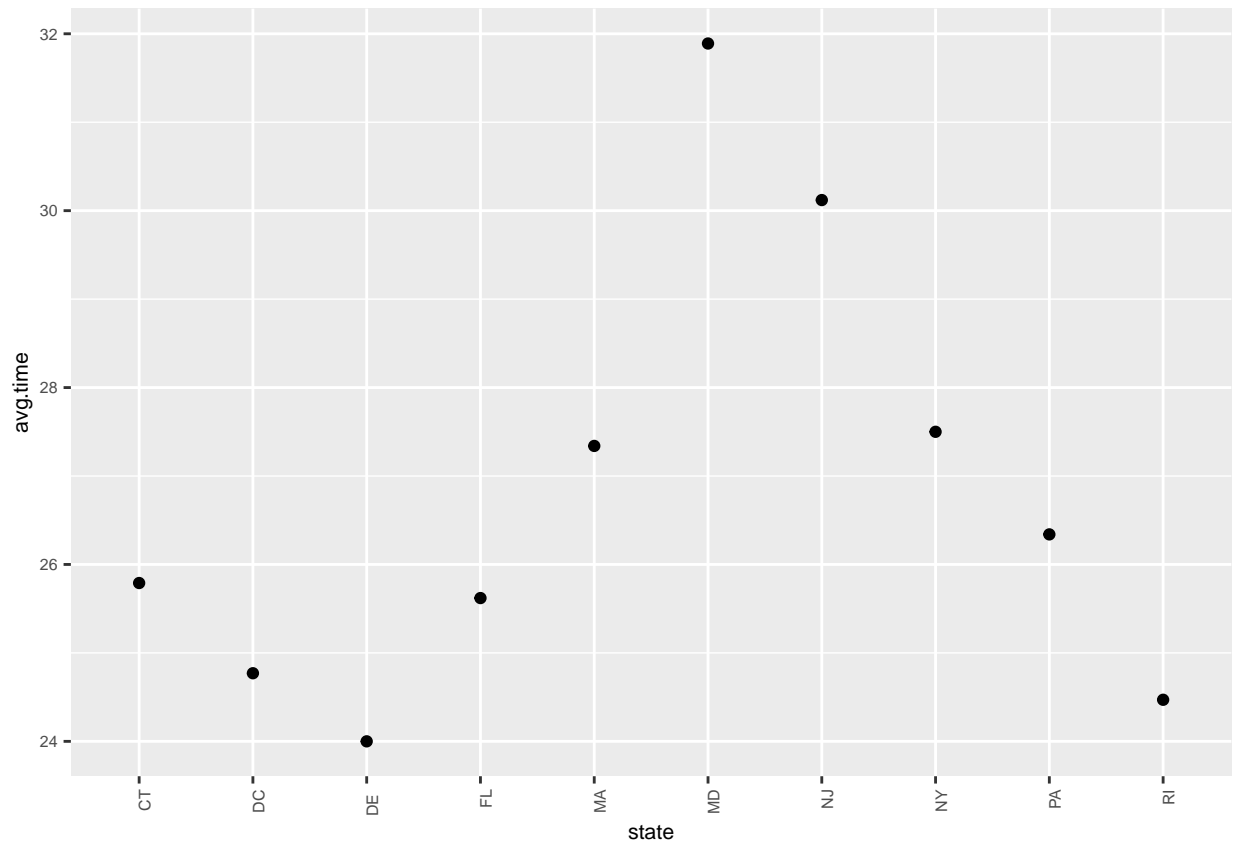
library(ggplot2)
ggplot(data2, aes(x = state, y = avg.time)) + geom_point() + theme(text = element_text(size=8),
  axis.text.x = element_text(angle=90, hjust=1))
```



```
stddev <- sd(data2$avg.time)
cat("Standard Deviation:", stddev, "\n")
```

```
## Standard Deviation: 2.532531
```

```
library(ggplot2)
p <- ggplot(data2, aes(x=state, y=avg.time)) + geom_point() +
  theme(text = element_text(size=8), axis.text.x = element_text(angle=90, hjust=1))
p + geom_smooth(method = "lm", formula = y~x)
```



*#Not sure why the line doesnt show up*

Maryland has the highest commute time out of all states, and the 6th most dense state. However, most of the other dense states have commute times similar to others, because of this, it can be said that there is no correlation between commute times and population density.