

# SF

October 29, 2024

```
knitr::opts_chunk$set(message = FALSE, warning = FALSE)

library(plotly)
library(dplyr)

# helper functions
standardize <- function(vector) {return((vector - mean(vector)) / sd(vector))}
convert_to_binary <- function(input_vector) {return(ifelse(input_vector != 0, 1, 0))}

#####
##### PREPARE THE DATA #####
#####

# loading the data
load("~/Desktop/Spring 2024/Sanfrancisco_data_analysis/Data_files/data_on_graph_with_covariates.rds")

data = data_on_graph_with_covariates %>%
  mutate(across(starts_with(c("class_", "upto")), list(ind = convert_to_binary))) %>%
  mutate(across(c("bus", "signal", "stop", "crossing"), ~round(., 5))) %>%
  mutate(across(c("density_per_hour", "SpeedLimit"), standardize))
```

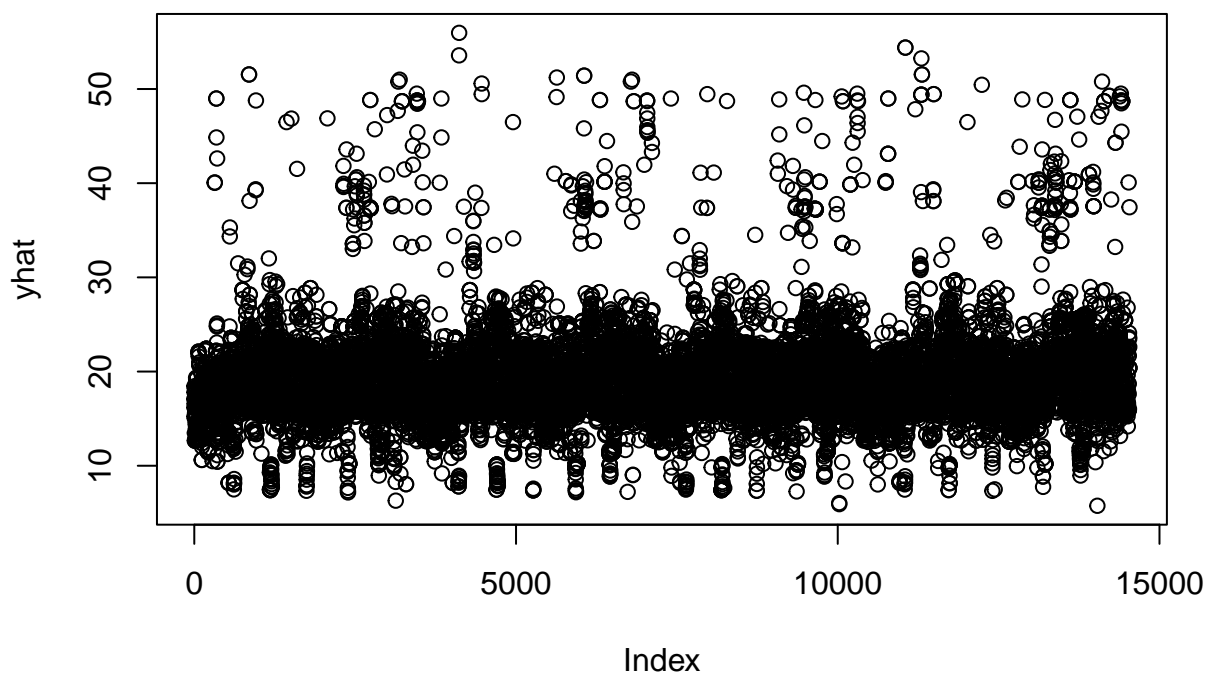
## 0.1 All covariates

```
res = lm(speed ~ SpeedLimit +
  density_per_hour +
  bus +
  signal +
  stop +
  crossing +
  upto1_ind +
  bus_number +
  signal_number +
  stop_number +
  crossing_number,
  data = data)
```

```
summary(res)
```

```
##
## Call:
## lm(formula = speed ~ SpeedLimit + density_per_hour + bus + signal +
##      stop + crossing + upto1_ind + bus_number + signal_number +
##      stop_number + crossing_number, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -53.572  -9.360   0.225   8.455  69.151
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    23.02078    0.20126  114.383 < 2e-16 ***
## SpeedLimit      3.57610    0.11289   31.679 < 2e-16 ***
## density_per_hour 0.62079    0.11235    5.525 3.34e-08 ***
## bus            -7.71252    1.15342   -6.687 2.37e-11 ***
## signal         -4.01630    0.34602  -11.607 < 2e-16 ***
## stop           -5.54105    0.66902   -8.282 < 2e-16 ***
## crossing        -2.92879    0.37166   -7.880 3.49e-15 ***
## upto1_ind       2.24544    0.40205    5.585 2.38e-08 ***
## bus_number      0.81119    0.64318    1.261  0.207
## signal_number   -0.10220    0.14116   -0.724  0.469
## stop_number     -0.35749    0.35374   -1.011  0.312
## crossing_number  0.02104    0.11595    0.181  0.856
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.43 on 14522 degrees of freedom
## Multiple R-squared:  0.1515, Adjusted R-squared:  0.1508
## F-statistic: 235.7 on 11 and 14522 DF, p-value: < 2.2e-16

yhat = res$fitted.values
plot(yhat)
```



## 0.2 Only SpeedLimit

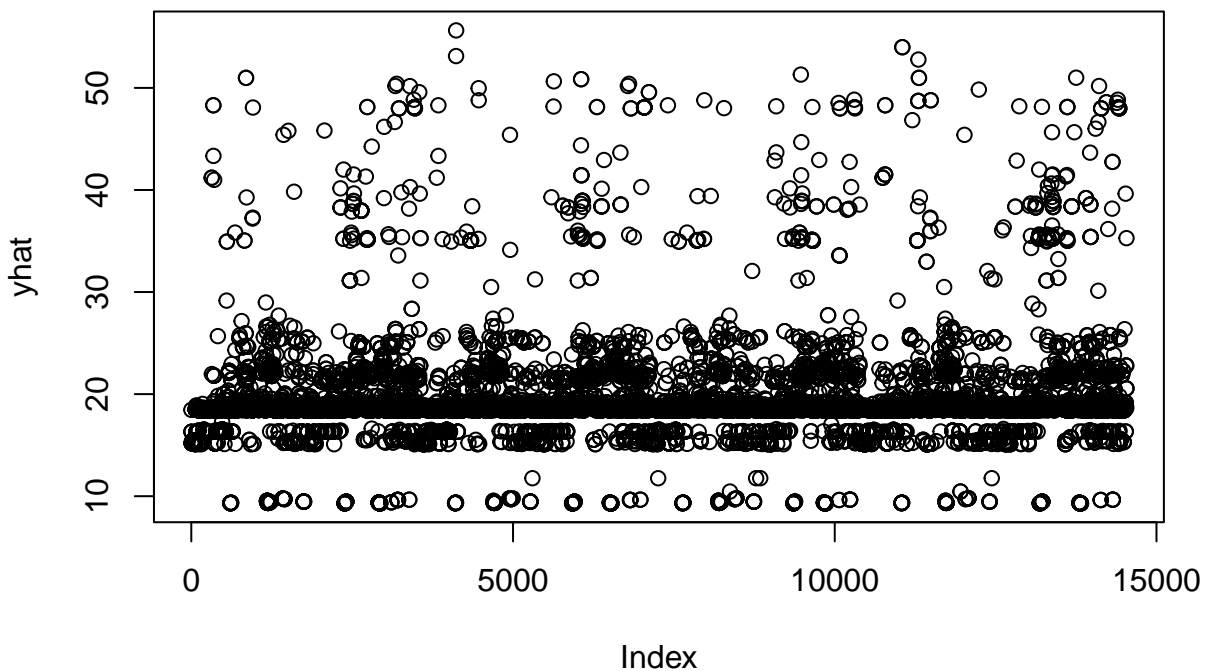
```
res = lm(speed ~ SpeedLimit +
  density_per_hour +
  #bus +
  #signal +
  #stop +
  #crossing +
  upto1_ind,
  #bus_number +
  #signal_number +
  #stop_number +
  #crossing_number,
  data = data)
```

```
summary(res)
```

```
##
## Call:
## lm(formula = speed ~ SpeedLimit + density_per_hour + upto1_ind,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -53.118  -9.884   0.282   8.856  73.070
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)      19.1491      0.1115 171.751 < 2e-16 ***
## SpeedLimit       3.9990      0.1140  35.077 < 2e-16 ***
## density_per_hour 0.6531      0.1130   5.779 7.67e-09 ***
## upto1_ind        2.6232      0.4059   6.463 1.06e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.7 on 14530 degrees of freedom
## Multiple R-squared:  0.114, Adjusted R-squared:  0.1138
## F-statistic: 623 on 3 and 14530 DF, p-value: < 2.2e-16
```

```
yhat = res$fitted.values
plot(yhat)
```



```
res = lm(speed ~ SpeedLimit,
          #density_per_hour +
          #bus +
          #signal +
          #stop +
          #crossing +
          #upto1_ind,
          #bus_number +
          #signal_number +
          #stop_number +
          #crossing_number,
          data = data)
```

```
summary(res)
```

```
##
```

```
## Call:
## lm(formula = speed ~ SpeedLimit, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -48.708  -9.283   0.373   8.420  72.794
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  19.3846     0.1057  183.35  <2e-16 ***
## SpeedLimit    4.4313     0.1057   41.91  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.75 on 14532 degrees of freedom
## Multiple R-squared:  0.1078, Adjusted R-squared:  0.1078
## F-statistic: 1757 on 1 and 14532 DF, p-value: < 2.2e-16

yhat = res$fitted.values
plot(yhat)
```

