

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.RData")

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)
```

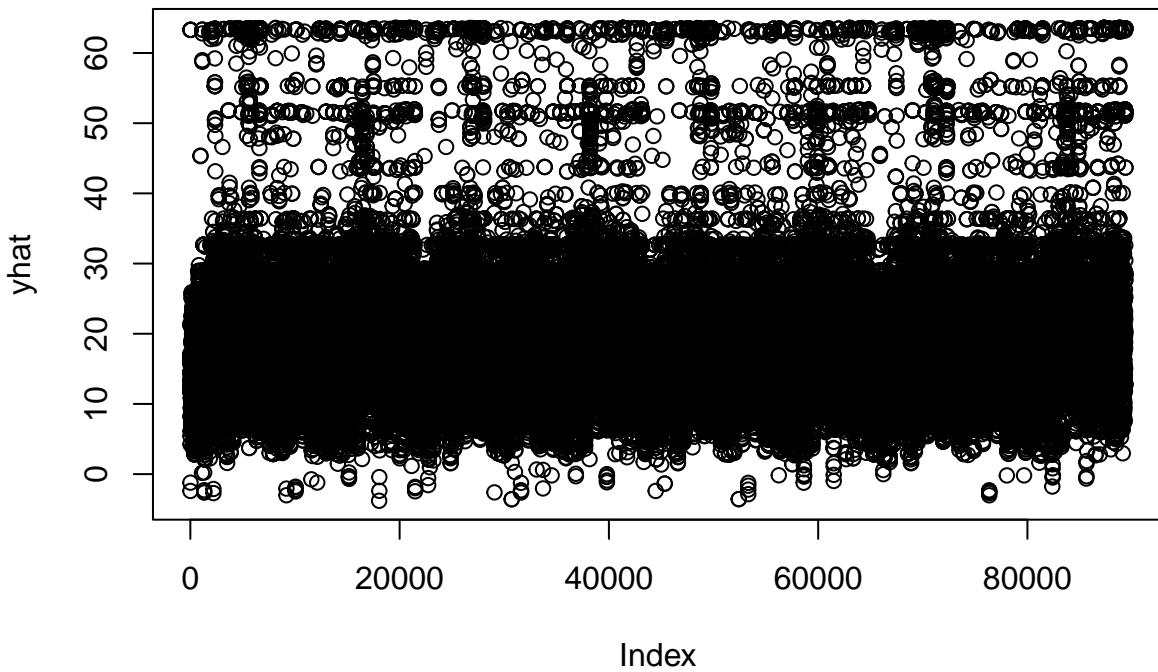
```

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 
## -63.545 -10.888   0.537  10.115  86.876 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 26.68057  0.08470 314.992 < 2e-16 ***
## SpeedLimit   5.58249  0.05709  97.783 < 2e-16 ***
## density_per_hour -0.28492  0.05250 -5.427 5.75e-08 ***
## bus          -8.31740  0.21816 -38.125 < 2e-16 ***
## signal        -6.61562  0.15839 -41.768 < 2e-16 ***
## stop          -6.58352  0.31567 -20.855 < 2e-16 ***
## crossing      -5.53488  0.16431 -33.685 < 2e-16 ***
## upto1_ind     8.25788  0.24119  34.238 < 2e-16 ***
## bus_number    0.06019  0.10605  0.568  0.5703    
## signal_number -0.15203  0.06838 -2.223  0.0262 *  
## stop_number    0.77877  0.14183  5.491 4.01e-08 *** 
## crossing_number 0.61584  0.05134 11.995 < 2e-16 *** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.43 on 89356 degrees of freedom
## Multiple R-squared:  0.2872, Adjusted R-squared:  0.2871 
## F-statistic: 3274 on 11 and 89356 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)

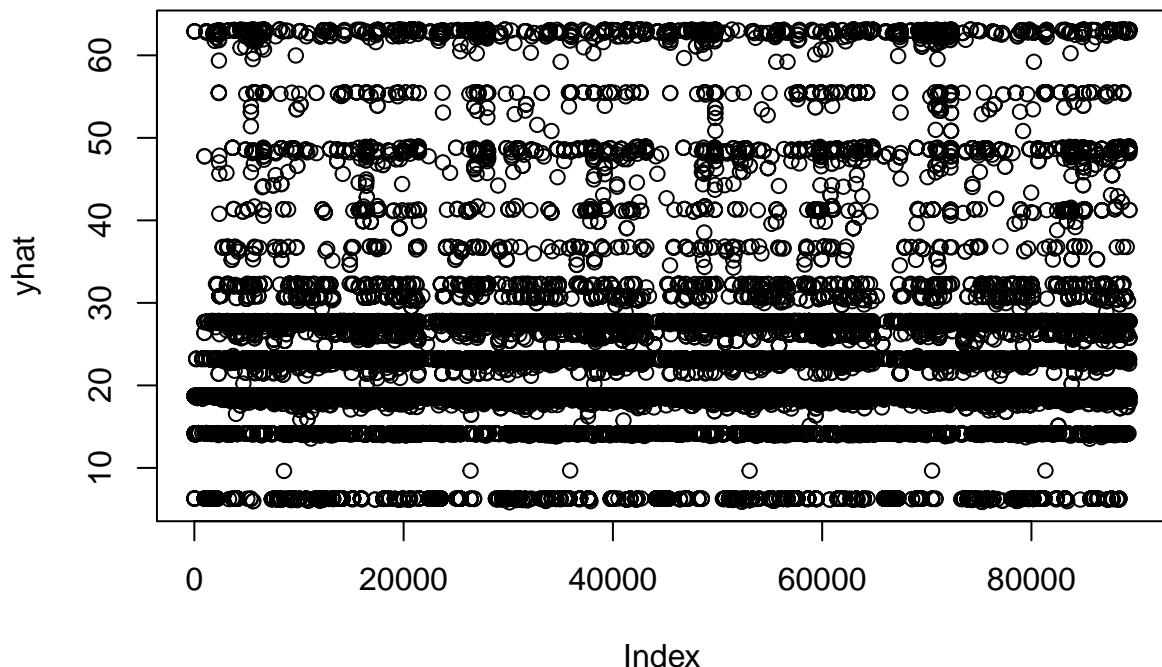
##
## Call:
## lm(formula = speed ~ SpeedLimit + density_per_hour + upto1_ind,
##      data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -63.115 -12.482    0.593  10.698  87.810 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  10.698    1.000  10.698  <2e-16 ***
## SpeedLimit -0.00015  0.00005 -3.000  0.0025 ** 
## density_per_hour 0.00015  0.00005  3.000  0.0025 ** 
## upto1_ind     0.00015  0.00005  3.000  0.0025 ** 
## bus           0.00015  0.00005  3.000  0.0025 ** 
## signal        0.00015  0.00005  3.000  0.0025 ** 
## stop          0.00015  0.00005  3.000  0.0025 ** 
## crossing      0.00015  0.00005  3.000  0.0025 ** 
## ---
## Signif. codes:  *** 0.001 ** 0.01 * 0.05 . 0.1 
## 
```

```

## (Intercept) 20.41619   0.05300 385.181 < 2e-16 ***
## SpeedLimit    6.84201   0.05789 118.195 < 2e-16 ***
## density_per_hour -0.24770   0.05495 -4.507 6.57e-06 ***
## upto1_ind      7.65322   0.25392 30.140 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.24 on 89364 degrees of freedom
## Multiple R-squared: 0.2044, Adjusted R-squared: 0.2043
## F-statistic: 7652 on 3 and 89364 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 
## -59.584 -12.449   0.426  10.982  87.149 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 20.85268   0.05125 406.9 <2e-16 ***
## SpeedLimit   7.56912   0.05125 147.7 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.32 on 89366 degrees of freedom
## Multiple R-squared:  0.1962, Adjusted R-squared:  0.1962 
## F-statistic: 2.181e+04 on 1 and 89366 DF,  p-value: < 2.2e-16

yhat = res$fitted.values
plot(yhat)

```

