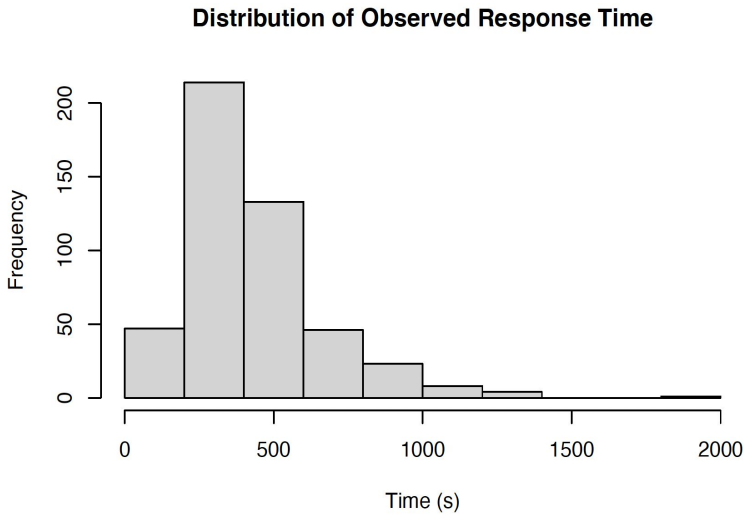
EMS Station EDA

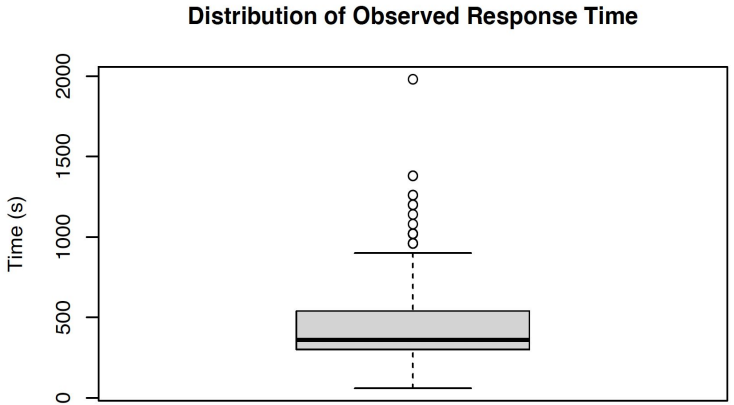
Group E: Ken Ye, Ejay Lin, Gorden Gao

Observed Response Time (Histogram)



- Removed NA and 0 values
- Unimodal
- Right-skewed
- We can see that the most common observed response time is at around 250-500 seconds, with a few outliers between 1750-2000 seconds.

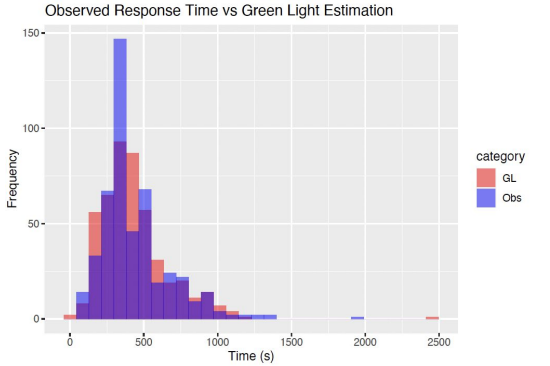
Observed Response Time Cont. (Box Plot)

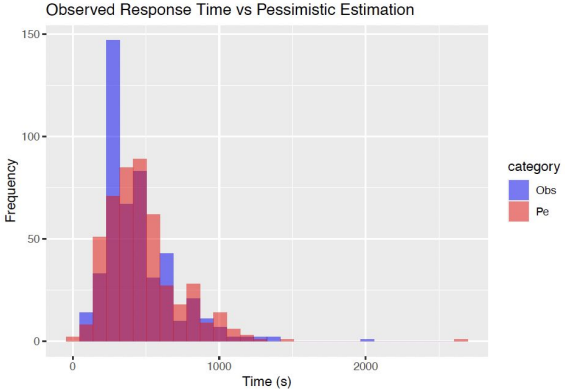


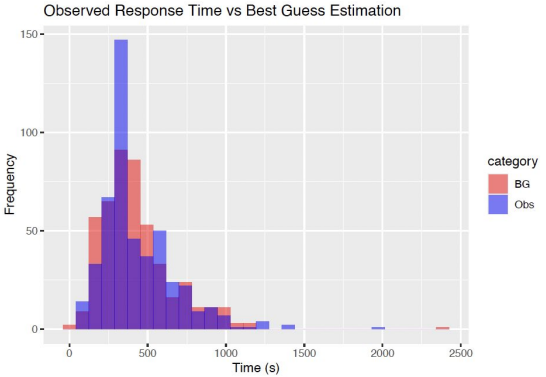
- We can see that the median observed response time is at about 350 seconds, and interquartile range is around 300-550 seconds.
- There are a few outliers above the range 60-950 seconds.

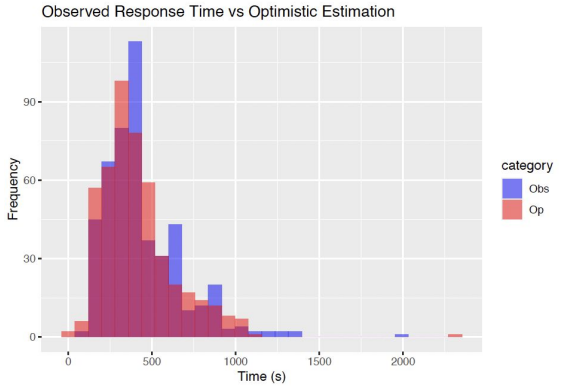
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
60	300	360	432	540	1980

Comparison with Estimated Response Time

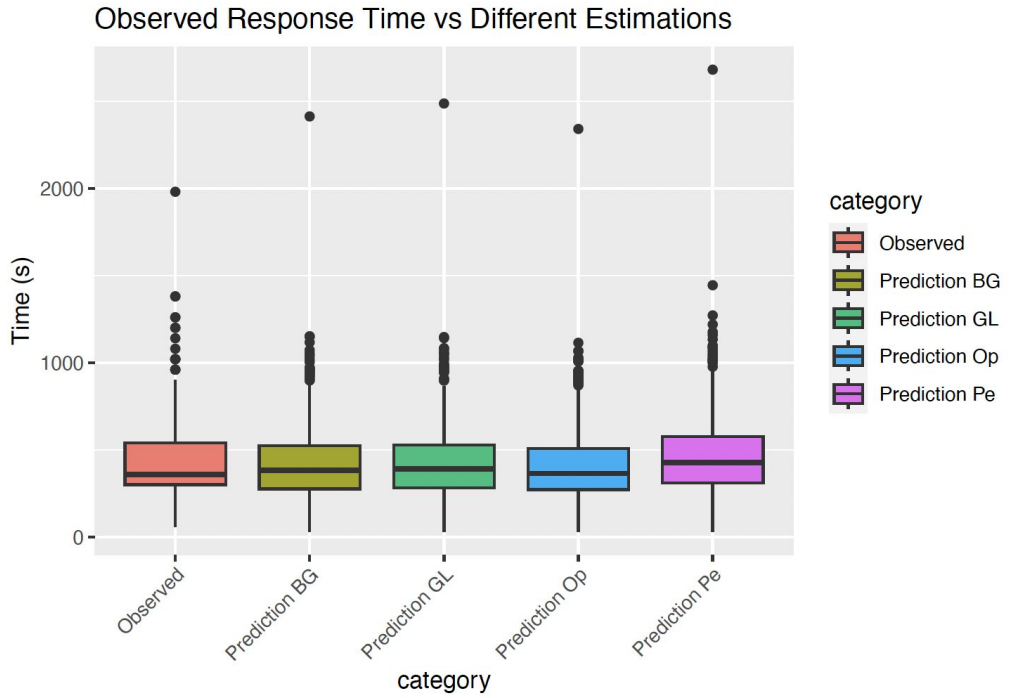








Visualization Cont.



RMSE

RMSE

```
rmse1 <- sqrt(mean((observed_df$value - pred1_df$value)^2))
rmse2 <- sqrt(mean((observed_df$value - pred2_df$value)^2))
rmse3 <- sqrt(mean((observed_df$value - pred3_df$value)^2))
rmse4 <- sqrt(mean((observed_df$value - pred4_df$value)^2))

print(rmse1)
```

```
## [1] 189.8238
```

```
print(rmse2)
```

```
## [1] 212.1679
```

```
print(rmse3)
```

```
## [1] 188.9788
```

```
print(rmse4)
```

```
## [1] 183.6995
```

The optimistic Google Map API has the best estimation among the four, based on the RMSE criteria.

Linear Regression (Observed vs Optimistic Est.)

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1369.36   -86.97   -20.68    56.54  1440.51
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    136.99639   16.41909   8.344 7.92e-16 ***
## pred4_df$value   0.70614    0.03454  20.443  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 170.9 on 474 degrees of freedom
## Multiple R-squared:  0.4686, Adjusted R-squared:  0.4674
## F-statistic: 417.9 on 1 and 474 DF,  p-value: < 2.2e-16
```