

London Fire Service Statistical Summary

Nithish Vengadesan

Section 1

This report fulfills the requests of Fire service managers and local politicians, performing the specific analysis

This data comes from the London Fire Brigade. A panel of Fire service managers and local politicians want to better understand some particular aspects of the costs and response times associated with incidents during the time period in the data.

Read Data

```
#Import Required Libraries
library(tidyverse)
library(dplyr)
library(gridExtra)
library(nycflights13)
library(emmeans)

options(width=100)
set.seed(9999) # Set the random seed to make sure you get the same results

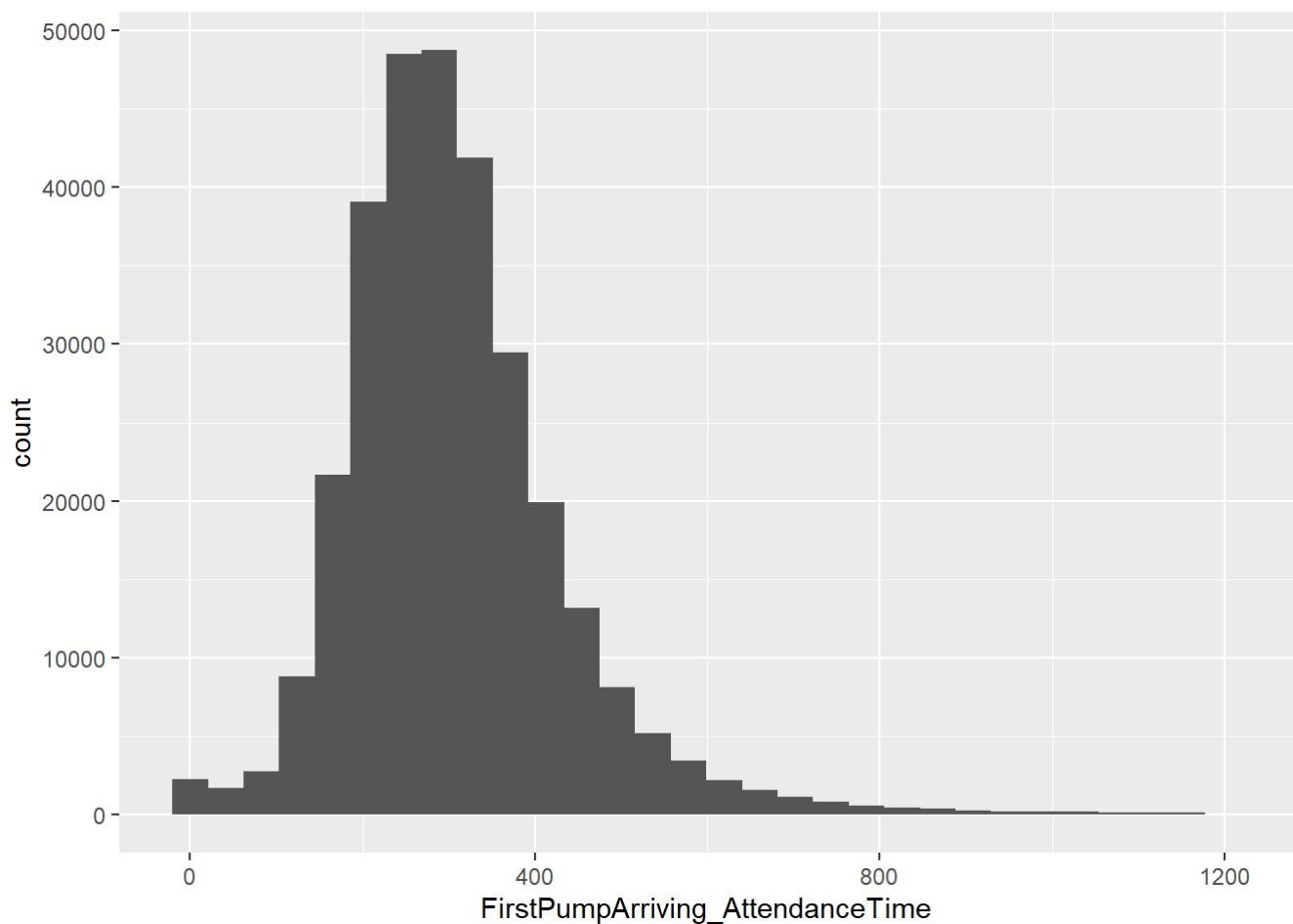
#Import Data Set
fire_data = read_csv("London_Fire_data.csv")
summary(fire_data)
```

## IncidentNumber	DateOfCall	CalYear	TimeOfCall	HourOfCall
## Length:322375	Length:322375	Min. :2019	Length:322375	Min. : 0.00
## Class :character	Class :character	1st Qu.:2019	Class1:hms	1st Qu.: 9.00
## Mode :character	Mode :character	Median :2020	Class2:difftime	Median :14.00
##		Mean :2020	Mode :numeric	Mean :13.42
##		3rd Qu.:2021		3rd Qu.:19.00
##		Max. :2022		Max. :23.00
##				
## IncidentGroup	StopCodeDescription	SpecialServiceType	PropertyCategory	PropertyType
## Length:322375	Length:322375	Length:322375	Length:322375	Length:322375
5				
## Class :character	Class :character	Class :character	Class :character	Class :character
## Mode :character	Mode :character	Mode :character	Mode :character	Mode :character
##				
##				
##				
##				
## AddressQualifier	Postcode_full	Postcode_district	UPRN	USRN
## Length:322375	Length:322375	Length:322375	Min. :0.000e+00	Min. : 420
0740				
## Class :character	Class :character	Class :character	1st Qu.:0.000e+00	1st Qu.:2040
0989				
## Mode :character	Mode :character	Mode :character	Median :0.000e+00	Median :2120
1121				
##			Mean :2.072e+10	Mean :2040
0837				
##			3rd Qu.:1.001e+10	3rd Qu.:2210
0813				
##			Max. :2.000e+11	Max. :9999
0422				
##				
## IncGeo_BoroughCode	IncGeo_BoroughName	ProperCase	IncGeo_WardCode	IncGeo_WardName
## Length:322375	Length:322375	Length:322375	Length:322375	Length:322375
## Class :character	Class :character	Class :character	Class :character	Class :character
## Mode :character	Mode :character	Mode :character	Mode :character	Mode :character
##				
##				
##				
##				
## IncGeo_WardNameNew	Easting_m	Northing_m	Easting_rounded	Northing_rounded
## Length:322375	Min. :503582	Min. :155998	Min. :503550	Min. :155950
## Class :character	1st Qu.:524924	1st Qu.:175804	1st Qu.:525150	1st Qu.:176050
## Mode :character	Median :530858	Median :180978	Median :530950	Median :181050
##	Mean :530634	Mean :180340	Mean :530667	Mean :180487
##	3rd Qu.:537035	3rd Qu.:185076	3rd Qu.:536350	3rd Qu.:185250
##	Max. :560461	Max. :200885	Max. :611150	Max. :302450
##	NA's :175667	NA's :175667		
##				
## Latitude	Longitude	FRS	IncidentStationGround	
## Min. : 0.00	Min. :-0.51	Length:322375	Length:322375	
## 1st Qu.:51.47	1st Qu.: -0.20	Class :character	Class :character	

```
## Median :51.51      Median :-0.12      Mode :character      Mode :character
## Mean   :51.36      Mean   :-0.12
## 3rd Qu.:51.55      3rd Qu.: -0.03
## Max.   :51.69      Max.    : 0.31
## NA's   :175667     NA's    :175667
## FirstPumpArriving_AttendanceTime FirstPumpArriving_DeployedFromStation
## Min.    : 1.0                      Length:322375
## 1st Qu.: 227.0                    Class :character
## Median  : 290.5                    Mode  :character
## Mean    : 308.1
## 3rd Qu.: 367.0
## Max.    :1199.0
## NA's    :19019
## SecondPumpArriving_AttendanceTime SecondPumpArriving_DeployedFromStation
## Min.    : 1.0                      Length:322375
## 1st Qu.: 293.0                    Class :character
## Median  : 363.0                    Mode  :character
## Mean    : 385.6
## 3rd Qu.: 450.0
## Max.    :1200.0
## NA's    :199385
## NumStationsWithPumpsAttending NumPumpsAttending PumpCount PumpHoursRoundUp
## Min.    : 1.0                      Min.    : 1.000      Min.    : 1.000      Min.    : 1.00
## 1st Qu.: 1.0                      1st Qu.: 1.000      1st Qu.: 1.000      1st Qu.: 1.00
## Median  : 1.0                      Median   : 1.000      Median   : 1.000      Median   : 1.00
## Mean    : 1.4                      Mean     : 1.571      Mean     : 1.619      Mean     : 1.37
## 3rd Qu.: 2.0                      3rd Qu.: 2.000      3rd Qu.: 2.000      3rd Qu.: 1.00
## Max.    :14.0                      Max.     :14.000      Max.     :250.000      Max.     :1203.00
## NA's    :3823                      NA's     :3823       NA's     :2008       NA's     :2111
## Notional Cost (£)      NumCalls
## Min.    : 333.0      Min.    : 1.000
## 1st Qu.: 339.0      1st Qu.: 1.000
## Median  : 346.0      Median   : 1.000
## Mean    : 471.9      Mean     : 1.306
## 3rd Qu.: 352.0      3rd Qu.: 1.000
## Max.    :407817.0    Max.     :175.000
... ..
```

Continuous Variable Visualisation

```
#check for data quality and outliers
ggplot(fire_data) +geom_histogram(aes(`FirstPumpArriving_AttendanceTime`))
```



The data is continuous and only few outliers or null value is present in few response which can be replaced with zero or removed.

Analysis of Cost on different Incidents

```
cost_data <- fire_data%>%
  group_by(IncidentGroup)%>%
  summarise(total_cost = sum(`Notional Cost (£)`, na.rm = TRUE), mean_cost = mean(`Notional Cost (£)`, na.rm = TRUE), Number_of_Alarms=n())

print(cost_data)
```

```
## # A tibble: 3 × 4
##   IncidentGroup total_cost mean_cost Number_of_Alarms
##   <chr>          <dbl>    <dbl>          <int>
## 1 False Alarm    61249812    378.          162299
## 2 Fire          43059576    838.           51443
## 3 Special Service 46812402    437.          108633
```

Distribution of Response time

```
fire_data <- mutate(fire_data, response_time =FirstPumpArriving_AttendanceTime ) %>%filter(!is.na(response_time))

mean_time_incidentgroup <- fire_data %>% group_by(IncidentGroup) %>% summarise(mean=mean(response_time,na.rm=T ),Frequency=n())

overall_meantime = mean(fire_data$FirstPumpArriving_AttendanceTime)
print(overall_meantime)
```

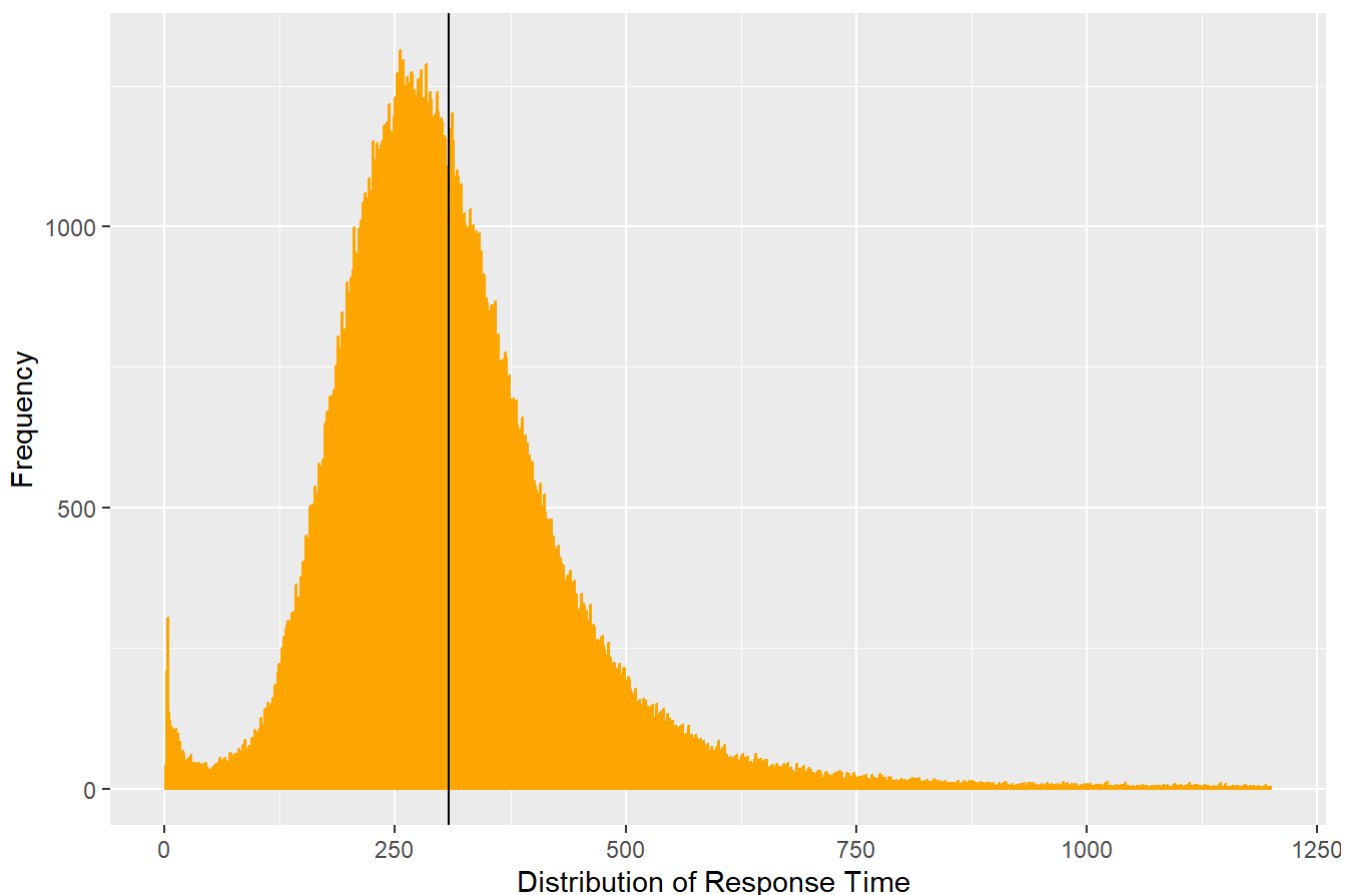
```
## [1] 308.0578
```

```
print(mean_time_incidentgroup)
```

```
## # A tibble: 3 × 3
##   IncidentGroup   mean Frequency
##   <chr>         <dbl>     <int>
## 1 False Alarm    299.     160995
## 2 Fire          319.       50921
## 3 Special Service 318.       91440
```

```
ggplot(data = fire_data, aes(FirstPumpArriving_AttendanceTime)) +
  geom_histogram(binwidth = 1, na.rm = TRUE , color = "orange") + geom_vline(xintercept = mean(
  fire_data$FirstPumpArriving_AttendanceTime, na.rm = T), color="black")+
  labs(x = "Distribution of Response Time ", y = "Frequency",title=" Distribution of Overall Mean Response Time")
```

Distribution of Overall Mean Response Time



Visualisation of Response time with respect to Incident Groups

```
ggplot(data = fire_data, aes(FirstPumpArriving_AttendanceTime, color = IncidentGroup)) +  
  geom_histogram(binwidth = 1, na.rm = TRUE ) +  
  geom_vline(xintercept = mean(fire_data$FirstPumpArriving_AttendanceTime, na.rm = T), color="black")+  
  facet_grid(IncidentGroup~.) +  
  xlim(0,700)+  
  labs(x = "Distribution of Response Time ", y = "Frequency of Incidents", title = "Mean Response  
Time for Different Incidents")
```



Summary Table of special service response times

```
spl_services <- fire_data[!is.na(fire_data$SpecialServiceType),]

spl_services<-spl_services %>%
  filter(IncidentGroup=="Special Service")%>%
  group_by(SpecialServiceType) %>%
  summarise(Frequency = n (), Mean_Response_Time = mean(FirstPumpArriving_AttendanceTime, na.rm = TRUE), Quantile10 = quantile(FirstPumpArriving_AttendanceTime, 0.10, na.rm = TRUE), Quantile90 = quantile(FirstPumpArriving_AttendanceTime, 0.90,na.rm = TRUE))

print(spl_services)
```

```
## # A tibble: 20 × 5
##   SpecialServiceType      Frequency Mean_Response_Time Quantile10 Quantile90
##   <chr>                <int>          <dbl>         <dbl>      <dbl>
## 1 Advice Only          1817            320.         188        462.
## 2 Animal assistance incidents 2157            341.         190        517.
## 3 Assist other agencies    4339            317.         182        462
## 4 Effecting entry/exit    23300            317.         182        470
## 5 Evacuation (no fire)     656            327.         192.       474.
## 6 Flooding             20428            328.         191        485
## 7 Hazardous Materials incident 2518            316.         185        463.
## 8 Lift Release          4506            307.         178        450
## 9 Making Safe (not RTC)    3344            321.         177        494
## 10 Medical Incident       2080            222.          8        433
## 11 No action (not false alarm) 7572            328.         188        485
## 12 Other rescue/release of persons 1196            331.         190.       482
## 13 Other Transport incident   858            318.         151.       491
## 14 Removal of objects from people 1526            208.          7        456
## 15 Rescue or evacuation from water 160            324.         176.       484.
## 16 RTC                 12077            318.         157        500
## 17 Spills and Leaks (not RTC) 2043            350.         195.       530
## 18 Stand By             154            309.         165.       472.
## 19 Suicide/attempts        707            323.         181        482
## 20 Water provision         1            245          245        245
```

Summary table of total responses

```
(fire_data.summary <- fire_data %>% summarise(mean_time=mean(response_time, na.rm=TRUE), sd_time=sd(response_time, na.rm=TRUE), N_time=n()))
```

```
## # A tibble: 1 × 3
##   mean_time sd_time N_time
##   <dbl>    <dbl> <int>
## 1    308.    129. 303356
```

Filtering out responses of Ealing and Greenwich

```
fire_data.E.G <- fire_data %>% filter(IncGeo_BoroughName %in% c("EALING", "GREENWICH"), na.rm
=TRUE)
#filtering Ealing and greenwich

fire_data.E.G.summarise<-fire_data.E.G %>% group_by(IncGeo_BoroughName) %>% summarise(Mean=m
ean(response_time , na.rm=TRUE), Frequency = n()) #summary table for Ealing and Greenwich

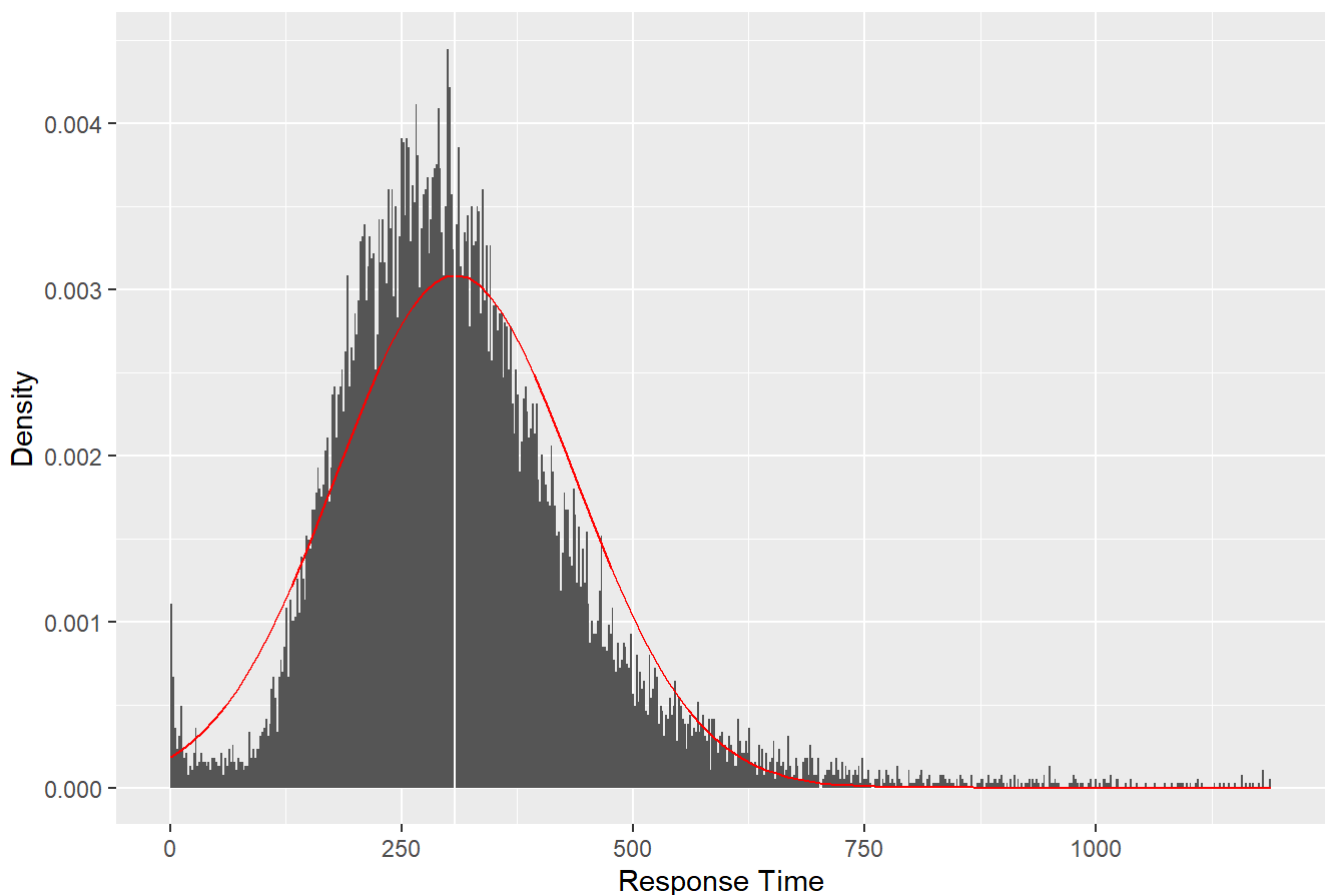
print(fire_data.E.G.summarise)
```

```
## # A tibble: 2 × 3
##   IncGeo_BoroughName Mean Frequency
##   <chr>             <dbl>     <int>
## 1 EALING             317.     10323
## 2 GREENWICH          311.      9124
```

Overall Response time graph

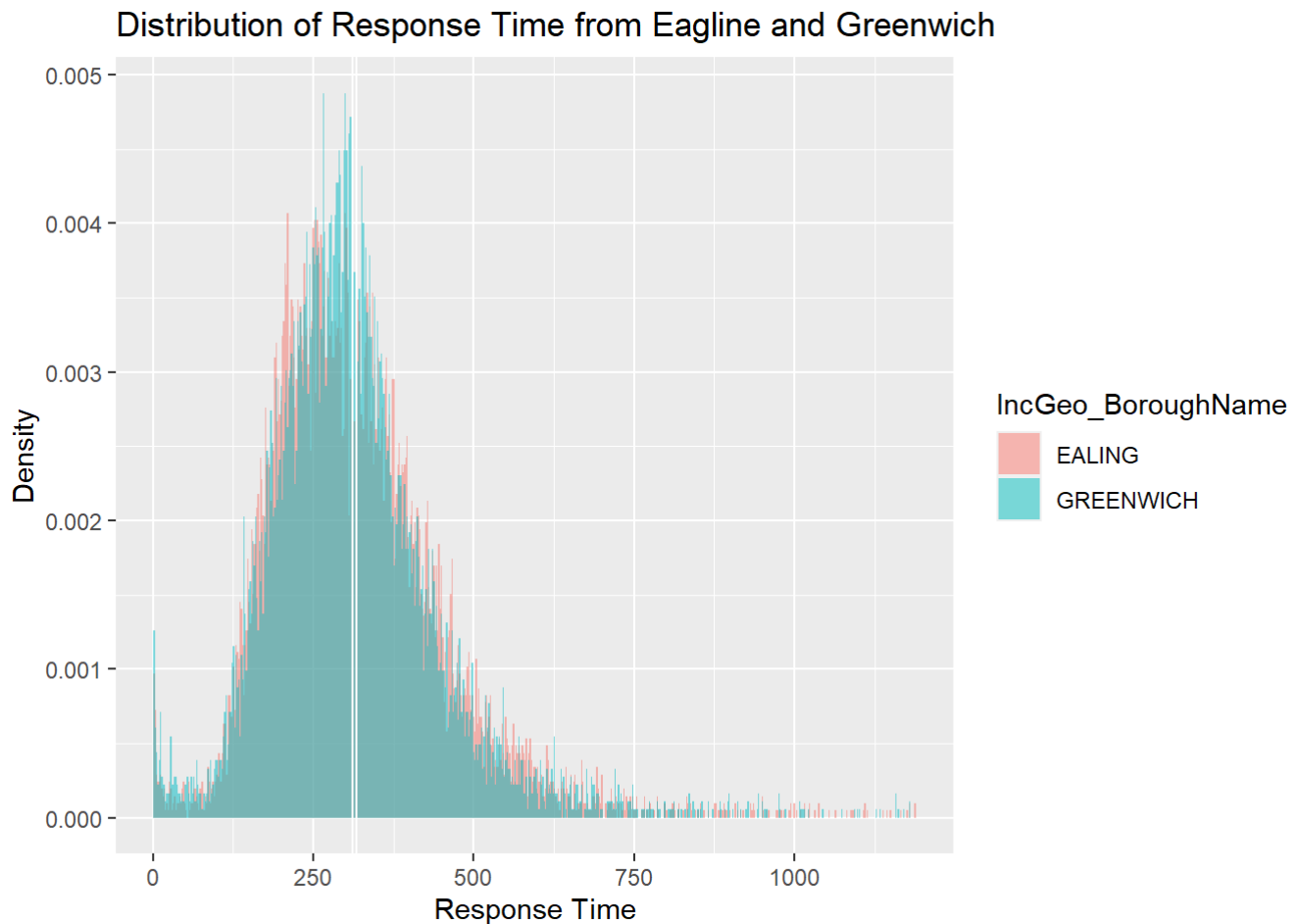
```
ggplot(fire_data.E.G, aes(x=response_time)) +
  geom_histogram(aes(y=..density..), binwidth=2) +
  stat_function(fun=function(x) {dnorm(x, mean=fire_data.summary$mean_time, sd=fire_data.su
mmmary$sd_time)}, col="red") + geom_vline(data=fire_data.summary, mapping=aes(xintercept=fire_
data.summary$mean_time), col="white") +
  labs(x="Response Time", y="Density", title="Distribution of Overall Response Time Graph")
```

Distribution of Overall Response Time Graph



Distribution of Response Time from Eagline and Greenwich

```
ggplot(fire_data.E.G, aes(response_time,..density.., fill=IncGeo_BoroughName)) + geom_histogram(binwidth=2,position="identity", alpha=.5) + labs(x="Response Time", y="Density", fill="Inc Geo_BoroughName", title="Distribution of Response Time from Eagline and Greenwich") + geom_vline(data=fire_data.E.G.summarise, mapping=aes(xintercept=fire_data.E.G.summarise$Mean), col="white")
```



Performing Two Sample T-Test

```
t_test<-t.test(  
  response_time~IncGeo_BoroughName,  
  data=fire_data.E.G  
)  
  
print(t_test)
```

```
##
## Welch Two Sample t-test
##
## data: response_time by IncGeo_BoroughName
## t = 2.8542, df = 19303, p-value = 0.00432
## alternative hypothesis: true difference in means between group EALING and group GREENWICH
is not equal to 0
## 95 percent confidence interval:
## 1.739777 9.368041
## sample estimates:
## mean in group EALING mean in group GREENWICH
## 316.9342 311.3803
```

Estimation

```
m.rt.by.IncGeo_BoroughName <- lm(response_time~IncGeo_BoroughName, data=fire_data.E.G)
( m.rt.by.IncGeo_BoroughName.emm <- emmeans(m.rt.by.IncGeo_BoroughName, ~IncGeo_BoroughName)
)
```

```
## IncGeo_BoroughName emmean SE df lower.CL upper.CL
## EALING 317 1.34 19445 314 320
## GREENWICH 311 1.42 19445 309 314
##
## Confidence level used: 0.95
```

```
( m.rt.by.IncGeo_BoroughName.contrast <- confint(pairs(m.rt.by.IncGeo_BoroughName.emm)) )
```

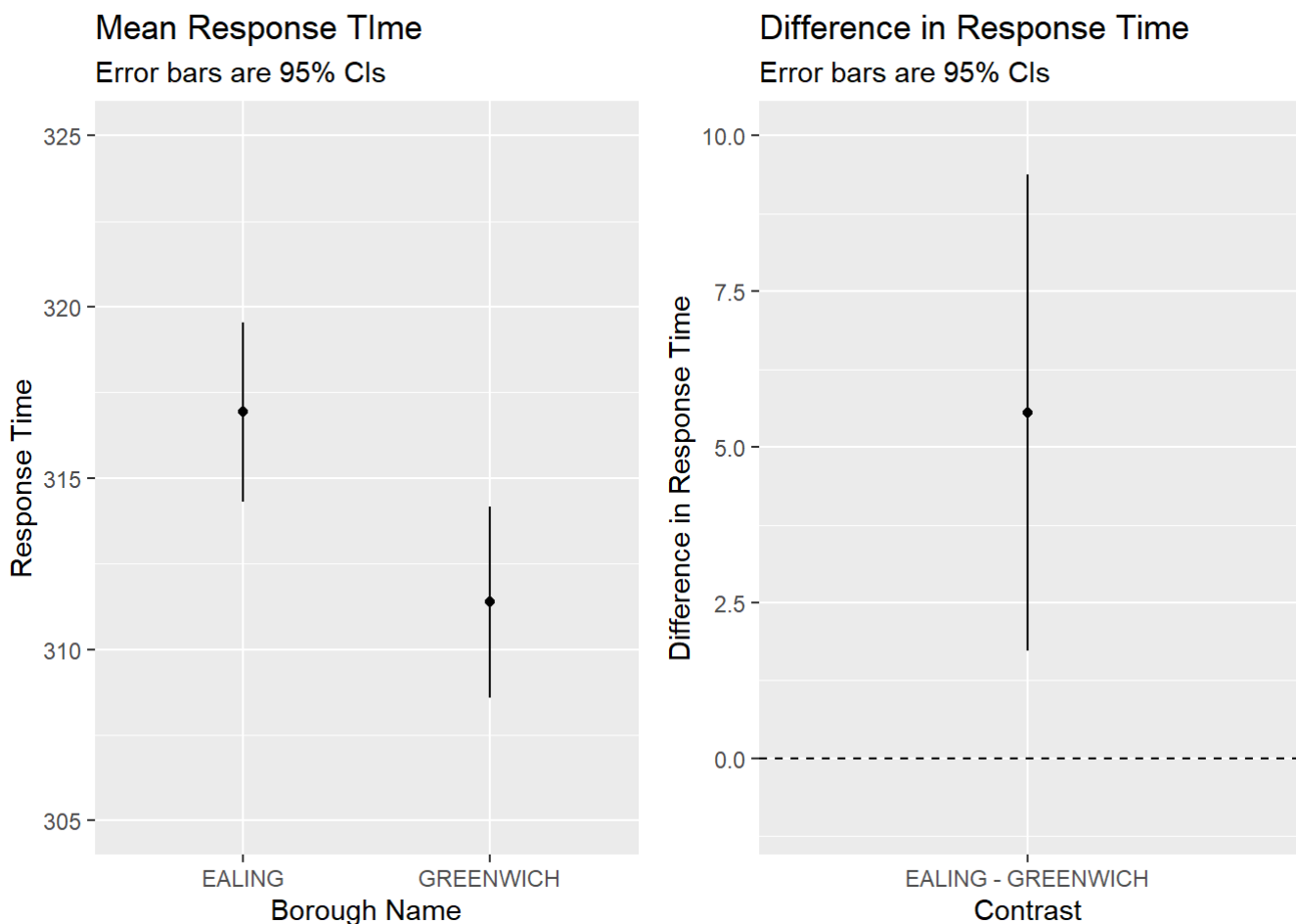
```
## contrast estimate SE df lower.CL upper.CL
## EALING - GREENWICH 5.55 1.95 19445 1.73 9.38
##
## Confidence level used: 0.95
```

Side-by-side plots CIs for the estimates for each group as well as the CI for the difference between groups

```

grid.arrange(
  ggplot(summary(m.rt.by.IncGeo_BoroughName.emm), aes(x=IncGeo_BoroughName, y=emmean, ymin=
lower.CL, ymax=upper.CL)) +
    geom_point() + geom_linerange() +
    labs(y="Response Time", x="Borough Name", subtitle="Error bars are 95% CIs", title="M
ean Response Time") + ylim(305,325),
  ggplot(m.rt.by.IncGeo_BoroughName.contrast, aes(x=contrast, y=estimate, ymin=lower.CL, ym
ax=upper.CL)) +
    geom_point() + geom_linerange() +
    labs(y="Difference in Response Time", x="Contrast", subtitle="Error bars are 95% CI
s", title="Difference in Response Time") + ylim(-1,10) +
    geom_hline(yintercept=0, lty=2),
  ncol=2
)

```



Section 2

This report presents the results of the analyses requested by the board. This used the data provided is London_Fire_data with almost 322375 incidents and 39 different variables. There was a small amount of missing data or inaccurate data entry the data was being replaced by zero or removed.

1) The costs of responding to fires and false alarms:

Total cost associated with all indecents including Fires, False and Speacial Services : 151121790GBP

Total Cost associated with actual Fires : 43059576 GBP

Total Cost associated with False Alarms : 61249812 GBP

Average cost spent on actual Fires : 837.9795 GBP

Average Cost Spent on False Alarms : 378.3796 GBP

Cost spent on False Fire Alarm is high as we get a lot of False Alarms counting to 162299

```
## # A tibble: 3 × 4
##   IncidentGroup total_cost mean_cost Number_of_Alarms
##   <chr>          <dbl>    <dbl>         <int>
## 1 False Alarm    61249812    378.         162299
## 2 Fire          43059576    838.          51443
## 3 Special Service 46812402    437.         108633
```

2) The distribution of response times

Overall Response time of all indecents including Fires, False and Speacial Services : 308.0578 sec

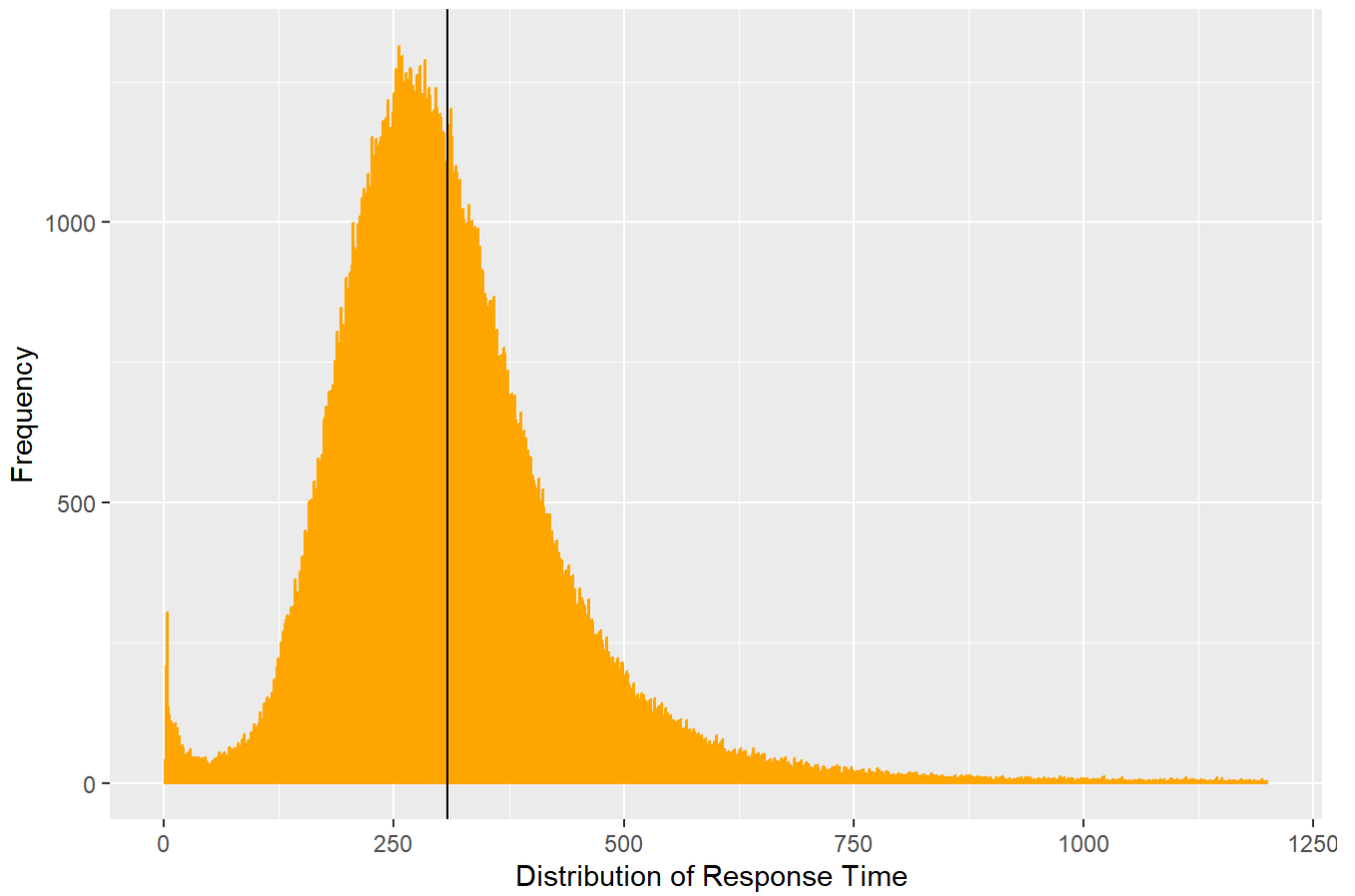
The Mean Response time for actual Fires : 319.4506 seconds

The Mean Response time for False Alarm : 298.9196 seconds

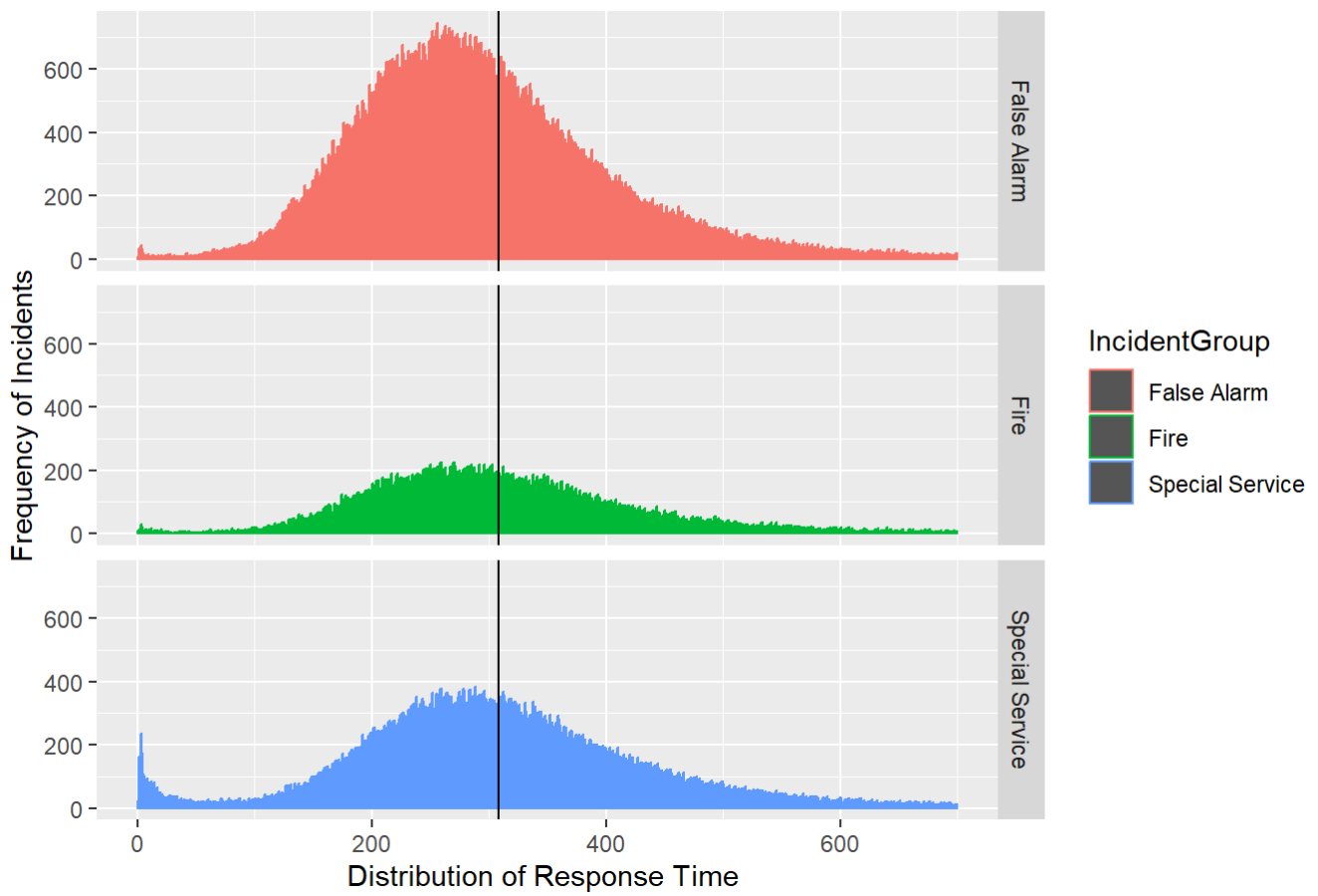
The Mean Response time for Special Services : 317.8027 seconds

The Graph depicts the Distribution of Response time for different Incidents and the black line indicates the mean value of overall Response time.

Overall Mean Response Time



Mean Response Time for Different Incidents



```
## # A tibble: 3 × 3
##   IncidentGroup    mean Frequency
##   <chr>          <dbl>     <int>
## 1 False Alarm    299.    160995
## 2 Fire          319.     50921
## 3 Special Service 318.     91440
```

Inference from Average Response time:

Avg Response time of Fires > Avg Response time of Special Services > Avg Response time of False Alarm

The Response time for Fires and Special Services does not vary much than that to False Alarm

3) Summary of special service response times

The count,mean response time ,10th percentile and 90th percentile for each type of incidents is summarized in table below for Special Services:

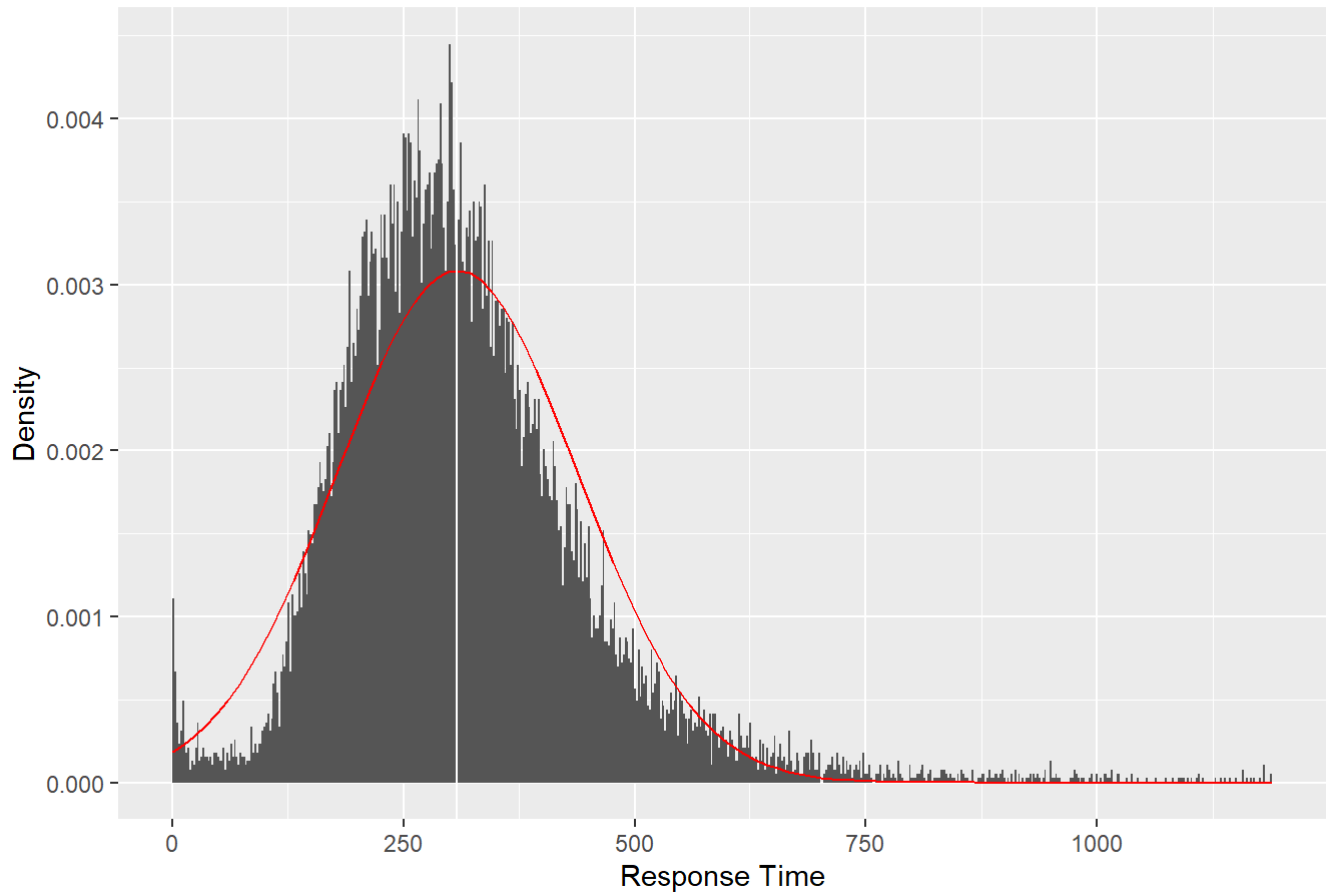
```
## # A tibble: 20 × 5
##   SpecialServiceType    Frequency Mean_Response_Time Quantile10 Quantile90
##   <chr>              <int>          <dbl>      <dbl>      <dbl>
## 1 Advice Only        1817            320.       188       462.
## 2 Animal assistance incidents 2157            341.       190       517.
## 3 Assist other agencies 4339            317.       182       462
## 4 Effecting entry/exit 23300            317.       182       470
## 5 Evacuation (no fire)   656            327.       192.      474.
## 6 Flooding            20428            328.       191       485
## 7 Hazardous Materials incident 2518            316.       185       463.
## 8 Lift Release         4506            307.       178       450
## 9 Making Safe (not RTC) 3344            321.       177       494
## 10 Medical Incident     2080            222.         8       433
## 11 No action (not false alarm) 7572            328.       188       485
## 12 Other rescue/release of persons 1196            331.       190.      482
## 13 Other Transport incident   858            318.       151.      491
## 14 Removal of objects from people 1526            208.         7       456
## 15 Rescue or evacuation from water 160            324.       176.      484.
## 16 RTC               12077            318.       157       500
## 17 Spills and Leaks (not RTC) 2043            350.       195.      530
## 18 Stand By           154            309.       165.      472.
## 19 Suicide/attempts     707            323.       181       482
## 20 Water provision         1            245        245       245
```

Conclusions drawn from the table :

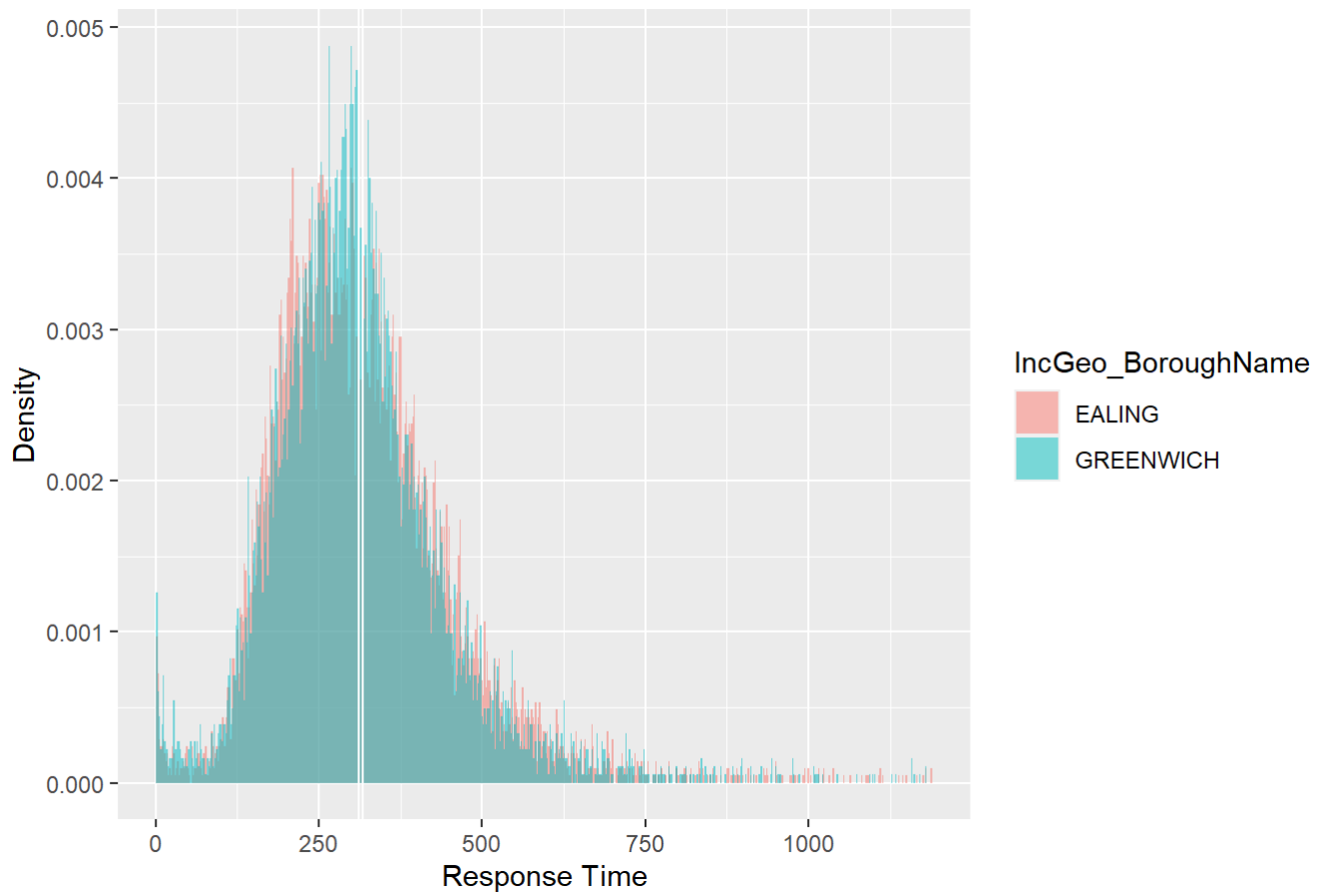
1. Effective entry/exit has the highest frequency of 23300.
- 2.Split and Leaks(not RTC)has the highest mean response time 350.3152 seconds and 50th percentile response time 530.0
- 3.Water Provision has the highest 10th percentile response time with 245.0 seconds

4) A t-test comparing Ealing and Greenwich

Distribution of overall Response Time



Distribution of Response Time from Ealing and Greenwich



```
##
## Welch Two Sample t-test
##
## data: response_time by IncGeo_BoroughName
## t = 2.8542, df = 19303, p-value = 0.00432
## alternative hypothesis: true difference in means between group EALING and group GREENWICH
is not equal to 0
## 95 percent confidence interval:
##  1.739777 9.368041
## sample estimates:
##      mean in group EALING mean in group GREENWICH
##                316.9342                311.3803
```

The Two Sample t-test analysis shows that the Ealing's mean response time of 316.93 seconds is significantly higher than the Greenwich average response time 311.38 Seconds , welch $t(19303) = 2.85$, $p=0.00432$ with a difference of 5.55 seconds

Estimation:

```
## IncGeo_BoroughName emmean SE df lower.CL upper.CL
## EALING                317 1.34 19445      314      320
## GREENWICH              311 1.42 19445      309      314
##
## Confidence level used: 0.95
```

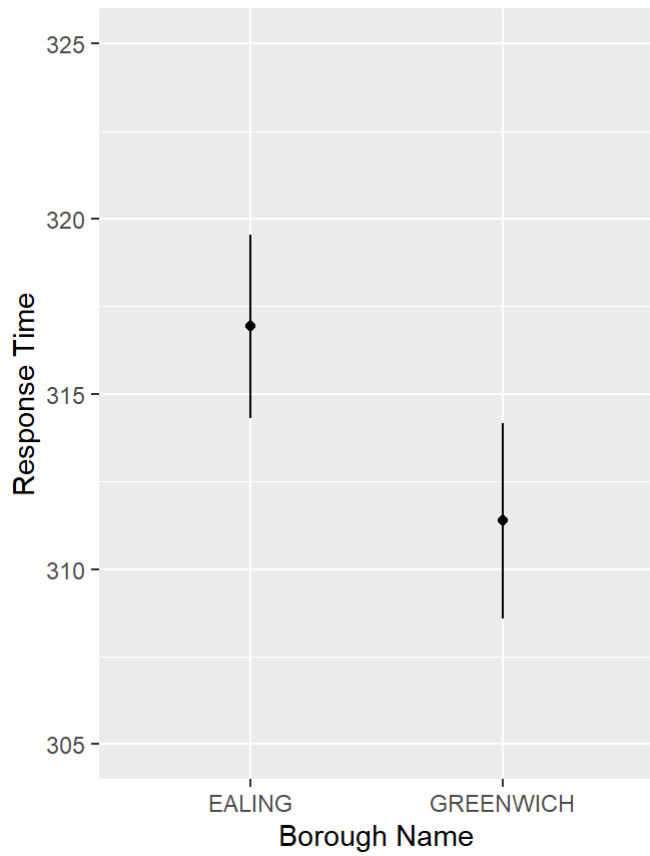
```
## contrast          estimate SE df lower.CL upper.CL
## EALING - GREENWICH    5.55 1.95 19445      1.73      9.38
##
## Confidence level used: 0.95
```

The mean response time for Ealing is 317 seconds 95% CI [314–320]. The mean response time for Greenwich is 311 seconds 95% CI [309–314]. The gain is 5.55 seconds 95% CI [1.73–9.38] smaller at Greenwich compared to Ealing.

Below image well describes the CI:

Mean Response Time

Error bars are 95% CIs



Difference in Response Time

Error bars are 95% CIs

