

# Air Traffic Delays Analysis

SALEH

4/18/2020

## 1. What is the pattern of arrival traffic and departure traffic delays with respect to days and weeks?

```
library(readr)

CA_delays <- read_csv("C:/Users/smos5/OneDrive - Kent State University/KENT m
y courese/Spring 2020/Analitycs in Practice/Assignment/Assignment 2/Sample_C
A_airtraffic_delays.csv")

## Warning: Missing column names filled in: 'X46' [46]
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   FlightDate = col_character(),
##   Marketing_Airline_Network = col_character(),
##   Operated_or_Branded_Code_Share_Partners = col_character(),
##   IATA_Code_Marketing_Airline = col_character(),
##   Originally_Scheduled_Code_Share_Airline = col_logical(),
##   DOT_ID_Originally_Scheduled_Code_Share_Airline = col_logical(),
##   IATA_Code_Originally_Scheduled_Code_Share_Airline = col_logical(),
##   Flight_Num_Originally_Scheduled_Code_Share_Airline = col_logical(),
##   Operating_Airline = col_character(),
##   IATA_Code_Operating_Airline = col_character(),
##   Tail_Number = col_character(),
##   Origin = col_character(),
##   OriginCityName = col_character(),
##   OriginState = col_character(),
##   OriginStateName = col_character(),
##   Dest = col_character(),
```

```
## DestCityName = col_character(),
## DestState = col_character(),
## DestStateName = col_character(),
## X46 = col_logical()
## # ... with 48 more columns
## )

## See spec(...) for full column specifications.
## Warning: 126 parsing failures.

## row          col          expected actual
file
## 1228 DivDistance      1/0/T/F/TRUE/FALSE 109      'C:/Users/smos5/OneDrive
- Kent State University/KENT my courese/Spring 2020/Analitycs in Practice/Ass
ignement/Assignment 2/Sample_CA_airtraffic_delays.csv'
## 1228 Div1Airport      1/0/T/F/TRUE/FALSE LAX      'C:/Users/smos5/OneDrive
- Kent State University/KENT my courese/Spring 2020/Analitycs in Practice/Ass
ignement/Assignment 2/Sample_CA_airtraffic_delays.csv'
## 1228 Div1AirportID    1/0/T/F/TRUE/FALSE 12892    'C:/Users/smos5/OneDrive
- Kent State University/KENT my courese/Spring 2020/Analitycs in Practice/Ass
ignement/Assignment 2/Sample_CA_airtraffic_delays.csv'
## 1228 Div1AirportSeqID 1/0/T/F/TRUE/FALSE 1289208 'C:/Users/smos5/OneDrive
- Kent State University/KENT my courese/Spring 2020/Analitycs in Practice/Ass
ignement/Assignment 2/Sample_CA_airtraffic_delays.csv'
## 1228 Div1WheelsOn     1/0/T/F/TRUE/FALSE 26      'C:/Users/smos5/OneDrive
- Kent State University/KENT my courese/Spring 2020/Analitycs in Practice/Ass
ignement/Assignment 2/Sample_CA_airtraffic_delays.csv'

## ....
## See problems(...) for more details.
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.2
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```

ArrDep_delays<- CA_delays %>% select("DayofMonth","DayOfWeek","Origin","DepDe
lay","DepDelayMinutes", "DepDel15","Dest","ArrDelay","ArrDelayMinutes","ArrDe
l15")

ArrDep_delays$DayofMonth<-as.Date(ArrDep_delays$DayofMonth, origin = "2018-01
-01")

ArrDep_delays$DayOfWeek<-as.factor(ArrDep_delays$DayOfWeek)

ArrDep_delays$DepDel15<-as.factor(ArrDep_delays$DepDel15)

ArrDep_delays$ArrDel15<-as.factor(ArrDep_delays$ArrDel15)

ArrDelay_per<-(ArrDep_delays$ArrDelay/CA_delays$ArrTime)*100

ArrDep_delays<-cbind(ArrDep_delays,ArrDelay_per)

```

## 1. What is the pattern of arrival traffic and departure traffic delays with respect to days and weeks?

```

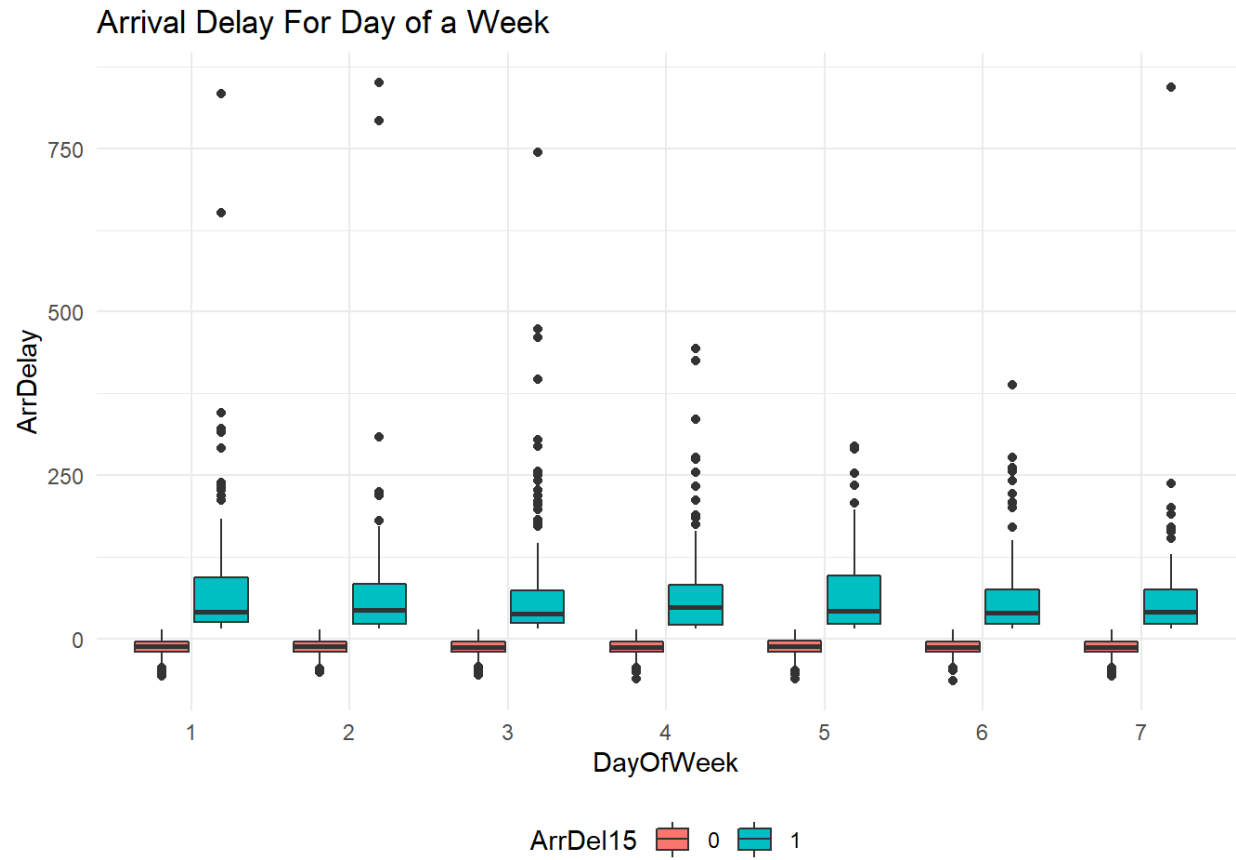
library(dplyr)

library(ggplot2)

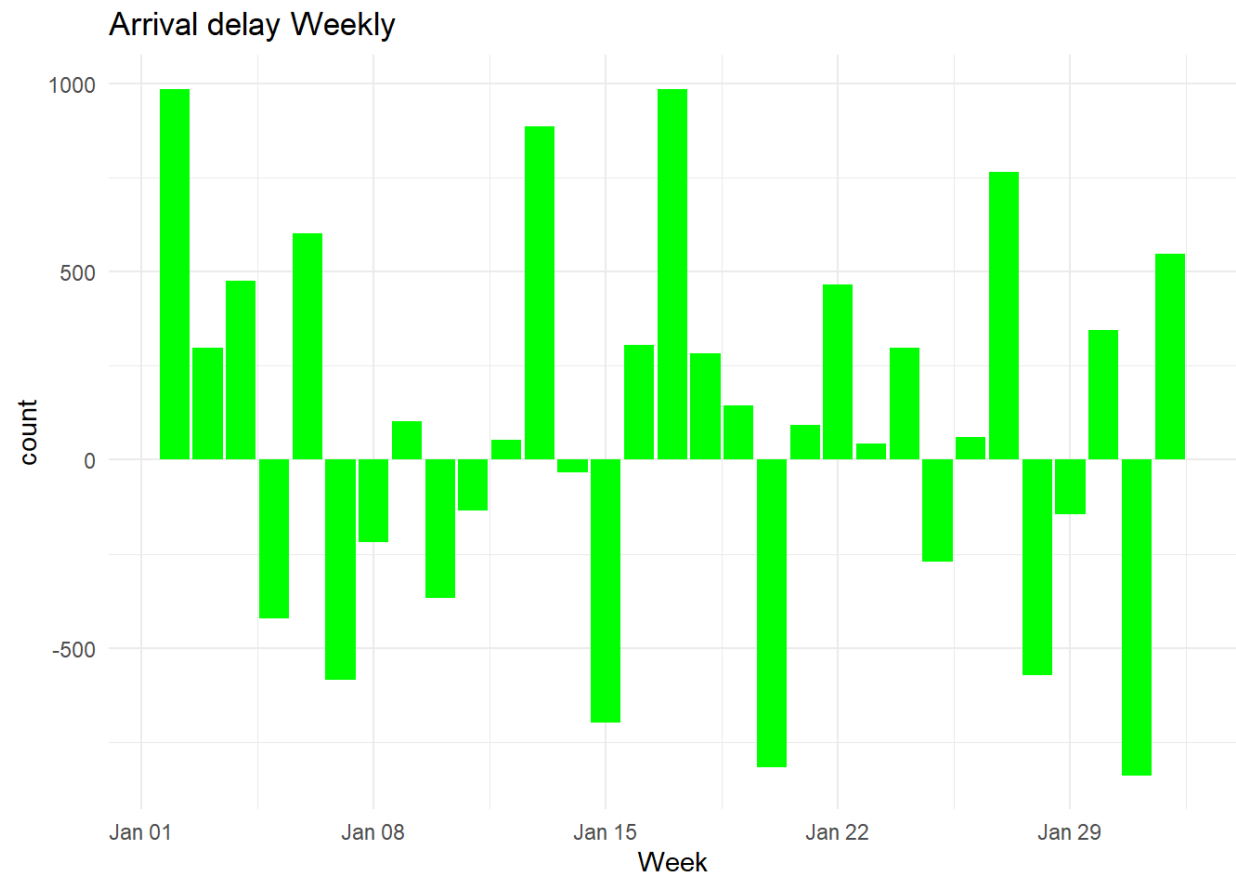
## Warning: package 'ggplot2' was built under R version 3.6.2

ArrDep_delays %>%
  filter(!is.na(ArrDelay)) %>%
  filter(!is.na(ArrDel15)) %>%
  ggplot() +
  aes(x = DayOfWeek, y = ArrDelay, fill = ArrDel15) +
  geom_boxplot() +
  scale_fill_hue() +
  labs(title = "Arrival Delay For Day of a Week ") +
  theme_minimal() +
  theme(legend.position = "bottom")

```

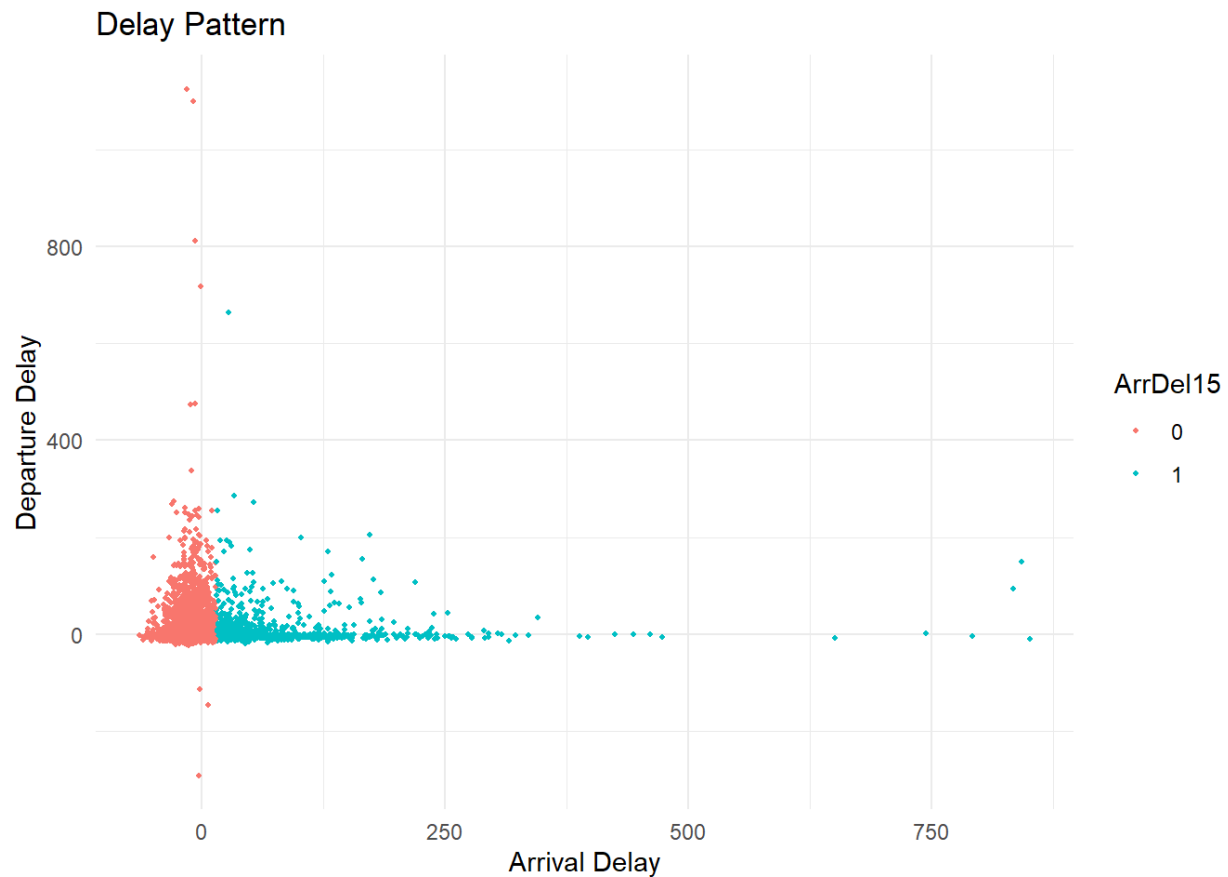


```
ArrDep_delays %>%  
  filter(!is.na(DepDelay)) %>%  
  filter(!is.na(ArrDelay)) %>%  
  ggplot() +  
  aes(x = DayofMonth, weight = ArrDelay) +  
  geom_bar(fill = "Green") +  
  labs(x = "Week", title = "Arrival delay Weekly") +  
  theme_minimal()
```



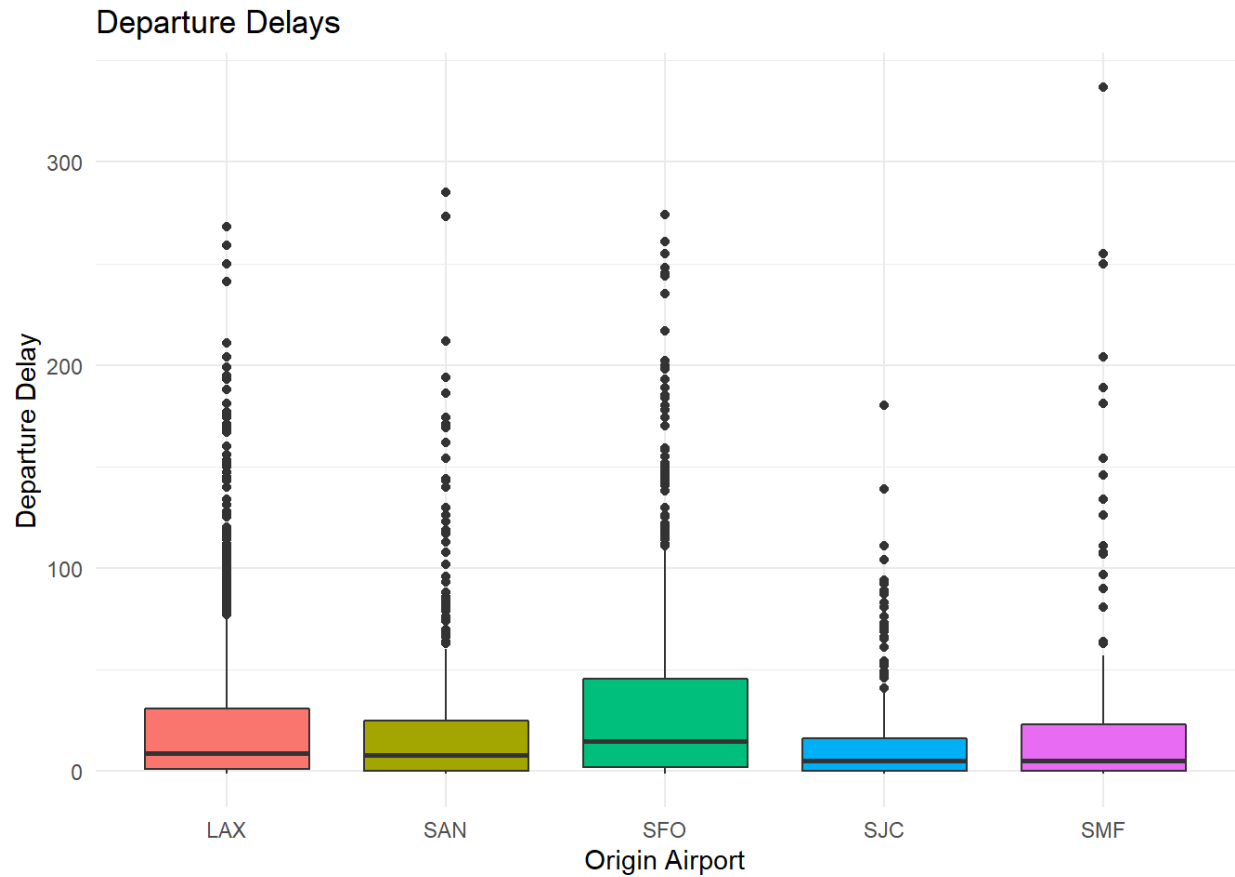
## 2. Can you interpret the traffic delays?

```
ArrDep_delays %>%  
  filter(!is.na(DepDelay)) %>%  
  filter(!is.na(ArrDelay)) %>%  
  filter(!is.na(ArrDel15)) %>%  
  ggplot() +  
  aes(x = ArrDelay, y = DepDelay, colour = ArrDel15) +  
  geom_point(size = 0.7) +  
    scale_color_hue() +  
  labs(x = "Arrival Delay", y = "Departure Delay", title = "Delay Pattern") +  
  theme_minimal()
```



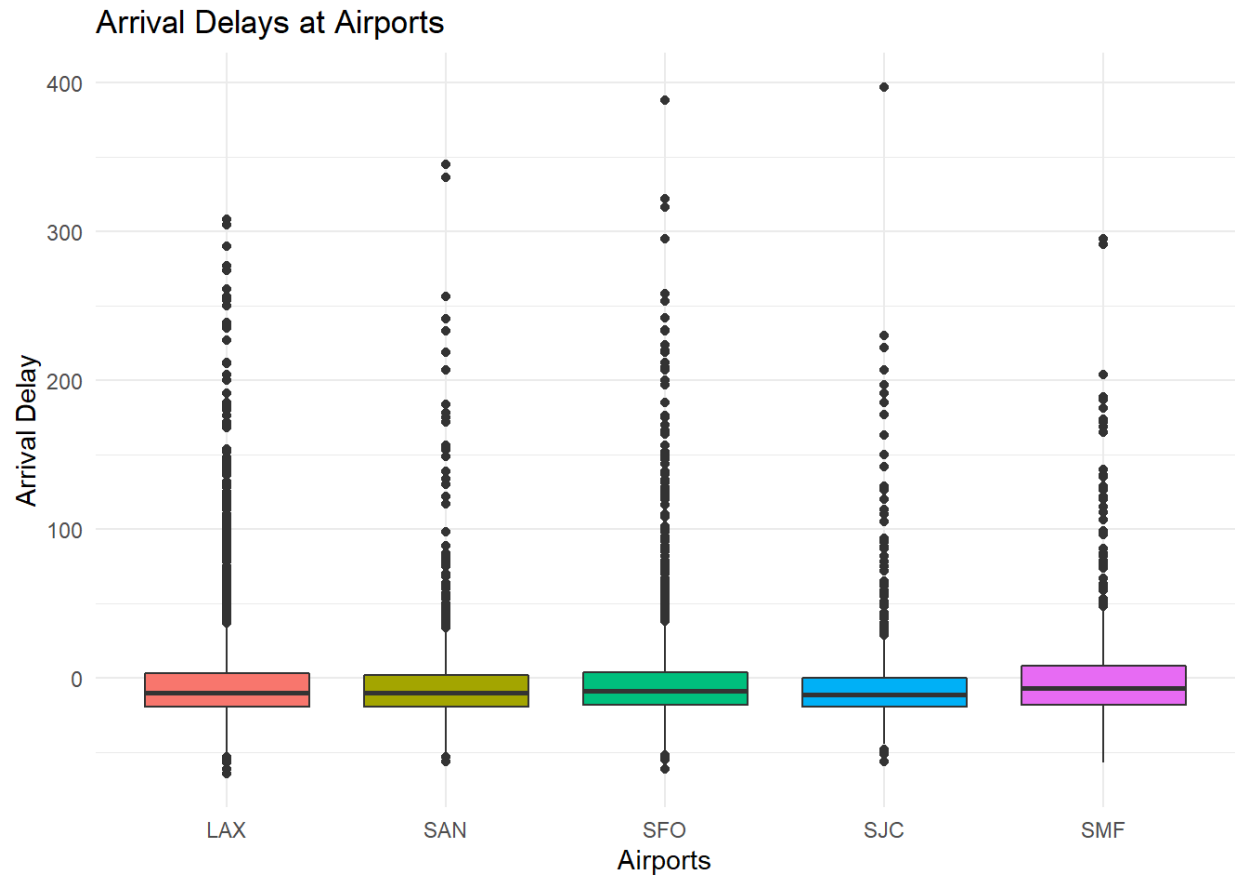
### 3. Which Airport ('Origin Airport') has highest departure delay?

```
ArrDep_delays %>%  
  filter(DepDelay >= -1L & DepDelay <= 403L & !is.na(DepDelay)) %>%  
  ggplot() +  
  aes(x = Origin, y = DepDelay, fill = Origin) +  
  geom_boxplot() +  
  scale_fill_hue() +  
  labs(x = "Origin Airport", y = "Departure Delay", title = "Departure Delays")  
) +  
  theme_minimal() +  
  theme(legend.position = "none")
```



#### 4. Which Airport has highest Arrival delay?

```
ArrDep_delays %>%  
  filter(ArrDelay >= -64L & ArrDelay <= 405L & !is.na(ArrDelay)) %>%  
  ggplot() +  
  aes(x = Origin, y = ArrDelay, fill = Origin) +  
  geom_boxplot() +  
  scale_fill_hue() +  
  labs(x = "Airports", y = "Arrival Delay", title = "Arrival Delays at Airports") +  
  theme_minimal() +  
  theme(legend.position = "none")
```



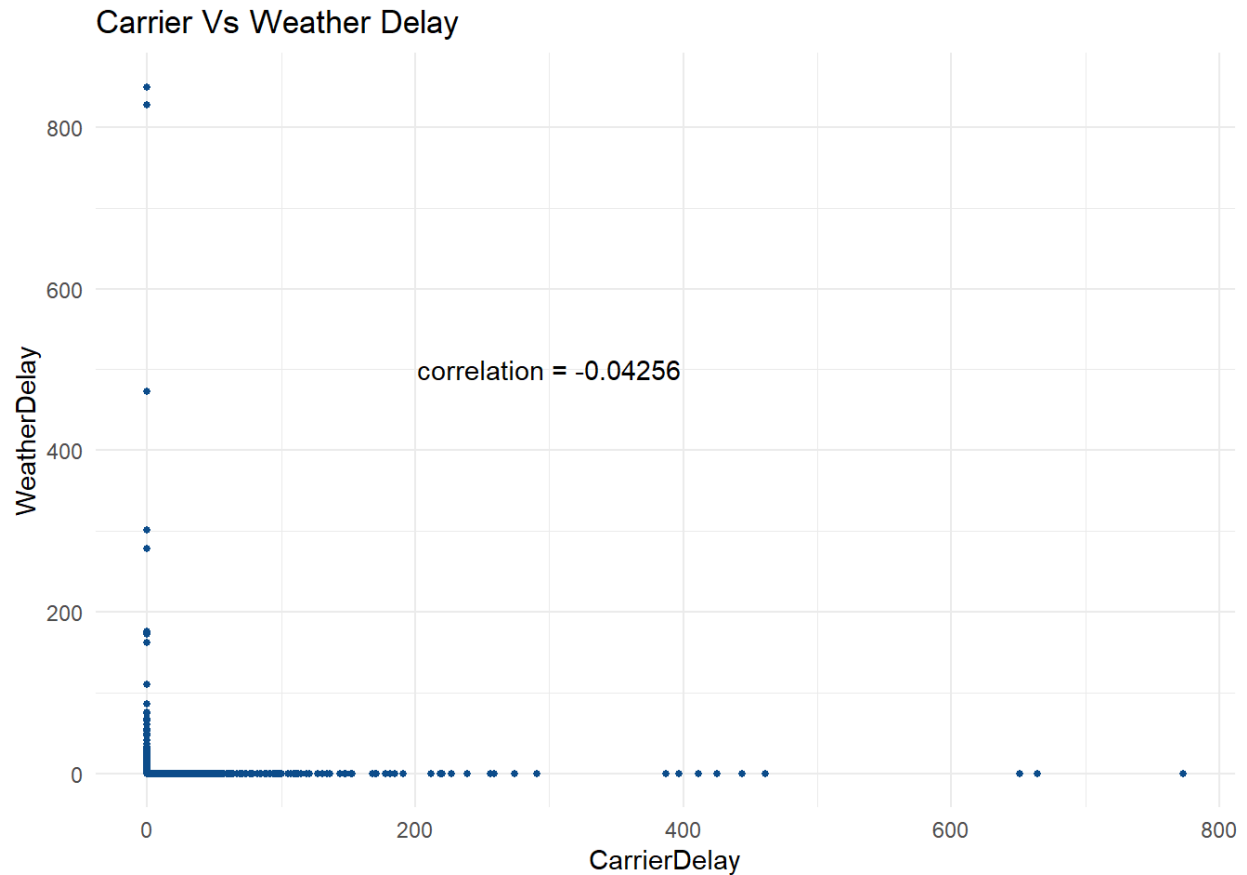
5. How do you relate the delay pattern to the distance travelled?

6. Is there any correlation between weather delay and carrier delay?

```
traffic_delay<-CA_delays %>% select('DayofMonth', 'CarrierDelay', 'WeatherDelay', 'NASDelay', 'SecurityDelay', 'LateAircraftDelay')
%>% filter(!is.na(CarrierDelay))

ggplot(traffic_delay) +
  aes(x = CarrierDelay, y = WeatherDelay) +
  geom_point(size = 1L, colour = "#0c4c8a") +
  labs(title = "Carrier Vs Weather Delay") +
  theme_minimal() + annotate("text", x = 300, y = 500, label = "correlation = -0.04256")
```





**7. What is the delay pattern you can find in respective states?**

**8. How many delayed flights were cancelled? (approximation)**

```
cancel_delays<-CA_delays %>% select('DepDelay', 'ArrDelay', 'Cancelled', 'Diverted')
```

```
cancel_delays %>% filter(Cancelled==1 & !is.na(Cancelled) & DepDelay > 0 & !is.na(DepDelay)) %>% count(Cancelled)
```

```
## # A tibble: 1 x 2
```

```
##   Cancelled     n
```

```
##       <dbl> <int>
```

```
## 1         1     44
```

```
cancel_delays %>%
```

```
  filter(DepDelay >= -292L & DepDelay <= 300L & !is.na(DepDelay)) %>%
```

```
  filter(Cancelled >= 1L & Cancelled <= 1L) %>%
```

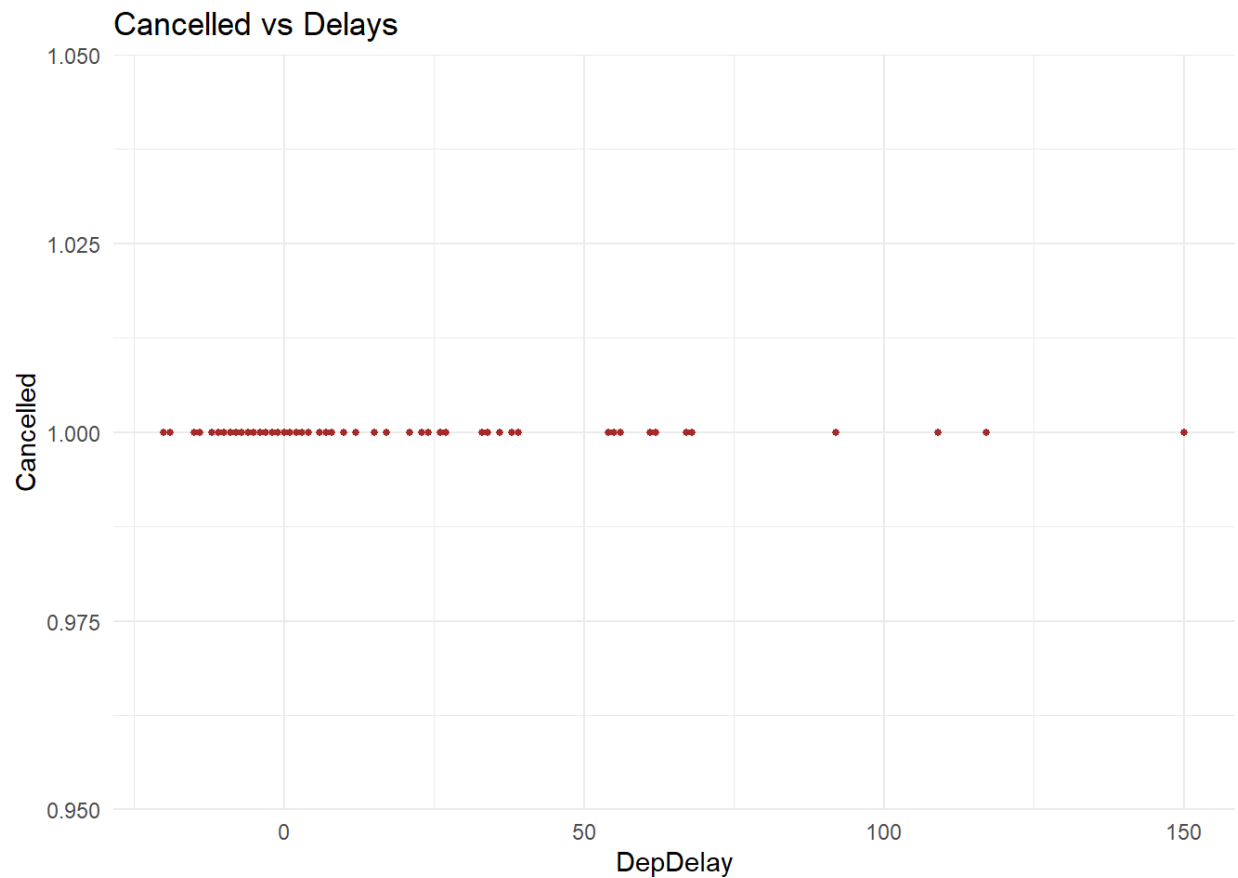
```
  ggplot() +
```

```
  aes(x = DepDelay, y = Cancelled) +
```

```
  geom_point(size = 1, colour = "brown") +
```

```
  labs(title = "Cancelled vs Delays") +
```

```
theme_minimal()
```



### 9. How many delayed flights were diverted? (approximation)

```
cancel_delays %>% filter(Diverted==1 & !is.na(Diverted) & ArrDelay>0 | DepDelay > 0 & !is.na(DepDelay)) %>% count(Diverted)
```

```
## # A tibble: 2 x 2
##   Diverted      n
##   <dbl> <int>
## 1       0  2091
## 2       1     3
```

### 10. What time of the day do you find Arrival delays?

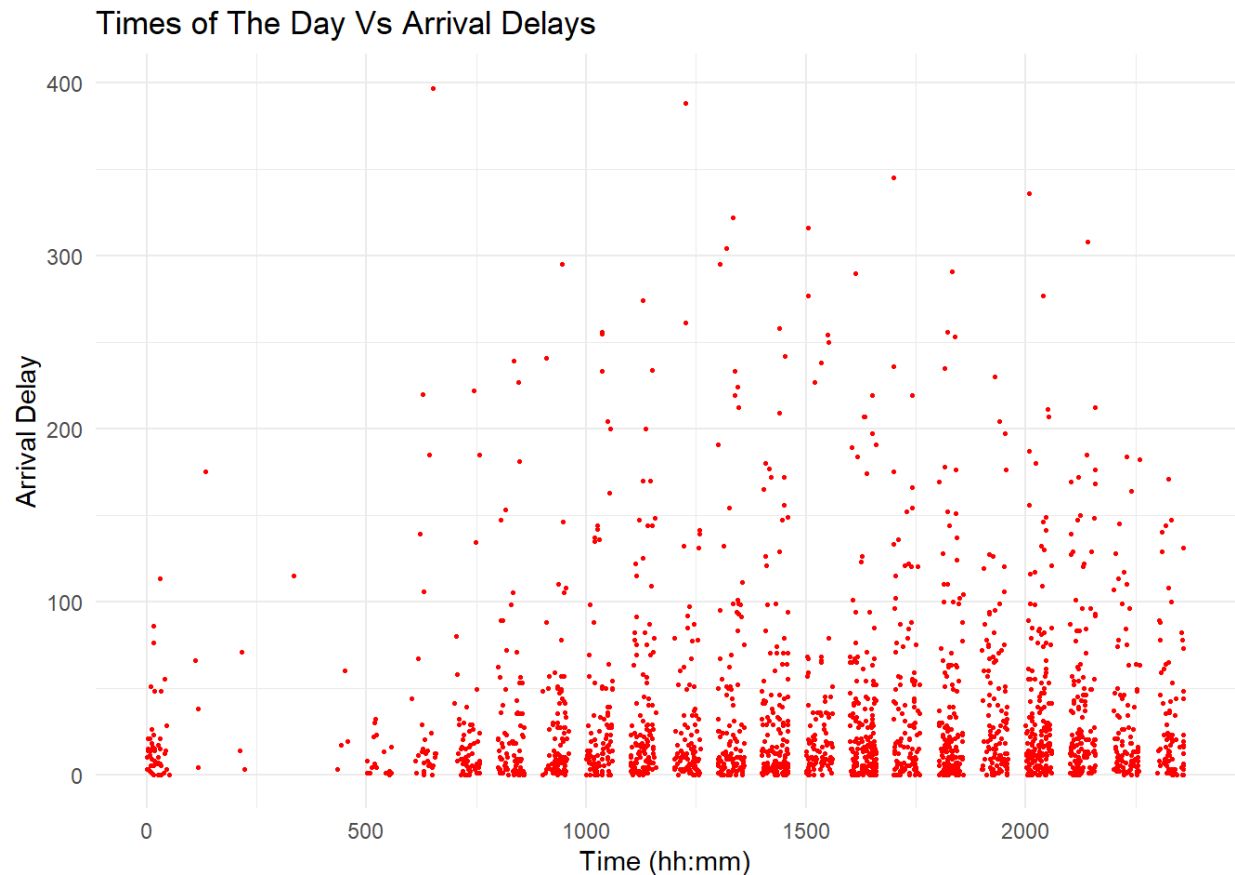
```
time_delays<-CA_delays %>% select('CRSDepTime', 'DepDelay', 'CRSArrTime', 'ArrDelay')
```

```
time_delays %>%
  filter(ArrDelay >= 0L & ArrDelay <= 410L & !is.na(ArrDelay)) %>%
  ggplot() +
```

```

aes(x = CRSArrTime, y = ArrDelay) +
geom_point(size = 0.5, colour = "RED") +
labs(x = "Time (hh:mm)", y = "Arrival Delay", title = "Times of The Day Vs Arrival Delays") +
theme_minimal()

```



## 11. What time of the day do you find Departure delays?

```

time_delays %>%
  filter(DepDelay >= -1L & DepDelay <= 417L & !is.na(DepDelay)) %>%
  ggplot() +
  aes(x = CRSDepTime, y = DepDelay) +
  geom_point(size = 0.5, colour = "#0c4c8a") +
  labs(x = "Time (hh:mm)", y = "Departure Delay", title = "Times of The Day Vs Departure Delays") +
  theme_minimal()

```

Times of The Day Vs Departure Delays

