Analyzing NYC Flight Data - HW2

Kinjal Ray

```
install.packages("nycflights13", repos = "http://cran.us.r-project.org")

##
## The downloaded binary packages are in
##
/var/folders/gs/1zv4dkj94hg76qd46gvyn_qw0000gn/T//RtmpysGkFj/downloaded_packa
ges

library(nycflights13)

## Warning: package 'nycflights13' was built under R version 3.5.2

nyc <- nycflights13::flights
```

1.

```
nyc$airgain <- nyc$arr_delay - nyc$dep_delay
```

PART A - Do Airlines Gain Time?

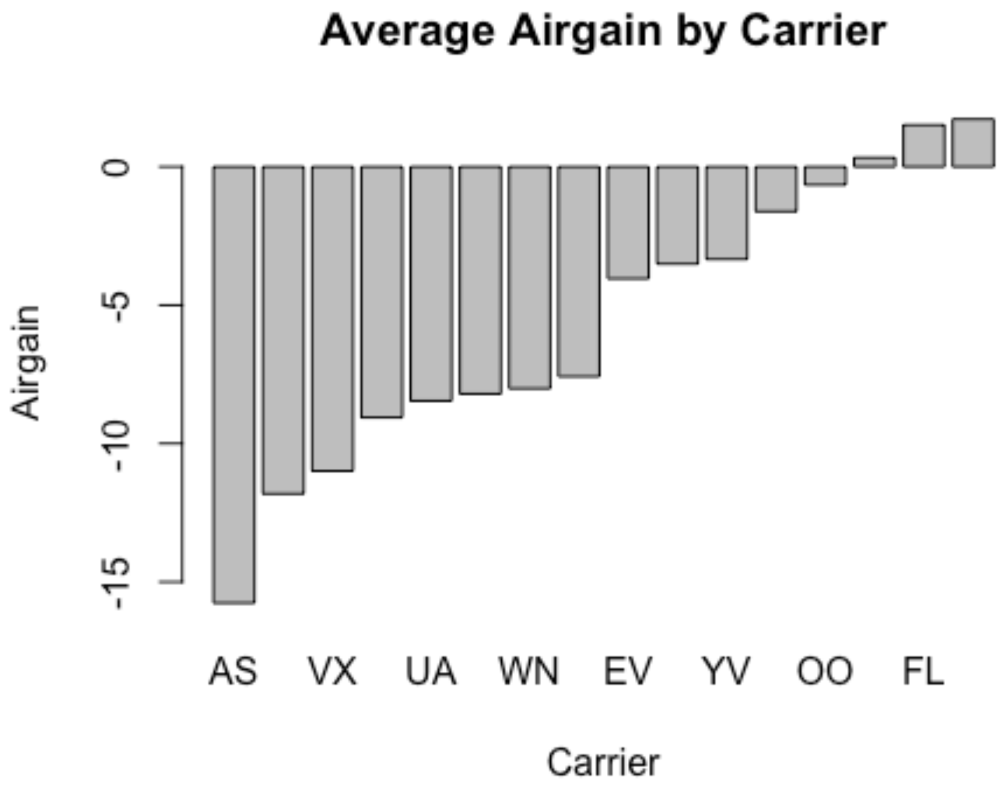
```
ifelse(mean(nyc$airgain, na.rm = TRUE) > 0, "NO, on average, airlines don't
gain time in the air", "YES, on average, airlines do gain time")

## [1] "YES, on average, airlines do gain time"
```

PART B - Average Airgain by Airline Carrier

```
avg_airgain <- tapply(nyc$airgain, nyc$carrier, mean, na.rm = TRUE)
```

PART C - Build bar chart

```
avg_airgain <- sort(avg_airgain)
barplot(avg_airgain, main = "Average Airgain by Carrier", xlab = "Carrier",
ylab = "Airgain")
```



2. Lowest Departure Delay, Best Airgain

```
avg_dep_delay <- tapply(nyc$dep_delay, nyc$origin, mean, na.rm = TRUE)
least_dep_delay <- names(avg_dep_delay[avg_dep_delay == min(avg_dep_delay)])

avg_airgain_bycarrier <- tapply(nyc$airgain, nyc$origin, mean, na.rm = TRUE)
best_airgain <- names(avg_airgain_bycarrier[avg_airgain_bycarrier ==
min(avg_airgain_bycarrier)])
```

LGA has the least amount of departure delay on average. JFK has the best airgain.

3. Best Orgination Airport based on Month of Travel

The best origination airport for the month of June is LaGuardia airport (LGA)

```
library(dplyr)

##
## 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
```

```
Best_Origin <- function(month_of_travel) {

month_of_travel <- switch (month_of_travel, "January" = 1, "February" = 2,
"March" = 3,"April"=  4, "May" = 5, "June" = 6, "July" = 7, "August" = 8,
"September" = 9, "Ocotber" = 10, "November" = 11, "December" = 12)

  by_month <- filter(nyc, nyc$month == month_of_travel)
  by_origin <- tapply(by_month$arr_delay, by_month$origin, mean, na.rm =
TRUE)
  best_choice <- names(by_origin[by_origin == min(by_origin)])
  return(best_choice)
}

Best_Origin("June")
```

```
## [1] "LGA"
```

4. Best Time of Day for Travel

The best time of day for travel is morning.

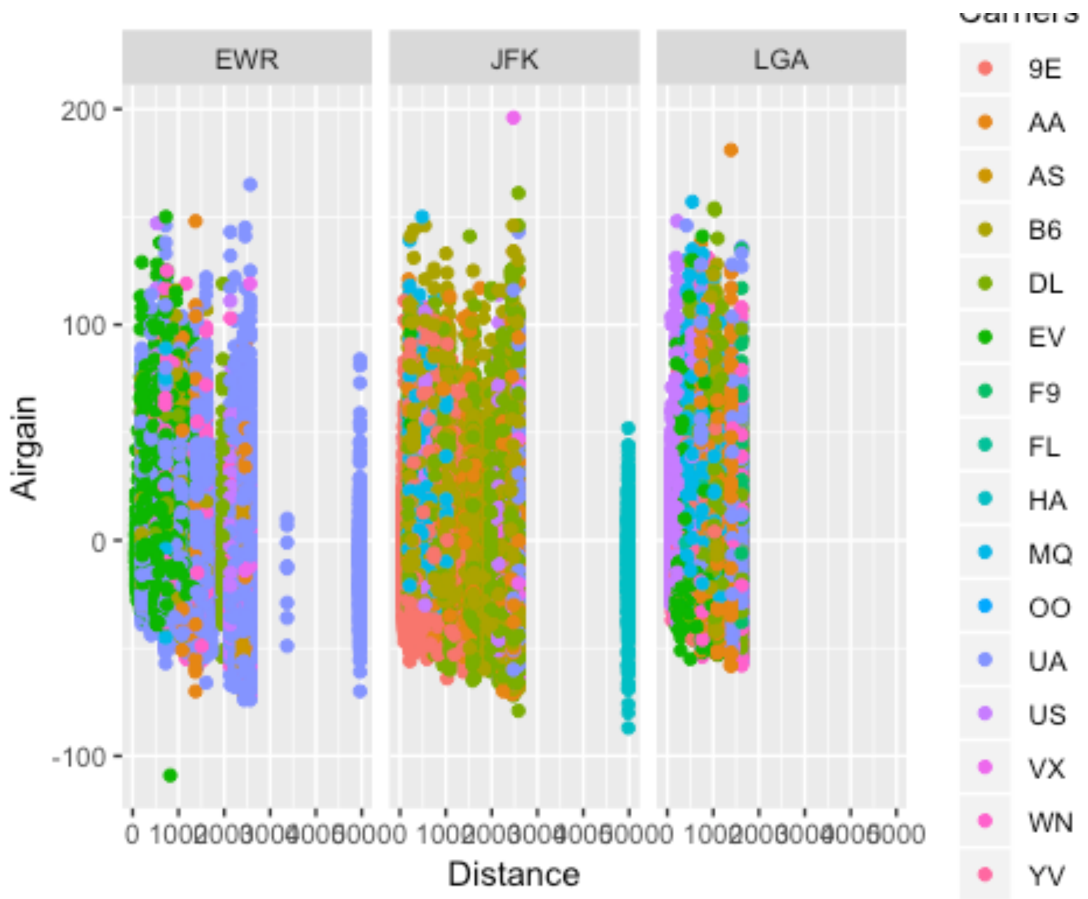
```
atl_only <- filter(nyc, nyc$dest == "ATL")
atl_only$timeOfDay <- ifelse(atl_only$hour >= 4 & atl_only$hour < 12,
"Morning",ifelse(atl_only$hour >= 12 & atl_only$hour < 16, "Afternoon",
"Evening"))

least_arr_delay <- tapply(atl_only$arr_delay, atl_only$timeOfDay, mean, na.rm
= TRUE)

names(least_arr_delay[least_arr_delay == min(least_arr_delay)])
```

```
## [1] "Morning"
```

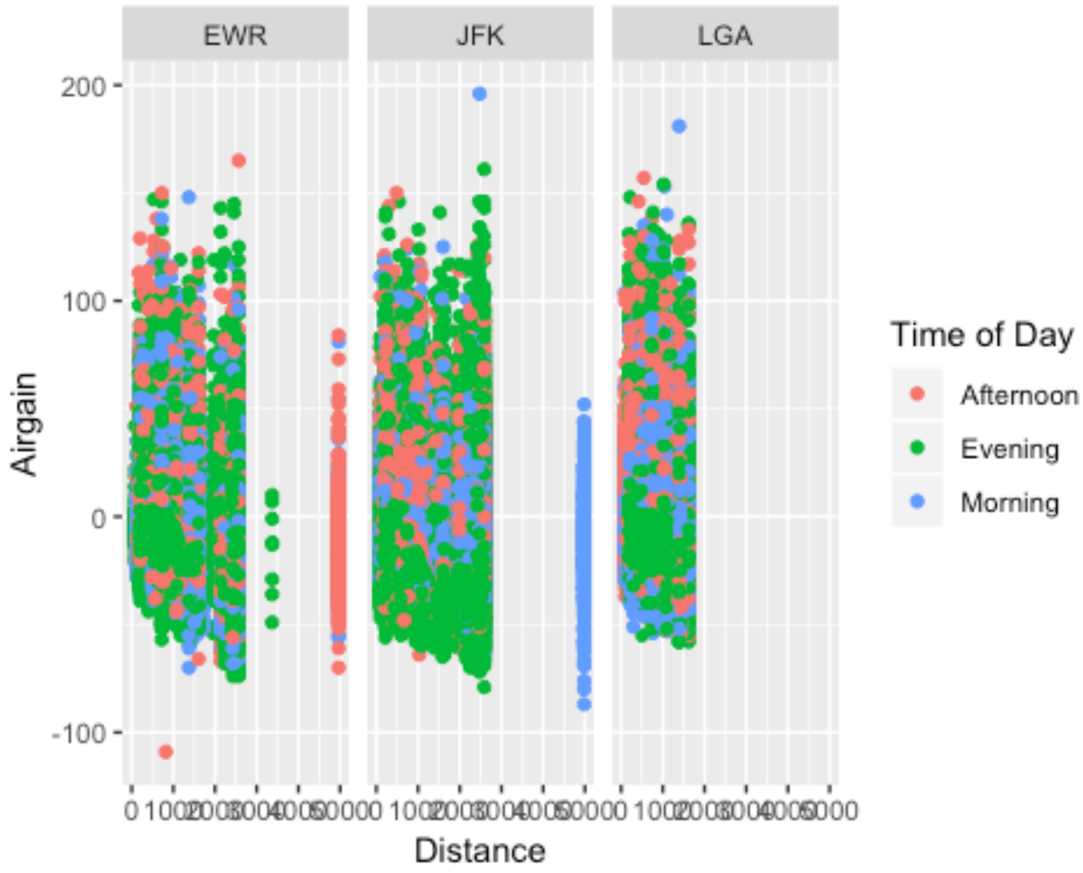
Relationship between Distance and Airgain

Plot 1:Grouped by Carrier Plot 2:Grouped by Time of Departure

```
library(ggplot2)
baseplot <- ggplot(nyc, aes(x = nyc$distance, y = nyc$airgain, color =
nyc$carrier)) + labs(x = "Distance", y = "Airgain", colour = "Carriers")

baseplot + geom_point(na.rm = TRUE) + facet_wrap(~ nyc$origin)
```



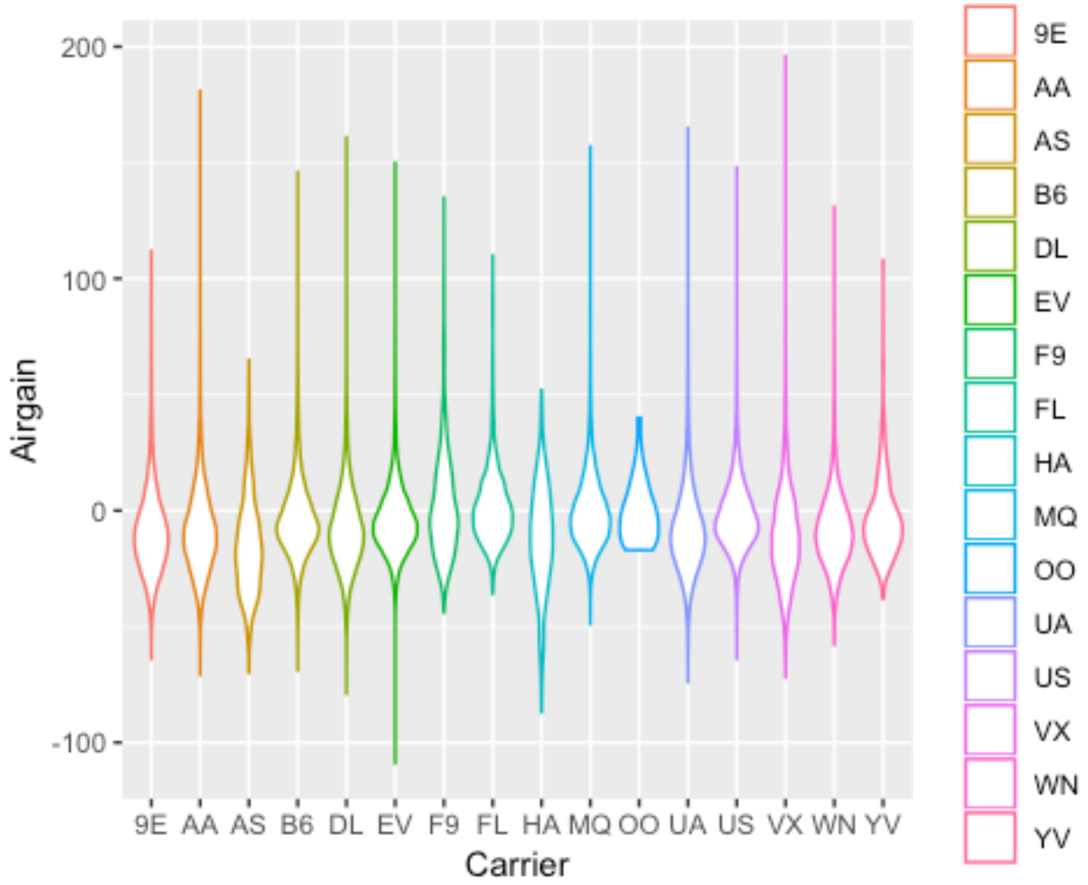
```r
nyc$timeOfDay <- ifelse(nyc$hour >= 4 & nyc$hour < 12, "Morning",
ifelse(nyc$hour >= 12 & nyc$hour < 16, "Afternoon", "Evening"))

baseplot2 <- ggplot(nyc, aes(x = nyc$distance, y = nyc$airgain, colour =
nyc$timeOfDay)) + labs(x = "Distance", y = "Airgain", colour = "Time of Day")

baseplot2 + geom_point(na.rm = TRUE) + facet_wrap(~ nyc$origin)
```



Airgain by Carrier

This violin plot showcasing the distribution of flight air gain amounts grouped by carrier. Although most carriers hover around the 0 air gain amount, the plot provides useful insight on performance of carriers. For example, the OO airline seems more consistent than other airlines, because it has much shorter tails. YV, EV, B6, and 9E predominantly have negative air gain, meaning they finish flights faster than expected.

```
ggplot(nyc, aes(x = nyc$carrier, y = nyc$airgain, colour = nyc$carrier)) +
geom_violin(na.rm = TRUE) + labs(x = "Carrier", y  = "Airgain", colour =
"Carrier")
```


200
100
0
-100
Airgain
9E
AA
AS
B6
DL
EV
F9
FL
HA
MQ
OO
UA
US
VX
WN
YV
Carrier
9E
AA
AS
B6
DL
EV
F9
FL
HA
MQ
OO
UA
US
VX
WN
YV