Project 1 - Visual

```
## 1. How many variables and how many observations in the data for CVG_Flights data set

cvg <- read.csv("CVG_Flights.csv", header = TRUE, stringsAsFactors = FALSE, na.strings = "")

nrow(cvg)
```

```
## [1] 7763
```

```
ncol(cvg)
```

```
## [1] 27
```

```
## 2. Missing value and visualizations

library(naniar)
```

```
## Warning: package 'naniar' was built under R version 3.6.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.3

##
## 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
```

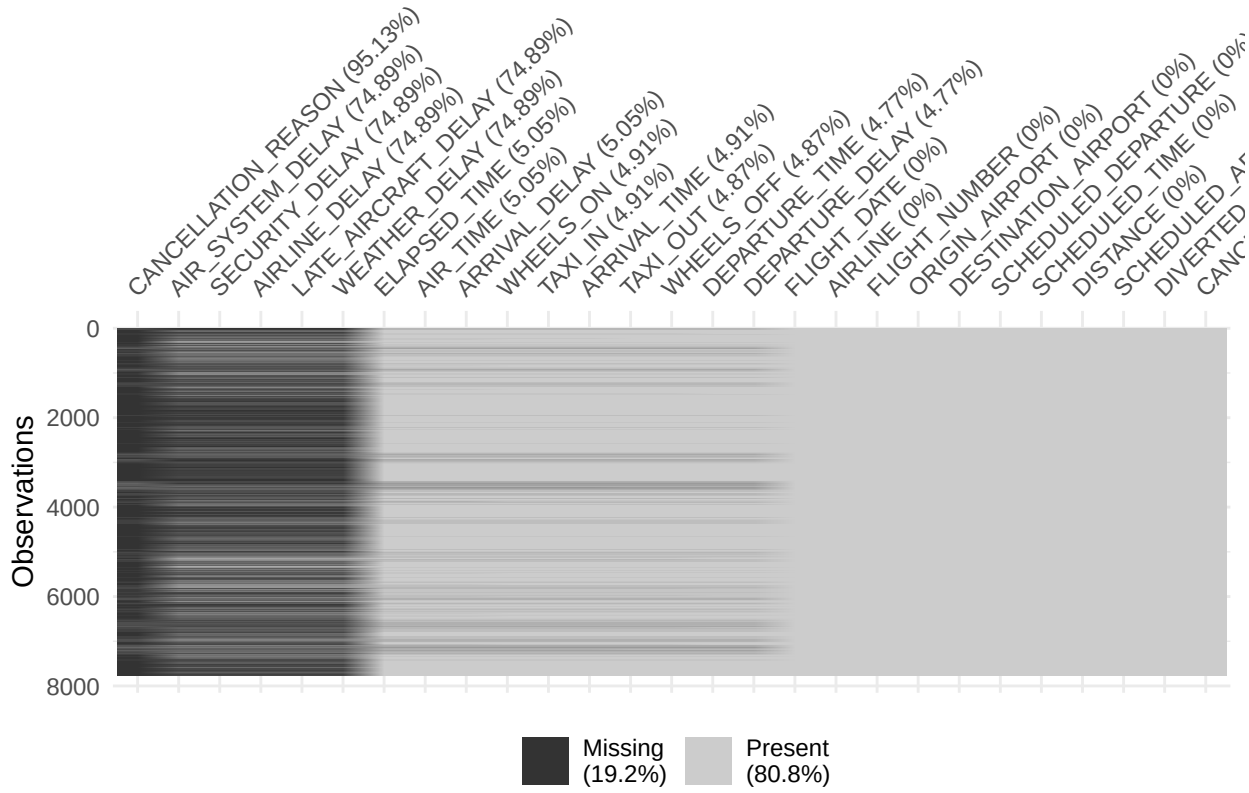
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

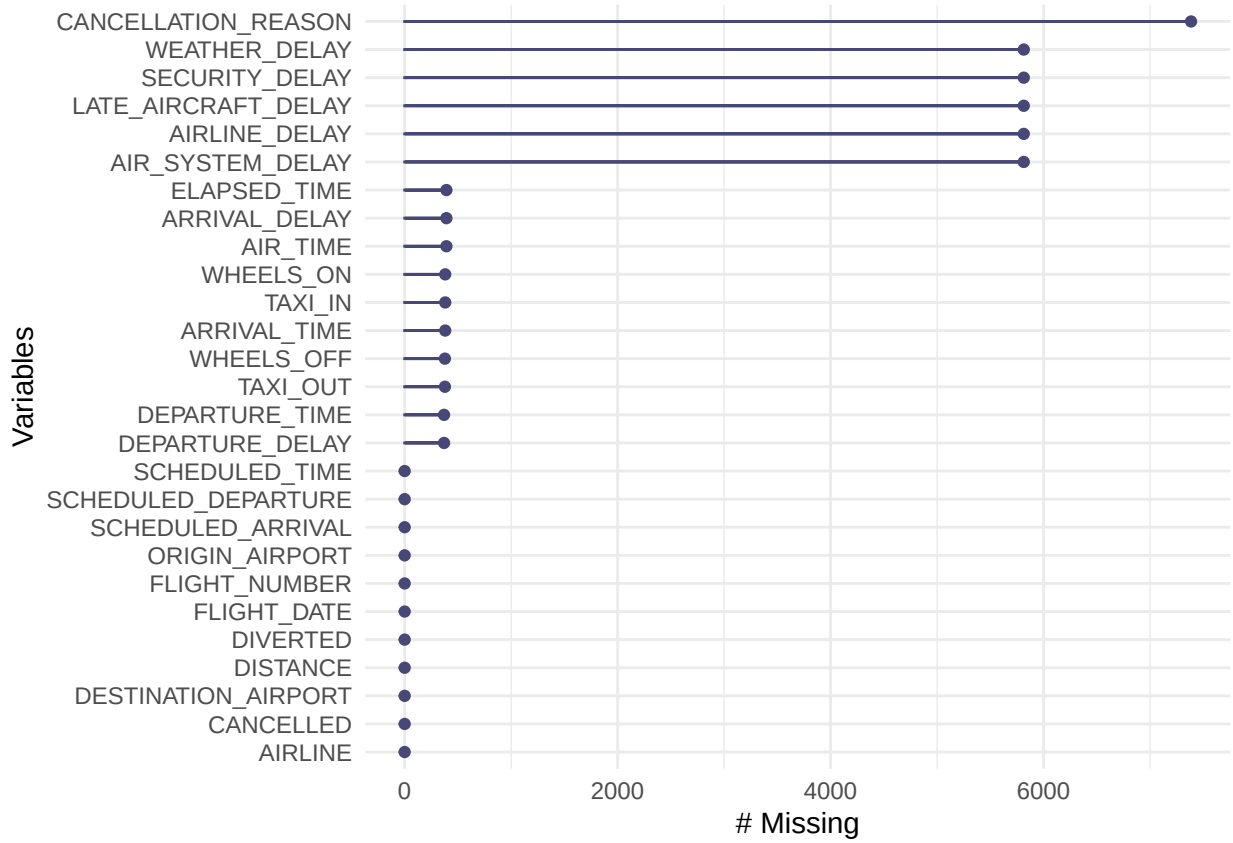
```
any_na <- sapply(cvg, function(cvg) sum(is.na(cvg)))
any_na
```

```
##         FLIGHT_DATE             AIRLINE       FLIGHT_NUMBER      ORIGIN_AIRPORT
##                   0                   0                   0                   0
## DESTINATION_AIRPORT SCHEDULED_DEPARTURE      DEPARTURE_TIME     DEPARTURE_DELAY
##                   0                   0                 370                 370
##            TAXI_OUT          WHEELS_OFF      SCHEDULED_TIME        ELAPSED_TIME
##                 378                 378                   0                 392
##            AIR_TIME            DISTANCE           WHEELS_ON             TAXI_IN
##                 392                   0                 381                 381
##   SCHEDULED_ARRIVAL        ARRIVAL_TIME       ARRIVAL_DELAY            DIVERTED
##                   0                 381                 392                   0
##           CANCELLED CANCELLATION_REASON     AIR_SYSTEM_DELAY      SECURITY_DELAY
##                   0                7385                5814                5814
##       AIRLINE_DELAY LATE_AIRCRAFT_DELAY     WEATHER_DELAY
##                5814                5814                5814
```

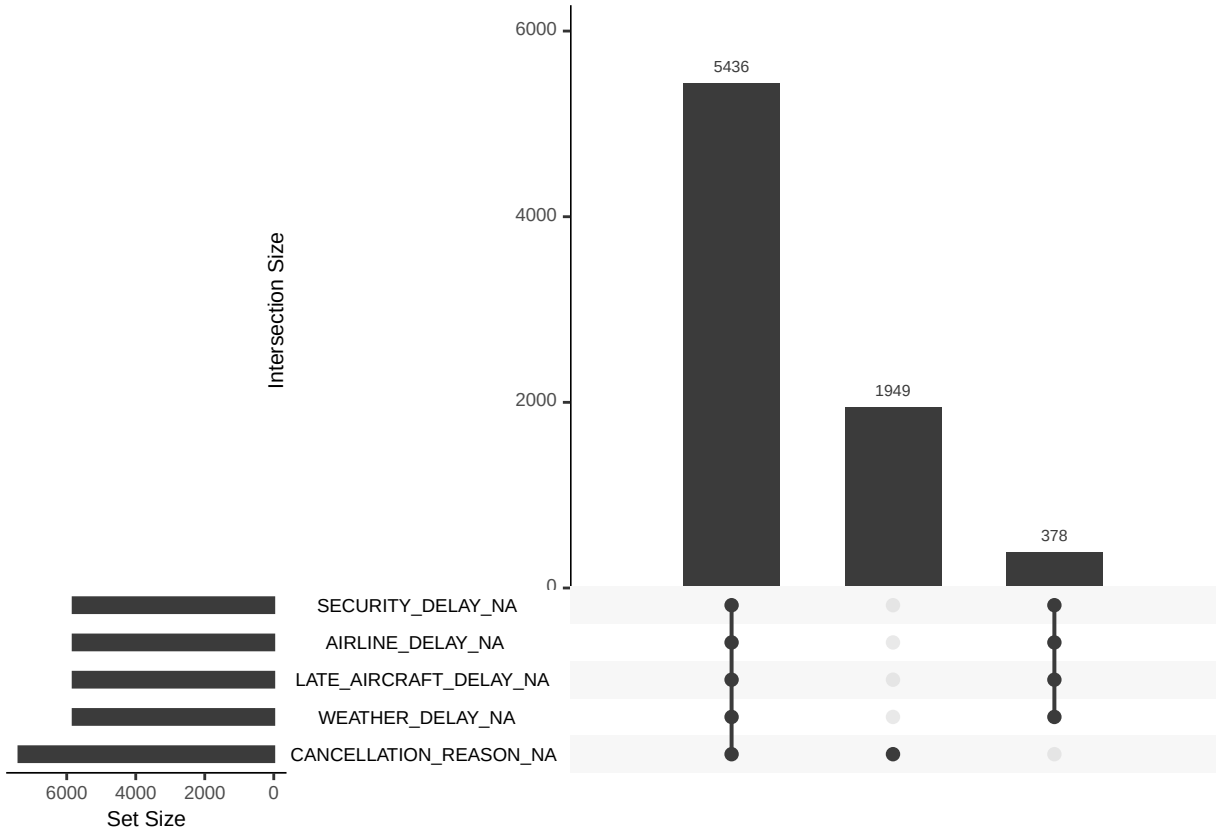
```
vis_miss(cvg, sort_miss = TRUE)
```




```
gg_miss_var(cvg)
```




```
gg_miss_upset(cvg)
```



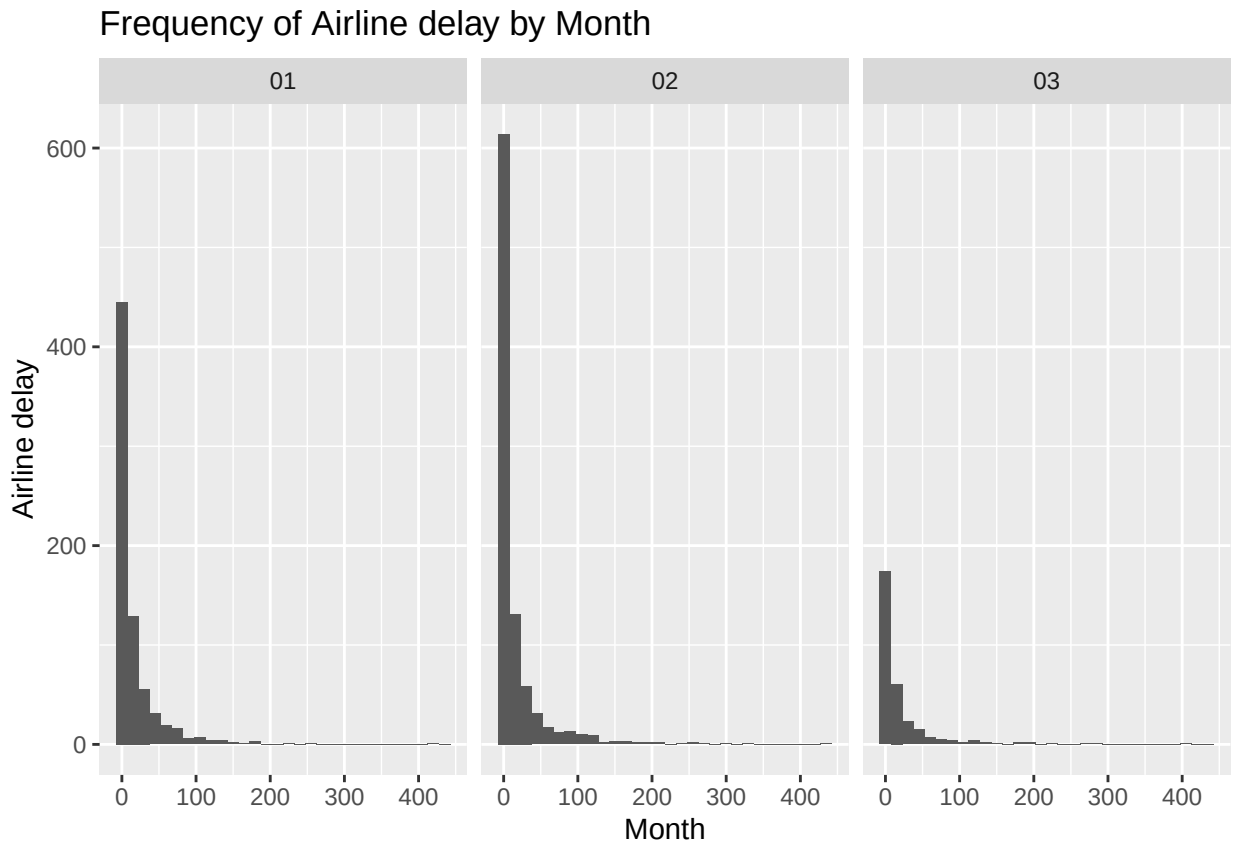

```
## 3. Summary statistic

summary(cvg)
```

```
##  FLIGHT_DATE          AIRLINE          FLIGHT_NUMBER  ORIGIN_AIRPORT
##  Length:7763        Length:7763        Min.   :  62   Length:7763
##  Class :character   Class :character   1st Qu.:1908   Class :character
##  Mode  :character   Mode  :character   Median :3246   Mode  :character
##                                        Mean   :3194
##                                        3rd Qu.:4536
##                                        Max.   :6297
##
##  DESTINATION_AIRPORT SCHEDULED_DEPARTURE DEPARTURE_TIME     DEPARTURE_DELAY
##  Length:7763         Length:7763         Length:7763        Min.   :-24.00
##  Class :character    Class :character    Class :character   1st Qu.: -5.00
##  Mode  :character    Mode  :character    Mode  :character   Median : -1.00
##                                                             Mean   : 12.35
##                                                             3rd Qu.: 12.00
##                                                             Max.   :708.00
##                                                             NA's   :370
##     TAXI_OUT       WHEELS_OFF        SCHEDULED_TIME   ELAPSED_TIME
##  Min.   :  1.00   Length:7763        Min.   : 60.0   Min.   : 49.0
##  1st Qu.: 12.00   Class :character   1st Qu.: 85.0   1st Qu.: 83.0
##  Median : 15.00   Mode  :character   Median :126.0   Median :123.0
##  Mean   : 18.52                      Mean   :128.1   Mean   :127.3
##  3rd Qu.: 21.00                      3rd Qu.:156.0   3rd Qu.:152.0
```

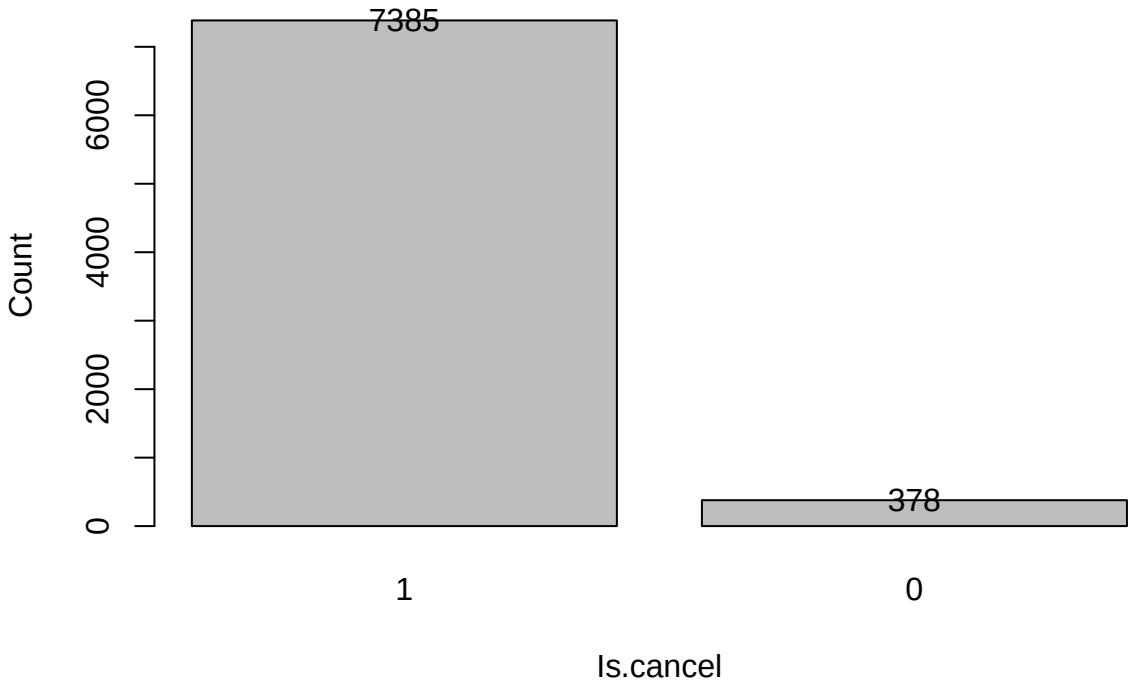
```
##  Max.   :122.00                     Max.   :313.0   Max.   :340.0
##  NA's   :378                                         NA's   :392
##     AIR_TIME        DISTANCE       WHEELS_ON           TAXI_IN
##  Min.   : 36.0   Min.   : 229.0   Length:7763        Min.   :  1.000
##  1st Qu.: 56.0   1st Qu.: 308.0   Class :character   1st Qu.:  5.000
##  Median : 99.0   Median : 589.0   Mode  :character   Median :  6.000
##  Mean   :100.4   Mean   : 675.9                      Mean   :  8.448
##  3rd Qu.:125.0   3rd Qu.: 871.0                      3rd Qu.:  9.000
##  Max.   :319.0   Max.   :2036.0                      Max.   :128.000
##  NA's   :392                                         NA's   :381
##  SCHEDULED_ARRIVAL  ARRIVAL_TIME       ARRIVAL_DELAY       DIVERTED
##  Length:7763        Length:7763        Min.   :-56.00   Min.   :0.000000
##  Class :character   Class :character   1st Qu.:-12.00   1st Qu.:0.000000
##  Mode  :character   Mode  :character   Median : -2.00   Median :0.000000
##                                        Mean   : 10.65   Mean   :0.001803
##                                        3rd Qu.: 16.00   3rd Qu.:0.000000
##                                        Max.   :716.00   Max.   :1.000000
##                                        NA's   :392
##    CANCELLED       CANCELLATION_REASON AIR_SYSTEM_DELAY SECURITY_DELAY
##  Min.   :0.00000   Length:7763         Min.   :  0.00   Min.   :0.00
##  1st Qu.:0.00000   Class :character    1st Qu.:  0.00   1st Qu.:0.00
##  Median :0.00000   Mode  :character    Median : 10.00   Median :0.00
##  Mean   :0.04869                       Mean   : 16.35   Mean   :0.01
##  3rd Qu.:0.00000                       3rd Qu.: 23.00   3rd Qu.:0.00
##  Max.   :1.00000                       Max.   :258.00   Max.   :9.00
##                                        NA's   :5814     NA's   :5814
##  AIRLINE_DELAY    LATE_AIRCRAFT_DELAY WEATHER_DELAY
##  Min.   :  0.00   Min.   :  0.00      Min.   :  0.000
##  1st Qu.:  0.00   1st Qu.:  0.00      1st Qu.:  0.000
##  Median :  0.00   Median :  2.00      Median :  0.000
##  Mean   : 17.06   Mean   : 23.24      Mean   :  3.873
##  3rd Qu.: 18.00   3rd Qu.: 30.00      3rd Qu.:  0.000
##  Max.   :435.00   Max.   :435.00      Max.   :638.000
##  NA's   :5814     NA's   :5814        NA's   :5814
```

```
d<- as.POSIXct(cvg$FLIGHT_DATE, format = "%m/%d/%Y")
cvg$dm <- format(d, "%m")

# Frequency of Airline delay faceted by Month

ggplot(cvg, aes(AIRLINE_DELAY)) + geom_histogram(bins = 30) + facet_wrap(~dm) + ylab("Airline delay") +
```

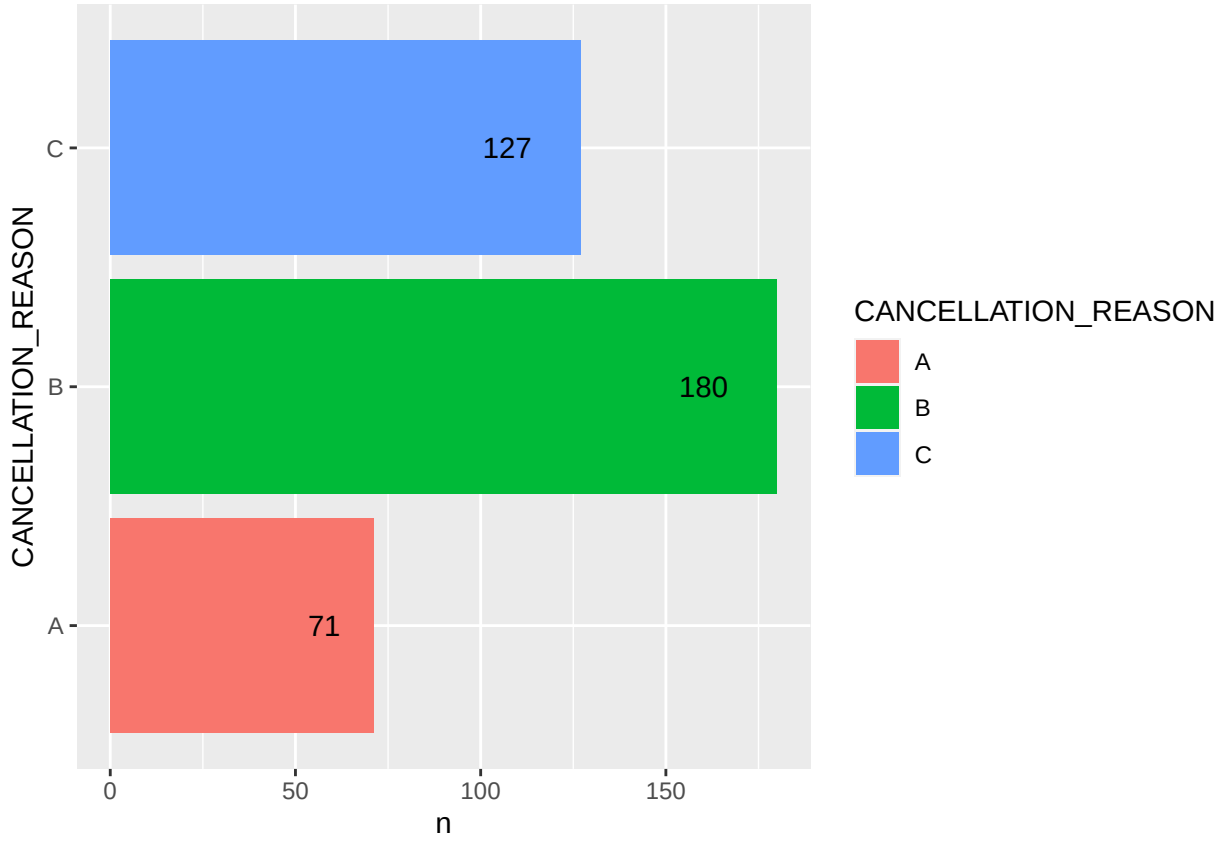
```
## Warning: Removed 5814 rows containing non-finite values (stat_bin).
```




```
# Numbers of Cancelled & Non-cancelled flights

counts <- aggregate(cvg$CANCELLED, by = list(cvg$CANCELLED), FUN = length)
names(counts) <- c("Is.cancel","Count")
p <- barplot(counts$Count, xlab = "Is.cancel", ylab = "Count", names.arg = c("1","0"))
text(p, counts$Count, labels= counts$Count, xpd=TRUE)
```




```
# Numbers of Cancellation reason

cancel <- cvg %>% filter(CANCELLATION_REASON != "NA") %>% count(CANCELLATION_REASON)
ggplot(cancel, aes(x = n, y = CANCELLATION_REASON, fill = CANCELLATION_REASON)) + geom_col() + geom_tex
```



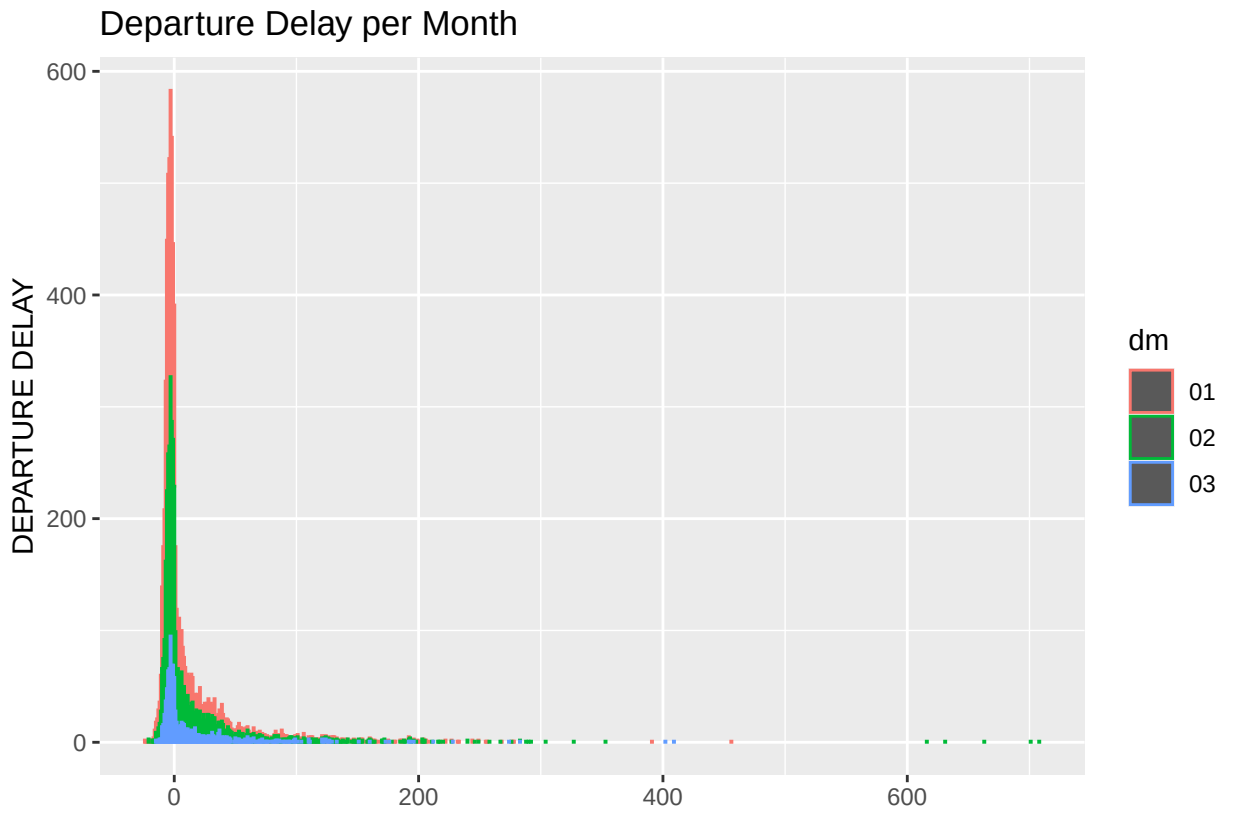

```
# Table for Departure delay & Arrival Delay in each Airport

origin <- aggregate(cvg$DEPARTURE_DELAY, by = list(cvg$ORIGIN_AIRPORT), FUN = sum, na.rm = TRUE)
colnames(origin) <- c("Airport", "Departure_delay")
origin$Departure_delay <- sort(origin$Departure_delay, decreasing = FALSE)

arrival <- aggregate(cvg$ARRIVAL_DELAY, by = list(cvg$DESTINATION_AIRPORT), FUN = sum, na.rm = TRUE)
colnames(arrival) <- c("Airport", "Arrival_delay")
arrival$Arrival_delay <- sort(arrival$Arrival_delay, decreasing = FALSE)

tbl <- cbind(origin,arrival[,2])
colnames(tbl) <- c("Airport","Depart_delay","Arr_delay")
tbl
```

```
##    Airport Depart_delay Arr_delay
## 1      ATL         -128     -1644
## 2      BDL         -123      -978
## 3      BNA          -27      -751
## 4      BOS          -20      -585
## 5      CLT          -12      -448
## 6      CVG          -11      -214
## 7      DEN          -10      -160
## 8      DFW           18      -152
## 9      DTW           51       -57
## 10     EWR           59       -54
## 11     FLL           67       -54
```

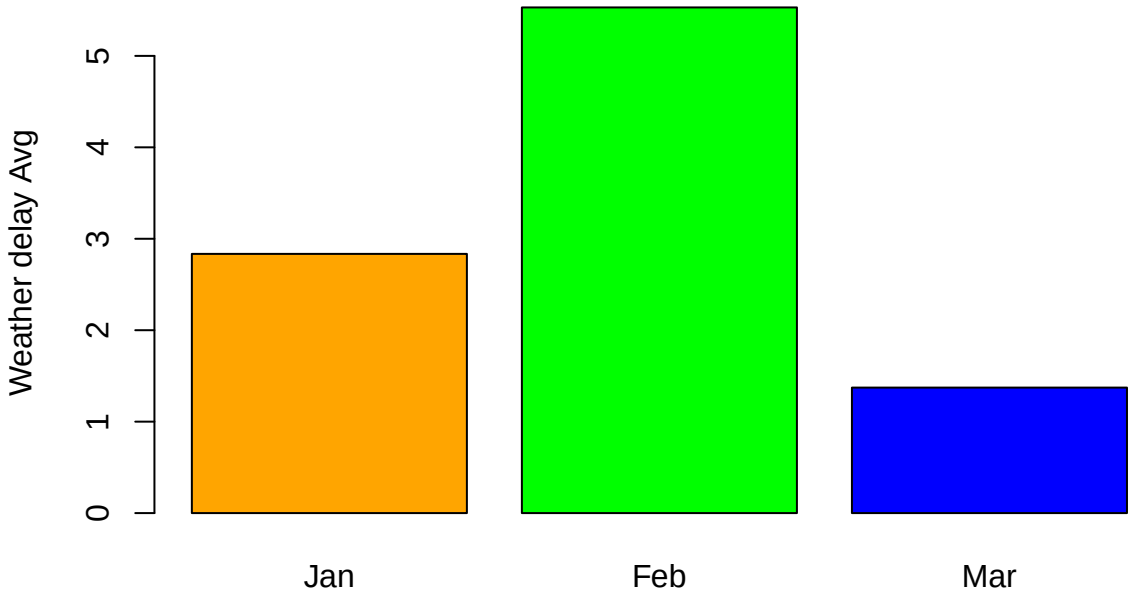
```
## 12     GRR          117       -45
## 13     IAD          223       -39
## 14     IAH          232         1
## 15     JFK          283        20
## 16     LAS          291        66
## 17     LAX          301       161
## 18     LGA          437       232
## 19     MCI          497       237
## 20     MCO          640       276
## 21     MEM          710       301
## 22     MIA          725       328
## 23     MKE          916       599
## 24     MSN         1109       606
## 25     MSP         1160       710
## 26     ORD         1201       845
## 27     PHX         1264       851
## 28     PIT         1271       980
## 29     RDU         1697      1115
## 30     RSW         1756      1214
## 31     SAN         1882      1236
## 32     SEA         2031      1606
## 33     SFO         2150      1865
## 34     SLC         2898      2352
## 35     STL         3169      3203
## 36     TPA         7227      5738
## 37     TTN        14418     14041
## 38     XNA        42798     45115
```

```
## 4. Visualize the association between some variable pairs

# Total Departure delay in Months

ggplot(cvg, aes(DEPARTURE_DELAY, col = dm)) + geom_bar() + ylab("DEPARTURE DELAY") + xlab("") + ggtitle
```

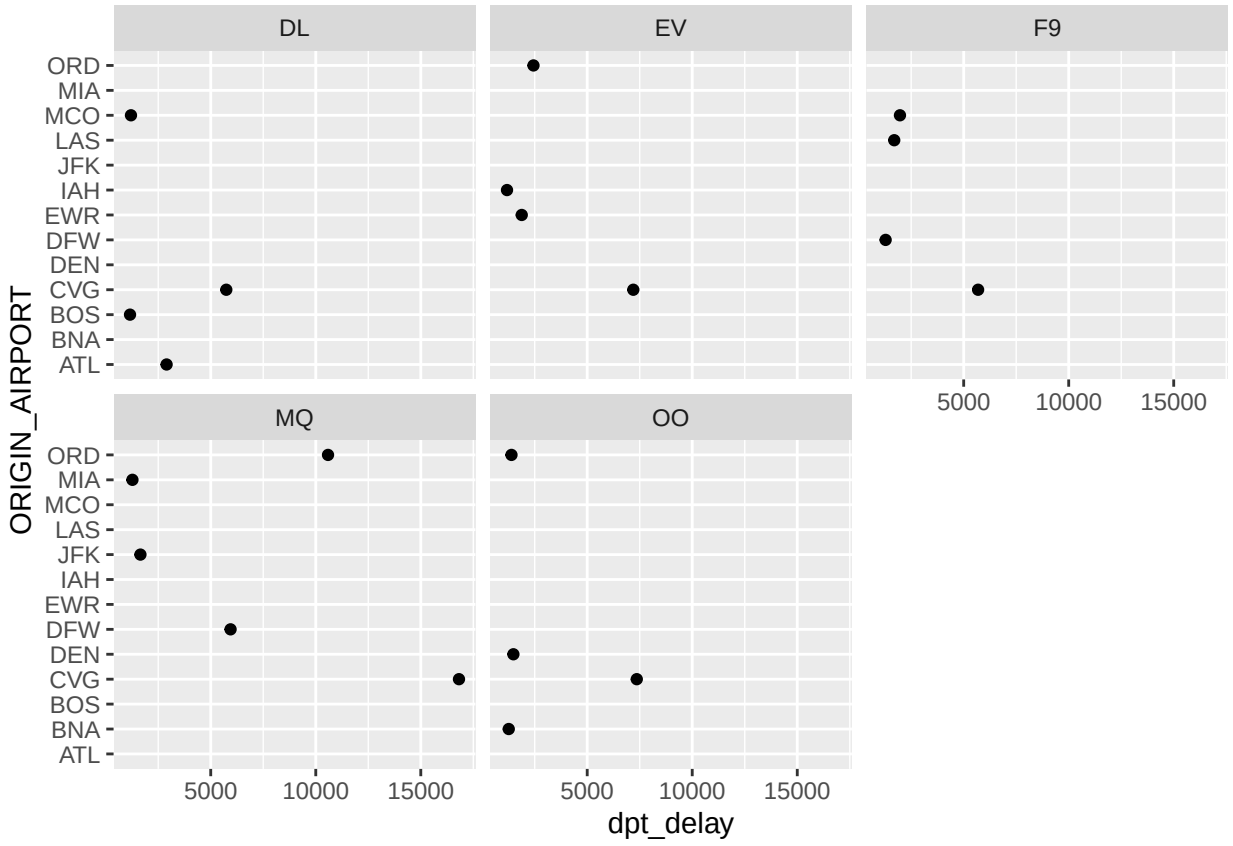
```
## Warning: Removed 370 rows containing non-finite values (stat_count).
```



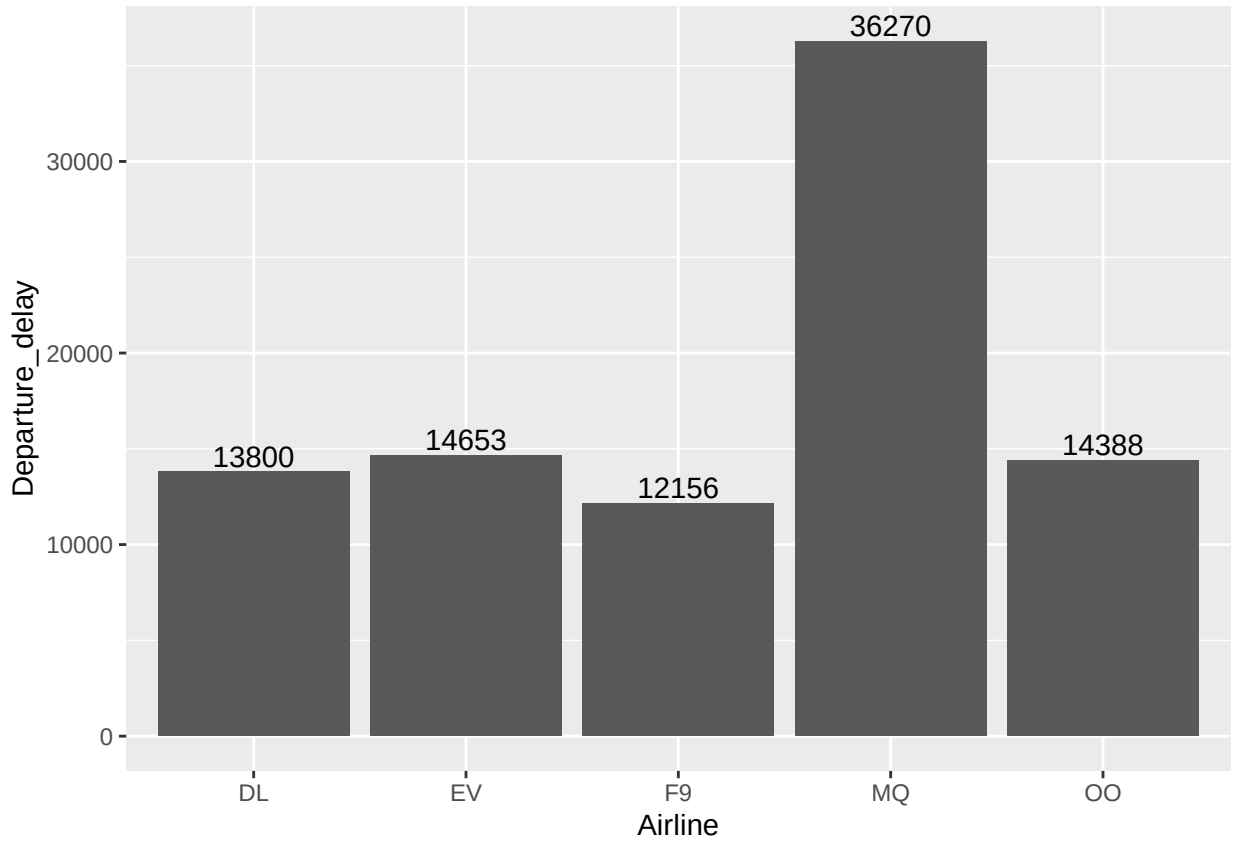

```
# Average delay due to weather in Months

weather <- cvg %>% group_by(dm) %>% filter(WEATHER_DELAY != "NA") %>% summarize(avg_weather_delay = mea

## `summarise()` ungrouping output (override with `.groups` argument)

barplot(weather$avg_weather_delay, names.arg = c("Jan","Feb","Mar"), ylab = "Weather delay Avg", col =
```



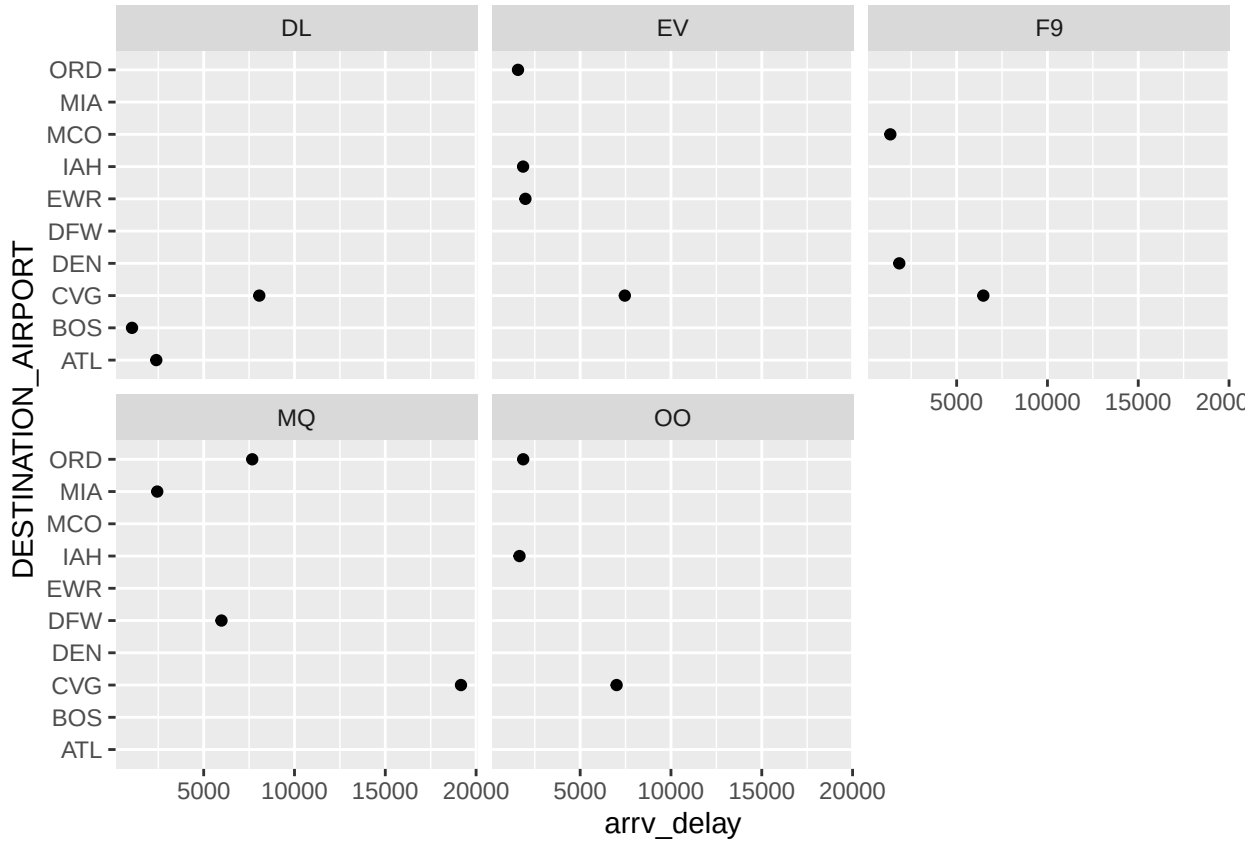

```
# Departure delay in each Origin airport, faceted by Airline

aa <- cvg %>% filter(DEPARTURE_DELAY != "NA") %>% group_by(ORIGIN_AIRPORT, AIRLINE) %>% summarize(dpt_d

## `summarise()` regrouping output by 'ORIGIN_AIRPORT' (override with `.groups` argument)

ggplot(aa, aes(x = dpt_delay, y = ORIGIN_AIRPORT)) + geom_point() + facet_wrap(~ AIRLINE)
```



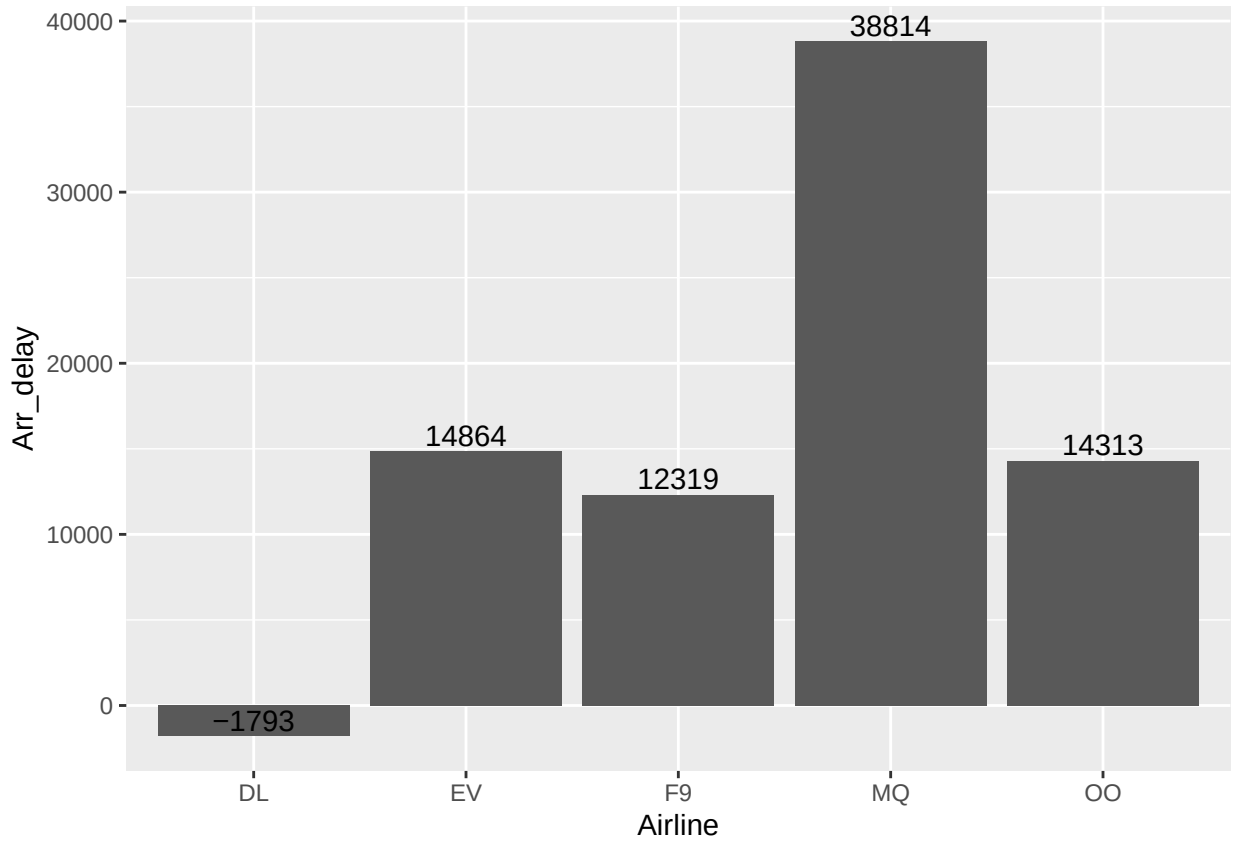

```
# Total departure delay by Airline

delay <- aggregate(cvg$DEPARTURE_DELAY, by = list(cvg$AIRLINE), FUN = sum, na.rm = TRUE)
colnames(delay) <- c("Airline", "Departure_delay")
ggplot(delay, aes(x = Airline, y = Departure_delay)) + geom_col() + geom_text(aes(label = Departure_dela
```




```
# Arrival delay in each Destination airport, faceted by Airline

bb <- cvg %>% filter(ARRIVAL_DELAY != "NA") %>% group_by(DESTINATION_AIRPORT, AIRLINE) %>% summarize(ar

## `summarise()` regrouping output by 'DESTINATION_AIRPORT' (override with `.groups` argument)

ggplot(bb, aes(x = arrv_delay, y = DESTINATION_AIRPORT)) + geom_point() + facet_wrap(~ AIRLINE)
```




```
# Total arrival delay by Airline

arr_delay <- aggregate(cvg$ARRIVAL_DELAY, by = list(cvg$AIRLINE), FUN = sum, na.rm = TRUE)
colnames(arr_delay) <- c("Airline", "Arr_delay")
ggplot(arr_delay, aes(x = Airline, y = Arr_delay)) + geom_col() + geom_text(aes(label = Arr_delay), vju
```




```
## 5. Visualize the association between some variable pairs conditional on some other variables

library(PerformanceAnalytics)
```

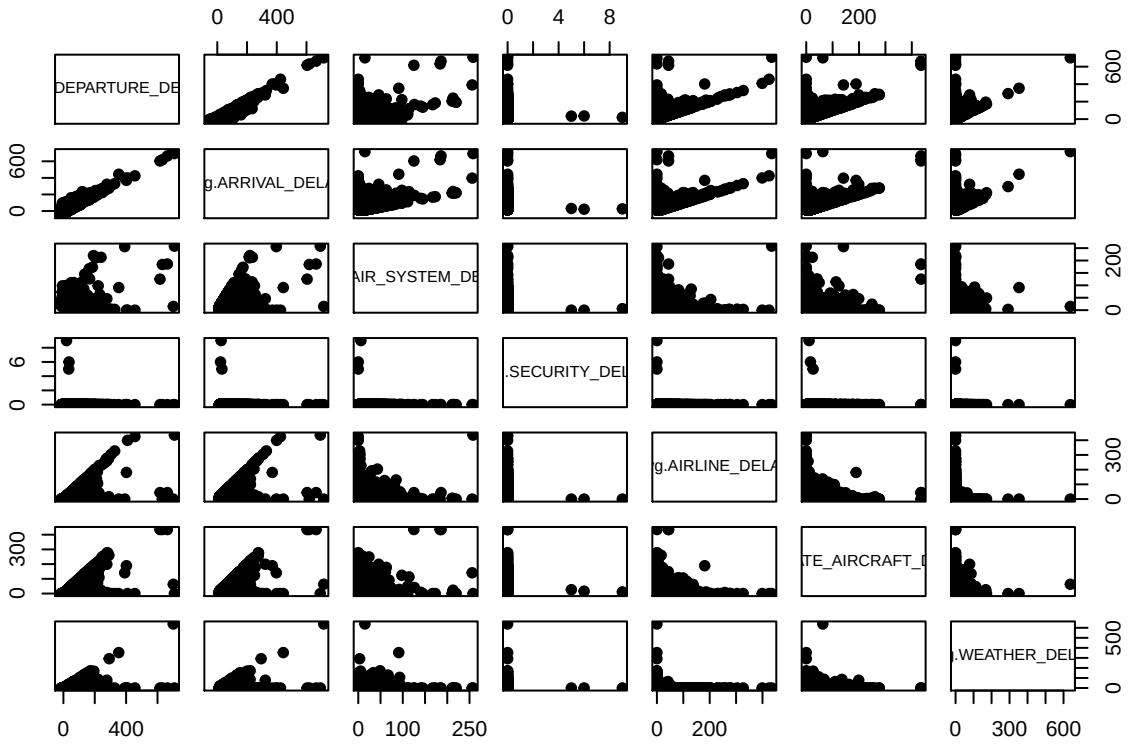
```
## Warning: package 'PerformanceAnalytics' was built under R version 3.6.3

## Loading required package: xts

## Warning: package 'xts' was built under R version 3.6.3

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 3.6.3

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

##
## Attaching package: 'xts'
```

```
## The following objects are masked from 'package:dplyr':
##
##     first, last

##
## Attaching package: 'PerformanceAnalytics'

## The following object is masked from 'package:graphics':
##
##     legend
```

```
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 3.6.3

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

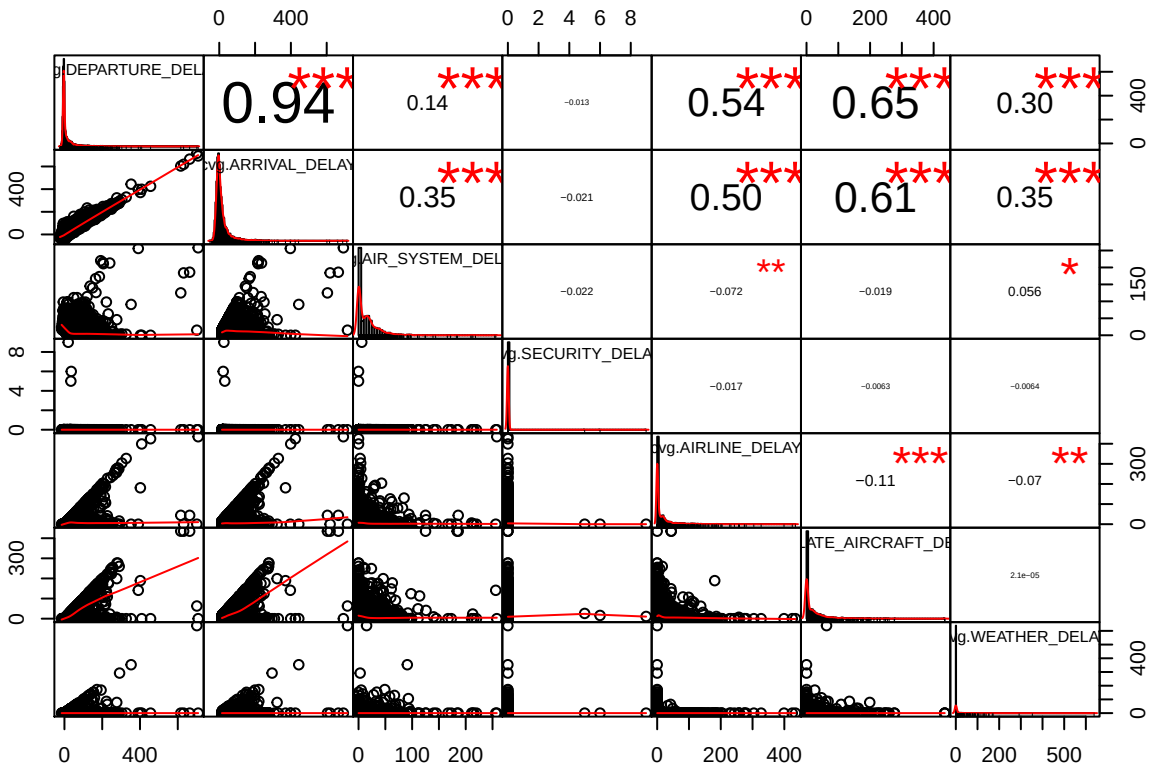
```
cvg.perf <- data.frame(cvg$DEPARTURE_DELAY, cvg$ARRIVAL_DELAY, cvg$AIR_SYSTEM_DELAY, cvg$SECURITY_DELAY
pairs(cvg.perf, pch = 19)
```



```r
round(cor(cvg.perf), digits = 3)
```

```
##                          cvg.DEPARTURE_DELAY cvg.ARRIVAL_DELAY
## cvg.DEPARTURE_DELAY                        1                 NA
## cvg.ARRIVAL_DELAY                         NA                  1
## cvg.AIR_SYSTEM_DELAY                      NA                 NA
## cvg.SECURITY_DELAY                        NA                 NA
## cvg.AIRLINE_DELAY                         NA                 NA
## cvg.LATE_AIRCRAFT_DELAY                   NA                 NA
## cvg.WEATHER_DELAY                         NA                 NA
##                          cvg.AIR_SYSTEM_DELAY cvg.SECURITY_DELAY
## cvg.DEPARTURE_DELAY                        NA                 NA
## cvg.ARRIVAL_DELAY                          NA                 NA
## cvg.AIR_SYSTEM_DELAY                        1                 NA
## cvg.SECURITY_DELAY                         NA                  1
## cvg.AIRLINE_DELAY                          NA                 NA
## cvg.LATE_AIRCRAFT_DELAY                    NA                 NA
## cvg.WEATHER_DELAY                          NA                 NA
##                          cvg.AIRLINE_DELAY cvg.LATE_AIRCRAFT_DELAY
## cvg.DEPARTURE_DELAY                     NA                      NA
## cvg.ARRIVAL_DELAY                       NA                      NA
## cvg.AIR_SYSTEM_DELAY                    NA                      NA
## cvg.SECURITY_DELAY                      NA                      NA
## cvg.AIRLINE_DELAY                        1                      NA
## cvg.LATE_AIRCRAFT_DELAY                 NA                       1
## cvg.WEATHER_DELAY                       NA                      NA
##                          cvg.WEATHER_DELAY
## cvg.DEPARTURE_DELAY                     NA
## cvg.ARRIVAL_DELAY                       NA
## cvg.AIR_SYSTEM_DELAY                    NA
## cvg.SECURITY_DELAY                      NA
## cvg.AIRLINE_DELAY                       NA
## cvg.LATE_AIRCRAFT_DELAY                 NA
## cvg.WEATHER_DELAY                        1
```

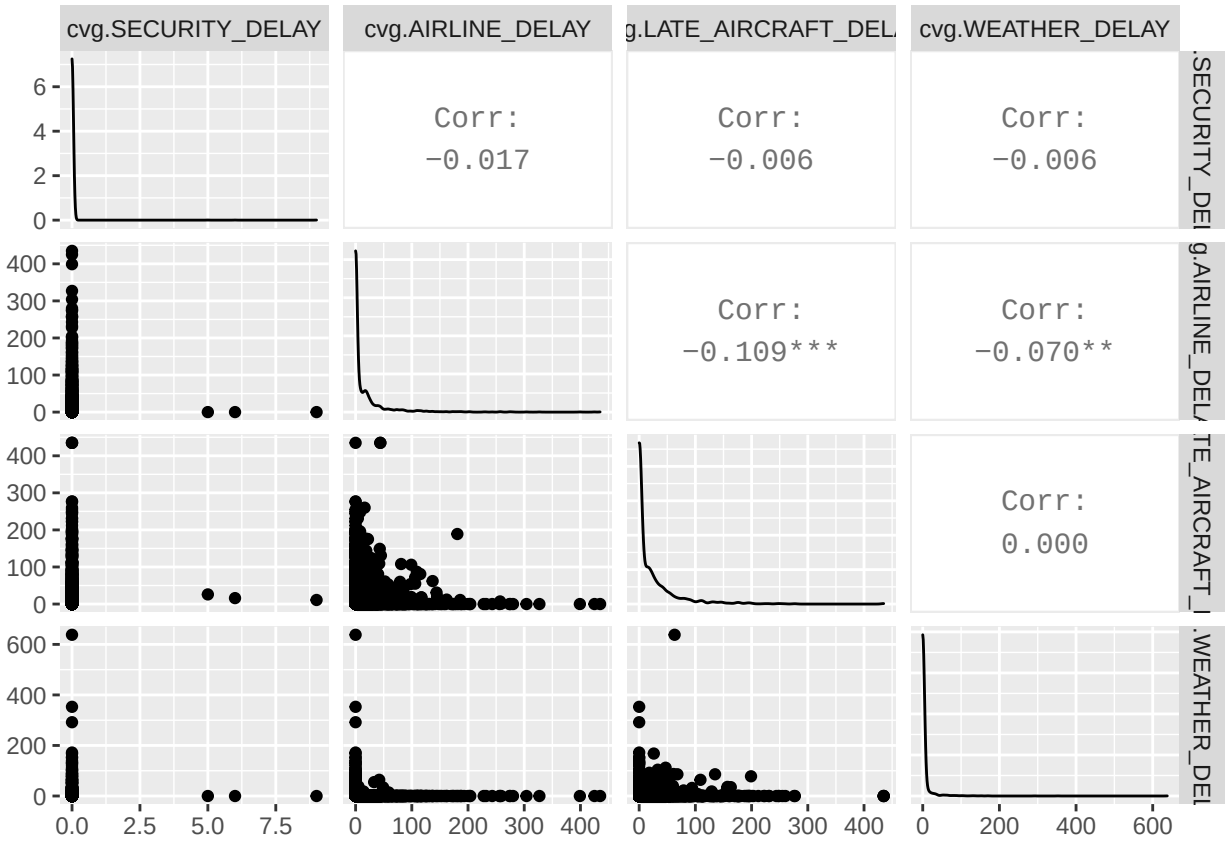
```r
chart.Correlation(R = cvg.perf, histogram = TRUE, pch = 19)
```




```
perf <- data.frame(cvg$SECURITY_DELAY, cvg$AIRLINE_DELAY, cvg$LATE_AIRCRAFT_DELAY, cvg$WEATHER_DELAY)
ggpairs(perf)
```

```
## Warning: Removed 5814 rows containing non-finite values (stat_density).

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 5814 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 5814 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 5814 rows containing missing values

## Warning: Removed 5814 rows containing missing values (geom_point).

## Warning: Removed 5814 rows containing non-finite values (stat_density).

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 5814 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 5814 rows containing missing values
```

```
## Warning: Removed 5814 rows containing missing values (geom_point).

## Warning: Removed 5814 rows containing missing values (geom_point).

## Warning: Removed 5814 rows containing non-finite values (stat_density).

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 5814 rows containing missing values

## Warning: Removed 5814 rows containing missing values (geom_point).

## Warning: Removed 5814 rows containing missing values (geom_point).

## Warning: Removed 5814 rows containing missing values (geom_point).

## Warning: Removed 5814 rows containing non-finite values (stat_density).
```




```
## 6. For airlines.csv and airports,csv data sets, how many variables and how many observations in the
# Airlines data set

airlines <- read.csv("airlines.csv",header = TRUE, stringsAsFactors = FALSE, na.strings = "")

summary(airlines)
```

```
##   IATA_CODE           AIRLINE
##  Length:14          Length:14
##  Class :character   Class :character
##  Mode  :character   Mode  :character
```

```
nrow(airlines)
```

```
## [1] 14
```

```
ncol(airlines)
```

```
## [1] 2
```

```
airl <- apply(airlines, 2,function(airlines) sum(is.na(airlines)))
airl
```

```
## IATA_CODE   AIRLINE
##         0         0
```

```
# Airports data set

airports <- read.csv("airports.csv",header = TRUE, stringsAsFactors = FALSE, na.strings = "")

summary(airports)
```

```
##   IATA_CODE           AIRPORT              CITY              STATE
##  Length:322         Length:322         Length:322         Length:322
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    COUNTRY             LATITUDE       LONGITUDE
##  Length:322         Min.   :13.48   Min.   :-176.65
##  Class :character   1st Qu.:33.65   1st Qu.:-110.84
##  Mode  :character   Median :39.30   Median : -93.40
##                     Mean   :38.98   Mean   : -98.38
##                     3rd Qu.:43.15   3rd Qu.: -82.72
##                     Max.   :71.29   Max.   : -64.80
##                     NA's   :3       NA's   :3
```

```
nrow(airports)
```

```
## [1] 322
```

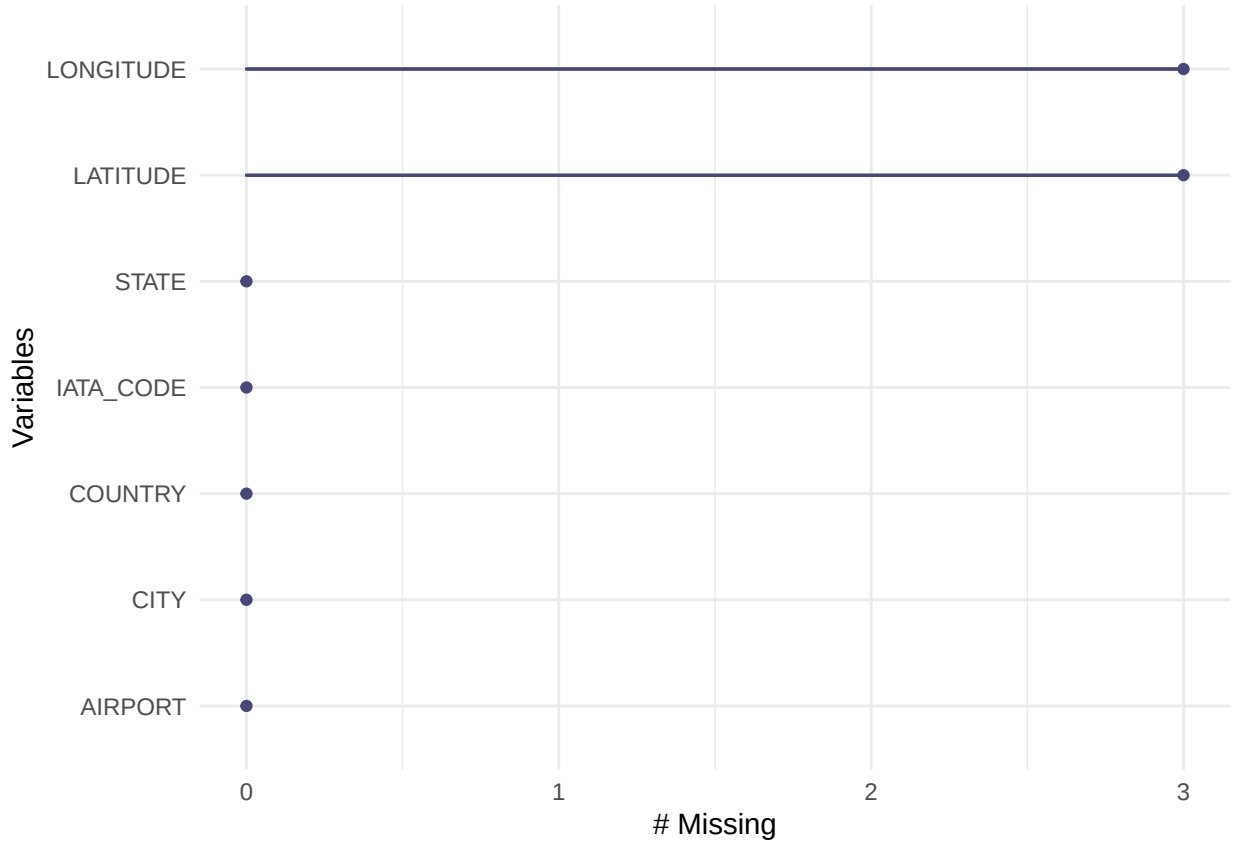
```
nrow(airlines)
```

```
## [1] 14
```

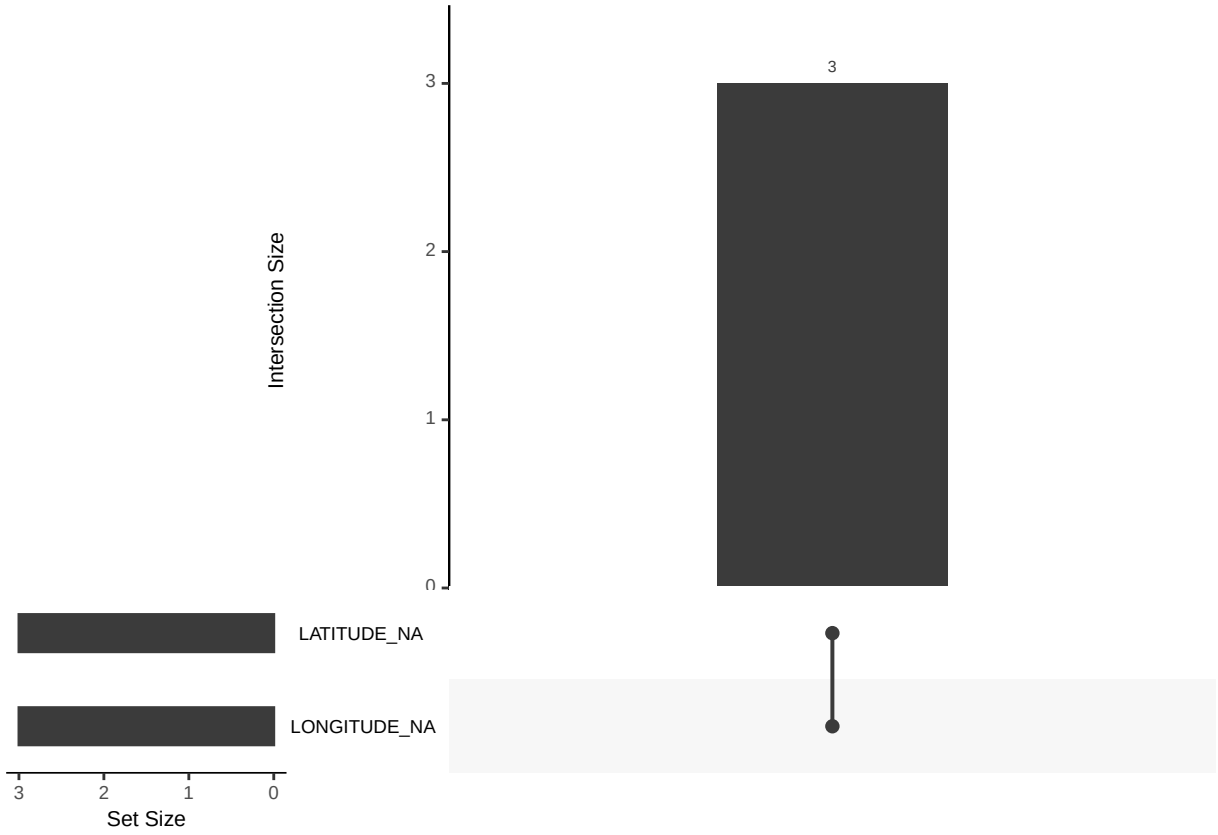
```
airp <- apply(airports, 2,function(airports) sum(is.na(airports)))
airp
```

```
## IATA_CODE   AIRPORT      CITY     STATE   COUNTRY  LATITUDE LONGITUDE
##         0         0         0         0         0         3         3
```

```
gg_miss_var(airports)
```




```
gg_miss_upset(airports)
```




```
## 7. Merge CVG_Flights.csv, airlines.csv, and airports.csv

cvg_airlines <- left_join(cvg, airlines, by = c("AIRLINE"="IATA_CODE"))
airlines_airports <- left_join(airlines, airports, by = "IATA_CODE")

head(cvg_airlines, 10)
```

```
##    FLIGHT_DATE AIRLINE FLIGHT_NUMBER ORIGIN_AIRPORT DESTINATION_AIRPORT
## 1     1/1/2015      MQ          3355            CVG                 MIA
## 2     1/1/2015      MQ          3597            CVG                 DFW
## 3     1/1/2015      OO          4811            BNA                 CVG
## 4     1/1/2015      DL          2315            CVG                 ATL
## 5     1/1/2015      F9          1308            IAD                 CVG
## 6     1/1/2015      DL          1502            CVG                 TPA
## 7     1/1/2015      MQ          3491            DFW                 CVG
## 8     1/1/2015      F9          1290            CVG                 FLL
## 9     1/1/2015      DL          2359            CVG                 RSW
## 10    1/1/2015      F9           935            TTN                 CVG
##    SCHEDULED_DEPARTURE DEPARTURE_TIME DEPARTURE_DELAY TAXI_OUT WHEELS_OFF
## 1           6:00:00 AM     5:56:00 AM              -4       15 6:11:00 AM
## 2           6:00:00 AM           <NA>              NA       NA       <NA>
## 3           6:45:00 AM     6:37:00 AM              -8       16 6:53:00 AM
## 4           7:10:00 AM     7:08:00 AM              -2       36 7:44:00 AM
## 5           7:15:00 AM     7:05:00 AM             -10       11 7:16:00 AM
## 6           7:20:00 AM     7:19:00 AM              -1        9 7:28:00 AM
## 7           7:20:00 AM           <NA>              NA       NA       <NA>
```

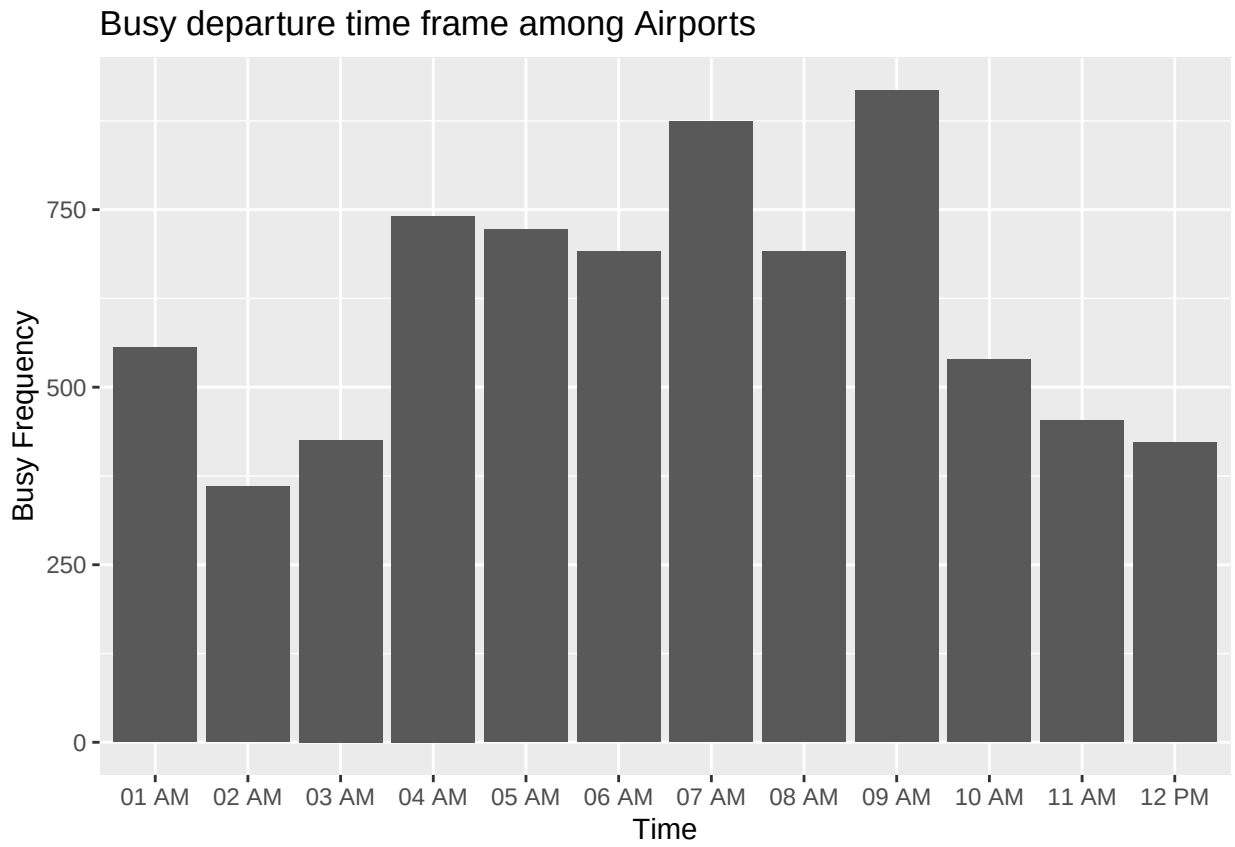
```
## 8           7:25:00 AM     7:23:00 AM             -2       10 7:33:00 AM
## 9           7:31:00 AM     7:22:00 AM             -9       27 7:49:00 AM
## 10          7:40:00 AM     7:51:00 AM             11        9 8:00:00 AM
##    SCHEDULED_TIME ELAPSED_TIME AIR_TIME DISTANCE  WHEELS_ON TAXI_IN
## 1             153          150      123      948 8:14:00 AM      12
## 2             165           NA       NA      812       <NA>      NA
## 3              72           66       45      230 8:38:00 AM       5
## 4             102          101       62      373 8:46:00 AM       3
## 5              95           91       75      388 8:31:00 AM       5
## 6             141          116      105      773 9:13:00 AM       2
## 7             128           NA       NA      812       <NA>      NA
## 8             150          143      126      932 9:39:00 AM       7
## 9             153          150      117      879 9:46:00 AM       6
## 10            127          119      103      532 9:43:00 AM       7
##    SCHEDULED_ARRIVAL ARRIVAL_TIME ARRIVAL_DELAY DIVERTED CANCELLED
## 1         8:33:00 AM   8:26:00 AM            -7        0         0
## 2         7:45:00 AM         <NA>            NA        0         1
## 3         8:57:00 AM   8:43:00 AM           -14        0         0
## 4         8:52:00 AM   8:49:00 AM            -3        0         0
## 5         8:50:00 AM   8:36:00 AM           -14        0         0
## 6         9:41:00 AM   9:15:00 AM           -26        0         0
## 7        10:28:00 AM         <NA>            NA        0         1
## 8         9:55:00 AM   9:46:00 AM            -9        0         0
## 9        10:04:00 AM   9:52:00 AM           -12        0         0
## 10        9:47:00 AM   9:50:00 AM             3        0         0
##    CANCELLATION_REASON AIR_SYSTEM_DELAY SECURITY_DELAY AIRLINE_DELAY
## 1                 <NA>               NA             NA            NA
## 2                    B               NA             NA            NA
## 3                 <NA>               NA             NA            NA
## 4                 <NA>               NA             NA            NA
## 5                 <NA>               NA             NA            NA
## 6                 <NA>               NA             NA            NA
## 7                    B               NA             NA            NA
## 8                 <NA>               NA             NA            NA
## 9                 <NA>               NA             NA            NA
## 10                <NA>               NA             NA            NA
##    LATE_AIRCRAFT_DELAY WEATHER_DELAY dm                    AIRLINE.y
## 1                   NA            NA 01 American Eagle Airlines Inc.
## 2                   NA            NA 01 American Eagle Airlines Inc.
## 3                   NA            NA 01        Skywest Airlines Inc.
## 4                   NA            NA 01         Delta Air Lines Inc.
## 5                   NA            NA 01       Frontier Airlines Inc.
## 6                   NA            NA 01         Delta Air Lines Inc.
## 7                   NA            NA 01 American Eagle Airlines Inc.
## 8                   NA            NA 01       Frontier Airlines Inc.
## 9                   NA            NA 01         Delta Air Lines Inc.
## 10                  NA            NA 01       Frontier Airlines Inc.
```

```
head(airlines_airports,10)
```

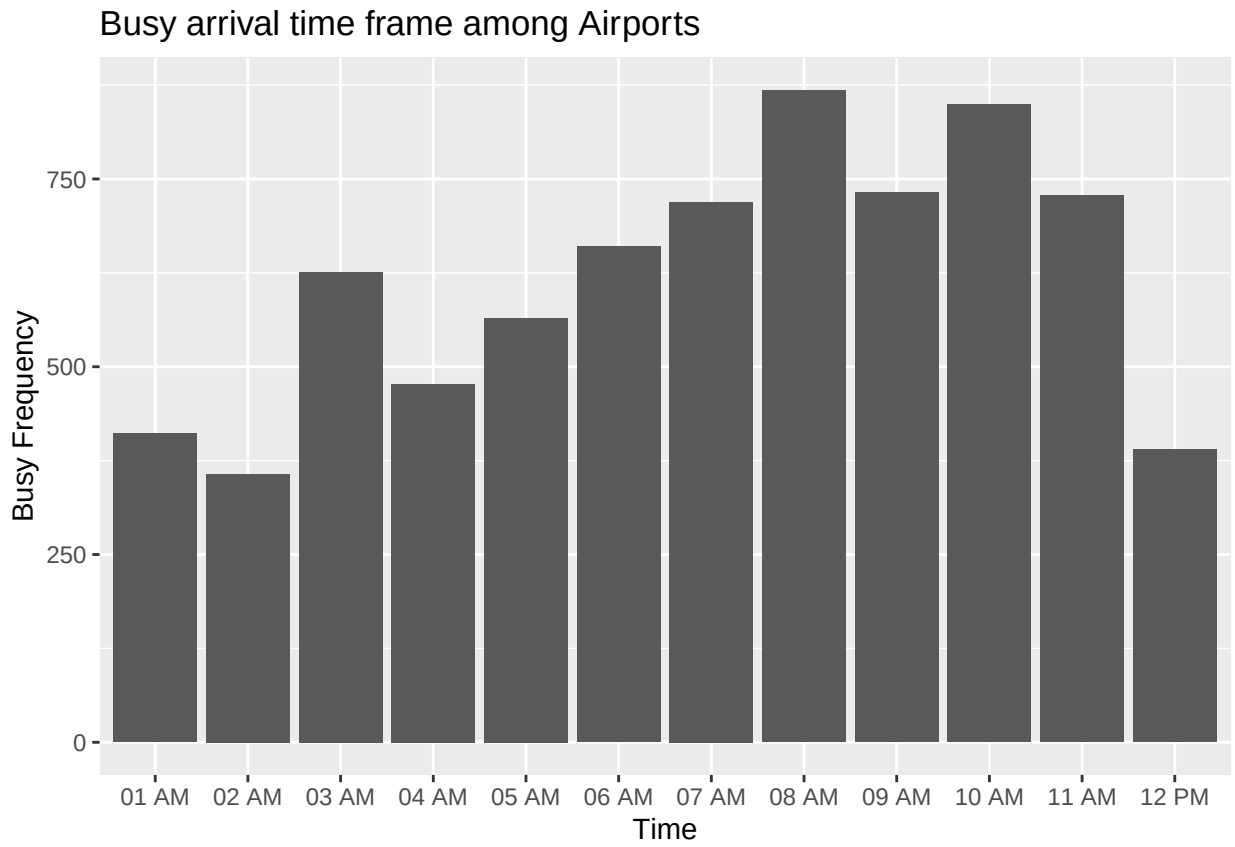
```
##    IATA_CODE                AIRLINE AIRPORT CITY STATE COUNTRY LATITUDE
## 1         UA  United Air Lines Inc.    <NA> <NA>  <NA>    <NA>       NA
## 2         AA American Airlines Inc.    <NA> <NA>  <NA>    <NA>       NA
## 3         US        US Airways Inc.    <NA> <NA>  <NA>    <NA>       NA
```

```
## 4         F9 Frontier Airlines Inc.    <NA> <NA>  <NA>    <NA>       NA
## 5         B6        JetBlue Airways    <NA> <NA>  <NA>    <NA>       NA
## 6         OO  Skywest Airlines Inc.    <NA> <NA>  <NA>    <NA>       NA
## 7         AS   Alaska Airlines Inc.    <NA> <NA>  <NA>    <NA>       NA
## 8         NK       Spirit Air Lines    <NA> <NA>  <NA>    <NA>       NA
## 9         WN Southwest Airlines Co.    <NA> <NA>  <NA>    <NA>       NA
## 10        DL   Delta Air Lines Inc.    <NA> <NA>  <NA>    <NA>       NA
##    LONGITUDE
## 1         NA
## 2         NA
## 3         NA
## 4         NA
## 5         NA
## 6         NA
## 7         NA
## 8         NA
## 9         NA
## 10        NA
```

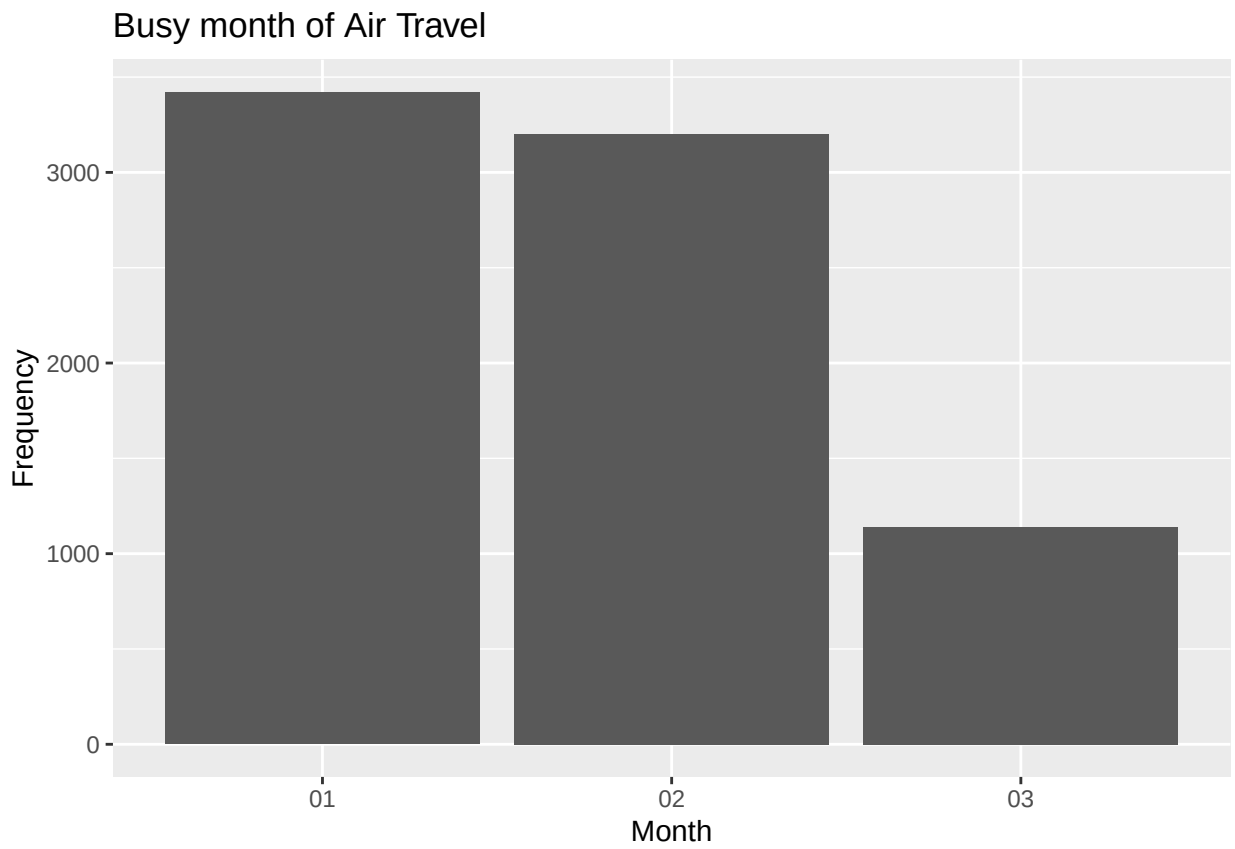
```r
## 8. Questions propose

#Which departure time frame is the busiest among airports

dh <- as.POSIXct(cvg$DEPARTURE_TIME, format = "%H:%M:%S %p")
cvg$dp <- format(dh, "%H %p")

cvg %>%
  filter(!is.na(dp)) %>%
  ggplot(aes(x= dp)) +
  geom_bar() + ylab("Busy Frequency") + xlab("Time") + ggtitle("Busy departure time frame among Airports
```




```
# Which arrival time frame is the busiest among airports

ah <- as.POSIXct(cvg$ARRIVAL_TIME, format = "%H:%M:%S %p")
cvg$ap <- format(ah, "%H %p")

cvg %>%
  filter(!is.na(ap)) %>%
  ggplot(aes(x= ap)) +
  geom_bar() + ylab("Busy Frequency") + xlab("Time") + ggtitle("Busy arrival time frame among Airports")
```

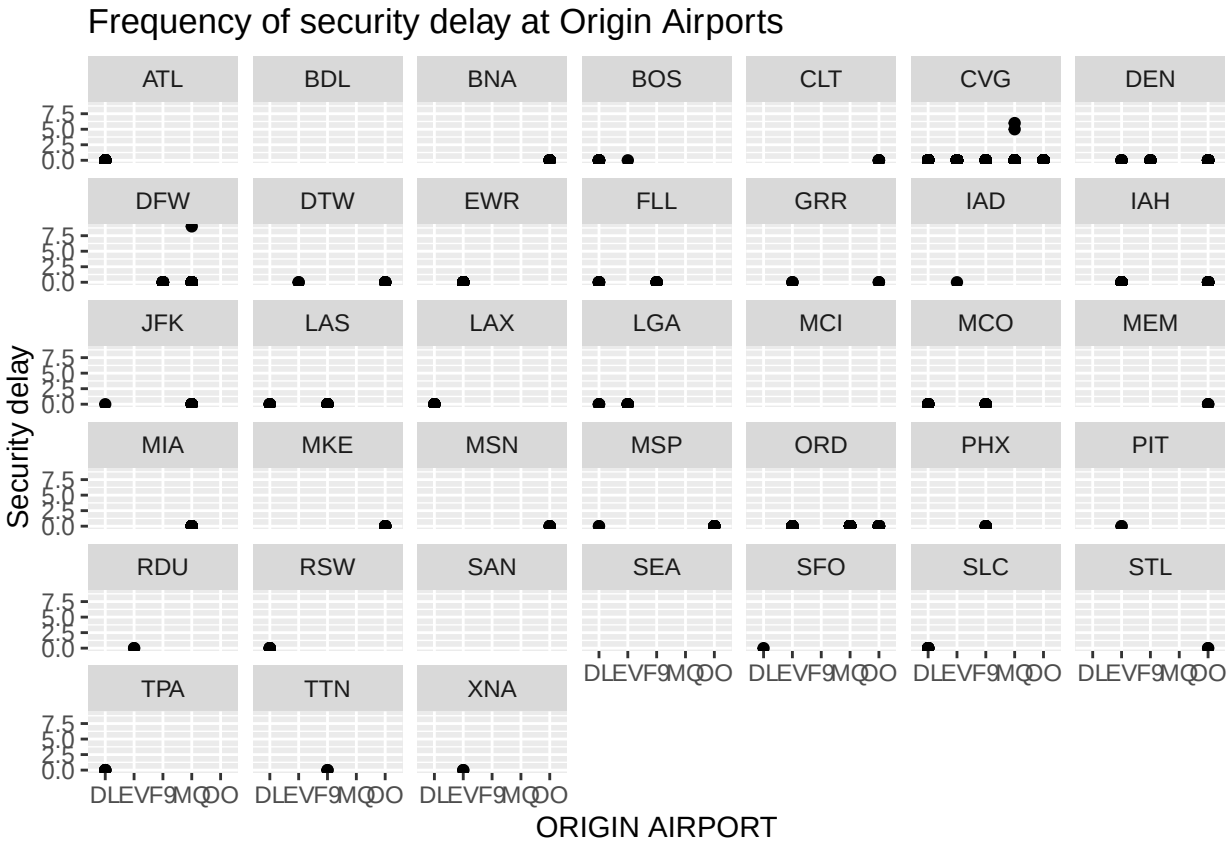



```
# Which month is the busiest months

d<- as.POSIXct(cvg$FLIGHT_DATE, format = "%m/%d/%Y")
cvg$dm <- format(d, "%m")

ggplot(cvg,aes(dm)) + geom_bar() + ylab("Frequency") + xlab("Month") + ggtitle("Busy month of Air Trave
```



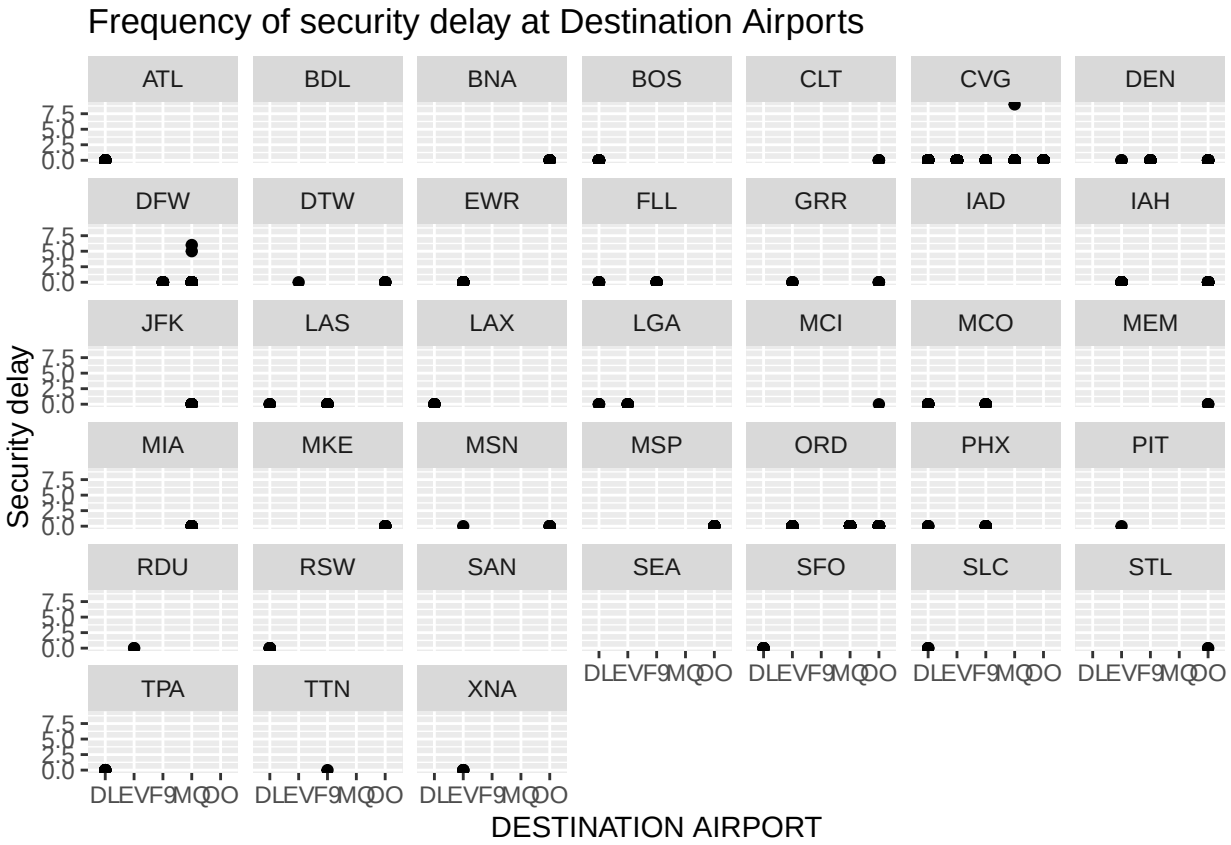

```
# Which airline has high security delay, at which origin airport

ggplot(cvg, aes(AIRLINE, SECURITY_DELAY)) + geom_point() + facet_wrap(~ORIGIN_AIRPORT) + ylab("Security
```

```
## Warning: Removed 5814 rows containing missing values (geom_point).
```




```
# Which airline has high security delay, at which destination airport

ggplot(cvg, aes(AIRLINE, SECURITY_DELAY)) + geom_point() + facet_wrap(~DESTINATION_AIRPORT) + ylab("Secu
```

```
## Warning: Removed 5814 rows containing missing values (geom_point).
```


Frequency of security delay at Destination Airports
ATL
BDL
BNA
BOS
CLT
CVG
DEN
DFW
DTW
EWR
FLL
GRR
IAD
IAH
JFK
LAS
LAX
LGA
MCI
MCO
MEM
MIA
MKE
MSN
MSP
ORD
PHX
PIT
RDU
RSW
SAN
SEA
SFO
SLC
STL
TPA
TTN
XNA
Security delay
0.0
2.5
5.0
7.5
DL
EV
F9
MQ
OO
DESTINATION AIRPORT