Sta 325 Final Project

Calleigh Smith, Hannah Bogomilsky, Hugh Esterson, Maria Henriquez, Mariana Izon

11/22/2020

```
library(readr)
library(dplyr)
library(tidyverse)
```

```
flights <- read_csv("data/flights.csv")
```

```
## Warning: 1 parsing failure.
##  row               col           expected actual                file
## 1143 CANCELLATION_CODE 1/0/T/F/TRUE/FALSE      A 'data/flights.csv'
```

```
unique(flights$OP_CARRIER)
```

```
## [1] "AA" "DL" "B6" "AS"
```

```
unique(flights$DEST)
```

```
##  [1] "LAX" "SFO" "SJC" "SAN" "PSP" "SMF" "OAK" "LGB" "ONT" "BUR"
```

```
class(flights$CARRIER_DELAY)
```

```
## [1] "numeric"
```

```
flights <- flights %>%
  mutate(CARRIER_DELAY = case_when(CARRIER_DELAY > 0 ~ 1),
         WEATHER_DELAY = case_when(WEATHER_DELAY > 0 ~ 1),
         NAS_DELAY = case_when(NAS_DELAY > 0 ~ 1),
         SECURITY_DELAY = case_when(SECURITY_DELAY > 0 ~ 1),
         LATE_AIRCRAFT_DELAY = case_when(LATE_AIRCRAFT_DELAY > 0 ~ 1))
```

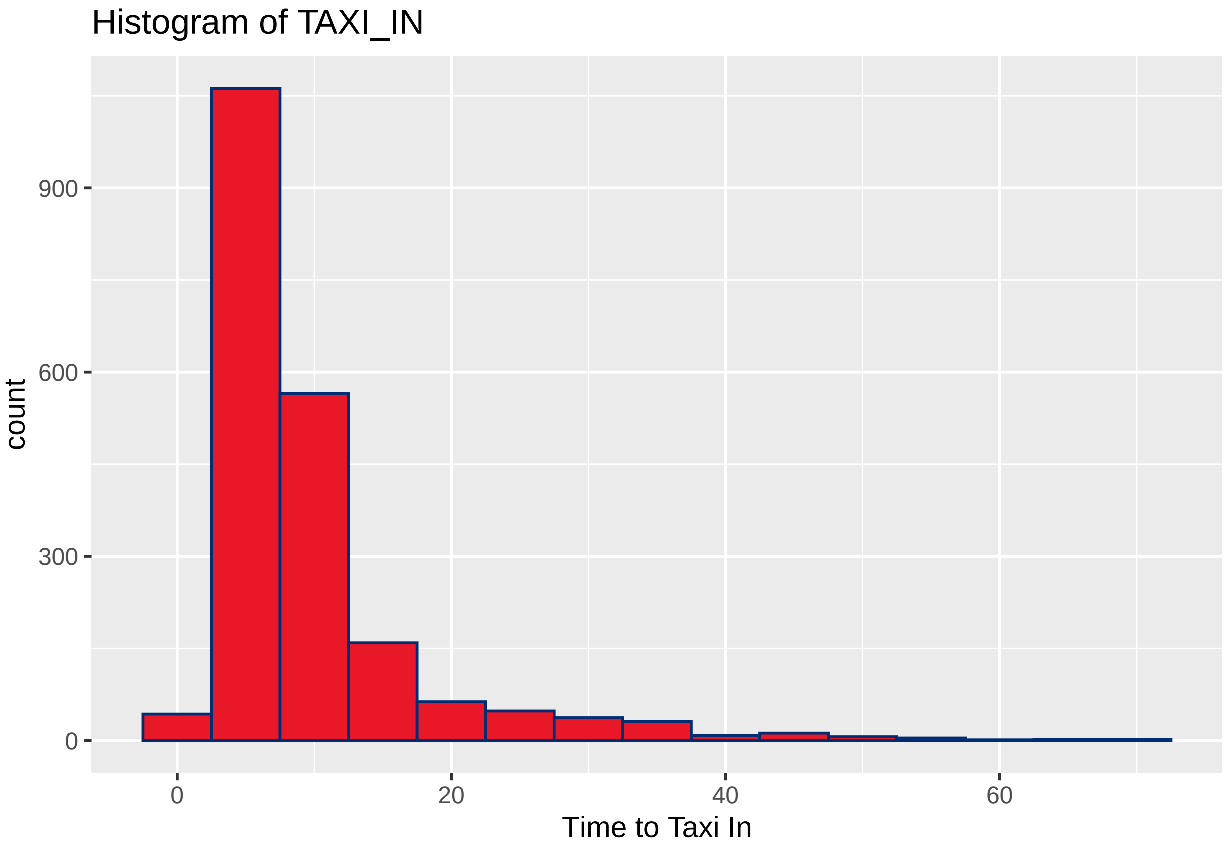
```
flights
```

```
## # A tibble: 2,044 x 34
##     YEAR MONTH DAY_OF_MONTH DAY_OF_WEEK FL_DATE    OP_CARRIER TAIL_NUM
##    <dbl> <dbl>        <dbl>       <dbl> <date>     <chr>      <chr>
##  1  2020     1            1           3 2020-01-01 AA         N110AN
##  2  2020     1            2           4 2020-01-02 AA         N111ZM
##  3  2020     1            3           5 2020-01-03 AA         N108NN
##  4  2020     1            4           6 2020-01-04 AA         N102NN
##  5  2020     1            5           7 2020-01-05 AA         N113AN
##  6  2020     1            6           1 2020-01-06 AA         N103NN
##  7  2020     1            7           2 2020-01-07 AA         N113AN
##  8  2020     1            8           3 2020-01-08 AA         N106NN
##  9  2020     1            9           4 2020-01-09 AA         N102NN
## 10  2020     1           10           5 2020-01-10 AA         N117AN
## # ... with 2,034 more rows, and 27 more variables: OP_CARRIER_FL_NUM <dbl>,
## #   ORIGIN <chr>, ORIGIN_CITY_NAME <chr>, DEST <chr>, DEST_CITY_NAME <chr>,
## #   CRS_DEP_TIME <dbl>, DEP_TIME <dbl>, DEP_DELAY <dbl>, TAXI_OUT <dbl>,
## #   WHEELS_OFF <dbl>, WHEELS_ON <dbl>, TAXI_IN <dbl>, CRS_ARR_TIME <dbl>,
## #   ARR_TIME <dbl>, ARR_DELAY <dbl>, CANCELLED <dbl>, CANCELLATION_CODE <lgl>,
## #   DIVERTED <dbl>, CRS_ELAPSED_TIME <dbl>, ACTUAL_ELAPSED_TIME <dbl>,
## #   AIR_TIME <dbl>, DISTANCE <dbl>, CARRIER_DELAY <dbl>, WEATHER_DELAY <dbl>,
## #   NAS_DELAY <dbl>, SECURITY_DELAY <dbl>, LATE_AIRCRAFT_DELAY <dbl>
```

Taxi Histograms

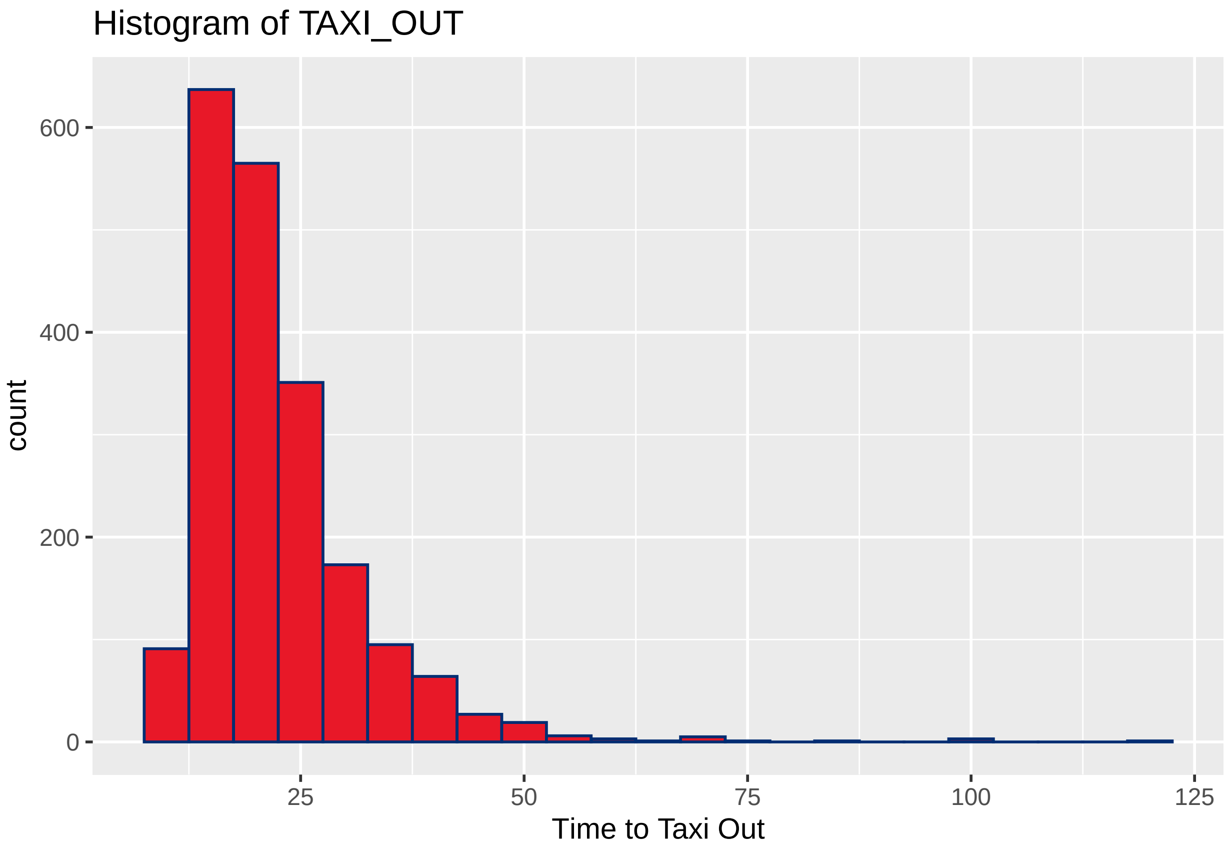
```
ggplot(data = flights, aes(x = TAXI_IN)) +
  geom_histogram(binwidth = 5, fill = "#E81828", color = "#002D72") +
  labs(x = "Time to Taxi In",
       title = "Histogram of TAXI_IN")
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```



```
ggplot(data = flights, aes(x = TAXI_OUT)) +
  geom_histogram(binwidth = 5, fill = "#E81828", color = "#002D72") +
  labs(x = "Time to Taxi Out",
       title = "Histogram of TAXI_OUT")
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```

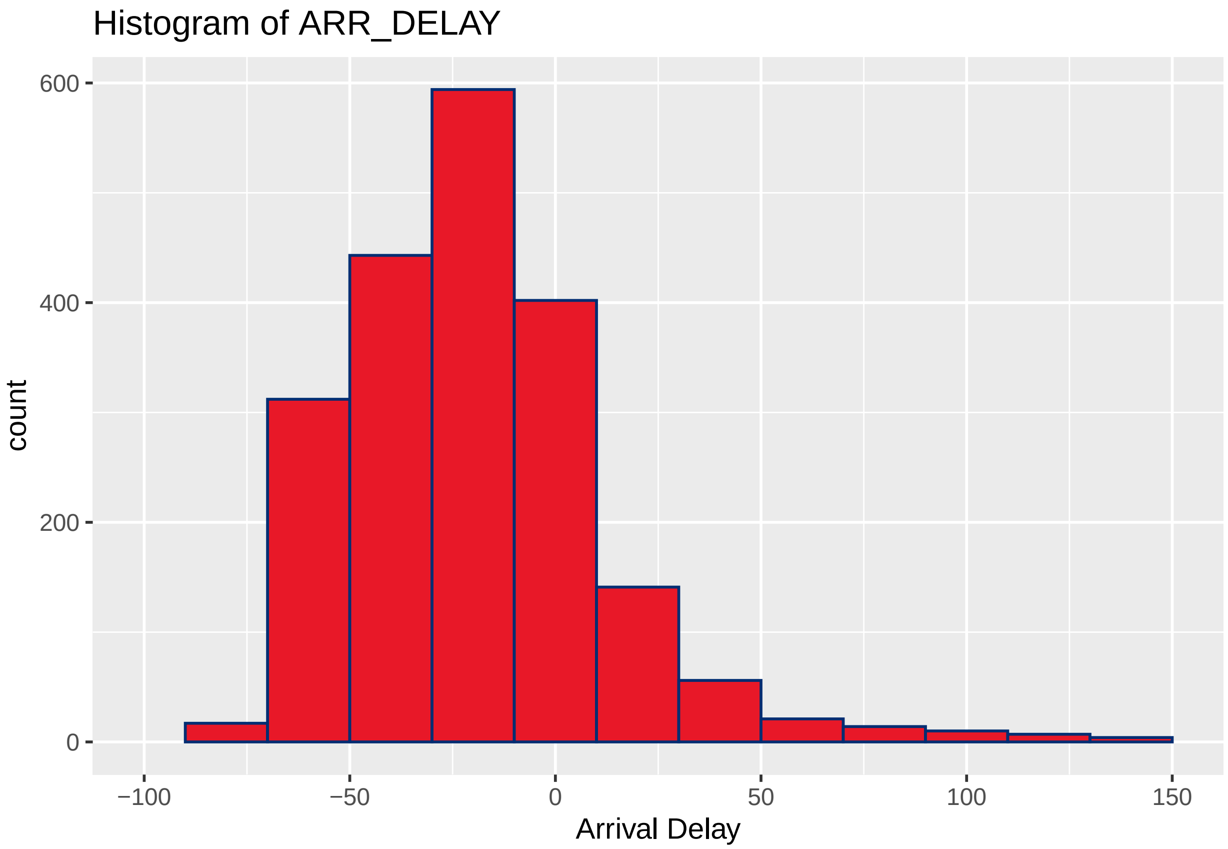


Delay Histograms

```
ggplot(data = flights, aes(x = ARR_DELAY)) +
  geom_histogram(binwidth = 20, fill = "#E81828", color = "#002D72") +
  xlim(-100, 150) +
  labs(x = "Arrival Delay",
       title = "Histogram of ARR_DELAY")
```

```
## Warning: Removed 22 rows containing non-finite values (stat_bin).
```

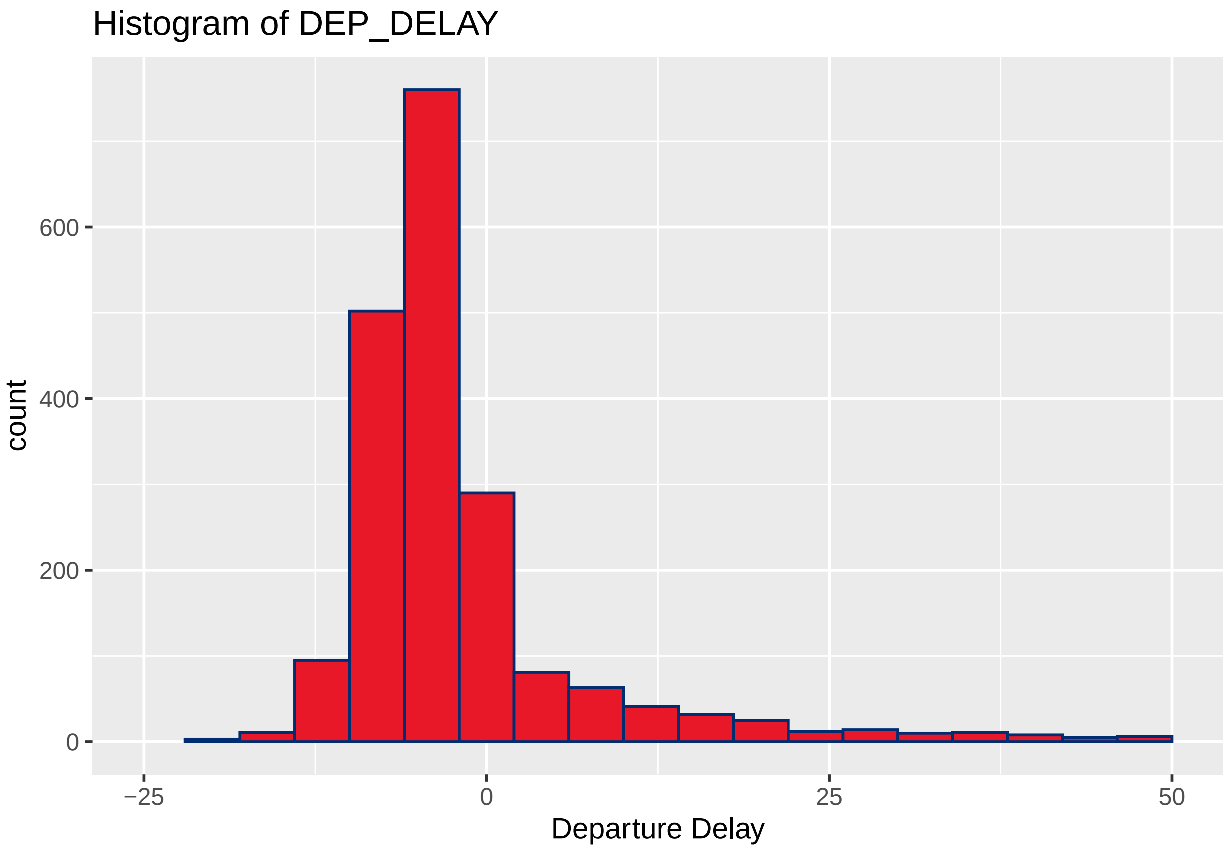
```
## Warning: Removed 1 rows containing missing values (geom_bar).
```



```r
ggplot(data = flights, aes(x = DEP_DELAY)) +
  geom_histogram(binwidth = 4, fill = "#E81828", color = "#002D72") +
  xlim(-25, 50) +
  labs(x = "Departure Delay",
       title = "Histogram of DEP_DELAY")
```

```
## Warning: Removed 75 rows containing non-finite values (stat_bin).

## Warning: Removed 1 rows containing missing values (geom_bar).
```


Histogram of DEP_DELAY
count
600
400
200
0
−25
0
25
50
Departure Delay