Lab3.2

AUTHOR
Dina Al Jibori

```
library(tidyverse)
```

```
── Attaching core tidyverse packages ────────────────────────
tidyverse 2.0.0 ──
✔ dplyr     1.1.2     ✔ readr     2.1.4
✔ forcats   1.0.0     ✔ stringr   1.5.0
✔ ggplot2   3.4.3     ✔ tibble    3.2.1
✔ lubridate 1.9.2     ✔ tidyr     1.3.0
✔ purrr     1.0.2
── Conflicts ──────────────────────────────────────────
tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>)
to force all conflicts to become errors
```

```
library(stat20data)
library(lubridate)
library(ggthemes)
glimpse(flights)
```

```
Rows: 113,013
Columns: 19
$ year           <dbl> 2020, 2020, 2020, 2020, 2020, 2020,
2020, 2020, 2020, 2…
$ month          <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1…
$ day            <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1…
$ dep_time       <dbl> 8, 29, 37, 41, 44, 48, 49, 506, 528,
540, 550, 551, 555…
$ sched_dep_time <dbl> 2359, 39, 40, 45, 2300, 56, 56, 515,
530, 536, 600, 555…
$ dep_delay      <dbl> 9, -10, -3, -4, 104, -8, -7, -9, -2, 4,
-10, -4, -9, -5…
$ arr_time       <dbl> 528, 356, 846, 908, 834, 641, 614,
1050, 812, 1303, 803…
$ sched_arr_time <dbl> 532, 420, 856, 913, 709, 658, 634,
1101, 820, 1332, 810…
$ arr_delay      <dbl> -4, -24, -10, -5, 85, -17, -20, -11,
-8, -29, -7, -26, …
```

```
$ carrier        <chr> "UA", "F9", "UA", "AA", "AA", "UA",
"UA", "UA", "WN", "…
$ flight         <dbl> 521, 162, 197, 794, 289, 168, 694, 710,
1310, 408, 54, …
$ tailnum        <chr> "N76522", "N342FR", "N17126", "N907AA",
"N165US", "N778…
$ origin         <chr> "SFO", "SFO", "SFO", "SFO", "SFO",
"SFO", "SFO", "SFO",…
$ dest           <chr> "AUS", "DEN", "EWR", "MIA", "PHL",
"ORD", "IAH", "IAH",…
$ air_time       <dbl> 175, 125, 285, 296, 271, 214, 189, 196,
87, 246, 110, 1…
$ distance       <dbl> 1504, 967, 2565, 2585, 2521, 1846,
1635, 1635, 646, 229…
$ hour           <dbl> 23, 0, 0, 0, 23, 0, 0, 5, 5, 5, 6, 5,
6, 6, 5, 6, 6, 6,…
$ minute         <dbl> 59, 39, 40, 45, 0, 56, 56, 15, 30, 36,
0, 55, 4, 0, 59,…
$ time_hour      <dttm> 2020-01-01 23:00:00, 2020-01-01
00:00:00, 2020-01-01 0…
```

```r
rm(list = ls())
```

Question 1

```r
# month <= 3 , month >= 5 ,
filter(flights, dest=='PDX' , month >= 3 & month <= 5)
```

```
# A tibble: 811 × 19
    year month   day dep_time sched_dep_time dep_delay
arr_time sched_arr_time
   <dbl> <dbl> <dbl>    <dbl>          <dbl>     <dbl>
<dbl>          <dbl>
 1  2020     3     1      648            700       -12
832            855
 2  2020     3     1      827            830        -3
1028           1025
 3  2020     3     1      838            845        -7
1008           1030
 4  2020     3     1      910            910         0
1123           1105
 5  2020     3     1      931            915        16
1110           1054
 6  2020     3     1     1037           1040        -3
1248           1239
 7  2020     3     1     1114           1115        -1
```

```
1246           1300
 8  2020     3     1     1235           1215        20
1422           1410
 9  2020     3     1     1424           1430        -6
1613           1620
10  2020     3     1     1443           1450        -7
1631           1645
# ℹ 801 more rows
# ℹ 11 more variables: arr_delay <dbl>, carrier <chr>, flight
<dbl>,
#   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>,
distance <dbl>,
#   hour <dbl>, minute <dbl>, time_hour <dttm>
```

There were 811 flights in 2020.

Question 2

```
flights <- flights %>% mutate(avg_speed = distance/(air_time/60
```

Question 3

```
arrange(flights, desc(dep_delay))
```

```
# A tibble: 113,013 × 20
    year month   day dep_time sched_dep_time dep_delay
arr_time sched_arr_time
   <dbl> <dbl> <dbl>    <dbl>          <dbl>     <dbl>
<dbl>          <dbl>
 1  2020     3     6     1407            907      1740
1722           1213
 2  2020     2    20     1604           1245      1639
1900           1538
 3  2020     3     2     1247           1140      1507
2049           1958
 4  2020     2    12      955            907      1488
1246           1215
 5  2020     1    24     1005            952      1453
1303           1245
 6  2020     3     6      816           1111      1265
1611           1914
 7  2020     2    29     1046           1403      1243
1221           1529
 8  2020     2    12      828           1253      1175
1647           2124
 9  2020     2    14      655           1125      1170
```

```
1015             1435
10  2020     1    13     1238            1800      1118
1423             1933
# ℹ 113,003 more rows
# ℹ 12 more variables: arr_delay <dbl>, carrier <chr>, flight
<dbl>,
#   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>,
distance <dbl>,
#   hour <dbl>, minute <dbl>, time_hour <dttm>, avg_speed
<dbl>
```

```
arrange(flights, dep_delay)
```

```
# A tibble: 113,013 × 20
    year month   day dep_time sched_dep_time dep_delay
arr_time sched_arr_time
   <dbl> <dbl> <dbl>    <dbl>          <dbl>     <dbl>
<dbl>          <dbl>
 1  2020     3    31     1930            2010       -40
2119           2222
 2  2020     3    28     1540            1615       -35
2329             48
 3  2020    11    19     2341              16       -35
509            558
 4  2020     3    20     1303            1334       -31
1424           1446
 5  2020     5    24      804             834       -30
1115           1144
 6  2020     3    30     1851            1920       -29
2033           2100
 7  2020     9    10     1834            1903       -29
1959           2030
 8  2020     4     3     2057            2125       -28
2223           2301
 9  2020     6    26     2045            2113       -28
2243           2312
10  2020     3    17     1708            1735       -27
1840           1909
# ℹ 113,003 more rows
# ℹ 12 more variables: arr_delay <dbl>, carrier <chr>, flight
<dbl>,
#   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>,
distance <dbl>,
#   hour <dbl>, minute <dbl>, time_hour <dttm>, avg_speed
<dbl>
```

The flight that holds the record for longest departure delay is flight 576 that was delayed by 29 hours and it's destination was Pheonix, Arizona. The flight that holds the record for shortest departure delay is flight 915 that left earlier by 0.67 hours (40 minutes) and it's destination was Spokane, Washington.

Question 4

```
flights %>%
    summarize(max_delay = max(dep_delay),
              min_delay = min(dep_delay),
              n = n())
```

```
# A tibble: 1 × 3
  max_delay min_delay      n
      <dbl>     <dbl>  <int>
1      1740       -40 113013
```

Question 5

```
flights %>%
    group_by(origin) %>%
    summarize(mean_delay = mean(dep_delay),
              n = n())
```

```
# A tibble: 2 × 3
  origin mean_delay     n
  <chr>       <dbl> <int>
1 OAK         0.430 28668
2 SFO         1.92  84345
```

The mean departure delay of flights leaving from Oakland is 0.43 hours and and San Francisco is 1.92 hours.

Question 6

```
flights %>%
    summarize(prop_no_delay = mean(dep_delay <= 0),
              n = n())
```

```
# A tibble: 1 × 2
  prop_no_delay      n
          <dbl>  <int>
1         0.812 113013
```

```
flights %>%
    filter(origin == "SFO") %>%
    summarize(prop_no_delay = mean(dep_delay <= 0),
              n = n())
```

```
# A tibble: 1 × 2
  prop_no_delay     n
          <dbl> <int>
1         0.818 84345
```

```
flights %>%
    filter(origin == "OAK") %>%
    summarize(prop_no_delay = mean(dep_delay <= 0),
              n = n())
```

```
# A tibble: 1 × 2
  prop_no_delay     n
          <dbl> <int>
1         0.792 28668
```

The proportion of the flights left on or ahead of schedule are 81.1%. For Oakland, the proportion of the flights left on or ahead of schedule are 81.8%. For SFO, the proportion of flights left on or ahead of schedule are 79.2%

Question 7

```
flights %>%
  filter(origin == "SFO" & month == 3 & year == 2020)
```

```
# A tibble: 11,536 × 20
    year month   day dep_time sched_dep_time dep_delay
arr_time sched_arr_time
   <dbl> <dbl> <dbl>    <dbl>          <dbl>     <dbl>
<dbl>          <dbl>
 1  2020     3     1       12           2359        13
856            830
 2  2020     3     1       24             29        -5
808            854
 3  2020     3     1       31             35        -4
539            556
 4  2020     3     1       32           2305        87
624            518
 5  2020     3     1       37             40        -3
626            646
```

```
 6  2020     3     1       42            45        -3
825            910
 7  2020     3     1       45          2330        75
923            745
 8  2020     3     1       58            59        -1
410            440
 9  2020     3     1      122          2355        87
639            535
10  2020     3     1      506           513        -7
822            845
# ℹ 11,526 more rows
# ℹ 12 more variables: arr_delay <dbl>, carrier <chr>, flight
<dbl>,
#   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>,
distance <dbl>,
#   hour <dbl>, minute <dbl>, time_hour <dttm>, avg_speed
<dbl>
```

11,536 flights left SFO during March 2020

Question 8

```
flights %>%
  filter(origin == "SFO" & month == 4 & year == 2020)
```

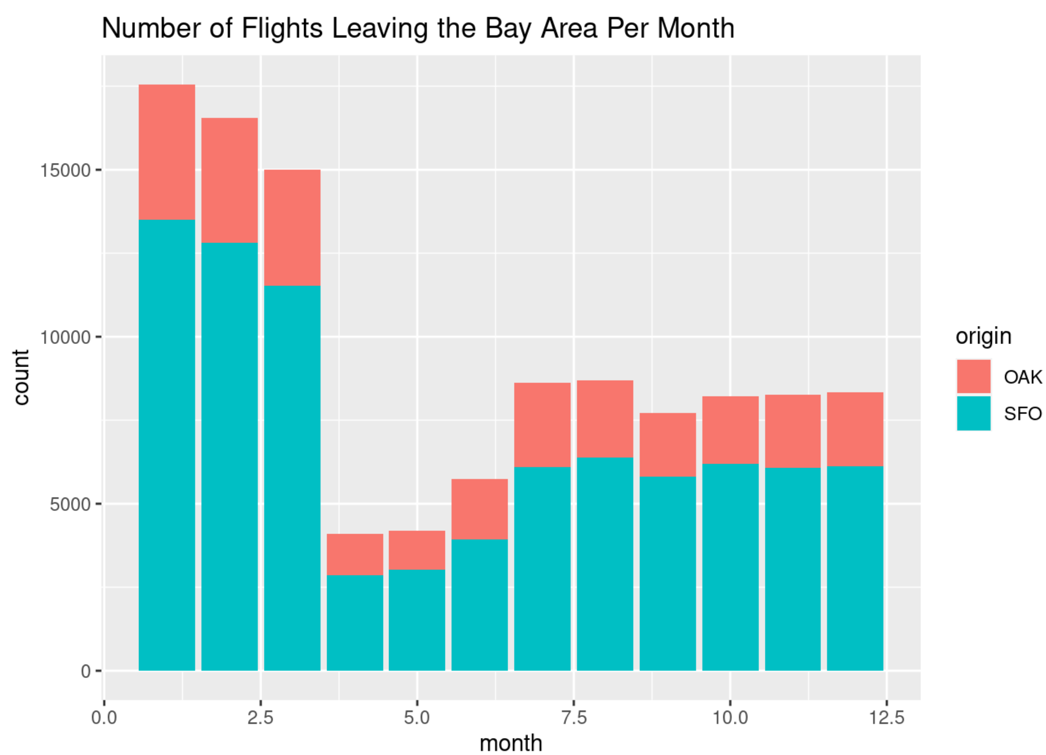
```
# A tibble: 2,854 × 20
    year month   day dep_time sched_dep_time dep_delay
arr_time sched_arr_time
   <dbl> <dbl> <dbl>    <dbl>          <dbl>     <dbl>
<dbl>          <dbl>
 1  2020     4     1       12          1950       262
138           2124
 2  2020     4     1       35            45       -10
837            912
 3  2020     4     1       36            47       -11
824            908
 4  2020     4     1      549           555        -6
901            925
 5  2020     4     1      549           600       -11
656            732
 6  2020     4     1      554           600        -6
752            802
 7  2020     4     1      555           600        -5
1103           1144
 8  2020     4     1      602           600         2
1318           1330
 9  2020     4     1      606           610        -4
```

```
733              725
10  2020     4     1      615              600         15
850              854
# ℹ 2,844 more rows
# ℹ 12 more variables: arr_delay <dbl>, carrier <chr>, flight
<dbl>,
#   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>,
distance <dbl>,
#   hour <dbl>, minute <dbl>, time_hour <dttm>, avg_speed
<dbl>
```

2,854 flights left SFO during March 2020

Question 9

```
ggplot(flights, aes(x = month, fill = origin)) +
  geom_bar() + labs(title = "Number of Flights Leaving the Bay
```



I do see the effect of the pandemic as there is a dramatic decrease in flights in April and May and then a slow rise as the COVID adjustment is happening.

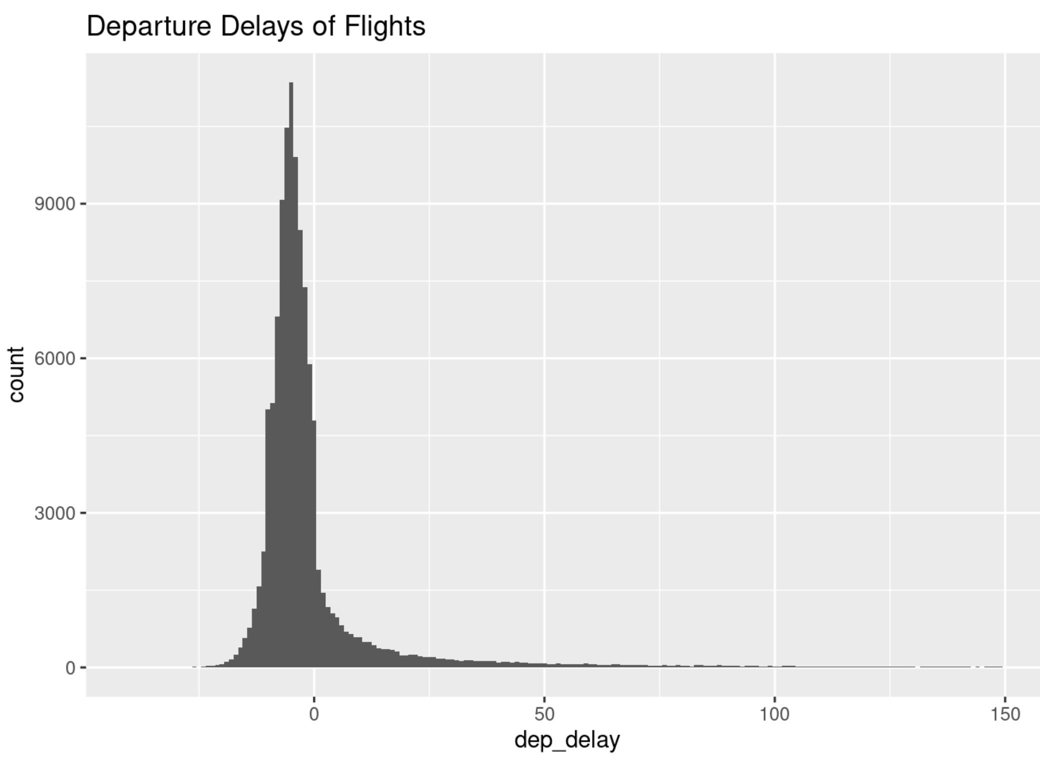
Question 10

Create a histogram showing the distribution of departure delays for all flights. Describe in words the shape and modality of the distribution and, using numerical summaries, (i.e. summary statistics) its center and spread. Be sure to use measures of center and spread that are most appropriate for this type of distribution. Also set the limits of the x-axis to focus on where most of the data lie.

```
ggplot(flights, aes(x = dep_delay)) +
geom_histogram(binwidth = 1) +
labs(title = "Departure Delays of Flights") + xlim(-40,150)
```

```
Warning: Removed 768 rows containing non-finite values
(`stat_bin()`).

Warning: Removed 2 rows containing missing values
(`geom_bar()`).
```




```
flights %>%
  summarize(median_delay = median(dep_delay), iqr_delay = IQF
```

```
# A tibble: 1 × 2
  median_delay iqr_delay
         <dbl>     <dbl>
1           -4         6
```

The shape of the distribution is unimodal and it's right skewed. This indicates that the majority of our departure delays are less than or right at 0. The median was the center of spread I used as it is the least affected by outliers. The median is -4 meaning that the most reoccuring time for departure delays is -4. The interquartile range of 6 tells us that half of our data lies within the range of 6 units which makes it pretty concentrated

Question 11

Add a new column to your data frame called before_times that takes values of TRUE and FALSE indicating whether the flight took place up through the end of March or after April 1st, respectively. Remake the histograms above, but now separated into two subplots: one with the departure delays from the before times, the other with the flights from afterwards. Can you visually detect any difference in the distribution of departure delays?

This is best done with a new layer called facet_wrap(). Learn about it's use by reading the documentation: https://ggplot2.tidyverse.org/reference/facet_wrap.html.

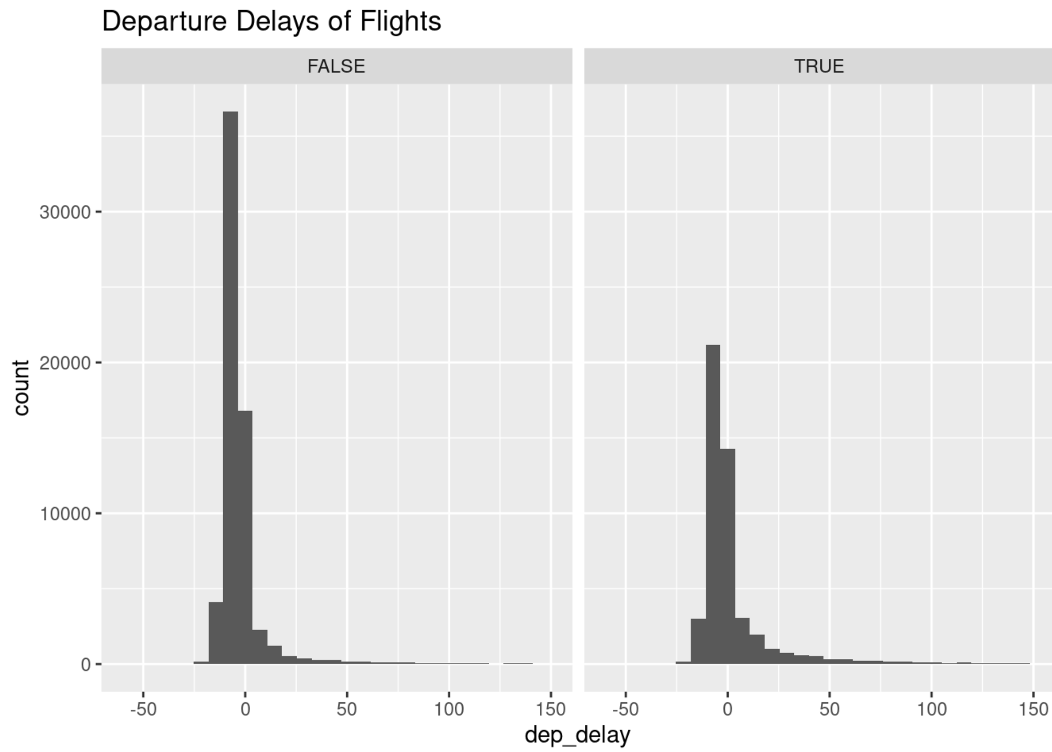
```
flights <- flights %>%
    mutate(before_times = month <= 3)

  p <- ggplot(flights, aes(x = dep_delay)) +
  geom_histogram()  +
  labs(title = "Departure Delays of Flights") + xlim(-60,150)
  p + facet_wrap(vars(before_times))
```

```
`stat_bin()` using `bins = 30`. Pick better value with
`binwidth`.

Warning: Removed 768 rows containing non-finite values
(`stat_bin()`).

Warning: Removed 4 rows containing missing values
(`geom_bar()`).
```



Yes I can detect differences in the the distribution of departure delays. The count of flights that occur after March 31st are higher than those occuring before. The general distribution stays the same but it appears that there were nearing 4000 flights after March while there were 2000 before. This is due to the time frame after march being a lot longer than three months measured in march.

Question 12

If you flew out of OAK or SFO during this time period, what is the tail number of the plane that you were on? If you did not fly in this period, find the tail number of the plane that flew JetBlue flight 40 to New York's JFK Airport from SFO on May 1st.

```
flights %>%
  filter(origin == 'SFO' & dest == 'JFK' & month == 5 & day ==
```

```
# A tibble: 1 × 21
   year month   day dep_time sched_dep_time dep_delay arr_time
sched_arr_time
  <dbl> <dbl> <dbl>    <dbl>          <dbl>     <dbl>    <dbl>
<dbl>
1  2020     5     1     1511           1520        -9     2304
2358
# ℹ 13 more variables: arr_delay <dbl>, carrier <chr>, flight
<dbl>,
```

```
#   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>,
distance <dbl>,
#   hour <dbl>, minute <dbl>, time_hour <dttm>, avg_speed
<dbl>,
#   before_times <lgl>
```

The tail number of the plane that flew JetBlue flight 40 to New York's JFK Airport from SFO on May 1st is N982JB.

Question 13

```
flights %>%
  group_by(carrier) %>%
    summarize(median_delay = median(dep_delay),IQR_Delay = IQR(
```

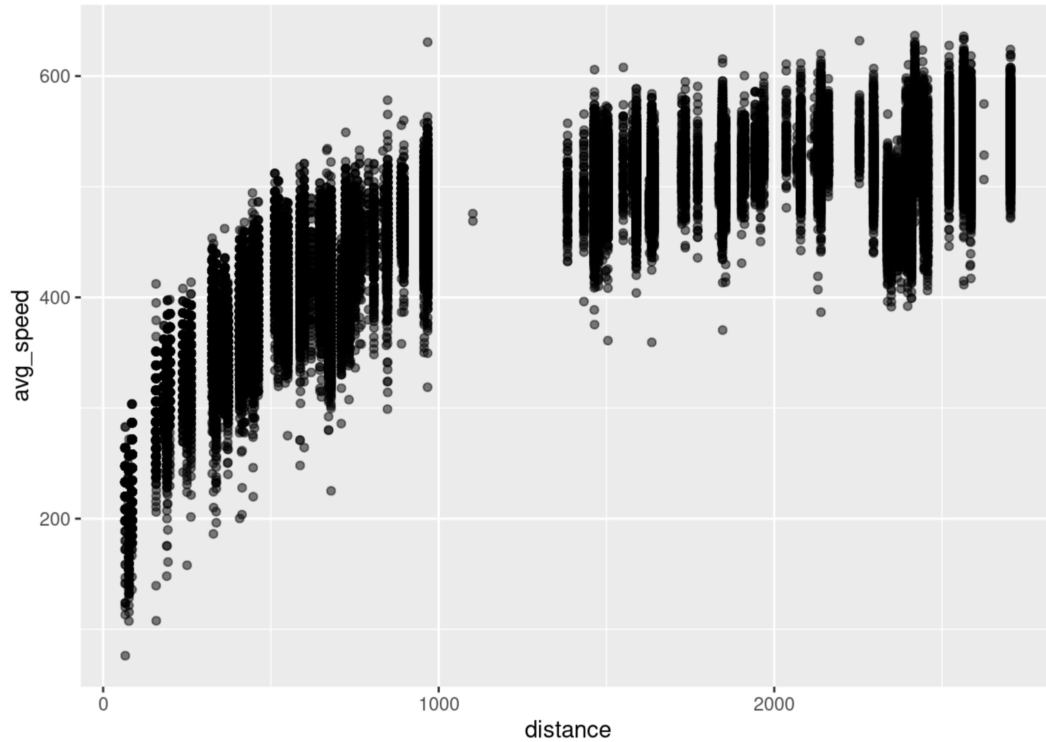
```
# A tibble: 12 × 3
   carrier median_delay IQR_Delay
   <chr>          <dbl>     <dbl>
 1 AA                -6         6
 2 AS                -7         9
 3 B6                -8         7
 4 DL                -5         5
 5 F9                -5         9
 6 G4                -7        14
 7 HA                -5         9
 8 NK                -4         6
 9 OO                -5         7
10 UA                -5         6
11 WN                -3         5
12 YV                -7        11
```

B6 has the lowest typical departure delay. DL has the least variable departure delays.

Question 14

Create a plot that captures the relationship of average speed vs. distance and describe the shape and structure that you see. What phenomena related to taking flights from the Bay Area might explain this structure?

```
flights <- flights %>% mutate(avg_speed = distance/(air_time/60
ggplot(flights, aes(x = distance, y = avg_speed)) +
    geom_jitter(alpha=0.5)
```



There is a slow exponential rise with it stabilizing at around 600 miles per hour. There's a blank gap in the middle of the graph which coincides with the estimated distance of the midwest areas from the Bay which typically recieve less flights. This means people are either travelling close by or going all the way to the east coast.

Question 15

```
flights %>%
  filter(origin == 'SFO') %>%
  group_by(month) %>%
  summarize(avg_delay = mean(dep_delay)) %>%
  arrange(desc(avg_delay))
```

```
# A tibble: 12 × 2
   month avg_delay
   <dbl>     <dbl>
 1     1     9.05
 2     2     8.04
 3     3     1.13
 4    10    -0.182
 5    12    -0.961
 6     7    -1.57
 7     9    -1.76
 8     6    -1.86
 9     8    -1.87
```

```
10     11    -1.91
11      5    -2.68
12      4    -3.79
```

```
flights %>%
  filter(origin == 'SFO') %>%
  group_by(month) %>%
  summarize(med_delay = median(dep_delay)) %>%
  arrange(desc(med_delay))
```

```
# A tibble: 12 × 2
   month med_delay
   <dbl>     <dbl>
 1     1        -3
 2     2        -3
 3     3        -5
 4     6        -5
 5     8        -5
 6     9        -5
 7    10        -5
 8    12        -5
 9     5        -6
10     7        -6
11    11        -6
12     4        -7
```

The month has the highest average departure delay is January. The highest median departure delay is also January.The highest median departure delay is more useful to know when deciding which month(s) to avoid flying if you particularly dislike flights that are severely delayed. This is due to the fact that an average can be easily swayed by outliers so an entire month can have no delays but a few very delayed flights and have a low average while a median isn't very swayed by outliers since it looks at what's in the middle of the data set.

Question 16

```
flights %>%
  group_by(tailnum) %>%
  summarize(max_distance = sum(distance)) %>%
  arrange(desc(max_distance))
```

```
# A tibble: 3,763 × 2
   tailnum max_distance
   <chr>          <dbl>
```

```
 1 N705TW        237912
 2 N969JT        236889
 3 N986JB        234235
 4 N983JT        233355
 5 N980JT        232910
 6 N984JB        228353
 7 N989JT        225875
 8 N968JT        223371
 9 N961JT        220298
10 N978JB        220023
# ℹ 3,753 more rows
```

Plane N705TW flew the max distance. Since the earth is 24,901 miles in circumference and the N705TW flew 237,912 miles then it flew 9 (9.55) times around the planet.

Question 17

```
flights %>%
  filter(avg_speed == max(avg_speed)) %>%
  group_by(tailnum) %>%
  arrange(desc(avg_speed))
```

```
# A tibble: 1 × 21
# Groups:   tailnum [1]
   year month   day dep_time sched_dep_time dep_delay arr_time
sched_arr_time
  <dbl> <dbl> <dbl>    <dbl>          <dbl>     <dbl>    <dbl>
<dbl>
1  2020     2    19     2302           2310        -8      616
657
# ℹ 13 more variables: arr_delay <dbl>, carrier <chr>, flight
<dbl>,
#   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>,
distance <dbl>,
#   hour <dbl>, minute <dbl>, time_hour <dttm>, avg_speed
<dbl>,
#   before_times <lgl>
```

I defined fastest as the plane with the highest average speed meaning it will consistently have the fastest speed. The plane that was the fastest is the N30913 which is the Boeing 787-8.

Question 18

```
flights <- flights %>%
mutate(day_of_week = wday(ymd(paste(year, month, day, set = "-'

flights %>%
  group_by(day_of_week) %>%
      filter(dest == 'SMF' & origin == 'SFO') %>%
        summarize(avg_delay= mean(dep_delay)) %>%
          arrange(avg_delay)
```

```
# A tibble: 7 × 2
  day_of_week avg_delay
  <ord>           <dbl>
1 Tue              1.99
2 Wed              2.19
3 Sun              5.12
4 Thu              5.38
5 Fri              8.77
6 Mon              9.48
7 Sat             14.5
```

```
flights %>%
     group_by(tailnum) %>%
      filter(dest == 'SMF' & origin == 'SFO') %>%
        summarize(avg_delay= mean(dep_delay)) %>%
          arrange(avg_delay)
```

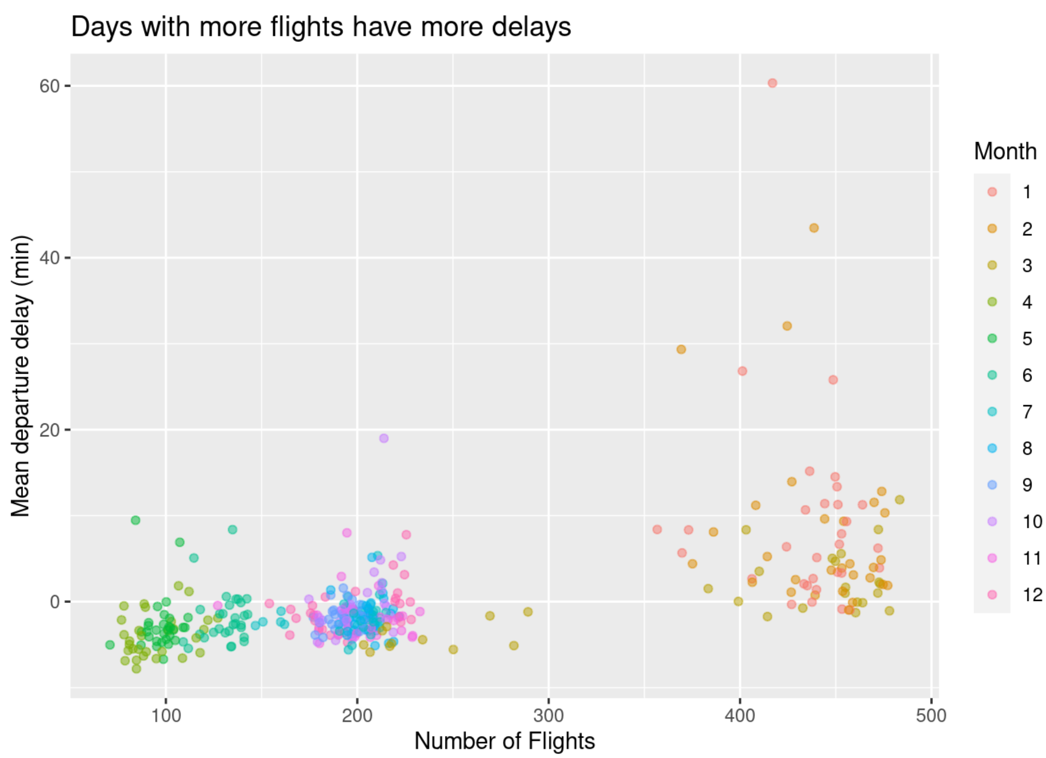
```
# A tibble: 230 × 2
   tailnum avg_delay
   <chr>       <dbl>
 1 N968JT      -16
 2 N977JE      -15
 3 N76517      -13
 4 N969JT      -12.2
 5 N432UA      -10
 6 N909EV      -10
 7 N963SW      -10
 8 N862AS       -9.5
 9 N451UA       -9
10 N860AS       -9
# ℹ 220 more rows
```

I am defining best as the one that has the smallest departure delay as that would be something I would want when travelling on an airline. The best day to travel is Tuesday as it has the lowest average departure delay. The best airline for travelling from Sacramento to San Fransisco is JetBlue as it has the lowet average departure delay.

Question 19

```
flights %>%
  filter(origin == 'SFO') %>%
  group_by(day, month, year) %>%
  summarize(n= n(), avg_delay = mean(dep_delay)) %>%
  ggplot(aes(x = n, y = avg_delay, color = factor(month))) +
  geom_jitter(shape = "circle", size = 1.5, alpha = 0.5) +
  labs(
    title = "Days with more flights have more delays",
    x = "Number of Flights",
    y = "Mean departure delay (min)",
    color = "Month"
  )
```

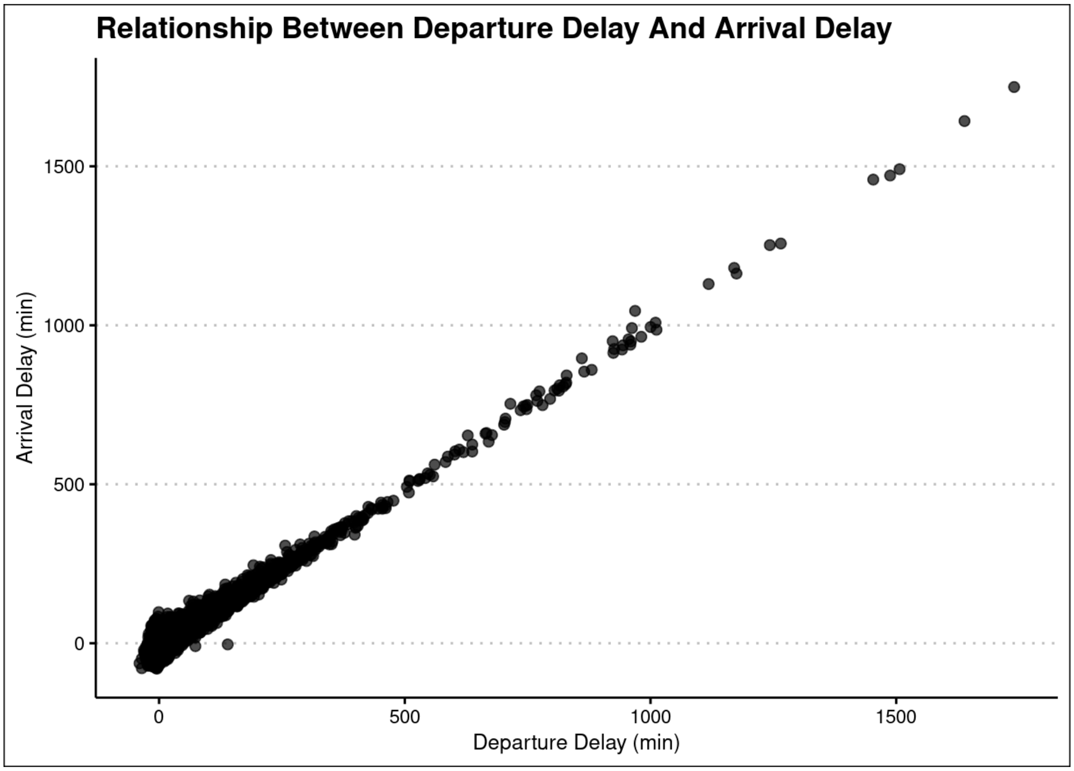
```
`summarise()` has grouped output by 'day', 'month'. You can
override using the
`.groups` argument.
```



Question 20

```
flights %>%
  ggplot(aes(x = dep_delay, y = arr_delay)) +
  geom_jitter(shape = "circle", size = 2, alpha = 0.7) + theme_
  labs(
    title = "Relationship Between Departure Delay And Arrival D
```

```
    x = "Departure Delay (min)",
    y = "Arrival Delay (min)"
  )
```




```
flights %>%
  mutate(y_hat = fitted(lm(arr_delay ~ dep_delay)),
         res = arr_delay - y_hat) %>%
  arrange(res)
```

```
# A tibble: 113,013 × 24
    year month   day dep_time sched_dep_time dep_delay
arr_time sched_arr_time
   <dbl> <dbl> <dbl>    <dbl>          <dbl>     <dbl>
<dbl>          <dbl>
 1  2020     2    17     1030            810       140
1142           1146
 2  2020     7     7     1944           1830        74
2042           2051
 3  2020     3    25     1011           1015        -4
1741           1901
 4  2020     2    13     1542           1545        -3
1824           1940
 5  2020     2    13     1554           1600        -6
1836           1955
 6  2020     3    10     1654           1655        -1
1838           1950
 7  2020     2    14      653            700        -7
```

```
943          1100
 8  2020     2    14     1603           1545        18
1850           1940
 9  2020     2    14      826            815        11
1108           1205
10  2020     2    14      800            800         0
1042           1150
# ℹ 113,003 more rows
# ℹ 16 more variables: arr_delay <dbl>, carrier <chr>, flight
<dbl>,
#   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>,
distance <dbl>,
#   hour <dbl>, minute <dbl>, time_hour <dttm>, avg_speed
<dbl>,
#   before_times <lgl>, day_of_week <ord>, y_hat <dbl>, res
<dbl>
```

The flight has the highest arrival delay given its departure delay is flight 889. The linear model predicted it's arrival delay would be 129.74 minutes but it was actually 133.74 minutes than the linear model prediction.

Question 21

Fit a multiple linear regression model that explains arrival delay using departure delay and the distance of the flight and print out the coefficients (the intercept and two slopes). Speculate as to why the sign (positive or negative) of the distance coefficient is what it is. Can we compare the coefficients for departure delay and distance to understand which has the stronger relationship? Why or why not?

```r
linear_model <- lm(arr_delay ~ dep_delay + distance, flights)
linear_model
```

```

Call:
lm(formula = arr_delay ~ dep_delay + distance, data = flights)

Coefficients:
(Intercept)    dep_delay     distance  
  -9.168277     0.999918    -0.001104  
```

The sign of the distance being negative means that there is a negative relationship between distance and arrival delay. This is probably due to certain airplanes being used for long distance flights that might be able to fly faster therefore being the cause to why arrival delay tends to decrease in relation to a higher distance. The coeeficiant of departure delay being 0.99 means there is a strong and positive relationship between departure and arrival delay. The coefficiant for distance tells us there’s a 0.0011 minute decrease for every mile so it has a small and negative relationship with departure delay.