

Assignment 4

Morgan McCullough

10/6/2021

```
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.0.5
##
## 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(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.0.5
## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.3      v purrr 0.3.4
## v tibble 3.1.1       v stringr 1.4.0
## v tidyr 1.1.3        v forcats 0.5.1
## v readr 2.0.2
## Warning: package 'ggplot2' was built under R version 4.0.5
## Warning: package 'tibble' was built under R version 4.0.5
## Warning: package 'tidyr' was built under R version 4.0.5
## Warning: package 'readr' was built under R version 4.0.5
## Warning: package 'purrr' was built under R version 4.0.5
## Warning: package 'forcats' was built under R version 4.0.5
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()

flights <- read.csv("C:/Users/morga/Desktop/assignment_4/flights.csv")
flightsDF <- data.frame(flights)
head(flightsDF)

##   year month day dep_time sched_dep_time dep_delay arr_time sched_arr_time
## 1 2013     1   1      517             515         2      830             819
## 2 2013     1   1      533             529         4      850             830
## 3 2013     1   1      542             540         2      923             850
```

```
## 4 2013      1      1      544      545      -1      1004      1022
## 5 2013      1      1      554      600      -6      812      837
## 6 2013      1      1      554      558      -4      740      728
##   arr_delay carrier flight tailnum origin dest air_time distance hour minute
## 1         11      UA   1545  N14228   EWR  IAH      227      1400     5     15
## 2         20      UA   1714  N24211   LGA  IAH      227      1416     5     29
## 3         33      AA   1141  N619AA   JFK  MIA      160      1089     5     40
## 4        -18      B6    725  N804JB   JFK  BQN      183      1576     5     45
## 5        -25      DL    461  N668DN   LGA  ATL      116       762     6      0
## 6         12      UA   1696  N39463   EWR  ORD      150       719     5     58
##           time_hour
## 1 1/1/2013 5:00
## 2 1/1/2013 5:00
## 3 1/1/2013 5:00
## 4 1/1/2013 5:00
## 5 1/1/2013 6:00
## 6 1/1/2013 5:00
```

```
airlinesDF <- data.frame(read.csv("C:/Users/morga/Desktop/assignment_4/airlines.csv"))
head(airlinesDF)
```

```
##   carrier      name
## 1      9E   Endeavor Air Inc.
## 2      AA American Airlines Inc.
## 3      AS   Alaska Airlines Inc.
## 4      B6   JetBlue Airways
## 5      DL   Delta Air Lines Inc.
## 6      EV ExpressJet Airlines Inc.
```

```
airportsDF <- data.frame(read.csv("C:/Users/morga/Desktop/assignment_4/airports.csv"))
head(airportsDF)
```

```
##   faa      name      lat      lon alt tz dst
## 1 04G   Lansdowne Airport 41.13047 -80.61958 1044 -5  A
## 2 06A   Moton Field Municipal Airport 32.46057 -85.68003 264 -6  A
## 3 06C   Schaumburg Regional 41.98934 -88.10124 801 -6  A
## 4 06N   Randall Airport 41.43191 -74.39156 523 -5  A
## 5 09J   Jekyll Island Airport 31.07447 -81.42778 11 -5  A
## 6 0A9   Elizabethton Municipal Airport 36.37122 -82.17342 1593 -5  A
##           tzone
## 1 America/New_York
## 2 America/Chicago
## 3 America/Chicago
## 4 America/New_York
## 5 America/New_York
## 6 America/New_York
```

```
planesDF <- data.frame(read.csv("C:/Users/morga/Desktop/assignment_4/planes.csv"))
head(planesDF)
```

```
##   tailnum year      type      manufacturer      model engines seats
## 1  N10156 2004 Fixed wing multi engine      EMBRAER EMB-145XR      2     55
## 2  N102UW 1998 Fixed wing multi engine      AIRBUS  INDUSTRIE  A320-214      2    182
## 3  N103US 1999 Fixed wing multi engine      AIRBUS  INDUSTRIE  A320-214      2    182
## 4  N104UW 1999 Fixed wing multi engine      AIRBUS  INDUSTRIE  A320-214      2    182
## 5  N10575 2002 Fixed wing multi engine      EMBRAER EMB-145LR      2     55
```

```
## 6 N105UW 1999 Fixed wing multi engine AIRBUS INDUSTRIE A320-214      2    182
##   speed      engine
## 1    NA Turbo-fan
## 2    NA Turbo-fan
## 3    NA Turbo-fan
## 4    NA Turbo-fan
## 5    NA Turbo-fan
## 6    NA Turbo-fan
```

```
weatherDF <- data.frame(read.csv("C:/Users/morga/Desktop/assignment_4/weather.csv"))
head(weatherDF)
```

```
##   origin year month day hour  temp  dewp humid wind_dir wind_speed wind_gust
## 1   EWR 2013     1   1     1 39.02 26.06 59.37     270   10.35702      NA
## 2   EWR 2013     1   1     2 39.02 26.96 61.63     250    8.05546      NA
## 3   EWR 2013     1   1     3 39.02 28.04 64.43     240   11.50780      NA
## 4   EWR 2013     1   1     4 39.92 28.04 62.21     250   12.65858      NA
## 5   EWR 2013     1   1     5 39.02 28.04 64.43     260   12.65858      NA
## 6   EWR 2013     1   1     6 37.94 28.04 67.21     240   11.50780      NA
##   precip pressure visib                time_hour
## 1      0    1012.0     10 2013-01-01T06:00:00Z
## 2      0    1012.3     10 2013-01-01T07:00:00Z
## 3      0    1012.5     10 2013-01-01T08:00:00Z
## 4      0    1012.2     10 2013-01-01T09:00:00Z
## 5      0    1011.9     10 2013-01-01T10:00:00Z
## 6      0    1012.4     10 2013-01-01T11:00:00Z
```

```
TAMPA <- data.frame(filter(flightsDF, dest == "TPA"))
TAMPA <- filter(TAMPA, month == "11")
TAMPA <- filter(TAMPA, day == "1")
TAMPA <- filter(TAMPA, year == "2013")
TAMPA <- filter(TAMPA, sched_arr_time >= 1200)
head(left_join(TAMPA, weatherDF, by = "day"))
```

```
##   year.x month.x day dep_time sched_dep_time dep_delay arr_time sched_arr_time
## 1   2013     11   1    1040          1029         11    1341          1323
## 2   2013     11   1    1040          1029         11    1341          1323
## 3   2013     11   1    1040          1029         11    1341          1323
## 4   2013     11   1    1040          1029         11    1341          1323
## 5   2013     11   1    1040          1029         11    1341          1323
## 6   2013     11   1    1040          1029         11    1341          1323
##   arr_delay carrier flight tailnum origin.x dest air_time distance hour.x
## 1        18      B6     925  N566JB   JFK  TPA      159     1005     10
## 2        18      B6     925  N566JB   JFK  TPA      159     1005     10
## 3        18      B6     925  N566JB   JFK  TPA      159     1005     10
## 4        18      B6     925  N566JB   JFK  TPA      159     1005     10
## 5        18      B6     925  N566JB   JFK  TPA      159     1005     10
## 6        18      B6     925  N566JB   JFK  TPA      159     1005     10
##   minute      time_hour.x origin.y year.y month.y hour.y  temp  dewp humid
## 1      29 11/1/2013 10:00      EWR  2013         1      1 39.02 26.06 59.37
## 2      29 11/1/2013 10:00      EWR  2013         1      2 39.02 26.96 61.63
## 3      29 11/1/2013 10:00      EWR  2013         1      3 39.02 28.04 64.43
## 4      29 11/1/2013 10:00      EWR  2013         1      4 39.92 28.04 62.21
## 5      29 11/1/2013 10:00      EWR  2013         1      5 39.02 28.04 64.43
## 6      29 11/1/2013 10:00      EWR  2013         1      6 37.94 28.04 67.21
```

```
##   wind_dir wind_speed wind_gust precip pressure visib      time_hour.y
## 1      270    10.35702        NA      0    1012.0    10 2013-01-01T06:00:00Z
## 2      250     8.05546        NA      0    1012.3    10 2013-01-01T07:00:00Z
## 3      240    11.50780        NA      0    1012.5    10 2013-01-01T08:00:00Z
## 4      250    12.65858        NA      0    1012.2    10 2013-01-01T09:00:00Z
## 5      260    12.65858        NA      0    1011.9    10 2013-01-01T10:00:00Z
## 6      240    11.50780        NA      0    1012.4    10 2013-01-01T11:00:00Z
```

```
library(maps)
```

```
## Warning: package 'maps' was built under R version 4.0.5
```

```
##
```

```
## Attaching package: 'maps'
```

```
## The following object is masked from 'package:purrr':
```

```
##
```

```
##      map
```

```
#2
```

```
# anti_join(flights,airports,by=c("dest"="faa")) returns all rows from flights
```

```
# that are not found in airports.
```

```
# anti_join(airports,flights,by=c("faa"="dest")) returns all rows from airports
```

```
# not found in flights.
```

```
filteredFlights <- data.frame(filter(flightsDF, flight >= 100))
```

```
head(filteredFlights)
```

```
##   year month day dep_time sched_dep_time dep_delay arr_time sched_arr_time
## 1 2013     1   1      517           515         2      830           819
## 2 2013     1   1      533           529         4      850           830
## 3 2013     1   1      542           540         2      923           850
## 4 2013     1   1      544           545        -1     1004          1022
## 5 2013     1   1      554           600        -6      812           837
## 6 2013     1   1      554           558        -4      740           728
##   arr_delay carrier flight tailnum origin dest air_time distance hour minute
## 1         11      UA   1545  N14228   EWR  IAH      227      1400     5      15
## 2         20      UA   1714  N24211   LGA  IAH      227      1416     5      29
## 3         33      AA   1141  N619AA   JFK  MIA      160      1089     5      40
## 4        -18      B6    725  N804JB   JFK  BQN      183      1576     5      45
## 5        -25      DL    461  N668DN   LGA  ATL      116       762     6       0
## 6         12      UA   1696  N39463   EWR  ORD      150       719     5      58
##           time_hour
## 1 1/1/2013 5:00
## 2 1/1/2013 5:00
## 3 1/1/2013 5:00
## 4 1/1/2013 5:00
## 5 1/1/2013 6:00
## 6 1/1/2013 5:00
```

```
# filtering negative delays because those are ahead of schedule, meaning not a
```

```
# concern of ours.
```

```
filteredFlights <- filter(filteredFlights, arr_delay > 0)
```

```
filteredFlights <- filter(filteredFlights, dep_delay > 0)
```

```
filteredFlights = select(filteredFlights, arr_delay, dep_delay, day)
```

```
delayedFlight <- data.frame(filter(filteredFlights, arr_delay > 20 | dep_delay > 20))
```

```
weatherDF <- filter(weatherDF, wind_speed > 15 )
weatherDelays <- data.frame(merge(weatherDF, delayedFlight))
```

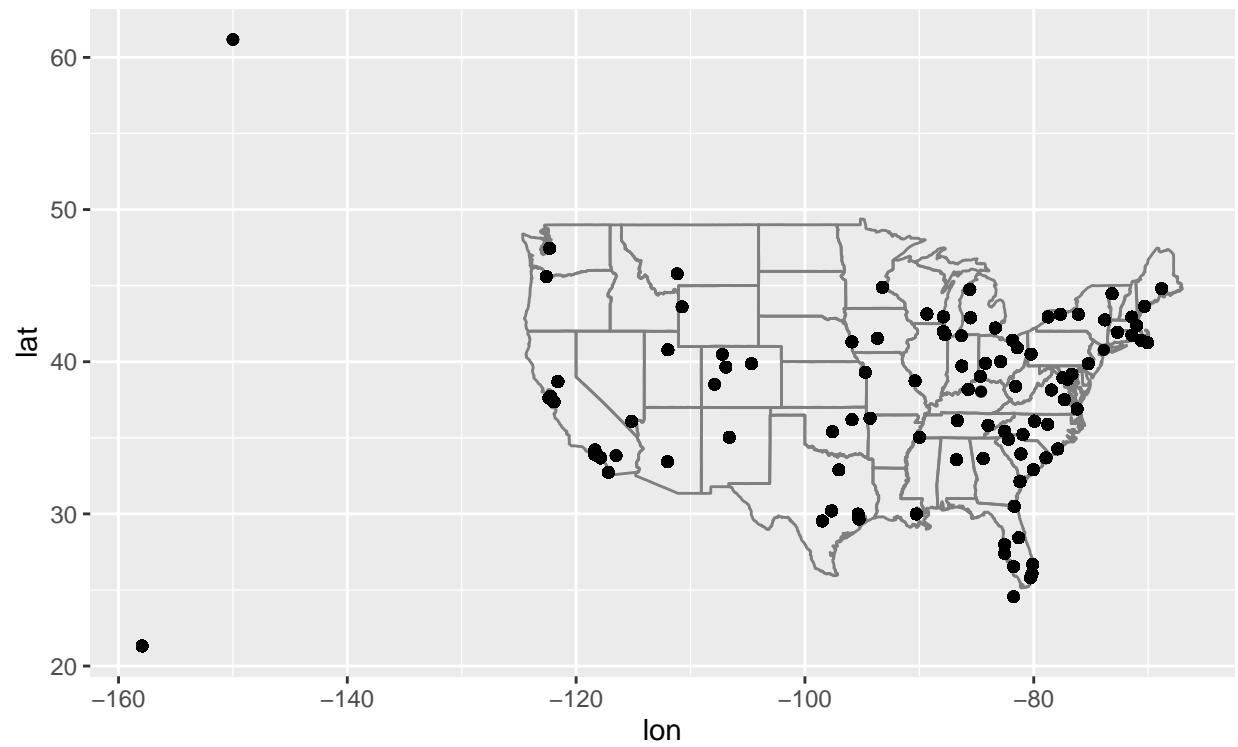
```
head(weatherDelays)
```

```
##   day origin year month hour temp dewp humid wind_dir wind_speed wind_gust
## 1    1     EWR 2013     1   13 39.2 28.4 69.67      330   16.11092      NA
## 2    1     EWR 2013     1   13 39.2 28.4 69.67      330   16.11092      NA
## 3    1     EWR 2013     1   13 39.2 28.4 69.67      330   16.11092      NA
## 4    1     EWR 2013     1   13 39.2 28.4 69.67      330   16.11092      NA
## 5    1     EWR 2013     1   13 39.2 28.4 69.67      330   16.11092      NA
## 6    1     EWR 2013     1   13 39.2 28.4 69.67      330   16.11092      NA
##   precip pressure visib            time_hour arr_delay dep_delay
## 1      0        NA    10 2013-01-01T18:00:00Z        33         2
## 2      0        NA    10 2013-01-01T18:00:00Z        32         8
## 3      0        NA    10 2013-01-01T18:00:00Z        12        24
## 4      0        NA    10 2013-01-01T18:00:00Z        44         2
## 5      0        NA    10 2013-01-01T18:00:00Z        21         2
## 6      0        NA    10 2013-01-01T18:00:00Z        30        47
```

*# From this, we can gather that the increase in wind speeds tend to cause
delays for both arriving and departing flights.*

```
flightsDF%>%left_join(airportsDF,c("dest"="faa"))%>%
  ggplot(aes(lon,lat))+
  borders("state")+
  geom_point()+
  coord_quickmap()
```

```
## Warning: Removed 7602 rows containing missing values (geom_point).
```



```
library(ggplot2)
library(maps)

covid <- data.frame(read.csv("C:/Users/morga/Desktop/assignment_4/covid19_vaccinations_USA.csv"))
names(covid)[1] <- 'State'
names(covid)[2] <- 'JJ'
names(covid)[3] <- 'Mod'
names(covid)[4] <- 'Pfiz'
```

{r} # wordcloud made in Wordle #