BSDS 100-Case Study 2

You are asked to submit both the R Markdown file and its pdf output.

Q1. Install the packages `nycflights13`, 'maps" and load the datasets:

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.3     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(maps)
```

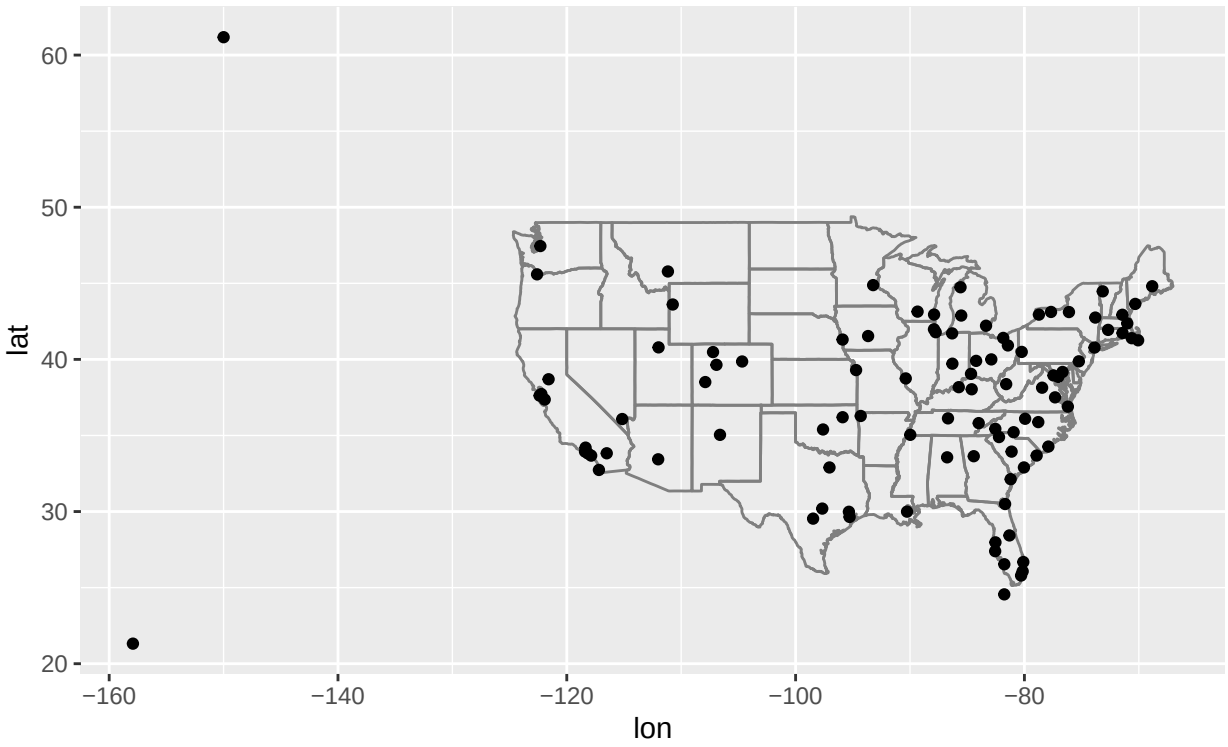
```
##
## Attaching package: 'maps'
```

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

```
library(nycflights13)
data(flights)
data(airlines)
data(airports)
data(planes)
data(weather)
```

Compute the average delay by destination, then join on the airports data frame so you can show the spatial distribution of delays. Here's an easy way to draw a map of the United States:

```
airports %>%
  semi_join(flights, c("faa" = "dest")) %>%
  ggplot(aes(lon, lat)) +
    borders("state") +
    geom_point() +
    coord_quickmap()
```



a) Use the size or color of the points to display the average delay for each airport.

b) Add the location of the origin and destination (i.e. the lat and lon) to flights.

c) Is there a relationship between the age of a plane and its delays?

d) What weather conditions make it more likely to see a delay?

e) What happened on June 13 2013? Display the spatial pattern of delays, and then use Google to cross-reference with the weather.

Q2. Using the `tweets.csv` data that is available on Canvas

(a) Identify all tweets with the word 'flight' in them.

(b) How many tweets end in a question mark?

(c) How many tweets have airport codes in them (assume any three subsequent capital letters are airport codes).

(d) Identify all tweets with URLs in them.

(e) Replace all instances of repeated exclamation points with a single exclamation point.

(f) Replace consecutive exclamation points, question marks, and periods with a single period, split the tweet on periods.