Lecture Assignment 15

Viraj Vijaywargiya

2022-05-20

```
library(nycflights13)
library(tidyverse)
```

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

```
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.6      v dplyr   1.0.8
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
```
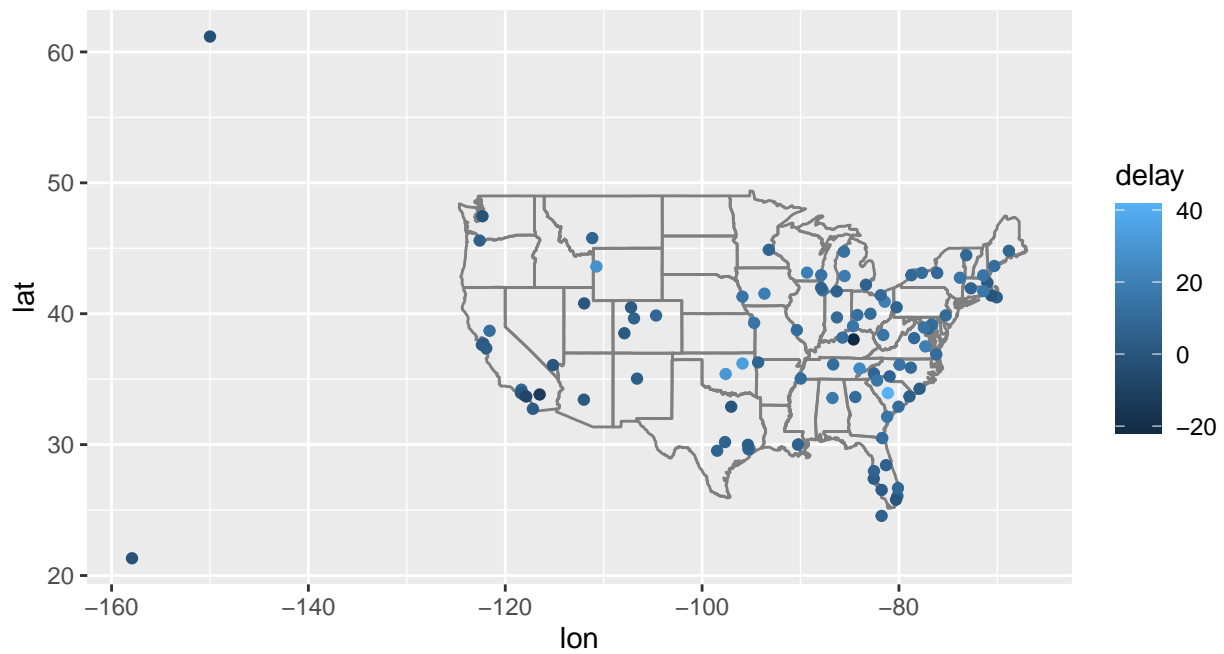
```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Part 13.4.6

Question 1

```r
avg_delays <-flights %>%
  group_by(dest) %>%
  summarise(delay = mean(arr_delay, na.rm = TRUE)) %>%
  inner_join(airports, by = c(dest = "faa"))

avg_delays %>%
  ggplot(aes(x = lon, y = lat, colour = delay)) +
  borders("state") +
  geom_point() +
  coord_quickmap()
```

Question 2

```
locations <- airports %>%
  select(faa, lat, lon)

flights %>%
  select(year:day, hour, origin, dest) %>%
  left_join(locations, by = c("origin" = "faa")) %>%
  left_join(locations, by = c("dest" = "faa"),
    suffix = c("Origin", "Dest"))
```

```
## # A tibble: 336,776 x 10
##     year month   day  hour origin dest  latOrigin lonOrigin latDest lonDest
##    <int> <int> <int> <dbl> <chr>  <chr>     <dbl>     <dbl>   <dbl>   <dbl>
## 1   2013     1     1     5 EWR    IAH        40.7     -74.2    30.0   -95.3
## 2   2013     1     1     5 LGA    IAH        40.8     -73.9    30.0   -95.3
## 3   2013     1     1     5 JFK    MIA        40.6     -73.8    25.8   -80.3
## 4   2013     1     1     5 JFK    BQN        40.6     -73.8      NA      NA
## 5   2013     1     1     6 LGA    ATL        40.8     -73.9    33.6   -84.4
## 6   2013     1     1     5 EWR    ORD        40.7     -74.2    42.0   -87.9
## 7   2013     1     1     6 EWR    FLL        40.7     -74.2    26.1   -80.2
## 8   2013     1     1     6 LGA    IAD        40.8     -73.9    38.9   -77.5
## 9   2013     1     1     6 JFK    MCO        40.6     -73.8    28.4   -81.3
## 10  2013     1     1     6 LGA    ORD        40.8     -73.9    42.0   -87.9
## # ... with 336,766 more rows
```
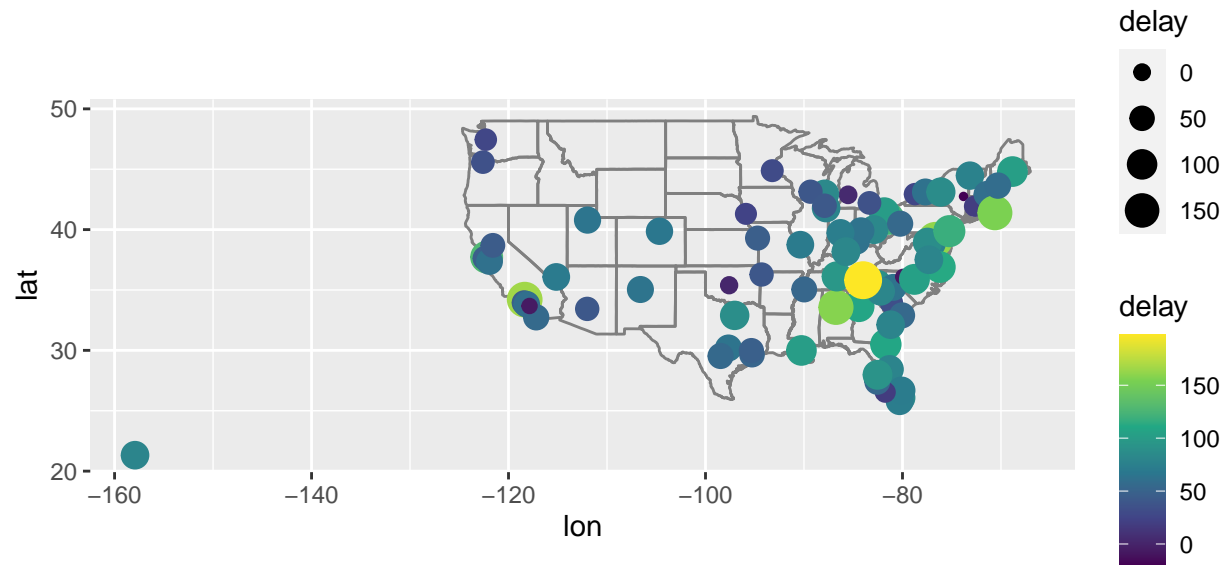
Question 5

From June 12 to June 13, there was a series of derechos, a widespread storm, in the southeastern and Midwest US.

```
flights %>%
  filter(year == 2013, month == 6, day == 13) %>%
  group_by(dest) %>%
  summarise(delay = mean(arr_delay, na.rm = TRUE)) %>%
  inner_join(airports, by = c("dest" = "faa")) %>%
  ggplot(aes(y = lat, x = lon, size = delay, colour = delay)) +
  borders("state") +
  geom_point() +
  coord_quickmap() +
  scale_colour_viridis_c()
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
```

3

As shown in the plot above, the largest delays were in the southeast and the Midwest, which was where the storm occurred across.

Part 13.5.1

Question 1

Flights having a missing tailnum means flight being cancelled, as they have msissing values of arr_time.

```
flights %>%
  anti_join(planes, by = "tailnum") %>%
  count(carrier, sort = TRUE)
```

```
## # A tibble: 10 x 2
##    carrier     n
##    <chr>   <int>
##  1 MQ      25397
##  2 AA      22558
##  3 UA       1693
##  4 9E       1044
##  5 B6        830
##  6 US        699
##  7 FL        187
##  8 DL        110
##  9 F9         50
## 10 WN         38
```

Tail numbers that don't have a matching record in planes are registered to American Airlines (AA) or Envoy Airlines (MQ).

Question 2

```
flights100 <- flights %>%
  filter(!is.na(tailnum)) %>%
  group_by(tailnum) %>%
  count() %>%
  filter(n >= 100)

flights %>%
  semi_join(flights100, by = "tailnum")
```

```
## # A tibble: 228,390 x 19
##     year month   day dep_time sched_dep_time dep_delay arr_time sched_arr_time
##    <int> <int> <int>    <int>          <int>     <dbl>    <int>          <int>
##  1  2013     1     1      517            515         2      830            819
##  2  2013     1     1      533            529         4      850            830
##  3  2013     1     1      544            545        -1     1004           1022
##  4  2013     1     1      554            558        -4      740            728
##  5  2013     1     1      555            600        -5      913            854
##  6  2013     1     1      557            600        -3      709            723
##  7  2013     1     1      557            600        -3      838            846
##  8  2013     1     1      558            600        -2      849            851
##  9  2013     1     1      558            600        -2      853            856
## 10  2013     1     1      558            600        -2      923            937
## # ... with 228,380 more rows, and 11 more variables: arr_delay <dbl>,
## #   carrier <chr>, flight <int>, tailnum <chr>, origin <chr>, dest <chr>,
## #   air_time <dbl>, distance <dbl>, hour <dbl>, minute <dbl>, time_hour <dttm>
```