Spatial mapping of flight delays

Abigail Castro, Pulkit Rampa, Ethan Thurmond, Eliza Goggins, Salem Keleta

Instructions: Split the tasks to be more efficient. All three parts can be done independently. Each of you should take responsibility for a part and finish it. The person whose name is first in the alphabetic order shares screen and merges all parts into the final R markdown version.

Load nycflights13

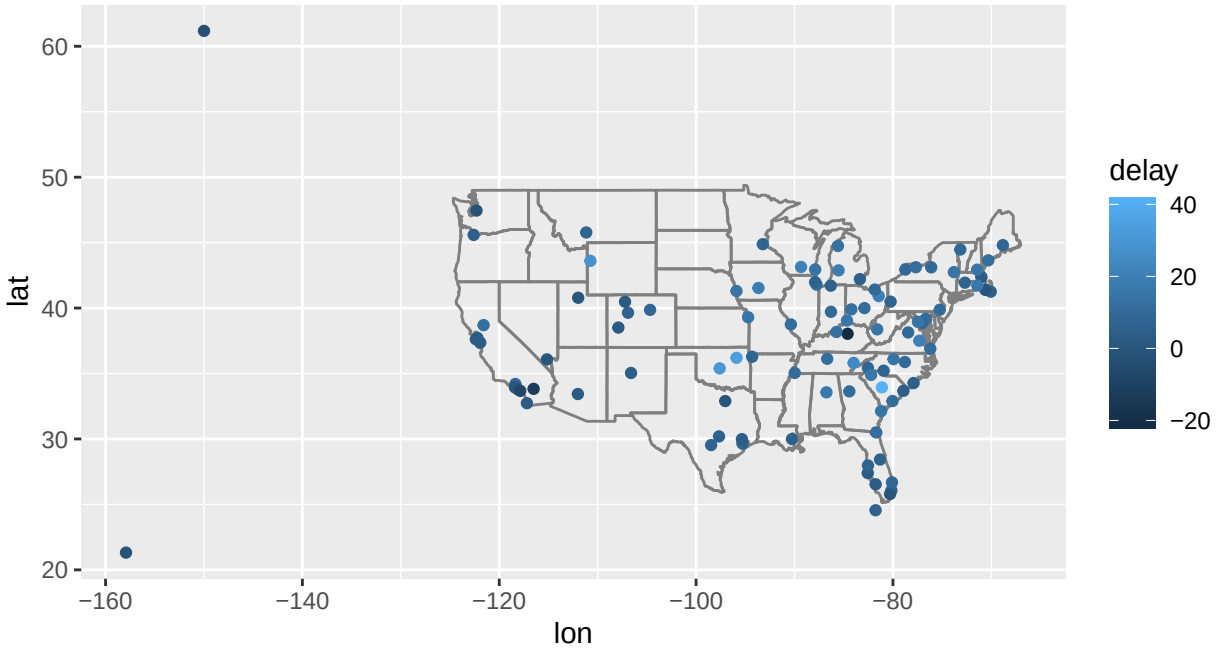
Spatial distribution of delays Abigail and Pulkit

1. Compute the average delay by destination, then join on the airports data frame so you can show the spatial distribution of delays. (See RDS 13.4.1 and 3.8.2)

```
avg_dest_delays <-
  flights %>%
  group_by(dest) %>%
  # arrival delay NA's are cancelled flights
  summarise(delay = mean(arr_delay, na.rm = TRUE)) %>%
  inner_join(airports, by = c(dest = "faa"))
```

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

```
avg_dest_delays %>%
  ggplot(aes(lon, lat, color = delay)) +
  borders("state") +
  geom_point() +
  coord_quickmap()
```



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

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

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

```
## # A tibble: 336,776 x 10
##     year month   day  hour origin dest  lat.x lon.x lat.y lon.y
##    <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
```

Relationship between the age of planes and delays Eliza and Ethan

1. Calculate the average delay of each plane over the whole year.

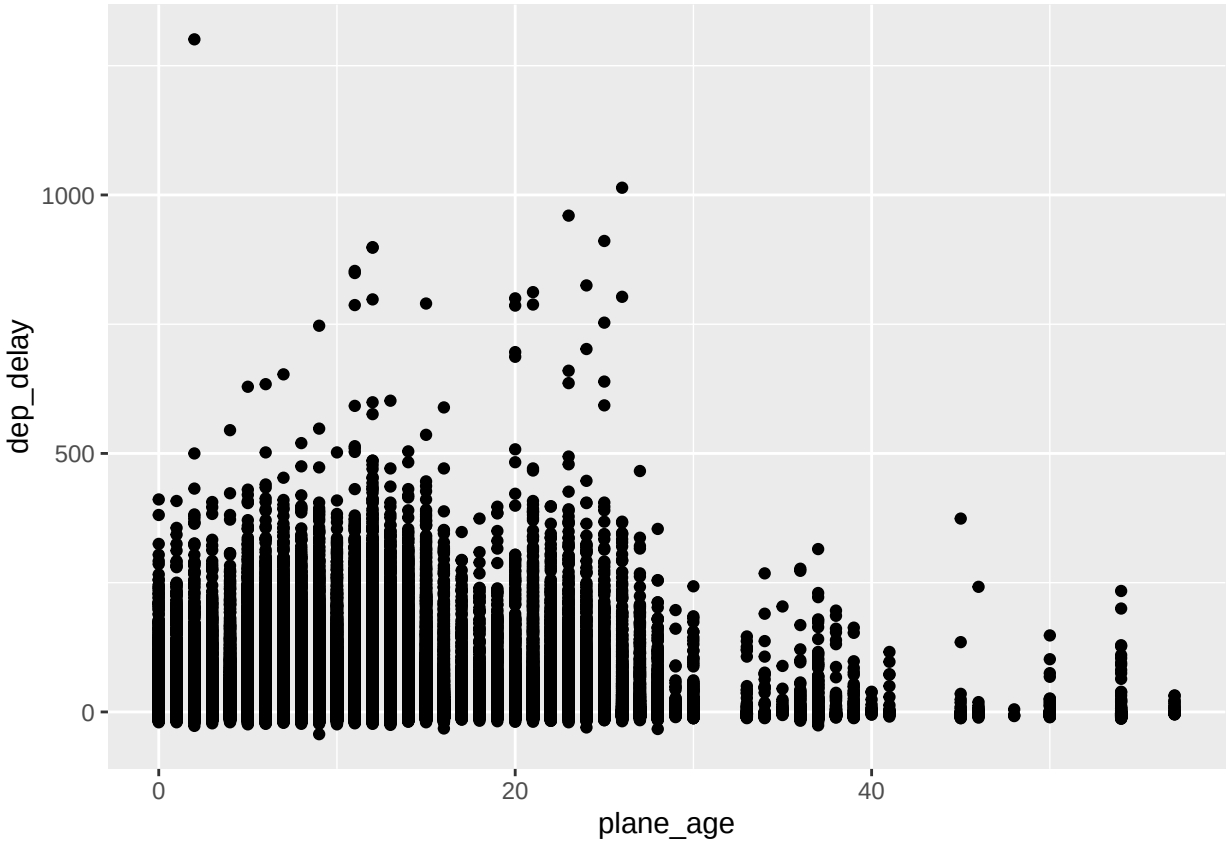
```
flights_and_planes <- flights %>% left_join(planes, by="tailnum") %>%
  rename(flight_year = year.x, manufacture_year = year.y)
flights_and_planes %>%
  group_by(tailnum) %>%
  filter(!is.na(dep_time)) %>%
  summarize(mean(dep_delay))
```

```
## # A tibble: 4,037 x 2
##    tailnum `mean(dep_delay)`
##    <chr>               <dbl>
##  1 D942DN             31.5
##  2 N0EGMQ              8.49
##  3 N10156             17.8
##  4 N102UW              8
##  5 N103US             -3.20
##  6 N104UW              9.94
##  7 N10575             22.7
##  8 N105UW              2.58
##  9 N107US             -0.463
## 10 N108UW              4.22
## # ... with 4,027 more rows
```

2. Plot the relationship between the age of a plane and its delays.

```
flights_and_planes %>% mutate(plane_age = flight_year - manufacture_year) %>%
  ggplot(aes(plane_age, dep_delay)) +
  geom_point()
```

```
## Warning: Removed 61980 rows containing missing values (geom_point).
```



3. Describe your observations. Is there a pattern? The trend pattern that we observed is that there tends to be a longer departure delay when dealing with newer planes and shorter departure delays dealing with older planes. However, there are more younger planes so that might skew the observations.

Relationship between weather and delays Salem

1. What weather conditions make it more likely to see a delay? Look at this source for inspiration: https://www.faa.gov/nextgen/programs/weather/faq/. Then identify the most relevant variables in the weather dataframe that relate to delays. Weather conditions: (during winter) airport surface winds, and low ceiling and visibility, peak during summer months (april-septermber) due to rain and thunderstorm

Relevant variables: wind speed, wind direction, visibility, precipitation, month

```
#weather
```

2. Join flights and weather. Calculate the average delay for each weather condition (e.g. low visibility, high wind speed, etc.). What conditions cause more delays? Describe your observations and patterns.

```
flight_weather <-
  flights %>%
  inner_join(weather, by = c(
    "origin" = "origin",
```

```
    "year" = "year",
    "month" = "month",
    "day" = "day",
    "hour" = "hour"
  ))
#flight_weather
```

Precipitation causes the most delays because almost any amount of precipitation results in some type of delay. Lower visibility is also results in higher delays.

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

```
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()
```

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