Course 2 Section 2.13 - Your turn

Jiaying Wu

13/10/2020

```r
# Load library
library(tidyverse)
library(nycflights13)
library(ggmap)
```

Delta flights

```r
# La Guardia airport
airports %>% filter(faa == "LGA")
```

```
## # A tibble: 1 x 8
##   faa   name          lat   lon   alt    tz dst   tzone
##   <chr> <chr>       <dbl> <dbl> <dbl> <dbl> <chr> <chr>
## 1 LGA   La Guardia  40.8 -73.9    22    -5 A     America/New_York
```

Confirming the point location

Confirm with a Google Map search that the geographical coordinates describing the location of La Guardia airport is in fact a point location in New York City.

Q1.Filter the flights data for for Delta Air Line flights from La Guardia in August

```r
# Filter Delta flights and create variable of La Guardia's location
delta <- flights %>%
  filter(carrier == "DL", origin == "LGA", month == 8) %>%
  left_join(airports, by = c("dest" = "faa"))
```

Q2.Create variables orig_lon and orig_lat to add La Guardia airport's location

```r
# Create variables to add the geographical coordinates of LGA's location
delta <- delta %>%
  mutate(orig_lon = filter(airports, faa == "LGA")$lon,
         orig_lat = filter(airports, faa == "LGA")$lat)
```

Q3.Download a map from the Google Maps Platform based on the following geographical coordinates -92.20562 (longitude) and 36.20259 (latitude)

```
# Geographical coordinates of a location of interest
location_US <- c(lon = -92.20562, lat = 36.20259)


# Enable Google services (need to obtain your Google Map API key first)
register_google(key = "Insert_your_API_key")


# Download map based on above coordinates (zoom of 4 gives us map of US)
map_US <- get_map(location_US, zoom = 4)
```

Q4.Overlay lines connected by the origin and destinations of Delta Air Line flights on this map

```
# Plot map
ggmap(map_US) +
  geom_segment(data = delta,
               aes(x = orig_lon, y = orig_lat, xend = lon, yend = lat),
               color = "navyblue",
               alpha = 0.2)
```
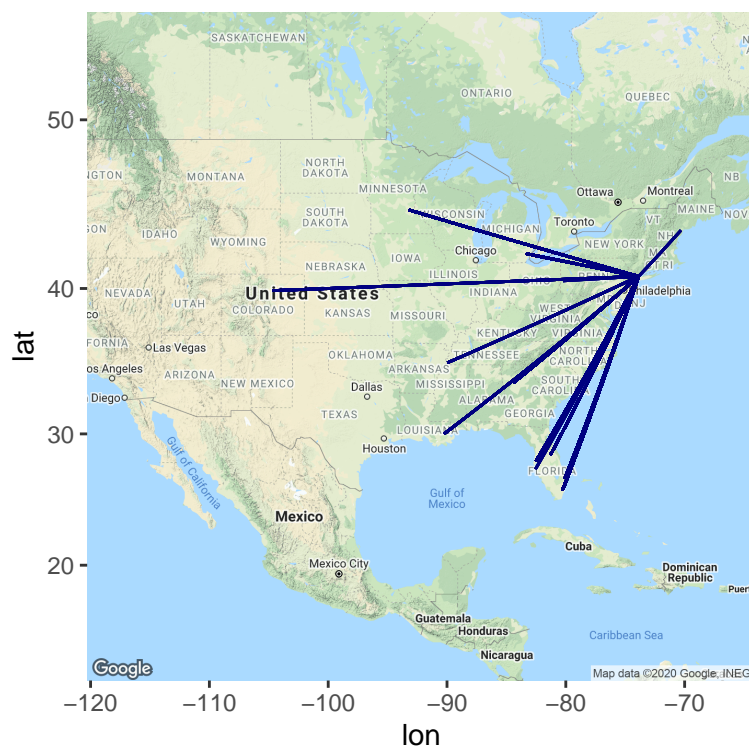

Airport delays

Using the weather and flights data set, perform an analysis to answer the question of whether the wind direction of strong winds affect airport operations.

Q5.join the weather data to the flights data

```
library(nycflights13)

flight_weather <- flights %>%
  left_join(weather, by = c("origin", "time_hour"))

flight_weather
```

```
## # A tibble: 336,776 x 32
##    year.x month.x day.x dep_time sched_dep_time dep_delay arr_time
##     <int>   <int> <int>    <int>          <int>     <dbl>    <int>
## 1    2013       1     1      517            515         2      830
## 2    2013       1     1      533            529         4      850
## 3    2013       1     1      542            540         2      923
## 4    2013       1     1      544            545        -1     1004
## 5    2013       1     1      554            600        -6      812
## 6    2013       1     1      554            558        -4      740
## 7    2013       1     1      555            600        -5      913
## 8    2013       1     1      557            600        -3      709
## 9    2013       1     1      557            600        -3      838
## 10   2013       1     1      558            600        -2      753
## # ... with 336,766 more rows, and 25 more variables: sched_arr_time <int>,
## #   arr_delay <dbl>, carrier <chr>, flight <int>, tailnum <chr>, origin <chr>,
## #   dest <chr>, air_time <dbl>, distance <dbl>, hour.x <dbl>, minute <dbl>,
## #   time_hour <dttm>, year.y <int>, month.y <int>, day.y <int>, hour.y <int>,
## #   temp <dbl>, dewp <dbl>, humid <dbl>, wind_dir <dbl>, wind_speed <dbl>,
## #   wind_gust <dbl>, precip <dbl>, pressure <dbl>, visib <dbl>
```

Q6.filter by La Guardia airport and high wind speeds (greater than 25 m/s)

```
flight_weather <- flight_weather %>%
  filter(origin == "LGA", wind_speed > 25)

flight_weather
```

```
## # A tibble: 1,611 x 32
##    year.x month.x day.x dep_time sched_dep_time dep_delay arr_time
##     <int>   <int> <int>    <int>          <int>     <dbl>    <int>
## 1    2013       1    20     1253           1300        -7     1438
## 2    2013       1    20     1254           1300        -6     1438
## 3    2013       1    20     1304           1305        -1     1452
## 4    2013       1    20     1304           1308        -4     1522
## 5    2013       1    20     1306           1315        -9     1504
## 6    2013       1    20     1314           1322        -8     1623
## 7    2013       1    20     1317           1325        -8     1454
## 8    2013       1    20     1327           1330        -3     1620
## 9    2013       1    20     1335           1335         0     1622
## 10   2013       1    20     1337           1350       -13     1546
## # ... with 1,601 more rows, and 25 more variables: sched_arr_time <int>,
## #   arr_delay <dbl>, carrier <chr>, flight <int>, tailnum <chr>, origin <chr>,
## #   dest <chr>, air_time <dbl>, distance <dbl>, hour.x <dbl>, minute <dbl>,
```

```
## #   time_hour <dttm>, year.y <int>, month.y <int>, day.y <int>, hour.y <int>,
## #   temp <dbl>, dewp <dbl>, humid <dbl>, wind_dir <dbl>, wind_speed <dbl>,
## #   wind_gust <dbl>, precip <dbl>, pressure <dbl>, visib <dbl>
```

Q7.plot departure delay against wind direction, perhaps focusing a restricted range of delay or using only a smoother instead of all the points.

```
ggplot(flight_weather, aes(x = wind_dir, y = dep_delay)) +
  geom_point(alpha = 0.2) +
  geom_smooth(se = FALSE) +
  labs(title = "Departure delay versus Wind direction at La Guardia airport", x = "Wind Direction", y =
```

```
## Warning: Removed 169 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 169 rows containing missing values (geom_point).
```


Departure delay versus Wind direction at La Guardia airport