Lab 04 - Visualizing spatial data

Status200OK

10/1/20

Load packages and data

```r
library(tidyverse)
library(dsbox)
```

```r
states <- read_csv("data/states.csv")
```

Exercise 1

There are 1643 rows and 6 columns in the `dennys` dataset.

Each row in the dataset represents a location of Denny's restaurants.The variables in the dataset are: address, city, state, zip, longitude, latitude.

Exercise 2

There are 909 rows and 6 columns in the `laquinta` dataset.

Each row in the dataset represents a location of La Quinta restaurants.The variables in the dataset are: address, city, state, zip, longitude, latitude.

Exercise 3

There are no Denny's restaurants outside the United States.

Not all La Quinta hotels are in the United States. There are La Quinta hotels in Canada, Mexico, Honduras, Turkey, Chile, and Colombia as well.

Exercise 4

To see if there is an establishment outside of the US, we could find out the range of the latitude and longitude of the US from the internet, and then check in the dataset if there is a longitude or latitude that falls outside of that range.

Another method could be checking if the state of the hotel is a state that is not in the US. The list of the states' abbreviation of the US can be found from the `states` dataset and with the use of this, states with the abbreviation that is not in the list can be filtered and stored.

Exercise 5

Now we will check for the locations of Denny's that are outside the US by filtering the state variable.

```
dennys %>%
  filter(!(state %in% states$abbreviation))
```

```
## # A tibble: 0 x 6
## # ... with 6 variables: address <chr>, city <chr>, state <chr>, zip <chr>,
## #   longitude <dbl>, latitude <dbl>
```

As seen in the tibble, there are 0 rows after we applied the filter, which means there are no Denny's outside the US.

Exercise 6

We will now add a new variable, or column, for the country of the Denny's locations, all of which are the United States.

```
dennys <- dennys %>%
  mutate(country = "United States")
```

Exercise 7

```
temp_laquinta <- laquinta %>%
  filter(!(state %in% states$abbreviation))
```

The locations of La Quinta hotels that are outside the US are shown below:

```
temp_laquinta
```

```
## # A tibble: 14 x 6
##    address                    city               state zip    longitude latitude
##    <chr>                      <chr>              <chr> <chr>      <dbl>    <dbl>
##  1 Carretera Panamericana Sur~ "\nAguascalientes" AG    20345      -102.     21.8
##  2 Av. Tulum Mza. 14 S.M. 4 L~ "\nCancun"         QR    77500      -86.8     21.2
##  3 Ejercito Nacional 8211     "Col\nPartido Igl~ CH    32528      -106.     31.7
##  4 Blvd. Aeropuerto 4001      "Parque Industria~ NL    66600      -100.     25.8
##  5 Carrera 38 # 26-13 Avenida~ "\nMedellin Colom~ ANT   0500~      -75.6      6.22
##  6 AV. PINO SUAREZ No. 1001   "Col. Centro\nMon~ NL    64000      -100.     25.7
##  7 Av. Fidel Velazquez #3000 ~ "\nMonterrey"      NL    64190      -100.     25.7
##  8 63 King Street East        "\nOshawa"         ON    L1H1~      -78.9     43.9
##  9 Calle Las Torres-1 Colonia~ "\nPoza Rica"      VE    93210      -97.4     20.6
## 10 Blvd. Audi N. 3 Ciudad Mod~ "\nSan Jose Chiap~ PU    75010      -97.8     19.2
## 11 Ave. Zeta del Cochero No 4~ "Col. ReservaTerr~ PU    72810      -98.2     19.0
## 12 Av. Benito Juarez 1230 B (~ "\nSan Luis Potos~ SL    78399      -101.     22.1
## 13 Blvd. Fuerza Armadas       "contiguo Mall La~ FM    11101      -87.2     14.1
## 14 8640 Alexandra Rd          "\nRichmond"       BC    V6X1~      -123.     49.2
```

There are 14 La Quinta hotels that are outside the US.
The states ON and BC are in Canada, ANT in Colombia, FM in Honduras, and AG, QR, CH, NL, UE, PU and SL are in Mexico.

Exercise 8

Now, adding countries' name in the `laquinta` dataset.

```r
laquinta <- laquinta %>%
  mutate(country = case_when(
    state %in% state.abb    ~ "United States",
    state %in% c("ON", "BC") ~ "Canada",
    state == "ANT"          ~ "Colombia",
    state == "FM"           ~ "Honduras",
    state %in% c("AG", "QR",
                 "CH", "NL",
                 "UE", "PU",
                 "SL")      ~ "Mexico"
  ))
```

Exercise 9

Now, filtering out only those La Quinta hotels that are in the US.

```r
laquinta <- laquinta %>%
  filter(country == "United States")
```

Now, lets find out which state has the most and least number of Denny's.

```r
dennys %>%
  count(state) %>%
  filter(n == max(n) | n == min(n)) %>%
  arrange(desc(n))
```

```
## # A tibble: 2 x 2
##   state     n
##   <chr> <int>
## 1 CA      403
## 2 DE        1
```

The tibble above shows that California has the largest number of Denny's and Delaware has the least number of Denny's.

Similarly, lets find out which state has the most and least number of La Quinta.

```r
laquinta %>%
  count(state) %>%
  filter(n == max(n) | n == min(n)) %>%
  arrange(desc(n))
```

```
## # A tibble: 2 x 2
##   state     n
##   <chr> <int>
## 1 TX      237
## 2 ME        1
```

The tibble above shows that Texas has the largest number of La Quinta and Maine has the least number of La Quinta.

The result is quite surprising because we, backed by the joke made famous by Mitch Hedberg, were expecting that the states which have largest and least number of the two establishments would be the same.

Exercise 10

Now, lets find out where are Denny's most densely located.

```
dennys %>%
  count(state ) %>%
  inner_join(states, by = c("state" = "abbreviation")) %>%
  mutate (persqmile= n/area) %>%
  arrange(desc(persqmile)) %>%
  top_n(1)
```

```
## Selecting by persqmile
```

```
## # A tibble: 1 x 5
##   state     n name                   area persqmile
##   <chr> <int> <chr>                 <dbl>     <dbl>
## 1 DC        2 District of Columbia   68.3    0.0293
```

The tibble above shows that District of Columbia has the most Denny's per square mile.

Similarly, lets find out where are La Quinta most densely located.

```
laquinta %>%
  count(state ) %>%
  inner_join(states, by = c("state" = "abbreviation")) %>%
  mutate (persqmile= n/area) %>%
  arrange(desc(persqmile)) %>%
  top_n(1)
```

```
## Selecting by persqmile
```

```
## # A tibble: 1 x 5
##   state     n name           area persqmile
##   <chr> <int> <chr>         <dbl>     <dbl>
## 1 RI        2 Rhode Island 1545.   0.00129
```

The tibble above shows that Rhode Island has the most La Quinta per square mile.

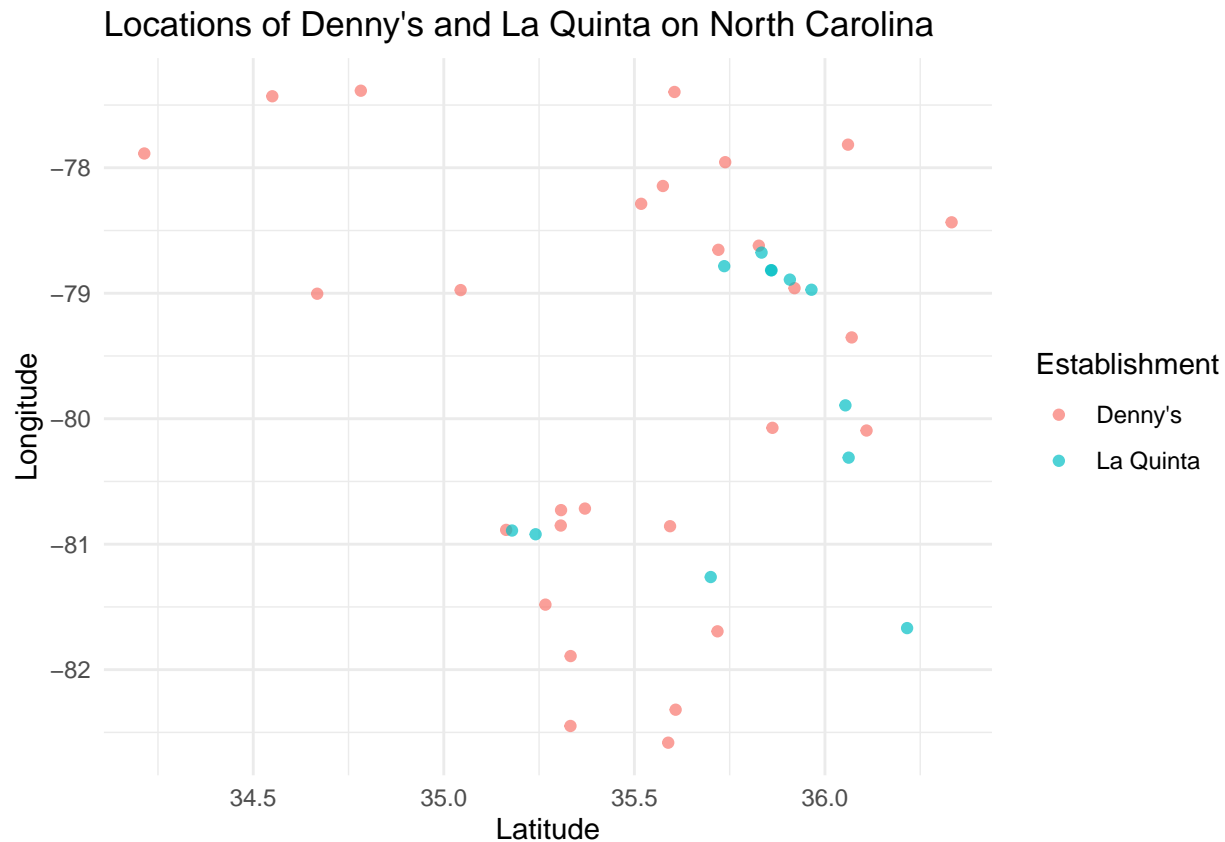**Exercise 11**

Now, merging the two datasets.

```
dennys <- dennys %>%
  mutate(establishment = "Denny's")
laquinta <- laquinta %>%
  mutate(establishment = "La Quinta")
dn_lq <- bind_rows(dennys, laquinta)
```

After this, lets visualize the spread of the two establishments within North Carolina.

```
dn_lq %>%
  filter( state == "NC") %>%
  ggplot(mapping = aes(x = longitude, y = latitude, color = establishment)) +
    geom_point(alpha= 0.7) +
    labs(title= "Locations of Denny's and La Quinta on North Carolina",
      x= "Longitude",
      y= "Latitude",
      color="Establishment") +
```

```
  coord_flip() +
  theme_minimal()
```

## Locations of Denny's and La Quinta on North Carolina

While there are few locations where Denny's and La Quinta seem to be together, in majority of the locations Denny's are found without any neighboring La Quintas. Hence the joke by Mitch Hedberg doesnt appear to hold here.

Exercise 12

Finally, lets visualize the spread of the two establishments within Texas.

```
dn_lq %>%
  filter( state == "TX") %>%
  ggplot(mapping = aes(x = longitude, y = latitude, color = establishment)) +
    geom_point(alpha= 0.5) +
    labs(title= "Locations of Denny's and La Quinta on Texas",
      x= "Longitude",
      y= "Latitude",
      color="Establishment") +
    coord_flip() +
    theme_minimal()
```

Locations of Denny's and La Quinta on Texas
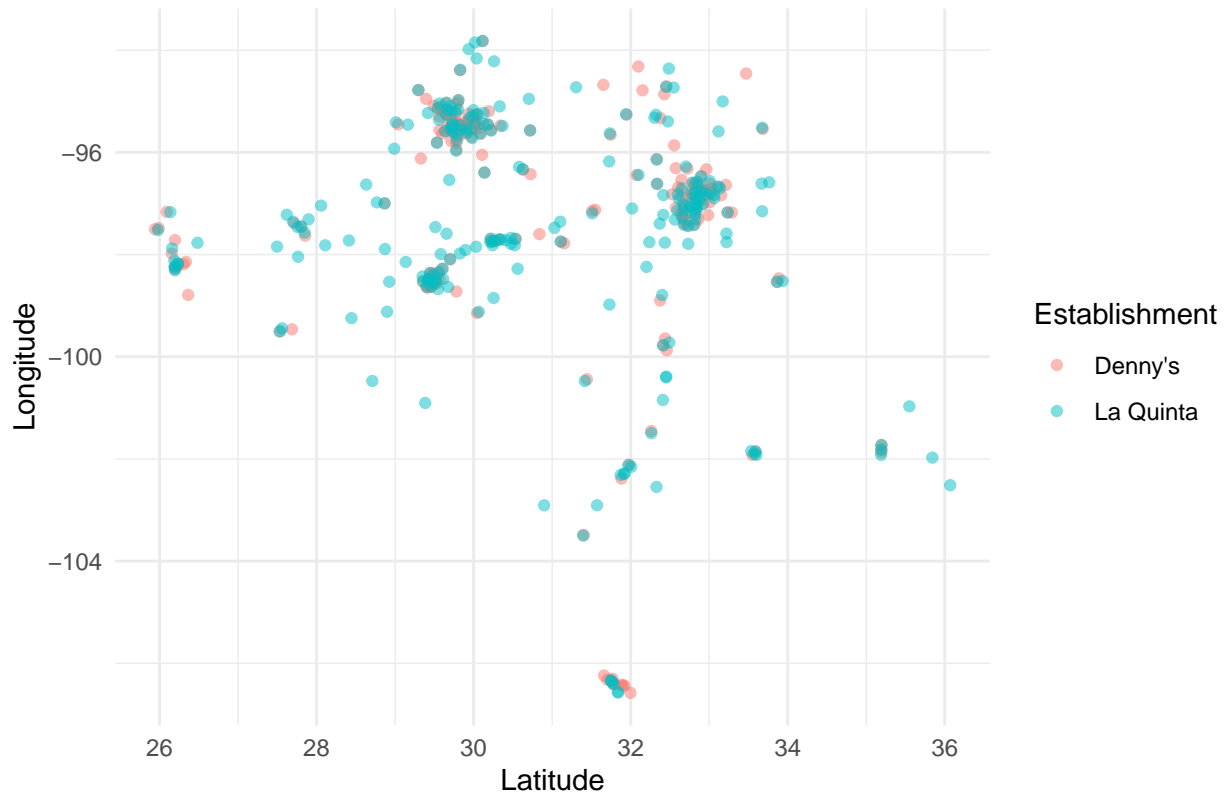
In almost all the locations, Denny's and La Quinta appear to be clustered together. Hence, the joke by Mitch Hedberg appear to hold here.