Lab 6.1 - Visualizing Spatial Data

Adham Rishmawi

Spring 2022

The late comedian Mitch Hedberg famously joked that "La Quinta" is Spanish for "next to Denny's". In this lab exercise, we explore the accuracy of his claim that Denny's, a casual diner chain that is open 24 hours, is actually co-located with La Quinta Inn and Suites, a hotel chain. This lab is inspired by a blog post by John Reiser entitled Mitch Hedberg and GIS.

Loading the Dataset

```
library(tidyverse)
```

We use three datasets, all stored in ./data:

- `dennys` — The locations of all Denny's restaurants:
  - *Scraped* from All Denny's Locations
  - Described here: Locations of Denny's restaurants
- `laquinta` — The locations of all La Quinta hotels:
  - *Scraped* from La Quinta Locations
  - Described here: Locations of La Quinta Inn motels
- `states` — The names of all the US states, including DC, plus some additional information.

Note that the first two datasets are represented in RDS format, which is an R-internal format for datasets that maintains data types. The last dataset is in CSV format and requires some

```
dennys <- read_rds("data/dennys.rds")
laquinta <- read_rds("data/laquinta.rds")
states <- read_csv("data/states.csv", col_types = cols(
  name = col_character(),
  abbreviation = col_character(),
  area = col_double()
))
```

The dataset descriptions on the RStudio-Education site are rather uninformative. So describe them here, including a data dictionary and their size.

```
obs <- nrow(dennys)
col <- ncol(dennys)
obs1 <- nrow(laquinta)
col1 <- ncol(laquinta)
obs2 <- nrow(states)
col2 <- ncol(states)
```

there is 1643 observations and 6 variables in Dennys data set/ there is 909 observations and 6 variables in laquinta data set/ there is 51 observations and 3 variables in states data set

-dennys and laqunita datasets have similar and understandable variables (address, city, state, zip, longitude, and latitude) -States are simliar too and have variables likes (Name, Abbreviation, and Area)

Focusing on US Locations

We limit our analysis to Denny"s and La Quinta locations in the United States.

Are there any Denny's or La Quinta locations that are outside the US? Answer this by filtering for observations where `state` is not in `states$abbreviation`. Here, the `%in%` operator matches the states listed in the `state` variable to those listed in the `states$abbreviation` vector. The `!` operator means **not**:

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

laquinta
```

```
## # A tibble: 895 x 6
##    address                          city           state zip   longitude latitude
##    <chr>                            <chr>          <chr> <chr>     <dbl>    <dbl>
##  1 793 W. Bel Air Avenue            "\nAberdeen"   MD    21001     -76.2     39.5
##  2 3018 CatClaw Dr                  "\nAbilene"    TX    79606     -99.8     32.4
##  3 3501 West Lake Rd                "\nAbilene"    TX    79601     -99.7     32.5
##  4 184 North Point Way              "\nAcworth"    GA    30102     -84.7     34.1
##  5 2828 East Arlington Street       "\nAda"        OK    74820     -96.6     34.8
##  6 14925 Landmark Blvd              "\nAddison"    TX    75254     -96.8     33.0
##  7 909 East Frontage Rd             "\nAlamo"      TX    78516     -98.1     26.2
##  8 2116 Yale Blvd Southeast         "\nAlbuquerqu~ NM    87106    -107.      35.1
##  9 7439 Pan American Fwy Northeast  "\nAlbuquerqu~ NM    87109    -107.      35.2
## 10 2011 Menaul Blvd Northeast       "\nAlbuquerqu~ NM    87107    -107.      35.1
## # ... with 885 more rows
```

Between laquinta and states, we see discrepancies of laquintas and their locations in Latin American countries like Colombia, Pru, and Puerto Rico!

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

```
## # A tibble: 1,643 x 6
##    address                          city           state zip   longitude latitude
##    <chr>                            <chr>          <chr> <chr>     <dbl>    <dbl>
##  1 2900 Denali                      Anchorage      AK    99503    -150.      61.2
##  2 3850 Debarr Road                 Anchorage      AK    99508    -150.      61.2
##  3 1929 Airport Way                 Fairbanks      AK    99701    -148.      64.8
##  4 230 Connector Dr                 Auburn         AL    36849     -85.5     32.6
##  5 224 Daniel Payne Drive N         Birmingham     AL    35207     -86.8     33.6
##  6 900 16th St S, Commons on Gree   Birmingham     AL    35294     -86.8     33.5
##  7 5931 Alabama Highway, #157       Cullman        AL    35056     -86.9     34.2
##  8 2190 Ross Clark Circle           Dothan         AL    36301     -85.4     31.2
##  9 900 Tyson Rd                     Hope Hull (Tys~ AL    36043     -86.4     32.2
## 10 4874 University Drive            Huntsville     AL    35816     -86.7     34.7
## # ... with 1,633 more rows
```

Between the dennys and states datasets, there are no discrepencies.

14 LAQUINTAS WERE FILTERED OUT!

Going forward we will work with the data from the United States only, so you can redefine `dennys` and `laquinta` to include only US locations. Include an indication of how many Denny's and La Quinta records are filtered out because they are not in the US.

Computing Frequencies

We now compute some density statistics for the two franchises.

By State

```
most_states_dennys <- dennys %>%
  group_by(state) %>%
  summarize(Num = n()) %>%
  arrange(desc(Num))

most_states_dennys
```

```
## # A tibble: 51 x 2
##    state   Num
##    <chr> <int>
##  1 CA      403
##  2 TX      200
##  3 FL      140
##  4 AZ       83
##  5 IL       56
##  6 NY       56
##  7 WA       49
##  8 OH       44
##  9 MO       42
## 10 PA       40
## # ... with 41 more rows
```

```
most_states_laquinta <- laquinta %>%
  group_by(state) %>%
  summarize(Num = n()) %>%
  arrange(desc(Num))

most_states_laquinta
```

```
## # A tibble: 48 x 2
##    state   Num
##    <chr> <int>
##  1 TX      237
##  2 FL       74
##  3 CA       56
##  4 GA       41
##  5 TN       30
##  6 OK       29
##  7 LA       28
##  8 CO       27
##  9 NM       19
## 10 NY       19
## # ... with 38 more rows
```

```
most_states_laquinta <- laquinta %>%
  group_by(state) %>%
  summarize(Num = n()) %>%
  arrange(Num)

most_states_laquinta
```

```
## # A tibble: 48 x 2
##    state   Num
##    <chr> <int>
##  1 ME        1
##  2 AK        2
##  3 NH        2
##  4 RI        2
##  5 SD        2
##  6 VT        2
##  7 WV        3
##  8 WY        3
##  9 IA        4
## 10 MI        4
## # ... with 38 more rows
```

```
most_states_dennys <- dennys %>%
  group_by(state) %>%
  summarize(Num = n()) %>%
  arrange(Num)

most_states_dennys
```

```
## # A tibble: 51 x 2
##    state   Num
##    <chr> <int>
##  1 DE        1
##  2 DC        2
##  3 VT        2
##  4 AK        3
##  5 IA        3
##  6 NH        3
##  7 SD        3
##  8 WV        3
##  9 LA        4
## 10 MT        4
## # ... with 41 more rows
```

The Most Denny: California (This makes total sense to me because Dennys was founded there!)

The Least Denny: Delaware (This did not surprise me because Delaware is located on the opposite side of the coast which would make it less known there)

The Most La Quinta: Texas (This also makes sense to me because La Quinta was dounded in San Antonio, Texas; making it demand more there.)

The Least La Quinta: Maine (Maine probably has a low tourist attraction which makes sense for there to be one La Qunita)

Compute which states have the most and fewest Denny's locations. Do the same for La Quinta. Describe the results and discuss whether there's anything interesting about them.

By Area

```
most_area_dennys <- dennys %>%
  count(state) %>%
  inner_join(states, by = c("state" = "abbreviation")) %>%
  arrange(desc(area/n))

most_area_dennys
```

```
## # A tibble: 51 x 4
##    state     n name            area
##    <chr> <int> <chr>          <dbl>
##  1 AK        3 Alaska       665384.
##  2 MT        4 Montana      147040.
##  3 SD        3 South Dakota  77116.
##  4 WY        4 Wyoming       97813.
##  5 IA        3 Iowa          56273.
##  6 ND        4 North Dakota  70698.
##  7 NE        5 Nebraska      77348.
##  8 LA        4 Louisiana     52378.
##  9 KS        8 Kansas        82278.
## 10 MS        5 Mississippi   48432.
## # ... with 41 more rows
```

```
most_area_laquinta <- laquinta %>%
  count(state) %>%
  inner_join(states, by = c("state" = "abbreviation")) %>%
  arrange(desc(area/n))

most_area_laquinta
```

```
## # A tibble: 48 x 4
##    state     n name            area
##    <chr> <int> <chr>          <dbl>
##  1 AK        2 Alaska       665384.
##  2 SD        2 South Dakota  77116.
##  3 ME        1 Maine         35380.
##  4 WY        3 Wyoming       97813.
##  5 MI        4 Michigan      96714.
##  6 MT        9 Montana      147040.
##  7 NE        5 Nebraska      77348.
##  8 ND        5 North Dakota  70698.
##  9 IA        4 Iowa          56273.
## 10 NV        8 Nevada       110572.
## # ... with 38 more rows
```

This dataset in a essence, reassure the validity of the joke that Mitch Herdberg made. There are 5+ countries that share the leader boards on each tibble which makes it fair to conclude such a joke!

Compute which states have the most Denny's locations *per thousand square miles*. Do the same for La Quinta. Describe the results and discuss whether there's anything interesting about them. *Note*: to do this, you'll need to *look up* the area of the states in the `states` dataset. Start with the table of counts by state that you just computed above and then *join* it with the table that has the area of each state. The states' areas are given in square miles; make a new column for the result of the unit conversion.

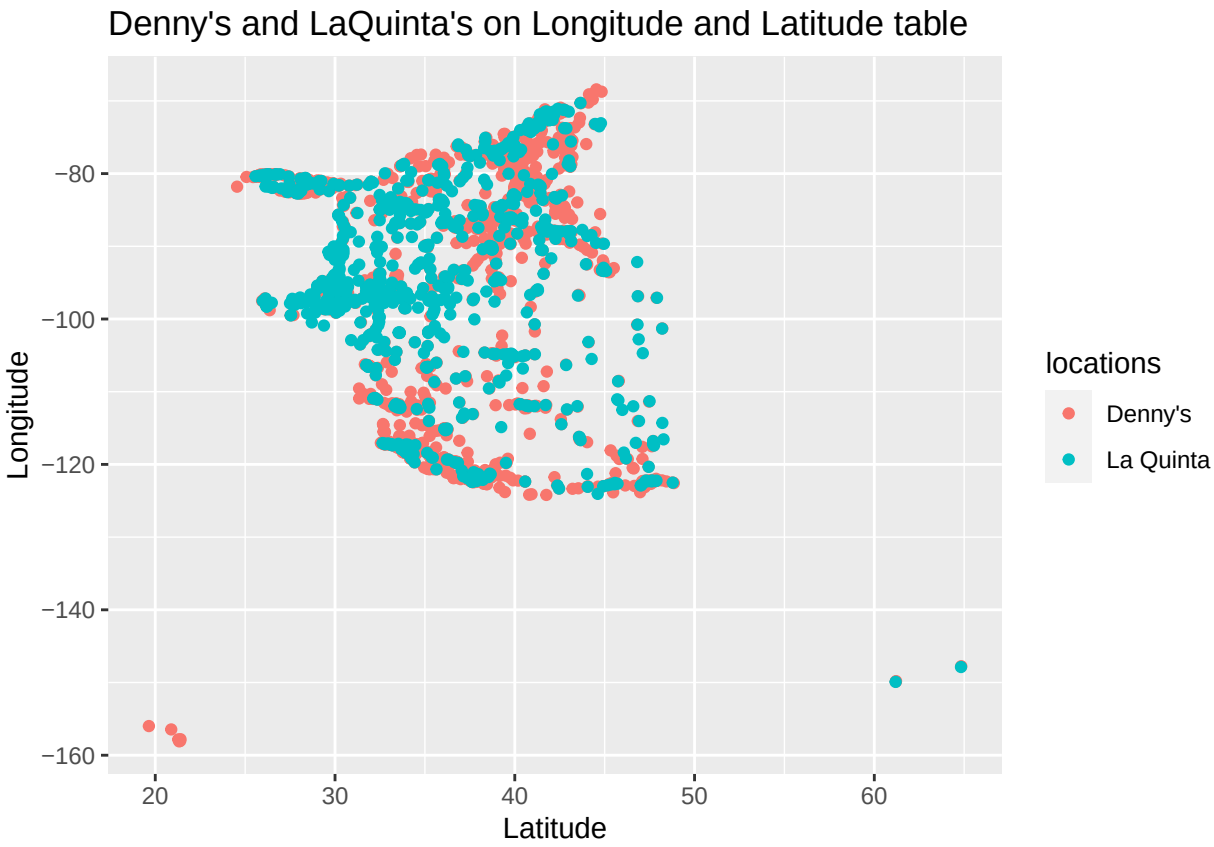
Plotting Locations

Finally, we plot the locations (latitude and longitude) of both establishments.

To do this, we'll put the two datasets together into a single data frame. First, we'll add an identifier variable called `establishment` to distinguish rows from the two datasets, and then we'll bind them with the `bind_rows` function (n.b., this binding works because the two tables have the same columns):

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

```
dn_lq %>%
  ggplot() +
  aes(x=latitude, y=longitude, color=establishment) +
  geom_point() +
  labs (
    x = "Latitude",
    y = "Longitude",
    title = "Denny's and LaQuinta's on Longitude and Latitude table",
    color="locations"
  )
```



Plot the locations of the two establishments using a scatter plot, and color the points by the establishment type. *Note*: the latitude is plotted on the x-axis and the longitude on the y-axis.

Drawing Conclusions

This graph proves that for most part, that all Denny and La quinta are co-located together. Which makes the joke even more funnier!

Note: the homework for this week concludes this analysis.