Lab 6 - La Quinta is Spanish for next to Denny's, Pt. 1

Visualizing spatial data

FATIMAH YOUSEF ALMOSAWI

Student ID: 2201002709

Load the needed packages. Note: dsbox is not yet on CRAN. For now, you need to install it before you load the library. Uncomment, by deleting # from, the two lines of code below.

```
install.packages("devtools")
devtools::install_github("rstudio-education/dsbox")
```

Load the libraries here.

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

To help with our analysis we will also use a dataset on US states, which is located in your repository's `data` folder.

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

Exercises

1. What are the dimensions of the Denny's dataset? (Hint: Use inline R code and functions like `nrow` and `ncol` to compose your answer.) What does each row in the dataset represent? What are the variables?

```
dim(dennys)
```

```
## [1] 1643    6
```

```
head(dennys)
```

```
## # A tibble: 6 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
```

The 'dennys' dataset has 1,643 rows and 6 columns. Each row represents location information about a denny's restaurant.

2. What are the dimensions of the La Quinta's dataset? What does each row in the dataset represent? What are the variables?

```
dim(laquinta)
```

```
## [1] 909   6
```

```
head(laquinta)
```

```
## # A tibble: 6 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
```

This dataset has 909 rows and 6 columns. Each row represents location information about laquinta's hotel.

Knit, *commit, and push your changes to GitHub with an appropriate commit message. Make sure to commit and push all changed files so that your Git pane is cleared up afterwards.*

3. Add a country variable to the Denny's and Laquinta's datasets and set all observations equal to `"United States"`. Remember, you can use the `mutate` function for adding a variable. Make sure to save the result of this as `dn` and lq, respectively, so that the stored data frame contains the new variable going forward. head(laquinta)

   dn <- dennys %>% + mutate(country = "United States") lq <- laquinta %>% + mutate(country = "United States")

For Denny's

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

For La Quinta

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

Knit, *commit, and push your changes to GitHub with an appropriate commit message. Make sure to commit and push all changed files so that your Git pane is cleared up afterwards.*

4. Which states have the most and fewest Denny's locations? What about La Quinta? Is this surprising? Why or why not?

   dn %>% + count(state) %>% + arrange(desc(n)) dn %>% + count(state) %>% + arrange((n)) lq %>% + count(state) %>% + arrange(desc(n)) lq %>% + count(state) %>% + arrange((n))

For Denny's - Most

```
dn %>%
    count(state) %>%
    arrange(desc(n))
```

```
## # A tibble: 51 x 2
##    state     n
##    <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
```

CA, which stands for California has the most Denny's locations. It is not surprising because it is the largest state in the US by population.

For Denny's - Fewest

```
dn %>%
    count(state) %>%
    arrange((n))
```

```
## # A tibble: 51 x 2
##    state     n
##    <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
```

DE, Delaware has the fewest locations. It is not surprising because it is one of the smallest US states by population.

For La Quinta - Most

```
lq %>%
    count(state) %>%
    arrange(desc(n))
```

```
## # A tibble: 59 x 2
##    state     n
##    <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 49 more rows
```

TX, Texas has the most La Quinta locations. The surprising thing is that Texas has more locations than California. This despite the fact that California has more population and has stronger economy than Texas

For La Quinta - Fewest

```
lq %>%
    count(state) %>%
    arrange((n))
```

```
## # A tibble: 59 x 2
##    state     n
##    <chr> <int>
##  1 AG        1
##  2 ANT       1
##  3 BC        1
##  4 CH        1
##  5 FM        1
##  6 ME        1
##  7 ON        1
##  8 QR        1
##  9 SL        1
## 10 VE        1
## # ... with 49 more rows
```

10 states with only one La Quinta Locations

5. Which states have the most Denny's locations per thousand square miles? What about La Quinta?

   dn %>% + count(state) %>% + inner_join(states, by = c("state" = "abbreviation")) lq %>% + count(state) %>% + inner_join(states, by = c("state" = "abbreviation")) dn <- dn %>% + mutate(establishment = "Denny's") lq <- lq %>% + mutate(establishment = "La Quinta") dn_lq <- bind_rows(dn, lq) ggplot(dn_lq, mapping = aes(x = longitude, y = latitude, color = establishment)) + + geom_point()

\textcolor{red}{We are interested in the number of locations per state, which we could find in Denny's or La Quinta's datasets, and the area of each state, which is only available in the `states` dataset, which we imported earlier from the Data folder using the code `states <- read_csv("data/states.csv")`.}

Joining Denny's and states

```
dn %>%
  count(state) %>%
  inner_join(states, by = c("state" = "abbreviation"))
```

```
## # A tibble: 51 x 4
##    state     n name                    area
##    <chr> <int> <chr>                  <dbl>
##  1 AK        3 Alaska               665384.
##  2 AL        7 Alabama               52420.
##  3 AR        9 Arkansas              53179.
##  4 AZ       83 Arizona              113990.
##  5 CA      403 California           163695.
##  6 CO       29 Colorado             104094.
##  7 CT       12 Connecticut            5543.
##  8 DC        2 District of Columbia     68.3
##  9 DE        1 Delaware              2489.
## 10 FL      140 Florida               65758.
## # ... with 41 more rows
```

Joining La Quinta's and states

```
lq %>%
  count(state) %>%
```

```
  inner_join(states, by = c("state" = "abbreviation"))
```

```
## # A tibble: 48 x 4
##    state     n name           area
##    <chr> <int> <chr>         <dbl>
##  1 AK        2 Alaska      665384.
##  2 AL       16 Alabama      52420.
##  3 AR       13 Arkansas     53179.
##  4 AZ       18 Arizona     113990.
##  5 CA       56 California  163695.
##  6 CO       27 Colorado    104094.
##  7 CT        6 Connecticut   5543.
##  8 FL       74 Florida      65758.
##  9 GA       41 Georgia      59425.
## 10 IA        4 Iowa         56273.
## # ... with 38 more rows
```

We need to add an identifier variable which we'll call `establishment` and set the value to `"Denny's"` and `"La Quinta"` for the `dn` and `lq` data frames, respectively.

```
dn <- dn %>%
  mutate(establishment = "Denny's")
```

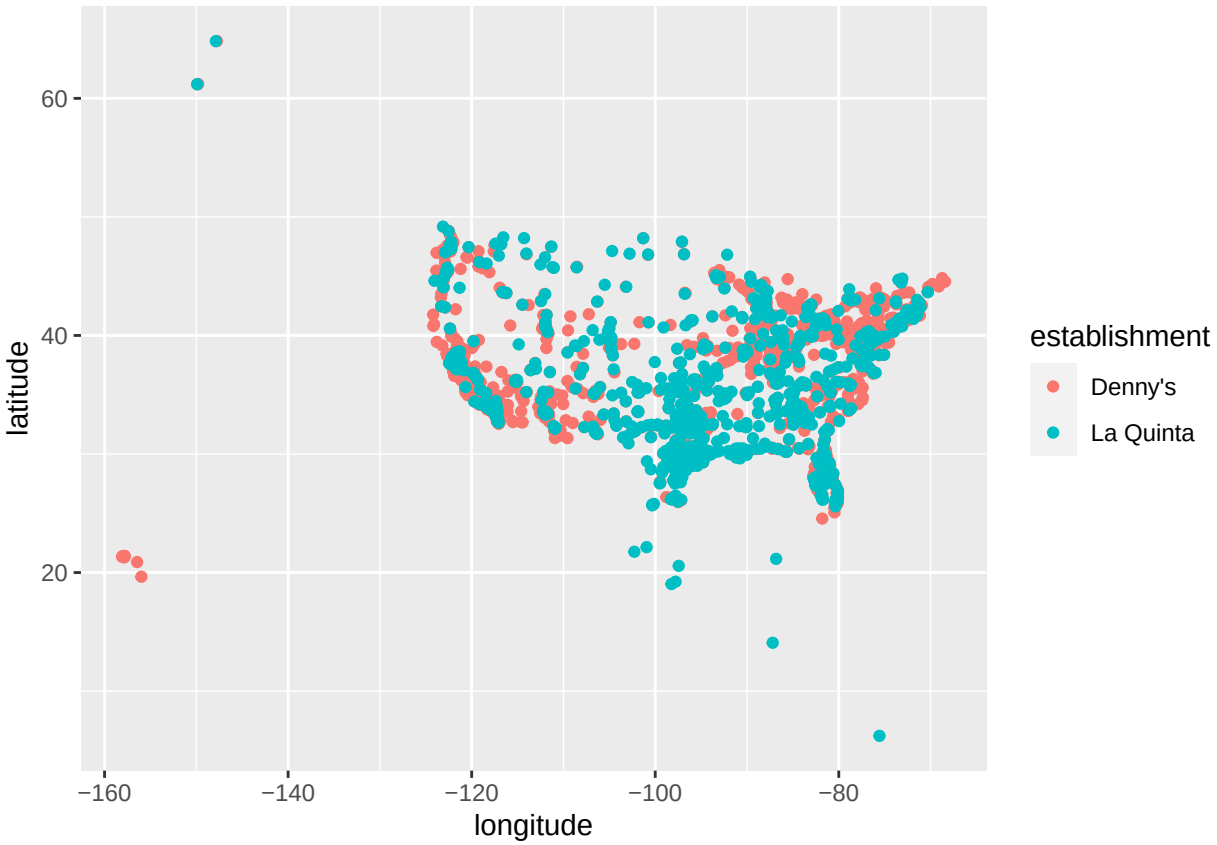
```
lq <- lq %>%
  mutate(establishment = "La Quinta")
```

Since the two data frames (`dn` and `lq`) have the same columns, we can easily bind them with the `bind_rows` function:

```
dn_lq <- bind_rows(dn, lq)
```

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

```
ggplot(dn_lq, mapping = aes(x = longitude, y = latitude, color = establishment)) +
    geom_point()
```



The graph figure shows that indeed there are many La Quinta hotels that located next to Denny's restaurants. We could also see from the figure that there is the presence of the two chains across the US as the figure almost depict the exact map of the country. The figure also indicates the presence of these chains in Alaska (top-left), Hawaii (bottom-left), and some cities in Central America(bottom-right)

Knit, *commit, and push your changes to GitHub with an appropriate commit message. Make sure to commit and push all changed files so that your Git pane is cleared up afterwards and review the md document on GitHub to make sure you're happy with the final state of your work.* Done LAB 6