

Homework 04 - Wrangling Spatial Data

Gloria Grace

Load packages and data

```
library(tidyverse)
```

```
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()  
))
```

Exercise 1

```
dn_ak <- dennys %>%  
  filter(state == "AK")  
nrow(dn_ak)
```

```
## [1] 3
```

There are 3 Denny's location in Alaska.

Exercise 2

```
lq_ak <- laquinta %>%  
  filter(state == "AK")  
nrow(lq_ak)
```

```
## [1] 2
```

There are 2 La quinta's location in Alaska.

Exercise 3

```
dn_lq_ak <- full_join(dn_ak, lq_ak, by = "state", suffix = c("_",  
dn_lq_ak
```

```
## # A tibble: 6 x 11  
##   address_dn city_dn state zip_dn longitude_dn latitude_dn  
##   <chr>      <chr>   <chr> <chr>          <dbl>         <dbl>  
## 1 2900 Denali Anchora... AK      99503        -150.         61.2  
## 2 2900 Denali Anchora... AK      99503        -150.         61.2  
## 3 3850 Debar... Anchora... AK      99508        -150.         61.2  
## 4 3850 Debar... Anchora... AK      99508        -150.         61.2  
## 5 1929 Airpo... Fairban... AK      99701        -148.         64.8  
## 6 1929 Airpo... Fairban... AK      99701        -148.         64.8  
## # ... with 3 more variables: zip_lq <chr>, longitude_lq <dbl>,
```

There are a total of 6 pairings locations which also 6 times of distance calculation that needs to be done.

Exercise 4

There are 6 observation in the koin dn_lq_ak data frame. The names of the variables in this data frame is: address_dn, city_dn, state, zip_dn, longitude_dn, latitude_dn, address_lq, city_lq, zip_lq, longitude_lq, latitude_lq.

Exercise 5

```
# Implementation from dsbox  
haversine <- function(long1, lat1, long2, lat2) {  
  # convert to radians  
  long1 = long1 * pi / 180  
  lat1 = lat1 * pi / 180  
  long2 = long2 * pi / 180
```

```

lat2 = lat2 * pi / 180

# Earth mean radius in km (WGS84 ellipsoid)
R = 6371.009

a = sin((lat2 - lat1)/2)^2 + cos(lat1) * cos(lat2) * sin((lon2 - lon1)/2)^2
d = R * 2 * asin(sqrt(a))
# Return the distance in km
return(d)
}

```

```

dn_lq_ak <- dn_lq_ak %>%
  mutate(distance = haversine(longitude_dn, latitude_dn, longitude_lq, latitude_lq))
dn_lq_ak

```

```

## # A tibble: 6 x 12
##   address_dn city_dn state zip_dn longitude_dn latitude_dn distance
##   <chr>      <chr>   <chr> <chr>          <dbl>        <dbl>
## 1 2900 Denali Anchorage AK    99503      -150.        61.2
## 2 2900 Denali Anchorage AK    99503      -150.        61.2
## 3 3850 Debar... Anchorage AK    99508      -150.        61.2
## 4 3850 Debar... Anchorage AK    99508      -150.        61.2
## 5 1929 Airpo... Fairbanks AK    99701      -148.        64.8
## 6 1929 Airpo... Fairbanks AK    99701      -148.        64.8
## # ... with 4 more variables: zip_lq <chr>, longitude_lq <dbl>, latitude_lq <dbl>,
## #   distance <dbl>

```

Exercise 6

```

dn_lq_ak %>%
  group_by(address_lq) %>%
  summarise(count = n(), dist_to_nearest = round(min(distance), 2))

```

```

## # A tibble: 2 x 3
##   address_lq          count dist_to_nearest
##   <chr>          <int>          <dbl>
## 1 3501 Minnesota Dr.      3            2.03
## 2 4920 Dale Rd           3            5.2

```

Exercise 7

Based on the result on exercise 6, the nearest La quinta location that is next to Denny's is 2.03 km and 5.20, which is possible if they are next to each other.

Exercise 8

```
compute_closest_dennys <- function(the_state) {  
  dn_state <- dennys %>%  
    filter(state == the_state)  
  lq_state <- laquinta %>%  
    filter(state == the_state)  
  
  # join to get all possible pairings  
  dn_lq_state <- full_join(dn_state, lq_state, by = "state", suffixes = c("_dn", "_lq"))  
  
  # compute distances  
  mutate(distance = haversine(longitude_dn, latitude_dn, longitude_lq, latitude_lq))  
  
  # compute distance to closest Denny's for each La Quinta location  
  group_by(address_lq, city_lq) %>%  
    summarise(dist_to_nearest = round(min(distance), digits = 2))  
}
```

Exercise 9

```
dn_lq_nc <- compute_closest_dennys("NC") %>%  
  arrange(dist_to_nearest)
```

`summarise()` has grouped output by 'address_lq'. You can override this with `group_by()`.

```
dn_lq_nc
```

```
## # A tibble: 12 x 3
## # Groups:   address_lq [12]
##   address_lq              city_lq          dist_to_ne
##   <chr>                <chr>
## 1 4900 South Tryon St      "\nCharlotte"
## 2 2211 Summit Park Ln     "\nRaleigh"
## 3 4414 Durham Chapel Hill Blvd "\nDurham"
## 4 1910 Westpark Dr        "\nDurham"
## 5 3127 Sloan Dr           "\nCharlotte"
## 6 191 Crescent Commons    "\nCary"
## 7 1001 Hospitality Ct     "\nMorrisville"
## 8 1001 Aerial Center Pkwy  "\nMorrisville"
## 9 1201 Lanada Rd          "\nGreensboro"
## 10 2020 Griffith Rd       "\nWinston-Salem"
## 11 1607 Fairgrove Church Rd "\nConover"
## 12 165 Hwy 105 Extension   "\nBoone"
```

```
dn_lq_nc %>%
  ungroup(address_lq) %>%
  summarise(min(dist_to_nearest), round(median(dist_to_nearest), 2))
```

```
## # A tibble: 1 x 2
##   `min(dist_to_nearest)` `round(median(dist_to_nearest), 2)`
##   <dbl>
## 1      1.78
```

Exercise 10

```
dn_lq_tx <- compute_closest_dennys("TX") %>%
  arrange(dist_to_nearest)
```

```
## `summarise()` has grouped output by 'address_lq'. You can override this with `groups = NULL`.
```

```
dn_lq_tx
```

```
## # A tibble: 237 x 3
## # Groups:   address_lq [237]
##   address_lq          city_lq          dist_to_nearest
##   <chr>              <chr>              <dbl>
## 1 601 Ave Q           "\nLubbock"          0.0
## 2 1402 Seawall Blvd   "\nGalveston"        0.0
## 3 3610 Santa Ursula Ave "\nLaredo"           0.0
## 4 7625 East Ben White Blvd "\nAustin"           0.0
## 5 1112 South Fort Hood St "\nKilleen"           0.0
## 6 28673 I-45 North     "\nThe Woodlands"    0.0
## 7 11033 Gateway Blvd West "\nEl Paso"           0.0
## 8 2950 Wood Ridge Blvd  "\nBrenham"           0.0
## 9 4105 West Airport Fwy "\nIrving"            0.0
## 10 1708 Interstate 40 East "\nAmarillo"          0.0
## # ... with 227 more rows
```

```
dn_lq_tx %>%
  ungroup(address_lq) %>%
  summarise(min(dist_to_nearest), round(median(dist_to_nearest), 2))
```

```
## # A tibble: 1 x 2
##   `min(dist_to_nearest)` `round(median(dist_to_nearest), 2)`
##   <dbl>
## 1      0.02
```

Exercise 11

```
dn_lq_ca <- compute_closest_dennys("CA") %>%
  arrange(dist_to_nearest)
```

```
## `summarise()` has grouped output by 'address_lq'. You can override this with `group_by()`.
```

```
dn_lq_ca
```

```
## # A tibble: 56 x 3
## # Groups:   address_lq [56]
```

```
##   address_lq           city_lq           dist_to_ne
##   <chr>              <chr>
## 1 190 N. 10th Street   "\nFowler"
## 2 14972 Sand Canyon Ave "\nIrvine"
## 3 1500 N. Cherry Ct    "\nTulare"
## 4 20 Airport Blvd     "\nSouth San Francisco"
## 5 8858 Spectrum Park Way "\nBakersfield"
## 6 3232 Riverside Dr    "\nBakersfield"
## 7 316 Pittman Road     "\nFairfield"
## 8 150 Bonita Rd        "\nChula Vista"
## 9 330 East Fir Ave     "\nFresno"
## 10 200 Jibboom St      "\nSacramento"
## # ... with 46 more rows
```

```
dn_lq_ca %>%
  ungroup(address_lq) %>%
  summarise(min(dist_to_nearest), round(median(dist_to_nearest), 2))
```

```
## # A tibble: 1 x 2
##   `min(dist_to_nearest)` `round(median(dist_to_nearest), 2)`
##   <dbl>
## 1      0.02
```

Exercise 12

Among all of the states that I examined, in California Mitch Hedberg's joke are most likely to hold true. The nearest distance between La quinta and Denny in california is 0.02 km and the median is 1.37 where I can conclude that there are many side-by-side in California because from the data it can be easily tell that they had the most nearest distance out of the other two states. ## Wrap-Up

GitHub permalink

<https://github.com/Calvin-DS202-21FA/portfolio-gloriagrce/blob/613d54a3f8a6ea5a2b108c99fd01cc90cda5f4a9/hw/hw04-wrangle-spatial/hw04-wrangle-spatial.Rmd>

Reflections

When working on this homework, I got stuck quite a while when I am on the function part of the code. I was frustrated on how to call the code and eventually I got it to work, first I tried with the mutate and assigning the function to a variable. But then when in the function there is a code where it adds dataframe, I got confused on how to call the function without the data frame exist from the start. I am not sure that I am right but I tried calling the function and assign it to a new data frame which it works.

For the second function too, when there is "return(dist_to_closest)" it said that dist_to_closest wasn't defined, which is why I tried myself to remove the variable from the summarise() function. But then, the data that came out was not what I want it to be. Therefore, I tried to remove the return and it works. Is this how suppose to be or I luckily get passed it?

It took me a couple of hours and I think I will have to focus more on calling and making a function in R studio.