DATA 202 Homework 3

Adham Rishmawi

2022-10-12

```
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()
))
```

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

```
[1] 3
```

1.there are 3 locations in Alaska (Dennys)

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

```
[1] 2
```

2.there are 2 locations in Alaska (laquinta)

```
dn_lq_ak <- full_join(dn_ak, lq_ak, by = "state", suffix = c("_dn", "_lq"))
dn_lq_ak
```

```
# A tibble: 6 x 11
  address_dn    city_dn state zip_dn longitude_dn latitude_dn address_lq city_lq
  <chr>         <chr>   <chr> <chr>         <dbl>       <dbl> <chr>      <chr>
1 2900 Denali   Anchor~ AK    99503         -150.        61.2 3501 Minn~ "\nAnc~
2 2900 Denali   Anchor~ AK    99503         -150.        61.2 4920 Dale~ "\nFai~
3 3850 Debarr ~ Anchor~ AK    99508         -150.        61.2 3501 Minn~ "\nAnc~
4 3850 Debarr ~ Anchor~ AK    99508         -150.        61.2 4920 Dale~ "\nFai~
5 1929 Airport~ Fairba~ AK    99701         -148.        64.8 3501 Minn~ "\nAnc~
6 1929 Airport~ Fairba~ AK    99701         -148.        64.8 4920 Dale~ "\nFai~
# ... with 3 more variables: zip_lq <chr>, longitude_lq <dbl>,
#   latitude_lq <dbl>
```

4. How many observations are in the joined dn_lq_ak data frame? What are the names of the variables in this data frame. :

the .x has to come from the x data and .y came from the second the full_join call. The reason for two different .y and .x is because there was a similar column in both and the system had to differentiate between them!

5. Add a new variable, called distance to the dn_lq_ak data frame that contains the distances between

each Denny's and La Quinta locations. Make sure to save the result back to dn_lq_ak so that you can use it later.

```
# Great-circle distance, 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

  # Compute the distance in km
  a = sin((lat2 - lat1)/2)^2 + cos(lat1) * cos(lat2) * sin((long2 - long1)/2)^2
  d = R * 2 * asin(sqrt(a))
  return(d)
}
```

```
dn_lq_ak|>
  mutate(distance = haversine(longitude_dn,latitude_dn,longitude_lq,latitude_lq))
```

```
# A tibble: 6 x 12
  address_dn    city_dn state zip_dn longitude_dn latitude_dn address_lq city_lq
  <chr>         <chr>   <chr> <chr>         <dbl>       <dbl> <chr>      <chr>
1 2900 Denali   Anchor~ AK    99503         -150.        61.2 3501 Minn~ "\nAnc~
2 2900 Denali   Anchor~ AK    99503         -150.        61.2 4920 Dale~ "\nFai~
3 3850 Debarr ~ Anchor~ AK    99508         -150.        61.2 3501 Minn~ "\nAnc~
4 3850 Debarr ~ Anchor~ AK    99508         -150.        61.2 4920 Dale~ "\nFai~
5 1929 Airport~ Fairba~ AK    99701         -148.        64.8 3501 Minn~ "\nAnc~
6 1929 Airport~ Fairba~ AK    99701         -148.        64.8 4920 Dale~ "\nFai~
# ... with 4 more variables: zip_lq <chr>, longitude_lq <dbl>,
#   latitude_lq <dbl>, distance <dbl>
```

6.

```
dn_lq_ak_mindist <- dn_lq_ak %>%
  group_by(address_lq) %>%
  summarise(min(dn_lq_ak$distance))
```

```
Warning: Unknown or uninitialised column: `distance`.
```

```
Warning in min(dn_lq_ak$distance): no non-missing arguments to min; returning
Inf
```

```
Warning: Unknown or uninitialised column: `distance`.
```

```
Warning in min(dn_lq_ak$distance): no non-missing arguments to min; returning
Inf
```

```
dn_lq_ak_mindist
```

```
# A tibble: 2 x 2
  address_lq          `min(dn_lq_ak$distance)`
  <chr>                                  <dbl>
1 3501 Minnesota Dr.                       Inf
2 4920 Dale Rd                             Inf
```

7. There are 3 la Quinta locations near Dennys and they are in Anchorage(2),Fairbanks

8.

```
lq_ak1 <- laquinta %>%
  filter(state == "MI")
nrow(lq_ak1)
```

```
[1] 4
```

```
dn_ak1 <- dennys %>%
  filter(state == "MI")
nrow(dn_ak1)
```

```
[1] 22
```

```
dn_lq_ak1 <- full_join(dn_ak1, lq_ak1, by = "state", suffix = c("_dn", "_lq"))
dn_lq_ak1
```

```
# A tibble: 88 x 11
   address_dn   city_dn state zip_dn longitude_dn latitude_dn address_lq city_lq
   <chr>        <chr>   <chr> <chr>         <dbl>       <dbl> <chr>      <chr>
 1 3310 Washte~ Ann Ar~ MI    48104         -83.7        42.3 41211 For~ "\nCan~
 2 3310 Washte~ Ann Ar~ MI    48104         -83.7        42.3 30847 Fly~ "\nRom~
 3 3310 Washte~ Ann Ar~ MI    48104         -83.7        42.3 12888 Ree~ "\nSou~
 4 3310 Washte~ Ann Ar~ MI    48104         -83.7        42.3 45311 Par~ "\nUti~
 5 4785 Beckle~ Battle~ MI    49015         -85.2        42.3 41211 For~ "\nCan~
 6 4785 Beckle~ Battle~ MI    49015         -85.2        42.3 30847 Fly~ "\nRom~
 7 4785 Beckle~ Battle~ MI    49015         -85.2        42.3 12888 Ree~ "\nSou~
 8 4785 Beckle~ Battle~ MI    49015         -85.2        42.3 45311 Par~ "\nUti~
 9 2033 Rawson~ Bellev~ MI    48111         -83.5        42.2 41211 For~ "\nCan~
10 2033 Rawson~ Bellev~ MI    48111         -83.5        42.2 30847 Fly~ "\nRom~
# ... with 78 more rows, and 3 more variables: zip_lq <chr>,
#   longitude_lq <dbl>, latitude_lq <dbl>
```

```
dn_lq_ak1|>
  mutate(distance = haversine(longitude_dn,latitude_dn,longitude_lq,latitude_lq))
```

```
# A tibble: 88 x 12
   address_dn   city_dn state zip_dn longitude_dn latitude_dn address_lq city_lq
   <chr>        <chr>   <chr> <chr>         <dbl>       <dbl> <chr>      <chr>
 1 3310 Washte~ Ann Ar~ MI    48104         -83.7        42.3 41211 For~ "\nCan~
 2 3310 Washte~ Ann Ar~ MI    48104         -83.7        42.3 30847 Fly~ "\nRom~
 3 3310 Washte~ Ann Ar~ MI    48104         -83.7        42.3 12888 Ree~ "\nSou~
 4 3310 Washte~ Ann Ar~ MI    48104         -83.7        42.3 45311 Par~ "\nUti~
 5 4785 Beckle~ Battle~ MI    49015         -85.2        42.3 41211 For~ "\nCan~
 6 4785 Beckle~ Battle~ MI    49015         -85.2        42.3 30847 Fly~ "\nRom~
 7 4785 Beckle~ Battle~ MI    49015         -85.2        42.3 12888 Ree~ "\nSou~
 8 4785 Beckle~ Battle~ MI    49015         -85.2        42.3 45311 Par~ "\nUti~
 9 2033 Rawson~ Bellev~ MI    48111         -83.5        42.2 41211 For~ "\nCan~
10 2033 Rawson~ Bellev~ MI    48111         -83.5        42.2 30847 Fly~ "\nRom~
# ... with 78 more rows, and 4 more variables: zip_lq <chr>,
#   longitude_lq <dbl>, latitude_lq <dbl>, distance <dbl>
```

```
dn_lq_ak_mindist1 <- dn_lq_ak1 %>%
  group_by(address_lq) %>%
  summarise(closest = min(dn_lq_ak1$distance))
```

```
Warning: Unknown or uninitialised column: `distance`.

Warning in min(dn_lq_ak1$distance): no non-missing arguments to min; returning
Inf

Warning: Unknown or uninitialised column: `distance`.

Warning in min(dn_lq_ak1$distance): no non-missing arguments to min; returning
Inf

Warning: Unknown or uninitialised column: `distance`.

Warning in min(dn_lq_ak1$distance): no non-missing arguments to min; returning
Inf

Warning: Unknown or uninitialised column: `distance`.

Warning in min(dn_lq_ak1$distance): no non-missing arguments to min; returning
Inf
```

the data from Michigan shows that there are way more locations od dennys/laquinta that are near each other (r wasnt working correctly but i know this code works by working on personal r)