

Lab 04 - La Quinta is Spanish for next to Denny's, Pt. 2

Alison Yao (yy2564) Oct 12, 2021

Load packages and data

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

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

Exercise 1

There are 3 Denny's locations in Alaska.

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

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

Exercise 2

There are 2 La Quinta locations in Alaska.

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

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

Exercise 3

There are $3 * 2 = 6$ pairs between all Denny's and all La Quinta locations in Alaska, so we need to calculate 6 distances.

Exercise 4

```
dn_lq_ak <- full_join(dn_ak, lq_ak, by = "state")
dn_lq_ak
```

```
## # A tibble: 6 × 11
##   address.x city.x state zip.x longitude.x latitude.x address.y city.
##   <chr>      <chr> <chr> <chr>      <dbl>      <dbl> <chr>      <chr>
## 1 2900 Denali Ancho... AK    99503      -150.      61.2 3501 Minn... "\nAn
## 2 2900 Denali Ancho... AK    99503      -150.      61.2 4920 Dale... "\nFa
## 3 3850 Debar... Ancho... AK    99508      -150.      61.2 3501 Minn... "\nAn
## 4 3850 Debar... Ancho... AK    99508      -150.      61.2 4920 Dale... "\nFa
## 5 1929 Airpo... Fairb... AK    99701      -148.      64.8 3501 Minn... "\nAn
## 6 1929 Airpo... Fairb... AK    99701      -148.      64.8 4920 Dale... "\nFa
## # ... with 2 more variables: longitude.y <dbl>, latitude.y <dbl>
```

There are 6 observations. The variables are address.x, city.x, state, zip.x, longitude.x, latitude.x, address.y, city.y, zip.y, longitude.y, and latitude.y.

Exercise 5

```
haversine <- function(long1, lat1, long2, lat2, round = 3) {
  # convert to radians
  long1 = long1 * pi / 180
  lat1 = lat1 * pi / 180
  long2 = long2 * pi / 180
  lat2 = lat2 * pi / 180

  R = 6371 # Earth mean radius in km

  a = sin((lat2 - lat1)/2)^2 + cos(lat1) * cos(lat2) * sin((long2 - long1)/2)^2
  d = R * 2 * asin(sqrt(a))

  return( round(d,round) ) # distance in km
}
```

Exercise 6

```
dn_lq_ak <- dn_lq_ak %>%
  mutate(distance = haversine(longitude.x, latitude.y, longitude.y, latitude.x))
dn_lq_ak
```

```
## # A tibble: 6 × 12
##   address.x   city.x state zip.x longitude.x latitude.x address.y   city.
##   <chr>      <chr> <chr> <chr>      <dbl>      <dbl> <chr>      <chr>
## 1 2900 Denali Ancho... AK    99503      -150.      61.2 3501 Minn... "\nAn
## 2 2900 Denali Ancho... AK    99503      -150.      61.2 4920 Dale... "\nFa
## 3 3850 Debar... Ancho... AK    99508      -150.      61.2 3501 Minn... "\nAn
## 4 3850 Debar... Ancho... AK    99508      -150.      61.2 4920 Dale... "\nFa
## 5 1929 Airpo... Fairb... AK    99701      -148.      64.8 3501 Minn... "\nAn
## 6 1929 Airpo... Fairb... AK    99701      -148.      64.8 4920 Dale... "\nFa
## # ... with 3 more variables: longitude.y <dbl>, latitude.y <dbl>, distance
```

Exercise 7

```
dn_lq_ak_mindist <- dn_lq_ak %>%
  group_by(address.x) %>%
  summarise(closest = min(distance))
dn_lq_ak_mindist
```

```
## # A tibble: 3 × 2
##   address.x      closest
##   <chr>          <dbl>
## 1 1929 Airport Way    5.01
## 2 2900 Denali        1.89
## 3 3850 Debarr Road   5.51
```

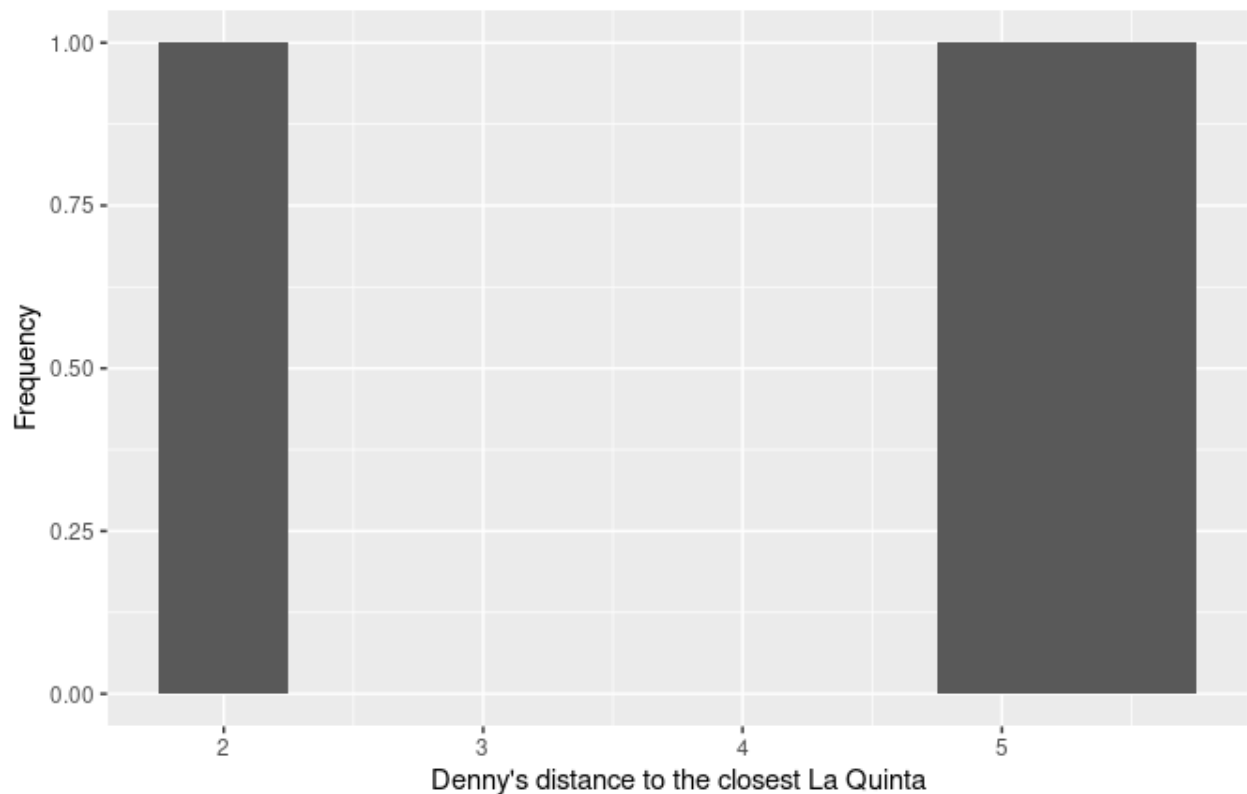
Exercise 8

The 3 closest distances ranges from 1.886 to 5.514, with a mean of 4.138, a medium of 5.014 and a standard deviation of 1.97.

```
dn_lq_ak_mindist %>%
  ggplot(aes(x = closest)) +
  geom_histogram(binwidth = 0.5) +
  labs(
    title = "Distances between Denny's and the nearest La Quinta locations",
    subtitle = "Data filtered by Alaska only",
    x = "Denny's distance to the closest La Quinta",
    y = "Frequency"
  )
```

Distances between Denny's and the nearest La Quinta locations in Alaska

Data filtered by Alaska only



```
dn_lq_ak_mindist %>%
  summarise(
    count = n(),
    min_dist = min(closest),
    mean_dist = mean(closest),
    median_dist = median(closest),
    max_dist = max(closest),
    range_dist = max(closest) - min(closest),
    SD_dist = sd(closest)
  )
```

```
## # A tibble: 1 × 7
##   count min_dist mean_dist median_dist max_dist range_dist SD_dist
##   <int>   <dbl>   <dbl>     <dbl>   <dbl>     <dbl>   <dbl>
## 1     3     1.89     4.14     5.01     5.51     3.63     1.97
```

Exercise 9

```
dn_nc <- dennys %>% filter(state == "NC") # 28
dn_nc
```

```
## # A tibble: 28 × 6
##   address          city          state zip longitude
##   <chr>          <chr>        <chr> <chr>     <dbl>
```

```
## 1 1 Regent Park Boulevard Asheville NC 28806 -82.6
## 2 7135 Nc #4 Battleboro NC 27809 -77.8
## 3 581 South Highway 9 Black Mountain NC 28711 -82.3
## 4 4541 Sunset Rd Charlotte NC 28216 -80.9
## 5 516 Tyvola Rd Charlotte NC 28217 -80.9
## 6 University of North Carolina At Charlotte NC 28223 -80.7
## 7 8031 Concord Mills Blvd Concord NC 28027 -80.7
## 8 7021 Highway 751, #901 Durham NC 27707 -79.0
## 9 5505 Raeford Rd Fayetteville NC 28304 -79.0
## 10 808 S Memorial Dr Greenville NC 27834 -77.4
## # ... with 18 more rows
```

```
lq_nc <- laquinta %>% filter(state == "NC") # 12
lq_nc
```

```
## # A tibble: 12 × 6
##   address          city      state zip longitude
##   <chr>          <chr>    <chr> <chr>    <dbl>
## 1 165 Hwy 105 Extension "\nBoone" NC 28607 -81.7
## 2 3127 Sloan Dr "\nCharlotte" NC 28208 -80.9
## 3 4900 South Tryon St "\nCharlotte" NC 28217 -80.9
## 4 4414 Durham Chapel Hill Blvd "\nDurham" NC 27707 -79.0
## 5 1910 Westpark Dr "\nDurham" NC 27713 -78.9
## 6 1201 Lanada Rd "\nGreensboro" NC 27407 -79.9
## 7 1607 Fairgrove Church Rd "\nConover" NC 28613 -81.3
## 8 191 Crescent Commons "\nCary" NC 27518 -78.8
## 9 2211 Summit Park Ln "\nRaleigh" NC 27612 -78.7
## 10 1001 Aerial Center Pkwy "\nMorrisville" NC 27560 -78.8
## 11 1001 Hospitality Ct "\nMorrisville" NC 27560 -78.8
## 12 2020 Griffith Rd "\nWinston-Salem" NC 27103 -80.3
```

```
dn_lq_nc <- full_join(dn_nc, lq_nc, by = "state") # 336
dn_lq_nc
```

```
## # A tibble: 336 × 11
##   address.x city.x state zip.x longitude.x latitude.x address.y city
##   <chr>    <chr> <chr> <chr>    <dbl>    <dbl> <chr>    <chr>
## 1 1 Regent ... Ashevi... NC 28806 -82.6 35.6 165 Hwy 1... "\nB
## 2 1 Regent ... Ashevi... NC 28806 -82.6 35.6 3127 Sloa... "\nC
## 3 1 Regent ... Ashevi... NC 28806 -82.6 35.6 4900 Sout... "\nC
## 4 1 Regent ... Ashevi... NC 28806 -82.6 35.6 4414 Durh... "\nD
## 5 1 Regent ... Ashevi... NC 28806 -82.6 35.6 1910 West... "\nD
## 6 1 Regent ... Ashevi... NC 28806 -82.6 35.6 1201 Lana... "\nG
## 7 1 Regent ... Ashevi... NC 28806 -82.6 35.6 1607 Fair... "\nC
## 8 1 Regent ... Ashevi... NC 28806 -82.6 35.6 191 Cresc... "\nC
## 9 1 Regent ... Ashevi... NC 28806 -82.6 35.6 2211 Summ... "\nR
## 10 1 Regent ... Ashevi... NC 28806 -82.6 35.6 1001 Aeri... "\nM
## # ... with 326 more rows, and 2 more variables: longitude.y <dbl>,
## # latitude.y <dbl>
```

```

dn_lq_nc <- dn_lq_nc %>%
  mutate(distance = haversine(longitude.x, latitude.y, longitude.y, latitude.y))
dn_lq_nc_mindist <- dn_lq_nc %>%
  group_by(address.x) %>%
  summarise(closest = min(distance))
dn_lq_nc_mindist # 28

```

```

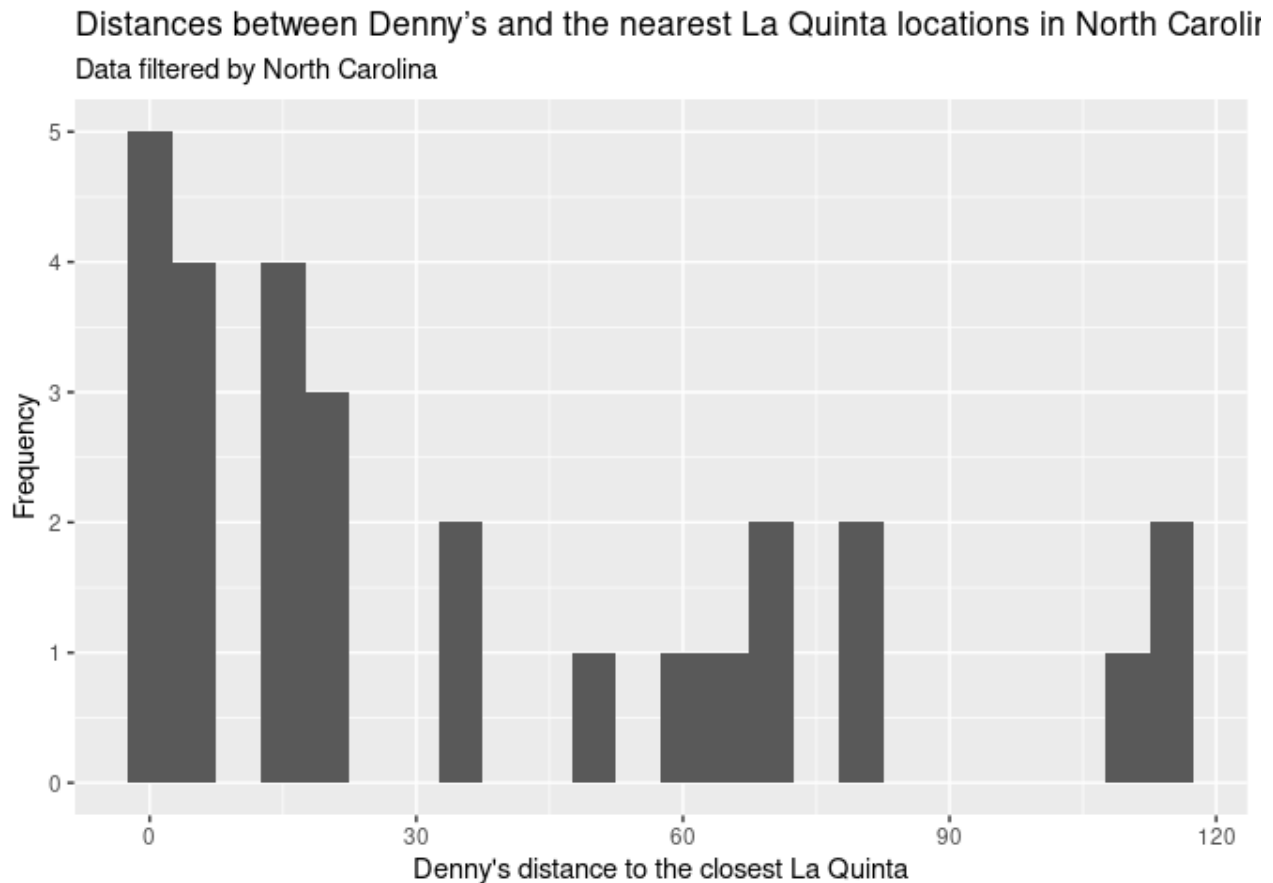
## # A tibble: 28 × 2
##   address.x          closest
##   <chr>          <dbl>
## 1 1 Regent Park Boulevard    81.9
## 2 101 Wintergreen Dr         2.84
## 3 103 Sedgehill Dr          16.1
## 4 1043 Jimmie Kerr Road     34.2
## 5 1201 S College Road       71.2
## 6 1209 Burkemount Avenue     2.31
## 7 1493 Us Hwy 74-A Bypass    20.0
## 8 1524 Dabney Dr            21.8
## 9 1550 Four Seasons         70.0
## 10 1800 Princeton-Kenly Road 47.9
## # ... with 18 more rows

```

```

dn_lq_nc_mindist %>%
  ggplot(aes(x = closest)) +
  geom_histogram(binwidth = 5) +
  labs(
    title = "Distances between Denny's and the nearest La Quinta locations",
    subtitle = "Data filtered by North Carolina",
    x = "Denny's distance to the closest La Quinta",
    y = "Frequency"
  )

```



The 28 closest distances ranges from 0.253 to 116.304, with a mean of 36.75639, a medium of 19.021 and a standard deviation of 37.96253.

```
dn_lq_nc_mindist %>%
  summarise(
    count = n(),
    min_dist = min(closest),
    mean_dist = mean(closest),
    median_dist = median(closest),
    max_dist = max(closest),
    range_dist = max(closest) - min(closest),
    SD_dist = sd(closest)
  )
```

A tibble: 1 × 7

	count	min_dist	mean_dist	median_dist	max_dist	range_dist	SD_dist
## 1	28	0.253	36.8	19.0	116.	116.	38.0

Exercise 10

```
dn_tx <- dennys %>% filter(state == "TX") # 200
dn_tx
```

```
lq_tx <- laquinta %>% filter(state == "TX") # 237
lq_tx
```

```
dn_lq_tx <- full_join(dn_tx, lq_tx, by = "state") # 47400
dn_lq_tx
```

https://af81d303fa69478baa7c40bfd49d038f.app.rstudio.cloud/file_show?path=%2Ftmp%2Ftmp0eE4hW%2Fpreview-235132ea0ca.html


```
## 10 120 East ... Abilene TX      79601      -99.6      32.4 9305 East... "\nA
## # ... with 47,390 more rows, and 2 more variables: longitude.y <dbl>,
## #   latitude.y <dbl>
```

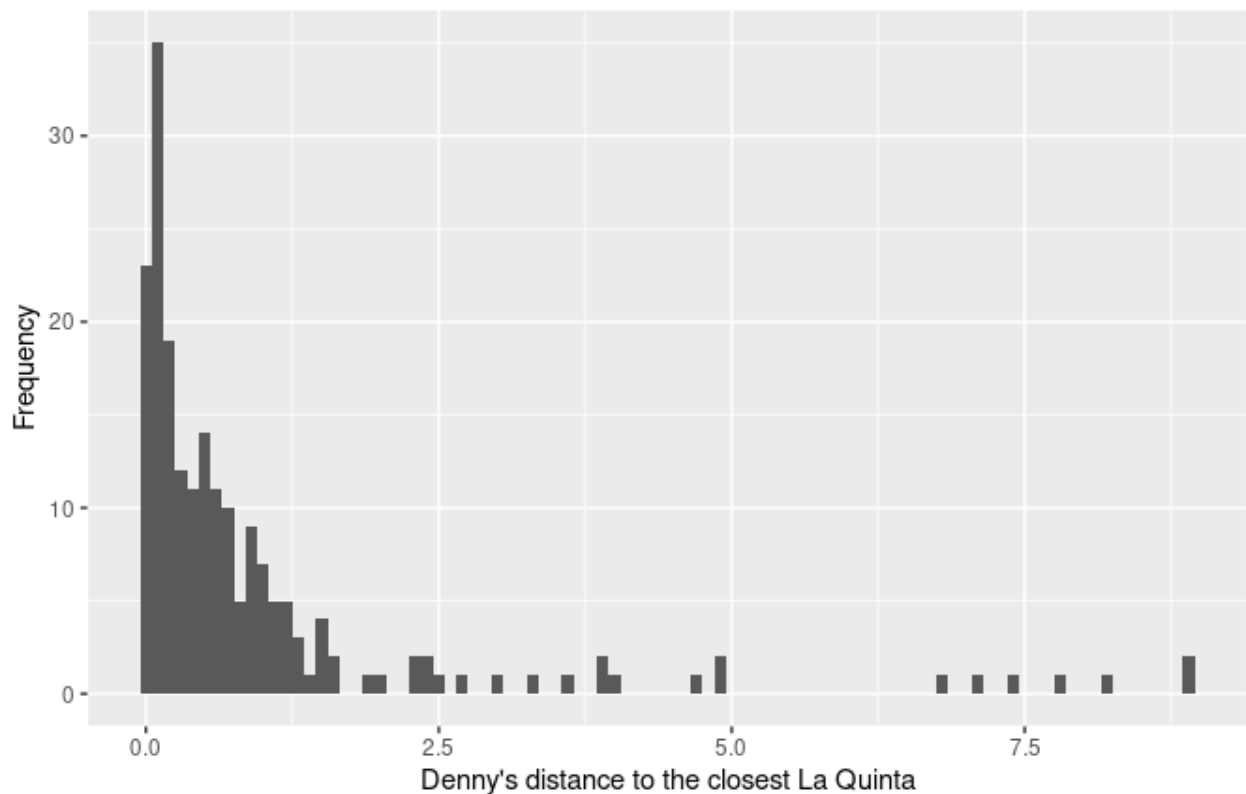
```
dn_lq_tx <- dn_lq_tx %>%
  mutate(distance = haversine(longitude.x, latitude.y, longitude.y, latitud
dn_lq_tx_mindist <- dn_lq_tx %>%
  group_by(address.x) %>%
  summarise(closest = min(distance))
dn_lq_tx_mindist # 200
```

```
## # A tibble: 200 × 2
##   address.x      closest
##   <chr>         <dbl>
## 1 100 Cottonwood      7.80
## 2 100 E Pinehurst    1.30
## 3 100 Us Highway 79 S  0.56
## 4 101 N Fm 707       8.86
## 5 1011 Beltway Parkway 1.91
## 6 1015 Spur 350 West  0.462
## 7 1015 West Main St   1.06
## 8 10367 Highway 59    1.96
## 9 10433 N Central Expwy 0.052
## 10 105 W 42nd St      6.80
## # ... with 190 more rows
```

```
dn_lq_tx_mindist %>%
  ggplot(aes(x = closest)) +
  geom_histogram(binwidth = 0.1) +
  labs(
    title = "Distances between Denny's and the nearest La Quinta locations",
    subtitle = 'Data filtered by Texas',
    x = "Denny's distance to the closest La Quinta",
    y = 'Frequency'
  )
```

Distances between Denny's and the nearest La Quinta locations in Texas

Data filtered by Texas



The 200 closest distances ranges from 0.001 to 8.892, with a mean of 0.95, a median of 0.452 and a standard deviation of 1.608521.

```
dn_lq_tx_mindist %>%
  summarise(
    count = n(),
    min_dist = min(closest),
    mean_dist = mean(closest),
    median_dist = median(closest),
    max_dist = max(closest),
    range_dist = max(closest) - min(closest),
    SD_dist = sd(closest)
  )
```

```
## # A tibble: 1 × 7
##   count min_dist mean_dist median_dist max_dist range_dist SD_dist
##   <int>   <dbl>   <dbl>     <dbl>   <dbl>     <dbl>   <dbl>
## 1    200    0.001     0.952     0.452     8.89      8.89    1.61
```

Exercise 11

Here, I am going to choose New York.

```
dn_ny <- dennys %>% filter(state == "NY") # 56
dn_ny
```

```
## # A tibble: 56 × 6
```

	address	city	state	zip	longitude	latitude
	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>
## 1	114 Wolf Road	Albany	NY	12205	-73.8	42.7
## 2	3920 Maple Road	Amherst	NY	14226	-78.8	43.0
## 3	176 Grant Ave	Auburn	NY	13021	-76.5	42.9
## 4	364 W Main St	Batavia	NY	14020	-78.2	43.0
## 5	1250 Upper Front St	Binghamton	NY	13901	-75.9	42.2
## 6	805 Pennsylvania Ave	Brooklyn	NY	11207	-73.9	40.7
## 7	2215 Delaware Ave	Buffalo	NY	14216	-78.9	42.9
## 8	4445 Main St	Buffalo	NY	14226	-78.8	43.0
## 9	5300 W Genessee St	Camillus	NY	13031	-76.3	43.0
## 10	160 Eastern Blvd	Canandaigua	NY	14424	-77.3	42.9

```
## # ... with 46 more rows
```

```
lq_ny <- laquinta %>% filter(state == "NY") # 19
lq_ny
```

```
## # A tibble: 19 × 6
```

	address	city	state	zip	longitude
	<chr>	<chr>	<chr>	<chr>	<dbl>
## 1	94 Business Park Dr	"\nArmonk"	NY	10504	-73.7
## 2	8200 Park Road	"\nBatavia"	NY	14020	-78.2
## 3	581 Harry L. Dr	"\nJohnson City"	NY	13790	-76.0
## 4	1229 Atlantic Ave.	"\nBrooklyn"	NY	11216	-74.0
## 5	533 3rd Ave	"\nBrooklyn"	NY	11215	-74.0
## 6	1412 Pitkin Ave	"\nBrooklyn"	NY	11233	-73.9
## 7	6619 Transit Rd	"\nWilliamsville"	NY	14221	-78.7
## 8	1749 Route 9	"\nClifton Park"	NY	12065	-73.8
## 9	540 Saw Mill River Rd	"\nElmsford"	NY	10523	-73.8
## 10	4317 Rockaway Beach Blvd.	"\nFar Rockaway"	NY	11692	-73.8
## 11	821 Stewart Ave	"\nGarden City"	NY	11530	-73.6
## 12	10 Aero Rd	"\nBohemia"	NY	11716	-73.1
## 13	833 New Loudon Rd	"\nLatham"	NY	12110	-73.8
## 14	17 West 32nd St	"\nNew York"	NY	10001	-74.0
## 15	31 West 71st Street	"\nNew York"	NY	10023	-74.0
## 16	6225 Niagara Falls Blvd	"\nNiagara Falls"	NY	14304	-79.0
## 17	16 Plaza Blvd	"\nPlattsburgh"	NY	12901	-73.5
## 18	37-18 Queens Blvd	"\nLong Island City"	NY	11101	-73.9
## 19	5394 Willow Place	"\nVerona"	NY	13478	-75.6

```
dn_lq_ny <- full_join(dn_ny, lq_ny, by = "state") # 1064
dn_lq_ny
```

```
## # A tibble: 1,064 × 11
##   address.x city.x state zip.x longitude.x latitude.x address.y city.
##   <chr>      <chr> <chr> <chr>      <dbl>      <dbl> <chr>      <chr>
## 1 114 Wolf ... Albany NY 12205 -73.8 42.7 94 Busine... "\nAr
## 2 114 Wolf ... Albany NY 12205 -73.8 42.7 8200 Park... "\nBa
## 3 114 Wolf ... Albany NY 12205 -73.8 42.7 581 Harry... "\nJo
## 4 114 Wolf ... Albany NY 12205 -73.8 42.7 1229 Atla... "\nBr
## 5 114 Wolf ... Albany NY 12205 -73.8 42.7 533 3rd A... "\nBr
## 6 114 Wolf ... Albany NY 12205 -73.8 42.7 1412 Pitk... "\nBr
## 7 114 Wolf ... Albany NY 12205 -73.8 42.7 6619 Tran... "\nWi
## 8 114 Wolf ... Albany NY 12205 -73.8 42.7 1749 Rout... "\nCl
## 9 114 Wolf ... Albany NY 12205 -73.8 42.7 540 Saw M... "\nEl
## 10 114 Wolf ... Albany NY 12205 -73.8 42.7 4317 Rock... "\nFa
## # ... with 1,054 more rows, and 2 more variables: longitude.y <dbl>,
## # latitude.y <dbl>
```

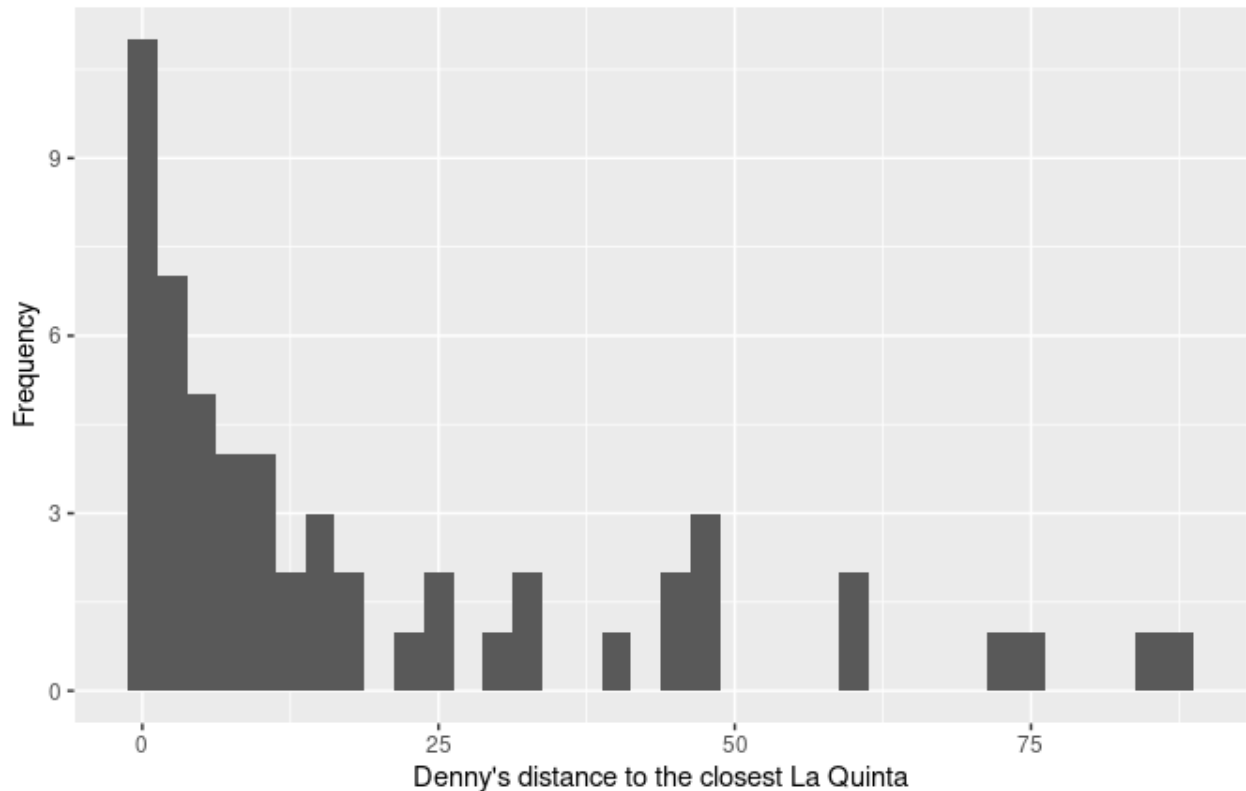
```
dn_lq_ny <- dn_lq_ny %>%
  mutate(distance = haversine(longitude.x, latitude.y, longitude.y, latitud
dn_lq_ny_mindist <- dn_lq_ny %>%
  group_by(address.x) %>%
  summarise(closest = min(distance))
dn_lq_ny_mindist # 56
```

```
## # A tibble: 56 × 2
##   address.x closest
##   <chr>      <dbl>
## 1 1 River St 16.2
## 2 103 Elwood Davis Road 16.5
## 3 10390 Bennet Road 25.7
## 4 1078 Glenwood Avenue 3.94
## 5 114 Wolf Road 0.804
## 6 1142 Arsenal St 2.62
## 7 1143 Deer Park Ave 13.9
## 8 118 Victory Highway 88.3
## 9 1250 Upper Front St 6.18
## 10 126 Troy Rd 0.513
## # ... with 46 more rows
```

```
dn_lq_ny_mindist %>%
  ggplot(aes(x = closest)) +
  geom_histogram(binwidth = 2.5) +
  labs(
    title = "Distances between Denny's and the nearest La Quinta locations",
    subtitle = 'Data filtered by New York',
    x = "Denny's distance to the closest La Quinta",
    y = 'Frequency'
  )
```

Distances between Denny's and the nearest La Quinta locations in New York

Data filtered by New York



The 56 closest distances ranges from 0.017 to 88.297, with a mean of 19.66382, a medium of 9.908 and a standard deviation of 23.8474.

```
dn_lq_ny_mindist %>%
  summarise(
    count = n(),
    min_dist = min(closest),
    mean_dist = mean(closest),
    median_dist = median(closest),
    max_dist = max(closest),
    range_dist = max(closest) - min(closest),
    SD_dist = sd(closest)
  )
```

```
## # A tibble: 1 × 7
##   count min_dist mean_dist median_dist max_dist range_dist SD_dist
##   <int>   <dbl>   <dbl>     <dbl>   <dbl>     <dbl>   <dbl>
## 1     56    0.017    19.7       9.91    88.3      88.3    23.8
```

Exercise 12

Among the states I examined, Mitch Hedberg's joke most likely to hold true in Texas. Because all the distances from Denny's to the closest La Quinta is within 9. The distribution is strongly positively skewed, therefore indicating that the majority of the

distances in this distribution are very small. Also, there are 200 Denny's in Texas, which is a lot more than the other states I observed. Therefore,