Lab 1

Carmen Canedo

Necessary Packages

```r
library(mdsr)
```

Loading in data

```r
walking <- read.csv("lab-1.csv", header = TRUE)
```

Cleaning data

Getting rid of unnecessary

```r
walking <- walking %>%
  select(-type, -desc, -name)
```

Making column names simpler

```r
walking <- walking %>%
  rename(altitude = altitude..ft.) %>%
  rename(speed = speed..mph.) %>%
  rename(distance_mi = distance..mi.) %>%
  rename(distance_int_ft = distance_interval..ft.)
```

Summary stats

We will be working with a full population

```r
columns <- c(walking$time, walking$latitude, walking$longitude, walking$altitude, walking$speed, walkin

sum_latitude <- favstats( ~ latitude, data = walking)
sum_latitude
```

```
##       min       Q1  median       Q3      max    mean           sd   n
##  36.18001 36.18029 36.1806 36.18091 36.18117 36.1806 0.0003514294 221
##  missing
##        0
```

```
sum_longitude <- favstats( ~ longitude, data = walking)
sum_longitude
```

```
##        min        Q1   median        Q3       max     mean           sd
##  -86.74297 -86.74294 -86.7429 -86.74286 -86.74278 -86.7429 4.645446e-05
##    n missing
##  221       0
```

```
sum_altitude <- favstats( ~ altitude, data = walking)
sum_altitude
```

```
##    min    Q1 median    Q3   max     mean       sd   n missing
##  500.8 505.8  511.2 512.4 516.1 509.3181 4.041696 221       0
```

```
sum_speed <- favstats( ~ speed, data = walking)
sum_speed
```

```
##  min    Q1 median    Q3 max     mean       sd   n missing
##    0 2.275    2.7 3.325  10 2.839545 1.226507 220       1
```

```
sum_distance_mi <- favstats( ~ distance_mi, data = walking)
sum_distance_mi
```

```
##  min    Q1 median    Q3   max       mean         sd   n missing
##    0 0.047   0.09 0.137 0.181 0.09093213 0.05260591 221       0
```

```
distance_int_ft <- as.numeric(walking$distance_int_ft)
sum_dist_int_ft <- favstats( ~ distance_int_ft, data = walking)
sum_dist_int_ft
```

```
##  min   Q1 median   Q3   max    mean       sd   n missing
##    0 3.14   4.01 5.29 14.58 4.32362 1.937173 221       0
```

Question 1: The standard deviation is larger for latitude.

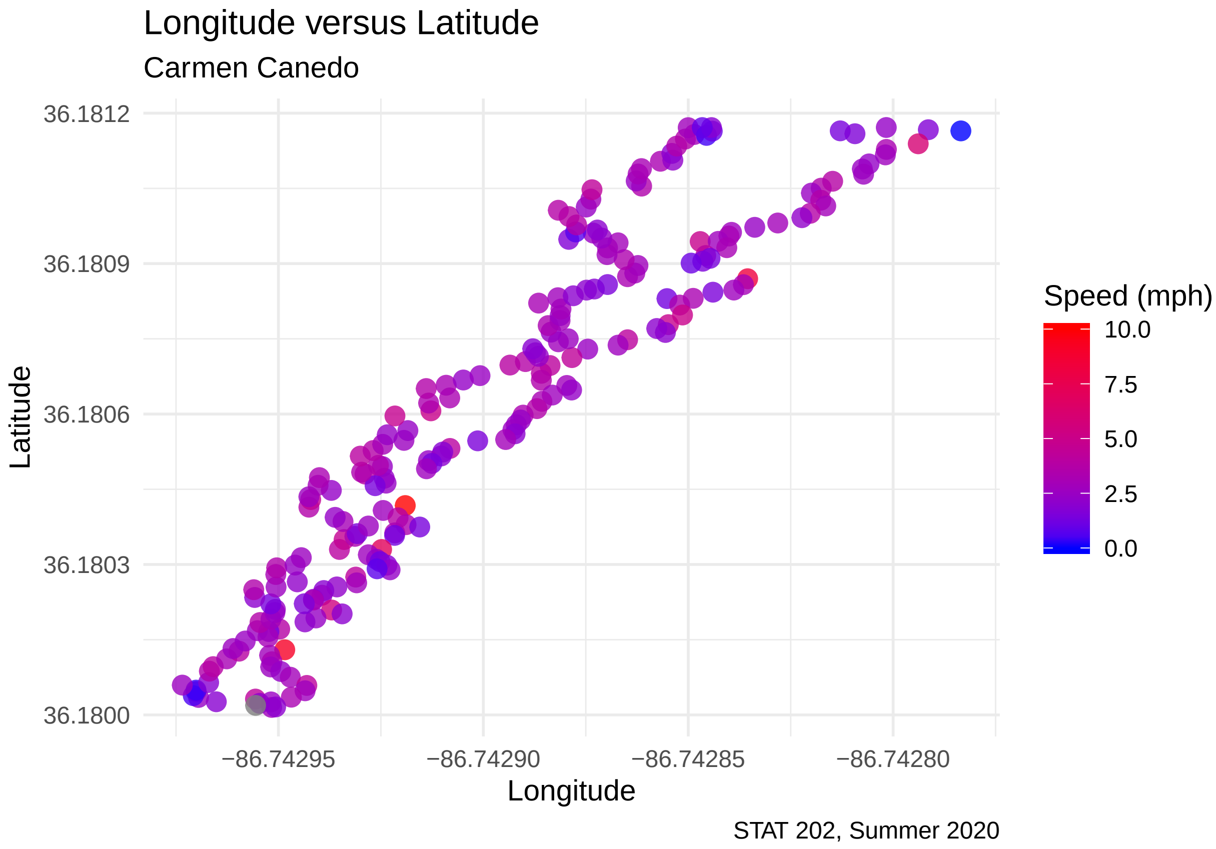
Question 2: This tells us that the latitude moves farther from the mean latitude.

Latitude v. Longitude Scatter Plot

x = longitude y = latitude group by = speed color scheme = red to blue scale

```
lat_v_long <- walking %>%
  ggplot(aes(x = longitude, y = latitude)) +
  geom_point(alpha = 0.8, aes(color = speed), size = 3) +
  scale_color_gradient(low = "blue", high = "red") +
  theme_minimal() +
  labs(title = "Longitude versus Latitude",
       subtitle = "Carmen Canedo",
       caption = "STAT 202, Summer 2020",
       x = "Longitude",
       y = "Latitude",
       color = "Speed (mph)")

lat_v_long
```



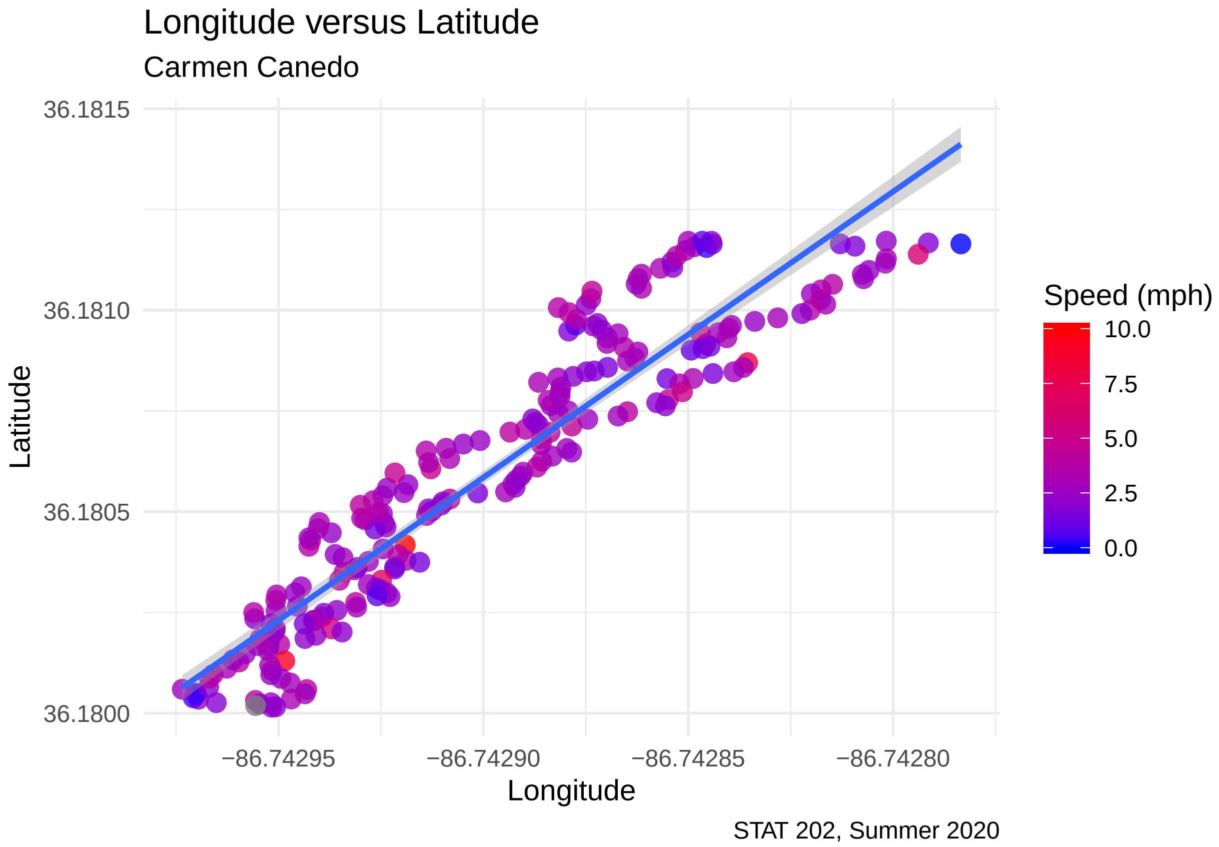
Adding Linear Regression

x = longitude y = latitude

Let's calculate the linear regression model

```
lat_v_long <- lat_v_long +
  geom_smooth(method = "lm")

lat_v_long
```



Details

- Equation for regression line and the correlation coefficient
- Is the line of best fit a good tool to estimate the path traveled? Why or why not?
- How does the correlation help you answer part b?