- Data import
- Histograms
- Scatterplot
- Scatterplot seperate
- Scatterplot with best fit line
- Scatterplot with best fit line seperate
- Questions

# data exploration 1

## Analysis of Environmental Data

Devon Parsons

# Data import

Loaded here, used read.csv(here(""))

```
require(here)
```

```
## Loading required package: here
```

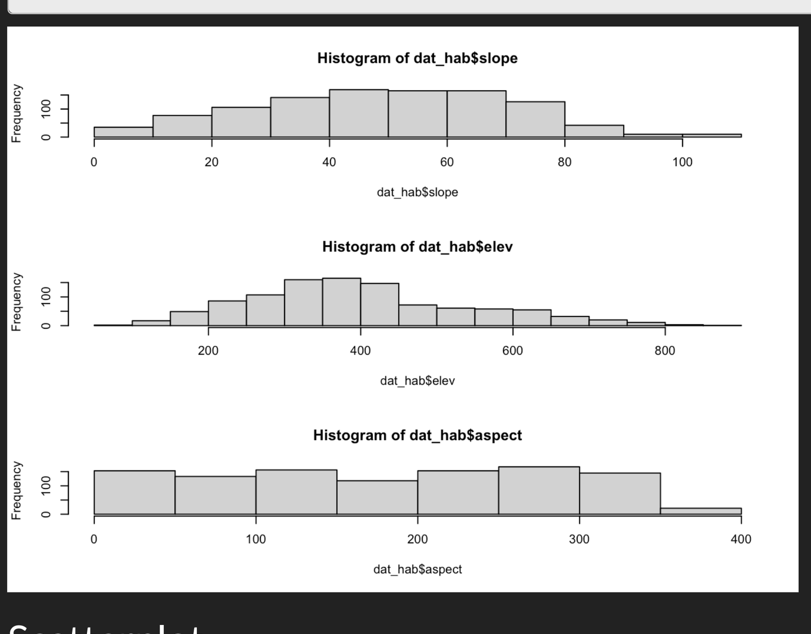
```
## here() starts at /Users/devonparsons/environmental_data
```

```
dat_hab <- read.csv(here("data", "hab.sta.csv"))

dat_bird <- read.csv(here("data", "bird.sta.csv"))
```

# Histograms

```
{
par(mfrow = c(3, 1))
hist(dat_hab$slope)
hist(dat_hab$elev)
hist(dat_hab$aspect)
}
```




# Scatterplot

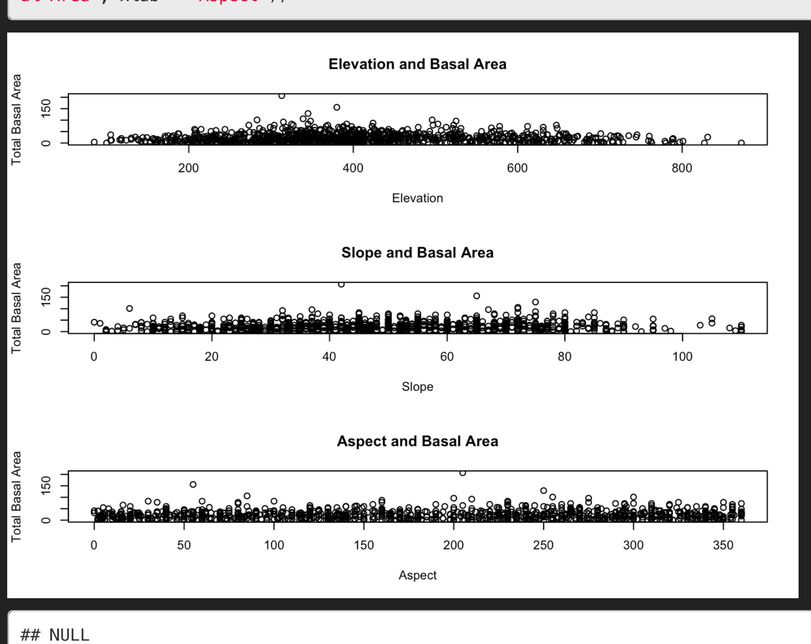
```
par(mfrow = c(3, 1))
(plot(x = dat_hab$elev, y = dat_hab$ba.tot, main = "Elevation and Basal Area", ylab = "Total Basal Area", xlab = "Elevation"))
```

```
## NULL
```

```
(plot(x = dat_hab$slope, y = dat_hab$ba.tot, main = "Slope and Basal Area", ylab = "Total Basal Area", xlab = "Slope"))
```

```
## NULL
```

```
(plot(x = dat_hab$aspect, y = dat_hab$ba.tot, main = "Aspect and Basal Area", ylab = "Total Basal Area", xlab = "Aspect"))
```

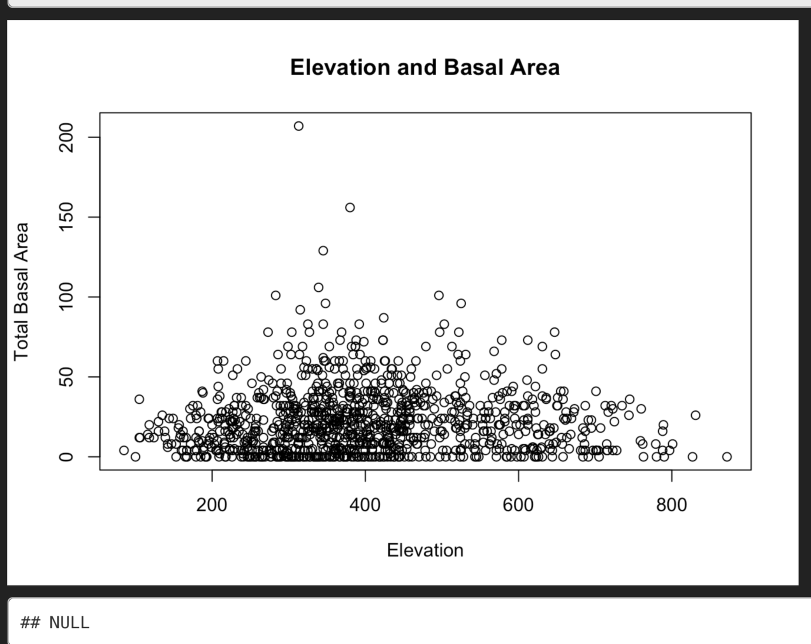



```
## NULL
```

# Scatterplot seperate

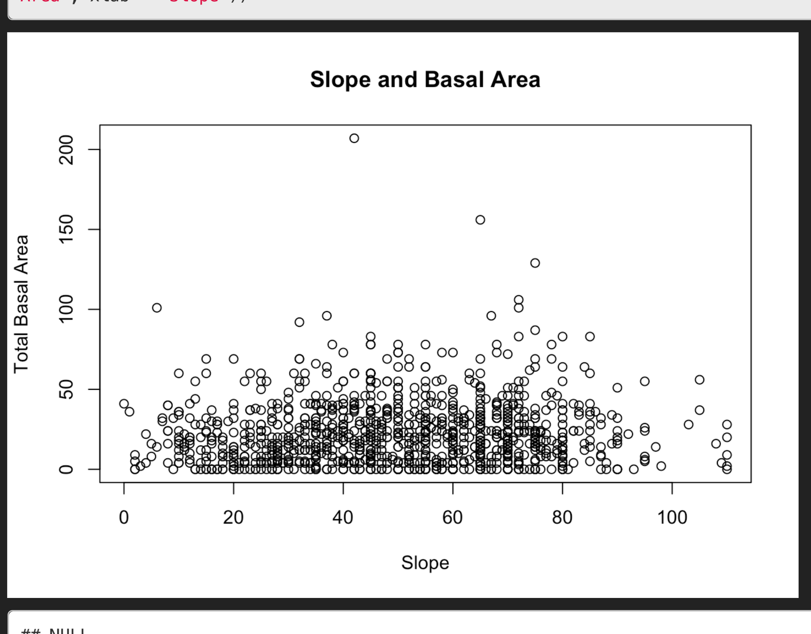
this chunk is for myself to get conclusions from because the par version is hard to look at

```
(plot(x = dat_hab$elev, y = dat_hab$ba.tot, main = "Elevation and Basal Area", ylab = "Total Basal Area", xlab = "Elevation"))
```



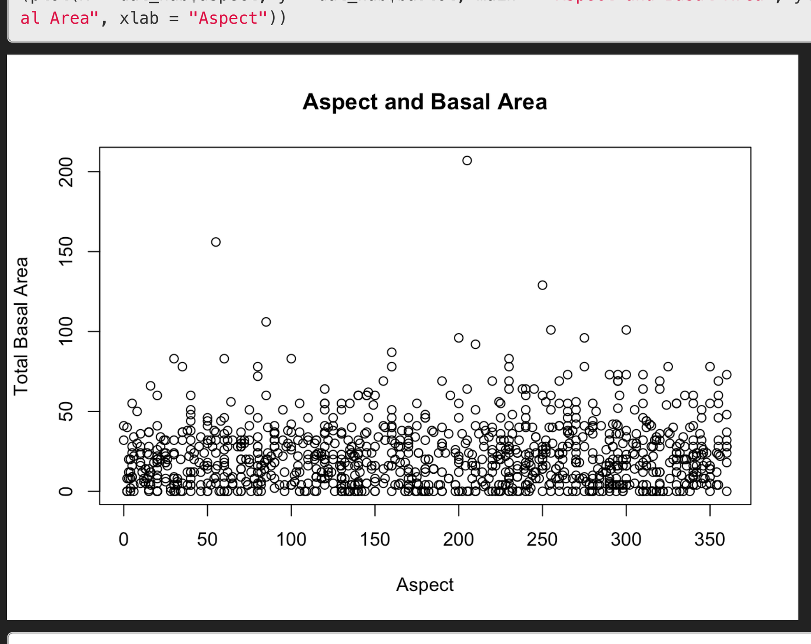

```
## NULL
```

```
(plot(x = dat_hab$slope, y = dat_hab$ba.tot, main = "Slope and Basal Area", ylab = "Total Basal Area", xlab = "Slope"))
```




```
## NULL
```

```
(plot(x = dat_hab$aspect, y = dat_hab$ba.tot, main = "Aspect and Basal Area", ylab = "Total Basal Area", xlab = "Aspect"))
```




```
## NULL
```

# Scatterplot with best fit line

```
par(mfrow = c(3, 1))
(plot(x = dat_hab$elev, y = dat_hab$ba.tot, main = "Elevation and Basal Area", ylab = "Total Basal Area", xlab = "Elevation"))
```

```
## NULL
```

```
data_center_x = mean(dat_hab$elev)
data_center_y = mean(dat_hab$ba.tot)
c(data_center_x, data_center_y)
```

```
## [1] 400.52008  22.42639
```

```
points(x = data_center_x, y = data_center_y, col = "red")
line_point_slope = function(x, x1, y1, slope)
{
  get_y_intercept =
    function(x1, y1, slope)
      return(-(x1 * slope) + y1)

  linear =
    function(x, yint, slope)
      return(yint + x * slope)

return(linear(x, get_y_intercept(x1, y1, slope), slope))
}
curve(
  line_point_slope(
    x,
    data_center_x,
    data_center_y,
    -0.1),
  add = TRUE)
(plot(x = dat_hab$slope, y = dat_hab$ba.tot, main = "Slope and Basal Area", ylab = "Total Basal Area", xlab = "Slope"))
```

```
## NULL
```

```
data_center_x = mean(dat_hab$slope)
data_center_y = mean(dat_hab$ba.tot)
c(data_center_x, data_center_y)
```

```
## [1] 50.26864 22.42639
```

```
points(x = data_center_x, y = data_center_y, col = "red")
line_point_slope = function(x, x1, y1, slope)
{
  get_y_intercept =
    function(x1, y1, slope)
      return(-(x1 * slope) + y1)

  linear =
    function(x, yint, slope)
      return(yint + x * slope)

return(linear(x, get_y_intercept(x1, y1, slope), slope))
}
curve(
  line_point_slope(
    x,
    data_center_x,
    data_center_y,
    -0.1),
  add = TRUE)

(plot(x = dat_hab$aspect, y = dat_hab$ba.tot, main = "Aspect and Basal Area", ylab = "Total Basal Area", xlab = "Aspect"))
```

```
## NULL
```

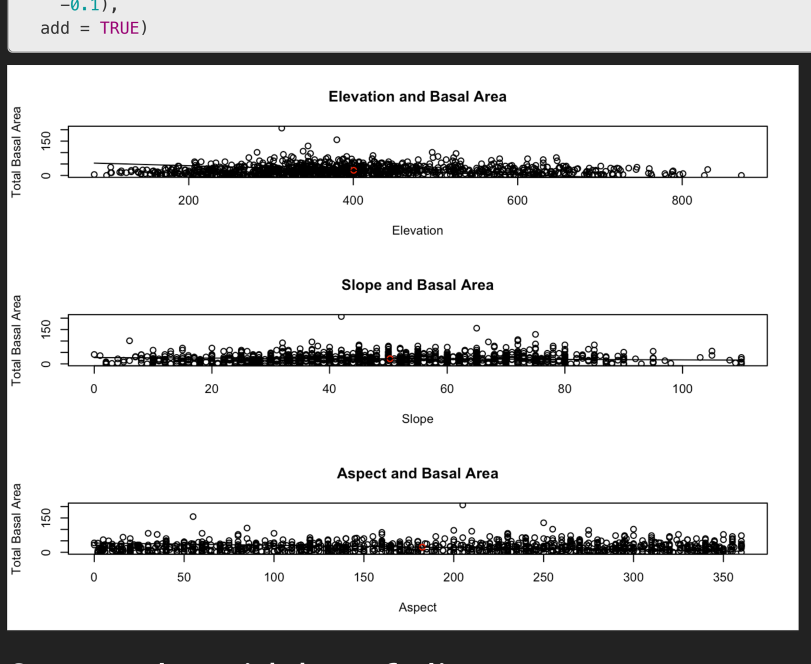
```
data_center_x = mean(dat_hab$aspect)
data_center_y = mean(dat_hab$ba.tot)
c(data_center_x, data_center_y)
```

```
## [1] 182.49140  22.42639
```

```
points(x = data_center_x, y = data_center_y, col = "red")
line_point_slope = function(x, x1, y1, slope)
{
  get_y_intercept =
    function(x1, y1, slope)
      return(-(x1 * slope) + y1)

  linear =
    function(x, yint, slope)
      return(yint + x * slope)

return(linear(x, get_y_intercept(x1, y1, slope), slope))
}
curve(
  line_point_slope(
    x,
    data_center_x,
    data_center_y,
    -0.1),
  add = TRUE)
```




# Scatterplot with best fit line seperate

```
(plot(x = dat_hab$elev, y = dat_hab$ba.tot, main = "Elevation and Basal Area", ylab = "Total Basal Area", xlab = "Elevation"))
```

```
## NULL
```

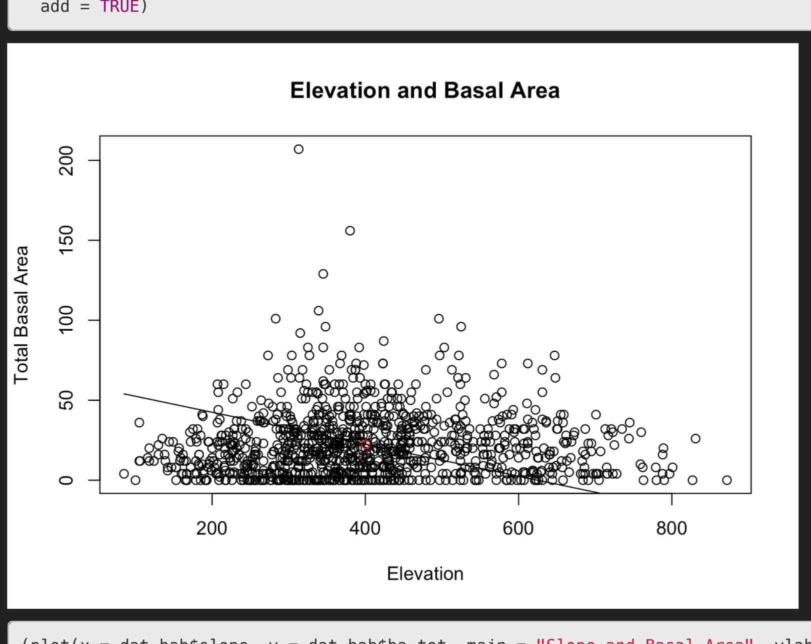
```
data_center_x = mean(dat_hab$elev)
data_center_y = mean(dat_hab$ba.tot)
c(data_center_x, data_center_y)
```

```
## [1] 400.52008  22.42639
```

```
points(x = data_center_x, y = data_center_y, col = "red")
line_point_slope = function(x, x1, y1, slope)
{
  get_y_intercept =
    function(x1, y1, slope)
      return(-(x1 * slope) + y1)

  linear =
    function(x, yint, slope)
      return(yint + x * slope)

return(linear(x, get_y_intercept(x1, y1, slope), slope))
}
curve(
  line_point_slope(
    x,
    data_center_x,
    data_center_y,
    -0.1),
  add = TRUE)
```




```
(plot(x = dat_hab$slope, y = dat_hab$ba.tot, main = "Slope and Basal Area", ylab = "Total Basal Area", xlab = "Slope"))
```

```
## NULL
```

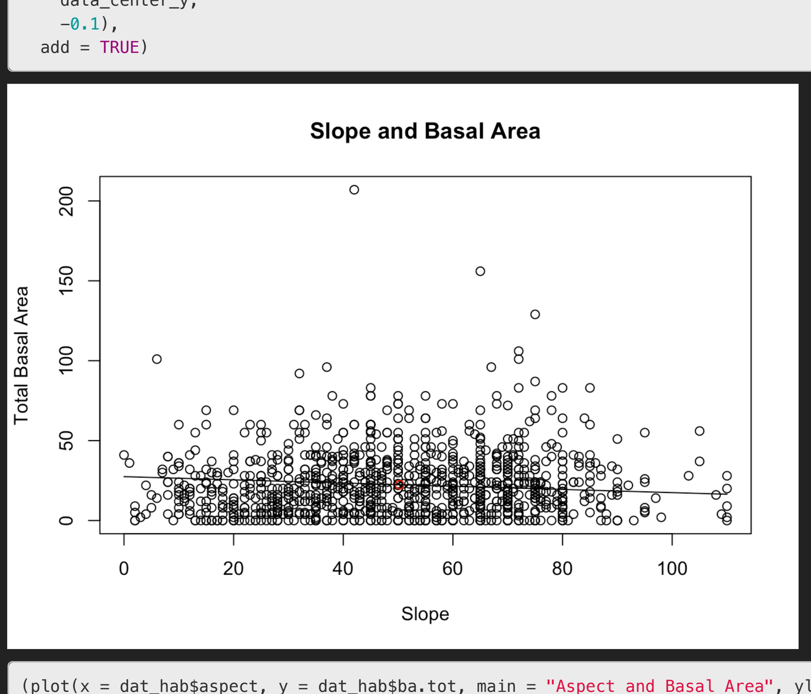
```
data_center_x = mean(dat_hab$slope)
data_center_y = mean(dat_hab$ba.tot)
c(data_center_x, data_center_y)
```

```
## [1] 50.26864 22.42639
```

```
points(x = data_center_x, y = data_center_y, col = "red")
line_point_slope = function(x, x1, y1, slope)
{
  get_y_intercept =
    function(x1, y1, slope)
      return(-(x1 * slope) + y1)

  linear =
    function(x, yint, slope)
      return(yint + x * slope)

return(linear(x, get_y_intercept(x1, y1, slope), slope))
}
curve(
  line_point_slope(
    x,
    data_center_x,
    data_center_y,
    -0.1),
  add = TRUE)
```




```
(plot(x = dat_hab$aspect, y = dat_hab$ba.tot, main = "Aspect and Basal Area", ylab = "Total Basal Area", xlab = "Aspect"))
```

```
## NULL
```

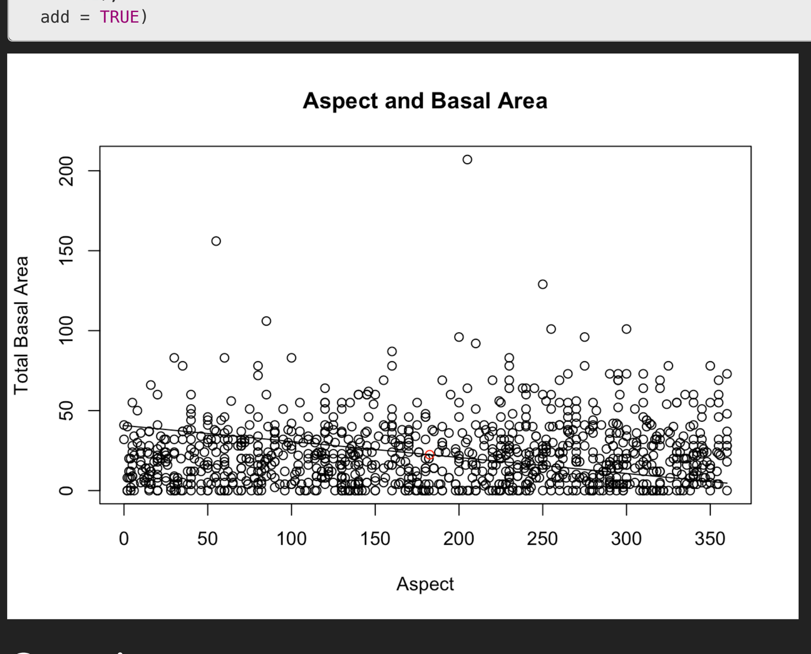
```
data_center_x = mean(dat_hab$aspect)
data_center_y = mean(dat_hab$ba.tot)
c(data_center_x, data_center_y)
```

```
## [1] 182.49140  22.42639
```

```
points(x = data_center_x, y = data_center_y, col = "red")
line_point_slope = function(x, x1, y1, slope)
{
  get_y_intercept =
    function(x1, y1, slope)
      return(-(x1 * slope) + y1)

  linear =
    function(x, yint, slope)
      return(yint + x * slope)

return(linear(x, get_y_intercept(x1, y1, slope), slope))
}
curve(
  line_point_slope(
    x,
    data_center_x,
    data_center_y,
    -0.1),
  add = TRUE)
```




# Questions

2. Elevations ranging from 300 to 450 m had the highest frequencies for sampling sites. There were close to no sites with elevations from 50 to 100 m, as well as no elevations past 800 m. Many of the sites had elevations from 200 to 500 meters. There were much more low elevation sampling sites than high elevation sampling sites. The shape of the elevation histogram leans much closer to the left side, while the right side gets much more sparse.
3. The unit of this slope data set is percent slope. But, the slope is ratio over rise, so a 100% slope really means a 45 degree slope.
4. Most sample sites were not flat. Only a small amount of sites had an slope of 0 to 10%, meaning there were almost no flat sites. Most sample sites had some amount of slope. The majority of sites were between a slope of 20 and 80%, giving us a pretty even mixture of steep and shallow slopes. The histogram with the variables frequency vs slope has a relatively normal distribution. Both ends of the x-axis (0 and 100) are on the lower frequencies. Whereas slopes ranging from 20 to 80 have higher frequencies. Slopes from 40 to 50 have the highest frequency within the histogram.
5. Aspect is the topological appearance of an object. Appearance to the eye when seen from a specific view. In this dataset, aspect is in units of degrees. So is the histogram comparing aspect and frequency, an aspect of 180 is really 180 degrees, which translates to south-facing.
6. The shape of the aspect histogram is pretty evenly distributed. The histogram blocks are very similar in frequency, except for 360 degrees which is completely north-facing. All of the sites were evenly distributed between every possible direction (SE, SW, NE, NW, etc.) The smallest frequency ofs sites were in the observed direction Northwest.
7. For my scatterplot of basal area vs elevation, I noticed a denser amount of plot points on the left side of the graph. At elevations less than 500 meters, there were densely packed data points around 25 m^2/ha total basal area. Higher amounts of total basal area showed points evenly distributed throughout the elevations, but were more sparse. The association line is linear, but it is not a good fit for the data because it does not include the points at the far right of the plot with higher elevation. For basal area vs slope, the points are evenly distributed along the whole plot, but it is more densely packed towards the center of the plot (slope between 20 and 80). The best fit line is linear and is a good fit for the data because all of the points are represented and there is an even amount of points on each side of the line. The point for best fit is almost directly in the center. For basal area vs aspect, the data points were very evenly distributed. Even the denser (darker) groups of point are evenly spread across the scatterplot. The best fit line is linear, and is a very good fit for the data as it reaches all along the x-axis, from and aspect of 0 to 360, or North to South.