**Statistics for Soil Survey Part 1 2018 Final Project**

Kim Cumella

**Objectives**

The purpose of this project is to present a model of soils data between two different plant communities in Lincoln County, WY. One group represents sites dominated by early sagebrush with low productivity. The other group includes sites with relatively high productivity and plant communities composed of Wyoming big sagebrush, mountain big sagebrush and basin big sagebrush (will refer to as larger sagebrush species). From observations in the field, it appears that the difference in the two plant communities is dictated by the clay content in the soil profile. The early sagebrush sites tend to have high clay near to the surface and the larger sagebrush sites have a relatively low clay percentage throughout. My goal is to show statistical analysis and visual representation to further support and communicate our ecological site assignments for these two different sites. This can be used to present findings to other offices and collaborate on ecological site concepts/surveying techniques. It could also be a useful model in digital soil mapping, when coupled with the unique spectral signatures of the different species.

R software will be used to compute statistics and visually present soils data between the two different groups of sites. Methods for statistical analysis in R will be referenced from the training webpages.

<http://ncss-tech.github.io/stats_for_soil_survey/chapters/>

<https://github.com/ncss-tech/stats_for_soil_survey>

**Results**

**Loading Pedon Data from NASIS**

The first step was to load one group of pedons (the early sagebrush sites being one group and the larger sagebrush sites the other group) in NASIS by using the Lab POINT - Pedon/Site/Transects by User Pedon ID (2100 max) Query. This allowed me to list each pedon in the group of interest so I wouldn’t have to load one pedon at a time or sort through a large volume of pedons in the selected set. If more efficient, it would be useful to learn how to specify these pedons directly in R. After the statistics were run on the first group of pedons, I repeated the process with the other group as the new selected set. To load the data into R, I used examples from the chapter on tabular data provided in the training.

# load required libraries

library(soilDB)

library(aqp)

# load data from a NASIS selected set

f <- fetchNASIS(from = 'pedons')

# what kind of object is this?

class(f)

# how many pedons

length(f)

**Statistical analysis of soil properties**

First, the horizons and sites were extracted and the command view(h) was used to view the

names of the different fields in the horizon table. Observing the field names from the site and

horizon tables was useful to adjust the commands from code in the guide (ex. genhz was

changed to hzname).

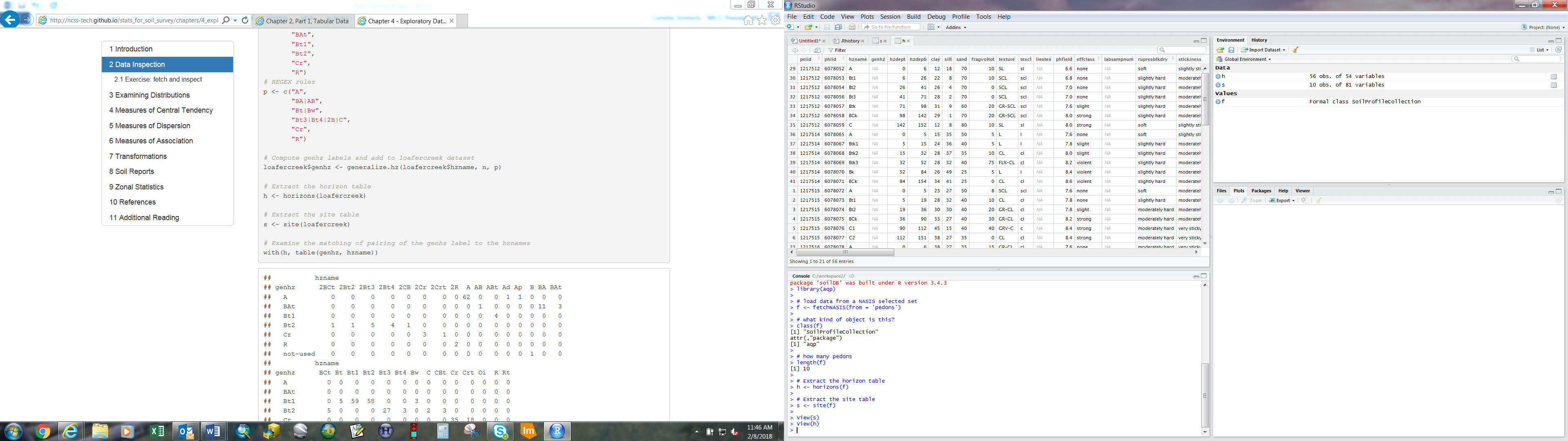
# Extract the horizon table

h <- horizons(f)

# Extract the site table

s <- site(f)

view(h)

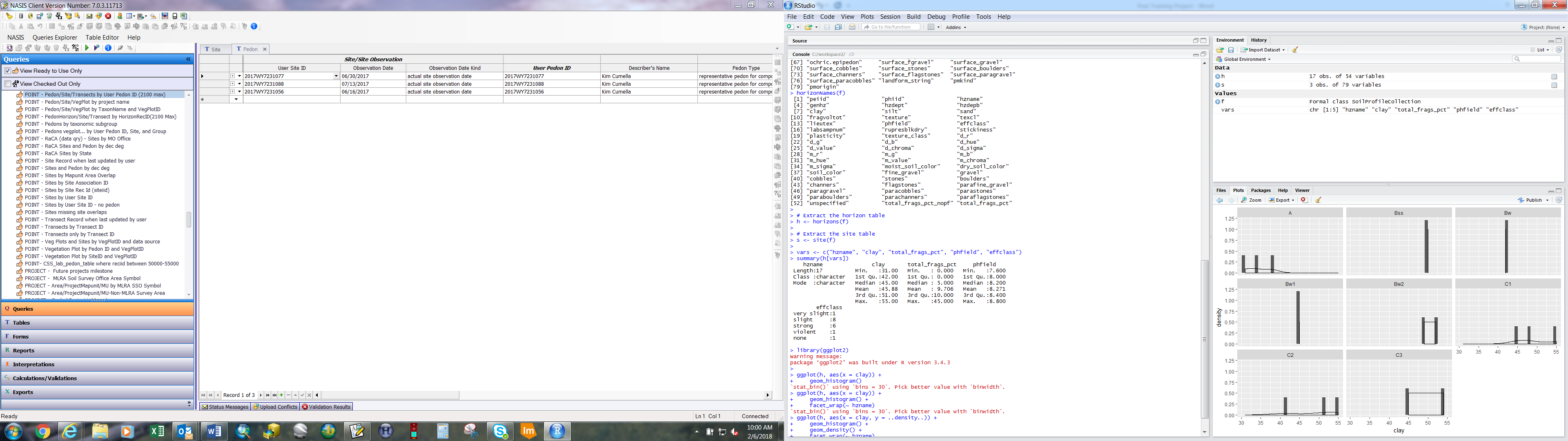


Once the horizon table was loaded, several soil properties were observed across the groups of pedons. Clay percentage was assumed to be the distinguishing factor but pH and fragment percentage were checked just to be sure. The main difference in characteristics between the sites was indeed clay percentage. The fragment percent maximum had a wide gap between the two but that was caused by an outlier in the data. I decided to disregard further analysis on pH and fragment percent since the differences weren’t significant across the two sites.

vars <- c("hzname", "clay", "total\_frags\_pct", "phfield")

summary(h[vars])

**Early Sagebrush Sites**



**Larger Sagebrush Sites**

hzname clay total\_frags\_pct phfield

Length:39 Min. :12.00 Min. : 0.000 Min. :6.600

Class :character 1st Qu.:22.50 1st Qu.: 0.000 1st Qu.:7.600

Mode :character Median :26.00 Median : 5.000 Median :8.000

Mean :25.44 Mean : 8.795 Mean :7.856

3rd Qu.:28.00 3rd Qu.:10.000 3rd Qu.:8.300

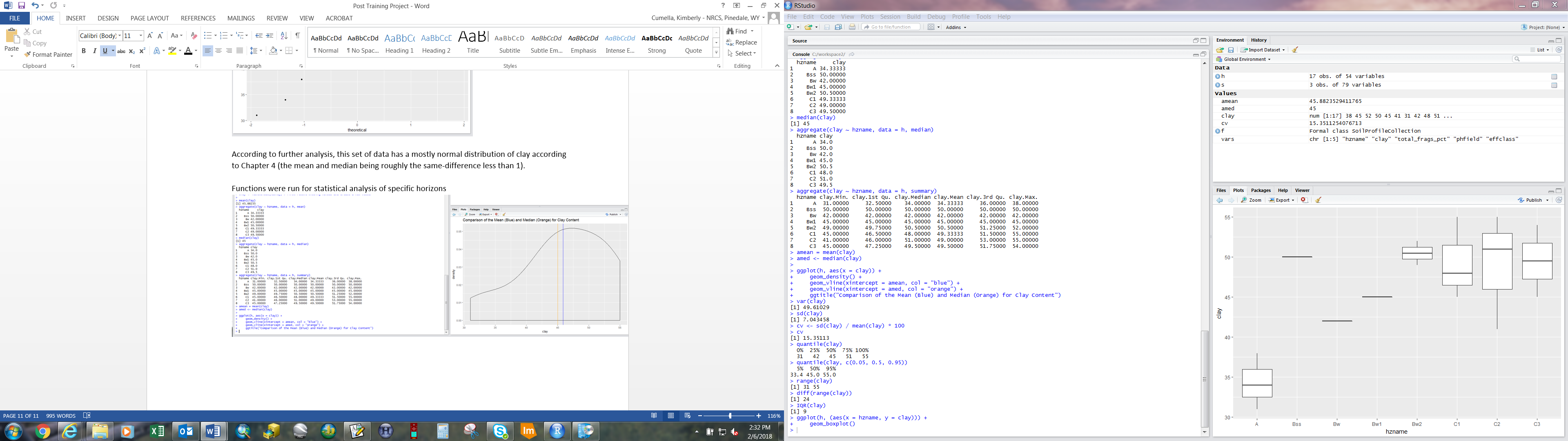
Max. :45.00 Max. :75.000 Max. :8.600

Boxplots were created to visualize the differences in clay distribution by graphically displaying the minimum, maximum, median, quartiles and outliers of the data. This method was referenced from the chapter on exploratory analysis in the training.

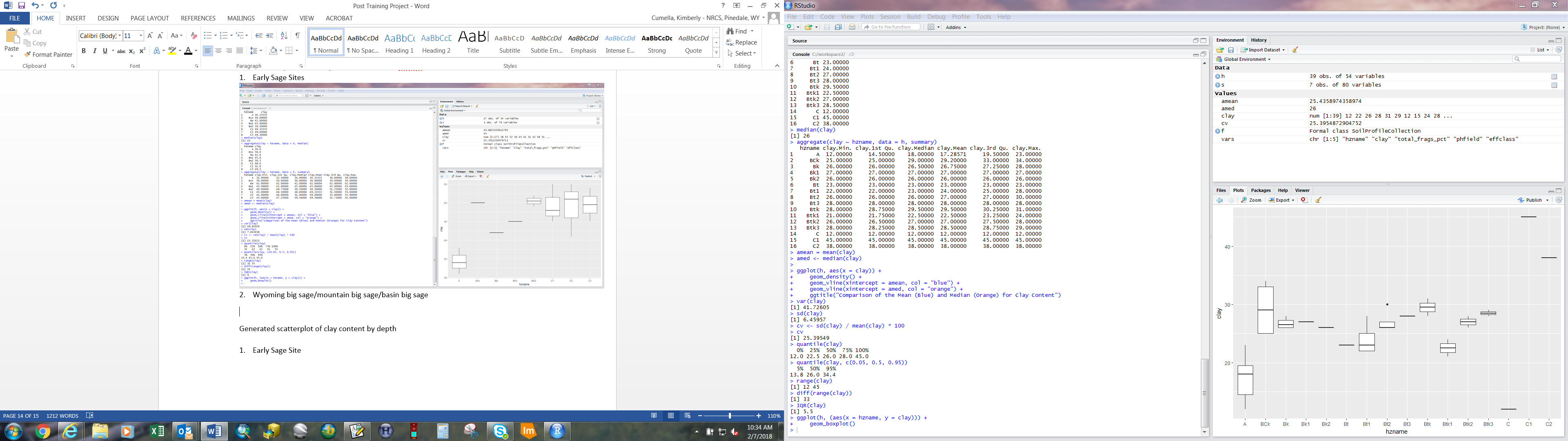
ggplot(h, (aes(x = hzname, y = clay))) +

geom\_boxplot()

**Early Sagebrush Sites**



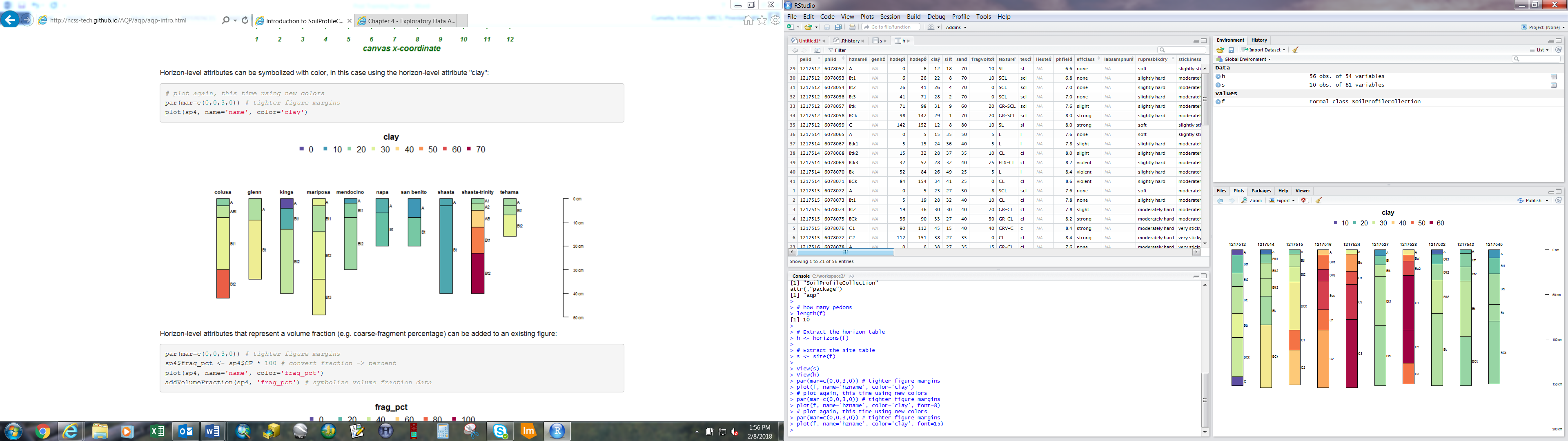
**Larger Sagebrush Sites**



The clay distribution in the profile was further visualized using methods from the SoilProfileCollection Object Introduction section of the guide.

par(mar=c(0,0,3,0)) # tighter figure margins

plot(f, name='hzname', color='clay')



12177516, 1217524 and 1217528 represent the early sage dominated sites. The other pedons are found at sites dominated by the larger sagebrush species. In the future it could be helpful to learn methods for adding brackets for grouping the different pedons based on site type.

**Plotting Location of Pedons**

To present the geographic location of the pedons, examples from the chapter on tabular data were used.

# load libraries

library(sp)

library(maps)

pedon.locations <- site(f)[, c('site\_id', 'x', 'y')]

coordinates(pedon.locations) <- ~ x + y

proj4string(pedon.locations) <- '+proj=longlat +datum=WGS84'

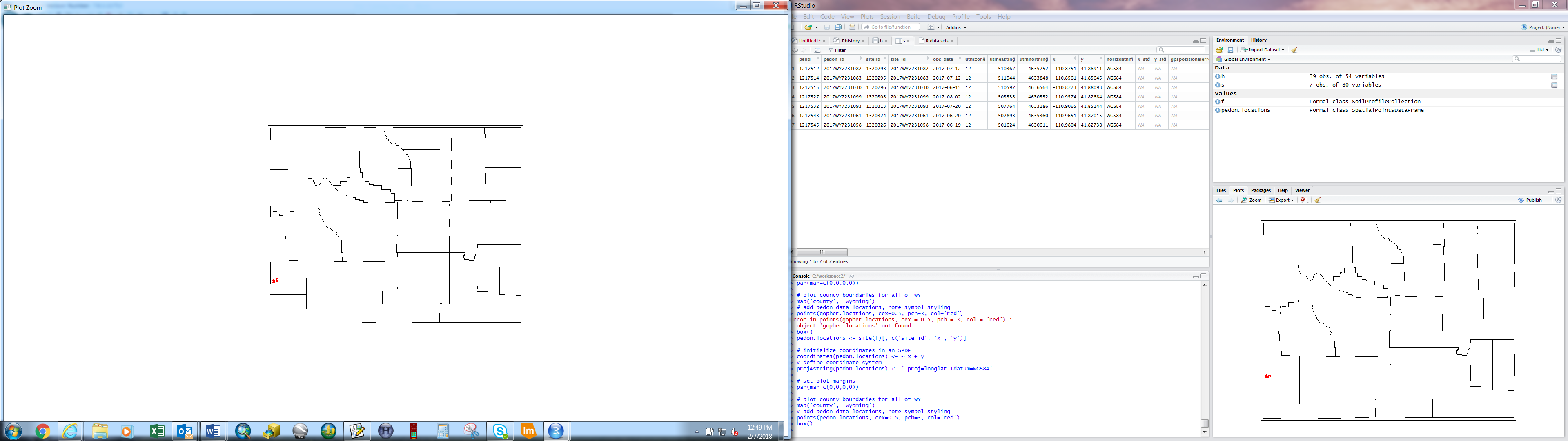
par(mar=c(0,0,0,0))

map('county', 'wyoming')

# add pedon data locations, note symbol styling

points(pedon.locations, cex=0.5, pch=3, col='red')

box()



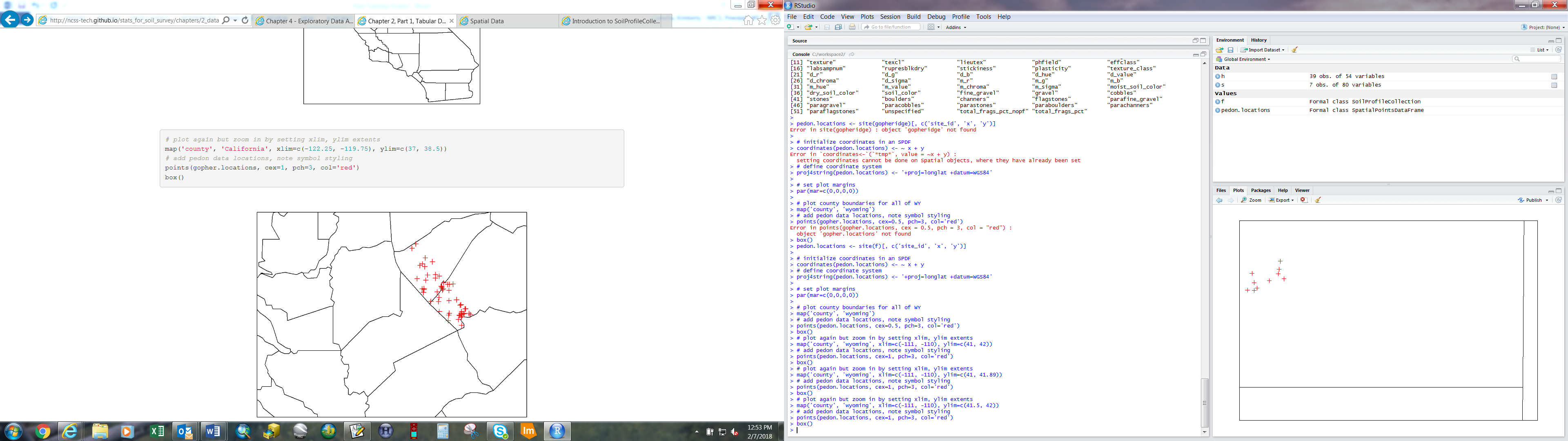
Xlim, ylim extents were set in order to zoom into the area of interest.

map('county', 'wyoming', xlim=c(-111, -110), ylim=c(41.5, 42))

# add pedon data locations, note symbol styling

points(pedon.locations, cex=1, pch=3, col='red')

box()



Lincoln County

**Conclusion**

The sites dominated by early sage with low productivity were confirmed to have a relatively high clay content throughout the soil profile. Initially I decided to check for differences in pH and fragment content in addition to clay percentage. The statistics, aside from an outlier in fragment percentage, didn’t show a significant enough difference for further examination. The analysis will help our office to survey more efficiently by solidifying the understanding that these breaks in vegetation are key in sampling distribution. When collaborating with other survey offices, it will help provide visualization of our mapping concepts. The results could also be used as a reference for potential digital soil mapping projects.