Assignment1

Margaret Walker

January 13, 2016

R Basics

Homework 1- Margaret Walker

1. What are the names of the columns in the dataset?

```
tgpp <- read.csv("http://dmcglinn.github.io/quant_methods/data/tgpp.csv", header = T)
names(tgpp)
```

First, I read in the data file from the website and named it tgpp. Then I used the function names to get the names of all the columns in the data set. The column names are printed below:

```
##  [1] "plot"      "year"      "record_id" "corner"    "scale"
##  [6] "richness"  "easting"   "northing"  "slope"     "ph"
## [11] "yrsslb"
```

2. How many rows and columns does the dataset have?

```
dim(tgpp)
```

In order to figure this out I used the function dim(). This function told me the dataset has 4080 rows and 11 columns. See code below:

```
## [1] 4080   11
```

3. What kind of object is each column?

```
sapply(tgpp, class)
```

For this question I used the sapply function in order to on the dataset. This applies the class function to each column in the dataset. See work below:

```
##      plot      year record_id    corner     scale  richness   easting
## "integer" "integer" "integer" "integer" "numeric" "integer" "integer"
##  northing     slope        ph    yrsslb
## "integer" "integer" "numeric" "numeric"
```

4. What are the values of the datafile for rows 1, 5, and 8 at columns 3, 7, and 10?

```
tgpp[1, 3]
```

To figure this out we will just subset the data. That is, we will use the forumula tgpp[row, column] for each of the three combinations and R will return the value. See code below:

```
## [1] 187
```

```
tgpp[5, 7]
```

```
## [1] 727000
```

```
tgpp[8, 10]
```

```
## [1] 6.9
```

5. Plot the relationship between scale and richess.

```
pdf("./richess_scale.pdf")
richavg <- tapply(tgpp$richness, tgpp$scale, mean)
sdv <- tapply(tgpp$richness, tgpp$scale, sd)
scales <- as.numeric(names(richavg))
plot(richavg ~ scales, xlab = expression(Scale(m^2)), ylab = "Richness", col = "blue",
    ylim = c(0, 90))
arrows(scales, richavg - sdv, scales, richavg + sdv, length = 0.05, angle = 90,
    code = 3, col = "blue")
dev.off()
```

I first found the mean of richness for each of the different values of scale. Then I also created a vector for the standard deviation of each of those means. I plotted the richness averages versus scale and used the arrow function to add standard deviation bars. This gives us both the uncertainty for each value as well as the overall trend. However, it looks like this relationship is non linear and transforming the data using a log transformation may work better. See code below:

```
## pdf
##   2
```

6. What happens to your plot when you set the log argument to 'xy'?

```
pdf("./logrichess_scale.pdf")
richavg <- tapply(tgpp$richness, tgpp$scale, mean)
scales <- as.numeric(names(richavg))
plot(log10(richavg) ~ log10(scales), lwd = 2, col = "blue", type = "o", xlab = expression(log[10] *
    scale(m^2)), ylab = expression(log[10] * richness), xlim = c(-2, 2), ylim = c(0,
    2))
dev.off()
```

Based on the last graph it looks like the relationship between scale and richness is not linear. Therefore, log transforming both variables could help us decipher the relationship between the two. When we create the same graph as above except taking the log of both x and y values we see a more linear relationship. That is, as scale increases so does richness.

```
## pdf
##   2
```