R Basics Homework 1

Caroline Tribble

Manual download and import of the data:

```
tgpp <- read.csv('https://raw.githubusercontent.com/dmcglinn/quant_methods/gh-pages/data/tgpp.csv', header = TRUE
)
```

This dataset represents the vascular plant species richness that was collected from the Tallgrass Prairie Preserve from 10 x 10 m quadrats. Species richness is simply the number of species that occur within a quadrat.

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

```
colnames(tgpp)
```

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

The column names are plot, year, record_id, corner, scale, richness, easting, northing, slope, pH and yrsslb.

2. How many rows and columns does this data file have?

```
dim(tgpp)
```

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

There are 4,080 rows and 11 columns of data contained in this file.

3. What kind of object is each data column? Hint: checkout the function sapply().

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

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

Each column contains either integer or numeric values.

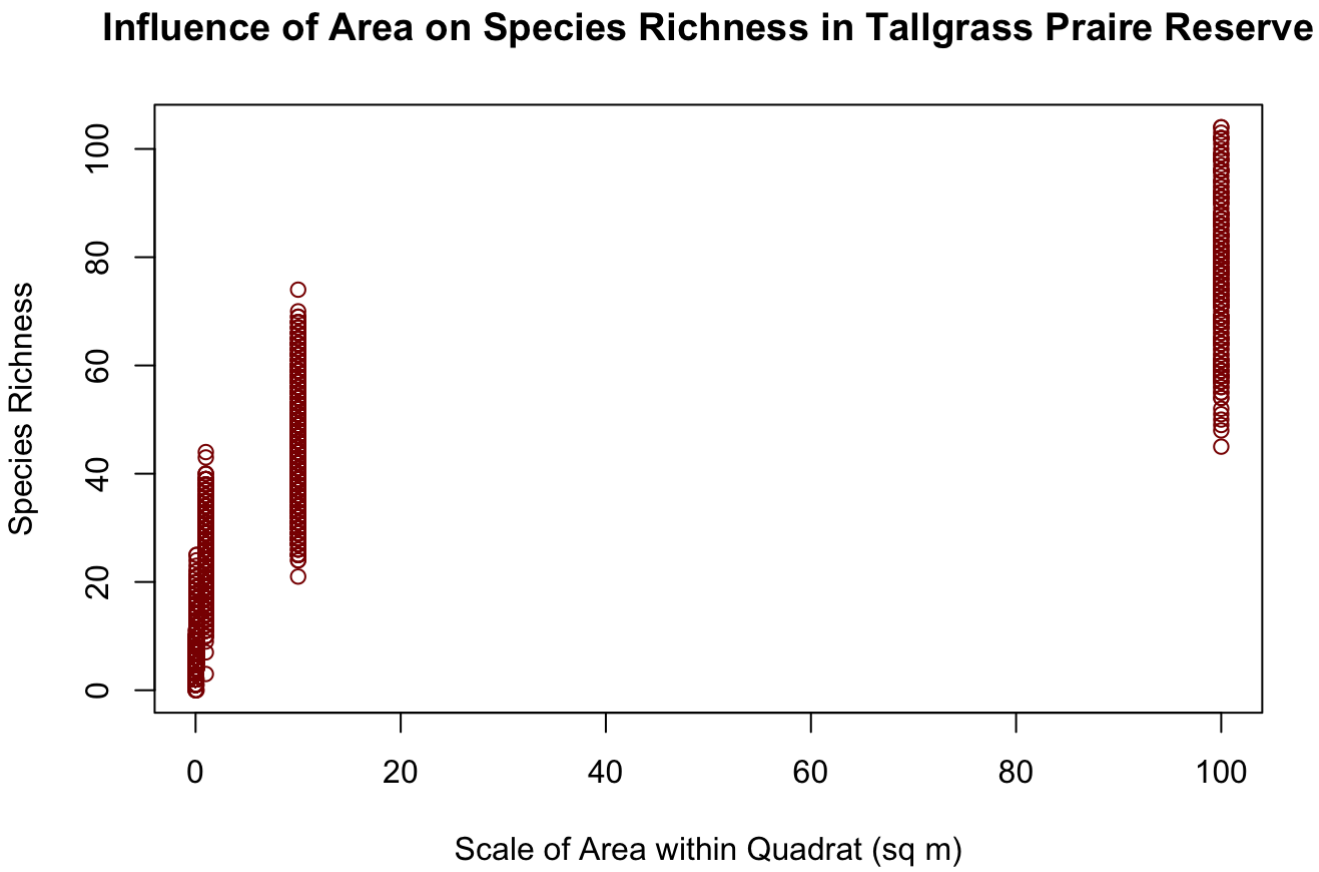
4. What are the values of the datafile for rows 1, 5, and 8 at columns 3, 7, and 10

```
tgpp[c(1,5,8),c(3,7,10)]
```

```
##   record_id easting  ph
## 1       187  727000 6.9
## 5       191  727000 6.9
## 8       194  727000 6.9
```

5. Create a pdf of the relationship between the variables “scale” and “richness”. Scale is the area in square meters of the quadrat in which richness was recorded. Be sure to label your axes clearly, and choose a color you find pleasing for the points. To get a list of available stock colors use the function colors(). Also see this link:

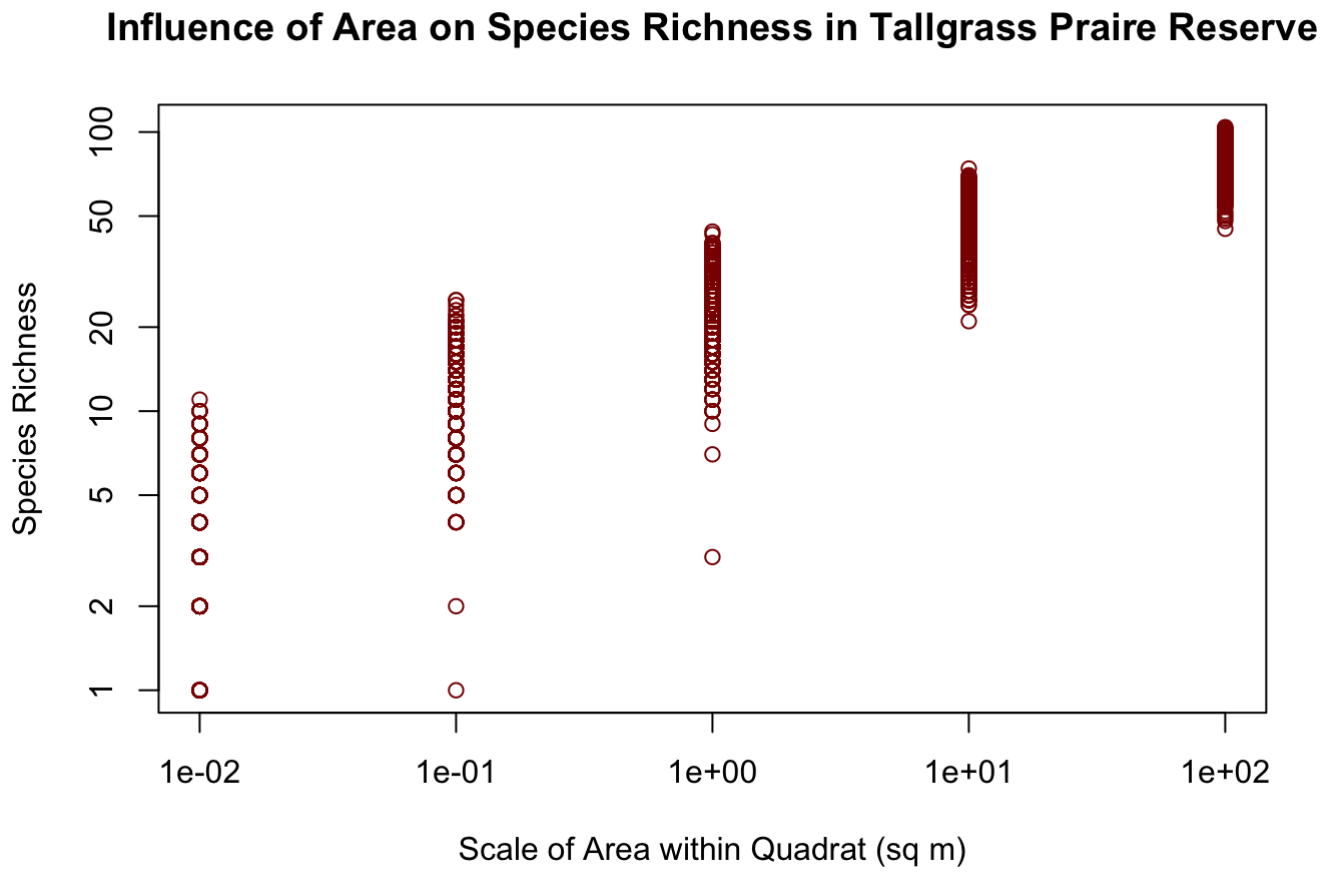
```
plot(tgpp$scale,tgpp$richness, type = 'p', col = 'dark red', xlab ='Scale of Area within Quadrat (sq m)', ylab =
'Species Richness',main = 'Influence of Area on Species Richness in Tallgrass Praire Reserve')
```



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

```
plot(tgpp$scale,tgpp$richness, type = 'p', col = 'dark red', xlab ='Scale of Area within Quadrat (sq m)', ylab =
'Species Richness',main = 'Influence of Area on Species Richness in Tallgrass Praire Reserve',log = 'xy')
```

```
## Warning in xy.coords(x, y, xlabel, ylabel, log): 4 y values <= 0 omitted
## from logarithmic plot
```



When log transforming the data you can see there is a clear relationship between your variables which was not as clearly visible before you transformed it. Species richness increases with the scale of area and shows this concave down, increasing trend. You can also see as area of the quadrat increases there is less variation in species richness. Another thing to note is when you log transform your data you cannot transform zeros or negative values and these are omitted from the plot. If you have a lot of zeros in your data one way to deal with this is using a log+1 transformation instead.

For creating a pdf of the plot:

```
pdf('/Users/caroline/Documents/Statistical Programming Class/HW1_Areaandrichness')
```