Homework1

Kevin Mack

1/14/2016

Introduction to R: Basics

We are working with a dataset of the species richness of vascular plants within the Tallgrass Prairie Preserve (TgPP).

```r
dat1=read.csv('http://dmcglinn.github.io/quant_methods/data/tgpp.csv', header=TRUE)
summary(dat1)
```

```
##       plot            year        record_id        corner     
##  Min.   :205.0   Min.   :1998   Min.   : 187   Min.   :1.00  
##  1st Qu.:225.0   1st Qu.:2001   1st Qu.:1207   1st Qu.:1.75  
##  Median :281.0   Median :2004   Median :2226   Median :2.50  
##  Mean   :275.8   Mean   :2004   Mean   :2226   Mean   :2.50  
##  3rd Qu.:317.5   3rd Qu.:2006   3rd Qu.:3246   3rd Qu.:3.25  
##  Max.   :350.0   Max.   :2009   Max.   :4266   Max.   :4.00  
##                                                NA's   :240   
##      scale            richness         easting          northing      
##  Min.   :  0.010   Min.   :  0.00   Min.   :727000   Min.   :4069000  
##  1st Qu.:  0.100   1st Qu.:  7.00   1st Qu.:729750   1st Qu.:4074750  
##  Median :  1.000   Median : 18.00   Median :731000   Median :4078500  
##  Mean   :  8.496   Mean   : 24.38   Mean   :731550   Mean   :4078000  
##  3rd Qu.: 10.000   3rd Qu.: 37.00   3rd Qu.:734000   3rd Qu.:4080250  
##  Max.   :100.000   Max.   :104.00   Max.   :738000   Max.   :4086000  
##                                                                       
##      slope            ph            yrsslb      
##  Min.   :1.00   Min.   :5.500   Min.   : 0.150  
##  1st Qu.:2.00   1st Qu.:6.000   1st Qu.: 0.250  
##  Median :3.50   Median :6.200   Median : 1.240  
##  Mean   :3.55   Mean   :6.333   Mean   : 2.079  
##  3rd Qu.:5.00   3rd Qu.:6.700   3rd Qu.: 3.212  
##  Max.   :8.00   Max.   :7.600   Max.   :11.220  
## 
```

1. This names of the columns in the dataset are plot, year, record_id, corner, scale, richness, easting, northing, slope, ph, and yrsslb. These headers are listed when viewing the data, but can also be obtained by the names() function.

```r
names(dat1)
```

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

2. This datafile has 11 columns and 4,080 rows.

```r
dim(dat1)
```

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

3. The dataset as a whole is a dataframe object. Each of the columns is a different object. Some are integers and some are numerics.

```r
class(dat1)
```

```
## [1] "data.frame"
```

```r
sapply(dat1,class)
```

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

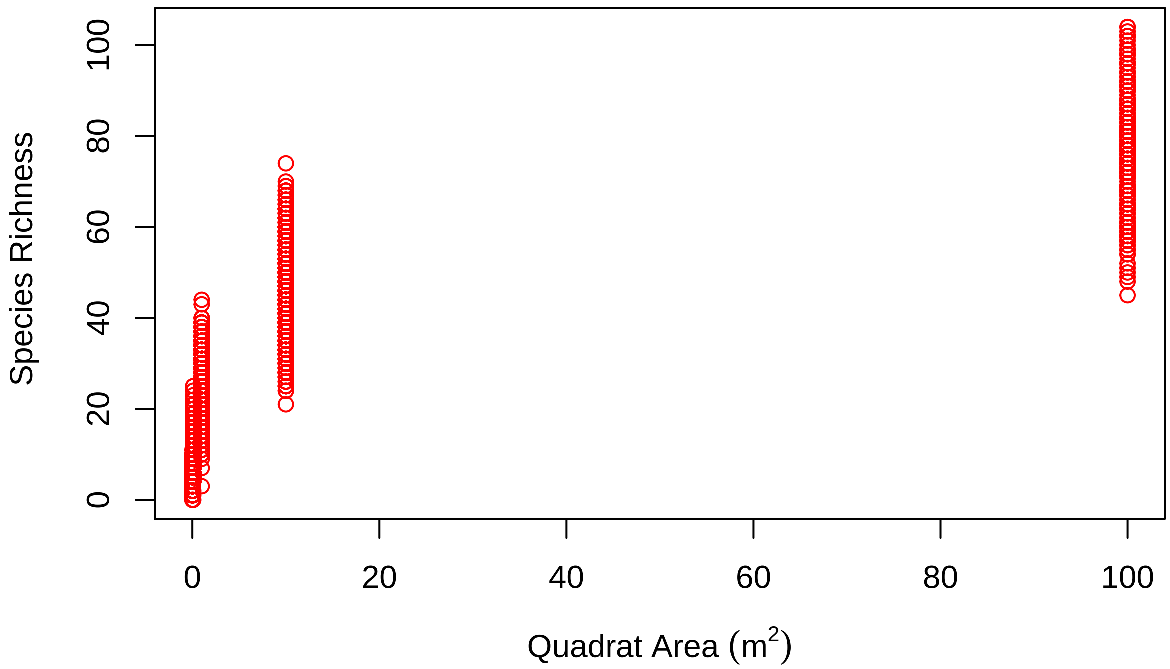
4. The values of rows 1, 5, and 8 at columns 3 (record_id), 7 (easting), and 10 (ph) are given below.

```r
dat1[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. Below I've plotted the relationship between scale and richness, or the Species Richness as a function of Quadrat Area. This plot is dificult to intrepret as species richness varies quite a bit for each given quadrat area.

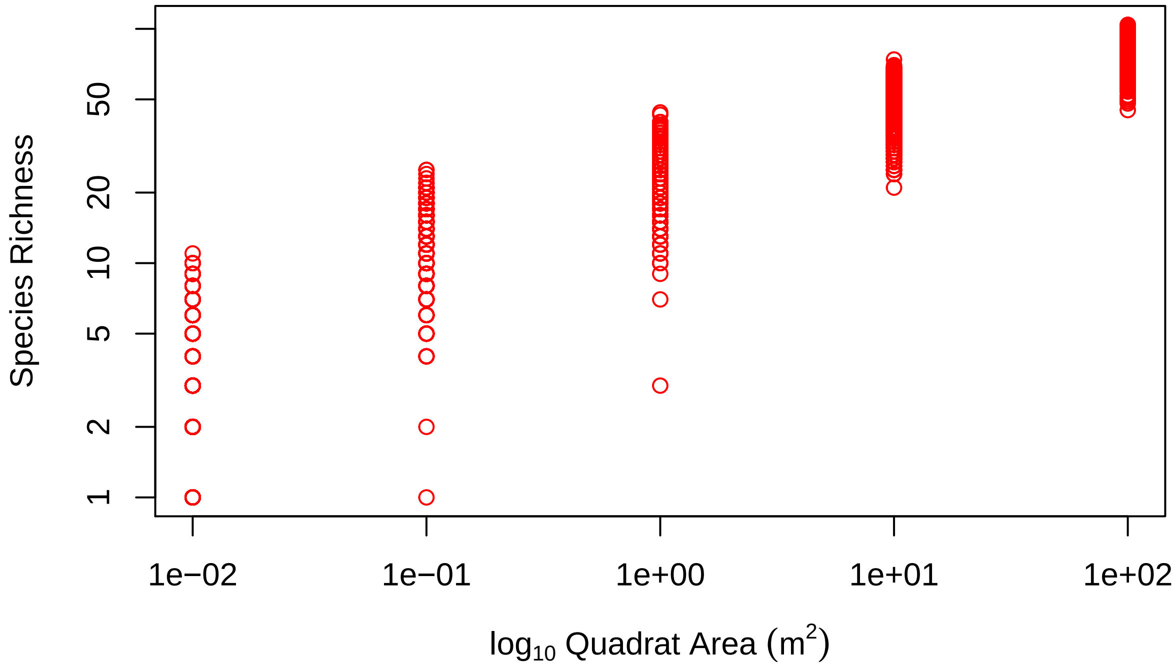
```r
plot(dat1$scale, dat1$richness,
     xlab=expression("Quadrat Area " (m^2)),
     ylab=" Species Richness",col="red")
```



log transforming the quadrat area (scale) makes a trend more apparent, but again intrepretation is hindered by variation in spices richness.

```
plot(dat1$scale, dat1$richness, log='xy',
     xlab=expression(log[10]*" Quadrat Area " (m^2)),
     ylab="Species Richness",col="red")
```

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

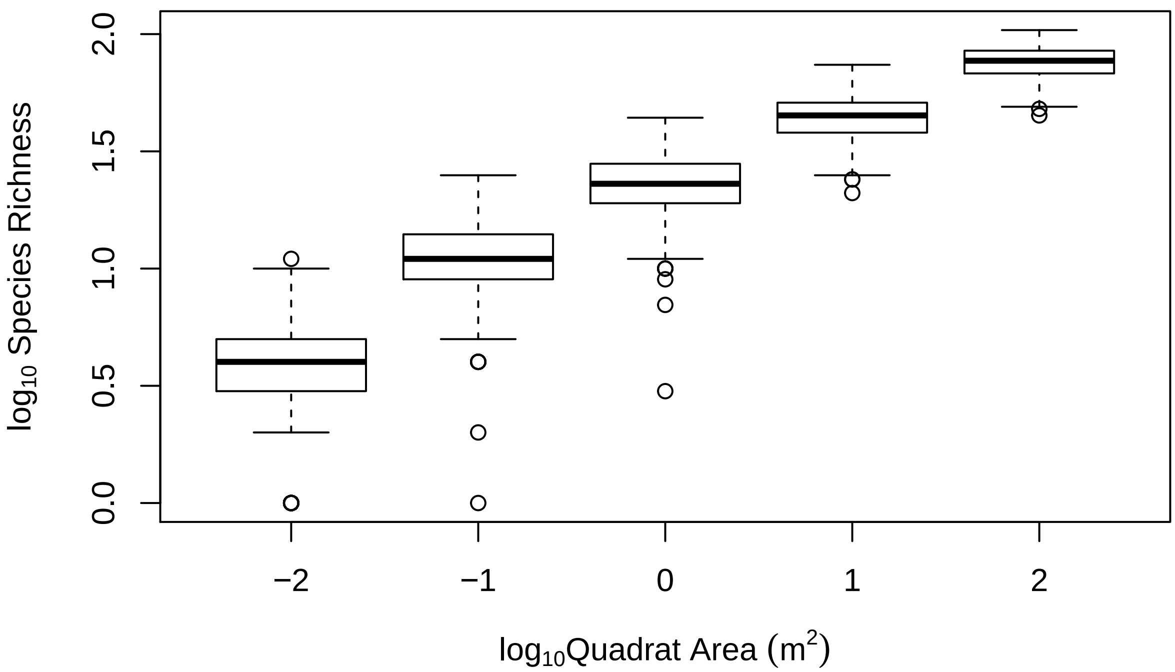


Based on the questions (and Dan's answers) on github, I generated the following two plots. The first is a boxplot with both variables log transformed. This shows the general trend that species richness increases with increasing quadrat area. Additionally, this plot visually shows the variability in species richness without overcrowding the figure. Personally, this is my prefered method of representing the data. The final figure is also log transformed, but represents the average species richness as a function of quadrat area. This figure best represents the trend, but does not capture variability in species richness.

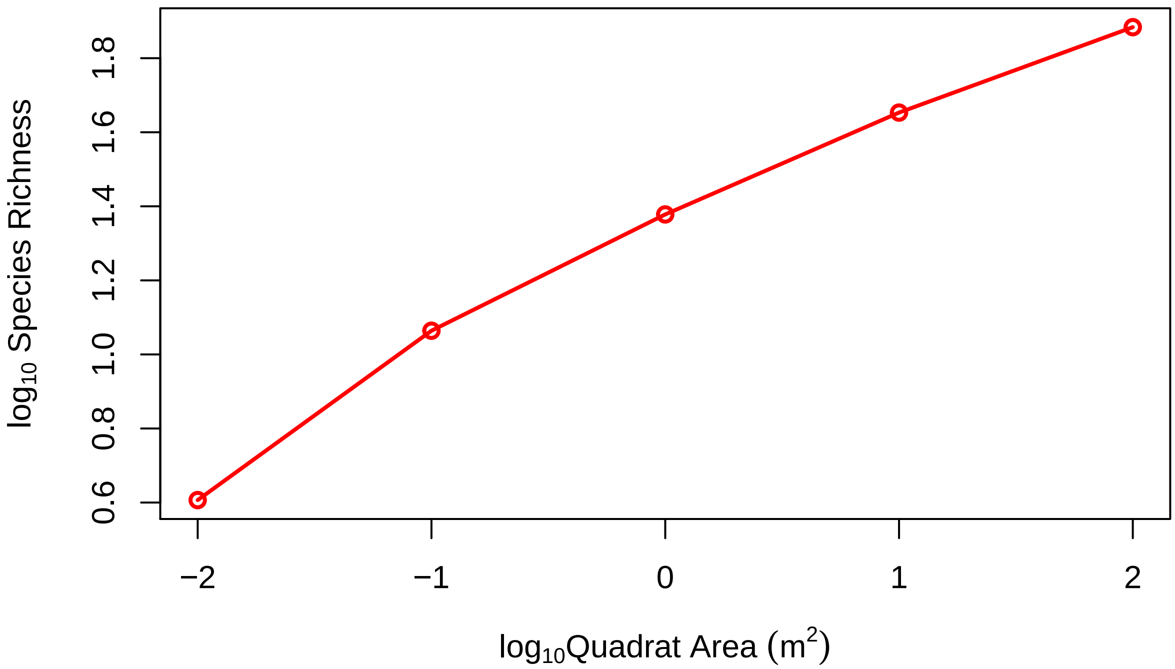
```
boxplot(log10(dat1$richness) ~ log10(dat1$scale),
        xlab=expression(log[10]*"Quadrat Area " (m^2)),
        ylab=expression(log[10]*" Species Richness"))
```

```
## Warning in bplt(at[i], wid = width[i], stats = z$stats[, i], out = z$out[z
## $group == : Outlier (-Inf) in boxplot 1 is not drawn
```

```
## Warning in bplt(at[i], wid = width[i], stats = z$stats[, i], out = z$out[z
## $group == : Outlier (-Inf) in boxplot 2 is not drawn
```




```
avg_richness = tapply(dat1$richness, dat1$scale, mean)
area = as.numeric(names(avg_richness))
plot(log10(area), log10(avg_richness), lwd=2, col='red', type="o",
     xlab=expression(log[10]*"Quadrat Area " (m^2)),
     ylab=expression(log[10]*" Species Richness"))
```


log10 Species Richness
0.6
0.8
1.0
1.2
1.4
1.6
1.8
−2
−1
0
1
2
log10Quadrat Area (m2)