KJPipho_PS_2

Krista Pipho

September 8, 2018

Question 1. Telomeres

```
# First we open the Telomeres file
telo <- read.csv("C:\\Users\\xenon\\Desktop\\R Studio 2018\\TELOMERES.csv")
telo
```

```
##     years telomere.length
## 1       1            1.63
## 2       1            1.24
## 3       1            1.33
## 4       2            1.50
## 5       2            1.42
## 6       2            1.36
## 7       2            1.32
## 8       3            1.47
## 9       2            1.24
## 10      4            1.51
## 11      4            1.31
## 12      5            1.36
## 13      5            1.34
## 14      3            0.99
## 15      4            1.03
## 16      4            0.84
## 17      5            0.94
## 18      5            1.03
## 19      5            1.14
## 20      6            1.17
## 21      6            1.23
## 22      6            1.25
## 23      6            1.31
## 24      6            1.34
## 25      7            1.36
## 26      6            1.22
## 27      8            1.32
## 28      8            1.28
## 29      8            1.26
## 30      7            1.18
## 31      7            1.03
## 32      8            1.10
## 33      8            1.13
## 34      8            0.98
## 35      9            1.24
## 36     10            0.85
## 37     10            1.05
## 38     12            1.15
## 39     12            1.14
```

After loading this file we can check the variable types.

```
# We can test them manually
is.numeric(telo$years)
```

```
## [1] TRUE
```

```
is.numeric(telo$telomere.length)
```

```
## [1] TRUE
```

```
# We can see that both variables are numeric, but must probe further to find the exact type
is.integer(telo$years)
```

```
## [1] TRUE
```

```
is.double(telo$years)
```

```
## [1] FALSE
```

```
is.integer(telo$telomere.length)
```

```
## [1] FALSE
```

```
is.double(telo$telomere.length)
```

```
## [1] TRUE
```

```
# Alternatively, we can have it return class types of variables directly
class(telo$years)
```

```
## [1] "integer"
```

```
class(telo$telomere.length)
```

```
## [1] "numeric"
```

From this we can see that class() is more efficient, but sometimes has limited precision (returning "numeric" rather than "double")

Once we know that both of our variables are numeric we can choose the best visualization strategy. For two quantitative variables, scatter plots are a good place to start.

```
plot(telo$years,telo$telomere.length, main="Child Illness Duration vs. Parental Telomere Length",
     xlab = "Years of Illness", ylab = "Telomere Length",
     col.lab='dark green',
     col.main = 'purple',
     cex.lab = 1.5,
     cex.main=1.5,
     lwd = 2)
abline(lm(telo$telomere.length~telo$years),
       col='red',
       lwd = 3,
       lty = "dotted")
```

Child Illness Duration vs. Parental Telomere Length
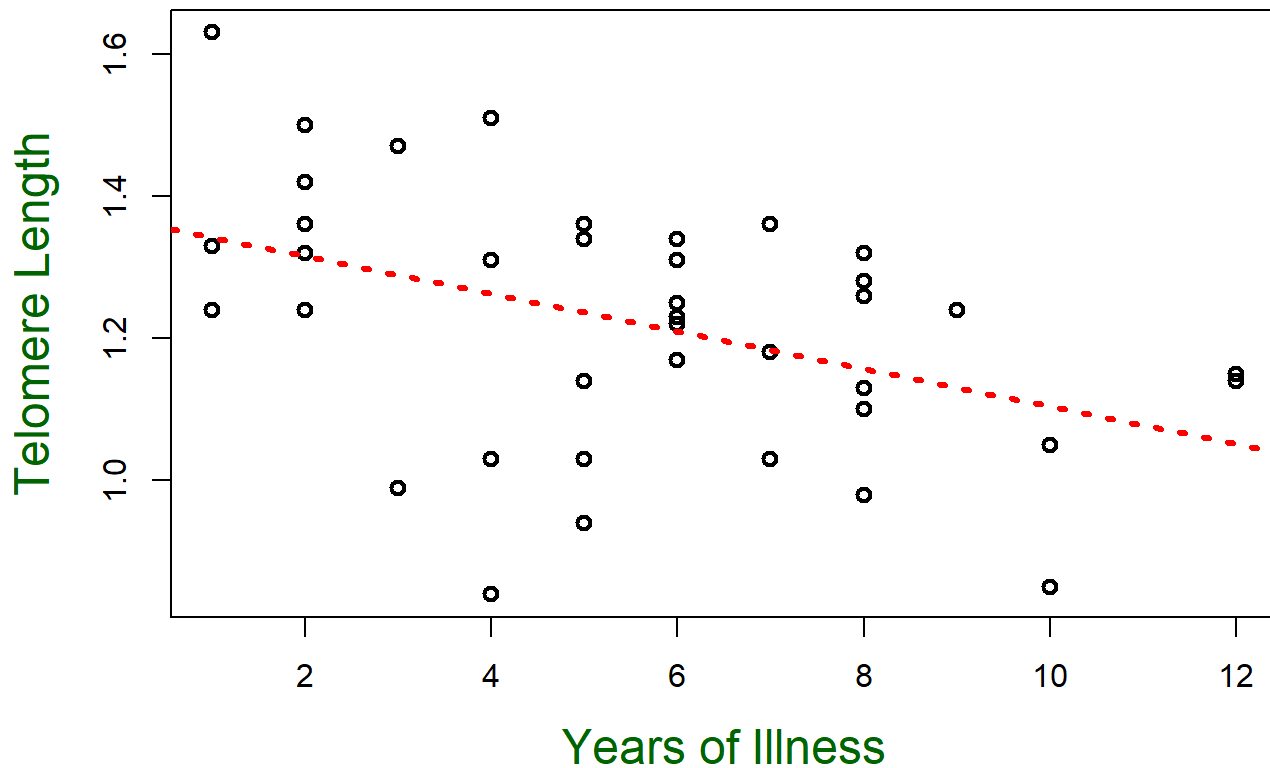
Using a trend line we can easy see a modest correlation between increased child illness duration and decreased parental telomere lengths.

Question 2. Firefly Spermatophores

Because the data from Whitlock and Schluter (edition 2), Chapter 2: Q19 (Page 58) is not already contained in a file we must manually input the values to a dataframe.

```
# We name the dataframe 'weights' because the data are weights in milligrams

weights <- c(0.047, 0.037, 0.041, 0.045, 0.039, 0.064, 0.064, 0.065, 0.079, 0.070, 0.066,
0.059, 0.075, 0.079, 0.090, 0.069, 0.066, 0.078, 0.066, 0.066, 0.055, 0.046,
0.056, 0.067, 0.075, 0.048, 0.077, 0.081, 0.066, 0.172, 0.080, 0.078, 0.048,
0.096, 0.097)

weights
```

```
##  [1] 0.047 0.037 0.041 0.045 0.039 0.064 0.064 0.065 0.079 0.070 0.066
## [12] 0.059 0.075 0.079 0.090 0.069 0.066 0.078 0.066 0.066 0.055 0.046
## [23] 0.056 0.067 0.075 0.048 0.077 0.081 0.066 0.172 0.080 0.078 0.048
## [34] 0.096 0.097
```

Here we have only one numerical variable. A histogram is an appropriate plot for displaying the frequency distribution of this data.

```
hist(weights, main="Weights of Spermatophores in Milligrams")
```


Weights of Spermatophores in Milligrams
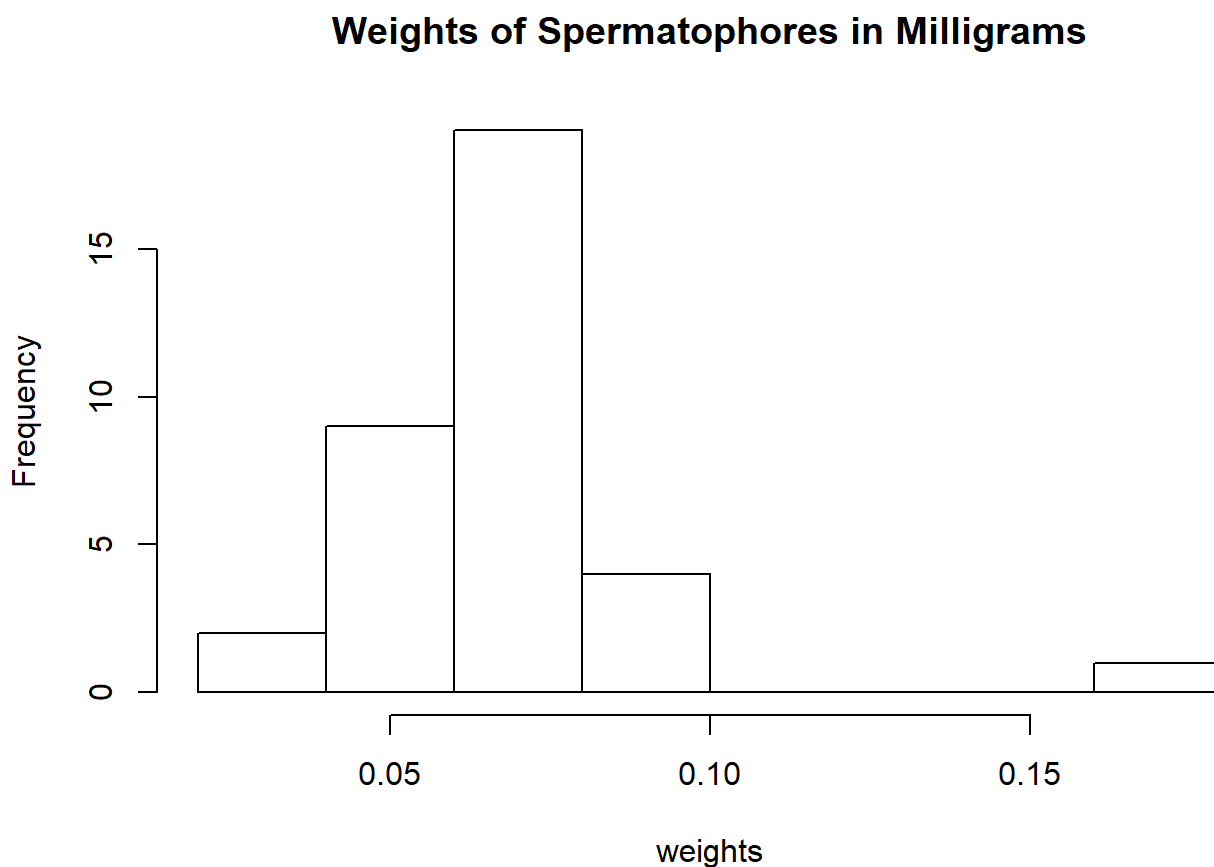
The distribution looks roughly normal, with most values falling between 0 and .1 mg. There is, however, one outlier of substantially higher mass.

Question 3. Using ggplot2

Remaking the scatter plot from question 1 with ggplot2

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

```
ggplot(data = telo) + geom_point(mapping = aes(x = telo$years,y = telo$telomere.length),color = "blue") +la
bs(title =" Child Illness Duration vs. Parental Telomere Length", x = "Years of Illness", y = "Telomere Len
gth") + geom_smooth(method = "lm",aes(x= telo$years,y=telo$telomere.length))
```

Child Illness Duration vs. Parental Telomere Length
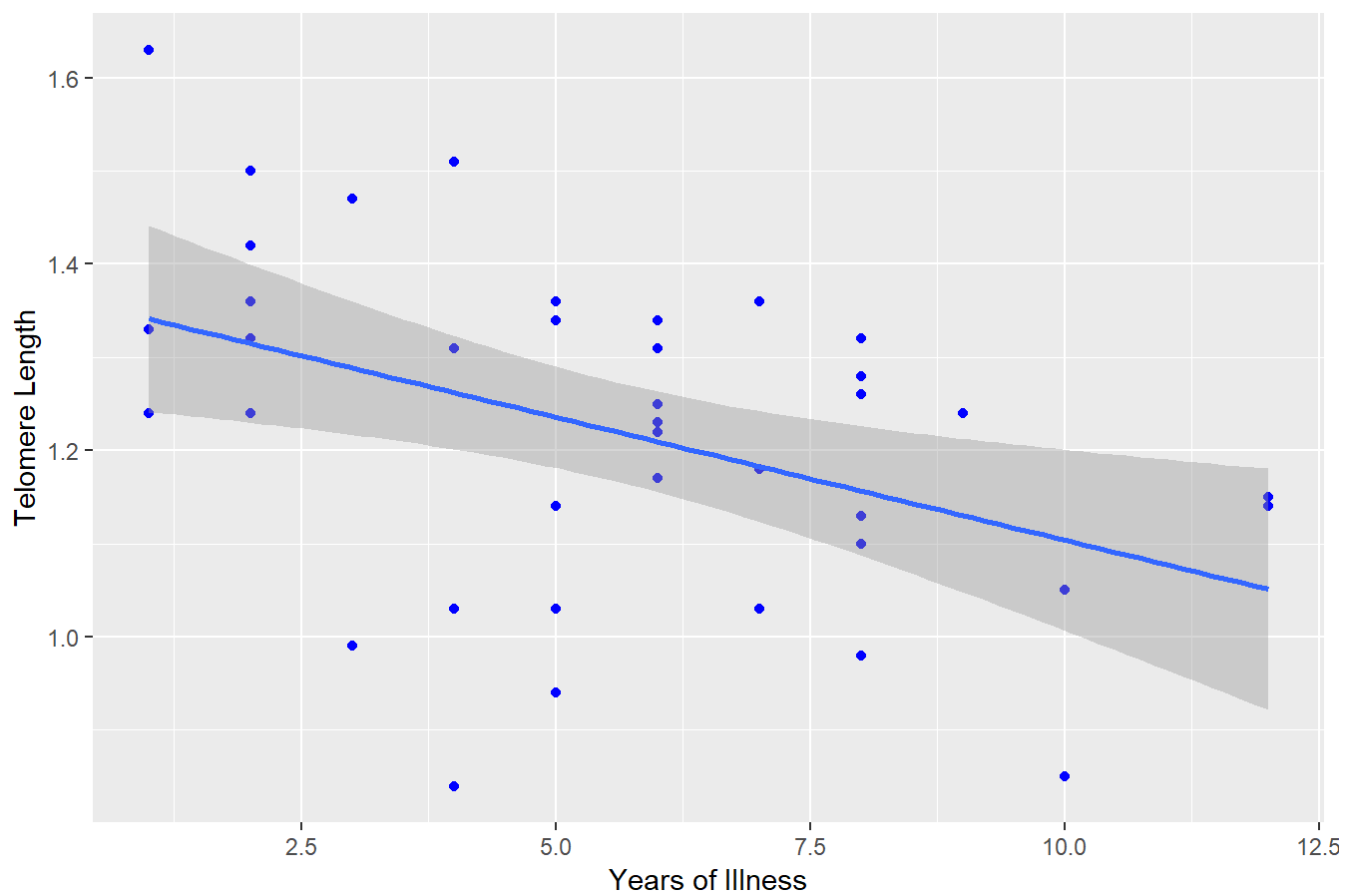
Remaking the histogram from question 2 with ggplot2

```
ggplot() + geom_histogram(binwidth=.01, mapping = aes(x = weights), color='blue')+labs(title="Weights of Sp
ermatophores in Milligrams")
```

Weights of Spermatophores in Milligrams