Abalone

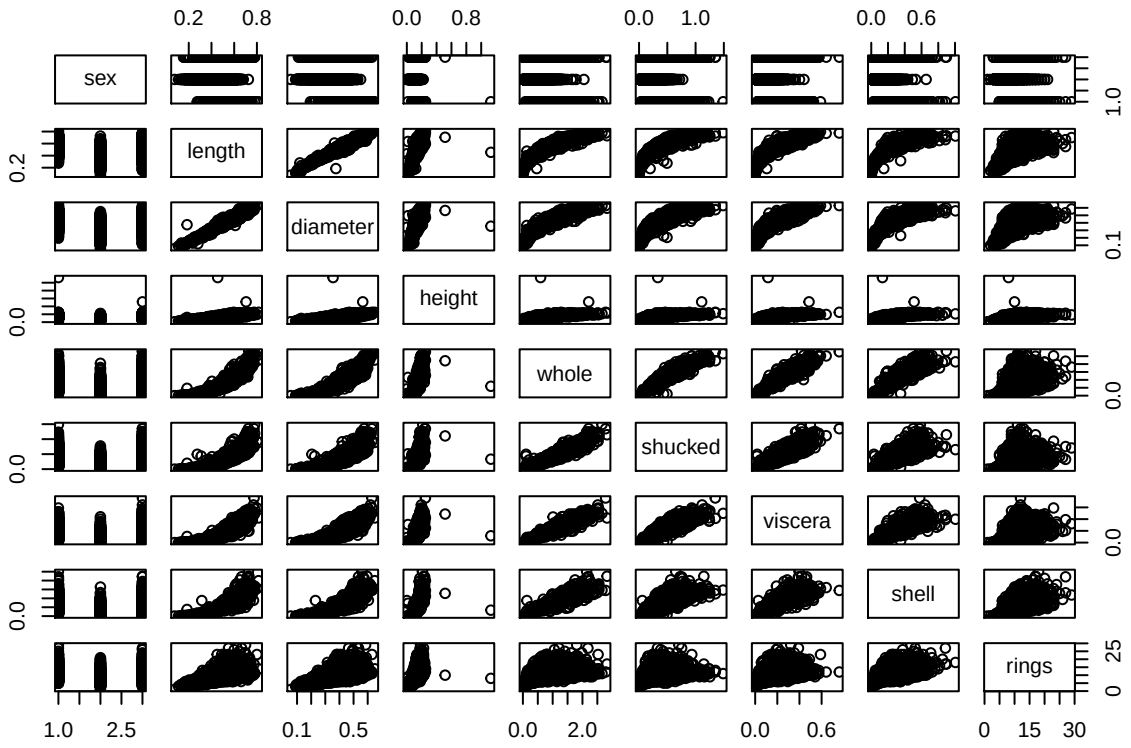
Lab 1

Brendan Alexander

9/27/2019

Import the data

```
data <- read.table("abalone.data",sep = ",")
head(data)
names(data) <- c("sex","length","diameter", "height", "whole", "shucked", "viscera", "shell", "rings")
head(data)
pairs(data)
```



Week 1: The proposal (1-3)

Team lead presenting: Madsen (grp 3)

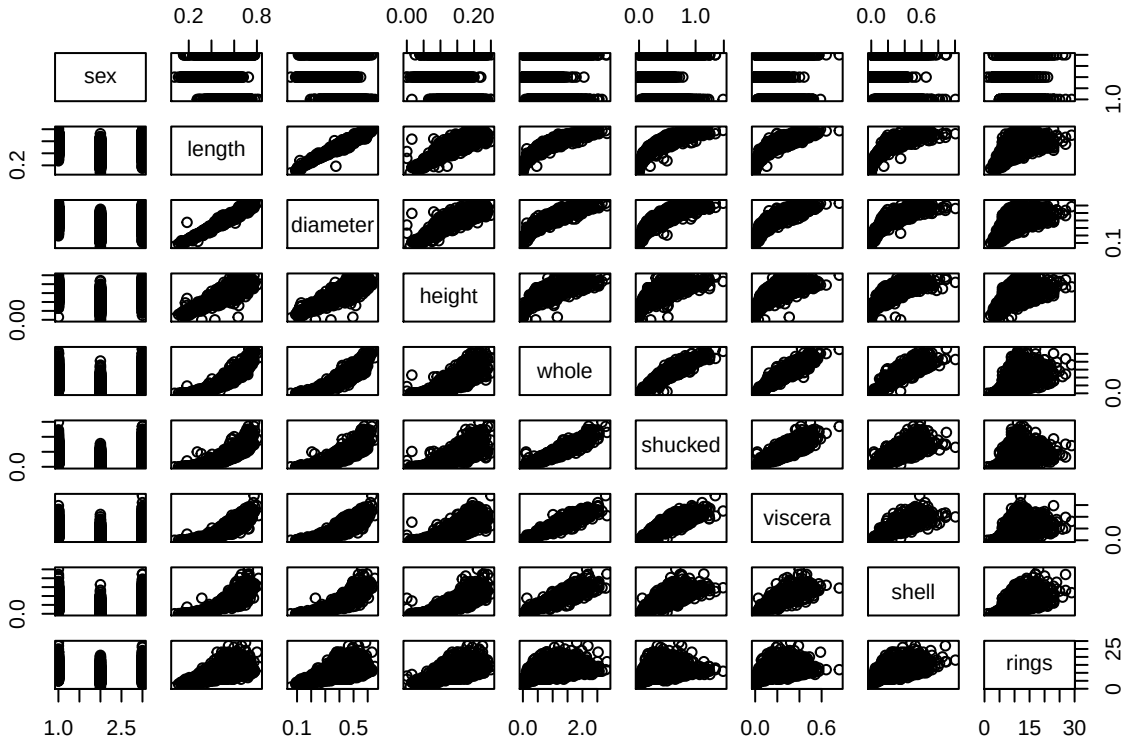
Proposal

1. Height: Outlier issue
   - check for outliers

- Pulled 2 outliers from height

2. Check pairs again

```
data1 <- data[-c(2052,1418),]
pairs(data1)
```



Suggestions:

Group 1

- Stepwise regression to evaluate all the variables
- Remove outliers from weight
- Reconsider sex as: mature vs immature

Group 2

- Sex as a dummy variable?
- Don't want to kill animals, stepwise regression,
  - How to estimate weights without killing?
  - Use whole weight, regress on shucked viscera and shell
    - derived variables from the weight

Group 4

- Correlations first, with `ring`
  - diameter has a strong correlation
  - combine length, diameter and height, make a derived variable

Week 1: The proposal (3-5)

Proposal

1. Inspect correlations
   - sex length and diameter highest correlation, no death
   -
2. Box plots
   - comparative boxplots for the different sex categories
3. Outliers?
   - hov?
     - if not homogeneous, transform the data
4. fit a model
   - rings (or age) ~ (nothing specified)
5. Derived variable –> mature = male + female vs infant
   - boxPlot against length and diameter
6. Derived variable –> volume = length$diameter$height
7. Multiple regression: whole ~ shucked + viscera + shell