Quick Summary of Covariate Data for Lab 05

Here, we only have 2 covariates so summarizing things is fairly simple. There is more one could do but the key is to understand the values that you have, consider whether there might be data errors, and to evaluate relationships between the covariates.

```r
# SwampSquirrel data summary
# read data
dat=read.table("http://www.montana.edu/rotella/502/SwampSquirrels.inp",
      header=FALSE,skip=2,
      colClasses=c("character","integer","numeric","numeric","character"))
# remove column of semi-colons
dat=dat[,-5]
# assign column names
colnames(dat) <- c("ch","freq","birth_date","tail_length")
# summarize the birthday and tail length covariates
summary(dat[,3:4])

##    birth_date    tail_length
##  Min.   :189   Min.   :227
##  1st Qu.:204   1st Qu.:236
##  Median :207   Median :243
##  Mean   :207   Mean   :240
##  3rd Qu.:212   3rd Qu.:245
##  Max.   :225   Max.   :246

# examine correlation between the 2 covariates
cor(dat[,3:4])

##             birth_date tail_length
## birth_date       1.000       0.347
## tail_length      0.347       1.000
```
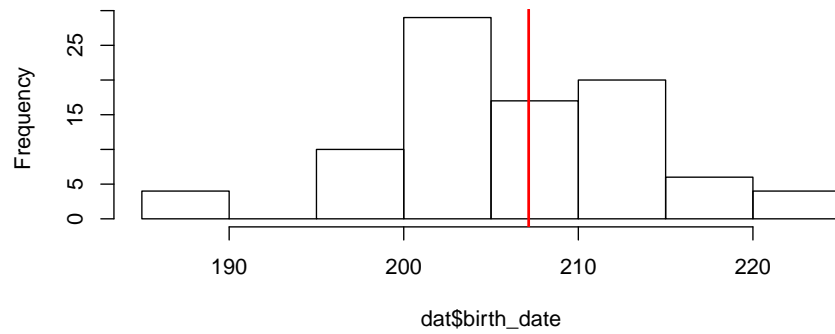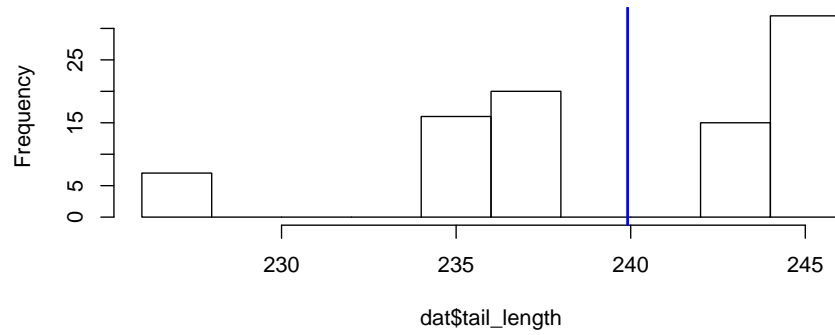
```
par(mfrow = c(2, 1))
hist(dat$birth_date)
abline(v = mean(dat$birth_date), col = "red", lwd = 2)
hist(dat$tail_length)
abline(v = mean(dat$tail_length), col = "blue", lwd = 2)
```
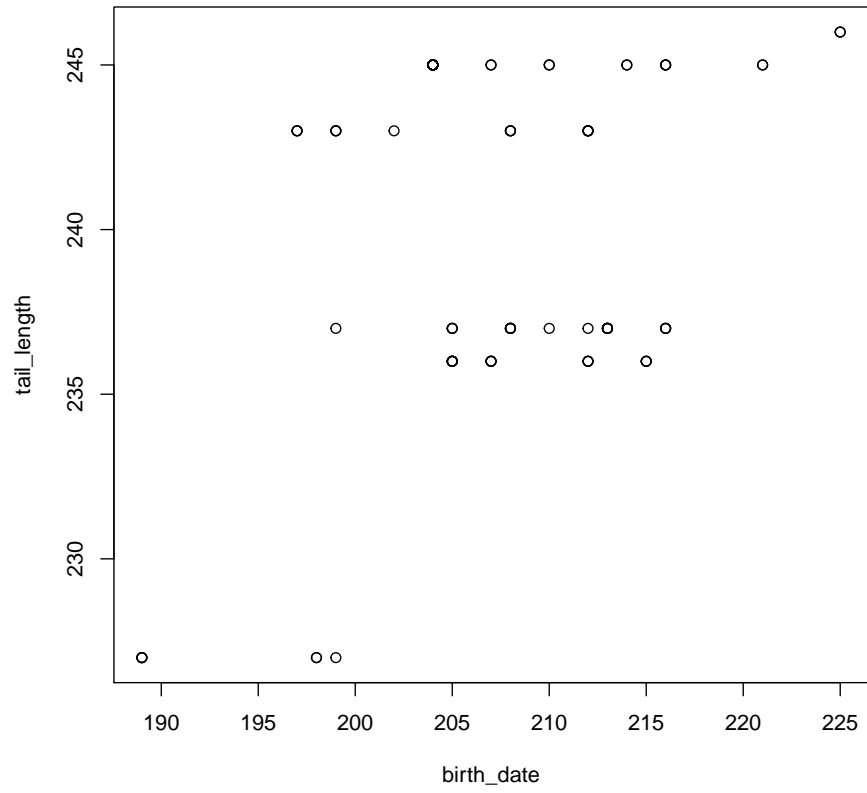
Histogram of dat$birth_date

Histogram of dat$tail_length


```
par(mfrow = c(1, 1))
```

```r
with(data = dat, plot(birth_date, tail_length))
```

You often want a table of summary statistics that includes specific statistics of interest that might not be included in the resuls of R's 'summary' function. If you do, you can write a function customized to your liking. Here's an example.

```r
# write a function called 'stats'
# provide names for the stats you want to report
stat.names=c("mean","sd","min","max",".025q",".5q",".975")
# record the number of stats you want to report
num.stats=length(stat.names)
stats=function(x){
      y=as.matrix(x)
      cols=dim(y)[2]
      b=matrix(NA,cols,num.stats)
      for (i in 1:cols){
        a=y[,i]
        b[i,]=t(rbind(mean(a),sd(a),min(a),max(a),
                t(t(quantile(a, probs = c(0.025,0.50,0.975))))))
      }
row.names(b)=names(x)
colnames(b)=stat.names
      b
}
stats(dat[3:4])

##               mean      sd min max  .025q .5q .975
## birth_date  207.2 7.189 189 225    189 207  221
## tail_length 239.9 5.379 227 246    227 243  245
```