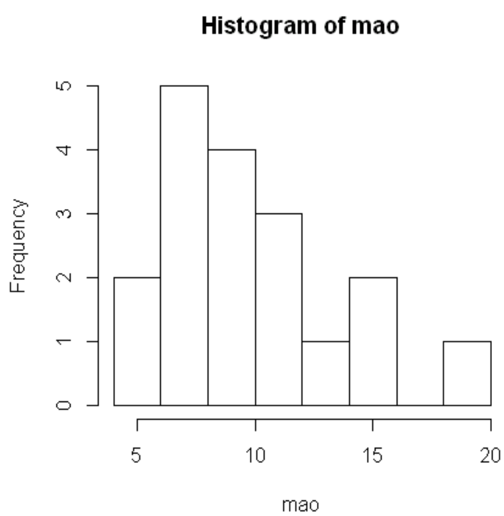
2.1.1(a)
(i) width (mm) last upper molar
(ii) numeric continuous
(iii) *Acropithecus rigidus* specimen
(iv) n=36

2.1.1(b)
(i) There are three: birthweight, date-of-birth, mother's race
(ii) numeric continuous, numeric discrete, categorical nominal
(iii) baby
(iv) n=65

2.2.3
R code for histogram:
```
mao=c( 6.8, 8.4, 8.7,11.9,14.2,18.8,9.9, 4.1, 9.7,12.7, 5.2, 7.8,7.8, 7.4, 7.3,10.6,14.5,10.7)
hist(mao)
```



One example of a tabular frequency distribution is

range	frequency
(4,6]	2
(6,8]	5
(8,10]	4
(10,12]	3
(12,14]	1
(14,16]	2
(16,18]	0
(18,20]	1

2.3.3
R code:
```
b=c(6.3,5.9,7.0,6.9,5.9)
mean(b)
median(b)
```

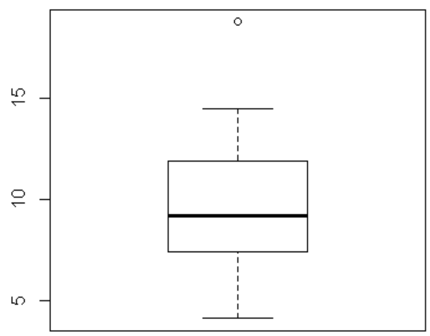
gives $\bar{y}$=6.4 nmol/gm (sample mean) and $\tilde{y}$ =6.3 nmol/gm (sample median).

2.3.11
R code:
```
p=c(5,7,7,8,8,8,9,9,9,10,10,10,10,10,10,10,10,10,11,11,11,11,11,11,11,11,12,12,12,12,12,13,13,
13,14,14)
median(p)
```

This gives $\tilde{y}$=10.5 piglets

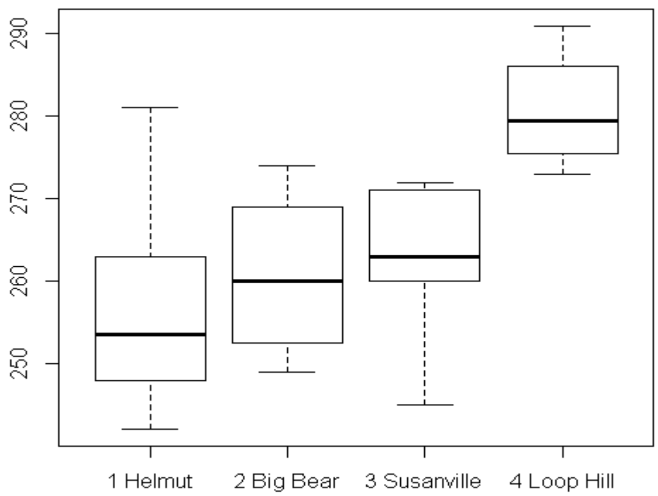
2.4.2
R code:
```
mao=c( 6.8, 8.4, 8.7,11.9,14.2,18.8,9.9, 4.1, 9.7,12.7, 5.2, 7.8,7.8, 7.4, 7.3,10.6,14.5,10.7)
summary(mao) # gives 5-number summary
IQR(mao) # gives interquartile range
```

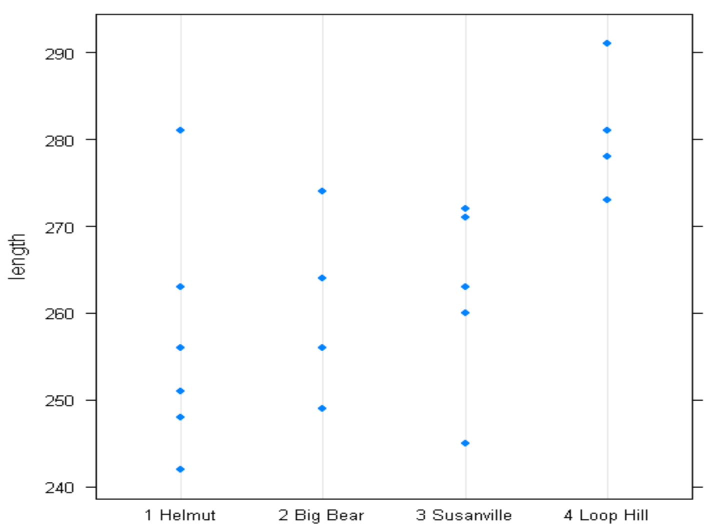
(a) $\tilde{y}$=9.2, Q1=7.5, Q3=11.6
(b) IQR=11.6-7.5=4.1
(c) upper fence is 11.6+1.5(4.1)=17.75
(d) boxplot(mao) gives



2.4.7
(a) IQR=127.42-113.59=13.83
lower fence = 113.59-1.5(13.83)=92.845
upper fence = 127.42+1.5*(13.83)=148.165
(b) There are no outliers; the smallest value 95.16 is larger than the lower fence, and the largest value 145.11 is smaller than the upper fence

2.5.2
The posted R code gives:



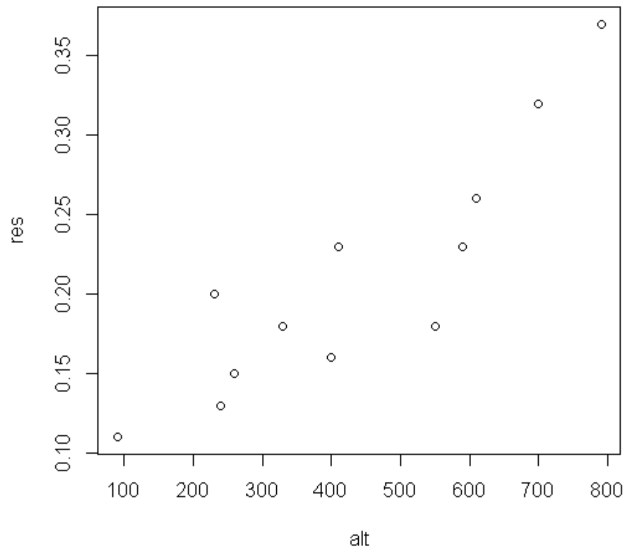


South to North makes more sense, so we can get a feel for what happens to the squirrels' length as we move north.

2.5.3

R code:

```
alt=c(90,230,240,260,330,400,410,550,590,610,700,790)
res=c(0.11,0.20,0.13,0.15,0.18,0.16,0.23,0.18,0.23,0.26,0.32,0.37)
plot(alt,res)
```



2.6.5
R code:

```
atp=c(1.45,1.19,1.05,1.07)
mean(atp)
sd(atp)
```

This gives ȳ=1.19 and s=0.184. You *can* do this by hand using the formulas.
ȳ=(1.45+1.19+1.05+1.07)/4=1.19 and
s=sqrt(((1.45-1.19)^2+(1.19-1.19)^2+(1.05-1.19)^2+(1.07-1.19)^2)/3)=0.184

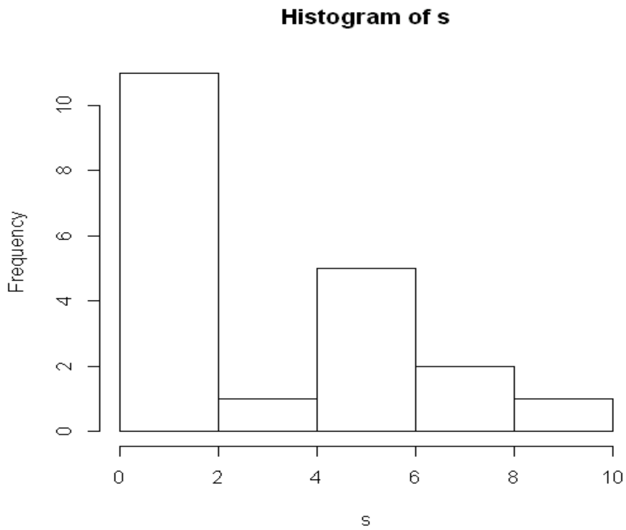
2.6.11
The intervals described are (57.9,138.7), (17.5,179.1), and (-22.9,219.5). The proportions in these intervals are:
26/36=72.2%, 34/36=94.4%, 36/36=100.0%.

2.6.12
These are reasonably close to the percentages given by the empirical rule, 68%, 95%, and >99%.

2.S.7
R code:

```
> s=c(5,0,9,6,0,0,5,0,6,1,5,0,0,0,0,7,0,0,4,7)
> mean(s)
[1] 2.75
> median(s)
[1] 0.5
```

Gives ȳ=2.75 seizures and ỹ =0.5 seizures. Note that mean pulled higher in direction of skew (skewed right).



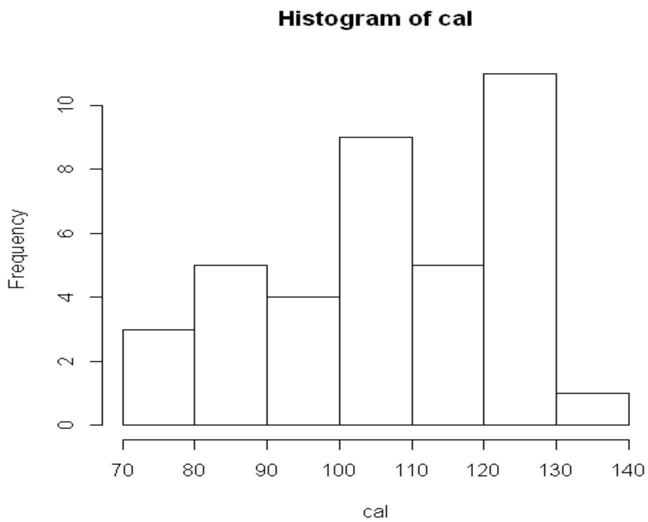
The most common observation is zero (also called the mode). This is an example of "zero inflated data", which also occurs with variables like medical costs in a month, earthquakes in a year, etc. Often scientists will separate the analysis of such data into the probability of a zero, and the distribution of positive outcomes.

2.S.17
This is a challenging problem!

From looking at the histograms, (a) and (b) will have mean and medians around 40, but (b) will have a larger standard deviation. (c) should have the mean pulled higher than the median. Given all this, (a) is output 2, (b) is output 4, and (c) is output 1.

2.S.19
R code:

```
>
cal=c(95,110,135,120,88,125,112,100,130,107,86,130,122,122,127,107,107,107,88,126,125,112
,78,115,78,102,103,93,88,110,104,122,112,80,121,126,90,96)
> summary(cal)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  78.00   95.25  108.50  107.90  122.00  135.00
> mean(cal)
[1] 107.8684
> sd(cal)
[1] 16.0778
> hist(cal)
```



The mean and median are both about 108 nM, so this is a typical value. The interval containing the middle 50% of the data has length 27 nM; an observation typically deviates from the mean by 16 nM. The data are very slightly skewed to the left.

2.S.20
The key here is the length of the box. The boxplot shows the middle 50% of the data are between, roughly, 21 and 37. The histogram that bests match this is (a).