Descriptive Statistics and T-Tests

So first thing's first we want to read in some data that we are going to do the analysis on.

Remember first thing's first set your working Directory

```
setwd()
```

read in the data called livestock.csv

```
Data <- read.csv('livestock.csv')
```

ok now we've got the data lets take a look at it

```
str(Data)
```

So we've got a couple data columns that we can use

```
mean(Data$livestock)
```

this gives us the mean of one of the columns now do it for the other

mode is not an inbuilt function in R because really no one uses the mode.

this next block of code creates our own function for mode, don't worry about it.

```
getmode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}
```

```
getmode(Data[,1])
```

well at least there is median in R, lets do that

```
median(Data$livestock)
```

also easy, do the other one.

How about standard deviation

```
sd(Data$livestock)
```

gosh this is easy

Lets work it out by hand

first we need the mean

```
mn <- mean(Data$livestock)
vc <- Data$livestock
var1<- (vc - mn)^2
var1<- sum(var1)/length(Data$livestock)
stdv <- sqrt(var1)
stdv
```

that's pretty close right!

how about confidence intervals.

confidence intervals are normally worked out for fitted models so we will do it by hand.

```
ci1 <- mn + 1.96*(sd(Data$livestock)/sqrt(length(Data$livestock)))
ci1
```

this is a confidence interval, but it is only one side.

```
ci2 <- mn - 1.96*(sd(Data$livestock)/sqrt(length(Data$livestock)))
ci2
```

there we go

time to do some t-tests

t-tests are done with functions again.

```
t.test(Data$livestock, Data$birds, paired = TRUE)
```

There we go we've compared livestock to birds, but we could also compare within livestock or bird

this is actually hard to do because it should be done in an anova but we can call specific parts of the data using a package

this line installs the package if you don't have it

```
install.packages('data.table')
```

this line loads it in

```
library(data.table)
```

this package creates "enhanced data frames"

this just turns it into something our package can use

```
Data <- data.table(Data)
```

lets pull out what we need

this is just saying for all the data

```
thornveldl <- Data[group == "thornveld"]$livestock
grasslandl <- Data[group == "grassland"]$livestock


t.test(thornveldl, grasslandl, paired = TRUE)
```

now try and do the same for the difference between birds

if we want to do it without loading in a new package we can just use the subset command

```
thornveldb <- subset(Data$birds, Data$group == 'thornveld')
grasslandb <- subset(Data$birds, Data$group == 'grasskand')
```

this achieves the exact same thing

now t test it

But there is a problem with doing lots of t-tests on things, the more tests you do the more likely you are to get false positive, there is a better way! more to follow.

read in the dataset wood.csv

this dataset is species, their regions and their diameter.

using t tests work out if there are differences between the means of diameter in the different regions