

# My Internship

By Mark De Lisio

# Why did I apply for the internship

- To gain a better understanding in statistics
- To gain knowledge and understanding in using statistical programs such as R studio
- To understand and experience the role a biometrician plays in experimental design
- To do something different and something out of my comfort zone to gain new skills

# First Week Monday

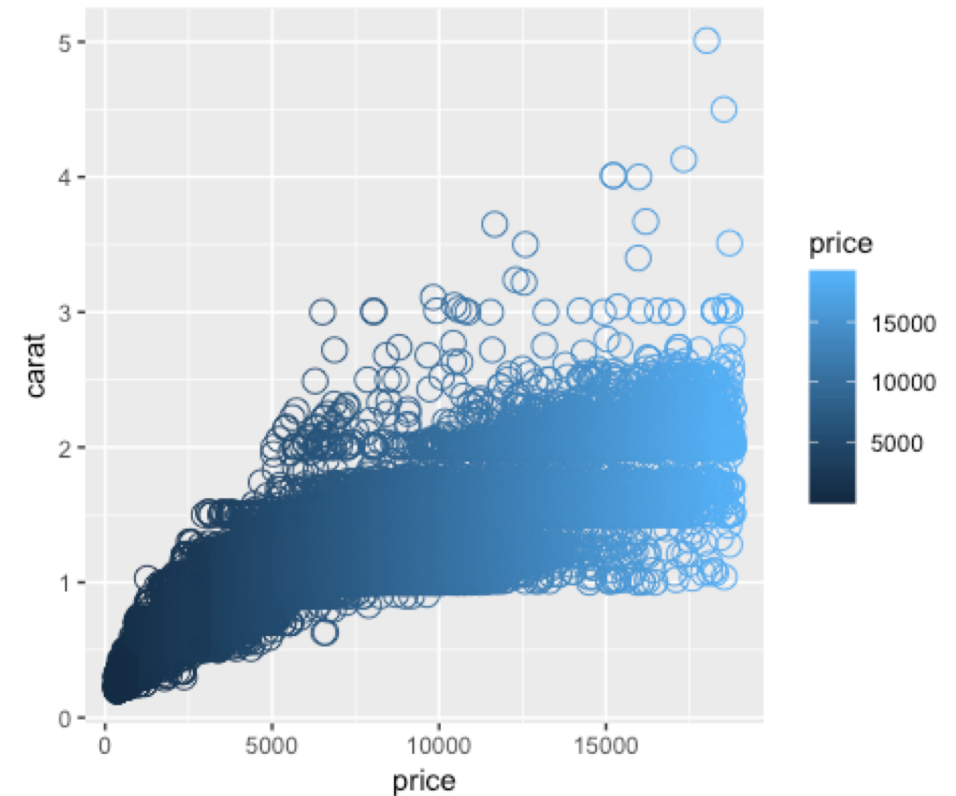
- Had the opportunity to participate in Sam's 'Introduction to R' workshop.
- Learnt the basics of R and how to use it efficiently
- Learnt how to write scrips
- Learnt how to use R for:
  - Data Management
  - Creating Functions and Vectors
  - Statistical tests
  - Basic Graphics

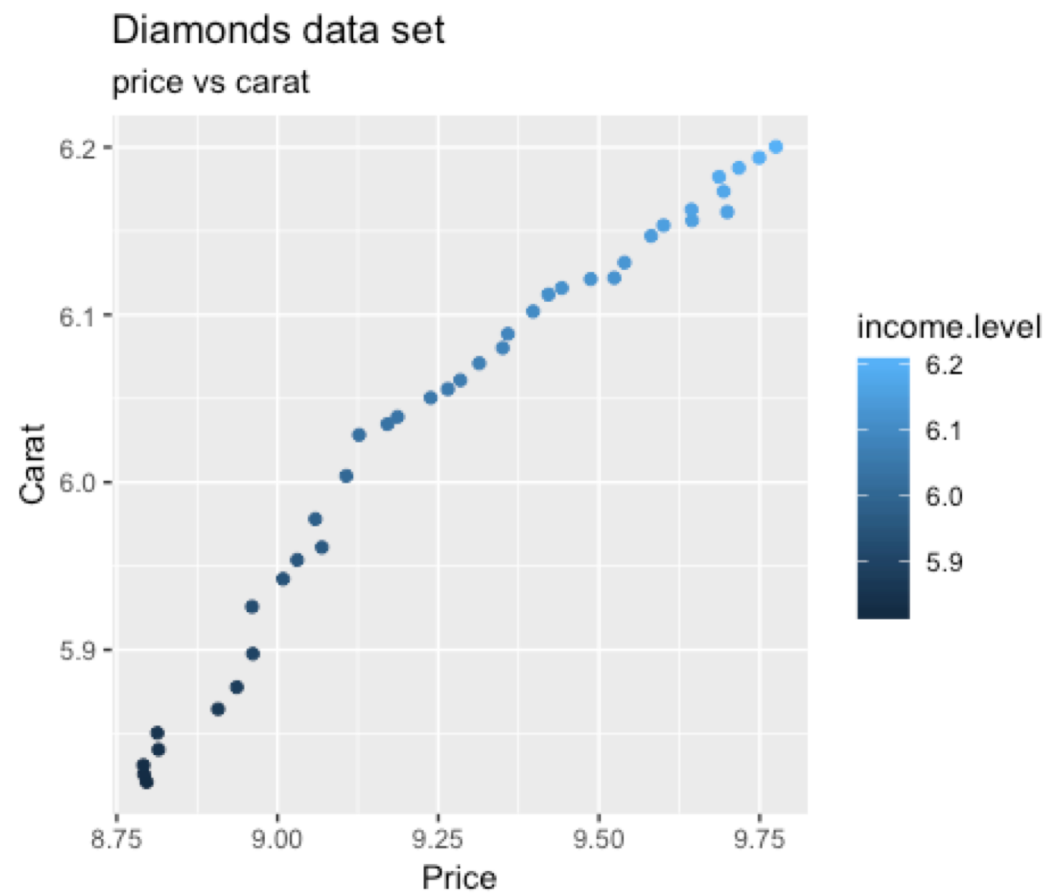
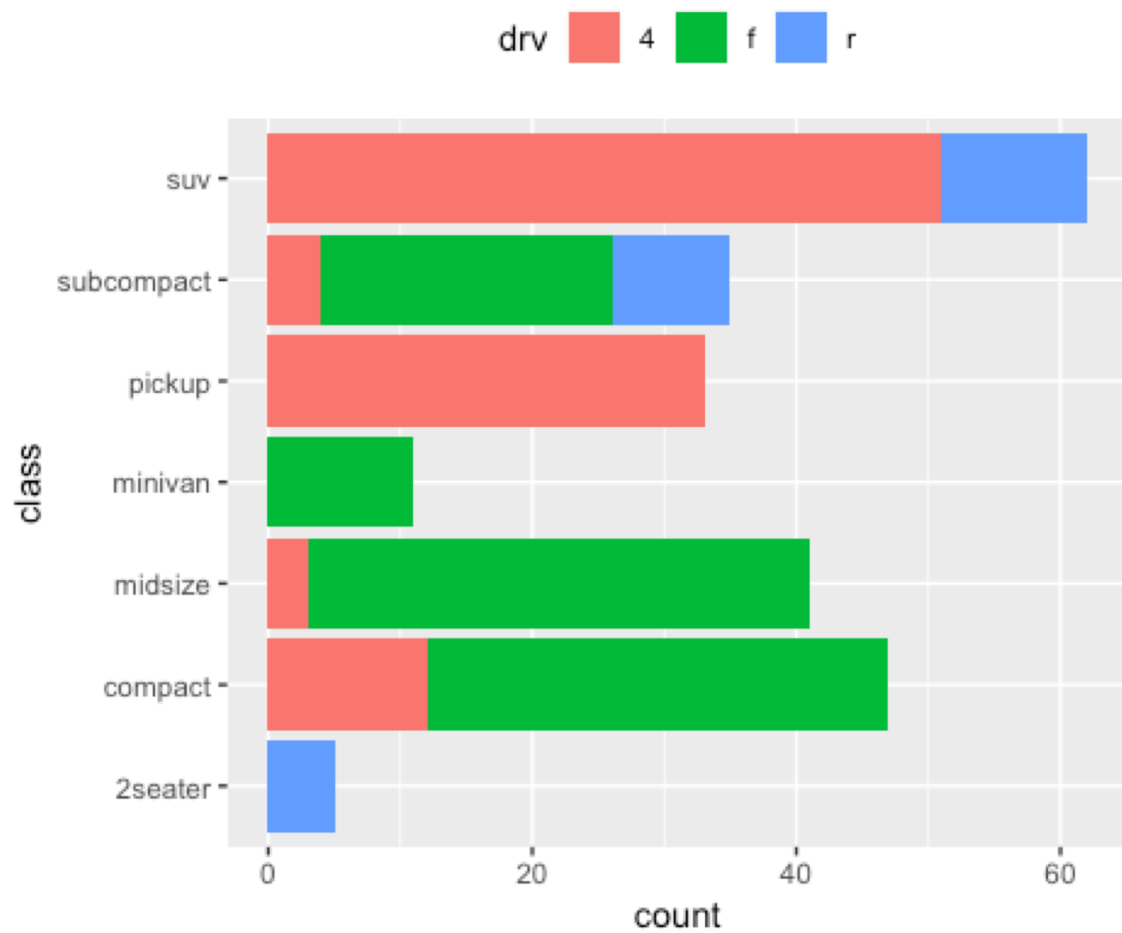
# First Week Tuesday

- Had the opportunity to participate in Sharon's 'Planning and designing an agronomic experiment workshop'.
- Learnt how to set up an agronomic experiment correctly
- Learnt how to use R to generate:
  - Completely randomised designs
  - Randomised complete block design
  - Latin square designs
  - Split-plot designs

# First Week Wednesday

- Learnt how to use ggplot
- Used an online tutorial
- Got to practice and improve skills using ggplot on different data sets and creating different graphs





# First Week Thursday

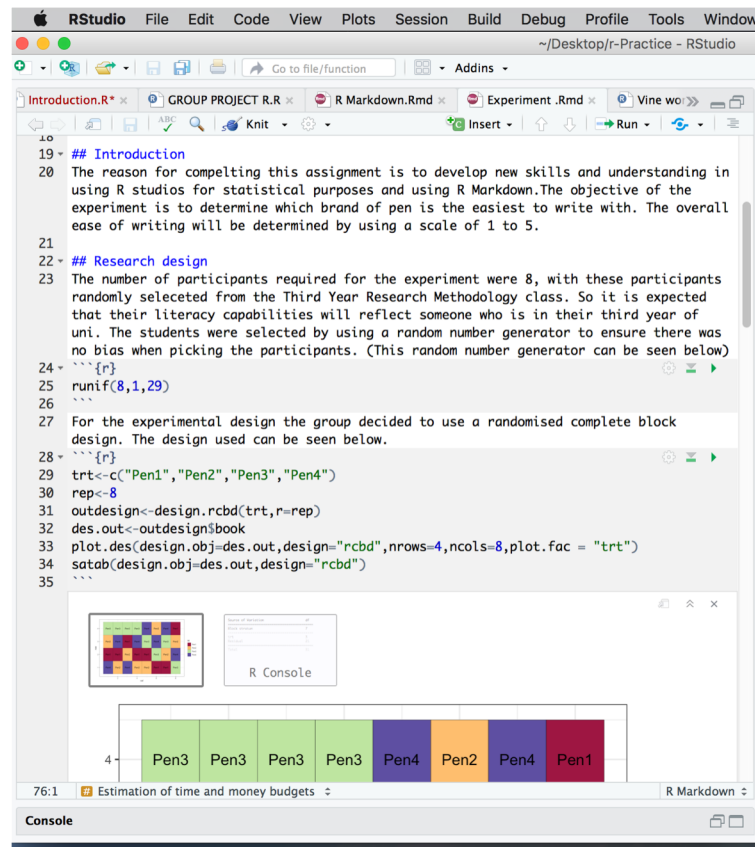
- Continued practicing with ggplot
- Used R to analyse our group project data set and see if they produced the same results as genestat (which they did)
- Continued improving skills with R

## Analysis of variance

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Variate: Overall_ease_of_writing_score					
Assessor	7	0.719	0.103	0.324	0.934489	Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Pen.type	3	9.594	3.198	10.089	0.000255 ***	Assessor stratum	7	0.7188	0.1027	0.32	
Residuals	21	6.656	0.317			Assessor.*Units* stratum					
						Pen_type	3	9.5938	3.1979	10.09	<.001
						Residual	21	6.6562	0.3170		
						Total	31	16.9688			

# First Week Friday

- Continued using R and improving skills
- Had a talk with Sam about R markdown and how to use it



```
19 # Introduction
20 The reason for compelling this assignment is to develop new skills and understanding in
    using R studios for statistical purposes and using R Markdown.The objective of the
    experiment is to determine which brand of pen is the easiest to write with. The overall
    ease of writing will be determined by using a scale of 1 to 5.
21
22 # Research design
23 The number of participants required for the experiment were 8, with these participants
    randomly selected from the Third Year Research Methodology class. So it is expected
    that their literacy capabilities will reflect someone who is in their third year of
    uni. The students were selected by using a random number generator to ensure there was
    no bias when picking the participants. (This random number generator can be seen below)
24 ```{r}
25 runif(8,1,29)
26 ```
27 For the experimental design the group decided to use a randomised complete block
    design. The design used can be seen below.
28 ```{r}
29 trt<-c("Pen1", "Pen2", "Pen3", "Pen4")
30 rep<-8
31 outdesign<-design.rcbd(trt,r=rep)
32 des.out<-outdesign$book
33 plot.des(design.obj=des.out,design="rcbd",nrows=4,ncols=8,plot.fac = "trt")
34 satab(design.obj=des.out,design="rcbd")
35 ```
```

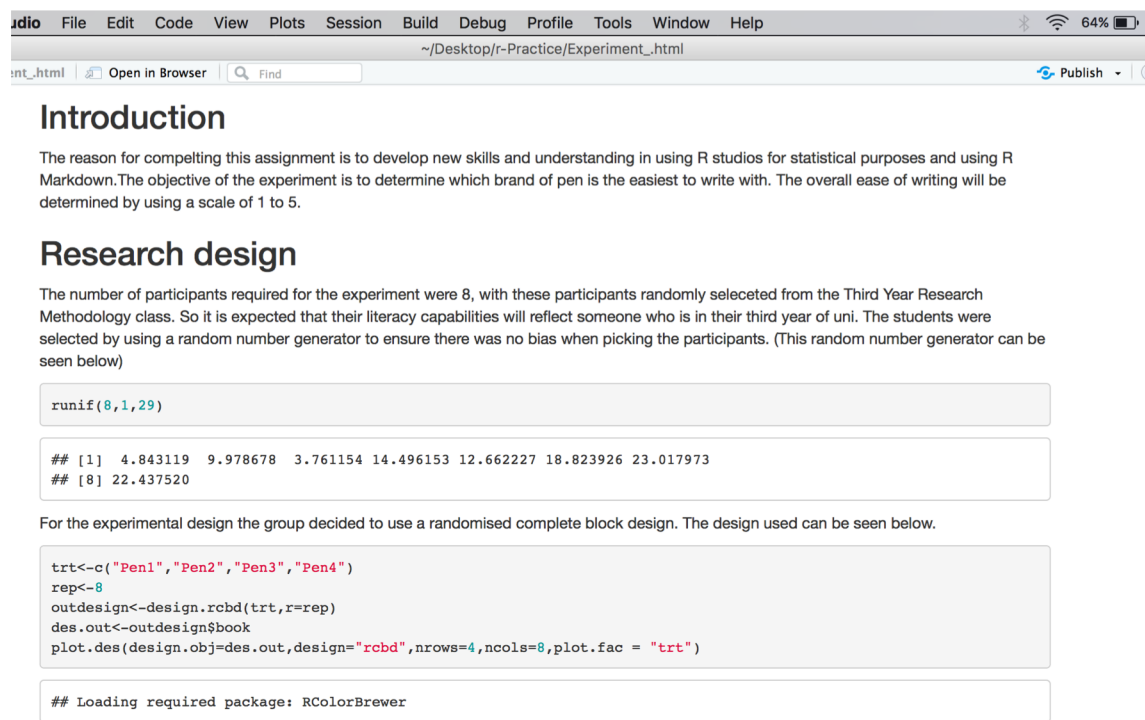
The screenshot shows the RStudio interface with a script editor containing R Markdown code. The code defines a randomized complete block design (rcbd) with 4 blocks and 8 treatments (Pen1, Pen2, Pen3, Pen4). Below the code, a visualization of the design is shown as a 4x8 grid of colored cells. The colors represent the pen brand used in each cell: Pen3 (green), Pen4 (purple), Pen2 (orange), and Pen1 (red). The grid shows that each pen is used exactly once in each of the 4 blocks.

Block	Pen1	Pen2	Pen3	Pen4
1	Pen1	Pen2	Pen3	Pen4
2	Pen2	Pen3	Pen4	Pen1
3	Pen3	Pen4	Pen1	Pen2
4	Pen4	Pen1	Pen2	Pen3



# Second Week Friday

- Used R markdown to rewrite our group project
- Continued using R to improve skills



**Introduction**

The reason for competing this assignment is to develop new skills and understanding in using R studios for statistical purposes and using R Markdown. The objective of the experiment is to determine which brand of pen is the easiest to write with. The overall ease of writing will be determined by using a scale of 1 to 5.

**Research design**

The number of participants required for the experiment were 8, with these participants randomly selected from the Third Year Research Methodology class. So it is expected that their literacy capabilities will reflect someone who is in their third year of uni. The students were selected by using a random number generator to ensure there was no bias when picking the participants. (This random number generator can be seen below)

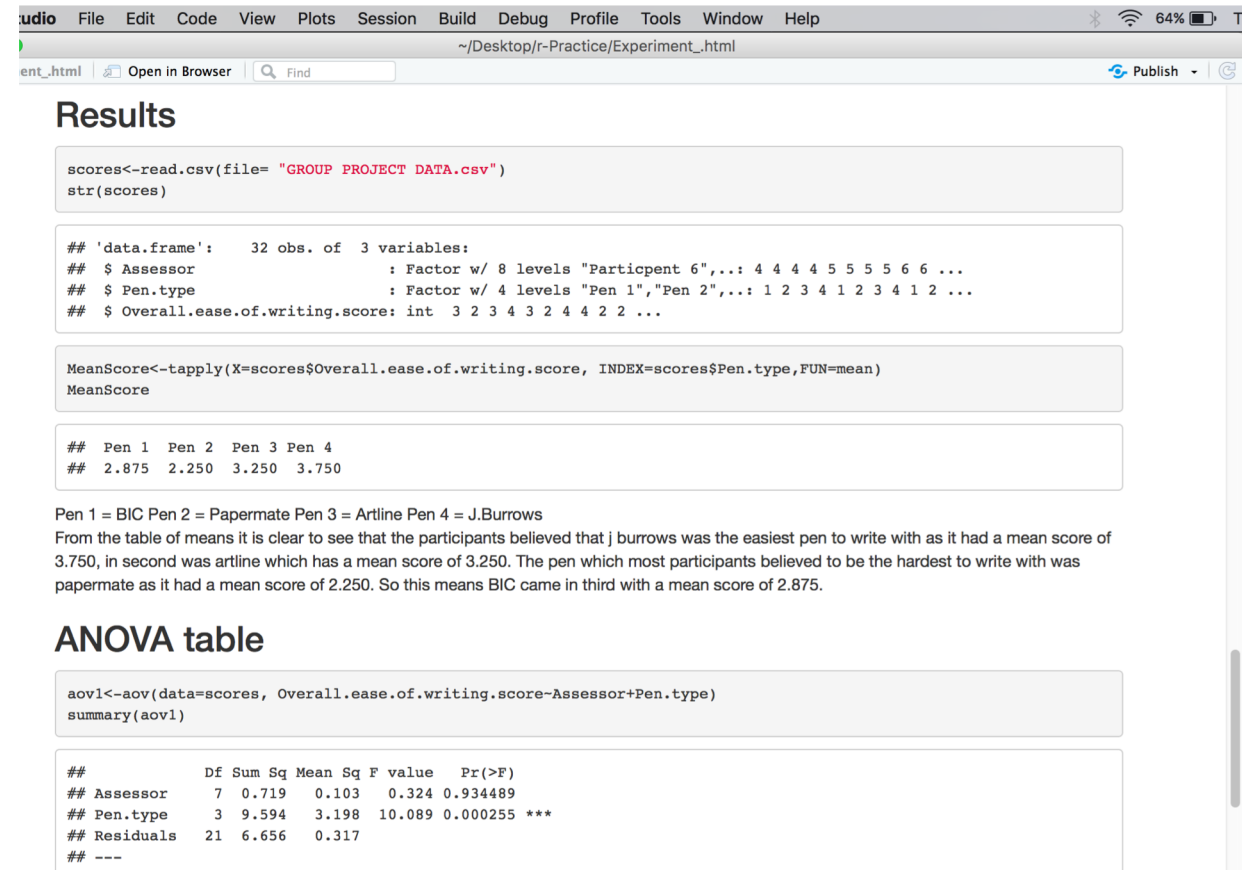
```
runif(8,1,29)
```

```
## [1] 4.843119 9.978678 3.761154 14.496153 12.662227 18.823926 23.017973  
## [8] 22.437520
```

For the experimental design the group decided to use a randomised complete block design. The design used can be seen below.

```
trt<-c("Pen1", "Pen2", "Pen3", "Pen4")  
rep<-8  
outdesign<-design.rcbd(trt,r=rep)  
des.out<-outdesign$book  
plot.des(design.obj=des.out,design="rcbd",nrows=4,ncols=8,plot.fac = "trt")
```

```
## Loading required package: RColorBrewer
```



**Results**

```
scores<-read.csv(file= "GROUP PROJECT DATA.csv")  
str(scores)
```

```
## 'data.frame': 32 obs. of 3 variables:  
## $ Assessor : Factor w/ 8 levels "Participant 6",...: 4 4 4 4 5 5 5 6 6 ...  
## $ Pen.type : Factor w/ 4 levels "Pen 1","Pen 2",...: 1 2 3 4 1 2 3 4 1 2 ...  
## $ Overall.ease.of.writing.score: int 3 2 3 4 3 2 4 4 2 2 ...
```

```
MeanScore<-tapply(X=scores$Overall.ease.of.writing.score, INDEX=scores$Pen.type,FUN=mean)  
MeanScore
```

```
## Pen 1 Pen 2 Pen 3 Pen 4  
## 2.875 2.250 3.250 3.750
```

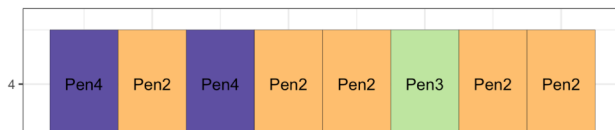
Pen 1 = BIC Pen 2 = Papermate Pen 3 = Artline Pen 4 = J.Burrows

From the table of means it is clear to see that the participants believed that j burrows was the easiest pen to write with as it had a mean score of 3.750, in second was artline which has a mean score of 3.250. The pen which most participants believed to be the hardest to write with was papermate as it had a mean score of 2.250. So this means BIC came in third with a mean score of 2.875.

**ANOVA table**

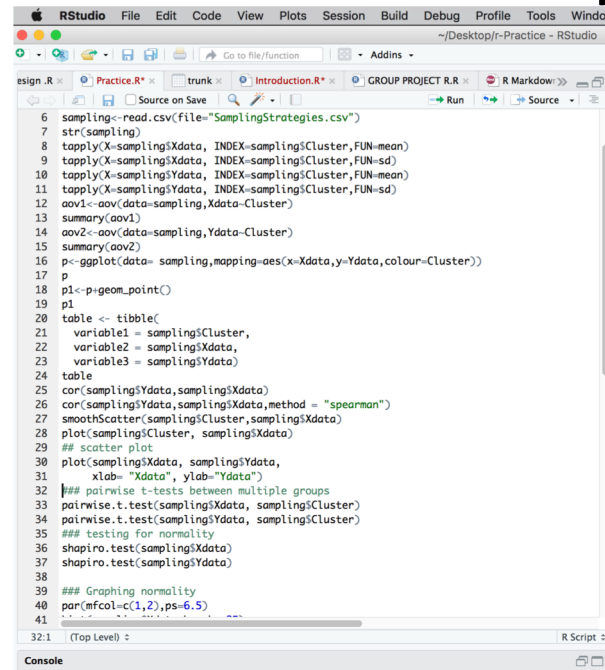
```
aov1<-aov(data=scores, Overall.ease.of.writing.score~Assessor+Pen.type)  
summary(aov1)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)      
## Assessor    7  0.719   0.103   0.324 0.934489      
## Pen.type     3  9.594   3.198  10.089 0.000255 ***  
## Residuals   21  6.656   0.317                  
## ---
```



# Third Week Monday

- Continued practicing in R to improve skills
- Had a talk with Richard about the principles of statistical inference in practical applications.
- Received new data sets from Peter to play around with.



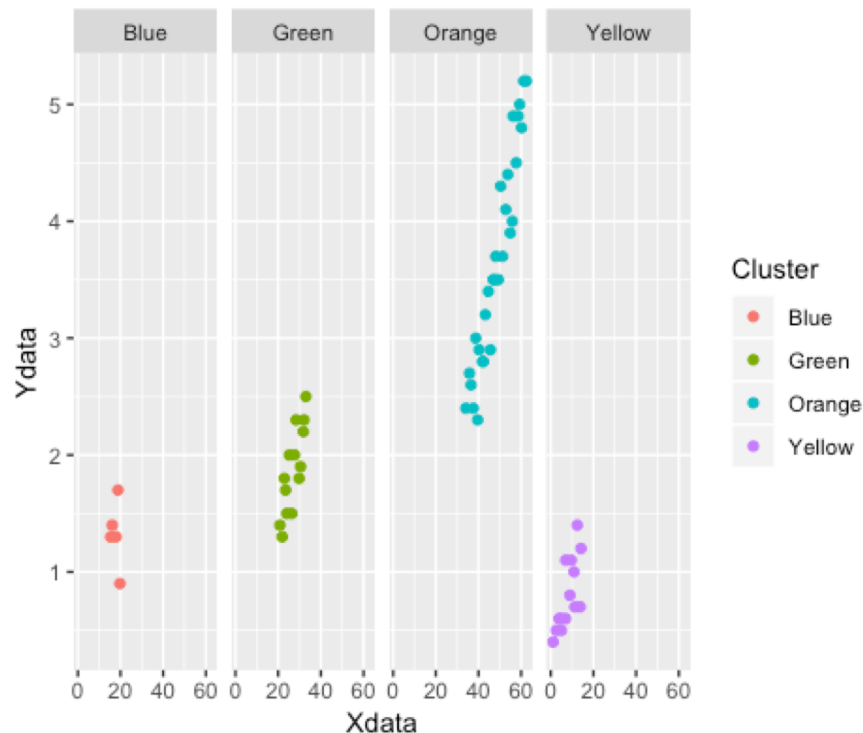
```
6 sampling<-read.csv(file="SamplingStrategies.csv")
7 str(sampling)
8 tapply(X=sampling$Xdata, INDEX=sampling$Cluster, FUN=mean)
9 tapply(X=sampling$Xdata, INDEX=sampling$Cluster, FUN=sd)
10 tapply(X=sampling$Ydata, INDEX=sampling$Cluster, FUN=mean)
11 tapply(X=sampling$Ydata, INDEX=sampling$Cluster, FUN=sd)
12 aov1<-aov(data=sampling, Xdata=Cluster)
13 summary(aov1)
14 aov2<-aov(data=sampling, Ydata=Cluster)
15 summary(aov2)
16 p<-ggplot(data= sampling, mapping=aes(x=Xdata, y=Ydata, colour=Cluster))
17 p
18 p1<-p+geom_point()
19 p1
20 table <- tibble(
21   variable1 = sampling$Cluster,
22   variable2 = sampling$Xdata,
23   variable3 = sampling$Ydata)
24 table
25 cor(sampling$Ydata, sampling$Xdata)
26 cor(sampling$Ydata, sampling$Xdata, method = "spearman")
27 smoothScatter(sampling$Cluster, sampling$Xdata)
28 plot(sampling$Cluster, sampling$Xdata)
29 ## scatter plot
30 plot(sampling$Xdata, sampling$Ydata,
31      xlab="Xdata", ylab="Ydata")
32 ## pairwise t-tests between multiple groups
33 pairwise.t.test(sampling$Xdata, sampling$Cluster)
34 pairwise.t.test(sampling$Ydata, sampling$Cluster)
35 ### testing for normality
36 shapiro.test(sampling$Xdata)
37 shapiro.test(sampling$Ydata)
38
39 ### Graphing normality
40 par(mfcol=c(1,2), ps=6.5)
41
```

# Third Week Tuesday

- Continued using the new data set to improve skills in R
- Had the opportunity to sit in and listen to Chris's PHD talk
- Had a talk with Peter about Sampling and the math's/ formulas behind it.
- Had a talk with Beata about how statistics can be used in the workforce and how she uses statistics for genetics.
- Had the opportunity to visit the plant accelerator and speak to Chris about the importance of experimental design

# Third Week Wednesday

- Continued improving skills in R
- Had the opportunity to sit in on a consultancy meeting
- Had a talk with Helena about the principles of experimental design



Third Week Thursday

# Reflection

- The internship has given me the opportunity to improve my skills in using statistical programs and allowed me to improve my understanding of statistical knowledge.
- It has also showed me the important role biometricians play in experiments and setting the experiments up correctly to get valid results.
- The internship experience will help me in the future as I look to pursue a career in research work.