INTERNSHIP PROGRAM

By: Xinyuan FU
20/07/2018

The University of Adelaide
Biometry Hub

INTRODUCTION

- The internship was based at Biometry Hub, University of Adelaide

- I completed the required 84 hours (02/07/2018 - 20/07/2018) internship by having statical course and learning how to use R language.

- Internship Supervisor: Wendy and Sam

AGENDA

- Internship mission
- Internship Tasks
- Achievement
- Skills Utilised
- What did I learn
- Internship Impact
- Conclusion

INTERNSHIP MISSION

- My mission:

Learn experimental designs and analysis

Further understanding of application data analysis to the reality.

Interested in the way of experimental design and data collection by the professional statisticians.

INTERNSHIP TASK

Week 1

- Courses:

An introduction to R and Rstudio

Planning and designing an agronomic experiment.

- Learning how to work with “tidyverse (ggplot2)”, “R markdown”

INTERNSHIP TASK

Week 2

- Review what we've done in the first week
- Use R markdown to do the Group Project Report in Research Methodology course

INTERNSHIP TASK

Week 3

- Classes:

Principles of statistical inference

Linear mixed model

Principles of experimental designs

Dataset

Types of Sampling (Randomised, Stratified and Cluster)

- Have a

Presentation from a PHD Student

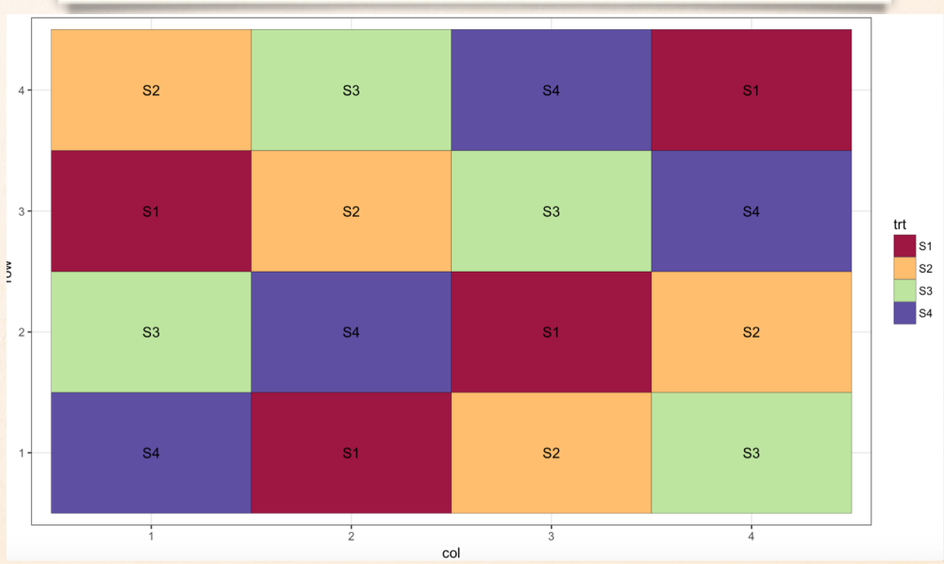
Consultant meeting from GRDC project

Tour at Plant Accelerator

ACHIEVEMENTS

- Planning and designing an experiment.

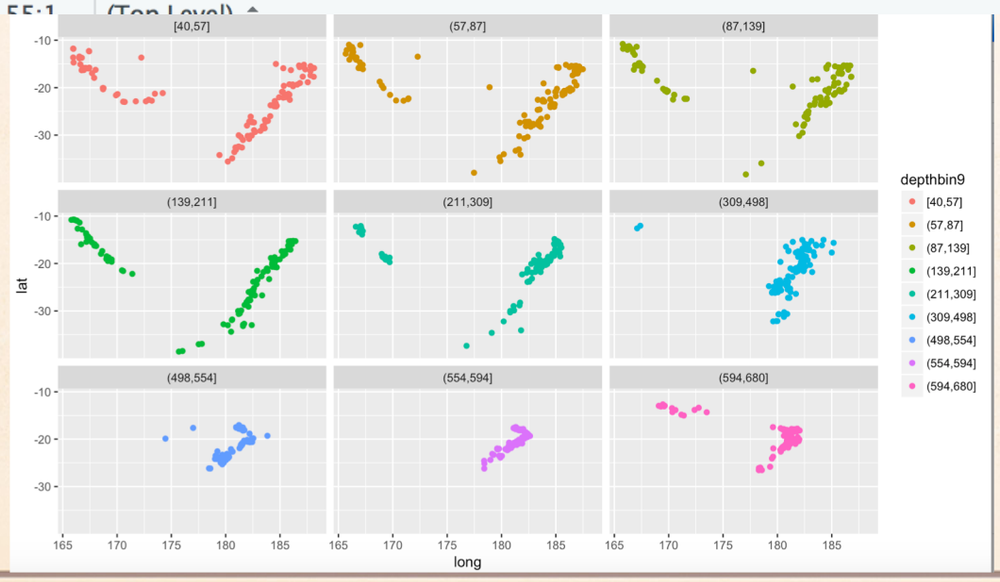
```
# Exercise 1
trt <- c(1,2,3,4)
rep <- 3
outdesign <- design.crd(trt, r=rep, serie = 0)
des.out <- outdesign$book
plot.des(design.obj = des.out, design = "crd", nrows = 4,
         ncols = 3, plot.fac = "trt")
satab(design.obj = des.out, design = "crd")

# Exercise 2
trt <- c(1,2,3,4,5,6,7)
rep <- 12
outdesign <- design.crd(trt, r=rep, serie = 0)
des.out <- outdesign$book
plot.des(design.obj = des.out, design = "crd", nrows = 6,
         ncols = 14, plot.fac = "trt")
satab(design.obj = des.out, design = "crd")

# RCBD
# Example 2

trt <- c(1,2,3,4,5,6,7,8,9,10,11)
rep <-4
outdesign <- design.rcbd(trt, r=rep, serie=0)
des.out <- outdesign$book
plot.des(design.obj = des.out, design = "rcbd", nrows = 11,
         ncols = 4, plot.fac = "trt")
```



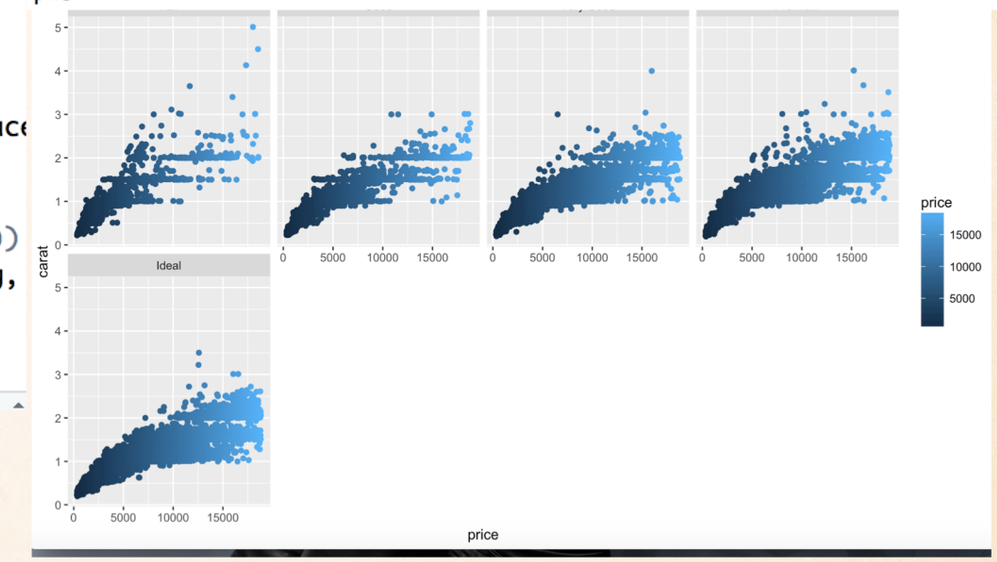
ACHIEVEMENTS

- ggplot 2

```
ggplot(data = quakes) + geom_histogram(aes(x = mag,  fill = 'grey30'),b
ggplot(quakes) + geom_histogram(aes(x = depth), bin = 7, fill = 'grey3(
ggplot(quakes) + geom_histogram(aes(x = stations), bin = 3, fill = 'gre
ggplot(quakes) + geom_point(aes(x = depth, y = mag))
ggplot(quakes) + geom_point(aes(x = mag, y = stations))

ggplot(quakes) + geom_point(aes(x = long, y = lat, color = mag)) + face

quakes$depthbin9 <- cut_number(quakes$depth, 9)
ggplot(quakes) + geom_point(aes(x = long, y = lat, color = depthbin9))
ggplot(data = quakes, x = lat, y = long, size = exp(mag), color = mag,

ggplot(quakes) + geom_point(aes(x = mag, y = stations, shape=18))
```



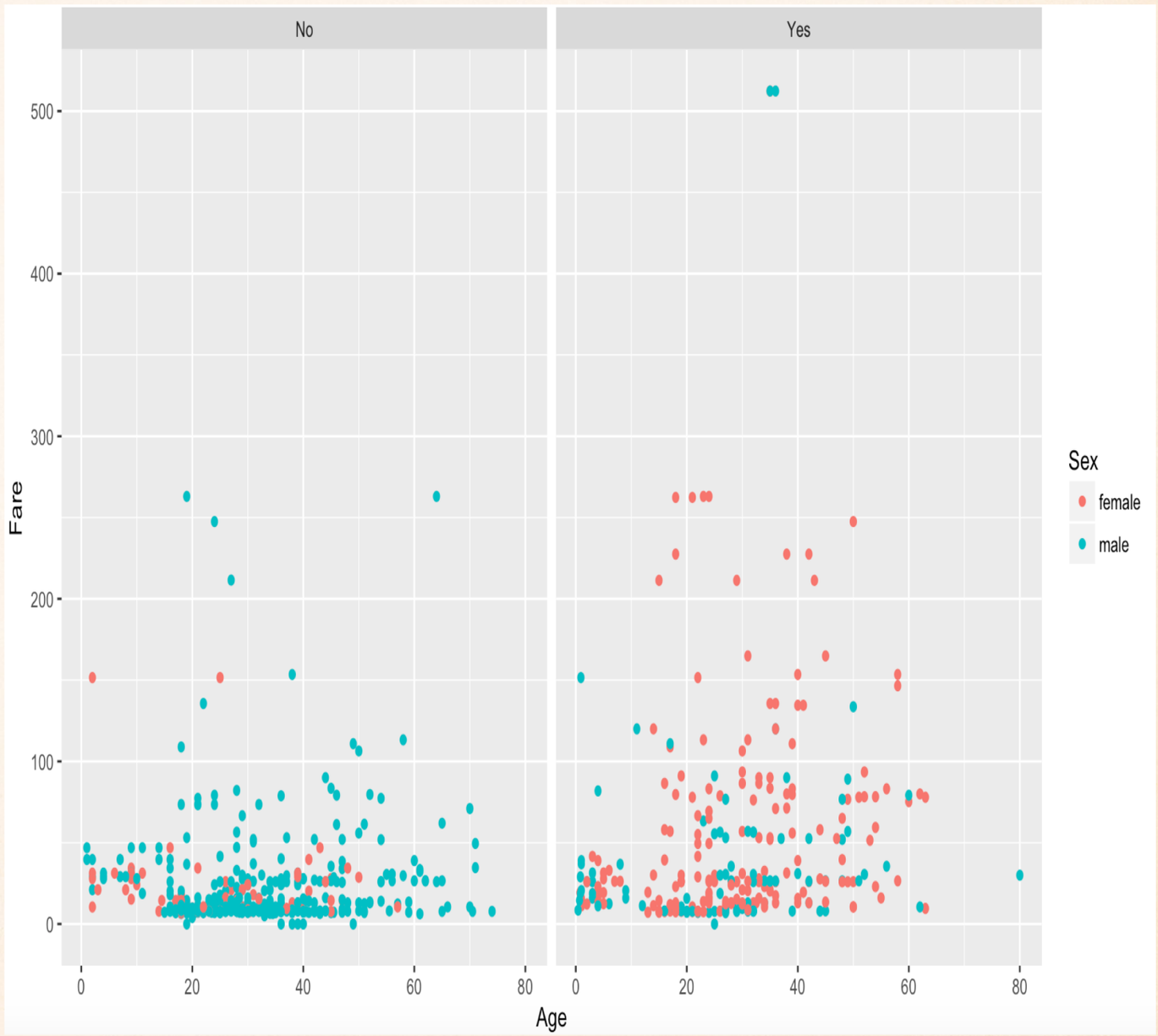

```
head(brewer.pal.info, 10)
p7 <- p1 + coord_cartesian(xlim=c(-5, 30000), ylim=c(-3, 10))
p7
p8 <- p1 + scale_x_continuous(breaks=seq(0, 30000, 10000),labels = scal
p8
p9 <- p1 + facet_wrap( ~ cut , ncol=4)
p9
p10 = p1 + facet_wrap(color ~ cut, ncol = 5)
p10
p11 <- p9 + geom_smooth(method = "lm")
p11
p13 <- p9 + geom_smooth(method = "lm", size = 0.1,colour = "black")
p13
```



ACHIEVEMENTS

- Tidyverse

```
26  passengers %>%
27    mutate(FamSize = Parch + SibSp) %>%
28    arrange(desc(FamSize))
29
30  # Turn numerical values of Survived column to "No" & "Yes" (new data fr
31  passengers1 <- passengers %>%
32    mutate(Survived = ifelse(Survived == 0, "No", "Yes"))
33  passengers1
34
35  # Plot barplot of passenger Sex
36  ggplot(passengers, aes(x = Sex), size= 0.5) +
37    geom_bar()
38
39  # Scatter plot of Age vs Fare
```



ACHIEVEMENTS

- R markdown

```
248 ```{r = experimental data analysis, results='asis',echo=FALSE,
    warning=FALSE, message=FALSE}
249 library(tidyverse)
250 project_2 <- read_csv("/Users/fxy1996/Desktop/Biometry/project
    2.csv")
251 summary(project_2)
252
253 aov1 <- aov(Sips~factor(People)+Drink, data=project_2)
254 aov1
255 summary(aov1)
256
257 ```
258
259 ```{r=plot, results='asis',echo=FALSE, warning=FALSE}
260 p <- ggplot(data = project_2, mapping = aes(x = People, y =
    Sips,colour = Drink))
261 p
262
263 p1 <- p + geom_point()
264 p1
265
266 p2 <- p1 + labs(title="Sips set", subtitle="People vs Sips",
```

ACHIEVEMENTS

Group Project

Hypothesis: There is no difference of the number of sips between drinking Coca-Cola and water.

Expected experimental design

Expected experimental design

Research objectives, research design, definition of treatments, external and internal validity arguments

1. Research objectives

Identify whether it is true that for fizzy drinks a smaller package is required to provide the same number of sips as for still water. The volume of the coca cola and water should be the same (100 ml for each). And each student should use straw to drink coca cola and water. The brand and flavour of the drinks should be the same throughout the experiment. The target population is the students in Waite campus.

2. Research design

Treatments	Factors	Levels	Labels
2x1x1x1=2 treatments	Suppliers	2	
	Type of coke	1	No sugar coca cola
	Type of water	1	Distilled water
	Drink size	1	100 ml

Subject selection: In this experiment, the casual employment should be over 16 years old with health body, for example, they should not have diabetes. The samples are students, which chosen from the Waite campus. The gender are equitable that 10 male and 10 female are selected.

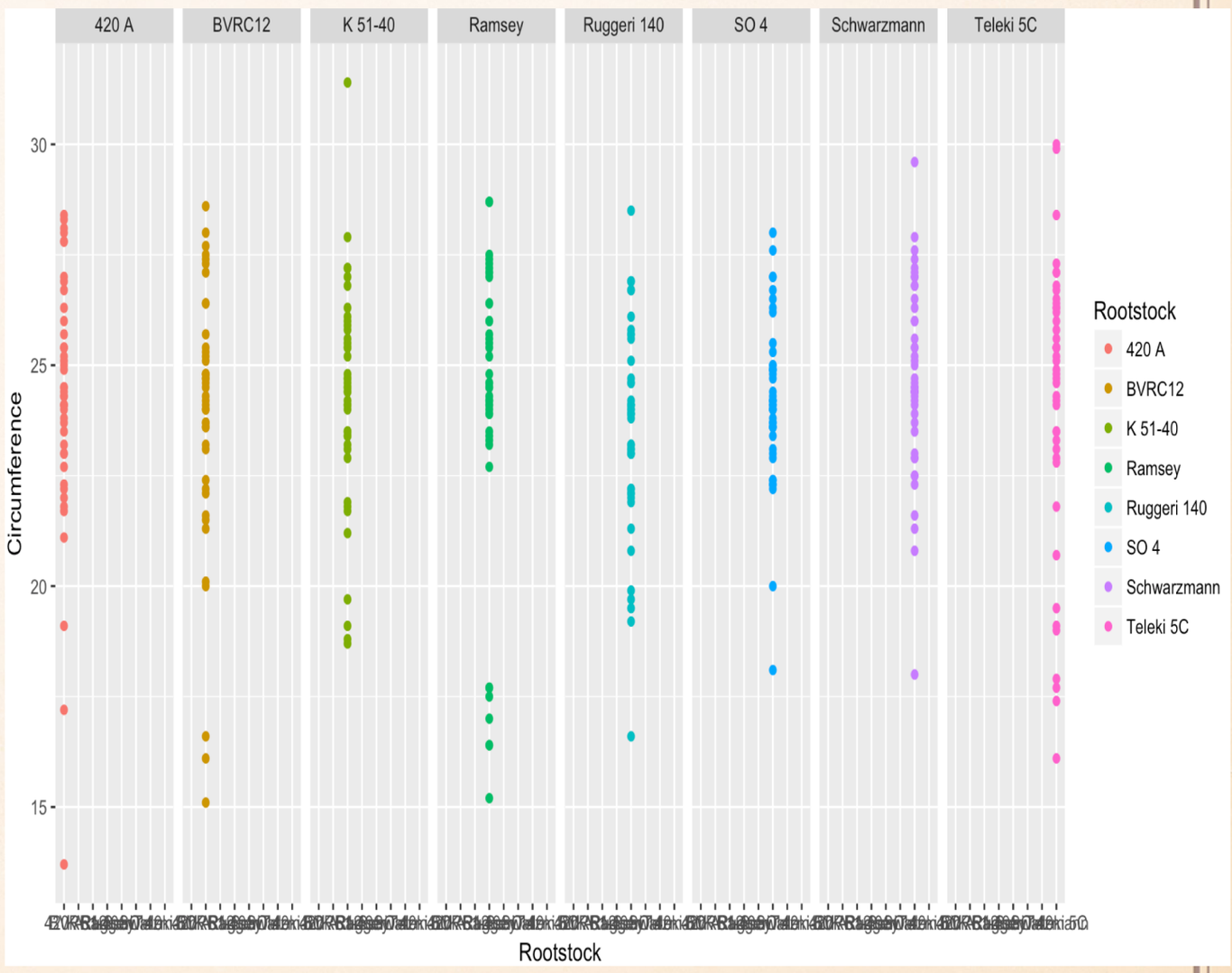
3. Definition of treatments, steps and ingredients

Definition: a set of conditions applied in a controlled and reproducible way to experimental material. And the treatments can be composed of several factors. In this experiment, coca cola, water, straws and cups are used. Step 1, we will randomly choose 10 male and 10 female in Waite campus. Step 2, supply either 100ml

ACHIEVEMENTS

- Dataset

```
aov3 <- aov(Circumference~Rootstock, data=Trunk)
aov3
aov4 <- aov(Circumference~Row, data=Trunk)
aov4
aov5 <- aov(Circumference~Panel, data=Trunk)
aov5

                    aes(x, y, ...)
####Rootstock
p <- ggplot(Trunk, aes(x = Rootstock, y = Circumference, color = Rootst
  geom_point()
p

p1 <- p+labs(title="Trunks", subtitle="Rootstock vs Circumference", y="

ggplot(Trunk, aes(x = Rootstock, y = Circumference, color = Rootstock))
  geom_point() +
  facet_grid(~Rootstock)

#### Row
p2 <- ggplot(Trunk, aes(x = Row, y = Circumference, color = Row)) +
```



ACHIEVEMENT

- Class

How to design an experiment

Select and control factor

Estimate sample size

Randomisation

Types of sampling

Source of variability

Animal genetic experiment: Linear mixed model

ACHIEVEMENT

- Presentation and tour

Know some present agricultural areas that are researching in Australia.

How biometry hub provides support for agricultural experiment data analysis.

How researchers use statistics in their experiment

SKILL UTILISED

- R language

tidyverse, drlyr, tibble…

WHAT DID I LEARN

- Soft skills

Utilisation of R language

- Statical skills

Reasonable design experiment

Analyse experimental data

INTERNSHIP IMPACT

- Long term advantages
- Critical thinking and analysis
- Easier to work in a team

CONCLUSION

- The internship in biometry is a great opportunity to improve personal and professional skills.
- It helps me prepare my further career.
- It is an unforgettable and valuable experience.
- Grateful to all staff in Biometry Hub for their patient help.

THANK YOU