

Biometry Hub Internship Presentation

BY ZHAOHAN (ELIJAH) LUO
24/07/2020

A Little About Myself

- International students from Xi'an, China
- Have been in Australia for more than 5 years
- Surf Lifesaver at Henley
- Member of University of Adelaide Boat Club
- Swimming coach

Week 1 - What I Have Done/Learnt

Rstudio

- Basic coding
- Function packages
- Data frames
- Graphics

Tidye verse / loops

- Sort complex data
- Application of loops

Experimental design

- Different types of experimental design and when to apply them

Rmarkdown

- Report writing

Reflection of Intro to R

- More careful when I am doing basic coding.
 - At the beginning, there were a lot of mistakes made due to wrong spelling or a lost comma
 - Keeps notes in the code
- Problem solving skills is significantly improved
 - Used to ask for help immediately after encountering an issue
 - Now I would look for issues myself first before asking for help
- I now understand how coding works

Reflection on talking with Russell about Functions

- The most important topic in coding is to manage the complexity of the code
- The code needs to be sorted in order
- Leave comments after codes
- Give functions descriptive names

Experimental Design

- Degrees of Freedom
- Four main experimental designs
 - Completely randomized design (CRD)
 - Randomized complete Block design (RCBD)
 - Latin square Design (LS)
 - Split-plot design
- Treatment Structure
 - Crossed
 - Nested

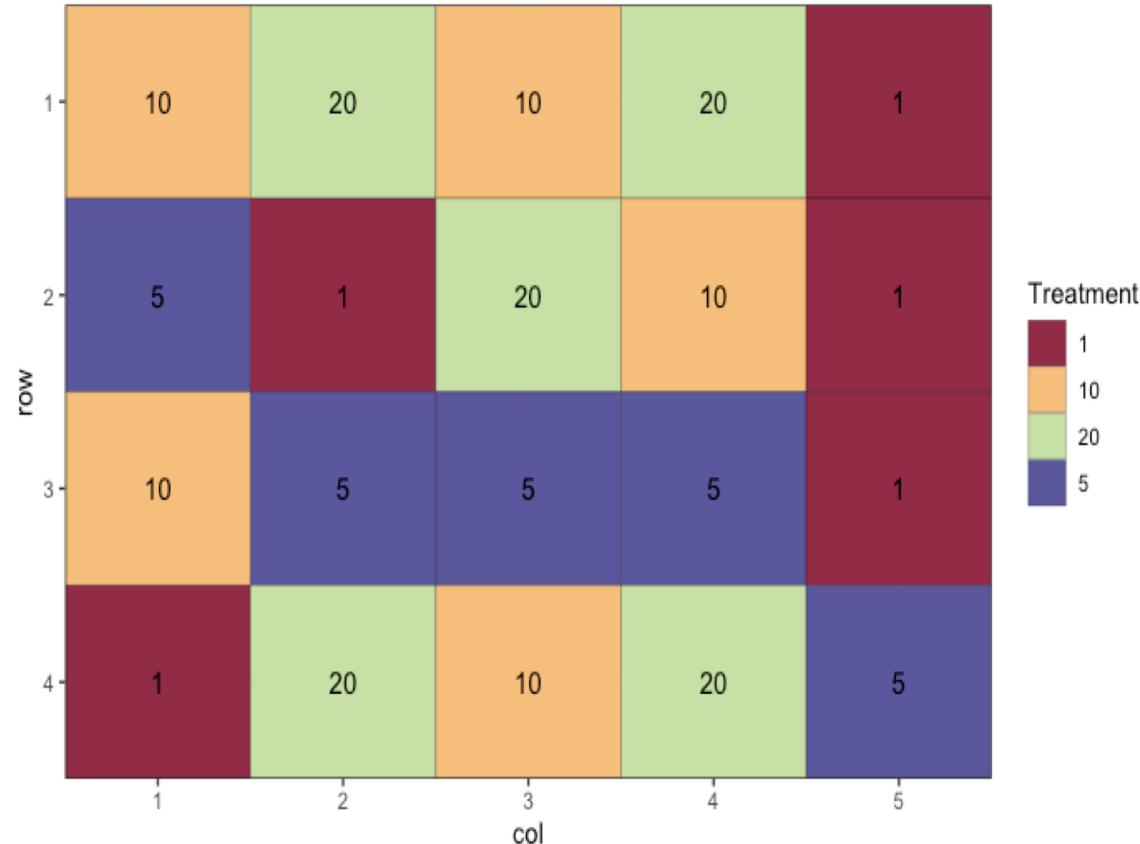
Degrees of Freedom

Experimental design

Treatment Structure

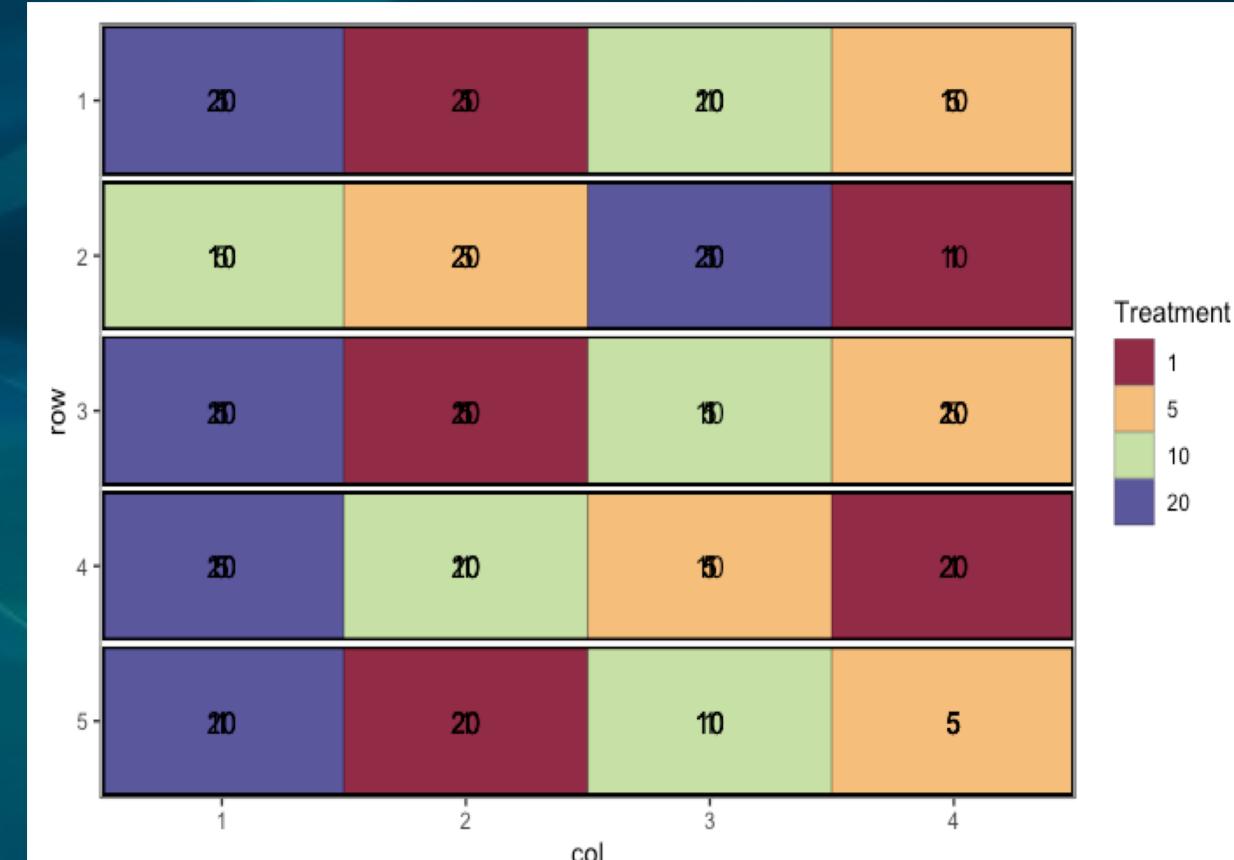
Completely Randomized Design

```
trt<-c(1,5,10,20) # the treatment levels  
rep<-5 # replications  
outdesign<-design.crd(trt, r=rep) # create design  
design.out<-des.info(design.obj = outdesign, nrows = 4, ncols =5)  
...
```



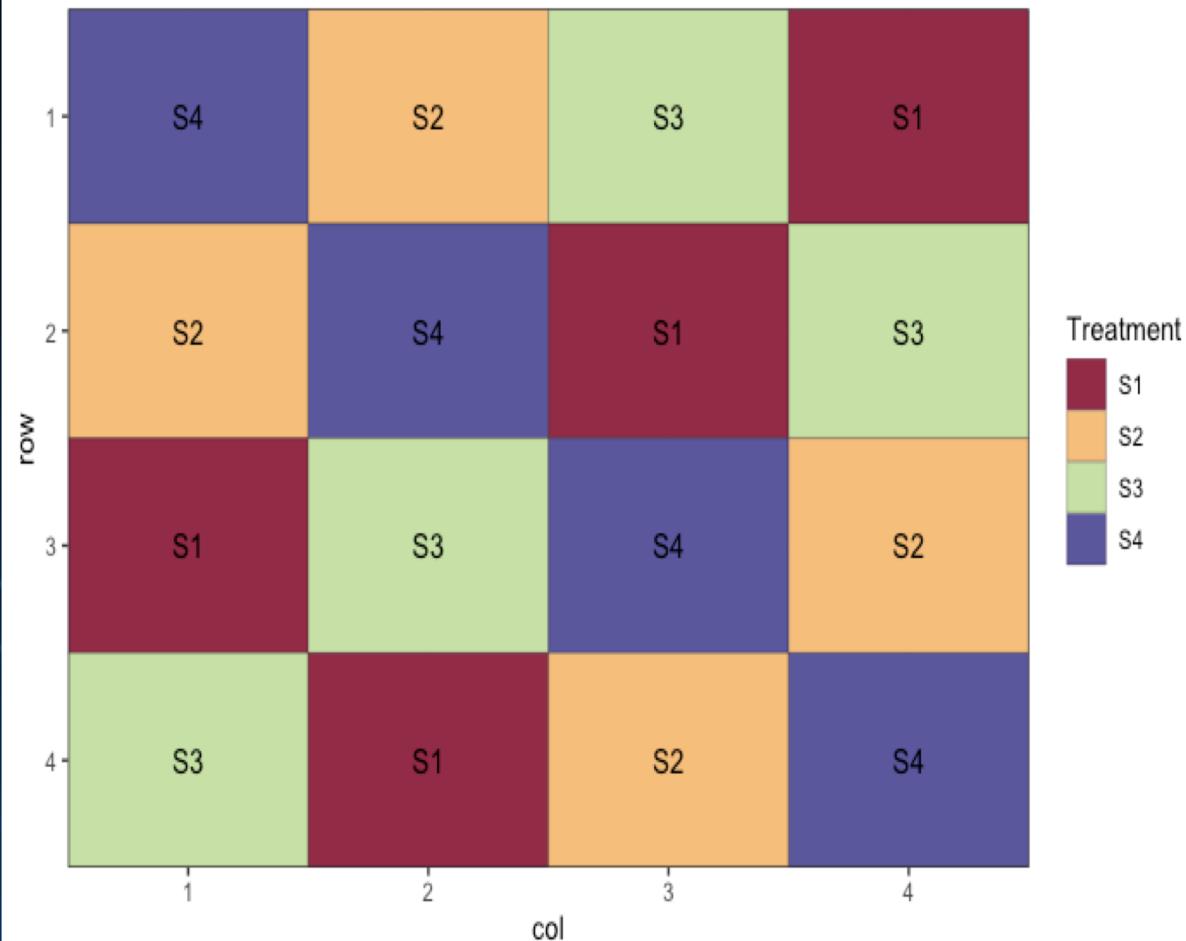
Randomized Complete Block Design

```
trt<-c(1,5,10,20) # the treatment levels  
rep<-20 # replications  
outdesign<-design.rcbd(trt, r=rep) # create design  
design.out<-des.info(design.obj = outdesign, nrows = 5, ncols =4,  
brows = 1, bcol = 4)
```



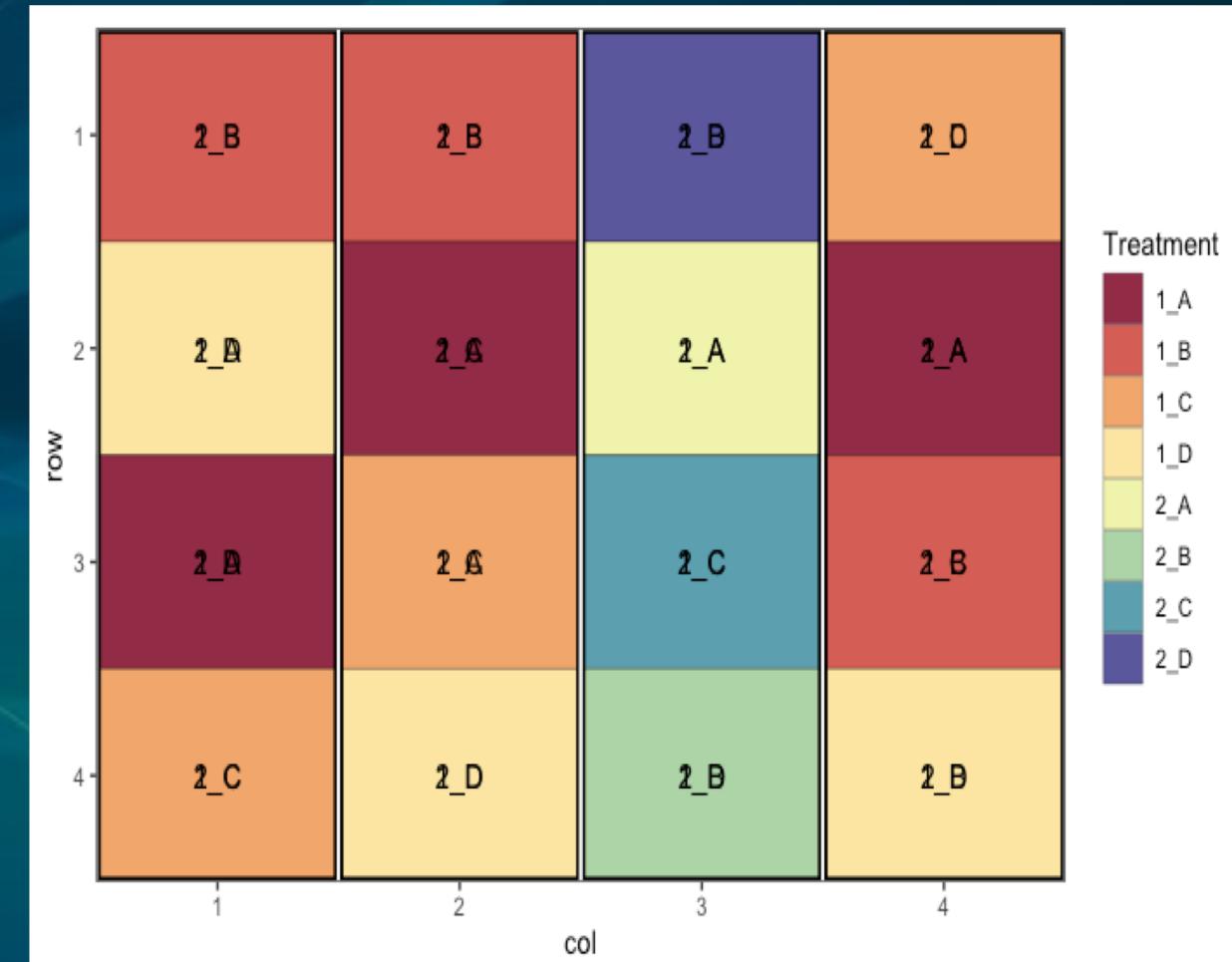
Latin square Design

```
trt<- c("S1", "S2", "S3", "S4")
outdesign<- design.lsd(trt)
des.out<-des.info(design.obj = outdesign, nrows = 4, ncols = 4)
```



Split-plot design

```
wholeplot<- c(1,2)
subplot<- c("A", "B", "C", "D")
outdesign <- design.split(wholeplot, subplot, r=4)
des.out <- des.info(design.obj = outdesign, nrows = 8, ncols = 4, brows = 4, bcols = 1)
```



Residuals Degrees of Freedom

- The rule of thumb of residual degrees of Freedom is 12 of an experiment
- More residual degrees of freedom, more accurate the analysis will be
- Choosing the right experimental structure is important as more complex experimental structure will reduce residual degrees of freedom

Reflection after doing experimental design

- It is important to choose the right design according to the subjects and variables tested
- I now understand the experimental design structures much better
 - Completely randomized design does not constrain the random allocation of treatments
 - Randomized complete block design, the number of experimental units in each block has to equal to the number of treatments (I did not know that before)
 - Latin square design's rows and columns must correspond to the number of treatment levels
 - Writing down for the aims, observations, arrangements, replicates, designs and blocking
 - It is best to draw the design out before coding in Rstudio

Rmarkdown

How to write reports in Rmarkdwon?

Open Rstudio, click the button on the top left of the screen

Choose Rmarkdown, then choose document, then word

Give the document a title and type in the author's name

Put '#' in front of the main heading

Put '##' in front of the sub-heading

Put '*' in front of dotpoints

Press 'return' after finishing the dot point then press 'tab' then '-' to add sub-dotpoint

Find and click 'Insert' in Rmarkdown script to insert tables or graphs

The screenshot shows the RStudio interface with three tabs open: "Week 1 Summary.Rmd", "model", and "Rmarkdown study.Rmd*". The "Rmarkdown study.Rmd*" tab is active and displays the following R Markdown code:

```
16 # Monday ----- Heading
17
18 ## Basics: ----- Sub-heading
19
20 * R studio requires accurate typing, word for word
21
22 * Run codes first after entering
23
24 ## Variables:
25 * Numerical: A <- 2, means "A" gets value 2, where A is the variable
26
27 ````{r}
28 A <- 2
29 ```````
30
31
32 * Text: city <- "Adelaide", means "city" gets "Adelaide", where city is the variable and there has
33 to be quotation marks around the text, such as "text".
34
35 * Important: the variables can be named as anything, therefore it is important to give it a
36 special/specific name.
37
38 ## Vectors:
39
40 * No matter numerical or text, c( ) needs to be employed to create/ start the vector, such as
41 - Vec.num <- c(1,3,2,7,9,15), or
```

Dot
point

R
script

Sub-
dotpoi
nt

The table of contents on the report

Table of Contents

Monday.....	2
Morning.....	2
Intro to R.....	2
Afternoon.....	3
Tuesday.....	4
Morning.....	4
Intro to R.....	4
Afternoon.....	9
Wednesday.....	9
Morning.....	9
Experimental Design.....	9
Afternoon.....	10
Experimental Design.....	10

Content of the report

Monday

Morning

Intro to R

Basics:

- R studio requires accurate typing, word for word
- Run codes first after entering

Variables:

- Numerical: `A <- 2`, means "A" gets value 2, where A is the variable
`A <- 2`
- Text: `city <- "Adelaide"`, means "city" gets "Adelaide", where city is the variable and there has to be quotation marks around the text, such as "text".
- Important: the variables can be named as anything, therefore it is important to give it a special/specific name.

All connected by hyperlinks

Reflection after doing Rmarkdown

- I learnt that reports can be written in Rstudio, which is much simpler and has more functions.
 - Will be very handy in the future and create well-structured and nice-looking reports
- The structure of the report is crucial for viewers to read and understand the report easily. Therefore, the structure needs to be clear

Personal Development

- Independent thinking skills
 - Search codes and packages online to solve problems
- Logical thinking
 - I am now better able to think about an issue step by step, rather than jump quickly from one idea to another
- Patience when coding
 - More patient to look for mistakes in codes and start over to identify errors, as well as trying different ways to solve the problem, rather than getting frustrated and stop thinking

Week 2 – What I Have Done/ Learnt

gg plots with Pete

- gg plot package
- Application of gg plots

Attending Hot Desk meeting with Julian and Annie

- Client talk about his experimental design
- Give out advises according to the information he provided

Exact Permutation test with Wendy

- Assume something is right, then prove it wrong

Attending StatsPD@Waite

- It is important to understand the experimental design structure
- A design needs to be analyzed in different ways to check consistency

Bioinformatics with Mario

- Hearing about information about different methods of analysis in genetics

Honors project talk with Meixuan

- Listen to her talk about her honors project

Flying drones with Pete

- Listen to Pete talk about his project – which is about machine learning

Zoom Meeting with Enqi

- Enqi talked about his honors project and his working experience after honor.

Website Building with Sam

- Learnt how to use blogdown in R
- Website building basics and uploading the website

gg plots

It is an effective tool with multiple functions to present data in an organized and logical way

```
install.packages("ggpubr")
install.packages("devtools")
library(ggpubr)
```

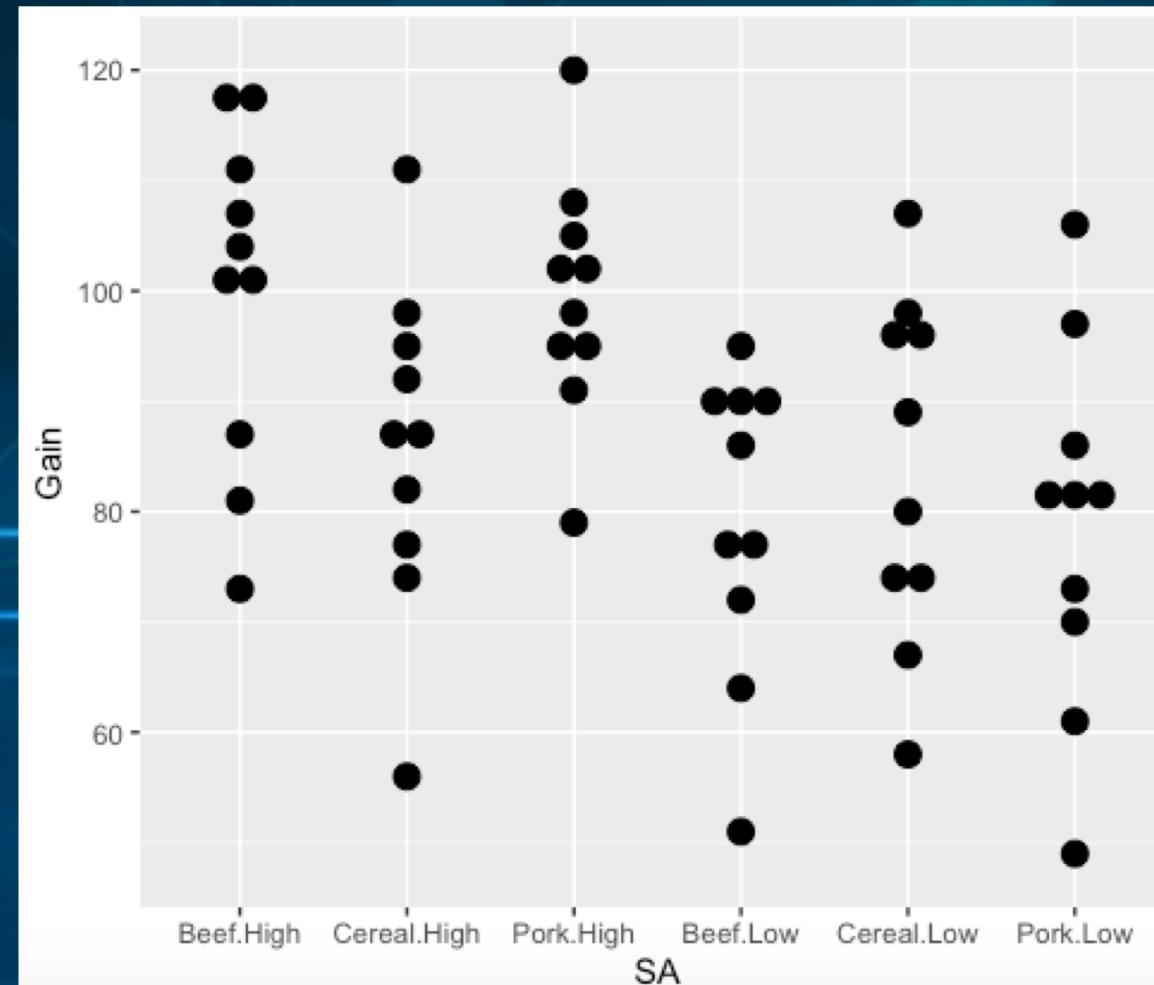
Always install the packages first

Reflection

- Always google when having issues finding the right code
- There are multiple different codes to achieve the same result
- Grouping is very important in gg plot

Scatter plot from rat factorial

```
df=subset(rats, select = -SNA)
rats<-df
rats
dplot1<-ggplot(rats, aes(x=SA, y=Gain)) +
  geom_dotplot(binaxis = 'y', stackdir = 'center')
```



StatsPD@Waite

- It is about a two-phased design experiment
 - Requires more raters in the second phase
 - The treatment effects depend on the block
 - The effect can be tested by combining observed MS to get the 'right' error variance or put the wright model in ASREML
- What I have learnt when analyzing data:
 - Understand the experimental design structure
 - Use simplified examples
 - Analyze data in different ways to check consistency
 - Look at the subsets of the data
 - Use ANOVA to develop formal model
 - Use ARSEML for final analysis

Reflection After the meeting

- Important to discuss your experimental designs and results with other statisticians to identify any potential errors and mistakes that you did not see, as well as bouncing new ideas
- The research in statistics can be challenging because it is easy to have flaws in the experimental design, such as mix in confounding variables or do not have enough treatments
- Every statistician need to have clear understandings of every decision of the experiment

Bioinformatics

- Multi-dimensional scaling analysis
- Regulatory-Network Analysis
- Software thresholding
- Topological Overlap Matrix
- Cluster-One modules
- Variable selection
- Machine learning

Reflection

- A lot of machine learning is involved in statistics
- Statistical concepts can be hard to understand especially when it is applied into other fields of science

Drone Flying

Gave me an insight into what future sampling and data collecting would look like



Web design

Open Rstudio,
choose 'File',
click 'New Project'

Choose
'New
Directory,
then choose
'Website
using
blogdown'

Name the
directory,
then e,
copy it to
'Hugo
theme'

Start editing
the
template

Copy it to
'Hugo
theme'

Chose a
theme from
the Hugo
website

My Web page: <https://zhaohanluo.netlify.app/>



Zhaohan (Elijah) Luo

Bachelor of Agricultural Science
University of Adelaide



Biography

Zhaohan (Elijah) Luo is a Chinese international student who's currently studying in the University of Adelaide. He is finishing off his final year of Bachelor of Agricultural Science degree and looking forward to doing an Honour's project with the Biomtery Hub of the university. Zhaohan has been in Australia to live and study for more than 5 years and has established many interests not only in academics, but also in athletics. He is a swimming coach, a surf lifesaver at Henley SLSC and a member of University of Adelaide Boat club.

Interests

- Horticulture
- Plant Genetics
- Experimental Design
- Data Analysis

Education

- 🎓 Certificate of Secondary Education, 2017
Adelaide High School
- 🎓 Bachelor of Agricultural Science, 2020
University of Adelaide

Reflection

- It is exciting and challenging because I have never thought of creating a web page before
 - Stop setting myself limits by saying I can't do something because I have never done it
 - Keep trying even after many failed attempts
- Experimenting is an important part of self-learning
 - I am used to doing what I am told and did not have much interest in exploring
 - It is a faster way of learning
- Important to save regularly and have backups
 - Sometimes if I stuff up too much I would have to start over

Week 3

Beata's Talk

Statistical approaches for the detection of genetic basis of agriculturally relevant traits

Paul's Talk

His experience of more than 20 years moving around the university in between different groups.

Olena's Talk

Honors project, summer internship, her aims of imbedding the R course into undergraduate studies

Beata's Talk

- Two analysis method:
 - The talk included QTL analysis, which is for the analysis of quantitative trait loci for bi-parental populations
 - Genome-wide association analysis, which is for panels of genotype carrying diverse germplasm
- Two approaches to QTL mapping
 - Wgaim – can model extraneous variation that's relevant to plant field trials and allows whole genome QTL analysis
 - Bayesian approach – gives out graphs to filter the strongest signals from samples

Reflection On Beata's Talks

- Statistics can be applied in almost every scientific field and there are many different methods to analyze data collected from the experiment
- However, the effectiveness of analytic methods is very much depending on the situation and what researchers are trying to identify. Therefore, it is important to try different methods when analyzing data to determine which is the best

Paul's talk

- More than 20 years of experience as a statistician
- Worked with different groups of people on Waite campus
 - Advantages of working with statisticians: bouncing new ideas around
 - Advantages of working with different groups: getting to know everyone
- It is important to consult with biometrists throughout a project rather than later towards the finishing of a project

Reflection on Paul's talk

- It is important to have good communications with statisticians throughout the whole project
- Working with different people have both advantages and disadvantages. Therefore it is important to develop different skills to minimize those disadvantages

Olena's Talk

- The structure and requirement of the honors project
- Importance of including R coding and statistics into every year of the agricultural science degree
- The benefit of doing honors project and the challenges

Reflection on Olena's talk

- Made me realize the importance of data analysis skills in the agricultural industry
- Many employer actually prefer students with statistical knowledge
- Doing honour's project is important in job seeking

Overall Reflection of Myself Throughout the Internship

- Through this internship, as listed in previous slides, I have learnt and developed multiple important skills which I did not have. And I believe me that those skills will be very beneficial to me in both of my studying and working careers.
- I also realized that I am a slow learner due to language difficulties and non-computing background. However, I learnt that even though it takes me a longer time to finish tasks, I will get the results if I keep trying.

Reflection on the internship

- I entered this internship without any previous coding skills and this internship involves a lot of self-learning, which can be challenging. However, I think this is good practice for students who are lacking the self-learning and independent thinking skills to develop those skills which will help them in the future. The working environment is very friendly and reduced my stress.
- However, I did not understand many parts of the talks (such as bioinformatics, exact permutation tests and detection of genetic basis of traits) I received from different statisticians due to language difficulty and relatively shallow statistics related knowledge. Even though it is good to receive different knowledge from different people to develop my overall understanding of statistics, I hope in future internships, students can have more time to practice the foundations of coding and grasp lower level concepts of statistics in different fields.
- Also I hope the mentors can speak a bit slow so people can have more time to digest those big words they just heard before moving on

At Last.....

I would like to thank everyone who I have worked throughout this internship. Thank you for being patient when introducing new knowledge and help me out when I need it. I hope to work with all of you again in the future!!