Biometry Hub Internship

By Ziyang Wang

01

Week 1

First glance

R-Studio

Introduction to R and RStudio

R package-ggplot2

Plot construction and customize layout

Experimental Design

Definitions and terms, different design types and practices

Communication

Attending Peter's presentation, and statistical meeting with Jing and Shiyu

R and RStudio

The introduction

RStudio 08/Jul/2019

- R Studio user interface
- Basic definitions (variables, packages, functions)
- Exploring data frames
- Data management
- Graphics

Learn More

Learn More

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

Test.R | Rintroduction.R | Genstat.R | workbook1.R | 08-reactiveValues.R

Source on Save | Run | Source

```
1  2+3+2
2  #Practice
3  #Practice1
4  8/2
5
6  #2
7  (210*1290)/(sqrt(1367^3))
8
9  #Find the answer
10 sqrt(9)
11 log(20)
12 exp(20)
13 abs(-10)
14 round(3.1298376198712,digits=5)
15
16 #Find the volume of a water tank
17 pi*(1.19^2)*4.28
18
19 2*a
20
21 #Practice
22
```

Calculations

9:17 (Top Level) R Script

Console | Terminal | Jobs

```
[1] 4
>
> #2
> (210*1290)/(sqrt(1367^3))
[1] 5.359894
>
> #Find the answer
> sqrt(9)
[1] 3
> log(20)
[1] 2.995732
> exp(20)
[1] 485165195
> abs(-10)
[1] 10
> round(3.1298376198712,digits=5)
```

Environment | History | Connections

User interface

Import Dataset | Global Environment

Values

city	chr [1:3] "Chengdu" "Queensland" "Tok
col	"unique"
h	num [1:5] 2 4 6 8 10
r	num [1:5] 1 2 3 4 5
T	int [1:5] 1 2 3 4 5
V	num [1:5] 6.28 50.27 169.65 402.12 78
vec.rep	chr [1:9] "Z" "Y" "W" "Z" "Y" "W" "Z"
x	num [1:20] 7.36 65.84 5.57 46.43 27.4
z	num [1:3] NA 3 5

Files | Plots | Packages | Help | Viewer

R: Data Input | Find in Topic

read.table {utils} R Docu

Data Input

Search for help

Description

Reads a file in table format and creates a data frame from it, with cases corresponding to lines and var fields in the file.

Usage

Packages

```
read.table(file, header = FALSE, sep = "", quote = "\"'",
           dec = ".", numerals = c("allow.loss", "warn.loss", "no.loss"),
           row.names, col.names, as.is = !stringsAsFactors,
           na.strings = "NA", colClasses = NA, nrows = -1,
           skip = 0, check.names = TRUE, fill = !blank.lines.skip,
           strip.white = FALSE, blank.lines.skip = TRUE,
           comment.char = "#",
           allowEscapes = FALSE, flush = FALSE,
           stringsAsFactors = default.stringsAsFactors(),
           fileEncoding = "", encoding = "unknown", text, skipNul = FALSE)

read.csv(file, header = TRUE, sep = ",", quote = "\"",
         dec = ".", fill = TRUE, comment.char = "", ...)

read.csv2(file, header = TRUE, sep = ";", quote = "\"",
          dec = ",", fill = TRUE, comment.char = "", ...)

read.delim(file, header = TRUE, sep = "\t", quote = "\"",
           dec = ".", fill = TRUE, comment.char = "", ...)
```

ggplot2

A handy R package

ggplot( )+layers

- Graphs creation
- numerical & categorical data
- Customization- color, symbol, size, and transparency.

Example plot 1

Example plot 2

Microsoft Office

The consequences of today are determined by the actions of the past. To change your future, alter your decisions today.

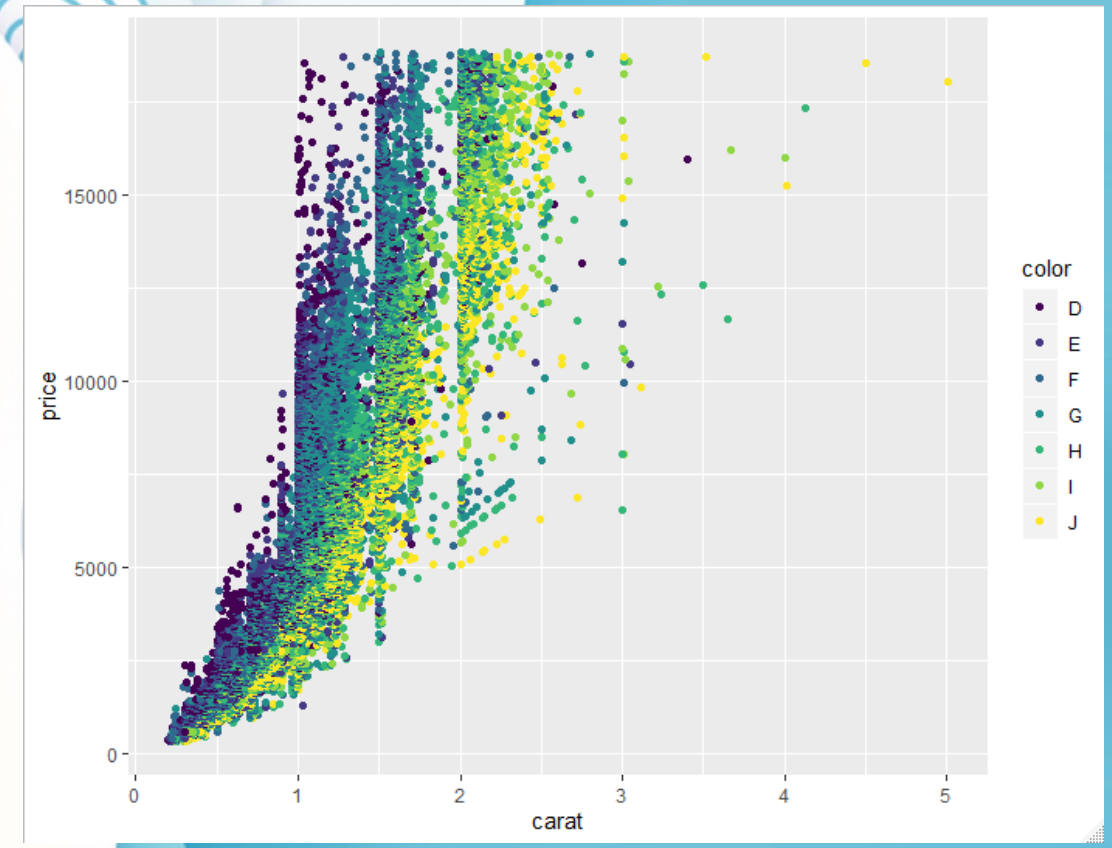
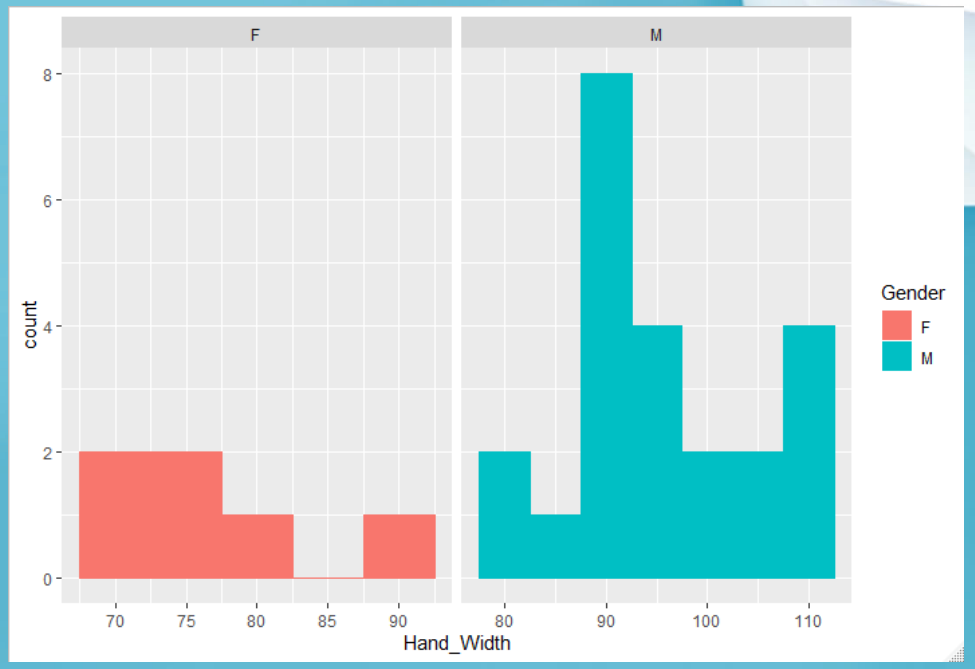
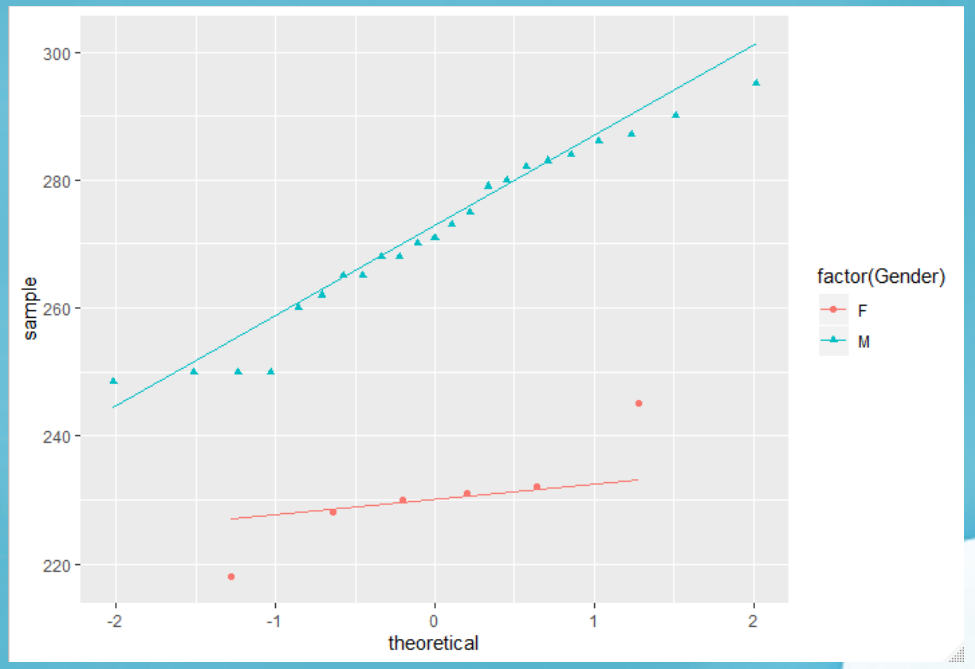


Figure 1. A scatter plot generated by ggplot() function using in-build R dataset "diamond"

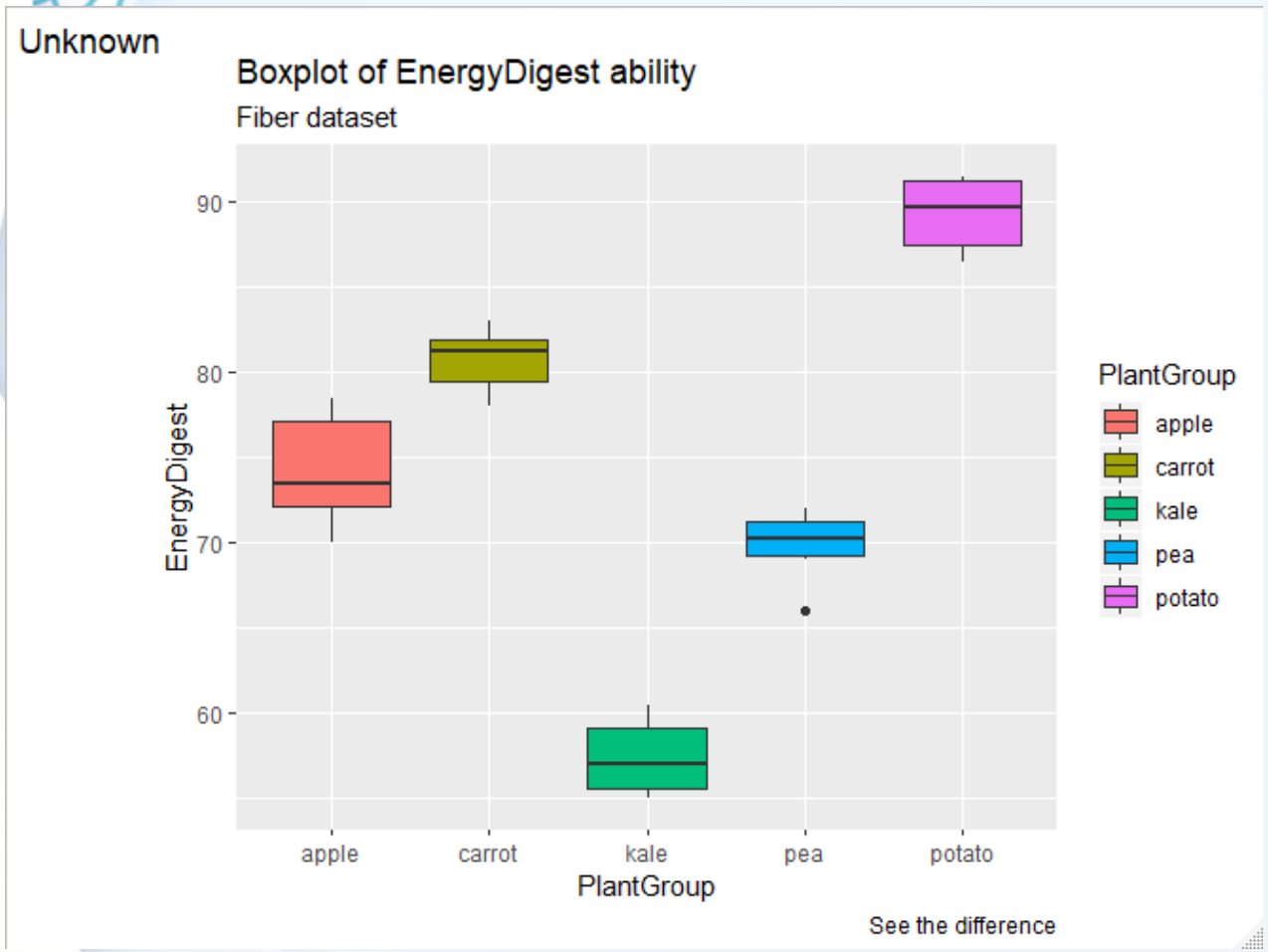
```
(p <- ggplot(data = diamonds,
mapping =
aes(x=carat,y=price))+
geom_point(aes(color=color)))
```



Histogram showing the hand width(mm) of females and males



Normal QQ-plot of the Foot length (mm) of females and males




```
fiber <- read.csv(
  file = "D:/Firefox download/RStudio/Energydigestability1.csv")

view(fiber)
```

```
(p0 <- ggplot(data = fiber,
              mapping = aes(PlantGroup,EnergyDigest)))

(p01 <- p0+ geom_boxplot(aes(fill=PlantGroup),
                         outlier.shape = 19,
                        outlier.size = 1.5))
(p02 <- p01 +
        labs(title = "Boxplot of EnergyDigest ability",
             subtitle = "Fiber dataset",
             caption = "See the difference",
             tag = "Unknown"))
```

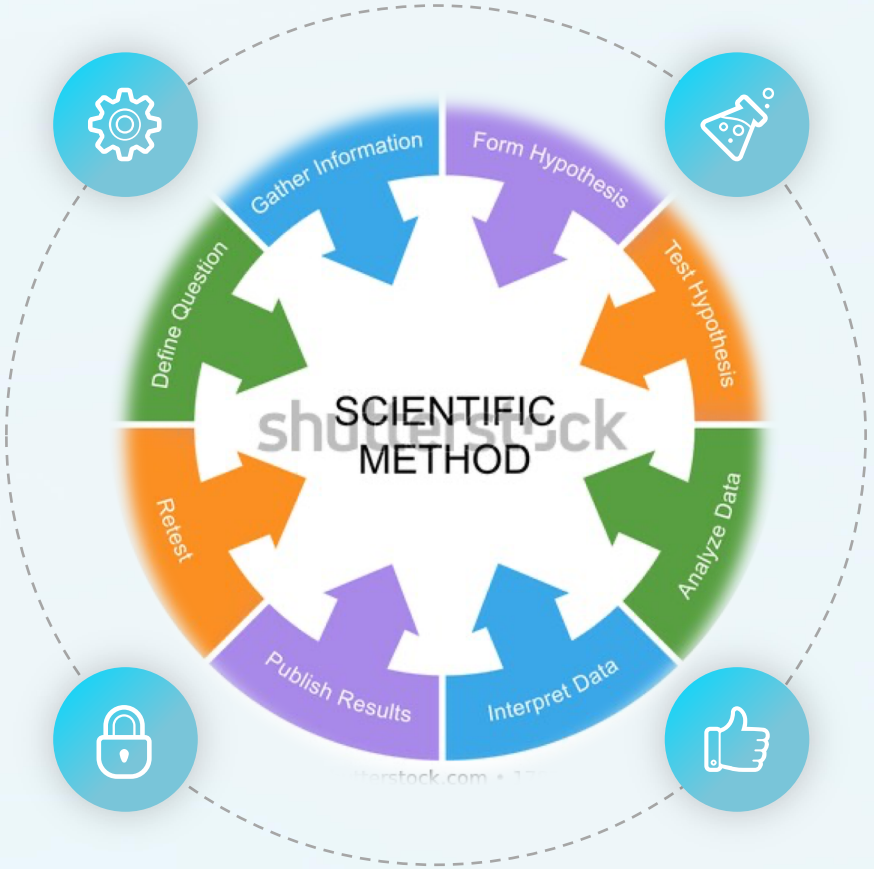
Boxplot showing the energy digestibility % for each plant diet

Agronomic experiment

The planning and designing

Terms and definitions

Populations, samples, treatments, experimental and observational units, replication



Designing experiments

CRD, RCBD, Latin Square, Factorial, Split-plot

Planning experiments

of treatment, levels of treatment, blocking, natural gradient

Design-agricolae R package

Use R coding to adopt your design
Residual degree of freedom

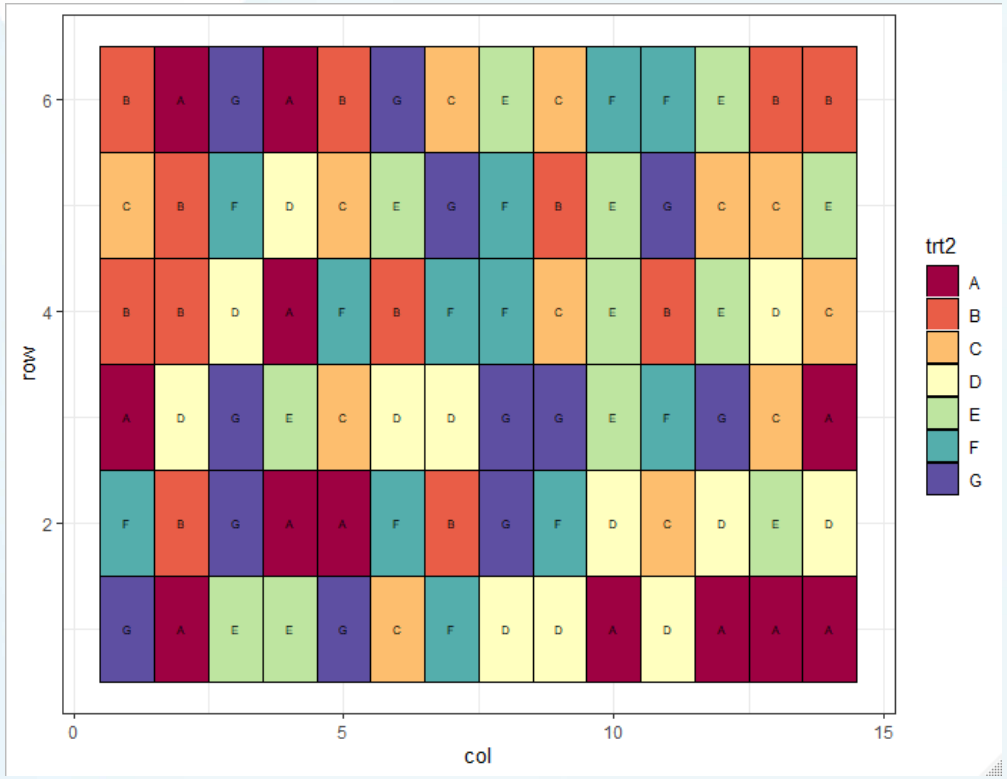


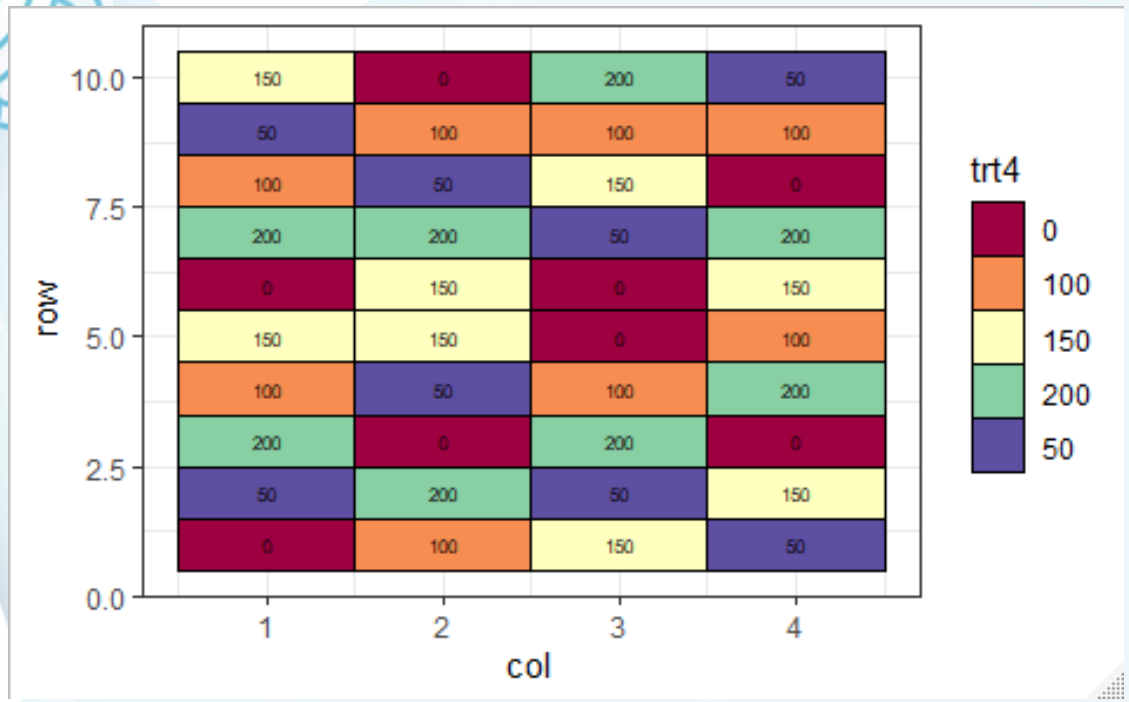
Figure 2. A completely randomized design

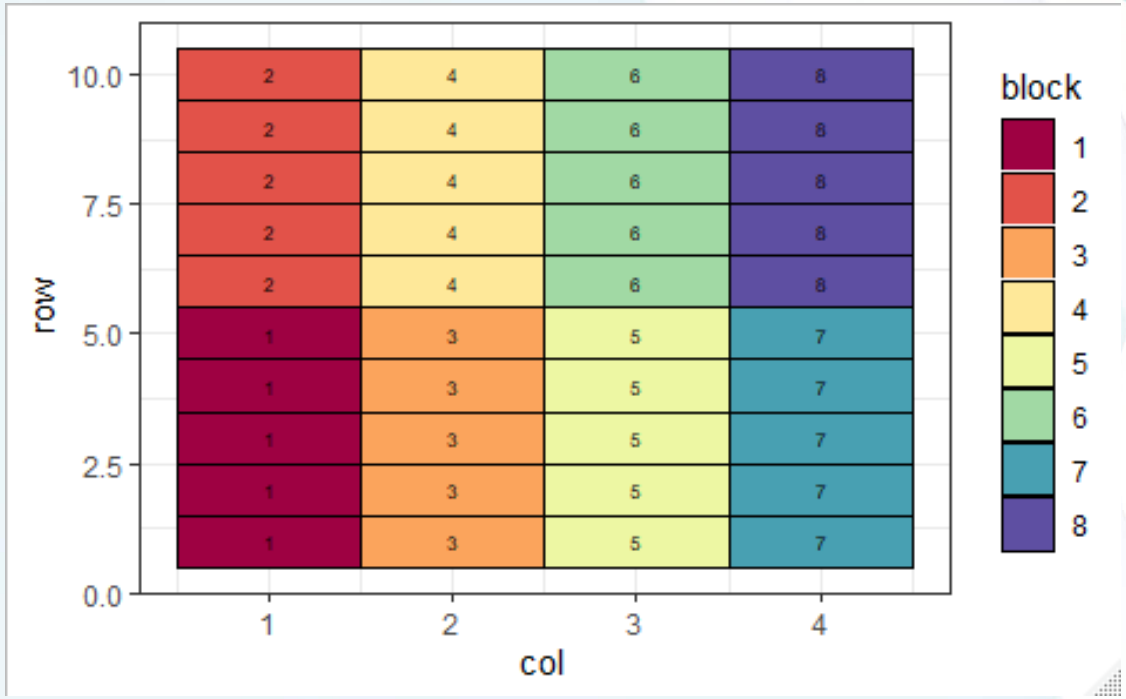
```
trt2 <- c("A","B","C","D","E","F","G")
rep2 <- 12
E2 <- design.crd(trt2, r = rep2)
des.e2 <- E2$book

plot.des(design.obj=des.e2, design="crd",
         nrows=6, ncols=14, plot.fac="trt2")
```

Source of Variation	df
trt2	6
Residual	77
Total	83

Figure 3. The skeletal ANOVA table

```
satab(design.obj = des.e2, design ="crd")
write.csv(des.e2, "e2.csv",row.names=FALSE)
```





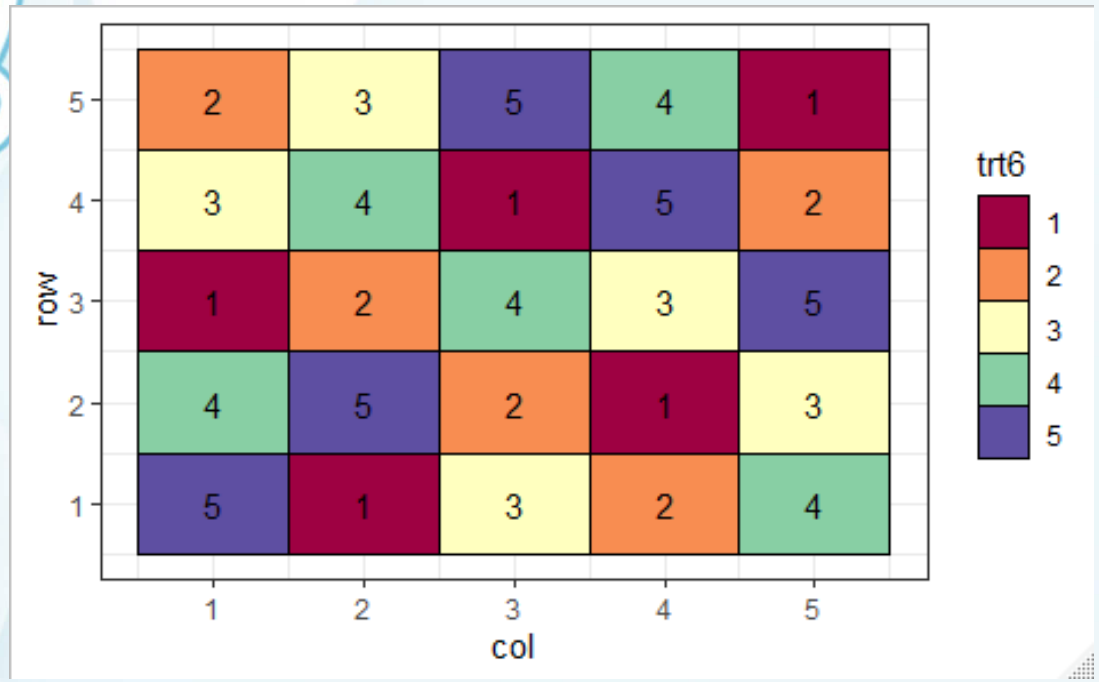

```
trt4 <- c("0","50","100","150","200")
rep4 <- 8
E4 <- design.rcbd(trt4, r = rep4)
des.e4 <- E4$book

plot.des(design.obj=des.e4, design="rcbd",
         nrows=10, ncols=4, plot.fac="trt4")
```

```
plot.des(design.obj=des.e4, design="rcbd",
         nrows=10, ncols=4, plot.fac="block")

satab(design.obj = des.e4, design ="rcbd")
```

Source of Variation	df
Block stratum	7
trt4	4
Residual	28
Total	39

Figure 4. A design frame based on Randomized Complete Block design

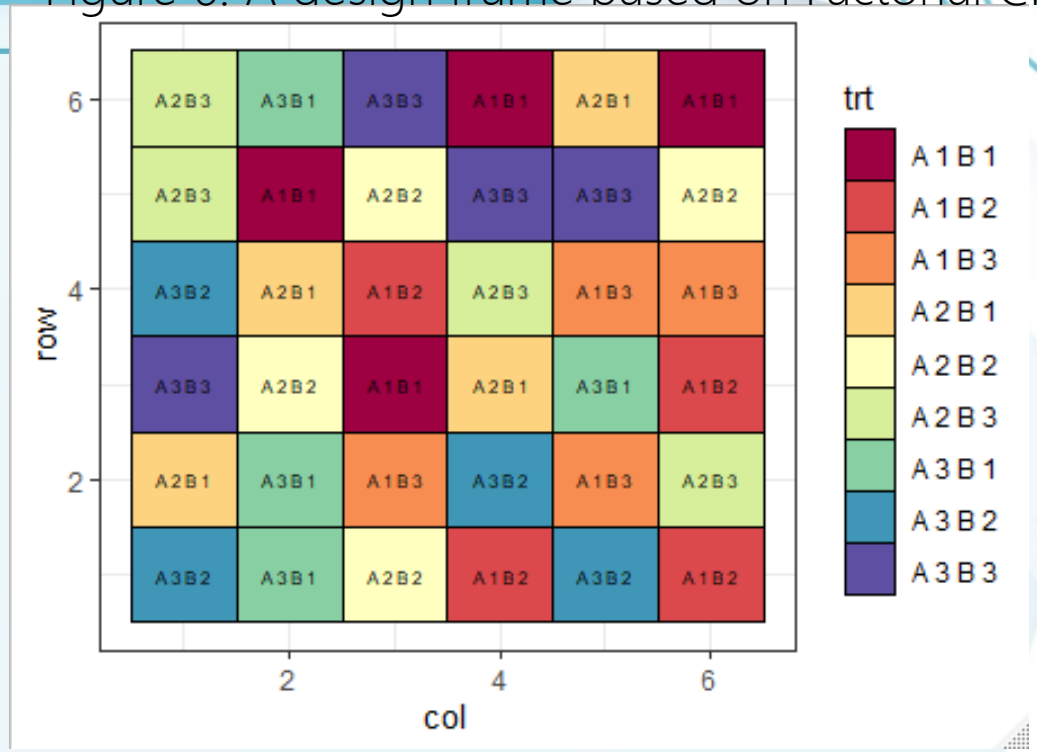
Figure 5. A design frame based on Latin Square




```
trt6 <- c("1","2","3","4","5")
E6 <- design.lsd(trt6)
des.e6 <- E6$book

plot.des(design.obj=des.e6, design="lsd",
         nrows=5, ncols=5, plot.fac="trt6")

satab(design.obj = des.e6, design ="lsd")

write.csv(des.e6, "e6.csv",row.names=FALSE)
```

Source of Variation	df
Row	4
Column	4
trt6	4
Residual	12
Total	24

Figure 6. A design frame based on Factorial CRD




```
trt7 <- c(3,3)
rep7 <- 4
E7 <- design.ab(trt7, r = rep7, design = "crd")
des.e7 <- E7$book

plot.des(design.obj=des.e7, design="fac",
         nrows=6, ncols=6, plot.fac="trt")

satab(design.obj = des.e7, design ="faccrd")

write.csv(des.e7, "e7.csv",row.names=FALSE)
```

Source of Variation	df
A	2
B	2
AB	4
Residual	27
Total	35

Meetings & Presentations

With biometry staffs

Statistical meeting is important for efficient communication in project, note taking skills are very important

Peter Josef Kasprzak

Presentation- Yield Estimation by Computer Program

Lachlan Mitchell

Digit recognition by MINIST

Jing and Shiyu

CSA project-plant pathogen WGCNA analysis-Transcriptional factor

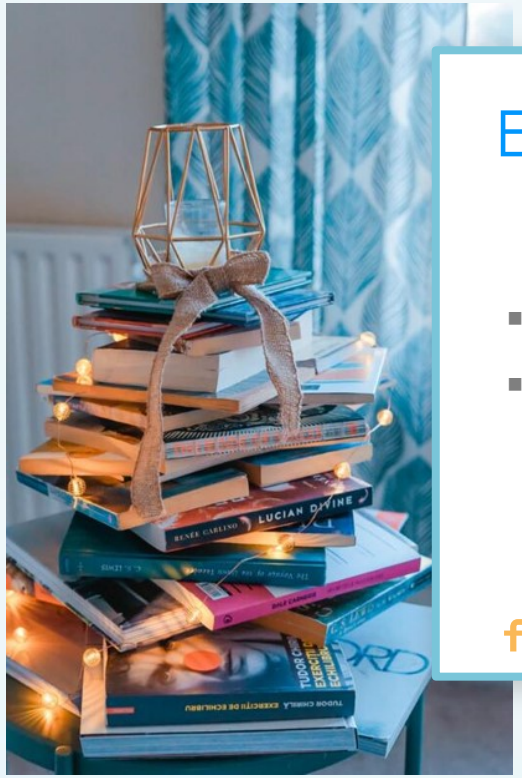
02

Week 2

Digging further

02 Week 2 overview

Further exploration



Exploring RStudio

- R vs GenStat 19th
- Shiny App

Statistical meetings

- Principles of experimental design
- Principles of statistical inference in practical applications
- The importance of sampling

02 Experimental design

Helena Oakey

Biofuel potential of barley straw

- From field biomass to ethanol – not economic
- Need more efficient sugar release – find the gene responsible for high sugar release

Spatial row-column design

- 648 elite varieties, 5 replicates
- no repeating of the same variety in the same row and column.

Field variation and Lab variation

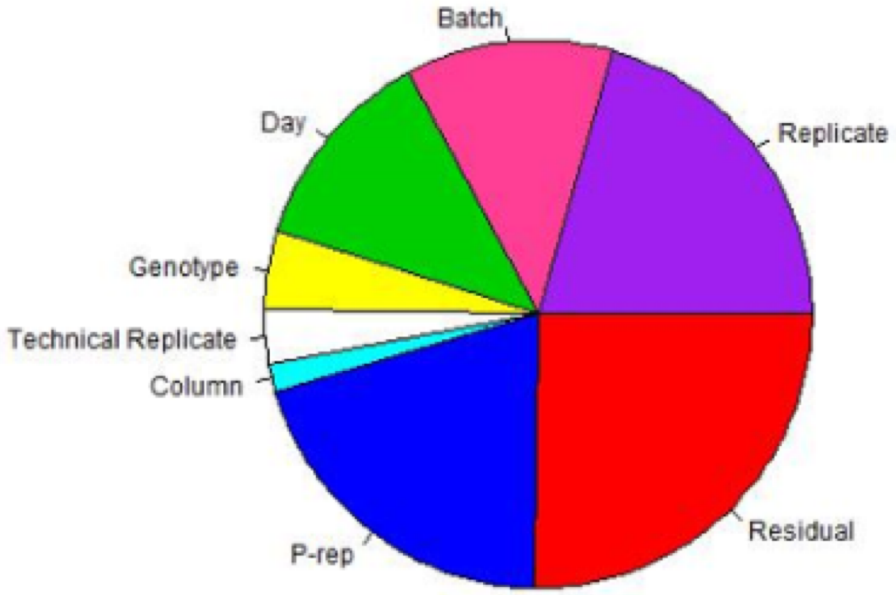
Samples re-randomized in the lab

Consider variation between various batches

Max - 20 varieties

Source of variation

In-lab re-randomization reduced the variation brought by varieties



TERM	%
Replicate	20.6
Batch	12.2
Day	12.3
Genotype	4.6
Technical Rep	3.2
Column	1.6
P-rep	20.2
Residual	25.3

02 Principles of statistical inference

Richard Jarrett

PDSA Cycle

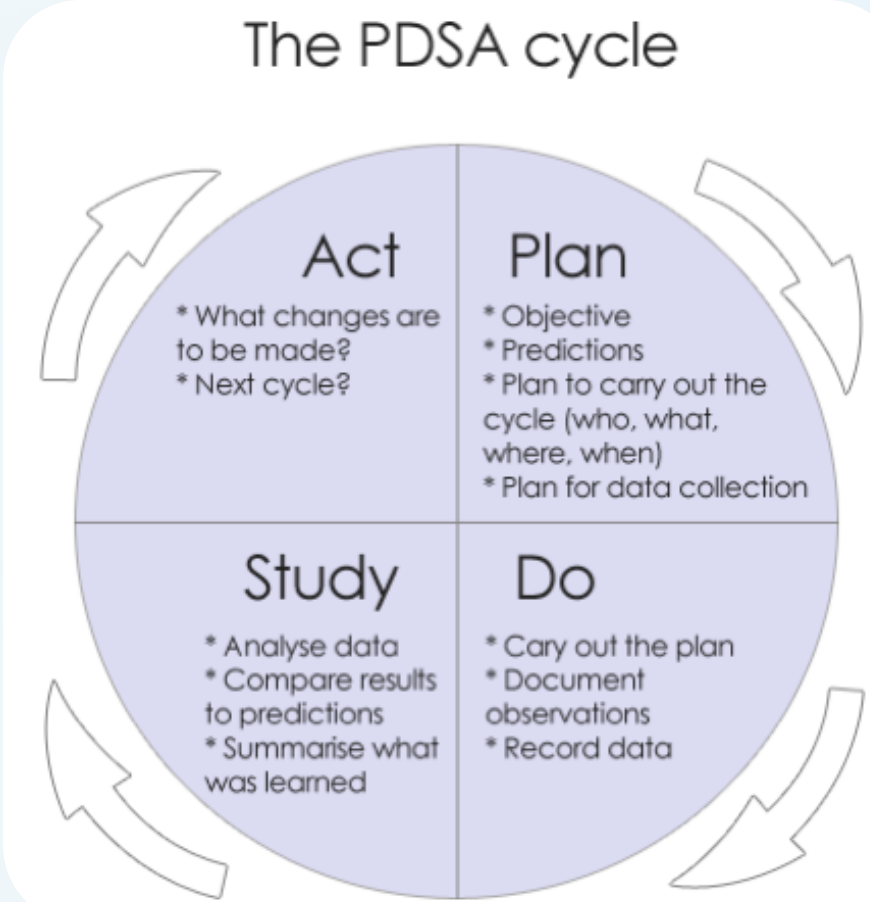
- Questions
- Scope
- Experiment conduction
- Analyze data and draw conclusion
- Implement changes

Accuracy of estimators

- Group means, SEs
- T-test

Blocking design

- grouping things that have similar result together



02 Sampling design

Peter Josef Kasprzak

Sampling and Randomisation

Why?

Infinite and Finite

design-based and model-based inference in survey sampling

Sampling types

Descriptive, analytical and pattern sampling

Importance to detect the errors of estimators, and use proper sampling protocol**s**

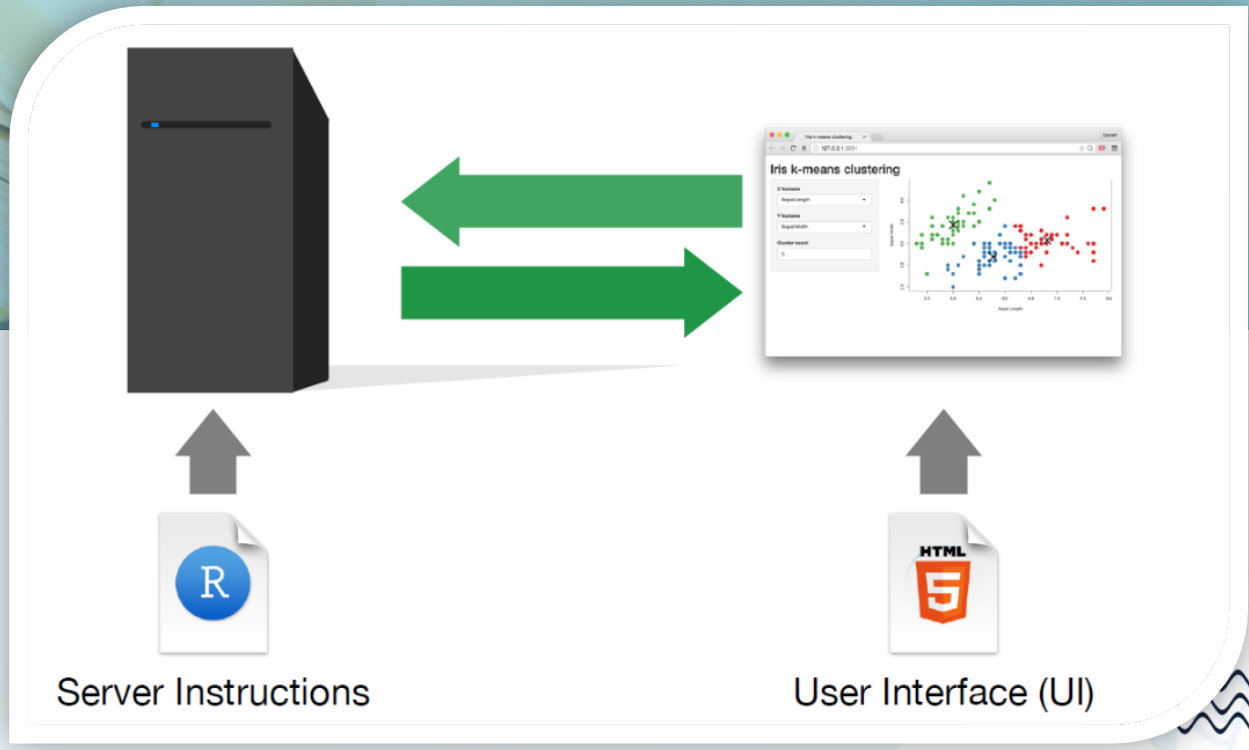
Central limit theorem & SLLN (Strong law of large numbers)

R Studio
Shiny app

Build interactive web apps straight from R

Shiny packages & Shiny dashboard

Customize with CSS and HTML



03 ShinyApp elements

App template

The shortest viable shiny app

```
library(shiny)
ui <- fluidPage()

server <- function(input, output) {}

shinyApp(ui = ui, server = server)
```

Input

Use input values with input$

Output

Save objects to display to output$

Display

Build objects to display with render*()


03
Week 3
Please enter your title

03

《Practical Statistics and Experimental Design for Plant and Crop Science》

Book reading

Basic statistical calculations

- Population/sample mean, median, variance.
- Corrected sum of squares(Sxx).
- SD,CV
- Weighted mean, harmonic mean

Type of variables

Continuous, discrete and categorical variables.

Basic summary

- Frequency distributions
- Histograms, boxplot, stem-leaf plot
- Quartiles and ranges

Linear mixed effect models

Sam Rogers

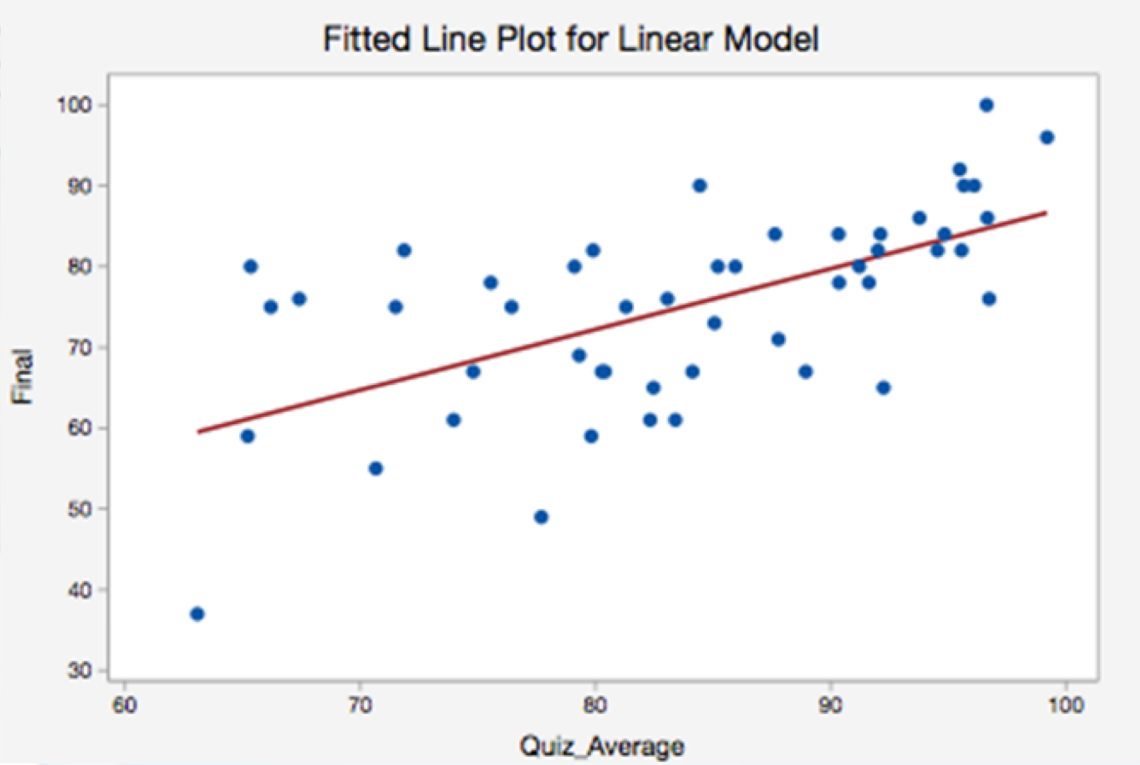
Linear models

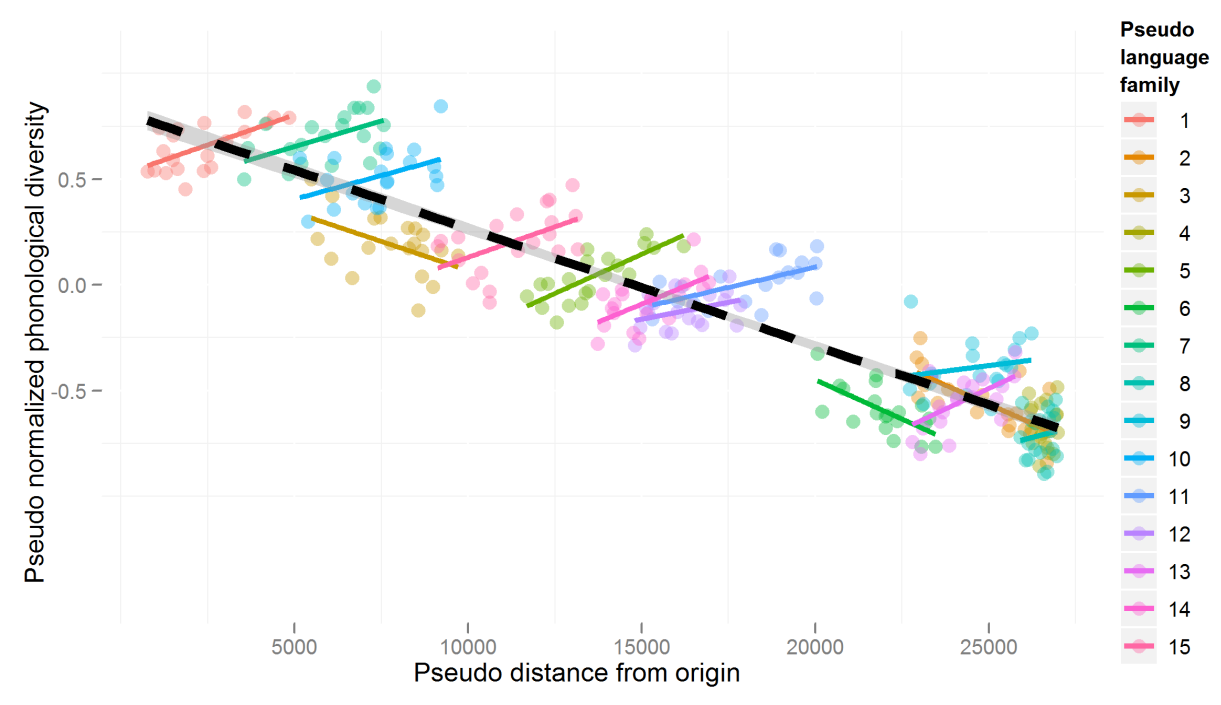
Linear model:
$Y=X\beta+\varepsilon$

Linear mixed model:
$Y=X\beta+Zu+\varepsilon$

Learn more

LME4	ASReml-R
Open source (free to use)	Close source (have to pay for access)
Cannot specify residual correlation structure	Can specify residual correlation structure
Cannot incorporate market based relationship matrices	Can incorporate market based relationship matrices
Limited variance structure available for random effect	Flexible variance structure available for random effect





Linear model

$Y=X\beta+\varepsilon$

Linear mixed model

$Y=X\beta+Z\upsilon+\varepsilon$

Optimal Design

Julian Taylor

Optimal Design

In the design of experiments, optimal designs are a class of experimental designs that are optimal with respect to some statistical criterion.

Package OD

A R package used to design proper experiments to reduce experimental cost

The End

THINKS FOR YOUR LISTING