test

Olamide Adu

## Introduction

This is a short introduction to R, and it promises to be great

# Import Data -------------------------------------------------------------  
  
library(tidyverse)

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.4 ✔ readr 2.1.5  
✔ forcats 1.0.0 ✔ stringr 1.5.1  
✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
✔ purrr 1.0.2   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

read\_csv("data/popdata.csv")

Rows: 189 Columns: 6  
── Column specification ────────────────────────────────────────────────────────  
Delimiter: ","  
chr (1): clone  
dbl (5): block, cutw, height, dia, fert  
  
ℹ Use `spec()` to retrieve the full column specification for this data.  
ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

# A tibble: 189 × 6  
 block cutw height dia clone fert  
 <dbl> <dbl> <dbl> <dbl> <chr> <dbl>  
 1 1 2.4 71 0.6 A 3  
 2 1 0.7 67 1.4 A 3  
 3 1 6.5 211 3.5 A 3  
 4 1 1.1 69 1 A 3  
 5 2 2 116 1.4 A 3  
 6 2 4.9 123 3.2 A 3  
 7 2 0.8 68 2.2 A 3  
 8 2 1.3 79 1.8 A 3  
 9 2 8.8 166 2.4 A 3  
10 2 2 91 2.5 A 3  
# ℹ 179 more rows

read\_csv("data/popdata2.csv", skip = 5)

Rows: 189 Columns: 6  
── Column specification ────────────────────────────────────────────────────────  
Delimiter: ","  
chr (1): clone  
dbl (5): block, cutw, height, dia, fert  
  
ℹ Use `spec()` to retrieve the full column specification for this data.  
ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

# A tibble: 189 × 6  
 block cutw height dia clone fert  
 <dbl> <dbl> <dbl> <dbl> <chr> <dbl>  
 1 1 2.4 71 0.6 A 3  
 2 1 0.7 67 1.4 A 3  
 3 1 6.5 211 3.5 A 3  
 4 1 1.1 69 1 A 3  
 5 2 2 116 1.4 A 3  
 6 2 4.9 123 3.2 A 3  
 7 2 0.8 68 2.2 A 3  
 8 2 1.3 79 1.8 A 3  
 9 2 8.8 166 2.4 A 3  
10 2 2 91 2.5 A 3  
# ℹ 179 more rows

poplar\_data <- read.csv("data/popdata2.csv", skip = 5)  
# Understanding Data

str(poplar\_data)

'data.frame': 189 obs. of 6 variables:  
 $ block : int 1 1 1 1 2 2 2 2 2 2 ...  
 $ cutw : num 2.4 0.7 6.5 1.1 2 4.9 0.8 1.3 8.8 2 ...  
 $ height: int 71 67 211 69 116 123 68 79 166 91 ...  
 $ dia : num 0.6 1.4 3.5 1 1.4 3.2 2.2 1.8 2.4 2.5 ...  
 $ clone : chr "A" "A" "A" "A" ...  
 $ fert : int 3 3 3 3 3 3 3 3 3 3 ...

summary(poplar\_data)

block cutw height dia   
 Min. :1.000 Min. : 0.10 Min. : 22.0 Min. :0.60   
 1st Qu.:2.000 1st Qu.: 1.10 1st Qu.:101.0 1st Qu.:2.00   
 Median :3.000 Median : 2.20 Median :263.0 Median :3.20   
 Mean :3.011 Mean : 3.78 Mean :248.9 Mean :2.86   
 3rd Qu.:4.000 3rd Qu.: 5.80 3rd Qu.:365.0 3rd Qu.:3.60   
 Max. :5.000 Max. :19.70 Max. :506.0 Max. :4.90   
 clone fert   
 Length:189 Min. :1.000   
 Class :character 1st Qu.:1.000   
 Mode :character Median :1.000   
 Mean :1.963   
 3rd Qu.:3.000   
 Max. :3.000

head(poplar\_data)

block cutw height dia clone fert  
1 1 2.4 71 0.6 A 3  
2 1 0.7 67 1.4 A 3  
3 1 6.5 211 3.5 A 3  
4 1 1.1 69 1.0 A 3  
5 2 2.0 116 1.4 A 3  
6 2 4.9 123 3.2 A 3

tail(poplar\_data)

block cutw height dia clone fert  
184 5 1.3 263 3.3 C 1  
185 5 5.1 323 3.8 C 1  
186 5 11.6 394 4.4 C 1  
187 5 6.8 336 4.1 C 1  
188 5 1.3 393 3.1 C 1  
189 5 6.0 339 3.8 C 1

dim(poplar\_data)

[1] 189 6

{## Should we convert

table(poplar\_data$block)

1 2 3 4 5   
37 39 37 37 39

names(poplar\_data)

[1] "block" "cutw" "height" "dia" "clone" "fert"

table(poplar\_data$clone)

A B C   
51 68 70

table(poplar\_data$fert)

1 3   
98 91

table(poplar\_data$cutw)

0.1 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.1 1.2 1.3 1.4 1.5 1.6 1.7   
 1 1 1 6 6 9 9 8 4 7 5 9 4 2 4 2   
 1.8 1.9 2 2.1 2.2 2.3 2.4 2.5 2.6 2.7 2.8 2.9 3.1 3.2 3.5 3.6   
 1 4 5 5 3 1 4 3 2 2 1 1 2 2 1 1   
 3.7 3.8 3.9 4 4.1 4.3 4.5 4.6 4.7 4.8 4.9 5.1 5.2 5.3 5.6 5.7   
 1 1 1 1 3 1 1 3 2 1 1 3 2 1 2 1   
 5.8 6 6.1 6.2 6.3 6.5 6.6 6.8 6.9 7 7.1 7.2 7.3 7.4 7.6 7.7   
 1 2 2 1 3 2 1 4 1 1 1 2 1 1 1 2   
 8.4 8.5 8.6 8.8 9.2 10.1 10.2 10.3 11.3 11.6 11.8 12.1 12.2 12.8 14.2 14.3   
 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1   
14.7 19.7   
 2 1

table(poplar\_data$height)

22 24 26 27 31 32 36 42 43 48 49 51 58 61 62 66 67 68 69 71   
 1 1 1 1 1 2 1 1 1 2 1 2 1 1 1 1 1 2 1 1   
 73 74 76 77 79 81 82 86 89 91 92 93 94 99 101 114 116 117 123 127   
 1 1 1 1 5 3 1 1 1 1 2 3 1 1 1 1 2 1 1 1   
128 144 154 161 163 164 166 173 182 186 188 197 208 211 213 218 221 227 233 234   
 1 2 1 1 1 1 1 1 1 3 1 1 1 1 1 1 2 1 1 1   
236 238 241 242 245 246 247 249 251 254 259 261 262 263 267 274 276 281 282 284   
 1 1 2 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2   
293 294 302 305 306 307 313 316 323 324 328 331 334 336 339 340 341 342 343 346   
 1 1 1 1 2 2 1 2 3 1 1 1 2 2 2 1 1 1 1 1   
347 348 349 353 356 357 361 364 365 366 371 373 377 378 382 386 391 393 394 397   
 1 1 3 2 1 1 1 1 1 3 3 2 1 1 1 1 2 2 3 1   
401 403 404 406 407 408 411 413 414 416 418 421 423 424 431 434 437 441 454 459   
 1 1 1 2 1 1 1 1 3 1 2 1 1 1 2 1 1 1 1 1   
499 506   
 1 1

## Data Types Explanation

poplar\_data$block <- factor(poplar\_data$block)  
poplar\_data$fert <- factor(poplar\_data$fert)  
poplar\_data$clone <- factor(poplar\_data$clone)  
  
summary(poplar\_data)

block cutw height dia clone fert   
 1:37 Min. : 0.10 Min. : 22.0 Min. :0.60 A:51 1:98   
 2:39 1st Qu.: 1.10 1st Qu.:101.0 1st Qu.:2.00 B:68 3:91   
 3:37 Median : 2.20 Median :263.0 Median :3.20 C:70   
 4:37 Mean : 3.78 Mean :248.9 Mean :2.86   
 5:39 3rd Qu.: 5.80 3rd Qu.:365.0 3rd Qu.:3.60   
 Max. :19.70 Max. :506.0 Max. :4.90

str(poplar\_data)

'data.frame': 189 obs. of 6 variables:  
 $ block : Factor w/ 5 levels "1","2","3","4",..: 1 1 1 1 2 2 2 2 2 2 ...  
 $ cutw : num 2.4 0.7 6.5 1.1 2 4.9 0.8 1.3 8.8 2 ...  
 $ height: int 71 67 211 69 116 123 68 79 166 91 ...  
 $ dia : num 0.6 1.4 3.5 1 1.4 3.2 2.2 1.8 2.4 2.5 ...  
 $ clone : Factor w/ 3 levels "A","B","C": 1 1 1 1 1 1 1 1 1 1 ...  
 $ fert : Factor w/ 2 levels "1","3": 2 2 2 2 2 2 2 2 2 2 ...

## Changing Data Types —————————————————-

# Descriptive Statistics ————————————————–

## Measure of Central Tendency ——————————————–

table(poplar\_data$block)

1 2 3 4 5   
37 39 37 37 39

mean(poplar\_data$cutw)

[1] 3.779894

median(poplar\_data$cutw)

[1] 2.2

mean(poplar\_data$height)

[1] 248.9153

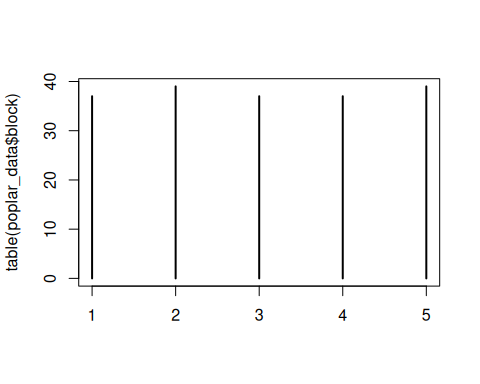
-------------------------------------------------  
var(poplar\_data$cutw)

[1] -13.07991

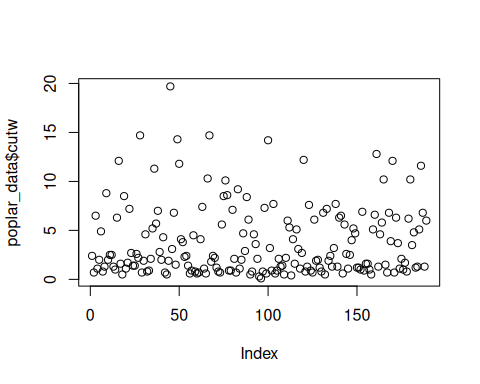
sd(poplar\_data$cutw)

[1] 3.616616

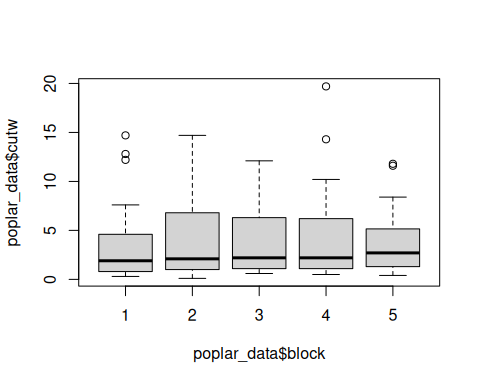
### Simple Visualization --------------------------------------------------  
plot(table(poplar\_data$block))



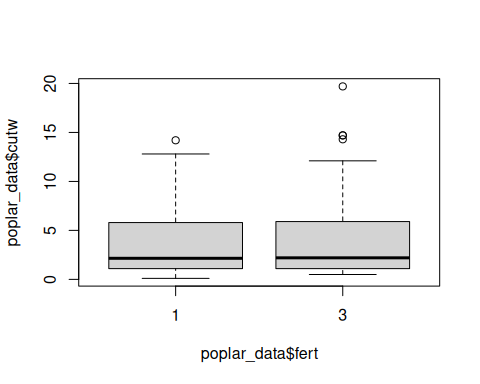
### Single variable -------------------------------------------------------  
plot(poplar\_data$cutw)



#### Cat variable ---------------------------------------------------------  
plot(poplar\_data$cutw ~ poplar\_data$block)



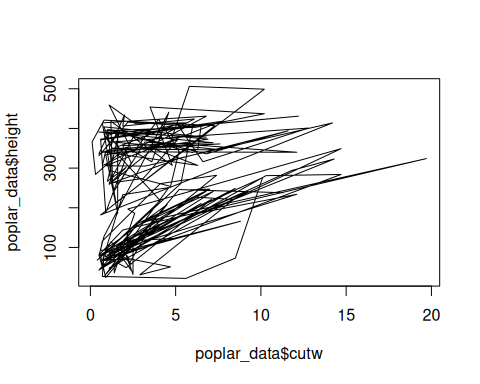
plot(poplar\_data$cutw ~ poplar\_data$fert)



#### Num variable ———————————————————

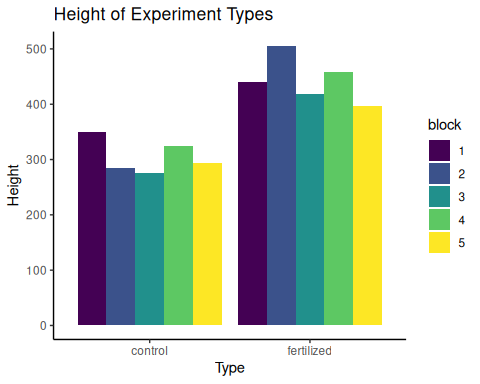
#### Num vs Num ——————————————-

plot(poplar\_data$cutw, poplar\_data$height, type = "l")



# Packages ----------------------------------------------------------------  
## Tidyverse --------------------------------------------------------------  
library(tidyverse)  
  
poplar\_data <- poplar\_data |>   
 mutate(  
 fert = ifelse(fert == 1, "fertilized", "control")  
 )  
## ggplot -----------------------------------------------------------------

poplar\_data |>   
 ggplot(aes(fert, height, fill = block)) +  
 geom\_col(position = "dodge") +  
 scale\_fill\_viridis\_d() +  
 labs(  
 x = "Type",  
 y = "Height",  
 title = "Height of Experiment Types"  
 ) +  
 theme\_classic()



poplar\_data |>   
 ggplot(aes(cutw, height, col = fert)) +  
 geom\_line() +  
 facet\_wrap(~block)



# Inferential Statistics with R -------------------------------------------  
  
## Brief Intro on Hypothesis Testing --------------------------------------  
  
## Calculating new variable -----------------------------------------------

## One Sample T-Test (compare with mean) ———————————-

t.test(poplar\_data$cutw, mu = 4)

One Sample t-test  
  
data: poplar\_data$cutw  
t = -0.83668, df = 188, p-value = 0.4038  
alternative hypothesis: true mean is not equal to 4  
95 percent confidence interval:  
 3.260946 4.298843  
sample estimates:  
mean of x   
 3.779894

t.test(poplar\_data$cutw, poplar\_data$height)

Welch Two Sample t-test  
  
data: poplar\_data$cutw and poplar\_data$height  
t = -25.209, df = 188.28, p-value < 2.2e-16  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 -264.3177 -225.9532  
sample estimates:  
 mean of x mean of y   
 3.779894 248.915344

## Two Sample T-Test ------------------------------------------------------  
  
### Lesser ----------------------------------------------------------------  
### Greater ---------------------------------------------------------------

## Anova ------------------------------------------------------------------  
anova(aov(poplar\_data$height ~ poplar\_data$block))

Analysis of Variance Table  
  
Response: poplar\_data$height  
 Df Sum Sq Mean Sq F value Pr(>F)  
poplar\_data$block 4 104269 26067 1.4744 0.2117  
Residuals 184 3253124 17680

anova(aov(poplar\_data$height ~ poplar\_data$clone))

Analysis of Variance Table  
  
Response: poplar\_data$height  
 Df Sum Sq Mean Sq F value Pr(>F)  
poplar\_data$clone 2 62644 31322 1.7682 0.1735  
Residuals 186 3294749 17714

anova(aov(poplar\_data$height ~ poplar\_data$fert))

Analysis of Variance Table  
  
Response: poplar\_data$height  
 Df Sum Sq Mean Sq F value Pr(>F)   
poplar\_data$fert 1 2191675 2191675 351.58 < 2.2e-16 \*\*\*  
Residuals 187 1165718 6234   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1