

# Lab 3

2025-04-24

## R Markdown

```
library(gridExtra)
pan <- c(12.70, 12.30, 12.40, 12.80, 12.60,
12.50, 12.20, 12.70, 12.40, 12.60)
grilled <- c(11.80, 11.55, 11.75, 11.65, 11.50,
11.75, 11.30, 12.10, 11.50, 11.45)

burgers_df <- data.frame(
  sat_fat = c(pan, grilled),
  method = c(rep("pan", length(pan)),
  rep("grilled", length(grilled)))
)
```

```
sample(burgers_df$sat_fat)
```

```
## [1] 12.30 12.40 12.80 11.50 12.20 11.45 11.75 12.70 12.70 11.65 11.75 11.80
## [13] 11.30 11.55 11.50 12.50 12.10 12.60 12.40 12.60
```

```
burgers_perm <- burgers_df
burgers_perm$sat_fat <- sample(burgers_perm$sat_fat)
```

```
perm_t <- NA

reps <- 10000

for(i in 1:reps){
  burgers_perm <- burgers_df
  burgers_perm$sat_fat <- sample(burgers_perm$sat_fat)
  perm_t[i] <- t.test(sat_fat ~ method,
  data=burgers_perm)$statistic
}
```

```
t.test(sat_fat ~ method, data=burgers_df)$statistic
```

```
##          t
## -9.408921
```

```
t0 <- t.test(sat_fat ~ method, data=burgers_df)$statistic
sum(abs(perm_t) >= abs(t0))/reps
```

```
## [1] 0
```

```
sum(abs(perm_t) >= abs(t0))/reps
```

```
## [1] 0
```

## Part 1: Two sample permutation test

### aPermutation test function

```
perm_test <- function(dataf, r) {  
  colnames(dataf)[1:2] <- c("y", "x")  
  perm_t <- numeric(r) # Storage for permutation t statistics  
  
  for(i in 1:r){  
    perm_t[i] <- t.test(sample(dataf$y) ~ dataf$x)$statistic  
  }  
  # Observed t statistic  
  t0 <- t.test(y ~ x, data=dataf)$statistic  
  # Calculate permutation p-value  
  p_value <- mean(abs(perm_t) >= abs(t0))  
  # Return results  
  return(list(observed_t = t0, permuted_t = perm_t, p_value = p_value))  
}  
  
two_sample <- perm_test(burgers_df, 40)  
print(two_sample)
```

```
## $observed_t  
##          t  
## -9.408921  
##  
## $permuted_t  
## [1] -1.10651494  0.46157872  0.15305651  0.55086038  1.72142364  0.28470187  
## [7] -0.41717121  0.02185129 -0.96351866 -2.38021242  1.01081776  0.10929125  
## [13]  2.58408665  0.06556083  0.73172910  1.20382023  0.46157872 -2.38021242  
## [19] -2.51469265  0.59576542  0.37289909 -0.41717121  2.58408665  0.24074768  
## [25] -0.96351866 -0.59576542 -0.46157872  1.15495548  0.24074768 -0.02185129  
## [31]  0.64086768 -0.59576542 -1.10651494  0.32874752 -0.50613668 -0.64086768  
## [37]  0.68618337  0.86991853  0.02185129  0.37289909  
##  
## $p_value  
## [1] 0
```

## 1b Testing your function

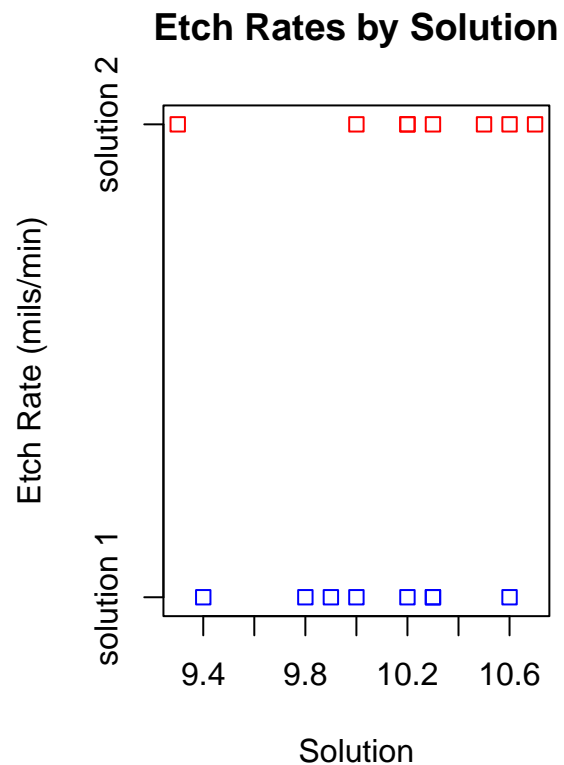
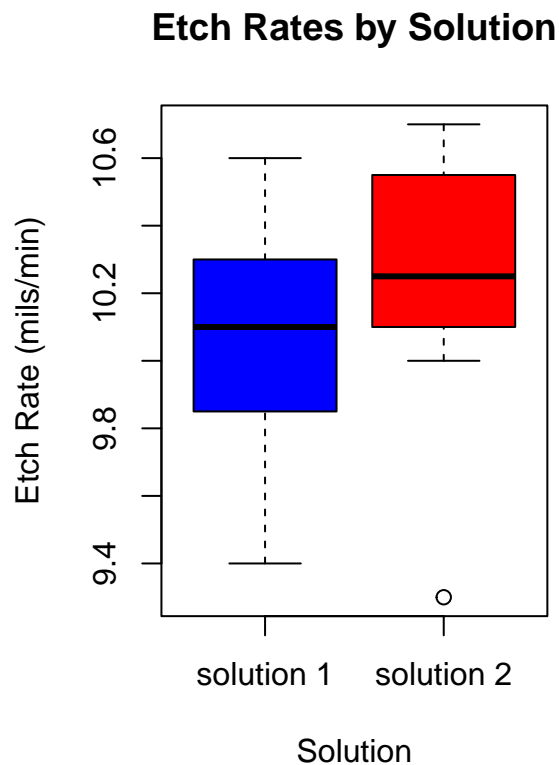
### (i) Wet Chemical Etch Rates

```
# Define the data for each solution group
solution1 <- c(9.9, 9.4, 10.0, 10.3, 10.6, 10.3, 9.8, 10.2)
solution2 <- c(10.6, 10.2, 10.2, 10.0, 9.3, 10.7, 10.5, 10.3)
# Combine the data into a single data frame for easy plotting
e_df <- data.frame(
  e_rate = c(solution1, solution2),
  method2 = c(rep("solution 1", length(solution1)), rep("solution 2", length(solution2)))
)

# For side by side plots 1 row and 2 columns
par(mfrow = c(1, 2))

#Box Plot for the solution
boxplot(e_rate ~ method2, data = e_df,
        main = "Etch Rates by Solution",
        xlab = "Solution",
        ylab = "Etch Rate (mils/min)",
        col = c("blue", "red"))

#Scatter plot (stripchart)
stripchart(e_rate ~ method2, data = e_df,
           main = "Etch Rates by Solution",
           xlab = "Solution",
           ylab = "Etch Rate (mils/min)",
           col = c("blue", "red"))
```



```
# one plot at a time
par(mfrow = c(1, 1))
```

## (ii) Non-normal data with equal means

```
#Load the means_equal.csv file
means_equal_file <- read.csv("means_equal.csv")
head(means_equal_file)
```

```
##           y           x
## 1 3.3083146 group1
## 2 0.9289040 group1
## 3 0.4571145 group1
## 4 0.2738723 group1
## 5 0.2958346 group1
## 6 0.7249473 group1
```

```
#Show columns y (quantitative) and x(group)
```

```
# For side by side plots 1 row and 2 columns
par(mfrow = c(1, 2))
#Box Plot
```

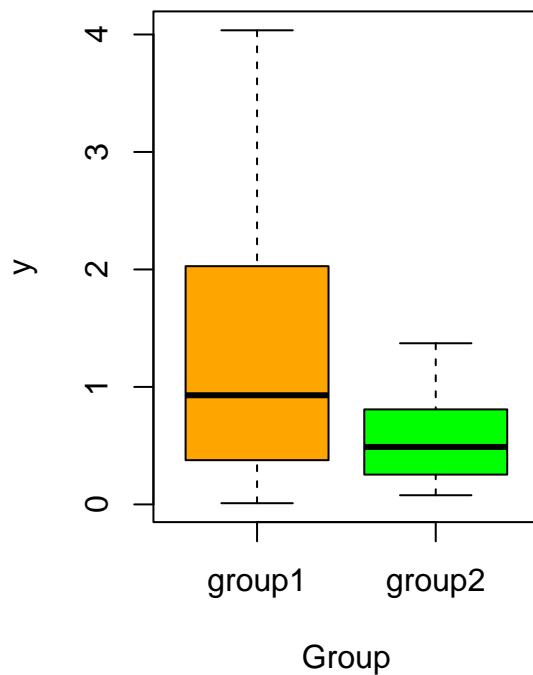
```

boxplot(y ~ x, data = means_equal_file,
        main = "Boxplot: y by group (x)",
        xlab = "Group",
        ylab = "y",
        col = c("orange", "green"))

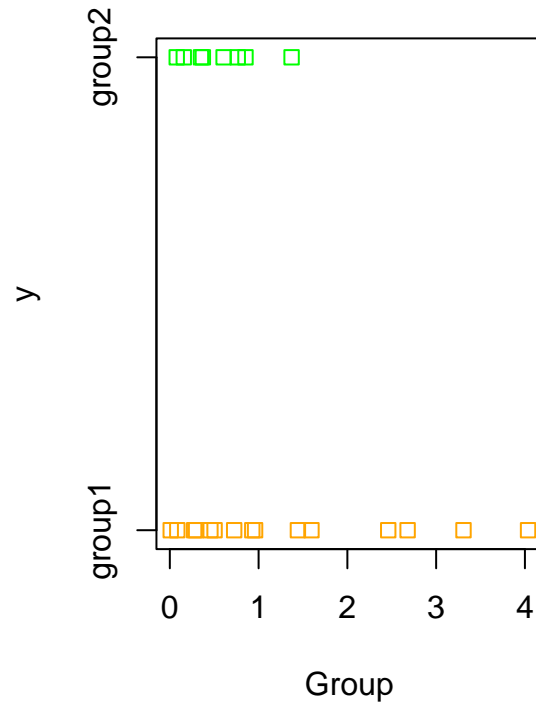
#Scatter plot
stripchart(y ~ x, data = means_equal_file,
           main = "Scatter plot: y by group (x)",
           xlab = "Group",
           ylab = "y",
           col = c("orange", "green"))

```

**Boxplot: y by group (x)**



**Scatter plot: y by group (x)**



```

# one plot at a time
par(mfrow = c(1, 1))

```

```

Non_normal_perm_equal_mean_result <- perm_test(means_equal_file, 40)
Non_normal_perm_equal_mean_result$p_value

```

```
## [1] 0.05
```

```
t.test(y ~ x, data = means_equal_file)$statistic
```

```
##          t
## 2.097349
```

```
t.test(y ~ x, data = means_equal_file)$p.value
```

```
## [1] 0.04963312
```

The Permutation test p-value: 0.075 t-test statistic: 2.097 t-test p-value: 0.0496

The permutation test for non-normal data with equal means have a p-value of 0.075, which is not statistically significant at the 0.05 level.

The t-test is 0.0496, which is statistically significant at the 0.05 level.

The t-test relies on assumptions of normality and equal variance. The permutation test does not rely on distributional assumptions and is more robust. This difference highlights why the permutation test is preferred in situations with non-normal data.

t-test ( $p = 0.0496$ ): does reject  $H_0$  so this is a Type I Error (false positive). Permutation test ( $p=0.075$ ): does not reject  $H_0$  so this is correct decision (no error).

This t-test is a Type I Error rate

### (iii) Non-normal data with different means

```
#Load the means_diff.csv file
means_diff_file <- read.csv("means_diff.csv")
head(means_diff_file)
```

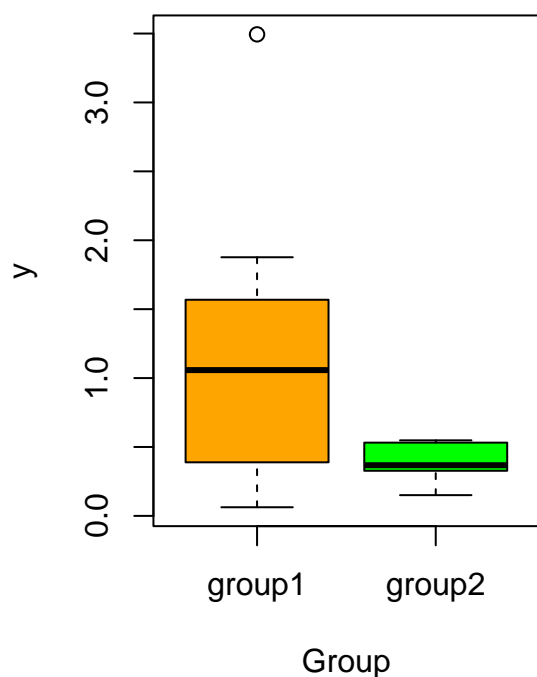
```
##           y           x
## 1 0.3877827 group1
## 2 1.8768018 group1
## 3 0.3897914 group1
## 4 1.2599098 group1
## 5 3.4941736 group1
## 6 0.9861160 group1
```

```
#Show columns y (quantitative) and x(group)
```

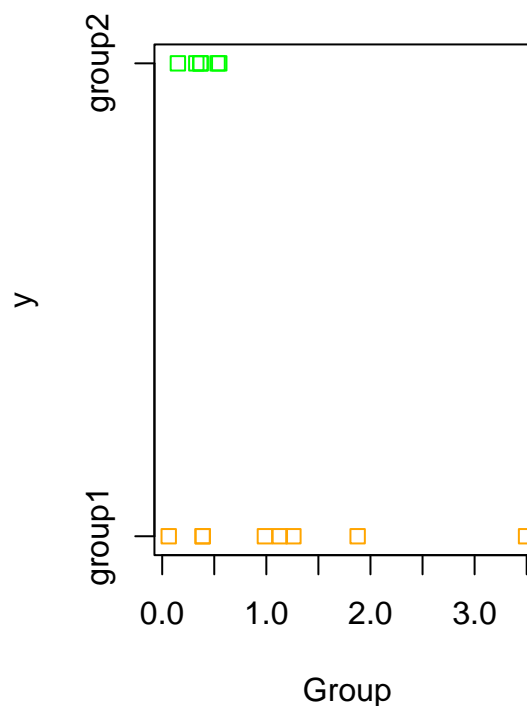
```
par(mfrow = c(1, 2))
#Box Plot
boxplot(y ~ x, data = means_diff_file,
        main = "Boxplot: y by group (x)",
        xlab = "Group",
        ylab = "y",
        col = c("orange", "green"))

#Scatter plot
stripchart(y ~ x, data = means_diff_file,
           main = "Scatter plot: y by group (x)",
           xlab = "Group",
           ylab = "y",
           col = c("orange", "green"))
```

**Boxplot: y by group (x)**



**Scatter plot: y by group (x)**



```
par(mfrow = c(1, 1))
```

```
Non_normal_perm_equal_diff_result <- perm_test(means_diff_file, 40)  
Non_normal_perm_equal_diff_result$p_value
```

```
## [1] 0.025
```

```
t.test(y ~ x, data = means_diff_file)$statistic
```

```
##          t  
## 2.084587
```

```
t.test(y ~ x, data = means_diff_file)$p.value
```

```
## [1] 0.07377473
```

The Permutation test p-value: 0.05 t-test statistic: 2.08 t-test p-value: 0.074

The permutation test for non-normal data with a true difference in means has a p-value of 0.05, which is right on the edge of statistical significance at the 0.05 level.

The t-test p-value is 0.074, which is not statistically significant at the 0.05 level.

The t-test relies on assumptions of normality and equal variance. The permutation test does not rely on distributional assumptions and is more robust.

This difference highlights why the permutation test is preferred in situations with non-normal data.

t-test ( $p = 0.074$ ): does not reject  $H_0$ , so this is a Type II Error (false negative). Permutation test ( $p=0.05$ ): does reject  $H_0$ , so this is the correct decision (no error).

This shows how assumption violations can lead to Type II Errors in the t-test, and why the permutation test is more reliable here.

## **Part 2: ANOVA on Therapy Dogs**

### **1) In your own words, describe the overall aim of the study. Specifically, in**

the abstract, it states that “The aim of this study was to assess the impact of client-canine contact on wellbeing outcomes.” Explain what they mean by this in more detail.

The goal of this study was to figure out if we touch a therapy dog during a canine-assisted intervention (CAI) actually leads to better mental health and wellbeing outcomes for students. When the authors say they were assessing the “impact of client-canine contact on wellbeing outcomes,” they mean they wanted to test how different levels of interaction — touching the dog, being near the dog without touching, or just talking to the handler (no dog) — affect things like stress, happiness, loneliness, and feeling connected to others.

They compared three different types of interaction: one where students could pet and physically engage with the dog, one where students were near the dog but couldn’t touch it, and one where there was no dog at all just the handler. The researchers wanted to see how these different experiences affected things like stress, happiness, loneliness, and how connected students felt to others.

As a result, they were trying to go beyond the usual question of “Do therapy dogs help people feel better?” and figure out why they help and if physical touch plays a key role in making those benefits happen.

### **2) How were subjects enrolled into this study? Describe any concerns with**

regard to generalizability to a broader population that are present based on how their study was conducted.

The subjects were undergraduate psychology students from a mid-sized Canadian university. They were in a online system and got course credit for participating. Students were randomly assigned to one of three groups but they self-selected to be part of the study. The concerns it might have was mostly female, young, and enrolled in psychology courses. Participation was conducted at only one university, which may not represent the general student population or broader public.

### **3) Regarding the dog therapy treatment of interest:**

- a) What are the three treatment groups that participants were assigned to, and exactly how were participants allocated to their treatment group?

The three groups are Direct Contact - students got to hang out with a therapy dog and were encouraged to actually pet and cuddle the dog, Indirect Contact students were in the same room as the dog and could see it up close, but they weren’t allowed to touch it and Handler-only students just talked with the dog’s handler, but there was no dog around at all.

b) If statistically significant differences are found between the treatment groups, would it be appropriate here to conclude that these differences are caused by the treatment itself? Why or why not? Yes, if statistically significant differences are found between the groups, we can say the treatment caused them. That’s because this study was a randomized controlled trial, which is basically the most reliable way to test whether something actually works. Since participants were randomly placed into groups, it helps make sure the groups were similar to start with. There are differences in outcomes like improvements in stress and happiness, it’s



probably because of the treatment, not because one group just happened to have more stressed-out students than another.

Also, since the only major difference between the groups was how they interacted with the dog, we can feel pretty confident that any changes were actually caused by the type of dog interaction they had.

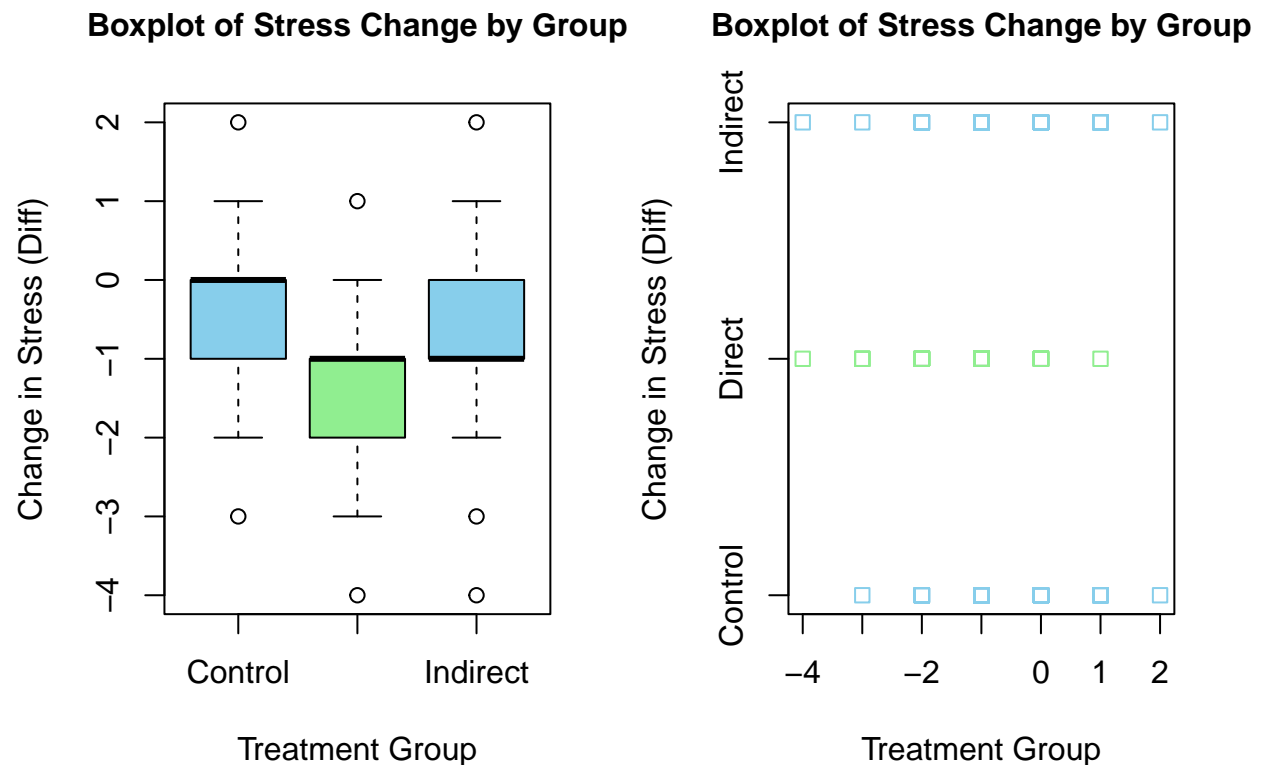
#### 4) Load the `dog_data_post.csv` data from Canvas in your R Markdown file.

```
# Load the data
dog_data <- read.csv("dog_data_post.csv")
head(means_diff_file)

##           y           x
## 1 0.3877827 group1
## 2 1.8768018 group1
## 3 0.3897914 group1
## 4 1.2599098 group1
## 5 3.4941736 group1
## 6 0.9861160 group1

par(mfrow = c(1, 2))
#side-byside boxPlot
boxplot(Diff ~ GroupAssignment, data = dog_data,
        main = "Boxplot of Stress Change by Group",
        xlab = "Treatment Group",
        ylab = "Change in Stress (Diff)",
        col = c("skyblue", "lightgreen"), cex.main = 1)

#Scatter plot
stripchart(Diff ~ GroupAssignment, data = dog_data,
           main = "Boxplot of Stress Change by Group",
           xlab = "Treatment Group",
           ylab = "Change in Stress (Diff)",
           col = c("skyblue", "lightgreen"), cex.main = 1)
```



```
par(mfrow = c(1, 1))
```

### State the null and alternative hypotheses

$\mu_1$  = mean Diff (stress change) for treatment group 1  $\mu_2$  = mean Diff (stress change) for treatment group 2

Null hypothesis ( $H_0$ ):  $\mu_1 = \mu_2$

Alternative hypothesis ( $H_A$ ):  $\mu_1 \neq \mu_2$

```
anova_result <- aov(Diff ~ GroupAssignment, data = dog_data)
summary(anova_result)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## GroupAssignment  2  33.06  16.529    20.72 4.04e-09 ***
## Residuals      281 224.16   0.798
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### Part 3: ANOVA Under the Hood

#### 1) Therapy Dogs again

```

library(readr)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:gridExtra':
##
##      combine

## The following objects are masked from 'package:stats':
##
##      filter, lag

## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union

summary <- dog_data %>% group_by(GroupAssignment) %>%
  summarize(y_bar_i = mean(Diff), S2_i = var(Diff), n_i = n())

a <- dim(summary)[1]
N <- dim(dog_data)[1]

SSE <- sum((summary$n_i - 1) * summary$S2_i)
MSE <- SSE / (N - a)

y_ddot <- mean(dog_data$Diff)
SST <- sum(summary$n_i * (summary$y_bar_i - y_ddot) ^ 2)
MST <- SST / (a - 1)

F0 <- MST/MSE
pval <- pf(F0, df1 = (a-1), df2 = (N-a), lower.tail = FALSE)

output <- matrix(NA, nrow = 2, ncol = 5)

output[1,1] <- (a-1)
output[1,2] <- SST
output[1,3] <- MST
output[1,4] <- F0
output[1,5] <- pval
output[2,1] <- (N-a)
output[2,2] <- SSE
output[2,3] <- MSE

output <- matrix(as.character(signif(output, 4)), nrow = 2)

colnames(output) <- c("Df", "Sum Sq", "Mean Sq", "F0", "p-value")
rownames(output) <- c("method", "Residuals")

knitr::kable(output)

```

	Df	Sum Sq	Mean Sq	F0	p-value
method	2	33.06	16.53	20.72	4.036e-09
Residuals	281	224.2	0.7977	NA	NA

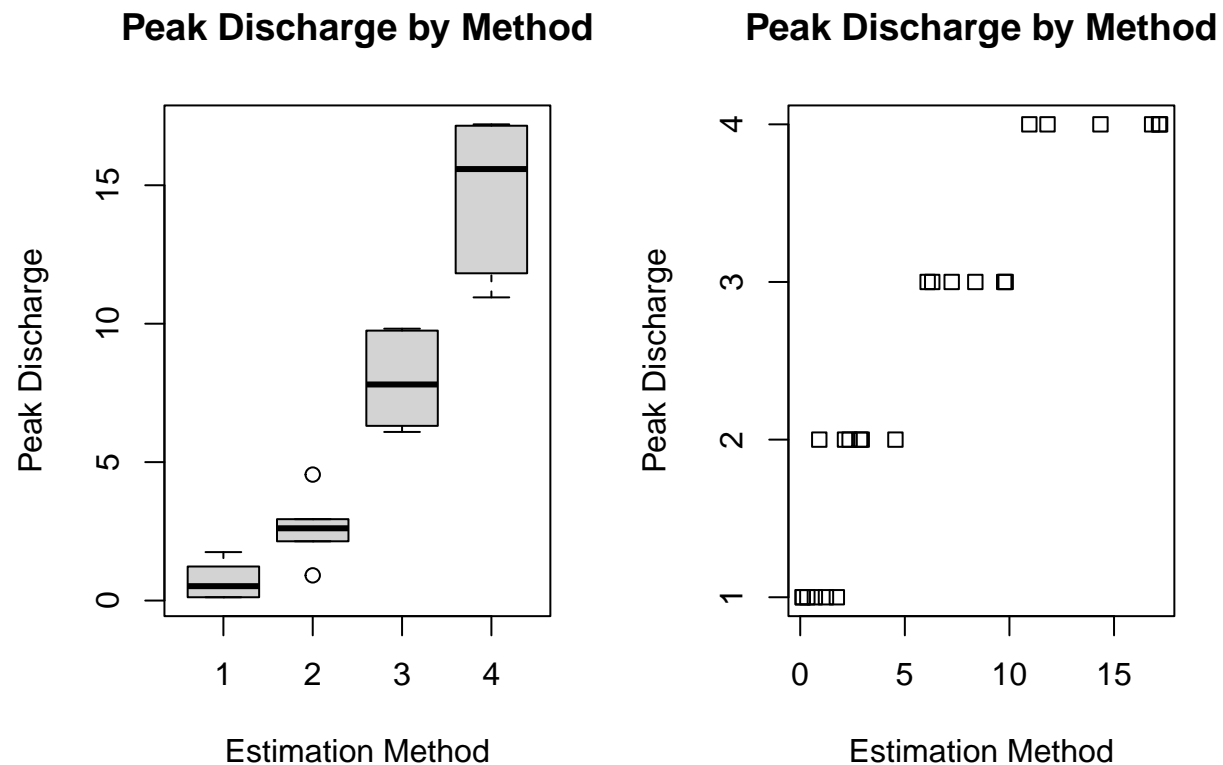
2)

a: Input the data, create dataframe, make side-by-side boxplots and scatter diagram

```
method <- rep(1:4, each = 6)
obs <- c(
  0.34, 0.12, 1.23, 0.70, 1.75, 0.12,
  0.91, 2.94, 2.14, 2.36, 2.86, 4.55,
  6.31, 8.37, 9.75, 6.09, 9.82, 7.24,
  17.15, 11.82, 10.95, 17.20, 14.35, 16.82
)
peak_df <- data.frame(Method = factor(method), Discharge = obs)

# Side-by-side boxplot and scatter plot
par(mfrow = c(1, 2))
boxplot(Discharge ~ Method, data = peak_df,
  main = "Peak Discharge by Method",
  xlab = "Estimation Method", ylab = "Peak Discharge")

stripchart(Discharge ~ Method, data = peak_df,
  main = "Peak Discharge by Method",
  xlab = "Estimation Method", ylab = "Peak Discharge")
```



```
par(mfrow = c(1, 1))
```

b): Run ANOVA using aov and show full output

```
anova_peak <- aov(Discharge ~ Method, data = peak_df)
summary(anova_peak)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Method      3  708.3   236.1    76.07 4.11e-11 ***
## Residuals   20   62.1     3.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c):

```
library(dplyr)
summary <- peak_df %>% group_by(Method) %>%
  summarize(y_bar_i = mean(Discharge), S2_i = var(Discharge), n_i = n())

a <- dim(summary)[1]
N <- dim(peak_df)[1]
```

```

SSE <- sum((summary$n_i - 1) * summary$S2_i)
MSE <- SSE / (N - a)

y_ddot <- mean(dog_data$Diff)
SST <- sum(summary$n_i * (summary$y_bar_i - y_ddot) ^ 2)
MST <- SST / (a - 1)

F0 <- MST/MSE
pval <- pf(F0, df1 = (a-1), df2 = (N-a), lower.tail = FALSE)

output <- matrix(NA, nrow = 2, ncol = 5)

output[1,1] <- (a-1)
output[1,2] <- SST
output[1,3] <- MST
output[1,4] <- F0
output[1,5] <- pval
output[2,1] <- (N-a)
output[2,2] <- SSE
output[2,3] <- MSE

output <- matrix(as.character(signif(output, 4)), nrow = 2)

colnames(output) <- c("Df", "Sum Sq", "Mean Sq", "F0", "p-value")
rownames(output) <- c("Group", "Residuals")

knitr::kable(output)

```

	Df	Sum Sq	Mean Sq	F0	p-value
Group	3	1998	665.9	214.5	2.256e-15
Residuals	20	62.08	3.104	NA	NA

## Including Plots

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.