BIOL 425 Final Report: My Own 150-character-or-so Long Informative Take-Home-Message Title (Do not use the original article title; a summary of your own analysis)

Your Name

5/1/2023

Table of Contents

# 1 Introduction

## 1.1 Background

Describe the research field & questions

## 1.2 Biological hypothesis

Specific to your data analysis (not the general hypothesis of the paper)

## 1.3 Significance

Why your question is important?

# 2 Materials and Methods

## 2.1 Samples

Specify the nature, source of the biological samples, sample size, positive and negative controls (if any)

## 2.2 Experimental procedure

Describe how the data were collected, e.g., the sequencing technology. Include computational protocols, e.g., software tools, R packages used to visualize and produce the data. Again, specific for your data set. Do NOT copy & paste from the paper

## 2.3 Statistical methods

Describe the statistical hypothesis, the kind of statistical tests (e.g., t-test, ANOVA, regression, PCA), the specific R packages used

# 3 Exploratory data analysis

Embed your R codes and show your work (code, plots, & comments)

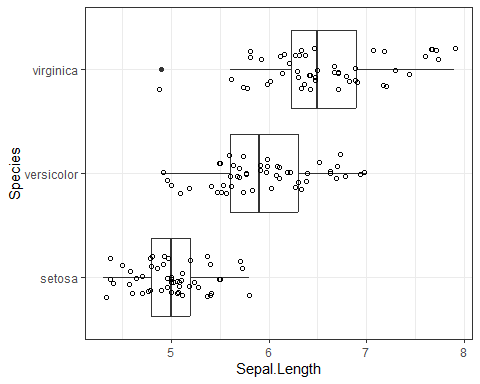
knitr::opts\_chunk$set(echo = TRUE)  
  
# load libraries  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.1 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── 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

# load data  
data("iris")  
  
# peak data  
glimpse(iris)

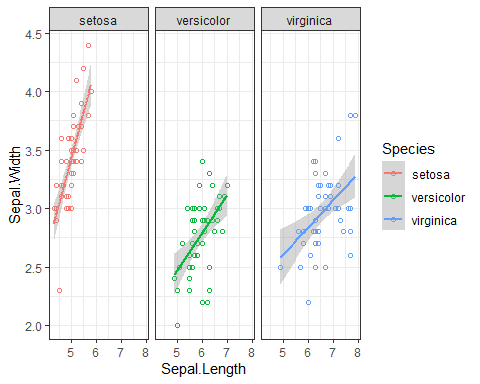
## Rows: 150  
## Columns: 5  
## $ Sepal.Length <dbl> 5.1, 4.9, 4.7, 4.6, 5.0, 5.4, 4.6, 5.0, 4.4, 4.9, 5.4, 4.…  
## $ Sepal.Width <dbl> 3.5, 3.0, 3.2, 3.1, 3.6, 3.9, 3.4, 3.4, 2.9, 3.1, 3.7, 3.…  
## $ Petal.Length <dbl> 1.4, 1.4, 1.3, 1.5, 1.4, 1.7, 1.4, 1.5, 1.4, 1.5, 1.5, 1.…  
## $ Petal.Width <dbl> 0.2, 0.2, 0.2, 0.2, 0.2, 0.4, 0.3, 0.2, 0.2, 0.1, 0.2, 0.…  
## $ Species <fct> setosa, setosa, setosa, setosa, setosa, setosa, setosa, s…

# boxplot to compare sepal length among species, which suggest significant differences that could be tested with an ANOVA  
iris %>% ggplot(aes(x = Species, y = Sepal.Length )) +  
 geom\_boxplot() +  
 geom\_jitter(shape = 1, width = 0.2) +  
 theme\_bw() +  
 coord\_flip()



# scatter plot to compare sepal length versus sepal width, which suggest significant correlation that could be tested with linear regression  
iris %>% ggplot(aes(x = Sepal.Length, y = Sepal.Width, color = Species)) +  
 geom\_point(shape = 1) +  
 geom\_smooth(method = "lm") +  
 facet\_wrap(~Species) +   
 theme\_bw()

## `geom\_smooth()` using formula = 'y ~ x'



# 4 Results

Show plots, each one with a legend/caption, explaining x- and y-axis, colors, points, sample sizes, and what’s the story you want to tell a reader

Make statistical conclusions

# 5 Conclusions

## 5.1 Biological conclusions

Does the study answer the question raised in the beginning? State the NOVEL findings.

## 5.2 Future work

What are the remaining questions to be answered? Suggest future work and directions.

# 6 References

## 6.1 Cite paper

List the full citation of the original paper, including DOI, e.g.,

Ronald L. Wasserstein & Nicole A. Lazar (2016) The ASA Statement on p-Values: Context, Process, and Purpose, The American Statistician, 70:2, 129-133, DOI: 10.1080/00031305.2016.1154108

## 6.2 Cite code repository (if available)

List the computational resource , e.g., the github link

Computer codes & data: <https://github.com/SchlossLab/Armour_Resolution_mBio_2021>

## 6.3 Cite data file

List the data source, e.g., the URL of the Excel sheet:

Data set: <https://genome.cshlp.org/content/suppl/2022/01/18/gr.275901.121.DC1/Supplemental_Table_S3.xls>