

# Homework 2

Insert your name here

## Table of contents

Question 1	2
Question 2	4
Question 3	4

<b>Appendix</b>	<b>5</b>
-----------------	----------

[Link to the Github repository](#)

---

**!** Due: Tue, Feb 14, 2023 @ 11:59pm

Please read the instructions carefully before submitting your assignment.

1. This assignment requires you to only upload a PDF file on Canvas
2. Don't collapse any code cells before submitting.
3. Remember to make sure all your code output is rendered properly before uploading your submission.

Please add your name to the author information in the frontmatter before submitting your assignment

For this assignment, we will be using the [Abalone dataset](#) from the UCI Machine Learning Repository. The dataset consists of physical measurements of abalone (a type of marine snail) and includes information on the age, sex, and size of the abalone.

## Question 1

💡 20 points

EDA using `readr`, `tidyr` and `ggplot2`

### 1.1 (5 points)

Load the “Abalone” dataset as a tibble called `abalone` using the URL provided below. The `abalone_col_names` variable contains a vector of the column names for this dataset (to be consistent with the R naming pattern). Make sure you read the dataset with the provided column names.

```
library(readr)
url <- "http://archive.ics.uci.edu/ml/machine-learning-databases/abalone/abalone.data"

abalone_col_names <- c(
  "sex",
  "length",
  "diameter",
  "height",
  "whole_weight",
  "shucked_weight",
  "viscera_weight",
  "shell_weight",
  "rings"
)

abalone <- ... # Insert your code here
```

---

### 1.2 (5 points)

Remove missing values and NAs from the dataset and store the cleaned data in a tibble called `df`. How many rows were dropped?

```
df <- ... # Insert your code here
```

---

### 1.3 (5 points)

Plot histograms of all the quantitative variables in a **single plot** <sup>1</sup>

```
... # Insert your code here
```

---

### 1.4 (5 points)

Create a boxplot of **length** for each **sex** and create a violin-plot of **diameter** for each **sex**. Are there any notable differences in the physical appearances of abalones based on your analysis here?

```
... # Insert your code for boxplot here
```

```
... # Insert your code for violinplot here
```

---

### 1.5

Create a scatter plot of **length** and **diameter**, and modify the shape and color of the points based on the **sex** variable. Change the size of each point based on the **shell\_wight** value for each observation. Are there any notable anomalies in the dataset?

```
... # Insert your code here
```

---

### 1.6

For each **sex**, create separate scatter plots of **length** and **diameter**. For each plot, also add a **linear** trendline to illustrate the relationship between the variables. Use the **facet\_wrap()** function in R for this, and ensure that the plots are vertically stacked **not** horizontally. You should end up with a plot that looks like this: <sup>2</sup>

---

<sup>1</sup>You can use the **facet\_wrap()** function for this. Have a look at its documentation using the help console in R


<sup>2</sup>Plot example for 1.6

## Question 2

 30 points

---

## Question 3

 50 points

---

## Appendix

### Session Information

Print your R session information using the following command

```
sessionInfo()
```

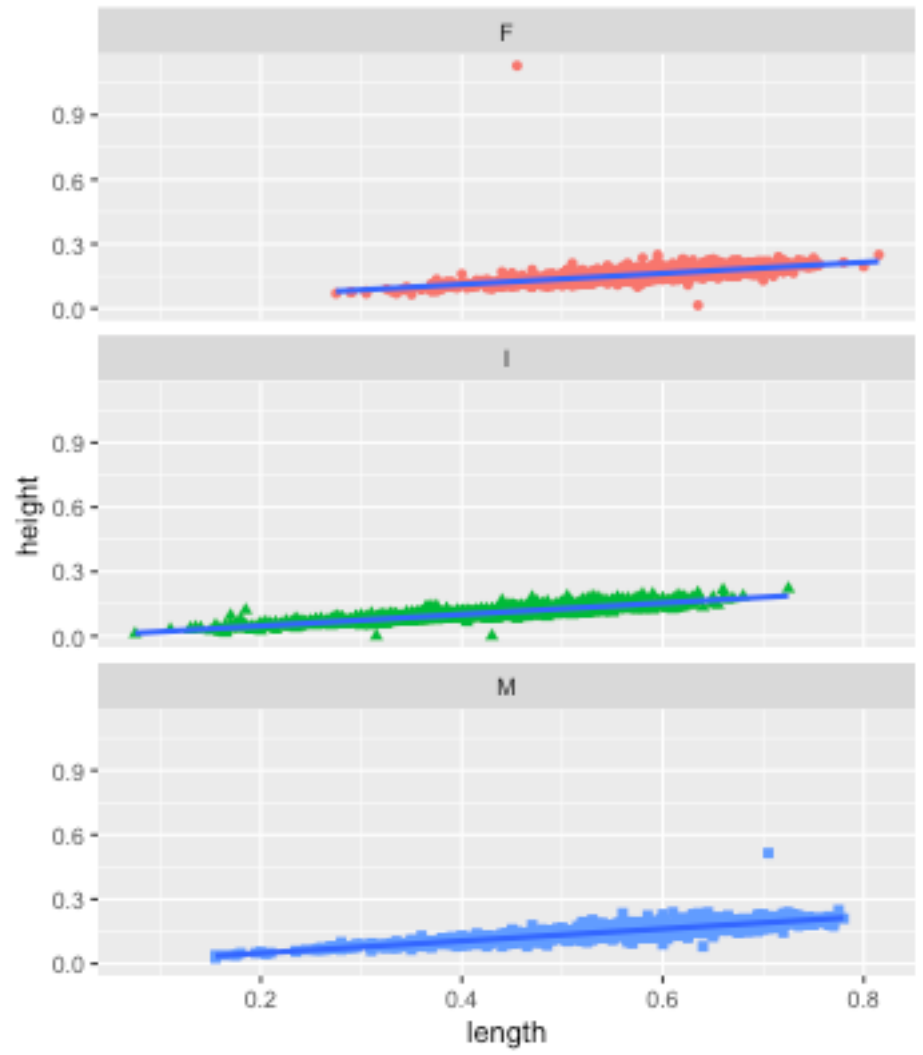
```
R version 4.2.2 (2022-10-31)
Platform: x86_64-apple-darwin22.1.0 (64-bit)
Running under: macOS Ventura 13.2

Matrix products: default
BLAS:   /usr/local/Cellar/openblas/0.3.21/lib/libopenblas-r0.3.21.dylib
LAPACK: /usr/local/Cellar/r/4.2.2_1/lib/R/lib/libRlapack.dylib

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] stats      graphics  grDevices datasets  utils      methods    base

loaded via a namespace (and not attached):
 [1] digest_0.6.31    lifecycle_1.0.3  jsonlite_1.8.4   magrittr_2.0.3
 [5] evaluate_0.20    rlang_1.0.6      stringi_1.7.12   cli_3.6.0
 [9] renv_0.16.0-53   vctrs_0.5.1      rmarkdown_2.20   tools_4.2.2
[13] stringr_1.5.0    glue_1.6.2       xfun_0.36        yaml_2.3.6
[17] fastmap_1.1.0    compiler_4.2.2   htmltools_0.5.4  knitr_1.41
```



:::{.hidden unless-format="html"}  
:::