HiPy: Introduction to R and ggplot2

This document will provide you with a series of challenges designed to introduce you to R and ggplot2. Each challenge is structured to encourage active learning, experimentation, and the use of ChatGPT as a learning tool. Hints, tips, example code snippets, and suggested ChatGPT prompts are included to guide the learning process.

Challenge 0: Install R and RStudio

If you haven’t got R and RStudio installed, then go and ask ChatGPT how to do this. It will provide you with exactly the information you need.

If you’re too impatient to install today, then try <https://posit.cloud> – an online version of RStudio that is pretty much identical to the desktop version.

## Challenge 1: Exploring the Iris Dataset

### Task:

Load the iris dataset and display the first 6 rows. Then, calculate and display summary statistics for each feature.

### Hints & Tips:

* Use the `head()` function to display the first few rows of a dataset.
* The `summary()` function provides summary statistics for each feature in the dataset.

### Example Code:

data(iris)  
head(iris)  
summary(iris)

### Suggested ChatGPT Prompts:

* How do I load built-in datasets in R?
* What is the iris dataset in R and what does it contain?
* Show me how to calculate summary statistics in R.

## Challenge 2: Creating a Boxplot

### Task:

Create a boxplot showing the distribution of sepal length for each iris species.

### Hints & Tips:

* Use `ggplot(data = iris, aes(x = Species, y = Sepal.Length)) + geom\_boxplot()` to create your boxplot.

### Example Code:

library(ggplot2)  
ggplot(data = iris, aes(x = Species, y = Sepal.Length)) + geom\_boxplot()

### Suggested ChatGPT Prompts:

* How do I use ggplot2 to make a boxplot in R?
* Explain the syntax of ggplot2 functions in R.

## Challenge 3: Generating a Scatterplot

### Task:

Create a scatterplot of sepal length vs. sepal width, color-coded by species.

### Hints & Tips:

* Utilize `aes(color = Species)` within your `ggplot()` function to color-code points.

### Example Code:

ggplot(data = iris, aes(x = Sepal.Length, y = Sepal.Width, color = Species)) + geom\_point()

### Suggested ChatGPT Prompts:

* How can I color code points in a ggplot scatterplot in R?
* Tips for enhancing scatterplots in ggplot2.

## Challenge 4: K-Means Clustering

### Task:

Perform k-means clustering on the iris dataset (using Sepal.Length and Sepal.Width) and visualize the clusters.

### Hints & Tips:

* Use the `kmeans()` function, specifying the number of centers (e.g., `kmeans(iris[,1:2], centers = 3)`).
* Plot the clusters by adding `geom\_point(aes(color = as.factor(clustering$cluster)))` to your scatterplot.

### Example Code:

set.seed(123) # for reproducibility  
clustering <- kmeans(iris[,1:2], centers = 3)  
ggplot(iris, aes(Sepal.Length, Sepal.Width)) + geom\_point(aes(color = as.factor(clustering$cluster)))

### Suggested ChatGPT Prompts:

* WTF is k-means clustering.
* How to perform k-means clustering in R?
* Visualizing k-means clustering results using ggplot2.