# Practical 1

March 27, 2025

## 1 Hands-On Exercises: Data Visualization with ggplot2

---

### 1.1 Exercise 1: Installing and Loading ggplot2

#### 1.1.1 Task:

1. Install `ggplot2` if it is not already installed.

2. Load the `iris` dataset and display the first few rows.

#### 1.1.2 Questions:

3. What information does `head(iris)` provide?

---

### 1.2 Exercise 2: Understanding the Grammar of Graphics

#### 1.2.1 Task:

1. Use `ggplot2` to create a basic scatter plot of **`Sepal.Length`** vs **`Sepal.Width`** from the `iris` dataset.

#### 1.2.2 Questions:

1. What are the five key components of the **Grammar of Graphics**?

2. What does the `aes()` function do in `ggplot2`?

3. What happens if you remove `aes()` from the plot?

---

### 1.3 Exercise 3: Creating Basic Plots

#### 1.3.1 Task:

1. **Scatter plot:** Create a scatter plot showing the relationship between **`Petal.Length`** and **`Petal.Width`**.

2. **Line plot:** Create a line plot showing **Sepal.Length** across observations.

3. **Bar plot:** Create a bar plot showing the count of each **Species** in the `iris` dataset.

#### 1.3.2 Questions:

1. What is the difference between `geom_point()`, `geom_line()`, and `geom_bar()`?

2. Why is `Species` used as a **categorical** variable in the bar plot?

3. What happens if you try to plot `geom_line()` without an x-axis variable?

---

### 1.4 Exercise 4: Faceting for Multi-panel Plots

#### 1.4.1 Task:

1. Create a faceted scatter plot of **Sepal.Length** vs **Sepal.Width**, with facets based on **Species**.

#### 1.4.2 Questions:

1. What does `facet_wrap(~ Species)` do in the plot?

2. What is the difference between `facet_wrap()` and `facet_grid()`?

3. How would you modify the facet layout to show **two columns** instead of automatic placement?

---

### 1.5 Exercise 5: Customizing ggplot2 Visualizations

#### 1.5.1 Task:

1. Modify your scatter plot by:
   - Changing the theme to **`theme_minimal()`**.
   - Adding a **title, axis labels, and a caption**.
   - Changing the **point color to blue**.

#### 1.5.2 Questions:

1. What are some common `ggplot2` themes?

2. How does `labs()` improve visualization clarity?

3. How can you remove **gridlines** from a plot?

---

## 1.6 Exercise 6: Working with Multiple Geoms

### 1.6.1 Task:

1. Modify the scatter plot by adding:
   - A **smooth trend line** using `geom_smooth(method = "lm")`.
   - Coloring the points based on `Species`.

### 1.6.2 Questions:

1. What does `geom_smooth()` do in a plot?
2. What happens if you remove `method = "lm"`?
3. How can you disable the confidence interval in `geom_smooth()`?

---

## 1.7 Exercise 7: Using Statistical Transformations

### 1.7.1 Task:

1. Create a bar plot that shows the **mean Sepal.Length** per **Species** using `stat_summary()`.
2. Overlay a **density curve** on a histogram of `Sepal.Length`.

### 1.7.2 Questions:

1. What does `stat_summary()` do in `ggplot2`?
2. How does `geom_density()` differ from `geom_histogram()`?
3. What happens when you change the `binwidth` in a histogram?

---

## 1.8 Final Challenge: Create Your Own Data Visualization

### 1.8.1 Task:

- Choose a dataset (e.g., `iris`, `diamonds`, or another dataset).
- Create a **multi-layered** plot with:
  - **Custom themes, colors, labels, and statistical layers**.
  - **Faceting or interactivity** (optional).

### 1.8.2 Questions:

1. What insights does your plot reveal?

2. How did you customize the plot?

3. What challenges did you face while creating your visualization?

[ ]: