Worksheet

Rohan Alexander

15 September, 2025

1. (The idea for this exercise is from Liza Bolton.) The following code makes a nice graph. Modify it to create the ugliest graph that you can.

```
from sklearn.datasets import load_iris
import matplotlib.pyplot as plt
import pandas as pd

# Load built-in Iris dataset
iris = load_iris(as_frame=True)
df = iris.frame
df["species"] = df["target"].map(dict(zip(range(3), iris.target_names)))

# Aesthetic setup
plt.style.use("seaborn-v0_8-whitegrid")
fig, ax = plt.subplots(figsize=(8, 6))

# Scatter plot by species
colors = {"setosa": "#1f77b4", "versicolor": "#ff7f0e", "virginica": "#2ca02c"}
for species, group in df.groupby("species"):
    ax.scatter(
        group["sepal length (cm)"],
        group["sepal width (cm)"],
        label=species,
        s=60,
        alpha=0.8,
        edgecolor="white",
        linewidth=0.8,
        color=colors[species],
    )

# Titles and labels
```

```
ax.set_title("Iris Sepal Dimensions", fontsize=16, weight="bold")
ax.set_xlabel("Sepal Length (cm)", fontsize=12)
ax.set_ylabel("Sepal Width (cm)", fontsize=12)

# Legend and layout
ax.legend(title="Species", frameon=True)
plt.tight_layout()
plt.show()
```

2. In small groups of two, please read the relevant chapter, and create a two-to-five slide summary of the most important points, using Quarto. Make a PR that will add your content to the class GitHub repo.

- Tufte, *The Visual Display of Quantitative Information*, Chapter 1.
- Healy, *Data Visualization*, Chapter 1.3-1.5.
- Cleveland, *The Elements of Graphing Data*, Chapter 2.2 and 2.3.
- Tukey, *EDA*, Chapter 5.

Homework:

- Make three graphs from three different Open Data Toronto datasets and email them to me.