DS 5110: Introduction to Data Management and Processing (Fall 2024) Roi Yehoshua
Student name: (Due) September 26, 2024

PS2: Matplotlib

1. Moore's Law is the observation that the number of transistors on CPUs approximately doubles every two years. Write a program that illustrates Moore's Law by comparing the actual number of transistors on CPUs between the years 1972 and 2012 (shown in Table 1), and that predicted by Moore's Law which may be stated mathematically as:

$$n_i = n_0 2^{(y_i - y_0)/2}$$

where n_0 is the number of transistors in the first year y_0, and n_i is the number of transistors in the ith year y_i.

Because the data cover 40 years, the values of n_i span many orders of magnitude, thus it is convenient to apply Moore's Law to the logarithm of n_i, which shows a linear dependence on y_i:

$$\log_{10} n_i = \log_{10} n_0 + \frac{y_i - y_0}{2} \cdot \log_{10} 2$$

Plot a graph that shows the observed vs. predicted number of CPU transistors for each year in logarithmic scale (see example in Figure 1).

Year	1972	1974	1978	1982	1985	1989	1993	1997
Transistors (in millions)	0.0025	0.005	0.029	0.12	0.275	1.18	3.1	7.5

Year	1999	2000	2003	2004	2007	2008	2012
Transistors (in millions)	24.0	42.0	220.0	592.0	1720.0	2046.0	3100.0

Table 1: Number of CPU transistors over the years

2. The Iris data set consists of 50 samples from each of three species of the Iris flower (Setosa, Versicolour, and Virginica), stored in a 150×4 NumPy array. Each sample consists of 4 features: sepal length, sepal width, petal length and petal width (all in centimeters). This data set can be downloaded using the Scikit-Learn library as follows:

```
from sklearn.datasets import load_iris

X, y = load_iris(return_X_y=True)
```

Here, `X` is a 150×4 matrix containing the features of each flower, and `y` contains the type of each flower (0, 1, or 2).

Create a scatter plot in which the x coordinate of each point corresponds to the sepal length, the y coordinate correspond to the sepal width, the size of the point represents

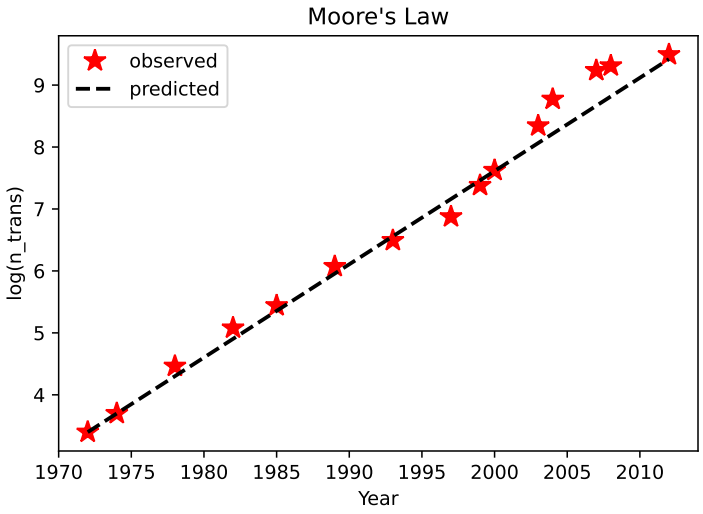

Figure 1: The predictions of Moore's law

the petal width, and the color represents the type of the flower (see example in Figure 2).

3. Write a program that displays a bar chart of letter frequencies in the English language, estimated by analyzing the text of Moby-Dick (`http://www.gutenberg.org/ebooks/2701`). Label the vertical bars by the letters (see example in Figure 3).

4. The market share of different web browsers has changed over the years. Suppose we have the following data for browser market shares in 2024:

Browser	**Market Share (%)**
Chrome	65.18
Safari	18.55
Edge	5.26
Firefox	2.74
Samsung Internet	2.56
Opera	2.15

Table 2: Browser market shares worldwide in 2024

Write a program that displays the market share of each browser as a pie chart. Make sure to:

- Include percentages on each slice of the pie chart.
- Separate the "Chrome" slice slightly from the rest of the pie (using the `explode` parameter) to highlight its dominance.
- Use different colors for each slice.
- Add a title to the chart.

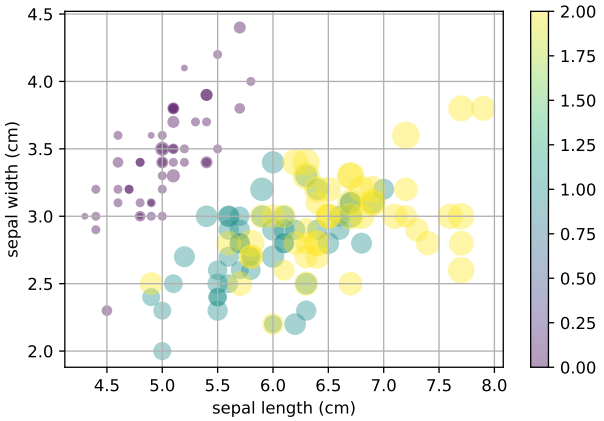

Figure 2: The Iris flower data set

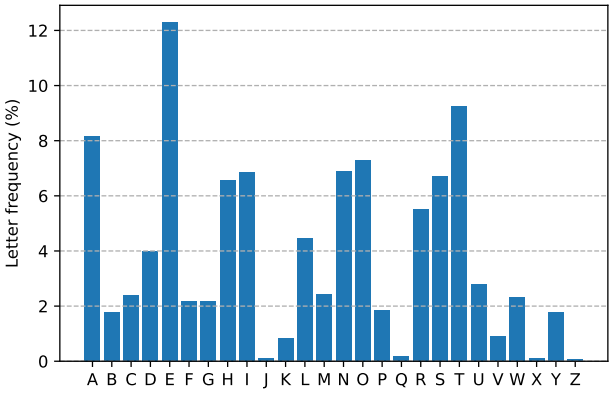

Figure 3: The frequencies of the English letters in a sample text file

5. The MNIST dataset consists of 70,000 handwritten digit images, each of size 28×28 pixels, with labels from 0 to 9. You can load this dataset using Scikit-Learn as follows:

```
from sklearn.datasets import fetch_openml

X, y = fetch_openml('mnist_784', return_X_y=True, as_frame=False)
```

where `X` contains the image data in a flattened form, and `y` contains the corresponding digit labels.

Write a program that plots the first 50 images in this dataset alongside their labels. Display the images in a grid with 5 rows and 10 columns, and add the corresponding digit label below each image.

6. Consider the mathematical function:

$$z = \sin(x^2 + y^2)$$

Write a program that creates a 3D surface plot of this function for x and y ranging from -5 to 5. Include a color bar that represents the value of z at each point, and use a colormap such as `coolwarm` for the plot.