

Assigned: 1-23-2023

Due Date: **2-1-2023 by Noon**

CS 6635/5635 Spring Semester 2023

Assignment 1 (Plotting and Graphing using Python and/or Matlab)

Use either Matplotlib in Python¹ or Matlab² to perform the following plotting:

Part 1: Generate your own data and visualize it [30 pts]

Create an array of 100 random numbers uniformly distributed between 0 and 1. Create a second array of 200 random numbers with Gaussian distribution between 1 and 100.

Note: You will need to find an appropriate mean and standard deviation for the Gaussian. It is okay if just a few of the numbers generated are outside the [1,100] range.

- 1 [4 pts] Create a **box plot** for visualization of both arrays.
- 2 [4 pts] Partition the data into 20 bins and create a **histogram** of both arrays using the 20 bins with a bar chart. You may not use a histogram function from a plotting library, however, you may use a bar chart function.
- 3 [6 pts] **Write** the arrays into a binary file. **Read** it back into an array. Visualize the arrays that were read in by plotting the **cumulative distribution function** as a **line graph**.
- 4 [4 pts] Create 2D arrays using **uniform random sampling** and **gaussian random sampling** with 5,000 points on [0,1] x [0,1]. Plot the arrays with a **scatter plot** and compare.

Note: You can use python's built in random.sample package.

4b) [8 pts] For both sampling arrays, generate 100 bins along both dimensions(think of counting the number of points in each grid cell). The output will be a 2D array of size 100x100. Show these arrays as images.

4c) [4 pts] Now plot both uniform and gaussian sampled arrays as **contour plots** with 10 levels.

Hint: Use *tricontourf* in matplotlib as it is unstructured data.

Part 2: Interesting data sets for visualization [20 pts]

- 1 [5 pts] Download the NOAA Land Ocean Temperature Anomalies Data Set: <https://my.eng.utah.edu/~cs6635/NOAA-Temperatures.csv>. Create a bar plot of the data. Include a label called "Year" along the x-axis and a label called Degrees F +/-

¹ Other Python visualization libraries, such as Seaborn, Pandas, and Plotly are also acceptable, but note that, for the purposes of plotting in Python, the TAs are most familiar with Matplotlib.

² Matlab is freely available for all University of Utah students: <https://www.mathworks.com/academia/tah-portal/university-of-utah-31156832.html>

From Average along the y-axis. Color each bar with either red/blue based on whether there is a positive/negative change in temperature. Describe trends in the data.

- 2 [5 pts] Download the dataset <https://my.eng.utah.edu/~cs6635/Breakfast-Cereals.xls> and generate a **radar chart** with 8 nutritional statistics for 3 cereals.
- 3 [10 pts] Five Thirty Eight maintains a server with many interesting datasets: <https://github.com/fivethirtyeight/data>. Choose two different data sets to visualize. Visualize each data set using what you think is the most appropriate plot for the given data. Choose between **Parallel Coordinates** and **Scatter Plot** for each data set and use each plotting type only once. It is also helpful to color lines or points according to certain features in the data set to get more insight. Describe the trends you can find in the data by experimenting with these plots.

Part 3: Questions on The Value of Visualization Paper [20 pts]

(<https://my.eng.utah.edu/~cs6635/Value-of-Visualization.pdf>)

- 1 [5 pts] Why is assessing value of visualizations important? What are the two measures for deciding the value of visualizations?
- 2 [5 pts] Briefly describe a mathematical model for the visualization block shown in Fig. 1.
- 3 [5 pts] State four parameters that describe the costs associated with any visualization technique.
- 4 [5 pts] What are the pros and cons of interactivity of visualizations?

Part 4: 3D scalar volume data sets (Only for CS 6635) [20 pts]

MATLAB/Python also can be used for analysis and visualization of 3D volume data sets, such as brain MRI images. Download the brain MRI data set from <https://my.eng.utah.edu/~cs6635/T2.nii.gz>. The data format is .nii with 320 x 320 x 256 dimensions. Load data in MATLAB/Python. Extract one slice for each axis (three slices total) from the volume and save them as images. Use at least two colormaps to show the three image set and describe the difference this choice makes.

- One library that can be used for this in Python is NiBabel: <https://nipy.org/nibabel/>

Extra Credit: Reimplement **Part 1 [5 pts]** and **Part 2 [5 pts]** in **BOTH** Python and Matlab. Did you find Python or Matlab easier to use? Why? Please include the replicated plots, it is not necessary to repeat the answers to any questions unless they have changed.

What to turn in:

Write a **report documenting your results**, including any necessary plots/figures, and answering any questions asked above. Be sure to explain any figures you submit and to write a **conclusion** at the end of your report. Your homework is primarily graded upon your report. Please submit your report on Canvas in PDF format.

- Your report should be in PDF format and should stand on its own
- It should describe the methods used.

- It should explain your results and contain figures.
- It should also answer any questions asked above.
- It should cite any sources used for information, including source code.

Note: Any figures/plots in the report should be captioned appropriately. Also be sure to include axis labels in all plots.

This homework assignment is due on **February 1, 2023 by 11:59 am**. If you don't understand these directions or have questions, please send questions to teach-cs6635@sci.utah.edu or come see one of the TAs or the instructor during office hours **well in advance of the due date**.

*If you wish, you may create a Jupyter Notebook³ with prose documenting/explaining your results, export that to PDF, and turn the PDF in as your report. This is an option for those who like using Jupyter, but it is not required.

3 See <https://www.datacamp.com/community/tutorials/tutorial-jupyter-notebook> for a tutorial