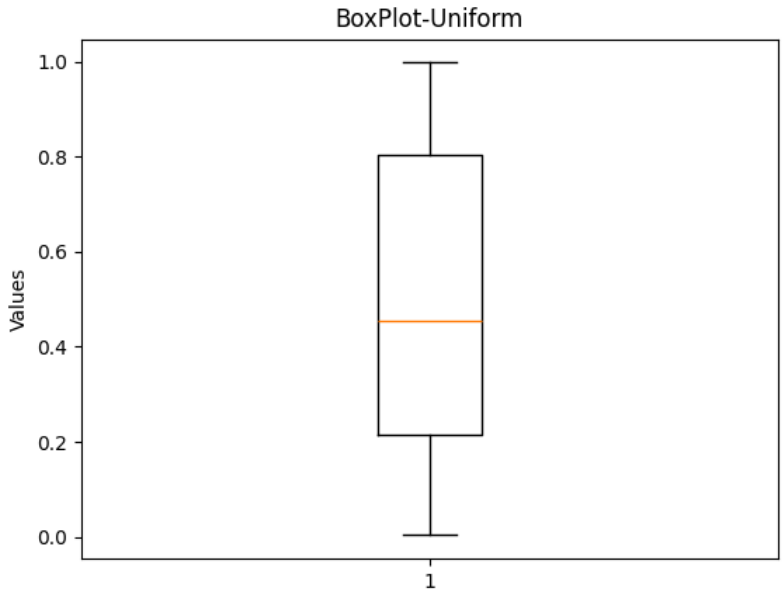
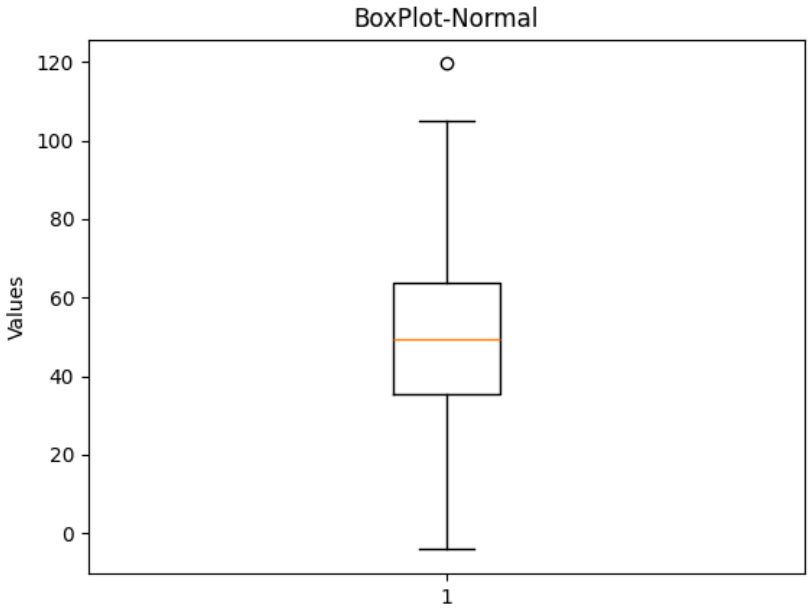
CS 6635 Assignment 1

Part 1:

1)

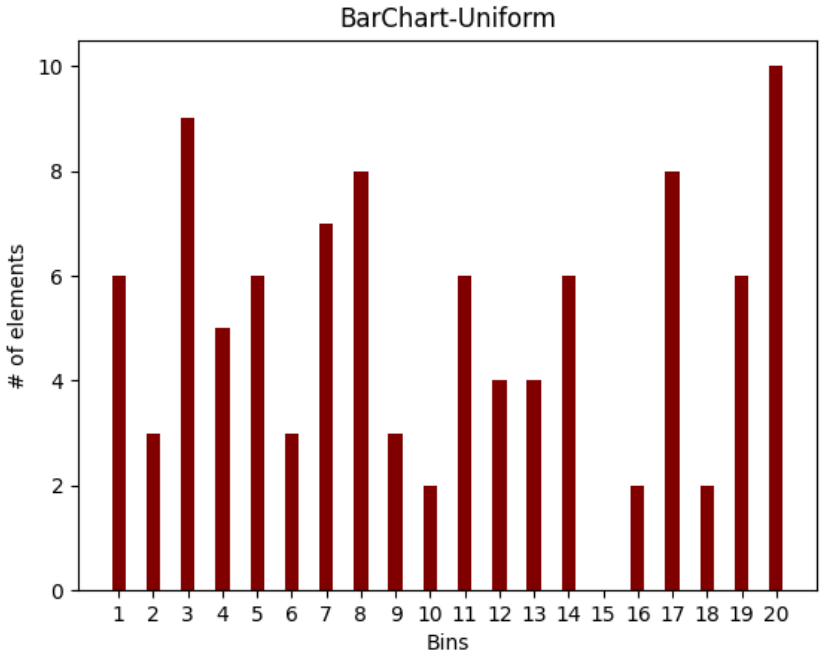


The above is the box plot for the array which has been sampled from the uniform distribution.

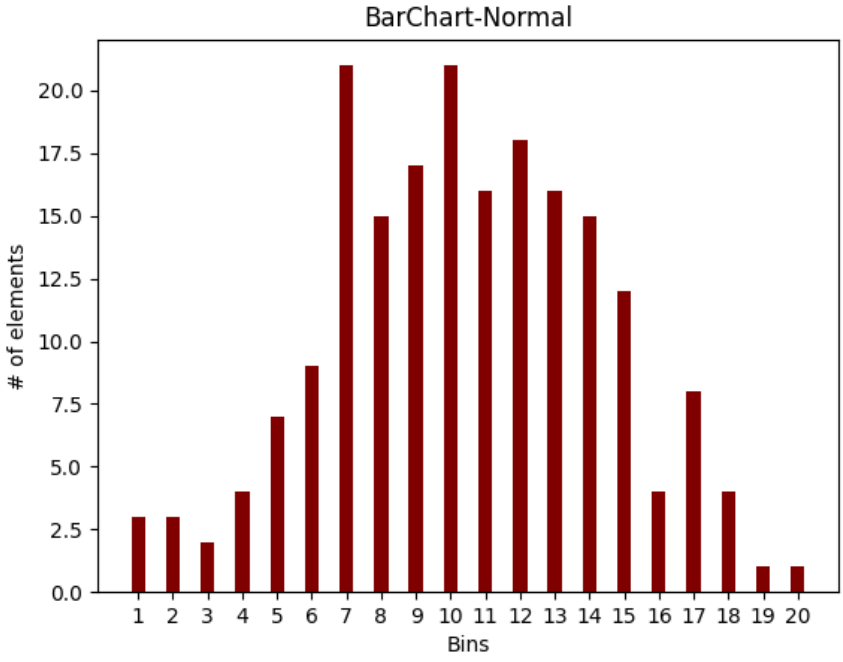


The above is the box plot for the array which has been sampled from the normal distribution, in particular each element of the array follows N(50,20).

2)



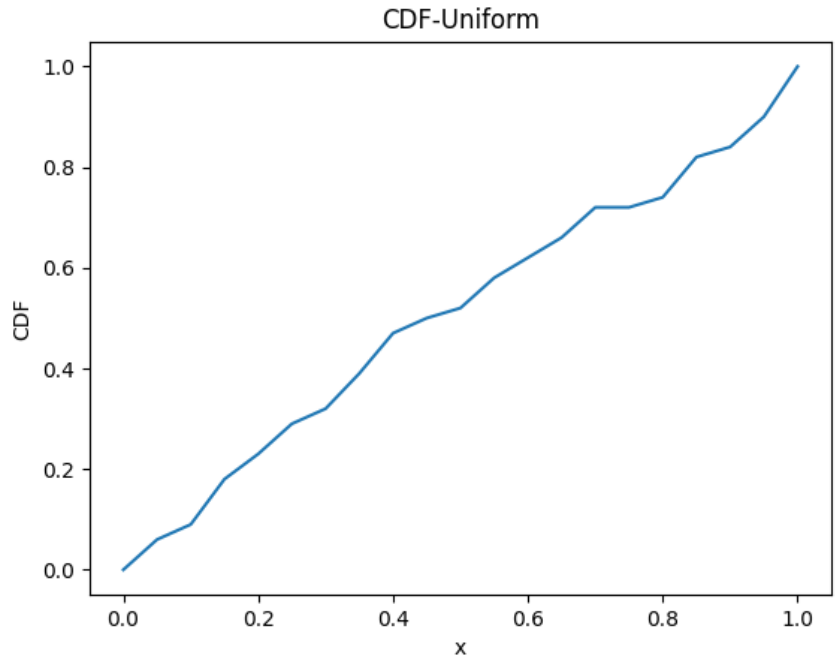
This is the Bar chart obtained by binning the previously obtained uniformly distributed array. It has been split into 20 bins by dividing the [0,1] interval into 20 equal parts.



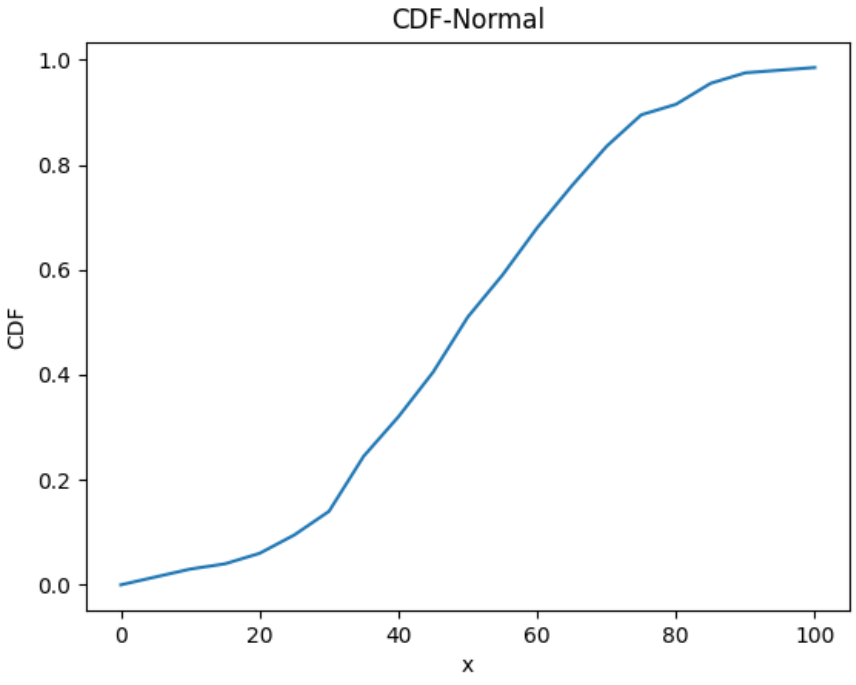
This is the Bar chart obtained by binning the previously obtained normally distributed array. It has been split into 20 bins by dividing the [1,100] interval into 20 equal parts.

3)

As the question mentions, the cdf graphs have been plotted by writing the arrays into a binary file and then reading them back.



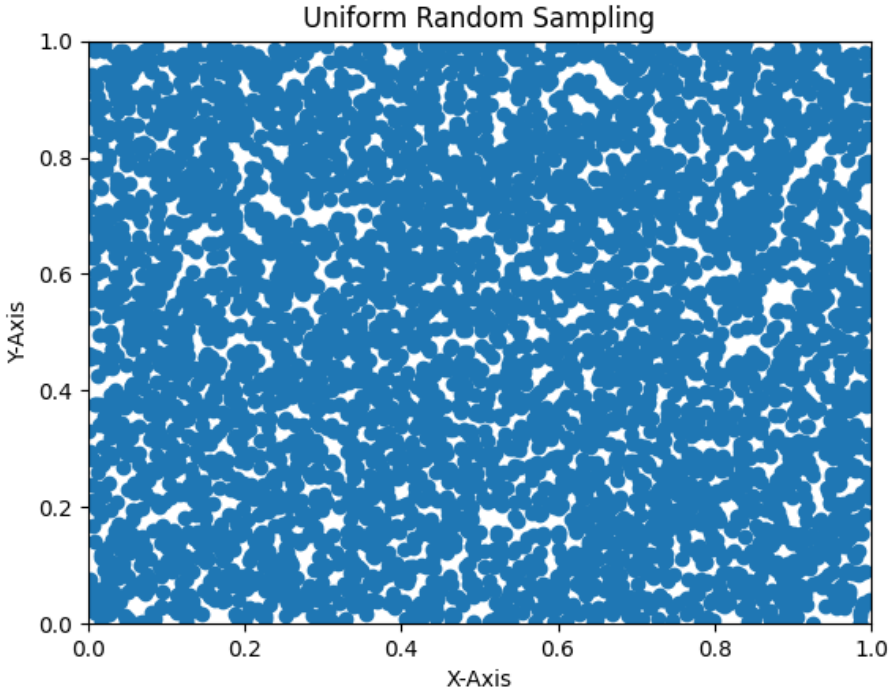
The above is the line graph of the cdf of the above mentioned uniformly sampled array.



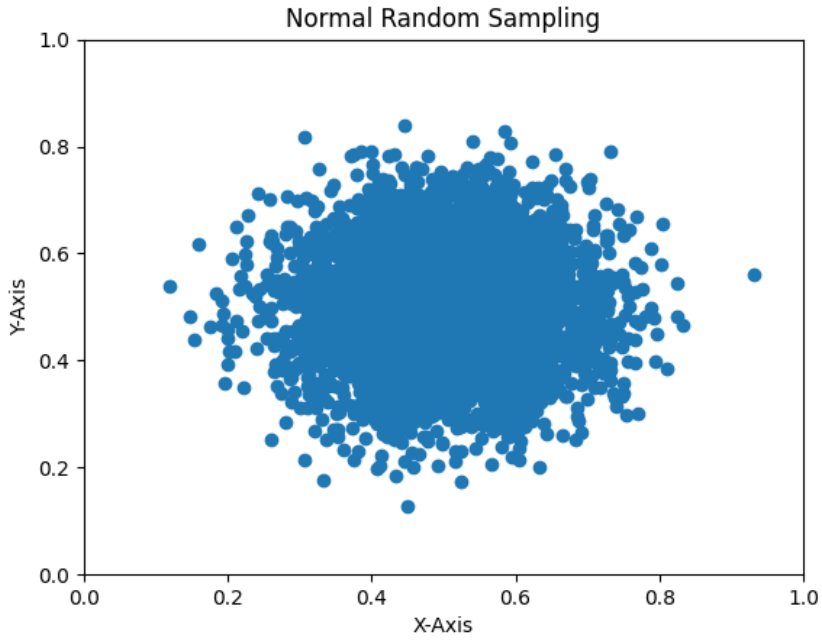
The above is the line graph of the cdf of the above mentioned normally sampled array.

4)

A)



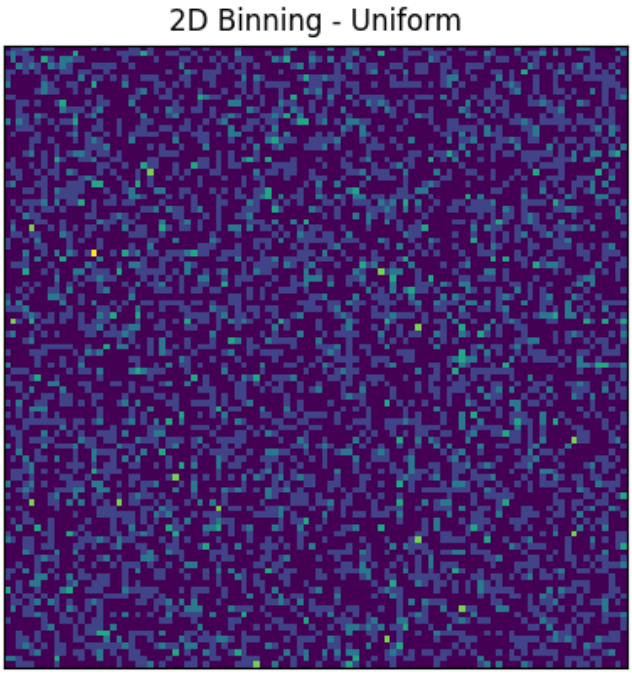
The above is the scatter plot obtained by uniform sampling in [0,1]x[0,1].



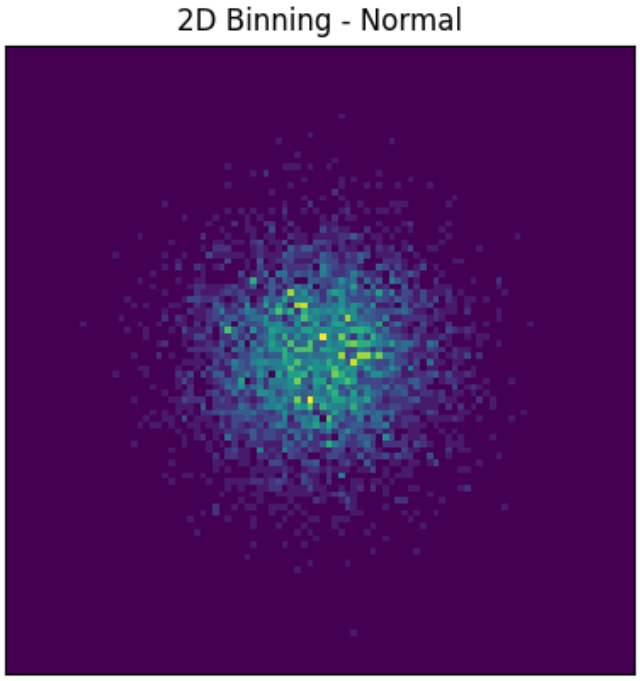
The above is the scatter plot obtained by normal sampling in [0,1]x[0,1], in particular the distribution followed is N(0.5,0.1).

It is to be noted that in the case of the uniform sampling the points are homogeneously scattered in the entire domain. However, in the case of normal sampling the points are concentrated at (0.5,0.5).

B)



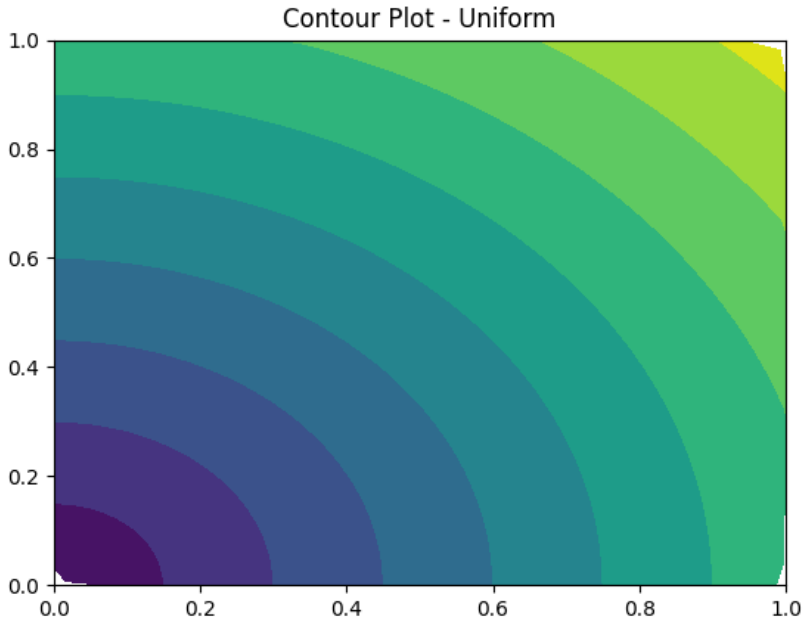
The above is the 100x100 image obtained by binning the uniformly sampled 2D array.



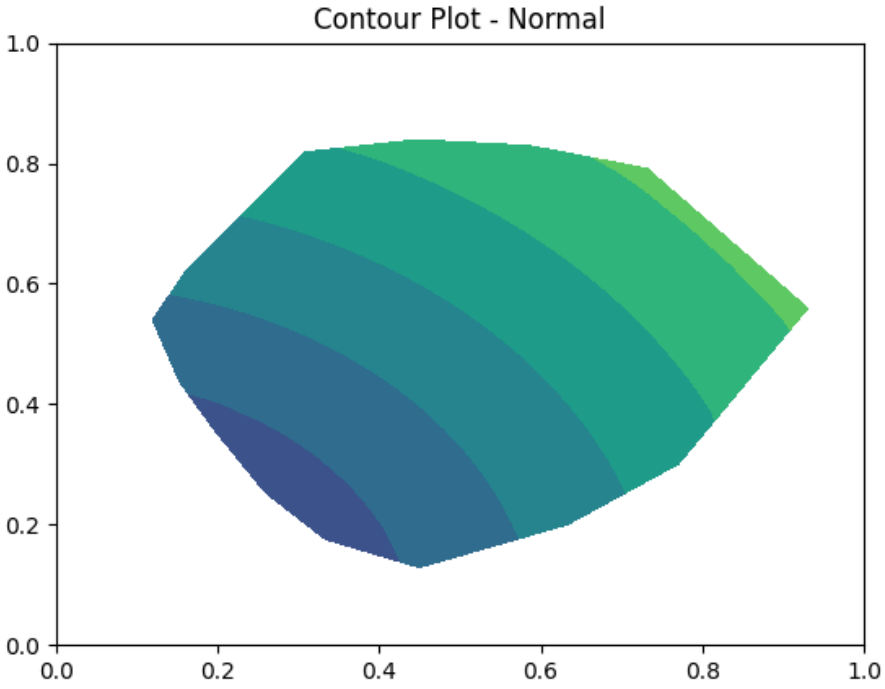
The above is the 100x100 image obtained by binning the normally sampled 2D array.

These images show that the points are densely concentrated in the centre in the case of normal sampling while it is quite homogenous in the case of uniform sampling.

C)



The above image is the contour plot of the uniformly sampled points.

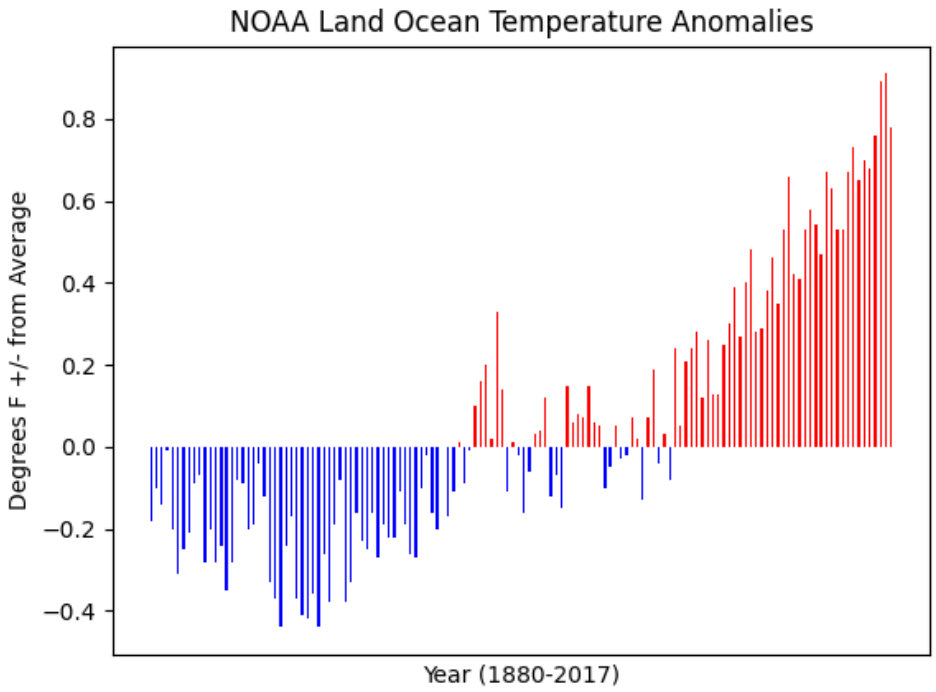


The above image is the contour plot of the normally sampled points.

To compute the contour plot, I have used the triangulation function available in matplotlib which computes the Delaunay triangulation of the point set. The scalar function defined on the set of points is their L2 norm. The 10 levels I have considered are [0,0.15,0.3,0.45,0.6,0.75,0.9,1.05,1.2,1.35,1.5].

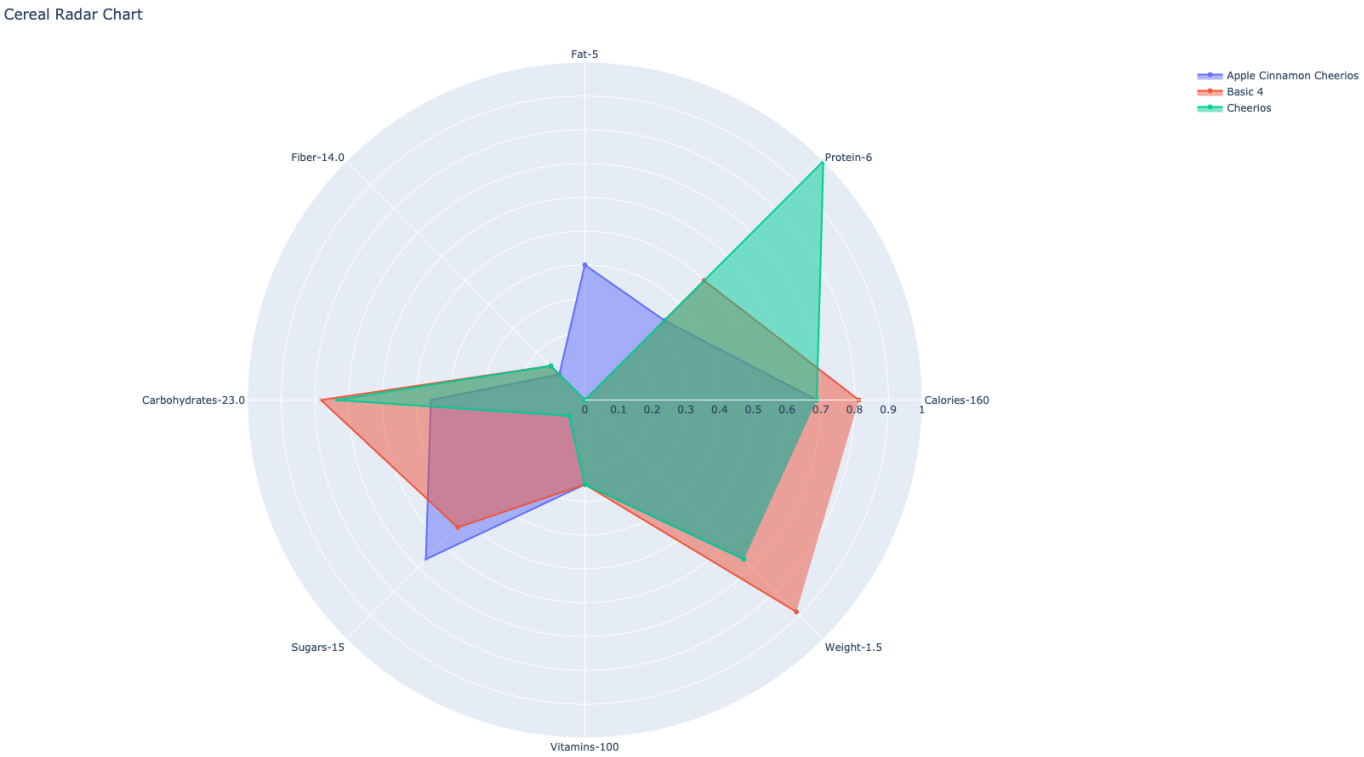
Part 2

1)



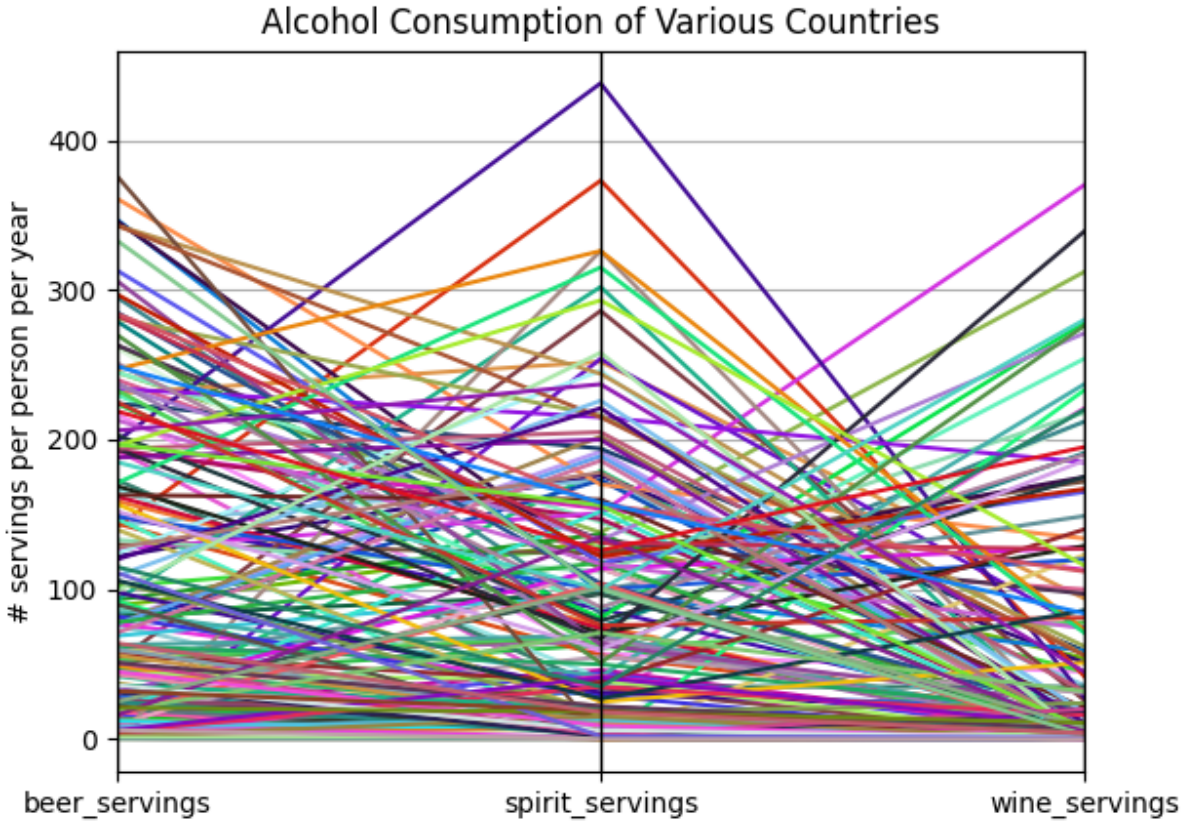
The general trend of the bar chart shows that the temperature in the recent past has been increasing and is significantly higher than the average temperature. This could be happening due to global warming.

2)



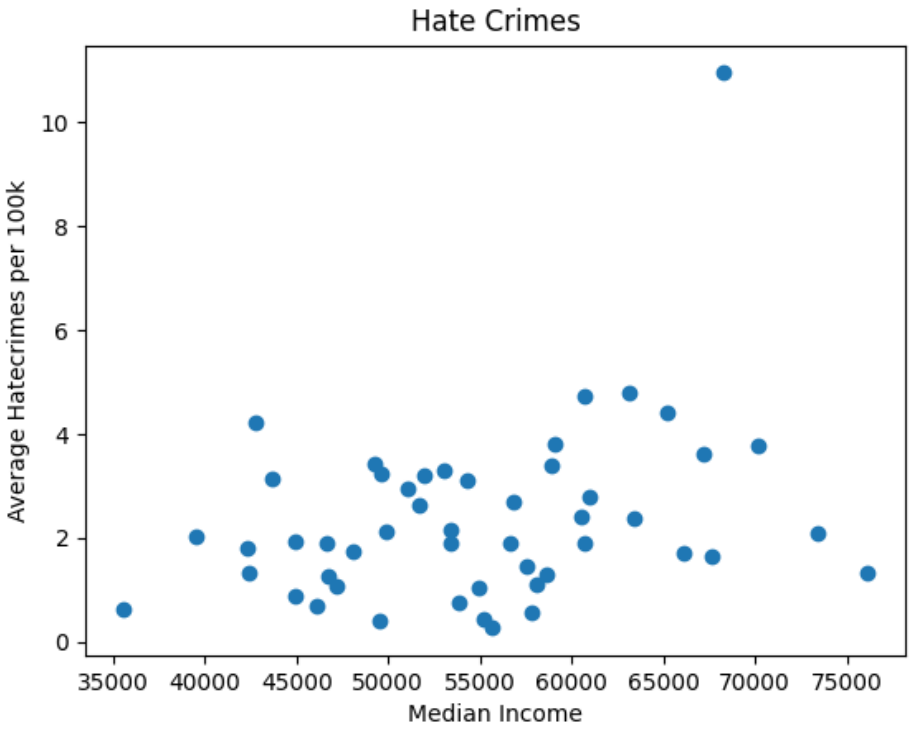
The above is the radar chart for three cereals and 8 nutritional statistics. The 8 statistics have been mentioned on the circumference of the circle along with the max value of that stat. So, for each cereal the statistic shown in the radar chart is % of the statistic wrt the maximum value. This was done because all the statistics have a different range and needed to be normalised.

3)



The above is a parallel coordinate plot depicting the alcohol consumption in various countries, where each polyline corresponds to a country.

Due to the presence of large amounts of crossings between beer servings and spirit servings, it can be inferred that in most countries beer and spirit are negatively correlated i.e only one of those two drinks are popular. The same observation can be made between Spirit and Wine.

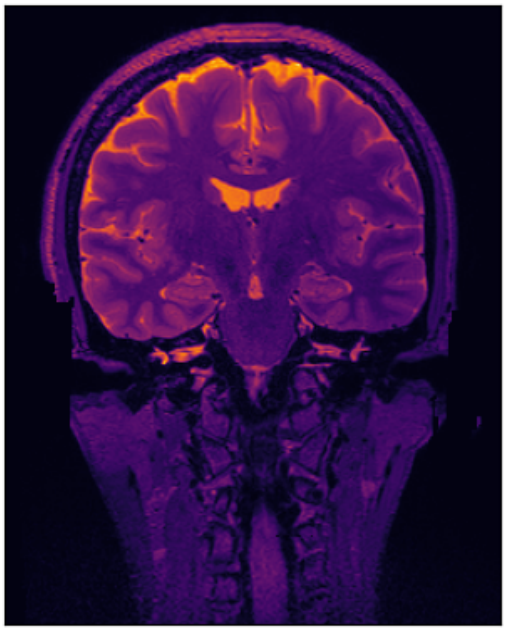


The following is a scatter plot between median income and average hate crimes where each point in the plot is a a state of the USA.

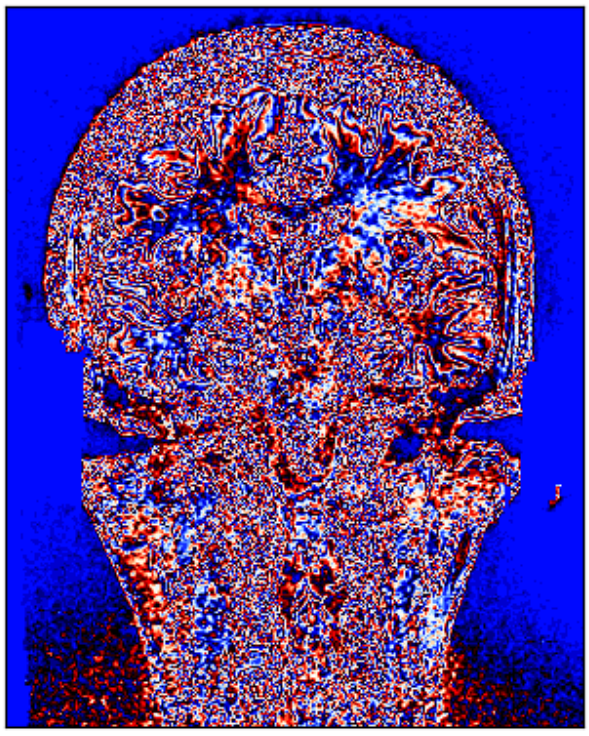
An observation which I found interesting is that, there does not seem to be any correlation between the income and the number of hate crimes. One would expect that income and crimes are negatively correlated but this visualisation clearly shows that such a relationship does not hold in the case of hate crimes.

Part 4

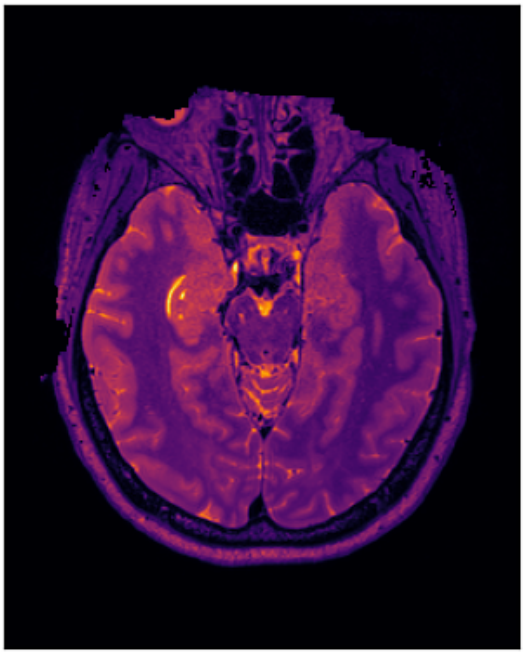
X-direction -150



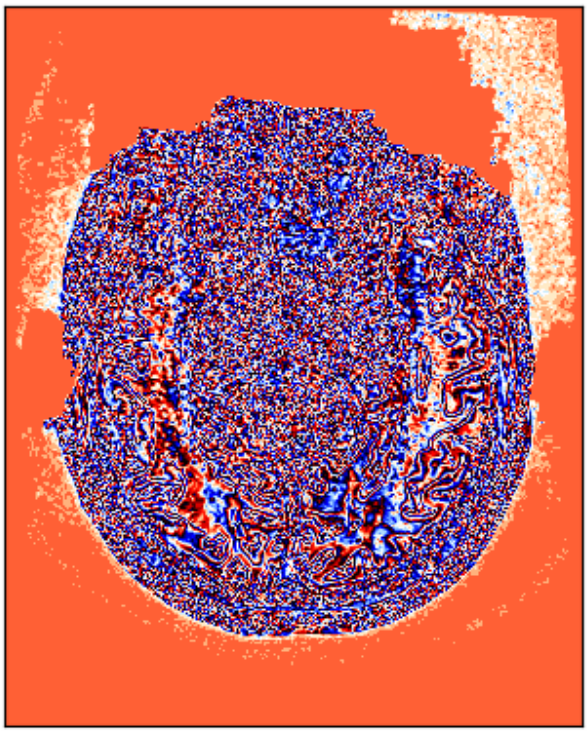
X-direction -150



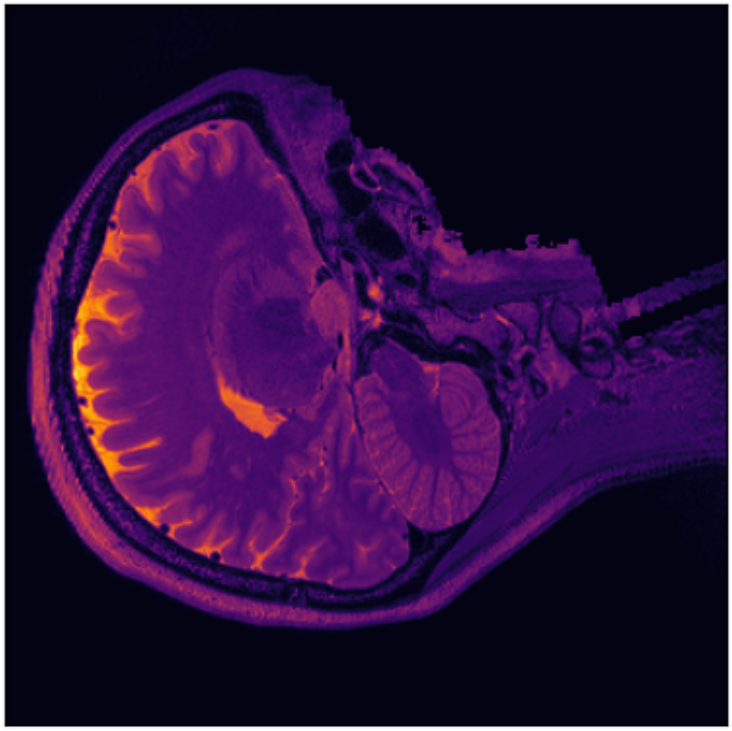
Y-direction -150



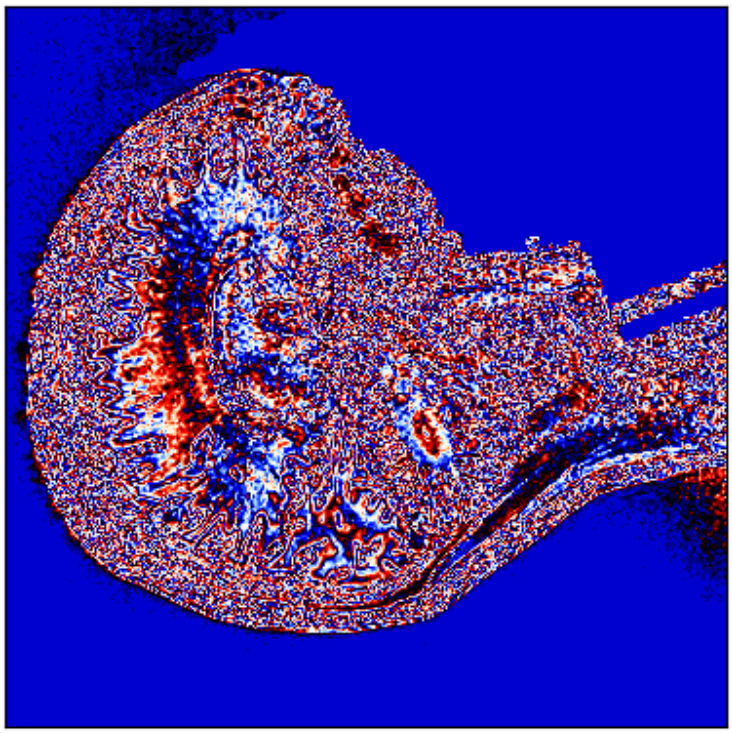
Y-direction -150



Z-direction -150



Z-direction -150



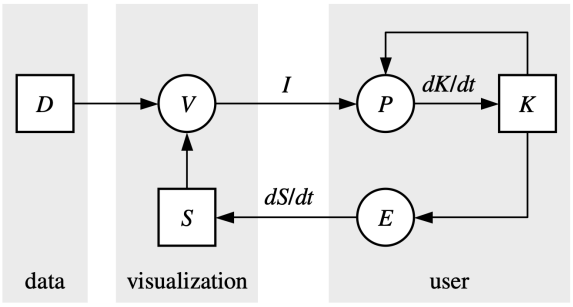
The above are sliced images from the given dataset. The title of each image describes the axis of slicing and the number of the slice.

The images have been generated using two colormaps namely, inferno and flag.
The importance of the choice of colormap is clearly evident from the images, ie in the case of the inferno colormap the details of the interior of the brain are clearly visible and is extremely useful to understand the brain's anatomy. But in the case of the flag colour map no information is visible and the visualization does not convey anything useful.
The inferno being a sequential colormap is extremely useful in this case, and any other sequential colormap would be quite useful in such a visualization.

PART3

1) It is essential to be able to assess the value of a visualization for making good choices of visualisations which could potentially be useful to a user. The two measures for assessing the value of a visualization are effectiveness and efficiency.

2)



Briefly, the given model says that given data D, this data is processed by a visualiser V which has a particular specification S at time t to output a time varying image I(t) = V(D,S,t). This I(t) is made available to the user who has perception P and knowledge K. The rate of increase of knowledge of the user is dK/dt = P(I,K) (this differential equation can be written up differently by opening up the integrand as well).
Further, in the case of interactive exploration, the specification S varies as dS/dt = E(K).

3) The costs are

A) Initial Development Cost
B) Initial Cost per User
C) Initial Cost Per Session
D) Perception and Exploration Cost.

4)

The cons of interaction are
A) allowing the user to modify the specifications will lead to subjectiveness making it quite hard to compare visualisations.
B) Interaction is expensive and will lead to a very high Perception and Exploration Cost.

The pros of interaction are
A) Interaction enhances the understanding of data. In cases where there is too much data, navigation and selection of data by the user is essential in interpreting and understanding the data.
B) Interaction is quite useful for the development of new methods of visualization, by exploring the entire 'visualization' space before coming up with the final prototype.

Conclusions

For generating the random arrays, I have used the np.random package. Most of the plots have been generated using matplotlib except for the radar chart and parallel coordinates, for which I used plotly and pandas respectively.

I just want to mention that, for all visualisations corresponding to the randomly generated 1D arrays, namely the box plots, histograms and cdf plots, could be made more close to the theoretical described distribution by increasing the number of samples. Even now, the general trends of the uniform and normal distributions are visible but they will be more clearer with an increase in the sample size.

The code and figures arein the following repo:

https://github.com/DhruvMeduri/CS6635.git