Note-taking assignment for Lecture week - 2
Beena Balakrishna

- The week 2 lecture started with a discussion on visualizing data that was taken from the following website https://projects.fivethirtyeight.com/mortality-rates-united-states/

- The following points were discussed in class based on the data from the previously mentioned website:
    1. What is the visualization trying to show?
       Responses :
        a. The data shows various causes of death in the United States of America and it's density in each county.
        b. A 'play' feature shows the rate of change of that particular cause of death in a given region for a time period (1980- 2014).
        c. Each case of death is denoted with a different colour. The colour scheme method of data representation is similar to a heat graph.
    2. What are it's methods?
       Responses:
        a. A colour scheme method is used to represent a given set of data.
        b. It was also noticed that some causes of death largely affected same regions.
        c. The data represented on the graph must've been collected from Census, medical reports, etc.
    3. What are the strengths/weaknesses?
       Responses:
        a. The data was easy to visualise and understand. Each data set was comprehended just by looking at the graph.
        b. Data was collected for each county and it recorded the various causes of death between the time period. This collection of data is huge but it was well represented by the graph.
        c. The weakness from the dataset is that the range of 'DEATHS PER 100K PEOPLE' changed for each cause of death. If a comparison was to be made to check which disease/reason was the cause of maximum or minimum deaths for a specific county for the year 2000 then it would be a task to do so.

- Some of the topics that were later discussed in the class were:
    1. How does drawing work?
    2. The different methods of palette used are mutation, histograms and aggregation and splitting.
    3. Different elements of a visualisation.

- Two videos were shown in class on to show how video games were initially designed and the following are the links to the videos:
    1. https://www.youtube.com/watch?v=qfDxiVpgjiM
    2. https://www.youtube.com/watch?v=crBo7voJTgY

- A brief amount of time was spent in discussing on how a line can be drawn with 5 parameters using Python code. The same line is then drawn on a 8x8 grid, 16x16 grid and 32x32 grid. The difficulties to fit a line in a grid system were exaplined.

```
start = (x0, y0)
end   = (x1, y1)
width = 1.0
```

- As the lecture progressed, the two main methods of representing an image were discussed and they are the raster representation and vector representation.

    1. Raster: This method of image representation uses pixels and each pixel was denoted by a colour or number of colours depending on the complexity of the image. Pixels take a small amount of storage space to denote data. The most common file formats for raster type of images are GIF (Graphic Interchange Format), JPG (Joint Photographic Experts Group) and PNG (Portable Network Graphics). The compression tools for a raster image are JPG (lossy compression technique) and PNG (lossless compression technique). If a raster image had to be edited then the software Paint and Photoshop could be used for efficient results.

    2. Vector: In this method of image representation, each component is associated with properties that is defined by a single point or a set of points that constitute a 'drawing'. Such an image can comprise of points, paths, patterns, shapes or text. The most common file formats are SVG (Scalable Vector Graphics) , PDF (Portable Document Format) and EPS (Encapsulated Post Script). Vector image is suggested to be used only when small set of data is to be represented. One of the efficient compression tools for vector image is text compression (lossless compression technique). The editing software for a vector image is an Illustrator and Inkscape.

Raster Representation

	1 Line	10 Lines	1000 Lines	1e6 Lines
600x600	45kb	45kb	45kb	45kb
1200x1200	180kb	180kb	180kb	180kb
2400x2400	720kb	720kb	720kb	720kb

(uncompressed, 1-bit images)

600 x 600 = 360,000 bits / 8 = 45,000 bytes = 45 kilobytes

Figure1: Data storage utilized for raster image

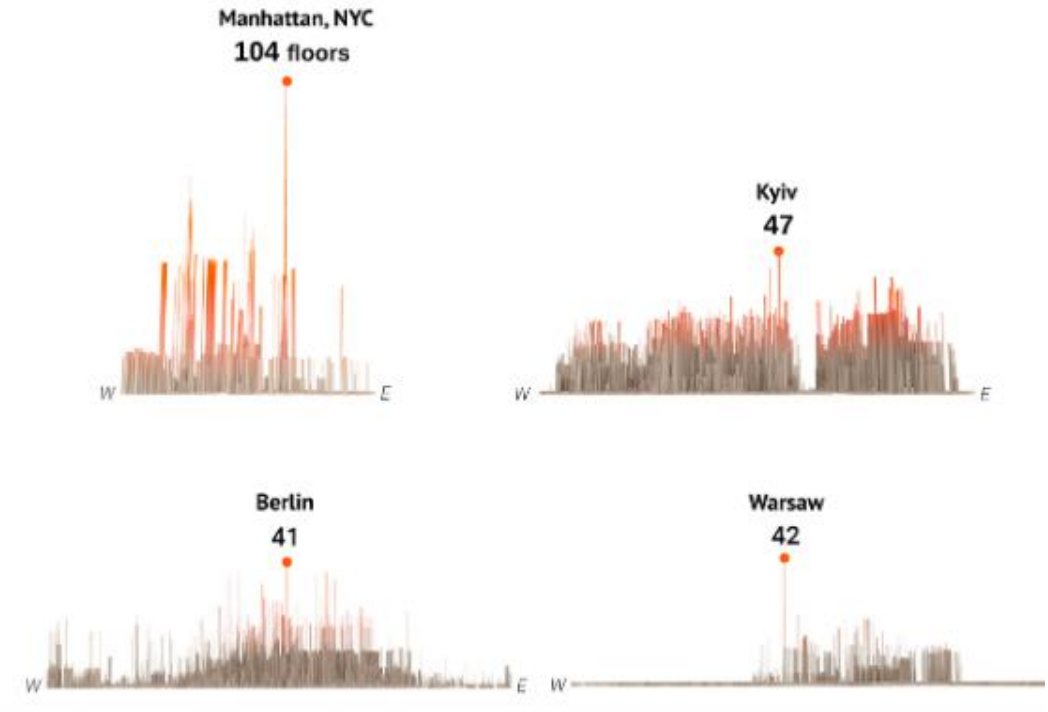
Vector Representation

	1 Line	10 Lines	1000 Lines	1e6 Lines
600x600	5 bytes	50 bytes	5kb	5mb
1200x1200	5 bytes	50 bytes	5kb	5mb
2400x2400	5 bytes	50 bytes	5kb	5mb

(uncompressed, single precision)

Figure2: Data storage utilized for vector image

- Mutation operation, a visualisation palette was discussed in details in this week's lecture. An example of a histogram was considered to explain the mutation operation as shown below:



1. What can we understand from these plots?
   Responses:
   a. The longest line in each histogram denotes the tallest building in the city.
   b. The location of the buildings can be found out by looking at the horizontal scale that shows the direction (West to East).
   c. The change in the colour of buildings in each city is based on the number of buildings and the average of height of all the buildings.

NOTE: Definitions of some important terminologies.

1. Fractal: Are a method of generating never-ending patterns. They are infinitely complex patterns that are self-similar across different scales. They are created by repeating a simple process over and over in an ongoing feedback loop. Driven by recursion, fractals are images of dynamic systems. Geometrically the nature is made up of fractals for example trees, mountains, clouds, etc.

2. Data binning: Is defined as a data pre-processing technique used to reduce the effect of minor observation errors. It is a method of quantisation in which original data values fall that fall in a small interval (bin) are replaced by an interval with a central value.

- In the last session of the lecture, Python code using matplotlib was discussed. The following link leads to the Notebook that includes the codes and graphs discussed in the lecture class
  https://uiuc-ischool-dataviz.github.io/spring2019/nbv.html?notebook_name=%2Fspring2019%2Fweek03%2Fexamples_week03_notes.ipynb