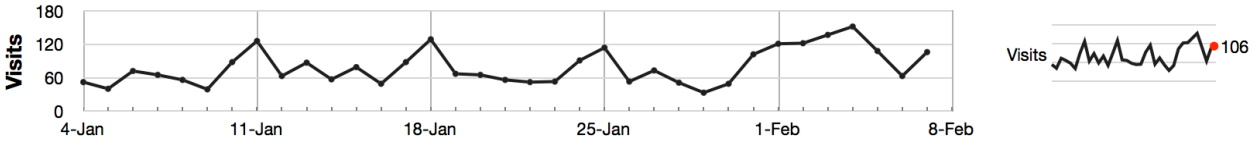
Shivangi Kochrekar
BE COMPS A
2018130020

Data Science, 2022
Tut 8: Information Visualization

Question 1

The data shown here are the number of visits to a university website for a particular statistics course. There are 90 students in the class.



1. What are the names (type) of the 2 plots shown?

   The first plot is a Time Series plot. The second plot is Sparkline plot. It shows exactly same data, in a more compact form without labeling axes.

2. List any 2 interesting features in these data.

   There is large amount of traffic in the first week of February. As the semester started in January, it is possible that there is some midterm test in February leading to rise in number of visits.
   There are 90 students in class but some days have more than 90 visits. This scenario indicates that students visit the site more than once per day, or there may be some visits to the site by external users.

Question 2

What are the names of the axes on a bar plot?

The x-axis represents the category axis and y-axis represents value axis. It maybe vice-versa in case of horizontal bar plot.

Question 3

Which types of features can the human eye easily pick out of a time series plot?

The human eye can easily pick up features of a time series plot such as upward and downward trends, gaps i.e., missing values, spikes, sinusoids (plots such as sine and cosine).

Question 4

Why is the principle of minimizing “data ink” so important in an effective visualization? Give an scientific or engineering example of why this important.

The data-ink ratio is the proportion of Ink that is used to present actual data compared to the total amount of ink (or pixels) used in the entire display. The principle of minimizing “data ink” reduces the time or work to interpret that plot, by eliminating elements that are non-essential to the plot’s interpretation.
It is important in situations that are time and safety critical, for example operation rooms in medical facilities, operator control rooms, etc.

Question 5

Describe what the main difference(s) between a bar chart and a histogram are.

Bar Chart	Histogram
Bar charts are used to compare variables	Histograms are used to show distributions of variables.
Bar charts plot categorical data.	Bar Charts plot quantitative data with ranges of the data grouped into bins or intervals.
In bar chart, bars can be reordered	In histogram, bars cannot be reordered.
There are spaces between the variables of a bar chart.	There are no spaces between the bars of a histogram since there are no gaps between the bins. An exception would occur if there were no values in a given bin but in that case the value is zero rather than a space.
The bars of bar charts typically have the same width.	The widths of the bars in a histogram need not be the same as long as the total area is one hundred percent if precents are used or the total count if counts are used.
The values in bar charts are given by the length of the bar.	The values in histograms are given by areas.