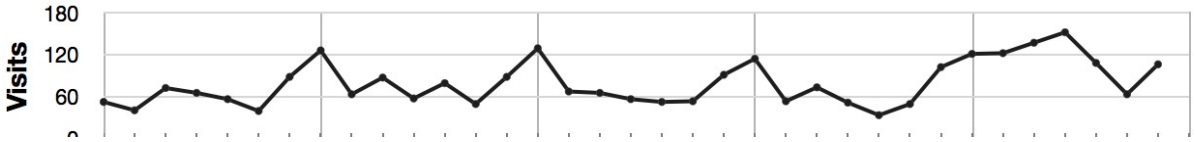
Data Science, 2022
Tut 8: Information Visualization

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?
2. List any 2 interesting features in these data.

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.

Features shown in the data are:

- A noticeable weekly cycle; probably assignments are due the next day!
- A sustained, high level of traffic in the first week February - maybe a midterm test.
- Some days have more than 90 visits, indicating that students visit the site more than once per day, or due to external visitors to the site.

Question 2

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

The category axis and value axis.

Question 3

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

Features such as sinusoids, spikes, gaps (missing values), upward and downward trends are quickly picked out by the human eye, even in a poorly drawn plot.

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.

It reduces the time or work to interpret that plot, by eliminating elements that are non-essential to the plot’s interpretation. Situations which are time or safety critical are examples, for example in an operator control room, or medical facility (operating room).

Question 5

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

- Histograms are used to show distributions of variables while bar charts are used to compare variables.

- Histograms plot quantitative data with ranges of the data grouped into bins or intervals while bar charts plot categorical data.

- Bars can be reordered in bar charts but not in histograms.

- 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. On the other hand, there are spaces between the variables of a bar chart.

- 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 percents are used or the total count if counts are used. Therefore, values in bar charts are given by the length of the bar while values in histograms are given by areas.