Data visualization is high band connection between computer to human brains. It's about understanding the data and making decisions.

Visualization Wheel (Alberto Cairo)

This wheel is a guide how to make a visualization based on our audience and find a right balance between density and lightness, novelty and redundancy, functionality and decoration. For example, here are some explanation of each line in wheel graph:

- Density: heavy amount of information is displayed
- Unidimensional: only 1 way to explore the data
- Redundancy: reusing the basic concept in graph to explain the same thing in multiple ways.
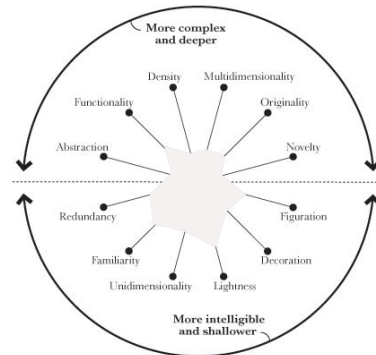
Graphical Heuristic, Data-Ink Ratio (Edward Tufte)

Heuristic is a process or rule to helps you make decision, by being practical rather than perfect or optimal. Edward Tufts talks about data ink ratio to chart junk. Data ink ratio is the amount of data-ink divided by the total ink required to print the graphics. Basically, he suggests removing those elements which don't add information to the graphics. Here is an example, in graph below, we need to remove the background, since it doesn't add any valuable information, remove the legend since it's already marked on x axis, remove the y axis since the table name shows what is showing already. Finally, by these changes we end up with cleaner and smaller image, shown on the right-hand side.
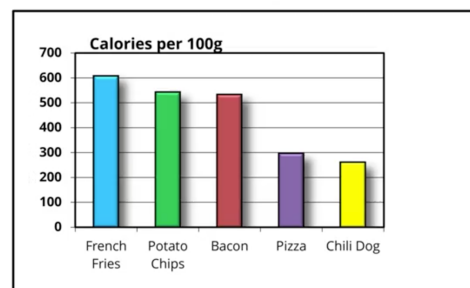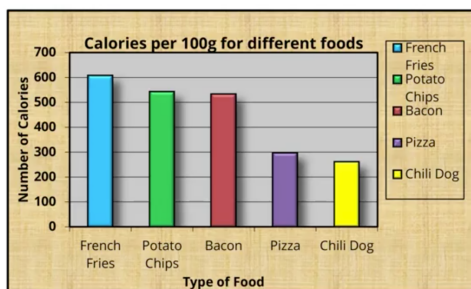
In best practice, it's good to replace the color with patterns. This is called hatching in graphics. Use color when you want to emphasize on specific part of graph, as shown in plot here. Moreover, showing the bar height in each bar, makes reading the plot easier.
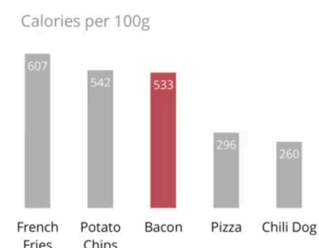

Chart Junk (Tufte)

There are three kind of junk chart. The first one is unintended optical chart. For example, excessive shading, patterning of chart features makes the human eye jump and cause visual fatigue. This phenomenon is called moire patterns. The second from of chart junk is the grid. Grid
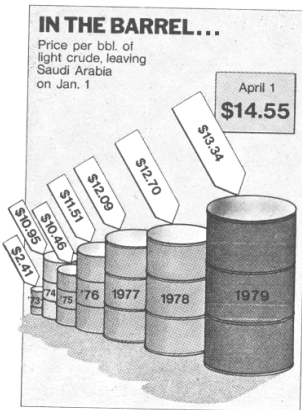
is unnecessary and causes competitions with the actual data shared. Direct labelling of data is a great way of reducing this form of chart junk. The last form of chart junk is the duck. You remember legs as how the trend of diamond is as shown in a picture. The research shows the duck chartjunk is easier to recall in a long run. Duck chartjunk are easier to remember the details, more attractive and nicer to look at.
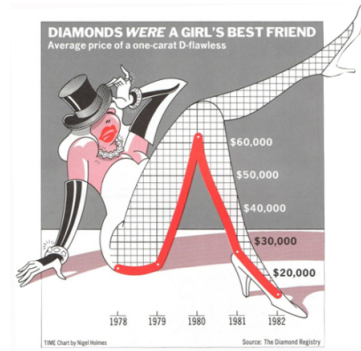
Sparkline (Tufte)

Sparkline are data words, an intriguing way to bridge the gap between the text and figures. This could be embedded in tables along with the data you described. Excel is using this technique in their spreadsheet. Here is an example of Sparkline in excel. They give general feeling about trending of the data. The last Tufte principle is lie factor

defined as the size of an effect shown in the graphic divided by the size of the effect actually in the data. For example, in the image below, the price is shown over years., but it's unclear whether the barrel sizes are different or not. It's unclear the price and size of gasoline changes overtime or it's a perspective showing the recent barrel bigger than the rest.

Qualities of Visualization (Cairo)
- Truthful: make sure the graphs are truthful and not misleading. Sometimes the truth is subjective (example: http://infovis.fh-potsdam.de/readings/Cairo2015.pdf)
- Functionality: try to increase functionality and ease of understanding of the charts.
- Beautiful: Beauty varies. Remember that different perception is based on different audience.
- Insightful: Present in a way that viewer says aha! the right figure drives home the result quickly.
- Enlightening: combination of previous four, but with a social ethical responsibility.

Procedural vs declarative information visualization method
In simple words, here is an easy example:
Procedural(Matplotlip):
1. Go to kitchen
2. Get Sugar, Milk and Tea,
3. Mix them and heat over fire till it boils
4. Put that in a cup and bring it to me

Declarative (like D3.js):

1. Get me a cup of tea.

In procedural language, you define the whole process and provide the steps how to do it. You just provide orders and define how the process will be served.

In declarative language, you just set the command or order, and let it be on the system how to complete that order. You just need your result without digging into how it should be done.

Matplotlip (Python Visualization)

This is a great tool for visualization. This has several backend. Backend is an abstraction layer which knows how to interact with operating environment.

the next layer of Matplotlib is artist layer. It's an abstract around drawing and layout primitives. This is heavily relying on axes, primitives and collections. Matplotlib also has a scripting layer, that mainly used by professionals (Pyplot)

- **Backend Layer**
  - *Deals with the rendering of plots to screen or files*
  - *In Jupyter notebooks we use the inline backend*
- **Artist Layer**
  - *Contains containers such as Figure, Subplot, and Axes*
  - *Contains primitives, such as a Line2D and Rectangle, and collections, such as a PathCollection*
- **Scripting Layer**
  - *Simplifies access to the Artist and Backend layers*

Matplotlib has so many options to use like scatterplot, line plots, bar charts and etc. here is an example how to follow the rule to dejunkify the plot.

Before and After: