

DATA-DRIVEN PROBLEM SOLVING IN MECHANICAL ENGINEERING

Data Visualization

MASOUD MASOUMI

ME 364 - Spring 2022

Department of Mechanical Engineering
The Cooper Union for the Advancement of Science and Art

February, 2022

Why Data Visualization?



Exploratory Data Analysis

Getting a hand on what our data really look like
and helps us to search for patterns in the data

Error Detection

Get to see the outlier points, insufficient cleaning
issues, and erroneous assumptions

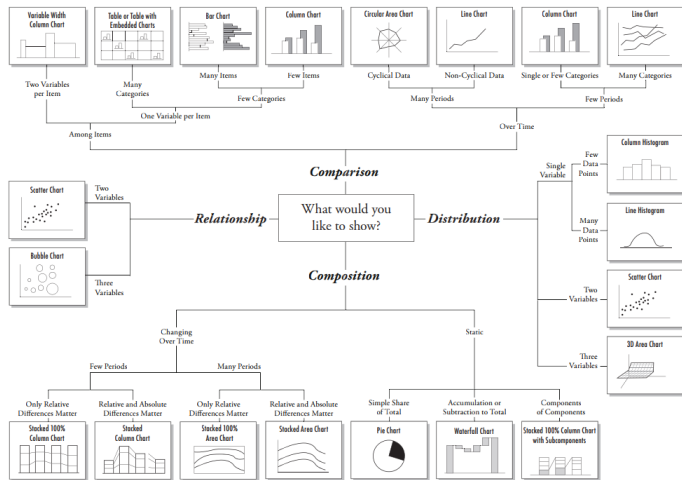
Communication

Present the results effectively, since a picture is
worth a thousands words!

What Type of Chart Should be Used?



You need to produce the right plot for a given data, not just the first thing that comes to mind!



www.ExtremePresentation.com
© 2009 A. Abdel - a.abdel@gmail.com

[link to this chart: <https://extremepresentation.typepad.com/files/choosing-a-good-chart-09.pdf>]



They have advantages:

- Precision representation
- Scale representation
- Multivariate visualization
- Heterogeneous data presentation

Best practices:

- Order columns to highlight importance
- Use *emphasis*, FONT, or color
- Avoid excessive-length column descriptive

Here is an [example](#) of a not great table converted into a much better [table](#)!

Dot and Line Plot



They provide the best way to visually represent a set of (x,y) points.

Best practices:

- Show the data points, not just fits!
- Accept and show uncertainty in the data
- Use [box-and-whisker](#) plots if needed ([example](#))
- Use color and hatching to distinguish lines
- Never connect points for categorical data

Scatter Plot



They are used to represent massive data sets with large number of points.

Best practices:

- Use the right-sized dots ([example](#))
- Do not use 3D scatter plots, unless showing a point could of a 3D object
- To represent 3D data, use bubble plots and change color, shape, size, and shading of the dots ([example](#))

Bar Plot and Pie Chart



Represent the relative proportions of categorical variables.

Partition a geometric whole (circle or bar) into areas proportional to the frequency of each group.

Best practices:

- Directly label slices of the pie
- Use bar charts for precise comparison
- Do not use 3D pie chart. It is chartjunk! ([example](#))

Histogram



They are plots of the observed frequency distribution.

They try to answer questions like, where is peak of the distribution? Is the mode near the mean? Is the distribution symmetric or skewed?

Best practices:

- Use the appropriate number of bins ([example](#))
- Turn the histogram into a probability density function or *pdf* ([example](#))