Elementary Plots
Why Should We Care?

- Everyone uses plotting
- But...most people ignore or are unaware of simple principles
- Default plotting tools are not always the best

- More importantly, it is easy to lie or deceive people with bad plots
Some examples of available plotting tools!

- [Grace](http://plasma-gate.weizmann.ac.il/Grace/)
- [gnuplot](http://www.gnuplot.info/)
- [QtiPlot](http://www.proindependent.com/pricing.html)
- [Excel](http://office.microsoft.com/en-us/excel/default.aspx)
- [The MathWorks](http://www.mathworks.com/)
- [AptPlot](http://www.aptplot.com/)
- [SigMaplot](http://www.sigmaplot.com/products/sigmaplot/sigmaplot-details.php)
- [Matplotlib](http://matplotlib.sourceforge.net/)
- [Wolfram](http://www.wolfram.com/)
What Can Plots Do?

- Data analysis and communication
  - In a simplistic view, plotting reduces a large amount of information to a smaller form that is more easily understood via certain graphical representation.
  - Reduction of the data to its simplest and cleanest form, such that the relationships inherent in the data (points) are easily perceived.
Examples of plots generated by a number of tools using their default setting

Default Excel Plot

Default Matplotlib/Matlab Plot

Default Pages Plot
Examples of plots generated by a number of tools using their default setting

- Why are they all different?
- What is good/bad about each?
These plots demonstrate two important points:

First, there is no obvious standard for what a plot should look like. This is easy to see by the differences in the axes and scale lines, the data rectangle inside the plot, and the actual representation of the data values.

Second, creating a plot is an iterative process that can not be generally applied to all types of data.

There are no magic formulas for creating a useful plot. However, some general principles have been advocated that can be applied to plots to improve their likelihood of being useful.
The information provided here should be considered as guidelines.

**PRINCIPLES OF PLOTTING**

*Visualizing Data* [Cleveland 93] and *Elements of Graphing Data* [Cleveland 94] by William S. Cleveland
Principles of Plotting

• Improving the vision
  – Improve the readability of the plot

• Improving the understanding
  – Ensure that the analysis of the plot is effectively communicated.
Improving the Vision

- Principle 1: Reduced clutter, Make data stand out
  - The main focus of a plot should be on the data itself, any superfluous elements of the plot that might obscure or distract the observer from the data needs to be removed.

Which one is better?
Improving the Vision

• Principle 2: Use visually prominent graphical elements to show the data.
  – Connecting lines should never obscure points and points should not obscure each other.
  – If multiple samples overlap, a representation should be chosen for the elements that emphasizes the overlap.
  – If multiple data sets are represented in the same plot (superposed data), they must be visually separable.
  – If this is not possible due to the data itself, the data can be separated into adjacent plots that share an axis.
Improving the Vision

• Principle 3: **Use proper scale lines and a data rectangle.**
  – Two scale lines should be used on each axis (left and right, top and bottom) to frame the data rectangle completely.
  – Add margins for data to make the plot prominent.
  – Tick-marks out and 3-10 for each axis.
Improving the Vision

• Principle 4: Reference lines, labels, notes, and keys.
  – Reference lines are only used to show the thresholds within data.
  – Only use them sparsely when necessary and don’t let them obscure data.
Improving the Vision

• Principle 4: Reference lines, labels, notes, and keys.
  – Only use them sparsely when necessary and don’t let them obscure data.
Improving the Vision

• Principle 5: Superposed data set
  – Symbols should be separable and data sets should be easily visually assembled.
Improving the Understanding

- Principle 1: Provide explanations and draw conclusions
  - A graphical representation is often the means in which a hypothesis is confirmed or results are communicated.
  - Describe everything, draw attention to major features, describe conclusions

Explain everything in the plot.
Do not let them guess.
Improving the Understanding

- Principle 2: Use all available space.
  - Fill the data rectangle as much as you can, only use zero if you need it *(for scientific data)*
Improving the Understanding

• Principle 3: Align juxtaposed plots
  – Make sure scales match and graphs are aligned
Improving the Understanding

- Principle 4: Use log scales when appropriate
  - Used to show percentage change, multiplicative factors and skewness
Improving the Understanding

- Principle 5: Bank to $45^\circ$ (optional!!!)
  - Optimize the aspect ratio of the plot
Summary of Principles

• Improve vision
  1. Reduced clutter, Make data stand out
  2. Use visually prominent graphical elements
  3. Use proper scale lines and a data rectangle
  4. Reference lines, labels, notes, and keys
  5. Superposed data set

• Improve understanding
  1. Provide explanations and draw conclusions
  2. Use all available space
  3. Align juxtaposed plots
  4. Use log scales when appropriate
  5. Bank to 45°
SIMPLE PLOTTING TECHNIQUES
Connected Symbol Plots

- The most common plotting technique
- Used to plot time series or other 1D data

![Average Monthly Temperature Graph](image1)

![Air temperature Graph](image2)
**Connected Symbol Plots**

**Symbols.** For noisy data that shows high frequency characteristics

**Connections.** For smooth data that shows low frequency characteristics

**Connected Symbols.** The symbols demonstrate the actual concentrations of the data, while the path that the data takes can be better followed using connections.
Connected Symbol Plots

**Symbols.** For noisy data that shows high frequency characteristics

**Connections.** For smooth data that shows low frequency characteristics

**Connected Symbols.** The symbols demonstrate the actual concentrations of the data, while the path that the data takes can be better followed using connections.
Dot Plots

- Similar in nature to bar charts or pie charts
- Should be used for quantitative labeled data

The data points do not have sequential relation!!

A dot plot showing the odds of dying.
Dot Plots

The values should normally be sorted such that the largest value is at the top.
- Exception: the data has an inherent order that must be preserved

A log scale should be used to reduce skewness in the data

A dot plot showing the odds of dying.
Dot Plots

- Real world data is not always univariate.
- To represent multi-dimensional data, a multiway dot plot can be used.

A dot plot showing the odds of dying.
Dot Plots

- A multiway dot plot is just several dot plots that share common labels and are juxtaposed such that they share an axis.

![Medal Rankings from 2008 Beijing Olympics](image-url)
Scatter Plots

- Scatter plots are used to show how one variable is affected by another, or correlated, in 2D data.

A scatter plot showing the biological principle of scaling for mammals. For each sample, the metabolic rate is plotted against the body mass to show a high correlation between the two variables. The points have also been labeled to provide additional information.
Scatter Plots

• If used properly, the correlation of the data can easily be discerned.

Scatter plots showing different levels (high, low, and no, respectively) of correlation for points generated with different magnitudes of randomness.
Scatter Plots

- It is often desirable to express the correlation as a line that provides the best fit for the data.

Linear regression using least squares fits a line to the data. The fit is good for high and low correlation (left and middle), but can result in problems in the case of outliers (right).
Scatter Plots

- As with dot plots, scatter plots can be used to represent data in higher dimensions. This is frequently done with a scatter plot matrix.
- This assigns each dimension of the plot to a single row and column in the matrix. The variables are then plotted against each other as a standard scatter plot for each entry in the matrix.
Histograms

- Histograms are a special type of bar charts used for plotting distributions in data.
- The horizontal axis represents fixed intervals of the data and the vertical axis represents the number of values that lie within the intervals.
Box Plots

• Box plots are typically used to represent the statistical variation in the data
Brief Overviews of Types of Graphs

2D Graphs
- Bar/Column
- Bar Dev
- Bar Left Y
- Bar Right Y
- Bar Top
- Bar X
- Box
- Detrended Probability
- Half-Normal Probability
- Hanging Bar Histograms
- Histograms
- Line
- Pie Charts
- Probability
- Probability-Probability
- Quantile-Quantile
- Range
- Scatterplots
- Sequential/Stacked
- Voronoi Scatterplot

3D Graphs
- Spectral
- Trace

3D Sequential Graphs
- Bivariate Histograms
- Box
- Range
- Raw Data Contour/Discrete
- Sequential Contour
- Sequential Surface
- Raw Data Spikes
- Raw Data Surface

3D Categorized Graphs
- Contour
- Deviation
- Scatterplots
- Space
- Spectral
- Surface

Ternary Categorized Graphs
- Ternary Contour/Area
- Ternary Contour/Line
- Ternary Scatterplot

4D/Ternary Graphs
- Scatterplots
- 3D Ternary
- Contour/Area
- Contour/Line
- 3D Deviation
- 3D Space

nD/Icon Graphs
- Chernoff Faces
- Columns
- Lines
- Pies
- Polygons
- Profiles
- Stars
- Sun Rays

3D XYZ Graphs
- Contour
- Deviation
- Scatterplots
- Space

2D Categorized Graphs
- Detrended Probability
- Half-Normal Probability
- Normal Probability
- Probability-Probability
- Quantile-Quantile

Matrix Graphs
- Columns
- Lines
- Scatterplot

http://www.statsoft.com/Textbook/Graphical-Analytic-Techniques