Data Analysis in Astronomy and Physics

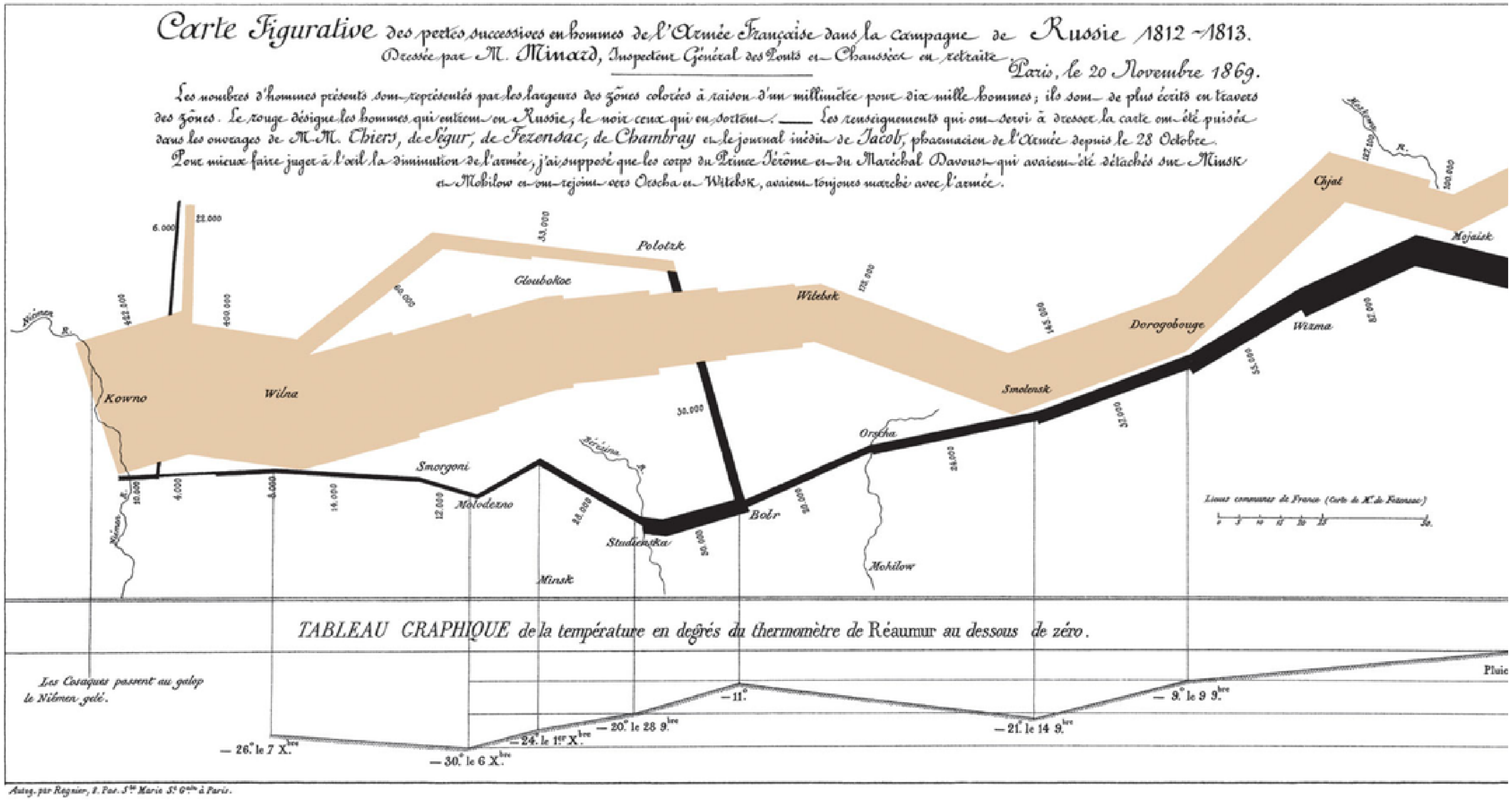
Lecture 2: Introduction to Data Visualization

M. Röllig - SS22

Introduction

The most promising method of understanding huge amount of data being generated by modern science is by visualization. It is estimated that 50% of the brain’s neurons are associated with vision. **“The purpose of [scientific] computing is insight, not numbers.”** Richard Hamming,1982. See also: https://www.masswerk.at/minard/

Out[ ]=



Introduction

Wikipedia: Charles Joseph Minard was a pioneer of the use of graphics in engineering and statistics. He is famous for his *Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813*, a flow map published in 1869 on the subject of Napoleon's disastrous Russian campaign of 1812.

- The graph displays several variables in a single two-dimensional image:
  - the size of the army - providing a strong visual representation of human suffering, e.g. the sudden decrease of the army's size at the battle crossing the Berezina river on the retreat;
  - the geographical co-ordinates, latitude and longitude, of the army as it moved;
  - the direction that the army was traveling, both in advance and in retreat, showing where units split off and rejoined;
  - the location of the army with respect to certain dates; and
  - the weather temperature along the path of the retreat, in another strong visualisation of events (during the retreat "one of the worst winters in recent memory set in"[1]).

Data Visualization

The classical definition of visualization is as follows: the **formation of mental visual images**, the act or process of **interpreting in visual terms** or of **putting into visual form**.

- umbrella term, usually covering both *information* and *scientific visualization*.
- anything that **converts data sources into a visual representation** (like charts, graphs, maps, sometimes even just tables).

Scientific Visualization

- visualization of scientific data that have close ties to real-world objects with spatial properties
- example might be visualizations of air flow over the wing of an airplane, or density structure of MHD simulations
- The goal is often to generate an image of something for which we have spatial information and combine that with data that is perhaps less directly accessible, like temperate or pressure data.
- The different scientific fields often have very specific conventions for doing their own types of visualizations.

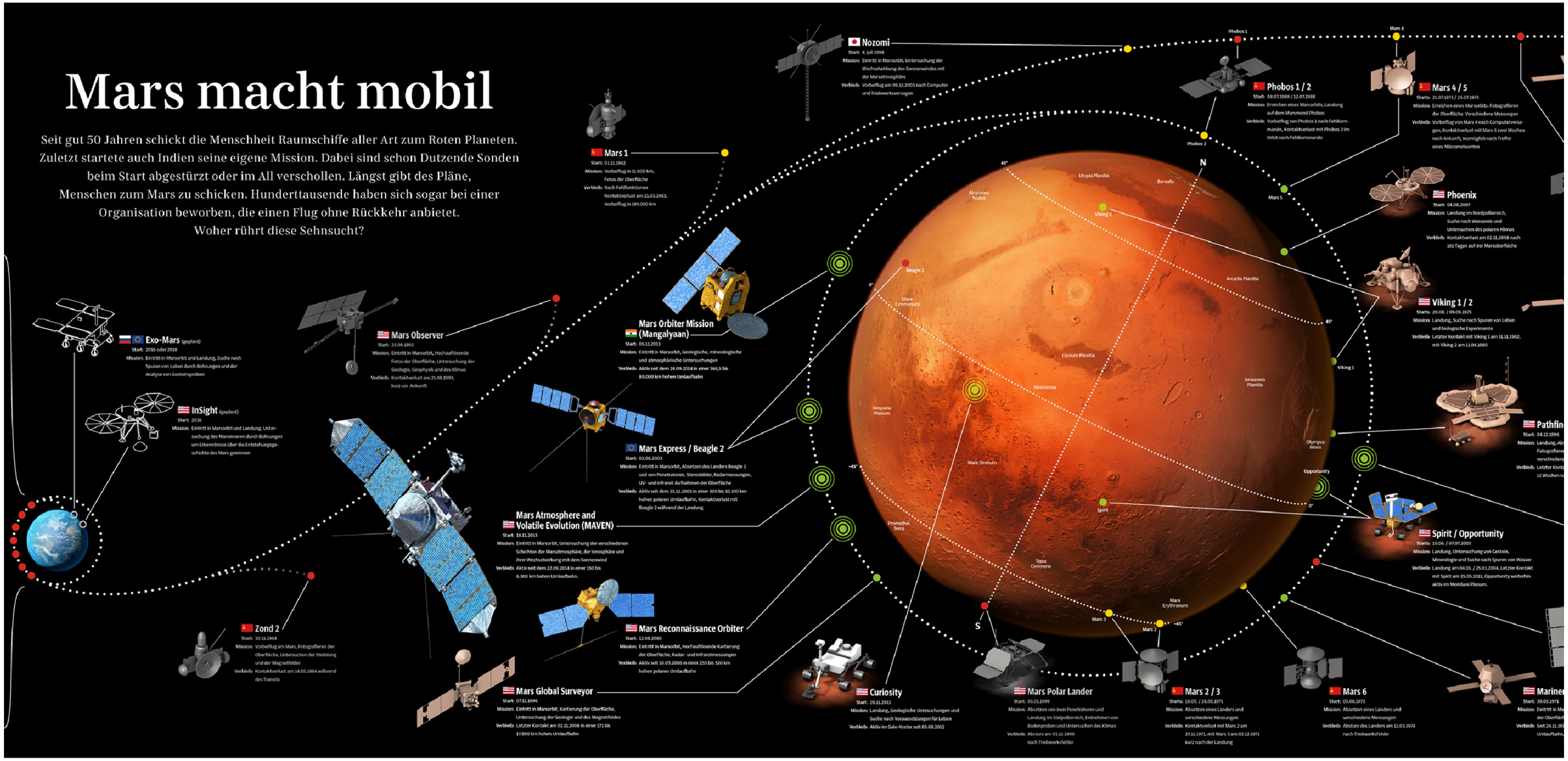
“The graphical representation of complex physical phenomena in order to assist scientific investigation and to make inferences that aren’t apparent in numerical form. Typical examples include processing of satellite photographs and 3D representations of molecules and fluids to examine their dynamics.[Usability First, 2003]”

Information Visualization

- broad term, covering most statistical charts and graphs but also other visual/spatial metaphors that can be used to **represent data sets that don’t have inherent spatial components**.

Infographic

- specific sort of genre of visualizations.
- Infographics have become popular on the web as a way of combining various statistics and visualizations with a narrative and, sometimes, a polemic.



Credits: http://arminschieb.com/mars-missions/

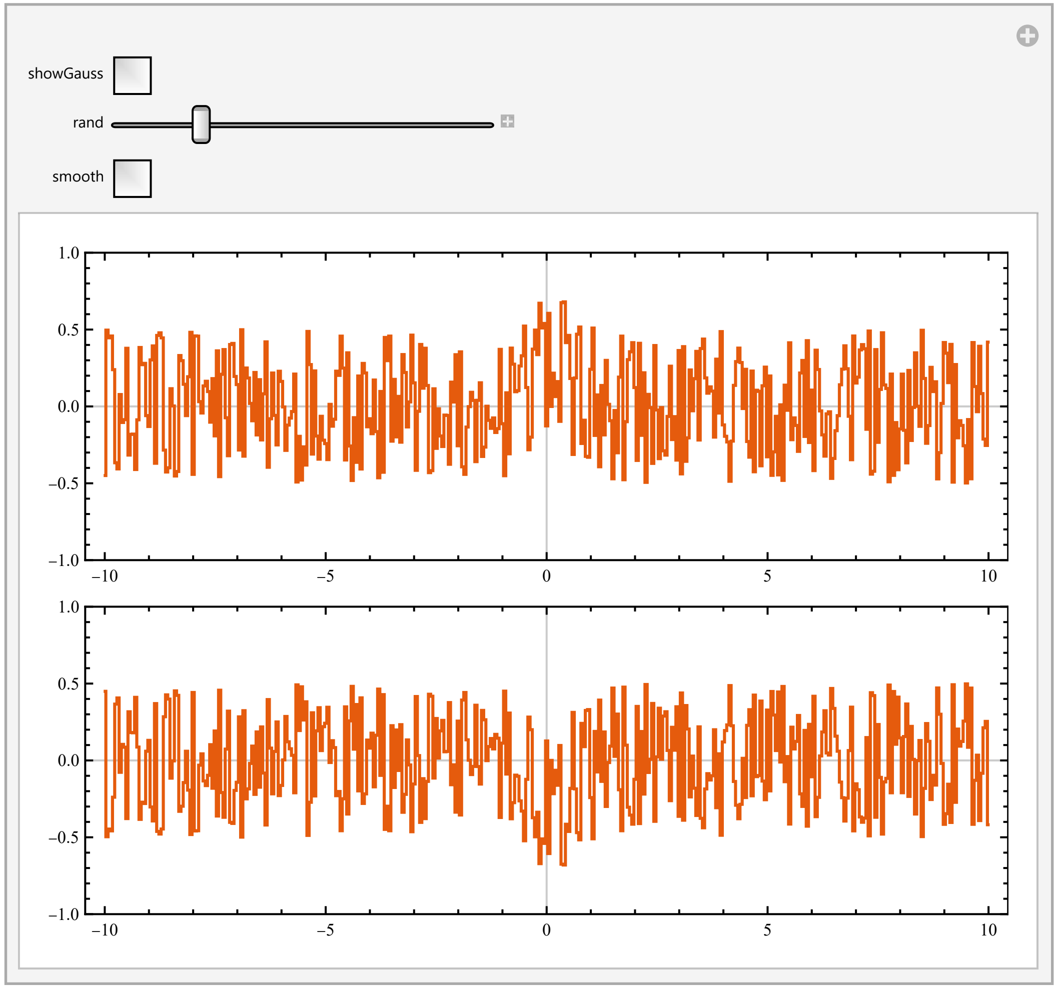
Visual Analytics

- the practice of using visualizations to analyze data.
- In some research, visualizations can support more formal statistical tests by allowing researchers to interact with the data points directly without aggregating or summarizing them.
- Even simple scatter plots, when the variables are chosen carefully, can show outliers, dense regions, bi-modalities, etc.
- In fields where the data themselves are visual (e.g., medical fields, astronomy), visual analytics may actually be the primary means of analyzing data.
- The process of analyzing data through visualization is itself studied by researchers in the visual analytics field.

Visual Analytics - Astronomy Examples

Detect spectral lines in noisy measurements. Visual trick: watch the spectrum upside-down

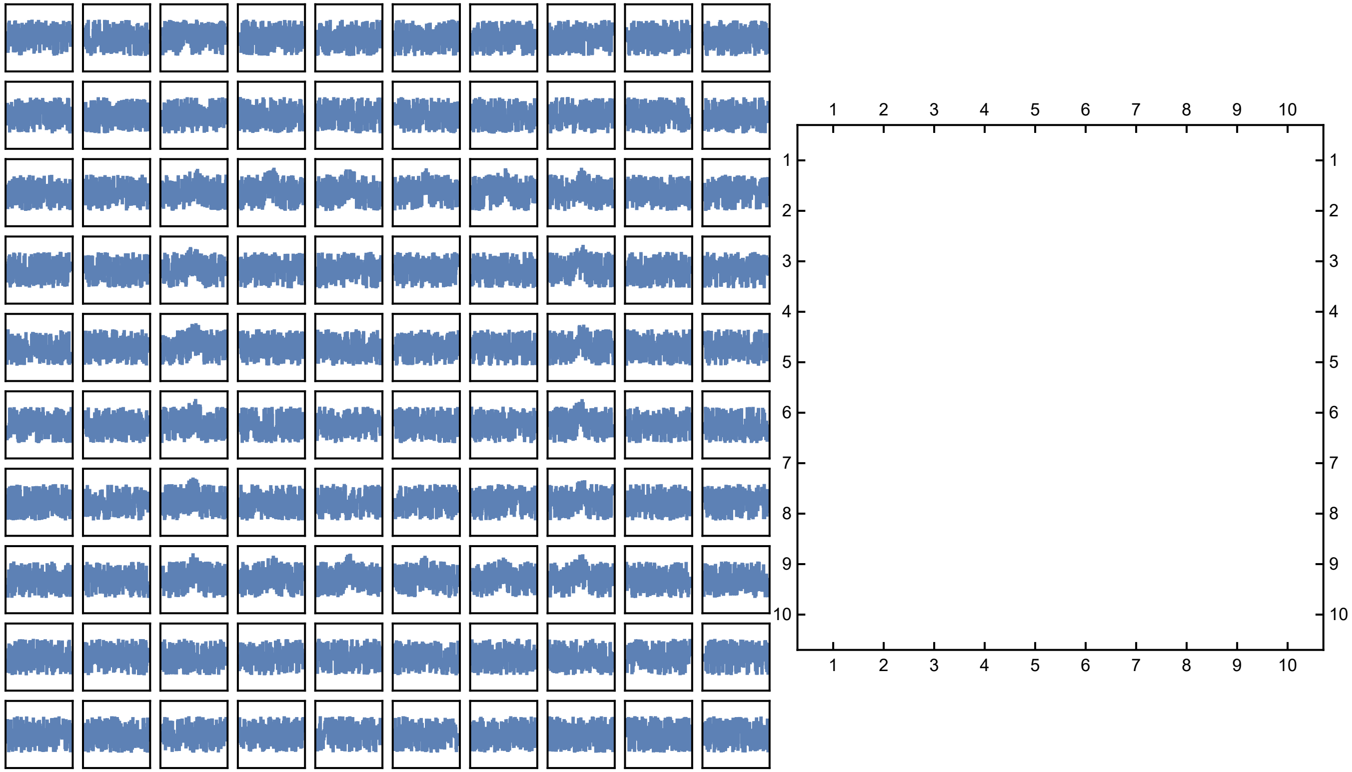
Out[ ]=



Visual Analytics - Astronomy Examples

Detect spectral lines in noisy measurements. Look for spatial correlations! Computing the maps shows non-random features.

Out[ ]=



Visual Analytics

“People use visual analytics tools and techniques to synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data; detect the expected and discover the unexpected; provide timely, defensible, and understandable assessments; and communicate assessment effectively for action. " (From: Thomas and Cook 2005, 2006)

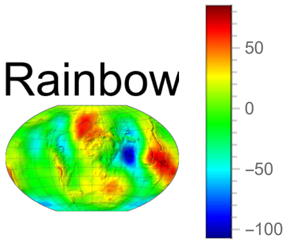
Visual analytics includes the following focus areas:

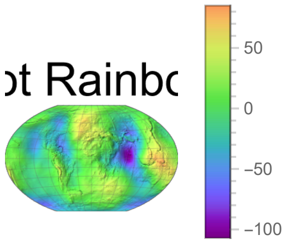
- analytical reasoning techniques that let users obtain deep insights that directly support assessment, planning, and decision making
- visual representations and interaction techniques that exploit the human eye's broad bandwidth pathway into the mind to let users see, explore, and understand large amounts of information simultaneously;
- data representations and transformations that convert all types of conflicting and dynamic data in ways that support visualization and analysis;
- techniques to support production, presentation, and dissemination of analytical results to communicate information in the appropriate context to a variety of audiences. "

Improve Information visibility in displayed data

NO RAINBOW!

Out[ ]=

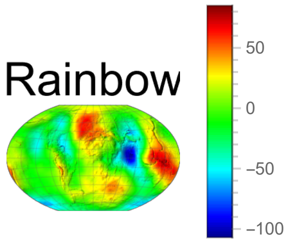


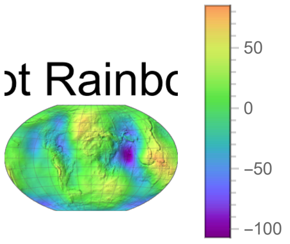


Improve Information visibility in displayed data

NO RAINBOW!

Out[ ]=



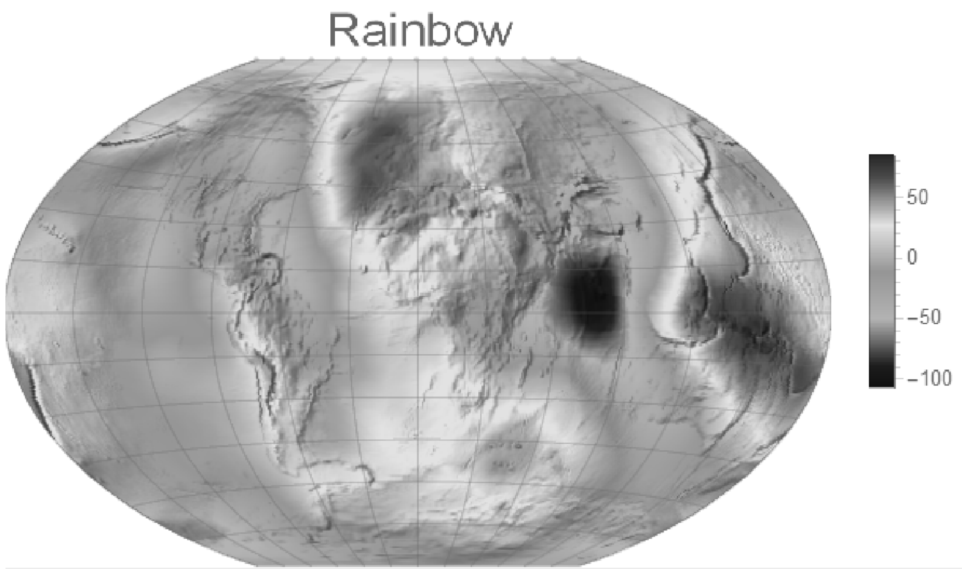


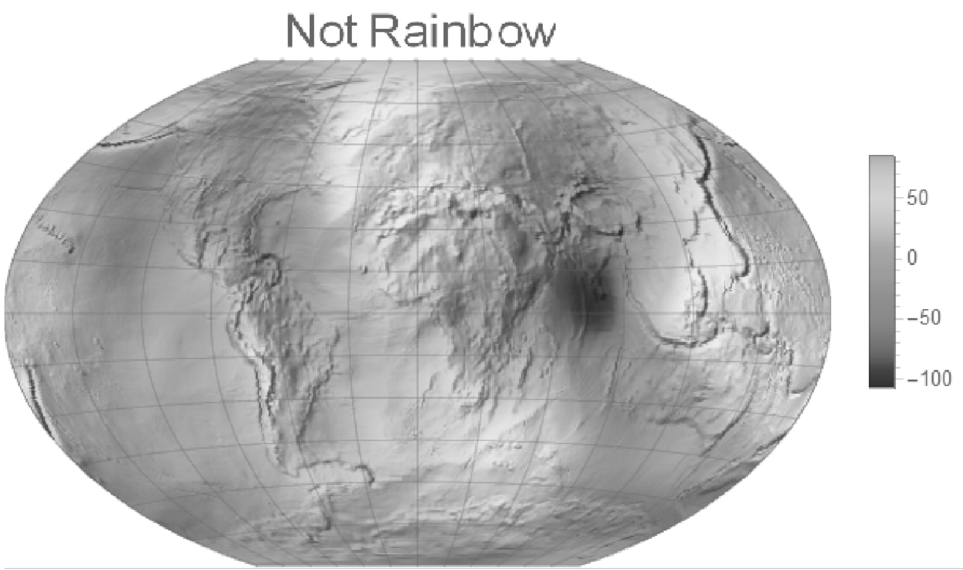
- Most rainbow schemes contain bands of almost constant hue with sharp transitions in-between, which are perceived as jumps in the data.
- In this example large areas are green without features, while for example the yellow line in northern Australia implies a sudden change that does not exist in reality.
- It is more important that a scheme is smooth than that it contains many colours.
- Colour-blind people have difficulty distinguishing some colours of the rainbow. In the right figure, the yellow spot in the middle of the Sahara is not visible in red-blind vision, making the green areas effectively even larger than they are in normal vision.

Improve Information visibility in displayed data

NO RAINBOW!

Out[ ]=

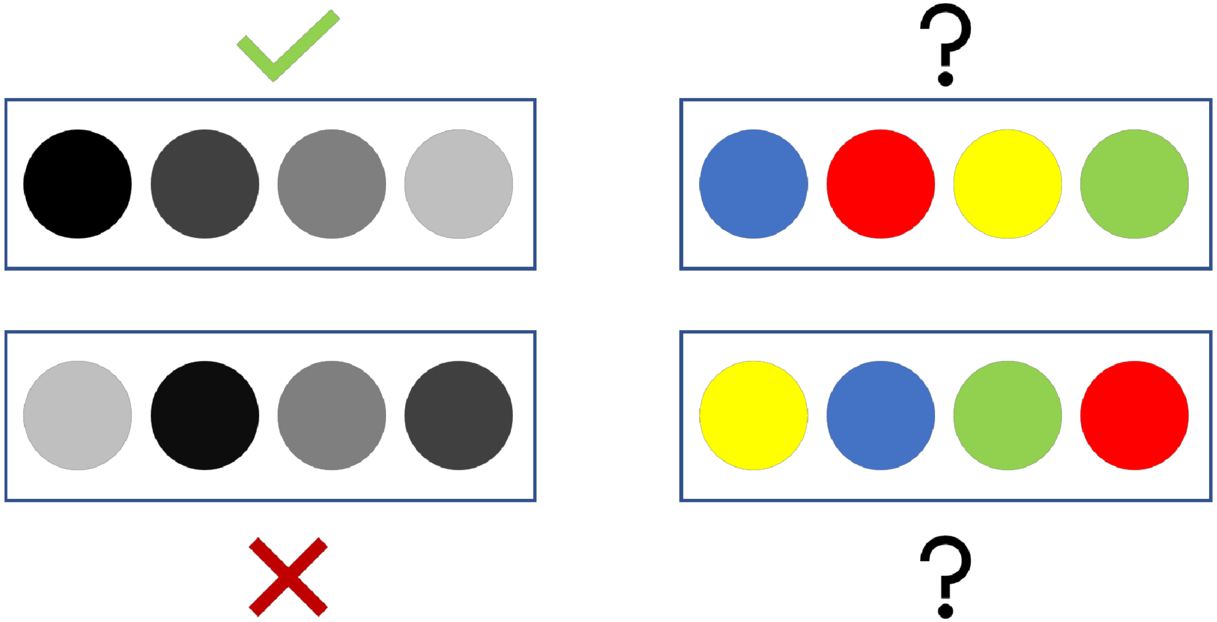




- The sunset scheme emphasizes more clearly the average (light colours) and the extremes (dark colours of contrasting hues).

Color vs. monochrome

Perceptional ordering



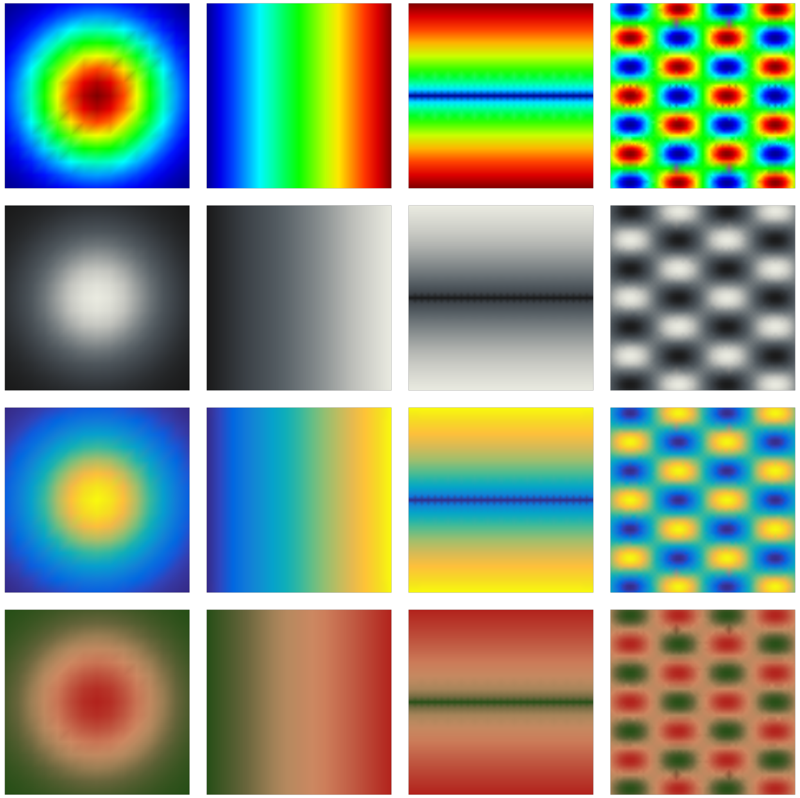
Perceptual ordering. (a) We can easily place the gray paint chips in order based on perception, (b) but cannot do this with the colored chips.

Credits: D. Borland and R. M. Taylor, “Rainbow Color Map (Still) Considered Harmful,” IEEE Computer Graphics and Applications, vol. 27, no. 2, pp. 14-17, Apr. 2007.

Color gradients

False and missed data gradients

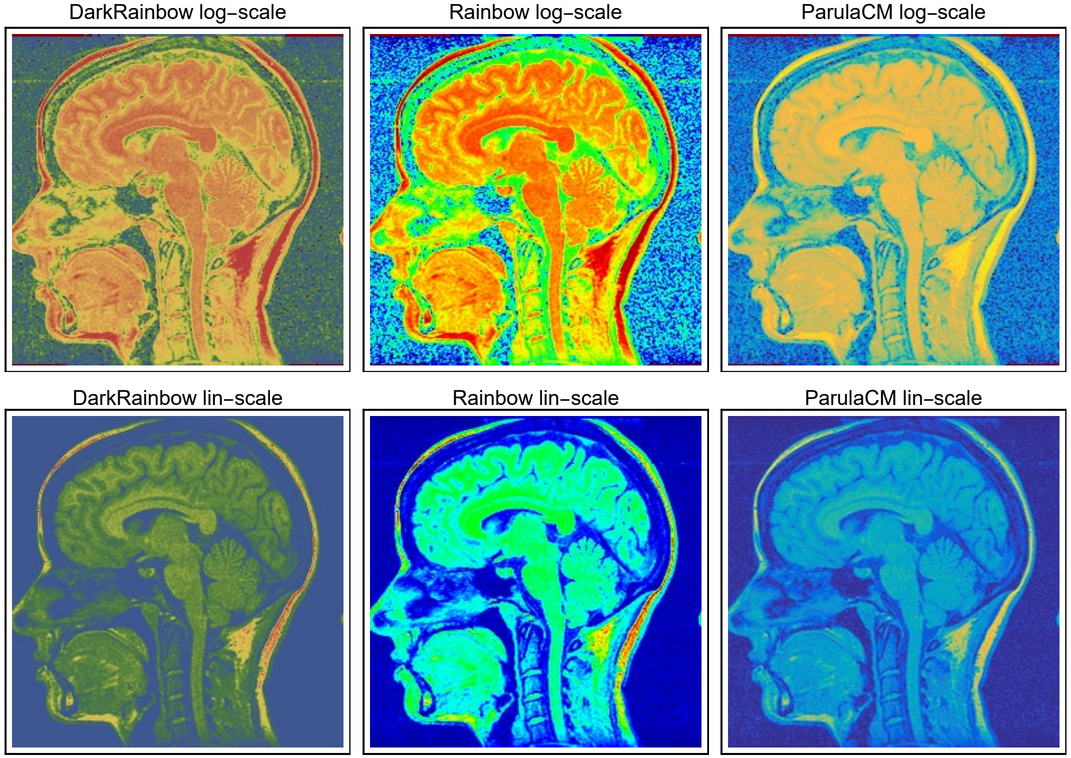
Out[ ]=



- Four data sets visualized with (row 1) rainbow, (row 2) gray-scale, (ro2 3) Parula, and (row 4) an green–red color maps.
- Apparent sharp gradients in the data in (row 1) are revealed as rainbow color map artifacts, not data features.

Color gradients

Out[ ]=



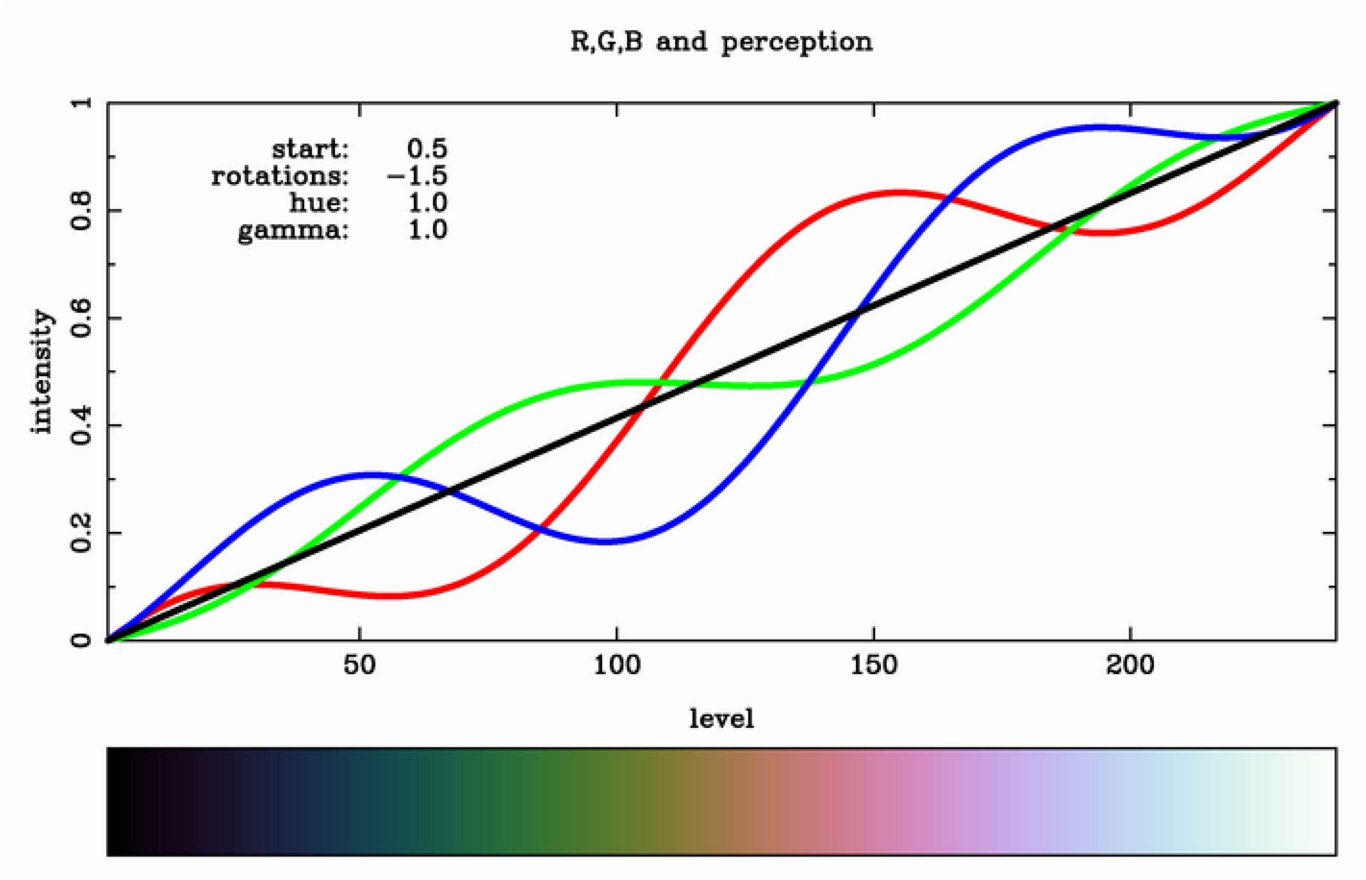
Different colorscales applied to a CT brain scan. Top row is log-scaled, bottom row linear.

Credits:ftp://medical.nema.org/medical/dicom/Multiframe/CT/

Cubehelix color scheme

Webpage: http://www.mrao.cam.ac.uk/~dag/CUBEHELIX/

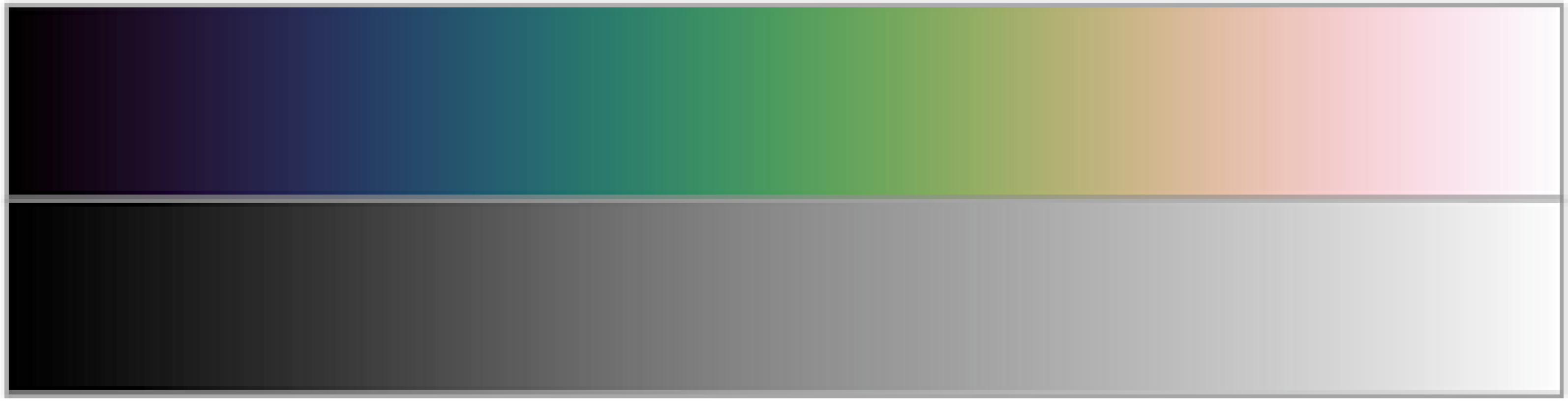
Many colour schemes used to display astronomical intensity images do not have an underlying increase in the perception of the brightness of the colours used (e.g. burning out to red for the high data values, but using yellow/green for intermediate data values, which are perceived as being brighter than the red).

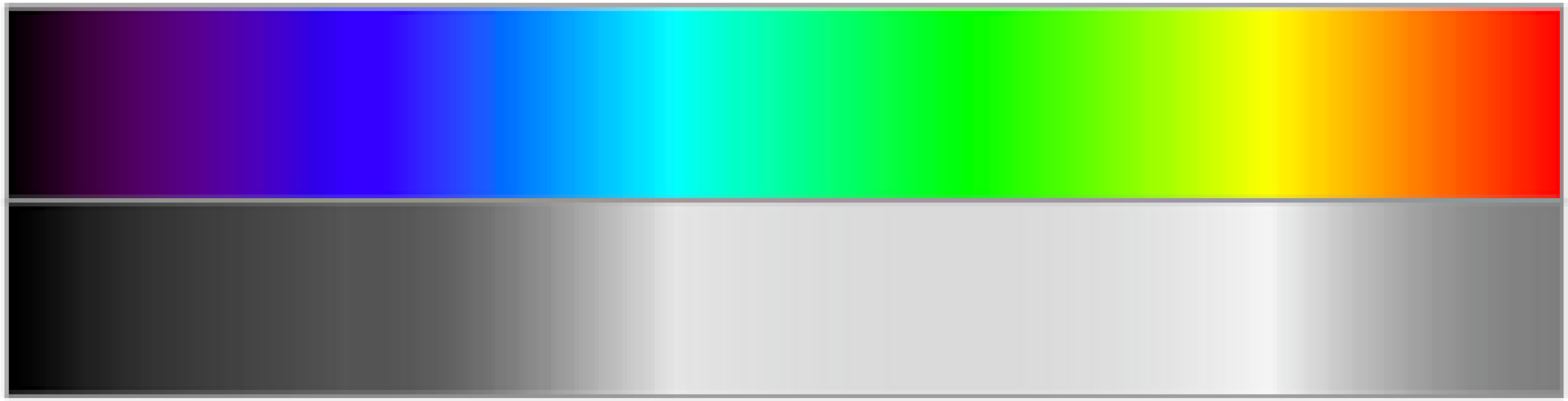


Advantages

(Credits: http://www.ifweassume.com/2013/05/cubehelix-or-how-i-learned-to-love.html)

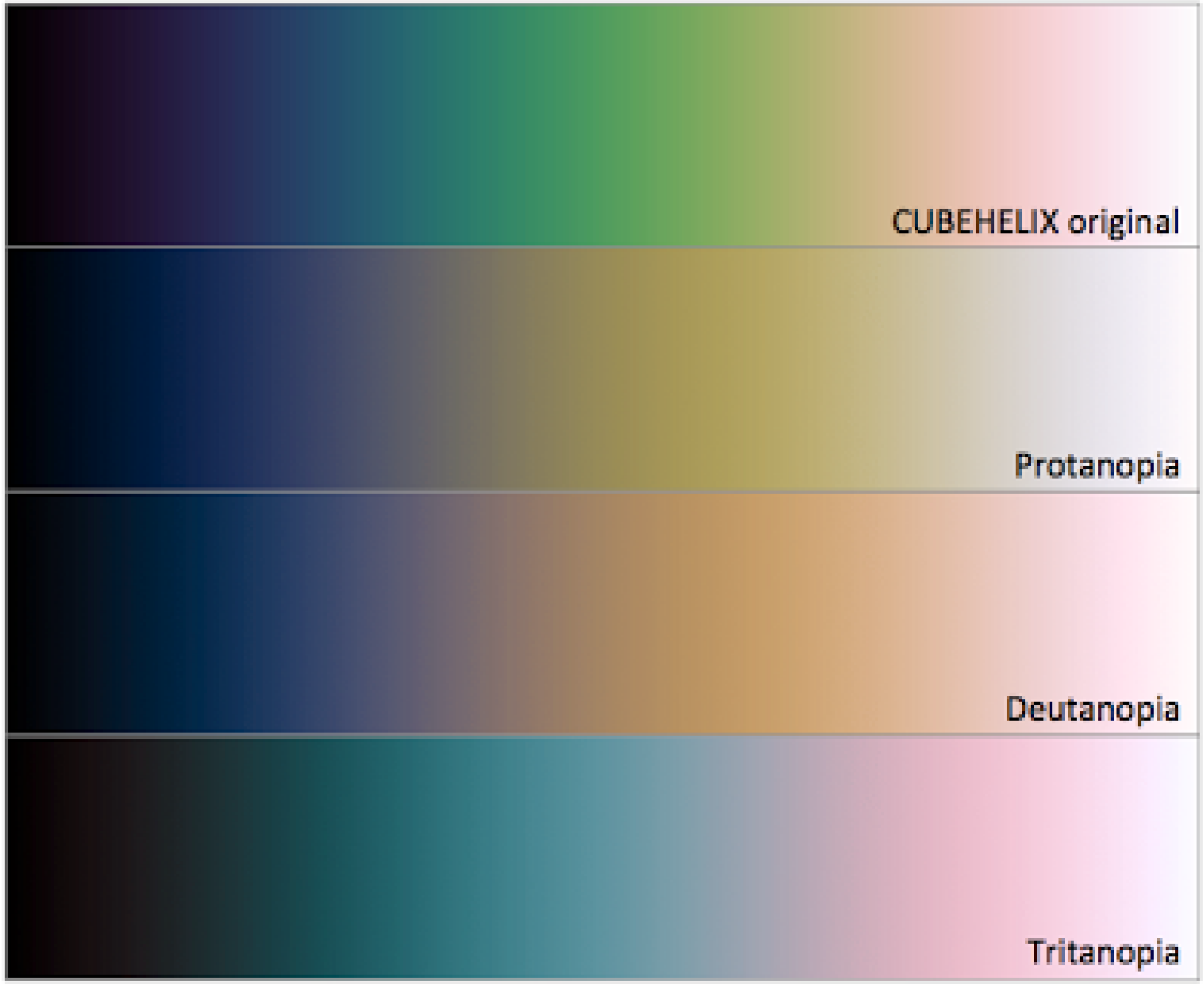
- it prints equally well in color and black & white





Advantages

- CUBEHELIX is pretty friendly for your colorblind readership! Here’s an example of the default CUBEHELIX color scheme, followed by simulations for how it is seen by people with three types of colorblindness



Common Static Visualization Types

1D/Linear

2D/Planar

3D/Volumetric

Temporal

nD/Multidimensional

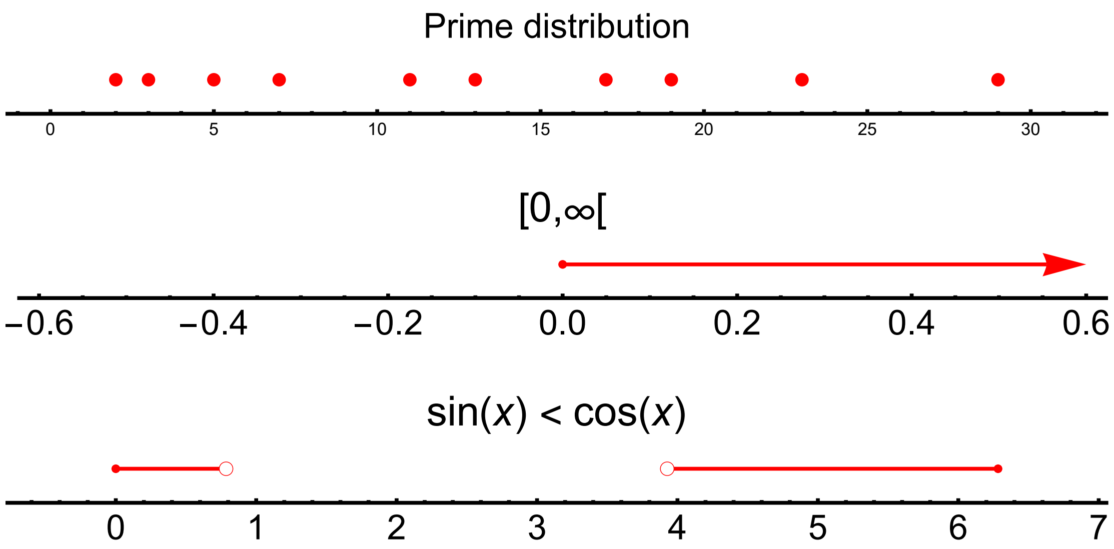
Tree/Hierarchical

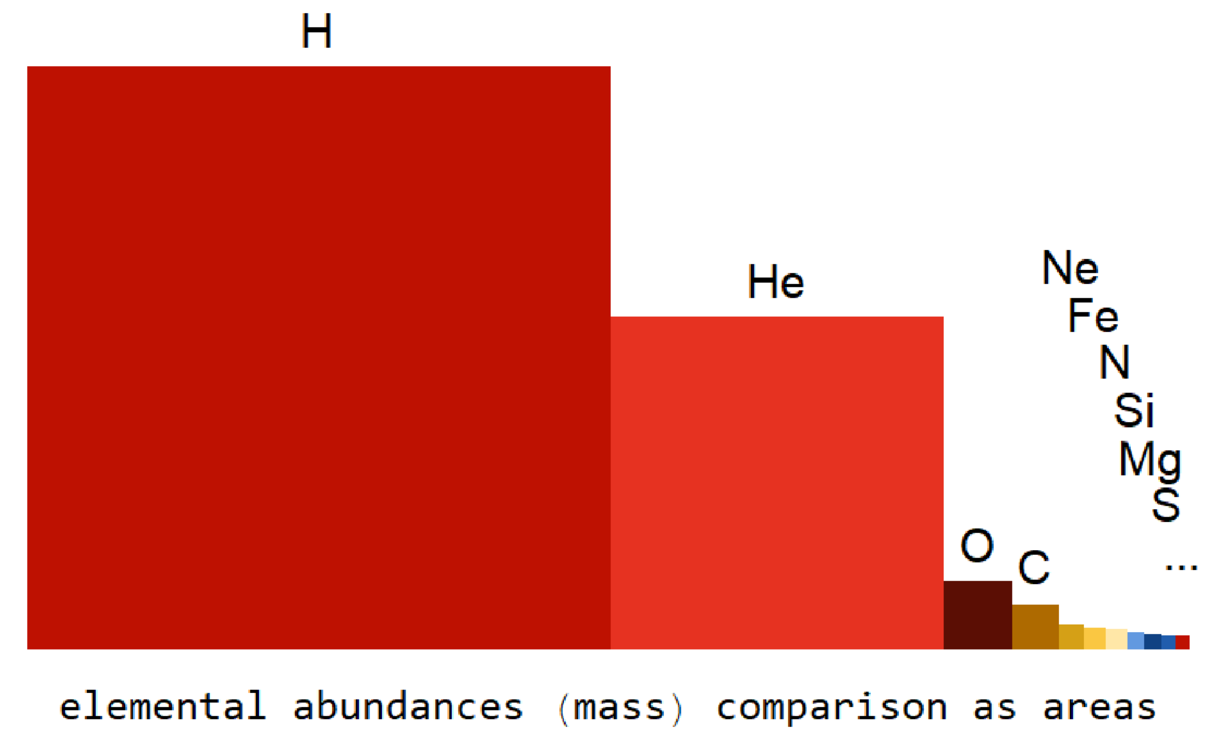
Network

1D/Linear

Examples

Out[•]=



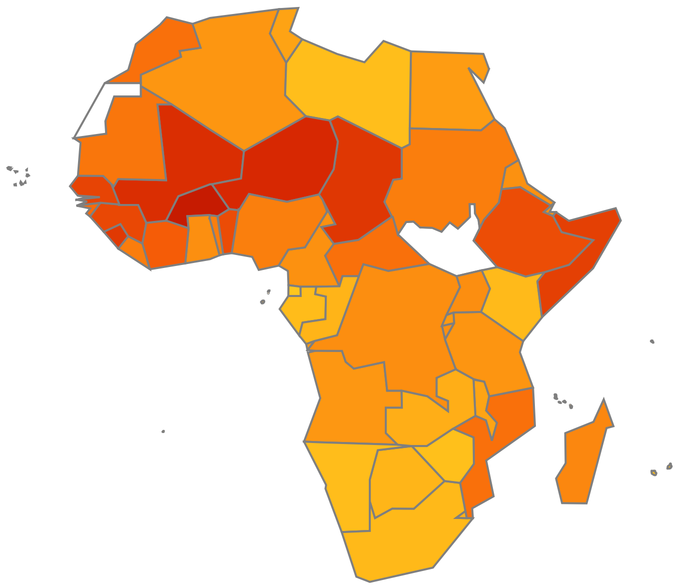

H
He
Ne
Fe
N
Si
Mg
S
O
C
...
elemental abundances (mass) comparison as areas

2D/Linear (geospatial)

Chloropleth map

A choropleth map (from Greek *χώρο* (“area/region”) + *πλήθος* (“multitude”)) is a thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map, such as population density or per-capita income. (Wikipedia)

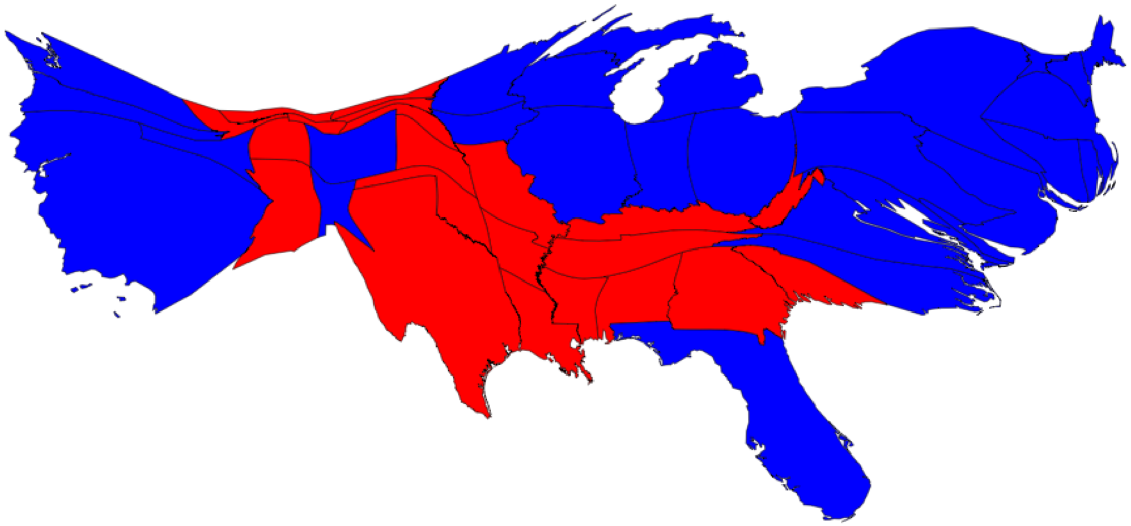
Example: Literacy fractions in Africa



Cartogram

A cartogram is a map in which some thematic mapping variable – such as travel time, population, or Gross National Product – is substituted for land area or distance. The geometry or space of the map is distorted in order to convey the information of this alternate variable.

Example: US state size proportional to population



Contour map (actually 3D data)

Contour lines of functions of 2 variables are curves along which the function has a constant value. In maps, contour lines join points of equal map value.

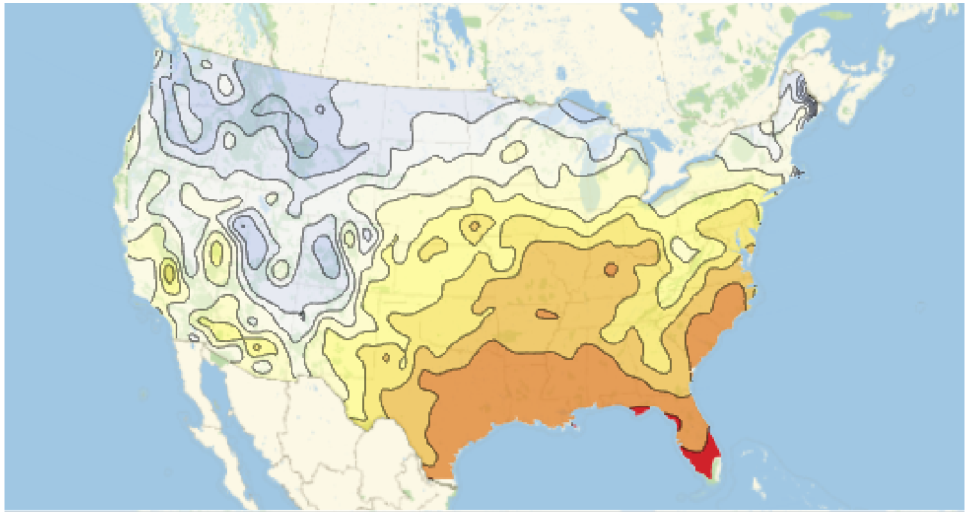
Contour lines are often given names starting with iso- (from greek “equal”). Some examples are :

isobar is a line of equal or constant pressure
isotherm is a line of constant temperature
isochore is a line of constant volume
isophote line of constant illuminance, i.e. irradiating flux.

Contour map (actually 3D data)

Example: temperature at sea level

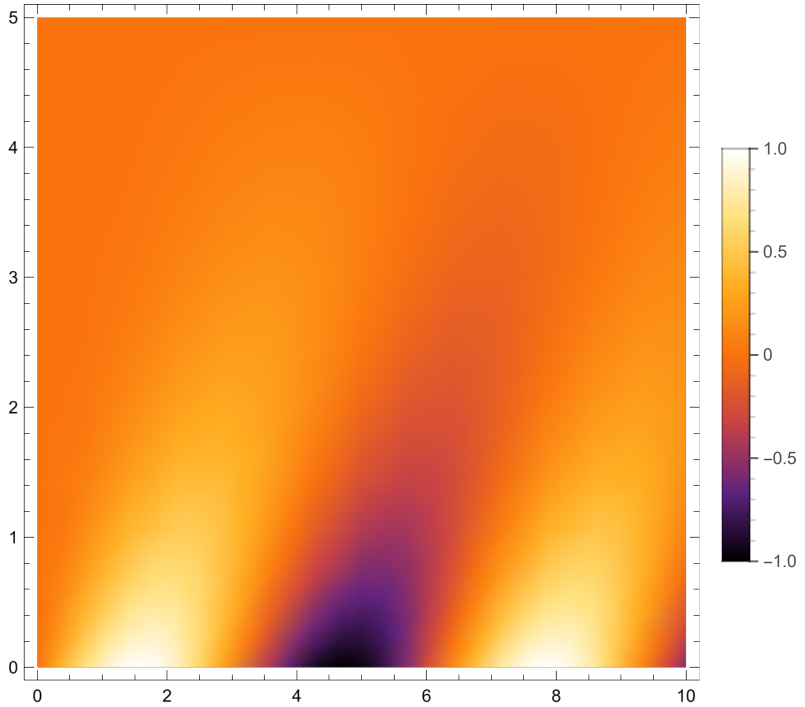
Out[ ]=



Density plot

A density plot visualizes a function $f(x, y)$ or a list of {x,y,z} data in terms of colors/gray scales.

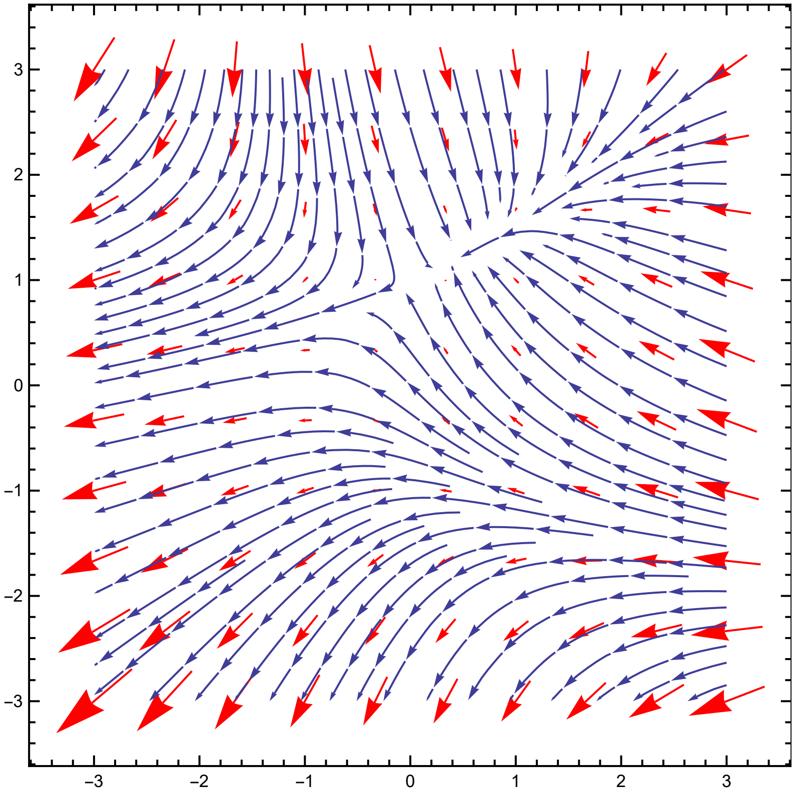
Example: solution to the heat equation in one dimension

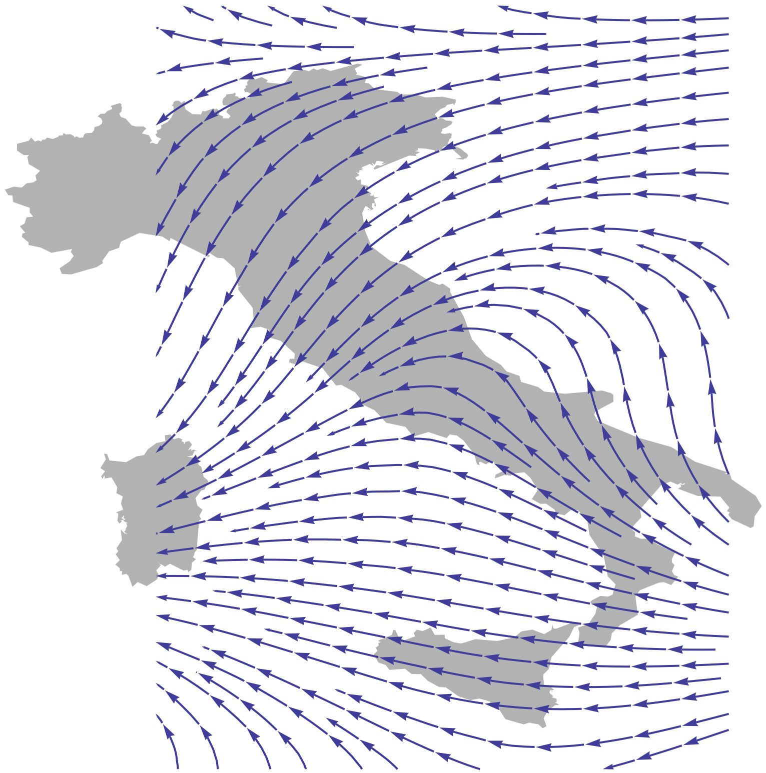


Stream plots (actually at least 4D)

Stream plots visualize a vector field $\{v_x, v_y\}$ as function of {x,y}.

Example:

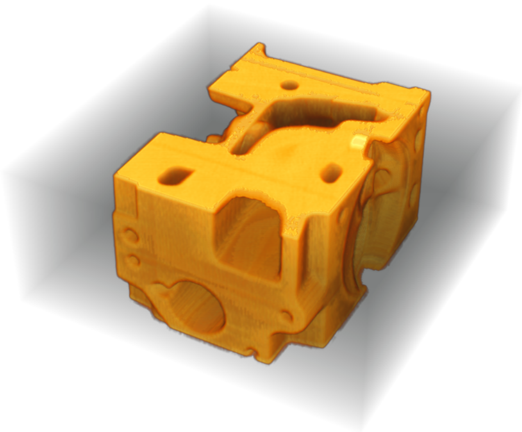


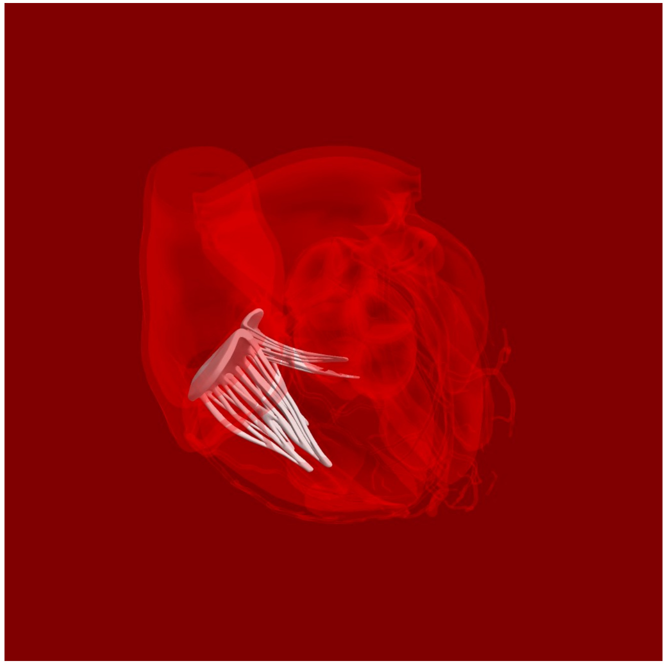


3D visualization

Volumetric/volume rendering

Out[•]=



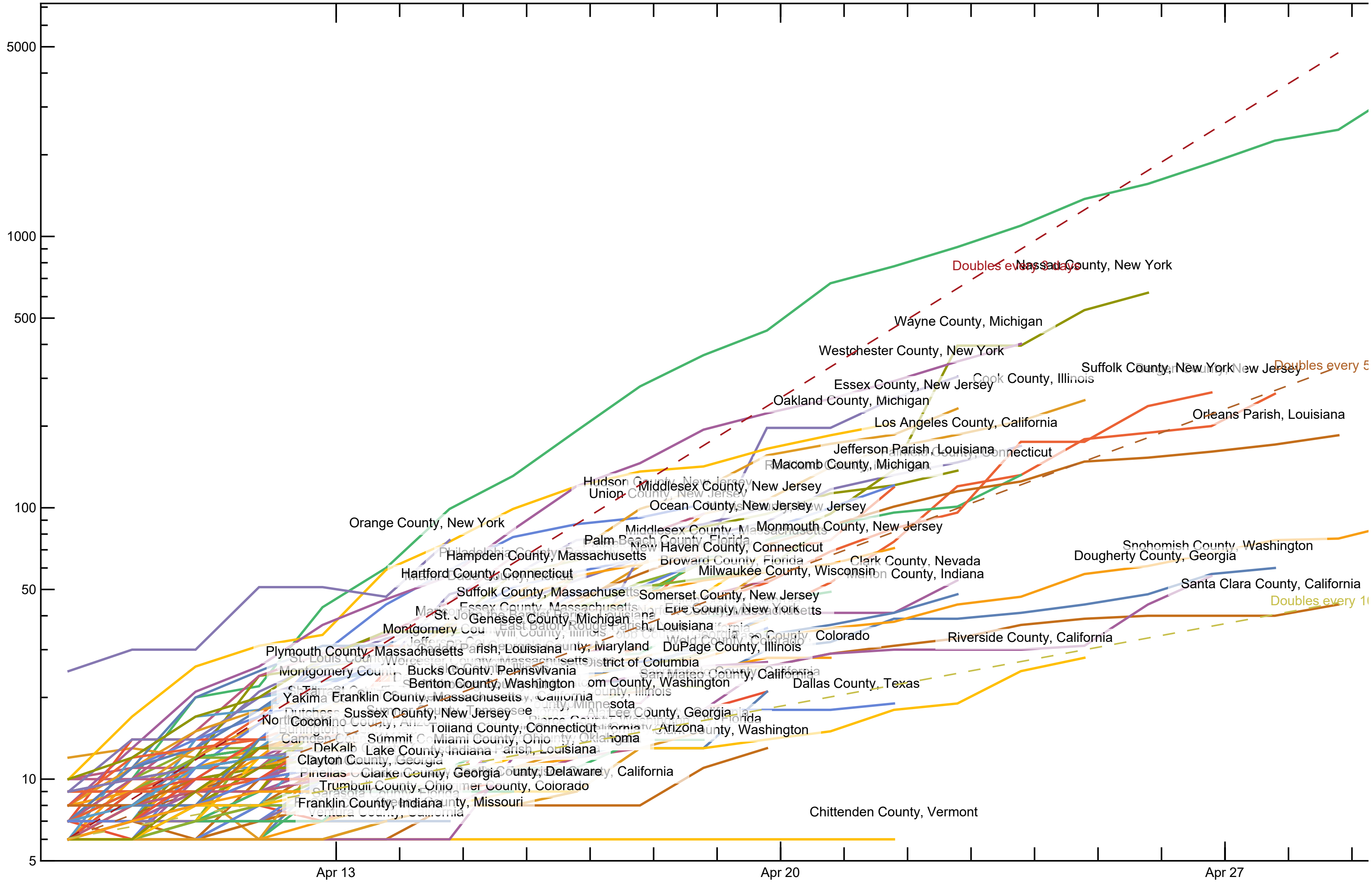


Temporal data

Time series

A time series is a sequence of data points, measured typically at successive points in time spaced at uniform time intervals

Out[•]=

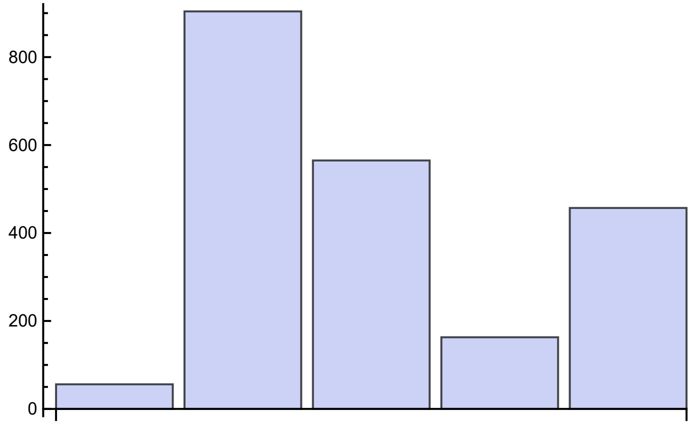

5000
1000
500
100
50
10
5
Apr 13
Apr 20
Apr 27
Doubles every 3 days
Nassau County, New York
Doubles every 5
Doubles every 10
Wayne County, Michigan
Westchester County, New York
Suffolk County, New York
Essex County, New Jersey
Cook County, Illinois
Oakland County, Michigan
Los Angeles County, California
Orleans Parish, Louisiana
Jefferson Parish, Louisiana
Macomb County, Michigan
Middlesex County, New Jersey
Ocean County, New Jersey
Monmouth County, New Jersey
Orange County, New York
New Haven County, Connecticut
Hampden County, Massachusetts
Clark County, Nevada
Milwaukee County, Wisconsin
Hartford County, Connecticut
Snohomish County, Washington
Dougherty County, Georgia
Santa Clara County, California
Suffolk County, Massachusetts
Somerset County, New Jersey
Genesee County, Michigan
Riverside County, California
DuPage County, Illinois
Plymouth County, Massachusetts
Bucks County, Pennsylvania
Benton County, Washington
Dallas County, Texas
Franklin County, Massachusetts
Sussex County, New Jersey
Lee County, Georgia
Tolland County, Connecticut
Miami County, Ohio
Lake County, Indiana
Clayton County, Georgia
Clarke County, Georgia
Trumbull County, Ohio
Franklin County, Indiana
Chittenden County, Vermont

nD/Multidimensional Plots

Category proportions, counts

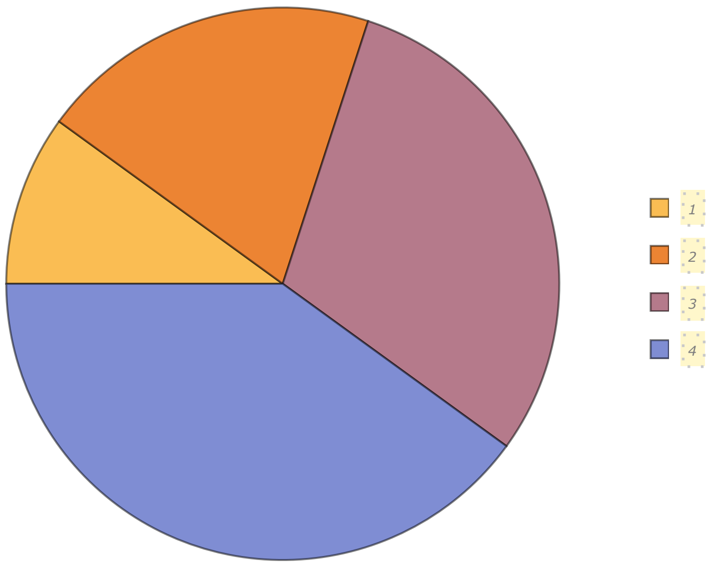
Bar Chart

A bar chart or bar graph is a chart with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar chart is sometimes called a column bar chart.



Pie Chart

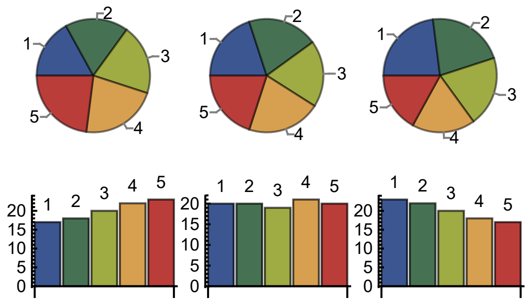
A pie chart is a circular chart divided into sectors, illustrating numerical proportion. In a pie chart, the arc length of each sector (and consequently its central angle and area), is proportional to the quantity it represents.



Pie Chart

Pie charts are a bad tool to convey numerical differences of data. Comparison by angle is less efficient than by length

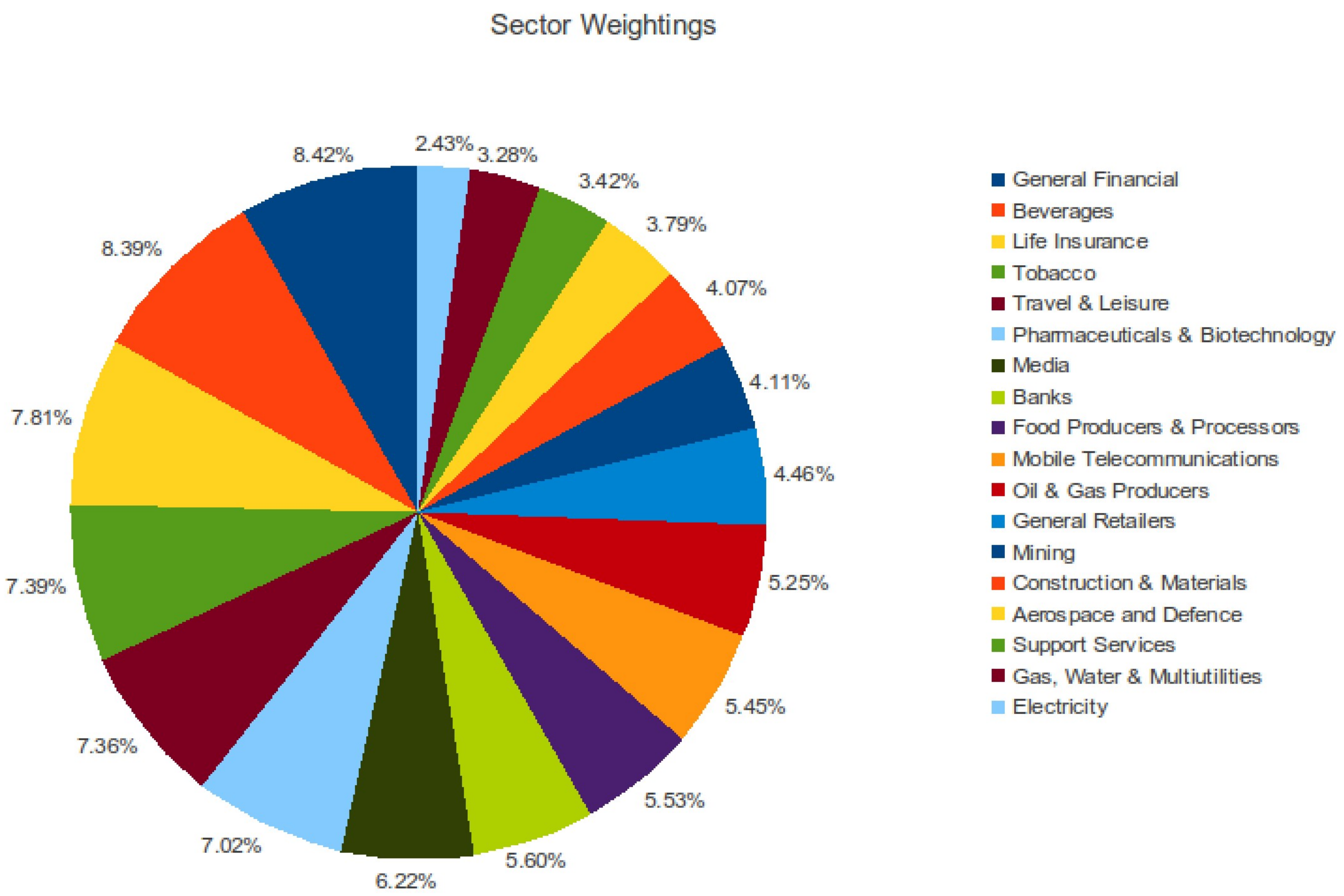
Out[ ]=



Pie Chart

Bad Example:

Out[ ]=



Pie Chart

Bad Example:

Histograms

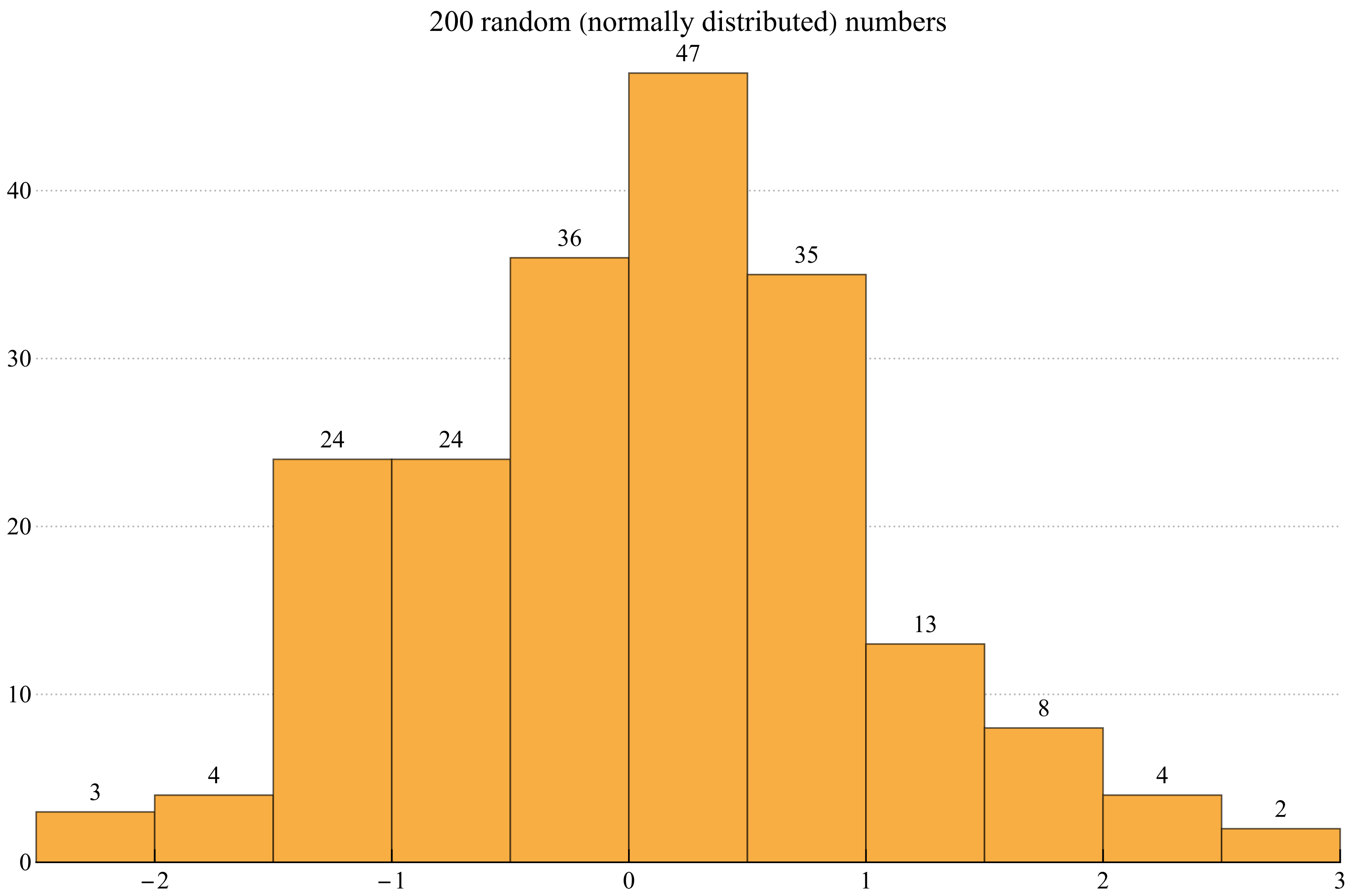
- A histogram is a graphical representation of the distribution of data. It is an estimate of the probability distribution of a continuous variable.
- A histogram is a representation of tabulated frequencies, shown as adjacent rectangles, erected over discrete intervals (bins), with an area proportional to the frequency of the observations in the interval.
- The height of a rectangle is also equal to the frequency density of the interval, i.e., the frequency divided by the width of the interval. The total area of the histogram is equal to the number of data.

A histogram is a function m_i that counts the number of observations that fall into each of the disjoint categories (known as *bins*). Thus, if we let n be the total number of observations and k be the total number of bins, the histogram m_i meets the following conditions:

$$\boxed{n = \sum_{i=1}^{k} m_i}$$

Histograms

Out[ ]=

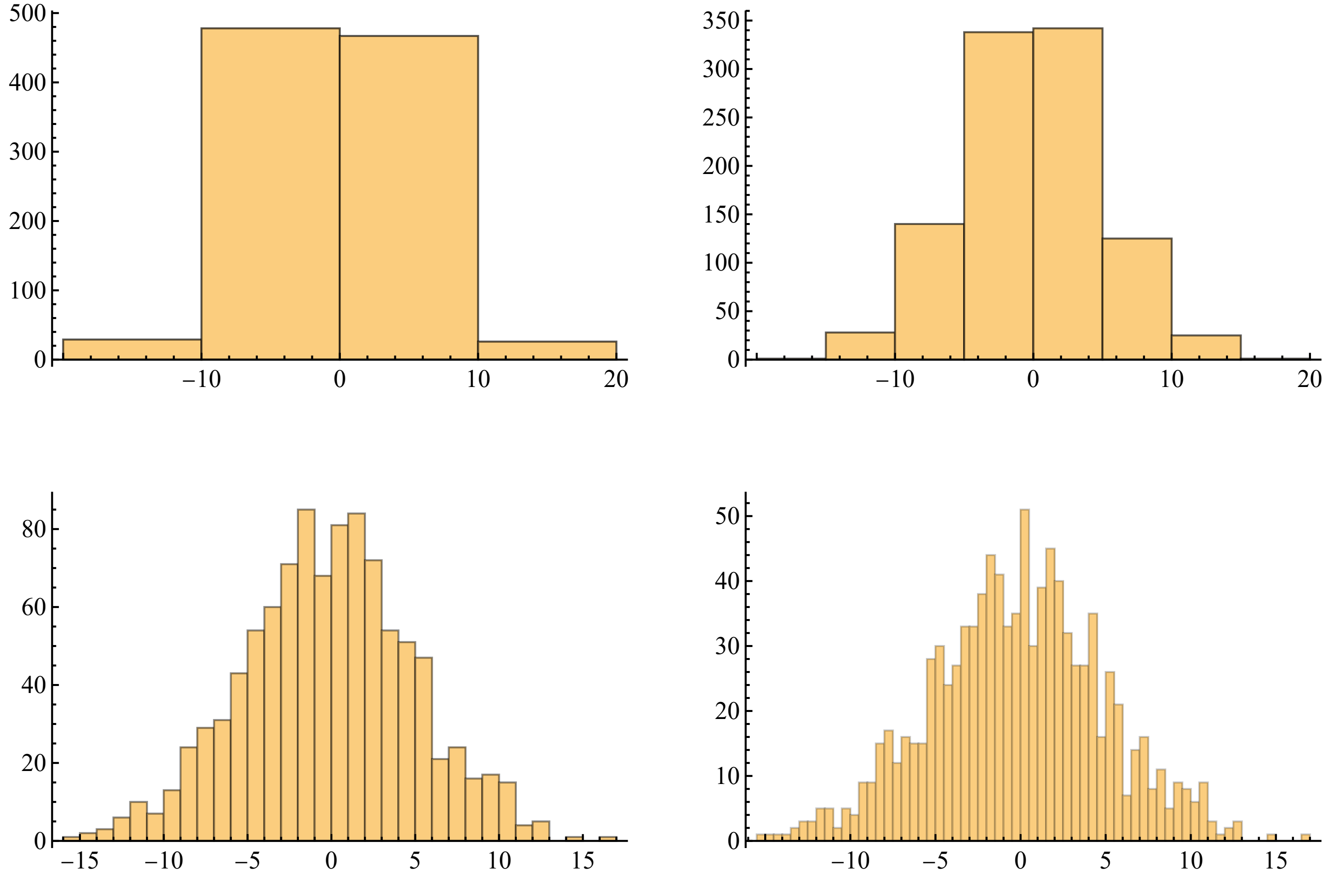


Histograms

Number of bins and width

There is no optimal number of bins. Different bin sizes can reveal different features of the data

Out[•]=

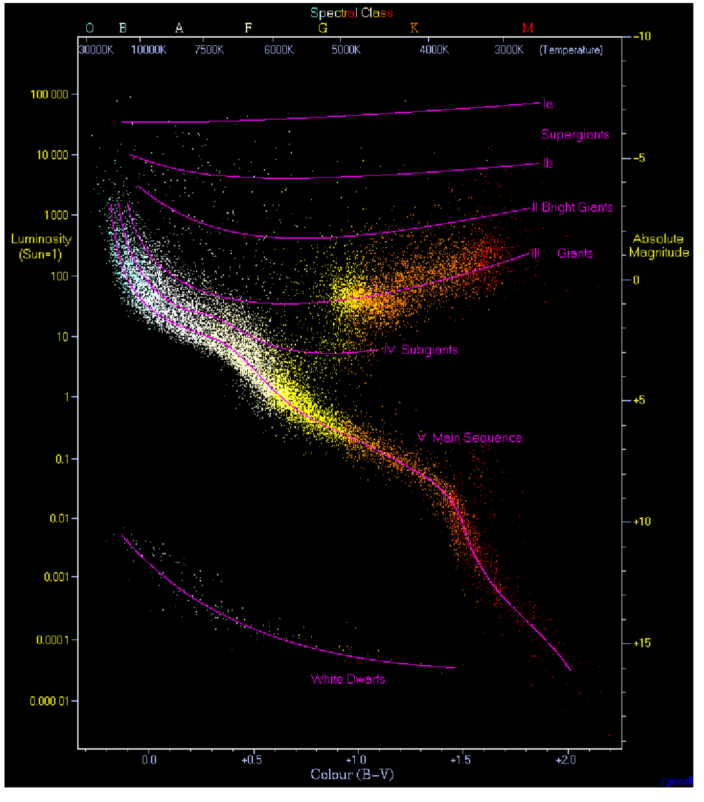


Relationships between variables

Scatter plot

A scatter plot, scatterplot, or scatter-graph is a type of mathematical diagram using Cartesian coordinates to display values for two variables for a set of data.

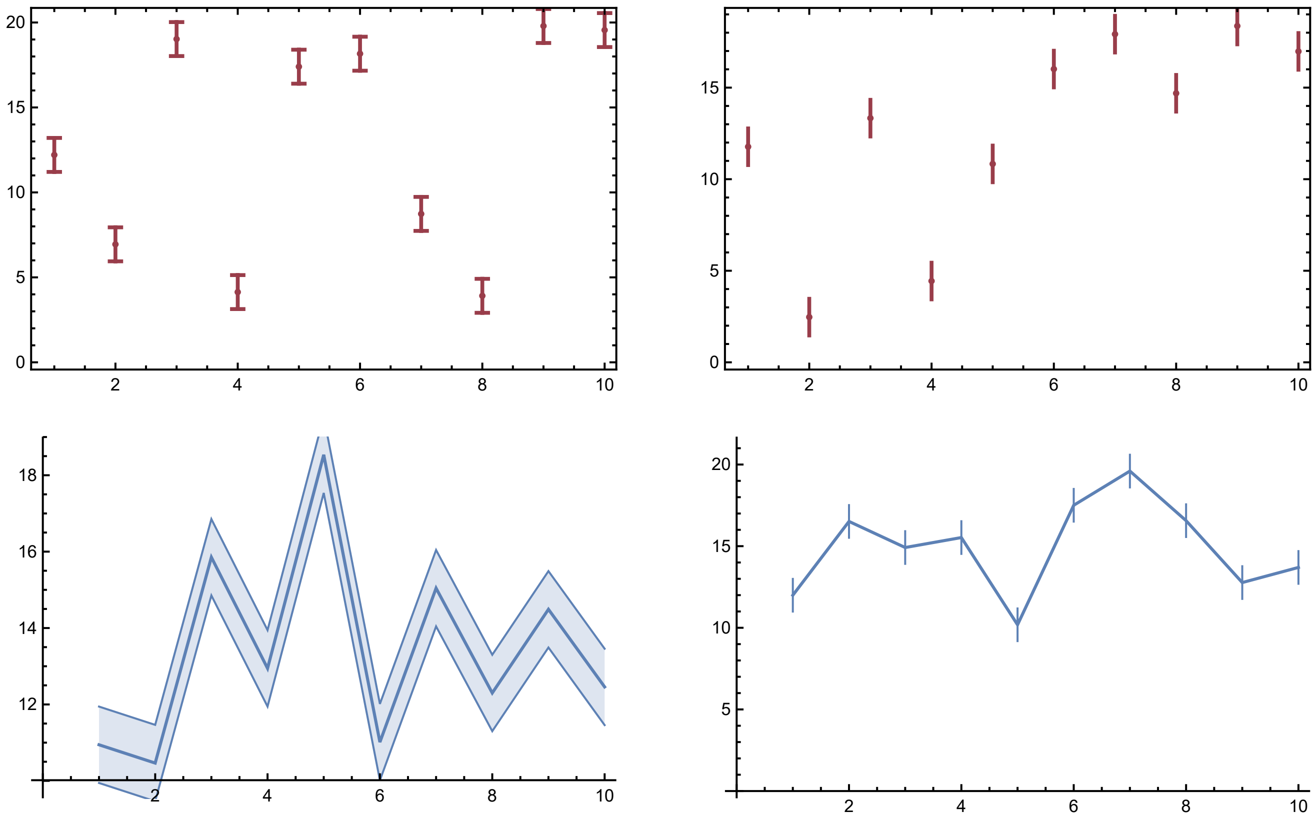
Example: Hertzsprung-Russel diagram



Scatter plot

Example: error bar plot

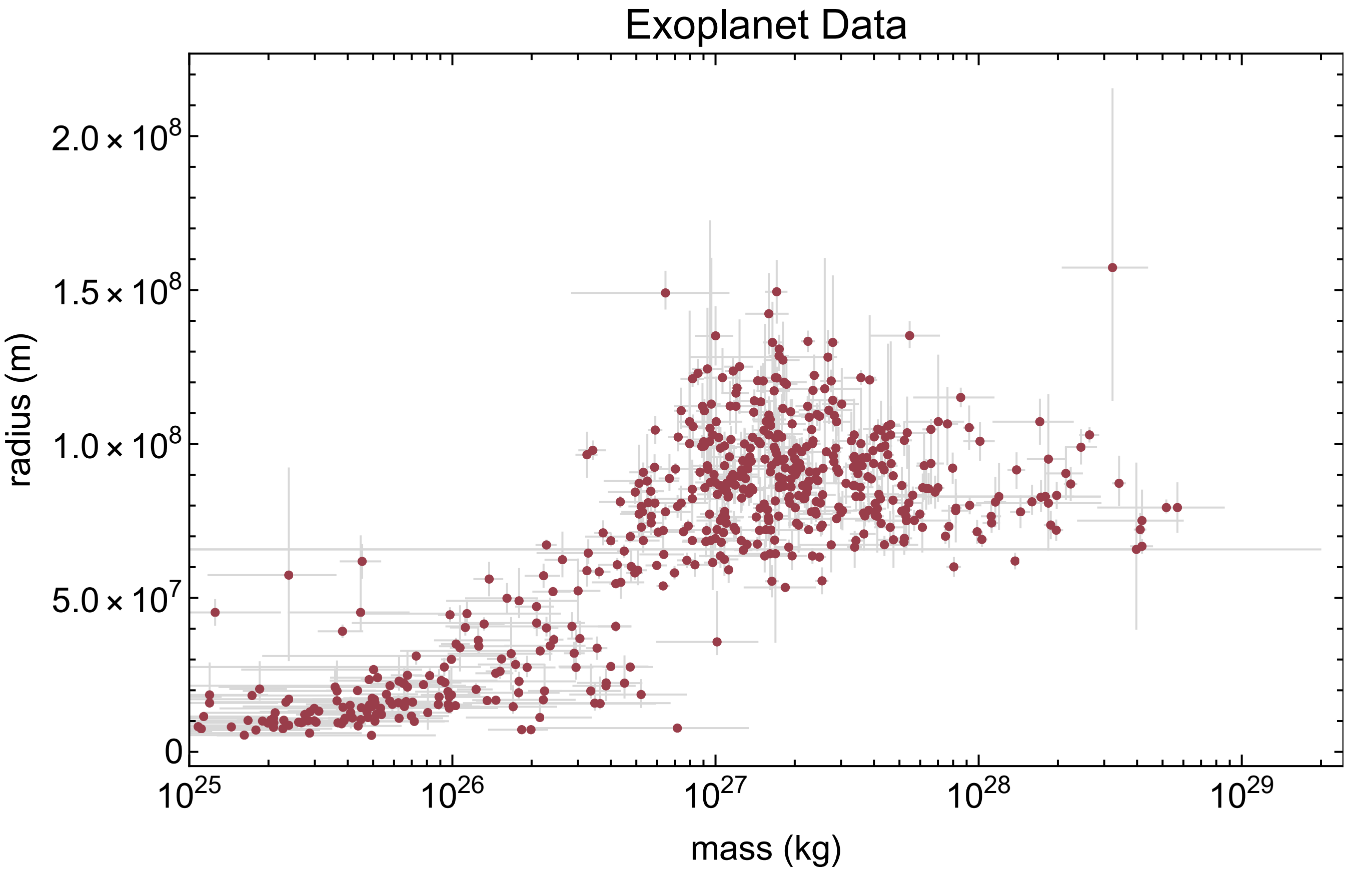
Out[ ]=



Scatter plot

When the data spreads across many orders of magnitude, a Log- or LogLog plot is used.

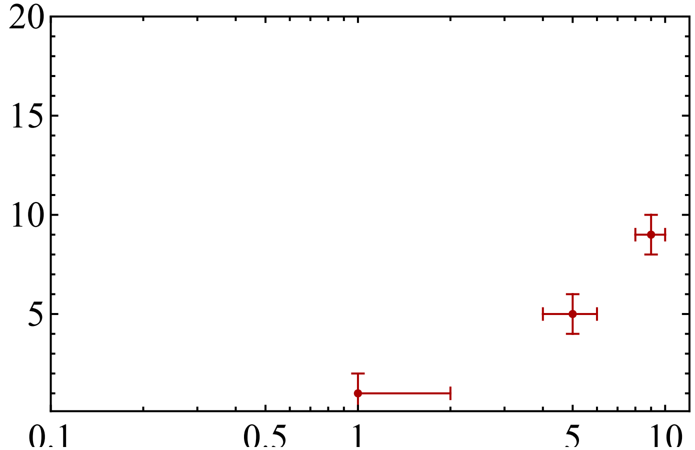
Out[ ]=

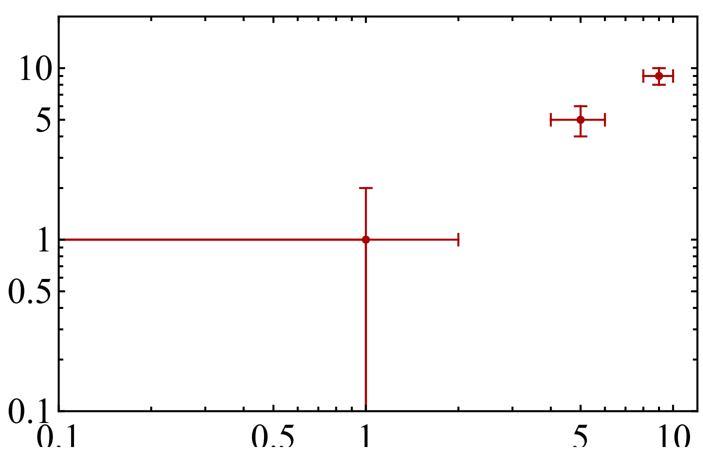


Scatter plot

When plotting error bars in logarithmic plots, remember that equally large absolute errors, change error bar length along the logarithmic axis.

Out[ ]=



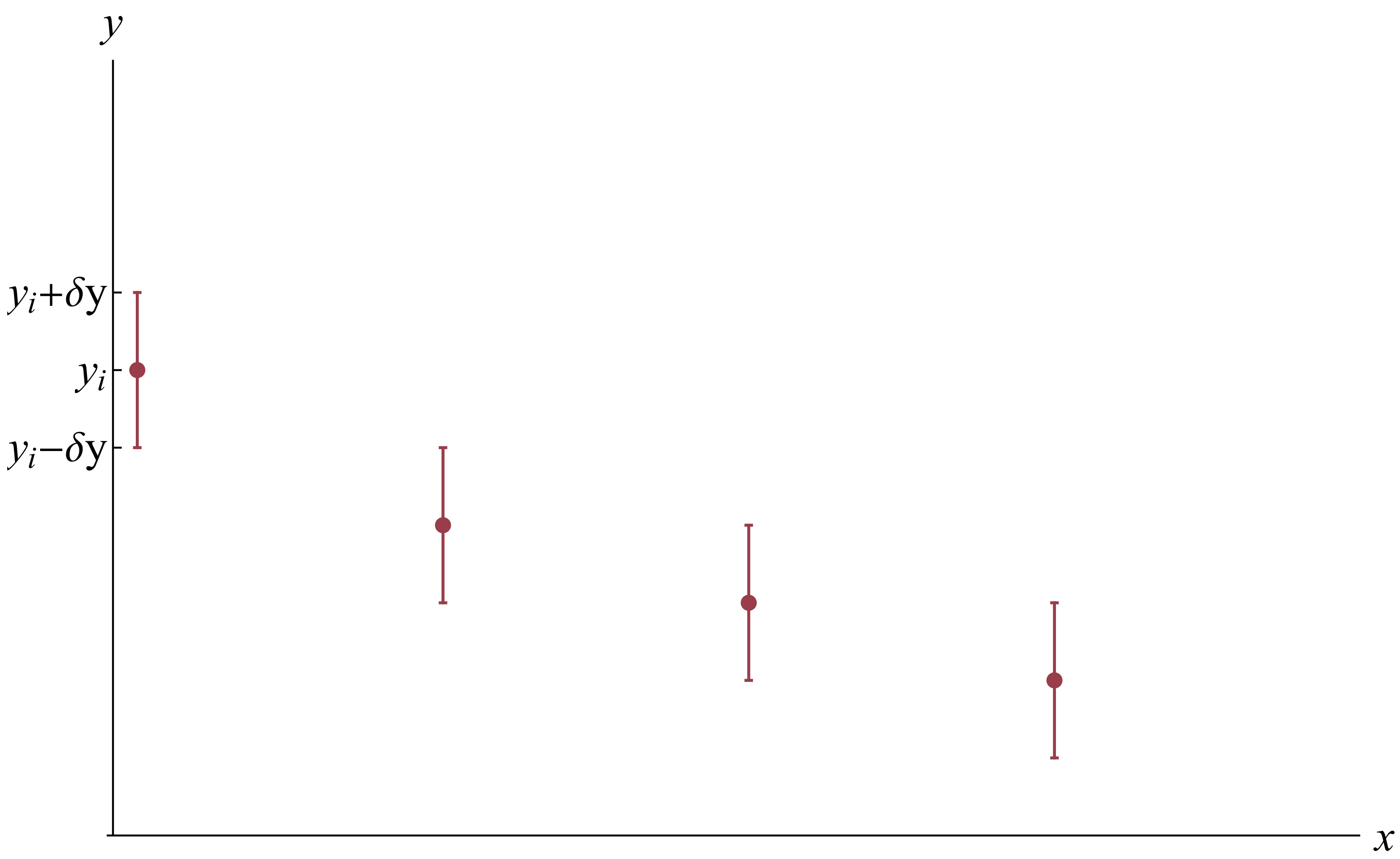


What is a better way to draw error bars in log-plots?

Absolute Error bars

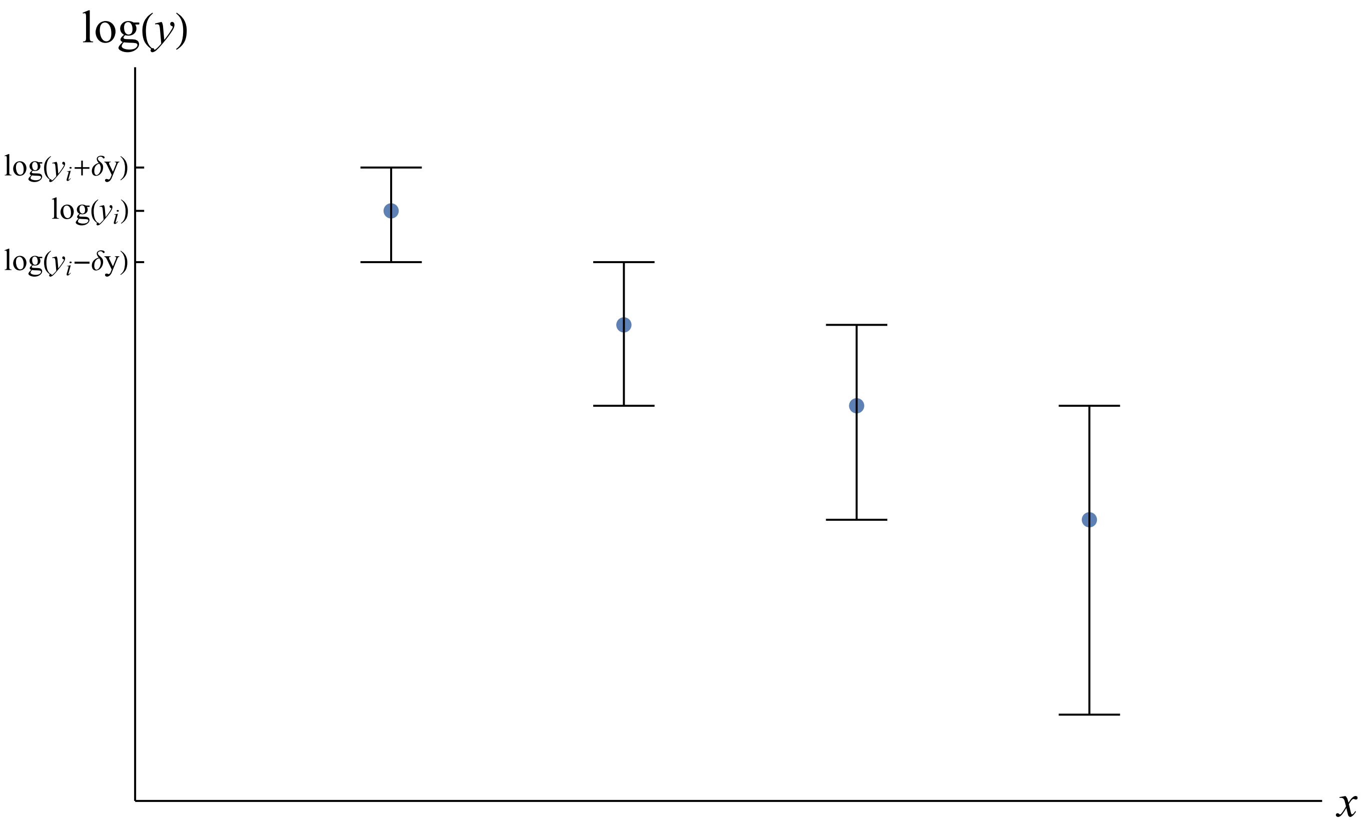
Suppose that one has a sufficient number of measurements to make an estimate of a measured quantity y and report its absolute error $\pm\delta y$.

The absolute error $\pm\delta y$ is represented on a Cartesian plot by extending lines of the appropriate size above and below the point y.



Absolute Error Bars

If plotted on a logarithmic plot, however, absolute error bars that are symmetric on a *y vs. x* plot become asymmetric; the lower portion is longer than the upper portion.



Here the absolute error is the same for all data points.

This gives a misleading view of measurement precision, especially when measured quantities vary by several orders of magnitude.

Error in Logarithmic Quantities

To represent error bars correctly on a log plot, one must recognize that the quantity being plotted, which we call z, is different than the measured quantity y.

Out[ ]//TraditionalForm=

$$z = \log(y)$$

The error δz is

Out[ ]//TraditionalForm=

$$\delta z = \delta(\log(y))$$

On the assumption of small errors, a differential analysis can be used

Out[ ]//TraditionalForm=

$$\delta z = dz = d[\log(y)] = \frac{1}{2.303}\,\frac{dy}{y} = \frac{0.434\,\delta y}{y}$$

Remember that $\log_a b = \frac{\log_c b}{\log_c a}$ and $\frac{d}{dx}\ln(x) = \frac{1}{x}$, then

Out[ ]//TraditionalForm=

$$\frac{d}{dy}\log(y) = \frac{d}{dy}\left(\frac{\ln(y)}{\ln(10)}\right) = \frac{1}{\ln(10)}\,\frac{1}{y} = \frac{1}{2.30259}\,\frac{1}{y} \Longrightarrow d\log(y) = \frac{1}{2.30259}\,\frac{dy}{y} = 0.434\,\frac{dy}{y}$$

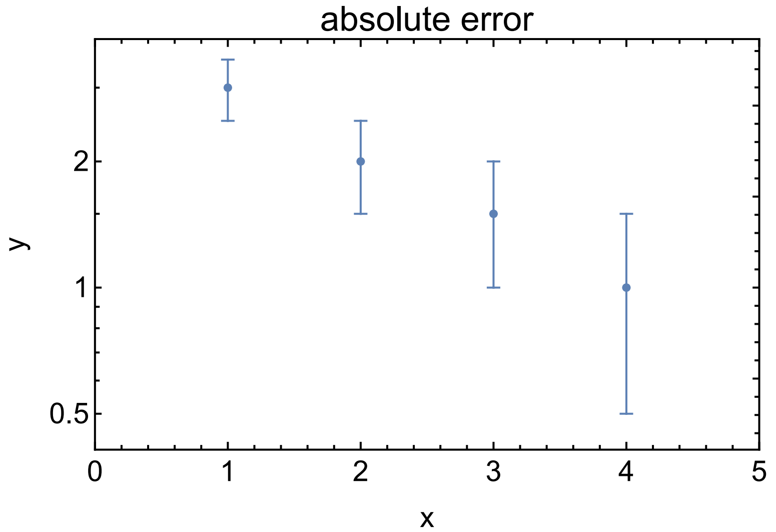
Log Error is Relative Error

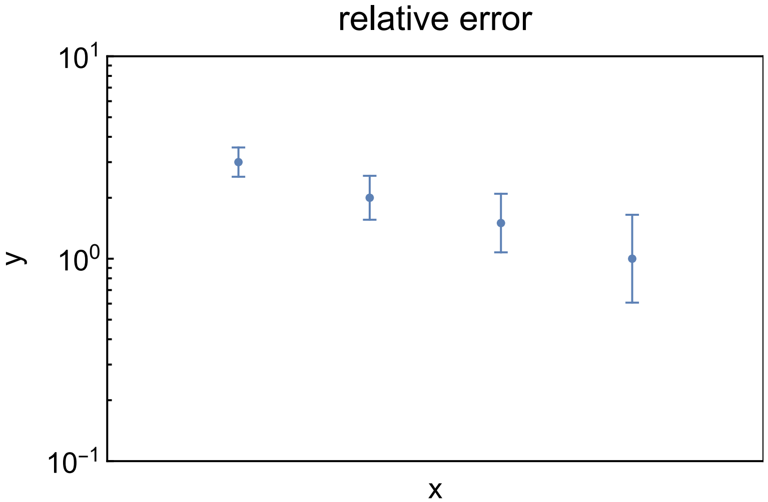
The error δz is thus given by the relative error in y

Out[ ]//TraditionalForm=

$$\delta z = 0.434 \frac{dy}{y}$$

Out[ ]=

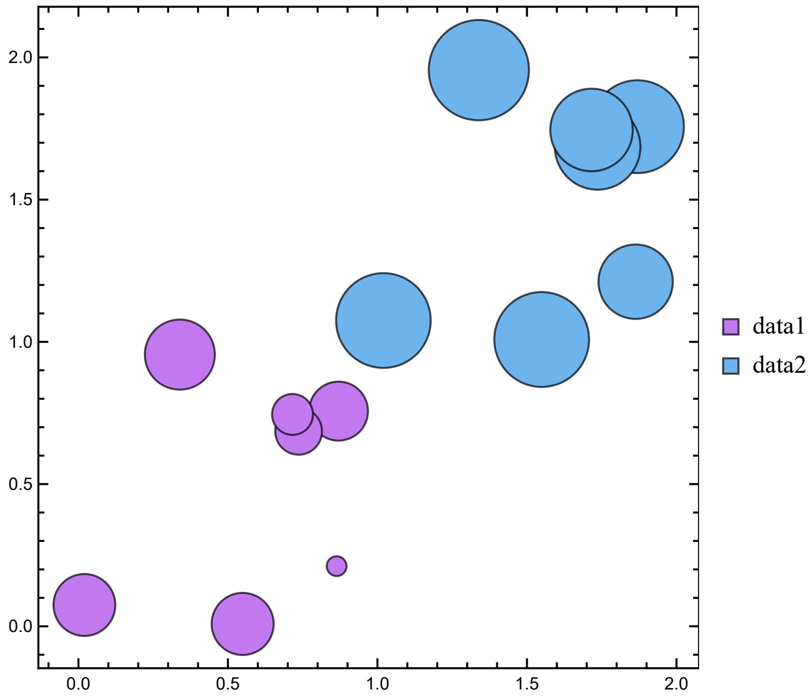




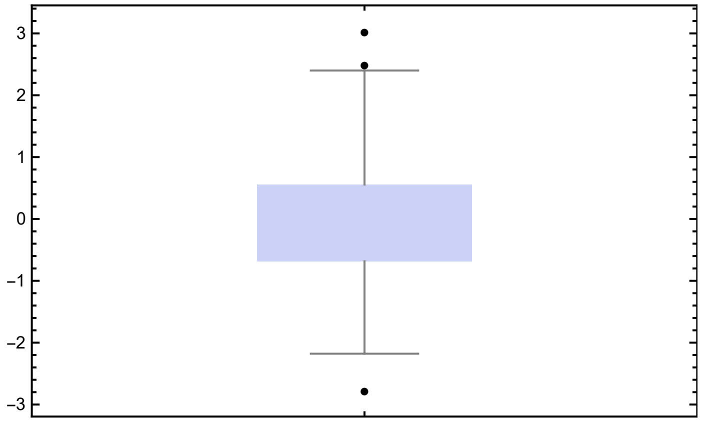
The error bars now display correctly on a logarithmic plot.

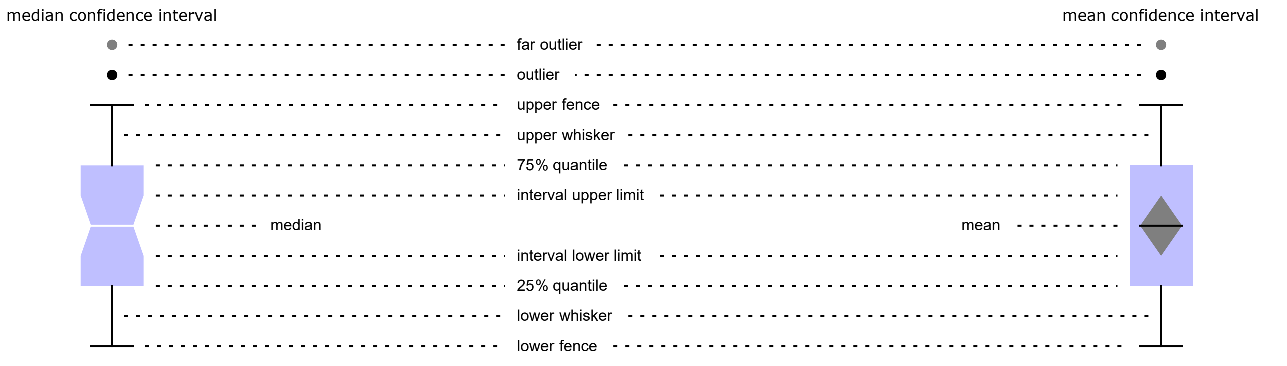
Bubble chart

A bubble chart is a type of chart that displays three dimensions of data. Each entity with its triplet (v_1, v_2, v_3) of associated data is plotted as a disk that expresses two of the v_i values through the disk's xy location and the third through its size.



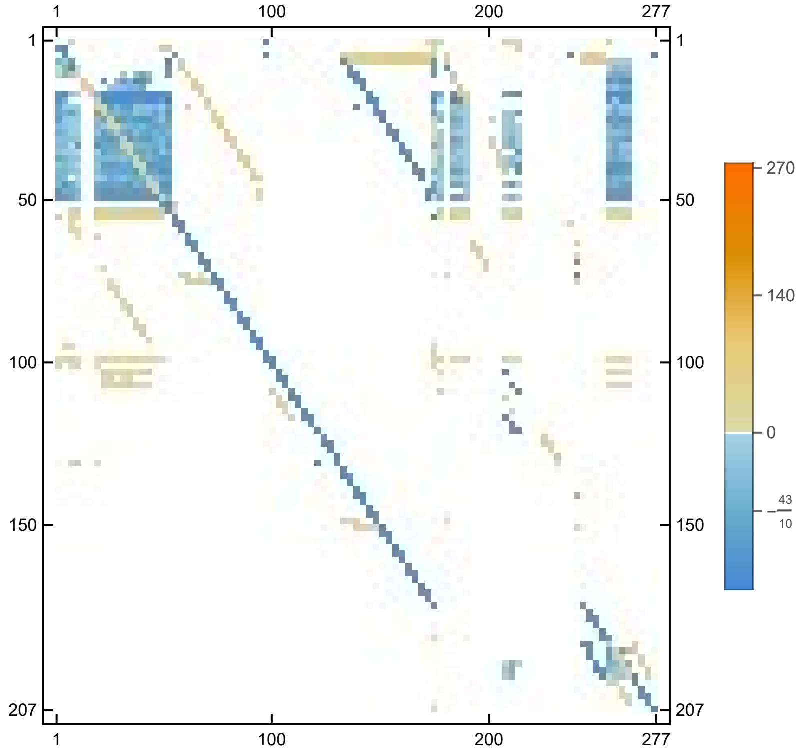
Box-Whisker Chart





Matrix plots

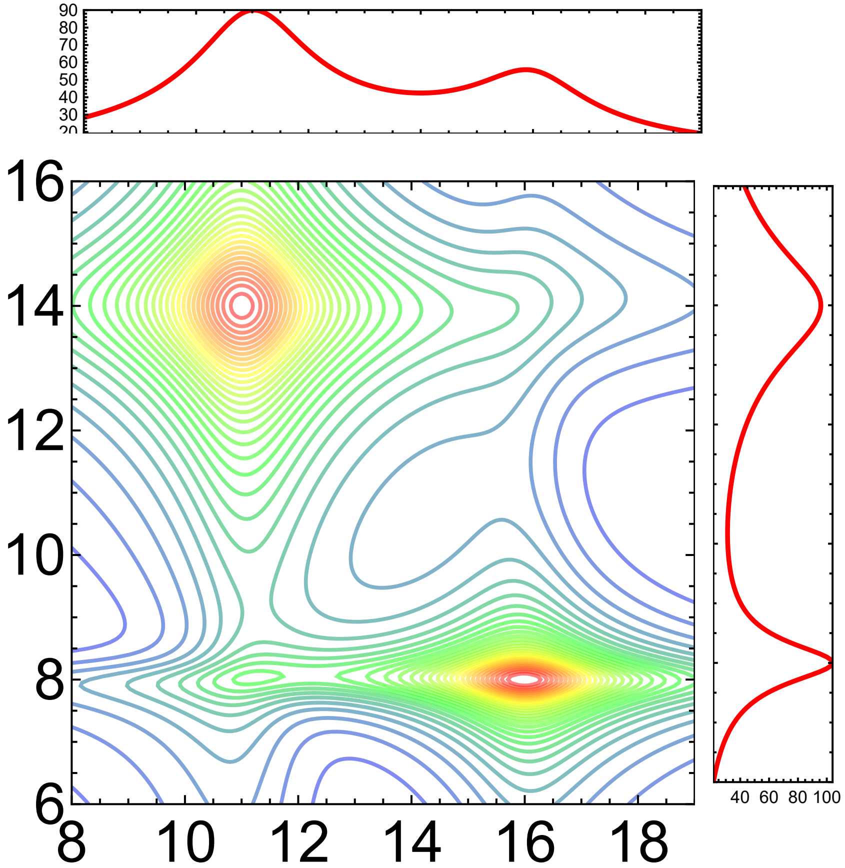
In matrix plots or array plots, the numerical values of a matrix are color coded and plotted in a {x,y} grid.



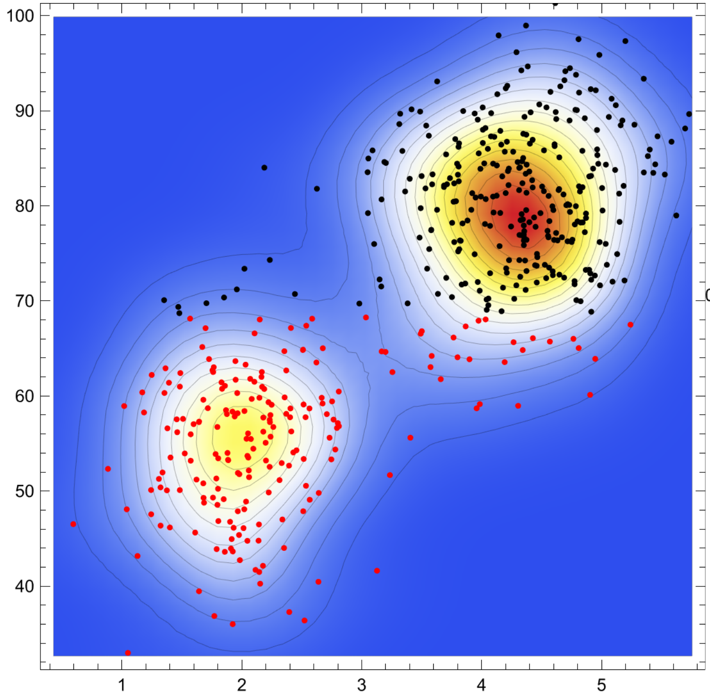
Combined plots

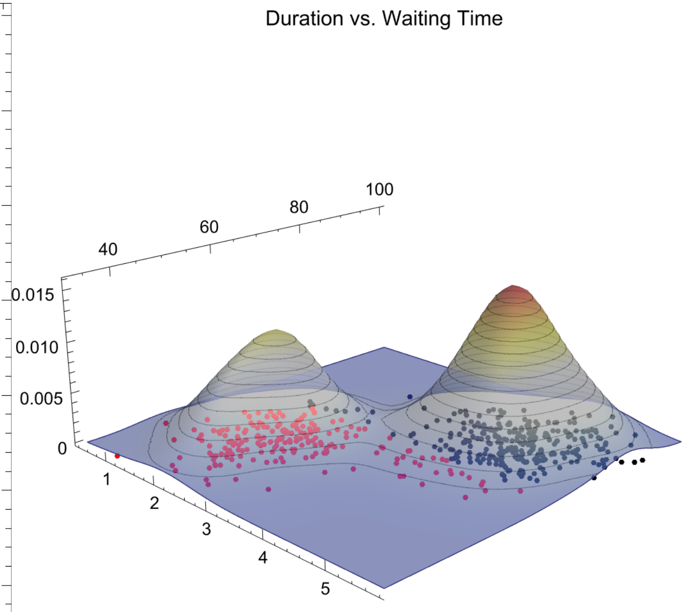
Example:

Out[•]=



Example: Bivariate clustered datasets

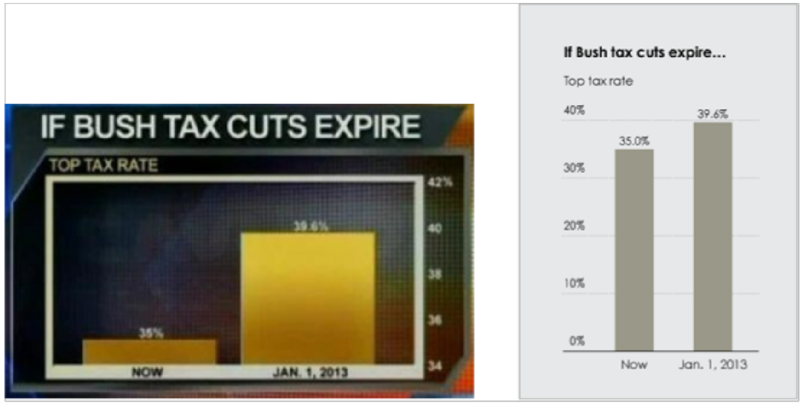




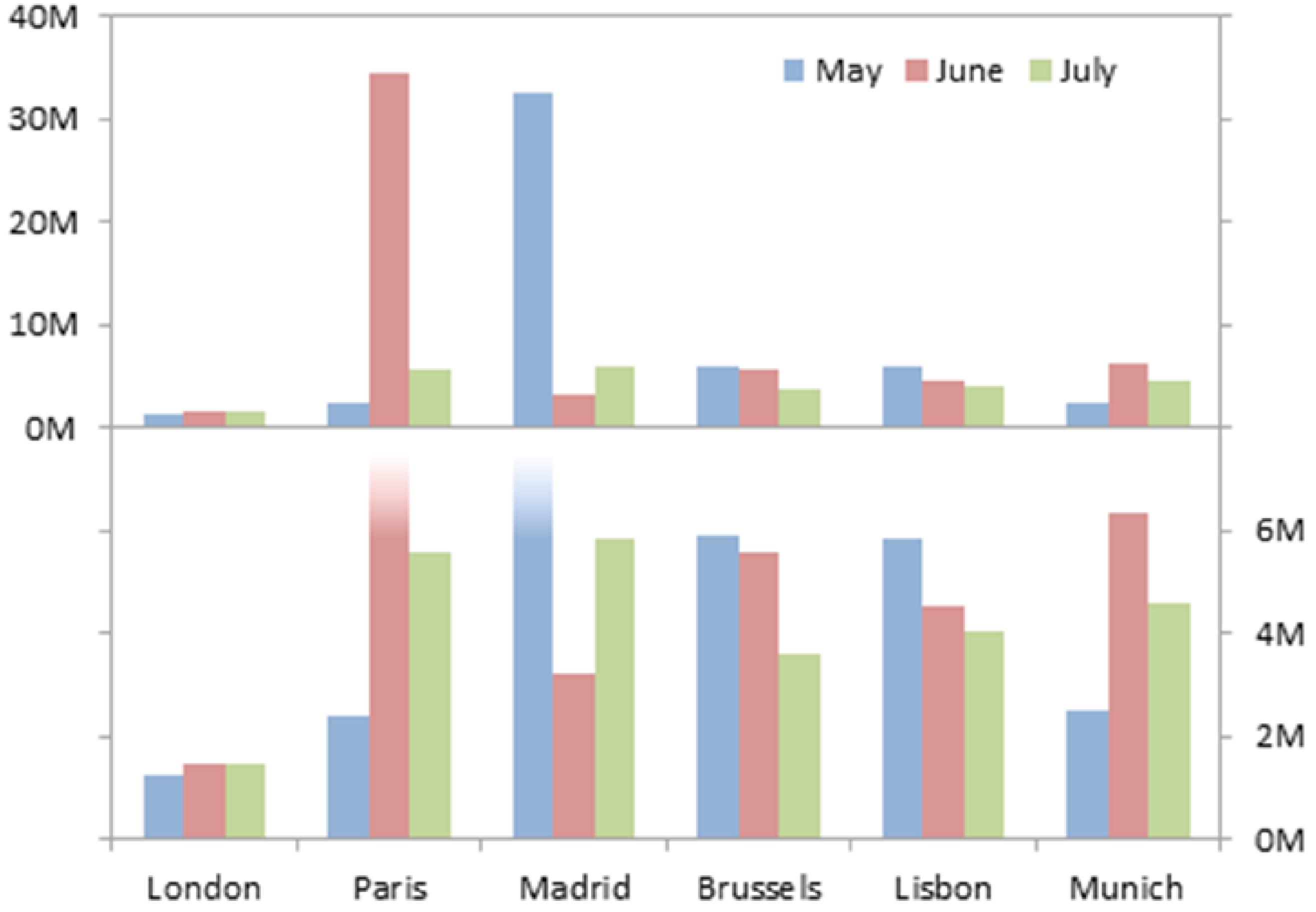
General guidelines - do’s and don’ts

Credits: Angela Zoss (angela.zoss@duke.edu)

Do - Use full axes - avoid distortion



Do - Use full axes - avoid distortion



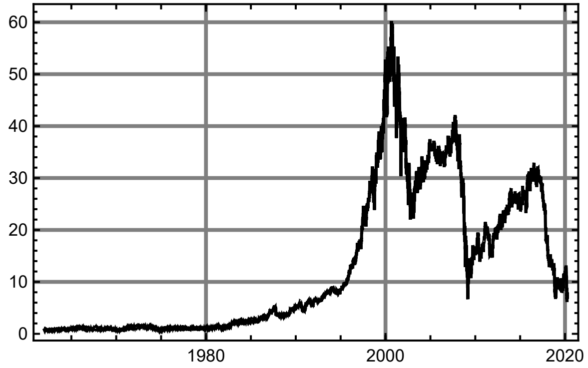
Do - Use full axes - maintain consistency

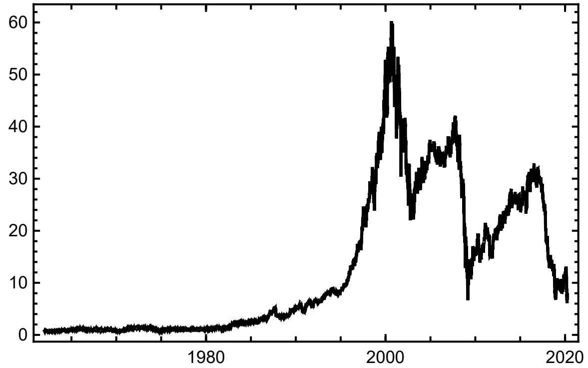
- Consistently use colors, symbols etc.
- When combining plots, make sure to use similar axes ranges, units, etc.

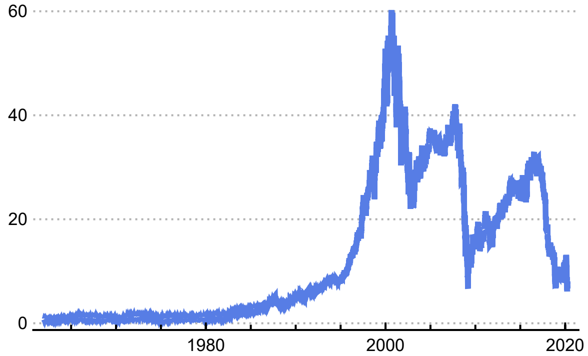
Keep it simple - simplify less important information.

Chart elements like gridlines, axis labels, colors, etc. can all be simplified to highlight what is most important/relevant/interesting.

Out[ ]=







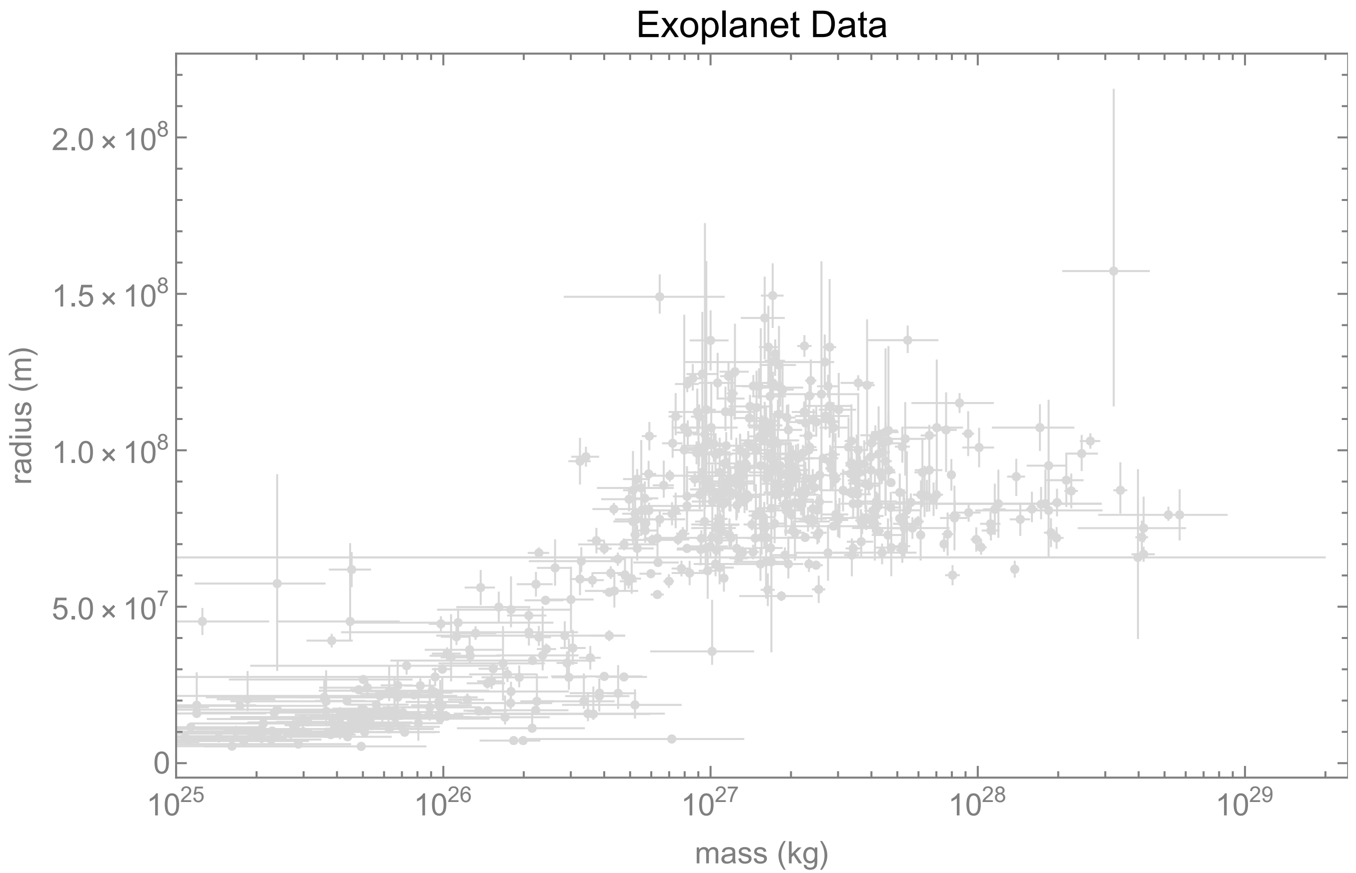
Take a step back - Do pass the squint test!

“When you squint at your page, so that you cannot read any of the text, do you still ‘get’ something about the page?”
You squint your eyes and make an assessment on the overall layout, of elements that stand out, the visual balance and other characteristics.

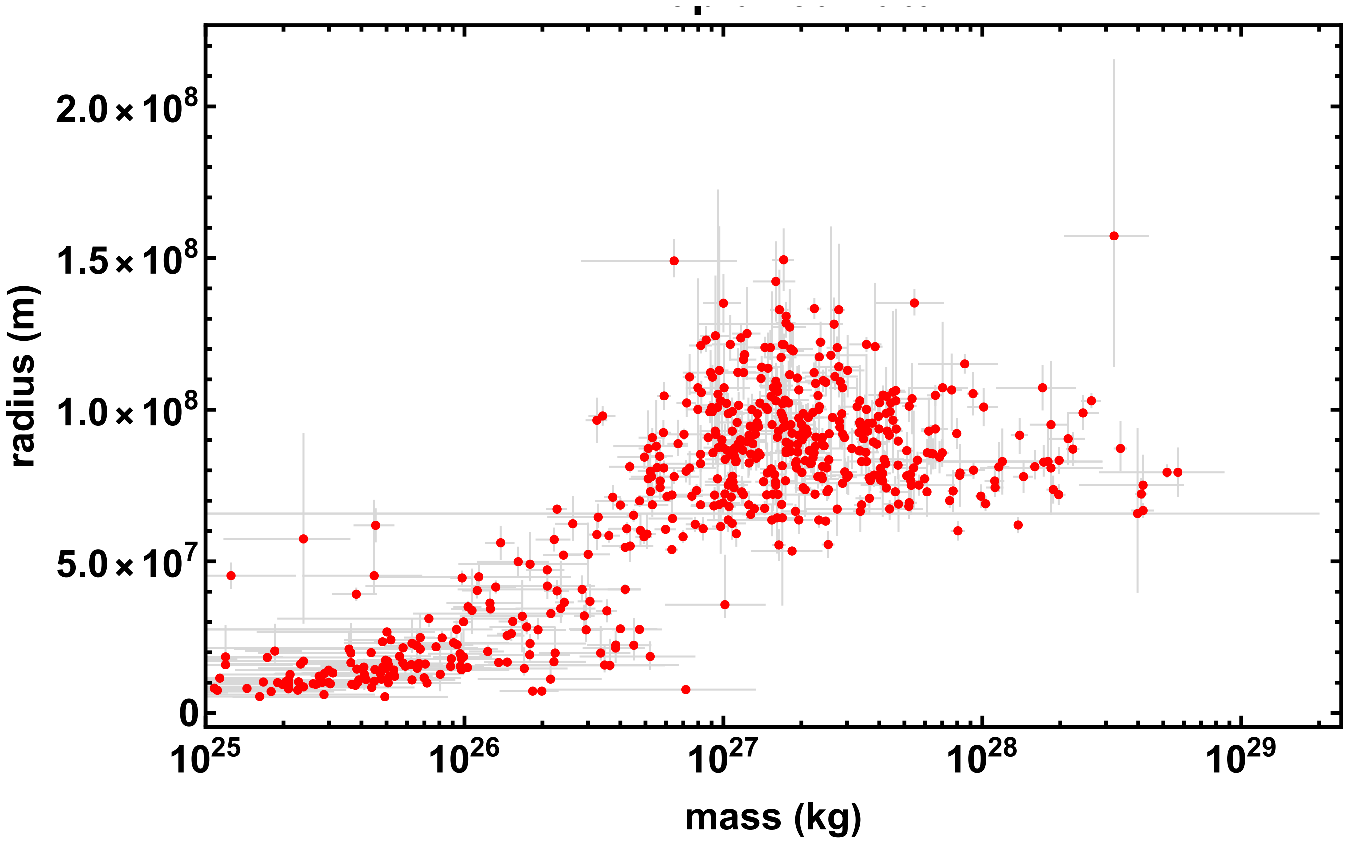
The squint test

- Which elements draw the most attention? What color pops out?
- Do the elements balance? Is there a clear organization?
- Do contrast, grouping, and alignment serve the function of the chart?
- Highly relevant for beamer/presentation.

Out[ ]=



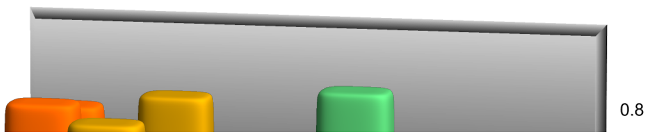
Exoplanet Data

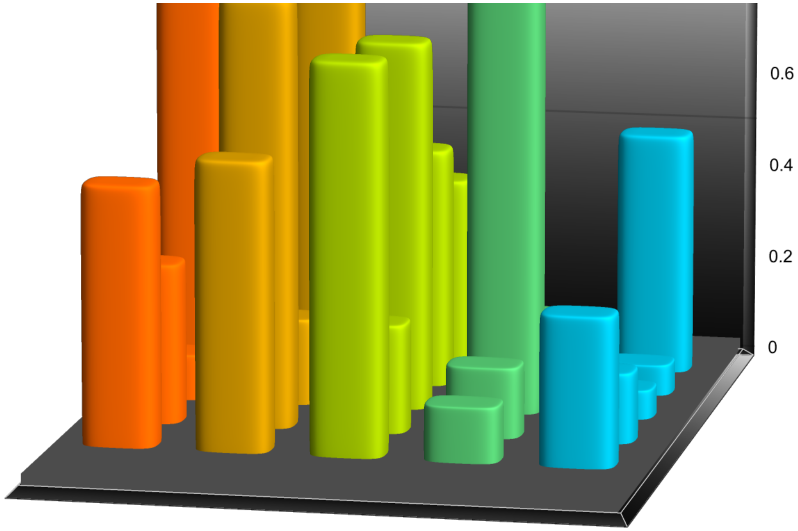

radius (m)
2.0 × 10^8
1.5 × 10^8
1.0 × 10^8
5.0 × 10^7
0
10^25
10^26
10^27
10^28
10^29
mass (kg)

Use 3D effects carefully

Studies show that 3D effects reduce comprehension. Blow apart effects likewise make it hard to compare elements and judge areas.

Out[ ]=




0.6
0.4
0.2
0

Colors: Don’t use more than (about) six colors.

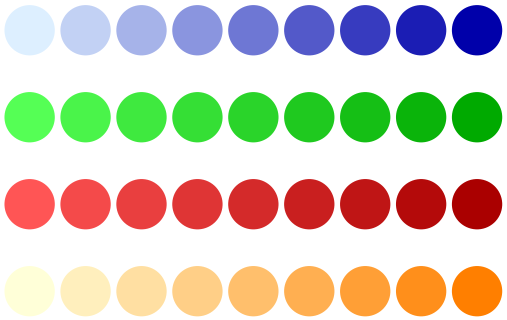
Using color categories that are relatively universal makes it easier to see differences between colors.

The more colors you need (that is, the more categories you try to visualize at once), the harder it is to do this.

Colors:

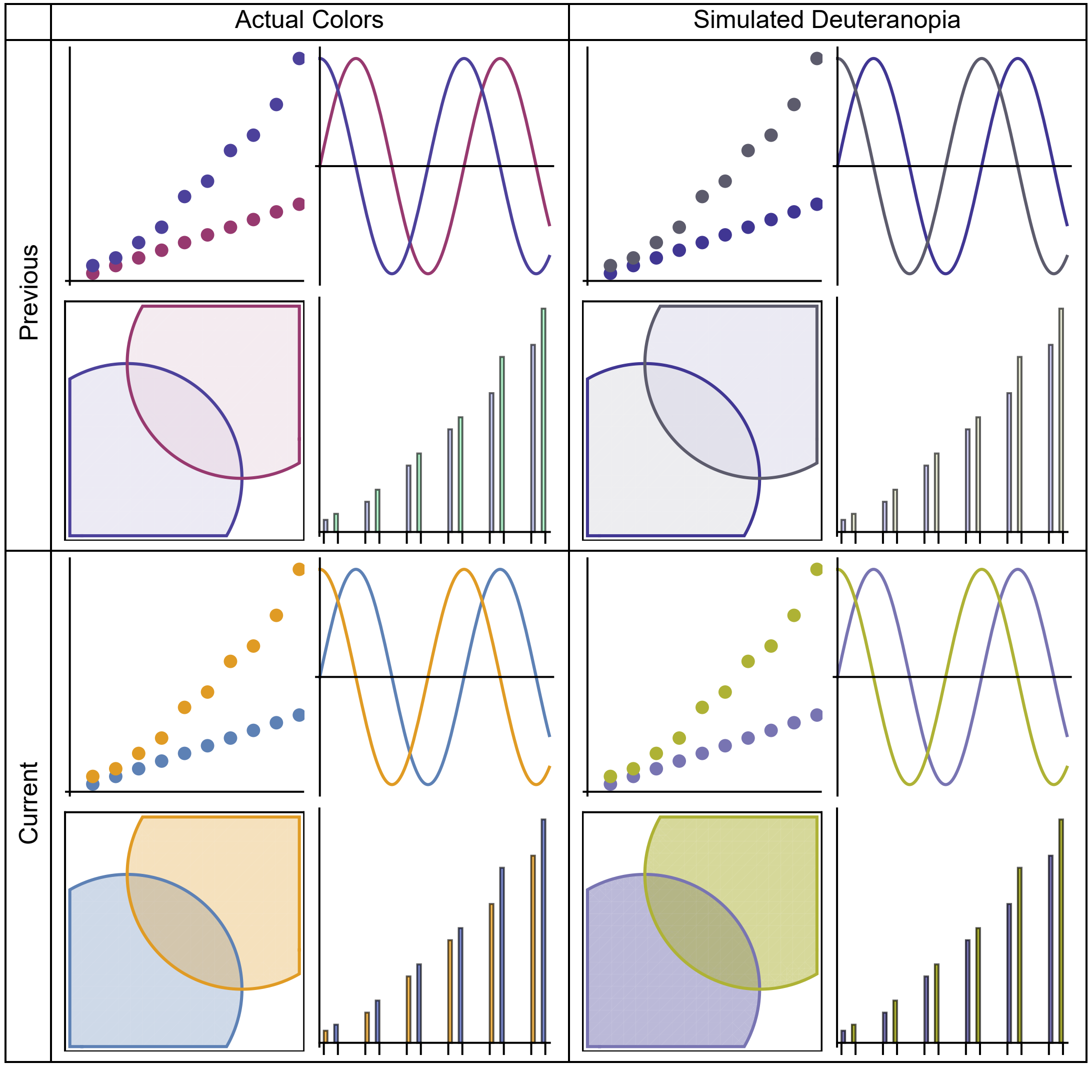
(If you want color to show a numerical value, use a range that goes from white to a highly saturated color in one of the universal color categories.)

Out[ ]=



Colors: Remember the color blind people

Out[ ]=

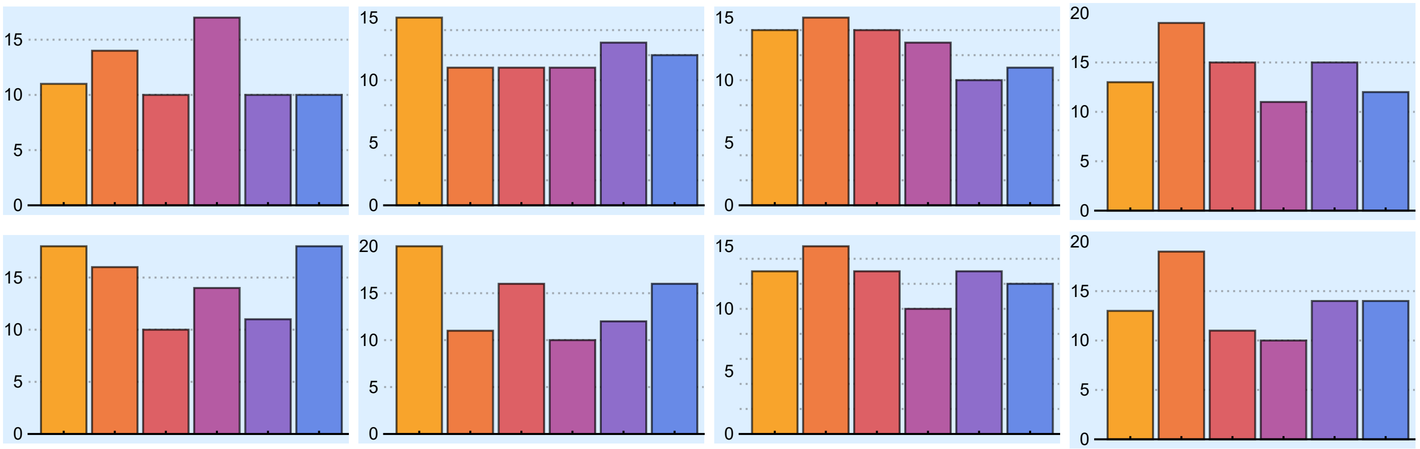


Colors: Also, test what it looks like in gray scale. (Vary both hue and saturation.)

Consistency: Don’t change “style” boats midstream.

One of the easiest ways to get the most out of charts is to rely on comparison to do the heavy lifting. Our visual system can detect anomalies in patterns. Try keeping the form of a chart consistent across a series so differences from one chart to another will pop out.

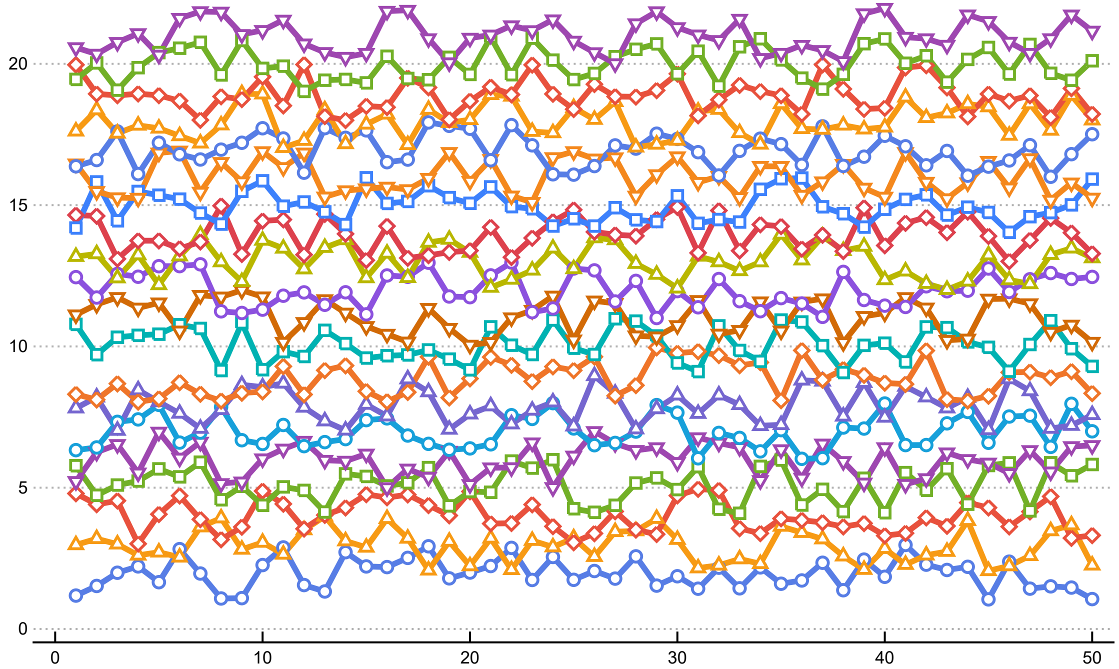
Out[ ]=

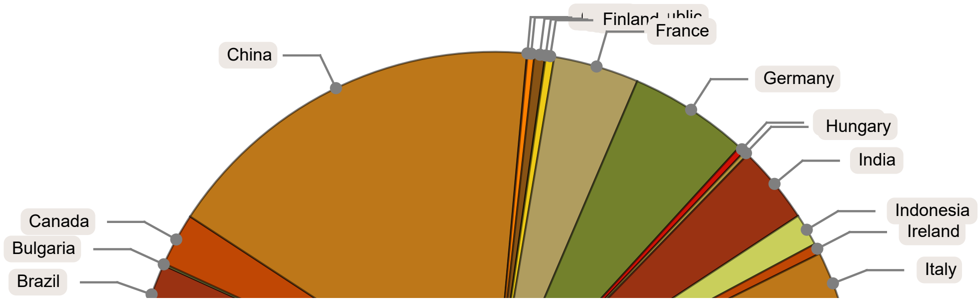


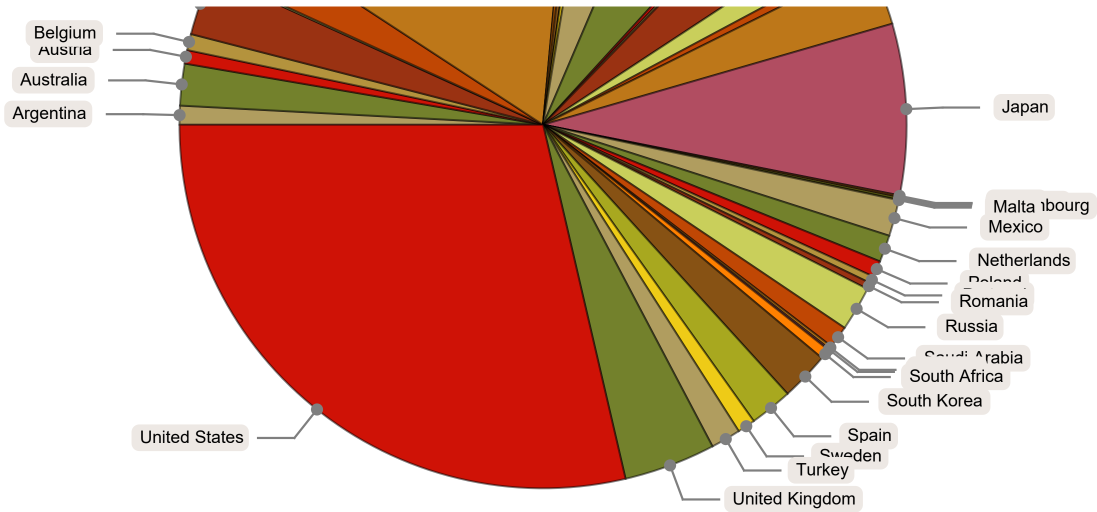
Finally, don’t overload the chart.

Adding too much information to a single chart eliminates the advantages of processing data visually; we have to read every element one by one!

Out[ ]=






Belgium
Austria
Australia
Argentina
Japan
Malta
Mexico
Netherlands
Romania
Russia
South Africa
South Korea
Spain
Turkey
United States
United Kingdom

References

- The Truthful Art: Data, Charts, and Maps for Communication (Voices That Matter), Alberto Cairo, ISBN-10: 0321934075
- Introduction to Data Visualization (http://guides.library.duke.edu/datavis)
- Definitions and Rationale for Visualization (http://www.siggraph.org/education/materials/HyperVis/visgoals/visgoal2.htm)(G. Scott Owen, HyperVis - Teaching Scientific Visualization Using Hypermedia, 1999)
- [Thomas and Cook, 2005] J.J. Thomas and K.A. Cook, eds., Illuminating the Path: The Research and Development Agenda for Visual Analytics, IEEE CS Press, 2005.
- [Thomas and Cook, 2006] James J. Thomas and Kristin A. Cook., IEEE Computer Graphics and Applications, 26(1):10-13, January/February, 2006.
- [Usability First, 2003] Usability First, Usability Glossary. Retrieved at: 2003. http://www.usabilityfirst.com/glossary/main.cgi?function=display_term&term_id=682

References

- ColorBrewer2 .0 http://colorbrewer2.org
- Colorblind Web Page Filter (http://colorfilter.wickline.org/)
- Green, D. A., 2011, `A colour scheme for the display of astronomical intensity images', Bulletin of the Astronomical Society of India, 39, 289. (2011BASI...39..289G at ADS.) or http://arxiv.org/pdf/1108.5083.pdf
- MYCARTA Blog (https://mycarta.wordpress.com/2012/05/29/the-rainbow-is-dead-long-live-the-rainbow-series-outline/)

Additional Material

Perception Issues of Visualization

Perception of Visuals

General Perception Issues

Visualization Rules

Perception of Visuals

Visualizations/Pictures

Information encoded in the entirety of graphical objects and their visual attributes as result of visualization process

Visual attributes

- mode (“flavor”) of presentation chosen, e.g. color, size, orientation
- clever choice of visual attributes is paramount to visualization process
- redundancy of visual attributes enhances interpretability

Visual attributes

- Interpretation of visual attributes
  - Innate reaction
    - natural to interpret
    - example: increasing brightness → increasing numerical values
    - pre-conscious/pre-attentive interpretation
  - Acquired reactions
    - acquired through education, usually more complex
    - example: color ranking, street/travel signs, isolines, isosurfaces
  - Illusory visual attributes
    - optical illusions, unwanted side effects

Color

- **psychophysical** process
  - **physics**: *relation to wavelengths*, spectral distribution and amount of light entering eye
  - **psychology**: *perceived sensation with no linear relation to physics*
- no complete theory (**three types of cones: R,G,B**; opponent theory)
- variety of color spaces: geometric descriptions of color gamut

Color

- **perceptual dimensions of color**: hue, saturation, intensity
  may be varied independently or in connection to each other
  - **hue**: “colors” of rainbow (relates to wavelength)
  - **saturation**: “paleness” of color is lack of saturation (relates to spectral distribution)
  - **intensity**: light/dark colors (relates to amount of light entering eye) - brightness

Hue

effective use for nominal data types and ordinal data types (color scale!)

When white light is incident on a surface some wavelength’s are absorbed and others are reflected. This gives us the perception of the color of the object.**The dominant wavelength is called the color or hue of the surface.**

Our eyes respond to two other sensations in addition to the dominant wavelength. These are the luminance or **brightness**, which is a function of the intensity (energy) of light and purity or **saturation**. Saturation is how washed out or “pure” the color is, e.g., pastels are not very pure. The chromaticity of light refers to the purity and dominant frequency.

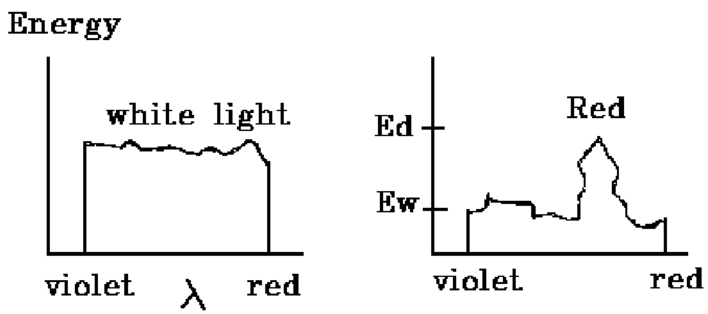


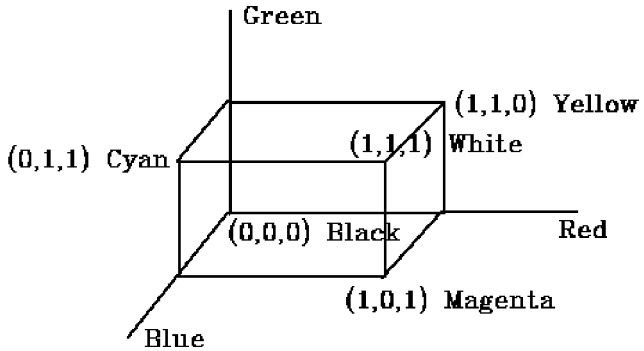
Fig: the **brightness** equals the area under curve or total energy. The purity is equal to Ed - Ew. For example light is 100% pure (Ed - Ew = 100) when Ew = 0 and 0% pure (Ed - Ew = 0) when Ew = Ed.

(source: http://www.siggraph.org/education/materials/HyperGraph/color/images/colorli1.gif)

RGB color model

In the RGB color model, we use **red**, **green**, and **blue** as the 3 primary colors. We don’t actually specify what wavelengths these primary colors correspond to, so this will be different for different types of output media, e.g., different monitors, film, videotape, slides, etc.

This is an **additive model** since the phosphors are emitting light. A subtractive model would be one in which the color is the reflected light. We can represent the RGB model by using a unit cube. Each point in the cube (or vector where the other point is the origin) represents a specific color. This model is the best for setting the electron guns for a CRT.



Note that for the “complementary” colors the sum of the values equals white light (1, 1, 1). For example:

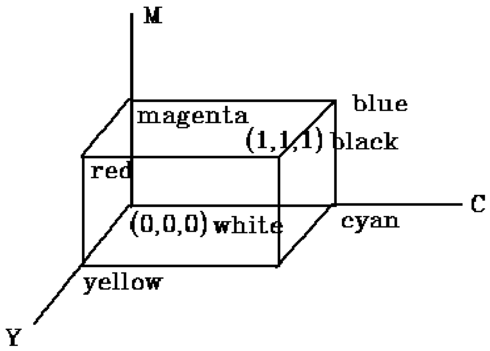
red (1, 0, 0) + cyan (0, 1, 1) = white (1, 1, 1)

green (0, 1, 0) + magenta (1, 0, 1) = white (1, 1, 1)

blue (0, 0, 1) + yellow (1, 1, 0) = white (1, 1, 1)

CMY color model

CRTs (cathode ray tubes) produce color by emission and uses the RGB model. Printers produce color by reflective light so it is a subtractive process and uses a model based on the colors: Cyan, Magenta, Yellow.

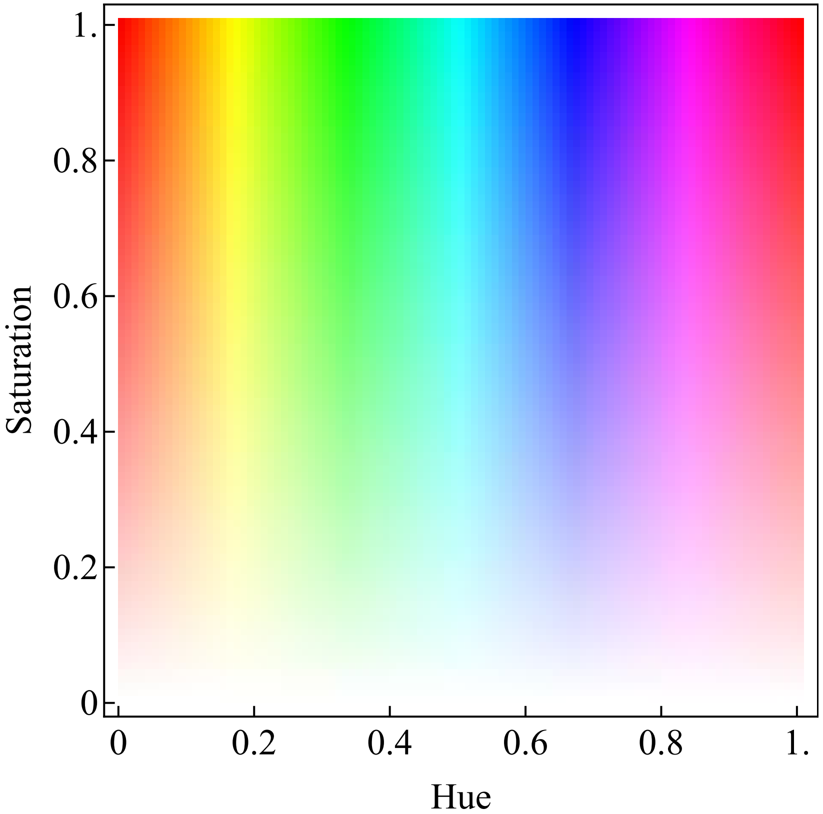


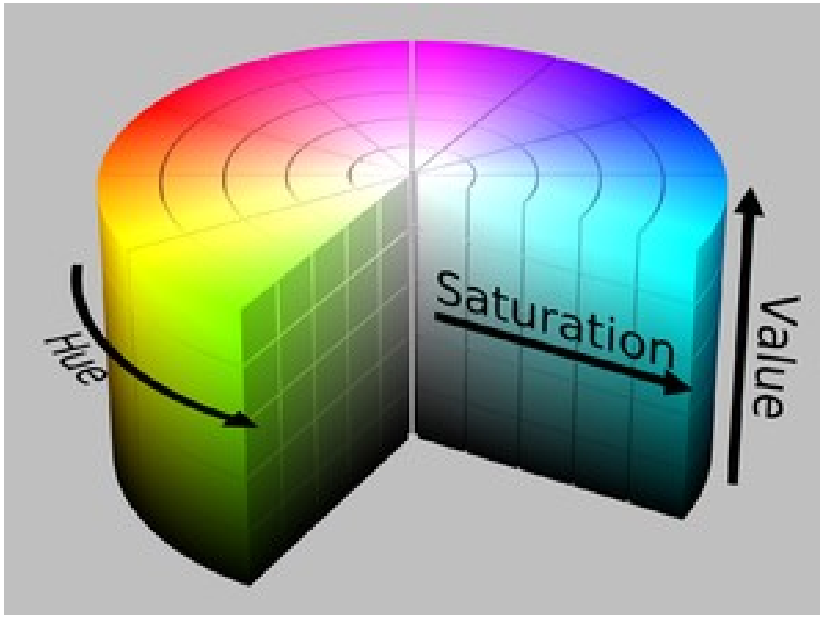
Remember that cyan = green + blue, so light reflected from a cyan pigment has no red component, i.e., the red is absorbed by cyan. Similarly magenta subtracts green and yellow subtracts blue. Printers usually use four colors: cyan, yellow, magenta and black. This is because cyan, yellow, and magenta together produce a dark gray rather than a true black.

$$\begin{pmatrix} C \\ M \\ Y \end{pmatrix} = 1 - \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

HSV color model

The HSV (**Hue**, **Saturation**, and **Value**) color model is more intuitive than the RGB color model. The user specifies a color (hue) and then adds white or black. There are 3 color parameters: Hue, Saturation, and Value. Changing the saturation parameter corresponds to adding or subtracting white and changing the value parameter corresponds to adding or subtracting black.





HSV color model

The saturation varies between $0.0 \leq s \leq 1.0$ and is the ratio of purity of a related hue to its maximum purity at $s = 1$. at s equals 0 is the gray scale, that is the diagonal of the RGB cube corresponds to v of the HSV hexcone. notice the complementary colors(red + cyan, blue + yellow, green + magenta) are diagonally opposite.

So to choose a color we do the following:
1. select pure hue (specifies H and sets S = V = 1)
To add black decrease V and/or to add white decrease S.

Intuitive color concepts

Artists start with a “pure color or hue”, then add black pigment to produce different shades. The more black pigment the darker the shade. They add white pigment and get different tints. Adding both black and white pigments gives different tones. If we look at the cross-section of the hexcone we can see the analogy with the artists model.

Emotional Response to color

Different colors evoke different reactions in viewers. Be aware that some of these reactions will be culturally specific.
Why are so many astronomy talks given with a blue background color?

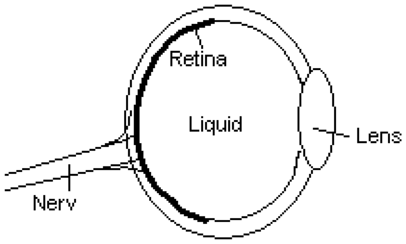
- Red - danger, stop, negative, excitement, hot
- Dark Blue - stable, calming, trustworthy, mature
- Light Blue - youthful, masculine, cool
- Green - growth, positive, organic, go, comforting
- White - pure, clean, honest Black - serious, heavy, death
- Gray - integrity, neutral, cool, mature Brown - wholesome, organic, unpretentious
- Yellow - emotional, positive, caution

Emotional Response to color

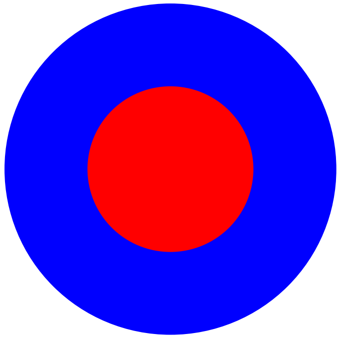
- Gold - conservative, stable, elegant
- Orange - emotional, positive, organic
- Purple - youthful, contemporary, royal
- Pink - youthful, feminine, warm
- Pastels - youthful, soft, feminine, sensitive
- Metallic - elegant, lasting, wealthy

Physiological Principles for the Effective Use of Color

Reference: “Physiological Principles for the Effective Use of Color”, G. Murch, IEEE CG&A, pp. 49-54, Nov., 1984.



Lens



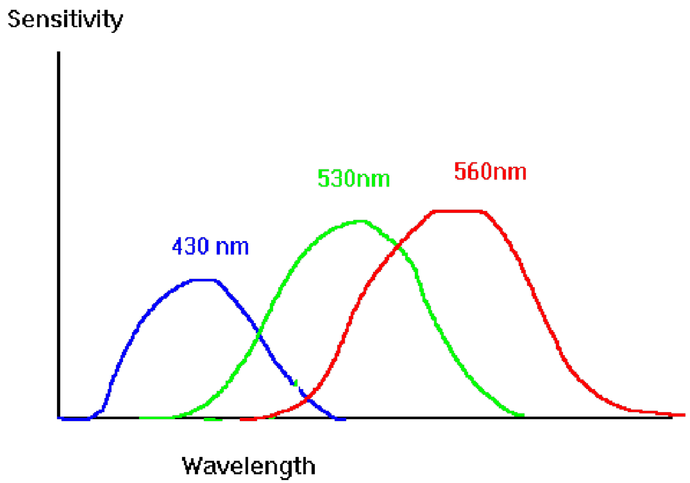
For a given lens curvature, longer wavelengths have a longer focal length, i.e., red is the longest focal length and blue is the shortest. Since to have an image focused on the retina, the lens curvature must change with wavelength with red light requiring the greatest curvature and blue light the least curvature. This means that if pure blue and pure red hues are intermixed, the lens is constantly changing shape and the eye becomes tired.

Physiological Principles for the Effective Use of Color

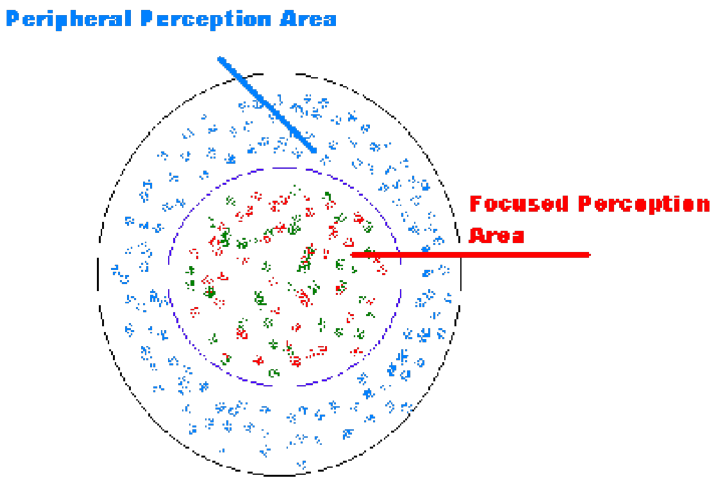
Retina

The retina contains the photo receptors that absorb photons and transmit chemical signals to the brain. There are two types: **rods** which are night-vision receptors and have no color dependency, and **cones**, which have color sensitivity and require a higher level of light intensity than the rods.

There are three types of **photopigments in the cones**; "**blue**" with a maximum sensitivity at *430 nm*, "**green**" with a maximum sensitivity at *530 nm*, and "**red**" at *560 nm* (this wavelength actually corresponds to yellow). Light at a single wavelength will partially activate all three types of cones, e.g. at a wave-length of 470 nm, it will be strong blue plus some red and green component.



Physiological Principles for the Effective Use of Color



The percentage of cones is not equal but is as follows: blue (4%), green (32%), and red (64%). In addition, the cones are differentially distributed in the retina. The center of the retina has a dense concentration of cones but no rods while the periphery has many rods but few cones. The color distribution is also asymmetrical. The center of the retina is primarily green cones, surrounded by red–yellow cones, with the blue cones being mainly on the periphery. **The center of the retina has no blue cones.**

We see objects by edge detection, where an edge can be created by a difference in color or brightness or both. Edges formed by color differences alone, with no brightness differences, appear fuzzy and unfocused, so we need to add changes in brightness to get sharp edges.

Color blindness

About nine percent of the population has some kind of color perception problem. The most common is red-green deficiency, which can arise from a deficiency of either the red or the green photopigments. These people have difficulty distinguishing any color that is dependent upon the red:green ratio. In general people will have slightly color perceptions and these may be extreme.

hints & guidelines

- small blue objects: disadvantage for short-wavelengths cones
- blue (cool colors): farther away, cooler, lower or negative values
- red (warm colors): nearer, warmer, higher and positive values; danger
- shape of object displayed with rainbow scale may not be readily apparent
- hues may change appearance on different backgrounds
- "color-blindness"
- ranking of hues not inherent
- discontinuous color scales

hints & guidelines

- Avoid the simultaneous display of highly saturated, spectrally extreme colors. Remember that this causes the lens to rapidly change shape and so tires the eyes. Desaturate the colors or else use colors that are close together in the spectrum.
- Pure blue should be avoided for text, thin lines, and small shapes. Since there are no blue cones in the center of the retina, these are difficult to see. But blue makes an excellent background color, e.g. for a computer display it tends to blur the raster lines.
- Avoid adjacent colors that differ only in the amount of blue. Since blue does not contribute to brightness, this creates fuzzy edges.
- Older operators need higher brightness levels to distinguish colors.
- Colors change in appearance as the ambient light level changes. The magnitude of a detectable change in color varies across the spectrum.

hints & guidelines

- It is difficult to focus upon edges created by color alone. Avoid red and green in the periphery of large displays.
- Opponent colors go well together.
- For color-deficient observers, avoid single color distinctions.

Saturation

- effective use of saturation for ordinal data types
- careful when interpreting saturation and brightness independently

Brightness

effective use of brightness for ordinal (and quantitative) data types

- hints
  - bright objects on dark background look bigger than dark objects on bright background
  - fading brightness gives impression of distance/depth
  - absolute brightness not perceived linearly
  - change of brightness not perceived linearly (Machband)
  - brightness contrast influences perception of brightness

Texture

effective use for nominal data types

- hints
  - careful with overlapping textures
  - textures may give rise to other impressions, e.g.density
  - include legend

Orientation

- hints
  - familiarity of shape often connected to orientation
  - things don’t look the same when their orientation is changed.
  - symmetry around vertical axis preferred
  - use various orientations to assure correct view of objects

Depth attributes

Use depth attributes to enhance the perception of 3-d structures

- fading brightness to show increasing depth
- perspective geometry to show increasing depth
- occlusion to distinguish back/front
- transparency/translucency to distinguish back/front
- change of brightness (shading) to simulate surfaces
- rotation/"rocking" to enhance 3-d perception
- stereo effect: anaglyph, shutter glasses, VR

Motion

- frame update rate to perceive motion
  - at least 10 frames/sec [BRY94]
- Examples
  - animation
  - flick two or more images to depict differences, similarities

Init

In[◦]:=
```
SetDirectory[NotebookDirectory[]];
```

In[◦]:=
```
SetOptions[$FrontEndSession, UnderoverscriptBoxOptions → {LimitsPositioning → False}]
```

In[◦]:=
```
ClearAll[equation]
Attributes[equation] = {HoldAll, HoldAllComplete};
equation[eq___] := TraditionalForm@Framed[HoldForm@Defer[eq], FrameStyle → Directive[Gray], FrameMargins → 10]
ClearAll[equationNoBox]
Attributes[equationNoBox] = {HoldAll, HoldAllComplete};
equationNoBox[eq___] := TraditionalForm[HoldForm@Defer[eq]]
```

In[◦]:=
```
equation[ℱ[k, "N", ρ] == 𝒫[X ≤ k] == Sum[Binomial["N", m] ρ^n (1 - ρ)^("N"-n), {n, 0, k}]]
```

Out[◦]//TraditionalForm=

$$\boxed{\mathcal{F}(k, \mathrm{N}, \rho) = \mathcal{P}(X \leq k) = \sum_{n=0}^{k} \binom{\mathrm{N}}{m} \rho^{n} (1-\rho)^{\mathrm{N}-n}}$$

In[◦]:=
```
equationNoBox[ℱ[k, "N", ρ] == 𝒫[X ≤ k] == Sum[Binomial["N", m] ρ^n (1 - ρ)^("N"-n), {n, 0, k}]]
```

Out[◦]//TraditionalForm=

$$\mathcal{F}(k, \mathrm{N}, \rho) = \mathcal{P}(X \leq k) = \sum_{n=0}^{k} \binom{\mathrm{N}}{m} \rho^{n} (1-\rho)^{\mathrm{N}-n}$$

In[◦]:=
```
Attributes[equationInvBox] = {HoldAll, HoldAllComplete};
equationInvBox[eq___] := TraditionalForm@Framed[Style[HoldForm@Defer[eq]], FrameStyle → Invisible]
```

In[ ]:=
```
Format[parens[e_]] := Style[DisplayForm@RowBox[{"(", MakeBoxes@e, ")"}], SpanMaxSize → Infinity]
Format[brackets[e_]] := Style[DisplayForm@RowBox[{"[", MakeBoxes@e, "]"}], SpanMaxSize → Infinity]
Format[parens[e_, i_]] := Style[DisplayForm@RowBox[{"(", MakeBoxes@e, ")"}], SpanMaxSize → Infinity, SpanMinSize → i]
Format[brackets[e_, i_]] := Style[DisplayForm@RowBox[{"[", MakeBoxes@e, "]"}], SpanMaxSize → Infinity, SpanMinSize → i]
 equationNoBox[Subscript[OverBar[parens[X / Y]], "geom"] == foo]
equation[Subscript[OverBar[brackets[X / Y]], "geom"] == foo]
equationNoBox[Subscript[OverBar[brackets[X / Y, 3]], "geom"] == foo]
```

Out[ ]//TraditionalForm=

$$\left(\frac{X}{Y}\right)_{\text{geom}} = \text{foo}$$

Out[ ]//TraditionalForm=

$$\boxed{\left[\frac{X}{Y}\right]_{\text{geom}} = \text{foo}}$$

Out[ ]//TraditionalForm=

$$\left[\frac{X}{Y}\right]_{\text{geom}} = \text{foo}$$

Special cells

In[ ]:=
```
ToPictureCell[expr_] := CellPrint[ExpressionCell[expr, "Picture", ShowStringCharacters → False]]
```

In[ ]:=
```
PicturePrint[expr_, opts : OptionsPattern[Cell]] := CellPrint[TextCell[expr, "Picture", opts, ShowStringCharacters → False]]
```

Set options

In[ ]:=
```
imgsize = 500;
fontfam = "Times";
fontsize = 13;
sty  = Directive[fontsize, FontFamily → fontfam];
```

In[ ]:=
```
SetOptions[#, BaseStyle → Thick, PlotStyle → ColorData[1, "ColorList"], ImageSize → imgsize, PlotRange → All,
     FillingStyle → Automatic, TicksStyle → Directive[fontsize, FontFamily → fontfam], Frame → True, Axes → False] & /@
   {Plot, ListPlot, LogPlot, LogLogPlot, ListLogLogPlot, LogLinearPlot, ListLogLinearPlot};
```

```
In[◦]:= SetOptions[Histogram, {ImageSize → Large, BaseStyle → {Directive[fontsize, FontFamily → fontfam]}}];

In[◦]:= SetOptions[{Histogram3D, DensityHistogram}, {ImageSize → Medium, BaseStyle → {Directive[fontsize, FontFamily → fontfam]}}];

In[◦]:= OverviewMouseover[header_, subs_] :=
  CellPrint@Cell[BoxData[ToBoxes@Mouseover[Framed[Style[header, "OverviewSection"], FrameMargins → {{15, 15}, {5, 5}},
        BoxFrame → 2, FrameStyle → None, RoundingRadius → 5],

       Evaluate@If[subs == "", Framed[Style[header, "OverviewSection", White],
          FrameMargins → {{15, 15}, {5, 5}}, BoxFrame → 2, FrameStyle → None, RoundingRadius → 5, Background → m8red[1]],
         Row[{Framed[Style[header, "OverviewSection", White], FrameMargins → {{15, 15}, {5, 5}}, FrameStyle → None,
            RoundingRadius → 5, Background → m8red[1]], " » ", Framed[Style[subs, "OverviewSection"], FrameMargins → {{15, 15},
              {5, 5}}, FrameStyle → m8red[1], BoxFrame → 2, RoundingRadius → 5, Background → White]}]]]], "OverviewSection"]

In[◦]:= LargeOverviewMouseover[header_, subs_] :=
  CellPrint@Cell[BoxData[ToBoxes@Mouseover[Framed[Style[header, "LargeOverviewSection"],
        FrameMargins → {{15, 15}, {5, 5}}, BoxFrame → 2, FrameStyle → None, RoundingRadius → 5],

       Evaluate@If[subs == "", Framed[Style[header, "LargeOverviewSection", White], FrameMargins → {{15, 15}, {5, 5}}, BoxFrame → 2,
          FrameStyle → None, RoundingRadius → 5, Background → m8red[1]], Row[{Framed[Style[header, "LargeOverviewSection", White],
            FrameMargins → {{15, 15}, {5, 5}}, FrameStyle → None, RoundingRadius → 5, Background → m8red[1]], " » ",
           Framed[Style[subs, "LargeOverviewSection"], FrameMargins → {{15, 15}, {5, 5}}, FrameStyle → m8red[1],
            BoxFrame → 2, RoundingRadius → 5, Background → White]}]]]], "LargeOverviewSection"]

In[◦]:= MapThread[LargeOverviewMouseover, Transpose[{{"What is Mathematica?", "Uses, community, prebuilt materials"},
    {"A Look at Workflow", "syntax, palettes, free-form input"}, {"Mathematica in the Classroom",
     "teaching and learning applications"}, {"Work with Data", "built-in curated data, your own data"}}]];
```

What is *Mathematica*?

A Look at Workflow

Mathematica in the Classroom

Work with Data

In[ ]:=
```
Clear[numberLine];
numberLine[ssize_] := {SeedRandom[1234];
  Block[{data = RandomReal[{-1, 1}, 20], mean, dev, sample, pos, smean, sdev},
   mean = Mean[data];
   dev = 1/Length[data] Total[(data - mean)^2];
   SeedRandom[];
   sample = RandomChoice[data, ssize];
   pos = Position[data, #] & /@ sample;
   smean = Mean[sample]; sdev = 1/Length[sample] Total[(sample - smean)^2];
   Sow[{smean, sdev}];
   Graphics[{
     PointSize[Large], Red, Point[{#, 0.05}] & /@ data,
     Black, Thick, Arrowheads[{-.03, .03}], Arrow[{{-1, 0}, {1, 0}}],
     Line[{{mean, 0.05}, {mean, -0.05}}],
     Line[{{mean + dev, 0.0}, {mean + dev, -0.05}}], Line[{{mean - dev, 0.0}, {mean - dev, -0.05}}],
     Circle[{#, 0.05}, 0.03] & /@ sample,
     Orange,
     Line[{{smean, 0.05}, {smean, -0.05}}],
     Line[{{smean + sdev, 0.0}, {smean + sdev, -0.05}}], Line[{{smean - sdev, 0.0}, {smean - sdev, -0.05}}],}, ImageSize → 700,
    PlotLabel → Row[{"μ=", mean, "   σ^2=", dev}]]]}
```

In[ ]:=
```
elements = {"H", "He", "O", "C", "Ne", "Fe", "N", "Si", "Mg", "S", "..."};
abundances = {739000, 240000, 10400, 4600, 1340, 1090, 960, 650, 580, 440, 500};
```

In[ ]:=
```
sides = Prepend[Sqrt[abundances] // N, 0]
xvalues = Accumulate[sides]
```

Out[ ]=
```
{0, 859.651, 489.898, 101.98, 67.8233, 36.606, 33.0151, 30.9839, 25.4951, 24.0832, 20.9762, 22.3607}
```

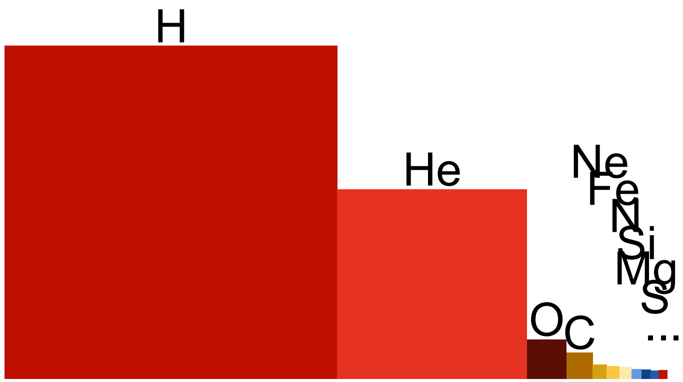
Out[ ]=
```
{0, 859.651, 1349.55, 1451.53, 1519.35, 1555.96, 1588.97, 1619.96, 1645.45, 1669.54, 1690.51, 1712.87}
```

In[ ]:=
```
Table[{ColorData[49][i], Rectangle[{0, 0}, {sides[[i + 1]], sides[[i + 1]]}]}, {i, 1, Length[elements]}]
GraphicsRow[Graphics[#, PlotRange → {{0, 1000}, {0, 1000}}, AspectRatio → 1, ImageSize → Small] & /@ %]
```

Out[ ]=
```
{{, Rectangle[{0, 0}, {859.651, 859.651}]}, {, Rectangle[{0, 0}, {489.898, 489.898}]},
 {, Rectangle[{0, 0}, {101.98, 101.98}]}, {, Rectangle[{0, 0}, {67.8233, 67.8233}]}, {, Rectangle[{0, 0}, {36.606, 36.606}]},
 {, Rectangle[{0, 0}, {33.0151, 33.0151}]}, {, Rectangle[{0, 0}, {30.9839, 30.9839}]},
 {, Rectangle[{0, 0}, {25.4951, 25.4951}]}, {, Rectangle[{0, 0}, {24.0832, 24.0832}]},
 {, Rectangle[{0, 0}, {20.9762, 20.9762}]}, {, Rectangle[{0, 0}, {22.3607, 22.3607}]}}
```

Out[ ]=


In[ ]:=
```
SeedRandom[111];
Labeled[
 Graphics[Table[{ColorData[20][i], Rectangle[{xvalues[[i]], 0}, {xvalues[[i + 1]], sides[[i + 1]]}], Text[Style[elements[[i]],
      Large, Black], {xvalues[[i]] + (xvalues[[i + 1]] - xvalues[[i]]) 0.5, If[i ≤ 4, sides[[i + 1]] + 50, 700 - (i - 3) * 70]}]},
   {i, 1, Length[elements]}]], "elemental abundances (mass) comparison as areas", Bottom]
```

Out[ ]=


elemental abundances (mass) comparison as areas

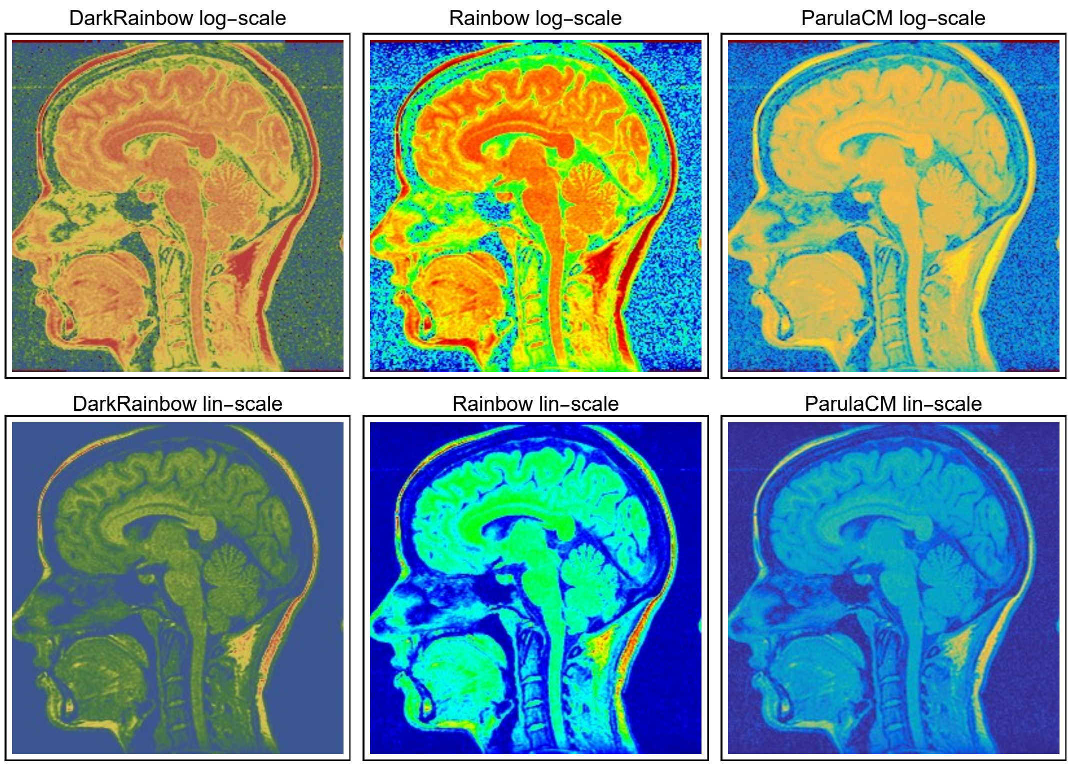
```
In[]:= (*Read in the numerical data*)Get["https://pastebin.com/raw/gN4wGqxe"]
ParulaCM = With[{colorlist = RGBColor @@@ parulaColors}, Blend[colorlist, #] &];
Cube1CM = With[{colorlist = RGBColor @@@ cube1Colors}, Blend[colorlist, #] &];
CubeYFCM = With[{colorlist = RGBColor @@@ cubeYFColors}, Blend[colorlist, #] &];
LinearLCM = With[{colorlist = RGBColor @@@ cube1Colors}, Blend[colorlist, #] &];
JetCM = With[{colorlist = RGBColor @@@ jetColors}, Blend[colorlist, #] &];
```

```
In[]:= braindata = Import["C:\\Users\\roell\\Downloads\\nemamfmr.imagesAB.tar\\DISCIMG\\IMAGES\\BRAINSAG", "DICOM"];
Length[braindata]
```

```
Out[]= 60
```

```
In[]:= LogarithmicScaling[x_, min_, max_] := Log[x / min] / Log[max / min]
```

Out[]=



```
In[ ]:= dat = {{"Alabama", "AL", "Montana", "MT"}, {"Alaska", "AK", "Nebraska", "NE"},
    {"Arizona", "AZ", "Nevada", "NV"}, {"Arkansas", "AR", "New Hampshire", "NH"}, {"California", "CA", "New Jersey", "NJ"},
    {"Colorado", "CO", "New Mexico", "NM"}, {"Connecticut", "CT", "New York", "NY"}, {"Delaware", "DE", "North Carolina", "NC"},
    {"Florida", "FL", "North Dakota", "ND"}, {"Georgia", "GA", "Ohio", "OH"}, {"Hawaii", "HI", "Oklahoma", "OK"},
    {"Idaho", "ID", "Oregon", "OR"}, {"Illinois", "IL", "Pennsylvania", "PA"}, {"Indiana", "IN", "Rhode Island", "RI"},
    {"Iowa", "IA", "South Carolina", "SC"}, {"Kansas", "KS", "South Dakota", "SD"}, {"Kentucky", "KY", "Tennessee", "TN"},
    {"Louisiana", "LA", "Texas", "TX"}, {"Maine", "ME", "Utah", "UT"}, {"Maryland", "MD", "Vermont", "VT"},
    {"Massachusetts", "MA", "Virginia", "VA"}, {"Michigan", "MI", "Washington", "WA"},
    {"Minnesota", "MN", "West Virginia", "WV"}, {"Mississippi", "MS", "Wisconsin", "WI"}, {"Missouri", "MO", "Wyoming", "WY"}};

In[ ]:= abbreviations = Flatten[(# /. {a_, b_, c_, d_} :> {b → a, d → c}) & /@ dat];
```