VSB TECHNICAL UNIVERSITY OF OSTRAVA | FACULTY OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE | DEPARTMENT OF COMPUTER SCIENCE

Data Visualization

460-4120

Fall 2024

Last update 25. 9. 2024

Data Visualization

- Lecturer
  - Tomas Fabian
  - Office

    room EA408, building of FEECS
  - Office hours

    Tuesday 13:00 – 15:00 (all other office hours are by appointment)
  - Email

    tomas.fabian@vsb.cz
  - Web site with additional materials

    http://mrl.cs.vsb.cz/people/fabian/vd_course.html

Course Targets and Goals

- Getting acquainted with advanced techniques of visualization of scientific and technical data (spatial and non-spatial data)
- Application of selected techniques to measured or simulated data
- Correct interpretation of visualization results
- Mastering selected tools for graphing, data visualization, and infographics

Course Prerequisites

- Basics of programming (e.g. C++, Python, JavaScript)
- Previous courses (definitely not compulsory, just good to have):
  - Fundamentals of Computer Graphics (ZPG)
  - Digital Image Processing (DZO)
  - Computer Graphics I (PG1)
- To be familiar with basic concepts of mathematical analysis, linear algebra, vector calculus, and statistics

Main Topics

- Methods of data representation
- Reconstruction methods of the signal (regular grid, scattered data)
- Mapping data on a color scale
- Charts and their visualization
- Visualization of scalar quantities (1D, 2D basic charts, height maps)
- Visualization of vector fields (arrows, streamlines, stream ribbon, LIC, etc.)
- Tensor data and their visualization
- Visualization of volumetric data
- Flow visualization
- Isosurfaces and isolines
- Illustrative visualization
- Visualization using virtual and augmented reality (Oculus Rift, HTC Vive, MS HoloLens)
- Graphic Engines and Tools

Organization of Semester and Grading

- Each lecture will discuss one main topic
- Given topic will be practically realized during the following exercise
- Typical exercise scenario: data from some domain → design of interpretation and visualization → realization → presentation
- The individual tasks from the exercise will be scored (see the list on the web site)
- You can earn up to 45 points in total (until the last week of semester)
- Final written test covering topics from the previous slide with subsequent discussion (credit week (pre-term) or during the exam period (TBA))
- You can earn up to 55 points

Study Materials

- Chun-houh Chen, Wolfgang Härdle, Antony Unwin, ***Handbook of Data Visualization***, ISBN: 978-3-540-33036-3, 936 pages, Springer, 2008.
- Alexandru C. Telea, ***Data Visualization: Principles and Practice***, Second Edition, ISBN: 978-1466585263, 617 pages, AK Peters, 2014.
- Charles D. Hansen and Chris Johnson, *The Visualization Handbook*, ISBN: 0-12-387582-x, 984 pages, Elsevier, 2004.
- Tamara Munzner, *Visualization Analysis and Design* , ISBN: 978-1466508910, 428 pages, AK Peters, 2014.
- Casey Reas and Ben Fry, *Processing: A Programming Handbook for Visual Designers and Artists*, ISBN: 978-0262182621, 712 pages, MIT Press, 2007.
- Stephen Few, *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, Second Edition, ISBN: 978-0970601971, 371 pages, Analytics Press, 2012.
- Claus O. Wilke, ***Fundamentals of data visualization: A primer on making informative and compelling figures***. O'Reilly Media, 2019. (online)
- Colin Ware, *Information Visualization: Perception for Design (Interactive Technologies)*, Fourth edition, 560 pages, Morgan Kaufmann, ISBN: 978-0128128756, 2024.

Data Visualization and Theory

- In general, graphics provide an excellent approach for exploring data
- Graphics have been used extensively in statistics for a long time
- There is not a substantive body of theory about the topic
- Knowledge is expressed in principles to be followed,

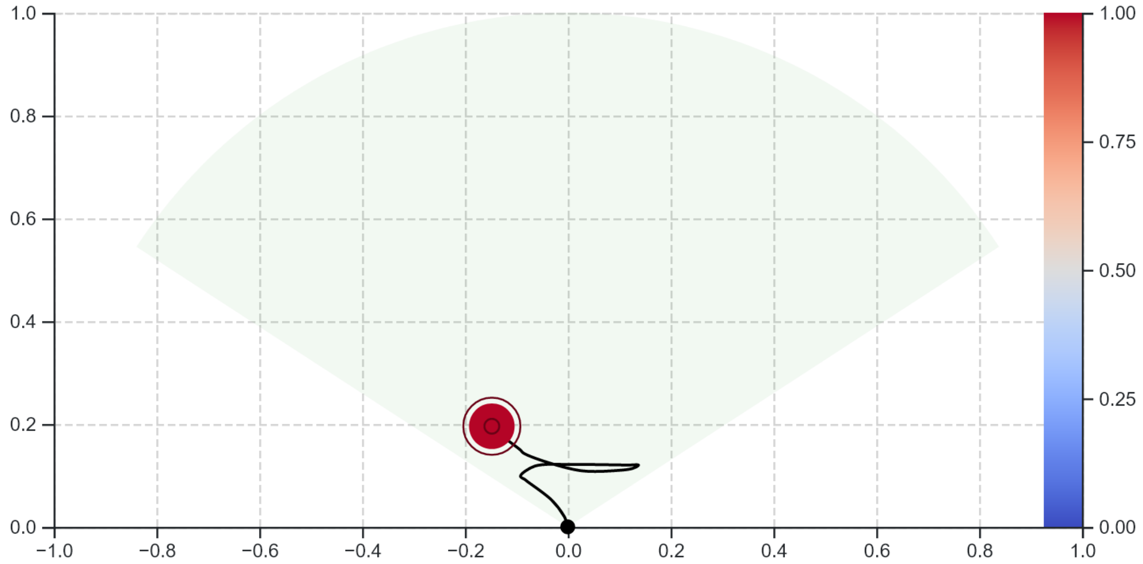
    e.g. Edward Tufte, *The visual display of quantitative information* (2001).

Presentation and Exploratory Graphics

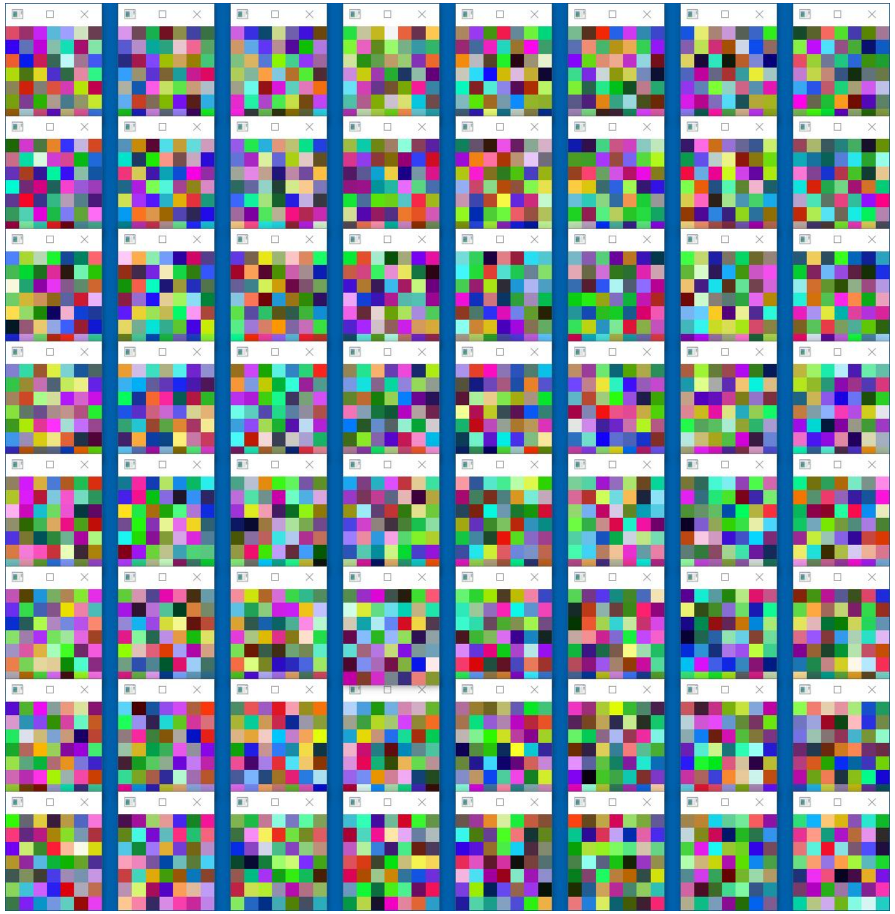
- *Presentation graphics* are like proofs of mathematical theorems
  - They should be of high (i.e. publication) quality and include complete definitions and explanations of the variables
  - They may give no hint as to how a result was reached
  - They should offer convincing support for its conclusion
- *Exploratory graphics* are used for looking for results
  - They should be fast and informative rather than slow and precise
  - They are not intended for presentation, so that detailed legends and captions are unnecessary

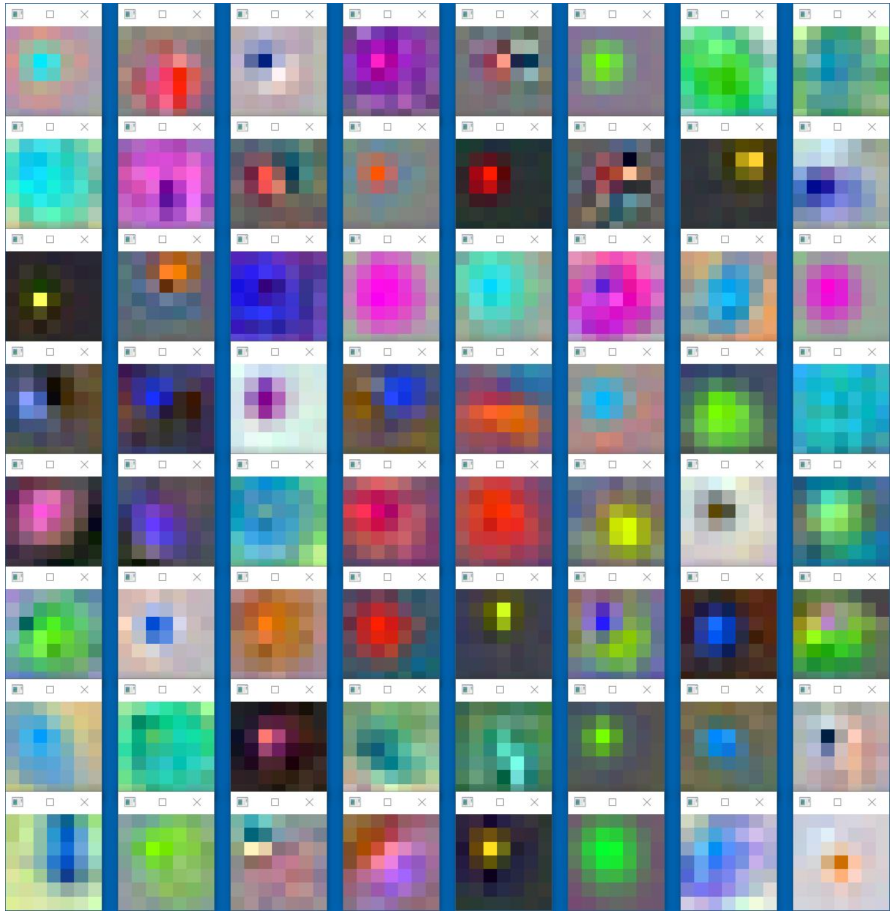
Handbook of Data Visualization, p. 5

Presentation Graphics Example

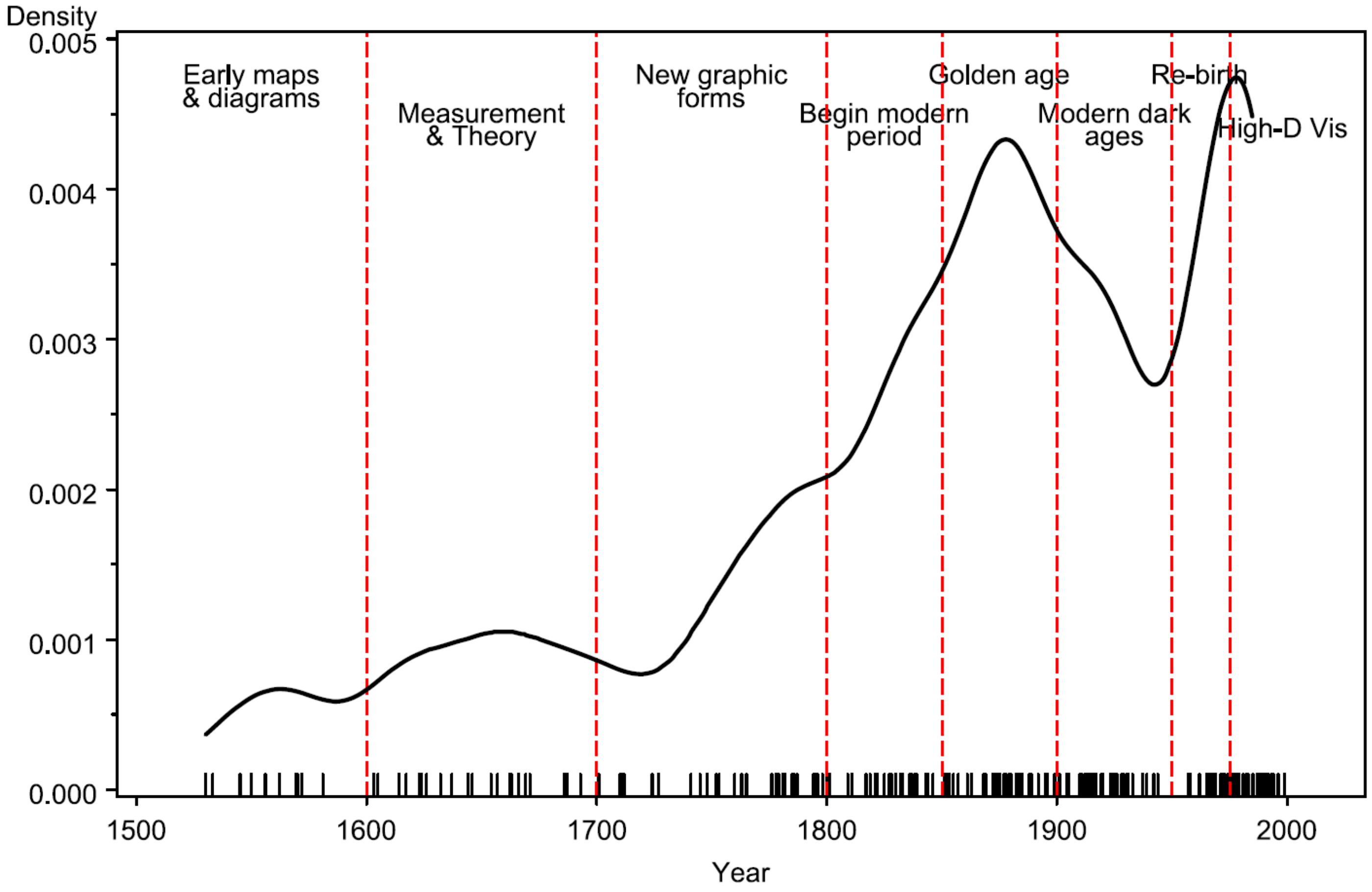
#	Typ	Otázka	Odpověď
1	PDS	Jak blízký vztah máš k informatice?	 Tah č. 1 o délce 0,69 proveden za 4,85 s. Konečná poloha (-0,15, 0,20) ve vzdálenosti 0,25 a směru -37,0° potvrzena za 17,63 s. Velikost 2,99 nastavena za 9,90 s. Teplota 1,00 nastavena za 8,22 s. Celkový čas strávený nad otázkou 35,76 s.

Exploratory Graphics Example





Milestones in the History: Developments



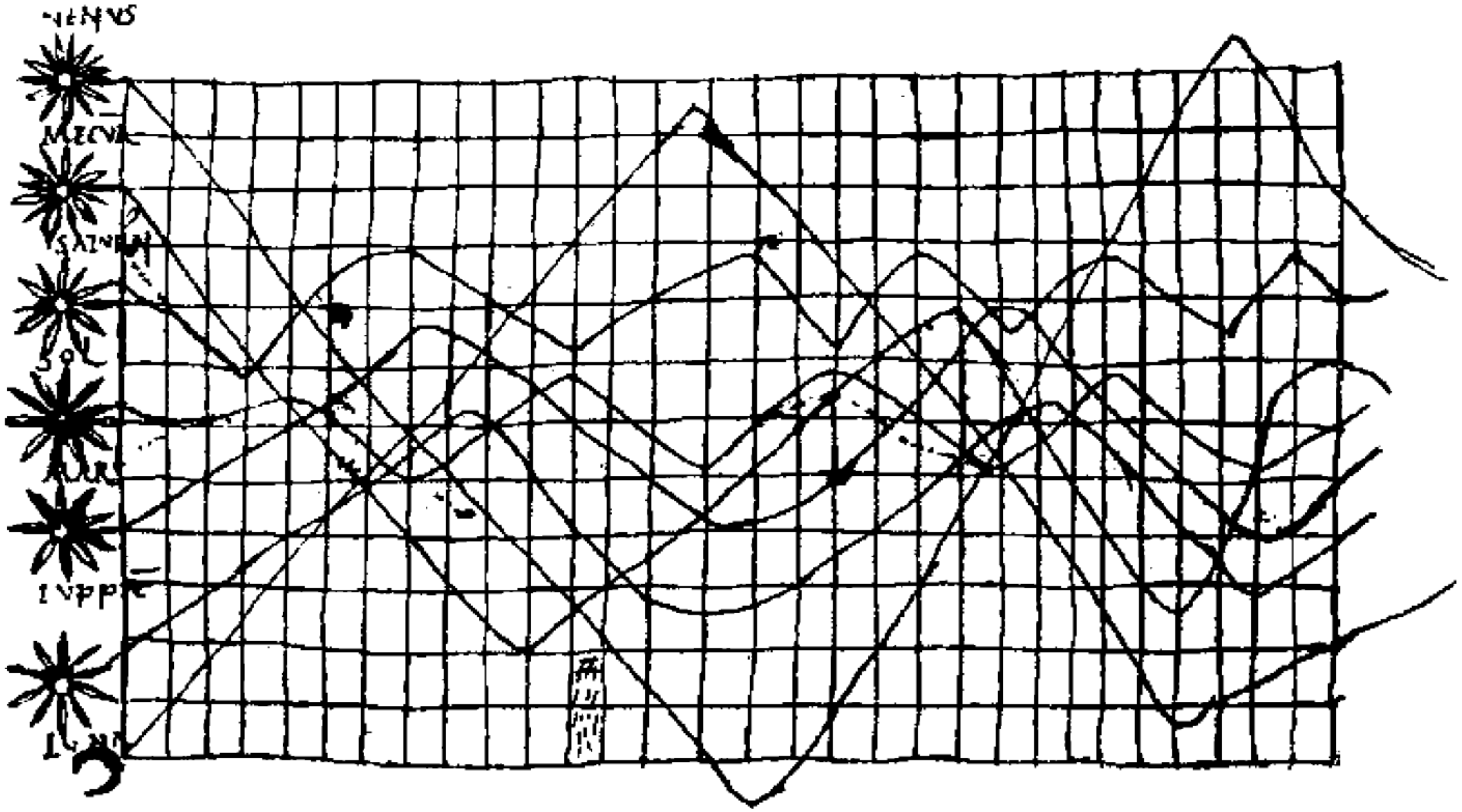
Time distribution of events considered milestones in the history of data visualization

Handbook of Data Visualization, p. 18

Milestones in the History: Maps and Proto Graphs

- Positions of stars and other celestial bodies
- Ancient Egyptian – idea of coordinates (200 B.C.)
- 10th century – depiction of quantitative information
- 14th century – plotting a theoretical function
- 16th century – development of techniques and instruments for precise observation and measurement lead to the beginning of data visualization
- 17th century – the rise of analytic geometry and coordinate systems, theories of errors of measurement and estimation, the birth of probability theory, the beginnings of demographic statistics, the dawn of practical application

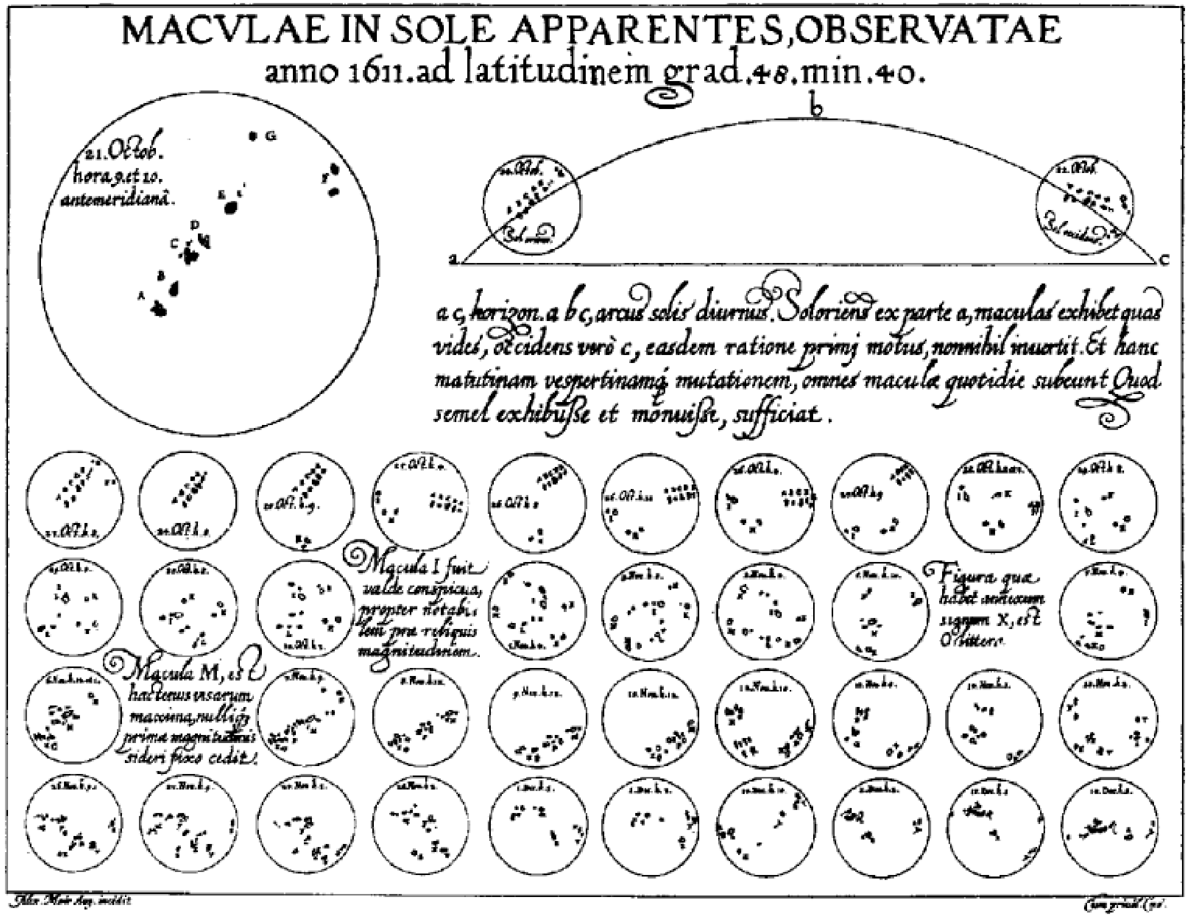
Milestones in the History: Multiple time-series graph



Planetary movements shown as cyclic inclinations over time, by an unknown astronomer, appearing in a 10th-century

Funkhouser (1936, p. 261)

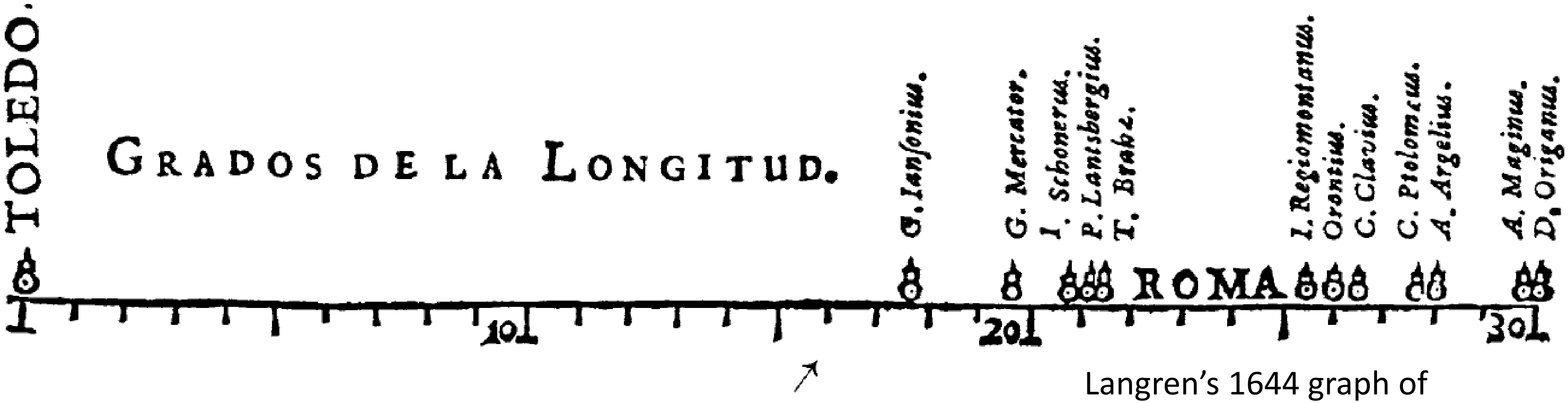
Milestones in the History: Small Multiples



Scheiner’s 1626 representation of the changes in sunspots over time. It uses the principle of "small multiples" - multiple images depict the recordings of sunspots over two months

Scheiner (1626 – 1630)

Milestones in the History: Small Multiples



Langren’s 1644 graph of determinations of the distance, in longitude, from Toledo to Rome.

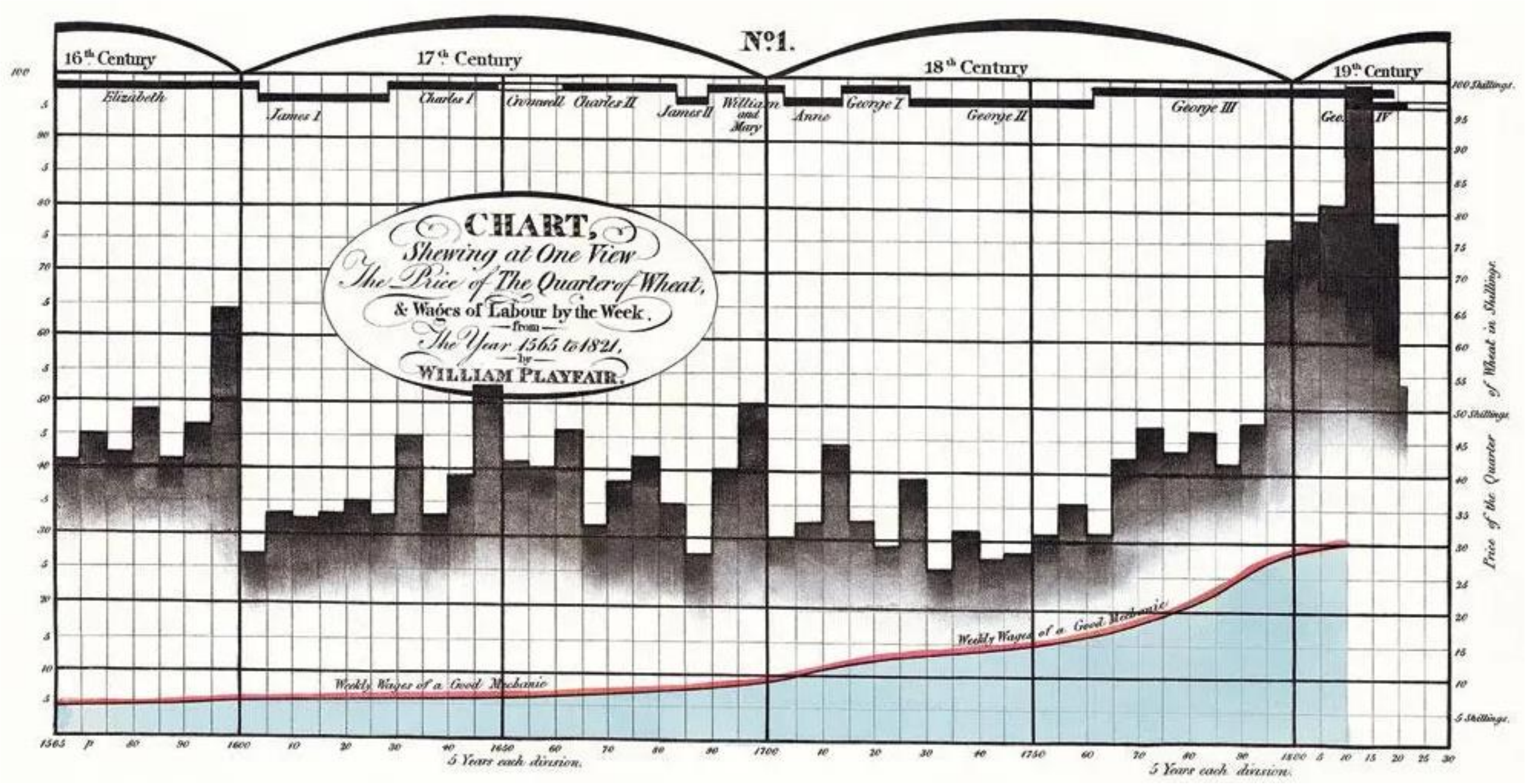
The correct distance is 16°30'.

Tute (1997, p. 15)

Milestones in the History: New Graphic Forms

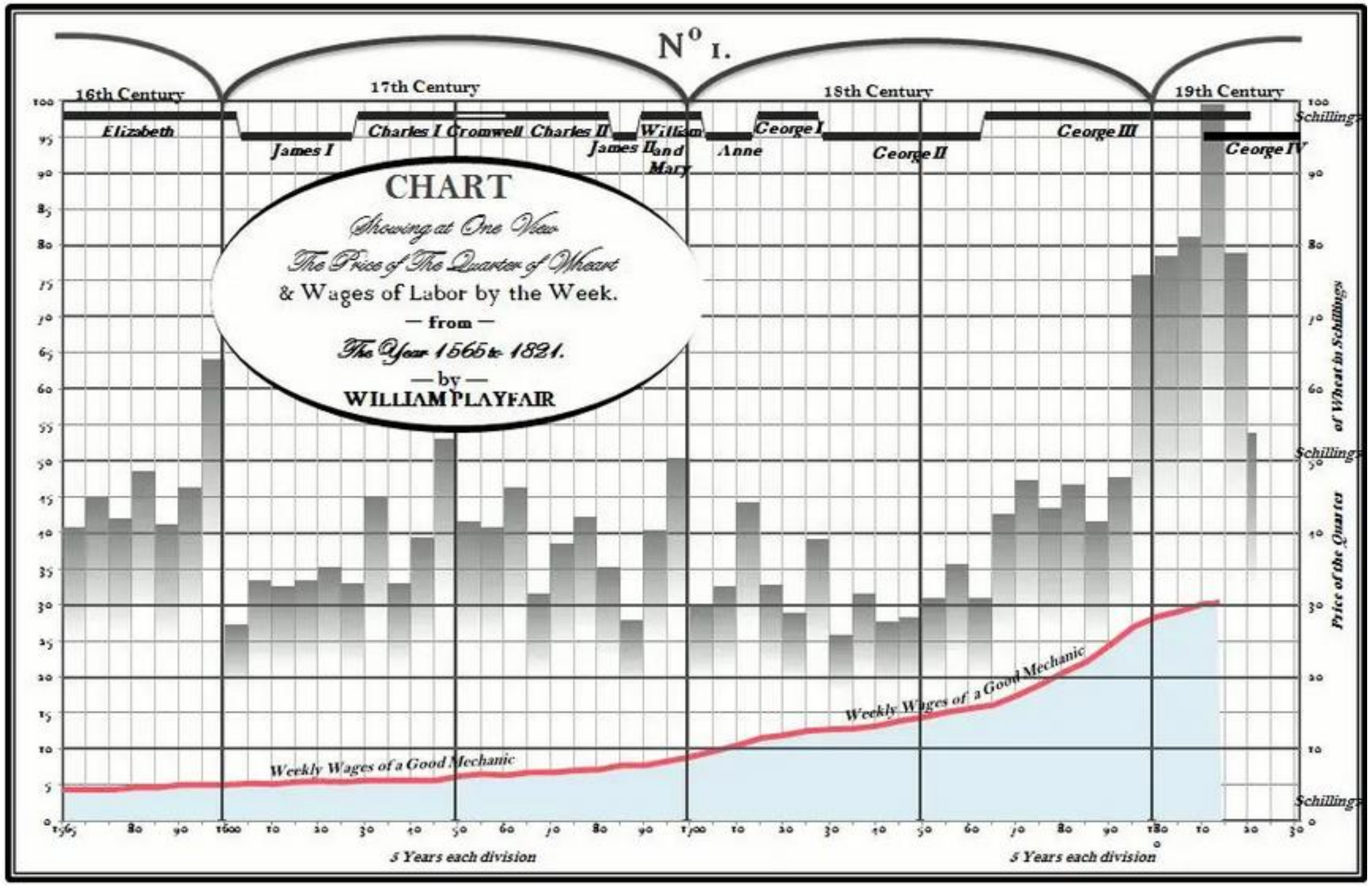
- 18th century – dawn of isolines and contours (cartography, magnetic field), timelines (e.g. timeline chart of biography of 2000 famous people), visual encoding of quantity (squares to compare the areas of states), three and four-color printing (subtractive RYBK similar to modern CMYK), lithography
- Ideas of curve fitting and interpolation (Lambert)
- William Playfair (1759-1823) – inventor of most of the graphical forms used today – line graph, bar chart, pie chart, and circle graph

Milestones in the History: New Graphic Forms



One of Playfair's most iconic images, "Chart Showing at One View the Price of the Quarter of Wheat, & Wages of Labour

Milestones in the History: New Graphic Forms



The same chart recreated in Microsoft Excel by Jorge Camoes

Milestones in the History: Iso Lines

First use of isolines to show contours of equal value (magnetic declination) on a coordinate grid.

Edmund Halley *(1701)*

Milestones in the History: Beginnings of Modern Graphics

- 19th century – explosive growth in statistical graphics and thematic mapping
- Histograms, line graphs, contour plots, scatter plots etc.
- Thematic mapping progressed from single maps to comprehensive atlases
- Use of continuous shadings (from white to black) to show the distribution

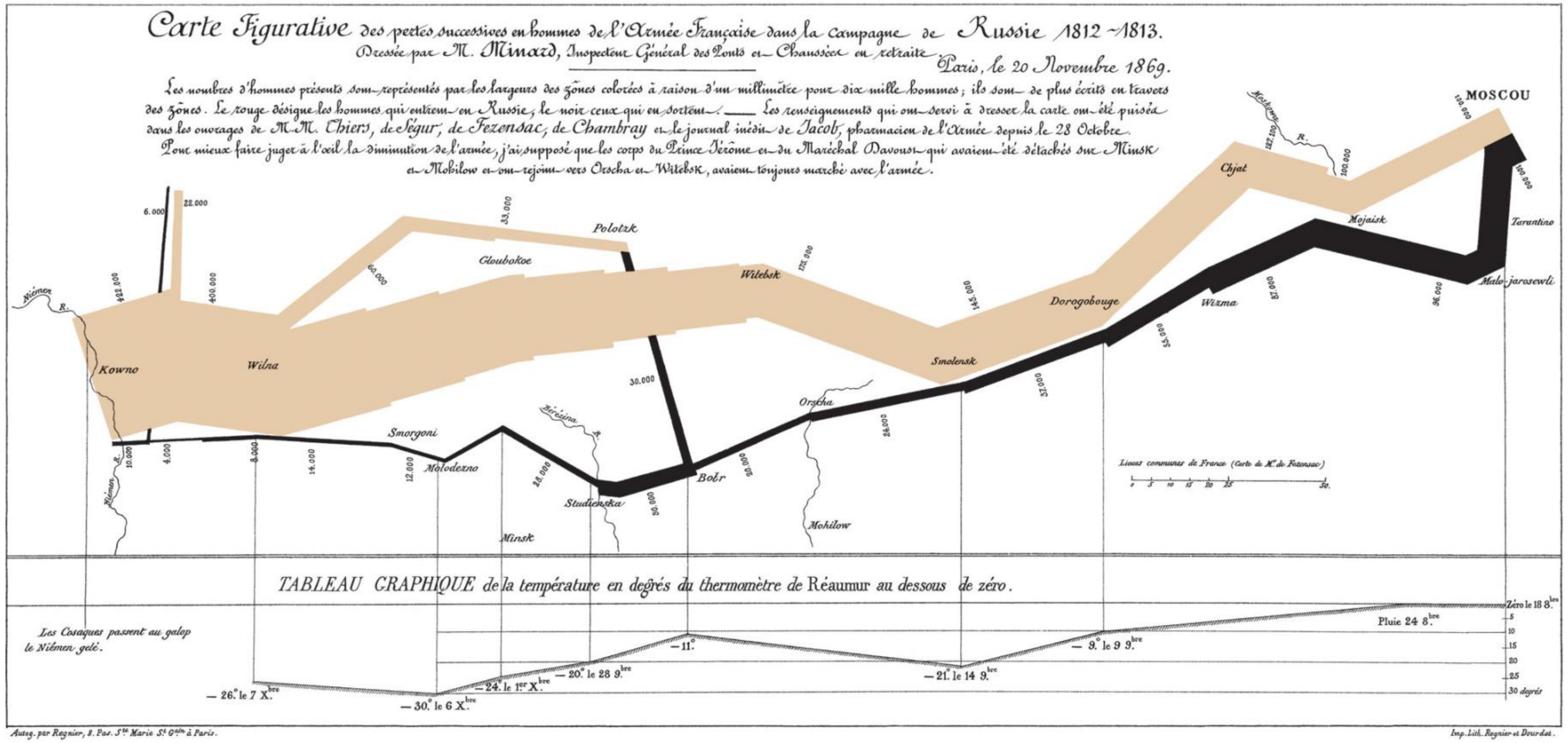
Milestones in the History: The Last 100 Years

- 1850-1900 The Golden Age of Statistical Graphics
  - 3D surface plots (both axonometric and stereograms)
  - Contour diagrams, showing isolevel curves of 3D surfaces
  - Flow lines of width proportional to quantities (Minard)
- 1900–1950: The Modern Dark Ages
  - There were few graphical innovations
  - A time of necessary dormancy, application and popularization rather than one of innovation
  - Statistical graphics became mainstream

Milestones in the History: The Last 100 Years

- 1950–1975: Rebirth of Data Visualization
  - Exploratory data analysis (EDA) – stem-leaf plots, boxplots,
- 1975–present: High-D, Interactive and Dynamic Data Visualization
  - New methods for visualizing high-dimensional data, scatterplot matrix, parallel coordinates plot, spreadplots (distribution dot plots)
  - Graphical techniques for discrete and categorical data
  - Increased attention to the cognitive and perceptual aspects of data display
  - Statistical and graphics software

Milestones in the History: Flow map



Minard's flow map graphic of Napoleon's March on Moscow (also called "the best graphic ever produced")

Charles Joseph Minard (1781-1870)

Map portrays the losses suffered by Napoleon's army in the Russian campaign of 1812. Beginning at the Polish-Russian border, the thick band shows the size of the army at each position. The path of Napoleon's retreat from Moscow in the bitterly cold winter is depicted by the dark lower band, which is tied to temperature and time scales.

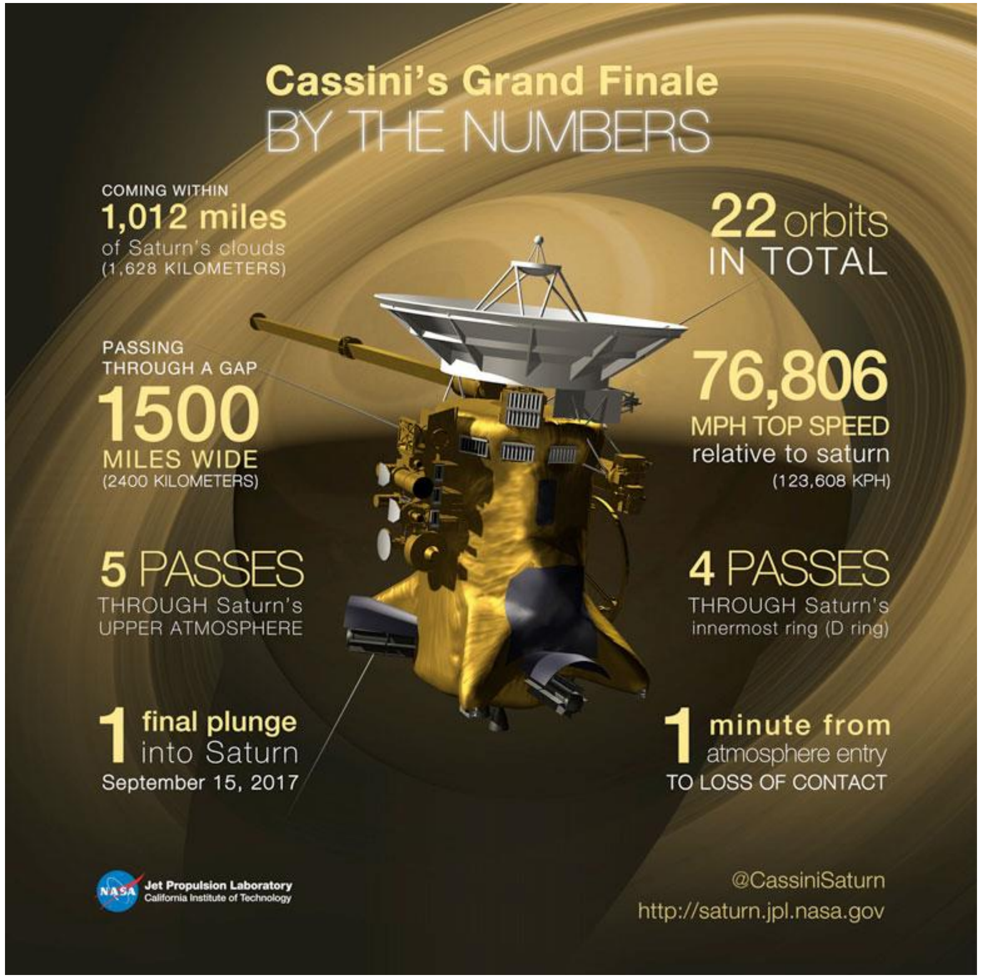
Milestones in the History

- A summary of milestones in the history of thematic cartography, statistical graphics, and data visualization

- An illustrated chronology of innovations by Michael Friendly and Daniel J. Denis

- http://euclid.psych.yorku.ca/SCS/Gallery/milestone/

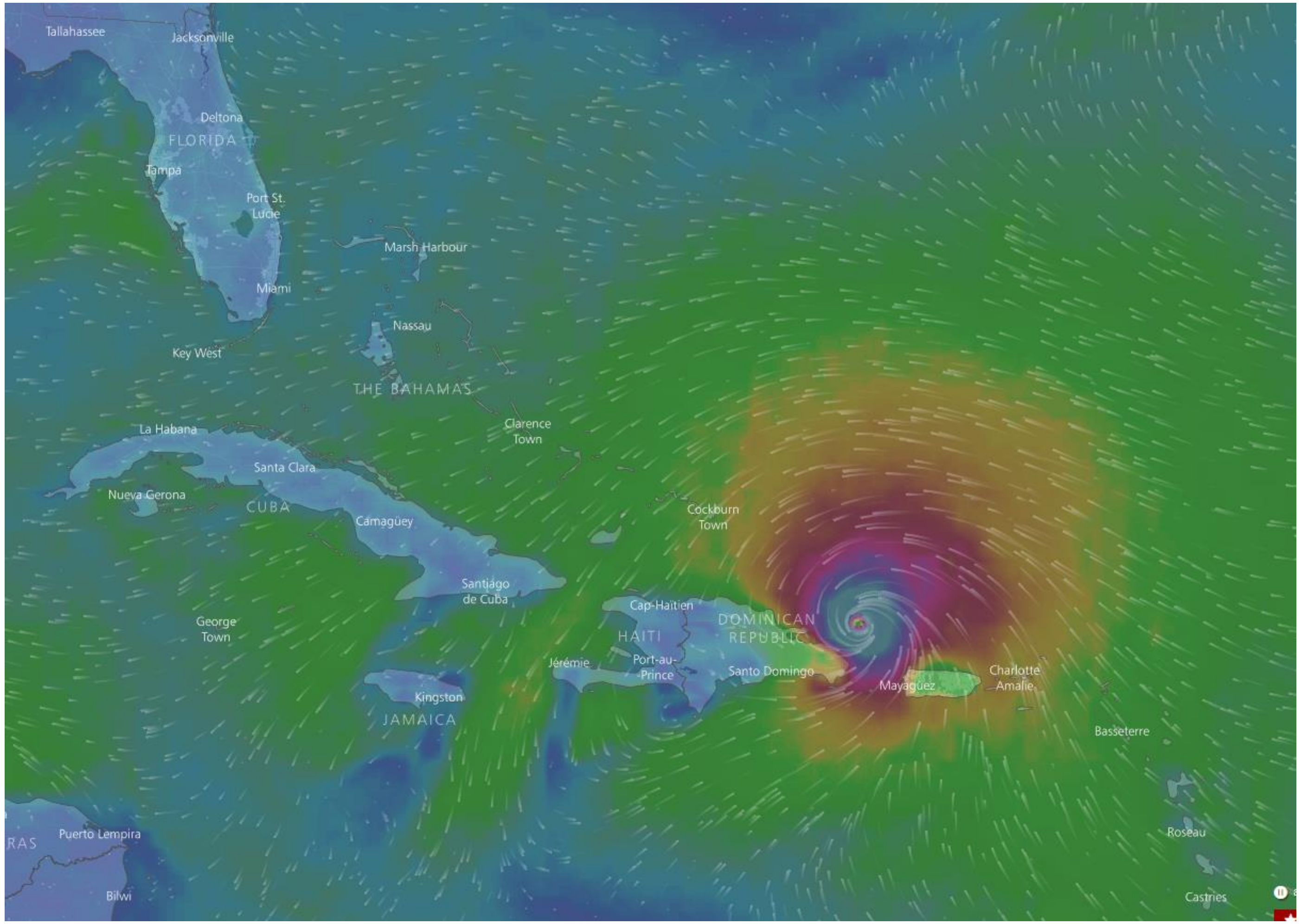
Examples



Cassini Grand Finale Dive
(Illustration)

www.jpl.nasa.gov

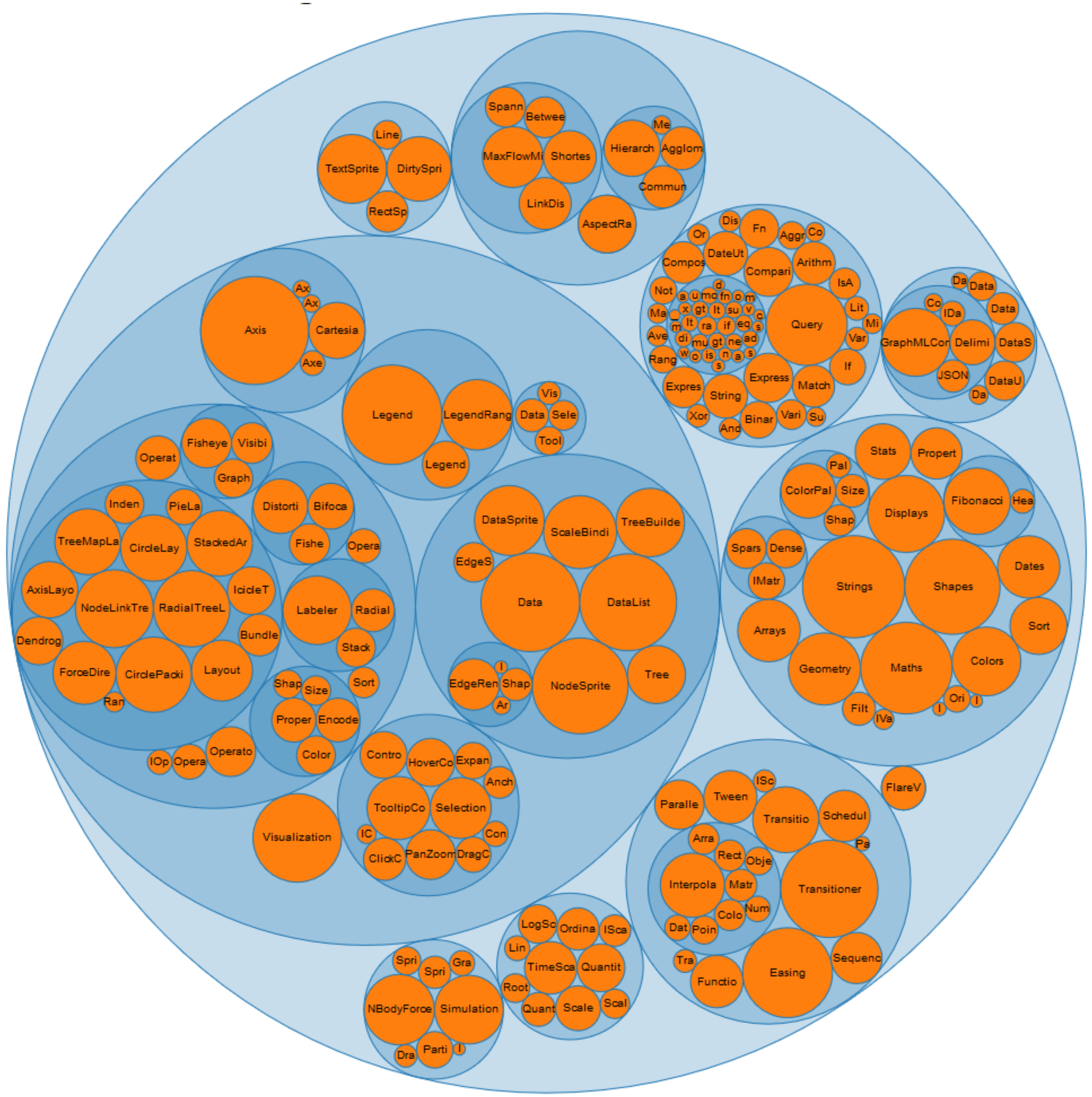
Examples



Wind map & weather forecast

www.windy.com

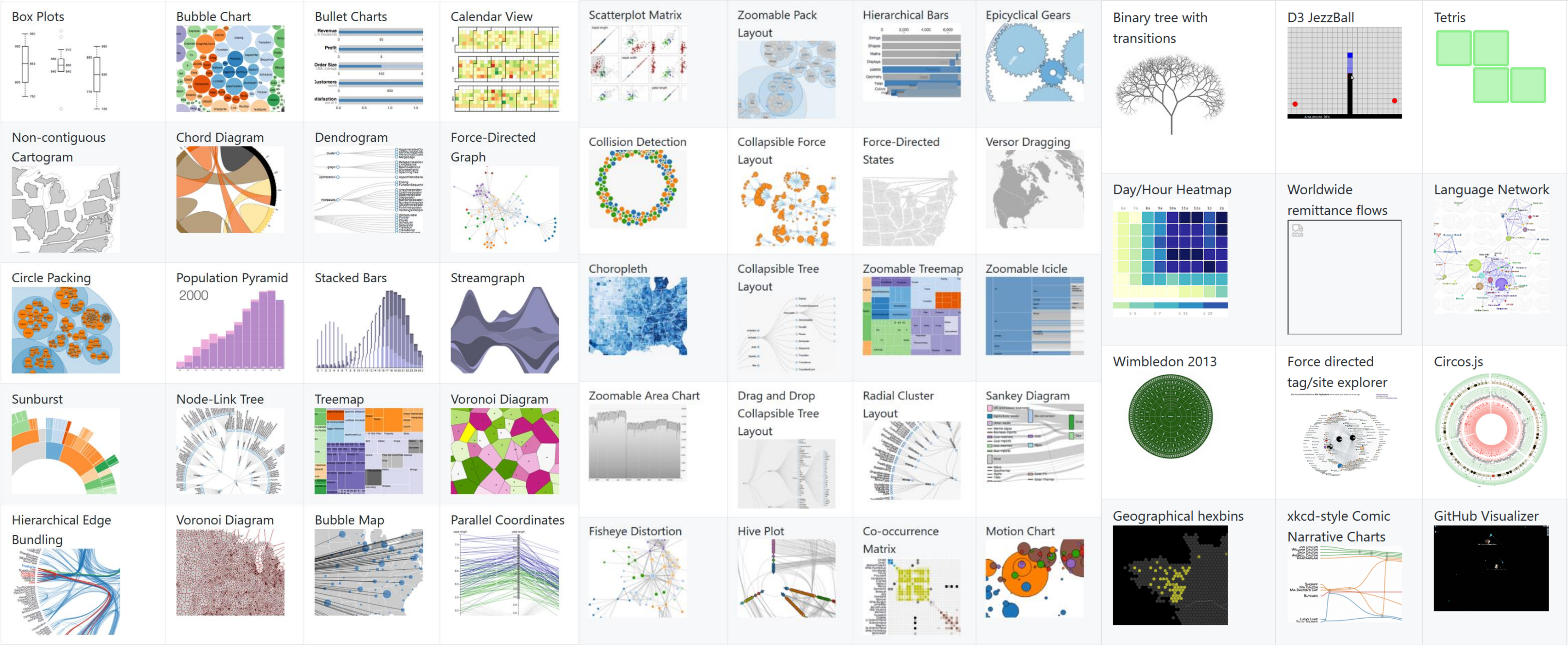
Examples



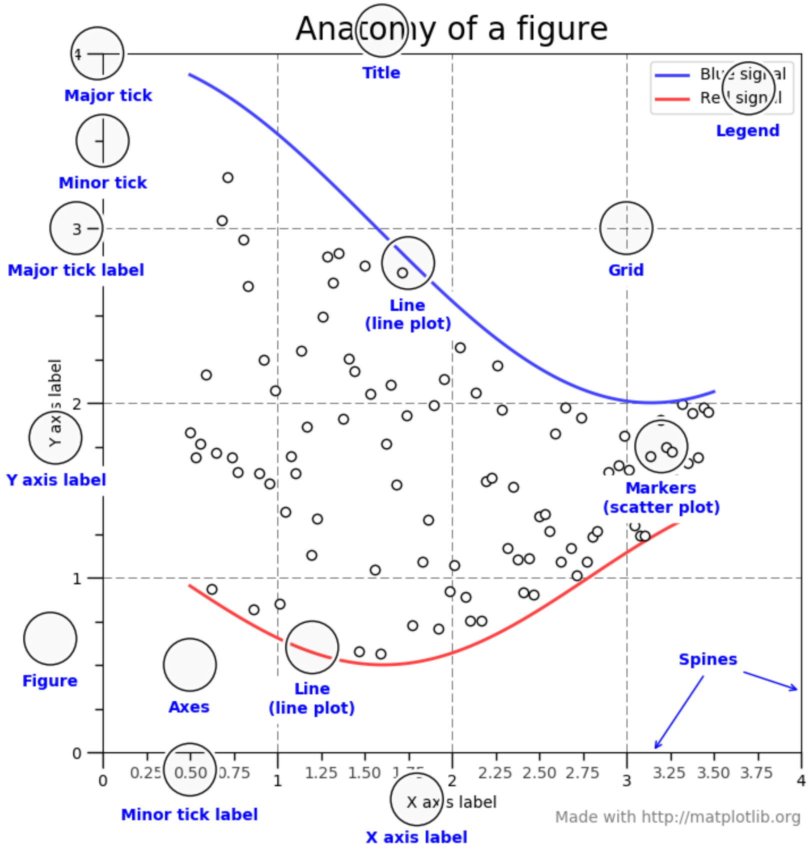
Enclosure diagrams use containment to represent the hierarchy

bl.ocks.org/mbostock/4063530

Examples


Box Plots
Bubble Chart
Bullet Charts
Calendar View
Non-contiguous Cartogram
Chord Diagram
Dendrogram
Force-Directed Graph
Circle Packing
Population Pyramid
2000
Stacked Bars
Streamgraph
Sunburst
Node-Link Tree
Treemap
Voronoi Diagram
Hierarchical Edge Bundling
Voronoi Diagram
Bubble Map
Parallel Coordinates
Scatterplot Matrix
Zoomable Pack Layout
Hierarchical Bars
Epicyclical Gears
Collision Detection
Collapsible Force Layout
Force-Directed States
Versor Dragging
Choropleth
Collapsible Tree Layout
Zoomable Treemap
Zoomable Icicle
Zoomable Area Chart
Drag and Drop Collapsible Tree Layout
Radial Cluster Layout
Sankey Diagram
Fisheye Distortion
Hive Plot
Co-occurrence Matrix
Motion Chart
Binary tree with transitions
D3 JezzBall
Tetris
Day/Hour Heatmap
Worldwide remittance flows
Language Network
Wimbledon 2013
Force directed tag/site explorer
Circos.js
Geographical hexbins
xkcd-style Comic Narrative Charts
GitHub Visualizer

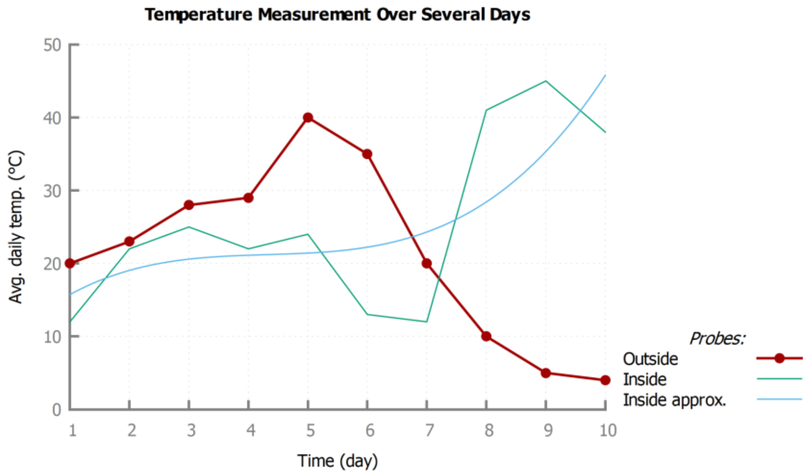
Examples



Anatomy of a figure
Title
Blue signal
Red signal
Legend
Major tick
Minor tick
Major tick label
Grid
Line
(line plot)
Y axis label
Markers
(scatter plot)
Figure
Axes
Line
(line plot)
Spines
Minor tick label
X axis label
Made with http://matplotlib.org

Excercise

- Gnuplot (see www.gnuplot.info) - portable command-line driven graphing utility for interactive visualization of mathematical functions and data
- You can find scripts with a few examples on my website

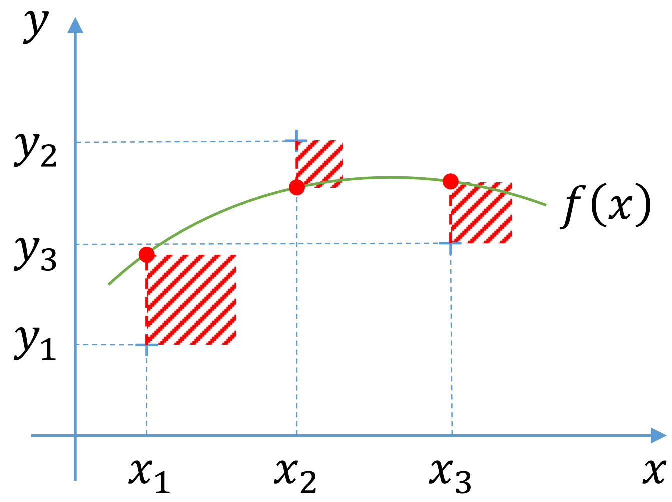


- Try to create your own time series visualization (e.g. temperature datasets) based on this example

Quadratic Regression

- Given data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ and quadratic function $f(x_i) = ax_i^2 + bx_i + c$
- The goal is to find parameters a, b, c that minimize the error E

$$E(a, b, c) = \sum_{i=1}^{n} (y_i - f(x_i))^2$$



- Note that the cost function of polynomial regression can also be taken to be mean squared error $E = 1/n \sum_{i=1}^{n} (y_i - f(x_i))^2$

In general, (local) maxima and **minima** of a continuous and differentiable function E is where $\frac{\partial E}{\partial a} = 0, \frac{\partial E}{\partial b} = 0, \frac{\partial E}{\partial c} = 0$

Quadratic Regression (cont.)

$$\frac{\partial E}{\partial a} = \sum_{i=1}^{n} 2\big(y_i - f(x_i)\big)\left(-x_i^2\right) = \sum x_i^4\, a + \sum x_i^3\, b + \sum x_i^2\, c - \sum x_i^2 y_i = 0$$

$$\frac{\partial E}{\partial b} = \sum_{i=1}^{n} 2\big(y_i - f(x_i)\big)\left(-x_i\right) = \sum x_i^3\, a + \sum x_i^2\, b + \sum x_i\, c - \sum x_i y_i = 0$$

$$\frac{\partial E}{\partial c} = \sum_{i=1}^{n} 2\big(y_i - f(x_i)\big)(-1) = \sum x_i^2\, a + \sum x_i\, b + \sum 1\, c - \sum y_i = 0$$

- Turn the system of equations into a matrix form … … and solve for a, b, c

$$M = \begin{bmatrix} \sum x_i^4 & \sum x_i^3 & \sum x_i^2 \\ \sum x_i^3 & \boxed{\sum x_i^2} & \boxed{\sum x_i} \\ \sum x_i^2 & \boxed{\sum x_i} & \boxed{n} \end{bmatrix} \begin{bmatrix} a \\ \boxed{b} \\ \boxed{c} \end{bmatrix} = \begin{bmatrix} \sum x_i^2 y_i \\ \boxed{\sum x_i y_i} \\ \boxed{\sum y_i} \end{bmatrix}$$

$$\overbrace{M^{-1} M}^{I} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = M^{-1} \begin{bmatrix} \sum x_i^2 y_i \\ \sum x_i y_i \\ \sum y_i \end{bmatrix}$$

Linear regression with $f(x_i) = bx_i + c$

M is invertibe iff $\det(M) \neq 0$

Weighted Sum of Squared Residuals

- Modified cost function

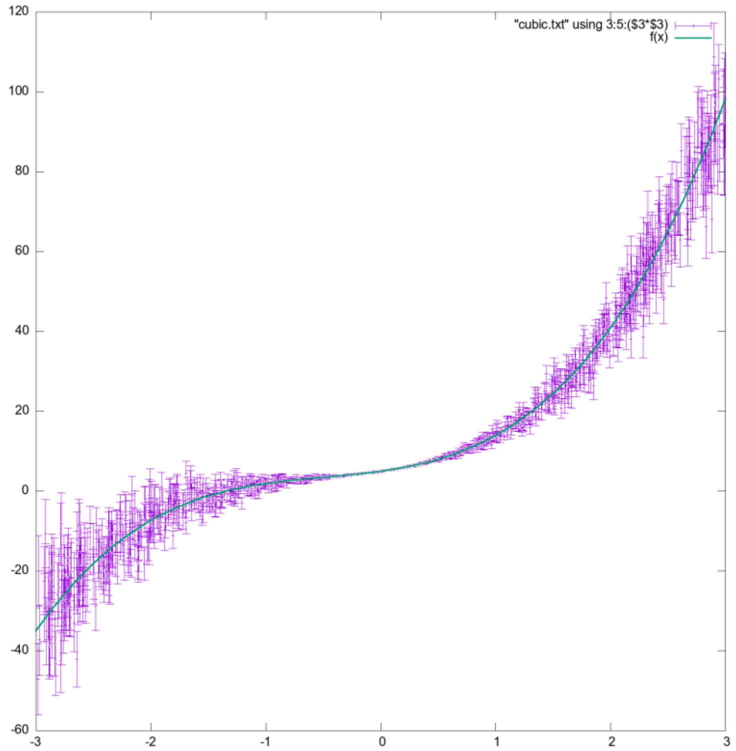
$$E(a, b, c) = \sum_{i=1}^{n} w_i (y_i - f(x_i))^2$$

- Weight parameter w_i equals to inverse of sigma square

$$w_i = \frac{1}{\sigma_i^2}$$

- Gnuplot script

```
f(x) = a*x**3 + b*x**2 + c*x + d
fit f(x) "cubic.txt" using 3:5:($3*$3) yerrors via a,b,c,d
plot "cubic.txt" using 3:5:($3*$3) smooth unique with yerror, f(x) with lines lw 3
```



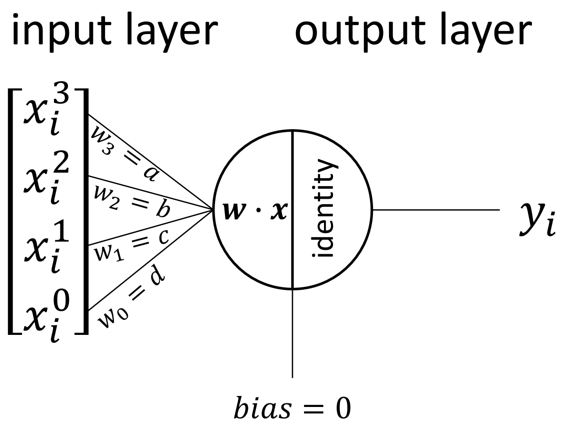
Technical Note About Normalization

- How to normalize x-axis for fitting in gnuplot

```
stats "data.txt" index 0 using 1:2 prefix "A"
f(x) = a*x**2 + b*x + c
fit f(x) "data.txt" index 0 using (($1)-A_min_x)/(A_max_x-A_min_x):2 via a,b,c
```

Cubic Regression with NN

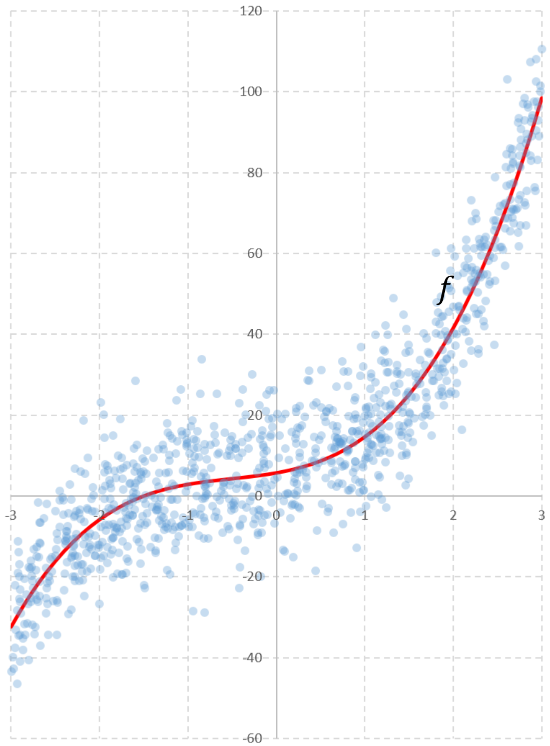
- Given data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ and cubic function $f(x_i) = ax_i^3 + bx_i^2 + cx_i + d$, we want to find optimal values for these four parameters via neural network optimization




```
Layer (type)              Output Shape            Param #
===========================================================
output (Dense)            (None, 1)               4
===========================================================
Trainable params: 4
___________________________________________________________
Features (inputs):
          x3         x2        x1   x0
0   0.077512   0.181794  0.426374   1
1   0.008647   0.042128  0.205251   1
2 -19.596714   7.268679 -2.696049   1

Labels (outputs):
         y
0   7.405902
1   5.964680
2 -23.171585

Epoch 1/100, loss: 1464.6207
...
Epoch 100/100, loss: 99.0606
```



final model params = [2.0186083, 3.0940003, 3.7165887, 5.4477468]