

故事中的统计学

天津医科大学
生物医学工程与技术学院

2016-2017 学年下学期 (春)
公共选修课

第三章 眼见未必为实的图形

伊现富 (Yi Xianfu)

天津医科大学 (TJMU)
生物医学工程与技术学院

2017 年 5 月



1

统计图形

- 引言
- 简介
- 著名设计者

2

知识拓展

- 埃舍尔
- 眼见未必为实

3

图说天下

1 统计图形

- 引言
- 简介
- 著名设计者

2 知识拓展

- 埃舍尔
- 眼见未必为实

3 图说天下



1

统计图形

- 引言
- 简介
- 著名设计者

2

知识拓展

- 埃舍尔
- 眼见未必为实

3

图说天下



统计数据分析

统计学与数据分析过程可大致分为两个组成部分：定量分析方法（Quantitative techniques）和图解分析方法（graphical techniques）。

定量分析方法

定量分析方法是指那套产生数值型或表格型输出的统计学操作程序；比如，包括假设检验、方差分析、点估计（point estimation）、可信区间以及最小二乘法回归分析。

这些手段以及与此类似的其他技术方法全都颇具价值，属于是经典分析方面的主流。



统计数据分析

统计学与数据分析过程可大致分为两个组成部分：定量分析方法（Quantitative techniques）和图解分析方法（graphical techniques）。

定量分析方法

定量分析方法是指那套产生数值型或表格型输出的统计学操作程序；比如，包括假设检验、方差分析、点估计（point estimation）、可信区间以及最小二乘法回归分析。

这些手段以及与此类似的其他技术方法全都颇具价值，属于是经典分析方面的主流。



图解分析方法

统计学与数据分析还有一大套我们一般称之为图解分析方法的统计学工具。这些工具包括散点图（scatter plot）、直方图、概率图（probability plot）、残差图（residual plot）、箱形图（box plot, box-and-whisker plot）、块图（block plot）以及双标图（biplot）。

探索性数据分析（Exploratory data analysis, EDA）就密切地依赖于这些手段以及与此类似的其他技术方法。图解分析操作程序不仅仅是在 EDA 背景下才使用的工具；在检验假设、模型选择（model selection）、统计模型验证、估计量（estimator）选择、关系确定、因素效应判定以及离群值（outlier）检出方面，此类图解分析工具还可以作为最佳捷径，用来深入认识数据集。此外，优质的统计图形还可以作为一种令人信服的沟通手段，用来向他人传达存在于数据之中的基本讯息。如果不采用统计图形，也就会丧失深入认识数据基础结构之一个或多个方面的机会。

图解式统计学方法的目标

- ① 探究数据集的内容
- ② 用于发现数据之中的结构
- ③ 检查统计学模型之中的假设
- ④ 沟通传达分析结果



- 1 简介
 - 引言
 - 简介
 - 著名设计者

- 2 知识拓展
 - 埃舍尔
 - 眼见未必为实
- 3 图说天下



统计图形

统计图形，又称为统计图、统计学图形、图解方法、图解技术、图解分析方法或图解分析技术，是指统计学领域当中用于可视化定量数据的信息图形。有时，人们也把统计图形与各种统计学表格统称为统计图表或统计学图表。



散点图

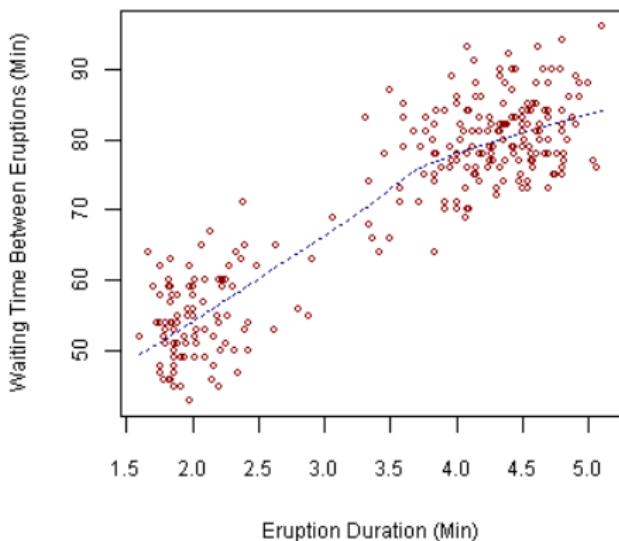
A scatter plot (also called a scatter graph, scatter chart, scattergram, or scatter diagram) is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data.

The data is displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis.

If the points are color-coded, one additional variable can be displayed.



Old Faithful Eruptions



Waiting time between eruptions and the duration of the eruption for the Old Faithful Geyser in Yellowstone National Park, Wyoming, USA. This chart suggests there are generally two “types” of eruptions: short-wait-short-duration, and long-wait-long-duration.

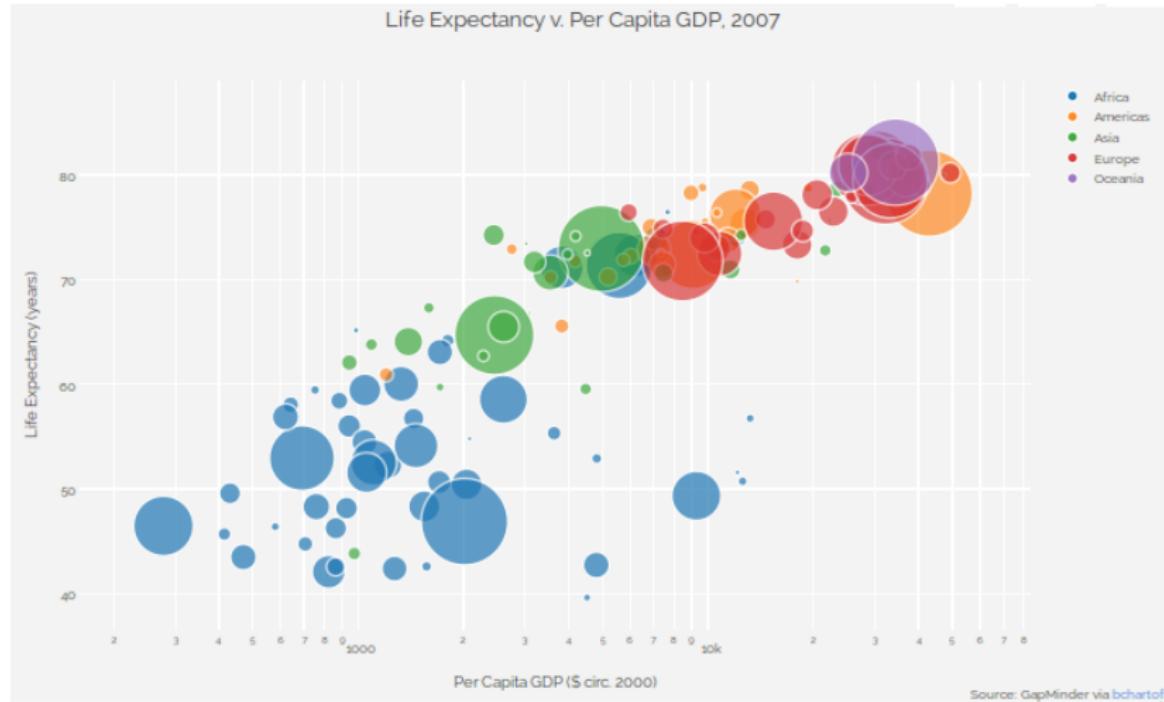


气泡图

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. Bubble charts can facilitate the understanding of social, economical, medical, and other scientific relationships.

Bubble charts can be considered a variation of the scatter plot, in which the data points are replaced with bubbles. As the documentation for Microsoft Office explains, "You can use a bubble chart instead of a scatter chart if your data has three data series that each contain a set of values. The sizes of the bubbles are determined by the values in the third data series.".

统计图形 | 气泡图



折线图

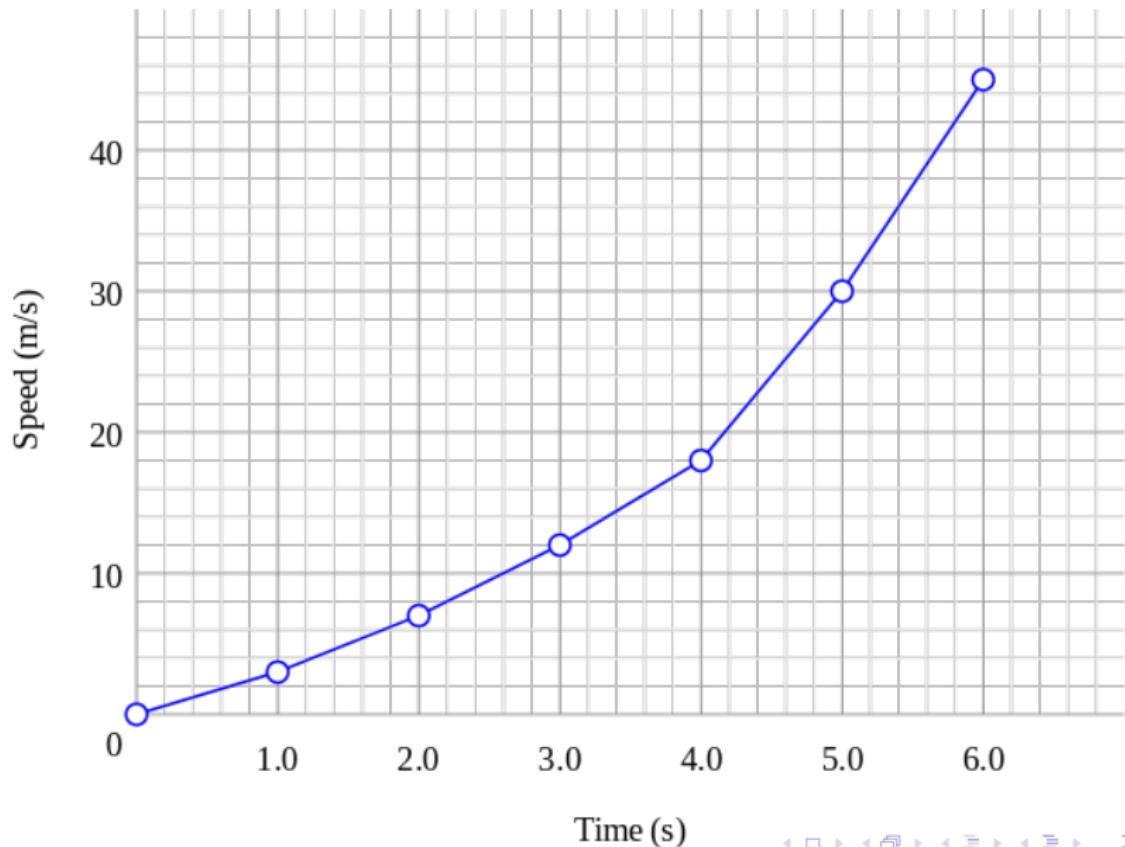
A line chart or line graph is a type of chart which displays information as a series of data points called ‘markers’ connected by straight line segments.

It is a basic type of chart common in many fields. It is similar to a scatter plot except that the measurement points are ordered (typically by their x-axis value) and joined with straight line segments.

A line chart is often used to visualize a trend in data over intervals of time –a time series –thus the line is often drawn chronologically. In these cases they are known as run charts.



统计图形 | 折线图



趋势图

趋势图 (run chart) , 也称运行图、链图、走势图, 是一类在时间序列中表达数据变量的统计图表, 通常该数据在一些工业或商业过程中可用来表达特定程序的表现。

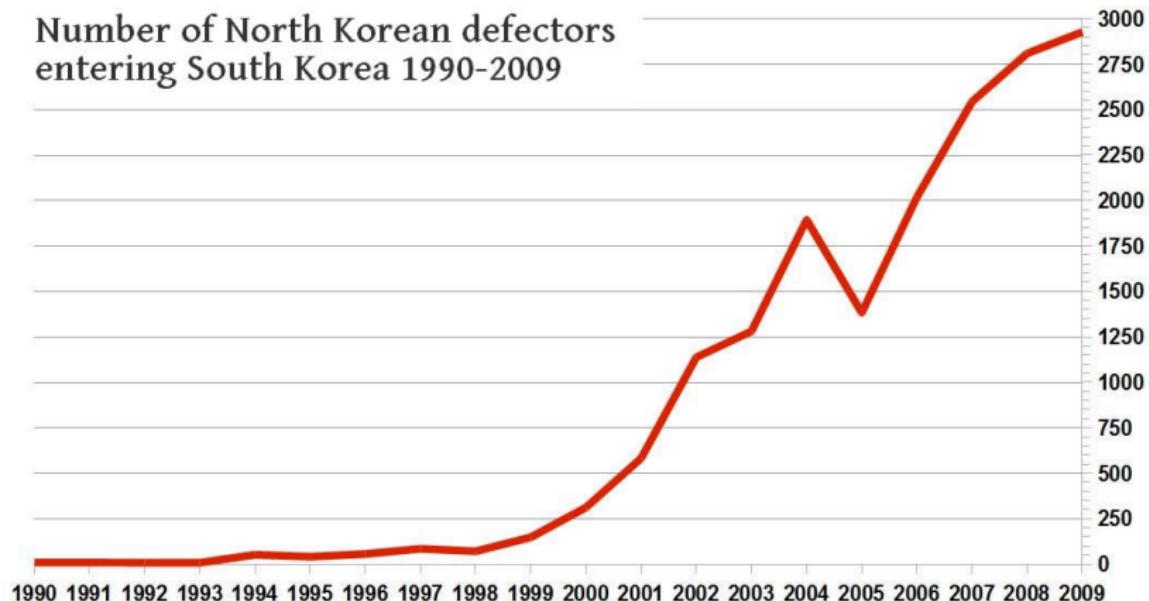
趋势图通常用于发现数据集中的异常现象, 其暗示数据的变化往往是时间及其他特定因素所导致的变异性。

A run chart, also known as a run-sequence plot is a graph that displays observed data in a time sequence. Often, the data displayed represent some aspect of the output or performance of a manufacturing or other business process. It is therefore a form of line chart.



统计图形 | 趋势图

Number of North Korean defectors
entering South Korea 1990-2009



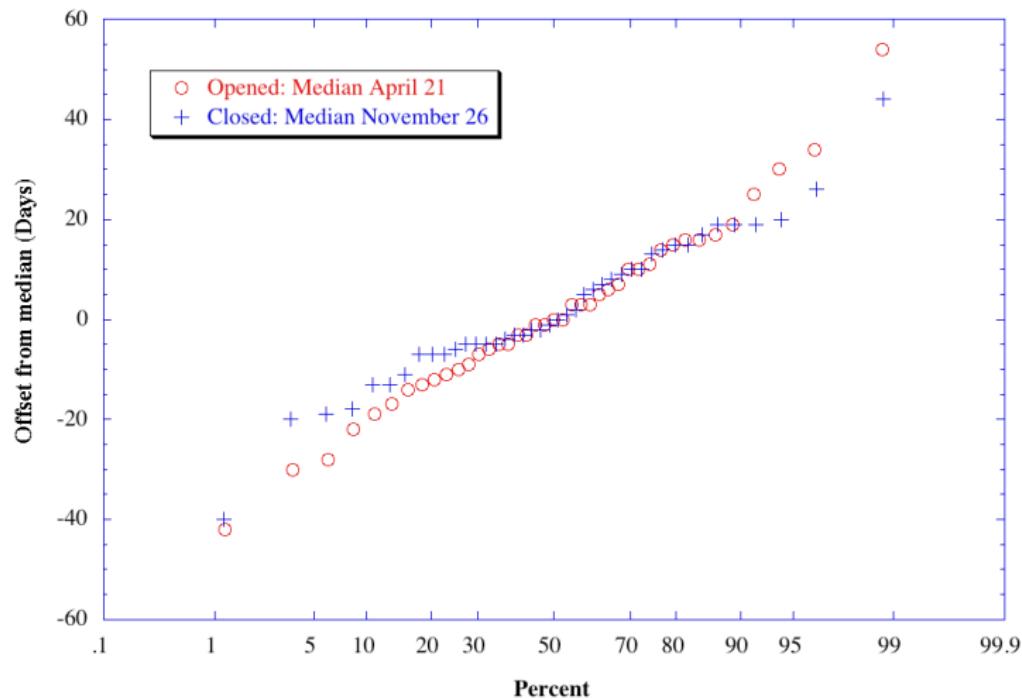
概率图

The normal probability plot is a graphical technique to identify substantive departures from normality. This includes identifying outliers, skewness, kurtosis, a need for transformations, and mixtures. Normal probability plots are made of raw data, residuals from model fits, and estimated parameters.

In a normal probability plot (also called a "normal plot"), the sorted data are plotted vs. values selected to make the resulting image look close to a straight line if the data are approximately normally distributed. Deviations from a straight line suggest departures from normality.

The normal probability plot is a special case of the Q-Q probability plot for a normal distribution. The theoretical quantiles are generally chosen to approximate either the mean or the median of the corresponding order statistics.

统计图形 | 概率图



Q-Q plot for first opening/final closing dates of Washington State Route 20 versus a normal distribution. Outliers are visible in the upper right corner.



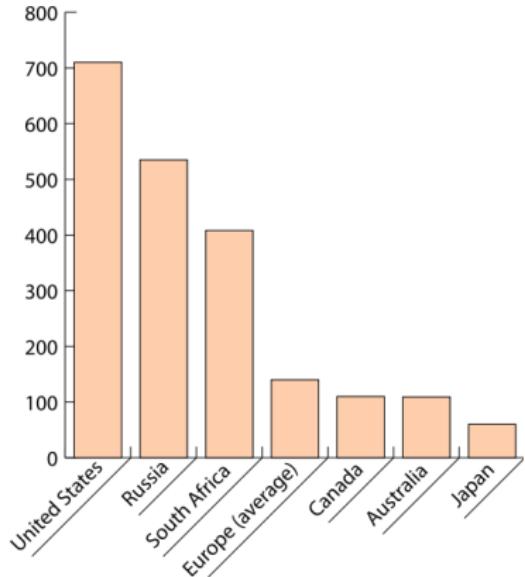
柱状图

长条图（bar chart）亦称条图（bar graph）、条状图、棒形图、柱状图，是一种以长方形的长度为变量的统计图表。长条图用来比较两个或以上的价值（不同时间或者不同条件），只有一个变量，通常利用于较小的数据集分析。长条图亦可横向排列，或用多维方式表达。

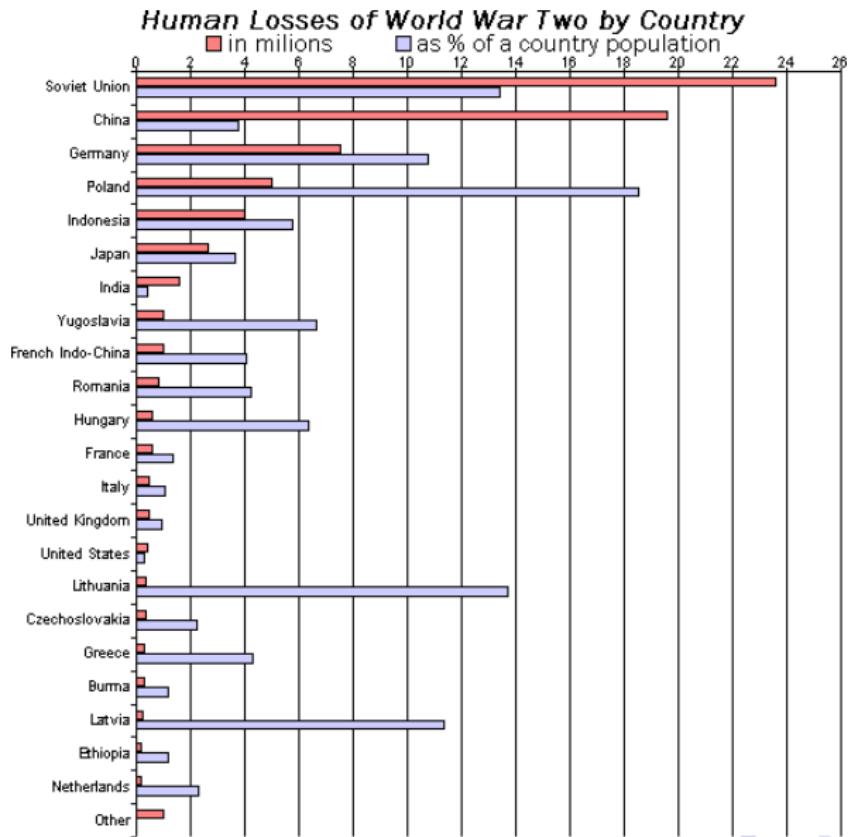
A bar chart or bar graph is a chart or graph that presents grouped data with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar chart is sometimes called a Line graph.

A bar graph is a chart that uses either horizontal or vertical bars to show comparisons among categories. One axis of the chart shows the specific categories being compared, and the other axis represents a discrete value. Some bar graphs present bars clustered in groups of more than one.

统计图形 | 柱状图



统计图形 | 柱状图



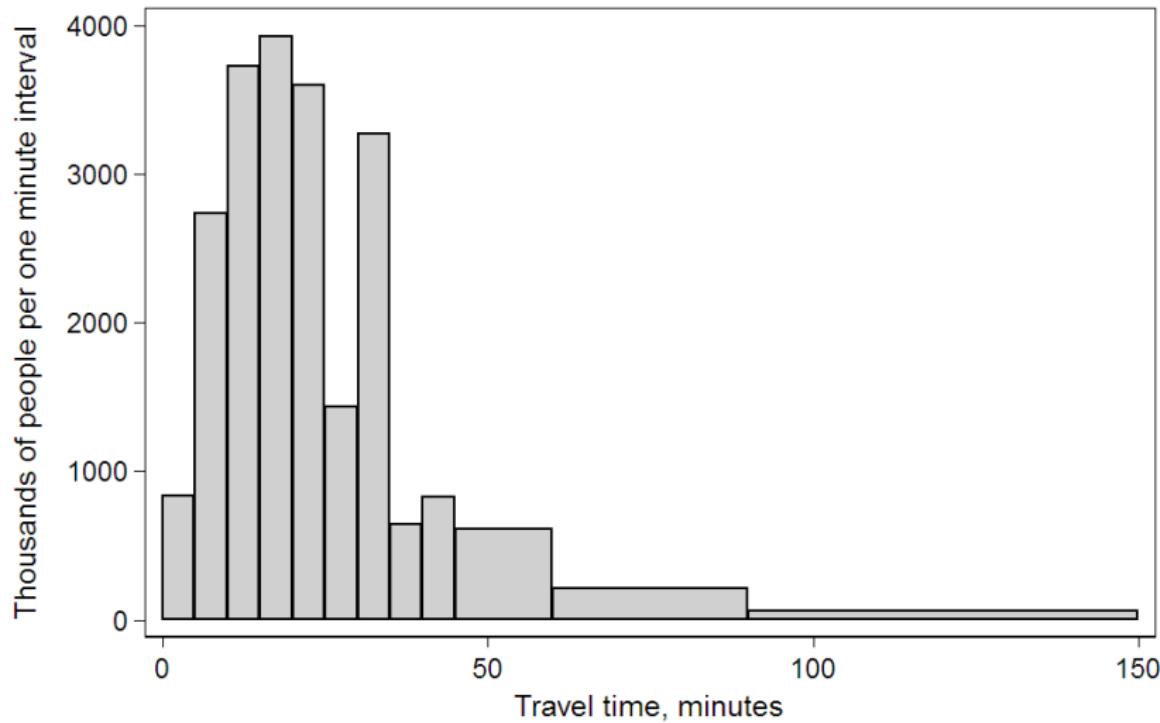
直方图

在统计学中，直方图（Histogram）是一种对数据分布情况的图形表示，是一种二维统计图表，它的两个坐标分别是统计样本和该样本对应的某个属性的度量。直方图是品质管理七大工具之一。

A histogram is a graphical representation of the distribution of numerical data. It is an estimate of the probability distribution of a continuous variable (quantitative variable) and was first introduced by Karl Pearson. It is a kind of bar graph.

To construct a histogram, the first step is to “bin” the range of values—that is, divide the entire range of values into a series of intervals—and then count how many values fall into each interval. The bins are usually specified as consecutive, non-overlapping intervals of a variable. The bins (intervals) must be adjacent, and are often (but are not required to be) of equal size.

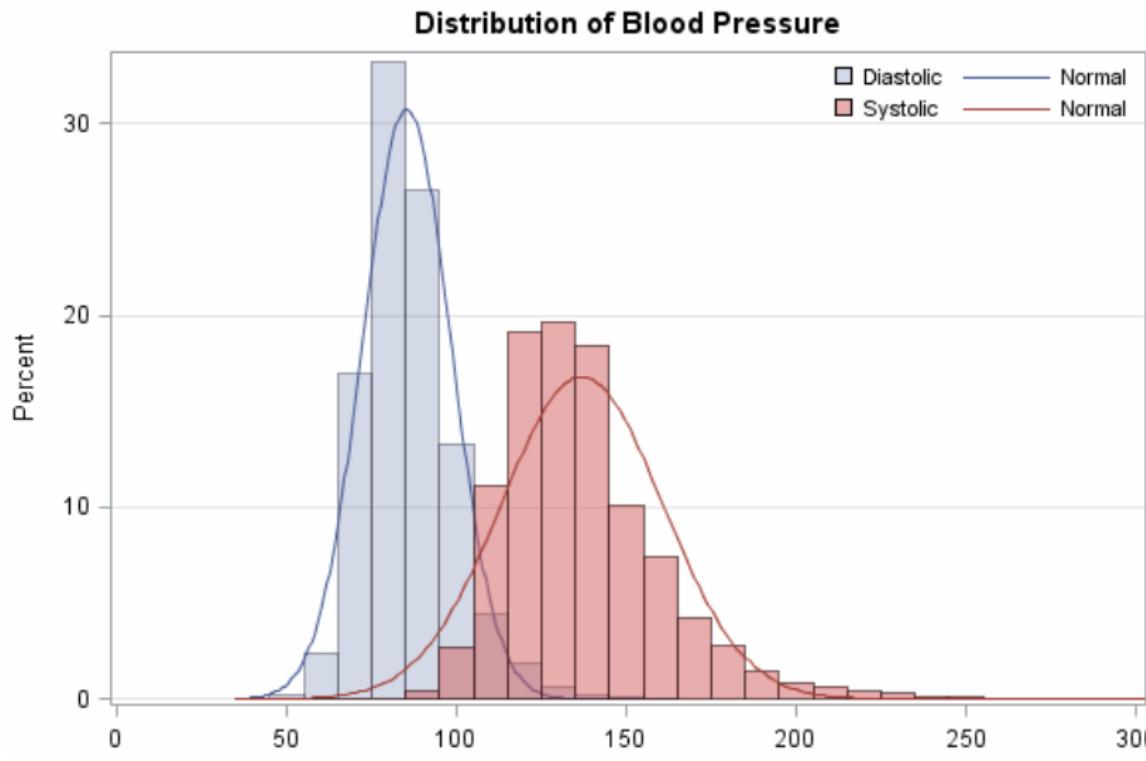
统计图形 | 直方图



Histogram of travel time (to work), US 2000 census. Area under the curve equals the total number of cases. This diagram uses Q/width from the table



统计图形 | 密度图



茎叶图

A stem-and-leaf display is often called a stemplot, but the latter term often refers to another chart type. A simple stem plot may refer to plotting a matrix of y values onto a common x axis, and identifying the common x value with a vertical line, and the individual y values with symbols on the line.

Unlike histograms, stem-and-leaf displays retain the original data to at least two significant digits, and put the data in order, thereby easing the move to order-based inference and non-parametric statistics.

A basic stem-and-leaf display contains two columns separated by a vertical line. The left column contains the stems and the right column contains the leaves.



统计图形 | 茎叶图

树 茎					树叶				
88	80	75	97	9	7	6	5		
57	69	74	96	9	4	4	3	2	0
86	79	68	56	8	8	6	5	5	
95	85	79	74	8	3	2	1	1	0
54	68	94	67	7	9	9	8	7	7
73	85	78	53	7	4	4	3	1	1
83	65	77	48	6	9	8	7	5	
94	71	93	82	5	7	6			
77	71	64	81	5	4	3			
76	70	92	64	4	8				
89	81	75	63	4					
75									

表(1)某校 98 级概率统计成绩

树叶					树茎					树叶				
0	2	2	3	4	4	7	9	6	5	0	0	1	2	3
5	5	6	7	7	8	8	9	4	3	5	5	6	7	8
5	5	6	7	7	8	8	9	8	9	8	9	8	8	7
2	3	3	4	4	9	7	7	4	4	4	4	4	2	1
1	2	3	4	4	6	7	7	4	1	1	1	1	0	0
5	6	5	6	5	5	6	6	3	3	0	5	4	4	3
								6	6	5	5	5	5	5

图 2 左右两侧分别是某校 98.(1),(2)班概率统计茎叶图成绩表



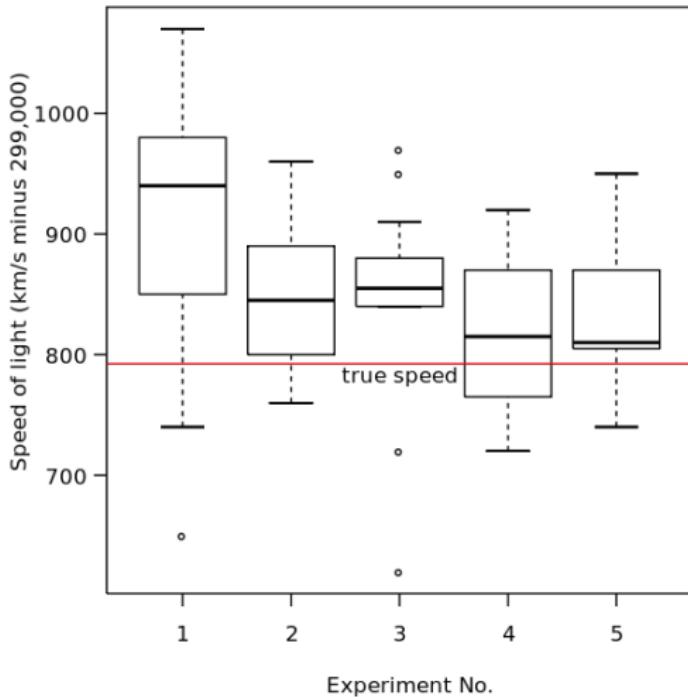
箱线图

箱形图 (Box-plot) , 又称为盒须图、盒式图、盒状图或箱线图, 是一种用作显示一组数据分散情况资料的统计图。因型状如箱子而得名。

箱形图于 1977 年由美国著名统计学家约翰 · 图基 (John Tukey) 发明。它能显示出一组数据的最大值、最小值、中位数、下四分位数及上四分位数。

In descriptive statistics, a Box-Plot or Box-and-Whiskers-Diagram is a convenient way of graphically depicting groups of numerical data through their quartiles. Box plots may also have lines extending vertically from the boxes (whiskers) indicating variability outside the upper and lower quartiles, hence the terms box-and-whisker plot and box-and-whisker diagram. Outliers may be plotted as individual points. Box plots are non-parametric: they display variation in samples of a statistical population without making any assumptions of the underlying statistical distribution.

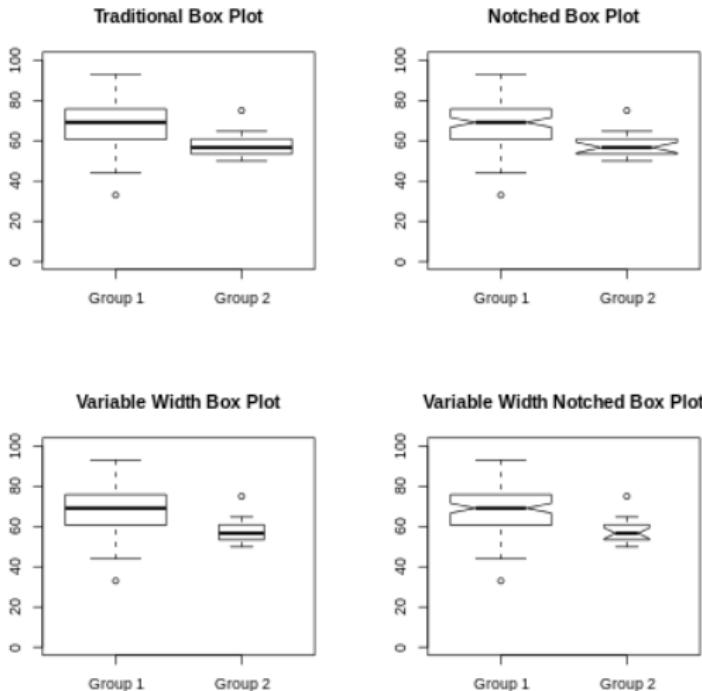
统计图形 | 箱线图



Box plot of data from the Michelson–Morley experiment.



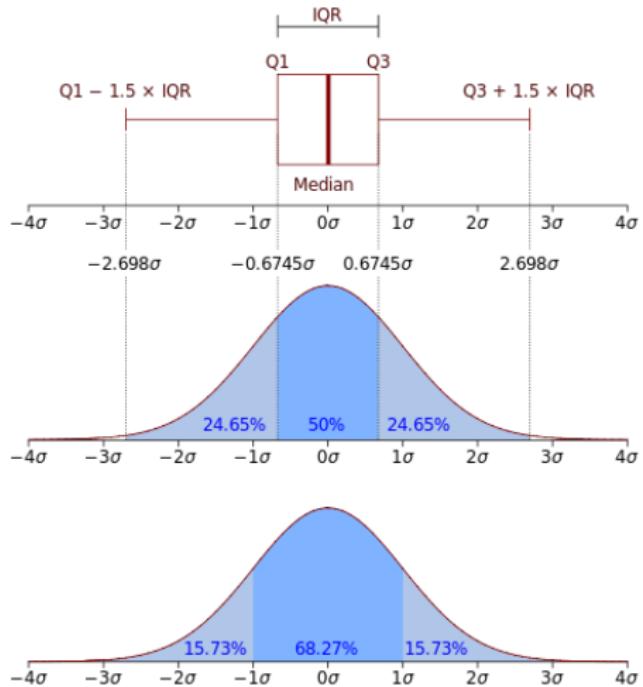
统计图形 | 箱线图



Four box plots, with and without notches and variable width.



统计图形 | 箱线图



Boxplot and a probability density function (pdf) of a Normal $N(0, 1\sigma^2)$ Population.



小提琴图

A violin plot is a method of plotting numeric data. It is similar to box plot with a rotated kernel density plot on each side.

The violin plot is similar to box plots, except that they also show the probability density of the data at different values (in the simplest case this could be a histogram). Typically violin plots will include a marker for the median of the data and a box indicating the interquartile range, as in standard box plots. Overlaid on this box plot is a kernel density estimation. Like box plots, violin plots are used to represent comparison of a variable distribution (or sample distribution) across different “categories”.

A violin plot is more informative than a plain box plot. In fact while a box plot only shows summary statistics such as mean/median and interquartile ranges, the violin plot shows the full distribution of the data. The difference is particularly useful when the data distribution is multimodal (more than one peak). In this case a violin plot clearly shows the presence of different peaks, their position and relative amplitude. This information could not be represented with a simple box plot which only reports summary statistics. The inner part of a violin plot usually shows the mean (or median) and the interquartile range. In other cases, when the number of samples is not too high, the inner part can show all sample points (with a dot or a line for each sample).

统计图形 | 小提琴图

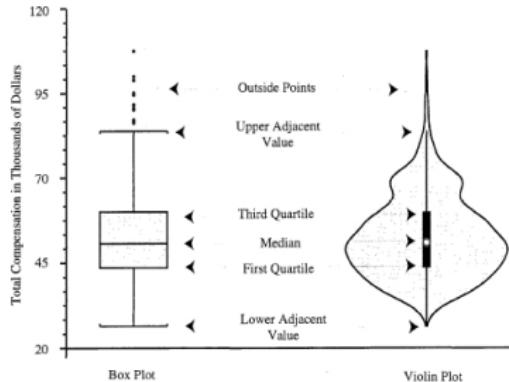
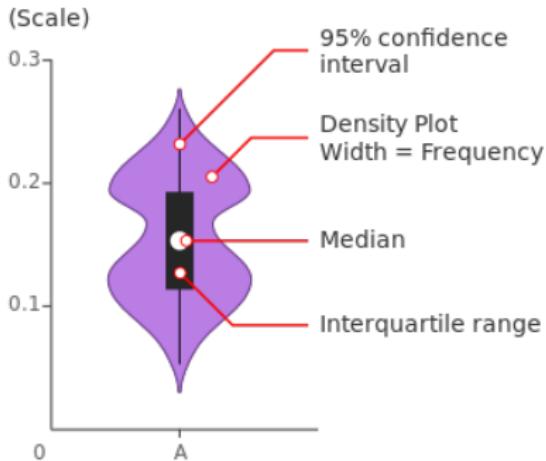
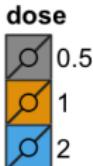
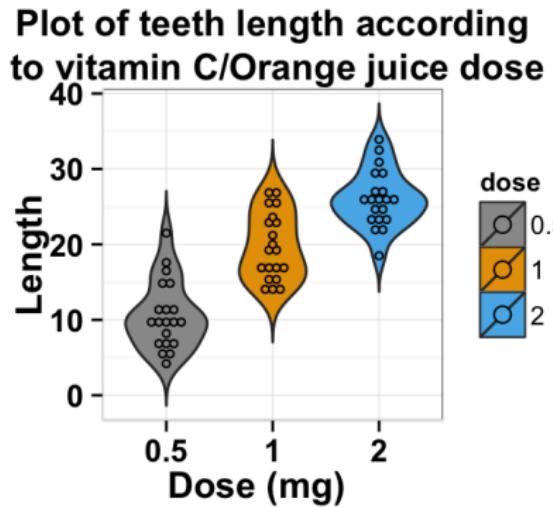


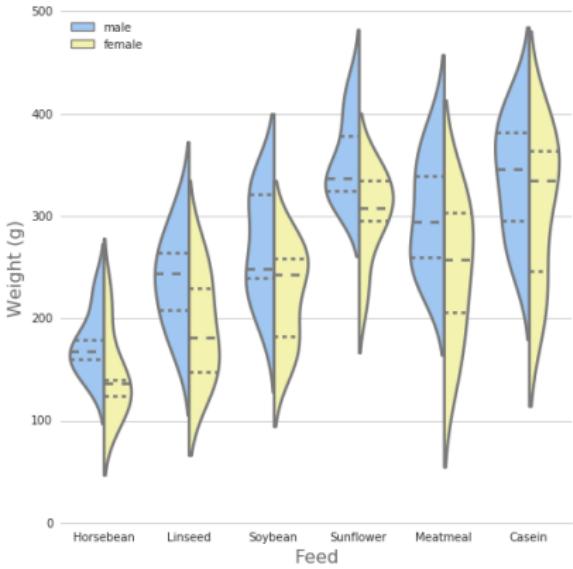
Figure 1. Common Components of Box Plot and Violin Plot. Total compensation for all academic ranks.



统计图形 | 小提琴图



Chick weights by feed type



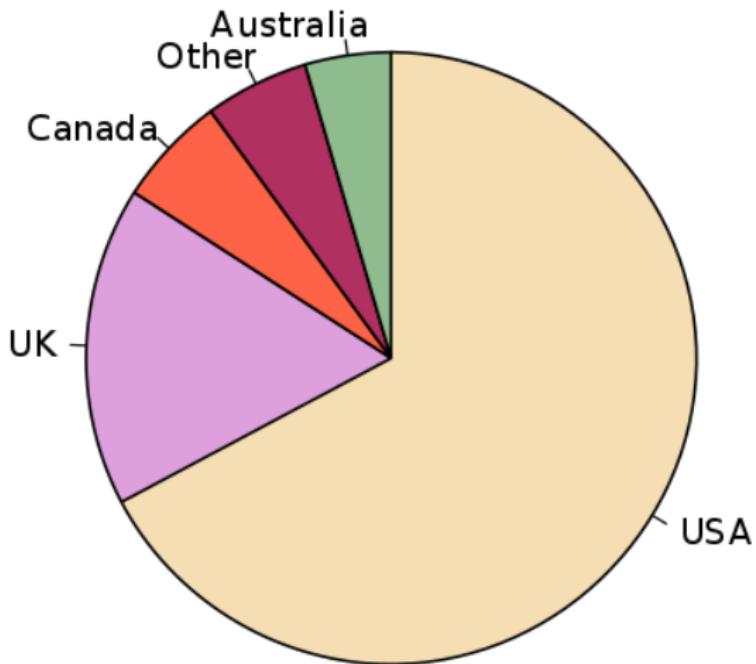
饼图

饼图，或称饼状图，是一个划分为几个扇形的圆形统计图表，用于描述量、频率或百分比之间的相对关系。在饼图中，每个扇区的弧长（以及圆心角和面积）大小为其所表示的数量的比例。这些扇区合在一起刚好是一个完全的圆形。顾名思义，这些扇区拼成了一个切开的饼形图案。

饼图在商业领域和大众媒体中几乎无处不在，但很少用于科技出版物。这是受到批评最多的图表之一，而很多统计学家建议避免使用这一图表。它们指出，在饼图中很难对不同的扇区大小进行比较，或对不同饼图之间数据进行比较。

在一些特定情况下，饼图可以很有效地对信息进行展示。特别是在想要表示某个大扇区在整体中所占比例，而不是对不同扇区进行比较时，这一方法十分有效。饼图在扇区所占比例达到总体的 25% 或 50% 时，可以很好地达到展示的目的。但通常，可能更多情况会采用其它图表如条形图或圆点图（dot plot），或非图表的方法如表格来表达信息。

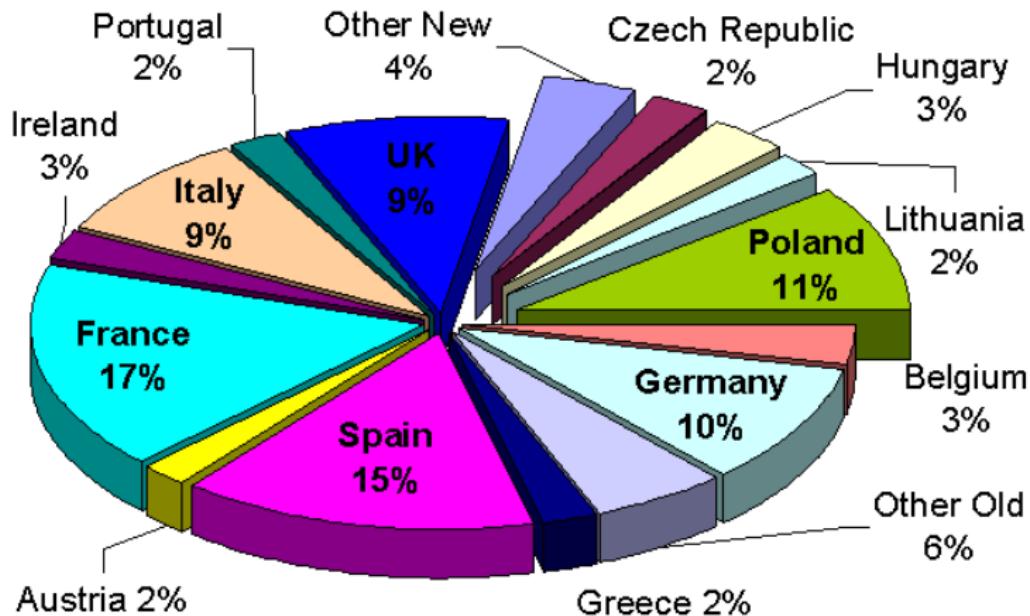
统计图形 | 饼图



英语为母语的人口分布饼图。



Percentage of EU Farm Land by Country



欧盟各国拥有欧盟农田的百分比。

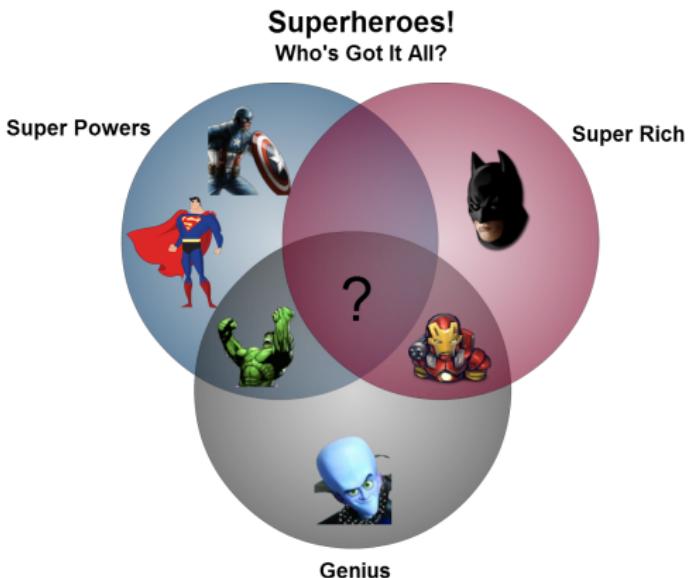
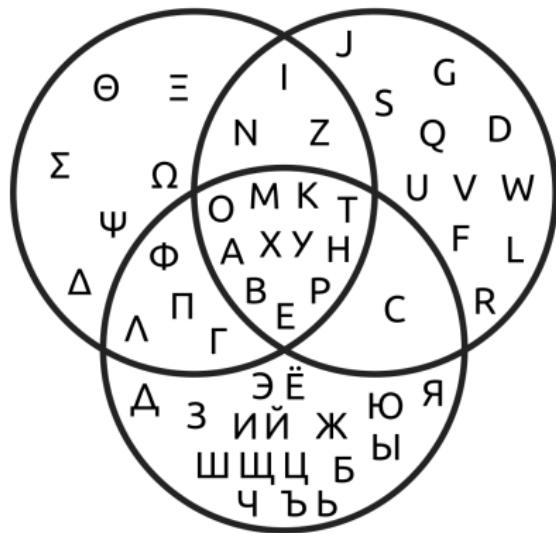


韦恩图

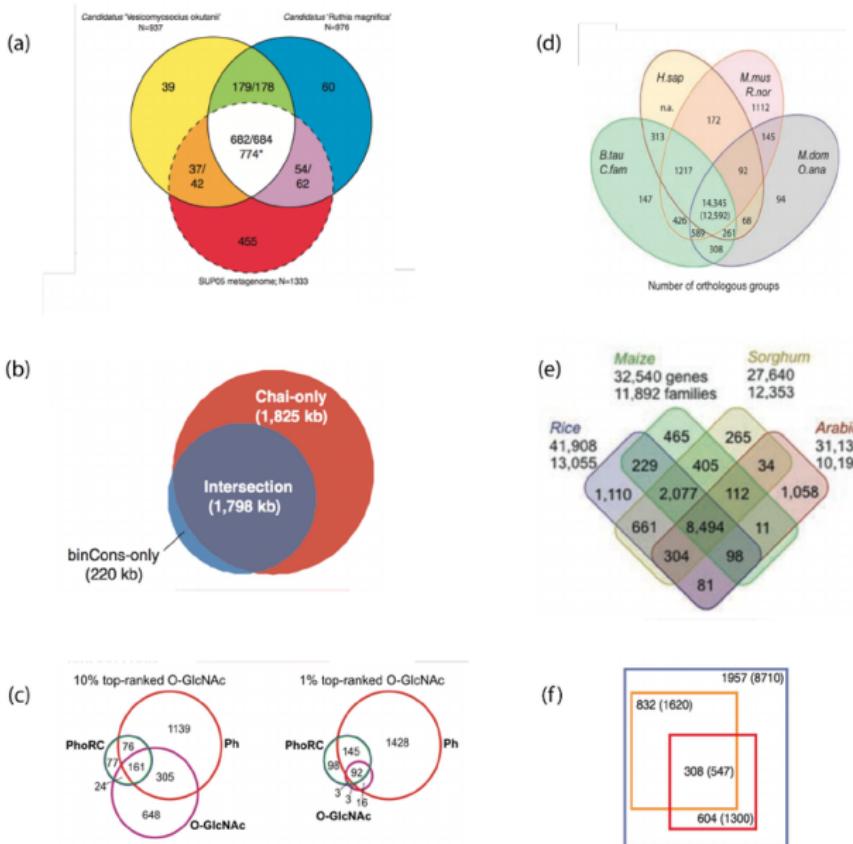
文氏图 (Venn diagram) , 或译 Venn 图、温氏图、维恩图、范氏图, 是在所谓的集合论 (或者类的理论) 数学分支中, 在不太严格的意义下用以表示集合 (或类) 的一种草图。它们用于展示在不同的事物群组 (集合) 之间的数学或逻辑联系, 尤其适合用来表示集合 (或) 类之间的“大致关系”, 它也常常被用来帮助推导 (或理解推导过程) 关于集合运算 (或类运算) 的一些规律。

A Venn diagram (also called primary diagram, set diagram or logic diagram) is a diagram that shows all possible logical relations between a finite collection of different sets. These diagrams depict elements as points in the plane, and sets as regions inside closed curves. A Venn diagram consists of multiple overlapping closed curves, usually circles, each representing a set. The points inside a curve labelled S represent elements of the set S, while points outside the boundary represent elements not in the set S. In Venn diagrams the curves are overlapped in every possible way, showing all possible relations between the sets. They are thus a special case of Euler diagrams, which do not necessarily show all relations. Venn diagrams were conceived around 1880 by John Venn. They are used to teach elementary set theory, as well as illustrate simple set relationships in probability, logic, statistics, linguistics and computer science.

统计图形 | 韦恩图



统计图形 | 韦恩图

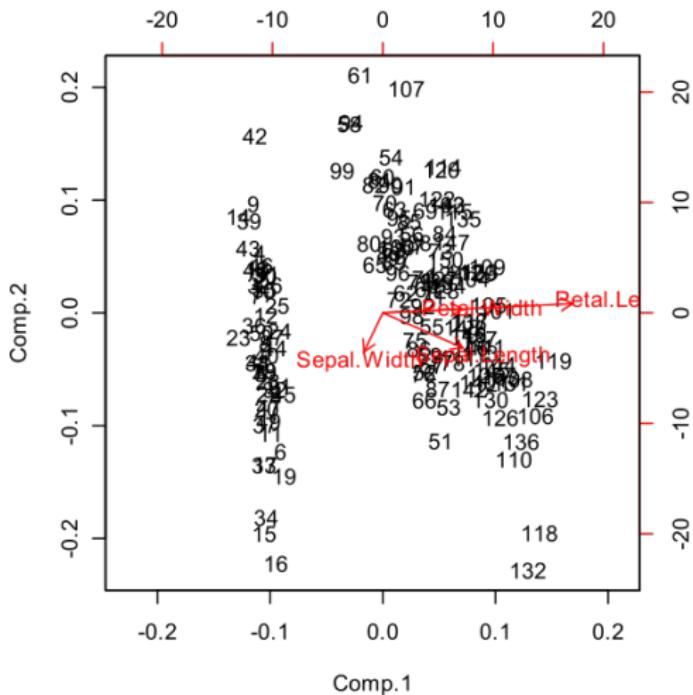


双标图

双标图 (Biplots) 是一类统计学的统计图形。双标图可以同时把抽样和资料矩阵变量中的数据用图表表示出来。抽样样本可以用向量、线性轴和非线性轨迹表达。在类别变量的案例中，类别水平点 (category level points) 可以用来代表其类别变量的水平值。一类广义的双标图可以适用呈现序数标量和区间标量。



统计图形 | 双标图



Biplot of Anderson's iris data set.



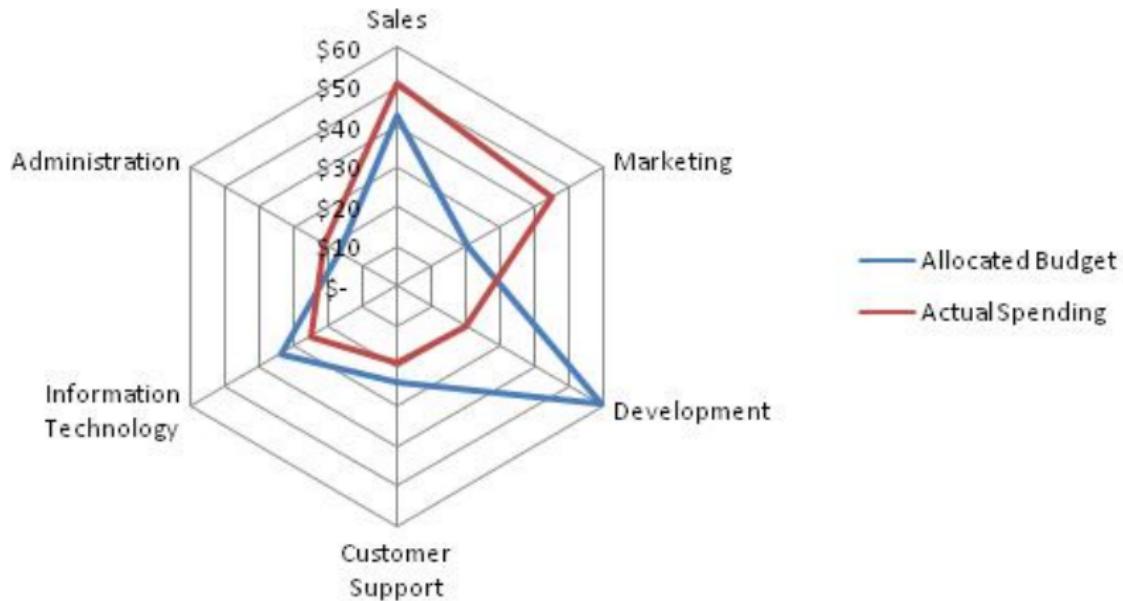
雷达图

A radar chart is a graphical method of displaying multivariate data in the form of a two-dimensional chart of three or more quantitative variables represented on axes starting from the same point. The relative position and angle of the axes is typically uninformative.

The radar chart is also known as web chart, spider chart, star chart, star plot, cobweb chart, irregular polygon, polar chart, or Kiviat diagram. It is equivalent to a parallel coordinates plot in polar coordinates.



统计图形 | 雷达图



This spider chart represents the allocated budget versus actual spending for a given organization.



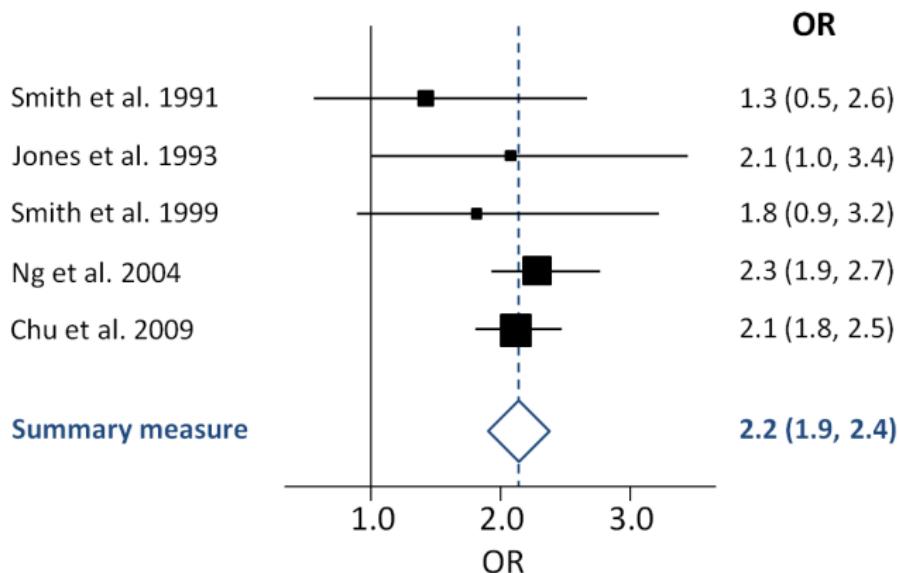
森林图

A forest plot, also known as a blobbogram, is a graphical display of estimated results from a number of scientific studies addressing the same question, along with the overall results.

It was developed for use in medical research as a means of graphically representing a meta-analysis of the results of randomized controlled trials. In the last twenty years, similar meta-analytical techniques have been applied in observational studies (e.g. environmental epidemiology) and forest plots are often used in presenting the results of such studies also.



统计图形 | 森林图



An example forest plot of five odds ratios (squares, proportional to weights used in meta-analysis), with the summary measure (centre line of diamond) and associated confidence intervals (lateral tips of diamond), and solid vertical line of no effect. Names of (fictional) studies are shown on the left, odds ratios and confidence intervals on the right.

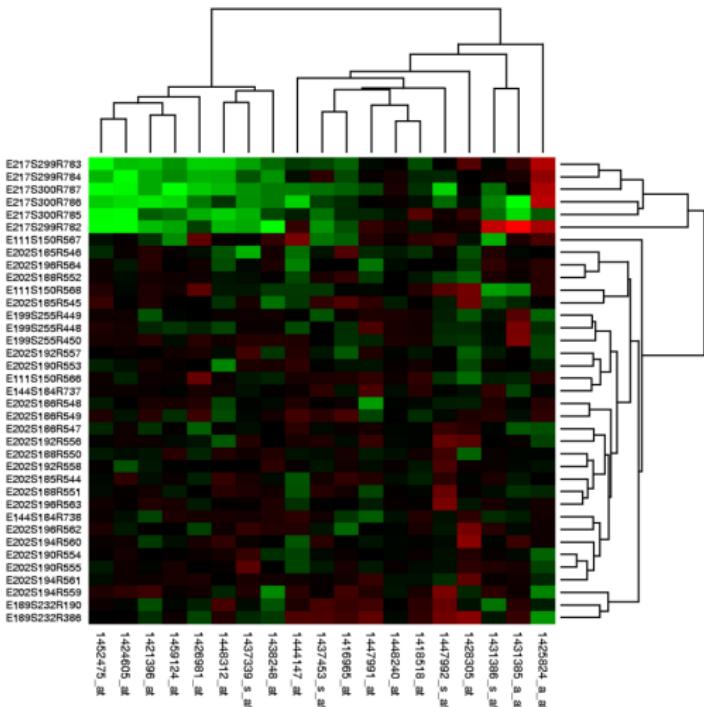


热图

A heat map (or heatmap) is a graphical representation of data where the individual values contained in a matrix are represented as colors. The term ‘heat map’ was originally coined and trademarked by software designer Cormac Kinney in 1991, to describe a 2D display depicting financial market information, though similar plots such as shading matrices have existed for over a century.



统计图形 | 热图



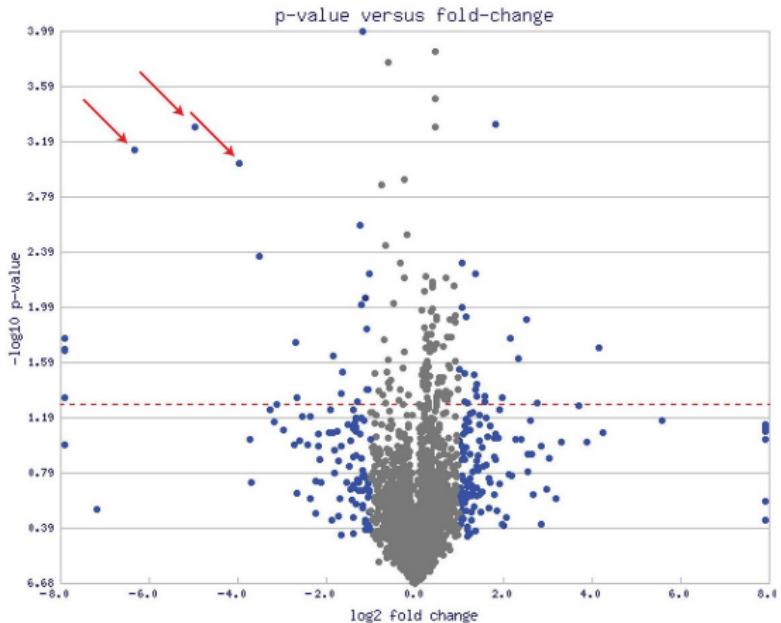
Heat map generated from DNA microarray data reflecting gene expression values in several conditions.

火山图

In statistics, a volcano plot is a type of scatter-plot that is used to quickly identify changes in large data sets composed of replicate data. It plots significance versus fold-change on the y and x axes, respectively. These plots are increasingly common in omic experiments such as genomics, proteomics, and metabolomics where one often has a list of many thousands of replicate data points between two conditions and one wishes to quickly identify the most meaningful changes. A volcano plot combines a measure of statistical significance from a statistical test (e.g., a p value from an ANOVA model) with the magnitude of the change, enabling quick visual identification of those data-points (genes, etc.) that display large magnitude changes that are also statistically significant.

A volcano plot is constructed by plotting the negative log of the p value on the y axis (usually base 10). This results in data points with low p values (highly significant) appearing toward the top of the plot. The x axis is the log of the fold change between the two conditions. The log of the fold change is used so that changes in both directions appear equidistant from the center. Plotting points in this way results in two regions of interest in the plot: those points that are found toward the top of the plot that are far to either the left- or right-hand sides. These represent values that display large magnitude fold changes (hence being left or right of center) as well as high statistical significance (hence being toward the top).

统计图形 | 火山图



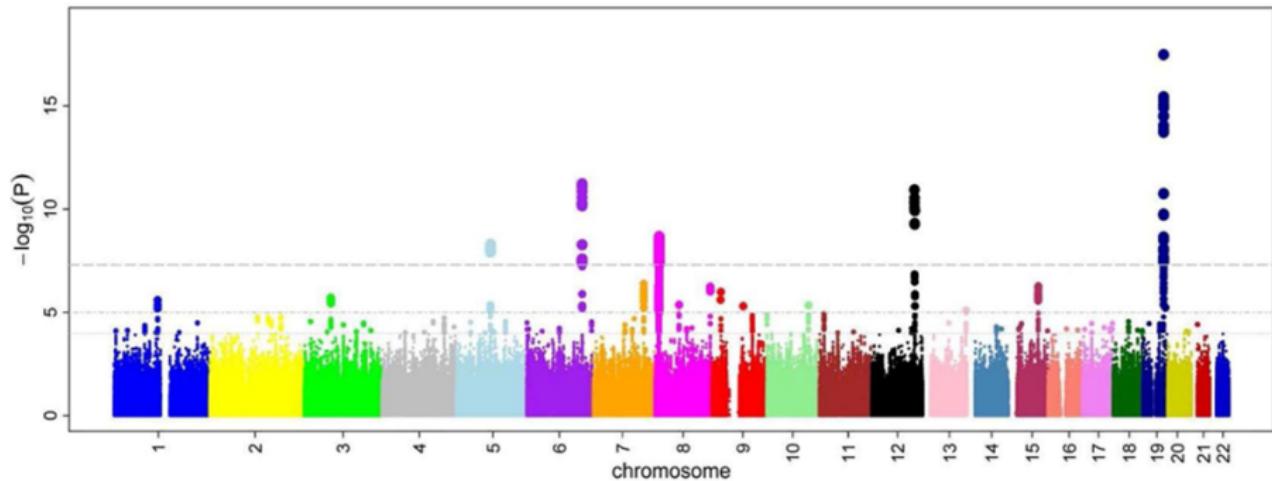
Volcano plot showing metabolomic data. The red arrows indicate points-of-interest that display both large magnitude fold-changes (x axis) and high statistical significance (- \log_{10} of p value, y axis). The dashed red line shows where $p = 0.05$ with points above the line having $p < 0.05$ and points below the line having $p > 0.05$. This plot is colored such that those points having a fold-change less than 2 ($\log_2 = 1$) are shown in gray.

曼哈顿图

A Manhattan plot is a type of scatter plot, usually used to display data with a large number of data-points - many of non-zero amplitude, and with a distribution of higher-magnitude values, for instance in genome-wide association studies (GWAS). In GWAS Manhattan plots, genomic coordinates are displayed along the X-axis, with the negative logarithm of the association P-value for each single nucleotide polymorphism (SNP) displayed on the Y-axis, meaning that each dot on the Manhattan plot signifies a SNP. Because the strongest associations have the smallest P-values (e.g., 10^{-15}), their negative logarithms will be the greatest (e.g., 15).



统计图形 | 曼哈顿图



An illustration of a Manhattan plot depicting several strongly associated risk loci.

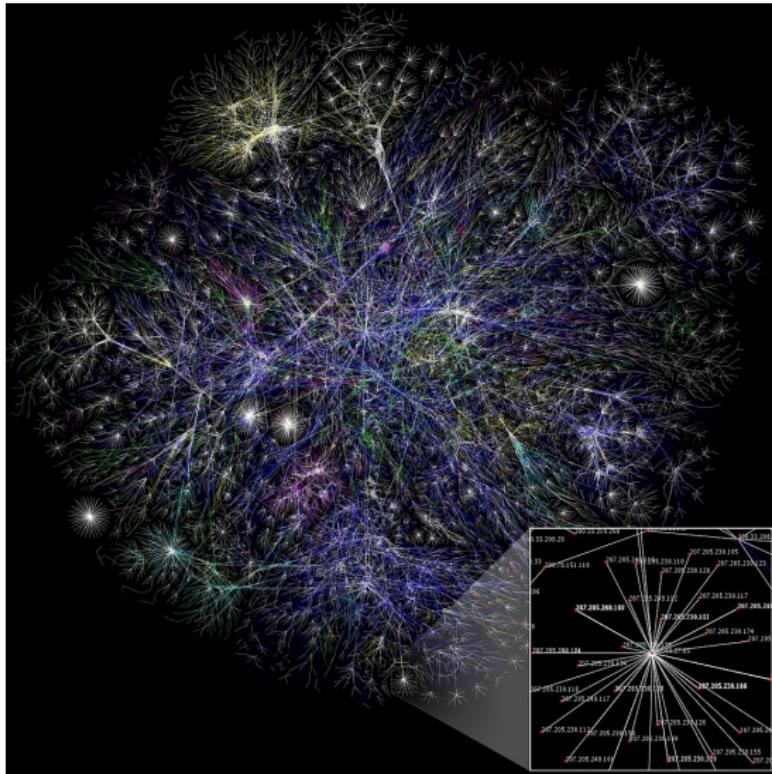


网络可视化

Graph drawing is an area of mathematics and computer science combining methods from geometric graph theory and information visualization to derive two-dimensional depictions of graphs arising from applications such as social network analysis, cartography, linguistics, and bioinformatics.

A drawing of a graph or network diagram is a pictorial representation of the vertices and edges of a graph. This drawing should not be confused with the graph itself: very different layouts can correspond to the same graph. In the abstract, all that matters is which pairs of vertices are connected by edges. In the concrete, however, the arrangement of these vertices and edges within a drawing affects its understandability, usability, fabrication cost, and aesthetics. The problem gets worse, if the graph changes over time by adding and deleting edges (dynamic graph drawing) and the goal is to preserve the user's mental map.

统计图形 | 网络可视化



2005年初因特网的部分映射，每条线表示两个IP地址以及两个节点之间的延迟。



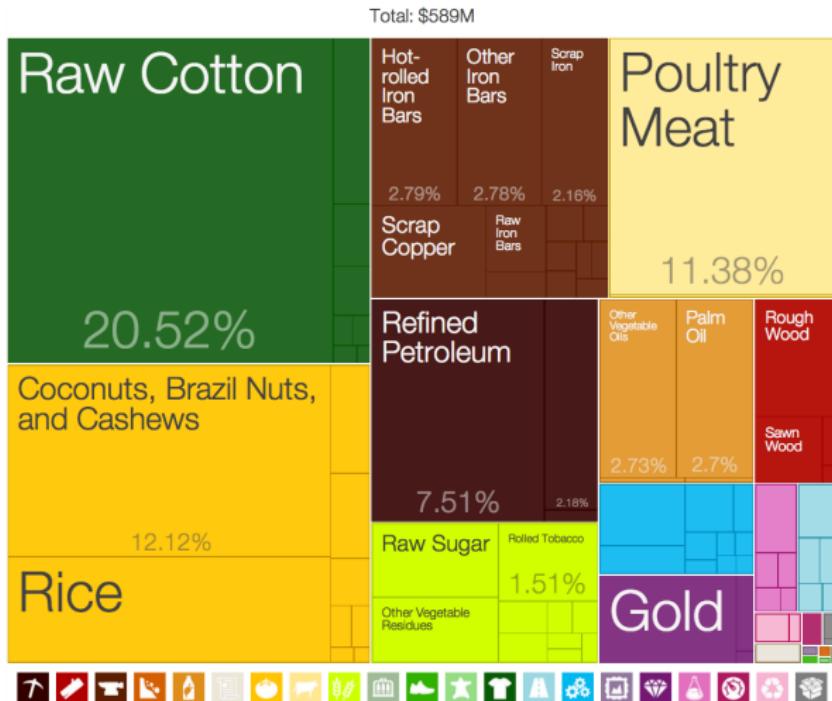
矩形树图

In information visualization and computing, treemapping is a method for displaying hierarchical data using nested rectangles.

Treemaps display hierarchical (tree-structured) data as a set of nested rectangles. Each branch of the tree is given a rectangle, which is then tiled with smaller rectangles representing sub-branches. A leaf node's rectangle has an area proportional to a specified dimension of the data. Often the leaf nodes are colored to show a separate dimension of the data.

When the color and size dimensions are correlated in some way with the tree structure, one can often easily see patterns that would be difficult to spot in other ways, such as if a certain color is particularly relevant. A second advantage of treemaps is that, by construction, they make efficient use of space. As a result, they can legibly display thousands of items on the screen simultaneously.

统计图形 | 矩形树图



Treemap of Benin's exports by product category, 2009. The Product Exports

Treemaps are one of the most recent applications of these kind of visualizations, developed by the Harvard-MIT Observatory of Economic Complexity.



示意地图

示意地图 (Cartogram) , 又称比较统计地图, 是将地图根据统计数据变形的得到示意图, 通常是将各个地理单位的面积扩大或缩小, 来表示有关数据的数值。

示意地图的用途在于视觉上展示统计数据, 在尽可能保持各数据点之间的相对位置之时, 亦透过数据点的实际面积来标示其数值, 使读者不会因为地理因素影响以产生错觉。

实例

2004 年美国总统选举的投票分布图, 若以传统的地理图表显示, 会看到小布什得到压倒性的支持。不过, 一但用示意地图把同一组数据重组, 就会发现原来因为大多数支持民主党的支持者都集中在沿岸的大城市, 所以在地理图中被挤至边沿。事实上, 双方的确是势均力敌的。

示意地图

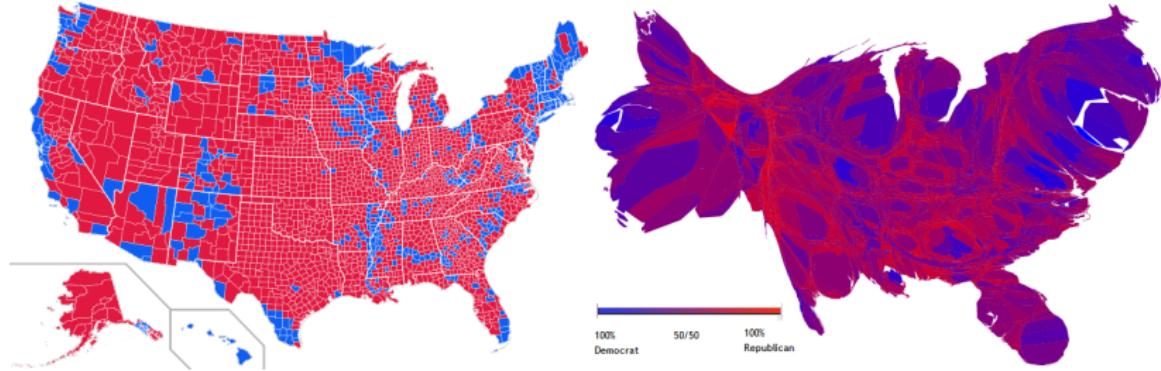
示意地图 (Cartogram) , 又称比较统计地图, 是将地图根据统计数据变形的得到示意图, 通常是将各个地理单位的面积扩大或缩小, 来表示有关数据的数值。

示意地图的用途在于视觉上展示统计数据, 在尽可能保持各数据点之间的相对位置之时, 亦透过数据点的实际面积来标示其数值, 使读者不会因为地理因素影响以产生错觉。

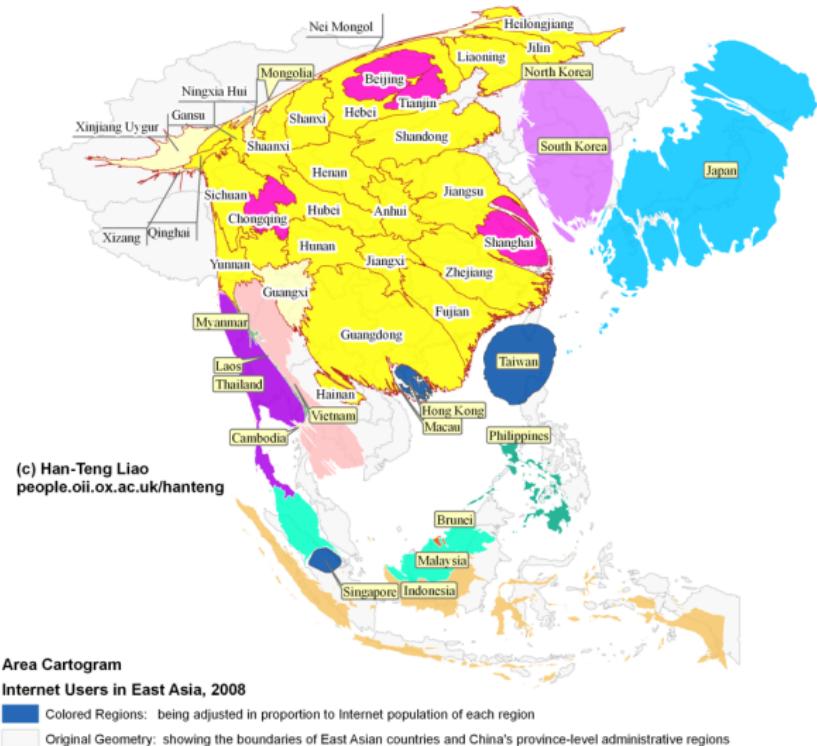
实例

2004 年美国总统选举的投票分布图, 若以传统的地理图表显示, 会看到小布什得到压倒性的支持。不过, 一但用示意地图把同一组数据重组, 就会发现原来因为大多数支持民主党的支持者都集中在沿岸的大城市, 所以在地理图中被挤至边沿。事实上, 双方的确是势均力敌的。

统计图形 | 示意地图



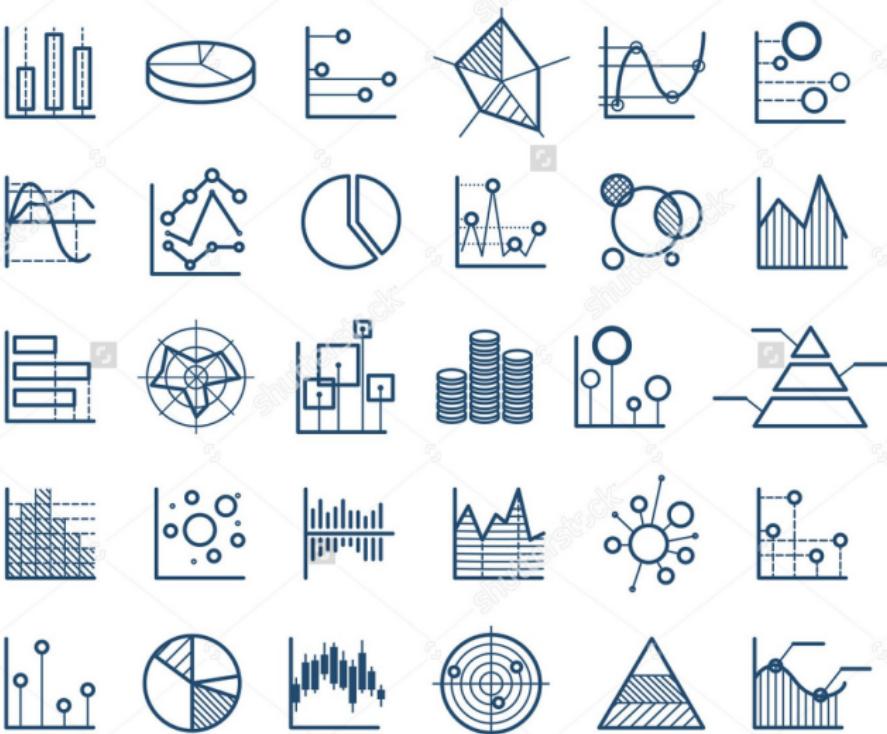
统计图形 | 示意地图



2008 年东亚网民分布的面积统计比较地图，面积愈大代表人数愈多。

统计图形 | 总结

STATISTICS CHART ICONS AND LINEAR BUSINESS MARKET DIAGRAMS



- 1 统计图形
 - 引言
 - 简介
 - 著名设计者

- 2 知识拓展
 - 埃舍尔
 - 眼见未必为实
- 3 图说天下



威廉 · 普莱费尔 (William Playfair)

发表了所谓的第一幅饼图以及众所周知的那张描绘英格兰进出口发展演变情况的图形。

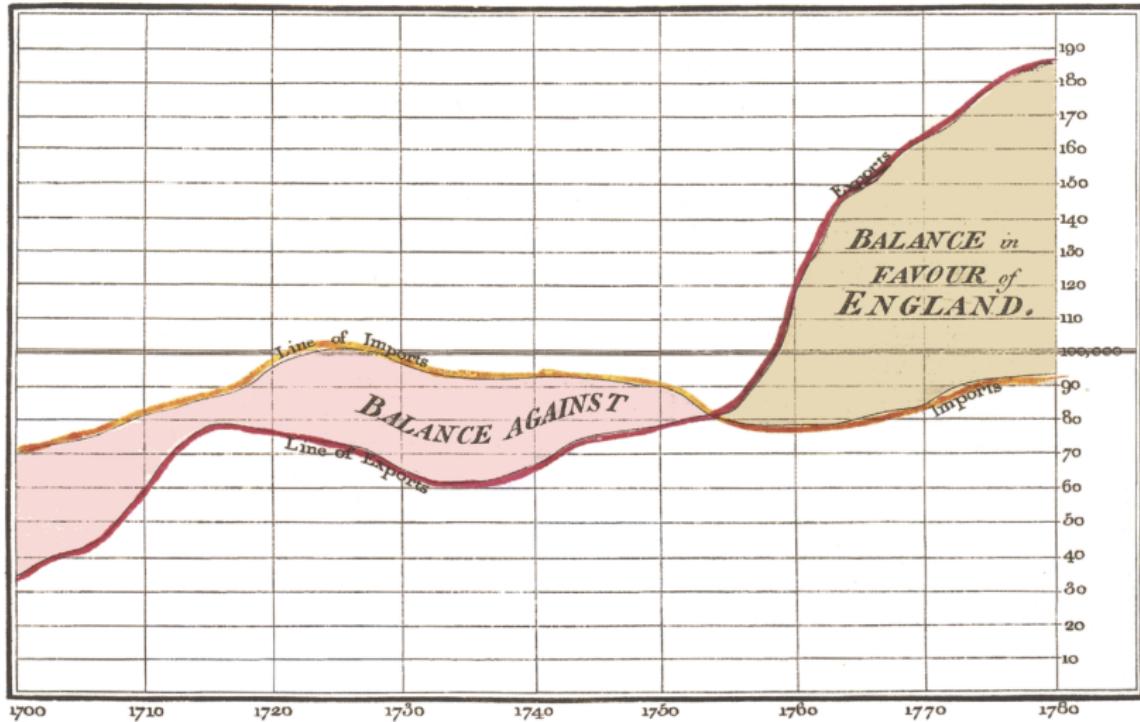
William Playfair was a Scottish engineer and political economist, the founder of graphical methods of statistics.

William Playfair invented several types of diagrams: in 1786 the line, area and bar chart of economic data, and in 1801 the pie chart and circle graph, used to show part-whole relations.



统计图形 | 设计者 | 威廉·普莱费尔

Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



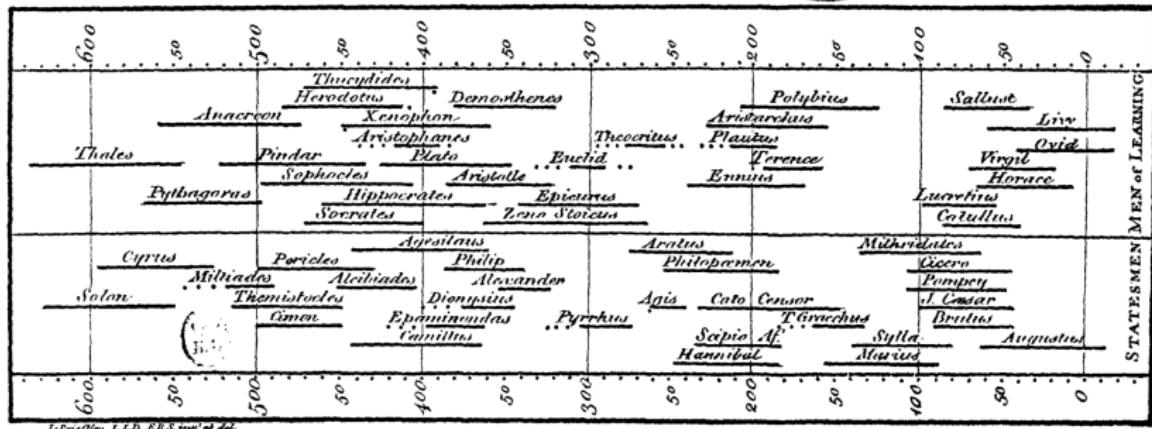
The Bottom line is divided into Years, the Right hand line into £10,000 each.

Published as the Act directs, 1st May 1786, by Wm. Playfair.

No. 60, Middlesex Strand, London.



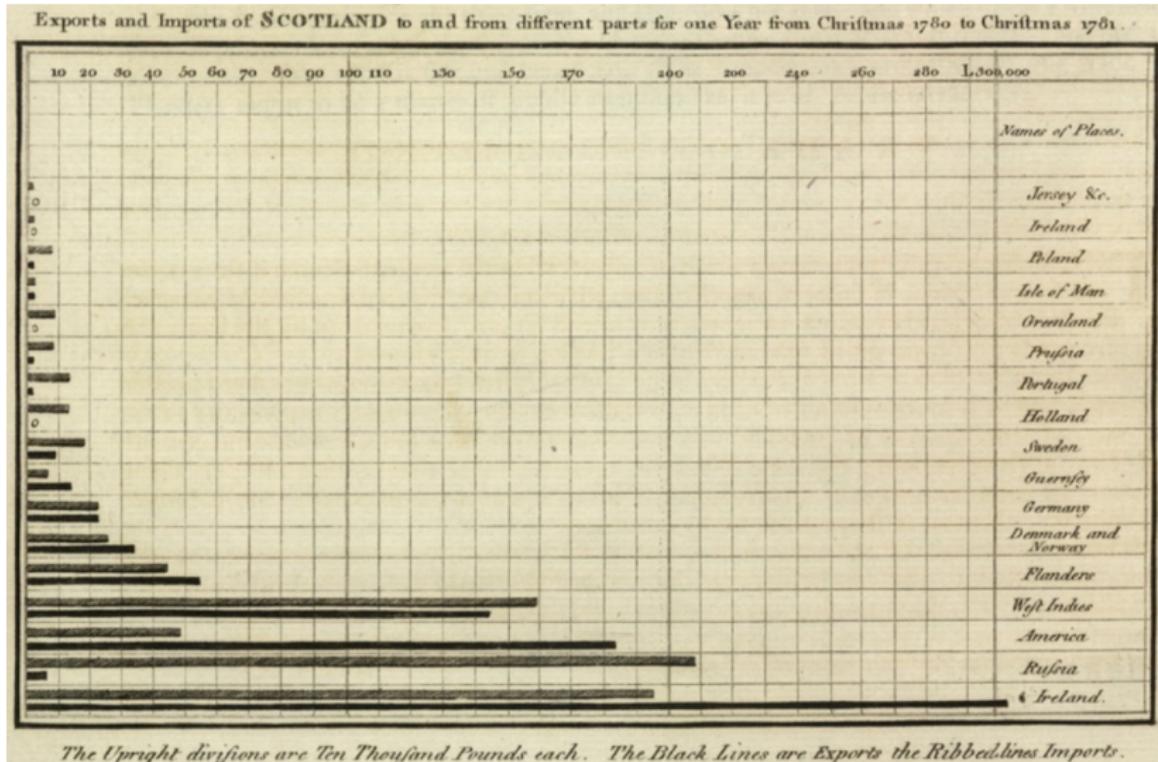
A Specimen of a Chart of Biography.

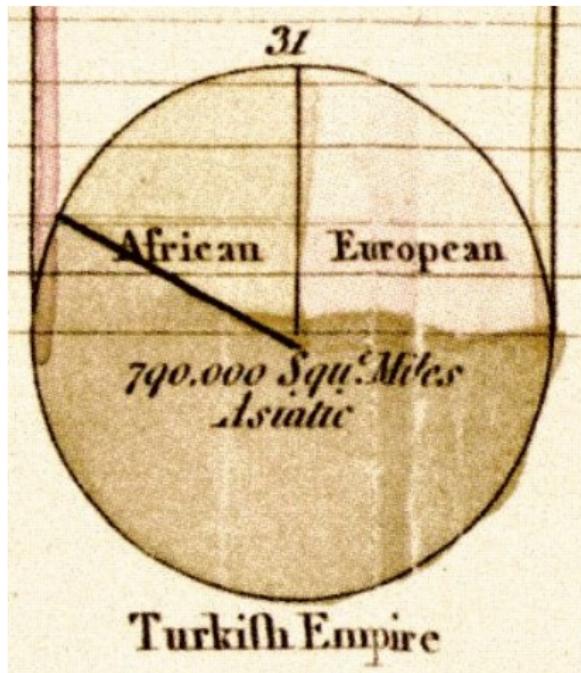


J. Priestley L. L. D. F.R.S. now ad dat.



统计图形 | 设计者 | 威廉·普莱费尔





Pie chart from Playfair's Statistical Breviary (1801), showing the proportions of the Turkish Empire located in Asia, Europe and Africa before 1789.



南丁格尔

弗罗伦斯·南丁格尔，(Florence Nightingale)，英国护士和统计学家。曾经运用统计图形来说服英国政府，以改善军队的卫生状况。

克里米亚战争时，她极力向英国军方争取在战地开设医院，为士兵提供医疗护理。她分析过堆积如山的军事档案，指出在克里米亚战役中，英军死亡的原因是在战场外感染疾病，及在战场上受伤后缺乏适当护理而伤重致死，真正死在战场上的人反而不多。她更用了圆形图以说明这些资料。

由于南丁格尔的贡献，让昔日地位低微的护士，社会地位与形象都大为提高，成为崇高的象征。“南丁格尔”也成为护士精神的代名词。



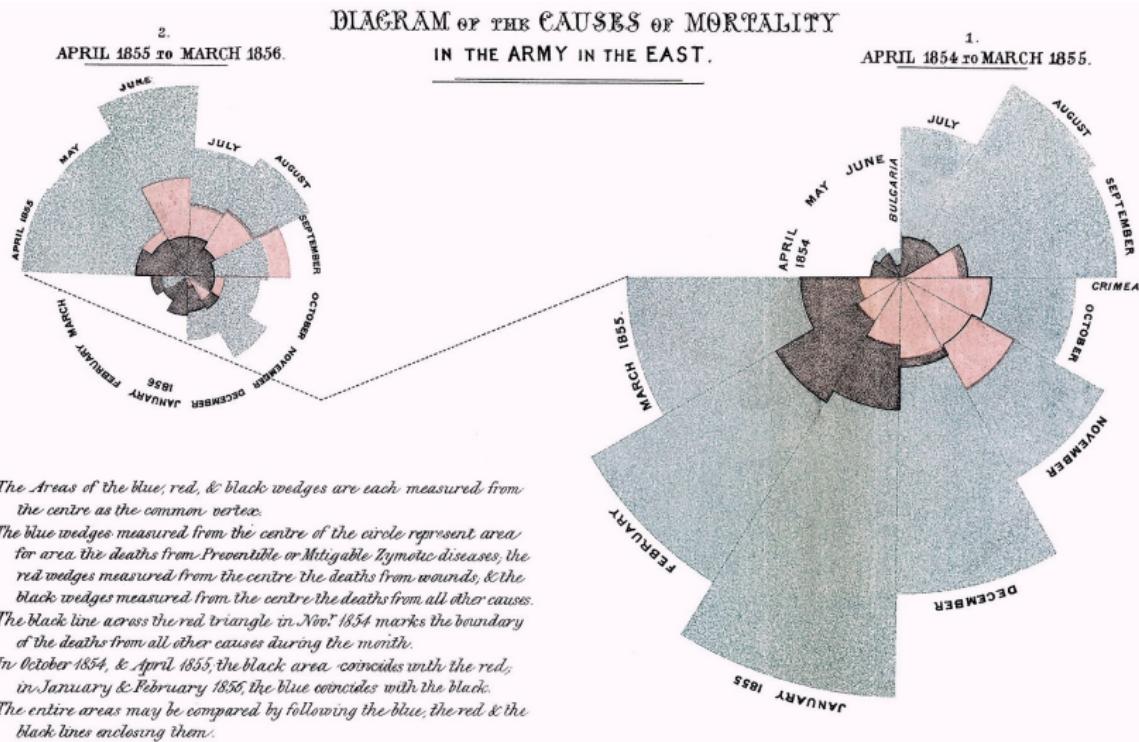
南丁格尔 vs. 统计

南丁格尔从小就显示出对数学的天分，后来，南丁格尔成为视觉表现和统计图形的先驱。她所使用的圆饼图，虽然在 1801 年由威廉普莱费尔所发明，它在当时仍是一个新颖的显示数据的方法。

南丁格尔被描述为“在统计的图形显示方法上，是一个真正的先驱”，她发展出极坐标图饼图的形式（polar area diagram），或称为南丁格尔玫瑰图（Nightingale rose diagram），相当于现代圆形直方图（circular histogram），以说明她在管理的野战医院内，病人死亡率在不同季节的变化。她使用极坐标图饼图，向不会阅读统计报告的国会议员，报告克里米亚战争的医疗条件。

在 1859 年南丁格尔被选为英国皇家统计学会的第一个女成员，她后来成为美国统计协会的名誉会员。





弗罗伦斯·南丁格尔绘制的“东部军队死亡原因统计图”。



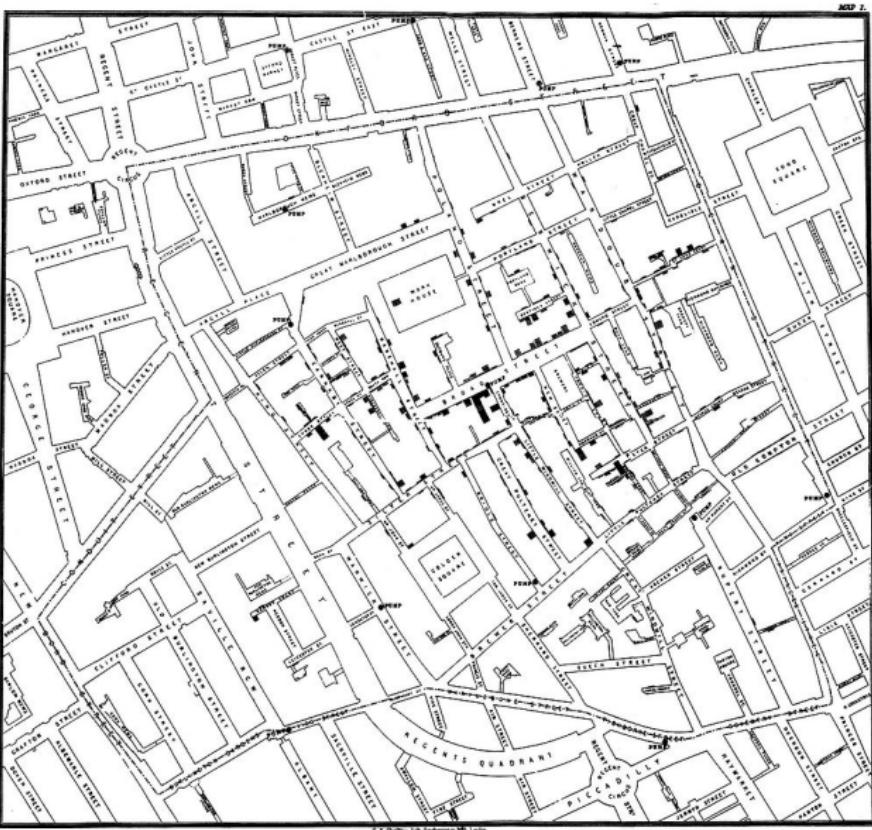
约翰·斯诺

绘制了 1854 伦敦霍乱死亡病例的分布图，从而发现了病源所在。斯诺的研究可以说是公共卫生学历史上一重大事件。

约翰·斯诺（John Snow），英国内科医生，因在 1854 年宽街霍乱爆发事件研究中作出重大贡献，被认为是麻醉学和公共卫生医学的开拓者。



统计图形 | 设计者 | 约翰·斯诺



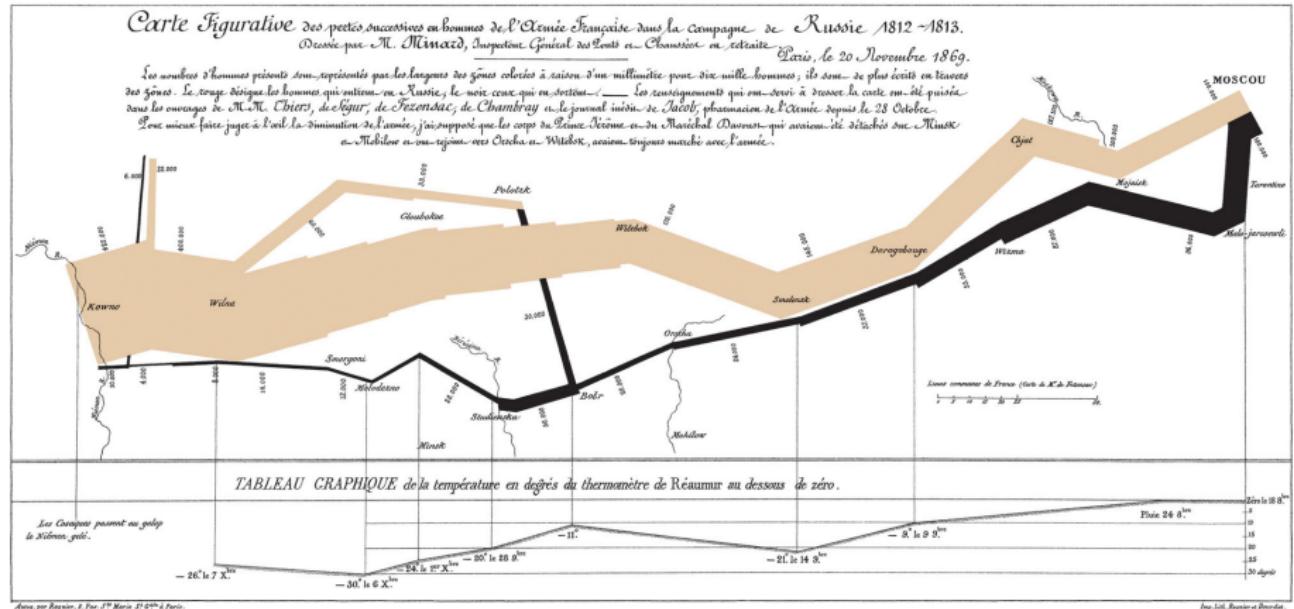
查尔斯·约瑟夫·密纳德

设计了大量的地图；其中，最为著名的一幅描绘了拿破仑入侵俄国的活动。

Charles Joseph Minard was a French civil engineer recognized for his significant contribution in the field of information graphics in civil engineering and statistics. Minard was, among other things, noted for his representation of numerical data on geographic maps.



统计图形 | 设计者 | 查尔斯·约瑟夫·密纳德



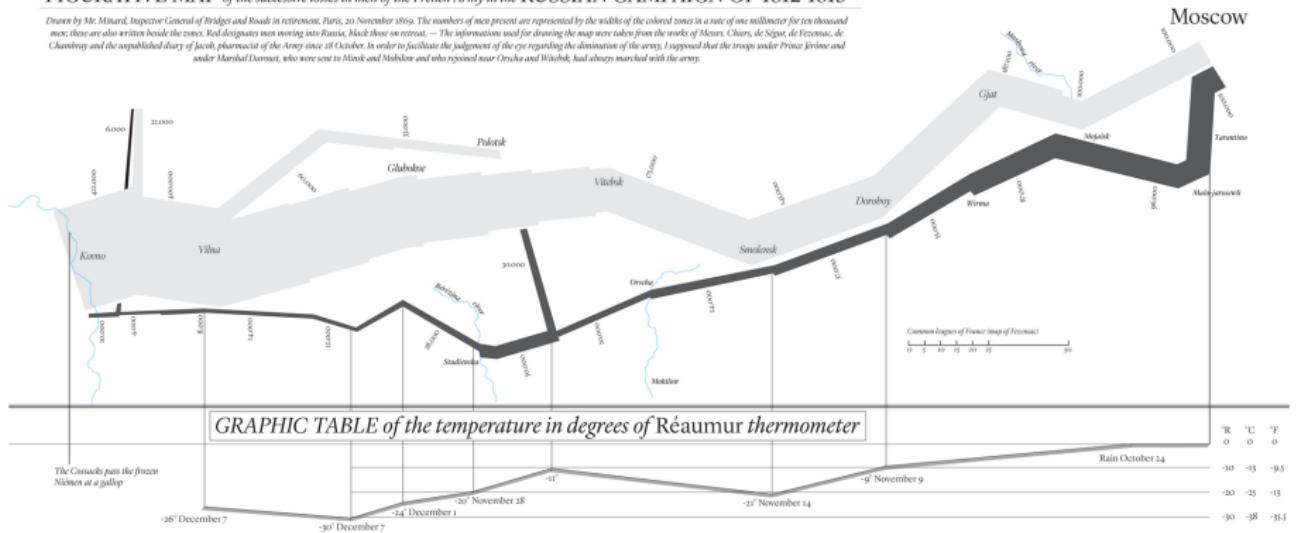
法国工程师查尔斯·约瑟夫·密纳德于 1861 年绘制的关于拿破仑入侵俄国的信息图形。



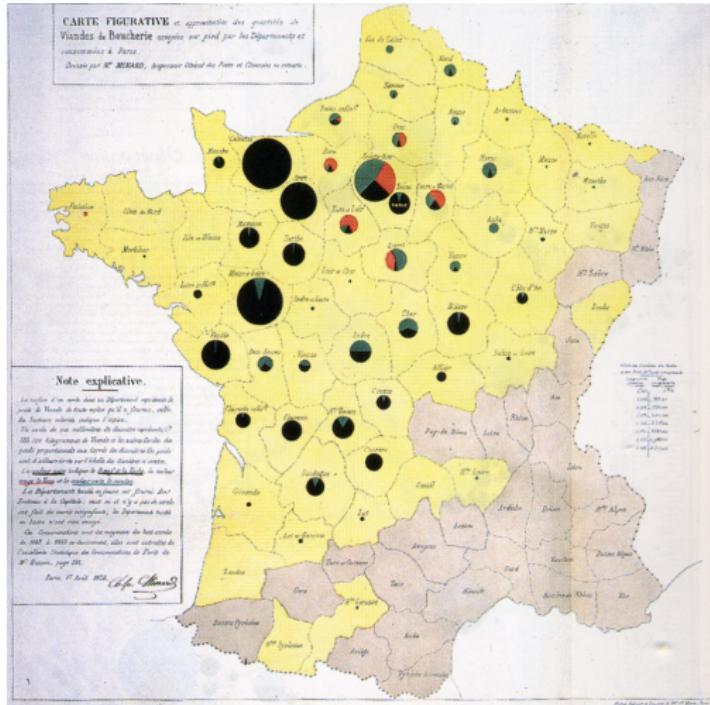
统计图形 | 设计者 | 查尔斯·约瑟夫·密纳德

FIGURATIVE MAP of the successive losses in men of the French Army in the RUSSIAN CAMPAIGN OF 1812-1813

Drawn by Mr. Minard, Inspector General of Bridges and Roads in retirement, Paris, 20 November 1869. The numbers of men present are represented by the widths of the colored zones in a rate of one millimeter for ten thousand men; these are also written beside the zones. Red designates men moving into Russia, black those in retreat. — The informations used for drawing the map were taken from the works of Messrs. Chicher, de Ségur, de Ferriac, de Chabrey and the unpublished diary of Jacob, pharmacist of the Army. In order to facilitate the judgment of the eye regarding the diminution of the army, I supposed that the troops under Prince Jérôme and under Marshal Davout, who were sent to Minsk and Mohilow and who rejoined near Oryol and Witebsk, had always marched with the army.



统计图形 | 设计者 | 查尔斯·约瑟夫·密纳德



Minard's map using pie charts to represent the cattle sent from all around France for consumption in Paris (1858).

奥图·纽拉特

设计了一种特殊类型的统计图形，称为同型图（isotype）。这种图形工具的具体目的就是，通过对群众进行形象的可视化教育，来实现社会变革。

奥图·纽拉特（Otto Neurath），奥地利人，是个科学家、社会学家及经济学家，在因纳粹而被迫逃离到英国前是维也纳学派的知名人物。

同型图

Isotype 图像符号，是奥图于 1925 年发表经过系统化设计的图画文字。“ISOTYPE”（International System of Typographic Picture Education）希望透过系统文的图像来取代文字，形成一种世界共通的语言，其本质类似于现在流行于网络的 MSN 表情符号等。

奥图曾试图用这种图画文字来取代沟通，虽然最后失败了，但是这种 Isotype 的概念，后来在 ICON、标示等设计领域，有着十分深远的影响。

奥图·纽拉特

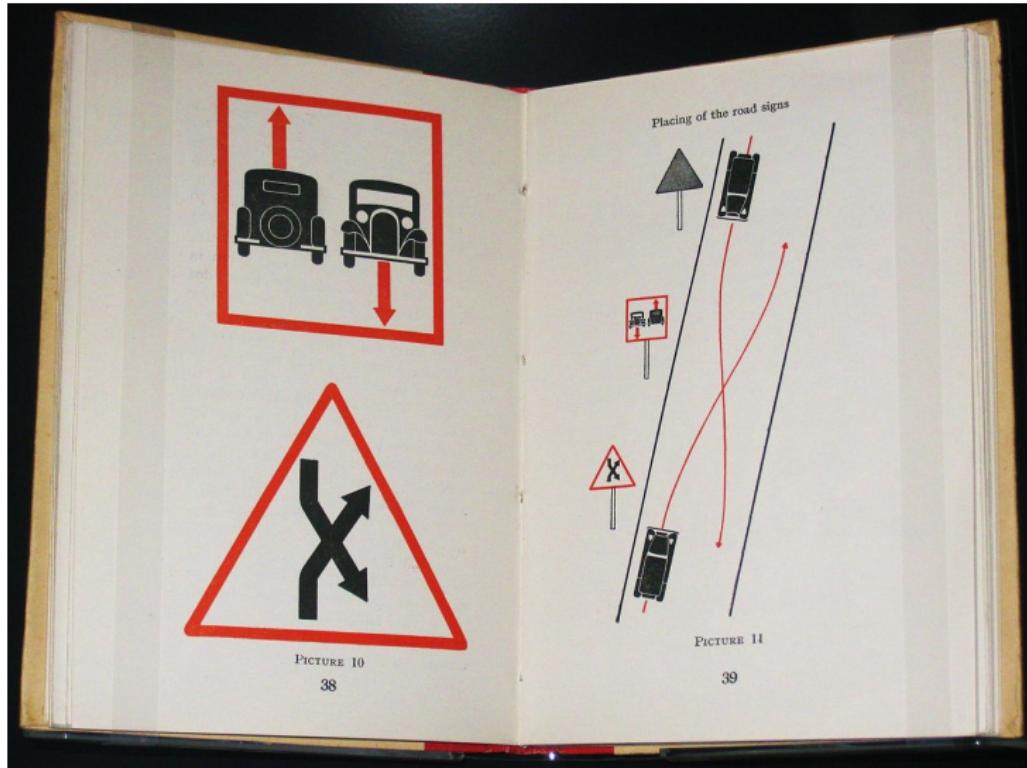
设计了一种特殊类型的统计图形，称为同型图 (isotype)。这种图形工具的具体目的就是，通过对群众进行形象的可视化教育，来实现社会变革。

奥图·纽拉特 (Otto Neurath)，奥地利人，是个科学家、社会学家及经济学家，在因纳粹而被迫逃离到英国前是维也纳学派的知名人物。

同型图

Isotype 图像符号，是奥图于 1925 年发表经过系统化设计的图画文字。“ISOTYPE” (International System of Typographic Picture Education) 希望透过系统文的图像来取代文字，形成一种世界共通的语言，其本质类似于现在流行于网络的 MSN 表情符号等。

奥图曾试图用这种图画文字来取代沟通，虽然最后失败了，但是这种 Isotype 的概念，后来在 ICON、标示等设计领域，有着十分深远的影响。



Pages from Neurath's International picture language, 1936.



- 1 统计图形
 - 引言
 - 简介
 - 著名设计者

- 2 知识拓展
 - 埃舍尔
 - 眼见未必为实
- 3 图说天下

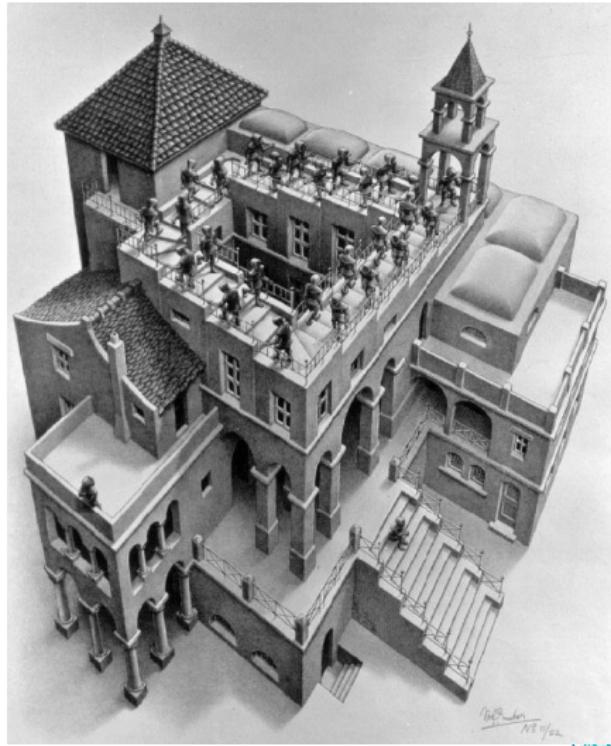
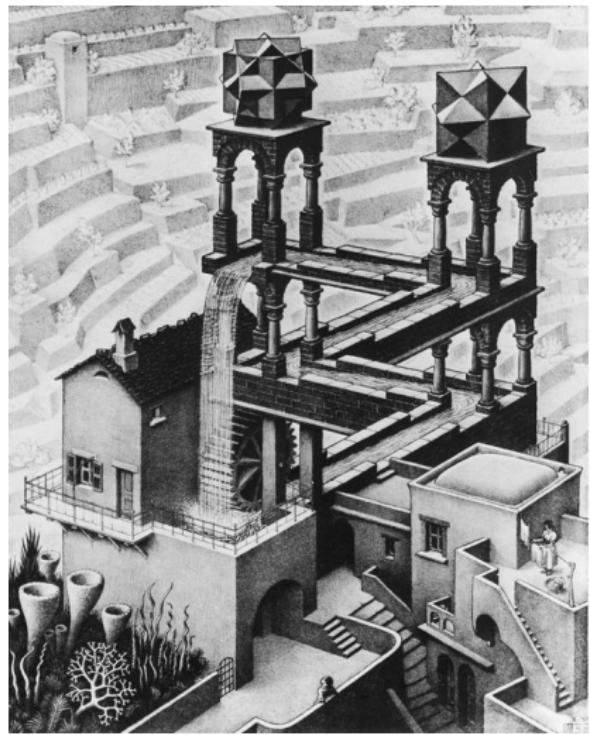


- 1 统计图形
 - 引言
 - 简介
 - 著名设计者

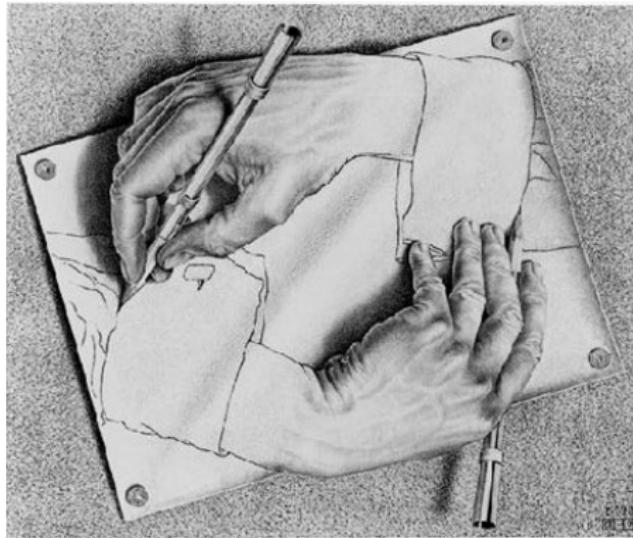
- 2 知识拓展
 - 埃舍尔
 - 眼见未必为实
- 3 图说天下



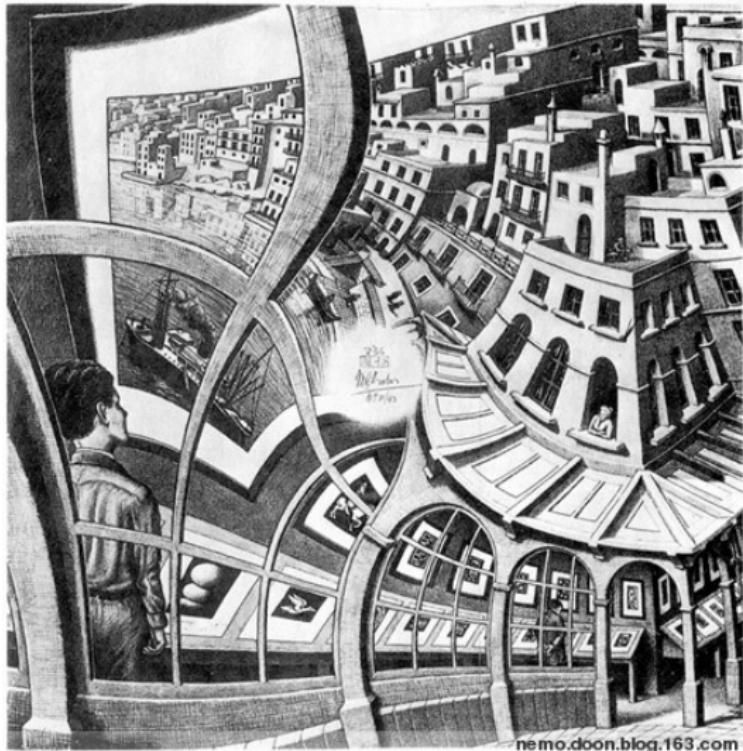
埃舍尔 | 作品



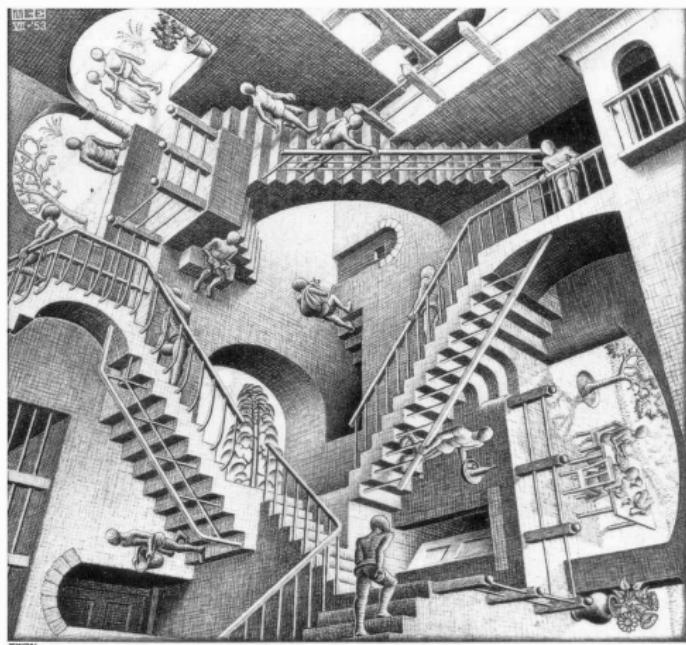
埃舍尔 | 作品



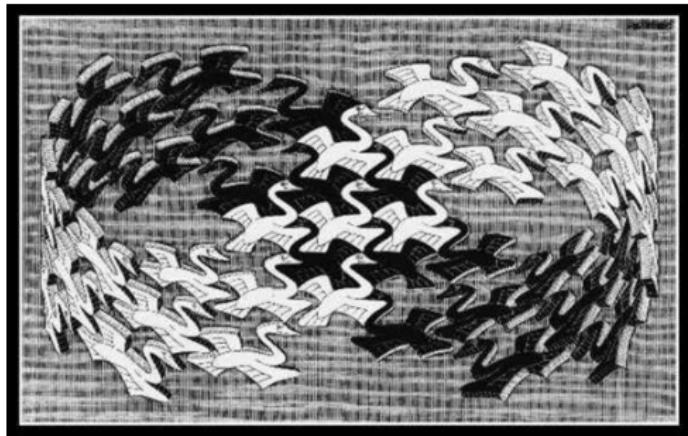
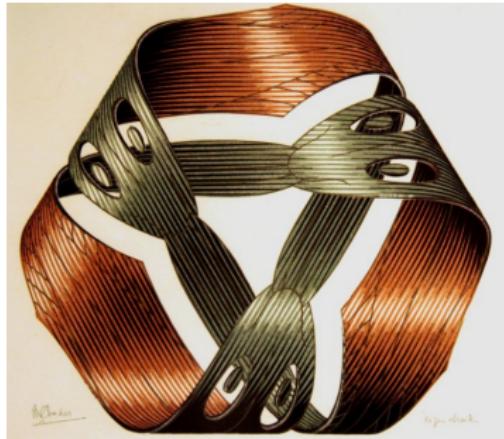
埃舍尔 | 作品



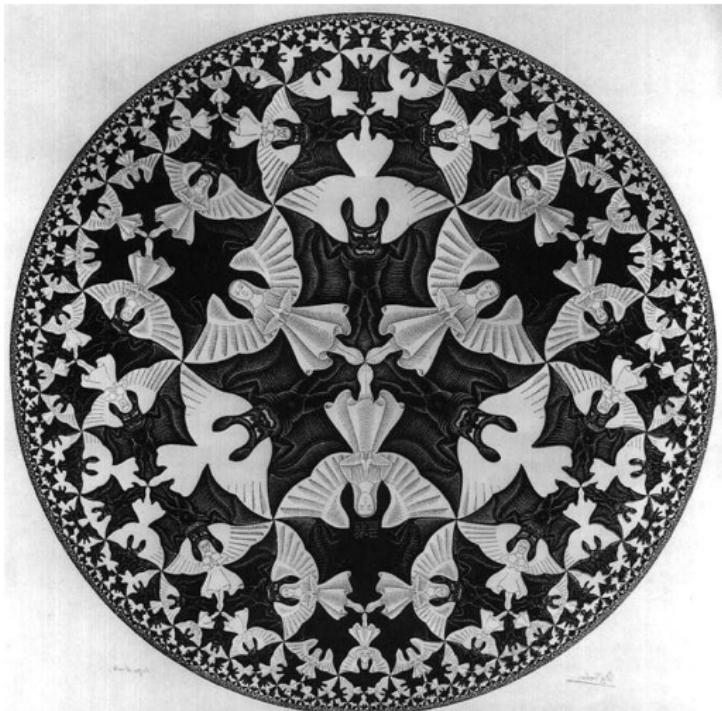
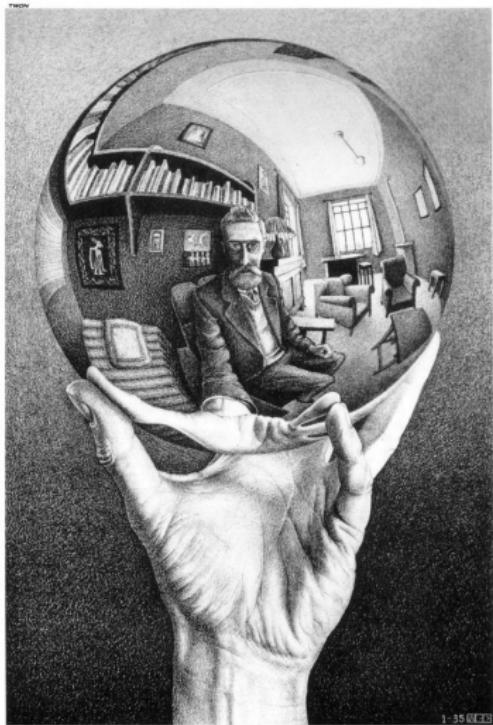
埃舍尔 | 作品



埃舍尔 | 作品

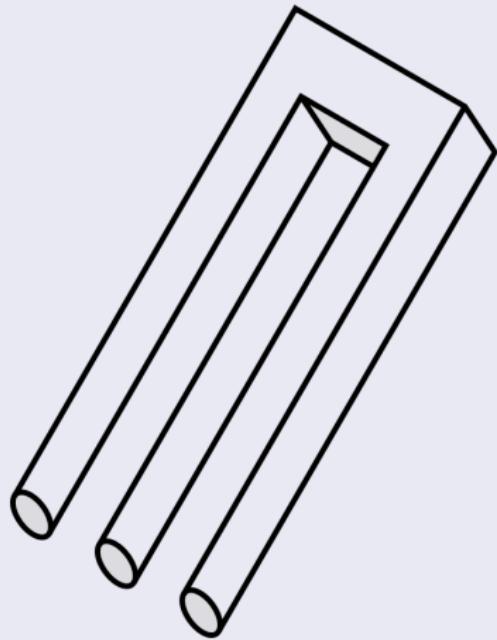


埃舍尔 | 作品

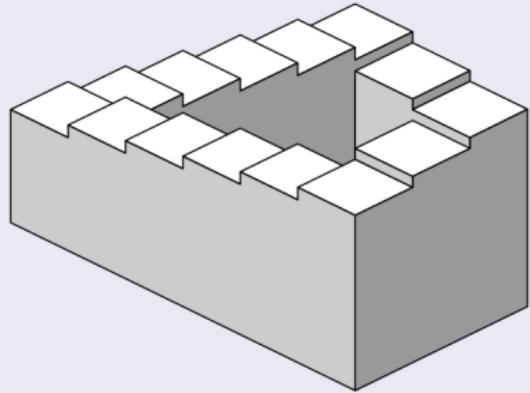


不可能的图形

魔鬼叉子

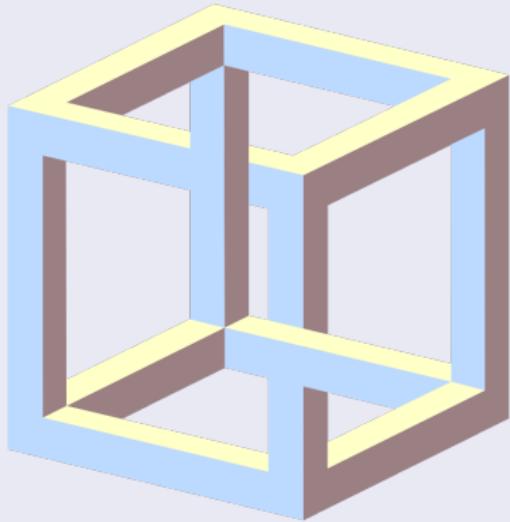


彭罗斯阶梯

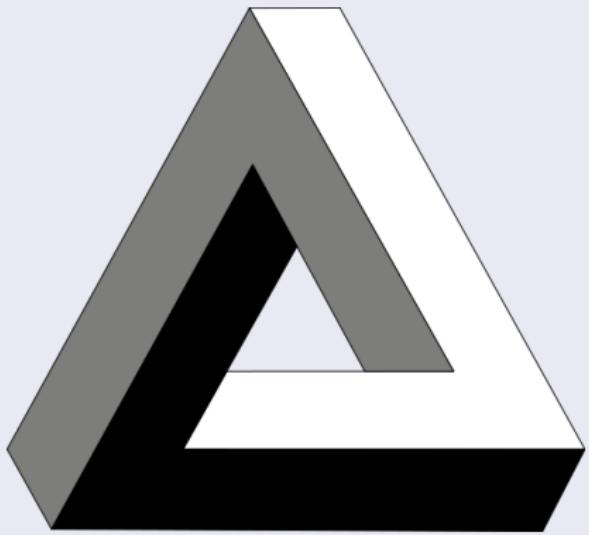


不可能的图形

不可能的方块



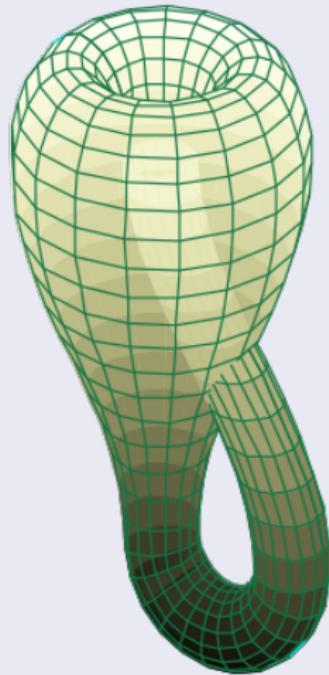
潘洛斯三角



莫比乌斯带



克莱因瓶



不可能的图片

不可能的图片

- 有趣的“不可能图形”：眼睛欺骗了你
- 不可能图片
- 不可能图形，眼见不一定为实
- 不可思议的幻觉图
- 不可能的物体





《哥德尔、埃舍尔、巴赫》

《哥德尔、埃舍尔、巴赫：集异璧之大成》（Gödel, Escher, Bach: an Eternal Golden Braid），是一本赢得普立兹奖的书。它是侯世达的著作，由 Basic Books 出版社在 1979 年出版的。这本书的二十周年版本在 1999 年发行，而且由侯世达加上新的前言。《集异璧之大成》（ISBN 7-100-01323-2）是商务印书馆在 1996 年出版的根据 1995 年英文版翻译的中文版。

本书的英文副标题意译为“一条永恒的金带”，其首字母与哥德尔、埃舍尔、巴赫三人的英文名字首字母 GEB 相同，而商务印书馆中文译本的副标题中的“集异璧”则与 GEB 谐音。本书一共有两篇，上篇译为“集异璧 GEB”，下篇译为“异集璧 EGB”。

本书主要讲述了逻辑学家哥德尔，艺术家埃舍尔，和作曲家巴赫的创造性的成就怎样交织在一起。正如作者所说：“我认识到，哥德尔、埃舍尔和巴赫只是用不同的方式来表达一样相同的本质。我尝试重现这种本质而写出这本书。”

侯世达

侯世达因其著作《哥德尔、埃舍尔、巴赫》获得普立兹奖（非小说类别）和美国国家图书奖（科学类别）。

侯世达定律

侯世达定律（Hofstadter's law）是一句自指的格言，由侯世达在《哥德尔、埃舍尔、巴赫》一书中提出：

侯世达定律：做事所花费的时间总是比你预期的要长，即使你的预期中考虑了侯世达定律。——侯世达，《哥德尔、埃舍尔、巴赫》

侯世达定律指做复杂任务需要花费的时间总是很难预计的。程序员经常会引用这一定律，特别是在进行有关提高效率的讨论时（如《人月神话》和极限编程）。其自指的特征反映了即便意识到任务的复杂性，预计花费的时间仍是困难的。

- 1 统计图形
 - 引言
 - 简介
 - 著名设计者

- 2 知识拓展
 - 埃舍尔
 - 眼见未必为实
- 3 图说天下



眼见未必为实



眼见未必为实



眼见未必为实



眼见未必为实



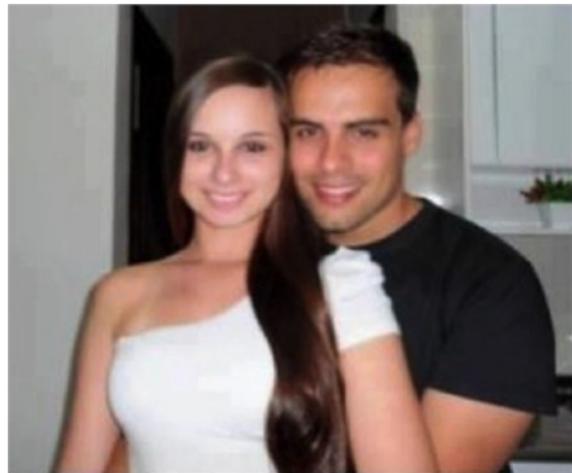
眼见未必为实



眼见未必为实



眼见未必为实



眼见未必为实



课后娱乐

- 27 张图告诉你眼见未必为实
- 15 Creative Venn Diagrams to Get You Thinking

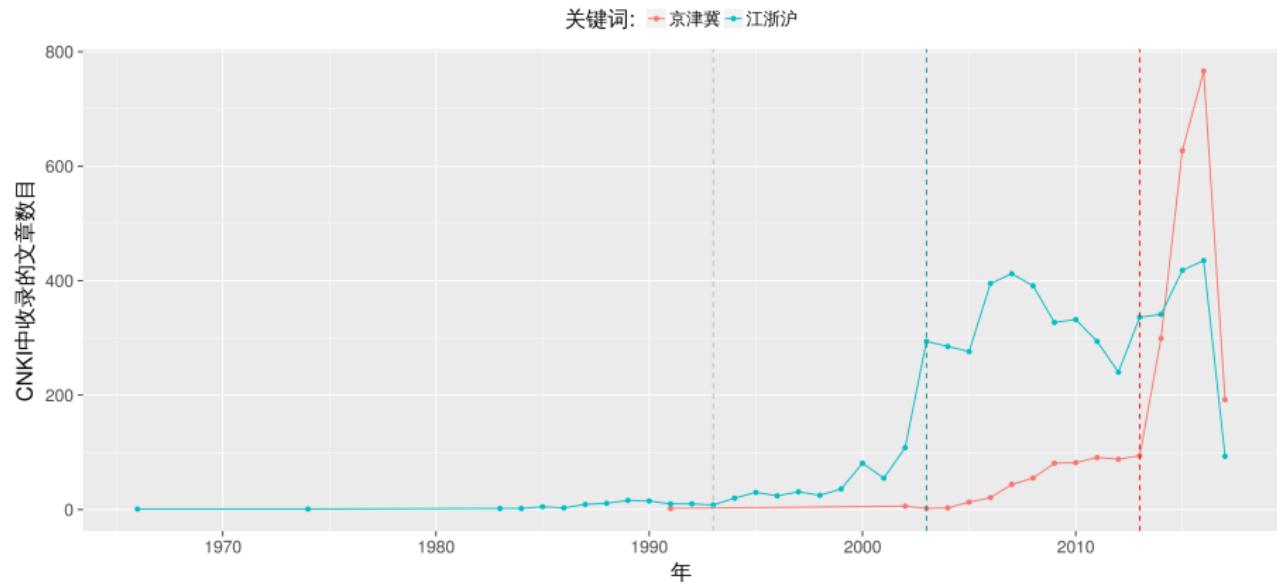


- 1 简介
- 引言
- 简介
- 著名设计者

- 2 知识拓展
 - 埃舍尔
 - 眼见未必为实
- 3 图说天下



京津冀 vs. 江浙沪



Powered by



T_EX L^AT_EX X_ET_EX Beamer

