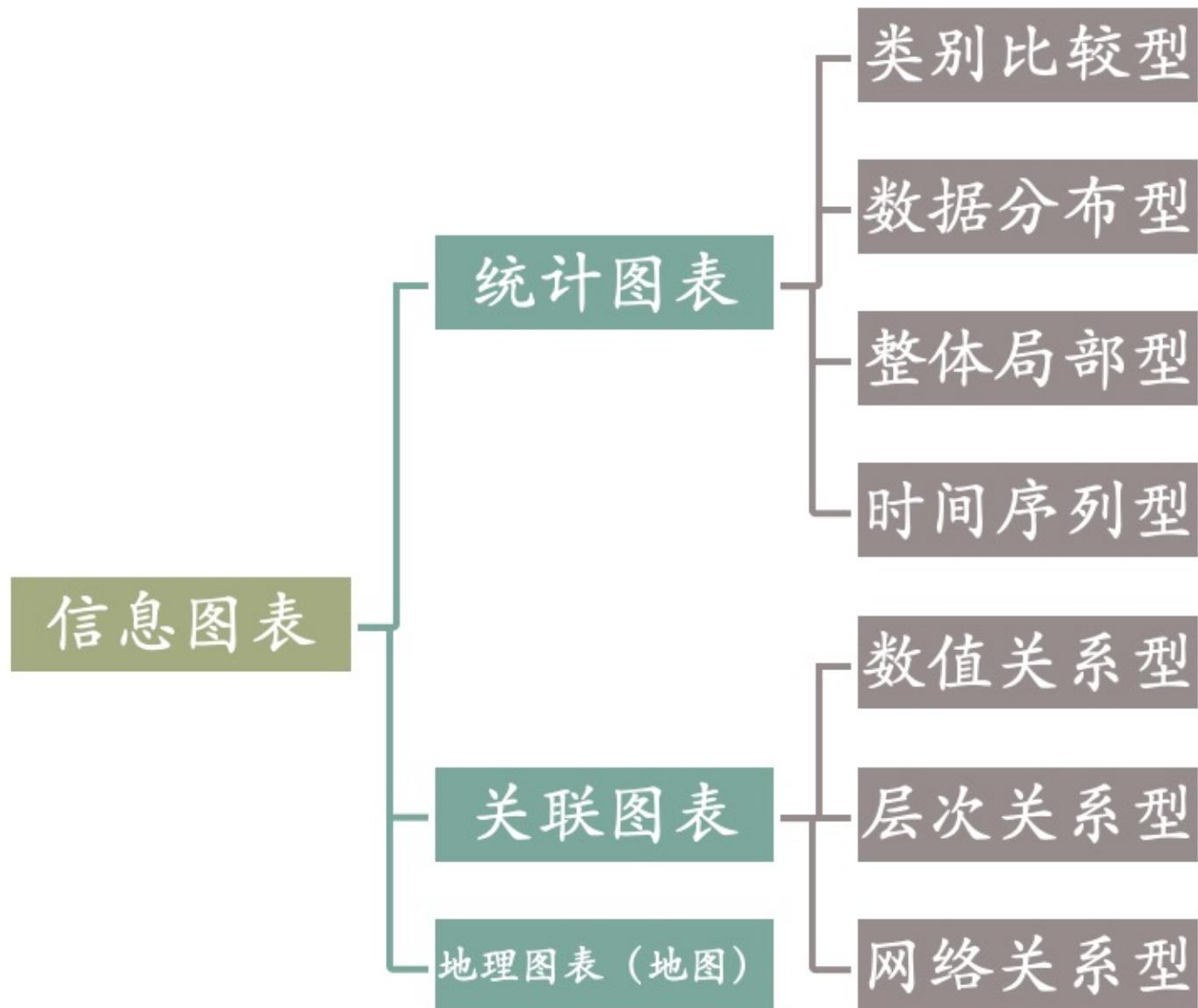


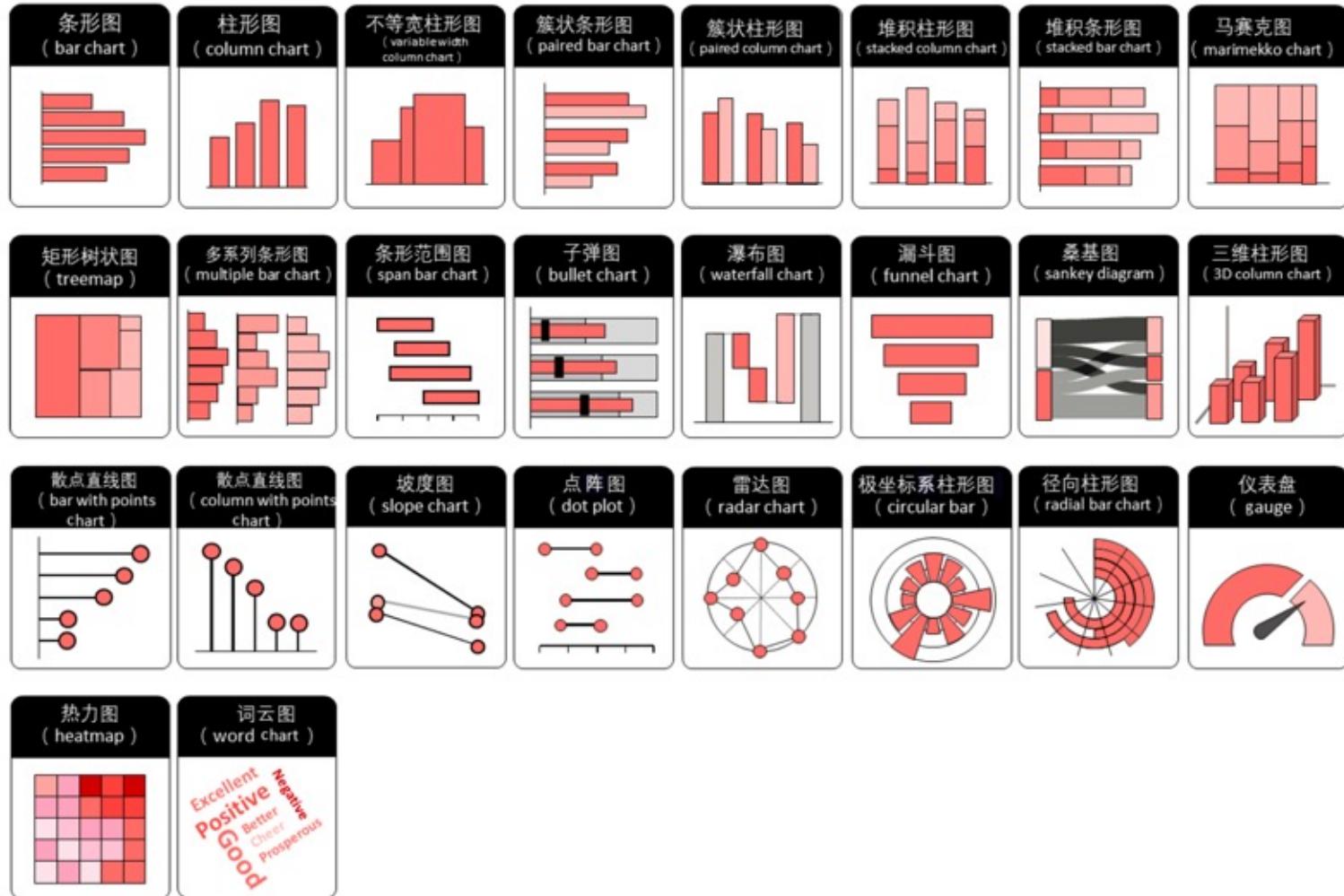
信息图表（继续）



信息图表

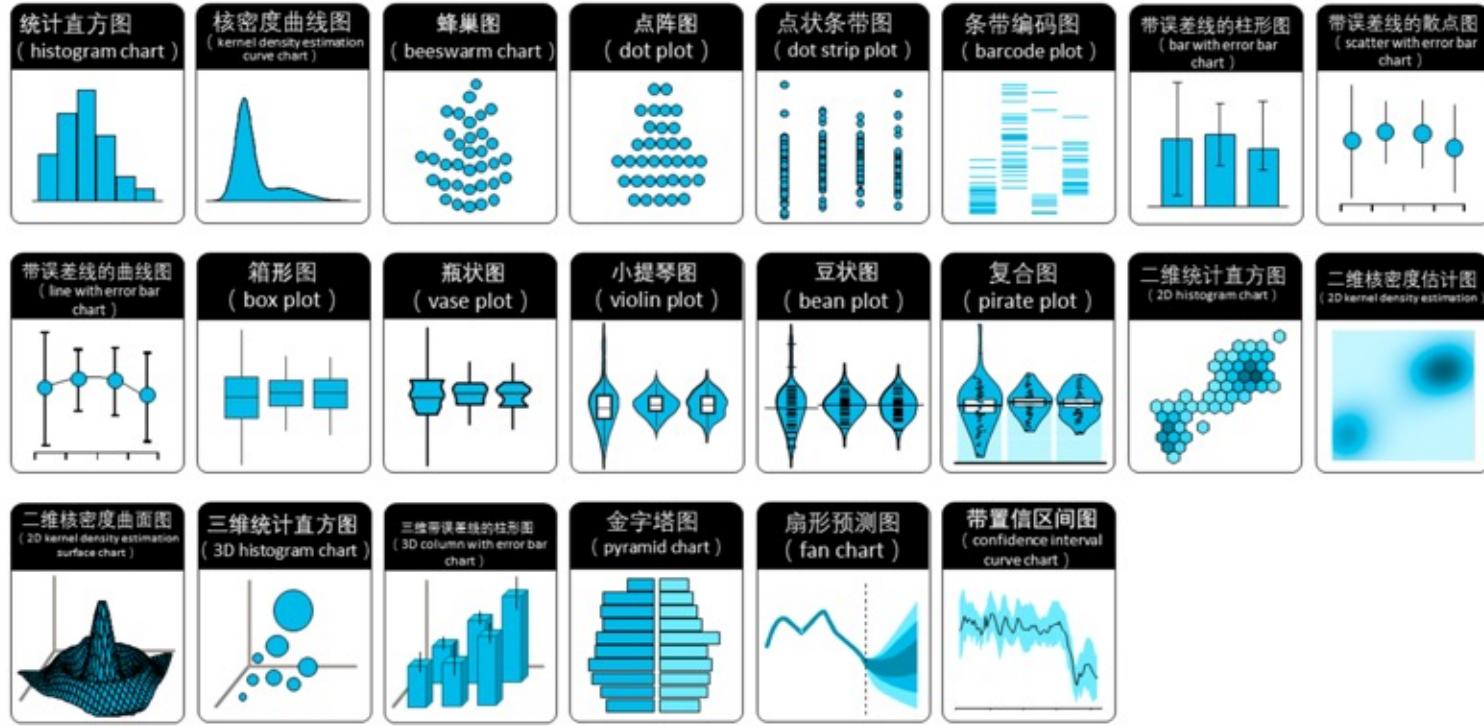


类别比较型图表



适用的变量数据类型：数值型vs类别型

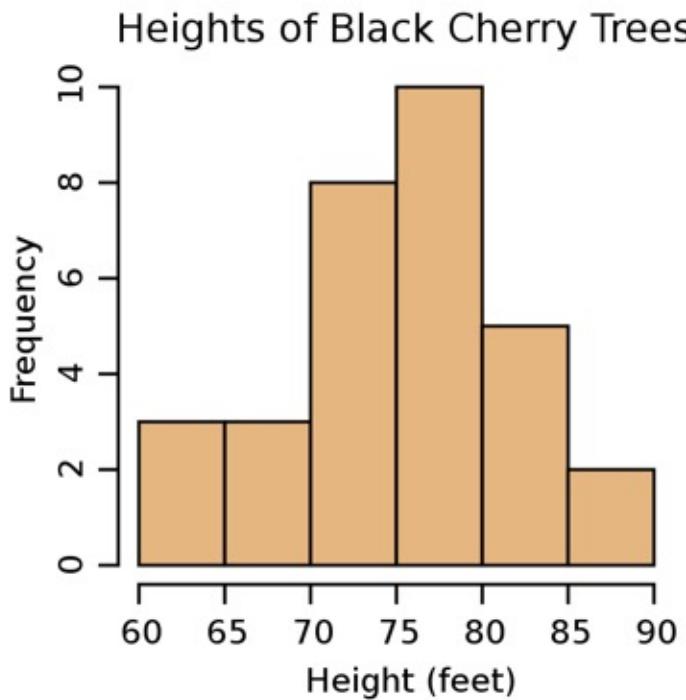
数据分布型图表



适用的变量数据类型：数值型（出现频率、分布规律）

直方图

- 统计直方图 (histogram, 频数分布直方图)
- X轴：数据的测量区间（分组）
- Y轴：各组对应的数据元频数M



直方图能解答的问题：

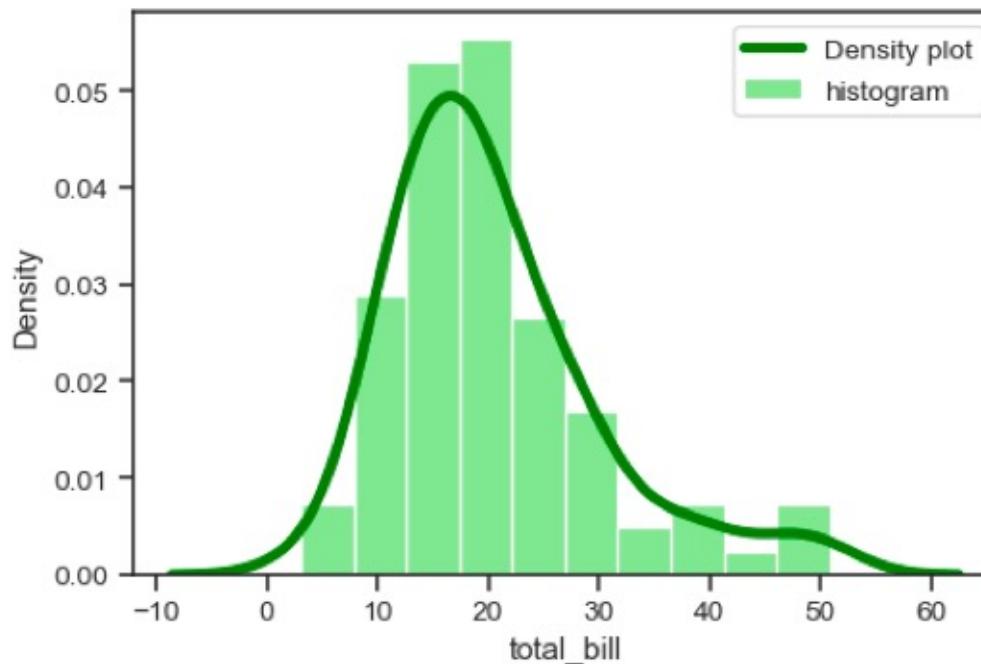
- 哪些组的数据比较集中？
- 哪些组的数据比较异常？
- 是不是有孤立的数据点？

直方图的基本参数：

- 组数：把数据分成了多少个测量区间。
- 组距：每一区间两个端点的距离，即每组的值域。
- 频数：各组内数据元的数量和组距之比。

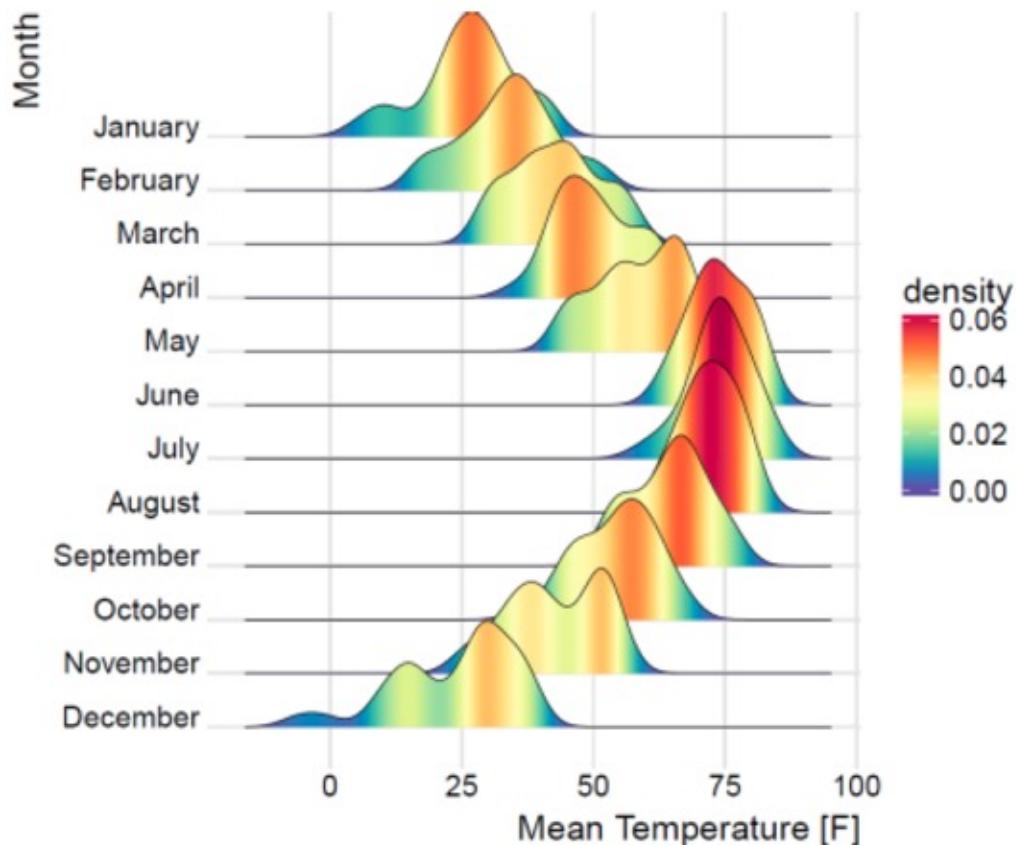
核密度曲线图（密度图）

- 和直方图相关，用于显示数据元在X轴的连续数据段内的分布状况（把直方图转化为平滑曲线，得出数据元更平滑的分布模式）。
- 不受分组数量的影响，能更好地体现数据分布的整体形状。



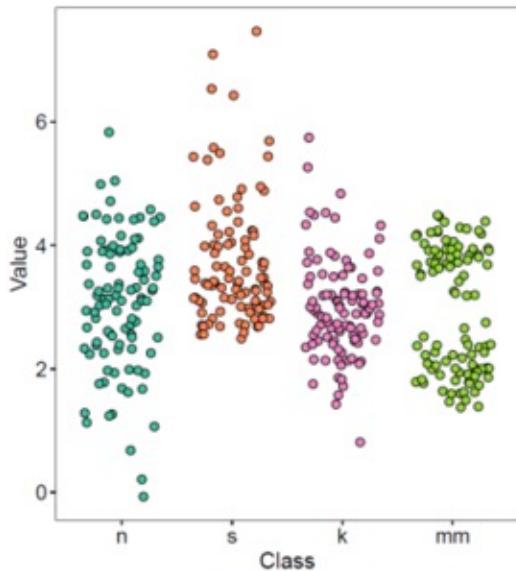
核密度曲线图（密度图）

密度图的变种：峰峦图——应用于多数据系列的核密度可视化

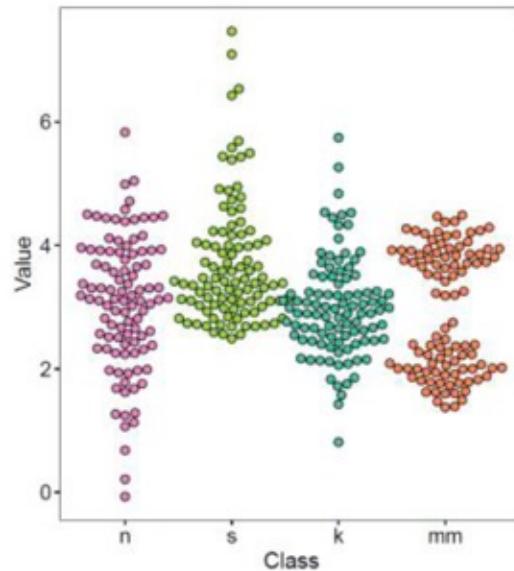


抖动散点图、蜂巢图和点阵图

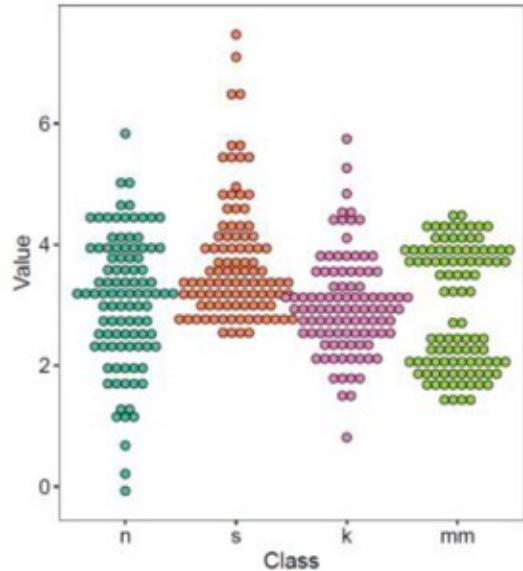
- 用散点图的方式展示数据的分布规律。
- X轴为分类变量，Y轴为数值变量。



(a) 抖动散点图



(b) 蜂巢图

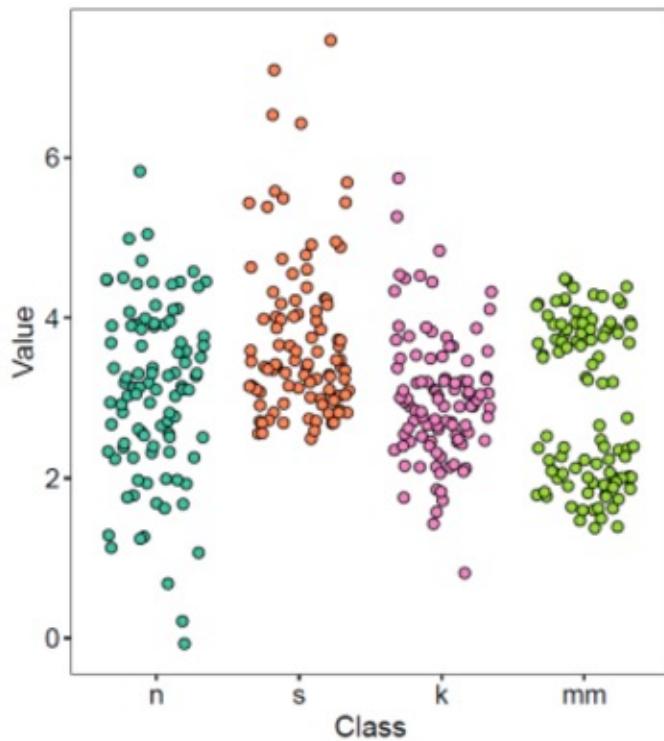


(c) 点阵图

抖动散点图、蜂巢图和点阵图

抖动散点图 (jitter chart) : 每个数据元的Y坐标为数值变量, X坐标在对应类别标签中心线两侧一定范围内随机生成。

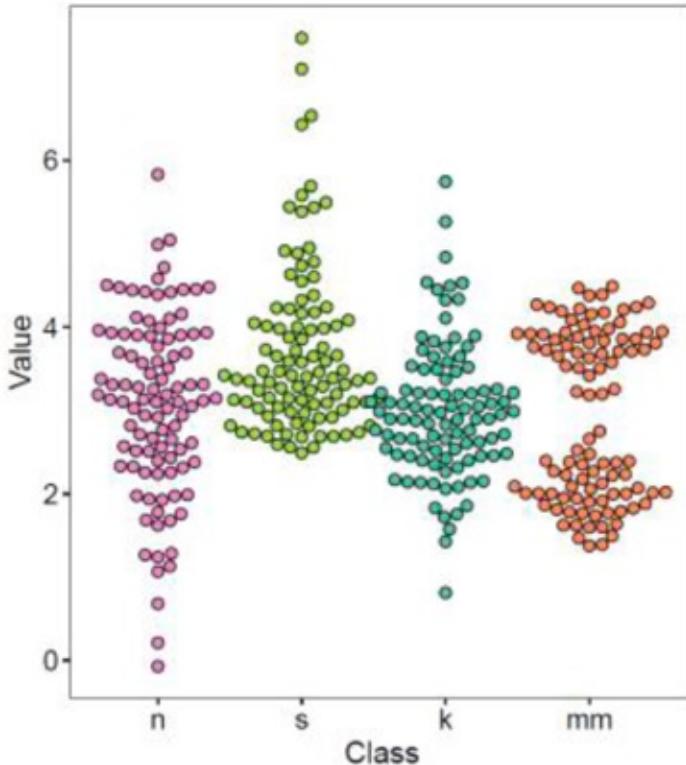
- 缺点: 容易有重合的点。



抖动散点图、蜂巢图和点阵图

蜂巢图（beeswarm chart）：数据元X坐标沿类别标签中心线向两侧对称地展开，并偏向斜上方。

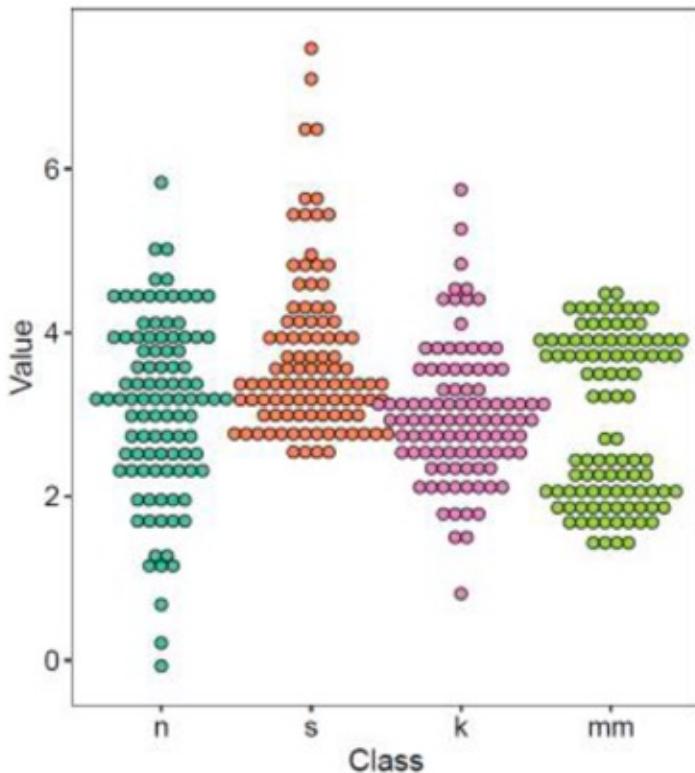
- 缺点：如果某个数值的数据点较多，其排布容易“跨区”



抖动散点图、蜂巢图和点阵图

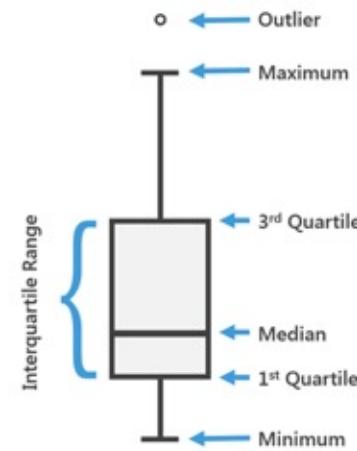
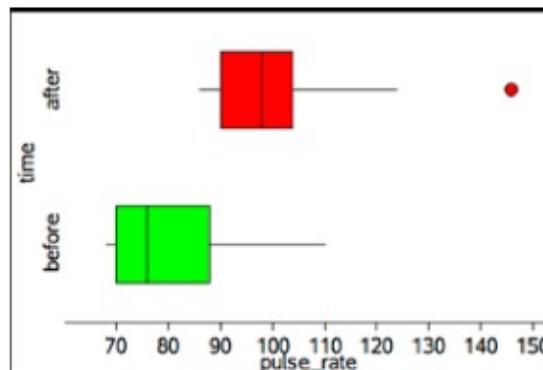
点阵图 (dot plot) , 数据元的X坐标沿类别标签中心线向两侧均匀且对称地展开，不偏向斜上方。

- 缺点：和蜂巢图类似，容易“跨区”。



箱线图

- 箱线图（box-whisker plot，也称箱须图、箱型图、盒子图）：不直接绘制数据元，而是先进行统计处理，再绘制统计结果。
- 显示一组数据的最大值、最小值、中位数，以及上下四分位数。反映一组或多组连续型数据的分布情况，包括其中心位置和散布范围。
 - 上、下四分位数之间的数据元分布（中间50%数据元）用“箱子”（box）显示（即四分位距IQR），并用线条展示中位数。
 - 上、下四分位数以外的数据元分布范围用线段显示，被称为“晶须”（whisker）。
 - 异常值（outlier）用离散点显示。

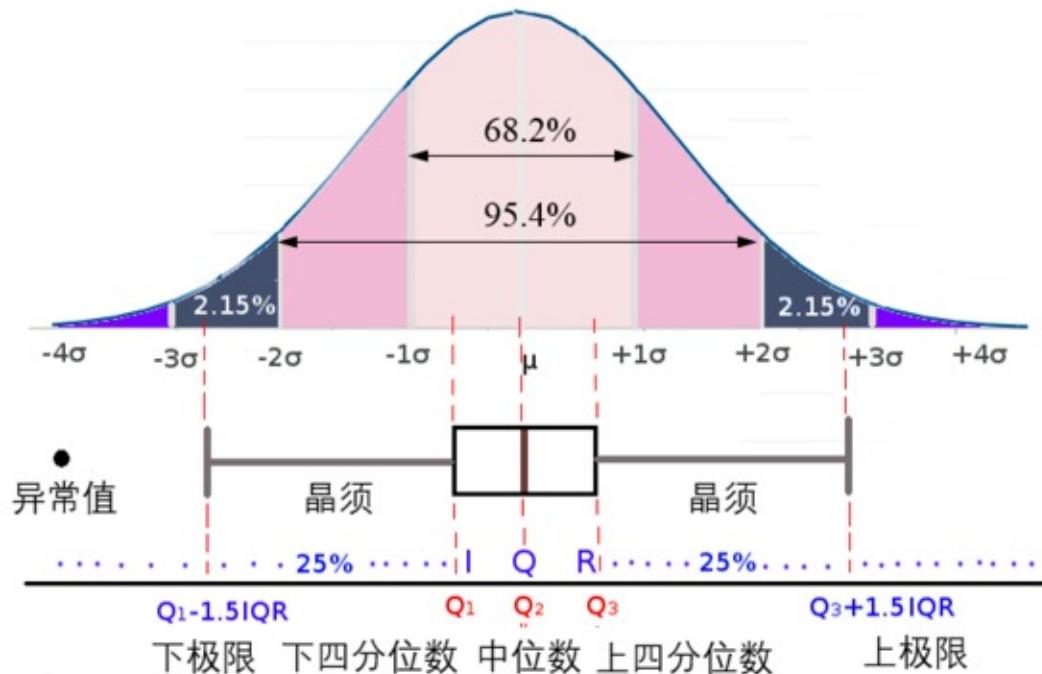


John W. Tukey

箱线图

四分位数 (quartile) : 把所有数值由小到大排列，并分成四等份，处于三个分割点位置的数值。中位数为其中之一。

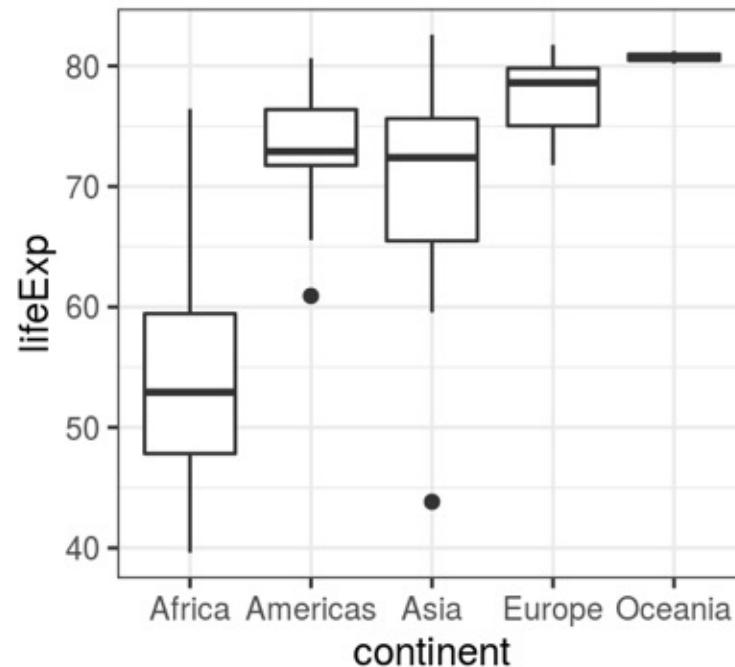
Q1和Q3的差距：四分位距 (Inter Quartile Range, IQR)。当数据符合正态分布时，IQR乘以0.7413为标准化四分位距 (Normalized IQR)，相当于标准差 (σ)。



箱线图

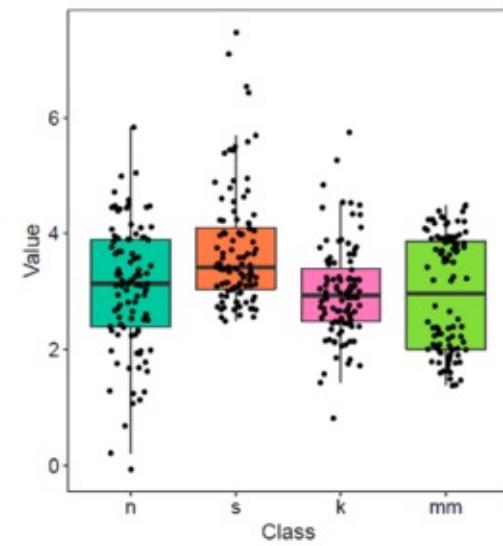
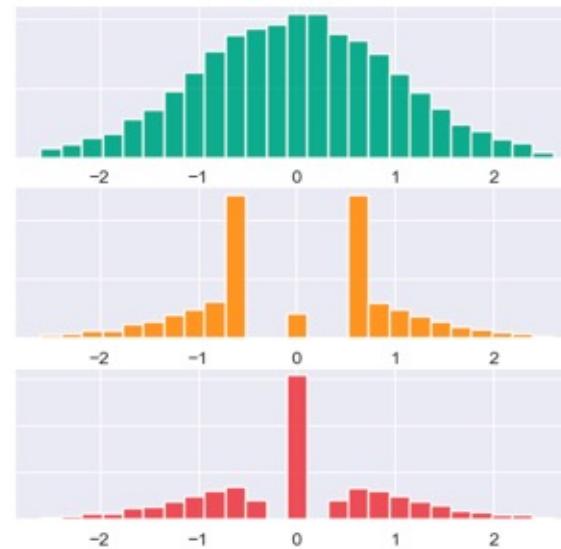
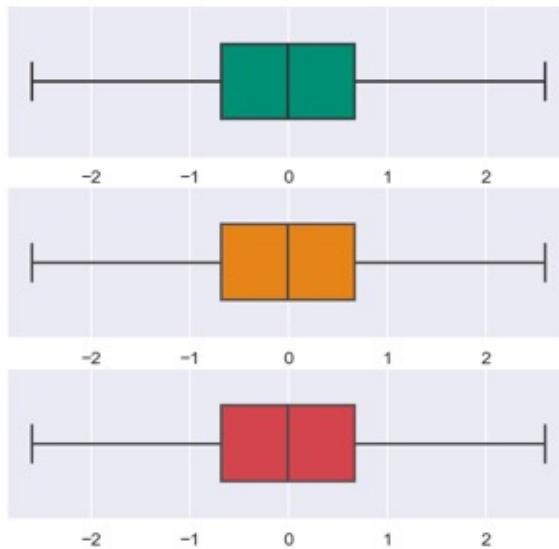
箱线图的用处

- 读取数据分布的关键信息（中位数和上下四分位数）
- 识别离群值（以及它们的数值）
- 观察数据分布是否对称
- 数据分布的紧密程度
- 对比各组数据



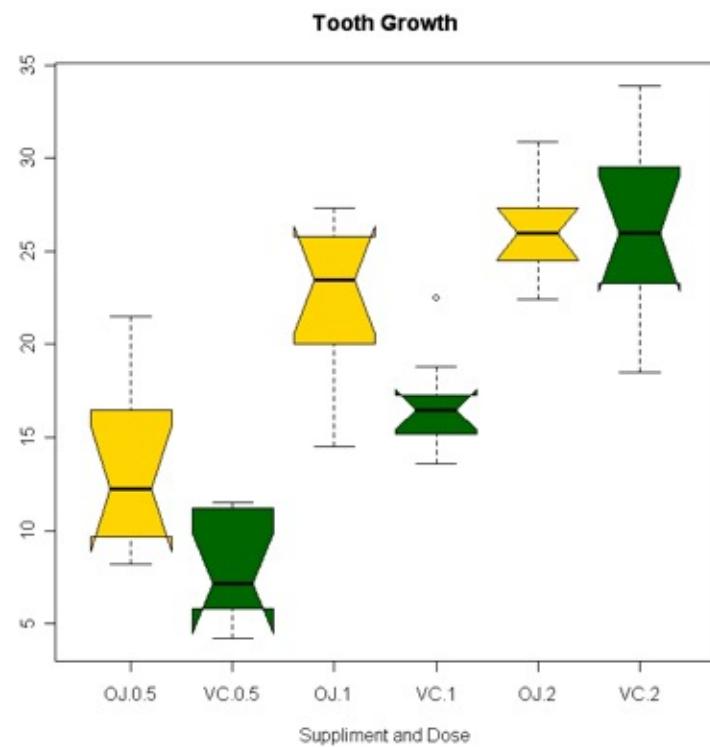
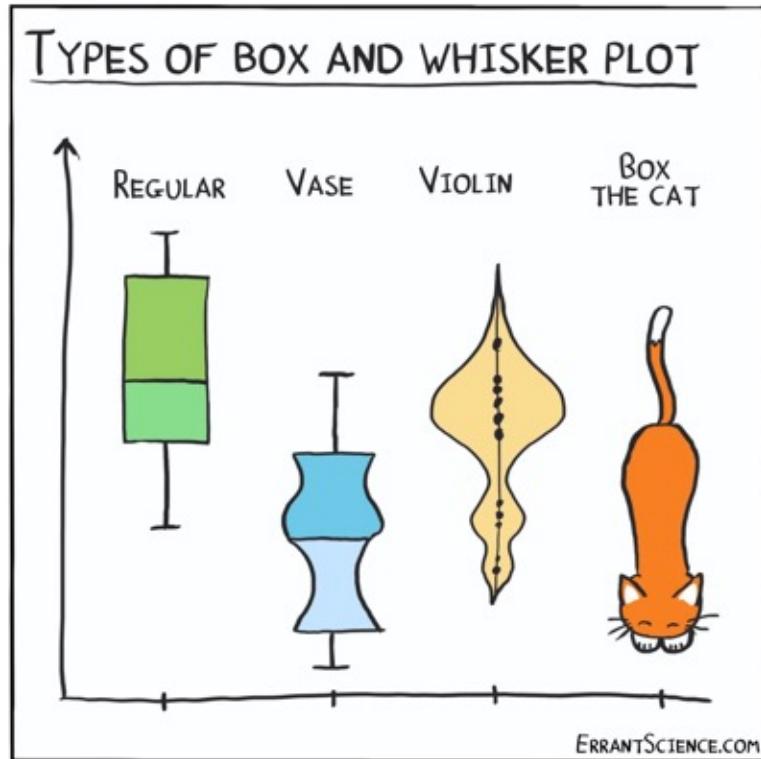
箱线图

- 缺点：不体现数据的具体分布，所以无法适用于双峰及多峰分布的数据
- 解决方案：
 - 与抖动散点图结合
 - 瓶状图、小提琴图



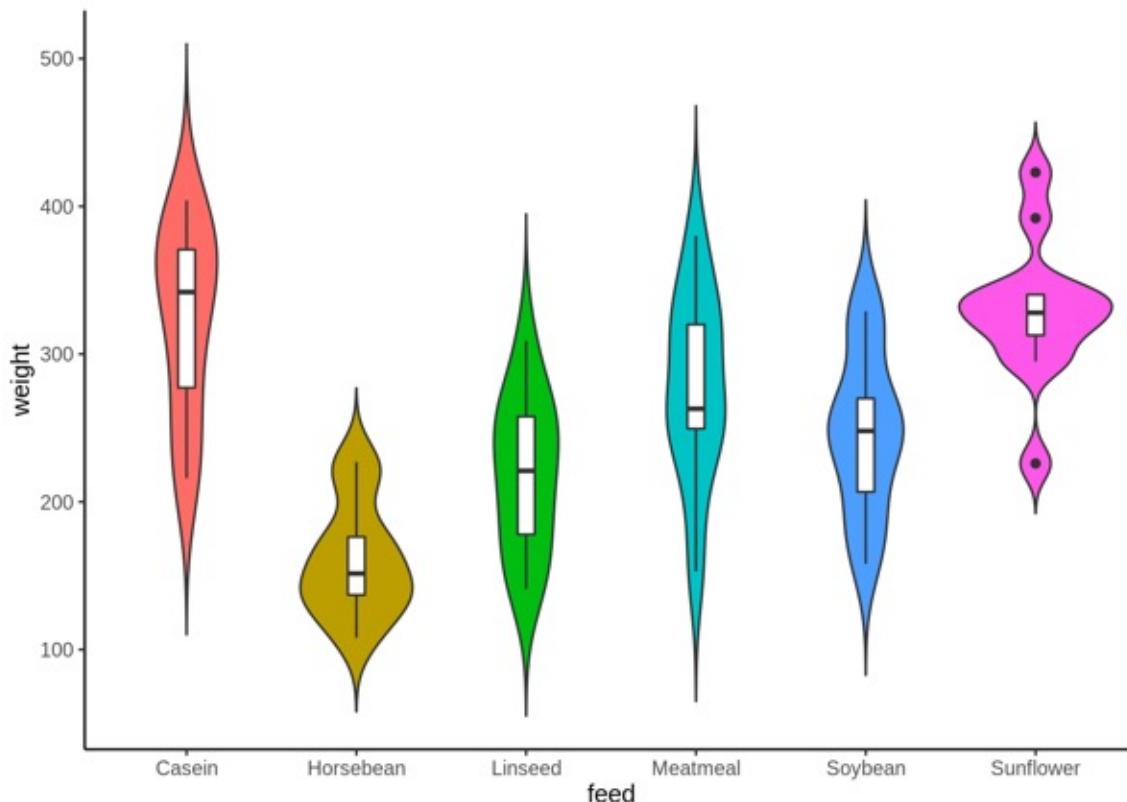
瓶状图和小提琴图

- 瓶状图 (vase plot) : 用核密度曲线替代箱形部分，显示数据的分布形状

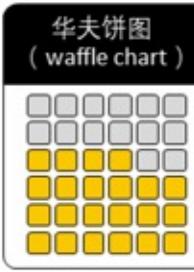
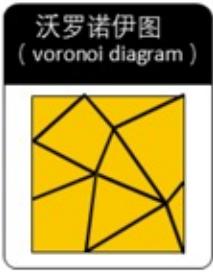
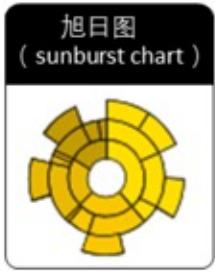
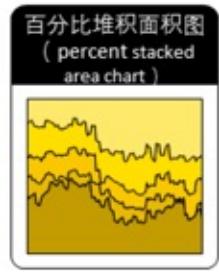
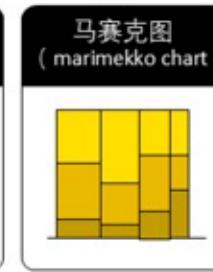
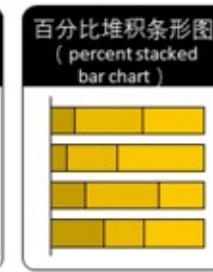
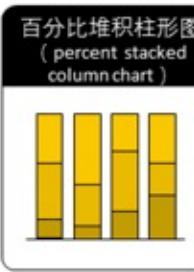
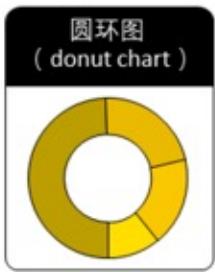
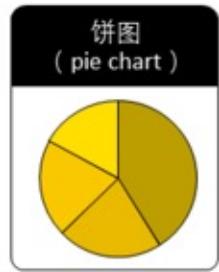


瓶状图和小提琴图

- 小提琴图 (violin plot) : 结合了箱形图和瓶状图的特征，显示数据的分布形状。



整体局部型图表

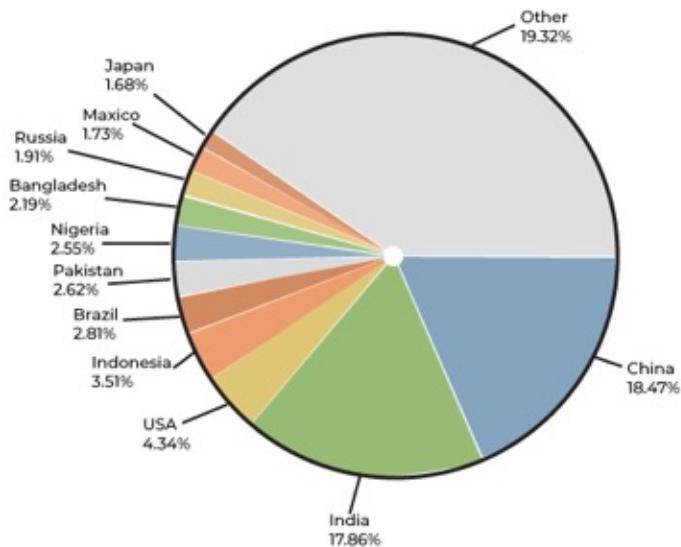


局部成分在整体中的占比

饼图

- 饼图 (pie chart) :通过弧度大小来对比整体中的各个分类。用于快速了解数据的占比分配。

Population Pie Chart



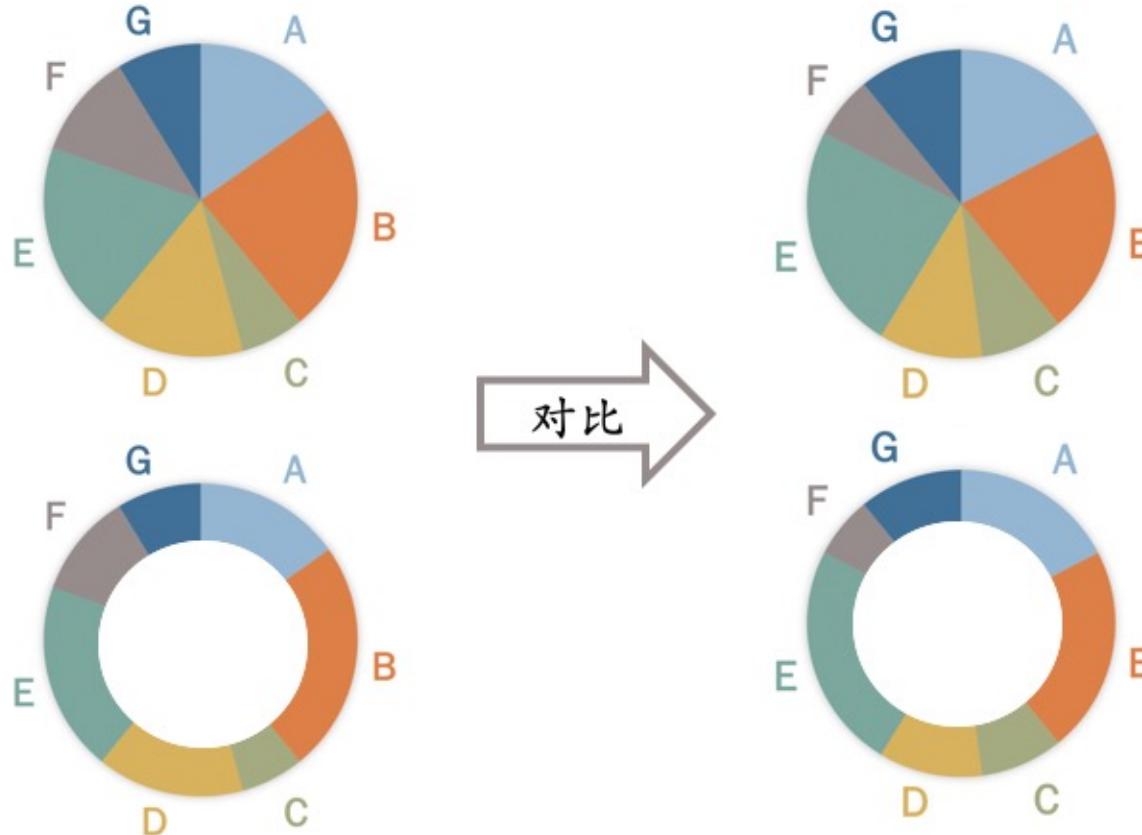
饼图的缺点:

- 不适用于分类过多的数据，原则上不可多于9个分类。
- 占据较大的画布空间。
- 多个饼图之间难以进行数值比较。
- 不适合展示多变量连续型数据的占比。

圆环图

又叫“甜甜圈图”(donut chart)，其本质是将饼图挖空。

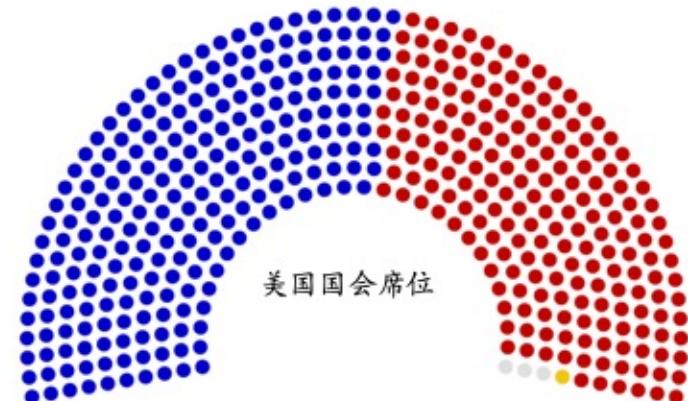
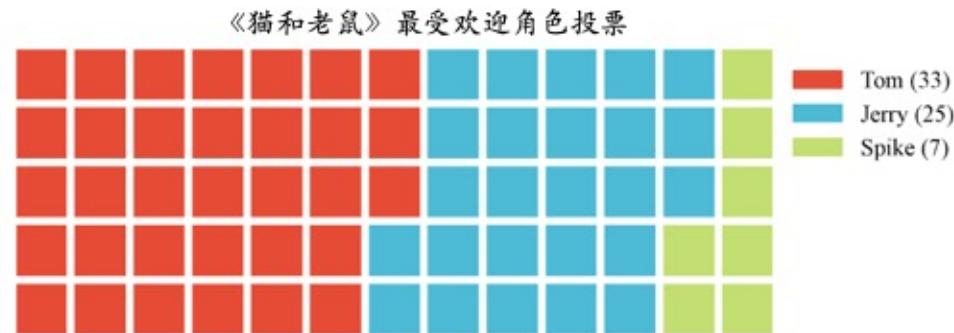
- 两个饼图之间的对比：对比的是扇形的面积。
- 两个圆环图之间的对比：对比的是圆弧的长度。
- 圆环图相对于饼图空间的利用率更高。



华夫饼图

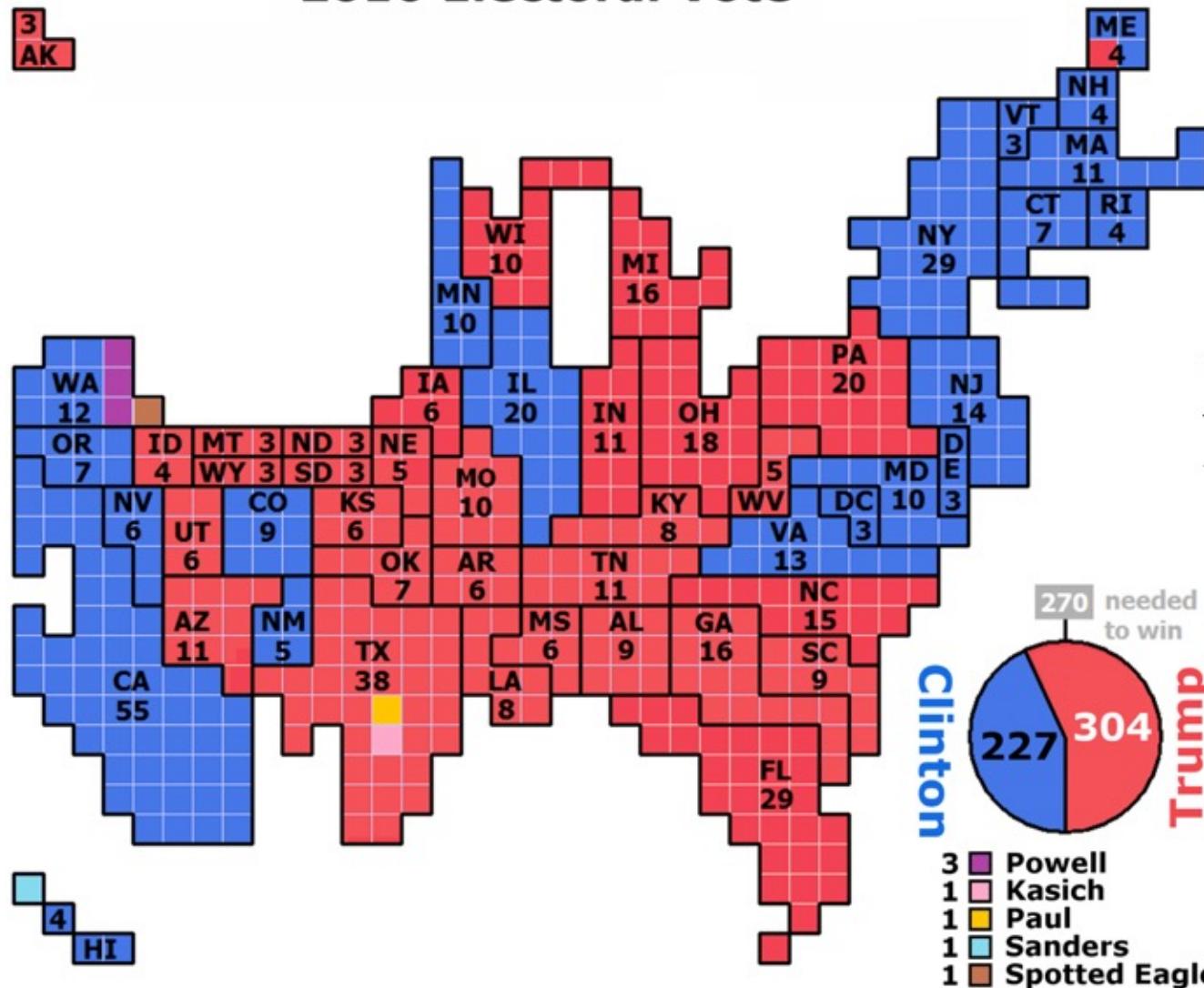
华夫饼图 (waffle chart) : 展示总数据的分组情况。

- 适用于快速展示数据集中不同类别的分布和比例。
- 有利于和其他数据集的分组比例进行比较。
- 常用于展示国会席位、投票状况、产品销量等内容。

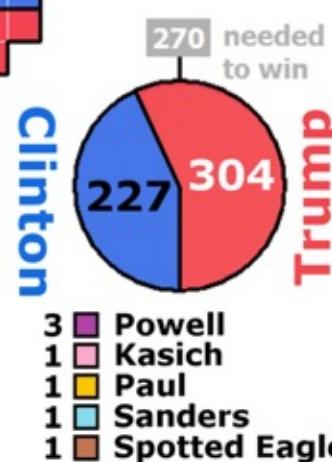


华夫饼图

2016 Electoral Vote



美国2016年大选
各州票数分布及投票情况



Pop Quiz

回答正确得2分
回答错误得1分
缺席不得分

1. 纹理是多种视觉变量的组合，其中不包括以下哪一种？

- A. 形状 | B. 颜色 | C. 方向 | D. 位置

2. 颜色类视觉通道中，哪一种最适于编码定性/分类属性？

- A. 饱和度 | B. 色调 | C. 亮度 | D. 对比度

3. 查看多个分类变量之间的差异及相关性，应该用哪种图表？

- A. 表格型热力图 | B. 雷达图 | C. 不等宽柱形图 | D. 词云图

4. 判断视觉通道表现力的标准不包括下面哪一项？

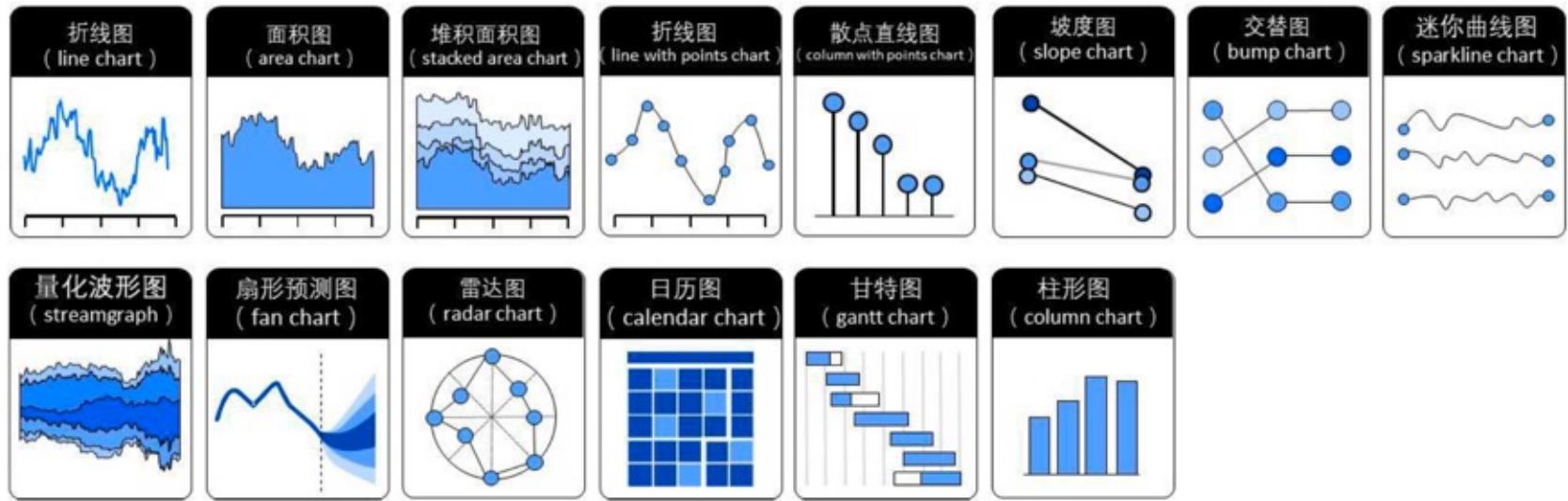
- A. 可分辨性 | B. 精确性 | C. 视觉突出性 | D. 统一性

5. 在小尺寸图像里，当饱和度作为视觉通道时，其层次不宜超过多少个？

- A. 2个 | B. 3个 | C. 4个 | D. 5个



时间序列型图表



时变数据

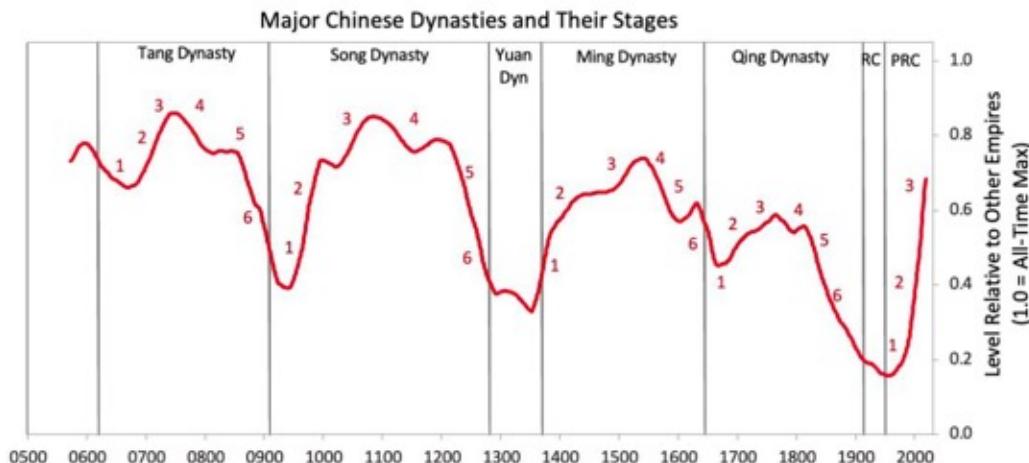
时变数据 (time-varying/temporal data)

- 时间序列数据 (time series data) [狭义的时变数据]
 - 和时间轴直接相关
 - 数据元可被视为事件，其发生时间可视为一个变量
 - 股票交易、历年气温变化、体育赛事的赛程、监控摄像头拍摄的视频
- 流数据 (streaming data) [特殊的时变数据]
 - 具有无限长度的时间轴
 - 以流量模式生成，常见于互联网和传感器网络
- 有序数据/顺序数据 (sequential data) [广义的时变数据]
 - 具有内在的先后顺序、排序逻辑
 - 其变化顺序可以映射为逻辑推进轴，类似于时间轴
 - DNA测序、元素周期、光谱
- 时变数据特点：体量大、维数高、变量多，类型丰富、分布范围广泛

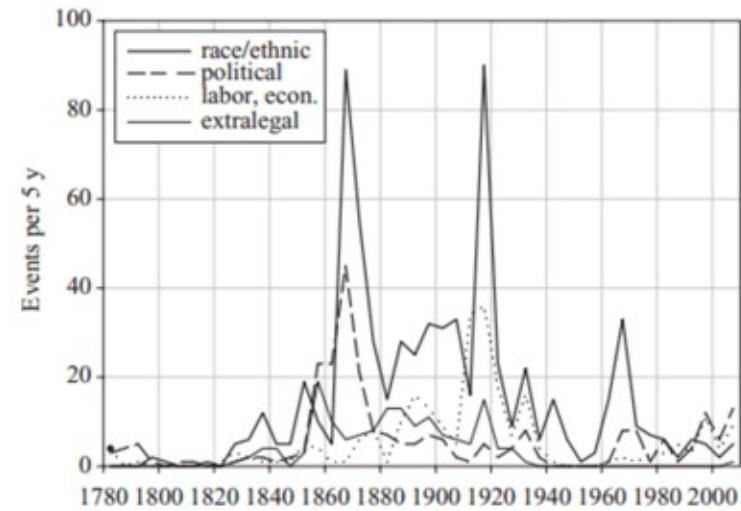
时变数据

对时序数据的可视化：发现历史、社会、经济规律。

- 历史动力学 (Cliodynamics)：人类社会的微观活动和历史事件构成了时变数据，其循环规律分为世俗周期（200-300年周期）和社会稳定周期（50年为周期）。



中国古代王朝的历史周期律

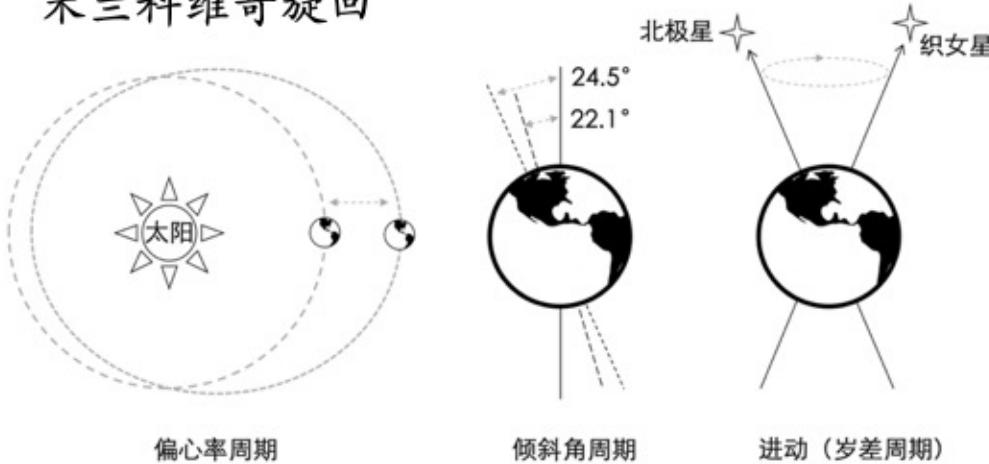


美国城市暴力活动的周期

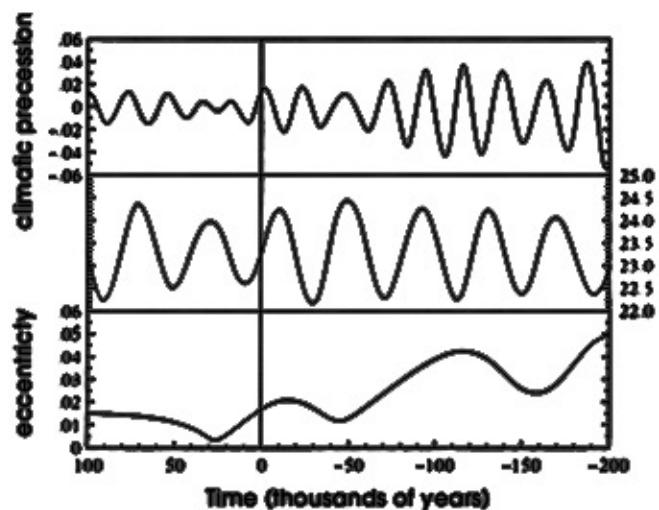
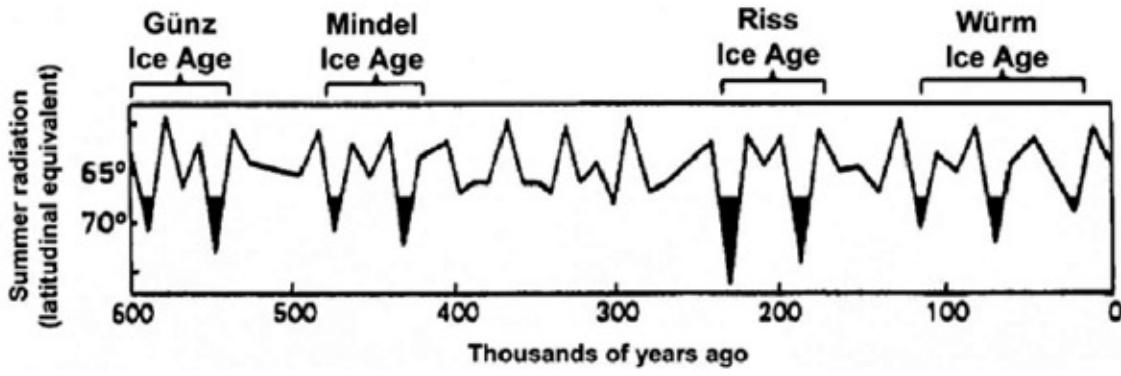
时变数据

对时序数据的可视化：发现自然界的规律。

米兰科维奇旋回



Milutin Milankovitch



时变数据

时变数据分析的内容

信号分解、模式挖掘、滤波、特征预测等方法

序列内的极值

序列间相似性

和其他数据的
匹配性

检索数据元

数据元的变化

数据元的出现
频次

相似片段
(周期)



数据量大时

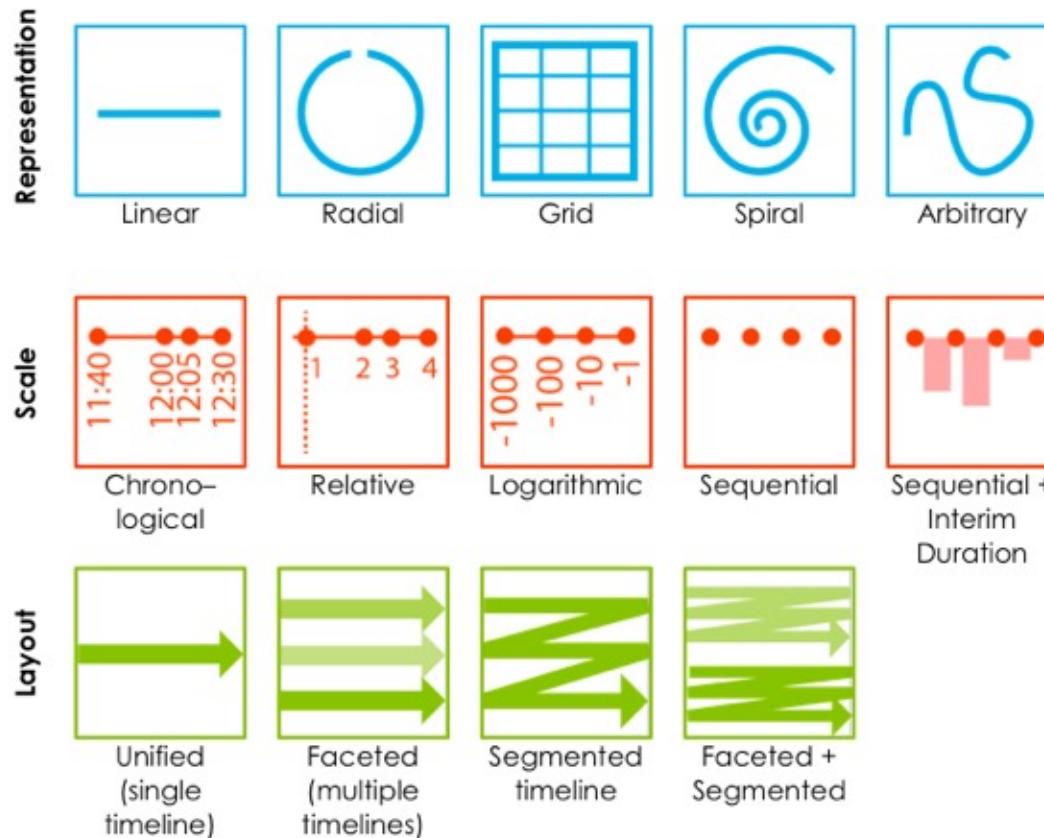
可视化

- 揭示数据中隐藏的特征模式
- 展示与时间相关的变化规律和趋势

时变数据

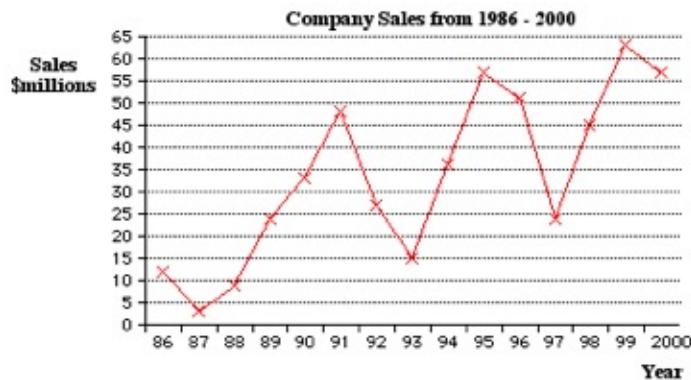
时变数据可视化的总体设计三大维度：

表达 (representation) 、比例 (scale) 、布局 (layout)

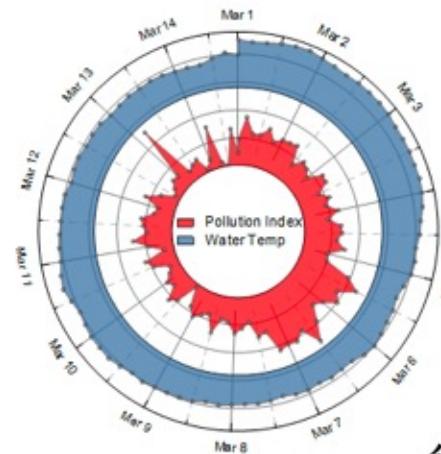


时变数据：表达

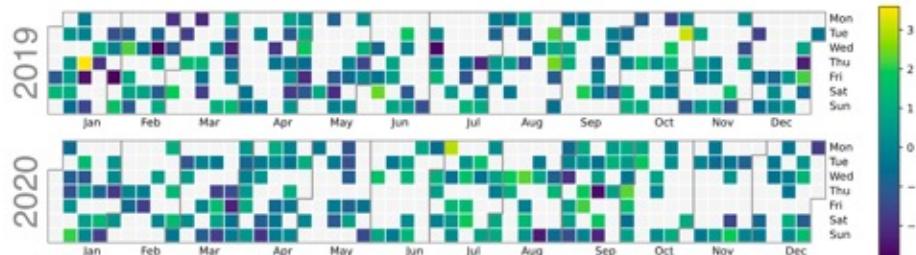
决定时间数据以什么样的形式展现在可视化作品中。



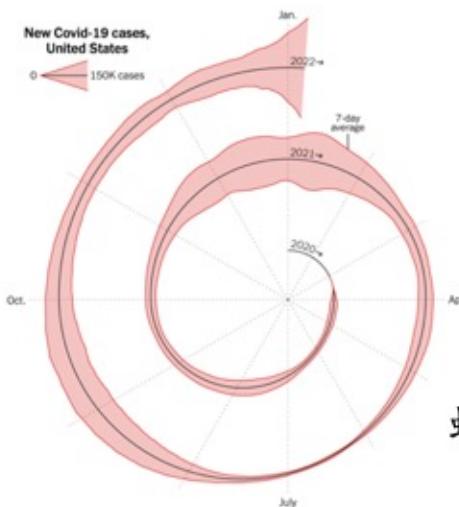
线性表达



径向表达



表格式表达

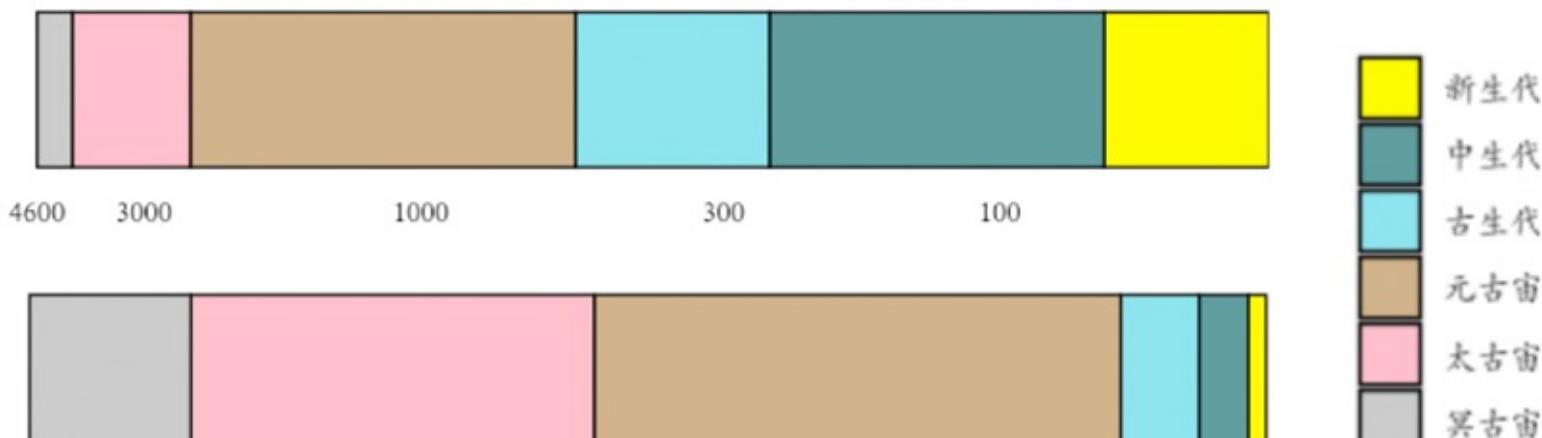


螺旋形表达

时变数据：比例

决定时序数据映射到可视化作品里的比例（线性、对数、绝对时间、相对时间等）

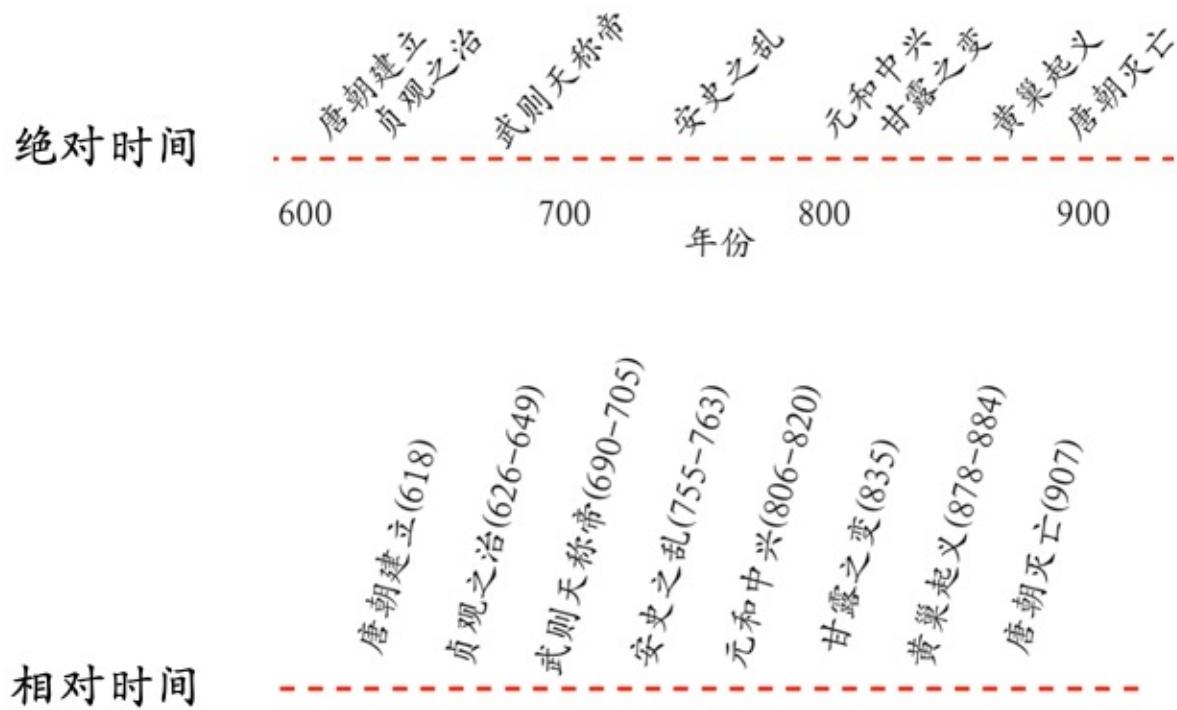
对数



线性

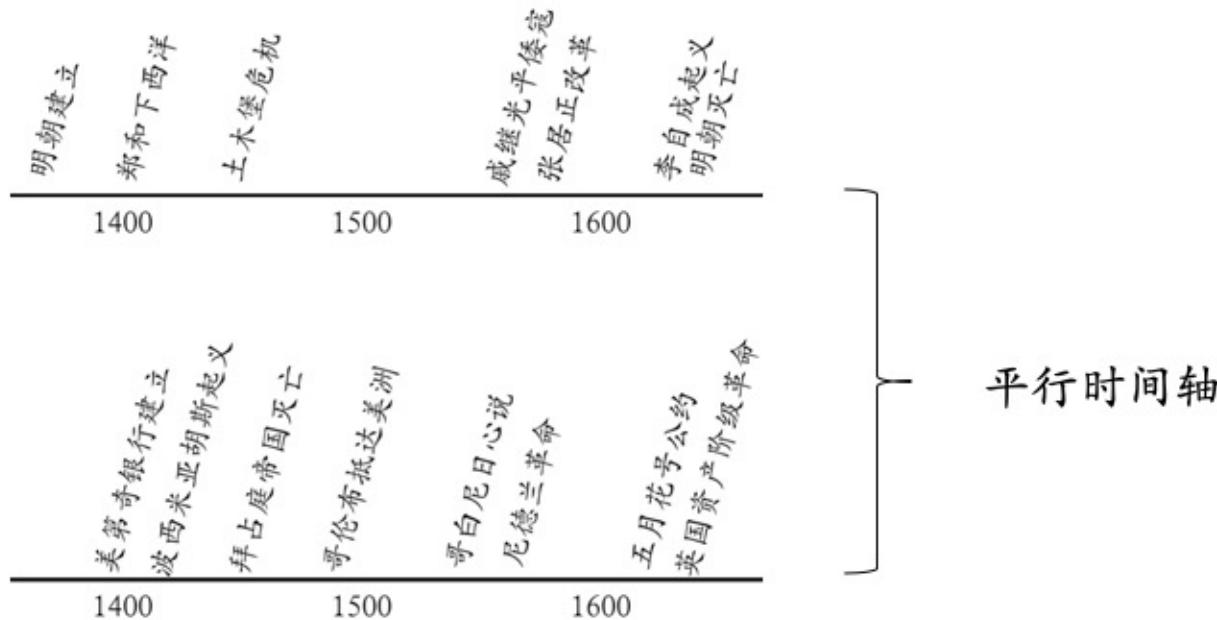


时变数据：比例



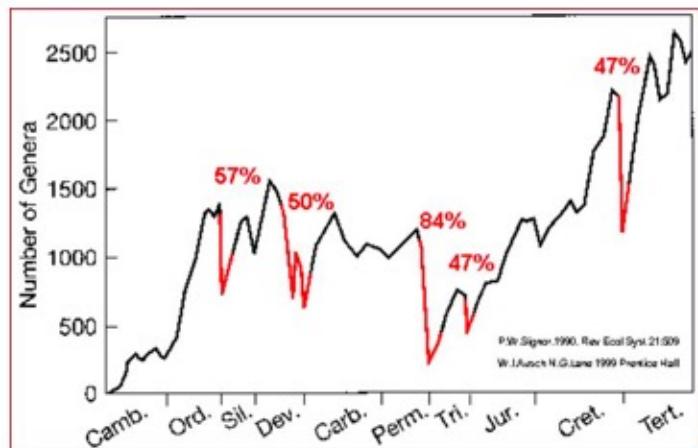
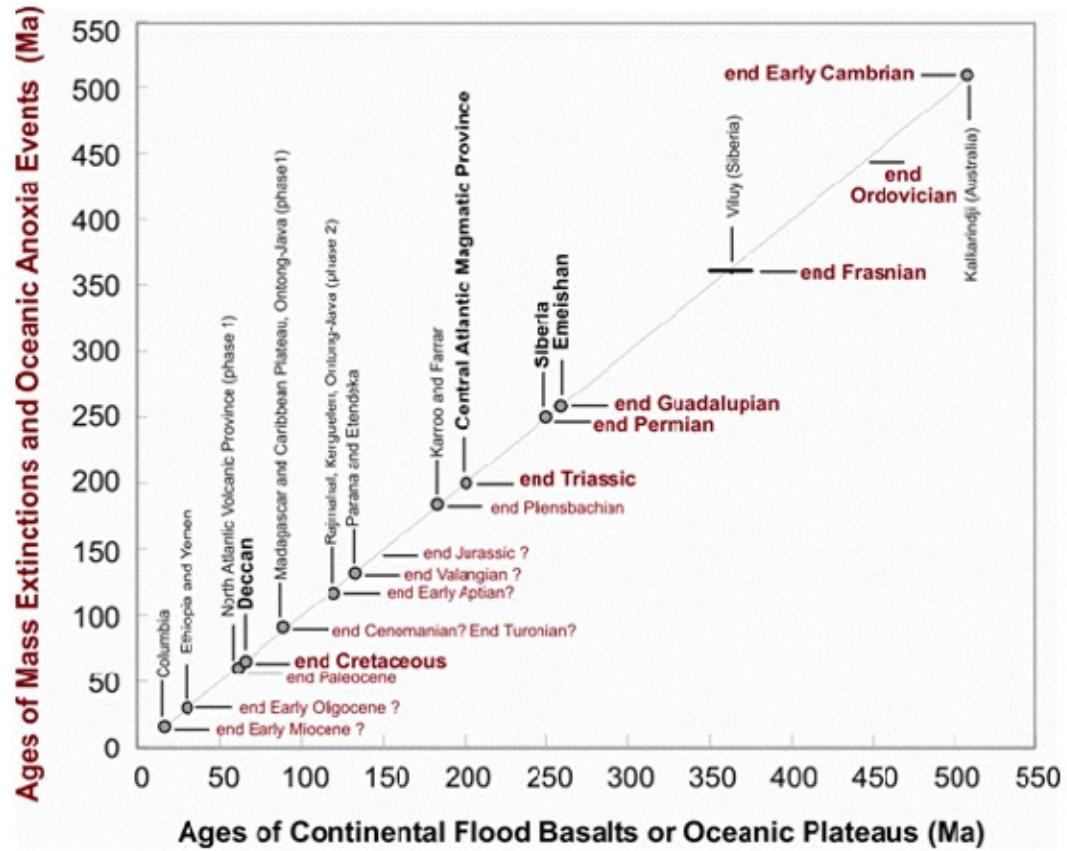
时变数据：布局

时序数据的排布方式（单条时间轴、平行时间轴、交叉时间轴、转折时间轴）



时变数据：布局

交叉时间轴

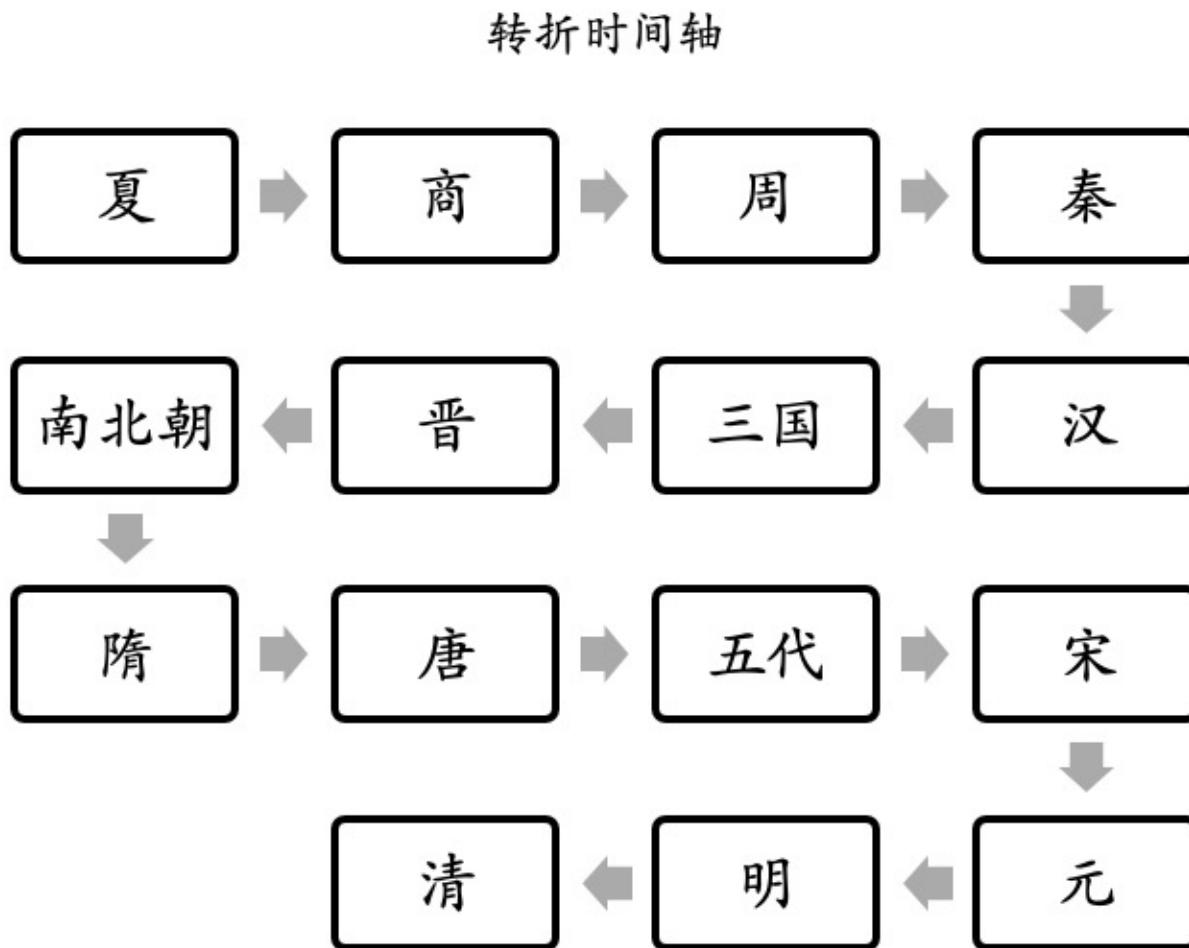


生物大灭绝



洪流玄武岩 flood basalt
大火成岩省 (LIPs)

时变数据：布局



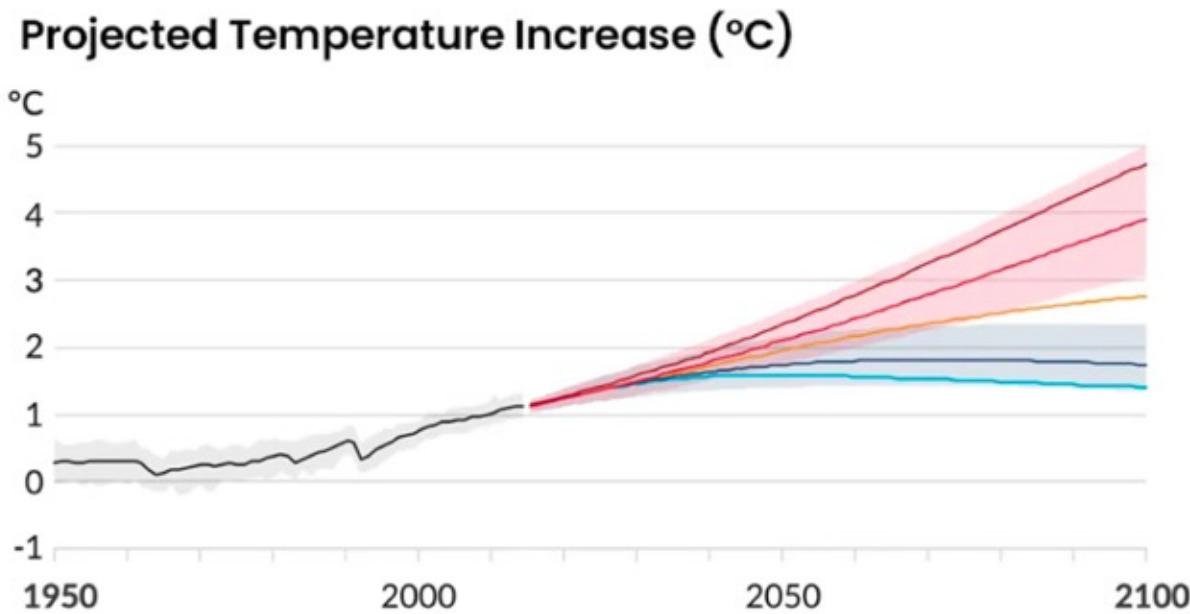
时间属性

对时间属性的刻画方式：

- 连续时间轴：线性时间域和循环时间域（周期时间域）
- 离散时间轴：时间基元和间隔
- 逻辑时间轴：相对时间、分支（分离并立、if 线）等

线性时间域

常用线性表达，定义原点和所有数据元在时间轴的取值范围。

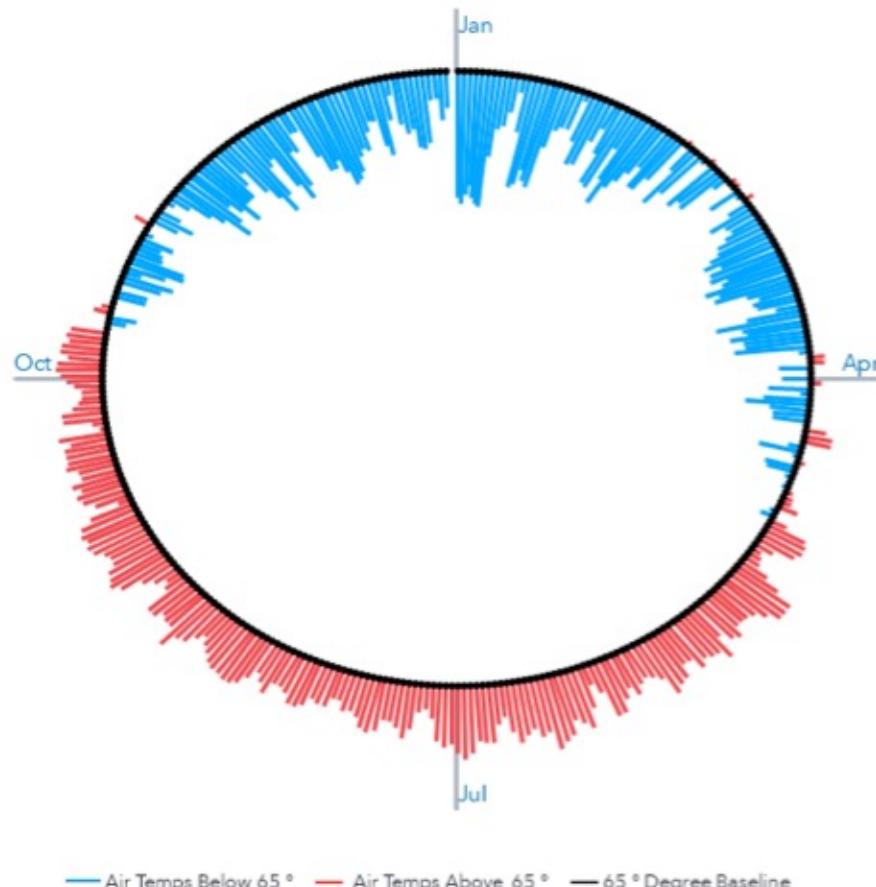


Five Scenarios of Fossil Fuel Burning

- Highest CO₂ amounts
- Medium to high CO₂ amounts
- Medium CO₂ amounts
- Smaller CO₂ amounts, then no increase in CO₂ late in the 21st century
- No increase in CO₂ beginning in 2050

循环时间域

在严格的循环时间域中，数据元间的相对顺序不存在严格的意义。常用径向或螺旋的表达方式。



时间基元和间隔

离散时间轴：展示时间基元，包括**时间点**（没有持续的概念）或局部连续的**时间段**（局部连续但总体离散）。时间基元之间为时间轴上的间隔。

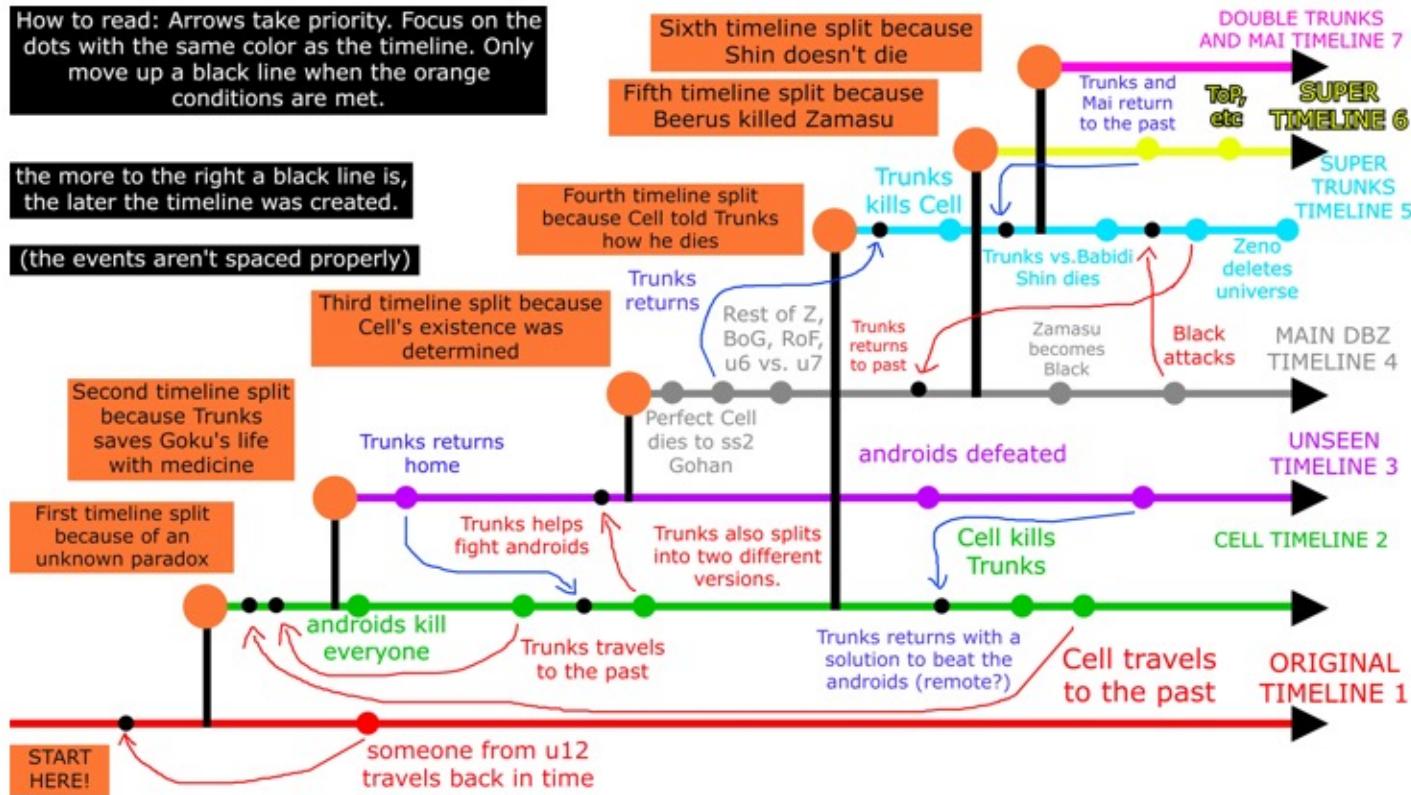
时间点

| 时间 | 事件 |
|------------|---------------|
| 581年 | 隋朝建立 |
| 589年 | 隋朝统一 |
| 618年 | 隋朝灭亡，唐朝建立 |
| 627—645年 | 玄奘西行 |
| 641年 | 文成公主嫁到吐蕃 |
| 754年 | 鉴真东渡到达日本 |
| 755—763年 | 安史之乱 |
| 907年 | 唐朝灭亡，五代十国时期开始 |
| 10世纪初 | 耶律阿保机建立契丹国 |
| 960年 | 北宋建立 |
| 11世纪中期 | 毕昇发明活字印刷术 |
| 1115年 | 完颜阿骨打建立金朝 |
| 1125年 | 金灭辽 |
| 1127年 | 金灭北宋，南宋建立 |
| 1206年 | 成吉思汗建立蒙古政权 |
| 1271年 | 忽必烈改国号为大元 |
| 1276年 | 元灭南宋 |
| 1368年 | 元朝灭亡，明朝建立 |
| 1405—1433年 | 郑和七次下西洋 |
| 1406—1420年 | 营建北京城 |
| 16世纪中期 | 戚继光抗倭 |

时间段

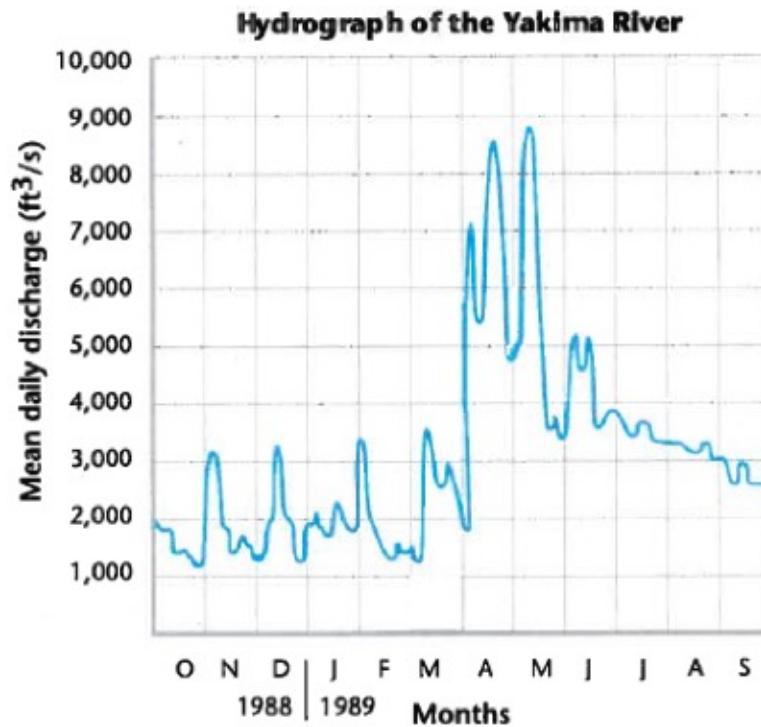
特殊时间轴

特殊时间轴：展示分支时间时间域，多角度时间域等特殊情况。



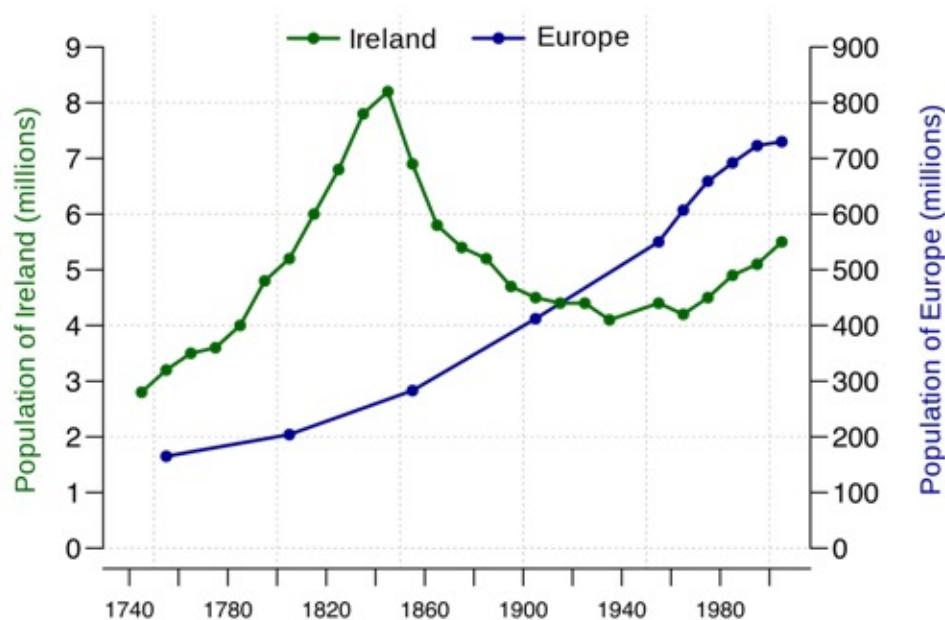
折线图

- 折线图 (line chart) 是时间序列图表里最常见的类别，通常被用于在连续时间域里显示另一个数值变量。
- 适于显示趋势和关系，展示事物的发展情况。
- 不建议使用三维折线图。



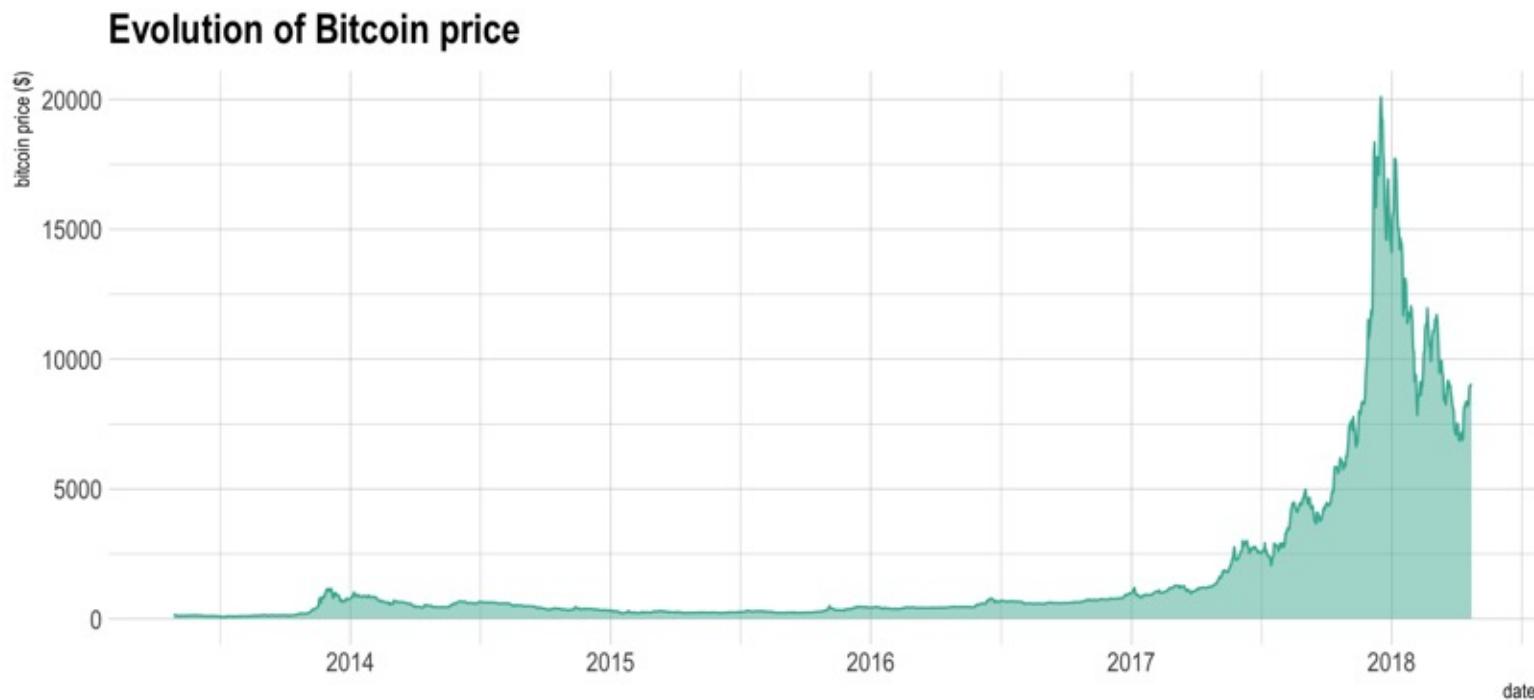
折线图

双Y轴：展示多个数值差异较大的时间序列。



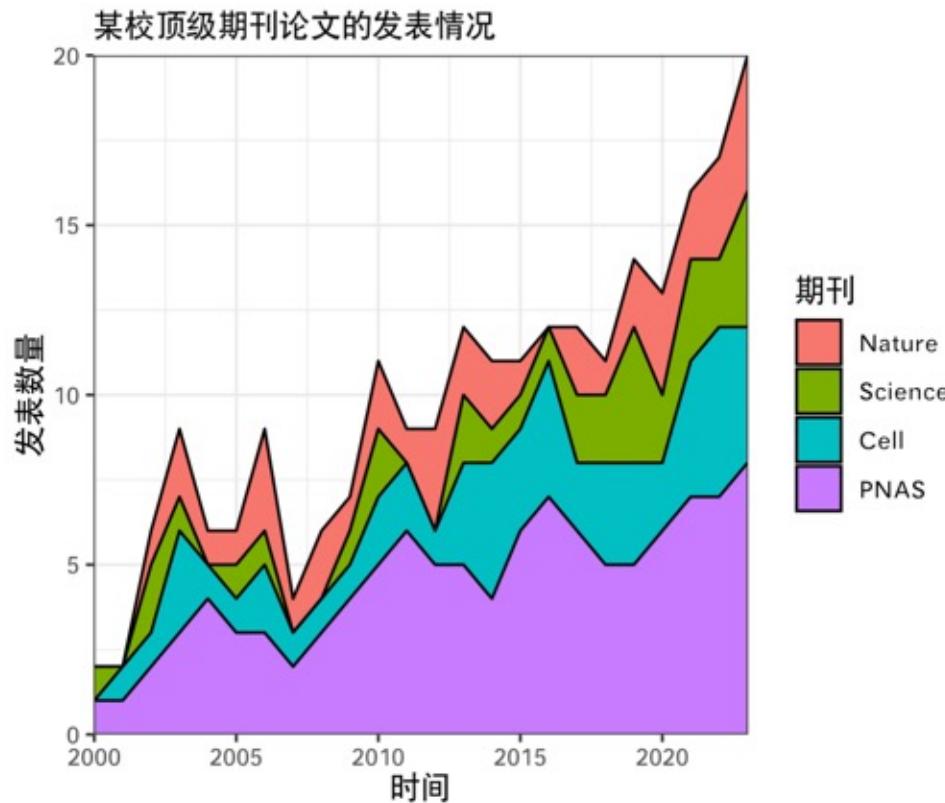
面积图

- 面积图 (area graph)：将折线与时间轴之间的区域用颜色或者纹理填充，可以更好地突出趋势信息，让图表在视觉上更加突出。



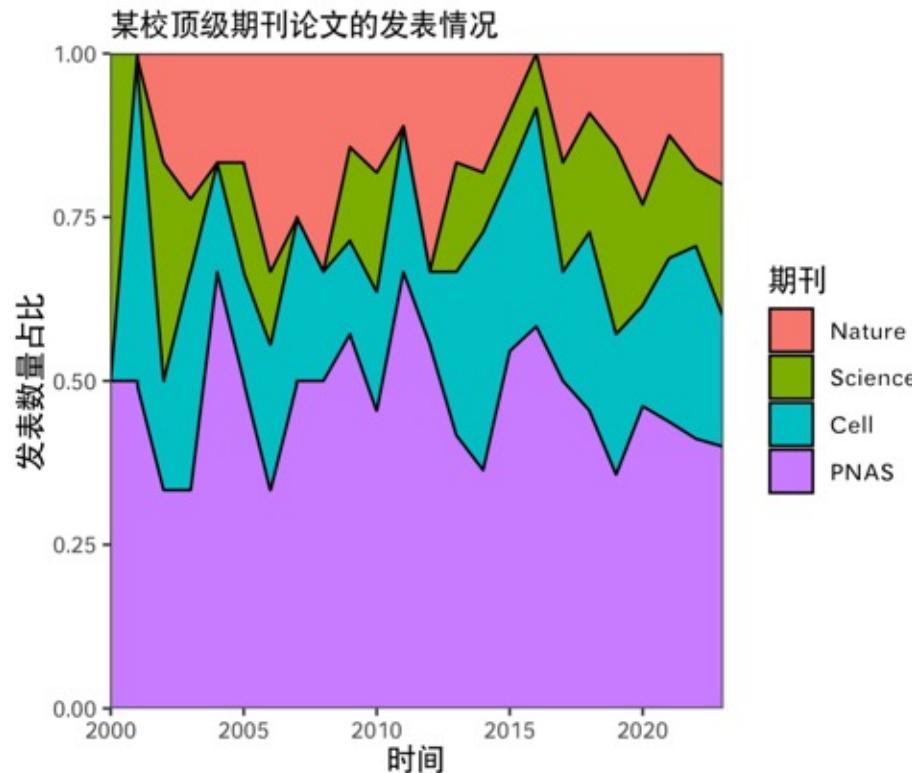
面积图

- 堆积面积图 (stacked area graph, 层叠面积图) : 同时显示多个时间序列, Y轴方向为累积关系, 最上层折线是所有数值变量在对应时间点的总和。适合展现不同时间序列的总量和分量的变化情况。



面积图

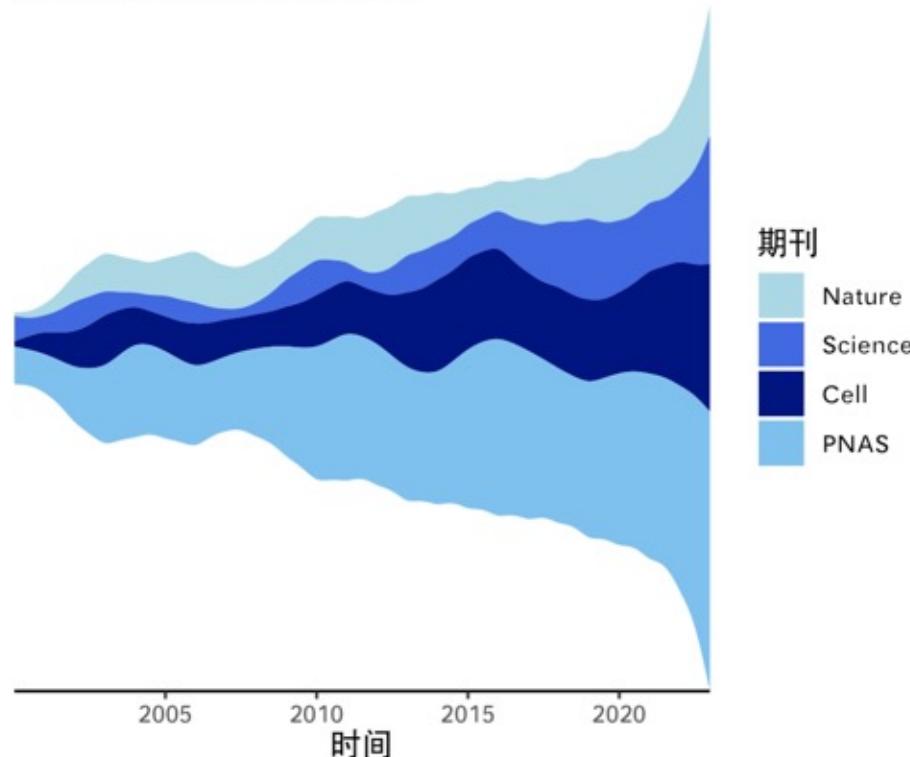
百分比堆积面积图 (fractional stacked area graph) : 进行归一化处理后的堆积面积图。



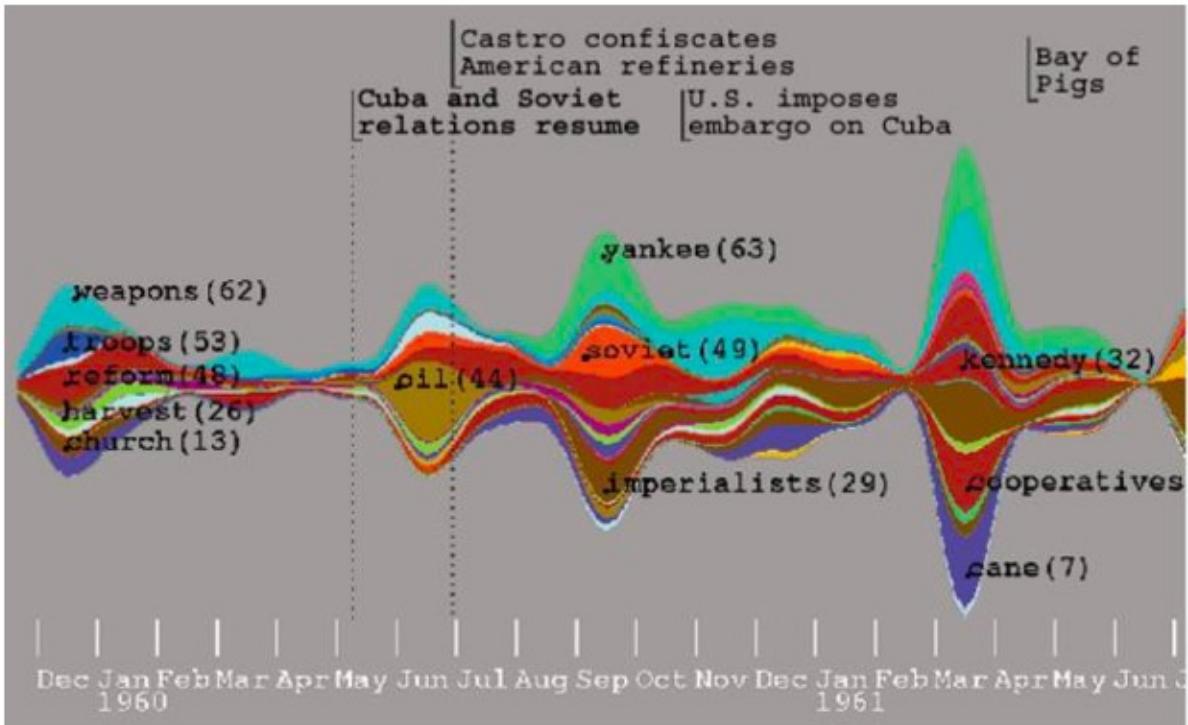
河流图

- 河流图（stream chart）：堆积面积图的一种变体，通过流动的形状来展示不同类别/序列随时间的变化情况。
- 将数据分散到中心基准线上，形成类似河流的形状，视觉冲击力较强。

某校顶级期刊论文的发表情况



河流图



Theme River: In Search of Trends, Patterns, and Relationships

交互式图表
可以克服

使用场景

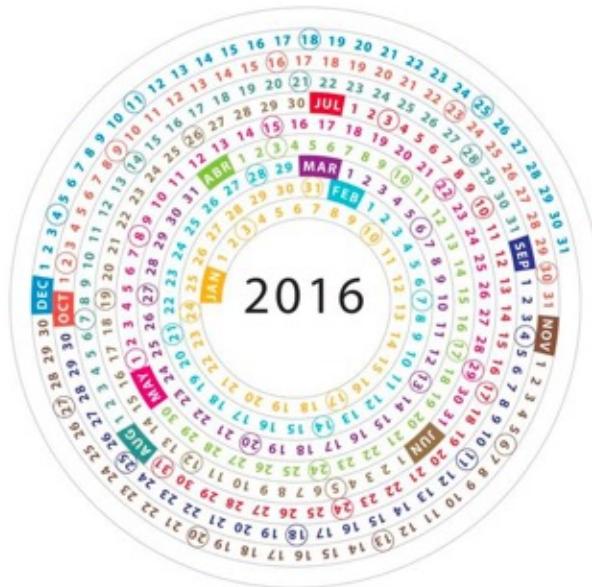
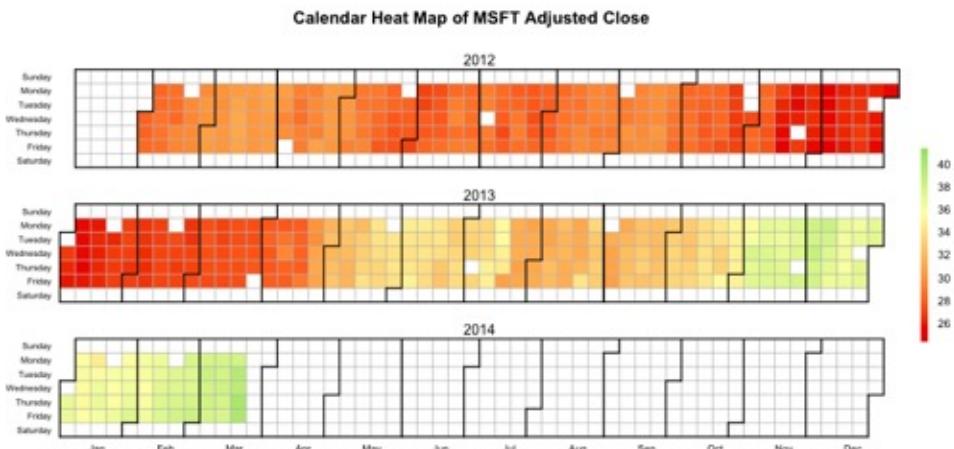
- 季节性峰值和谷值代表周期性模式
- 可显示大量资产在一段时间内的波动率

缺点

- 不易读
- 显示大型数据集时易显得混乱
- 数值小的类别可能会被“淹没”
- 无法读取每精确数值，只能看比例

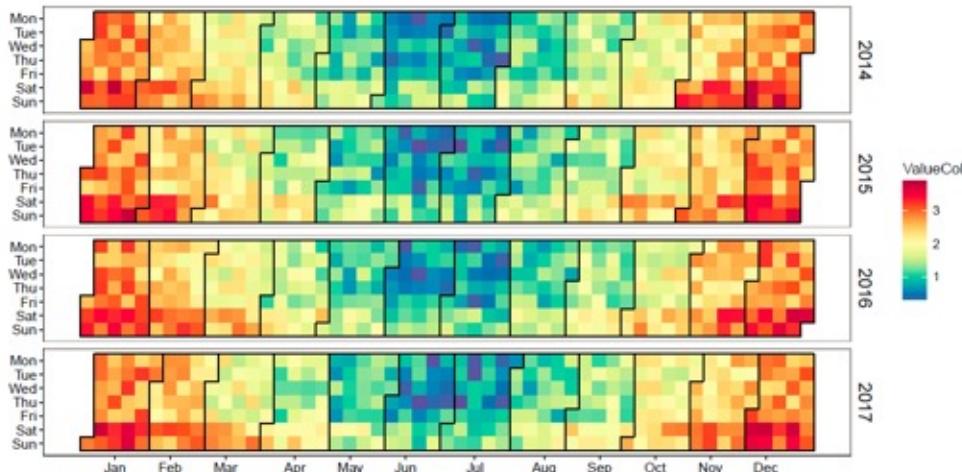
日历图

日历图 (calendar chart) : 将时间属性和日历对应, 符合常人理解时间的习惯。一般采用表格映射的方式。

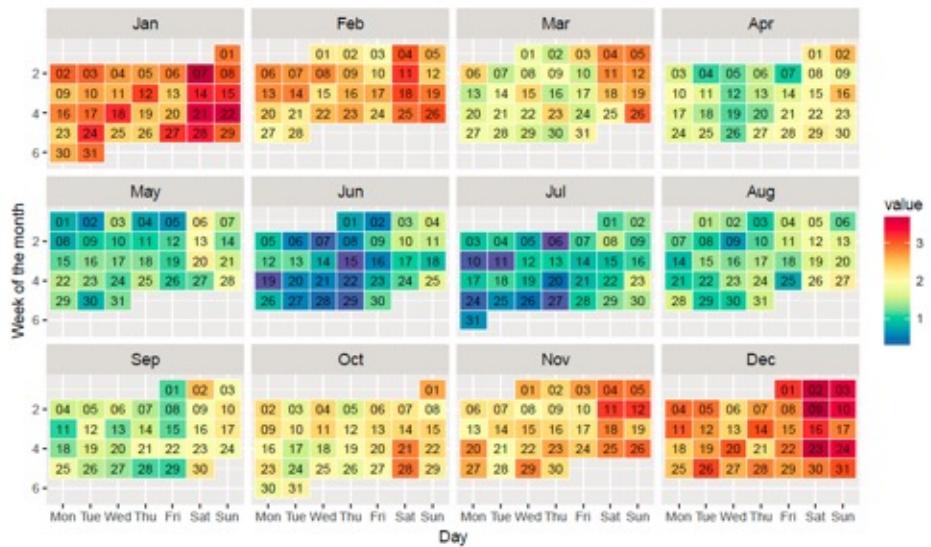


日历图

以年为单位



以月为单位



日历图



Shan Ye

yeshancqcq

A data-driven geoscientist at CUGB

Edit profile

At 29 followers · 11 following

University of Wisconsin-Madison

美国

yeshan@umich.edu

yeshan-geo.github.io

Achievements



Pinned

EarthCubeGeochron/IRD Dive Public

Mining documents with IRD data over the last 5 millions years.

HTML ⭐ 3 ⚡ 2

next-server Public

Forked from GIS-Info/next-server

"GIS海外院校和项目信息"下一代网站的后端部分

Python

EarthCubeGeochron/Sparrow Public

A software tool and schema+API spec for connecting laboratory measurements to data consumers

Python ⭐ 14 ⚡ 3

Customize your pins

GIS-Info/next-web Public

"GIS海外院校和项目信息"下一代网站的web前端部分

Vue ⭐ 8 ⚡ 17

GIS-Info/GIS-Info.github.io Public

GISphere Guide | GIS海外留学信息

SCSS ⭐ 151 ⚡ 38

TerraMap Public

A tool under development

C++

326 contributions in the last year



Contribution settings ▾

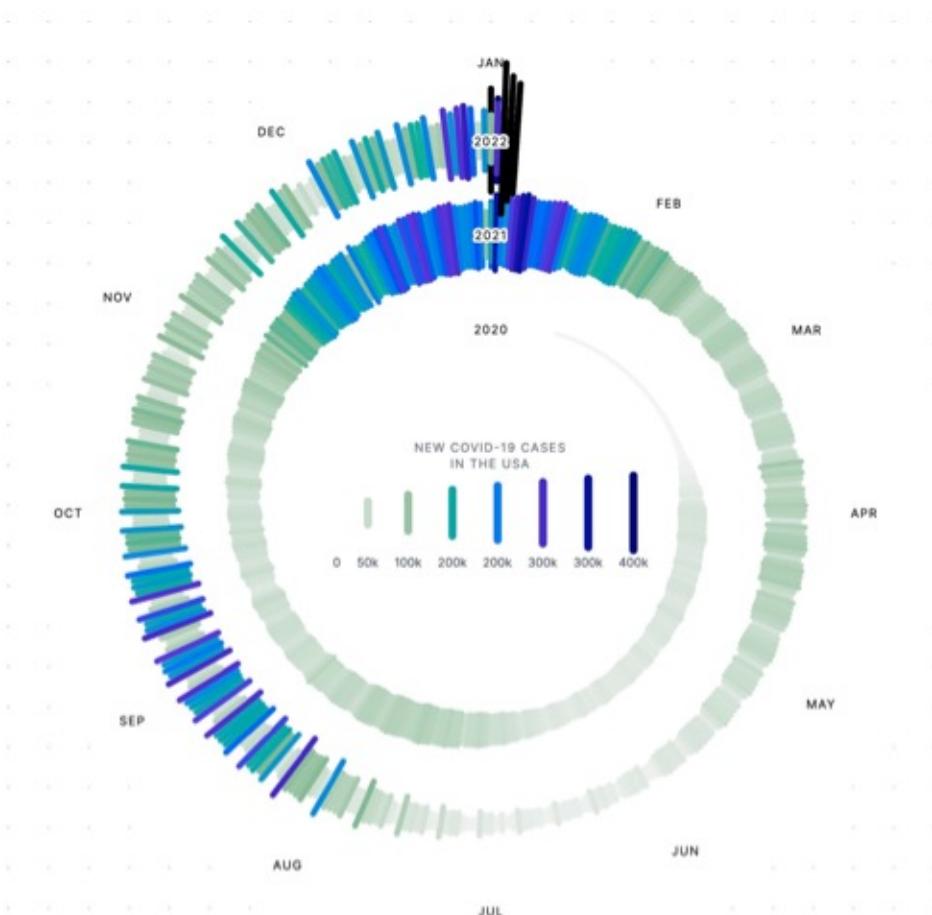
2023



GitHub 日历图

螺旋图

螺旋图 (spiral chart) : 沿阿基米德螺旋线 (Archimedes spiral) 描绘时间序列数据。图表从螺旋形的中心点开始向外发展。



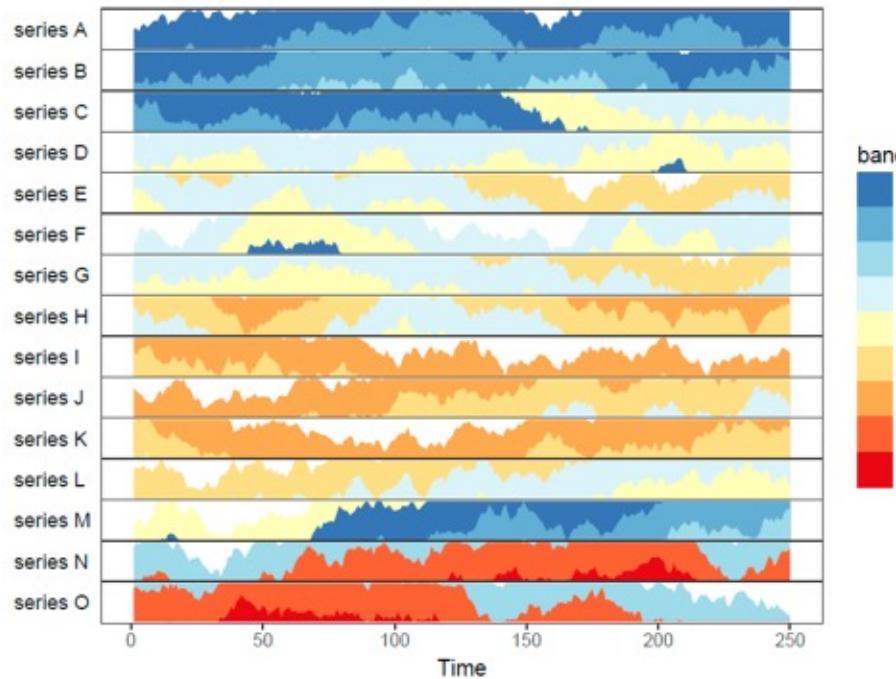
适用场景

- 节省空间，可用于显示时间跨度大的数据
- 每一圈的刻度差相同，适合展示周期性数据

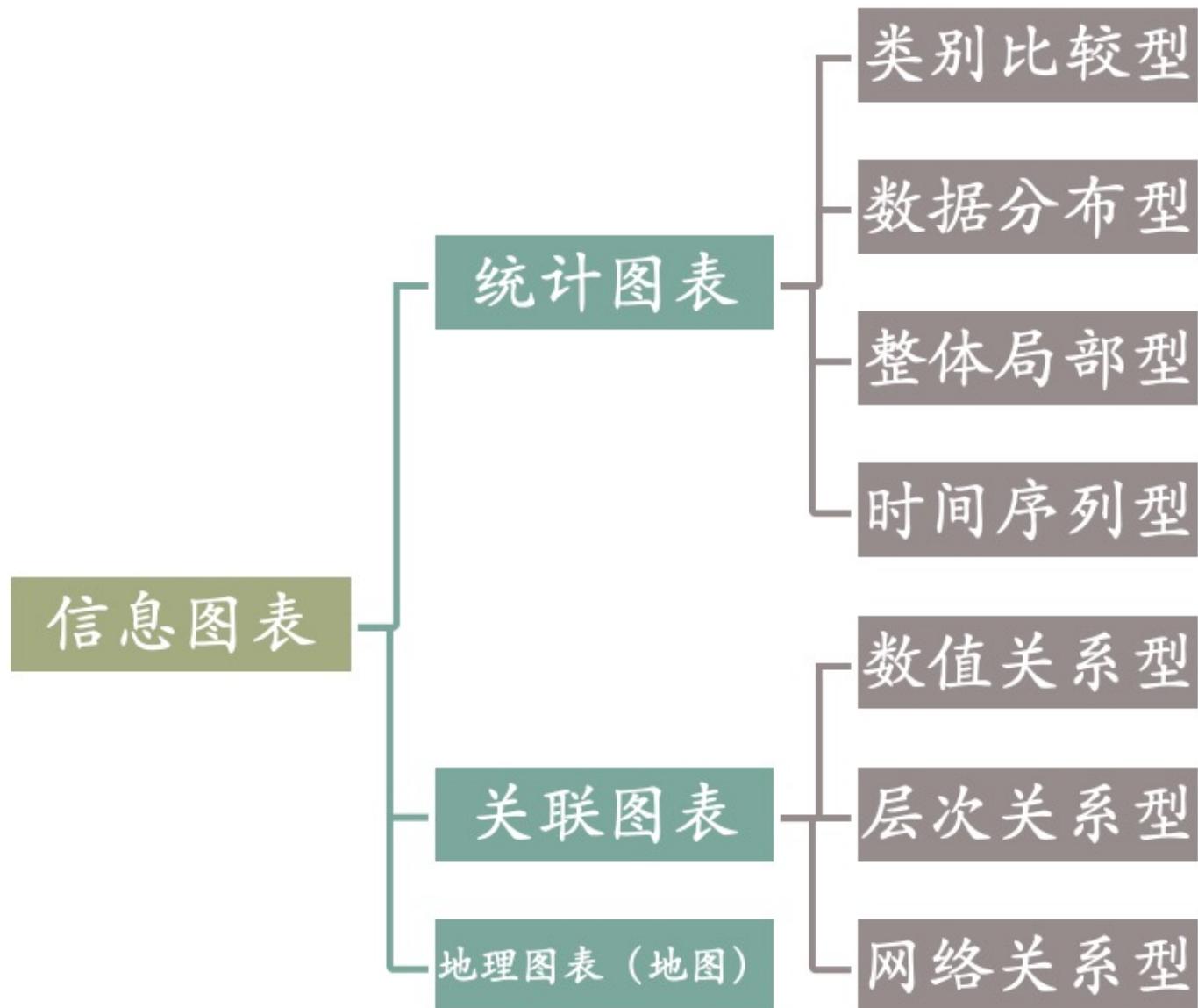
地平线图

地平线图 (horizon graph) : 将不同的时变数据系列纵向排列后进行展示。最初被金融投资经理用于展示同一时间的股票数据。现在的使用场景包括：

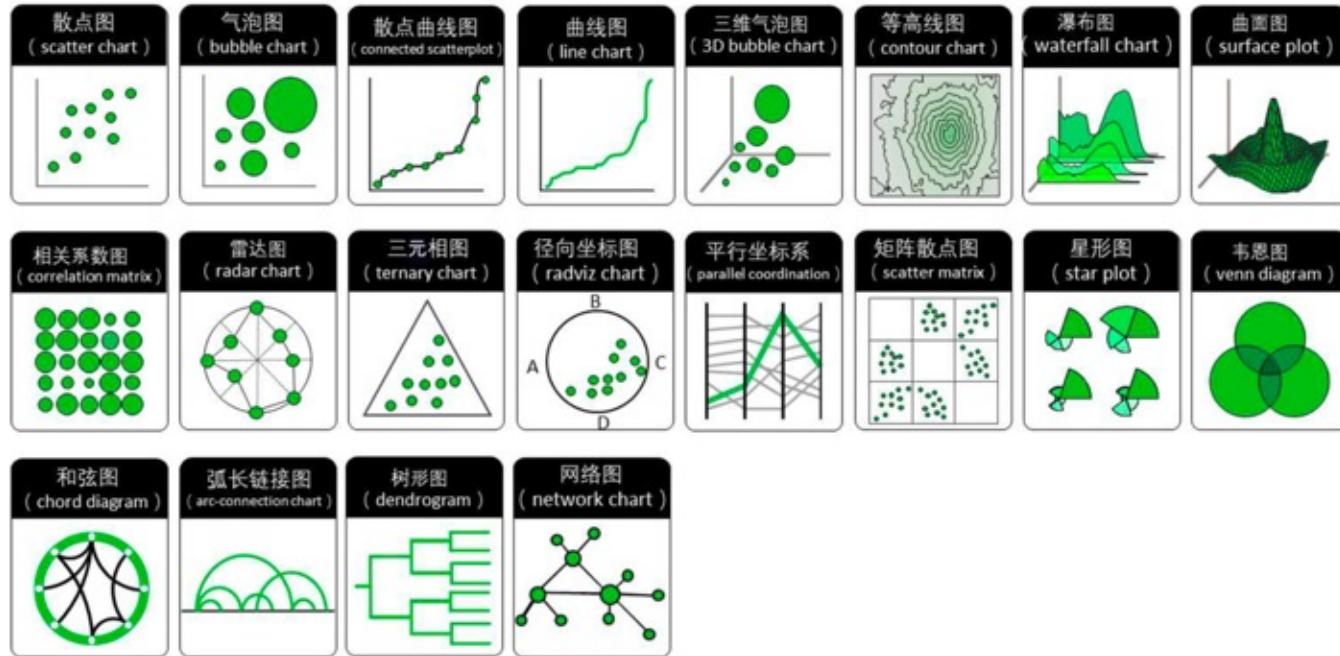
- (1) 辨识异常数据、异常变化和主要的数据规律；
- (2) 在合理的精度范围内观察每个数据系列随时间的变化；
- (3) 比较不同数据系列的数值情况。



信息图表



关联图表



数值关系型图表

- 变量一般都为数值型
- 变量为1~3个：散点图、气泡图、曲面图等
- 变量多于3个：平行坐标系、矩阵散点图、径向坐标图、星形图、切尔若夫脸谱图等

层次关系型数据

- 数据个体之间的层次关系
- 包含关系、从属关系两种
- 节点链接图、树形图、冰柱图、旭日图、圆填充图、矩形树状图

网络关系型图表

- 展示不具备层次结构的关系数据
- 关系更加复杂、自由、灵活
- 桑基图、弦图、节点链接图、弧长链接图、蜂箱图

散点图和气泡图

经典气泡图可视化作品：两百年来的世界人口变化

