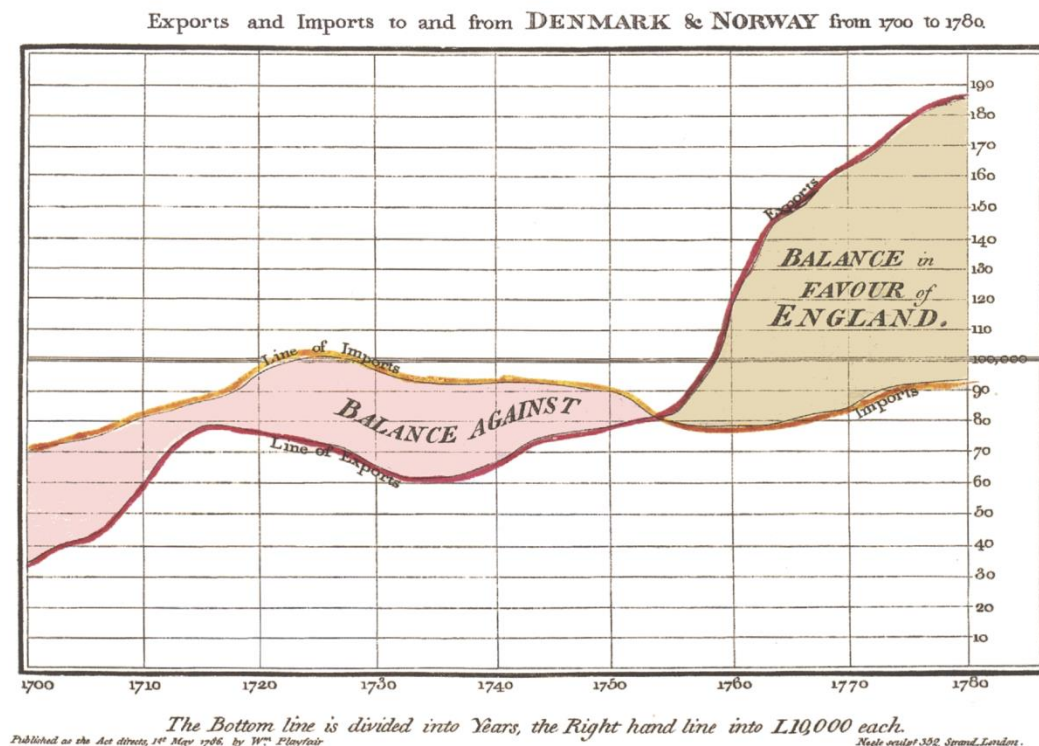


统计图形与数据可视化

Yu Lijia

悠久而古老的统计图形

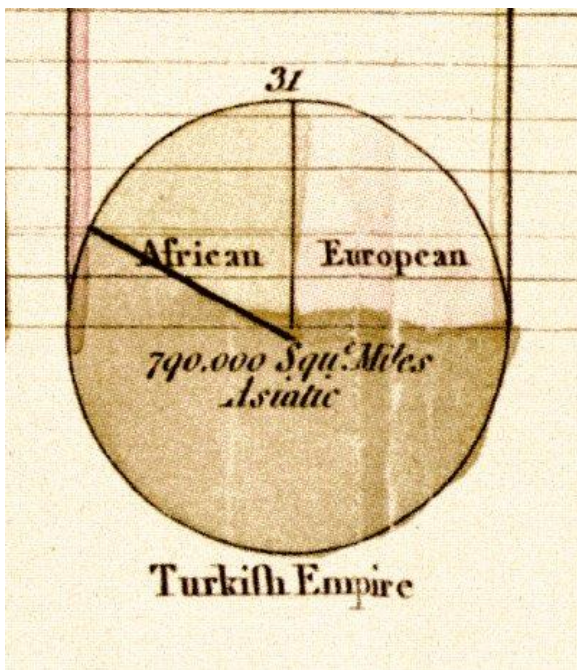
▶ 饼图和线图的起源



这幅图主要展示了1700年至1780年间英格兰的进出口时序数据，左边表明了对外贸易对英格兰不利，而随着时间发展，大约1752年后，对外贸易逐渐变得有利。

悠久而古老的统计图形

▶ 饼图和线图的起源



这幅图展示了土耳其帝国在三大洲的国土面积分布。

William Playfair

- ▶ 饼图和线图都是当今社会中常用的统计图形，它们是由有着“统计图形奠基人”之称的苏格兰工程师兼政治经济学家William Playfair发明的。



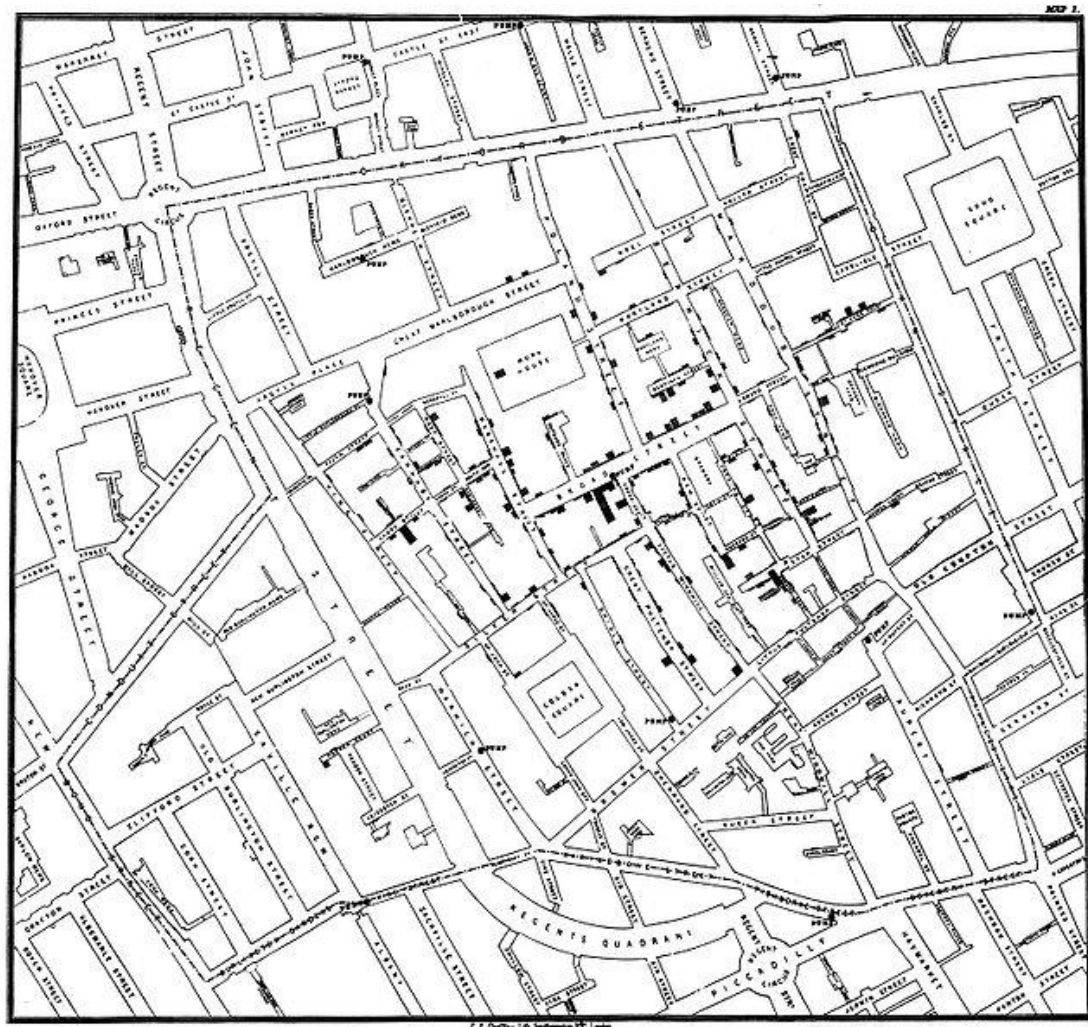
悠久而古老的统计图形

▶ 霍乱传染之谜

医生兼自学成才的科学家John Snow通过彻底调查这种致命疾病的根源。

这幅图形揭示了一个重要现象，就是死亡发生地都在街道中部一处水源（水井）周围，市内其它水源周围极少发现死者。进一步

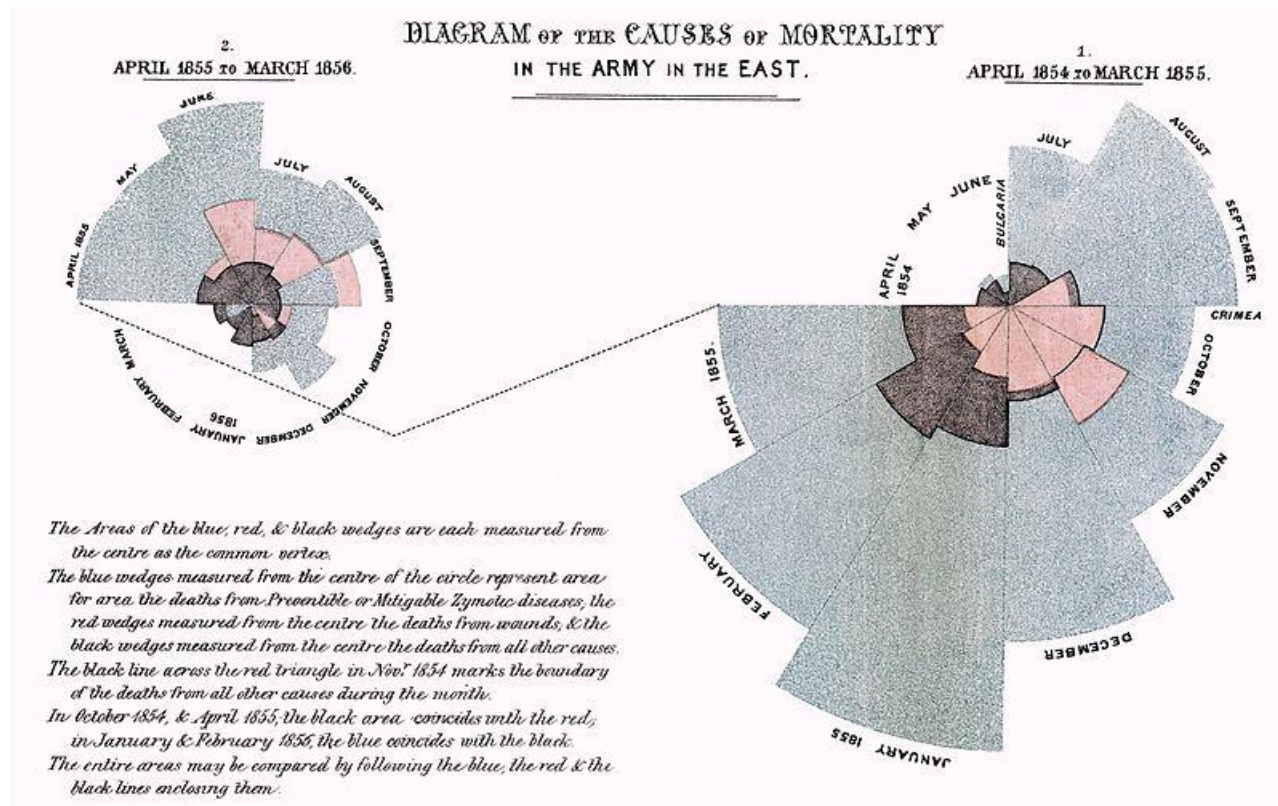
调查他发现这些死者都饮用过这里的井水。



悠久而古老的统计图形

▶ 提灯女士的玫瑰图

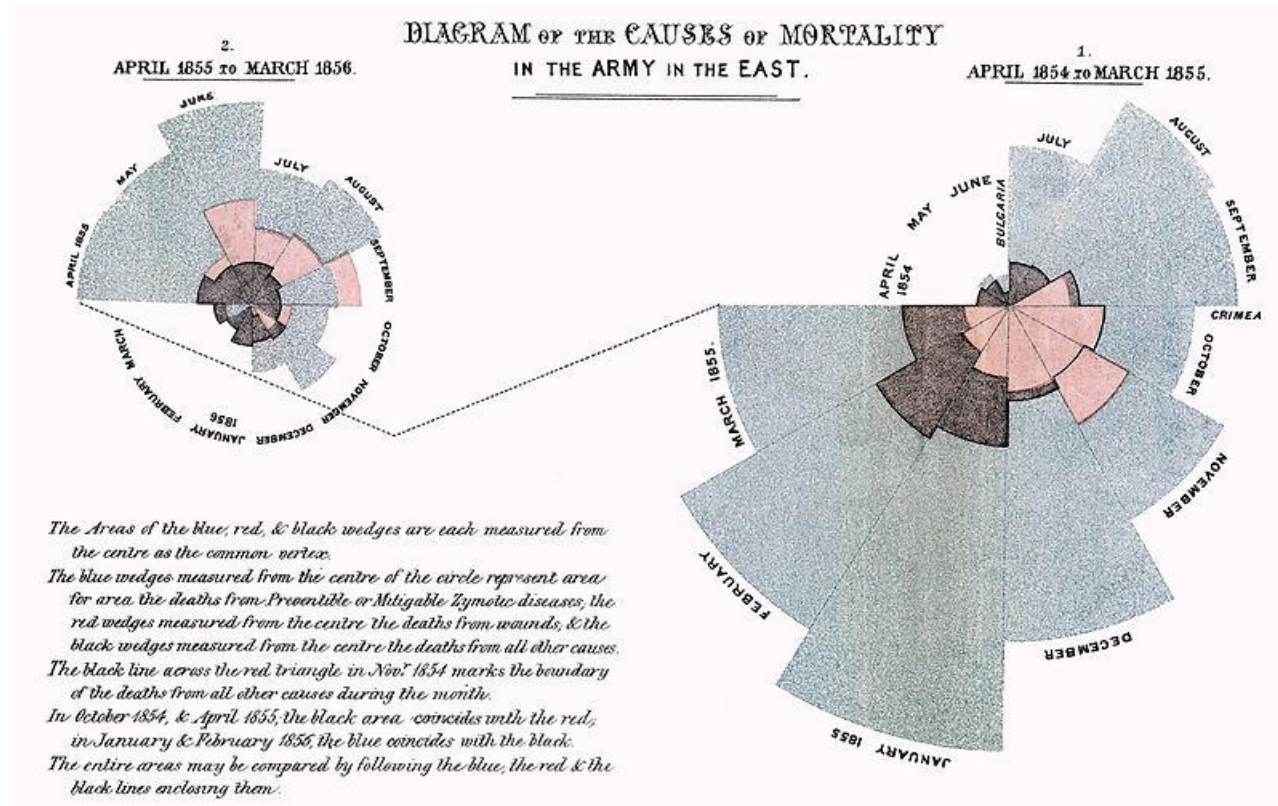
的极坐标面积图：两幅图分别是1854年和1855年的军队伤亡人数，一年12个月恰好可以将极坐标分为12等分，每一瓣代表一个月。图中用颜色标记出了三种死亡原因。



悠久而古老的统计图形

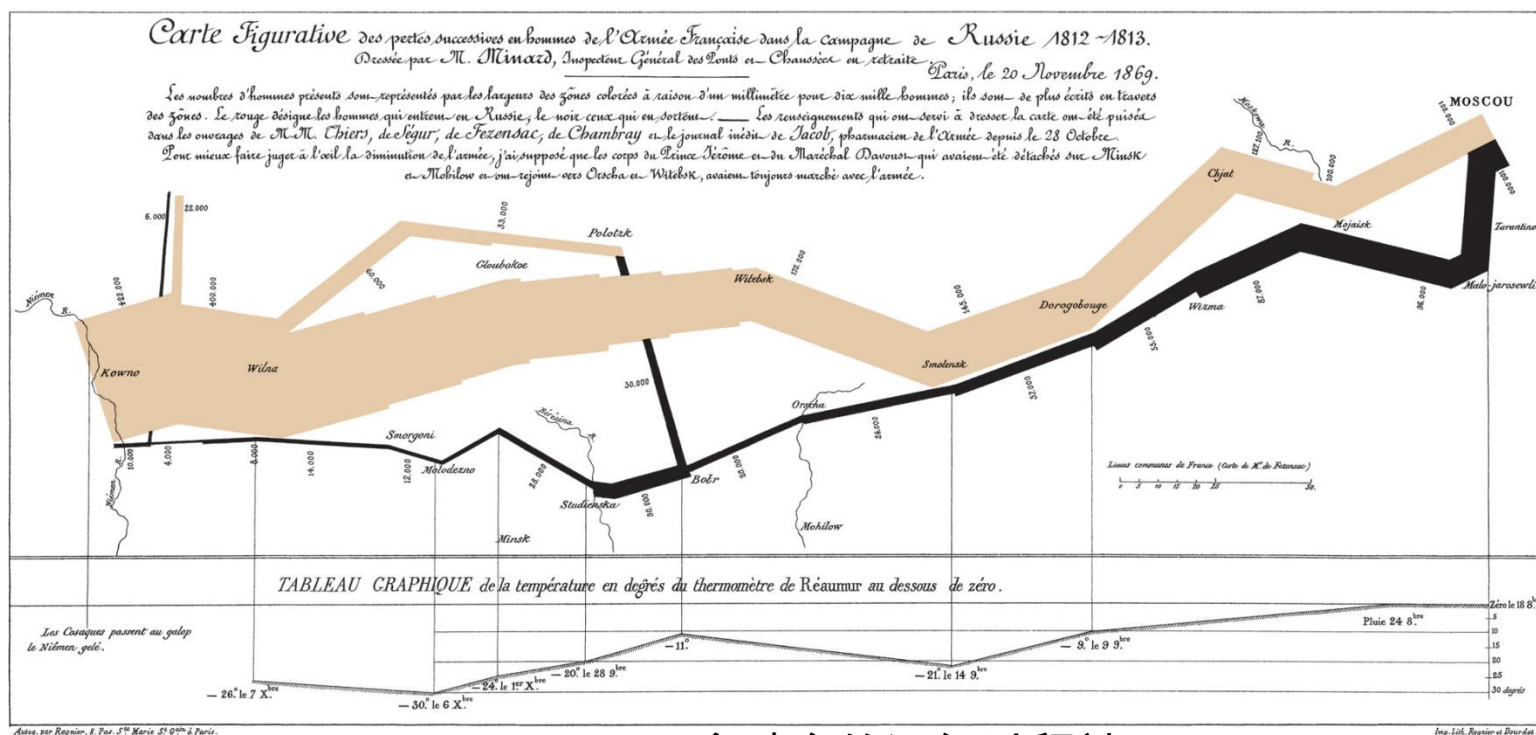
▶ 提灯女士的玫瑰图

南丁格尔的极坐标面积图：两幅图分别是1854年和1855年的军队伤亡人数，一年12个月恰好可以将极坐标分为12等分，每一瓣代表一个月。图中用颜色标记出了三种死亡原因。



悠久而古老的统计图形

▶ 拿破仑的俄罗斯远征

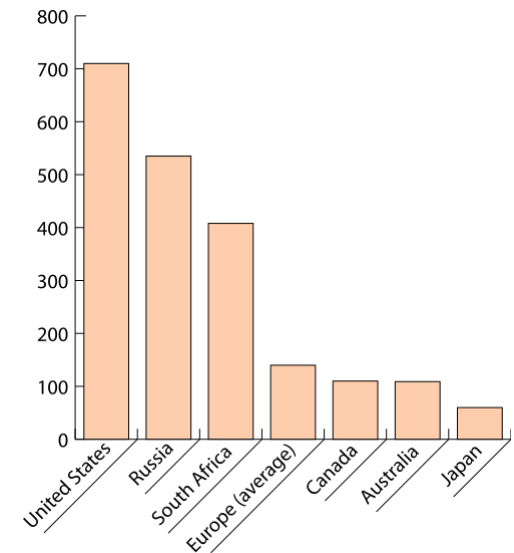
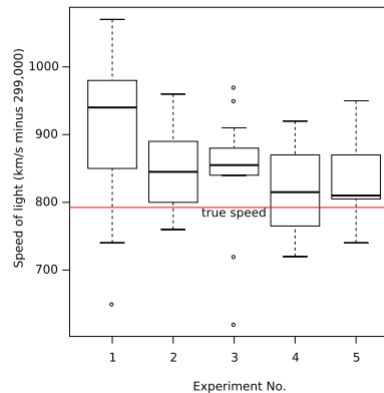
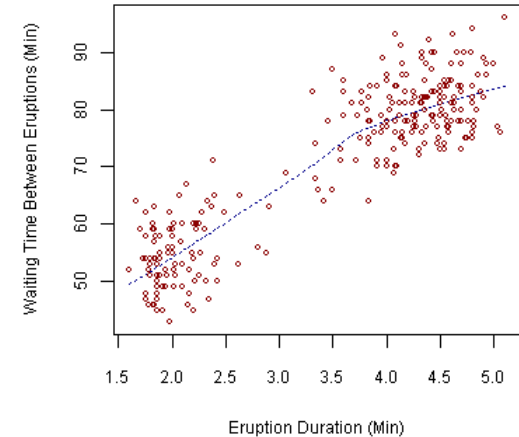


拿破仑的远征过程被Charles Joseph Minard用一张二维平面图形记录了下来，Minard是一位法国工程师，他以在工程和统计中应用图形而闻名。

What is Statistical graphics?

- ▶ Scatter plot(散点图)
- ▶ Bar chart(条形图)
- ▶ Pie chart(饼图)
- ▶ Box-plot(箱线图)
- ▶ Biplot(双标图)

Old Faithful Eruptions

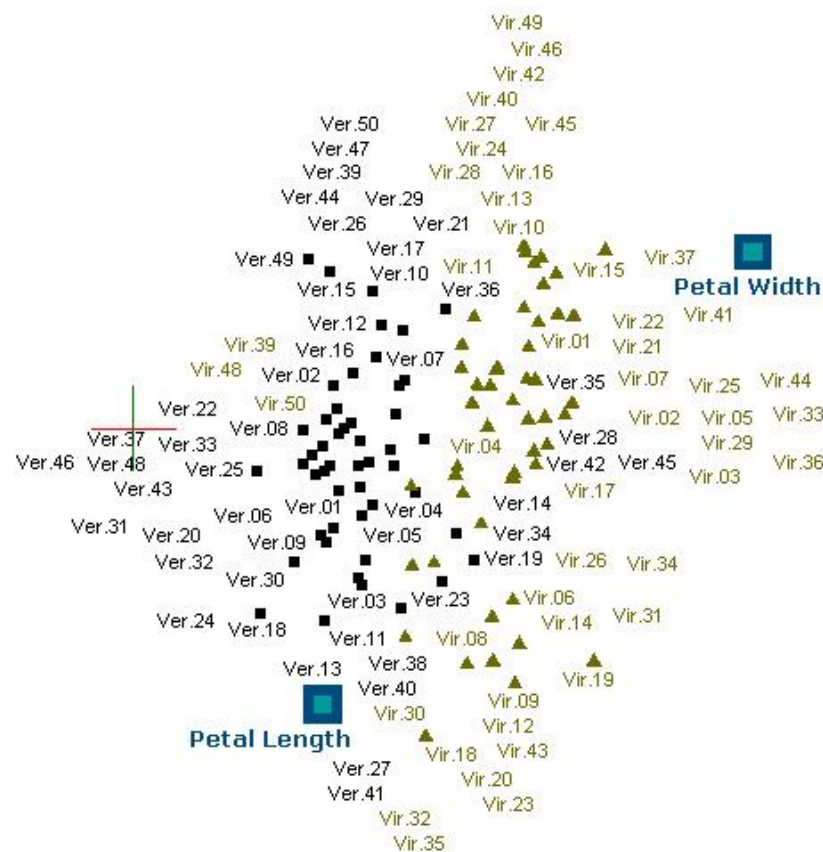
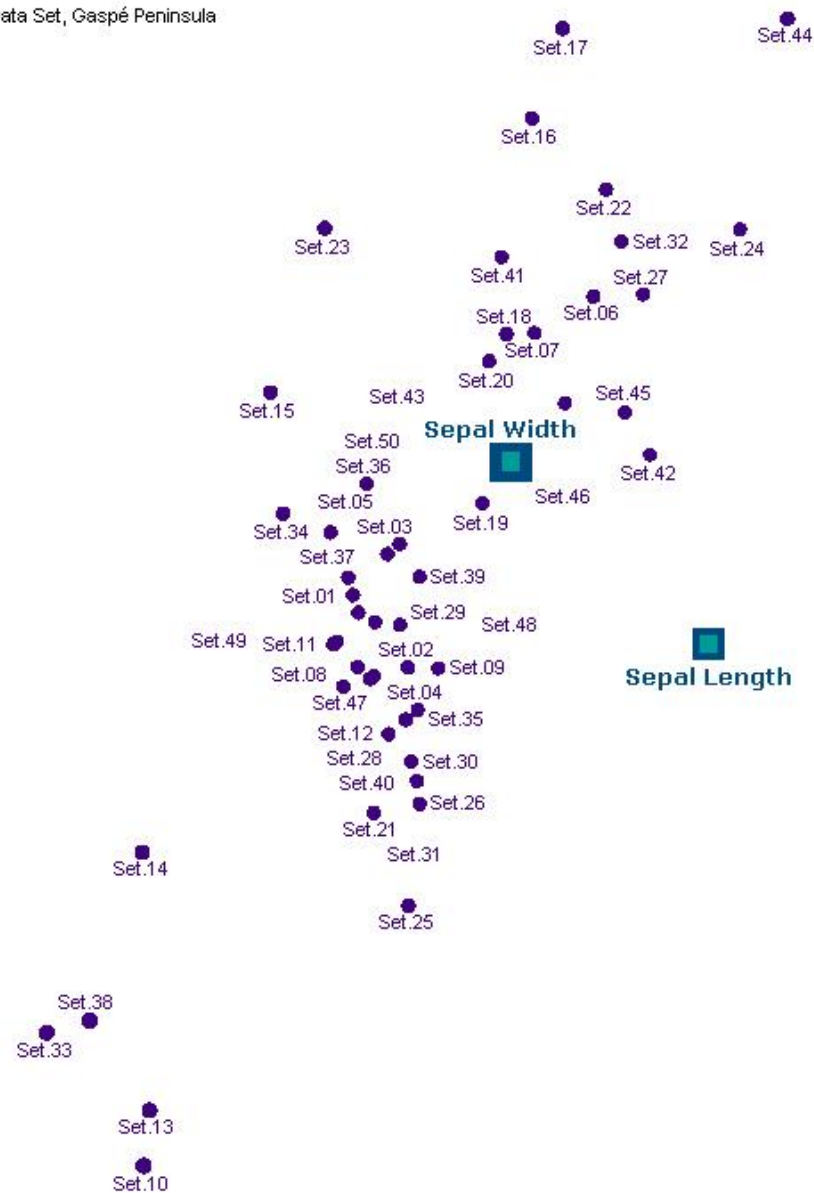


鸢尾花卉与双标图

Iris Flower Data Set, Gaspé Peninsula



Spectramap Biplot, $R^2=100\%$



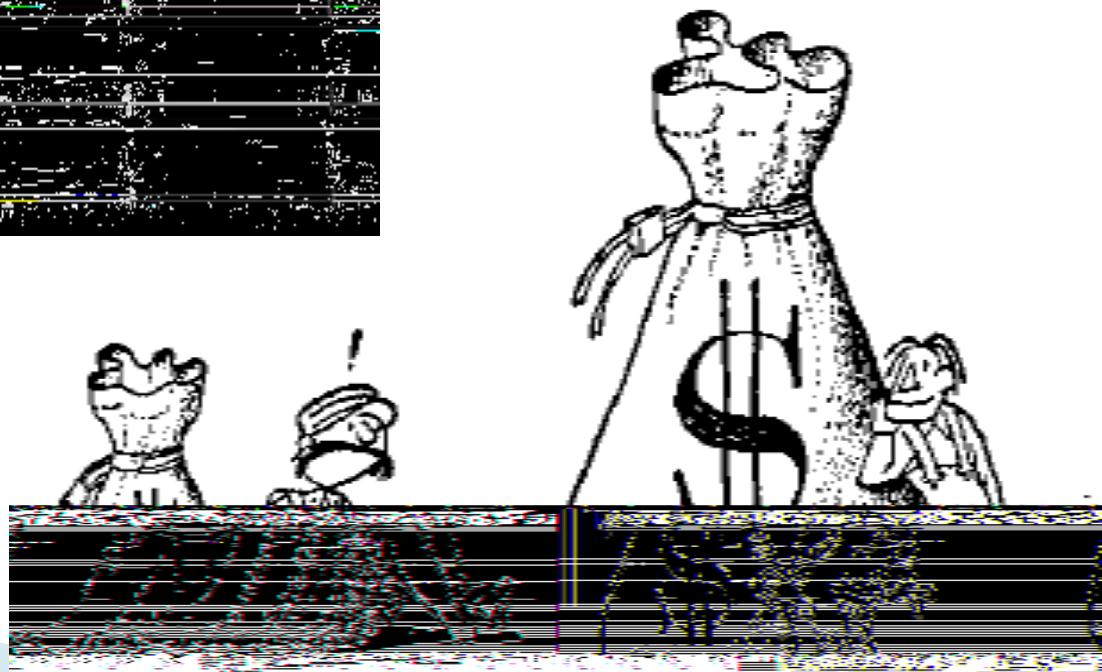
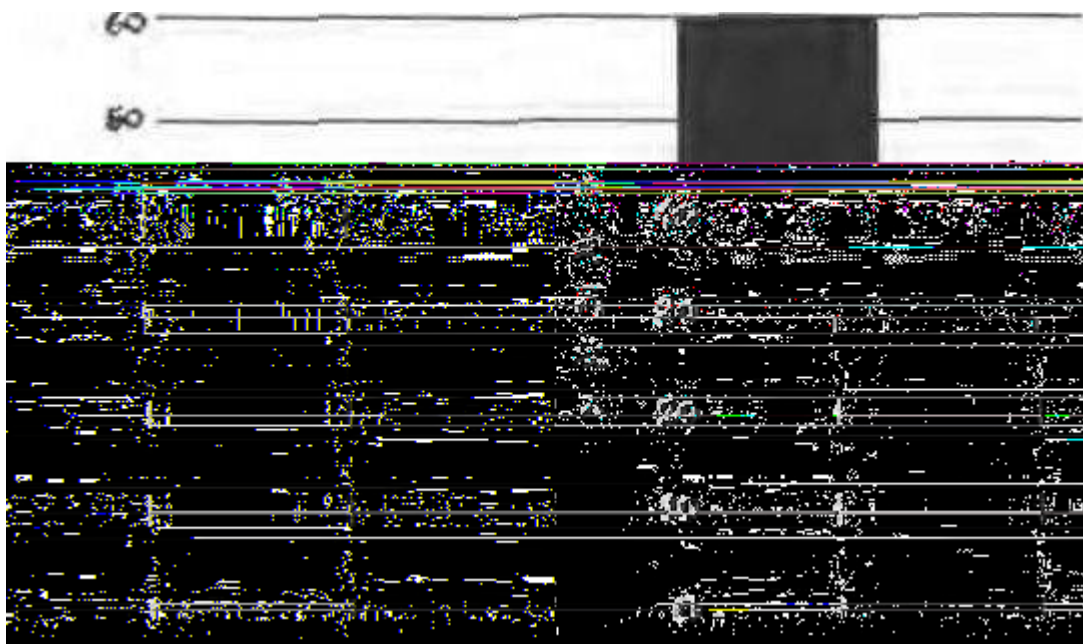
What is Data visualization?



Beautiful Visualization

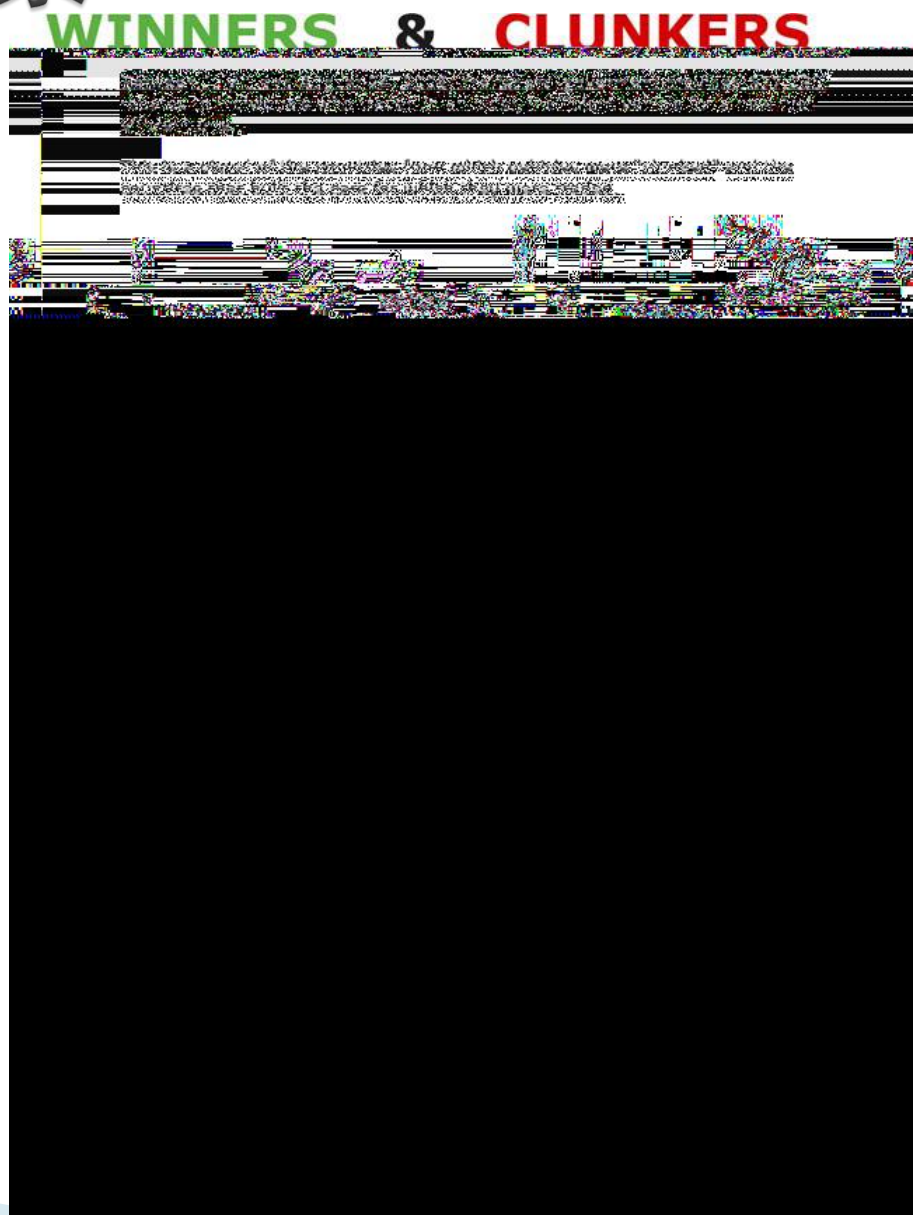
Looking at Data Through the Eyes of Experts

一个失败的例子

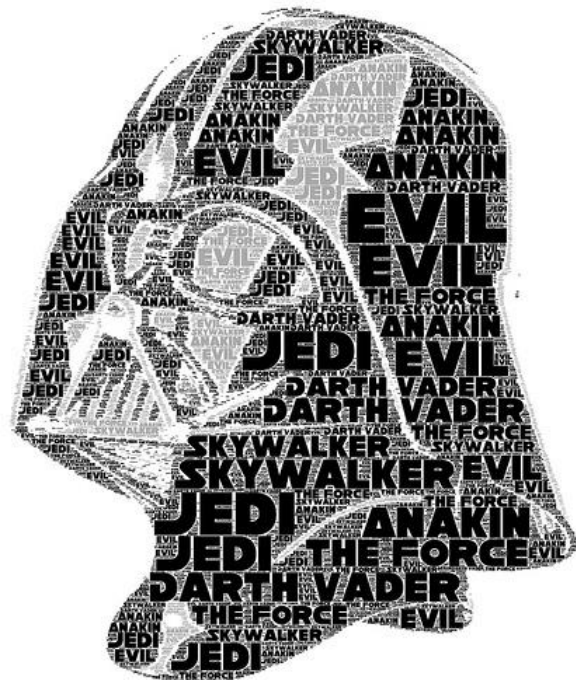
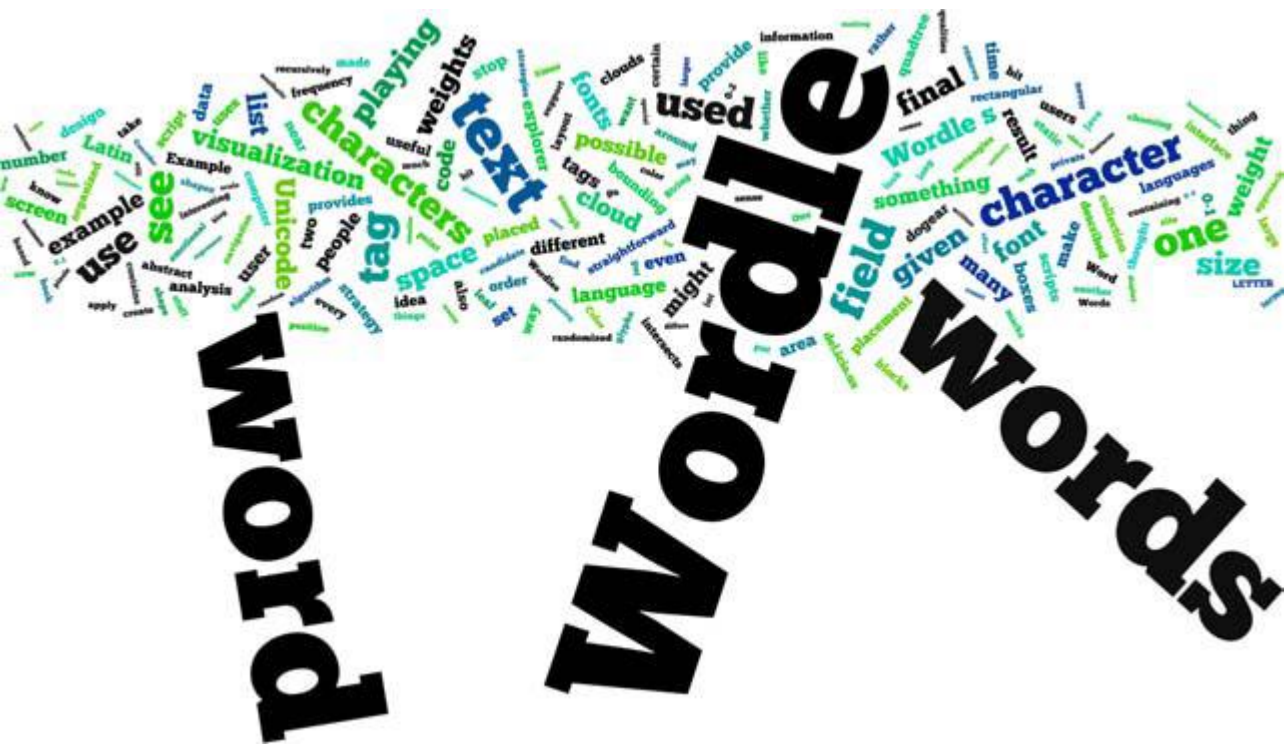


数据可视化的三要素

- ▶ 数据信息
- ▶ 有效性和精确性
- ▶ 艺术的加工



标签云



地理信息可视化



The greatest value
of a picture is when
it forces us to
notice what we
never expected to
see.

——John Tukey



参考资料

- ▶ 现代统计图形——谢益辉
- ▶ How to lie with statistics——Darrell Huff
- ▶ Beautiful Visualization——Julie Steele and Noah Iliinsky
- ▶ Wikipedia

谢谢！

