

数据可视化

之 基础篇

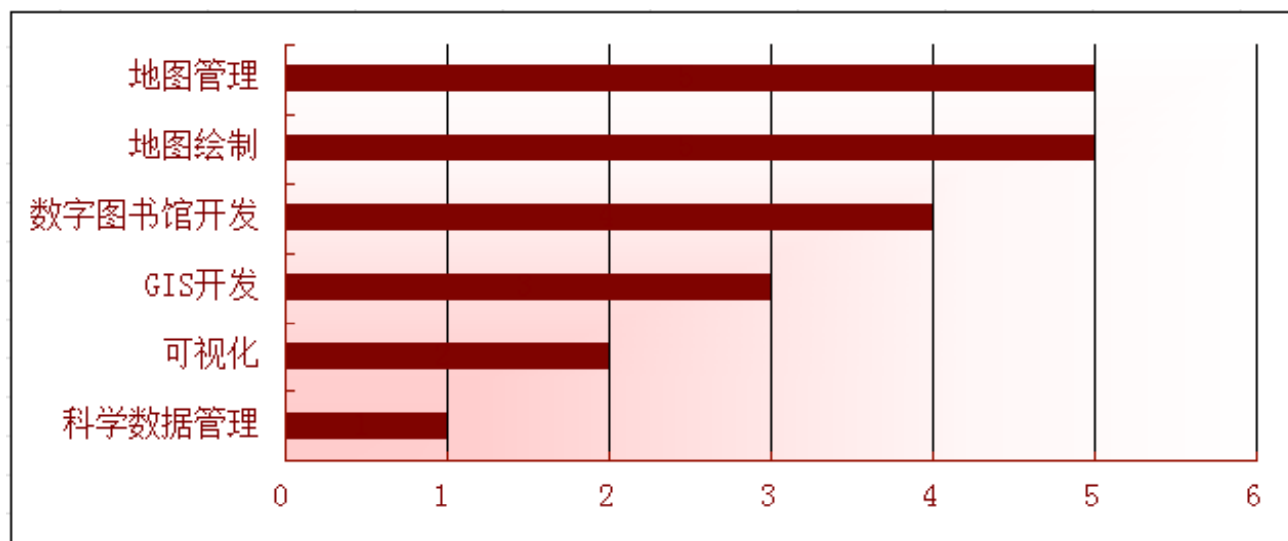
香江

信息分馆学科服务部

赖剑菲

2013-5-23

我的工作领域



70%

视觉思维





世了神曲[上传] 图片联盟 www.tplm123.com

提纲



- 几个概念

- 几种类型

- 几种方法

- 几点注意

提纲



■ 几个概念

■ 几种类型

■ 几种方法

■ 几点注意

► 大数据 (Big Data)

大数据是指无法在一定时间内用传统数据库软件工具对其内容进行抓取、管理和处理的数据集合

1. Volume

数据量巨大

全球在2010年正式进入ZB时代，IDC预计到2020年，全球将总共拥有35ZB的数据量

2. Variety

结构化数据、半结构化数据和非结构化数据

如今的数据类型早已不是单一的文本形式，订单、日志、音频，能力提出了更高的要求

3. Value

沙里淘金，价值密度低

以视频为例，一部一小时的视频，在连续不间断监控过程中，可能有用的数据仅仅只有一两秒。如何通过强大的机器算法更迅速地完成数据的价值“提纯”是目前大数据汹涌背景下亟待解决的难题

4. Velocity

实时获取需要的信息

大数据区别于传统数据最显著的特征。如今已是ZB时代，在如此海量的数据面前，处理数据的效率就是企业的生命

可视化

利用计算机图形学和图像处理技术，将数据转换成图形或图像在屏幕上显示出来，并进行交互处理的理论、方法和技术。

它涉及到计算机图形学、图像处理、计算机视觉、计算机辅助设计等多个领域，成为研究数据表示、数据处理、决策分析等一系列问题的综合技术。

可视化

- 通过使用电脑，对数据进行交互的可视表达以增强认知 [Card, Mackinlay hneiderman '98]
 - 通常被认为是一个生成图形图像的过程，实为认知过程
 - 形成某个物体的感知图像
 - 强化认知理解
 - 可视化的目的是洞悉，而非图像
 - 洞察：发现，决策，解释，分析，探索，学习
- (State Key Lab of CAD&CG, Zhejiang University, 陈为)

数据可视化

—借助于图形化手段,清晰有效地传达与沟通信息。（维基百科）

—起源于1960s计算机图形学，用计算机创建图形图表，可视化数据，直接呈现数据的属性和变量。

—数据采集、处理、保存设备长足发展，数据可视化应用领域增加，表现形式不断变化，数据可视化的概念边界不断扩大。

数据可视化

—关于数据之视觉表现形式的研究；
数据表现形式被定义为一种以某种概要形式抽提出来的信息，包括相应信息单位的各种属性和变量。

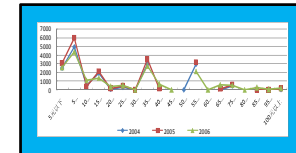
一份图表的产生过程

学号	姓名	性别	出生日期	身高	体重
1001001	张明	男	1979-01-01	175.0	65.0
1001002	李娜	女	1980-02-15	160.0	50.0
1001003	王强	男	1978-03-10	180.0	70.0
1001004	赵敏	女	1981-04-20	155.0	45.0
1001005	孙伟	男	1979-05-05	170.0	60.0
1001006	周丽	女	1982-06-18	165.0	55.0
1001007	吴昊	男	1977-07-22	185.0	75.0
1001008	郑芳	女	1980-08-08	158.0	48.0
1001009	陈杰	男	1978-09-12	172.0	62.0
1001010	林娜	女	1981-10-25	162.0	52.0
1001011	黄磊	男	1979-11-01	178.0	68.0
1001012	周娜	女	1982-12-10	160.0	50.0
1001013	吴昊	男	1977-12-20	182.0	72.0
1001014	李娜	女	1980-01-01	155.0	45.0

提炼信息

选择
图表

制作
图表



•数据表格

•理解数据

•确定比较关系

•准备作图数据

•数据图表

•分析数据

•确定图表类型

•制作和美化图表

•提炼信息

•确定图表处理

•检查确认图表

方式

•明确主题

几个取决因素

- 从数据中分析提炼的信息
- 信息所属的数据关系种类
- 想通过图表表达的观点
- 想要强调的重点

提纲



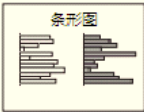
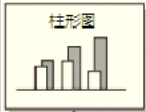
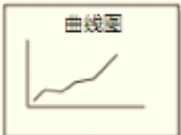
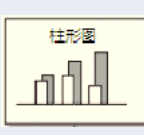
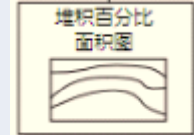
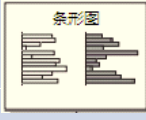
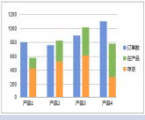
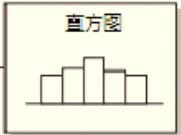
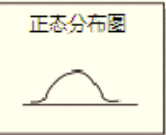
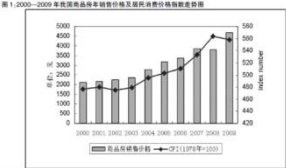
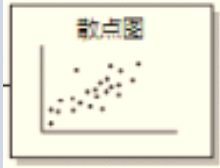
- 几个概念

- **几种类型**

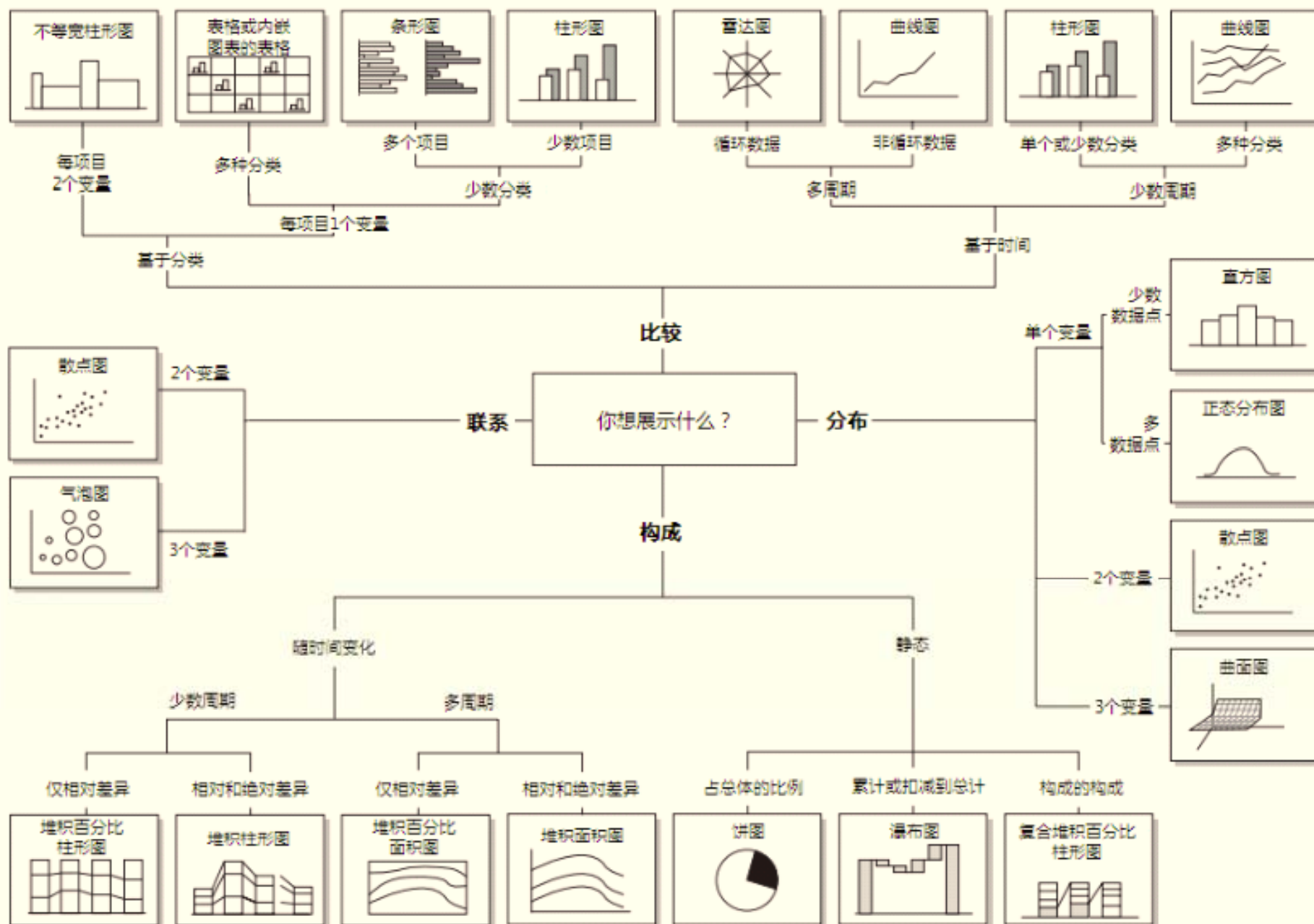
- 几种方法

- 几点注意

数据间的关系

数据关系	应用场景举例	适用的图表类型
分类比较	各科目各班级的平均分比较 排名：最畅销的公共选修课 TOP10	   
时间序列	1年中12个月的个人消费支出变化趋势	  
总体构成	个人消费支出结构 学校教学系列/研究系列专业教师的构成	   
频次分布	某次测试成绩分区的学生数分布	 
关联关系	毕业生的工资与学历之间是否存在关系 房地产价格与CPI的关系	 

图表类型选择指南



图表的基本类型及其选择

- 曲线图：用来反映随时间变化的趋势
- 柱形图：用来反映分类项目之间的比较，也可以反映时间趋势
- 条形图：用来反映分类项目之间的比较
- 饼图：用来反映构成，即部分占总体的比例
- 散点图：用来反映相关性或分布关系
- 地图：用来反映地理区域之间的分类比较

提纲



- 几个概念

- 几种类型

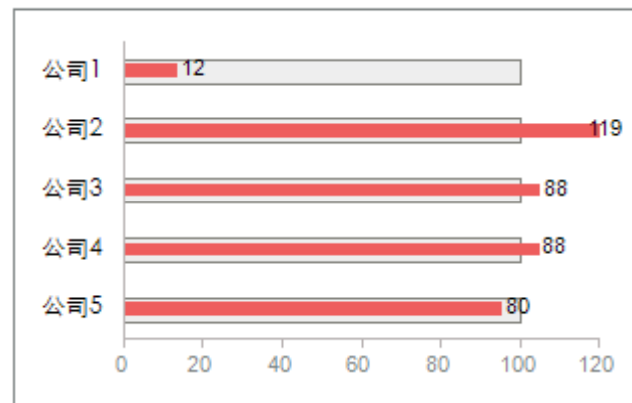
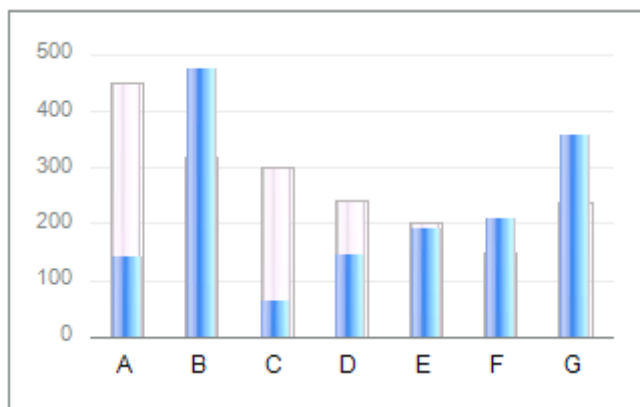
- 几种方法

- 几点注意

图表类型选择

● 柱形图 VS 条形图

- 都可表示分类项目的比较
- 当分类项目的标签文本比较长的时候，柱形图的标签会出现重叠、倾斜，而此时条形图可很好地解决问题
- 眼睛似乎更易于比较水平条形的长度



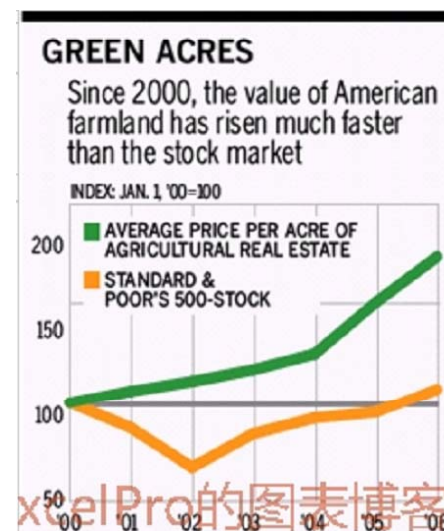
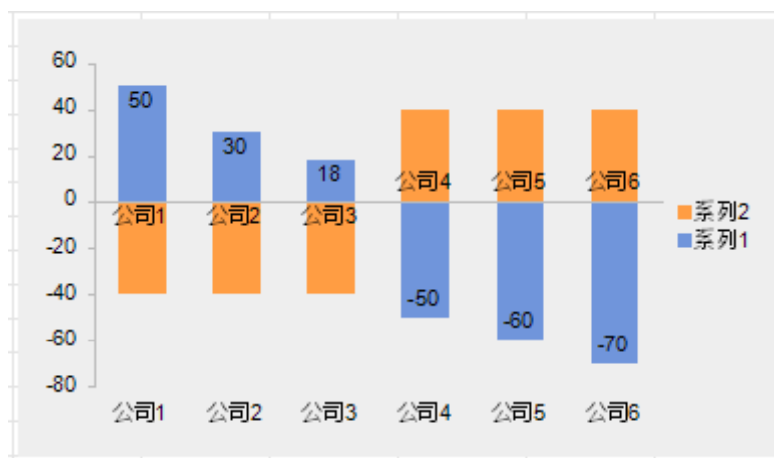
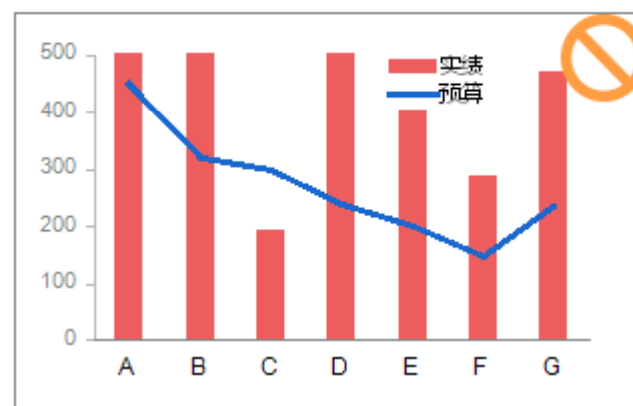
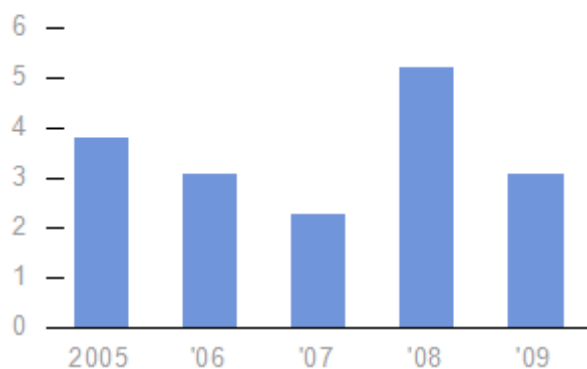
图表类型选择

● 柱形图 VS 曲线图

- 都可表示时间序列的趋势，建议采用曲线图反映趋势
- 柱形图更强调各数据点的值及其之间差异，曲线图更强调起伏变化的趋势印象，而带数据点的曲线图则兼备二者特点
- 柱形图更适于表现离散型时间序列，曲线图更适于表现连续型时间序列；数据点少可用柱形图，反则可用曲线图
- 需要放大波动幅度而适于非零起点坐标时，建议使用曲线图

图表类型选择

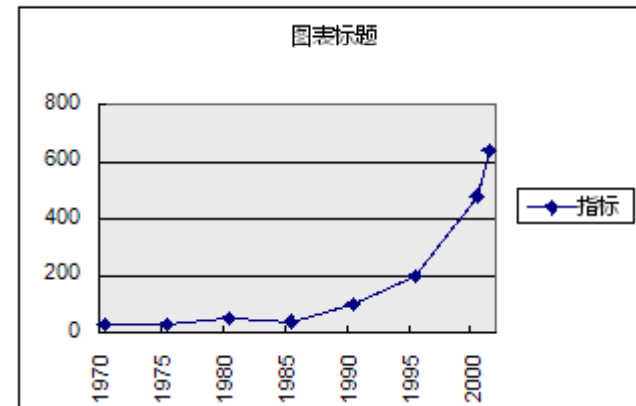
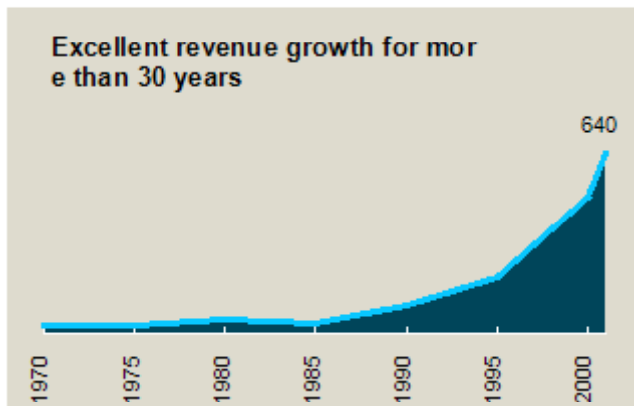
● 柱形图 VS 曲线图



图表类型选择

● 面积图 VS 曲线图

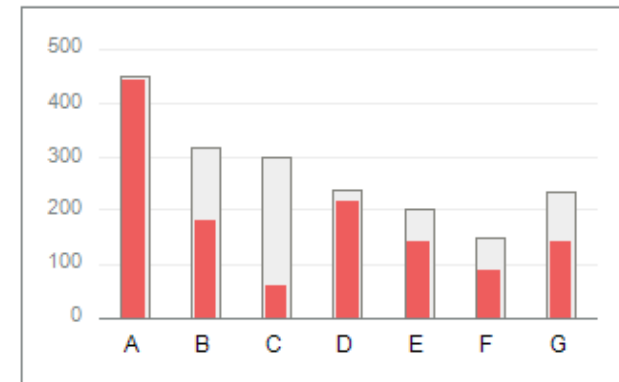
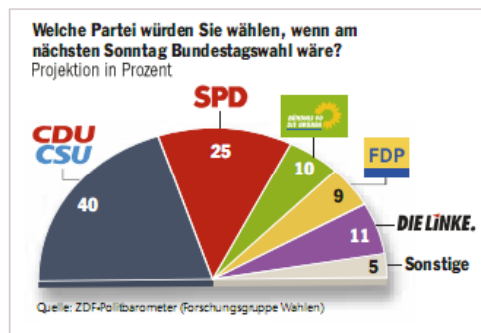
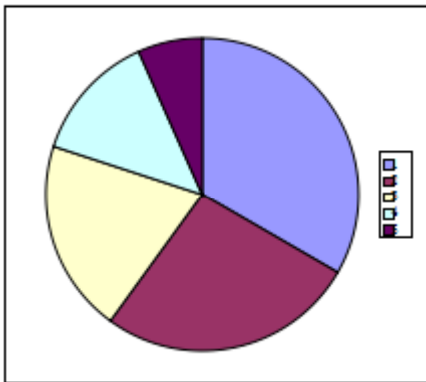
- 都可表示时间序列的趋势，当只有一个数据系列时，两者完全等价
- 面积图除可表示数据系列的发展趋势外，还可表达或比较数值随时间的积累总量



图表类型选择

● 饼图 VS 条形图

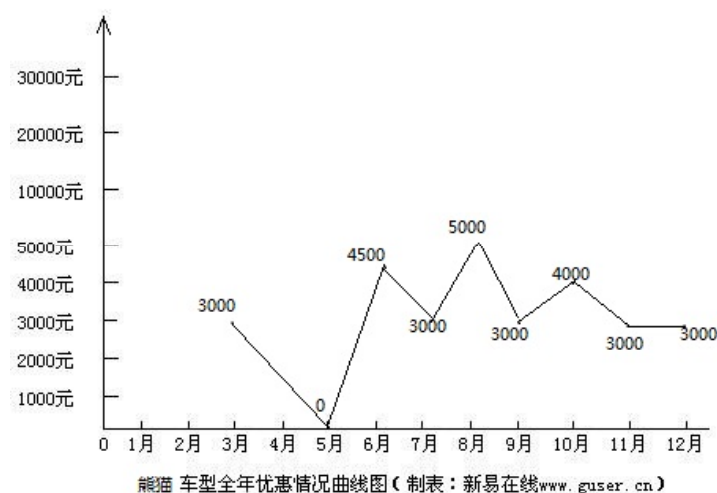
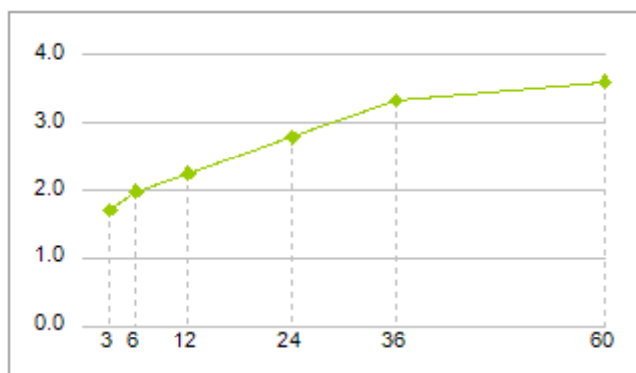
- 都可表达数据量大小的差异
- 饼图非常适宜于表示某种现象或某个事物各构成因素的百分比率，营造一种整体和构成的清晰印象
- 条形图更适宜于比较数据点的差异而非构成



图表类型选择

● 连线的散点图 VS 曲线图

- 都可用于表示基于时间序列的数据值变化
- 带连线的散点图可用来替代制作一个表现趋势的曲线图，多在X轴不等距间隔效果时使用，如不同存款期限的利息水平。



提纲



- 几个概念

- 几种类型

- 几种方法

- 几点注意

图表制作原则

- 要有明确的作用
- 简洁易懂
- 适当的注释说明
- 美观、得体

图表制作原则

●柱形图：

- 数据系列和数据点不能太多，否则应考虑其他图表类型或分解作图
- 比较分类项目时，若分类标签文字过长导致重叠或倾斜，请改用条形图

●条形图：

- 作图前最好先将数据排序，让最长的条形在最上面

图表制作原则

●曲线图：

- 曲线线条要足够粗，粗过所有非数据元素，以起到醒目提示的视觉效果
- 曲线不能太多，避免凌乱
- 不宜作分类数据的比较

●饼图：

- 分类项应少于5项或者7项，太细的归于其他
- 数据要从大到小排序，最大扇区从12点位置开始，顺时针旋转，符合阅读习惯

图表制作原则

●曲线图：

- 曲线线条要足够粗，粗过所有非数据元素，以起到醒目提示的视觉效果
- 曲线不能太多，避免凌乱
- 不宜作分类数据的比较

●饼图：

- 分类项应少于5项或者7项，太细的归于其他
- 数据要从大到小排序，最大扇区从12点位置开始，顺时针旋转，符合阅读习惯

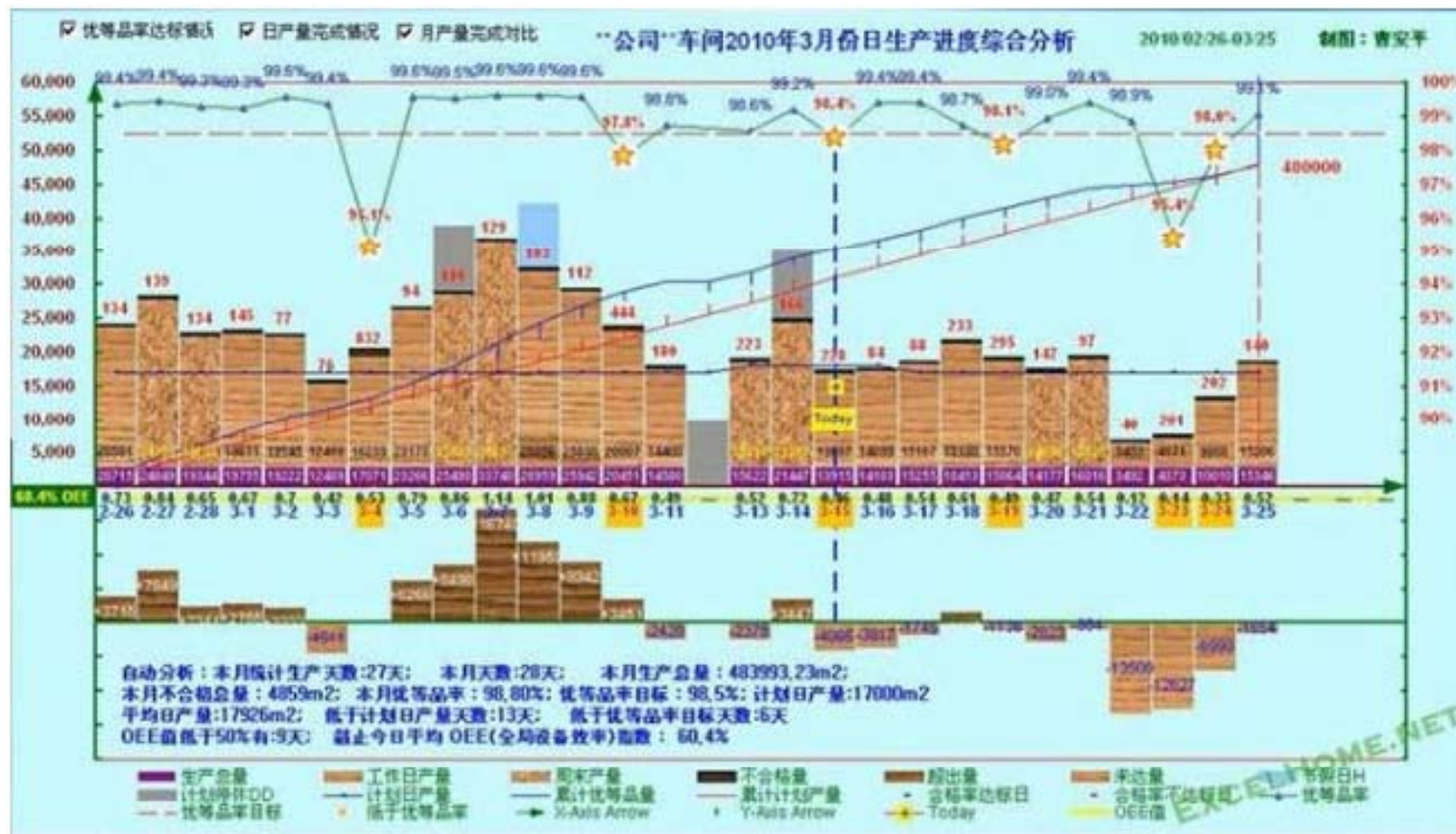
避免几个误区

●频繁改变图表类型



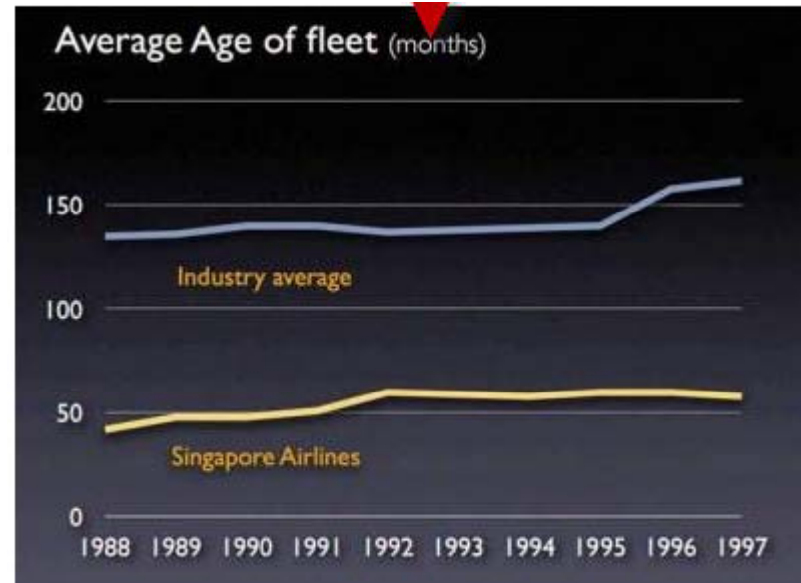
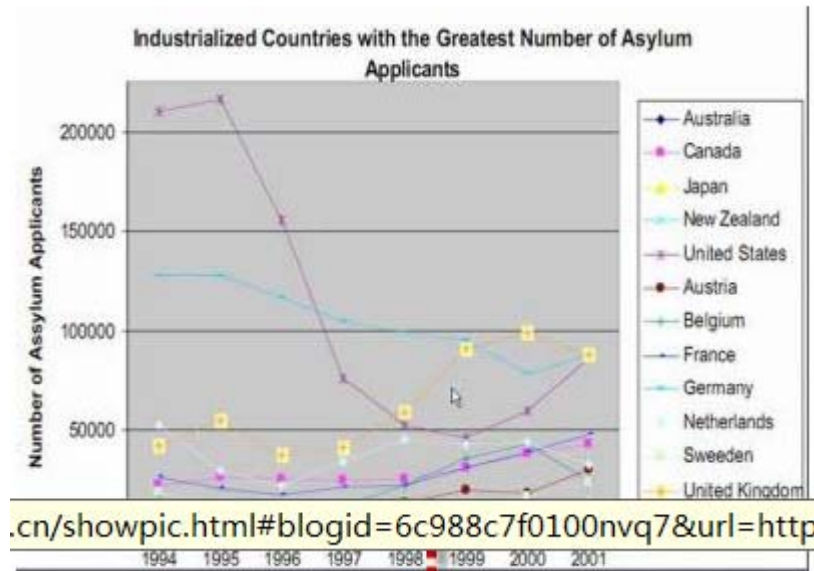
避免几个误区

●过于复杂的图表



避免几个误区

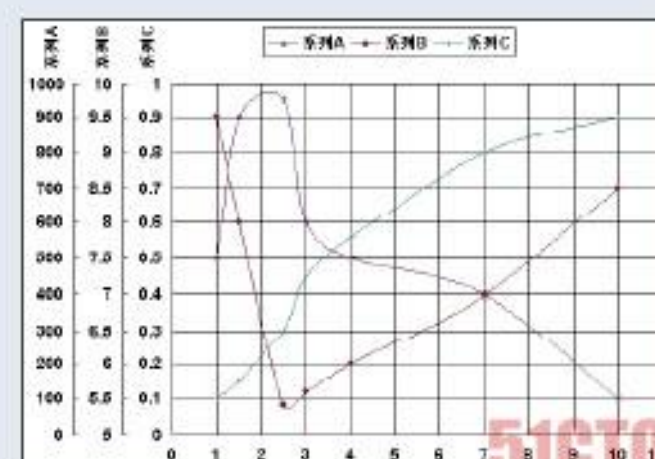
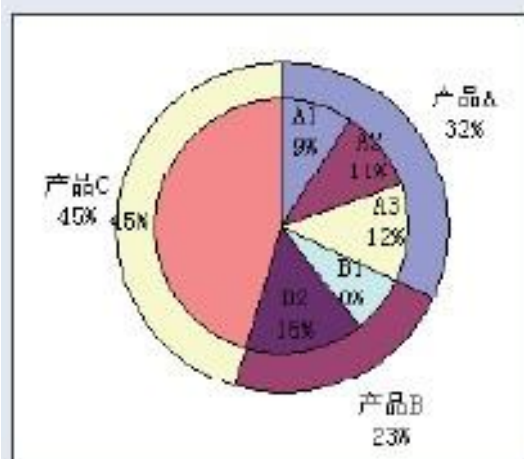
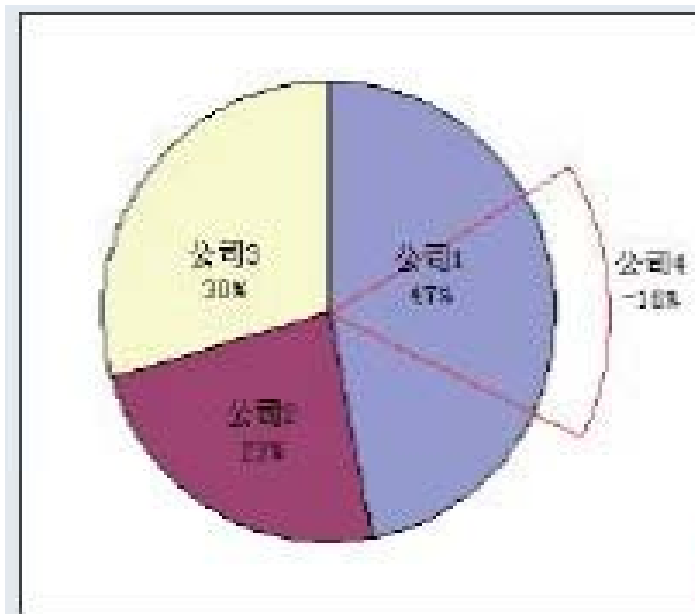
●过于复杂的图表



避免几个误区

●错误的图表

- 有负数的饼图
- 双层的饼图
- 扇区不同半径的饼图
- 多Y轴的图表



51670

避免几个误区

●不必要的图表

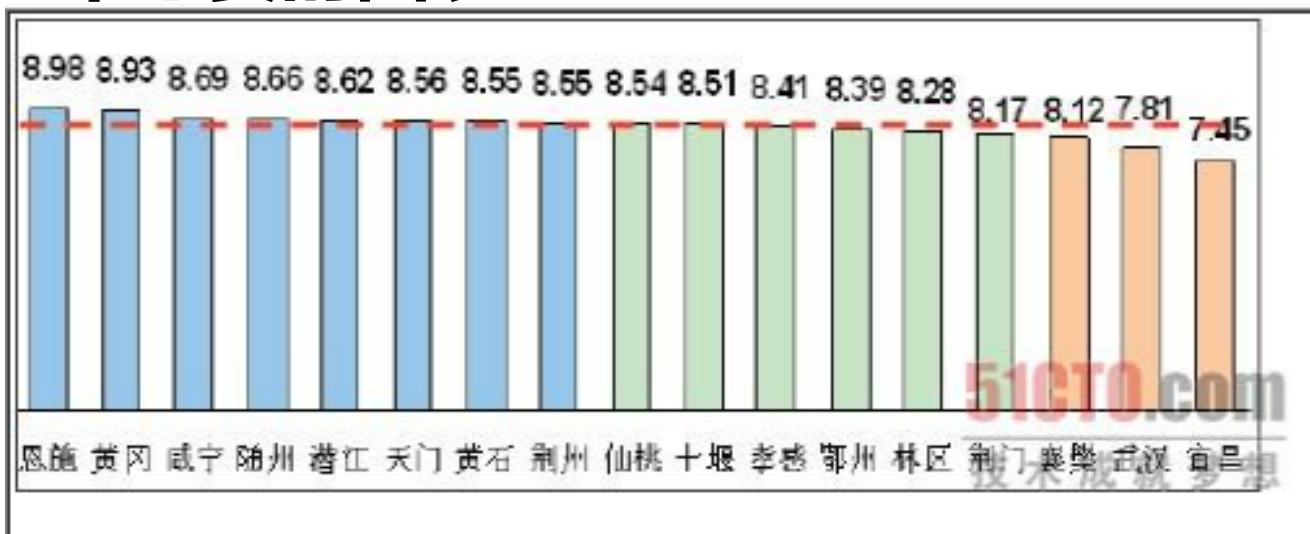


图5-10 只用数字反映关键指标，准确而且印象突出 例图来源：《商业周刊》

最后

- 记录、分析、挖掘信息

- 启发思维

- 知识发现

Q & A

■ 信息分馆读者QQ群: 150029344

■ EMAIL: jflai@lib.whu.edu.cn

■ 推荐阅读书目：

□ 《用图表说话》，基恩（美）

□ 《用数字说话》，库米（美）

□ 《Excel图表之道》，刘万祥

□ 《用地图说话》，刘万祥

Thank You

!