

描述统计分析

数据的度量、频数与概率

讲师：萨缪尔 Samuel



萨缪尔老师

网易数据分析教研负责人；

前盛大游戏战略规划总监、前腾讯游戏商业智能中心Leader

上海交通大学高金硕士、《哈佛管理导师》外部导师

- **知乎大V：「萨缪尔」** 主要聚集与商业分析、行业与战略分析、Python数据分析等
- **BAT** 互联网巨头商业洞察分析 & 咨询公司战略咨询背景
- 擅长行业趋势研究和战略管理咨询工具，**为20多家上市公司提供战略发展决策建议**

课程

亮点

1

数据的类型与度量

2

频率与概率

3

可视化图表

数据的类型与度量

数据类型		举例	含义
定性数据	定类测量	性别、颜色、血型	不能对其进行等级划分和排列
	定序测量	餐厅的星级、学历、星礼卡	对定序测量的数据进行计算是没有意义的。
定量数据	离散型	定距测量	数值并不存在直接的倍数差异
	连续型	定比测量	频次
		定距测量	温度、身高、血压
		定比测量	距离、重量、收入、速度



频数与概率

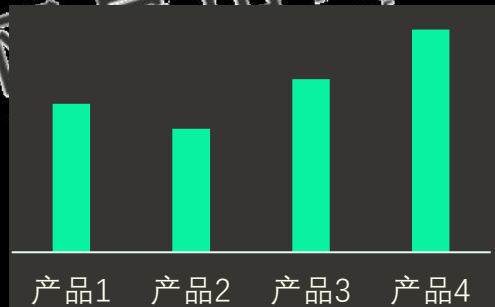
绩效	频数	概率	累计频数
A	5	5%	5
B	30	30%	35
C	50	50%	85
D	10	10%	95
合计	100	100%	100

频数、概率、累计频数

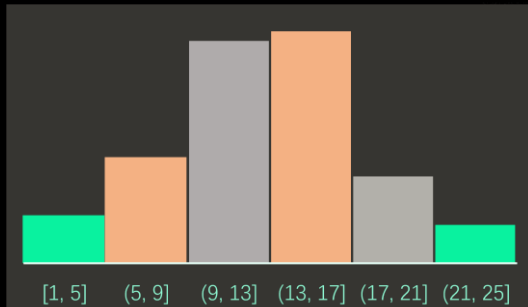
注意：频率总和必须等于1（或是100%），因为每个独立频率都是总频率的一部分。最后一个类别的累计频数必须总是等于频数总和，因为它代表该类别及其先前类别的频数的总和。

累计频数只有在数据分类有明确顺序时才有意义，也就是说对于定序、定距、定比测量的数据才可以使用累计频数，而对于定类测量的数据则不可以使用累计频数

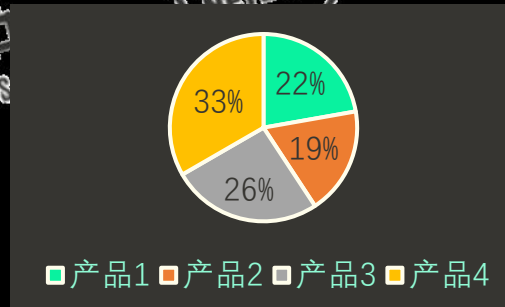
作图



柱状图



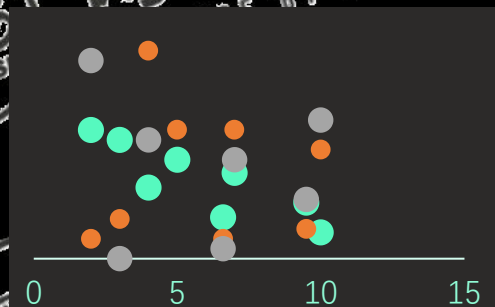
直方图



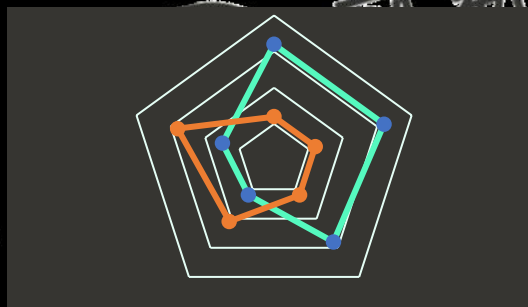
饼图



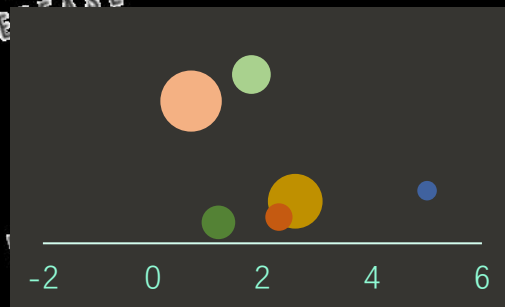
折线图



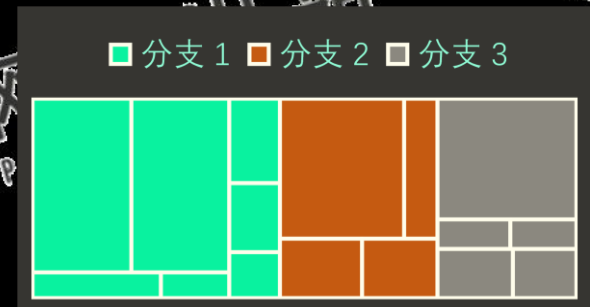
散点图



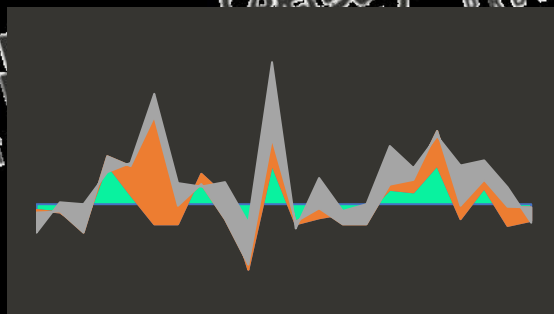
雷达图



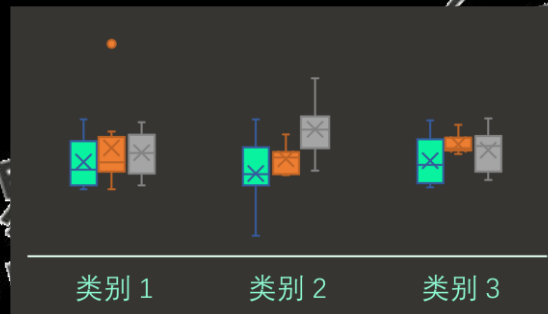
气泡图



树状图



面积图



箱型图

示例：选择合适的图表完成下表

2015年全球主要国家温室气体排放情况

国 家	CO ₂ 排放量/千吨	人均排放量/吨	国 家	CO ₂ 排放量/千吨	人均排放量/吨
世界总计	35 270 000	---	印度尼西亚	510 000	2.6
中 国	10 330 000	7.4	沙特阿拉伯	490 000	16.6
美 国	5 300 000	16.6	巴 西	480 000	2.0
欧 盟	3 740 000	7.3	英 国	480 000	7.5
印 度	2 070 000	1.7	墨西哥	470 000	3.9
俄罗斯	1 800 000	12.6	伊 朗	410 000	5.3
日 本	1 360 000	10.7	澳大利亚	390 000	16.9
国际运输业	1 070 000	---	意大利	390 000	6.4
德 国	840 000	10.2	法 国	370 000	5.7
韩 国	630 000	12.7	南 非	330 000	6.2
加拿大	550 000	15.7	波 兰	320 000	8.5

课程总结

- 1 数据的类型分为定性数据与定量数据，对于不同的数值类型选取的测量方法不同
- 2 频率总和必须等于1，累计频数只有在数据分类有明确顺序时才有意义
- 3 熟悉不同类型的可视化图表对应的使用场景

谢谢观看

