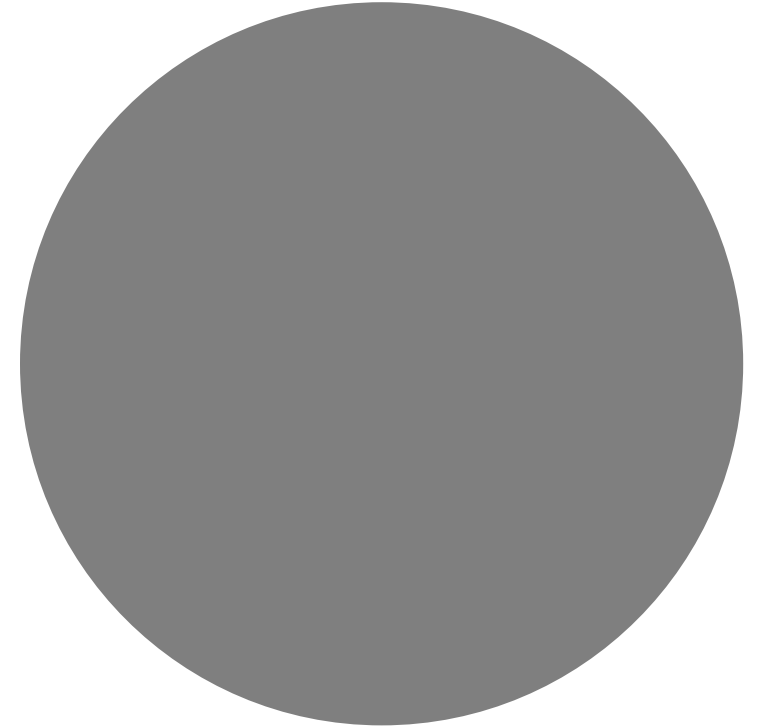


# Lesson 10: Data Visualization

---

Introduction to ggplot2 in R

Chris Qi



# ggplot2绘图系统

- 层 (Layer)

Data	感兴趣的变量 (data frame)
Aesthetics	x-axis / y-axis / color / fill / size / labels / alpha / shape / linear width / linear type
Geometries	point / line / histogram / bar / boxplot
Facets	columns / rows
Statistics	binning / smoothing / descriptive / inferential
Coordinates	cartesian / fixed / polar / limits
Themes	non-data ink

# ggplot2绘图系统

- 绘图函数
  - **qplot()**
    - 类似于Base系统的plot(), 参数包含aesthetics/geom/facet...
    - 隐藏了绘图实现的细节
  - **ggplot()**
    - 是核心, 可以实现qplot()无法实现的功能
    - 调用ggplot()本身并不能实现绘图, 要在其基础上添加层(如geom\_point())才可以

# ggplot2 绘图实践

# 探索性数据分析

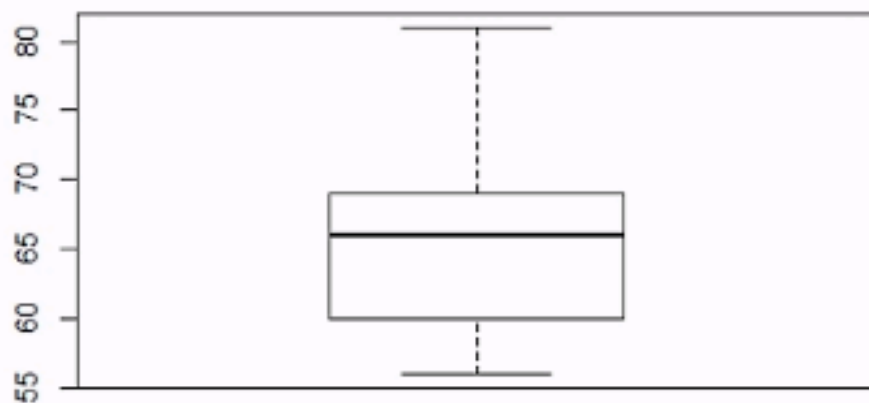
- **目的**
  - 了解数据特征、找到数据中的模式、形成分析策略
  - 图与数字互相验证、帮助发现错误、用于交流结果

# 探索性数据分析

- 特点
  - 快速、通常呈现在屏幕设备
  - 不需要过分注重图是否漂亮
  - 请注重**六大作图原则**

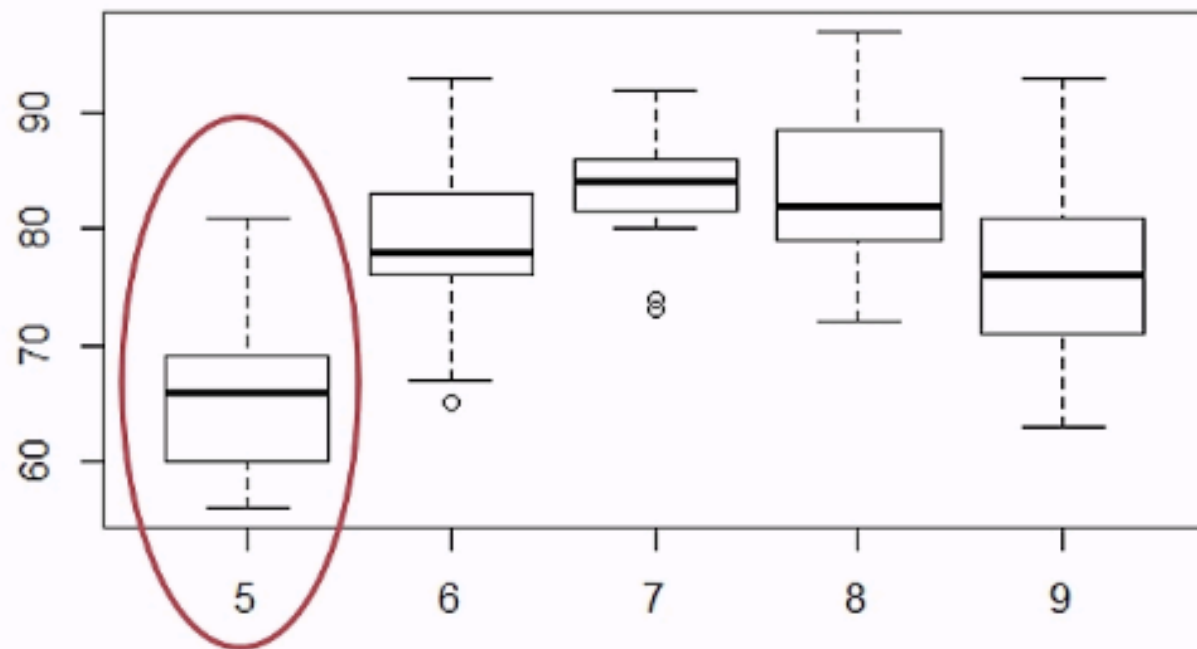
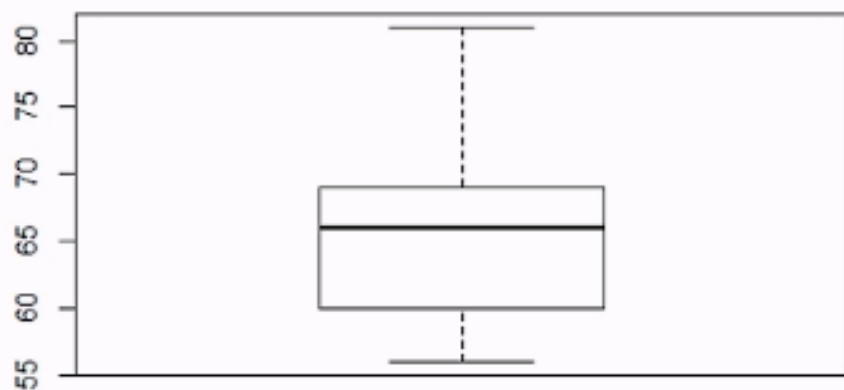
# 分析性作图的六大原则

- 凸显比较 (谁和谁比?)



# 分析性作图的六大原则

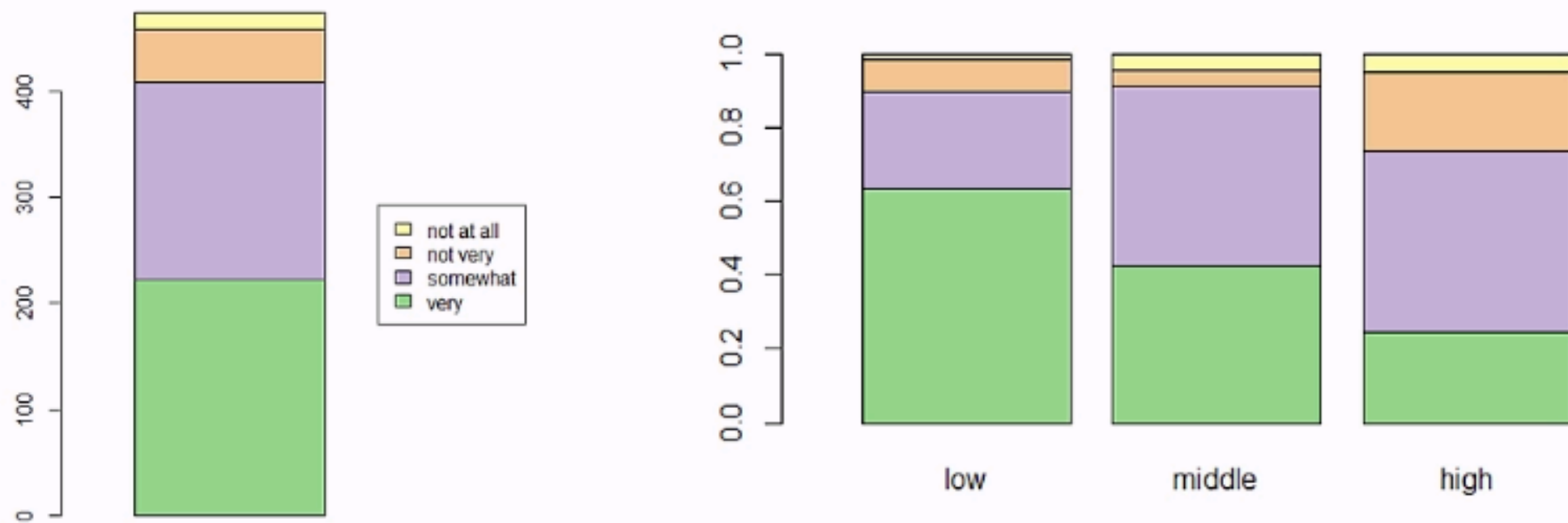
- 凸显比较 (谁和谁比?)





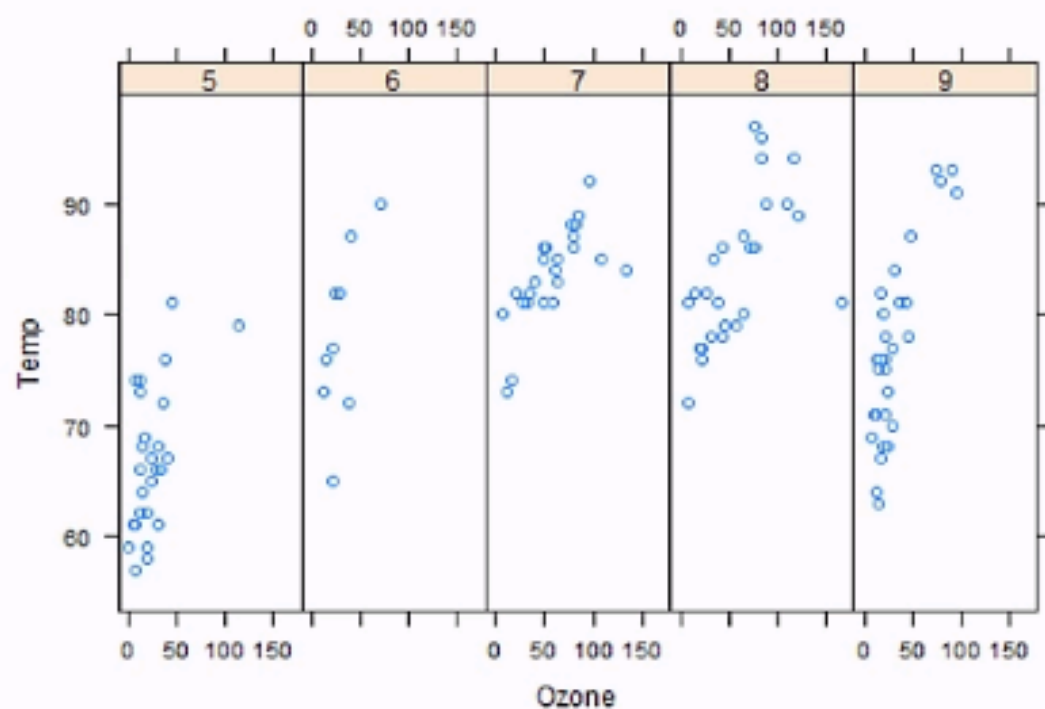
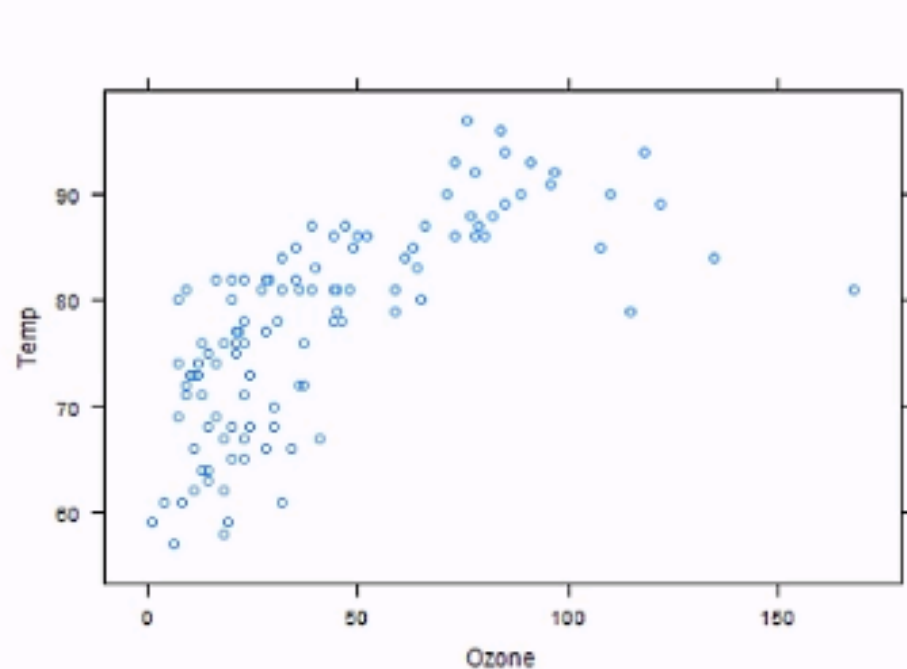
# 分析性作图的六大原则

- 凸显机制 (因果/机制/解释)



# 分析性作图的六大原则

- 凸显多元性 (>2个变量、逃离扁平)

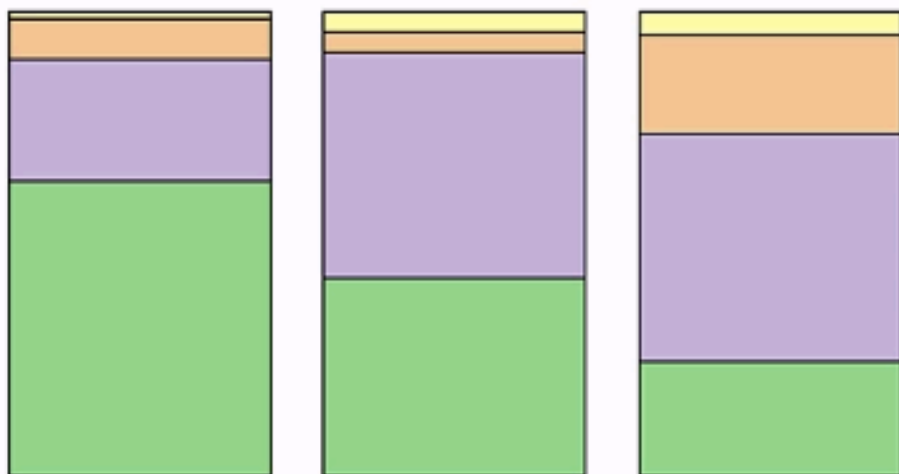


# 分析性作图的六大原则

- 整合证据
  - 整合文字、数字、图、表等
  - 用多种方式显示数据的特征
  - 不要让工具主宰分析

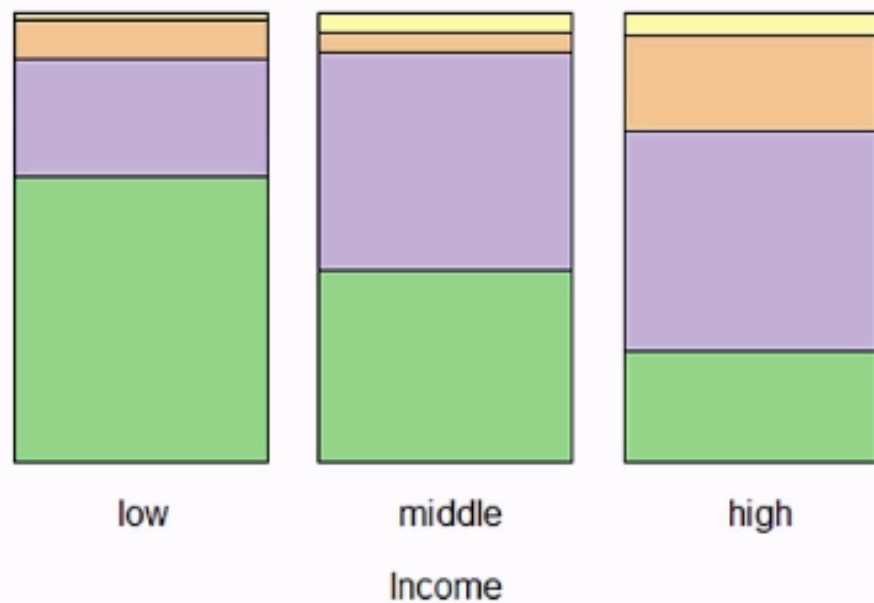
# 分析性作图的六大原则

- 使用适当的图标、尺度等
  - 完备性、一图胜千言



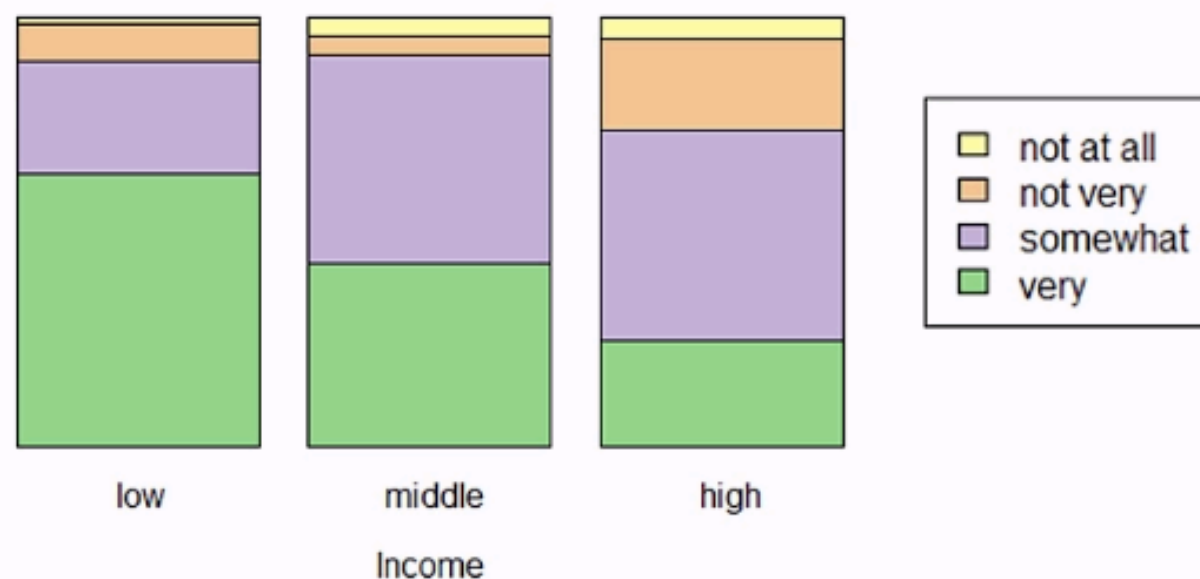
# 分析性作图的六大原则

- 使用适当的图标、尺度等
  - 完备性、一图胜千言



# 分析性作图的六大原则

- 使用适当的图标、尺度等
  - 完备性、一图胜千言



# 分析性作图的六大原则

- 内容是王道
  - 研究问题的重要性 > 作图漂亮

## 获取帮助

- 如何问问题
  - 操作系统、版本、哪一步产生的错误、预期是什么、得到的结果是什么、其他有用的信息
  - 例如：Win7 R 3.2.0 lm() “seg fault on large data frame”
- Google