



数据可视化知识点回顾

➤ 基础知识点回顾：

条形图、饼图、直方图、核密度图、箱线图和点图。

➤ 中级知识点回顾：

散点图、气泡图、折线图、相关图、马赛克图。

实训项目

数据1介绍:

选取人类死亡率数据库（HMD，2007），提供了1951-2005年瑞典人口信息。这是个多变量的数据集，变量描述如下表：

变量名称	描述
Year	年份：1951,1952，...,2005
Age	年龄：0,1,...,109
Female_Exp	女性生存人口数
Male_Exp	男性生存人口数
q_female	女性死亡率
q_male	男性死亡率
Female_death	女性死亡人数
Male_death	男性死亡人数
L_female_exp	对数女性生存人口数
L_male_exp	对数男性生存人口数

实训项目

题目1:

1.通过读取文件death rate.csv获取数据保存到df中，简单分析数据，获取共有数据多少条，是否有缺失值或是异常值,若存在这样的数据，将这些数据剔除;对于死亡率来说，它的值域是 $0 < q \leq 1$ 。

【注：1-6题中的死亡率，只考虑男性的死亡率】

2.绘制散点图，分别展示年龄、年份与男性的死亡率（对数即取log）的关系。

3.绘制年龄与对数生存人数的散点图，分析这2个量的关系。

4.绘制直方图来观察一下男性死亡人数的分布。

5.绘制**男性的对数死亡人数即（Male_death 的对数）**的直方图，来观察男性的对数死亡人口数的分布情况。

6.计算df的各变量的相关系数，并画出相关图。【用corrgram包】

实训项目

数据2介绍:

选取堪培拉房价交易指数数据，仅通过该数据进行一些可视化分析，数据显示如下表：

变量名称	描述
date	日期
index	房价指数（HPI）
year	年
month	月

实训项目

题目2:

- 1.通过读取文件House-handle.csv获取数据保存到houseIndex中。
- 2.数据探索，绘制一张图表来展示1990到2011年的HPI的变化情况，横轴是时间（可以是数据的第一列），纵轴是HPI值。
- 3.绘制一张图，展示每个月的HPI的增长量，表示为delta,在0的位置添加参考线。
【注：增长量，可以用下一条减上一条来计算；第一条的上一条的HPI值可以认为是1】
- 4.为了进一步了解HPI的波动，计算其每个月的增长率。绘图时，增长率为正数的月份由加号（“+”）表示，为负的用（“o”）表示。
- 5.对HPI增长率建立表格，其中每一行代表一个月份，每一列代表一个年份，显示前四年的数据（HPI增长率舍入到小数点后4位）；并绘制一个HPI的平均年增长率和HPI的平均月增长率（全部年份的年增长率（列平均）和月增长率（行平均））。
- 6.绘制一个箱线图，来查看HPI的增长率的分布情况。

提交内容：提交代码和结果的文件或是将题目、代码和绘制的结果，整理到word文档中，文件名：R第三次实训-姓名-学号。

THANKS

