Data Analysis Tools and Practice(Using R)

2018.03.22

R复习01

毛志来
datanalysis2018@126.com

课堂测试时间

R复习01

课堂测试04

25分钟

- 使用鸢尾花数据iris
  1）先用names()观察其结构，然后用花瓣长度和宽度做散点图
  2）在plot函数里面添加细节。修改点的形状和颜色由白色空心圆换成红色雪花；修改坐标轴名称并添加标题"relationship between width and length of Iris petal"。
- 使用 airquality 数据
  1）绘温度 Temp 直方图，加一个横坐标"Temperature",加一个标题"The Distribution of Temperature"
  2）频数变频率，并设置颜色为绿色
  3）四幅图放在一个面板里，两个一排。并使用MASS包的trueHist函数画出频率直方图：
  第一幅图，airquality里温度变量的直方图（频数）
  第二幅图，airquality里该变量的直方图（频率）并添加密度曲线，填充红色
  第三幅图，airquality里风速变量的直方图（频数）
  第四幅图，airquality里该变量的直方图（频率），并添加密度曲线，填充蓝色
- 使用mtcars里的mpg做箱图，给箱图添加坐标轴：x轴为"Number of Cylinders"，y轴为="Miles Per Gallon"标题"Car Milage Data"。根据不同cyl变量下mpg的箱线图，并添加x轴"Number of Cylinders",y轴"Miles Per Gallon"
- 按要求作图：
  1)创建字符向量colors,元素为"green","orange","brown";创建字符向量months,元素为"一月","二月","三月","四月","五月";创建字符向量regions,元素为"东部地区","西部地区","南部地区";创建矩阵values,元素为值2,9,3,11,9,4,8,7,3,12,5,2,8,10,11，要求3行5列
  2)使用矩阵values创建推叠的条形图，添加标题为"总收入"，x轴名称为"月份"，y轴名称为"收入"，条形图的标签为字符向量months(使用names.arg参数)，推叠台型图的颜色设置为创建的字符向量colors
  3)添加图例，内容为字符向量regions，分别对应条形图中的三种颜色

R复习01

Rmarkdown

RmarkDown的环境准备：https://miktex.org/download

Shiny:http://yanping.me/shiny-tutorial/

练习-0023

- 宅男Jason某日统计了DC超级英雄电影中6个英雄巨头杀敌人数如下：

superhero_kills <- c(“superman kills 1030 enemies.”, “wonderwoman kills 206 enemies.”, “aquaman kills 32 enemies.”, “cyborg kills 17 enemies.”, “batman kills 4 enemies.”, “the flash kills 0 enemies.”)

他想考考自己的女朋友Rian知不知道这些超级英雄到底打败了多少敌人，但是懒得把里面的数字一个一个改成X。他想到在Data Maniac的课堂上学习了sub（）函数，并且了解到在R中，\s代表空格，([0-9]+)代表任何出现的数字。他希望可以得到结果如下：

[1] “superman kills X enemies.” “wonderwoman kills X enemies.” “aquaman kills X enemies.” “cyborg kills X enemies.”
[5] “batman kills X enemies.” “the flash kills X enemies.”

R复习01 练习-0024

- Data Maniac 班上有很多可爱的同学，他们的信息被我们偷偷收集了，以下是班上14名同学的基本信息：

  name <- c(“Jennifer”, “Thalia”, “Ken”, “Elaine”,"Jason“,”Chris“,”Lily“,”Odelia“,”Martin“,”Isabel“,”Jane“,”Connie“,”Elisa“，”Cherry")

  age <- c(16, 17, 14, 17 , 29 , 19 , 21, 18, 19, 23, 17, 19, 22, 21)

  hair <- c(“black”, “green”, “black”, “brown”, “white”, “black”, “purple”, “black”, “blue”, “black”, “green”, “silver”, “green”, “black”)

  1）为了更好的分析，我们用以下方法把这些数据转变为data.frame，并命名为classmates。

  2）你需要协助Jason用课上学到的dplyr包完成以下任务： 1. 同学中选出所有19岁及以下的； 2. 在上一个任务的基础上选出黑色头发的； 3. 接着根据年龄从大到小将这些黑发及19岁以下的同学进行排序； 4. 计算出满足以上条件同学的平均年龄及最小年龄,分别命名为mean_age1和

  3）计算出所有14位同学的平均年龄，最大及最小年龄，分别命名为mean_age2和max_age及min_age2。

R复习01 练习-0025

- 使用数据集airquality回答下列问题
  1）使用str()函数来观察airquality这个数据的变量有那些:
  2）用函数计算第三个变量（风速）的平均值，最小值，最大值和标准差:
  3）使用pdf("mygraph.pdf") 将上面的图形保存到你的作业文件夹（本地硬盘）
  4）用plot()函数创建风速与风度的散点图：添加回归曲线和标题“Weather in NYC”：

- 使用R自带的数据集cars画出散点图，颜色设置为彩虹色，形状为编码为1:10的图形。主标题为“speed and diantance”，主标题颜色为蓝色，主标题缩放比例为1.5，字体为2，副标题为“scatter plot”，副标题颜色为灰色，主标题缩放比例为1.2

- 画出数据框cars的speed列的频率直方图，主标题为“speed hist”，主标题颜色为蓝色，主标题缩放比例为1.5，字体为2，副标题为“histogram exercise”，副标题颜色为灰色，主标题缩放比例为1.2，y轴范围为0到0.1
  添加密度曲线，要求颜色为红色，线段类型为虚线，宽度为2

- 使用R数据集VADeaths,查看这个数据集，画出各个年龄段死亡率的箱型图，要求并排排列，颜色为前4个彩虹色，添加图例，图例名称为VADeaths的列名，y轴范围为0到100，主标题为“VADeaths barplot”，主标题颜色为蓝色，主标题字体为2，副标题为“barplot exercise”，副标题颜色为灰色，主标题缩放比例为1.5

R复习01 练习-0026

●dapengde_DummyR_PM25.csv是2003 年 8 月在北京城区的三个高度（8 米，100 米，325 米）测得的 PM2.5 的质量浓度日变化的统计数据，共 4 列 25 行。

1）请画出一条折线表示h8和time的关系，要求是"time"和"pm2.5" 分别是x轴的名称和y轴的名称，lty=1（表示line的type为1，表示直线） y轴的范围是0到200.

2）在上图增加一条折线(使用lines()函数)表示h100和time的关系，要求颜色为红色，线型为虚线(lty=2)

3）在上图中增加图例来表示上边画的两条折线，其中图例位置为（x=15，y=180）位置处，内容为8m和100m,两条折线分别为黑色直线和红色虚线。

4）画出x轴，刻度指定为和时间相对应的24个小时。

5）与h8和h100两条折线相对应，画出其对应的y轴均值的水平线。

R复习01　　练习-0027

图表表示的某种商品上一周与本周销量的对比图，请根据表格中的数据创建矩阵，并完成那个下列的作图要求：

1) 将各组数据用条形图表示，要求水平、并列的方式，上周和本周的颜色分别为黄色和红色，不添加坐标轴
2) 在底部添加水平坐标轴
3) 在左侧添加垂直坐标轴，要求在位置2,5,8,11,14,17,20处，标签为'Mon"Tue"Wed"Thur"Fri"Sat"Sun',不显示刻度

	pre	now
1	113	123
2	134	145
3	123	136
4	145	178
5	123	113
6	234	167
7	145	220

谢谢！

毛志来
datanalysis2018@126.com