

课程复习II



课堂测试时间

- 1、查看数据集mtcars，根据要求作图：
 - (1)画出car与GDP关系，设置图形属性颜色由region决定，并将省份名称标注在图上，注意避免省份标签的文本不会重叠；
 - (2)画一个块面图表示car与GDP的关系，要求根据region分组，两行
 - (3)画出car与GDP关系的拟合直线，直线两侧不显示置信区间
 - (4)画出car与GDP关系的折线图
 - (5)画出region变量的条形图
 - (6)画出region变量的直方和频率折线图，指定组距为0.5
 - (7)使用region作为分类变量，画出GDP变量的画箱线图
 - (8)如果我们只想看人均GDP在2至4之间，平均汽车保有量在10至20之间的省份的散点图，可以使用xlim()、ylim()函数限定坐标轴范围。X轴标签为'Per
 - (9)画出region变量的条形图，要求根据region来进行颜色填充，不显示图例，组距宽度为1，xlab,ylab为空，反转x,y坐标

- 2、使用ggplot2包的数据集mpg,画出cty和displ关系的散点图，并添加随机抖动
- 3、根据要求绘制地图：
 - (1)加载ggplot2和maps包，获取数据euro指定map为“UK”,“France”,“Spain”,“Germany”,“Italy”
 - (2)使用euro绘制地图，表示x轴表示long变量，y轴表示lat变量，使用group变量分组，填充颜色根据region来指定，geom_polygon来绘制边，颜色指定为黑色，调色板，y轴范围为(40, 60)(使用scale_y_continuous()函数)，x轴范围为(-25, 25)(使用scale_x_continuous()函数),标题为Euorpe's Big Five Football Leagues

- `ggplot()`, 图层
 - * `data; mapping; geom; stat; position; aes(); layer();`
- `geom_xxx`:
 - * `point; line; path; bar; histogram; smooth; density; jitter; text; line; line; abline; tile; area; polygon;`
- `stat_xxx`:
 - * `identity; smooth; function; boxplot; density; quantile; sum; unique; bin; stat_bin2d`
- 其余:
 - * `fill; bins; colour; group; labs; binwidth; shape; alpha; maps;`

课堂测试复习

- 1、10位同学的姓名、性别、年龄、身高、体重数据如下：

| Name | Sex | Age | Height | Weight |
|-----------------|-----|-----|--------|--------|
| Alice Harden | F | 13 | 56.5 | 84.0 |
| Sandy Muller | F | 11 | 51.3 | 50.5 |
| Sharon Wilshere | F | 15 | 62.5 | 112.5 |
| Tammy Wenger | F | 14 | 62.8 | 102.5 |
| Alfred Ferguson | M | 14 | 69.0 | 112.5 |
| Duke klopp | M | 14 | 63.5 | 102.5 |
| Guido Conte | M | 15 | 67.0 | 133.0 |
| Robert Mourinho | M | 12 | 64.8 | 128.0 |
| Thomas Bryant | M | 11 | 57.5 | 85.0 |
| William Curry | M | 15 | 66.5 | 112.0 |

- 1) 根据以上信息构造一个数据框，数据框的名称为stuinfos
- 2) 分别计算全体学生年龄、身高、体重的和、平均值、标准差，并且设置数据的输出格式为小数点后两位
- 3) 依姓氏和名字对数据集进行排序：
- 4) 将上述数据写成(`write.table()`)一个纯文本的文件（文件名为class.txt），并用函数`read.table()`读取文件中的身高和体重数据

- 1、数据集alpe_d_huez2描述了环法自行车赛期间Alpe d'Huez赛段的最快时间，以及关于年份和吸毒指控的背景信息。绘制出车手最快时间的分布。使用a) 直方图和b) 箱线图显示它们。
- 2、mtcars是datasets包中的数据集。请使用str()函数了解这个数据集的构成，并输出数据集，然后按要求画图：
 - * a. 我们要设置一个蓝色背景和红色的点或线。我们应该使用什么命令
 - * b. 画出cyl和mpg关系的散点图，并将结果输出为plot.png，要求输出为白底，360px*360px,点的大小为72
- 3、obama_vs_mccain数据集描述了2008年美国总统选举中的各州投票信息，以及关于收入，失业，种族和宗教的背景信息。
 - * a. 画出收入Income和参加选举比例Turnout之间的关系散点图。提示：Turnout存在Na值。
 - * b. 将上述图形点的形状为黑色实心三角形(17)
 - * c. 数据集中有一个因子类型的列regions,请画出每个地区region下的收入Income和参加选举比例Turnout之间的关系散点图。要求设置布局为5列，行优先。

- 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。(1)、使用矩阵values创建推叠(堆积)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), 推叠条形图的颜色设置为创建的字符向量colors; (2)、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。
- 2、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。

- 3、右图表示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 |

- 1、查看数据集mtcars， 根据要求作图：
 - (1)分别使用qplot、ggplot函数画出mpg和wt关系的散点图；
 - (2)使用三种方式画出mpg列的直方图,同时在使用qplot和ggplot时指定每个小圆柱体的宽度是4；
 - (3)使用三种绘图函数画出mpg变量的密度曲线。
- 2、使用datasets包中的数据集中的pressure， 查看其数据并按要求画图：
 - (1)请画出pressure和temperature关系的曲线图；
 - (2)分别使用qplot和ggplot画出pressure和temperature关系的散点图和折线图。
- 3、使用datasets中的数据集中的ToothGrowth， 完成如下的绘图要求：
 - (1)以supp变量作为分类,分别使用三种绘图函数画出len变量的箱型图。
- 4、使用ggplot2包中数据集mpg， 完成练习：
 - (1)使用mpg数据集定义一个ggplot对象，表示hwy与cty的关系；
 - (2)画一个散点图，指定颜色有year列来指定，并在上边绘图的基础上画出平滑的拟合曲线；
 - (3)继续使用(1)中定义的ggplot对象画散点图，使用class来指定颜色，displ指定大小，透明度；指定为0.5,position指定为抖动，在散点图的基础上添加拟合曲线；
 - (4)使用qplot画出hwy与cty的关系的散点图，并根据year变量分面，同时添加拟合曲线。

练习

- 某校测的19名学生的四项指标：性别、年龄、身高（cm）、体重（磅），具体见0022_student.CSV，要求：
 - * 绘出体重对于身高的散点图
 - * 绘出不同性别情况下，体重与身高的散点图
 - * 绘出不同年龄段的体重与身高的散点图
 - * 绘出不同性别和不同年龄段的体重与身高的散点图
-
- 0022_height01.txt, 画直方图
 - 0022_height02.txt, 画箱式图
 - 0022_marriage.txt, 画散点图
 - 0022_language.txt, 画条形图（母语和日常使用）
 - 0022_language.txt, 画饼图（世界主要语种使用人数比例）

- 从0018_grade.csv中读取两班成绩
- 计算每个班级的均值和标准方差
- 计算每个人的标准化成绩，添加到数据中，写到0023grade.txt中
- 分别画出来两班成绩和标准成绩的箱线图
- 在一张图中画出两班成绩和标准成绩的箱线图

- 模拟产生100个学号 (1300022001到1300022100)
- 模拟产生三个科目的成绩，要求第一科最大值99，最小值70；第二科平均值81， $sd=7$ ，最大值100；第三科平均值83， $sd=18$ ，最大值100
- 把学号和三科成绩组成一个数据框，显示数据框内容
- 求每个学生的总分、平均分
- 针对三科成绩、总分、平均分，分别做饼图、直方图、条形图，箱线图
- 分别用par和layout把多个图放在一个图中显示：同一个数据的不同类的图形，不同数据的同一类，不同数据的不同图形

```
plot(rnorm(1000), col="red")
```

- 使用上面的语句，练习颜色的各种表示方法
 - 使用Par和layout函数，分别实现不同颜色的多个图形组合，2*2，3*3，1*1*2*3等
-

课件第12页，citysales.csv

- 输入现有代码，看显示结果
- 用rainbow、top.colors、cm.colors、gray、terrian.colors替换heat.colors，看执行效果
- 练习课件第23页的颜色参数
- 添加图例

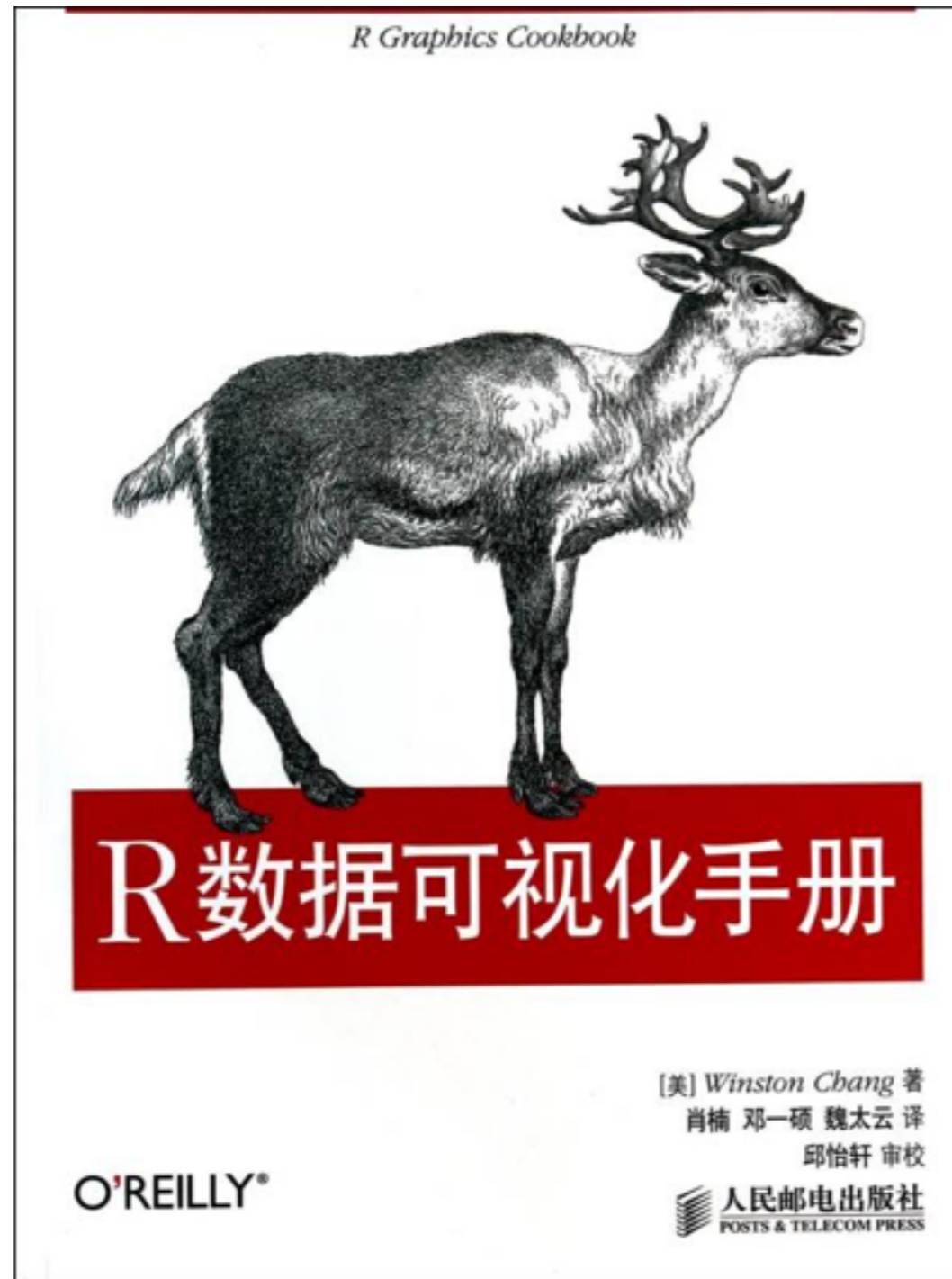
cityrain.csv

- 用不同颜色画出不同城市的线图
- 用不同符号画出不同城市的线图
- 用不同颜色画出不同城市的散点图
- 用不同符号画出不同城市的散点图
- 分别加上图例
- 用par和layout把前面四个图放在一张图中，分别为
 $2*2$ ， $1*4$ ， $1+2+1$

- `gdp_long.txt`
 - 做折线图（网格、特殊线，图例的不同位置）
 - 条形图（正常、堆积、横向、颜色宽度等、显示数字、误差线）
-
- `cityrain.csv`
 - 做折线图（边界标注，`slide`，`mar`和`bty`的含义）

- 用qplot和ggplot重新做前面所有画图的练习题
- 0022、0023、0024、0025、0026、0029
- 课堂测试06、课堂测试07

- R数据可视化手册的各章，熟悉所有例子。



大作业

- 完成课后大作业0003
- 按照要求提交

谢谢!

孙惠平

sunhp@ss.pku.edu.cn