

talk01 练习与作业

目录

| | | |
|-----|-----------------------|---|
| 0.1 | 练习和作业说明 | 1 |
| 0.2 | talk01 内容回顾 | 1 |
| 0.3 | 练习和作业说明 | 2 |
| 0.4 | 练习与作业：用户验证 | 2 |
| 0.5 | 练习/作业 1 | 2 |
| 0.6 | 练习/作业 2 | 2 |

0.1 练习和作业说明

将相关代码填写入以 “{r}” 标志的代码框中，运行并看到正确的结果；

完成后，用工具栏里的”Knit” 按键生成 PDF 文档；

将生成的 PDF 改为：姓名-学号-talk01 作业.pdf，并提交到老师指定的平台/钉群。

0.2 talk01 内容回顾

- R 背景介绍
- 几个简单示例
 - R 流行度调查
 - R job trends

0.3 练习和作业说明

将相关代码填写入以 “{r}” 标志的代码框中，运行并看到正确的结果；

完成后，用工具栏里的 **Knit** 按键生成 PDF 文档；

将得到的 PDF 文件更名为：姓名-学号-talk## 作业.pdf。

将 word 文档提交到老师指定的平台/钉群。

0.4 练习与作业：用户验证

请运行以下命令，验证你的用户名。

如你当前用户名不能体现你的真实姓名，请改为拼音后再运行本作业！

```
Sys.info()[["user"]]
```

```
## [1] "wchen"
```

```
Sys.getenv("HOME")
```

```
## [1] "/Users/wchen"
```

0.5 练习/作业 1

安装 R 和 RStudio 最新版；

用 RStudio 打开本脚本，按提示操作安排相应的包；

成功运行本脚本，并输出作业；

0.6 练习/作业 2

重现 talk01 中的两个图。

1. 安装需要的包：

2. R 的流行性调查

运行以下代码，得到 R 的流行性调查；

注意输入文件：chaper01_preface_scholarly_impact_2012.4.9.csv 的正确路径；

注意理解每行代码的意义；

```
library("ggplot2"); library("reshape2");

dat <- read.csv(file = "../data/talk01/chaper01_preface_scholarly_impact_2012.4.9.csv")

cols.subset <- c("Year", "JMP", "Minitab", "Stata", "Statistica", "Systat", "R");
Subset <- dat[, cols.subset];
ScholarLong <- melt(Subset, id.vars = "Year");
names(ScholarLong) <- c("Year", "Software", "Hits");

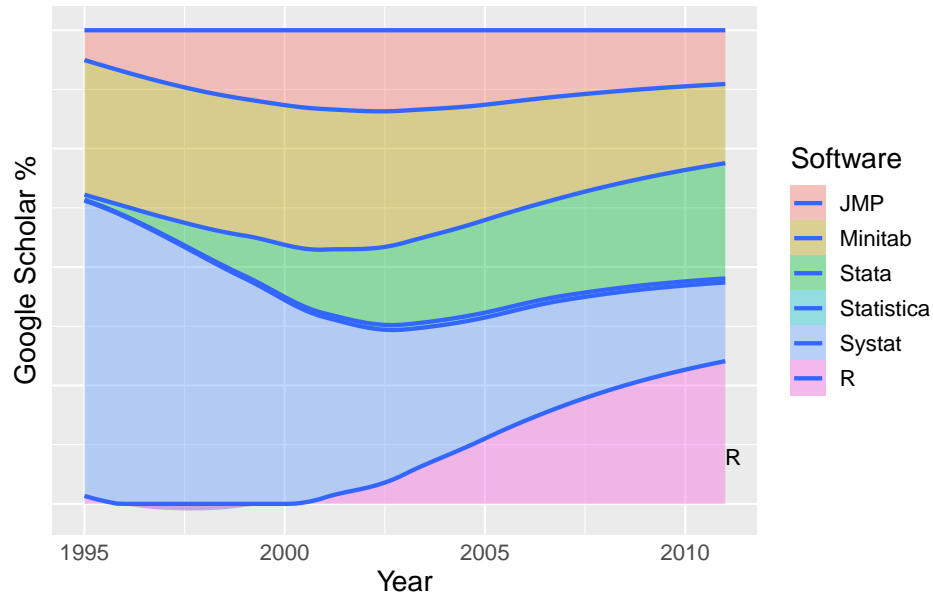
plot1 <-
  ggplot(ScholarLong, aes(Year, Hits, group=Software)) + # 准备
  geom_smooth(aes(fill=Software), position="fill", method="loess") + # 画图
  ggtitle("Market share") + # 设置图标题
  scale_x_continuous("Year") + # 改变 X 轴标题
  scale_y_continuous("Google Scholar %", labels = NULL) +
  theme(axis.ticks = element_blank(), text = element_text(size=14)) +
  guides(fill=guide_legend( title = "Software", reverse = F )) +
  geom_text(data = data.frame( Year = 2011, Software = "R", Hits = 0.10 ),
    aes(label = Software), hjust = 0, vjust = 0.5);

plot1; ## 画图

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Stacking not well defined when not anchored on the axis
```

Market share



3. R 的招聘趋势

运行以下代码，得到 R 的招聘趋势

注意输入文件：chaper01_preface_scholarly_impact_2012.4.9.csv 的正确路径；

注意理解每行代码的意义；

```
library("ggplot2"); ## 主作图包

##2. -- 读取数据 --
dat <- read.table(file = "../data/talk01/chaper01_preface_indeed_com_stats_2015.txt",
                  header = T, as.is = T);

##3. 处理数据
dat$date <- as.Date(dat$date); ## 把第一列改为日期

# 根据 job 对 software 进行调整
dat <- transform(dat, software = reorder(software, job));
```

```
plot2 <-
  ggplot( dat, aes( date, job, group = software, colour = software) ) +
    geom_line( size = 0.8 ) +
    ggtitle("Job trends (data from indeed.com)") + # 设置图标题
    xlab("Year") + ylab("%") +
    # 改变字体大小; 要放在 theme_grey() 后面
    theme( text = element_text(size=14) ) +
    guides(colour=guide_legend( title = "Tool", reverse = TRUE )) +
    scale_colour_brewer(palette="Set1") + # 改变默认颜色
    geom_text(data = dat[dat$date == "2015-01-01" & dat$software %in% c("R"), ],
              aes(label = software), hjust = 0, vjust = 0.5);
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
```

```
plot2;
```

