

数据的读入操作

周世祥

2020年5月20日

数据读入

备注：本节内容涉及大量路径的设定，此处以windows系统路径为例，读者在运行时需要根据自己的情况调整！

清空工作空间

```
rm(list = ls())
```

从txt中读入

输入完整路径，可以顺利读入文件,下面命令可读入存在Downloads文件夹中的movie文件，用户请根据自己的文件路径修改运行

```
movie_txt = read.table("K:/Data-Science-in-Action-R-Tools-and-Case-Studies-master3/R与千寻（代码+数据）/第二章 R语言数据操作/2.2 数据读入/2.2.1 结构化数据读入/movie.txt", header = T, fileEncoding = "UTF-8")
head(movie_txt)
```

```
##           name boxoffice doubanscore type duration  showtime
## 1           叶问3   77060.44         6.4 动作      105  2016/3/4
## 2           美人鱼  338583.26         6.9 喜剧       93  2016/2/8
## 3      女汉子真爱公式   6184.45         4.5 喜剧       93 2016/3/18
## 4 西游记之孙悟空三打白骨精 119956.51         5.7 喜剧      120  2016/2/8
## 5           澳门风云三 111693.89         4.0 喜剧      112  2016/2/8
## 6      功夫熊猫3   99832.53         7.7 喜剧       95 2016/1/29
##  director    star1 index1      star2 index2
## 1    叶伟信    甄子丹  11385      张晋   4105
## 2    周星驰    邓超   41310      林允   9292
## 3    郭大雷    赵丽颖 181979      张翰  44277
## 4    郑保瑞    郭富城  12227      巩俐   8546
## 5      王晶    周润发  16731      刘德华 30277
## 6    吕寅荣  杰克布莱克    178 安吉丽娜朱莉  1540
```

用getwd获得工作目录

```
getwd()
```

```
## [1] "K:/Data-Science-in-Action-R-Tools-and-Case-Studies-master3/R与千寻（代码+数据）/第二章 R语言数据操作/2.2 数据读入/2.2.1 结构化数据读入"
```

先把movie文件转移到该工作目录下，即可采用以下命令直接用文件名读取

```
movie_txt = read.table("movie.txt", header = T, fileEncoding = "UTF-8")
head(movie_txt)
```

```
##              name boxoffice doubanscore type duration  showtime
## 1              叶问3   77060.44         6.4  动作      105  2016/3/4
## 2              美人鱼  338583.26         6.9  喜剧       93  2016/2/8
## 3          女汉子真爱公式    6184.45         4.5  喜剧       93  2016/3/18
## 4  西游记之孙悟空三打白骨精 119956.51         5.7  喜剧      120  2016/2/8
## 5              澳门风云三 111693.89         4.0  喜剧      112  2016/2/8
## 6              功夫熊猫3   99832.53         7.7  喜剧       95  2016/1/29
##  director      star1 index1      star2 index2
## 1    叶伟信      甄子丹  11385        张晋  4105
## 2    周星驰      邓超   41310        林允  9292
## 3    郭大雷      赵丽颖 181979        张翰 44277
## 4    郑保瑞      郭富城  12227        巩俐  8546
## 5      王晶      周润发  16731        刘德华 30277
## 6  吕寅荣  杰克布莱克    178  安吉丽娜朱莉  1540
```

从csv中读入

有了分隔符，世界一片清朗

```
tes = read.table("电影数据.csv", header = T, sep = ",", fileEncoding = "UTF-8"); head(tes)
```

```
##              name boxoffice doubanscore type duration  showtime
## 1              叶问3   77060.44         6.4  动作      105  2016/3/4
## 2              美人鱼  338583.26         6.9  喜剧       93  2016/2/8
## 3          女汉子真爱公式    6184.45         4.5  喜剧       93  2016/3/18
## 4  西游记之孙悟空三打白骨精 119956.51         5.7  喜剧      120  2016/2/8
## 5              澳门风云三 111693.89         4.0  喜剧      112  2016/2/8
## 6              功夫熊猫3   99832.53         7.7  喜剧       95  2016/1/29
##  director      star1 index1      star2 index2
## 1    叶伟信      甄子丹  11385        张晋  4105
## 2    周星驰      邓超   41310        林允  9292
## 3    郭大雷      赵丽颖 181979        张翰 44277
## 4    郑保瑞      郭富城  12227        巩俐  8546
## 5      王晶      周润发  16731        刘德华 30277
## 6  吕寅荣  杰克布莱克    178  安吉丽娜朱莉  1540
```

没有分隔符，世界一片混沌

```
(tes = read.table("电影数据.csv", header = T, fileEncoding = "UTF-8"))
```

```
##          name boxoffice doubanscore type duration showtime director star1 index1 star2 index2
## 1          叶问3 77060.44, 6.4, 动作, 105, 2016/3/4, 叶伟信, 甄子丹, 11385, 张晋, 4105
## 2          美人鱼 338583.26, 6.9, 喜剧, 93, 2016/2/8, 周星驰, 邓超, 41310, 林允, 9292
## 3          女汉子真爱公式 6184.45, 4.5, 喜剧, 93, 2016/3/18, 郭大雷, 赵丽颖, 181979, 张翰, 44277
## 4  西游记之孙悟空三打白骨精 119956.51, 5.7, 喜剧, 120, 2016/2/8, 郑保瑞, 郭富城, 12227, 巩俐, 8546
## 5          澳门风云三 111693.89, 4.0, 喜剧, 112, 2016/2/8, 王晶, 周润发, 16731, 刘德华, 30277
## 6          功夫熊猫3 99832.53, 7.7, 喜剧, 95, 2016/1/29, 吕寅荣, 杰克布莱克, 178, 安吉丽娜朱莉, 1540
## 7  北京遇上西雅图之不二情书 78341.38, 6.5, 喜剧, 131, 2016/4/29, 薛晓路, 汤唯, 13499, 吴秀波, 77260
## 8          谁的青春不迷茫 17798.89, 6.4, 爱情, 108, 2016/4/22, 姚婷婷, 白敬亭, 14759, 郭姝彤, 755
## 9          睡在我上铺的兄弟 12561.55, 5.0, 爱情, 95, 2016/4/1, 张琦, 陈晓, 13251, 秦岚, 9549
## 10         冰河追凶 4262.14, 5.6, 动作, 102, 2016/4/15, 徐伟, 梁家辉, 6911, 佟大为, 5614
## 11         梦想合伙人 8058.15, 3.8, 剧情, 101, 2016/4/29, 张太维, 姚晨, 7315, 唐嫣, 66756
## 12         我的新野蛮女友 3336.83, 3.4, 喜剧, 107, 2016/4/22, 赵根植, 宋茜, 81163, 车太贤, 1789
## 13         刑警兄弟 3005.96, 5.2, 喜剧, 97, 2016/4/22, 戚家基, 黄宗泽, 9823, 金刚, 4010
## 14         大唐玄奘 3271.44, 5.1, 剧情, 90, 2016/4/29, 霍建起, 黄晓明, 32595, 徐峥, 10318
## 15         我的特工爷爷 32009.37, 5.3, 动作, 99, 2016/4/1, 洪金宝, 洪金宝, 9148, 刘德华, 30277
## 16         火锅英雄 36624.84, 7.3, 犯罪, 95, 2016/4/1, 杨庆, 陈坤, 5763, 白百何, 10585
## 17         百鸟朝凤 8686.14, 8.0, 剧情, 108, 2016/5/6, 吴天明, 陶泽如, 1139, 李岷城, 3290
## 18         夜孔雀 3260.42, 4.7, 爱情, 84, 2016/5/20, 戴思杰, 刘亦菲, 58355, 刘烨, 11248
## 19         钢刀 924.86, 4.3, 动作, 94, 2016/5/20, 阿甘, 何润东, 11822, 李学东, 521
```

专用函数read.csv

```
movie_csv = read.csv("电影数据.csv", fileEncoding = "UTF-8"); head(movie_csv)
```

```
##          name boxoffice doubanscore type duration showtime
## 1          叶问3 77060.44          6.4 动作      105 2016/3/4
## 2          美人鱼 338583.26          6.9 喜剧       93 2016/2/8
## 3          女汉子真爱公式 6184.45          4.5 喜剧       93 2016/3/18
## 4  西游记之孙悟空三打白骨精 119956.51          5.7 喜剧      120 2016/2/8
## 5          澳门风云三 111693.89          4.0 喜剧      112 2016/2/8
## 6          功夫熊猫3 99832.53          7.7 喜剧       95 2016/1/29
##  director      star1 index1      star2 index2
## 1  叶伟信      甄子丹 11385      张晋 4105
## 2  周星驰      邓超 41310      林允 9292
## 3  郭大雷      赵丽颖 181979      张翰 44277
## 4  郑保瑞      郭富城 12227      巩俐 8546
## 5    王晶      周润发 16731      刘德华 30277
## 6  吕寅荣 杰克布莱克      178 安吉丽娜朱莉 1540
```

从xls,xlsx中读入数据

```
# install.packages("readxl")
library("readxl")
# 其中col_names参数仍然是为了设定是否把第一行当做变量名
movie_excel = data.frame(read_excel("电影数据.xlsx", col_names = T)); head(movie_excel)
```

```
##           name boxoffice doubanscore type duration  showtime
## 1          叶问3   77060.44         6.4 动作      105  2016/3/4
## 2          美人鱼 338583.26         6.9 喜剧       93  2016/2/8
## 3      女汉子真爱公式   6184.45         4.5 喜剧       93  2016/3/18
## 4 西游记之孙悟空三打白骨精 119956.51         5.7 喜剧      120  2016/2/8
## 5          澳门风云三 111693.89         4.0 喜剧      112  2016/2/8
## 6          功夫熊猫3   99832.53         7.7 喜剧       95  2016/1/29
##  director    star1 index1      star2 index2
## 1   叶伟信    甄子丹  11385      张晋   4105
## 2   周星驰    邓超   41310      林允   9292
## 3   郭大雷    赵丽颖 181979      张翰  44277
## 4   郑保瑞    郭富城  12227      巩俐   8546
## 5     王晶    周润发  16731      刘德华 30277
## 6   吕寅荣  杰克布莱克    178  安吉丽娜朱莉   1540
```

非结构化数据——文本数据的读入

```
rm(list = ls())
```

1.读入简单文本数据

```
novel = read.csv("novel.csv", fileEncoding = "UTF-8")
head(novel)
```

##	人气排序	小说名称	作者	小说类型	总点击数	会员周点击数	总字数	评论数	评分
## 1	1	一念永恒	耳根	仙侠小说	4383898	10691	1155534	435429	9.8
## 2	2	斗战狂潮	骷髅精灵	仙侠小说	1678379	36587	422116	23159	10.0
## 3	3	天影	萧鼎	仙侠小说	1248708	32019	373763	25253	9.8
## 4	4	不朽凡人	鹅是老五	仙侠小说	2457382	9610	995669	146715	9.9
## 5	5	玄界之门	忘语	仙侠小说	3736897	6709	1784999	238113	9.8
## 6	6	龙王传说	唐家三少	玄幻小说	2968846	3080	1552654	293934	9.8
##	小说性质	写作进程	授权状态	更新时间					
## 1	公众作品	连载中	A级签约	2016/10/23 11:50					
## 2	公众作品	连载中	A级签约	2016/10/22 17:05					
## 3	公众作品	连载中	A级签约	2016/10/23 10:40					
## 4	公众作品	连载中	A级签约	2016/10/22 20:50					
## 5	公众作品	新书上传	A级签约	2016/10/23 10:15					
## 6	公众作品	新书上传	A级签约	2016/10/23 7:00					

内容简介

1
一念成沧海，一念化桑田。一念斩千魔，一念诛万仙。??? 唯我念……永恒??? 这是耳根继《仙逆》《求魔》《我欲封天》后，创作的第四部长篇小说《一念永恒》

2
双月当空，无限可能的英魂世界 \n??? 孤寂黑暗，神秘古怪的嬉命小丑 \n??? 百城联邦，三大帝国，异族横行，魂兽霸幽 \n??? 这是一个英雄辈出的年代，人类卧薪尝胆重掌地球主权，孕育着进军高纬度的野望！ \n??? 重点是……二年级的废柴学长王同学，如何使用嬉命轮盘，撬动整个世界，学妹们，请注意，学长来了！！ \n??? 斗战一群：21222419（两千人战力群） \n??? 斗战二群：12962047 \n??? 骷髅的微信公共号：kuloujingling00 \n??? 新浪微博：骷髅精灵

3
阴阳分天地，五行定乾坤。 \n??? 天穹之下岁月沧桑的中土神州，正是仙道昌盛的时代，亿万生灵欣欣向荣。 \n??? 纵横千万里间，总有人间一幕幕悲欢离合，在恢弘长生的仙道中上演着。 \n??? 有光便有暗，天穹之下光辉之中，仍有沉默的影子悄然前行着…… \n??? 新书上线！精彩万分！请各位书友多多投票支持！另外，大家可以添加微信公众号zhuxianxiaoding（诛仙萧鼎），QQ官方群 176378308 进行交流。

4
在这里，拥有灵根才能修仙，所有凡根注定只是凡人。 \n??? 莫无忌，只有凡根，一介凡人！ \n??? 是就此老去，还是不甘？ ???

5
天降神物！异血附体！ ??? 群仙惊惧！万魔退避！ ??? 一名从东洲大陆走出的少年。 ??? 一具生死相依的红粉骷髅。 ??? 一个立志成为至强者的故事。 ??? 一段叱咤星河，大闹三界的传说。 ??? 忘语新书，已完本《凡人修仙传》《魔天记》。

6
伴随着魂导科技的进步，斗罗大陆上的人类征服了海洋，又发现了两片大陆。魂兽也随着人类魂师的猎杀无度走向灭亡，沉睡无数年的魂兽之王在星斗大森林最后的净土苏醒，它要带领仅存的族人，向人类复仇！ \n??? 唐舞麟立志要成为一名强大的魂师，可当武魂觉醒时，苏醒的，却是…… \n??? 旷世之才，龙王之争，我们的龙王传说，将由此开始。 \n???

2.用readtable读入文本

文本数据普通读法.

微博有8个指标：脱敏的微博名，所在地，性别，粉丝数，关注人数，微博数，创建时间及个人描述。

```
# test = read.table("weibo.txt", sep = "\t") #运行有错误提示的
# read.table只能读完整的横行竖列的数据，当某些字段存在空字段，就无法构成完整的表格，因此需要告诉R某些行并没有该表格每行规定的8个元素。
# 空字段需要fill参数来填满
# test = read.table("weibo.txt", sep = "\t", fill = T)
# 看看最后一行发生了什么：警告信息
# Warning message:
# In scan(file = file, what = what, sep = sep, quote = quote, dec = dec, :
# EOF within quoted string 引号括起来的字符串有问题，文中有很多\n\t的换行制表符，记录都挤在一起了
#test = read.table("weibo.txt", sep = "\t", fill = T)
#test[92, ]
# 设置quote = ""意思是禁止所有引用符号
weibo = read.table("weibo.txt", sep = "\t", fill = T, quote = "", fileEncoding = "UTF-8")
# stringsAsFactors将文本转化为字符，strip.white将字符中的前后空格去掉
weibo = read.table("weibo.txt", sep = "\t", fill = T, quote = "", strip.white = T, stringsAsFactors = F, fileEncoding = "UTF-8")
head(weibo)
```

```
##      V1      V2 V3      V4 V5      V6      V7
## 1 熊粉1 北京 朝阳区 f 371068 183 2862 Wed Nov 24 15:54:44 +0800 2010
## 2 熊粉2 甘肃 白银 m      7 314      2 Fri Oct 07 16:42:39 +0800 2011
## 3 熊粉3 海外 英国 f      686 600 1089 Sun Apr 18 22:12:07 +0800 2010
## 4 熊粉4 上海 杨浦区 f 728201 202      11 Mon Aug 31 15:38:13 +0800 2009
## 5 熊粉5 北京 东城区 f      508 336 536 Mon Mar 22 22:54:18 +0800 2010
## 6 熊粉6      上海 m      577 499 698 Thu May 06 00:42:56 +0800 2010
##                                     V8
## 1 职业占星师，占星专栏作者，心理学、神秘学探索者。
## 2
## 3
## 4      日常更新在微信：znxmyzn 工作联系： 助理余小姐
## 5      愤青 宝妈 白领 普通青年 伪小资
## 6
```

```
# 查看第75至80行数据
```

```
weibo[75:80, ]
```

```
##                                V1          V2 V3
## 75                            熊粉75 北京 东城区 m
## 76                            熊粉76 北京 东城区 m
## 77                            熊粉77 北京 朝阳区 m
## 78 敬请关注【创业人】杂志官方微信，让创业更轻松，资讯更宽广。
## 79                            订阅...
## 80                            熊粉78 北京 海淀区 m
##          V4    V5    V6                                V7
## 75    6288 1630 10853 Sun Sep 27 14:59:32 +0800 2009
## 76      470 489    620 Sun Apr 25 20:45:37 +0800 2010
## 77 322963 1550 12048 Tue Mar 15 15:19:37 +0800 2011
## 78      NA    NA    NA
## 79      NA    NA    NA
## 80    287   384   1055 Sun Feb 06 13:54:42 +0800 2011
##                                V8
## 75                            文化创意产业先行者，战略研究者，企业家之友
## 76                            不管世界末日来不来，都要活好当下每一秒。
## 77 凭借一本杂志 专业品鉴您的创业生活！【创业人】与您分享创业经验。
## 78
## 79
## 80
```

```
# 处理缺失数据
weibo = weibo[substr(weibo$V1, 1, 2) == "熊粉", ]
```

3.用readLines读入文本

read.table对读入的数据格式要求非常高，既要表格齐整，又要注意各种符号。readlines则不同，参数只有一个文件名。

```
# weibo1是个字符向量
weibo1 = readLines("weibo.txt", encoding = "UTF-8")
head(weibo1)
```

```
## [1] "熊粉1 \t北京 朝阳区 \tf \t371068 \t183 \t2862 \tWed Nov 24 15:54:44 +0800 2010 \t职业占
星师，占星专栏作者，心理学、神秘学探索者。"
## [2] "熊粉2 \t甘肃 白银 \tm \t7 \t314 \t2 \tFri Oct 07 16:42:39 +0800 2011 \t"
## [3] "熊粉3 \t海外 英国 \tf \t686 \t600 \t1089 \tSun Apr 18 22:12:07 +0800 2010 \t"
## [4] "熊粉4 \t上海 杨浦区 \tf \t728201 \t202 \t11 \tMon Aug 31 15:38:13 +0800 2009 \t日常更新
在微信：zmxmyzn 工作联系： 助理余小姐"
## [5] "熊粉5 \t北京 东城区 \tf \t508 \t336 \t536 \tMon Mar 22 22:54:18 +0800 2010 \t愤青 宝妈
白领 普通青年 伪小资"
## [6] "熊粉6 \t上海 \tm \t577 \t499 \t698 \tThu May 06 00:42:56 +0800 2010 \t"
```

可以看到文本的真实结构，“”就是制表符。

```
# 使用字符分割函数将weibo1分开
tmp = strsplit(weibo1, " \t")
class(tmp)
```

```
## [1] "list"
```

```
# 查出每个list的元素长度，来查看异常值
l1 = sapply(tmp, length)
# 长度为0和1是异常行，长度为7和8的可以采用
table(l1)
```

```
## l1
##  0  1  7  8
##  2  3 26 74
```

列表显示原数据有的长度为0，即空行，有的长度为1，即对应一个空格和一个异常字符；长度为7的行缺少最后一列描述，长度为8的是正常行。因此在长度为7的行后补一个空格。

```
# 检验是否均在最后一行缺失，hi即储存所有长度为7的行的最后一个元素
hi = c()
for(i in 1:26)
{
  show(i)
  hi[i] = tmp[l1 == 7][[i]][7]
}
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
## [1] 9
## [1] 10
## [1] 11
## [1] 12
## [1] 13
## [1] 14
## [1] 15
## [1] 16
## [1] 17
## [1] 18
## [1] 19
## [1] 20
## [1] 21
## [1] 22
## [1] 23
## [1] 24
## [1] 25
## [1] 26
```

```
tmp[l1 == 7][1]
```



```
## [[1]]
## [1] "熊粉2"           "甘肃 白银"
## [3] "m"              "7"
## [5] "314"            "2"
## [7] "Fri Oct 07 16:42:39 +0800 2011"
```

```
# 对含有7个字符补充一个空字符,使得最后选择数据框较完整
tmp[l1 == 7] = lapply(tmp[l1 == 7], function(x) c(x, ""))
tmp[l1 == 7][1]
```

```
## [[1]]
## [1] "熊粉2"           "甘肃 白银"
## [3] "m"              "7"
## [5] "314"            "2"
## [7] "Fri Oct 07 16:42:39 +0800 2011" ""
```

```
# 将原来含有7个字符和8个字符的合在一起
infoDf = as.data.frame(do.call(rbind, tmp[l1 == 7 | l1 == 8]), stringsAsFactors = F)
colnames(infoDf) = c("name", "location", "gender", "Nfollowers",
                    "Nfollow", "Nweibo", "createTime", "description")
head(infoDf)
```

```
##      name      location gender Nfollowers Nfollow Nweibo
## 1 熊粉1 北京 朝阳区      f      371068      183    2862
## 2 熊粉2 甘肃 白银      m           7      314         2
## 3 熊粉3 海外 英国      f        686      600    1089
## 4 熊粉4 上海 杨浦区      f      728201      202         11
## 5 熊粉5 北京 东城区      f        508      336     536
## 6 熊粉6      上海      m        577      499     698
##                                createTime
## 1 Wed Nov 24 15:54:44 +0800 2010
## 2 Fri Oct 07 16:42:39 +0800 2011
## 3 Sun Apr 18 22:12:07 +0800 2010
## 4 Mon Aug 31 15:38:13 +0800 2009
## 5 Mon Mar 22 22:54:18 +0800 2010
## 6 Thu May 06 00:42:56 +0800 2010
##                                description
## 1 职业占星师，占星专栏作者，心理学、神秘学探索者。
## 2
## 3
## 4 日常更新在微信：znxmyzn 工作联系：助理余小姐
## 5 愤青 宝妈 白领 普通青年 伪小资
## 6
```

4.readLines的其他用法

```
yitian = readLines("倚天屠龙记.Txt", encoding = "UTF-8")
yitian[1:10]
```

```
## [1] "          一      天涯思君不可忘"
## [2] "      “春游浩荡，是年年寒食，梨花时节。白锦无纹香烂漫，玉树琼苞堆雪。静夜沉沉，”
## [3] “浮光霭霭，冷浸溶溶月。人间天上，烂银霞照透彻。浑似姑射真人，天姿灵秀，意气殊高”
## [4] “洁。万蕊参差谁信道，不与群芳同列。浩气清英，仙才卓犖，下土难分别。瑶台归去，洞”
## [5] “天方看清绝。””
## [6] “      作这一首《无俗念》词的，乃南宋末年一位武学名家，有道之士。此人姓丘，名处机”
## [7] “，道号长春子，名列全真七子之一，是全真教中出类拔萃的人物。《词品》评论此词道：”
## [8] ““长春，世之所谓仙人也，而词之清拔如此”。这首词诵的似是梨花，其实词中真意却是”
## [9] “赞誉一位身穿白衣的美貌少女，说她“浑似姑射真人，天姿灵秀，意气殊高洁”，又说她”
## [10] ““浩气清英，仙才卓犖”，“不与群芳同列”。词中所颂这美女，乃古墓派传人小龙女。”
```

```
# 分段
# 在每个字符中找至少有一个空格的标号
para_head = grep("\\s+", yitian, perl = T)
para_head[1:10]
```

```
## [1] 1 2 6 14 20 22 28 55 61 81
```

```
# 首先构造一个矩阵，第一列是一段之首的序号，第二列是一段之尾的序号
cut_para1 = cbind(para_head[1:(length(para_head) - 1)], para_head[-1] - 1)
head(cut_para1)
```

```
##      [,1] [,2]
## [1,]    1    1
## [2,]    2    5
## [3,]    6   13
## [4,]   14   19
## [5,]   20   21
## [6,]   22   27
```

```
# 编写一个函数，将属于一段的文字粘贴起来
yitian_para = sapply(1:nrow(cut_para1), function(i) paste(yitian[cut_para1[i, 1]:cut_para1[i,
2]], collapse = ""))
yitian_para[1:4]
```

```
## [1] "          一      天涯思君不可忘"
## [2] "      “春游浩荡，是年年寒食，梨花时节。白锦无纹香烂漫，玉树琼苞堆雪。静夜沉沉，浮光霭
## [3] “浮光霭霭，冷浸溶溶月。人间天上，烂银霞照透彻。浑似姑射真人，天姿灵秀，意气殊高洁。万蕊参差谁信道，不
## [4] “洁。万蕊参差谁信道，不与群芳同列。浩气清英，仙才卓犖，下土难分别。瑶台归去，洞天方看清绝。””
## [5] “      作这一首《无俗念》词的，乃南宋末年一位武学名家，有道之士。此人姓丘，名处机，道号长
## [6] “春子，名列全真七子之一，是全真教中出类拔萃的人物。《词品》评论此词道：“长春，世之所谓仙人也，
## [7] “而词之清拔如此”。这首词诵的似是梨花，其实词中真意却是赞誉一位身穿白衣的美貌少女，说她“浑似姑
## [8] “射真人，天姿灵秀，意气殊高洁”，又说她“浩气清英，仙才卓犖”，“不与群芳同列”。词中所颂这美
## [9] “女，乃古墓派传人小龙女。她一生爱穿白衣，当真如风拂玉树，雪裹琼苞，兼之生性清冷，实当得起“冷浸
## [10] “溶溶月”的形容，以“无俗念”三字赠之，可说十分贴切。长春子丘处机和她在终南山上比邻而居，当年一
## [11] “见，便写下这首词来。”
## [12] “      这时丘处机逝世已久，小龙女也已嫁与神雕大侠杨过为妻。在河南少室山山道之上，却另有一个
## [13] “少女，正在低低念诵此词。这少女十八九岁年纪，身穿淡黄衣衫，骑着一头青驴，正沿山道缓缓而上，心
## [14] “中默想：“也只有龙姊姊这样的人物，才配得上他。”这一个“他”字，指的自然是神雕大侠杨过了。她也
## [15] “不拉缰绳，任由那青驴信步而行，一路上山。过了良久，她又低声吟道：“欢乐趣，离别苦，就中更有痴儿
## [16] “女。君应有语，渺万里层云，千山暮雪，只影向谁去？””
```

参考文献

朱雪宁，《R语言从数据思维到数据实战》，中国人民大学出版社，2019