

当我们在R中读入了一个txt\csv等格式的数据时，即自成了一个数据框

数据框就是一个表格，它的每一列可以存放不同类型的数据，因此，它最接近于实际数据的形态

一、创建数据框

```
1 > birthday<-c(1997,1998,1971,1999,2000,2001,2003)
2 > gender<-c("男","男","女","男","女","女","男")
3 > people<-data.frame(name,birthday,gender);people
4   name birthday gender
5 1  张三      1997     男
6 2  李四      1998     男
7 3  王大妈    1971     女
8 4   小二      1999     男
9 5   小五      2000     女
10 6   小六      2001     女
11 7   小七      2003     男
```

二、筛选、引用

```
1 > people[3,1]
2 [1] 王大妈
3 Levels: 李四 王大妈 小二 小六 小七 小五 张三
4 > people[3,1]#查看第3行第1列的人物信息
5 [1] 王大妈
6 Levels: 李四 王大妈 小二 小六 小七 小五 张三
7 > people[3,] #查看第3行的人物信息
8   name birthday gender
9 3 王大妈      1971     女
10 > people$birthday
11 [1] 1997 1998 1971 1999 2000 2001 2003
12 > people[,2]#查看第2列人物的出生年份
13 [1] 1997 1998 1971 1999 2000 2001 2003
14 > new_shuju<-people[people$birthday==2000,];new_shuju
15   name birthday gender
16 5 小五      2000     女
17 > #选择出生年份为2000年的数据
18
19 > new_shuju<-people[people$gender=="男"&people$birthday>1998,]
```

```

20 > new_shuju
21   name birthday gender
22 4 小二      1999    男
23 7 小七      2003    男

```

三、增列、合并

```

1 #1.增列
2 > nation<-c("汉族","汉族","回族","藏族","汉族","汉族","苗族")
3 > #给数据框添加一列数据：民族
4 > people$nat<-nation
5 > people$nation<-nation
6 > #给数据框的增列命名了2次，因此要删除其中的一列
7 > people[, -5]
8   name birthday gender  nat
9 1 张三      1997    男 汉族
10 2 李四      1998    男 汉族
11 3 王大妈    1971    女 回族
12 4 小二      1999    男 藏族
13 5 小五      2000    女 汉族
14 6 小六      2001    女 汉族
15 7 小七      2003    男 苗族

```

```

1 #2.合并
2 > x1<-people[1:3,];x1
3   name birthday gender  nat nation
4 1 张三      1997    男 汉族  汉族
5 2 李四      1998    男 汉族  汉族
6 3 王大妈    1971    女 回族  回族
7 > x2<-people[4:7,];x2
8   name birthday gender  nat nation
9 4 小二      1999    男 藏族  藏族
10 5 小五      2000    女 汉族  汉族
11 6 小六      2001    女 汉族  汉族
12 7 小七      2003    男 苗族  苗族
13 > people.nat<-merge(x1,x2,by="nat");people.nat
14   nat name.x birthday.x gender.x nation.x name.y birthday.y gender.y nation.y
15 1 汉族 张三      1997    男    汉族 小五      2000    女    汉族

```

16	2	汉族	张三	1997	男	汉族	小六	2001	女	汉族
17	3	汉族	李四	1998	男	汉族	小五	2000	女	汉族
18	4	汉族	李四	1998	男	汉族	小六	2001	女	汉族

```
1 #显示匹配不到的样本，只需要在merge()函数中多传入一个参数all.x=T即可
2 > #all.x=T,将前一个数据框people中nat列所有的值做合并，匹配不到赋值NA
3 > x3<-x2[1:3,];x3
4   name birthday gender  nat nation
5 4 小二      1999     男 藏族   藏族
6 5 小五      2000     女 汉族   汉族
7 6 小六      2001     女 汉族   汉族
8
9 > people.nat<-merge(x1,x3,by="nat",all.x = T);people.nat
10    nat name.x birthday.x gender.x nation.x name.y birthday.y gender.y nation.y
11 1 汉族   张三      1997     男   汉族   小五      2000     女   汉族
12 2 汉族   张三      1997     男   汉族   小六      2001     女   汉族
13 3 汉族   李四      1998     男   汉族   小五      2000     女   汉族
14 4 汉族   李四      1998     男   汉族   小六      2001     女   汉族
15 5 回族 王大妈      1971     女   回族   <NA>      NA      <NA>   <NA>
```