



暨南大學
JINAN UNIVERSITY



多元统计分析及R语言建模

第3章 多元数据直观表示及R使用

王斌会 教授



●内容：

包括条图、箱尾图、星相图、脸谱图、调和曲线图等图形及R语言使用。

●要求：

要求学生了解多元数据的直观表示方法及多变量图形的一些特点，并掌握一些复杂的多元数据的图示技术。

3 多元数据直观表示及R使用 → 3.1 数据直观分析简述

说明与举例

● 说明：

图形有助于对所研究数据的直观了解，如果能把一些多元数据直接绘图显示，便可从图形一目了然看出多元变量之间的关系。

● 举例：

【例 3-1】为了研究全国31个省、市、自治区2007年城镇居民生活消费的分布规律，根据调查资料做区域消费类型划分。

3 多元数据直观表示及R使用 → 3.1 数据直观分析简述

● 数据输入：数据R语言读取

#在mvstats4.xls:d3.1中选取A1:I32，拷贝

```
X=read.table('clipboard',header=T) ; X
```

● 直观分析：图示法

对上述数据直接做条图意义不大，通常需
要对其统计量（如均值、中位数等）做直观分
析。

● 数据输出：

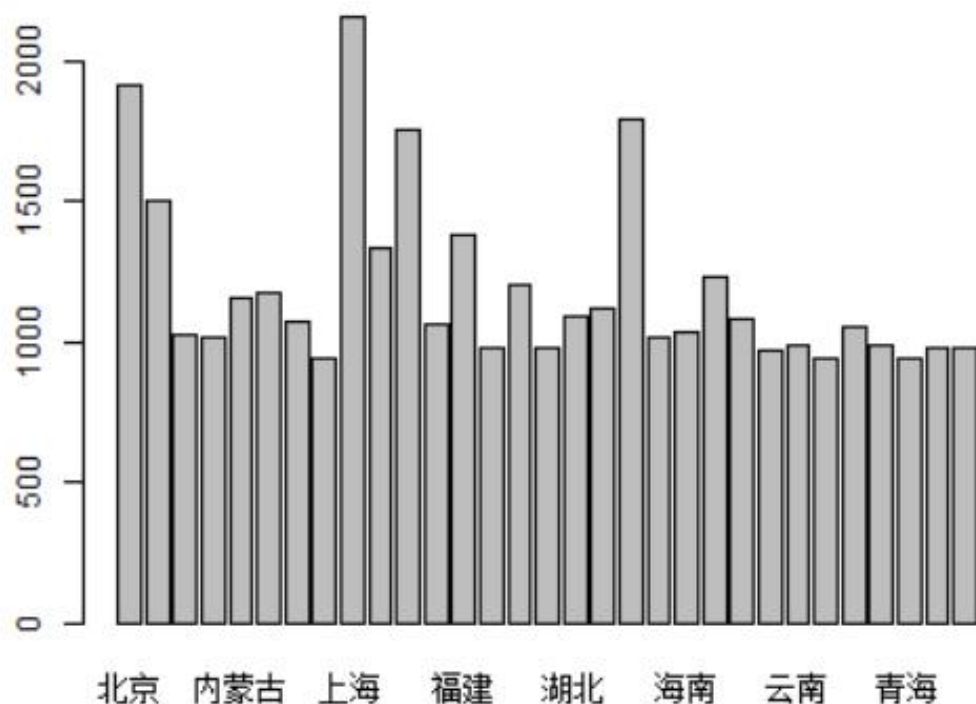
```
R Console
> X=read.table('clipboard',header=T); X
      食品  衣着  设备  医疗  交通  教育  居住  杂项
北京 4934.1 1512.9 981.1 1294.1 2328.5 2384.0 1246.2 649.7
天津 4249.3 1024.2 760.6 1164.0 1309.9 1639.8 1417.5 463.6
河北 2789.9  975.9 546.8  833.5 1010.5  895.1  917.2 266.2
山西 2600.4 1064.6 477.7  640.2 1028.0 1054.1  991.8 245.1
内蒙古 2824.9 1396.9 561.7  719.1 1123.8 1245.1  941.8 468.2
辽宁 3560.2 1017.7 439.3  879.1 1033.4 1052.9 1047.0 400.2
吉林 2842.7 1127.1 407.4  854.8  873.9  997.8 1062.5 394.3
黑龙江 2633.2 1021.5 355.7  729.6  746.0  938.2  784.5 310.7
上海 6125.5 1330.1 959.5  857.1 3153.7 2653.7 1412.1 763.8
江苏 3928.7  990.0 707.3  689.4 1303.0 1699.3 1020.1 377.4
浙江 4892.6 1406.2 666.0  859.1 2473.4 2158.3 1168.1 467.5
安徽 3384.4  906.5 465.7  554.4  891.4 1170.0  850.2 309.3
福建 4296.2  940.7 645.4  502.4 1606.9 1426.3 1261.2 376.0
江西 3192.6  915.1 587.4  385.9  733.0  973.4  728.8 294.6
山东 3180.6 1238.3 661.0  708.6 1333.6 1191.2 1027.6 325.6
河南 2707.4 1053.1 549.1  626.6  858.3  936.6  795.4 300.2
湖北 3456.0 1046.6 550.2  525.3  903.0 1120.3  857.0 242.8
湖南 3243.9 1017.6 603.2  668.5  986.9 1285.2  869.6 315.8
广东 5056.7  814.6 853.2  752.5 2966.1 1994.9 1444.9 454.1
广西 3398.1  656.7 491.0  542.1  932.9 1050.0  803.0 277.4
海南 3546.7  452.9 520.0  503.8 1401.9  837.8  819.0 210.9
重庆 3674.3 1171.2 706.8  749.5 1118.8 1237.4  968.5 264.0
四川 3580.1  949.7 562.0  511.8 1074.9 1031.8  690.3 291.3
贵州 3122.5  910.3 463.6  354.5  895.0 1036.0  718.7 258.2
云南 3562.3  859.7 280.6  631.7 1034.7  705.5  673.1 174.2
西藏 3836.5  880.1 271.3  272.8  866.3  441.0  628.4 335.7
陕西 3063.7  910.3 513.1  678.4  866.8 1230.7  831.3 332.8
甘肃 2824.4  939.9 505.2  564.3  861.5 1058.7  768.3 353.7
青海 2803.5  898.5 484.7  613.2  785.3  953.9  641.9 331.4
宁夏 2760.7  994.5 480.8  646.0  859.0  863.4  910.7 302.2
新疆 2760.7 1183.7 475.2  598.8  890.3  896.8  737.0 331.8
```


3 多元数据直观表示及R使用 → 3.2 均值条图及R使用

● 条图：（1）样品（行）均值条图

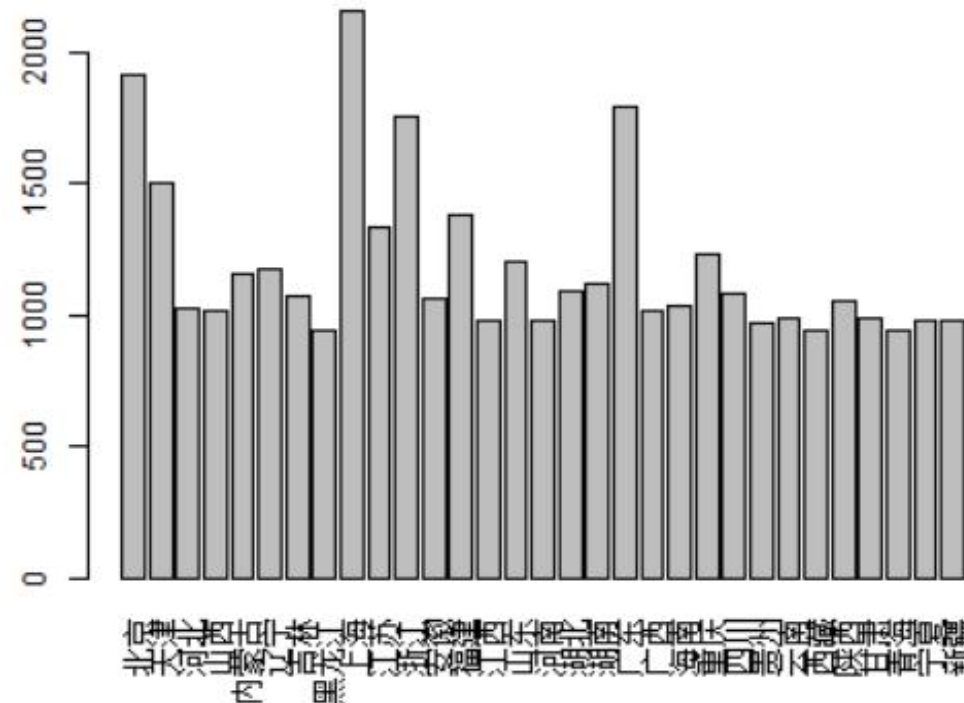
#按行做均值条形图

```
barplot(apply(X,1,mean))
```



#修改横坐标位置

```
barplot(apply(X,1,mean),las=3)
```

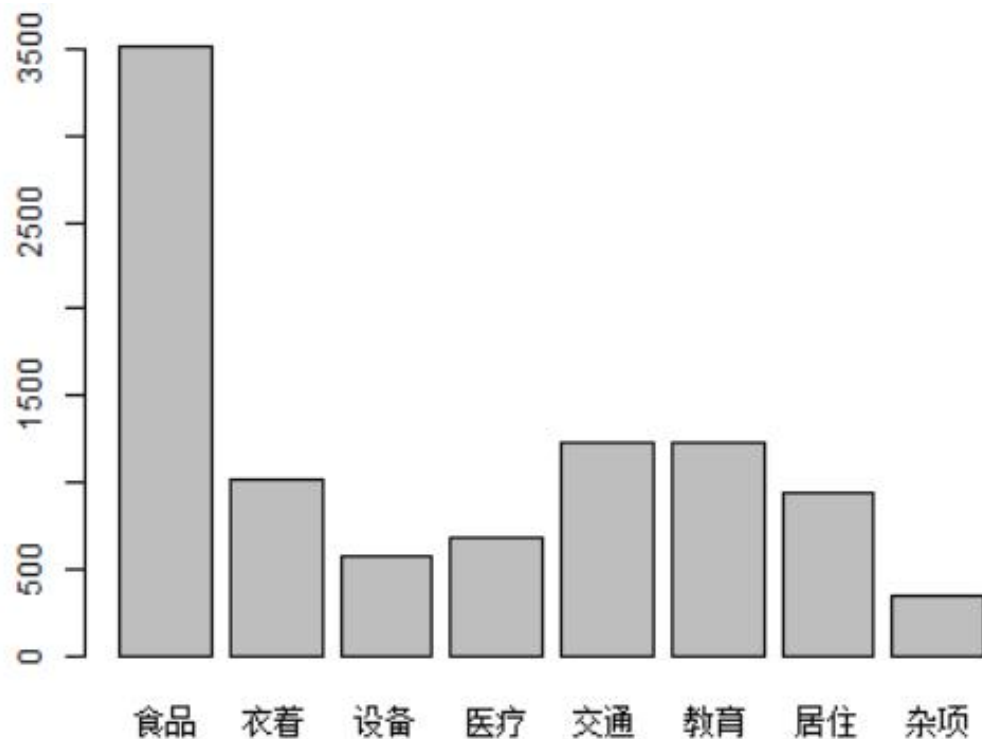


3 多元数据直观表示及R使用 → 3.2 均值条图及R使用

● 条图：（2）变量（列）均值条图

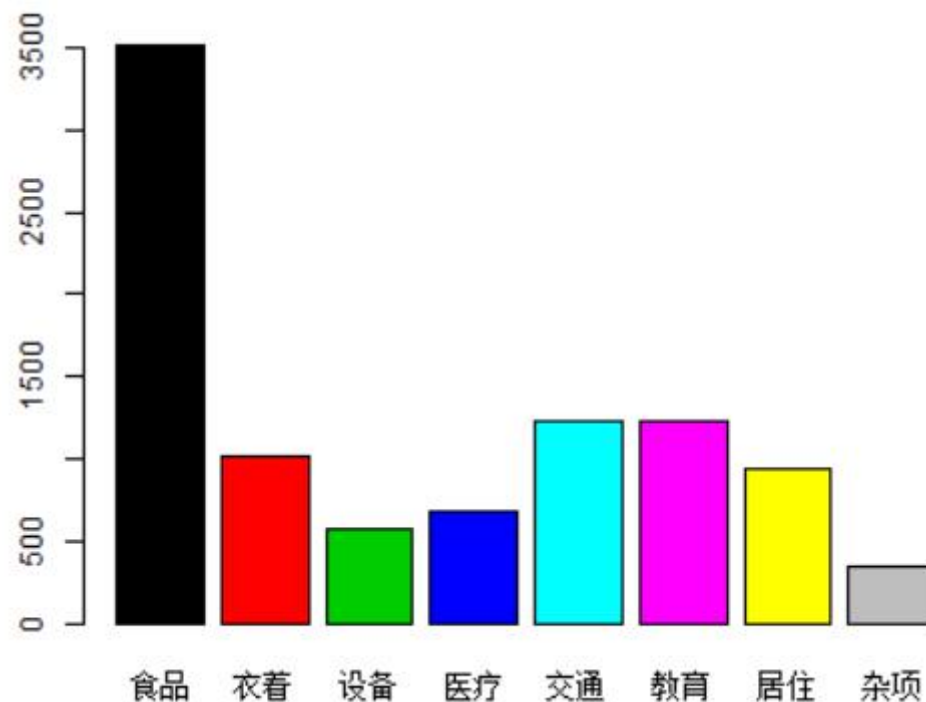
#按列做均值图条形

```
barplot(apply(X,2,mean))
```



#按列做彩色均值图条形图

```
barplot(apply(X,2,mean),col=1:8)
```

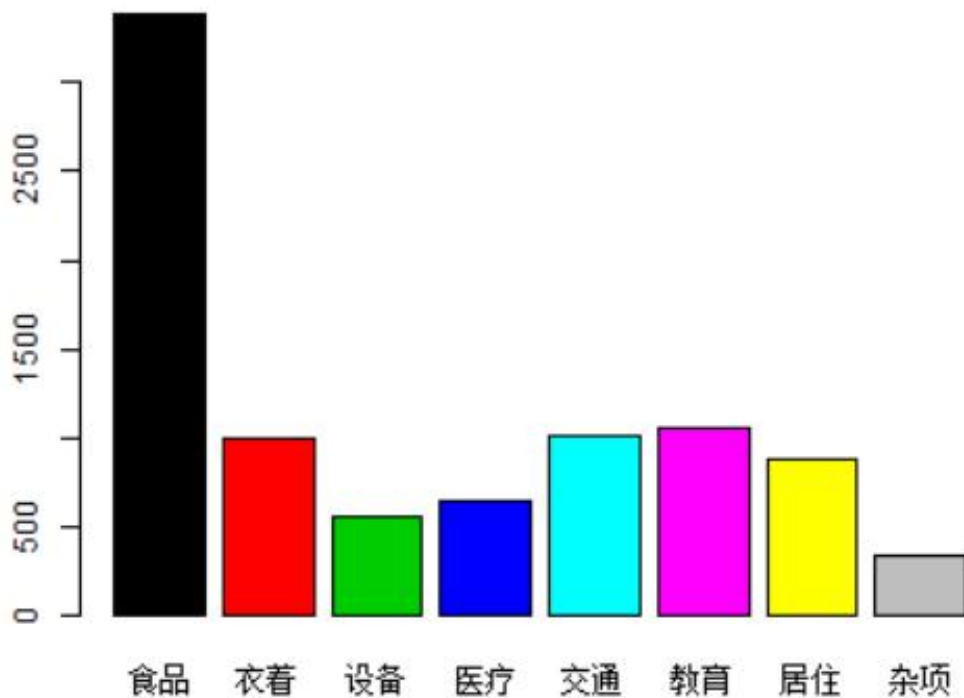


3 多元数据直观表示及R使用 → 3.2 均值条图及R使用

● 条图：(2) 变量(列) 中位数条图

#按列做中位数条形图

```
barplot(apply(X,2,median),col=1:8)
```



● 饼图：(2) 变量(列) 圆图

#按列做均值饼图

```
pie(apply(X,2,mean))
```



3 多元数据直观表示及R使用 → 3.3 箱尾图及R使用

- 箱尾图

- 用途：

Tukey提出的箱尾图由箱子和其上引出的两个尾组成，这种图用来表示在一定时间内一个班成绩的变化、物体位置的变化、原材料的变化、产品标准的变化等。

- 特征：

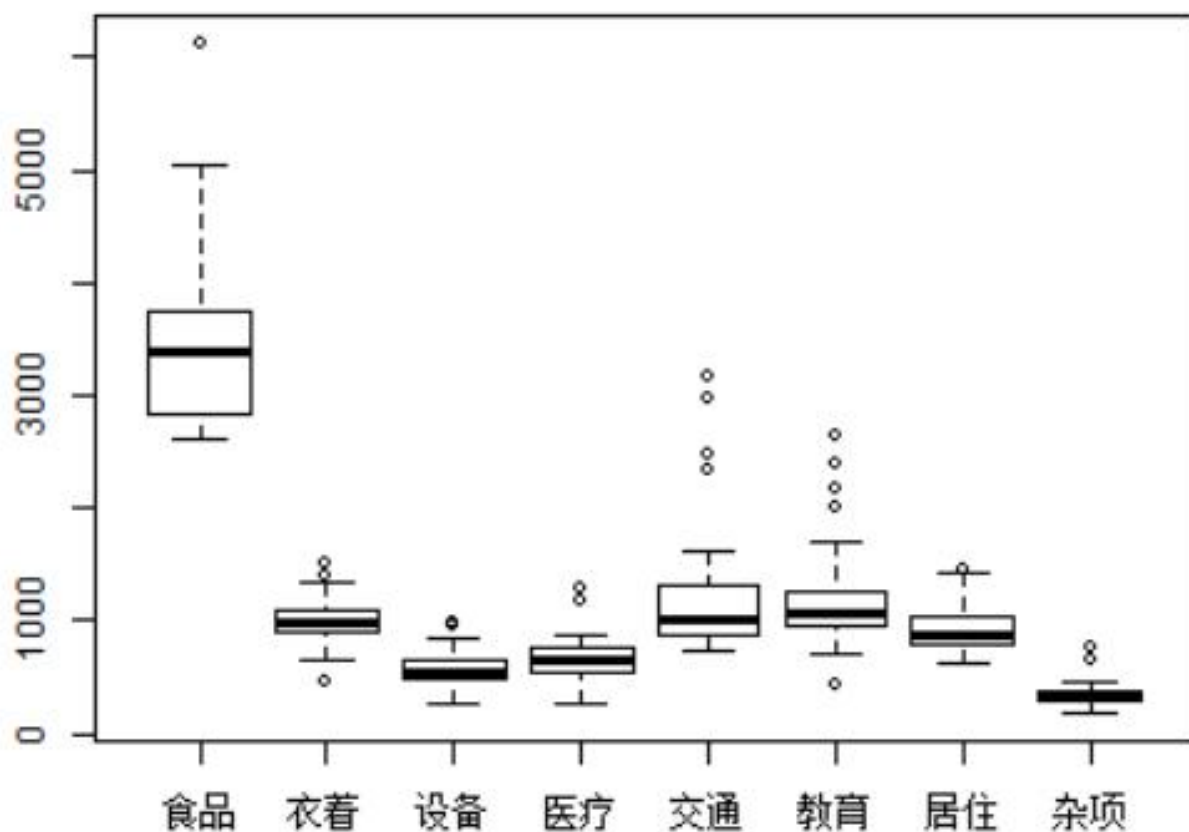
箱尾图可以比较清晰地表示数据的分布特征，它由4部分组成。

- R语言函数：

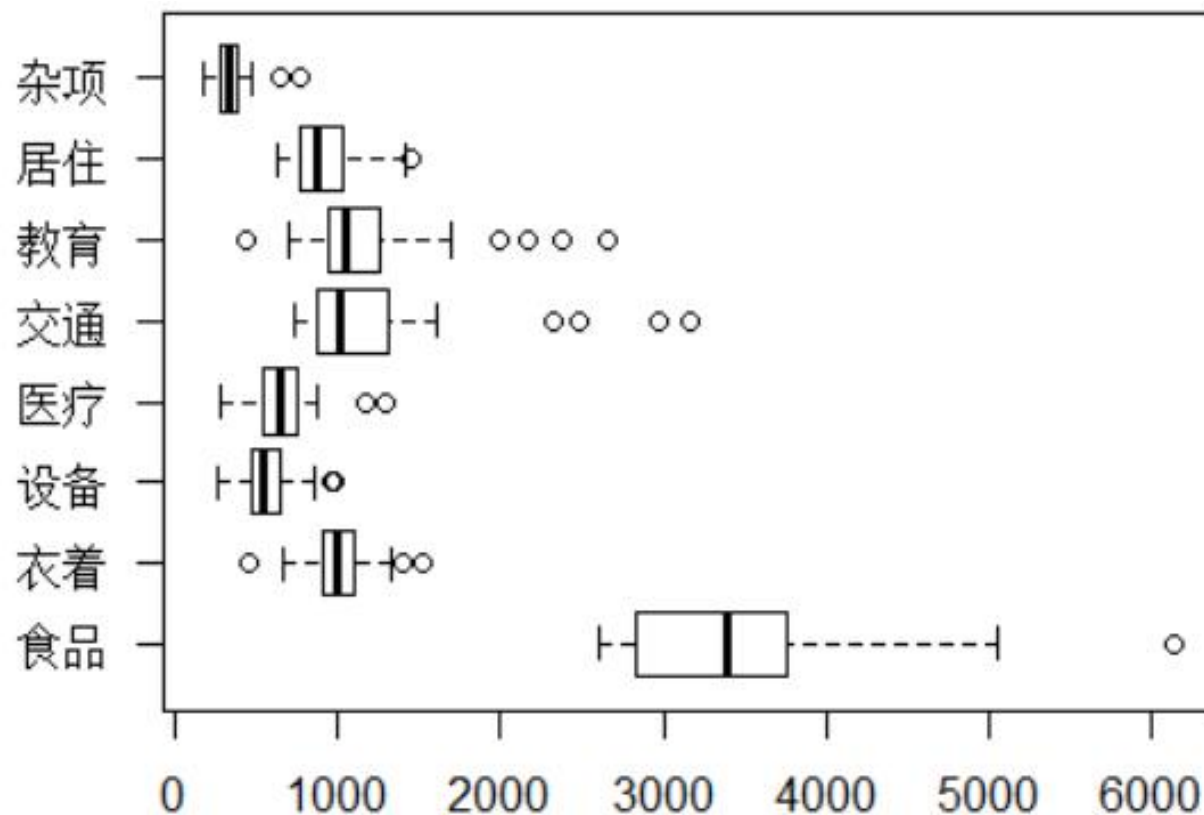
boxplot(X,...)

3 多元数据直观表示及R使用 → 3.3 箱尾图及R使用

`boxplot(X)` #按列做垂直箱线图



`boxplot(X, horizontal=T)` #水平箱线图



3 多元数据直观表示及R使用 → 3.4 星相图及R使用

● 星相图

● 用途：

它将每个变量的各个观察单位的数值表示为一个图形， n 个观察单位就有 n 个图，每个图的每个角表示每个变量。

● 特征：

星相图是雷达图的多元表示形式。

● R语言函数：

`stars(X, draw.segments = FALSE, key.loc = NULL, ...)`

3 多元数据直观表示及R使用 → 3.4 星相图及R使用

#简单星相图

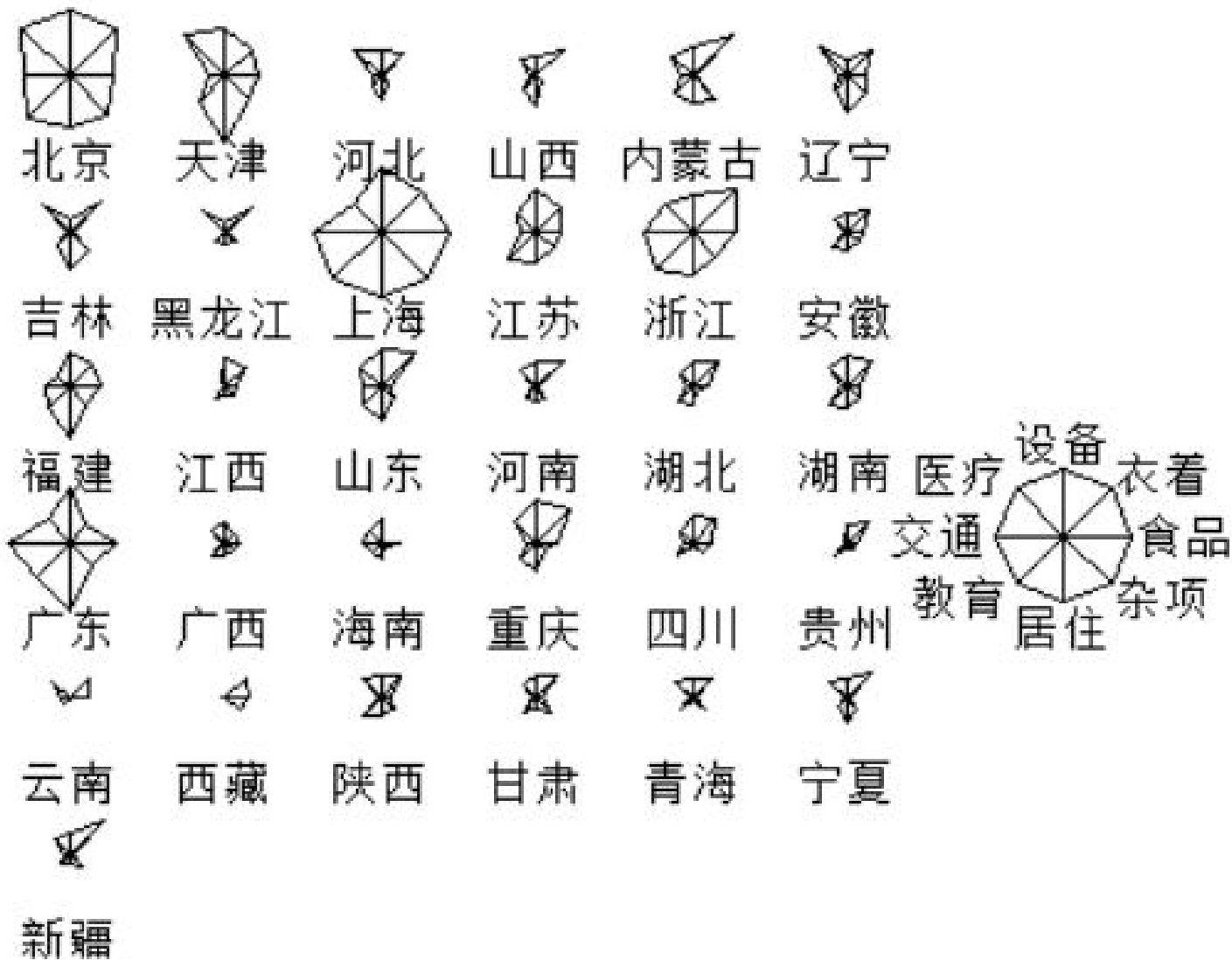
stars(X)



3 多元数据直观表示及R使用 → 3.4 星相图及R使用

#带图例的星相图

```
stars(X,key.loc=c(17,7))
```



3 多元数据直观表示及R使用 → 3.4 星相图及R使用

#带图例度彩色星相图

```
stars(X,key.loc=c(17,7),
      draw.segments=T)
```

