**《数据可视化技术》**

**上机实验指导**

**实验一 数据可视化——数据处理**

**【实验目的与要求】**

1、掌握文本文件数据导入方法

2、掌握网络数据源导入方法

3、掌握数据清洗的方法

4、掌握数据加工方法

5、掌握数据抽样方法

**【实验内容】**

学号 姓名 出生日期 年龄 籍贯 民族 入学年份

2016001 陈明 1997-7-1 20 四川成都 汉族 2016

2016002 李雪 1998-11-2 19 湖北武汉 回族 2015

2016002 李雪 1998-11-2 19 湖北武汉 回族 2015

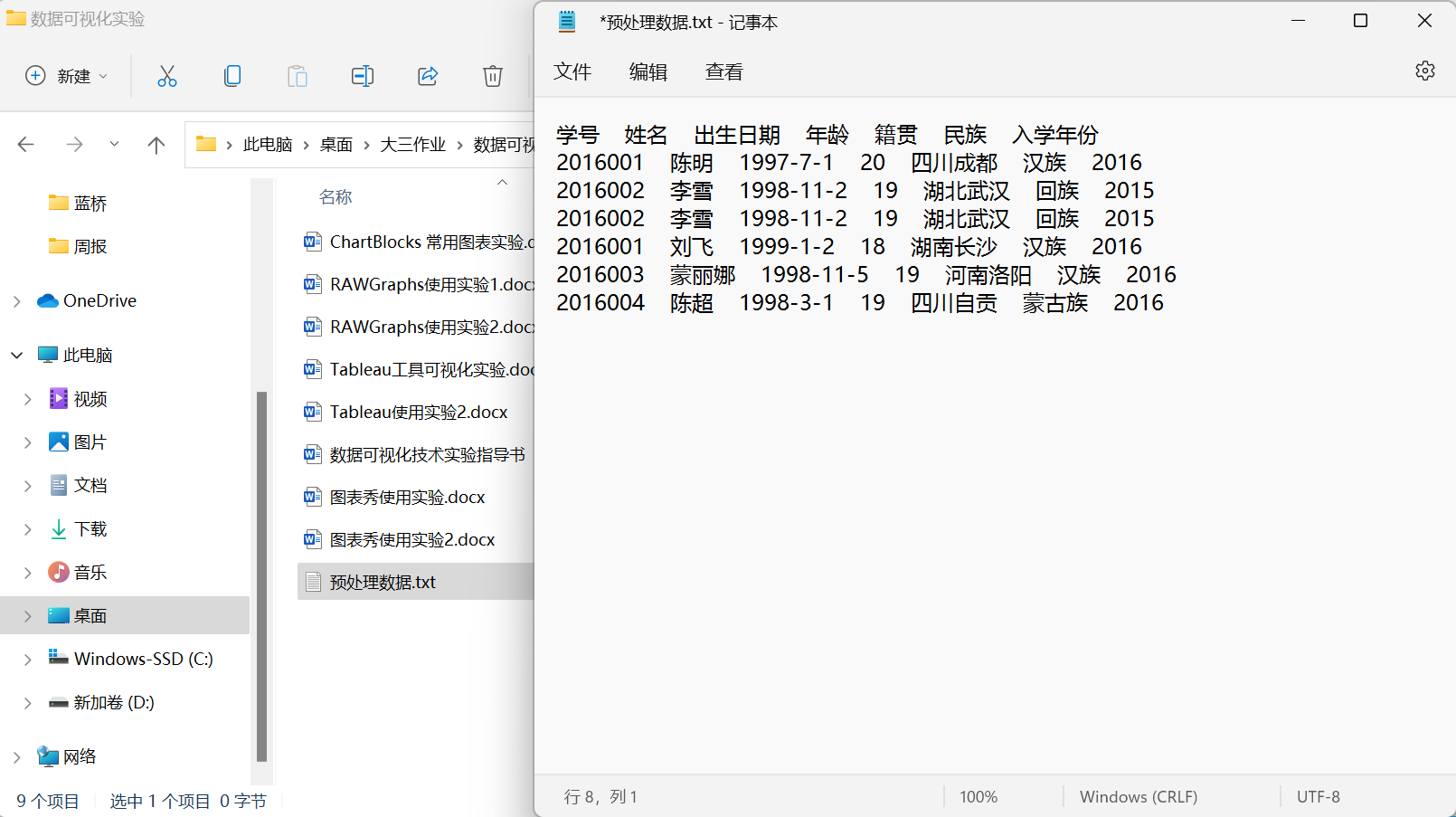
2016001 刘飞 1999-1-2 18 湖南长沙 汉族 2016

2016003 蒙丽娜 1998-11-5 19 河南洛阳 汉族 2016

2016004 陈超 1998-3-1 19 四川自贡 蒙古族 2016

1. 请将以上数据保存为文本文件，并将数据导入到excel表中，命名为学生信息表。

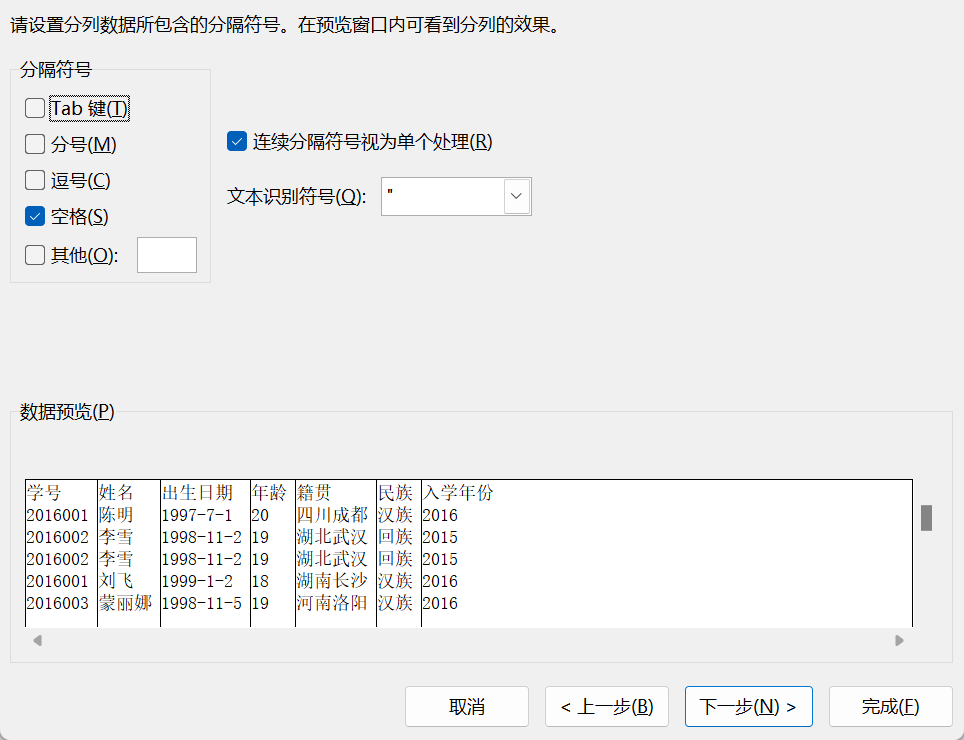
**数据另存为文本文件**



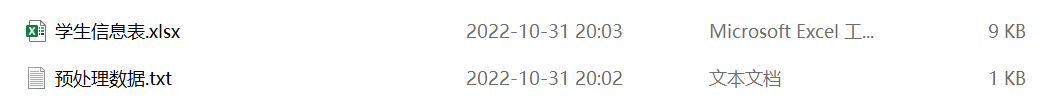
**将文本文件导入到excel**



**使用空格进行分隔**



**另存为学生信息表.xlsx**



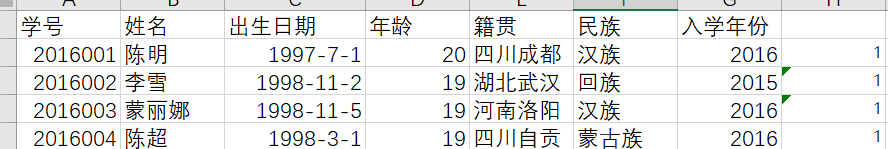
1. 请使用函数法实现对于重复数据的识别和删除。（使用COUNTIF函数）

方法1：





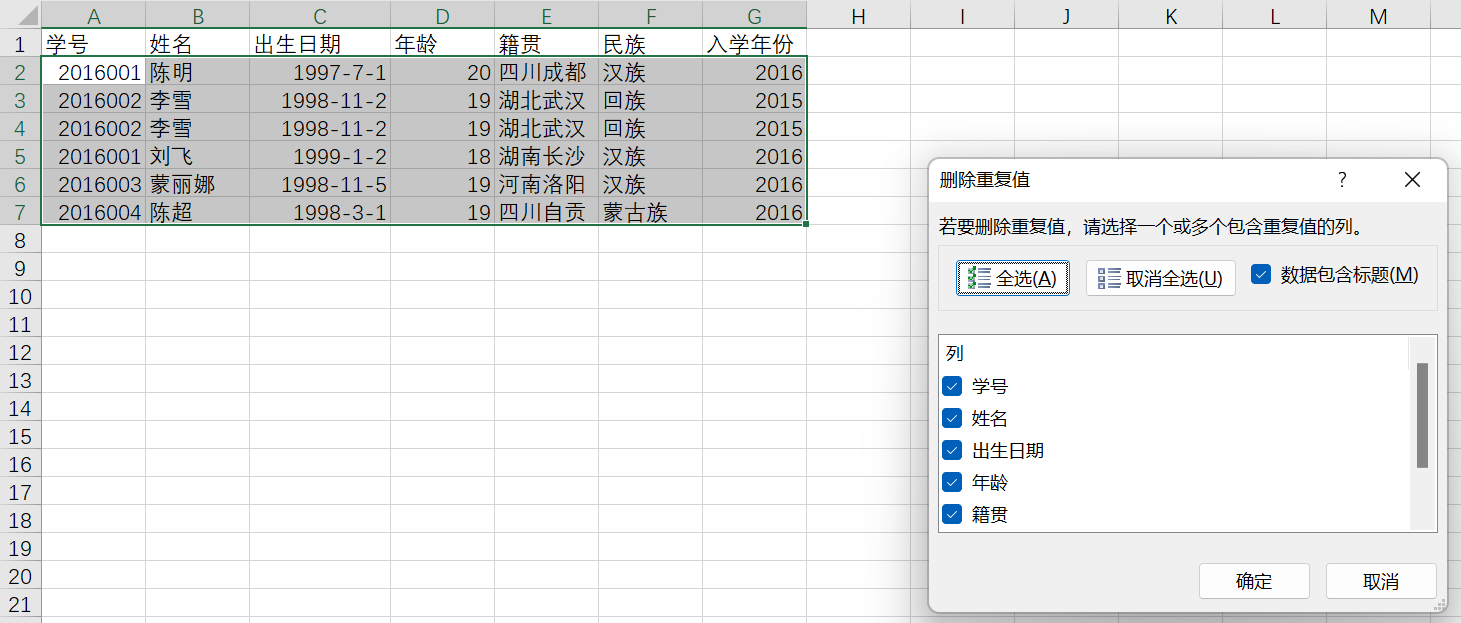
由于学号存在两个一样的，因此手动删除H列为2的两行重复值

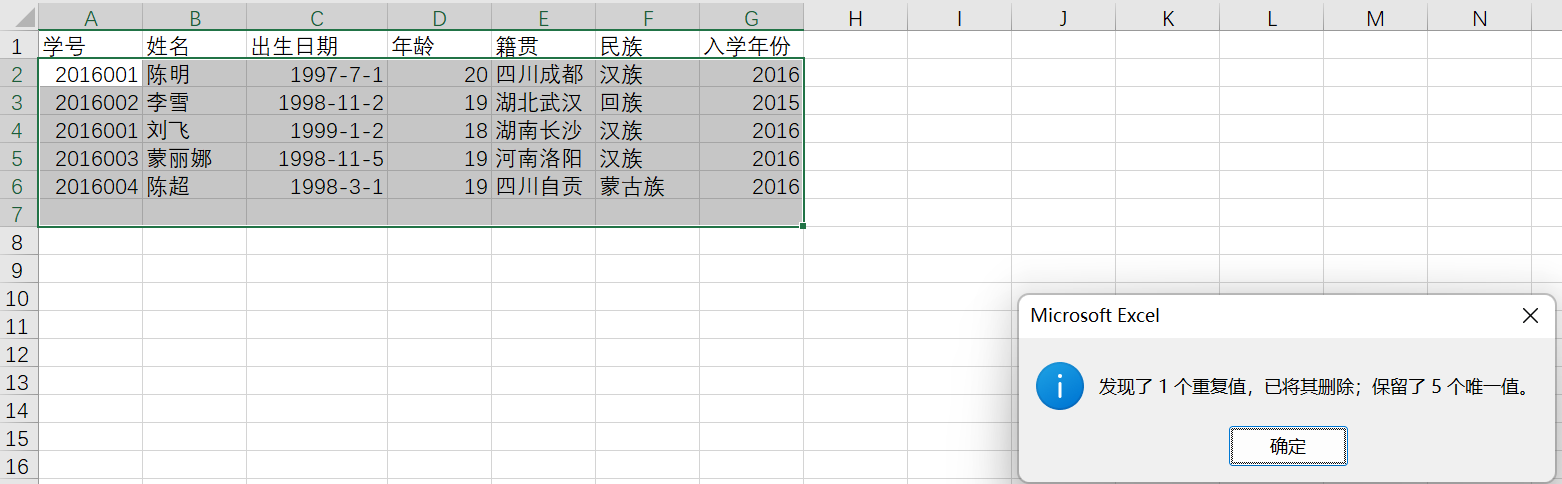


方法2：

（也可以使用excel的数据--删除重复值的功能）

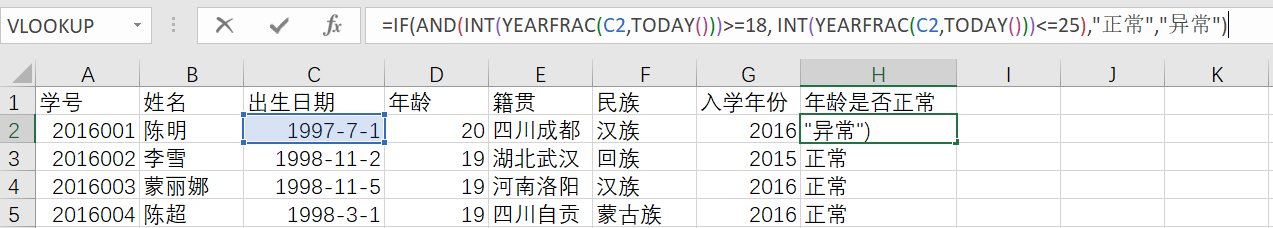






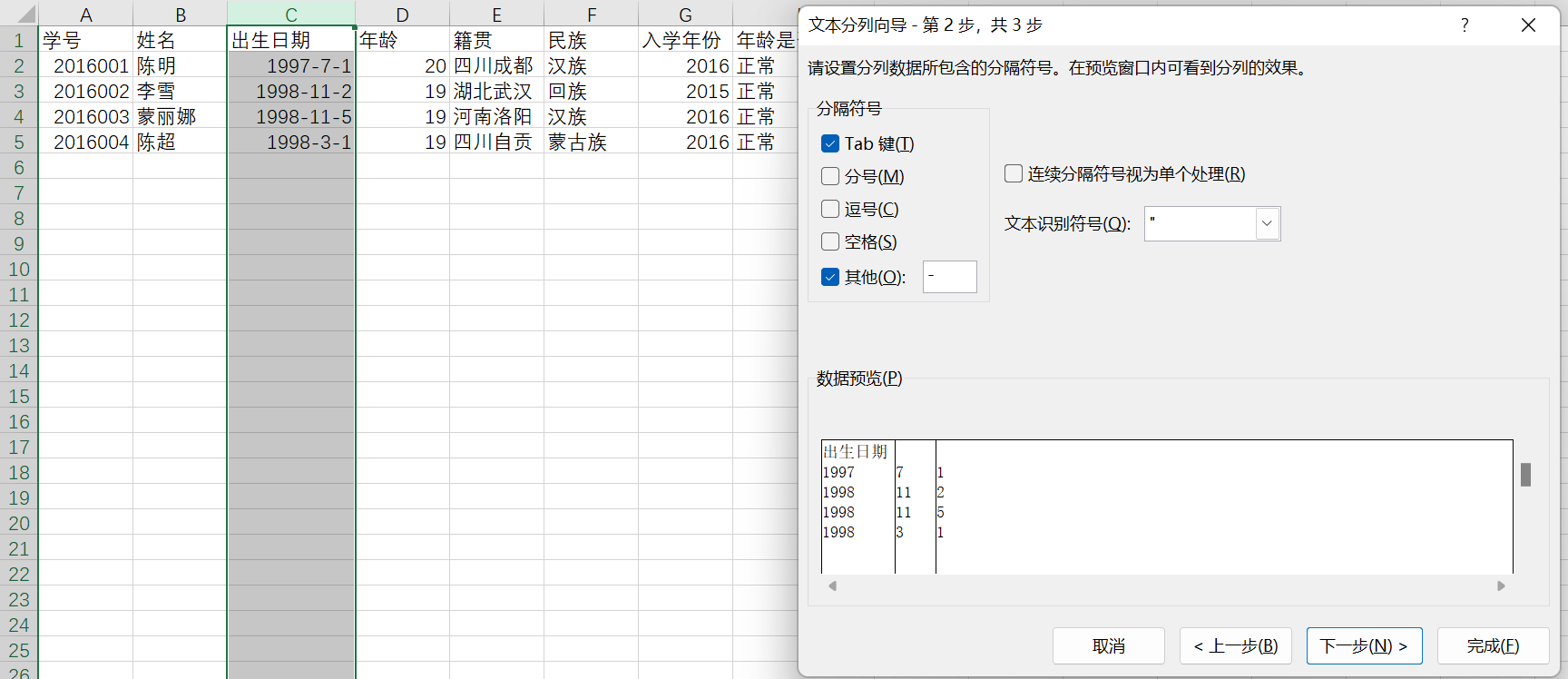
这个方法只删除了一行

1. 请增加一列，该列列名为：判断年龄是否合理，判断学生的年龄，是否在18岁到25岁之间，是，则该列单元显示 正常，否则，显示 异常。



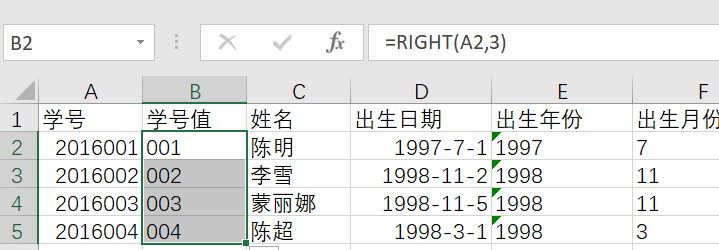
使用TODAY()函数、YEARFRAC函数计算年龄，使用AND进行判断年龄是否在区间内，再使用IF函数输出正常及异常

1. 请使用数据抽取里的菜单法，提取出学生出生年份、月份和日期列信息。





1. 在学号列右侧新增1列，列名为学号值，列的内容取学号列的后3位的值。



1. 将表中的籍贯和民族列合并到 新增的”省份和民族“列中。



7、请使用VLOOKUP函数查找下面这两个学号的学生的籍贯。

2016001

2016004



1. 新增1列，列名为 就读时长（年），使用Datedif函数统计出每个学生的入学时长



以上内容都需要进行截图！

**实验二 Excel数据可视化**

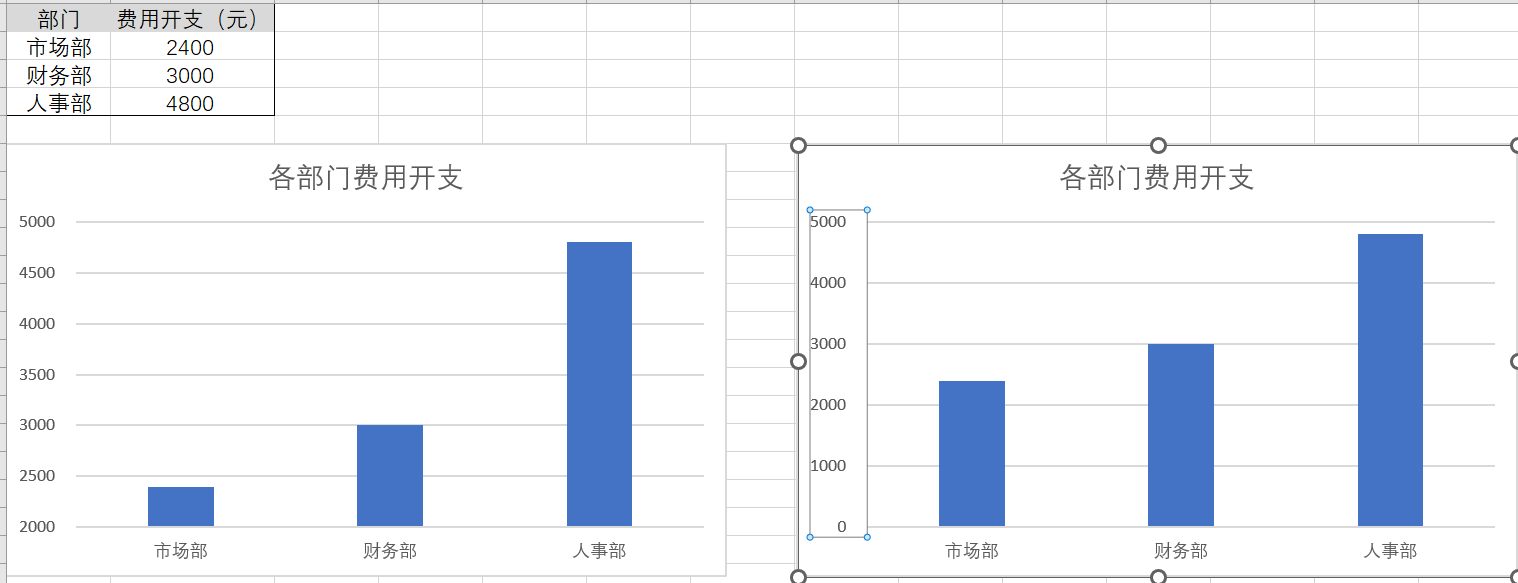
【实验目的】

使用Excel实现数据的可视化（数据可不采用这里的，可以自己设置，截图部分上传到QQ群作业里，至少截图5个）

【实验内容】

1. **用Excel绘制直方图**

* 直方图：又称质量分布图、柱状图，是一种统计报告图，也是表示资料变化情况的主要工具。直方图由一系列高度不等的纵向条纹或线段表示数据分布的情况，一般用横轴表示数据类型，纵轴表示分布情况。
* 实验数据和结果

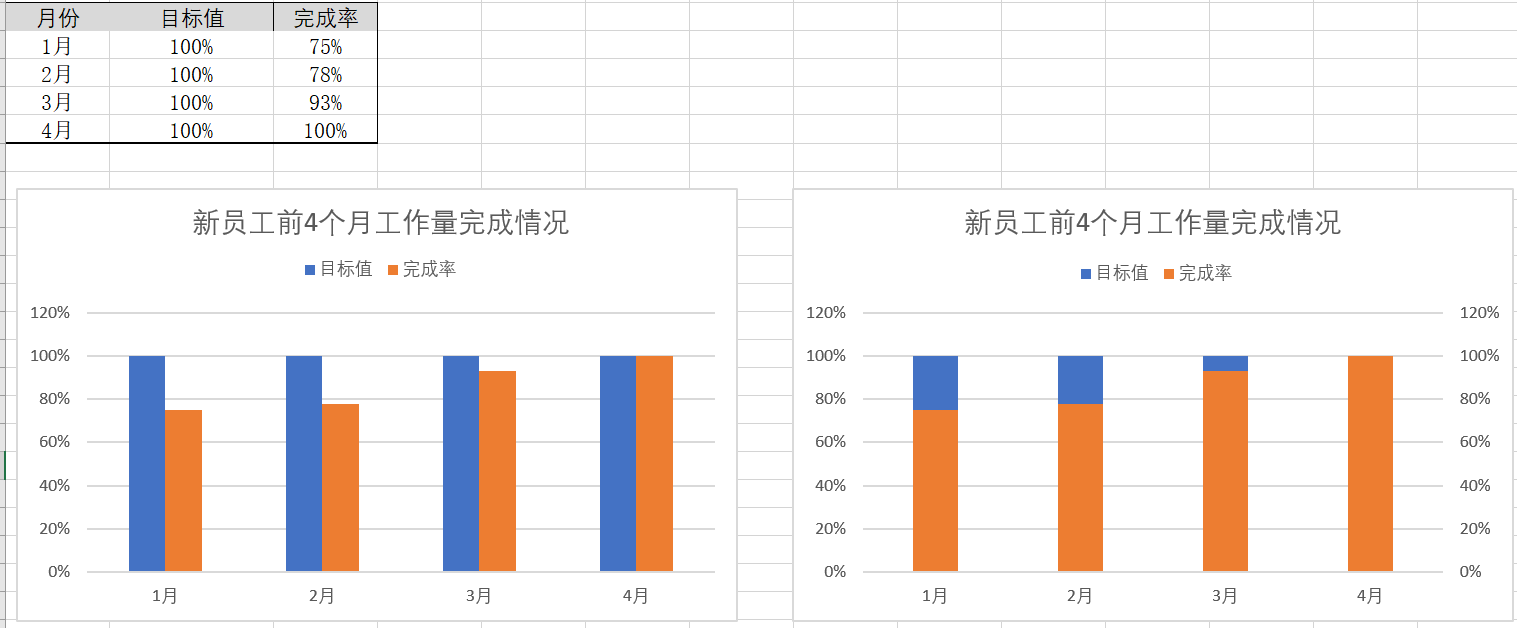


实验设置（office2013或wps均可，wps设置坐标轴格式）：





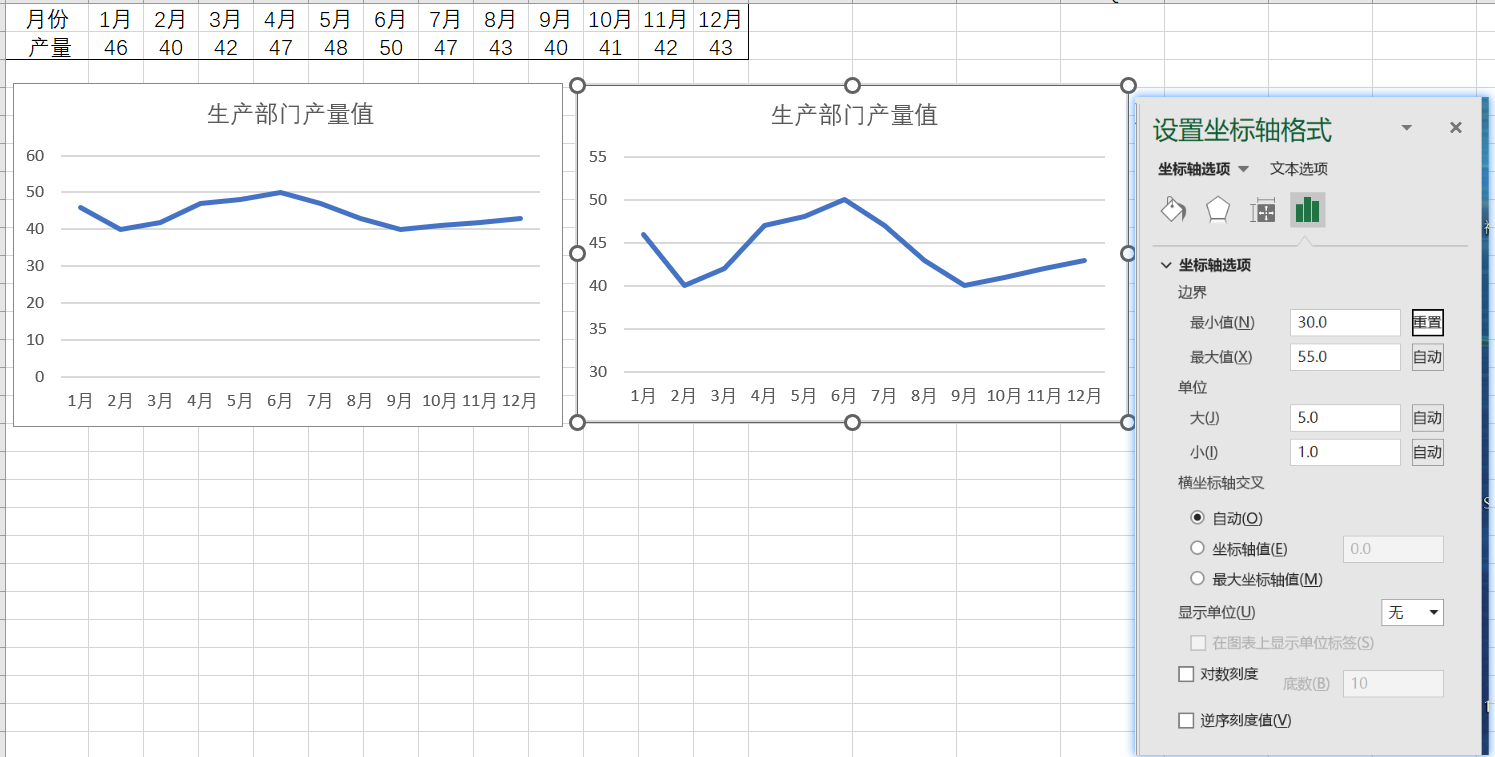
（设置数据系列的间距等）

1. 绘制百分比柱形堆积图

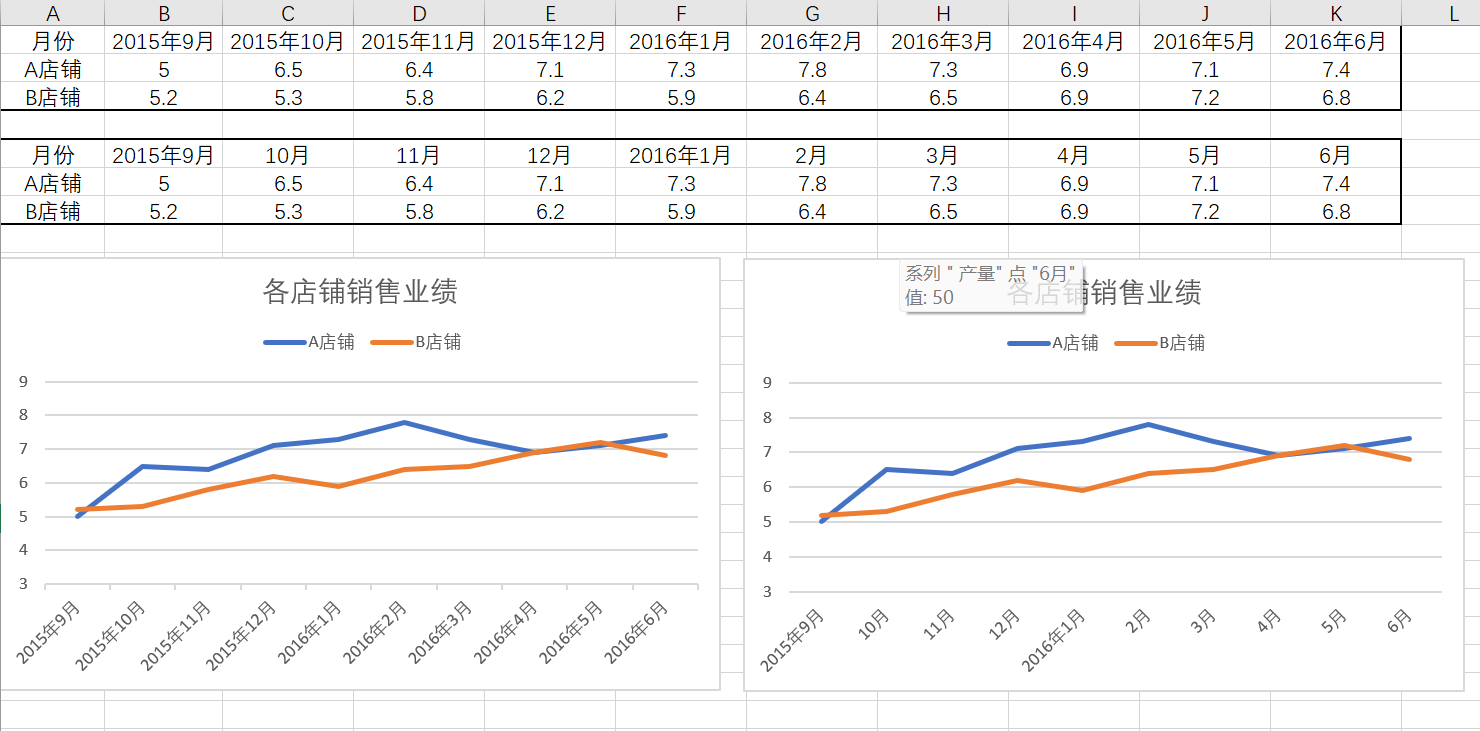
|  |  |  |
| --- | --- | --- |
| 月份 | 目标值 | 完成率 |
| 1月 | 100% | 75% |
| 2月 | 100% | 78% |
| 3月 | 100% | 93% |
| 4月 | 100% | 100% |

（PS：数据系列绘制选择次坐标轴）

1. 折线图

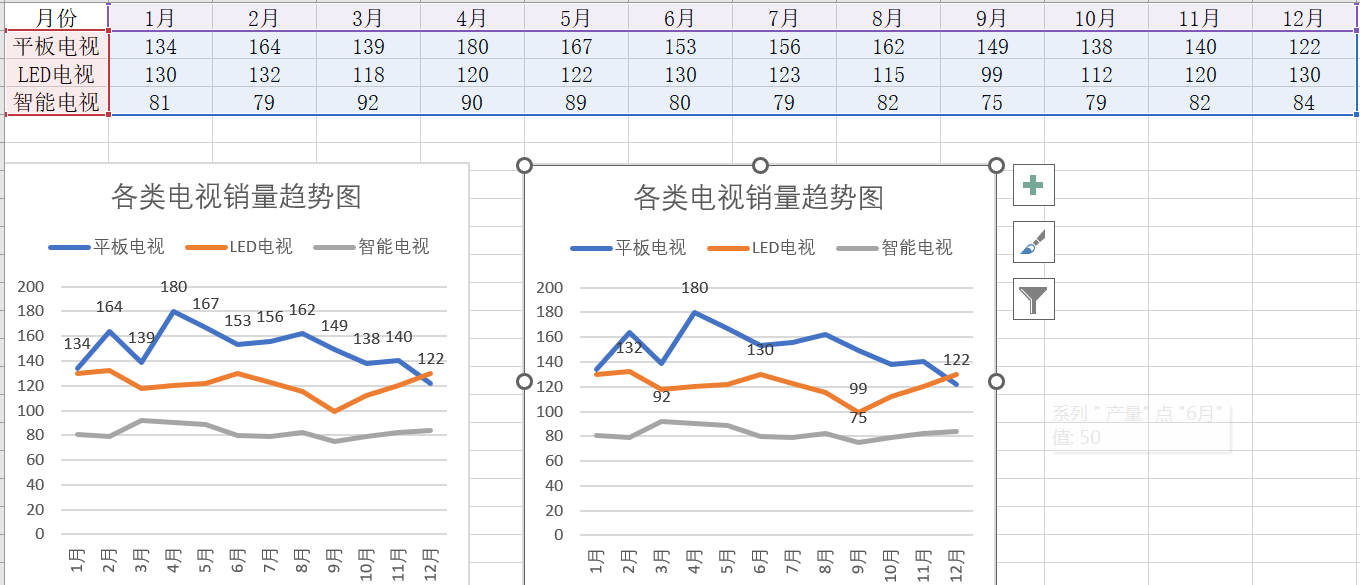
3.1 单折线

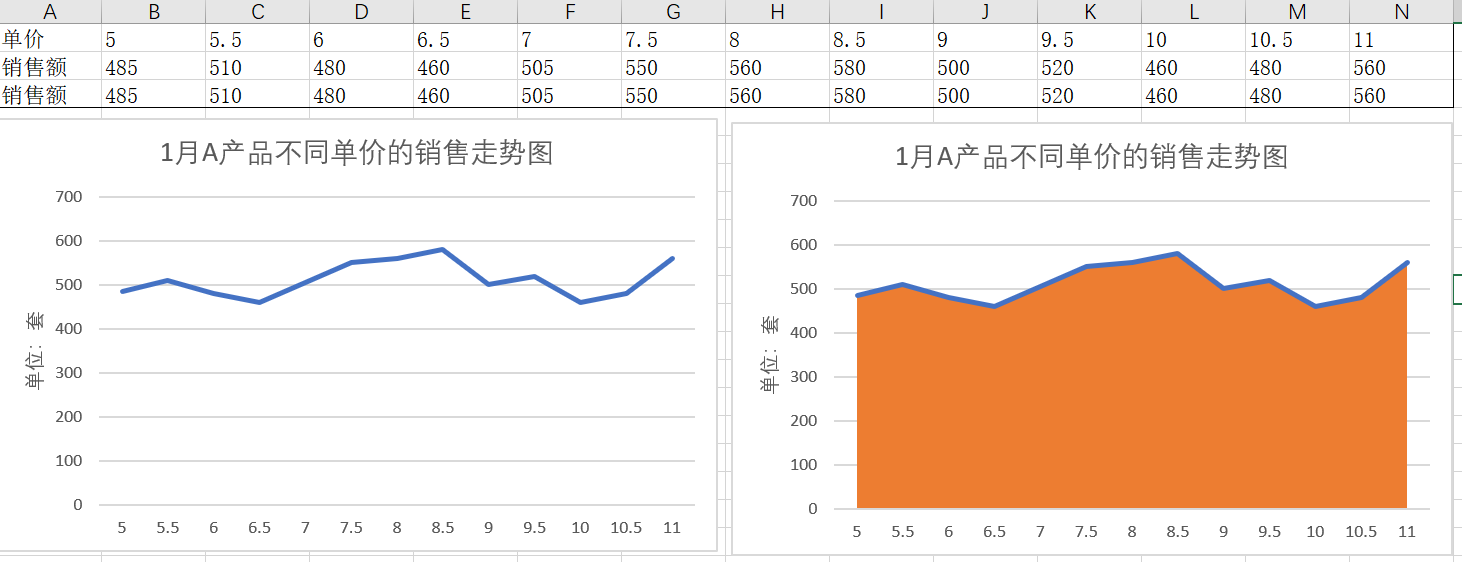
单击Y轴坐标，打开坐标轴格式窗格，在“坐标轴选项”下输入边界最小值“30”，边界最大值“50”，然后输入主要单位值“5”，结果如右图所示。

 3.2 双折线图

**3.3** 三折线并加突出显示折线图中的数据点

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 月份 | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 | 7月 | 8月 | 9月 | 10月 | 11月 | 12月 |
| 平板电视 | 134 | 164 | 139 | 180 | 167 | 153 | 156 | 162 | 149 | 138 | 140 | 122 |
| LED电视 | 130 | 132 | 118 | 120 | 122 | 130 | 123 | 115 | 99 | 112 | 120 | 130 |
| 智能电视 | 81 | 79 | 92 | 90 | 89 | 80 | 79 | 82 | 75 | 79 | 82 | 84 |

（加数据标签） （删除不需要显示的数据标签）

3.4 折线图+面积图

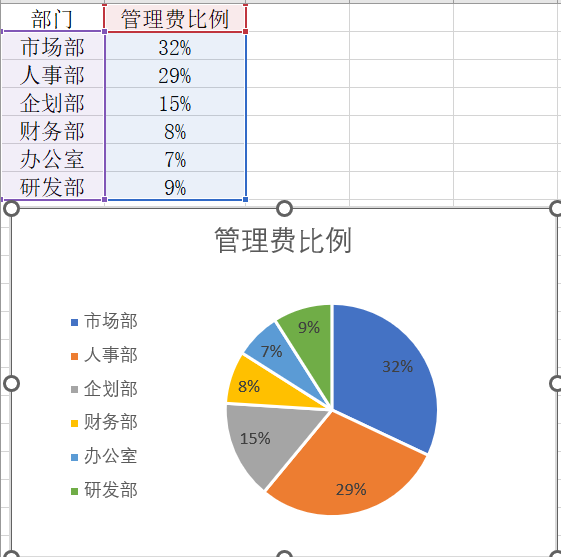
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 单价 | 5 | 5.5 | 6 | 6.5 | 7 | 7.5 | 8 | 8.5 | 9 | 9.5 | 10 | 10.5 | 11 |
| 销售额 | 485 | 510 | 480 | 460 | 505 | 550 | 560 | 580 | 500 | 520 | 460 | 480 | 560 |
| 销售额 | 485 | 510 | 480 | 460 | 505 | 550 | 560 | 580 | 500 | 520 | 460 | 480 | 560 |

（wps选中折线，选择更改图表类型，自定义组合里修改另外一个销售额的曲线为面积图）

1. 圆饼图

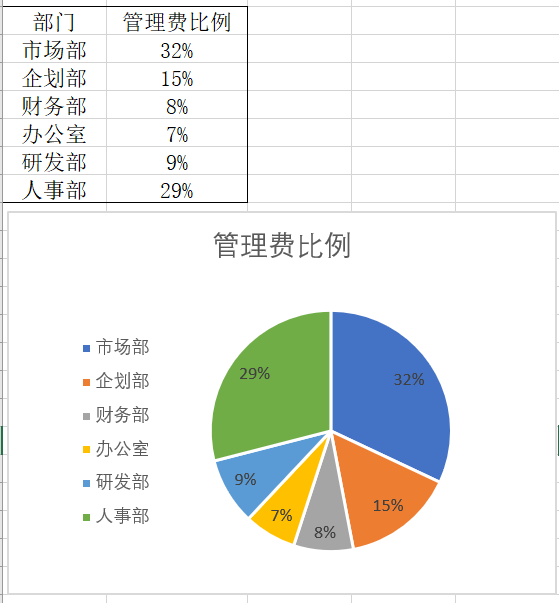
4.1 圆饼图

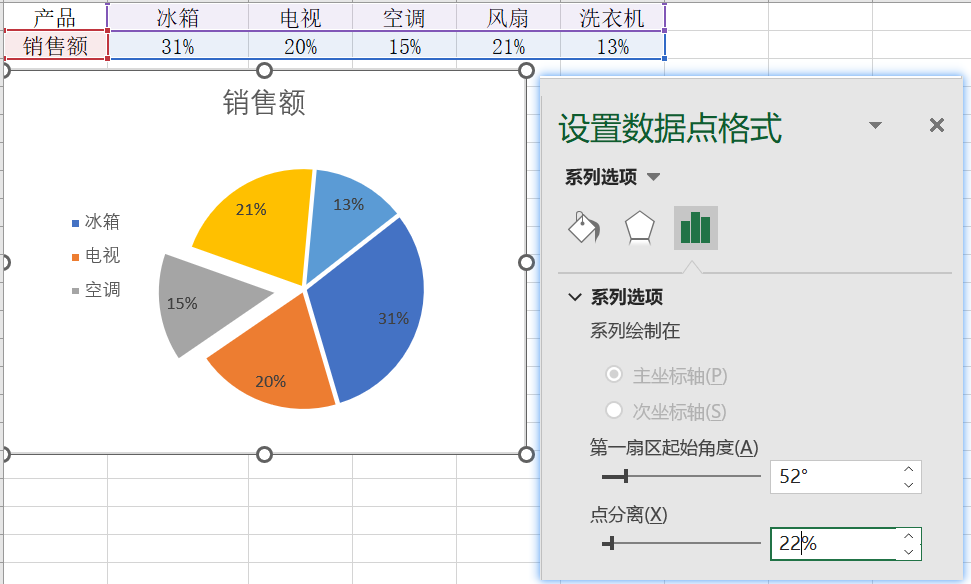
|  |  |
| --- | --- |
| 部门 | 管理费比例 |
| 市场部 | 32% |
| 人事部 | 29% |
| 企划部 | 15% |
| 财务部 | 8% |
| 办公室 | 7% |
| 研发部 | 9% |



一般而言，圆饼图的切片分布一般是将数据较大的两个扇区设置在水平方向的左右两侧。因此，可通过更改数据源的排序顺序改变圆饼图切片的分布位置外，还可以对圆饼图切片进行旋转，使圆饼图的两个较大扇区分布在左右两侧。如

|  |  |
| --- | --- |
| 部门 | 管理费比例 |
| 市场部 | 32% |
| 企划部 | 15% |
| 财务部 | 8% |
| 办公室 | 7% |
| 研发部 | 9% |
| 人事部 | 29% |



**4.2 分离圆饼图**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 产品 | 冰箱 | 电视 | 空调 | 风扇 | 洗衣机 |
| 销售额 | 31% | 20% | 15% | 21% | 13% |

**(wps设置！，excel2013则双击圆饼图打开“设置数据系列格式”窗格，再单击需要被强调的扇区（系列为“空调”），然后在“系列选项”组下设置“点爆炸型”的百分比值为“22%”，即将所选中的扇区单独分离出来。由于分离的扇区显示在图表下方，需要调整“第一扇区起始角度”值“53”来改变扇区位置，使其显示在图表的左边区域）**

1. **散点图**

**5.1 平滑线联系散点图**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 年龄 | 16 | 19 | 21 | 23 | 26 | 28 | 30 | 32 | 33 | 36 | 40 | 44 | 50 | 54 |
| 网购金额 | 410 | 600 | 1600 | 2050 | 2450 | 3200 | 3200 | 3450 | 4050 | 2520 | 1850 | 500 | 450 | 250 |

