

泰坦尼克号乘客数据可视化

概要

泰坦尼克号是一艘奥林匹克级邮轮，于 1912 年 4 月处女航时撞上冰山沉没，泰坦尼克号由位于[爱尔兰岛贝尔法斯特](#)的哈兰德与沃尔夫造船厂兴建，是当时最大的客运轮船。1912年4月14日，船上时间夜里11时40分，泰坦尼克号撞上[冰山](#)；4月15日凌晨2时20分，船体断裂成两截后沉入[大西洋](#)，船上1500多人丧生。泰坦尼克号海难为和平时期死伤人数最惨重的海难之一。——[维基百科 - 泰坦尼克号](#)

设计

本次可视化的数据样本来源于 <https://www.kaggle.com/c/titanic/data>，共包含了 891 条数据，而总乘客人数为 2224，所以这个样本数据已经能很好的代表总体数据。

根据数据集的特点，我选择使用柱状图和饼图来向读者传达乘客的幸存情况，以及乘客幸存率和男女比例、座位等级之间的关系。

本次的可视化分析图表采用 D3 + Dimple 绘制，鼠标移动到图表上时可以自动显示相关信息。所以并没有额外添加标签来显示相关信息。

泰坦尼克号乘客生存和死亡情况

首先，我先对样本数据进行聚合处理：

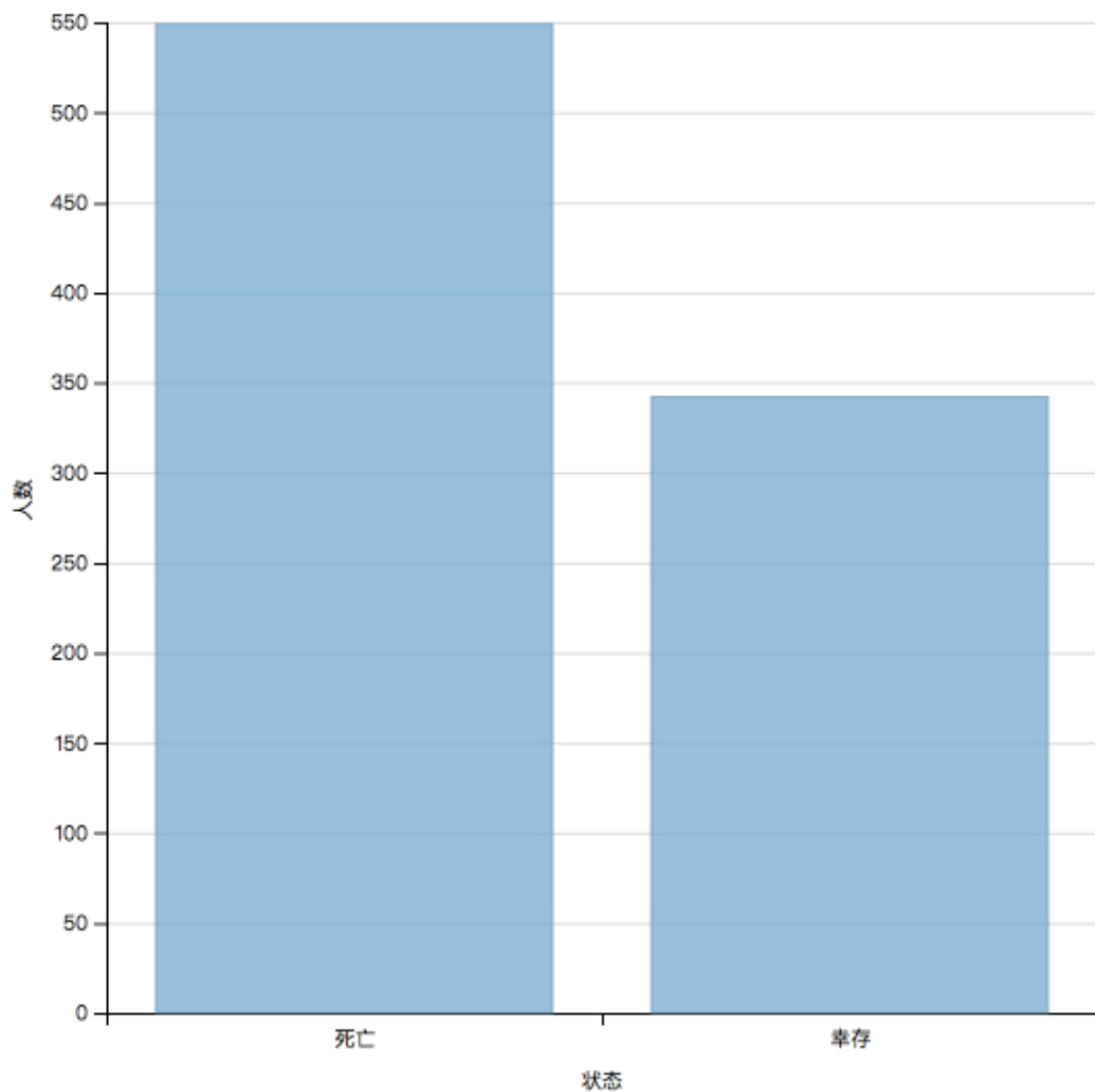
```
var nests = d3.nest()  
  .key(function(d) { return d.Survived; })  
  .rollup(function(peoples) { return peoples.length; })  
  .entries(data);
```

根据乘客的 Survived 状态，分成两组数据，然后使用 Dimple 画柱状图：

```
var svg = dimple.newSvg("body", width, height);
var chart = new dimple.chart(svg, nests);
var x = chart.addCategoryAxis("x", "key");
var y = chart.addMeasureAxis("y", "value");
var s = chart.addSeries(null, dimple.plot.bar);
x.title = "状态"
y.title = "人数"
chart.draw();

svg.append("text")
  .attr("x", (width/2))
  .attr("y", 30)
  .attr("text-anchor", "middle")
  .style("font-size", "16px")
  .text("泰坦尼克号乘客生存和死亡情况");
```

泰坦尼克号乘客生存和死亡情况



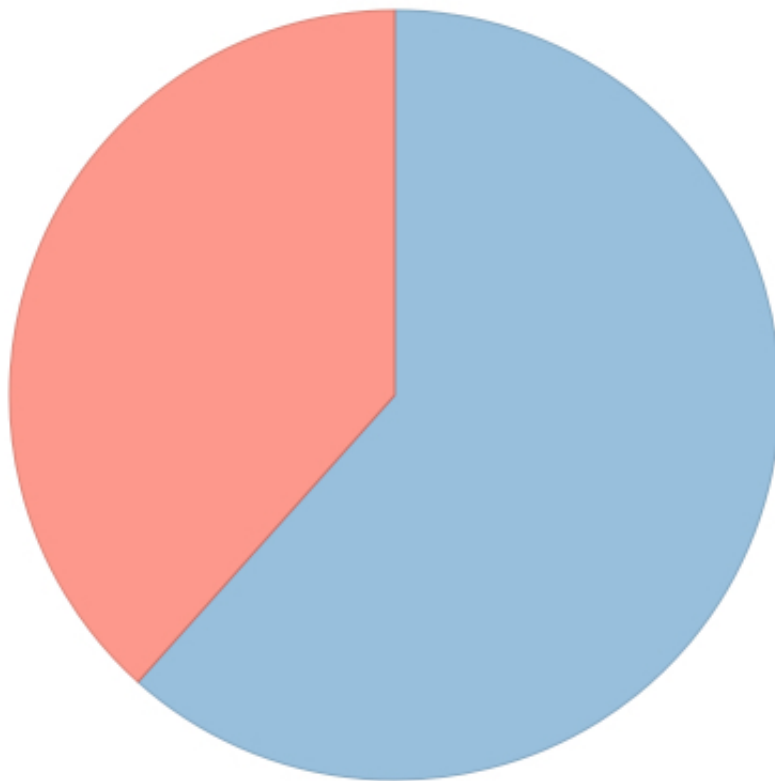
柱状图可以很直观的看出来生存和死亡的乘客具体数量。使用饼图可以很直观的看出来生存和死亡乘客的比例：

```
var svg_ticket = dimple.newSvg("body", width, height);
svg_ticket.append("text")
.attr("x", (width/2))
.attr("y", 30)
.attr("text-anchor", "middle")
.style("font-size", "16px")
.text("泰坦尼克号乘客的座位等级情况");

var chart_ticket = new dimple.chart(svg_ticket, nest_pclass);
var x_ticket = chart_ticket.addCategoryAxis("x", "key");
var y_ticket = chart_ticket.addMeasureAxis("y", "value");
var s_ticket = chart_ticket.addSeries(null, dimple.plot.bar);
x_ticket.title = "座位等级"
y_ticket.title = "人数"
chart_ticket.draw();
```

乘客生存死亡比例饼图如下：

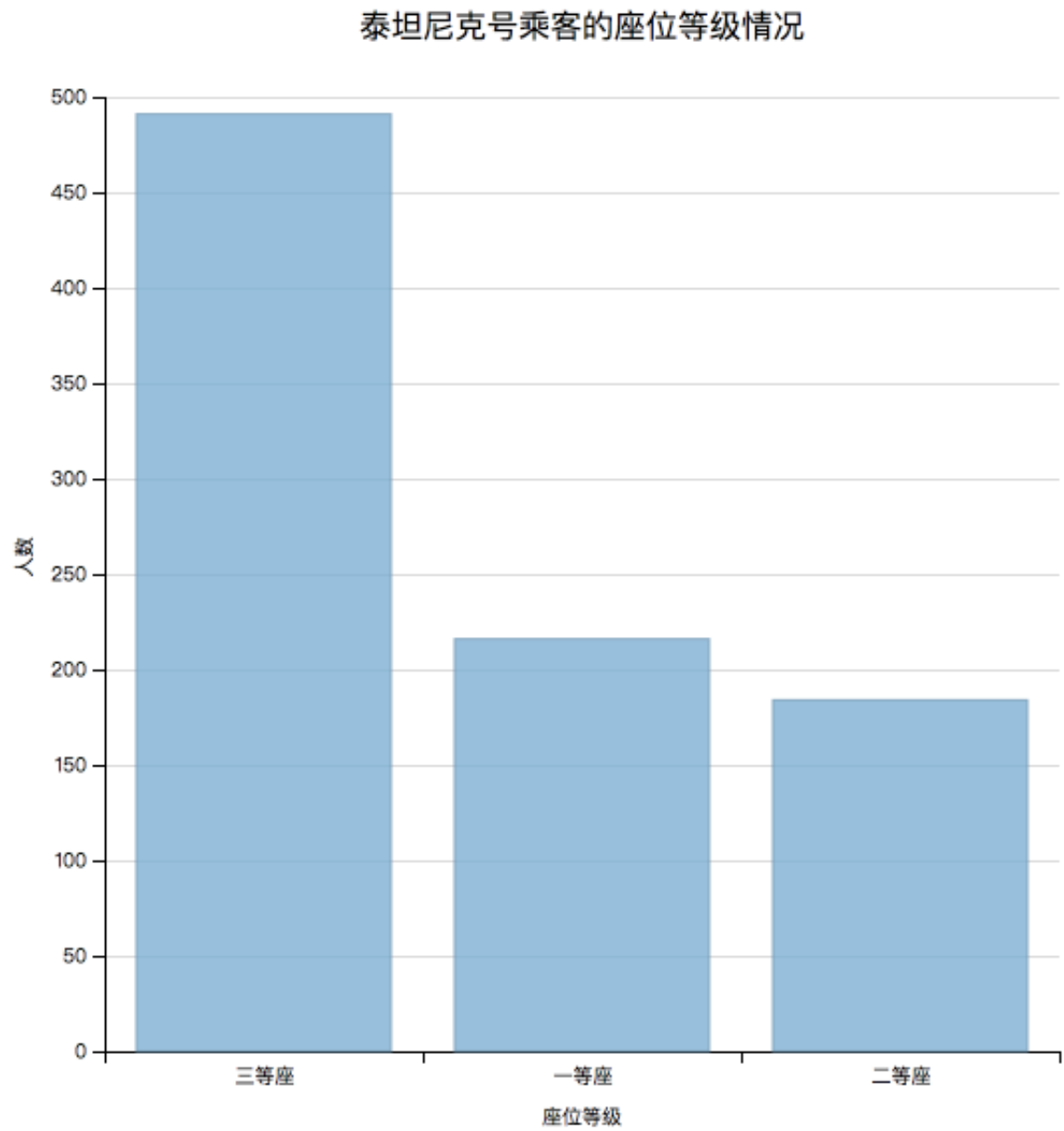
泰坦尼克号乘客生存和死亡比例



泰坦尼克号乘客的座位等级情况

我还注意到了泰坦尼克号上有三种不同的座位，分布于不同的座舱，那么不同的座位对于生存率是否会影响呢？

下面是不同座位的数量统计：



以及对应的不同座位的乘客生还率：

一等座：63%，二等座：47%，三等座：24%

一等座乘客生存和死亡比例 二等座乘客生存和死亡比例 三等座乘客生存和死亡比例



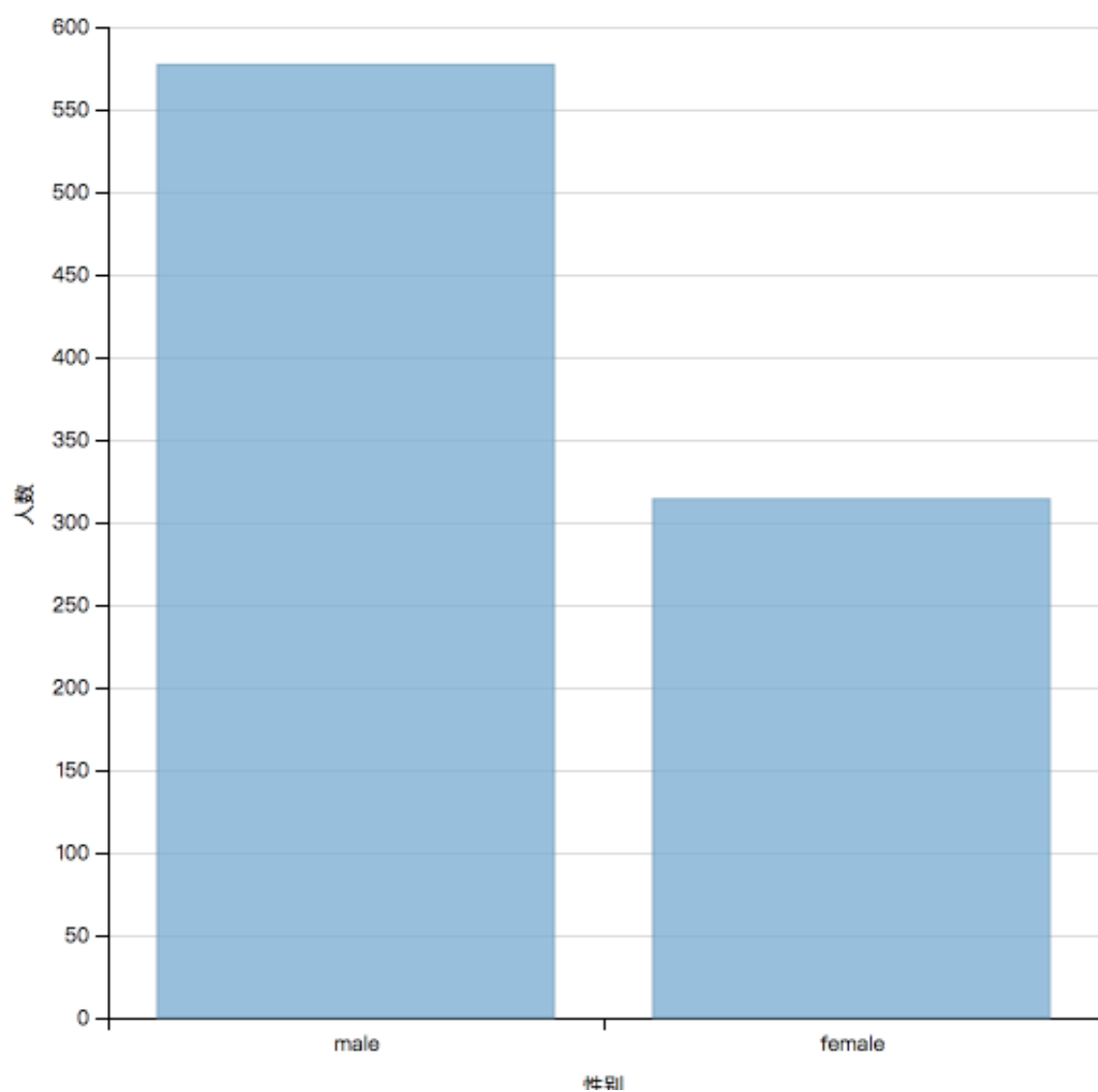
可以很轻易的看出来，座位越好，生还率越高。

泰坦尼克号乘客的性别情况

除了座位等级，乘客的性别是否对生还率有影响呢？下面我对乘客的性别做了可视化。

乘客的性别数量：

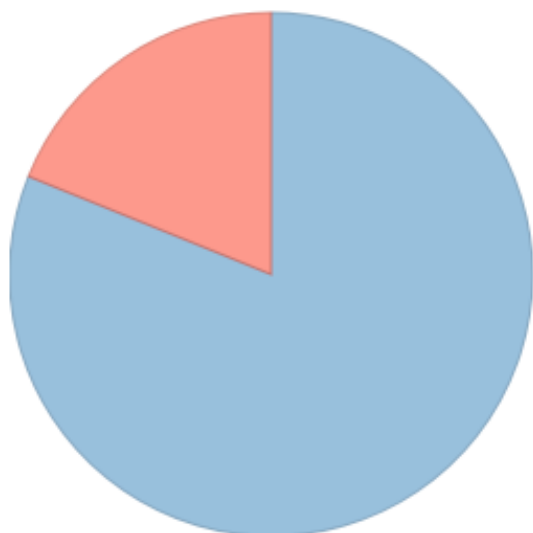
泰坦尼克号乘客的男女性别情况



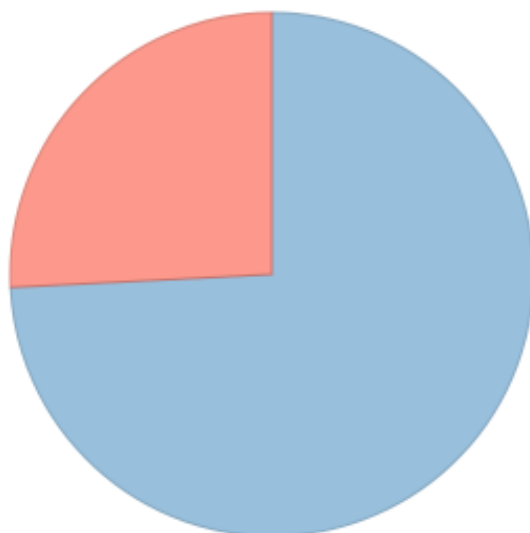
男女分别对应的生还率：

男性：19%，女性：74%

男性乘客的幸存死亡比例



女性乘客的幸存死亡比例



女性的生还比例远高于男性。由于泰坦尼克号的沉没时间较长，救助时间也比较长，体现了在灾难发生时，**女士优先** 的原则。

反馈

我向好友分享了我的可视化，并收到了一些反馈：

- 仅仅对乘客的生还率做可视化并不能体现出数据中有价值的地方。

改进：增加了不同座位等级和男女比例对生存率影响的可视化。

参考资源

- Dimple: dimplejs.org
- d3js.org: d3js.org
- <http://pkuwwt.github.io/d3-tutorial-cn/about.html>
- <http://www.ruanyifeng.com/blog/2014/11/basic-charts.html>
- <http://alignedleft.com/tutorials/d3>
- https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Global_Objects
- <https://bl.ocks.org/mbostock/3886208>
- <http://jsfiddle.net/NikhilS/3hgra/>
- <https://stackoverflow.com/questions/17226576/creating-a-text-labeled-x-axis-with-an-ordinal-scale-in-d3-js>

