

# 大数据可视化基础

## Task4：数据可视化大屏

### 第五组

- 成员：
- 2111408 网络空间安全学院 周钰宸
- 2111698 网络空间安全学院 于泽林
- 2111334 人工智能学院 刘宗桓
- 2111325 金融学院 韩啸宇
- 2112920 金融学院 肖一正

### 任务要求

请上传你的数字化大屏，说明你阐述的数据问题

### 任务准备

本次我们任务的主题是当代年轻人阅读情况统计

### 数据收集

我们首先通过问卷星的方式，由两位金融学院的同学，设计并在校内收集了100名在校南开学生的阅读书籍情况。统计结果如下：

	A	B	C	D	E	F	G	H
1		城市	每年阅读书籍数量	每年花费金额	主要购书平台	读书目的	最喜欢的一本书	
2	1	四川	4-6本	200元以上	网购	兴趣爱好	撒哈拉的故事	
3	2	北京	4-6本	0-50元	网购	兴趣爱好	白鹿原	
4	3	吉林	0-3本	50-100元	书店	为中华之崛起而读书	棋王·树王·孩子王	
5	4	北京	0-3本	200元以上	书店	提升专业知识	杨绛译唐吉珂德	
6	5	海南	10本以上	50-100元	书店	兴趣爱好	第七天	
7	6	天津	7-10本	200元以上	网购	为中华之崛起而读书	明朝那些事	
8	7	海南	10本以上	100-200元	书店	兴趣爱好	当代英国概况	
9	8	广东	0-3本	100-200元	网购	丰富精神世界	房思琪的初恋乐园	
10	9	宁夏	7-10本	100-200元	网购	丰富精神世界	国富论	
11	10	贵州	4-6本	200元以上	书店	为中华之崛起而读书	共产党宣言	
12	11	陕西	7-10本	50-100元	书店	为中华之崛起而读书	呼兰河传	
13	12	天津	4-6本	50-100元	网购	提升专业知识	张枣的诗	
14	13	云南	0-3本	50-100元	网购	为中华之崛起而读书	红高粱	
15	14	青海	0-3本	50-100元	网购	提升专业知识	第七天	
16	15	江苏	10本以上	50-100元	书店	为中华之崛起而读书	目送	
17	16	海南	7-10本	100-200元	网购	兴趣爱好	国富论	
18	17	山东	4-6本	0-50元	书店	丰富精神世界	红高粱	
19	18	河南	10本以上	0-50元	网购	为中华之崛起而读书	我们仨	
20	19	甘肃	0-3本	200元以上	书店	丰富精神世界	平凡的世界	
21	20	云南	10本以上	200元以上	书店	兴趣爱好	边城	
22	21	安徽	0-3本	100-200元	网购	提升专业知识	自在独行	
23	22	澳门	10本以上	50-100元	书店	提升专业知识	受活	
24	23	天津	0-3本	50-100元	网购	兴趣爱好	边城	
25	24	河南	4-6本	200元以上	网购	为中华之崛起而读书	自在独行	
26	25	澳门	4-6本	50-100元	书店	丰富精神世界	大秦帝国	
27	26	湖北	10本以上	200元以上	网购	丰富精神世界	傲慢与偏见	
28	27	安徽	0-3本	0-50元	书店	丰富精神世界	城南旧事	
29	28	湖南	4-6本	0-50元	书店	提升专业知识	南渡北归	

相关问题如下：

### NKU 读书情况

- 1.你是哪里人?
- 2.每年阅读书籍数量
  - 虚拟变量 1: 0-3 本
  - 虚拟变量 2: 4-6 本
  - 虚拟变量 3: 7-10 本
  - 虚拟变量 4: 10 本以上
- 3.每年花费金额
  - 虚拟变量 1: 0-50 元
  - 虚拟变量 2: 50-100 元
  - 虚拟变量 3: 100-200 元
  - 虚拟变量 4: 200 元以上
- 4.主要购书平台
  - 虚拟变量 1: 线上
  - 虚拟变量 2: 书店
- 5.读书目的
  - 虚拟变量 1: 提升专业知识
  - 虚拟变量 2: 兴趣爱好
  - 虚拟变量 3: 丰富精神世界
  - 虚拟变量 4: 为中华之崛起而读书
- 6.最喜欢的一本书

然后为了方便进行可视化处理，我们需要将上述问卷星调查结果中的一些文本类型的数据转换为数值类型。通过Python代码处理后结果如下：

1	ID	购书平台	每年买书的花销	居住城市	最喜欢的书	最喜欢的作者			
2	1001	京东	1200	北京	鲁迅全集	鲁迅			
3	1002	当当	1500	上海	活着	余华			
4	1003	亚马逊	800	广州	三体	刘慈欣			
5	1004	当当	1000	深圳	解忧杂货店	东野圭吾			
6	1005	京东	900	成都	哈利波特	J.K. 罗琳			
7	1006	天猫	1100	杭州	白夜行	东野圭吾			
8	1007	当当	950	武汉	小王子	圣埃克苏佩里			
9	1008	亚马逊	850	重庆	挪威的森林	哈尔德·赫曼			
10	1009	京东	1300	南京	围城	钱钟书			
11	1010	当当	1000	西安	百年孤独	加西亚·马尔克斯			
12	1011	京东	1200	北京	鲁迅全集	鲁迅			
13	1012	当当	1500	上海	活着	余华			
14	1013	亚马逊	800	广州	三体	刘慈欣			
15	1014	当当	1000	深圳	解忧杂货店	东野圭吾			
16	1015	京东	900	成都	哈利波特	J.K. 罗琳			
17	1016	天猫	1100	杭州	白夜行	东野圭吾			
18	1017	当当	950	武汉	小王子	圣埃克苏佩里			
19	1018	亚马逊	850	重庆	挪威的森林	哈尔德·赫曼			
20	1019	京东	1300	南京	围城	钱钟书			
21	1020	当当	1000	西安	百年孤独	加西亚·马尔克斯			
22	1021	京东	1200	北京	鲁迅全集	鲁迅			
23	1022	当当	1500	上海	活着	余华			

## 图片处理



为了能够让我们的可视化大屏效果更加好，我们在网上找到了一个图书的图片，并通过如下代码进行了透明度的处理：

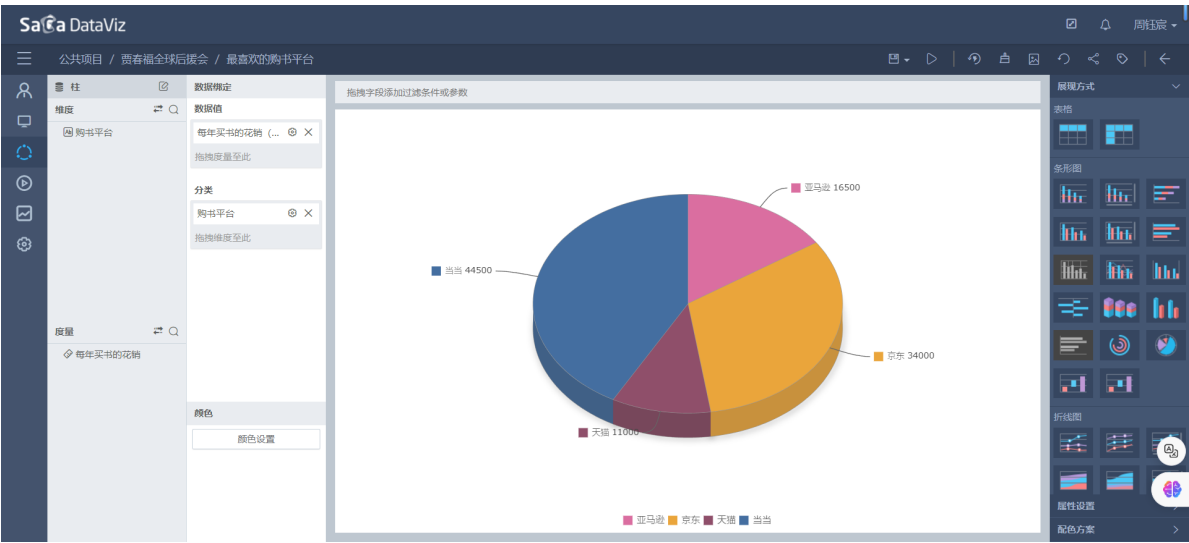
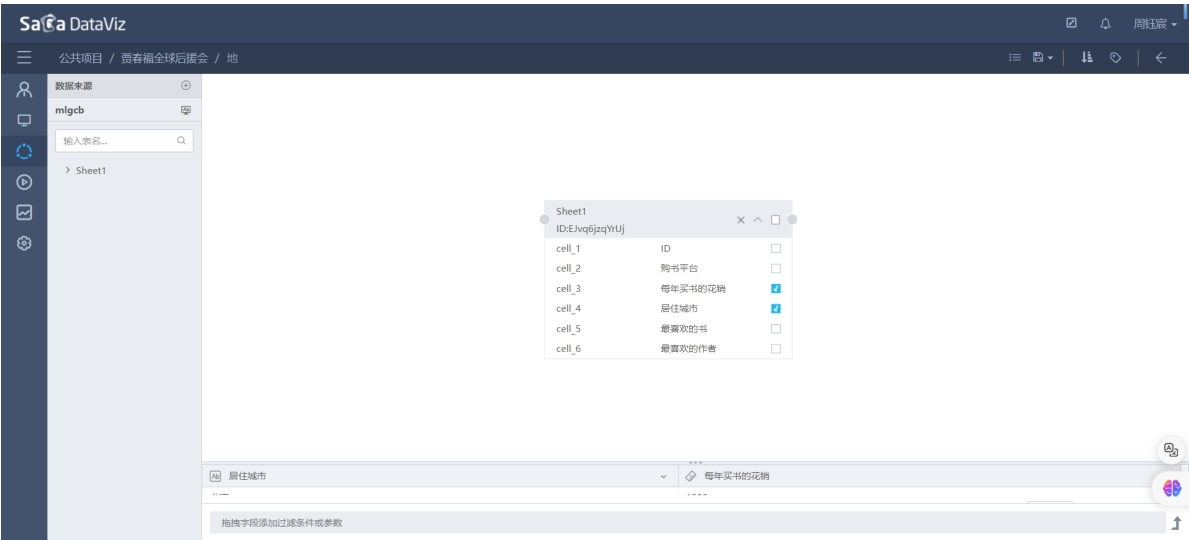
```
1  from PIL import Image
2
3  # Load the image
4  img_path = "data/Book.webp"
5  img = Image.open(img_path)
6
7  # Convert the image to RGBA if it's not already in that mode
8  img = img.convert("RGBA")
9
10 # Retrieve the data of the image
11 datas = img.getdata()
12
13 # Create a new data list with modified alpha values
14 new_data = []
15 for item in datas:
16     # Change the alpha value (reduce opacity to 50%)
17     new_data.append(item[:-1] + (int(item[-1] * 0.5),))
18
19 # Update the image data with the new reduced opacity data
20 img.putdata(new_data)
21
22 # Save the modified image with reduced opacity
23 new_img_path = "data/Book_reduced_opacity.webp"
24 img.save(new_img_path)
```

处理效果如下：

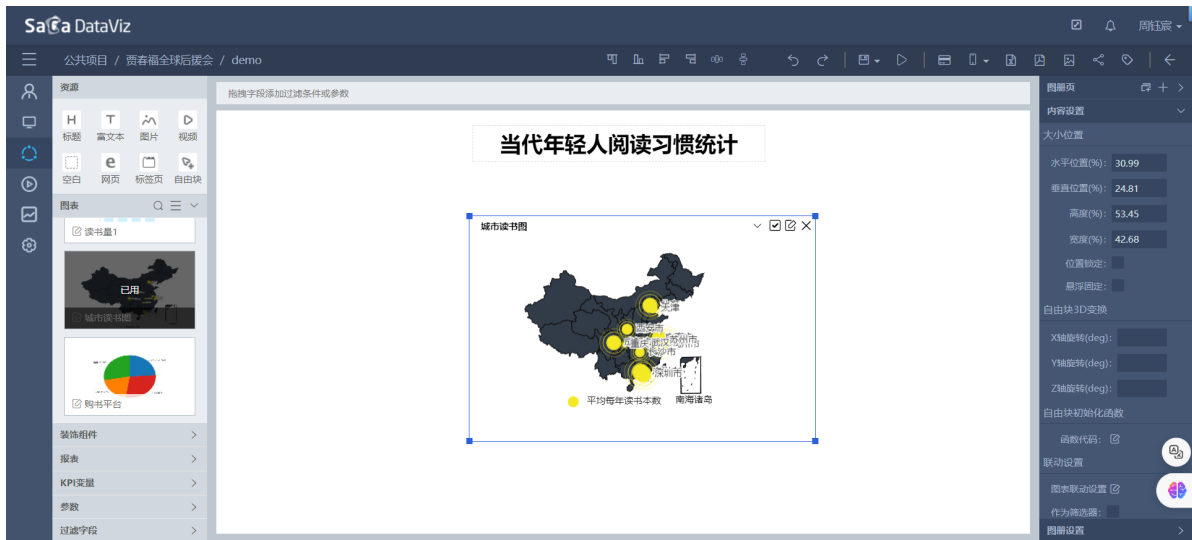


## 任务过程

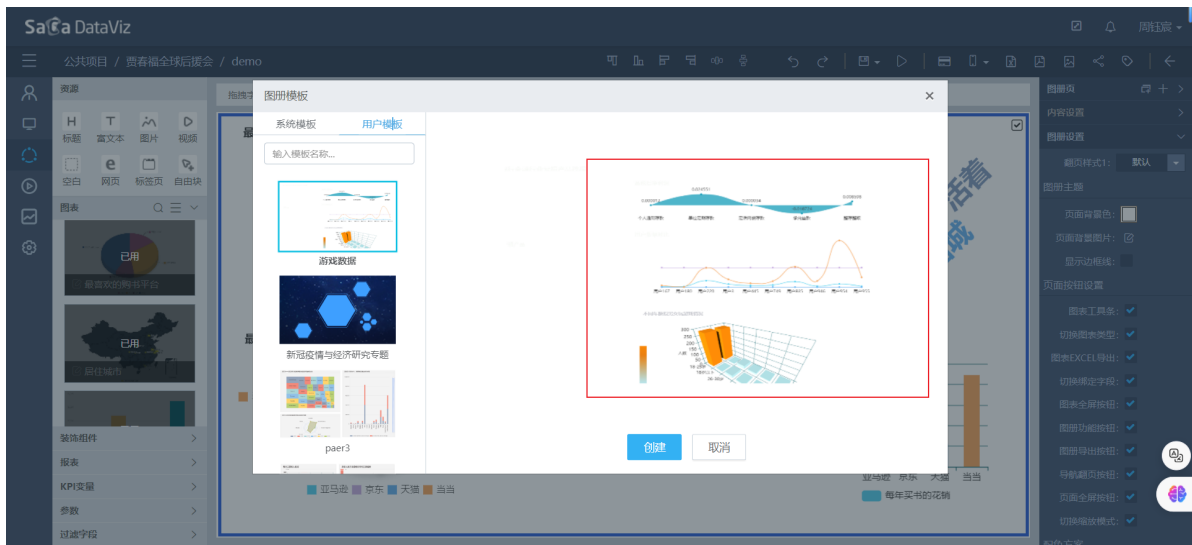
接下来我们利用数据依次制作每张子图：



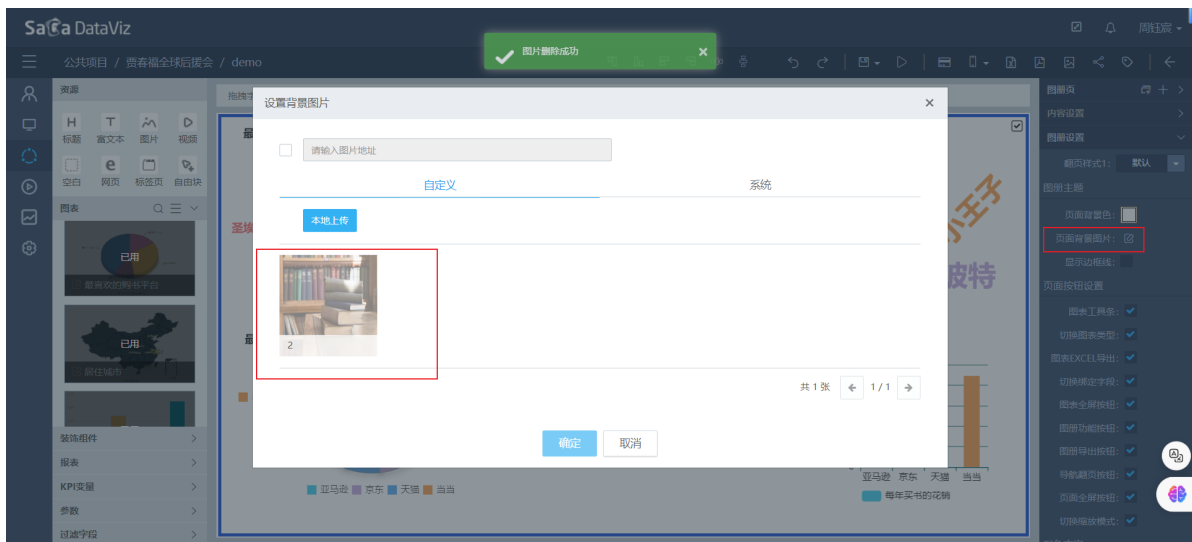
使用自由块组成可视化大屏：



选择可视化模板美化效果：



利用制作好的背景图片：



选择模板添加标题栏，最终呈现效果如下：



## 任务结果展示



从上图结果中可以看出：

1. **最喜欢的作者：**日本著名推理小说家东野圭吾大获全胜，另外还有著名科幻小说刘慈欣等。
2. **最喜欢的购书平台：**可以看到当当和京东占据大头，亚马逊和天猫占较小部分。
3. **最喜欢的书籍：**几乎都平行占据位置，其中白夜行等可以和作者相互联系起来。
4. **每年买书的花销：**可以看到在当当网中大家购买的花销是最大的，另外花销也和购书平台喜爱度呈现正相关和累计效果。
5. **居住城市：**我们本次调研的南开学生在南方的居多，也是因为我们组员大部分来自北方，想要多了解下南方的同学读书情况。

我们的可视化大屏链接为：<https://e.idataviz.com/src/q.html?s=nankai-e3a60cd632ed28c4>

二维码为：



## 任务感受

本次任务我们五个人通过小组合作，各发挥所长，完成了本次任务。我们收获颇丰，具体而言：

1. 通过分工安排，各取所长，我们收获了效率和沟通交流的合作能力，这对我们未来学习和工作都很重要。
2. 通过亲自制作问卷星并收集数据，我们锻炼了解决实际问题的能力。
3. 通过使用Dataviz进行可视化，我们了解了该可视化工具的强大和食用。

感谢老师审阅！