作业2-1 报告

本次作业分为两部分：

1. 处理源数据
    - 手动删除多余属性列
    - 统一属性名（如将“通话时长”修改为“通信时长”等）
    - 利用Python脚本“countTime.py”将通信时长单位统一为秒
    - 手动标记添加“亲密性”属性。“1”代表“亲密”，“-1”代表“不亲密”
    - 手动标记添加“性别”属性。“1”代表“男性”，“0”代表“女性”，“2”代表“不详”
2. 计算统计值和绘制盒图
    - 编写处理数据的Python脚本
    - 读取csv文件并有效获取数据
    - 计算每个通信方的通信总时长
    - 计算“通信总时长”这组数据的平均数、中位数、五数概括以及通信次数的众数等统计值
    - 针对“通信总时长”数据绘制盒图

运行说明：

- 所需Python库：matplotlib
- 将数据文件放到该项目根目录内，命名为“myCallsData.csv”
- 运行Python脚本“preProgress.py”脚本即可

程序功能介绍：

函数说明

- csv2dict：将csv文件内数据读取到Python字典中，第一个参数为字典中的key（本程序中以对方手机号码为key），第二个参数为想要获取的属性或结合本程序逻辑的分析种类

- getAverage：获得平均值，参数为待求数组
- getPercentageValue：获得百分位数（用于计算四分位数，中位数等），第一个参数为待求数组，第二个参数为百分位

逻辑说明

1. 获取‘对方号码’-‘通信时长’的字典，对字典进行排序，并将通信时长赋值给一个数组
2. 对有序的通信时长数组进行计算，得到平均数、中位数、四分位数等统计值
3. 获取‘对方号码’-‘通信次数’的字典，对字典进行降序排序，获得最大的通信次数，即通信次数众数
4. 获取标记为“亲密关系”的人的通信时长数据相关的字典，进行相似处理
5. 对“通信总时长”和“亲密人通信总时长”两组数据绘制盒图

作业成果展示截图：

起始时间	通信时长	通信方方式	对方号码	通信地点	通信类型	亲密性	性别
2017/5/2 21:21	47	被叫	18935375789	黑龙江哈尔	国内通话	1	1
2017/5/1 21:11	1269	被叫	18221187502	黑龙江哈尔	国内通话	1	1
2017/5/1 20:34	11	被叫	18221187502	黑龙江哈尔	国内通话	1	1
2017/4/30 21:02	183	被叫	18845897013	黑龙江哈尔	国内通话	-1	
2017/4/29 13:09	35	被叫	18545529543	黑龙江哈尔	国内通话	-1	
2017/4/19 16:12	18	被叫	18545120810	黑龙江哈尔	国内通话	-1	
2017/4/16 12:21	17	被叫	15546530029	黑龙江哈尔	国内通话	1	1
2017/4/15 21:12	154	被叫	18935375789	黑龙江哈尔	国内通话	1	1
2017/4/15 14:02	272	被叫	15235797188	黑龙江哈尔	国内通话	1	1
2017/4/14 21:06	88	被叫	13835795019	黑龙江哈尔	国内通话	1	0
2017/4/14 18:55	154	被叫	13546716025	黑龙江哈尔	国内通话	1	0
2017/4/14 17:46	18	被叫	15504602637	黑龙江哈尔	国内通话	1	0
2017/4/11 14:32	34	主叫	15546530029	黑龙江哈尔	国内通话	1	1
2017/4/1 22:33	124	被叫	18935375789	黑龙江哈尔	国内通话	1	1
2017/3/30 21:59	42	被叫	18800463842	黑龙江哈尔	国内通话	1	1
2017/3/29 13:59	26	被叫	18800463842	黑龙江哈尔	国内通话	1	1
2017/3/27 11:52	22	被叫	18246012393	黑龙江哈尔	国内通话	-1	
2017/3/21 09:45	22	被叫	15204690814	黑龙江哈尔	国内通话	-1	
2017/3/17 18:38	87	被叫	18845720978	黑龙江哈尔	国内通话	1	1
2017/3/16 18:54	19	被叫	18545529543	四川成都	国内通话	-1	
2017/3/16 11:31	51	被叫	13982083318	四川成都	国内通话	-1	0
2017/3/16 10:54	61	被叫	13982083318	四川成都	国内通话	-1	0
2017/3/10 17:11	61	被叫	15776611862	黑龙江哈尔	国内通话	-1	
2017/3/8 14:04	52	被叫	15546530029	黑龙江哈尔	国内通话	1	1
2017/3/8 12:09	67	被叫	18045000153	黑龙江哈尔	国内通话	-1	
2017/3/7 17:47	15	主叫	15546530029	黑龙江哈尔	国内通话	1	1
2017/3/7 12:57	28	被叫	13110107108	黑龙江哈尔	国内通话	1	0
2017/3/7 12:55	12	被叫	18245177131	黑龙江哈尔	国内通话	-1	
2017/3/7 12:31	21	被叫	18245177131	黑龙江哈尔	国内通话	-1	
2017/3/6 12:03	25	主叫	13110107108	黑龙江哈尔	国内通话	1	1
2017/3/5 10:08	20	主叫	15204692757	黑龙江哈尔	国内通话	-1	
2017/3/5 09:04	20	被叫	13313666968	黑龙江哈尔	国内通话	-1	
2017/3/4 10:44	34	被叫	15546530029	黑龙江哈尔	国内通话	1	1
2017/3/4 09:51	30	主叫	13946113823	黑龙江哈尔	国内通话	-1	0
2017/3/4 08:57	20	主叫	13946113823	黑龙江哈尔	国内通话	-1	0
2017/3/3 22:24	49	被叫	13817949464	黑龙江哈尔	国内通话	-1	
2017/3/3 10:14	70	主叫	18245019564	黑龙江哈尔	国内通话	1	1
2017/3/2 20:29	147	被叫	15326665870	黑龙江哈尔	国内通话	-1	1

对数据源进行处理后部分数据截图

```
$ python preProgress.py
Start preProgressing the data
总通信时长均值：  281
总通信时长中位数：  65
总通信时长五数概括：   min: 9;   q1: 33;   middle: 65;   q3: 195;   max: 3711
总通信次数众数：  40,  通信方为：13835795019
```

程序运行后，统计值输出

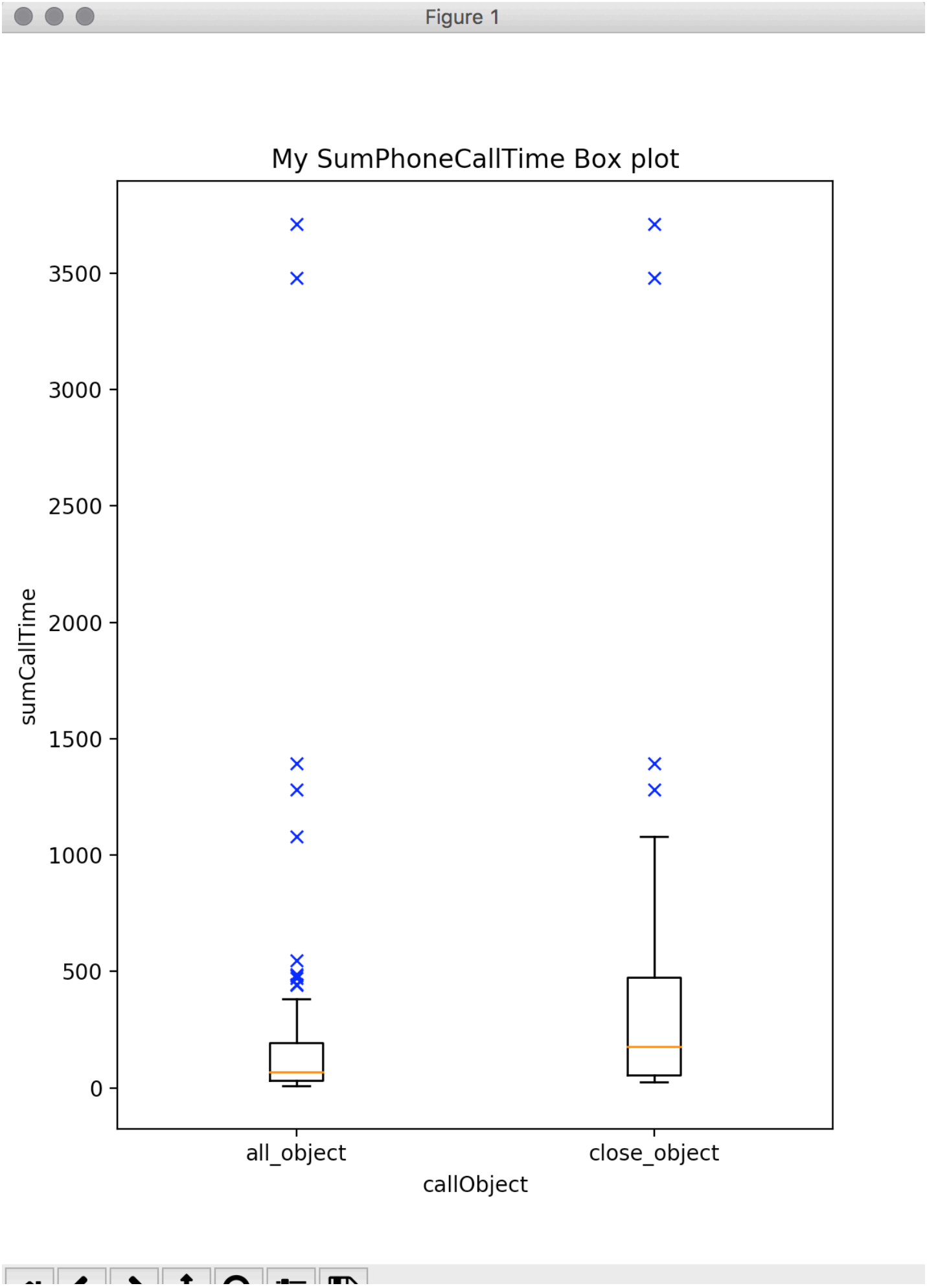
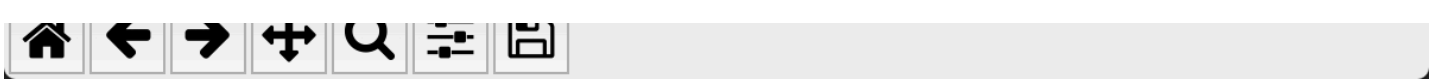


Figure 1
My SumPhoneCallTime Box plot
3500
3000
2500
2000
1500
1000
500
0
sumCallTime
all_object
close_object
callObject



盒图展示：左边为全部通信人，右边为标记为亲密关系的联系人