作业1.2

## 程序说明

* 程序使用julia编程语言编写，运行前需要安装julia运行时环境(类似Java程序需要安装JRE才能运行一样)
* gbk2utf8.py是将Excel导出的csv的文件编码转成utf8的小程序，便于后续分析
* preprocess.jl是主程序，对数据进行基本的处理并且输出基本统计量

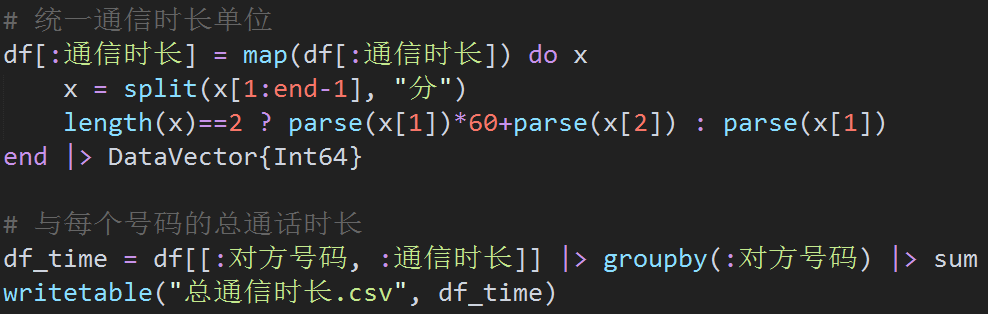
## 数据文件说明

格式: 除rawdata.csv外，所有文件都是UTF8编码，LF换行符(Unix)，因此使用Windows自带的记事本以及Excel打开可能会出现乱码、没有换行等情况。使用Sublime, Notepad++等编辑器打开可以看到正确格式

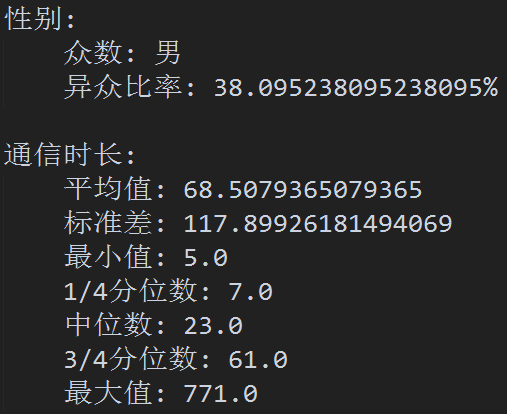
* rawdata.csv是Excel直接导出的文件
* rawdata.csv.utf8是将rawdata.csv编码转成UTF8后的文件，内容上是相同的
* 总通信时长.csv是与每个电话号码的通话时间之和
* processed.csv是处理后的数据
* plot\*.svg是画的盒图
* statistics.txt是输出的各列的统计量，对于通话时长有均值、五数概括等，其它列只有众数和异种比率

## 截图

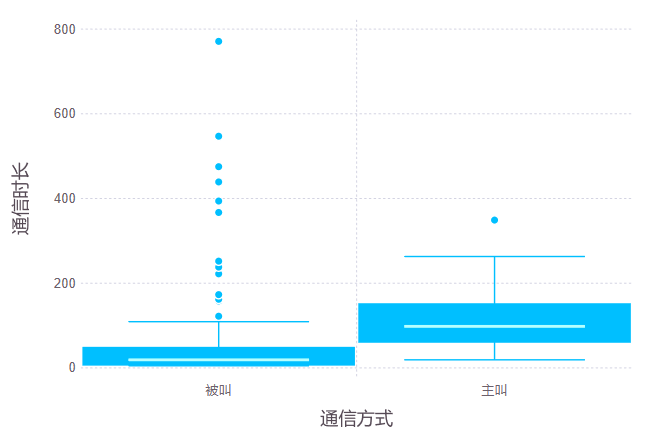
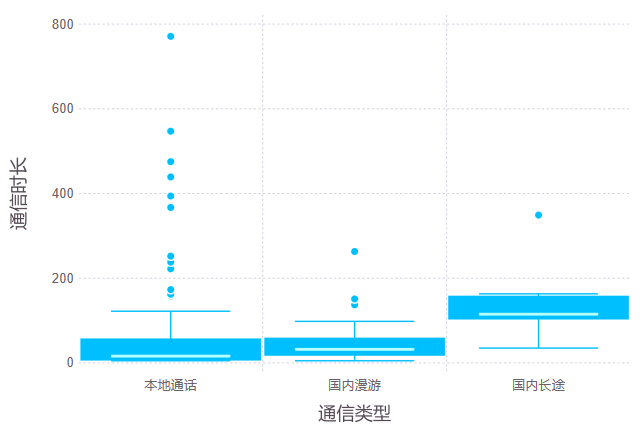
代码片段(preprocess.jl):



统计量片段(statistics.txt):



盒图(plot2.svg, plot4.svg):

本地、被叫通话时长偏低的部分原因是有大量的外卖电话，平均10秒不到，远低于正常通话时间。