Python作业报告

1. 代码实现介绍

共有两个py文件：population.py和consumptionlevel.py，分别用来爬取人口数据和居民消费水平数据并作图，两个文件实现方法类似，故以下仅以population.py为例阐述实现方法。

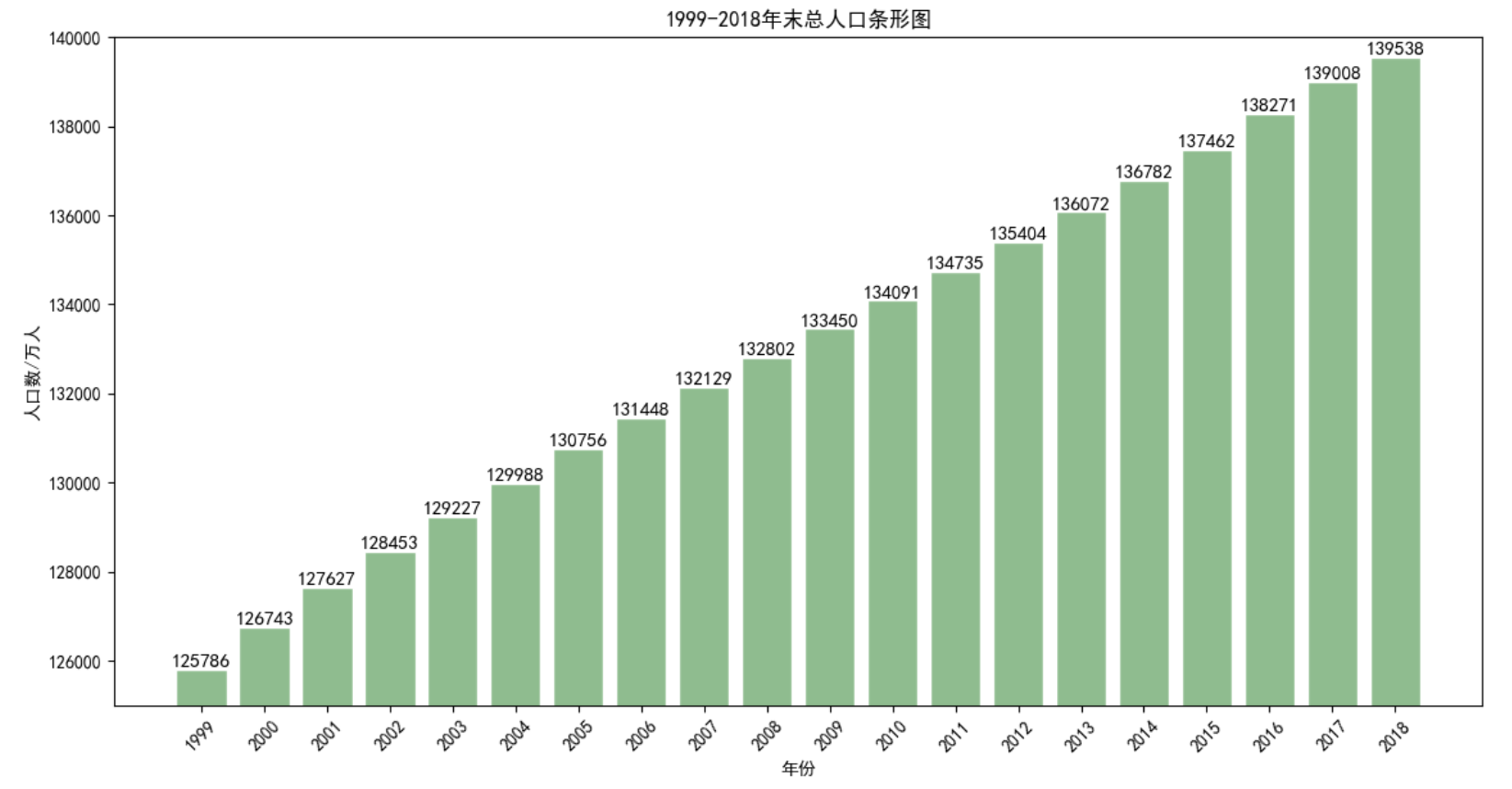
Population.py中有gettime、getpopulation、plotdata这几个主要方法：

gettime是用来获取当前时间的时间戳以用来填充爬取数据时的参数。

getpopulation是用来爬取数据并存入本地数据库的方法。首先根据国家统计局的网页参数填充keyvalue参数，建立一个session并在session的基础上用post方法爬取数据，然后修改keyvalue中的‘dfwds’参数以爬取过去20年的所有数据，再在session的基础上用get方法爬取数据并存入r中。随后利用json.loads从r.text中提取想要的数据，分别根据不同数据的标签提取年份、总人口、男性人口和女性人口数据并逆序存放（按时间从早到晚排序）。随后是向sqlite数据库中写入数据，利用sqlite3.connect连接数据库population.db（不存在时自动创建），在数据库中创建table，并用insert向table中插入数据。

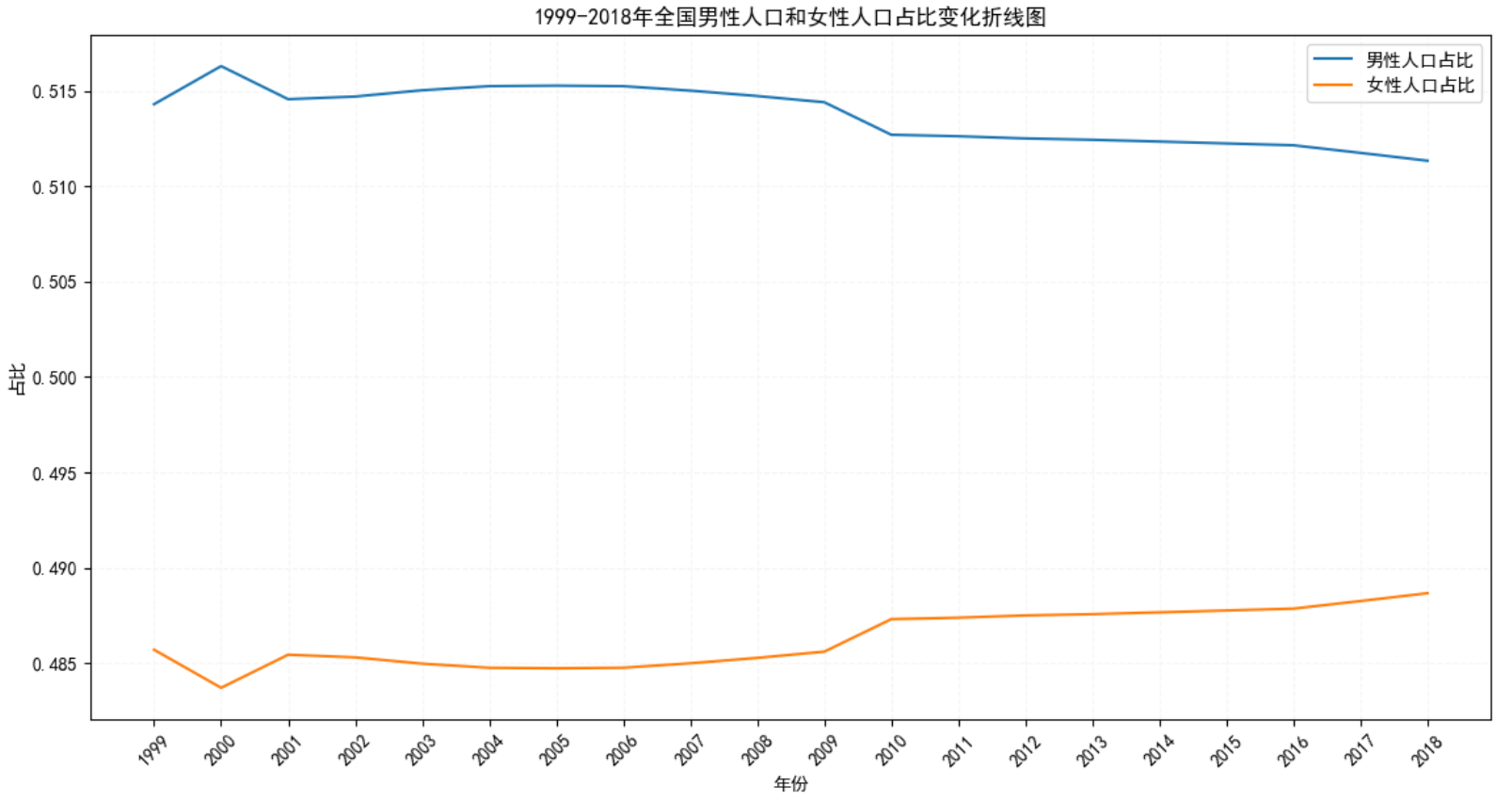
plotdata方法是用来作图以更好地分析数据的方法，其中有plot1和plot2方法分别绘制不同的图。首先从数据库中提取所有数据并存入data中，这是一个高维list，从中分别提取出不同数据。plot1是作年份-总人口条形图，plot2是作年份-男女人口占比折线图。

1. 数据分析
2. 年份——全国总人口条形图



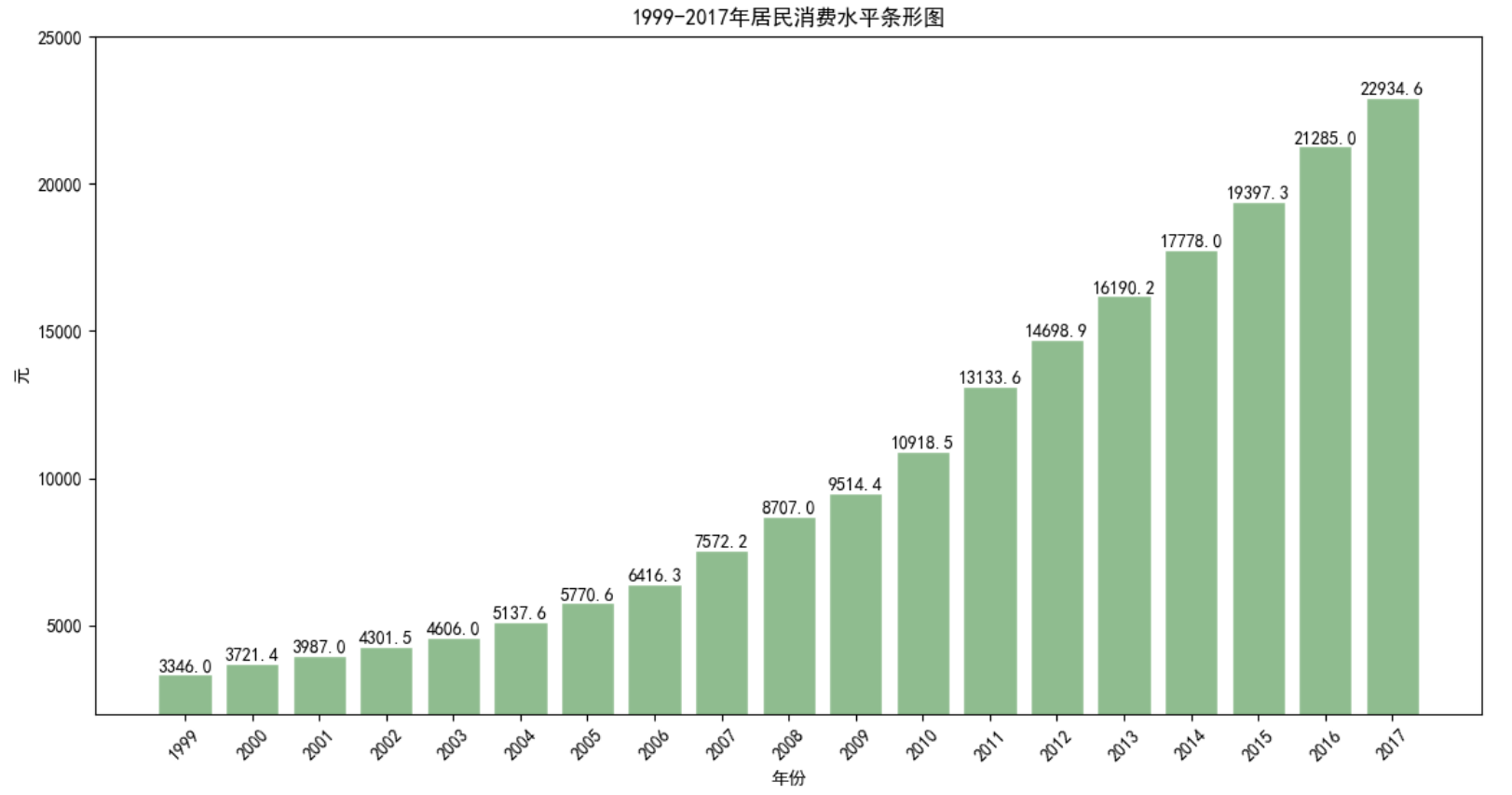
数据分析：从上图可以看出，从1999年到2018年，全国总人口一直处于上升趋势，到2018年末时，全国总人口达到139538万人，接近14亿。

1. 年份——全国男女性人口占比折线图



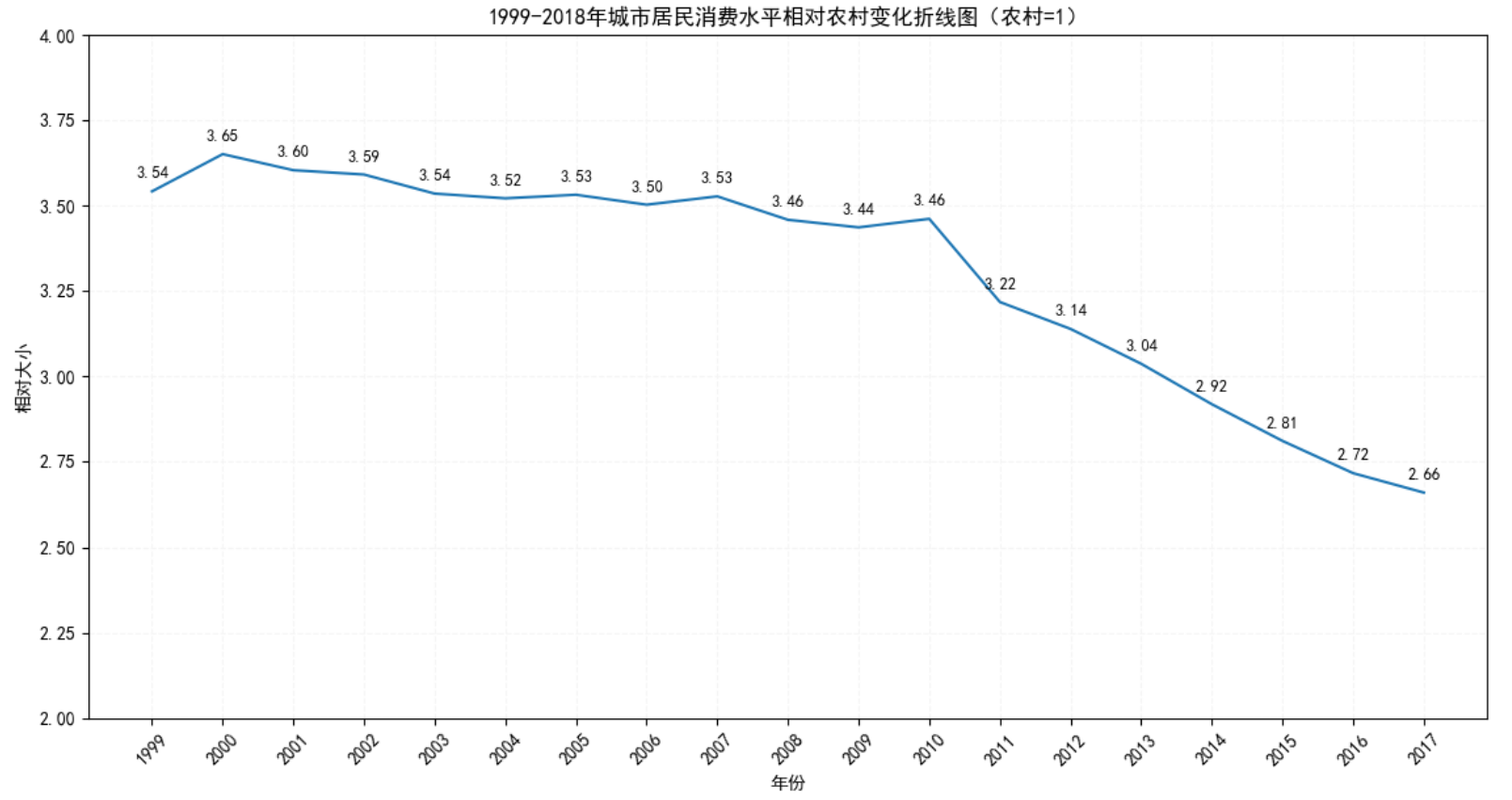
数据分析：从上图可以看出，从2006年到2018年，男女人口差距呈现逐渐缩小的趋势，说明人们重男轻女的思想观念在逐渐减轻。

1. 年份——居民消费水平条形图



数据分析：从上图可以看出，从1999年到2017年，居民消费水平逐渐上升，并且2007年以后增长速度相比以前变快，说明人们的生活水平在逐渐提高。

1. 年份——城市居民消费水平相对农村变化折线图（农村=1）



数据分析：从上图可以看出，从1999年到2017年，城镇相对农村居民消费水平总体呈下降趋势，特别是在2010年后下降变快，说明城镇和农村的居民消费水平差距在逐渐减小，但是目前差距仍然很大。

1. 总结与体会

在学习python一段时间后，对python也比较熟悉了，加上这次作业难度不是特别高，所以完成的过程比较顺利。遇到的问题是刚开始爬取数据的时候没有建立session，导致只能爬取某一年的数据或者爬不到数据，后来通过建立session并通过分次爬取数据的方法解决了这一问题。

这次作业让我学会了如何通过python爬取网页数据，熟悉了sqlite的使用，学会了将数据存入本地数据库并读取及利用matplotlib作出统计图形等，加大了对python使用的熟练程度。