

Python 作业报告

一. 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:

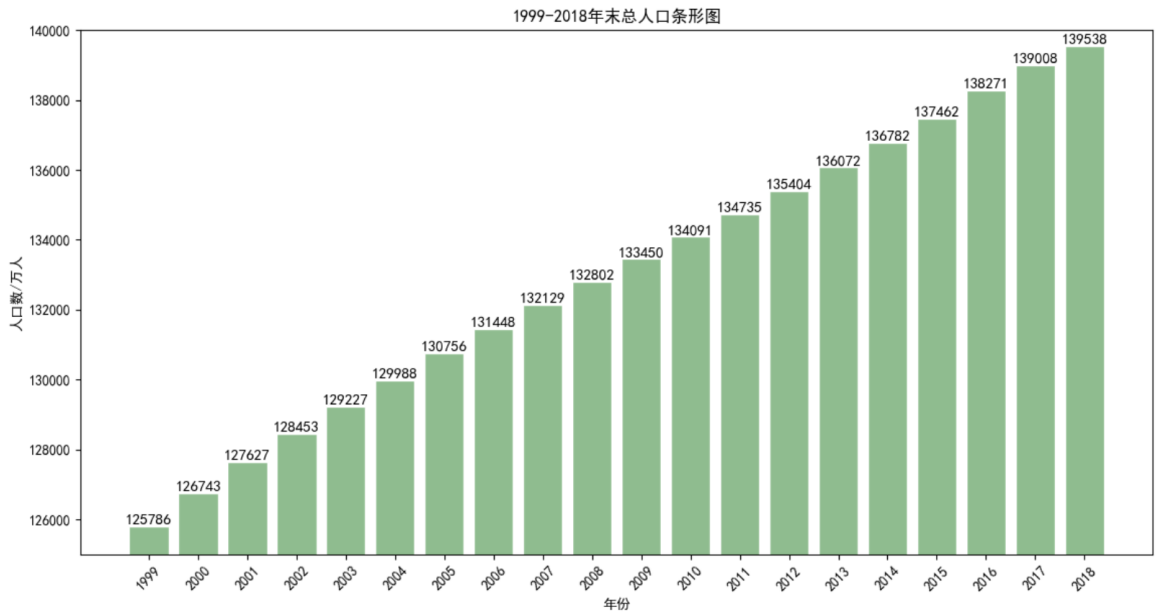
gettime 是用来获取当前时间的时间戳以用来填充爬取数据时的参数。

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 中提取想要的信息, 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。

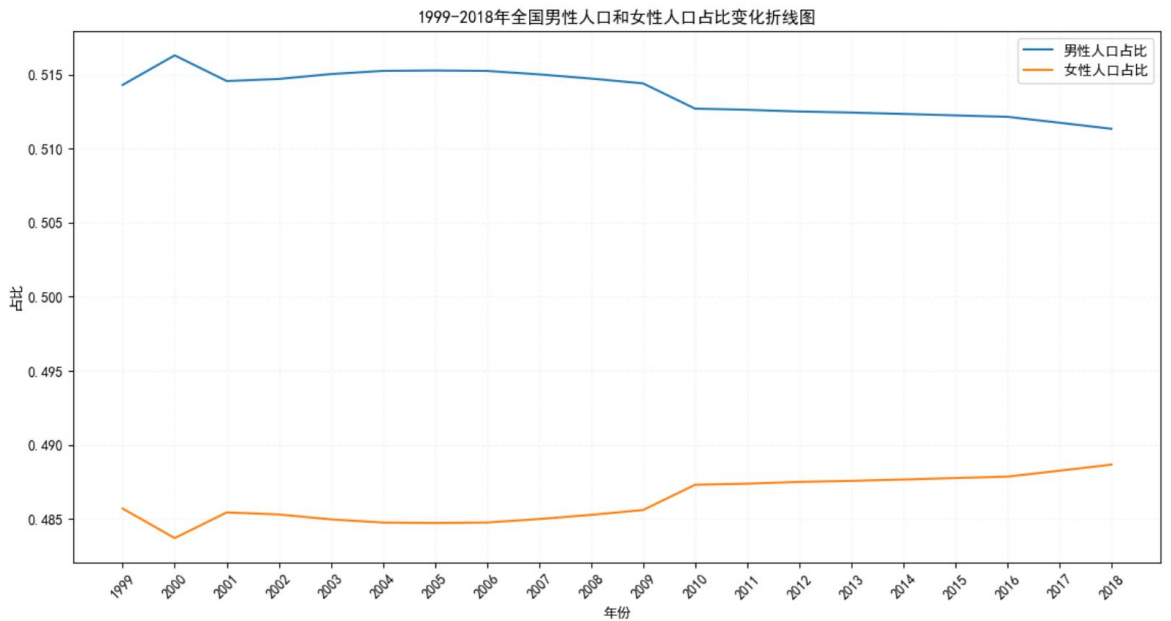
二. 数据分析

1. 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。

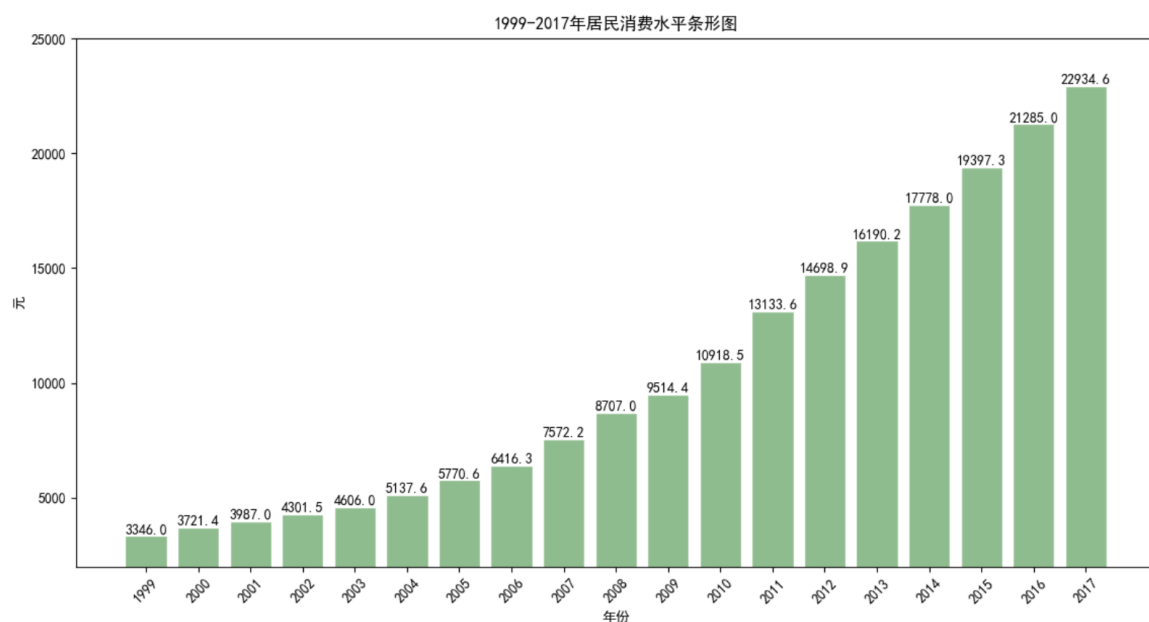
2. 年份——全国男女性人口占比折线图



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

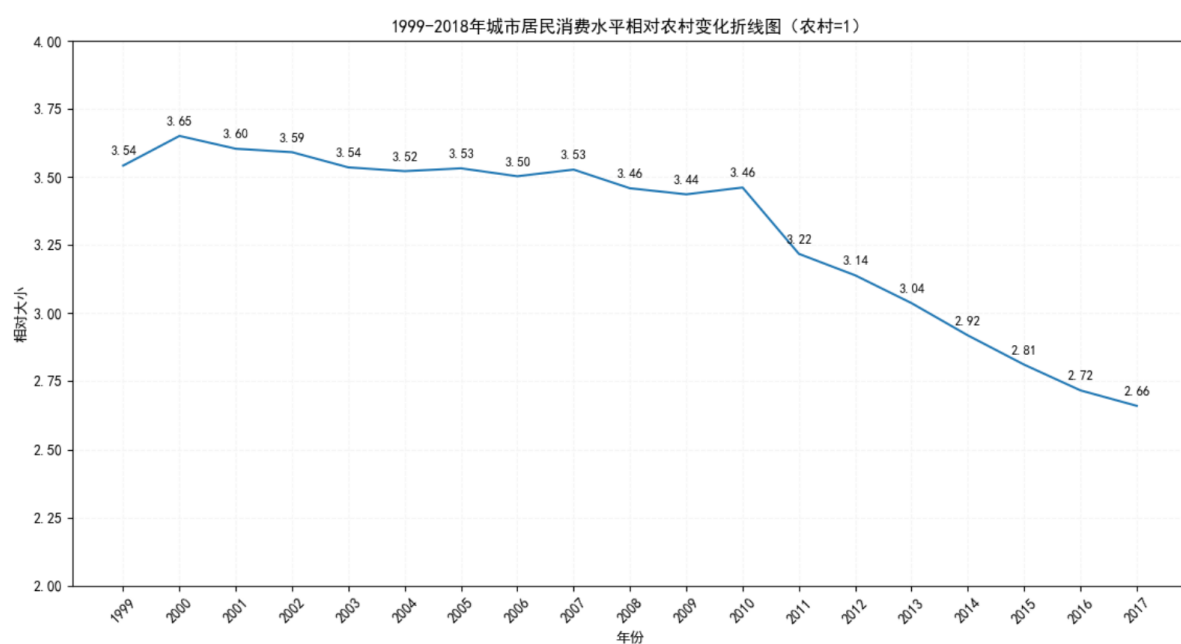
的趋势，说明人们重男轻女的思想观念在逐渐减轻。

3. 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。

4. 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。

三. 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。