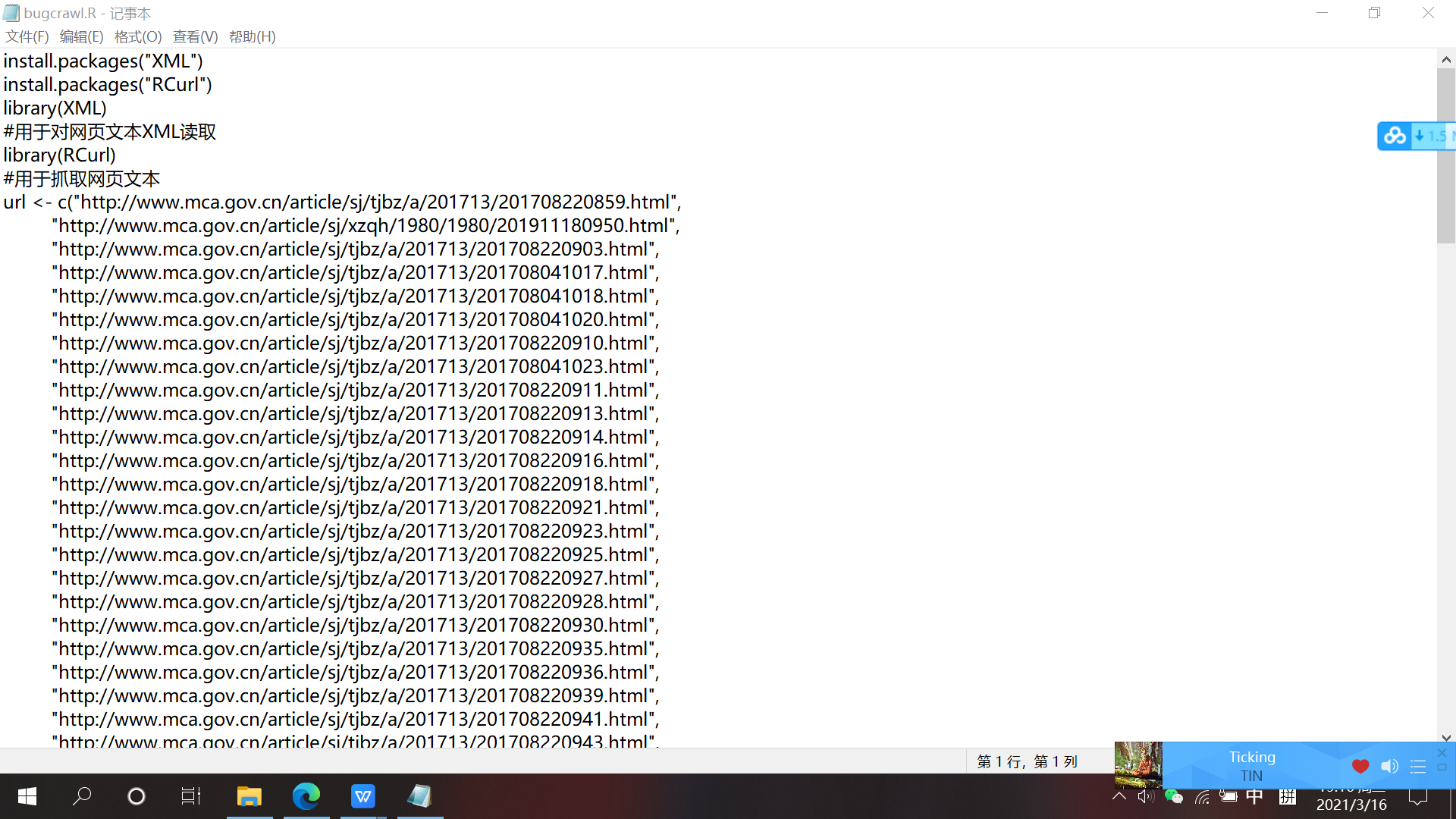
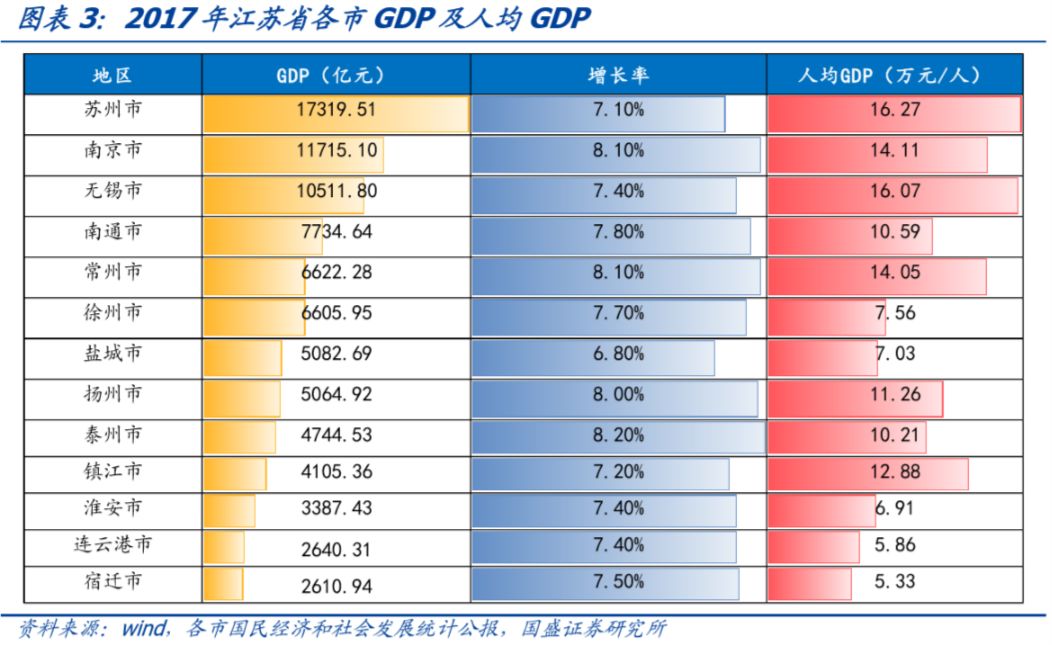
通过民政局网站http://www.mca.gov.cn/article/sj/xzqh/统计了1986-2019年间上海、江苏、浙江和山东四个省地级市（及部分直辖县）的行政区划代码、名称及简称，并整理成csv形式。

使用RCurl与XML实现对网络文本的爬虫抓取，按照URL码获取网页XML脚本文件并通过R语言字符串剪切分割合并数组的方式实现自动剔除多余数据，获取地级市数据并合并为表，后期人工进行校验核对工作。在过程中出现几个典型问题在于2013年包含了所有市和县数据而且file类型不好打开，导致这一年必须手动操作，而且2011、2012年格式与其它几年不同，给校验过程带来了困难。好在有各年的修改情况披露文件可以参考分析。



后期寻找数据，获取了2017年有关江苏省GDP,市政情况和债务情况的数据，以及浙江省各年的浙江省发电量和货运量，如下图



通过以上数据对已有的程序进行debug操作，并发现了在字符串的处理和解码文本时出现的全NULL情况，进而导致了程序结果错误。并在自己的数据中发现了处理时的错误。

本次SRT参与带来最大的感想是自己与科研的差距和在计算机方面的不足，同时这次宝贵的经历也带给我巨大的帮助。例如第一次在Github与学长学姐们共同合作，让我深刻理解到了协同编程与交互的重要性，同时也带来了巨大便利。这些都让我在后来IT方向实习时更好的融入了公司的项目学习和参与。第二则是在 R语言上感受到了自己存在的不足，一方面是用包和各种R式风格的编程技巧不熟练，在阅读学长学姐们的代码时令我常常陷入沉思，而很多重要工作，如function的编写和代码debug解决bug都因我学艺不精而难以胜任。但是同时这也化作了我更加学习R语言的动力，也在本次学习中让我更加体会到R这一数据处理软件的强大之处和魅力所在。最后，则是我的数据直观感受得到了显著提升。作为信息大数据方向交叉性学科专业的学生，我从原先拘泥于编程技巧思路，甚至在R中妄图通过指针类比设计来处理省级变动转变为了一种对大数据表格的分析和处理，抓住异同而非用程序一了百了的感受直觉。

本次SRT的学习经历受益匪浅。更让我难忘的，是各位学长学姐们热心的帮助和导师的幽默和严谨敏锐的洞察力，激励着我在学术道路上继续前行。