



时空数据分析与挖掘实习

——实习日志一

学 院：遥感信息工程学院

班 级：2006 班（20F10）

姓 名：马文卓

学 号：2020302131249

老 师：田扬戈

时 间：2023 年 5 月 15 日

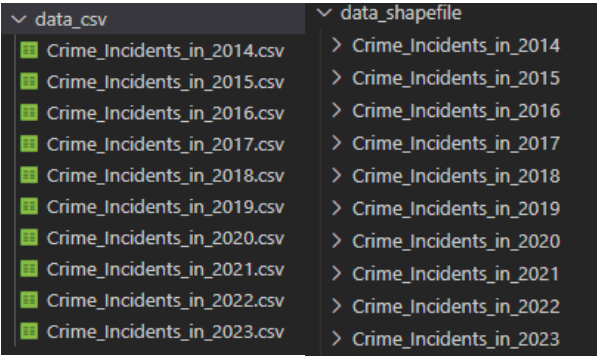
1. 今日进度

- ① 数据下载与解析
- ② 犯罪数量随年份变化分析
- ③ 犯罪位置热力图分析

2. 阶段结果与分析

(1) 数据下载与解析

本次实习我们使用的是从<https://opendata.dc.gov> 开源数据网站中下载的华盛顿特区（WashingtonDC）犯罪事件数据。值得注意的是，我从该网站上下载了十年的（2014-2023）华盛顿特区犯罪数据（两种形式：csv 和 shapefile），总共 308766 个犯罪事件作为本次实习的原始数据集。



其中，一个犯罪案件包含这样 24 个字段（其中标红的为关键字段），如下表所示：

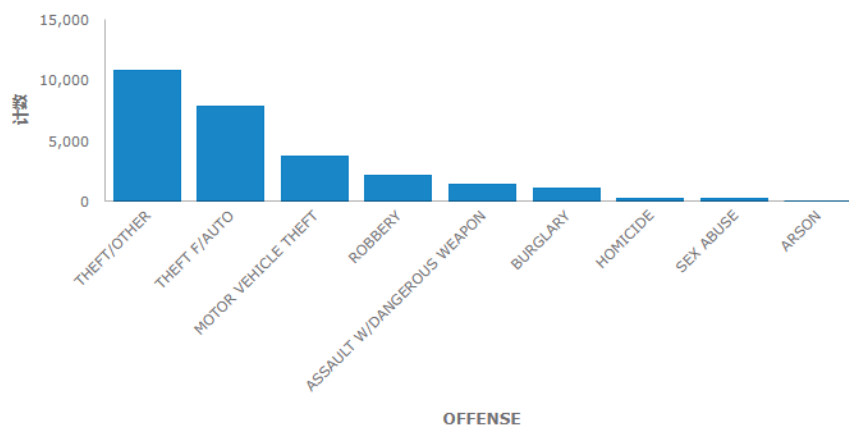
Key	Definition	Description
CCN	Criminal Complaint Number	刑事诉讼编号（警方给每个案件分配的唯一编号）
REPORT_DAT	Crime Report Date	案件被警方报道的时间（可能晚于发生时间）
SHIFT	MPD Shift	警局成员在案件报到时的值班。白班一般在早上 7 点到下午 15 点之间(军事时间)；晚上 15 点到 23 点，午夜 23 点到 7 点。如果是未知的，则该字段将显示“UNK”
METHOD	Type of weapon used to commit crime	执行犯罪时用的武器
OFFENSE	Crime Offense	犯罪行为的定义（种类）
BLOCK	Block Name	街道名
XBLOCK	Block X Coordinate	犯罪事件街区的 X 坐标(质心)，参考马里兰州平面 NAD 1983 米
YBLOCK	Block Y Coordinate	犯罪事件街区的 Y 坐标(质心)，参考马里兰州平面 NAD 1983 米
WARD	District Ward Identifier	地区分区编号
ANC	Advisory Neighborhood Commission Identifier	咨询居委会编号

DISTRICT	Police District	警区号
PSA	Police Service Areas	警察服务区号
NEIGHBORHOOD_CLUSTER	Neighborhood Cluster	社区聚类号
BLOCK_GROUP	Census Block Group	普查组号
CENSUS_TRACT	Census Tract	普查区号
VOTING_PRECINCT	Voting Precinct	投票选区
X	X Coordinate	犯罪事件的 X 坐标, 参考马里兰州平面 NAD 1983 米
Y	Y Coordinate	犯罪事件的 Y 坐标, 参考马里兰州平面 NAD 1983 米
LATITUDE	Latitude	犯罪事件发生的纬度
LONGITUDE	Longitude	犯罪事件发生的经度
BID	Business Improvement Districts	商业区号
START_DATE	Crime Start Date	犯罪开始日期
END_DATE	Crime End Date	犯罪结束日期
OBJECTID	Internal feature number	自动生成的连续唯一整数

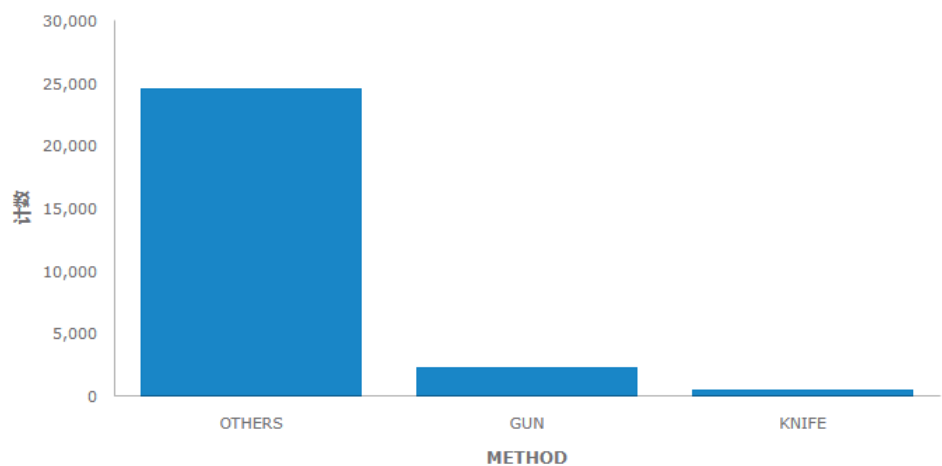
其中值得我们注意的是, 统计者将犯罪类型分为两个主要的犯罪类别: 暴力犯罪和财产犯罪。暴力犯罪包括杀人、性虐待、使用危险武器攻击和抢劫。暴力犯罪可以通过使用的武器进一步搜索: 枪支, 非枪支, 或两者兼而有之。财产犯罪包括入室盗窃、机动车盗窃、车辆盗窃、盗窃(其他)和纵火。如下表:

Category	Offense	Description
Property Crime	THEFT/OTHER	盗窃 (其他)
	THEFT F/AUTO	车辆盗窃
	MOTOR VEHICLE THEFT	机动车盗窃
	BURGLARY	入室盗窃
	ARSON	纵火
Violent Crime	ASSAULT W/DANGEROUS WEAPON	危险武器攻击
	HOMICIDE	杀人
	SEX ABUSE	性虐待
	ROBBERY	抢劫

以 2022 年的数据为例, 我们可以看到盗窃罪的数量是最多的, 纵火罪相对较少。整体而言财产犯罪远多于暴力犯罪。

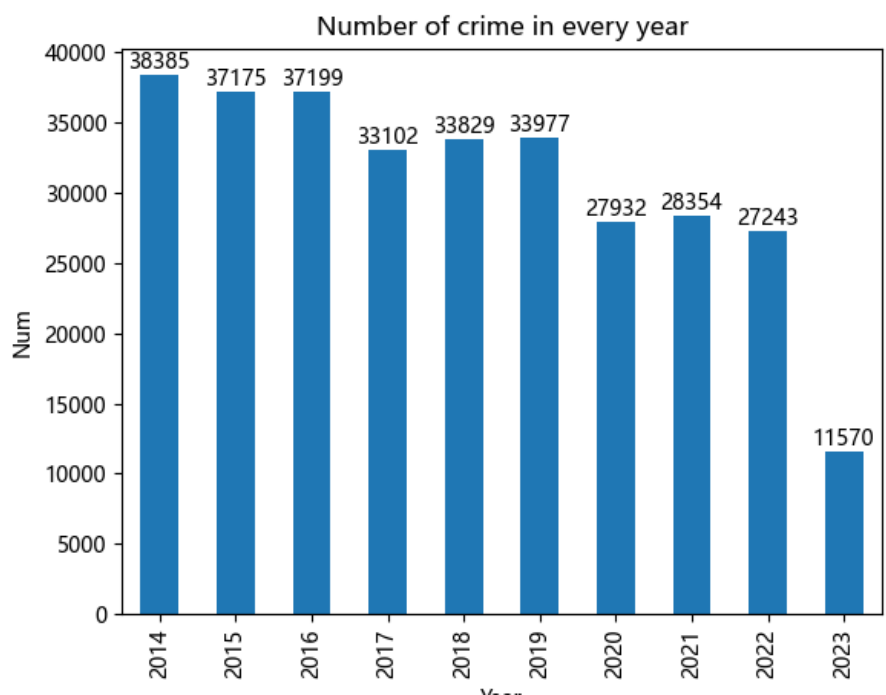


从犯罪手段来看，使用枪支和刀的还是少数，更多的还是利用其他手段。



(2) 犯罪数量随年份变化分析

为了更进一步的了解犯罪数据的时间分布情况，我还做了十年来犯罪数量随年份变化的柱状图。如下图所示：

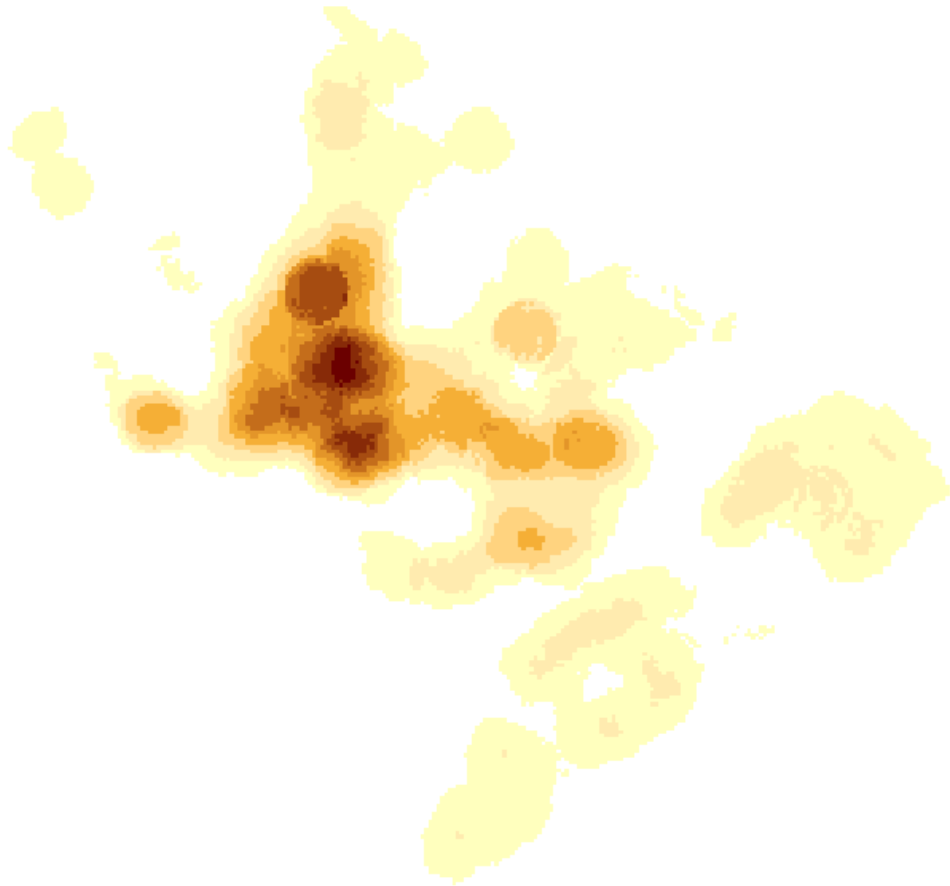


其中 2023 年由于只统计到 5 月份，所以数量较少。整体上犯罪数据数量呈下降趋势。值得注意的是在 2016-2017 和 2019-2020 有两次较为明显的下降，后者可能是由于疫情导致，前者的原因有待考究。下面的新闻其实可以验证我们分析的正确性。

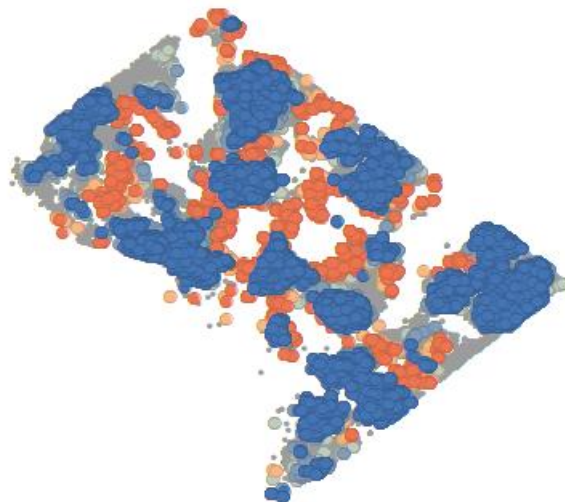
根据美国联邦调查局FBI报道，2015和2016连续两年出现犯罪率上涨升的情况，到了2017年美国各地报告了125万起暴力犯罪，每10万美国居民中有383起暴力犯罪，比2016年略有下降。

(3) 热力图分析

为了了解犯罪位置的空间分布情况，我对 2014-2023 十年来 WashingtonDC 的犯罪数据汇总起来绘制了热力图，如下图所示。通过这份热力图我们可以清晰的得到犯罪高发的区域（颜色较深的地方）。



同时我还做了热点分析，使用的是余弦距离作为指标，效果如下：、



3. 遇见问题与解决方案

主要遇到的一个问题时，在下载 WashingtonDC 的底图时，由于面积太大无论是从 Openstreetmap 上下载还是采用 Arcgis 插件都无法完成。解决方案是，后续将图层导入到 QGIS 当中直接加载底图进行可视化。

然后还有一个问题就是对于 ArcGIS 的操作不太熟，不过在阅读文档资料之后就比较简单了。