<https://open.toronto.ca/dataset/real-estate-asset-inventory/>

<https://open.toronto.ca/dataset/highrise-residential-fire-inspection-results/>

<https://open.toronto.ca/dataset/covid-19-cases-in-toronto/> \*\*\*

犯罪率是一个重要的城市安全指数，每年多伦多警局都会公布城市犯罪数据。犯罪数据中不仅包括犯罪类型，还包括受害者信息。 然而根据 the Municipal Freedom of Information and Protection of Privacy Act， 多伦多警局需要采取必要措施以保护涉案人员的隐私。因此，公布的犯罪案例往往以一种匿名的形式。 这些犯罪案例中，受害者分布则有利于提醒生活中多伦多地区的人们提高防范意识。

Carrington 对加拿大2000-2041年犯罪水平估计显示，加拿大犯罪水平与人口年龄结构相关，随着加拿大的老龄化加剧，预测所有类型的犯罪会随着下降，而青少年特有的抢劫和入室盗窃等犯罪将会下降得更快。

对1999年加拿大统计局的General Social Survey (GSS) 数据分析指出，15岁以上的加拿大人中有25% 是至少一项犯罪行为的受害者，而在1993年这个比例只有23%，出现了上升趋势。 但从1993年到1999年间，性侵犯，抢劫，殴打，入室盗窃等犯罪率变化不明显，而盗窃个人财务的犯罪率出现增加的趋势。

gaetz2010surviving 对多伦多街头青年的受害情况调查发现， 街头青年成为受害者主要是因为无家可归的青年的脆弱性，司法和庇护系统没有对街头青年生活提供有效的干预。

通过上述literature review分析我们发现关于最近多伦多地区Victims of Crimes 特征研究的文献较少。 多伦多地区作为加拿大人口最密集区域之一，且每年接纳大量的移民。对近期的Victims of Crimes 特征研究可以帮助市政府制定合理的政策以保护潜在的Victims of Crimes，从而有效的降低多伦多地区的犯罪率。

应该向读者传达对数据集的感觉。 所有变量都应进行彻底检查和解释。 解释是否存在可以使用的类似数据集以及为什么不可以。 如果构建了变量，那么应该提及这一点，并且应该提及值得注意的高级清洁方面，但是本节应该关注目的地，而不是旅程。 通过包括所有观察结果的图表（可能还包括表格）以及对这些图表和这些数据的其他特征的讨论，了解变量的样子非常重要。 还应包括摘要统计数据以及变量之间的任何关系。 如果这变得太详细，那么可以使用附录。

## 这个数据来自于Toronto Open data 数据库中的Police Annual Statistical Report。关于**Victims of Crimes**数据每年更新一次，最新的更新发生在Nov 28, 2023。 该数据库由多伦多警察局发布，具有很高的价值且数据准确。 虽然我们也可以通过GSS数据估计**Victims of Crimes， 但GSS数据可能会存在较大的误差。因此这个数据研究多伦多Victims of Crimes 是不可替代的。**

这个数据集主要针对个人受害者，原始数据包括1110 observations and 9 variables. 在data clean 过程中，首先忽略 observation ID，其次考虑category 变量所有value 都是“Crimes Against the Person”， 忽略 crimes category. 最后，我们研究主要focus 在不同犯罪类型中受害人情况，而不对assault type 进行研究，因此变量中的Assault subtype 变量也被舍去。经过data cleaning 过程，我们最终得到用于分析的数据包括1110 observations and 6 variables. The definition of these 6 variables are shown in Table 1.

在2019年之前，the number of Victims of Crimes 出现上升趋势，然而在2020年，the number of Victims of Crimes 出现明显下降，从最高点29000 跌落到24500。 随后这个数值持续增加，并在2022年达到28000。 2020 年加拿大发生COVID-19 疫情，可以作为解释the number of Victims of Crimes trend 剧烈波动的原因。

当考虑Victims of Crimes性别之后，我们发现在2019年之前， 男性Victims多于女性Victims， 且都呈现上升趋势，但是2020年之后，男性Victims 少于女性Victims，且存在上升趋势。 性别不明的Victims仅仅占有很少的一部分。

多伦多地区受害者主要是因为遭受 assault造成，约占整体受害者3/4 以上。 Assault 产生的受害者人数也是在2019年之前保持上升趋势，在2020年出现明显下降，2020 年之后继续上升。 但是其他类型的受害者人数趋势则各异，比如robbery 造成的受害者在2017年达到最高，之后2017-2021 则出现下降趋势，然而2022年robbery 造成的受害者人数出现转折，出现大幅增加。 Sexual violation 的受害者人数在2019 年之前一直低于robbery 造成的受害者人数，且出现上升趋势。但2020年之后， Sexual violation 的受害者人数 超越了robbery 造成的受害者人数。上述曲线分析可以看出，2020年是一个转折点，受害者人数和类型都发生剧烈变化。

受害者中，绝大部分都是成人，但不可忽视的是每年依然有很少部分的受害者为youth and child. Adult and Youth crimes 的人数也是在2020年drop 到最低值，并持续上升的趋势。

考虑到绝大部分受害者都是adults, 我们对年龄进行细分，并得到图5所示结果. 在所有年龄段中，25 to 34 Victims 的人数最多，其次是18-24年龄的Victims。 我们发现，受害者主要集中在18-54 年龄段。这个年龄段人处于工作阶段，往往需要更多的收入以维持生活。 分析不同年龄段趋势我们发现，在2020年以前发生后，18-24 和65+ 的Victims 下降最多，但是2020年之后，这两个年龄段的上升趋势也是最大的。 值得庆幸的是，child (<12) 的vicitms 人数整体是下降趋势，说明多伦多地区读child 保护更加有效。

通过上述数据分析，我们发现2014-2022 区间多伦多地区的Victims of Crimes整体趋势是上升的，说明多伦多地区的治安存在下降的风险。 同时我们发现，政府政策会影响犯罪率。比如在COVID-19 期间的隔离政策使得 Victims of Crimes 出现明显的下降， 而随着隔离政策的放宽，Victims of Crimes 的人数也随之上升。 同时，我们发现Victims of Crimes存在明显的年龄特征，成年人年龄在18-54岁时Victims 的高发年龄区间。 这个年龄段的人，需要进行工作以维持生计，不可避免的要接触更多的陌生人，从而导致成为Victims of Crimes的风险增加。 同时我们要意识到，2020 后Victims of Crimes 的分布和2020年之前的区别，从而多伦多政府可以及时调整政策以抑制犯罪的发生。

多伦多治安问题时关乎居民生活满意的重要指标，而受害人分布则直接反映多伦多治安问题。 本文使用多伦多警局提供的Victims of Crimes summary 数据from 2014 to 2022 研究Victims 趋势。结果显示the number of Victims 出现明显的上升趋势，且在COVID-19 发生前后the number of Victims的分布出现明显变化。 政府2020 年COVID-19 的政策虽然减少了number of Victims，但疫情过后出现的number of Victims反弹必须引起政府注意。

Some brief comments/suggestions for the final submission:

* Submission seems to be missing final pieces. Consider having an actual discussion.
* I would recommend tightening the language in the whole paper.
* The abstract could use a little bit of editing.
* Some clarity could be added around the overall aim of the paper.
* Make sure the Intro section has all the components (see rubric) — here i'm thinking also of "structure of the paper".
* Tables can have more informative captions. In table 2, cause of death is redundant. Also consider relative proportions in addition to counts to get point across
* Figure 1 could be reconsidered. For instance, apart from drug toxicity other causes of death are hard to tell apart in graph.