# Should we be more Worried about Processed Sugar than Violent Crime?

***COMP 30780 Data Science in Practice***


Marcus Gray (13378861), Evin Kierans (17451696)

School of Computer Science

University College Dublin


*07/05/2021*


**Abstract**. In this project we aim to see how the harm of processed sugar can be compared to the harm of violent crime, and how they do compare. Using diabetes as a proxy for sugar, we use various metrics to compare the harm of diabetes and violent crime. We can do this using our key dataset from the Global Health Data Exchange where we can access and compare such data on a variety of diseases and conditions. We undertake our analyses on multiple geographical levels from countries world-wide to metropolitan areas. We examine how the harm caused by diabetes and violent crime is developing over time world-wide and look for demographic groups that seem to be suffering disproportionately. We also look at regions within countries and see how the harm caused by diabetes and violent crime distributes within them. We then take on a case study on the United States. We do this by adding data that gives us information on financial expenditure, official diabetes diagnosis data, official crime rates, and population estimates. We look at whether particular state financial expenditures in the United States reduces the death and burden of diabetes and violent crime. We also go a level down geographically and explore metropolitan areas. Here we investigate whether there is a substantial change in the harm caused by violent crime relative to diabetes compared to larger areas with smaller population densities. We find that the harm caused by diabetes seems to be getting increasingly worse while the harm for violent crime seems to be reducing overall. We also find that this holds true within countries, and in fact only a very small percentage of demographic groups have a higher rate of harm by diabetes than from violent crime. We find a correlation between higher percentage spending on justice systems, and lower crime rates per population. Lastly, in metropolitan areas we see no substantial change in the death rates due to violence versus diabetes.

**Declaration**. We, Marcus Gray, 13378861 and Evin Kierans, 17451696, declare that this assignment is our own work and that we have correctly acknowledged the work of others. This assignment is in accordance with University and School guidance[1] on good academic conduct in this regard.

[1] See https://www.cs.ucd.ie/sites/default/files/cs-plagiarism-policy_august2017.pdf

# 1. Introduction

Many more people die from diabetes world-wide than those who die as victims of violent crime. This is an established fact, and it suggests that we should be more worried about dying from diabetes as opposed to dying from violent crime, and perhaps be more worried about diabetes more in general than violent crime. However, this statistic and what it suggests have a severe and obvious contrast to the societal attention given to fatal events of violent crime over that of diabetes.

There could be many reasons for this, and there it is easy to imagine that the factors and processes involved to what information society pays attention to is complex. One could argue, for example, that the reason we see more news of events of violent crime is that such events tend to be nuanced in an interesting way in their details depending on the place, and the people involved. Whereas with a disease like diabetes the story does not change so much, rather it is a story repeated over and over and despite being tragic it soon loses its novelty. Another reason is that violent conflict, even on the level of individuals, may have implications to wider society as it can indicate a struggle of power between different people and different values. Such conflicts can be representative of wider issues in society, and in severe cases can spark large collective movements of more violence.

However, there are reasons to think that this difference in attention is not so justified. Evolution has made physical conflict and danger innately interesting to homo sapiens, and it captures our attention so well that it is difficult to find a set of movies, tv programmes, or fictional books that have their plot untouched by it. In addition, whatever the cause of a greater proportion of news and stories based on violent events, it may be influencing people and large to think that this sort of threat is among the most prominent by the frequency in which it arises. Consumption of sugar is also a major factor in someone developing diabetes. But sugar is also a key ingredient in a vast number of food products and surely contributes to enormous amounts of revenue world-wide. Is public denial, marketing, and vested interests inhibiting the attention diabetes deserves?

The point is that the perspective on the magnitude of the threat of violent crime versus diabetes may be greatly distorted. Does the statistic mentioned above give this accurate perspective? Does it give a true representation of the comparative harm? Or is the reality more in line with societal attention than that and there is an aspect to the threat of violent crime that is being overlooked by this statistic? What also makes this issue interesting is that consumption of sugar, which is voluntary, is a major contributor to developing diabetes, whereas being a victim of a crime is not.

People tend to prefer to have an accurate perspective to the threats to their wellbeing. In this project we want to find out how the harm from diabetes and violent crime can be measured - in more ways than total deaths - and how they could be compared to create a more comprehensive and accurate picture..

The remainder of the report is organized as follows. In the next section, Motivations & Objectives, we clarify what we aim to achieve with this project and provide some information on previous related work done by others, as well as enumerate through our own research questions we attempt to answer our initial project title. After this, we cover our preprocessing, and data acquisition stages. That then takes us on to our actual analysis and results for all four of our research questions. We then discuss some ethical considerations we have, as well as limitations, before moving on to our conclusion and division of work.

# 2. Motivations & Objectives

Our objective for this project was to investigate whether the significance of the harm caused by diabetes was as large, and as simple, as this statistic suggested. To do this we utilise various, reputable metrics of harm over multiple geographic regions. In addition we examined the United States in a case study where we investigate how government financing has on alleviating the harm of violent crime. We aimed to discover how we could compare the harm caused by violent crime and processed sugar, and how this harm compared.

## 2.1. Background & Motivations

There were several interesting studies done looking into both the harm of sugar, as well as a person's chances of being a victim of violent crime that we found while researching this topic. Among them, we found three that stood out as more relevant than others, and we took it onto ourselves to produce a more rounded and inclusive final report than these studies.

The first of these studies was done by Harvard University, called "The Sweet Danger of Sugar". It referenced a lot of statistics to do with and relating to sugar consumption, as well as death as a result of it. However, they didn't reference any metrics such as Disability Adjusted Life Years (DALYs), or Years of Life Lost (YLLs), as a result of sugar consumption. They do mention Diabetes, but rather as a potential side effect, rather than as a means of calculating results.

The second of these studies was from the BBC - it was essentially a crime calculator. Although different, we found this inspiring, as it was handling a lot of data in a similar way we wished to. However, a severe limitation of this was that it was localised to the UK only. It also only relates to crime, nothing about Diabetes (which is to be expected of a crime calculator).

The final study we found that was interesting was one that would significantly contribute to our overall project title. "Your risk of being a crime victim likely not as high as you think", by the Legal Examiner. We found that this fact, coupled with how deadly a disability like diabetes can be, would prove to make a very interesting project.

When we discovered the Global Health Data Exchange, and the vast resources they had available, we knew we had landed onto something good. It was through this website that we discovered the inspiration for our first research question - potential metrics for measuring harm. Disability Adjusted Life Years (DALYs) and Years of Life Lost (YLLs) were very useful assets to stumble across, and they are used throughout the project.

## 2.2. Research Questions

Our research questions are ordered in terms of geographical scale. Research questions 1 and 2 have a wide scope, including many countries worldwide, while research questions 3 and 4 are both part of a case study using data from the United States. We chose the United States as they are a massive information media hub, with an abundance of statistics available for analysis on the matter.

### 2.2.1. RQ1: How can we measure and compare the harm caused by diabetes and violent crime? Are we suffering more or less over time world-wide? Are there particular sub-demographic groups that suffer disproportionately?

For our first research question, we set out looking to create a basis and a standard for us to use and continue with in our following questions. Building on this, we also elected to include brief examples of these metrics in action, through the use of time series, as well as dividing our results by sub-demographic groups.

a) The first part of the question relates to how we can measure, and compare, the harm between the two factors. Through the use of Disability Adjusted Life Years (DALYs), and Years of

Life Lost (YLLs), we examine which is more appropriate for portraying what kind of information.

b) The second part of the question relates to time, and the effects on a world-wide level. Here, we look at several time series that depict interesting narratives of how each of these two factors (diabetes and violent crime) have changed over time, and which is causing more harm.

c) The final part of this question refers to sub-demographic groups, meaning different people in different places and different ages. We meticulously examine multiple age ranges and locations with relation to the harm caused by both violent crime and diabetes.

#### 2.2.2. RQ2: How does the harm of diabetes versus violent crime distribute within countries?

a) What features are present in demographic groups that suffer greater harm from interpersonal violence than diabetes?

In our second research question we continue our analysis to regions within countries. Here we look at how the harm of diabetes and violent crime distributes within seven different countries. These include the United States, Japan, Kenya, Indonesia, Mexico, Brazil, and the United Kingdom.
We then explore what features are common across the demographic groups with the highest deaths by violent crime, with a closer look at Brazil in particular.

#### 2.2.3. RQ3: How effective is government financing at reducing the death and burden caused by interpersonal violence and diabetes on the population?

For this research question, we aim to examine what effects government action, through financial expenditure, have on the crime rates, and if there is a correlation between this and the size of the population being examined. For this question, we decided to do a case study specifically on the United States, breaking it down on a State-by-State level, including the District of Columbia too.

#### 2.2.4. RQ4: Do high-crime areas reveal a substantial difference in the ratio of death rates from diabetes versus violence?

In our last research question we continue with the United States and we go a level down geographically. Looking at Metropolitan Statistical Areas and see how the deaths from violent crime changes to diabetes in smaller areas that have greater population densities.

# 3. Data Wrangling

We collected data from a variety of sources. There was a stark difference in how much work was needed for data preparation. We will go over the key datasets we used in this project, and the challenging aspects to preparing this data for analysis.

## 3.1. Data Acquisition

Our primary dataset has come from the Global Health Data Exchange (GHDx). According to the GHDx website;

"The Institute for Health Metrics and Evaluation (IHME) is an independent global health research center at the University of Washington. The Global Health Data Exchange (GHDx) is a data catalog created and supported by IHME." [2] (GHDx , 2021).

[2] See http://ghdx.healthdata.org/about-ghdx

This dataset allowed us to measure the harm of both diabetes and violent crime, worldwide and on a within-country-level basis, through the use of metrics such as deaths, Years of Life Lost, and Disability-Adjusted Life Years. These metrics quantify the burden of disabilities, diseases and deaths.

Data can be obtained from the organisations website where they have a data collection tool that generates and sends a user a dataset via email. A user has a selection of parameters they can adjust to create the dataset desired. In practice a generated dataset is likely to be delivered as multiple datasets of the same format, which the user has to concatenate themselves.

We also collected data from the Federal Bureau of Investigation[3]. This government agency has a host of datasets regarding crime rates in the US on its official website available to download. It is difficult to think where one may get more accurate data relating to crime across multiple states in the US than from this agency.

We also collected diabetes diagnosis percentage data from the Centers for Disease Control and Prevention for the states[4]. This had diagnosis percentages for the US not only for states, but for the counties within them as well. The latest data we could get from this source was from the year 2018.

We obtain data from the US Census Bureau[5] based on financial expenditure of states (and smaller territories in the United States).

To calculate accurate population density we again obtain data from the US Census Bureau on population estimates and land area. The US census is carried out only once every 10 years (the last census with publicly available information was taken in 2010). As our diabetes diagnosis dataset is based on 2018, the population density measures were about 8 years out of date. Since we had the population estimates we decided to use that to get a population density estimate as well.

Lastly, we collected a dataset that allowed us to construct a choropleth map.

## 3.2. Data Cleaning & Preparation

The complexity of preprocessing required for our research questions varied considerably. For research question 3 and 4 in particular, they were very fastidious when it came to data preparation.

### 3.2.1. **RQ1**.

Processing the data for RQ1 wasn't strenuous by any means, however, it was tedious. To begin with, the data collection had to be done piece by piece, as a result of limitations of the Global Health Data Exchange website. The only notable part once all the data was downloaded, was that they had to be concatenated, a fairly straightforward procedure. However, due to the various datasets being massive, it does take some time to concatenate them into the combined CSV.

### 3.2.2. **RQ2**.

Data preparation for RQ2 was similar to RQ1 given that the dataset came from the GHDx data tool. Simple actions such as removing rows with kidney disease and eliminating some unneeded columns was necessary. Data on the UK had to be done slightly differently, however, as there were some regions included that overlapped. Rows containing other metrics, such as incidence, were also excluded from the cleaned dataset.

---

[3] See: https://www.fbi.gov/services/cjis/ucr (FBI, n.d.)
[4] See: https://www.cdc.gov/diabetes/data/index.html (Centers for Disease Control and Protection, 2019)
[5] See: https://www.bjs.gov/index.cfm?ty=tp&tid=161 (Bureau of Justice Statistics, n.d.)

#### 3.2.3. **RQ3**.

This research question was very tedious in terms of preprocessing. The data was primarily from the Bureau of Justice Statistics. It seemed to be made for an excel spreadsheet, and as such there were many commas and comments to remove. In the end, it was nearly unrecognisable from what it once was, but it was usable.

Additionally, we used a dataset from the CDC (Centre for Disease Control and Prevention), called the Diabetes Atlas. Using this, we could get state-by-state knowledge on the diabetes rate. And finally, for the violent_crime_by_state_df.csv file, I made an adjustment and added in a CrimePerPop column - a missing metric that now allowed us to examine the crime rate of the population.

#### 3.2.4. **RQ4**.

The population estimates and land area datasets had to be joined by their respective states, counties, and cities to allow population densities for these areas to be calculated. To calculate population density correctly we referred to the definition given by the US Census Bureau on population density;

"Total population within a geographic entity (for example, United States, state, county, place) divided by the land area of that entity measured in square kilometers or square miles. Density is expressed as both "people per square kilometer" and "people per square mile" of land area."

The preprocessing for the MSA Crime dataset proved challenging. A snippet of this dataset can be seen in the table below (Table 1). As you can see, this code seems to be optimised for user readability than for data science. Our first goal with this dataset is to isolate all the counties and cities that make up each Metropolitan area. We then have to join each area with the land area from our land dataset to calculate the overall population density. We also have to do the same thing for the diagnosed diabetes percentage of that area from our DiabetesAtlas data.

| **Crime in the United States** | | | | |
|---|---|---|---|---|
| by Metropolitan Statistical Area, 2018 | | | | |
| Metropolitan Statistical Area | Counties/principal cities | Population | **Violent crime** | Murder and nonnegligent manslaughter |
| **Abilene, TX M.S.A.**[2] | | **170,417** | | |
| | Includes Callahan, Jones, and Taylor Counties | | | |
| | City of Abilene | 122,480 | 591 | 8 |
| | Total area actually reporting | 100.0% | 670 | 9 |
| | Rate per 100,000 inhabitants | | 393.2 | 5.3 |
| **Akron, OH M.S.A.**[3] | | **704,283** | | |
| | Includes Portage[3] and Summit Counties | | | |
| | City of Akron | 197,690 | 1,704 | 38 |
| | Total area actually reporting | 96.2% | 2,509 | 45 |
| | Estimated total | 100.0% | 2,547 | 45 |
| | Rate per 100,000 inhabitants | | 361.6 | 6.4 |

*Table 1. Sample of CrimeInUSbyMSA dataset*

This task was more complicated than it first appeared. In the 'Counties/principle cities' column we see that there are cells that begin with the string 'Includes' and then a list of counties. It appeared that these row cells would correspond to the counties that made up a Metropolitan area. However, it turns out that there are some cities that make up metropolitan areas in the United States that are independent in their own right, and not part of any counties. In order to be identified correctly when joining with the other datasets, each name had to be superseded by the correct term - either county or city. It became clear that there would have to be careful string processing. For example in the table cells below (Table 2), we see that there is a list of counties, then followed by a list of cities. The other related cells followed this format. Each county had to be followed by 'County' and each city had to be followed by 'City'.

| | |
|---|---|
| **Richmond, VA M.S.A.** | Includes Amelia, Charles City, Chesterfield, Dinwiddie, Goochland, Hanover, Henrico, King and Queen, King William, New Kent, Powhatan, Prince George, and Sussex Counties and Colonial Heights, Hopewell, Petersburg, and Richmond Cities |

*Table 2. Example of a MSA including counties and independent cities*

To further complicate matters - it turns out that some of the metropolitan areas don't exist just in one state as one may expect, but multiple. And we learned that we could not just rely on the county name as many states have counties with the same names. This meant that we had to include the state code as part of the identifier of the county or city (Table 3).

| | |
|---|---|
| **Memphis, TN-MS-AR M.S.A.**[2, 3] | Includes Crittenden County, AR; DeSoto, Marshall, Tate, and Tunica[3] Counties, MS; and Fayette, Shelby, and Tipton Counties, TN |

*Table 3. Example of a MSA residing in multiple states*

This led to 4 separate categories to process: single state MSA counties, single state MSA cities, multi-state MSA counties, and multi-state MSA cities. When this was done all counties and cities could be joined with their respective diagnosis percentages, population totals, population densities, and homicide numbers.

The American data and diabetes data were straightforward to process. The Crime by State dataset was a similar format to the Crime by MSA but simpler and easier to do after processing the MSA dataset.

# 4. Data Analysis & Results

Below we explain our analysis for each of our research questions followed by the results obtained. Before we do so we need to clearly explain some of the metrics we use in our analysis. There are different ways to measure the burden of disease. Included in our research questions is the use of metrics such as Years of Life Lost (YLL) and Disability Adjusted Life Years (DALY). Years of Life Lost is simply the estimation of premature death given the age and life expectancy of the deceased person. A DALY is an attempt to measure suffering caused by living with a condition and the years of life lost combined. According to the WHO;

"One DALY represents the loss of the equivalent of one year of full health. DALYs for a disease or health condition are the sum of the years of life lost due to premature mortality (YLLs) and the years lived with a disability (YLDs) due to prevalent cases of the disease or health condition in a population[6]." (WHO, 2021)

[6] See: https://www.who.int/data/gho/indicator-metadata-registry/imr-details/158

## 4.1. RQ1. How can we measure and compare the harm caused by diabetes and violent crime? Are we suffering more or less over time world-wide? Are there particular sub-demographic groups that suffer disproportionately?

In order to do any sort of study or comparison into the harm of sugar consumption versus that of violent crime, one must first establish a metric for harm. Thus, in this research question, we aim to establish and examine which metrics are most appropriate for measuring both violent crime and sugar consumption, through Diabetes as a medium for the latter. In particular, there are three metrics we will examine; DALYs (Disability Adjusted Life Years), YLLs (Years of Life Lost), and Deaths.

The second part of this question pertains to an across-time assessment of how things have changed. In order to do this, we needed to pick different demographic groups to examine. We chose to compare men of age 20-24, and compare them to men of age 60-64, expecting to see diabetes be less dangerous than violent crime in the first group, and vice versa in the second.

### 4.1.1. Datasets

For this research question, most of the data we required was readily available from the Global Health Data Exchange website, split into various snippets for each country. Each snippet had extensive information pertaining to DALYs, YLLs, Deaths, Diabetes, Interpersonal Violence, years, age, sex and location. For the most part, we only use values in the dataset that are 10-15 years old, which is actually only a portion of the data, as it goes right back as far as 1990.

### 4.1.2. Approach

For us to accurately decide which metrics were applicable for our project, we must make a comparison between them. So that's what we did; we looked at various locations, notably Ireland, and compared the measured harm caused by both diabetes and interpersonal violence, side by side, on annually based bar charts, with each different metric (YLLs, DALYs, Deaths).

Once this was established, we went for a deeper look at various sub-demographic groups, examining their effect over time. To do this, we chose various age groups, genders, and locations and compared results.

Finally, we looked deeper into the various sub-demographic groups that we used to observe the effect over time from part 2. Our findings in this section were expected to be a predictable distribution across the various ages, for all of Deaths, YLLs, and DALYs. However, some of our results here may be surprising.

### 4.1.3. Results

**Part 1:** For the first part of this research question, we set out to answer the question of which metric, DALYs, YLLs, or Deaths, we should use for the majority of this project. To be brief, there was no clear answer, and as with everything, more is better. We found that, depending on the context, it was usually prudent to include more than one metric, especially when looking at a metric that's difficult to measure, such as harm. To show this simply, we created the same visualisation, 3 times, one for each metric.

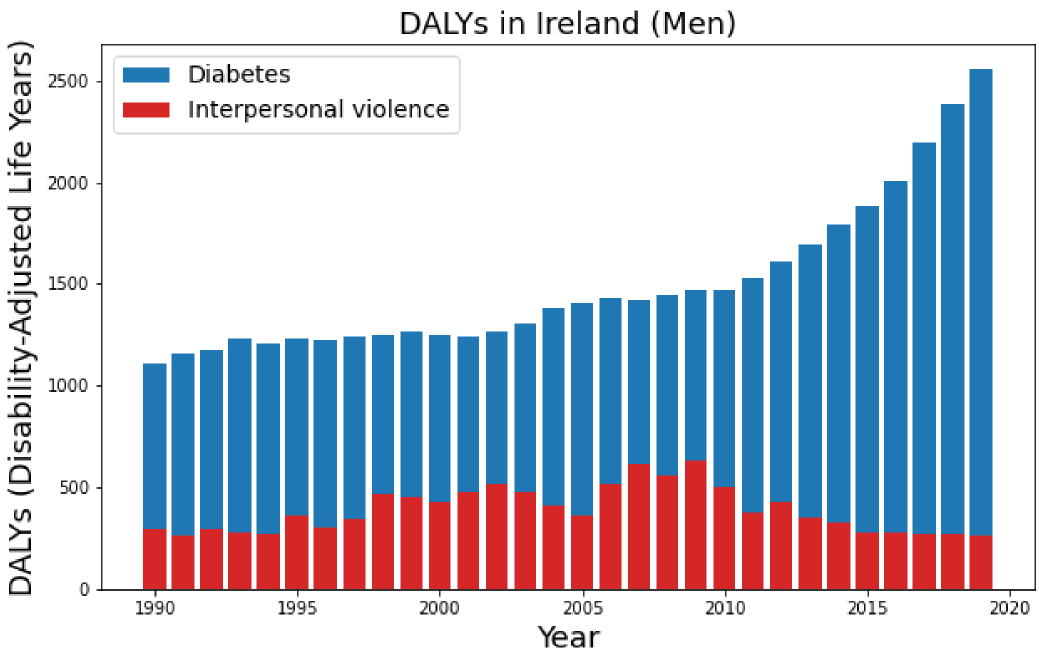


*Figure 1 DALYs of men in Ireland (years)*

As you can see from this visualisation, when we look at DALYs of men in Ireland as a result of both diabetes and interpersonal violence, we can see a clear trend of decrease in terms of interpersonal violence's harm, and an increase in that of diabetes. However, when we look at something that would be more widely recognised, death, this puts a temporary haze on the matter.

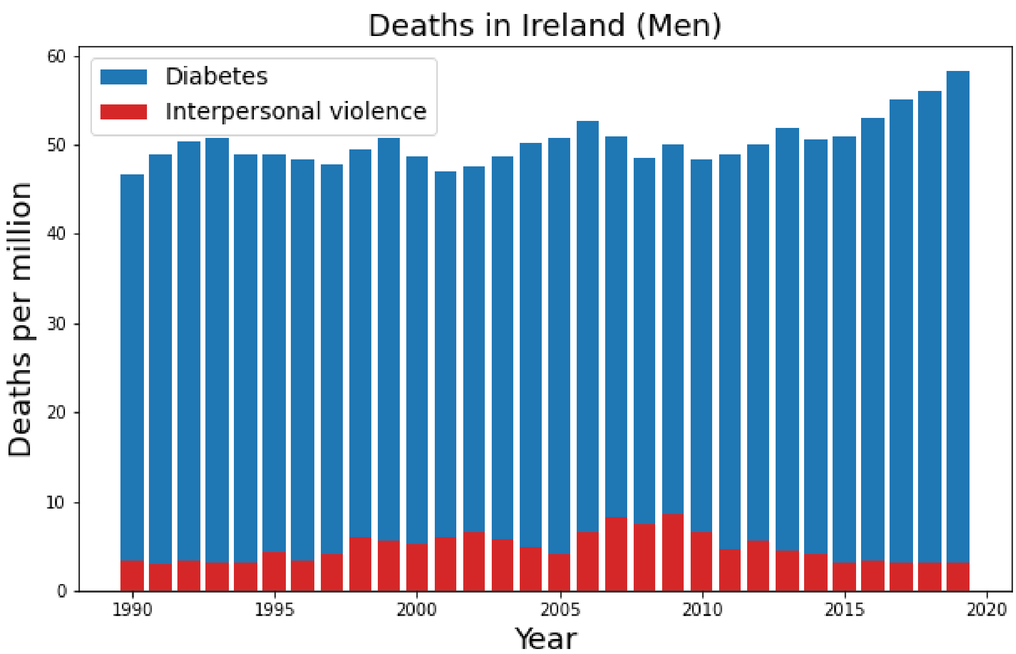


*Figure 2 Deaths from Diabetes vs Interpersonal Violence in Ireland (years)*

This visualisation paints a slightly different picture. It's quite apparent, that at least in terms of death, the numbers have remained relatively similar across the board, within an expected margin of variance. It is fair to say that when we only take in the last 10 years, there is certainly a downward trend in interpersonal violence. It is just not as prominent as it would be if we used DALYs to show it. We also have to remember that these are two different metrics, one is measuring the actual harm as a result of disability, the other is a death toll, so a difference is absolutely expected.

While this may also be the case for diabetes, diabetes affects people before just instant death. Because of this, we can conclude via the two of these visualisations of the metrics that not only are more people dying from diabetes, but more people are being affected overall by diabetes. Thus it wouldn't be illogical to conclude that more people are being diagnosed with diabetes in general.

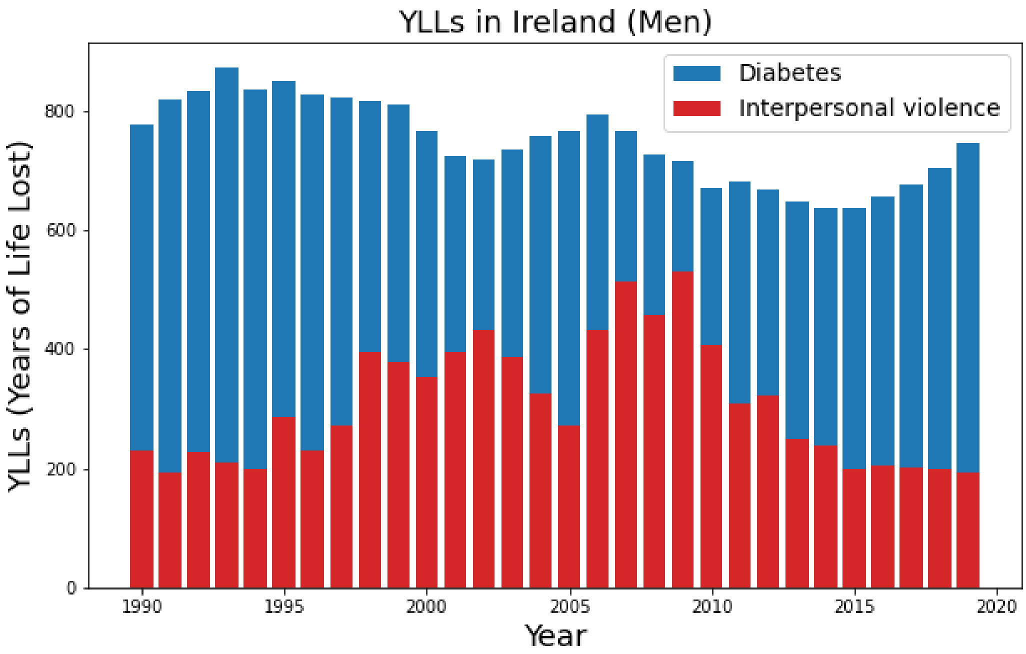


*Figure 3 YLLs of Men in Ireland (years)*

To finish out the first part, we look at YLLs (years of life lost). This is, essentially, a combination of the two metrics of Death and DALYs. However, against expectations, this metric showed another perspective. It showed, as clear as day, a fall in the years taken from people affected by violent crime, but it also puts a new perspective on the effects of diabetes. This visualisation in relation to diabetes, suggests that despite the rising diagnostics and deaths, less years of life have been lost as a result of it. This suggests that we have maybe gotten better at assisting people suffering from diabetes over time, or that we've at least made it easier for people with diabetes to manage their lifestyles through various technological advantages.

**Part 2:** Moving on to the second part of this research question, how have these changed over time? Well, we've touched on it briefly in the previous section, but we'll take a deeper look into diabetes here. We won't focus too much on interpersonal violence for this, as it's quite self-explanatory that over time, crime is falling as the quality of life and life expectancy rises.

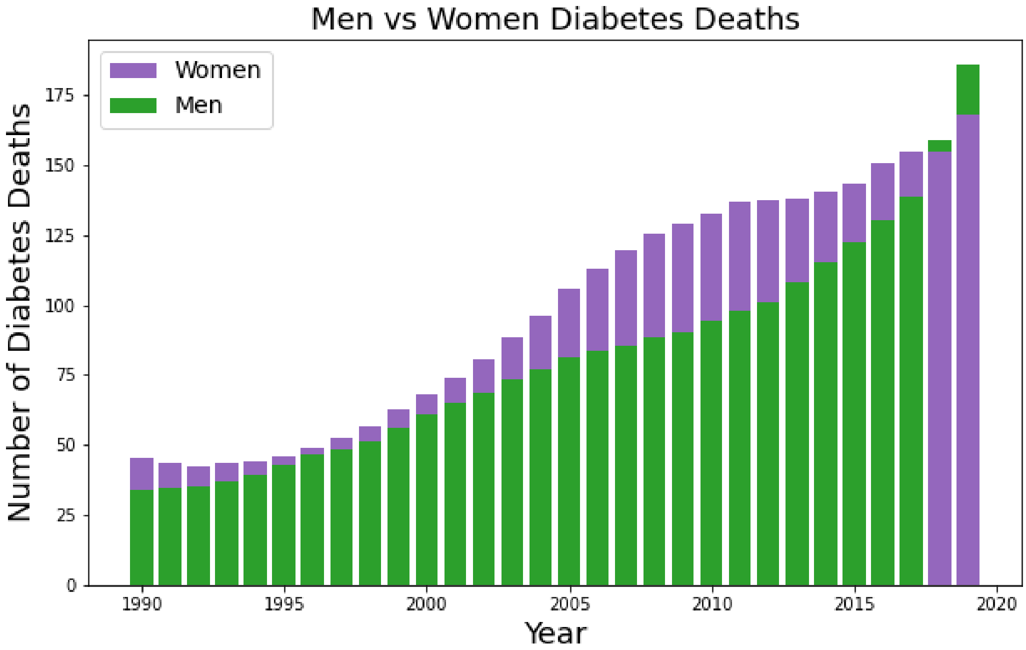


*Figure 4 Men vs Women death from Diabetes (years)*

This visualisation was a bit of a head scratcher at first. How could women, for nearly 20 years, consistently be dying in higher numbers than men from diabetes, only to suddenly… not be? It wasn't until we looked into it further that we found that older men are more likely to suffer from diabetes. This is due to men having a higher chance of developing hyperglycemia as they get older. This, coupled with the fact that life expectancy is constantly lengthening, suggests that there must be more older-men now than there were in 2017, and thus more men developing hyperglycemia and Diabetes.

Another possible explanation for the overall rise in diabetes worldwide is due to the abundance of misinformation that came from the USA in the 50's and 60's regarding sugar. They typically painted fat as the main threat, while raking in money from all-high sugar sales. This undoubtedly had a massive impact on the older age categories, who would have been the youth of back then.

**Part 3:** For part 3 of RQ1, we look at different sub-demographic groups, with respect to not only time, but also the various metrics we discussed earlier. The goal here was to examine whether there were sub-demographic groups that were at a higher risk of either diabetes or violent crime than others.

To begin with, let's look specifically at diabetes in regards to ages for the year of 2018 specifically.

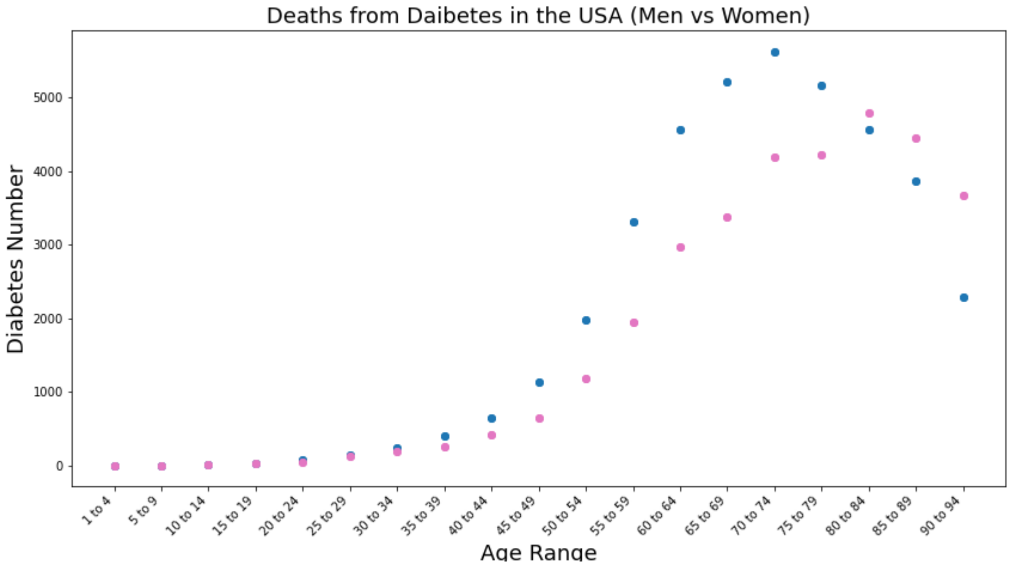


*Figure 5 Death from Diabetes in the USA (Men vs Women, by Age)*

As is quite apparent in this visualisation, men of age 70 to 74 are, by a considerable margin - likely one of the big contributing factors to men having the higher death rate from diabetes in 2018. As mentioned previously, this is likely a combination of the misinformation regarding sugar consumption that used to be a lot more abundant than it is today, mixed with the increased life expectancy. There are a few things of note in this visualisation that are of particular interest. For example, when comparing the deaths from diabetes in people over the age of 40, there is a massive difference between that number and the number below 40. So this is something to keep in mind as we look at more age-related visualisations moving forward.

Next, we will look at the comparison between interpersonal violence and diabetes at some specific age categories over time, specifically in the USA, the english-speaking world's largest source of media news.

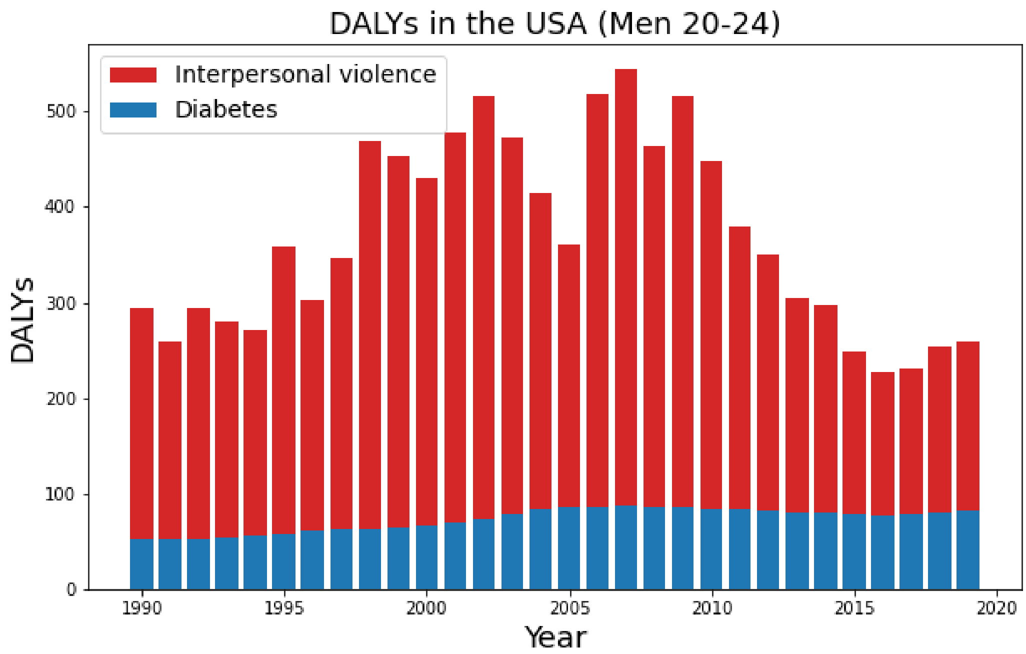


*Figure 6 DALYs of men in the USA (age 20-24)*

What we see here is a volatile variance for interpersonal violence, spiking in the early and mid 2000's, while DALYs from diabetes victims remain very steady, with only a slight incline overall. Dominantly, this visualisation is all interpersonal violence, with diabetes being a very small portion.

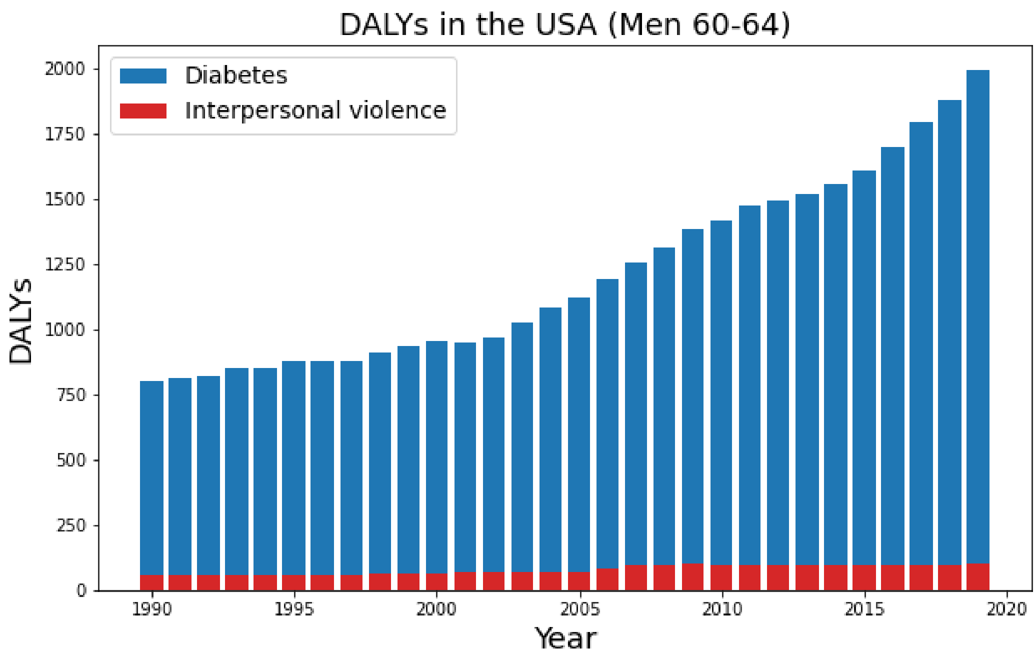


*Figure 7 DALYs of men in the USA (age 60-64)*

This flips entirely when we take an older age category, such as that of men between the ages of 60 and 64. Here, we see interpersonal violence being a largely insignificant inconvenience when compared to diabetes. Furthermore, we see a very clean-cut steady climb in the DALYs of diabetes for people in this age category. It is incredibly interesting that the climb in DALYs for diabetes is so consistent here, with only 1 or two years counting less than their previous, and even those that do, by a miniscule margin, and not at all in the last 15 years of the dataset.

When we look at DALYs for diabetes in women, across various age groups over the years, we see a relatively similar story:

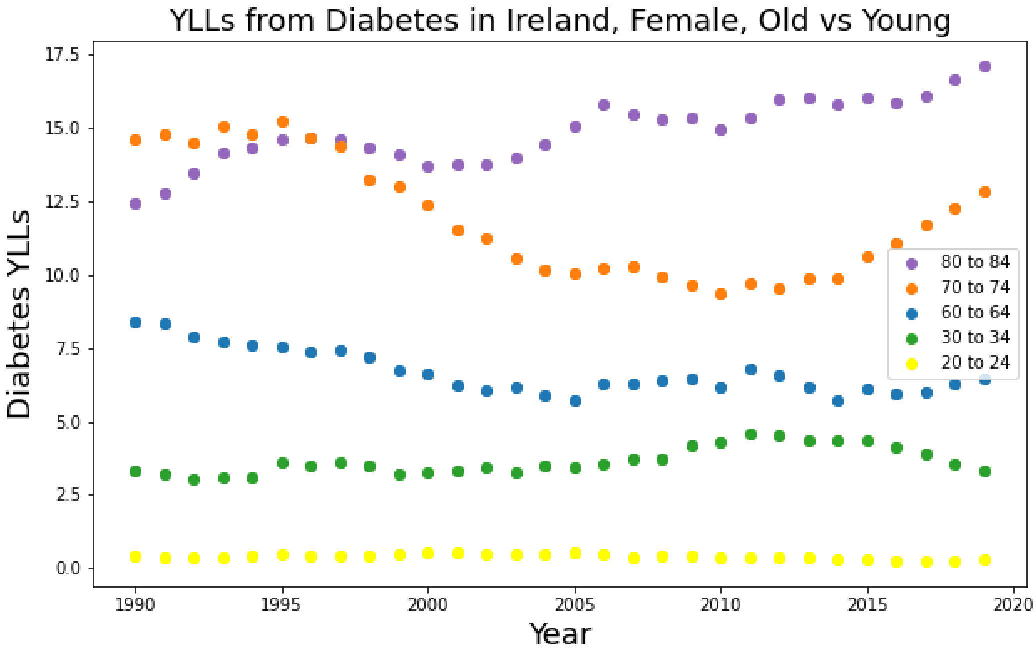


*Figure 8 DALYs from Diabetes from women in Ireland (years)*

What's interesting about this visualisation, is that we see that the age category of women that are most affected in terms of DALYs are that of 40-45. Doctors in Ireland normally start to look for signs of naturally developing diabetes in women at that age. This is partially due to the menopause, and the rise of blood sugars that coincide with this. Women of this age would have higher DALYs as a result due to these changes, and having to get used to them.

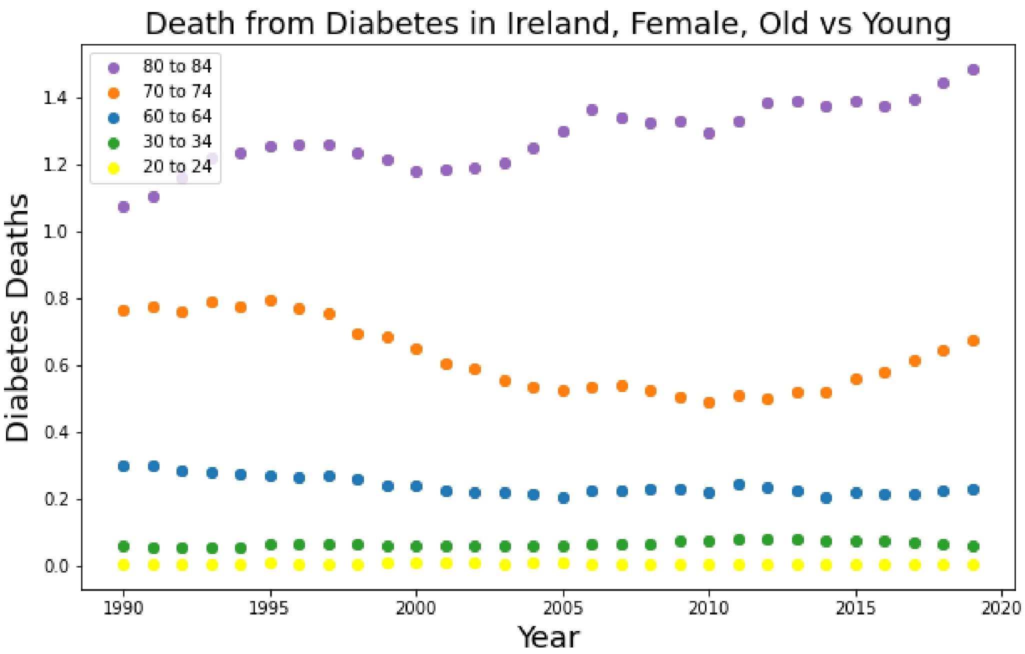


*Figure 9 Deaths from Diabetes from women in Ireland (years)*

When we look at deaths, however, we see a more expected outcome. This is a common occurrence when we use death as a metric, so a lot of the visualisations for it are quite self-explanatory. One thing worth pointing out is the correlation in each age's rise and fall compared to other ages of that year. The rises and falls coincide with each other, even across massively different age ranges, such as the rise of 80 to 84 and 60 to 64 in the early 2010's.

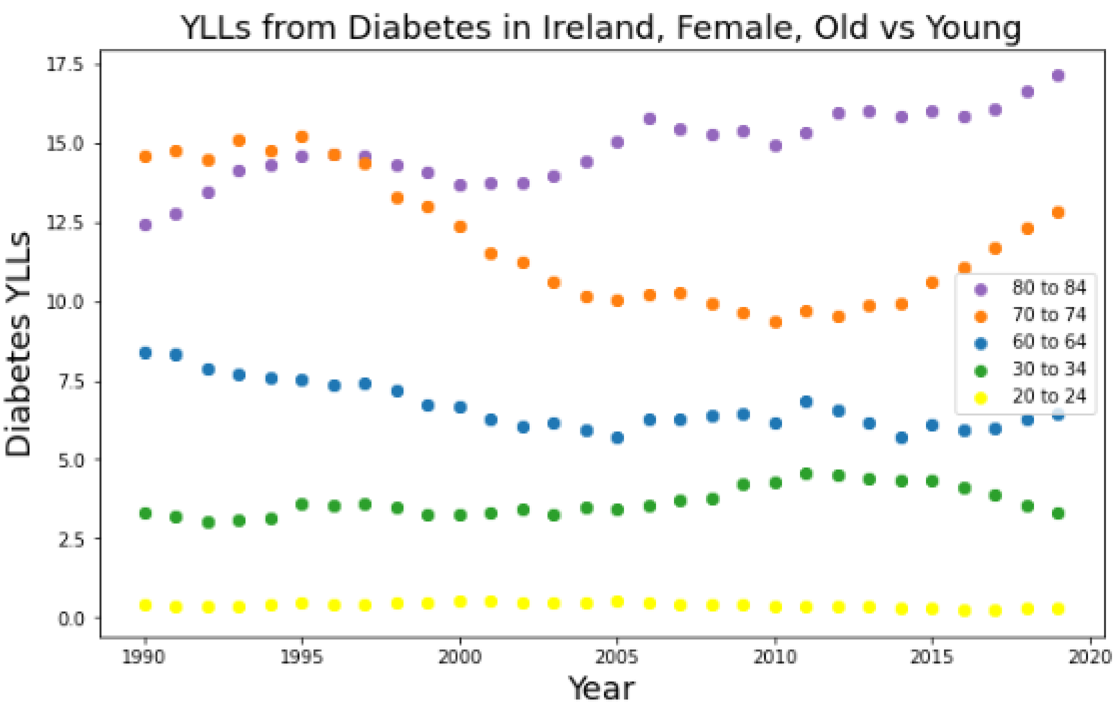


*Figure 10 YLLs from Diabetes from women in Ireland (years)*

YLLs produce mostly predictable results too, but with a couple of very interesting pieces of note. If we look at the change over time of the category of 70 to 74, we can see a fall in the years of life lost (up until the last few years, at least). At the beginning, women in their 70's were losing more years of their life than women in their 80's. A weird thing to think about, until you realise that there were likely a lot more women in their 70's than in their 80's back then. When looking at such a large time period, we must keep life expectancy in mind, especially since life expectancy in Ireland only broke 80 in the early 2000's.

#### 4.1.4. Discussion

We found some very interesting statistics throughout this research question. Most notably, the sudden change in the number of diabetes deaths worldwide when compared by gender, and how, in the last 2 years of the dataset, seemingly inexplicably, men have taken the lead and by a considerable margin. We hypothesize that this is due to an increase in expected life worldwide, combined with the fact that the older men get, the more likely they are to be diagnosed with diabetes, more so than women. When coupling these two facts together, we get a larger population of elderly men, who are at a higher risk of diabetes, and thus likely dying from it too.

Another interesting discussion point in this research question pertains to the DALYs of men with diabetes in Ireland. This discovery leads to what is very likely a flaw when discussing DALYs with regards to ages specifically. When we look at that visualisation, we can see that men between the ages of 30 to 34 and men between the ages of 80 to 84, collectively, have a remarkably similar number of DALYs. This is misleading because of sample size. There are a lot more men between the ages of 30 and 34 in Ireland than that of 80 to 84. This would, by that logic, point to men between the ages of 30 and 34 having a much higher DALYs count than those of 80 to 84. So why are they the same?

Well that's where the nature of DALYs comes into play - men between the ages of 80 to 84, although fewer in numbers, are suffering on an individual basis a lot more than that of men between 30 and 34. Purely by coincidence, the metric of DALYs puts these two sub-demographic groups neck and neck, despite so many factors separating them. Hence why we concluded that DALYs, Deaths and YLLs are not very useful by themselves, and are best used in tandem with at least one other.

## 4.2. RQ2. How does the harm of diabetes vs violent crime distribute within countries?

- **What features are present in demographic groups that suffer greater harm from interpersonal violence than diabetes?**

Following from RQ1, we now look to regions within countries themselves. Data used in this analysis involves regions within seven different countries. These include the United States, Japan, Kenya, Brazil, Mexico, Indonesia, and the United Kingdom. These countries are diverse in terms of their global location, income levels, and total deaths due to violence and diabetes. We also extend our analysis by exploring what features are common to the demographic groups that have high death rates due to violence.

### 1. Datasets

The dataset for this research question was obtained by the GHDx data tool. It contains data on interpersonal violence and diabetes mellitus on the regions within our selected countries for every year from 2010 to 2019.

We also used map data to help construct a choropleth map of Brazil.

## Approach

We use three main metrics in our analysis for this research question: deaths, Years of Life Lost (YLLs), and Disability Adjusted Life Years (DALYs). To compare these measurements between countries and regions in a meaningful way we use each of these three measurements in rates per 100,000.

We began our exploratory analysis by looking at the distribution of the burden of diabetes and violence for regions within each country based on the metrics mentioned above. We compare these distributions for all regions in each country as a whole and also for demographic groups based on age and gender. We made similar analysis for regions within each country in turn. This process continued until we had a comparison of distributions that seemed to provide an important insight to the project objective.

In addition, we made various calculations that let us compare how our different metrics of burden compared for the harm of diabetes and violent crime. For example, we could calculate how many years of life the average victim of diabetes versus violent crime could expect to lose. With these numbers, we can then compare the expected years of lost life for victims of diabetes versus violent crime. Calculations like these help us to understand how to more accurately interpret our results using our chosen metrics.

We also examine the demographic groups that have a higher death rate due violence and we see what features are common among them. We select this high death-by-violence group by finding the demographic groups, split on age, gender, region, and year for over a ten year period, that have a higher death rate due to violence than the average death rate due to diabetes across all countries. We can then look at what features common with these demographic groups and see how consistently regions have sustained such groups through time.

We look further into the country with the greatest number of high death-by-violence demographic groups out of all seven countries. We then calculate these groups again for this country by selecting those that have a higher death rate due to violence than its national diabetes death rate average, and see how these demographic groups distribute within the regions of the country. We look at the frequency a region has sustained at least one of these high death-by-violence demographic groups over a ten year period.

## Results

We find that the average victim of diabetes can expect to lose just over 13 years of lost life. In contrast, an average victim of violent crime can expect to lose over 42 years of lost life. From these measurements we can see that the average victim of violence can expect to lose over three times years of life lost more than a victim of diabetes. If we run these calculations for each country we see that in each case there are significantly more YLLs that individual victims can expect from violence over diabetes. In Indonesia victims of violent crime can expect to lose 2.6 times the years of life lost to victims of diabetes, the lowest of all seven countries. In the United Kingdom victims of violent crime can expect to lose just over 4 times the years of life lost compared to the average victim of diabetes, the highest out of all the countries.

In the graph below (Figure 11), we can see a disparity between the YLL of violence and diabetes, with diabetes contributing to a greater loss. This general pattern holds for all the countries in this analysis.

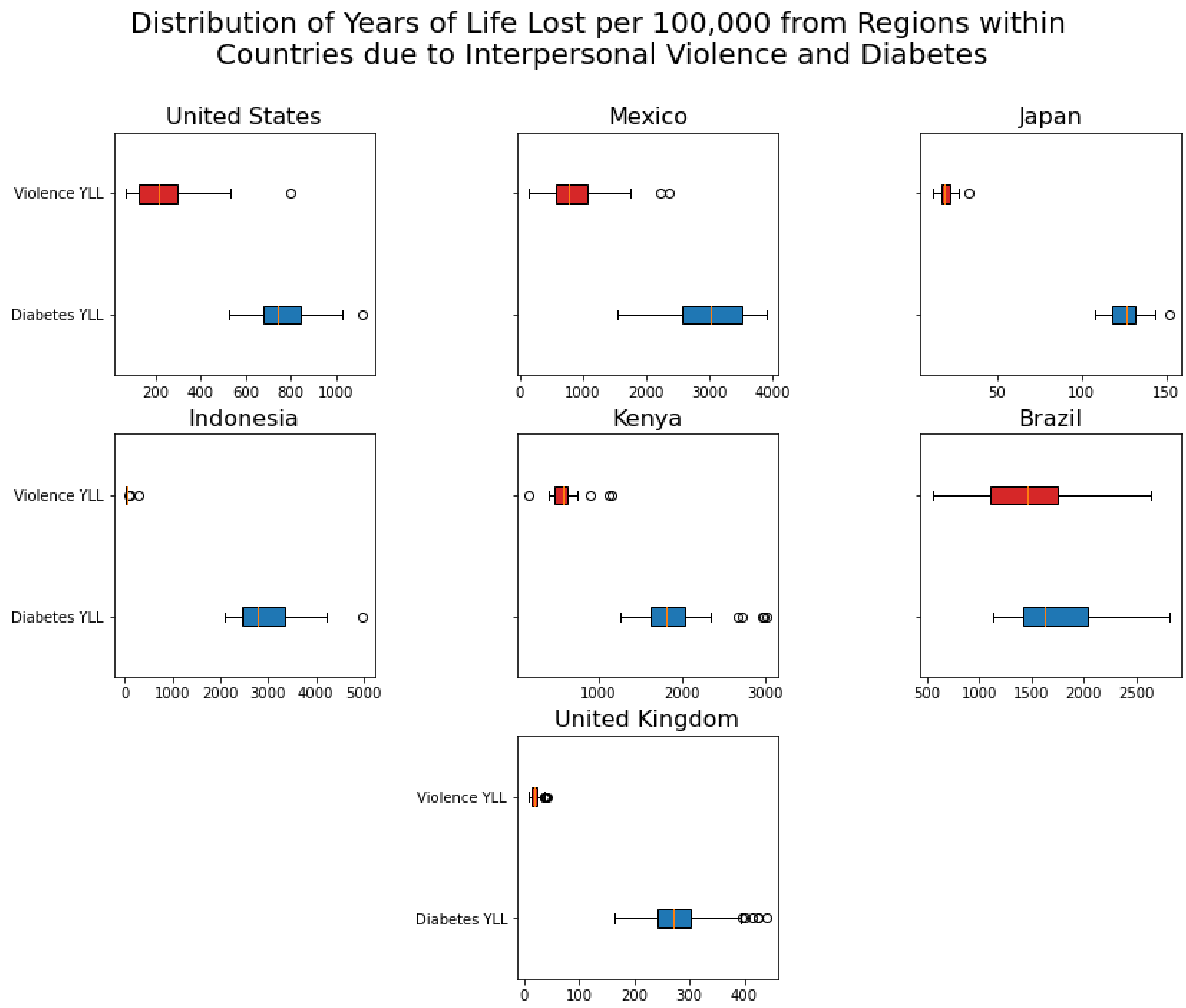


*Figure 11 Distribution of YLLs between regions in countries*

Furthermore, we found that only 0.6 percent of demographic groups based on gender, age, and region, for each year over a ten year period, had a higher death rate from violent crime than from the average diabetes death rate. Most of these demographic groups reside in Brazil.

It was at this point we decided to extend this analysis by looking at the features common to these demographic groups that had these high rates of deaths due to violence. In terms of gender, we find that all high violence demographic groups are male. In terms of age, we see a sharp rise after age 14, a peak at ages 20-24, and then a sharp decline from age 34. Location, however, is a bit more complicated.

We find that there is a greatly uneven distribution of these demographic groups between our countries. Three of our seven countries, the UK, Japan, and Indonesia, do not have any demographic groups that have a higher death rate due to violence than the overall diabetes average. In the remaining four countries we again see an uneven distribution of the high death-by-violence demographic groups. Most of these reside in a single country, Brazil, which holds 1123 out of the total 1827 of these groups. Mexico holds 664, Kenya holds 37, and the United States only holds three. Discovering this large concentration of high death-by-violence demographic groups in Brazil is what prompted us to look at this country more closely.

In the states of Brazil, we find that, when we recalculate its high death-by-violence demographic groups by using the country's own diabetes average death rate, if a region has at some stage sustained a high death-by-violence demographic group, it is likely to do so across time. This is shown in the graph below (Figure 12).

It's important to note that this does not take into account the density of high death-by-violence demographic groups in each region. i.e. how many high death-by-violence demographic groups are present in the region at one time.

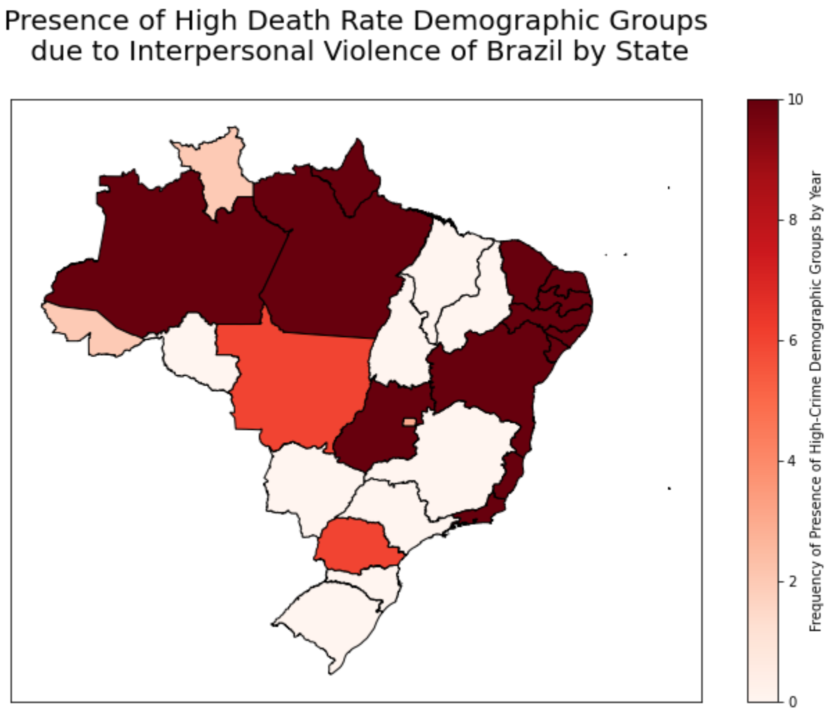


*Figure 12 Frequency of presence of high death-by-violence demographic groups in Brazil*

As we can see, the high death-by-violence demographic groups do not distribute evenly across the country. In the dark shaded areas we see regions that have had at least one such group each year for the past 10 years. The blank regions, however, have not sustained a high death-by-violence demographic groups during this time period.

Despite Brazil having many more high death-by-violence demographic groups when taken with all countries collectively, when we calculate high death-by-violence demographic groups using Brazil's own diabetes average we see that the high death-by-violence demographic groups only make up 1.98% of Brazil's total demographic groups over the 10 year period.

If we come back to all seven countries collectively again, we find out of 394 total regions within all countries only 49 have sustained a high death-by-violence demographic group at least once in the 10 year period. Furthermore, we calculated that the frequency of a region within a country that sustains such a group has an average frequency of 7.79. This means that if any region has sustained a high death-by-violence demographic group from our seven countries, we can expect it to sustain this kind of demographic group for 7-8 years.

## Discussion

The distribution of YLL on regions within countries seems to be the most relevant finding in terms of the project objective. The average victim of violence can expect far greater YLL than a victim of diabetes. This means that YLL as a metric heavily favours violent crime over diabetes. If it was shown that the harm caused by diabetes was still greater using this metric, it could be a strong case for the greater harm due to diabetes.

Despite YLL's heavily favouring violence, we still see that diabetes contributes to the greater loss of years of life lost for all countries. This is also despite the vast differences between these countries. Although there is much less YLL per person for victims of diabetes, the sheer number of victims of diabetes overwhelms the numbers on violence. (Another perspective on this can be seen in the appendix, Figure 19). This is why we think this result says a lot in terms of our project objective. We find a similar disparity in the distribution of deaths and DALY's, with less disparity of deaths and greater disparity of DALY's. It is YLLs however, a more robust measurement then DALY's that favours violence over diabetes, that seems to send a powerful message here.

By visualising the distributions we can get a sense of the magnitude of the disparity that we cannot get with just averages. We can see that some countries have no overlap of distributions whatsoever. It does seem that the higher rates of violence in a country correspond to overlapping distributions. It's also revealing that only 0.6 % of overall demographic groups have a higher death rate due to violence than the average diabetes death rate. It appears that demographic groups that suffer more harm from violent crime are the vast minority.

We also can see that although the country is an important factor in terms of whether you are more in danger of diabetes or violent crime, it is not the full story geographically speaking. There are only 49 regions overall from all seven countries that have sustained high death-by-violence demographic groups at least once. What this and Brazil's high death-by-violence demographic group distribution suggests is that there tends to be a consistency in the location of these demographic groups. Where you are in the country makes a big difference, especially since locations that sustain high violence demographic groups and those that don't tend to keep doing so over time.

(Graphs used for this analysis were constructed with help of publicly available code. Links in project reference file).

## 4.3. RQ3. How effective is government financing at reducing the death and burden caused by interpersonal violence and diabetes on the population?

### 4.3.1. Datasets

The data for this research question came from the CDC Diabetes Atlas, and the BJS (Bureau of Justice Statistics)[7]. The latter of these two required a lot of preprocessing in order for it to be usable, but the CDC Diabetes Atlas was good to go. We would have liked an additional dataset pertaining to expenditure on diabetes in the US, but unfortunately, nothing like this seemed to be available.

[7] https://www.cdc.gov/diabetes/data/index.html, (Centers for Disease Control and Protection, 2019)

### 4.3.2. Approach

We took a slightly different approach in this research question, opting to do a case study into the United States of America specifically, primarily because media representation was such a large part of the background of this project, and the biggest producer of media representation in the world is the United States. Additionally, we thought a state-by-state breakdown, taking population into consideration, would create some interesting visualisations that could prove to be very informative, while also enabling us to assess whether or not government financing was at all effective at reducing the death and burden caused by interpersonal violence in particular.

### 4.3.3. Results

As it is very difficult to prove a causality through data alone, we want to make it clear that any suggested causality is purely speculation in these results. We cannot, one way or the other, prove whether government financing in relation to violent crime is as a preventative measure, or as a reactionary measure. All we can look at are the facts, the crime rates of a time frame, and the expenditure that went into that period.

Throughout these visualisations, we have chosen to highlight 6 states of particular interest. These are the three highest populated; California, Texas and Florida, and the three lowest populated; New Mexico, the District of Columbia, Alaska. Obviously, the District of Columbia is a District and not a state, but it proves to be an interesting control region due to its small population, so bear with us.

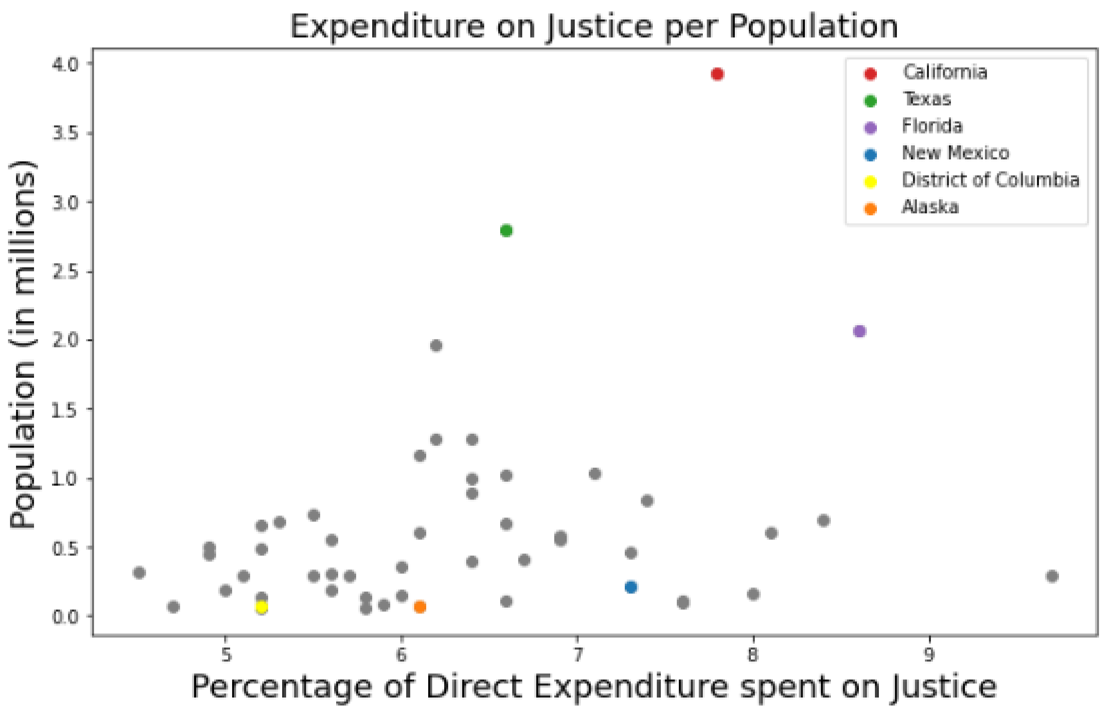


*Figure 13 Percentage expenditure on the US Justice System of Total Expenditure (per State)*

What we're looking at here is each state's population, in millions, compared against the percentage of each state's government's total expenditure on their justice system. As you can see, it is quite well distributed across the graph, the population seems to have very little effect on the percentage of the financing that went towards the judicial system.

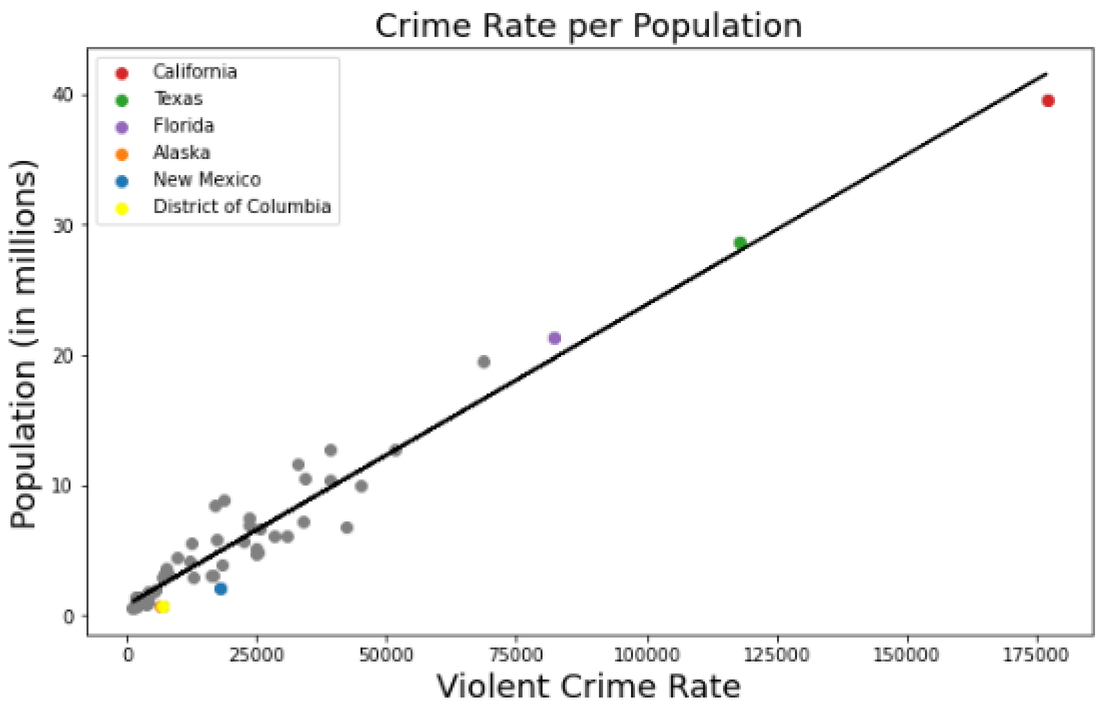


*Figure 14 Crime rate per population*

This changes however, when we look at crime rate in relation to the population. It still remains within expectations for now, with the higher populated states having higher crime rates. In this visualisation, there is a very strong correlation here, with a correlation coefficient of 0.975. So, we can be relatively confident that the higher a population is, the higher count of violent crimes we can expect.

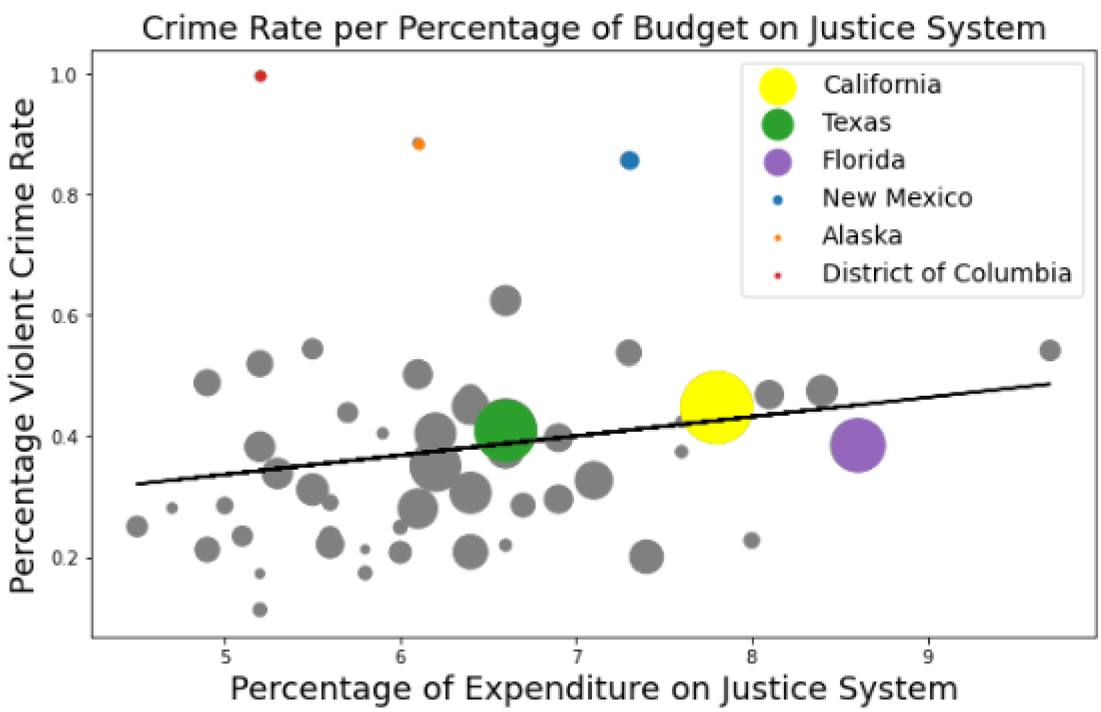


*Figure 15 Crime Rate of State vs Percentage Expenditure on Justice System (sized by Population)*

Where it gets interesting, is when we compare the state government's percentage expenditure to the crime rate of each population. Here, it seems there is a much less impactful correlation coefficient of 0.198. This is because the correlation coefficient can't take the population size into consideration, as I have done with the size of each of the points. As we can see, there definitely is a stronger correlation here when we look at the bigger populations, and it is the smaller ones that are the big outliers here. Alaska, Columbia and New Mexico in particular stand way out, single handedly reducing the correlation coefficient by a significant amount. From this, we can conclude that depending on the size of a state's population, the greater correlation there is between the crime rate of that state and it's expenditure on the judicial system. However, as mentioned previously, we can't show whether or not this is reactionary of preventative spending.

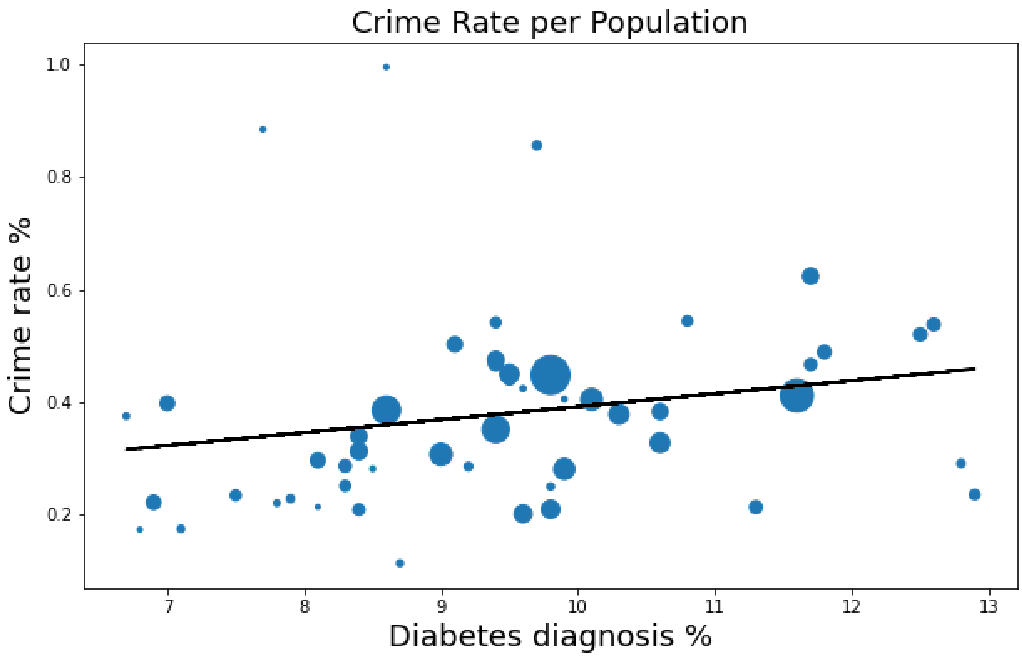


*Figure 16 Crime rate vs Diabetes Diagnosis rate (size by Population)*

Now to tie all this into our overall project - bringing in diabetes. Unfortunately, we were unable to find any specific records relating to state's spending on diabetes prevention specifically, they were all very medically general. So, let's assume for now that there is a relatively consistent spending percentage across all the states, as there is for the judicial spending, varying only by 5% at most. We can then make the conclusion that there is *likely* a similar correlation between government expenditure on diabetes prevention/treatment and that of diabetes diagnostics. This visualisation has a correlation coefficient of 0.201, but this is again skewed by the lower population outliers than the visualisation has to account for itself.

### 4.3.4. Discussion

What we found in this research question wasn't very surprising - there is a correlation between government expenditure on violent crime and the crime rate of that respective state. What was interesting, however, was the effect that the population of these states had on the outcome, and on the correlation in general. If we didn't think to include this, we could have had some very different results, and misleading visualisations with regards to where a correlation may have lay. What we found, however, was that states with smaller populations have very volatile crime rates - this is likely due to the smaller sample size of say, New Mexico (2 million), to California (~40 million).

## 4.4. RQ4. Do high-crime areas reveal a substantial difference in the ratio of death rates from diabetes versus violence?

In our last research question we continue this case study on the United States. As mentioned at the beginning of our report, we already know that more people die from diabetes worldwide, but what happens if we drop down another geographical level and compare the harm caused by diabetes and violence at metropolitan statistical areas (MSAs) - urban areas with generally higher population densities? In this question we aim to find out whether these urban areas tell a different story regarding the harm caused by diabetes versus violent crime.

1. **Datasets**

From the GHDx we obtain data on diabetes deaths for American states.

We obtain data from the Federal Bureau of Investigation on the crime rates by state and by metropolitan area.

We obtain Diabetes Atlas data from the Centers for Disease Control and Prevention on diabetes diagnosis percentages for states and the counties and other areas within them.

From the US Census Bureau we obtain data on population estimates and land area.

1. **Approach**

While we have violent crime data for metropolitan areas, we do not have data on deaths due to diabetes for these areas. Therefore, to answer this question it is necessary to calculate a diabetes death estimate for the metropolitan areas and compare it to the official crime data of these areas.

We create this estimate by using the population of each state, diabetes diagnosis percentage data for each state, and the corresponding deaths due to diabetes for each state. Using this data, we can calculate the number of people diagnosed with diabetes in each state. Then we can calculate the deaths per total population diagnosed in each state and do this for all states. Next, we get the average of this to calculate our estimate number. We multiply this number by any population of people diagnosed with diabetes to create an estimated number of people who die from diabetes in that population.

We have population data and diabetes diagnosis data for each area within our MSA's. We use this to calculate the number of people diagnosed for the whole MSA. We then use our estimate number to calculate our estimated deaths due to diabetes for each MSA.

We also create an estimate of population density for each MSA by dividing the population estimate of the counties and areas that make up MSA's with their respective land areas. We get the average of these population densities to get the overall population density for the MSA.

We then use a t-test to see whether there is a statistical significance between the estimated diabetes deaths and the violence deaths for MSA's.

We take the diabetes deaths estimate and deaths from violent crime as a proportion of the population of each MSA. This lets us compare the deaths from violent crime against diabetes in these areas. We then rank the metropolitan areas by their deaths due to violence, to get the most violent areas, and see if there is a substantial difference in the deaths due to diabetes versus violent crime in these areas.

We claim to have found a substantial difference if we find a non-trivial proportion of such areas to have a greater death rate due to violence than diabetes in their respective metropolitan areas.

## 1. Results

The results of the t-test are as shown below (Figure 17). Here we have a p value < 0.01, meaning that the differences between these two violent crime data in the MSA's and the diabetes death estimates in the MSA's are statistically significant.

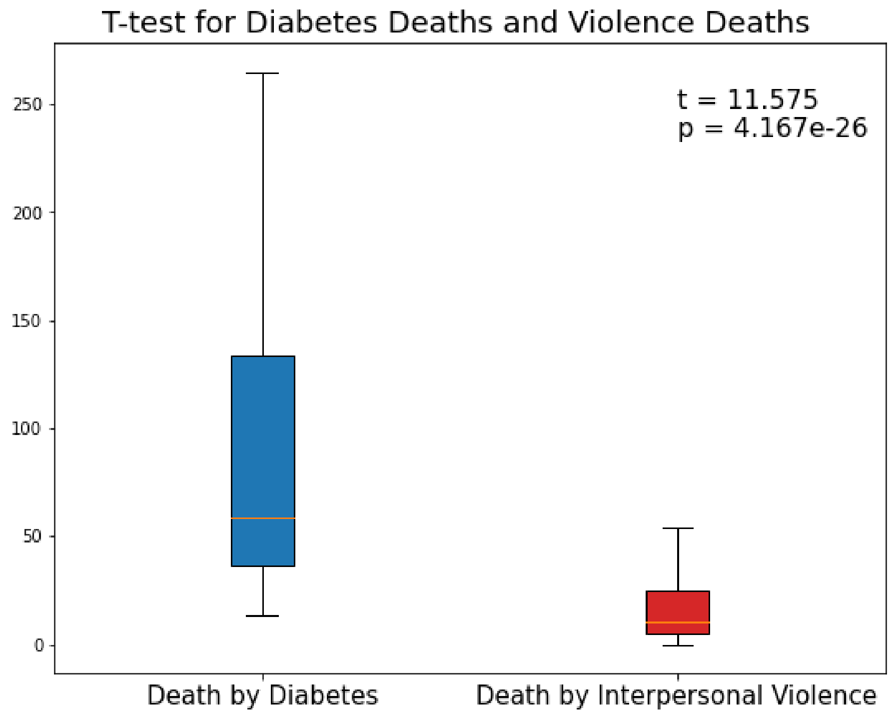


*Figure 17 T test results for death by diabetes and interpersonal violence data*

The results for the research question are easily demonstrated by the graph below (Figure 18). This graph shows 20 metropolitan areas from over 300 of such areas analysed, ranked by deaths due to violent crime per population. As we can see, even though the death rates due to violence of the most violent metropolitan areas begin to creep up to their respective rate of diabetes deaths, we can not say that we see any substantial change.

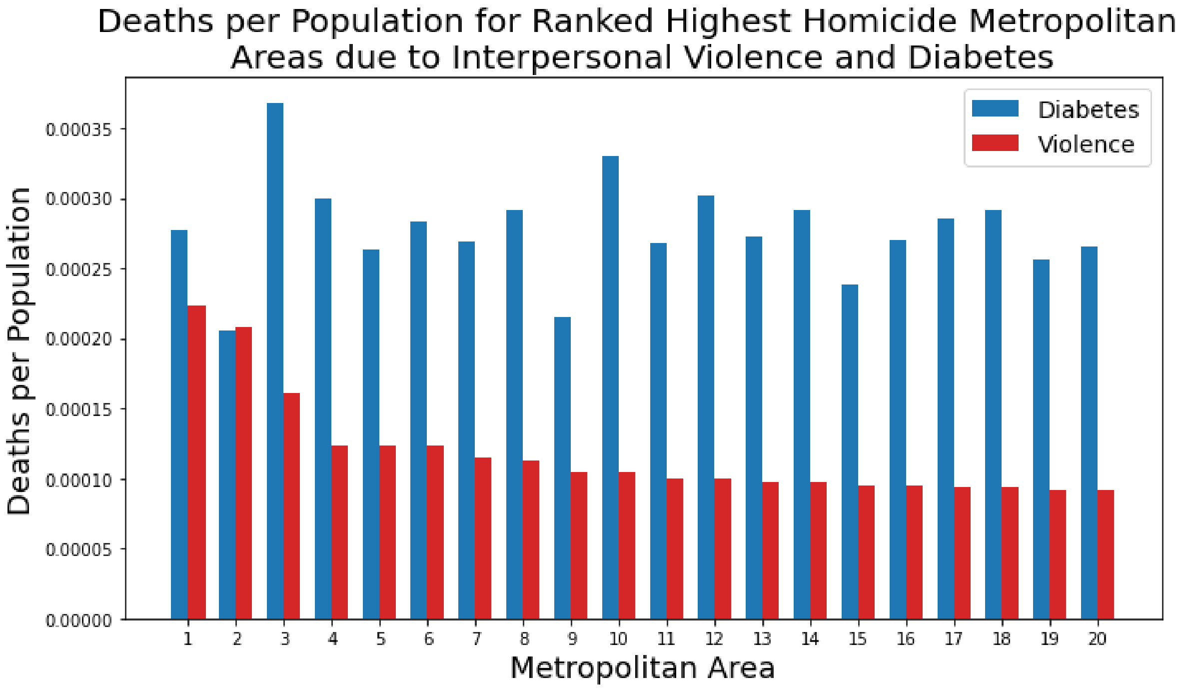


*Figure 18 Metropolitan Areas ranked by homicides per population*

**1. Discussion**

We found this result somewhat surprising, considering that some of the metropolitan areas we included had within them some of the highest homicide rates in the world. It may be the case that the geographical locations are still large enough that the high crime areas within them get balanced out by their low crime surroundings.

We had also assumed that there would be a steady convergence in deaths due to violence and diabetes with increasing population densities. We envisioned that there may have been a crossover point where violent crime overtook diabetes in deaths after a certain population density. However, analysis found no evidence of such a convergence. Our use of population density was also motivated as a way to tie in state, metropolitan areas, and cities together to analyse them on the same plane.

When we began this research question we intended to expand our diabetes death estimates to cities and expand our analysis on an even smaller, more densely populated geographical area. However, we soon realised that we would be probably overextending the estimate we already have. It would be interesting to get diabetes death data on such smaller geographical areas to judge it against our estimates here, and to continue to bring cities into this analysis.

In undertaking this analysis we had an interesting finding relating to the variance of the deaths caused by diabetes compared to those caused by violent crime. This analysis was taken from a single point in time and narrowed to a single country. With this in mind we had implicitly assumed that the variance caused by violence would be greater than diabetes, especially since we were looking at smaller, more urban areas. We expected greater variance in violent crime given the starkly different reputations different areas have for such things. We thought that the variance of diabetes would be far more uniform, given that many factors remained the same. But does not seem to be the case. The variance of diabetes seems to be a little greater, and more evenly spread, than for violent crime, which skews to the right. We wonder if we have had our perceptions distorted in a similar way that mentioned in the introduction of this report. A visualisation of this can be seen in the appendix (Figure 20).

# 5. Discussion

Before we continue with our conclusion we wish to discuss some topics that we feel are relevant to our project.

## 5.1. Ethical Considerations

In this project we use data on diabetes and violent crime from populations around the world. We also examine different sub-demographic groups based on gender, age, and location.. We do this to explore the comparative harm of violence and diabetes on these groups to meet our project's objective: to see whether people should be more worried about processed sugar than violent crime. We do not aim to associate any sub demographic group with any characterisation or pattern of behaviour. To emphasize this point, we look to the limitations of the GHDx dataset, which has provided this information on these demographic groups. In the case of geographic areas we see that some countries and regions within countries have starkly different rates of death due to both violence and diabetes. While our speculations in our discussions of our results may imply possible reasons for these observations, we do not claim to know why this is the case or what the underlying factors are that are causing them, nor is it our aim to do so.

Furthermore, this dataset provides information on the victims of interpersonal violence and diabetes. In terms of interpersonal violence it does not provide any information whatsoever on the perpetrators that cause death by violence. It’s therefore worth emphasizing that not only is it not the aim of this project to identify any categorisations of bad actors, but the analysis undertaken in this project is objectively insufficient to do so. Lastly, in the case of both death due to violence and diabetes, apart from basic demographic information of the victims, and the fact that the violence is interpersonal in nature, we have virtually no information on the conditions and context in which the events contributing to the data occurred i.e. deaths due to violence or diabetes.

Before we talk about our conclusions we want to explain what we are not saying. In our project we explore the harm from two different threats to human wellbeing. Whether we find that, in general, one of these threats is greater than the other, we do not argue that the ‘lesser’ threat is still not significant. This goes for the harm it causes for individuals as well as the institutions that deal with such harms such as medical organisations focused on diabetes and justice departments that aim to keep individuals in our societies safe from violent crime. These organisations have a role in a functioning society that stretches further than the current individuals interacting with them and their importance should not be denigrated. In terms of diabetes and violent crime we do not argue that people should be worried about just one or the other. This goes both for people in general and for the demographic groups found to be subject to high rates of death from either diabetes or interpersonal violence. Both are threats to our wellbeing, even if from one perspective one threat is more so.

We also understand that there are different perspectives in relation to this issue. While we take a higher level view and look at this issue in terms of whole populations, we wish to state that our objective is not to understate or denigrate the individual suffering caused by victims of diabetes or violent crime, whatever the conclusions of this project may be. The objective of this project, and how we believe it can help society, is to give people a more comprehensive view of two significant threats to human wellbeing and put them in perspective that aids to inform..

## Limitations

Our project has some limitations of course. There was a lack of data on the amount of direct expenditure that went towards any sort of healthcare relating specifically to diabetes, they were all too medically general. We would have liked to show a direct comparison between preventative expenditure on Diabetes and Violent Crime in RQ3, but this limitation prevented us from doing so.

We were also unable to obtain any data of diabetes deaths from cities in the United States to expand our analysis from Metropolitan areas. Considering that we know that some cities in America have the highest homicide rates in the world, we think it would make interesting research to investigate whether the story of the harm of diabetes versus violent crime changes at this geographic level.

# 6. Conclusions & Future Work

In this project we aimed to investigate whether the picture of harm caused by violent crime and diabetes could really be summed up by a single statistic. We used various, reputable, metrics of harm to measure and compare the total burden of violent crime and diabetes on multiple geographic and demographic levels. In addition, we explored deeper using data from the United States to help us understand the relationship between judicial and law enforcement expenditure to that of crime rates in each state.

We found that through the use of Deaths, YLLs and DALYs, we can create a very insightful overview of a metric for harm caused by interpersonal violence, and diabetes. We found that when using any of these metrics, it was advised that you investigate with more than one to better understand what is going on.

With respect as to whether or not we are suffering more or less now from diabetes and violent crime than previously, the answer depends on which you're looking at. Diabetes, consistently, has been on the rise in the last 20 years of this dataset, whereas violent crime seems to be slowly falling in recent years. Previous trends show it rising and falling often, but in some areas, the current dip is the lowest that has ever been seen.

Historically, we found that women were more likely to die in higher numbers than men from diabetes. However, this is seemingly no longer accurate, as in the last 2 years of the dataset, men died more, as they seemed to overtake women and continue to die in growing numbers year by year. As speculated, this is likely due to the fact that women tend to develop diabetes at an earlier age than men, but as life expectancy grows, so too do the number of men being diagnosed with diabetes at an older age. This, coupled with the mass misinformation surrounding sugar and its harmful effects back in the 50's and 60's, has likely had a massive knock-on effect where today's elderly are suffering much more from diabetes than the youth. This is also likely due to the increased possibility of developing hyperglycemia the older one gets.

When we analysed trends worldwide, we found that while more people do still suffer from diabetes overall, the actual harm caused on an individual level basis to the average victim of the respective cause was far greater for that of violent crime. This finding continues in regions within countries, and we find that despite the average victim of violence losing over 3 times the years of life the average victim of diabetes does, diabetes still contributes to the greatest loss of life, due to the sheer numbers of victims.

In relation to the effect financial expenditure has on the crime rate of a state or district, we found that there is definitely a correlation between the two factors. However, whether or not this trend was preventative, or reactionary, it is impossible to tell from just the data alone. Realistically, we can hypothesize that it is likely a combination of both. In relation to diabetes, as previously mentioned, any takeaways would be purely speculation. Again, we can hypothesize that there is a correlation between government spending and diabete due to the consistent trend between population and expenditure, but it is just a supposition. We can only conclude that there is a correlation between crime rate with respect to the population size, and the percentage of a government's total annual expenditure.

We found no substantial difference in the harm caused by diabetes versus crime in metropolitan areas. Diabetes still brings more deaths in these areas, despite some of these areas containing regions within them that have very high homicide rates. Again, this result presses for an expansion of this analysis to just cities themselves.

Overall, we believe we were successful in achieving what we set out to do - we have sufficiently shown that Processed Sugar, in the form of Diabetes, is a much deadlier deathbringer and life harmer than that of violent crime. We have shown that, in general, we should be more worried about the consumption of processed sugars, due to the alarming increase in the harm it has worldwide.

# 7. Responsibilities

Research Questions:

- RQ1 - Evin
- RQ2 - Marcus
- RQ3 - Evin
- RQ4 - Marcus

Report Writing:

- Abstract - Marcus
- Introduction - Both
- Motivation & Objectives - Both
- Data Acquisition - Marcus
- Ethical Considerations - Marcus
- Limitations - Both
- Conclusions & Future Work - Evin

# 8. Bibliography


1. GHDx (2021). *About the GHDx*. GHDx.

   http://ghdx.healthdata.org/about-ghdx

2. WHO (2021). *Disability-adjusted life years (DALYs).* World Health Organisation. https://www.who.int/data/gho/indicator-metadata-registry/imr-details/158

# 9. Appendix

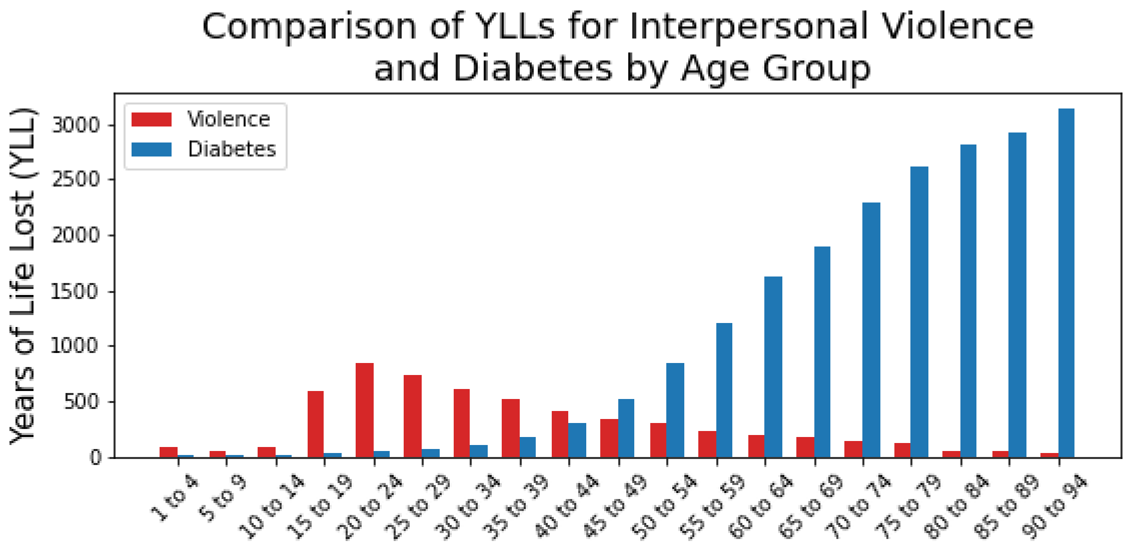


*Figure 19. Comparison of YLLs for interpersonal violence and diabetes by age)*

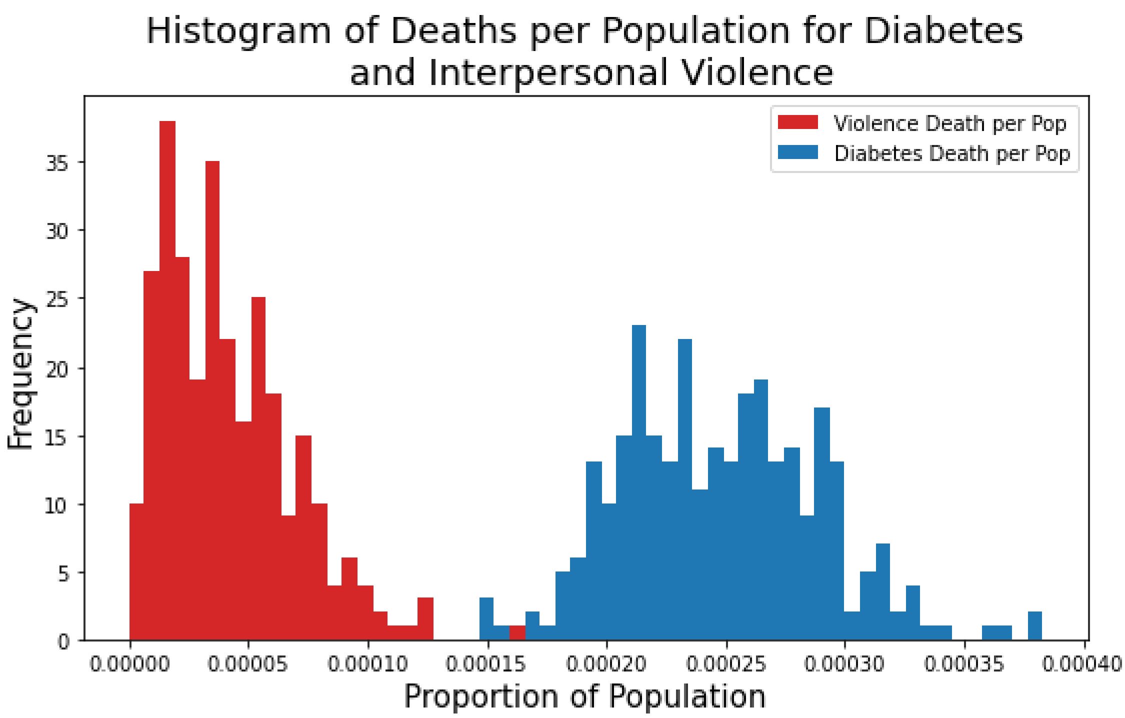


*Figure 20. Histogram of deaths per population for diabetes and interpersonal violence*