

## **SCIE 201 - Group 22**

### **Group Project**

Nelson Wong  
Wasiur Rahman

#### **Group Members**

Arman Yazdanbod  
Fatem Khani  
Muhannad Nouri  
Zoya Younis

## **Section A:**

Throughout this semester we have been equipped with the knowledge to follow steps concerning data analysis. For this group project, we have been asked to analyze our choice of datasets to show a clear understanding of all the steps in the data analysis pipeline. We have decided to study the impact of population growth on the crime rate. To accomplish this task, we will analyze two datasets from Statistics Canada. Our first goal is to clean, analyze and visualize the population dataset to see the general trend in the population across Canada, and specifically the distribution of population across each respective Province and/or Territory. We want to find out what percentage of Canadian Population reside in each given region of the country. Next, we want to clean and analyze the “crime rate” data set. As we were unable to find a reliable source for a crime rate of Canada in the same time period as our population dataset, we decided to derive our own crime rate based on guilty criminal cases across Canada and for each Province and/or Territory. We want to gather information about the population, and about the number of guilty criminal cases in order to compare them and look for trends and correlations. We hypothesize that with higher population growth, comes higher percentage of guilty criminal cases. On the surface, this could seem like common knowledge but without the right figures and statistics backing it up, it cannot be confirmed. This brings us to the task on hand; The ability to accurately organize, clean, and analyze the data so that we could create visuals to clearly present our findings.

## **Section B:**

We decided to utilize the datasets provided from Statistics Canada as it was not possible for us to gather sufficient and accurate data ourselves across Canada to answer the questions on hand. The source for our data, Statistics Canada, is considered to be highly reliable and thus enabling us to analyze it with confidence that it is accurate. We obtained the data through a method of crowdsourcing. Our two datasets are as follows:

1. [Estimates of population, Canada, Provinces and Territories](#)
2. [Court, adult cases by type of sentence, total guilty cases, by province and territory](#)

The first dataset will be analyzed to look at population across Canada and its distribution per each Province and/or Territory. We want to compare the population size of each Province and/or Territory to find out what percentage of the overall population they each represent. The second dataset will be used to address the crime rate. We will look at the total number of guilty criminal cases across Canada and compare that to the total number of guilty cases in each Province and/or Territory to find out what percentage of total crime they each represent. Finally, we will compare these percentages to see if there is a trend between population size growth and an increase in crime.

## **Section C:**

The datasets obtained are clean in the sense that they do not include any spelling, spacing or formatting issues. We uploaded the data to OpenRefine and searched for duplicates, misspellings, capitalization and inconsistent formatting. The datasets passed all these tests. Thus there was no need to clean the data. However, there were values in both datasets that were not necessary for us to answer the questions.

The first dataset included the population size of Canada over a year and half period leading to the end of 2016. This dataset displays Canada's population as a whole, and then further broken down into each province and territory. The only values needed from this dataset were Canada's population as a whole at the end of 2016, and those values for each Province and/ or Territory at the end of 2016. As a result, using Microsoft Excel we organized the data by removing all other unnecessary values while keeping the overall organization and format of the data set intact.

This is shown in the images displayed below:

### **Dataset 1:**

Geography	2016		2017		
	Q3	Q4	Q1	Q2	Q3
Canada	36,264,604	36,415,721	36,474,968	36,560,776	36,708,083
Newfoundland and Labrador	530,305	530,618	530,097	529,204	528,817
Prince Edward Island	149,472	150,099	150,271	150,769	152,021
Nova Scotia	948,618	951,345	950,990	952,016	953,869
New Brunswick	757,384	758,606	758,198	758,148	759,655
Quebec	8,321,888	8,344,608	8,351,714	8,366,022	8,394,034
Ontario	13,976,320	14,053,658	14,084,797	14,124,305	14,193,384
Manitoba	1,318,115	1,323,550	1,327,883	1,331,960	1,338,109
Saskatchewan	1,148,588	1,153,590	1,156,342	1,159,549	1,163,925
Alberta	4,236,376	4,250,816	4,261,116	4,272,398	4,286,134
British Columbia	4,757,658	4,778,963	4,783,461	4,795,891	4,817,160
Yukon	38,086	38,170	38,209	38,273	38,459
Northwest Territories <sup>5</sup>	44,617	44,406	44,452	44,584	44,520
Nunavut <sup>5</sup>	37,177	37,292	37,438	37,657	37,996

### Transposed Dataset 1:

	Q4
Canada	36,415,721
Newfoundland and Labrador	530,618
Prince Edward Island	150,099
Nova Scotia	951,345
New Brunswick	758,606
Quebec	8,344,608
Ontario	14,053,658
Manitoba	1,323,550
Saskatchewan	1,153,590
Alberta	4,250,816
British Columbia	4,778,963
Yukon	38,170
Northwest Territories <sup>5</sup>	44,406
Nunavut <sup>5</sup>	37,292

The second dataset was uploaded to OpenRefine and tested with the same precision. Furthermore, this dataset was organized and presented in a clean manner, and thus there was no need to make any changes. The second dataset had values that were of no need to our analysis similar to the first dataset. Thus the dataset was uploaded onto excel and the extra values were removed. These extra values included specific data on the type of crime committed:

### Dataset 2:

	2015/2016						
	Total guilty cases	Custody	Conditional sentence	Probation	Fine	Restitution	Other sentences <sup>1</sup>
	number						
<b>Total offences</b>	<b>220,233</b>	84,241	8,022	94,505	69,842	5,432	164,788
Total <i>Criminal Code</i>	<b>192,005</b>	76,573	7,188	85,052	58,655	5,388	150,578
<i>Criminal Code</i> (without traffic)	<b>155,846</b>	70,233	6,530	79,160	30,622	5,210	124,110
Crimes against the person	<b>40,219</b>	15,129	1,973	28,804	4,153	512	32,680
Homicide	<b>108</b>	81	0	12	3	0	73
Attempted murder	<b>32</b>	22	0	4	2	1	17

**Transposed Dataset 2:**

Canada	Total Guilty crime cases
Ontario	64611
Quebec	48697
Alberta	35850
British Colombia	24786
Saskatchewan	15506
Manitoba	12146
Nova Scotia	6876
New Brunswick	5109
Newfoundland and Labrador	3297
Nanuvut	986
Prince Edward Island	898
North West Territories	845
Yukon	635

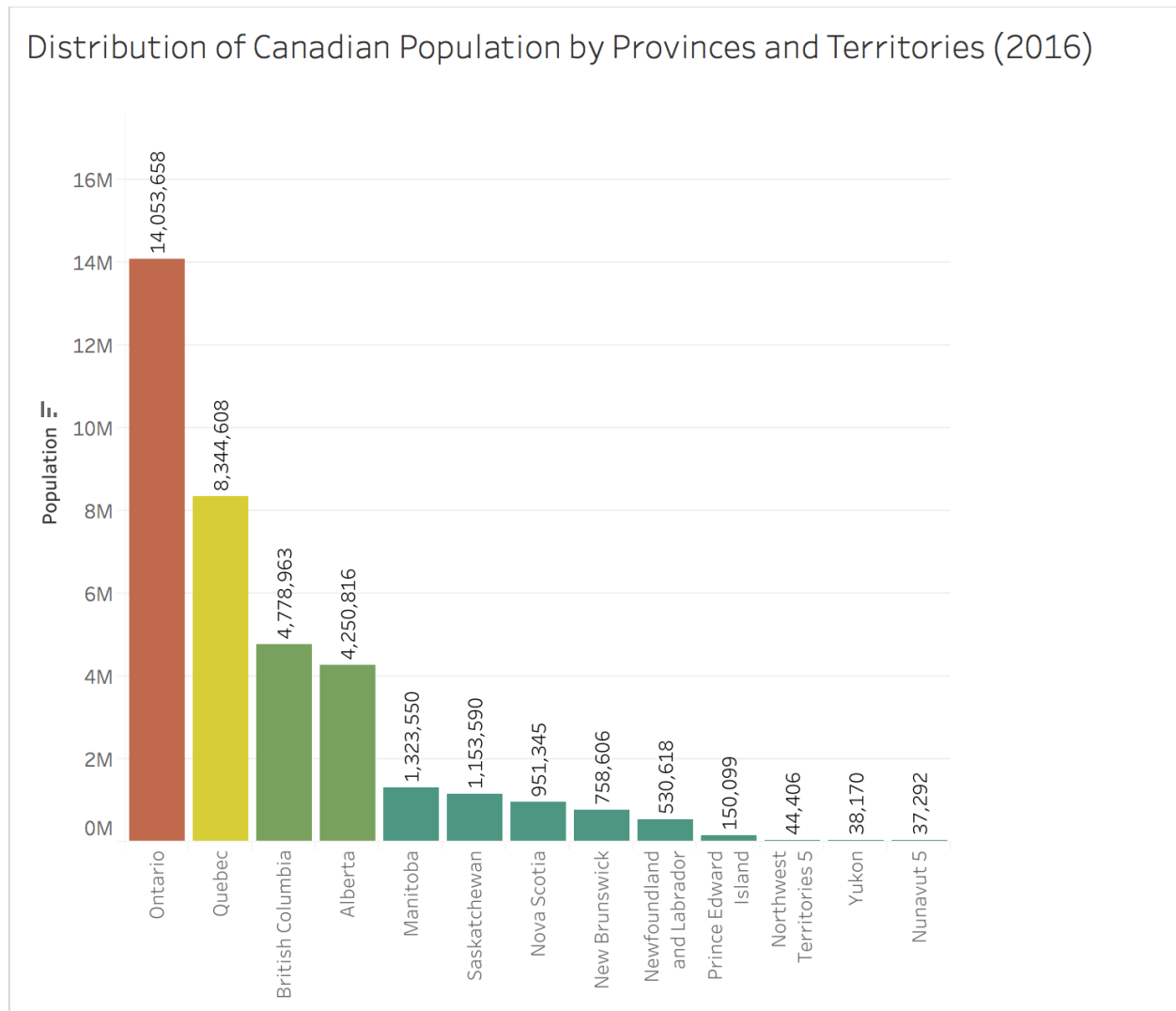
### **Section D:**

1. What is the Estimated Canadian Population size for each Province and/or Territory at the end of 2016?
2. What Percentage of the Canadian population is represented by each Province and/or Territory?
3. What is the total number of guilty criminal cases across Canada, per each Province and/or Territory at the end of 2016?
4. What Percentage of total criminal cases does each Province and/or Territory represent?
5. How does the population of each Province / Territory compared to their respective number of guilty criminal cases?
6. How does the percentage population of each Province / Territory compared to their percentage share of guilty criminal cases?

### **Section E:**

After making sure the data is clean on OpenRefine, we uploaded the data to Microsoft Excel and transposed it and excluded all the values that were not necessary for each given question. Next, we uploaded the needed values into a separate Excel sheet, saved it and uploaded it into Tableau. On Tableau, with a simple drag and pull we placed our values on their respective axis, made sure every visualization was formatted correctly, and each variable was labeled.

### Visualization For Question 1: (Tableau)

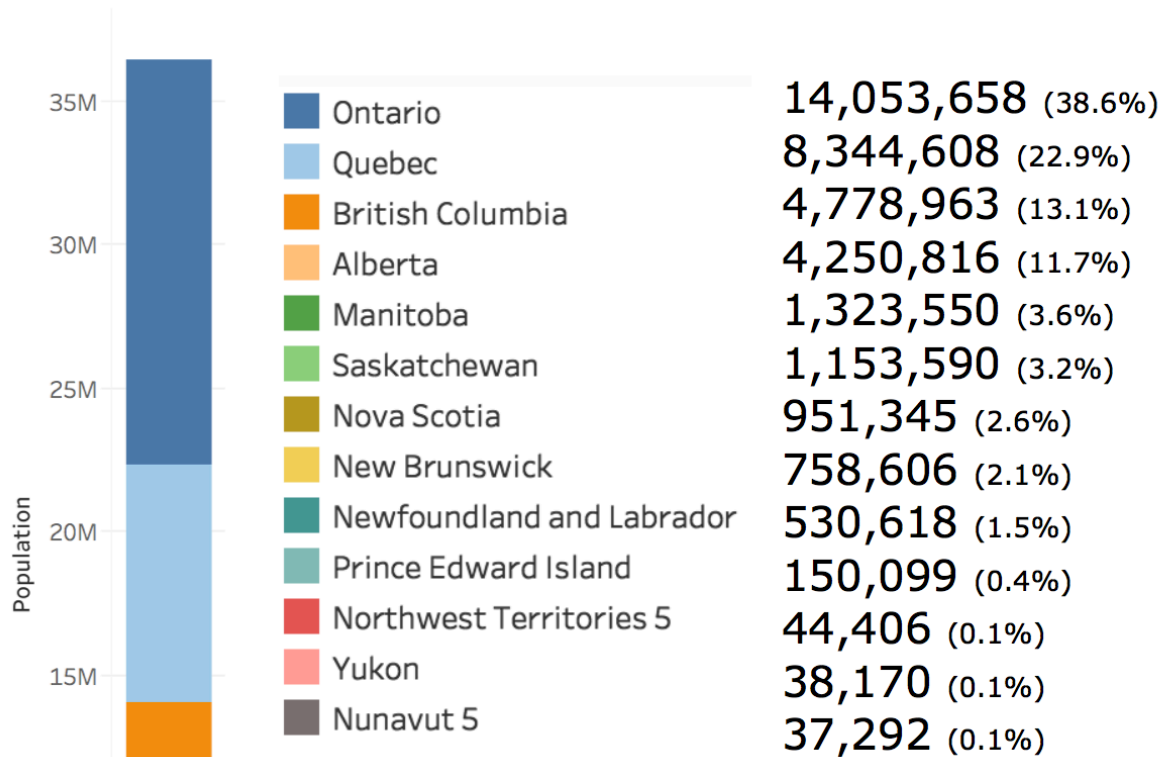


This visualization allows us to easily compare the population size of each Province / Territory to one another. It is sorted in descending order from the largest population size on the left hand side (Ontario) to the smallest population size on the right hand side (Nunavut). This diagram was chosen as it is accurately labelled with the numeric population size of each Province / Territory, the name of that Province / Territory, and further colour coded by value range to show different colours when the numbers vary too much from one another.



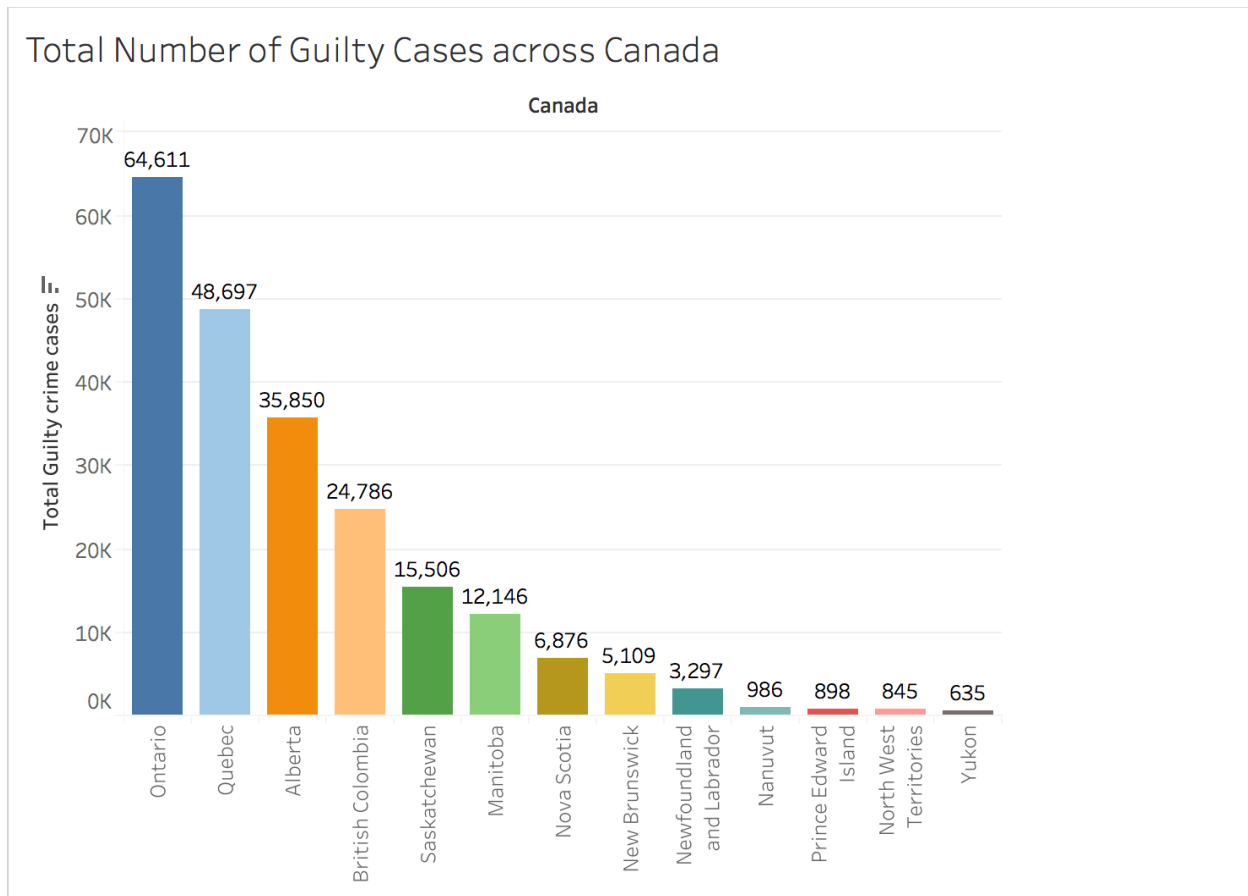
## Visualization for Question 2:

### Distribution of Canada's Population



This visual permits us to compare the distribution of population across each Province and/or Territory across Canada. We decided to use this method of presentation as a pie chart did not allow for an axis on the left hand side so that each value could be labeled. Also, on this chart, the legend can be much more descriptive and easier to understand.

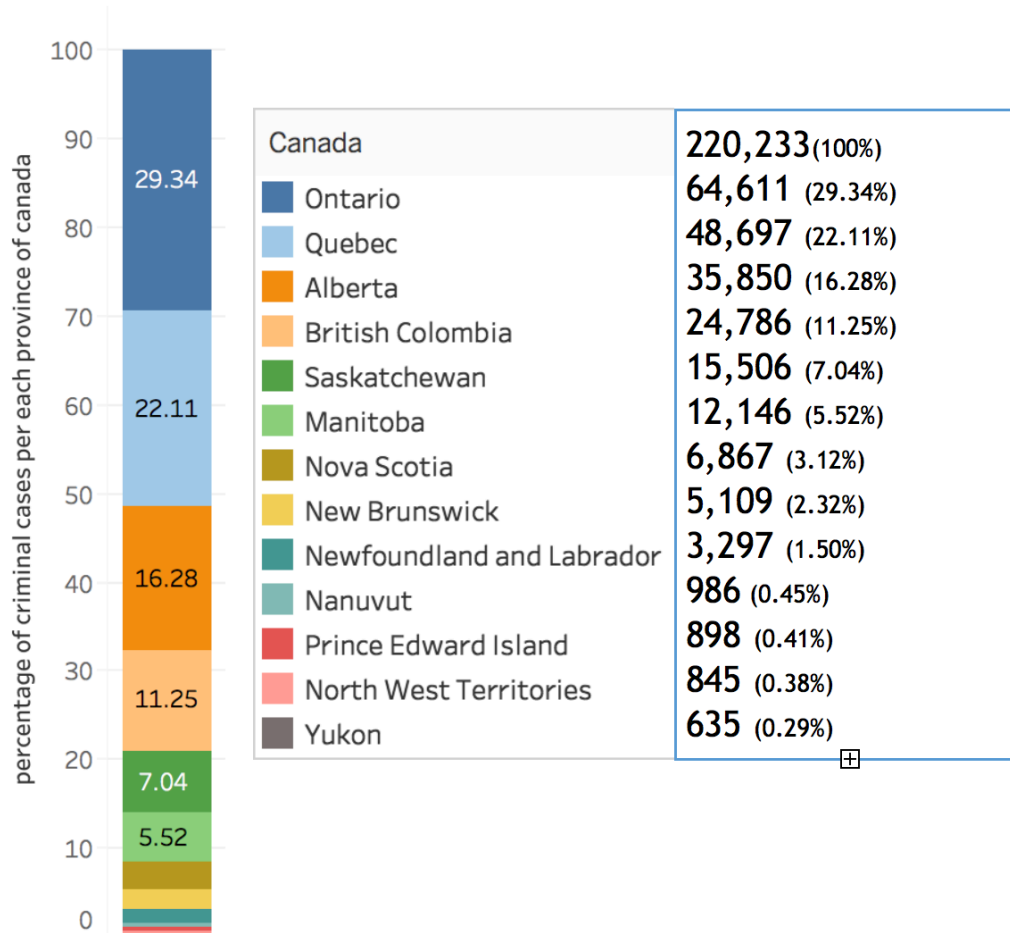
### **Visualization for Question 3:**



This visual shows the distribution of total number of guilty criminal cases across Canada per each respective Province / Territory. It is sorted in descending order from the highest number of criminal court cases on the left (Ontario) to the smallest number of criminal court cases on the right (Yukon). This method of presentation was used as this bar chart clearly shows the difference between each Province / Territory, the y-axis clearly shows each value mark, the x-axis clearly has every Province / Territory titled, each bar is labeled with their numeric value and furthermore, the colour coding clearly differentiates between each Province and/or Territory.

#### Visualization for Question 4:

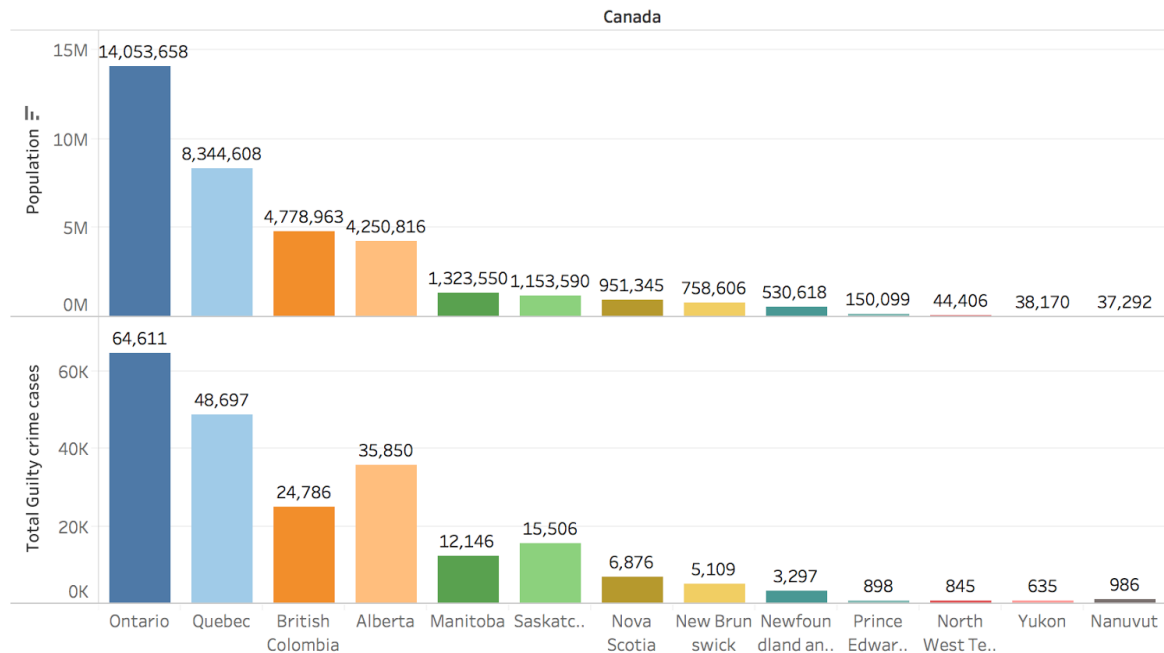
Percentage of Total Guilty Criminal Cases per Province



This visual shows the percentage share of guilty criminal cases per each Province / Territory. We decided against the pie chart because the difference between Ontario, Quebec, and Alberta were not presented as well as they have been on this visual. Furthermore, the inability to be able to create such a detailed legend and y-axis value marks on a pie chart made us choose this method of visualization. This diagram clearly shows the percentage share of each Province / Territory based on the scale on the y-axis, quantitative presentation in the legend and differentiates between each Province / Territory by colour coding.

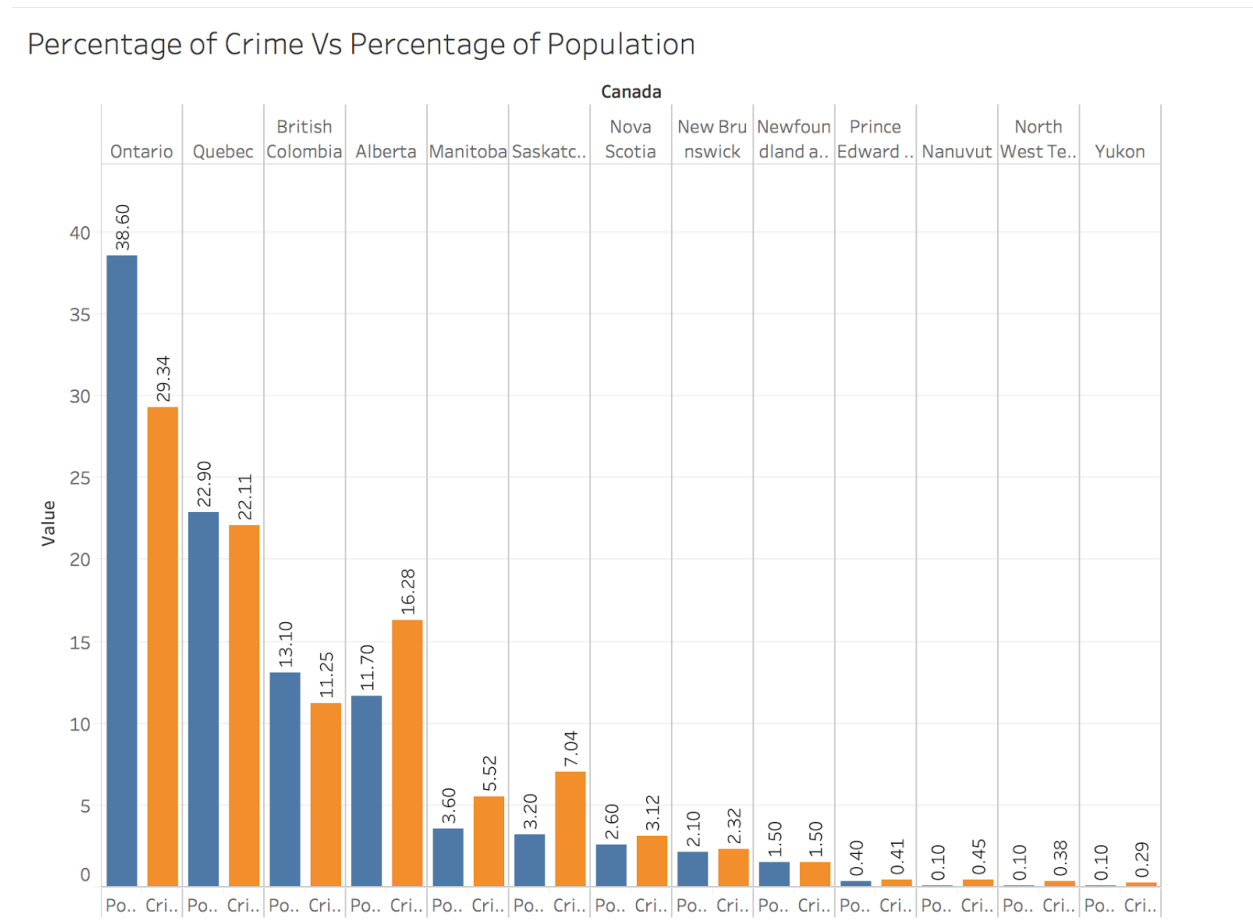
### **Visualization for Question 5:**

Total Number of Guilty Criminal Cases Vs. Population



This visual was used to compare the population size of each Province / Territory to their respective number of guilty criminal cases. The values are sorted in descending order based on population, with the highest on the left hand side and the lowest on the right hand side. This diagram shows that Alberta, Saskatchewan and Nunavut are showing higher ranking in number of guilty criminal cases than their rank in population size. This could mean that they have a higher crime rate but these two quantitative values cannot be compared 1:1 and a correct correlation cannot be derived from them. Thus bringing us to our last question that allows for a much more direct comparison when we look at the percentage share of each Province / Territory.

## Visualization for Question 6:



In this visual, we compare the percentage share of population vs “crime rate” for each given Province / Territory. This diagram has been sorted in descending order from highest on the left to the lowest on the right. The y-axis shows the value for each bar in percents, and each bar has a numeric value further showing its exact value. The x-axis is labeled to show that the blue bars present the population values and the orange bars present the values for “crime rate”. Each pair of blue and orange bar is further labeled at the top so that the values for each Province / Territory is distinguishable. This diagram allows for a more direct comparison between the two variables.

## **Conclusion:**

We were able to confirm our hypothesis that there is a general linear trend between population size and “crime rate”. The two variables, the variables being population growth and crime rate, are proportional to each other. We are able to conclude that as the population size increases, so does the total number of criminal guilty cases. As a result, this contributes to a larger crime rate. As previously stated, on the surface this hypothesis could be construed as common knowledge, but proving it without the correct tools would have not been possible. Clear understanding of the data pipeline has enabled us to efficiently collect, clean and analyze data in order for us to answer our questions in a scientifically sound manner.

### **References**

1. Statistics Canada, CANSIM, Table [252-0056](#) - Court, adult cases by type of sentence, total guilty cases, by province and territory, Catalogue no. [85-002-X](#).
2. Statistics Canada, Table [051-0005](#) - Estimates of population, Canada, provinces and territories, quarterly (persons), CANSIM (database).