Data Viz – Assignment 4

[Personal income tax rates and credits - 2022 - Ontario Data Catalogue](https://data.ontario.ca/en/dataset/personal-income-tax-rates-and-credits/resource/1ee6b3a7-c357-46c2-8aef-27f10ed7126e)

https://data.ontario.ca/en/dataset/personal-income-tax-rates-and-credits/resource/1ee6b3a7-c357-46c2-8aef-27f10ed7126e

What software did you use to create your data visualization?

Excel

Who is your intended audience?

Individuals who file yearly income tax in Canada

What information or message are you trying to convey with your visualization?

The percentage of tax that the individual will be required to pay based upon the income level. The individual's income level will determine which bracket he/she belongs to, and after determining that, they will understand their tax percentage

What design principles (substantive, perceptual, aesthetic) did you consider when making your visualization? How did you apply these principles? With what elements of your plots?

Aesthetic – I chose colours that were not overwhelming or deep to look at. A palette which is not distracting away from the actual data itself. The draw the audience's eye's to the actual bar's itself and hopefully not wander too much away, I used thicker border than the default. Regarding the x and y axis labels, I bolded it as well as made the font size bigger so that it will be sufficiently sized to be able to view without having to squint and pay more attention than what is absolutely needed. Lastly, I put the actual raw data percentage on top of each bar so that the audience will know what the exact value is rather than having to guess based on the tick mark and y-axis values. [1]

Substantive: The actual raw data provided in this dataset are not averages, but actual specific values that correspond with each of the tax brackets, so therefore, the visualization does accurately and honestly present the data in a manner that audience can comprehend without deception. [2]

Perceptual: The visualization is trying to convey that with increased income, the amount of tax they will be having to pay will be incrementally higher, not taking into account the various deductibles and tax credits that may be available to the individual. [3]

---

[1] Data Visualization: Graphing our Data: Choosing the Right Visualization (Pg 21, 22)
[2] Data Visualization: Graphing our Data: Choosing the Right Visualization (Pg 21, 22)
[3] Data Visualization: Graphing our Data: Choosing the Right Visualization (Pg 21, 22)

How did you ensure that your data visualizations are reproducible? If the tool you used to make your data visualization is not reproducible, how will this impact your data visualization?

With excel it is very hard to make the visualization reproducible as there is no coding. To those who are aware of how to use excel, it would not be too hard to replicate since the values and text are in a convenient location where they can create, but to one who has no knowledge about how to use excel, it may be challenging without walking them through it. [4]

The lack of reproducibility will not impact the specific data visualization that I created, but if its not reproducible by others, its hard to convey the information to the mass's and therefore, it would only be the individual's who are looking at my visualization who would be able to comprehend what is being conveyed. [5]

How did you ensure that your data visualization is accessible?

1. Choosing a colour pallet that does the sway away individual's who are colour blind[6]
2. Choosing a font type that is one of the commonly used, and font size that is sufficiently large enough so that people can make out.[7]
3. Putting the actual percent value on top of the bar's so that the audience knows what the actual value is.[8]
4. Using a proper chart title that sufficiently describes what the visualization is about.[9]

Who are the individuals and communities who might be impacted by your visualization?

Every individual who has to file taxes in Canada will be impacted by this visualization.

How did you choose which features of your chosen dataset to include or exclude from your visualization?

This was based on simplicity more than anything else, and the amount of information to the most amount of people I though it may impact. If too many features were included, such as the various deductibles and credits, it would have taken away from the actual information that I wanted to convey, which was simply "which tax bracket is one in". Overall aim was to lessen the cognitive load[10]

What 'underwater labour' contributed to your final data visualization product?

The government who has to determine the tax brackets for the different income levels, and individuals involved in aggregating all the data that are in this dataset

---

[4] Data Visualization – First Steps: Reproducible Data Visualization (Pg 12-15)
[5] Data Visualization – First Steps: Reproducible Data Visualization (Pg 12-15)
[6] Data Visualization – Visualization with Purpose: Accessible Data Visualisation (Pg 17-25)
[7] Data Visualization – Visualization with Purpose: Accessible Data Visualisation (Pg 27-29)
[8] Data Visualization – Visualization with Purpose: Accessible Data Visualisation (Pg 27-29)
[9] Data Visualization – Visualization with Purpose: Accessible Data Visualisation (Pg 32)
[10] Data Visualization – Graphing our Data: Choosing the Right Visualization (Pg 33,34)

What software did you use to create your data visualization?

R Studio

Who is your intended audience?

All residents of Ontario

What information or message are you trying to convey with your visualization?

There is an upward trend with regards to population from 2021 to 2046 in all age groups, but only the high-growth category shows a greater upward growth I population compared to reference.

What design principles (substantive, perceptual, aesthetic) did you consider when making your visualization? How did you apply these principles? With what elements of your plots?

Aesthetic – Larger, bold, black font, and Aerial font was used for the title, axis, facet, and legend titles. This was done so that audience would be able clearly distinguish what the variables were and the data that is trying to be conveyed.[11]

Facet was used because there are two different groups (Low growth and high growth) being compared to the reference. A clear distinction between the categories wanted to be made, and therefore, a facet was used.[12]

Different colour palette was used to make clear differentiation between the different age groups. In the raw data, there is data for each year from 2021 to 2046, but it felt unnecessary to use all since it would be too much information to visualize, and the same overall result could be represented by only using three time plots (2021, 2035, 2046).[13]

Substantive: Accurate and honest results are being conveyed as it is using the raw data from the dataset, and the trend from beginning to end is not intentionally and/or unintentionally being manipulated in a manner that would mislead the audience to believing something other than what is being portrayed, which is the general uptrend in population from 2021 to 2046, especially in the high-growth scenario. [14]

Perceptual: There is a uptrend in population from 2021 to 2046 for each of the groups, but not too much different when comparing BETWEEN the reference to the low growth and high growth for the 0-14 and 65+ age group. On the other hand, the 15-64 age

---

[11] Data Visualization – Visualization with Purpose: Accessible Data Visualisation (Pg 27-29)
[12] Data Visualization – Visualization with Purpose: Accessible Data Visualisation (Pg 27-29)
[13] Data Visualization – Visualization with Purpose: Accessible Data Visualisation (Pg 27-29)
[14] Data Visualization – Visualization with Purpose: Accessible Data Visualisation (Pg 27-29)

groups shows a consider growth trend when comparing from reference to the high-growth group.[15]

How did you ensure that your data visualizations are reproducible? If the tool you used to make your data visualization is not reproducible, how will this impact your data visualization?

We can ensure reproducibility by making the R-markdown code available to whoever wants to view it. Without the code, it would be very hard to replicate. [16]

How did you ensure that your data visualization is accessible?

Larger text size, and bold font were used so that the text would stand out to the audience. Similarly a standard Aerial font type was used. A good colour palette was used to increase contrast, and account for individuals who are colour-blind. A grey background was used rather than a clear white so that a defined space was used to attract the eyes to the visualization. Each of the bar's are marks with a boarder to attract the audience's eye's to the data rather than have them wander.[17]

Who are the individuals and communities who might be impacted by your visualization?

Many groups are impacted by the rise in population, but its the politician's specifically who need to decide what to do and how to account for the increase in population, such that demand does not outweigh the supply of various things, such as various services, programs, an housing. Although it's the politicians who are going to do the hard labour of building a province that accounts for increased population, it will be the actual individual's who will be impacted because they will need to navigate through all the changes.

How did you choose which features of your chosen dataset to include or exclude from your visualization?

Although data is available for each year from 2021 to 2046, a conscious decision was made to only use three of the years (beginning, middle, and end) so not to overcrowd and make the visualization too busy to look at and prevent the stress of cognitive load. [18]

Similarly, age group was available for each year, but a conscious decision was made to only use the three age groups, to account for (youth, middle age, and older age). This was done also to decrease the cognitive load and not to overwhelm what the visualization is trying to convey to the audience. [19]

What 'underwater labour' contributed to your final data visualization product?
The main group of people's who go unnoticed are the analysts that predict the population and put a value on them for each of the future years for each age and sex group.

---

[15] Data Visualization – Visualization with Purpose: Accessible Data Visualisation (Pg 27-29)
[16] Data Visualization – First Steps: Reproducible Data Visualization (Pg 12-15)
[17] Data Visualization – Visualization with Purpose: Accessible Data Visualisation (Pg 27-29)
[18] Data Visualization – Graphing our Data: Choosing the Right Visualization (Pg 33,34)
[19] Data Visualization – Graphing our Data: Choosing the Right Visualization (Pg 33,34)