Getting to Know Canadian PRs

CPSC 436V: Information Visualization
Project Milestone 3

April 14, 2021
Group 1

Ansel Hartanto
15984164 | o8c1b | ansel

Adrian Lewczuk
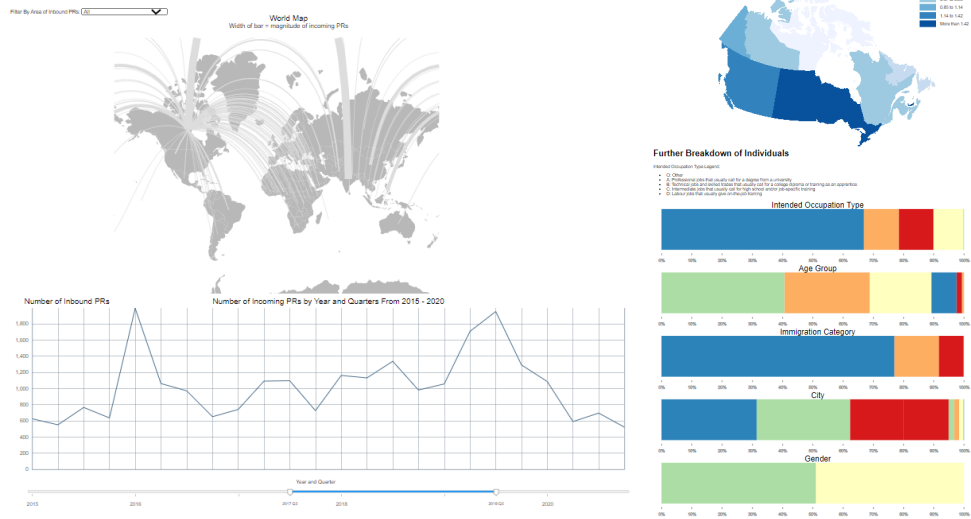41700162 | p1a1b | alewczuk

Alvin Lo
38125167 | p8c1b | ubcalvin

Table of Contents

1. Overview

Teaser Image

Project Summary

Canada is often referred to as a country of immigrants, and for good reason. From 2017-2018, immigration accounted for 80% of our population increase, and in 2019 we experienced the highest level of permanent resident admissions in recent history ([2020 Annual Report to Parliament on Immigration](#)). But who are these permanent residents?

This is the question we set out to answer when creating *Getting to Know Canadian PRs*, our visualization for this project. Despite being a seemingly straightforward question, it remained largely unanswered. We believe that the lack of readily available and easily digestible information on this topic could be contributing to stereotyping and any negative sentiment towards immigrants. Although this is a constant issue, it becomes especially apparent around elections when immigration policy is up for debate.

Getting to Know Canadian PRs demystifies our permanent residents (PRs) by visualizing publicly available information about their:

- countries of citizenship;
- intended destinations;
- intended occupations;
- immigration categories;

- ages;
- genders;
- and immigration trends in general.

We believe it can help combat our identified problem by educating the general public.

2. Data

Data Description

Our core data comes from 7 datasets containing information about approximately 1.7 million permanent residents admitted into Canada between 2015 and 2020. After processing, they contain the following attributes:

Attribute	Example	Type	Cardinality/Range
Date (Quarter)	2015 Q1	Ordinal	24
Country (of Citizenship)	India	Categorical	100
Region (of Country of Citizenship)	Western Asia	Categorical	20
Province/Territory (Intended)	Ontario	Categorical	14
Population (of Province/Territory)	528,361	Quantitative	36,219 - 14,734,014
Age Group	15-29	Ordinal	6
Gender	Female	Categorical	2
Intended Occupation Type	A	Categorical	5
Immigration Category	Sponsored Family	Categorical	5
City (Intended)	Yellowknife	Categorical	172

Additionally, some attributes are derived in real time based on user interaction:
- Aggregations
  - Our datasets do not include any totals; they are all computed. For example, summations across geographical areas or time periods.
- Percentages
  - Our datasets only include absolute values; relative percentages are computed. For example, percentages of new PRs relative to provincial/territorial populations.

Data Source(s)

Our core data is publicly published on a quarterly basis by the Canadian government:
- [Admissions of PRs by Country of Citizenship](#)
- [Admissions of PRs by Province/Territory and Age](#)
- [Admissions of PRs by Province/Territory and City](#)
- [Admissions of PRs by Province/Territory and Gender](#)
- [Admissions of PRs by Province/Territory and Immigration Category](#)
- [Admissions of PRs by Province/Territory and Intended Occupation](#)
- [Province/Territory Population Estimates](#)

However, we also accessed other public sources as needed:
- [Kaggle: Country to Continent](#)
- [Github (phil-pedruco): GeoJSON of Countries](#)
- [Github (holtzy): GeoJSON of Countries](#)

Data Processing

We are working with historical data, meaning it is not live and only needed to be processed once. In theory, the same transformations could be repeated on new data so that the visualization could be updated each quarter (when new data becomes available).

In short, the following changes were made:
- Standardization
  - All datasets were transformed to have the same column setup (one per quarter of each year) and row setup (one per country or one per province-attribute pair).
- Completeness
  - Due to privacy issues, "--" was used for any values greater than 0, but less than 5. Since all the other numbers in the datasets are rounded to the nearest 5, we decided to round these values down to 0 anywhere they appeared.
- Accuracy
  - Totals in the datasets were calculated from unrounded subtotals and then rounded, meaning the sum of subtotals didn't always equal the totals. To fix this, we removed all totals and opted to calculate totals ourselves when needed.
- Simplification
  - Some data was narrowed down, and some was grouped. For instance, we opted to only include the top 100 countries out of the 200 contributing PRs to Canada in the past 5 years and to group hundreds of job codes into just 5 job types (based on the same NOC 2011 classification). Also, all irrelevant data was removed, like monthly subtotals or immigration sub-categories.
- Mapping
  - Multiple datasets were used to accomplish some tasks. For example, we mapped countries from government datasets to GeoJSON centroids and global regions.

3. Goals and Tasks

Intended Tasks

Our overarching goal was to educate the general public about who Canadian PRs are, especially around election times when immigration policy is up for debate and may change depending on election results. We wanted our visualization to help viewers understand the facts and help voters make informed decisions.

More specifically, we wanted to facilitate the ability of viewers to:
- Discover the distribution of inbound PRs across countries of citizenship and identify any outliers
- Discover the trend of PR admissions over time
- Compare the relative PR intakes between provinces/territories and identify the extremes both over time and at a given time
- Lookup/browse (1) breakdowns of PR characteristics and (2) the trend of PR admissions over time for a province/territory

4. Visualization

Visualization Interface

Our interface is made up of four views: World Map, Line Chart, Canadian Map, and Stacked Bars. The Line Chart view acts as the "hub" of the interface, separating the World Map view above it from the Canadian Map and Stacked Bars views below it.

The Line Chart view is linked unidirectionally to the World Map and Stacked Bars views, and bidirectionally to the Canadian Map view. The Canadian Map view is also linked unidirectionally to the Stacked Bars view.
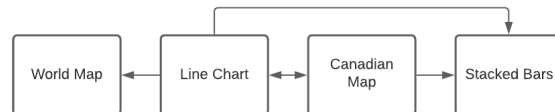
World Map View (Innovative View)

The World Map view visualizes the origins of admitted PRs admitted, by displaying arcs connecting their countries of citizenship to Canada. The goal of this view is to display the diversity of their origins, by showing a largely uniform distribution across 100 countries. It is intended to be consumed at a high level, but some surface-level details can be explored.
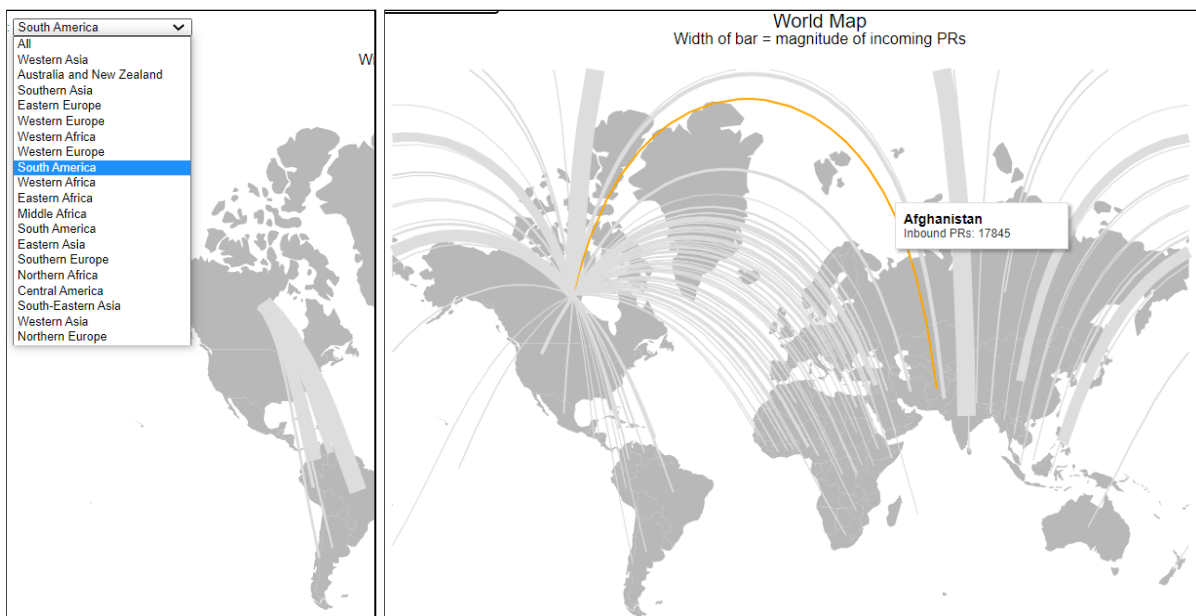
Marks

- **Line (connecting)** mark encodes PRs admitted into Canada
- **Interlocking area** mark encodes countries of citizenship

Channels

- **Width (1D) size** encodes amount of PRs from a given country
- **Spatial region** encodes geographic locations of countries

Interactivity

- **Dropdown (UI)** combats visual clutter and information density by allowing viewers to select a global region and only see the arcs for countries in that region. The arc widths also change based on the visible data, allowing them to identify the least and most influential countries within the selected region.
- **Tooltip** allows viewers to extract the exact number of inbound PRs and ensure that a given arc corresponds to a given country. Moreover, the arc color change on hover helps combat occlusion by allowing viewers to "scroll through" arcs and see each one.

Design Rationale

- We opted to only include half of the countries whose citizens were admitted into Canada as PRs to help reduce crowding on the map. The 100 countries that contributed the least total PRs were removed. Their totals were at least 5 times smaller than the average.
- The arcs intentionally connect to the center of each country, including Canada. This view focuses on the global distribution of inbound PRs which is based on confirmed citizenship information. This view purposely does not include the destinations of PRs (province/territory, city) as this information is based on intent and is too granular. In doing so, we have kept it as the Overview in the Overview/Detail Multiform relationship with the Canadian map.
- The data in this view is heavily geographical, and we chose to leverage a world map to best visualize it. The familiarity of the map makes it easy for viewers to quickly find regions/countries of interest to them and understand Canada's global intake of PRs.
- Arc widths are not great for direct comparisons between countries, but this wasn't our concern. We wanted to emphasize outliers, those countries with far larger amounts of inbound PRs, and arc widths do this effectively.
- We also opted to use arcs instead of direct straight lines to avoid the visual clutter that would result from directly overlapping line marks.

Line Chart View

The Line Chart view visualizes the admission of PRs into Canada over time, by displaying one data point per quarter. The slider acts as the x-axis of the graph, and allows viewers to choose a timeframe by selecting a starting and ending quarter. The goal of this view is to provide an overview of trends over time and act as the focal point of linkage across the entire interface.
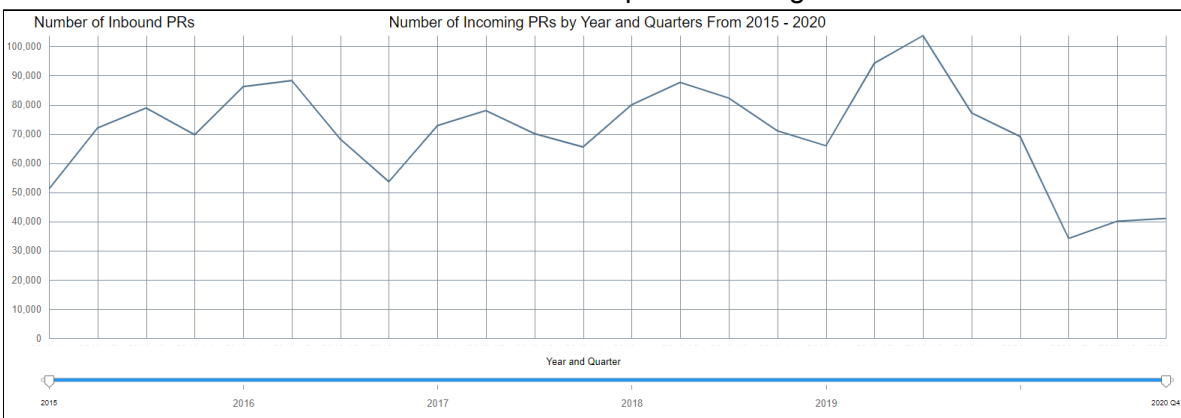
Marks

- **Line (connecting)** mark
- **Point** mark encodes
  PRs admitted into Canada

Channels

- **Position on common scale** encodes time (quarter)
- **Position on common scale** encodes number of inbound PRs

Interactivity

- **Slider (UI)** allows viewers to explore a particular time period between two quarters. Only the selected data is visible in each of the World Map, Canadian Map, and Stacked Bars views with the default being all available quarters. Note that the legend in the Canadian Map view changes depending on time frame as well.

Design Rationale

- A line chart is a perfect and obvious choice for visualizing trends over time, and by lining up the slider with the x-axis we are able to make it clear which timeframe is selected.
- We decided on one data point per quarter because it was a good middle ground between having monthly (too dense) or yearly (too sparse) data points.
- This view has a Multiform, Overview/Detail relationship with the other views.

Canadian Map View

The Canadian Map view visualizes the split of PRs between provinces/territories based on the amount of incoming PRs as a proportion of population. The goal of this view is to highlight any inequities in intake across the country, and allow viewers to dive deeper into specific regions.
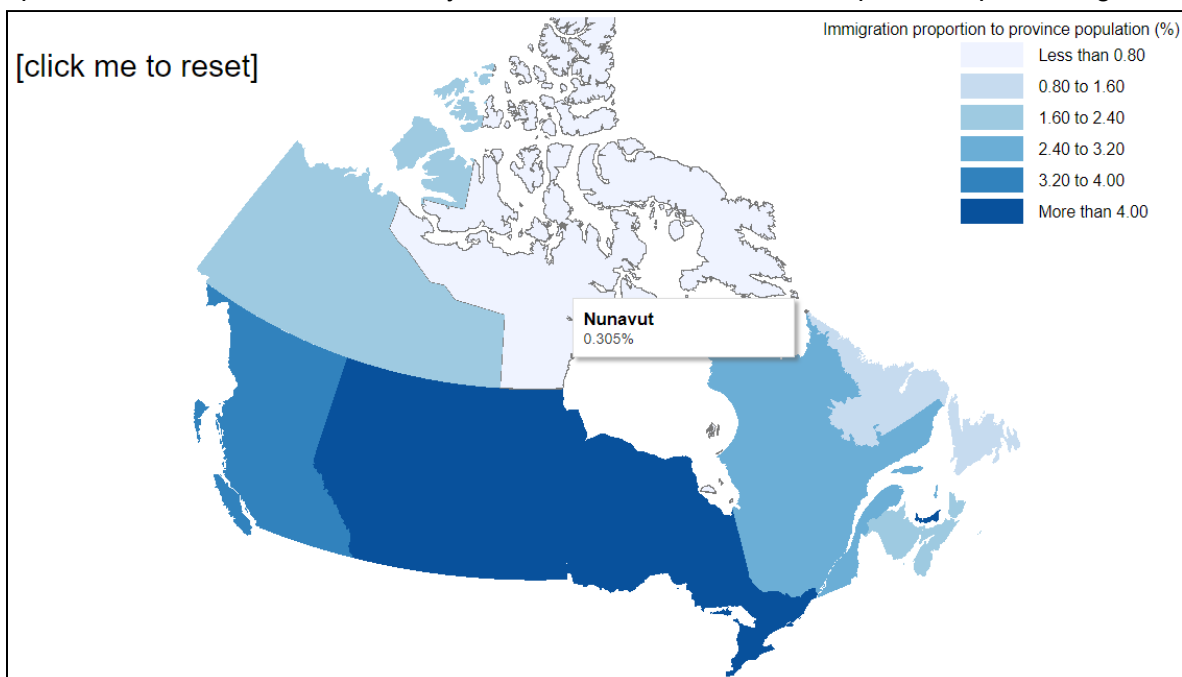
Marks

- **Interlocking area** mark encodes Canadian provinces/territories

Channels

- **Spatial region** encodes geographical location
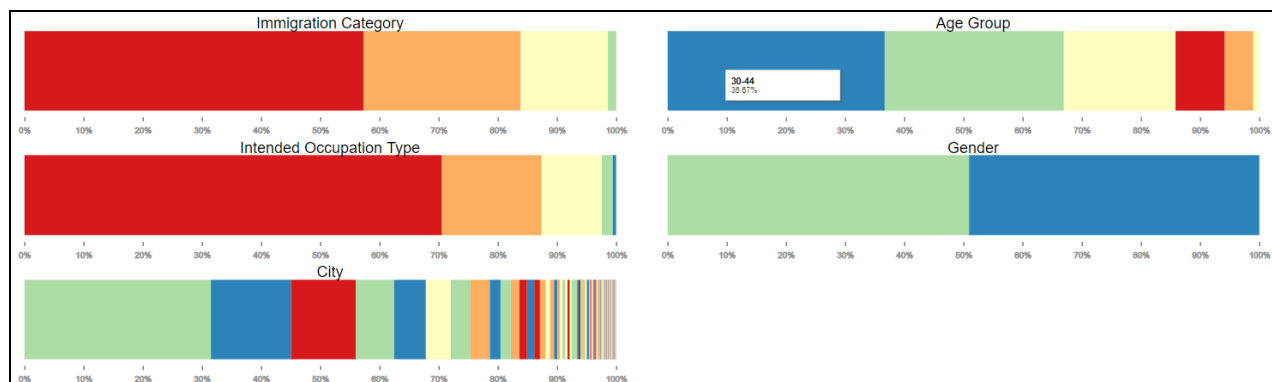- **Color saturation** encodes amount of PRs (% of population)

Interactivity

- **Tooltip** allows viewers to extract the exact amount of inbound PRs (as a percentage of population) and confirm that they are looking at the correct province/territory.
- Viewers are able to look up or browse details for a particular province/territory by clicking on it. This then filters the data in the Line Chart and Stacked Bars views to only include information on the PRs intending to live in that province/territory. This filtering is in addition to any filtering by time (quarter). By default, information for all of Canada is included, and viewers can return to this default by clicking on "[click me to reset]".

Design Rationale

- There is an inherent tradeoff between the familiarity of a choropleth map and the more accurate decoding of a symbol map. We opted for the former and felt it was a better fit, especially since there were a small number of bins that our sequential color palette needed to show.
- By having a dynamic legend, we are able to always display intake percentages that are relative between provinces/territories. The aim of this view is to see which areas are taking in less or more PRs than others, rather than to study absolute values. Also note that the bins shown by the legend are divided into equal sizes to avoid inaccuracy.
- For the same reason as above, we also chose to calculate intake amounts as percentages of population. Intuitively, it is more impactful for a small province like PEI to accept 4,000 PRs than a large province like Ontario.

Stacked Bars View

The Stacked Bars view visualizes more detailed breakdowns of PRs across categories like age, gender, and intended occupation. The goal of this view is to provide additional details about admitted PRs.

Marks

- **Line** mark encodes PRs admitted into Canada

- **Color hue** encodes each of: city, age group, immigration category, intended occupation type, gender

Channels

- **Width (1D) size** encodes relative amount of PRs

Interactivity

- **Tooltip** allows viewers to extract the exact amount of inbound PRs (as a percentage) in each category and acts as a legend as well to explain which mark is for which category.

Design Rationale

- We opted to include cities despite a large number of bins, particularly when the data is not filtered by province/territory. This is because this bar chart still clearly shows the cities taking in the most PRs, as well as the overall variety in cities. For instance, in Ontario, Toronto is the intended destination of almost 80% of PRs, but it is apparent that tens of other cities are also considered.
- We used occupation types rather than occupations, to aggregate and limit the number of bins in this bar chart to 5 instead of hundreds. We also used high-level immigration categories rather than low-level, also to reduce the number of bins to a more easily interpretable amount.
- It would be helpful to have persisting legends, and for the same colours to represent the same categories (e.g. for the gender chart, if male was always yellow and female was always green instead of refreshing).
- We used stacked bar charts because relative values as breakdowns were more important than absolute values and we wanted to keep the part-to-whole relationship.
- We acknowledge that the use of a categorical rainbow hue palette may not have been ideal. However, we chose to do so simply because of the different color saturations that allow the bins to be easily told apart.

Usage Scenario

Thomas wants to be an informed voter in the upcoming public election. Immigration is a particularly important issue for him. He lives in Moncton, NB which has a small population of under 100,000 people and he has been hearing a lot of grumbling about large increases in immigrants within his community. He wants to get his facts straight before deciding which party's immigration policy he agrees with most.

Thomas finds *Getting to Know Canadian PRs* and begins to explore.

The first thing he notices is that PRs are being admitted into Canada from all over the world. India seems to be a bit of an outlier with far more immigrants than other countries, but overall there is a lot of diversity in the citizenship of inbound PRs.
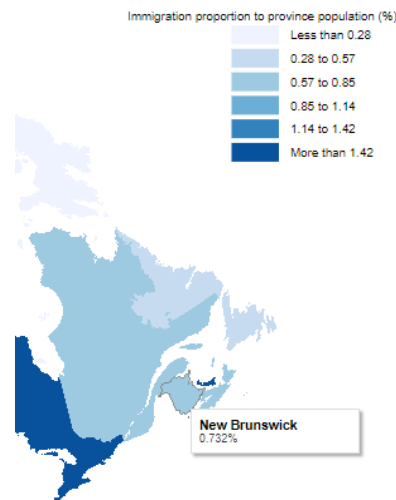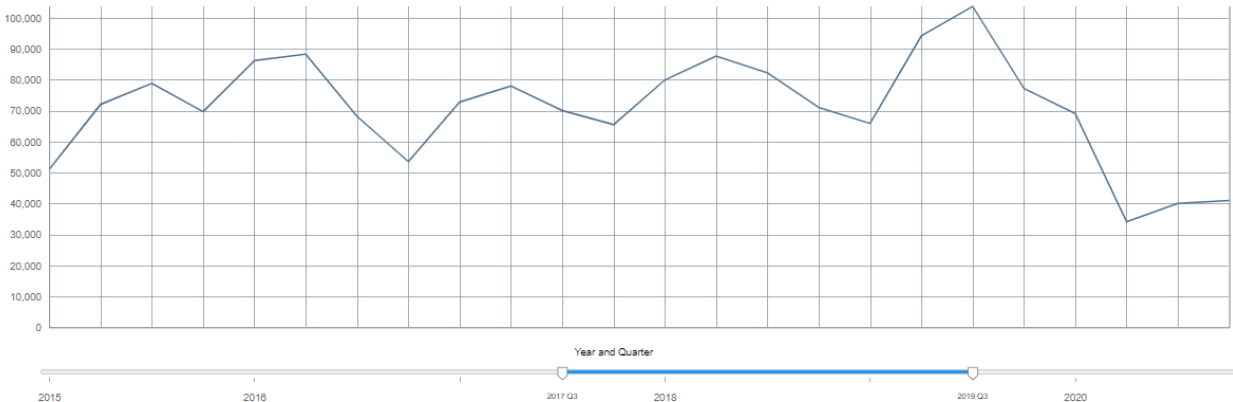
Out of interest, he takes a closer look at certain global regions like Western Europe.

India
Inbound PRs: 329075

Filter By Area of Inbound PRs: Western Europe

He then scrolls down and observes that the trend over the past few years has been fairly consistent, with some ups and downs along the way. It appears that there was a rather steep decline at the end of 2019, perhaps due to COVID-19. He's especially interested in the 2 years before, and so he sets the slider to start at 2017 Q3 and end at 2019 Q3.

Immigration proportion to province population (%)

Less than 0.28
0.28 to 0.57
0.57 to 0.85
0.85 to 1.14
1.14 to 1.42
More than 1.42

He scrolls down further and sees that his province of New Brunswick actually took in proportionally less (not more) PRs than many other provinces/territories at only 0.732% of the provincial population.

He clicks on New Brunswick and notices that the line chart has changed and now shows the trend specific to the province. He notices that there was actually an upward trend prior to COVID.
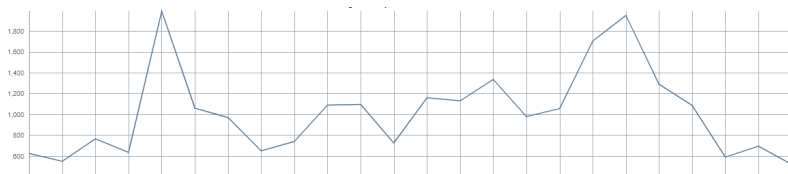
New Brunswick
0.732%

Finally, he scrolls down to the bottom of the page to see some more details about the PRs admitted into New Brunswick.
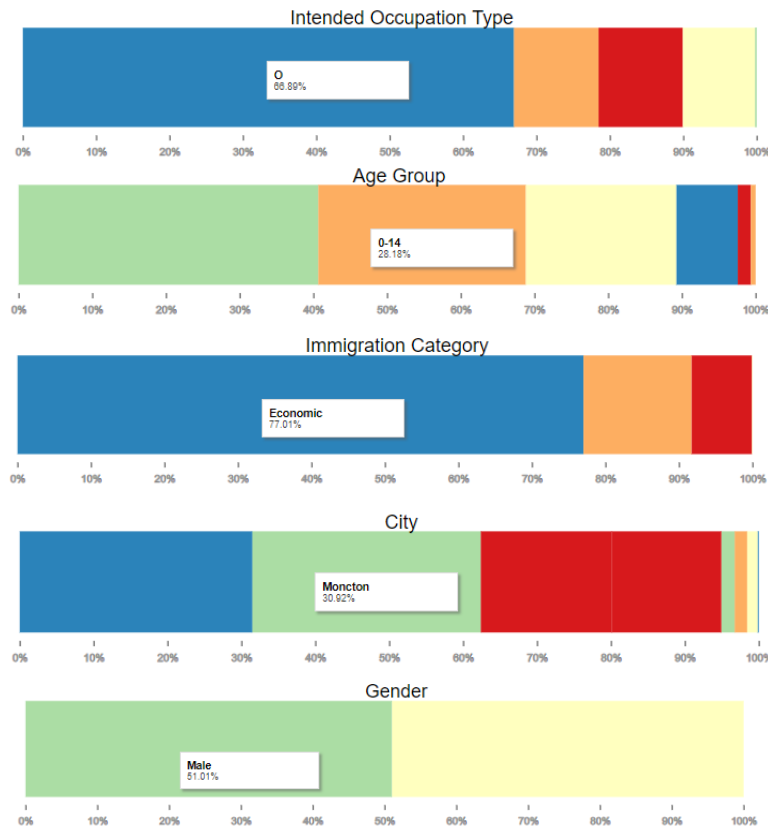
He is not surprised that his city of Moncton is the second most popular intended destination in his province, nor is he surprised that there is a roughly even split between male and female PRs.

He is concerned about the almost 70% of PRs who do not have a specific intended occupation type (Other) and plans to do some further research in this area. He is however, delighted to discover that almost 80% of PRs fall into the "Economic" immigration category and that many families are choosing New Brunswick for their new home, as children ages 0-14 make up the second largest age group of PRs.

Overall, Thomas has learned a lot and feels that he is now armed with the information he needs to decide which party's platform he aligns with well.

5. Credits

Sources

Use	Source	Explanation
Reference Only	D3 in Depth: Geographic	Learned how to use geoGenerator.centroid()
Reference Only	Github (phil-pedruco): countries.geo.json	Checked how to cross-reference two files in d3
Reference Only	CodeSandbox: d3-linked-charts-dispatcher	Referenced how to use dispatcher
Adapted tutorial code on bar	CodeSandbox: areachart.js CodeSandbox: d3-static-bar-chart	Line chart, bar chart Figured out how to separate out the

and line charts		different bins for bar chart and how to size bars based on length. Used our own data to plot length of bars Similarly, recalled how to plot points on coordinates but plotted our own data values
Reused global filter	CodeSandbox: Case Study on Measles and Vaccines	Reused global filter from case study with code from main.js. Adapted to match our own filtering conditions and capabilities
Reused range slider	Bl.ocks.org (johnwalley): d3-simple-slider Observable (mbostock): d3-simple-slider	Reused range slider code but adapted to use tick steps for purposes of our data.
World map code inspiration	D3.js Graph Gallery: Most basic connectionmap in d3.js	World map resource for our chart
Provinces code boilerplate	Social Innovation Simulation: Making Maps on D3 Bl.ocks.org (AldermanAxe): Canada Map	Boilerplate code to set up provinces chart and the path drawing of such provinces

6. Reflection

Project Development

Our project has followed a fairly straightforward and linear development path. We believe this is largely thanks to our ability to be realistic when planning, and our preference to underpromise and overdeliver rather than to overpromise and underdeliver. From the start, we set out to visualize an interesting topic, provide insight on it, and get a good grade, but we were never ambitious enough to try to make it into the Hall of Fame.

We began by individually brainstorming potential ideas (including relevant datasets), so that we could choose between a variety of viable options together. Once we chose our topic (Canadian PRs), we worked together to brainstorm and draft the proposal so that we would be on the same page. From there, we simply assigned tasks to each team member based on their preference, skills, and availability, and checked in regularly or when issues arose.

Changes to Visualization Goals

Our overarching visualization goals did not change throughout, but some minor goals did:
- The Candian Map was originally intended to be similar to a hexbin map, with each province/territory represented as an equally sized square. We later considered a symbol map as well, but ultimately (having taken our skillset and time constraints into consideration) we decided that a choropleth map was best suited for our intended tasks.

- The Stacked Bars view was originally made of pie charts, which we later learned were quite useless (especially in terms of comparison between parts of the whole) leading us to choose stacked bar charts instead.

Changes to Technical Goals

Again, our goals didn't change much, but some components originally left vague were decided:
- The dropdown in the World Map view was meant to filter by continent. We thought that a numeric filtering might be more helpful (e.g. countries with PRs > x), but opted to save on time and compromise by using the slightly more useful and more difficult region filter.
- The slider in the Line Chart view only allowed for selection of a single quarter at first, but we opted to invest the time and effort into the ability to select any amount of quarters. This was harder to implement, but we felt it provided significantly more value.

Other Questions

- Our original proposal was very realistic, and we didn't encounter any issues requiring major workarounds or leading us to abandon/replace a feature.
- If we were to begin a new project, we would likely invest more time at the start into the details rather than relying on us figuring it out as we go. There were times we thought we were all on the same page, but in reality we had all made different assumptions that were never explicitly stated. We would also prioritize a more accurate work breakdown.

7.Project Management & Team Assessment

Team Member Contributions

Milestone 1

Ansel	Adrian	Alvin
● Overview and usage scenario	● Data description ● Visualization description ● Sketches	● Work breakdown and schedule

Milestone 2

● Initialization of views	● Data processing	● Writeup

Milestone 3

● Programming formatting design/changes ● Code cleanup/commenting	● Final report ● Minor coding support	● Implementing interactivity and data filtering ● Code cleanup/commenting

Work Breakdown & Schedule

Overall, we were able to stick to our weekly deadlines, but some tasks took either less (green cells) or more (yellow cells) time than initially planned. Tasks in red are new additions that we did not explicitly allocate time to, but did end up needing a fair amount of it.

In general:
- Our time estimates were good - we were able to use time saved on certain tasks towards other tasks that needed more. We feel that our estimates were reasonable.
- It took us longer than anticipated (and multiple attempts) to wrangle all the data into a consistent format ready for consumption.
- We should have foreseen that there would be a fair amount of bugs given the complexity of this project compared to individual assignments.
- Write ups were often left as the last task of a milestone, with the idea that the more complete the code is the easier it'll be to describe it. However, it would have been better to make a template/skeleton early and fill it in as we go.

Week of	Ansel	Adrian	Alvin
March 8th	Proposal (3,4, and 6) Formal discussion on how our CSV should look like	Proposal (3,4, and 5) Formal discussion on how our CSV should look like	Proposal (3,4, and 6) Formal discussion on how our CSV should look like
Hours (Estimated)	5	8	5
Hours (Actual)	7	7	8
March 15th	Development of time slider and proportionate geographical chart First draft development of all charts	Development of pie charts and proportionate geographical chart Comprehensive CSV data preprocessing and cleaning	Finding code written in previous assignments that can be re-used or referenced to Development of immigration movement to Canada chart
Hours (Estimated)	10	10	10
Hours (Actual)	13	10	5
March 22nd	Ensure and check that best Github practices (pull requests, branches, etc) are being met Finish development of time slider and proportionate geographical chart Review Alvin's work for internal peer review First draft development of all charts	Finish development of pie charts and proportionate geographical chart Review Ansel's work for internal peer review Comprehensive CSV data preprocessing and cleaning	Finish development of pie charts and proportionate geographical chart Review Adrian's work for internal peer review
Hours (Estimated)	12	10	10
Hours (Actual)	15	8	5

March 29th	Polishing of assigned views for WIP milestone Tooltip development for pie charts Work with Adrian to ensure the main proportional view chart is finished	Polishing of assigned views for WIP milestone Evaluate timeline of tasks to ensure we are on track and suggest changes if behind/ahead of schedule Submit Milestone 2 Start on report and project planning	Polishing of views for WIP milestone Build references README to ensure we are citing work where necessary and giving credit as appropriately Development of tooltip for immigrant flows chart Interactivity linking of all charts and slider work
Hours (Estimated)	10	10	12
Hours (Actual)	5	6	15
April 5th	Finish tooltip development for pie charts Review Alvin and Adrian's work on immigration flows chart tooltip development and draft a series of quality checks Various bug fixes	*Final Report comprehensive compilation* *Draft presentation style and script for the final presentation conducted virtually over Zoom to course 436V staff*	Review Ansel's work on pie chart tooltip development and draft a series of quality checks Submit M3 to Canvas and appropriate Github repo Interactivity linking of all charts and slider work
Hours (Estimated)	10	10	9
Hours (Actual)	8	11	10
April 12th	Practice presentation	Practice presentation	Practice presentation
Hours (Estimated)	3	3	3
Hours (Actual)	0.5	0.5	0.5