

Developing Data Products: The best Canadian state to live - based on the individual income statistics

Background

This is an example of Shiny web app with associated supporting documentation, aiming to answer to the following question:

What is the best Canadian state to live, according to the desired salary?

User can **Enter the desired salary**, and after hitting **Submit**, the dashboard will show number of tax filers per each province which obtained that salary. In addition, user can see other important information obtained from data manipulations: **Income per Source**, **Income per Range**, as well as three Key Performance Indicator (KPI), namely **Top Income Range**, **Total Number of Tax Filers**, top **Predicted Income** per source.

Data

For this web app, data representing *Individual statistics by tax filing method for all returns filed, economic characteristics, 2014 tax year, Canada Revenue Agency* - recorded by the *Open Data Government of Canada* and released at 2018-06-27 (downloaded on January 25, 2019): <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1110004601> were sorted, cleaned and manipulated.

These are most recent data on this subject (to date), published by the Government of Canada. The original file “11100046.csv” contains 18241 observations, with data for the year 2014, described by the following dimensions: Geography (161 items: St. John’s, Newfoundland and Labrador; Bay Roberts, Newfoundland and Labrador; Grand Falls-Windsor, Newfoundland and Labrador; Corner Brook, Newfoundland and Labrador; etc.); Filing method (Paper; Netfile; Efile); Major source of income (7 items: Employment; Investment; Pension; Self-employment; etc.); Income range (4 items: Less than \$25,000; Between \$25,000 and \$49,999; Between \$50,000 and \$99,999; \$100,000 and Above); Complexity of T1 return (2 items: Simple; Complex); Number of Tax filers, etc.

Data from the original file were sorted and summarized, such that cities from “Geography” variable, were grouped by provinces, taking the name of the new variable “State”. Some data from territories were not included as population is small and hence number of tax filers is not significant compared with other provinces.

The finalized data file used for this app, namely **Income_Canada_2014.csv**, has 16,635 observations and 5 variables: **State**, **Source.Income**, **Income.Range**, **Number.Tax.Filers**, **Average.Income**.

Because subsettings are used in this study, find below some details about variables:

- **State** has 12 categories: *AB, BC, MB, NB, NL, NS, NT, ON, PE, QC, SK, YT*.
- **Source.Income** has 7 categories: *Employment, Investment, Pension, Self-employment, Other sources of income, Multiple sources of income, No income*.
- **Income.Range** has 4 categories: *Less than \$25,000; Between \$25,000 and \$49,999; Between \$50,000 and \$99,999; \$100,000 and Above*.
- **Average.Income** was used for building reactive output, in order to have a numeric value for income, which can be compared with salary inserted by the user.

App Overview

- Both essential parts of a Shiny app: ui.R and server.R were build. They are available in the github repository: <https://github.com/lilianabraescu/Developing-Data-Products-The-best-Canadian-state-to-live>

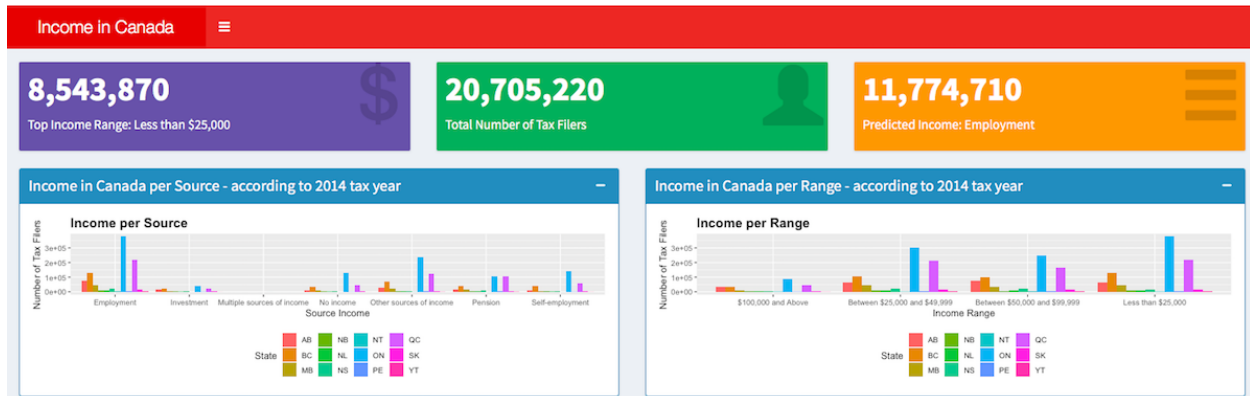


Figure 1: Income per Source (left) & Income per Range (right)

- Some screenshots of the app:
- *Income per Source* and *Income per Range* are plotted for all tax filers from Canada (2014 tax year), and the KPIs *Top Income Range*, *Total Number of Tax Filers*, top *Predicted Income* per source were computed.

Plot show that Ontario (ON) and Quebec (QC) are in the top for all income sourcesm and all income ranges.

- *What is the best Canadian state to live, if user enter \$50,000 (by default) desired salary?*
- *What is the best Canadian state to live, if user enter \$110,000 desired salary?*

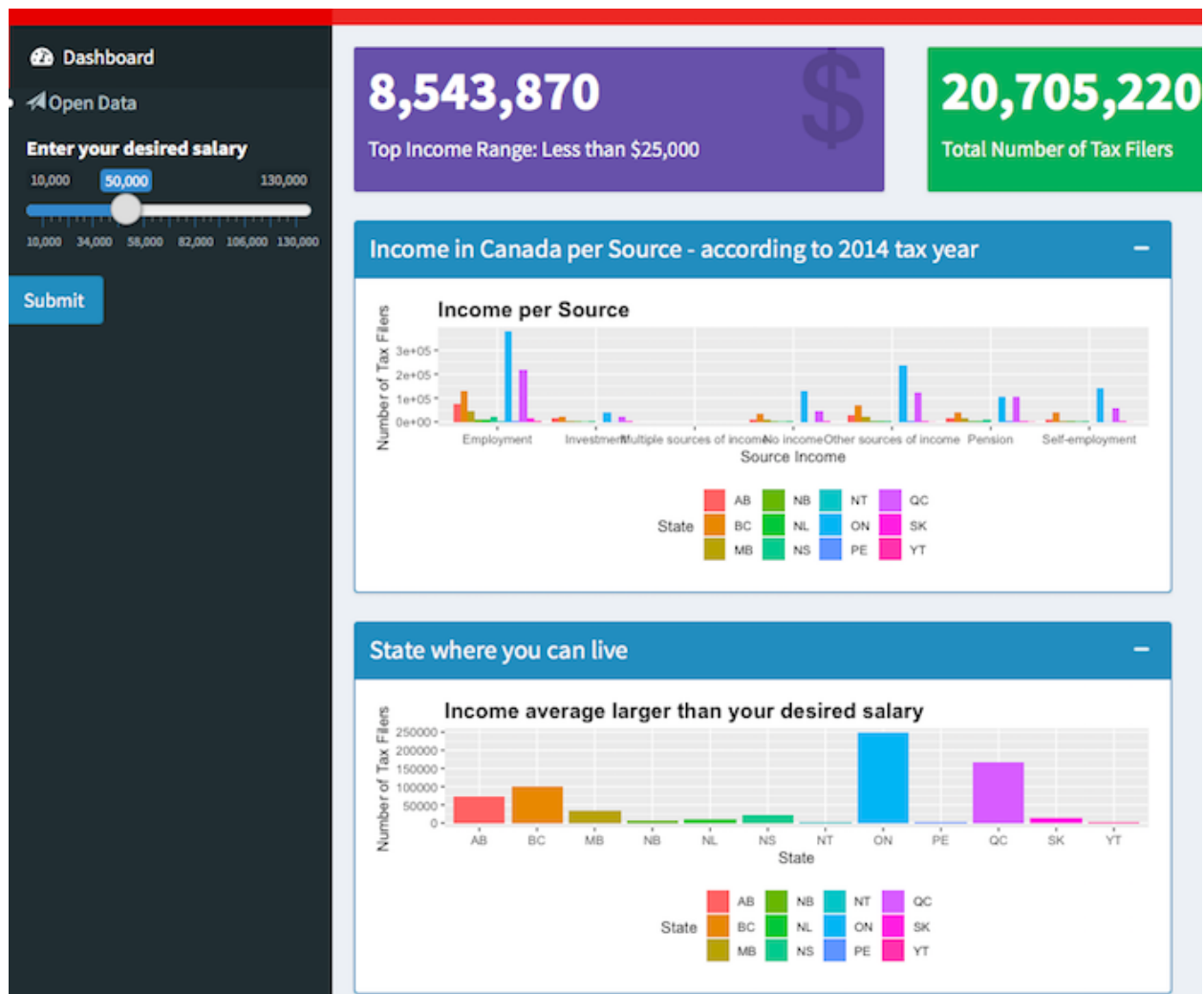


Figure 2: Income average larger than \$50,000 (bottom)

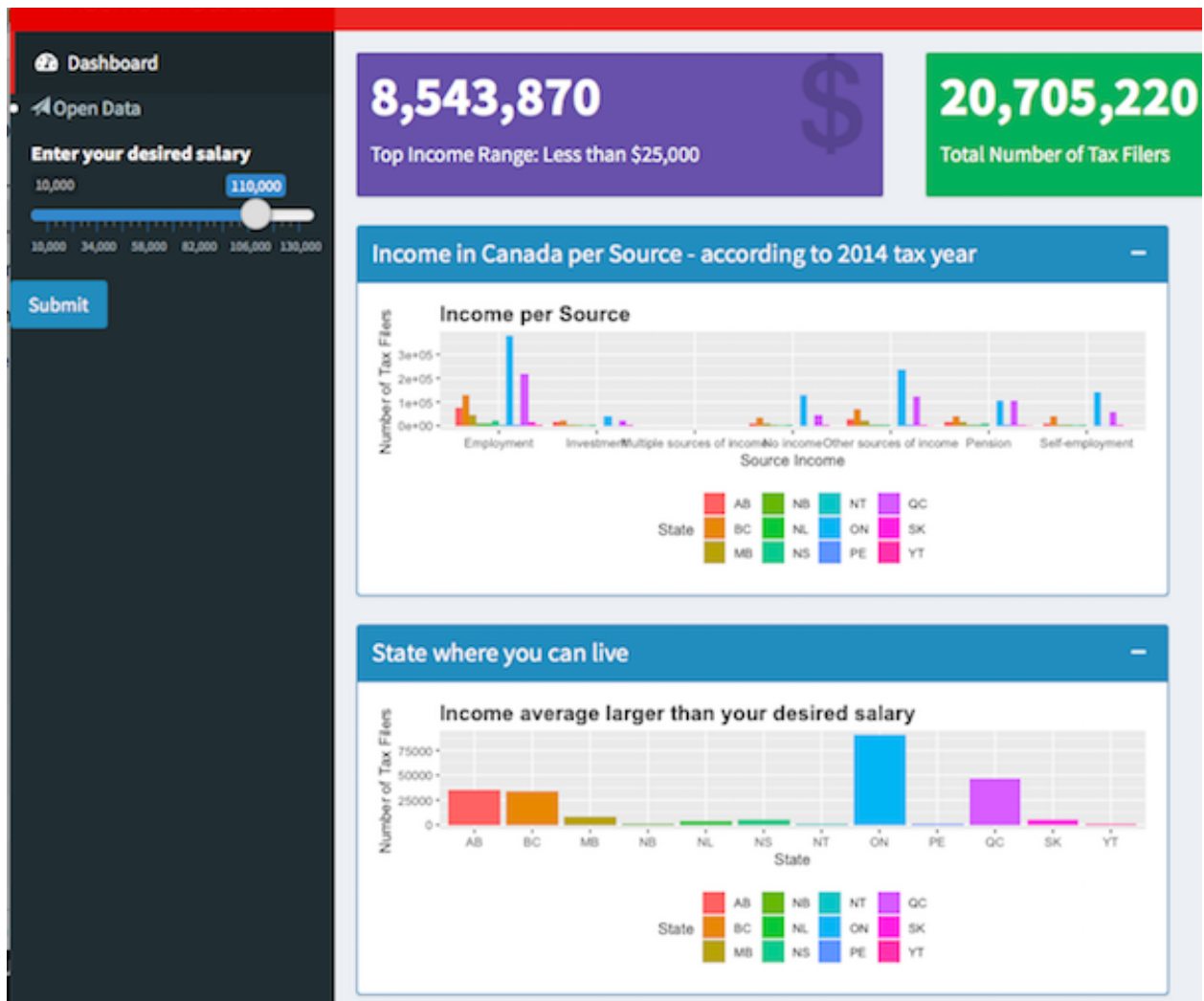


Figure 3: Income average larger than \$110,000 (bottom)