Technical Documentation: BC Demographic Survey: DIP Data Linkage Rates

Table of Contents

PROJECT: Evaluating the BC Demographic Survey Data in the DIP

PREPARED BY:
Amelia Lowery & Julie Hawkins & Lindsay Fredrick
Data Science Partnerships Program
BC Stats BC Data Systems
Ministry of Citizens' Services

DATE: 2024-04-19

Purpose

This document summarizes...

Project Overview

Methodology

Overview

Linked Variables Summary

Caveats

Linked Individual Demographics

Caveats

Software

This analysis is implemented in the R and python programming languages Van Rossum and Drake Jr (1995). The code used to generate this analysis was reviewed by three data scientists. Key tools used to complete this work include the Apache Arrow project (Richardson et al. 2021), the tidyverse (Wickham et al. 2019) and the interal SAE package dippy (**dippy?**).

GitLab

All code is stored under the git version control system and shared inside the secure environment in these GitLab repos:

- Creation of Demographic Survey Variables:
  https://projectsc.popdata.bc.ca/shares/arda-demographic-survey
- Creation of Linkage Statistics:
  https://projectsc.popdata.bc.ca/shares/linkage-litmus-test

References

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Jonathan Keane, Romain François, Jeroen Ooms, and Apache Arrow. 2021. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Van Rossum, Guido, and Fred L Drake Jr. 1995. *Python Tutorial*. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.