Technical Documentation: Preventing & Reducing Homelessness Integrated Data Project

Table of Contents

Purpose 2
Background Materials 2
Study Population Attributes 2
Flows Into and Out of Homelessness 2
Appendix 4
Software 4
Raw Data 4
References 6

PROJECT: Preventing & Reducing Homelessness Integrated Data Project

PREPARED BY:
Data Science Partnerships Program
Digital Platforms and Data Division
Office of the Chief Information Officer
Ministry of Citizens' Services

DATE: 2022-04-05

DISCLAIMER: All inferences, opinions, and conclusions drawn in this data summary report and data slides are those of the authors, and do not reflect the opinions or policies of the Data Innovation Program or the Province of British Columbia.

Purpose

This document summarizes the analytical methods used to generate the *Preventing and Reducing Homelessness Integrated Data* project estimate of the homeless population of British Columbia with data from the year 2021. It supplements the 2019 and 2020 technical documents where new analytical methods were used. For a comprehensive overview of analytics methods used, please see the 2019 and 2020 Technical Documents. This update remains focused on implementing a cross-agency analytic definition of homelessness based on administrative data.

Background Materials

Study Population Attributes

Demography

This project had three sources of age and gender data for the study population: BCH shelter clients data, SDPR Income Assistance data and the B.C. Central Demographic data. This section outlines the hierarchy of those data sources and the process to determine which to use.

Date of Birth

As in 2019 and 2020, age was calculated as the age (as an integer) as of December 31 which is the end of the definition period. To meet conservative statistical disclosure goals, ages were grouped into three groups: 24 and under, 25 to 55, and 55 and over. Two additional age categories were created and used for the 2021 analysis: 24 and under, 25-34, 35-44, 45-54, and 55 and over. To compare age distributions of the homelessness population relative to the general population, population estimates by age from Statistics Canada were used.

Flows Into and Out of Homelessness

Three movement categories were defined to aid in describing how individuals cycle into and out of homelessness cohort over the 3 years of data (2019-2021) available:

1. *new_arrival* represents those individuals who had not been identified in the homeless cohort data in any of the previous years.
2. *left_cohort* represents those individuals who were identified in the homeless cohort data in 2019 or 2020 but not in subsequent years of the data
3. *re_entered* represents those individuals who were identified in the homeless cohort in 2019, were not identified in the cohort in 2020 and then returned to the cohort in 2021.

Due to the limited years of homeless cohort data available, an unknown number of individuals were identified as *new_arrival* in 2019, *left_cohort* in 2021 and *re_entered* in 2019 and 2020. A value of 0 was chosen to represent the group totals in calculations.

Some movement throughout time within sub-populations is expected. It was generally handled by calculating an individual's movement grouping for a given year before assigning their *age*, *gender* and *geograph*ic region for that year. An individual's most recent gender was used as their assigned gender indicator for all years. The region in which an individual experienced the most number of months of homelessness was assigned their geographic region for that year. By following this method, movement group subtotals also captured movement within a subpopulation over time.

Appendix

This appendix is provided as specific documentation of the data analysis conducted on provisioned Data Innovation Program data in the secure analytics environment.

Software

This analysis is implemented in the R programming language (R Core Team 2021). 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), cansim (von Bergmann and Shkolnik 2021), dipr (Albers and Hazlitt 2020) and the R package targets (Landau 2021) for project organization. All code is stored under the git version control system and shared inside the secure environment in these GitLab repos:

- Parquet and restating: https://projectsc.popdata.bc.ca/shares/hl-data-to-parquet
- Application of definition and creation of output group: https://projectsc.popdata.bc.ca/shares/hl-cohort

Raw Data

All data was converted from compressed fixed width files into parquet files for ease of analysis. Significant testing against Population Data B.C. data provisioning metrics occurred to ensure that conversions were done accurately.

Social Development and Poverty Reduction Data

Involvement data

- filenames: “idosdpr2018-2019.bceainvolvement.A.dat.gz”, “idosdpr2020.bceainvolvement.A.dat.gz”
- columns used: ym, fileid, deprltncd, birthdt_yymm

NFA data

- filenames: “idosdpr2018-2019.bceanfa.B.dat.gz”, “idosdpr2020.bceanfa.B.dat.gz”
- columns used: ym, fileid, nfa, csdname
- subsetting: only rows where nfa == 1

BC Housing Data

Client Data

- filename: “bchousing_hifis2017-2020.hifis_clients.A.dat.gz”
- columns used: gender, dobyyyy, dobmm

Shelter Stays Data

- filename: “bchousing_hifis2017-2020.hifis_clnts_shlt_stays.B.dat.gz”
- columns used: shelter_stay_start_date, shelter_stay_end_date, shelter_org_id, shelter_census_sub_division, clientid

Shelter Attributes

- filename: Shelter attributes_HIFIS.xlsx
- columns used: org_id, shelter_type

Extra ID data - filename: “bchousing_hifis2017-2020.hifis_extra_clntid_popid_xlk.A.dat.gz” - columns used: clientid

Demographic data

- filename: “demographics1986-2020.B.dat.gz”
- columns used: sex, dobyyyy, dobmm

References

Albers, Sam, and Stephanie Hazlitt. 2020. *Dipr: Provide Functions to Efficiently Import SRE Data*.

Landau, William Michael. 2021. "The Targets r Package: A Dynamic Make-Like Function-Oriented Pipeline Toolkit for Reproducibility and High-Performance Computing." *Journal of Open Source Software* 6 (57): 2959. https://doi.org/10.21105/joss.02959.

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.

von Bergmann, Jens, and Dmitry Shkolnik. 2021. *Cansim: Accessing Statistics Canada Data Table and Vectors*. https://CRAN.R-project.org/package=cansim.

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.

BC Housing Corporation. [creator]. Private Market Rent Supplements for Eligible Seniors and Working Families. E04. Data Innovation Program, Province of British Columbia [publisher]. Data Extract. Approver Year (2019).

BC Housing Corporation. [creator]. Shelter and Homeless Outreach Private Market Rent Supplements. E03a. Data Innovation Program, Province of British Columbia [publisher]. Data Extract. Approver Year (2019).

Ministry of Social Development and Poverty Reduction. [creator]. BC Employment and Assistance (BCEA). E05. Data Innovation Program, Province of British Columbia [publisher]. Data Extract. Approver Year (2019).

Ministry of Health. [creator]. BC Vital Events and Statistics. E03. Data Innovation Program, Province of British Columbia [publisher]. Data Extract. Approver Year (2019).

Ministry of Health. [creator]. Central Demographics Files. E03. Data Innovation Program, Province of British Columbia [publisher]. Data Extract. Approver Year (2019).

Ministry of Health. [creator]. Registration and Premium Billing (RPBLite). E02. Data Innovation Program, Province of British Columbia [publisher]. Data Extract. Approver Year (2019).

Ministry of Health. [creator]. COVID Testing Data. E02. Data Innovation Program, Province of British Columbia [publisher]. Data Extract. Approver Year (2019).