Economic Commission for Europe

Conference of European Statisticians
Group of Experts on Measuring Poverty and Inequality
Geneva, Switzerland, 8-9 December 2022
Workshop on Harmonization of Poverty Statistics to Measure SDG 1 and 10
Geneva, Switzerland, 7 December 2022

Agenda item: Social policies and data

Providing new statistics for population often missing from statistical system in Canada

Note by Statistics Canada[1]

Abstract

In the effort to develop statistics through household surveys, census, and administrative data, it is clear that some sub-populations are more difficult than others to study. In fact, because they are difficult to reach and include, some sub-populations may be excluded from the statistical picture altogether. This report describes three recent initiatives at Statistics Canada that demonstrate how new strategies can provide more information about populations at risk of exclusion from the traditional statistical methods. The first initiative assesses the coverage of vulnerable populations in administrative and census datasets to document the number of people who are not represented in the tax filing population and/or census enumeration in the years leading up to their deaths. A second initiative explores the implementation of a low-income measure for the population living in the Canada's northern territories and for First Nations living on reserve, which is now included in standard products from the 2021 Census. A third initiative combines data from the Canadian census with multiple administrative data sources to calculate annual labour market indicators for the population living on reserve not covered in the monthly Labour Force Survey.

---

[1] Andrew Heisz, Christine Laporte, Alex Miller and Eric Olson, Statistics Canada

Introduction

In their data collection, statistical agencies always make every effort to reach as much of the population as possible. The ability to use administrative data as well as data integration has made it possible to obtain relevant information that has often been difficult to obtain. However, some populations are still more challenging to describe than others.

In Canada, hard-to-enumerate groups include Canadians who speak neither English nor French, those who are uncomfortable with online questionnaires, single-person households, vulnerable populations and mobile populations (Statistics Canada, 2020). These populations may include recent immigrants and non-permanent residents. Young people, especially single men, can be very mobile, and are less likely to participate in surveys or to file taxes (Messacar, 2018). Some vulnerable persons such as older persons living alone, people with health concerns and persons living in unsafe neighborhoods communicating with a stranger like a census employee may also be challenging. Populations living in First Nations communities (also known as "on-reserve") and in the Northern territories of Canada are often excluded from surveys due to their remoteness or other reasons. Finally, populations that live in remote areas can also be hard to reach.

The objective of this report is to describe three recent initiatives at Statistics Canada (including one conducted by partners in academia with Statistics Canada's assistance). These demonstrate how new data strategies and simple adjustments can provide enhanced information about populations at risk of exclusion from the traditional collection methods.

1. Using vital statistics to assess the coverage of vulnerable populations in administrative and census data

Those who are living in poverty, those who are homeless and those who are living on the margin of the society often experience significant barriers to participation in surveys or the census and may fail to file tax returns. As a result, they may be absent from many government statistics, making it difficult to pinpoint the size and composition of this group. In turn, those absent from administrative data face an added challenge in Canada as taxfiling serves as a gateway for the distribution of a number of transfer payments to individuals and therefore, some vulnerable persons may not access benefits to which they are entitled and would need. In short, their absence from the statistical data limits the ability of policymaking to address the difficulties facing those most in need. Identifying the sociodemographic characteristics and the proportion of Canadian that fall into this category is difficult, as, by definition, they are missing from the data.

Record linkage provides one way to assess the absence of persons from one-file or another. In several instances Statistics Canada has undertaken record linkages for this purpose and studied the patterns, or provided the data to academic researchers to do the same. In one example, Green et al. (2021) searched for the existence of presence and absence patterns from government data sets such as administrative tax forms and Census[2]. They start from the premise that while people may be absent from Census and tax filings, the vital statistics records should be relatively complete in recording deaths. Therefore, death

---

[2] In a second example, Sanmartin et al. (2021) explored the linkage rates between the numerous data sources used to build an analytic file to study overdose deaths. One finding to come from this study was that only 40% of persons who died from overdose could be linked to annual employment (T4) and/or social assistance (T5007) files.

records may provide some insight into the size and characteristics of populations missing from other data.

Figure 1: Using death records to identify those missing from census and tax
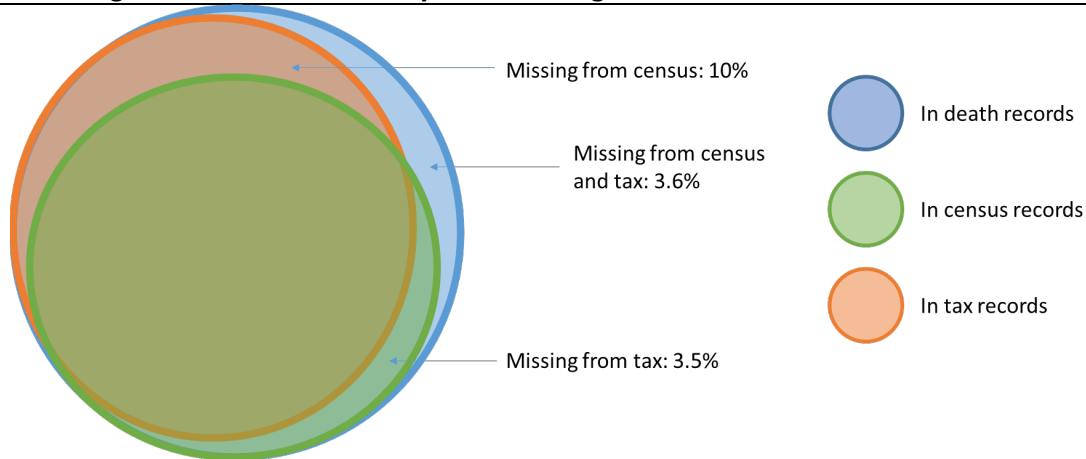
Source: Adapted from Green et al., 2021.

Their research proceeded in several steps. First, they estimated the number of individuals present in the vital statistics database who were not present in the administrative data prepared through the tax and transfer systems[3]. Then, they investigated the proportion of these individuals that died from "deaths of despair"[4] (deaths due to suicide, overdose and alcoholic liver diseases). Their results found 3.5% of people were in census but not the tax data, 10% of people were in the tax (T1) data but not in census and 3.6% of deaths (CVSD) were in neither census nor tax. Among persons dying from deaths of despair, early results found that deaths of despair make up 4.5% of deaths in British Columbia compared to 17% among the population absent from both census and tax filings. Figure 1 provides an illustration or presence or absence of individuals in each data set. In this Venn diagram, overlaps signify of people common to datasets, while boundaries signify those not common. The tax file is represented by two circles as two different versions of the tax file were examined.

Pinpointing the socio-demographic characteristics of people who die pre-maturely from deaths of despairs could improve the ability of policymakers to preemptively identify those who are most vulnerable and ultimately prevent unnecessary deaths. Being able to first identify those individuals in a systematic way, is an innovation that opens new possibilities in health policy and harm reduction.

2. Adding the low-income measure for the on-reserve and territorial population in Census 2021

In the next two examples, we describe how Statistics Canada has sought to improve the number of statistics available for First Nations people who live on reserve, and people living in the Northern territories of Canada[5]. Over the years, Statistics Canada has developed partnerships with national

---

[3] In this case, it is possible that persons may be present in two datasets but are marked as absent due to a failure to link the records due to incomplete information. Record linkages are usually conducted on name, address and date of birth variables or using government identifiers.

[4] These are defined by Case and Deaton (2021) as deaths due to suicide, overdose and alcoholic liver diseases.

[5] By way of background, the 2021 Census counted 1.8 million Indigenous people, accounting for 5.0% of the total population in Canada including more than 1 million First Nations people (1,048,405), 624,220 Métis and 70,545 Inuit living in Canada. Among status First Nations people, just over two-fifths (40.6%) lived on reserve in 2021.

Indigenous organisations and local leadership in order to reach their communities and encourage their participation in the statistical system, particularly through the quinquennial census of population. However, the Indigenous population remains difficult to reach - especially those residing in First Nations communities (also known as "on reserve"). Figure 2 shows the tremendous geographic diversity of reserves.

Until Census 2021, low-income statistics have not been made widely available for persons living on reserve or in the Northern territories. Heisz (2019) and Harding and St-Denis (2021) introduced the concept of measuring low income (based upon the international standard 50% of median equivalent household income, which in Canada is called the "low-income measure" (LIM)) using the 2011 National Household Survey. Up to that point, low-income statistics were not available in these regions due to the idea that substantial in-kind transfers (for example, subsidized housing) and economic activities based on sharing and own consumption such as fishing, hunting and farming would make interpretation of the measurement difficult. Of course, these considerations would apply to any area of Canada, even though they might apply more so in the previously excluded areas, and this reason was not considered to be strong enough to continue to exclude these areas[6].

Figure 2: Location of First Nations Communities (reserves – green circles) and Inuit Settlements (white circles)
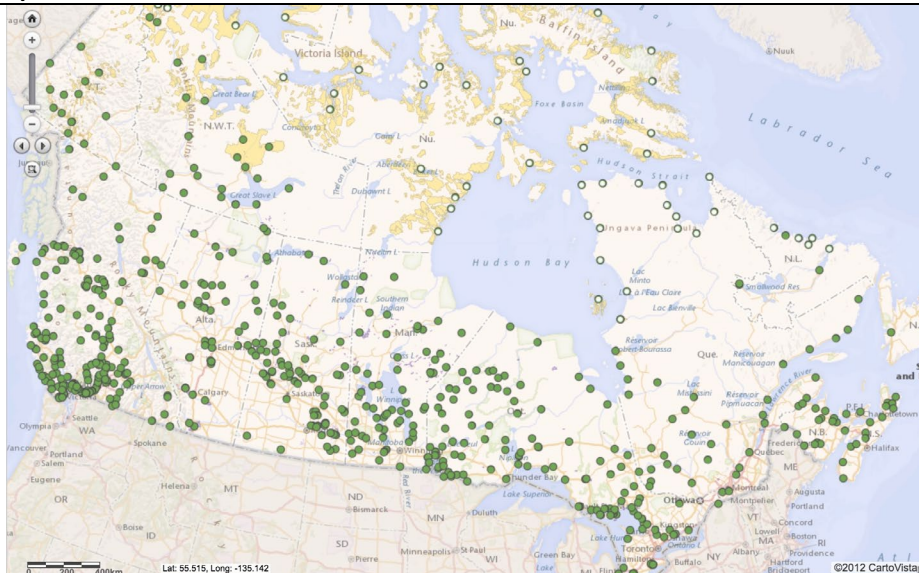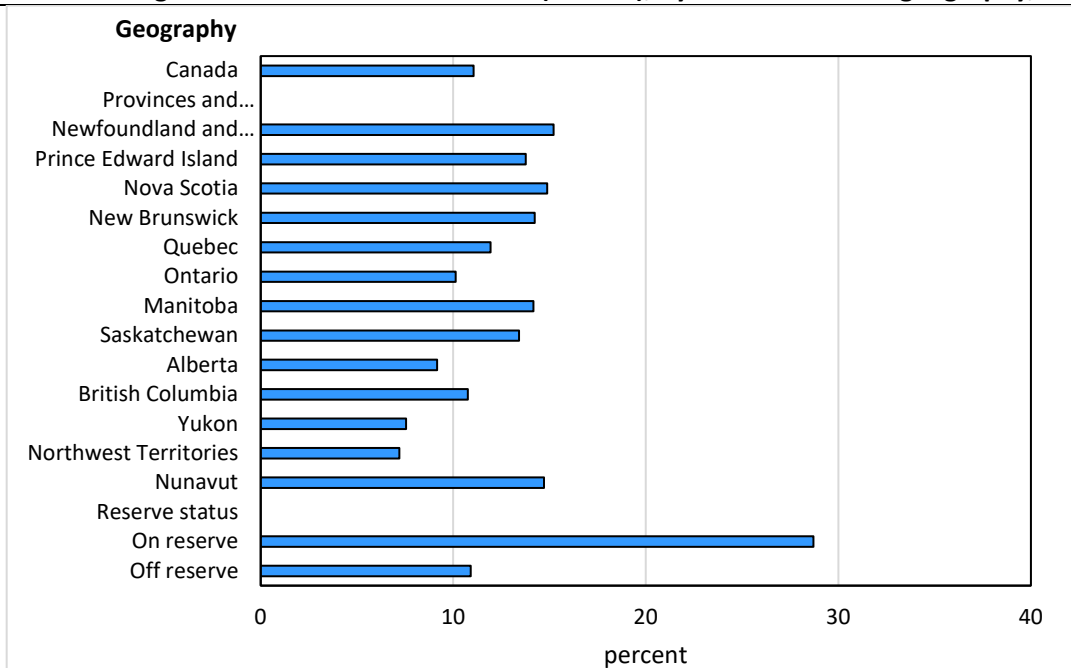

Source: Statistics Canada

There are three northern territories in Canada: Yukon territory (population=40,232 in 2021), Northwest territory (41,070) and Nunavut (36,858); Canada's total enumerated population was 36,991,981).

[6] Canada has other measures of low income that are produced regularly as part of the income program. Notably, Canada's official poverty rate, the Market Basket Measure is not yet available for the on-reserve geography. It is available for the Yukon and Northwest territories, and a version is currently being developed for Nunavut (Gustajtis, Lam and McDermott, 2021). The Market Basket Measure was declared Canada's official measure of poverty in 2018. The MBM thresholds are defined as the cost of a basket of goods and services required for a modest basic standard of living for a reference family of two adults and two children. The MBM threshold are adjusted for other family size using the equivalence scale, i.e., the square root of economic family size. The thresholds are calculated for 53 different geographic areas, in part, from local prices of goods and services.

As part of the 2021 Census program, Statistics Canada extended the coverage of the LIM to the on-reserve and territorial population in standard products, introducing the LIM as an initial low-income concept for the on-reserve and territorial population.[7] Increased coverage demonstrates an elevation of the circumstances experienced on reserve and in the territories. This new statistic provides an important reference point to the rest of Canada.

Figure 2 shows the percentage of respondents in low income by geography of residence for Canada, provinces, and territories as well as for those residing on and off reserves. Important geographic variation can be observed. The low-income rate for Canada is 11.1%. The on-reserve population has a low-income rate of 28.7%, three times higher than the total Canadian population. The values for the territories also show significant contrasts: the populations in Yukon (7.6%) and the Northwest Territories (7.2%) are below the national proportion; however, in Nunavut (14.7%) is above.

Figure 3: Percentage of individuals in low-income (LIM-AT), by selected level of geography, 2020

Source: Statistics Canada, Census of Population, 2021

Important considerations when using the LIM on reserve or in the territories

Part of the move to including the territories and reserves to the Low-income Measure included providing documentation regarding interpretation. Statistics Canada described four potential considerations regarding data quality and interpretation issues in using the LIM for the population residing in the territories and on reserve.

---

[7] From 2011, customs tables were made on request for the on-reserve and territorial population using the census or the National Household Survey.

1. **Regional differences:** The LIM does not consider the cost of living and therefore cannot take into account regional differences in the capacity of some population to attain a specific standard of living.
2. **Income measures:** The standard definition of income by Statistics Canada does not include the value of goods or services received in-kind, such as subsidies to housing and does not include the value of goods and services produced for own consumption, such as food produced from hunting or fishing. These values might however be significant for Indigenous or Northern communities.
3. **Incomplete enumeration and non-response:** Challenges in data collection such as non-response and incomplete enumeration of reserves must be considered when interpreting low-income statistics. For example, in 2021, there were 63 incompletely enumerated reserves. If the population living in those communities are different from other communities, the observed results may not be generalizable to the full population of interest.   In a similar way, if non-responding households are significantly different from responding ones, there are risks of bias.
4. **Quality of income measurement:** Income data in our statistical programs relies mostly on the tax data collected by the Canadian Revenue Agency. Some of the income earned by persons living on reserve is tax-exempt and there may be under-reporting due to that circumstance. Income data for non-tax filers is often imputed from other available information. However, if there is larger incidence of non-tax filing on reserve, this could affect the quality of the estimate.

While these considerations demonstrate the limitations to the low-income data for populations on reserve and in the territories, publishing the data is an important starting point. The circumstances faced in those regions continue to be underrepresented in the data, particularly as it applies to the relative distribution of income in Canada. This work illustrates the opportunity that ongoing progress in data integration and collection can provide.

3. Labour market indicators for the populations living in First Nations communities using census and administrative data

The final initiative leverages administrative data to provide annual information on labour market indicators. It uses census data to identify the on-reserve Indigenous population and administrative data to provide information on employment income, among other things.

Accurate and timely labour market and sociodemographic information is necessary for designing and implementing policies for the benefit of a region. In Canada, data pertaining to First Nations communities (reserves) and their populations are often limited. While there exist various timely surveys and other statistical programs producing labour market information within Canada, these usually do not include reserves in their coverage. The Census of Population provides information on all Canadians including First Nations, Metis and Inuit. However, this information is collected only once every 5 years.

This third initiative attempts to address the issue described above through a novel record linkage approach. By combining the Census of Population which includes sociodemographic and geographic variables with the Longitudinal Workers File (LWF) which provides annual income tax and earnings data, annual labour market indicators can be constructed for the population of interest and are comparable over periods of time.

For this project, labour market indicators are calculated for those who live in a First Nations community and identify as single-identity First Nation in the census.

Record linkage and analytical approach

The design of the data is based on the following principle: census data provide detailed information on the place of residence and sociodemographic characteristics of Canadians (such as Indigenous identity) but is conducted only once every five years. Administrative data from the Canadian Revenue Agency (CRA) are available on an annual basis and are a reliable source of employment and income data but provide a limited amount of information on the characteristics of individuals and their place of residence. Linkage between the two data sources is performed to provide detailed labour market information for individuals with specific sociodemographic characteristics on an annual basis.

A "panel approach" is used:

1. The population enumerated in the census acts as the population of analysis. The census data provide accurate geographical information on the place of residence of people (such as residing on or off reserve) and other sociodemographic information such as Indigenous identity. The analytical sample is drawn from the census.
2. Up to five years of LWF annual data can be linked to the census data. This provides the variables necessary to create annual labour market indicators between census years for the analytical sample.

Every year, an updated year of LWF data is added to the linkage as more tax data become available. Furthermore, with every new census, a new analytical sample can be drawn, and a new panel can be initiated. This allows a refresh of the sample as several characteristics of analytical sample members are measured at census years only[8].

Target population and selected indicators

The latest analytical sample is composed of all 2016 Census respondents residing on reserve on the census date (May 10, 2016). For the indicators to be comparable over time, the analytical sample needs to be comparable from year to year. The analytical sample is therefore composed of annual cross-sections of 2016 Census respondents who were 20 to 64 *on the year an indicator was calculated*. For example, the analytical sample in 2016 will include those who were 20 to 64 years old in 2016. Likewise, the analytical sample in 2020 will include those who were 20 to 64 in 2020. Since the sample is drawn from 2016 Census respondents, this means that the analytical sample in 2020 is comprised of census respondents who were 16 to 60 on the census date (May 10, 2016).

The population of interest is the working-age First Nations population residing in First Nations communities. Although the most used definition of the working age includes those between 15 and 64 years old, the lower age bound for this analysis is 20 years old. This restriction is necessary because individuals under 15 are out of scope for several questions in the census, including those on

---

[8] In addition to the LWF which has a two-year lag in the production of data, earnings administrative data files (T4) are added which allow for preliminary results for some of the indicators. These are subject to change because self-employment information is not included in the T4 data. However, less than 2% of the on-reserve population is self-employed exclusively, according to LWF data meaning that the employment rate from T4 files will therefore show a slight negative bias.

employment and education, and an individual who is 20 years old in 2020 was 16 in the 2016 Census. Furthermore, the age range is also restricted to those between 20 and 64 years old because the linkage rate for individuals below 20 years old is very low. This is due to the fact that younger individuals are less likely to be employed and to have a SIN, especially among the on-reserve population.

Four types of variables are used in the production of the tables: (1) fixed sociodemographic variables from the census thought to be relatively stable; (2) annual employment indicator variables from the LWF; (3) annual income indicator variables from the LWF; and (4) longitudinal indicator variables from the LWF (table 1).

Figure 4 shows annual employment rates developed by this program for 70 Ontario[9] reserves. Employment rates range from below 10 percent to over 70 percent and are different for men and women.

Caveats and challenges

The use of the LWF annual data and the census, provided once every five years, leads to some challenges.

1. The employment status information from this project provides no measures of labour force participation, labour force activity, or unemployment.
2. The LWF, like other administrative tax data in Canada, measures annual income. It provides a summary of work history for the year and does not collect information about hours worked. It does not distinguish serial employment from multiple job holders.
3. The LWF does not allow for accurate identification of residence within a First Nations community nor of Indigenous identity, nor of education. Information on these concepts is provided solely every five years with the census. Therefore, stability assumptions need to be made relative to unchanged Indigenous status, place of residence and other sociodemographic information between censuses.

---

[9] Ontario is a province of Canada.

| Table 1:	
Overview of labour market indicators and socio-economic characteristics used in the on reserve labour market indicators initiative	
Indicators or characteristics	**Description**
1. Socio-economic characteristics from the census (typically frozen)	
Age*	Defined as age at the last birthday before Census day
Sex	Male or female
Provinces and territories	Selected Canadian provinces and territories. There are no reserves in Yukon and Nunavut
Metropolitan influence zone	Refers to the census metropolitan area (CMA), census agglomeration (CA), or non-CMA/CA of current residence on Census day, as well as the degree of CMA/CA influence of the place of
Education	Education refers to the highest certificate, diploma or degree obtained as of Census day
Field of study	Field of study refers to the predominant discipline or area of learning of a person's highest completed postsecondary education
Occupation	Occupations held by respondents during the census reference week
Activity limitation	Number of conditions and/or difficulties a respondent reported having which limited them from doing certain activities
2. Annual employment and job characteristics indicators	
Employment status	Presence of employment income
Employment and record linkage status	Presence of T4 record and/or self-employment income in T1, and record linkage status
Industry (NAICS sector), main employer	Distribution by NAICS sector 2-digits (2017) assigned to worker's business enterprise (national level)
Firm size, main employer	Estimated number of employees in employer's business (national level)
Number of jobs	Number of jobs based on the number of unique employee-employer pairs observed in a given year. The presence of self-employment income from the T1 is counted as a single distinct job and added to the number of T4s
Source of employment income	Employment income source (employee, self-employed, mixed)
3. Annual income indicators	
Total employment income	Sum of income from wages and salaries in all jobs held this year, and net self-employment income (mean, 10th percentile, median, 90th percentile)
4. Longitudinal employment indicators	
Tenure	Number of consecutive years a person has received a T4 from the same employer
Number of years employed	Number of years employed over the past five years (including the ongoing year; non-consecutive, may include part-year work)
Employed status flow type	Transitions in annual employed status (between with employment income and without)
Hiring rate	Linked to a T4 associated with a specific business on a given year and no T4 associated with that business in previous year
Separating rate	Linked to a T4 associated with a specific business on a given year and no T4 associated with that business the following year

These challenges introduce the possibility that the quality of estimates decreases for years further away from the base census year. Labour market information, absent from the existing labour force survey, is essential to understanding inequalities in standards of living across Canada. Employment and labour income, along with numerous other factors, are central to reducing low-income and poverty among indigenous communities.

Similar techniques might eventually be used to have an annual estimate of the prevalence of low income, but the cohabitants required for computing household income would have to be identified in the census as they are not in administrative data. A hypothesis of perfectly stable household arrangements over up to five years is a little less credible and could limit the interpretation and coherence of the results should we choose to add up individual incomes every year as if the household has not changed.

Conclusion

The pursuit of complete data and indicator coverage by statistical agencies across the world is ongoing. These three initiatives provide a glimpse into some of the strategies Statistics Canada uses to fill these data gaps. The growing importance of data in decision and policymaking means part of a poverty reduction strategy should involve developing statistics for those who are missing from key data sets.

Figure 4: Percentage of on-reserve First Nations population 20-64 years old with annual employment by census subdivision and gender, 2019

Percentage of on-reserve First Nations population, 20-64 years old with annual employment by census subdivision and gender, 2019
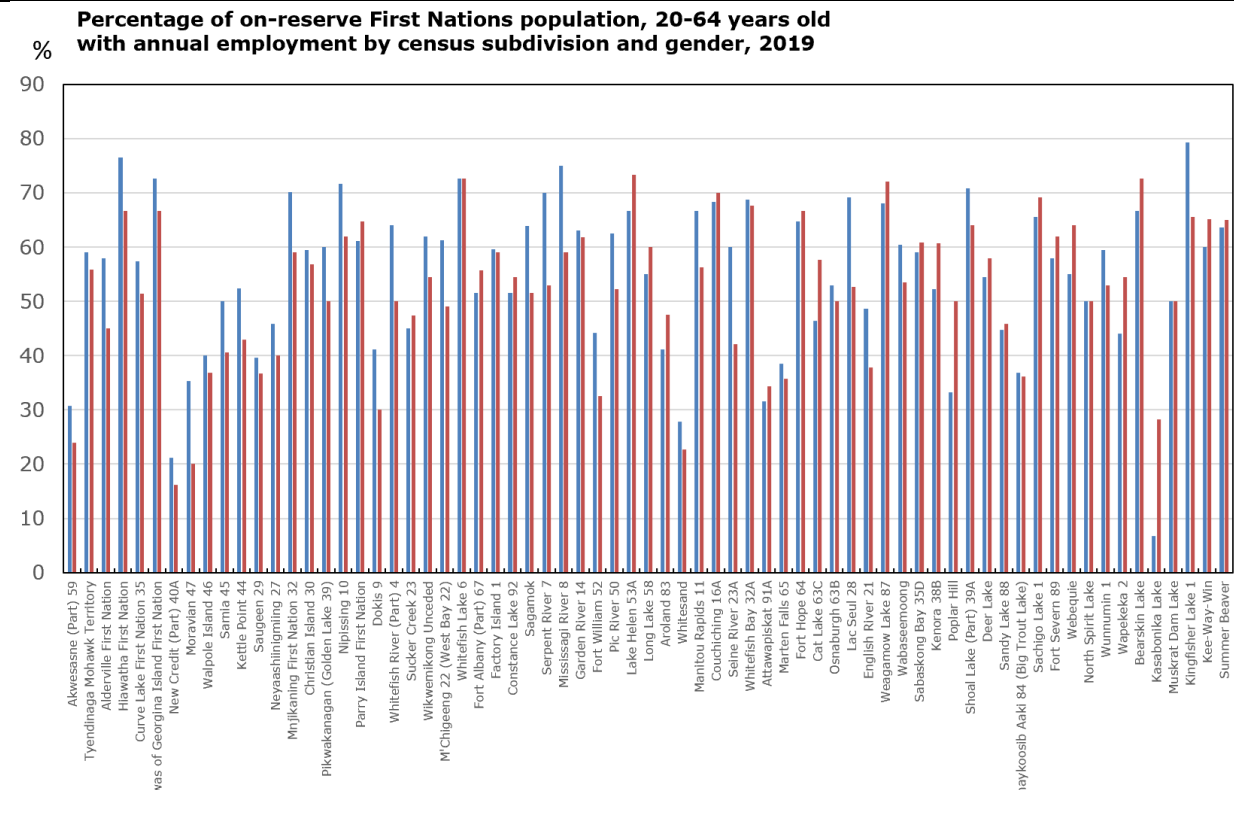
Source: Statistics Canada

References

Case, A., & Deaton, A. (2021). *Deaths of Despair and the Future of Capitalism*. Princeton University Press.

Green, D. A., Gutierrez, P., Milligan, K., & Snowberg, E. (2021). Basic income: Characteristics related to presence in and absence from the tax system. *Research Paper for the BC Basic Income Panel*. Retrieved from: Basic Income: Characteristics Related to Presence in and Absence from the Tax System (bcbasicincomepanel.ca)

Heisz, Andrew. (2019). Interpreting low-income statistics for the on-reserve and territorial populations using Census and National Household Survey data. Statistics Canada. Income Statistic Division. Income Research Paper Series. Retrieved from: Interpreting low-income statistics for the on-reserve and territorial populations using Census and National Household Survey data (statcan.gc.ca)

Messacar, D. (2018). Big Tax Data and Economic Analysis: Effects of Personal Income Tax Reassessments and Delayed Tax Filing. Analytical Studies: Methods and References. Statistics Canada. Catalogue no. 11-633-X — No. 012. Statistics Canada. Ottawa. Retrieved from: Big Tax Data and Economic Analysis: Effects of Personal Income Tax Reassessments and Delayed Tax Filing (statcan.gc.ca)

Sanmartin et al. (2021). Statistics Canada British Columbia Opioid Overdose Analytical File: Technical Report. Statistics Canada. Analytical Studies: Methods and References. Retrieved from: Statistics Canada British Columbia Opioid Overdose Analytical File: Technical Report (statcan.gc.ca)

Statistics Canada (2020). Painting a Portrait of Canada: The 2021 Census of Population (statcan.gc.ca), Statistics Canada. Catalogue number 98-26-0001, issue 2020001. Ottawa. 93 p.

Statistics Canada (2022). Income Reference Guide, Census of Population, 2021. Statistics Canada. 98-500-X2021004. Ottawa. Income Reference Guide, Census of Population, 2021 (statcan.gc.ca)

St-Denis, X. and Harding, A. (2021). Low-income statistics for the population living on reserve and in the North using the 2016 Census. Statistics Canada. Income Research Paper Series. ISBN 978-0-660-38552-5. Retrieved from: Low-income statistics for the population living on reserve and in the North using the 2016 Census (statcan.gc.ca)