# Harnessing the “Power From Data!”

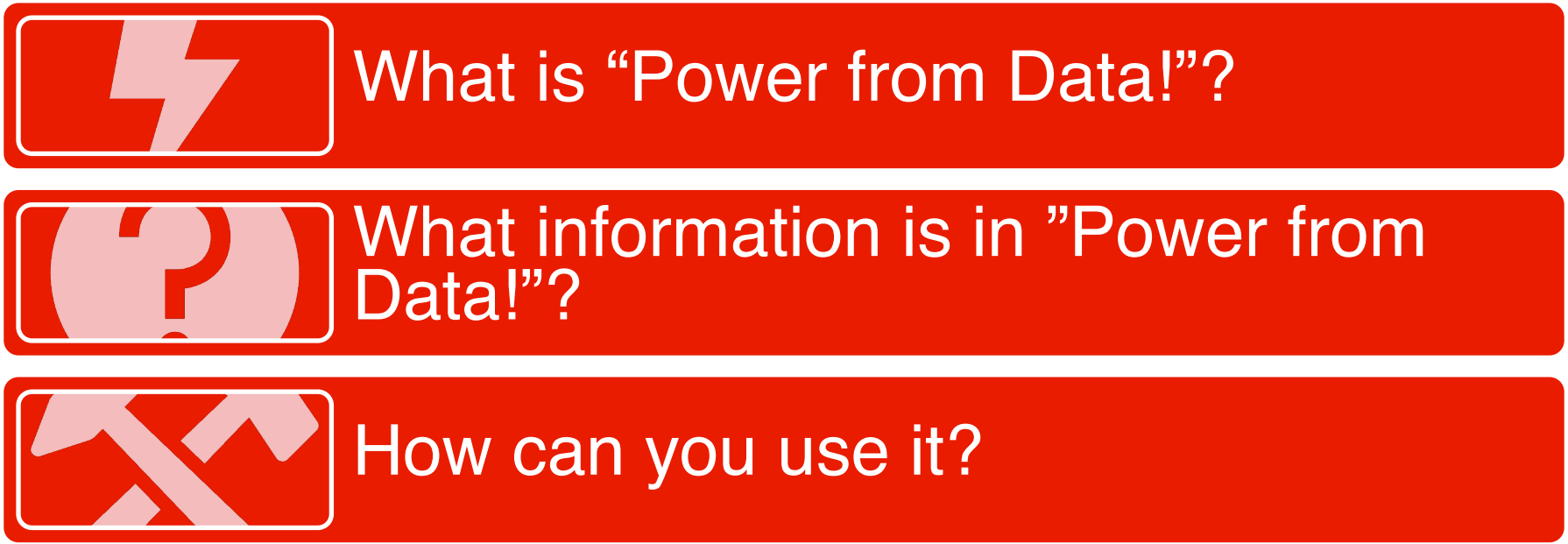
A Lightning Talk for Government Information Days

Tuesday, December 12th, 2023

Alparslan Ozturk

Carleton University
MacOdrum Library

# What are we talking about today?




Icons Source: Microsoft Powerpoint

# What is "Power from Data!”?

- Data and statistical literacy training tool developed by Statistics Canada
- For anyone and everyone
- Building confidence in:
    - Using data and statistical information
    - Understanding data and statistical information
    - Knowing the importance of data and statistical information
- Based on the Australian Bureau of Statistics' 2nd edition of “Statistics - A Powerful Edge!”

# What information is in “Power From Data!”?

- Five sections of information covering the fundamentals of data and statistical literacy
- 1 Data, statistical information and statistics
- 2 Sources of data
- 3 Data gathering and processing
- 4 Data exploration
- 5 Data visualization

# Data, statistical information and statistics

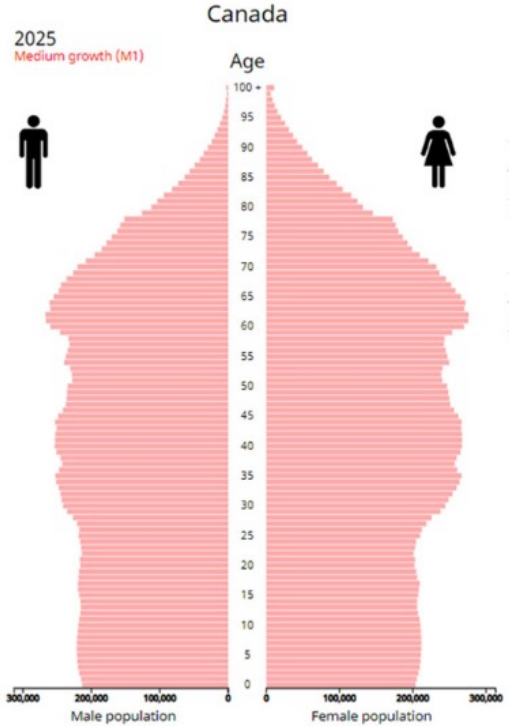
- **Definitions**
- **Terminologies**
- **Data quality**


1 Data, statistical information and statistics

- 1.1 Definitions
- 1.2 Examples of statistical information
- 1.3 Data quality
- 1.4 Exercises
- 1.5 Answers


Source: https://www150.statcan.gc.ca/n1/edu/power-pouvoir/toc-tdm/5214718-eng.htm

**Figure 1.2.1**
**Population pyramid of the Canadian population projected in 2025**




**Number**

| | 0 to 14 | 15 to 64 | 65 + |
|---|---|---|---|
| Male | 3,277,775 | 11,887,465 | 3,804,240 |
| Female | 3,132,445 | 12,146,870 | 4,354,150 |

**Proportion**

| | 0 to 14 | 15 to 64 | 65 + |
|---|---|---|---|
| Male | 17.3 | 62.7 | 20.1 |
| Female | 16.0 | 61.9 | 22.2 |

**Projections (2017 to 2036)**

The various projection scenarios are constructed by combining a number of assumptions regarding the future evolution of each of the components of population growth. They present a series of distinct, but plausible results in terms of the evolution of the size and the age and sex distribution of the Canadian population, in order to reflect the uncertainty associated with the future.

- ○ Low growth
- ◉ Medium growth (M1)
- ○ High growth

## Population pyramids

Figure 1.2.1 is an example of a dynamic age—sex pyramid of Canadian population projections for 2025, based on an assumption of medium growth. It consists of two horizontal histograms, placed side by side, indicating the number of persons in each single year of age. Men are shown on the left histogram and women are shown on the right, as stated by convention.

View the product Historical Age Pyramid for other years and other projection scenarios.

Age–sex pyramids are commonly used to present statistical information on the age and sex composition of a population. This chart clearly shows the aging "Baby Boomers."

Source: https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch1/examples-exemples/5214854-eng.htm

# Sources of data

- **Data sources**
- **Data providers**
- **Data users**
- **Type of data**


2 Sources of data

- 2.1 Data providers and data users
- 2.2 Type of data
- 2.3 Exercises
- 2.4 Answers


Source: https://www150.statcan.gc.ca/n1/edu/power-pouvoir/toc-tdm/5214718-eng.htm

Groups and organizations that use data, statistical information and statistics include the following:

- **Governments:** Federal, provincial and local governments need information on the population and the economy, among other things, to help them develop, implement and monitor socio-economic and environmental programs and other functions of government, such as licensing and regulation. It helps governments make decisions on issues such as where to build hospitals, locate services, or how much money to raise through taxation. It also allows the public, opposition politicians and interest groups to measure a government's performance in decision making and to hold it accountable if it does not meet the criteria.
- **Businesses:** Canadian businesses require information about the local, provincial and national economy, the unmet needs of a population, and trends in society. Data helps them make decisions about employing people, marketing their products and opening new offices, warehouses and factories. Data are also required for businesses to carry out their operations, such as billing, inventory and supply management.
- **Community groups:** These organizations need information about a wide variety of subjects, such as health and population distribution of Indigenous people, or the number and location of Canadian immigrants who require English or French language skills. Sporting clubs may want information about attendance at games or the number of young people in their local area.
- **Academics and researchers:** Data is required by those carrying out studies and other analyses, in a variety of roles. Data may be used in planning research work (e.g. in what community to conduct a study) or to back up research claims and other hypotheses (e.g. do historic data support a correlation between higher average temperatures and increased flooding).
- **Individuals:** Everyone, from students to pensioners, needs some form of information at some time during their lives. The information may be used to complete an essay, a major project or simply to satisfy one's curiosity.

Source: https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch2/5214779-eng.htm

# Data gathering and processing

- Planning
- Sampling
- Collecting
- Processing
- Estimation


3 Data gathering and processing
- 3.1 Planning
- 3.2 Sampling
  - 3.2.1 Selection of a sample
  - 3.2.2 Probability sampling
  - 3.2.3 Non-probability sampling
- 3.3 Collecting
  - 3.3.1 Data collection methods
  - 3.3.2 Questionnaire design
  - 3.3.3 Role of interviewers
- 3.4 Processing
  - 3.4.1 Coding
  - 3.4.2 Capture
  - 3.4.3 Editing
  - 3.4.4 Imputation
  - 3.4.5 Record linkage
- 3.5 Estimation
  - 3.5.1 Weighting
  - 3.5.2 Sampling error
  - 3.5.3 Non-sampling error
- 3.6 Quality management
- 3.7 Exercises
- 3.8 Answers


Source: https://www150.statcan.gc.ca/n1/edu/power-pouvoir/toc-tdm/5214718-eng.htm

# Design weight

The first step in estimation is assigning a weight to each sampled unit. The **design weight** ( $w_d$ ), which is the average number of units in the population that each sampled unit represents, is the inverse of its inclusion probability ($\pi$) in the sample.

$$w_d = 1/\pi$$

If the inclusion probability is 1/50, then each selected unit represents on average 50 units in the population and the design weight is $w_d = 50$ .

Some sample designs assign the same design weights for all units in the sample, while others give different design weights to sampled units for various reasons, such as improving precision or reducing cost.

## Example 1: Simple Random Sample

Suppose there are *N* =100 Grade 12 (or secondary 5) students in a high school. A simple random sample of size *n* =25 students is selected, and the selected students are invited to complete a questionnaire about their career plan.

- The inclusion probability is:
  $\pi = n/N = 25/100 = 1/4$ .
- The design weight is:
  $w_d = 1/\pi = 1/\frac{1}{4} = 4$ .

Each student selected in the simple represents four students of the school.

Source: https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch6/5214809-eng.htm

# Exploring data

- **Data exploration tools**
- **Variable types**
- **Calculations and computations**


4 Data exploration
- 4.1 Data exploration tools
- 4.2 Types of variables
- 4.3 Frequency distribution
- 4.4 Measures of central tendency
  - 4.4.1 Calculating the mean
  - 4.4.2 Calculating the median
  - 4.4.3 Calculating the mode
- 4.5 Measures of dispersion
  - 4.5.1 Calculating the range and interquartile range
  - 4.5.2 Visualizing the box and whisker plot
  - 4.5.3 Calculating the variance and standard deviation
- 4.6 Exercises
- 4.7 Answers


Source: https://www150.statcan.gc.ca/n1/edu/power-pouvoir/toc-tdm/5214718-eng.htm

Table of contents

# 4.5.3 Calculating the variance and standard deviation

Unlike range and interquartile range, variance is a measure of dispersion that takes into account the spread of all data points in a data set. It's the measure of dispersion the most often used, along with the standard deviation, which is simply the square root of the variance. The variance is mean squared difference between each data point and the centre of the distribution measured by the mean.

## Example 1 – Calculation of variance and standard deviation

Let's calculate the variance of the follow data set: 2, 7, 3, 12, 9.

The first step is to calculate the mean. The sum is 33 and there are 5 data points. Therefore, the mean is 33 ÷ 5 = 6.6. Then you take each value in data set, subtract the mean and square the difference. For instance, for the first value:

$(2 - 6.6)^2 = 21.16$

The squared differences for all values are added:

21.16 + 0.16 + 12.96 + 29.16 + 5.76 = 69.20

The sum is then divided by the number of data points:

69.20 ÷5 = 13.84

The variance is 13.84. To get the standard deviation, you calculate the square root of the variance, which is 3.72.


**Topic navigation**

- 4 Data exploration
  - 4.5 Measures of dispersion
    - 4.5.1 Calculating the range and interquartile range
    - 4.5.2 Visualizing the box and whisker plot
    - 4.5.3 Calculating the variance and standard deviation


Source: https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch12/5214891-eng.htm

# Data visualization

- **Using various visualizations**
- **Differences of use**
- **Examples of use cases**

5 Data visualization

- 5.1 Using graphs
- 5.2 Bar chart
- 5.3 Pictograph
- 5.4 Pie chart
- 5.5 Line chart
- 5.6 Scatterplot
- 5.7 Histogram
- 5.8 Exercises
- 5.9 Answers

Source: https://www150.statcan.gc.ca/n1/edu/power-pouvoir/toc-tdm/5214718-eng.htm

**Chart 5.3.2**
**Purchasing power of the Canadian dollar, 2000 to 2020**

2000 = $1.00

2005 = $0.89

2010 = $0.81

2015 = $0.75

2020 = $0.70

Source: https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch9/picto-figuratifs/5214825-eng.htm

# What information is in “Power From Data!”? Cont..

- Importance and utility of each section
- Exercises and answers per section covering all subjects
- Encourages use of Statistics Canada products
- Fundamentals of data and statistical literacy

5 Data Visualization

## 5.8 Exercises

1. The number of basketball games attended by 50 season ticket holders were: 15, 10, 17, 11, 15, 12, 13, 16, 12, 14, 14, 16, 15, 18, 11, 16, 13, 17, 12, 16, 18, 15, 17, 15, 19, 13, 14, 17, 16, 15, 12, 11, 17, 16, 15, 10, 14, 15, 13, 16, 18, 15, 17, 11, 14, 17, 15, 14, 13, 16.
   a. Tally the data and present them in a cumulative frequency table. Answer 1a
   b. Draw a vertical bar chart. Answer 1b
   c. Describe the data set using the five-number summary, the range and the interquartile range. These concepts have been presented in section 4 on data exploration. Answer 1c

Source: https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch10/exer/5214860-eng.htm

# How Can You Harness the “Power From Data!”?

Levels of experience in data and statistical literacy

Effective starting point

Fundamentals for your toolkits

Importance of data and statistical literacy as a skill

# Statistics Canada - Finding Data

*The following are carefully curated resources to help you with find what you need on the StatCan website. It contains and describes helpful data resources and our Data Table Find Guides. Check back often as this page will be updated regularly! And if there's anything you think we should add, let us know by emailing DataServices@Carleton.ca*

2021 Census Research Kit | Canada at a Glance | Canada Year Books | DataBytes | Data Hubs | Data Literacy Training | Data Portals | Data Table Find Guides | Data Visualizations: Interactive | DLI Survival Guide | Fact Sheets | Infographics | InfoGuide | Maps | Power From Data! | StatCan Webpage Presentation | StatsCAN Plus

## 2021 Census Research Kit

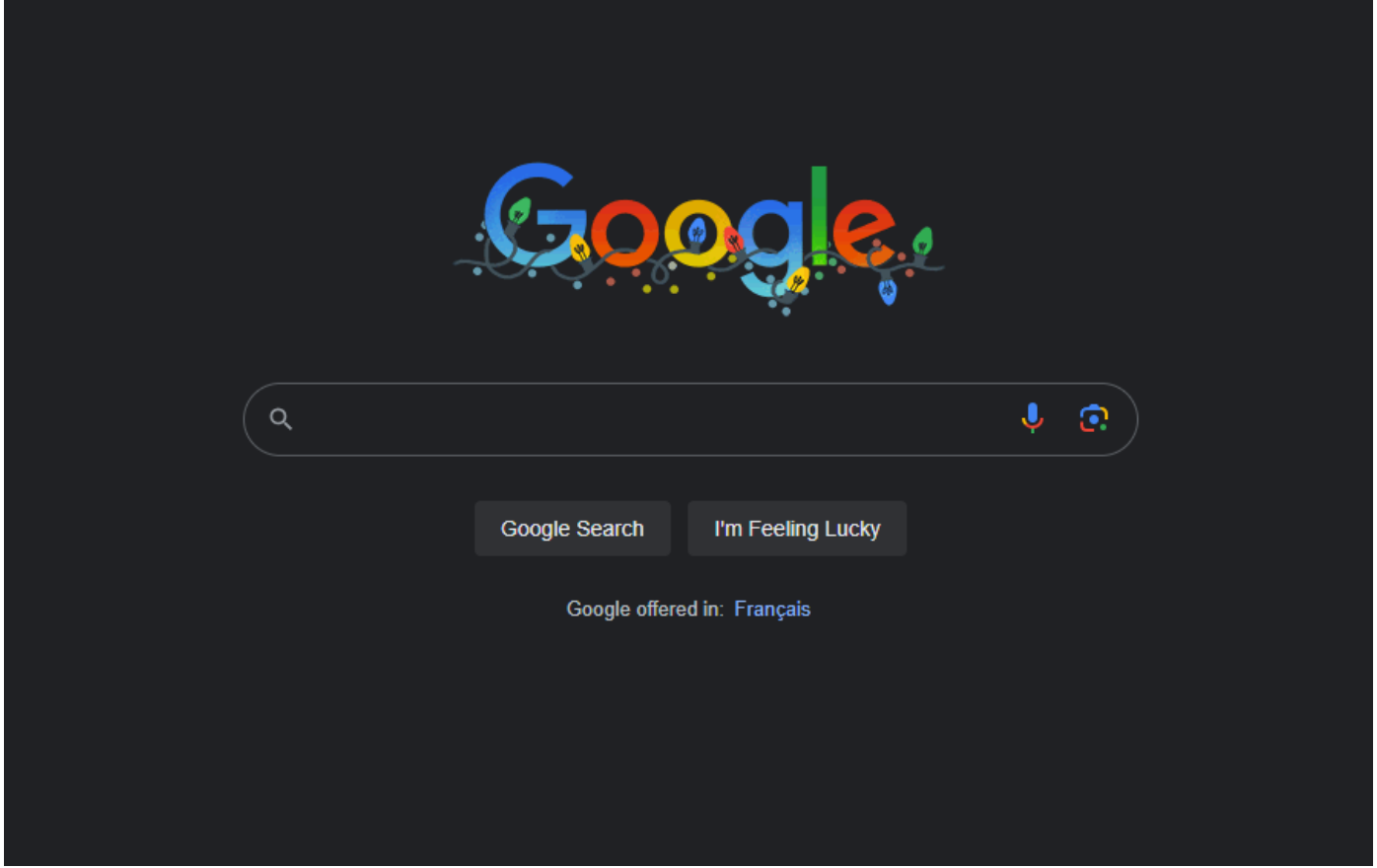
- This toolkit will help familiarize you with the Census and key sources of data and other information. It will show you: topics covered by the 2021 Census, examples of research involving Census data, and Census 2021 reference tools and sources for your next research project.

**Contact**

Data Services
dataservices@carleton.ca



Source: https://library.carleton.ca/guides/help/statistics-canada-finding-data




Source: https://www.google.com/

# Thank you for your time!

alparslan.Ozturk@carleton.ca

Carleton University
MacOdrum Library

# Links to resources

- Power From Data!
- “Statistics - A Powerful Edge!”
- Statistics Canada – Finding Data
- Google