# Exploring Statistics Canada Data Resources

Oct. 29, 2026

1:30 - 3:30 p.m.

Online

Sherman Centre for Digital Scholarship



**Stream**

**Data Analysis and Visualization**



Photo Credit:
Campus Fall 2025 by Georgia Kirkos

## Land Acknowledgement

McMaster University is located on the traditional Territories of the Mississauga and Haudenosaunee Nations, and within the lands protected by the “Dish With One Spoon” wampum agreement.

Sherman Centre for Digital Scholarship

A joint initiative of McMaster's Faculty of Humanities and McMaster University Libraries

McMaster University

# Code of Conduct

The Sherman Centre and the McMaster University Libraries are committed to fostering a supportive and inclusive environment for its presenters and participants.

As a participant in this session, you agree to support and help cultivate an experience that is collaborative, respectful, and inclusive, as well as free of harassment, discrimination, and oppression. We reserve the right to remove participants who exhibit harassing, malicious, or persistently disruptive behaviour.

Please refer to our code of conduct webpage for more information:
**scds.ca/code-of-conduct/**

# Session Recording and Privacy

This session is being recorded with the intention of being shared publicly via the web for future audiences. In respect of your privacy, participant lists will not be shared outside of this session, nor will question or chat transcripts.

Questions asked via the chat box will be read by the facilitator without identifying you. Note that you may be identifiable when asking a question during the session in an audio or visual format.

# SCDS Certificate of Attendance

The Sherman Centre's certificate program recognizes attendance at our workshops.

The certificate complements degree training, supports the development of critical DS competencies, and formalizes core skills fostered by the workshop series.

**General Certificate:** Any 7 SCDS workshops

**Stream-focused Certificate** (Data Analysis and Visualization, Digital Scholarship, or Research Data Management):
Any 3 SCDS workshops and 4 stream-specific workshops.

See all information (including frequently asked questions and how to authenticate your attendance) at **scds.ca/certificate-program**

## General Certificate



Any 7 SCDS workshops

## Stream-focused Certificate



Any 3 SCDS workshops
+ 4 stream-specific workshops

# Fall 2026: Upcoming Workshops

## Data Analysis Support Hub

**October 14 -** Multivariable Analysis with R
**October 15 -** Introduction to Python Programming
**October 29 -** Exploring Statistics Canada Data Resources
**November 4 -** Introduction to R Programming

## Research Impact

**October 19 -** DORA: Tracking the Impact of Non-Traditional Research Outputs
**November 10 -** DORA: Visualizing Research Impact
**February 10 -** Introduction to Bibliometric Analysis

## Research Data Management

**October 13 -** Data Management Plan Bootcamp (In-Person)
**October 21 -** Best Practices for Managing Data in your Research (Fall)
**November 17 -** Setting Up a Research Project with Data from Humans and Preparing for Ethics

## Do More with Digital Scholarship

**October 20 -** Creating Small and Ugly Open Educational Resources (OER)
**October 23 -** Social Media Scraping and Analysis with Zeeschuimer and 4Cat
**December 10 -** Introduction to SQL

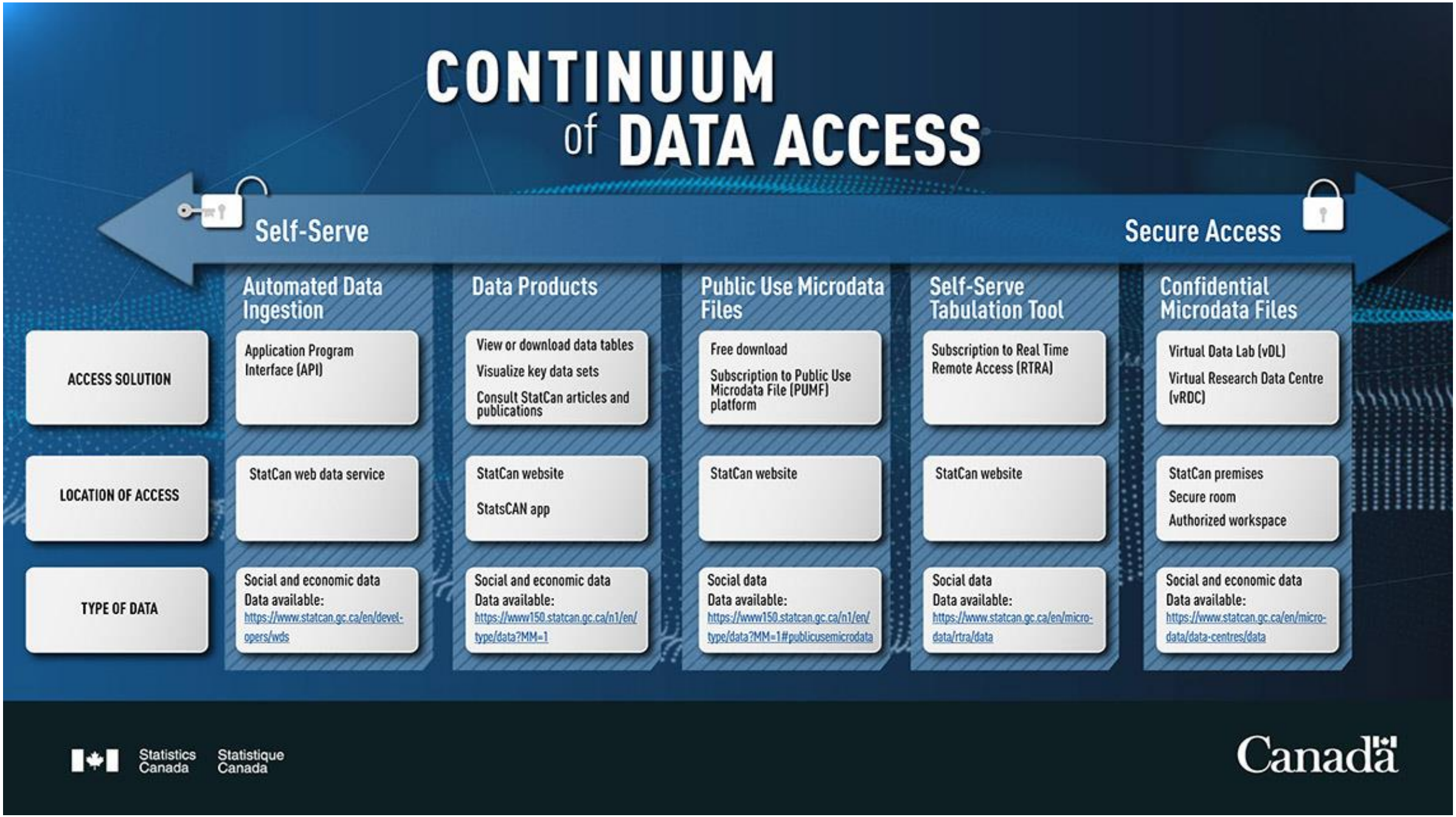
Register for Upcoming Workshops: [scds.ca/events](scds.ca/events)

# Learning Objectives

- By the end of the workshop, participants will be able to:
  - Identify the major types of Statistics Canada data products.
  - Distinguish between **data tables, surveys, PUMFs, Census data, and microdata**.
  - Search Statistics Canada effectively using **concepts and variables**, rather than just dataset titles.
  - Find and interpret a Statistics Canada data table.
  - Customize and download data from **Statistics Canada Data tables/Data Explorer**.
  - Find survey documentation and understand what a variable actually measures.
  - Locate and download a **Public Use Microdata File (PUMF)**.
  - Understand the difference between **aggregate data and microdata**.
  - Identify when data require access through a **Research Data Centre (RDC)**.
  - Recognize common problems involving geography, time periods, units, missing values, and data definitions.

# What is Statistics Canada?

- Statistics Canada is more than Census data
  - How many of you have used Statistics Canada before?
  - How many of you have downloaded a Statistics Canada CSV file?
  - How many of you have opened a Statistics Canada dataset and had absolutely no idea what half the variables meant?




**Credit:**
**https://www.statcan.gc.ca/en/microdata?HPA=1**

# The Statistics Canada Data Ecosystem

- Four ways to work with Statistics Canada data
  - **Published tables**
    - What is the number?
  - **Census data**
    - Who lives where?
  - **Public Use Microdata Files (PUMFs)**
    - What characteristics are associated with individual respondents?
  - **Research Data Centres**
    - The PUMF doesn't contain enough detail.

# The Most Important Concept: Table vs. Microdata

## Published table

| Province | Employment rate |
|---|---|
| Ontario | 61.2% |
| Quebec | 62.1% |
| Alberta | 67.4% |

## Microdata

| Person | Age | Sex | Province | Employment | Education |
|---|---|---|---|---|---|
| 001 | 24 | F | ON | Yes | University |
| 002 | 37 | M | AB | Yes | College |
| 003 | 52 | F | ON | No | High school |

You can analyze the published numbers.

But you don't know anything about the individual respondents behind those numbers.

Now you can investigate relationships between variables.

**- Tables answer questions about populations.**
**- Microdata allow you to investigate relationships between characteristics.**

# Data Liberation Initiative (DLI)

- DLI provides access to Stat Can's Standard products, Databases, Public use Microdata files and Geographic information files.
  - Main focus is on Socio-Economic data: Health, Education/Literacy, Labor Market, Income, Travel, Justice, Census of Population etc.
  - Databases such as the Small Area Business and Labour Database, Inter-Corporate Ownership, Financial Performance Indicators, Trade data etc.
  - An enhanced line of Census products
  - Aggregated data on subject such as Justice and Education
  - $5,000 worth of custom tabulations from Stat Can's Data Service Centres
  - All standard geographic files and databases including PCCF & PCCF+

# Data Liberation Initiative (DLI)

- Metadata from DLI surveys are marked-up in DDI/XML format for discoverability at the variable level from **Odesi Data Portal**.
- DLI members have support through a very active listserv.
- Currently over 77 subscribing institutions -- McMaster University Library is part of Stat Can DLI program.

# Hands-On Example -- Census

- **Research Question**
  - How has the population of Hamilton changed, and what does the age structure look like?

    - **Step 1 -- Find Census data**
      - Population by age and gender

    - **Step 2 -- Change the geography**
      - Compare – Hamilton, Toronto, Mississauga, Ontario, Canada

    - **Step 3 — Change the year**
      - Investigate – 2016 → 2021

# Mini Challenge

- Find the population aged 65+ for your home municipality in the 2021 Census.

# Hands-On Example -- Labour Force Survey

- **Research Question**
  - What has happened to Canada's unemployment rate recently?

# Hands-On Example -- Labour Force Survey

- Open:
  - **Table 14-10-0287-01 –** “Labour force characteristics, monthly, seasonally adjusted and trend-cycle.”
  - **Find** – Employment, Unemployment Rate, Participation Rate
  - **Change**: April → May → June → July → August
  - The table provides monthly observations and can be downloaded in CSV and other formats.

# Exercise

- **Question**
  - **Did Canada's unemployment rate increase or decrease over the selected period?**
  - Compare Canada with Ontario.
  - Compare youth unemployment with the overall unemployment rate.
  - What might explain the difference?

# Don't Trust the Number Yet

- **What is 6.5%?**
  - **Possible answers:** unemployment rate, annual rate, monthly rate, Ontario rate, Canadian rate, youth rate, seasonally adjusted rate
  - The correct answer is:
    - **We don't know until we understand the metadata.**
  - Explore –
    - **Who, What, Where, When, How it’s measured, Unit, Denominator, Seasonally adjusted**

# Public Use Microdata Files (PUMF)

- Each Public Use Microdata File is based on a corresponding master data file. The modifications performed by Statistics Canada before the PUMF is released ensure that the risk of breaching confidentiality has been removed. Since the results of any analysis performed do not have to be scrutinized before they are released, the file is considered "Public".

- Modifications made to the Master files for conversion to PUMFs may include **collapsing of variables** (e.g., age groups instead of individual years of age); **collapsing variables into one variable** (e.g., multiple language questions collapsed into one language variable for analysis); **suppressing variables** (although the variable is part of the master file, it will not show up in the public file); and **removing outliers** (removing cases that are extremes - often used with income).

- By using these techniques to anonymise the files, combining variables will not result in the user identifying a respondent from any given survey.

# Public Use Microdata Files (PUMF)

- **PUMF ≠ the entire survey**
  - **Survey → Master file → PUMF**
- The PUMF is a carefully modified version designed to protect confidentiality.
- Statistics Canada notes that PUMFs contain non-aggregated records but are modified and reviewed to reduce the risk of identifying individuals or businesses.

# Public Use Microdata Files (PUMF)

- Therefore:
    - **If you don't see a variable in the PUMF, don't automatically assume Statistics Canada didn't collect it.**
- It may exist in the master file but not be available in the PUMF.
- That's when researchers may need an **RDC** or another access route (such as RTRA).

# Public Use Microdata Files (PUMF)

- **Benefits**
  - Free
  - Very few restrictions on access & use of the data
  - No approval process to access the data

- **Limitations**
  - Content is limited (screened and grouped for confidentiality)
  - Not all surveys have a PUMF
  - PUMFs are cross-sectional, i.e., represent data collected at one point in time

# Real Time Remote Access (RTRA)

- On-line remote access facility allowing users to run SAS programs, in almost real-time, against microdata located in a central and secure location.
- Researchers using the RTRA system do not gain direct access to the microdata and cannot view the content of the microdata. Instead, users submit SAS programs to extract results in the form of frequency tables.
- As RTRA researchers cannot view the microdata, becoming a deemed employees of Stat Can is no longer necessary. Hence, rapid access to “confidential” microdata files.
- Using a secure username and password, the RTRA provides around the clock access to survey results through EFT from any computer with internet access.
- Confidentiality of the micro data is automated in the RTRA system, eliminating the need for manual intervention and allowing for rapid access to results.

# Real Time Remote Access (RTRA)

- **Benefits:**
  - Access from any computer with internet access using a secure username and password - No travel to RDCs
  - Few conditions on access and use of data
  - Full Master Files available; Confidentiality automated
  - Deemed employee status not required
- **Limitations:**
  - Descriptive statistics only **(Frequency, Mean, Median, Percentiles, Percentage distribution, Ratios, Proportions)**
  - Can only be used with SAS
  - Not all data sets available under RTRA service – currently 105 data sets (no access to Census Microdata)
  - Costs – 2 accounts for $5,000; unlimited access for $10,000
- https://www.statcan.gc.ca/eng/rtra/rtra

# Odesi Data Portal

- Web-based data extraction system delivered through  Scholars Portal, provides access to diverse, quality, numeric data sets including microdata survey collection from DLI, demographic data from Statistics Canada and polling data from Gallup and other organizations.

- Facilitates the exploration (searching, data manipulation, creation of summary statistics, graphing and export) of multiple, sophisticated data sets.

- Access is open to all OCUL institutions and is controlled by IP address.

- https://odesi.ca

# Hands-On Example -- Canadian Community Health Survey (CCHS)

- **Research Question**
- Is self-reported mental health associated with age group among Canadians aged 18 and older?

# Hands-On Example -- Canadian Community Health Survey (CCHS)

- **Step 1** -- Find the CCHS 2022 PUMF
- Download or work with a copy on Odesi using Data Explorer.
- **Never download a microdata file and immediately start analyzing it. Read the documentation first.**

# Hands-On Example -- Canadian Community Health Survey (CCHS)

- **Step 2** -- Find the Variable
  - Find the variable related to **Self-perceived mental health** (excellent/very good, good, fair/poor).
  - Find a variable related to **age group**.

# Hands-On Example -- Canadian Community Health Survey (CCHS)

- Look at the response categories
  - For example, the distinction between:
    - Excellent
    - Very good
    - Good
    - Fair
    - Poor
  - **matters enormously for analysis.**

# Hands-On Example -- Canadian Community Health Survey (CCHS)

- What percentage of respondents report:
  - Excellent/Very good mental health?

- How does this differ by:
  - age
  - gender
  - income group

- Does the pattern appear different among physically active respondents?

# Search by Variable, Not Dataset

- Imagine someone asks:
  - **“I need data about women's health.”**

- Don't search:
  - **women's health dataset**

- Instead break the research question into variables:
  - **Population**
  - Women aged 18–49
  - **Outcome**
  - Health status
  - **Exposure**
  - Employment

# Search by Variable, Not Dataset

- **Other variables**
  - income
  - education
  - access to health care
  - reproductive health
  - geographic location

- Now search for surveys containing those variables.

# The Workflow

**Research question**

↓

**Concepts**

↓

**Variables**

↓

**Dataset**

↓

**Documentation**

↓

**Data**

↓

**Analysis**

# Exercises – Try on Your Own

- **Housing -- How has the cost of housing changed in Canada?**
  - Find: relevant table, geography, time period, unit, download option
- **Employment -- What is the unemployment rate for young Canadians?**
  - Find: survey, table, age group, period, measure
- **Health -- What Canadian data could be used to study physical activity and health?**
  - Find: survey, PUMF or table, relevant variables, documentation
- **Immigration -- What Census data could be used to examine immigration and education?**
  - Find: table, variables, geography, Census year

# The "Can I Use This Dataset?" Checklist

- Before downloading data, ask...
  - **1. Does it contain my population?**
    - Who was surveyed?
  - **2. Does it contain my variables?**
    - Don't assume based on the dataset title.
  - **3. Is the geography appropriate?**
    - Canada? Province? CMA? Municipality? Census tract?
  - **4. Is the time period appropriate?**
    - 2021? 2022? 2025? Monthly?
  - **5. Is the data aggregate or microdata?**

# The "Can I Use This Dataset?" Checklist

- Before downloading data, ask...
  - **6. What are the units?**
    - Count? Percentage? Rate? Dollars?
  - **7. What is the denominator?**
    - This is particularly important for percentages and rates.
  - **8. Are there weights?**
    - For survey microdata, this can be critical.
  - **9. Are there missing values?**
  - **10. Can I legally/accessibly use the data?**
    - Open data, PUMF, DLI-restricted, RDC, etc.

# The Statistics Canada Research Workflow

**RESEARCH QUESTION**

↓

**DEFINE CONCEPTS**

↓

**IDENTIFY VARIABLES**

↓

**FIND THE DATASET**

↓

**READ THE DOCUMENTATION**

↓

**CHECK DATA AVAILABILITY**

↓

**DOWNLOAD / ACCESS**

↓

**CLEAN THE DATA**

↓

**ANALYZE & VISUALIZE**

↓

**INTERPRET THE RESULTS**

↓

**DOCUMENT & REPRODUCE**

# How to Cite Statistics Canada Products

- This electronic product shows how to cite all standard products offered by Statistics Canada. In addition to books, articles, analytical studies and periodicals, this work covers non-traditional products such as data files, tables, maps, databases and other Statistics Canada data products. It specifies which basic elements of information - author name, title, year of publication, periodicity, etc. - must be included in a bibliographic reference. Specific examples illustrate the citation rules.

- Citing Statistics Canada Products

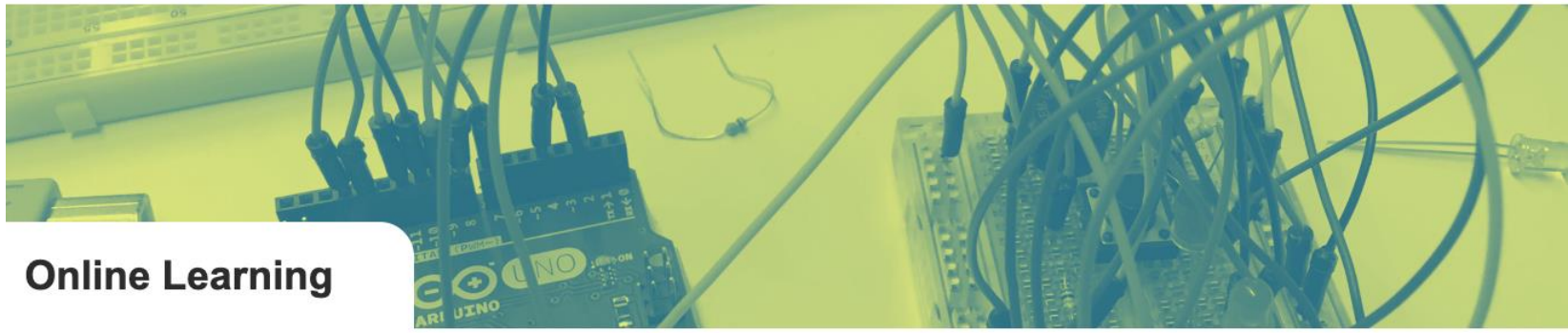
# Recommended Resources

- Statistics Canada -- Data
- Statistics Canada -- Data Tables
- Statistics Canada -- Census Data
- Statistics Canada -- Data Liberation Initiative
- Statistics Canada -- Browse Census Tables by Variable

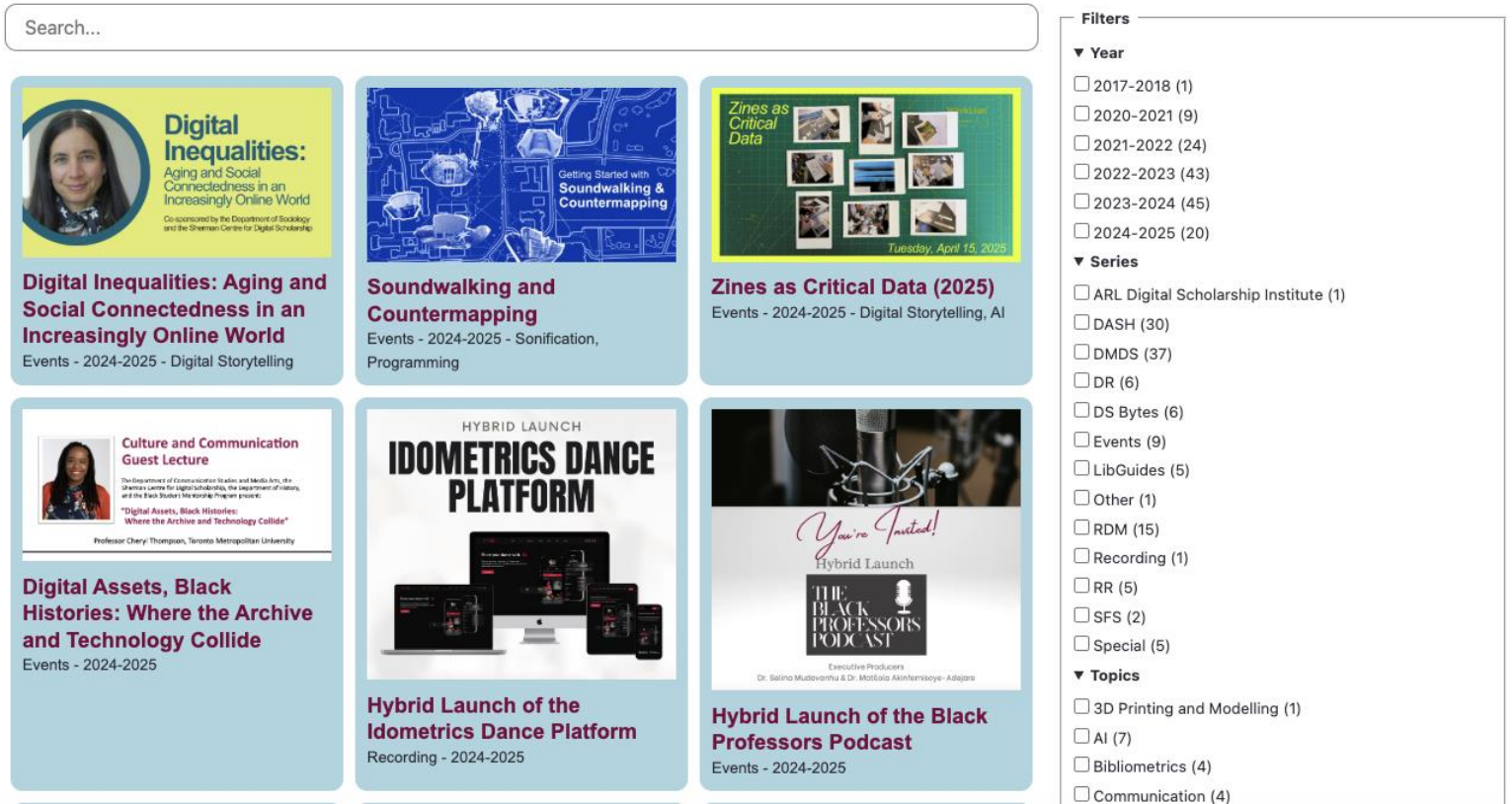
# Digital Learning Catalogue

Learn new digital skills at your own pace with 120+ workshop recordings and online modules. Explore topics like AI, machine learning, 3D printing and modelling, GIS, digital storytelling, sonification, and data visualization to expand your research possibilities.

**scds.ca/online-learning**




New skills are just a click away. These workshop recordings and online modules will teach you about data analysis, digital scholarship, innovative software, and more. You can also view a text list of these resources, sorted by original workshop or event series.



# Tell us what you think | Verify your attendance

**Provide feedback on this session and verify your participation at**
https://u.mcmaster.ca/verification

**This session's verification code:**

oh-canada

scds.ca
scds@mcmaster.ca

A joint initiative of McMaster's Faculty of Humanities and McMaster University Libraries

McMaster University

Sherman Centre for Digital Scholarship

Sherman Centre for Digital Scholarship

# Get In Touch

**Send SCDS an Email:**
scds@mcmaster.ca

**Register for a Workshop:**
https://u.mcmaster.ca/scds-workshops

**Subscribe to our Newsletter:**
https://u.mcmaster.ca/sign-up

**Schedule a Consultation:**
https://scds.ca/consultations/



A joint initiative of McMaster's Faculty of Humanities and McMaster University Libraries

McMaster University