Finding & evaluating statistics
INI235: Comparing Cities Assignment

Nicholas Worby
Government Information & Statistics Librarian
nicholas.worby@utoronto.ca
What you need:

• 2 cities (2 Census Metropolitan Areas)
• A theme
• 6 data points for comparison

Class research guide: [http://guides.library.utoronto.ca/INI235](http://guides.library.utoronto.ca/INI235)
Which one is the CMA?
What is/are data?

• ...?

• ...?
Data

Datum (pl. data):

“factual information...used as a basis for reasoning, discussion or calculation”
(Merriam-Webster)
Dataset

A collection of facts (observations) that can be manipulated as a unit using a computer
Microdata

A dataset containing the individual facts (observations) that were recorded.

For example:
• Individual responses to a survey
• Individual behavioral observations
• Hourly temperature readings at various weather stations
• Firm-level research & development spending
• Agriculture production by farm
Data terminology

Variable
• A characteristic being observed, that can take a range of values
• Observations may be numeric (e.g. temperature reading, Canadian dollars) or they may be descriptive (e.g. marital status, occupation)

Case
• The unit of the analysis (e.g. if 20,000 people respond to a survey, then a microdata file will have 20,000 “cases” or rows)
• Often noted as n=20,000, where n is the number of cases
**Microdata**

Column: One variable or Question

Row: The unit of observation

Cell: One observation of a variable
**Statistical table (aggregate data)**

Job satisfaction by age group and sex, Canada, provinces and regions occasional (percent unless otherwise noted)

<table>
<thead>
<tr>
<th>Data table</th>
<th>Add/Remove data</th>
<th>Manipulate</th>
<th>Download</th>
<th>Related Information</th>
<th>Help</th>
</tr>
</thead>
</table>

The data below is a part of CANSIM table 121-0002. Use the Add/Remove data tab to customize your table.

**Selected items [Add/Remove data]**

Geography $^6$ = Canada  
Age group $^8$ = Total, 15 years and over  
Sex = Both sexes

<table>
<thead>
<tr>
<th>Response $^2$</th>
<th>Estimates $^9$</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total, job satisfaction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of persons (persons x 1,000)</td>
<td>15,167</td>
<td></td>
</tr>
<tr>
<td>Percentage of persons</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Low 95% confidence interval, percent $^3$</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>High 95% confidence interval, percent $^3$</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Very satisfied or satisfied</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of persons (persons x 1,000)</td>
<td>12,730</td>
<td></td>
</tr>
<tr>
<td>Percentage of persons</td>
<td>83.9</td>
<td></td>
</tr>
<tr>
<td>Low 95% confidence interval, percent $^3$</td>
<td>82.5</td>
<td></td>
</tr>
<tr>
<td>High 95% confidence interval, percent $^3$</td>
<td>85.0</td>
<td></td>
</tr>
</tbody>
</table>
Statistics

Collections of calculated figures produced through statistical inquiry – counts, tallies, totals, means/averages, percentages

For example:
• Percentage of teenagers who smoke cigarettes
• Median household income by city
• Approval rates for a political candidate
Which do I need?

Key factor: “How much/many?” vs “Why?”

For example:

• “I need data on the number of Chinese immigrants in Toronto” ← Statistics
• “I’m exploring the relationship between income level and incidence of diabetes” ← Microdata
Statistics Canada

• Primary federal department responsible for collecting data and publishing statistics about Canada and Canadians
• Majority of data collected through surveys of Canadians and Canadian businesses
• Major projects
  – Census of Population
  – Census of Agriculture
  – Many surveys on a range of topics
  – Partnering with other government departments (e.g. Canada Revenue Agency) on collecting/storing administrative data/statistics
Methods of data collection

• Enumeration (census)
  – An official count or survey every person/object

• Survey
  – A study which asks questions of a sample of the population

• Administrative
  – Counting things or people as they use government services or pass particular locations (e.g. hospital visits, imports/exports, tax returns filed)
  – Universities are secondary users – data is used for a different purpose than it was collected for (e.g. collection of taxes yields income distribution statistics)
Census of Population

• What topics do census statistics cover?

<table>
<thead>
<tr>
<th>Age</th>
<th>Labour market activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Occupation</td>
</tr>
<tr>
<td>Marital status</td>
<td>Education</td>
</tr>
<tr>
<td>Family composition</td>
<td>Language of work</td>
</tr>
<tr>
<td>Dwellings and households</td>
<td>Place of work</td>
</tr>
<tr>
<td>Language</td>
<td>Mode of transportation to work</td>
</tr>
<tr>
<td>Immigration</td>
<td>Ethnic origin</td>
</tr>
<tr>
<td>Citizenship</td>
<td>Income</td>
</tr>
<tr>
<td>Migration</td>
<td>Housing and shelter costs</td>
</tr>
<tr>
<td>Aboriginal peoples</td>
<td>Place of birth</td>
</tr>
</tbody>
</table>
Demo Census Profiles

• [http://guides.library.utoronto.ca/INI235](http://guides.library.utoronto.ca/INI235)
• Go to “Urban Statistics” page & select first link
• Search your city by name
• Browse the full list of variables or toggle by topic
Demo Census Data Tables

• [http://guides.library.utoronto.ca/INI235](http://guides.library.utoronto.ca/INI235)
• Go to “Urban Statistics” page & select second link
• Select a topic & browse tables
Argh.. too much stuff! Choosing a theme

• Consider it a strategy for connecting variables & picking cities
• What information is available and what questions can it answer?
• Look at the variables. What interests you? What is a surprise?
• What connects variables? What other variables might you need for context?
Develop a search strategy

• Make note of possible search terms for each variable
• Think about synonyms, broader & narrower terms
• Be prepared for trial and error, as every database works a bit differently
Exercise

• Form groups of twos or threes

  Use the class research guide:
  http://guides.library.utoronto.ca/INI235

• You’re exploring poverty in the Toronto & Montreal CMAs

• Decide together which variables might be useful for this topic (pick 3)

• See if you can find Census tables containing your variables using the Census Data Tables or the Census Profiles

• Which variables did you choose?
What did they mean by that?

Non-census StatCan statistics for CMAs
Citing statistics

• Check bottom of page for “How to cite” instructions


• Use the StatCan citation guide’s reference builder
• Arrange all the required elements based on APA style guide (see page 29)
• Author / date for in text citations (see “Two or More Works by Same Author in Same Year” by OWL @ Purdue)
Getting more help

- Email me: nicholas.worby@utoronto.ca
- Ask Us Desk (2nd floor Robarts 11-5, M-Sat)
- Map & Data Library Reference Desk (5th floor Robarts, 11-5 M-F)
- Book a research consultation
- Ask Chat