

# Working with Open Data



## Handout

### ACQUIRING (FINDING) OPEN DATASETS




	<p><b>City of Vancouver Open Data Catalogue</b>  <a href="http://data.vancouver.ca">data.vancouver.ca</a></p>
	<p><b>Province of B.C. Open Data Catalogue</b>  <a href="http://data.gov.bc.ca">data.gov.bc.ca</a></p>
	<p><b>Government of Canada Open Government Portal</b>  <a href="http://open.canada.ca/en">open.canada.ca/en</a></p>
	<p><b>Research Data Canada</b>          Research Data Canada works with stakeholders to ensure research data is available to support innovation that benefits all Canadians.  <a href="http://www.rdc-drc.ca">www.rdc-drc.ca</a></p>
	<p><b>The Yale University Open Data Access (YODA)</b>          The Yale University Open Data Access (YODA) Project's mission is to advocate for the responsible sharing of clinical research data, open science, and research transparency.  <a href="http://yoda.yale.edu">yoda.yale.edu</a></p>
	<p><b>Instagram Open Data</b>          The Instagram API Platform can be used to build non-automated, authentic, high-quality apps and services.  <a href="http://www.instagram.com/developer">www.instagram.com/developer</a></p>
	<p><b>The Facebook Graph API</b>          The Graph API is the primary way to get data in and out of Facebook's platform.  <a href="http://developers.facebook.com/docs/graph-api">developers.facebook.com/docs/graph-api</a></p>

	<p><b>OpenWeatherMap</b>  OpenWeatherMap's weather station API collects data from thousands of remote and ground sensors distributed all around the world and is expanding permanently.  <a href="http://openweathermap.org">openweathermap.org</a></p>
	<p><b>Environment and Climate Change Canada</b>  Available through the Government of Canada Open Government Portal.  <a href="http://open.canada.ca/data/en/dataset?organization=ec">open.canada.ca/data/en/dataset?organization=ec</a></p>
	<p><b>The Guardian Data Blog</b>  The Guardian Data Blog regularly posts visualizations and makes data available through a Google docs format.  <a href="http://www.theguardian.com/data">www.theguardian.com/data</a></p>
	<p><b>New York Times Developer Network</b>  A great API and a really good explorer to access any article in the publication.  <a href="http://developer.nytimes.com">developer.nytimes.com</a></p>
	<p><b>Sports Reference</b>  A group of sites providing both basic and sabermetric statistics and resources for sports fans everywhere.  <a href="http://www.sports-reference.com">www.sports-reference.com</a></p>
	<p><b>InfoChimps</b>  A data platform and marketplace for sharing and selling data of all types.  <a href="http://www.infochimps.com">www.infochimps.com</a></p>
	<p><b>Qlik DataMarket</b>  Platform solutions for self-service data visualization, guided analytics applications, embedded analytics and reporting.  <a href="http://www.qlik.com/us/products/qlik-data-market">www.qlik.com/us/products/qlik-data-market</a></p>




	<p><b>CKAN (Comprehensive Knowledge Archive Network)</b>          CKAN is a powerful data management system that makes data accessible – by providing tools to streamline publishing, sharing, finding and using data.  <a href="http://ckan.org">ckan.org</a></p>
	<p><b>Google Advanced Search</b>  <a href="http://www.google.ca/advanced_search">www.google.ca/advanced_search</a></p>

## EXTRACTING (DOWNLOADING) OPEN DATASETS



	<p><b>Down Them All!</b>          An open-source freeware download manager/accelerator extension for Mozilla Firefox. It enables users to download all the links, images or embedded objects contained in a webpage.  <a href="http://addons.mozilla.org/en-US/firefox/addon/downthemall">addons.mozilla.org/en-US/firefox/addon/downthemall</a></p>
	<p><b>PostgreSQL</b>          PostgreSQL is an open source object-relational database system. It is free and open-source software, released under the terms of the PostgreSQL License, a permissive free-software license.  <a href="http://www.postgresql.org">www.postgresql.org</a></p>
	<p><b>MySQL</b>          MySQL Community Edition is the freely downloadable version of the world's most popular open source database. It is available under the GPL license and is supported by a huge and active community of open source developers.  <a href="http://www.mysql.com">www.mysql.com</a></p>




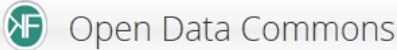

## CLEANING & PROCESSING OPEN DATASETS

	<p><b>Open Refine</b> A free, open source, power tool for working with messy data. <a href="http://openrefine.org">openrefine.org</a></p>
---	---

## PUBLISHING OPEN DATASETS

	<p><b>WordPress</b> An open source online publishing platform where you can create the perfect website or blog for your business, personal brand, event, or portfolio. <a href="http://wordpress.com/create-website">wordpress.com/create-website</a></p>
	<p><b>CKAN (Comprehensive Knowledge Archive Network)</b> CKAN is a powerful data management system that makes data accessible – by providing tools to streamline publishing, sharing, finding and using data. <a href="http://ckan.org">ckan.org</a></p>
	<p><b>Qlik DataMarket</b> Platform solutions for self-service data visualization, guided analytics applications, embedded analytics and reporting. <a href="http://www.qlik.com/us/products/qlik-data-market">www.qlik.com/us/products/qlik-data-market</a></p>
	<p><b>InfoChimps</b> A data platform and marketplace for sharing and selling data of all types. <a href="http://www.infochimps.com">www.infochimps.com</a></p>
<p><code>{"logo": "API Evangelist"}</code></p>	<p><b>The API Evangelist</b> API 101: information, tutorials, and tools to create your own API. <a href="http://apievangelist.com">apievangelist.com</a></p>



	<p><b>API Webinar Series</b>  Learn How to Use and Create APIs  An interactive webinar designed for beginners—the only thing you'll need to know is how to use a Google Spreadsheet. Gray Brooks, Senior API Strategist at General Services Administration's (GSA) Digital Services Innovation Center, and Jed Sundwall, co-founder of Measured Voice, will help you understand what an API is, how you can use them, and how to develop API-driven apps or recommend the creation of APIs at your own agency.  <a href="https://www.youtube.com/search?q=API+Webinar+Series">youtube.com (search "API Webinar Series")</a></p>
	<p><b>W3C on GitHub: Working With Metadata</b>  This document provides Best Practices related to the publication and usage of data on the Web.  <a href="https://github.com/w3c/dpub-metadata">github.com/w3c/dpub-metadata</a></p>
	<p><b>Open Data Commons</b>  Open Data Commons exists to provide legal solutions for open data.  Licenses Include:</p> <ul style="list-style-type: none"> <li>• Public Domain Dedication and License (PDDL) “Public Domain for data/databases”</li> <li>• Open Database License (ODC-ODbL) “Attribution Share-Alike for data/databases”</li> <li>• Attribution License (ODC-By) “Attribution for data/databases”</li> </ul> <a href="https://opendatacommons.org">opendatacommons.org</a>
	<p><b>Creative Commons</b>  Creative Commons provides free, easy-to-use copyright licenses to make a simple and standardized way to give the public permission to share and use your creative work—on conditions of your choice.  <a href="https://creativecommons.org">creativecommons.org</a></p>



## LET'S PRACTICE!



1.) Open up a Web Browser (e.g. Mozilla Firefox).

2.) Do a Google Search for “Ontario library open data”.

3.) Click on the “Ontario public library statistics | Ontario.ca” link or type <https://www.ontario.ca/data/ontario-public-library-statistics> into the address bar.

4.) Scroll down to “2016” from the listing of years.

5.) Click on the “CSV” link to open the dataset. When the pop-up window opens, select “Open with – Microsoft Office Excel (default)” and click “OK”.

6.) Clean up the data

- We only want the “CO.2.E English Print Volumes Held” statistics for Ontario libraries in year 2016 (column BF), so all other data can be deleted:
  1. Highlight all other columns to the right of column BF
  2. Delete the highlighted columns by right-clicking and selecting “delete” from the menu
  3. Follow steps above to delete all other columns (except column A) to the left of column BF
  4. *Be sure to keep column A “Library Full Name”*
- We only want to use the data showing the English print volumes statistics from libraries that start with the letter A, so delete all the rows after the libraries that start with A by highlighting the rows you want to delete, right-clicking, and selecting “delete” from the menu.
- Delete any rows that have the value “0” or do not have a number value in the English Print Volumes Held column.

7.) Save the Excel sheet as a CSV file by clicking “File” and “Save”. When the Save As pop-up box appears, name your file, select “CSV (Comma delimited) (\*.csv)” from the “Save As Type” drop-down menu. Select where you would like to save the file, then click “Save”.

8.) Open a new browser window. Google “Datawrapper” and select the “Datawrapper: Create Charts and Maps” link or type < <https://www.datawrapper.de/>> into the address bar.



9.) Click the blue “Create a Chart” button

10.) Select the “. . . or upload a CSV file” button

11.) Locate your saved CSV file and click “Open”

12.) Make sure Datawrapper interprets your data correctly

- Text should be in black and numbers should be in blue
- Click Column B: Under “Number Format” select to divide numbers by “1 thousandth (milli)” and round number to “0 (1,235)”
- If everything looks ok, click “Proceed” or number 3 “Visualize”

13.) Select the Chart Type (bar, column, pie, etc.). If the chart doesn’t look right, try clicking on the “Refine” tab to adjust your labels, values, colours, etc.

14.) Click the “Annotate” tab to add a title, description, or notes

15.) When you are done, click the number 4 “Publish & Embed” Link (if you submit your email address, you will be given the embedded code for your chart for you to use as desired).



## LEARN MORE ...

	<p><b>Learning Express Library</b>          The Popular Software Skills Centre offers learners tutorials on some of the most prevalent computer programs, including Microsoft Excel. Users will need to log-in with their valid Vancouver Public Library card and create a login and password.  <a href="http://www.vpl.ca/digital-library/learning-express-library">www.vpl.ca/digital-library/learning-express-library</a></p>
	<p><b>Lynda.com</b>          Online education database offering over 3,000 video courses by top experts in the fields of technology, business, and creativity.  <a href="http://www.vpl.ca/lynda">www.vpl.ca/lynda</a></p>
	<p><b>w3schools</b>          W3Schools is a web developers' site, with tutorials and references on many web development languages and most aspects of web programming.  <a href="http://www.w3schools.com">www.w3schools.com</a></p>
	<p><b>School of Data Online Courses</b>          A network of individuals and organizations working on empowering civil society organizations, journalists and citizens with skills they need to use data effectively.  <a href="http://schoolofdata.org/courses">schoolofdata.org/courses</a></p>
	<p><b>Open Data Stack Exchange</b>          A question and answer site for developers and researchers interested in open data.  <a href="http://opendata.stackexchange.com">opendata.stackexchange.com</a></p>





## VPL RESOURCES

<b>VPL Events Guide</b>	<a href="http://www.vpl.ca/events">www.vpl.ca/events</a>
<b>Lynda.com</b> <i>(Open Data: Unleashing Hidden value)</i>	<a href="http://www.vpl.ca/lynda">www.vpl.ca/lynda</a>
<b>VPL Information Services</b>	We are here to help! If you have any further questions, please ask us! Phone: 604-331-3603 / Email: <a href="mailto:info@vpl.ca">info@vpl.ca</a>
<b>VPL Subject Guides</b> Search <i>By Subject</i> > Technology & Computers	<a href="http://guides.vpl.ca">guides.vpl.ca</a>
<b>VPL One to One Training</b> individual help on topics covered in VPL courses.	Call 604-331-3603 to book an appointment.
<b>VPL Tech Cafe</b> A drop-in computer lab on Level 3 for practicing new skills with information staff assistance.	Central Library / 2:00 - 3:00 pm, Tuesdays & Thursdays

