

DATA 301 Introduction to Data Analytics Open Data

Dr. Ramon Lawrence
University of British Columbia Okanagan
ramon.lawrence@ubc.ca

DATA 301: Data Analytics (2)



What is Open Data?

Open Data is the movement to make data freely available to all with no restrictions on use or copyright.

Governments have been major supporters and providers of open data as data collected by governments is primarily done to benefit its citizens.

Corporations and other organizations are both producers and consumers of open data.

DATA 301: Data Analytics (3)

Open Data in Canada

Federal, provincial, and local governments have all been involved in the open data movement.

Canadian Federal government: <http://open.canada.ca/en>

- How to use: <http://open.canada.ca/en/working-data>
- Statistics Canada: <http://www.statcan.gc.ca/eng/rdc/data>

British Columbia government: <http://www.data.gov.bc.ca/>

City of Kelowna:

<https://www.kelowna.ca/city-services/city-maps-open-data/open-data-catalogue>

DATA 301: Data Analytics (4)

Open Data in Canada

DATA 301: Data Analytics (5)

Open Data in BC

DATA 301: Data Analytics (6)

Open Data in Kelowna

Dataset	ESOB	SRP	DSRP	ESM	CDV	Other	Request	Request	Request
Reserve Network									
Business Licenses	ESOB	SRP	DSRP	ESM	CDV				
City Boundary	ESOB	SRP	DSRP	ESM	CDV				
Centers 1m 2015	ESOB	SRP	DSRP	ESM	CDV				
Canal Contact Information					CDV				

Open Data in United States

United States government: <https://www.data.gov/>

Individual states have their own open data sites as well.

- Example: Washington state: <https://data.wa.gov/>

United States: Data.gov



Open Data Worldwide

UK: <http://data.gov.uk>

The World Bank: <http://data.worldbank.org/>

- Financial information and statistics

United Nations: <http://data.un.org/>

OECD: <https://data.oecd.org/>

Open Data Aggregators

There are many sites that aggregate open data sets (and some data sets for a cost). A Canadian based site is Quandl (<http://www.quandl.com>).

Kaggle provides many data sets and competitions and techniques for data analytics and machine learning.

<https://www.kaggle.com/datasets>

Open Data from Companies

Many companies either have public data or application programming interfaces (APIs) that allow people to use their data.

- Google: <https://www.google.com/publicdata/directory> (public data explorer) and <https://developers.google.com/maps/> (Google Maps API)
- Facebook: <https://developers.facebook.com/> (API)
- reddit: <https://www.reddit.com/dev/api> (API)
- Twitter: <https://dev.twitter.com/rest/public> (API)
- Amazon: <https://aws.amazon.com/public-data-sets/> (public data sets) and <https://developer.amazon.com/> (API for developers)
- Best Buy: <https://developer.bestbuy.com/> (API)

Try it: Open Data

Explore the federal, provincial, and City of Kelowna data sets to discover "something interesting". Report to your neighbors and to the class.

From any Canadian government open data site, retrieve a data set and analyze and visualize it using one of our tools: Excel, R, Python, Tableau.

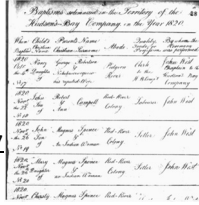
Open Data in Psychology and Social Sciences

Archaeology:

- Archaeology Data Service: <http://archaeologydataservice.ac.uk/>
- Many museums have online exhibits and open data.

Psychology:

- Journals increasing requiring open data sets.
- List of open data sites at: <http://guides.library.ucla.edu/c.php?g=180221&p=1188487>



History

- Digital Archive Database Project (UBC): <http://dadp.ok.ubc.ca>

Google Analytics

Google Analytics is an analysis service for tracking, optimizing, and understanding user interaction with a web site/service.

Using Google analytics is important for all business, but especially web companies, that rely on users interacting with their site to generate revenue and sales.

Google analytics helps identify and improve content to make it more accessible to potential customers.

- Very important skill set for business owners and managers.

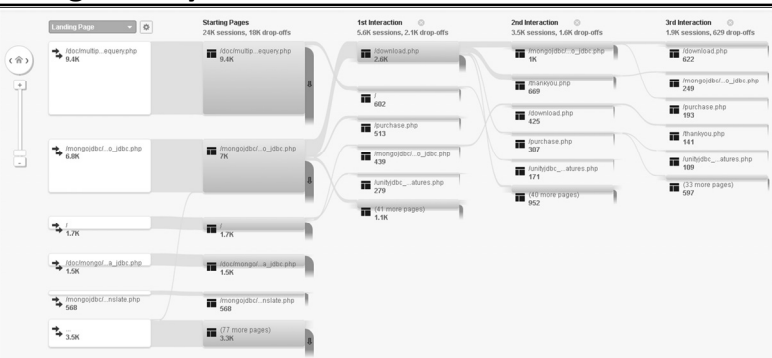
Google Analytics - Audience Overview

Google Analytics - Traffic Dashboard



Google Analytics - Behaviour Flow

Google Adwords



Google AdWords is a service to provide advertisements during searches and as display advertisements on web sites and in apps.

- Primary source of revenue for Google. <https://www.google.ca/adwords/>
- Companies bid on **keywords** and display opportunities that are presented by Google and affiliated sites.

Terminology:

- **Ad Impression** - display of an advertisement. Pricing in cost-per-thousand impressions or cost per mille (CPM).
- **Click through** - user clicks on an advertisement (and directly to new location)
- **Click through rate** - fraction of impressions that are clicked on
- **Pay-per-click (PPC)** - companies are billed on each click of an advertisement. The pricing depends on the bid amount and the desirability of the ad location.

Conclusion

Open Data is the movement to make data freely available to all with no restrictions on use or copyright.

Open data has been widely supported by governments and companies wishing to engage users (and developers) with their services.

Data analysts should use open data to help with their analysis whenever available.

Researchers are often responsible for making their publications and data available in an open fashion.

Google provides services for analytics and advertising that are valuable to understand as a business or site looking for user traffic.

Objectives

- Define open data and explain the motivations for making data "open".
- List some of the governments and organizations that provide data in an open fashion.
- Use open data sets when applicable when performing data analysis.
- Explain the role of Google Analytics and Google AdWords. Compare and contrast what these two services provide.