

Obtaining Data

DATA 201: Thinking With Data
Winter 2021

Jonathan Hudson, Ph.D
Instructor
Department of Computer Science
University of Calgary

Friday, January 15, 2021



Why do we want data?

Ethics — Key Points

- Ethics boards
- Informed consent
- Storing data
- Accessing data
- Anonymizing data
- Deleting data

How to obtain data?

Collecting Data

- Collect
- Generate
- Find / Extract

Collecting Data

- Collect
 - Observation, surveys, logging, sensors, crowdsourcing
- Generate
- Find / Extract

Collecting Data

- Collect
 - Observation, surveys, logging, sensors, crowdsourcing
- Generate
 - Simulations, experiment
- Find / Extract

Collecting Data

- Collect
 - Observation, surveys, logging, sensors, crowdsourcing
- Generate
 - Simulations, experiment
- Find / Extract
 - Open corpuses, data retailers, Application Program Interfaces (APIs), scraping the Web

Observation



Surveys



Qualtrics: Online Survey So x
www.qualtrics.com 1-800-340-9194

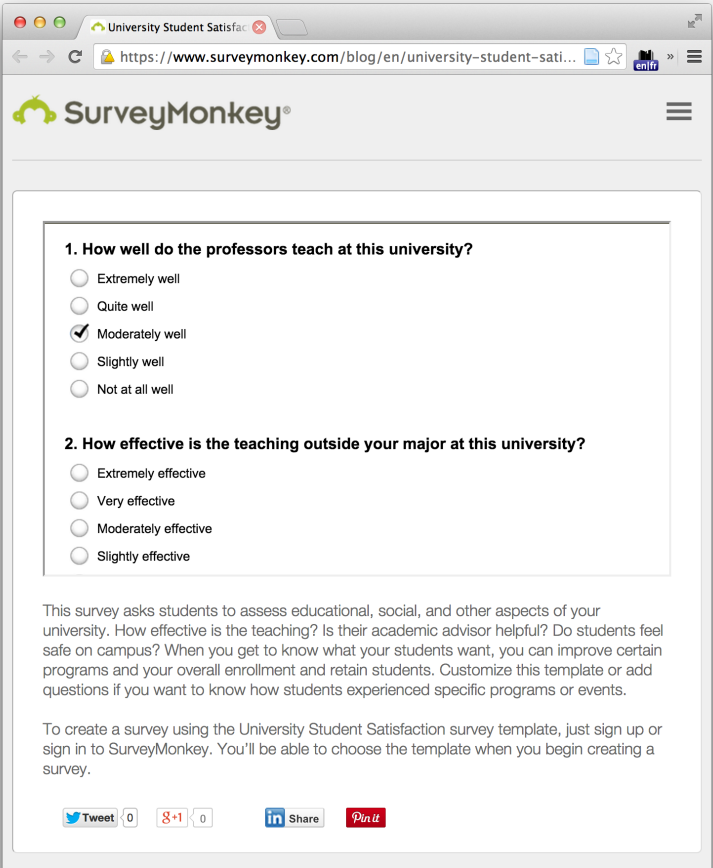
Ask Questions.
Get Answers. Act Fast.

Collect → Analyze → Act

Create your free account

Chat w/ Sales

The Ridiculously Powerful Insight Platform



University Student Satisfac x
https://www.surveymonkey.com/blog/en/university-student-sati...

SurveyMonkey

1. How well do the professors teach at this university?

- Extremely well
- Quite well
- Moderately well
- Slightly well
- Not at all well

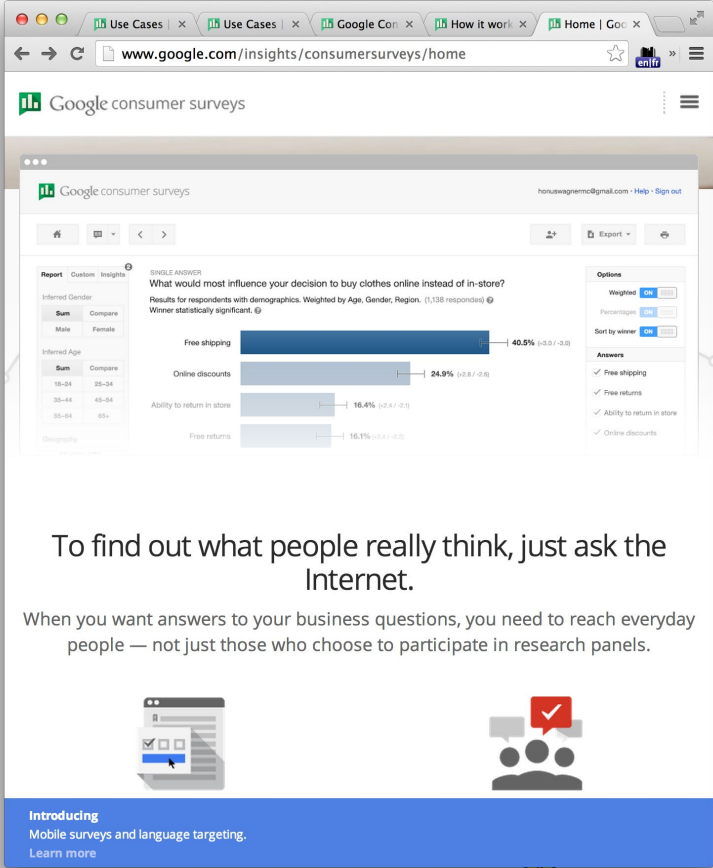
2. How effective is the teaching outside your major at this university?

- Extremely effective
- Very effective
- Moderately effective
- Slightly effective

This survey asks students to assess educational, social, and other aspects of your university. How effective is the teaching? Is their academic advisor helpful? Do students feel safe on campus? When you get to know what your students want, you can improve certain programs and your overall enrollment and retain students. Customize this template or add questions if you want to know how students experienced specific programs or events.

To create a survey using the University Student Satisfaction survey template, just sign up or sign in to SurveyMonkey. You'll be able to choose the template when you begin creating a survey.

Tweet 0 +1 0 Share Pin it



Use Cases x Use Cases x Google Con x How it work x Home | GoC x

www.google.com/insights/consumersurveys/home

Google consumer surveys

Report Custom Insights

What would most influence your decision to buy clothes online instead of in-store?
Results for respondents with demographics. Weighted by Age, Gender, Region. (1,538 responded) Winner statistically significant. (0)

Response	Percentage
Free shipping	40.5%
Online discounts	24.9%
Ability to return in store	16.4%
Free returns	16.1%

To find out what people really think, just ask the Internet.

When you want answers to your business questions, you need to reach everyday people — not just those who choose to participate in research panels.

Introducing Mobile surveys and language targeting. Learn more

General Guidelines

- Clear purpose/goal
- Every question is necessary
- Inform about survey length
- Simple and direct questions
- Logical order
- Avoid biased questions
- Use simple language
- Consistent with rating scale
- Test your survey
- Offer incentives

Month	Treatment	Pressure
March	Control	165
March	Placebo	163
March	300 mg	166
March	450 mg	168
April	Control	162
April	Placebo	159
April	300 mg	161
April	450 mg	163
May	Control	164

Surveys

Thank you for being a client of RBC.

This is a friendly reminder regarding an email we sent you recently.

You recently dealt with RBC over the phone.

We've shortened our survey to just 7 minutes and would like your thoughts on how we can better serve you.

Please complete the entire survey no later than **Wednesday, July 5, 2017.**

[Start the Survey](#)

Thank you for your time!

RBC Royal Bank and Ipsos

This survey is for service quality purposes only, you will not be asked to reveal personal account information. The information and feedback you provide will be shared with RBC to help them improve the service they provide to you and will be held in the strictest confidence by the bank.

Surveys

Dear Colleague,

You recently received an invitation to participate in a survey to tell us about the teaching culture at the University of Calgary. The purpose of this research is to develop a new survey instrument that can be used to provide feedback to universities on their teaching cultures, and how they can strengthen support for learning and teaching over time.

The research is funded by the Social Sciences and Humanities Research Council (SSHRC) and is being conducted at nine Canadian universities. The University of Calgary Conjoint Faculties Research Ethics Board has approved this study.

The survey will take 20 minutes to complete and is available until February 28, 2018.

Survey participation is confidential and voluntary; your decision regarding participation will not affect your position as a University of Calgary employee. You may decline to respond to any question by leaving it blank. If you have any questions about the research, please email the researchers at TCPS@ucalgary.ca or visit the [Institutional Teaching Culture Study](#) website

[Take the Institutional Teaching Culture Perception Survey](#)

Or copy and paste the URL below into your internet browser:

https://ucalgary.ca1.qualtrics.com/jfe/form/SV_57Q

In order for the questions to display properly, we strongly encourage you to complete this survey on a computer.

Thank you for considering this invitation.

Sincerely,

Lynn Taylor (former vice-provost teaching and learning)

Surveys



Surveys



Surveys

- Which movie genre do you think you like the most?
- What is your favourite movie genre?

Surveys

- Do you use ASCII to write emails?
- Do you use plain text to write emails?

Surveys

- Wouldn't you like fries as a side dish?
- Do you like fries as a side dish?

Surveys

I am very satisfied with the service.

1

Strongly
Disagree

2

3

4

5

Strongly
Agree

Surveys

I am ~~very~~ satisfied with the service.

1

Strongly
Disagree

2

3

4

5

Strongly
Agree

Likert Scale

1

Strongly
Disagree

2

Disagree

3

Neutral

4

Agree

5

Strongly
Agree

Likert Scale

Please indicate how much you agree or disagree with each of these statements:

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

The president is doing a good job.	1	2	3	4	5
The Congress is doing a good job.	1	2	3	4	5
The Secretary of Defense is doing a good job.	1	2	3	4	5

Likert Scale

Likert
Item

Please indicate how much you agree or disagree with each of these statements:

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

The president is doing a good job.	1	2	3	4	5
The Congress is doing a good job.	1	2	3	4	5
The Secretary of Defense is doing a good job.	1	2	3	4	5

Likert Scale

Please indicate how much you agree or disagree with each of these statements:

Likert
Item

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

The president is doing a good job.	1	2	3	4	5

Likert
Item

The Congress is doing a good job.	1	2	3	4	5
The Secretary of Defense is doing a good job.	1	2	3	4	5

Likert Scale

Please indicate how much you agree or disagree with each of these statements:

Likert
Item

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
--	----------------------	----------------------	-------------------------------------	-------------------	-------------------

The president is
doing a good job.

1

2

3

4

5

Likert
Item

The Congress is
doing a good job.

1

2

3

4

5

Likert
Item

The Secretary of
Defense is doing a
good job.

1

2

3

4

5

Likert Scale

Please indicate how much you agree or disagree with each of these statements:

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

The president is doing a good job.	1	2	3	4	5
The Congress is doing a good job.	1	2	3	4	5
The Secretary of Defense is doing a good job.	1	2	3	4	5

Likert Item

Response Level = 5

Likert Scale

Please indicate how much you agree or disagree with each of these statements:

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

The president is doing a good job.	1	2	3	4	5
The Congress is doing a good job.	1	2	3	4	5
The Secretary of Defense is doing a good job.	1	2	3	4	5

Likert Scale

Likert Item

Response Level = 5

Likert-Type Item

How often do you eat out?

1

2

3

4

5

Never

Sometimes

Average

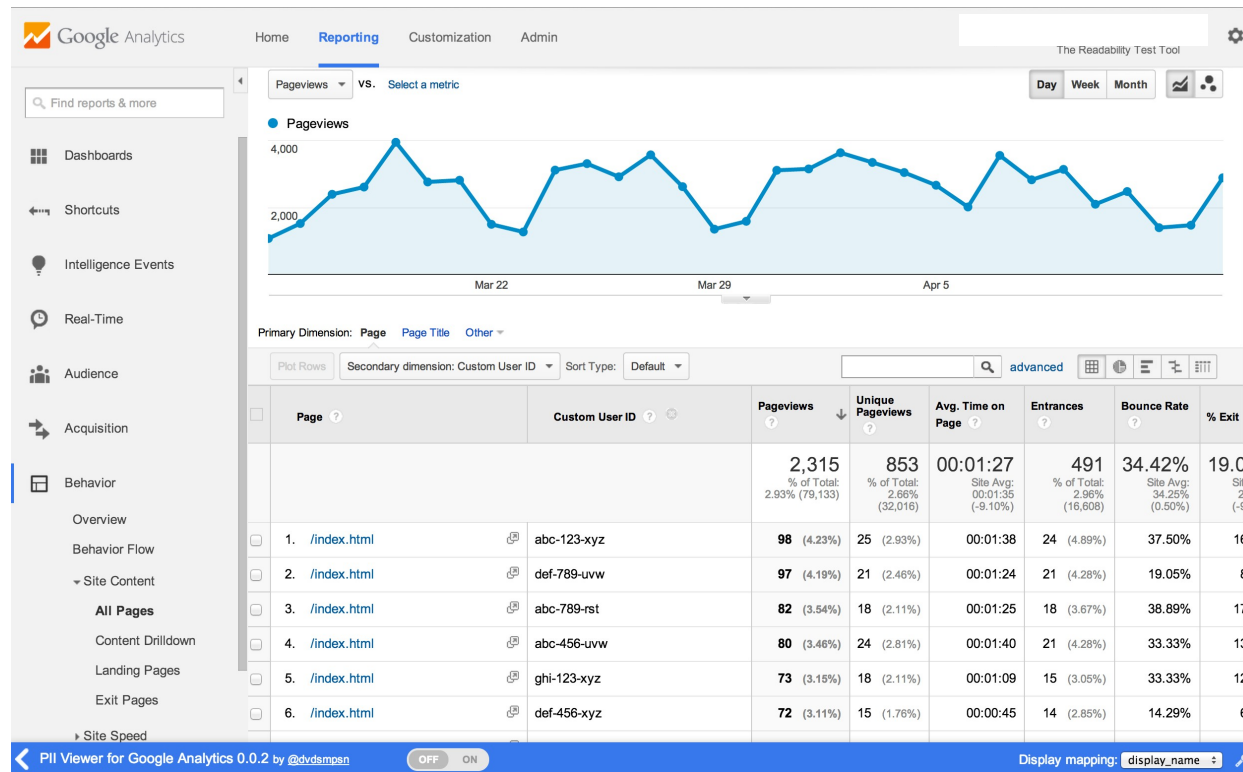
Often

Very often

Other Types of Questions

- Closed-ended questions
- Open-ended questions

Logging

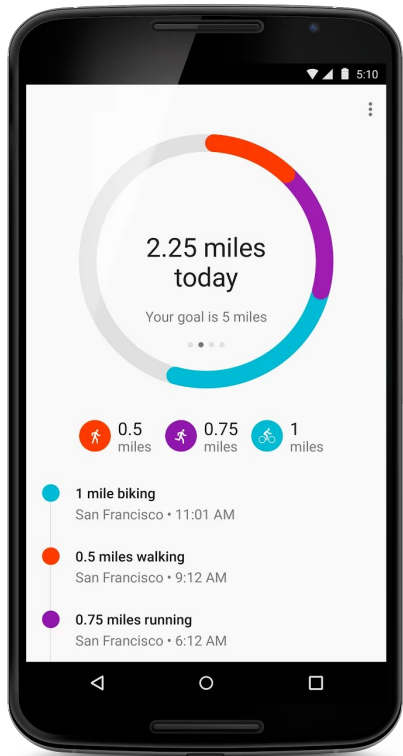


Google Analytics



Open Web Analytics

Sensors



Crowdsourcing

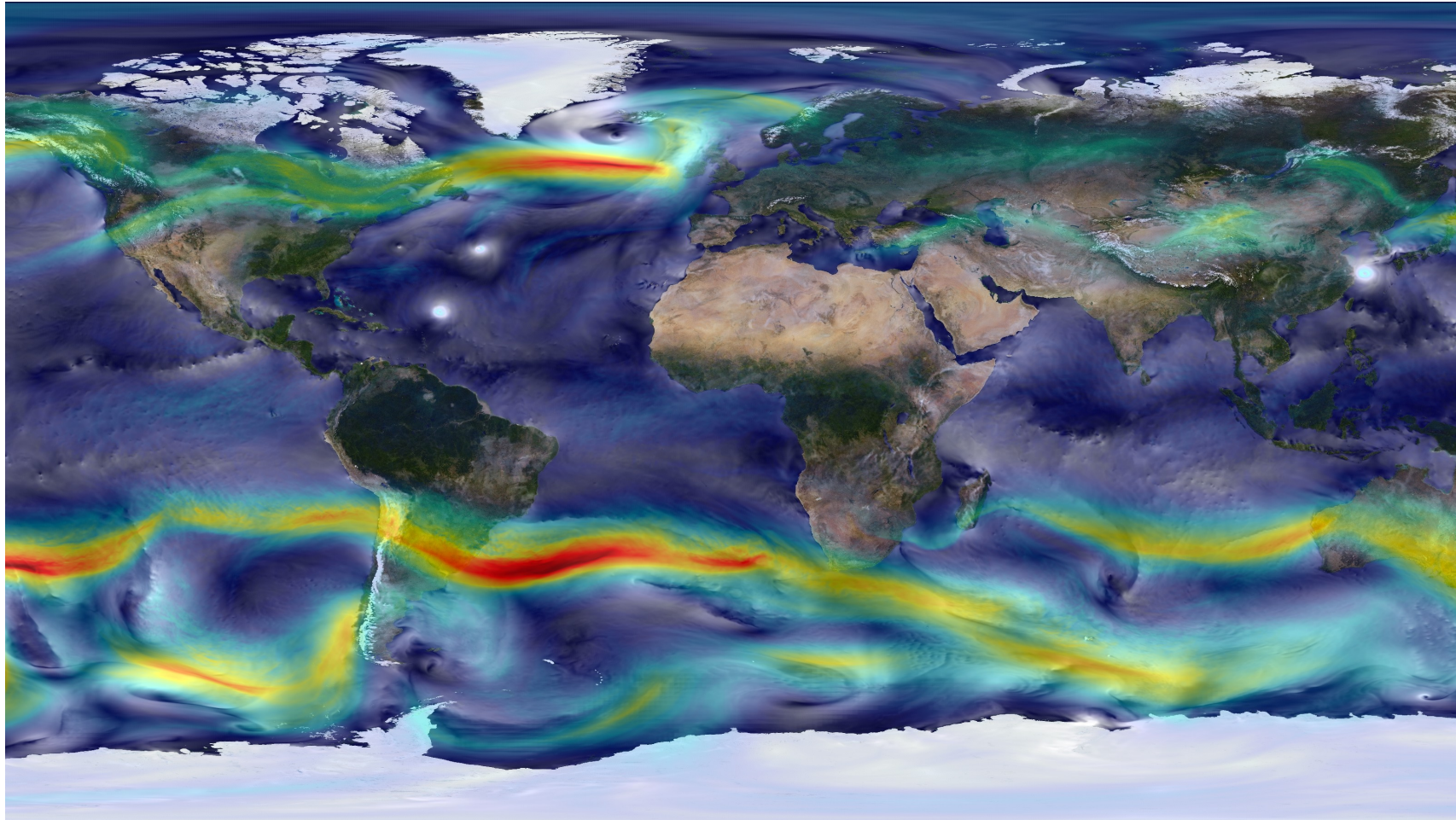
The screenshot shows the Amazon Mechanical Turk interface. The browser address bar displays <https://www.mturk.com/mturk/welcome>. The main heading is "HITs containing 'short survey'" with "11-20 of 49 Results". Below this, there are sorting options: "Sort by: HITs Available (most first)" and a "GO!" button. Navigation links include "Show all details", "Hide all details", and "First << Previous < 1 2 3 4 5 > Next >> Last".

HIT Title	Requester	HIT Expiration Date	Reward	Time Allotted	HITs Available
Answer a short survey about Work Team Dynamics	Whitney Ohmer	Oct 12, 2014 (2 weeks 5 days)	\$0.25	60 minutes	1
Answer a short survey about Work Team Dynamics	Whitney Ohmer	Oct 12, 2014 (2 weeks 5 days)	\$0.25	60 minutes	1
Short Survey	David Tannenbaum	Oct 12, 2014 (2 weeks 5 days)	\$0.10	60 seconds	1
Short survey about website experience (on average it takes 13 minutes)	peter slattery	Oct 6, 2014 (1 week 6 days)	\$1.50	40 minutes	1
Take a short survey	Jonathan Pirc	Oct 6, 2014 (1 week 6 days)	\$1.50	30 minutes	1
Answer a FUN and EASY survey about on consumer decision making!	CJJ	Oct 6, 2014 (1 week 6 days)	\$0.25	55 minutes	1

At the bottom of the page, there is a footer with links: "FAQ | Contact Us | Careers at Mechanical Turk | Developers | Press | Policies | Blog". Below these links is the copyright notice: "©2005-2014 Amazon.com, Inc. or its Affiliates". On the right side of the footer, it says "An amazon.com company".

Generating Data

Simulations

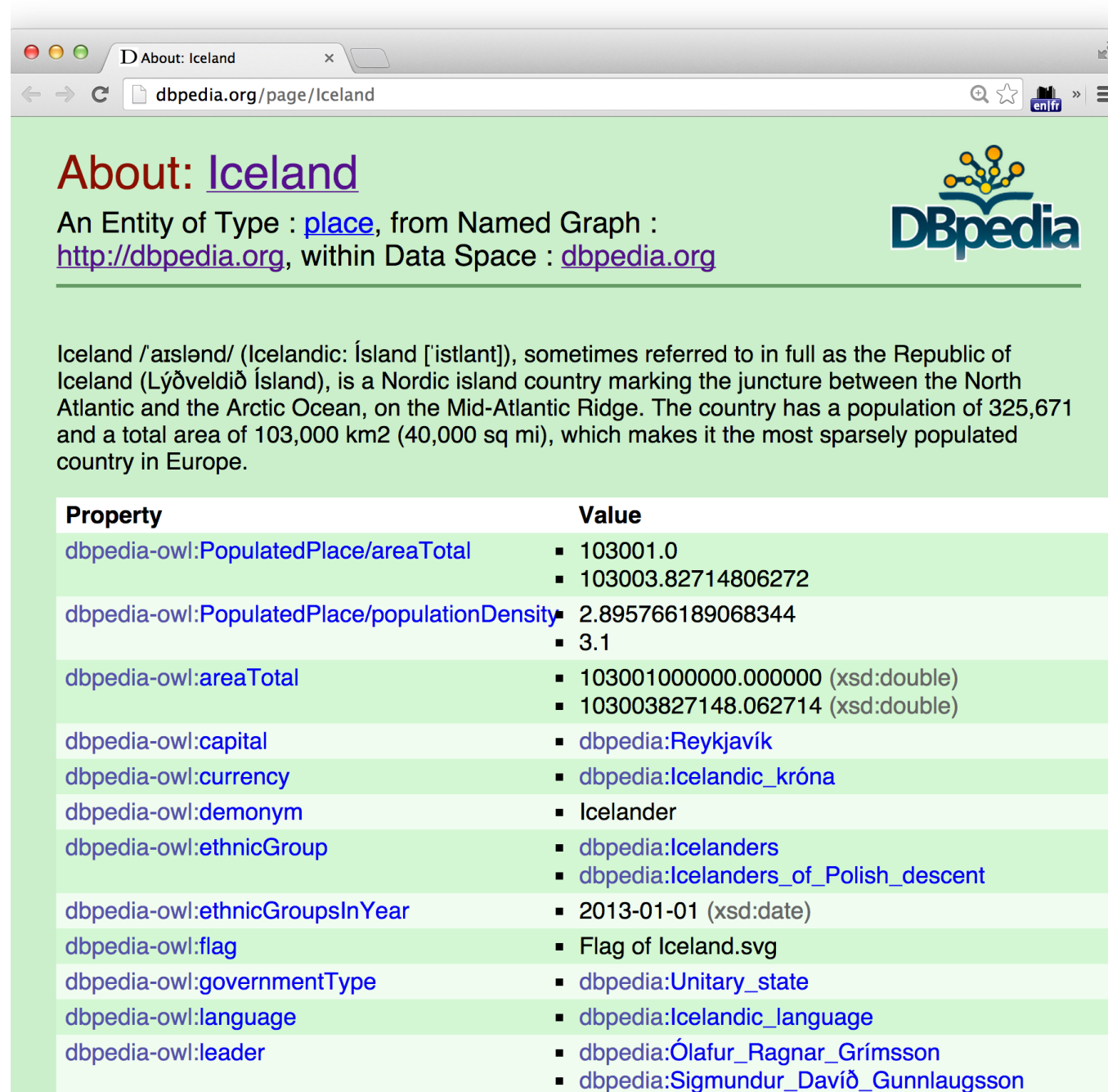


Experiment




Extracting / Finding Data

DBpedia



About: Iceland

An Entity of Type : [place](#), from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)



Iceland /aɪslənd/ (Icelandic: Ísland [ˈistlant]), sometimes referred to in full as the Republic of Iceland (Lýðveldið Ísland), is a Nordic island country marking the juncture between the North Atlantic and the Arctic Ocean, on the Mid-Atlantic Ridge. The country has a population of 325,671 and a total area of 103,000 km² (40,000 sq mi), which makes it the most sparsely populated country in Europe.


Property	Value
dbpedia-owl:PopulatedPlace/areaTotal	<ul style="list-style-type: none">103001.0103003.82714806272
dbpedia-owl:PopulatedPlace/populationDensity	<ul style="list-style-type: none">2.8957661890683443.1
dbpedia-owl:areaTotal	<ul style="list-style-type: none">103001000000.000000 (xsd:double)103003827148.062714 (xsd:double)
dbpedia-owl:capital	<ul style="list-style-type: none">dbpedia:Reykjavík
dbpedia-owl:currency	<ul style="list-style-type: none">dbpedia:Icelandic_króna
dbpedia-owl:demonym	<ul style="list-style-type: none">Icelander
dbpedia-owl:ethnicGroup	<ul style="list-style-type: none">dbpedia:Icelandersdbpedia:Icelanders_of_Polish_descent
dbpedia-owl:ethnicGroupsInYear	<ul style="list-style-type: none">2013-01-01 (xsd:date)
dbpedia-owl:flag	<ul style="list-style-type: none">Flag of Iceland.svg
dbpedia-owl:governmentType	<ul style="list-style-type: none">dbpedia:Unitary_state
dbpedia-owl:language	<ul style="list-style-type: none">dbpedia:Icelandic_language
dbpedia-owl:leader	<ul style="list-style-type: none">dbpedia:Ólafur_Ragnar_Grímssondbpedia:Sigmundur_Davíð_Gunnlaugsson

Wikidata

https://www.wikidata.org/wiki/Wikidata:Main_Page

English Not logged in Talk Contributions Create account Log in

Main Page Discussion Read View source View history Search



WIKIDATA

- Main page
- Community portal
- Project chat
- Create a new item
- Item by title
- Recent changes
- Random item
- Nearby
- Help
- Donate

Print/export

- Create a book
- Download as PDF
- Printable version

Tools

- What links here
- Related changes
- Special pages
- Permanent link
- Page information
- Wikidata items

open collaborative

multilingual free

Welcome to Wikidata

the free knowledge base with 16,194,852 data items that anyone can edit

[Introduction](#) • [Project Chat](#) • [Community Portal](#) • [Help](#)

Welcome!

Wikidata is a free, linked database that can be read and edited by both humans and machines.

Wikidata acts as central storage for the structured data of its

Learn about data

New to the wonderful world of data? [Develop and improve your data literacy through content](#) designed to get you up to speed and feeling comfortable with the fundamentals in no time.

Google books Ngram Viewer

Graph these comma-separated phrases: case-insensitive

between and from the corpus with smoothing of . [Search lots of books](#)



(click on line/label for focus)

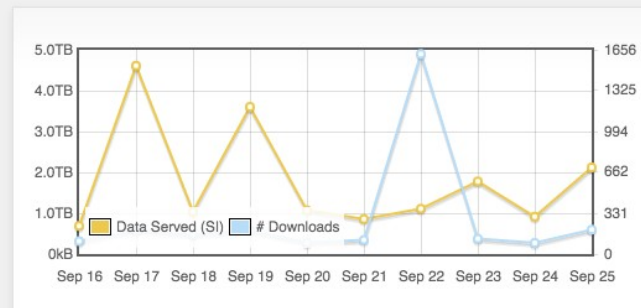
Search in Google Books:

- | | | | | | |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|----------------------------|
| 1800 - 1845 | 1846 - 1974 | 1975 - 1985 | 1986 - 1993 | 1994 - 2000 | chocolate |
| 1800 - 1840 | 1841 - 1913 | 1914 - 1924 | 1925 - 1990 | 1991 - 2000 | vanilla |
| 1800 - 1820 | 1821 - 1840 | 1841 - 1852 | 1853 - 1977 | 1978 - 2000 | strawberry |

Run your own experiment! Raw data is available for download [here](#).

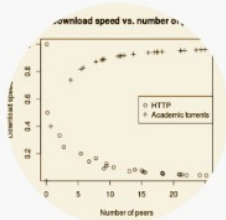
Making 15.72TB of research data available!

We've designed a distributed system for sharing enormous datasets - for researchers, by researchers. The result is a scalable, secure, and fault-tolerant repository for data, with blazing fast download speeds. Contact us at contact@academictorrents.com.



[View popular!](#)

[Upload a dataset!](#)



Accelerate your hosting for free with our academic BitTorrent infrastructure!



Distribute your public data globally for free to ensure it is available forever!

Sponsored by: **Institute for Reproducible Research** and **Whatbox** and

Ads!
Disable your ad-blocker!



Sports Reference | Sports Stats, fast, easy, and up-to-date

A group of sites providing both basic and sabermetric statistics and resources for sports fans everywhere. Our aim is to be the easiest-to-use, fastest, most complete sources for sports statistics anywhere.

Baseball-Reference.com (Mobile)

Baseball statistics from 1871 to the present for major league [players](#), [teams](#), and [leagues](#). Complete [postseason](#) and [managerial](#) data is included as well. [Minor league stats](#) back to 1877, and [box scores](#), gamelogs and splits back to 1912. Updated daily.



Basketball-Reference.com

Professional basketball statistics for every season from [Abdul-Jabbar](#) to [Zaslofsky](#). Find statistics for your favorite [player](#), [team](#), or [league](#). The site also includes sections for [coaches](#), [awards](#), [leaders](#), and the [playoffs](#). A user favorite is the [Play Index](#) custom leaderboards. Updated daily.



Pro-Football-Reference.com

Football statistics and game results for all of NFL, AFL, AAFC & APFA history. The site includes [player](#), [Coach](#), and [team](#) stats, games scores for each team, Pro Bowl selections for each season, [extensive leaderboards](#), and every [draft pick](#) ever selected. Our [play index](#) allows for custom searches through every NFL box score since 1960, every TD in NFL history and every play since 1994. Additionally, you can search through every play in Super Bowl history. Updated the morning after every NFL game.



Hockey-Reference.com

Hockey statistics from the creators of *Total Hockey* to our website. Find statistics for our favorite [player](#) from [Aalto](#) to [Zyuzin](#), [team](#), or [league](#). The site also includes sections for [coaches](#), [drafts](#), [awards](#), [leaders](#), and the [playoffs](#). Analyze players using [the Play Index](#).



College Football @ Sports-Reference.com

College Football game results for every major school back to [Princeton vs Rutgers](#) in 1869. [Player Stats](#) for [recent players schools](#), or [conferences](#) back to 1956. The site also includes sections for [coaches](#), [awards](#), [leaders](#), and [bowl games](#). Use the [Play Index](#) to create your own custom searches.



College Basketball @ Sports-Reference.com

College Basketball Stats and Game Result. Game results for every game featuring two major schools since 1949. [Player Stats](#) from [Stephen Curry](#) to [Bill Walton](#). Stats for [schools](#) and [conferences](#). The site also includes sections for [coaches](#), [awards](#), [leaders](#), and the [postseason](#). Use the [Play Index](#) to create your own custom searches.



New to this site? [Start Here](#)

[DataBank](#) [Microdata](#) [Data Catalog](#)

This page in: [English](#) [Español](#) [Français](#) [العربية](#) [中文](#)

World Bank Open Data

Free and open access to global development data

Search data e.g. GDP, population, Indonesia

Browse by [Country](#) or [Indicator](#)

MOST RECENT

[New country classifications by income level: 2017-2018](#)

World Bank Data Team, Jul 01, 2017

[Q2 2017 Update of World Development Indicators Available](#)

World Bank Data Team, Jul 01, 2017

[Chart: Globally, The Number of People Without Access to Electricity is Falling](#)

T. Khokhar, Jun 30, 2017

[Chart: Globally, Over 1 Billion](#)

[View all news](#)

[View all blogs](#)

WHAT YOU CAN LEARN WITH OPEN DATA

[Poverty headcount ratio at \\$1.90 a day \(2011 PPP\) \(% of population\)](#)

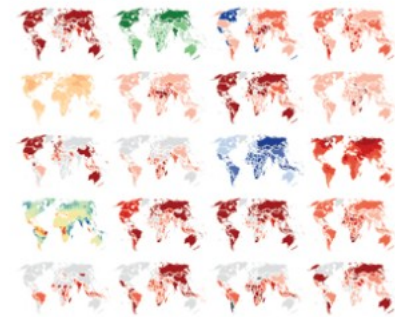


Extreme Poverty

The proportion of the world's population living in extreme poverty has dropped significantly

Atlas of Sustainable Development Goals 2017

From World Development Indicators



A Visual Guide to the 17 SDGs

Apr 17, 2017

Find, compare and share the latest OECD data: charts, maps, tables and related publications ...

Search

Browse by **topic** or **country**

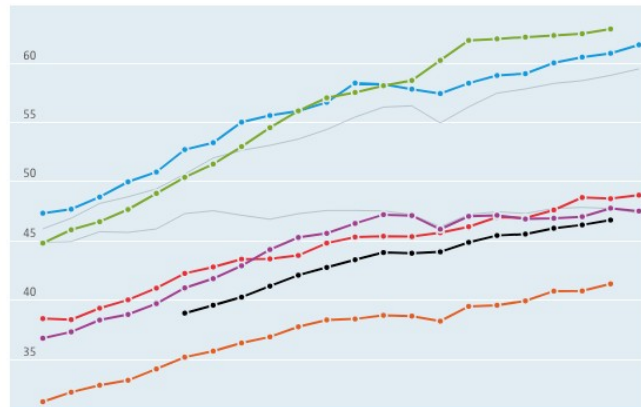
Search tips
Catalogue of OECD databases

OECD Data

Featured Charts

Labour productivity differences across OECD countries remain significant. Labour productivity growth continues its long-term decline, including in some G7 economies.

GDP per hour worked Total, US dollars, 1995 – 2016



Labour productivity and utilisation

Labour productivity, Annual growth rate (%), 2000 – 2016



Latest news

Statistical news releases

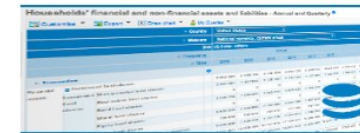
See recent statistical news releases.

Compare your income

Try OECD's Compare your income to check your perception of income inequality against reality!

Statistical resources

Database access



Make your own queries on large databases in our data warehouse, OECD.Stat.

Query databases

Statistical tools

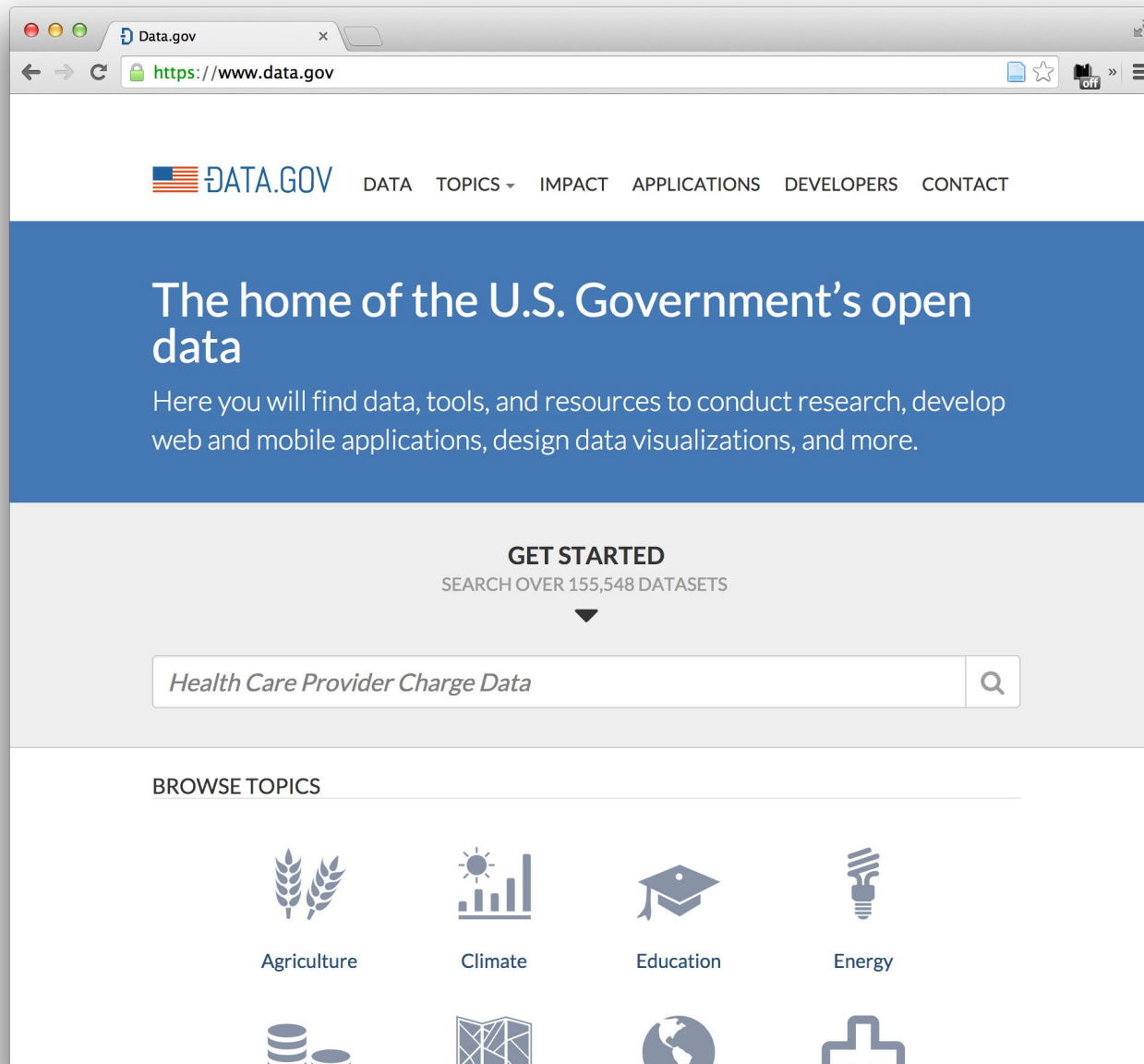
Learn about OECD methodologies and stay informed through statistical newsletters.

Go to statistical tools

Developer API

Learn about OECD's application programming interface (API) for accessing data.

Go to API for developers



The screenshot shows a web browser window with the URL `data.gov.uk/data/search`. The page header includes the **DATA.GOV.UK** logo with the tagline "Opening up Government" and navigation links for Home, Data, Apps, and Interact. A search bar at the top right contains the text "Search for data...".

The main content area features a secondary search bar with the text "Search for data..." and a magnifying glass icon. Below it, a green button labeled "Show Search Facets »" is visible. The search results are displayed as "19422 Results" and are sorted by "Popularity".

Three search results are listed:

- Live traffic information from the Highways Agency**: A result from the Highways Agency, described as "Live traffic information data showing traffic information on the strategic road network in England, maintained by the Highways Agency. Update: 12th August 2013 Following a change of...". It includes format options: NII, PDF, and XML, and is categorized under "Transport".
- Learning Aim Reference Service**: A result from the Skills Funding Agency, described as "Learning Aim Reference Service (LARS) service will offer a 'Quick Search' facility, allowing users to search by most commonly used fields (the current full set of search fields will still...". It includes format options: CSV, HTML, PDF, and Zip, and is categorized under "Education".
- English Indices of Deprivation 2010**: A result from the Department for Communities and Local Government, described as "The English Indices of Deprivation 2010 provide a relative measure of deprivation at small area level across England. Areas are ranked from least deprived to most deprived on seven...". It includes format options: CSV, HTML, and XLS, and is categorized under "Society".

The browser's address bar at the bottom shows `data.gov.uk/data/search#`.

data.gov.be

data.gov.be/fr/datasets

nl fr de en

Beta

Data.gov.be

HOME CONDITIONS D'UTILISATION DATA APPS IDÉES FORUM CONTACT LIENS

Liste de sets de données disponibles comme "open data".

Catégorie Type Granularité

- Tout - - Tout - - Tout - Appliquer

Titre	Catégorie	Type	Granularité
Zones de stationnement voirie 2013	Mobilité	Téléchargement	
Usages TIC des ménages wallons	TIC	Téléchargement, Service web	
Usages TIC des citoyens wallons	Population, Economie, TIC	Téléchargement, Service web	
UDP Mars 2013 par commune	Energie, Pouvoirs publics	Téléchargement	
UDP Mai 2013 par commune	Energie, Pouvoirs publics	Téléchargement	
UDP Janvier 2013 par habitant	Population, Energie	Téléchargement	
UDP Janvier 2013 par commune	Energie, Pouvoirs publics	Téléchargement	
UDP Février 2013 par commune	Energie, Pouvoirs publics	Téléchargement	
UDP Evolution Mensuelle	Energie	Téléchargement	
data.gov.be/fr/apps	Mobilité	Téléchargement	



- Jobs ▾
- Immigration ▾
- Travel ▾
- Business ▾
- Benefits ▾
- Health ▾
- Taxes ▾
- More services ▾

Home → [Open Government](#)

101 101 Open Data

Search open data that is relevant to Canadians, learn how to work with datasets, and see what people have done with open data across the country.

Follow:

Search through our Open Government Portal.

Find

[Search through our Open data portal](#)

Looking for data about Government of Canada services, financials, national demographic information or high resolution maps? Discover that and more through our open data portal, your one-stop shop for Government of Canada open datasets.

[Open Data Inventory](#)

Review the open data inventories submitted by Government of Canada departments and agencies.

[Working with data and application programming interfaces](#)

Learn more about how to use a variety of data formats, including application programming interfaces (APIs).

[Apps gallery](#)

Browse and download apps that were built both by the Canadian public and the Government of Canada that use open data.

[Open maps](#)

Explore the Government of Canada's geospatial data, services, and applications and create customized maps.

[Open data 101](#)

Learn what open data is, how to get the most out of it, and explore the principles that are guiding the Government of Canada's open data initiatives.

[Suggest a dataset](#)

Can't find the dataset that you are looking for? Let us know!

[Submit your app](#)

Have you made an application using Government of Canada open data or open information? If so, let us know and we will share it with Canadians in our [apps gallery](#).



OPEN GOVERNMENT PROGRAM

Open Data

Search a listing of Government of Alberta open datasets. Can't find what you are looking for? [Suggest a resource.](#)

Search Open Data

Enter search terms...

Refine Results

Topics

- Energy and Natural Resources (1364)
- Health and Wellness (357)
- Economy and Finance (158)
- Population and Demography (137)
- Agriculture (128)
- Employment and Labour (123)
- Business and Industry (105)
- Society and Communities (63)
- Arts, Culture and History (62)
- Government (54)

[SHOW MORE](#)

Publisher

- Alberta Energy Regulator (1339)
- Health (308)
- Treasury Board and Finance (243)
- Agriculture and Forestry (123)
- Economic Development and Trade (80)

2,308 results

Sort Results by

Relevance

Average Weekly Earnings (including overtime), Alberta

OPENDATA
CSV PDF

LAST MODIFIED
June 29, 2017

DESCRIPTION
This Alberta Official Statistic describes the average weekly earnings of all payroll employees in Alberta. Earnings include overtime and are seasonally adjusted (current dollars).

TAGS

AOS AWE ALBERTA OFFICIAL STATISTICS OFFICIAL STATISTIC SEPH

Provincial Base Map Series - 1:1 000 000 Scale Maps



OPENDATA
PDF

LAST MODIFIED
June 28, 2017

DESCRIPTION
This cartographic quality series of 1:1 000 000 scale colour maps cover the provincial extent of Alberta. The primary provincial base map displays the Alberta Township System (ATS), major hydrographic features, municipalities, major roads, railways and select geoadministrative features (parks,...


TAGS

CARTOGRAPHY EARTH SCIENCES GIS GEOGRAPHIC DATA
GEOGRAPHY GEOMATICS MAP TOPOGRAPHY TOPONYMY


Home Browse Data Citizen Dashboard Discussion Forum About Developer Info Help

[f](#)
[t](#)
[You Tube](#)
[in](#)
••
[Instagram](#)
Sign In




Open Calgary


Happy 150th Canada! Welcome to Open Calgary, your gateway to The City of Calgary's open data portal and citizen dashboard. In the spirit of openness, accountability and transparency, Open Calgary was created to facilitate the sharing of information, spark innovative ideas and foster a sense of collaboration among citizens.




Browse Data



Citizen Dashboard




Help




Solar Potential Images

Explore solar potential images of buildings in Calgary.



Historic Rainfall Data

Check out our rainfall gauge data since 1998.



2015 Normalized Bike Volumes

See 2015 data on two way bike volumes in the downtown core.



Data Sources for Data Science

Non-GIS and non-statistical data sources, primarily to support the Data Science Certificate

[Home](#)[Subject-specific data sources](#)[Directories and search engines](#)[Licenses and permissions](#)[Submit your data](#)

This guide

This guide links to sources of non-GIS and non-statistical data, including research data. It was created primarily to support the Data Science Certificate program, and mostly links to sources for raw data files.

Do you have a resource to suggest for this guide? Email [Jennifer Lee](#), Science Librarian.

For GIS (Geographic Information System), Statistics Canada, and City of Calgary data, see the [Spatial and Numeric Data Services guides](#).

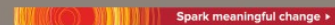
[!\[\]\(f507db636256ac11a5525ef93ec6b8d7_img.jpg\) Print Page](#)

Libraries and Cultural Resources | [Email](#) | [SMS](#) | [403.220.8895](#) | [Live Chat](#)



[Privacy Policy](#)
[Website feedback](#)

energize
THE CAMPAIGN FOR EYES HIGH



University of Calgary
2500 University Drive NW
Calgary, AB T2N 1N4
CANADA

Copyright © 2017



Tweepy

[Read the Docs](#)

[Discuss with Others](#)

[Report a Bug](#)

Tweepy

An easy-to-use Python library for accessing the Twitter API.

 Fork

1,552

 Star

3,630

The New York Times Developer Network

All the APIs Fit to POST

You already know that NYTimes.com is an unparalleled source of news and information. But now it's a premier source of data, too — why just read the news when you can hack it?

Getting Started

The Times Developer Network is our API clearinghouse and community. Here's how to get started:

1. Request an [API key](#)
2. Read the API documentation, [FAQ](#) and [Terms of Use](#)
3. Use the API Tool associated with each API to experiment without writing code

Using The NYT APIs

[Terms of Use](#)

Before you can use New York Times APIs, you must agree to the Terms of Use.

[Attribution Guidelines and Restrictions](#)

Please review these guidelines before you use our APIs.

[Branding Guidelines](#)

Please review the Branding guidelines before you use our APIs, this includes information on the use of our logo, image, and branding components.

[API Key Registration](#)

Ready to start coding? Request a key for each API you want to use.

Spotify Developer

Music, meet code. Powerful APIs, SDKs and widgets for simple and advanced applications



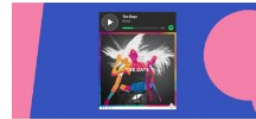
Spotify Web API

Our Web API lets your applications fetch data from the Spotify music catalog and manage user's playlists and saved music. [Read more](#)



Web API Console

Quickly test requests to our Web API, get back live JSON data and formatted cURL requests, with full OAuth and scope support.



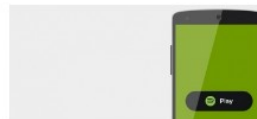
Spotify Play Button

Add a Spotify Play Button to your blog, website, or social page—all your fans have to do is hit "Play" to enjoy the music. [Read more](#)



Spotify iOS SDK

(Beta) The Spotify iOS SDK makes it easy to add streaming, user authentication and other Spotify features to iOS apps. [Read more](#)



Spotify Android SDK

(Beta) The Spotify Android SDK makes it easy to add audio streaming, user authentication and other Spotify features to Android apps. [Read more](#)



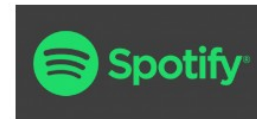
Integrate Spotify Connect into your product

Want to integrate Spotify Connect to give your users a seamless music experience over wifi? [Read more](#)



Developer Showcase

Here are just a few of the latest applications that have been built with our APIs, SDKs, and other developer tools. [Read more](#)



Design Resources

Hello partner. Welcome to our hub for partner guidelines and assets. We want to make it easy for you to integrate Spotify in your app while respecting our brand and legal/licensing restrictions. Note that by using these resources, you accept

[Read More](#)



Steam Web API Documentation

Valve provides these APIs so website developers can use data from Steam in new and interesting ways. They allow developers to query Steam for information that they can present on their own sites. At the moment the only APIs we offer provide item data for Team Fortress 2, but this list will grow over time.

Steam Web APIs available

ISteamNews: Steam provides methods to fetch news feeds for each Steam game.

ISteamUserStats: Steam provides methods to fetch global stat information by game.

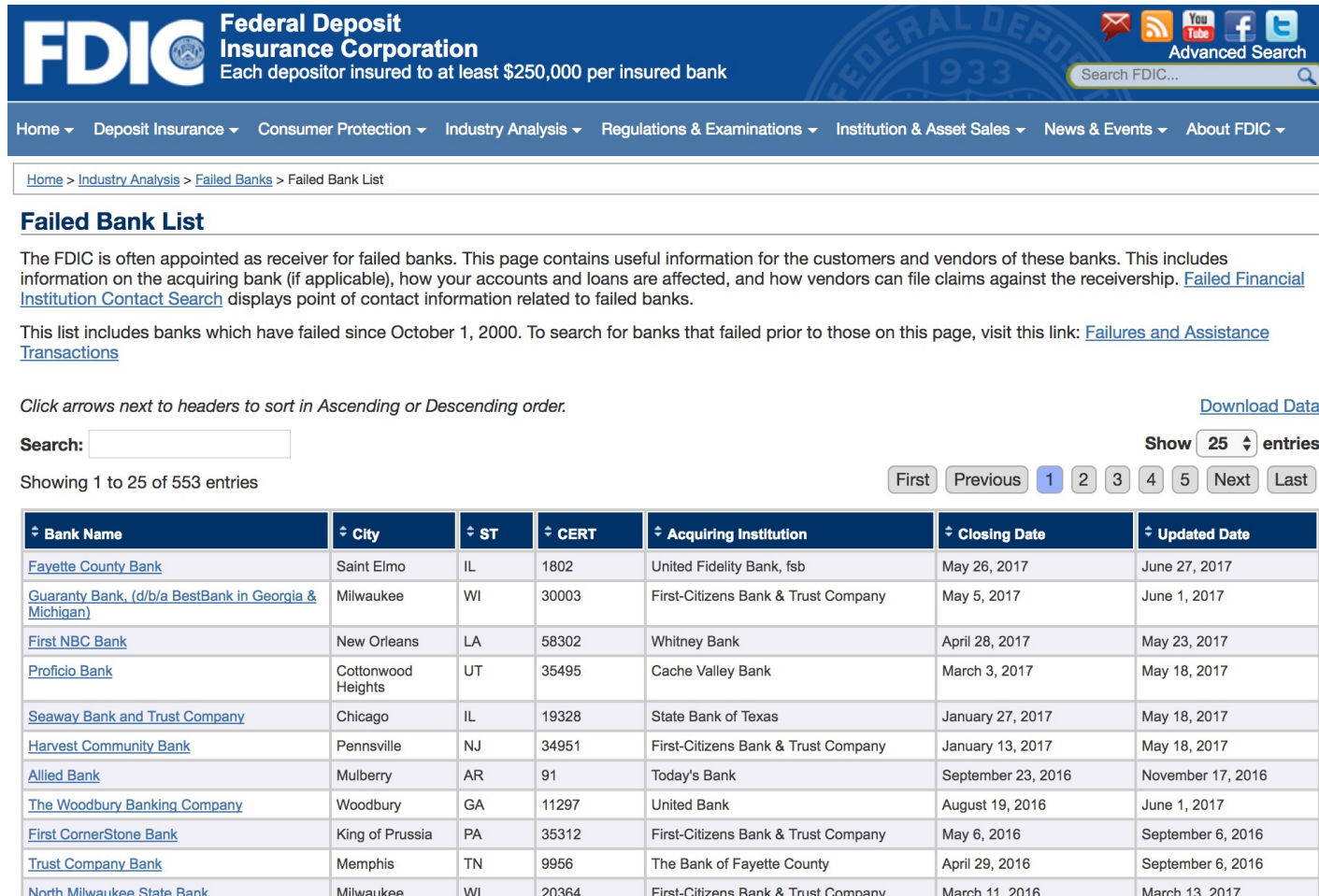
ISteamUser: Steam provides API calls to provide information about Steam users.

ITFItems_440: Team Fortress 2 provides API calls to use when accessing player item data.

Obtaining an Steam Web API Key

All use of the Steam Web API requires the use of an API Key. You can acquire one **by filling out this form**. Use of the APIs also requires that you agree to the **Steam API Terms of Use**.

Scraping



The screenshot shows the FDIC website's 'Failed Bank List' page. At the top, the FDIC logo and 'Federal Deposit Insurance Corporation' are displayed, along with the tagline 'Each depositor insured to at least \$250,000 per insured bank'. A navigation menu includes links for Home, Deposit Insurance, Consumer Protection, Industry Analysis, Regulations & Examinations, Institution & Asset Sales, News & Events, and About FDIC. A search bar is located in the top right corner.

The main content area features a breadcrumb trail: Home > Industry Analysis > Failed Banks > Failed Bank List. Below this is the section title 'Failed Bank List'. A paragraph explains that the FDIC is often appointed as receiver for failed banks and provides information on the acquiring bank, account status, and vendor claims. A link for 'Failed Financial Institution Contact Search' is provided. Another paragraph notes that the list includes banks failed since October 1, 2000, and provides a link for 'Failures and Assistance Transactions'.

Below the text, there is a search input field, a 'Download Data' link, and a 'Show 25 entries' dropdown. A pagination bar shows 'First', 'Previous', '1', '2', '3', '4', '5', 'Next', and 'Last' buttons.

Bank Name	City	ST	CERT	Acquiring Institution	Closing Date	Updated Date
Fayette County Bank	Saint Elmo	IL	1802	United Fidelity Bank, fsb	May 26, 2017	June 27, 2017
Guaranty Bank, (d/b/a BestBank in Georgia & Michigan)	Milwaukee	WI	30003	First-Citizens Bank & Trust Company	May 5, 2017	June 1, 2017
First NBC Bank	New Orleans	LA	58302	Whitney Bank	April 28, 2017	May 23, 2017
Proficio Bank	Cottonwood Heights	UT	35495	Cache Valley Bank	March 3, 2017	May 18, 2017
Seaway Bank and Trust Company	Chicago	IL	19328	State Bank of Texas	January 27, 2017	May 18, 2017
Harvest Community Bank	Pennsville	NJ	34951	First-Citizens Bank & Trust Company	January 13, 2017	May 18, 2017
Allied Bank	Mulberry	AR	91	Today's Bank	September 23, 2016	November 17, 2016
The Woodbury Banking Company	Woodbury	GA	11297	United Bank	August 19, 2016	June 1, 2017
First CornerStone Bank	King of Prussia	PA	35312	First-Citizens Bank & Trust Company	May 6, 2016	September 6, 2016
Trust Company Bank	Memphis	TN	9956	The Bank of Fayette County	April 29, 2016	September 6, 2016
North Milwaukee State Bank	Milwaukee	WI	20364	First-Citizens Bank & Trust Company	March 11, 2016	March 13, 2017



	A	B	C	D	E	F	G
1	=ImportHtml("http://www.fdic.gov/bank/individual/failed/banklist.html", "table", 0)						
2	Northern Star Bank	Mankato	MN	34983	BankVista Bank of Southern California,	12/19/2014	1/5/2015
3	Frontier Bank, FSB D/B/A El Paseo Bank	Palm Desert	CA	34738	N.A.	11/7/2014	11/21/2014
4	The National Republic Bank of Chicago	Chicago	IL	916	State Bank of Texas	10/24/2014	1/5/2015
5	NBRS Financial	Rising Sun	MD	4862	Howard Bank	10/17/2014	11/10/2014
6	GreenChoice Bank, fsb	Chicago	IL	28462	Providence Bank, LLC	7/25/2014	1/5/2015
7	Eastside Commercial Bank	Conyers	GA	58125	Community & Southern Bank	7/18/2014	1/5/2015
8	The Freedom State Bank	Freedom	OK	12483	Ava State Bank & Trust Company	6/27/2014	7/18/2014
9	Valley Bank	Fort Lauderdale	FL	21793	Landmark Bank, National Association	6/20/2014	7/28/2014

ROBERT KOSARA / NOVEMBER 15, 2009

The Simple Way to Scrape an HTML Table: Google Docs

Raw data is the best data, but a lot of public data can still only be found in tables rather than as directly machine-readable files. One example is the FDIC's [List of Failed Banks](#). Here is a simple trick to scrape such data from a website: Use Google Docs.

The table on that page is even relatively nice because it includes some JavaScript to sort it. But a large table with close to 200 entries is still not exactly the best way to analyze that data.

After some digging around – and even considering writing my own throw-away extraction script –, I remembered having read something about [Google Docs](#) being able to [import tables from websites](#). And indeed, it has a very useful function called `ImportHtml` that will scrape a table from a page.

MORE DATA ARTICLES

[Temperature Baseline Differences](#)

[A Look At Climate Data](#)

[Interactively Explore Climate Data](#)

[A Browser for Data.gov](#)

[Data Is A Dish Best Served Raw](#)

RECENTLY POPULAR

[Understanding Pie Charts](#)

[The Simple Way to Scrape an HTML](#)

[Table: Google Docs](#)

[Data: Continuous vs. Categorical](#)

[What is Visualization? A Definition](#)

[You Only See Colors You Can Name](#)

[EuroVis 2017 Conference Report. Part 2](#)

[EuroVis 2017 Conference Report. Part 1](#)



SUBSCRIBE VIA EMAIL

Data Formats

Data Formats

- Structured Data:
- Unstructured Data:
- Semi-structured Data:

Data Formats

- Structured Data: spreadsheet or database
- Unstructured Data:
- Semi-structured Data:

Data Formats

- Structured Data: spreadsheet or database
- Unstructured Data: raw text, streaming data, images, videos
- Semi-structured Data:

Data Formats

- Structured Data: spreadsheet or database
- Unstructured Data: raw text, streaming data, images, videos
- Semi-structured Data: somewhere in between

CSV (Comma-Separated Value)

```
1 firstName, lastName, age, streetAddress, city
2 John, Smith, 25, 21 2nd Street, New York, NY, 10021
```

firstName	lastName	age	streetAddress	city	state	postalCode	homePhoneNumber	faxPhoneNumber
John	Smith	25	21 2nd Street	New York	NY	10021	212 555-1239	646 555-4567

JSON (JavaScript Object Notation)

```
{
  "firstName": "John",
  "lastName": "Smith",
  "age": 25,
  "address": {
    "streetAddress": "21 2nd Street",
    "city": "New York",
    "state": "NY",
    "postalCode": "10021"
  },
  "phoneNumber": [
    {
      "type": "home",
      "number": "212 555-1239"
    },
    {
      "type": "fax",
      "number": "646 555-4567"
    }
  ]
}
```

firstName	lastName	age	streetAddress	city	state	postalCode	homePhoneNumber	faxPhoneNumber	gender
John	Smith	25	21 2nd Street	New York	NY	10021	212 555-1239	646 555-4567	male

XML (eXtensible Markup Language)

```
<person>
  <firstName>John</firstName>
  <lastName>Smith</lastName>
  <age>25</age>
  <address>
    <streetAddress>21 2nd Street</streetAddress>
    <city>New York</city>
    <state>NY</state>
    <postalCode>10021</postalCode>
  </address>
  <phoneNumbers>
    <phoneNumber type="home">212 555-1234</phoneNumber>
    <phoneNumber type="fax">646 555-4567</phoneNumber>
  </phoneNumbers>
  <gender>
```

firstName	lastName	age	streetAddress	city	state	postalCode	homePhoneNumber	faxPhoneNumber	gender
John	Smith	25	21 2nd Street	New York	NY	10021	212 555-1239	646 555-4567	male

Handling Data

Storing Data

Storing Data

- Backup
- Password protect or encrypt sensitive information

Provenance

- Keep track of where and when data was collected
- Record data collection / processing steps

IP and Copyright

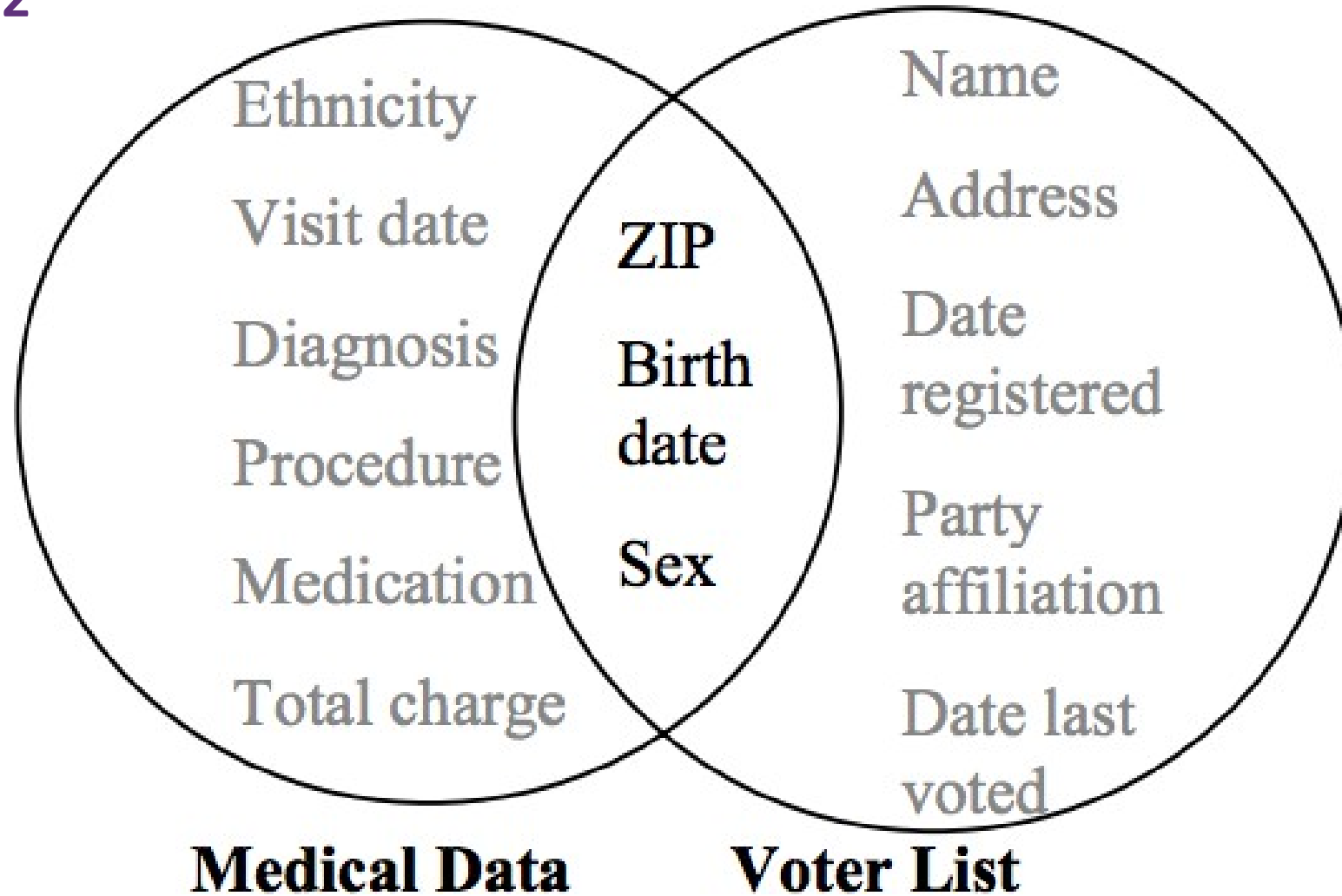
- Know who owns the data
- Do you need to publish or (re)share data?
- Do NOT violate copyright

Privacy

- Any information that could be used to identify individuals is sensitive
- May have legal repercussions for releasing it
- Anonymize data before sharing

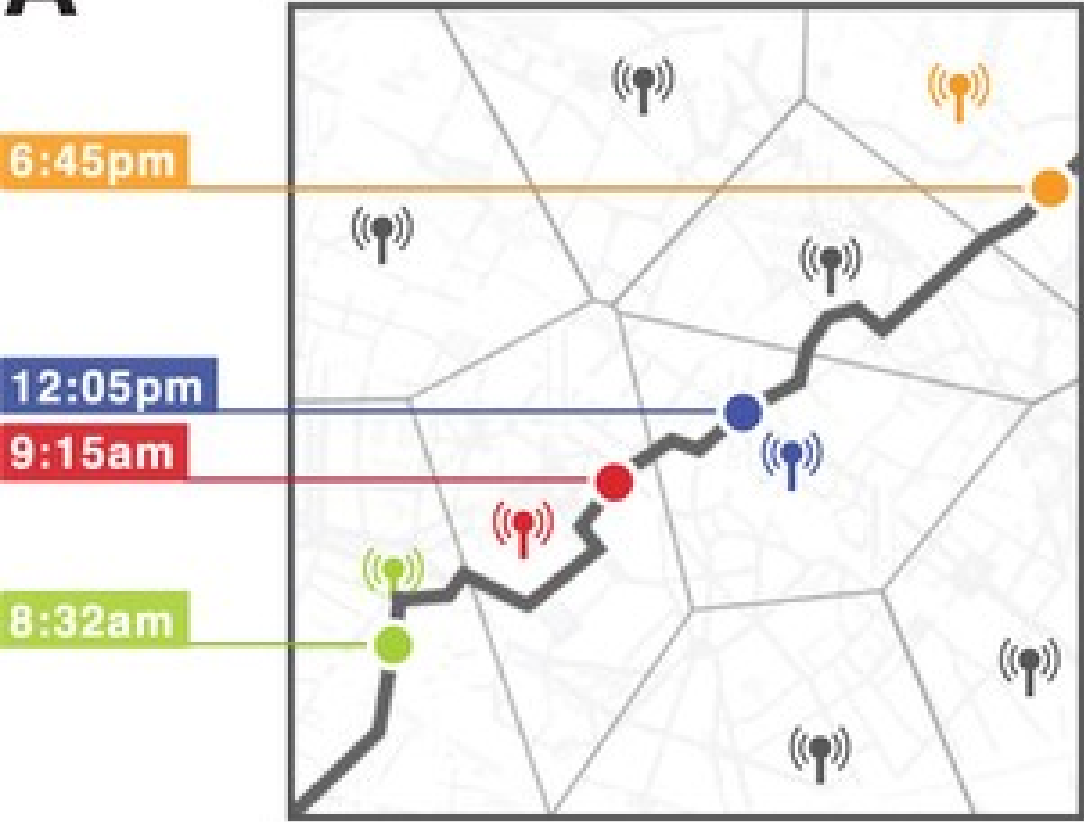
How to anonymize data?

L. Sweeney. 2002

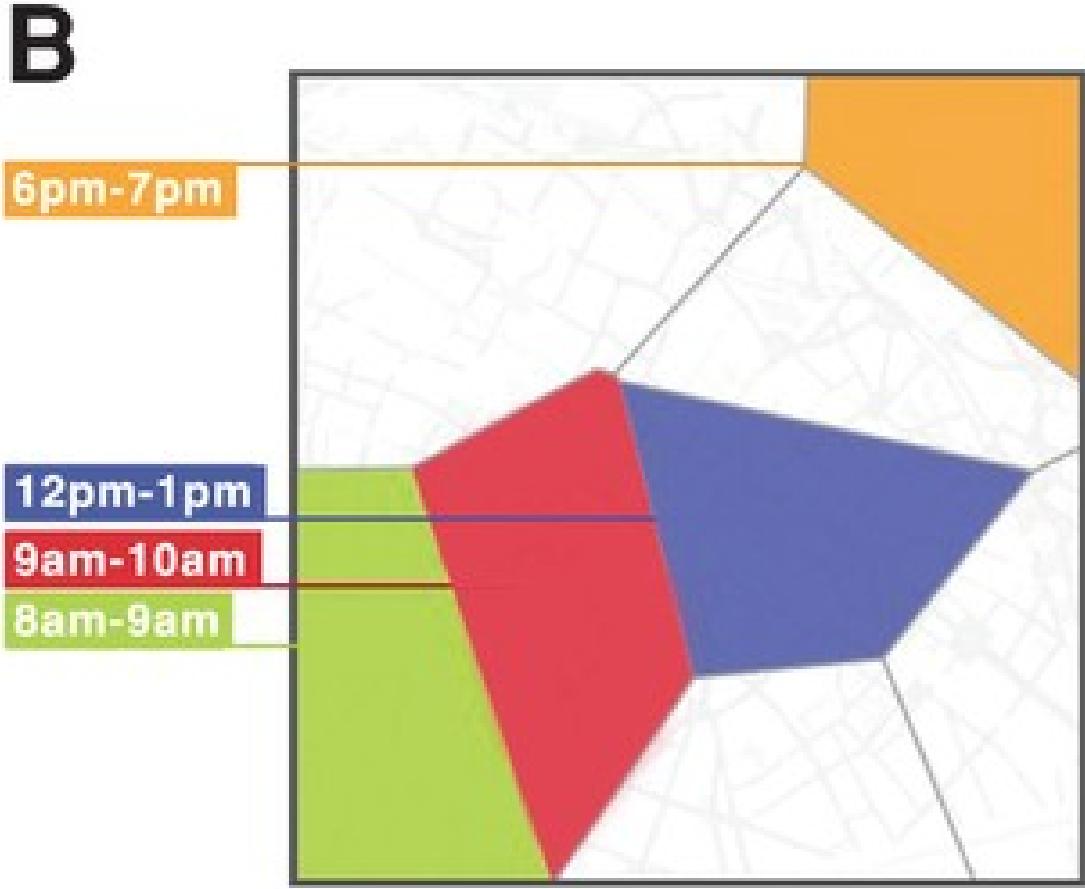
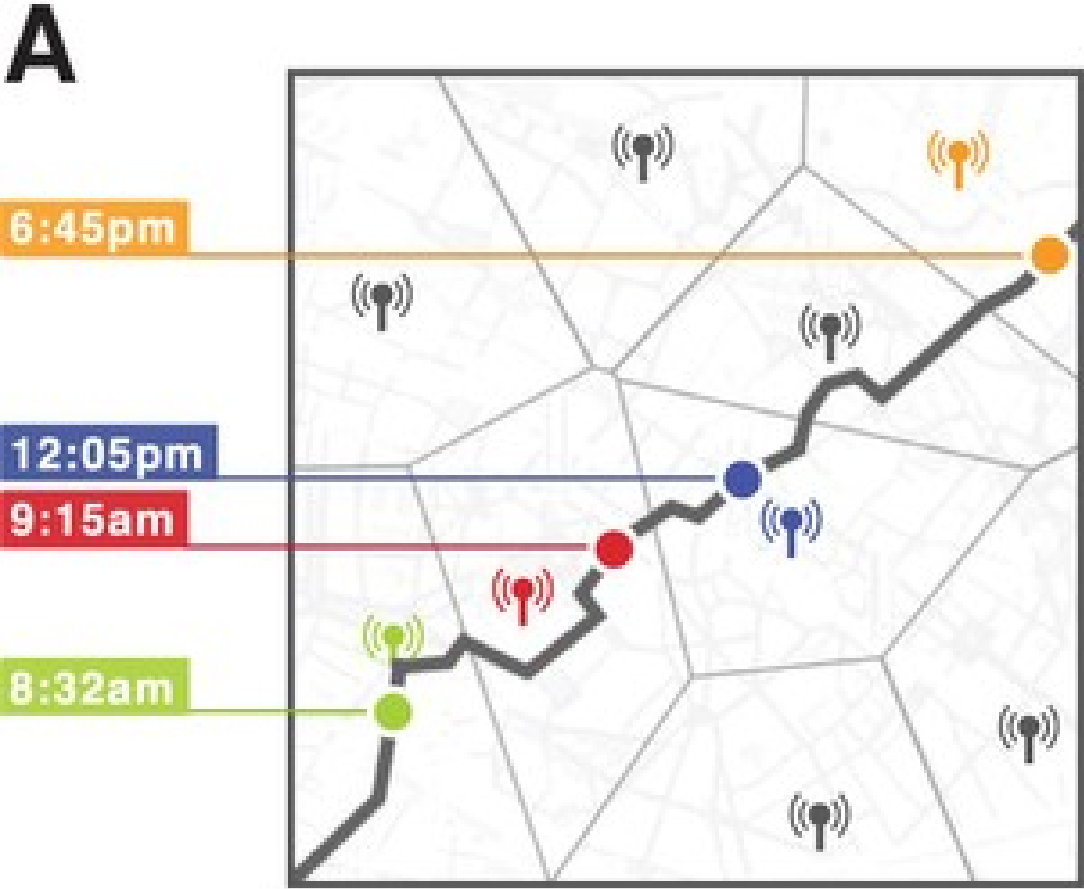


de Montjoye et al. 2013

A

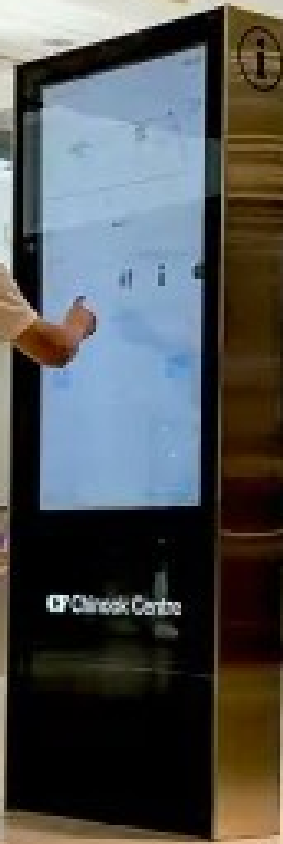


de Montjoye et al. 2013



Ethics

WEATHER



NORDSTR



Onward to ... Data Cleaning

Jonathan Hudson
jwhudson@ucalgary.ca
<https://pages.cpsc.ucalgary.ca/~jwhudson/>



UNIVERSITY OF
CALGARY