

# A Brief Introduction to Federal, State & Local Public Data



Image: Jon Snow's cholera map via John Snow Archive and Research Companion

Brandon Locke | @brandontlocke | CC-BY

# What do we mean by public data?

A very broad definition for our purposes today:

Information created by or pertaining to local, state, & federal government or its citizens

- Datasets created by governments through research or records created through standard operations
- Records and government documents in formats amenable to computation
- Datasets created by citizens to document government operations
-

[FILTER](#) » [OPTIONS](#) » [REVIEW](#)
[APPLY FILTERS](#) 
[HOW TO USE THE DATA FINDER](#) 
[GEOGRAPHIC LEVELS](#)


OR



1930 OR 1940 OR 1950 OR 1960 OR 1970 OR 1980 OR 1990 OR 2000 OR 2010

[TOPICS](#)
[DATASETS](#)
[RESET FILTERS](#)
[SELECT DATA](#) 
[13354 SOURCE TABLES](#)
[380 TIME SERIES TABLES](#)
[632 GIS BOUNDARY FILES](#)

 PAGE  OF 19    

VIEW 1 - 20 OF 380

POPULARITY	TABLE NAME	GEOGRAPHIC INTEGRATION	YEARS	GEOGRAPHIC LEVELS
	<a href="#">Total Population</a>	Nominal	1790, 1800, 1810, 1820, 1830, 1840, 1850, 1860, 1870, 1880, 1890, 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990, 2000, 2010	STATE, COUNTY
	<a href="#">Total Population</a>	Nominal	1970, 1980, 1990, 2000, 2010, 2008-2012	STATE, COUNTY, TRACT, CTY_SUB, PLACE
	<a href="#">Total Population</a>	Nominal	1980, 1990, 2000, 2010, 2008-2012	NATION, REGION, DIVISION, STATE, COUNTY, TRACT, CTY_SUB, PLACE
	<a href="#">Total Population</a>	Start	1990, 2000, 2010	STATE, COUNTY, TRACT, BLCK_GRP, CTY_SUB, PLACE, CD10TH-112TH, CBSA, URB_AREA, ZCTA



Have you used public data for anything?

 michigan state main library

GO

193 Records

Date Range: 10-3-2017 to 10-9-2017 (7 Days)



## FILTERS



## SUMMARY



## WHAT



## WHERE



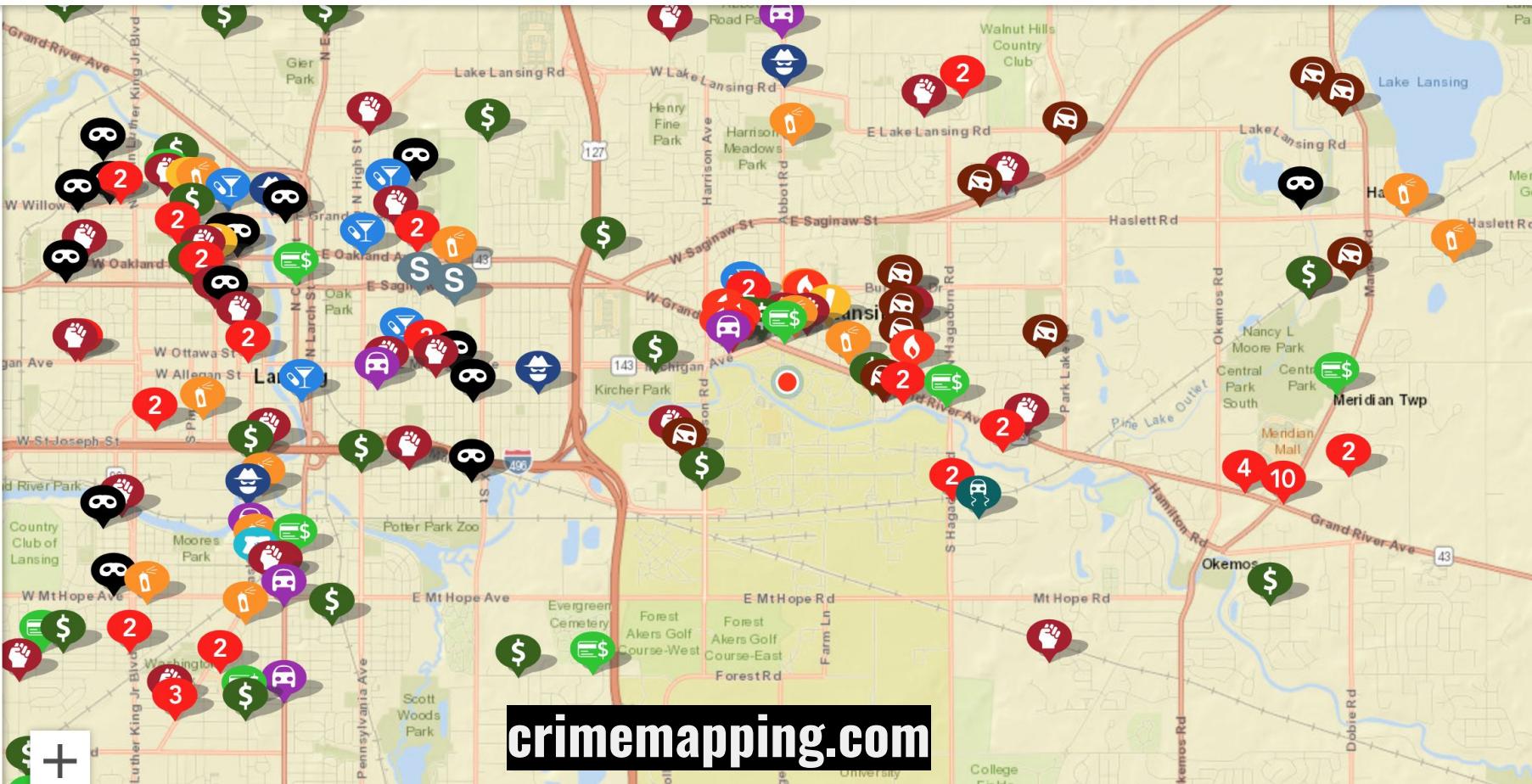
## WHEN



## REPORT



## CHARTS



crimemapping.com

# Bicycle Collision Analytics (Lincoln, NE)

Many of the graphs on this page are now clickable to get even more data out of them. Try it out!

In March 2015, I was hit by a car while on my bike in a bike path crosswalk, and was ticketed for failure to yield the right of way. I got interested in bicycle safety and, as a software engineer by trade, I decided to download, parse, and quantify as much collision data as possible from the Lincoln Police Department.

Last updated: 2017-08-07 11:08  
By Chris St. Pierre

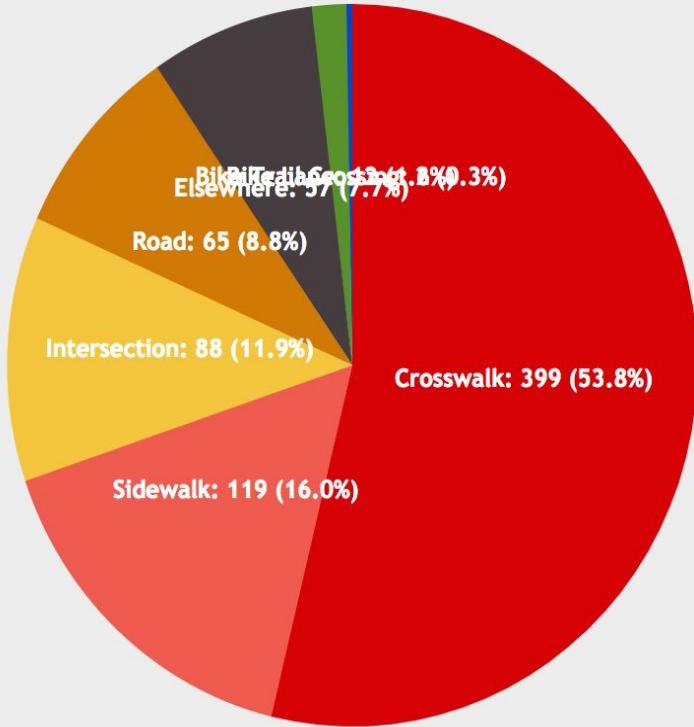
## Interpreting These Data

In the data below, bike-car collisions are categorized based on where they happened. The following locations are tracked:

- **Bike Lane:** Collision happened while a person on a bicycle was riding in a bike lane, including in an intersection.
- **Bike Trail:** Collision happened while a person on a bicycle was riding on a bike trail, not in an intersection.
- **Bike Trail Crossing:** Collision happened while a person on a bicycle was using a crosswalk between two bike trail segments, or continuing past the end of a bike trail over a crosswalk to the start of a sidewalk. This does not include cyclists crossing perpendicularly to a bike trail, unless they are continuing on another bike trail on the other side.
- **Crosswalk:** Collision happened while a person on a bicycle was using a crosswalk between two sidewalks; that is, attempting to cross a road or intersection either in a painted crosswalk that is not part of a bike trail, or crossing an intersection from one sidewalk to another. This does not include cyclists who are riding on the street and pass through the intersection without using a crosswalk, nor cyclists in a crosswalk that is part of a bike trail.
- **Elsewhere:** Collision happened elsewhere (alleyways, parking lots, etc.). This also includes ride-outs at mid-block and other collisions that happened on the road, but where the cyclist was not riding on the road as such.
- **Intersection:** Collision happened at an intersection on the road, not using a crosswalk. Does not include cyclists who entered the intersection on a bike trail.

[stpierre.github.io/crashes](http://stpierre.github.io/crashes)

## Collision locations



The majority of collisions involving a bicycle (399/756) happened in a crosswalk.

Large numbers of collisions also involve people riding in the street, and people riding in sidewalks. (Most sidewalk accidents involve cars turning into private drives, but a few featured cars jumping the curb.) A smattering involves people riding elsewhere, which includes parking lots, or crossing streets in the middle of blocks (i.e., away from crosswalks). Collisions in bike lanes are very rare indeed, but that likely reflects the rarity of bike lanes more than anything.

Clearly, crosswalks are far and away the most dangerous place for cyclists. No matter what other riding patterns are, cyclists almost certainly spend orders of magnitude more time out of crosswalks, so more than half the accidents happening in crosswalks is pretty clearly damning.

# What types of datasets exist?



image via: Nathan Zencey, Sunlight Foundation

<https://sunlightfoundation.com/2017/09/11/whos-at-the-popular-table-our-analysis-found-which-open-data-the-public-likes/>

# US Open Data Census

NUMBER OF PLACES **129**

NUMBER OF DATASETS **279**

NUMBER OF OPEN DATASETS **155**

PERCENTAGE OPEN **55%**

		Crime	Parcels	Budget	Construction Permits	Zoning (GIS)	Service Requests (311)	Transit	Business Listings	Restaurant Inspections	Web Analytics	Spending	Public Buildings	Procurement Contracts	Lobbyist Activity	Campaign Finance Contributions	Code Enforcement Violations	Property Assessment	Asset Disclosure	Property Deeds	Total Score
1	<a href="#">Las Vegas, NV</a>	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	1790	
2	<a href="#">Austin, TX</a>	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	1700	
3	<a href="#">Los Angeles, CA</a>	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	1670	
4	<a href="#">New York City, NY</a>	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	1650	
5	<a href="#">Santa Monica, CA</a>	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	2016	None	2016	2016	None	2016	2016	2016	1460	

- Asset Disclosure
- Budget
- Business Listings
- Campaign Finance Contributions
- Code Enforcement Violations
- Construction Permits
- Crime
- Lobbyist Activity
- Parcels
- Procurement Contracts
- Property Assessment
- Property Deeds
- Public Buildings
- Restaurant Inspections
- Service Requests (311)
- Spending
- Transit
- Zoning (GIS)

More information here: [us-city.census.okfn.org/faq/](http://us-city.census.okfn.org/faq/)

## Topics of the most popular open datasets

- 1  **Police and crime:** Police incidents; jail bookings; police station locations; crime statistics
- 2  **Transportation:** Taxi licenses; transit data; traffic counts; road infrastructure data; parking data
- 3  **Emergency calls:** Police, fire, and EMS responses; 911 calls; response times; incident reports
- 4  **Development:** Commercial developments; housing developments; property data; housing affordability
- 5  **Building safety:** Building permits; safety permits; certificates of occupancy
- 6  **Finance:** Revenue; spending; employee salaries; capital budgets; payments
- 7  **Elections:** Election results; polling locations; campaign finance reports
- 8  **Businesses and licenses:** Business licenses; liquor licenses; vendor, contract, and procurement data
- 9  **Inspections and service requests:** Restaurant health inspections; 311 requests; code violations
- 10  **Education:** Schools information; student health data; after-school programs; library locations

image/data via: Nathan Zencey,  
Sunlight Foundation

<https://sunlightfoundation.com/2017/09/11/whos-at-the-popular-table-our-analysis-found-which-open-data-the-public-likes/>

 Search

97 Results

Sort by  ▾**Authority** ▾

Official

Community

**Categories** ▾

Economy

Education

Environment

## Park Cemetery Alphabetical Log

[History](#)

Dataset

City of Marquette, Park Cemetery

Tags [marquette](#), [park cemetery](#), [michigan](#)[API Docs](#)Updated  
September 13, 2017Views  
4,241

## Shiawassee River



COMMUNITY

Environment

Tags [shiawassee river michigan](#)Updated  
September 13, 2017Views  
1,376**data.michigan.gov & michigan.gov/openmichigan**



### National Wetlands Inventory (from Michigan Department of Environmental Quality)

Shared by [michigan\\_admin](#)

This data set represents the extent, approximate location and type of wetlands and deepwater habitats in the conterminous United States. These data delineate the areal extent of wetlands and surface waters as defined by Cowardin et al. (1979). Certain wetland habitats are excluded from the Nat...



### Land Cover Circa 1800 (from Michigan Open Data)

Shared by [michigan\\_admin](#)

Landuse circa 1800 is a statewide database for Michigan based on original surveyors tree data and descriptions of the vegetation and land between 1816 and 1856. This is a fully attributed version that contains all vegetation codes (see attribute information). The database creators recognize that the...



### Potential Wetland Restoration (from Michigan Department of Environmental Quality)

Shared by [michigan\\_admin](#)

[Michigan Potential Wetland Restoration](#)[More Metadata](#)

8 attributes | 670117 locations |



### MDOT Cargo Ports (from MDOT Open Data)

Shared by [MDOT\\_GIS](#)

MDOT's Cargo Ports layer includes location and in/outbound tonnage for each cargo port. The data is used for the state long range plan, freight planning, economic analysis studies, and for various requests from local agencies. Please see the MDOT Metadata Form for additional information....

17 attributes | 40 locations |



### MDOT Carpool Lots (from MDOT Open Data)

Shared by [MDOT\\_GIS](#)

The Michigan Carpool Parking Lot Program (sometimes referred to as Park and Ride) began as a pilot program in 1974 with 11 carpool parking lots. Today there are over 240 carpool parking lots located across the state providing nearly 9,000 parking spaces. In addition to lots owned and main...

22 attributes | 259 locations |



### Metropolitan Planning Organizations (MPO) (from MDOT Open Data)

Shared by [MDOT\\_GIS](#)

Metropolitan Planning Organizations (MPO) are boundaries that define specific urbanized areas along with adjacent municipalities that are accepted into the MPO. MPOs serve as localized areas for transportation planning and forecasting traffic demand models. Please see the MDOT Metadata Form fo...

13 attributes | 14 locations |

# Who uses civic data and how do they use it?



## New York City: Targeting Illegal Building Conversions Inspections

[More](#)

In New York City, dangerous illegal building conversions are a big problem and public safety risk. The government opened and integrated data that had previously been siloed within by several departments and analyzed it to increase accurate inspection targeting and better allocate scarce resources where they will do the most good.



## Asheville, North Carolina: Empowering Startups

[More](#)

CIO Jonathan Feldman started “Open Data Day” to bring together citizens journalists, business people, elected officials and others together making opening data a collective issue. Led in partnership with Venture Asheville, the event gave Asheville a chance to highlight business opportunities that arise from publishing open data. One local employer with a staff of 35 uses publicly-available data from municipalities, adds their analysis, and sells it to insurance companies.



## Louisville: Restaurant Inspection Scores on Yelp

Louisville Metro Government published restaurant inspection data in an open, standardized format set by a cohort of other cities, third party organizations, and industry. Yelp was then able to consume the data easily and publish it alongside its popular online restaurant reviews, making the data more useful and accessible to the average citizen.

Big Data Visualization and Society - Riyadh

## STUDENT PROJECTS



Based on the theoretical constructs outlined by the project, the context of Riyadh, and the technical skills learned throughout the course, while simultaneously taking advantage and building on top of the work that has been generated through the MIT-KACST collaboration over the past few years, the students were asked to identify and tackle a policy issue, explore it and develop media to facilitate its representation and a broader conversation around such issues.

The projects primarily dealt with issues surrounding the transportation networks of Riyadh, and were developed over the course of the spring semester, first through research of the policy issues in the region and the development of technical skills, and then through an 8 week-long final project that put the different skills together into an interactive online platform. Students formed 5 teams of 2, and one team of 3, and went through an iterative design process, with weekly design critics, and extensive technical support outside of the classroom. The course culminated with a public review of the tools and materials developed.



# Mapping Prejudice

Events

Contact

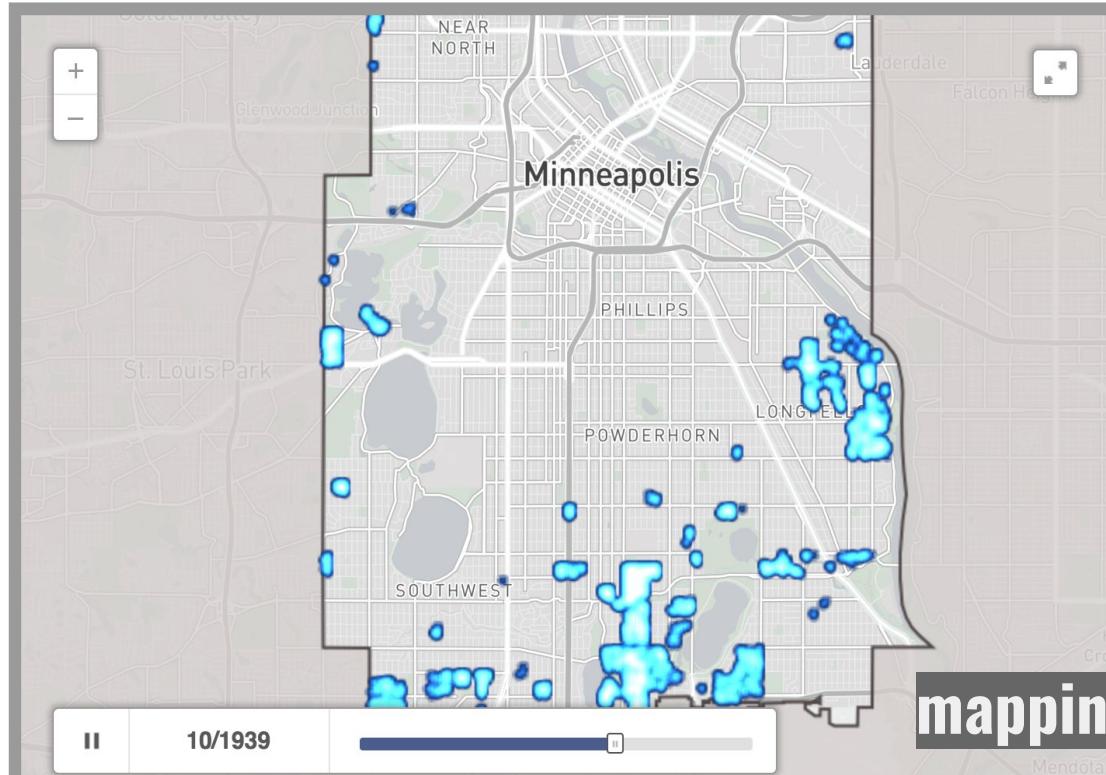
Historyapolis

What Are  
Covenants?

About Us

Get Involved

Media



[mappingprejudice.org](http://mappingprejudice.org)

# About Data for Black Lives

Data for Black Lives is a group of activists, organizers, and mathematicians committed to the mission of **using data science to create concrete and measurable change in the lives of Black people.**

Since the advent of computing, big data and algorithms have penetrated virtually every aspect of our social and economic lives. These new data systems have tremendous potential to empower communities of color. Tools like statistical modeling, data visualization, and crowd-sourcing, in the right hands, are powerful instruments for fighting bias, building progressive movements, and promoting civic engagement.

But history tells a different story, one in which data is too often wielded as an instrument of oppression, reinforcing inequality and perpetuating injustice. Redlining was a data-driven enterprise that resulted in the systematic exclusion of Black communities from key financial services. More recent trends like predictive policing, risk-based sentencing, and predatory lending are troubling variations on the same theme. Today, discrimination is a high-tech enterprise.

# Challenges and Threats

image: [Flickr user Blude](#)

# Challenges and Threats - Lack of Organization

1. Information you want or need is not included
2. Public data is often held in legacy formats and systems
  - Not machine readable or not ready for use
  - Available for visualization but not download
  - Access can be hampered by low bandwidth, crashes, corrupted files, etc.
3. Not well documented/uses confusing codes
4. Data isn't regularly updated, or formats & schemas change from period to period
5. Data does not conform to regular formats and styles and naming conventions

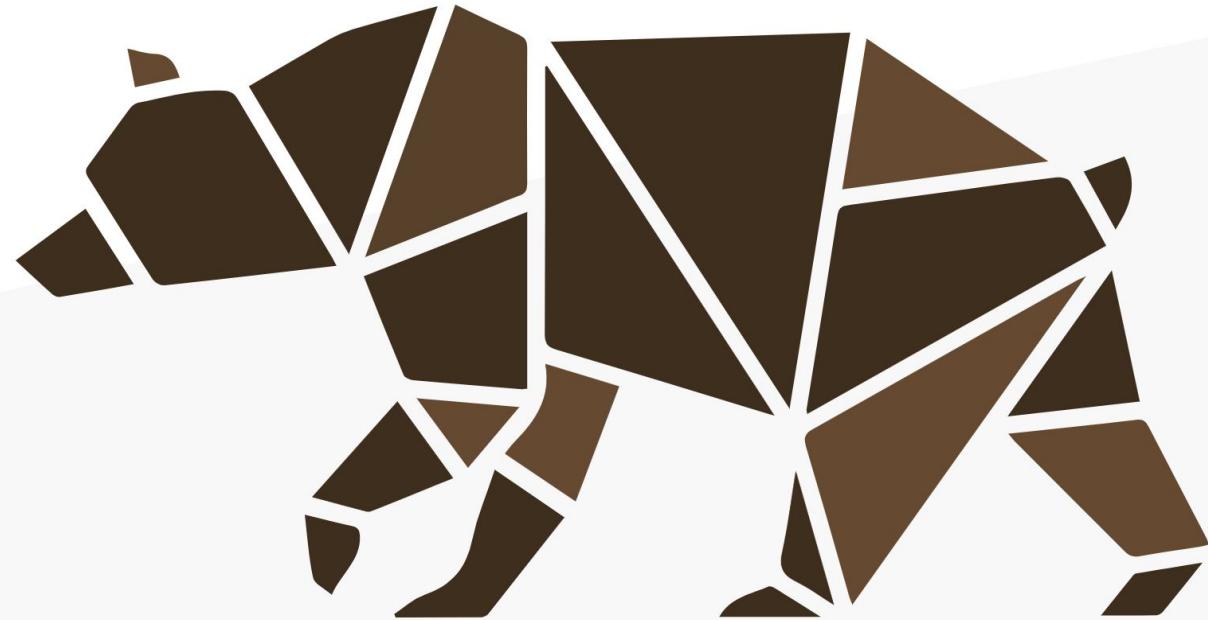


# Download California campaign finance data with ease

An open-source archive serving up daily downloads from CAL-ACCESS, California's database tracking money in state politics.

[Learn more »](#)

[Download latest data](#)





# Frequently asked questions

Questions and answers about this project and the underlying the CAL-ACCESS database.

## Why is all this necessary?

California's jumbled, dirty and difficult campaign-finance database, known as CAL-ACCESS, sprawls across 80 database tables and weighs in at more than 650 megabytes.

A significant effort is required to understand its esoteric structure and prepare the records for meaningful analysis.

That barrier blocks anyone seeking to interpret the information. The challenge of untangling the database requires weeks of study and significant guesswork, discouraging most analysts from daring—and raising the risk that those who do will make a critical error.

That is the problem we aim to solve by leading an open-source effort to perfect the numerous transformations, filters and computer operations necessary to refine the raw data into an easy-to-use product.

**Why is all this  
necessary?**

---

**What is the California  
Civic Data Coalition?**

**How far back does  
the CAL-ACCESS  
database go?**

**Do you offer all  
tables in the CAL-  
ACCESS database?**

**Do you offer all of the  
available data?**

**Do you modify the  
source data?**



# Download California



# Ingham County Clerk

Ingham County Courthouse • 341 S. Jefferson • P.O. Box 179  
Mason, MI 48854 • 517.676.7201



[DOCUMENT SEARCH](#) | [HELP](#) | [ABOUT](#) | [LOGOUT public](#)

[View Image](#)

**SUMMARY PAGE  
CANDIDATE COMMITTEE**

RECEIPTS	Column I This Period	Column II Cumulative this election cycle
3. Contributions		
a. Itemized (Schedule 1A - Column 6)	(3a.) \$ <u>34,101.97</u>	
b. Unitemized (less than \$20.01 each - no Schedule)	(3b.) \$ <u>NOT APPLICABLE</u>	
c. Subtotal of "Contributions"	(3c.) \$ <u>34,101.97</u>	
4. Other Receipts (Schedule 1A -1, Column 6)	(4.) \$ <u>9,500.00</u>	
5. TOTAL CONTRIBUTIONS AND OTHER RECEIPTS (Add Line 3c + Line 4)	(5.) \$ <u>43,601.97</u>	(18.) \$ <u>34,876.97</u>
<b>IN-KIND CONTRIBUTIONS &amp; EXPENDITURES</b>		(19.) \$ <u>9,500.00</u>
6. In-Kind Contributions (Schedule 1-IK, Column 7)	(6.) \$ <u>0</u>	(20.) \$ <u>44,376.97</u>
7. In-Kind Expenditures (Schedule 1B-IK, Column 6)	(7.) \$ <u>0</u>	(21.) \$ <u>0</u>
<b>EXPENDITURES</b>		(22.) \$ <u>0</u>
8. Expenditures		
a. Itemized (Schedule 1B, Column 6)	(8a.) \$ <u>39,579.97</u>	

# Challenges and Threats - Laws & politicians

1. FOIA immunity
2. Laws reducing the information collected or made available
  - H.R. 1305 - American Community Survey Act Would likely undercount minority populations in federal databases
  - Senate bill S.103 and House bill H.R.482 would prohibit funding from being used "to design, build, maintain, utilize, or provide access to a Federal database of geospatial information on community racial disparities or disparities in access to affordable housing."
3. Reduced funding/support for infrastructure
4. Changes to Title 44
  - See [diglib.org/archives/14551/](http://diglib.org/archives/14551/)