OPEN DATA, SIMPLY SEEN

@abhinemani | CAPP 30130 | University of Chicago

CIVIC TECH: KEYS TO SUCCESS

Empathy	*Lean*	*Data*	*Users*	*Sustainability*
➤ Build with, not for	➤ It's more than just a technical process	➤ Data "lives" everywhere	➤ ~~If you build it they will come~~	➤ Consider your options: startup, open source, non-profit, academic, etc
➤ Talk to users	➤ Expect failure	➤ Be ready to scrape	➤ Listen and learn	
➤ Understand behavior	➤ Only build what's necessary	➤ Open data is your friend	➤ Define metrics	➤ Tech -> Policy
➤ Learn what's out there	➤ Test, test, test	➤ ETLs are your best friend	➤ Find partners	➤ Expect to train everyone
				➤ Document everything

CIVIC TECH: KEYS TO SUCCESS

Empathy	*Lean*	*Data*	*Users*	*Sustainability*
➤ Build with, not for	➤ It's more than just a technical process	➤ Data "lives" everywhere	➤ ~~If you build it they will come~~	➤ Consider your options: startup, open source, non-profit, academic, etc
➤ Talk to users	➤ Expect failure	➤ Be ready to scrape	➤ Listen and learn	
➤ Understand behavior	➤ Only build what's necessary	➤ Open data is your friend	➤ Define metrics	➤ Tech -> Policy
➤ Learn what's out there	➤ Test, test, test	➤ ETLs are your best friend	➤ Find partners	➤ Expect to train everyone
				➤ Document everything

AGENDA

➤ Case Studies:

    ➤ Immigration in LA

    ➤ Toys for Tots

➤ Scraping

➤ Mapping

    ➤ Socrata

    ➤ Google Fusion Tables

    ➤ Mapbox

IMMIGRATION IN LA

Mapping through a problem with open data

4

Home   News   I-Team   Weather   Entertainment   •••

LIVE TV   62°   Connect

Local   Watch Live Newscast   U.S. & World   California News   Sports   Car Chases   Business   In the Weeds   Health   Tech   Weird   Weather   NewsConference

After DACA Announcement, LA Lawyer Warns of Immigration Scams

So-called experts have sought to take advantage of changes in immigration laws

By Gerardo Bonilla

WEATHER FORECAST

WEATHER ALERTS   View all

Los Angeles, CA **Change** ⌄

62° Scattered Clouds
Feels Like 62°

Radar   Forecast   Maps

IMMIGRATION POLICY / EDUCATION: CASE STUDY

➤ Policy: Changing federal rules around immigration policy and enforcement (2015)

➤ Problem: Fraud and misinformation taking advantage of residents and discouraging participation

➤ Approach

  ➤ Program: City partners with accredited institutions to host official educational workshops

  ➤ User needs

    ➤ "Which workshops can I trust?"

    ➤ "How do I find ones close to me?"

  ➤ Intervention

    ➤ Data capture from partnering organizations (simple excel file)

    ➤ Post data to open data portal

    ➤ Create simple, mobile-friendly web app/map for residents to find nearby workshops

WORKSHOPS POSTED ON OPEN DATA

MYIMMIGRATIONLA (DEPRECATED)

CURRENT IMMIGRATION WORKSHOP VISUALIZATION (SOCRATA)
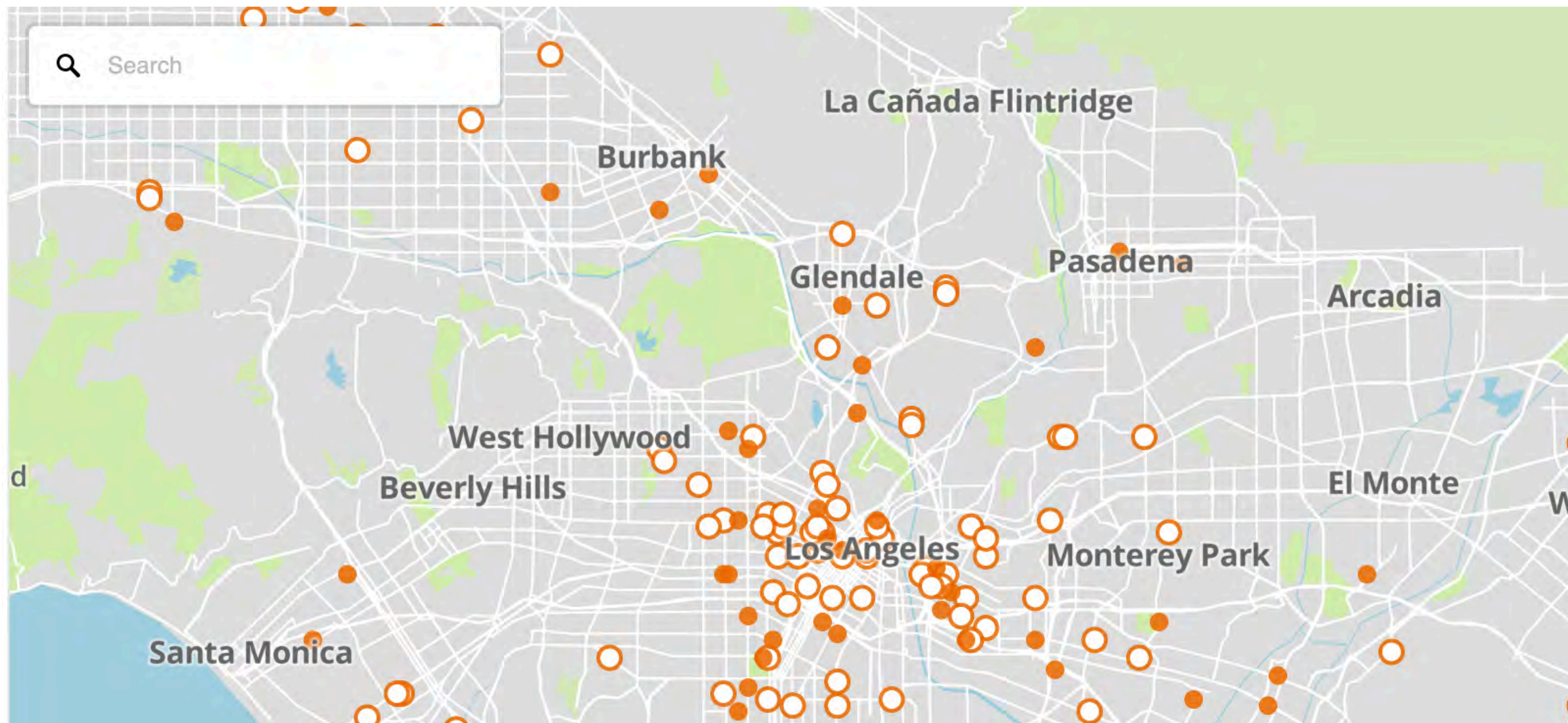
PRO-TIP:
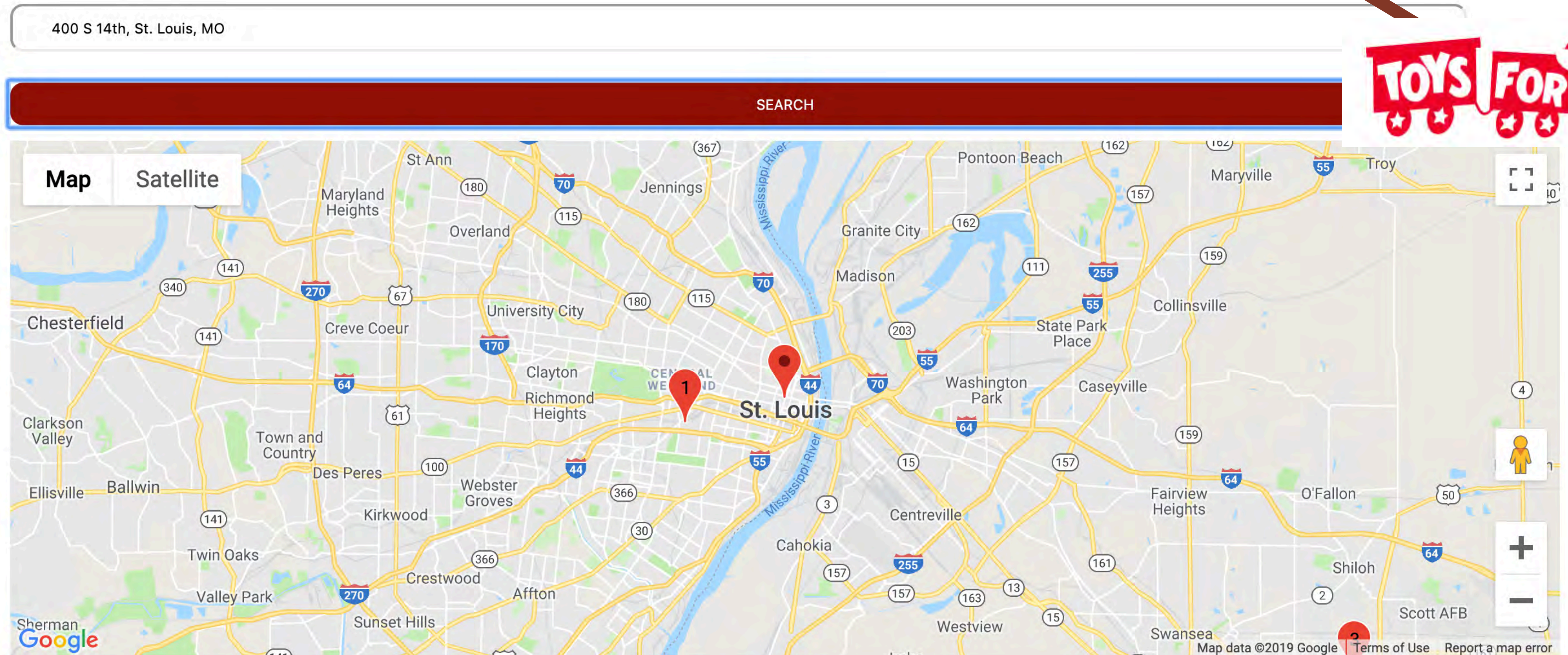# PUT DATA ON A (USEFUL) MAP.

TOYS FOR TOTS

Sometimes you have to scrape by…

TOYS FOR TOTS

NOTE: Locations on the map DO NOT represent the physical location of a campaign. DO NOT appear at the location displayed on the map. Campaigns will notify applicants of final distribution locations if your application can be fulfilled.

400 S 14th, St. Louis, MO

SEARCH

TOYS TO TOTS: CASE STUDY

➤ Problem: Difficult to find where to donate toys during the holiday season

➤ Approach

  ➤ User needs

    ➤ This website isn't helpful. (Sadly.)

    ➤ "How do I find donation locations close to me?"

  ➤ Intervention

    ➤ Data scraping from Toys for Tots webpage

    ➤ Aggregate and geocode the data with Google Fusion Tables

    ➤ Use open source, GitHub-hosted, mobile-friendly search and map app

GETTING THE DATA

➤ "Data" typically shared on flyers or in-person

➤ Main website did not have a CSV, API, or any machine-readable format

➤ Data had to be scraped from main website

  ➤ Data was posted as an html table

  ➤ Developer used PHP Scraper Gouette

  ➤ More information

DATA HOSTED ON GOOGLE FUSION TABLES

DONATEIT.TO BY JOHN KRAMLICH
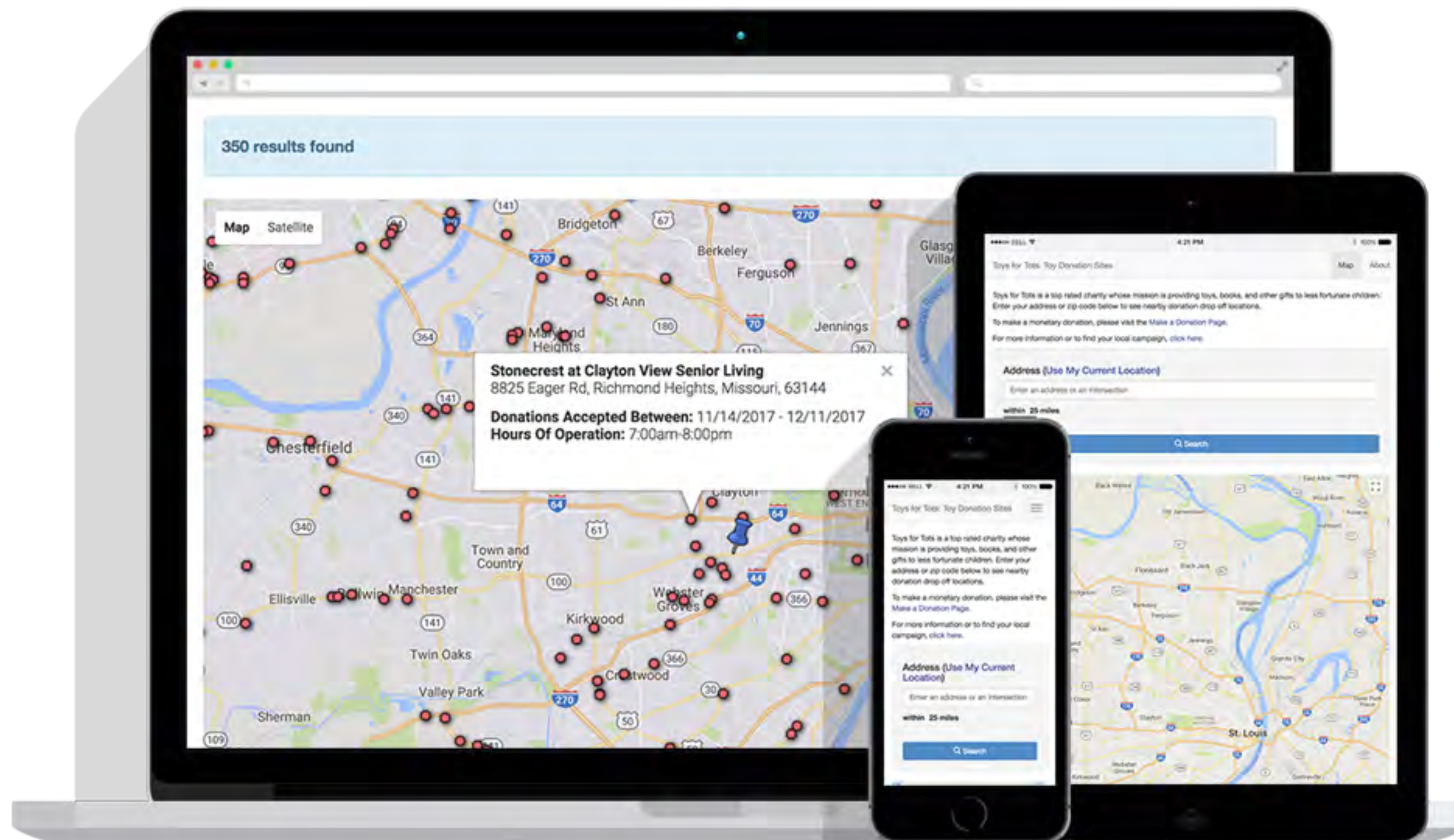
GOOD CIVIC TECH IS SOMETIMES SHORT-LIVED.

A BIT OF HISTORY

Data capture

FOIA & Scrapers

Personal Databases

CSV, MySQL, etc

Data validation

Geocoding, cleanup, etc

Source Data (base)

Webpages, PDFs
Unstructured data

HOW IT WAS
(AND STILL IS, A LITTLE)

Data "Product"

Apps, maps, etc

@abhinemani

SCRAPING

Getting something you can work with…

INTERACTIVE BEAUTIFULSOUP TUTORIAL

Link

BEAUTIFULSOUP EXAMPLE WITH WORKING CODE

Link

IN-BROWSER SCRAPING: GREPSER.COM

2019 Festival & Event Dates

Search Events

DCASE Homepage > Festival and Event Dates

Department:

Cultural Affairs and Special Events

People We Serve:

Businesses & Professionals

(event highlights only, subject to change)

2019

African American Designers in Chicago: Art, Commerce and the Politics of Race

Through March 3, 2019
Chicago Cultural Center, Chicago Rooms

www.chicagoculturalcenter.org

Featuring work from a wide range of practices includin architectural signage, illustration, graphic design, exhib this exhibition is the first to demonstrate how African A image of the black consumer and the work of the black American advertising/consumer culture. *African Ameri Commerce and the Politics of Race* is funded in part by American Art and The Richard H. Driehaus Foundation, an exploration of Chicago's art and design legacy.

Chicago Cultural Center Music, Theater & Dance, Exhibitions and More

Year-round

www.chicagoculturalcenter.org

Discover the free visual art and performance programs Center by the Chicago Department of Cultural Affairs a partners.

grepsr   How It Works   Services   Customers   Pricing   Blog   Contact Us

Web Scraping Service Platform That's Effortless

We get it. You're tired of learning and configuring complicated tools. Plus, it's taking way more time to structure and make data useable.

Grepsr m
capture, r

Get S

chrome web store

abhi.nemani@gmail.com ⌄

Home > Extensions > Grepsr - Web Scraping Tool

Grepsr - Web Scraping Tool

Offered by: https://www.grepsr.com

★★★★☆ 46 | Productivity | 👤 12,268 users

Add to Chrome

FROM WEBPAGES TO STRUCTURED DATA

FROM WEBPAGES TO SPREADSHEETS (AND SCHEDULED UPDATES)

| | Data Preview | API | JSON | RSS Feed | Schedule Crawl | Data Delivery | Download | Configure & Run |

Showing 1 to 25 of 25 records   April 30th 2019 05:29 pm

title	Date	URL	description
African American Designers in Chicago: Art,	Through March 3, 2019 Chicago Cultural Center, Chicago Rooms	http://www.chicagoculturalcenter.org	Featuring work from a wide range of practices including cartooning, sign painting, architectural signage, illustration, graphic design, exhibit design and product design, this exhibition is the first to demonstrate how African American designers remade the image of the black consumer and the work of the black artist in this major hub of American advertising/consumer culture. African American Designers in Chicago: Art,
Chicago Cultural Center Music, Theater &	Year-round	http://www.chicagoculturalcenter.org	Discover the free visual art and performance programs presented at the Cultural Center by the Chicago Depar
Maxwell Street Market	Sundays, Year Round, 7am–3pm 800 S. Desplaines St.	http://www.maxwellstreetmarket.us	
Under the Picasso	Year-round Daley Center, 50 W. Washington St.	http://www.underthepicasso.us	The D weekd
Juicebox	1st & 3rd Fridays (Chicago Cultural Center) 1st & 3rd Saturdays (Garfield	http://www.chicagoculturalcenter.org	With e geare
Year of Chicago Theater	January-December 2019 Citywide		Mayor Theat theatr

Chicago Events | chicago.gov   Edit In Chrom

| REPORT | SUPPORT | ASSETS | ACTIVITY | TEAM | SETTINGS |

‹   April 2019   ›

SUN	MON	TUE	WED	THU	FRI	SAT
28	29	30				

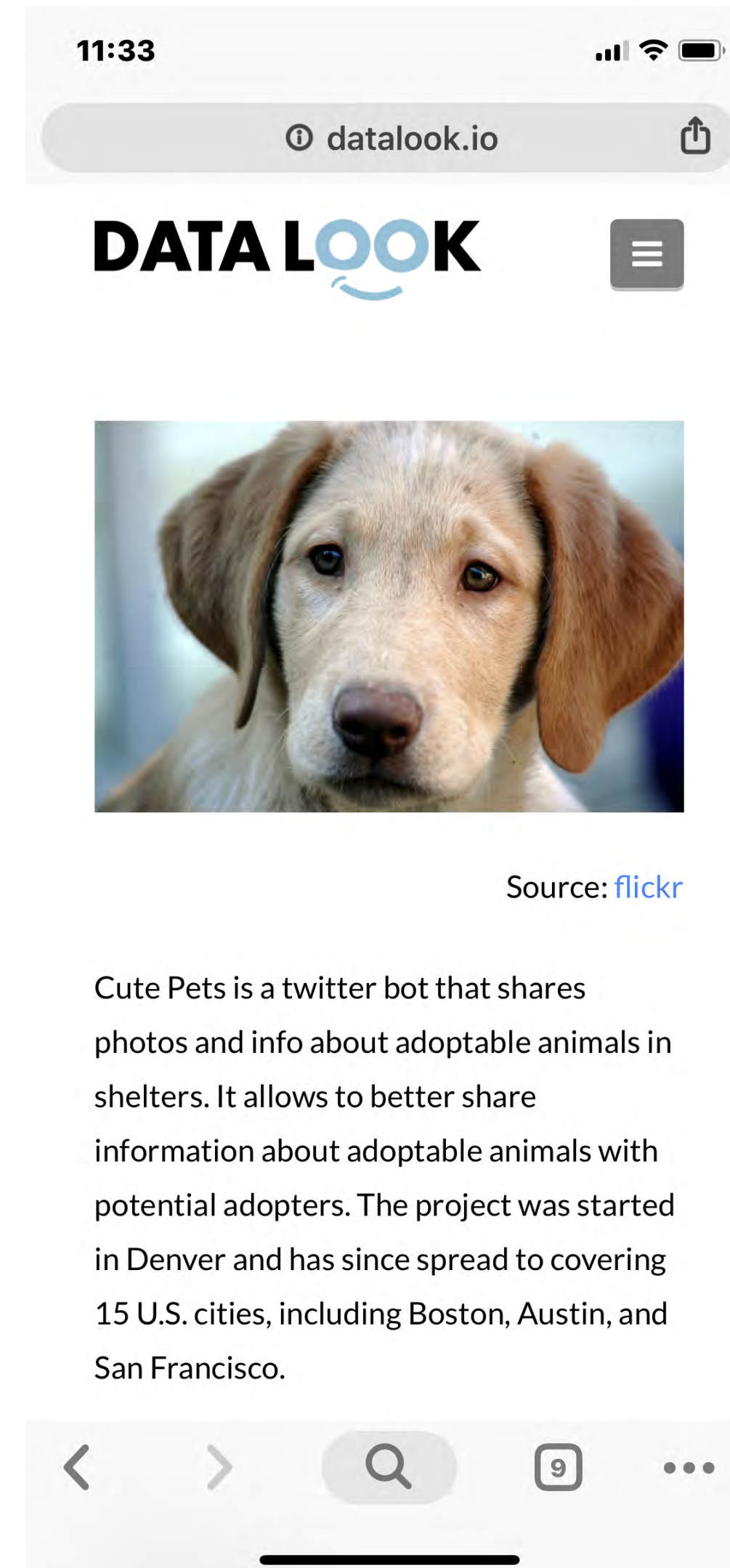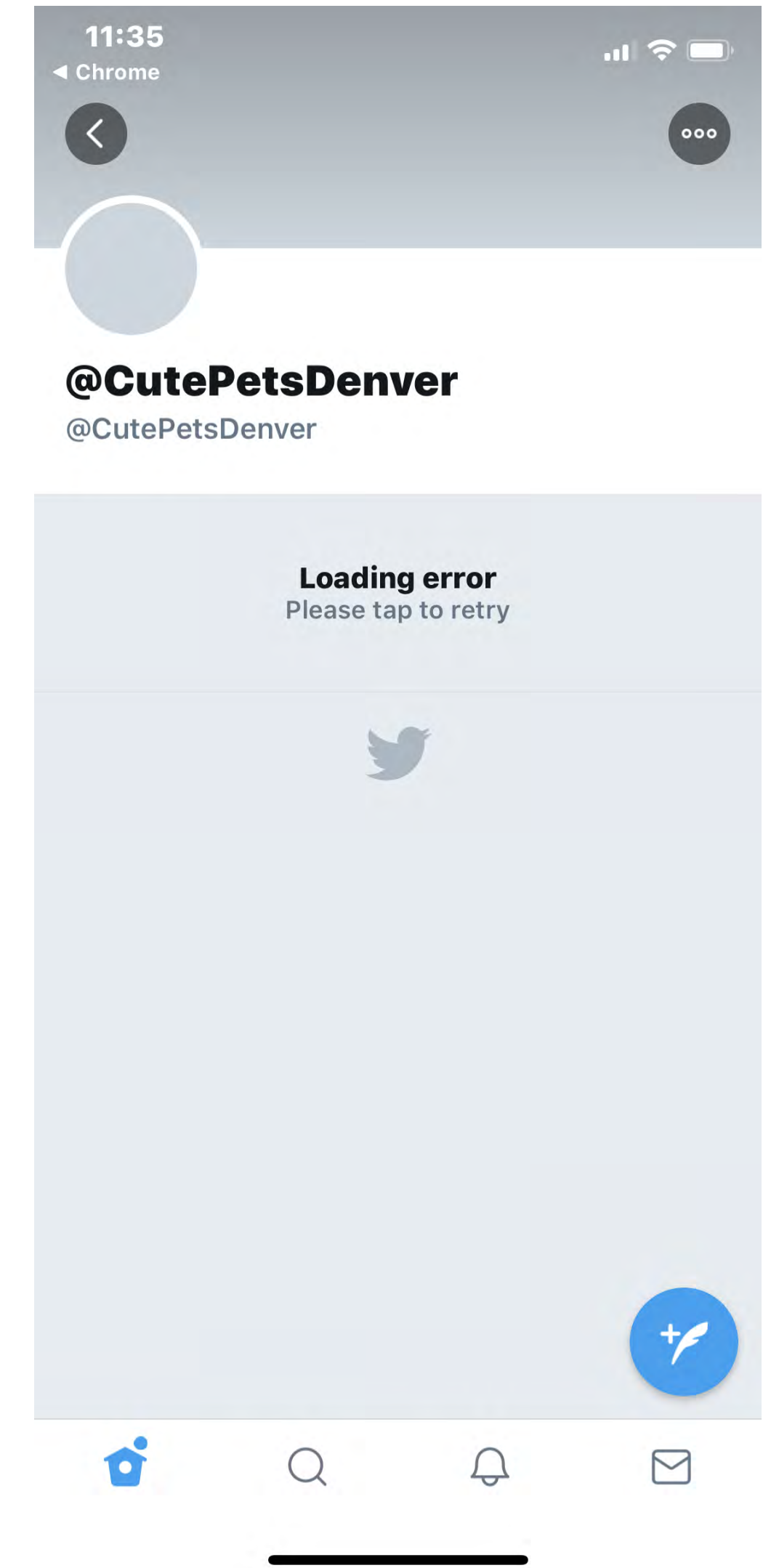| Data Preview | API | JSON | RSS Feed | Schedule Crawl | Data Delivery | Download | Configure & Run |

Schedule your web scraping on-demand!

Want to run or re-schedule your scrape on a specific time of a day or date? No problem. Click on the "Add Schedule" button and simply choose the data and time form the calendar and then save it. Yep, it's that simple!

Okay, got it

CUTE PETS: THE OPPORTUNITY AND PROBLEM WITH SCRAPING

Denver Pets
@CutePetsDenver

They call me Tabitha. I am a spayed female white bichon frise/mix.
PetHarbor.com/detail.asp?ID=…

7:11 AM - 18 Sep 2015

*A ream built a twitter bot powered by a web scraper to tweet out pictures of adoptable pets. **It was redeployed a dozen times.** But the webpage changed and the code wasn't maintained so the scraper and the bot broke.*

11:33
datalook.io

DATA LOOK

Source: flickr

Cute Pets is a twitter bot that shares photos and info about adoptable animals in shelters. It allows to better share information about adoptable animals with potential adopters. The project was started in Denver and has since spread to covering 15 U.S. cities, including Boston, Austin, and San Francisco.

11:33
govtech.com

gt

DATA

What Open Data is Doing for Denver Dogs

A new Twitter bot gathers up doggie data from the Denver Animal Shelter's website and works to find potential adopters.

BY JASON SHUEH / MAY 8, 2014

STINKDOG PHOTOS

11:35
Chrome

@CutePetsDenver
@CutePetsDenver

Loading error
Please tap to retry

Data
Centralization

Data
Consumption

Datastore

Source Data (base)

MODERN
OPEN DATA WORKFLOW:
MAKES SCRAPING LESS ESSENTIAL
BUT STILL SOMETIMES NECESSARY

Data "Product"

@abhinemani

GEOCODING & WEB MAP MAKING

Addresses are just the start

GEOCODING

....
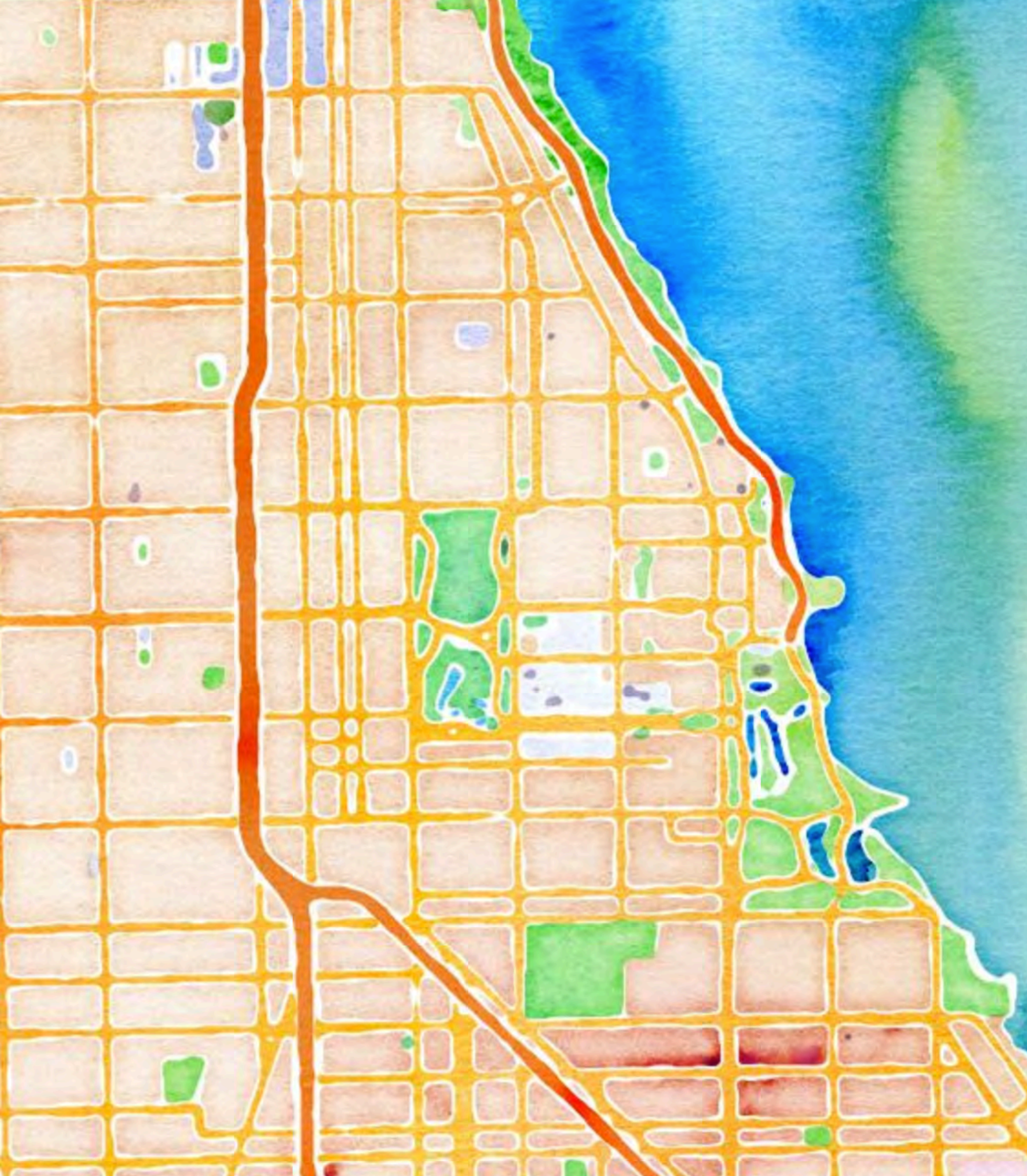
➤ Geocoding is the process of going from a street address to a "coordinate plane"

   ➤ Typically (and ideally) latitude & longitude (ie lat/long)

➤ Web mapping services require lat/long

   ➤ Addresses can be ambiguous

      ➤ Same streets in many cities

➤ Geocoding services can be expensive

   ➤ Popular geocoding services: Google Maps, CARTO, Esri, Mapquest

   ➤ Some governments/companies have their own geocoding services

THREE (FREE) MAPPING OPTIONS

➤ Socrata

  ➤ Government must have Socrata

  ➤ Government must have set up geocoding/geo-referencing

➤ Google Maps / Fusion Tables

  ➤ Fusion Tables requires little-to-no coding, but will be deprecated EOY

  ➤ Google Maps costs $ and requires database/coding

➤ Mapbox

  ➤ Data must have lat/long already

  ➤ Requires some coding

    ➤ Javascript

MAP MAKING WITH SOCRATA

1) *Select dataset*

2) *"Visualize"*

3) *Select Map*

4) *Save draft (requires account)*

5) *Click "share" icon to embed or share link*

MAP MAKING WITH FUSION TABLES

1. Download excel-file / CSV with some address data

2. Open fusion tables and upload data file

3. Identify which column has location data

4. Create map view (and wait for geocoding)

   1. Provide disambiguation help (eg State, County, etc) if possible

5. Customize tooltip and icons

6. Publish

MAP MAKING WITH FUSION TABLES

MAP/APP MAKING WITH FUSION TABLES: EDER'S SEARCHABLE MAP

Searchable Map Template

Map | About

This is a demo of Derek Eder's Searchable Map Template using Google Fusion Tables. Use this section to give a little context to your map.

Address (find me)

Enter an address or an intersection

within 1/2 mile

Custom filters and views

Take a look at the wiki to see how to add your own custom filters and views like:

- Checkboxes
- Radio buttons
- Drop down lists
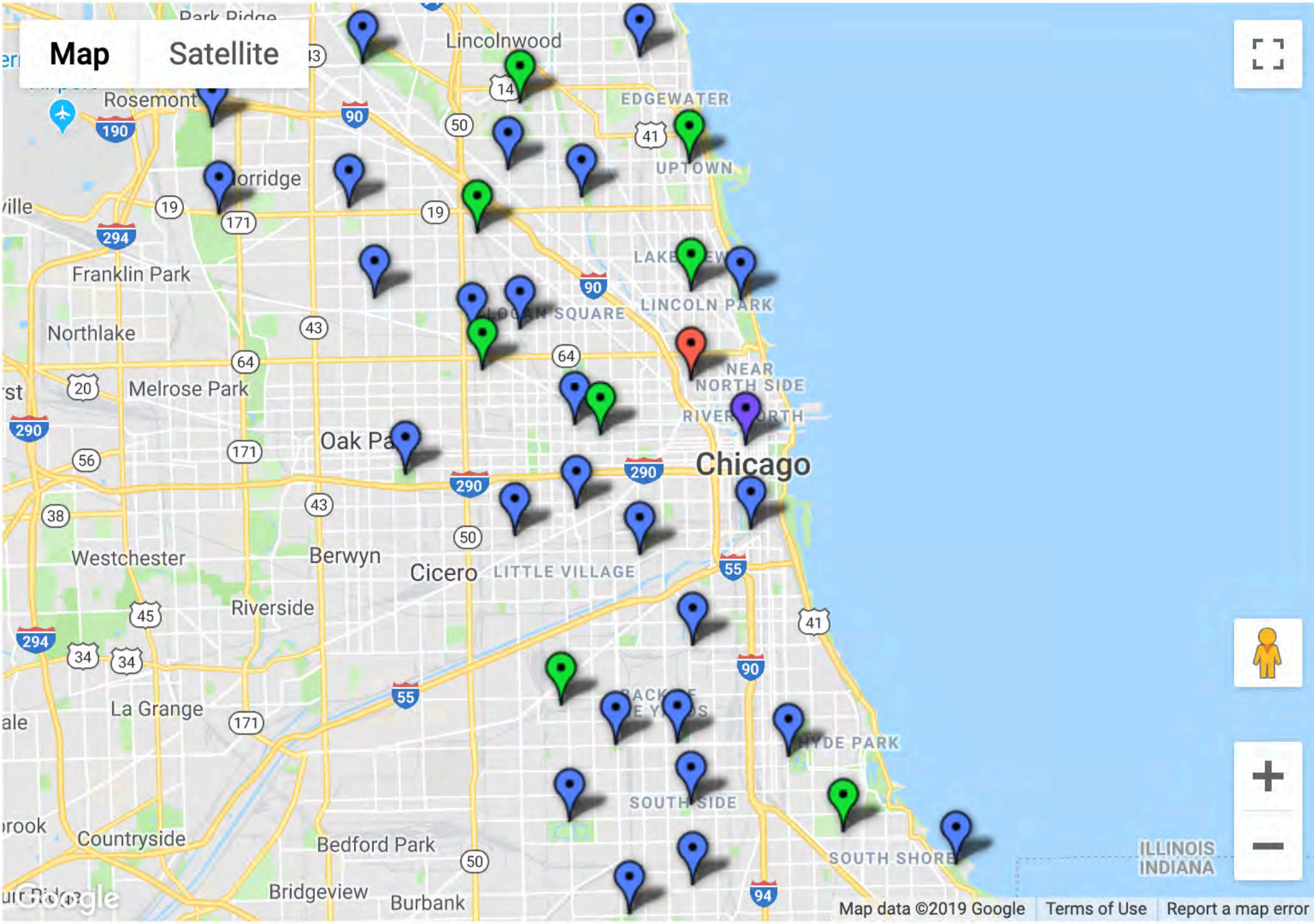- Text searches
- Sliders
- Date sliders
- Results lists

Q Search   C Reset

41 results found

derekeder.github.io/FusionTable-Map-Template/index.html

Map | Satellite

Chicago

Searchable Map Template by Derek Eder.

Map data ©2019 Google | Terms of Use | Report a map error

Features

- full screen, iframe and content templates
- display up to 100,000 map points
- address search (with variable radius and geocomplete)
- geolocation (find me!)
- results count
- RESTful URLs for sharing searches
- ability to easily add additional search filters (checkboxes, sliders, etc)
- mobile and tablet friendly using responsive design
- built with HTML, CSS and Javascript - no server side code required
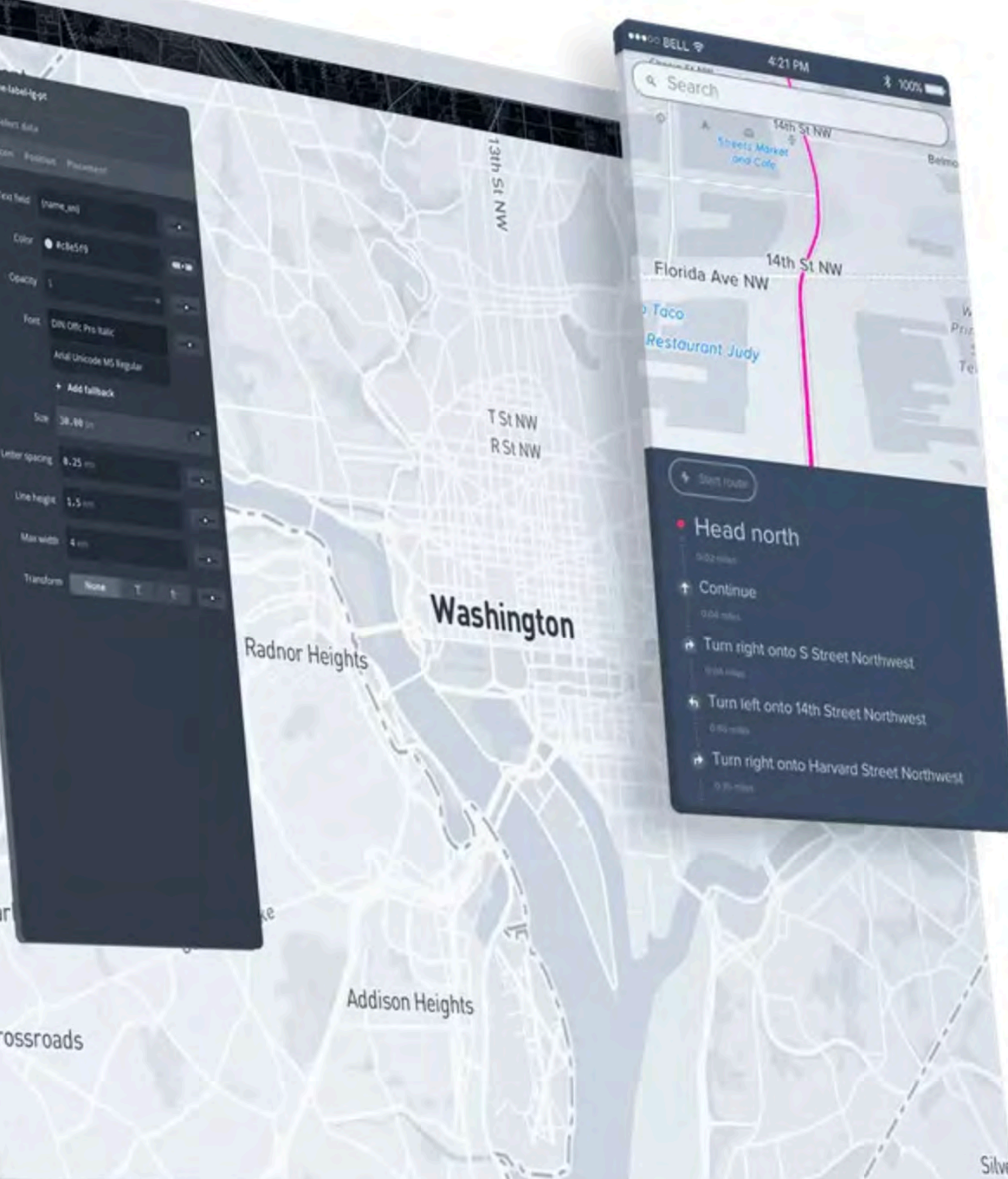
To embed, you can use the following code on your page:

```
<iframe
    style="border-style: none;"
    src="/path/to/map-template/index_iframe.html"
    width="600"
    height="950" >
</iframe>
```

Instructions Link: https://github.com/derekeder/FusionTable-Map-Template

MAPBOX FOR MAPPING

....

➤ MapBox is built on open source code and open data

➤ Free for testing / development

➤ Built initially for map styling; now serves as (mostly) full-service map web service

➤ Three kinds of core elements

  ➤ **Datasets**: Geo-coded datafiles

      ➤ GeoJSON, JSON, CSV, KML, etc

  ➤ **Tilesets**: Vector-files with base-layers and custom datasets

  ➤ **Styles**: Designed / styled maps for use in app-development

MAPBOX: SEARCHABLE / EMBEDDABLE MAP

Instructions Link: https://docs.mapbox.com/help/tutorials/tilequery-healthy-food-finder/

LECTURE 10 ASSIGNMENT (ASSIGNED 5/1)

➤ Writing Assignment / Midterm (more on next slide)

➤ Memo to the New Mayor: 1 page, single-spaced; Due 5/13

➤ Project Work

➤ Request GitHub Student Pack: https://education.github.com/pack

➤ *Begin* to follow instructions on creating interactive map
(next element of Civic Hacker Toolkit):
https://docs.mapbox.com/help/tutorials/tilequery-healthy-food-finder/

➤ Code samples for reference:

➤ mapbox-searchable-map: https://github.com/abhinemani/mapbox-searchable-map

➤ fusiontables-searchable-map: https://github.com/abhinemani/fusiontables-searchable-map

➤ JKAN Pages: fusiontables embed, mapbox-map, Mapbox-embed

WRITING ASSIGNMENT: MEMO TO THE NEW MAYOR

➤ By now, you should have read perspectives on civic data from dozens of industry leaders and practitioners, heard about how it has been implemented by the government and used by residents, and even worked with some yourself. To assess your comprehension of this material, you are to write a **1-page, single-spaced memo/briefing** for the new mayor of Chicago on the role you see technology playing in the city government.

➤ **This can focus *on any or all* of the topics covered to date: data-driven decision making and policy, lean principles, user needs (empathy), open data, and community engagement.** You should answer three questions: **1) what's working; 2) what's not; and 3) what (if anything) to focus on for the next 4 years?** You should include what you have learned about Chicago, as well as other cities/governments regarding **policies, programs, systems, and tools**. Your grade will be measured on grasp of the materials, as well as evidence of your own thinking/perspective.

➤ Due: Wednesday, 5/13, at the beginning of class