



Visualyze

Philippe Yu, Piyush Sharma, Mack Li



Dataset

- US Communities Crime Dataset 2018
- Interested in crime-related data and distribution
- Community data on population, age, income,
and crime committed



Solution

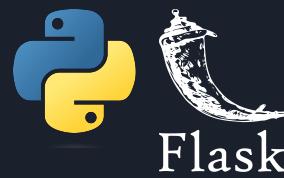
- Wanted to develop a **web application**
 - **Import Data** (extensible to all map data)
 - **Visualize** map-based data with popups
- Allow for **analysis** of data through visualization

Tech Stack

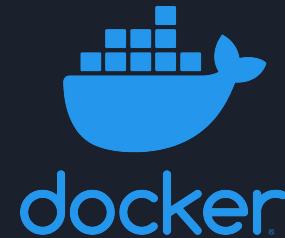
Frontend



Backend



Container





DEMO



Organization

The project is divided into three sprints

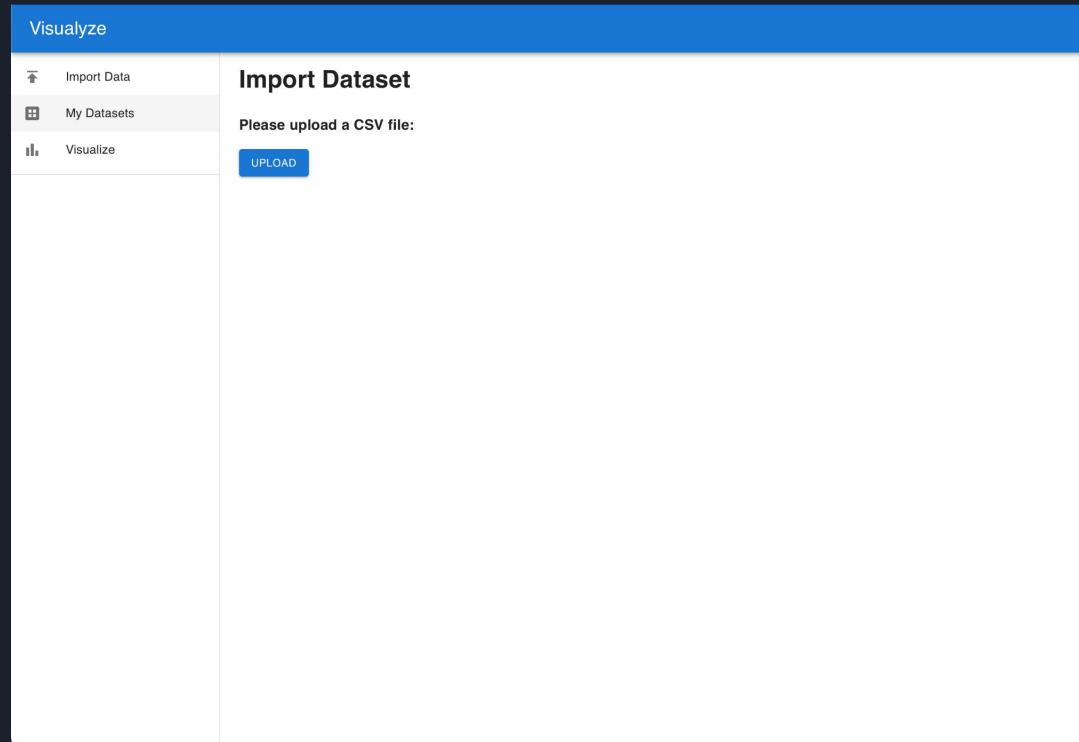
Sprint 1 (Oct 09)	Sprint 2 (Nov 03)	Sprint 3 (Dec 08)
<p>Deciding on ideas, framework</p> <ul style="list-style-type: none">• Frontend will be React• Backend will be Py + Flask• Containerization is Docker• Provisional database schema• Provided a test application	<p>Basic Flask, React, Github Actions</p> <ul style="list-style-type: none">• Map is open-source Leaflet• Layout decided to contain sidebar and main view• Three tabs<ul style="list-style-type: none">◦ Import◦ My Dataset◦ Visualization• Many features nonfunctional	<p>Functional application</p> <ul style="list-style-type: none">• Map now shows area and heatmap• Implementation of multi-selection• Non-functional components now made functional<ul style="list-style-type: none">◦ Lots of logging!◦ Bug tracking and subtle improvements!

Sprint 1

rowid	communityName	state	countyCode	communityCode	population	householdsize	racepctblack	racePctWhite	racePctAsian	racePctHispanic	agePct12t21	agePct12t29	agePct
1	Berkeley Heights	NJ	39	5320	11980	3.1	1.37	91.78	6.5	1.88	12.47	21.44	10.93
2	Marple	PA	45	47616	23123	2.82	0.8	95.57	3.44	0.85	11.01	21.3	10.48
3	Tigard	OR			29344	2.43	0.74	94.33	3.43	2.35	11.36	25.88	11.01
4	Gloversville	NY	35	29443	16656	2.4	1.7	97.35	0.5	0.7	12.55	25.2	12.19
5	Bemidji	MN	7	5068	11245	2.76	0.53	89.16	1.17	0.52	24.46	40.53	28.69
6	Springfield	MO			140494	2.45	2.51	95.65	0.9	0.95	18.09	32.89	20.04
7	Norwood	MA	21	50250	28700	2.6	1.6	96.57	1.47	1.1	11.17	27.41	12.76
8	Anderson	IN			59459	2.45	14.2	84.87	0.4	0.63	15.31	27.93	14.78
9	Fargo	ND	17	25700	74111	2.46	0.35	97.11	1.25	0.73	16.64	35.16	20.33
10	Waco	TX			103590	2.62	23.14	67.6	0.92	16.35	19.88	34.55	21.62
11	Sherman	TX			31601	2.54	12.63	83.22	0.77	4.39	15.73	28.57	15.16
12	San Pablo	CA			25158	2.89	21.34	49.42	17.21	26.78	13.65	28.82	13.23
13	Bowling Green	KY			40641	2.54	12.18	86.39	1.12	0.68	21.51	36.83	23.96
14	Pine Bluff	AR			57140	2.74	53.52	45.65	0.49	0.43	16.51	28.17	14.68
15	New Ulm	MN	15	46042	13132	2.53	0.06	99.21	0.47	0.59	14.0	25.03	12.83
16	Maplewood	MN	123	40382	30954	2.69	2.52	94.39	2.03	1.55	12.07	25.43	11.42
17	Enfield	CT	3	25990	45532	2.85	2.65	95.72	1.04	2.28	11.86	27.51	12.36
18	Glendale	CA			180038	2.62	1.3	74.02	14.14	20.96	12.04	26.68	12.37
19	Worthington	OH			14869	2.67	2.28	94.74	2.67	0.74	13.71	20.33	9.48
20	Arlington	TX			261721	2.6	8.41	82.64	3.92	8.91	14.18	32.78	15.14
21	Plymouth	MN	53	51730	50889	2.77	1.61	95.66	2.04	1.02	13.13	26.94	12.19
22	New York	NY			7322564	2.6	28.71	52.26	7.0	24.36	13.06	27.46	13.09
23	Marina	CA			26436	3.34	18.97	53.6	20.84	10.73	16.16	37.22	19.09
24	Lebanon	NH	9	41300	12183	2.36	0.41	97.55	1.55	0.91	11.29	26.87	12.2
25	Rockledge	FL			16023	2.63	13.79	83.94	1.42	2.4	11.91	23.09	10.33
26	Rogers	AR			24692	2.54	0.06	97.72	0.77	1.86	12.84	25.81	11.93
27	Bellaire	TX			13842	2.35	0.41	94.65	1.98	7.95	8.23	18.3	6.98
28	El Cajon	CA			88693	2.7	2.92	87.36	2.82	13.97	13.77	30.92	15.15

Over 150 fields...

Sprint 2 Uploading datasets



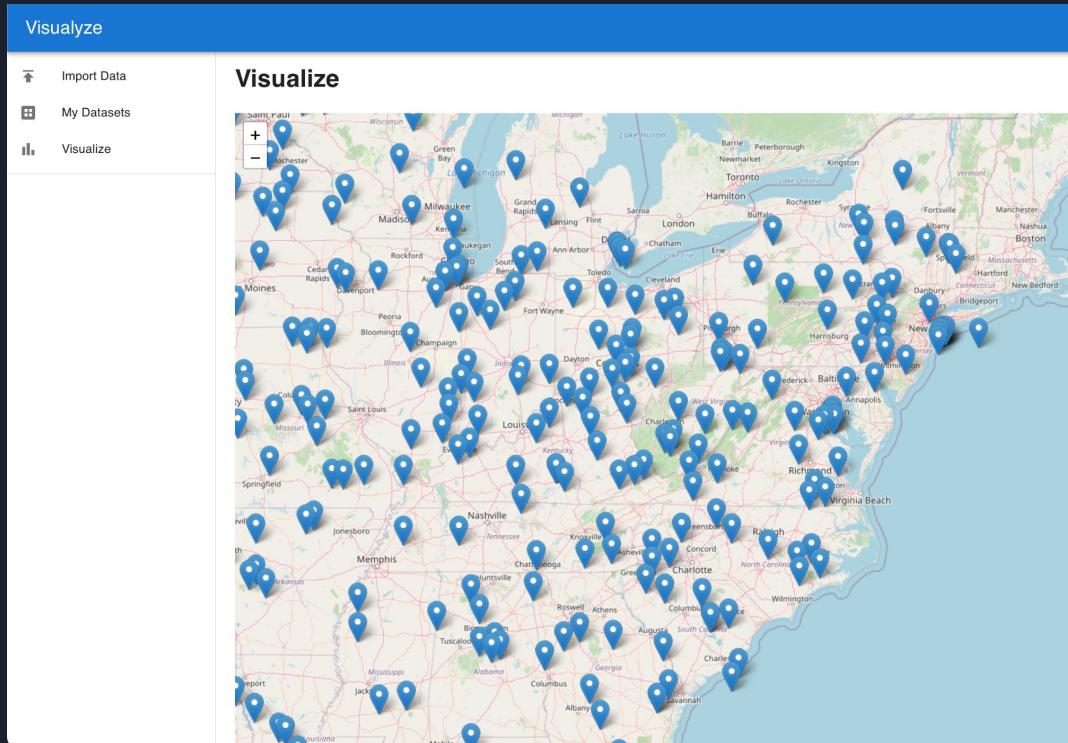
My datasets (S2)

The image shows two side-by-side screenshots of a web application interface. The left screenshot displays a navigation bar with 'Visualize' at the top, followed by three menu items: 'Import Data', 'My Datasets' (which is selected), and 'Visualize'. Below the menu is a section titled 'My Datasets' containing a single item: 'Dataset 1'. The right screenshot shows a browser window with the URL 'localhost:5000/datasets'. The browser's title bar also includes 'Philippe' and 'Diana - Playing'. The developer tools Network tab is open, showing a list of dataset entries. Each entry consists of a coordinate pair enclosed in brackets, such as '[40.658287, -74.421615]'. There are approximately 100 such entries listed.

Dataset	Coordinates
Dataset 1	[40.658287, -74.421615]
Dataset 2	[40.657853, -74.421038]
Dataset 3	[40.657771, -74.420772]
Dataset 4	[40.657004, -74.420462]
Dataset 5	[40.657842, -74.420166]
Dataset 6	[40.658029, -74.419784]
Dataset 7	[40.658084, -74.419711]
Dataset 8	[40.658537, -74.419293]
Dataset 9	[40.65877, -74.419077]
Dataset 10	[40.658935, -74.418839]
Dataset 11	[40.659171, -74.418594]
Dataset 12	[40.659775, -74.417793]
Dataset 13	[40.659874, -74.417742]
Dataset 14	[40.659889, -74.417739]
Dataset 15	[40.66031, -74.417646]
Dataset 16	[40.660713, -74.417448]
Dataset 17	[40.660922, -74.417273]
Dataset 18	[40.660999, -74.417216]
Dataset 19	[40.66129, -74.416927]
Dataset 20	[40.661806, -74.416199]
Dataset 21	[40.661819, -74.416169]
Dataset 22	[40.662053, -74.415665]
Dataset 23	[40.662152, -74.415341]
Dataset 24	[40.662284, -74.414742]
Dataset 25	[40.662306, -74.414281]
Dataset 26	[40.662306, -74.413631]
Dataset 27	[40.662327, -74.413335]
Dataset 28	[40.662393, -74.413138]
Dataset 29	[40.662459, -74.412939]
Dataset 30	[40.662664, -74.412655]
Dataset 31	[40.662405, -74.412405]
Dataset 32	[40.66314, -74.412037]
Dataset 33	[40.663266, -74.411857]
Dataset 34	[40.663661, -74.411431]
Dataset 35	[40.663692, -74.411402]
Dataset 36	[40.663962, -74.411158]
Dataset 37	[40.66437, -74.410789]
Dataset 38	[40.664743, -74.410522]
Dataset 39	[40.665331, -74.41006]
Dataset 40	[40.665611, -74.409743]
Dataset 41	[40.665797, -74.40944]
Dataset 42	[40.665885, -74.409258]
Dataset 43	[40.665951, -74.409122]
Dataset 44	[40.666068, -74.408836]
Dataset 45	[40.66628, -74.40832]
Dataset 46	[40.666505, -74.40773]
Dataset 47	[40.666692, -74.407362]
Dataset 48	[40.667021, -74.406901]
Dataset 49	[40.667274, -74.406713]
Dataset 50	[40.667746, -74.406468]
Dataset 51	[40.667845, -74.406366]
Dataset 52	[40.667982, -74.406136]
Dataset 53	[40.66801, -74.405927]
Dataset 54	[40.667982, -74.405681]
Dataset 55	[40.667921, -74.405515]
Dataset 56	[40.667798, -74.405281]
Dataset 57	[40.667702, -74.405097]
Dataset 58	[40.667614, -74.404874]
Dataset 59	[40.667432, -74.404492]
Dataset 60	[40.667185, -74.404117]
Dataset 61	[40.666888, -74.403766]
Dataset 62	[40.666822, -74.403689]
Dataset 63	[40.666752, -74.403606]
Dataset 64	[40.666722, -74.403571]
Dataset 65	[40.666537, -74.403535]
Dataset 66	[40.666954, -74.403468]
Dataset 67	[40.667256, -74.403496]
Dataset 68	[40.667638, -74.403396]
Dataset 69	[40.6682016, -74.4027582]
Dataset 70	[40.669238, -74.40157]
Dataset 71	[40.66919, -74.401591]
Dataset 72	[40.669996, -74.401576]
Dataset 73	[40.670435, -74.401541]
Dataset 74	[40.670492, -74.4016838]
Dataset 75	[40.670635, -74.401541]
Dataset 76	[40.670764, -74.4014216]
Dataset 77	[40.671132, -74.401012]
Dataset 78	[40.671561, -74.400386]
Dataset 79	[40.6716815, -74.4002018]
Dataset 80	[40.671769, -74.400068]
Dataset 81	[40.672148, -74.399671]
Dataset 82	[40.6724398, -74.39948]
Dataset 83	[40.672533, -74.399419]
Dataset 84	[40.672829, -74.399254]
Dataset 85	[40.673071, -74.399144]
Dataset 86	[40.6732328, -74.3990232]
Dataset 87	[40.673351, -74.398935]
Dataset 88	[40.6735267, -74.3986366]
Dataset 89	[40.6735693, -74.3984695]
Dataset 90	[40.6735741, -74.3982947]
Dataset 91	[40.6736399, -74.3980906]
Dataset 92	[40.6737067, -74.3980306]
Dataset 93	[40.6737928, -74.3980548]
Dataset 94	[40.674031, -74.39817]
Dataset 95	[40.674213, -74.398155]
Dataset 96	[40.674311, -74.398033]
Dataset 97	[40.674328, -74.39786]
Dataset 98	[40.674223, -74.397247]
Dataset 99	[40.674229, -74.396872]
Dataset 100	[40.674251, -74.396756]

Not functional...

Visualization map (S2)



Import Datasets (S3)

The screenshot shows the Visualize web application interface. At the top, there is a blue header bar with the title "Visualize". Below the header is a navigation sidebar on the left containing three items: "Import Data" (with a file icon), "My Datasets" (with a folder icon), and "Visualize" (with a bar chart icon). The main content area is titled "Import Datasets" and contains the instruction "Please select or drag and drop spreadsheets:". Below this instruction is a blue "UPLOAD" button. The entire main content area is enclosed in a dashed rectangular border.

My datasets (S3)

Visualize

Import Data

My Datasets

Visualize

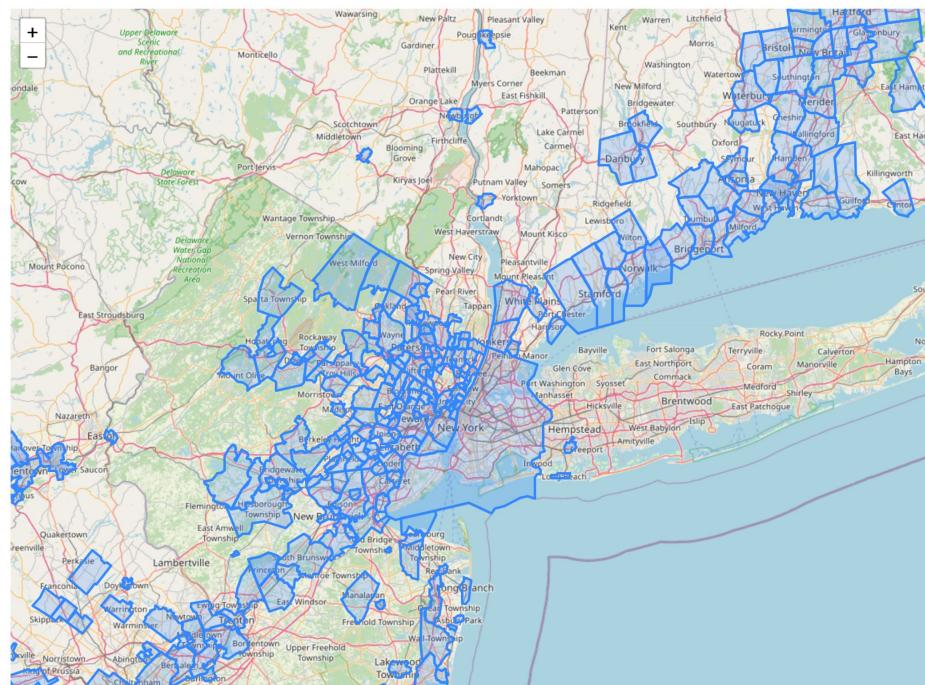
By Falcon, Piyush, Mack

small.db

communityName	countyCode	state	population	householdsize	murders	rapes	robberies	assaults	burglaries	larcenies	autoTheft	arsons	ViolentCrimesPerPop
Suffolk	800	VA	52141	2.82	9	37	160	304	646	1974	254	26	919.33
Martinsville	690	VA	16162	2.36	2	7	22	78	164	980	54	9	682.23
Lynchburg	680	VA	66049	2.63	3	45	107	351	560	2316	184	50	753.42
Manassas	683	VA	27957	2.95	0	10	55	32	141	1067	86	6	301.31
Fredericksburg	630	VA	19027	2.55	1	9	27	63	102	631	32	12	446.77
Vienna		VA	14852	2.8	1	3	4	7	66	379	24	4	94.07
Colonial Heights	570	VA	16064	2.52	0	6	15	18	88	718	43	1	234.77
Charlottesville	3	VA	13057	2.34	0	28	80	180	384	1934	149	15	699.49
Newport News	700	VA	170045	2.66	22	103	71	1334	2101	7267	684	63	845.63
Harrisonburg	660	VA	30707	2.98	1	7	17	26	166	1214	44	9	151.78
Salem	775	VA	23756	2.59	1	2	9	13	85	607	23	9	102.2
Herndon		VA	16139	2.96	0	3	10	10	42	556	21	2	127.65
Poquoson	735	VA	11005	2.92	0	1	1	40	19	150	5	2	355.9
Roanoke	770	VA	96397	2.35	11	32	217	305	118	4398	315	57	578.8
Radford	750	VA	15940	3.06	0	4	3	51	59	339	14	7	363.7
Hopewell	670	VA	23101	2.56	2	12	41	120	230	1009	37	33	708.53
Winchester	840	VA	21947	2.42	0	10	24	137	216	1494	54	7	711.46
Hampton	650	VA	133793	2.69	14	49	329	251	962	5532	561	73	455.92
Williamsburg	830	VA	11530	3.32	0	3	19	5	48	390	29	1	212.83
Christiansburg		VA	15004	2.5	0	6	4	10	86	716	38	0	112.94
Petersburg	730	VA	38386	2.61	11	26	221	293	810	1720	225	8	1332.66
Waynesboro	820	VA	18549	2.45	4	5	5	50	136	597	57	5	337.96

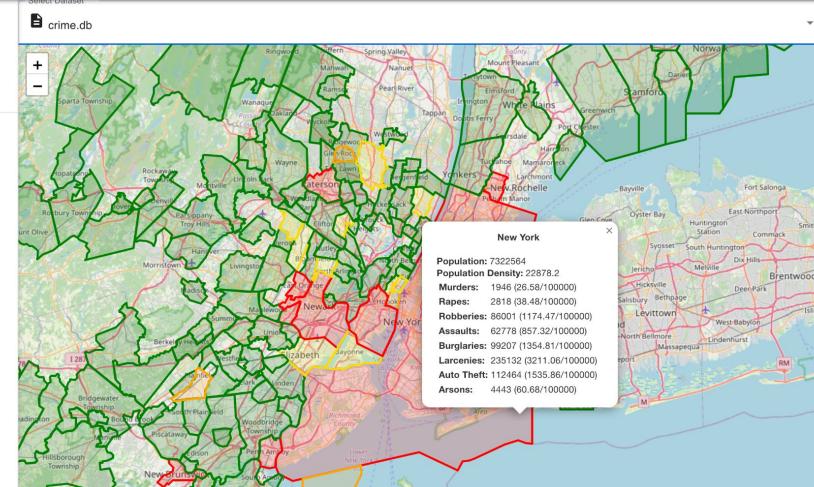
Visualization map (S3)

Visualize

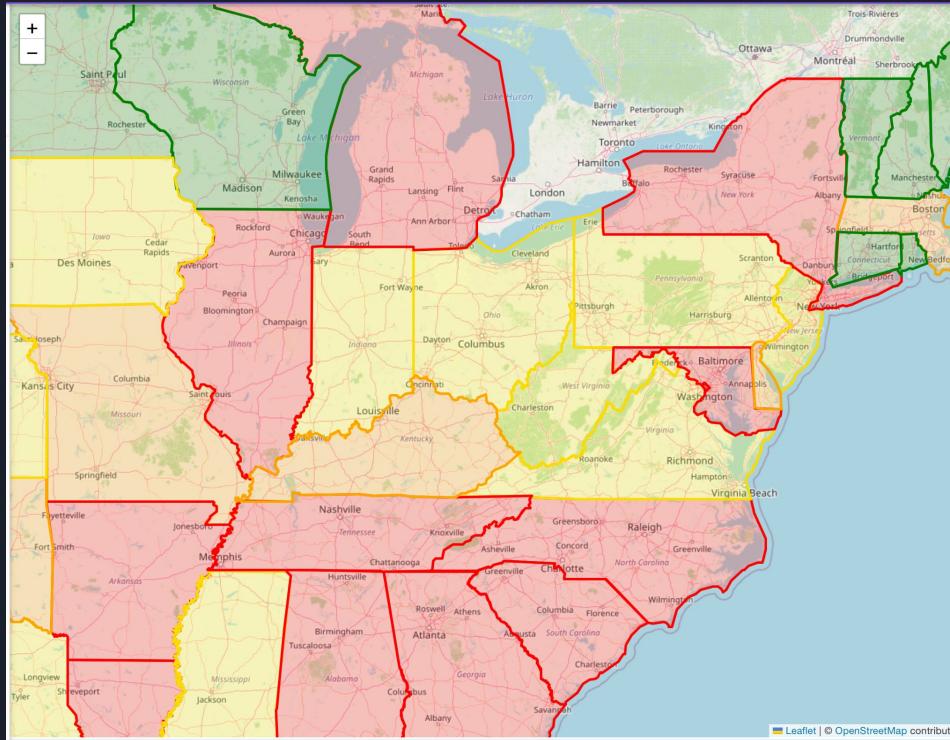


Visualize

- Import Data
- My Datasets
- Visualize



State level map (S3)





Contributions

Philippe

- Responsible for geolocations, boundaries, unit tests, and Map Leaflet

Piyush

- Responsible for React Dashboard UI, Import Datasets, Heatmap, Flask/GeoJson

Mack

- Responsible for data integrity, logging, and My Datasets multi selection



Testing

We tested our program using PyTest (backend) and logging (front).

Acceptance Criteria:

- The info stored on the front end matches exactly those on the backend including the attributes. The state level data are correctly summed, with special instructions on handling nulls.

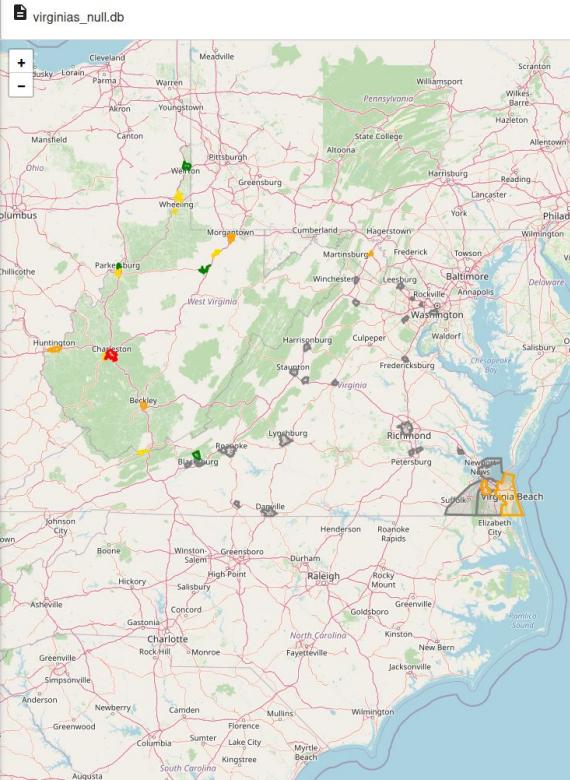
Example:

- Running Pytest to determine that there are 48 out of 50 states in the “sample” dataset.

Visualize

-  Import Data
-  My Datasets
-  Visualize

By Falcon, Piyush, Mack



```
Network Storage Inspector Console Debugger Performance Memory > 159
Filter Output Errors Warnings Logs Info Debug CSS XHR Requests Map.js:119
18:28:16.982 200: Map States
Map.js:119
18:28:17.949 Map: Zoom (8)
Map.js:33
18:28:17.950 Map: Loaded state markers on demand
Map.js:27
18:28:19.115 ↵ Object { rows: (2215) [...] }
Map.js:79
18:28:19.116 ⓘ Error processing mapdata information: Fonddu Lac ; 0 ; 0 ; [2]
Map.js:86
18:28:19.117 ⓘ Error processing mapdata information: Mc Allen ; 0 ; 0 ; [2]
Map.js:86
18:28:19.118 ⓘ Error processing mapdata information: Mc Kinney ; 0 ; 0 ; [2]
Map.js:86
18:28:19.119 ⓘ Error processing mapdata information: Mc Minnville ; 0 ; 0 ; [2]
Map.js:86
18:28:19.120 ⓘ Error processing mapdata information: Red Bank ; 0 ; 0 ; [2]
Map.js:86
18:28:19.121 ⓘ Error processing mapdata information: Middle ; 0 ; 0 ; [2]
Map.js:86
18:28:19.122 ⓘ Error processing mapdata information: South Kings ; 0 ; 0 ; [2]
Map.js:86
18:28:19.123 ⓘ Error processing mapdata information: East Lampeter ; 0 ; 0 ; [2]
Map.js:86
18:28:19.124 ⓘ Error processing mapdata information: Lower Paxton ; 0 ; 0 ; [2]
Map.js:86
18:28:19.125 ⓘ Error processing mapdata information: West Hempfield ; 0 ; 0 ; [2]
Map.js:86
18:28:19.126 ⓘ Error processing mapdata information: Mc Minnville ; 0 ; 0 ; [2]
Map.js:86
18:28:19.127 ⓘ Error processing mapdata information: Mc Alester ; 0 ; 0 ; [2]
Map.js:86
18:28:19.128 ⓘ Error processing mapdata information: Estonborough ; 0 ; 0 ; [2]
Map.js:86
18:28:19.129 ⓘ Error processing mapdata information: South Orange Village ; 0 ; 0 ; [2]
Map.js:86
18:28:19.130 ⓘ Error processing mapdata information: Cityof Orange ; 0 ; 0 ; [2]
Map.js:86
18:28:19.125 ⓘ Error processing mapdata information: Little Egg Harbor ; 0 ; 0 ; [2]
Map.js:86
18:28:19.126 ⓘ Error processing mapdata information: Mc Comb ; 0 ; 0 ; [2]
Map.js:86
18:28:19.126 ⓘ Error processing mapdata information: Water ; 0 ; 0 ; [2]
Map.js:86
18:28:19.127 ⓘ Error processing mapdata information: La Grange ; 0 ; 0 ; [2]
Map.js:86
18:28:19.127 ⓘ Error processing mapdata information: Water ; 0 ; 0 ; [2]
Map.js:86
18:28:19.128 ⓘ Error processing mapdata information: Grover City ; 0 ; 0 ; [2]
Map.js:86
18:28:19.126 ⓘ Error processing mapdata information: TwentyNine Palms-Morongo Valley ; 0 ; 0 ; [2]
Map.js:86
18:28:21.208 Map: Controller
Map.js:25
18:28:22.320 200: Map States
Map.js:119
18:28:24.453 Map: Controller
Map.js:25
18:28:26.820 Map: Zoom (?)
Map.js:33
18:28:27.350 Map: Controller
Map.js:25
18:28:28.241 200: Map States
Map.js:119
18:28:29.994 Map: Controller
Map.js:25
18:28:35.884 ⓘ crimedata.db
Visualize.js:65
18:28:35.885 ↵ Array(1) ["virginias_reduced.db", "virginias_nullall.db", "crimedata.db", "blank.db", "virginias_null.db", "virginias.db", "virginias_switch.db", "nullcrime.db", "small_nullpop.db", "crimedata_corruptstate.db", ...]
Visualize.js:65
18:28:35.890 Sidebar rendered
Sidebar.js:21
18:28:36.032 Reloading map with dataset virginias_null.db
Map.js:206
18:28:36.033 Map: Updated
Map.js:50
18:28:36.431 Reloading map with dataset virginias_null.db
Map.js:206
18:28:36.431 Map: Updated
Map.js:50
18:28:36.627 Map: Controller
Map.js:25
18:28:36.640 200: Map Cities
Map.js:70
18:28:36.662 ↵ Object { rows: (47) [...] }
Map.js:79
18:28:36.712 Map: Controller
Map.js:25
18:28:36.773 ↵ Object { rows: (47) [...] }
Map.js:79
18:28:36.937 Map: Controller
Map.js:25
>>
```

Testing & Logging



Mapping Problems

Dataset issues

Original dataset had incorrect lat and lon

When determining map boundaries of certain cities, lat and lon were swapped with eachother

Gathering this data requires using the openstreetmap api, this is rate limited to 1/sec.
This means upload will be considerably slower

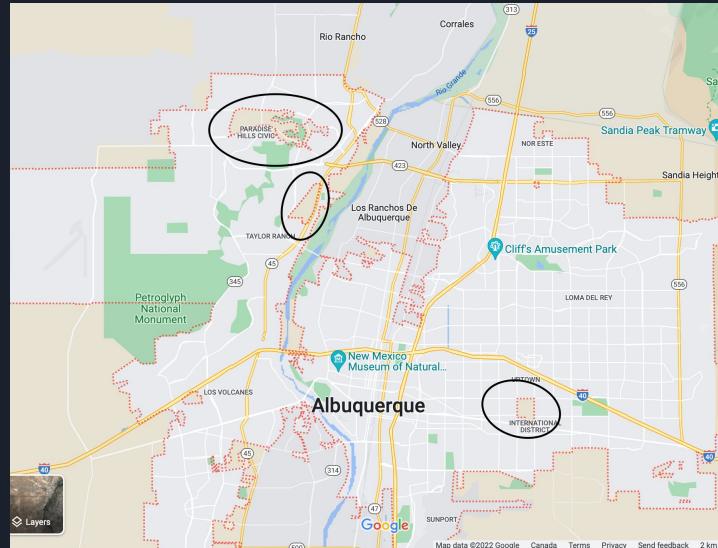
Speed

For large datasets, loading can be very slow and laggy at the city level

Strange boundaries

Many US cities have unincorporated areas not part of the actual city. This made cities not be continuous.

Ex:





Future Improvements

- Our current schema involves over 150 fields, many of them quite extraneous and not for immediate use.
- This means CSV files must be of a specific format, and special procedures must be done to handle CSV files that do not conform, both on the front and back end.
- Null values also may need special consideration.
- The sheer number of fields means live-updating will be very difficult, so all data will be cached.



Future Improvements

- Cross platform, docker issues, dataset issues, quirks with cities overlapping with each other, not being continuous
- Possible meeting irregularities
- Flow on how to generate dataset can be somewhat awkward; especially that the geojson part can be time consuming and limited by the api call limit
- Thus, db are stored locally in cache



Questions?