OpenUGRC: Leveraging OpenStreetMap for Statewide Data

Erik Neemann

11 May 2022

gis.utah.gov/presentations

gis.utah.gov

Overview

- Intro - Utah Open Source Places
- Motivation & Uses
- More Info & Data Quirks
- How it's Built
- How to Participate

Utah Geospatial Resource Center

gis.utah.gov

Intro - Utah Open Source Places

- New addition to UGRC's Statewide Geographic Information Datasource (SGID)
- Represents places of interest in Utah as point data
  - Tailored toward businesses and community locations
  - Intended to minimize overlap with GNIS Place Names
- Curated from OpenStreetMap (OSM) data
  - Currently contains over 23,800 features
- Updated frequently (every other week)
- Continually growing…over 400 features added since the initial data build in early March
- As a derivative product, Open Source Places carries the OSM data license: Open Database License (ODbL) v1.0

© OpenStreetMap contributors

Utah Geospatial Resource Center

UGRC's creation of the Utah Open Source Places dataset does not imply endorsement by, or affiliation with, the OpenStreetMap Foundation

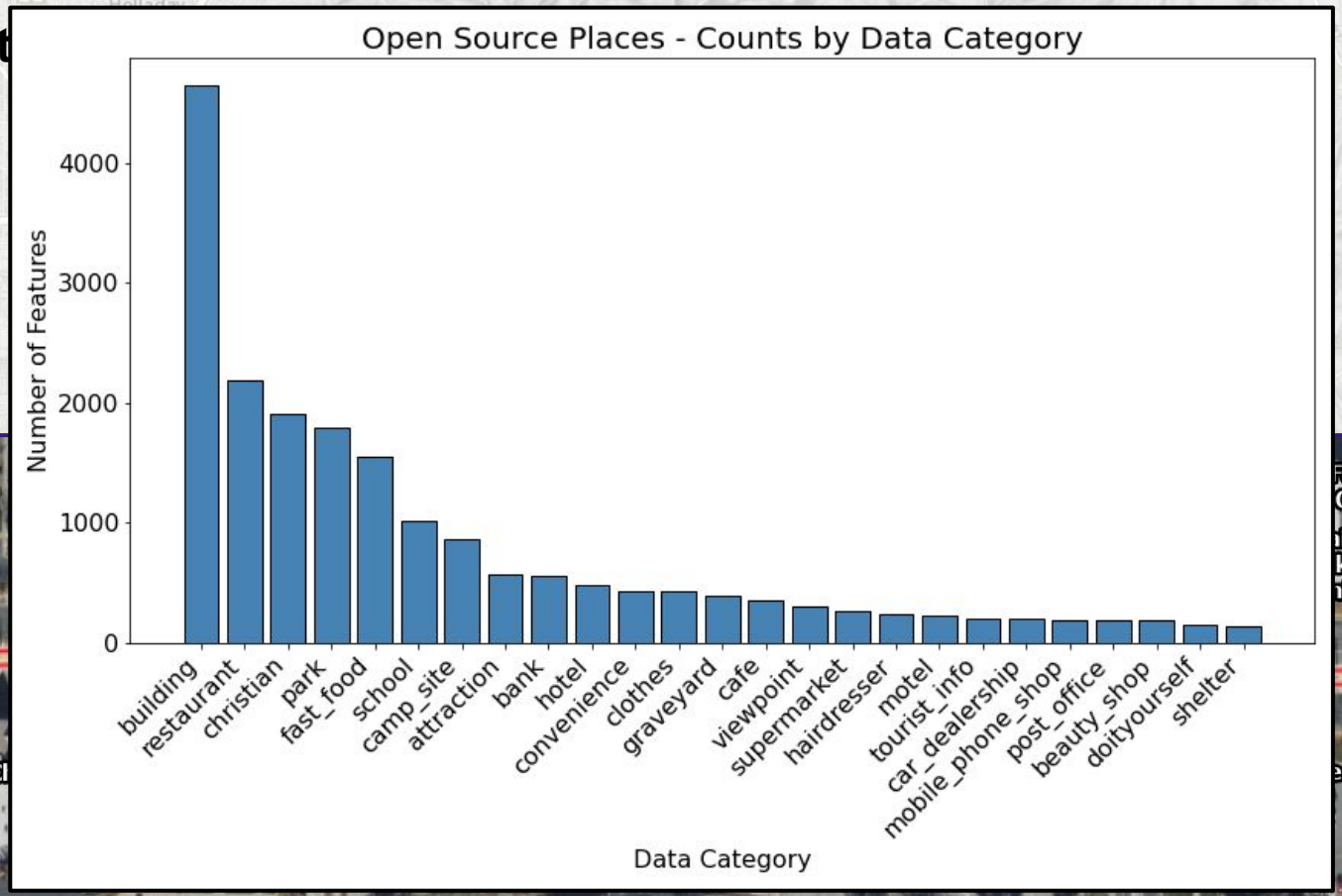
What's in there?

- Almost anything you'll find in OSM data
  - businesses
  - schools
  - viewpoints
  - parks
  - restaurants
  - banks
  - car dealerships
  - shops
  - churches
  - hotels
  - campgrounds
  - attractions

Utah Geospatial Resource Center

What's in there?

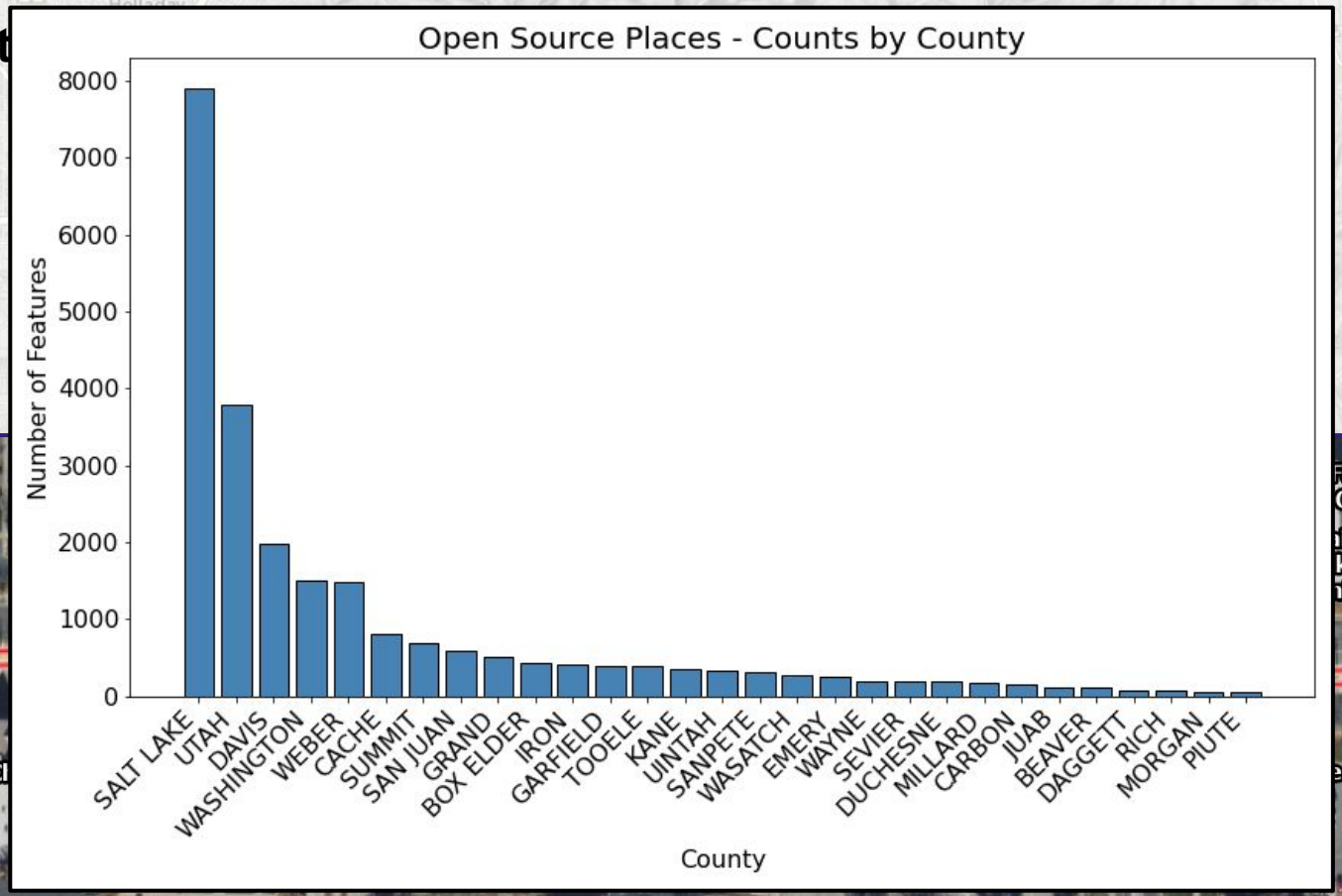
- **Almost**
  -
  -
  -
  -



Utah Geospatial Resource Center

What's in there?

- **Almost**
  -
  -
  -
  -



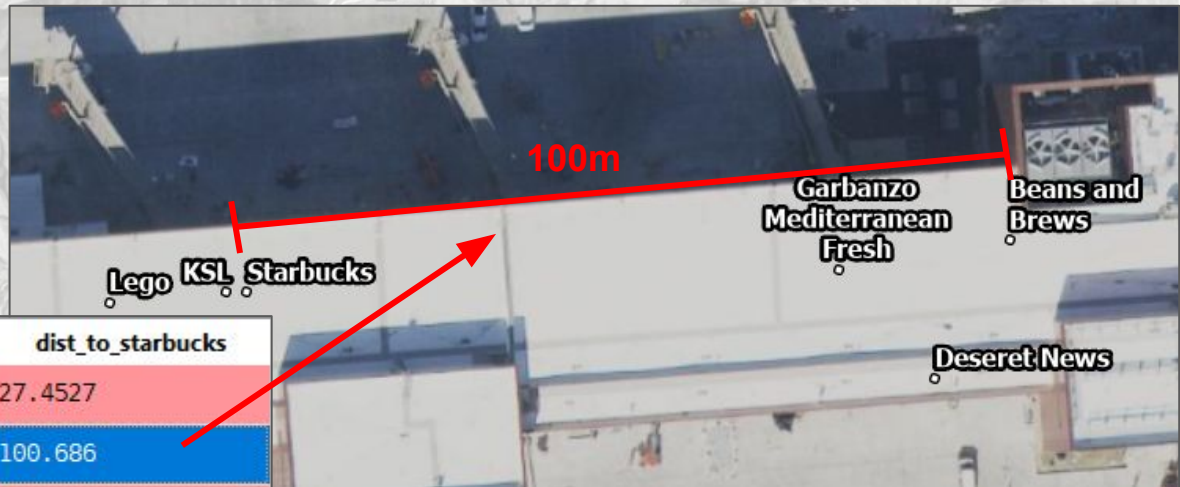
Utah Geospatial Resource Center

Motivation

- Original idea born out of the need for a statewide "common places" layer that could be used by 911 dispatchers to help locate callers
  - Quickly realized there were many other uses for the data
  - And more reasons to create the data
- Make OSM data more accessible
- Encourage growth of OSM data in Utah
  - Anyone can create an account, add, edit, improve OSM data
- Build upon OSM data with info from UGRC's SGID
- Allow users to directly contribute to a UGRC data layer

How can the data be used?

- Numerous potential uses for government agencies, citizens, researchers
- Analyses of business types or specific businesses
  - What cities have the highest fast food/population ratio?
  - How many Beans & Brews are within a half mile radius of Starbucks?
- Place locations can help 911 call-takers & emergency responders/managers
- Analysis of 'community locations' (tobacco regulation laws)
- Quality control other datasets for missing/closed locations
  - schools
  - fire stations
  - liquor stores
  - cemeteries, etc., etc.



osm_id	shape	name	dist_to_starbucks
2403976399	POINT (426496.6941…	Beans & Brews Coffee House	27.4527
8072353912	POINT (416551.4011…	Beans and Brews	100.686

Other info

- **Similarities to other SGID layers**
  - May include places present in other SGID datasets (parks, school, libraries, fire stations, etc.)
  - Available in variety of ways and formats
    - ArcGIS Online, OpenData, OpenSGID, web services, downloads, etc.
- **Differences from other SGID layers**
  - Not authoritative data
  - Different license (ODbL 1.0 instead of our usual CC BY 4.0)
  - Any user can contribute!
- **Main OSM data attributes:**

osm_id	category	name	amenity	cuisine	tourism	shop
3998254042	cafe	La Barba Coffee	cafe	<Null>	<Null>	<Null>
3998254041	jeweller	Morgan Jewelers	<Null>	<Null>	<Null>	jewelry
2328421314	museum	Discovery Gateway Children's Museum	<Null>	<Null>	museum	<Null>
145043325	park	Olympic Legacy Plaza	<Null>	<Null>	<Null>	<Null>
2327662158	restaurant	Flemming's Prime Steakhouse	restaurant	steak_house	<Null>	<Null>
23335190	stadium	vivint arena	<Null>	<Null>	<Null>	<Null>

Data quirks

- Expect some typos, missing locations, closed locations, naming differences, etc.
  - Be creative with definition queries:

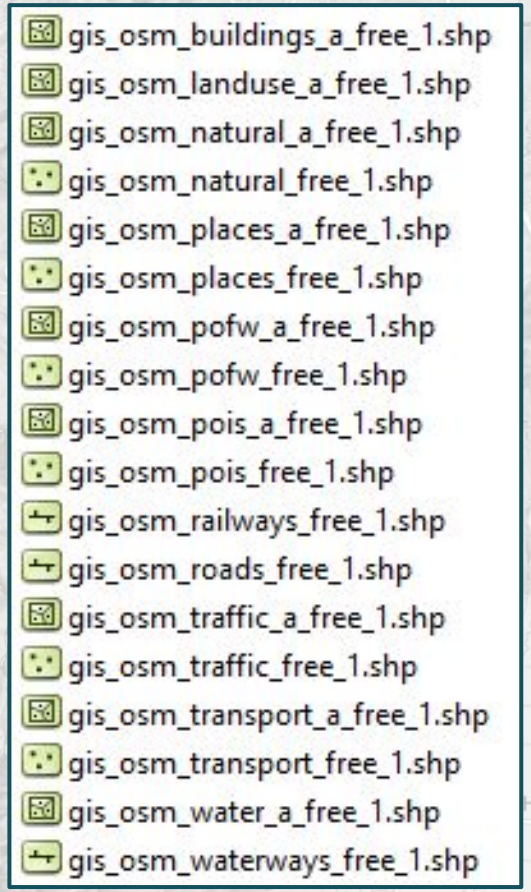
    `name like '%eans%rews%'`
- Several locations simply have a category of 'building'
  - Good candidate for OSM edits!
- Often a European 'flavour' to OSM data
  - centre (instead of center, ex: sports_centre, garden_centre)
  - theatre (instead of theater)
  - caravan_site (instead of rv_park or trailer_park)
  - chemist (instead of pharmacy)
  - tyre_shop (instead of tire_shop)
- Up to two addresses are provided
  - `osm_addr` - address straight from OSM data (25.8%)
  - `ugrc_addr` - nearest address point (within 25m) from statewide database (59.4%)

name	
Beans & Brews Coffee House	⭐preferred
Beans and Brews	
Beans & Brews Coffee House	
Beans & Brews Coffee House	
Beans & Brews Coffee House	
Beans & Brews Coffee House	
Beans & Brews Coffeehouse	

Utah Geospatial Resource Center

How's it built? - Python script details

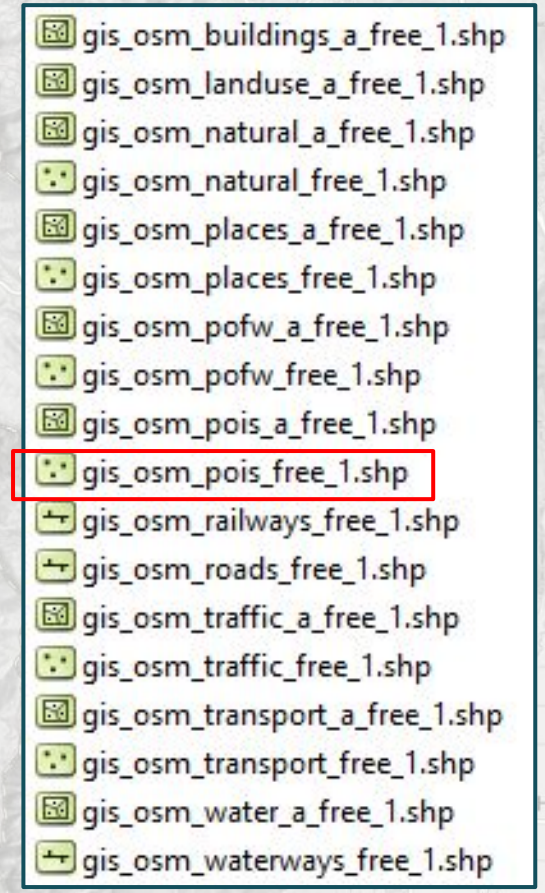
- Nightly OSM archive downloaded from Geofabrik (18 shapefiles)
- Several datasets combined into a single point layer
  - Specific 'categories' and data attributes are selected
  - Must have information in 'name' field
  - Polygons converted to internal centroids
  - Spatial filtering (3 or 10 meters) helps prevent duplicates
- Attributes are added from SGID polygons
  - `county`, `city`, `zip`, (census) `block_id`
- Nearby addresses are added from SGID address pts
- Additional attributes are joined from full OSM data
  - Pulled via Overpass API query into pandas dataframe
  - Joined using the 'osm_id'
    - `amenity`, `cuisine`, `tourism`, `shop`
    - `website`, `phone`, `open_hours`



Utah Geospatial Resource Center

How's it built? - Python script details

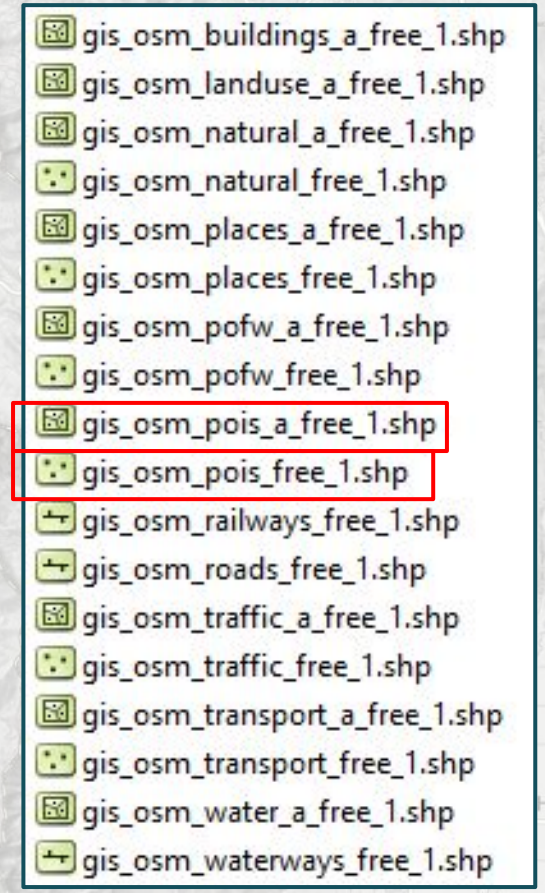
- Nightly OSM archive downloaded from Geofabrik (18 shapefiles)
- Several datasets combined into a single point layer
  - Specific 'categories' and data attributes are selected
  - Must have information in 'name' field
  - Polygons converted to internal centroids
  - Spatial filtering (3 or 10 meters) helps prevent duplicates
- Attributes are added from SGID polygons
  - `county`, `city`, `zip`, (census) `block_id`
- Nearby addresses are added from SGID address pts
- Additional attributes are joined from full OSM data
  - Pulled via Overpass API query into pandas dataframe
  - Joined using the 'osm_id'
    - `amenity`, `cuisine`, `tourism`, `shop`
    - `website`, `phone`, `open_hours`



Utah Geospatial Resource Center

How's it built? - Python script details

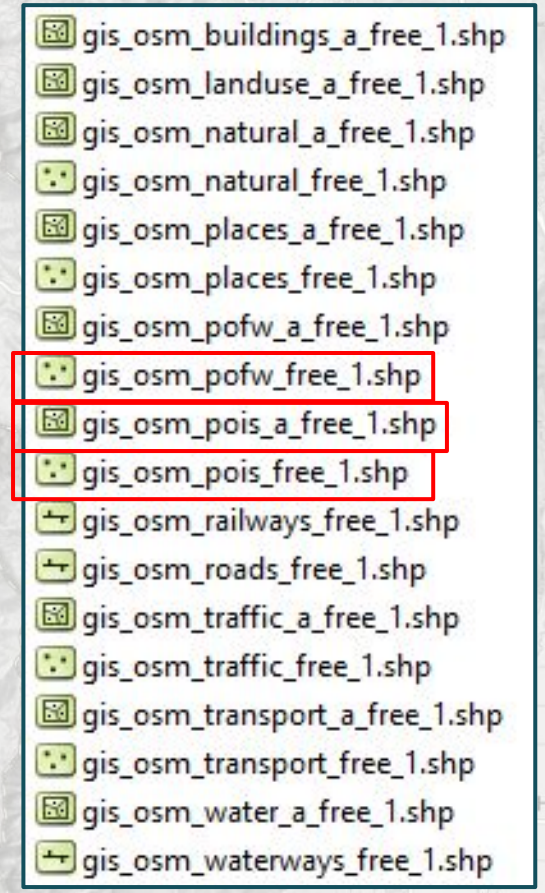
- Nightly OSM archive downloaded from Geofabrik (18 shapefiles)
- Several datasets combined into a single point layer
  - Specific 'categories' and data attributes are selected
  - Must have information in 'name' field
  - Polygons converted to internal centroids
  - Spatial filtering (3 or 10 meters) helps prevent duplicates
- Attributes are added from SGID polygons
  - `county`, `city`, `zip`, (census) `block_id`
- Nearby addresses are added from SGID address pts
- Additional attributes are joined from full OSM data
  - Pulled via Overpass API query into pandas dataframe
  - Joined using the 'osm_id'
    - `amenity`, `cuisine`, `tourism`, `shop`
    - `website`, `phone`, `open_hours`



Utah Geospatial Resource Center

How's it built? - Python script details

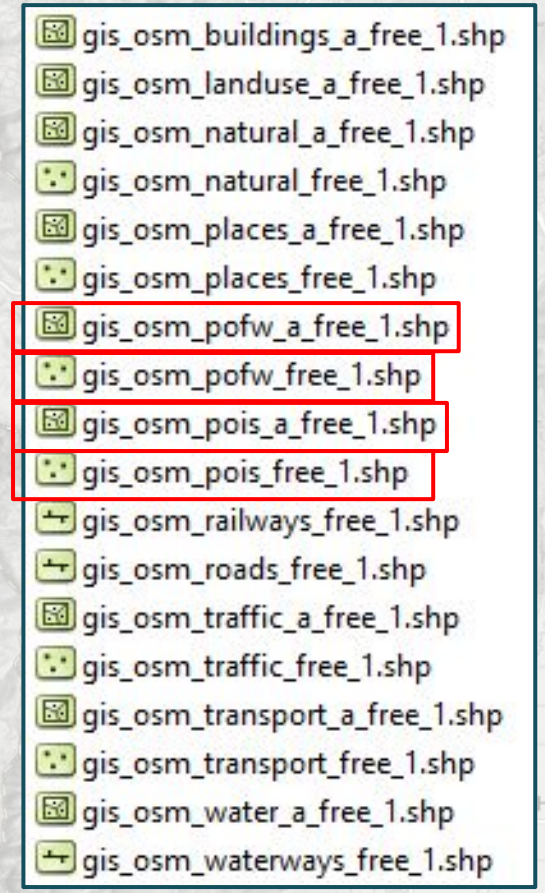
- Nightly OSM archive downloaded from Geofabrik (18 shapefiles)
- Several datasets combined into a single point layer
  - Specific 'categories' and data attributes are selected
  - Must have information in 'name' field
  - Polygons converted to internal centroids
  - Spatial filtering (3 or 10 meters) helps prevent duplicates
- Attributes are added from SGID polygons
  - `county`, `city`, `zip`, (census) `block_id`
- Nearby addresses are added from SGID address pts
- Additional attributes are joined from full OSM data
  - Pulled via Overpass API query into pandas dataframe
  - Joined using the 'osm_id'
    - `amenity`, `cuisine`, `tourism`, `shop`
    - `website`, `phone`, `open_hours`



Utah Geospatial Resource Center

How's it built? - Python script details

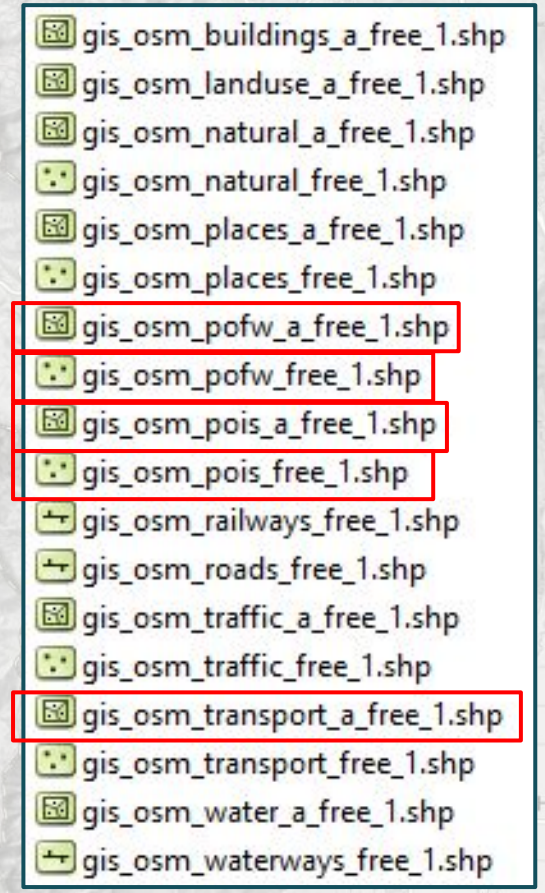
- Nightly OSM archive downloaded from Geofabrik (18 shapefiles)
- Several datasets combined into a single point layer
  - Specific 'categories' and data attributes are selected
  - Must have information in 'name' field
  - Polygons converted to internal centroids
  - Spatial filtering (3 or 10 meters) helps prevent duplicates
- Attributes are added from SGID polygons
  - `county`, `city`, `zip`, (census) `block_id`
- Nearby addresses are added from SGID address pts
- Additional attributes are joined from full OSM data
  - Pulled via Overpass API query into pandas dataframe
  - Joined using the 'osm_id'
    - `amenity`, `cuisine`, `tourism`, `shop`
    - `website`, `phone`, `open_hours`



Utah Geospatial Resource Center

How's it built? - Python script details

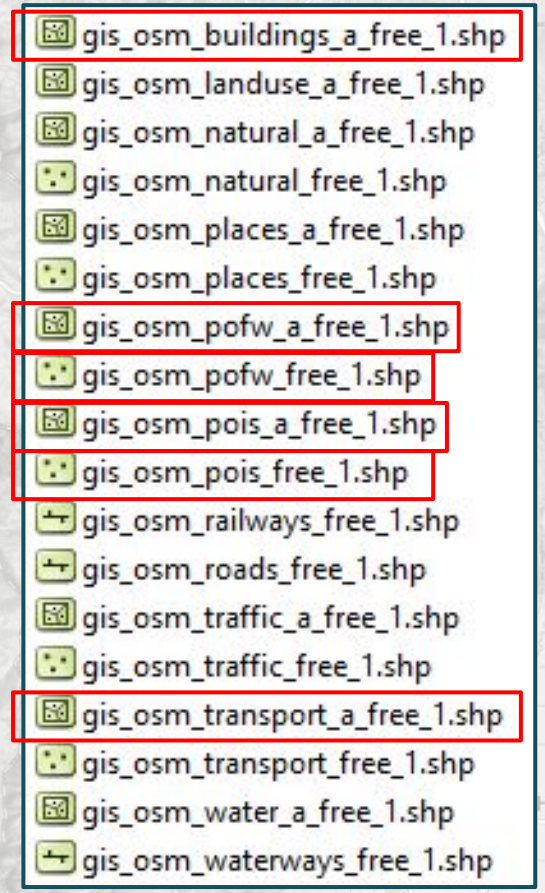
- Nightly OSM archive downloaded from Geofabrik (18 shapefiles)
- Several datasets combined into a single point layer
  - Specific 'categories' and data attributes are selected
  - Must have information in 'name' field
  - Polygons converted to internal centroids
  - Spatial filtering (3 or 10 meters) helps prevent duplicates
- Attributes are added from SGID polygons
  - `county`, `city`, `zip`, (census) `block_id`
- Nearby addresses are added from SGID address pts
- Additional attributes are joined from full OSM data
  - Pulled via Overpass API query into pandas dataframe
  - Joined using the 'osm_id'
    - `amenity`, `cuisine`, `tourism`, `shop`
    - `website`, `phone`, `open_hours`



Utah Geospatial Resource Center

How's it built? - Python script details

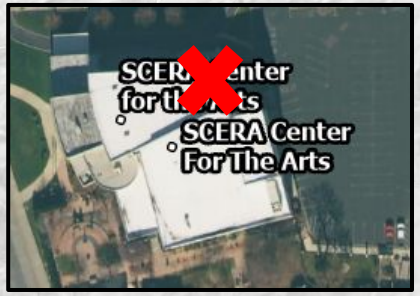
- Nightly OSM archive downloaded from Geofabrik (18 shapefiles)
- Several datasets combined into a single point layer
  - Specific 'categories' and data attributes are selected
  - Must have information in 'name' field
  - Polygons converted to internal centroids
  - Spatial filtering (3 or 10 meters) helps prevent duplicates
- Attributes are added from SGID polygons
  - `county`, `city`, `zip`, (census) `block_id`
- Nearby addresses are added from SGID address pts
- Additional attributes are joined from full OSM data
  - Pulled via Overpass API query into pandas dataframe
  - Joined using the 'osm_id'
    - `amenity`, `cuisine`, `tourism`, `shop`
    - `website`, `phone`, `open_hours`

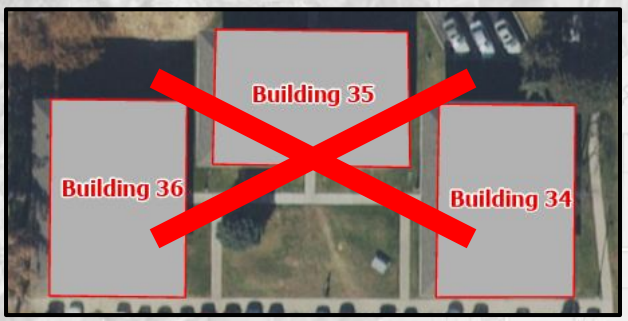


Utah Geospatial Resource Center

Python script - data cleanup

- Additional deduplication based on `name` and `block_id`
  - `block_id` field used like a spatial index
  - `name` and `block_id` combined into a single, lower-case string
  - Second or more occurrences of combined string are deleted from final data
  - Prevents two features with same name from existing in the same census block
- Removal of short or numeric-dominated 'names'
  - 'building' is stripped out of the `name` field
  - Feature is deleted if:
    - `len(name) < 3`
    - `name` is >50% numeric characters
  - Prevents apartment buildings (without context) or homes from appearing
    - 3A
    - 1228
    - Building 2
    - 675B

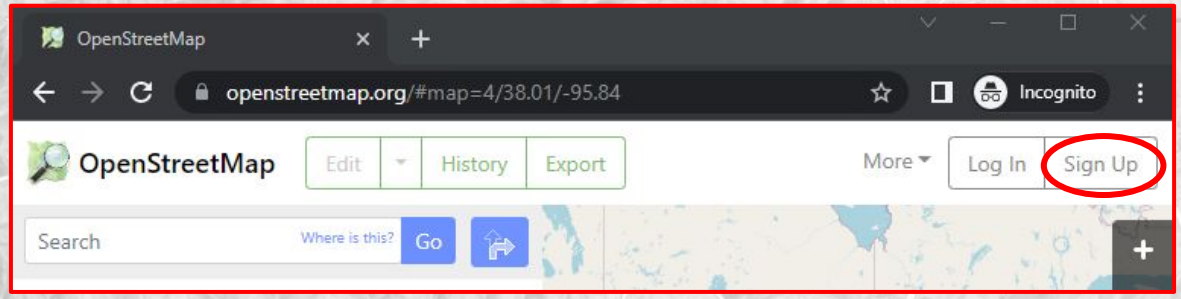




Utah Geospatial Resource Center

How to participate?

- 1) Create an OSM account
- 2) Log in and make edits



Data Editors

- Web-based iD editor
- Desktop and Mobile (Android and iOS)
  - JOSM, Vespucci, GoMap!!

OSM Community

- OSM Calendar shows upcoming events: https://osmcal.org/?in=United%20States
- Slack community: https://slack.openstreetmap.us/
  - Specifically, the `#local-utah` channel

Utah Geospatial Resource Center

How to participate?

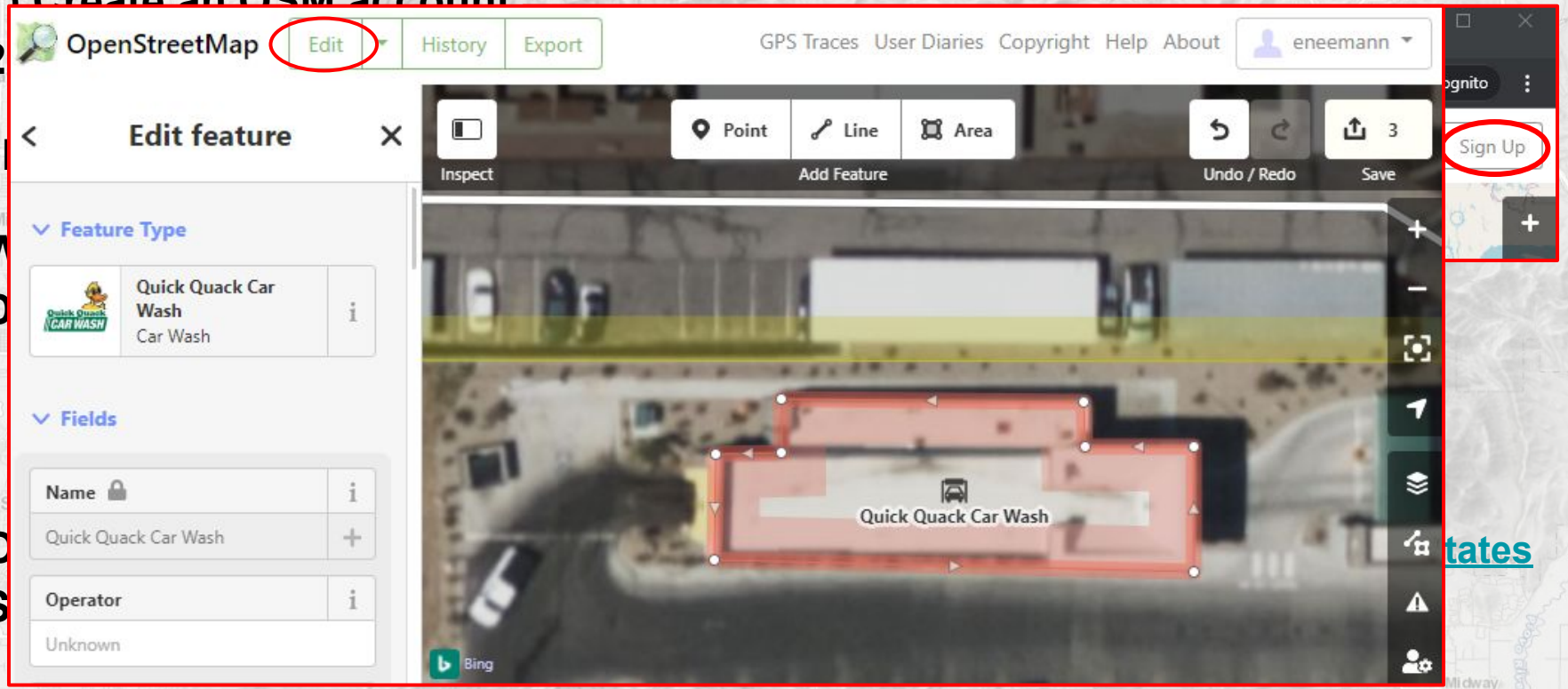
- 1) Create an OSM account
- 2

Data

- W
- D

OSM

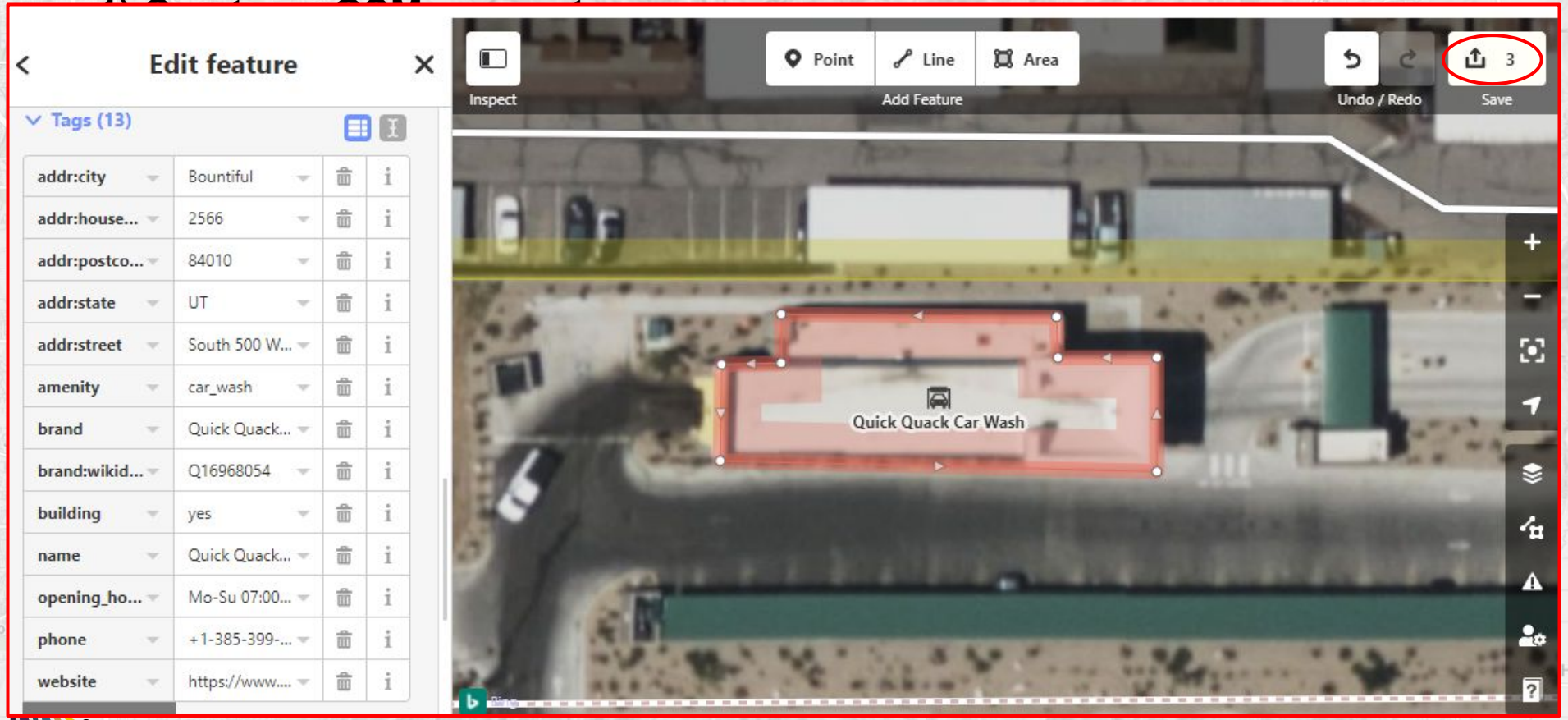
- C
- S



tates

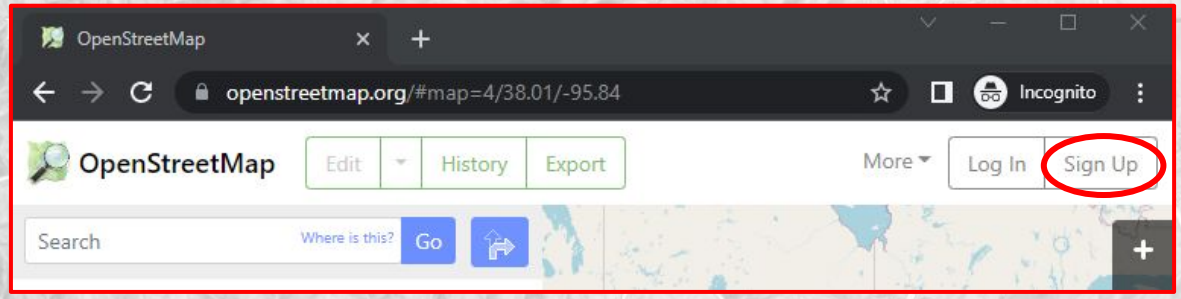
Utah Geospatial Resource Center

How to participate?



How to participate?

- 1) Create an OSM account
- 2) Log in and make edits



Data Editors

- Web-based iD editor
- Desktop and Mobile (Android and iOS)
  - JOSM, Vespucci, GoMap!!

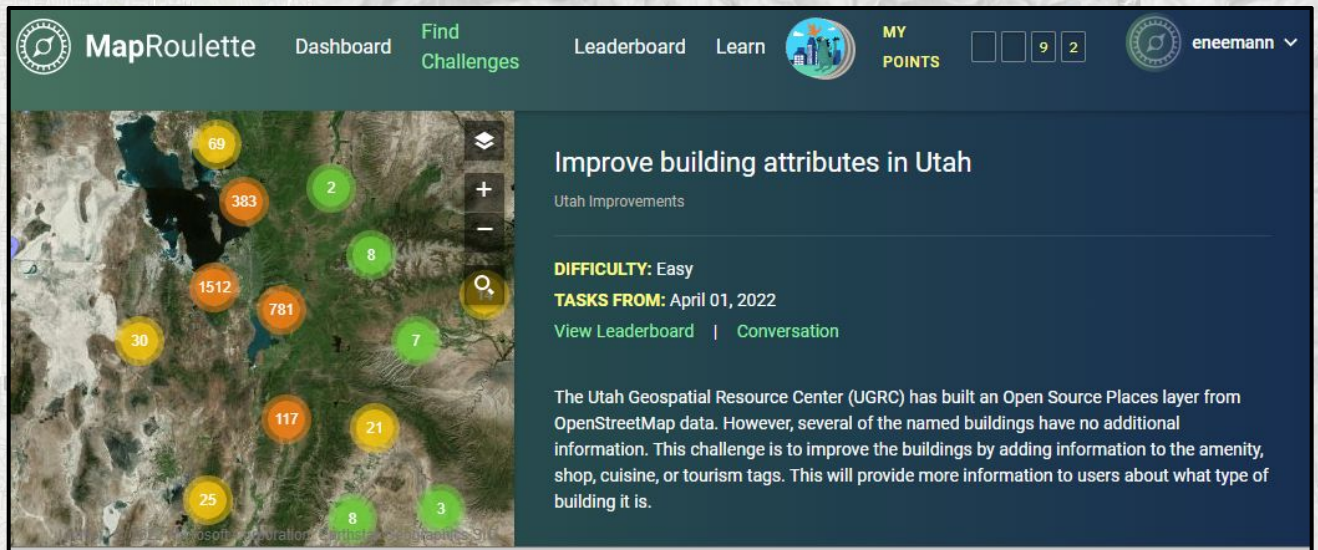
OSM Community

- OSM Calendar shows upcoming events: https://osmcal.org/?in=United%20States
- Slack community: https://slack.openstreetmap.us/
  - Specifically, the `#local-utah` channel

Utah Geospatial Resource Center

How to participate? - Map Roulette!

- Tool that serves random micro-tasks for OSM data improvements
- Challenges are created by users with specific tasks
- Other users can participate and improve the data
- Help 'Improve building attributes in Utah'!
  - https://maproulette.org/browse/challenges/26977
  - Add informational tags (`amenity`, `shop`, `cuisine`, `tourism`) or specify building type



Utah Geospatial Resource Center

MapRoulette Dashboard Find Challenges Leaderboard Learn MY POINTS 97 eneemann

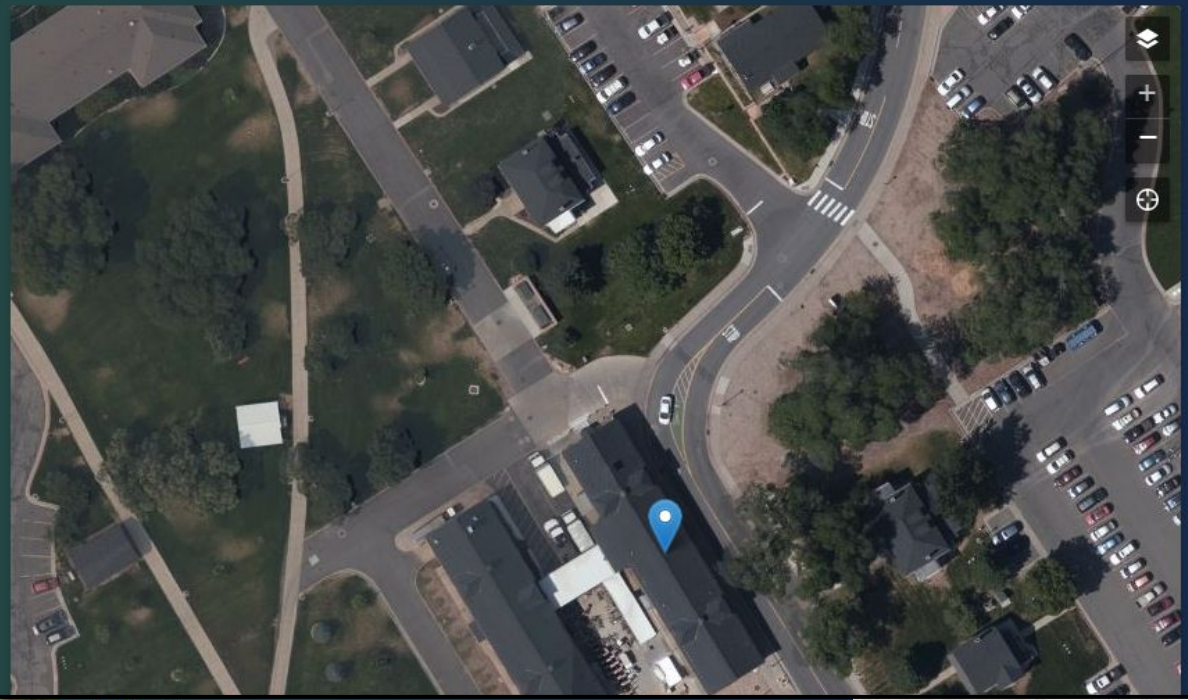
Improve building attributes in Utah

Utah Improvements

INSTRUCTIONS

tags, as necessary. Use the Wiki pages to guide your tag updates:

- https://wiki.openstreetmap.org/wiki/Key:building
- https://wiki.openstreetmap.org/wiki/Key:amenity
- https://wiki.openstreetmap.org/wiki/Key:shop
- https://wiki.openstreetmap.org/wiki/Key:cuisine
- https://wiki.openstreetmap.org/wiki/Key:tourism



COMPLETION

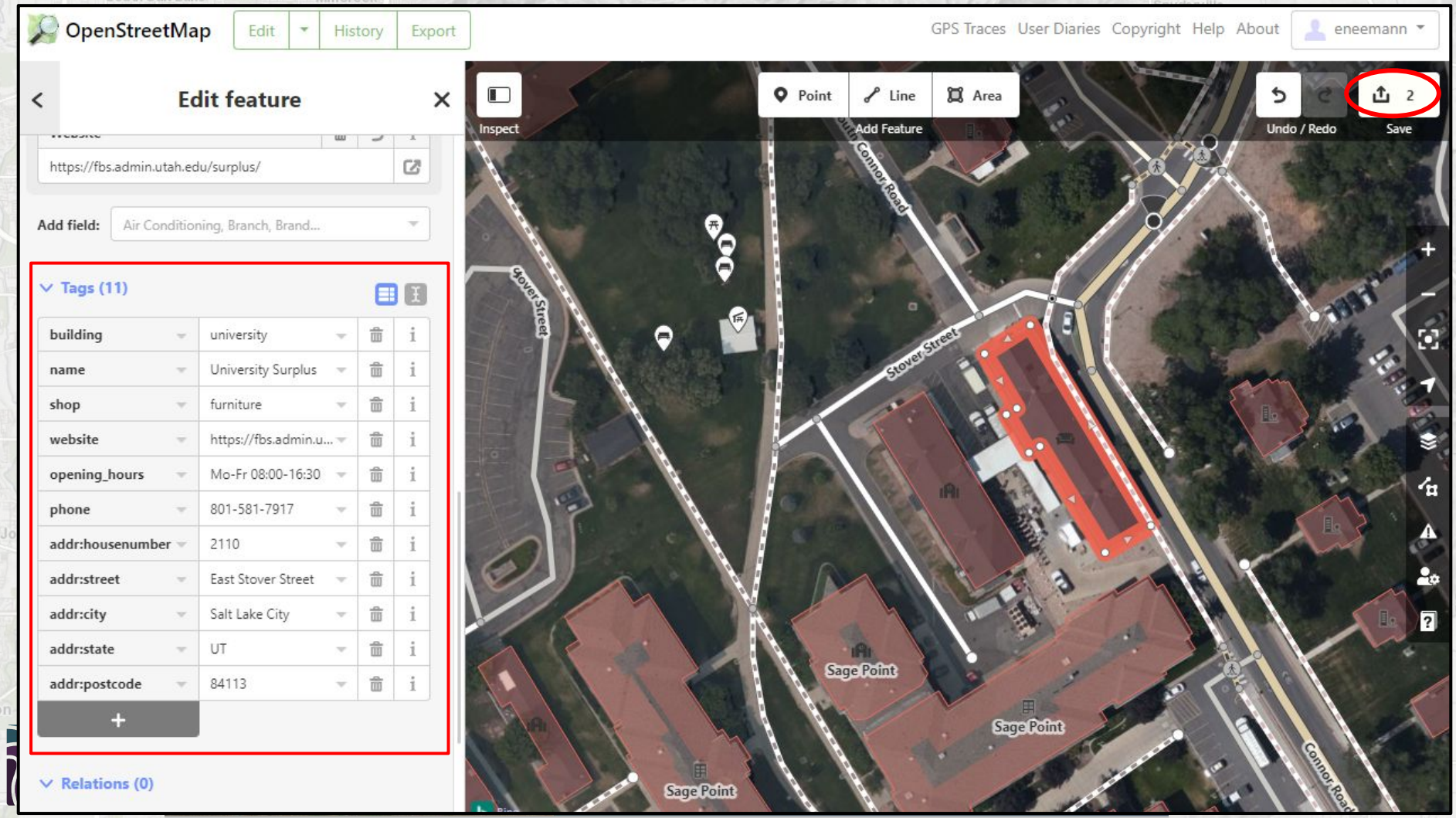
Add MR Tags

Current Editor: **Edit in iD (web editor)**

Edit | Not an Issue

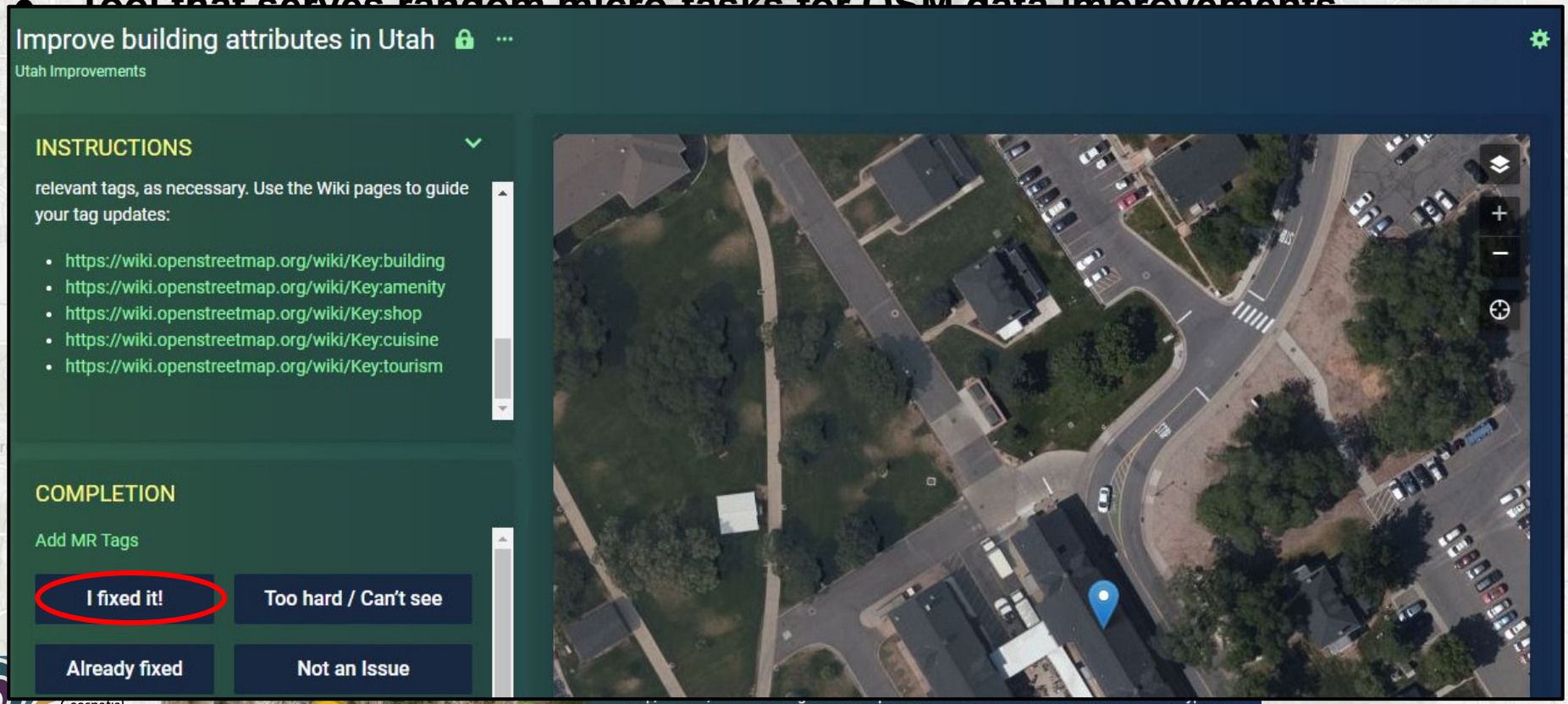
Skip | Other...

building it is.


OpenStreetMap
Edit
History
Export
GPS Traces
User Diaries
Copyright
Help
About
eneemann
Edit feature
https://fbs.admin.utah.edu/surplus/
Add field:
Air Conditioning, Branch, Brand...
Tags (11)
building
university
name
University Surplus
shop
furniture
website
https://fbs.admin.u...
opening_hours
Mo-Fr 08:00-16:30
phone
801-581-7917
addr:housenumber
2110
addr:street
East Stover Street
addr:city
Salt Lake City
addr:state
UT
addr:postcode
84113
Relations (0)
Inspect
Point
Line
Area
Add Feature
Undo / Redo
2
Save
Stover Street
Sage Point
Connor Road

How to participate? - Map Roulette!

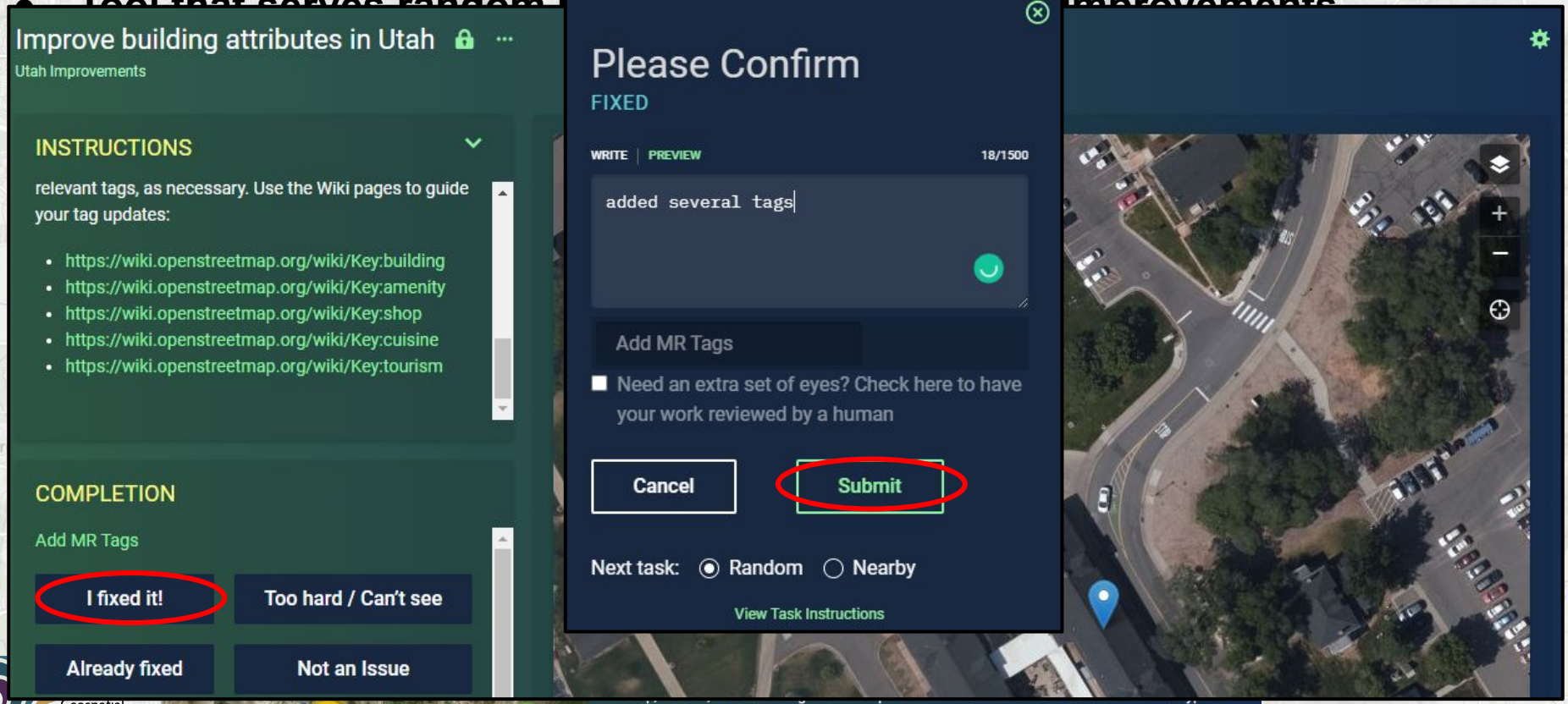
- Tool that serves random micro-tasks for OSM data improvements



Geospatial Resource Center

How to participate? - Map Roulette!

- Tool that serves random [illegible] improvements



Geospatial Resource Center

Final words

- Explore the data!
- Participate!
  - Create an OSM account
  - Make edits and updates to improve the OSM data
    - It will appear in UGRC's data within ~2 weeks
- Try the MapRoulette challenge!
- Have fun!

[Data Page]

[Blog Post]

Utah Geospatial Resource Center

Questions?

Erik Neemann

email: eneemann@utah.gov

twitter: @Erik_UGRC

gis.utah.gov/presentations

gis.utah.gov