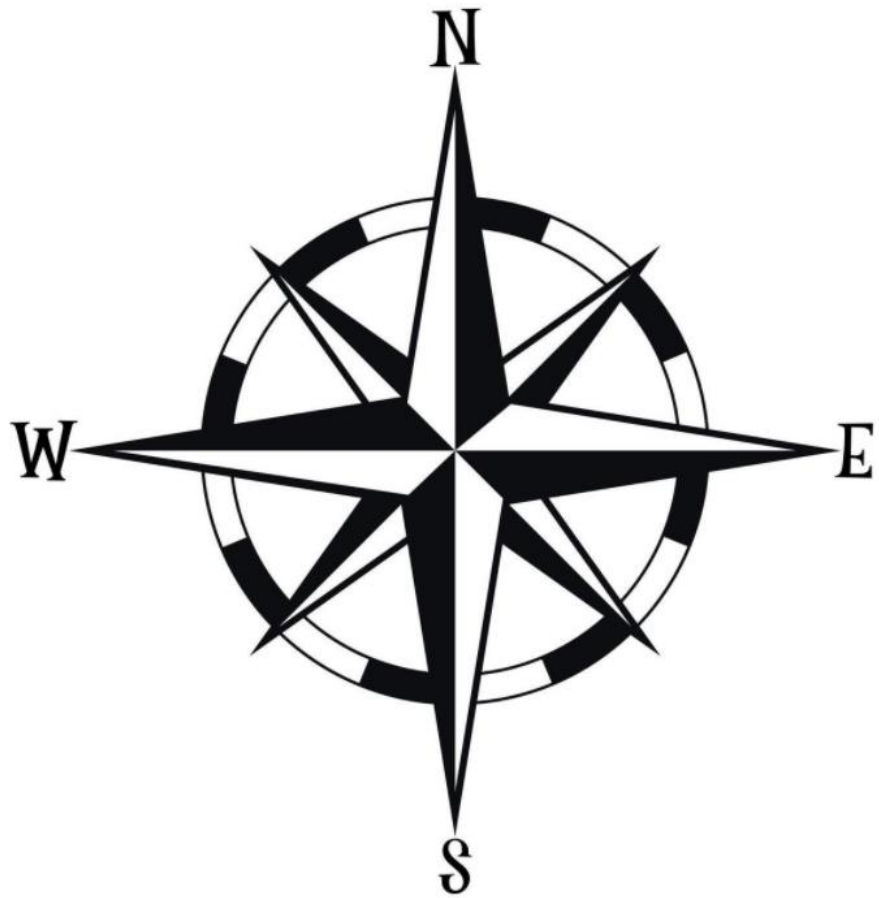
Data in Context

Wrangling Spatial Data with GIS



Sean Connin

02/24/21

https://etsy.me/37GTNCD

How Good is Your Data?

https://www.maproomblog.com/2020/01/xkcd-all-south-americas/

Strategic Wrangling

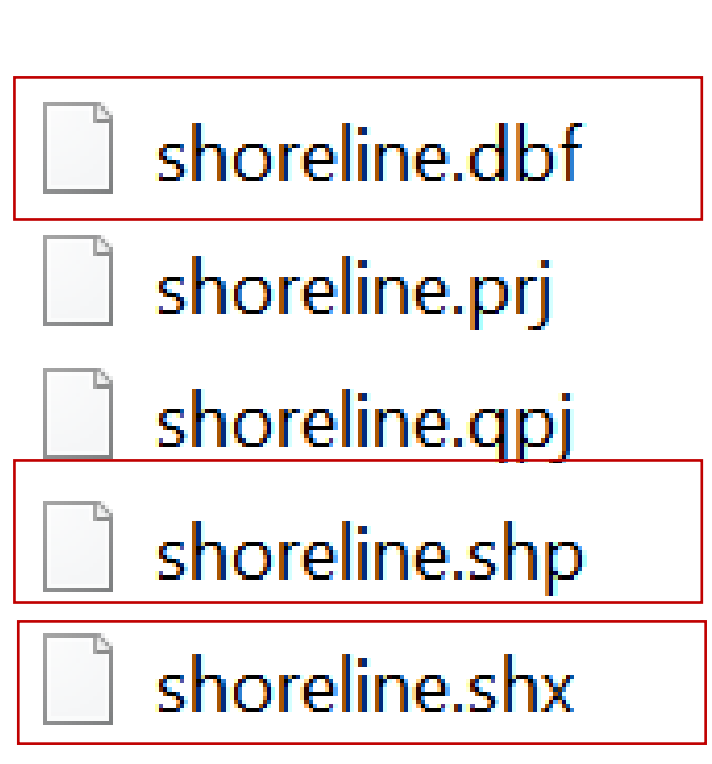
1. Data Formats
2. Data Problems
3. Cleanup Using QGIS

Coordinates

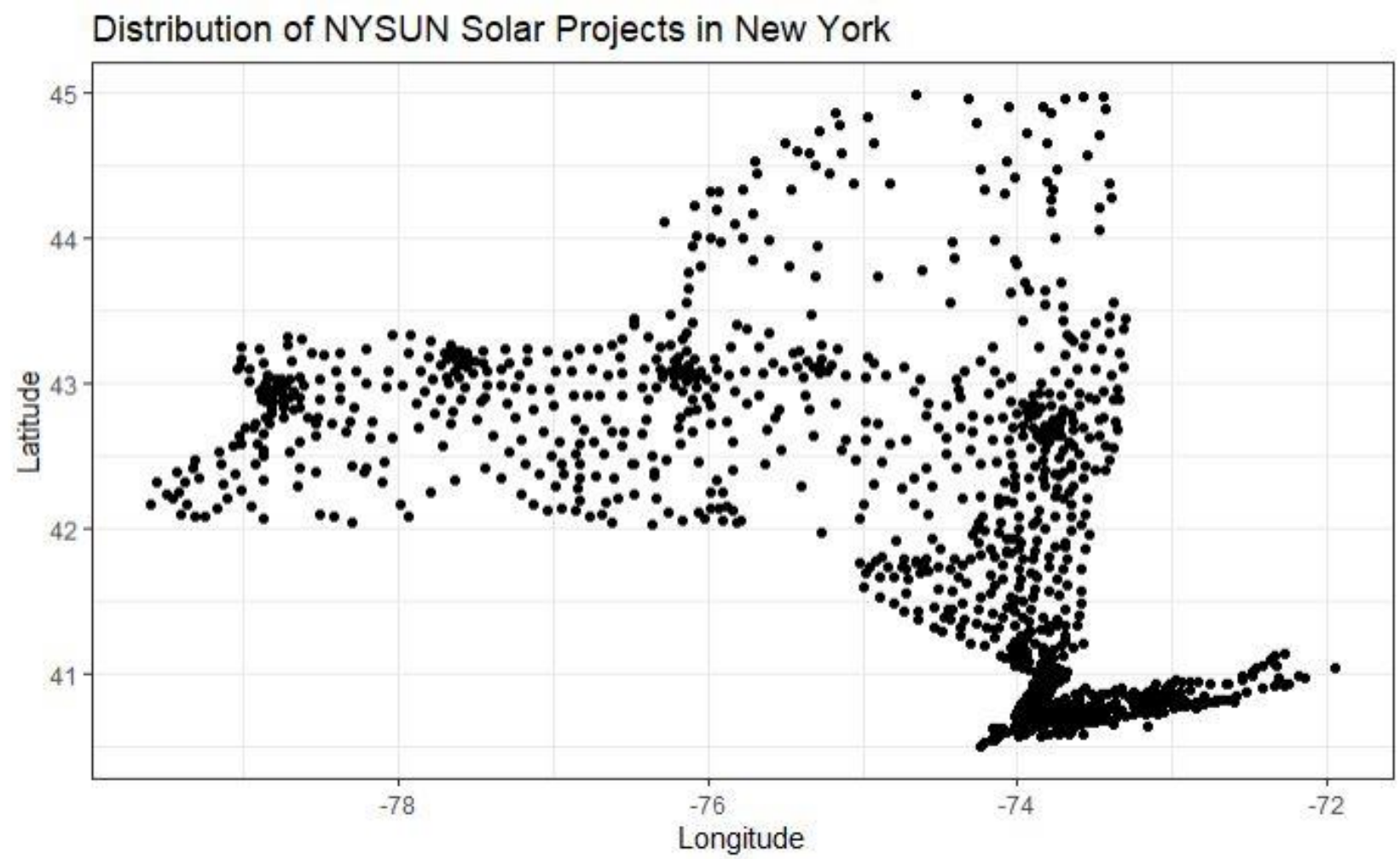
AD	AE
Latitude	Longitude
43.07772	-73.7592
43.0235	-74.3949
43.21472	-75.4547
44.32885	-75.4672
43.0235	-74.3949
42.85284	-73.7834
43.07772	-73.7592
43.03031	-74.6284
43.21472	-75.4547
44.73587	-75.2788
43.09785	-73.5
43.09785	-73.5
42.841	-73.8527

Shapefiles

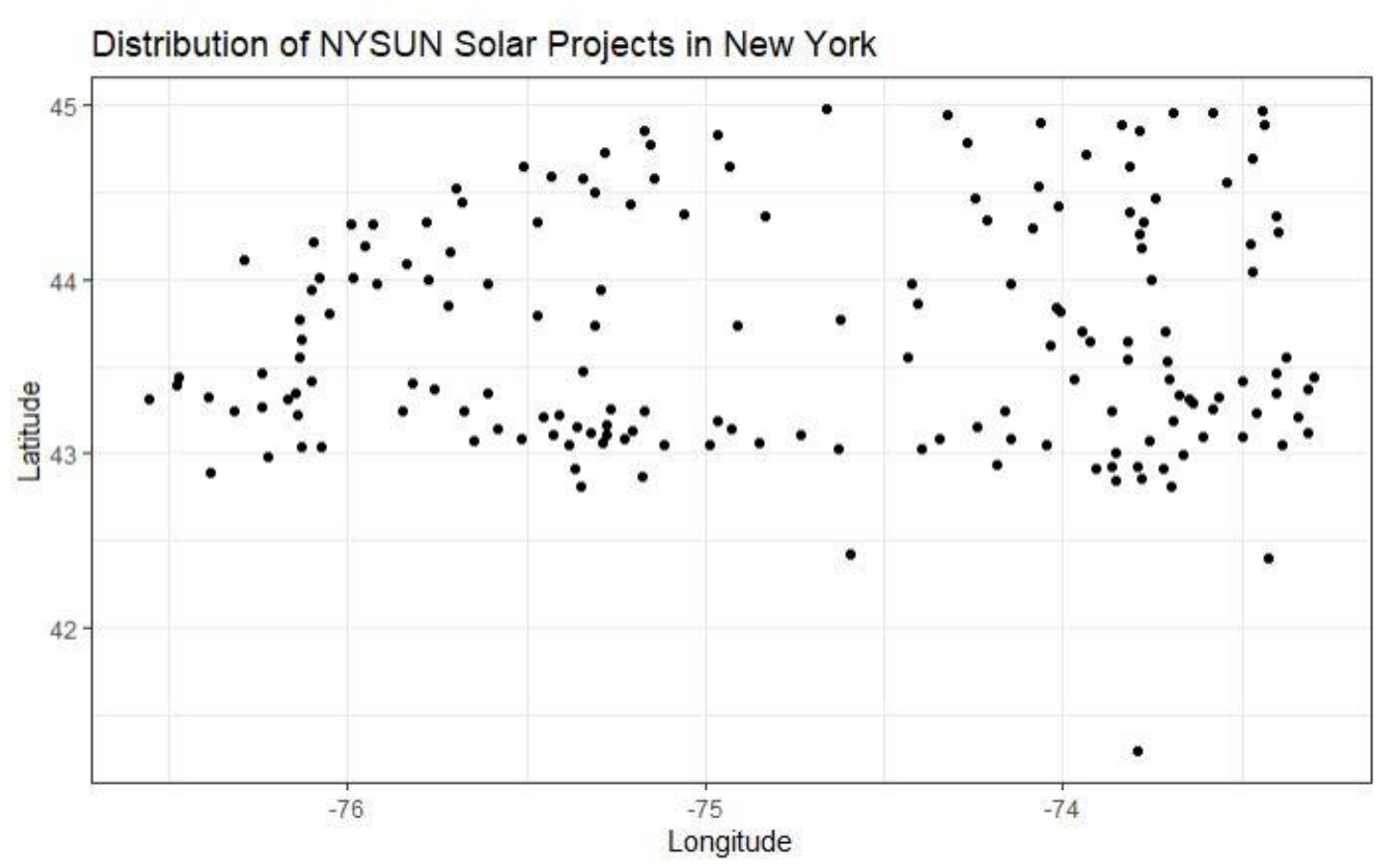
- Developed by ESRI
- Designed for vector data and attributes
- An open format for file transfer



```{r}

#read in awesome NYSUN solar data

state <- read.delim("https://raw.githubusercontent.com/sconnin/607DIC/main/state.csv",
 sep=",", stringsAsFactors = FALSE)%>%mutate_at( vars('Latitude', 'Longitude'), as.numeric)

 map <- ggplot() +
    geom_point(data = state, aes(x = Longitude, y = Latitude)) + # Add coordinate data
    theme_bw() + # Change the plot theme
    ggtitle("Distribution of NYSUN Solar Projects in New York") + # Give the plot a title
    xlab("Longitude") + # Change x axis title
    ylab("Latitude") # Change y axis title
    map


```



```{r}
# save selected county data as csv

adks<-state%>%filter(County %in% c("Clinton","Essex","Franklin","Fulton",
  "Jefferson", "Hamilton", "Herkimer","Lewis","Oneida","Oswego","Saratoga",
  'St. Lawrence","Washington","Warren"), na.rm=TRUE)

adks%>%kbl()%>%kable_material(c("striped"))

 map <- ggplot() +
    geom_point(data = adks, aes(x = Longitude, y = Latitude)) + # Add coordinate data
    theme_bw() + # Change the plot theme
    ggtitle("Distribution of NYSUN Solar Projects in New York") + # Give the plot a title
    xlab("Longitude") + # Change x axis title
    ylab("Latitude") # Change y axis title
    map

write_csv(adks, path="C:\\Users\\seanc\\Desktop\\DIC\\adks.csv")

```

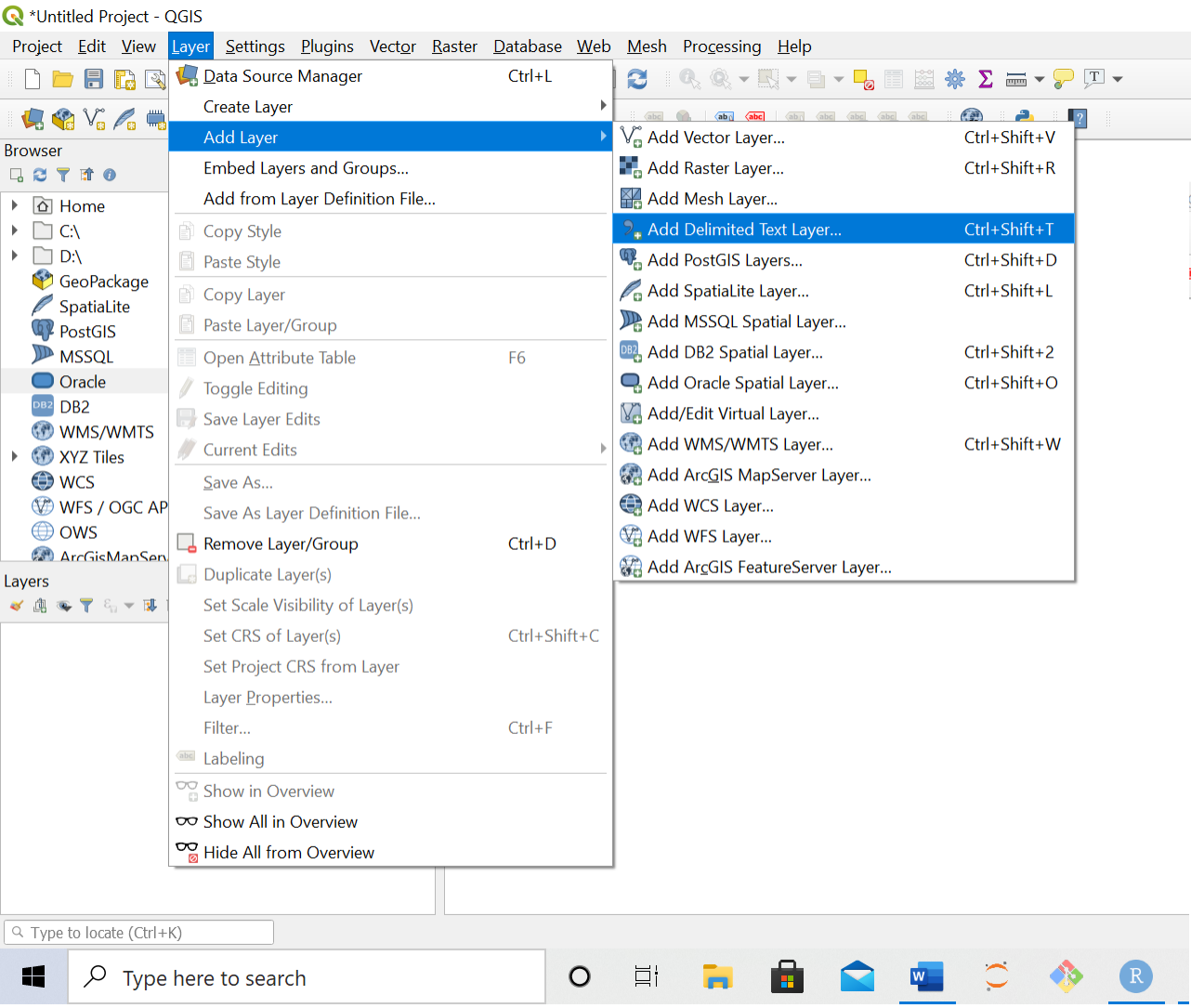


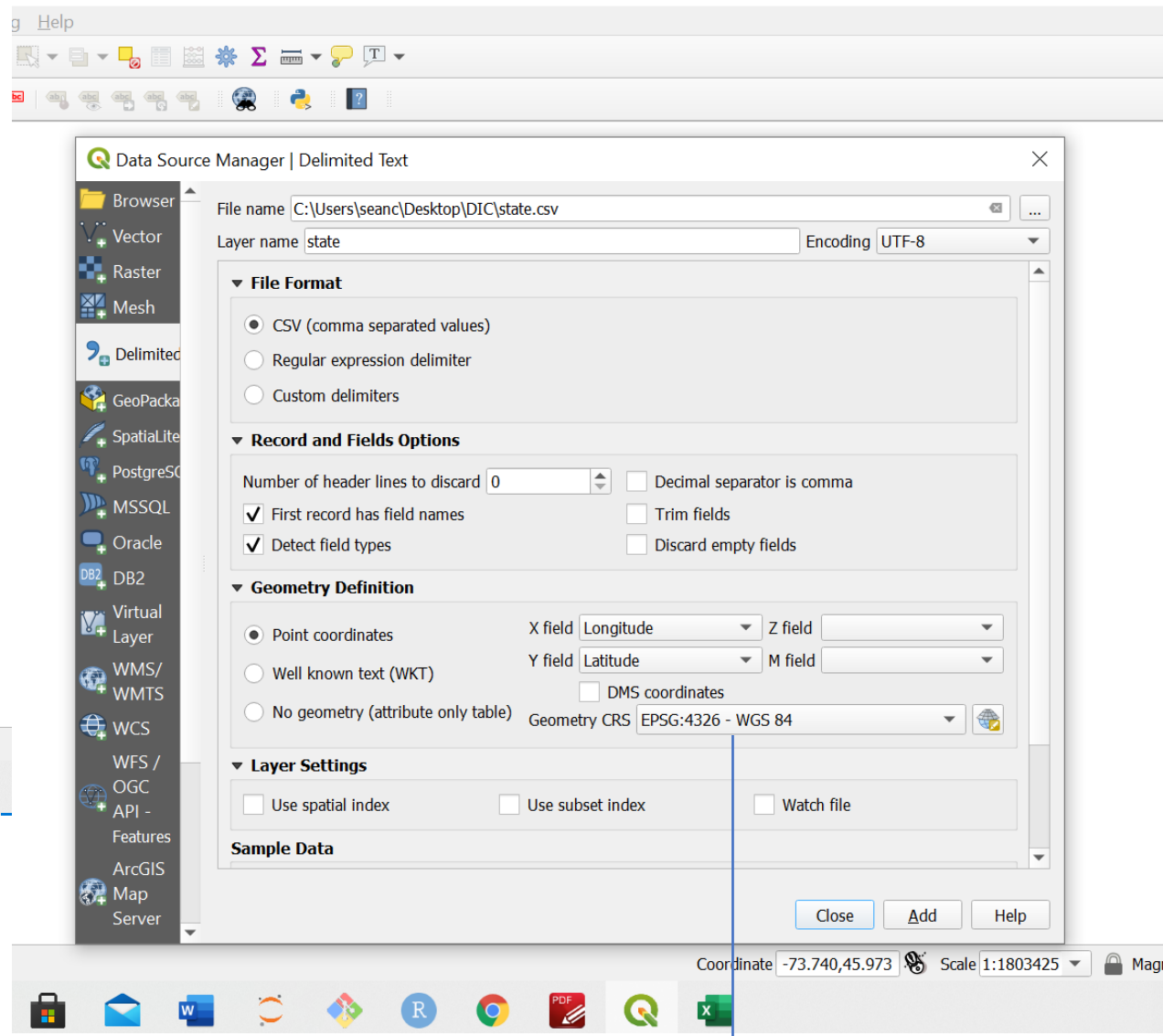
City	County	State	Zip.Code	Sector	Latitude	Longitude
Peru	Clinton	NY	12972	Non-Residential	44.56058	-73.54486
Saranac	Clinton	NY	12981	Non-Residential	44.64703	-73.81446
Mooers	Clinton	NY	12958	Non-Residential	44.96220	-73.58000
Plattsburgh	Clinton	NY	12901	Non-Residential	44.70168	-73.46871
Chazy	Clinton	NY	12921	Non-Residential	44.88821	-73.43534
Mooers Forks	Clinton	NY	12959	Non-Residential	44.96021	-73.69187
Mooers Forks	Clinton	NY	12959	Non-Residential	44.96021	-73.69187
Plattsburgh	Clinton	NY	12901	Non-Residential	44.70168	-73.46871

QGIS

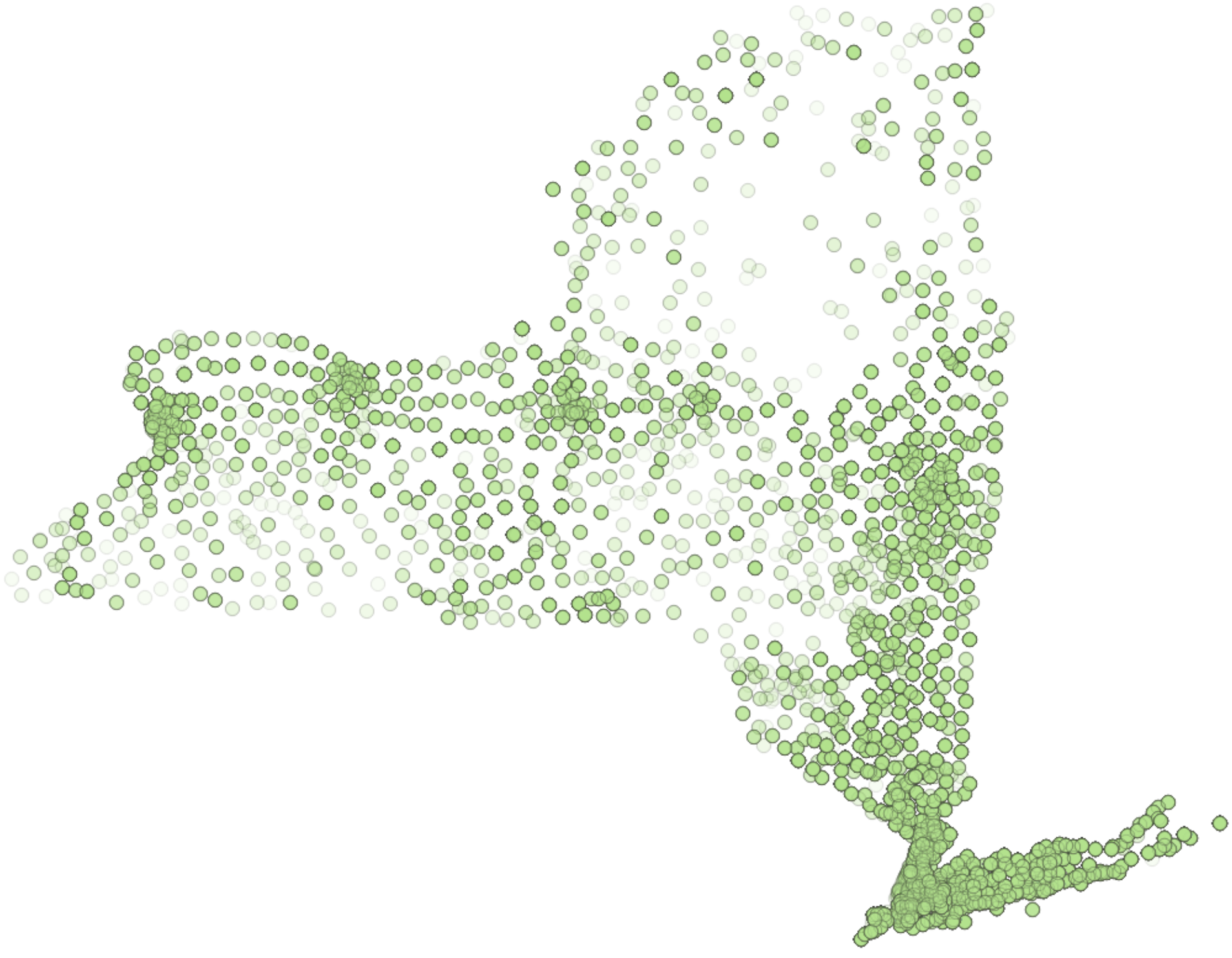
A Free and Open Source Geographic Information System

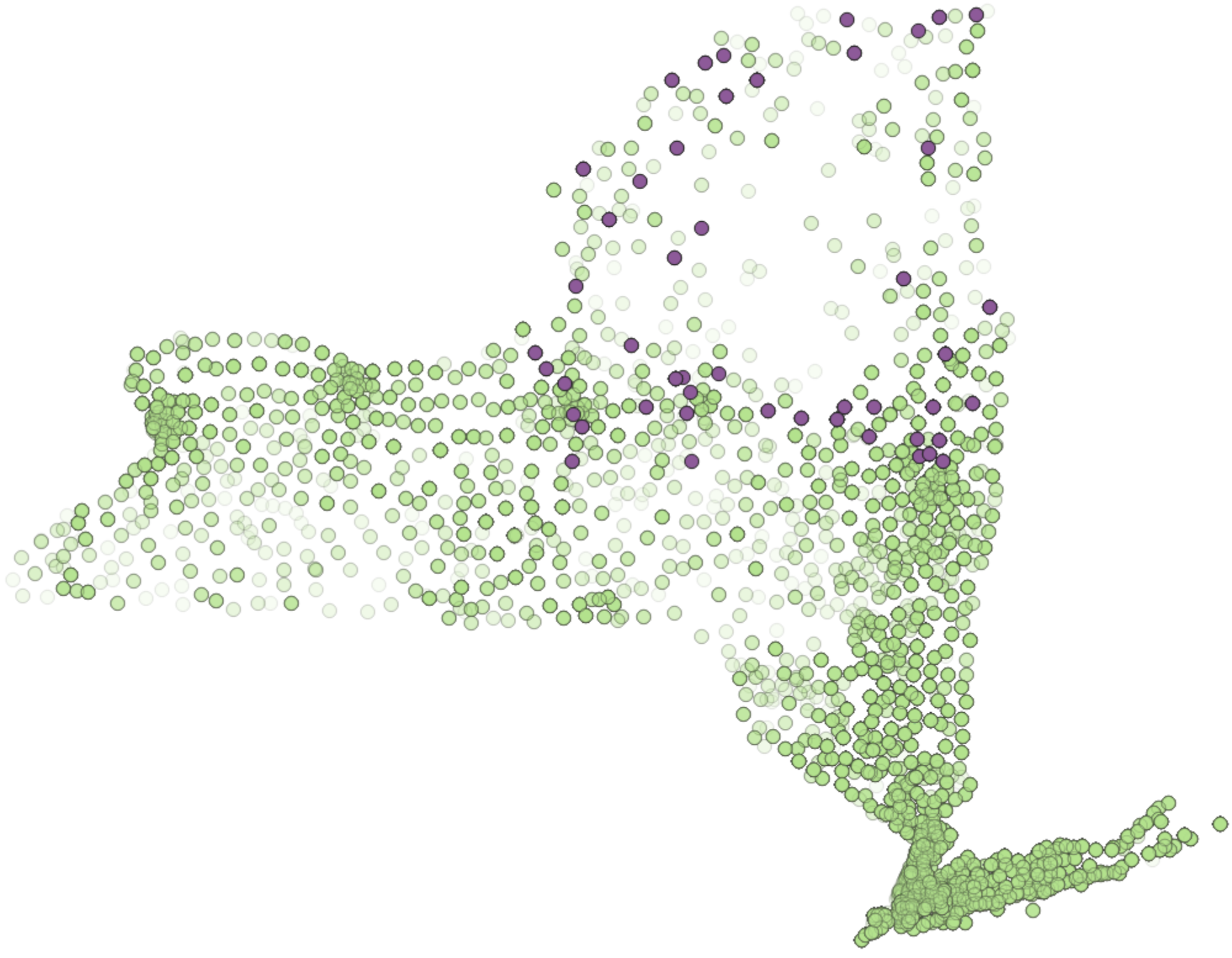
https://www.qgis.org/en/site/

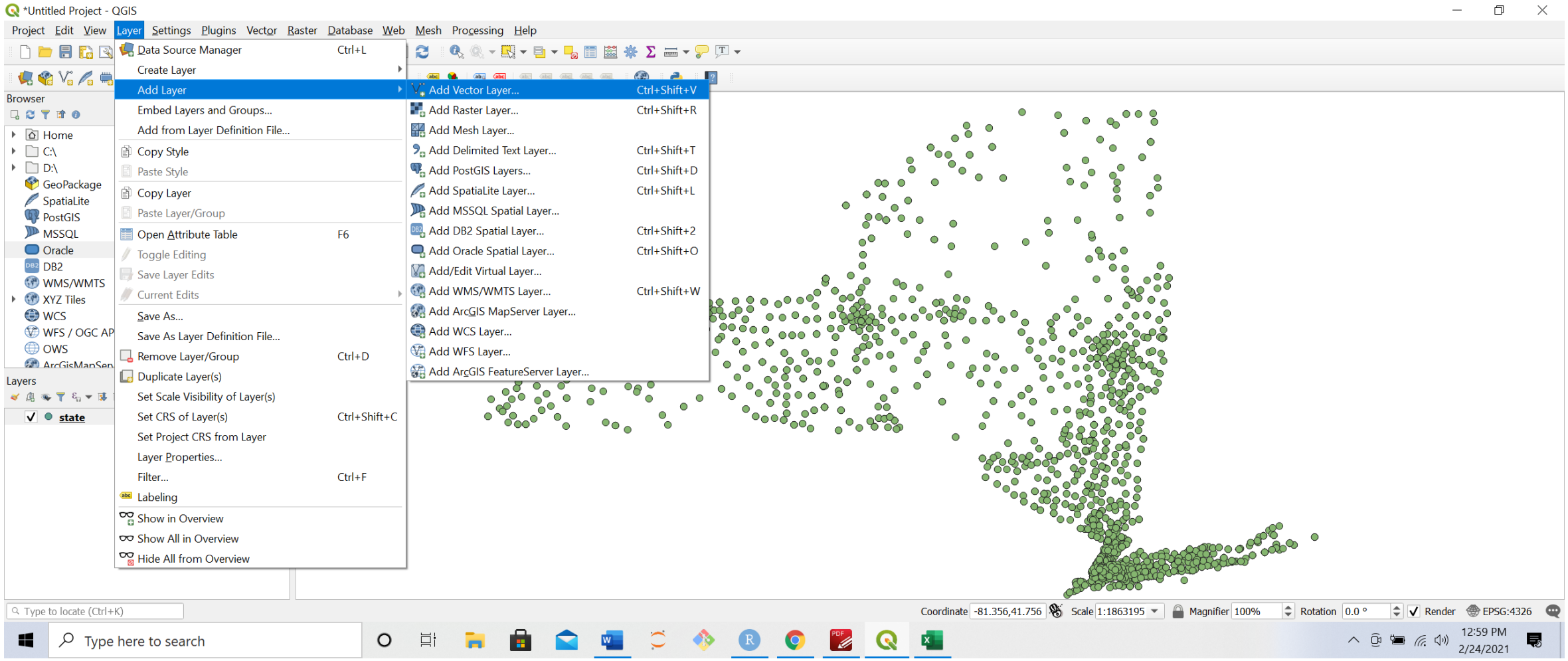


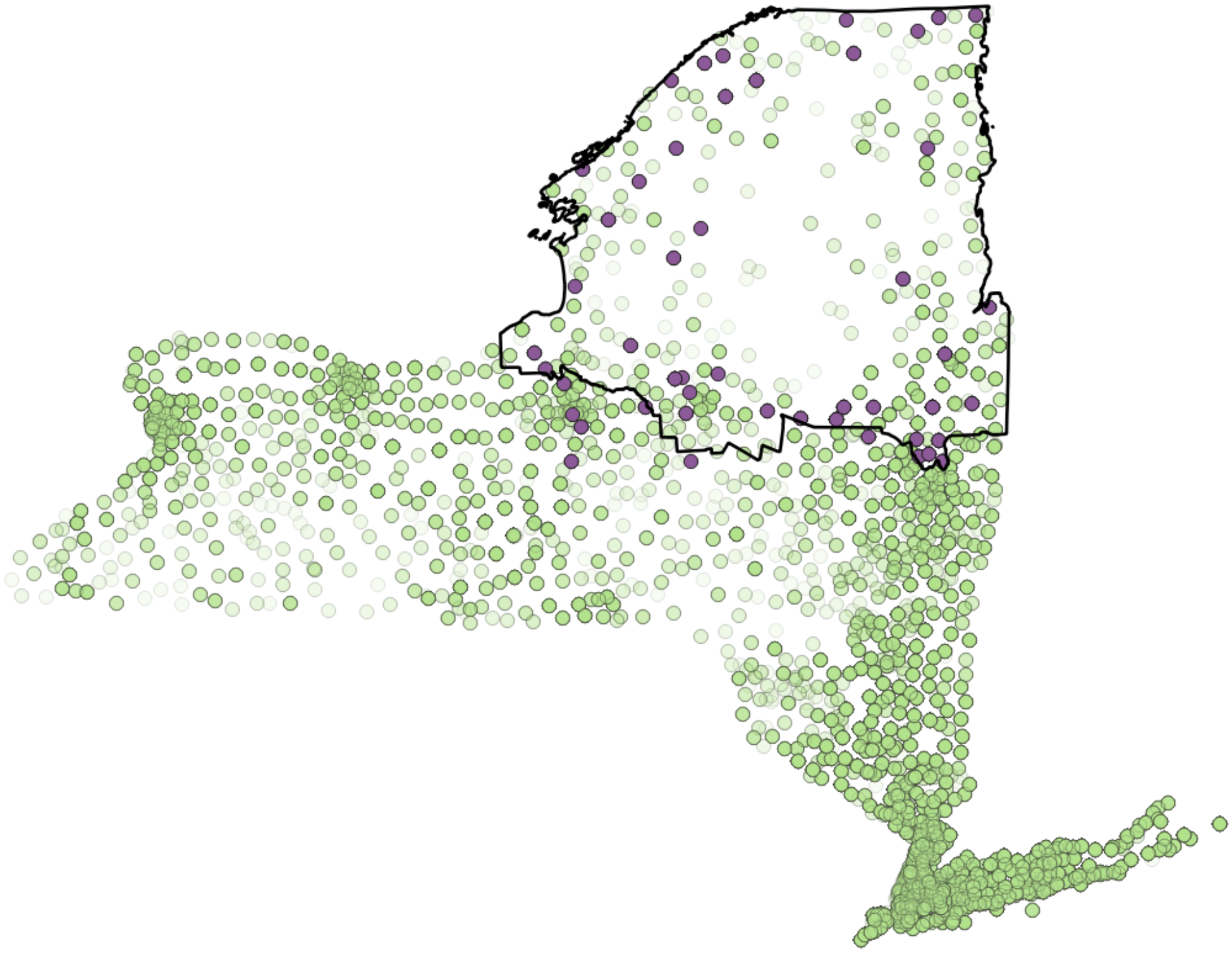


Coordinate Reference System	Authority ID
WGS 84	EPSG:4326






*Untitled Project - QGIS
Project Edit View Layer Settings Plugins Vector Raster Database Web Mesh Processing Help
Data Source Manager Ctrl+L
Create Layer
Add Layer
Embed Layers and Groups...
Add from Layer Definition File...
Copy Style
Paste Style
Copy Layer
Paste Layer/Group
Open Attribute Table F6
Toggle Editing
Save Layer Edits
Current Edits
Save As...
Save As Layer Definition File...
Remove Layer/Group Ctrl+D
Duplicate Layer(s)
Set Scale Visibility of Layer(s)
Set CRS of Layer(s) Ctrl+Shift+C
Set Project CRS from Layer
Layer Properties...
Filter... Ctrl+F
Labeling
Show in Overview
Show All in Overview
Hide All from Overview
Add Vector Layer... Ctrl+Shift+V
Add Raster Layer... Ctrl+Shift+R
Add Mesh Layer...
Add Delimited Text Layer... Ctrl+Shift+T
Add PostGIS Layers... Ctrl+Shift+D
Add SpatiaLite Layer... Ctrl+Shift+L
Add MSSQL Spatial Layer...
Add DB2 Spatial Layer... Ctrl+Shift+2
Add Oracle Spatial Layer... Ctrl+Shift+O
Add/Edit Virtual Layer...
Add WMS/WMTS Layer... Ctrl+Shift+W
Add ArcGIS MapServer Layer...
Add WCS Layer...
Add WFS Layer...
Add ArcGIS FeatureServer Layer...
Browser
Home
C:\
D:\
GeoPackage
SpatiaLite
PostGIS
MSSQL
Oracle
DB2
WMS/WMTS
XYZ Tiles
WCS
WFS / OGC AP
OWS
Layers
state
Type to locate (Ctrl+K)
Coordinate -81.356,41.756
Scale 1:1863195
Magnifier 100%
Rotation 0.0 °
Render
EPSG:4326
Type here to search
12:59 PM
2/24/2021



DIC - QGIS

Project Edit View Layer Settings Plugins Vector Raster Database Web Mesh Processing Help

Geoprocessing Tools
Geometry Tools
Analysis Tools
Research Tools
Data Management Tools

Buffer...
Clip...
Convex Hull...
Difference...
Dissolve...
Intersection...
Symmetrical Difference...
Union...
Eliminate Selected Polygons...

Browser

Favorites
Spatial Bookmarks
Project Home
Home
C:\
D:\
GeoPackage
SpatiaLite
PostGIS
MSSQL
Oracle
DB2
WMS/WMTS
XYZ Tiles
WCS

Layers

- Intersect for 14 Counties
- ANCA 14 Counties Shoreline Dissolved - SP
- 14 County Commercial Projects
- NY Solar Queue

Type to locate (Ctrl+K)

Type here to search

Intersection

Parameters Log

Input layer

Intersect for 14 Counties [EPSG:4326]

Selected features only

Overlay layer

ANCA 14 Counties Shoreline Dissolved - SP

Selected features only

Input fields to keep (leave empty to keep all fields)

Intersection

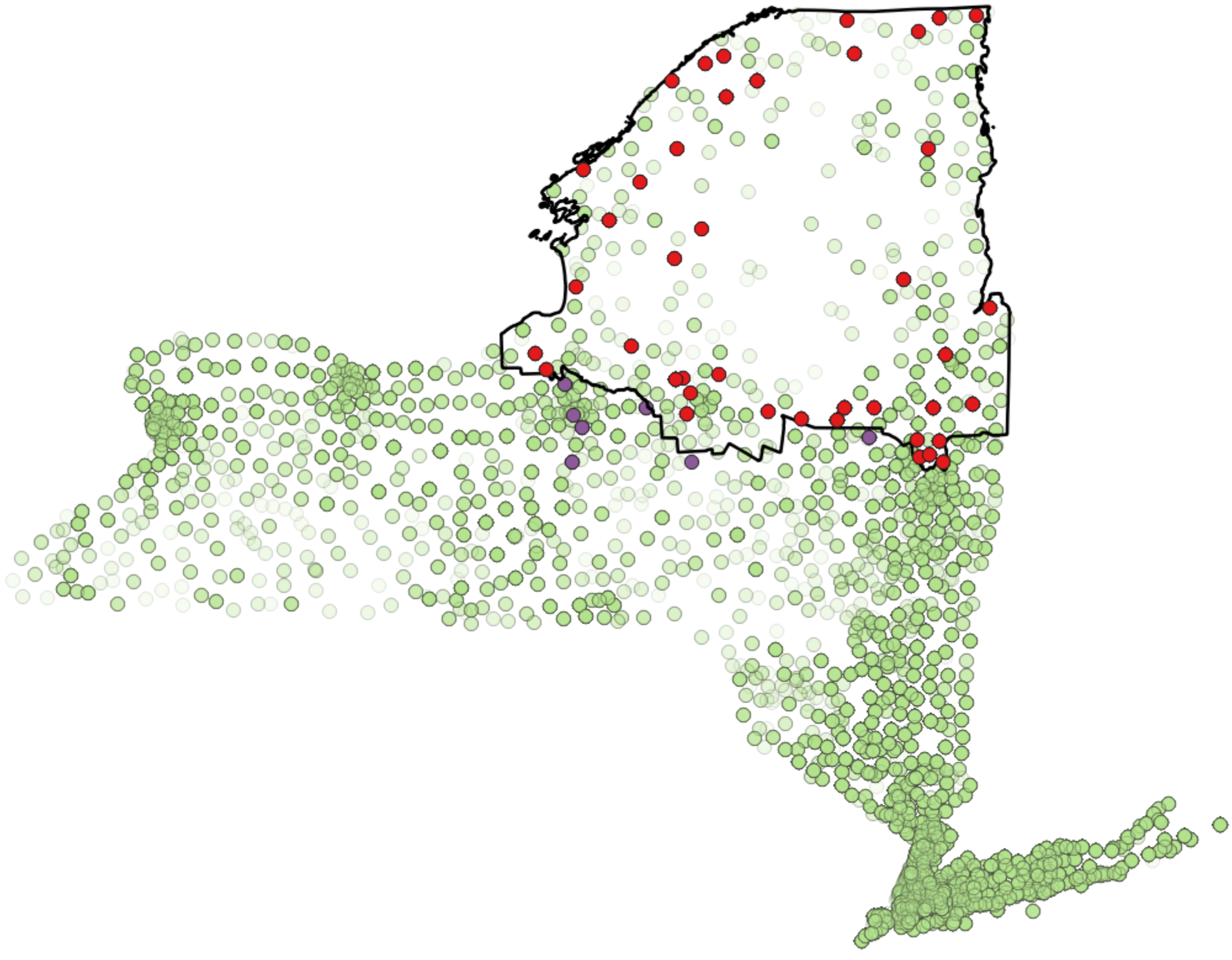
This algorithm extracts the overlapping portions of features in the Input and Overlay layers. Features in the output Intersection layer are assigned the attributes of the overlapping features from both the Input and Overlay layers.

0%

Cancel

Run as Batch Process...

Run Close Help



*DIC - QGIS

Project Edit View Layer Settings Plugins Vector Raster Database Web Mesh Processing Help

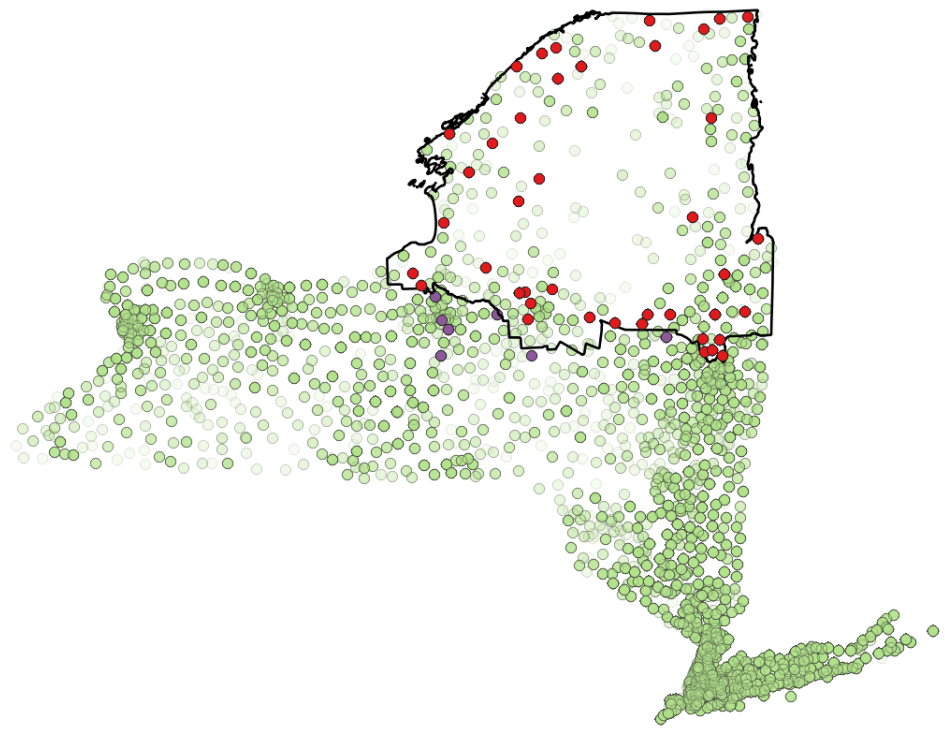
Browser

Favorites
Spatial Book
Project Hom
Home
C:\
D:\
GeoPackage
SpatiaLite
PostGIS
MSSQL
Oracle
DB2
WMS/WMTS
XYZ Tiles
WCS

Zoom to Layer
Zoom to Selection
Show in Overview
Show Feature Count
Copy Layer
Rename Layer
Duplicate Layer
Remove Layer...
Open Attribute Table
Toggle Editing
Filter...
Change Data Source...
Set Layer Scale Visibility...
Set CRS
Make Permanent...
Export
Styles
Properties...

Layers

Intersection
ANCA 14 Counties Shoreline Dissolved - SP
14 County Commercial Projects
NY Solar Queue



Type to locate (Ctrl+K) 1 legend entries removed. Coordinate -1749476,1617059 Scale 1:3747363 Magnifier 100% Rotation 0.0 ° Render ESRI:102715

Type here to search

Intersection :: Features Total: 92, Filtered: 92, Selecte...

	Reporting Period	Project Number	City	County	
1	03/31/19	31275A-1	Greenfield	Saratoga	NY
2	03/31/19	49762A-1	Johnstown	Fulton	NY
3	03/31/19	33493A-1	Rome	Oneida	NY
4	03/31/19	50970-1	Gouverneur	St. Lawrence	NY

Show All Features

*DIC - QGIS

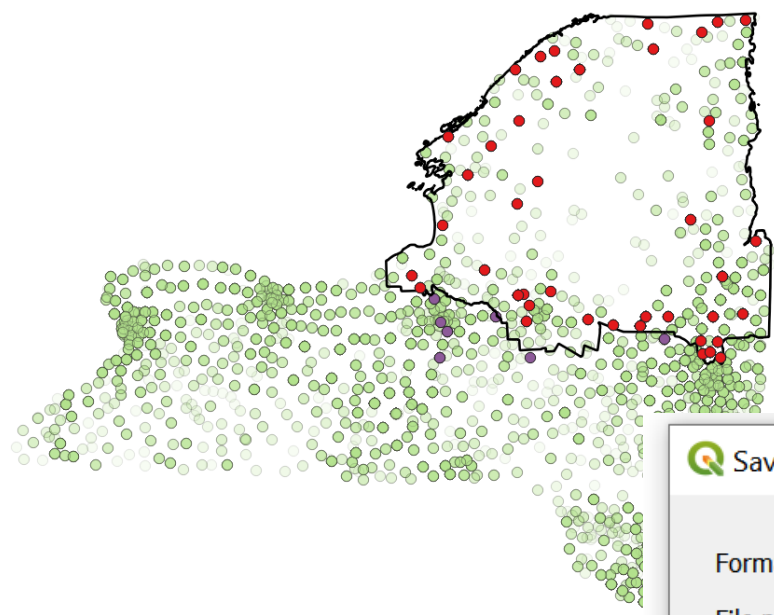
Project Edit View Layer Settings Plugins Vector Raster Database Web Mesh Processing Help

Browser

- Favorites
- Spatial Bookr
- Project Home
- Home
- C:\
- D:\
- GeoPackage
- SpatiaLite
- PostGIS
- MSSQL
- Oracle
- DB2
- WMS/WMTS
- XYZ Tiles
- WCS

Layers

- Intersection
- ANCA 14 Counties Shoreline Dissolved - SP
- 14 County Commercial Projects
- NY Solar Queue



Zoom to Layer
Zoom to Selection
Show in Overview
Show Feature Count
Copy Layer
Rename Layer
Duplicate Layer
Remove Layer...
Open Attribute Table
Toggle Editing
Filter...
Change Data Source...
Set Layer Scale Visibility...
Set CRS
Make Permanent...
Export
Styles
Properties...

Save Features As...
Save Selected Features As...
Save as Layer Definition File...
Save as QGIS Layer Style File...

Type to locate (Ctrl+K)
1 legend entries removed.
Coordinate

Type here to search

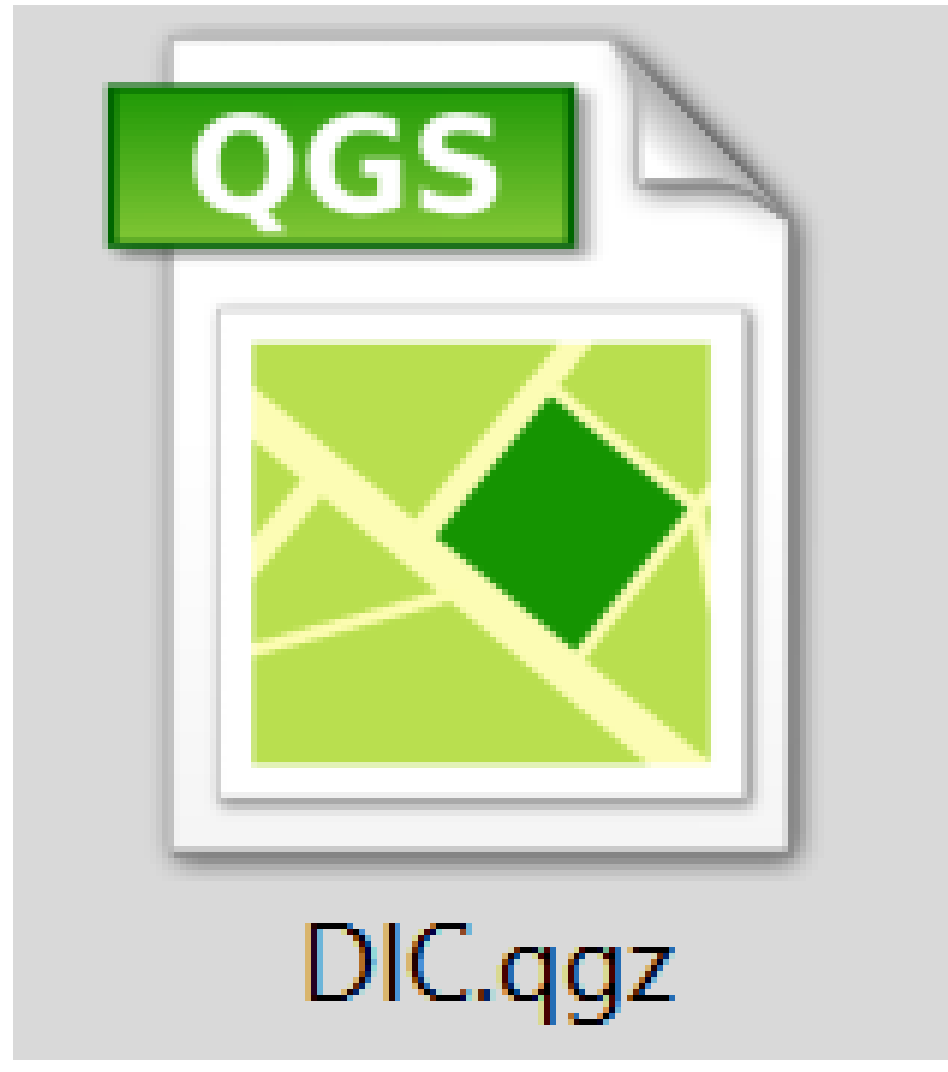
Save Vector Layer as...

Format: Comma Separated Value [CSV]
File name: C:\Users\seanc\Desktop\DIC\clean_adks.csv
Layer name:
CRS: EPSG:4326 - WGS 84

Encoding: UTF-8
Save only selected features
Select fields to export and their export options
Geometry
Geometry type: Automatic
Force multi-type
Include z-dimension
Extent (current: layer)
Layer Options
CREATE_CSVT: NO

Add saved file to map
OK
Cancel
Help

QGIS - Projects



Simple Features for R

A package that provides simple features access for R. Package sf:

- represents simple features as records in a `data.frame` or `tibble` with a geometry list-column
- represents natively in R all 17 simple feature types for all dimensions (XY, XYZ, XYM, XYZM)
- interfaces to GEOS to support geometrical operations including the DE9-IM
- interfaces to GDAL, supporting all driver options, `Date` and `POSIXct` and list-columns
- interfaces to PRØJ for coordinate reference system conversions and transformations
- uses well-known-binary serialisations written in C++/Rcpp for fast I/O with GDAL and GEOS
- reads from and writes to spatial databases such as PostGIS using DBI
- is extended by pkg lwgeom for further liblwgeom/PostGIS functions, including some spherical geometry functions

https://r-spatial.github.io/sf/