

API-231 / GIS-PubPol

Meeting 17 (Lab Exercise + Walk-Through 1)

Yuri M. Zhukov
Visiting Associate Professor of Public Policy
Harvard Kennedy School

March 28, 2024

Goal: explain geographic variation in Islamic State violence

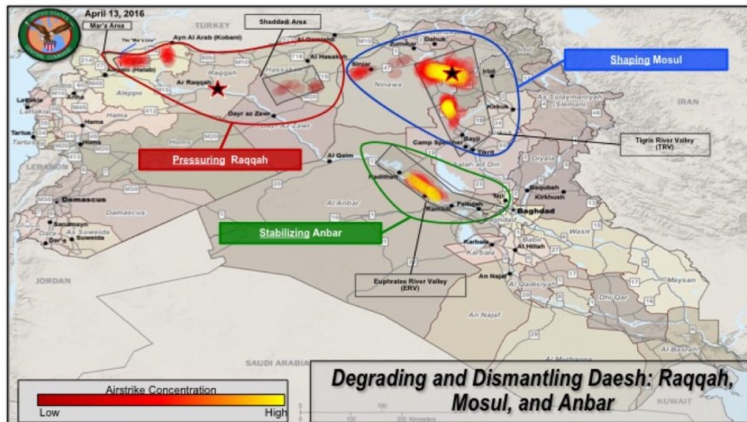


Figure 1: Operation Inherent Resolve

There is no formal Lab Exercise this week, but if you would like to replicate this in QGIS, follow along (for walk-through in R, see replication code on Canvas)

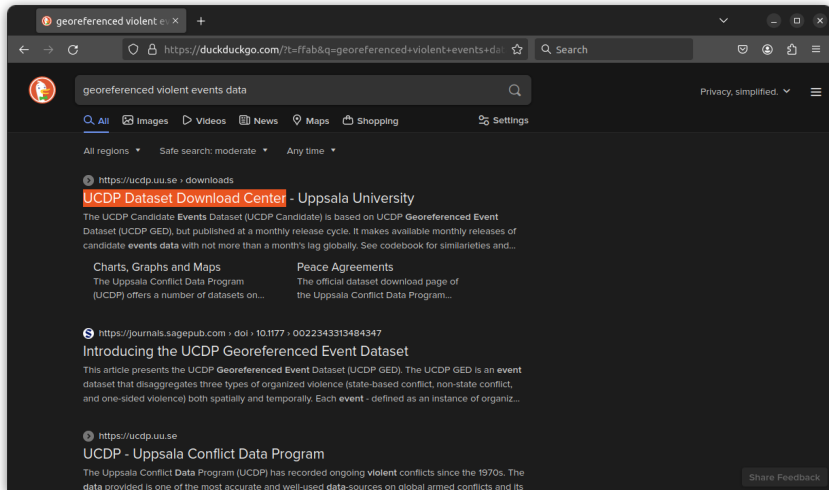
Overview of lab exercise

1. Data
 - a) Collect raw data from open sources online
 - b) Pre-processing to integrate data to common set of spatial units
2. Analysis
 - a) Export pre-processed data as comma-separated file
 - b) Run regression models in R

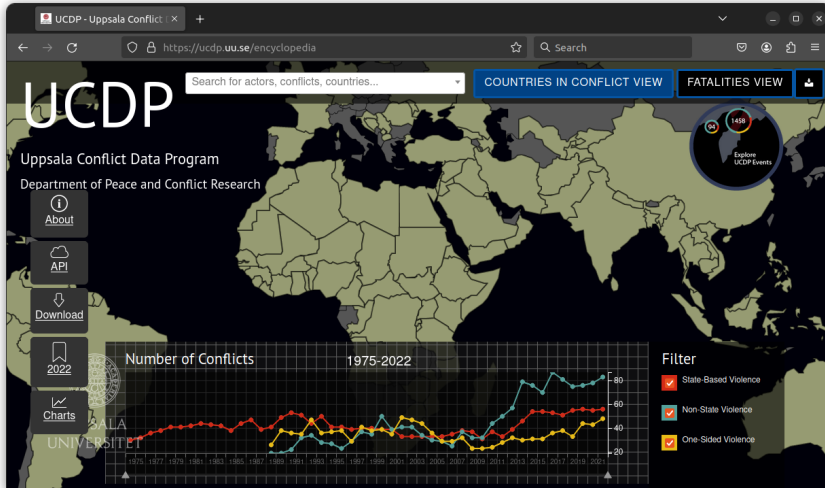
Data

Collection

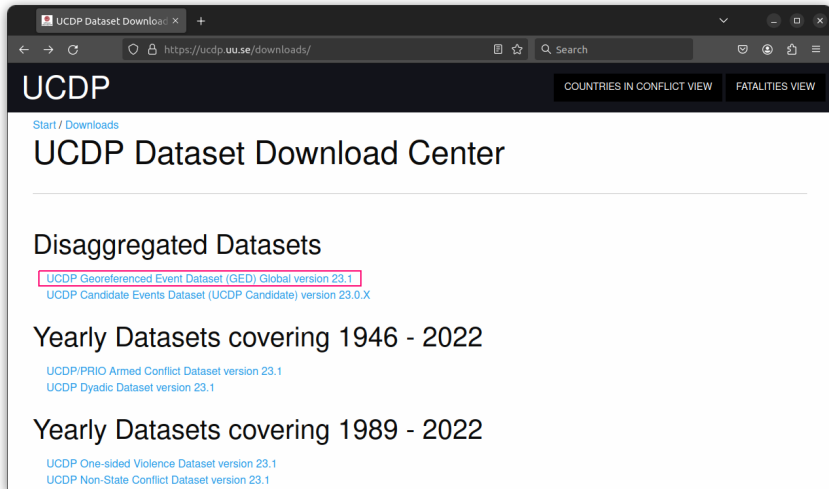
Data on **political violence** are easily found online. If we search for georeferenced violent events data, a top result is for “UCDP Dataset Download Center”



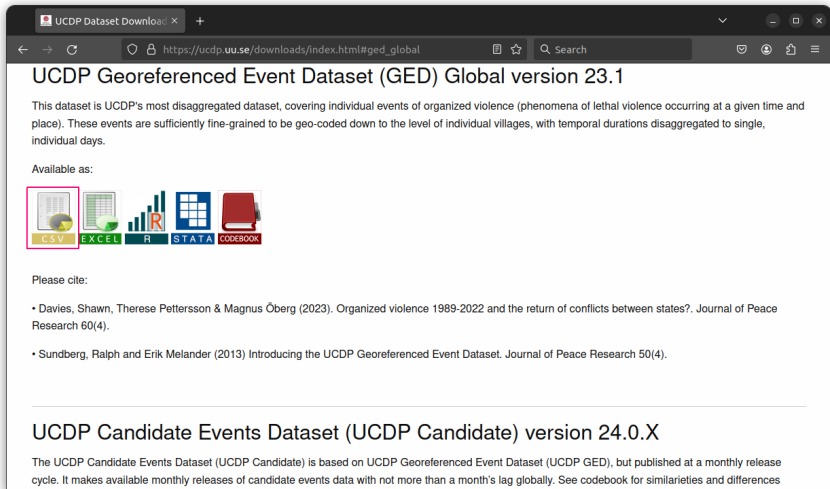
The Uppsala Conflict Data Program (UCDP) is one of the world's leading providers of open-source data on violence and armed conflict, based at Uppsala University



Downloadable data are located at ucdp.uu.se/downloads.
Click on “Georeferenced Event Dataset (GED)”



We have used GED before, for Problem Set 02 (Afghanistan and Yemen).
Download GED as a csv file and extract it from the zip archive



The screenshot shows a web browser window with the address bar displaying https://ucdp.uu.se/downloads/index.html#ged_global. The page title is "UCDP Georeferenced Event Dataset (GED) Global version 23.1". The main text describes the dataset as UCDP's most disaggregated dataset, covering individual events of organized violence, geo-coded down to the level of individual villages, with temporal durations disaggregated to single, individual days.

Available as:

Below this text, there are five icons representing different file formats: CSV, EXCEL, R, STATA, and CODEBOOK. The CSV icon is highlighted with a red box.

Please cite:

- Davies, Shawn, Therese Pettersson & Magnus Öberg (2023). Organized violence 1989-2022 and the return of conflicts between states?. Journal of Peace Research 60(4).
- Sundberg, Ralph and Erik Melander (2013) Introducing the UCDP Georeferenced Event Dataset. Journal of Peace Research 50(4).

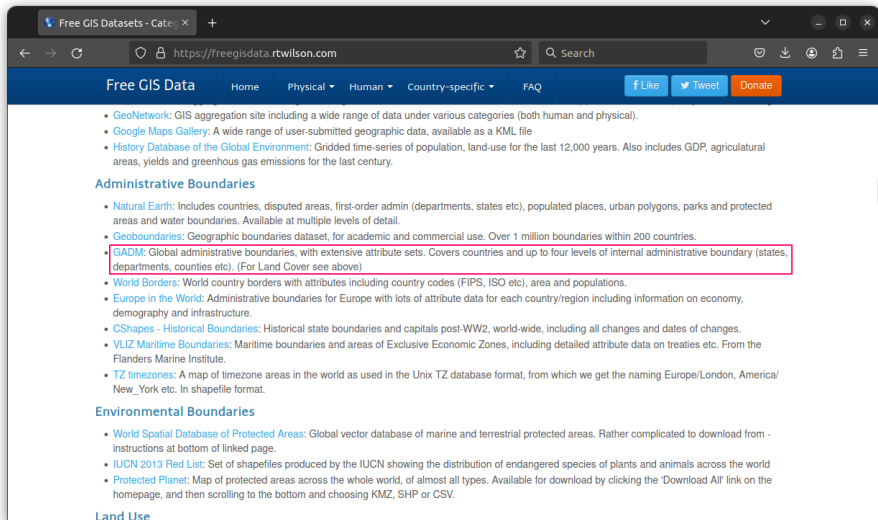
UCDP Candidate Events Dataset (UCDP Candidate) version 24.0.X

The UCDP Candidate Events Dataset (UCDP Candidate) is based on UCDP Georeferenced Event Dataset (UCDP GED), but published at a monthly release cycle. It makes available monthly releases of candidate events data with not more than a month's lag globally. See codebook for similarities and differences

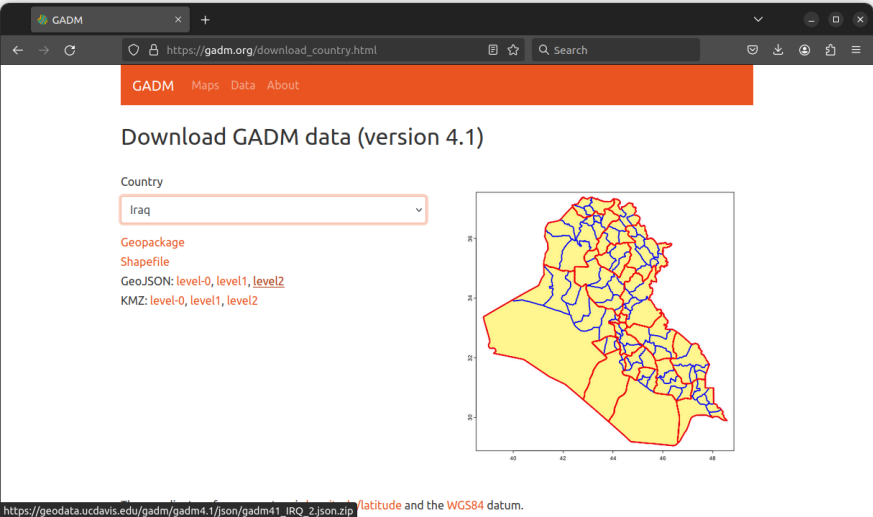
There are about 317K events here, each of which has point coordinates

	AB	AC	AD	AE	AF
			latitude	longitude	geom_wkt
1	adm_1	adm_2			
2	Kabul province	Kabul district	34.531094	69.162796	POINT (69.162796 34.531094)
3	Kabul province	Kabul district	34.564444	69.217222	POINT (69.217222 34.564444)
4	Nangarhar province	Jalalabad district	34.428844	70.45575	POINT (70.45575 34.428844)
5	Kabul province	Kabul district	34.531094	69.162796	POINT (69.162796 34.531094)
6	Nangarhar province	Kabul district	34.33333	70.41667	POINT (70.41667 34.33333)
7	Kunduz province		36.75	68.75	POINT (68.75 36.75)
8	Baghlan province	Khinjan district	35.315833	69.038889	POINT (69.038889 35.315833)
9	Kabul province	Kabul district	34.531094	69.162796	POINT (69.162796 34.531094)
10	Baghlan province	Khinjan district	35.31603	69.03871	POINT (69.03871 35.31603)
11	Herat province		34.5	62	POINT (62 34.5)
12	Kabul province	Kabul district	34.531094	69.162796	POINT (69.162796 34.531094)
13	Baghlan province	Khinjan district	35.315833	69.038889	POINT (69.038889 35.315833)
14	Kabul province	Kabul district	34.531094	69.162796	POINT (69.162796 34.531094)
15	Kabul province	Kabul district	34.531094	69.162796	POINT (69.162796 34.531094)
16	Kabul province	Kabul district	34.531094	69.162796	POINT (69.162796 34.531094)
17	Parwan province		35	68.91667	POINT (68.91667 35)
18	Kabul province	Kabul district	34.531094	69.162796	POINT (69.162796 34.531094)
19			33	65	POINT (65 33)
20	Kabul province	Kabul district	34.531094	69.162796	POINT (69.162796 34.531094)
21	Kabul province	Dih Sabz district	34.564444	69.217222	POINT (69.217222 34.564444)
22	Parwan province	Jabalussarai district	35.118366	69.235695	POINT (69.235695 35.118366)
23	Kabul province	Kabul district	34.531094	69.162796	POINT (69.162796 34.531094)
24	Kabul province	Kabul district	34.531094	69.162796	POINT (69.162796 34.531094)

For data on **administrative boundaries**, Rick Wilson's "Free GIS Data" site has links to multiple data sources. Of these, we will use `gadm.org`



From GADM's website, we can get data on sub-national administrative boundaries. Download the level2 GeoJSON file for Iraq...



The screenshot shows a web browser window with the GADM website. The address bar shows the URL `https://gadm.org/download_country.html`. The page has an orange header with the GADM logo and navigation links: Maps, Data, and About. The main heading is "Download GADM data (version 4.1)". Below this, there is a "Country" dropdown menu with "Iraq" selected. To the right of the dropdown, there are links for "Geopackage", "Shapefile", and "GeoJSON: [level-0](#), [level1](#), [level2](#)". Below these links, it says "KMZ: [level-0](#), [level1](#), [level2](#)". To the right of the text is a map of Iraq with its administrative boundaries highlighted in red and yellow. At the bottom of the page, there is a link to the GeoJSON file: `https://geodata.ucdavis.edu/gadm/gadm4.1/json/gadm41_IRQ_2.json.zip` and a note about the datum: "latitude and the WGS84 datum."

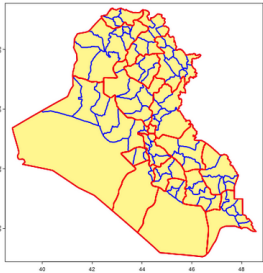
GADM Maps Data About

Download GADM data (version 4.1)

Country

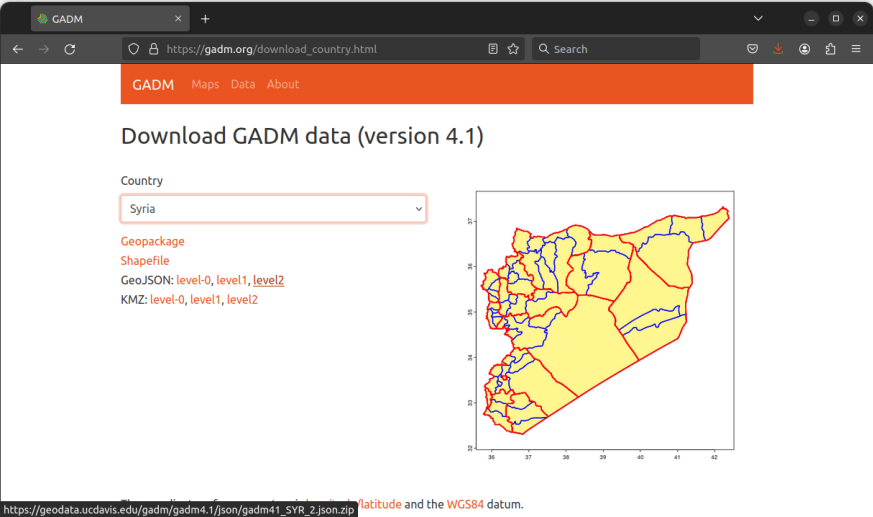
Iraq

Geopackage
Shapefile
GeoJSON: [level-0](#), [level1](#), [level2](#)
KMZ: [level-0](#), [level1](#), [level2](#)



https://geodata.ucdavis.edu/gadm/gadm4.1/json/gadm41_IRQ_2.json.zip latitude and the WGS84 datum.

... and for Syria



The screenshot shows a web browser window with the URL https://gadm.org/download_country.html. The page has an orange header with the GADM logo and navigation links: GADM, Maps, Data, and About. The main heading is "Download GADM data (version 4.1)". Below this, there is a "Country" dropdown menu with "Syria" selected. To the right of the dropdown is a map of Syria with its administrative boundaries highlighted in red and blue. Below the dropdown, there are links for different data formats: Geopackage, Shapefile, GeoJSON: [level-0](#), [level1](#), [level2](#), and KMZ: [level-0](#), [level1](#), [level2](#). At the bottom of the page, there is a link to the GeoJSON data for Syria: https://geodata.ucdavis.edu/gadm/gadm4.1/json/gadm41_SYR_2.json.zip, followed by the text "latitude and the WGS84 datum."

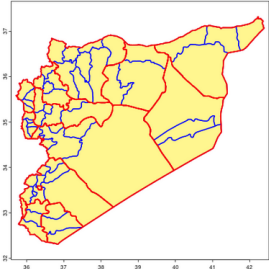
GADM Maps Data About

Download GADM data (version 4.1)

Country

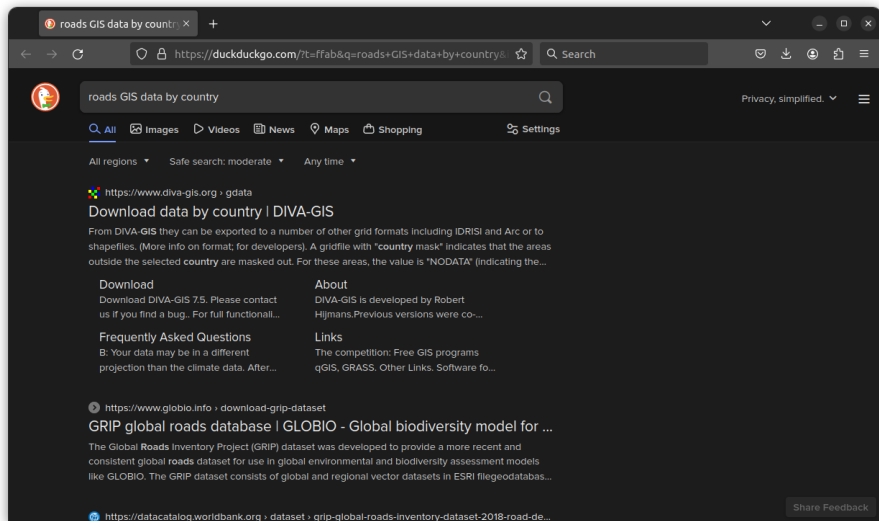
Syria

Geopackage
Shapefile
GeoJSON: [level-0](#), [level1](#), [level2](#)
KMZ: [level-0](#), [level1](#), [level2](#)

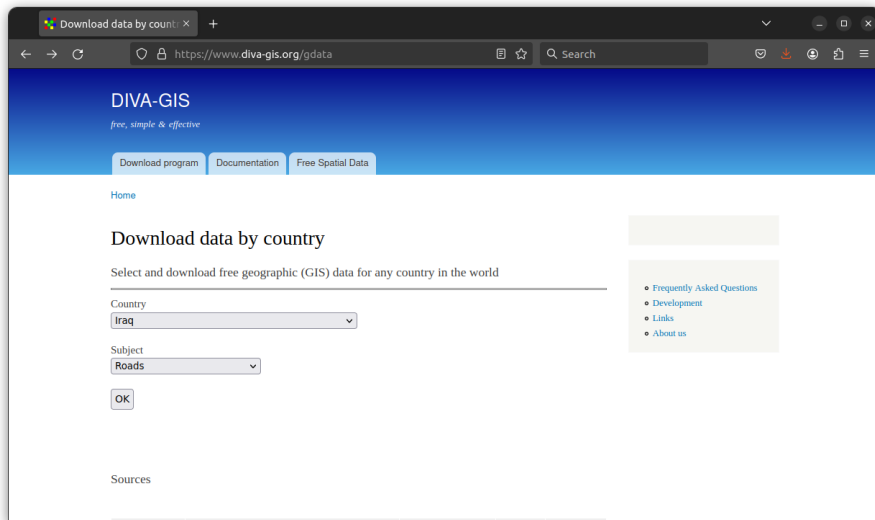


https://geodata.ucdavis.edu/gadm/gadm4.1/json/gadm41_SYR_2.json.zip latitude and the WGS84 datum.

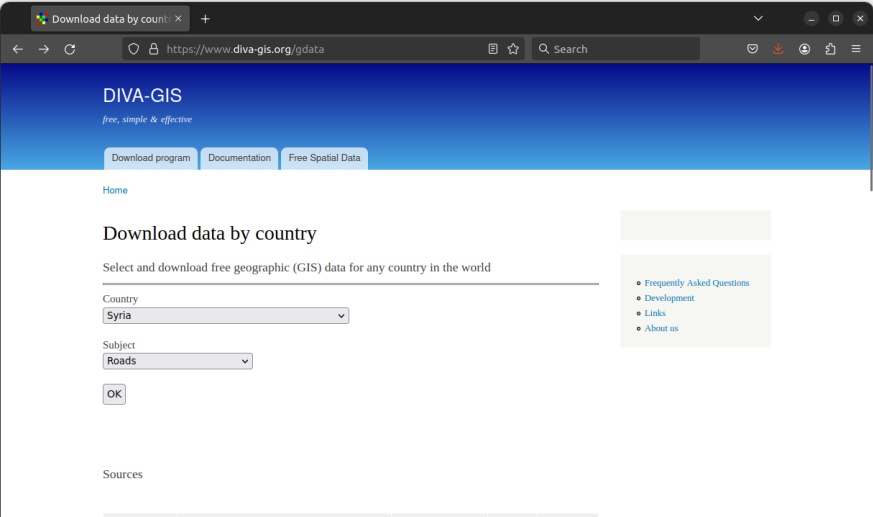
Multiple data sources exist for **roads**. Among the easiest to get are those from diva-gis.org/gData



These are country-level road layers from “Digital Chart of the World”,
download-able as zipped shapefiles



Download the roads data for both Iraq and Syria here



The screenshot shows a web browser window with the URL <https://www.diva-gis.org/gdata>. The page has a blue header with the text "DIVA-GIS" and the tagline "free, simple & effective". Below the header are three tabs: "Download program", "Documentation", and "Free Spatial Data". The main content area is titled "Download data by country" and includes the instruction "Select and download free geographic (GIS) data for any country in the world". There are two dropdown menus: "Country" with "Syria" selected and "Subject" with "Roads" selected. An "OK" button is located below the dropdowns. On the right side of the page, there is a sidebar with a list of links: "Frequently Asked Questions", "Development", "Links", and "About us".

Download data by country

Select and download free geographic (GIS) data for any country in the world

Country
Syria

Subject
Roads

OK

Sources

- [Frequently Asked Questions](#)
- [Development](#)
- [Links](#)
- [About us](#)

For **population data**, we can go back to Rick Wilson's page to find the link for "Gridded Population of the World (GPW)"

The screenshot shows a web browser window with the URL <https://freegisdata.rtwilson.com>. The page has a blue header with navigation links: Home, Physical, Human, Country-specific, and FAQ. There are also social media buttons for Facebook, Twitter, and a Donate button. The main content area is divided into sections: Wars, Conflict and Crime; Population; and Buildings, Roads and Points of Interest. The Population section is highlighted with a red box, and the link to the Gridded Population of the World (GPW) dataset is highlighted with a red box. The link text is: [Gridded Population of the World: Includes raw population, population density, both historic, current and predicted.](#)

data consists of polygons outlining reservoirs at high spatial resolution with extensive metadata about the dam and reservoir. **Registration required**

Wars, Conflict and Crime

- [ACLED: Armed Conflict Location and Event Data](#) - containing all reported conflict events in 50 countries in the developing world. Data from 1997 to present, and in Afghanistan and Pakistan from 2006 until present.
- [Uppsala Conflict Data Programme - Georeferenced Event Database](#): Locations of instances of political violence in Africa and Asia.
- [Global Terrorism Database](#): A database of terrorist events (both domestic and international) across the world from 1970-2008, including location and attribute information.
- [Peace Research Institute Oslo](#): A range of data including armed conflict locations, replication data, arms trade flows and resource datasets.

Population

- [Gridded Population of the World: Includes raw population, population density, both historic, current and predicted.](#)
- [Global Rural-Urban Mapping Project](#): Based on the above, but includes information on rural and urban population balances.
- [WorldPop](#): High-resolution, contemporary data on population across Africa, Asia and Central/Southern America. Combines the AfriPop, AmeriPop and AsiaPop projects.
- [High Resolution Settlement Layer](#): Estimates of human population at approximately 30m resolution, for 2015.
- [Facebook High Resolution Population Density Maps](#): High resolution (~30m) population density maps for over 150 countries.
- [Large Urban Areas 1950-2050](#): Historic, current and future estimates of populations in large urban areas of the world.
- [Global Urban Extent](#): Maps showing urban extent across the world, at 500m resolution, derived from MODIS images. **Requires email to author to download**
- [GeoHive](#): Population and country statistics. *Not provided in GIS data formats, but can easily be converted from CSV*

Buildings, Roads and Points of Interest

- [OpenStreetMap](#): Crowd-sourced data for the whole world consisting of most things you'd find on a standard local paper map: points of interest, buildings, roads and road names, ferry routes etc.
- [OSM Metro Extracts](#): City-sized extracts of the OpenStreetMap dataset, updated weekly for cities across the world
- [POI Factory](#): Point of Interest files originally designed for use in GPS units, but they can be loaded into a GIS fairly easily. Widely varying quality, and includes such things as shop and business locations (eg. all Tesco stores, all McDonald's restaurants) as well as places of worship, speed

sedac.ciesin.columbia.edu/gpw/

This will take us to the Center for International Earth Science Information Network (CIESIN) at Columbia, which hosts a large number of global, free GIS datasets

The screenshot shows a web browser window displaying the SEDAC website. The address bar shows the URL <https://sedac.ciesin.columbia.edu/data/collection/gpw-v4>. The page header includes the NASA logo and the text "EARTH DATA" and "Other DAACs". The main header features the SEDAC logo and the text "SOCIOECONOMIC DATA AND APPLICATIONS CENTER (SEDAC)" and "A Data Center in NASA's Earth Observing System Data and Information System (EOSDIS) — Hosted by CIESIN at Columbia University". The navigation menu includes links for DATA, MAPS, THEMES, RESOURCES, SOCIAL MEDIA, ABOUT, and HELP. The main content area is titled "Gridded Population of the World (GPW), v4" and includes a sidebar with links for Collection Overview, Methods, Data Sets (9), Map Gallery (97), Map Services (104), Citations, and FAQs. The main text area contains an introduction to the GPW collection, stating that it models the distribution of human population on a continuous global raster surface. A map of the world showing population density is displayed, with a legend indicating "GPWv4 Data Quality Indicators, v4.11_Mean Administrative Unit Area". The map shows a global distribution of population density, with higher densities in the Eastern Hemisphere. The page number "2 of 104" is visible at the bottom of the map. A search bar is located in the top right corner of the page.

Gridded Population of the World (GPW), v4

Follow Us: [Twitter](#) [Facebook](#) [YouTube](#) [RSS](#) | Share: [Twitter](#) [Facebook](#)

Collection Overview

Introduction

The Gridded Population of the World (GPW) collection, now in its fourth version (GPWv4), models the distribution of human population (counts and densities) on a continuous global raster surface. Since the release of the first version of this global population surface in 1995, the essential inputs to GPW have been population census tables and corresponding geographic boundaries. The purpose of GPW is to provide a spatially disaggregated population layer that is compatible with data sets from social, economic, and Earth science disciplines, and computer processing. It provides globally consistent and

GPWv4 Data Quality Indicators, v4.11_Mean Administrative Unit Area

2 of 104

Waiting for widget.uservoice.com...

Find in page ☐ Highlight All ☐ Match Case ☐ Match Diacritics ☐ Whole Words

Navigate to GPW's "Population Count v4.11" dataset, and download the files for 2010 (you will need to create an account to login)

The screenshot shows a web browser window with the URL <https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count>. The page is titled "Gridded Population of the World (GPW), v4" and "Population Count, v4.11 (2000, 2005, 2010, 2015, 2020)". A modal notice box is displayed in the center, stating: "You are **required** to login to download data or maps. Click 'LOGIN' to proceed to log in or to register. If you click 'CANCEL', you may browse the page but you will still be required to login to download data or maps." The notice box has "LOGIN" and "CANCEL" buttons. The background page shows a sidebar with "Data Sets (9)" and "Map Gallery (97)", and a main content area with "Downloads" and "View Recommended Citation(s)". The bottom of the page includes a footer with the URL and a "feedback and support" link.

The data are available for multiple years, in multiple formats and resolutions. We will be using the 15-minute GeoTiff raster for 2010 (latest before ISIS emerged in 2014)

The screenshot shows a web browser window with the URL <https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population>. The page is titled "GPW v4 Population" and features a sidebar with navigation links: Map Gallery (97), Map Services (104), Citations, FAQs, Documentation, What's New in GPWv4, What's New in Revision 10, What's New in Revision 11, What is UN WPP-adjusted population data?, Multimedia, SEDAC Hazards Mapper, Population Estimation Service, Acknowledgments, and COVID-19 Impact.

The main content area contains the following text:

The GPW v4 population data are available at the native 30 arc-second resolution and four lower resolutions: 2.5 arc-minute, 15 arc-minute, 30 arc-minute, and 1 degree. The netCDF-4 files are available at all resolutions except 30 arc-second. The data are stored in WGS84, geographic coordinate system (latitude/longitude).

Each downloadable is a compressed zip file containing either the global GeoTIFF (.tif) or ASCII (.asc) for the year and resolution of the estimate, or the netCDF-4 (.nc) containing all years of the estimate at a selected resolution, as well as data quality layers and ancillary files. A separate documentation zip file contains PDF documentation, an Excel file (.xlsx) with country-level information and sources, and a text file (.txt) with a log of changes to the data set by version.

The netCDF-4 file format is only available for the All Years Combined category and is not available at 30 arc-second resolution.

Please select all required fields and press the Create Download button:

Temporal: Single Year | FileFormat: GeoTiff | Resolution: 15 Minute ... | Create Download

Files Selected: 1 Download Size: 1.063MB

- ☐ Select All
- ☐ Year 2000
- ☐ Year 2005
- ☒ Year 2010
- ☐ Year 2015
- ☐ Year 2020

feedback and support

Also from CIESIN, we can download data on **cropland** (“Croplands, v1 (2000)”)

The screenshot shows a web browser window displaying the SEDAC (Socioeconomic Data and Applications Center) website. The browser's address bar shows the URL: <https://sedac.ciesin.columbia.edu/data/set/aglands-croplands>. The page header features the NASA logo and the text "EARTH DATA" and "Other DAACs". The main header includes the SEDAC logo and the text "SOCIOECONOMIC DATA AND APPLICATIONS CENTER (SEDAC)" and "A Data Center in NASA's Earth Observing System Data and Information System (EOSDIS) — Hosted by CIESIN at Columbia University". A navigation bar contains links for DATA, MAPS, THEMES, RESOURCES, SOCIAL MEDIA, ABOUT, and HELP. The main content area is titled "Global Agricultural Lands" and features a sidebar with links for Collection Overview, Methods, Data Sets (2), Map Gallery (14), Map Services (2), and Citations. The main content area displays the "Croplands, v1 (2000)" data set, with tabs for Set Overview, Data Download, Maps, Map Services, and Metadata. The "Purpose" section states: "To provide data on the extent of croplands for research on human-environment interactions." The "Abstract" section states: "The Global Croplands data set represents the proportion of land areas used as cropland (land used for the cultivation of food) in the year 2000. Satellite data from Moderate Resolution Imaging Spectroradiometer (MODIS) and Satellite Based Observation of Land Use (SBOL) are used." A world map titled "Global Agricultural Lands: Croplands, 2000" is displayed, showing the distribution of croplands globally. The page number "1 of 1" and a "feedback and support" link are visible at the bottom right.

Croplands, v1: Global Ag x +

<https://sedac.ciesin.columbia.edu/data/set/aglands-croplands>

Search SEDAC...

DATA MAPS THEMES RESOURCES SOCIAL MEDIA ABOUT HELP

Global Agricultural Lands

Follow Us: [Twitter](#) [Facebook](#) [YouTube](#) [RSS](#) | Share: [Twitter](#) [Facebook](#)

Collection Overview

Methods

Data Sets (2)

| Croplands, v1 (2000)

Show All...

Map Gallery (14)

Map Services (2)

Citations

Croplands, v1 (2000)

Set Overview Data Download Maps Map Services Metadata

Purpose:

To provide data on the extent of croplands for research on human-environment interactions.

Abstract:

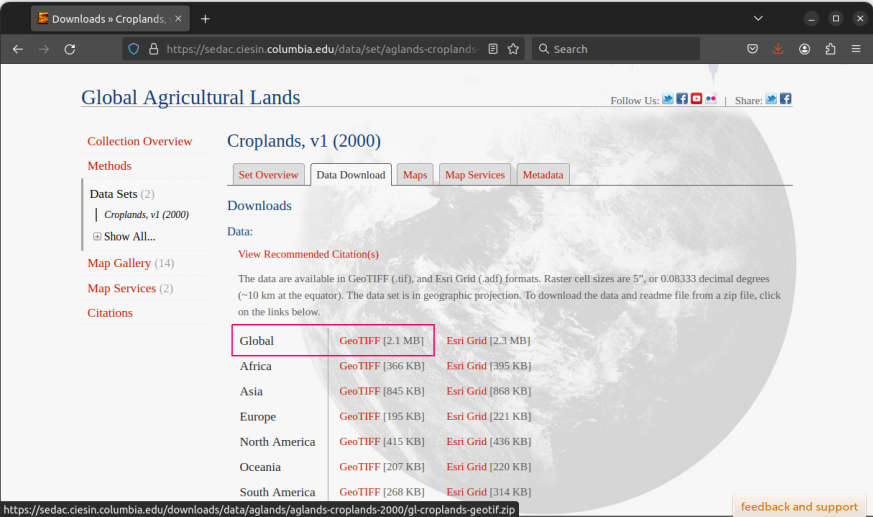
The Global Croplands data set represents the proportion of land areas used as cropland (land used for the cultivation of food) in the year 2000. Satellite data from Moderate Resolution Imaging Spectroradiometer (MODIS) and Satellite Based Observation of Land Use (SBOL) are used.

Global Agricultural Lands: Croplands, 2000

1 of 1

feedback and support

Download the global raster, in GeoTiff format



Global Agricultural Lands

Follow Us: [Twitter](#) [Facebook](#) [YouTube](#) [RSS](#) | Share: [Twitter](#) [Facebook](#)

Collection Overview

Methods

Data Sets (2)

- Croplands, v1 (2000)

Show All...

Map Gallery (14)

Map Services (2)

Citations

Croplands, v1 (2000)

[Set Overview](#) [Data Download](#) [Maps](#) [Map Services](#) [Metadata](#)

Downloads

Data:

View Recommended Citation(s)

The data are available in GeoTIFF (.tif), and Esri Grid (.adf) formats. Raster cell sizes are 5", or 0.08333 decimal degrees (~10 km at the equator). The data set is in geographic projection. To download the data and readme file from a zip file, click on the links below.

Global	GeoTIFF [2.1 MB]	Esri Grid [2.3 MB]
Africa	GeoTIFF [366 KB]	Esri Grid [395 KB]
Asia	GeoTIFF [845 KB]	Esri Grid [868 KB]
Europe	GeoTIFF [195 KB]	Esri Grid [221 KB]
North America	GeoTIFF [415 KB]	Esri Grid [436 KB]
Oceania	GeoTIFF [207 KB]	Esri Grid [220 KB]
South America	GeoTIFF [268 KB]	Esri Grid [314 KB]

<https://sedac.ciesin.columbia.edu/downloads/data/aglands/aglands-croplands-2000/gl-croplands-geotif.zip>

[feedback and support](#)

While we're here, let's also download data on **hydroelectric dams** ("Global Reservoir and Dam (GRanD), v1")

The screenshot shows a web browser window displaying the SEDAC (Socioeconomic Data and Applications Center) website. The browser's address bar shows the URL: <https://sedac.ciesin.columbia.edu/data/set/grand-v1-dams-rev>. The website header includes the NASA logo and the text "SOCIOECONOMIC DATA AND APPLICATIONS CENTER (SEDAC)". Below the header is a navigation bar with links: DATA, MAPS, THEMES, RESOURCES, SOCIAL MEDIA, ABOUT, and HELP. The main content area is titled "Global Reservoir and Dam (GRanD), v1". On the left side, there is a sidebar with links: Collection Overview, Methods, Data Sets (2), Map Gallery (20), Map Services (2), Citations, and Project Website. The main content area shows the "Dams, v1.01 (2011)" collection overview. It includes a "Purpose:" section stating: "To provide a geographically explicit and reliable database of dams for the scientific community." and an "Abstract:" section stating: "The Global Reservoir and Dam Database, Version 1, Revision 01 (v1.01) contains 6,862 records of reservoirs and their associated dams with a cumulative storage capacity of 6,197 cubic km. The dams were geospatially...". A world map titled "Global Reservoir and Dam Database: Dams" is displayed on the right side of the main content area. The map shows the distribution of dams globally, with a high concentration in Asia and Europe. The map is labeled "1 of 1" and has a "feedback and support" button.

Dams, v1.01: Global Rese X +

https://sedac.ciesin.columbia.edu/data/set/grand-v1-dams-rev

EARTHDATA Other DAACs

NASA SOCIOECONOMIC DATA AND APPLICATIONS CENTER (SEDAC)
A Data Center in NASA's Earth Observing System Data and Information System (EOSDIS) — Hosted by CIESIN at Columbia University

Search SEDAC... Data

DATA MAPS THEMES RESOURCES SOCIAL MEDIA ABOUT HELP

Global Reservoir and Dam (GRanD), v1

Follow Us: | Share: |

Collection Overview Dams, v1.01 (2011)

Methods Set Overview Data Download Maps Map Services Metadata

Purpose:
To provide a geographically explicit and reliable database of dams for the scientific community.

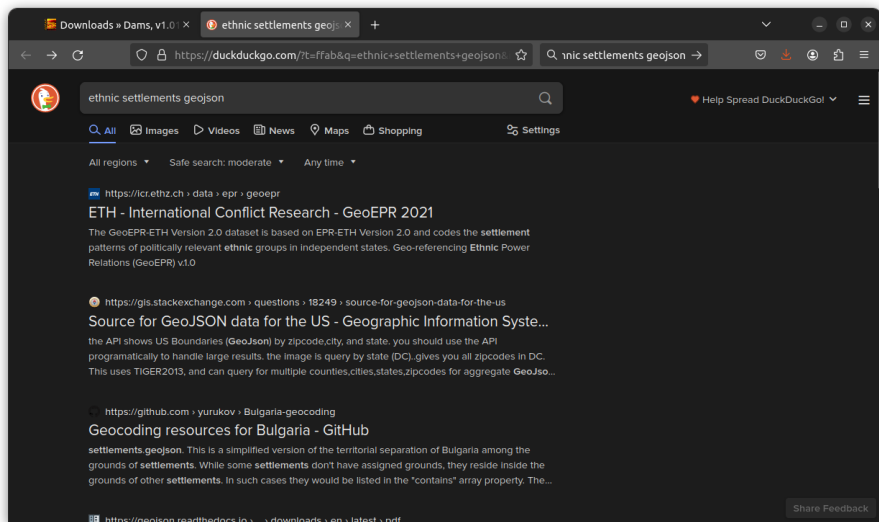
Abstract:
The Global Reservoir and Dam Database, Version 1, Revision 01 (v1.01) contains 6,862 records of reservoirs and their associated dams with a cumulative storage capacity of 6,197 cubic km. The dams were geospatially

Global Reservoir and Dam Database: Dams

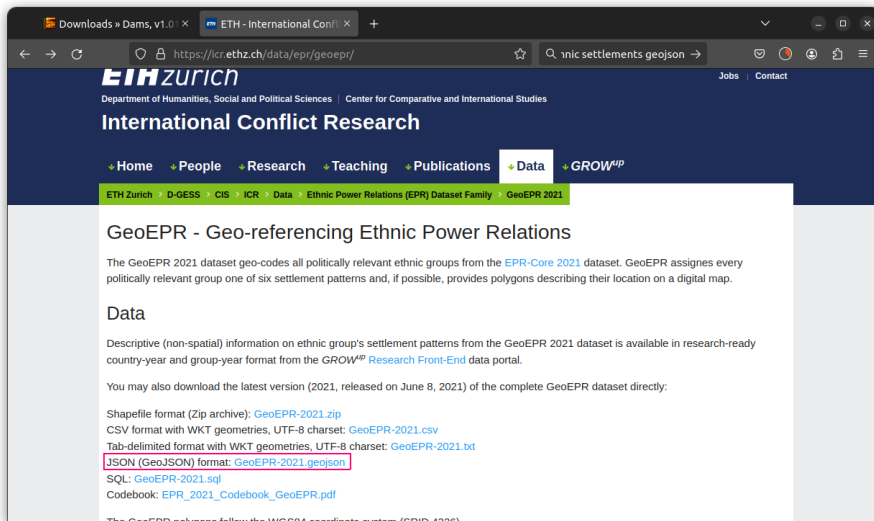
1 of 1

feedback and support

For data on **sectarian divisions**, a relatively reliable global source is the “Ethnic Power Relations” (geoEPR) dataset — we used this in Problem Set 02



Specifically, we will download the `GeoEPR-2021.geojson` file, which contains polygons for each ethno-religious group (`icr.ethz.ch/data/epr/geoepr`)



Here is the full list of data sources and links:

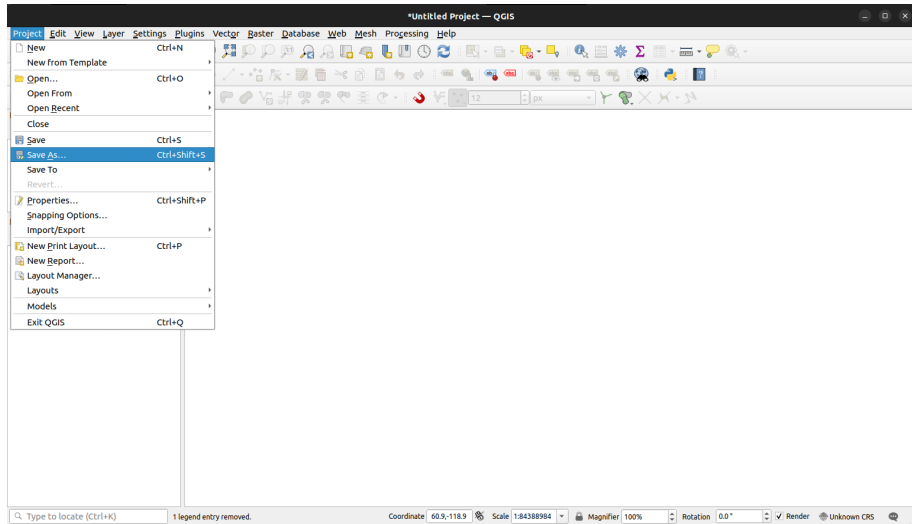
Category	Type	Format	Data source
ISIS violence	Table (non-geo)	.csv	UCDP GED
Administrative units	Vector (polygons)	.geojson	GADM
Roads	Vector (polylines)	.shp	DIVA-GIS
Population	Raster	.tif	CIESIN
Cropland	Raster	.tif	CIESIN
Dams	Vector (points)	.shp	CIESIN
Sectarian divisions	Vector (polygons)	.geojson	ETH-Zurich

These are all in the WT01.zip file posted on Canvas.

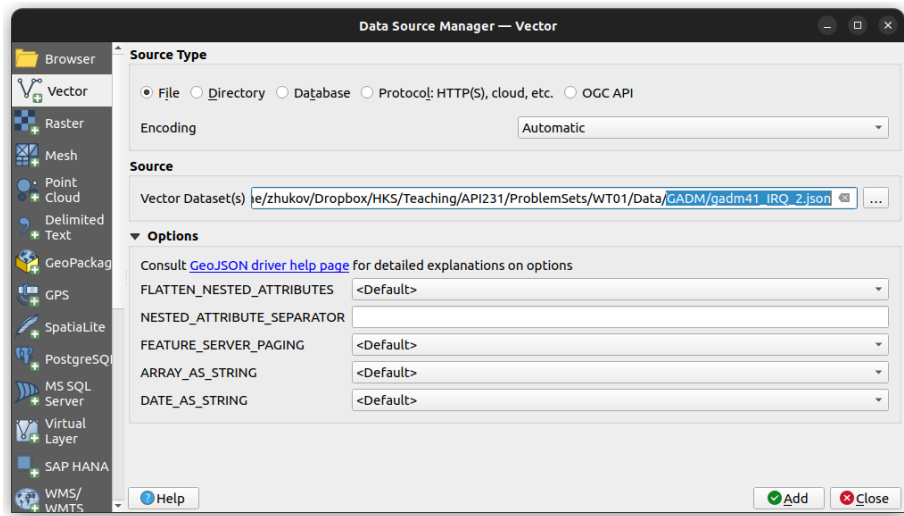
Pre-processing

Always save your progress!

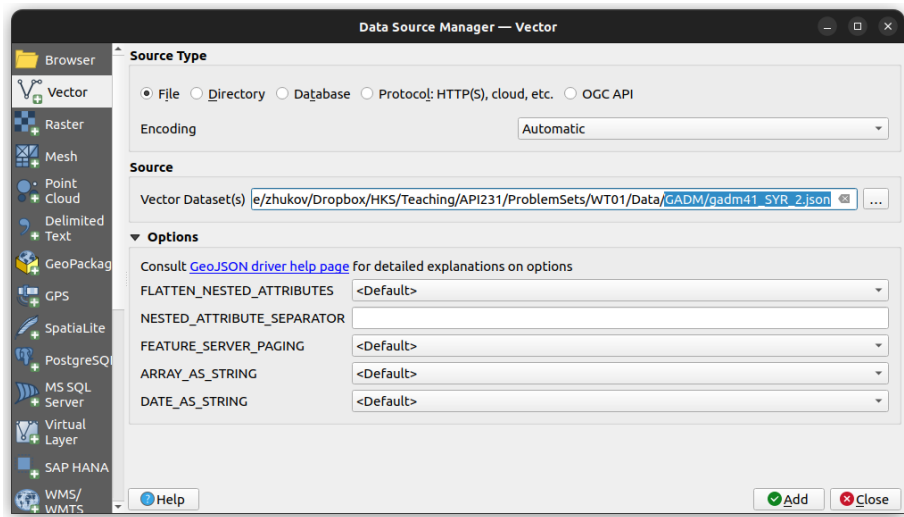
Go to Project → Save As...



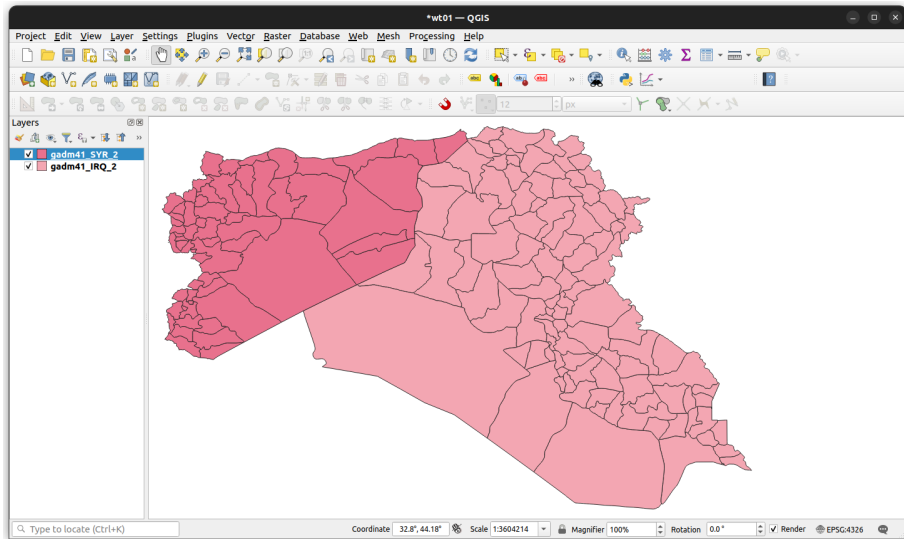
Load *administrative boundaries* (Layer → Add Layer → Add Vector Layer).
For **Iraq**: gadm41_IRQ_2.json file in Data/GADM folder.



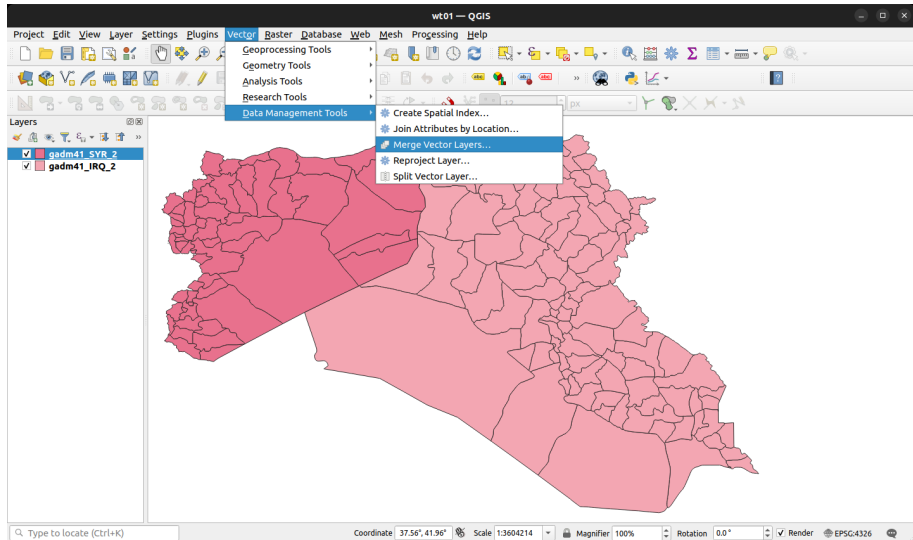
For **Syria**: gadm41_SYR_2.json file in Data/GADM folder.



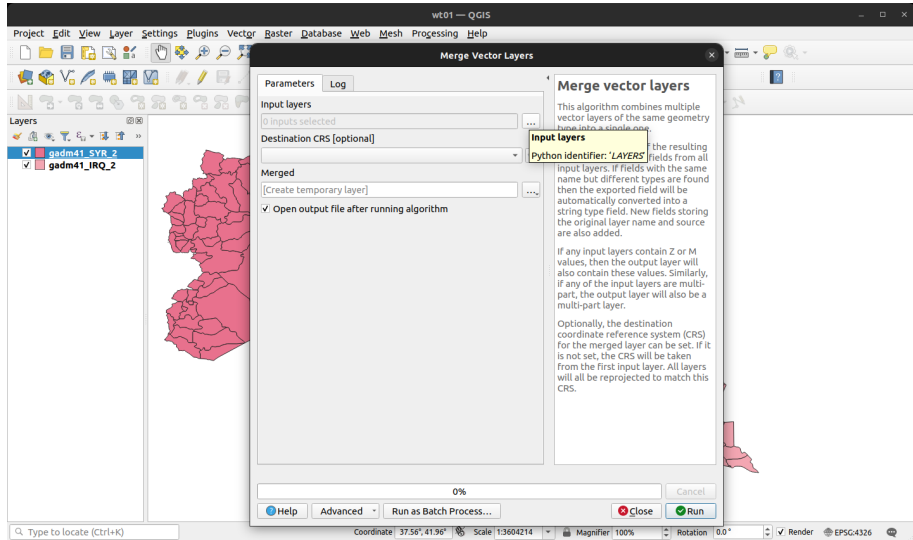
The two sets of district boundaries should be visible in the project window. Let's **merge them into a single layer**, with districts for both countries.



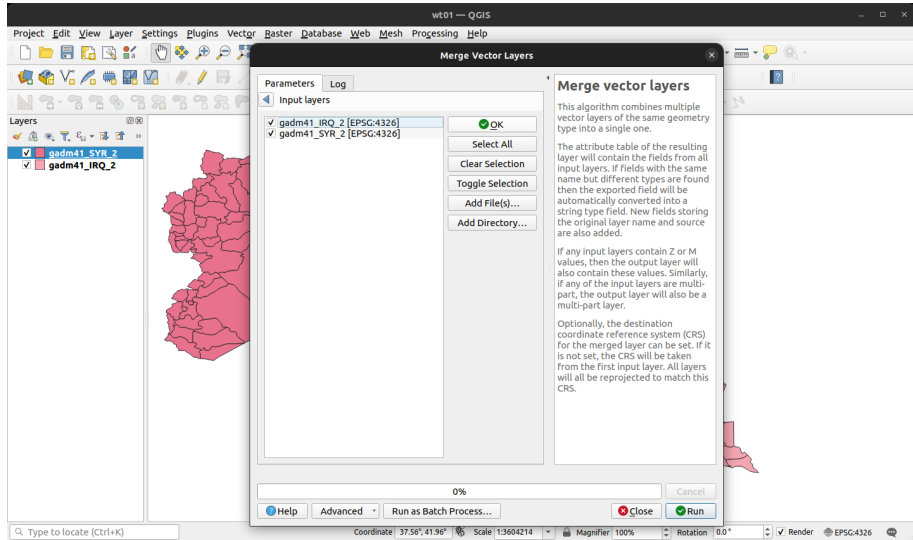
Open the Merge Vector Layers tool (Vector → Data Management Tool → Merge Vector Layers).



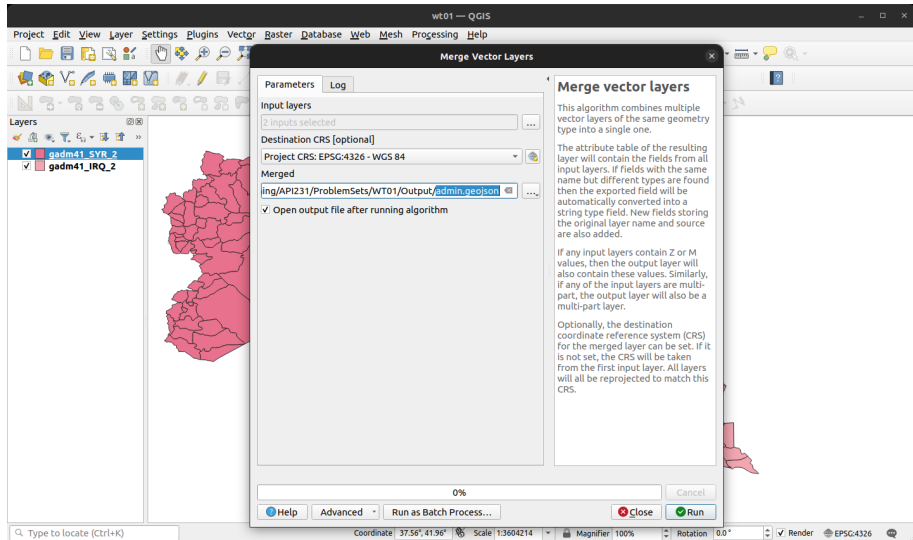
Click on the [...] box next to Input layers



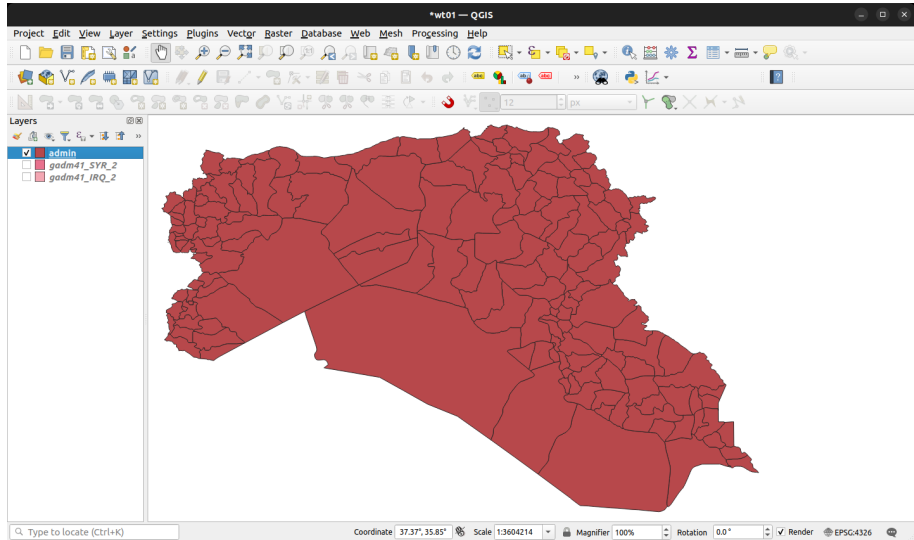
Check the boxes next to `gadm41_IRQ_2` and `gadm41_SYR_2`. Click OK.



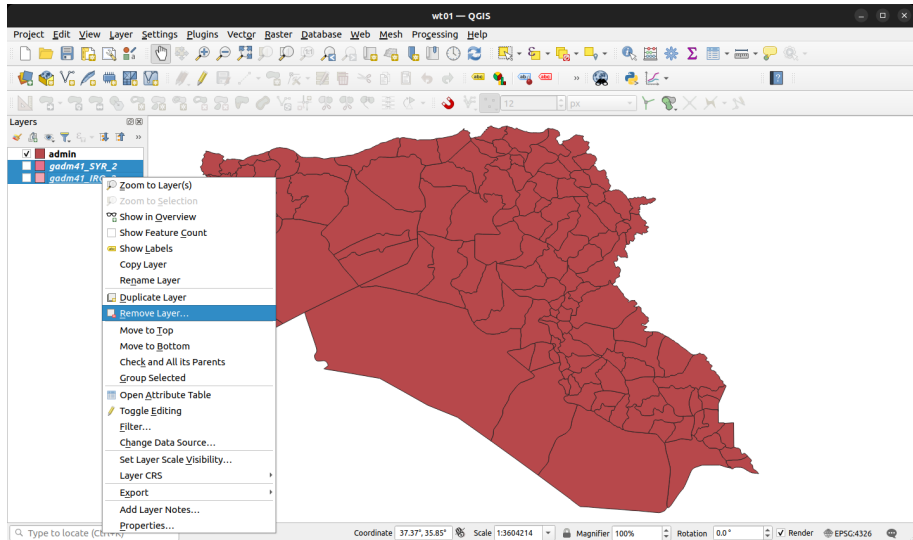
Click on the [...] box next to Merged. Find a location in which to save the output, and name the file `admin.geojson`. Click Run



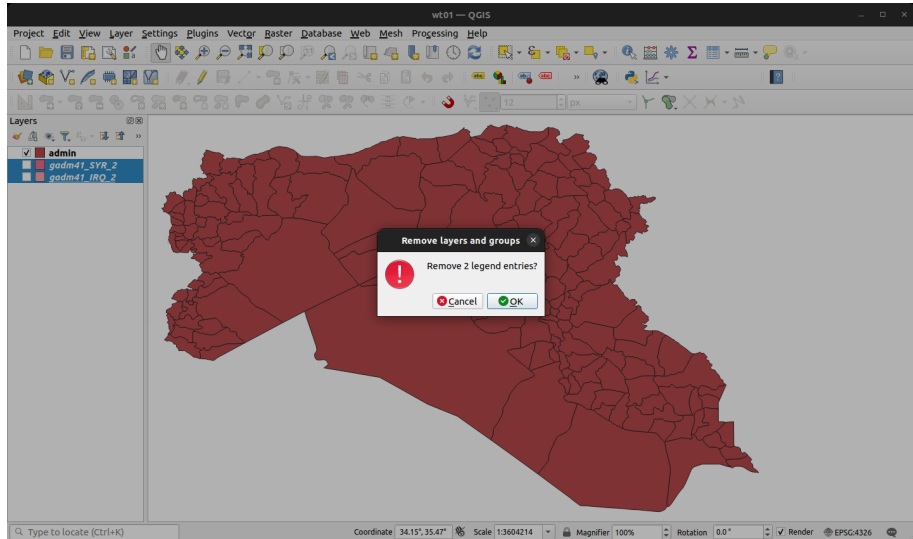
The merged layer should appear in the main project window.



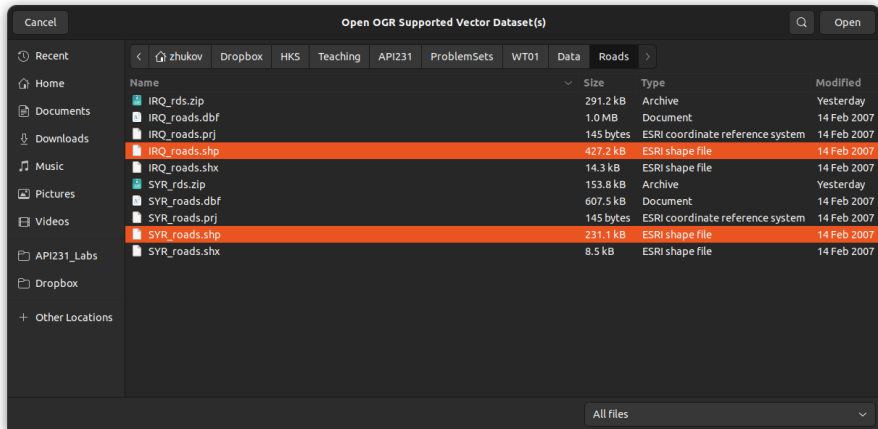
You can remove the gadm41_IRQ_2 and gadm41_SYR_2 layers from the project (Right-click on layer(s) in Layers menu → Remove Layer...)



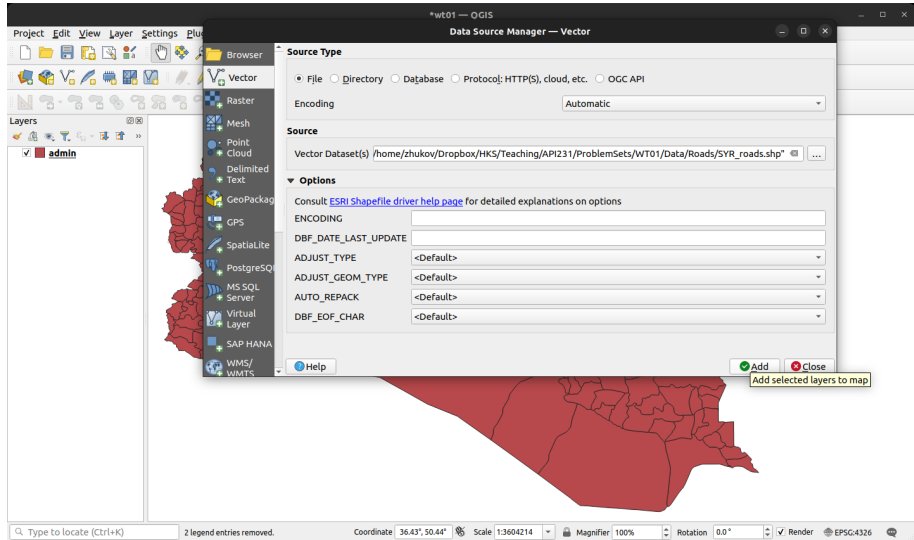
To conserve memory, it's good practice to remove data objects we are not actively using



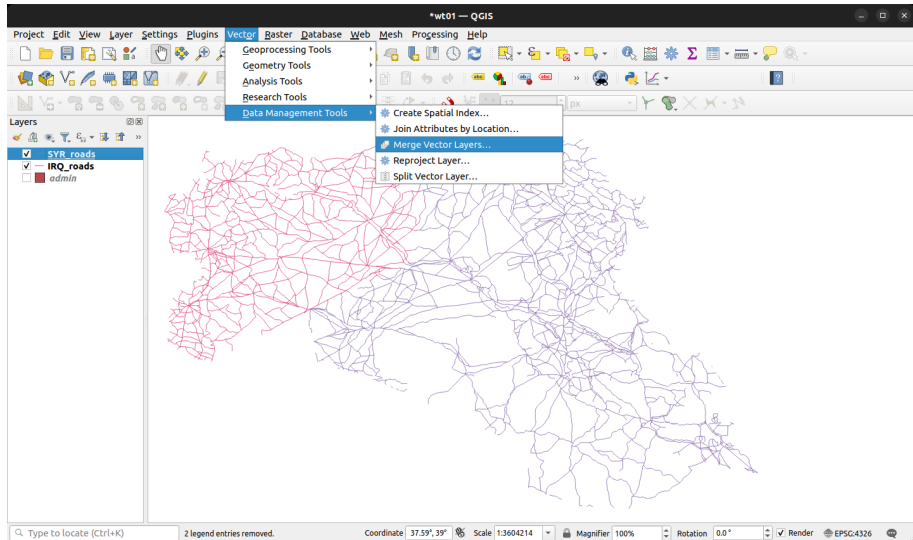
Load *roads data* (Layer → Add Layer → Add Vector Layer).
Try opening **two datasets at the same time**: IRQ_roads.shp and SYR_roads.shp files in Data/Roads



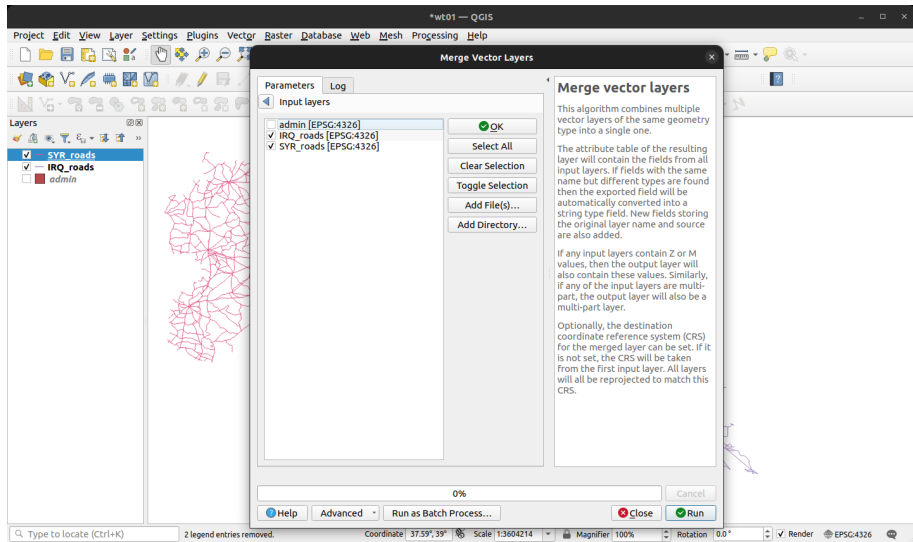
Paths to both files should appear in the Source box. Click Add



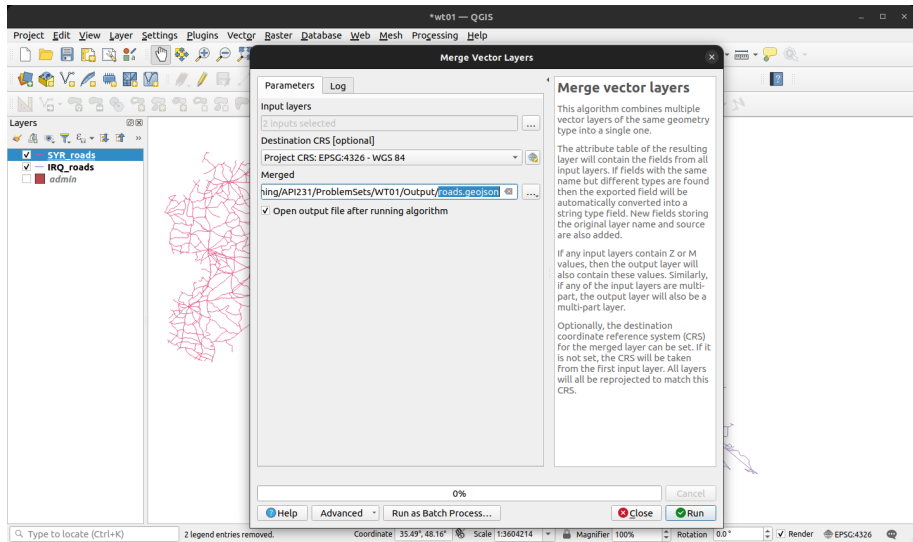
Let's repeat the Merge Vector Layers procedure for these two layers (Vector → Data Management Tool → Merge Vector Layers).



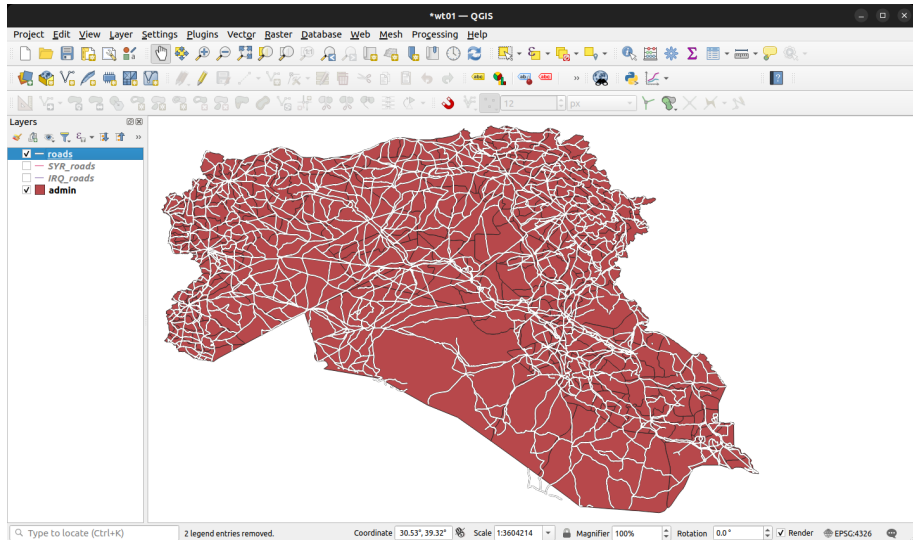
This time, perform the merge on the IRQ_roads and SYR_roads layers



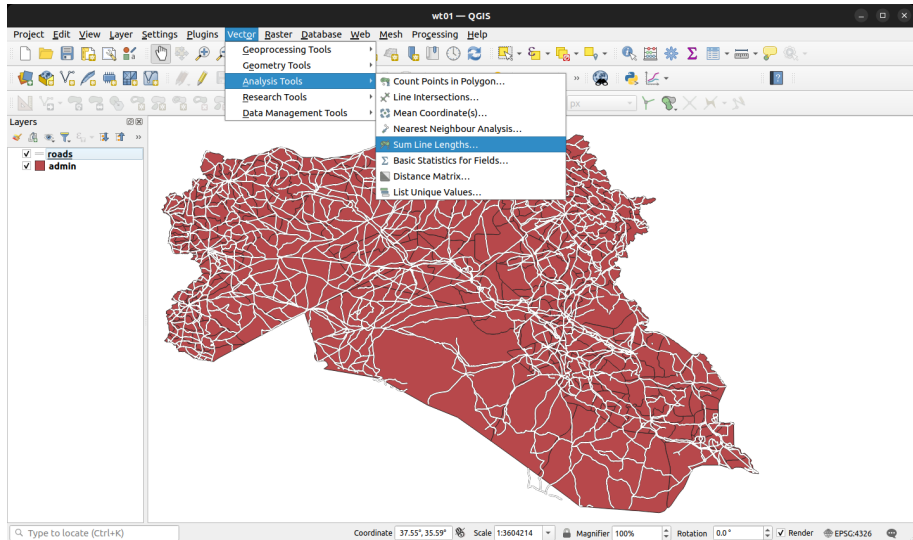
Save the output file as roads.geojson



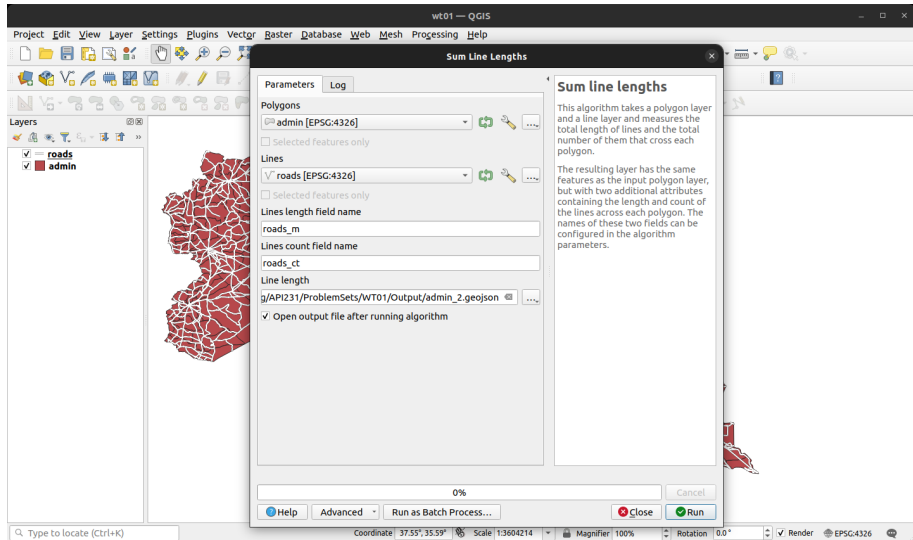
As before, you can remove the two country-specific roads layers (we no longer need them)



Let's **calculate each district's road density** with the Sum Line Lengths tool
(Vector → Analysis → Sum Line Lengths)



Set Polygons = admin, Lines = roads, Line length field name = roads_m,
Lines count field name = roads_ct. Save as admin_2.geojson

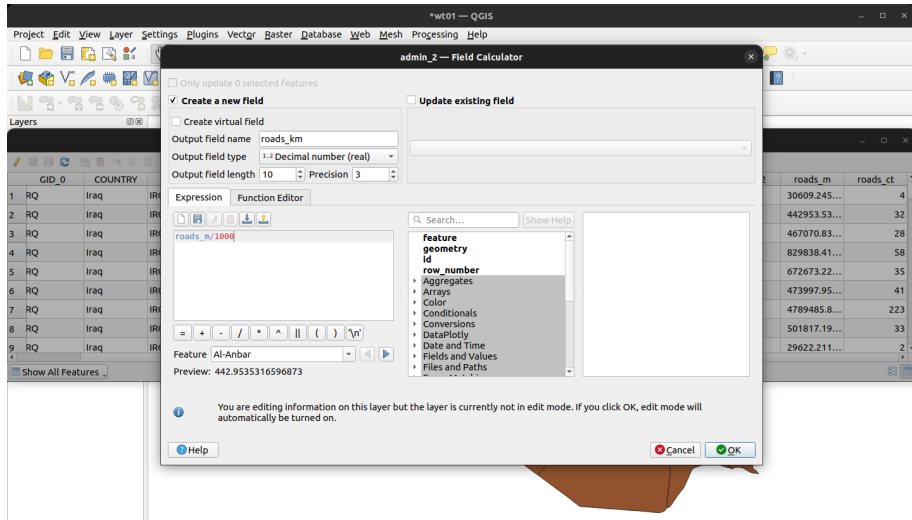


Open the attribute table for admin_2. The roads_m and roads_ct variables should be in the table. Let's convert from **meters to kilometers**

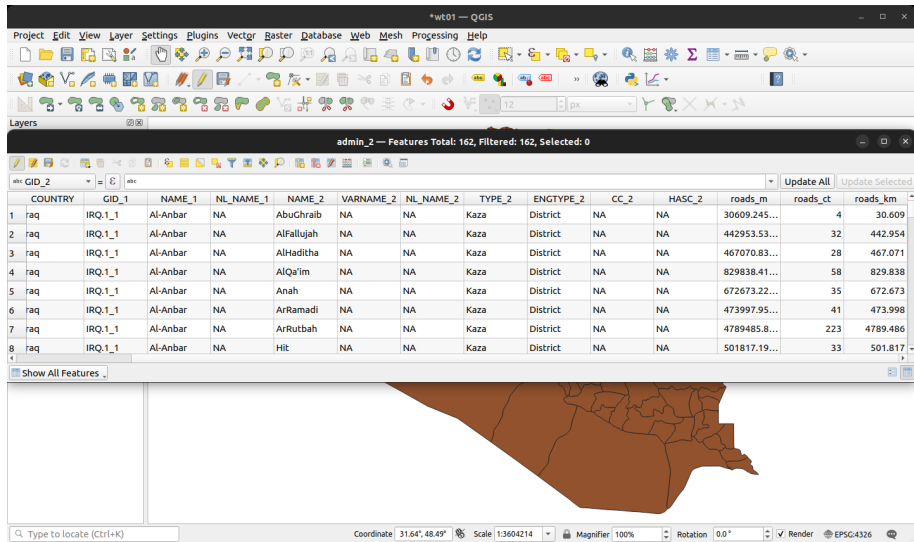
The screenshot shows the QGIS interface with the attribute table for the 'admin_2' layer open. The table contains 162 features, all filtered and none selected. The table has 15 columns: GID_0, COUNTRY, GID_1, NAME_1, NL_NAME_1, NAME_2, VARNAME_2, NL_NAME_2, TYPE_2, ENGTTYPE_2, CC_2, HASC_2, roads_m, and roads_ct. The first 9 rows are visible, showing data for various districts in Iraq. Below the table is a map view of Iraq, with a search bar and coordinate information at the bottom.

GID_0	COUNTRY	GID_1	NAME_1	NL_NAME_1	NAME_2	VARNAME_2	NL_NAME_2	TYPE_2	ENGTTYPE_2	CC_2	HASC_2	roads_m	roads_ct
1 RQ	Iraq	IRQ.1_1	Al-Anbar	NA	AbuGhraib	NA	NA	Kaza	District	NA	NA	30609.245...	4
2 RQ	Iraq	IRQ.1_1	Al-Anbar	NA	AlFallujah	NA	NA	Kaza	District	NA	NA	442953.53...	32
3 RQ	Iraq	IRQ.1_1	Al-Anbar	NA	AlHaditha	NA	NA	Kaza	District	NA	NA	467070.83...	28
4 RQ	Iraq	IRQ.1_1	Al-Anbar	NA	AlQa'im	NA	NA	Kaza	District	NA	NA	829838.41...	58
5 RQ	Iraq	IRQ.1_1	Al-Anbar	NA	Anah	NA	NA	Kaza	District	NA	NA	672673.22...	35
6 RQ	Iraq	IRQ.1_1	Al-Anbar	NA	ArRamadi	NA	NA	Kaza	District	NA	NA	473997.95...	41
7 RQ	Iraq	IRQ.1_1	Al-Anbar	NA	ArRutbah	NA	NA	Kaza	District	NA	NA	4789485.8...	223
8 RQ	Iraq	IRQ.1_1	Al-Anbar	NA	Hit	NA	NA	Kaza	District	NA	NA	501817.19...	33
9 RQ	Iraq	IRQ.1_1	Al-Anbar	NA	Kadhimiya	NA	NA	Kaza	District	NA	NA	29622.211...	2

Open Field Calculator for admin_2. Create new field, roads_km, of type Decimal number (real). For the Expression, write roads_m/1000. The Output preview should show a number with decimals. Click OK



The resulting roads_km field should be visible in the attribute table

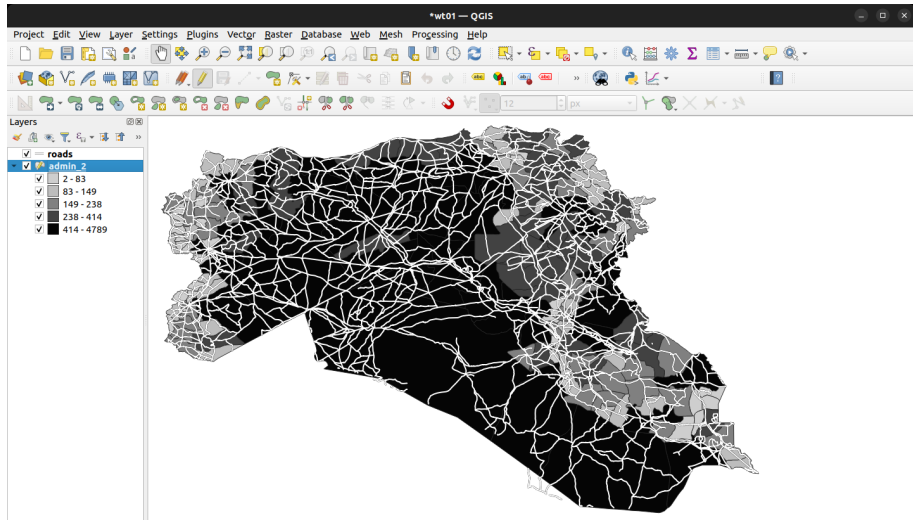


The screenshot displays the QGIS application window with the 'admin_2' layer selected. The attribute table is open, showing 162 features. The table structure is as follows:

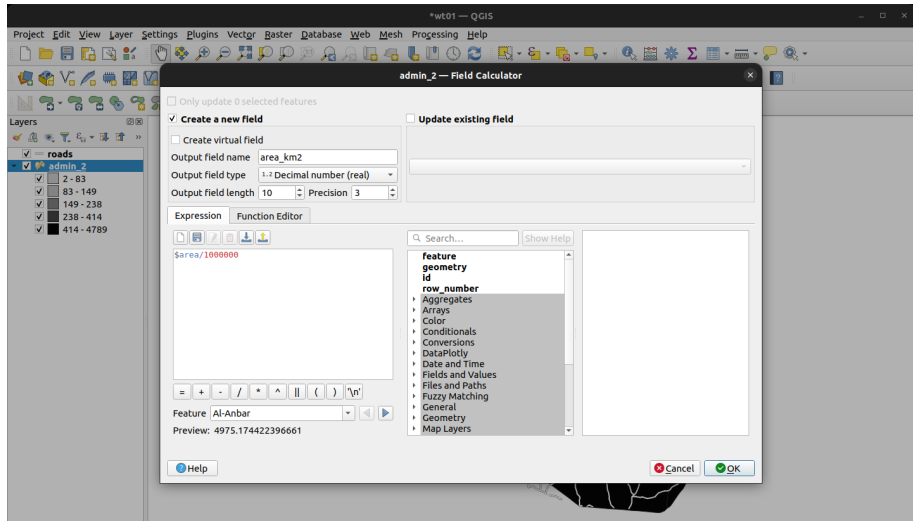
	COUNTRY	GID_1	NAME_1	NL_NAME_1	NAME_2	VARNAME_2	NL_NAME_2	TYPE_2	ENGTYPE_2	CC_2	HASC_2	roads_m	roads_ct	roads_km
1	raq	IRQ_1_1	Al-Anbar	NA	AbuGhraib	NA	NA	Kaza	District	NA	NA	30609.245...	4	30.609
2	raq	IRQ_1_1	Al-Anbar	NA	AlFallujah	NA	NA	Kaza	District	NA	NA	442953.53...	32	442.954
3	raq	IRQ_1_1	Al-Anbar	NA	AlHaditha	NA	NA	Kaza	District	NA	NA	467070.83...	28	467.071
4	raq	IRQ_1_1	Al-Anbar	NA	AlQa'im	NA	NA	Kaza	District	NA	NA	829838.41...	58	829.838
5	raq	IRQ_1_1	Al-Anbar	NA	Anah	NA	NA	Kaza	District	NA	NA	672673.22...	35	672.673
6	raq	IRQ_1_1	Al-Anbar	NA	ArRamadi	NA	NA	Kaza	District	NA	NA	473997.95...	41	473.998
7	raq	IRQ_1_1	Al-Anbar	NA	ArRutbah	NA	NA	Kaza	District	NA	NA	4789485.8...	223	4789.486
8	raq	IRQ_1_1	Al-Anbar	NA	Hilt	NA	NA	Kaza	District	NA	NA	501817.19...	33	501.817

The map view at the bottom shows a brown-shaded area representing the administrative region of Al-Anbar in Iraq. The status bar at the bottom indicates the coordinate system (EPSG:4326) and other project settings.

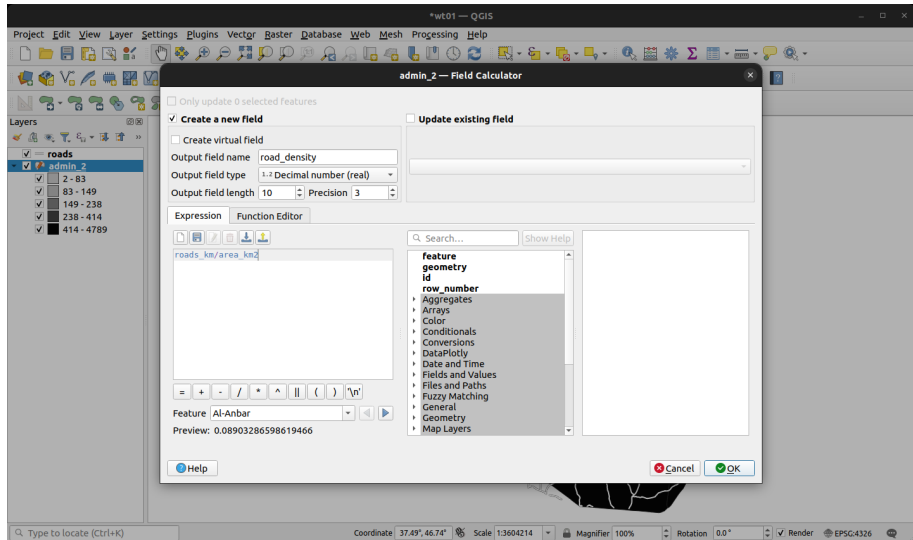
If we plot this variable, we can see its main limitations: smaller districts have less road length, so this becomes a proxy for district size. Let's create a **road density** measure that takes into account district size



Back in the Field Calculator, create a new field called `area_km2` of type Decimal number (real). For Expression, write `$area/1000000` (divide by 1M to convert from m^2 to km^2). Click OK



Open the Field Calculator again. Create a new field called `road_density` of type Decimal number (real). For Expression, write `roads_km/area_km2`



Check to make sure road_density is added to attribute table

*wt01 -- QGIS

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

Layers

admin_2 — Features Total: 162, Filtered: 162, Selected: 0

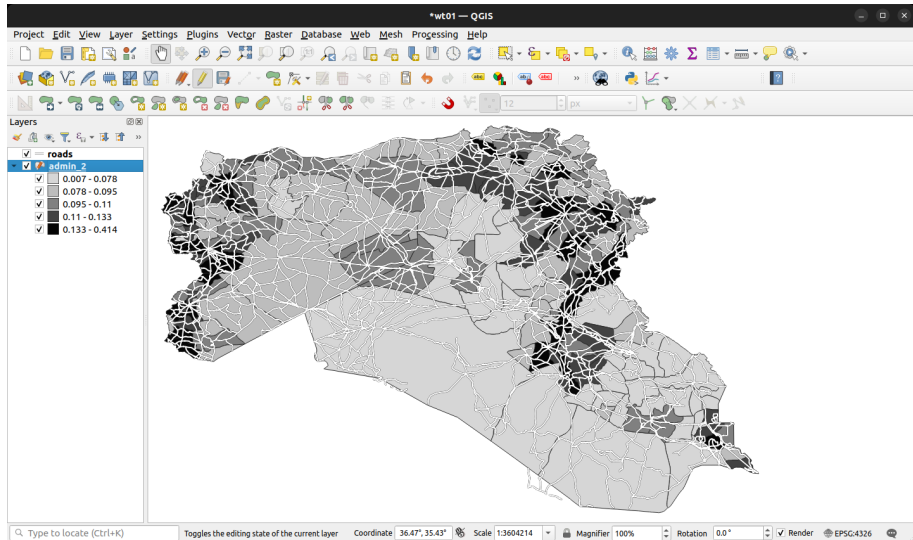
abc GID_2 ▾ = { abc

	NAME_1	NL_NAME_1	NAME_2	VARNAME_2	NL_NAME_2	TYPE_2	ENGTYPE_2	CC_2	HASC_2	roads_m	roads_ct	roads_km	area_km2	road_density
1	Al-Anbar	NA	AbuGhraib	NA	NA	Kaza	District	NA	NA	30609.245...	4	30.609	411.653	0.074
2	Al-Anbar	NA	AlFallujah	NA	NA	Kaza	District	NA	NA	442953.53...	32	442.954	4975.174	0.089
3	Al-Anbar	NA	AlHaditha	NA	NA	Kaza	District	NA	NA	467070.83...	28	467.071	4517.457	0.103
4	Al-Anbar	NA	AlQa'im	NA	NA	Kaza	District	NA	NA	829838.41...	58	829.838	8378.068	0.099
5	Al-Anbar	NA	Anah	NA	NA	Kaza	District	NA	NA	672673.22...	35	672.673	9891.892	0.068
6	Al-Anbar	NA	ArRamadi	NA	NA	Kaza	District	NA	NA	473997.95...	41	473.998	6803.200	0.07
7	Al-Anbar	NA	ArRutbah	NA	NA	Kaza	District	NA	NA	4789485.8...	223	4789.486	92452.662	0.052
8	Al-Anbar	NA	Hit	NA	NA	Kaza	District	NA	NA	501817.19...	33	501.817	6824.326	0.074

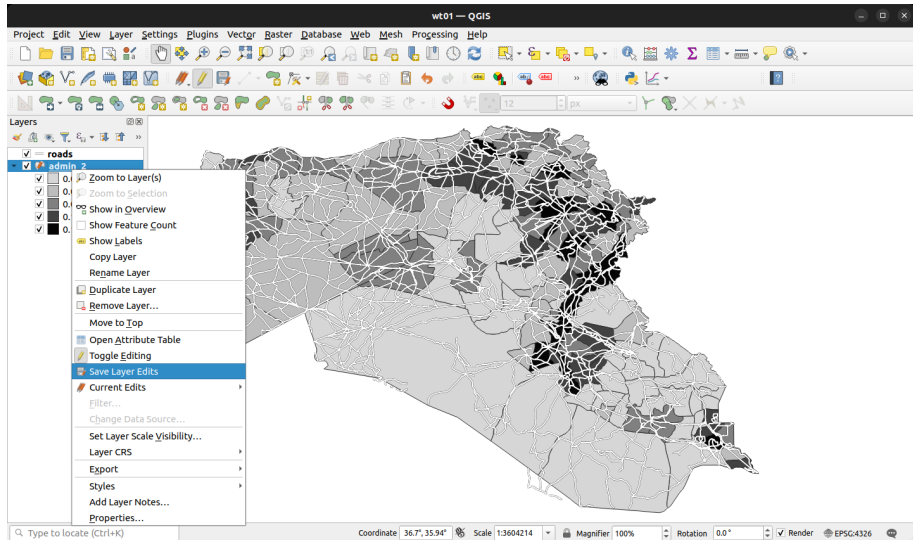
Show All Features

Type to locate (Ctrl+K) Toggles the editing state of the current layer Coordinate 35.96°, 35.5° Scale 1:3604214 Magnifier 100% Rotation 0.0° Render EPSG:4326

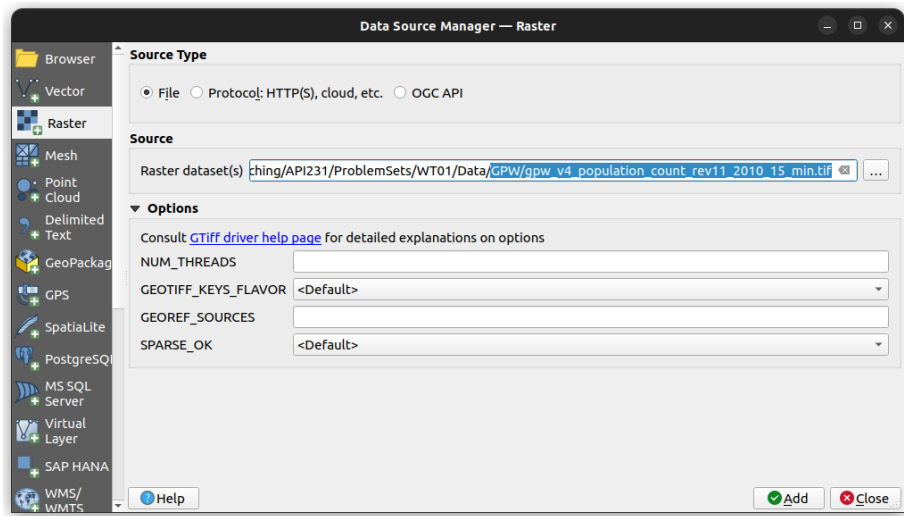
If we plot this variable, we see the opposite pattern from before (smaller, urban districts have higher road density)



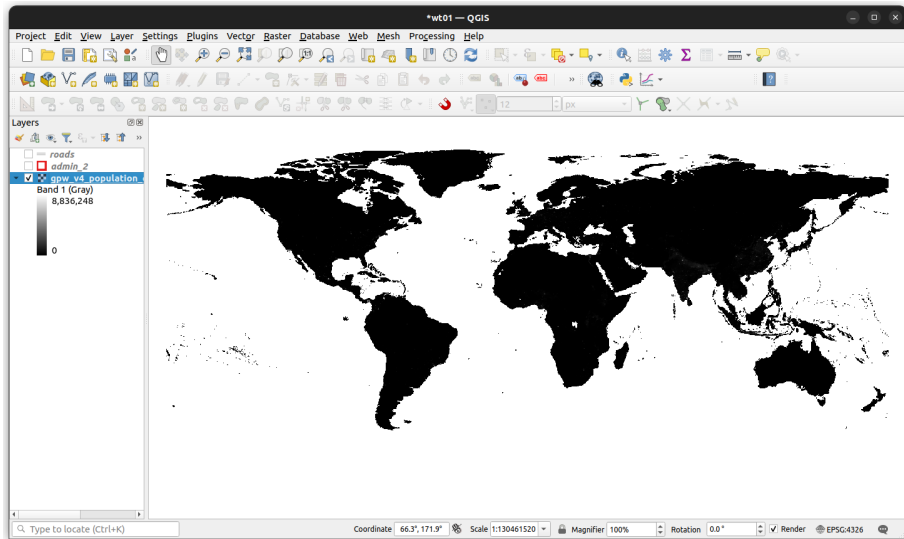
Save the edits you just made to the admin_2 layer (right click → Save Layer Edits)



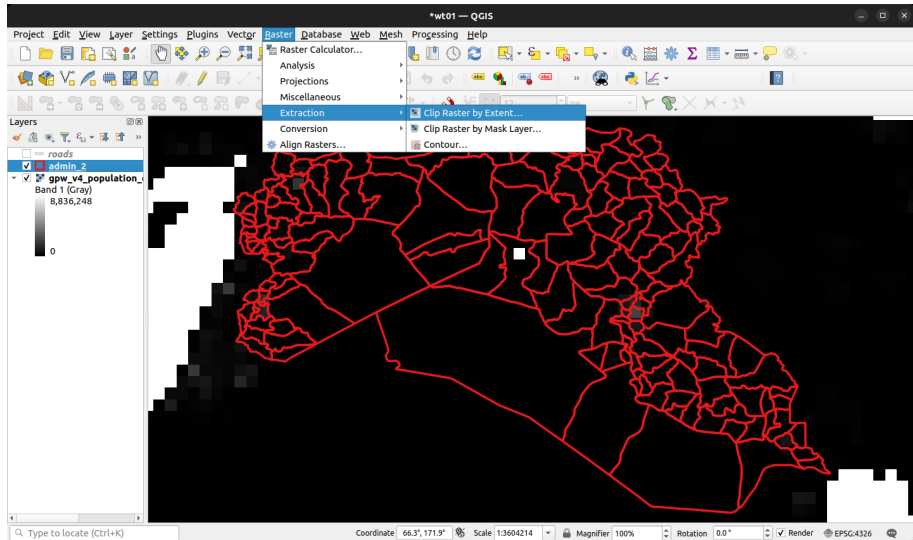
Load *population* raster (Layer → Add Layer → Add Raster Layer).\ Open the `gpw_v4_population_count_rev11_2010_15_min.tif` file in Data/GPW folder.



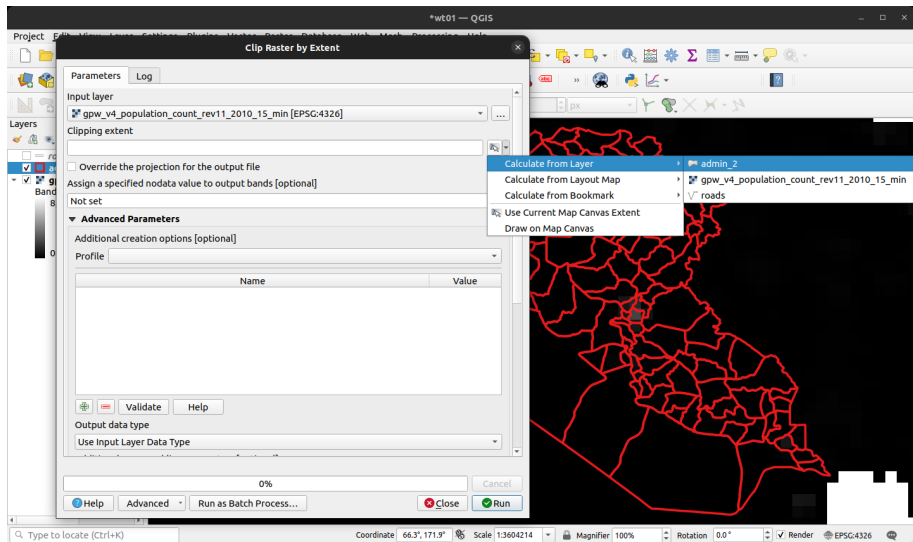
The (global) population layer should now be visible. Let's **extract** a **subset** of just the part of this layer that overlaps with Syria and Iraq.



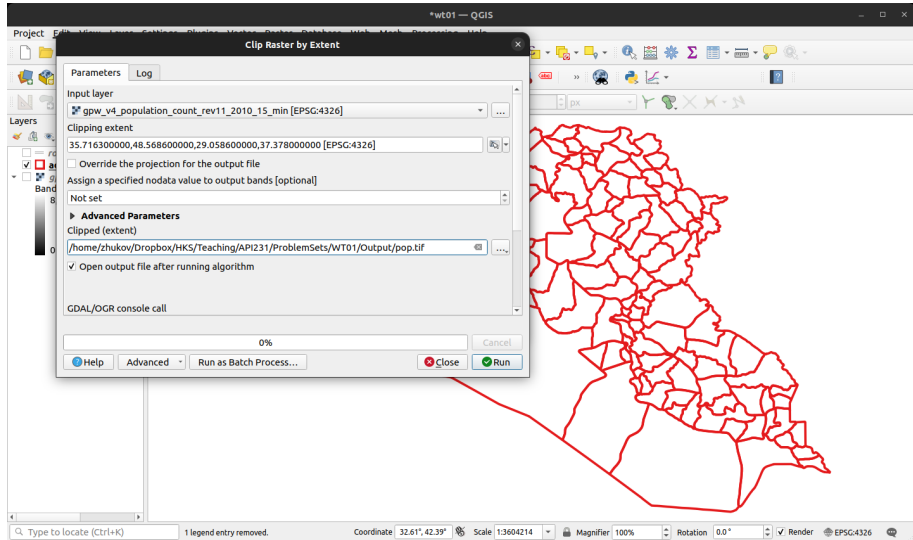
Open the Clip Raster by Extent tool (Raster → Extension → Clip Raster by Extent)



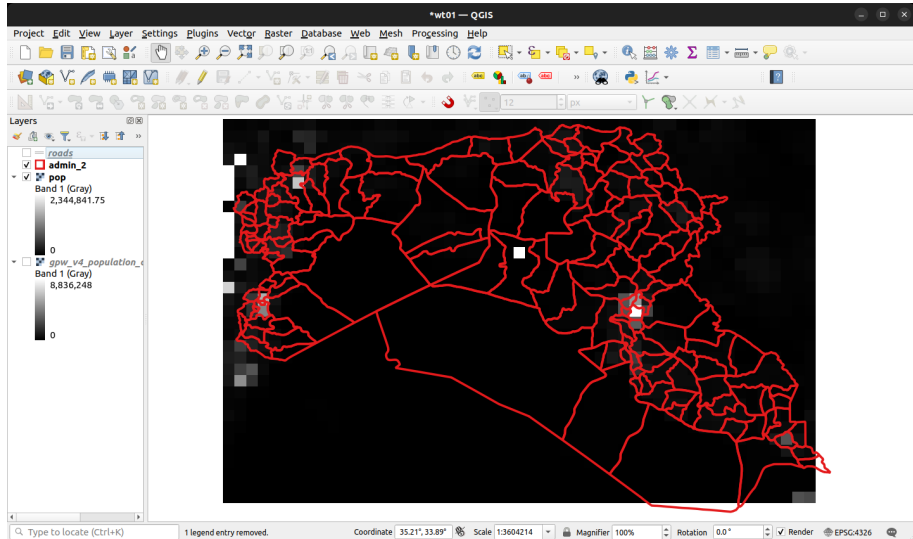
Set Input layer = gpw_v4_population_count_rev11_2010_15_min. Click the [...] button next to Clipping extent → Calculate from Layer → admin_2



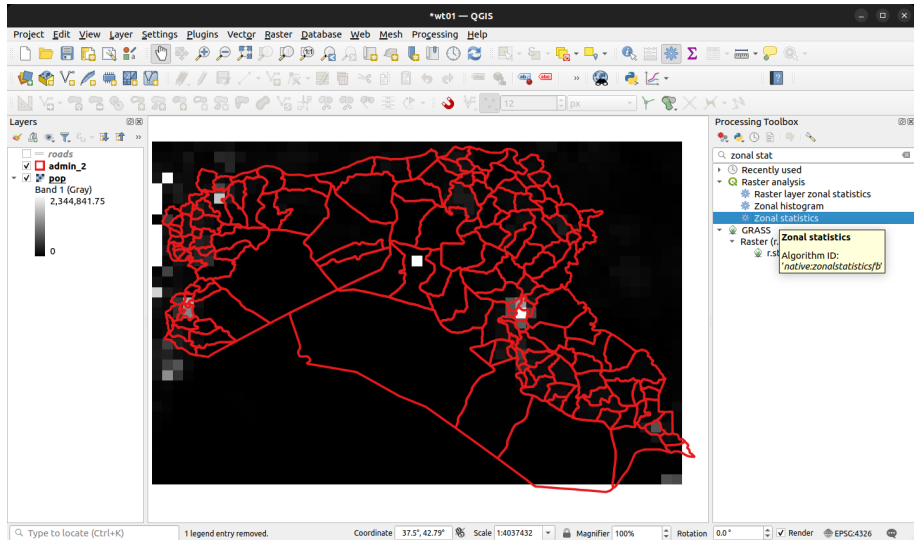
Save the file as `pop.tif`. Accept defaults for all other parameters, and click Run



A new layer called pop should appear. You can remove the original, global gpw_v4... layer

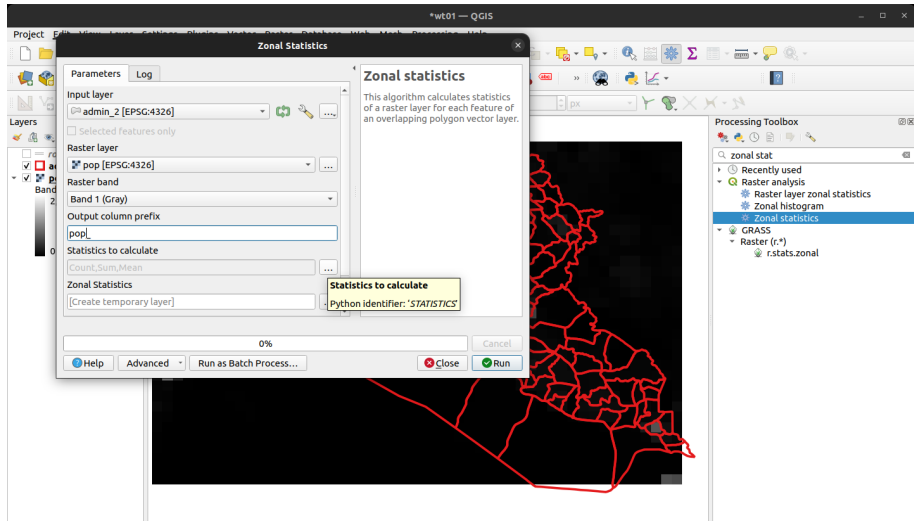


Let's **calculate the population size of each district**. Open the Processing Toolbox and open the Zonal statistics tool in the Raster Analysis menu

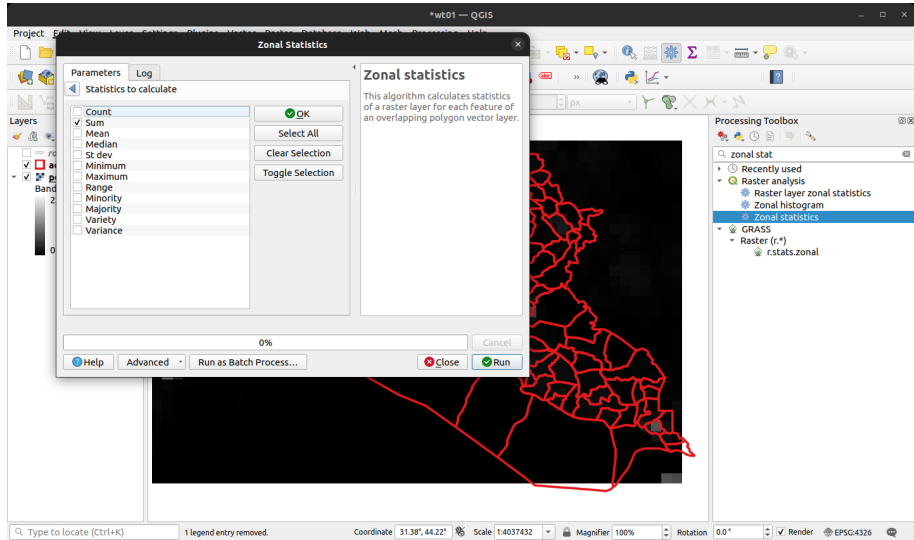


In the Zonal Statistics tool, set Input layer=admin_2, Raster layer=pop, and Output column prefix=pop_.

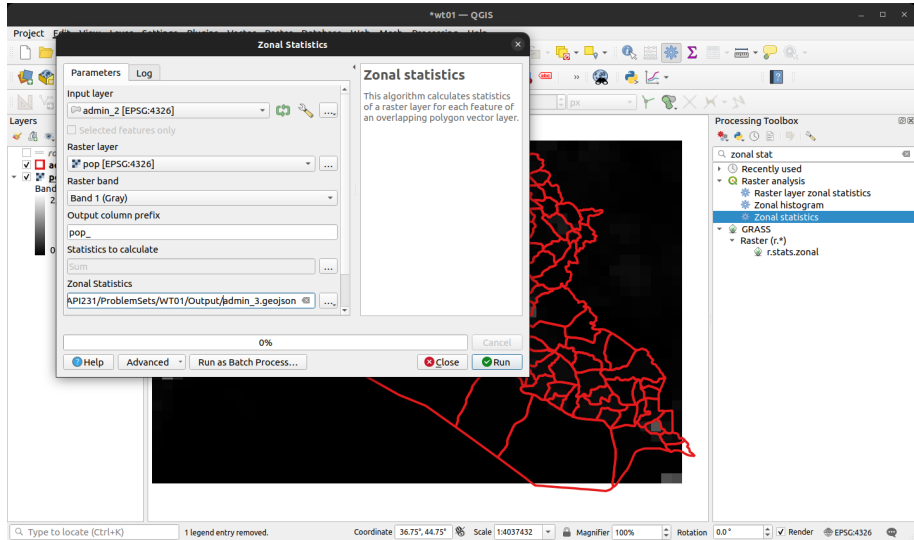
Click the [...] button next to Statistics to calculate



Check the box next to **Sum**. Leave all others un-checked. Click **OK**



Save the zonal statistics output as admin_3.geojson. Click Run

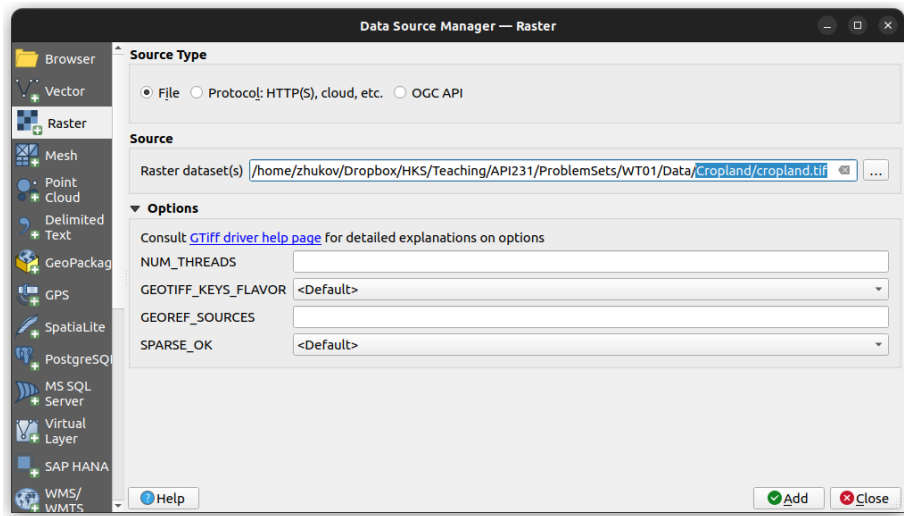


This should have generated a new layer, `admin_3`, with a new field, `pop_sum` (check the layer's attribute table)

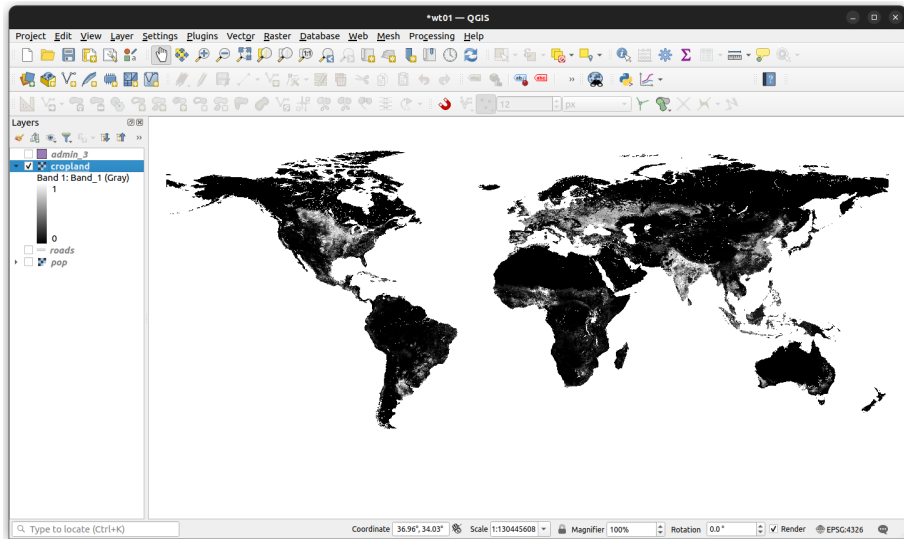
The screenshot shows the QGIS interface with the 'admin_3' layer selected in the Layers panel. The attribute table for 'admin_3' is displayed at the bottom, showing 162 features. The table includes columns for administrative data and a new 'pop_sum' column. The 'pop_sum' column for the first row shows 'Real NULL'.

NL_NAME_1	NAME_2	VARNAME_2	NL_NAME_2	TYPE_2	ENGTYPE_2	CC_2	HASC_2	roads_m	roads_ct	roads_km	area_km2	road_density	pop_sum
1	IA	AbuGhraib	NA	Kaza	District	NA	NA	30609.245...	4	30.609	411.653	0.074	pop_sum Real NULL
2	IA	AlFallujah	NA	Kaza	District	NA	NA	442953.53...	32	442.954	4975.174	0.089	87777.922...
3	IA	AlHaditha	NA	Kaza	District	NA	NA	467070.83...	28	467.071	4517.457	0.103	164966.32...
4	IA	AlQa'im	NA	Kaza	District	NA	NA	829838.41...	58	829.838	8378.068	0.099	64325.820...
5	IA	Anah	NA	Kaza	District	NA	NA	672673.22...	35	672.673	9891.892	0.068	517215.77...
6	IA	ArRamadi	NA	Kaza	District	NA	NA	473997.95...	41	473.998	6803.2	0.07	105344.98...
7	IA	ArRutbah	NA	Kaza	District	NA	NA	4789485.8...	223	4789.4861	92452.662	0.052	122194.71...
8	IA	Hit	NA	Kaza	District	NA	NA	501817.19...	33	501.817	6824.326	0.074	355362.47...
9	IA	Kadhimiya	NA	Kaza	District	NA	NA	29622.211...	2	29.622	270.137	0.11	

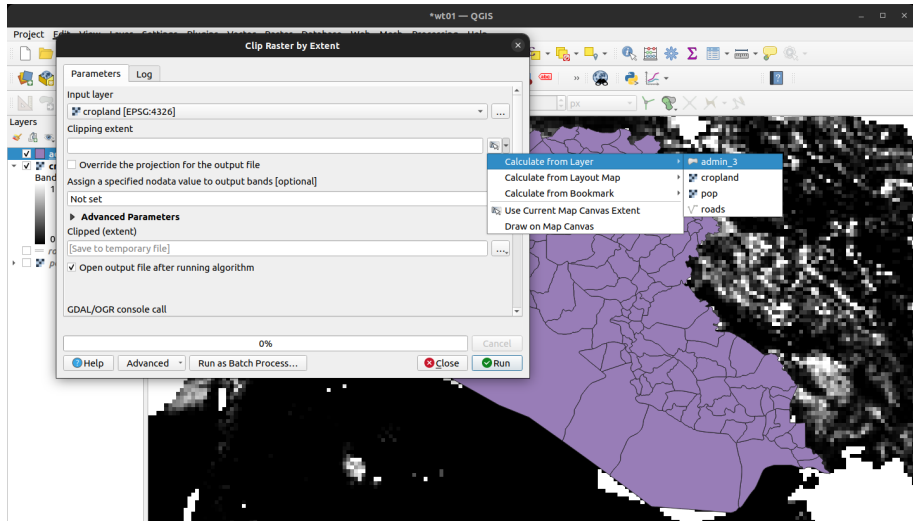
Load *cropland* raster (Layer → Add Layer → Add Raster Layer).
Open the *cropland.tif* file in Data/Cropland folder.



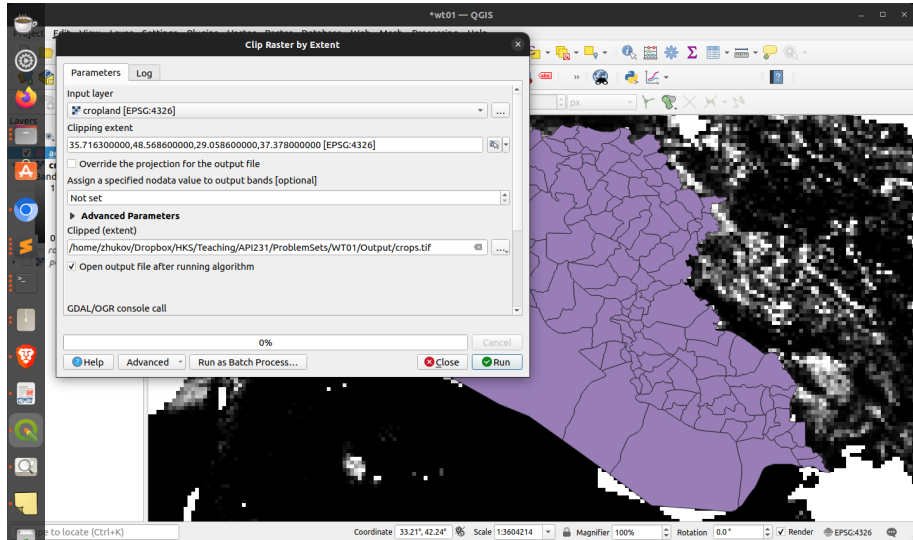
The (global) crops layer should now be visible. Let's **extract a subset** of it, as we just did for population



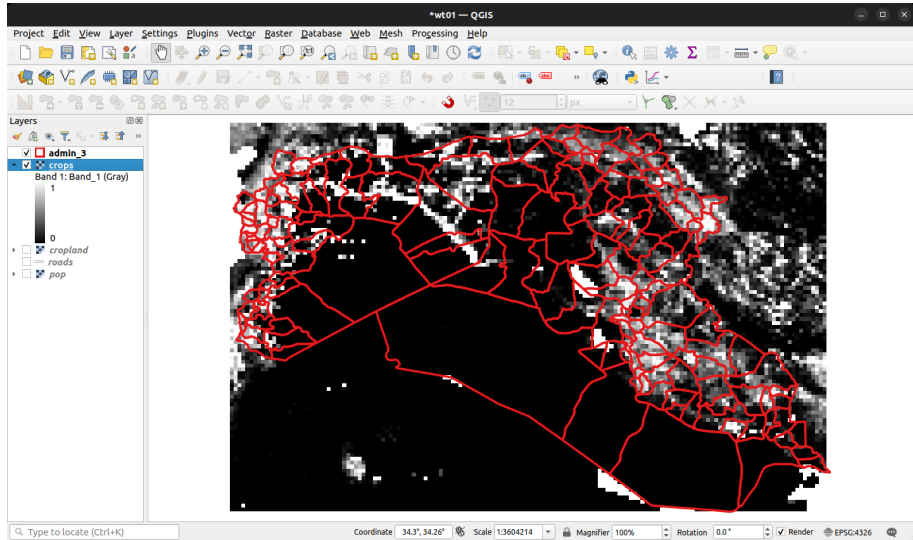
Open the Clip Raster by Extent tool (Raster → Extension → Clip Raster by Extent). Set Input layer = cropland. Click the [...] button next to Clipping extent → Calculate from Layer → admin_3



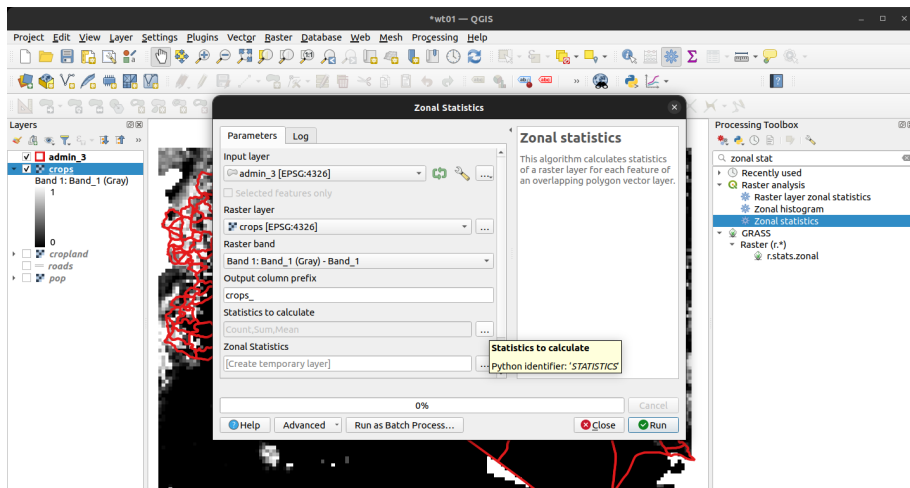
Save the output to a file called `crops.tif`



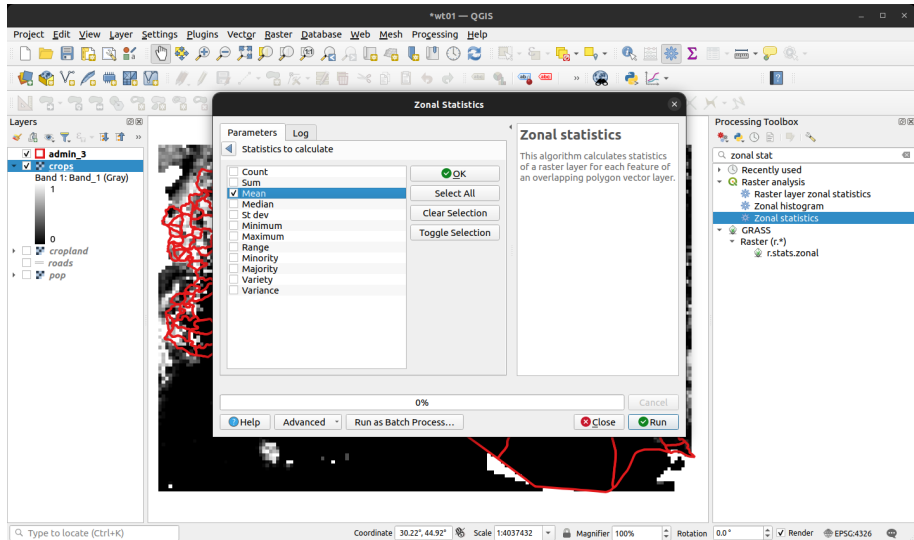
Make sure everything looks right. Remove the original, global cropland layer



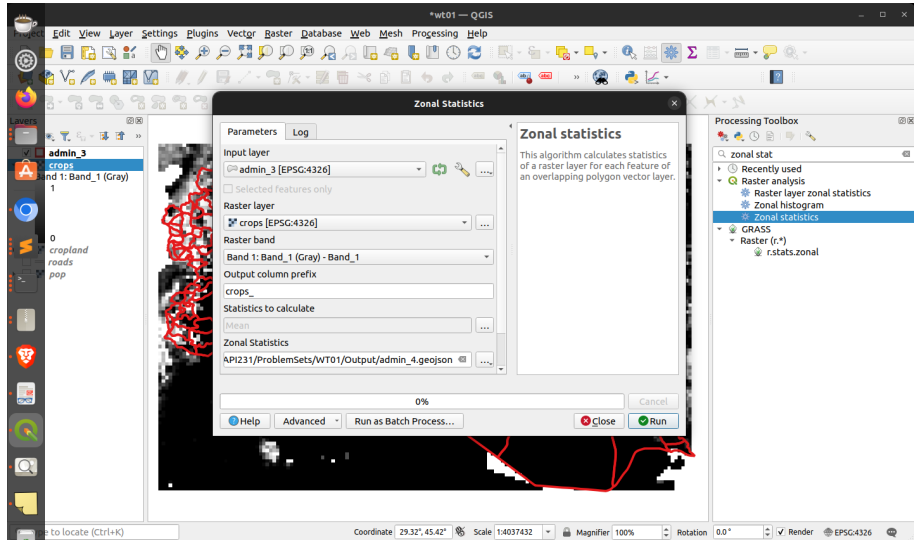
Let's **calculate crop cultivation in each district**. Open the Zonal statistics tool in Processing Toolbox. Set Input layer=admin_3, Raster layer=crops, and Output column prefix=crops_. Click the [...] button next to Statistics to calculate



Check the box next to Mean. Leave all others un-checked. Click OK, then Run on the next screen



Save the output to a file called admin_4.geojson

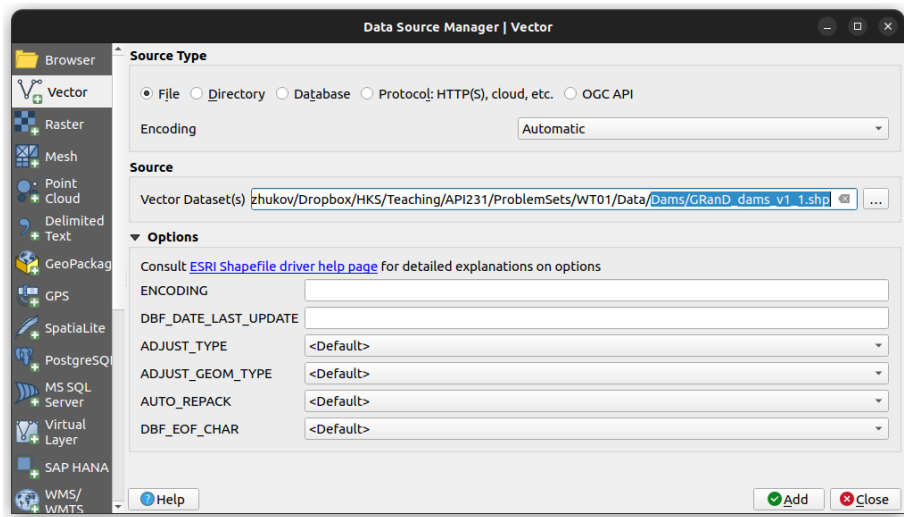


This should have generated a new field, `crops_mean` in the new `admin_4` layer
(check the layer's attribute table)

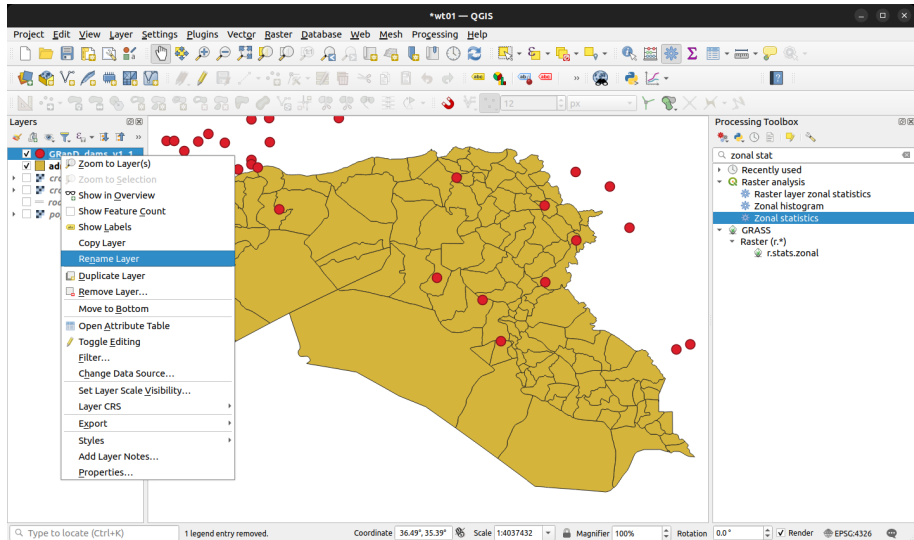
The screenshot shows the QGIS interface with the `admin_4` layer selected in the Layers panel. The attribute table for `admin_4` is displayed, showing 162 features. The table includes a new field named `crops_mean` with a data type of `Real` and a value of `NULL` for the first feature. The Processing Toolbox on the right shows the `zonal statistics` tool selected.

	NAME_2	VARNAME_2	NL_NAME_2	TYPE_2	ENGTYPE_2	CC_2	HASC_2	roads_m	roads_ct	roads_km	area_km2	road_density	pop_sum	crops_mean
1	buGhraib	NA	NA	Kaza	District	NA	NA	30609.245...	4	30.609	411.653	0.074	188651.46...	0
2	lFallujah	NA	NA	Kaza	District	NA	NA	442953.53...	32	442.954	4975.174	0.089	614252.99...	0
3	lHadhitha	NA	NA	Kaza	District	NA	NA	467070.83...	28	467.071	4517.457	0.103	87777.922...	0
4	lQa'im	NA	NA	Kaza	District	NA	NA	829838.41...	58	829.838	8378.068	0.099	164966.32...	0
5	knah	NA	NA	Kaza	District	NA	NA	672673.22...	35	672.673	9891.892	0.068	64325.820...	0
6	rRamadi	NA	NA	Kaza	District	NA	NA	473997.95...	41	473.998	6803.2	0.07	517215.77...	0
7	rRutbah	NA	NA	Kaza	District	NA	NA	4789485.8...	223	4789.4861	92452.662	0.052	105344.98...	0
8	lit	NA	NA	Kaza	District	NA	NA	501817.19...	33	501.817	6824.326	0.074	122194.71...	0
9	adhimiya	NA	NA	Kaza	District	NA	NA	29622.211...	2	29.622	270.137	0.11	355362.47...	0

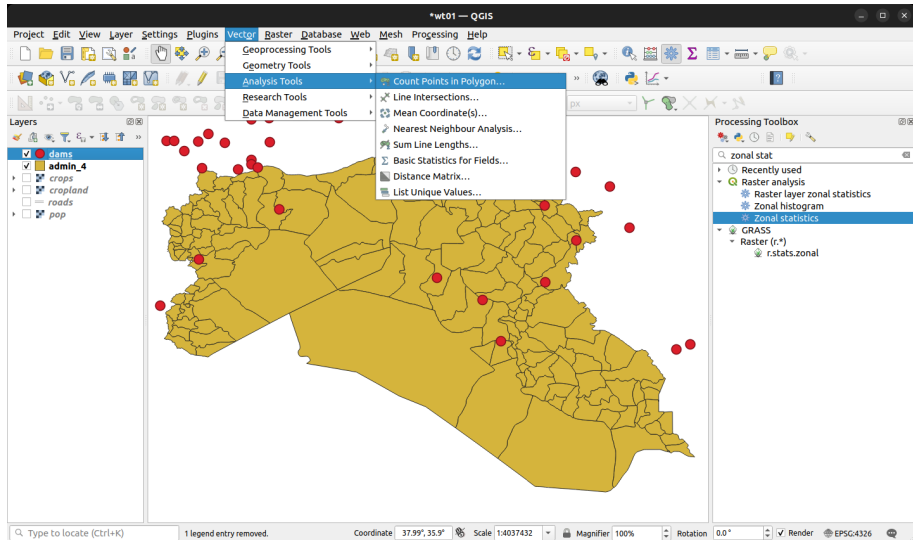
Load *hydroelectric dam* locations (Layer → Add Layer → Add Vector Layer).
Open the GRanD_dams_v1_1.shp file in Data/Dams folder.



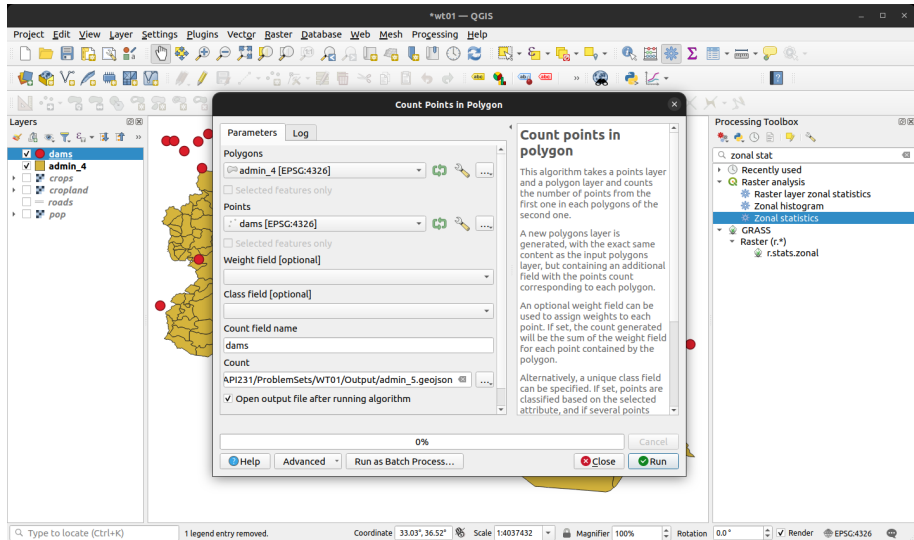
The dam locations should appear on the project window. Rename it from GRanD_dams_v1_1 to dams



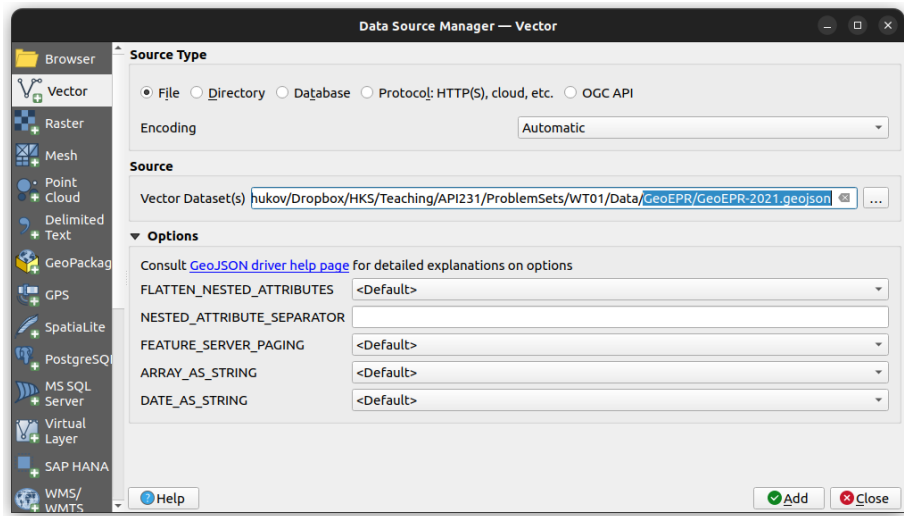
Let's calculate the **number of dams per district**. Open the Count Points in Polygon tool (Vector → Analysis Tools → Count Points in Polygon)



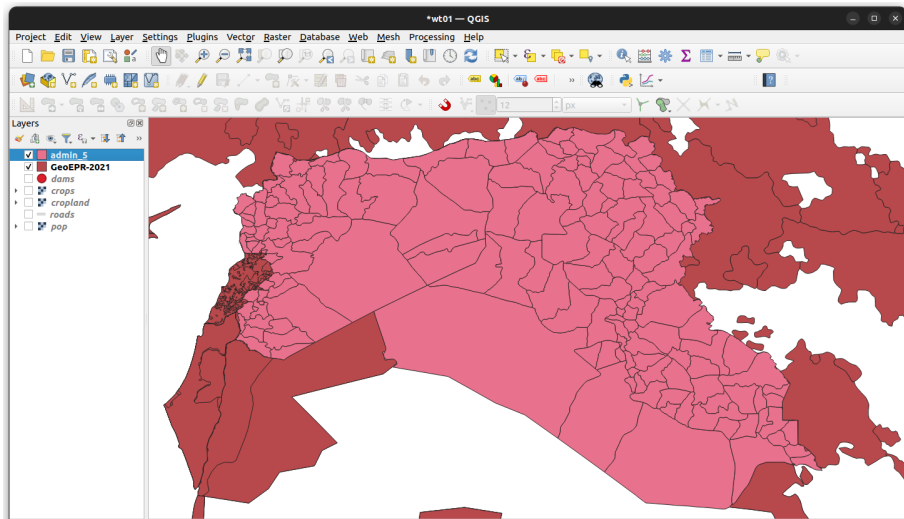
Select Polygons = admin_4, Points = dams. Name the count field dams, and save the output file as admin_5.geojson. Click Run



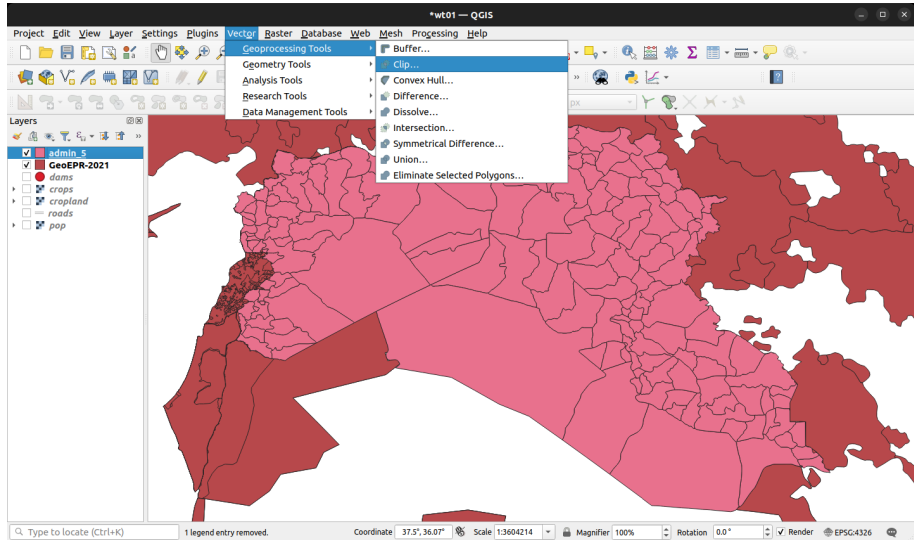
Add the *sectarian divisions* data to the project, using Add Vector Layer...\
Load the GeoEPR-2021.geojson file in Data/GeoEPR folder



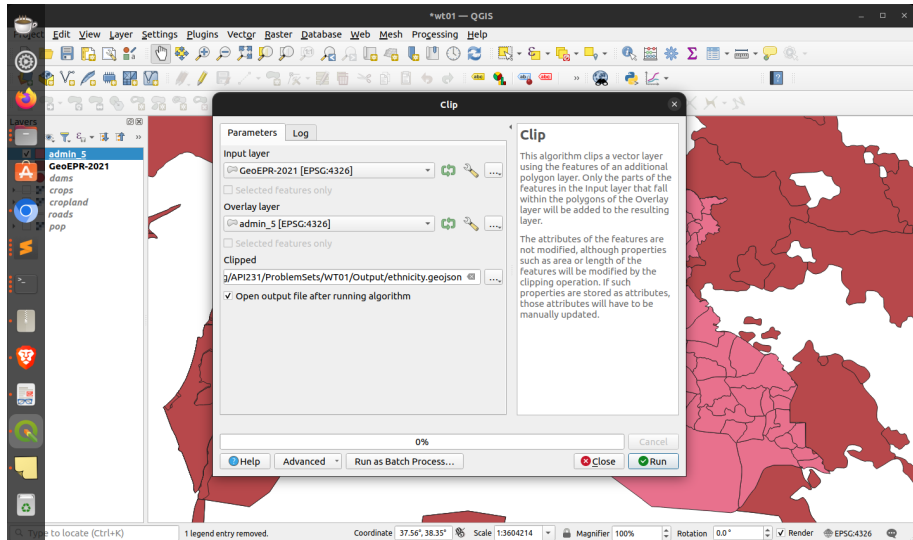
The (global) ethnicity layer should appear in the project window. Let's **calculate the proportion of each district populated by Sunni Arabs**. Let's first extract the subset that overlaps with our study area.



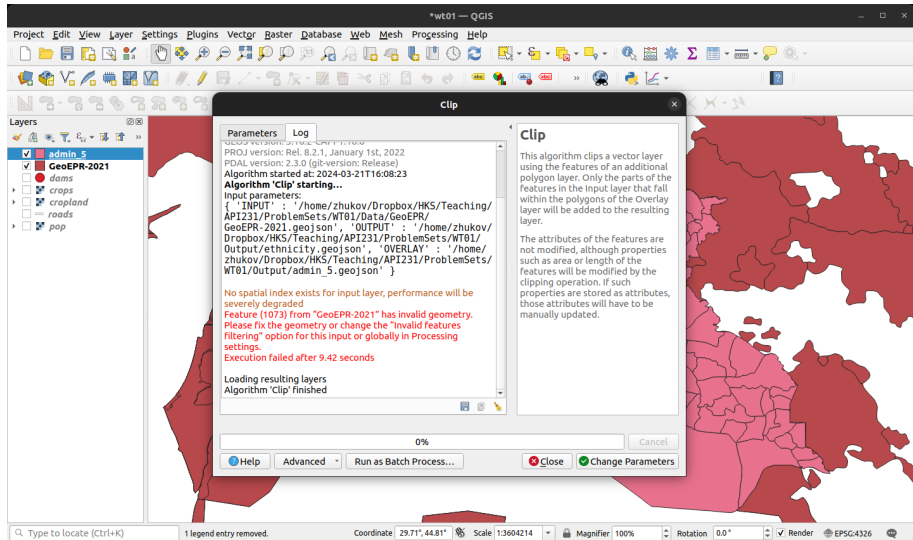
Open the Clip (vector) tool (Vector → Geoprocessing Tools → Clip)



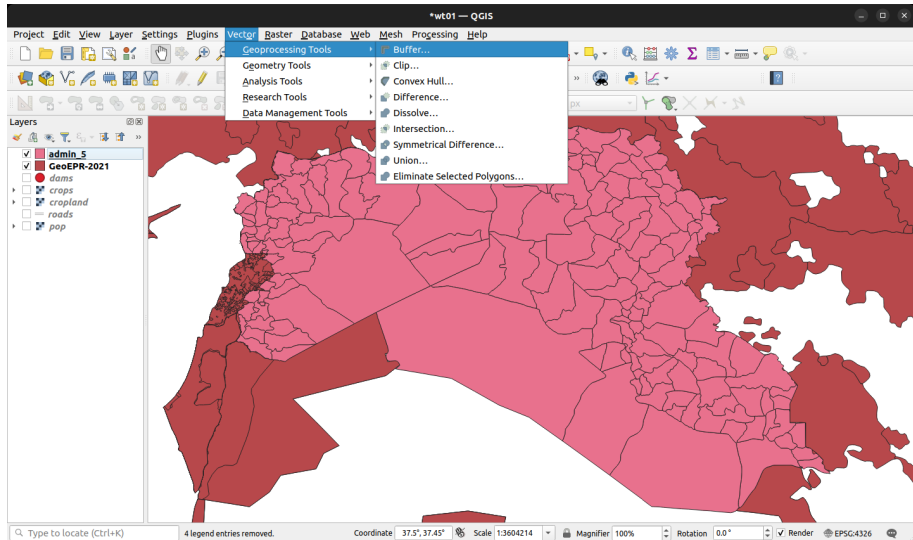
Set Input layer = GeoEPR, Overlay layer = admin_5. Save the file as ethnicity.geojson. Click Run



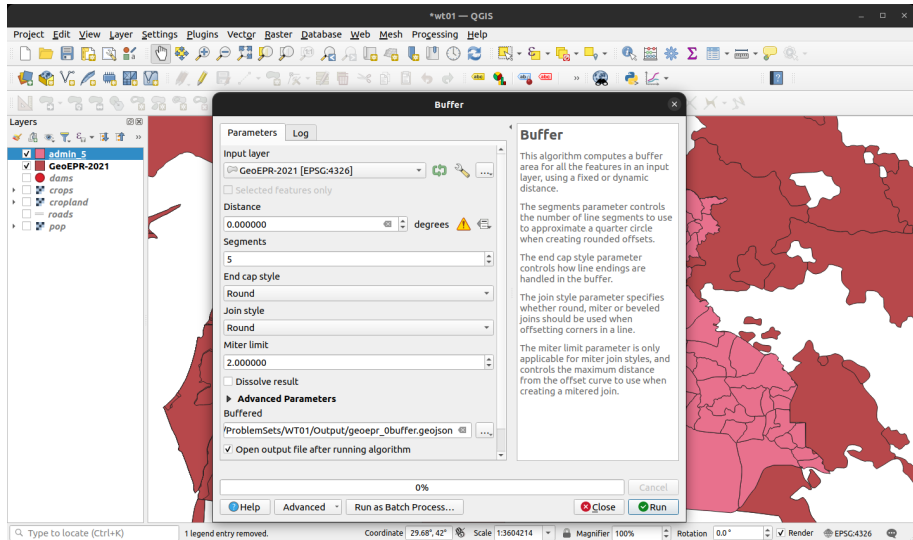
Uh oh! **Execution failed!** The error message says that GeoEPR-2021 contains “invalid geometries”. How do we fix this?



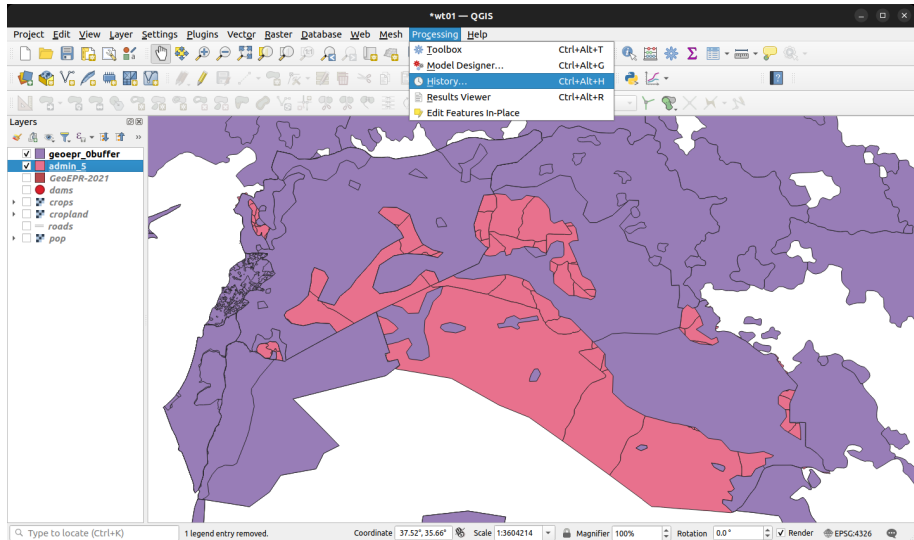
Many “invalid geometry” problems can be solved with the “zero buffer trick”. Go to Vector → Geoprocessing Tools → Buffer...



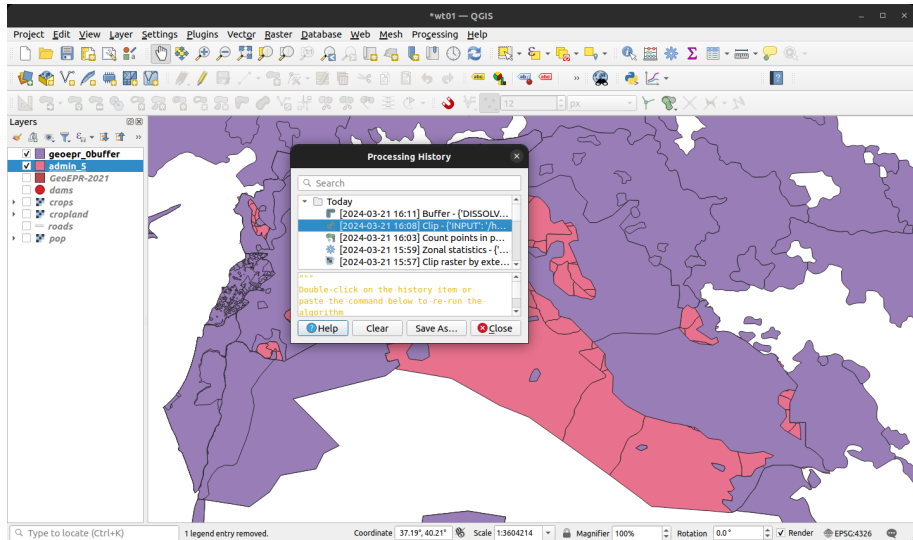
Set Input layer = GeoEPR-2021 and Distance = 0. Keep the defaults for the other parameters and save the file as `geoopr_0buffer.geojson`. Click Run



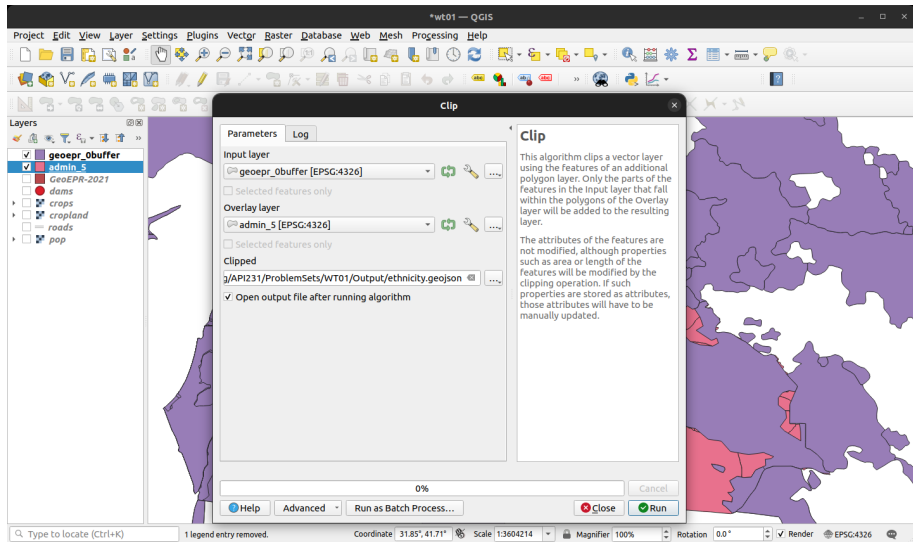
Now let's try the clipping operation again. To save time, go to Processing → History...



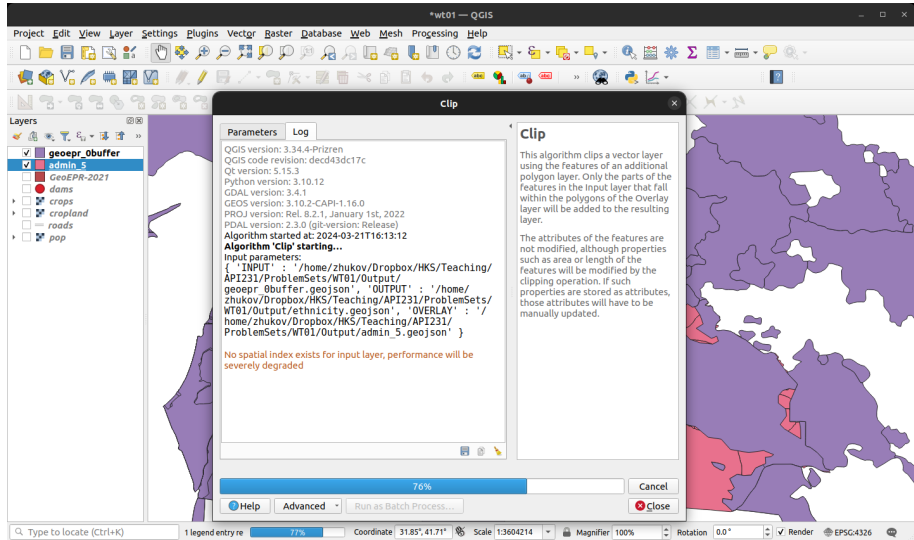
In the Processing History window, double-click on Clip (it should be the second-most recent item)



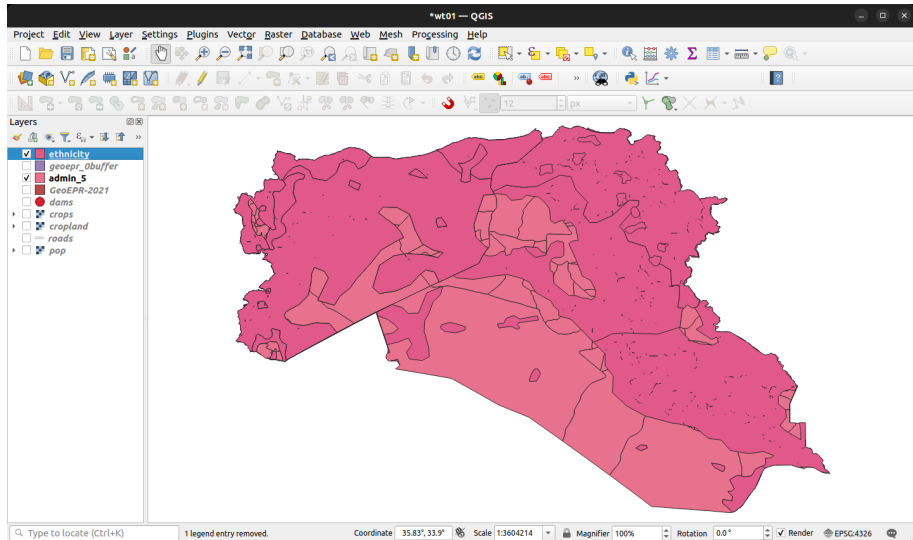
Set the Input layer = `geoopr_0buffer`. Click Run



This may take a few minutes, but should finish without errors. . .



The clipped layer should appear in the project window. You can remove the full GeoEPR layer to save memory



Look at the group field in the attribute table for ethnicity. There are multiple features for several groups (Kurds, Alawites, etc.). Let's **extract Sunni Arab polygons** only

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

Layers

- ☒ ethnicity
- ☒ admin_5
- ☐ dams
- ☐ crops
- ☐ cropland
- ☐ roads
- ☐ pop

ethnicity — Features Total: 46, Filtered: 46, Selected: 0

	gwid	statename	from	to	group	groupid	gwgroupid	umbrella	sqkm	type
1	645	Iraq	1946	2021	Shi'a	2000	64502000	NULL	108150	Regional & ...
2	645	Iraq	1946	2021	Sunni Arabs	3000	64503000	NULL	96757	Regional & ...
3	645	Iraq	1946	2021	Kurds	1000	64501000	NULL	72420	Regional & ...
4	652	Syria	1946	1966	Sunni Arabs	2000	65202000	NULL	143248	Regionally ...
5	652	Syria	1967	1967	Sunni Arabs	2000	65202000	NULL	143248	Regionally ...
6	652	Syria	1968	2021	Sunni Arabs	2000	65202000	NULL	142138	Regionally ...
7	663	Jordan	1946	1946	Jordanian ...	1000	66301000	NULL	89208	Statewide
8	663	Jordan	1947	2021	Jordanian ...	1000	66301000	NULL	89208	Statewide
9	630	Iran (Persia)	1946	2021	Arabs	1000	63001000	NULL	186282	Regional & ...

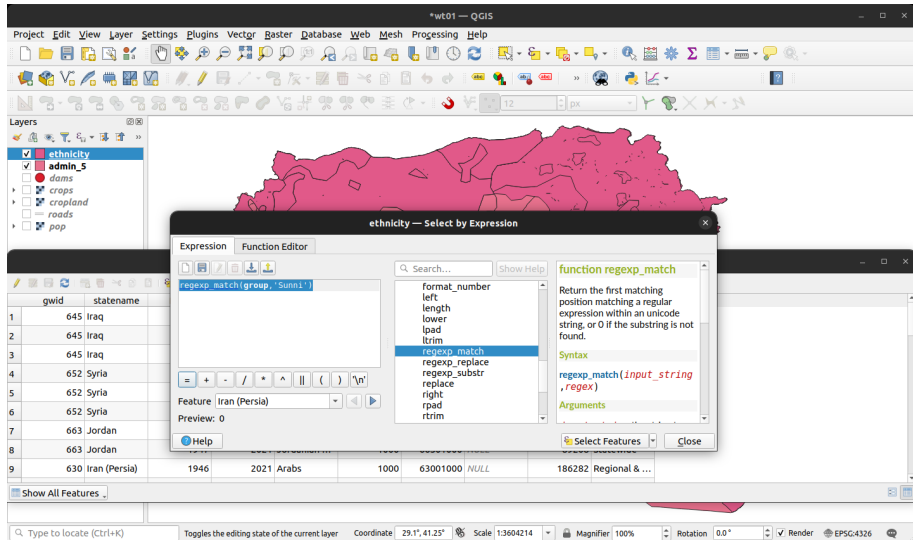
Show All Features

Open the Select by Expression tool (Edit menu → Select → Select Features by Expression)

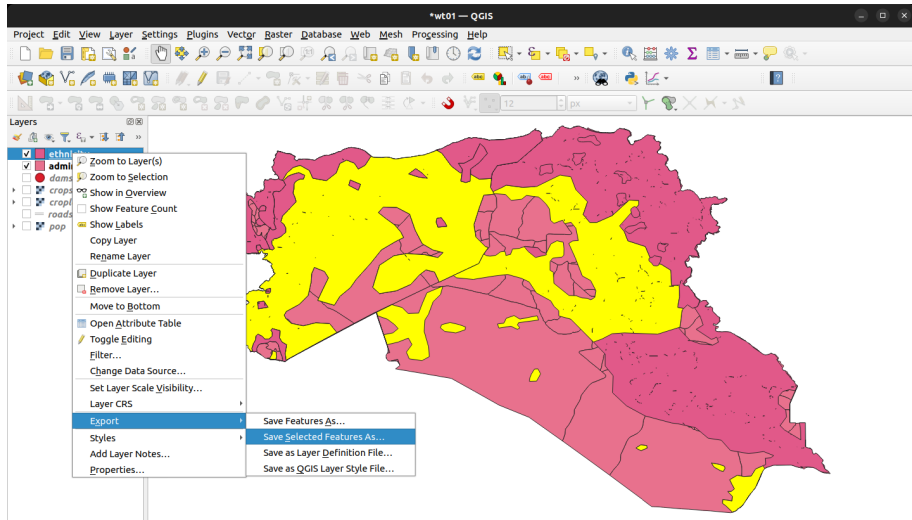
The screenshot shows the QGIS interface with the 'Select by Expression' tool open. The 'Layers' panel on the left lists several layers, with 'ethnicity' and 'admin_5' selected. The main map area displays a pink-shaded region. Below the map, a status bar indicates 'ethnicity — Features Total: 46, Filtered: 46, Selected: 0'. The 'Select by Expression' dialog box is open, showing a table of features. The table has columns: gwid, statename, groupid, gwgroupid, umbrella, sqkm, and type. The first row is highlighted, and the 'Select features using an expression' button is visible.

	gwid	statename	groupid	gwgroupid	umbrella	sqkm	type
1	645	Iraq	2000	64502000	NULL	108150	Regional & ...
2	645	Iraq	3000	64503000	NULL	96757	Regional & ...
3	645	Iraq	1000	64501000	NULL	72420	Regional & ...
4	652	Syria	2000	65202000	NULL	143248	Regionally ...
5	652	Syria	2000	65202000	NULL	143248	Regionally ...
6	652	Syria	2000	65202000	NULL	142138	Regionally ...
7	663	Jordan	1000	66301000	NULL	89208	Statewide
8	663	Jordan	1000	66301000	NULL	89208	Statewide
9	630	Iran (Persia)	1000	63001000	NULL	186282	Regional & ...

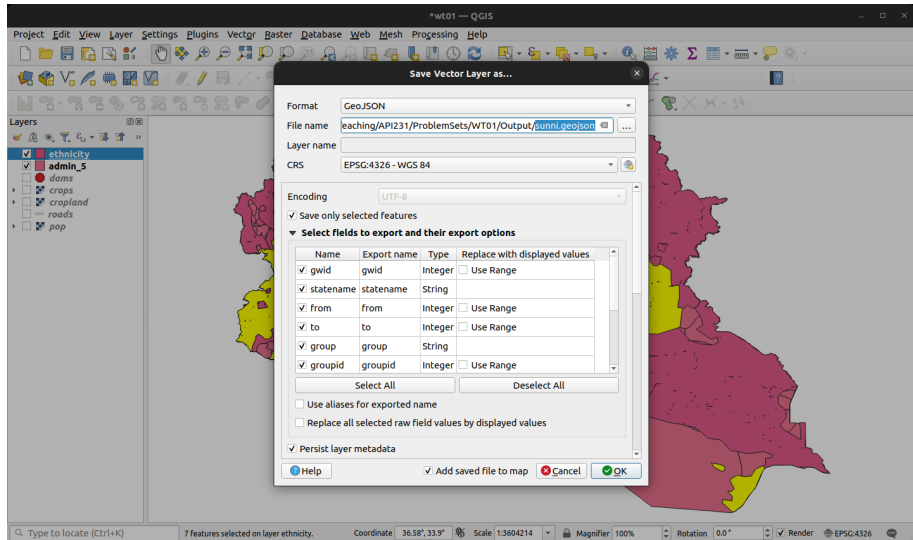
Let's use **regular expressions** to extract the polygons we need. Set Expression = `regexp_match(group, 'Sunni')`. Click Select Features



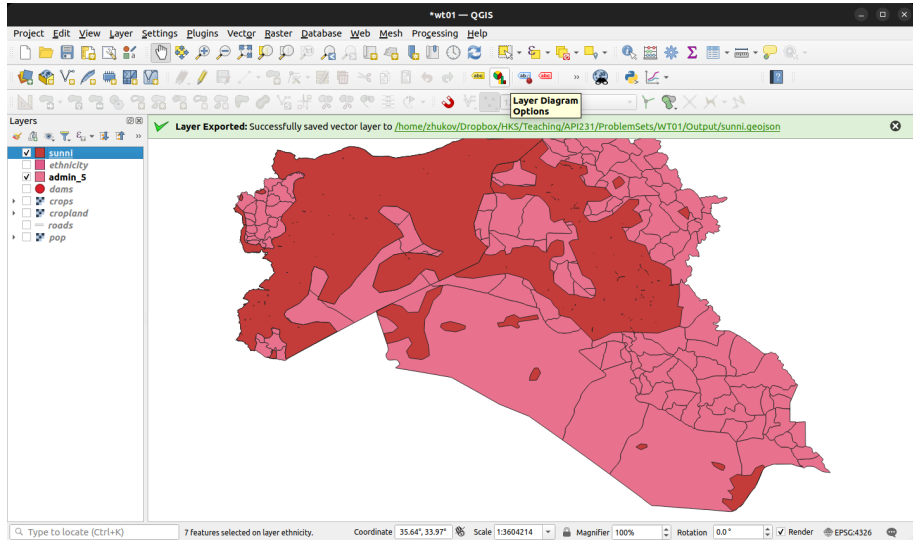
The selected polygons should appear yellow on the map. Let's **extract the selection into a new layer**. Right-click ethnicity in the Layer menu, go to Export→Save Selected Features As



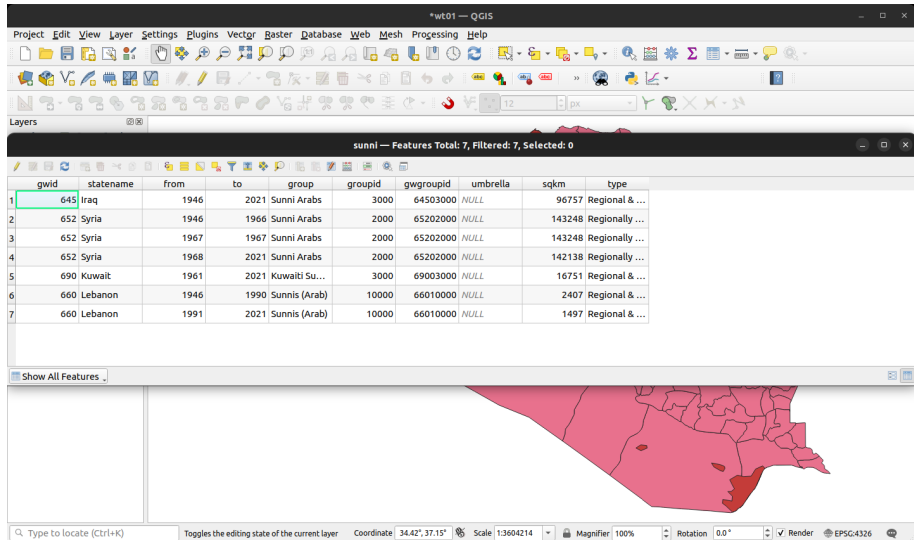
Save the layer as `sunni.geojson`. Make sure the box is checked next to Save only selected features. Click OK



The extracted selection should appear. Hide the previous layer ethnicity



The attribute table for sunni should now contain only features where the group field contains the word Sunni. Let's **consolidate these features into one polygon**



*wt01 — QGIS

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

Layers

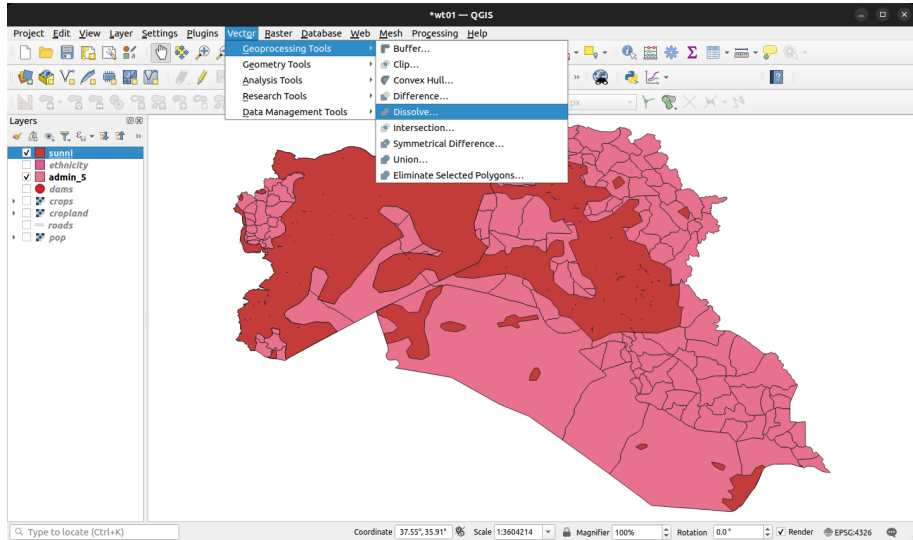
sunni — Features Total: 7, Filtered: 7, Selected: 0

	gwid	statename	from	to	group	groupid	gwgroupid	umbrella	sqkm	type
1	645	Iraq	1946	2021	Sunni Arabs	3000	64503000	NULL	96757	Regional & ...
2	652	Syria	1946	1966	Sunni Arabs	2000	65202000	NULL	143248	Regionally ...
3	652	Syria	1967	1967	Sunni Arabs	2000	65202000	NULL	143248	Regionally ...
4	652	Syria	1968	2021	Sunni Arabs	2000	65202000	NULL	142138	Regionally ...
5	690	Kuwait	1961	2021	Kuwaiti Su...	3000	69003000	NULL	16751	Regional & ...
6	660	Lebanon	1946	1990	Sunnis (Arab)	10000	66010000	NULL	2407	Regional & ...
7	660	Lebanon	1991	2021	Sunnis (Arab)	10000	66010000	NULL	1497	Regional & ...

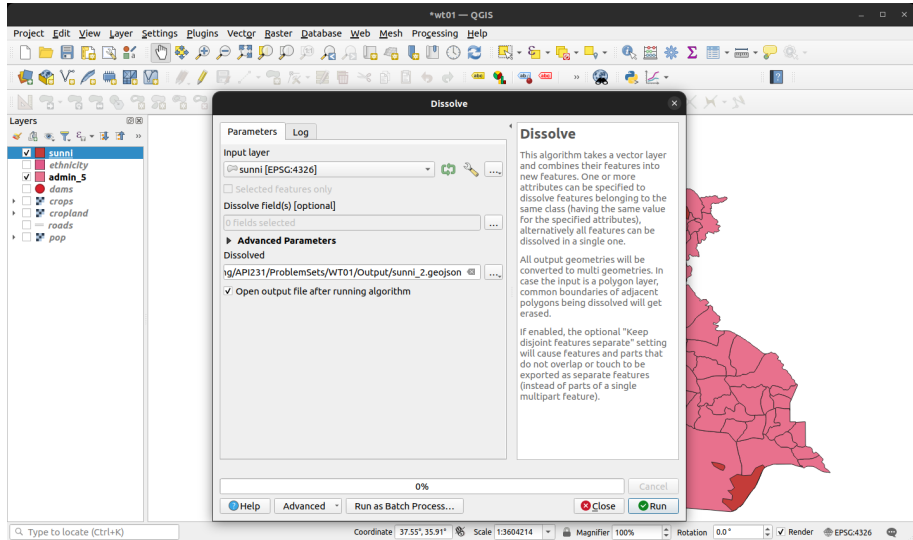
Show All Features

Type to locate (Ctrl+K) Toggles the editing state of the current layer Coordinate 34.42°, 37.15° Scale 1:3604214 Magnifier 100% Rotation 0.0° Render EPSG:4326

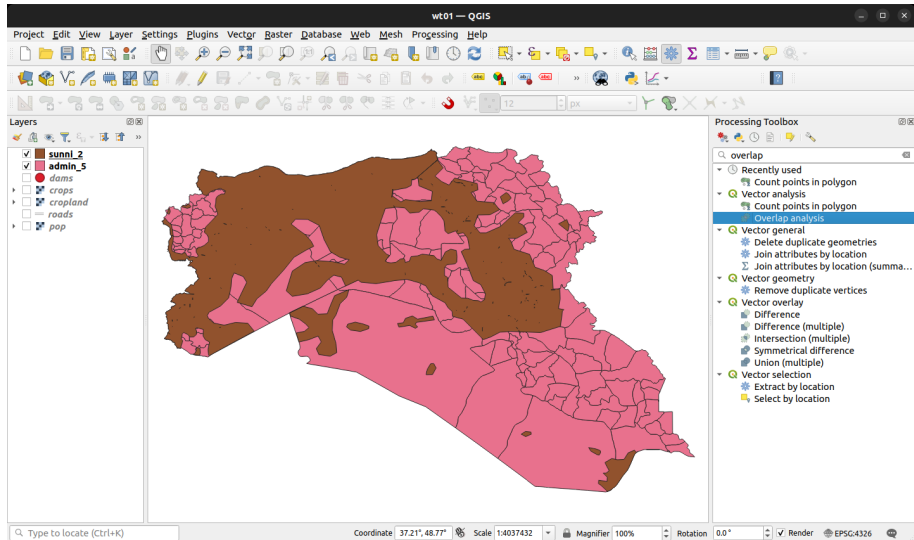
Open the Dissolve tool (Vector → Geoprocessing Tools → Dissolve).



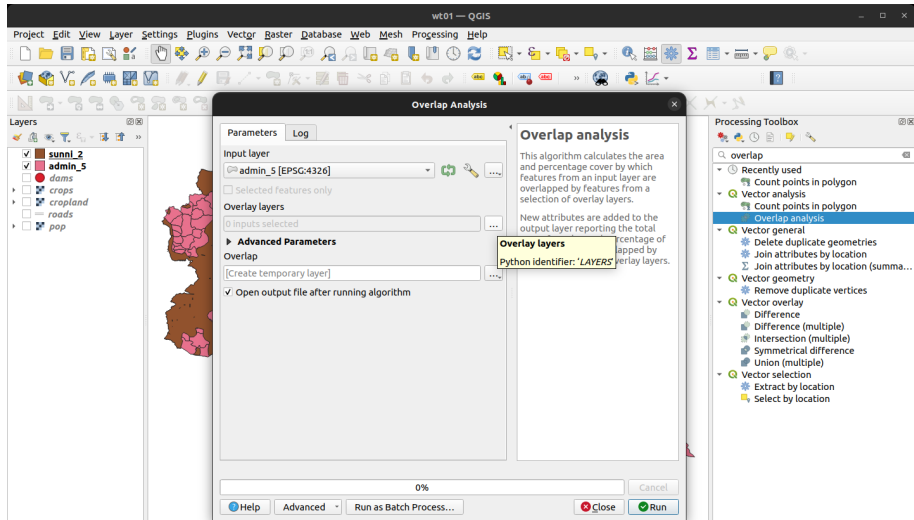
Set Input layer=sunni. Save the output as sunni_2.geojson



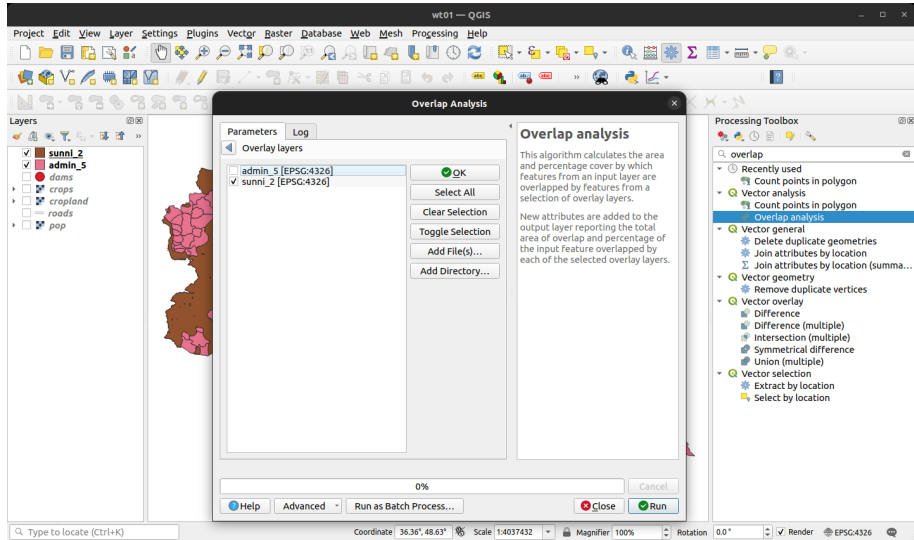
The dissolved polygon should appear in the project window. Now let's **calculate the proportion of each district populated by Sunnis**



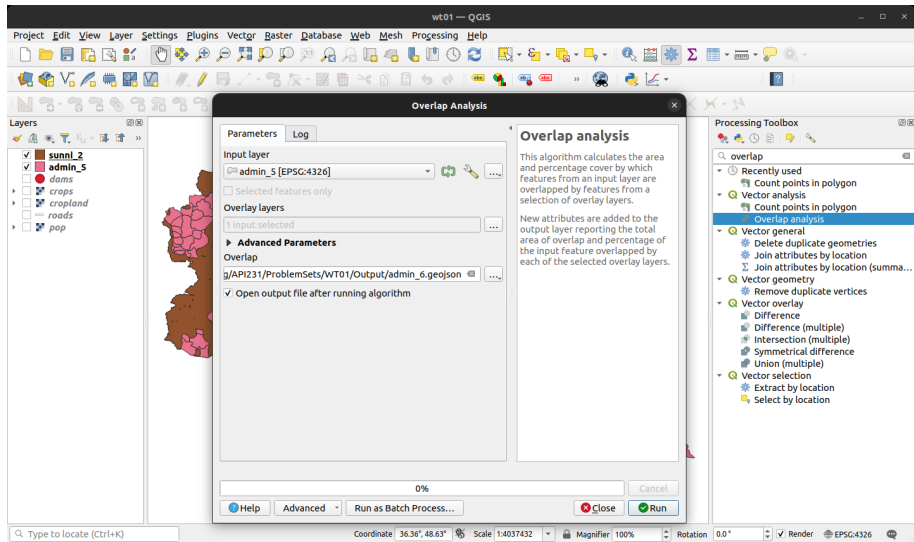
Open the **Overlap Analysis** tool (in Processing Toolbox → Vector Analysis). Set Input layer = `admin_5` and click the [...] button next to **Overlay layers**.



Check the box next to ✓ sunni_2. Click OK



Save the output as admin_6.geojson and click Run



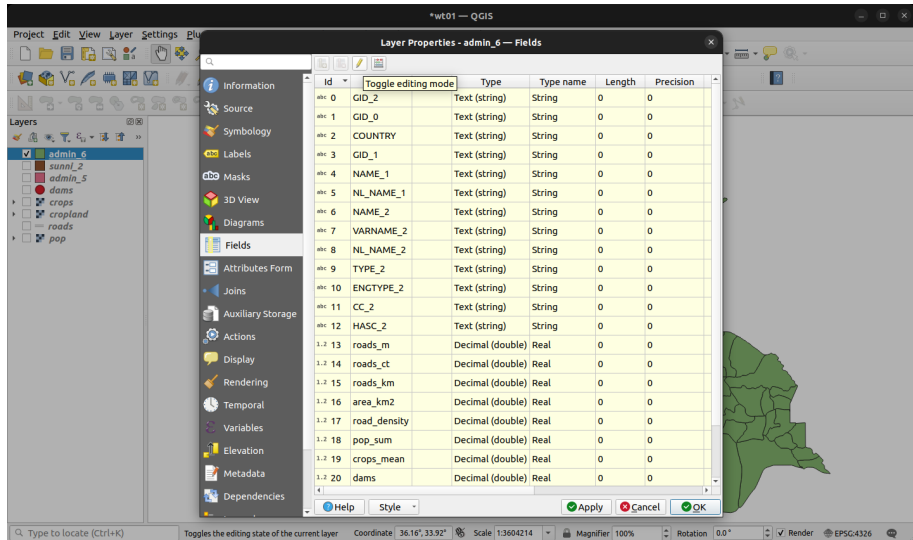
By default, the overlap fields in admin_6 will be named sunni_2_area and sunni_2_pc. Let's **change the name of these fields**

The screenshot shows the QGIS interface with the 'admin_6' layer selected. The table view displays the following data:

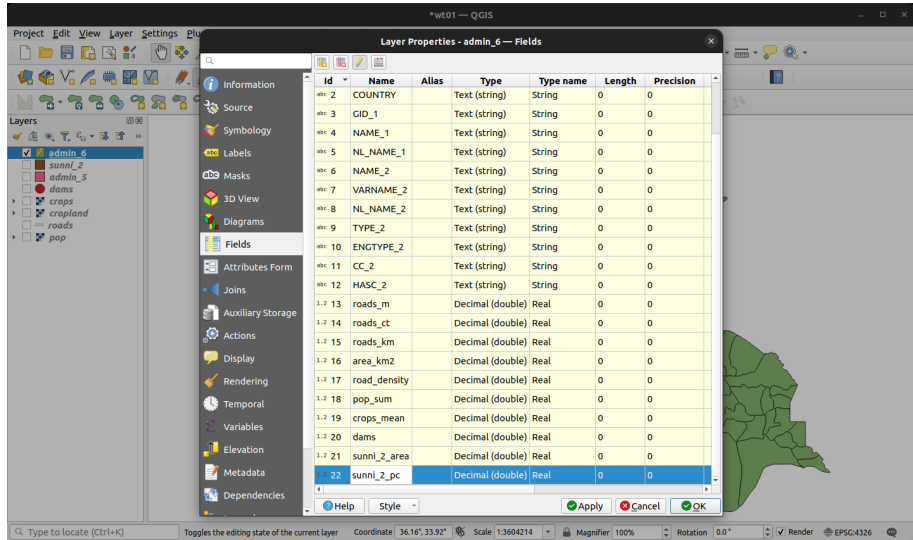
	NL_NAME_2	TYPE_2	ENGTTYPE_2	CC_2	HASC_2	roads_m	roads_ct	roads_km	area_km2	road_density	pop_sum	crops_mean	dams	sunni_2_area
1	IA	Kaza	District	NA	NA	30609.245...	4	30.609	411.653	0.074	188651.46...	0.7981666...	0	411652769...
2	IA	Kaza	District	NA	NA	442953.53...	32	442.954	4975.174	0.089	614252.99...	0.2081481...	0	388156802...
3	IA	Kaza	District	NA	NA	467070.83...	28	467.071	4517.457	0.103	87777.922...	0.0104193...	1	333214851...
4	IA	Kaza	District	NA	NA	829838.41...	58	829.838	8378.068	0.099	164966.32...	0.0559915...	0	364696736...
5	IA	Kaza	District	NA	NA	672673.22...	35	672.673	9891.892	0.068	64325.820...	0.0037536...	0	397983571...
6	IA	Kaza	District	NA	NA	473997.95...	41	473.998	6803.2	0.07	517215.77...	0.0782413...	1	412710402...
7	IA	Kaza	District	NA	NA	4789485.8...	223	4789.4861	92452.662	0.052	105344.98...	0.0002525...	0	116592763...
8	IA	Kaza	District	NA	NA	501817.19...	33	501.817	6824.326	0.074	122194.71...	0.0201562...	0	263652954...
9	IA	Kaza	District	NA	NA	29622.211...	2	29.622	270.137	0.11	355362.47...	0.5387999...	0	270137409...

The map view below the table shows a geographical area with a pinkish-red overlay, representing the 'sunni_2_area' field.

Open the layer Properties for admin_6, go to the Fields tab and click on the “pencil” button (Toggle editing mode)



Scroll down to sunni_2_pc and double-click on its name



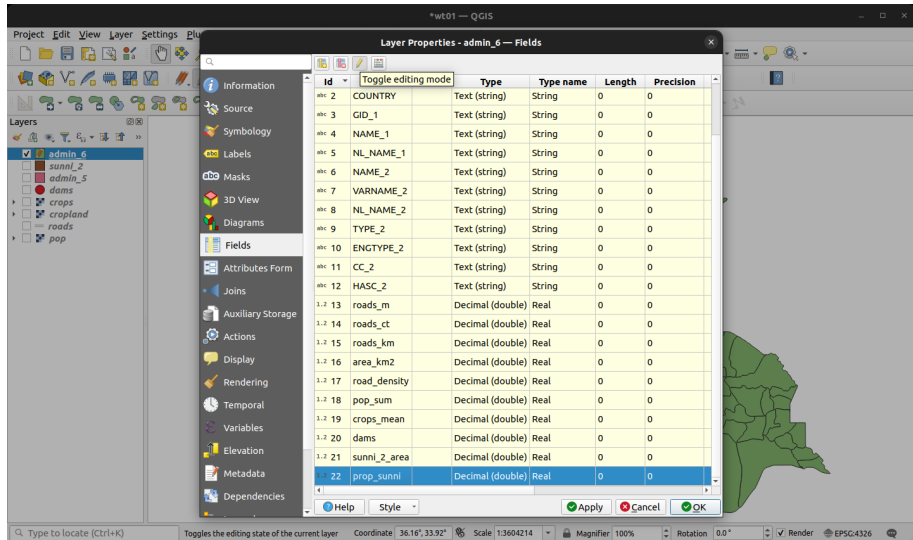
Rename the field prop_sunni

The screenshot shows the QGIS interface with the 'Layer Properties - admin_6 - Fields' dialog open. The 'Fields' tab is active, displaying a table of fields. The field 'prop_sunni' is highlighted in blue, indicating it is selected for editing. The table lists various fields with their IDs, names, aliases, types, type names, lengths, and precisions.

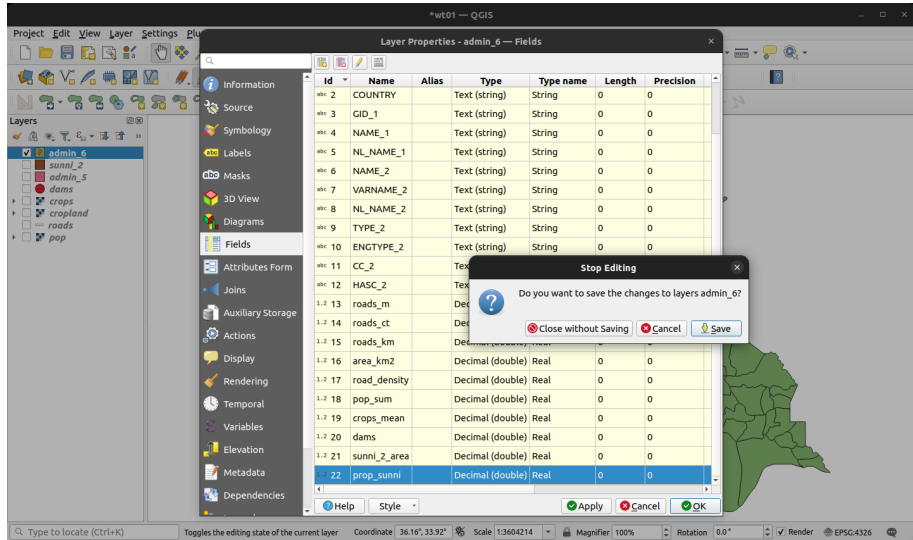
Id	Name	Alias	Type	Type name	Length	Precision
abc 2	COUNTRY		Text (string)	String	0	0
abc 3	GID_1		Text (string)	String	0	0
abc 4	NAME_1		Text (string)	String	0	0
abc 5	NL_NAME_1		Text (string)	String	0	0
abc 6	NAME_2		Text (string)	String	0	0
abc 7	VARNAME_2		Text (string)	String	0	0
abc 8	NL_NAME_2		Text (string)	String	0	0
abc 9	TYPE_2		Text (string)	String	0	0
abc 10	ENGTYPE_2		Text (string)	String	0	0
abc 11	CC_2		Text (string)	String	0	0
abc 12	HASC_2		Text (string)	String	0	0
1.2 13	roads_m		Decimal (double)	Real	0	0
1.2 14	roads_ct		Decimal (double)	Real	0	0
1.2 15	roads_km		Decimal (double)	Real	0	0
1.2 16	area_km2		Decimal (double)	Real	0	0
1.2 17	road_density		Decimal (double)	Real	0	0
1.2 18	pop_sum		Decimal (double)	Real	0	0
1.2 19	crops_mean		Decimal (double)	Real	0	0
1.2 20	dams		Decimal (double)	Real	0	0
1.2 21	sunni_2_area		Decimal (double)	Real	0	0
1.2 22	prop_sunni		Decimal (double)	Real	0	0

The 'Layers' panel on the left shows the following layers: admin_6 (checked), sunni_2, admin_5, dams, crops, cropland, roads, and pop. The status bar at the bottom indicates the coordinate is 36.16°, 33.92°, scale is 1:3604214, magnifier is 100%, rotation is 0.0°, and the render status is checked.

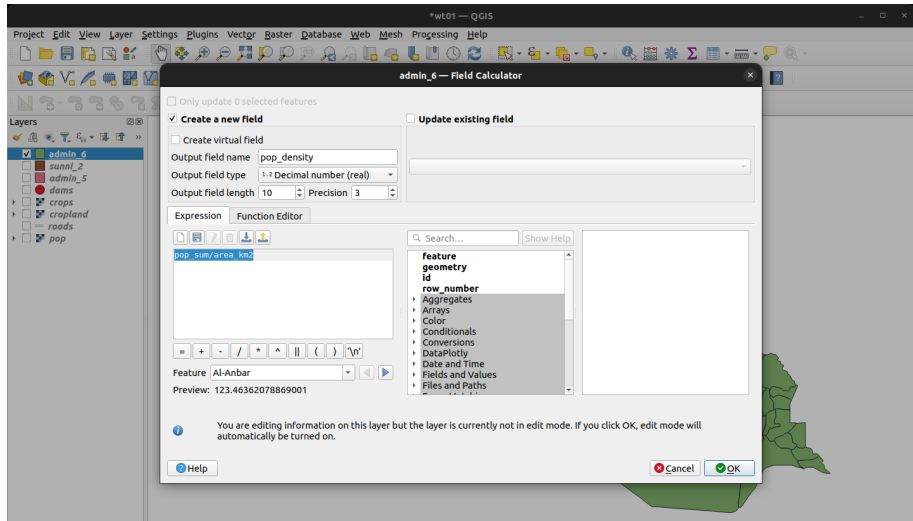
Hit Enter key (or Return key) to commit the name change. Click on the “pencil” again to leave editing mode



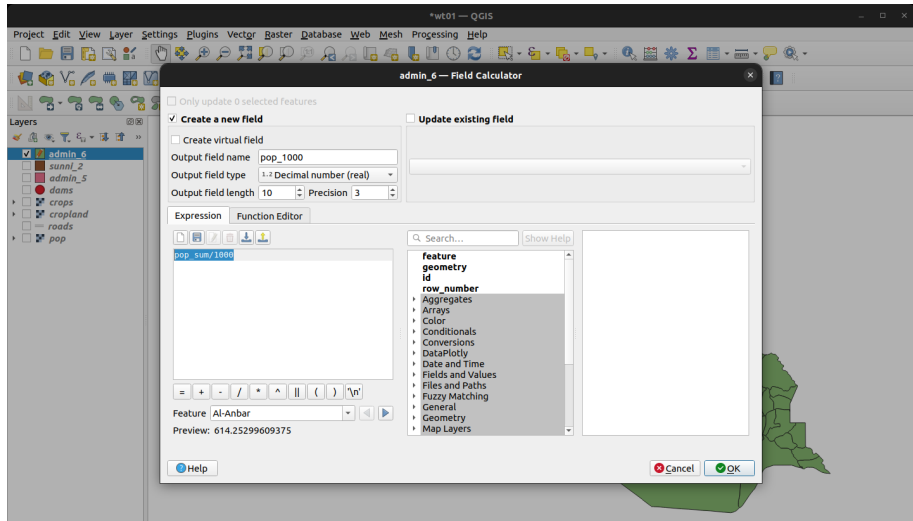
When prompted, save the changes you just made to admin_6



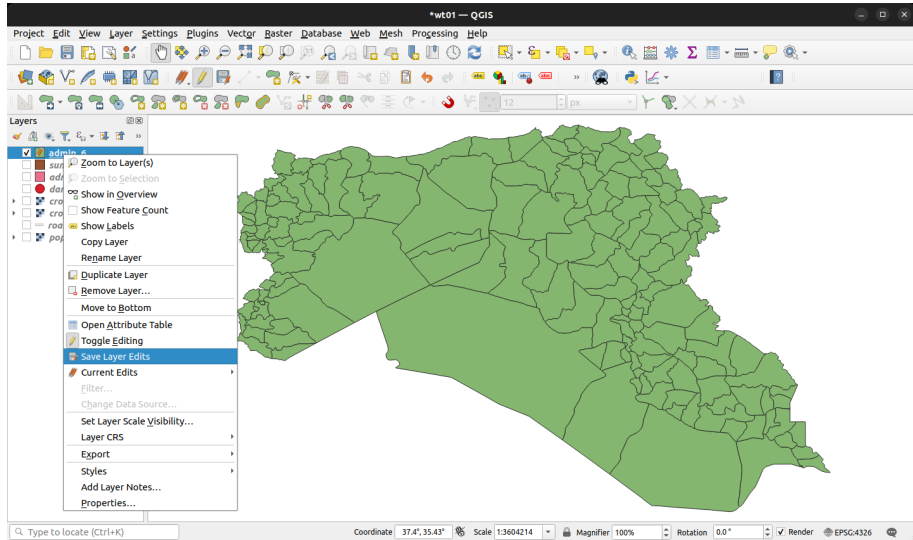
While we're at it, let's create a couple additional fields that could come in handy in the analysis. Create a **population density** field, named `pop_density`, of type Decimal number (real), with Expression set to `pop_sum/area_km2`



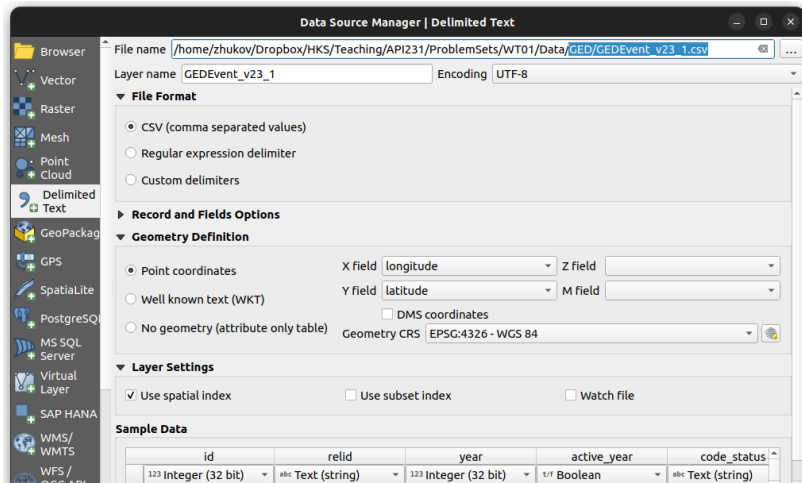
Let's also create a **rescaled population** field (1000's of residents). Name it pop_1000, with type Decimal number (real). Set Expression to pop_sum/1000



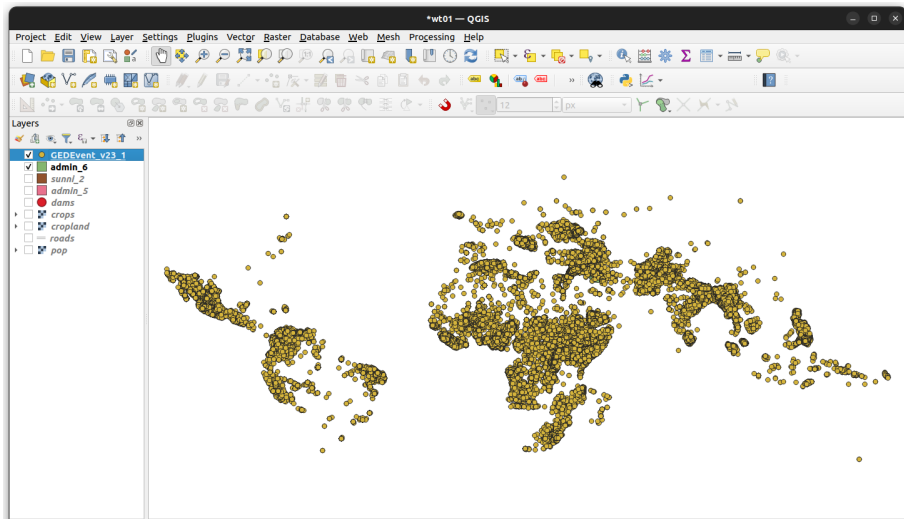
Save the layer edits to admin_6!



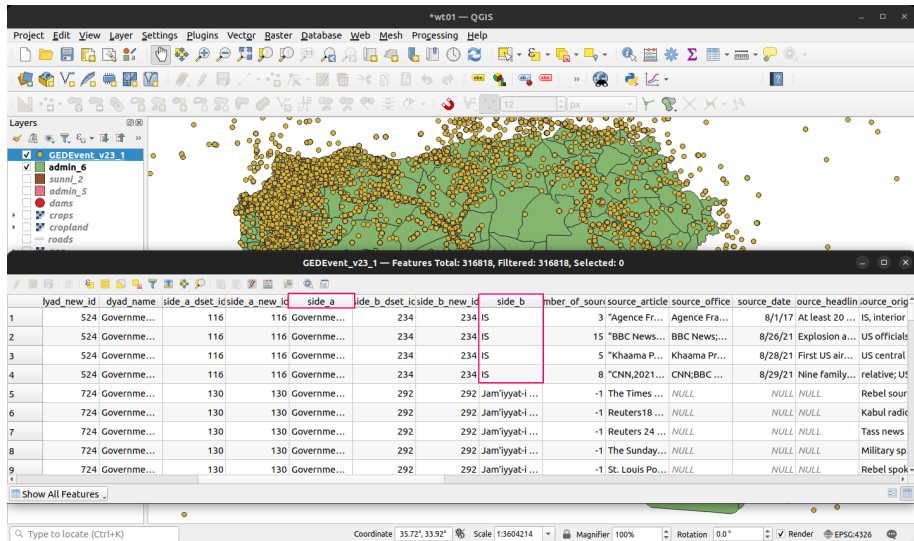
Add the *Islamic State violence data* to the project, using Add Delimited Text Layer.... Load the GEDEvent_v23.csv file in Data/GED folder. Set X field = longitude and Y field = latitude. Check the box next to ✓Use spatial index



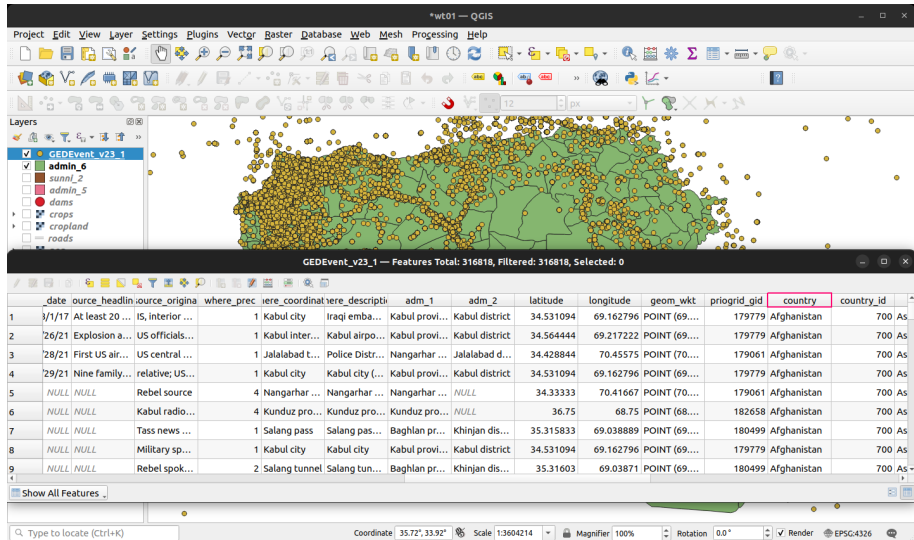
The (global) GED violent events layer should appear. There are several hundred thousand points here. We need to **extract events in Syria and Iraq involving the Islamic State**



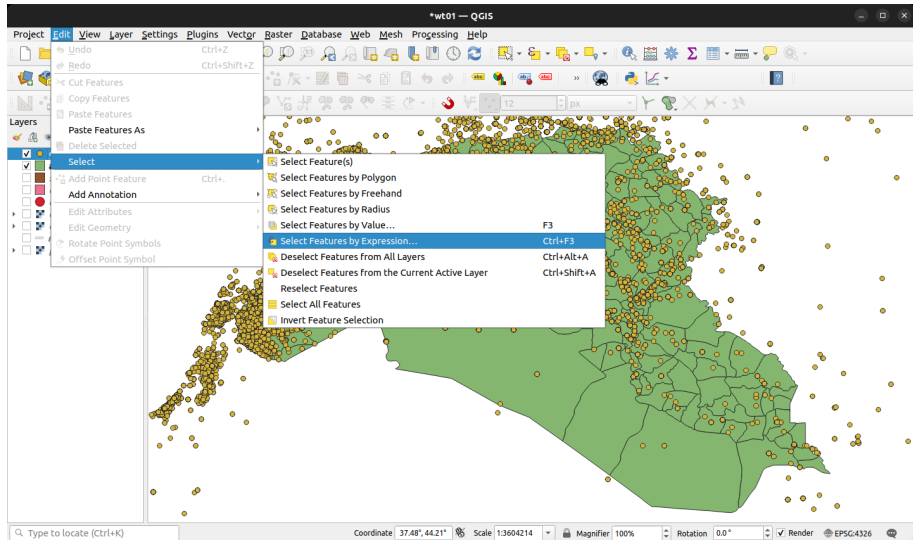
To figure out how to extract this data subset (by actor and location), let's explore the attribute table. We see fields for `side_a` and `side_b`, which list actors



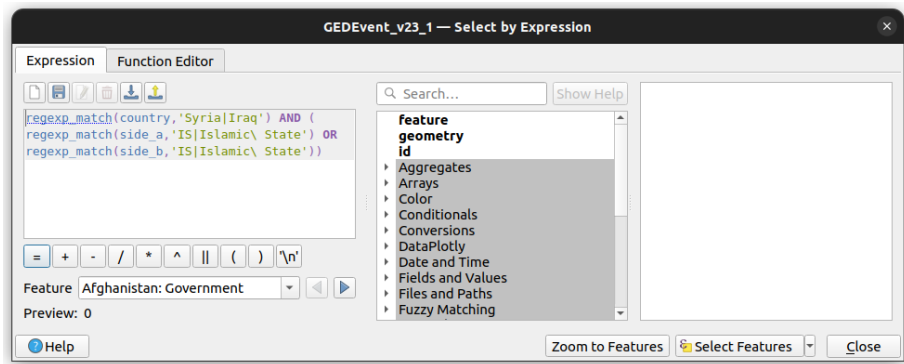
There is also a country field. So, we need to select points where country is Iraq or Syria and the Islamic State is side_a (or side_b)



Highlight the GED layer and go to Edit → Select → Select by Expression...

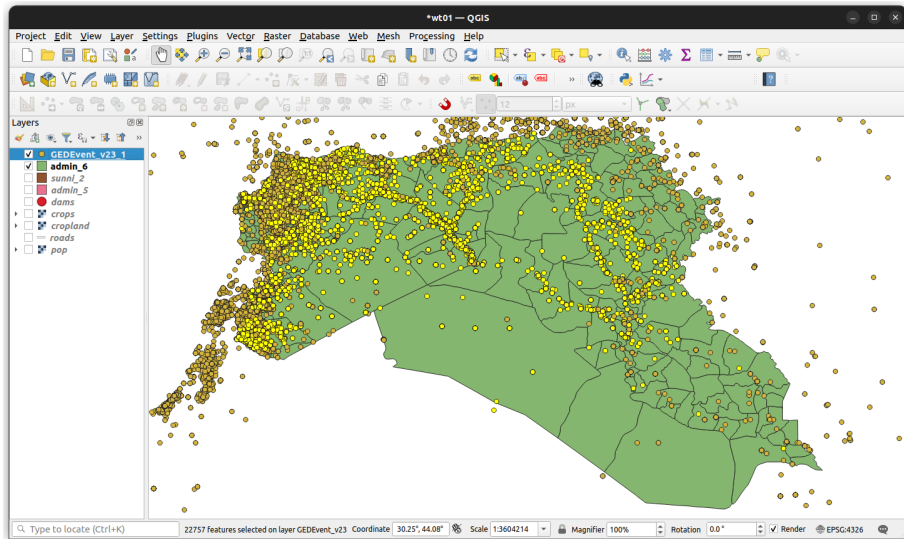


Let's combine regular expressions with logical operators. Set Expression to `regexp_match(country,'Syria|Iraq')` AND (`regexp_match(side_a,'IS|Islamic State')` OR `regexp_match(side_b,'IS|Islamic State')`)

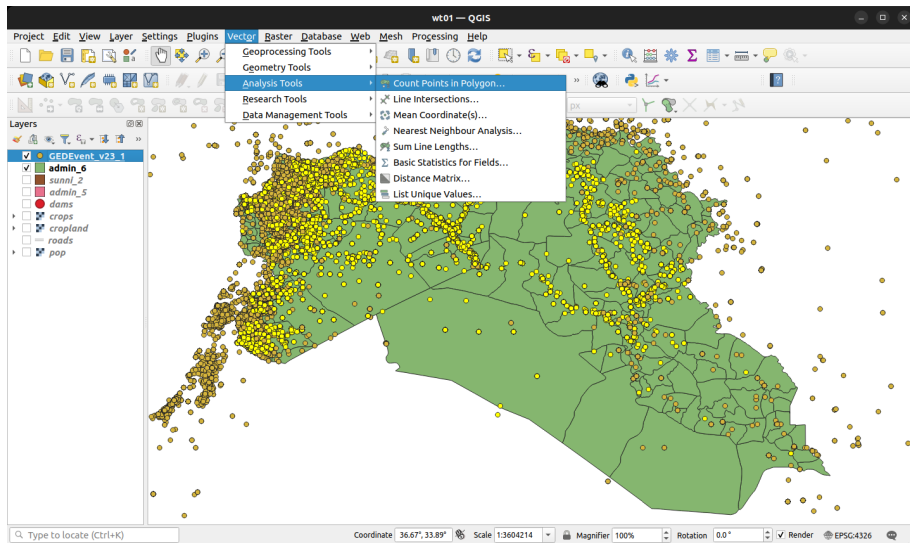


The vertical slash | is a regular expression for “OR”. So, in English this expression means ‘features where the field country contains “Iraq or Syria” and either side_a or side_b contains “IS or Islamic State”’

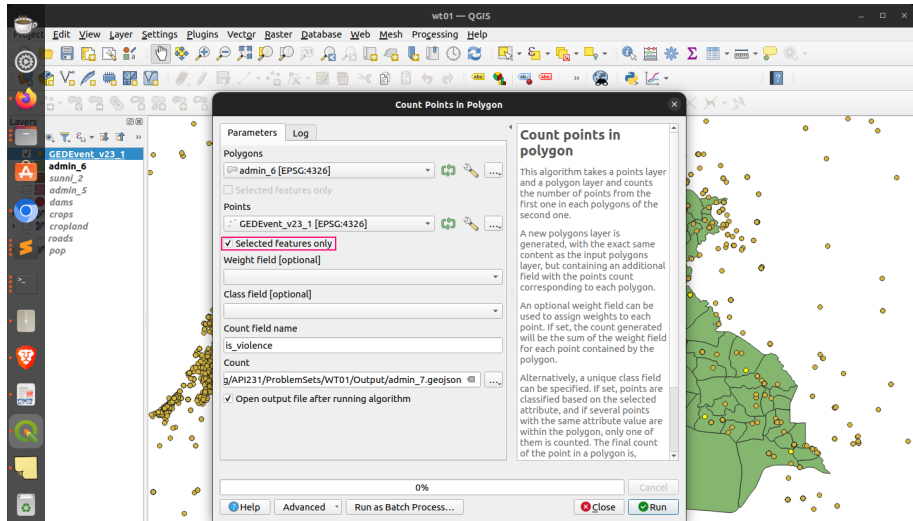
This should select a little over 22,000 events. Now let's calculate the **number of ISIS attacks per district**



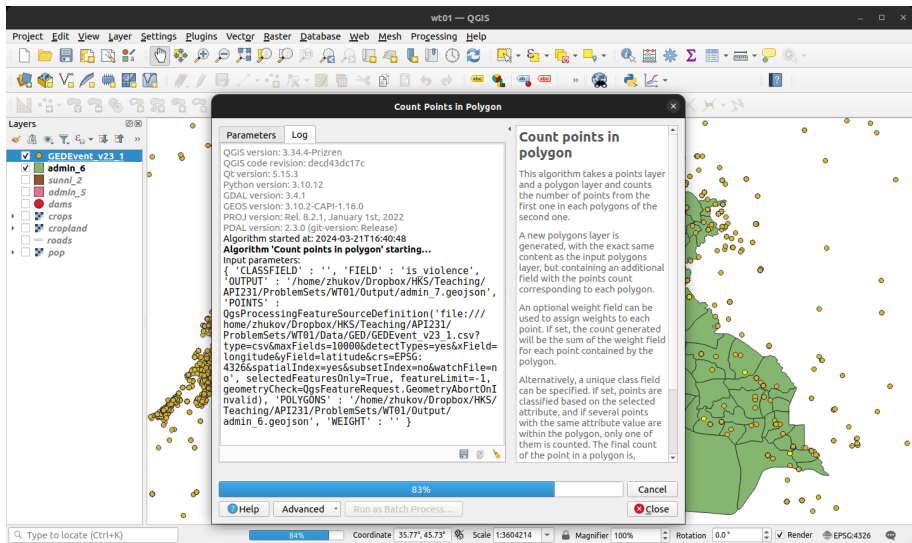
Open the Count Points in Polygon tool



Select Polygons = admin_6, Points = GEDEvent_v23_1. Make sure the box is checked next to Selected Features Only for the points. Name the count field is_violence, and save the output file as admin_7.geojson. Click Run



This may take a minute or two to run



The attribute table for this new layer should include all the new fields we have generated. Let's now **export the table as CSV** for further analysis

*wt01 — QGIS

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

Layers

- ☒ admin_7
- ☐ GEDEvent_v23_1
- ☐ admin_6
- ☐ sunni_2
- ☐ admin_5
- ☐ dams
- ☐ crops
- ☐ cropland

admin_7 — Features Total: 162, Filtered: 162, Selected: 0

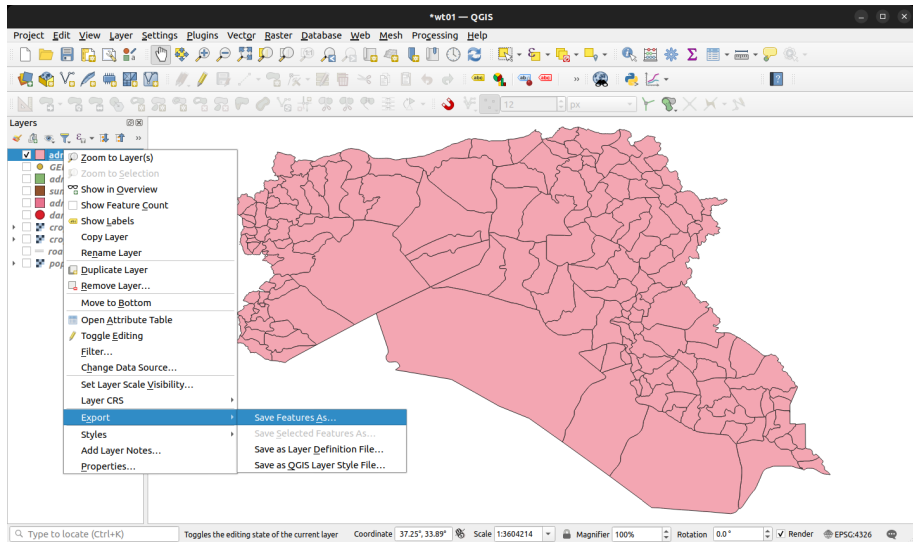
	HASC_2	roads_m	roads_ct	roads_km	area_km2	road_density	pop_sum	crops_mean	dams	sunni_2_area	prop_sunni	pop_density	pop_1000	is_violence
1	4A	30609.245...	4	30.609	411.653	0.074	188651.46...	0.7981666...	0	411652769...	99.999944...	458.278	188.651	8
2	4A	442953.53...	32	442.954	4975.174	0.089	614252.99...	0.2081481...	0	388156802...	78.018732...	123.464	614.253	396
3	4A	467070.83...	28	467.071	4517.457	0.103	87777.922...	0.0104193...	1	333214851...	73.761597...	19.431	87.778	49
4	4A	829838.41...	58	829.838	8378.068	0.099	164966.32...	0.0559915...	0	364696736...	43.529933...	19.69	164.966	122
5	4A	672673.22...	35	672.673	9891.892	0.068	64325.820...	0.0037536...	0	397983571...	40.233312...	6.503	64.326	28
6	4A	473997.95...	41	473.998	6803.2	0.07	517215.77...	0.0782413...	1	412710402...	60.664154...	76.025	517.216	307
7	4A	4789485.8...	223	4789.4861	92452.662	0.052	105344.98...	0.0002525...	0	116592763...	12.611077...	1.139	105.345	245
8	4A	501817.19...	33	501.817	6824.326	0.074	122194.71...	0.0201562...	0	263652954...	38.634283...	17.906	122.195	103
9	4A	29622.211...	2	29.622	270.137	0.11	355362.47...	0.5387999...	0	270137409...	100.00000...	1315.49	355.362	37

Show All Features

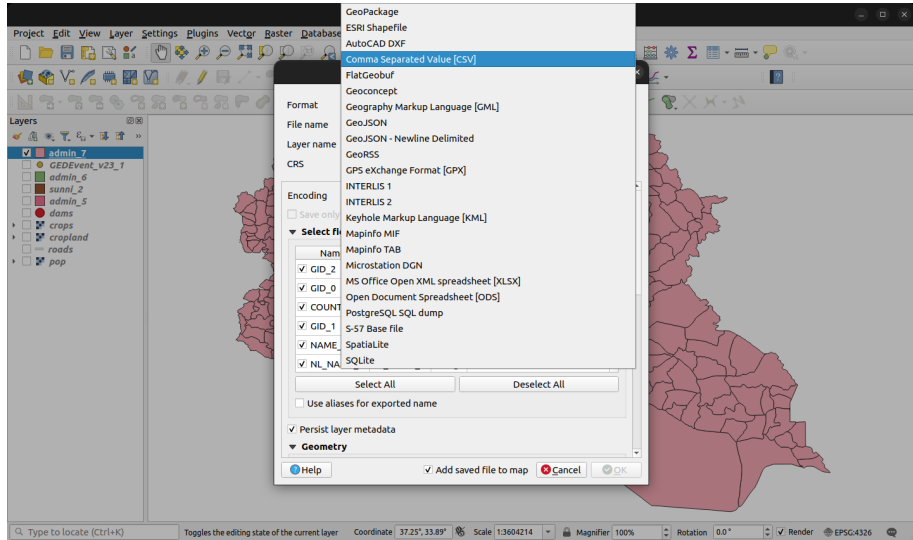
Type to locate (Ctrl+K)

Coordinate 36.62°, 34° Scale 1:3604214 Magnifier 100% Rotation 0.0° Render EPSG:4326

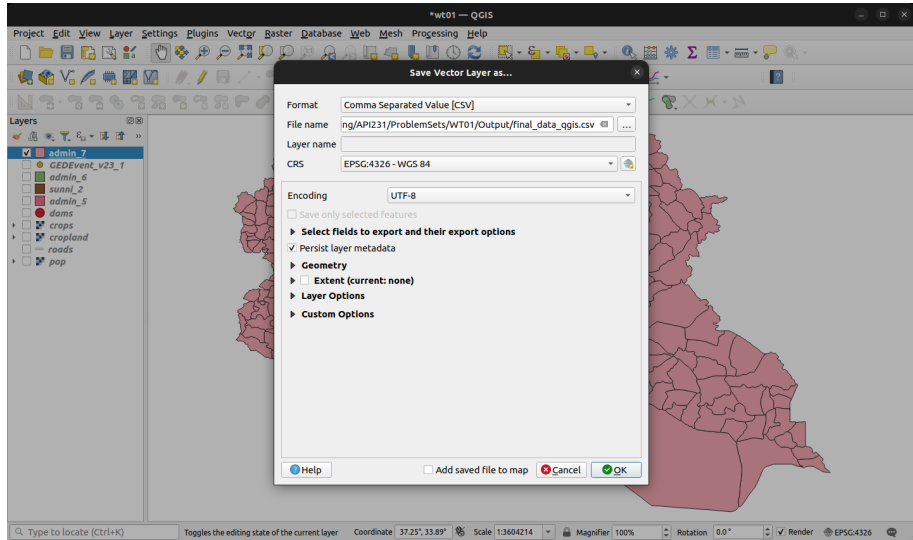
Right-click admin_7 in the Layer menu, go to Export→Save Features As.



Save the layer as an Comma Separated Values file



Name the file `final_data_qgis.csv`. Click OK



Analysis

Loading R packages

To conduct a regression analysis of these data in R, we will be using the stargazer packages

```
library(stargazer)
```

NOTE: The replication code for all of the preceding steps R is in wt08_demo.R on RStudio Cloud, and in WT01.zip (posted on Canvas).

Regression models

Now we're finally able to proceed to the *analysis stage*. For this we will need to **open the CSV file** we just created in R.

This code chunk imports the `final_data_qgis.csv` file into an object called `X`, and then lists the variable names:

```
X = read.csv("Output/final_data_qgis.csv")
names(X)
```

```
## [1] "GID_2"      "GID_0"      "COUNTRY"    "GID_1"      "NAME_1"
## [6] "NL_NAME_1"  "NAME_2"     "VARNAME_2"  "NL_NAME_2"  "TYPE_2"
## [11] "ENGTYPE_2"  "CC_2"       "HASC_2"     "roads_m"    "roads_ct"
## [16] "roads_km"   "area_km2"   "road_density" "pop_sum"    "crops_mean"
## [21] "dams"       "sunni_2_area" "prop_sunni"  "pop_density" "pop_1000"
## [26] "is_violence"
```

All the variables we created seem to be here.

Let's now *run some regression models!*

Quick refresher: our **regression analysis** will test 5 hypotheses at once

$$\text{violence}_i = \beta_1 \text{road density}_i + \beta_2 \text{population}_i + \beta_3 \text{cropland}_i \\ + \beta_4 \text{dams}_i + \beta_5 \text{Sunni presence}_i + \epsilon_i$$

where

- violence_i is the observed number of ISIS attacks in district i (is_violence)
- $\text{road density}_i, \dots, \text{Sunni presence}_i$ are explanatory variables (road_density, pop_1000, crops_mean, dams, prop_sunni)
- β are coefficient estimates corresponding to each Hypothesis

Hypothesis	Expectation	Observation
1. Power projection	$\beta_1 < 0$	?
2. Demographics	$\beta_2 > 0$	?
3. Political economy	$\beta_3 < 0$	?
4. Key infrastructure	$\beta_4 > 0$	?
5. Sectarian divisions	$\beta_5 > 0$	?

The first model (mod1) is an **Ordinary Least Squares** model that regresses a logarithmically-transformed dependent variable $\log(\text{is_violence} + 1)$ on all of the explanatory variables that correspond to our hypotheses. The log-transform is useful here because the `is_violence` variable is highly skewed

```
mod1 = lm(log(is_violence+1) ~ road_density + pop_1000 + crops_mean +  
          dams + prop_sunni, data=X)  
summary(mod1)
```

```
##  
## Call:  
## lm(formula = log(is_violence + 1) ~ road_density + pop_1000 +  
##     crops_mean + dams + prop_sunni, data = X)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -4.8209 -1.1849 -0.1103  0.9159  3.6005   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)  1.2767764  0.3438320   3.713 0.000284 ***  
## road_density  1.0261092  3.0441561   0.337 0.736513      
## pop_1000      0.0012820  0.0003496   3.667 0.000336 ***  
## crops_mean   -1.7270999  0.4814247  -3.587 0.000446 ***  
## dams          1.0476508  0.5331035   1.965 0.051168 .      
## prop_sunni    0.0348957  0.0030465  11.454 < 2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##
```

The second (mod2) is a **Generalized Linear Model with a quasi-Poisson link**. This parameterization is designed to accommodate dependent variables that are (over-dispersed) event counts

```
mod2 = glm(is_violence ~ road_density + pop_1000 + crops_mean +  
           dams + prop_sunni, data=X, family="quasipoisson")  
summary(mod2)
```

```
##  
## Call:  
## glm(formula = is_violence ~ road_density + pop_1000 + crops_mean +  
##      dams + prop_sunni, family = "quasipoisson", data = X)  
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept)  4.1398937  0.4813895   8.600 7.95e-15 ***  
## road_density -7.5459241  3.9548121  -1.908  0.0582 .  
## pop_1000      0.0010212  0.0002261   4.518 1.23e-05 ***  
## crops_mean   -2.7261978  0.6513259  -4.186 4.74e-05 ***  
## dams          0.7926585  0.3703707   2.140  0.0339 *  
## prop_sunni    0.0287344  0.0047586   6.038 1.10e-08 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for quasipoisson family taken to be 345.4133)  
##  
##      Null deviance: 69006  on 161  degrees of freedom  
## Residual deviance: 33381  on 156  degrees of freedom  
## AIC: NA  
##
```

The third and fourth models (mod3, mod4) add province-level **fixed effects** (NAME_1), which allow each province (e.g. Anbar) to have a different baseline level of violence

```
mod3 = lm(log(is_violence+1) ~ road_density + pop_1000 + crops_mean +  
          dams + prop_sunni + NAME_1, data=X)  
mod4 = glm(is_violence ~ road_density + pop_1000 + crops_mean +  
          dams + prop_sunni + NAME_1, data=X, family="quasipoisson")
```

mod3 is OLS, mod4 is quasi-Poisson

We can use the `stargazer()` command to export the models' coefficient estimates into a formatted table, which you could add to a paper or report:

```
stargazer::stargazer(mod1,mod2,mod3,mod4,type = "text",  
  keep = c("road_density","pop_1000","crops_mean",  
    "dams","prop_sunni"),  
  add.lines = list(c("Province FE","N","N","Y","Y")))
```

```
##
## =====
##                               Dependent variable:
## -----
##      log(is_violence + 1)      is_violence      log(is_violence + 1)      is_violence
##      OLS                      glm: quasipoisson      OLS                      glm: quasipoisson
##                               link = log              link = log
##      (1)                      (2)                      (3)                      (4)
## -----
## road_density                  1.026                  -7.546*                  2.898                  -11.922***
##                               (3.044)                  (3.955)                  (3.679)                  (4.533)
##
## pop_1000                     0.001***                 0.001***                 0.001***                 0.001***
##                               (0.0003)                 (0.0002)                 (0.0003)                 (0.0002)
##
## crops_mean                   -1.727***                 -2.726***                 -2.684***                 -2.537***
##                               (0.481)                  (0.651)                  (0.528)                  (0.606)
##
## dams                         1.048*                   0.793**                  0.805*                   0.880***
##                               (0.533)                  (0.370)                  (0.459)                  (0.298)
##
## prop_sunni                   0.035***                 0.029***                 0.016***                 0.012***
##                               (0.003)                  (0.005)                  (0.005)                  (0.004)
##
## -----
## Province FE                  N                      N                      Y                      Y
## Observations                 162                  162                  162                  162
## R2                          0.570
## Adjusted R2                 0.556
## Residual Std. Error         1.543 (df = 156)
## F Statistic                  41.351*** (df = 5; 156)
##                               1.204 (df = 125)
##                               13.079*** (df = 36; 125)
## =====
## Note:                        *p<0.1; ***p<0.05; ***p<0.01
```

What does this tell us about whether the data support our Hypotheses?

Hypothesis	Expectation	Confirm? (OLS)	Confirm? (QP)
1. Power projection	$\beta_{\text{road_density}} < 0$	✗	✓
2. Demographics	$\beta_{\text{pop_1000}} > 0$	✓	✓
3. Political economy	$\beta_{\text{crops_mean}} < 0$	✓	✓
4. Key infrastructure	$\beta_{\text{dams}} > 0$	✓	✓
5. Sectarian divisions	$\beta_{\text{prop_sunni}} > 0$	✓	✓