

SEST-6577

Geographic Information Systems for Security Studies

Lab 09 (+ Walk Through 1)

Yuri M. Zhukov
Associate Professor
Georgetown University

October 31, 2024

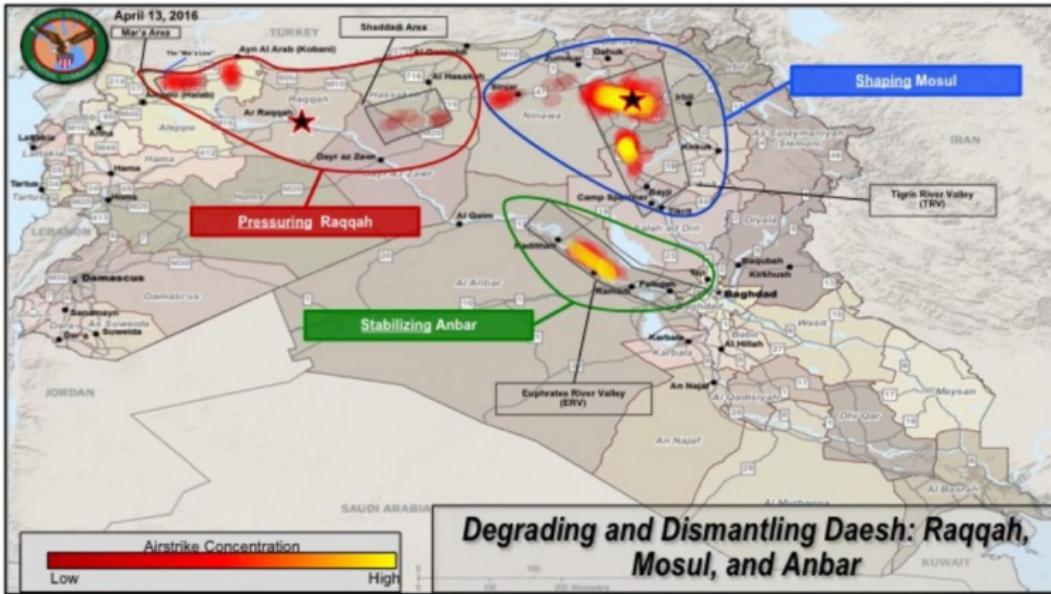
Goal: explain geographic variation in Islamic State violence

Figure 1: Operation Inherent Resolve

There is no formal Problem Set 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

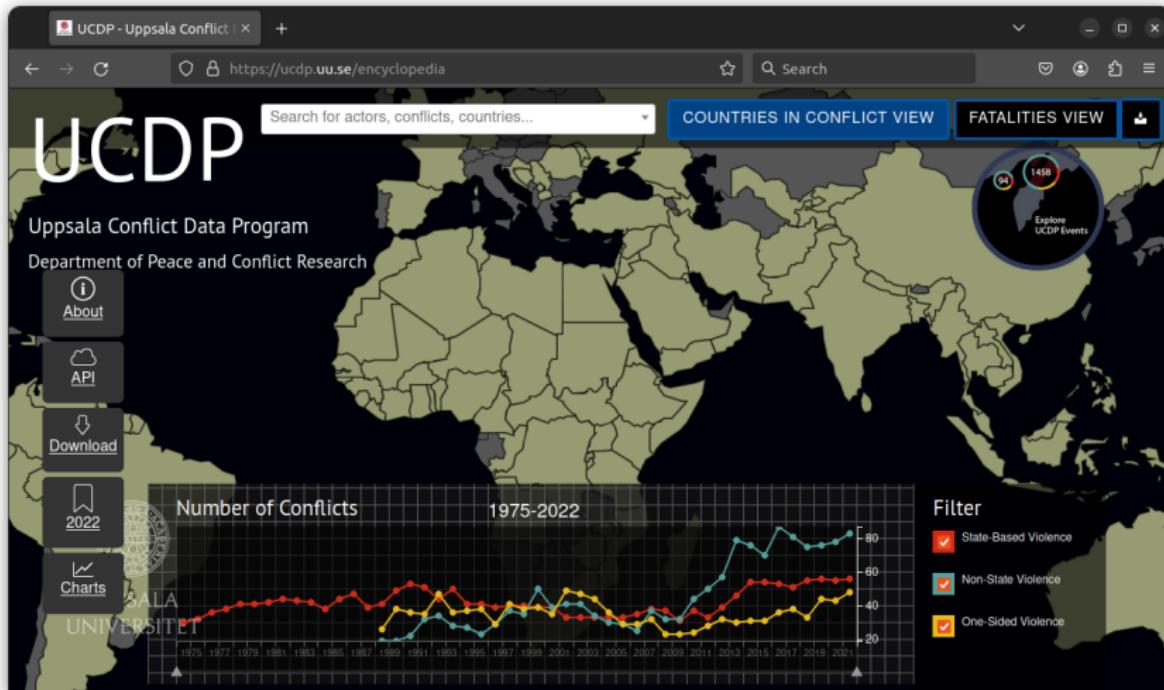
Collection

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 screenshot shows a web browser window with the following details:

- Search Bar:** The search term "georeferenced violent events data" is entered.
- Results:** The first result is a link to the "UCDP Dataset Download Center - Uppsala University".
 - Summary:** 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...
 - Related Links:** Charts, Graphs and Maps (The Uppsala Conflict Data Program offers a number of datasets on...); Peace Agreements (The official dataset download page of the Uppsala Conflict Data Program...).
- Second Result:** A link to an article titled "Introducing the UCDP Georeferenced Event Dataset".
 - Summary:** This article presents the UCDP Georeferenced Event Dataset (UCDP GED). The UCDP GED is an event dataset that disaggregates three types of organized violence (state-based conflict, non-state conflict, and one-sided violence) both spatially and temporally. Each event - defined as an instance of organiz...
- Third Result:** A link to the "UCDP - Uppsala Conflict Data Program".
 - Summary:** The Uppsala Conflict Data Program (UCDP) has recorded ongoing violent conflicts since the 1970s. The data provided is one of the most accurate and well-used data-sources on global armed conflicts and its...

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



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

The screenshot shows a web browser window with the title "UCDP Dataset Download". The URL in the address bar is <https://ucdp.uu.se/downloads/>. The page content includes a navigation bar with "COUNTRIES IN CONFLICT VIEW" and "FATALITIES VIEW". Below this, there are sections for "Disaggregated Datasets" (with links to "UCDP Georeferenced Event Dataset (GED) Global version 24.1" and "UCDP Candidate Events Dataset (UCDP Candidate) version 24.0.X"), "Yearly Datasets covering 1946 - 2023" (with links to "UCDP/PRI Armed Conflict Dataset version 24.1" and "UCDP Dyadic Dataset version 24.1"), and "Yearly Datasets covering 1989 - 2023" (with links to "UCDP One-sided Violence Dataset version 24.1", "UCDP Non-State Conflict Dataset version 24.1", and "UCDP Battle-Related Deaths Dataset version 24.1"). The "UCDP Georeferenced Event Dataset (GED) Global version 24.1" link is highlighted with a red border.

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 title "UCDP Dataset Download". The URL in the address bar is https://ucdp.uu.se/downloads/index.html#ged_global. The main content is titled "UCDP Georeferenced Event Dataset (GED) Global version 24.1". A descriptive text states: "This dataset is UCDP's most disaggregated dataset, covering individual events of organized violence (phenomena of lethal violence occurring at a given time and place). These events are sufficiently fine-grained to be geo-coded down to the level of individual villages, with temporal durations disaggregated to single, individual days." Below this, a section titled "Available as:" lists five formats: CSV (selected), EXCEL, R, STATA, and CODEBOOK. There is also a "Please cite:" section with two references:

- Davies, Shawn, Garoun Engström, Therese Pettersson & Magnus Öberg (2024). Organized violence 1989–2023, and the prevalence of organized crime groups. *Journal of Peace Research* 61(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 between the two products.

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

GEDEvent_v23_1.csv - LibreOffice Calc

File Edit View Insert Format Styles Sheet Data Tools Window Help

Liberation Sans 10 pt

AB latitude AC

	AB	AC	AD	AE	AF
1	adm_1	adm_2	latitude	longitude	geom_wkt
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		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	Hirat 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	Jabalussaraj 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	69.162706	POINT (69.162706 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`

The screenshot shows a web browser window with the title "Free GIS Datasets - Category". The URL is <https://freegisdata.rtwilson.com>. The page header includes "Free GIS Data", "Home", "Physical", "Human", "Country-specific", "FAQ", and social media sharing buttons for Facebook, Twitter, and a "Donate" button.

The main content area lists several data sources:

- GeoNetwork: GIS aggregation site including a wide range of data under various categories (both human and physical).
- Google Maps Gallery: A wide range of user-submitted geographic data, available as a KML file
- History Database of the Global Environment: Gridded time-series of population, land-use for the last 12,000 years. Also includes GDP, agricultural areas, yields and greenhouse gas emissions for the last century.

Administrative Boundaries

- Natural Earth: Includes countries, disputed areas, first-order admin (departments, states etc), populated places, urban polygons, parks and protected areas and water boundaries. Available at multiple levels of detail.
- Geoboundaries: Geographic boundaries dataset, for academic and commercial use. Over 1 million boundaries within 200 countries.
- GADM**: Global administrative boundaries, with extensive attribute sets. Covers countries and up to four levels of internal administrative boundary (states, departments, counties etc). (For Land Cover see above) (This item is highlighted with a red border)
- World Borders: World country borders with attributes including country codes (FIPS, ISO etc), area and populations.
- Europe in the World: Administrative boundaries for Europe with lots of attribute data for each country/region including information on economy, demography and infrastructure.
- CShapes - Historical Boundaries: Historical state boundaries and capitals post-WW2, world-wide, including all changes and dates of changes.
- VLIZ Maritime Boundaries: Maritime boundaries and areas of Exclusive Economic Zones, including detailed attribute data on treaties etc. From the Flanders Marine Institute.
- TZ timezones: A map of timezone areas in the world as used in the Unix TZ database format, from which we get the naming Europe/London, America/New_York etc. In shapefile format.

Environmental Boundaries

- World Spatial Database of Protected Areas: Global vector database of marine and terrestrial protected areas. Rather complicated to download from - instructions at bottom of linked page.
- IUCN 2013 Red List: Set of shapefiles produced by the IUCN showing the distribution of endangered species of plants and animals across the world
- Protected Planet: Map of protected areas across the whole world, of almost all types. Available for download by clicking the 'Download All' link on the homepage, and then scrolling to the bottom and choosing KMZ, SHP or CSV.

[Land Use](#)

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 following details:

- Title Bar:** GADM
- Address Bar:** https://gadm.org/download_country.html
- Header:** GADM (highlighted in orange), Maps, Data, About
- Main Content:** Download GADM data (version 4.1)
- Country Selection:** A dropdown menu set to "Iraq".
- File Formats:** Geopackage, Shapefile, [GeoJSON: level-0, level1, level2](#), KMZ: level-0, level1, level2
- Map Preview:** A map of Iraq showing its administrative divisions at level 2, with regions colored yellow and boundaries in blue.
- Link at the bottom:** https://geodata.ucdavis.edu/gadm/gadm4.1/json/gadm41 IRQ_2.json.zip

... and for Syria

The screenshot shows a web browser window for the GADM website (https://gadm.org/download_country.html). The page title is "GADM". The main content is titled "Download GADM data (version 4.1)". A dropdown menu labeled "Country" has "Syria" selected. Below the dropdown are download links for "Geopackage", "Shapefile", "GeoJSON: level-0, level1, level2", and "KMZ: level-0, level1, level2". To the right is a map of Syria with its administrative divisions outlined in blue and filled with yellow. The map includes a coordinate grid with latitude from 30 to 38 and longitude from 36 to 42.

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

The screenshot shows a web browser window with the following details:

- Address Bar:** https://duckduckgo.com/?t=ffab&q=roads+GIS+data+by+country&..
- Search Query:** roads GIS data by country
- Search Engine:** DuckDuckGo
- Results:**
 - Result 1:** [Download data by country | DIVA-GIS](https://www.diva-gis.org/gdata)

Description: From DIVA-GIS they can be exported to a number of other grid formats including IDRISI and Arc or to shapefiles. (More info on format; for developers). A gridfile with "country mask" indicates that the areas outside the selected **country** are masked out. For these areas, the value is "NODATA" (indicating the...)

Links:
 - Download:** Download DIVA-GIS 7.5. Please contact us if you find a bug. For full functionall...
 - About:** DIVA-GIS is developed by Robert Hijmans. Previous versions were co...
 - Frequently Asked Questions:** B: Your data may be in a different projection than the climate data. After...
 - Links:** The competition: Free GIS programs qGIS, GRASS. Other Links. Software fo...
 - Result 2:** [GRIP global roads database | GLOBIO - Global biodiversity model for ...](https://www.globio.info/download-grid-dataset)

Description: The Global Roads Inventory Project (GRIP) dataset was developed to provide a more recent and consistent global roads dataset for use in global environmental and biodiversity assessment models like GLOBIO. The GRIP dataset consists of global and regional vector datasets in ESRI filegeodatabase...
- Bottom Right:** Share Feedback button

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

The screenshot shows a web browser window titled "Data — DIVA-GIS". The URL is <https://diva-gis.org/data.html>. The page header includes "DIVA-GIS DATA" and navigation links for "Documentation", "Contents", and "Climate data for DIVA-GIS". The main content area is titled "Data" and contains instructions: "Select and download free geographic (GIS) data for any country in the world
". Below this are three dropdown menus: "Country" set to "Iraq", "Subject" set to "Roads", and "Format". A link "Download [IRO_rds.zip](#)" is provided. At the bottom, there is a section titled "Sources" with a table listing data sources for Iraq.

Subject	Description	Source	Format	Resolution
Administrative areas (boundaries)	Country outlines and administrative subdivisions for all countries. The level of subdivisions that is available varies between countries	GADM	Vector (area)	-
Roads	Digital chart of roads, rivers, and lakes. Separate files for line and area	Digital Chart of the World	Vector (line and area)	-

Download the roads data for both Iraq and Syria here

The screenshot shows a web browser window titled "Data — DIVA-GIS". The URL in the address bar is <https://diva-gis.org/data.html>. The page header includes "DIVA-GIS" and "DATA". A navigation bar at the top right contains links to "Documentation", "Contents", and "Climate data for DIVA-GIS". Below the header, a section titled "Data" is displayed. It contains a message: "Select and download free geographic (GIS) data for any country in the world
". There are three dropdown menus: "Country" set to "Syria", "Subject" set to "Roads", and "Format". Below these, a link "Download [SYR_rds.zip](#)" is shown. A "Sources" section follows, featuring a table with two rows. The first row has columns for "Subject" (Administrative areas (boundaries)), "Description" (Country outlines and administrative subdivisions for all countries. The level of subdivisions that is available varies between countries), "Source" (GADM), "Format" (Vector (area)), and "Resolution" (-). The second row has columns for "Subject" (Roads), "Description" (Digital vector lines, and lakes. Separate files for line features), "Source" (Digital Chart of the World), "Format" (Vector (line and area)), and "Resolution" (-). The URL for the second row is https://geodata.ucdavis.edu/diva/rds/SYR_rds.zip.

Subject	Description	Source	Format	Resolution
Administrative areas (boundaries)	Country outlines and administrative subdivisions for all countries. The level of subdivisions that is available varies between countries	GADM	Vector (area)	-
Roads	Digital vector lines, and lakes. Separate files for line features	Digital Chart of the World	Vector (line and area)	-

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

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 it includes such things as shop and business locations (eg. all Tesco stores, all McDonald's restaurants) as well as places of worship, speed

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 Socioeconomic Data and Applications Center (SEDAC) website. The URL is https://sedac.ciesin.columbia.edu/datacollection/gpw-v4. The page title is "Gridded Population of the World (GPW), v4". On the left, there's a sidebar with links like Collection Overview, Methods, Data Sets (9), Map Gallery (97), Map Services (104), Citations, and FAQs. The main content area has a heading "Introduction" followed by a detailed description of the GPW collection. To the right, there's a map titled "GPWv4 Data Quality Indicators, v4.11_Mean Administrative Unit Area" showing global population density in purple. At the bottom, there's a search bar and some navigation buttons.

Gridded Population of the World (GPW), v4

Collection Overview Introduction

Methods Data Sets (9) Map Gallery (97) Map Services (104) Citations FAQs

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 remote sensing. 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". A prominent modal dialog box in the center of the page displays a "Notice" message: "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." Below the message are two buttons: "LOGIN" (highlighted with a red border) and "CANCEL". The background of the page shows the "Socioeconomic Data and Applications Center (SEDAC)" logo and navigation links for DATA, MAPS, MEDIA, ABOUT, and HELP.

Gridded Population of the World (GPW), v4

Collection Overview

Population Count, v4.11 (2000, 2005, 2010, 2015, 2020)

Set Overview Data Download Maps Map Services Documentation Metadata

Methods

Data Sets (9)

Population Count, v4.11 (2000, 2005, 2010, 2015, 2020)

Show All...

Map Gallery (97)

Map Services (104)

Citations

feedback and support

<https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count-rev11/data-download#>

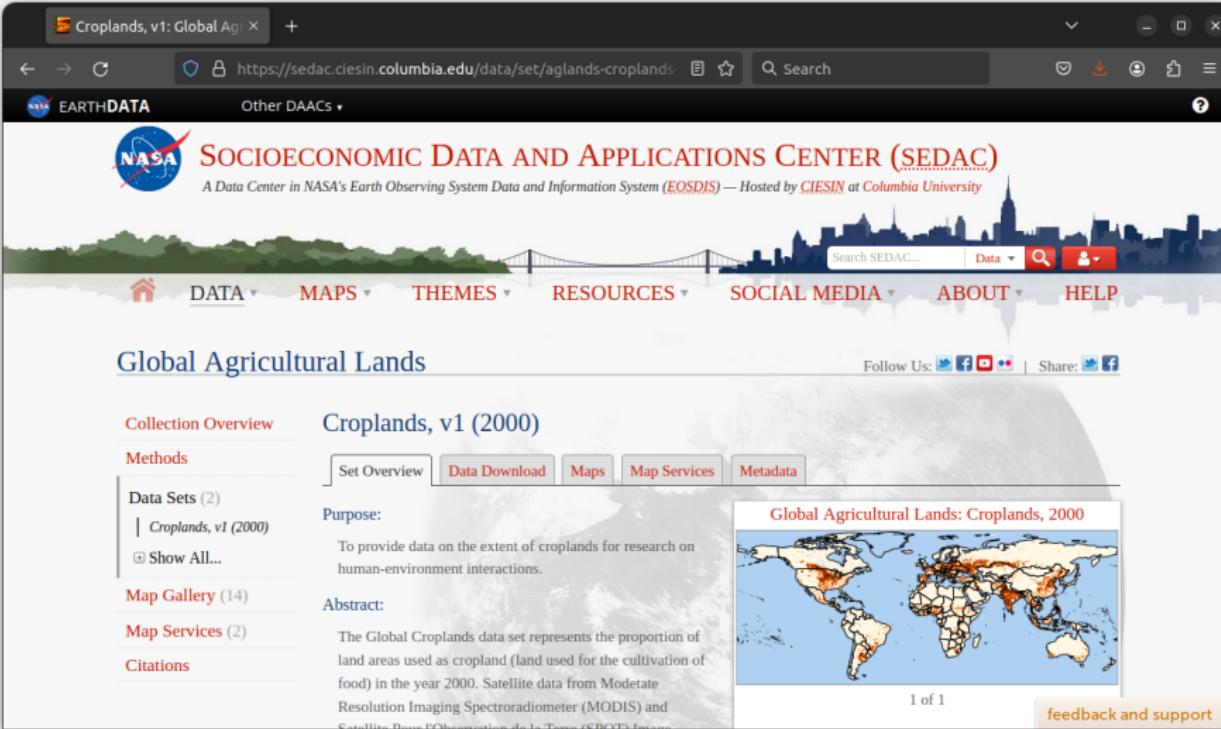
The data are available for multiple years, in multiple formats and resolutions. We will be using the 15-minute GeoTiff raster for 2010 (latest before ISIS emerged in 2014)

A screenshot of a web browser window titled "Downloads » Population". The URL is <https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population>. The page content is as follows:

- Map Gallery (97)**: 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).
- Map Services (104)**: 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.
- Citations**: The netCDF-4 file format is only available for the All Years Combined category and is not available at 30 arc-second resolution.
- FAQs**: Please select all required fields and press the Create Download button.
- Documentation**: Temporal: Single Year, FileFormat: GeoTiff, Resolution: 15 Minute ... **Create Download**
- What's New in GPWv4**: Please select all required fields and press the Create Download button.
- What's New in Revision 10**: Please select all required fields and press the Create Download button.
- What's New in Revision 11**: Please select all required fields and press the Create Download button.
- What is UN WPP-adjusted population data?**: Files Selected: 1 Download Size: 1.063MB
 - Select All
 - Year 2000
 - Year 2005
 - Year 2010
 - Year 2015
 - Year 2020
- Multimedia**
- SEDAC Hazards Mapper**
- Population Estimation Service**
- Acknowledgments**
- COVID-19 Visualizer**

feedback and support

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



The screenshot shows the Socioeconomic Data and Applications Center (SEDAC) website. The header includes the NASA Earthdata logo, a search bar, and links for "Other DAACs". Below the header is a banner featuring the SEDAC logo and a city skyline silhouette. The main navigation menu includes links for DATA, MAPS, THEMES, RESOURCES, SOCIAL MEDIA, ABOUT, and HELP. The left sidebar contains links for Collection Overview, Methods, Data Sets (2), Map Gallery (14), Map Services (2), and Citations. The main content area is titled "Global Agricultural Lands" and features a sub-section for "Croplands, v1 (2000)". This section includes tabs for Set Overview, Data Download (which is selected), Maps, Map Services, and Metadata. Below these tabs, the "Purpose" is described as providing data on the extent of croplands for research on human-environment interactions. The "Abstract" explains that the data set represents the proportion of land areas used as cropland in 2000, using satellite data from MODIS. To the right, there is a world map titled "Global Agricultural Lands: Croplands, 2000" showing agricultural land distribution. A footer at the bottom right includes a "feedback and support" link.

Download the global raster, in GeoTiff format

Global Agricultural Lands

Croplands, v1 (2000)

Collection Overview

Methods

Data Sets (2)

- | Croplands, v1 (2000)
- ⊕ Show All...

Map Gallery (14)

Map Services (2)

Citations

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 displaying the Socioeconomic Data and Applications Center (SEDAC) website. The URL is <https://sedac.ciesin.columbia.edu/data/set/grand-v1-dams-rev>. The page title is "Dams, v1.01: Global Reservoir and Dam Database". The main content area is titled "Global Reservoir and Dam (GRanD), v1". On the left, there is a sidebar with links to "Collection Overview", "Methods", "Data Sets (2)", "Map Gallery (20)", "Map Services (2)", "Citations", and "Project Website". The main content area includes tabs for "Set Overview", "Data Download", "Maps", "Map Services", and "Metadata". Below these tabs, there is a "Purpose" section stating: "To provide a geographically explicit and reliable database of dams for the scientific community." There is also an "Abstract" section describing the dataset. To the right, there is a world map titled "Global Reservoir and Dam Database: Dams" showing the locations of dams across the globe. At the bottom right of the map, it says "1 of 1" and "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

The screenshot shows a search results page from DuckDuckGo. The search query is "ethnic settlements geojson". The results include:

- ETH - International Conflict Research - GeoEPR 2021**
The GeoEPR-ETH Version 2.0 dataset is based on EPR-ETH Version 2.0 and codes the settlement patterns of politically relevant ethnic groups in independent states. Geo-referencing Ethnic Power Relations (GeoEPR) v1.0.
- Source for GeoJSON data for the US - Geographic Information System...**
the API shows US Boundaries (GeoJson) by zipcode, city, and state. you should use the API programmatically to handle large results. the image is query by state (DC), gives you all zipcodes in DC. This uses TIGER2013, and can query for multiple counties, cities, states, zipcodes for aggregate GeoJso...
- Geocoding resources for Bulgaria - GitHub**
settlements.geojson. This is a simplified version of the territorial separation of Bulgaria among the grounds of settlements. While some settlements don't have assigned grounds, they reside inside the grounds of other settlements. In such cases they would be listed in the "contains" array property. The...

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

The screenshot shows a web browser window with the URL <https://icr.ethz.ch/data/epr/geoepr/>. The page is titled "International Conflict Research" and features a navigation menu with "Data" selected. Below the menu, a breadcrumb trail shows the path: ETH Zurich > D-GESS > CIS > ICR > Data > Ethnic Power Relations (EPR) Dataset Family > GeoEPR 2021. The main content area is titled "GeoEPR - Geo-referencing Ethnic Power Relations". It describes the dataset as geo-coding politically relevant ethnic groups from the EPR-Core 2021 dataset and provides polygons for their locations. A "Data" section lists various formats available for download, including Shapefile, CSV, Tab-delimited, SQL, and JSON (GeoJSON). The "JSON (GeoJSON) format: [GeoEPR-2021.geojson](#)" link is highlighted with a red box.

Here is the full list of data sources and links:

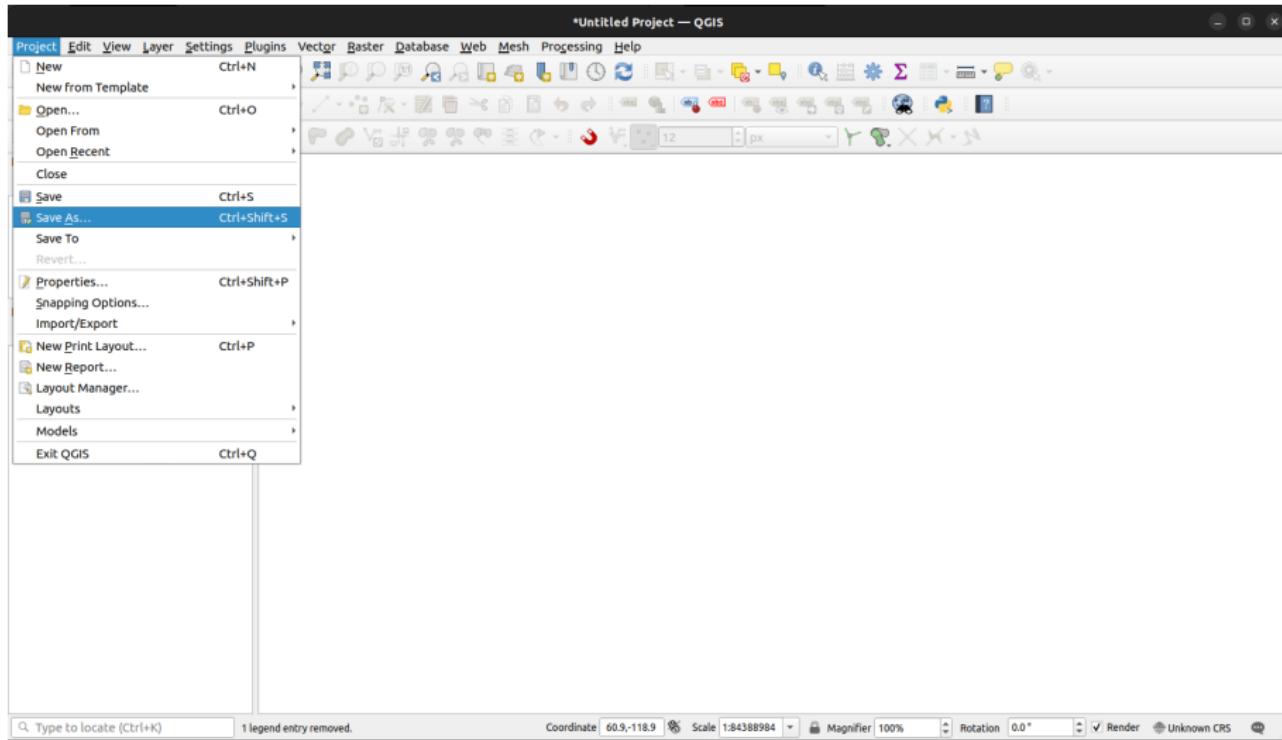
Category	Type	Format	Data source
ISIS violence	Table (non-geo)	.csv	UCDP GED
Administrative units	Vector (polygons)	.geojson	GADM
Roads	Vector (polylines)	.shp	DIVA-GIS
Population	Raster	.tif	CIESIN
Cropland	Raster	.tif	CIESIN
Dams	Vector (points)	.shp	CIESIN
Sectarian divisions	Vector (polygons)	.geojson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

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

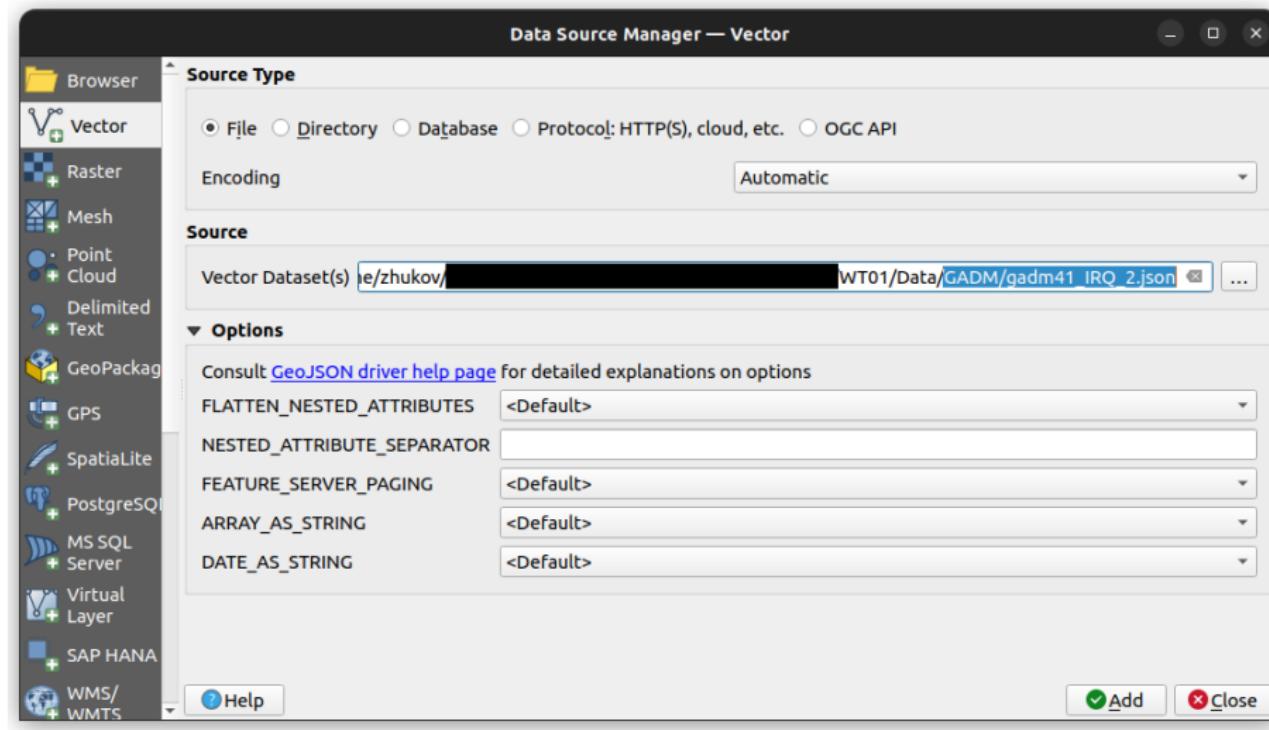
Pre-processing

Always save your progress!

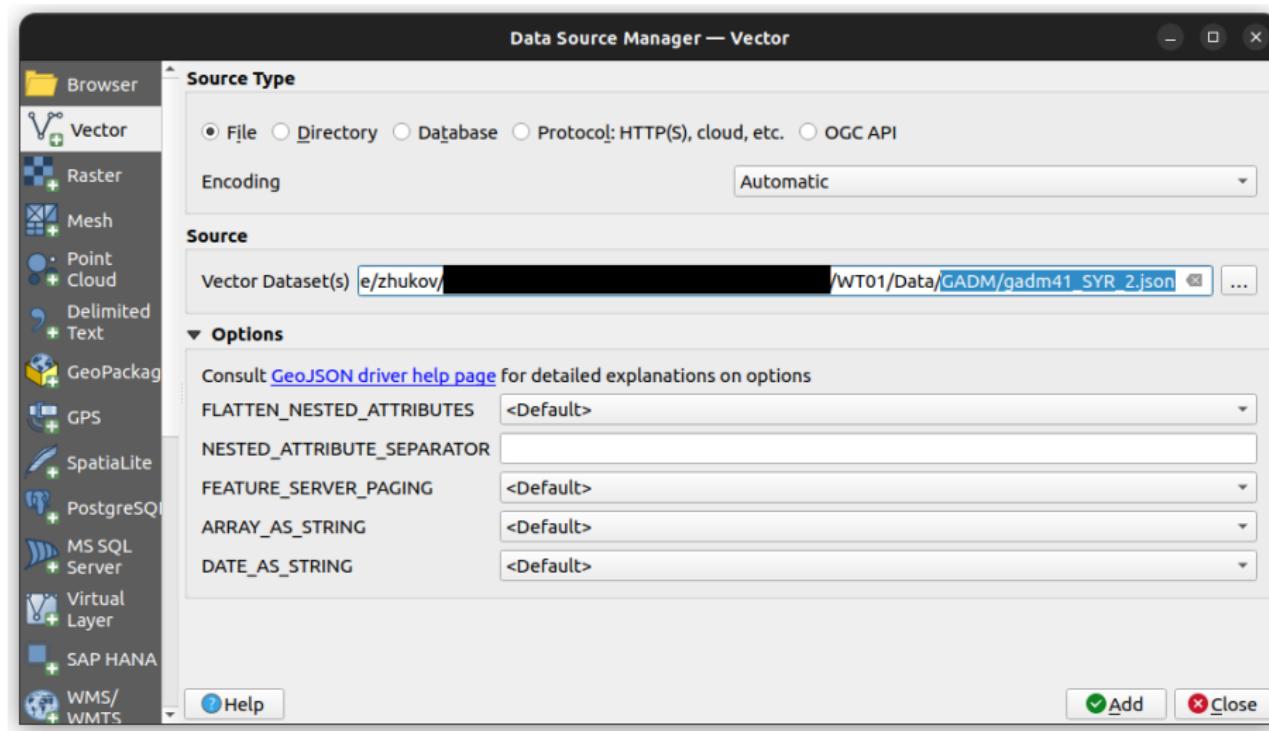
Go to Project → Save As...



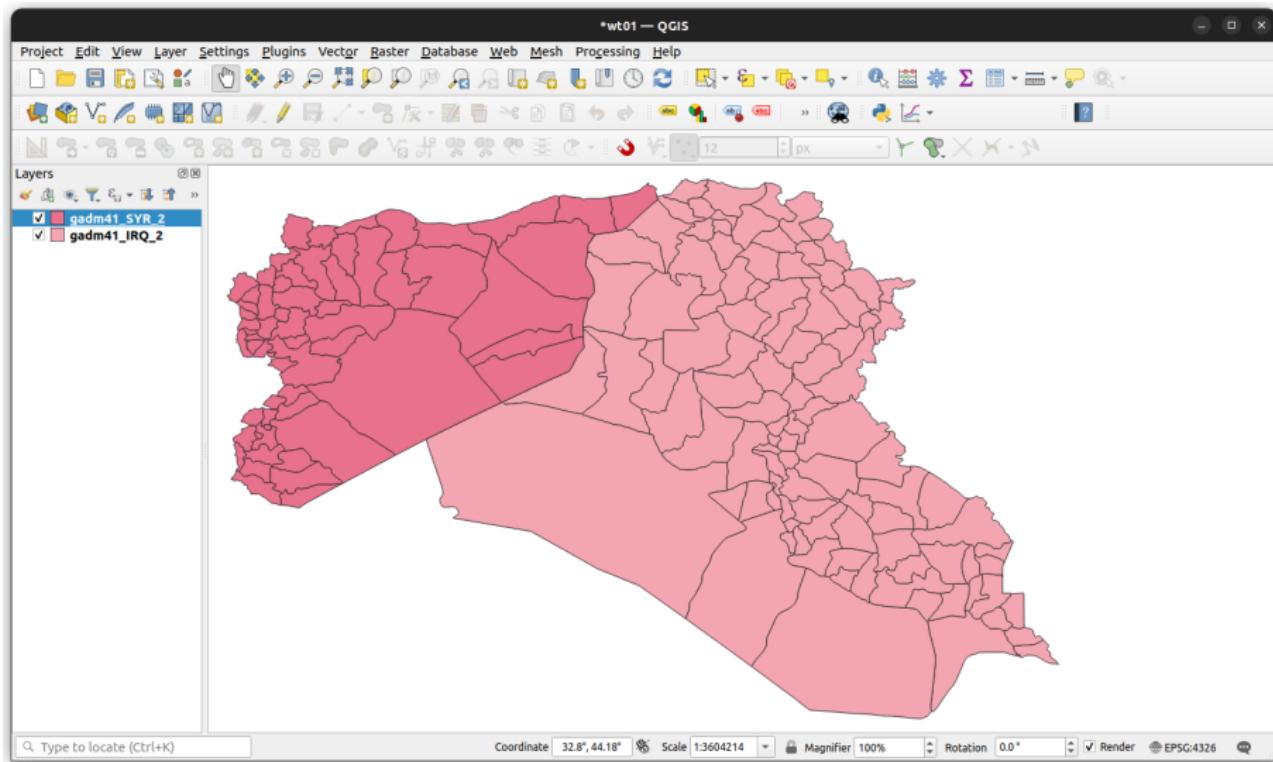
Load *administrative boundaries* (Layer → Add Layer → Add Vector Layer).
For **Iraq**: gadm41 IRQ_2.json file in Data/GADM folder.



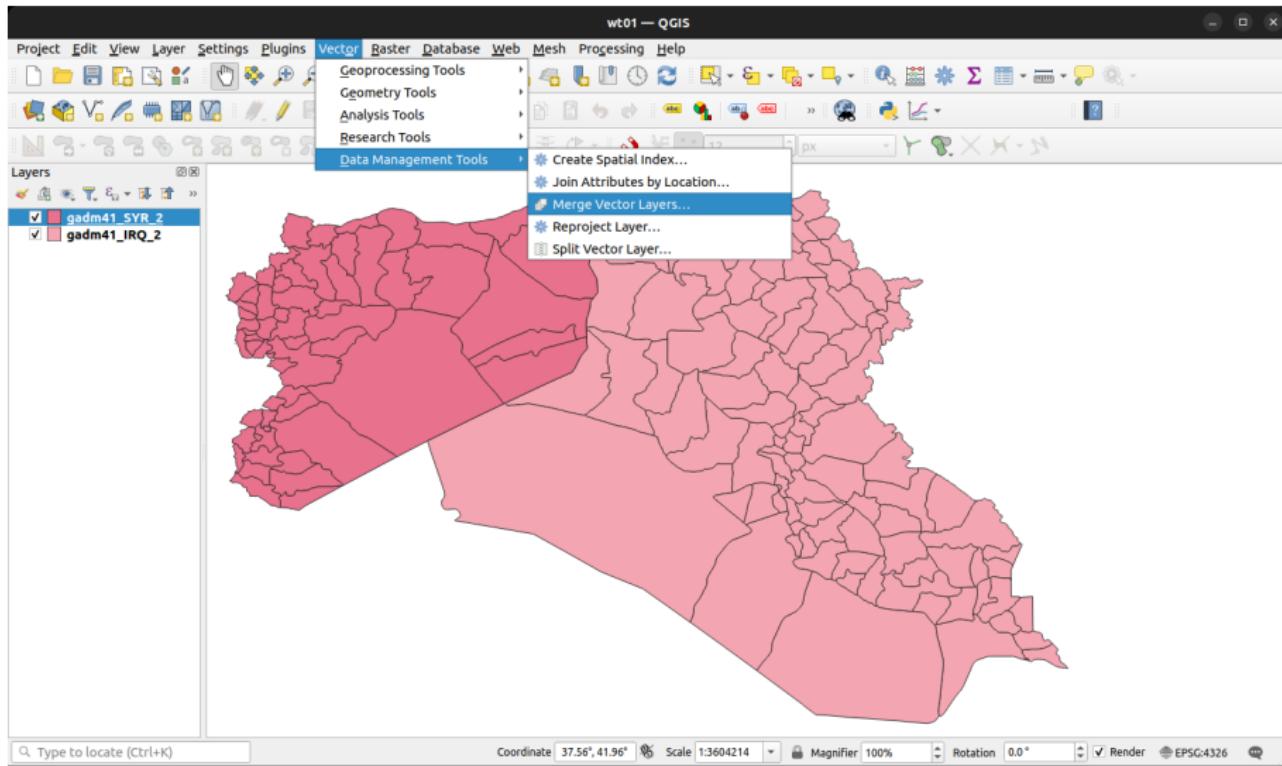
For Syria: gadm41_SYR_2.json file in Data/GADM folder.



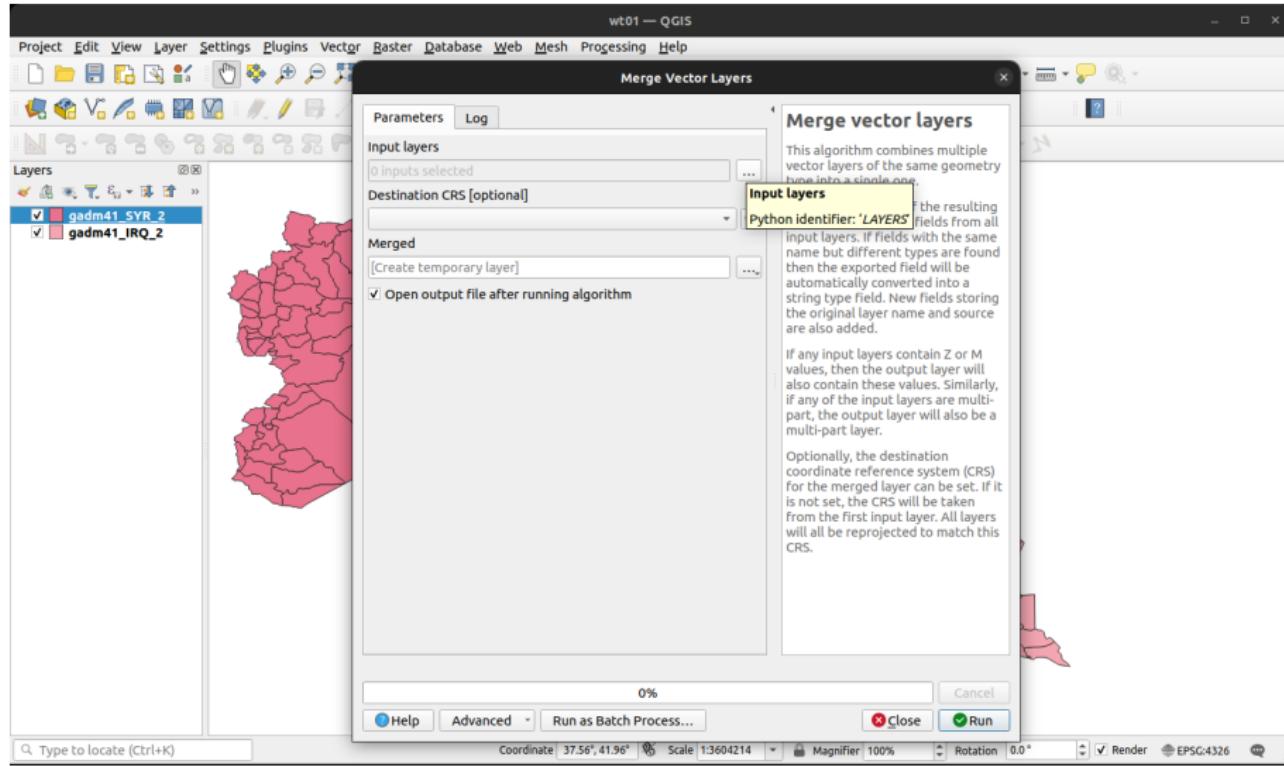
The two sets of district boundaries should be visible in the project window.
Let's **merge them into a single layer**, with districts for both countries.



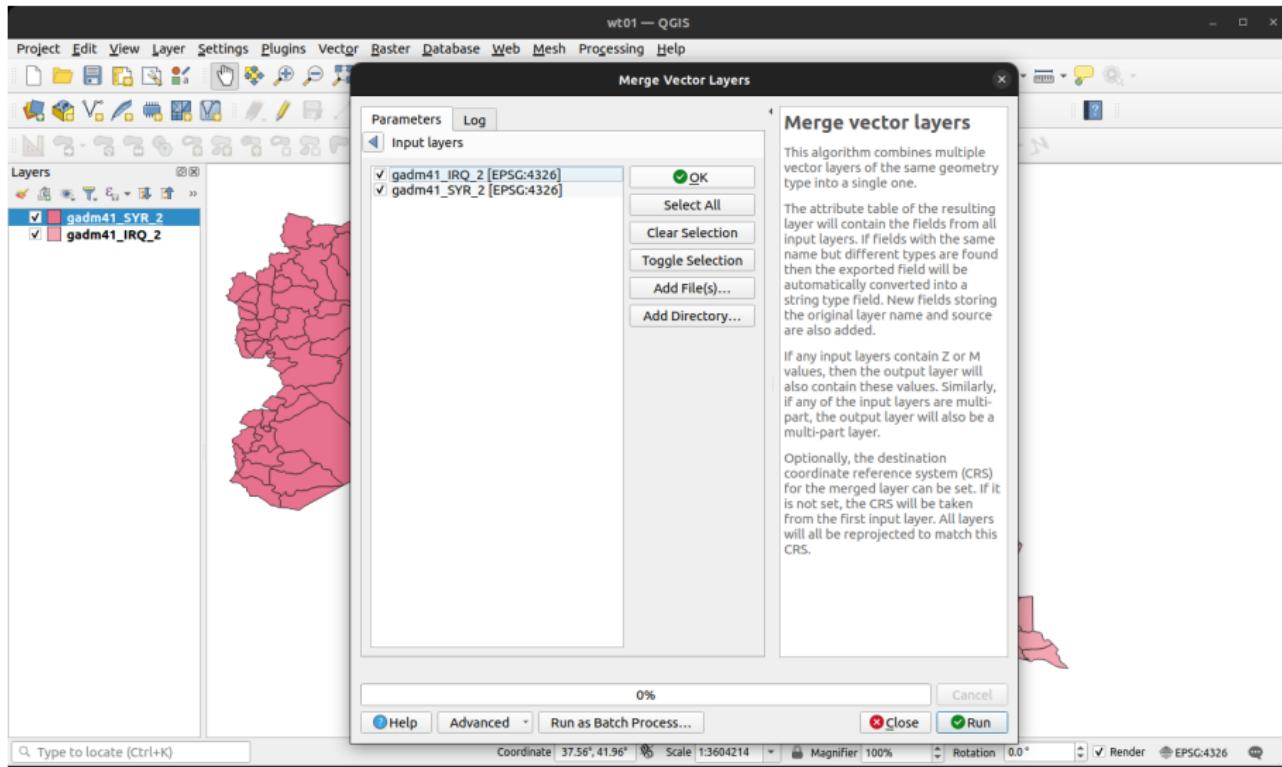
Open the Merge Vector Layers tool (Vector → Data Management Tool → Merge Vector Layers).



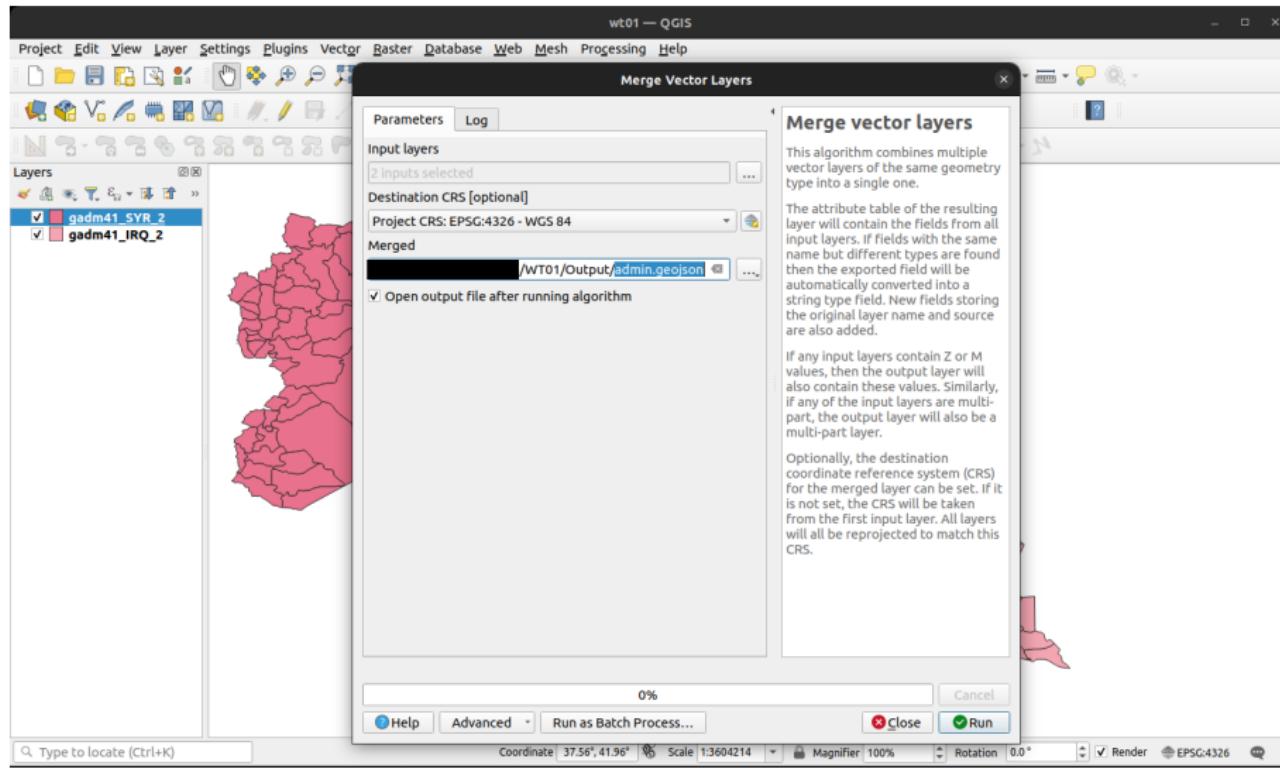
Click on the [...] box next to Input layers



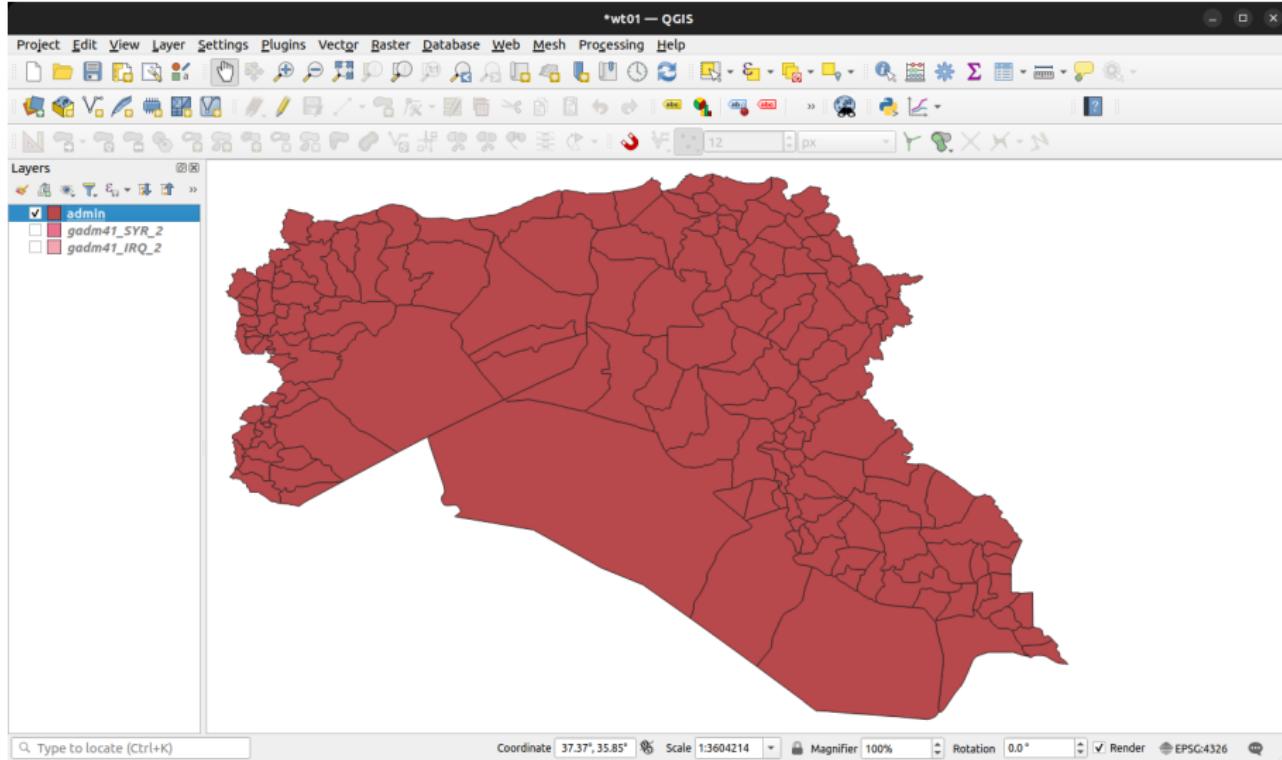
Check the boxes next to `gadm41 IRQ_2` and `gadm41 SYR_2`. Click OK.



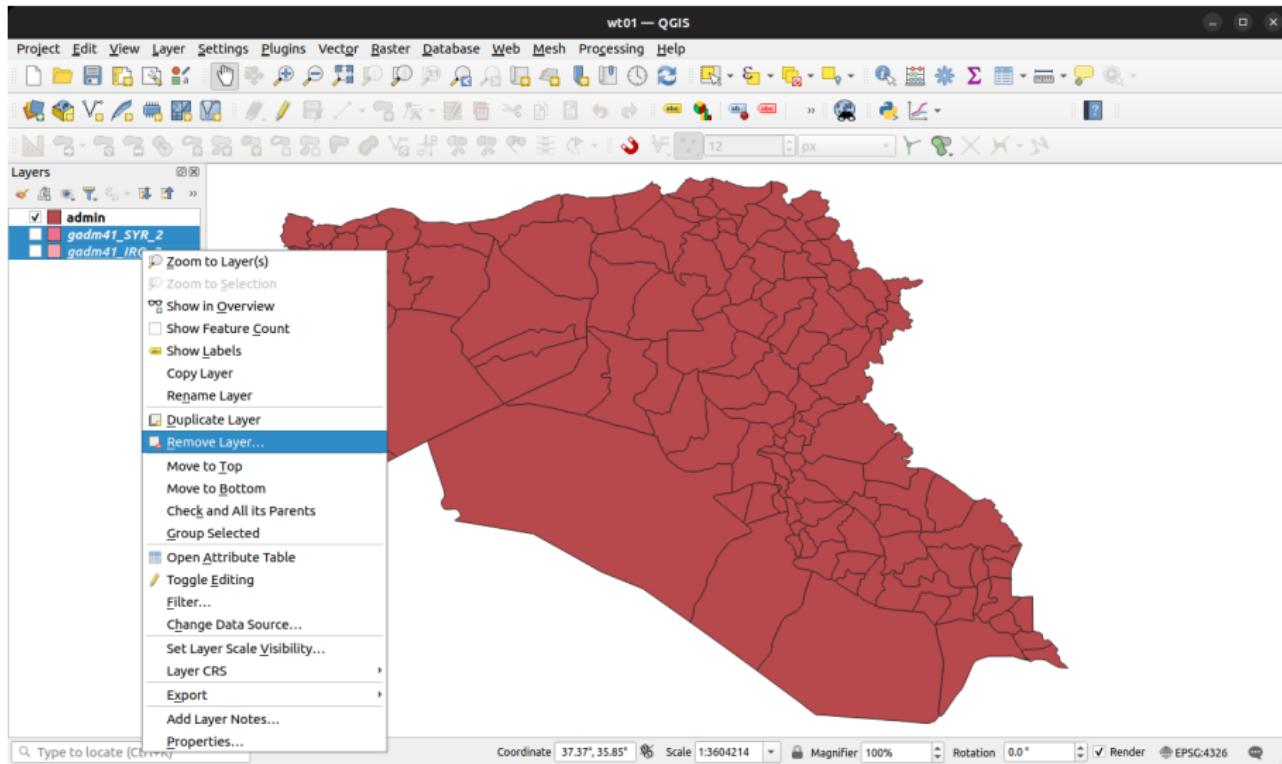
Click on the [...] box next to Merged. Find a location in which to save the output, and name the file admin.geojson. Click Run



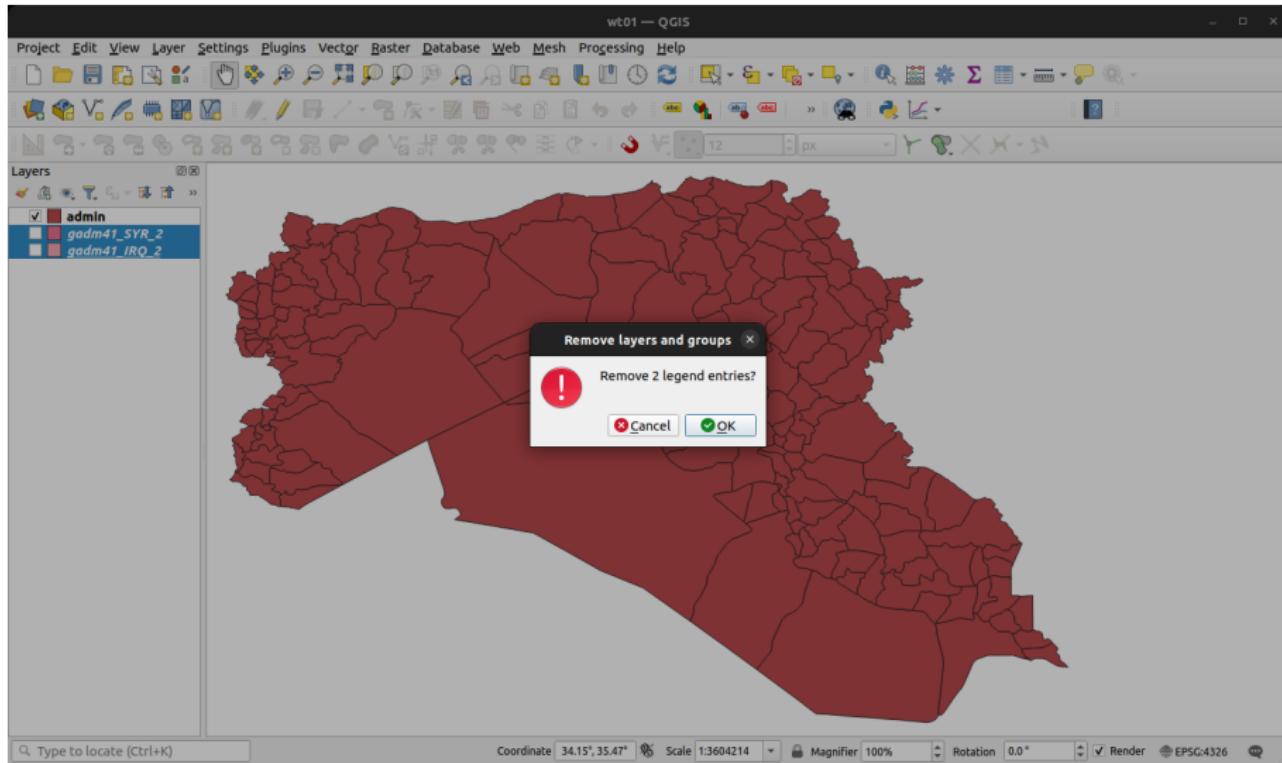
The merged layer should appear in the main project window.



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...)

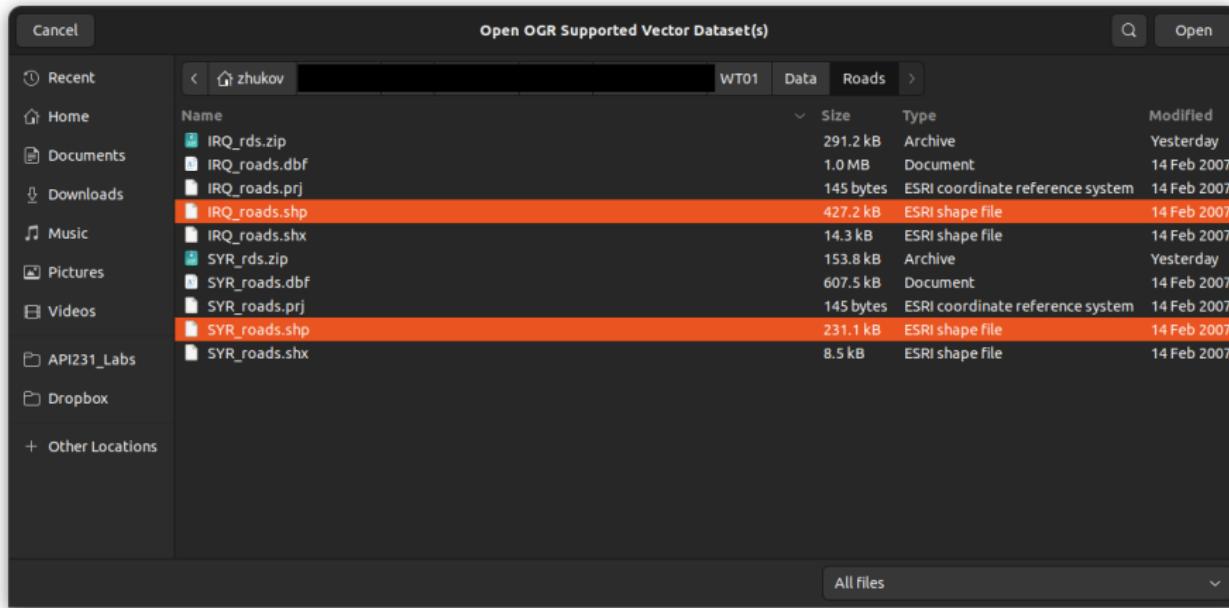


To conserve memory, it's good practice to remove data objects were are not actively using

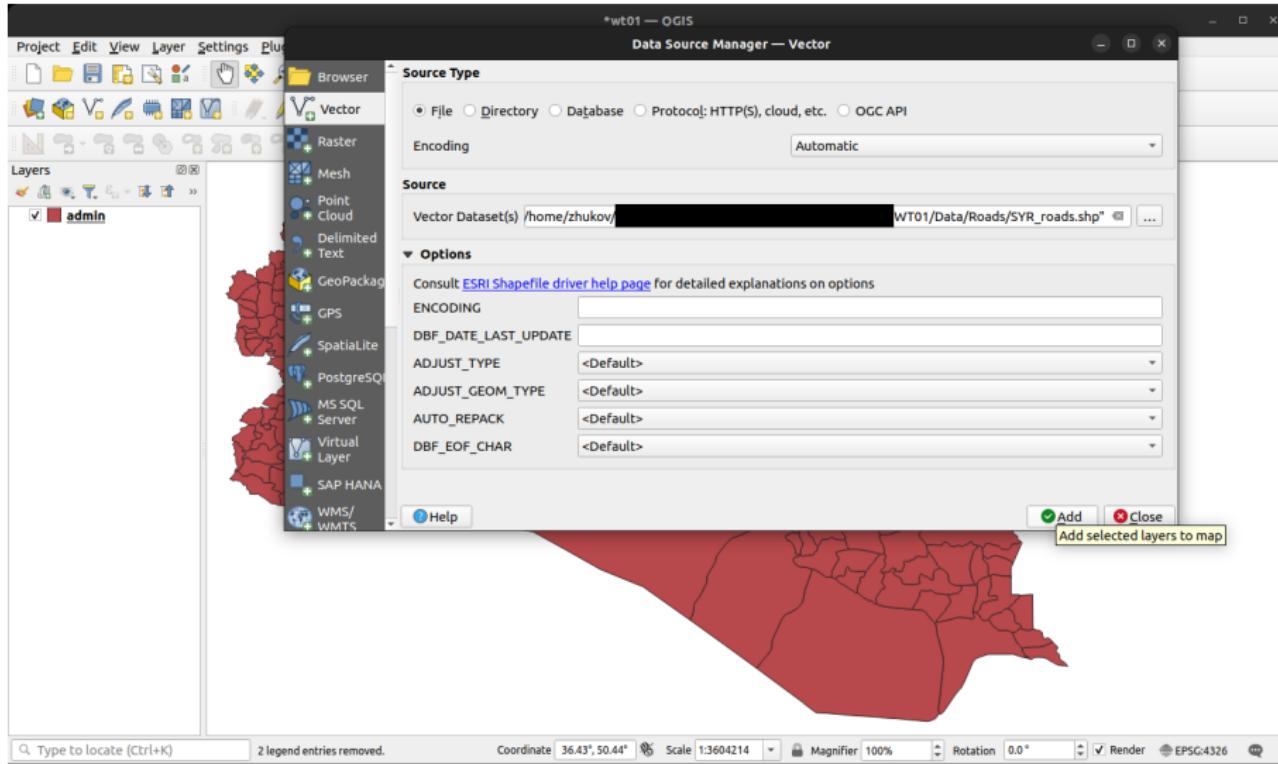


Load *roads data* (Layer → Add Layer → Add Vector Layer).

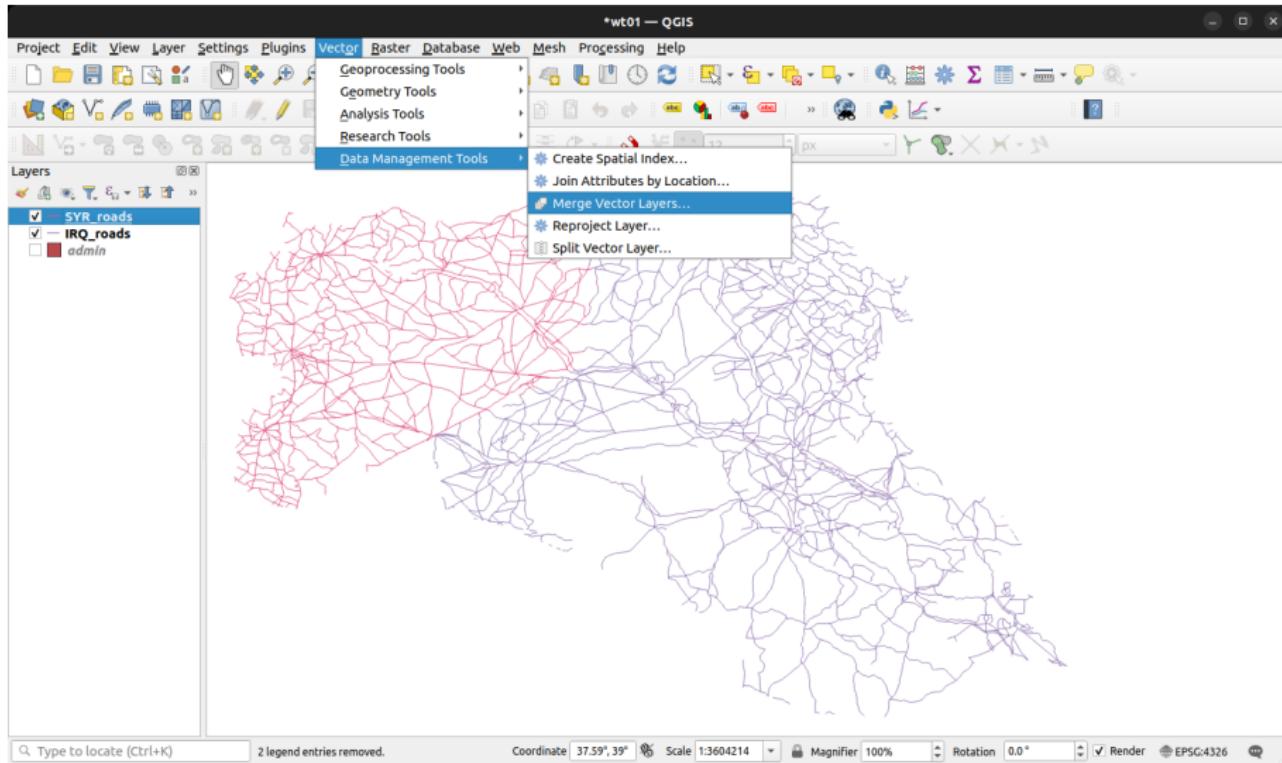
Try opening **two datasets at the same time**: IRQ_roads.shp and SYR_roads.shp files in Data/Roads



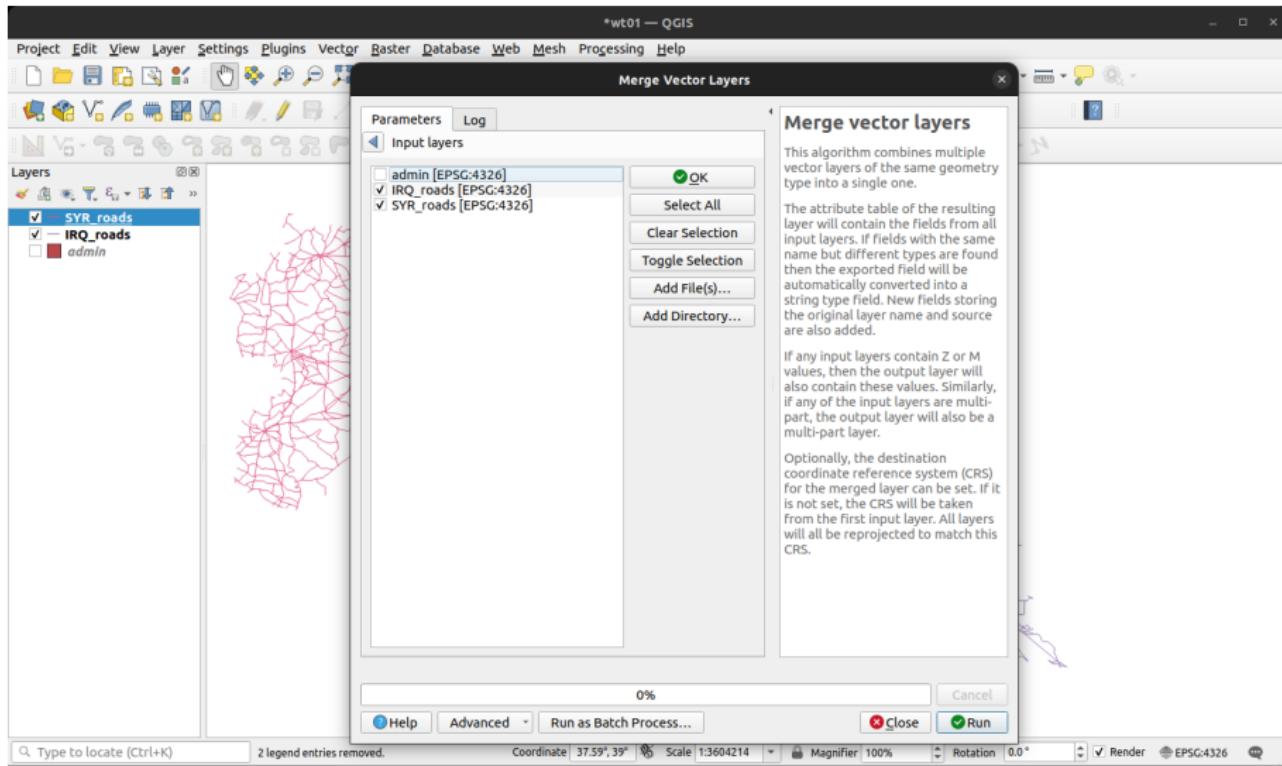
Paths to both files should appear in the Source box. Click Add



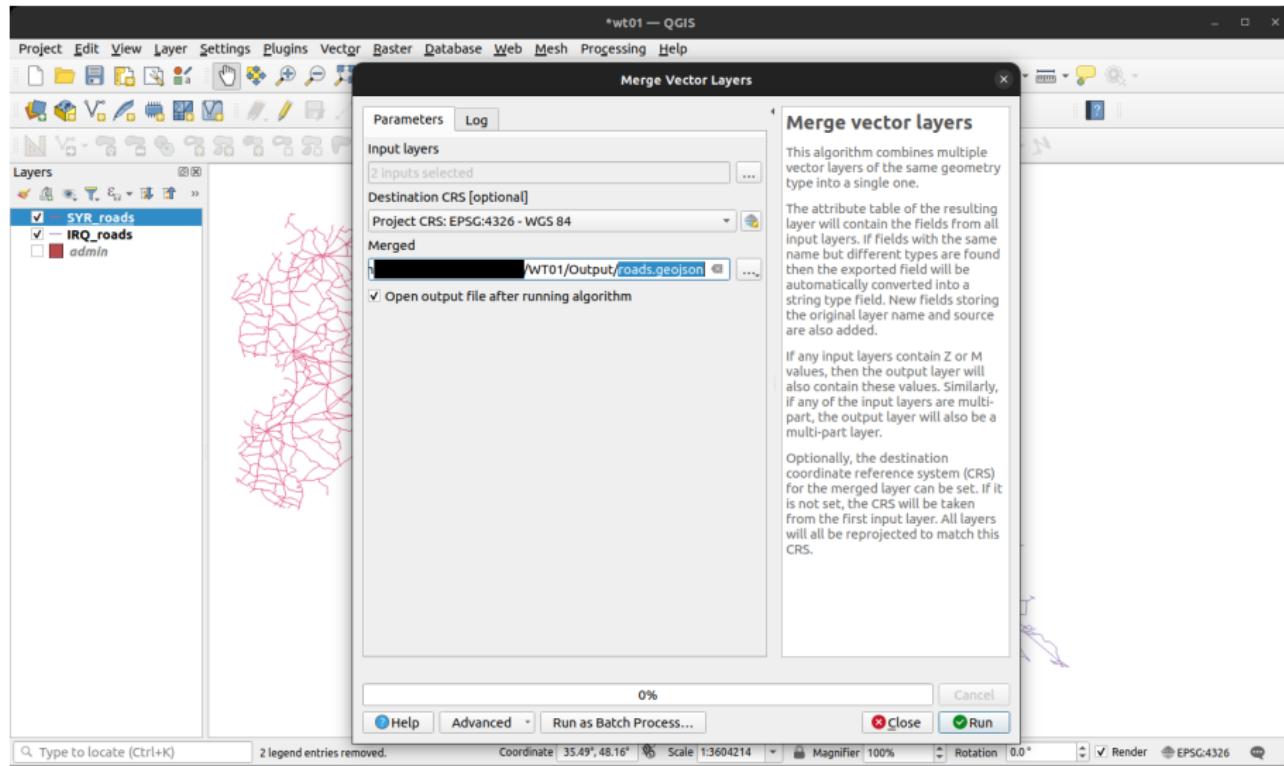
Let's repeat the Merge Vector Layers procedure for these two layers (Vector → Data Management Tool → Merge Vector Layers).



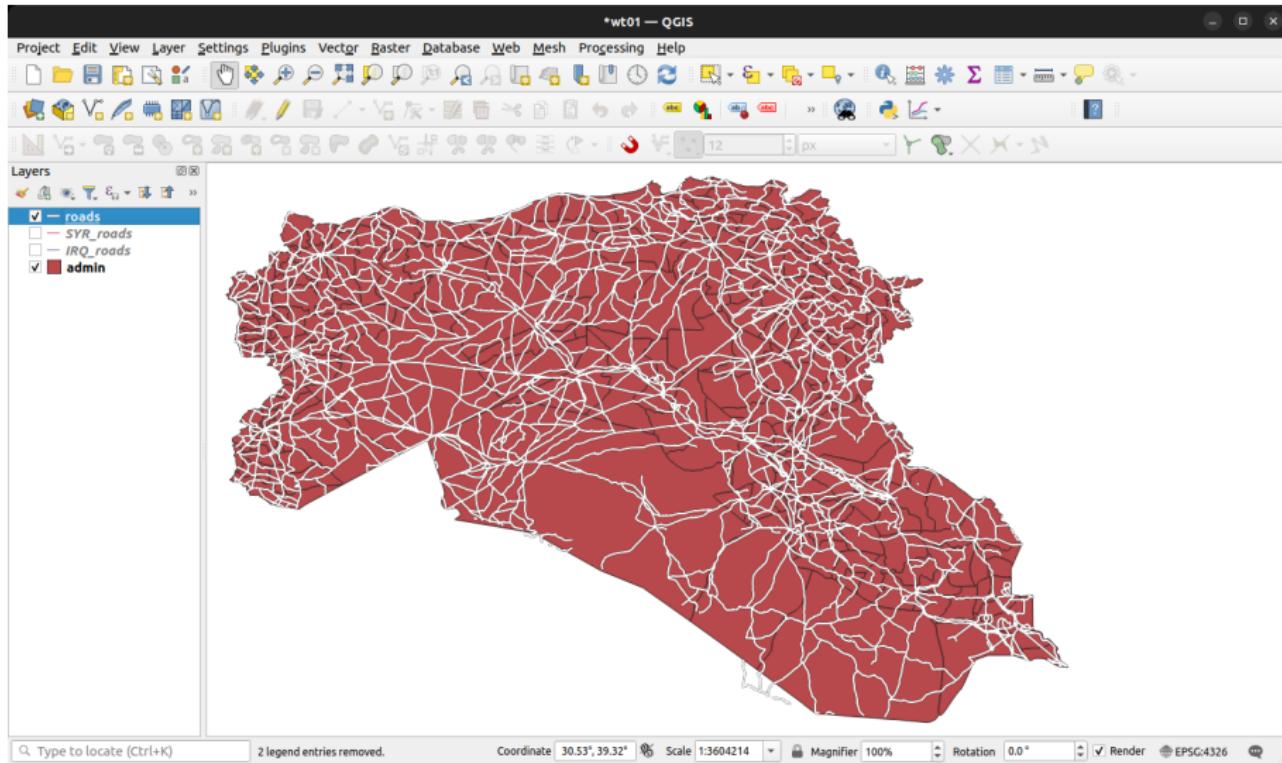
This time, perform the merge on the IRQ_roads and SYR_roads layers



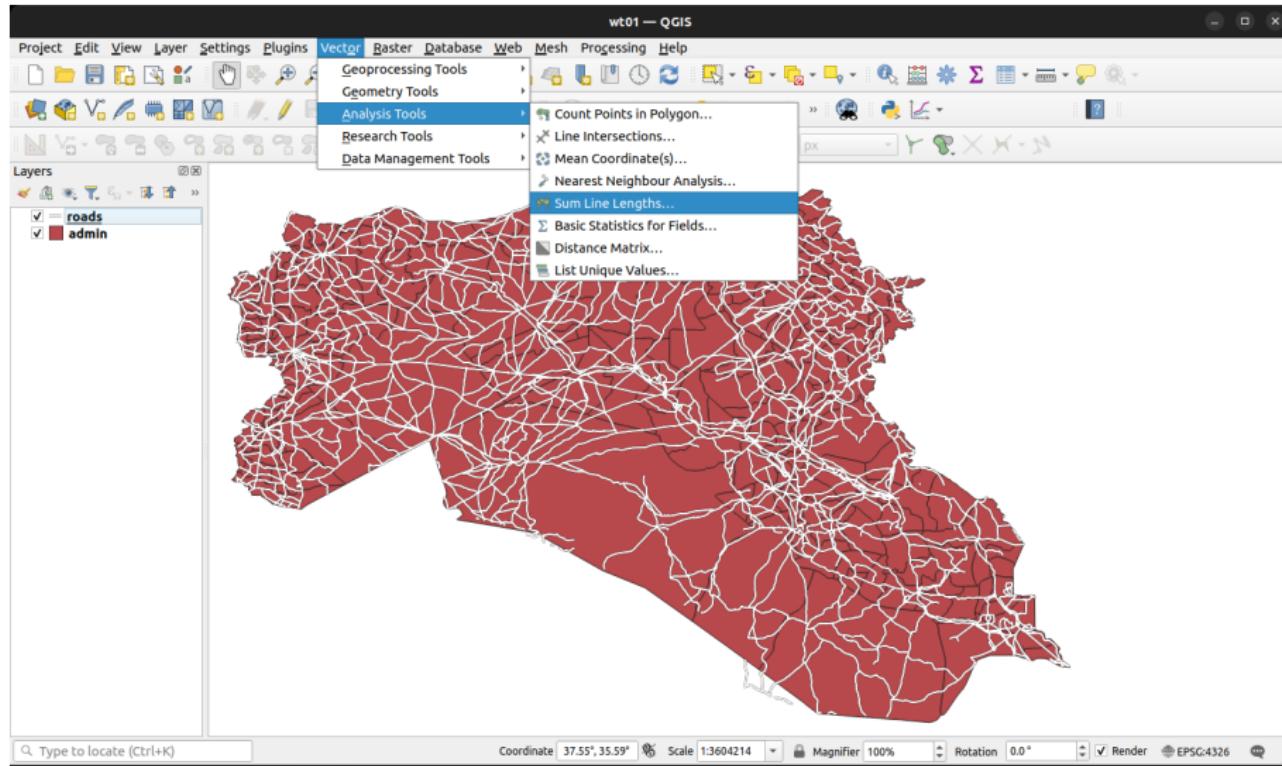
Save the output file as roads.geojson



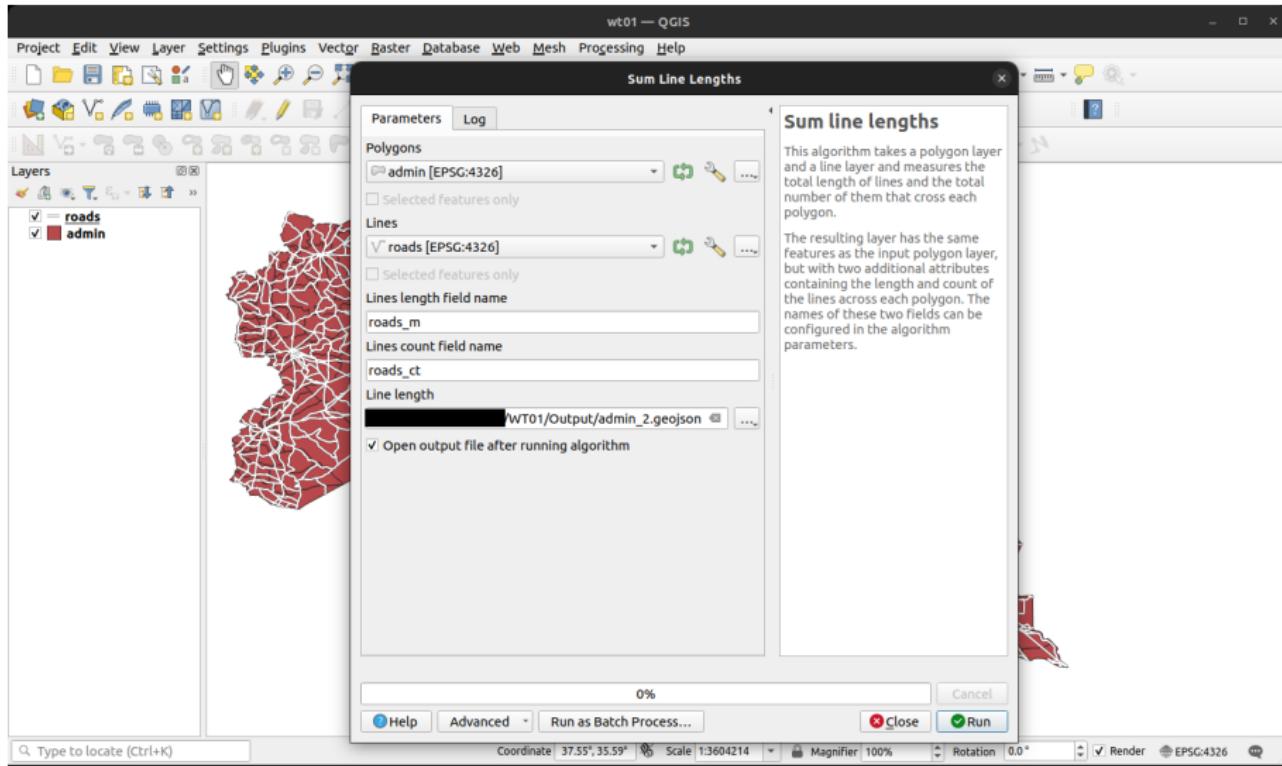
As before, you can remove the two country-specific roads layers (we no longer need them)



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



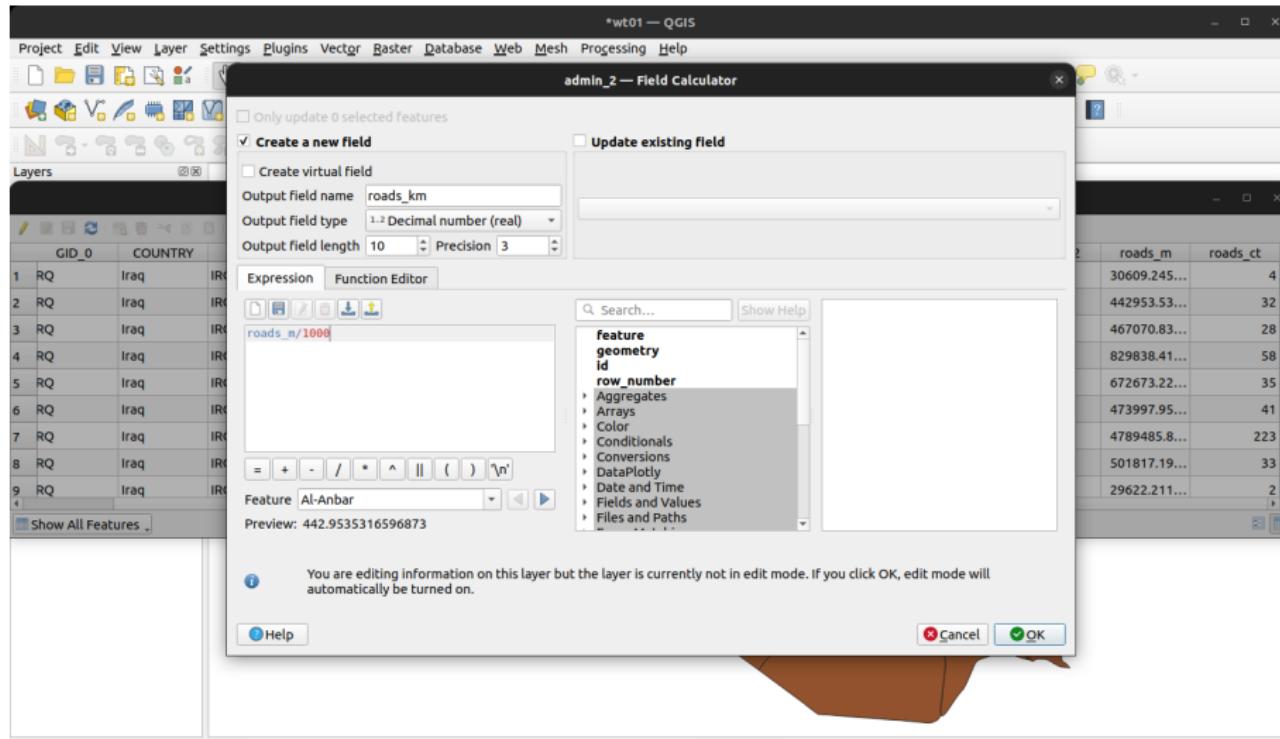
Open the attribute table for admin_2. The roads_m and roads_ct variables should be in the table. Let's convert from **meters to kilometers**

The screenshot shows the QGIS application interface. At the top, the title bar reads "wt01 — QGIS". Below it is the menu bar with options: Project, Edit, View, Layer, Settings, Plugins, Vector, Raster, Database, Web, Mesh, Processing, Help. The toolbar below the menu bar contains various icons for editing, selection, and analysis. The "Layers" panel on the left shows a single layer named "admin_2". The main workspace displays a map of Iraq with administrative boundaries. Below the map is the attribute table for "admin_2", which includes columns for 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 table lists 162 features, with the first few rows shown below:

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	IRQ_1_1	Al-Anbar	NA	AbuGhraib	NA	NA	Kaza	District	NA	NA	30609.245...	4
2	RQ	IRQ_1_1	Al-Anbar	NA	AlFallujah	NA	NA	Kaza	District	NA	NA	442953.53...	32
3	RQ	IRQ_1_1	Al-Anbar	NA	AlHaditha	NA	NA	Kaza	District	NA	NA	467070.83...	28
4	RQ	IRQ_1_1	Al-Anbar	NA	AlQa'lim	NA	NA	Kaza	District	NA	NA	829838.41...	58
5	RQ	IRQ_1_1	Al-Anbar	NA	Anah	NA	NA	Kaza	District	NA	NA	672673.22...	35
6	RQ	IRQ_1_1	Al-Anbar	NA	ArRamadi	NA	NA	Kaza	District	NA	NA	473997.95...	41
7	RQ	IRQ_1_1	Al-Anbar	NA	ArRutbah	NA	NA	Kaza	District	NA	NA	4789485.8...	223
8	RQ	IRQ_1_1	Al-Anbar	NA	Hit	NA	NA	Kaza	District	NA	NA	501817.19...	33
9	RQ	IRQ_1_1	Al-Anbar	NA	Kadhimiya	NA	NA	Kaza	District	NA	NA	29622.211...	2

At the bottom of the interface, there is a search bar labeled "Type to locate (Ctrl+K)" and a set of coordinate and scale controls.

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

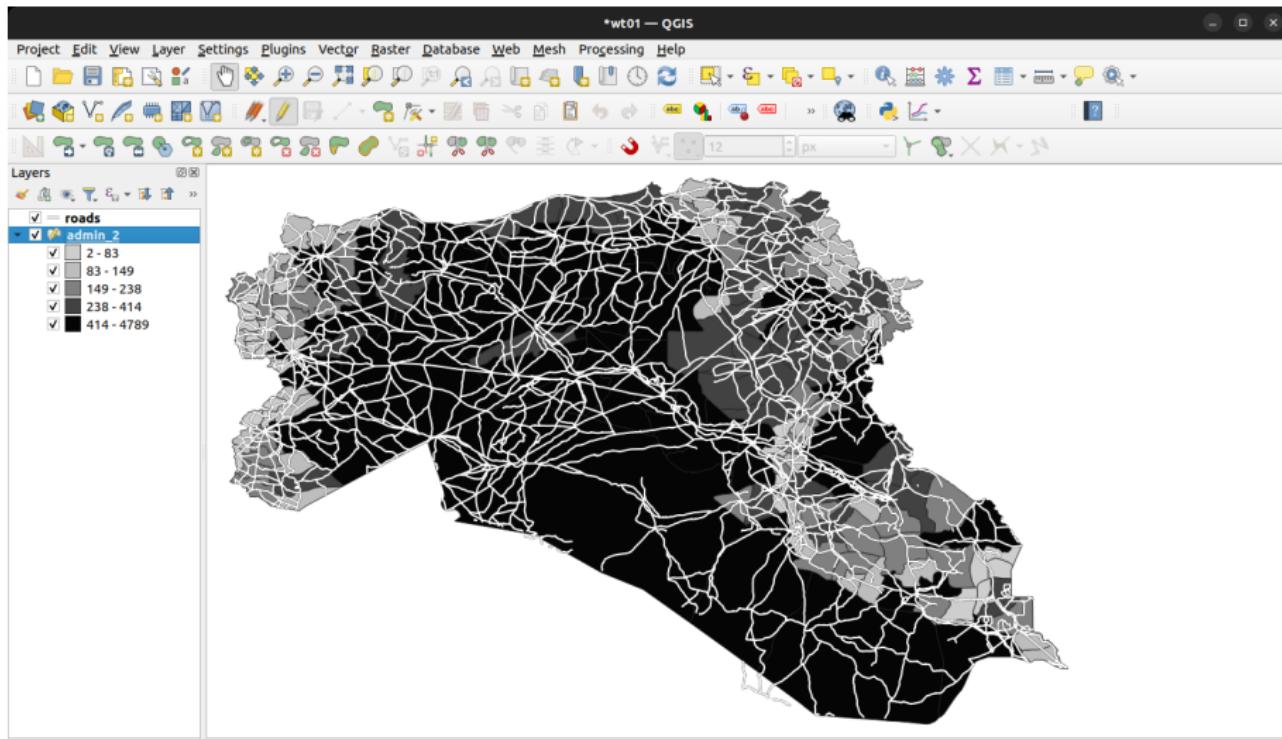


The resulting `roads_km` field should be visible in the attribute table

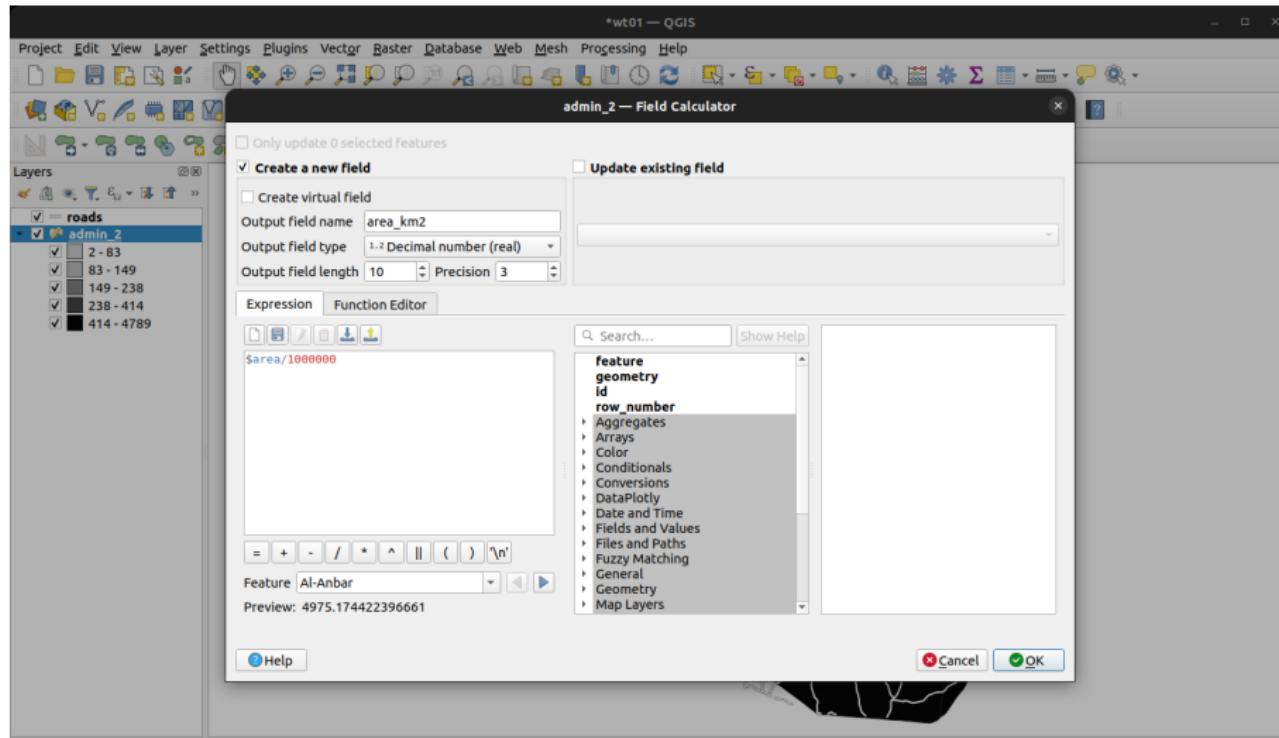
The screenshot shows the QGIS application interface. The title bar reads "#wt01 — QGIS". The menu bar includes Project, Edit, View, Layer, Settings, Plugins, Vector, Raster, Database, Web, Mesh, Processing, and Help. The toolbar contains various icons for file operations, selection, measurement, and analysis. The layers panel shows a single layer named "admin_2". The main window displays a map of Iraq with district boundaries. Below the map is an attribute table for the "admin_2" layer. The table has 16 columns: 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, and roads_km. The data shows 162 features, all filtered and selected. The "roads_km" column contains values such as 30609.245..., 442953.53..., 467070.83..., etc. The bottom of the screen shows the QGIS status bar with coordinates (31.64°, 48.49°), scale (1:3604214), magnifier (100%), rotation (0.0°), render status, and EPSG:4326 projection.

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	roads_km
Iraq	IRQ.1_1	Al-Anbar	NA	AbuGhraib	NA	NA	Kaza	District	NA	NA	30609.245...	4	30.609
Iraq	IRQ.1_1	Al-Anbar	NA	AlFallujah	NA	NA	Kaza	District	NA	NA	442953.53...	32	442.954
Iraq	IRQ.1_1	Al-Anbar	NA	AlHaditha	NA	NA	Kaza	District	NA	NA	467070.83...	28	467.071
Iraq	IRQ.1_1	Al-Anbar	NA	AlQa'im	NA	NA	Kaza	District	NA	NA	829838.41...	58	829.838
Iraq	IRQ.1_1	Al-Anbar	NA	Anah	NA	NA	Kaza	District	NA	NA	672673.22...	35	672.673
Iraq	IRQ.1_1	Al-Anbar	NA	ArRamadi	NA	NA	Kaza	District	NA	NA	473997.95...	41	473.998
Iraq	IRQ.1_1	Al-Anbar	NA	ArRutbah	NA	NA	Kaza	District	NA	NA	4789485.8...	223	4789.486
Iraq	IRQ.1_1	Al-Anbar	NA	Hit	NA	NA	Kaza	District	NA	NA	501817.19...	33	501.817

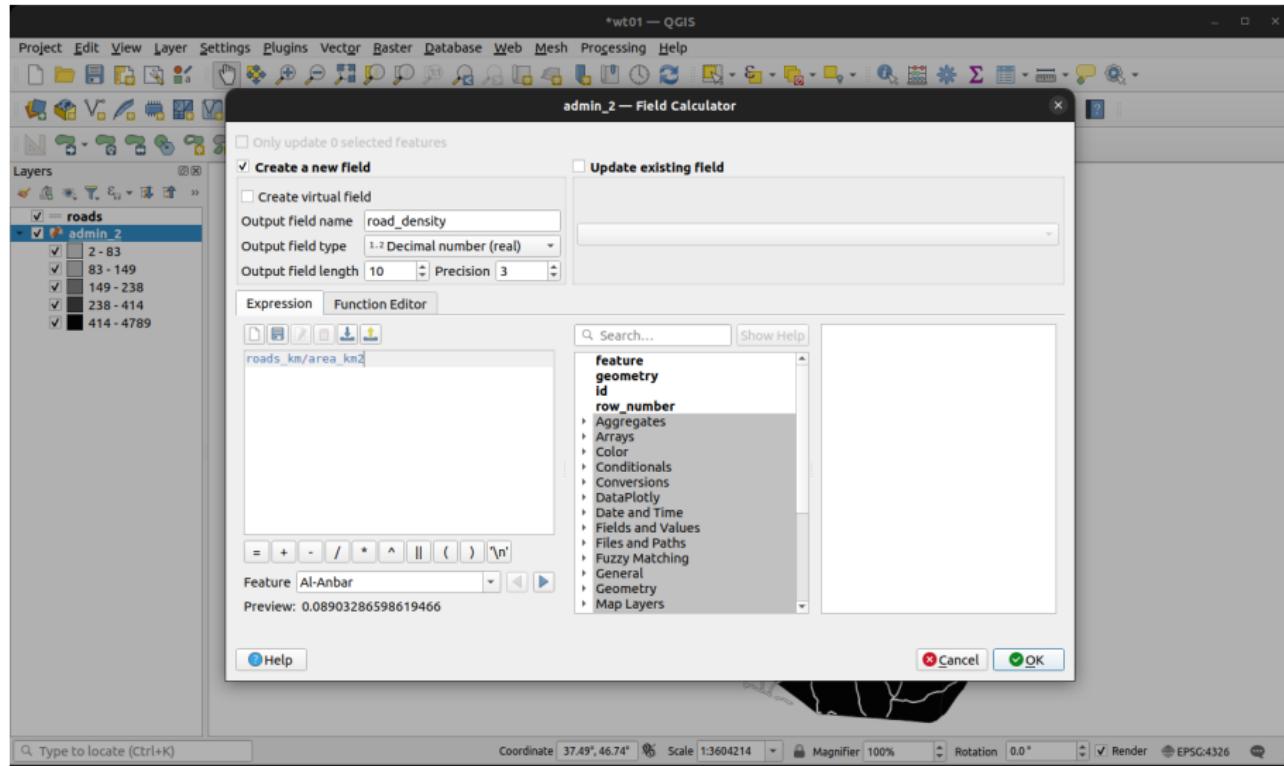
If we plot this variable, we can see its main limitations: smaller districts have less road length, so this becomes a proxy for district size. Let's create a **road density** measure that takes into account district size



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 m² to km²). 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



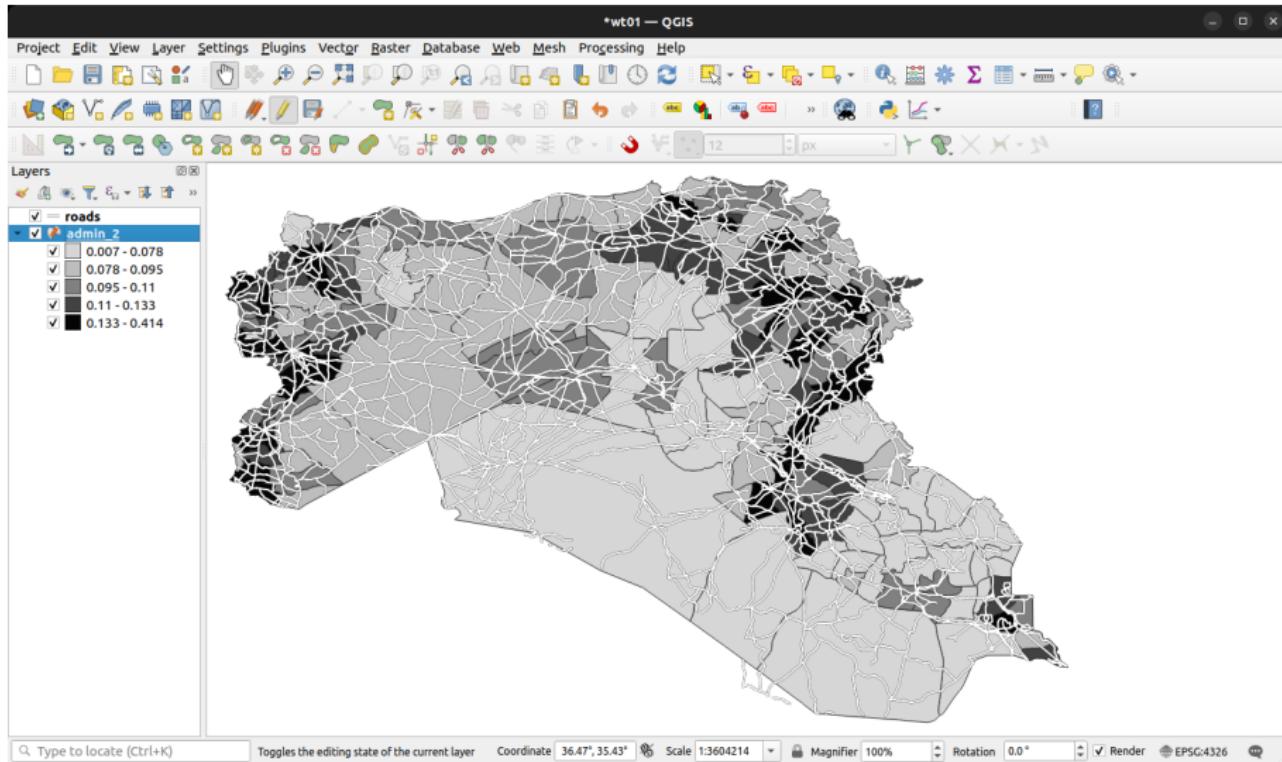
Check to make sure road_density is added to attribute table

The screenshot shows a QGIS interface with the title bar "#wt01 — QGIS". The menu bar includes Project, Edit, View, Layer, Settings, Plugins, Vector, Raster, Database, Web, Mesh, Processing, and Help. The toolbar contains various icons for file operations, selection, measurement, and analysis. The layers panel shows the 'admin_2' layer is selected, with a status bar indicating Features Total: 162, Filtered: 162, Selected: 0. Below the toolbar is a search bar with the placeholder 'abc'. The main window displays an attribute table for the 'admin_2' layer, which includes columns for ID, NAME_1, NL_NAME_1, NAME_2, VARNAME_2, NL_NAME_2, TYPE_2, ENCTYPE_2, CC_2, HASC_2, roads_m, roads_ct, roads_km, area_km2, and road_density. The table lists 162 features, with the first few rows shown below:

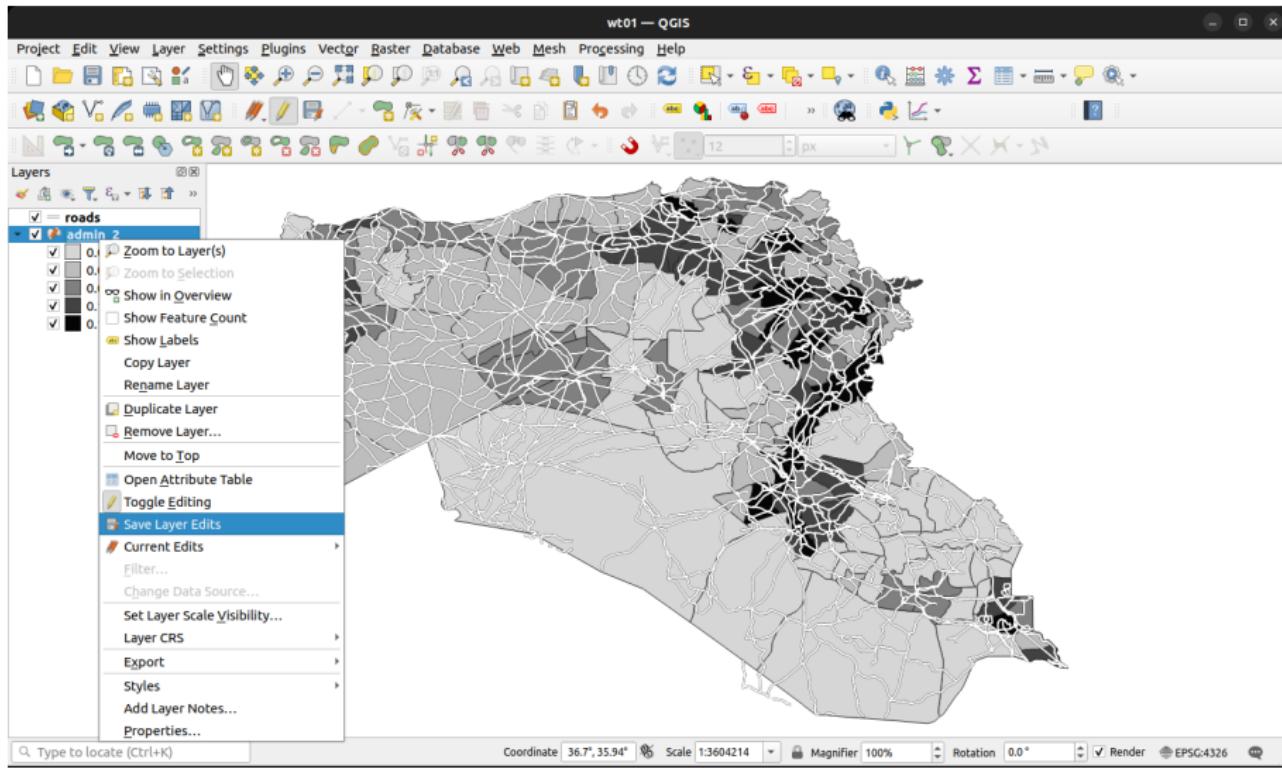
ID	NAME_1	NL_NAME_1	NAME_2	VARNAME_2	NL_NAME_2	TYPE_2	ENCTYPE_2	CC_2	HASC_2	roads_m	roads_ct	roads_km	area_km2	road_density
1	Al-Anbar	NA	AbuGhraib	NA	NA	Kaza	District	NA	NA	30609.245...	4	30.609	411.653	0.074
2	Al-Anbar	NA	AlFallujah	NA	NA	Kaza	District	NA	NA	442953.53...	32	442.954	4975.174	0.089
3	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

Below the table is a map view showing the administrative boundaries of Al-Anbar Governorate and a detailed road network. The map uses grayscale shading for different regions. At the bottom of the interface are various toolbars and status indicators.

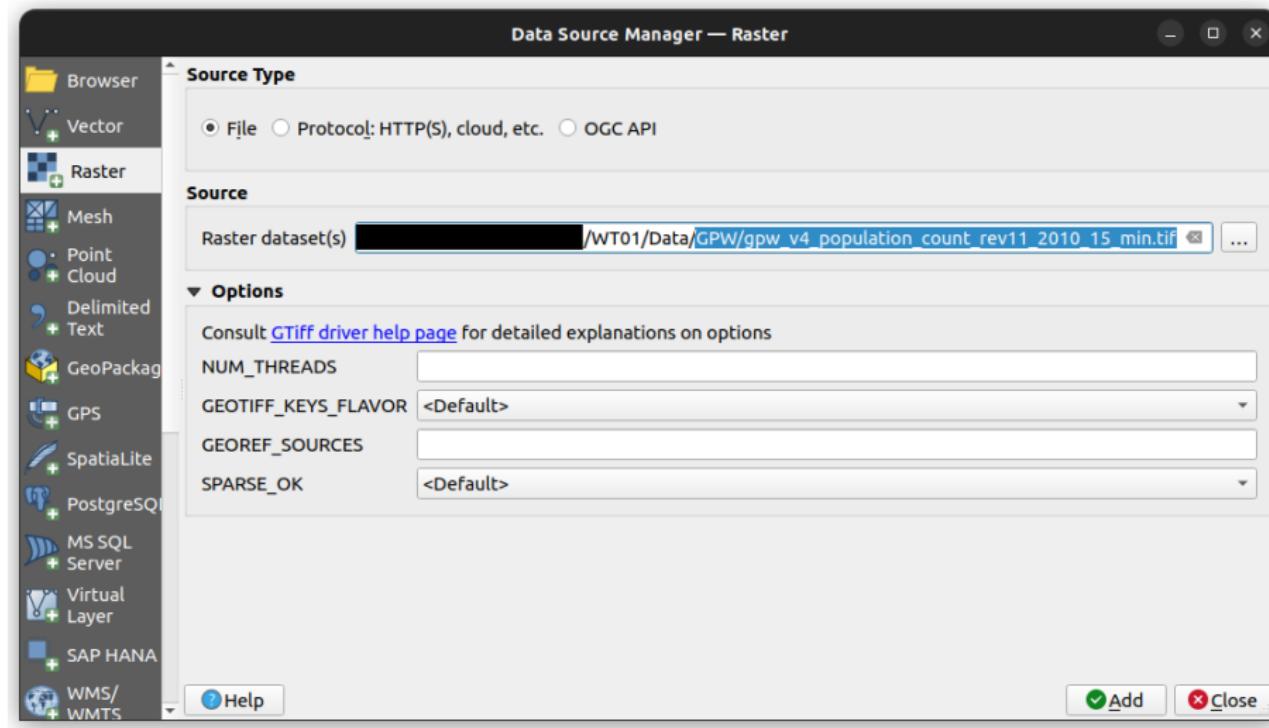
If we plot this variable, we see the opposite pattern from before (smaller, urban districts have higher road density)



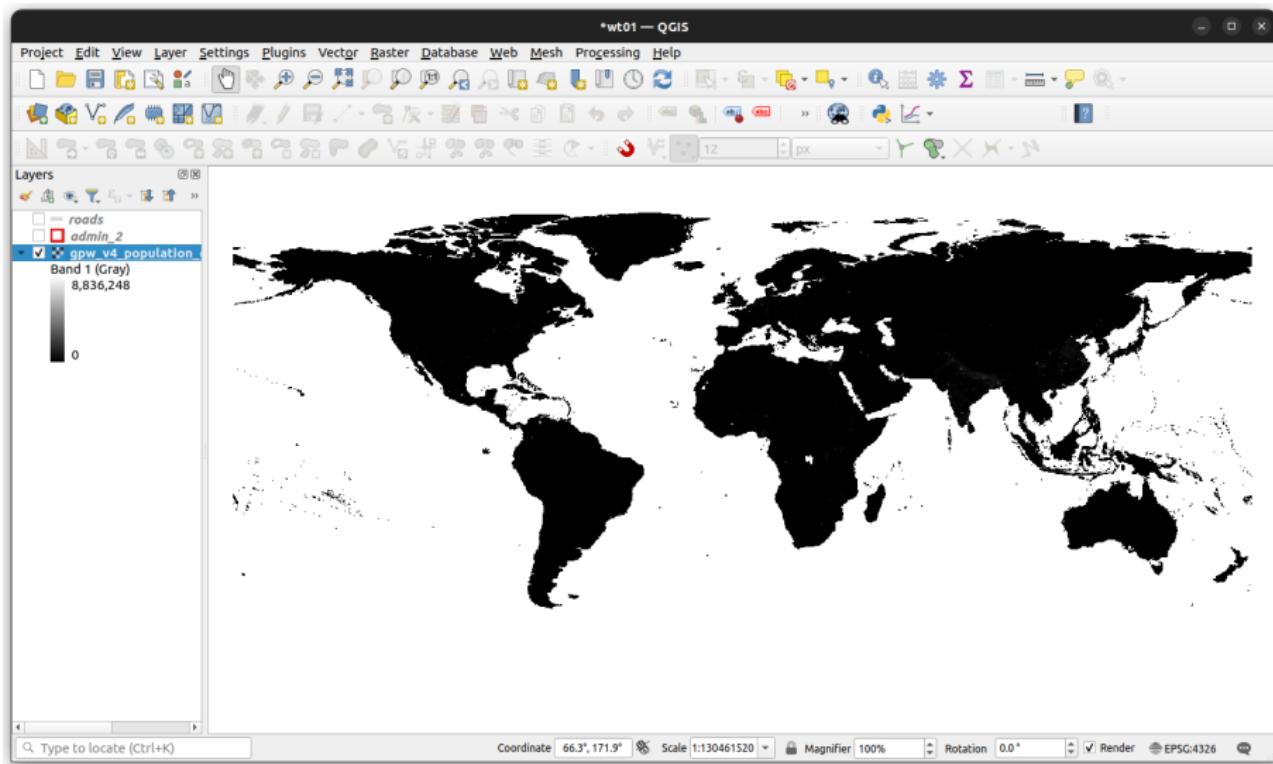
Save the edits you just made to the admin_2 layer (right click → Save Layer Edits)



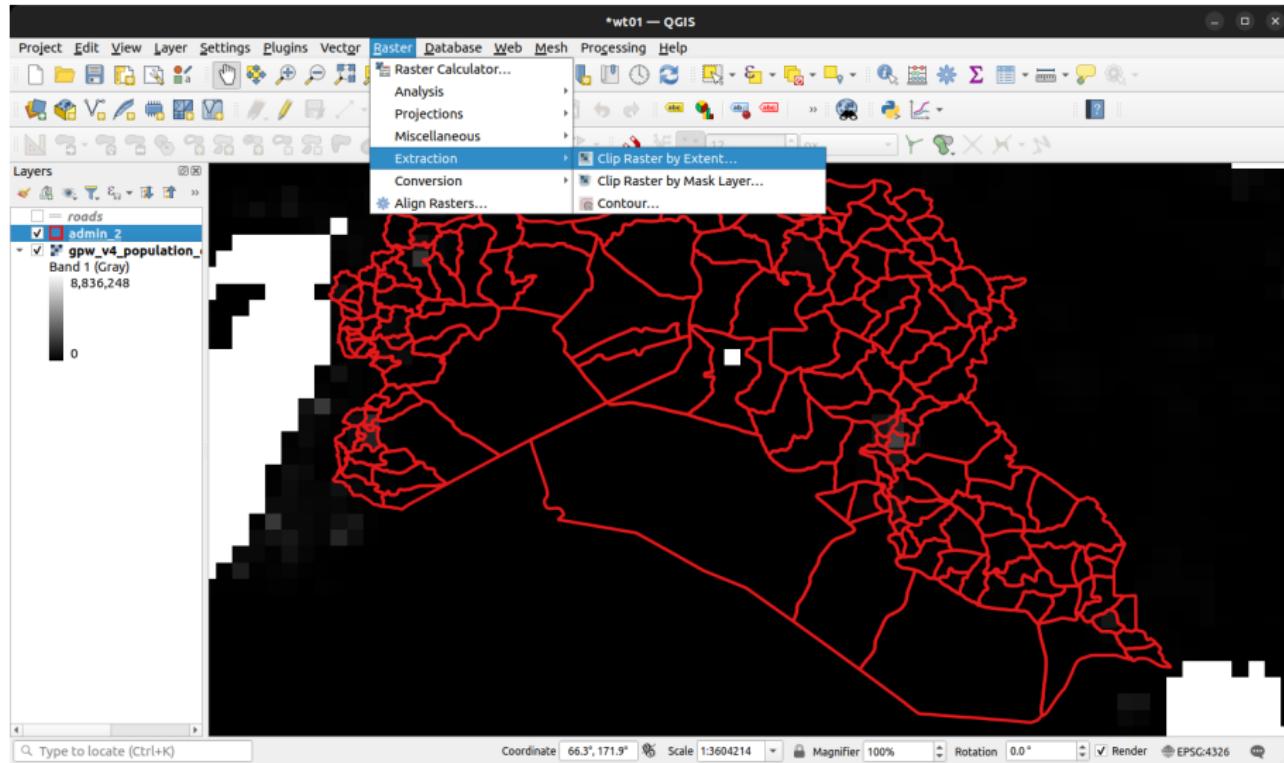
Load *population* raster (Layer → Add Layer → Add Raster Layer).\\ Open the gpw_v4_population_count_rev11_2010_15_min.tif file in Data/GPW folder.



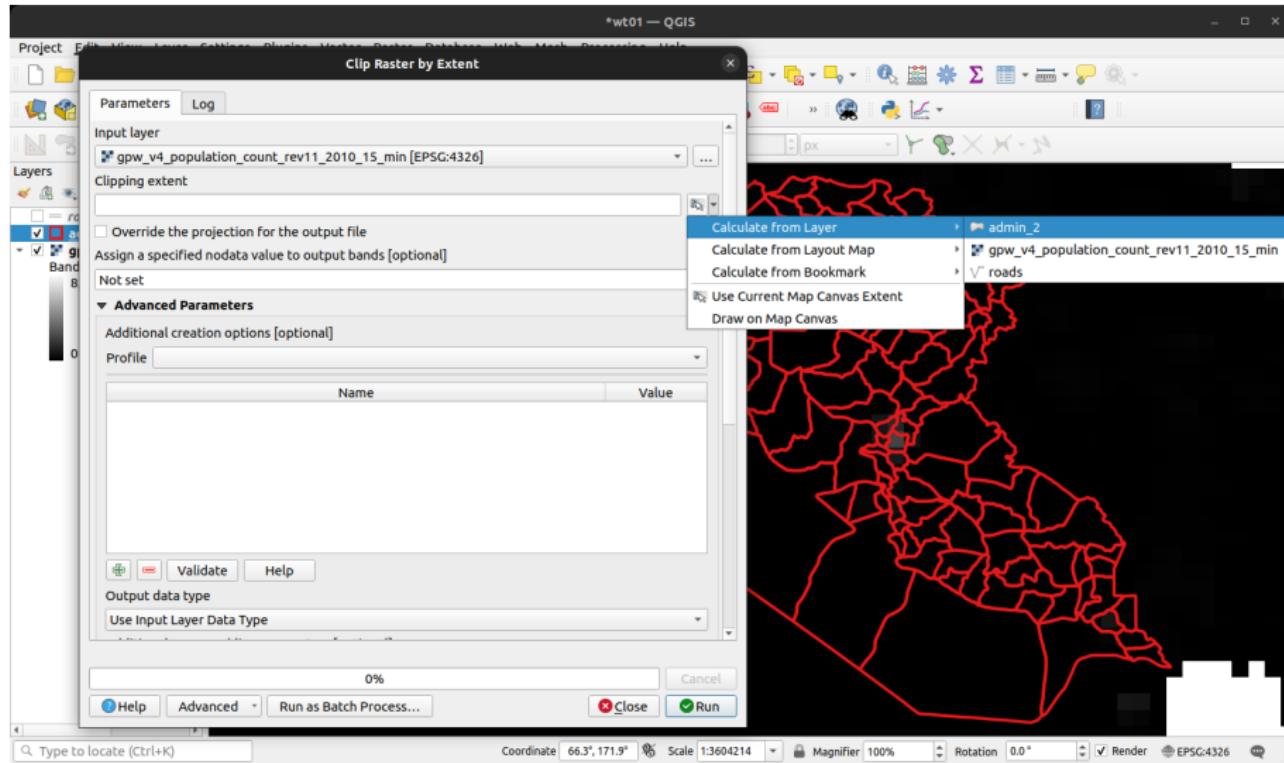
The (global) population layer should now be visible. Let's **extract a subset** of just the part of this layer that overlaps with Syria and Iraq.



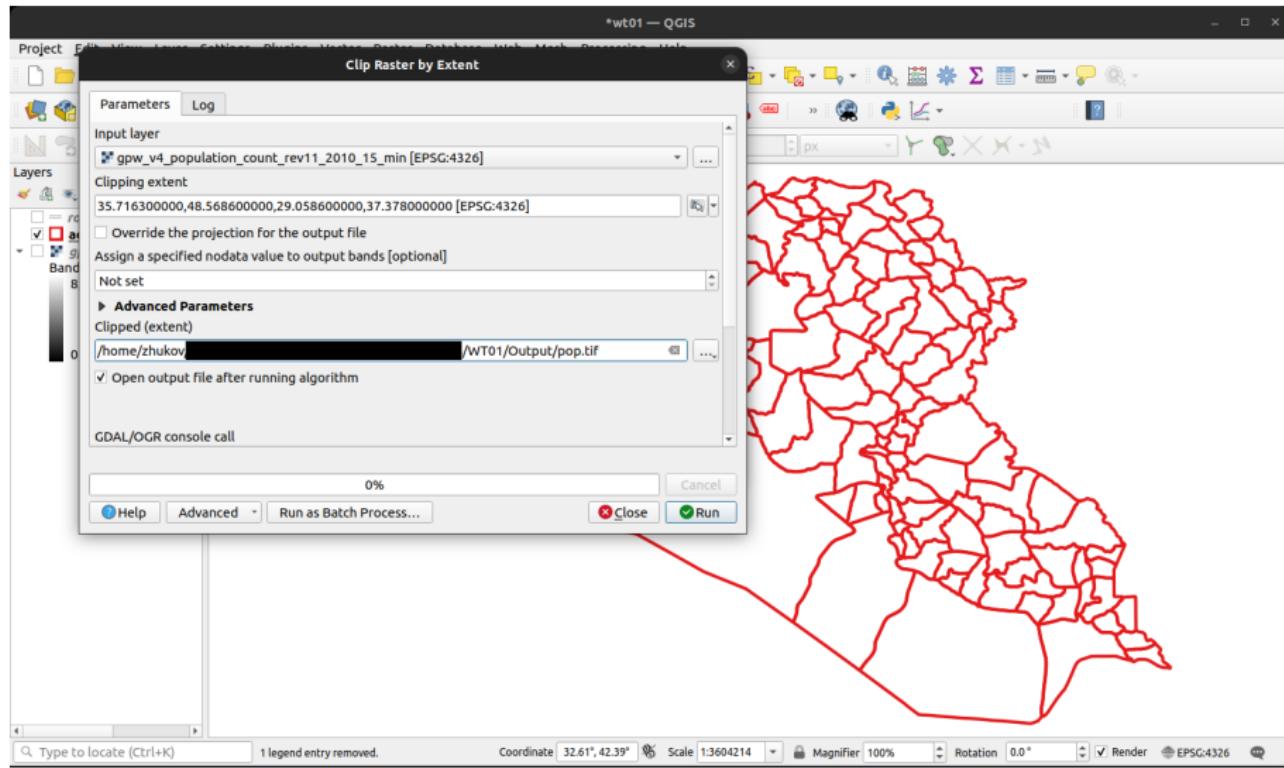
Open the Clip Raster by Extent tool (Raster → Extension → Clip Raster by Extent)



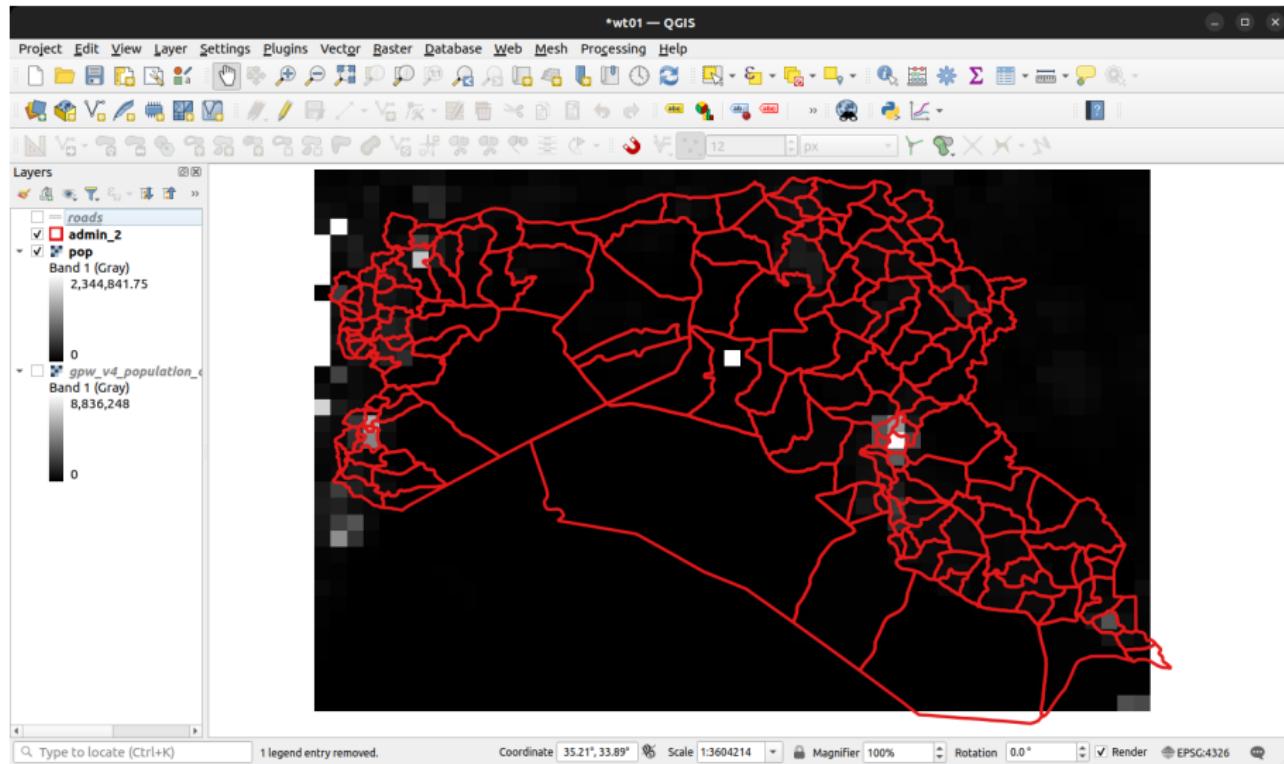
Set Input layer = gpw_v4_population_count_rev11_2010_15_min. Click the [...] button next to Clipping extent → Calculate from Layer → admin_2



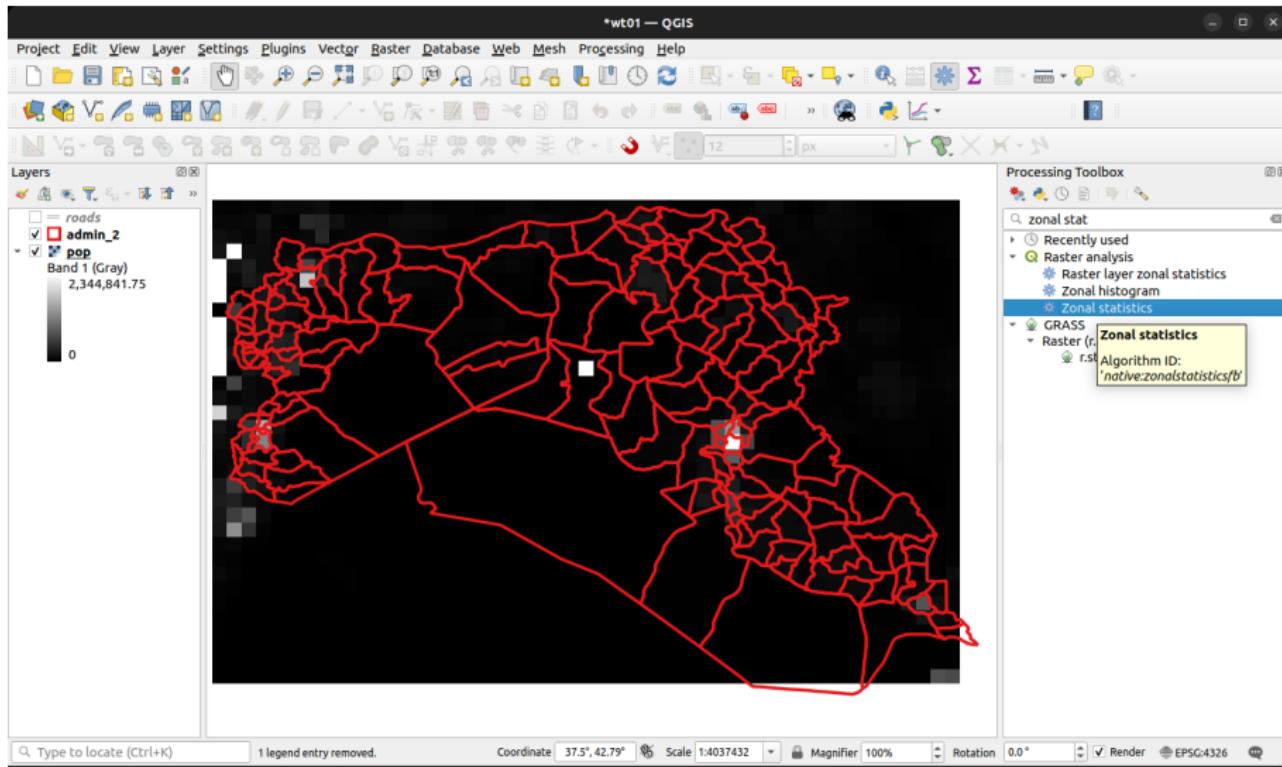
Save the file as pop.tif. Accept defaults for all other parameters, and click Run



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

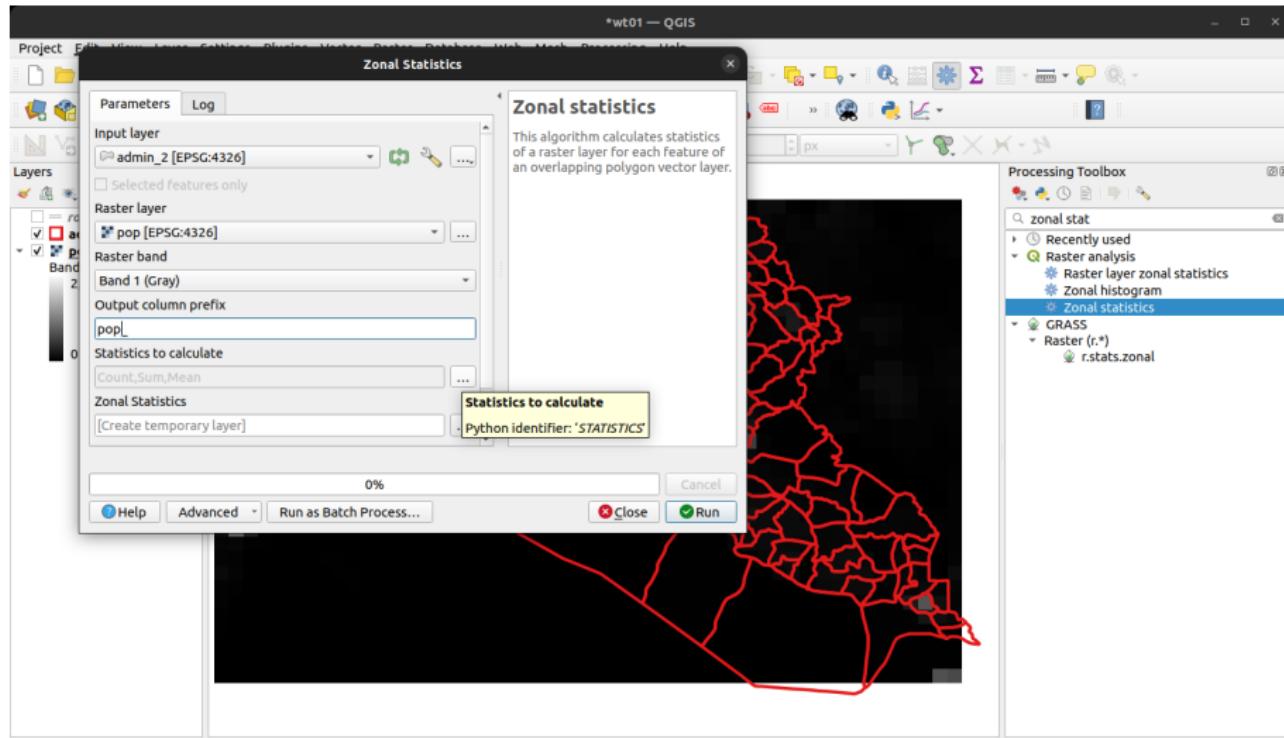


Let's calculate the population size of each district. Open the Processing Toolbox and open the Zonal statistics tool in the Raster Analysis menu

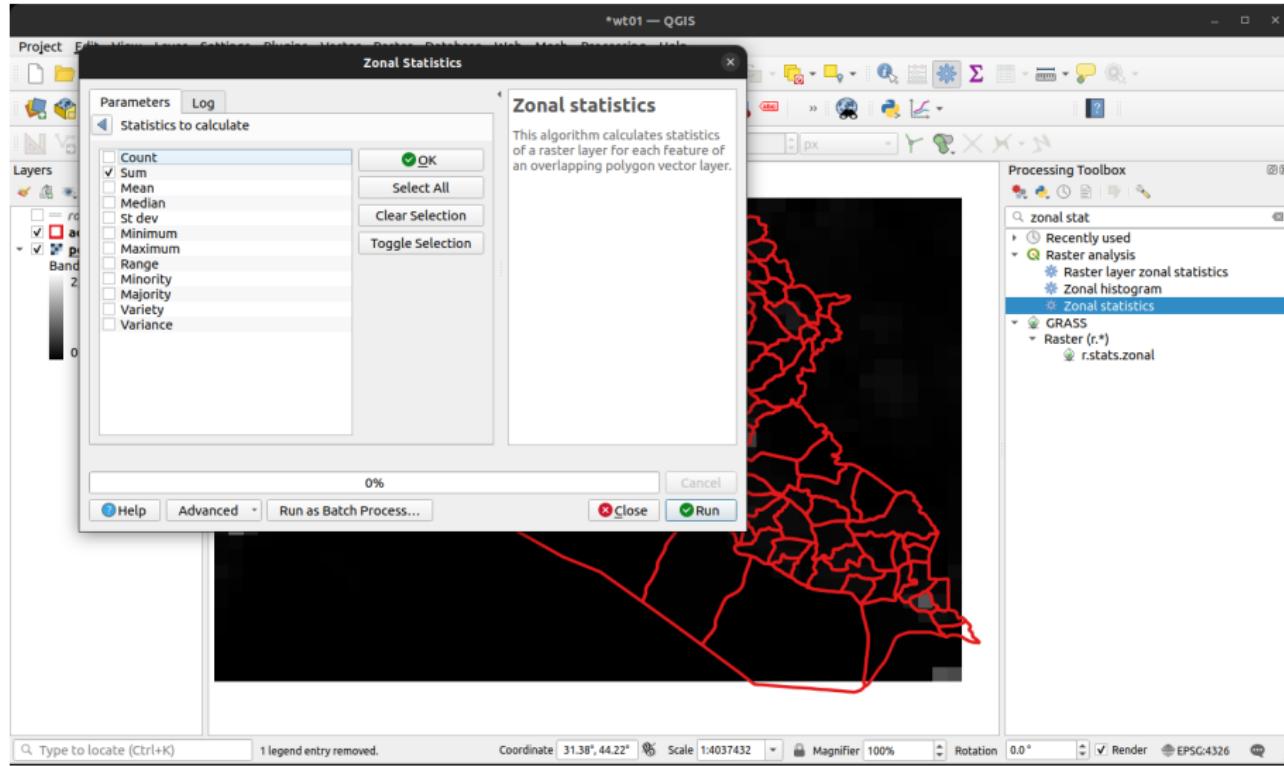


In the Zonal Statistics tool, set Input layer=admin_2, Raster layer=pop, and Output column prefix=pop_.

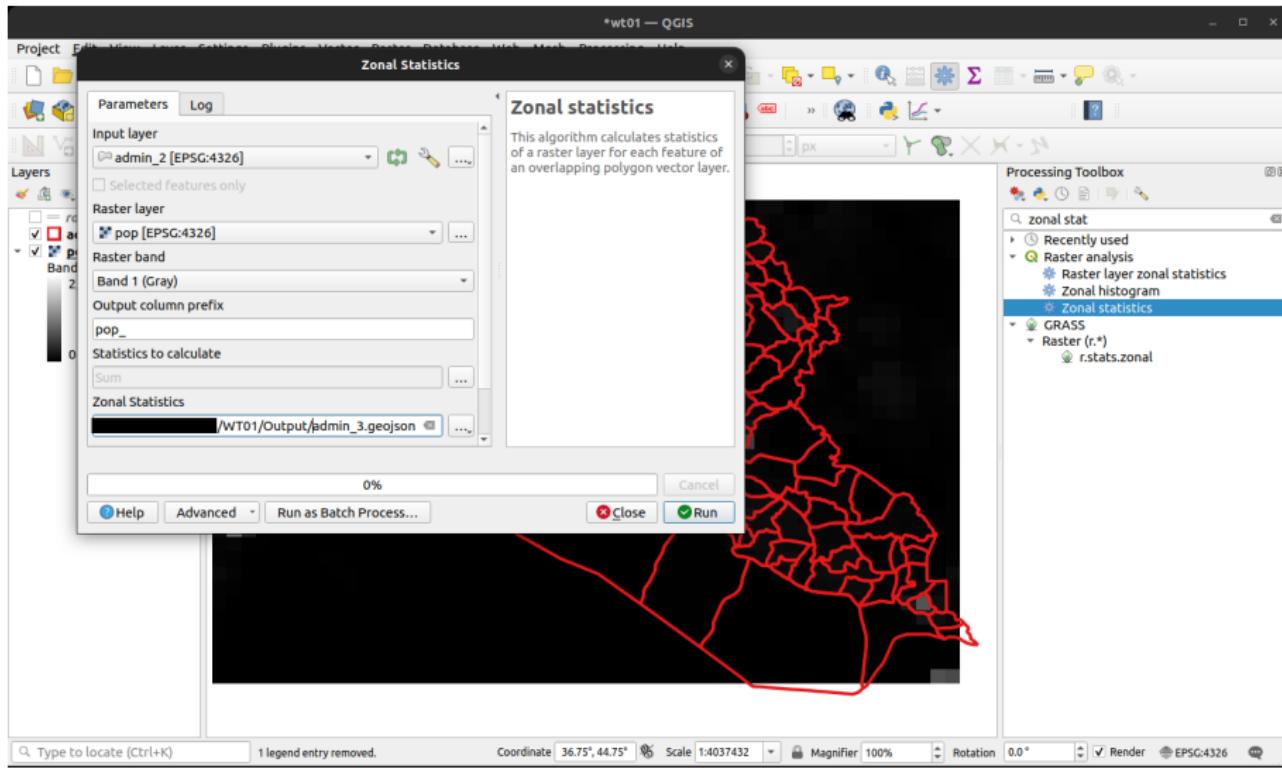
Click the [...] button next to Statistics to calculate



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)

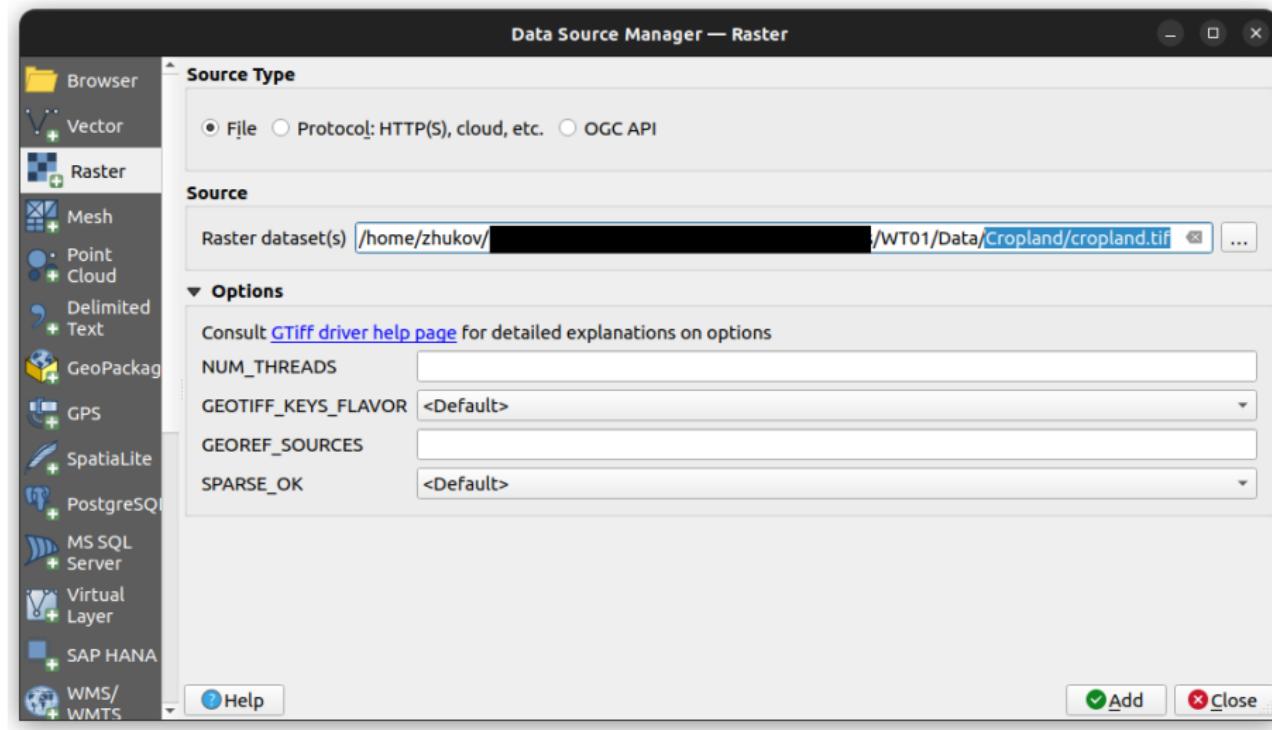
The screenshot shows the QGIS interface with the title bar "wt01 — QGIS". The menu bar includes Project, Edit, View, Layer, Settings, Plugins, Vector, Raster, Database, Web, Mesh, Processing, and Help. The toolbar has various icons for selection, measurement, and editing. The Layers panel on the left shows three layers: "admin_3" (selected), "roads", and "pop". The map canvas displays a purple-shaded map of Jordan divided into administrative districts. The Processing Toolbox on the right is open, showing the "zonal stat" group expanded, with "Zonal statistics" selected. The attribute table for "admin_3" is open at the bottom, showing 162 features. The "pop_sum" column is highlighted in yellow.

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

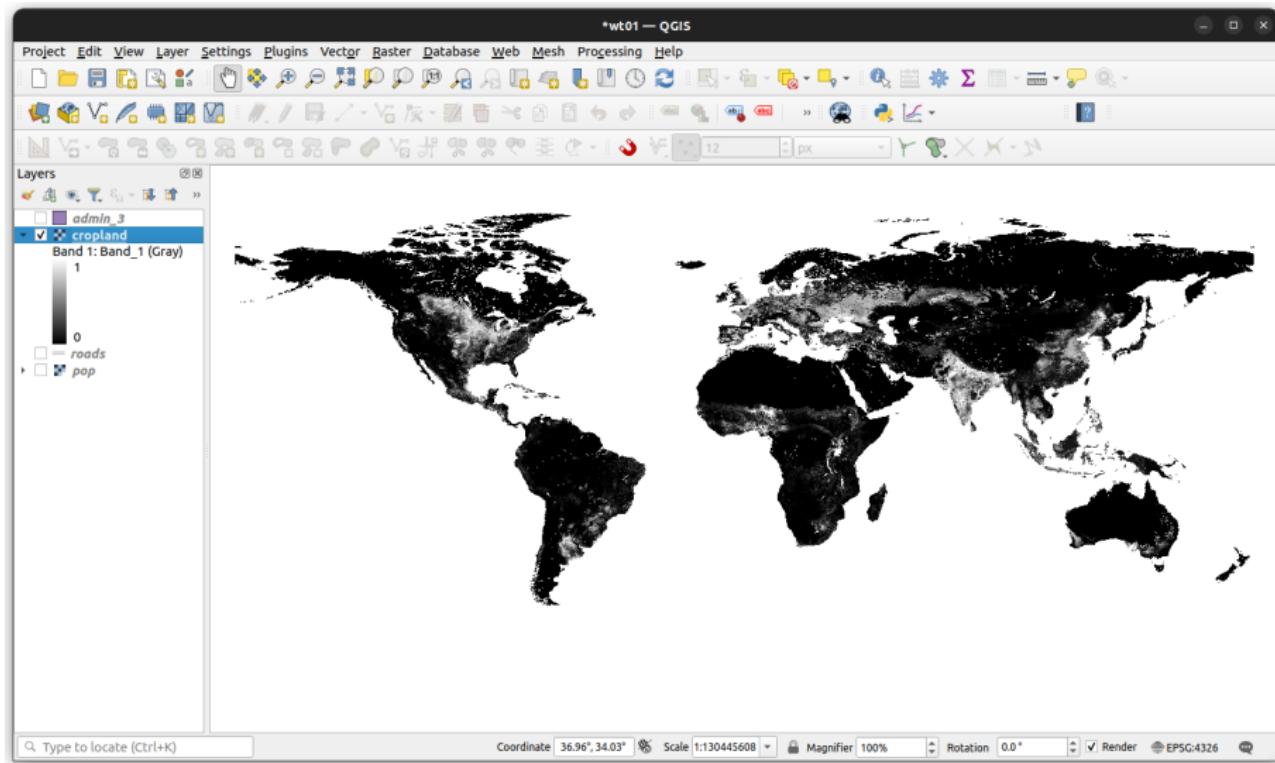
Legend entry removed.

Coordinate: 37.94°, 43.75° | Scale: 1:4037432 | Magnifier: 100% | Rotation: 0.0° | Render | EPSG:4326

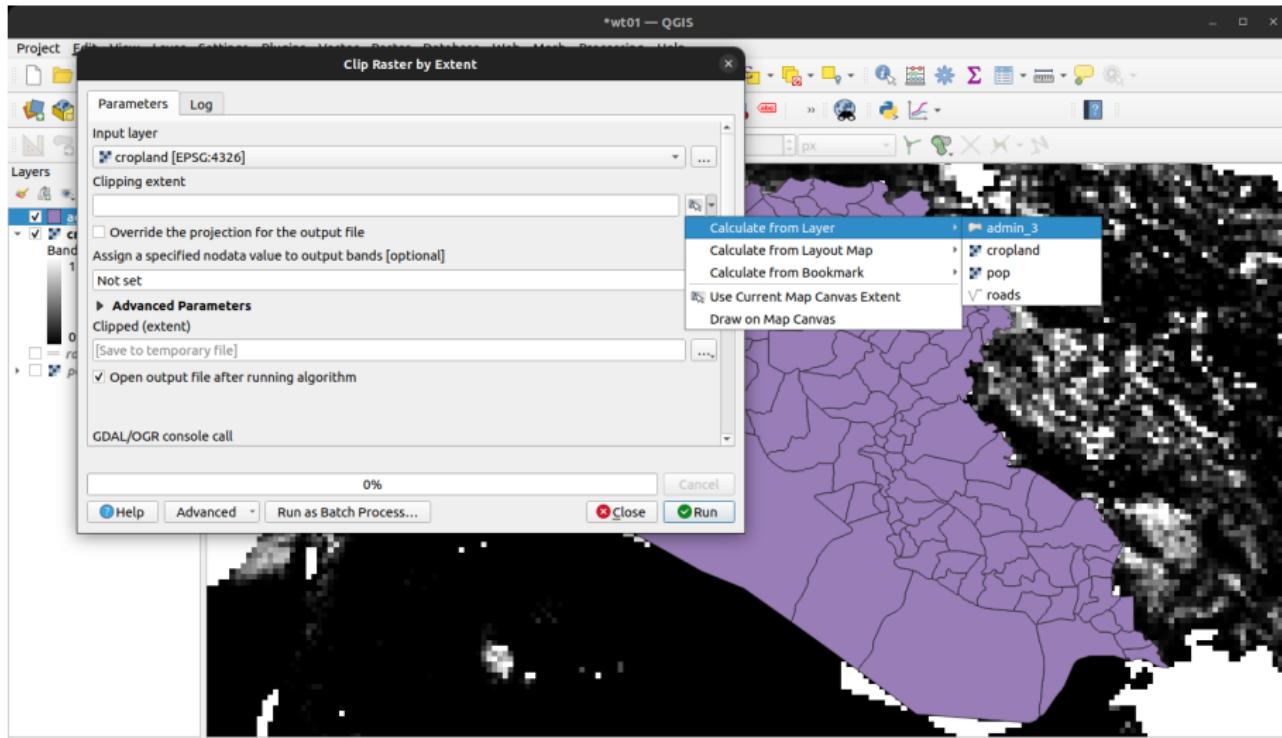
Load *cropland* raster (Layer → Add Layer → Add Raster Layer).
Open the *cropland.tif* file in Data/Cropland folder.



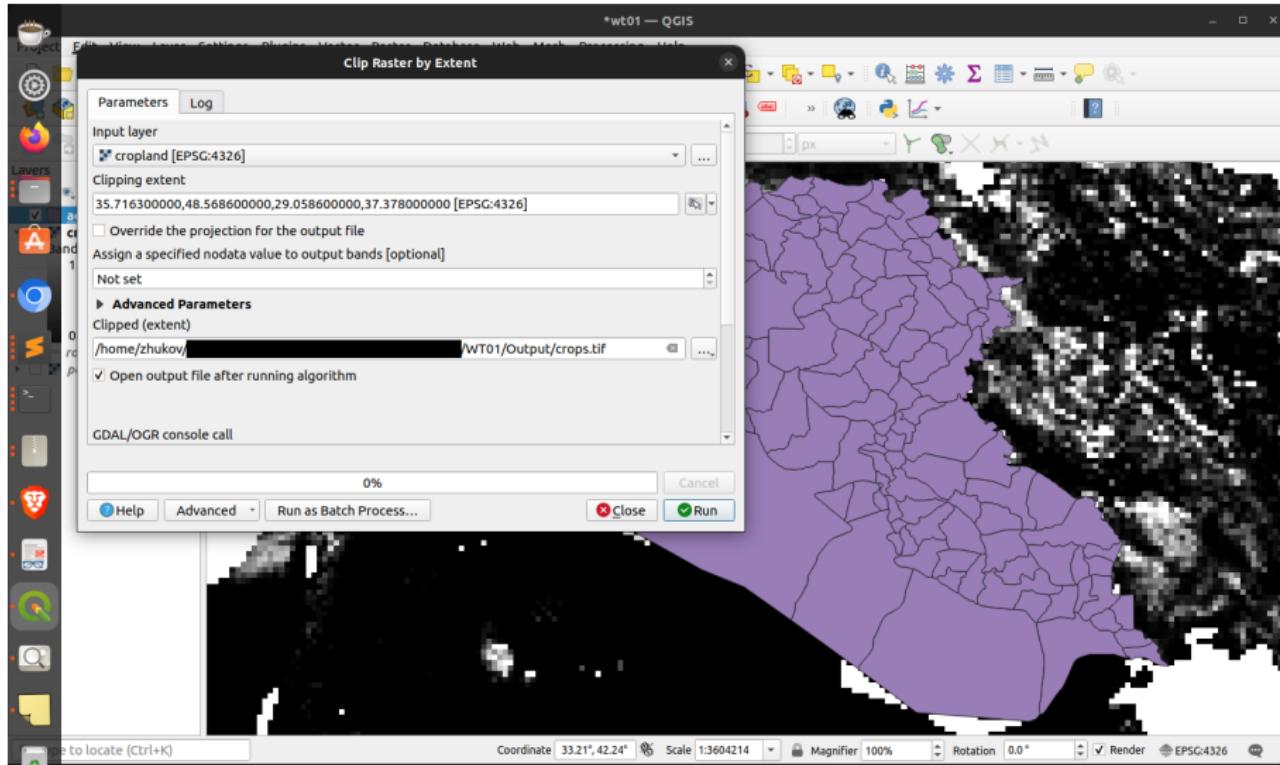
The (global) crops layer should now be visible. Let's **extract a subset** of it, as we just did for population



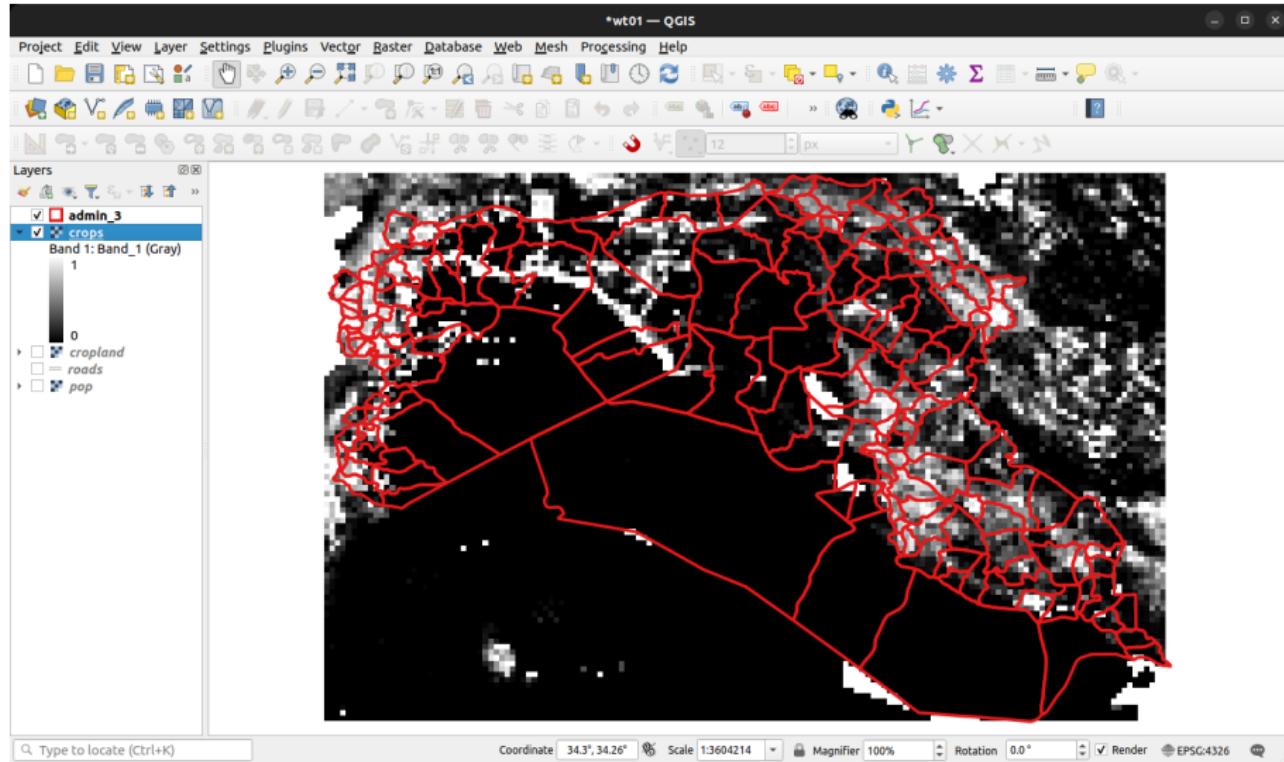
Open the Clip Raster by Extent tool (Raster → Extension → Clip Raster by Extent). Set Input layer = cropland. Click the [...] button next to Clipping extent → Calculate from Layer → admin_3



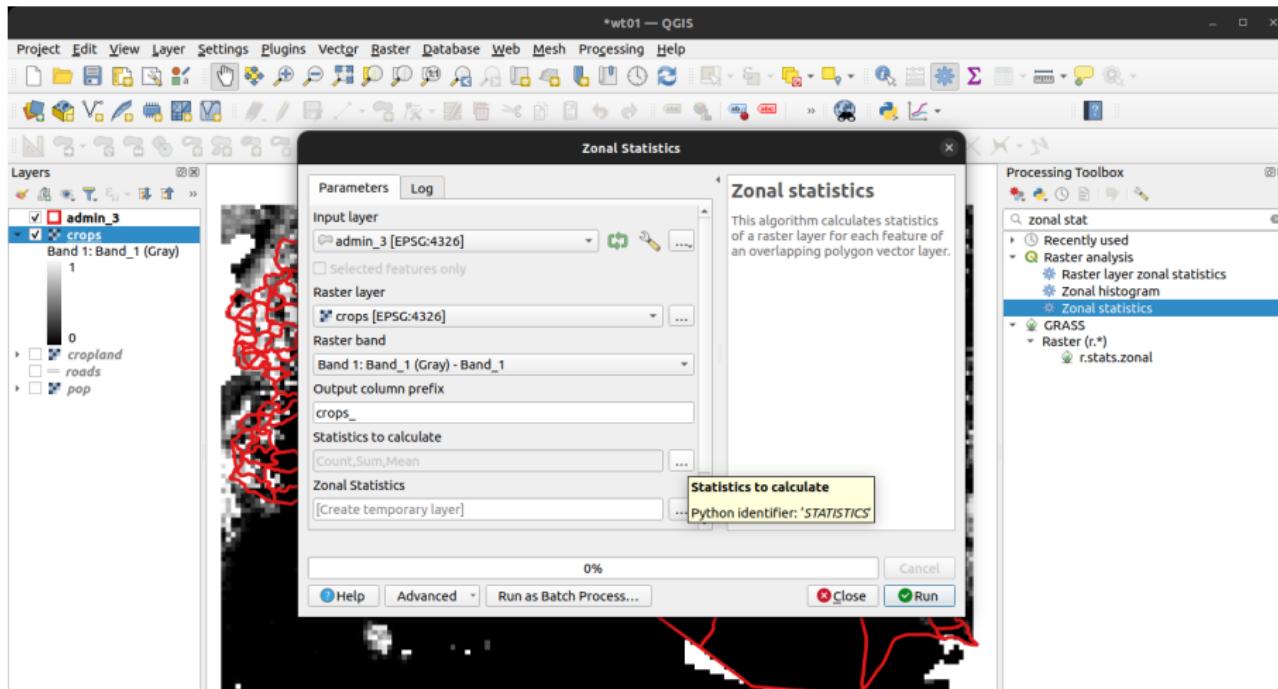
Save the output to a file called crops.tif



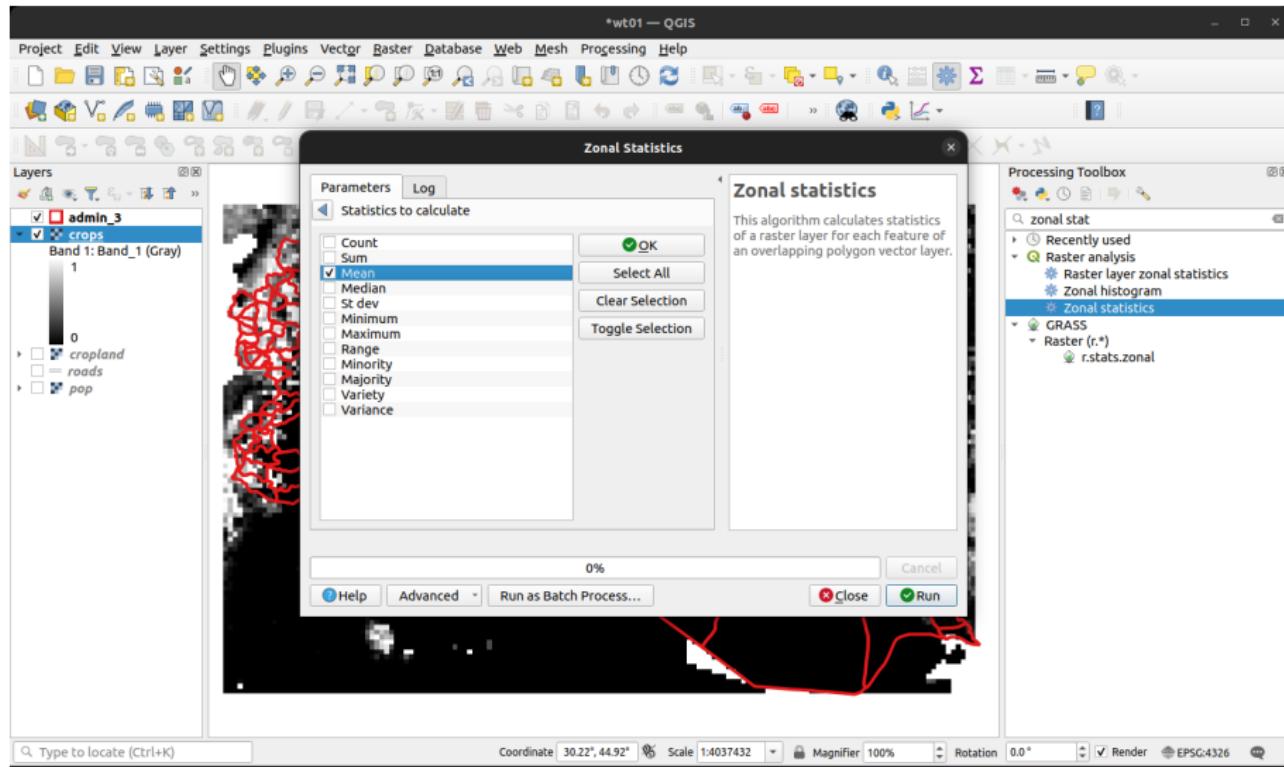
Make sure everything looks right. Remove the original, global cropland layer



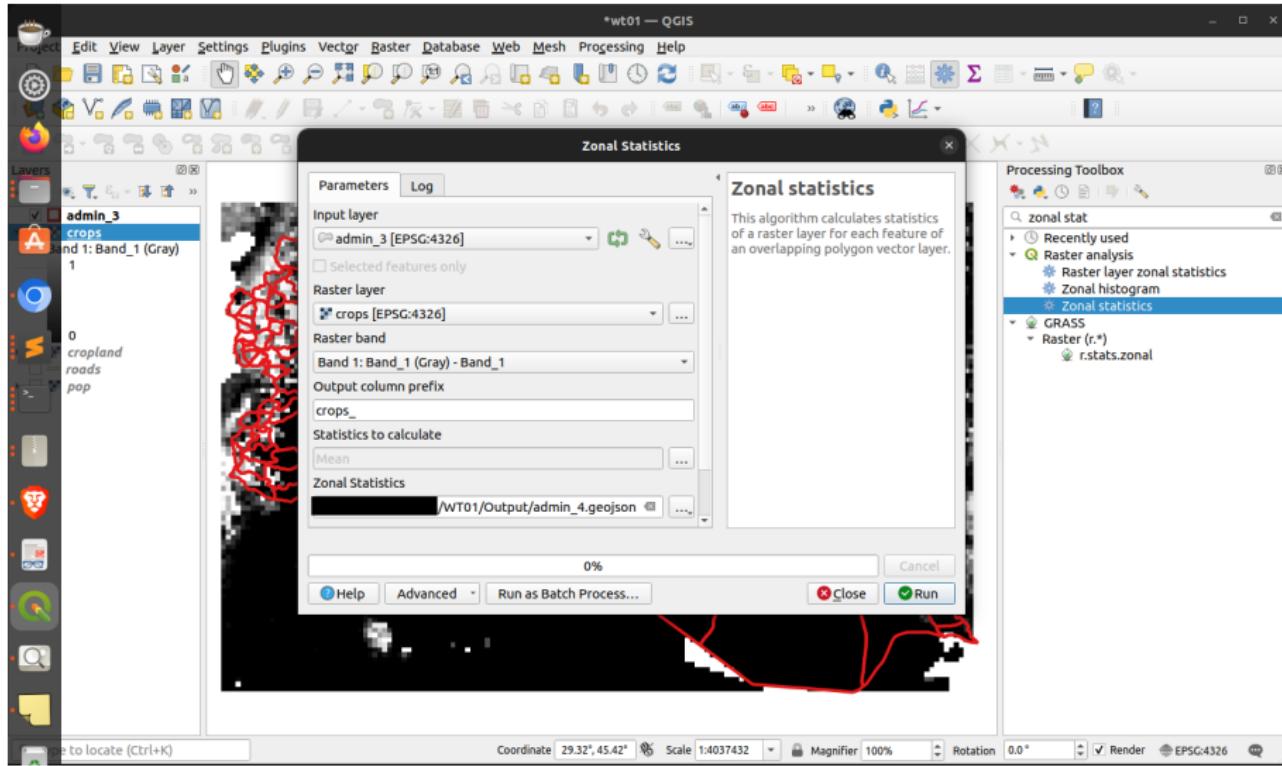
Let's **calculate crop cultivation in each district**. Open the Zonal statistics tool in Processing Toolbox. Set Input layer=admin_3, Raster layer=crops, and Output column prefix=crops_. Click the [...] button next to Statistics to calculate



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

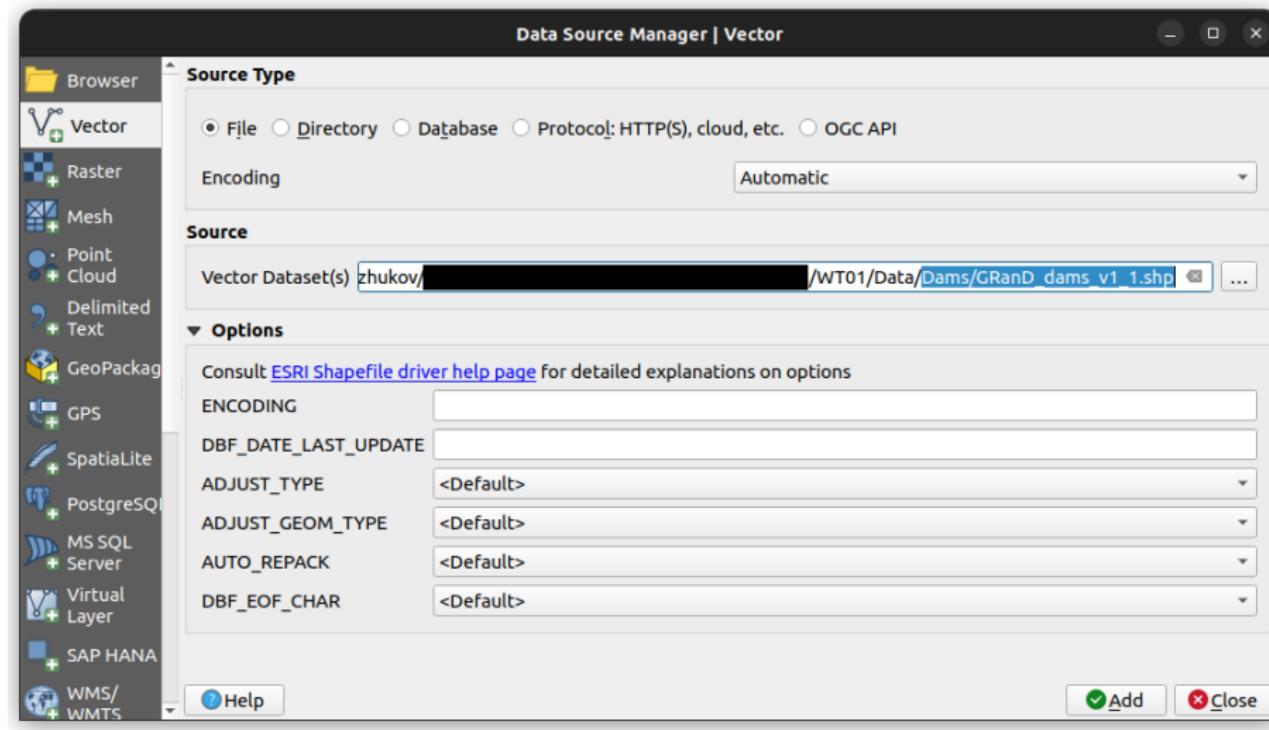


This should have generated a new field, `crops_mean` in the new `admin_4` layer (check the layer's attribute table)

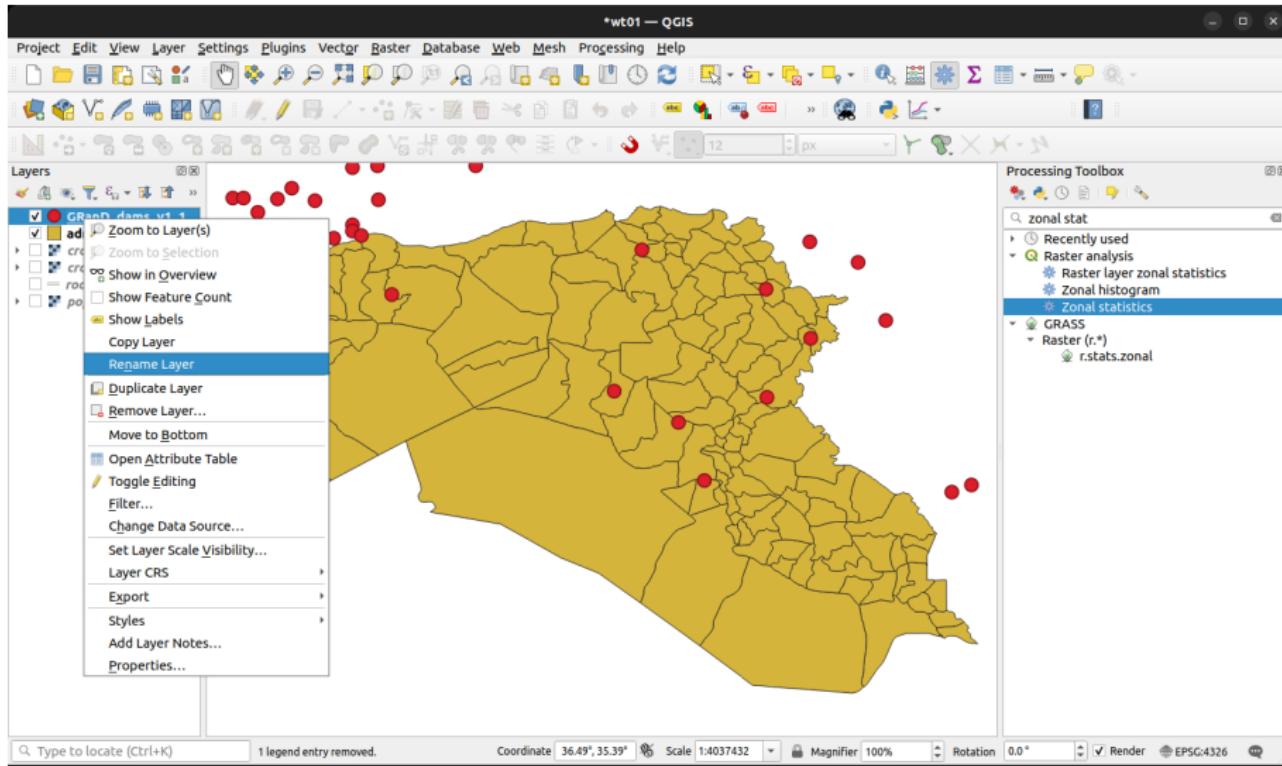
The screenshot shows a QGIS interface with the following components:

- Project Bar:** Standard QGIS menu items: Project, Edit, View, Layer, Settings, Plugins, Vector, Raster, Database, Web, Mesh, Processing, Help.
- Toolbar:** Various tools for selection, measurement, and editing.
- Layers Panel:** Shows the layers: `admin_4` (checked), `admin_3`, and `crops`. The `crops` layer is selected, showing a grayscale raster where yellow represents land cover.
- Processing Toolbox:** Contains recent items, Raster analysis (selected), Zonal statistics, and GRASS.
- Map View:** Displays a map of Iraq with yellow-shaded administrative boundaries. A status bar at the bottom indicates coordinates (35.94°, 35.46°) and scale (1:4037432).
- Attribute Table:** Shows the data for the `admin_4` layer. The columns include: NAME_2, VARNAME_2, NL_NAME_2, TYPE_2, ENGTTYPE_2, CC_2, HASC_2, roads_m, roads_ct, roads_km, area_km2, road_density, pop_sum, and crops_mean. The last column, `crops_mean`, contains values like 0, 614252.99..., and Real NULL.
- Bottom Bar:** Includes a search bar ("Type to locate (Ctrl+K)"), coordinate and scale controls, and a render button.

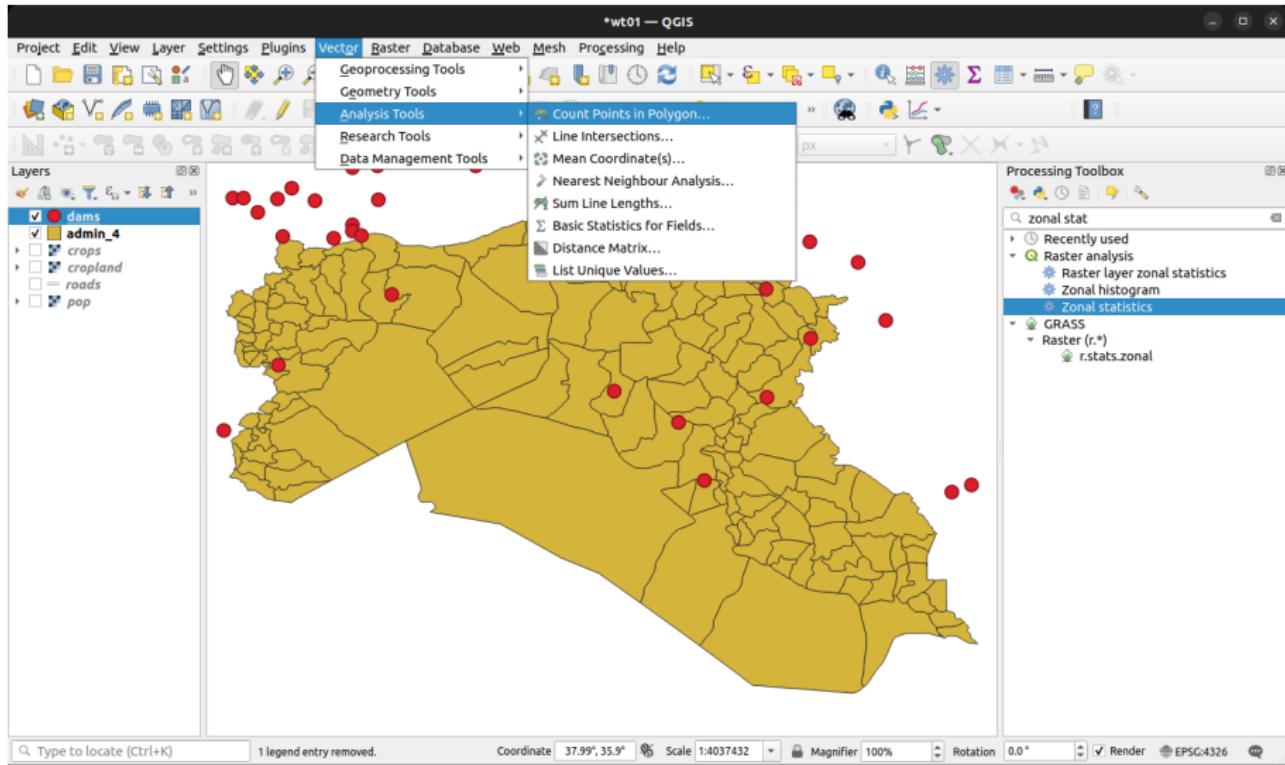
Load *hydroelectric dam* locations (Layer → Add Layer → Add Vector Layer).
Open the GRanD_dams_v1_1.shp file in Data/Dams folder.



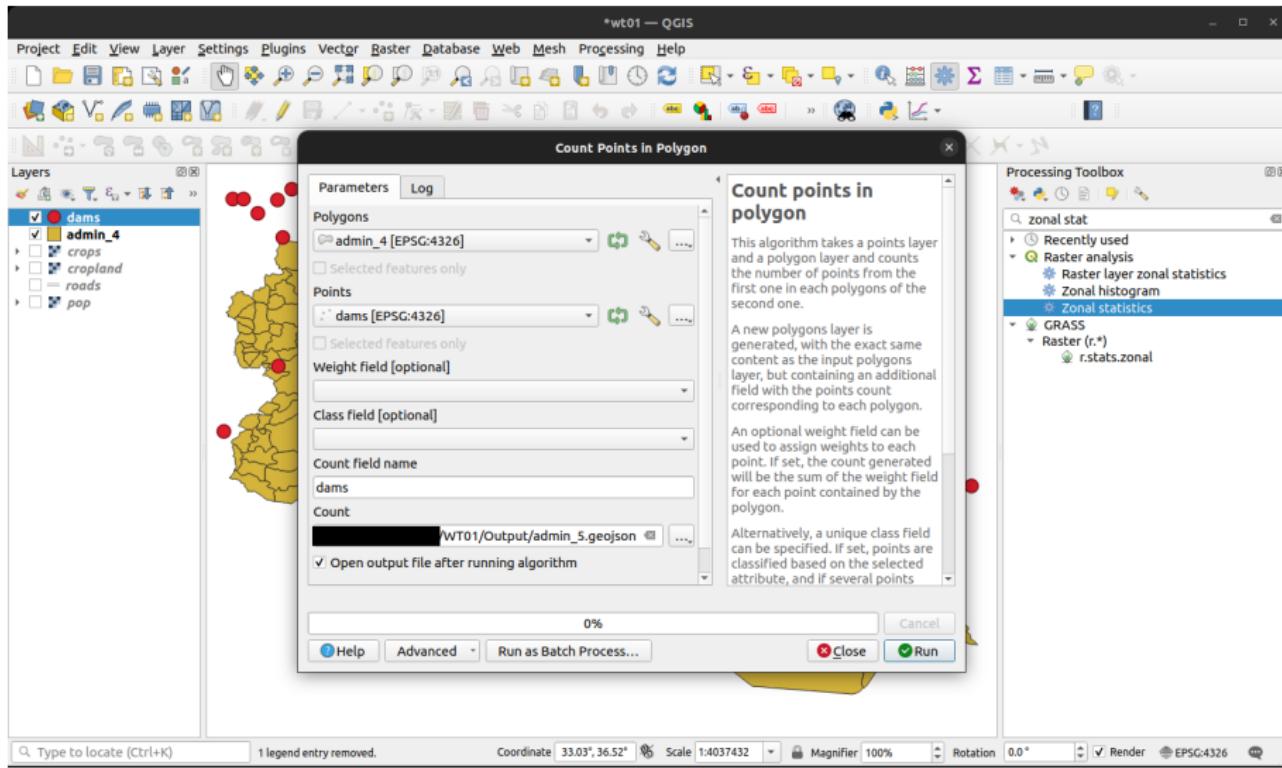
The dam locations should appear on the project window. Rename it from GRanD_dams_v1_1 to dams



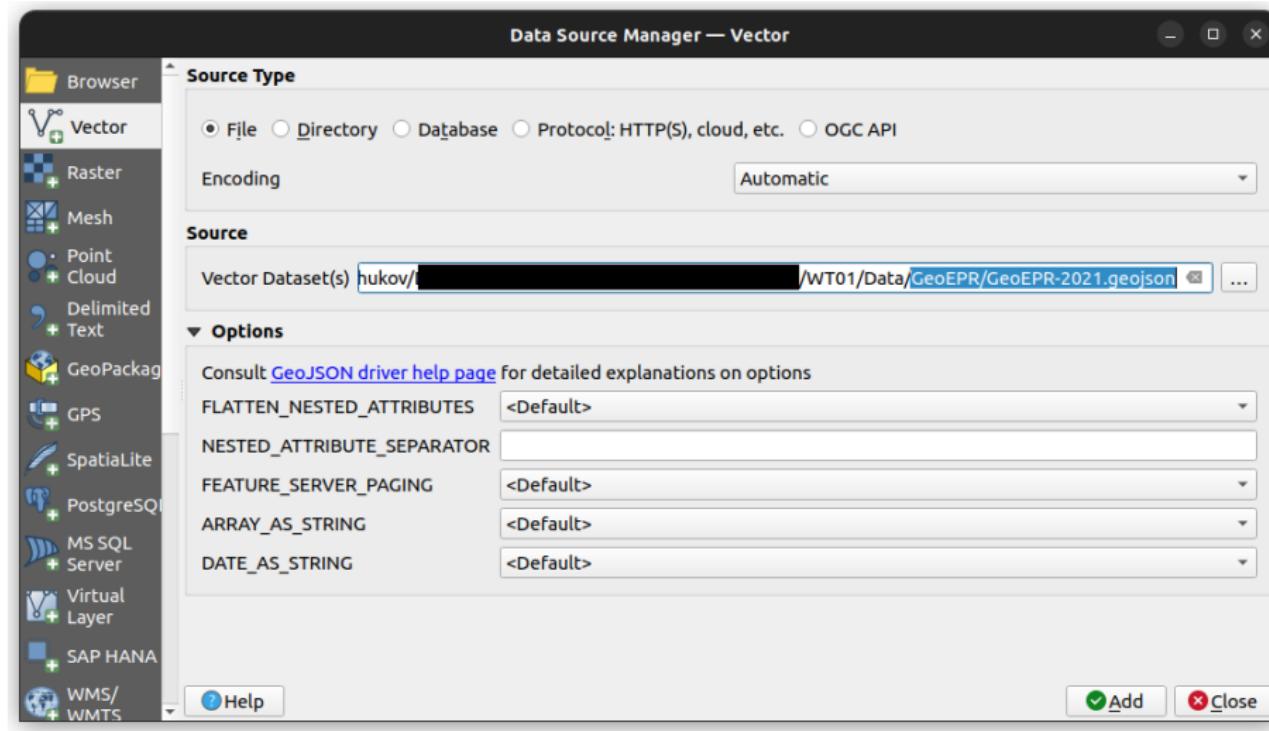
Let's calculate the **number of dams per district**. Open the Count Points in Polygon tool (Vector → Analysis Tools → Count Points in Polygon)



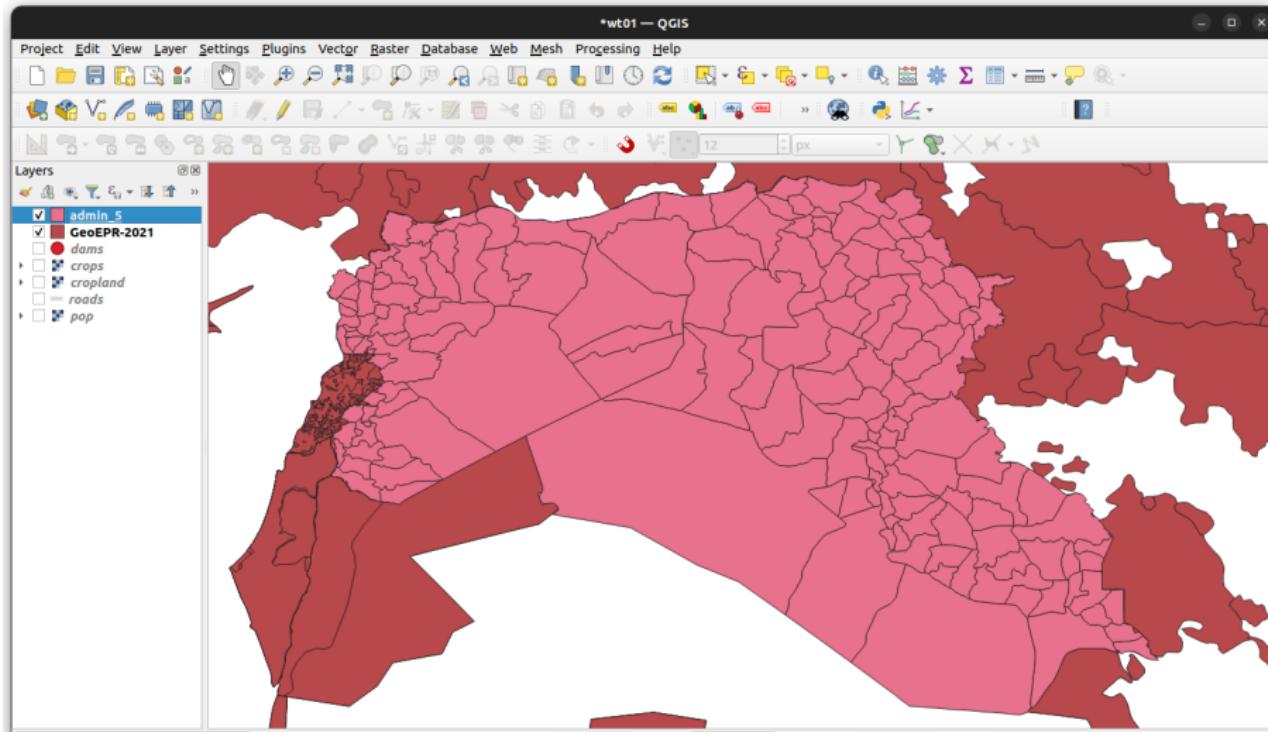
Select Polygons = admin_4, Points = dams. Name the count field dams, and save the output file as admin_5.geojson. Click Run



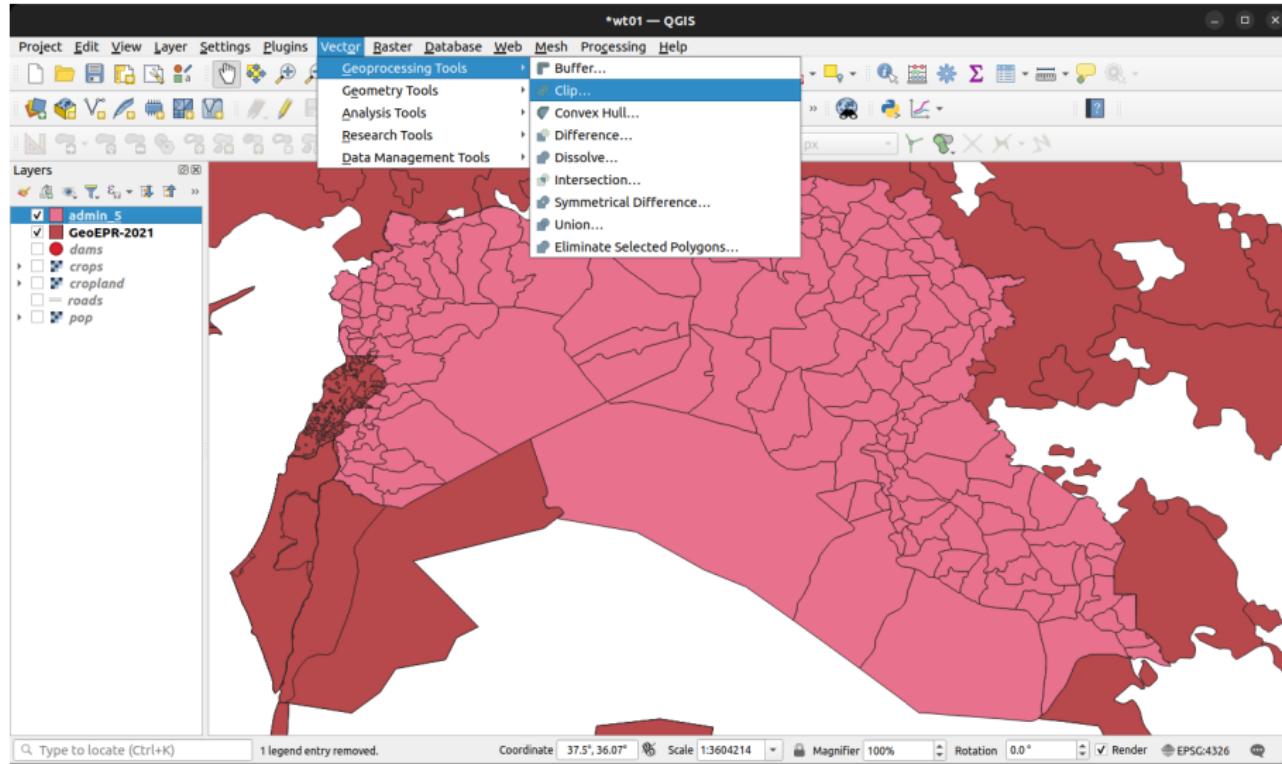
Add the *sectarian divisions* data to the project, using Add Vector Layer....\\ Load the GeoEPR-2021.geojson file in Data/GeoEPR folder



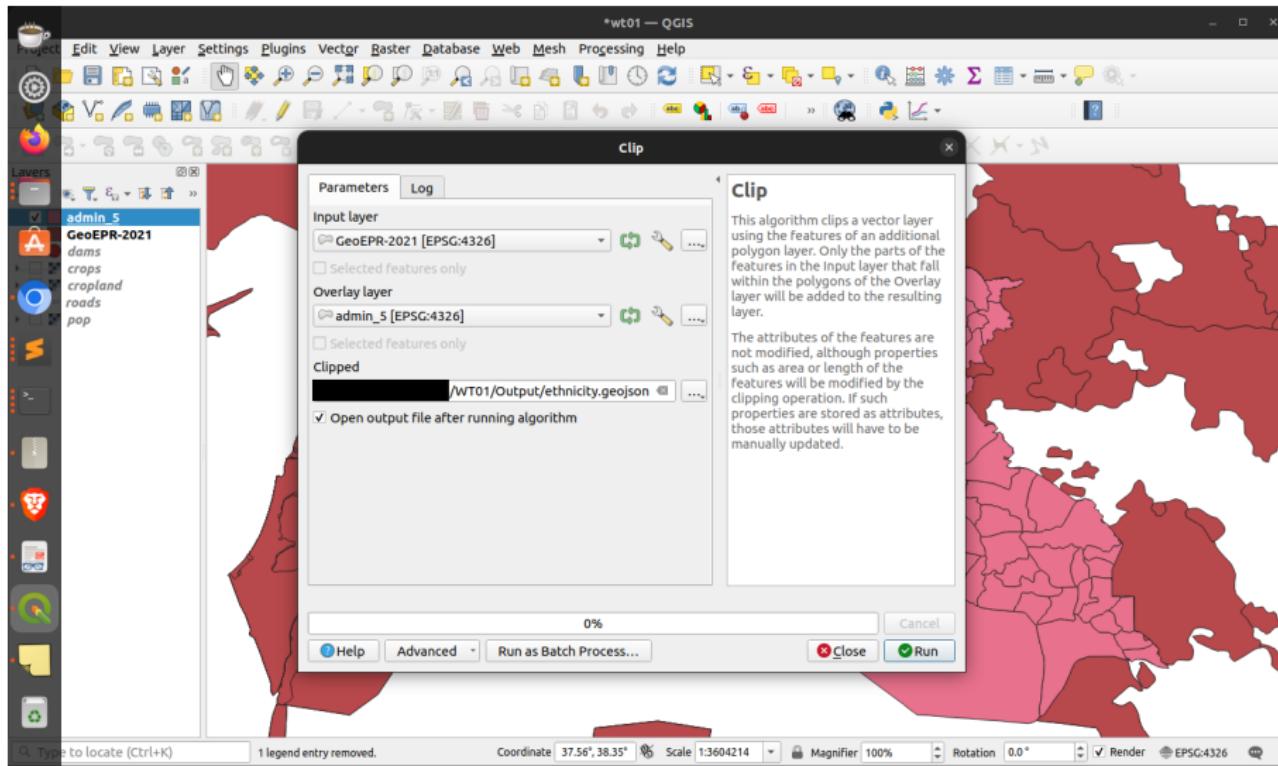
The (global) ethnicity layer should appear in the project window. Let's **calculate the proportion of each district populated by Sunni Arabs**. Let's first extract the subset that overlaps with our study area.



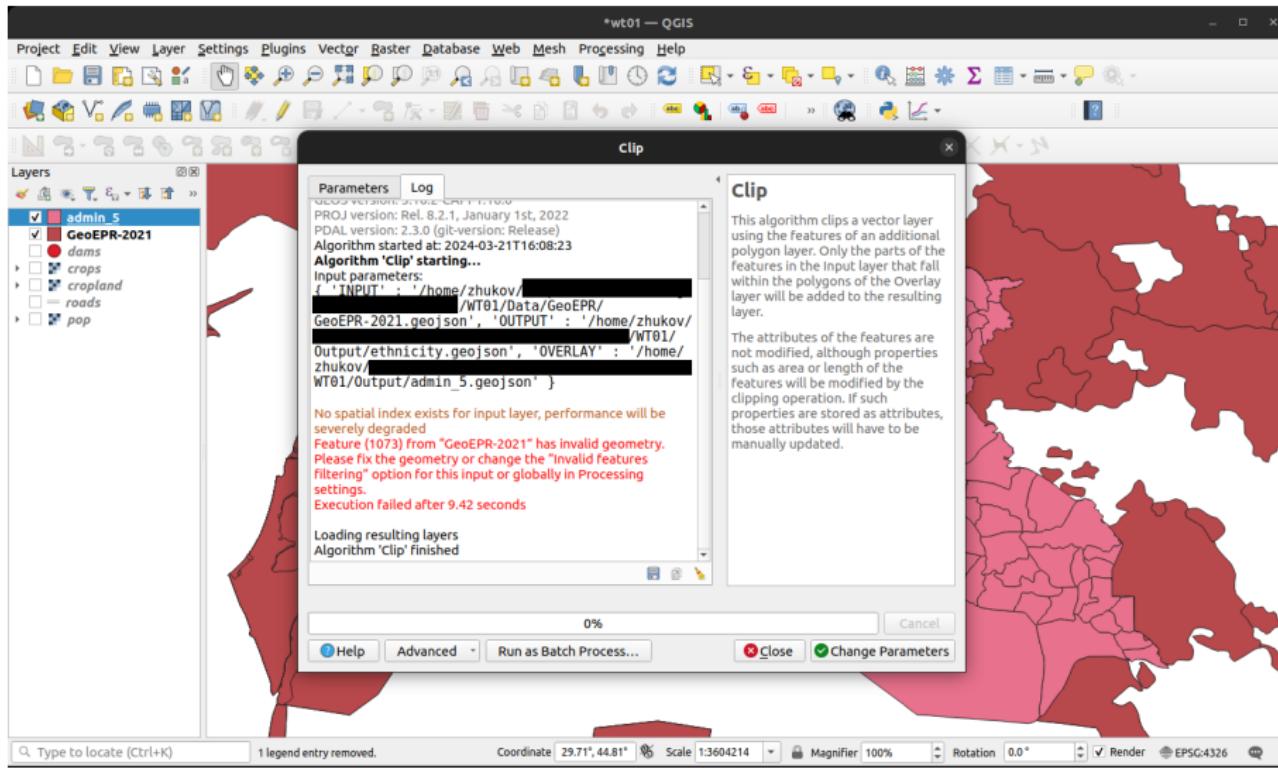
Open the Clip (vector) tool (Vector → Geoprocessing Tools → Clip)



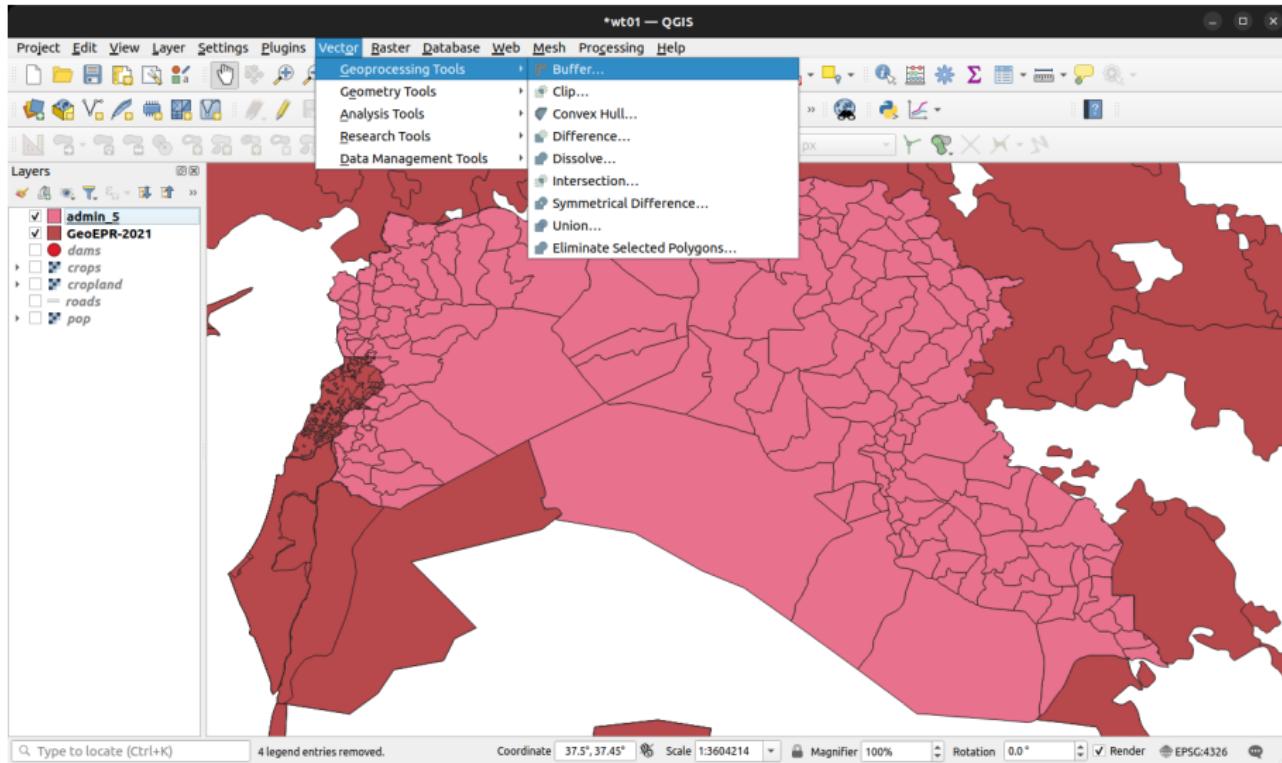
Set Input layer = GeoEPR, Overlay layer = admin_5. Save the file as ethnicity.geojson. Click Run



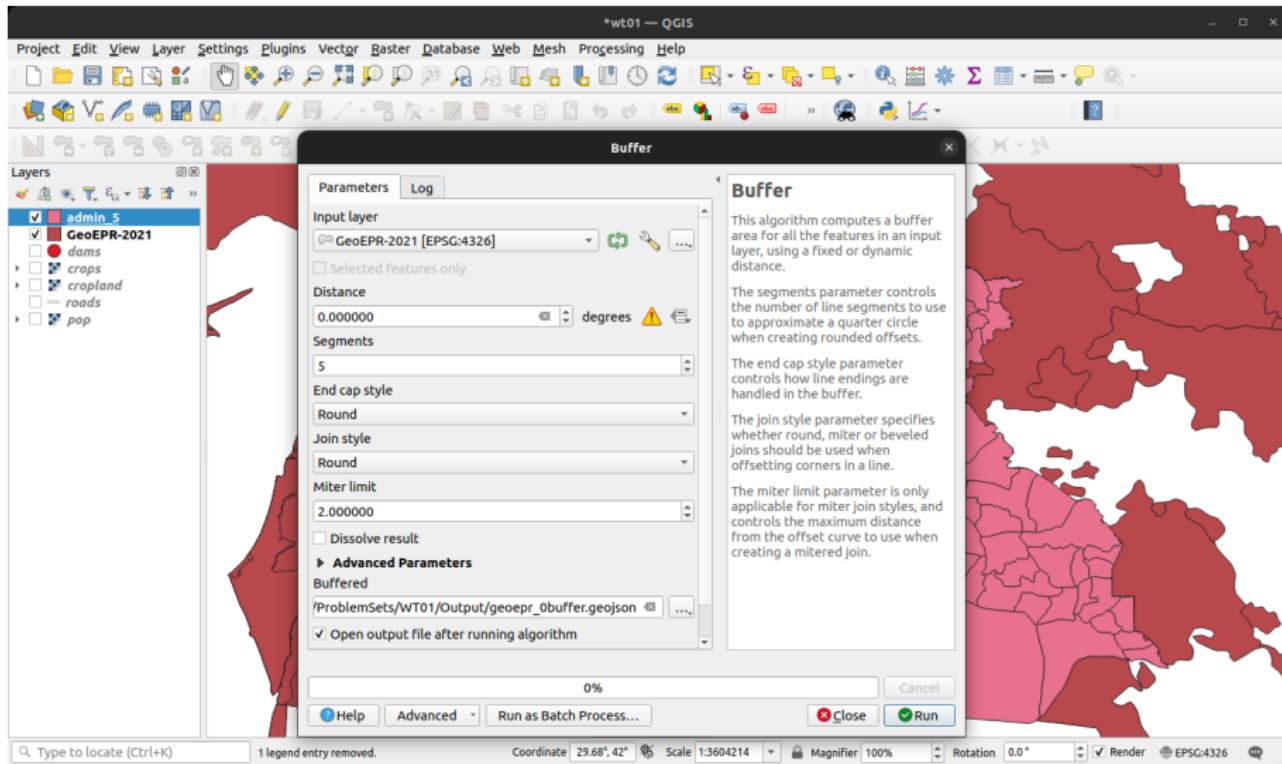
Uh oh! **Execution failed!** The error message says that GeoEPR-2021 contains "invalid geometries". How do we fix this?



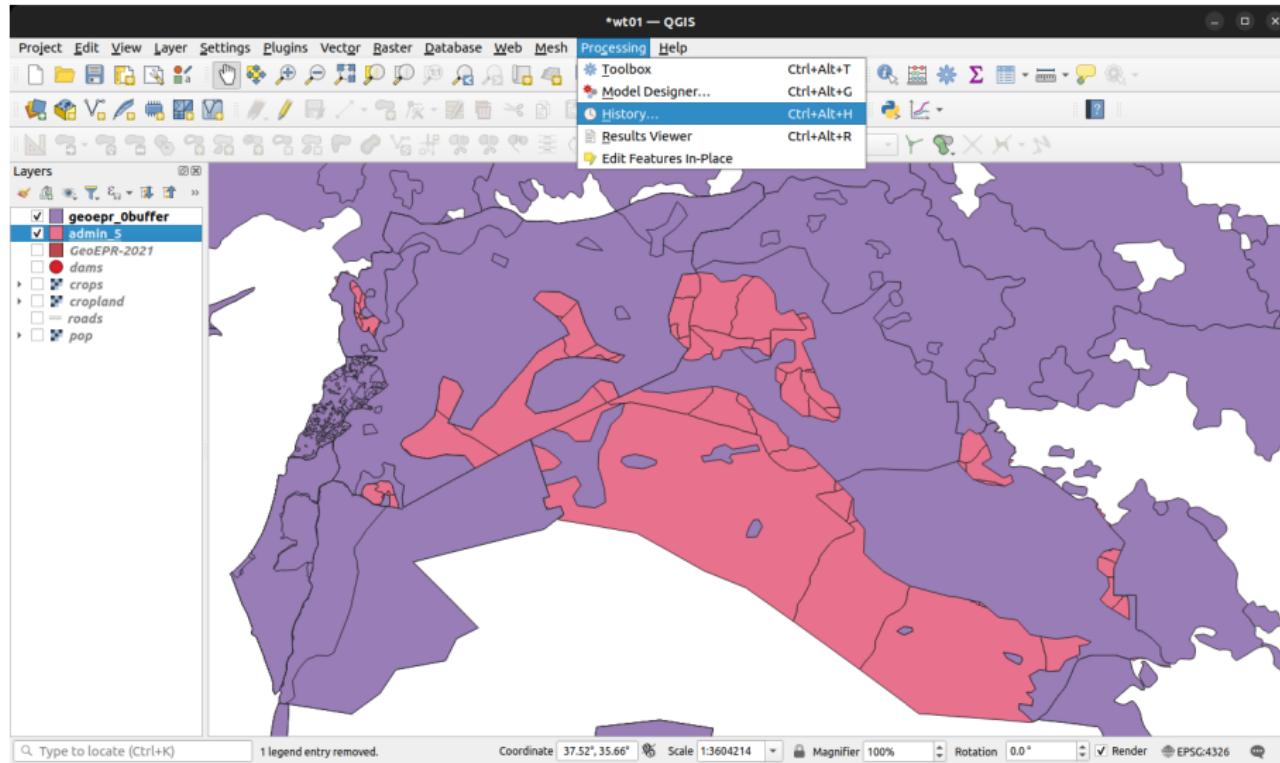
Many “invalid geometry” problems can be solved with the “zero buffer trick”. Go to Vector → Geoprocessing Tools → Buffer...



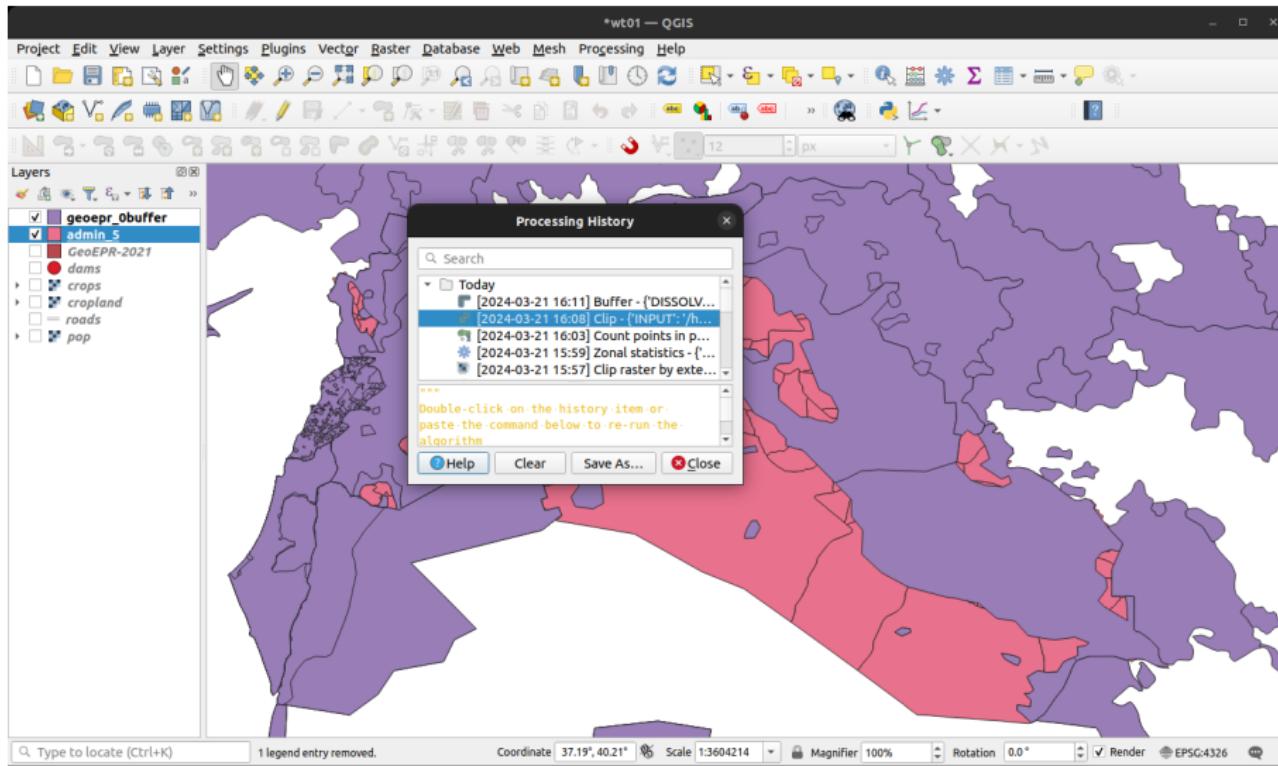
Set Input layer = GeoEPR-2021 and Distance = 0. Keep the defaults for the other parameters and save the file as geoepr_0buffer.geojson. Click Run



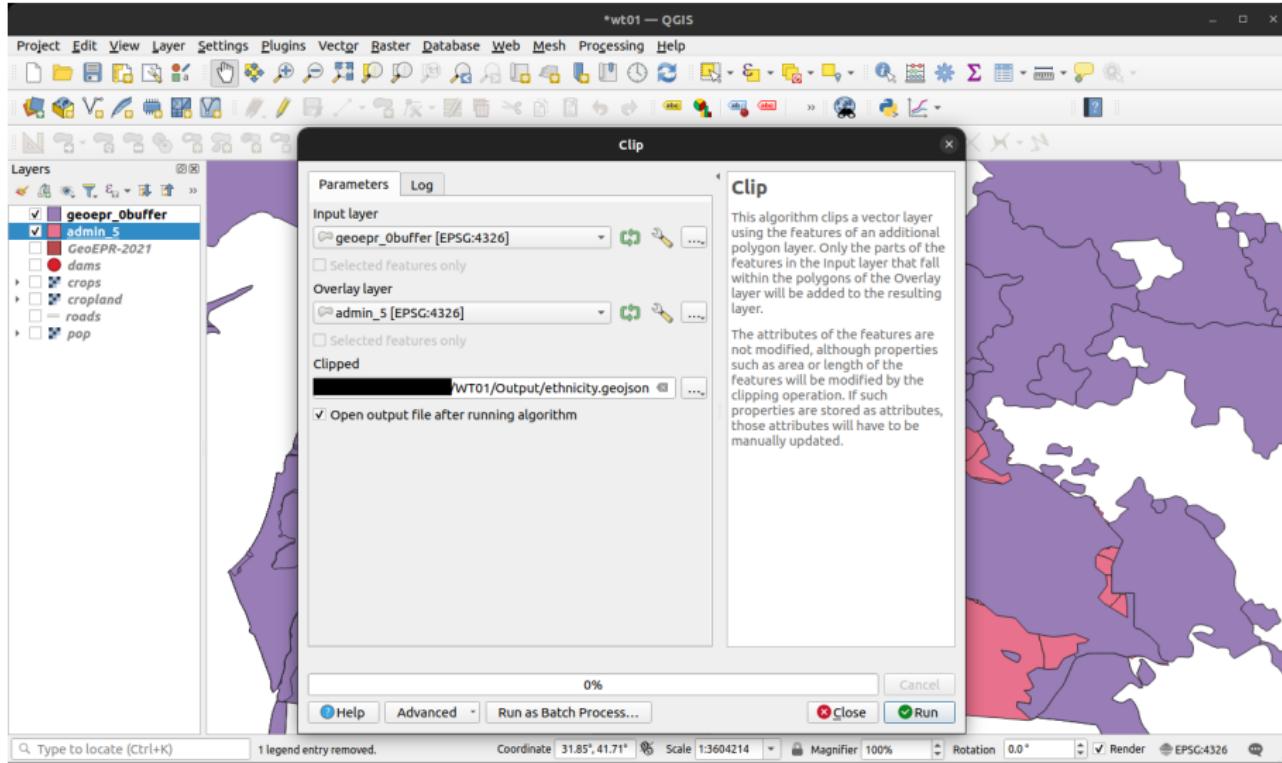
Now let's try the clipping operation again. To save time, go to Processing → History...



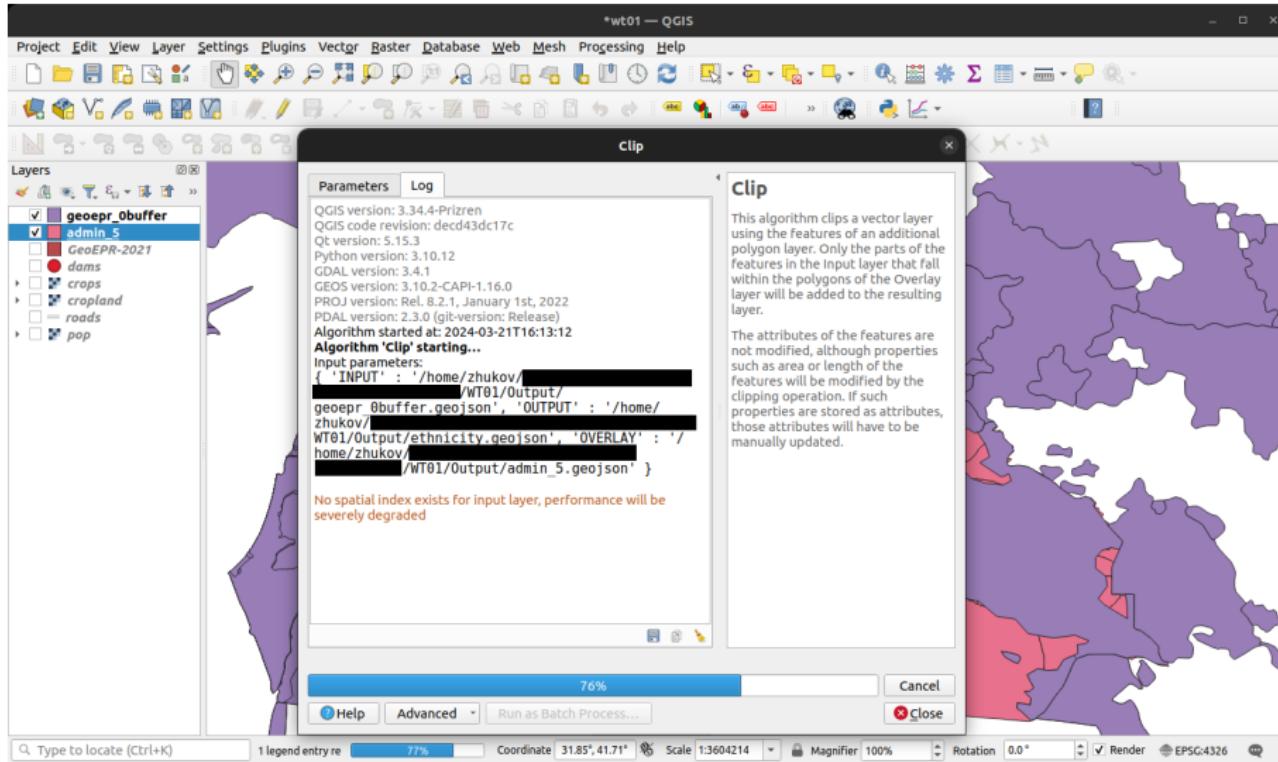
In the Processing History window, double-click on Clip (it should be the second-most recent item)



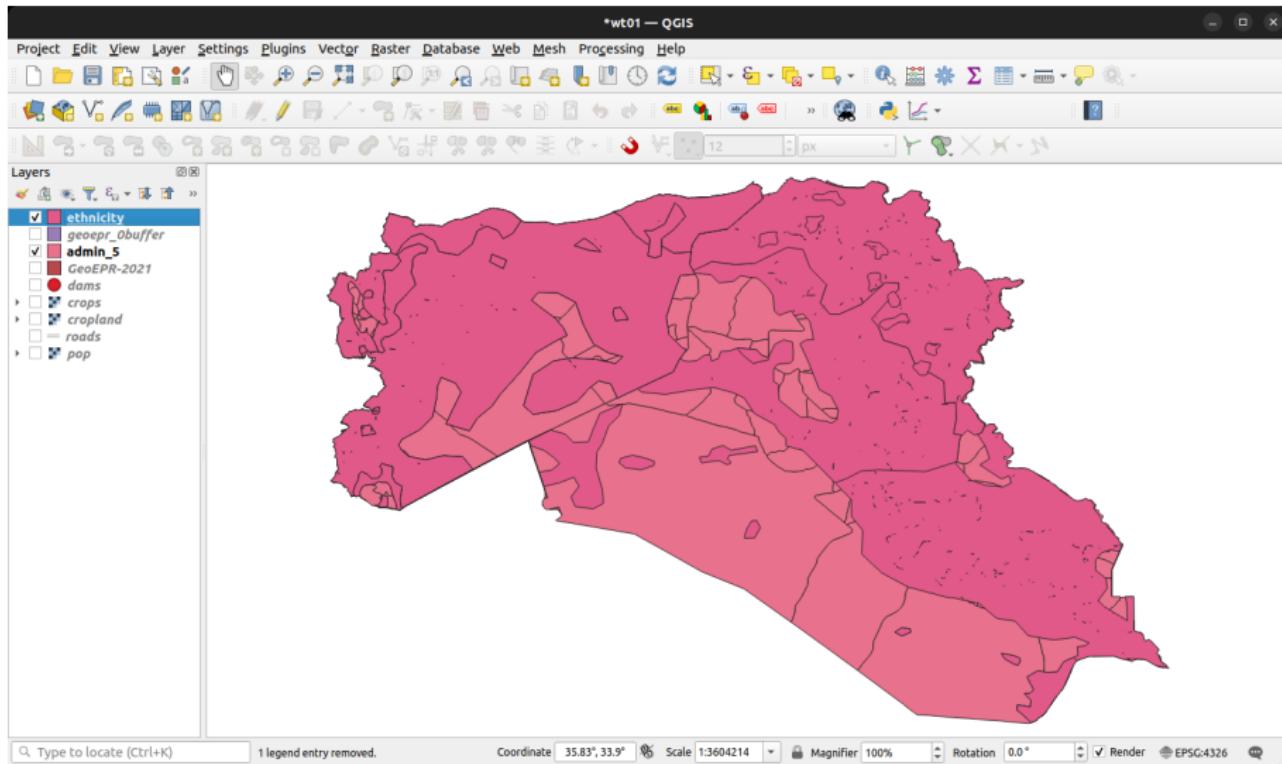
Set the Input layer = geoepr_0buffer. Click Run



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

The screenshot shows the QGIS interface with a map view and a detailed attribute table below it.

Map View: The map displays several polygon layers. The most prominent layer is labeled "ethnicity", which is highlighted in the layers panel. Other visible layers include "admin_5", "dams", "crops", "cropland", "roads", and "pop". The "ethnicity" layer consists of numerous small, irregular pink polygons representing different ethnic groups across a geographic area.

Attribute Table: The table shows the following columns and data for the selected "ethnicity" layer:

gwid	statename	from	to	group	groupid	gwgroupid	umbrella	sqkm	type
1	645	Iraq	1946	2021	Shi'a String	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 & ...

A green box highlights the first row (gwid 1) in the table, and the status bar at the bottom of the table window indicates "Show All Features".

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

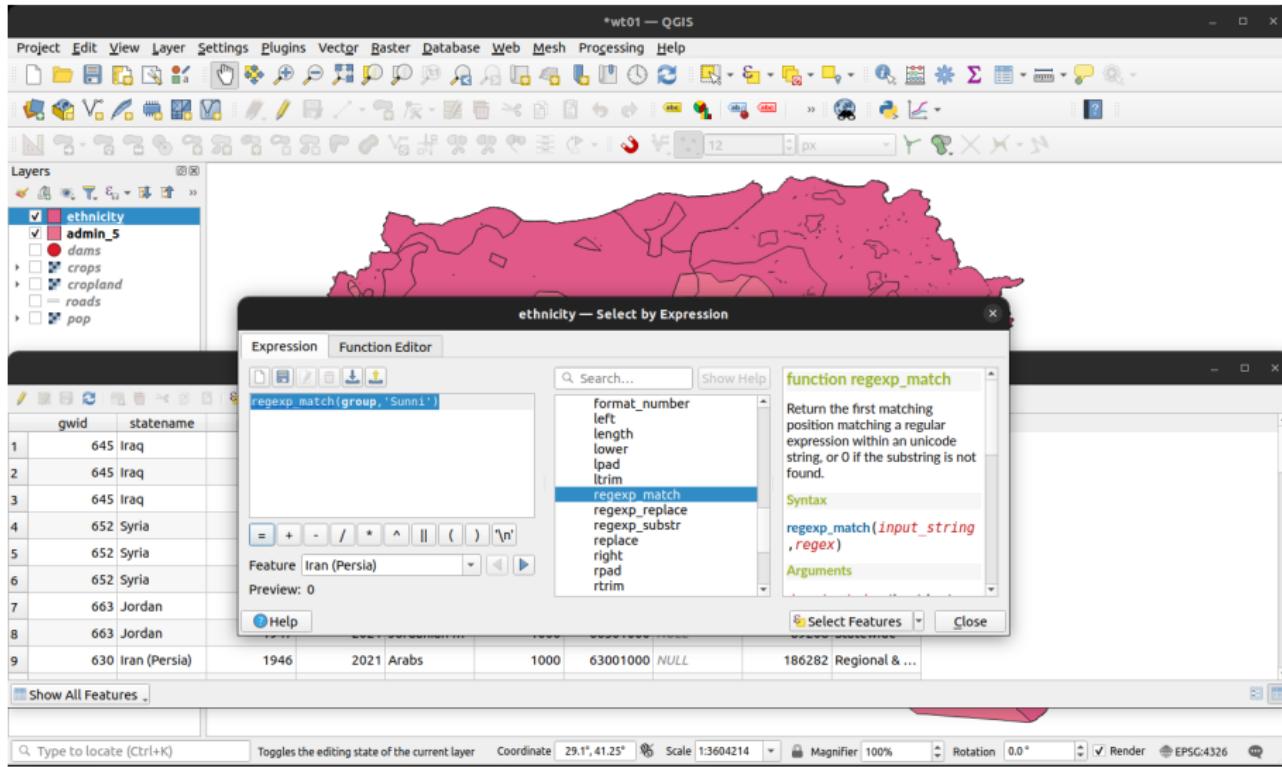
The screenshot shows the QGIS application window titled "wt01 — QGIS". The interface includes a top menu bar with Project, Edit, View, Layer, Settings, Plugins, Vector, Raster, Database, Web, Mesh, Processing, and Help. Below the menu is a toolbar with various icons for editing and analysis. On the left is a "Layers" panel listing several vector layers: "ethnicity" (selected), "admin_5", "dams", "crops", "cropland", "roads", and "pop". The main canvas displays a map of Iraq with administrative boundaries colored in shades of pink and purple. A status bar at the bottom of the map area indicates "ethnicity — Features Total: 46, Filtered: 46, Selected: 0".

Below the map is a "Select features using an expression" dialog box. The table below lists 9 selected features:

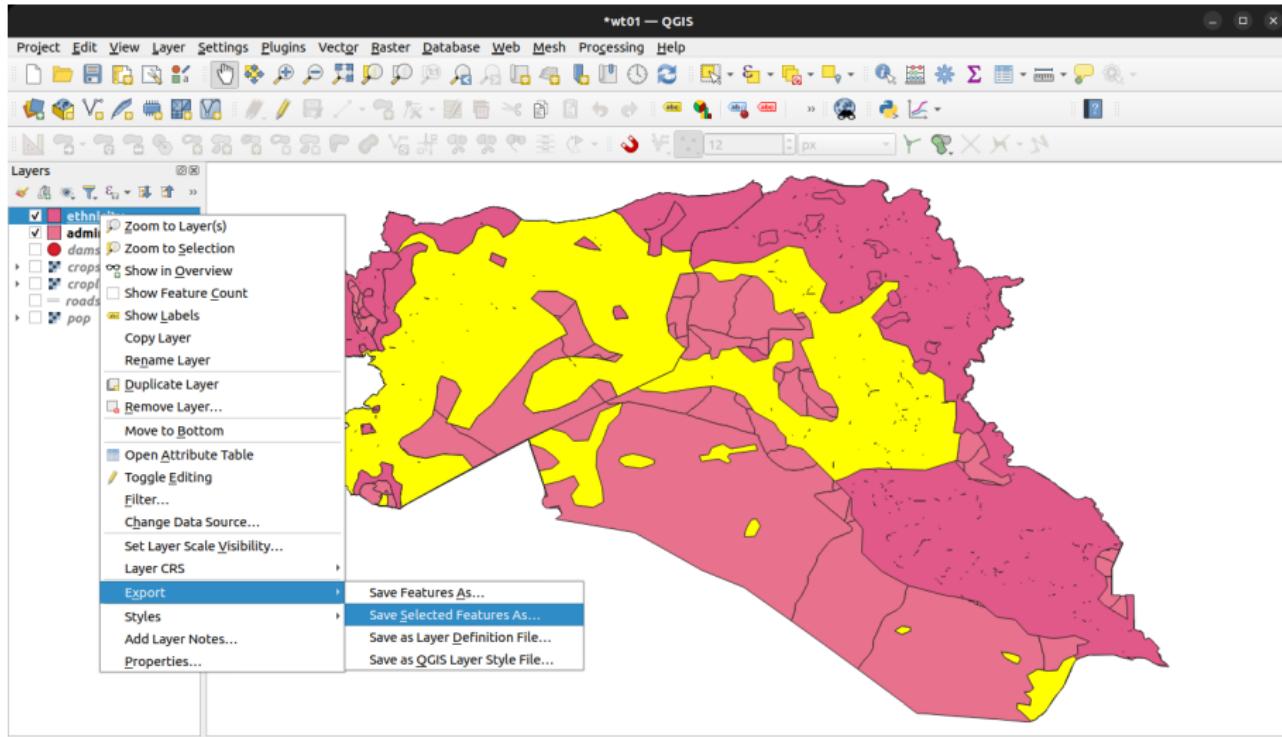
gwid	statename	groupid	gwgroupid	umbrella	sqkm	type	
1	645 Iraq	1946	2021 Shi'a Arabs	2000	64502000	NULL	108150 Regional & ...
2	645 Iraq	1946	2021 Sunni Arabs	3000	64503000	NULL	96757 Regional & ...
3	645 Iraq	1946	2021 Kurds	1000	64501000	NULL	72420 Regional & ...
4	652 Syria	1946	1966 Sunni Arabs	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 & ...

At the bottom of the table, there is a "Show All Features" button. The status bar at the very bottom of the screen shows "Type to locate (Ctrl+K)" and other coordinate and scale information.

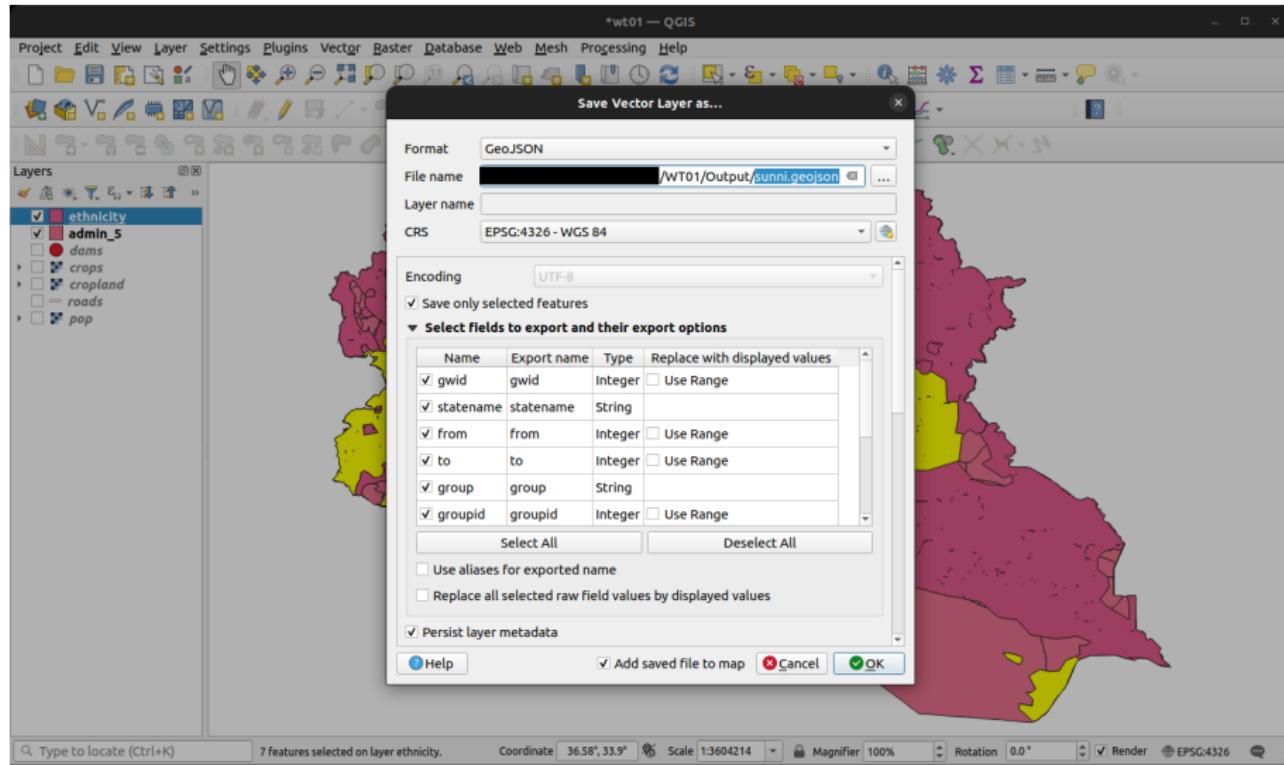
Let's use **regular expressions** to extract the polygons we need. Set Expression = `regexp_match(group, 'Sunni')`. Click Select Features



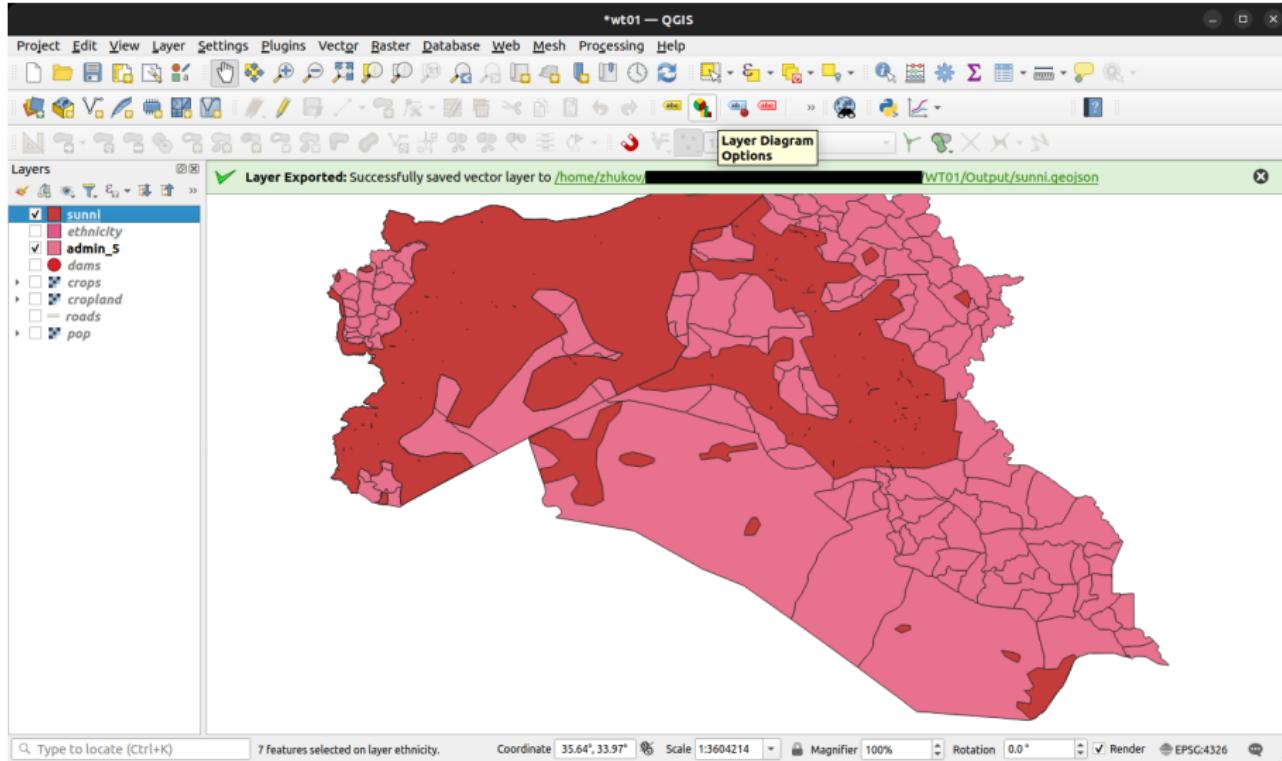
The selected polygons should appear yellow on the map. Let's **extract the selection into a new layer**. Right-click ethnicity in the Layer menu, go to Export→Save Selected Features As



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

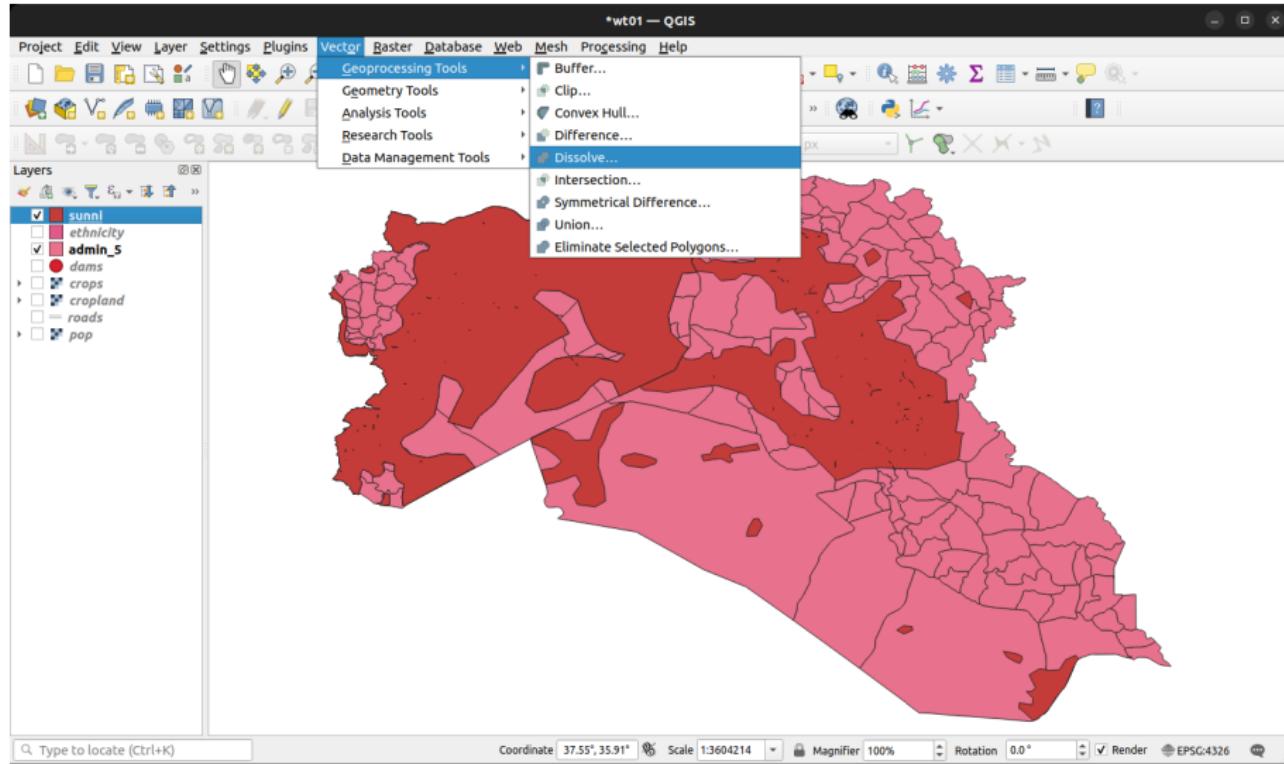
sunni — Features Total: 7, Filtered: 7, Selected: 0

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 & ...

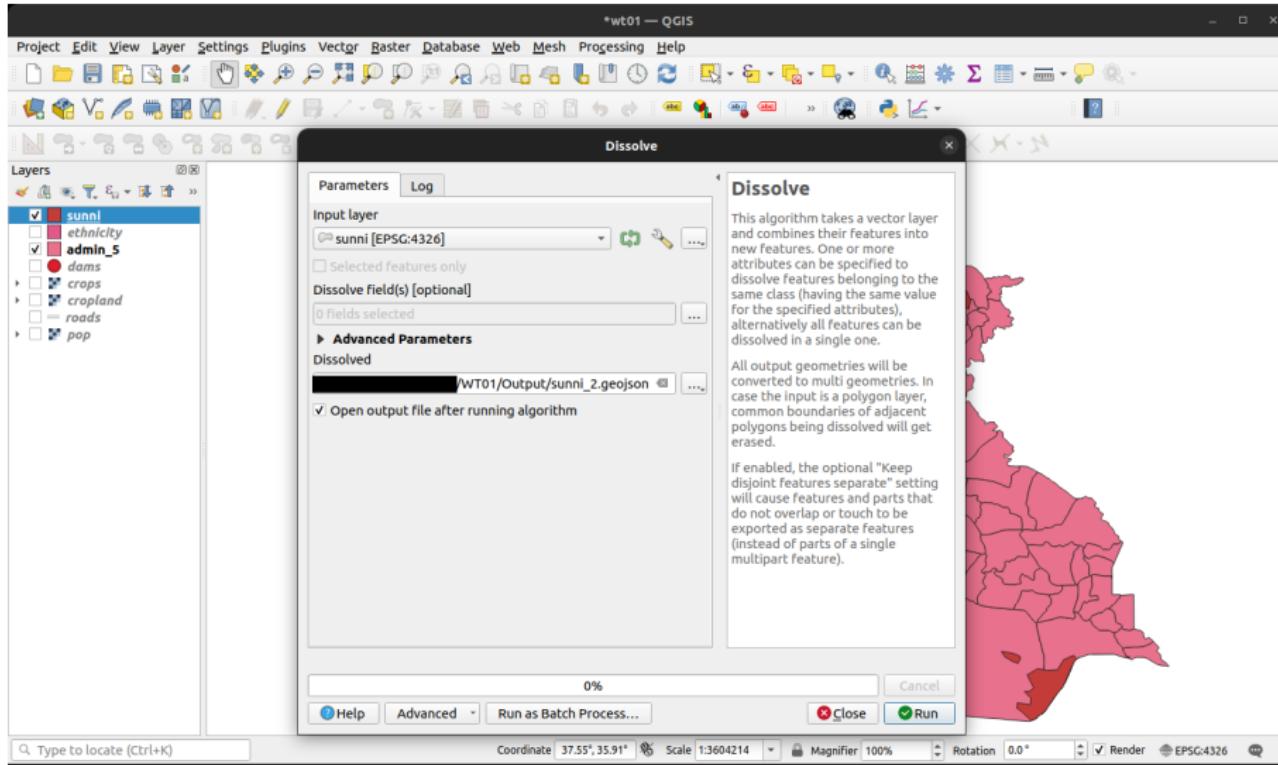
Show All Features

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

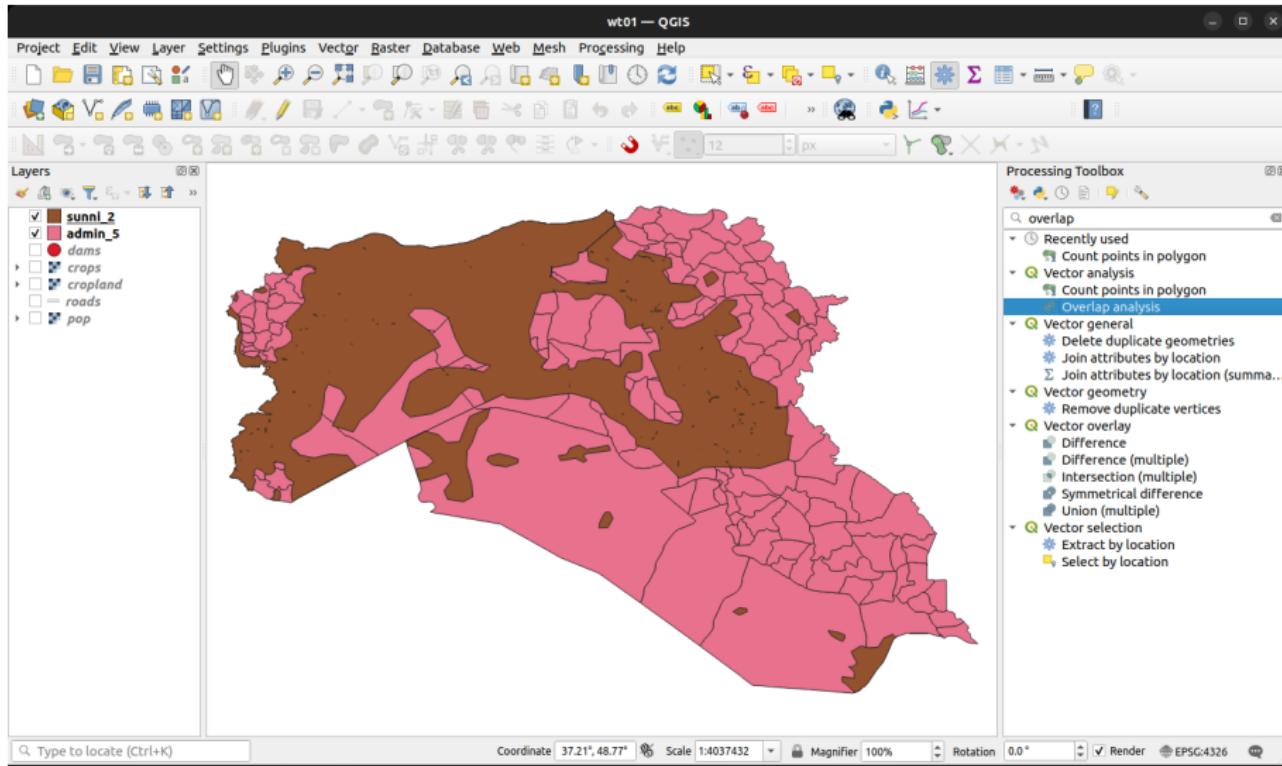
Open the Dissolve tool (Vector → Geoprocessing Tools → Dissolve).



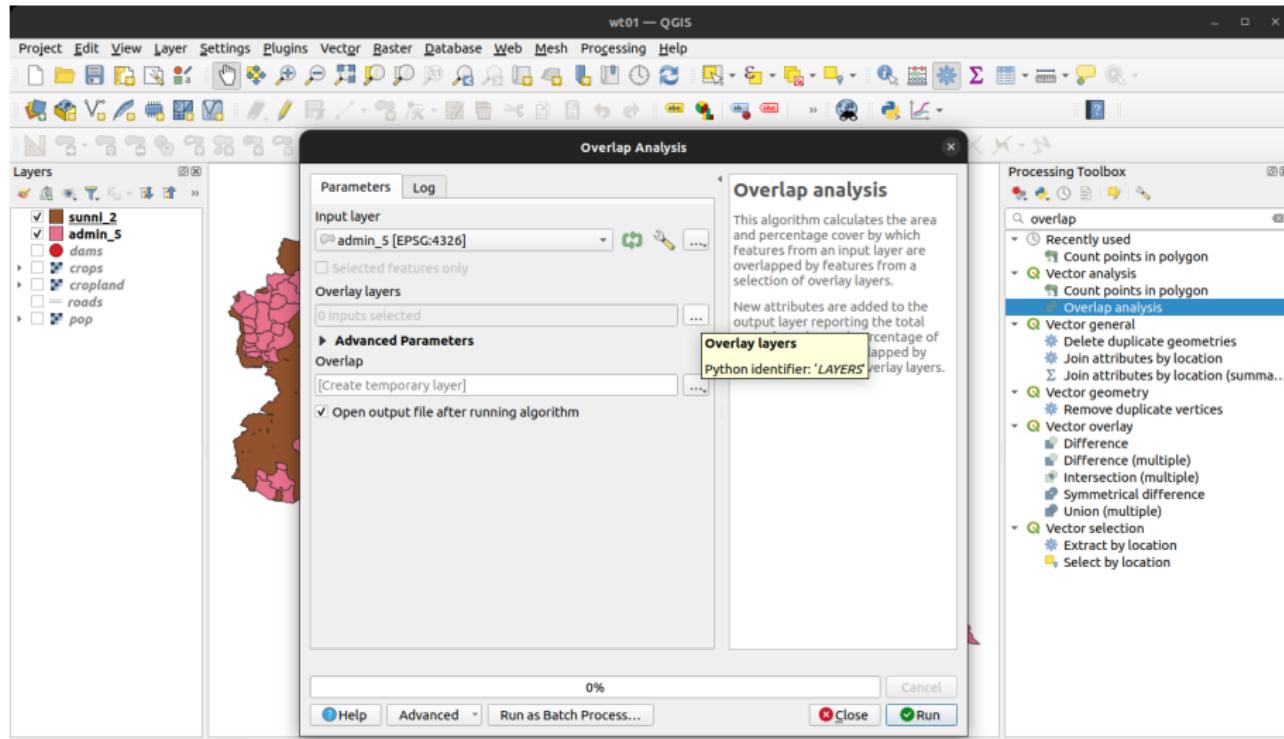
Set Input layer=sunni. Save the output as sunni_2.geojson



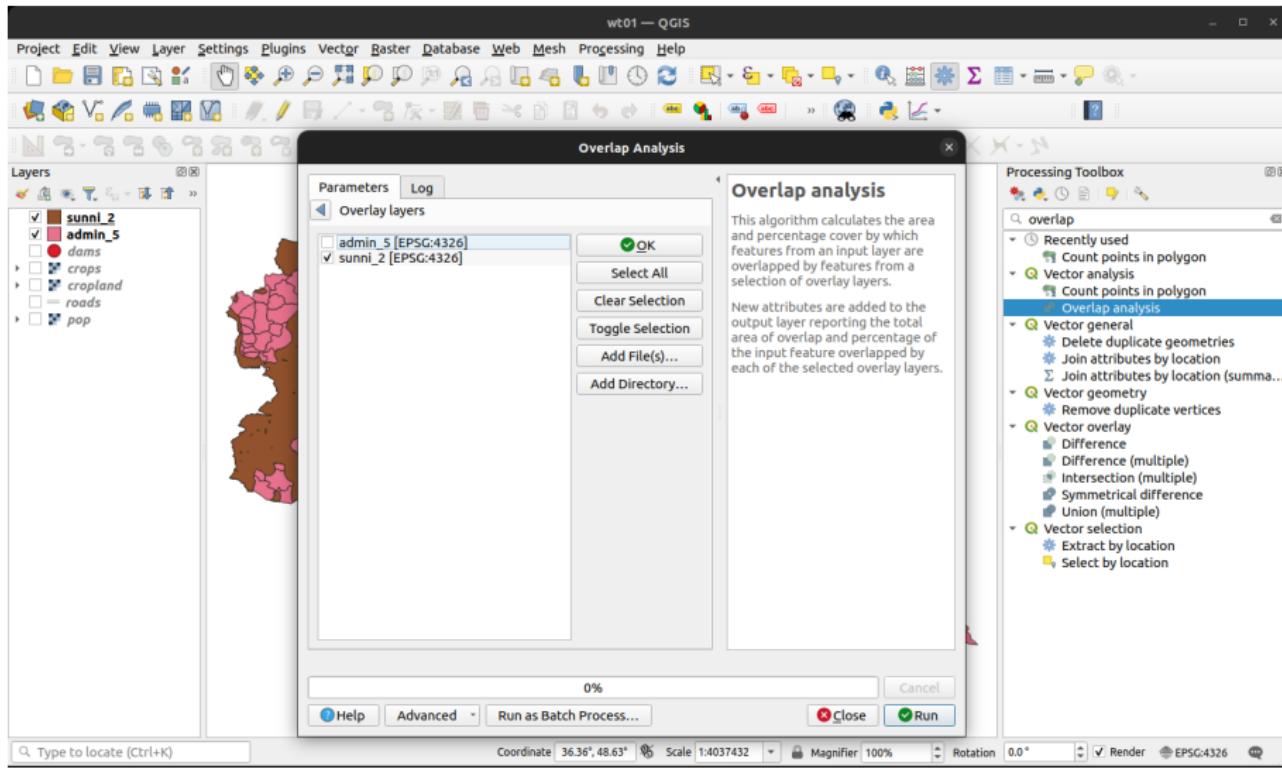
The dissolved polygon should appear in the project window. Now let's **calculate the proportion of each district populated by Sunnis**



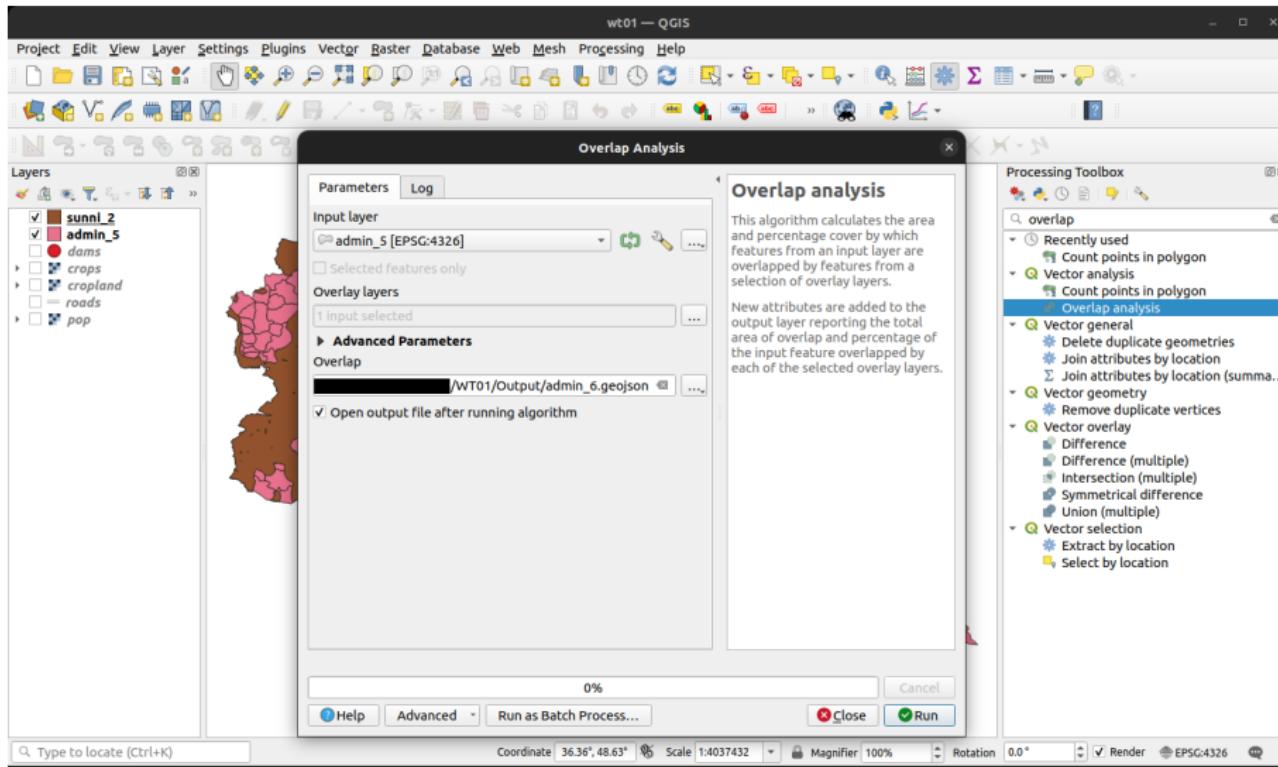
Open the Overlap Analysis tool (in Processing Toolbox → Vector Analysis). Set Input layer = admin_5 and click the [...] button next to Overlay layers



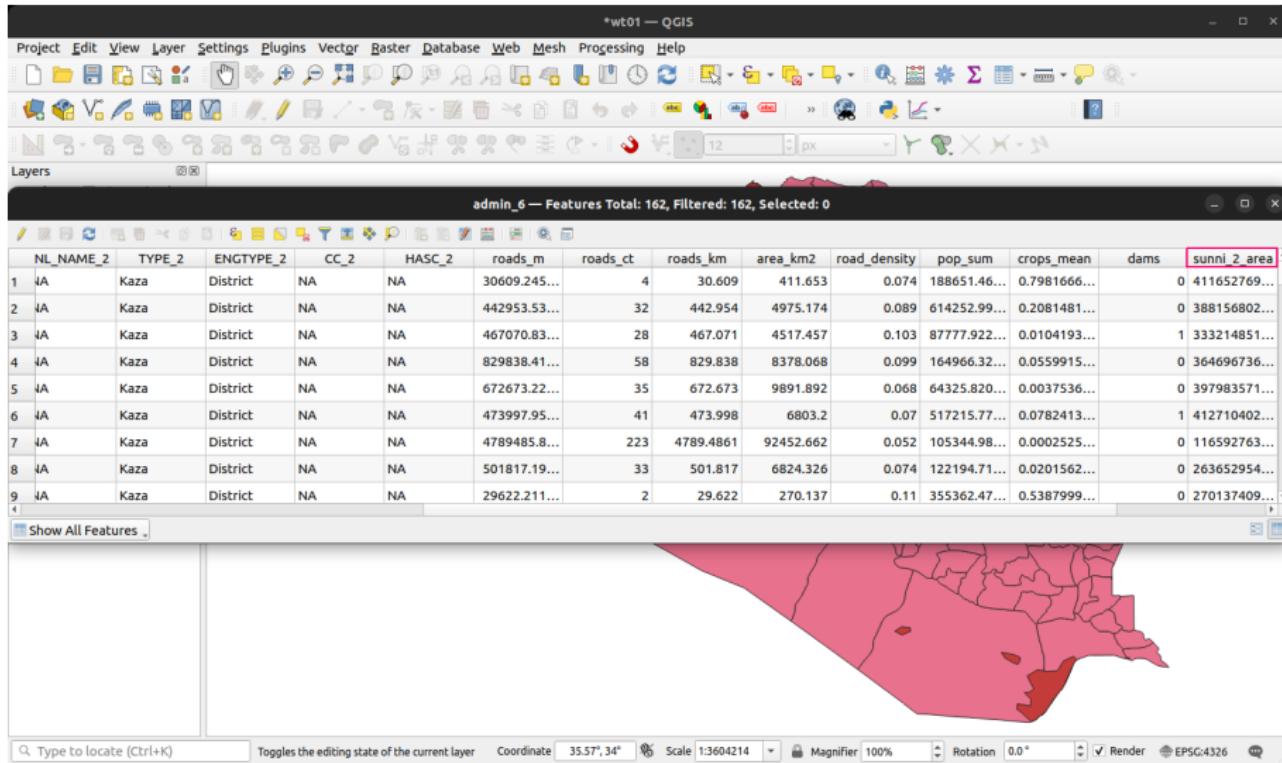
Check the box next to sunni_2. Click OK



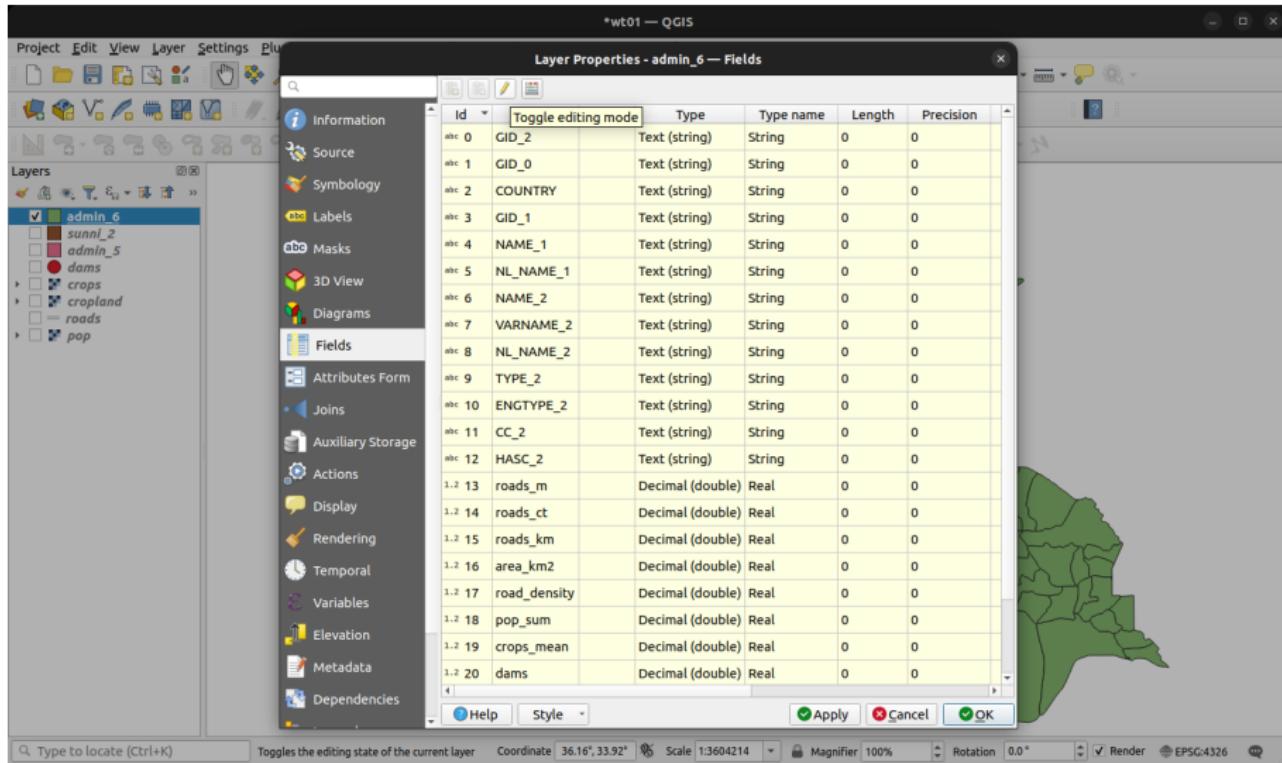
Save the output as admin_6.geojson and click Run



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**



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

The screenshot shows the QGIS application interface with the title bar "wt01 — QGIS". The main window displays a map of a region divided into several green administrative units. On the left, the "Layers" panel shows a list of layers: admin_6 (checked), sunni_2, admin_5, dams, crops, cropland, roads, and pop. The "Fields" tab is selected in the "Layer Properties" dialog box, which is centered over the map. The dialog contains a table with the following data:

ID	Name	Alias	Type	Type name	Length	Precision
2	COUNTRY		Text (string)	String	0	0
3	GID_1		Text (string)	String	0	0
4	NAME_1		Text (string)	String	0	0
5	NL_NAME_1		Text (string)	String	0	0
6	NAME_2		Text (string)	String	0	0
7	VARNAME_2		Text (string)	String	0	0
8	NL_NAME_2		Text (string)	String	0	0
9	TYPE_2		Text (string)	String	0	0
10	ENGTYPE_2		Text (string)	String	0	0
11	CC_2		Text (string)	String	0	0
12	HASC_2		Text (string)	String	0	0
13	roads_m		Decimal (double)	Real	0	0
14	roads_ct		Decimal (double)	Real	0	0
15	roads_km		Decimal (double)	Real	0	0
16	area_km2		Decimal (double)	Real	0	0
17	road_density		Decimal (double)	Real	0	0
18	pop_sum		Decimal (double)	Real	0	0
19	crops_mean		Decimal (double)	Real	0	0
20	dams		Decimal (double)	Real	0	0
21	sunni_2_area		Decimal (double)	Real	0	0
22	sunni_2_pc		Decimal (double)	Real	0	0

The "sunni_2_pc" row is highlighted with a blue selection bar at the bottom of the table. The "OK" button is visible at the bottom right of the dialog.

Rename the field prop_sunni

The screenshot shows the QGIS application interface with the title bar "rwt01 — QGIS". The main window displays a map of a region divided into administrative units. On the left, the "Layers" panel shows several layers: "admin_6" (selected), "sunni_2", "admin_5", "dams", "crops", "cropland", "roads", and "pop". A context menu is open over the "admin_6" layer. In the center, the "Layer Properties - admin_6 - Fields" dialog box is open, listing the following fields:

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

The "prop_sunni" field is currently selected, as indicated by the blue highlight. The dialog box includes buttons for "Help", "Style", "Apply", "Cancel", and "OK".

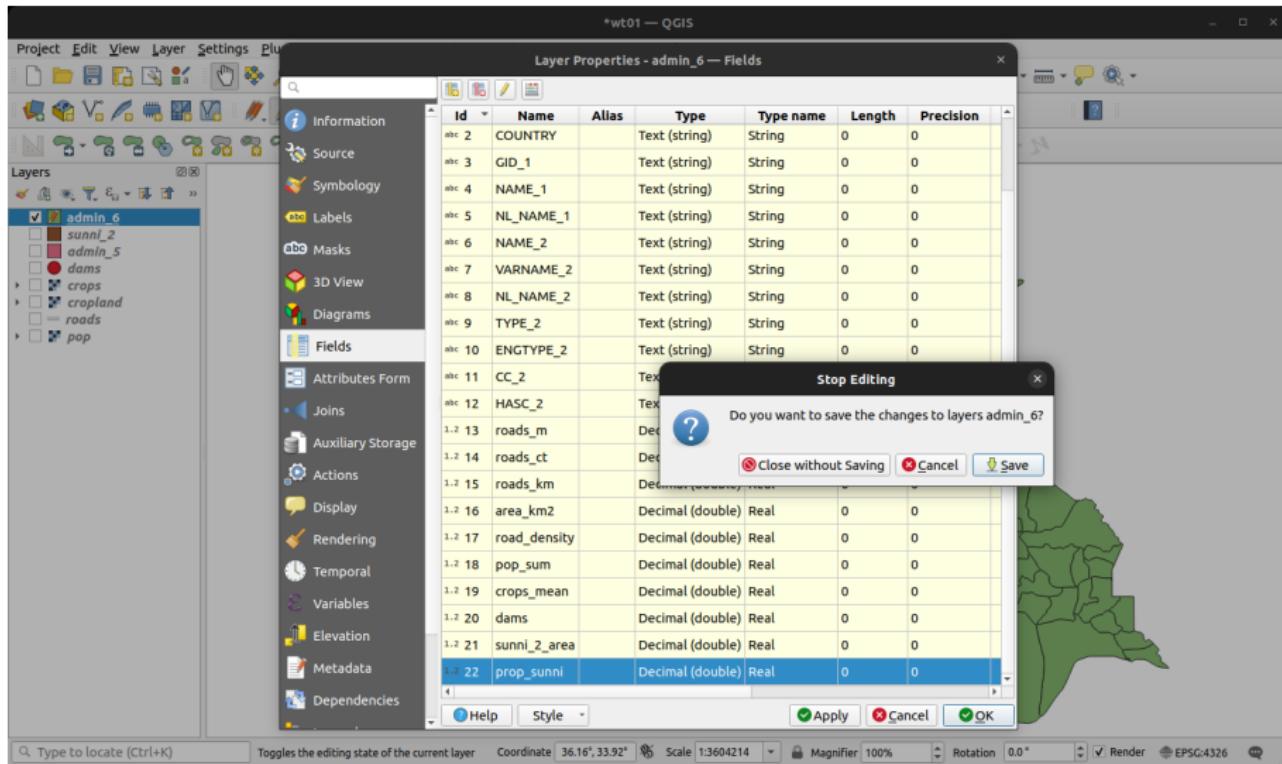
Hit Enter key (or Return key) to commit the name change. Click on the “pencil” again to leave editing mode

The screenshot shows the QGIS interface with the title bar "wt01 — QGIS". On the left, the "Layers" panel lists several layers: sunni_2, admin_5, dams, crops, cropland, roads, and pop. The "admin_6" layer is selected and highlighted with a blue border. In the center, the "Layer Properties - admin_6 - Fields" dialog is open. This dialog displays a table of fields with the following data:

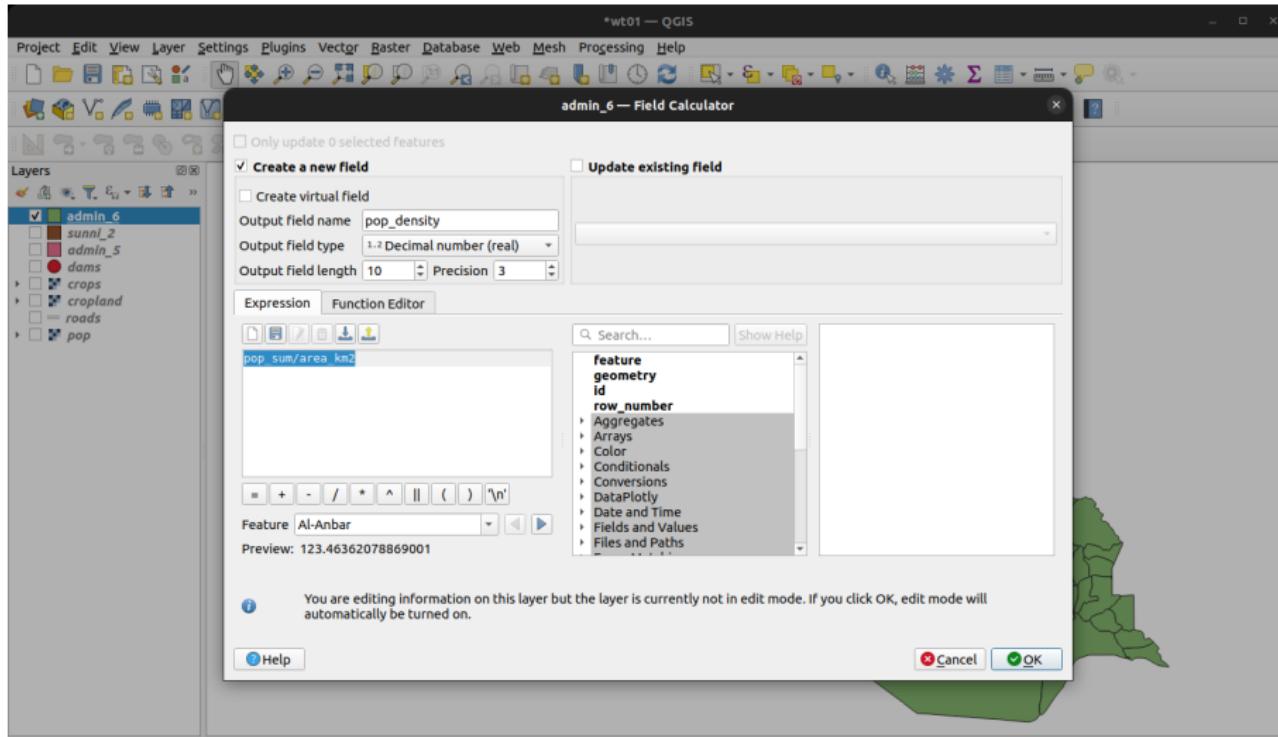
ID	Name	Type	Type name	Length	Precision
abc_2	COUNTRY	Text (string)	String	0	0
abc_3	GID_1	Text (string)	String	0	0
abc_4	NAME_1	Text (string)	String	0	0
abc_5	NL_NAME_1	Text (string)	String	0	0
abc_6	NAME_2	Text (string)	String	0	0
abc_7	VARNAME_2	Text (string)	String	0	0
abc_8	NL_NAME_2	Text (string)	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

At the bottom of the dialog, there are three buttons: "Help", "Style", "Apply", "Cancel", and "OK". The "OK" button is highlighted with a blue border. To the right of the dialog, a map view shows a green polygon representing the "admin_6" layer.

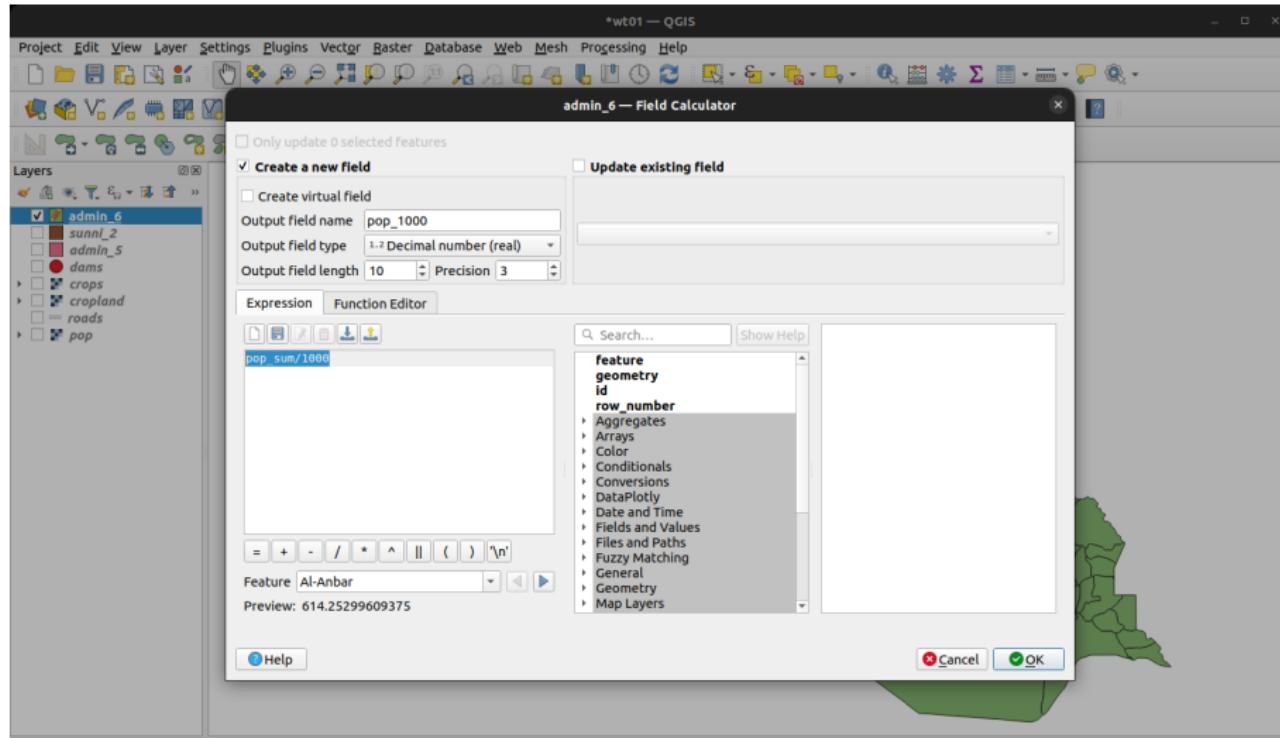
When prompted, save the changes you just made to admin_6



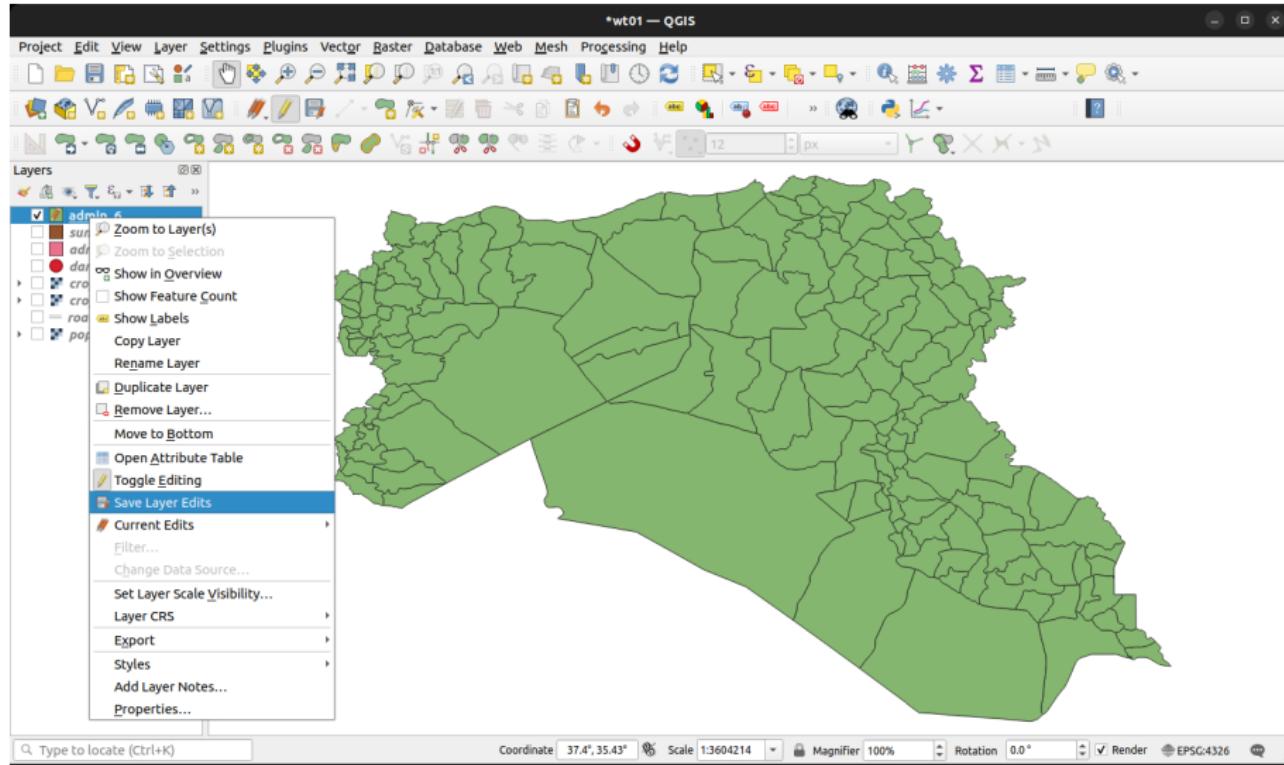
While we're at it, let's create a couple additional fields that could come in handy in the analysis. Create a **population density** field, named `pop_density`, of type Decimal number (real), with Expression set to `pop_sum/area_km2`



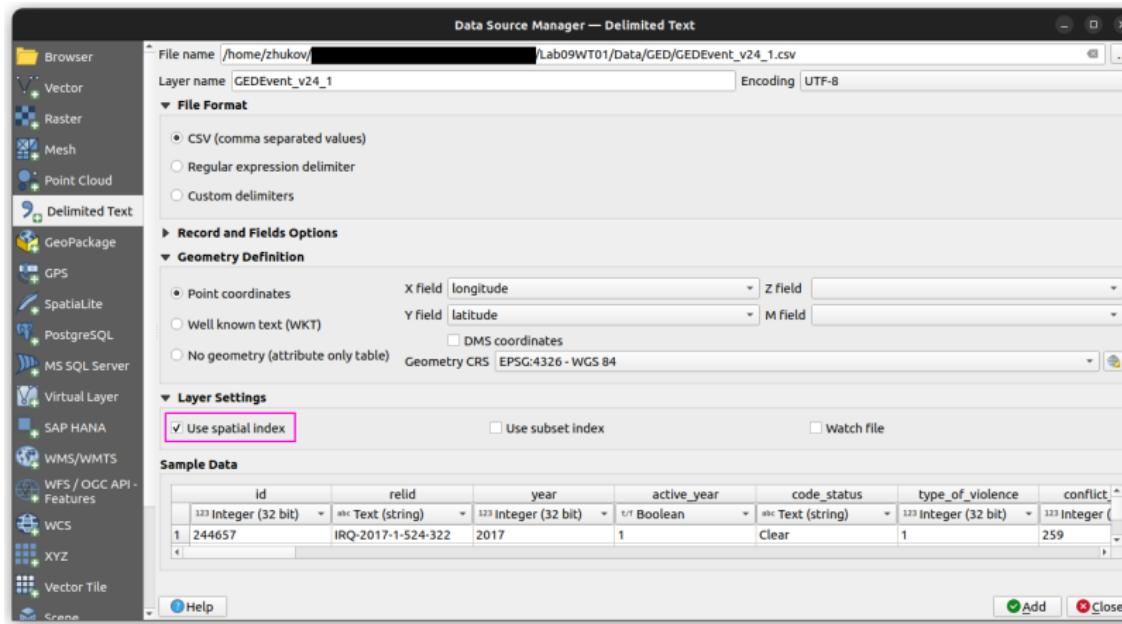
Let's also create a **rescaled population** field (1000's of residents). Name it **pop_1000**, with type **Decimal number (real)**. Set Expression to **pop_sum/1000**



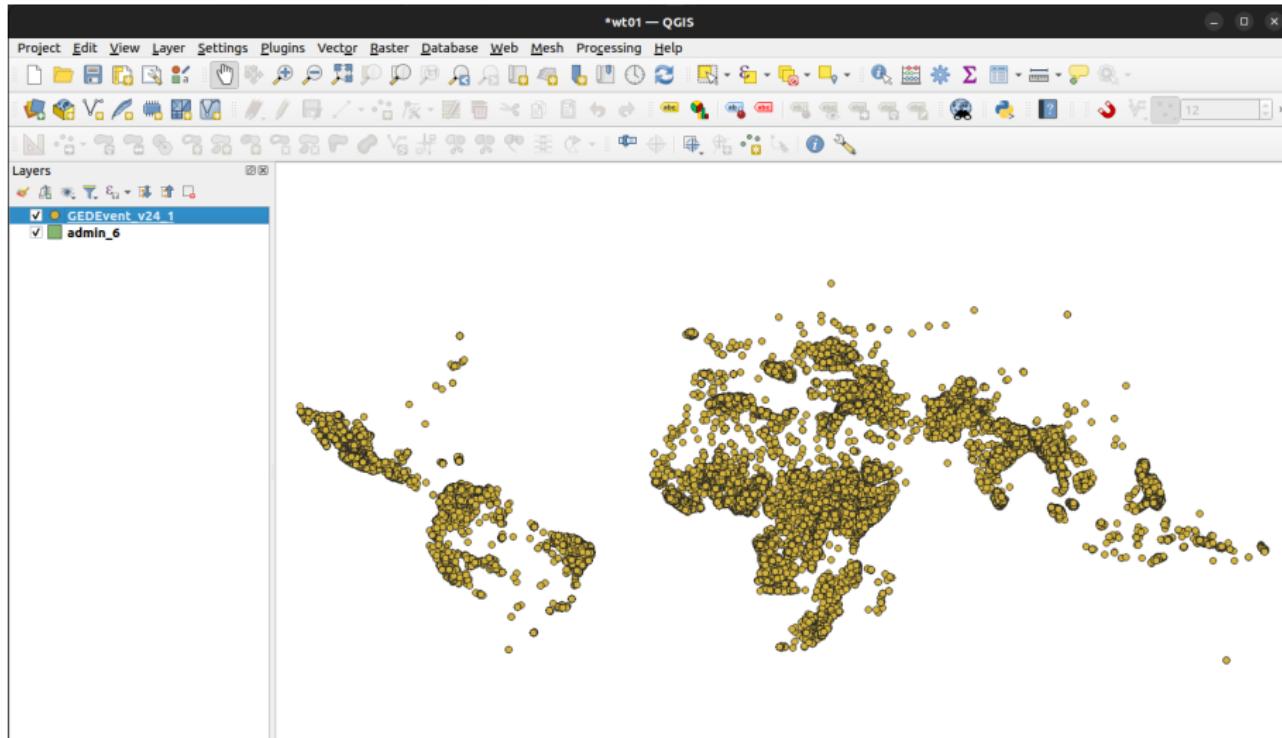
Save the layer edits to admin_6!



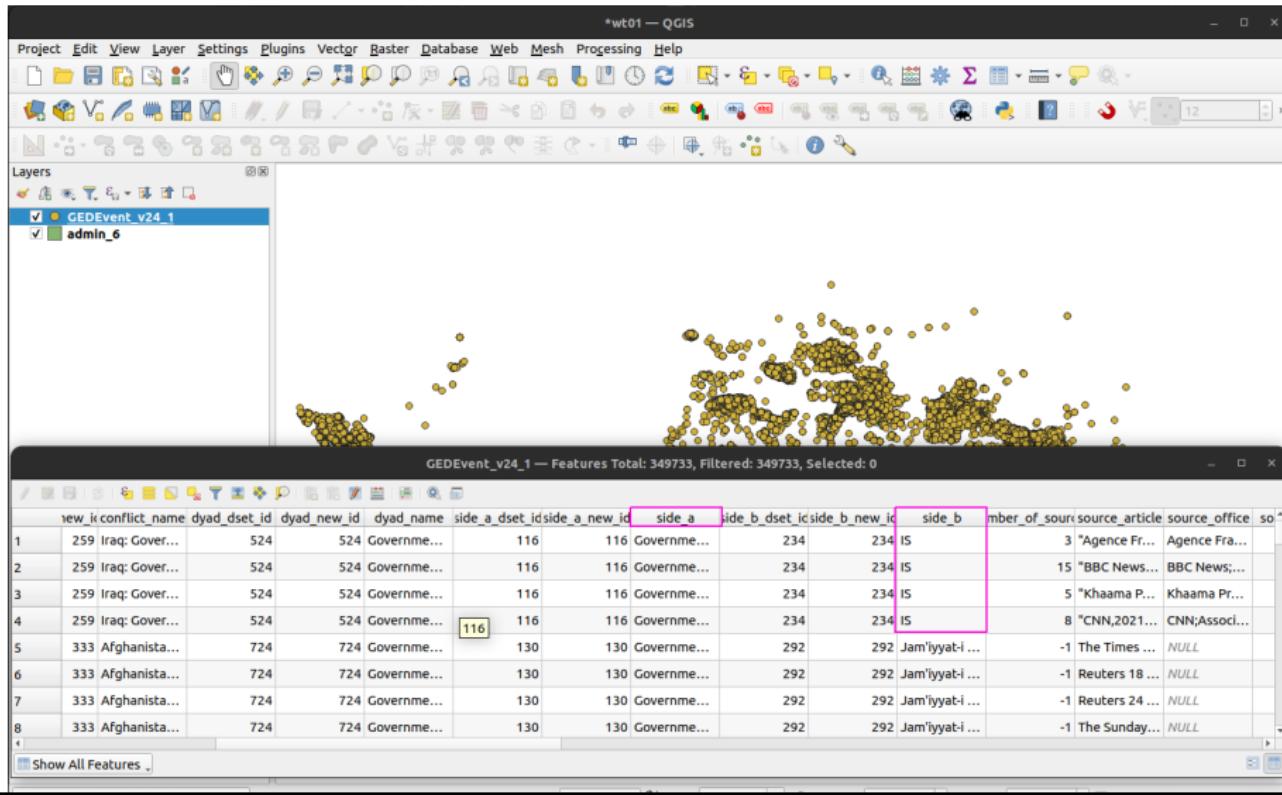
Add the *Islamic State violence data* to the project, using Add Delimited Text Layer.... Load the GEDEvent_v24.csv file in Data/GED folder. Set X field = longitude and Y field = latitude. Check the box next to Use spatial index



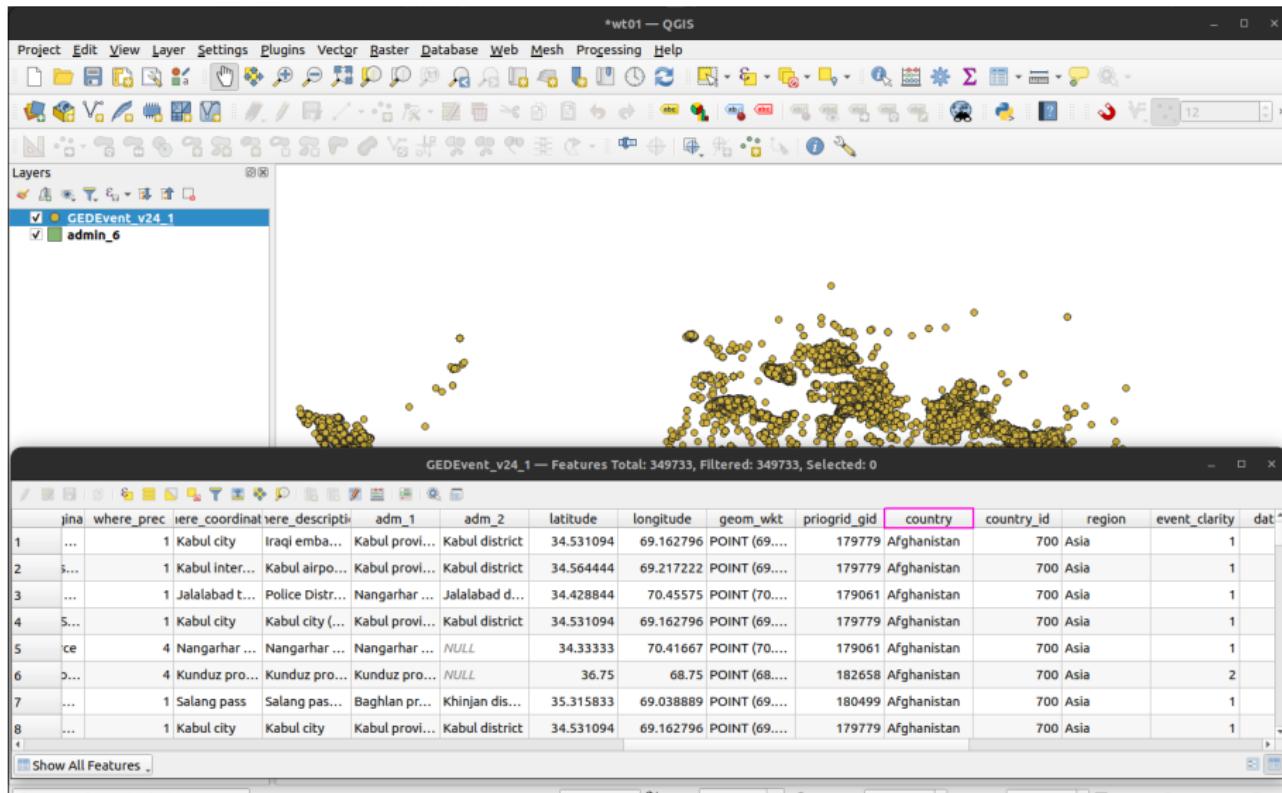
The (global) GED violent events layer should appear. There are several hundred thousand points here. We need to **extract events in Syria and Iraq involving the Islamic State**



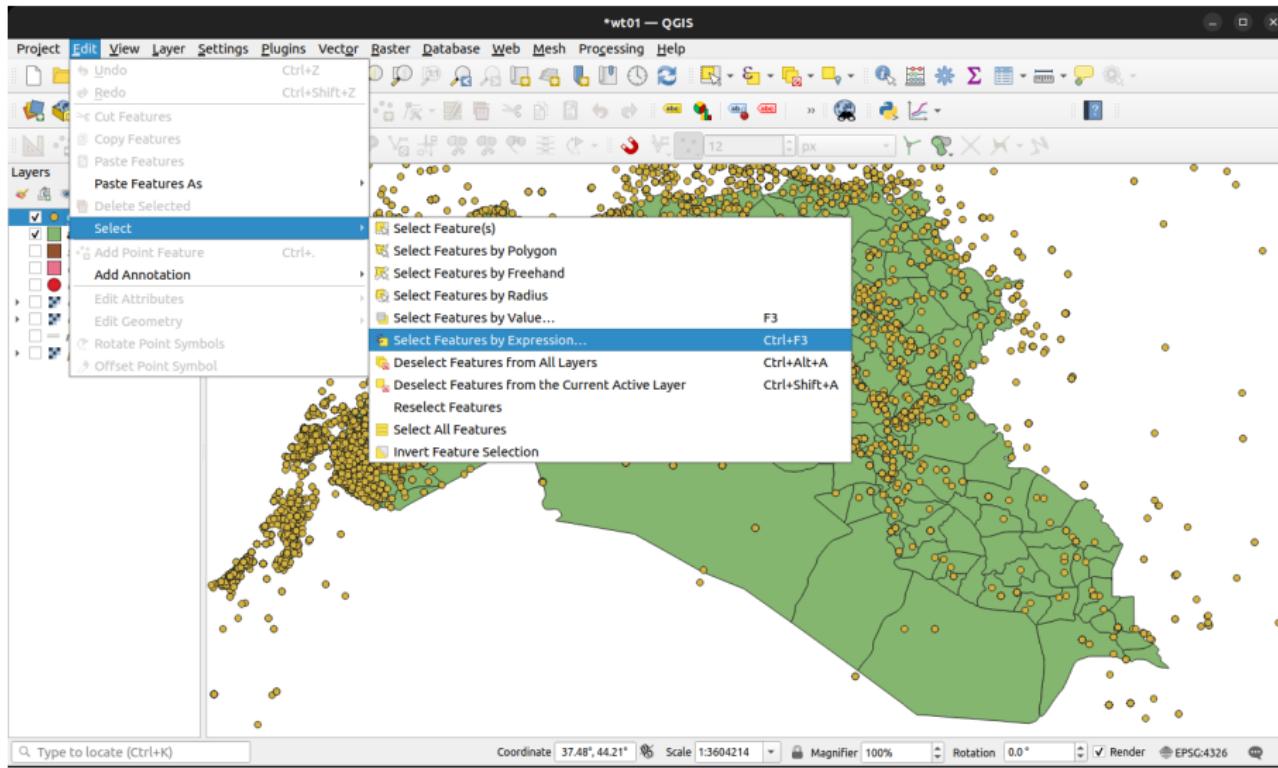
To figure out how to extract this data subset (by actor and location), let's explore the attribute table. We see fields for `side_a` and `side_b`, which list actors



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)

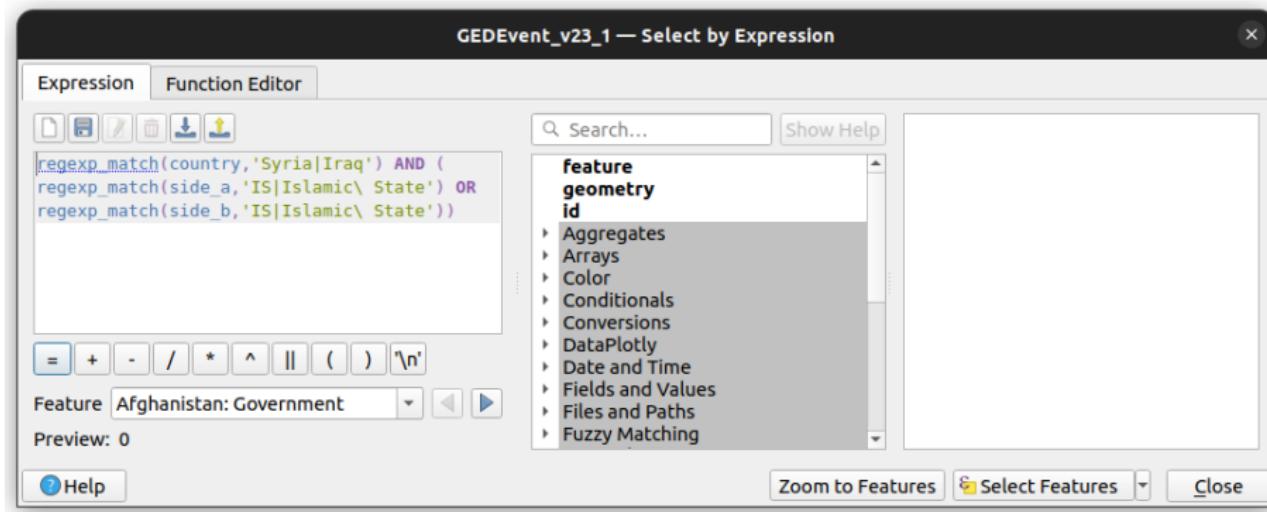


Highlight the GED layer and go to Edit → Select → Select by Expression...



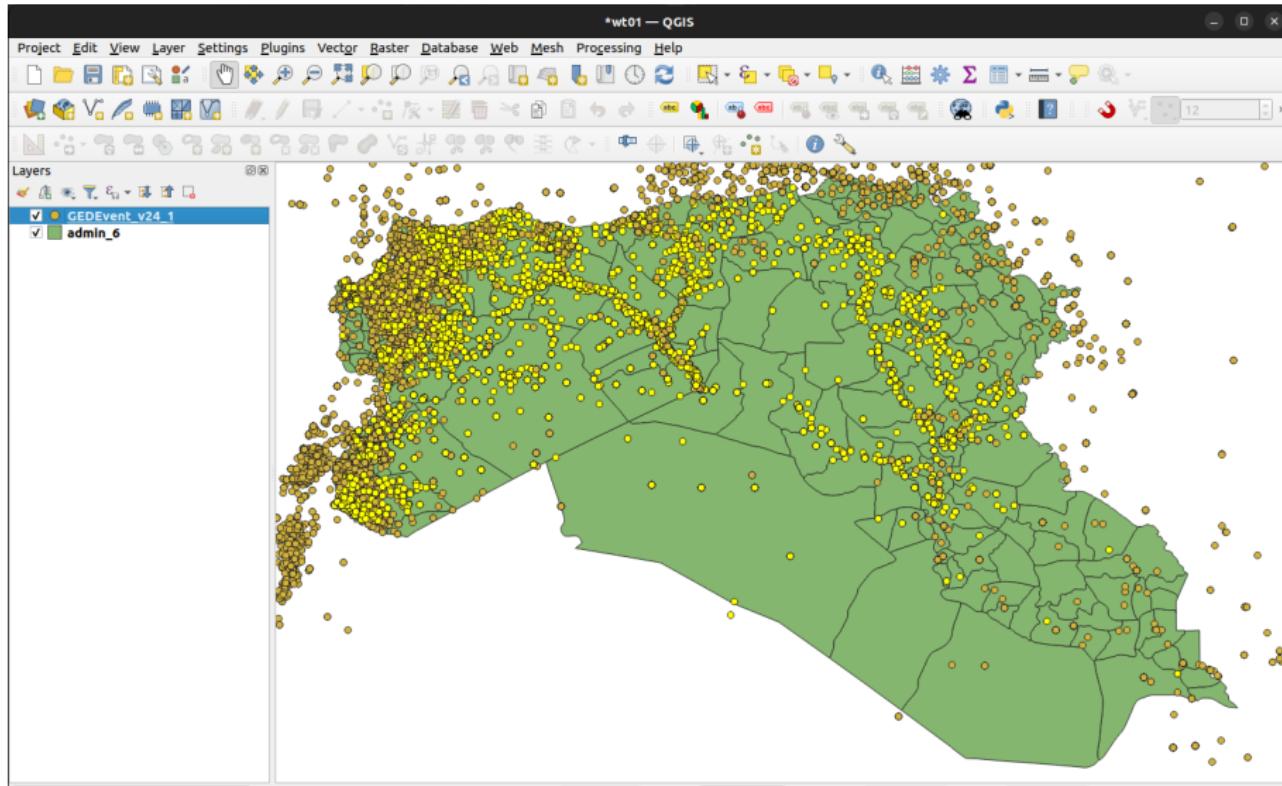
Let's combine regular expressions with logical operators. Set Expression to

```
regexp_match(country, 'Syria|Iraq') AND (
regexp_match(side_a, 'IS|Islamic State') OR
regexp_match(side_b, 'IS|Islamic State'))
```

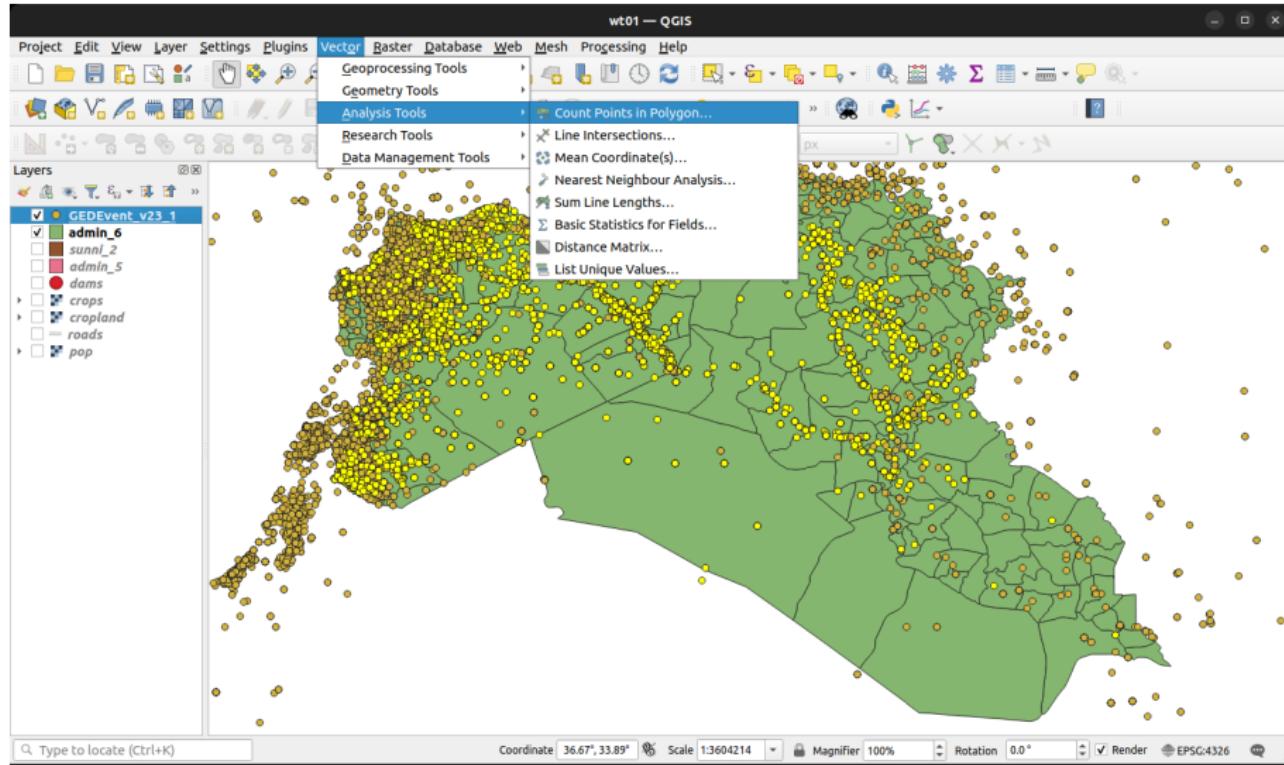


The vertical slash | is a regular expression for “OR”. So, in English this expression means ‘features where the field country contains “Iraq or Syria” and either side_a or side_b contains “IS or Islamic State”’

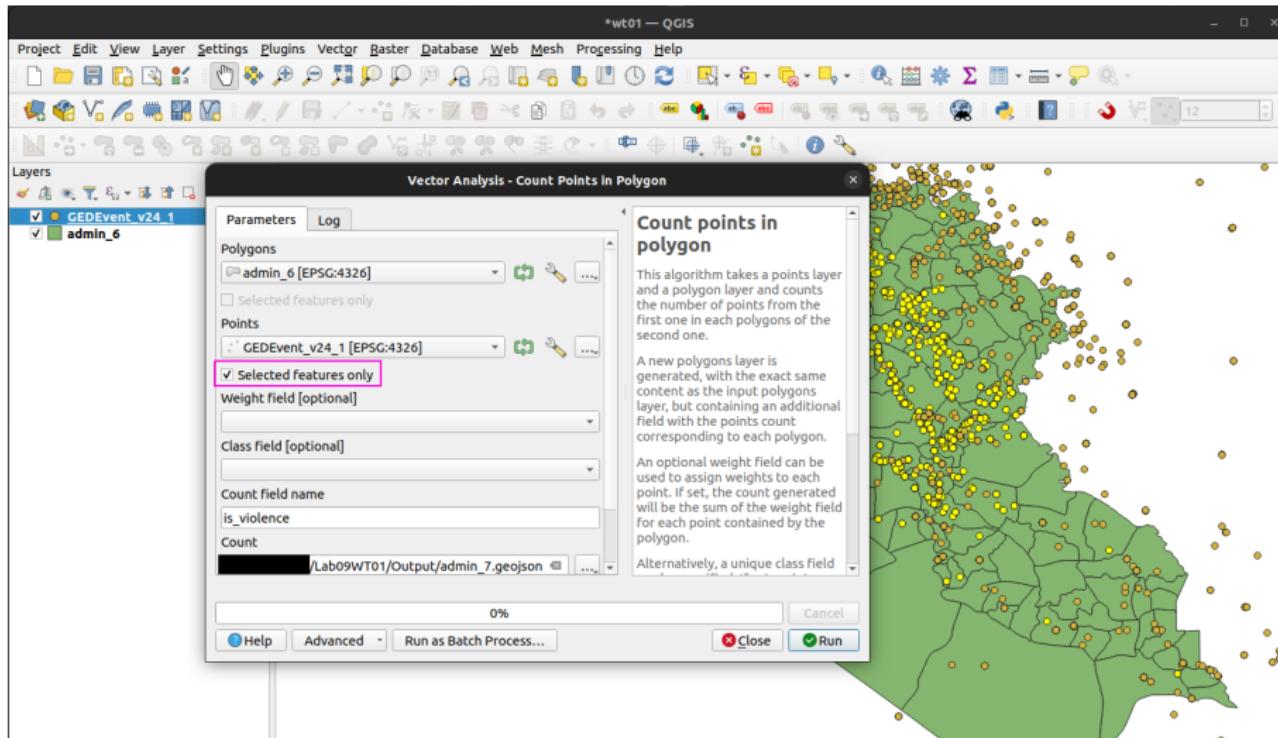
This should select a little over 23,000 events. Now let's calculate the **number of ISIS attacks per district**



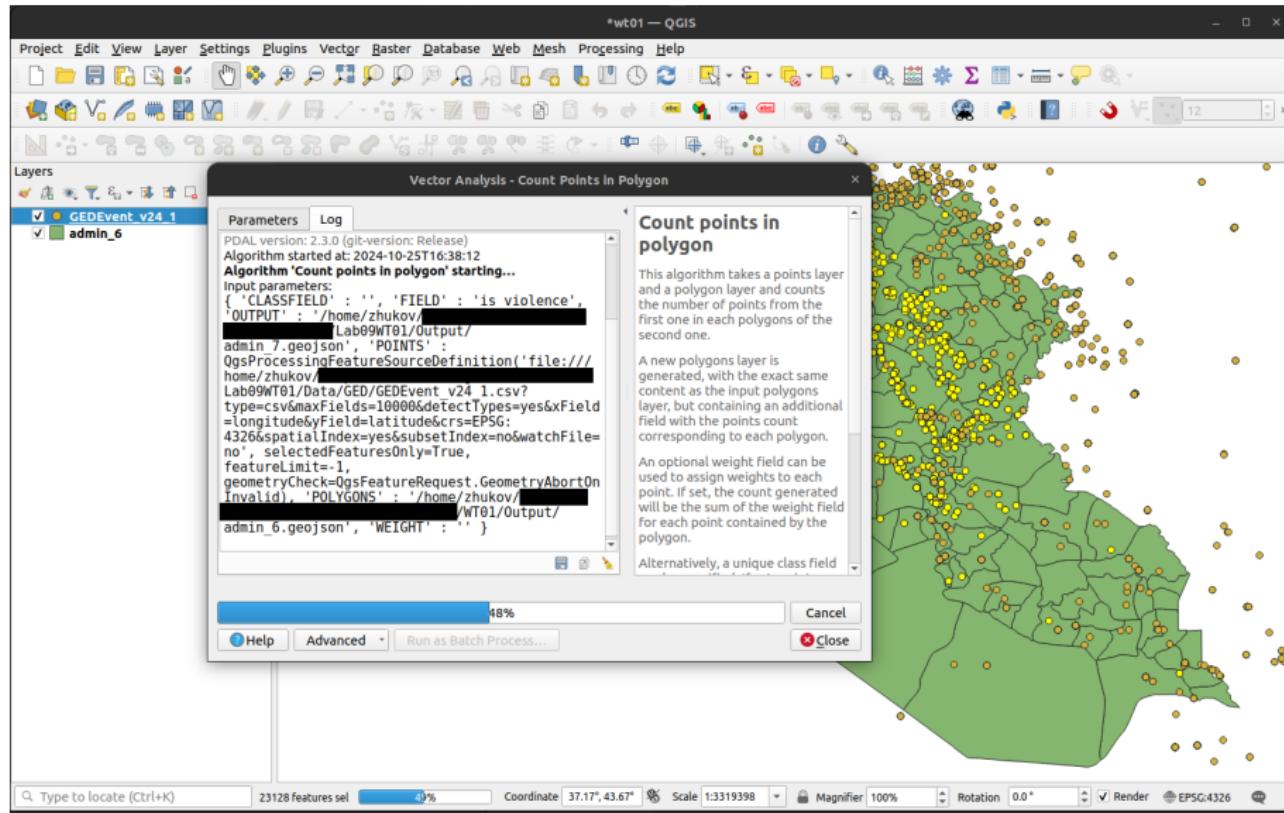
Open the Count Points in Polygon tool



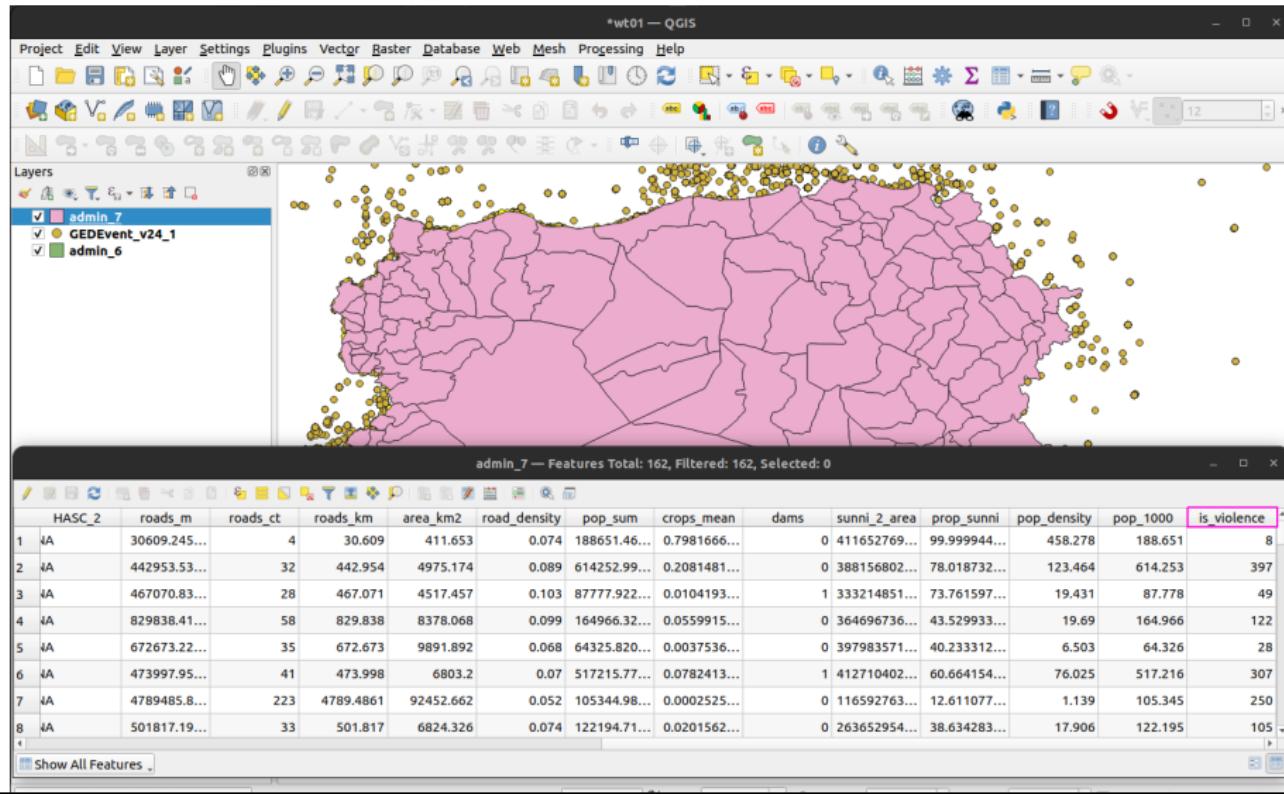
Select Polygons = admin_6, Points = GEDEvent_v24_1. Make sure the box is checked next to Selected Features Only for the points. Name the count field is_violence, and save the output file as admin_7.geojson. Click Run



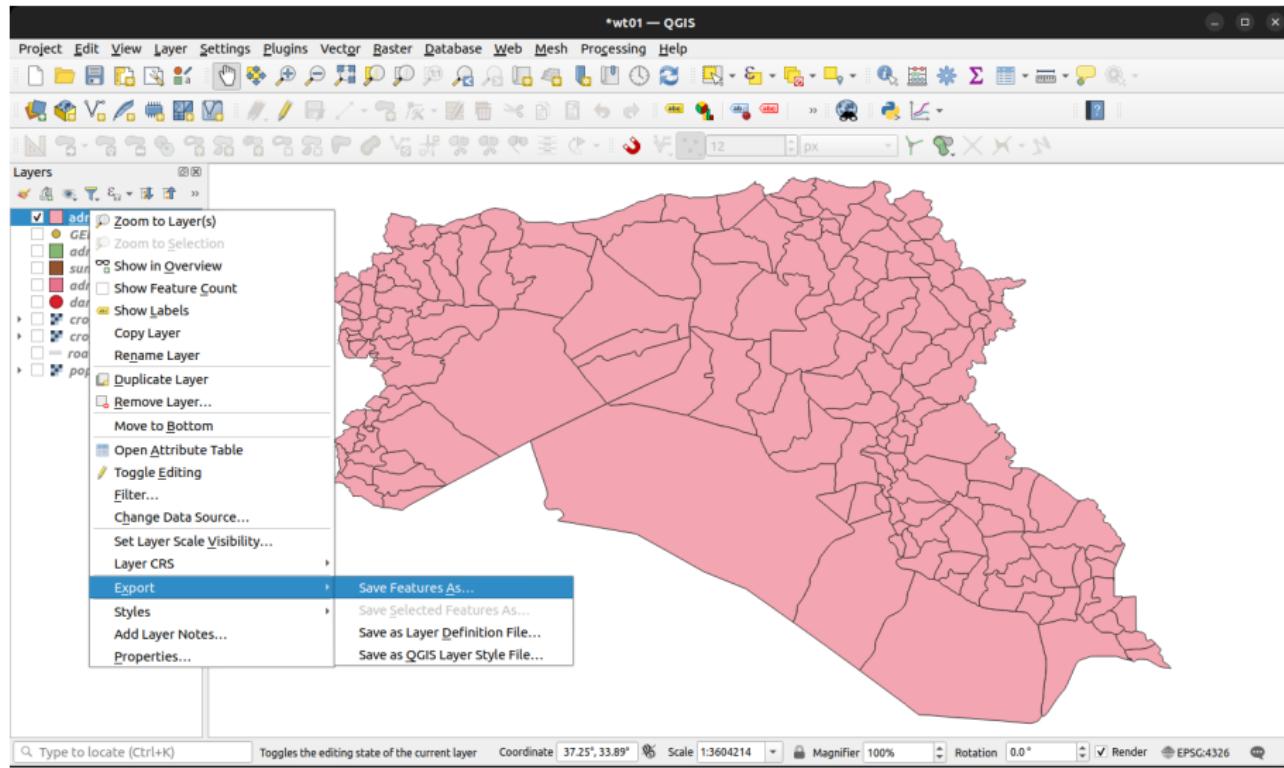
This may take a minute or two to run



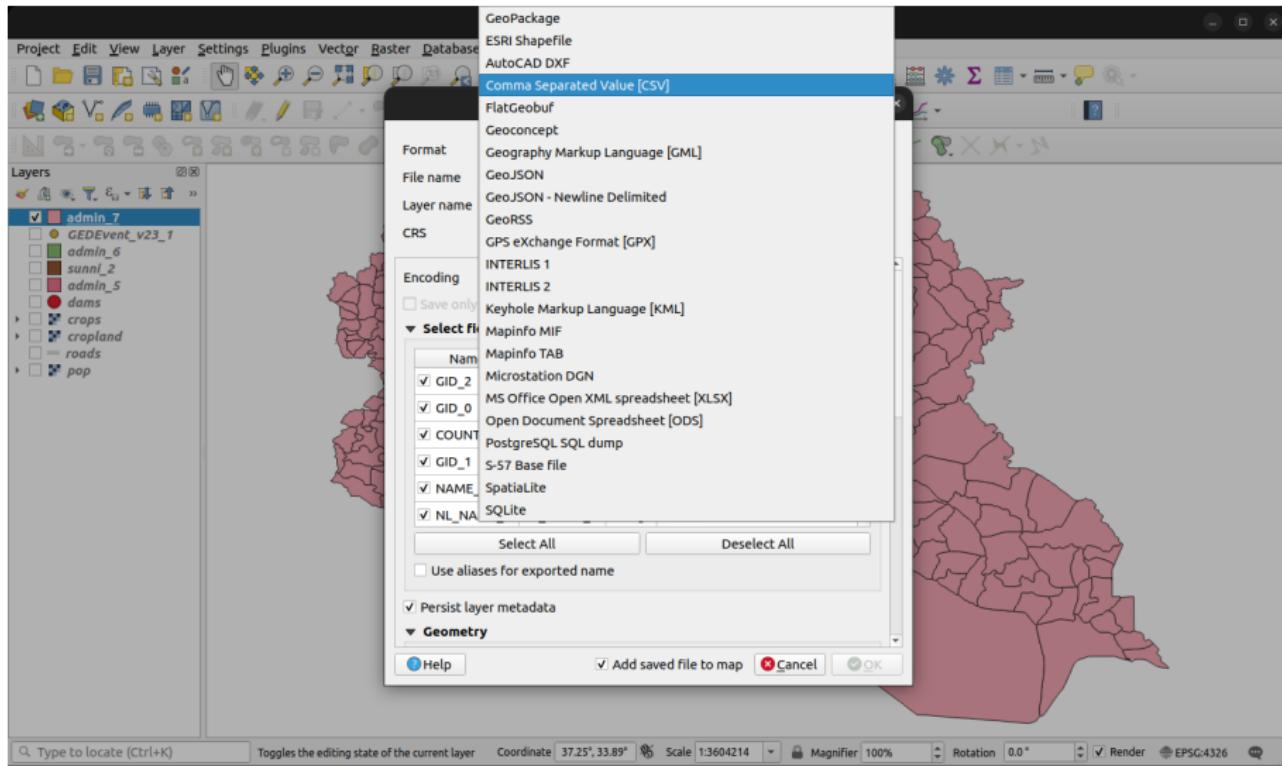
The attribute table for this new layer should include all the new fields we have generated. Let's now **export the table as CSV** for further analysis



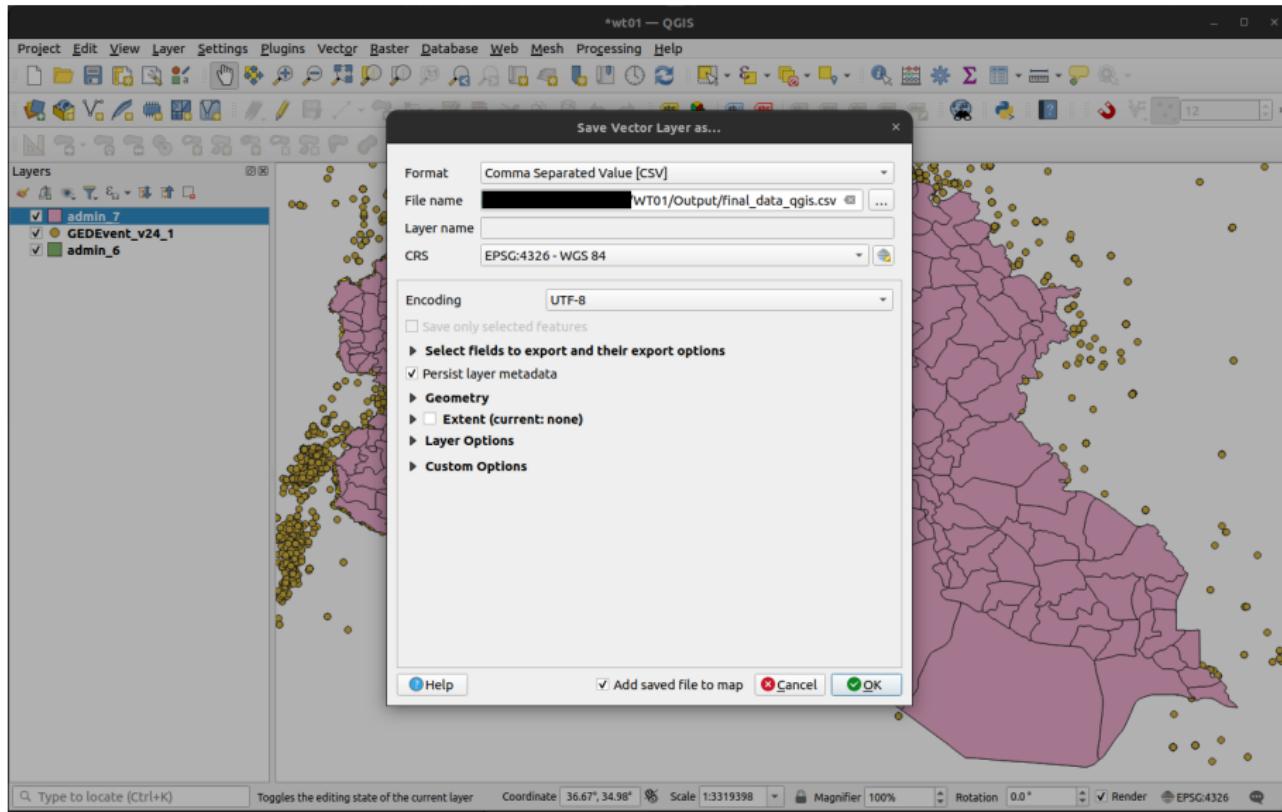
Right-click admin_7 in the Layer menu, go to Export→Save Features As.



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] "ENGTTYPE_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

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

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(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  0.0261092  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  
##          (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          0.790  
## Adjusted R2         0.556          0.730  
## Residual Std. Error  1.543 (df = 156)          1.204 (df = 125)  
## F Statistic         41.351*** (df = 5; 156)          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	$\beta_{\text{prop_sunni}} > 0$	✓	✓