

SEST-6577

# **Geographic Information Systems for Security Studies**

## Lab 09 (+ Walk Through 1)

Yuri M. Zhukov  
Associate Professor  
Georgetown University

November 4, 2025

**Goal:** explain geographic variation in Islamic State violence

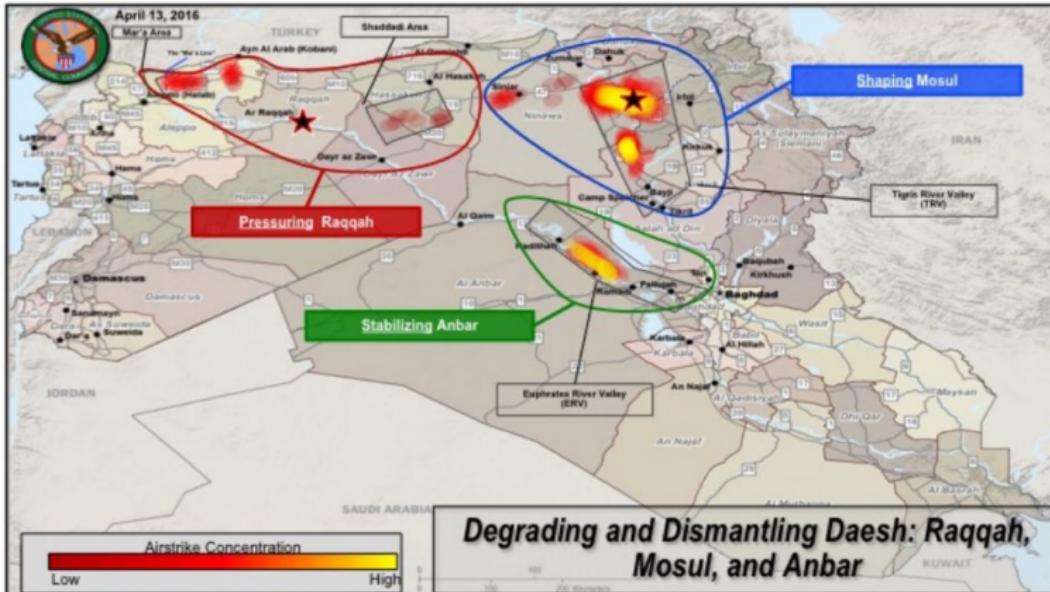


Figure 1: Operation Inherent Resolve

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 or Python, 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 (or Python)

# 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”

georeferenced violent events data

All Images Videos News More

Protected All regions Safe search: moderate Any time

Not many results contain violent

Search only for georeferenced "violent" events data?

**Uppsala Conflict Data Program**  
<https://ucdp.uu.se/downloads>

**UCDP Dataset Download Center**

UCDP Georeferenced Event Dataset (GED) Global version 25.1 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...

Uppsala Conflict Data Program  
The Uppsala Conflict Data Program...  
(UCDP) has recorded ongoing violent...

**FAOSTAT**  
<https://data.apps.fao.org/catalog/iso/820b82b2-ecb8-4137-85ef-812af620724b>

Georeferenced Event Dataset - UCDP - Datasets - "FAO catalog"

The UCDP Georeferenced Event Dataset (GED) provides detailed records of individual episodes of organized violence, coded by approximate date and location. Each event includes information on conflict type, actors, and estimated fatalities, and is disaggregated to daily resolution and subnational...

**Wolfram Data Repository**  
<https://datarepository.wolframcloud.com/resources/Global-Events-of-Organized-Violence>

Global Events of Organized Violence | Wolfram Data Repository

Global Events of Organized Violence Georeferenced dataset of individual events of organized violence from the Uppsala Conflict Data Program Uppsala Conflict Data Program This dataset is UCDP's most...

Share Feedback

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](http://ucdp.uu.se/downloads).  
Click on “Georeferenced Event Dataset (GED)”

The screenshot shows a web browser window with the title "UCDP Dataset Download" and the URL "ucdp.uu.se/downloads/". The main content area is titled "UCDP Dataset Download Center". It features several sections: "Disaggregated Datasets" with links to "UCDP Georeferenced Event Dataset (GED) Global version 25.1" (which is highlighted with a red box) and "UCDP Candidate Events Dataset (UCDP Candidate) version 25.0.X"; "Yearly Datasets covering 1946 - 2024" with links to "UCDP/IPRIO Armed Conflict Dataset version 25.1" and "UCDP Dyadic Dataset version 25.1"; "Yearly Datasets covering 1989 - 2024" with links to "UCDP One-sided Violence Dataset version 25.1", "UCDP Non-State Conflict Dataset version 25.1", and "UCDP Battle-Related Deaths Dataset version 25.1"; and "Other UCDP Datasets". The top right of the page has buttons for "COUNTRIES IN CONFLICT VIEW" and "FATALITIES VIEW".

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 "ucdp.uu.se/downloads/index.html#ged\_global". The main content area has a heading "Disaggregated Datasets" and a sub-heading "UCDP Georeferenced Event Dataset (GED) Global version 25.1". A descriptive text follows: "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 (highlighted with a red border), EXCEL, R, STATA, and CODEBOOK. At the bottom, there is a "Please cite:" section with three bullet points:

- Davies, S., Pettersson, T., Sollenberg, M., & Öberg, M. (2025). Organized violence 1989–2024, and the challenges of identifying civilian victims. *Journal of Peace Research*, 62(4).
- Sundberg, Ralph and Erik Melander (2013) Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research* 50(4).
- UCDP is part of and funded by DEMSCORE, national research infrastructure grant 2021-00162 from the Swedish Research Council.

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

GEDEvent\_v25\_1.csv - LibreOffice Calc

File Edit View Insert Format Styles Sheet Data Tools Window Help

Liberation Sans 10 pt

AD1:AF1      latitude

Y	Z	AA	AB	AC	AD	AE	AF	AG
1	where_prec	where_coordinates	where_description	adm_1	adm_2	latitude	longitude	geom_wkt
2	1	Kabul city	Iraqi embassy in K	Kabul province	Kabul district	34.53109	69.1628	POINT (69.162796 34.531094)
3	1	Kabul international	Kabul airport (Abb	Kabul province	Kabul district	34.56444	69.217222	POINT (69.217222 34.564444)
4	1	Jalalabad town	Police District 7 of Nangarhar provin	Jalalabad district	34.42884	70.45575	POINT (70.45575 34.428844)	1790
5	1	Kabul city	Kabul city (district	Kabul province	Kabul district	34.53109	69.1628	POINT (69.162796 34.531094)
6	4	Nangarhar province	Nangarhar provin	Nangarhar province		34.33333	70.41667	POINT (70.41667 34.33333)
7	4	Kunduz province	Kunduz province	Kunduz province		36.75	68.75	POINT (68.75 36.75)
8	1	Salang pass	Salang pass (on th	Baghlan province	Khujan district	35.31583	69.03889	POINT (69.038889 35.315833)
9	1	Kabul city	Kabul city	Kabul province	Kabul district	34.53109	69.1628	POINT (69.162796 34.531094)
10	2	Salang tunnel	Salang tunnel (nea	Baghlan province	Khujan district	35.31603	69.03871	POINT (69.03871 35.31603)
11	4	Hirat province	Hirat province	Hirat province		34.5	62	POINT (62 34.5)
12	1	Kabul city	Kabul city	Kabul province	Kabul district	34.53109	69.1628	POINT (69.162796 34.531094)
13	1	Salang pass	Salang pass	Baghlan province	Khujan district	35.31583	69.03889	POINT (69.038889 35.315833)
14	1	Kabul city	Kabul city	Kabul province	Kabul district	34.53109	69.1628	POINT (69.162796 34.531094)
15	1	Kabul city	Kabul city	Kabul province	Kabul district	34.53109	69.1628	POINT (69.162796 34.531094)
16	1	Kabul city	Kabul city (Kherkh	Kabul province	Kabul district	34.53109	69.1628	POINT (69.162796 34.531094)
17	4	Panwan province	Panwan province (P	Panwan province		35	68.91667	POINT (68.91667 35)
18	1	Kabul city	Kabul city	Kabul province	Kabul district	34.53109	69.1628	POINT (69.162796 34.531094)
19	6	Afghanistan				33	65	POINT (65 33)
20	1	Kabul city	Kabul city	Kabul province	Kabul district	34.53109	69.1628	POINT (69.162796 34.531094)
21	1	Kabul international	Kabul airport	Kabul province	Dih Sabz district	34.56444	69.21722	POINT (69.217222 34.564444)
22	2	Jabalussaraj town	Hajjan village (nea	Panwan province	Jabalussaraj dist	35.11837	69.2357	POINT (69.235695 35.118366)
23	1	Kabul city	Kabul city (in or ne	Kabul province	Kabul district	34.53109	69.1628	POINT (69.162796 34.531094)
24	1	Kabul city	Kabul city	Kabul province	Kabul district	34.53109	69.1628	POINT (69.162796 34.531094)
25	1	Kabul city	Kabul city	Kabul province	Kabul district	34.53109	69.1628	POINT (69.162796 34.531094)
26	2	Kabul city	Kabul city (just out	Kabul province	Kabul district	34.53109	69.1628	POINT (69.162796 34.531094)
27	1	Kabul city	Kabul city (Kheir Kt	Kabul province	Kabul district	34.53109	69.1628	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`

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](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](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 under "Country" is set to "Syria". To the right is a map of Syria with its administrative divisions outlined in blue and red. Below the map is a URL: [https://geodata.ucdavis.edu/gadm/gadm4.1/json/gadm41\\_SYR\\_2.json.zip](https://geodata.ucdavis.edu/gadm/gadm4.1/json/gadm41_SYR_2.json.zip).

Multiple data sources exist for **roads**. Among the easiest to get are those from [diva-gis.org/gData](https://diva-gis.org/gData)

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

- Search Bar:** The search term "roads GIS data by country" is entered.
- Address Bar:** The URL is
- Results:** The first result is a link to "Download data by country | DIVA-GIS".
  - 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..."
- Second Result:** A link to "GRIP global roads database | GLOBIO - Global biodiversity model for ...".
  - 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 Navigation:** Includes links to "Share Feedback" and "https://datacatalog.worldbank.org / dataset / grip-global-roads-inventory-dataset-2018-road-de..."

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 the following text: "Select and download free geographic (GIS) data for any country in the world<br>". 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:

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)	-
Land cover	Digital land cover maps, and lakes. Separate files for line	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" followed by navigation links: "Documentation :: Contents :: Climate data for DIVA-GIS >". Below the header, a section titled "Data" contains the following text: "Select and download free geographic (GIS) data for any country in the world<br>". There are three dropdown menus: "Country" set to "Syria", "Subject" set to "Roads", and "Format". A link "Download [SYR\\_rds.zip](#)" is provided. A "Sources" section at the bottom contains a table with two rows:

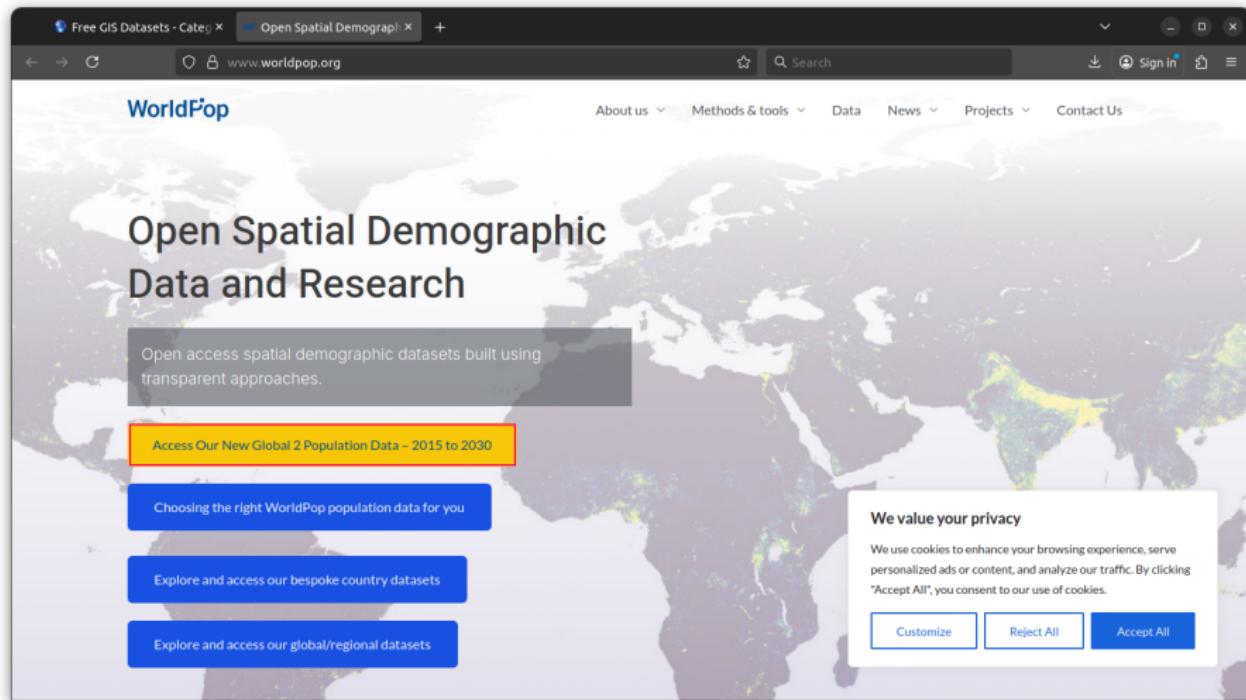
Subject	Description	Source	Format	Resolution
Administrative areas (boundaries)	Country outlines and administrative subdivisions for all countries. The level of subdivisions that is available varies between countries	GADM	Vector (area)	-
Land cover	Digital vector files, and lakes. Separate files for line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

For **population data**, we can go back to Rick Wilson's page. There are many options, but let's try "WorldPop"

Free GIS Data | Home | Physical | Human | Country-specific | FAQ | Like | Tweet | Donate

- [WorldCereal](#): global cereal crop maps at 10m resolution, plus more to come in future
- Lakes, Oceans and other Water Sources**
  - [Coastal Water Quality](#): Quality of coastal waters across the globe measured by chlorophyll concentrations from SeaWiFS satellite. Data for 1998 and 2007.
  - [Global Reservoir and Dam Database](#): Geographically-referenced data on all reservoirs with a storage capacity of more than 0.1 cubic kilometres. The 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 population 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](#)
- 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 coverage, but includes such things as shop and business locations (eg. all Tesco stores, all McDonald's restaurants) as well as places of worship, speed cameras etc. Registration is required. [To download data in a GIS-ready form choose Garmin CSV format on the download page](#). The CSV file will contain Latitude and Longitude in WGS-84 co-ordinates, as well as descriptions.

On the landing page, go to “Access... data”.



Let's grab the global mosaic file.

The screenshot shows a web browser window for the WorldPop Hub at [hub.worldpop.org/project/categories?id=3](http://hub.worldpop.org/project/categories?id=3). The page title is "WorldPop Hub". The main content discusses population datasets, mentioning "Global2" datasets and "Bespoke methods for individual countries (WOPR)". A red box highlights the first item in a list of datasets:

- Global mosaics 2015-2030 ( 1km resolution ) R2025A v1
- Individual countries 2015-2030 ( 100m resolution ) R2025A v1
- Individual countries 2015-2030 ( 1km resolution ) R2025A v1
- Bespoke methods for individual countries (WOPR)

At the bottom, there is a "Privacy and Cookies Policy" notice with "Accept and Continue" and "Learn More" buttons.

The earliest population raster they have is from 2015 (1 year after IS campaign began). Suboptimal, but we'll take it. Click on Data and Resources.

WorldPop :: Population C x + hub.worldpop.org/geodata/listing?id=137 DATA | CONTACT

## WorldPop Hub

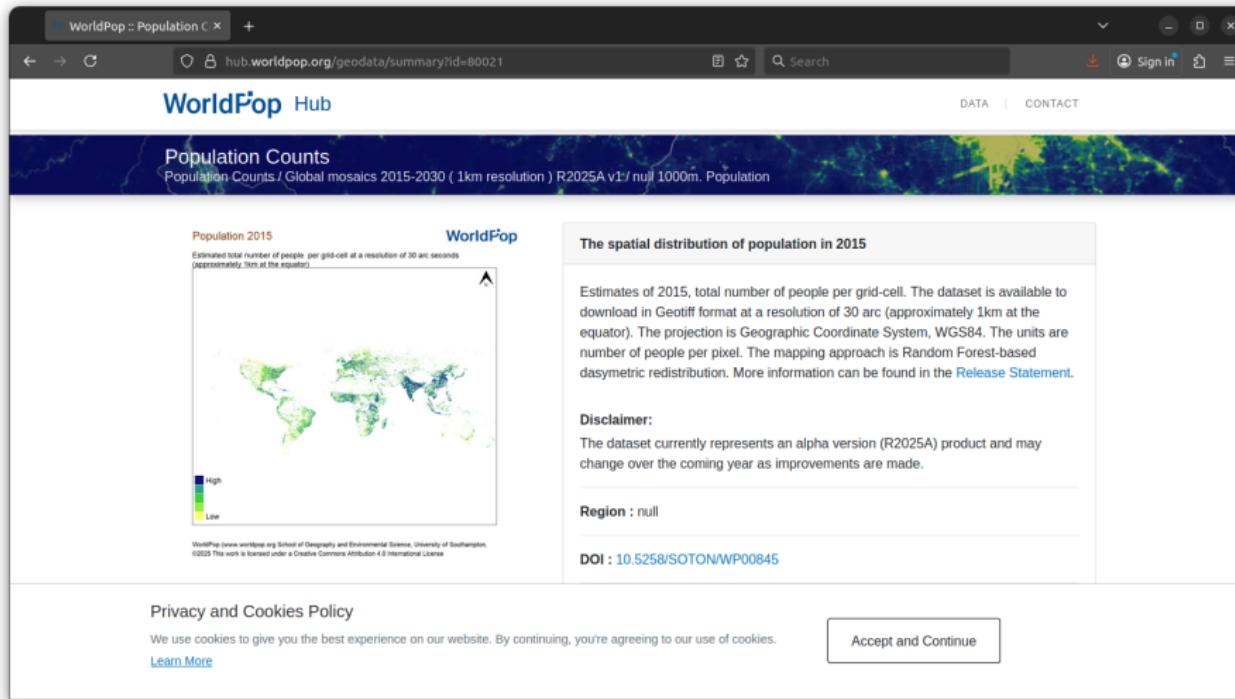
Mosaiced 1km resolution global datasets using 100m resolution population count datasets. More information can be found in the [Release Statement](#).

**Suggested Citation:**  
Bondarenko M., Priyatikanto R., Tejedor-Garavito N., Zhang W., McKeen T., Cunningham A., Woods T., Hilton J., Cihan D., Nosatiuk B., Brinkhoff T., Tatem A., Sorichetta A.. *The spatial distribution of population in 2015-2030 at a resolution of 30 arc (approximately 1km at the equator) R2025A version v1. Global Demographic Data Project - Funded by The Bill and Melinda Gates Foundation (INV-045237). WorldPop - School of Geography and Environmental Science, University of Southampton. DOI:10.5258/SOTON/WP00845*

Show 25 rows	entries	Search ...			
Continent	Country	Year	Geo Type	RES	Data & Resources
World	null	2015	Population	1km	<a href="#">Data &amp; Resources</a>
World	null	2016	Population	1km	<a href="#">Data &amp; Resources</a>
World	null	2017	Population	1km	<a href="#">Data &amp; Resources</a>
World	null	2018	Population	1km	<a href="#">Data &amp; Resources</a>
World	null	2019	Population	1km	<a href="#">Data &amp; Resources</a>
World	null	2020	Population	1km	<a href="#">Data &amp; Resources</a>

**Privacy and Cookies Policy**  
We use cookies to give you the best experience on our website. By continuing, you're agreeing to our use of cookies.  
[Learn More](#) [Accept and Continue](#)

Here we see a preview of the data and, if we scroll down...



... a link to download the data.

The screenshot shows a web browser window for the WorldPop Hub. The URL in the address bar is [hub.worldpop.org/geodata/summary?id=80021](https://hub.worldpop.org/geodata/summary?id=80021). The page displays metadata for a dataset, including a DOI (10.5258/SOTON/WP00845), a date of production (2025-09-01), and a recommended citation. A prominent green button at the bottom left allows users to "Download Entire Dataset / 274.87 MB". Below the main content, there is a cookie consent banner with options for "Privacy and Cookies Policy", "Accept and Continue", and "Learn More". The full URL of the dataset page is visible at the bottom of the browser window: [data.worldpop.org/GIS/Population/Global\\_2015\\_2030/R2025A/2015/0\\_Mosaicked/v1/1km\\_us/constrained/global\\_pop\\_2015\\_CN\\_1km\\_R2025A\\_UA\\_v1.tif](https://data.worldpop.org/GIS/Population/Global_2015_2030/R2025A/2015/0_Mosaicked/v1/1km_us/constrained/global_pop_2015_CN_1km_R2025A_UA_v1.tif).

Also at the RT Wilson page, we can find data on **cropland** (e.g., “Earthstat”)

The screenshot shows a web browser displaying the 'Free GIS Data' website at [freegisdata.rtwilson.com](http://freegisdata.rtwilson.com). The search bar at the top contains the term 'cropland'. Below the search bar, there are several sections of data:

- Land Cover**: A link to see above.
- 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**:
  - (See also [Land Cover](#) above)
  - Global Land Use Dataset**: Gridded data at 0.5 degree resolution showing population density, potential natural vegetation, grazing land extent, built-up land extent, crop extent (for 18 major crops) and land suitability for cultivation.
  - Human Influence and Footprint**: Human Influence Index and Human Footprint calculated from various factors which exert human influence on ecosystems, for example population distribution, urban areas, navigable rivers etc. Available at 30 arc-second resolution.
  - Global Agricultural Lands**: Extent and intensity of use of agricultural lands (both [cropland](#) and pasture) in 2000 from MODIS and SPOT images and agricultural inventory data.
  - Global Irrigated Area and Rainfed Crops Areas**: Vector mapping of global irrigated [cropland](#) and rainfed [cropland](#).
  - Crop Calendar GIS**: Gridded data on planting dates and harvesting dates across the world for 19 crops. Available at 5 minute and 0.5 degree resolutions.
  - EarthStat: Agricultural Land Use and potential use**: A number of GIS datasets on agricultural land use, including global [cropland](#) and pasture from 1700 to 2007, harvested areas and yields for 175 crops, and global fertiliser application rates.
  - European Urban Morphological Zones**: Data derived from the CORINE landcover dataset showing all sets of urban areas lying less than 200m apart.
  - Global Human Settlement Layer**: Multi-temporal data on built-up areas, including population, settlements and degree of urbanisation.
  - Global Urban Footprint**: Global map of built-up areas at 12m resolution.
  - Bing Global Building Footprints**: Global building footprints, derived by machine learning from Bing aerial imagery, and released by Microsoft
  - WorldCereal**: global cereal crop maps at 10m resolution, plus more to come in future
- Lakes, Oceans and other Water Sources**:
  - Coastal Water Quality**: Quality of coastal waters across the globe measured by chlorophyll concentrations from SeaWiFS satellite. Data for 1998 and 2007.

The first dataset we see here is a global raster from 2000. This pre-dates the IS campaign (good) by about 14 years (less good). This likely overstates how much cropland was still active in 2014. Suboptimal, but OK for an exploratory analysis.

The screenshot shows a web browser window with the URL <http://www.earthstat.org>. The page is titled "EarthStat - GIS data for all". The main content area has a green header with the text "What is EarthStat?". Below it, a paragraph describes EarthStat's mission to serve geographic data sets that help solve the grand challenge of feeding a growing global population while reducing agriculture's impact on the environment. A collaboration between the Global Landscapes Initiative at the University of Minnesota's Institute on the Environment and the Land Use and Global Environment lab at the University of British Columbia. The main content area also features a section titled "Cropland and Pasture Area in 2000" with a subtext explaining that agricultural activities have dramatically altered our planet's land surface. It mentions creating a global data set of croplands and pastures circa 2000 by combining agricultural inventory data and satellite-derived land cover data. Two buttons are visible: "Download Data (Geotiff 18.8MB)" and "More about this Data".

For data on **hydroelectric dams**, we can do a quick web search. This quickly points us to a dataset called “Global Reservoir and Dam (GRanD)”.

The screenshot shows a search results page from DuckDuckGo. The search query is "global dams database". The top result is a link to [globaldamwatch.org](https://www.globaldamwatch.org/database), titled "GDW consensus global database - GlobalDamWatch.org". The snippet describes it as a new river barrier and reservoir database developed by the Global Dam Watch (GDW) consortium. Below this, there's a link to "Global Dam Watch" which is described as an international collaboration. Another result is from "Nature" about a "Global Dam Tracker". The final result at the bottom is the "The Global Reservoir and Dam Database" from "waterdata", which is highlighted with a red border. This database is described as containing 6,862 records of reservoirs and their associated dams with a cumulative storage capacity of 6,197 cubic km.

global dams database

duckduckgo.com/?l=ffab&q=global+dams+database&ia=web

All Images Videos News More

Protected All regions Safe search: moderate Any time

globaldamwatch.org https://www.globaldamwatch.org / database

GDW consensus global database - GlobalDamWatch.org

GDW consensus global database The GDW (V) consensus global database Here, we present a new river barrier and reservoir database developed by the Global Dam Watch (GDW) consortium that integrates, harmonizes, and augments existing global datasets to support large-scale analyses.

Global Dam Watch

Global Dam Watch is an international collaboration which aims to improve o...

Nature https://www.nature.com / articles / s41597-023-02008-2

Global Dam Tracker: A database of more than 35,000 dams with...

Therefore, a global dam database that improves our understanding of the net impact of dams is crucial for informing a more sustainable and equitable approach to economic development.

World Bank Water Data https://wbbwaterdata.org / dataset / the-global-reservoir-and-dam-database

The Global Reservoir and Dam Database - waterdata

The Global Reservoir and Dam Database The Global Reservoir and Dam Database, Version 1, Revision 01 (V1.0) contains 6,862 records of reservoirs and their associated dams with a cumulative storage capacity of 6,197 cubic km. The dams were geospatially referenced and assigned to polygons depicting...

Share Feedback

Look for a link to download the data.

The screenshot shows a web browser window with the title "global dams database at" and "The Global Reservoir and Dam Database". The URL is "wbwaterdata.org/dataset/the-global-reservoir-and-dam-database". The page content includes:

- ORGANIZATION**: A world map showing dam locations.
- OPENNESS**: A rating of 0 out of 5 stars.
- SOCIAL**: Links to Twitter and Facebook.
- LICENSE**: Creative Commons Non-Commercial (Any).
- DATA AND RESOURCES**:
  - Global Reservoir and Dam dataset**: An "Explore" button.
  - Direct link to GRAND Data**: An "Explore" button.
  - Future Hydropower Reservoirs and Dams**: An "Explore" button, with a "More Information" and "Go to resource" link highlighted by a red box.
- SHOW TABS**: A section with a unique URL: "ln.sync.com/dl/bd47eb6b0/anhxalkr-62pmrgtq-k44xf84f-pyz4atkkm".

Seems legit... (scan ZIPs for viruses before extracting)

The screenshot shows a web browser window with three tabs open:

- global dams database at
- The Global Reservoir and
- GRanD\_Version\_1\_3.zip | +

The main content area displays a file named "GRanD\_Version\_1\_3.zip" with a size of 42.7 MB. A note says "View not available." Below it is a "Download" button.

To the right, there is a "Comments" section:

- A message from "You are not signed in." with a "Sign in to comment" button.
- A reply from "harold.dellafaille@engie.com.ee" dated Oct 1, 2021, 6:50 AM, with a "Reply" button.
- A reply from "lesimuchu@163.com" dated Feb 20, 2023, 7:59 AM, with a "Reply" button.

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](https://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. A red box highlights the "GeoEPR-2021.geojson" link in the "Data" section.

**GeoEPR - Geo-referencing Ethnic Power Relations**

The GeoEPR 2021 dataset geo-codes all politically relevant ethnic groups from the [EPR-Core 2021](#) dataset. GeoEPR assignes every politically relevant group one of six settlement patterns and, if possible, provides polygons describing their location on a digital map.

## Data

Descriptive (non-spatial) information on ethnic group's settlement patterns from the GeoEPR 2021 dataset is available in research-ready country-year and group-year format from the [GROW<sup>4P</sup> Research Front-End](#) data portal.

You may also download the latest version (2021, released on June 8, 2021) of the complete GeoEPR dataset directly:

Shapefile format (Zip archive): [GeoEPR-2021.zip](#)  
CSV format with WKT geometries, UTF-8 charset: [GeoEPR-2021.csv](#)  
Tab-delimited format with WKT geometries, UTF-8 charset: [GeoEPR-2021.txt](#)  
JSON (GeoJSON) format: [GeoEPR-2021.geojson](#)  
SQL: [GeoEPR-2021.sql](#)  
Codebook: [EPR\\_2021\\_Codebook\\_GeoEPR.pdf](#)

The GeoEPR polygons follow the WGS84 coordinate system (EPSG:4326)

Here is the full list of data sources and links:

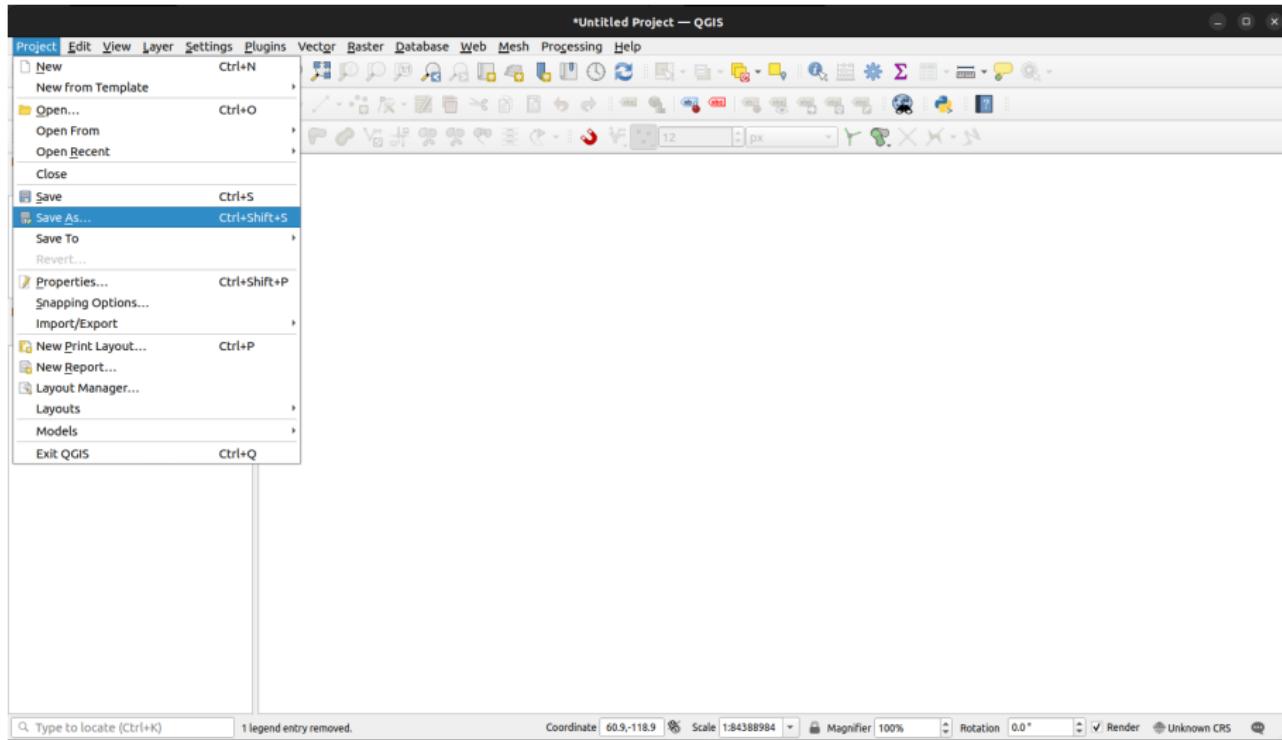
Category	Type	Format	Data source
ISIS violence	Table (non-geo)	.csv	UCDP GED
Administrative units	Vector (polygons)	.geojson	GADM
Roads	Vector (polylines)	.shp	DIVA-GIS
Population	Raster	.tif	WorldPop
Cropland	Raster	.tif	EarthStat
Dams	Vector (points)	.shp	GRanD
Sectarian divisions	Vector (polygons)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

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

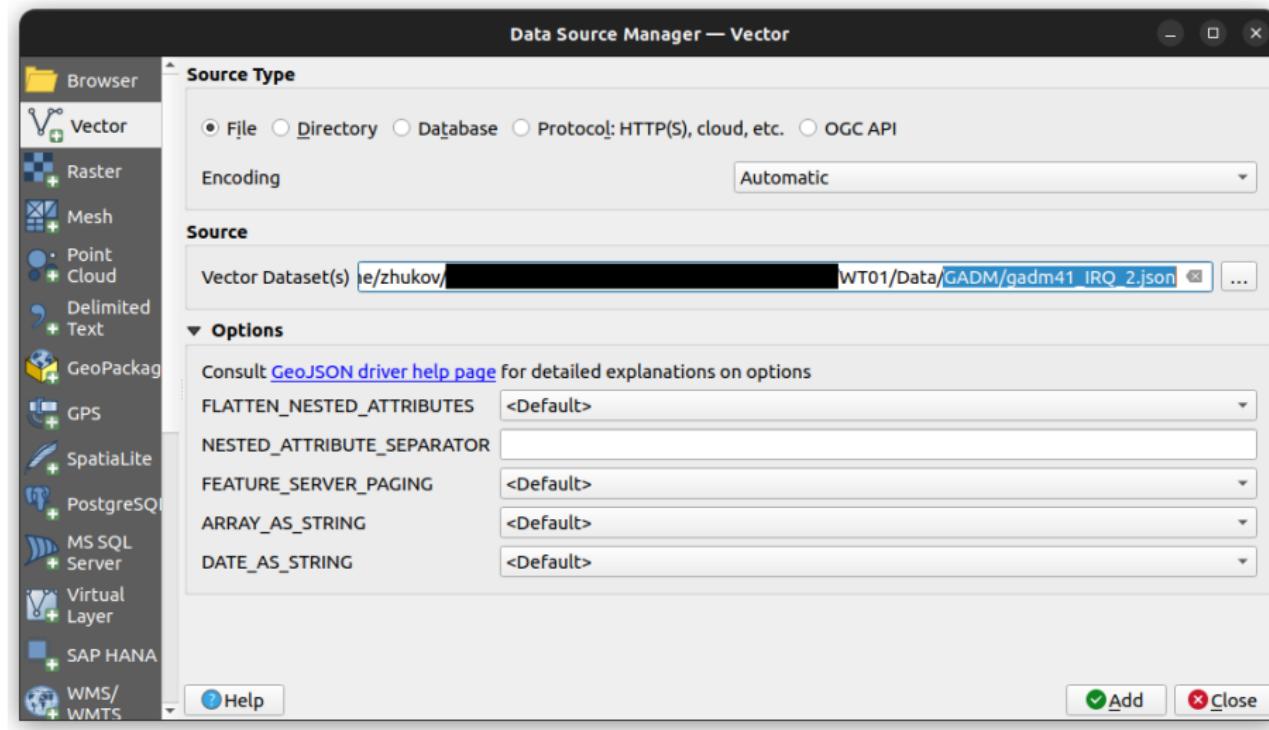
# Pre-processing

# Always save your progress!

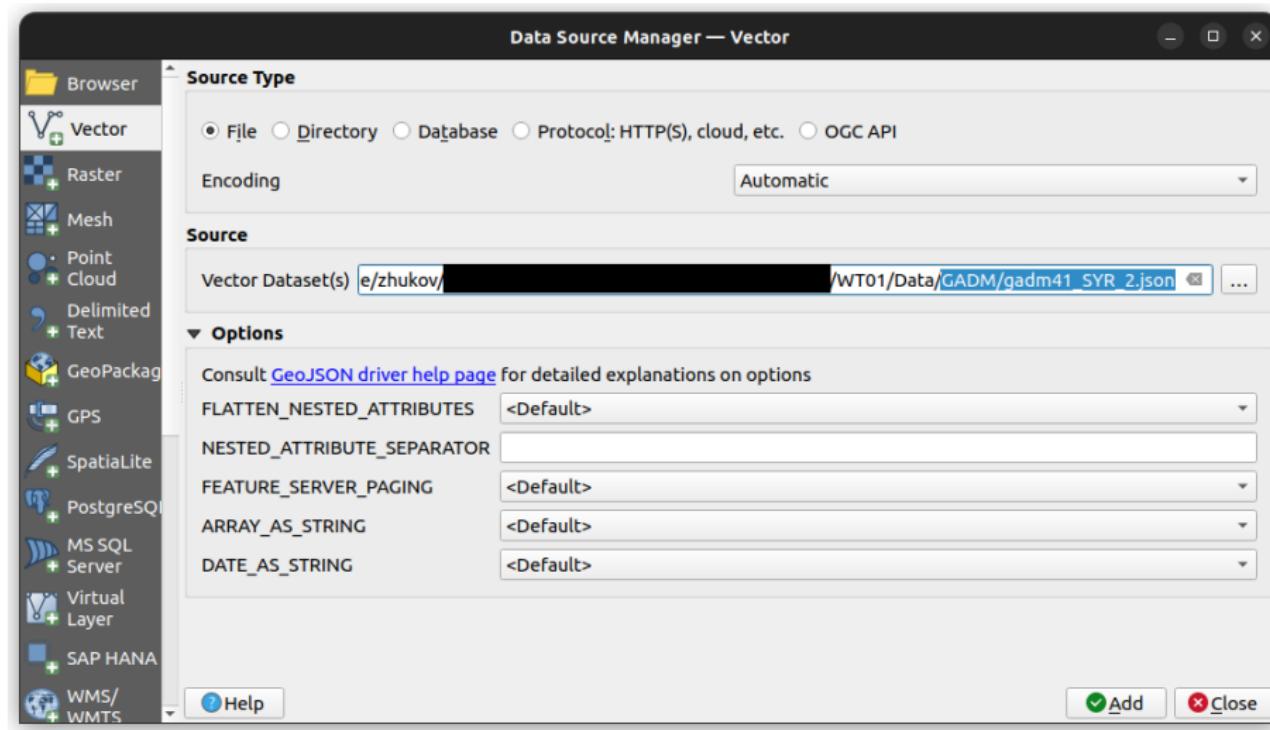
Go to Project → Save As...



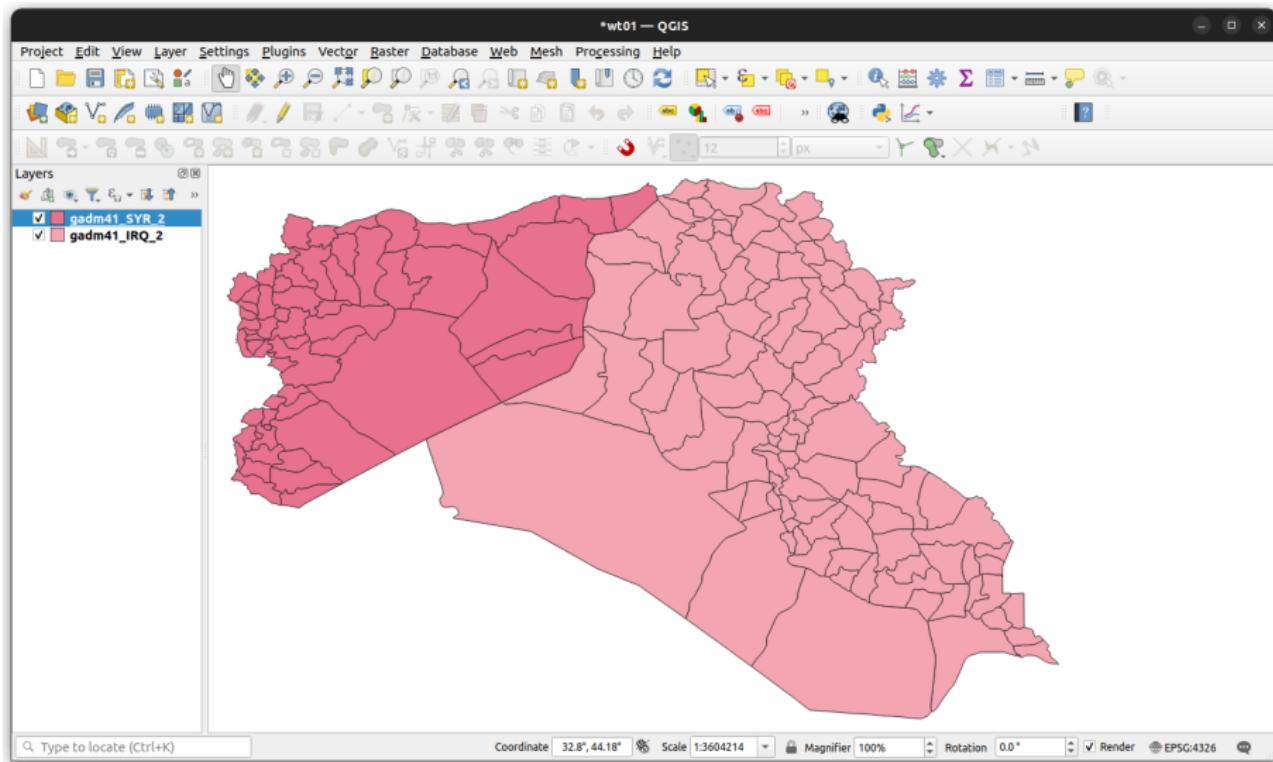
Load *administrative boundaries* (Layer → Add Layer → Add Vector Layer).  
For **Iraq**: gadm41 IRQ\_2.json file in Data/GADM folder.



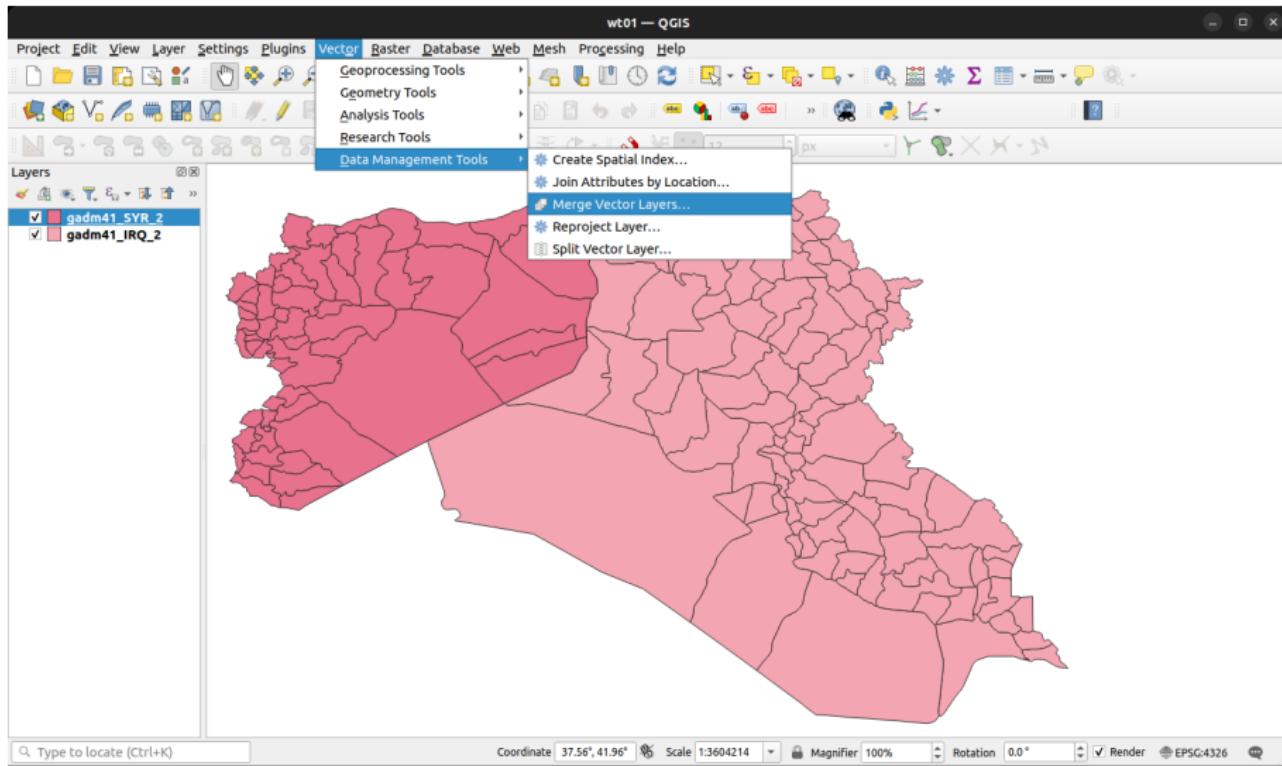
For Syria: gadm41\_SYR\_2.json file in Data/GADM folder.



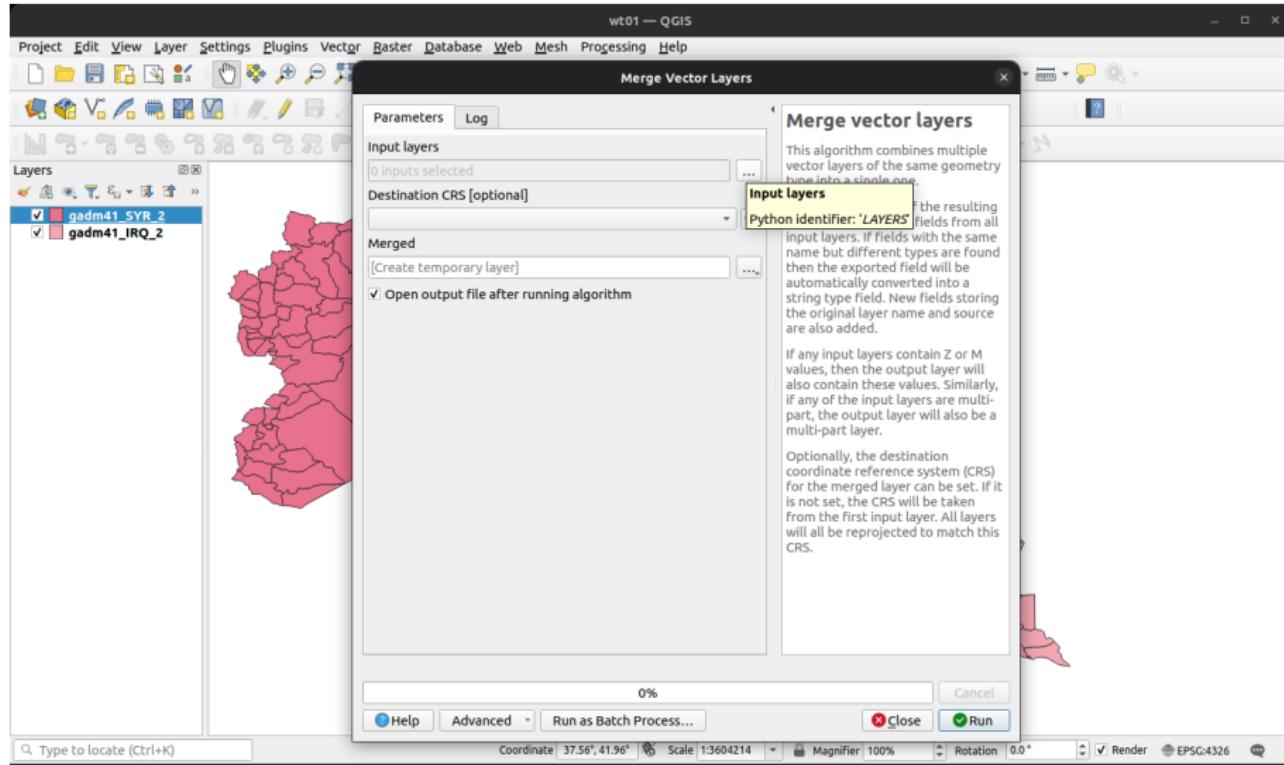
The two sets of district boundaries should be visible in the project window.  
Let's **merge them into a single layer**, with districts for both countries.



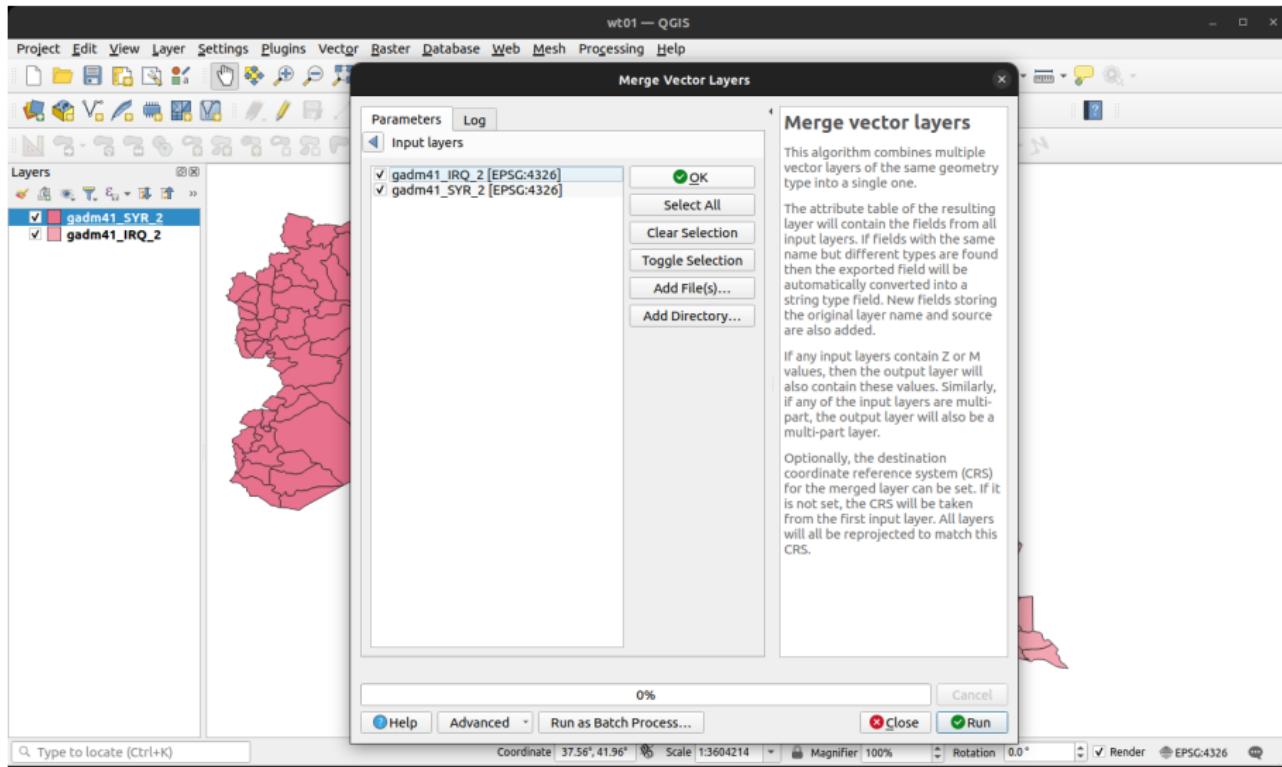
Open the Merge Vector Layers tool (Vector → Data Management Tool → Merge Vector Layers).



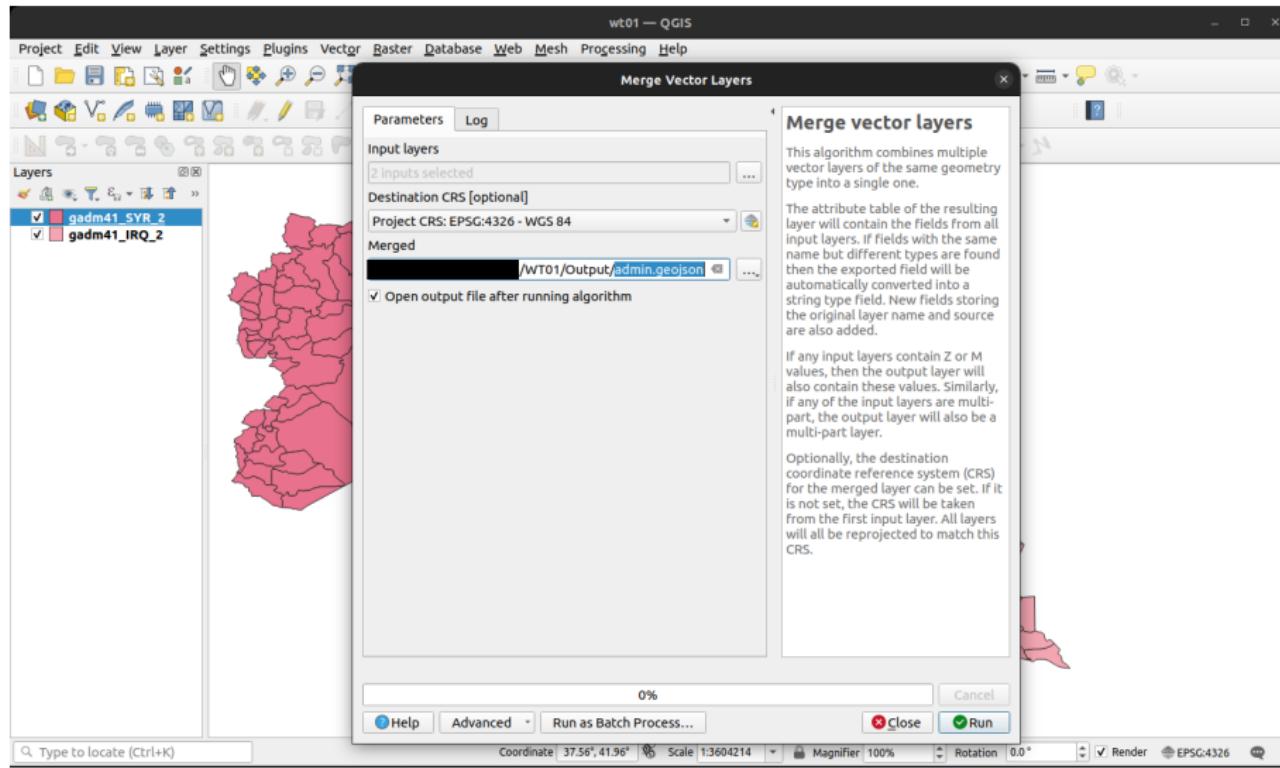
Click on the [...] box next to Input layers



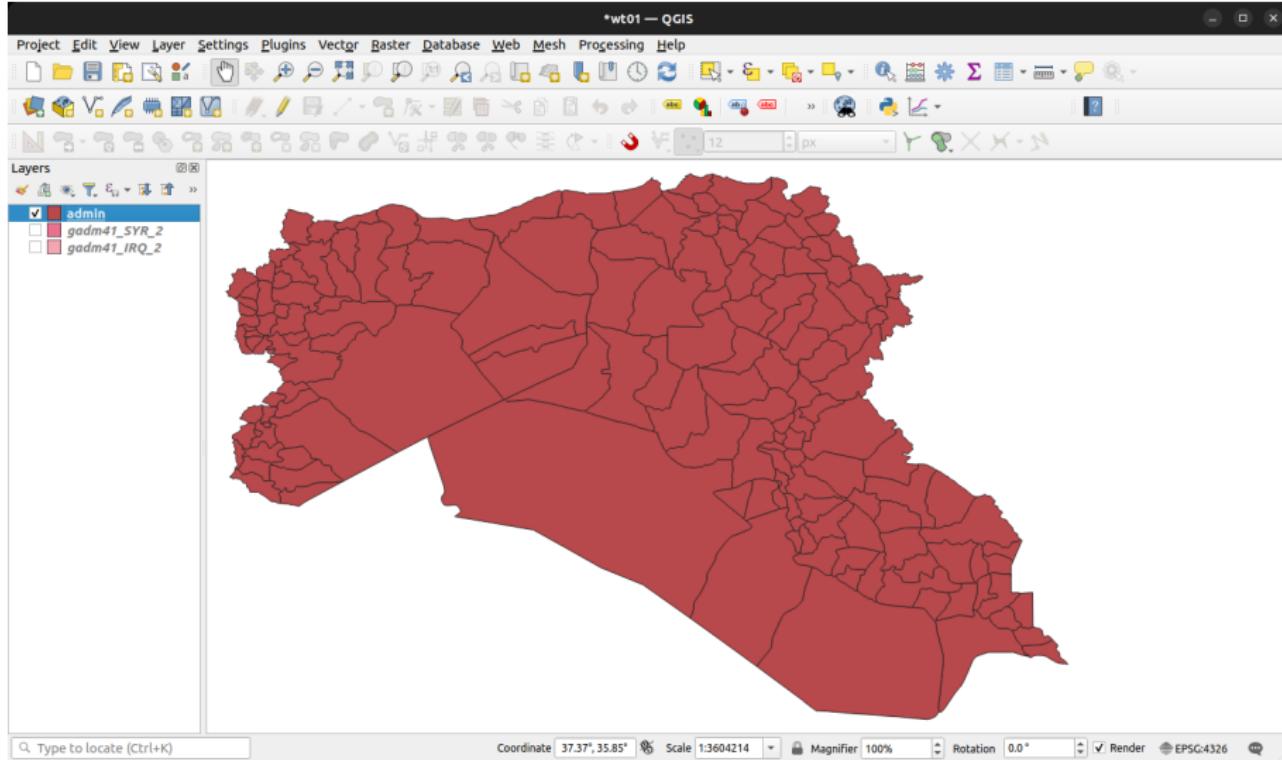
Check the boxes next to `gadm41 IRQ_2` and `gadm41 SYR_2`. Click OK.



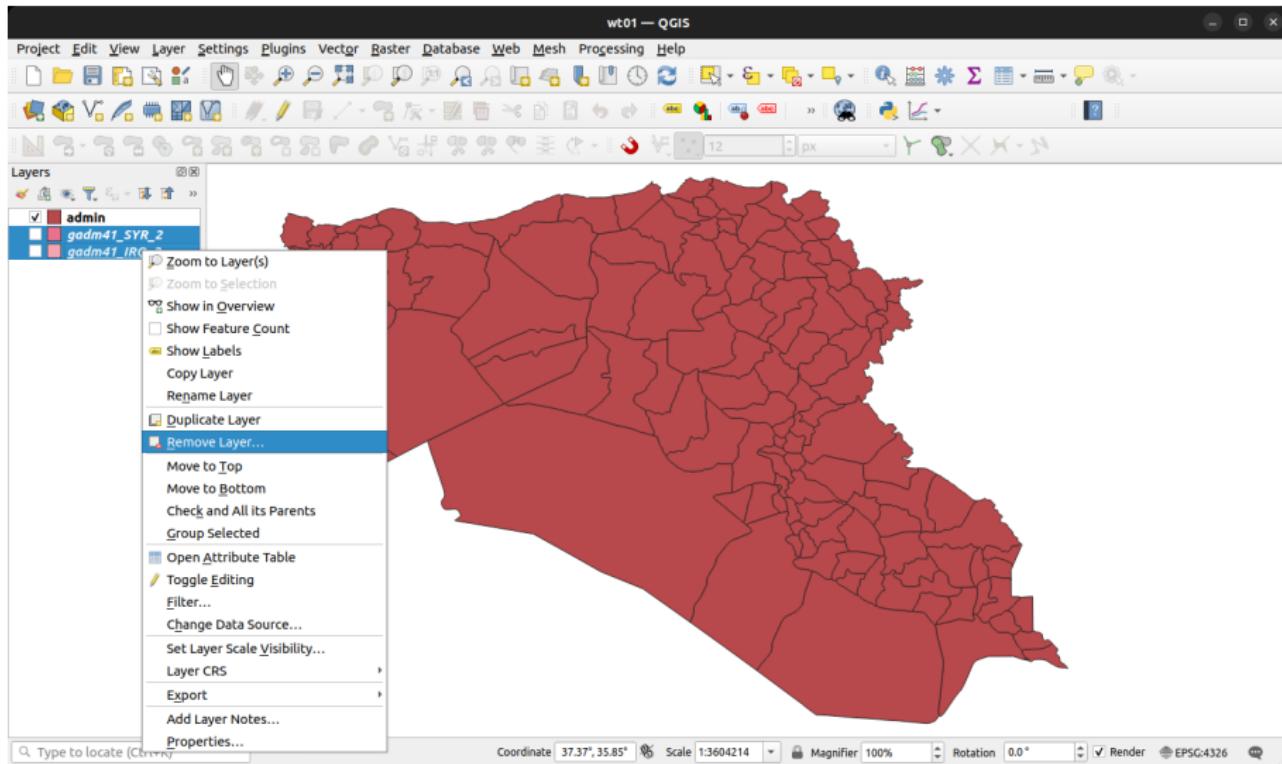
Click on the [...] box next to Merged. Find a location in which to save the output, and name the file admin.geojson. Click Run



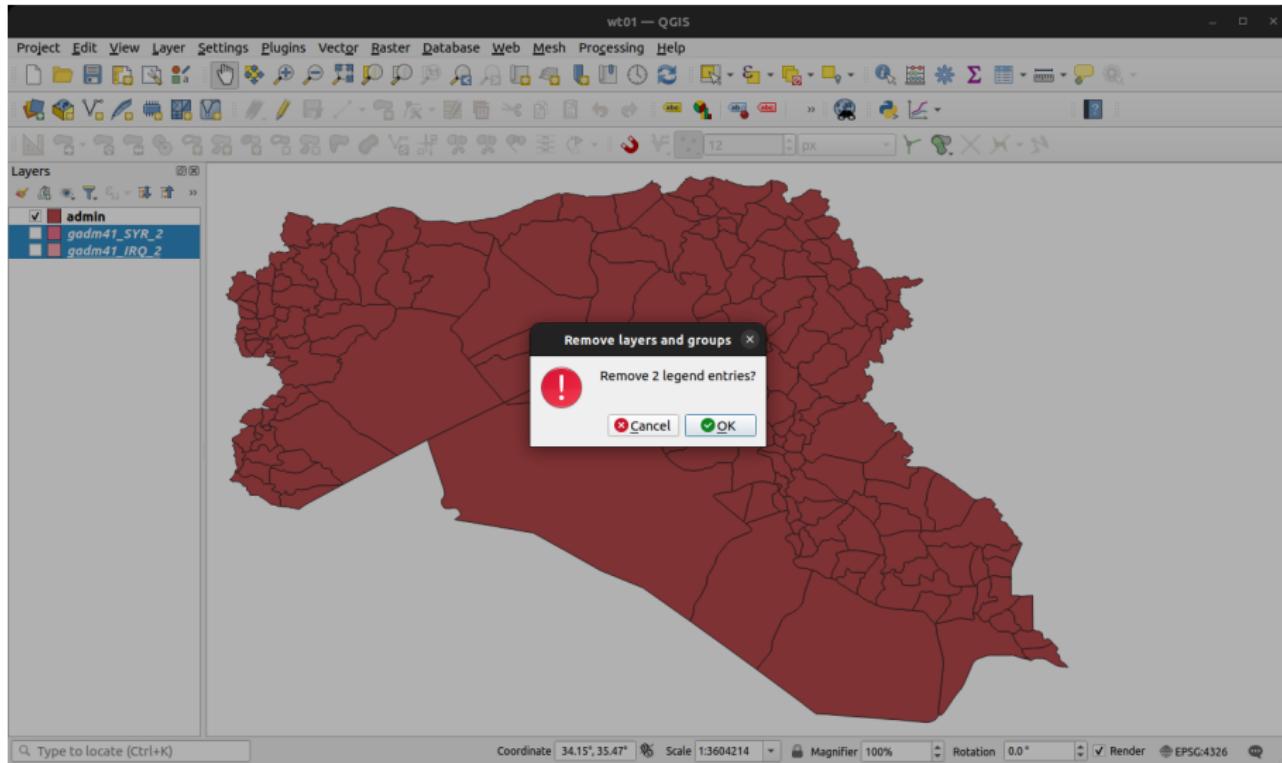
The merged layer should appear in the main project window.



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...)

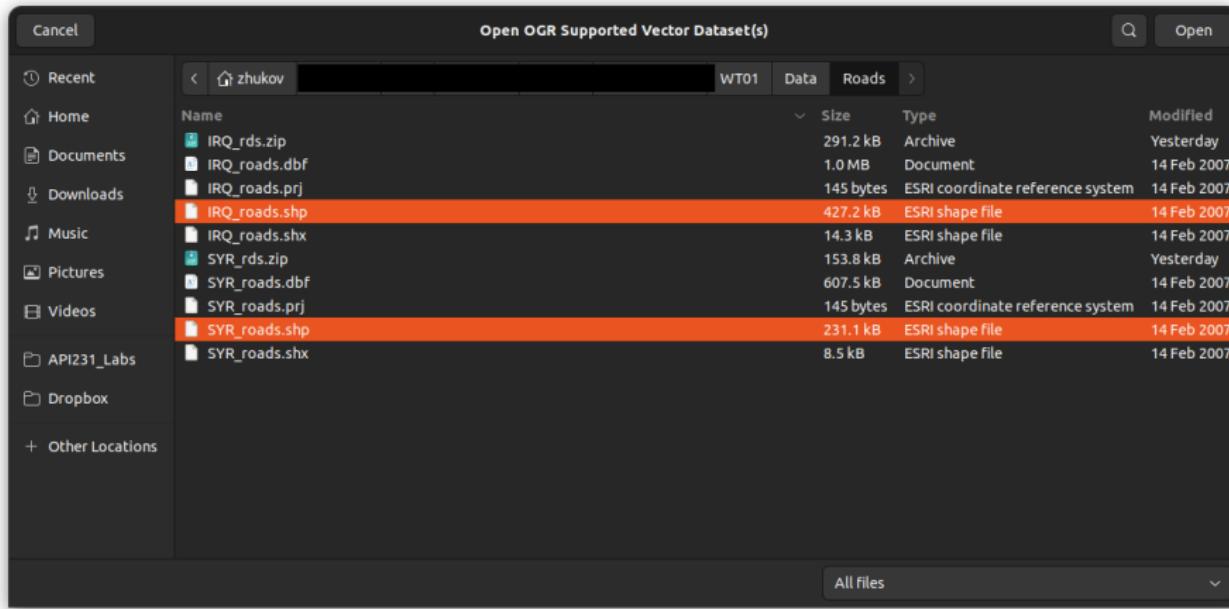


To conserve memory, it's good practice to remove data objects were are not actively using

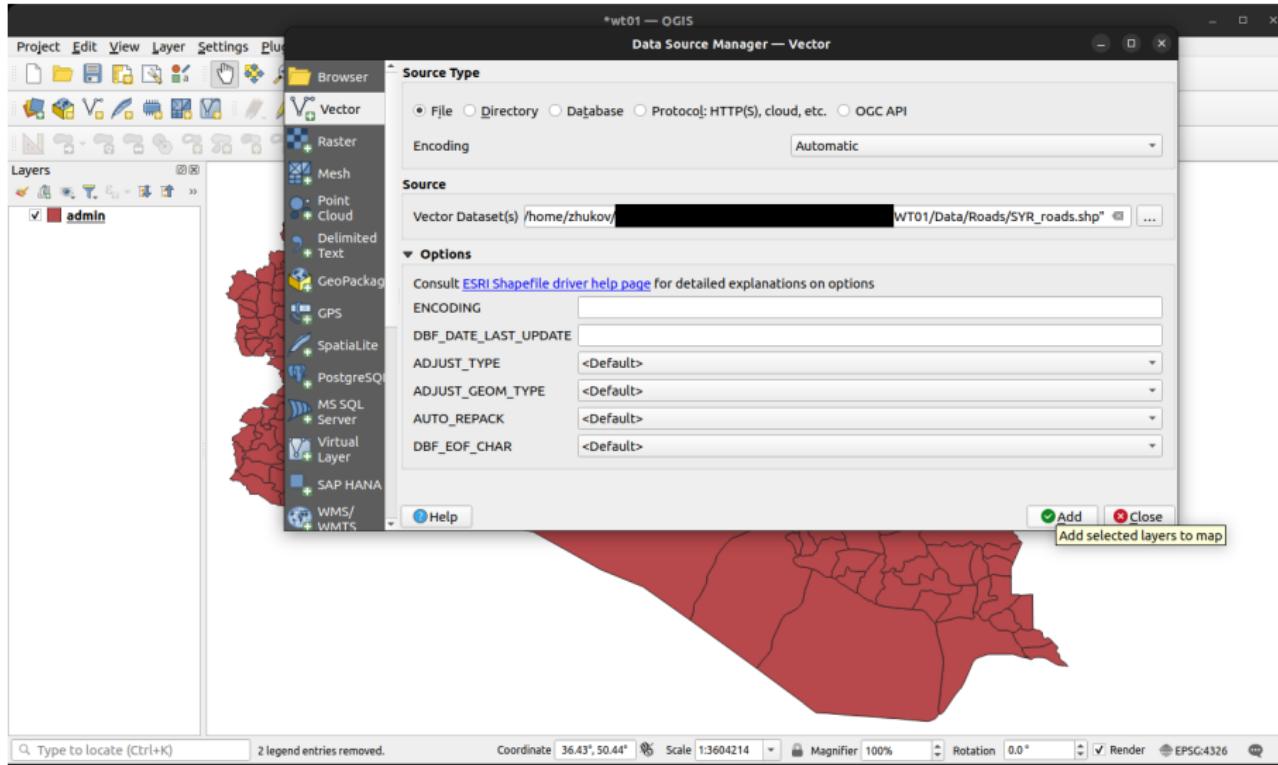


Load *roads data* (Layer → Add Layer → Add Vector Layer).

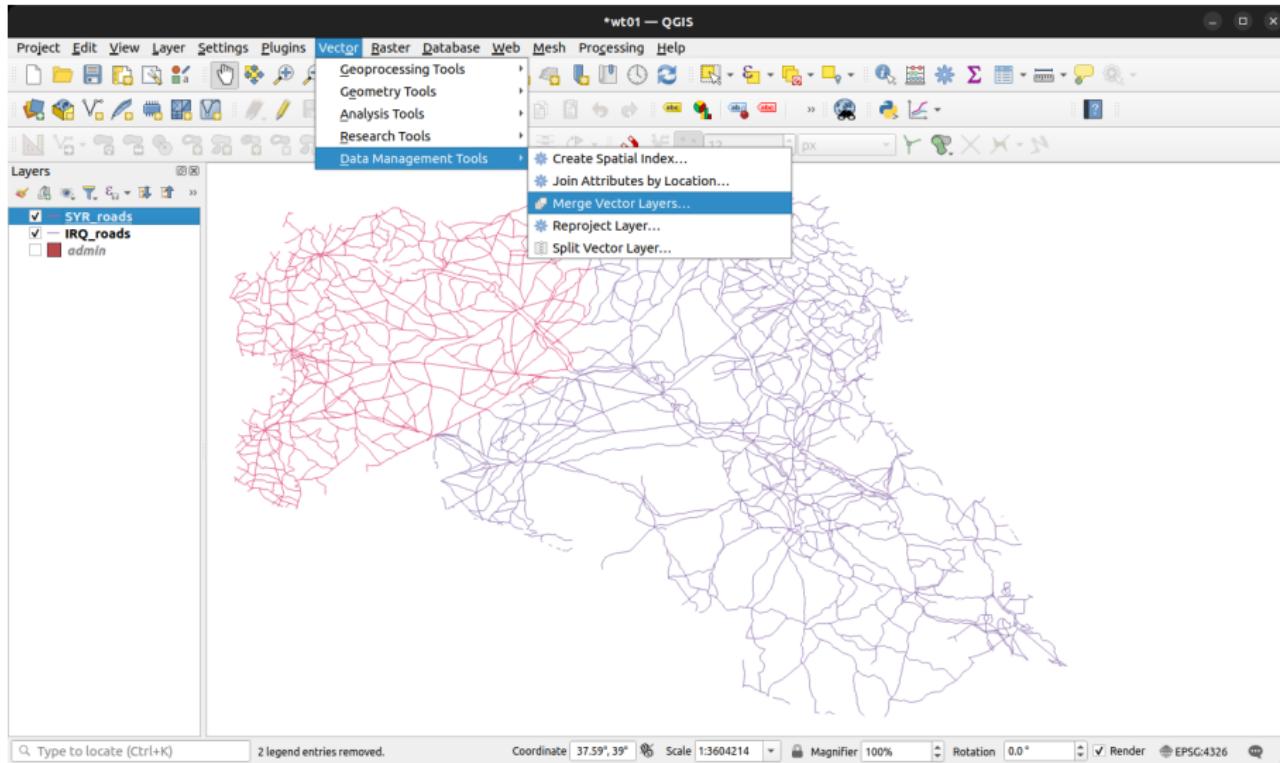
Try opening **two datasets at the same time**: IRQ\_roads.shp and SYR\_roads.shp files in Data/Roads



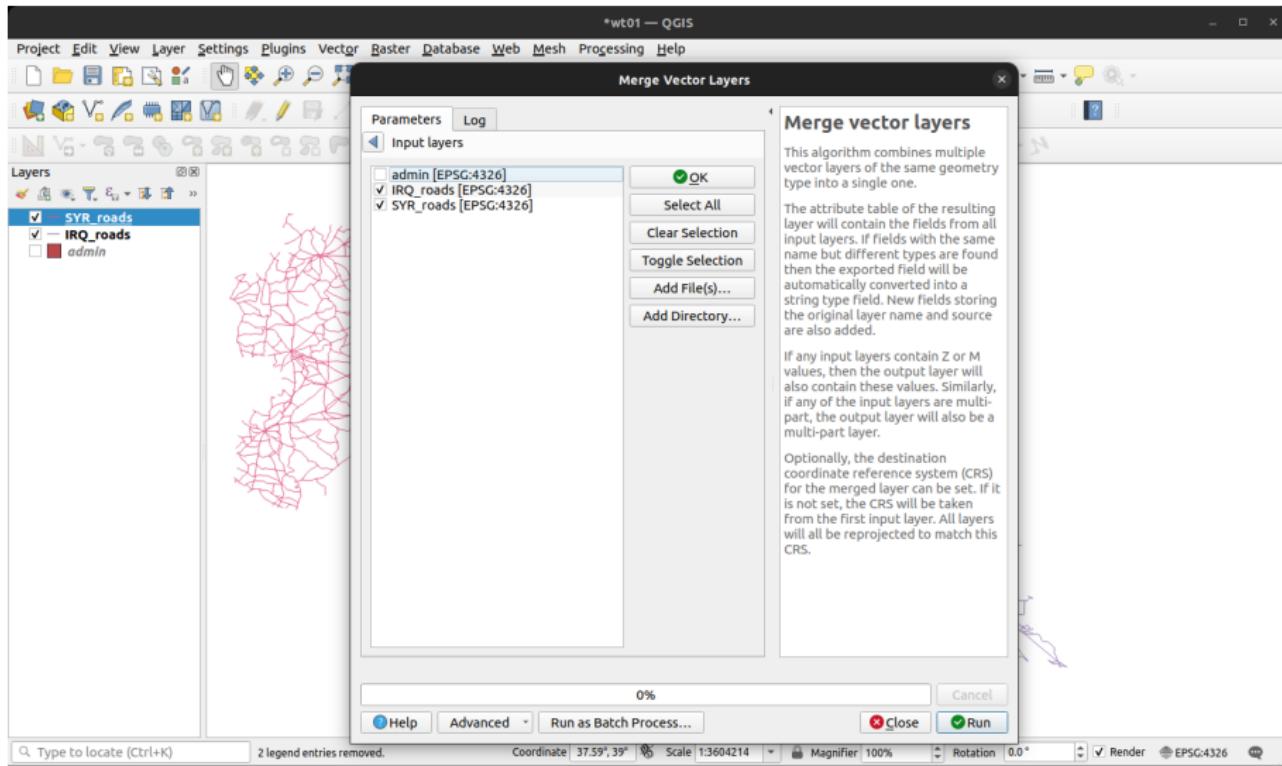
Paths to both files should appear in the Source box. Click Add



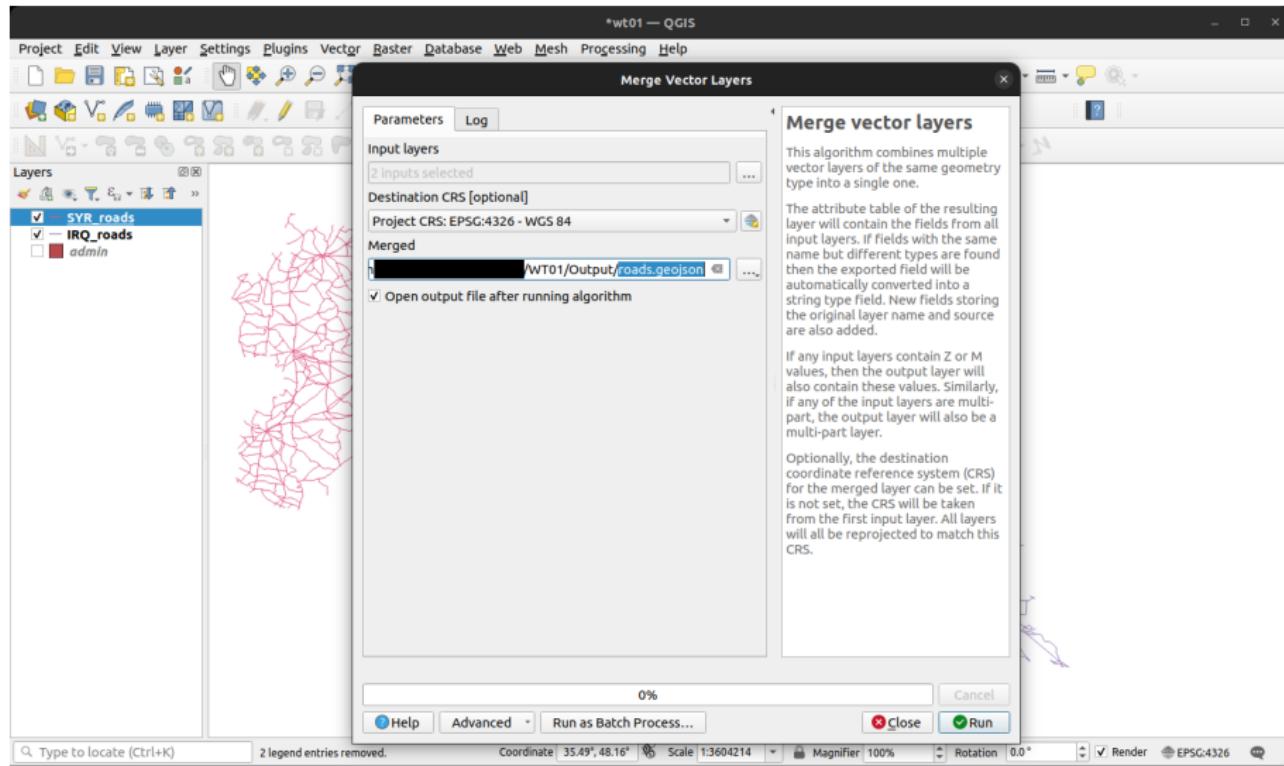
Let's repeat the Merge Vector Layers procedure for these two layers (Vector → Data Management Tool → Merge Vector Layers).



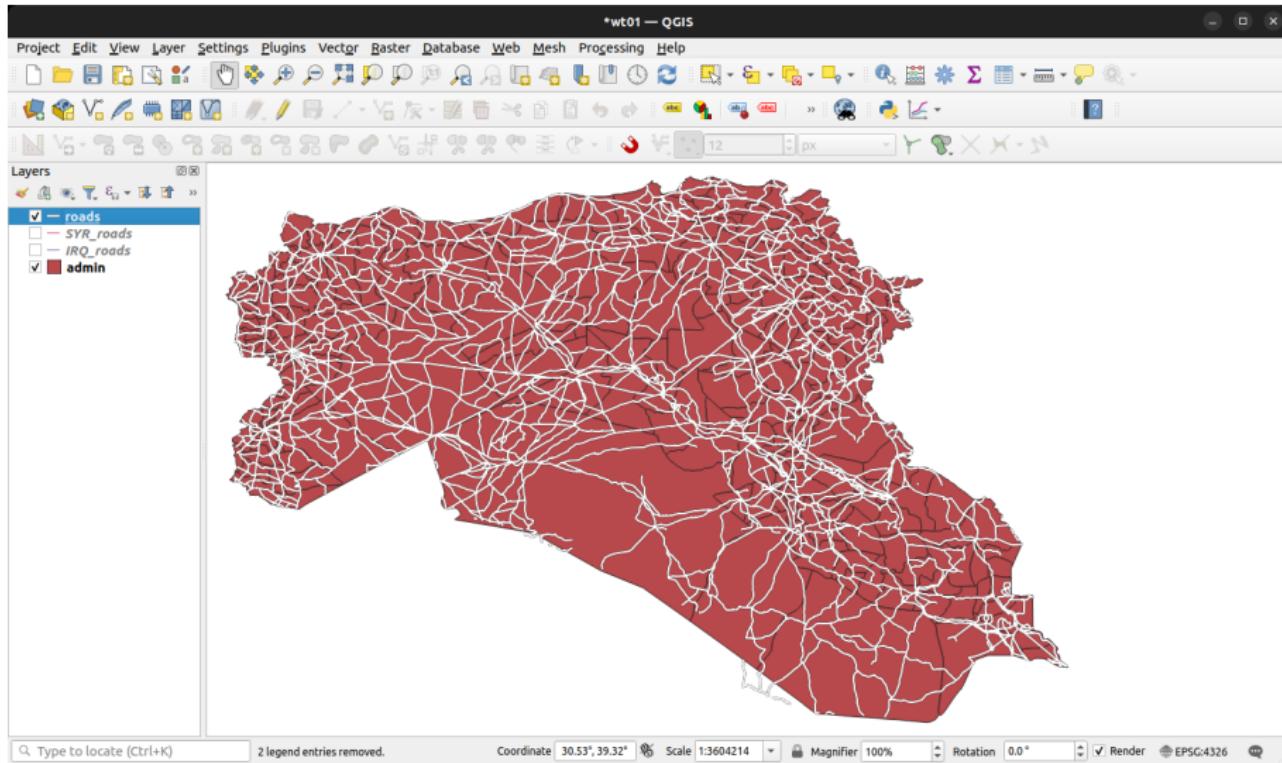
This time, perform the merge on the IRQ\_roads and SYR\_roads layers



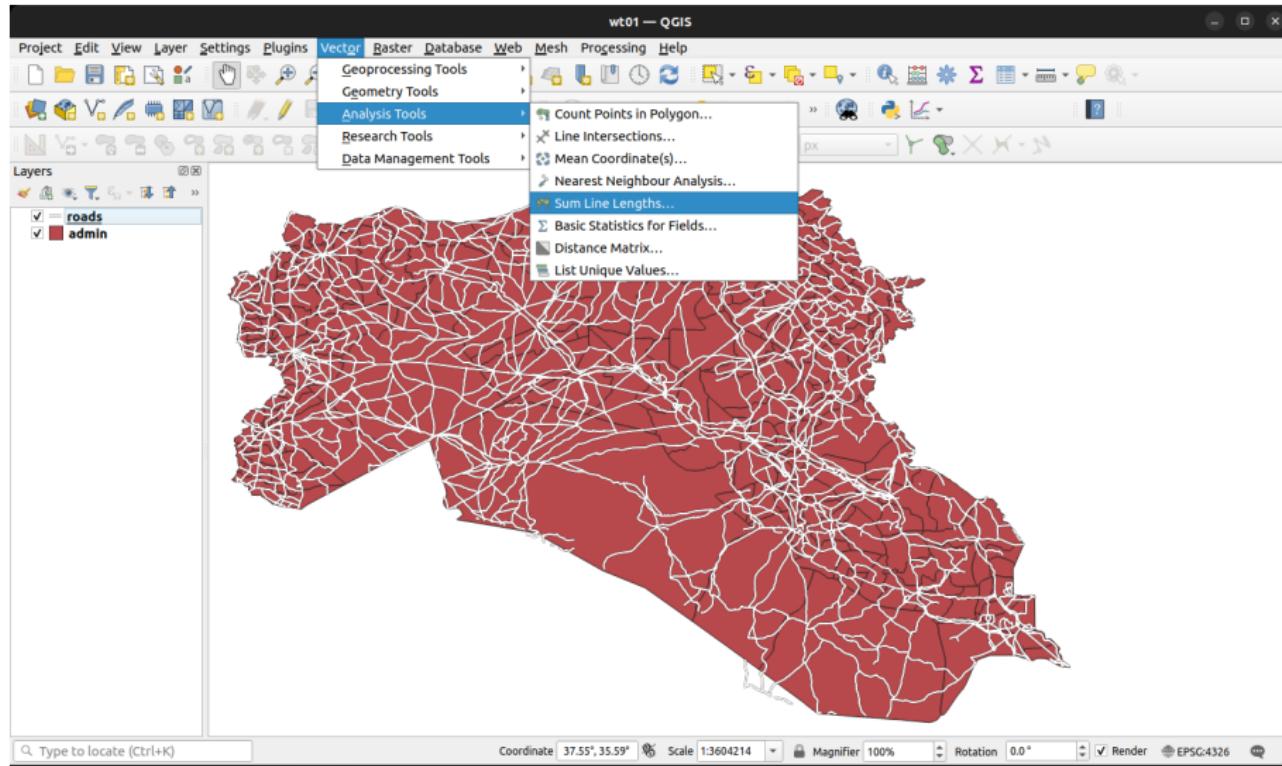
Save the output file as roads.geojson



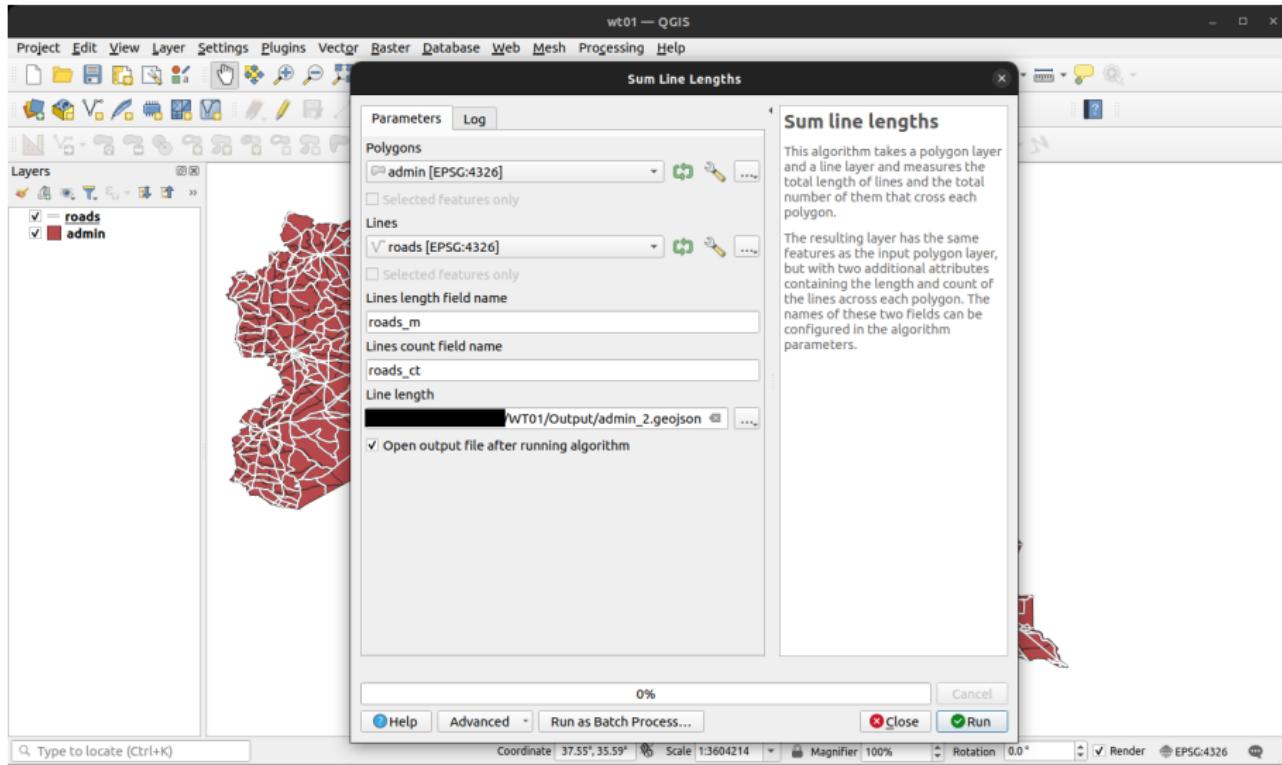
As before, you can remove the two country-specific roads layers (we no longer need them)



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



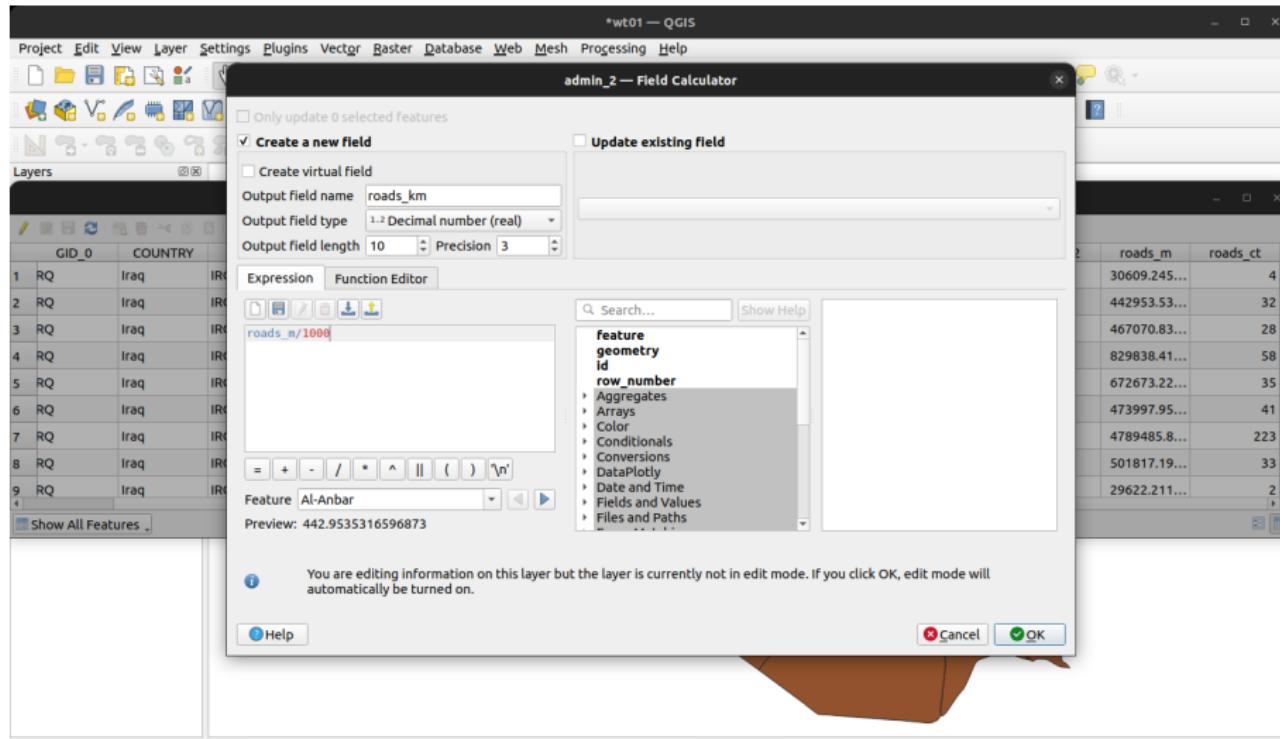
Open the attribute table for admin\_2. The roads\_m and roads\_ct variables should be in the table. Let's convert from **meters to kilometers**

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

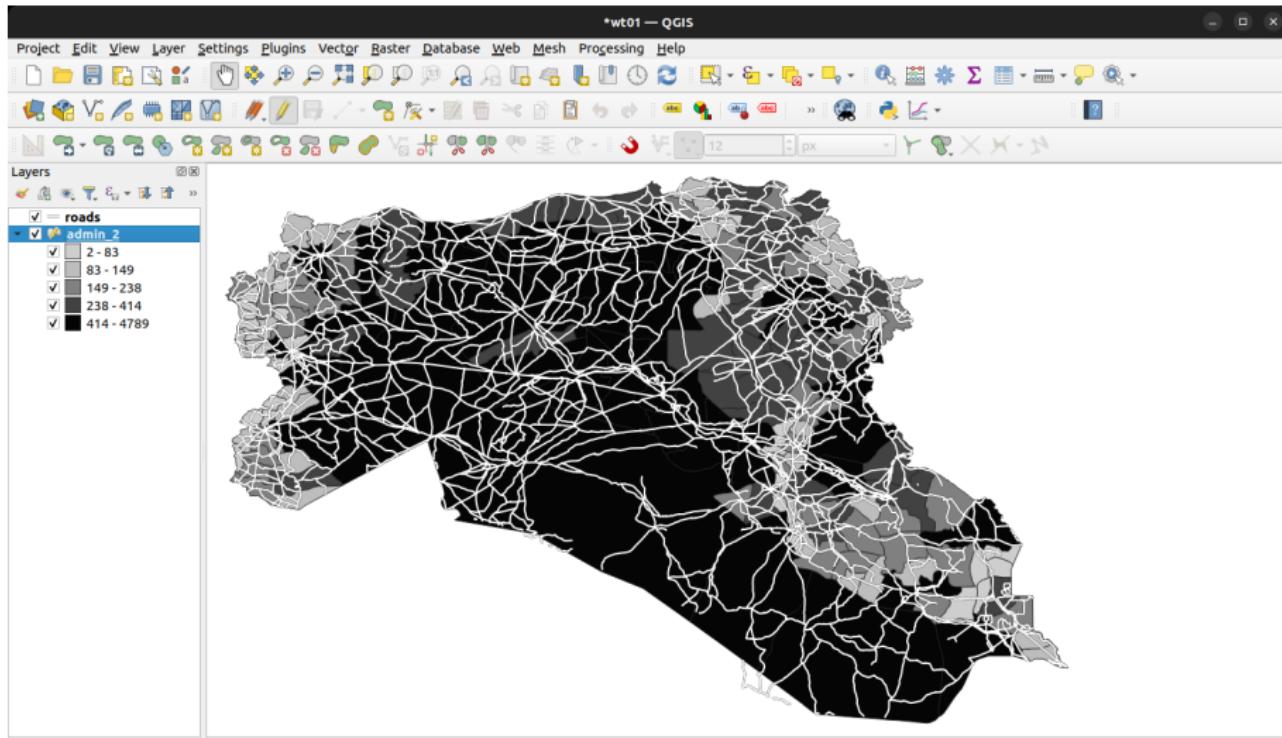


The resulting `roads_km` field should be visible in the attribute table

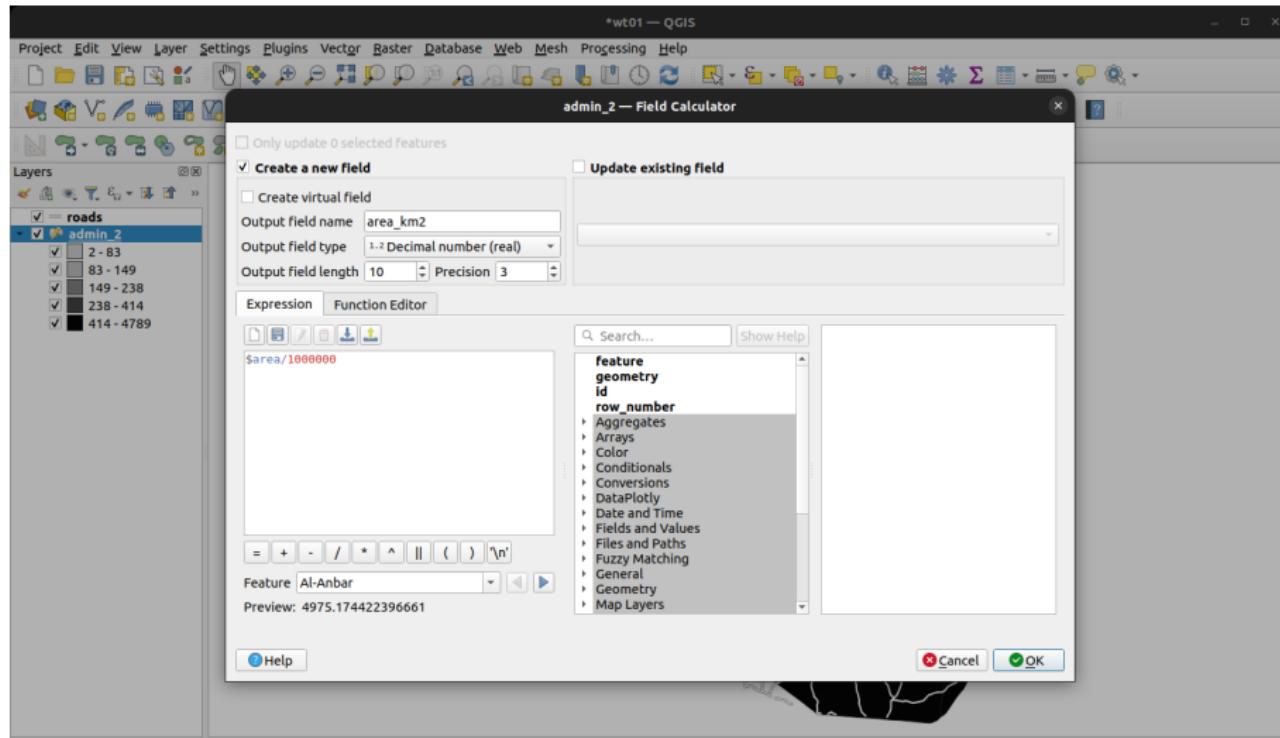
The screenshot shows the QGIS application interface. The title bar reads "#wt01 — QGIS". The menu bar includes Project, Edit, View, Layer, Settings, Plugins, Vector, Raster, Database, Web, Mesh, Processing, and Help. The toolbar contains various icons for file operations, selection, measurement, and 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

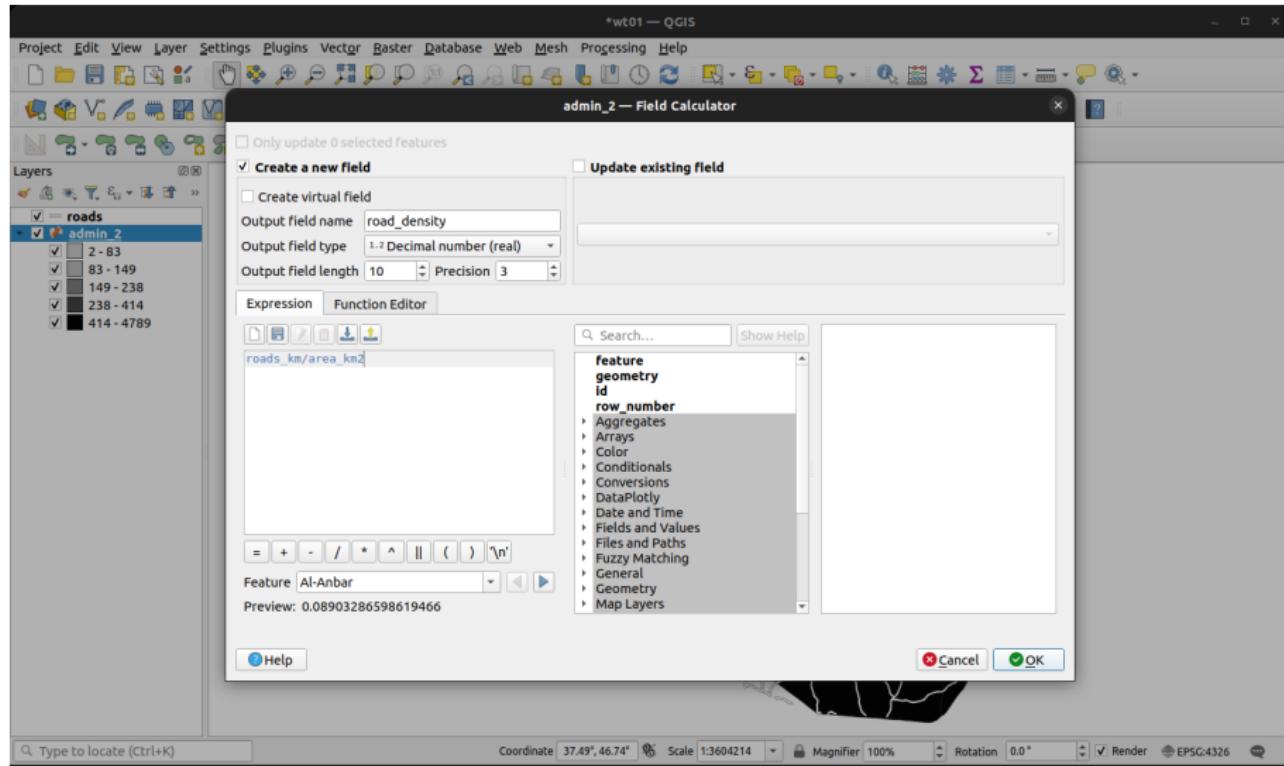
If we plot this variable, we can see its main limitations: smaller districts have less road length, so this becomes a proxy for district size. Let's create a **road density** measure that takes into account district size



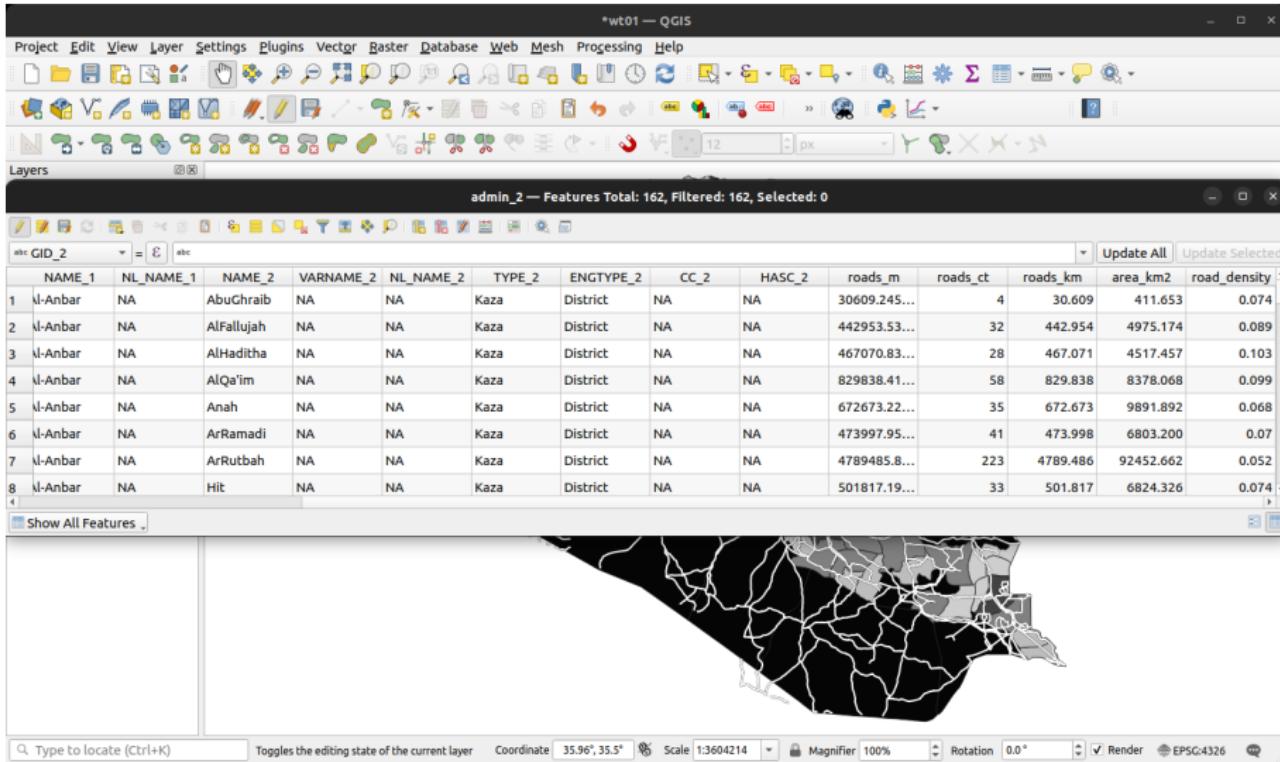
Back in the Field Calculator, create a new field called `area_km2` of type Decimal number (real). For Expression, write `$area/1000000` (divide by 1M to convert from m<sup>2</sup> to km<sup>2</sup>). Click OK



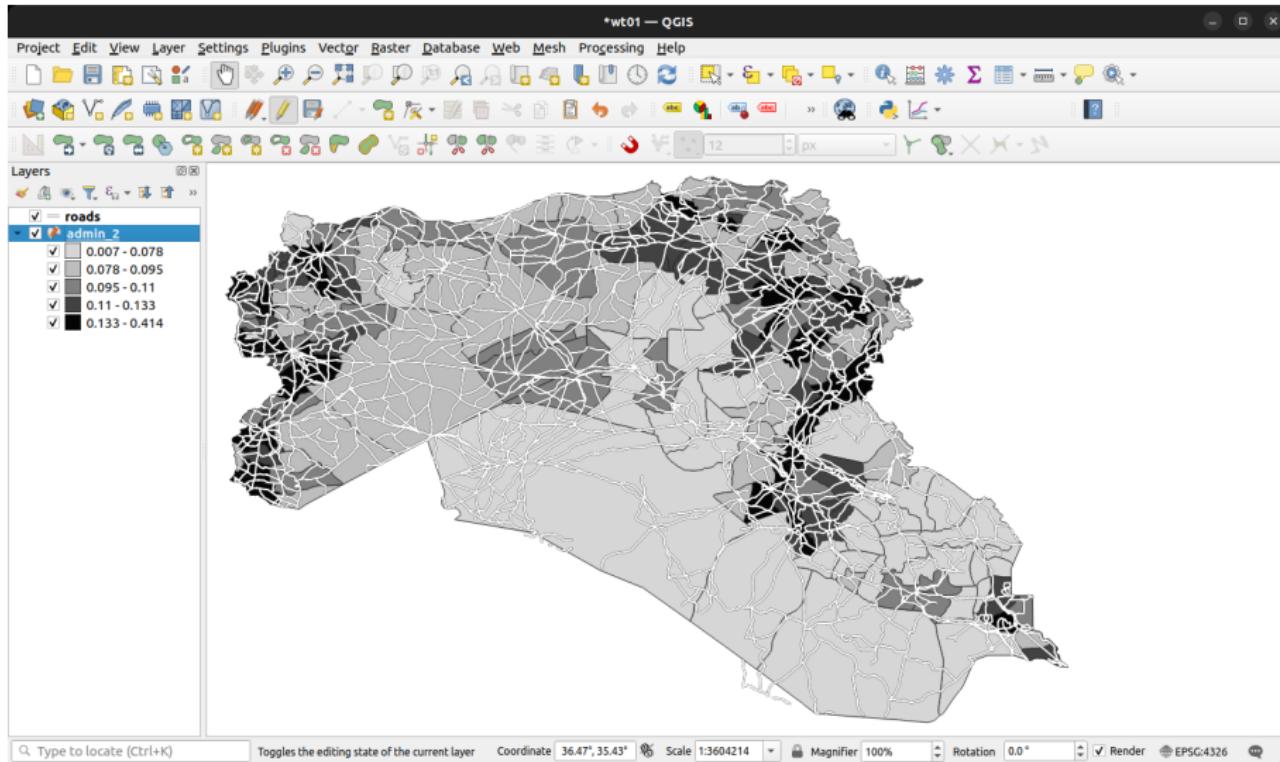
Open the Field Calculator again. Create a new field called road\_density of type Decimal number (real). For Expression, write roads\_km/area\_km2



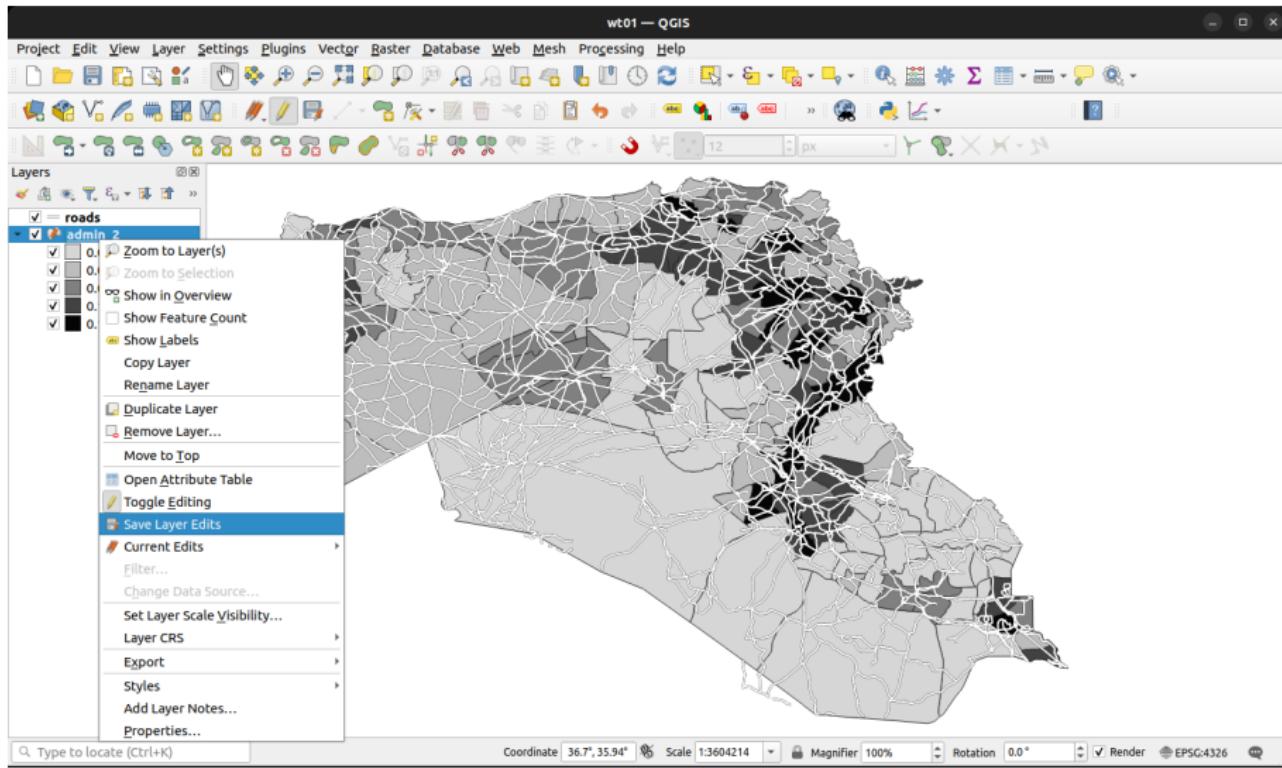
Check to make sure road\_density is added to attribute table



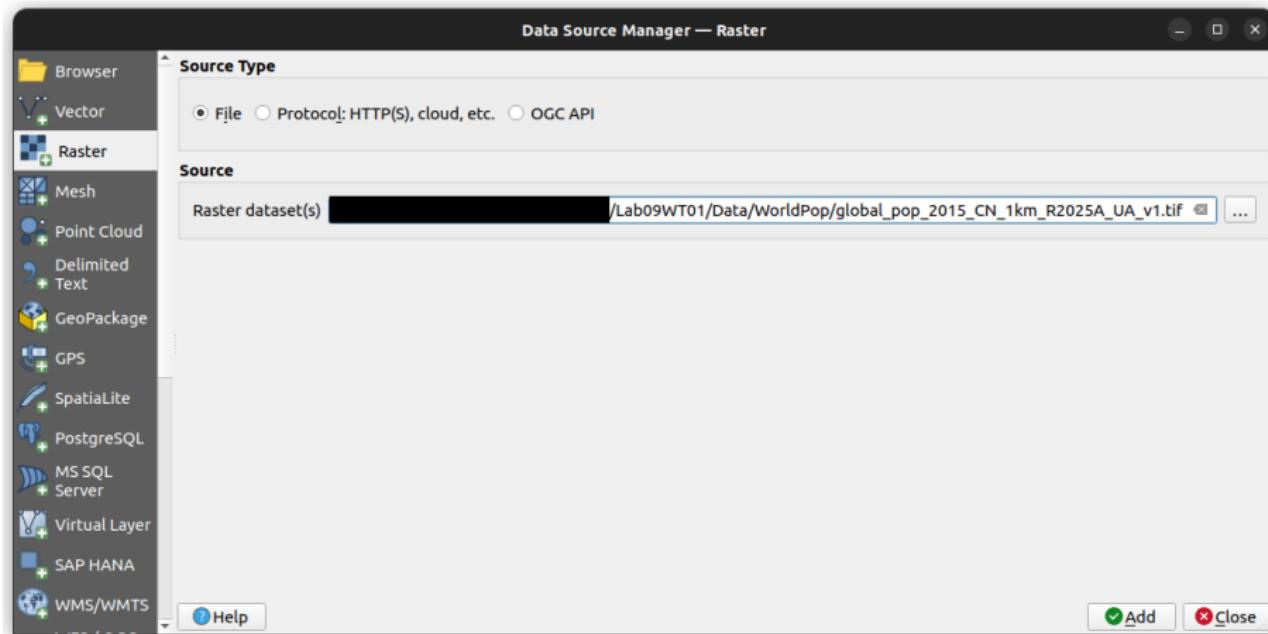
If we plot this variable, we see the opposite pattern from before (smaller, urban districts have higher road density)



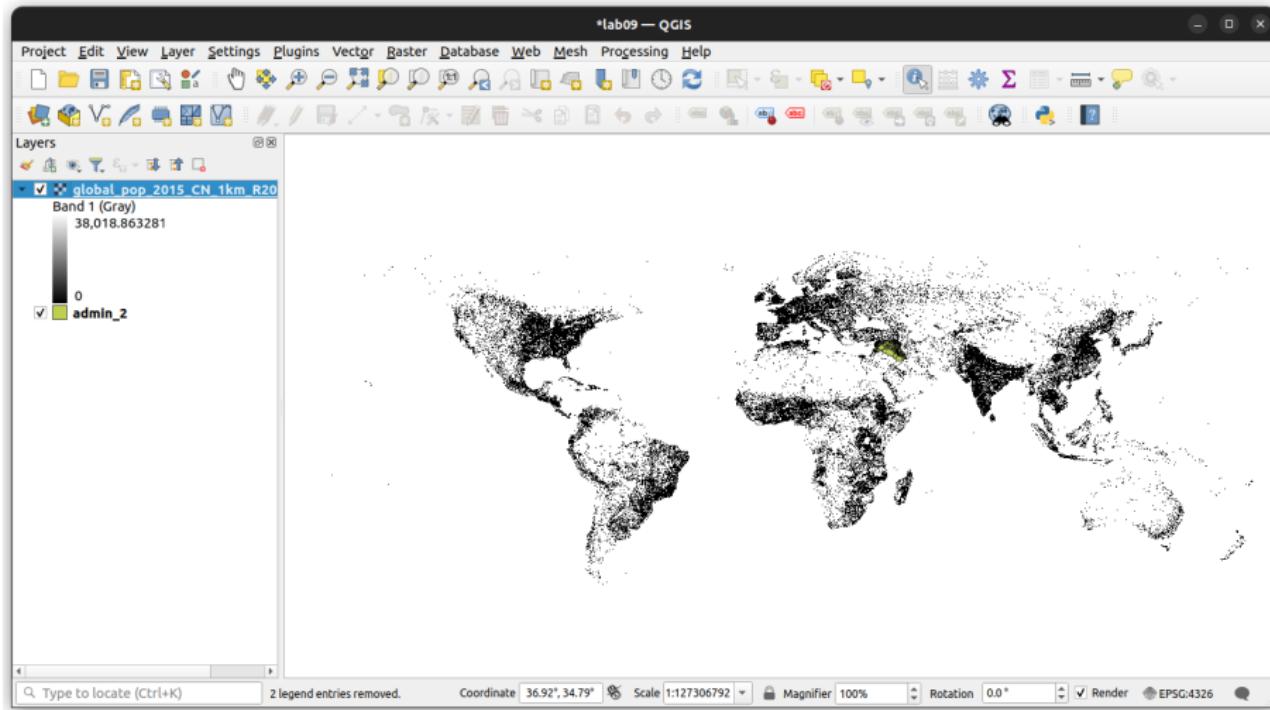
Save the edits you just made to the admin\_2 layer (right click → Save Layer Edits)



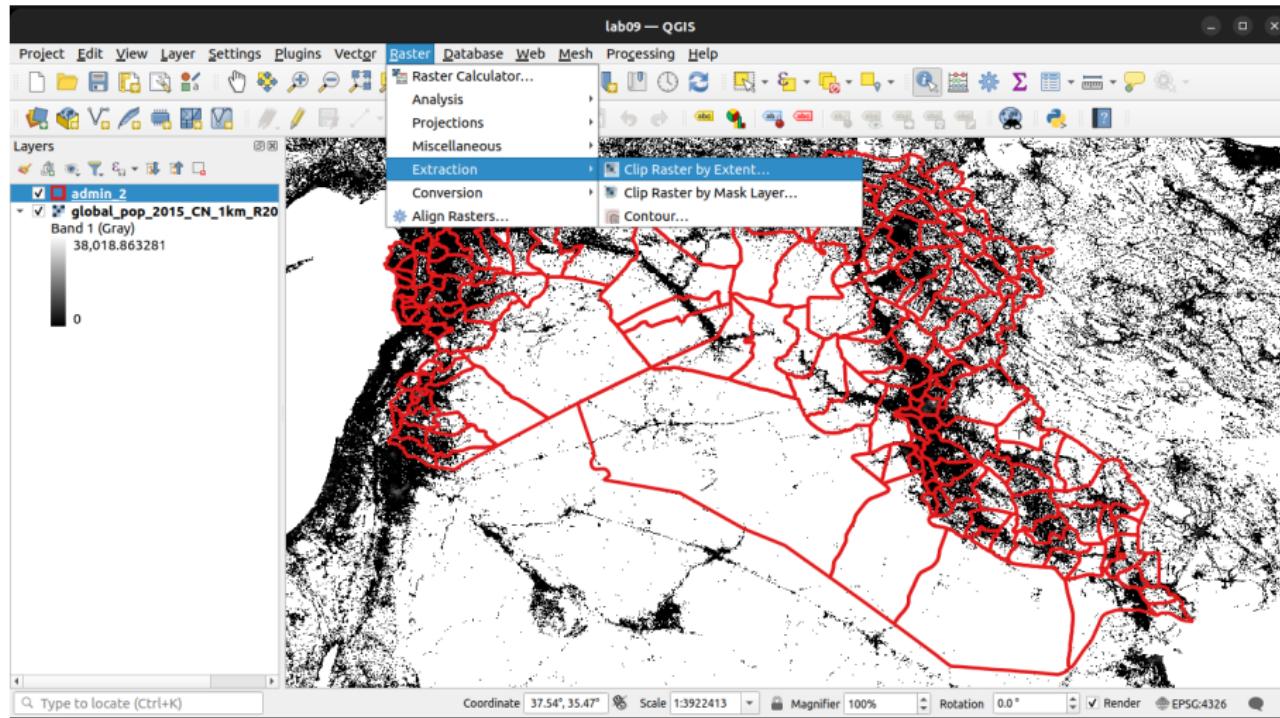
Load *population* raster (Layer → Add Layer → Add Raster Layer).\\ Open the global\_pop\_2015\_CN\_1km\_R2025A\_UA\_v1.tif file in Data/WorldPop folder.



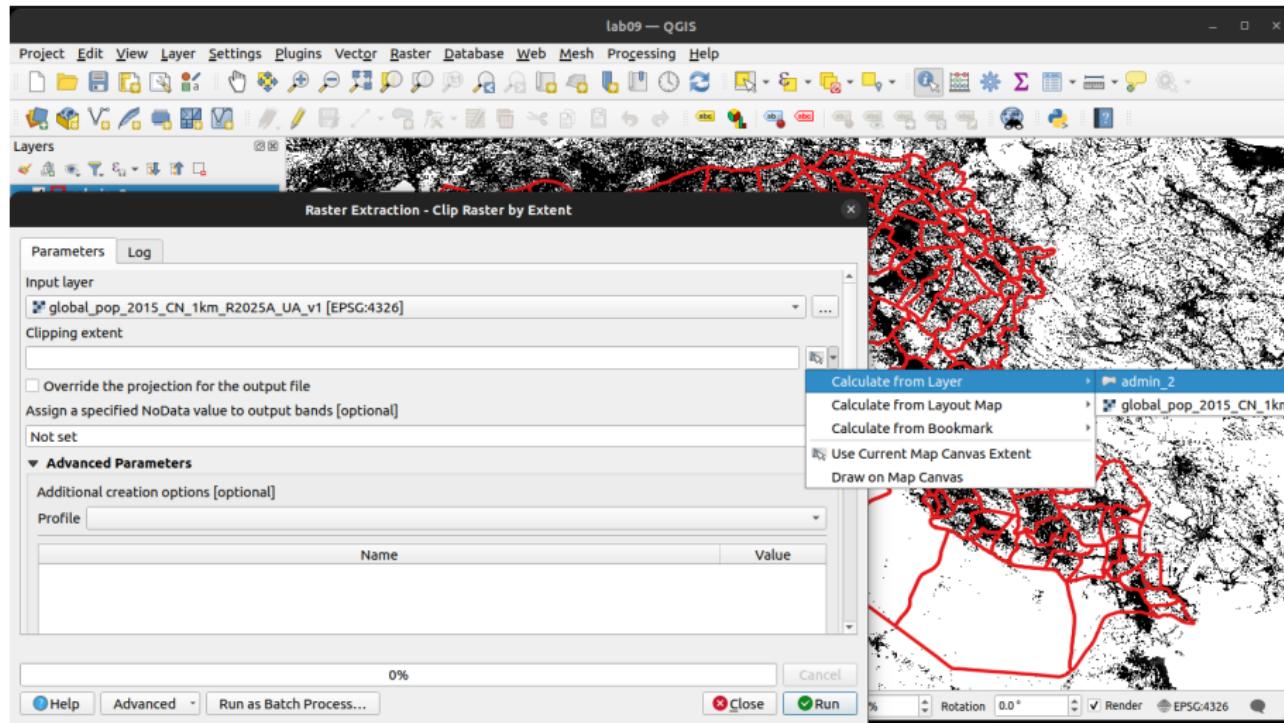
The (global) population layer should now be visible. Let's **extract a subset** of just the part of this layer that overlaps with Syria and Iraq.



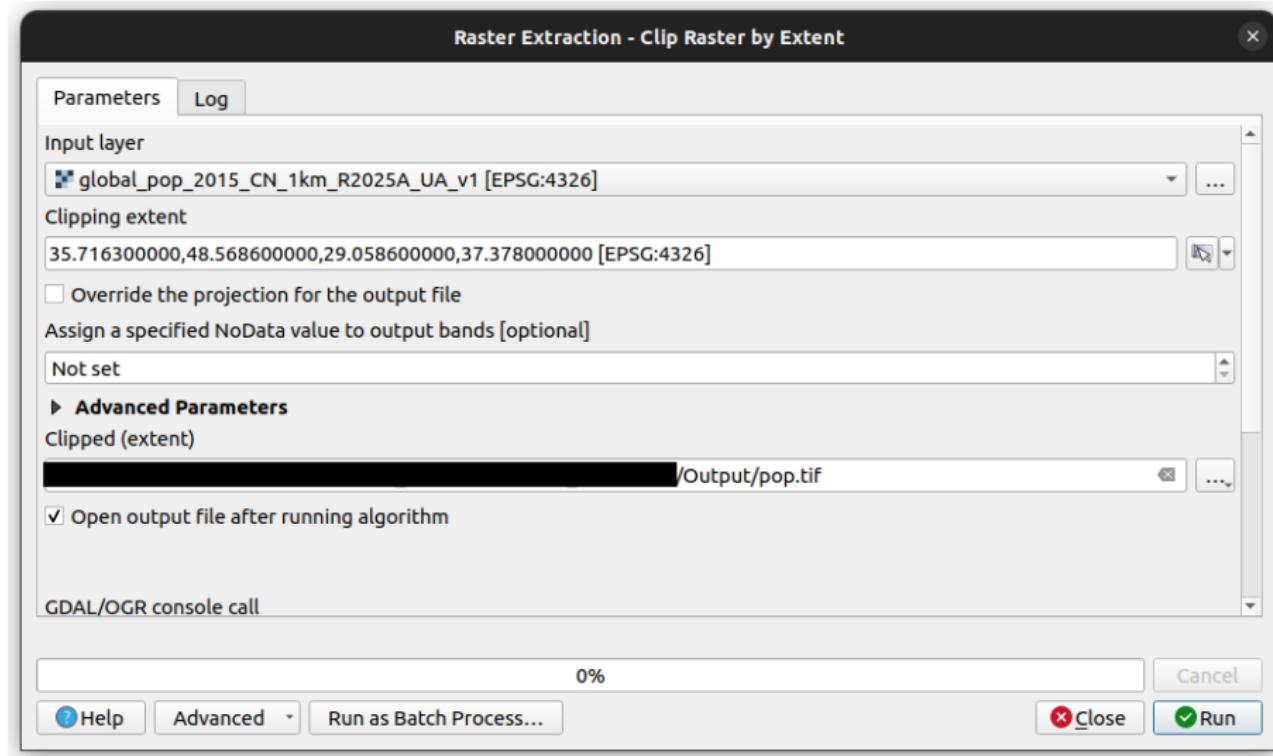
Open the Clip Raster by Extent tool (Raster → Extraction → Clip Raster by Extent)



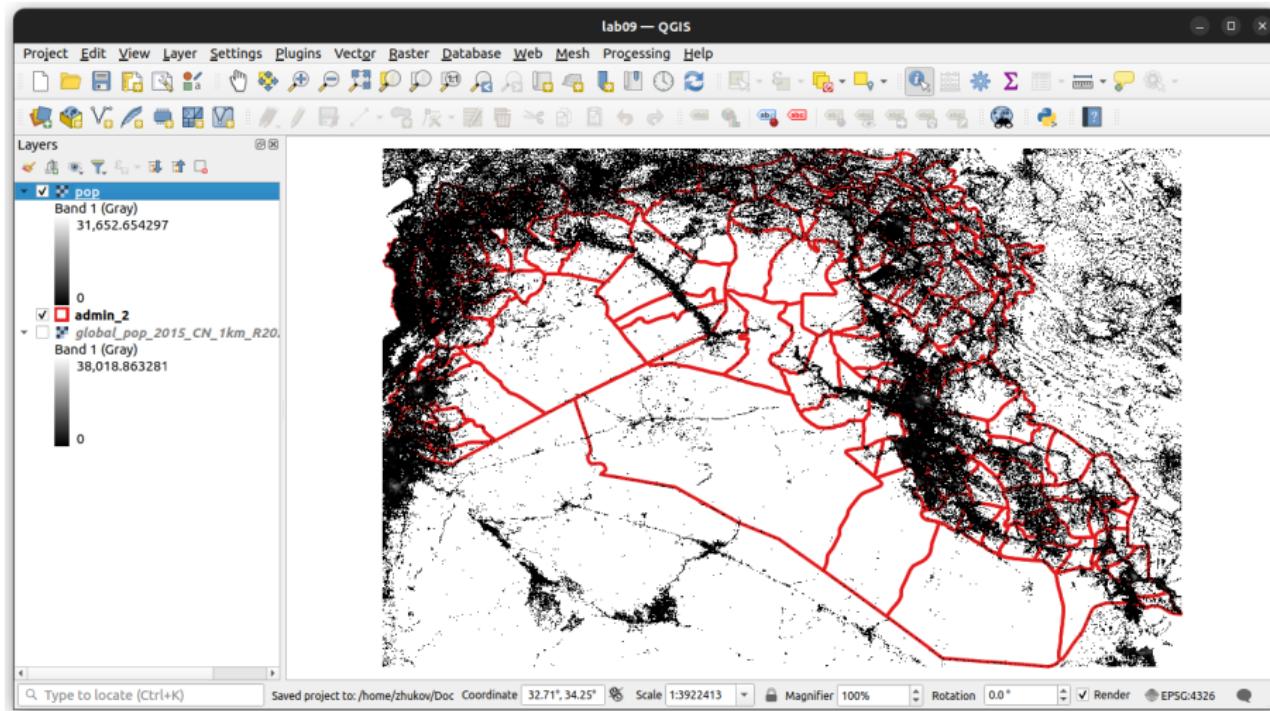
Set Input layer = gpw\_v4\_population\_count\_rev11\_2010\_15\_min. Click the [...] button next to Clipping extent → Calculate from Layer → admin\_2



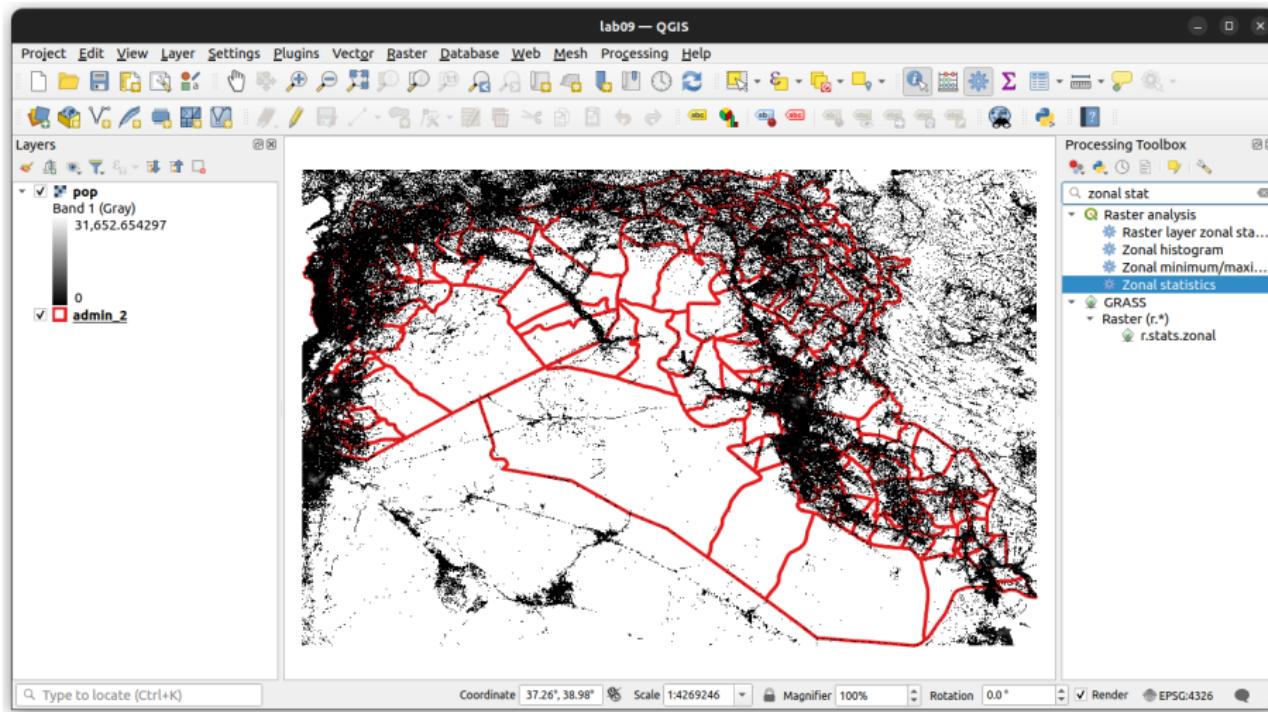
Save the file as pop.tif. Accept defaults for all other parameters, and click Run



A new layer called pop should appear. You can remove the original, global gpw\_v4... layer

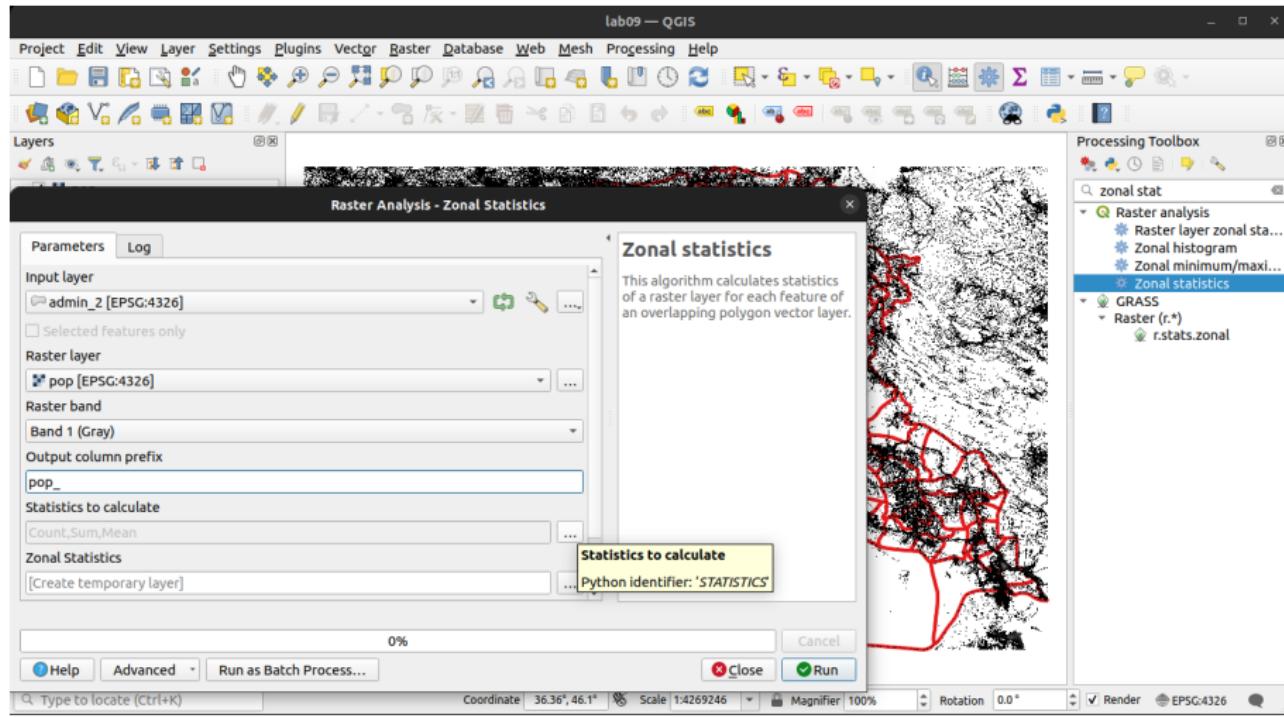


Let's calculate the population size of each district. Open the Processing Toolbox and open the Zonal statistics tool in the Raster Analysis menu

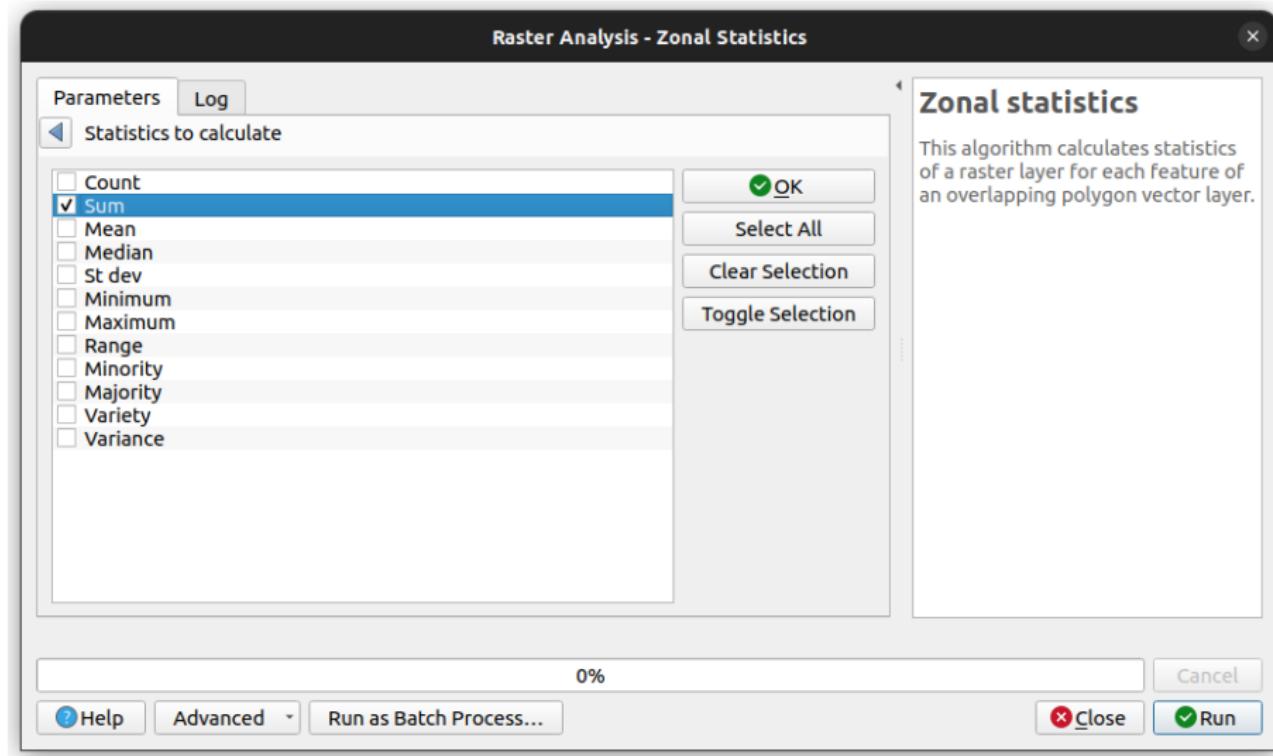


In the Zonal Statistics tool, set Input layer=admin\_2, Raster layer=pop, and Output column prefix=pop\_.

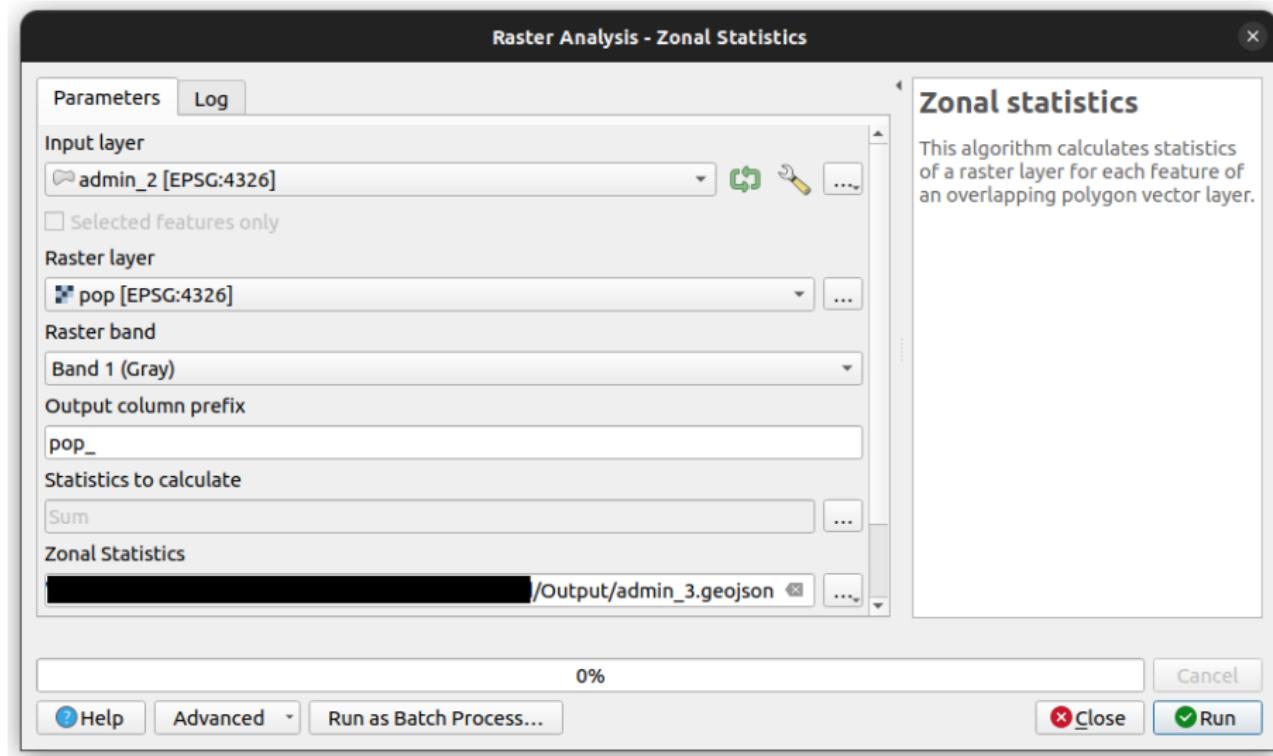
Click the [...] button next to Statistics to calculate



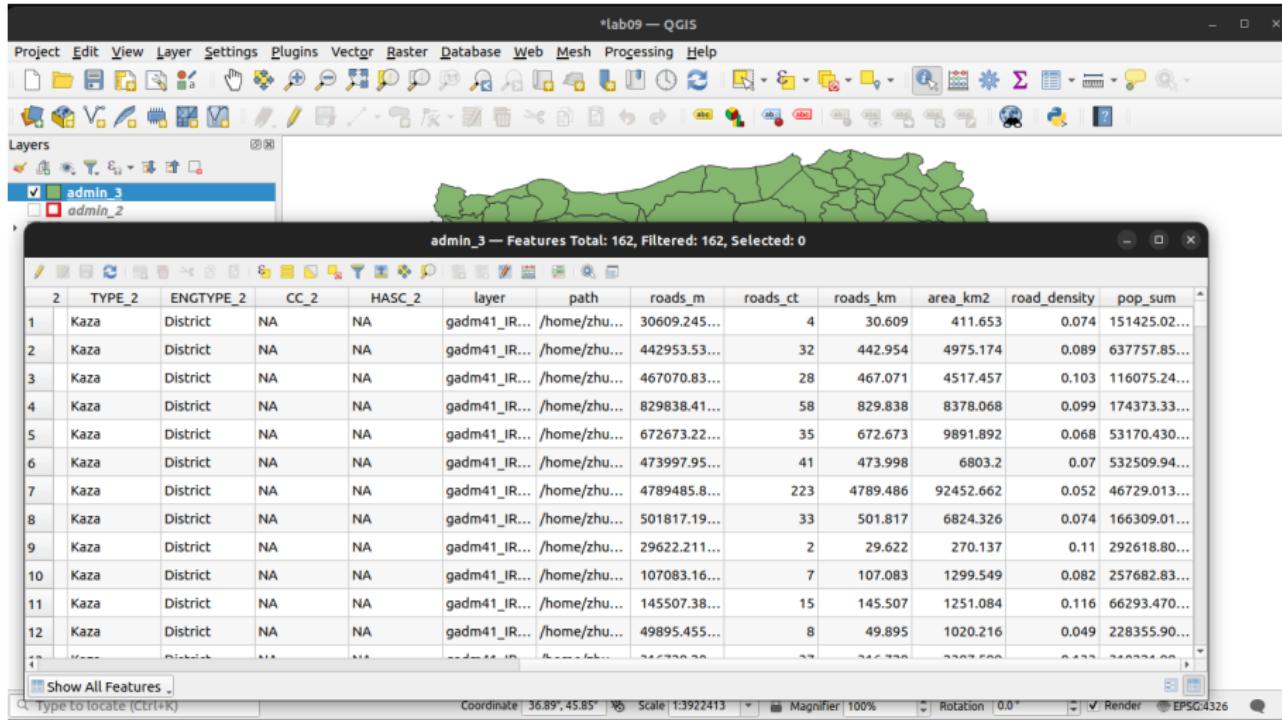
Check the box next to Sum. Leave all others un-checked. Click OK



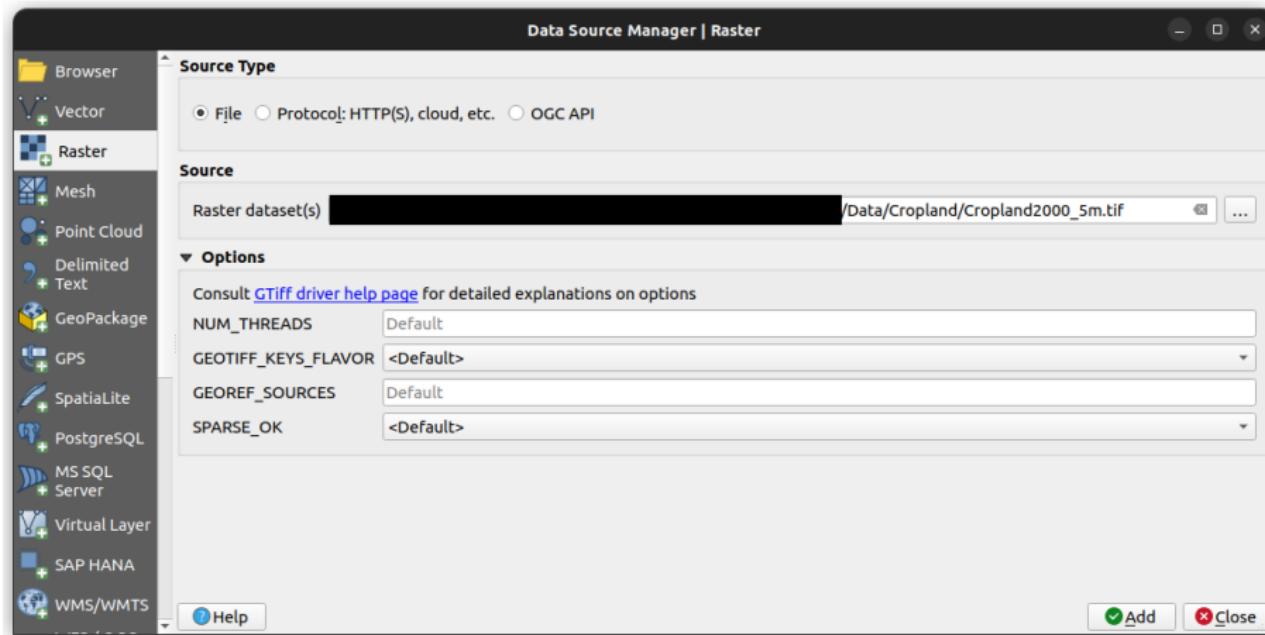
Save the zonal statistics output as admin\_3.geojson. Click Run



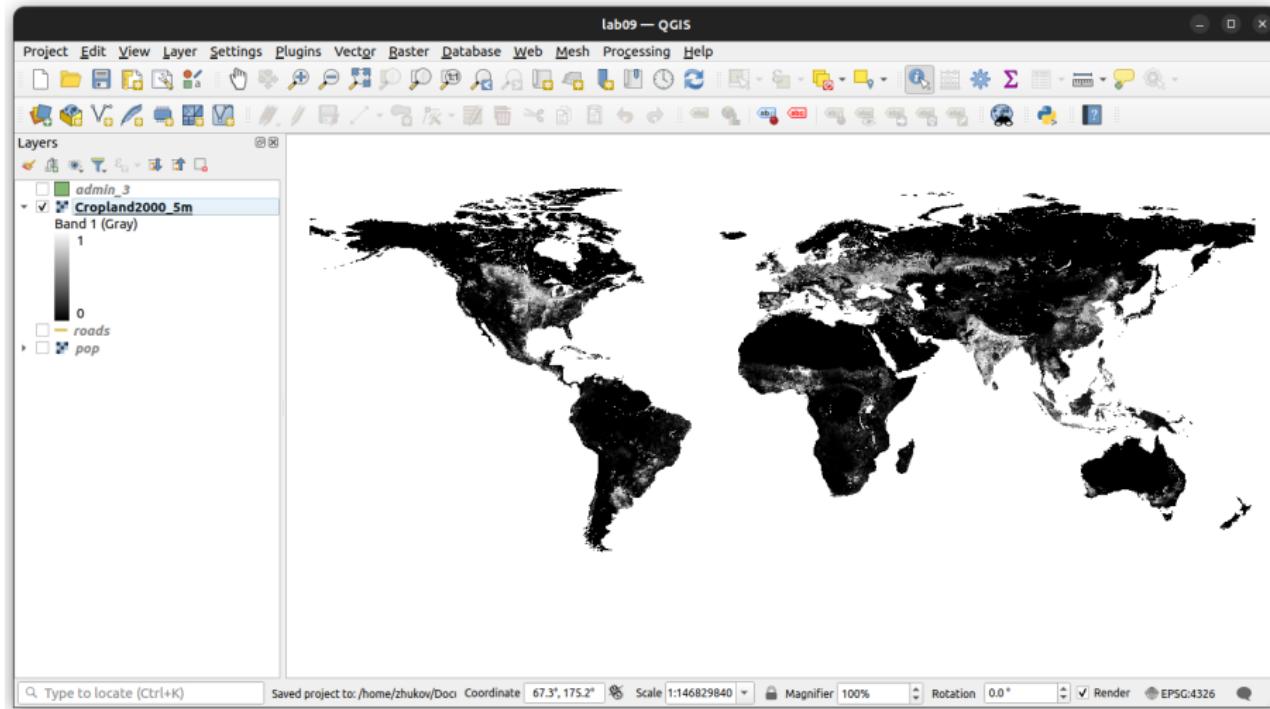
This should have generated a new layer, admin\_3, with a new field, pop\_sum (check the layer's attribute table)



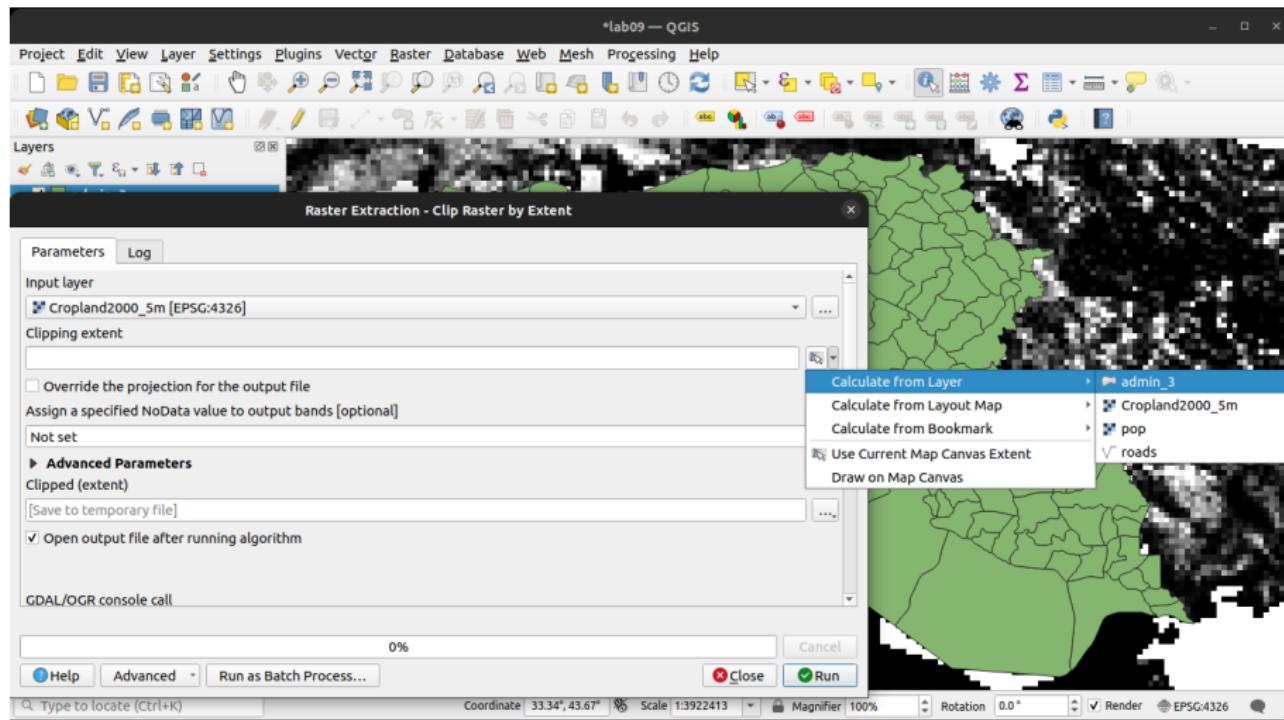
Load *cropland* raster (Layer → Add Layer → Add Raster Layer).  
Open the *cropland.tif* file in Data/Cropland folder.



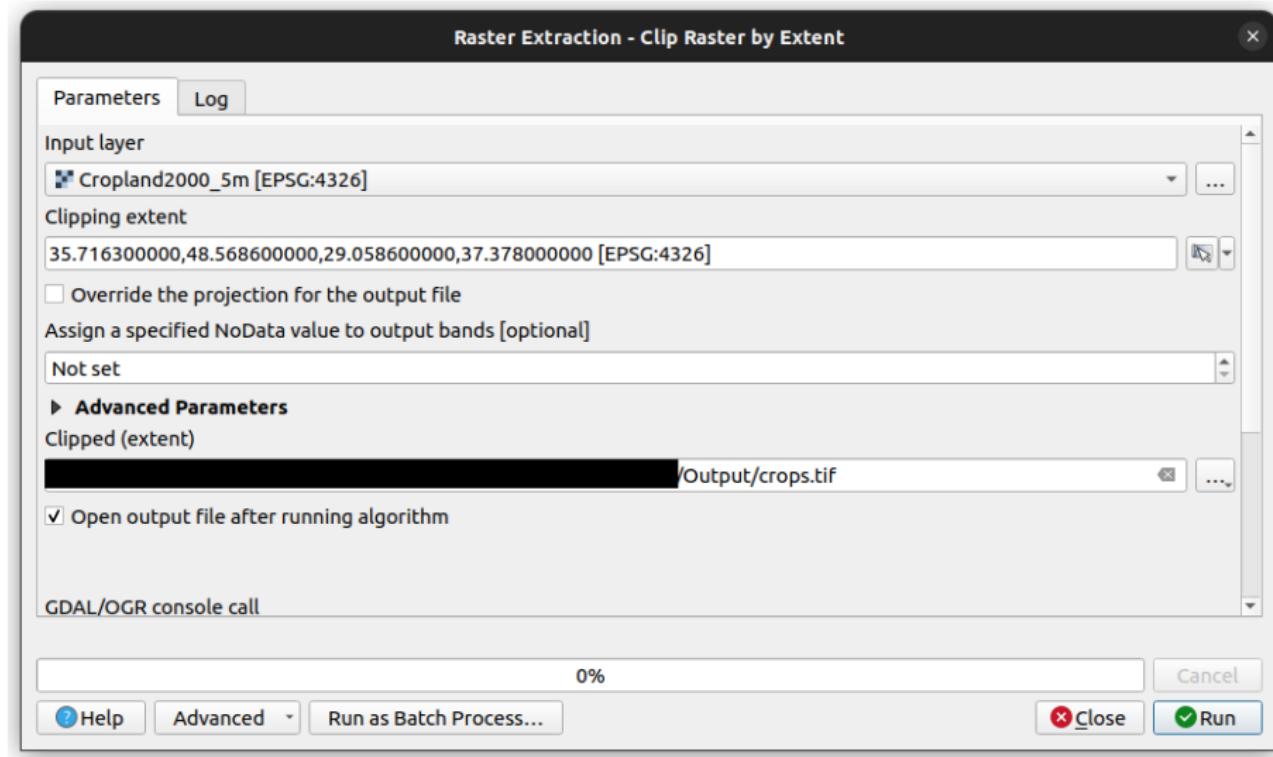
The (global) crops layer should now be visible. Let's **extract a subset** of it, as we just did for population



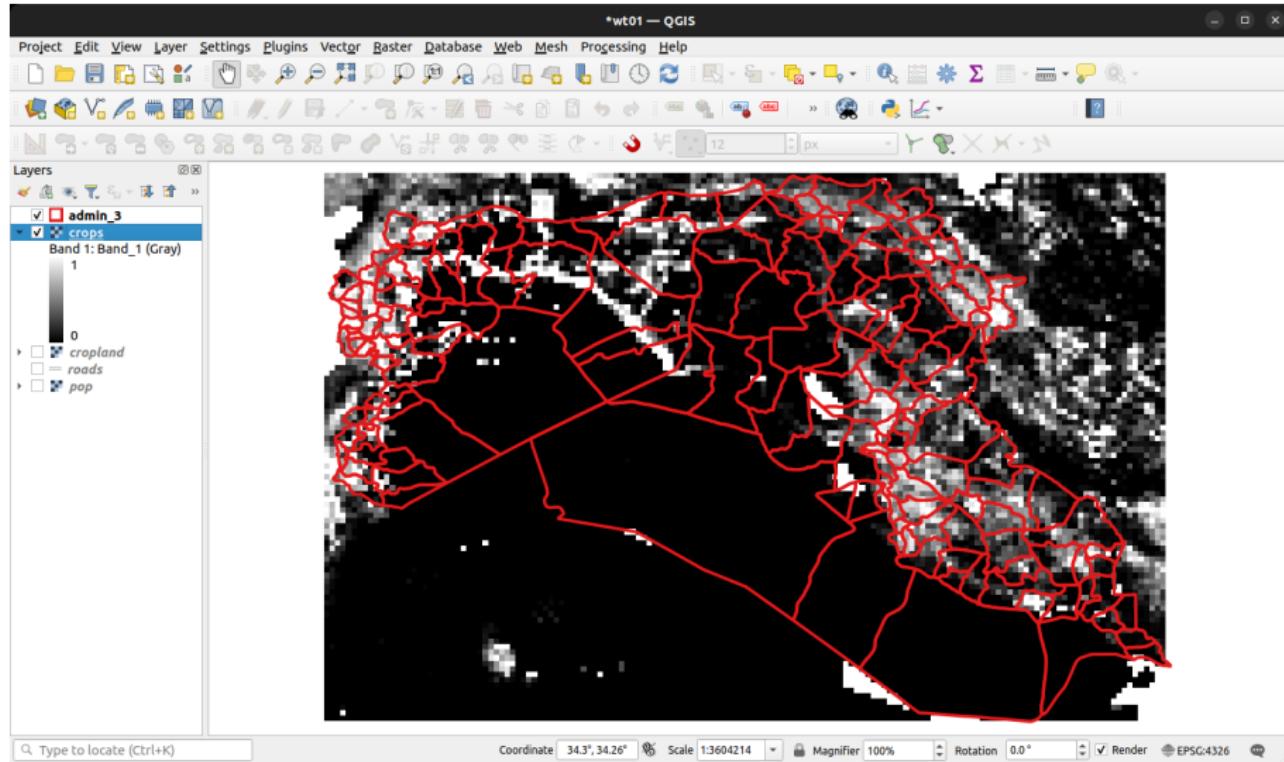
Open the Clip Raster by Extent tool (Raster → Extraction → Clip Raster by Extent). Set Input layer = cropland. Click the [...] button next to Clipping extent → Calculate from Layer → admin\_3



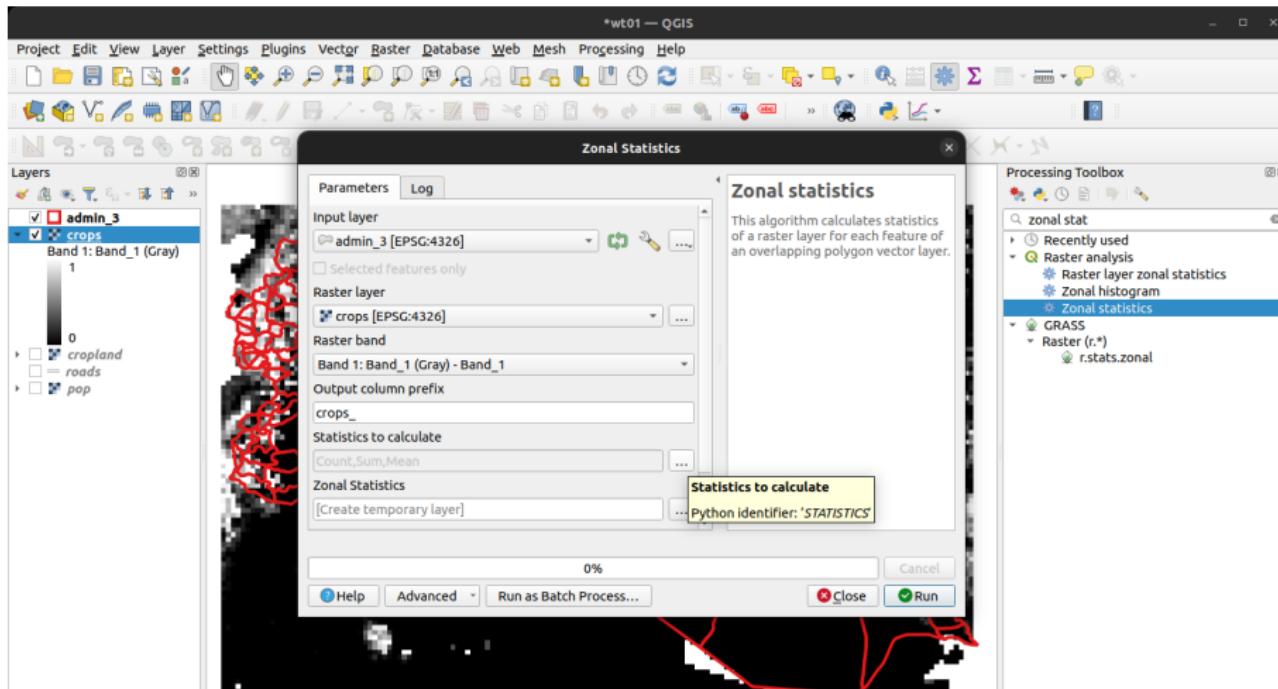
Save the output to a file called crops.tif



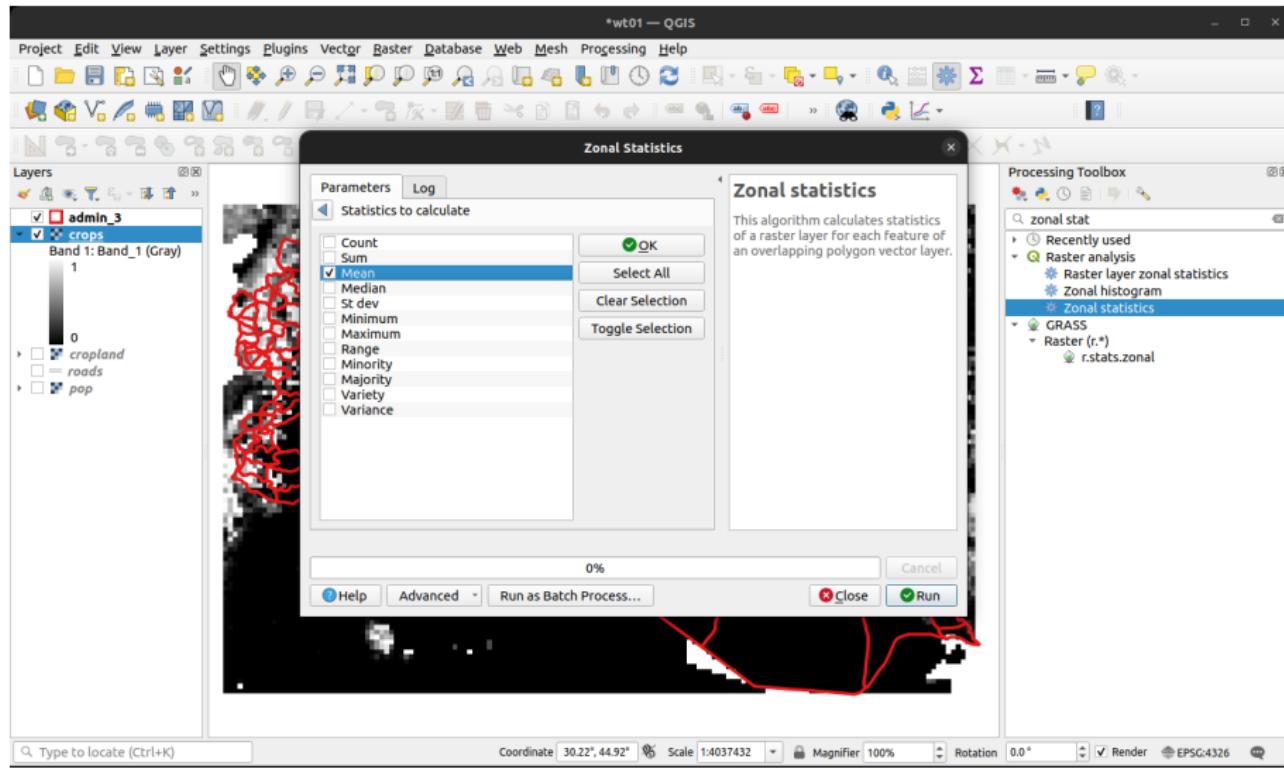
Make sure everything looks right. Remove the original, global cropland layer



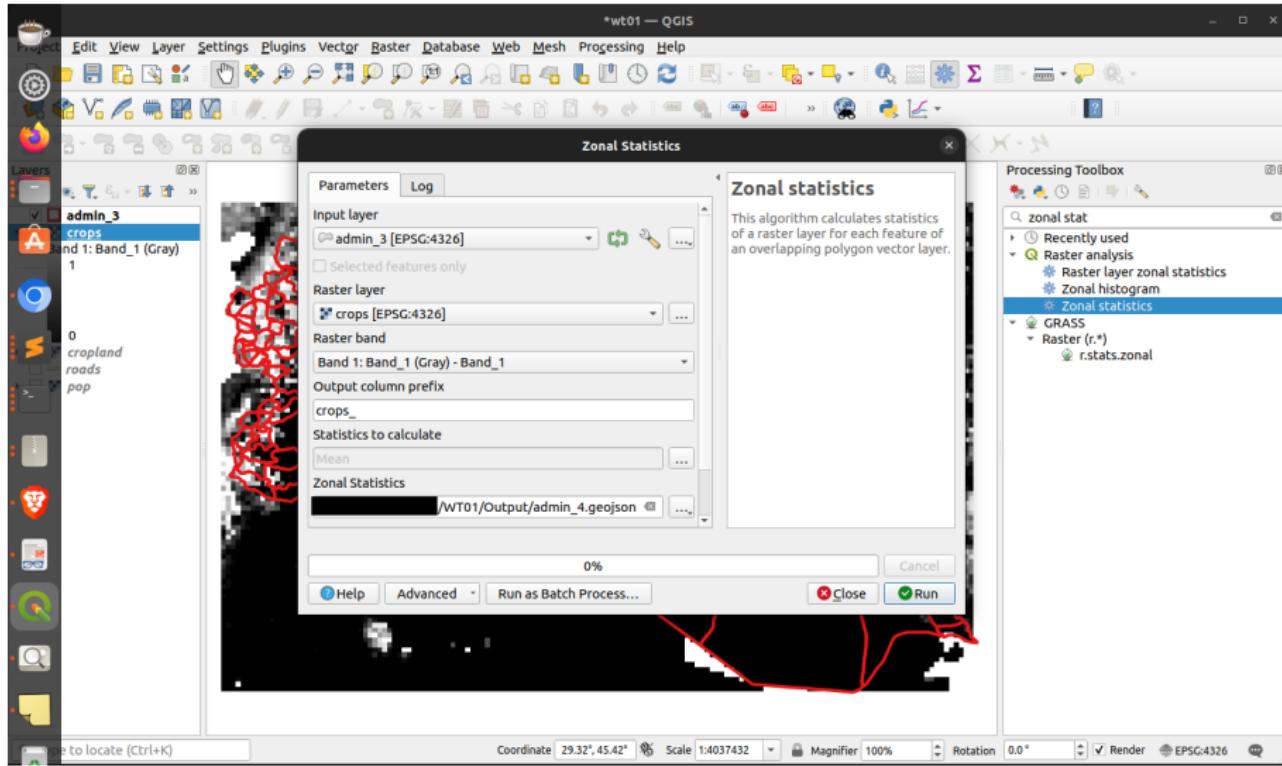
Let's **calculate crop cultivation in each district**. Open the Zonal statistics tool in Processing Toolbox. Set Input layer=admin\_3, Raster layer=crops, and Output column prefix=crops\_. Click the [...] button next to Statistics to calculate



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

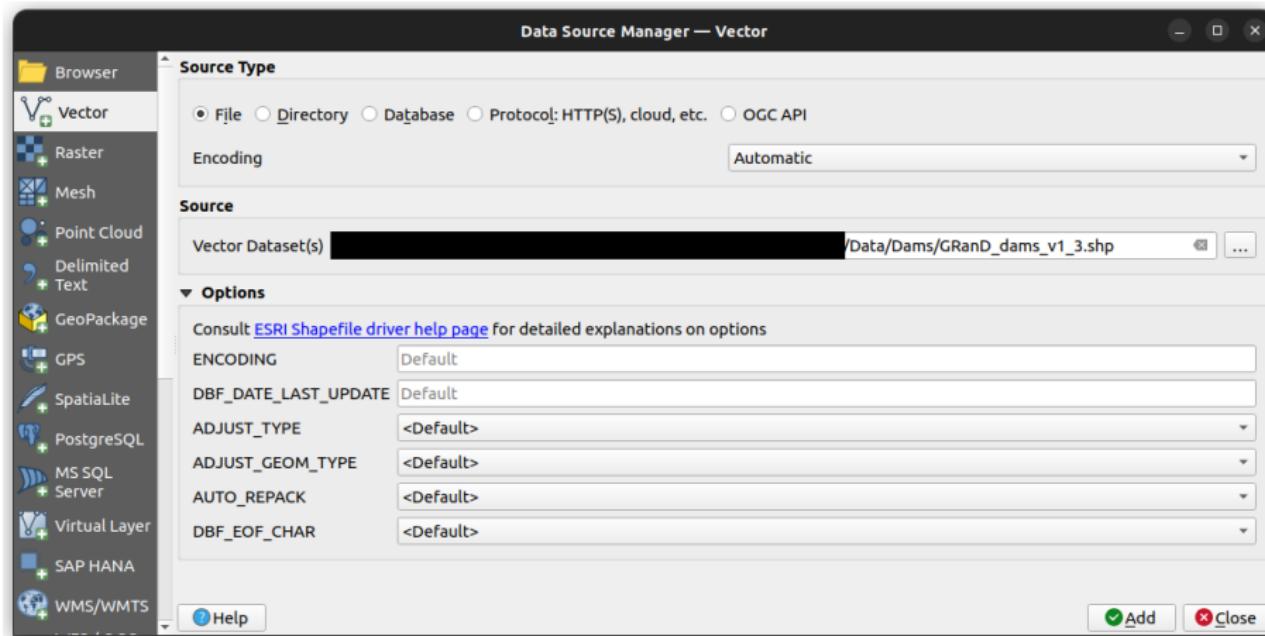


This should have generated a new field, `crops_mean` in the new `admin_4` layer (check the layer's attribute table)

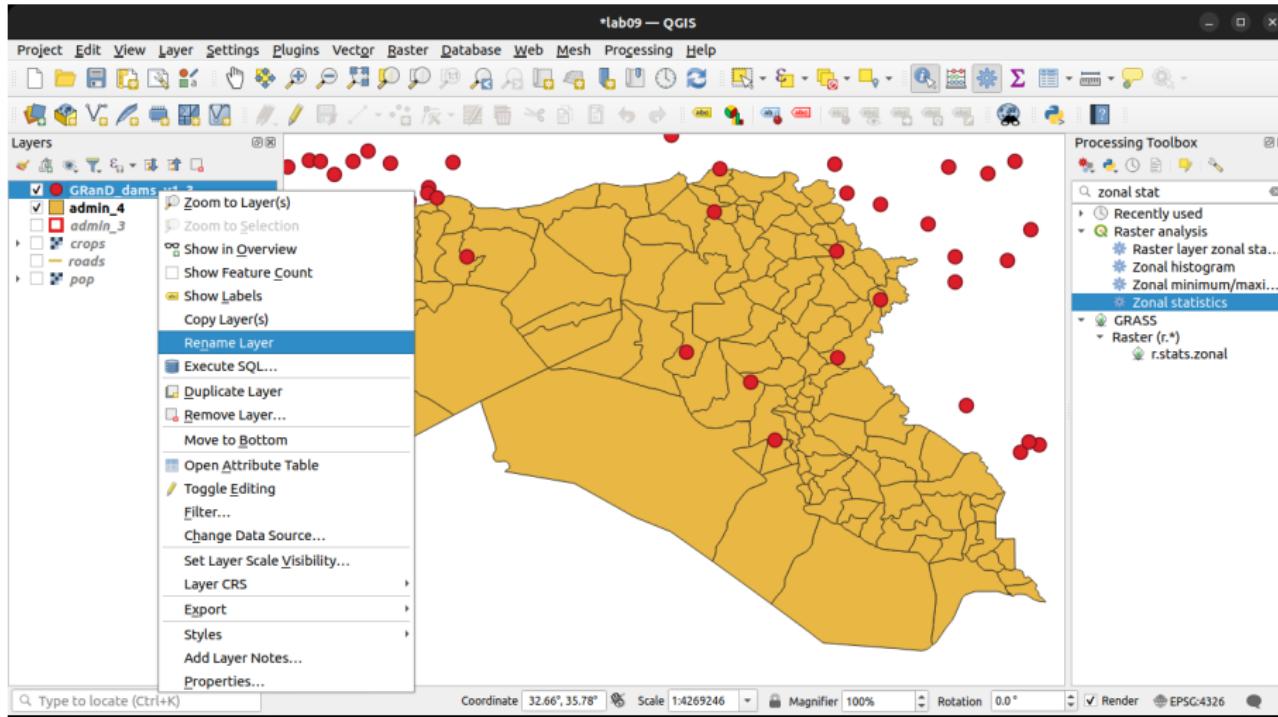
The screenshot shows the QGIS interface with the following components:

- 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 the crop area.
- Processing Toolbox:** Contains recent items, Raster analysis (selected), Zonal statistics, and GRASS.
- Map View:** Displays a map of Iraq with yellow-shaded administrative boundaries and a grayscale crop raster overlay.
- 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, 0.0104193..., etc.
- Bottom Bar:** Includes a search bar, coordinate display (35.94°, 35.46°), scale (1:4037432), magnifier, rotation, render checkbox, and EPSG code (EPSG:4326).

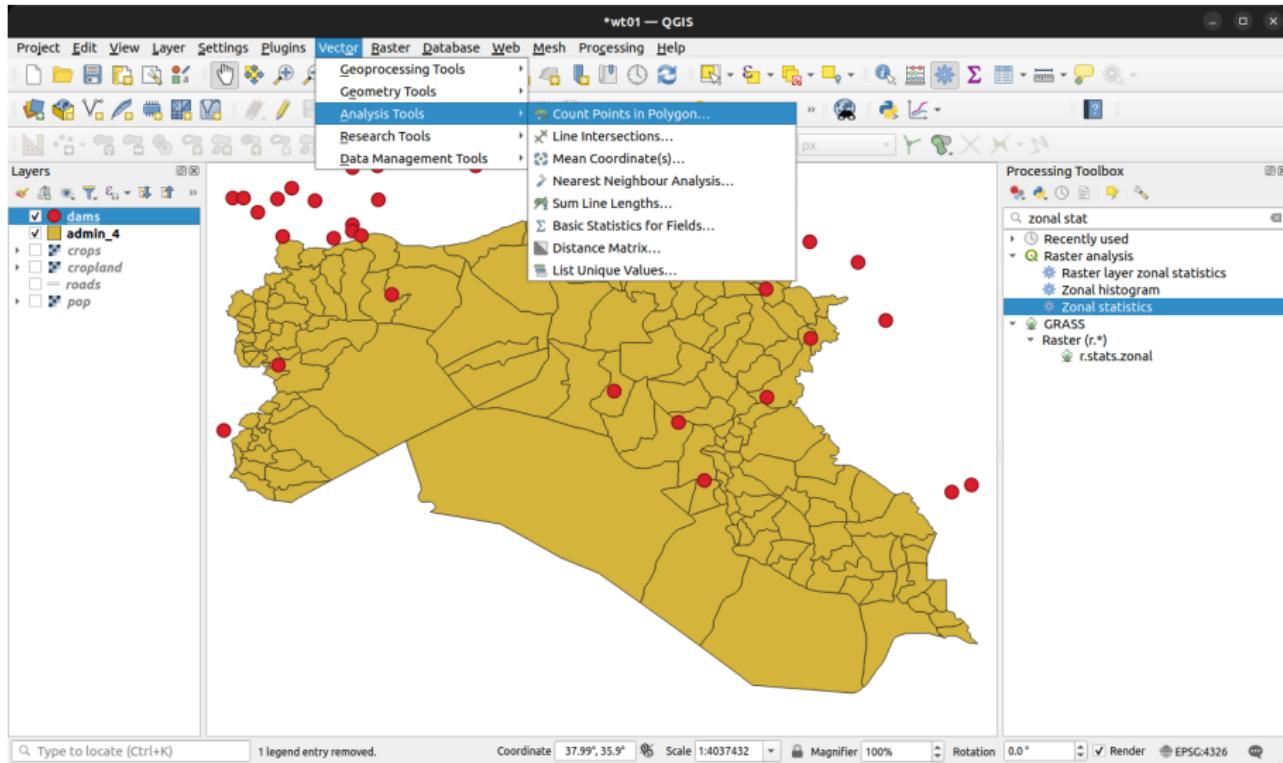
Load *hydroelectric dam* locations (Layer → Add Layer → Add Vector Layer).  
Open the GRanD\_dams\_v1\_3.shp file in Data/Dams folder.



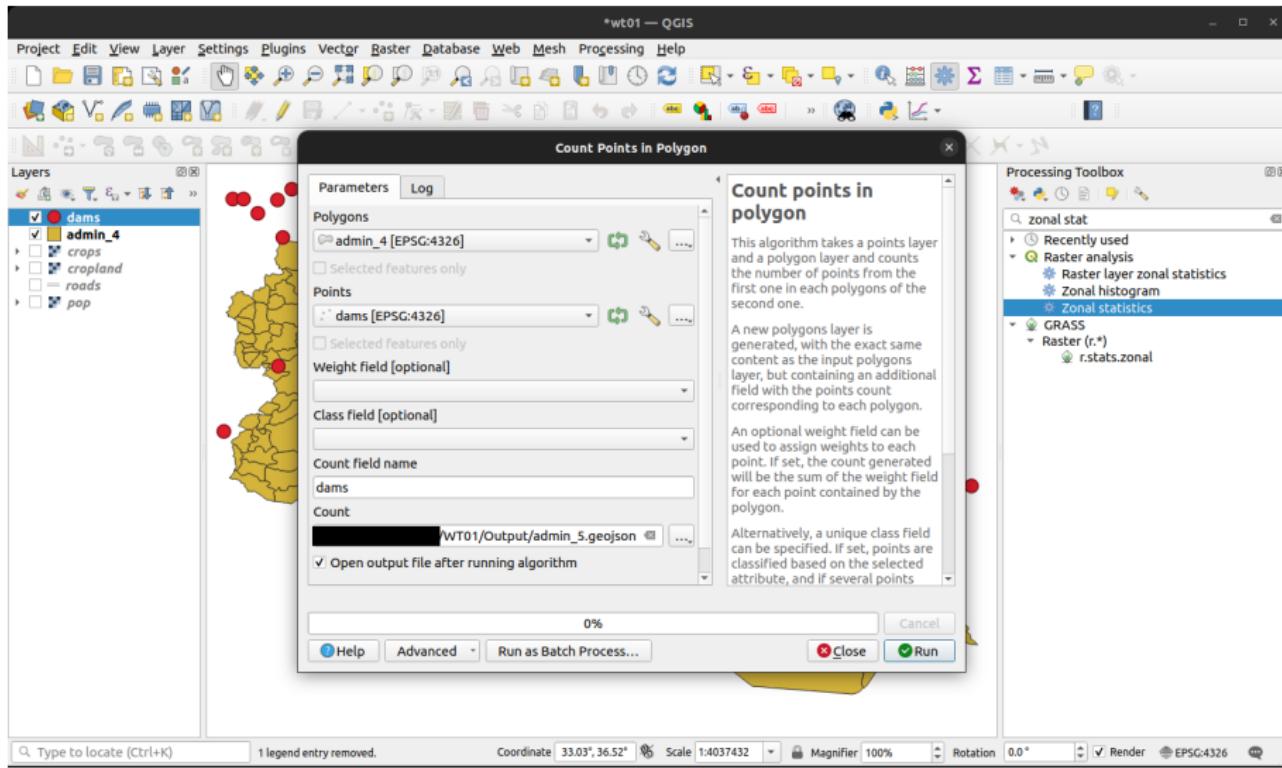
The dam locations should appear on the project window. Rename it from GRanD\_dams\_v1\_3 to dams



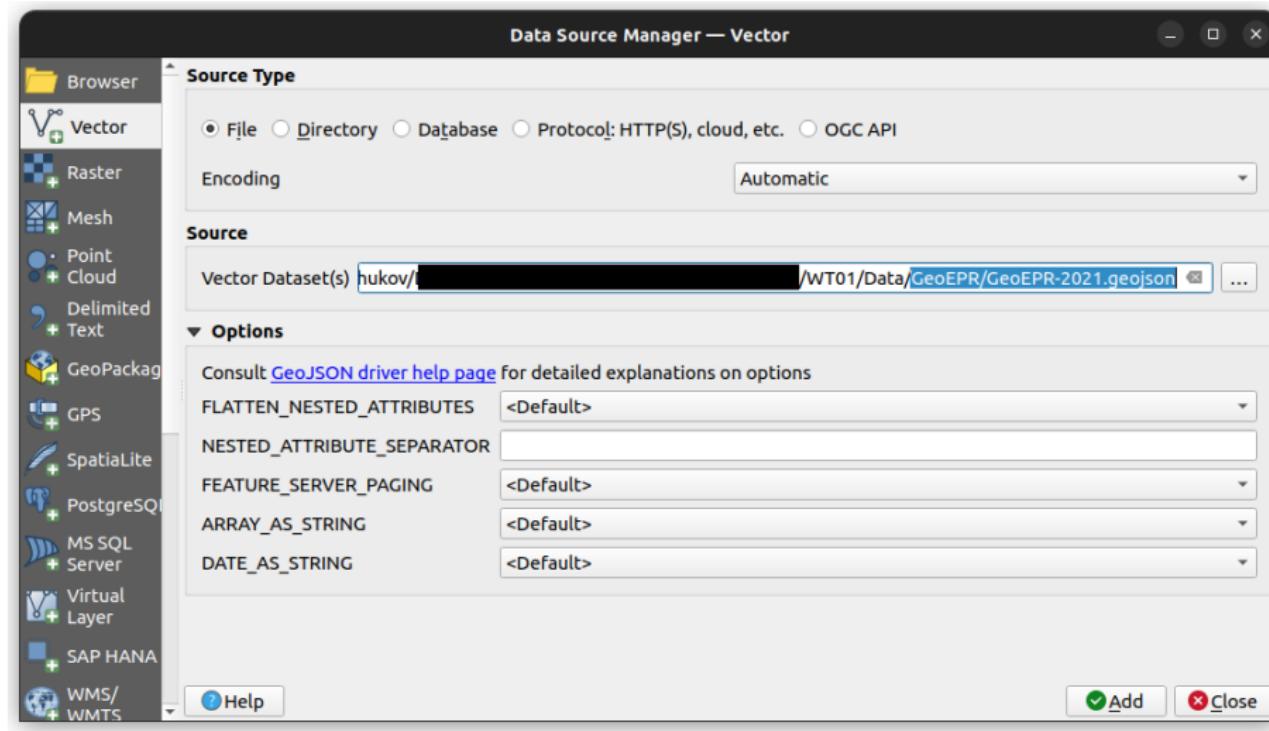
Let's calculate the **number of dams per district**. Open the Count Points in Polygon tool (Vector → Analysis Tools → Count Points in Polygon)



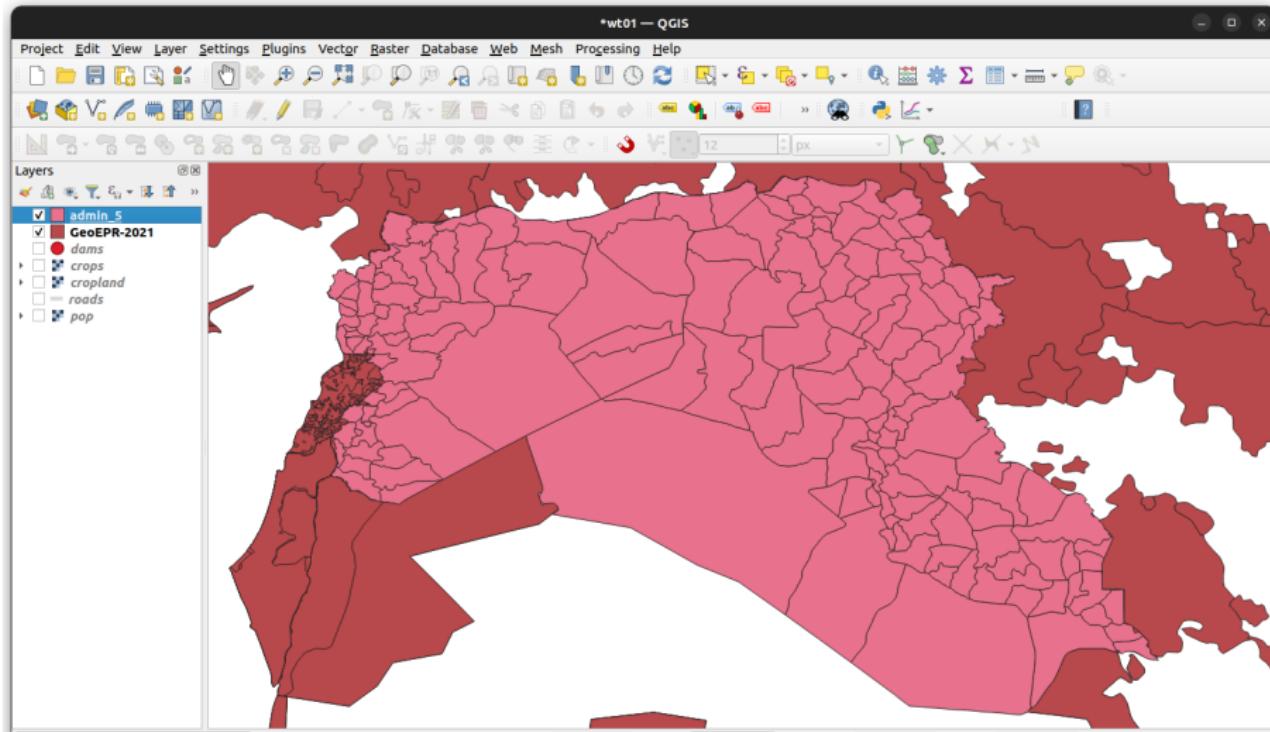
Select Polygons = admin\_4, Points = dams. Name the count field dams, and save the output file as admin\_5.geojson. Click Run



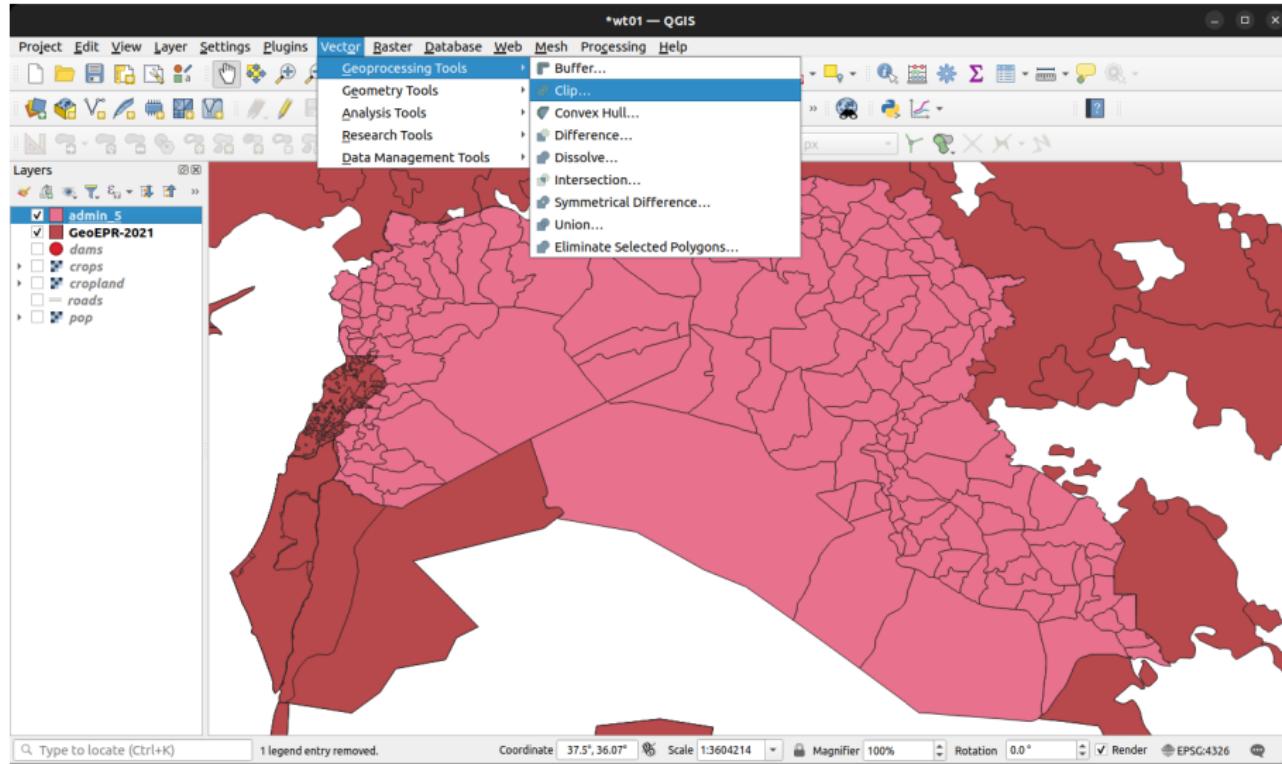
Add the *sectarian divisions* data to the project, using Add Vector Layer....\\ Load the GeoEPR-2021.geojson file in Data/GeoEPR folder



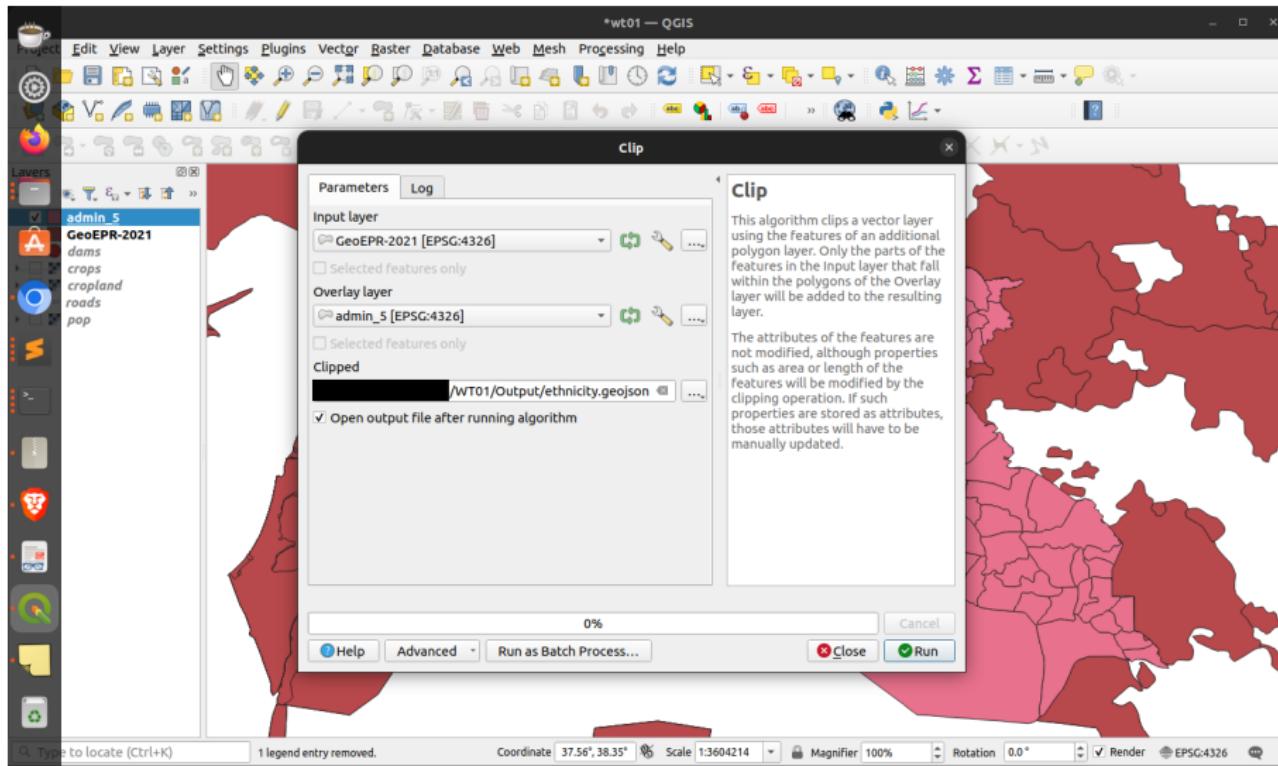
The (global) ethnicity layer should appear in the project window. Let's **calculate the proportion of each district populated by Sunni Arabs**. Let's first extract the subset that overlaps with our study area.



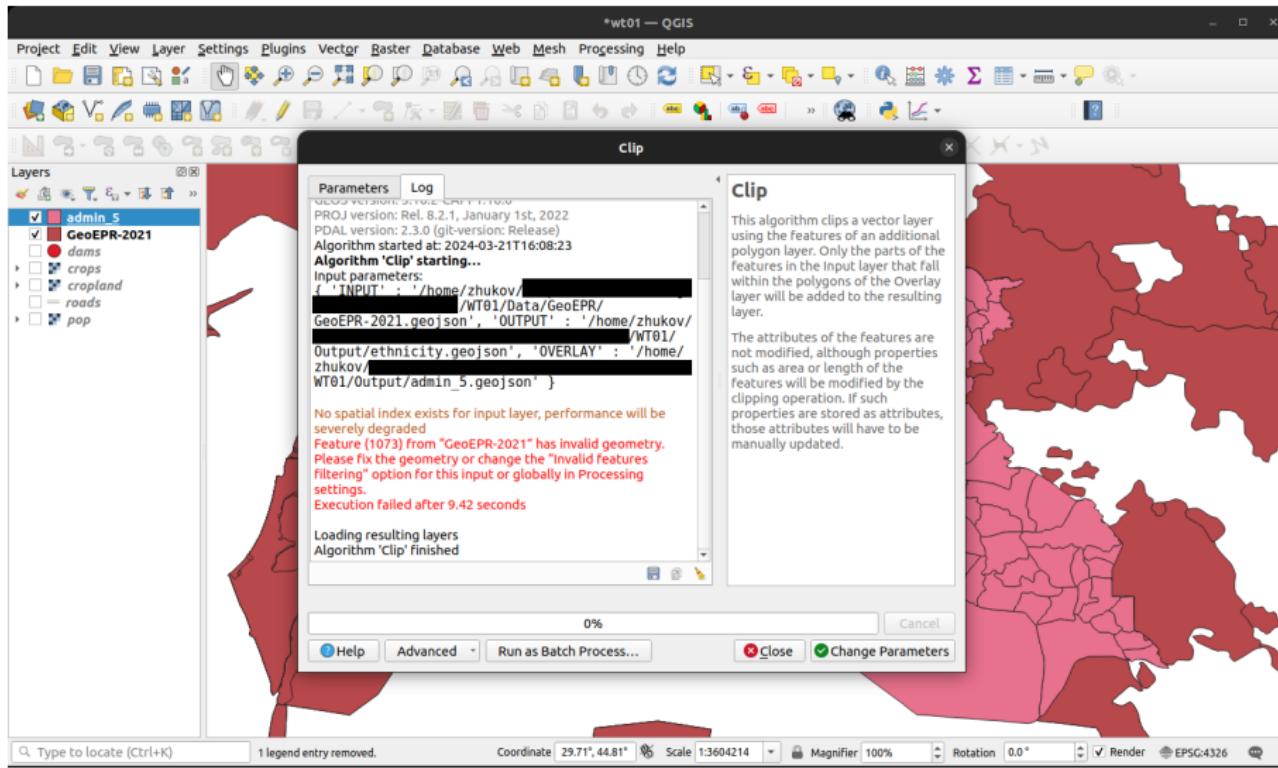
Open the Clip (vector) tool (Vector → Geoprocessing Tools → Clip)



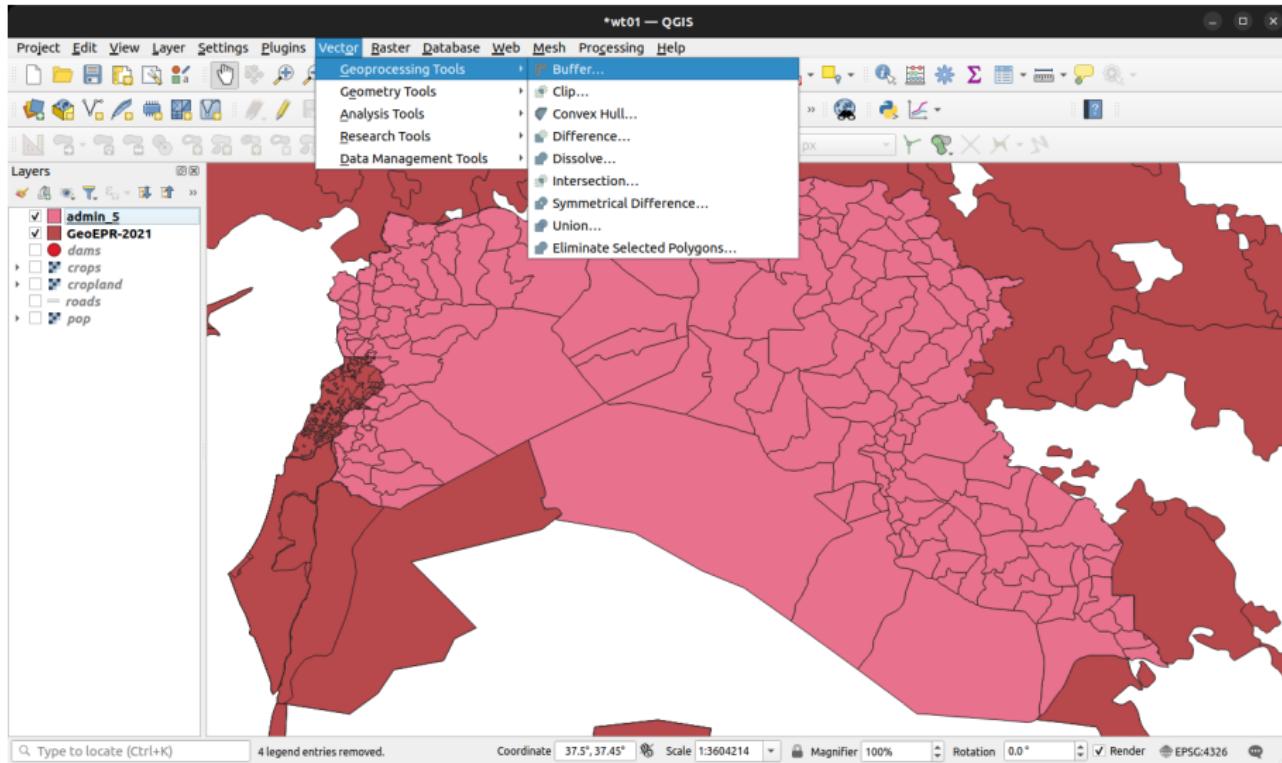
Set Input layer = GeoEPR, Overlay layer = admin\_5. Save the file as ethnicity.geojson. Click Run



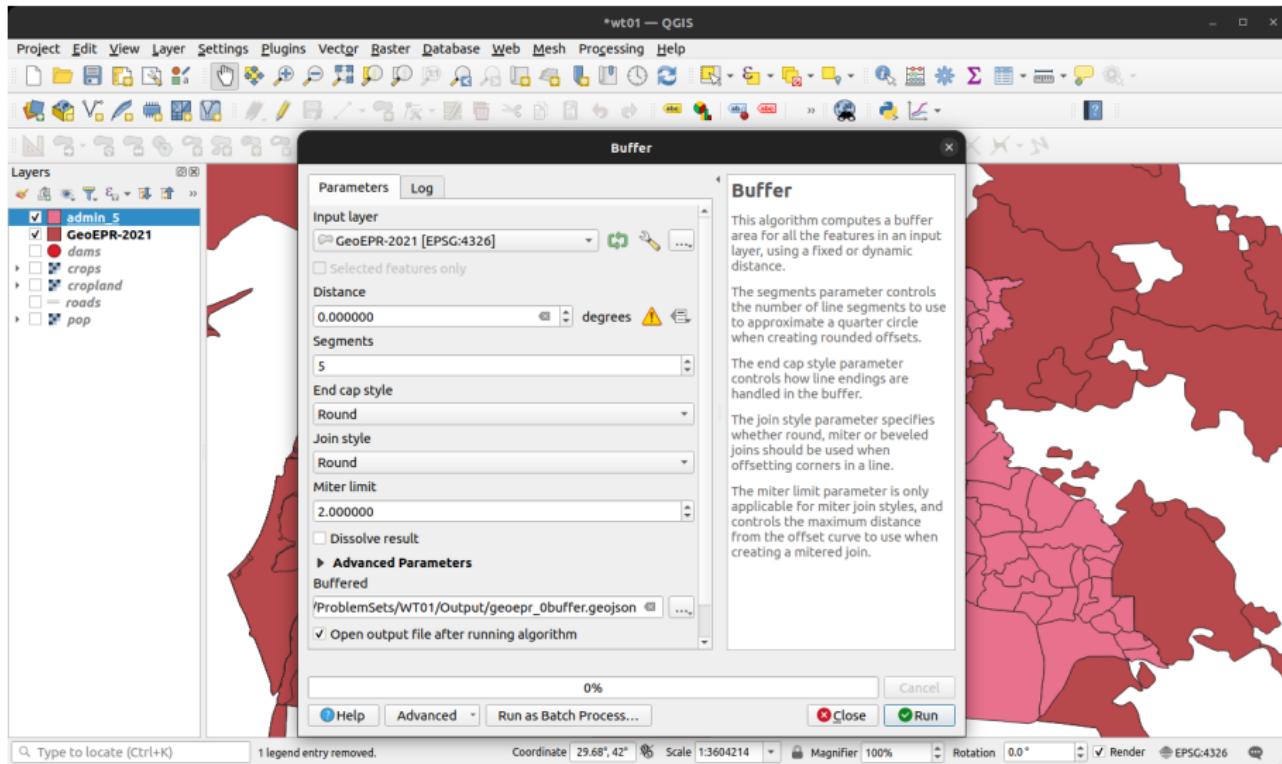
Uh oh! **Execution failed!** The error message says that GeoEPR-2021 contains "invalid geometries". How do we fix this?



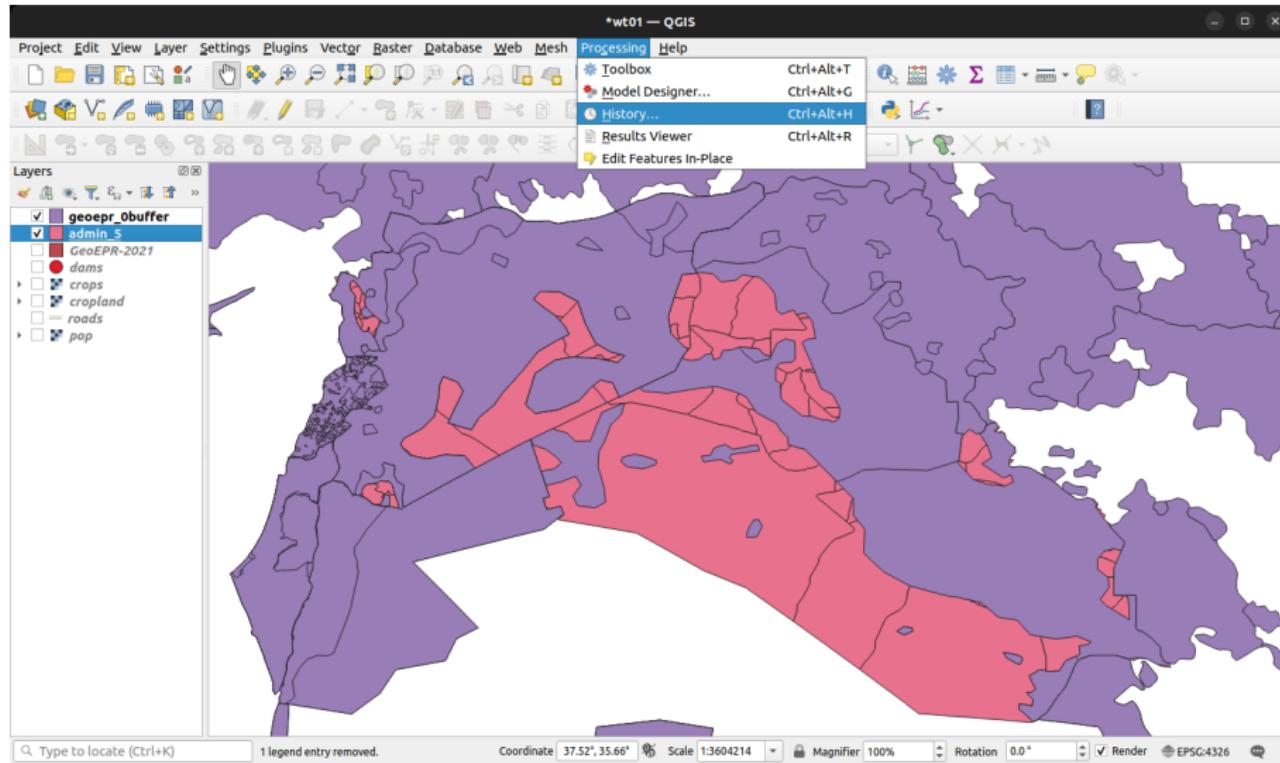
Many “invalid geometry” problems can be solved with the “zero buffer trick”. Go to Vector → Geoprocessing Tools → Buffer...



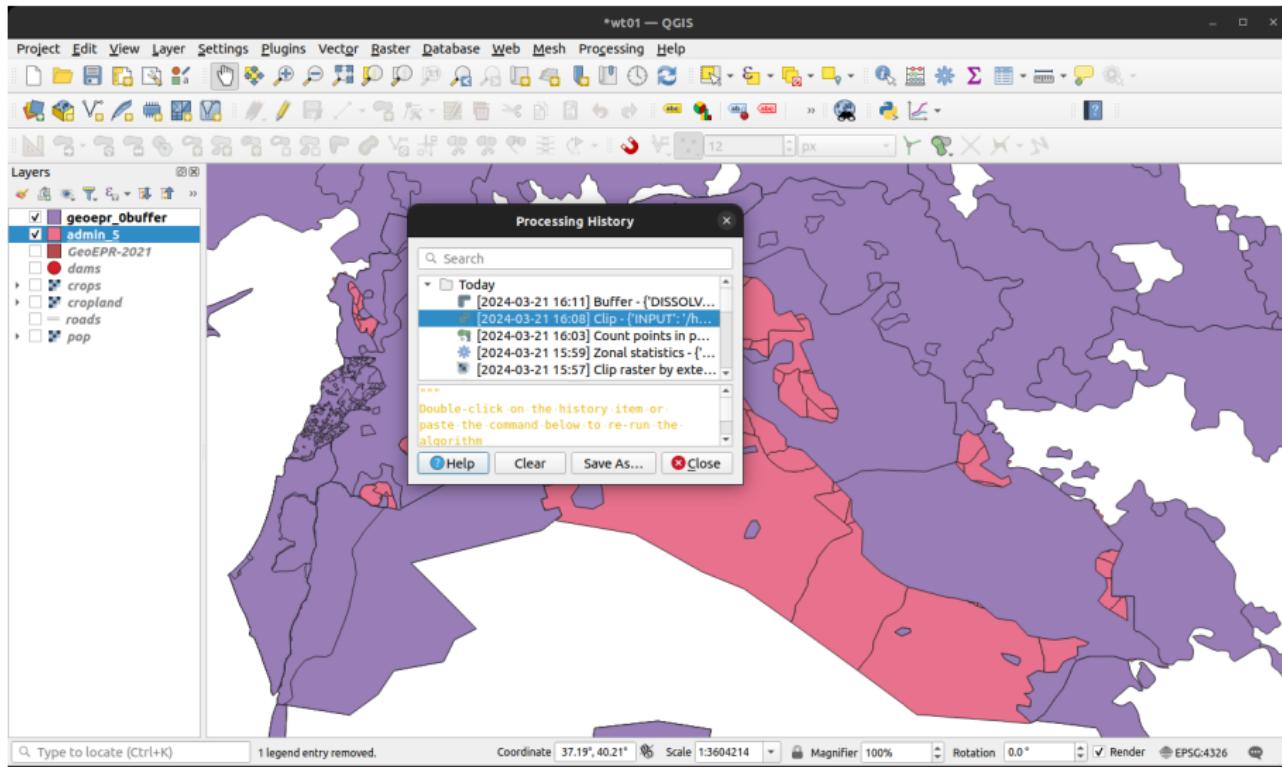
Set Input layer = GeoEPR-2021 and Distance = 0. Keep the defaults for the other parameters and save the file as geoepr\_0buffer.geojson. Click Run



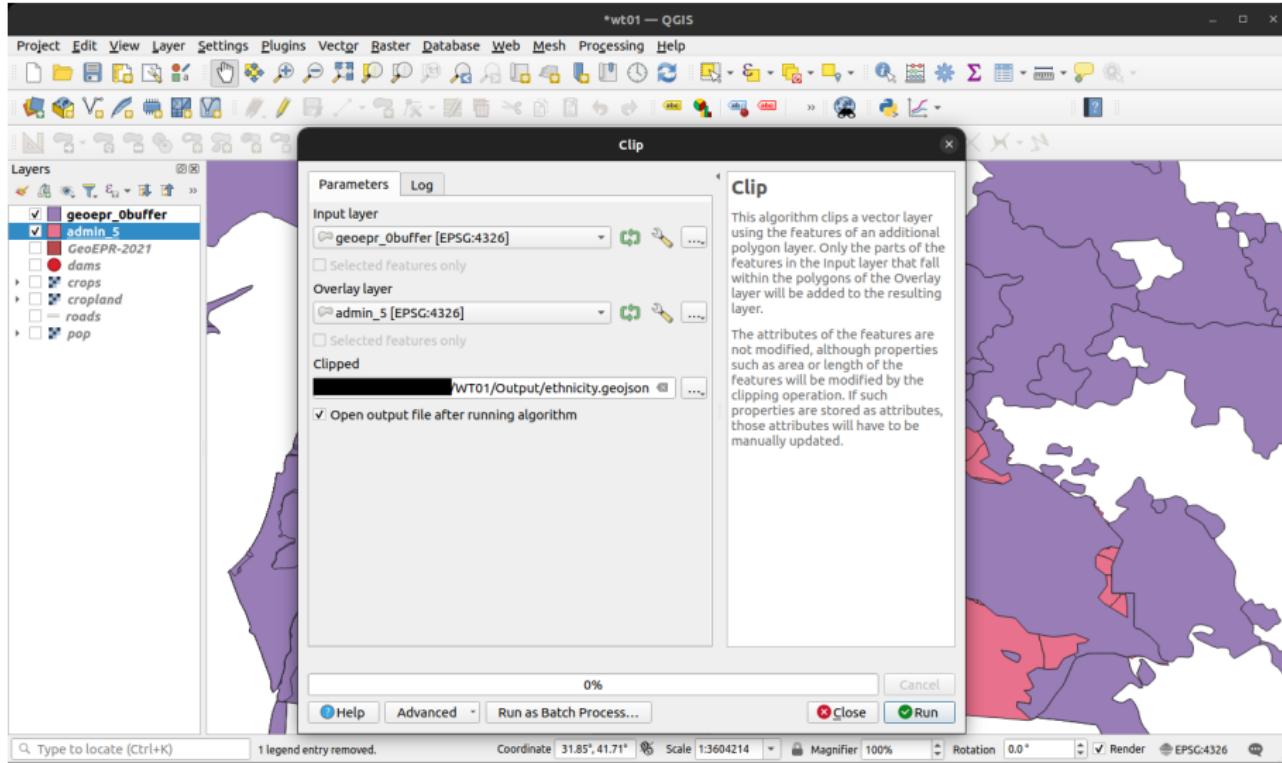
Now let's try the clipping operation again. To save time, go to Processing → History...



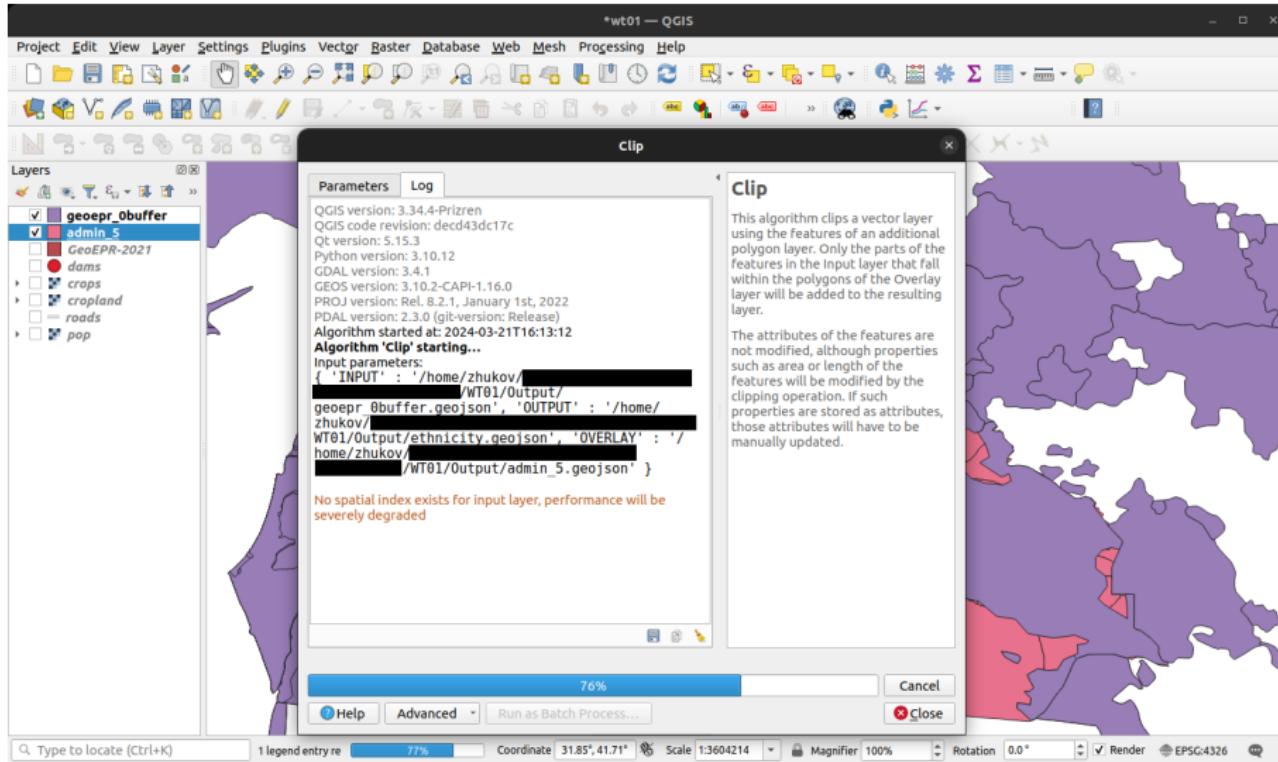
In the Processing History window, double-click on Clip (it should be the second-most recent item)



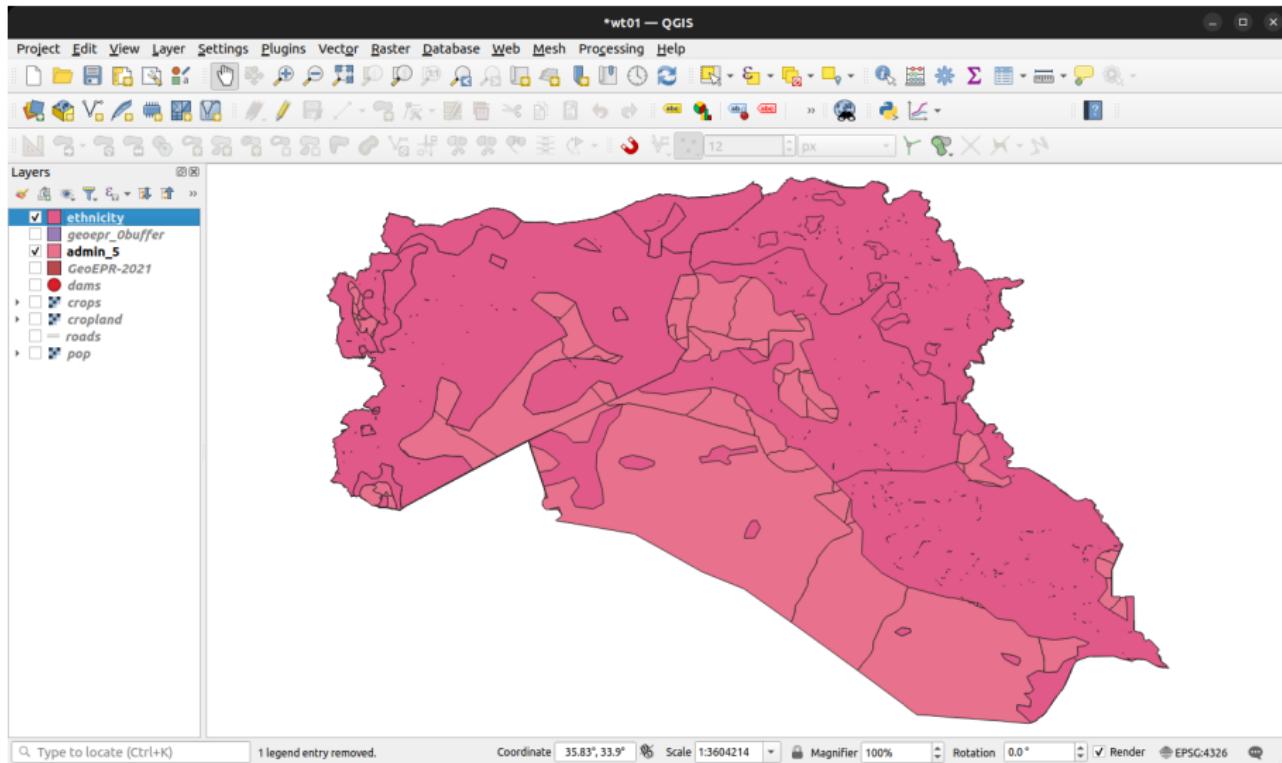
Set the Input layer = geoepr\_0buffer. Click Run



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 'ethnicity', which is colored pink. Other visible layers include 'admin\_5' (red), 'dams' (orange), 'crops' (light blue), 'cropland' (light green), 'roads' (yellow), and 'pop' (purple). The map shows a complex network of these features across a geographic area.

**Attribute Table:** The table is titled 'ethnicity — Features Total: 46, Filtered: 46, Selected: 0'. It contains the following columns:

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 row in the table is highlighted with a yellow background, showing gwid 1, statename Iraq, from 1946, to 2021, group Shi'a String, groupid 2000, gwgroupid 64502000, umbrella NULL, sqkm 108150, and type Regional & ...

**Bottom Bar:** A button labeled 'Show All Features' is visible at the bottom left of the table area.

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

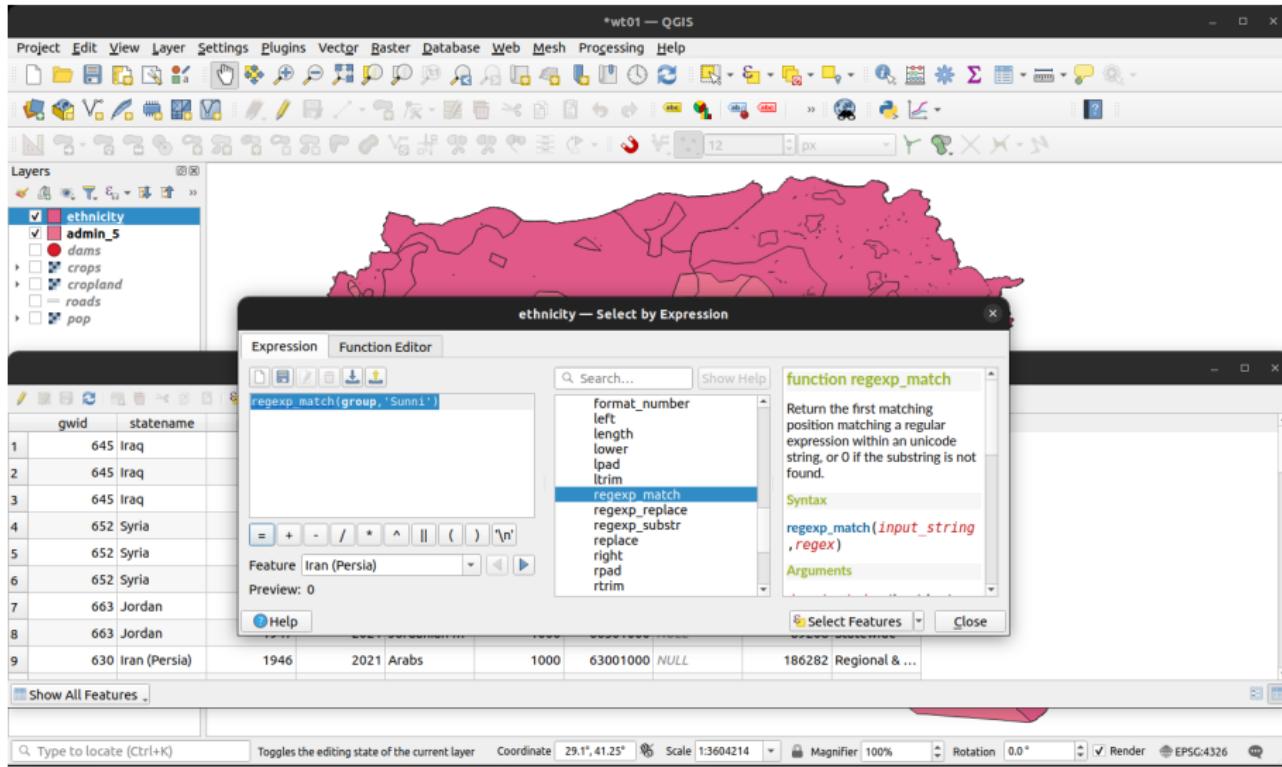
The screenshot shows the QGIS application window titled "wt01 — QGIS". The interface includes a top menu bar with Project, Edit, View, Layer, Settings, Plugins, Vector, Raster, Database, Web, Mesh, Processing, and Help. Below the menu is a toolbar with various icons for editing and analysis. On the left 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 main window contains a table with the following data:

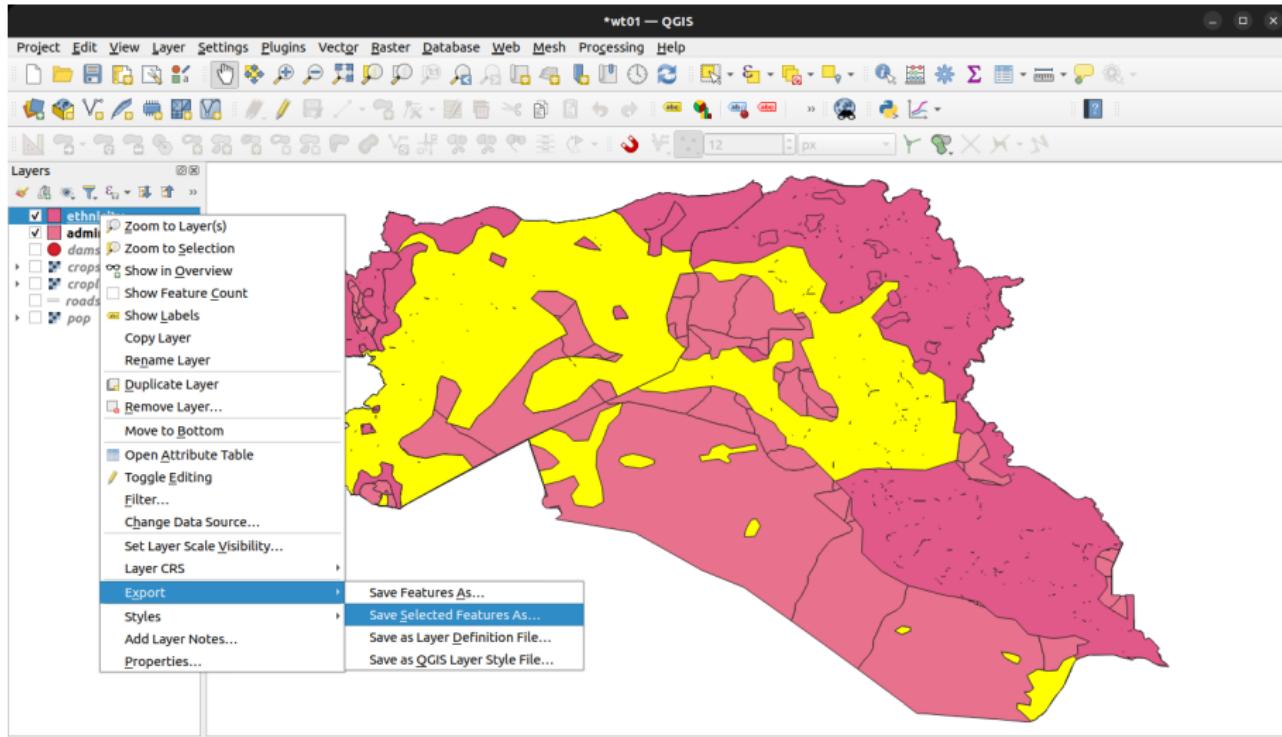
gwid	statename	groupid	gwgroupid	umbrella	sqkm	type
1	645 Iraq	1946	2021 Shi'a Arabs	2000	64502000 NULL	108150 Regional & ...
2	645 Iraq	1946	2021 Sunni Arabs	3000	64503000 NULL	96757 Regional & ...
3	645 Iraq	1946	2021 Kurds	1000	64501000 NULL	72420 Regional & ...
4	652 Syria	1946	1966 Sunni Arabs	2000	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 button labeled "Show All Features". The status bar at the very bottom of the screen includes "Type to locate (Ctrl+K)", "Toggles the editing state of the current layer", "Coordinate 29.1°, 41.25°", "Scale 1:3604214", "Magnifier 100%", "Rotation 0.0°", "Render", and "EPSG:4326".

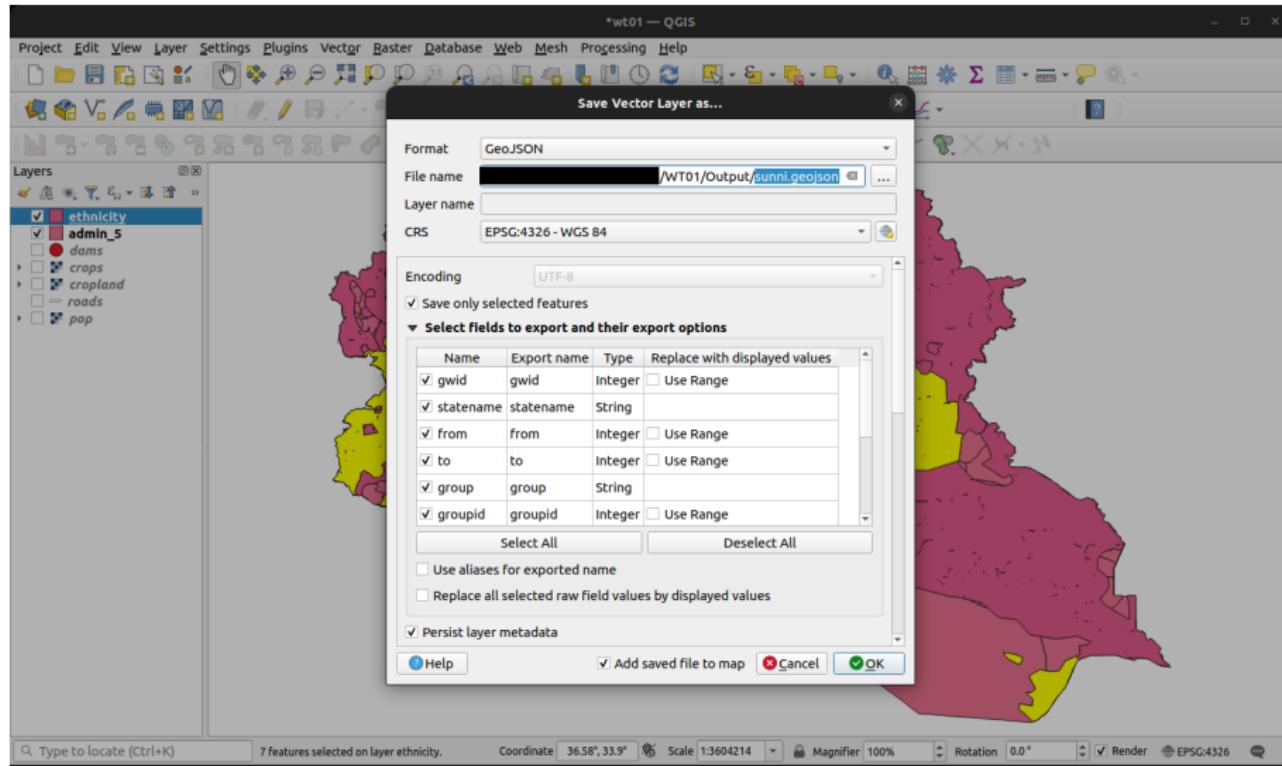
Let's use **regular expressions** to extract the polygons we need. Set Expression = `regexp_match(group, 'Sunni')`. Click Select Features



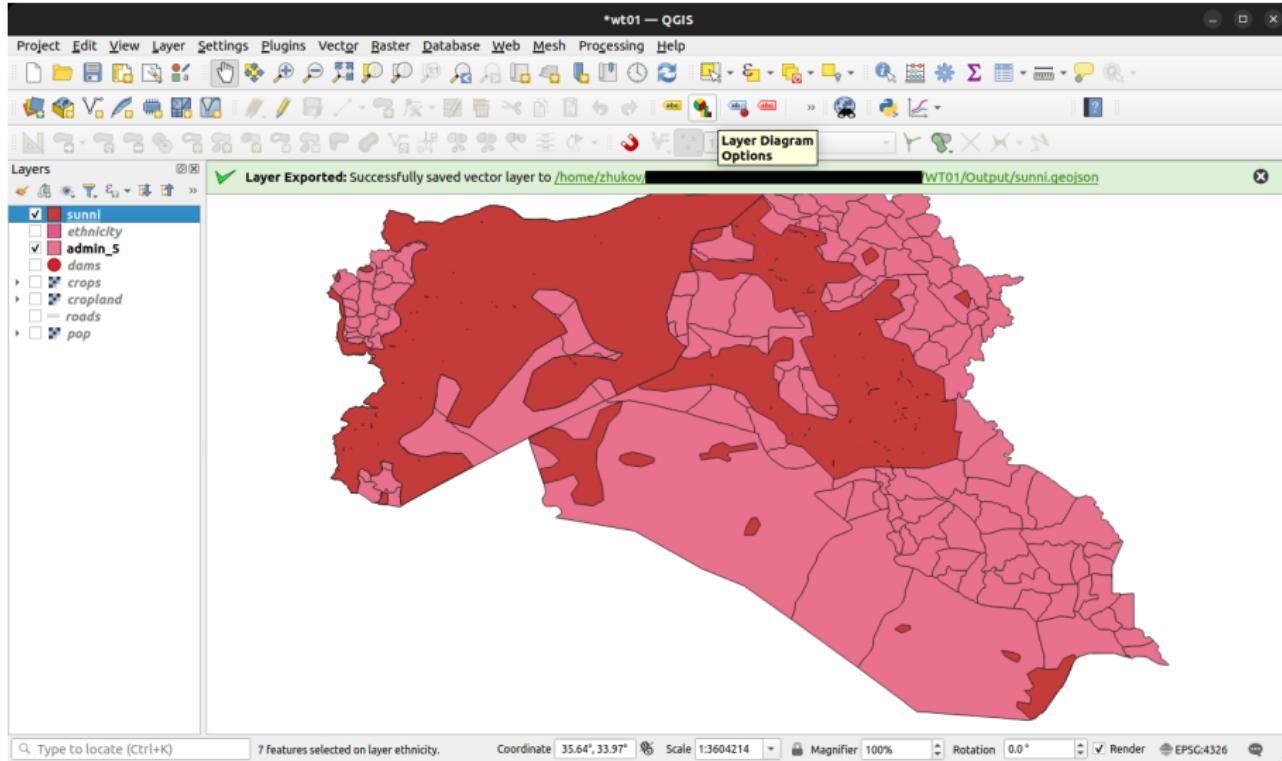
The selected polygons should appear yellow on the map. Let's **extract the selection into a new layer**. Right-click ethnicity in the Layer menu, go to Export→Save Selected Features As



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

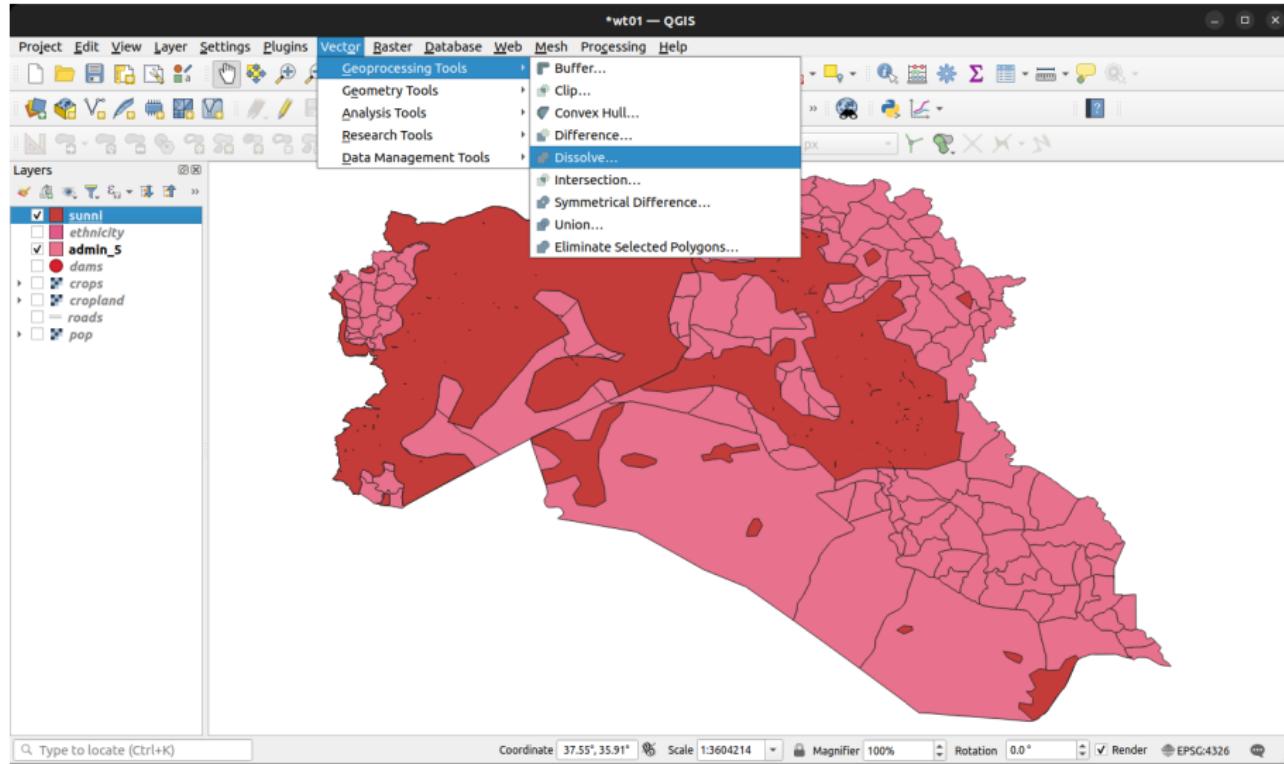
sunni — Features Total: 7, Filtered: 7, Selected: 0

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 & ...

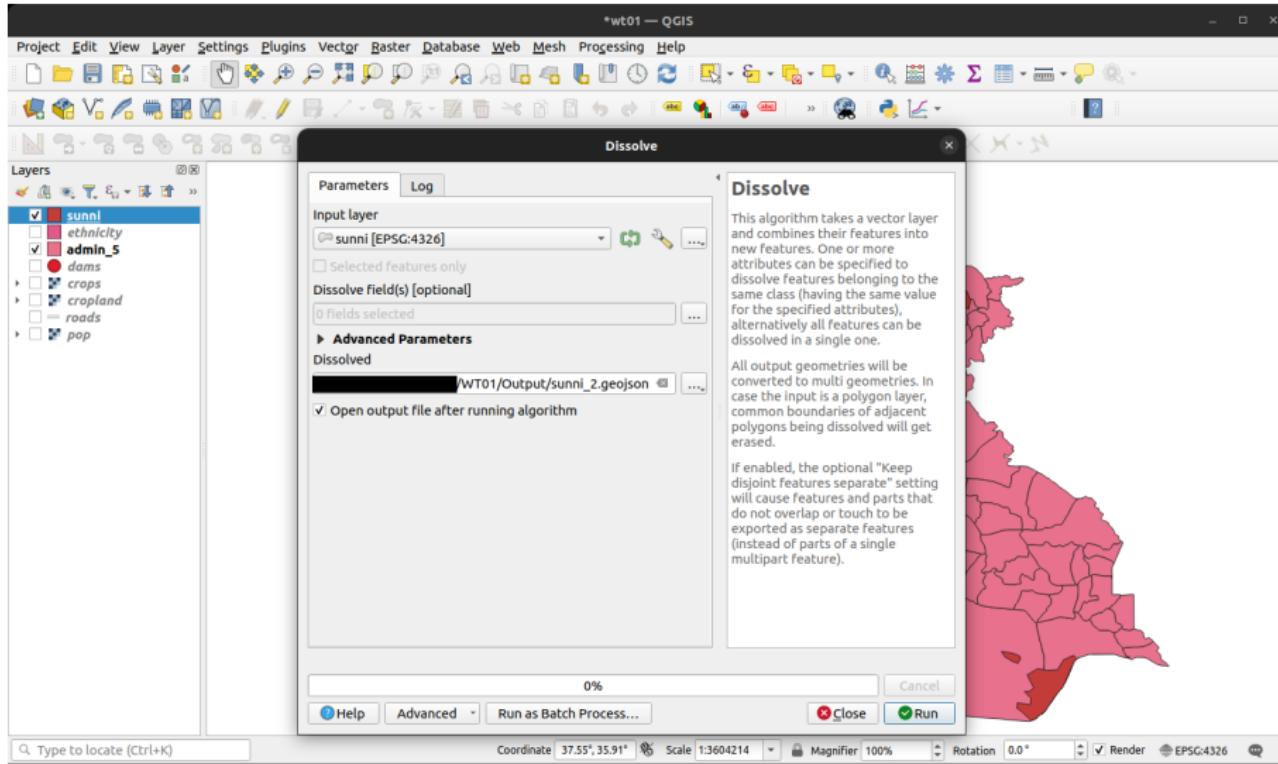
Show All Features

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

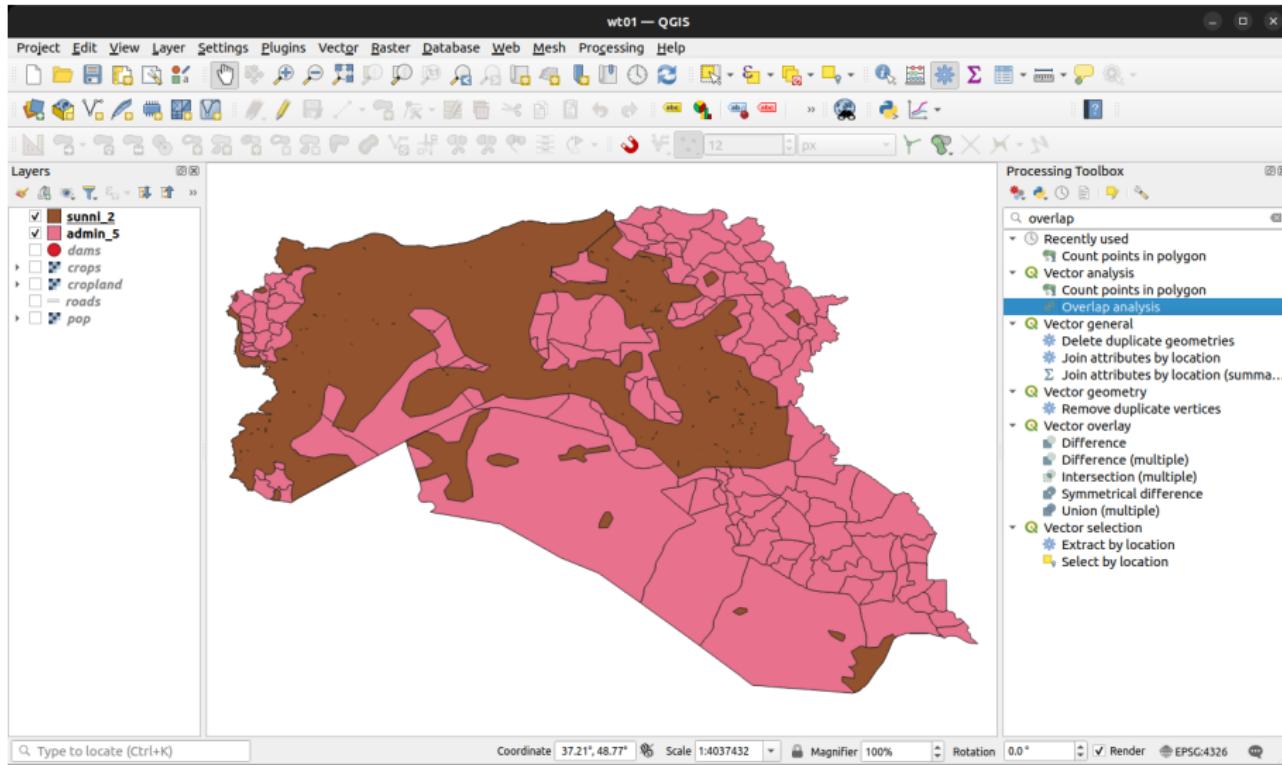
Open the Dissolve tool (Vector → Geoprocessing Tools → Dissolve).



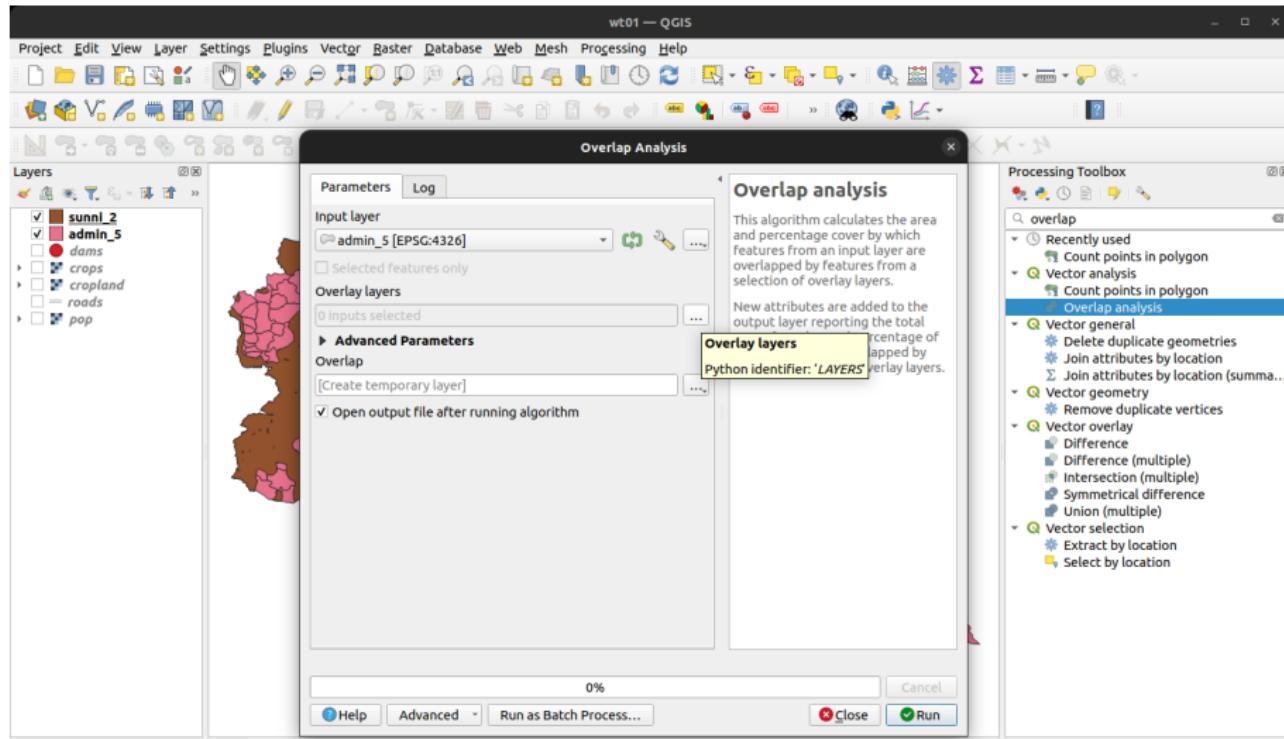
Set Input layer=sunni. Save the output as sunni\_2.geojson



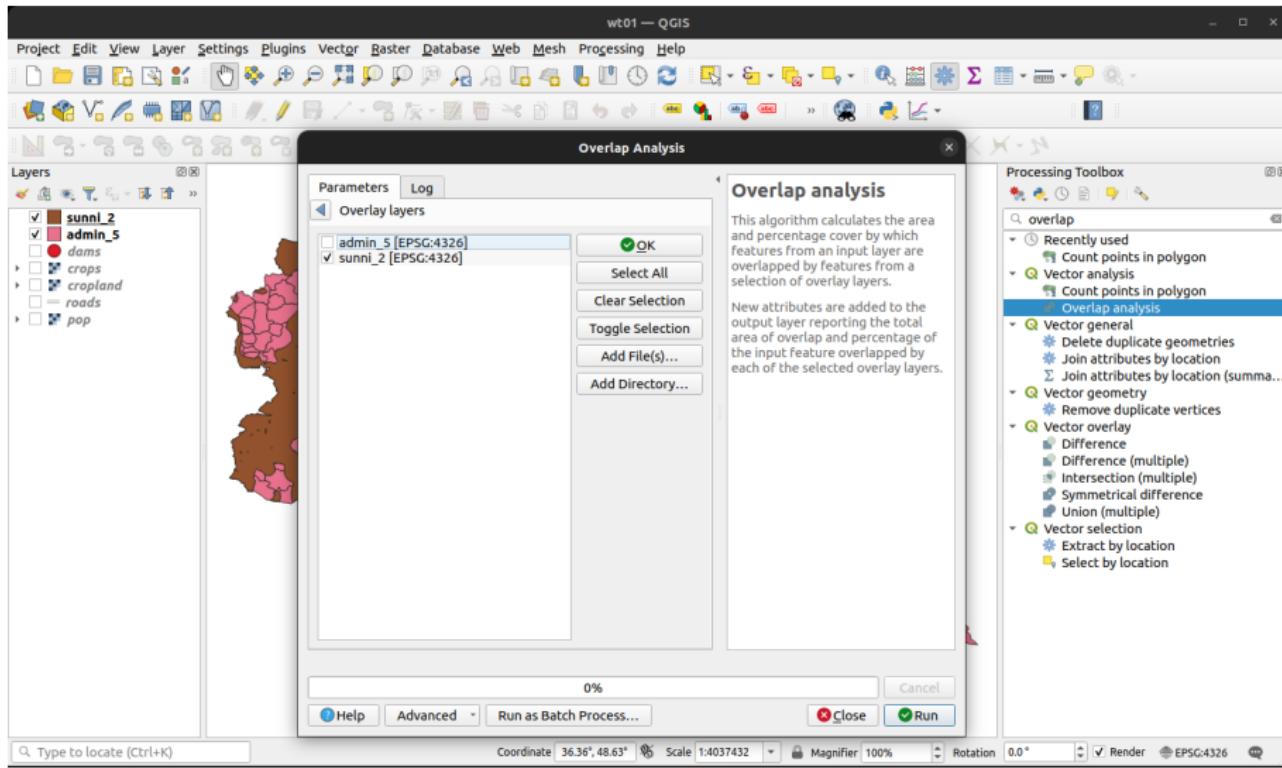
The dissolved polygon should appear in the project window. Now let's **calculate the proportion of each district populated by Sunnis**



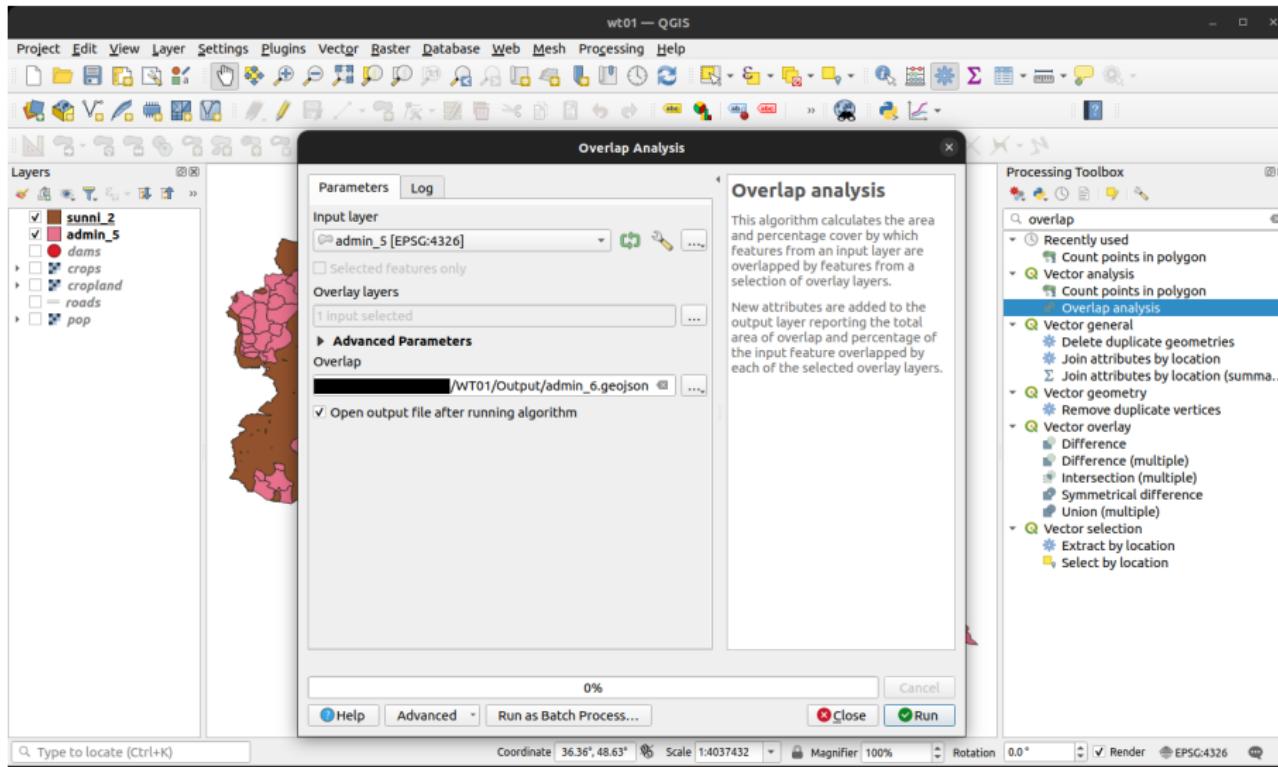
Open the Overlap Analysis tool (in Processing Toolbox → Vector Analysis). Set Input layer = admin\_5 and click the [...] button next to Overlay layers



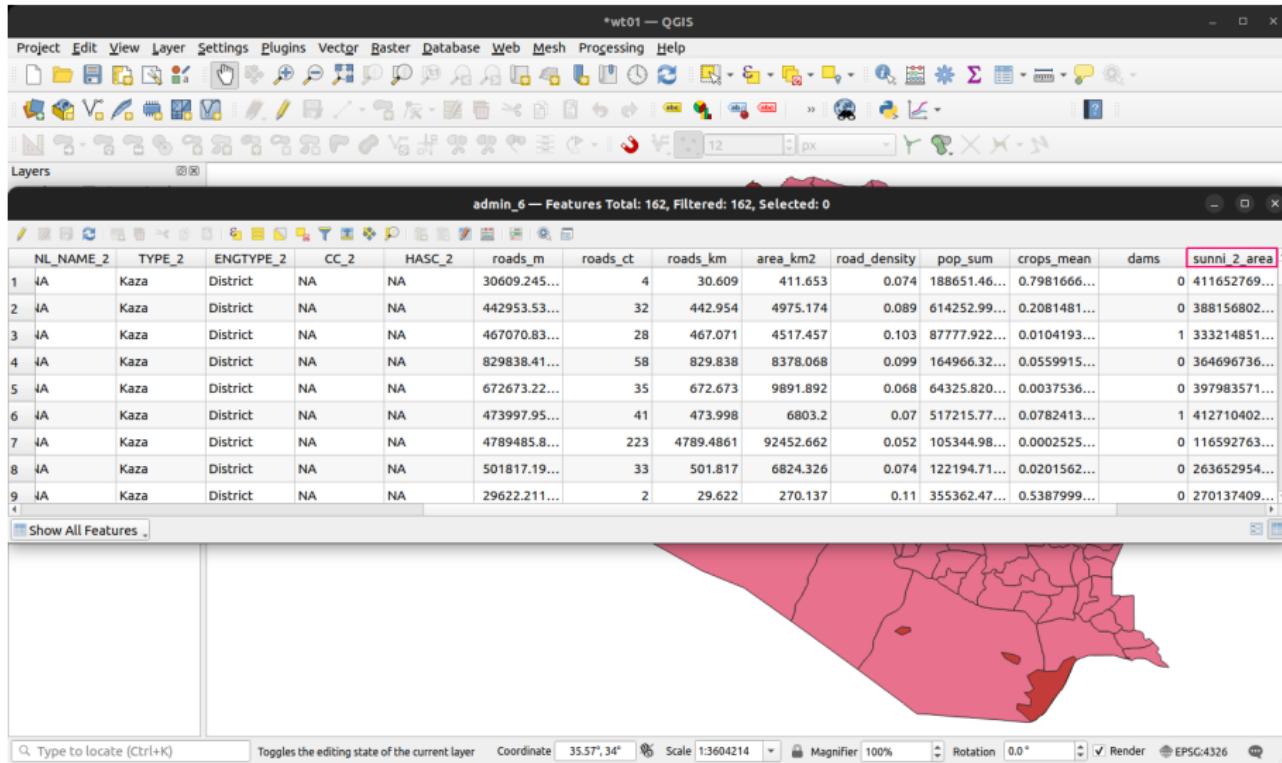
Check the box next to  sunni\_2. Click OK



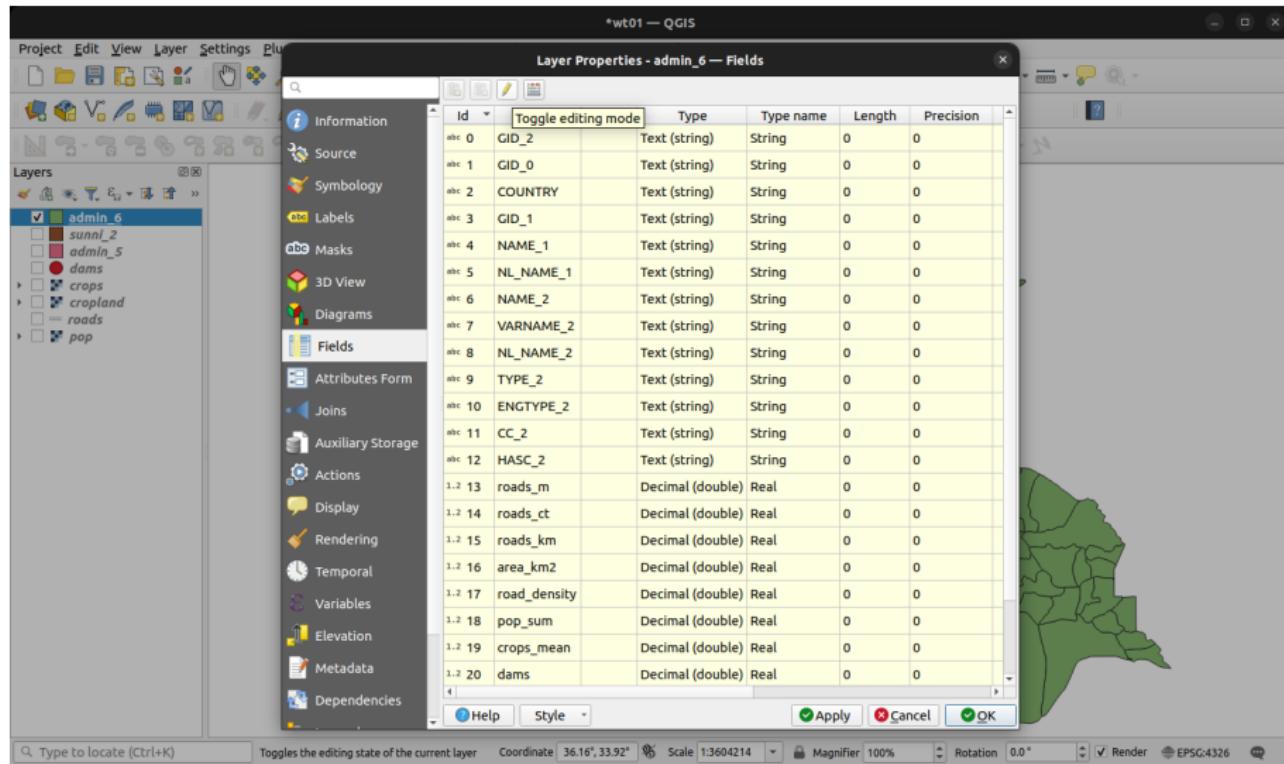
Save the output as admin\_6.geojson and click Run



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 "wt01 — QGIS". The left sidebar displays a tree view of layers: admin\_6 (checked), sunni\_2, admin\_5, dams, crops, cropland, roads, and pop. The main window shows a map of a region divided into green administrative units. A "Layer Properties - admin\_6 — Fields" dialog box is open in the center. The "Fields" tab is selected, showing 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 background. At the bottom of the dialog, there are buttons for Help, Style, Apply, Cancel, and OK.

# 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 several green administrative units. On the left, the "Layers" panel shows a tree structure with layers like "admin\_6", "sunni\_2", "admin\_5", etc. The "Fields" tab is selected in the sidebar. A dialog box titled "Layer Properties - admin\_6 - Fields" is open, showing a table of 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" 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.

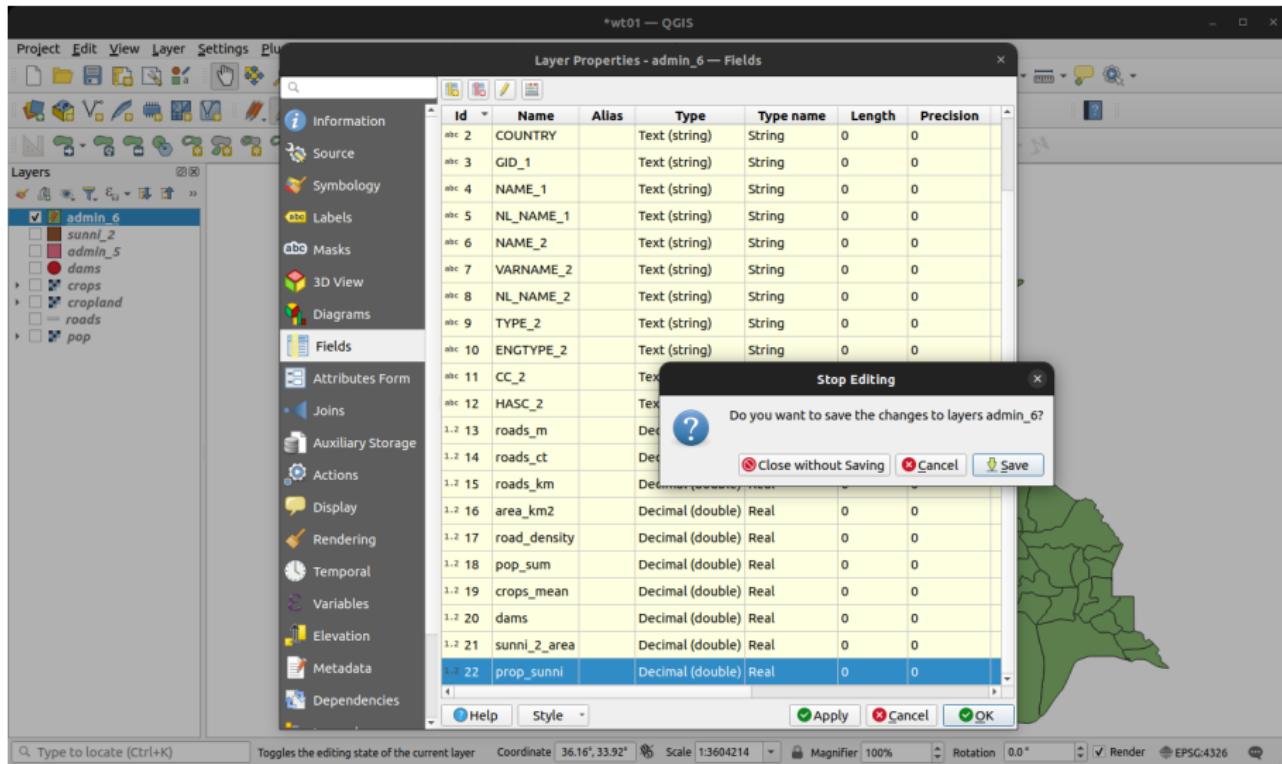
Hit Enter key (or Return key) to commit the name change. Click on the “pencil” again to leave editing mode

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:

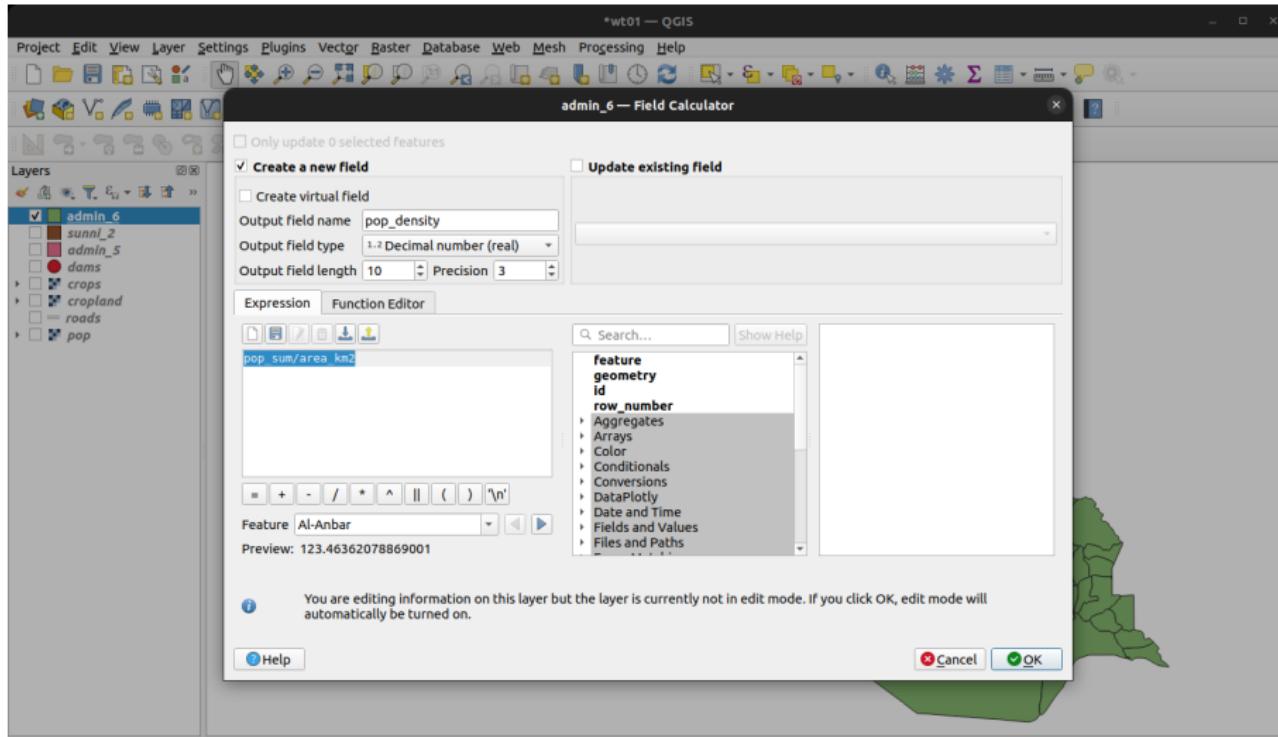
ID	Name	Type	Type name	Length	Precision
abc_2	COUNTRY	Text (string)	String	0	0
abc_3	GID_1	Text (string)	String	0	0
abc_4	NAME_1	Text (string)	String	0	0
abc_5	NL_NAME_1	Text (string)	String	0	0
abc_6	NAME_2	Text (string)	String	0	0
abc_7	VARNAME_2	Text (string)	String	0	0
abc_8	NL_NAME_2	Text (string)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

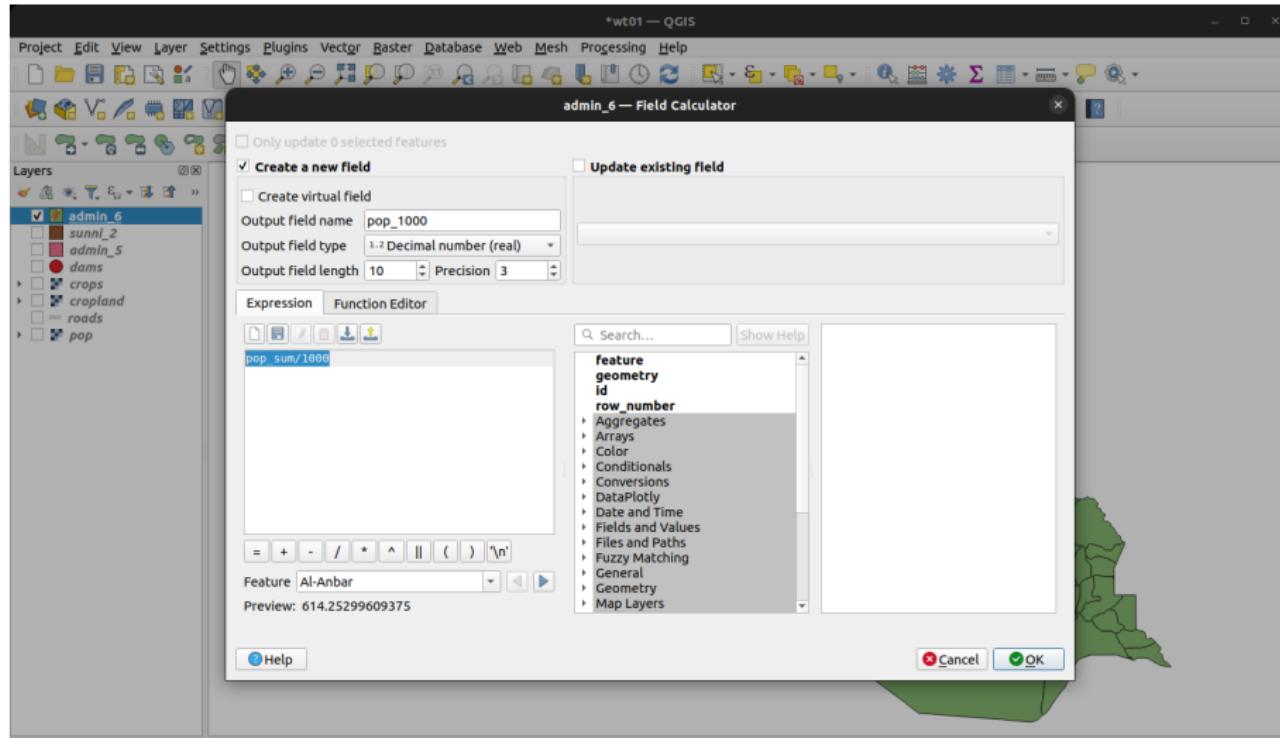
When prompted, save the changes you just made to admin\_6



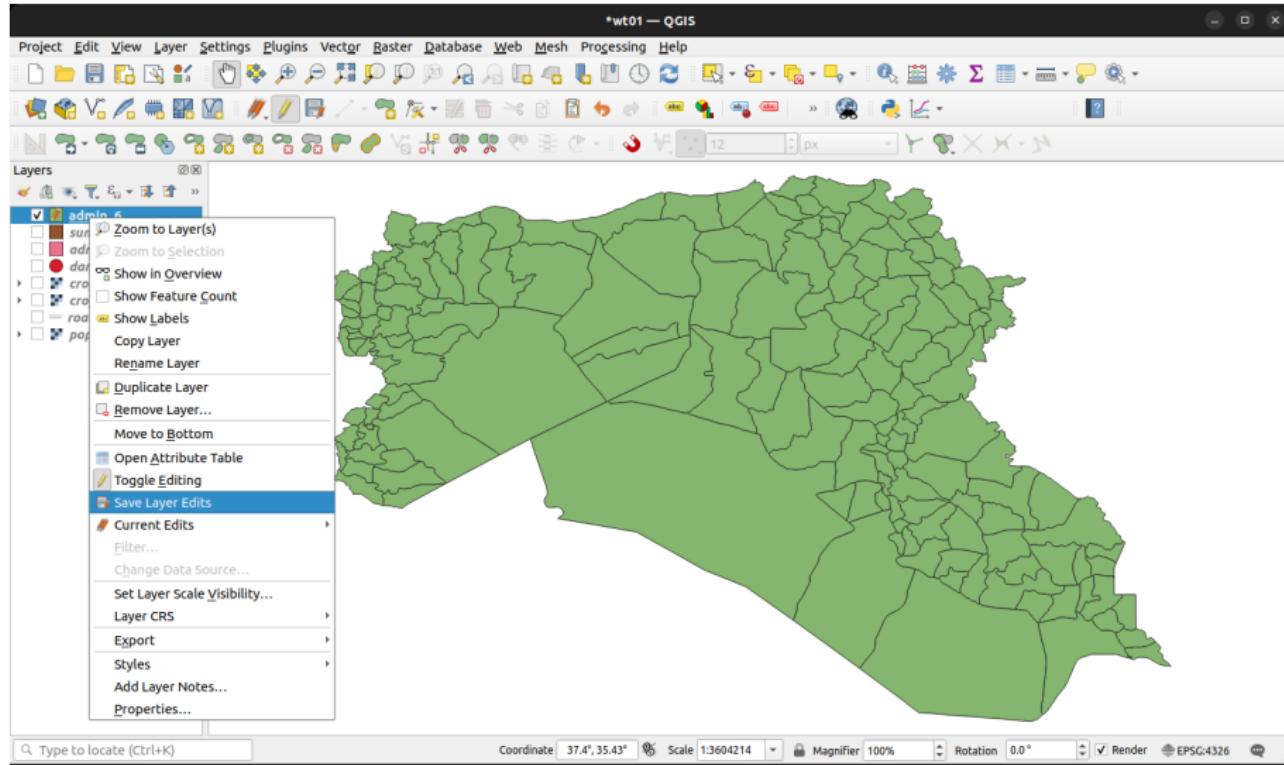
While we're at it, let's create a couple additional fields that could come in handy in the analysis. Create a **population density** field, named `pop_density`, of type Decimal number (real), with Expression set to `pop_sum/area_km2`



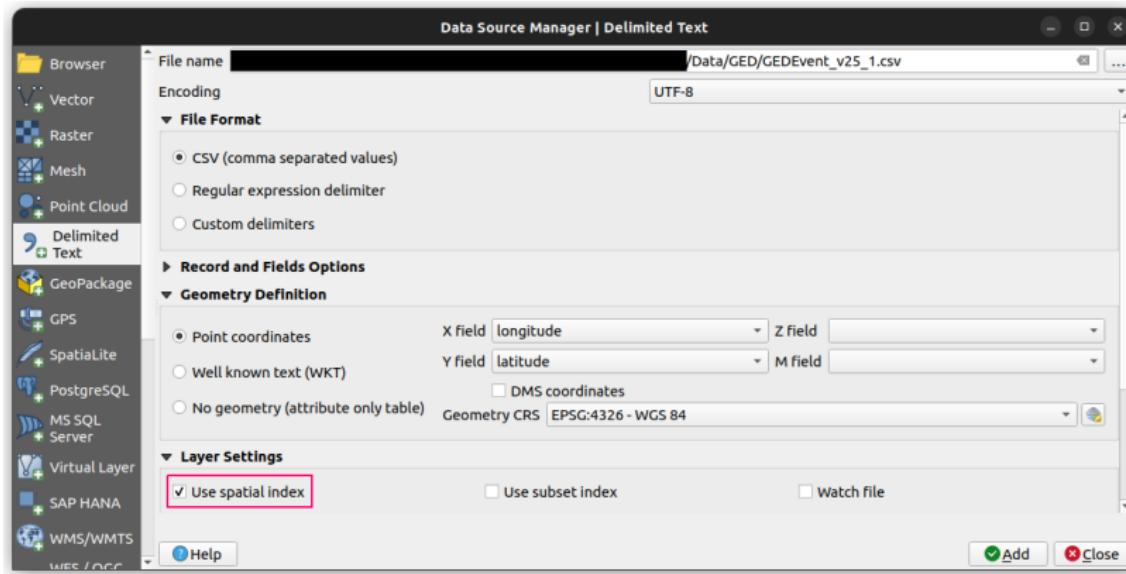
Let's also create a **rescaled population** field (1000's of residents). Name it **pop\_1000**, with type **Decimal number (real)**. Set Expression to **pop\_sum/1000**



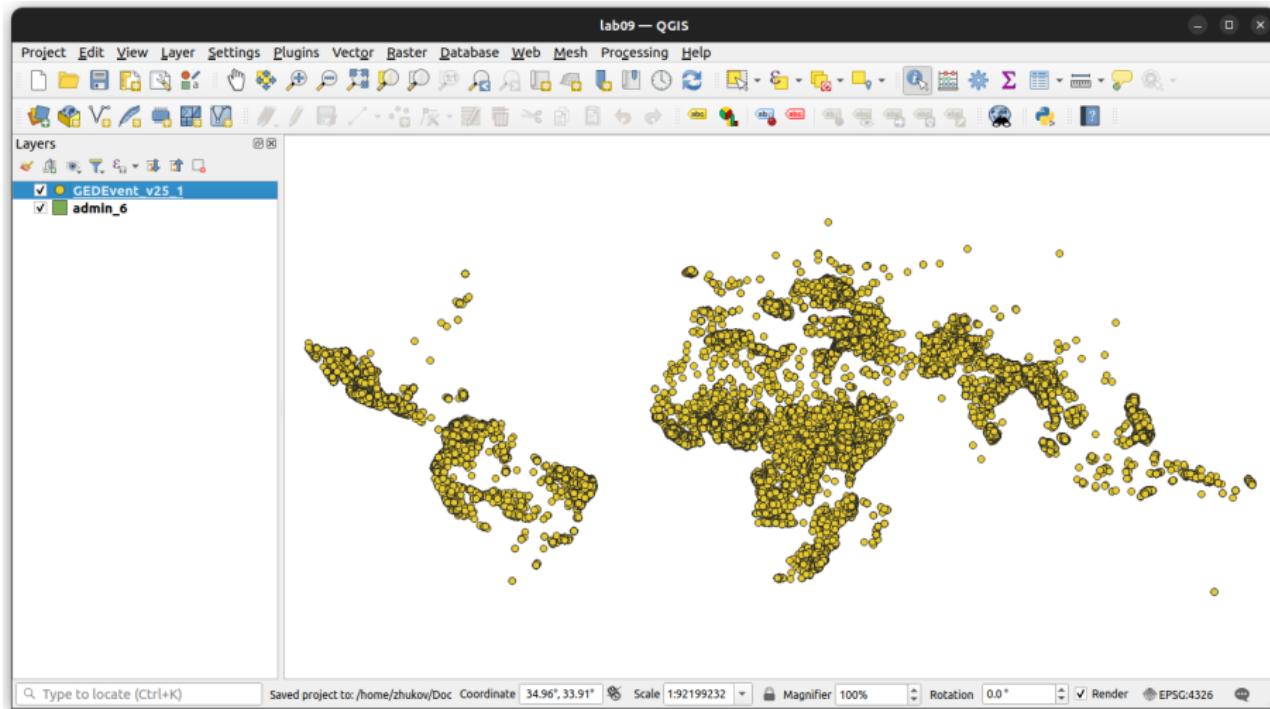
Save the layer edits to admin\_6!



Add the *Islamic State violence data* to the project, using Add Delimited Text Layer.... Load the GEDEvent\_v25\_1.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

GEDEvent\_v25\_1 — Features Total: 385918, Filtered: 385918, Selected: 0

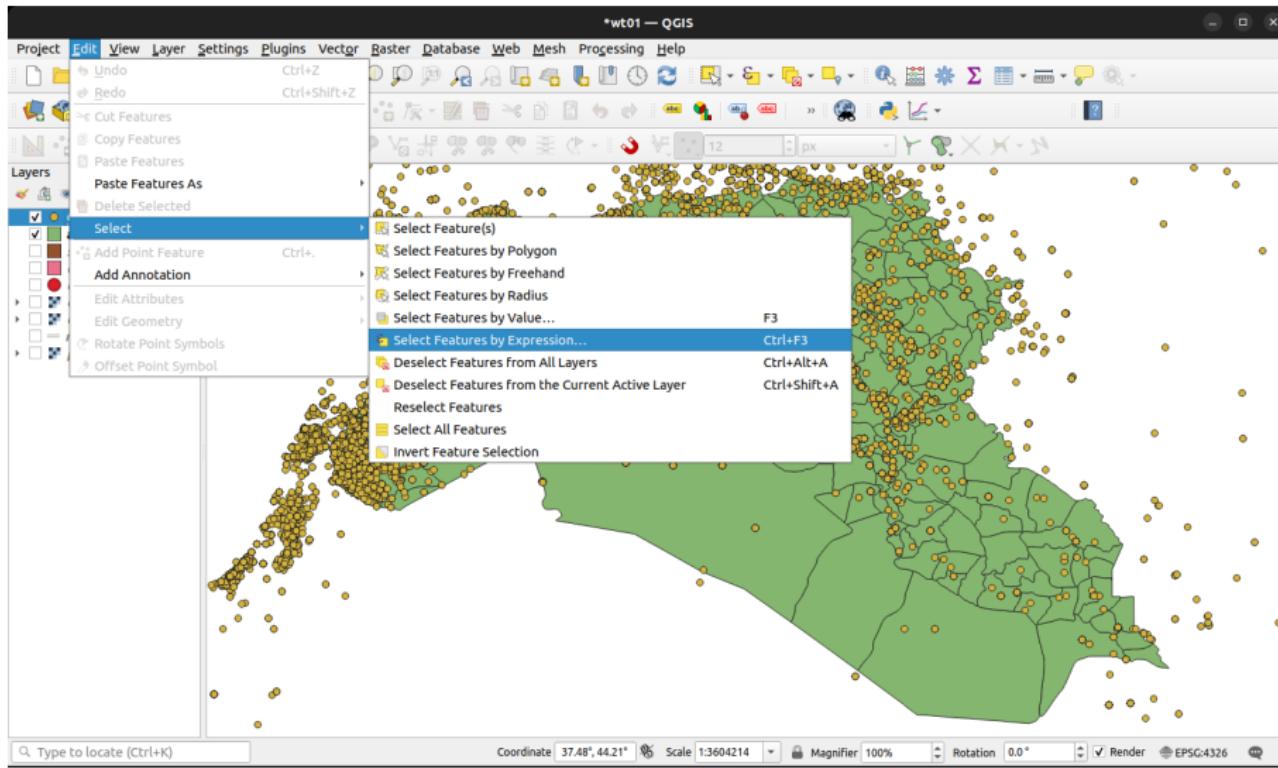
dset_id	dyad_new_id	dyad_name	side_a_dset_id	side_a_new_id	side_a	side_b_dset_id	side_b_new_id	side_b	number_of_sources	source_article	source_office	source_publisher
1	524	524 Governme...	116	116 Governme...	234	234	15		3	"Agence Fr...	Agence Fra...	
2	524	524 Governme...	116	116 Governme...	234	234	15		15	"BBC News...	BBC News;...	
3	524	524 Governme...	116	116 Governme...	234	234	15		5	"Khaama P...	Khaama Pr...	
4	524	524 Governme...	116	116 Governme...	234	234	15		8	"CNN,2021...	CNN;Reute...	
5	724	724 Governme...	130	130 Governme...	292	292	Jam'iyyat-i ...		-1	The Times ...	NULL	
6	724	724 Governme...	130	130 Governme...	292	292	Jam'iyyat-i ...		-1	Reuters 18 ...	NULL	
7	724	724 Governme...	130	130 Governme...	292	292	Jam'iyyat-i ...		-1	Reuters 24 ...	NULL	
8	724	724 Governme...	130	130 Governme...	292	292	Jam'iyyat-i ...		-1	The Sunday...	NULL	
9	724	724 Governme...	130	130 Governme...	292	292	Jam'iyyat-i ...		-1	St. Louis Po...	NULL	
10	724	724 Governme...	130	130 Governme...	292	292	Jam'iyyat-i ...		-1	Reuters 5 F...	NULL	
11	724	724 Governme...	130	130 Governme...	292	292	Jam'iyyat-i ...		-1	Reuters 8 F...	NULL	
12	724	724 Governme...	130	130 Governme...	292	292	Jam'iyyat-i ...		-1	The Washin...	NULL	
13	724	724 Governme...	130	130 Governme...	292	292	Jam'iyyat-i ...		-1	Washington	NULL	

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)

GEDEvent\_v25\_1 — Features Total: 385918, Filtered: 385918, Selected: 0

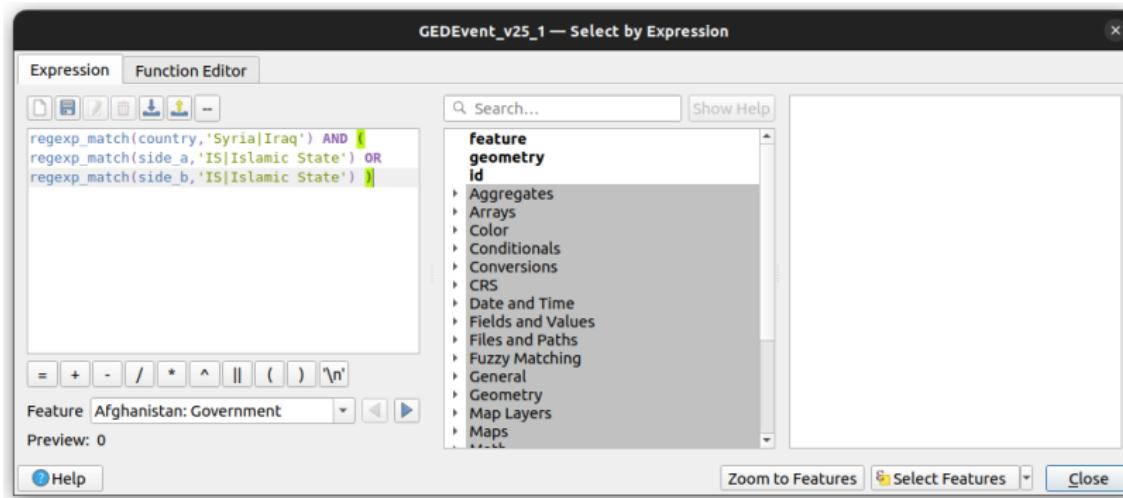
#	name	where_prec	where_coordinate	where_descript	adm_1	adm_2	latitude	longitude	geom_wkt	priogrid_gid	country	country_id
1	terior ...	1	Kabul city	Iraqi emba...	Kabul provi...	Kabul district	34.531094	69.162796	POINT (69....	179779	Afghanistan	700 Asia
2	fficials...	1	Kabul inter...	Kabul airpo...	Kabul provi...	Kabul district	34.564444	69.217222	POINT (69....	179779	Afghanistan	700 Asia
3	ntral ...	1	Jalalabad t...	Police Distr...	Nangarhar ...	Jalalabad d...	34.428844	70.45575	POINT (70....	179061	Afghanistan	700 Asia
4	ve; US...	1	Kabul city	Kabul city (...	Kabul provi...	Kabul district	34.531094	69.162796	POINT (69....	179779	Afghanistan	700 Asia
5	l source	4	Nangarhar ...	Nangarhar ...	Nangarhar ...	NULL	34.33333	70.41667	POINT (70....	179061	Afghanistan	700 Asia
6	l radio...	4	Kunduz pro...	Kunduz pro...	Kunduz pro...	NULL	36.75	68.75	POINT (68....	182658	Afghanistan	700 Asia
7	news ...	1	Salang pass	Salang pas...	Baghlan pr...	Khinjan dis...	35.315833	69.038889	POINT (69....	180499	Afghanistan	700 Asia
8	try sp...	1	Kabul city	Kabul city	Kabul provi...	Kabul district	34.531094	69.162796	POINT (69....	179779	Afghanistan	700 Asia
9	Ispok...	2	Salang tunnel	Salang tun...	Baghlan pr...	Khinjan dis...	35.31603	69.03871	POINT (69....	180499	Afghanistan	700 Asia
10	news ...	4	Hirat provi...	Hirat provi...	Hirat provi...	NULL	34.5	62	POINT (62 ...	179765	Afghanistan	700 Asia
11	news ...	1	Kabul city	Kabul city	Kabul provi...	Kabul district	34.531094	69.162796	POINT (69....	179779	Afghanistan	700 Asia
12	t offic...	1	Salang pass	Salang pass	Baghlan pr...	Khinjan dis...	35.315833	69.038889	POINT (69....	180499	Afghanistan	700 Asia
13	l ...	1	Kabul city	Kabul city	Kabul provi...	Kabul district	34.531094	69.162796	POINT (69....	179779	Afghanistan	700 Asia

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



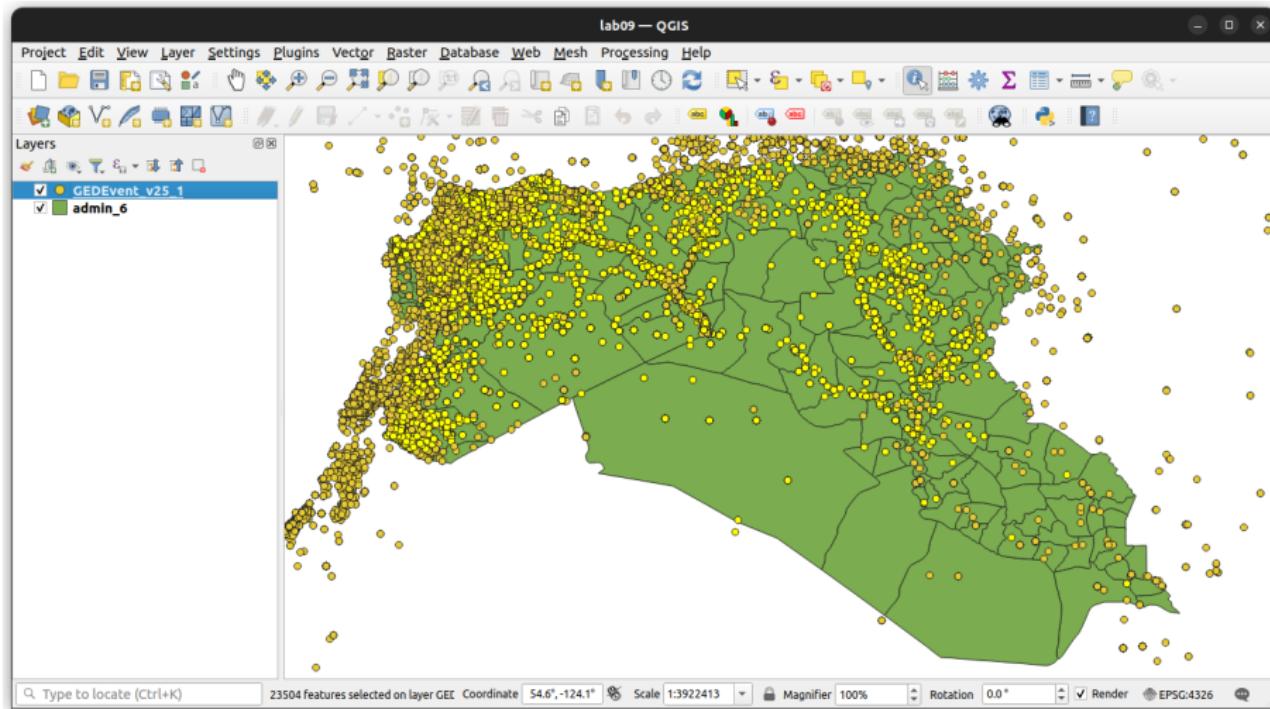
**Let's combine regular expressions with logical operators.** Set Expression to

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

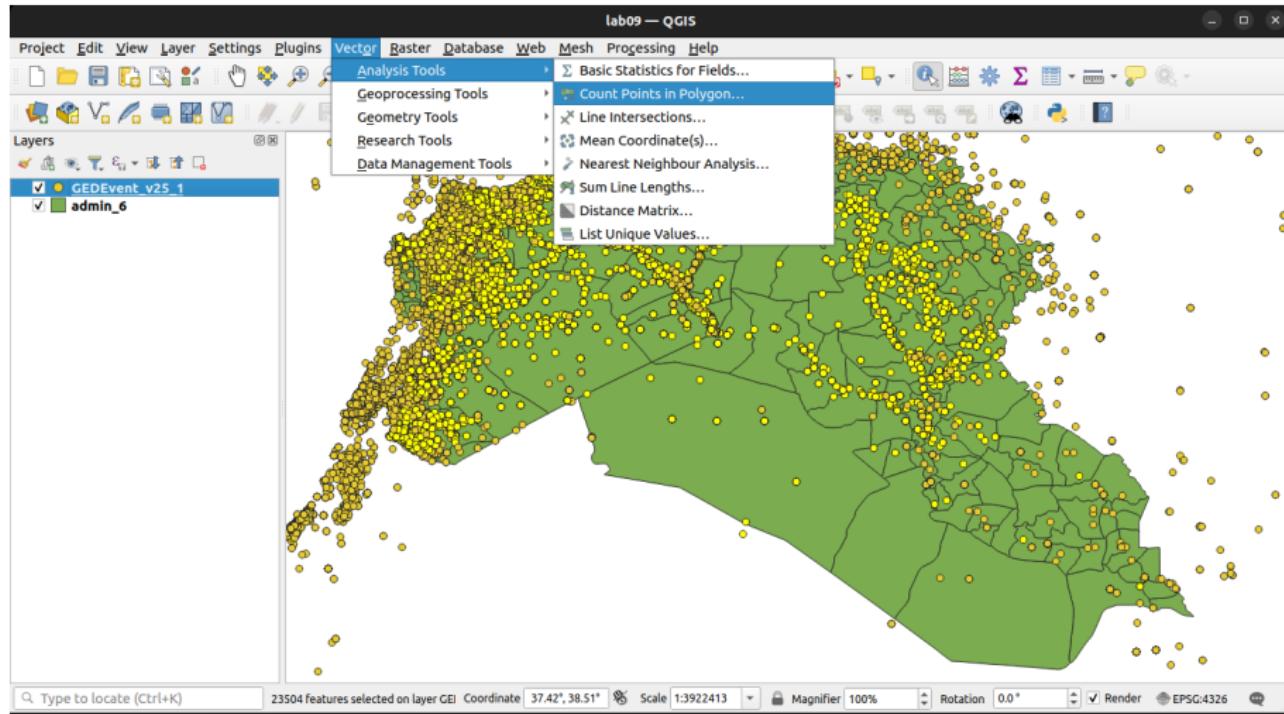


The vertical slash | is a regular expression for “OR”. So, in English this expression means ‘features where the field country contains “Iraq or Syria” and either side\_a or side\_b contains “IS or Islamic State” ’

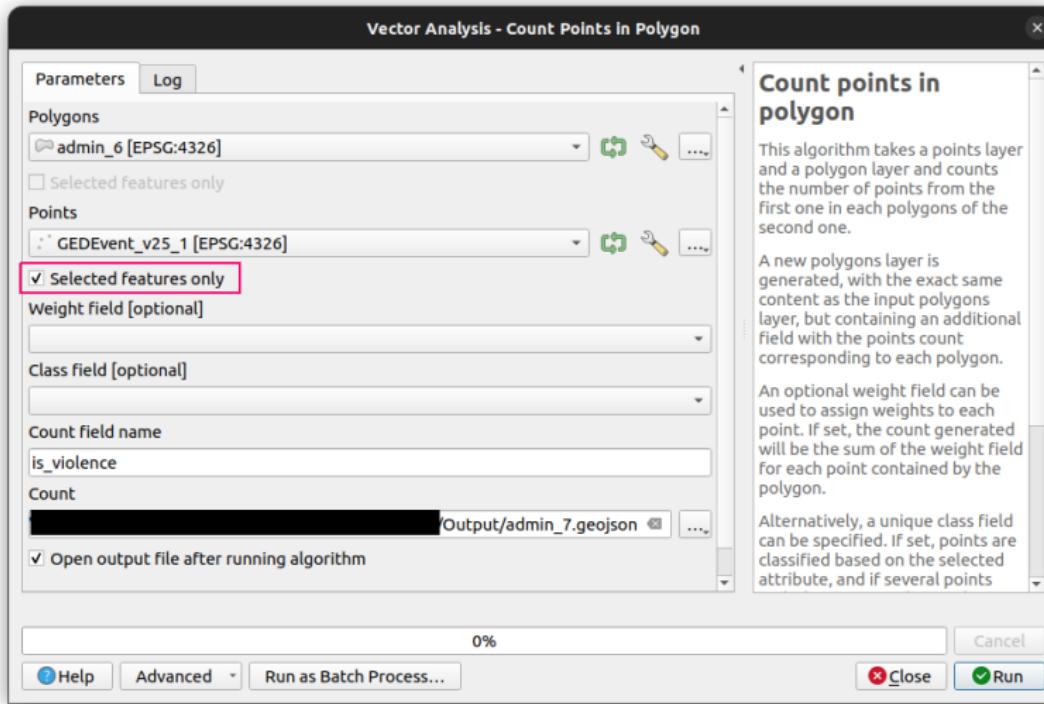
This should select a little over 23,000 events. Now let's calculate the **number of ISIS attacks per district**



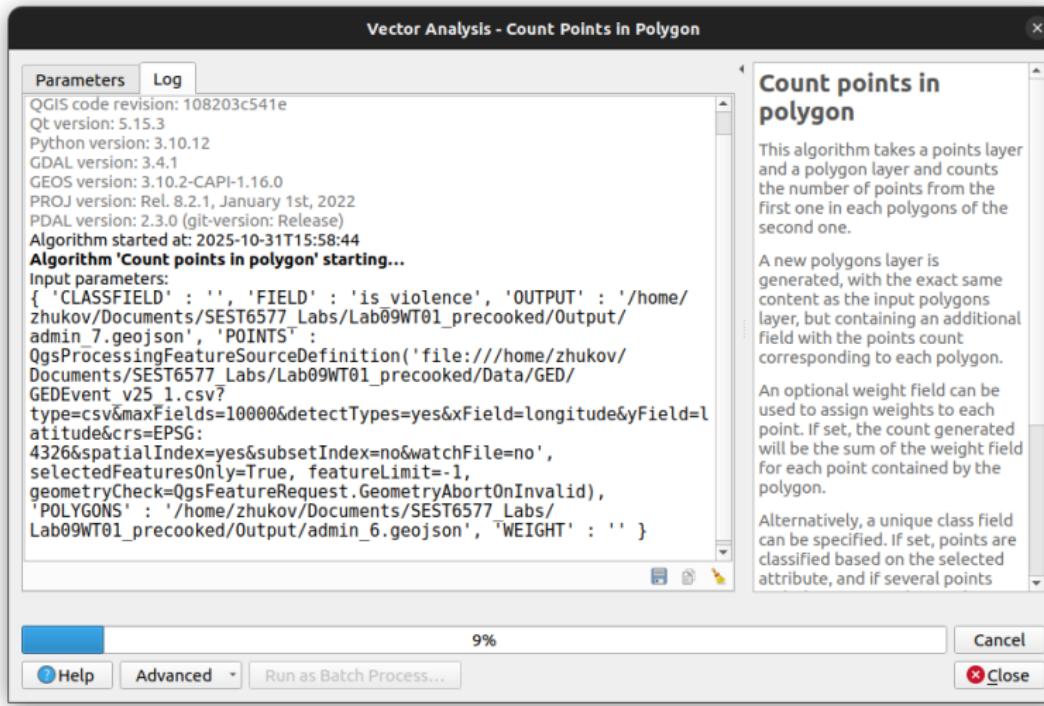
## Open the Count Points in Polygon tool



Select Polygons = admin\_6, Points = GEDEvent\_v25\_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

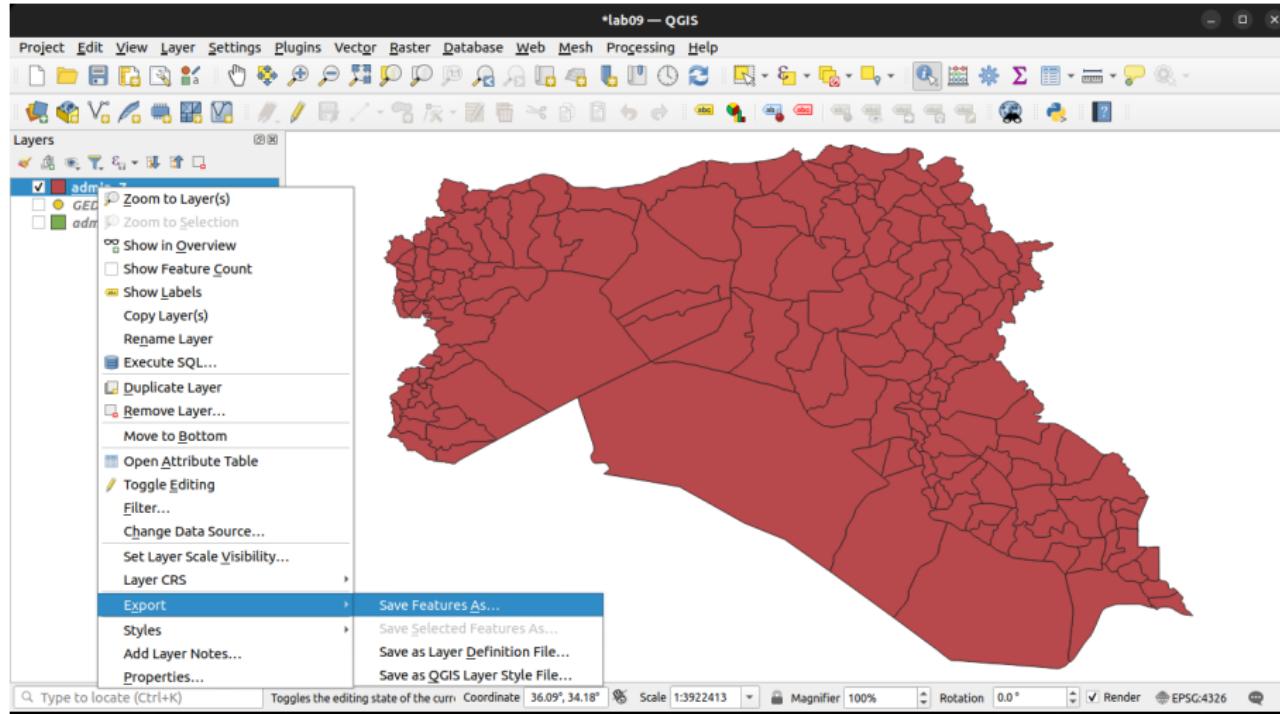


The attribute table for this new layer should include all the new fields we have generated. Let's now **export the table as CSV** for further analysis

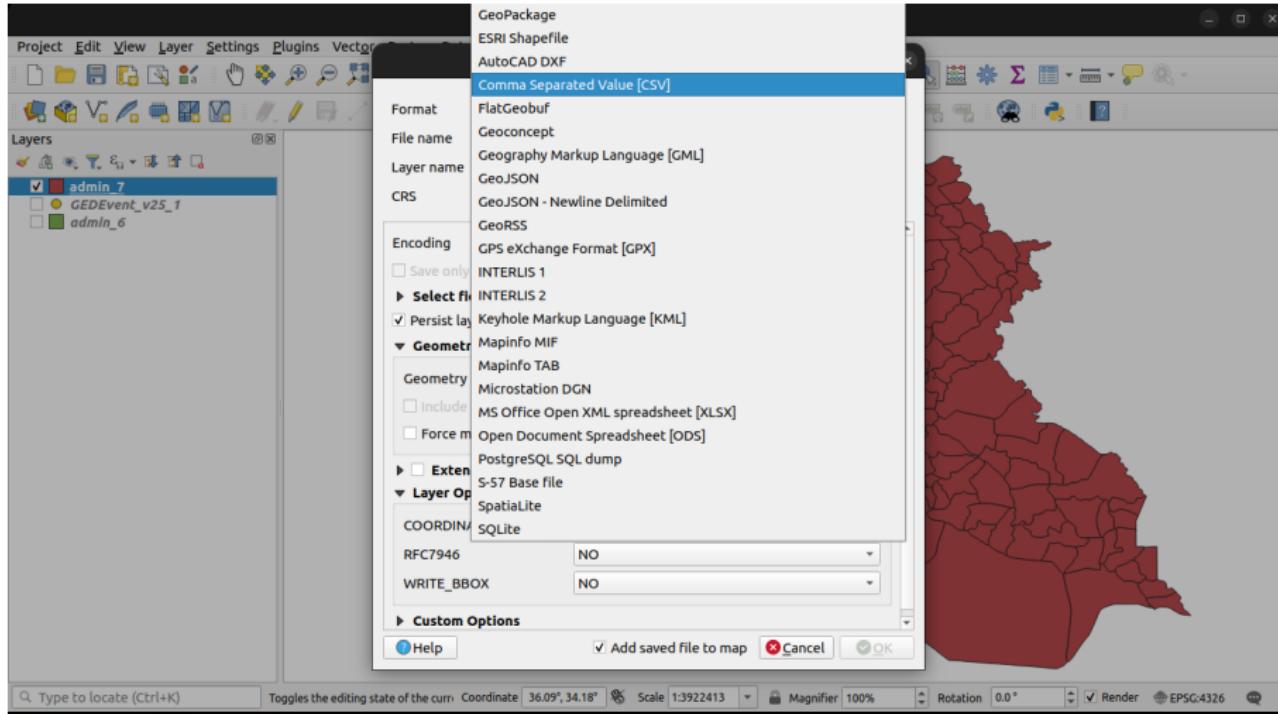
admin\_7 — Features Total: 162, Filtered: 162, Selected: 0

	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	...	4	30.609	411.653	0.074	151425.02...	0.8470610...	0	411652769...	99.999944...	367.846	151.425
2	...	32	442.954	4975.174	0.089	637757.85...	0.1972729...	0	388156802...	78.018732...	128.188	637.758
3	...	28	467.071	4517.457	0.103	116075.24...	0.0082528...	1	333214851...	73.761597...	25.695	116.075
4	...	58	829.838	8378.068	0.099	174373.33...	0.0584957...	0	364696736...	43.529933...	20.813	174.373
5	...	35	672.673	9891.892	0.068	53170.430...	0.0037176...	0	397983571...	40.233312...	5.375	53.17
6	...	41	473.998	6803.2	0.07	532509.94...	0.0789077...	1	412710402...	60.664154...	78.273	532.51
7	...	223	4789.486	92452.662	0.052	46729.013...	0.0002533...	0	116592763...	12.611077...	0.505	46.729
8	...	33	501.817	6824.326	0.074	166309.01...	0.0212115...	0	263652954...	38.634283...	24.37	166.309
9	...	2	29.622	270.137	0.11	292618.80...	0.6072135...	0	270137409...	100.00000...	1083.224	292.619
10	...	7	107.083	1299.549	0.082	257682.83...	0.0333288...	0	0	0	198.286	257.683
11	...	15	145.507	1251.084	0.116	66293.470...	0.0225227...	0	0	0	52.989	66.293
12	...	8	49.895	1020.216	0.049	228355.90...	0.1162594...	0	0	0	223.831	228.356
13	...	27	346.720	3267.500	0.122	216224.00...	0.0479620...	0	0	0	120.071	346.722

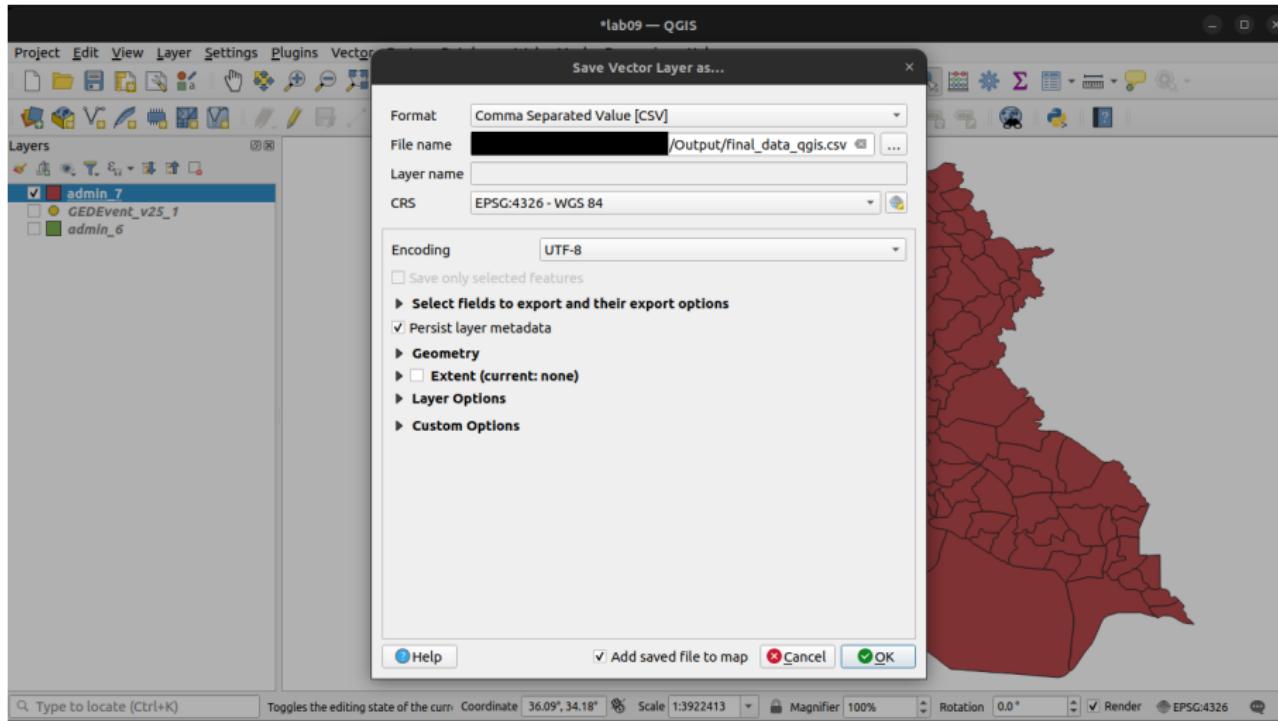
Right-click admin\_7 in the Layer menu, go to Export→Save Features As.



## 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 package

```
library(stargazer)
```

NOTE: The replication code for all of the preceding steps R is in `wt01_demo.R` on RStudio Cloud, and in `Lab09WT01.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"         "layer"          "path"
## [16] "roads_m"         "roads_ct"        "roads_km"       "area_km2"      "road_density"
## [21] "pop_sum"         "crops_mean"      "dams"           "sunni_2_area"   "prop_sunni"
## [26] "pop_density"    "pop_1000"        "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

- $\text{violence}_i$  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`)
- $\beta$  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 
## -4.5068 -1.0639 -0.1906  0.9280  3.8392 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 1.4918950  0.3455419   4.318 2.79e-05 ***
## road_density -1.4859463  3.1746426  -0.468 0.640390    
## pop_1000     0.0009083  0.0002421   3.752 0.000247 ***  
## crops_mean   -1.4268796  0.4836122  -2.950 0.003662 **  
## dams         1.1105982  0.5346231   2.077 0.039408 *   
## prop_sunni   0.0363425  0.0029937  12.140 < 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) 4.3622385  0.5788847   7.536 3.70e-12 ***
## road_density -9.2950240  5.8355866  -1.593  0.11323
## pop_1000      0.0004305  0.0001670   2.578  0.01087 *
## crops_mean    -2.1724732  0.7841517  -2.770  0.00628 **
## dams          0.8329988  0.4094748   2.034  0.04361 *
## prop_sunni    0.0302367  0.0051981   5.817 3.29e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 429.0671)
##
## Null deviance: 71658  on 161  degrees of freedom
## Residual deviance: 38816  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       -1.486        -9.295       -1.845      -12.958***  
##                   (3.175)        (5.836)       (3.525)      (4.943)  
##  
## pop_1000            0.001***     0.0004**     0.002***     0.001***  
##                   (0.0002)        (0.0002)       (0.0003)      (0.0002)  
##  
## crops_mean         -1.427***     -2.172***     -2.204***    -1.721***  
##                   (0.484)        (0.784)       (0.485)      (0.644)  
##  
## dams                1.111**      0.833**      0.647       0.603*  
##                   (0.535)        (0.409)       (0.430)      (0.315)  
##  
## prop_sunni          0.036***     0.030***     0.017***     0.016***  
##                   (0.003)        (0.005)       (0.005)      (0.005)  
##  
## -----  
## Province FE          N             N             Y             Y  
## Observations        162          162          162          162  
## R2                  0.569          0.819  
## Adjusted R2          0.555          0.767  
## Residual Std. Error   1.547 (df = 156)          1.121 (df = 125)  
## F Statistic          41.196*** (df = 5; 156)      15.695*** (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	$\beta_{\text{prop\_sunni}} > 0$	✓	✓