Module: Vector Analysis

Now that you have edited a few features, you must want to know what else one can do with them. Having features with attributes is nice, but when all is said and done, this doesn't really tell you anything that a normal, non-GIS map can't.

The key advantage of a GIS is this: *a GIS can answer questions.*

For the next three modules, we'll endeavor to answer a *research question* using GIS functions. Suppose you are a property developer, and you want to build a new residential development. You need to buy a farm to build it on. Your market research has given you the following criteria. The farm must:

- be at least a certain area in size;
- be within the Swellendam town region;
- have a road running through it for easy access;
- be within a certain distance from a school;
- be on a flat surface for easy construction of new buildings; and
- stay out of the shadow of the mountains in winter.

Within the next few modules, we'll harness the power of GIS analysis tools to locate suitable farm properties for this new residential development.

7.1 Lesson: Reprojecting and Transforming Data

Let's talk about Coordinate Reference Systems (CRSs) again. We've touched on this briefly before, but haven't discussed what it means practically.

The goal for this lesson: To reproject and transform vector datasets.

7.1.1 Follow along: Projections

The CRS that all the data as well as the map itself are in right now is called WGS84. This is a very common Geographic Coordinate System (GCS) for representing data. But there's a problem, as we will see.

- Save your current map.

- Then open the map of the world which you'll find under `exercise_data/world/world.qgs`.

- Zoom in to South Africa by using the *Zoom In* tool.

- Try setting a scale in the *Scale* field, which is in the *Status Bar* along the bottom of the screen. While over South Africa, set this value to `1:5000000` (one to five million).

- Pan around the map while keeping an eye on the *Scale* field.

Notice the scale changing? That's because you're moving away from the one point that you zoomed into at `1:5000000`, which was at the center of your screen. All around that point, the scale is different.

To understand why, think about a globe of the Earth. It has lines running along it from North to South. These longitude lines are far apart at the equator, but they meet at the poles.

In a GCS, you're working on this sphere, but your screen is flat. When you try to represent the sphere on a flat surface, distortion occurs, similar to what would happen if you cut open a tennis ball and tried to flatten it out. What this means on a map is that the longitude lines stay equally far apart from each other, even at the poles (where they are supposed to meet). This means that, as you travel away from the equator on your map, the scale of the objects that you see gets larger and larger. What this means for us, practically, is that there is no constant scale on our map!

To solve this, let's use a Projected Coordinate System (PCS) instead. A PCS "projects" or converts the data in a way that makes allowance for the scale change and corrects it. Therefore, to keep the scale constant, we should reproject our data to use a PCS.

7.1.2 Follow along: "On the Fly" Reprojection

QGIS allows you to reproject data "on the fly". What this means is that even if the data itself is in another CRS, QGIS can project it as if it were in a CRS of your choice.

- To enable "on the fly" projection, click on the *CRS Status* button in the *Status Bar* along the bottom of the QGIS window:

- In the dialog that appears, check the box next to *Enable 'on the fly' CRS transformation*.

- Type the word `global` into the *Filter* field. One CRS (*NSIDC EASE-Grid Global*) will appear in the list below.

- Click on it to select it, then click `OK`.

- Notice how the shape of South Africa changes. All projections work by changing the apparent shapes of objects on Earth.

- Zoom in to a scale of `1:5000000` again, as before.

- Pan around the map.

- Notice how the scale stays the same!

"On the fly" reprojection is also used for combining datasets that are in different CRSs.

- Deactivate "on the fly" reprojection again:

  – Click on the *CRS Status* button again.

  – Uncheck the *Enable 'on the fly' CRS transformation* box.

  – Clicking *OK*.

- Add another vector layer to your map which has the data for South Africa only. You'll find it as `exercise_data/world/RSA.shp`.

What do you notice?

The layer isn't visible! But that's easy to fix, right?

- Right-click on the layer in the *Layers list*.

- Select *Zoom to Layer Extent*.

OK, so now we see South Africa... but where is the rest of the world?

It turns out that we can zoom between these two layers, but we can't ever see them at the same time. That's because their Coordinate Reference Systems are so different. The *continents* dataset is in *degrees*, but the *RSA* dataset is in *meters*. So, let's say that a given point in Cape Town in the *RSA* dataset is about `4 100 000` meters away from the equator. But in the *continents* dataset, that same point is about `33.9` degrees away from the equator.

This is the same distance - but QGIS doesn't know that! You haven't told it to reproject the data. So as far as it's concerned, the version of South Africa that we see in the *RSA* dataset has Cape Town at the correct distance of `4 100 000` meters from the equator. But in the *continents* dataset, Cape Town is only `33.9` *meters* away from the equator! You can see why this is a problem.

QGIS doesn't know where Cape Town is *supposed* to be - that's what the data should be telling it. If the data tells QGIS that Cape Town is `34` meters away from the equator and that South Africa is only about `12` meters from north to south, then that is what QGIS will draw.

To correct this:

- Switch *Enable 'on the fly' CRS transformation* on again as before.

- Zoom to the extents of the *RSA* dataset.

Now, because they're made to project in the same CRS, the two datasets fit perfectly:
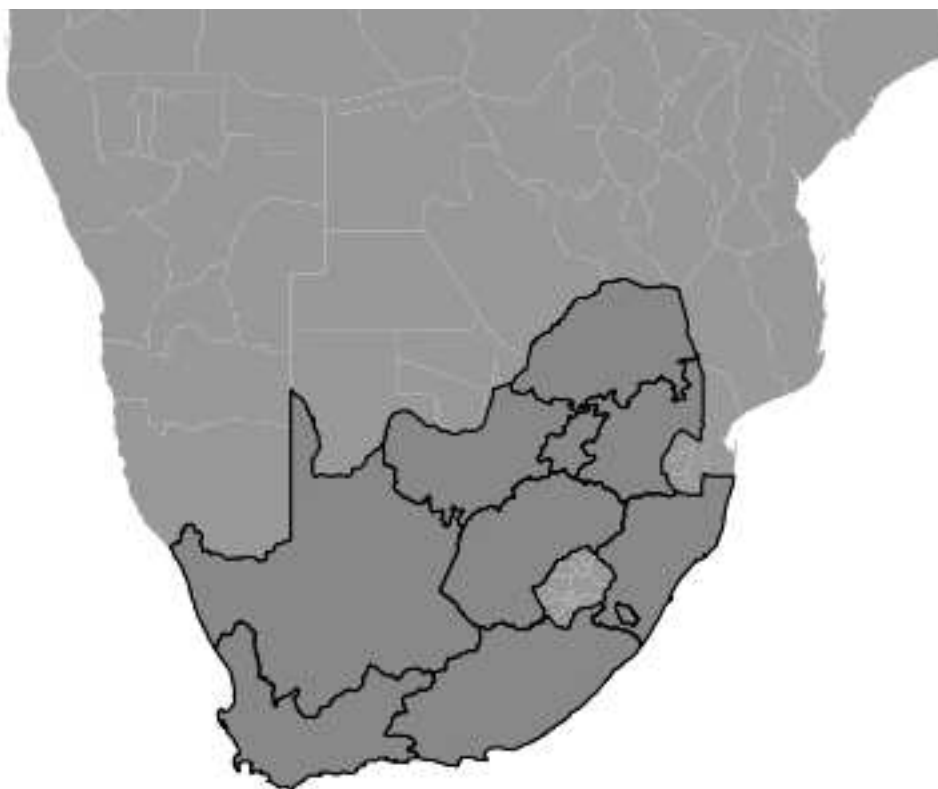
When combining data from different sources, it's important to remember that they might not be in the same CRS. "On the fly" reprojection helps you to display them together.

7.1.3 Follow along: Saving a Dataset to Another CRS

Remember when you calculated areas for the farms in the *Classification* lesson? You did it so that you could classify the farms according to area.

- Open your usual map again (containing the Swellendam data).
- Open the attribute table for the *rural* layer.
- Scroll to the right until you see the `AREA` field.

Notice how the areas are all very small; basically zero. This is because these areas are given in degrees - the data isn't in a Projected Coordinate System. In order to calculate the area for the farms in square meters, the data has to be in square meters as well. So, we'll need to reproject it.

But it won't help to just use "on the fly" reprojection. "On the fly" does what it says - it doesn't change the data, it just reprojects the layers as they appear on the map. To truly reproject the data itself, you need to export it to a new file using a new projection.

- Right-click on the *rural* layer in the *Layers list*.

- Select *Save As...* in the menu that appears. You will be shown the *Save vector layer as...* dialog.

- Click on the *Browse* button next to the *Save as* field.

- Navigate to `exercise_data/` and specify the name of the new layer as `rural_reprojected.shp`.

- Leave the *Encoding* unchanged.

- Change the value of the *Layer CRS* dropdown to *Selected CRS*.

- Click the *Browse* button beneath the dropdown.

- The *CRS Selector* dialog will now appear.

- In its *Filter* field, search for `33S`.

- Choose *WGS 84 / UTM zone 33S* from the list.

- Click *OK*.

The *Save vector layer as...* dialog now looks like this:
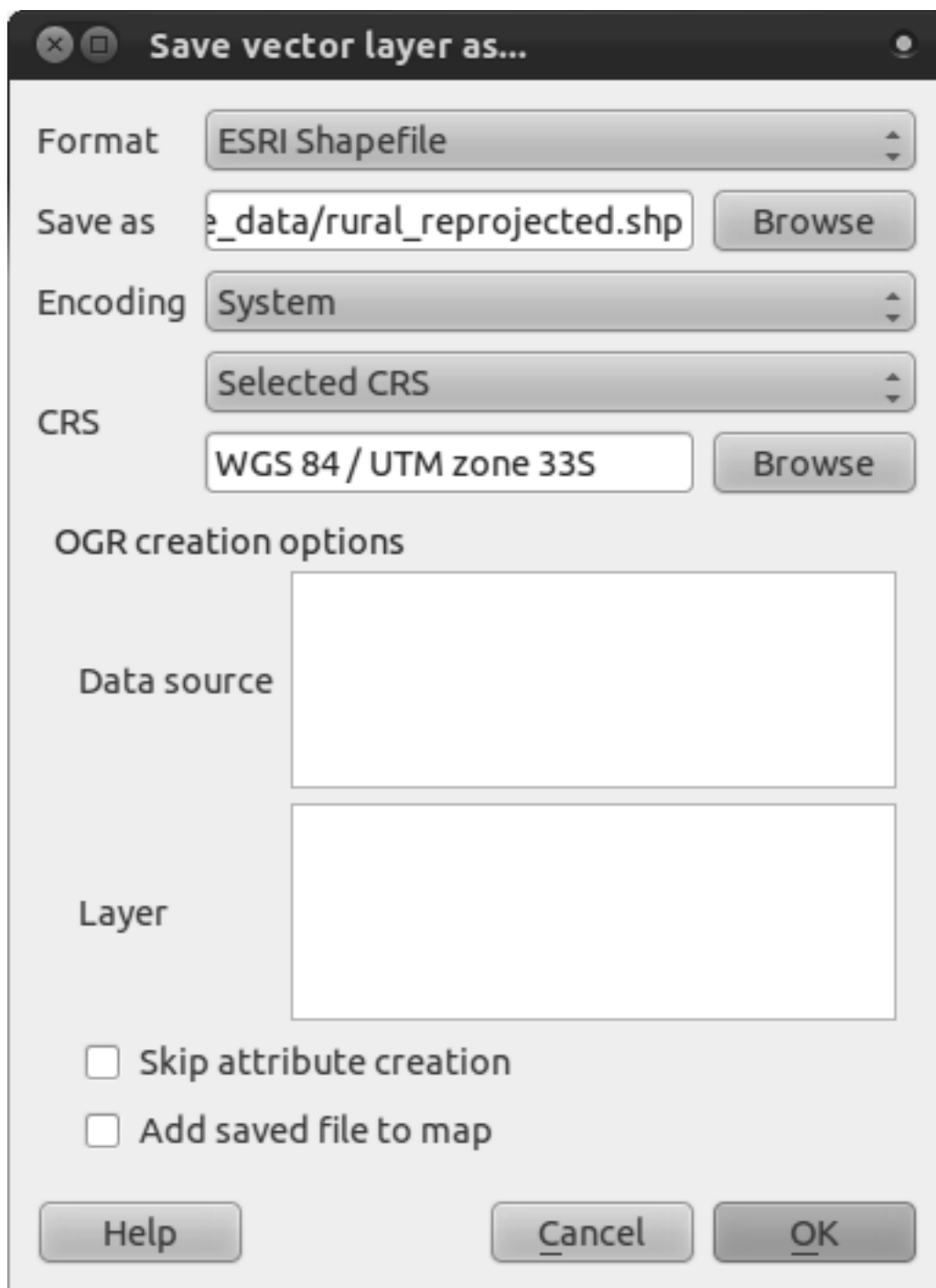
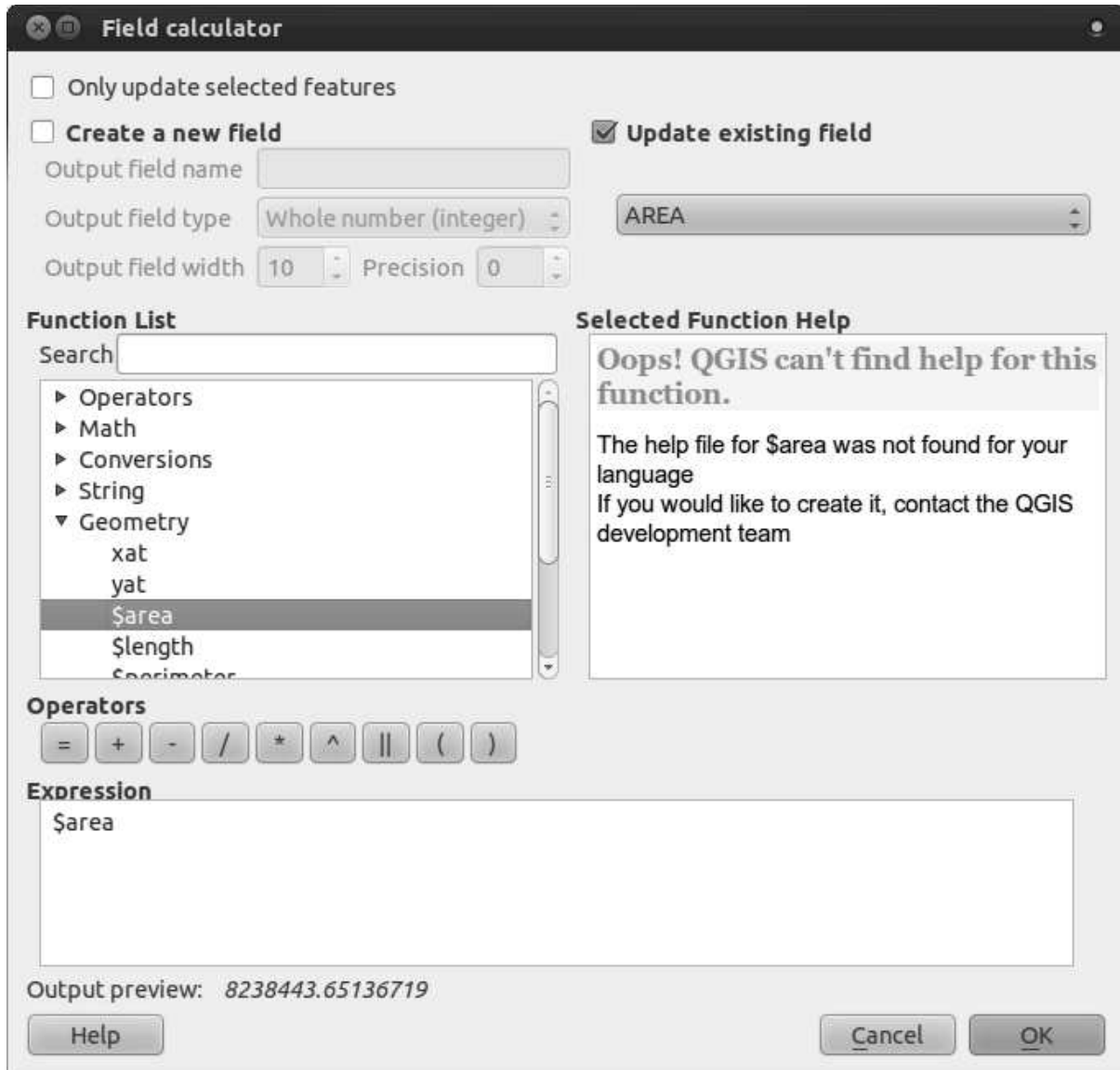- Click *OK* and after an instant, you should be presented with:
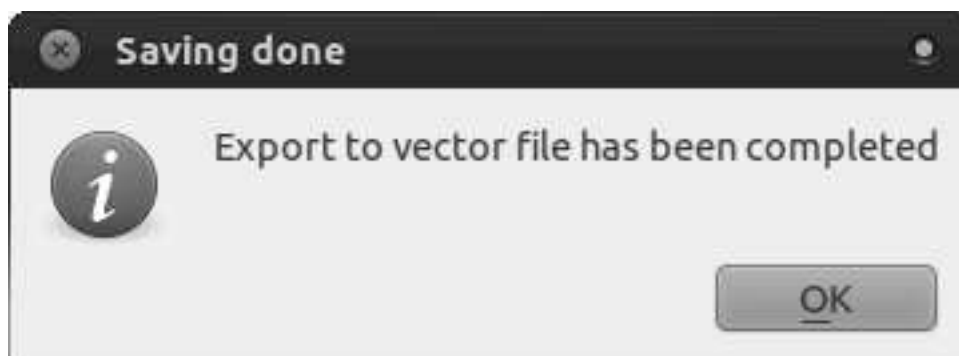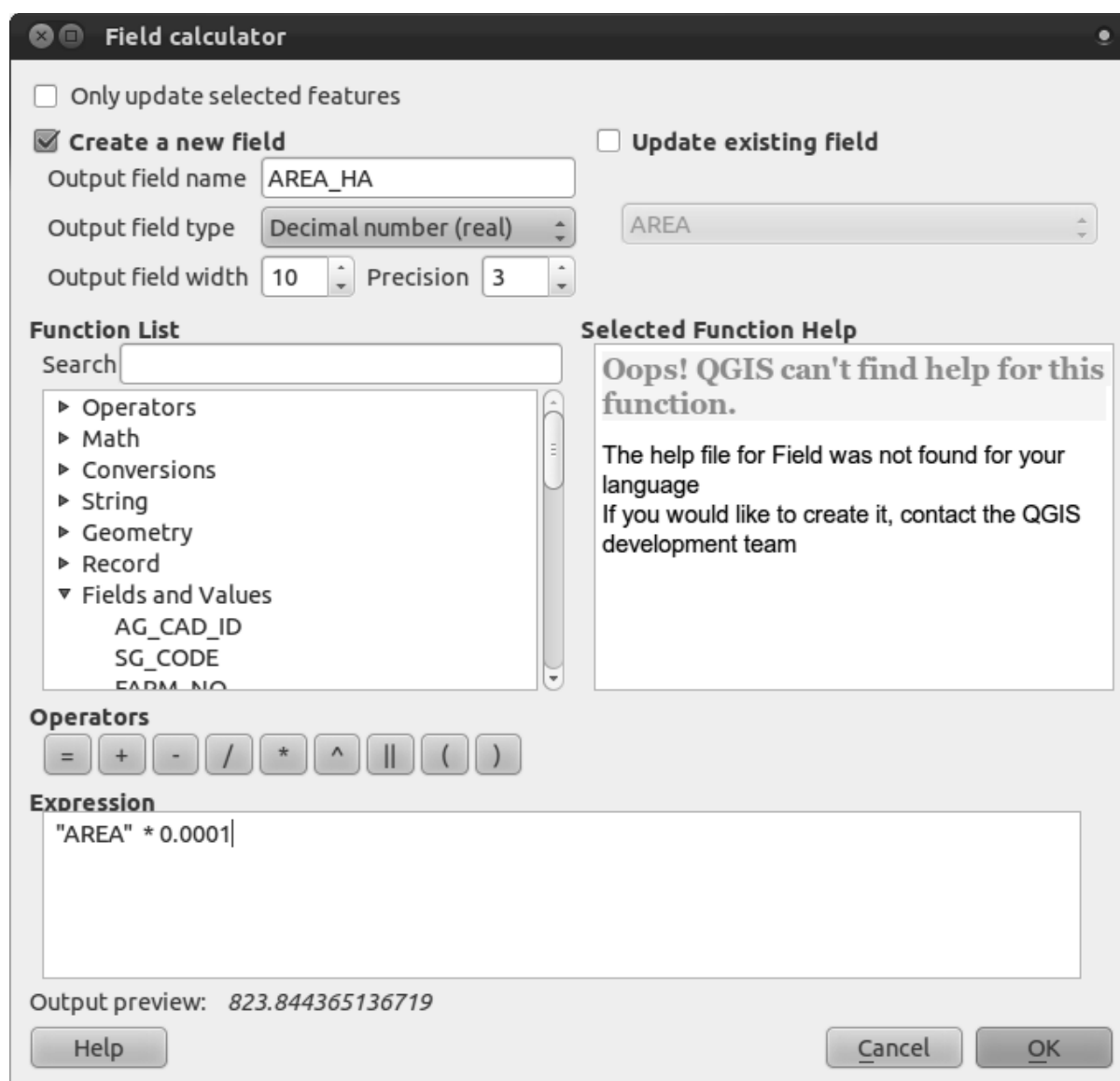
- Click *OK*.

- Start a new map:

Refer back to the lesson on *Classification* to remember how you calculated areas.

- Update the AREAS field by running the same expression as before:

This will update the AREAS field with the areas of the farms in square meters.

- To calculate the area in hectares, do this:

Look at the new values in your attribute table. This is much more useful, as people actually quote property areas in hectares, not in degrees. And projecting the data in an appropriate projection before calculating the area will actually give you the area in hectares. This is why it's a good idea to reproject your data, if necessary, before calculating areas, distances, and other values that are dependent on the spatial properties of the layer.

7.1.4 Follow along: Creating Your Own Projection

There are many more projections than just those included in QGIS by default. You can also create your own projections.

- Start a new map.
- Load the `world/oceans.shp` dataset.
- Go to *Settings → Custom CRS...* and you'll see this dialog:

- Click on the button with the star icon to create a new projection. You'll notice that the name and parameters are now blank.

An interesting projection to use is called `Van der Grinten I.`

- Enter its name in the *Name* field.

This projection represents the Earth on a circular field instead of a rectangular one, as most other projections do.

- For its parameters, use the following string:

```
+proj=vandg +lon_0=0 +x_0=0 +y_0=0 +R_A +a=6371000 +b=6371000
+units=m +no_defs
```

- Click the *Save* button:

- Click *OK*.

- Enable "on the fly" reprojection.

- Choose your newly defined projection (search for its name in the *Filter* field).

- On applying this projection, the map will be reprojected thus:

7.1.5 In conclusion

Different projections are useful for different purposes. By choosing the correct projection, you can ensure that the features on your map are being represented accurately.

7.1.6 Further reading

Materials for the *Advanced* section of this lesson were taken from this article[1].

Further information on Coordinate Reference Systems is available here[2].

7.1.7 What's next?

In the next lesson you'll learn how to analyze vector data using QGIS' various vector analysis tools.

7.2 Lesson: Vector Analysis

Vector data can also be analyzed to reveal how different features interact with each other in space. There are many different analysis-related functions in GIS, so we won't go through them all. Rather, we'll pose a question and try to solve it using the tools that QGIS provides.

The goal for this lesson: To ask a question and solve it using analysis tools.

7.2.1 The GIS Process

Before we start, it would be useful to give a brief overview of a process that can be used to solve any GIS problem. The way to go about it is:

1. State the Problem
2. Get the Data
3. Analyze the Problem
4. Present the Results

7.2.2 The problem

Let's start off the process by deciding on a problem to solve. Let's say you're a property developer, and you need to buy a farm to build a new residential development for people who have had enough of Cape Town and Johannesburg. You've done some market research, and you've come up with the following requirements that a farm needs to satisfy:

1. It needs to be in Swellendam.
2. You don't want to have to build a long access road (maximum distance 500 meters), and it has to be easily accessible from main routes.

---

[1]http://tinyurl.com/75b92np
[2]http://linfiniti.com/dla/worksheets/7_CRS.pdf

3. The people who make up the target market will probably have children, so it must be within a reasonable distance (by car) from a school.

4. The farm must be between 100 and 150 hectares in area.

7.2.3 The data

To answer these questions, we're going to need the following data:

1. The farms we're going to be choosing from.

2. The roads that need to be running through these farms.

3. The location of schools.

For this exercise the data has been provided already, but in reality you may need to find a provider for the datasets in question. In South Africa, the Surveyor-General's Office is a good source of the kind of data you'd find on 1: 50 000 topographic map sheets, which includes all of the data we need for this particular problem.

7.2.4 Follow along: Start a Project

So now that we know what we want to do, let's start doing it!

- Start a new QGIS project.
- Start adding new layers. They're all to be found in the `exercise_data/projected_data` folder.
- Add the `schools_33S`, `roads_33S` and `farms_33S` layers.

They'll probably look very familiar. They're the same data we've been working with all along, except that they're now in a Projected Coordinate System. This is necessary for analysis, for reasons we've covered in the previous lesson. (We want to be able to calculate distances in meters and areas in hectares, not in degrees.)

Once you've got these layers in your *Layers list*:

- Rename the layers by right-clicking on them and selecting the *Rename* option.
- Give them the new, less messy names `roads`, `schools` and `farms`.
- Save your map under *exercise_data*, as `analysis.qgs`.
- In your operating system's file manager, create a new folder under *exercise_data* and call it `residential_development`. This is where you'll save the datasets that will be the results of the analysis functions.

Now that we've got the data, let's analyze the problem!

## 7.2.5 ⟨tag⟩ Follow along: Analyzing the Problem: Farms in Swellendam

The first criterion we're facing is that the farm needs to be in Swellendam. So let's tell QGIS to only show us the farms that are, in fact, in Swellendam!

- Right-click on the *farms* layer in the *Layers list*.

- Select the option *Query....* This gives you a new *Query Builder* dialog.

- Scroll down in the *Fields* list on the left of this dialog until you see the field TOWN.

- Click on it once.

- Click the *All* button underneath the *Values* list:
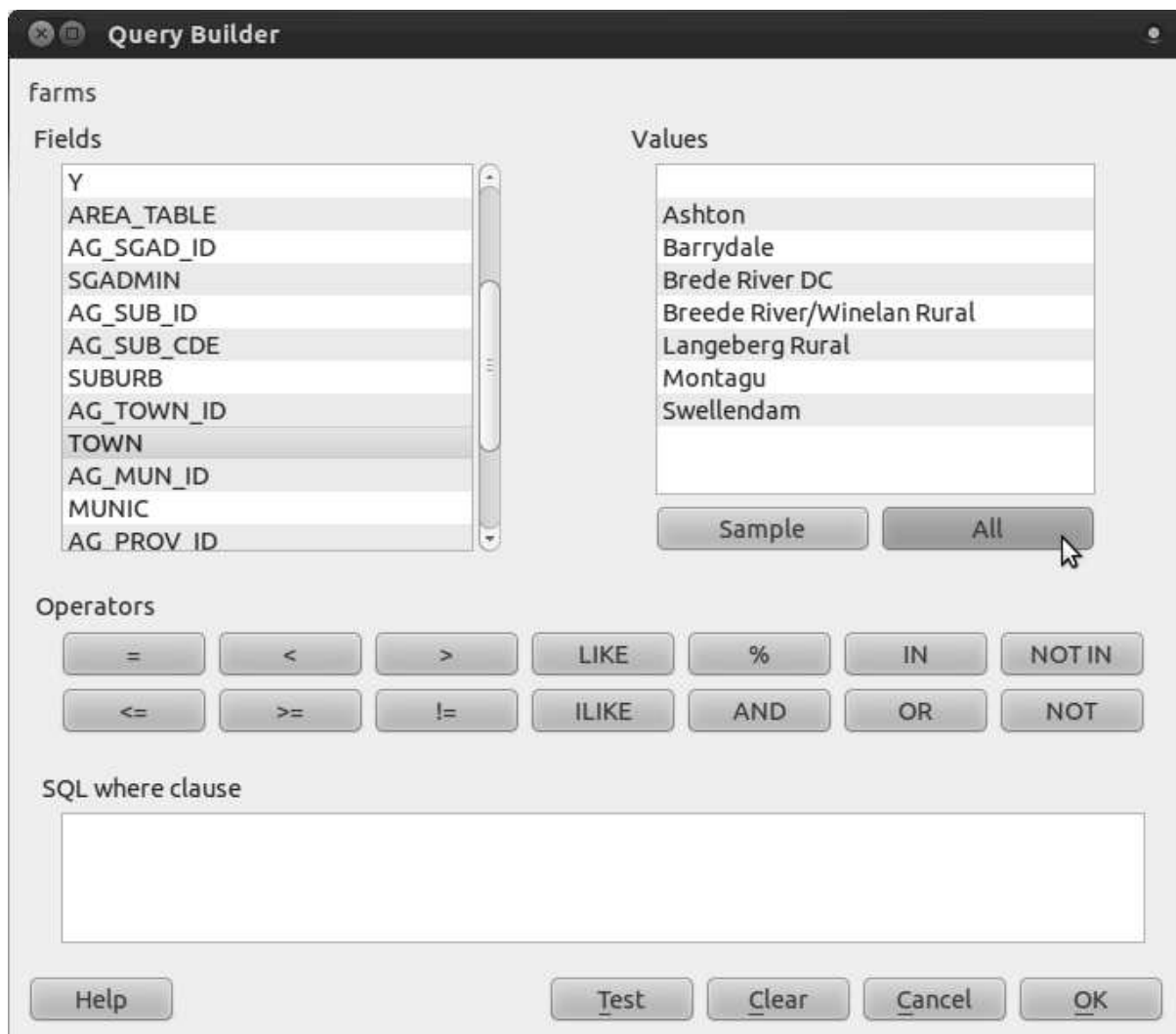
Now we need to tell QGIS to only show us the farms where the value of TOWN is equal to Swellendam.

- Double-click the word TOWN in the *Fields* list.

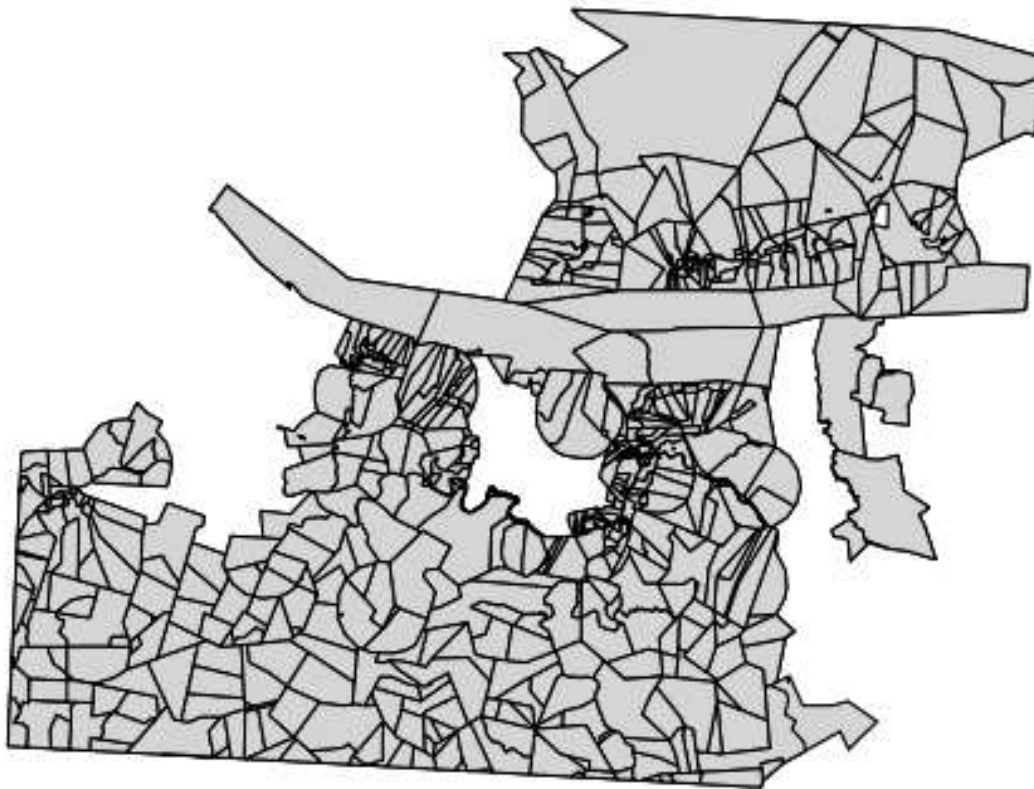- Watch what happens in the *SQL where clause* field below:
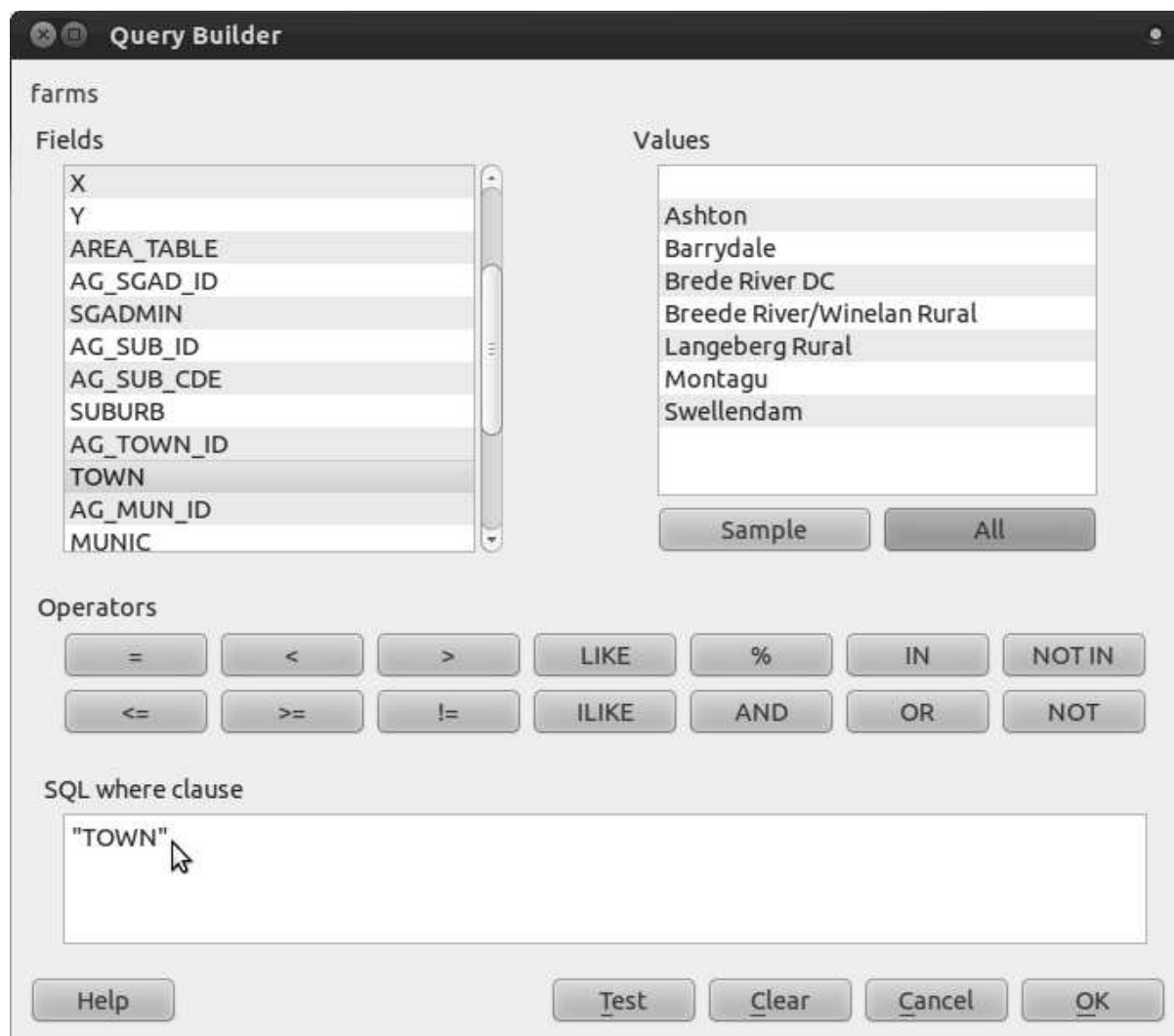
The word `"TOWN"` has appeared! To build the rest of the query:

- Click the = button (under *Operators*).
- Double-click the value `Swellendam` in the *Values* list.

Your query is this:

And now your *farms* layer looks like this:

For further analysis, it's better if we've got these farms as a separate layer.

- Right-click on the *farms* layer and click *Save As...*.

- Next to the *Save as* field in the dialog that appears, click the *Browse* button.

- Save the layer under `exercise_data/residential_development/`, as `swellendam_farms.shp`.

- Check the *Add saved file to map* box in the *Save vector layer as...* dialog.

- Click *OK*. It will tell you that `Export to vector file has been completed`.

- Click *OK*.

Now we've got the *farms* twice, but we only need the one we just created.

- Right-click on the old *farms* layer and remove it.

7.2.6 ⬤ Try yourself... find important roads

We have a similar problem with the roads; the only roads we're taking into account are the main ones.

- Build a query for the *roads* layer, like you did above for the *farms* layer. You want only the types `primary`, `secondary`, `tertiary` and `trunk`, so you need to build this

query:

```
"TYPE" = 'primary' OR "TYPE" = 'secondary' OR "TYPE" =
'tertiary' OR "TYPE" = 'trunk'
```

You can either use the approach above, where you double-clicked values and clicked buttons, or you can copy and paste the command above.

Check your results

7.2.7 🏷 Try yourself... find high schools

Your future customers will want to send their children to high schools.

- Use the same approach as before and build this query for the *schools* layer:

```
"has_high_s" = 'y'
```

You should have only 3 out of the 4 original schools left.

- Save the resulting layer under `exercise_data/residential_development/` as `high_schools.shp`.

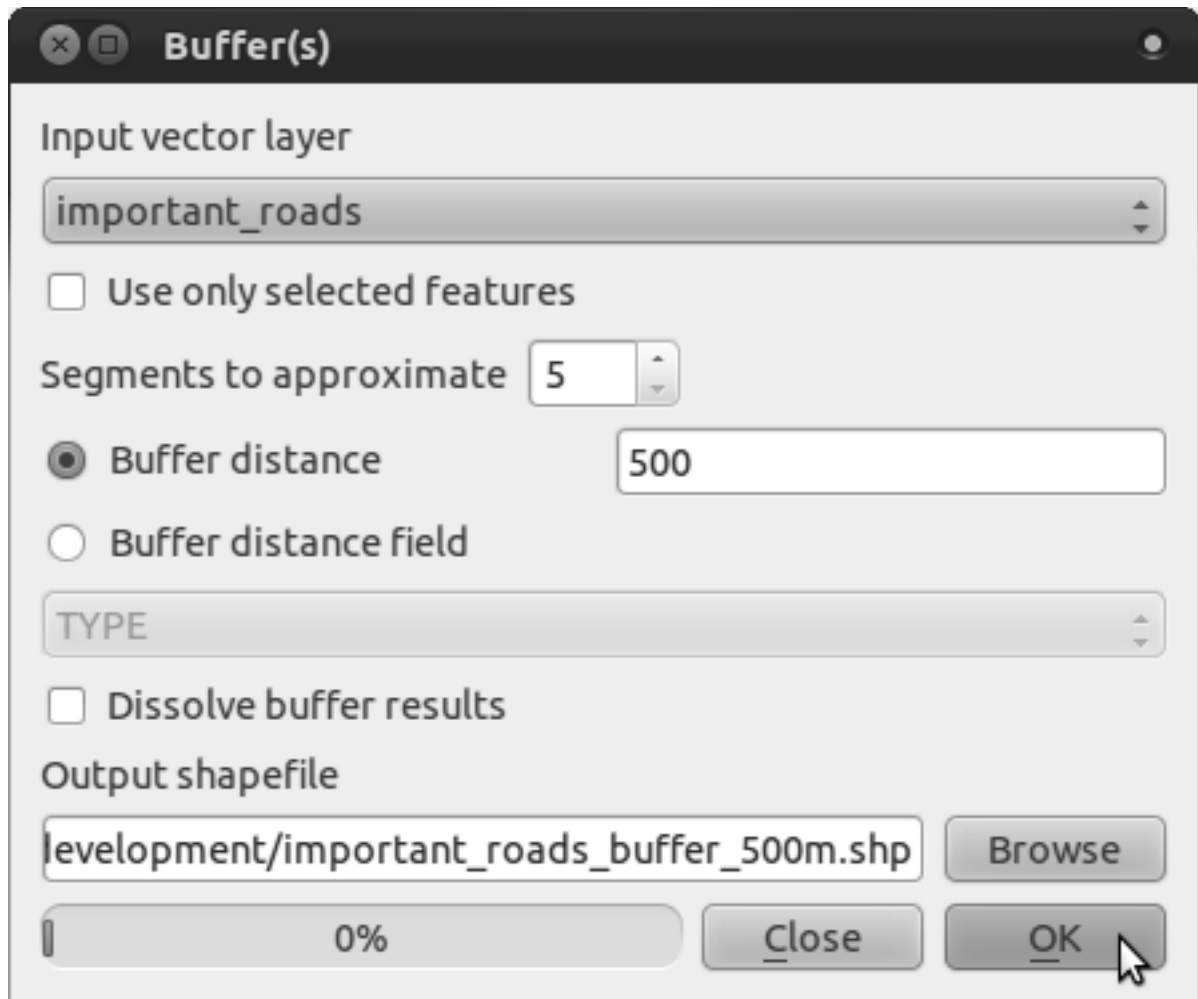## 7.2.8 🏷 Follow along: Analyzing the Problem: Distances From Schools and Roads

QGIS allows you to calculate distances from any vector object.

- Make sure that only the *important_roads* and *swellendam_farms* layers are visible, to simplify the map while you're working.

- Click on the *Vector → Geoprocessing Tools → Buffer(s)* tool:
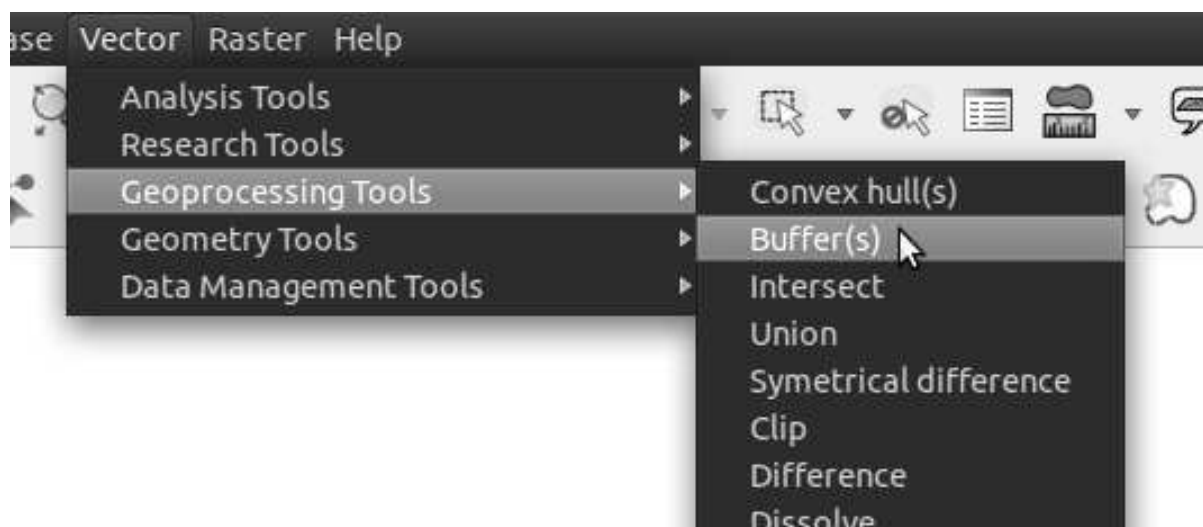
This gives you a new dialog.

- Set it up like this:

The *Buffer distance* is in meters because our input dataset is in a Projected Coordinate System. This is why we needed to use projected data!

- Save the resulting layer under `exercise_data/residential_development/` as `important_roads_buffer_500m.shp`.

- Click *OK* and it will create the buffer.

- When it asks you if it should "add the new layer to the TOC", click *Yes*. ("TOC" stands for "Table of Contents", by which it means the *Layers list*).

Now you've got this:

OK, that looks interesting! Those are all the areas that are within 500 meters of an important road.

But look closer: it looks like there are joins between different sections of road! That's not nice. Remove that layer and let's try again:
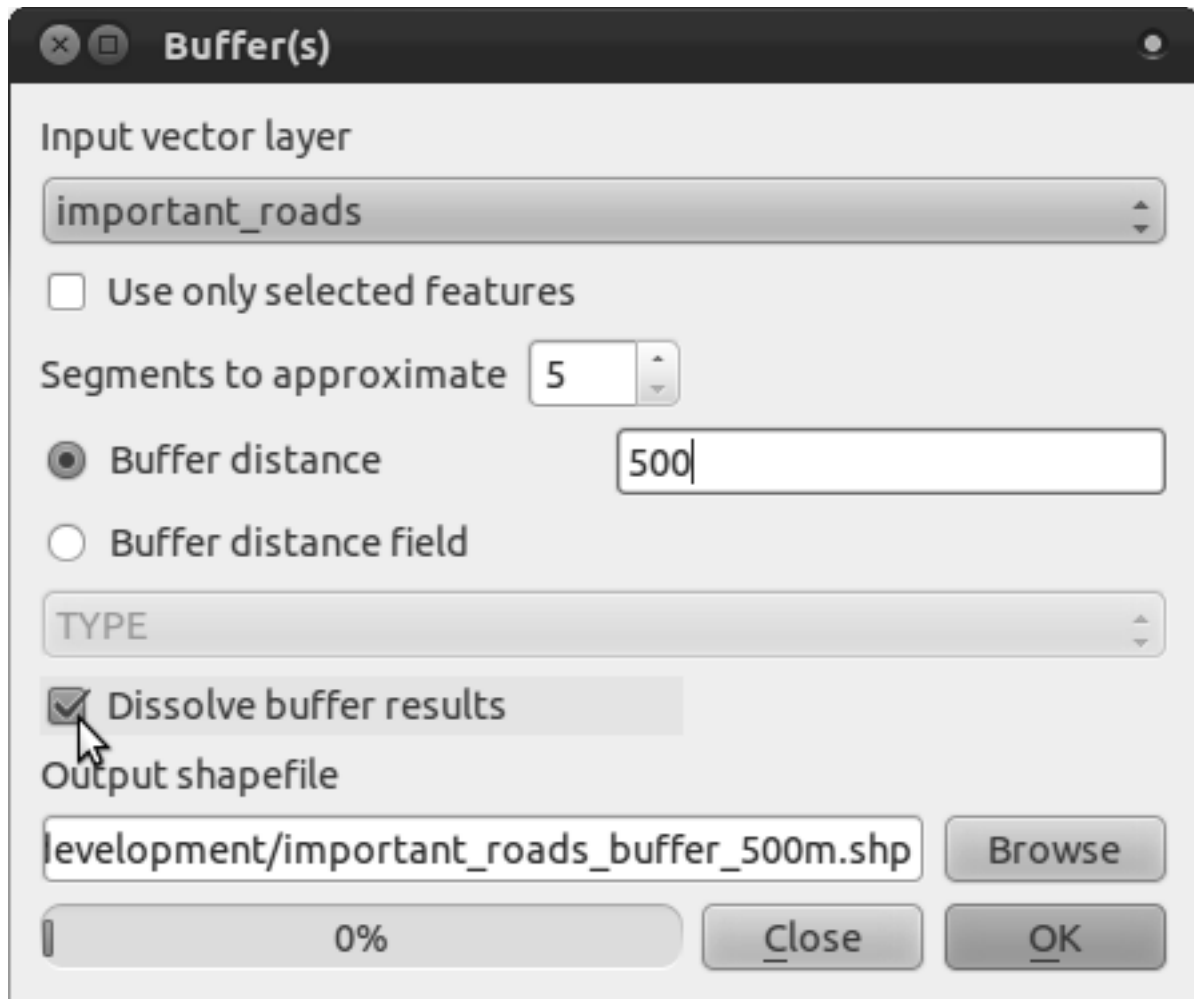
- Note that we're now checking the *Dissolve buffer results* box.

- Save the output under the same name as before (click *Yes* when it asks your permission to overwrite the old one).

- Click *OK*.

Once you've added the layer to the *Layers list*, it will look like this:

That's more like it! Now there are no unnecessary subdivisions.

### 7.2.9   Try yourself... distance from high schools

- Use the same approach as above and create a buffer for your schools.

It needs to be `10 km` in radius, and saved under the usual directory as `high_schools_buffer_10km.shp`.

Check your results

7.2.10   Follow along: Overlapping Areas

Now we have areas where the road in 500 meters away and there's a school within 10 km (as the crow flies). But obviously, we only want the areas where both of these criteria are satisfied at once! To do that, we'll need to use the *Intersect* tool. Find it under *Vector → Geoprocessing Tools → Intersect*. Set it up like this:
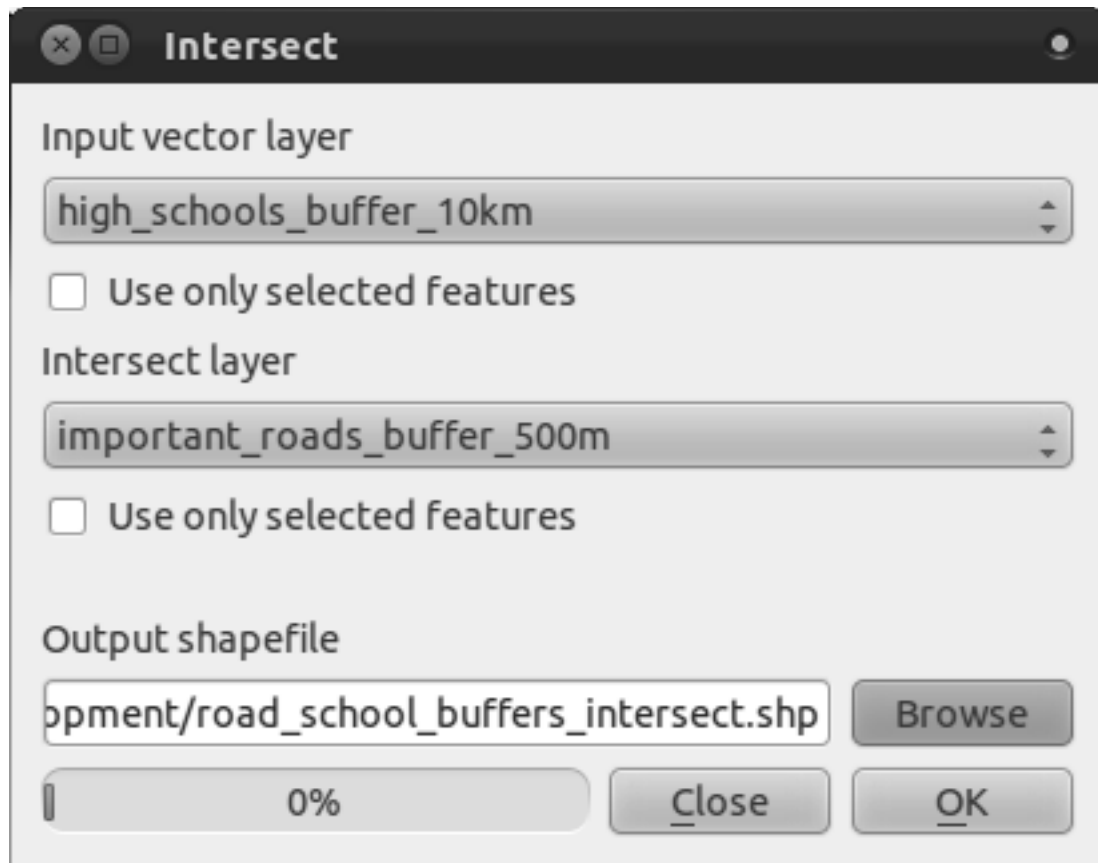
The two input layers are the two buffers; the save location is as usual; and the file name is `road_school_buffers_intersect.shp`. Once it's set up like this, click *OK* and add the layer to the *Layers list* when prompted.

In the image below, the blue areas show us where both distance criteria are satisfied at once!

You may remove the two buffer layers and only keep the one that shows where they overlap, since that's what we really wanted to know in the first place.
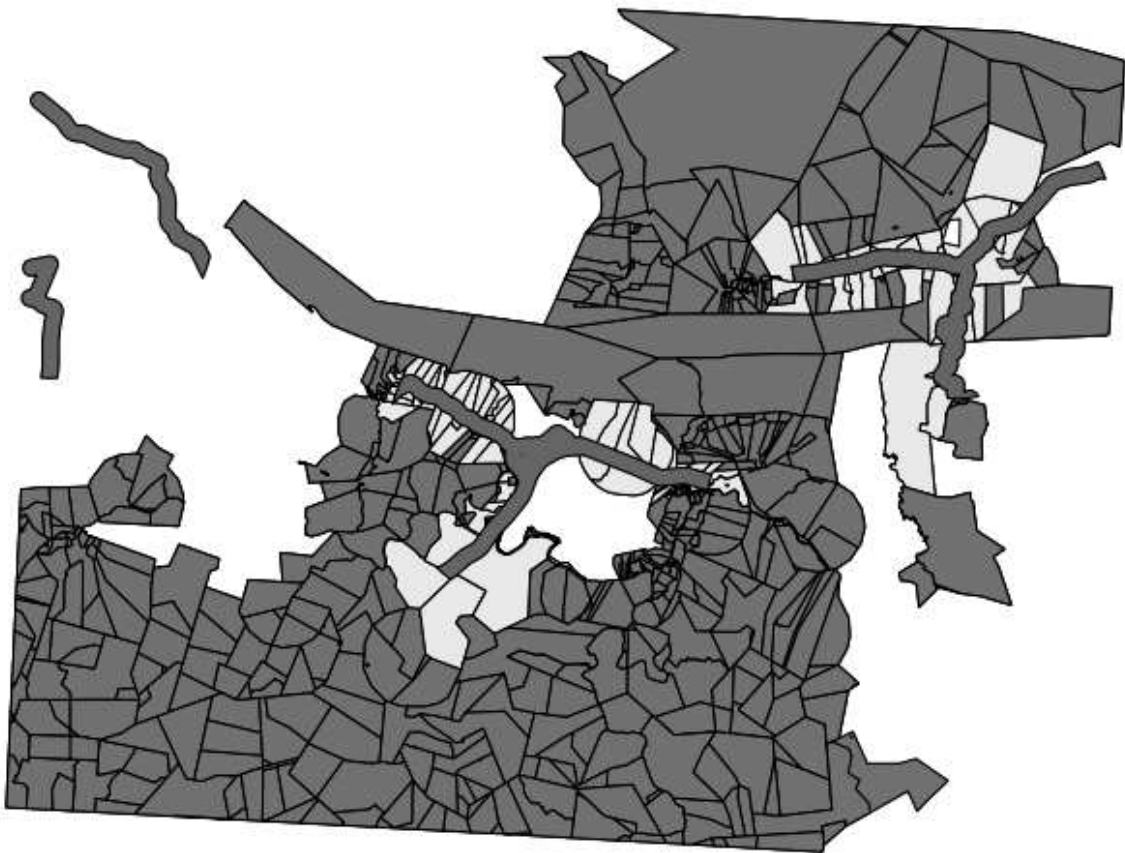
7.2.11 ✎ Follow along: Select the Farms

Now you've got the area that the farms must overlap. Next, you want to select the farms in that area.

- Click on the menu entry *Vector → Research Tools → Select by location*. A dialog will appear.
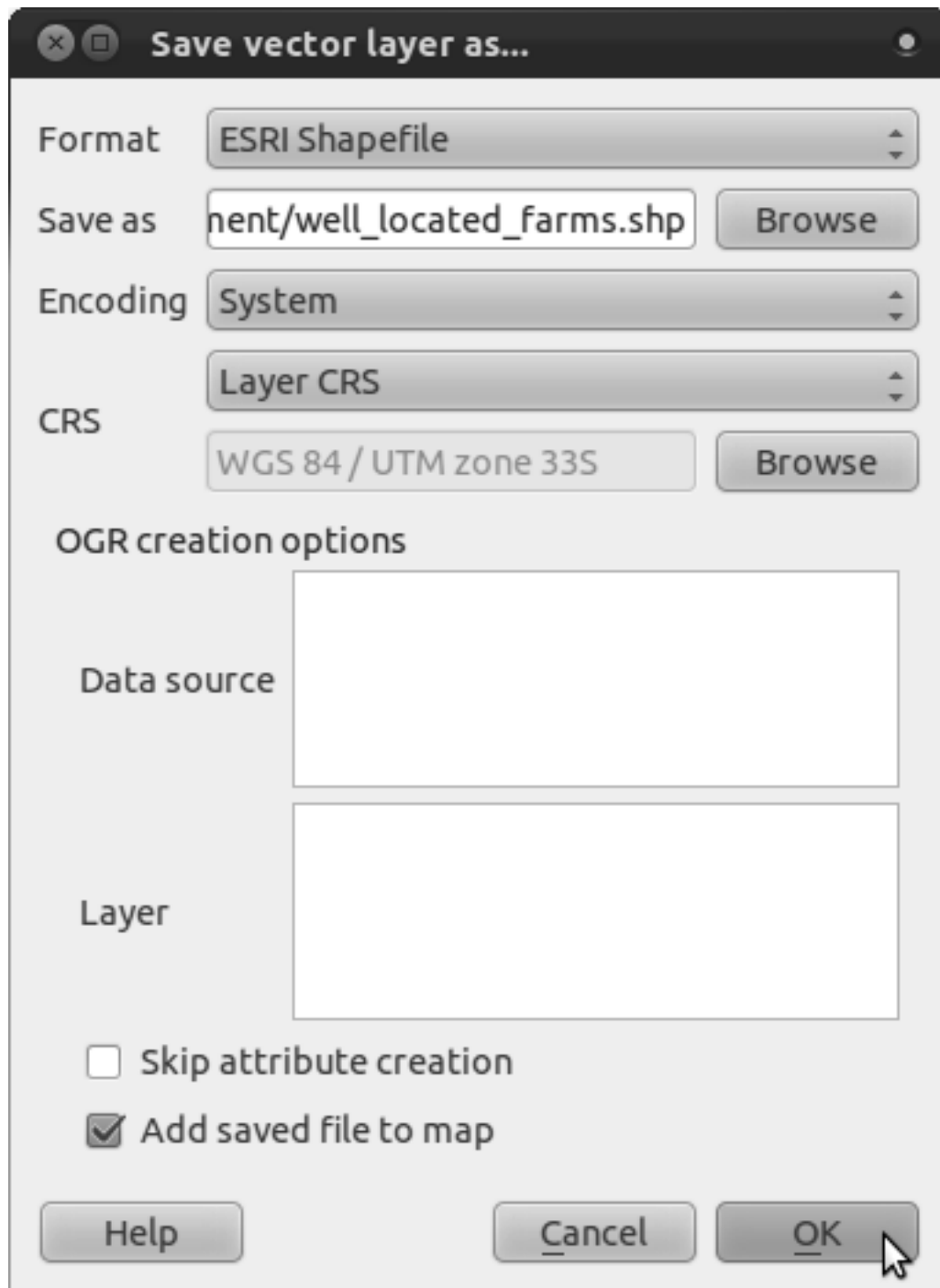- Set it up like this:
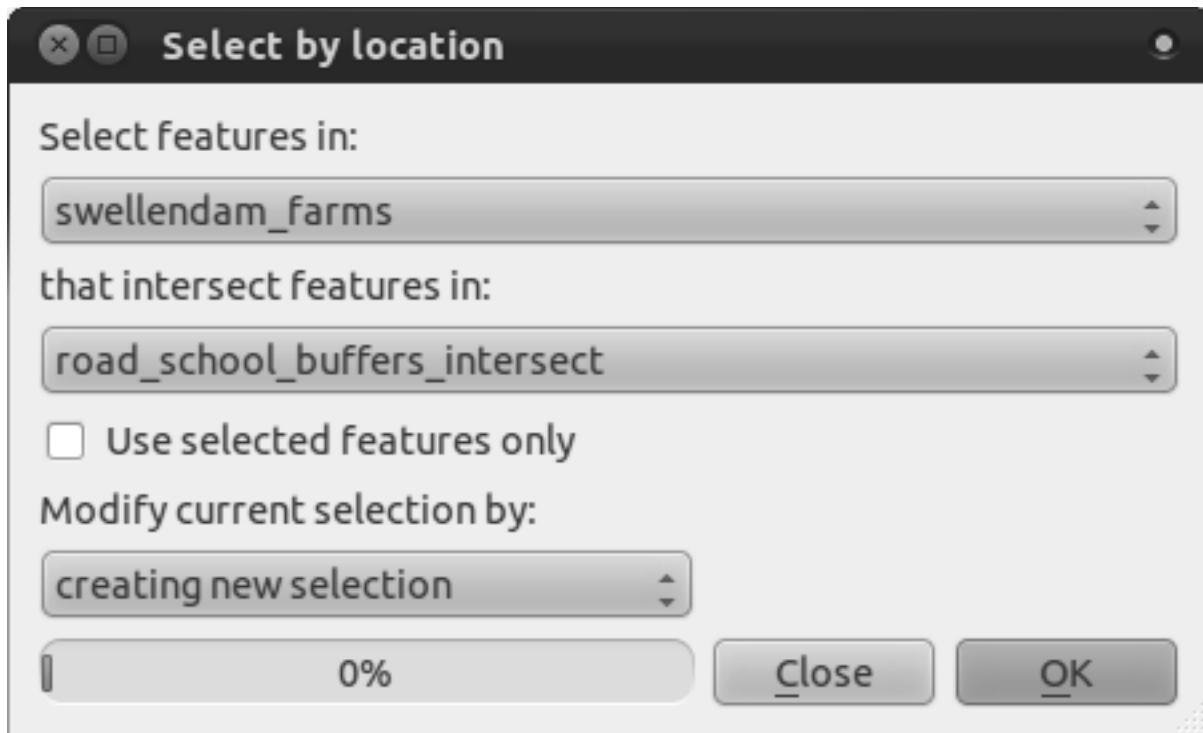
- Click *OK* and you'll see the results:

The yellow farms are the selected ones. Let's save them as a new layer.

- Right-click on the *swellendam_farms* layer in the *Layers list*.

- Select *Save Selection As....*
- Set the dialog up like this:

- The file name is `well_located_farms.shp`.
- Click *OK*.

Now you have the selection as a separate layer.

7.2.12 🏷 Follow along: Select Farms of the Right Size

To see which farms are the right area (between 100 and 150 ha), we first need to calculate their areas in hectares.

- Open the attribute table for the *well_located_farms* layer.

- Enter edit mode:

- Start the field calculator:

- Set it up like this:
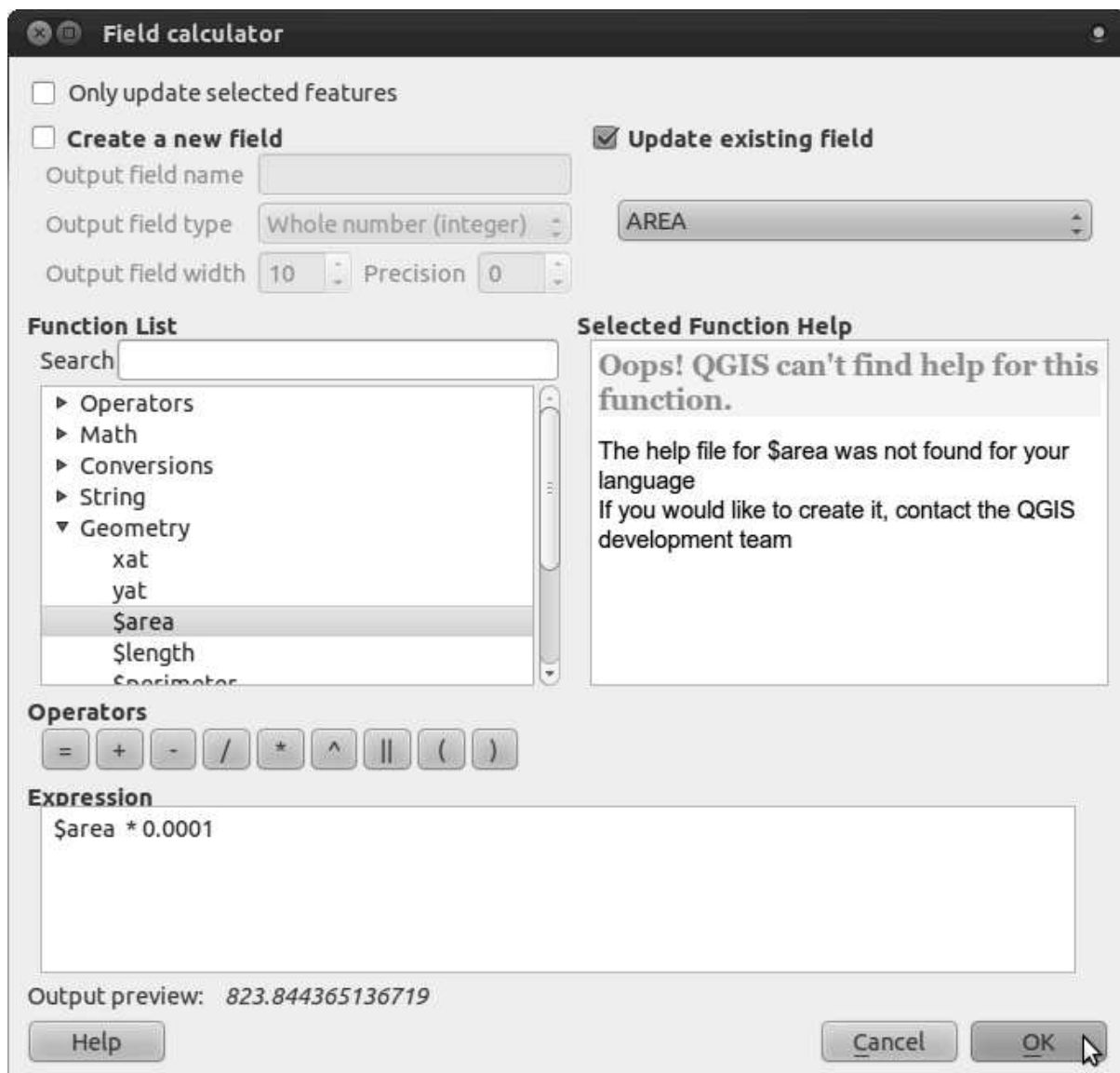
- If you can't find *AREA* in the list, try this instead:

- Click *OK*.

- Scroll to the right of the attribute table; your AREA field now has areas in hectares for all the farms in your *well_located_farms* layer.

- Click the edit mode button again as above, and save your edits when prompted.

- Do a query:

- Click *OK*. Your results should look like this:

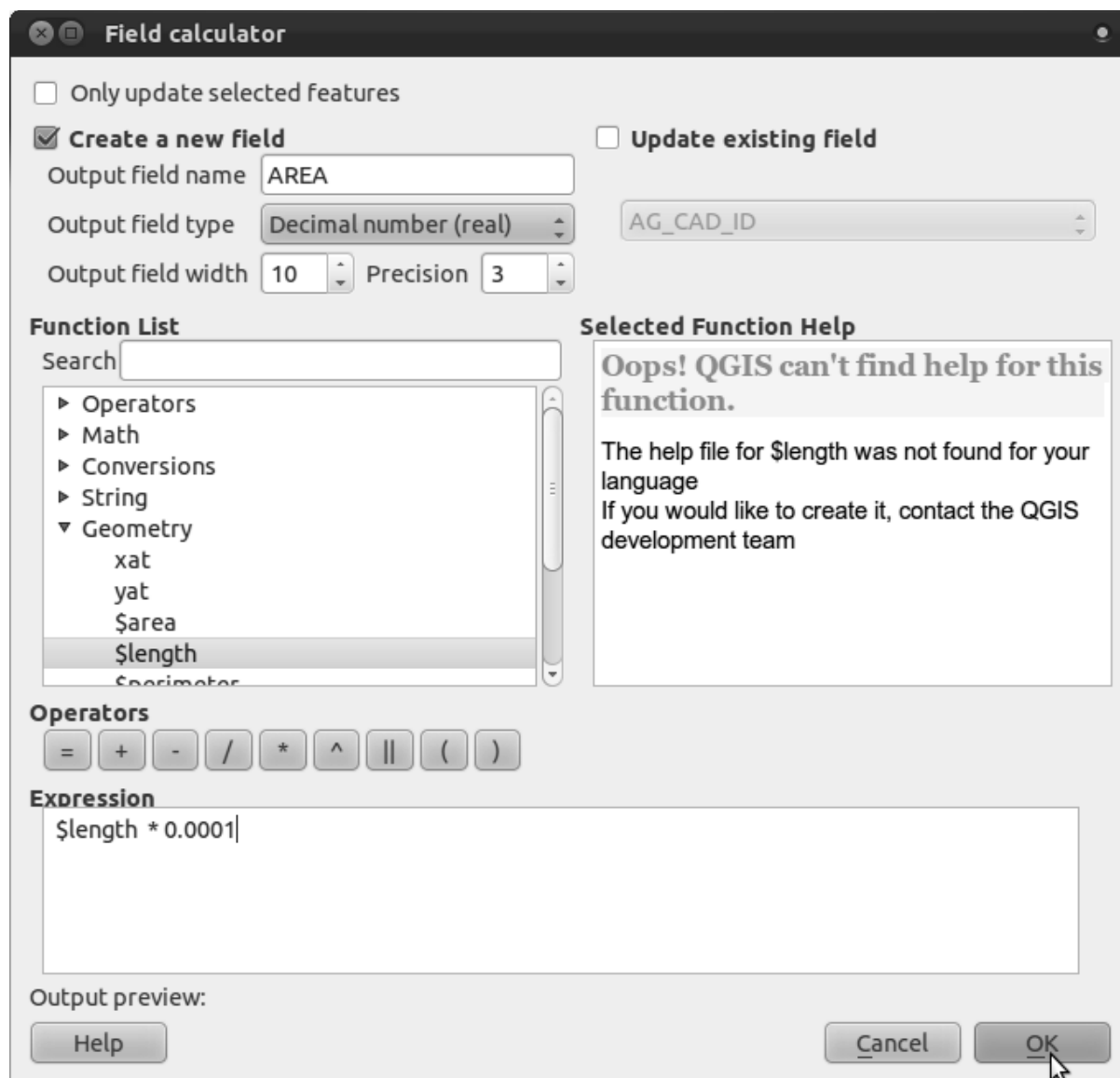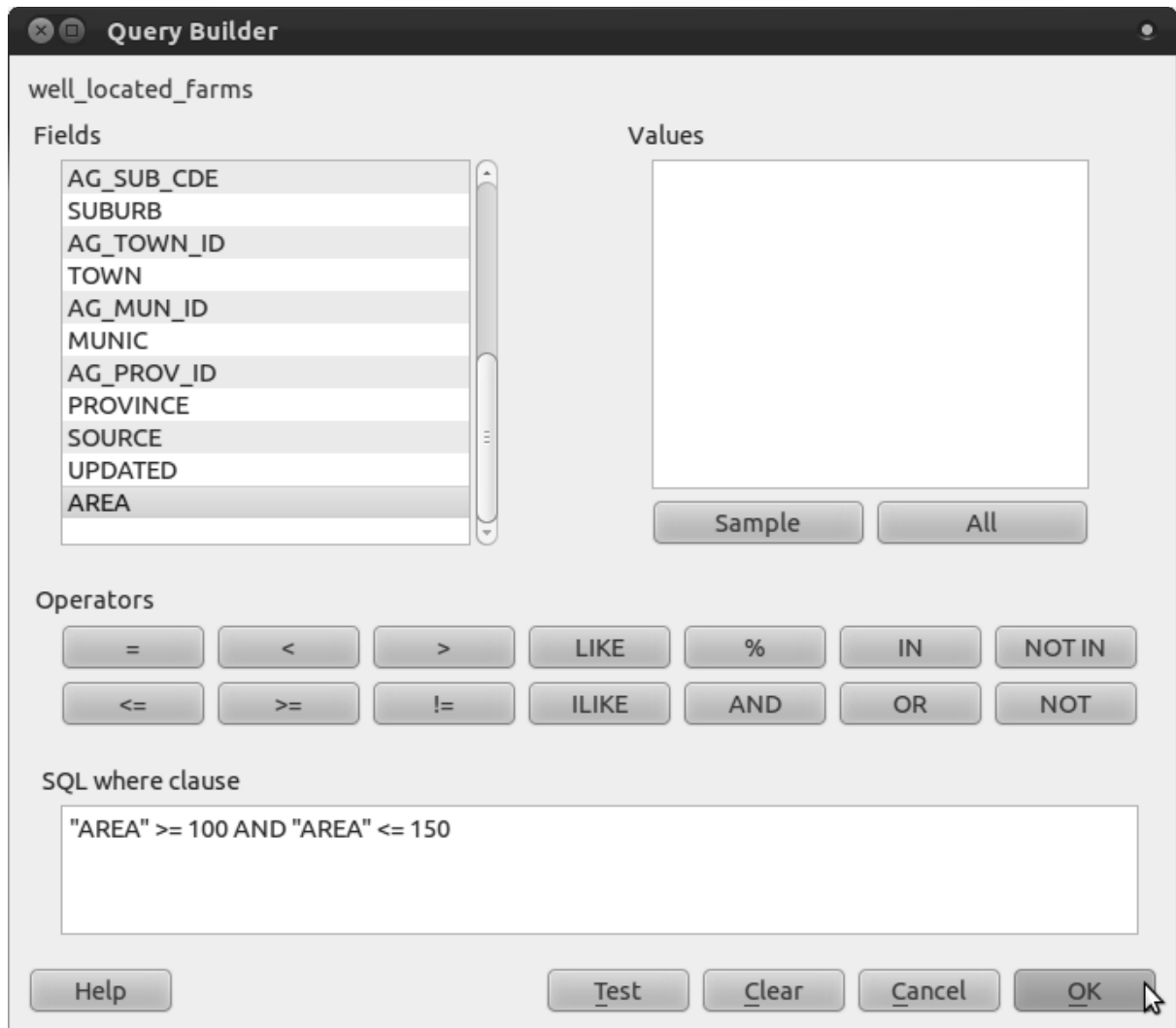That's it! Those red farms are the ones satisfying your criteria. Well done!

7.2.13 🏷 Try yourself...

- Save your solution as a new layer, using the approach you learned above for doing so. The file should be saved under the usual directory, with the name `solution.shp`.

7.2.14 In conclusion

Using the GIS problem-solving approach together with QGIS vector analysis tools, you were able to solve a problem with multiple criteria quickly and easily.

7.2.15 What's next?

In the next lesson, we'll look at how to calculate the shortest distance along the road from one point to another.

7.3 Lesson: Network Analysis

Calculating the shortest distance between two points is a commonly cited use for GIS. QGIS ships with this tool, but it's not visible by default. In this brief lesson, we'll show you what you need to get started.

The goal for this lesson: To activate, configure and use the *Road Graph* plugin.

7.3.1 ⬤ Follow along: Activate the Tool

QGIS has many plugins that add to its basic functions. Many of these plugins are so useful that they ship along with the program straight out of the box. They're still hidden by default, though. So in order to use them, you need to activate them first.

To activate the *Road Graph* plugin:

- Start the *Plugin Manager* by clicking on the QGIS main window's menu item *Plugins →
  Manage Plugins*. A dialog appears.

- Select the plugin like this:
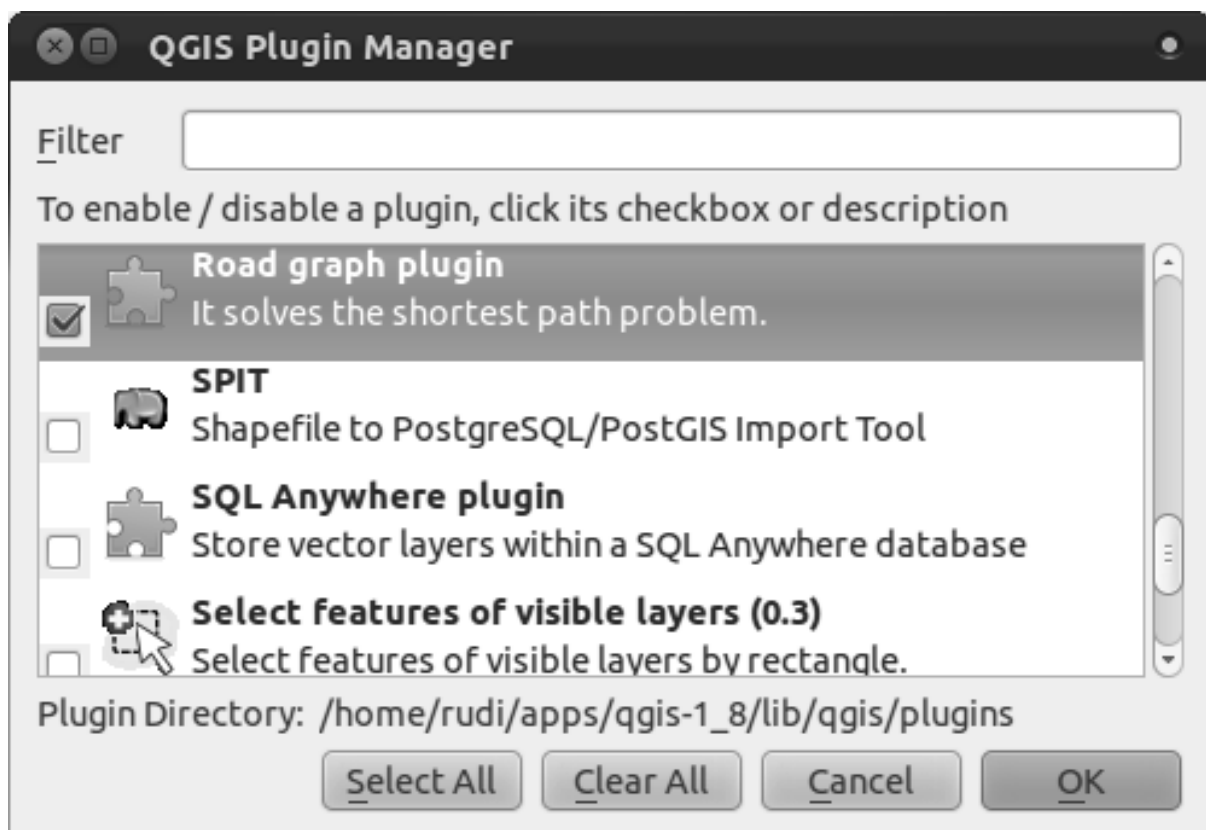
- Click *OK* on the *Plugin Manager* dialog.

- To see the plugin in your interface, go to *View → Panels* and ensure that *Shortest path*
  has a check mark next to it.

This panel will appear in your interface:

7.3.2 🏷 Follow along: Configure the Tool

To have a layer to calculate on, first save your current map. Then create a new map and load the layer *exercise_data/projected_data/roads_33S.shp*.

Since so many different configurations are possible when analyzing networks, the plugin doesn't assume anything before you've set it up. Unfortunately, this means that it won't do anything at all if you don't set it up first.

- Click on the menu item *Plugins → Road graph → Road graph settings*. A dialog will appear.
- Make sure it's set up like this (use defaults unless otherwise specified):
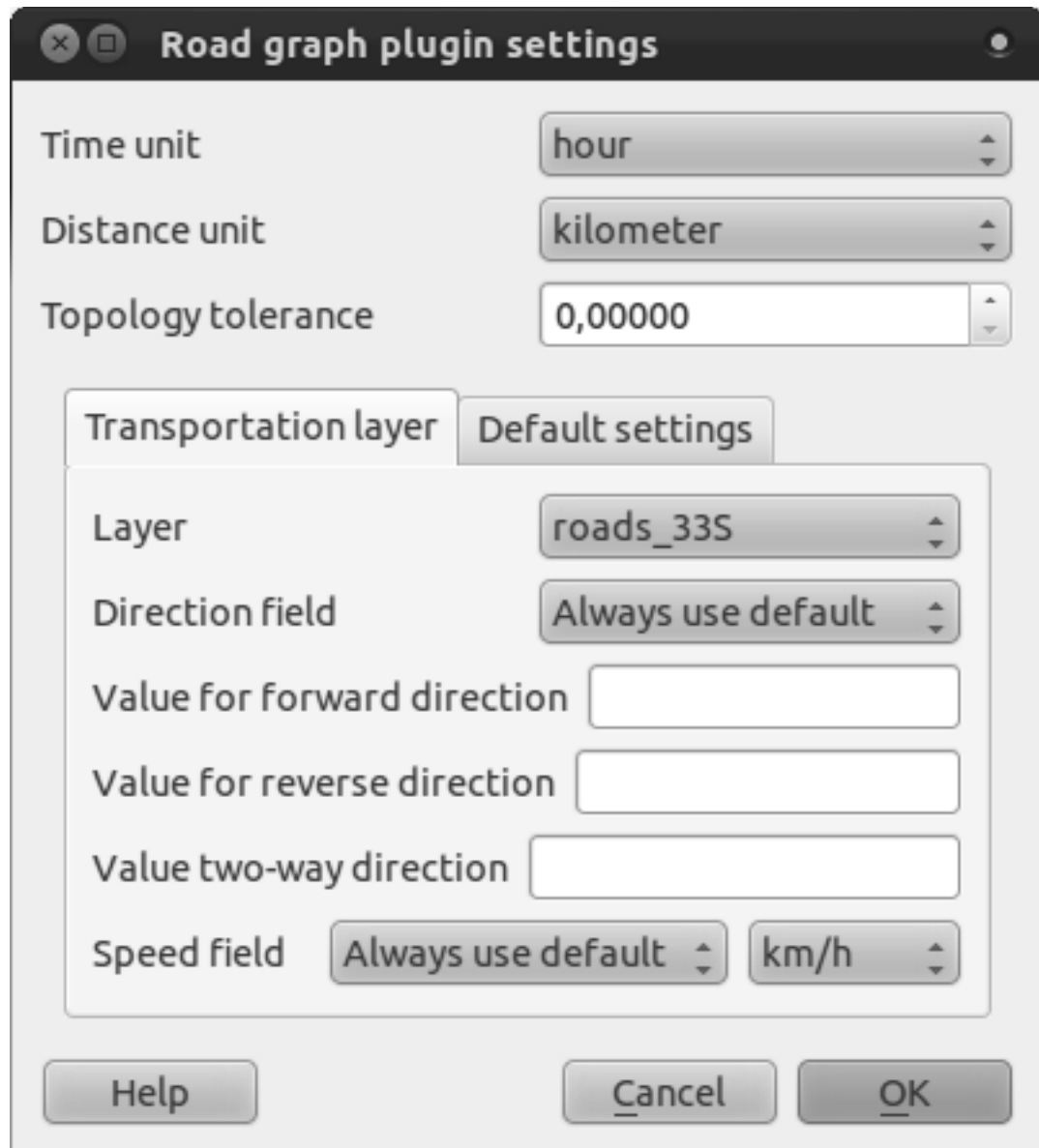
- *Time unit*: *hour*

- *Distance unit*: *kilometer*

- *Layer*: *roads_33S*

- *Speed field*: *Always use default / km/h*

- *Direction*: *Two-way direction*
- *Speed*: *60*

7.3.3   Follow along: Use the Tool

Let's say you want to get from Barrydale to Bonnievale:
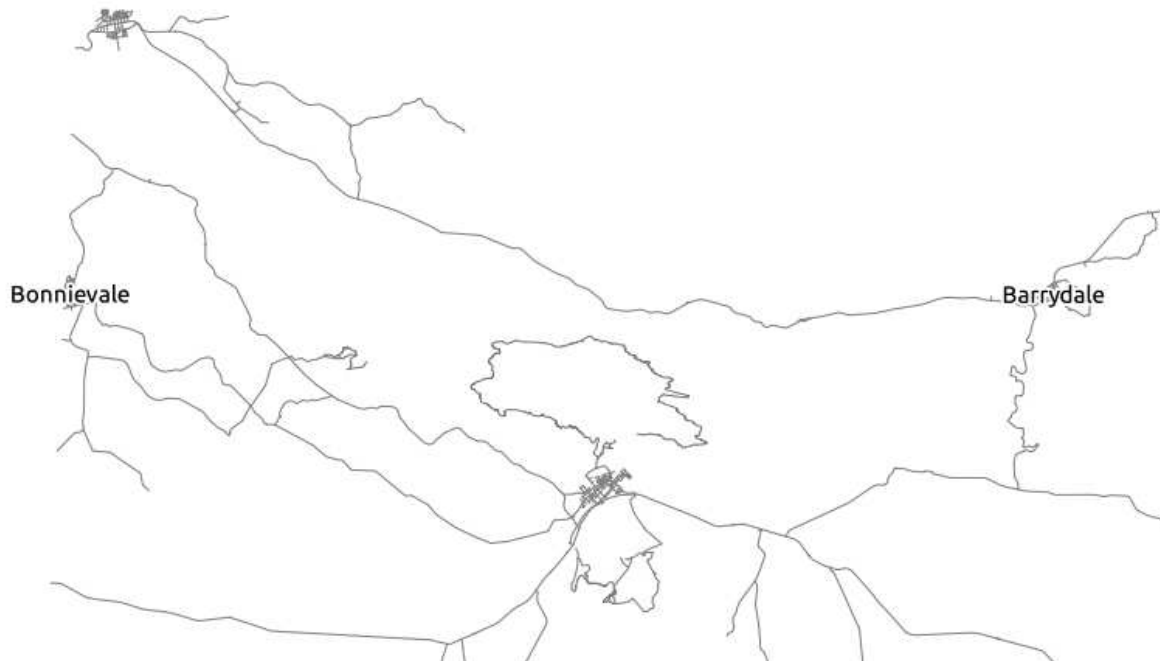
- In the plugin panel, click on the *Capture Point* button next to the *Start* field:

- Click somewhere in Barrydale.

- Use the *Capture Point* button next to the *Stop* field and capture a point somewhere in Bonnievale.

- Click on the *Calculate* button to see the solution:

7.3.4 Follow along: Using Criteria

Note: Section developed by Linfiniti and S Motala (Cape Peninsula University of Technology)

- Add the layer *exercise_data/projected_data/places_33S.shp*.

- Add the aerial photographs under *exercise_data/raster/*.

- Zoom to the Swellendam area (the town / cluster of roads in the center of the map).

- Select only the roads that are in the categories `trunk`, `primary` or `tertiary` by running this attribute selection query on the road layer:

```
"TYPE" = 'trunk' OR "TYPE" = 'primary' OR "TYPE" = 'tertiary'
```

- Right-click on the *roads_33S* layer in the *Layers list*.

- Save it out to a new file, `roads_subset`. Only the currently visible features will be saved.

- Remove the original *roads_33S* layer from the *Layers list*.

- Open the attribute table for the *roads_subset* layer and enter edit mode:

- Add a new column:

- Call this new column `SPEED`, and give it the type *Whole number (integer)* with a width of 3.

- In the main window, activate the *Select Features by Rectangle* tool:

- Select these roads:

(To select more than one road, hold the `ctrl` button and drag a box across any road that you want to include in the selection.)

  • In the attribute table, click on the *Show selected only* box.

  • Set the `SPEED` value for all the selected streets to `60`:

In context, this means that you're setting the speed limit on those roads to `60 km/h`.

  • Select the highway outside of town:

- Set the SPEED value for all the selected streets to 120.

- Close the attribute table, save your edits, and exit edit mode.

- Check the *Vector → Road graph → Road graph settings* to ensure that it's set up as explained previously in this lesson, but with the *Speed* value set to the SPEED field you just created.

- In the *Shortest path* panel, click the *Start point* button.

- Set the starting point to where the dirt road meets the on/off-rqamp at the edge of town on the east.

- Set the end point to the T-junction west of town.

- In the *Criterion* drop-down list in the *Shortest path* panel, select *Length*.

- Click *Calculate*. The route will be calculated for the shortest distance:

Notice the values of *Length* and *Time* in the *Shortest path* panel.

- Set the *Criterion* to *Time*.

- Click *Calculate* again. The route will be calculated for the shortest time:

You can switch back and forth between these criteria, recalculating each time, and note the changes in the *Length* and *Time* taken. Remember that the assumption being made to arrive at the time taken to travel a route does not account for acceleration, and assumes that you will be traveling at the speed limit at all times. In a real situation, you may want to split roads into smaller sections and note the average or expected speed in each section, rather than the speed limit.

7.3.5 Try yourself...

- Digitize the roads in Railton.

These are the main roads, type *tertiary*, speed `60`:

And these are the smaller streets, type *residential*, speed 40:

- Use the *Shortest path* tool to calculate the shortest and fastest ways to get from the southern extreme of town to the highway.

If, on clicking *Calculate*, you're getting an error stating that a path could not be found, then make sure that the roads you digitized actually meet each other. If they're not quite touching, either fix them by modifying the features, or set the *Topology tolerance* (*Road graph plugin settings*). If they're passing over each other without intersecting, use the *Split features* tool to "split" roads at their intersections:

Remember that the *Split features* tool only works in edit mode on selected features, though!

7.3.6 In conclusion

Now you know how to use the *Road Graph* plugin to solve shortest-path problems.

7.3.7 What's next?

Next you'll see how to run spatial statistics algorithms on vector datasets.

7.4 Lesson: Spatial Statistics

Note: Lesson developed by Linfiniti and S Motala (Cape Peninsula University of Technology)

Spatial statistics allow you to analyze and understand what is going on in a given vector dataset. QGIS includes several standard tools for statistical analysis which prove useful in this regard.

The goal for this lesson: To know how to use QGIS' spatial statistics tools.

7.4.1 Follow along: Create a Test Dataset

In order to get a point dataset to work with, we'll need a point dataset. In order to mimic a real dataset, let's create a random set of points.

To do so, you'll need a polygon dataset defining the extents of the area you want to create the points in.

We'll use the area covered by streets.

- Create a new empty map.

- Add the `roads_33S` layer from the `exercise_data/projected_data` directory, as well as the `srtm_41_19.tif` raster found in `exercise_data/raster/SRTM/`.

- Use the *Convex hull(s)* tool (available under *Vector → Geoprocessing Tools*) to generate an area enclosing all the roads:
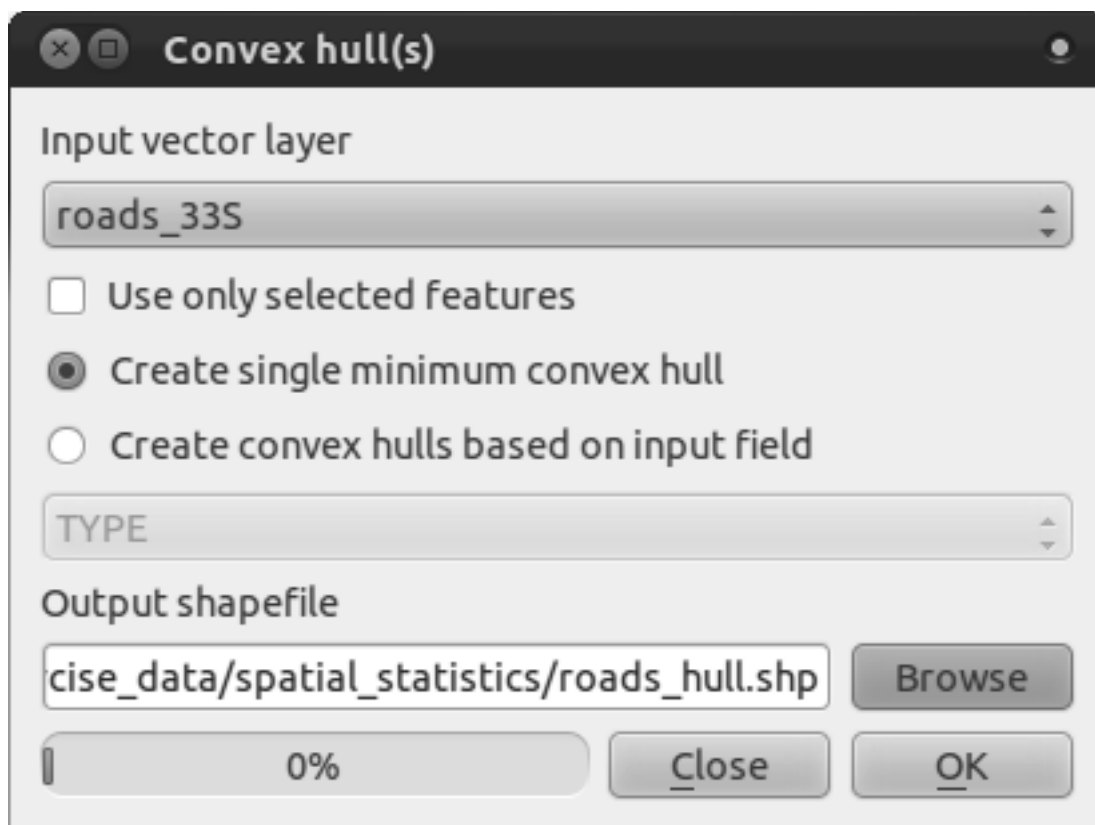
- Save the output under `exercise_data/spatial_statistics/` as `roads_hull.shp`.

- Add it to the TOC (*Layers list*) when prompted.

Creating random points

- Create random points in this area using the tool at *Vector → Research Tools → Random points*:
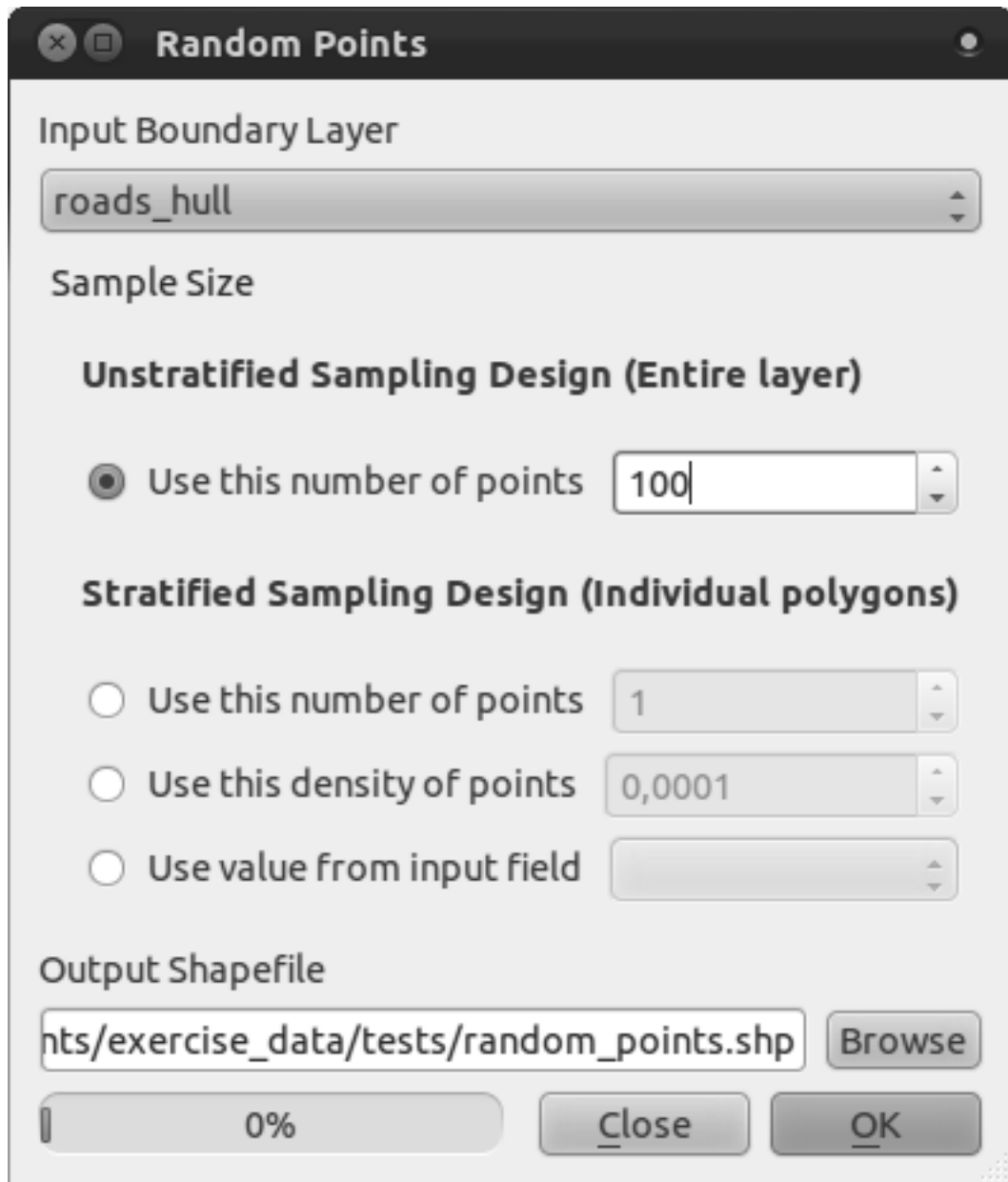
- Save the output under `exercise_data/spatial_statistics/` as `random_points.shp`.
- Add it to the TOC (*Layers list*) when prompted:

Sampling the data

- To create a sample dataset from the raster, you'll need to use the *Point sampling tool* plugin.

- Refer ahead to the module on plugins if necessary.

- Search for the phrase `point sampling` in the *Plugin Installer* and you will find the plugin.

- Note that it is in a third-party repository at time of writing, so you'll need to have third-party repositories enabled.

- As soon as it has been activated with the *Plugin Manager*, you will find the tool under *Plugins → Analyses → Point sampling tool*:

General   Fields   About

Layer containing sampling points:

random_points ▾

Layers with fields/bands to get values from:

random_points : ID (source point)
roads_hull : TYPE (polygon)
roads_hull : NAME (polygon)
roads_hull : ONEWAY (polygon)
roads_hull : LANES (polygon)
srtm_41_19.tif : Band 1 (raster)

Output point vector layer:

[   ] Browse

☒ Add created layer to the TOC

Status:
Complete the input fields and press OK...   OK   Close

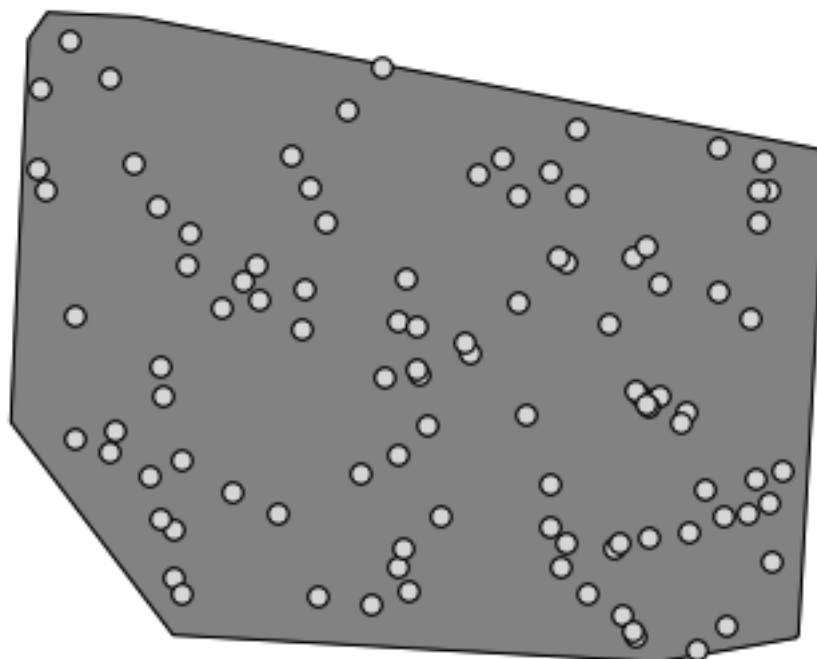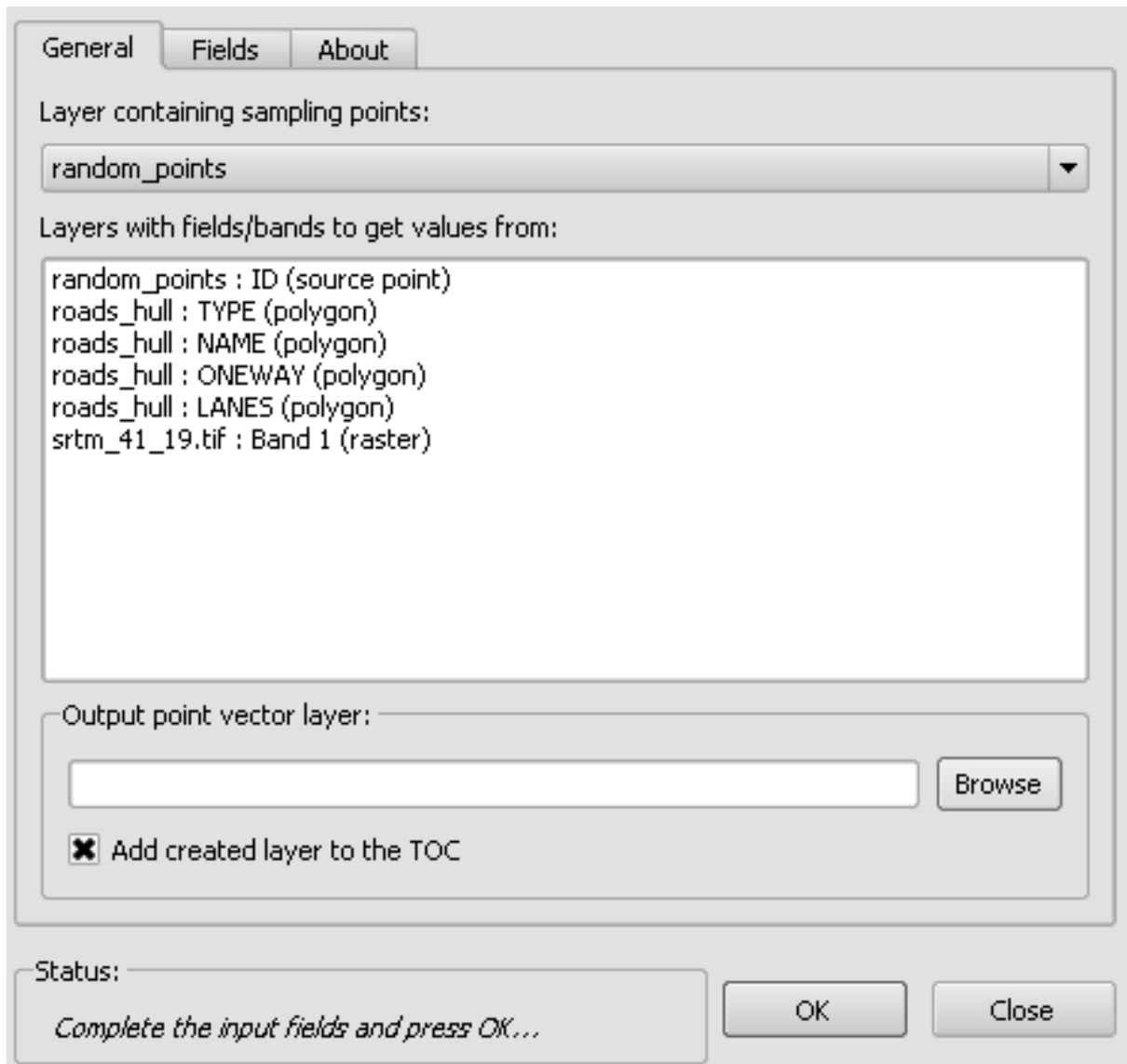- Select *random_points* as the layer containing sampling points, and the SRTM raster as the band to get values from.
- Save the output under `exercise_data/spatial_statistics/` as `random_samples.shp`.
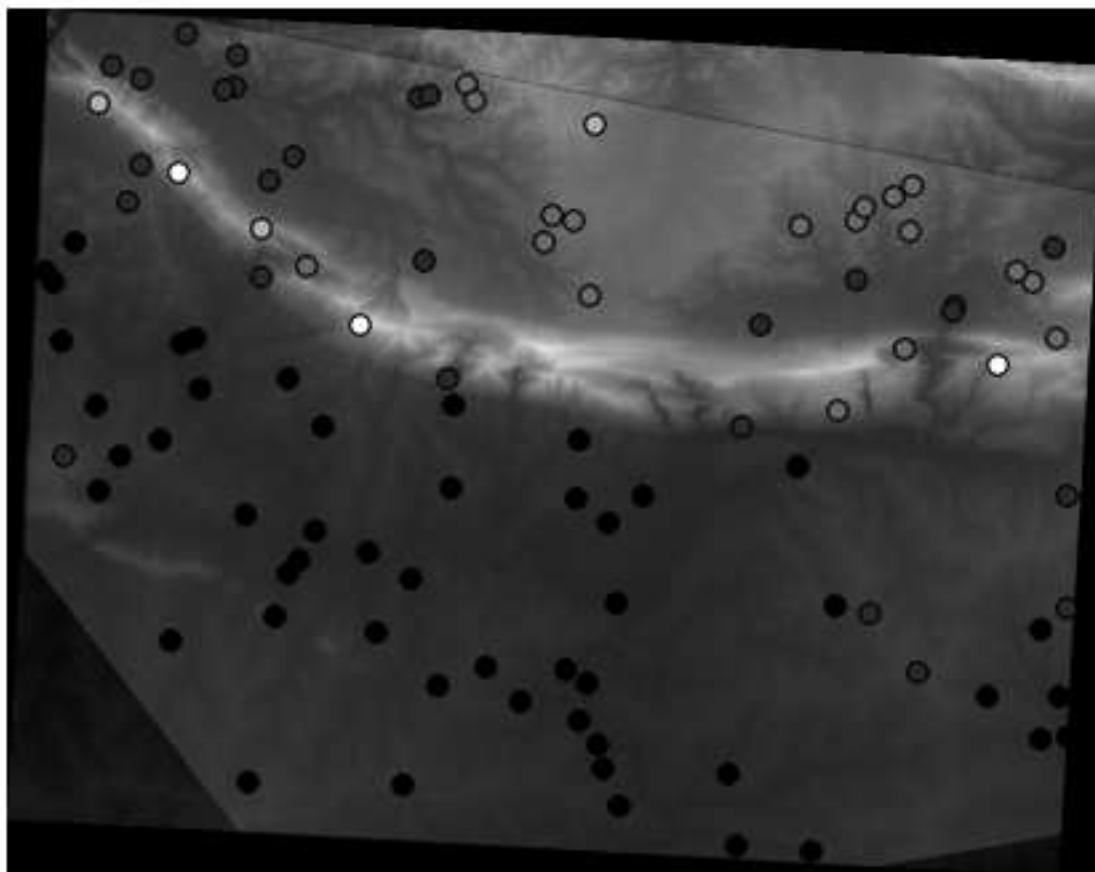- Add it to the TOC (*Layers list*) when prompted.

A possible sample layer is shown here:

The sample points are classified by their value such that darker points are at a lower altitude.

You'll be using this sample layer for the rest of the statistical exercises.

7.4.2 🏷 Follow along: Basic Statistics

Now get the basic statistics for this layer.

- Click on the *Vector → Analysis Tools → Basic statistics* menu entry.
- In the dialog that appears, specify the *random_samples* layer as the source.
- Click *OK*. You'll get results like this:

Parameter	Value
Mean	343.9
StdDev	254.482474839
Sum	34390.0
Min	34.0
Max	1226.0
N	100.0
CV	0.739989749458
Number of unique values	91
Range	1192.0
Median	269.0

Input Vector Layer

random_samples

☐ Use only selected features

Target field

srtm_41_19

Statistics output

Press Ctrl+C to copy results to the clipboard

OK

0%   Close

Note: You can copy and paste the results into a spreadsheet. The data uses a (colon :) separator.

	A	B
1	**Mean**	343.9
2	**StdDev**	254.4824748
3	**Sum**	34390
4	**Min**	34
5	**Max**	1226
6	**N**	100
7	**CV**	0.739989749
8	**Number of unique valu▸**	91
9	**Range**	1192
10	**Median**	269

- Close the plugin dialog when done.

To understand the statistics above, refer to this definition list:

Mean The mean (average) value is simply the sum of the values divided by the amount of values.

StdDev The standard deviation. Gives an indication of how closely the values are clustered around the mean. The smaller the standard deviation, the closer values tend to be to the mean.

Sum All the values added together.

Min The minimum value.

Max The maximum value.

N The amount of samples/values.

CV The spatial[3] covariance[4] of the dataset.

Number of unique values The number of values that are unique across this dataset. If there are 90 unique values in a dataset with N=100, then the 10 remaining values are the same as one or more of each other.

Range The difference between the minimum and maximum values.

Median If you arrange all the values from least to greatest, the middle value (or the average of the two middle values, if N is an even number) is the median of the values.

7.4.3 Follow along: Compute a Distance Matrix

- Create a new point layer in the same projection as the other datasets (`WGS 84 / UTM 33S`).

- Enter edit mode and digitize three point somewhere among the other points.

---

[3] http://en.wikipedia.org/wiki/Spatial_covariance
[4] http://en.wikipedia.org/wiki/Covariance

- Alternatively, use the same random point generation method as before, but specify only three points.

- Save your new layer as `distance_points.shp`.

In order to compare layers using statistical tools, a layer needs to have a unique ID field. Let's create one for your *random_samples* layer.

- Open the attribute table for this layer.

- Enter edit mode.

- Open the raster calculator.

- Create a new field called `id`, with an integer field type and a width of `3`.

- Look in the *Function list* under the category *Record*, and double-click on the item *$id*, so that it appears in the *Expression* field below.

- Click *OK* to generate a unique ID for this layer.

To generate a distance matrix using these points:

- Open the tool *Vector → Analysis Tools → Distance matrix*.

- Select the *distance_points* layer as the input layer, and the *random_samples* layer as the target layer.

- Set it up like this:

Input point layer

distance_points ▼

Input unique ID field

ID ▼

Target point layer

random_samples ▼

Target unique ID field

id ▼

Output matrix type

○ Linear (N*k x 3) distance matrix

○ Standard (N x T) distance matrix

● Summary distance matrix (mean, std. dev., min, max)

☒ Use only the nearest (k) target points   10 ▲▼

Output distance matrix

cuments/exercise_data/spatial_statistics/distance_matrix.csv   Browse

0%   OK   Close

- Click *OK* to generate the distance matrix.

- Open it in a spreadsheet program to see the results. Here is an example:

	A	B	C	D	E
1	InputID	MEAN	STDDEV	MIN	MAX
2	0	6099.53439317	2405.74088027	1464.79913222	9455.35096813
3	1	5753.00028096	2446.08407655	387.773659106	8609.09758864
4	2	8171.02431172	2254.5699261	4465.30253163	10950.2008259

7.4.4 🏷 Follow along: Nearest Neighbor Analysis

To do a nearest neighbor analysis:

- Click on the menu item *Vector → Analysis Tools → Nearest neighbor analysis*.
- In the dialog that appears, select the *random_samples* layer and click *OK*.
- The results will appear in the dialog's text window, for example:

Parameter	Value
Observed mean distance	2823.45817848
Expected mean distance	2856.05544691
Nearest neighbour index	0.988586612187
N	100
Z-Score	-0.218346109061

Note: You can copy and paste the results into a spreadsheet. The data uses a (colon **:**) separator.

7.4.5 🏷 Follow along: Mean Coordinates

To get the mean coordinates of a dataset:

- Click on the *Vector → Analysis Tools → Mean coordinate(s)* menu item.
- In the dialog that appears, specify *random_samples* as the input layer, but leave the optional choices unchanged.
- Specify the output layer as `mean_coords.shp`.
- Click *OK*.
- Add the layer to the *Layers list* when prompted.

Let's compare this to the central coordinate of the polygon that was used to create the random sample.

- Click on the *Vector → Geometry Tools → Polygon centroids* menu item.
- In the dialog that appears, select *roads_hull* as the input layer.
- Save the result as `center_point`.
- Add it to the *Layers list* when prompted.

As you can see from the example below, the mean coordinates and the center of the study area (in orange) don't necessarily coincide:
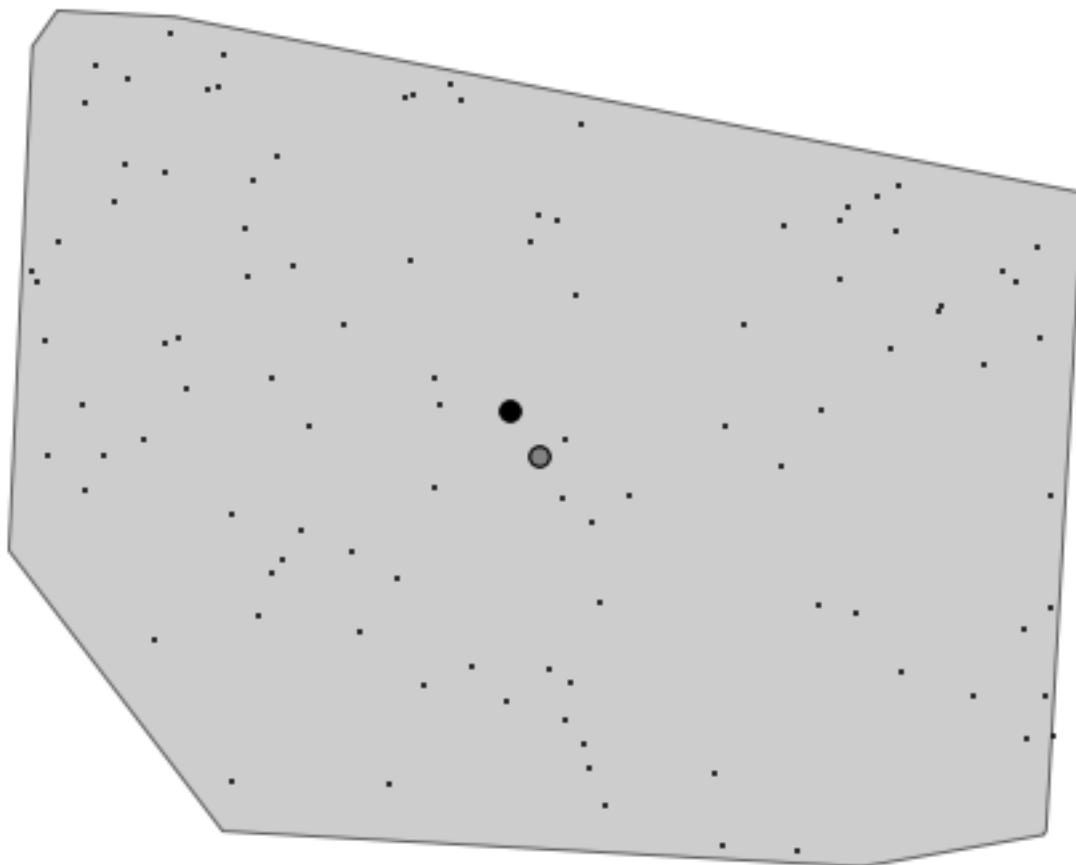
7.4.6 [tag icon] Follow along: Image Histograms

The histogram of a dataset shows the distribution of its values. The simplest way to demonstrate this in QGIS is via the image histogram, available in the *Layer Properties* dialog of any image layer.

- In your *Layers list*, right-click on the SRTM DEM layer.

- Select *Layer Properties*.

- Choose the tab *Histogram*. You will see a graph describing the frequency of values in the image.

- You can export it as an image:
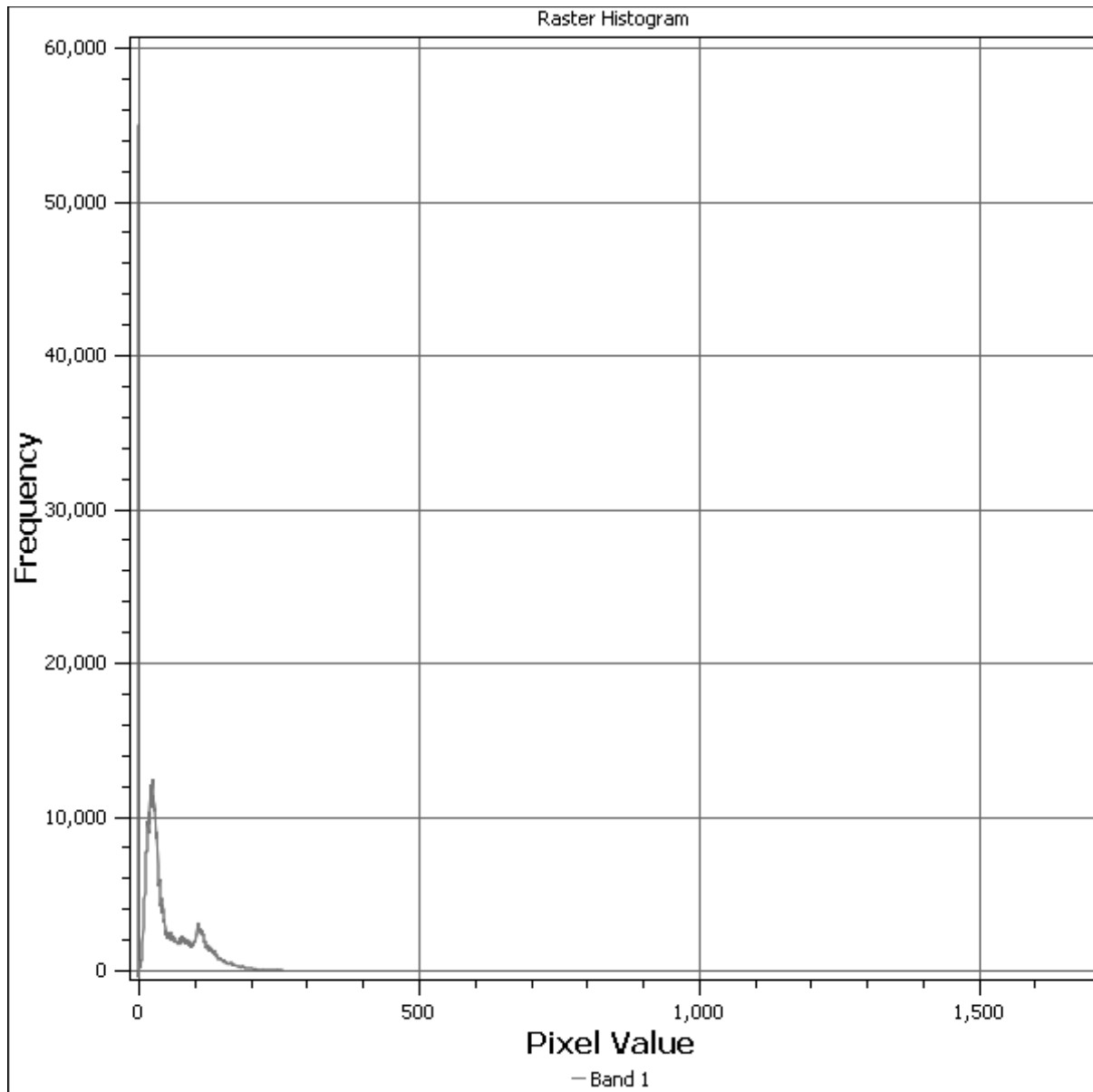
- Look at the metadata for this layer (under the *Metadata* tab).

The mean value is `332.8`, and the maximum value is `1699`! But those values don't show up on the histogram. Why not? It's because there are so few of them, compared to the abundance of pixels with values below the mean. That's also why the histogram extends so far to the right, even though there is no visible red line marking the frequency of values higher than about `250`.

Therefore, keep in mind that a histogram shows you the distribution of values, and not all values are necessarily visible on the graph.

- (You may now close *Layer Properties*.)

7.4.7 Follow along: Spatial Interpolation

Let's say you have a collection of sample points from which you would like to extrapolate data. For example, you might have access to the *random_samples* dataset we created earlier, and

would like to have some idea of what the terrain looks like.

To start, launch the *Grid (Interpolation)* tool by clicking on the *Raster → Analysis → Grid (Interpolation)* menu item.

- In the *Input file* field, select `random_samples`.

- Check the *Z Field* box, and select the value field (not the `id` field).

- Set the *Output file* location to `exercise_data/spatial_statistics/interpolation.tif`

- Check the *Algorithm* box and select *Inverse distance to a power*.

- Set the *Power* to `5.0` and the *Smoothing* to `2.0`. Leave the other values as-is.

- Check the *Load into canvas when finished* box and click *OK*.

- When it's done, click *OK* on the dialog that says `Process completed`, click *OK* on the dialog showing feedback information (if it has appeared), and click *Close* on the *Grid (Interpolation)* dialog.

Here's a comparison of the original dataset (left) to the one constructed from our sample points (right). Yours may look different due to the random nature of the location of the sample points.
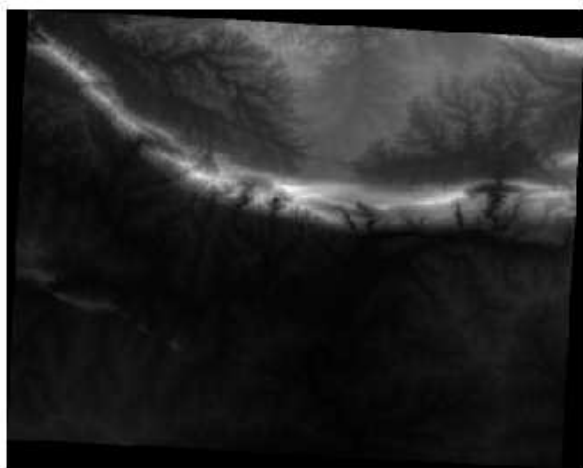

As you can see, 100 sample points aren't really enough to get a detailed impression of the terrain. It gives a very general idea, but it can be misleading as well. For example, in the image above, it is not clear that there is a high, unbroken mountain running from east to west; rather, the image seems to show a valley, with high peaks to the west. Just using visual inspection, we can see that the sample dataset is not representative of the terrain.

## 7.4.8 🏷 Try yourself...

- Use the processes shown above to create a new set of `1000` random points.

- Use these points to sample the original DEM.

- Use the *Spatial Interpolation* tool on this new dataset as above.

- Set the output filename to `interpolation_1000.tif`, with *Power* and *Smoothing* set to `5.0` and `2.0`, respectively.

The results (depending on the positioning of your random points) will look more or less like this:
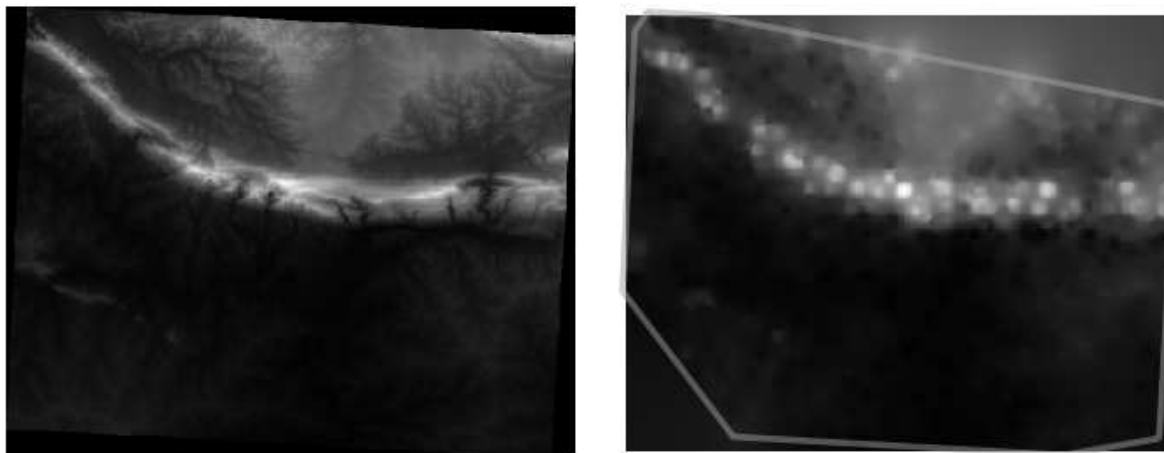
The border shows the *roads_hull* layer (which represents the boundary of the random sample points) to explain the sudden lack of detail beyond its edges. This is a much better representation of the terrain, due to the much greater density of sample points.

Here is an example of what it looks like with 10 000 sample points:
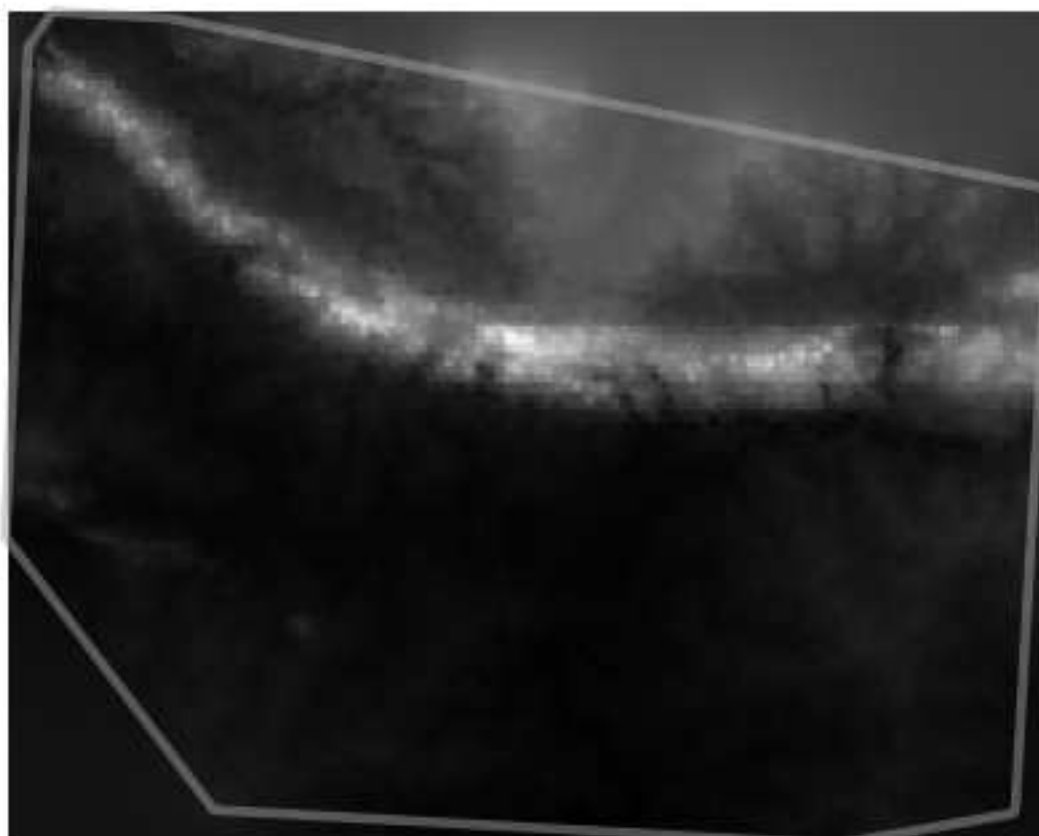
Note: It's not recommended that you try doing this with 10 000 sample points if you are not working on a fast computer, as the size of the sample dataset requires a lot of processing time.

7.4.9 🏷 Follow along: Installing SEXTANTE

The SEXTANTE plugin allows you to access various plugin tools from within a single interface. It is standard in QGIS starting with version 1.8. If you're unsure whether you have SEXTANTE, check if it's marked as installed in your *Plugin Installer*.

Assuming you have it installed (and enabled in the *Plugin Manager*):

- Activate SEXTANTE by enabling the *View → Panels → SEXTANTE Toolbox* menu entry. The toolbox looks like this:
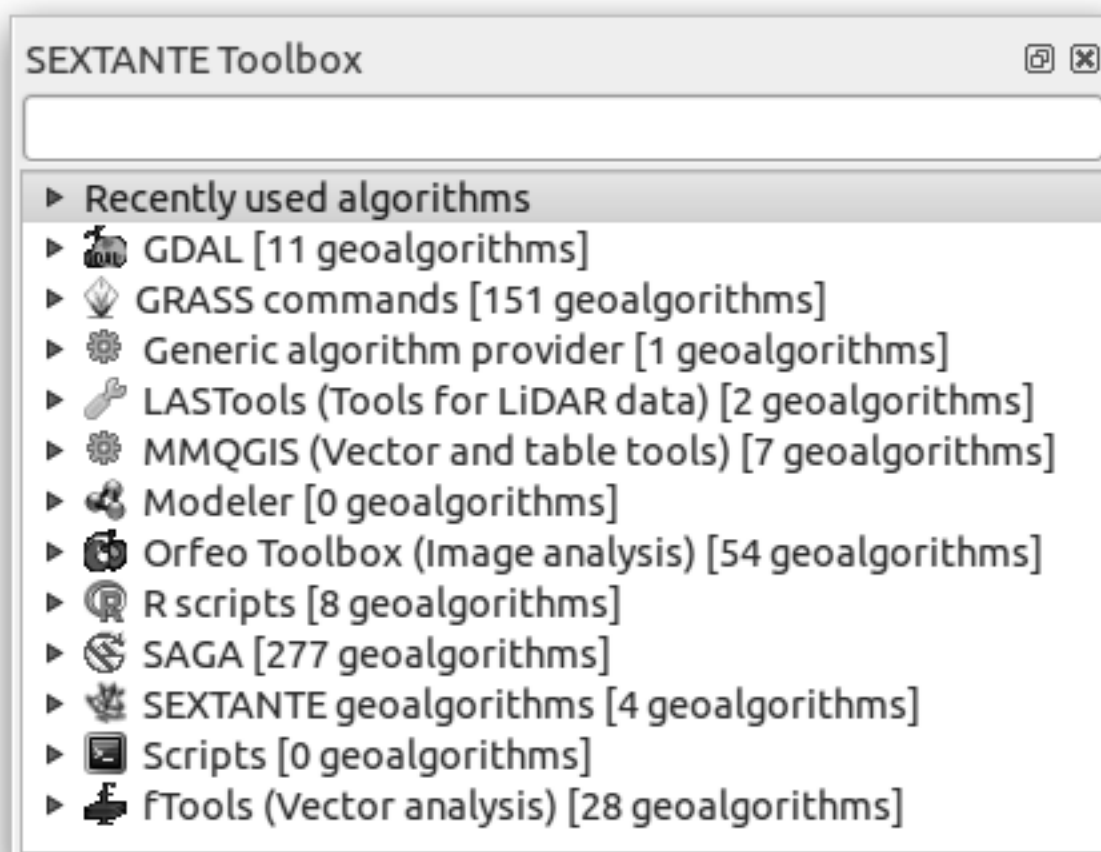
You will probably see it docked in QGIS to the right of the map. Note that the tools listed here are links to the actual tools. SEXTANTE doesn't have many tools of its own. To have access to spatial statistics tools, you will need to install the SAGA GIS program.

7.4.10 🏷 Follow along: Installing SAGA

SAGA is a GIS program with many useful analysis functions. To make use of these functions in QGIS via SEXTANTE, you need to install SAGA first.

On Windows

Included in your course materials you will find the SAGA installer for Windows.

- Start the program and follow its instructions to install SAGA on your Windows system. Take note of the path you are installing it under!

Once you have installed SAGA, you'll need to configure SEXTANTE to find the path it was installed under.

- Click on the menu entry *Analysis → SAGA options and configuration*.

- In the dialog that appears, expand the *SAGA* item and look for *SAGA folder*. Its value will be blank.

- In this space, insert the path where you installed SAGA.

On Ubuntu

- Search for *SAGA GIS* in the *Software Center*, or enter the phrase `sudo apt-get install saga-gis` in your terminal. (You may first need to add a SAGA repository to your sources.)

- QGIS will find SAGA automatically, although you may need to restart QGIS if it doesn't work straight away.

After installing

Now that you have installed and configured SAGA, its functions will become accessible to you.

7.4.11 Follow along: Spatial Point Pattern Analysis

For a simple indication of the spatial distribution of points in the *random_samples* dataset, we can make use of SAGA's *Spatial Point Pattern Analysis* tool.

- In the *SEXTANTE Toolbox*, find this tool under *SAGA → Geostatistics → Spatial Point Pattern Analysis*.

- Double-click on it to open its dialog.

- It produces three outputs, and so will require three output paths.

- Save these three outputs under `exercise_data/spatial_statistics/`, using whatever file names you find appropriate.
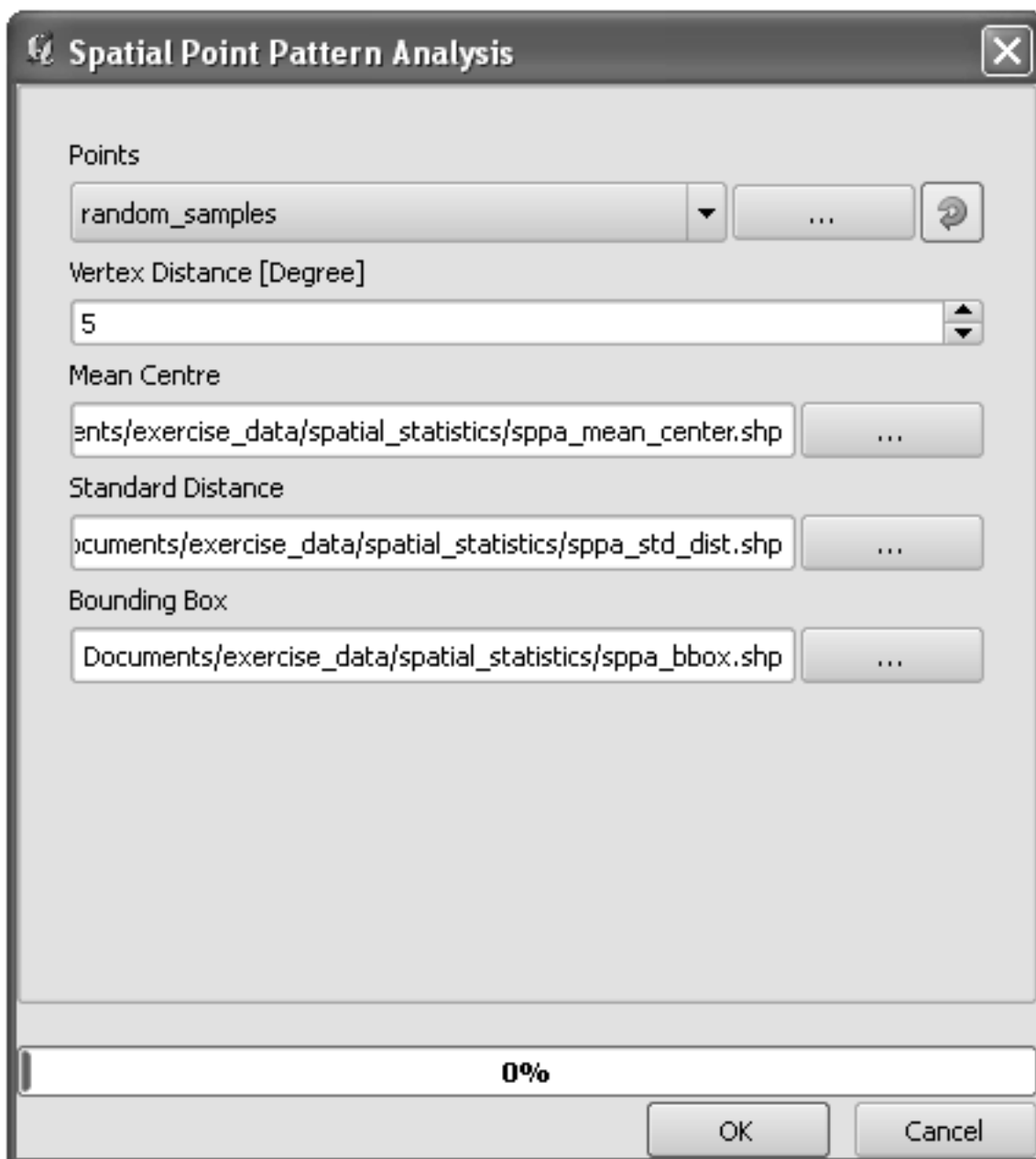
The output will look like this (the symbology was changed for this example):
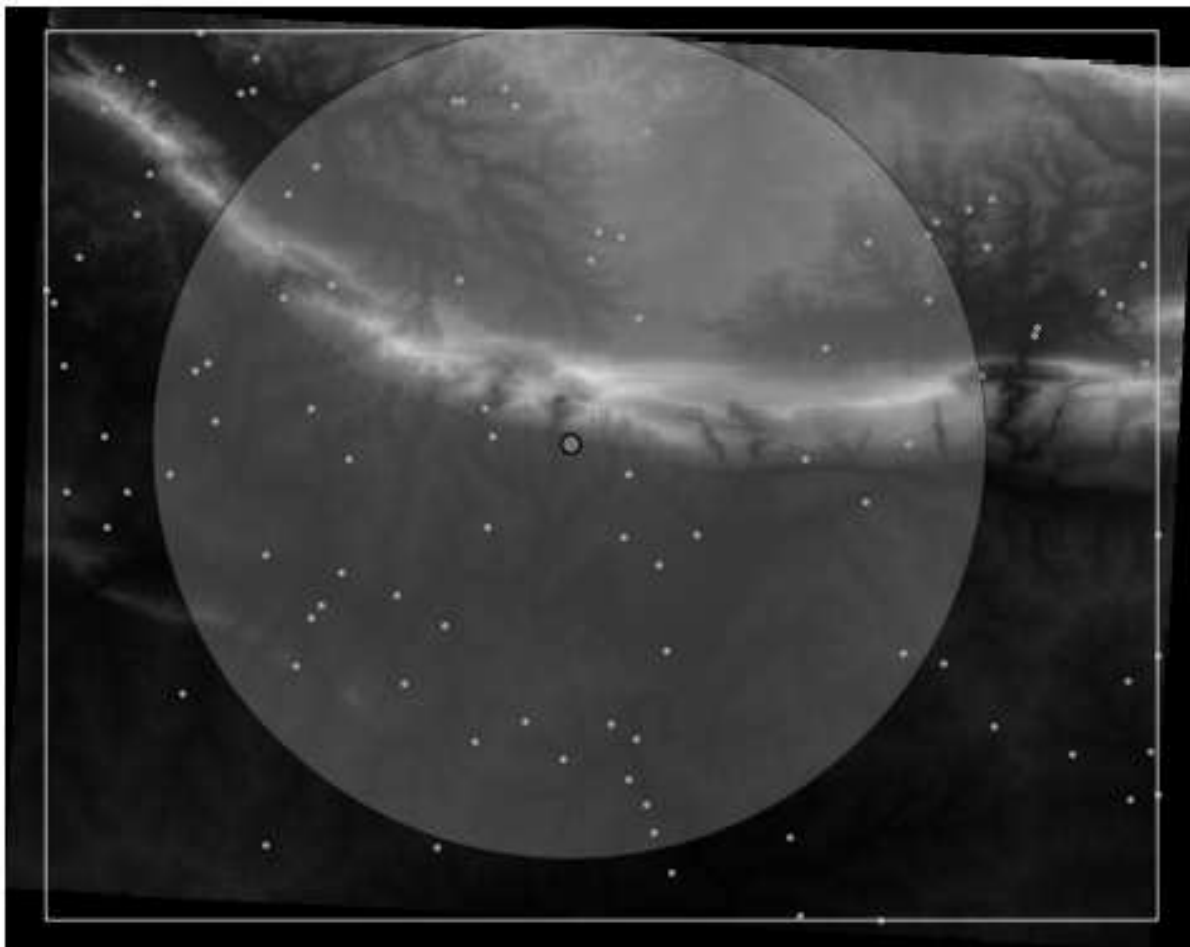
The red dot is the mean center; the large circle is the standard distance, which gives an indication of how closely the points are distributed around the mean center; and the rectangle is the bounding box, describing the smallest possible rectangle which will still enclose all the points.

7.4.12 Follow along: Minimum Distance Analysis

Often, the output of an algorithm will not be a shapefile, but rather a table summarizing the statistical properties of a dataset. One of these is the *Minimum Distance Analysis* tool.

- Find this tool in the *SEXTANTE Toolbox* as *SAGA → Geostatistics → Minimum Distance Analysis*.

It does not require any other input besides specifying the vector point dataset to be analyzed.

- Choose the *random_points* dataset.

- Click *OK*. On completion, a DBF table will appear in the *Layers list*.

- Select it, then open its attribute table. Although the figures may vary, your results will be in this format:

	NAME ▽	VALUE
0	Mean Average	2823.45817848
1	Minimum	424.0860061
2	Maximum	9773.35250512
3	Standard Deviation	1662.40681133
4	Duplicates	0

7.4.13 In conclusion

QGIS allows many possibilities for analyzing the spatial statistical properties of datasets.

7.4.14 What's next?

Now that we've covered vector analysis, why not see what can be done with rasters? That's what we'll do in the next module!