

Crime Mapping and Spatial Analysis

Lab Manual

John Palmer

2020-03-10

Contents

Cover	7
Introduction	9
Accessing QGIS	9
Accessing Stata	9
Accessing R	10
Lab Instructions	11
Lab 1: 9 January 2020	11
Lab 2: 23 January 2020	12
Lab 3: 5 February 2020	12
Lab 4: 20 February 2020	13
Lab 5: 5 March 2020	13
I Crime Mapping	15
1 Getting Started with QGIS	17
1.1 Beginning a new project	17
1.2 Changing the language	17
1.3 Installing plugins	17
1.4 Adding OpenStreetMap tiles	18
1.5 Placing points on the map	18
1.6 Placing polygons on the map	18
2 Mapping Basics	21
2.1 Mapping crime locations	21
2.2 Working with Multiple Layers	26
2.3 Printing or Exporting a Map	27
3 Creating Vector Data From Raster Base	29
3.1 Obtaining a raster	29
3.2 Creating a new vector layer.	31
3.3 Adding features	31

4	Adjusting Vector Layer Appearance	33
4.1	Opening the properties window	33
4.2	Changing fill color, border and style	33
4.3	Adding labels	34
5	Adding OpenStreetMap vectors	35
5.1	Downloading OSM data	35
5.2	Labeling the vector data	37
5.3	Filtering the vector data	37
6	Mapping Aggregated Data	39
6.1	Choropleth Maps	39
6.2	Proportional Symbols	41
6.3	Pie Charts	42
II	Spatial Analysis	43
7	Getting Started with QGIS	45
8	Getting Started with Stata	47
8.1	Menus	47
8.2	Commands	47
8.3	Do-Files	48
9	Getting Started with R	51
9.1	Using R from a Terminal	51
9.2	Text Editor	52
9.3	IDE	52
9.4	Basic Concepts	52
10	Buffers and Distances	53
10.1	Getting the Data	53
10.2	Mapping the Data	54
10.3	Drawing Buffers Around the Schools	54
10.4	Counting Points in Polygons	54
10.5	Calculating Distances	54
10.6	Joining and Exporting Data	55
11	Cluster Analysis	57
11.1	Using QGIS	57
11.2	Using Stata	60
11.3	Using R	60
12	Autocorrelation and Spatial Regression	63
12.1	Loading the Data into Stata	63
12.2	Creating Spatial Weights Matrix	65

<i>CONTENTS</i>	5
12.3 Regression and Spatial Regression	68

Cover

Introduction

This manual is intended to function as a companion to the *Crime Mapping and Spatial Analysis* labs as well as a stand-alone reference for working with QGIS, Stata, and R for mapping and analysis. Each chapter walks you through a particular task or set of tasks, usually with an exercise to serve as an example. These are organized according to the class lab sessions but can also be followed independently or used as a reference.

Accessing QGIS

QGIS is free and open source software, which you can download from <https://qgis.org>. At the time of writing, version 3.4 is the most recent long-term stable release and this manual is based on that version. When installing QGIS, leaving all of the default options as they are should be fine for purposes of this manual. The Windows installation includes a number of different choices for running QGIS; you should simply select the normal desktop version.

Accessing Stata

Stata is proprietary software, which cannot be freely downloaded if have not purchased a license. However, as a UPF student, you have free access to Stata on the UPF lab computers and through UPF's *MyApps* platform. The latter requires an internet connection and has some limitations in terms of resources, but you can access it from outside UPF using your own computer and a web browser.

To access Stata through *MyApps*, open a web browser and go to <https://myapps.upf.edu>. Log in using your UPF user ID (uxxxxx) and the password you use for logging into UPF computers (i.e. your birthdate, written as DDMMYYYY). Remember that this is not necessarily the password you use for Campus Global; your browser may have that Campus Global password stored, in which case you will need to manually change it each time. On the login page, you will need

to choose between using *vWorkspace Connector* or simply through your web browser with HTML5. The first option allows you to access local files on your computer, but it can be slow and “buggy”, particularly on Macs. To follow this approach, click the *Install* button on the page you reach after logging in. The second option can be faster, but any files you want to use will need to be first uploaded through the UPF MyCloud. To follow this approach, click the *Continue* button on the page after login, and then flip the *HTML5* switch to the on position at the top right of your screen.

Accessing R

R is free and open source software, which you can download from <https://cran.r-project.org/>. You will normally want to install the latest release from that site, and to then install a good text editor or Integrated Development Environment (IDE). An excellent choice for the latter is the (free) Open Source Edition of RStudio, which you can download from <https://www.rstudio.com>. R and RStudio are also both available on the lab computers. This manual will assume that you are running RStudio.

Lab Instructions

This section provides instructions for each lab session

Lab 1: 9 January 2020

The core of criminology research involves observing the world around you, thinking about it, and communicating your thoughts to others. Computers can be useful tools for this, and most of the labs will focus on how to make use of them. Yet, computers can also get in the way. They can distort our observations and encourage us to think and communicate in certain ways. For this lab, therefore, we will start by actually going outside into the real world, making observations, and using a pen and paper to think about them and communicate them.

Your task for this lab is to observe the courtyard in the middle of the Jaume I building, think about the following questions, and draw a map to explain your thoughts to someone else:

- How many people are in the courtyard? Note, of course, that you will need to make some decision about time to answer this question: How many people right now? How many people over the course of an hour? Think about how the way you refine the question in this way will affect not just your answer but also the methods you use for reaching it.
- Where are these people? In other words, how are these people you observe in the courtyard spatially distributed? (Again, time comes into play here, so we might be better off asking how they are spatio-temporally distributed.)
- What are these people doing and where are they doing it? For example, are people reading, looking at their phones, talking, walking, running...? Where are these activities taking place? For activities that involve movement, how can you best communicate how this movement is occurring?
- What aspects of the physical environment (benches, walls, entry and exit points...) seem to play an important role in the social activity you are observing?

We will proceed with this in several steps:

1. Conduct an initial assessment of the courtyard and develop a strategy for answering the questions and communicating them on a map. You can do this in groups or on your own.
2. We will then reconvene as a class to discuss your strategies.
3. Now return to the courtyard and carry out your observations, analysis, and mapping. You can work in groups but you will each need to produce a set of maps that convey your findings. These maps should be drawn by hand.
4. We will then reconvene again to discuss your finished maps.
5. Finally, we will start using QGIS using the *Getting Started with QGIS* section of this manual.

At the end of the lab, you should hand in your hand-drawn maps. They will be evaluated based on your ideas, not based on how well you draw, how neat they are etc. The goal here is to get you to think through what how to communicate this information visually. How you then execute this technically, will form much of what we cover in the coming labs.

Lab 2: 23 January 2020

In this lab, we will learn how to add create a map using point data. Follow the instructions in Section 2 to create a map of criminal complaints in New York City using open data from the New York City police department. Set up the map for printing using a layout that includes not only the map but also a scale bar and a legend. Export this as a PNG file.

Next, explore sources of point data that relate to your research interests and that you will consider using for your final mapping project. Create a new map using one of these sources and export it as a PNG file.

Submit both PNG files through Aula Global.

Lab 3: 5 February 2020

In this lab we will learn how to create your own cartographic vector data, how to use and modify existing sources of cartographic vector data, and how to combine data with polygons to make choropleth maps and to add proportional symbols and charts. Follow the instructions in Sections 3, 4, 5, and 6 to make two maps that you will need to submit through Aula Global:

1. A map of the area around this building, which includes a polygon that you will draw over the Jaume I courtyard, and street information from Open Street Maps.
2. A map of Europe showing criminal offense statistics.

Submit both PNG files through Aula Global.

Lab 4: 20 February 2020

In this lab we will learn how to use buffers and distances for spatial analysis. Follow the instructions in Section 10 and upload the csv file that you export in the end to Aula Global.

Lab 5: 5 March 2020

This lab is divided into two parts: In part 1 we will learn how to implement k-means clustering using QGIS. In part 2, we will learn how to use Stata to test for spatial autocorrelation and fit spatial regression models.

1. Download or reuse any set of point data that you want. Load this into Stata as a points layer, and then use this to work through Section 11 to cluster the data using QGIS. Use colors to indicate the clusters on your map. Export the color-coded, clustered map as a png file and submit it through Aula Global.
2. Start with the dataset (a csv file) that you generated in the previous lab (number of bars and number of battery reports within the buffer drawn around each school in Chicago). If you cannot find this, you can quickly recreate it or ask for a copy. Use this dataset to work through Section 12. Put your results into a Word document, including a brief interpretation of the regression models, and upload this through Aula Global.

Part I

Crime Mapping

Chapter 1

Getting Started with QGIS

1.1 Beginning a new project

In QGIS, “projects” are used to store the settings you are using. It is helpful to start a new project for each map or set of related maps that you are making. This way, you can return to your work later and start from where you left off. Note, however, that the project file does not store the data you are using or the maps you generate. These must be stored as separate files.

To start a new project use the menu item Project → New.

1.2 Changing the language

When you start QGIS it will normally be set to the default language of your computer’s operating system. To change the language, use the menu item Settings → Options. In the options menu select the General tab. Check the box Override system locale and select the language you want. You will need to restart QGIS for the change to go into effect.

1.3 Installing plugins

QGIS makes it easy for the QGIS development team as well as anyone else to write “plugins,” which add particular features or functions to the program. Many of these are extremely useful.

To install a plugin, use the menu item Plugins → Manage and Install Plugins.... This will bring up a window with a long list of available plugins on the QGIS official repository. (There are other plugin repositories that individual authors

maintain but for now we will use only the official one.). Use the search bar or scroll down the list to find and install the plugin you want.

1.4 Adding OpenStreetMap tiles

To quickly make use of existing map tiles from OpenStreetMap, go to Layer → Data Source Manager. Click the Browser folder at the top left. In the main part of the window, find the XYZ Tiles category and click it. You should see and OpenStreetMap item within this category now. Double click on this and you will the OpenStreetMap tiles will be added to your map.

1.5 Placing points on the map

The most straightforward (although least efficient) way to mark points on your map is to do it manually.

To do this, you first need to create a new map layer using Layer → Create layer → New Shapefile Layer.... In the window that pops up, you can choose to make this a layer of points, lines, or polygons (see Geometry Type selector, third from top in version 3.4). You can also modify the character encoding (File Encoding selector), the Coordinate Reference System (selector just below the File Encoding selector), and the data fields that will be associated with this new layer.

For now, keep the default values (which will produce a point layer), choose a directory and file name (top row: File name) and then click OK.

In order to add a point, move the cursor over your new layer in the Layers Panel and right-click on it. Then select Toggle Editing. Now select Edit → Add Point Feature. Click on the map wherever you want to add a point. Each time you click, you will need to provide attributes for the fields you created when you created the layer. For now, since we used the defaults, we need only an ID, which requires an integer value. Give this whatever (integer) value you want.

1.6 Placing polygons on the map

In addition to points, it may be useful to add polygons to a map in order to communicate information about structures and areas. To do this, start again by creating a new vector layer using Layer → Create Layer → New Shapefile Layer.... In the window that pops up, choose Polygon as the type, and add at least one new field in which you will store information associated with each feature that you draw. Save the new layer as an SHP file in a directory where you will be able to find it again.

The new layer you have added should appear in the Layers Panel. Right click on this new layer and select Toggle Editing. You will now see a set of new buttons and available menu items. Now select the menu item Edit → Add Polygon Feature. When you hover the mouse pointer over the map you will see that it now has the shape of a cross-hairs. Left click to add sequential points that define the polygon you want to add. When you have added the last point, right click the mouse and a dialog box will appear asking you to add the fields associated with this new feature. Remember that what you add will be limited by the field type you chose (and the ID field requires whole numbers by default).

Chapter 2

Mapping Basics

2.1 Mapping crime locations

2.1.1 Downloading Data

To get started, we will use the New York City Police Department’s (NYPD’s) most recent complaint data as a way to construct a basic map of reported crime locations. This data set is available on New York City’s open data website and described as including “all valid felony, misdemeanor, and violation crimes reported to the [New York City Police Department [NYPD] for all complete quarters so far this year.” At the time of writing, the temporal scope of this data is all of 2019 (since there are not yet any completed quarters in 2020.)

To start exploring the data, go to the website at:

<https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Current-Year-To-Date-/5uac-w243>

The website allows you to view or visualize the data directly online, as well as to export the data and to access it through an API. You could export the entire dataset using Export button:

If you do this, however, you may end up with a very large file and this may slow things down in terms of download and processing time. For purposes of this first exercise, instead click the View Data button, as this will allow you to filter the data before downloading it:

Clicking on the View Data button will take you to a new page that includes a Filter button. Click on this and then click the Add a New Filter Condition button. You can easily filter the data by any of its fields (or by multiple fields). For this exercise, filter the OFNS_DESC field (which is the description of the offense, as recorded by the NYPD) for FELONY ASSAULT:

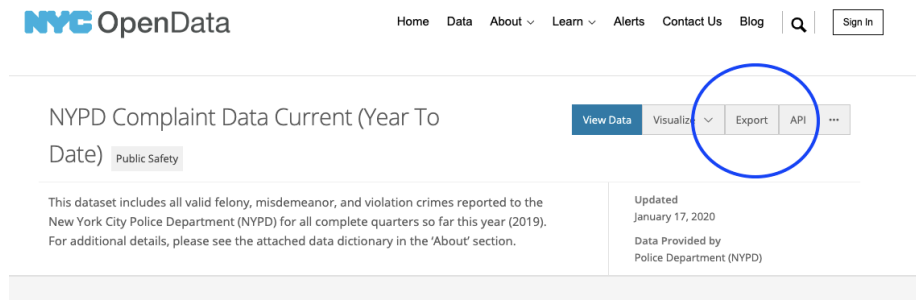


Figure 2.1: Location of Export button on the NYPD Complaint Data page of the NYC open data website

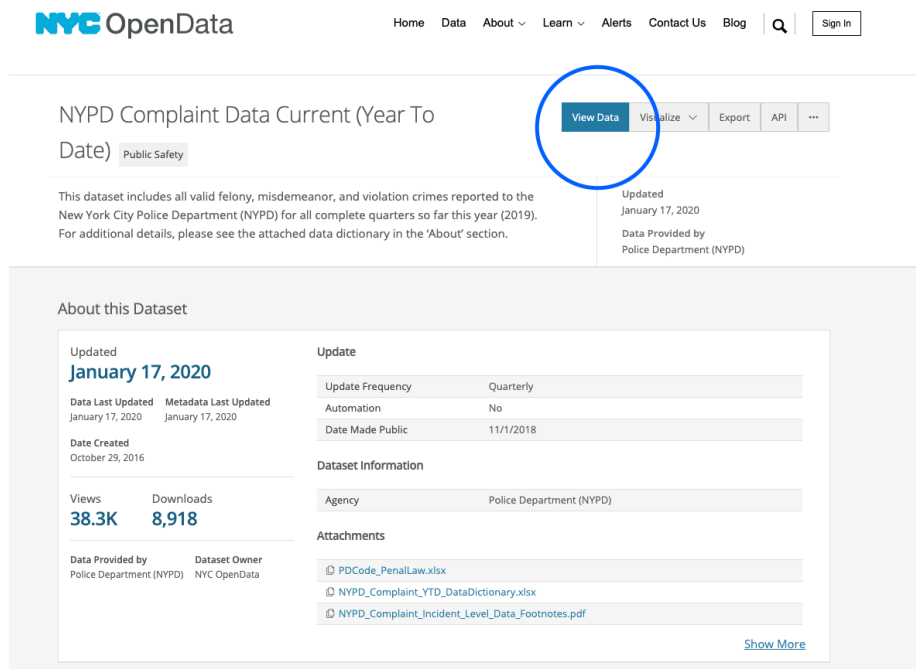


Figure 2.2: Location of the View Data button

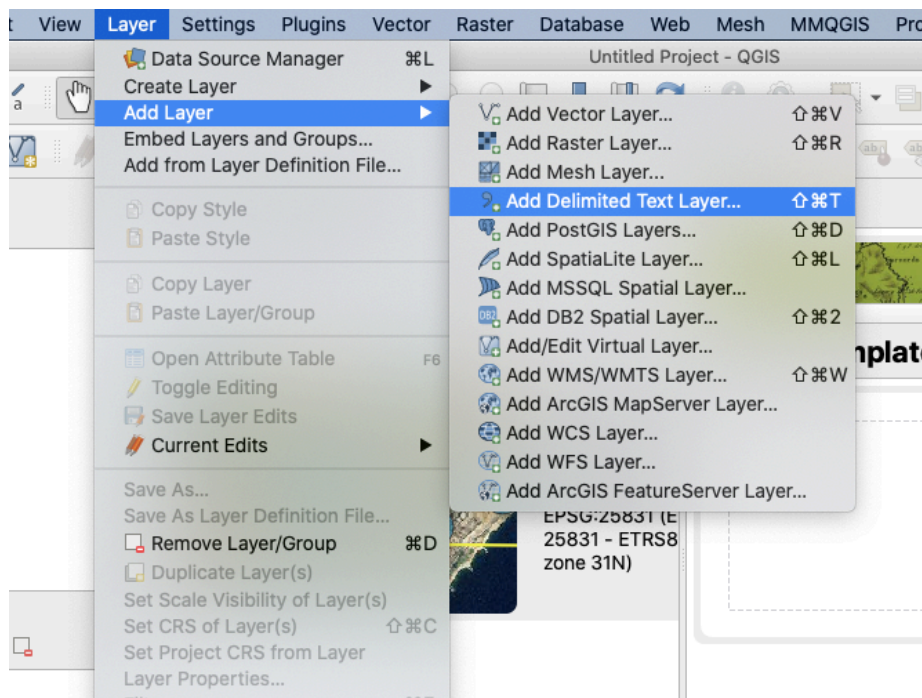


Figure 2.5: Screenshot of the Add Delimited Text Layer menu option in QGIS 3.10 for MacOS

In the dialog that comes up, click on the ... button to the right of the File name box at the top. Select the data file you just downloaded and saved.

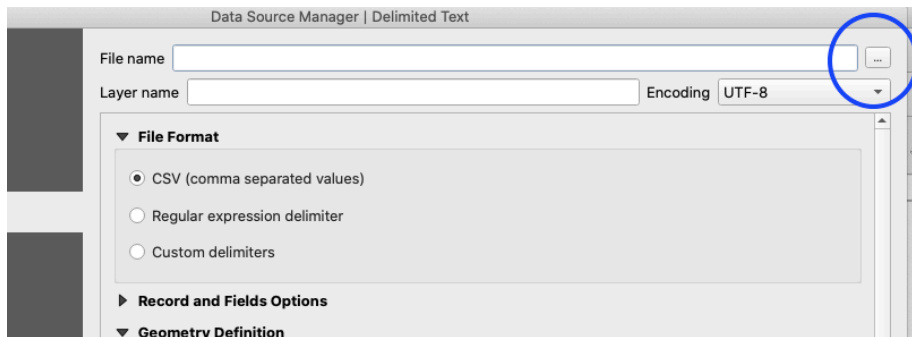


Figure 2.6: Choosing file in the dialog (button at top right with blue circle around it.)

Once you select the file in the dialog box, you will see that it is listed in the File name box. Note that you can also choose a Layer name. This is how the layer you are creating will be named on the layers bar of your QGIS interface. You will need to ensure that the proper File format is selected (in this case CSV) and you will need to provide information about geometry: In the Geometry definition box, make sure Point coordinates is selected. Make sure that the X field and Y field boxes contain the correct field names from your data for the x and y coordinates (in this case, the correct fields are Longitude and Latitude).

Now look at the Coordinate Reference System (CRS) selector, which is labelled Geometry CRS. This is a crucial setting because selecting the wrong CRS will lead to your layer being incorrectly placed on the map (or sometimes not visible at all). For points recorded as Latitude and Longitude, the correct CRS is EPSG 4326 - WGS 84, so make sure this is selected. We will talk more about other CRSs and how to determine which one your data relies on.

Note: Some of these settings may already be correctly set because QGIS automatically tries to figure them out from your file. So in many cases you will simply need to go through and double check them.

Once you have ensured that the settings are correct, look at the preview shown in the Sample Data section at the bottom of the window. You should see a table with data in rows and columns with headers. If you do not see this, go back and check the settings again.

Note: One common problem when importing CSV data is that QGIS may not be properly recognizing the character used to separate columns. CSV stands for comma-separated values, and in many systems this means that the columns are separated with commas (",") (as the name implies). In some systems, however, the default is to use semi-colons (";") instead of commas - and this is often the default on computers and software localized for Spain. Thus, for example, if you

save an Excel sheet as a CSV file on a UPF lab computer, you may find that it is saved using semi-colons rather than commas. You can easily see whether it is commas or semi-colons (or something else) by simply opening your CSV file in a text editor and looking at it closely. If it is not a comma, you will need to specify which character it is by choosing the Custom delimiters setting and then either checking one of the preset boxes or simply typing in the character in the Others box.

Once you have finished with the settings, click the Add button at the bottom of the window.

2.1.3 Adding data using an API

We just added data by first downloading it to your computer because you will often need to work with data files in this way. Data providers, however, are increasingly making it possible to access their data sources using an Application Programming Interface (API). This can simplify things for you, particularly when you will be working with a source frequently and will want the most updated data. We will discuss ways of working with APIs later on.

2.2 Working with Multiple Layers

When you first add the NYPD Complaint Data (or any other data source), you should be able to see it in the QGIS mapping window. Often, however, you will want to combine it with additional layers. For example, the NYPD Complaint Data will be much more useful if it is overlaid on top of a map of New York City. You can add such a map in many ways, but for now add Open Street Map tiles using the instructions in Section 1.4.

One issue you will sometimes need to deal with is finding the features in a particular layer you have added. For example, if you have first added Open Street Map tiles and then added points, it may be that your map window is focused on part of the world where there are no points. To view your new layer, you will need to pan and zoom to get to the location where it actually has data - in this case New York City. You can do that quickly by right-clicking on the layer in the Layers panel and selecting Zoom to Layer at the very top of the menu.

You may also need to make sure that the layer is not being covered by another layer. The order of the layers in the Layers Panel is the order in which they appear on the map (imagine actual layers of paper or transparent paper with features on a physical map). You can drag layers up and down in that list to change the order.

To learn more about each point marked on the map, select View → Identify Features and click on a point. You will see information in the Identify Results

panel on the right (although you may need to expand that panel to see the values associated with each feature).

2.3 Printing or Exporting a Map

What you see in the QGIS console is useful for creating, modifying, and analyzing your map. If you want to share your map with someone or publish it as part of a report, you will want to use the print composer to produce high-quality output.

2.3.1 Setting up the Print Layout

Select Project → New Print Layout. In the dialog box that pops up, you can give this print layout a title or leave the box blank and click OK for the default title. You will now see a new window with a canvas that you can use to compose exactly how you want your map to look.

2.3.2 Adding a map

The most obvious thing you will want on your composed map is the map itself. Select Add Item → Add Map and then use the mouse to draw a rectangle on the canvas where you want the map to appear. Notice that the map that will now appear in that box is essentially a mirror image of the map you currently have in the main console (although the dimensions may be different depending on how you draw the box). After drawing the box, you can still move it around using the Edit → Select/Move Item tool, and you can move the map around within it using the Edit → Move Content tool. Notice that you can add more than one map to your canvas. This makes it possible to include inset maps, for example, showing where your detailed map lies at a larger scale.

2.3.3 Adding a scale bar

You may wish to add a scale bar to your map, and you can do this easily using Add Item → Add Scale Bar. Again, you will need to click on the map to position and resize the scale bar. You will notice that you also have many options to choose from in changing its appearance using the Item properties tab on the right.

2.3.4 Adding a legend

To add a legend, select Add Item → Add Legend and use the mouse to draw a rectangle where you want the legend to appear. The legend will be automatically

filled with the information in your layers list, but you can modify this using the Item properties tab on the right. Select the legend you have added and scroll down the Item properties to the Legend Items section. Uncheck Auto update. Now select the layers you do not wish to see in the legend and click the minus sign at the bottom to remove them. Next click the item(s) that you want in the legend and use the editing button to edit the text that will appear. In the Main properties section you can also change the title and alignment of the legend text.

2.3.5 Exporting your map

When you are happy with your map, use Layout → Export as Image or Layout → Export as PDF or Layout → Export as SVG to export your map to a single file, which you can then share, embed in a report, or print.

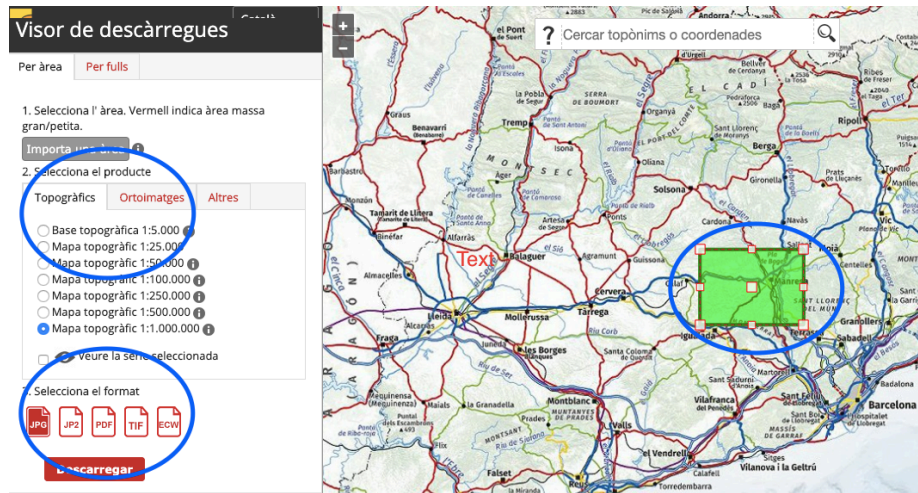
Chapter 3

Creating Vector Data From Raster Base

It will often be necessary to create vector data – points, lines, or polygons – by tracing features on an existing raster map. This chapter will walk you through that process.

3.1 Obtaining a raster

We will start by finding and downloading a raster base. For example, you could use a high resolution orthophoto, such as those available at <http://www.icc.cat/appdownloads/index.html>. When you first enter this site, it will be set to download topographical maps. Change this setting to orthophoto by clicking on the “Ortoimatges” tab in the left sidebar. Also change the format option in the bottom of the left sidebar to “tif” (for GeoTiff), and select the highest resolution: 1:2.500. Now drag the green square in the mapping area to the region you want to download. For this exercise, zoom in and select UPF’s Jaume I building, reducing the size of the green box as you do this so that it just covers the building itself. (Note that with this resolution selecting a large area will result in a very large file.) Once you have the green box in place, click the Descarregar button and enter your email address. The platform will send a link to the file to your email within a few moments and you can then click to download it.



Adding raster to map

By saving the file in GeoTiff format you have ensured that it includes CRS information that QGIS can read automatically. Add the layer using Layer → Add Layer → Add Raster Layer... and then selecting the file you downloaded. Depending on the CRS of your current map, QGIS may then give you a choice of how to transform your raster's CRS to the map's CRS. You can select one of these or you can go back to your map and change it's CRS to match the raster's.

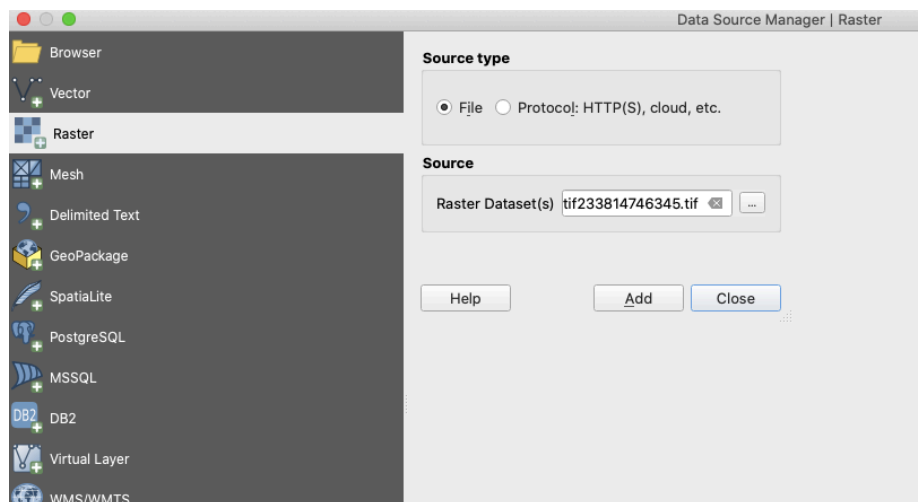


Figure 3.1: Adding raster layer.)

3.2 Creating a new vector layer.

Create a new vector layer using Layer → Create Layer → New Shapefile Layer.... In the window that pops up, you can choose to make this a layer of points, lines, or polygons (see Geometry Type selector, third from top). You can also modify the character encoding (File Encoding selector), the CRS (selector just below the File Encoding selector), and the data fields that will be associated with this new layer. For this exercise, choose Polygon as the type, and add at least one new field in which you will store information associated with each feature that you draw. For this exercise, we will be using the layer to indicate the approximate number of people who you observed in the courtyard during our first lab, so create a Whole Number field and give it a name (e.g. N). In the File section specify the file name and directory where you will save this new layer. Then click OK.

3.3 Adding features

The new layer you have added should appear in the Layers Panel. Right click on this new layer and select Toggle Editing. You will now see a set of new buttons and available menu items. Select Edit → Add Polygon Feature. When you hover the mouse pointer over the map you will see that it now has the shape of a cross-hairs. Left click to add sequential points that define the polygon you want to add. When you have added the last point, right click the mouse and a dialog box will appear asking you to add the fields associated with this new feature. Remember that what you add will be limited by the field type you chose (and the ID field requires whole numbers by default). For this exercise, draw a rectangle over UPF's Jaume I courtyard.

Chapter 4

Adjusting Vector Layer Appearance

QGIS gives you control over how your vector layers appear: You can adjust color, fill style, transparency, labeling, and many other characteristics.

4.1 Opening the properties window

To do this, you need to start by opening the Layer Properties window. Right click on the layer you want to adjust and select Properties. You will see a new window with a set of tabs on the left.

4.2 Changing fill color, border and style

To change the layer's color, select the Symbology tab. If this is a polygon layer, you will see a box with Fill and under this Simple fill. Click on Simple fill and then scroll down and click the Fill color selector and choose a color. Notice that you have different options for picking the color, including choosing one from a pallet or gradient, entering HSV or RGB values, or entering an HTML code. You can also choose the level of opacity, which lets you make semitransparent features. (Note that this can be set for the fill color itself or for the entire layer.)

You can also change the border and fill style using the Stroke color and Fill style selectors.

4.3 Adding labels

To add labels to your layer, select the Labels tab. At the very top of the window, change the selector from No labels to Single labels. Just underneath this selector, you will see a selector for Value. Use this to select the field you wish to use for your labels. If you created a field in the exercise above for number of people, you can use this to automatically display this information on each feature. Alternatively, you might have a field for name or ID. The idea here is that QGIS can automatically use one of the fields to generate labels. Try making labels and click OK to see how they look. Then return to the Properties window to explore the various options you have for adjusting the labels' appearance.

Chapter 5

Adding OpenStreetMap vectors

One useful way to add existing vector information to a map, while also making it possible to customize its appearance, is to download data from OpenStreetMap (OSM). The process will be to download the OSM file for a given area, then open it, choose the layers that you want, filter these layers according to feature type, and then modify appearances. To do this, you will need to install the OSMDownloader plugin. Open the Plugins menu, choose Manage and Install Plugins.... Search for “OSMDownloader” and install it. Notice that a small button now appears on the top left, just above the left pane. This is what you will use for downloading the vector data.

5.1 Downloading OSM data

Start with some raster base map that lets you navigate to a place of interest (OSM map tiles, an orthophoto etc.). Zoom in to a small area (if the area is too large, the download size can be huge). For purposes of this exercise, zoom to the area immediately around UPF’s Ciutadella campus (no more than a few blocks). Now click the OSMDownload button at the top left and drage the cursor on the map to select the rectangle for which you will download vector data.

In the window that appears, select Save File and choose a name and directory for the files you will download. Also click the Load Layer After Download box. Then click OK. Once the data is downloaded, you will see a new window with a list of layers that this file contains. You can select some or all of these. When you hit OK you will see each of these layers appear on your map and in your Layers Panel.

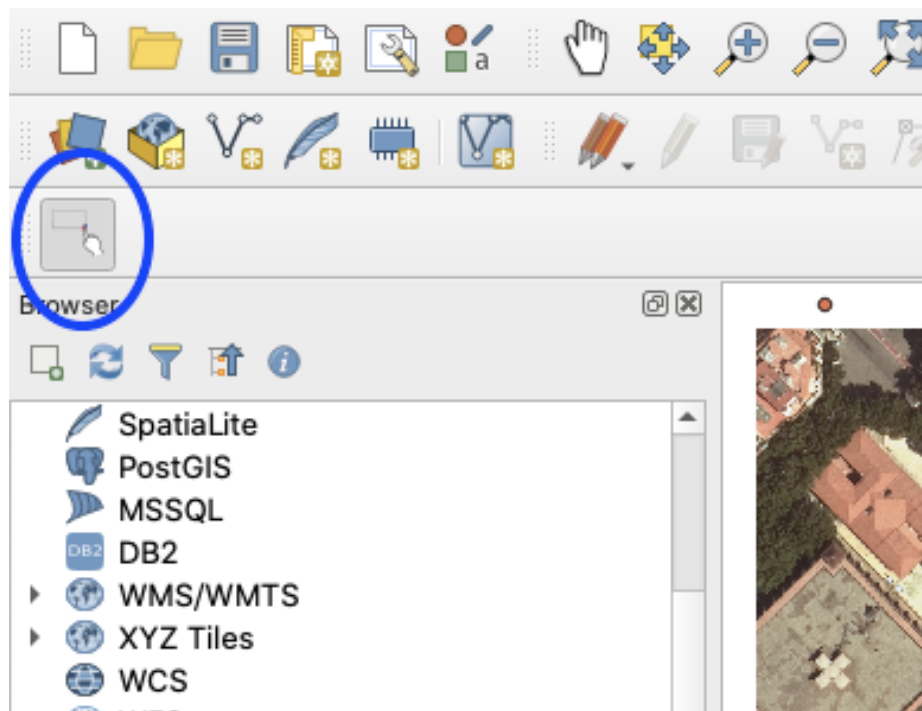


Figure 5.1: Adding raster layer.)

5.2 Labeling the vector data

You can add labels to this new layer just as you did above. You may find, however, that there is too much information and that you want to filter it somehow. You can filter the labels themselves using the labels tab. Or you can filter the features in the layer as follows.

5.3 Filtering the vector data

To filter a layer, right click on it and it and choose Filter. This will bring up a new window in which you can write a query to filter the features in this layer according to their attribute fields. To do this, you need to know something about what these fields contain, and you can see this by clicking on a field and hitting the Sample button. You can then create a filter by double clicking on a field name, clicking an operator, and then double clicking a value. You will see the query being built as you do this in the box at the bottom. For example “highway”=‘tertiary’ would filter the features to include only those for which the “highway” field has the value (character string in this case) ‘tertiary’.

Another way of filtering, which is often easier, is to start by coloring the features according to some field of interest. Open the Properties window for the layer. In the Symbology tab, select Categorized in the selector at the top. Then choose the field you want to use for this in the Value selector. Hit the Classify button and QGIS will assign a different random color to each different value that appears in this field. Click OK to see how this looks on the map. You can now click the field in the Layers Panel to expand it to show each unique value and the color assigned to it. You can click any of these to remove or return it to the map.

Chapter 6

Mapping Aggregated Data

6.1 Choropleth Maps

6.1.1 Downloading data

For this exercise, we will use 2013 data extracted from the Eurostat database Recorded offences by offence category - police data. You can download it directly from that site, but to make things easier I have already done some data cleaning and you can download the cleaned data here:

[Download Data](#)

6.1.2 Download and Unzip Shapefiles

We will map this data using the Eurostat Countries 2013 shapefiles at the 1:3 million scale. Download them here:

[Download Shapefiles](#)

The file you have downloaded is a zipped archive of shapefiles. Find the directory in which you saved it and double click on the file to unzip it.

6.1.3 Add Country Boundaries to Map

Add the country boundaries using Layer → Add Layer → Add Vector Layer.

In the Data Source Manager window that pops up, leave Source type as it is (the default is File with System encoding), and click the button in the Source section to select the shapefiles you want to add. Navigate to the directory in which you unzipped these files look at what you have. You will see a number of sets of files with identical names. Each set represents a different layer. We will

be using CNTR_RG_03M_2013, but go ahead and add all of them just to see what they look like. To do this, sort the files by type and then hold down the Shift key while clicking on all of the files that have the .shp extension. (QGIS will automatically open all associated files, which have the same names but different extensions).

After selecting the files, click “Open” in the two dialog windows, and the shapes should appear on the map. (Remember that the order of the layers in the layer panel controls the order on the map and that you may need to change this to view certain layers.)

6.1.4 Adding the data

Add the data using Layer → Add Layer → Add Delimited Text Layer...

In the File Name box, select the data file (for this exercise, it is crim_off_cat_rates.csv if you did not change the name when saving it). In the File format box, choose CSV if it is not already chosen. Look in the preview area below and see if the data is appearing in an organized way, with column headings and values. (It should be for this exercise, but if you use a different data set in the future and it is not appearing correctly, the first thing to try is selecting the Custom delimiters option for File format and trying the Semicolon option or something else.) For Geometry definition, select No geometry (attribute only table). This data source does not have coordinates. It simply has data by country name and you will need to link those country names to your shapes in the following step. Click OK.

6.1.5 Join the country polygons with the data.

In the layers panel, right click on the country boundary layer (which will be called CNTR_RG_03M_2013 if you have not changed the names) and select Properties. In the properties window, select the Joins tab.

To join the data to the polygons, there needs to be a field in the data table that contains a set of unique identifiers corresponding to a set of the same unique identifiers in the polygon attribute table. For this exercise, the GEO field in the data table has country codes that correspond to the same country codes in the polygon attribute table’s CNTR_ID field. (When you are using your own data sources, this will often require some initial cleaning. For example, you may need to convert country codes to make sure they match, or you may need to deal with complications like states, like Yugoslavia, that have broken apart or those, like Germany, that have formed through unification.)

Click the plus button at the bottom left. For Join layer, select your data you want to join (for this exercise it is crim_off_cat_rates). For Join field, select the field in the data table that you will be using for this join. In this case, GEO.

For the Target field, select the field in the polygon attribute table that you will be using for the join. In this case, CNTR_ID). Click OK and then Apply.

6.1.6 Color the polygons according to crime data

In the Style panel, choose Graduated in the selector at the very top. Now select for Column the field you want to represent on the map. Since you have joined the data with this polygon layer, you will be able to choose from the fields in the data table as well as in your polygon attribute table. For this exercise, choose crim_off_cat_rates_Theft). Select a Color ramp that you like. Under the Classes tab, select a choice for Mode. You can experiment with different options and see how they appear. Now click the Classify button and then click OK.

6.1.7 Clean the map

Use the layers panel to remove any unnecessary layers that are distracting (for instance, if you included all of the shape files in the Eurostat country data). Use the zoom and pan tools to focus the map on the area of interest to you.

6.2 Proportional Symbols

The choropleth map you just produced represents quantitative data using colors. But it is limited to showing one field from your data. What if you want to show additional information? One option would be to add a symbol on top of each color, making the size proportional to the additional values you want to display. In other words, one field will be displayed by color and another by size.

If you added all of the shapefiles that you downloaded in the last exercise, then you already have everything you need. Put a check next to the CNTR_LB_2013 layer in your layers panel to display a point for each country on top of the choropleth. You will use these points to place your proportional symbols. Right-click on this layer and select Properties. Use the same approach as in the last exercise to join this layer to your data table.

6.2.1 Style the point symbols using graduated sizes

In the Style panel, select Graduated at the top and choose the Column you want to represent here. For this exercise, choose crim_off_cat_rates_TOTAL. For Method now select Size. Choose a Mode and click Classify. You will need to play around with classification modes as well as minimum and maximum sizes to create symbols that are useful for explaining your data. Click OK to display the result on your map.

6.3 Pie Charts

To include even more information on your map, you may want to include some small charts that are then placed on top of geographic features. for this exercise we will use pie charts.

Return to your point layer properties and now select the Diagrams tab. At the top, select Pie chart. For attributes, select all of the individual crime categories that you want to include. Be careful not to include overlapping information here. Use the plus button to move all of these to the Assigned attributes list. You can also size your pie charts proportionally to another variable using the size tab here.

Part II

Spatial Analysis

Chapter 7

Getting Started with QGIS

If you already have QGIS set up for mapping, there is nothing more you need to do to get started with spatial analysis. Most of the tools you will need can be found under the Vector, Raster, and Processing menus.

Chapter 8

Getting Started with Stata

There are a number of ways to use Stata, and your choice will affect the overall workflow of your analysis. These include using the menus, typing commands directly in the console, and using do-files.

8.1 Menus

The most intuitive approach for new users is to use the menus. Stata's comprehensive and well-organized menus are one of the features that makes it attractive, and you can learn a lot about what Stata has to offer simply by exploring the menus options. One drawback to using the menus, though, is that this makes it cumbersome to repeat steps from your work – something that you often need to do in statistical analysis, and that you should always aim to make possible for others to do so that your work is replicable. Stata has a feature that can help with this, which is the option to create a log file that records everything you have done. This is something you should always do, even if working with the menus, as it allows you to go back and figure out exactly what you did, identify errors, etc. But going through a log file and repeating the commands can be tedious if you need to repeat your work multiple times. It is possible to create a “command log,” which contains only the commands you typed and thus can be more easily repeated, but these repetitions will require that you move away from the menu approach in any case.

8.2 Commands

An alternative to the menus is to type commands directly in the command console at the bottom. If you are using Stata in interactive mode (for instance on a server without a GUI), this is often the only possibility, as there will be

no menus. This can save time when you start to learn the commands. It is also good for quick analysis that you do not intend to repeat. And you can still log your work, just as with the menus. Ultimately, though, for most analysis you will want to keep track of your work in a more systematic way, and this is where do-files come in.

8.3 Do-Files

Do-Files are simply text files in which you store the commands you want to execute (also called a script). The key is that you can get Stata to execute these commands directly from the file. This can be done by selecting a line or set of lines and pressing *Ctrl-D*. (You can also press *Ctrl+D* without anything selected in order to execute all lines, or you can press *Ctrl+Shift+D* to execute all lines from wherever your cursor is to the bottom. If you forget this or want to see more options, just look at the *Tools* menu in the do-file editor.) The most efficient way to work with Stata is to use a do-file and to also log your work. This way, you end up with a do-file that you can re-run (and modify) whenever you need to repeat your work, along with an exact log of what lines you ran and Stata's output. While working, you may skip around your do-file and run particular lines, but the goal is to produce a file that can be run from start to finish to accomplish your analysis. You can then give this to someone else and they should be able to use it to replicate what you have done (and check for errors).

There are certain useful pieces of information and commands that you should include at the beginning and end of any do-file. These include information about what the file is intended to do, who wrote it, what version of Stata it was written with, and the commands to start logging to the log file and then (at the end) to stop logging. The Stata version is particularly useful because as new versions of Stata are released, they all are all written to be compatible with older ones and to be able to run older do-files exactly as they were intended. Thus, if you write a do-file using Stata 15 and specify this in your do-file, you can insure that all future versions of Stata will run it using Stata 15 (and thus continue to exactly replicate your work). Here is an example:

```
// Spatial analysis of Chicago Crime data
// John Palmer, 20 February 2020

version 15
clear
capture log close
log using chicago_analysis, text replace

// Here you would start your analysis for example:
display 2+2
```



```
log close
```


Chapter 9

Getting Started with R

R is a programming language and software environment for statistical computing and graphics. It is “free software,” which means you have the freedom to run, copy, distribute, study, change and improve R. It is also free in terms of price, and you can download it from its official home page at <https://r-project.org/>.

One of the powerful thing about R is the enormous ecosystem of packages that people have written for it. An R package is a bundle of code, often along with data, documentation and tests, which runs in R and can be shared. There are thousands of packages for R, written by people all over the world in a wide variety of disciplines, and mostly distributed under free and open source licenses. You can easily install R packages from (CRAN)[<https://r-project.org/>] as well from GitHub and other code repositories.

To install R, go to <https://cran.r-project.org/> and install the latest release that is compatible with your system. From time to time you may wish to upgrade the R on your computer in order to take advantage of new features or to ensure support for new packages. If you already have numerous packages, you may want to copy them into the new version to avoid having to download them again. It is sometimes useful to have multiple versions of R installed side-by-side and this is possible to do.

9.1 Using R from a Terminal

The Read-Eval-Print Loop is the most basic way of interacting with R directly. You can get to this by simply typing R from a terminal window. You can also run R files (non-interactively) from a terminal.

9.2 Text Editor

You can use any text editor to create, store, and edit R scripts (like do-files in Stata). It is nice to have features like code highlighting and a built-in REPL. Choices include Emacs, vim, and sublime.

9.3 IDE

An Integrated Development Environment combines a text editor with a set of other useful tools for debugging, automation, and other tasks. The open source version of R Studio is the obvious choice here. You can download it from <https://rstudio.com/products/rstudio/>. Using R with R Studio will be somewhat closer to your experience with Stata's GUI, but with some important differences. On one hand, you can do less with menus. You cannot, for example, use the menus to carry out statistical analysis. On the other hand, R Studio has many features that you will not find in Stata's GUI and these actually make your life much easier when you learn them. For example, when you start typing code into an R script, R Studio will offer you autocompletion when it recognizes words you are typing. You can also start typing a word and press the tab key to trigger this. And while you cannot look for the commands you need in your statistical analysis in the menus, you can easily look at the help files and other documentation related to them by typing `help([command])`. For example, `help(var)` brings up the help file for the `var` function and other related functions for computing variance, covariance, and correlation.

9.4 Basic Concepts

This concept of a function is an important one that you need to get used to for working with R (and most other programming languages). You already use functions in QGIS and Stata but you may not think of them as such. Functions are algorithms that do things. Think of them as verbs. You often provide a function with arguments, which it then uses. The other concept you need to understand is that of objects. As the name suggests, objects are things that have characteristics. Think of them as nouns. You will often pass an object to a function as one of the function's arguments and the function will then transform the object in some way or use it to compute a value, which it then returns.

Another important concept in R is that of the package or library. You have already encountered this concept with QGIS in the form of plugins and in Stata in the form of ado files. Much of the power of R comes from all of the packages that have been written for it, as already discussed above. You can find and keep track of packages in R Studio using the Packages tab at the bottom right.

Chapter 10

Buffers and Distances

To learn how to use buffers and distances for analyzing data in QGIS, we will work through an example similar to the one we talked about in class. For this exercise, start by going to Chicago’s open data portal and downloading the following datasets:

10.1 Getting the Data

First, go to the crimes database at:

<https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2/data>

Use the filter tab to select only crimes that were reported after 1 January 2020 (Date is after...) and that have the Primary Type listed as “BATTERY”. Export this data as a csv file and save it on your computer.

Now, go to the public schools database at:

<https://data.cityofchicago.org/Education/Chicago-Public-Schools-School-Locations-SY1920-Map/xtgb-6hqg>

Export the public school location data as a csv file and save it on your computer.

Finally, go to the bars database at:

<https://data.cityofchicago.org/Community-Economic-Development/bars/x7bm-cj66>

Filter this data to include only bar licenses that expire after 1 January 2020, export it and save it on your computers.

10.2 Mapping the Data

Now put all of this data on a map in QGIS using the techniques you have already learned. Each dataset is a csv file, so you will need to add it using Add Delimited Text Layer and specifying the coordinates. In each case, the coordinates are given in latitude and longitude.

10.3 Drawing Buffers Around the Schools

To draw buffers around the schools, first make sure you have the MMQGIS plugin installed and install it if you don't. Go to the MMQGIS menu at the top of your screen and choose Create → Create Buffers. In the dialog that opens up set the radius that you want to use for your buffers and the units that this will be in. Think about a good choice based on the research questions we discussed yesterday. Then select an output shapefile (button of dialog) location where you have write access and will remember where it is. Click OK to calculate the buffers.

10.4 Counting Points in Polygons

To count the number of batteries and the number of bars in each school's buffer, select Vector → Analysis Tools → Count Points in Polygon. In the dialog that opens, select your new buffer layer as the polygons and select one of the points layers that you want to count. Give the count field a name that will make sense to you later (like "N_batteries" or "N_bars"). Run the operation and then do this again for the other point layer you want to count. (Note that each time you run this, it will create a polygon layer with the counts. Is you use that layer as your polygons for the next count, then you will end up with both counts in the same attribute file, which will save you time. Alternatively, you can do each separately and then join them.)

10.5 Calculating Distances

To calculate distances from each school to the nearest bar, go to the Vector → Analysis Tools → Distance Matrix. Choose as the input layer your school points as as your target layer the bar points. Make sure you have a unique field selected for each. For Use only the nearest (k) target points type 1. Run the operation and wait for it to finish.

One complication to note: Distances are calculated in units that depend on the CRS of the two layers. In this case, both layers simply use latitude and longitude, so your distances will end up being in degrees of latitude and longitude. You will

be better off with meters, which you can achieve by first reprojecting your layers to UTM before carrying out the operation. For purposes of this lab, however, leave it as it is.

10.6 Joining and Exporting Data

If you want to end up with a file that you can analyze later in Stata, R, or some other statistical tool, you will want to export your data from QGIS. But you first need to join your distance table with the polygon layer in which you have your counts. Do this using the Join tab in the properties dialog for the polygon layer. Select the unique ID for each layer and carry out the join. If you do it correctly, you should see the distances in the attribute table for the polygons. If that is the case, then export your polygon layer as a csv file and save this for further analysis in our next lab. (Also turn it in as your lab assignment for this week.)

Chapter 11

Cluster Analysis

K-means clustering is a simple approach for partitioning observations into a set of k clusters. The method within-cluster variances based on the following equation:

$$\arg \min \sum_{i=1}^k \sum_{x_j \in s_i} (x_j - \mu_i)^2$$

11.1 Using QGIS

To perform k-means clustering using QGIS, first make sure you have the point data you wish to cluster loaded as a vector layer. Then open the Processing → Toolbox and double-click on Vector analysis → K-means clustering (this is in side bar on the right of the QGIS window).

When you do this, a menu will come up allowing you to select the points layer you wish to perform the clustering on and to specify the number of clusters (k), the field and layer name that will be used to store the cluster IDs in the attribute table, and other options.

Once you choose the options you want and click Run, QGIS will perform the clustering operation. The results will be stored in a new points layer with the name you selected in the previous step. Initially you will not see much difference, just a change in color of the points as this new layer is overlayed on top of the previous one. However, you will see that this new layer has a cluster ID field in the attributes table, and you can use this to color the points by selecting the Categorized option in Properties → Symbolization and choosing your cluster ID variable as the Value for the categorization. (Remember to hit the Classify button.)

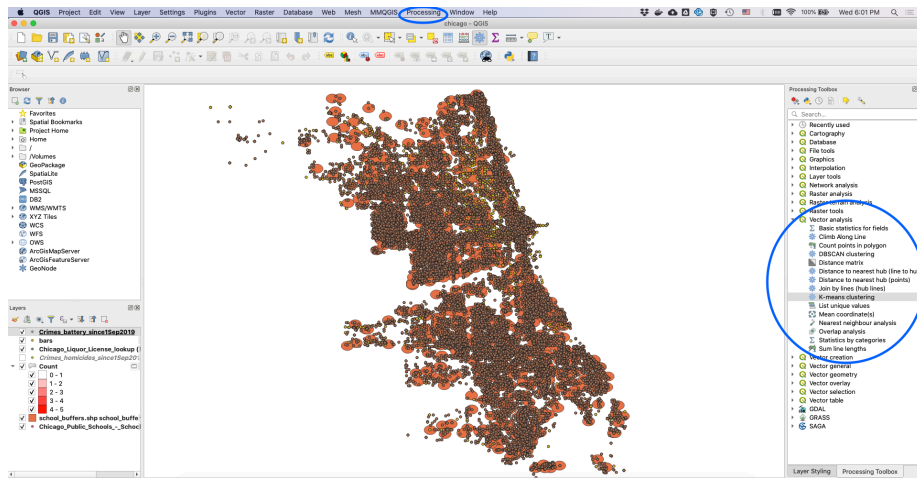


Figure 11.1: Screenshot of the Processing Toolbox in QGIS 3.10 for MacOS

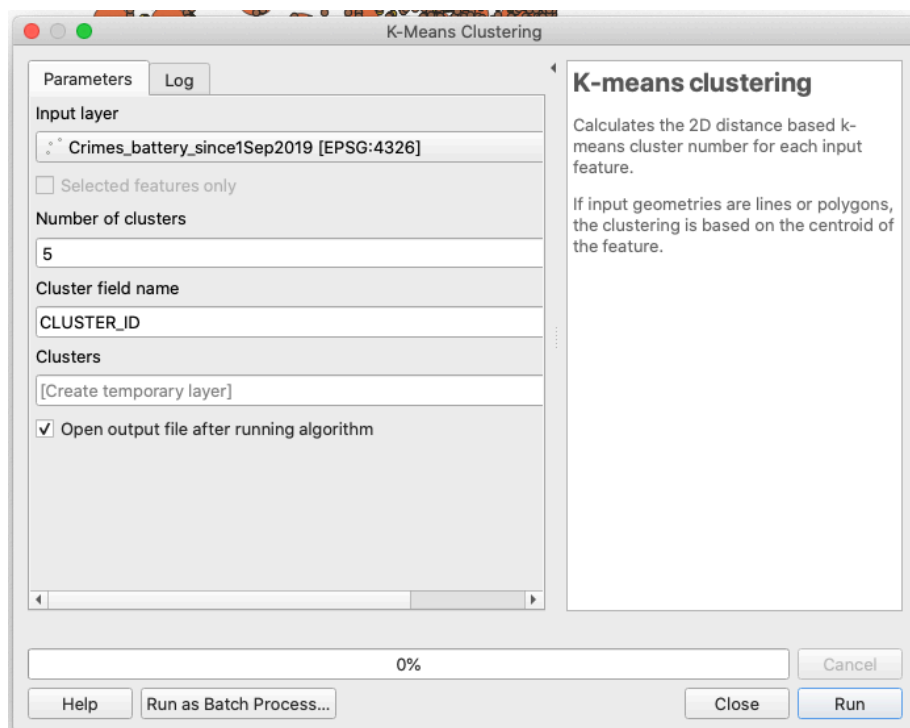


Figure 11.2: Screenshot of the k-means clustering menu

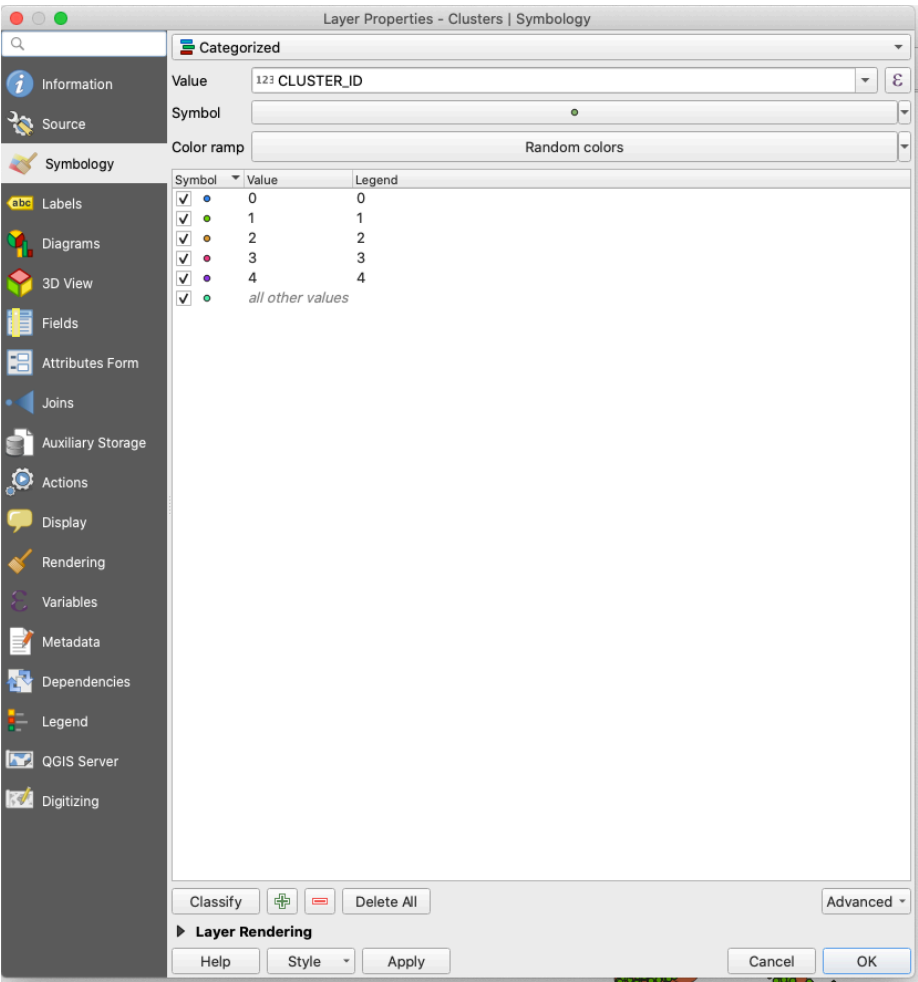


Figure 11.3: Screenshot of the properties menu for setting up color coding of clustered data

The final result should look something like this:

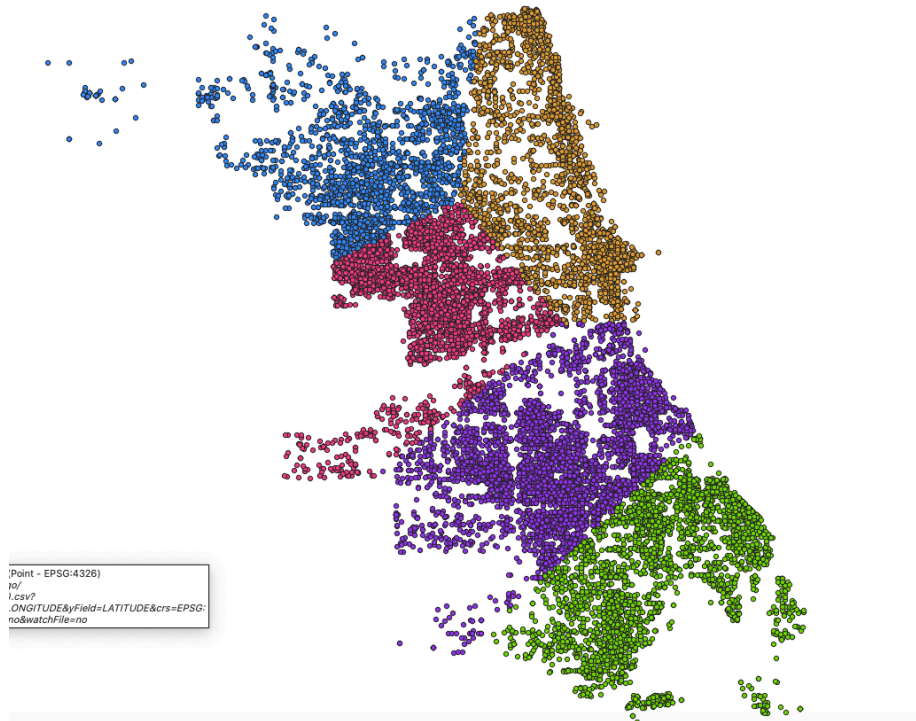


Figure 11.4: Screenshot of clustered data coded by color

11.2 Using Stata

To perform kmeans clustering in Stata, you can use the command *cluster kmeans*, followed by the list of variables you want to cluster on. For spatial clustering, use the x and y coordinates. The command lets you specify the number of clusters as an option, *k(#)*.

```
cluster kmeans xcoordinate ycoordinate, k(8)
```

11.3 Using R

This example is designed to give you a taste of R by walking you through what you would need to do for a very basic clustering analysis.

```
# first clear everything from the environment
rm(list=ls())
```

```
# now load necessary libraries
library(tidyverse)

# Now load some data. In this case, we will grab a subsample of the NYPD crime data from a URL:
D = read_csv("http://johnpalmer.github.io/CrimeMapping2020-Slides/data/NYPD_Complaint_Map__Year_t

# to begin with, take a look at the column names in this dataset:
names(D)

# you can also quickly look at the first few rows:
head(D)

# or open the whole thing as if it were a spreadsheet:
view(D)

# for clustering, let's first prepare the data so that we only have X and Y coordinates and remove any NA
D = D %>% select(X="X_COORD_CD", Y="Y_COORD_CD") %>% filter(!is.na(Y), !is.na(X))

# now the clustering function, which takes as arguments, the data and the number of clusters
clusters = kmeans(D,9)

# how many clusters should we use?
K=20
ss = rep(NA, K)
for (i in 1:K) {
  clusters=kmeans(D, centers=i)
  ss[i] = sum(clusters$withinss)
}

plot(1:K, ss, type="b")

# let's try it with 15 clusters and see how it looks in a plot:
clusters = kmeans(D,15)
plot(D, col=clusters$cluster, pch=20)
```


Chapter 12

Autocorrelation and Spatial Regression

This Section walks you through the analysis of spatial autocorrelation using the global Moran's I, and the fitting of spatial regression models with lagged independent variables, lagged dependent variables, and/or lagged errors. It uses as an example the dataset that you created in Section 10, which you should have saved as a csv file. (If you cannot find this, just ask and I will give you a copy to work with.)

12.1 Loading the Data into Stata

The first step is to load the data into Stata. If this is the first time you are working with a csv file in Stata or you have forgotten the command or simply don't want to type the file path, an easy way to load the file is through the menus: File → Import → Text Data.

You can then navigate your file system to select the file and import it. Of course, the more efficient way to do this is with the command and, better yet, to put this into a do-file so that you can re-run everything again any time. Notice that if you import the file with the menu, you will then see the command (including the path you need) in the Stata console, so you could simply copy this into your do-file. The command should look something like this (modifying the actual path for your system):

```
import delimited Z:\chicago\school_buffer_bar_battery_counts.csv
```

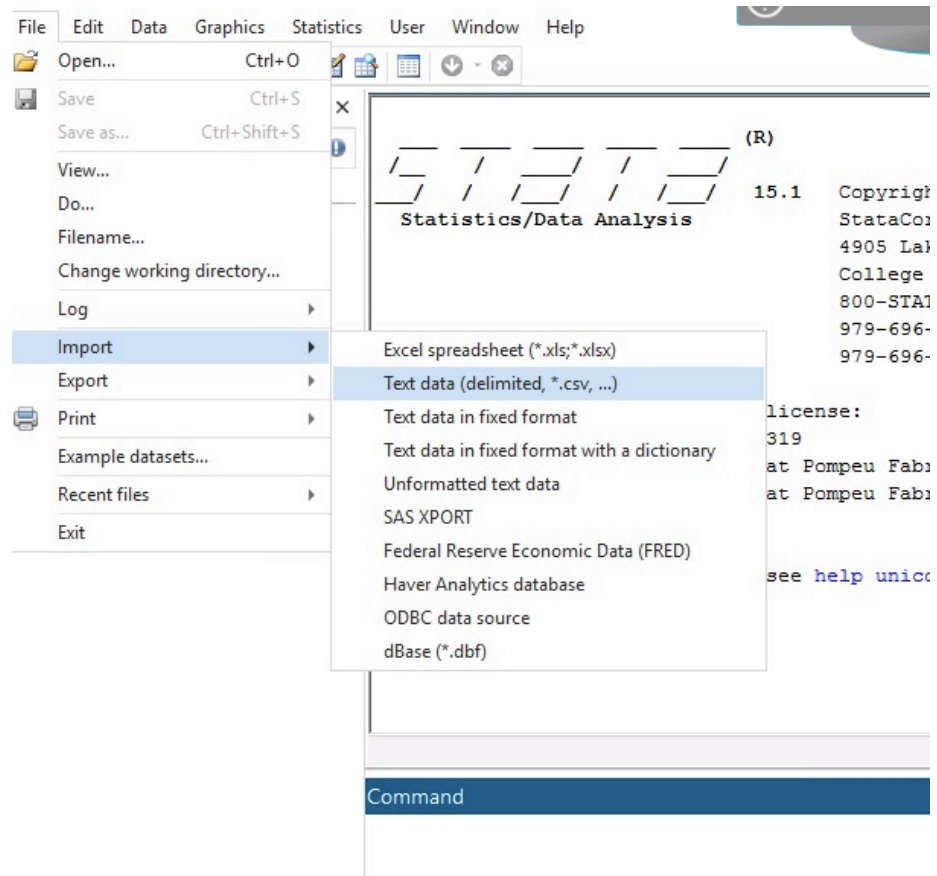


Figure 12.1: Screenshot of the Stata menu for importing csv files

12.2 Creating Spatial Weights Matrix

Once you have the data imported, you need to tell Stata that this is spatial data and create a spatial weights matrix. Again, you can do this with the menus: Statistics → Spatial Autoregressive Models.

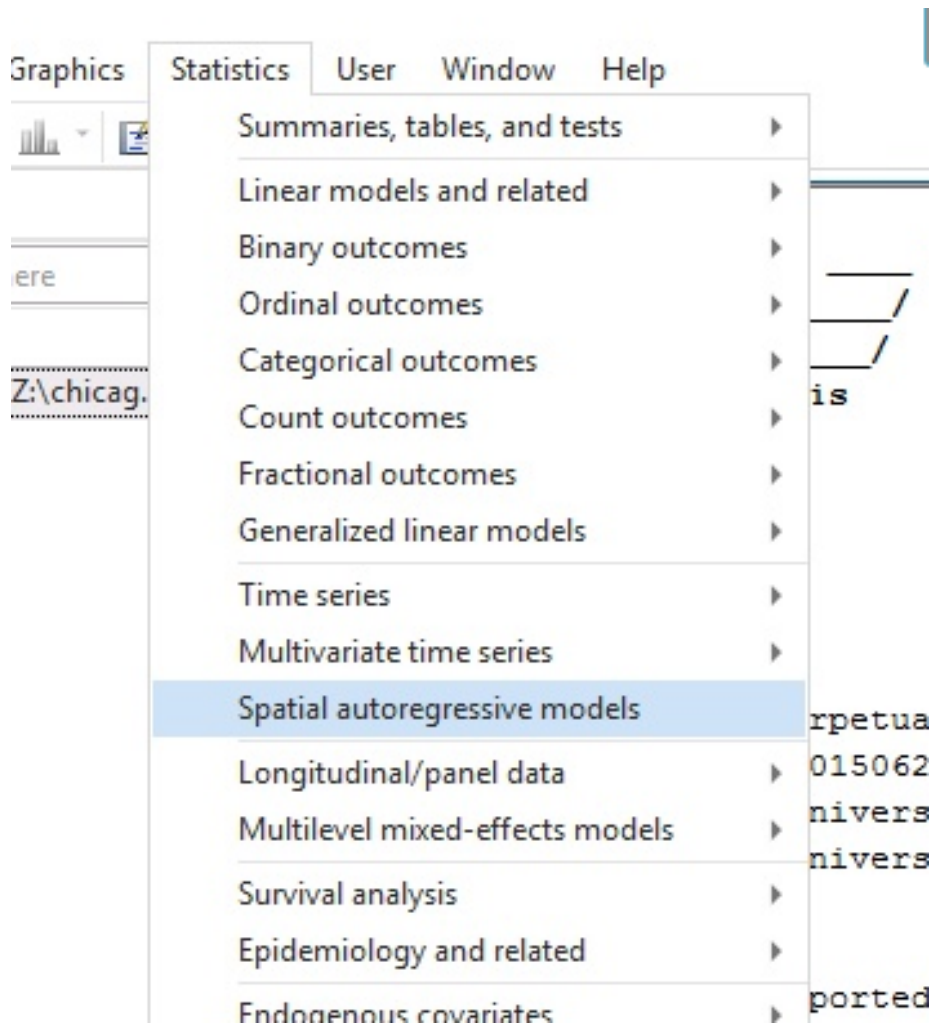


Figure 12.2: Screenshot of Stata's spatial regression menu

From the menu that pops up, you will see that you have the choice of using an external shapefile or simply using the information in the file that is currently in memory. Choose this second option, select a variable with unique values for the ID option (in this case, use `school_id`), specify the x and y coordinates (in this case, simply x and y), indicate whether these are longitudes and latitudes (in this

case they are), and choose whether you want distances reported in kilometers or miles. Then click Submit.

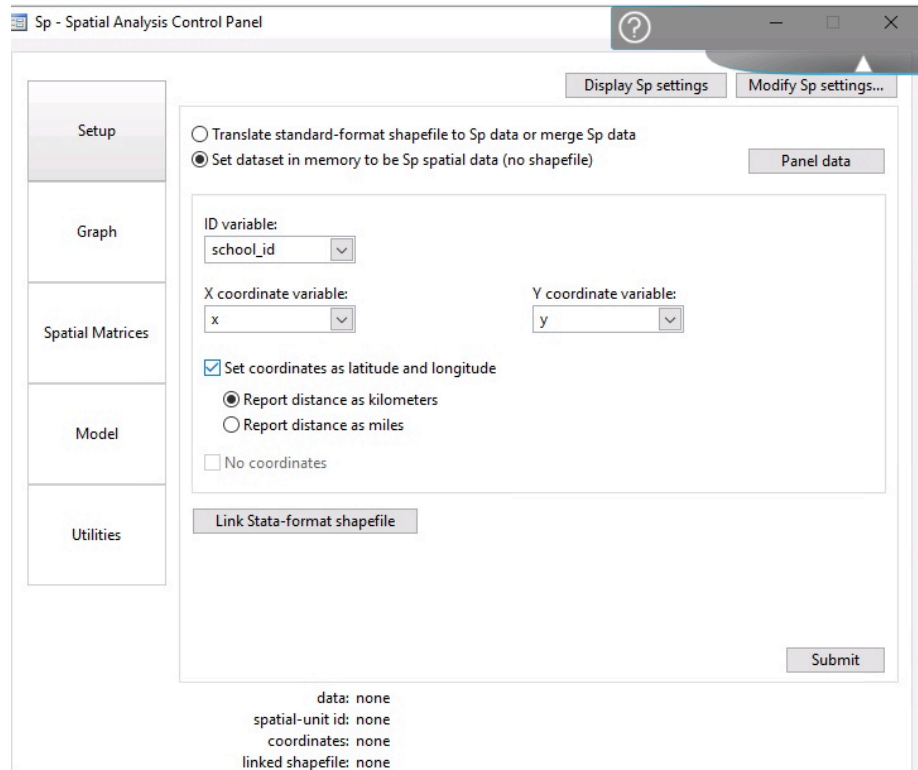


Figure 12.3: Screenshot of Stata's spatial regression menu

Now, without closing the menu, move to the Spatial Matrices tab. Here is where you can set up the weights matrix that you will be using. In this case you want to select Create standard matrix. Consider whether it makes more sense, given your data, to create a contiguity matrix or an inverse-distance matrix and choose the option in the menu accordingly. (As you will see, only one option actually works with this dataset. Think about why that is.) Give your matrix a name (which you will use later) and, if you have chosen an inverse-distance matrix, then decide if you want to set a threshold distance beyond which the weights will drop to 0. (You can leave the normalization method set at the default.)

Now hit Submit and Stata will create your matrix. Look in the console and you will see the command that you need in order to do this directly from a do-file next time:

```
spset school_id, coord(x y) coordsys(latlong)
```

```
spmatrix create idistance W, vtruncate(1) replace
```

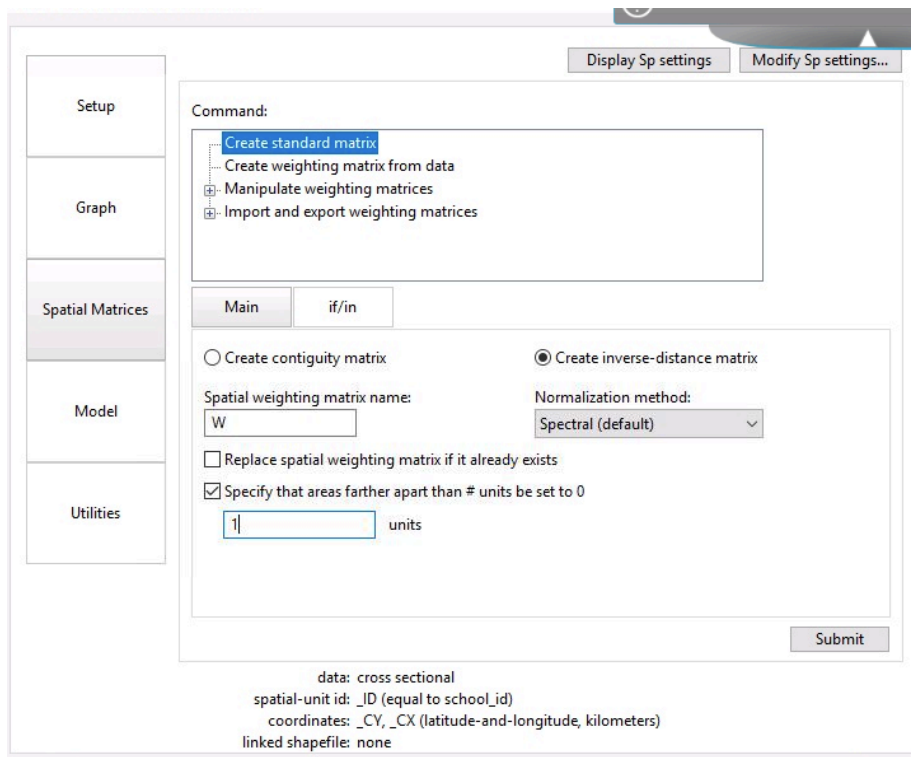


Figure 12.4: Screenshot of Stata's spatial regression menu: Spatial Weights Tab

12.3 Regression and Spatial Regression

Now that you have the spatial weights matrix created, you can start fitting spatial autoregressive models and also testing for autocorrelation. For this, we will move directly to the commands (rather than going through the menus) to make things more efficient (but feel free to explore the menus to see all of the possibilities).

Start with a non-spatial ordinary least squares regression of the number of batteries within each school's buffer on the number of bars in that buffer:

```
reg n_batteries nBars
```

Interpret the results and then test for spatial autocorrelation using the global Moran's I (specifying the name you chose for your weights matrix):

```
estat moran, errorlag(W)
```

Now fit the spatial regression models we discussed in class, using the following commands:

For lagging the dependent variable:

```
spregress n_batteries nBars, gs2sls dvarlag(W)
```

For lagging an independent variable:

```
spregress n_batteries nBars, gs2sls ivarlag(W: nBars)
```

For lagging the error term:

```
spregress n_batteries nBars, gs2sls errorlag(W)
```

Fit several combinations of these models and interpret the results. As we discussed in class, one useful way of interpreting the effects in these models is by using the following postestimation command:

```
estat impact
```