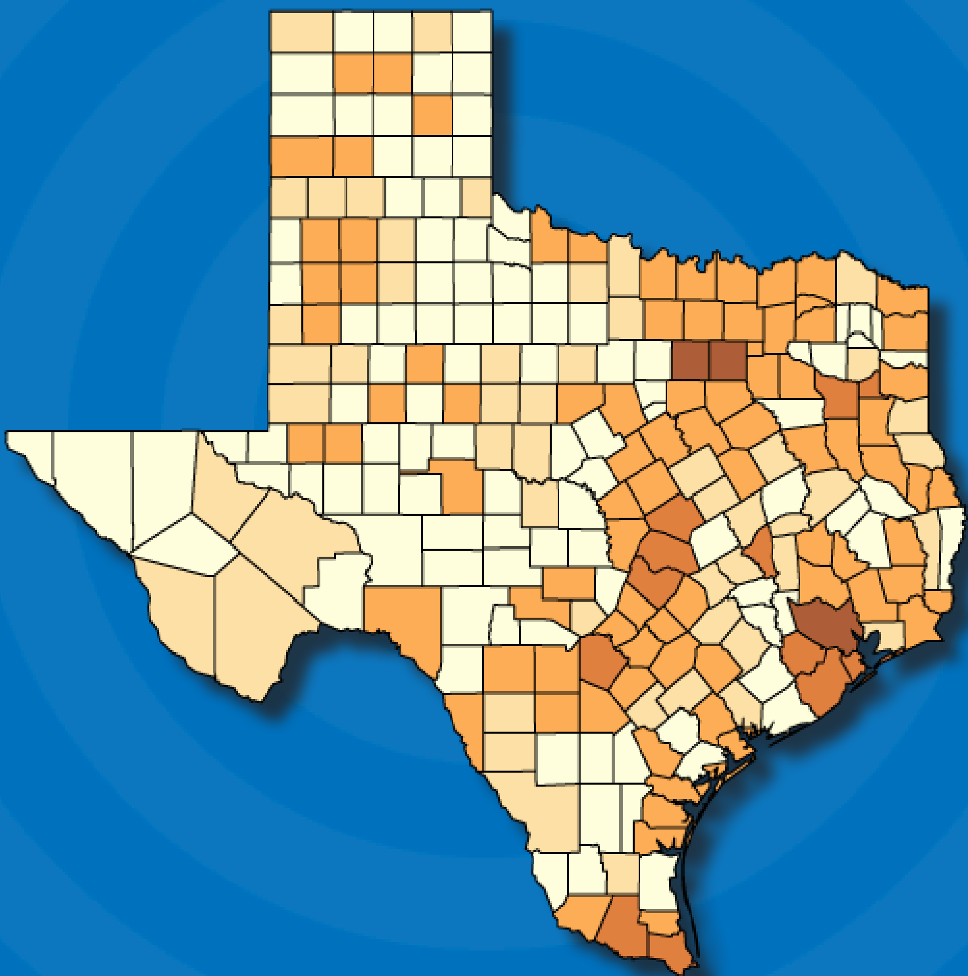
The Beginner's Guide to GIS

An introduction to the power of mapping.



The Beginner's Guide to GIS

Published by MangoMap Limited

Bristol, United Kingdom

www.mangomap.com

Contents

The Beginner's Guide to GIS 6

What is GIS? 8

What does a GIS do? 10
Do I need a GIS? 11
How Much Does GIS Cost? 12
Can I do GIS for myself? 13
How long does GIS take to learn? 14
How much of this book do I *really* need to read? 15

GIS Mapping 16

Map Types 18
Stacking Layers 24
Map Interactivity 25
Map Tools 26
The Anatomy of a GIS Map 29

GIS Data 30

Introducing the Shapefile 32
Other Common GIS Data Formats 34
Sourcing GIS Data 36
Preparing GIS Data 38

GIS Software 42

Types of GIS Software 44

Getting Started with Desktop GIS 50

Desktop GIS: The Basics 51
Popular Desktop GIS Software 60

Web GIS 62

The Scalpel vs the Swiss Army Knife 65
Story Maps 66
Will Web GIS Replace Desktop GIS? 67

Tutorials 68

Create a Web Map from a Spreadsheet Containing Regions 70
Create a Web Map from Addresses in a Spreadsheet 84
How To Remove Unwanted Regions From Your Map Data 100
How to Turn a List of Locations into an Online Heat Map 108

The Beginner’s Guide to GIS

New to GIS? Don’t know where to start? This beginners guide to GIS provides all the information that you need to understand what GIS is, and how it can be a great asset for you, and your organization.

As you have purchased this book, I'm guessing that the term GIS or web GIS has popped up at work. Maybe your organization has decided to start using GIS to make better sense of their data, or you were looking into ways to map your data, and that search led you to the term "GIS".

If like me, the first thing you did when you heard this term was head over to Wikipedia, you will have been confronted with lots of complex jargon and even some scary looking algorithms:

$$A = 270^\circ + \arctan\left(\frac{\left(\frac{\partial z}{\partial x}\right)}{\left(\frac{\partial z}{\partial y}\right)}\right) - 90^\circ\left(\frac{\left(\frac{\partial z}{\partial y}\right)}{\left|\frac{\partial z}{\partial y}\right|}\right)$$

Formula for calculating aspect in GIS
source: Wikipedia

Now, before you get disheartened or fall asleep, I have some good news for you!

GIS isn't that complicated, what's more it can even be fun. Once you've finished this short book not only will you understand what GIS is all about, but you will have the background knowledge and vocabulary required to discuss GIS and its associated tools with confidence.

You will even learn the basics of operating a GIS, preparing map data and sharing interactive maps in a web GIS.

What is GIS?

`GIS = Geographic Information Systems`

GIS is a very broad term, and trying to get a consistent definition can be tricky. Ask ten different GIS users and you will likely get ten different answers.

It could be argued—and the propeller heads often do—that *any* digital data that contains location based information is, in fact, a GIS.

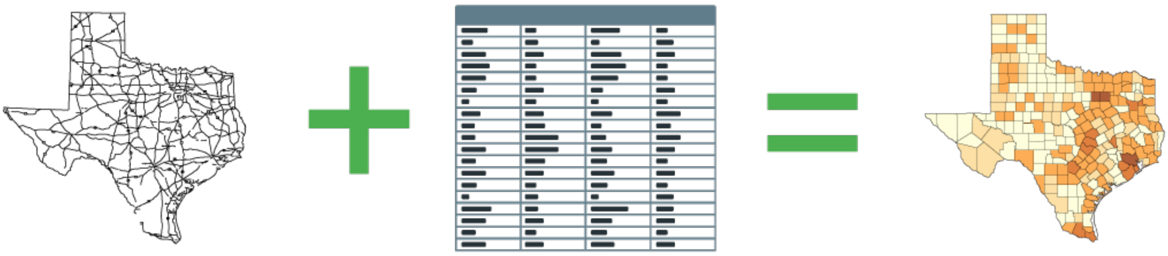
This location information in the GIS industry is called “spatial data” and it could be an address, coordinates containing latitude and longitude, or complex three dimensional geometry.

But you’re not here to get into semantics, or bogged down in jargon. I’m guessing you are here because you want to know what the majority of people mean when they say “GIS”, and more importantly:

How it can be of value to you and your organization.

WHAT DOES A GIS DO?

The truth is that it can do a lot of things, but here's the short answer: A GIS allows you to visualize your data as a map.



Maps + Data = GIS

We're visual creatures that possess an innate ability to visualize patterns. Patterns that might take us hours to identify in a spreadsheet can often be identified in an instant when displayed in a more visually engaging format like a graph, chart, or in this case a map.

There are many innovative ways that your data can be displayed on a map. It could be plotting markers, color coding locations based on an a data value or using heat maps to identify clusters and patterns in your data, the possibilities and potential insights are literally endless.

Data can be visualized in endless ways

What's more, GIS systems aren't static. They allow us to ask complex questions—or "queries" as they are called in GIS speak—anytime we like. A GIS system can answer these questions instantly by modifying colors, shapes or highlighting locations on the map.

DO I NEED A GIS?

The truth is a that a GIS is merely a tool, it's just a means to an end. Most people don't need a GIS, what they really need is the answer to a question. A GIS simply provides us with a more efficient mechanism of obtaining that answer.

So the question to ask yourself is this; will the benefits of being able to visualize my data on a map and quickly answer questions that are location specific, provide me with a return on the time taken to get up to speed on GIS?

I'll give you a clue:

If data is a core part of your business, then the answer starts with "y", ends in "s" and contains an "e" in the middle.

By the time you've finished this book — it's pretty short, so shouldn't take you long — I'm confident that you will be excited about the possibilities that GIS holds for your business.

Location is a powerful element of your data that, until now, you have likely been neglecting. You obviously know the what and when, and you probably know the how. But do you really understand the where?

I'll guarantee there are patterns in your data that have been right under your nose this whole time, but you just haven't been able to see them. A GIS system will shine a bright light on those patterns and make them clear for all to see.

HOW MUCH DOES GIS COST?

So I can hear you thinking:

> ***"That all sounds great, but be straight with me—am I going to need to sell a kidney?"***

I can't say that's never happened, but you'll be pleased to know that most of us that benefit from GIS still possess two functioning kidneys!

There are some great free and open source tools available, lots of affordable desktop and web tools (Mango, nudge nudge, wink wink) and also some very expensive solutions at the top that service very niche use-cases.

Later in the book we will be taking a more in-depth look at the various open source and paid GIS software solutions, both on the desktop and the web.

In most cases the main cost of GIS will be your investment of time, rather than software licensing.

Which brings us neatly onto the next question.

CAN I DO GIS FOR MYSELF?

If you listen to the consultants, then a long winded buzzword-filled answer will eventually arrive at a negative.

But isn't that always the case with consultants?

The truth is that you can do it yourself using freely available tools and some self-study.

And I know, because that's the way I learned GIS.

I'm a software engineer by trade, and managed to make it well into my twenties completely oblivious to the mysterious world of GIS.

Then one day I'm in a meeting with a client. They mention to me that they would like move their GIS system online and enquired whether or not I could do it.

Like every young and overly keen consultant, I nodded away as if knew what he was talking about and assured him that I could deliver a superb online IGS, hang-on, or was it GIS?

After the meeting I combed the internet and began to realise what I had got myself into. This GIS stuff looked pretty tricky to say the least—jargon galore, and very complicated looking tools.

Luckily, I persevered and managed to learn just enough to make it to the next meeting and maintain a very thin veneer of understanding and confidence.

That was ten years ago. Since then I've been a partner in a GIS consultancy company and more recently founded Mango, a web tool that makes it easy for non-coders to create and share GIS web applications.

I'm effectively a GIS immigrant, I didn't learn it in college — I stumbled across it completely by mistake. I think this gives me a unique perspective, and I'm confident that you too can quickly learn enough GIS to provide tangible value to your work.

And I can assure you of this: GIS is a great skill to have under your belt. It always looks good on a CV because it's something that most companies know they should be doing more of, but haven't had the personnel to get the ball rolling.

HOW LONG DOES GIS TAKE TO LEARN?

Photo by Jamie Street on Unsplash

Now I'm not going to pretend that GIS is simple, it can in some instances be extremely complex, but like most things it follows the 80/20 rule: you only really need to learn 20% of what's possible in order to enjoy 80% of the benefits.

If you tried to learn every aspect of GIS it could take a lifetime, a far better approach is to get a good idea of what's possible and then zero in on the parts that will be of benefit to *you and your business.*

That's where this book comes in, this book isn't going to teach you to become a GIS expert. My objective is to give you an introduction to what's possible and get you acquainted with some common use cases, tools and terminology.

This will be the foundation you need to zero in on the parts that will benefit you the most.

For example, a common GIS use case is business analysis. In this case you will be most interested in how to transform tabular data—like spreadsheets—into a map-ready format, and then build visualizations that highlight sales trends or sales territories.

What you don't need to understand is how to process LIDAR data, or orthorectify an aerial photograph. Don't worry if you don't understand what those things mean, truth be told, I don't really understand them either.

You see, as with all technical professions there are some professionals that like to make it all sound much more complicated than it really is. But don't let them put you off, most of these things that sound so complex can be learned in an afternoon.

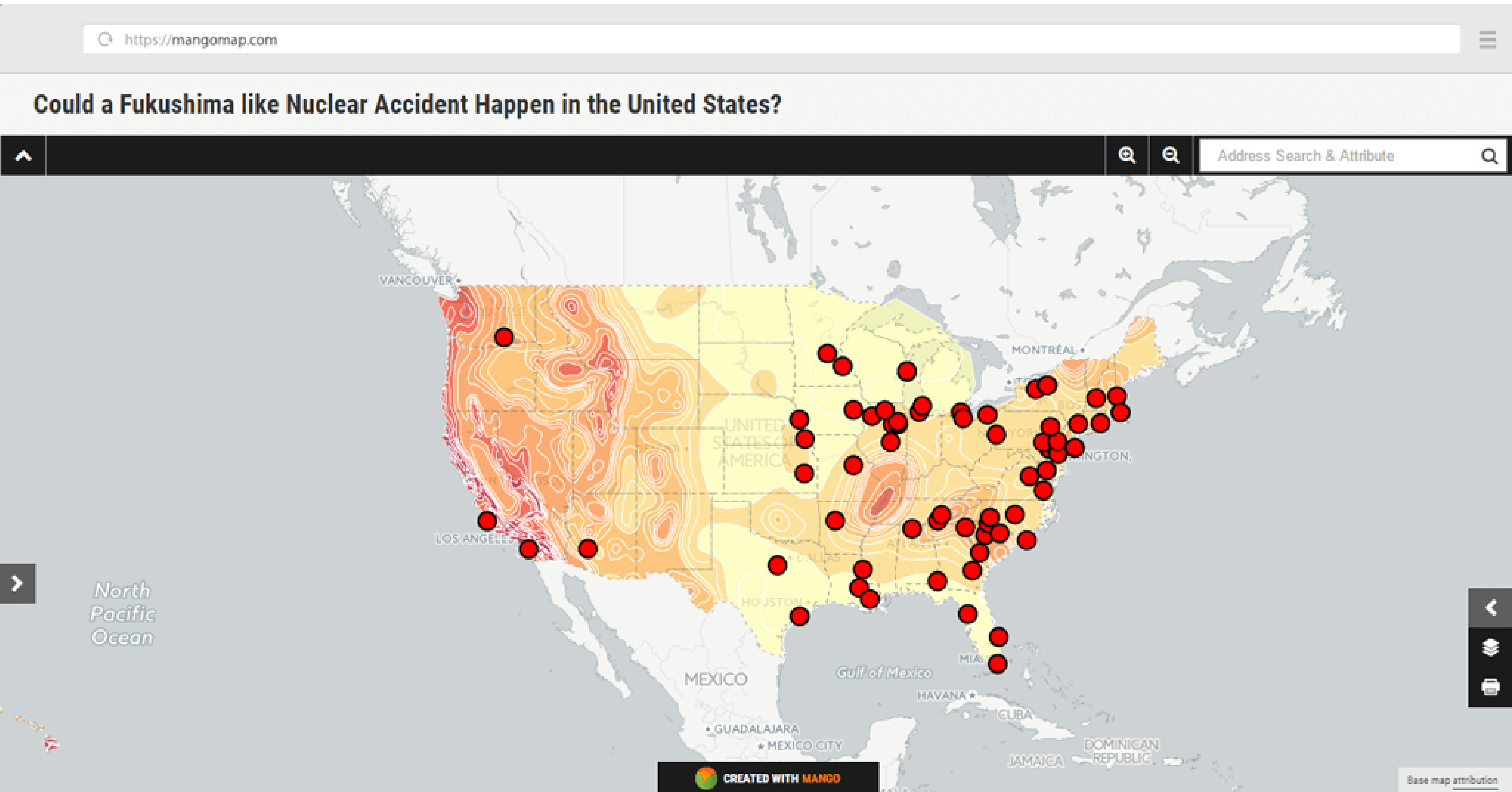
HOW MUCH OF THIS BOOK DO I *REALLY* NEED TO READ?

If you are serious about understanding the benefits of GIS, I recommend reading this guide front-to-back.

I've tried to make it as concise as possible and easy to understand, and once you're done, your understanding of GIS and its possibilities will have a solid foundation.

GIS Mapping

A GIS has one purpose:
To allow you to quickly and easily answer questions about your data that relate to location.



The most common use of a GIS is is to produce data visualizations in the form of a map. The idea being that each visualization is communicating something that might not be obvious from the raw data alone.

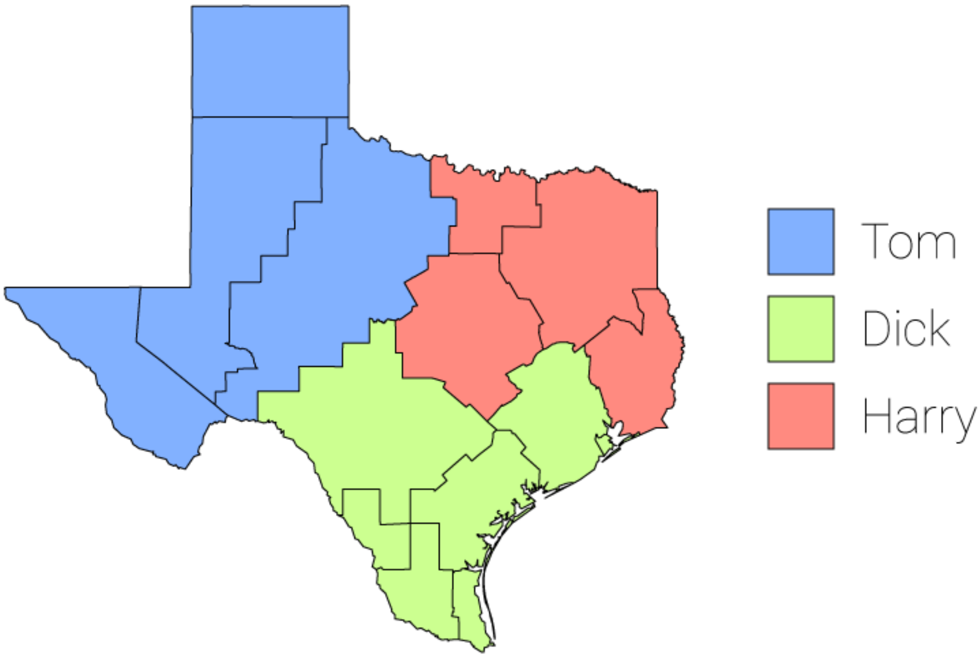
Let’s go ahead and look at some common examples.

MAP TYPES

THE CATEGORY MAP

The most common type of GIS map, and in many way the most simple is the category map.

A category map allows us to visualise which category each location belongs to. The example below shows sales territories, categorized by dealer. By looking at the legend we can quickly see which sales territories belong to which dealer.



A Category map showing sales territories for Tom, Dick, and Harry

LET'S TAKE A LOOK UNDER THE HOOD

A common misconception for those new to GIS, is that maps — like the category map above — are created in the same way that a graphic designer might create an illustration; i.e., starting with the outlines and then clicking on each area to change the color based on which dealer it belongs to.

To an extent that is what's happening, except the GIS is doing the work rather than you. You see a GIS is data driven, and we use rules to dictate the color or appearance of each "feature" on the map.

JARGON ALERT: Feature

In GIS each location/thing on the map is generically referred to as a “feature”.

So whether we are mapping States, Walmart outlets, or rivers; each individual State, Walmart outlet or river can be referred to as a “feature”.

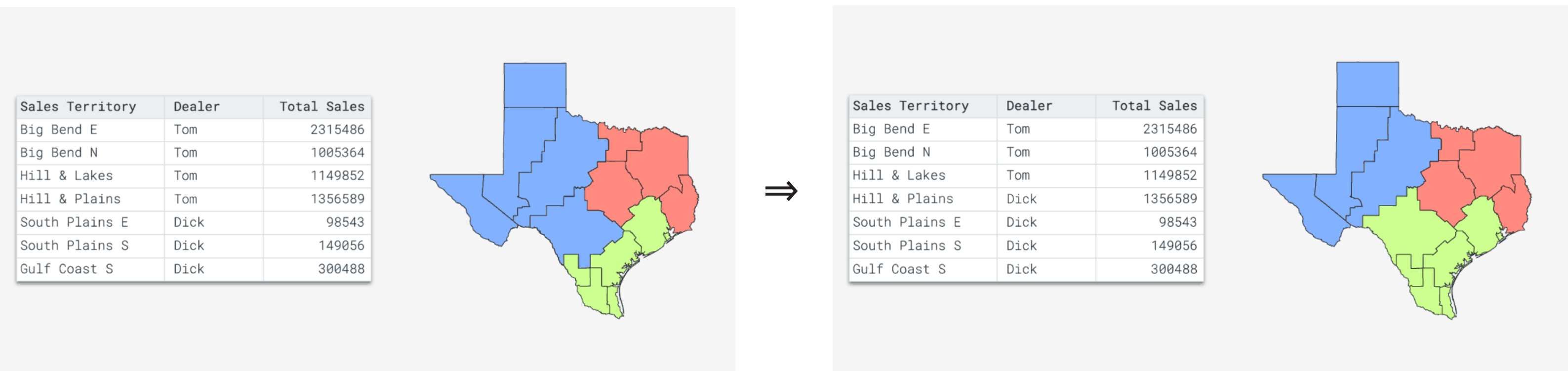
The rules — or “styles” as we call them in GIS speak — are based on data. For example our data looks like this:

Sales Territory	Dealer	Total Sales
Big Bend E	Tom	2315486
Big Bend N	Tom	1005364
Hill & Lakes	Tom	1149852
Hill & Plains	Dick	1356589
South Plains E	Dick	98543
South Plains S	Dick	149056
Gulf Coast S	Dick	300488

Territory sales data

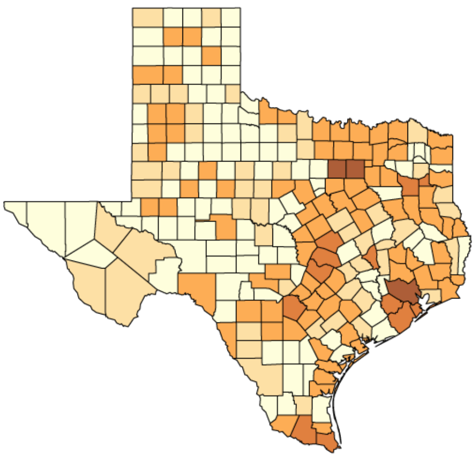
Each row in the table represents a feature on the map. So we can tell our GIS that we want all sales territories that belong to Tom to be colored red and our GIS will dutifully oblige.

In future if Tom gains more sales territories, we don’t need to touch the map. We just change the dealer value of a territory to Tom’s name and the map will magically update itself.



Update your data, and the map updates automatically!

THE QUANTITY MAP



Each feature in the map below is color coded based on sales. The darker color indicates more sales and lighter colors represent fewer sales.

This type of map is generally called a quantity map and if you are feeling particularly nerdy you can also call it a choropleth map. Sometimes this type of map will be referred to as a heat map, but this is incorrect terminology and shortly we will be looking at what a heat map actually is in GIS.

This type of visualization is useful for identifying patterns and trends based on location and is most often used for business analysis and demographics mapping.

Just like the category map above, this map is data driven and uses rules based on a number "attribute" in order to style each feature.

JARGON ALERT: Attribute

Let's continue to use the spreadsheet paradigm. Whilst features are the equivalent of rows, attributes are the equivalent of columns.

So for example the "dealer" attribute for a sales territory could be Tom, Dick, or Harry.

Creating the rules for quantities is a little more complex than it is for categories. Our legend will show a color ramp with each color on the ramp representing a number range.

Vertical quantity legend showing class breaks

Each color on this color ramp is commonly referred to as a class break. To define our class breaks we will need to tell the GIS two things; the first is how many class breaks we want and the second is is the number range for each class break.

NERD CORNER: Number Classification

When building a quantity legend there are lots of ways to decide the number range of each class break. Let's take a look at the options that are available in most GIS systems:

MANUAL CLASS BREAKS

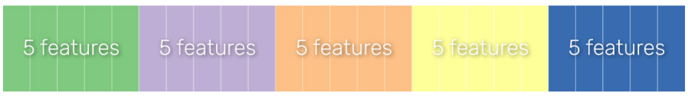
This is the easiest, you just manually key in the number range for each class break. This type of class break is useful when there are pre-existing ranges that are used within your organization for classification purposes.

EQUAL INTERVAL CLASS BREAKS

The range is divided by the number of class breaks.

Equal interval class breaks don't work well when the data contains extreme outliers, or heavily clustered data distribution. For example a dataset with 99 features with a value between 1 and 5 and a single feature with a value of 50, would produce a map that is almost uniformly a single color.

QUANTILES



Quantiles will adjust each class break to ensure that each class break contains exactly the same number of features.

This type of class ramp produces the most visually appealing map, as each color in the ramp is used the same number of times on the map, but it can also be confusing for end users that don't understand how the data is being presented.

JENKS

The Jenks optimization method, also called the Jenks natural breaks classification method, is a data clustering method designed to determine the best arrangement of values into different classes.

This is done by seeking to minimize each class's average deviation from the class mean, while maximizing each class's deviation from the means of the other groups.

In plain talk, it strikes a good balance between being pretty and sensibly distributed. If your data doesn't fit nicely into equal intervals, then this or manual breaks is your best bet.

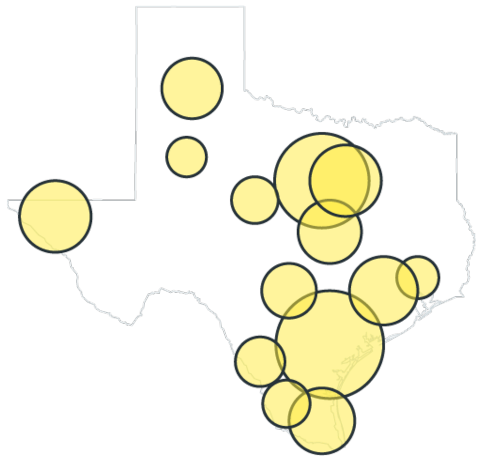
HEADS & TAILS

Now we're getting into hardcore mathematics.

Head/tail breaks is a clustering algorithm scheme for data with a heavy-tailed distribution such as power laws and lognormal distributions.

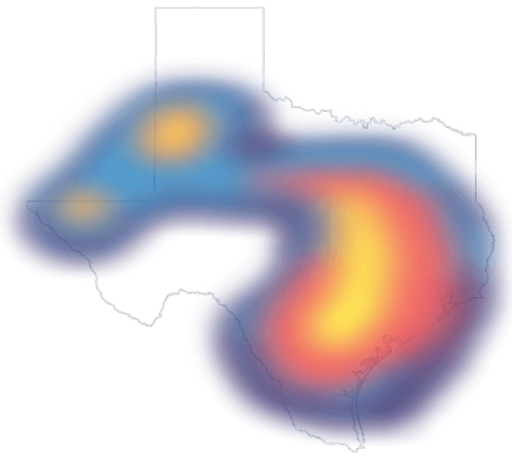
If you need this type of break, then the chances are that you already know what it means. If not, you can safely ignore it.

BUBBLE MAP



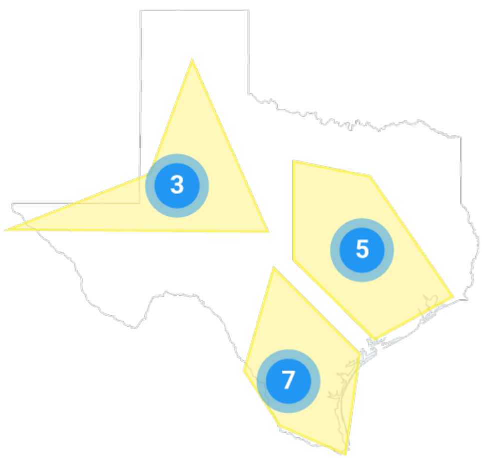
Also known as a "graduated marker map", a bubble map works in exactly the same way as a quantity map. But uses a range of bubble sizes rather than colors to visualize the variance.

HEAT MAP



A heat map is used in instances where our location data is so dense and tightly packed that we can't make visual sense of it. Heat maps are ordinarily used when mapping "points".

CLUSTER MAP



A cluster map solves the same problem as the heat map. It's a way to make sense of point layers that have a large number of features, but instead cluster nearby points together into a single point.

This cluster point usually uses a combination of color, size and labelling to communicate how many separate points each cluster encompasses.

JARGON ALERT:

Points, lines, & polygons

There are three basic geometry types used in GIS: Points, Lines, and Polygons.

POINTS

A point is a single locations on the map represented by a coordinate such as a lat/lon position. Points are used to map locations when the boundary of the location isn't important, such a pushpins on Google Maps.

LINES

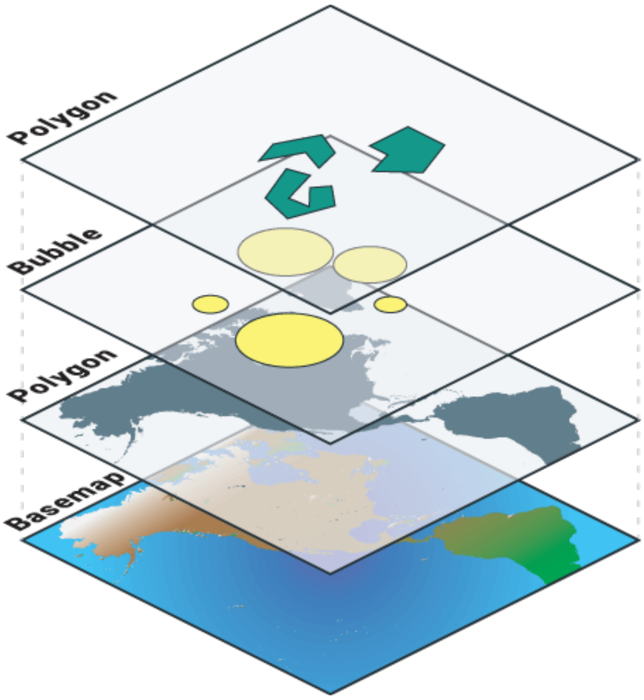
Lines are a series of interconnecting points. Lines are used to map linear features such as roads and rivers.

POLYGONS

Polygons are the same as lines except that the first point in the series of points and the last are always connected to form a closed loop. Polygons are used for boundary data such as country borders and property boundaries

STACKING LAYERS

A GIS map is often made up of multiple layers. The map types that we looked at above can be stacked on top of each other into a single map, each map in that stack is referred to as a “layer”.

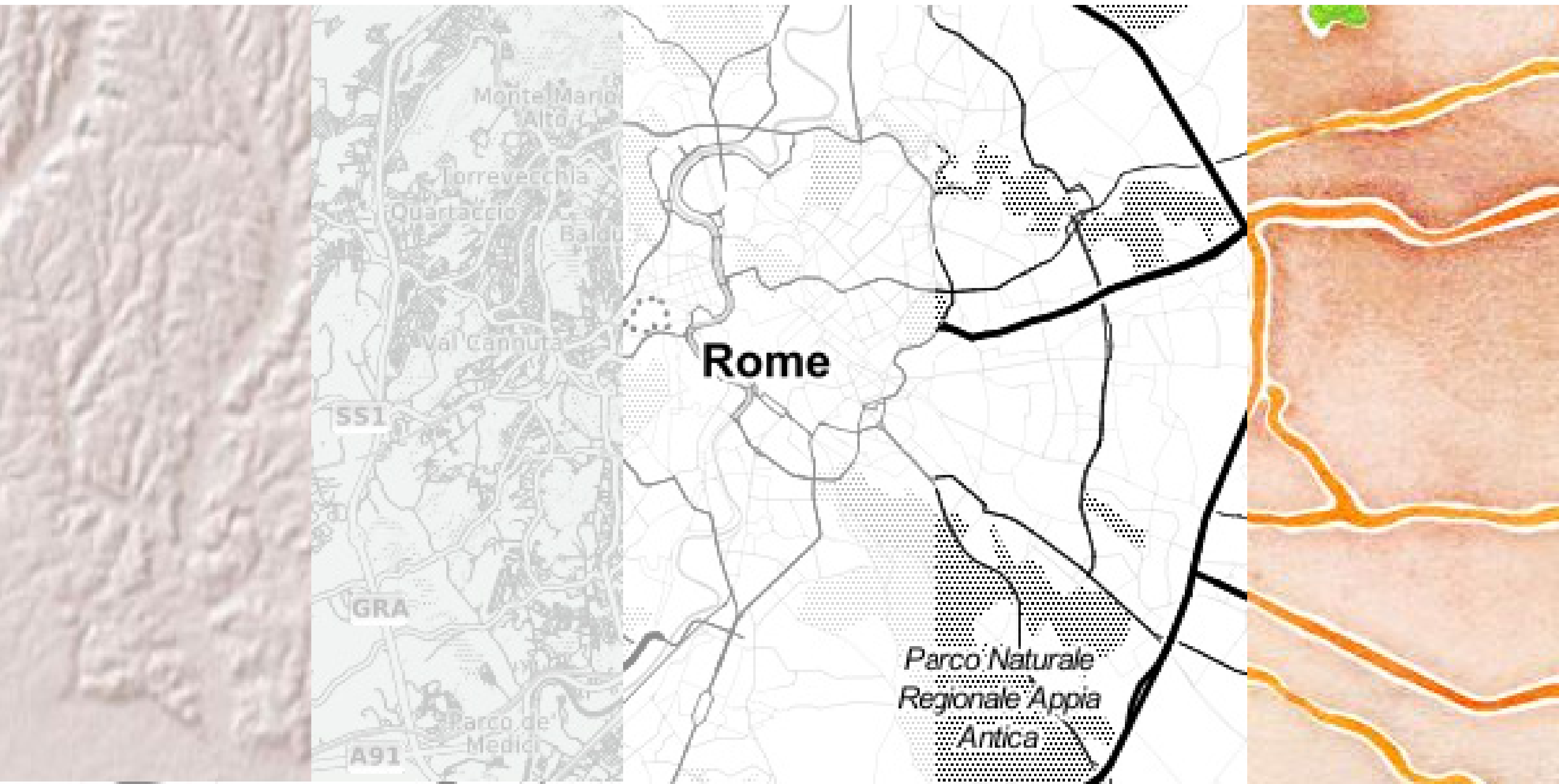


In most GIS systems each layer can be turned on and off in the map legend or moved up and down in the stack.

For example, in one map you might have one layer that contains the boundaries of each dealers territory, another layer that shows sales volume by county and another point layer showing the location of each dealers premises. The end user can then toggle each layer on and off as required.

Stacked map layers

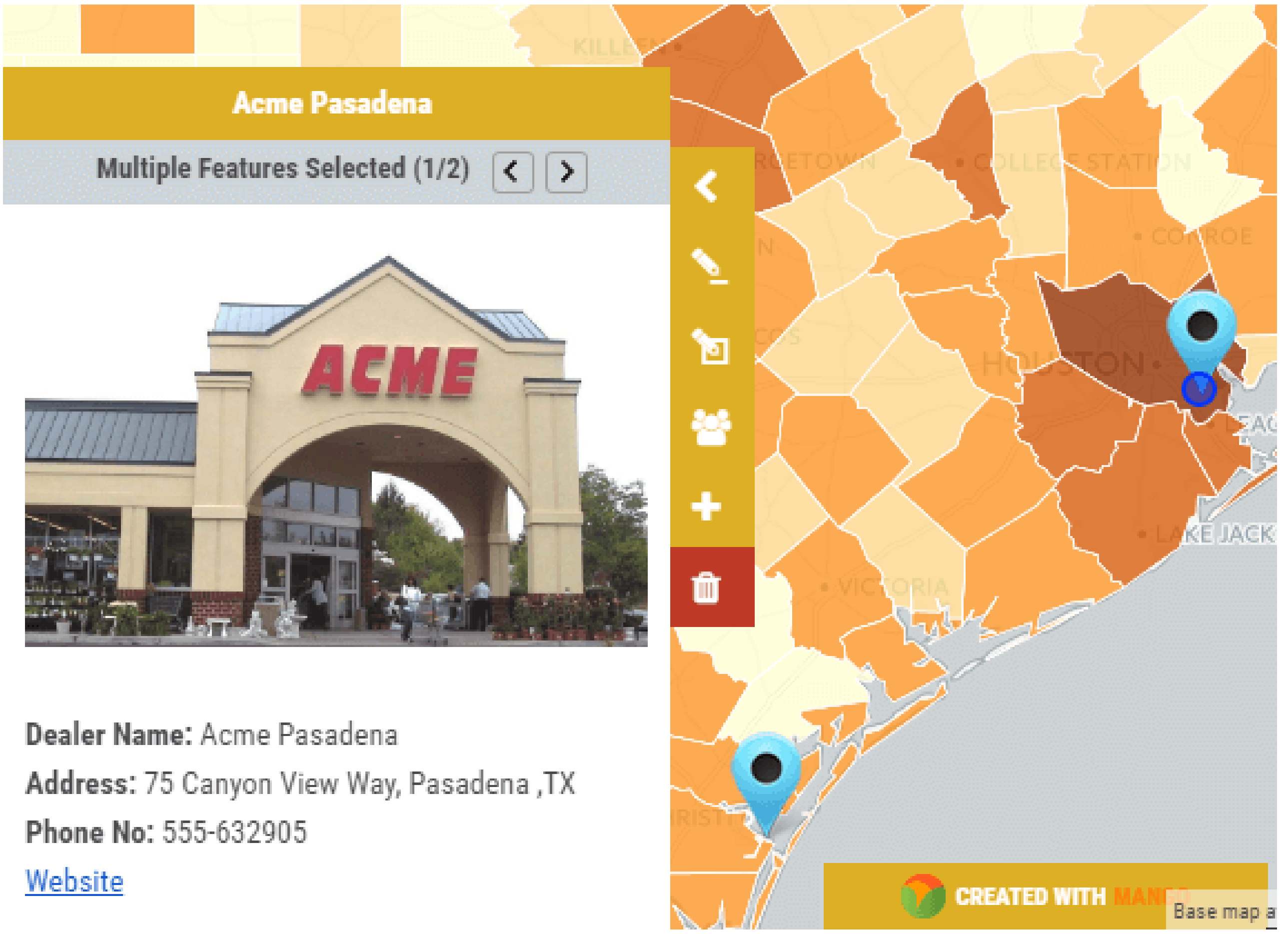
In web GIS systems it is also common to use what is known as a base map. A base map is a map similar to the street maps and satellite maps that you are used to seeing on Google Maps. Base maps give your map layers additional context and act as supplemental data.



MAP INTERACTIVITY

The layers on the map are only one half of the story. What separates a GIS map from a printed map is interactivity. The basic controls of most GIS systems work in the same way as Google Maps, which I'm sure you are already familiar with.

The map can be panned and zoomed to find an area of interest and then we can click on items on the map to find out additional information, in GIS this is called an "identify".



An identify popup

JARGON ALERT:

Identify

When we click on a feature in a GIS map and are presented with the attribute information in a table or popup for the clicked feature, this is referred to as an "identify". As in, "perform an identify" or "I identified the feature".

MAP TOOLS

Map visualizations are just part of the story. What really separates a GIS system from a printed map is user interactivity.

As I said earlier, GIS is a very broad subject and there are some very niche GIS systems and tools, but let's take a look at some of the tools that are common to most GIS systems, both on the desktop and the web.

MAP QUERY TOOLS

A GIS query tool, or selection tool as it is often know, allows us to filter a dataset based on criteria that we define. For example we could say "show me all of the counties in which we made more than $50,000 in sales last year", the GIS system would then highlight those features on the map.

Query tools allow maps users to do more than what the map maker envisioned when they styled the layers.

When making the map we will try and guess what data and visualizations the end user will find interesting and make sure we include layers that highlight those data points.

A query tool future-proofs our applications by allowing our users to build their own queries that we may not have envisaged when we first created the map.

In desktop GIS systems the query tools are usually very powerful and capable of building large and complex queries, the downside is that they also have a steep learning curve.

Web based GIS systems tend to have query tools that are more limited but much more intuitive, making them suitable for sharing with users that have no GIS experience.

This distinction is true across most features that are available on desktop and web GIS systems.

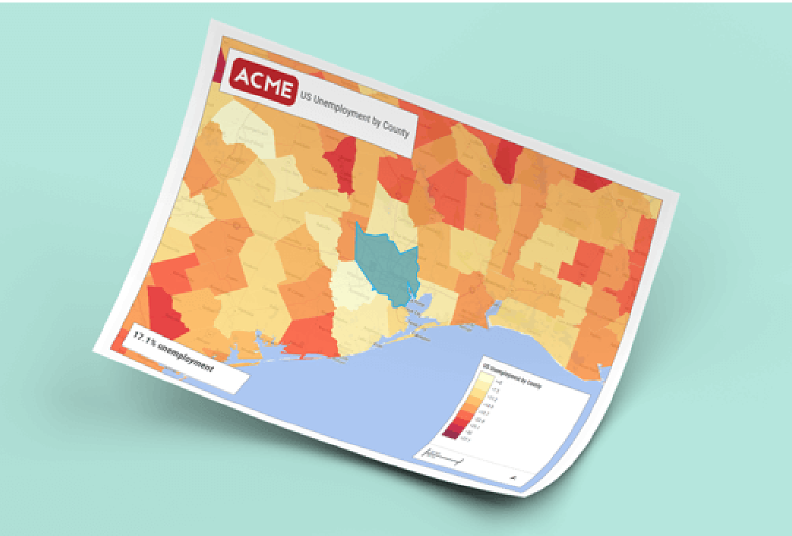
A desktop GIS is great for a power user that has time to invest in GIS learning activities, such as reading this book.

A web GIS system however, is a much better choice if you want to share your maps with a larger, more distributed audience. An audience which likely doesn't have access to a desktop GIS system, nevermind the time required in order to learn how one functions.

PRINTING MAPS

Printing maps is still a common task in GIS, especially on desktop systems.

It's still common for companies to have a GIS department that's responsible for creating maps and maintaining the GIS system. When non-GIS users require a map they will ask the GIS team to create a print or a PDF.



Slowly the internet is changing this workflow. It's now becoming increasingly common for organizations to make their maps available to staff and clients via a user friendly web GIS.

This is preferred, as good web GIS systems don't require any training or specific tools beyond a web browser.

SEARCH

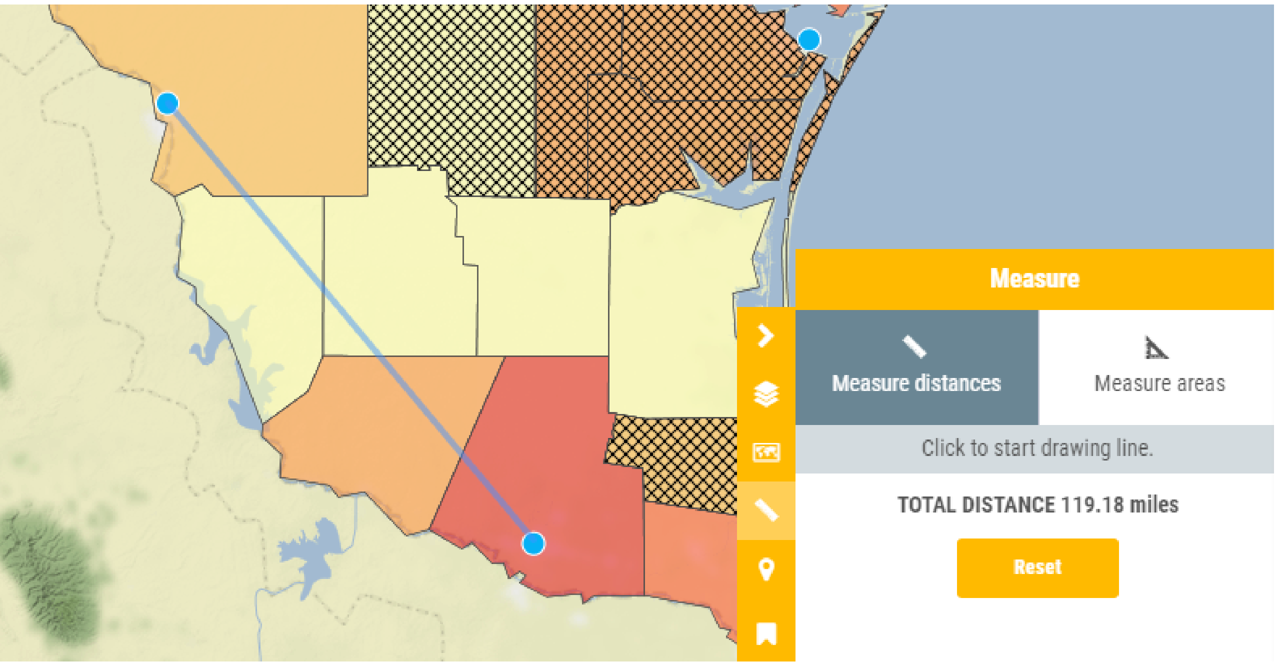
Maps can be searched by address or by values in the attributes of their layers.

Search is an area web the web tools tend to outperform desktop GIS systems. The web systems offer Google style fuzzy search that can search across multiple layers and list the results based on the quality of the match.

Search on desktop GIS systems still tends to be much more clunky and involve first choosing the layer and column within the layer that you wish to search.

MEASURING & DISTANCE TOOLS

With any map it's important to get a sense of perspective. Scale bars on a map let us know what distance those on screen pixels represent in the real world, and measuring tools allow us to accurately measure paths or areas on screen.



A web GIS measure tool

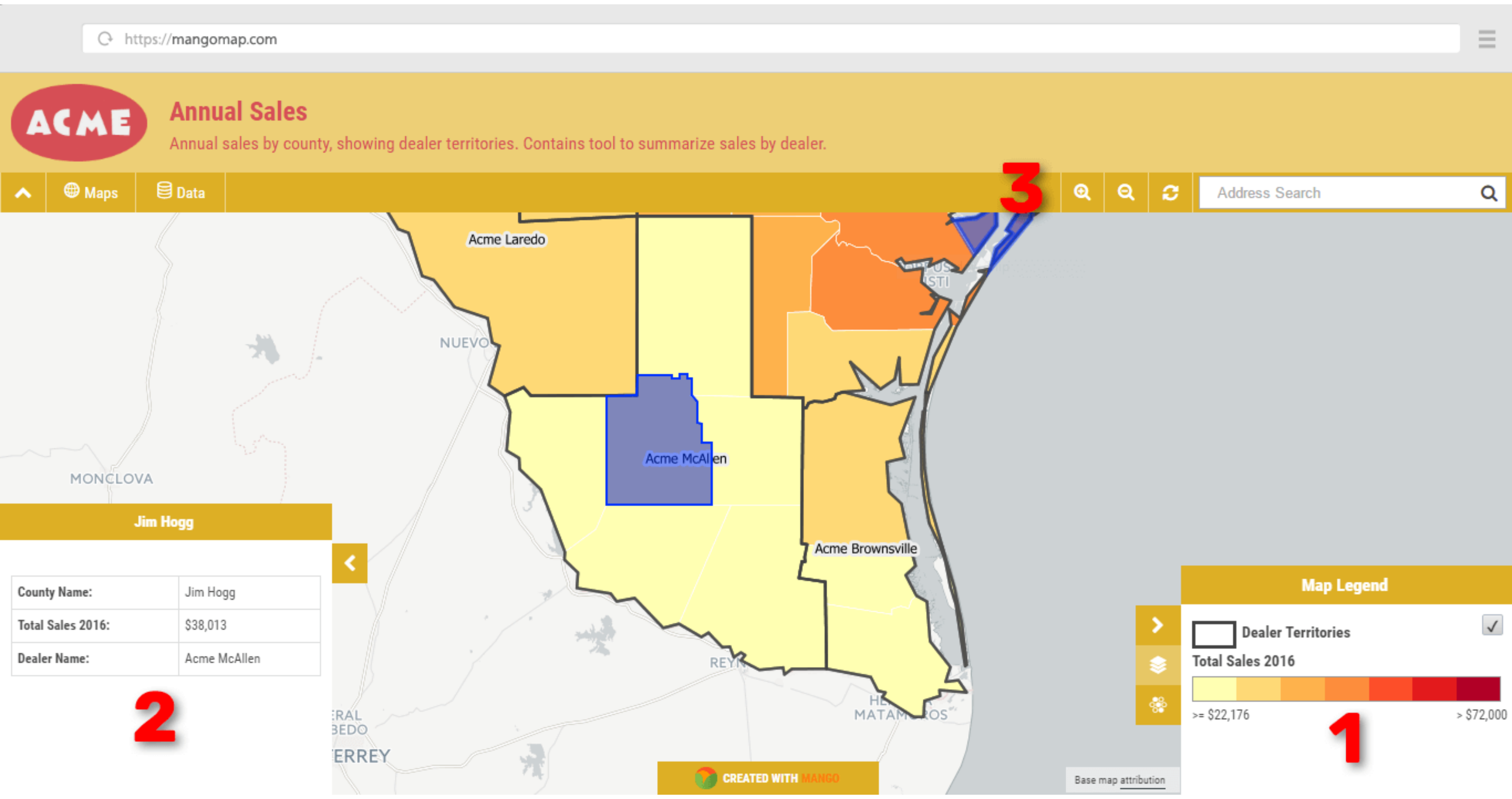
IS THAT IT?

Nope, the tools listed here are just the tip of the iceberg, and will only give you a taste of what's possible. Most GIS systems include a much larger range of tools.

And niche products will contain tools that are geared towards doing complex calculations based on geometry and the spatial relationship between features. We will be covering these features in more depth later in the book.

THE ANATOMY OF A GIS MAP

Every GIS program is different but they do tend to share some common DNA. Most GIS systems will have the following things in common:



1 LEGEND

The legend is where we make sense of what we are seeing on the map. The legend will list all of the layers and also show us the symbols and class breaks used in the styling of each layer.

2 INFO WINDOW

Sometimes it is displayed as a popup, other times it's show in a sidebar. The info window is used to display supplemental information, for example when you click on a map to identify a feature, the attribute data for the selected feature will be displayed here.

3 TOOLBAR

The toolbar will be displayed on the top or the sides of the interface. In desktop GIS systems these toolbars will be large and comprehensive, in web based systems they tend to be smaller and much more focused.

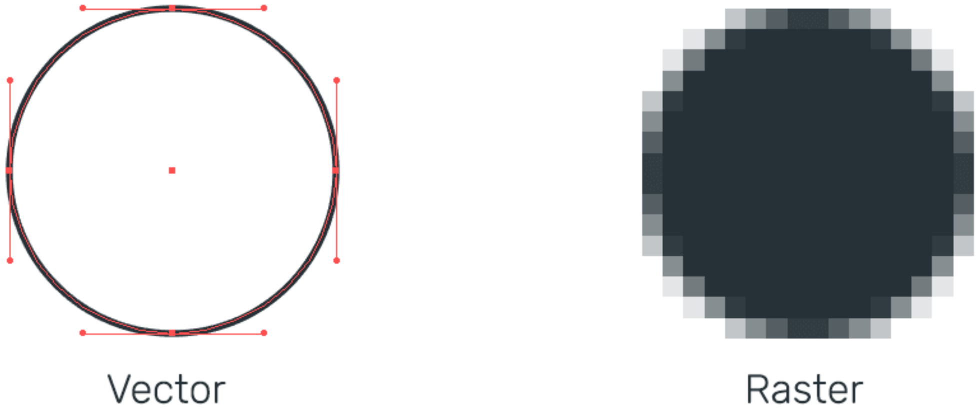
GIS Data

Now we understand what a GIS is and what is can do, the next step is to understand how a GIS is made.

But first, you'll need to take a crash course in GIS data.

If you think the term GIS is vague, then you haven't seen anything yet. There are a dizzying array of formats used for storing GIS data.

Before we delve into the various formats let's take a look at some fundamentals. Primarily there are two main types of GIS data: vector and raster.



Vector and Raster data

VECTOR DATA

You can think of vector data as instructions for how to render data. The best way to visualize it is to think of it as a spreadsheet with columns that contain your regular data, but in addition it always has an extra column called "geometry".

That column contains one or more coordinates that describe how to draw the point, line or polygon that represents that feature on the face of the earth.

RASTER DATA

If vector data is abstract, raster data is literal. Raster data is a bitmap image such as a TIFF or JPEG. This format is usually used for satellite imagery, aerial photography, elevation models and topographic maps.

INTRODUCING THE SHAPEFILE

The Shapefile is the most common format in GIS. It's a vector format that can be read by almost all GIS systems.

The name Shapefile is a little deceptive because the file is made up of at least four parts. The .SHP, the .DBF, the .PRJ and the .SHX.

It's not important that you remember what's in each part of of a Shapefile, but I think a brief explanation will help you better understand how GIS data is structured in general.

.SHP

The .shp file contains the geometry data of each feature in the shapefile.

.DBF

The .dbf is a dBase file which contains the attribute data for all of the features in the dataset. The dBase file is similar to a sheet in a spreadsheet, and can even be opened in Excel.

.SHX

The .shx is the spatial index. It allows GIS systems to find features within the .shp file more quickly.

.PRJ

The .prj is the projection file. It contains information about the "projection" and "coordinate system" the data uses.

GEOMETRY TYPE

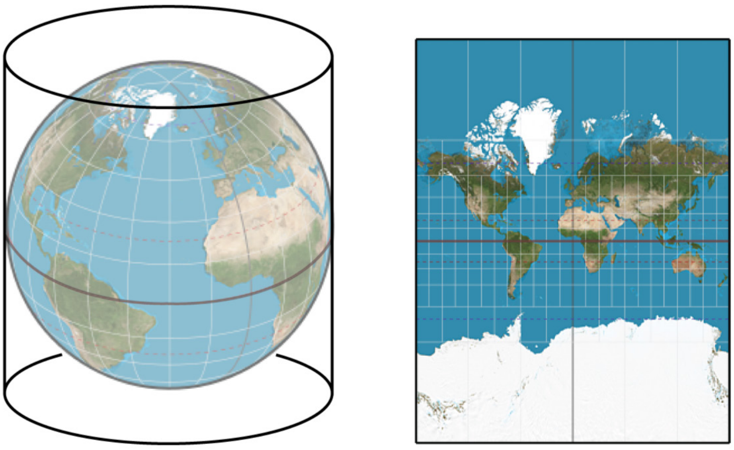
Every Shapefile can only contain one geometry type. This means that every feature in the dataset will be either a point, a line or a polygon. You can't have a dataset that contains a mixture of geometry types.

Most beginner and intermediate level GIS users never need to look any further than the Shapefile for storing and sharing map data.

So that wraps up our introduction to the Shapefile.

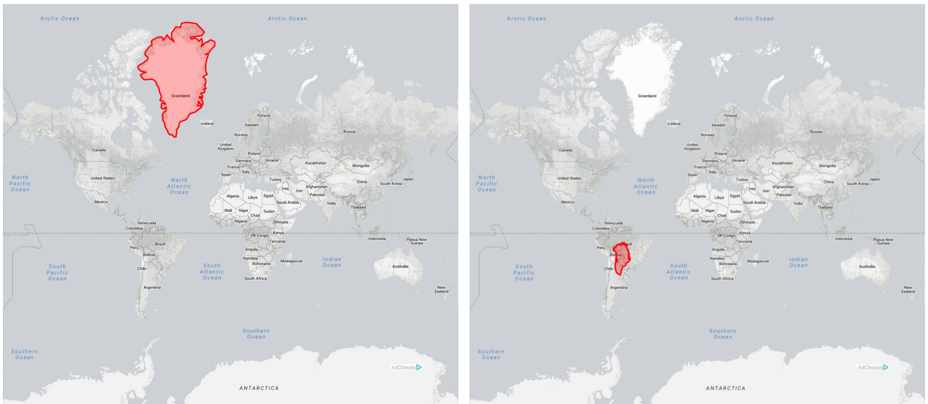
JARGON ALERT:
Projections & Coordinate Systems

You could fill an entire book on this subject, but the short answer is that the earth is a three-dimensional sphere, and your computer screen is two dimensional and flat. So in order to display the earth on your screen, it needs to be flattened.



Map projections distort the true size of countries

The flattening process creates distortions, and this is the reason that on some maps Greenland looks the same size as the whole of South America.



— Greenland isn't as big as you think it is, thanks to projection distortion. Left shows the projection distortion of Greenland, and right shows it's 'true size' when compared to South America. (thetruesize.com)

There are many different formulas of flattening the earth, each designed to cause less distortion in specific places on earth.

You don't need to understand how this process works, as the data you use will already have the correct coordinate system. And if you are making a new dataset, the default coordinate system used in most GIS systems (WGS84) will be suitable 99% of the time.

OTHER COMMON GIS DATA FORMATS

There a lots of other formats used in GIS. Each with their own distinctive benefits and drawbacks. Here's a quick list of other common formats that you might come across:

CSV [Comma-Separated Values]

Although the CSV isn't exclusively a mapping format, it is often used in mapping. The beauty of the CSV is its simplicity. This simplicity means they can be read by almost any program including the Excel or Google Docs.

It's literally a text file where columns are separated by commas and rows are separated by line breaks. When used in mapping, two extra columns are added to hold the x and y, or lat and lon. For mapping purposes this format is only really used for sharing point layers. The downside of the CSV is that they are very easy to break. Just one comma in the wrong place and the file becomes unreadable.

FileGDB [File GeoDatabase]

A file geodatabase (FileGDB) is a collection of files in a folder on disk that can store, query, and manage both spatial and nonspatial data.

This is a popular format amongst advanced GIS users. However, despite originally being touted as the favourite to replace the old but entrenched Shapefile as the de facto standard for sharing GIS data, the FileGDB never gained the popular support that many believed it would. The main reason being its lack of support amongst open source GIS platforms.

TAB
[MapInfo TAB]

This format is very similar to the Shapefile and is the default format used by the MapInfo desktop GIS system made by Pitney Bowes.

KML
[keyhole Markup Language]

This is the format most likely to be known by non-GIS users, as it is the default file format of Google Earth.

Unlike the other datasets covered here, KML does more than just store geometry and attribute data, it also contains lots of configuration options for Google Earth maps. This extra information, however, makes KML less portable, as the additional information is only relevant to Google Earth and isn't of any value to other GIS systems.

GeoJSON

JSON, or to give it its full name *JavaScript Object Notation*, is a light-weight data interchange format. It's primarily used by software developers due to the ease with which it can processed by web applications.

GeoJSON is a form of JSON that also contains geometry data. It's not often used as a format for sharing spatial data for human consumption but is very popular as an output for API's (application programming interface).

GeoTIFF

The GeoTIFF is the most widely supported raster data format. TIFF is a bitmap image format similar to GIF, PNG or JPEG.

A GeoTIFF is just a regular TIFF that also contains special metadata that allows us to know where it should be placed on a map. A GeoTIFF is an uncompressed format. There are many other raster formats that offer compression to reduce the filesize, but these tend to be proprietary formats that require additional paid software to use.

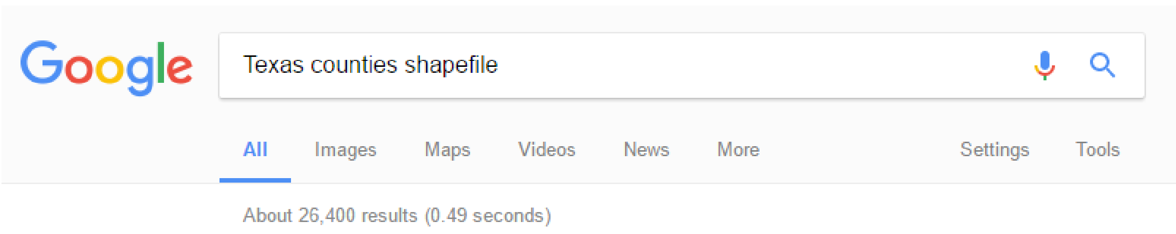
SOURCING GIS DATA

If your data corresponds to commonly agreed boundaries, such as states, counties, or any other administrative boundary, you're in luck! The good news is that there is a ton of useful map data for use, completely free and without restriction, with spatial datasets covering just about anything you can think of.

The bad news is that there isn't one single authoritative source of map data. Locating data will require a little time, a little patience, and a search engine.

When looking for spatial data it's best to start with Google, and simply search for the name of the dataset you want followed by the word Shapefile. Of all spatial data formats, shapefiles are the most ubiquitous, so you're more likely to get a larger range of results than searching for some other formats.

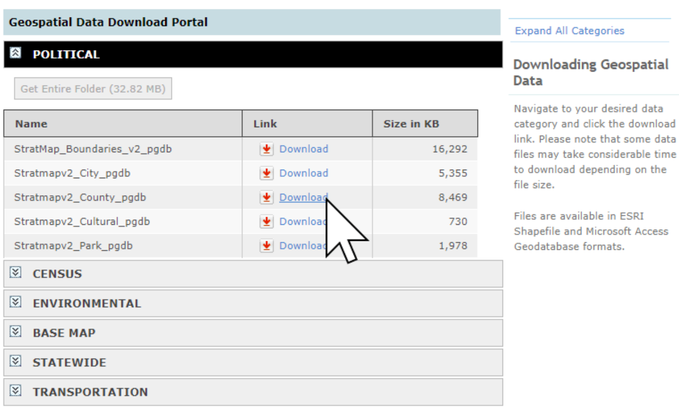
When searching you may come across other formats such as KML, GeoJSON, Tab Files or GeoTiff's, but Shapefile's are usually the easiest to find.



Search for "texas counties shapefile"

Shapefiles are usually provided as a zip containing the various files for the Shapefile.

Many map dataset portal don't offer example maps of the datasets they offer, so it's usually necessary to download the data and open it in a desktop GIS to see what the dataset contains.



If you don't have any luck with Google, then the next step is to be more direct.

In the US, Canada, and Europe, most national and local government websites will have a GIS section that contains GIS data available for download. These sites are often not very user friendly, so finding and downloading things can be a challenge. Perseverance is the key!

If all else fails, you can try contacting people who might have access to the data and asking if you can use it. For example if you wanted to map the forests of California and couldn't find the relevant GIS data, you could try emailing the state forestry office to see if they can share their data.

Most open data available for download online is available for use without restriction, but sometimes restrictions are in place on redistribution of the data, or simply that attribution of the data source is required. Either way, crediting the source when using third party data on your maps is the polite thing to do.

PREPARING GIS DATA

Later in the book I will be showing you the exact steps for preparing map data using a desktop GIS. But first I want to introduce to you the general concepts.

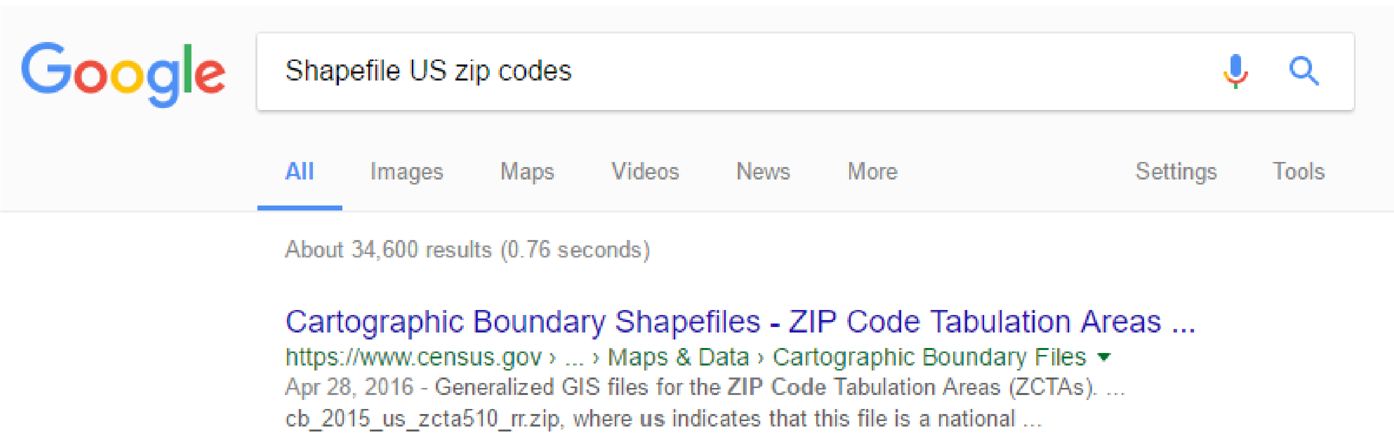
So let's just jump straight in and take a look at the most common data preparation tasks.

TABLE JOINS

So, here's the scenario: You have a spreadsheet that contains your data listed by zip code, and you would like to build a quantity map to visualize the distribution of the sales.

To do this you are going to need to find zip code map data and then join your spreadsheet to the map data so that it can be used in a GIS.

So first you go and find the data by Googling "Shapefile US zip codes".



Find data to perform a table join

Once you have sourced the data you would open it in your desktop GIS and use the table join features to merge the datasets.

The table join works by choosing a column from the map data and the spreadsheet that contain shared unique values; e.g. the zip code.

The GIS system will then find all records that match and join them together by appending the columns from your spreadsheet to the end of the attributes for the map data.

	A	B	C
1	agent	sales_2016	ZIP
2	Timothy Frey	877576.40	75148
3	Triston Clark	89692.73	75149
4	Myles Cantu	758083.43	75150
5	Moses Peters	551055.08	75152
6	Braelyn Bruce	377208.46	75153
7	Trevon Lyons	380568.72	75154
8	Lillie Nelson	935424.30	75155
9	Quintin Mooney	159779.61	75156
10	Jordan Trevino	199607.29	75158
11	Nia Norman	324737.85	75159
12	Nadia Williamson	976853.68	75160
13	Gordon Adkins	517472.40	75161
14	Lilly Stanley	34500.48	75163
15	Kody Singleton	666045.02	75165
16	Joseph Shepherd	28042.12	75166
17	Lindsey Kaufman	201091.39	75167
18	Victoria Golden	11028.81	75169
19	Kolby Gilmore	846637.76	75172
20	Emelia Macdonald	732627.11	75173
21	Valentino Randolph	625285.78	75180
22	Jameson Briggs	70923.15	75181
23	Rhys Barber	210752.57	75182
24	Cash Hughes	510476.84	75189

Your data

ZIP	PO_NAME	STATE	SUMBLKPOF	POP2004	POP04_SQN	SQMI	Shape_Leng	Shape_Area
75148	MALAKOFF	TX	5702	6387	73	87.45	1.2064818849	0.0216479238
75149	MESQUITE	TX	54841	58591	3527.5	16.6	0.3782267628	0.0041410976
75150	MESQUITE	TX	55260	60001	4950.6	12.11	0.2767509834	0.0030216607
75152	PALMER	TX	3058	3800	109.4	34.75	0.8433105223	0.0086301590
75153	POWELL	TX	118	124	13.1	9.39	0.2306721414	0.0023235743
75154	RED OAK	TX	25924	29508	622.3	47.4	0.7630706274	0.0117844495
75155	RICE	TX	2586	2777	24.5	113.51	0.8119969044	0.0281222265
75156	MABANK	TX	15862	17894	282.3	63.38	1.2545171559	0.0157142235
75158	SCURRY	TX	3869	5188	58.2	88.95	0.7901722214	0.0221016022
75159	SEAGOVILLE	TX	14358	16357	254.5	64.23	0.7658073848	0.015985136
75160	TERRELL	TX	19666	21532	173.8	123.91	1.1693710566	0.0308843102
75161	TERRELL	TX	5738	6170	49.5	124.72	0.8810466899	0.0310742559
75163	TRINIDAD	TX	2466	2717	61	44.53	0.8330758698	0.0110259547
75165	WAXAHACHIE	TX	28927	33154	205	161.74	1.4818443938	0.0401223256
75166	LAVON	TX	498	580	140	4.13	0.1900276964	0.0010359268
75167	WAXAHACHIE	TX	4764	5561	65.2	85.09	1.0612073603	0.0211166157
75169	WILLS POINT	TX	12154	13873	78.5	176.78	1.3008105876	0.044033153
75172	WILMER	TX	2786	2931	335.3	8.74	0.2891238559	0.0021749607
75173	NEVADA	TX	4039	5151	107.8	47.78	0.8581072184	0.0119457635
75180	MESQUITE	TX	19523	20969	2702.1	7.75	0.2598665827	0.0019337994
75181	MESQUITE	TX	16388	20080	1154.7	17.39	0.3451357559	0.0043326652
75182	SUNNYVALE	TX	3007	3733	188.3	19.82	0.3313749426	0.0049414024
75189	ROYSE CITY	TX	10417	12712	119.5	106.27	1.3089330495	0.0265419292

Map data

Joining spreadsheet data with geospatial data

GEOCODING ADDRESSES

This is the most common scenario for organizations looking to map business intelligence data, where the data contains addresses.

The process of transforming an address into a coordinate is called geocoding, and is offered by most desktop GIS systems.

It works by passing the address to a geocoding service that stores data about the exact location of addresses.

The geocoder outputs a Shapefile that contains your spreadsheet data in the form of attributes and the lat/lon coordinates that can be plotted on as points on the map for each record in the new dataset.

	A	B	C	D	E
1	Address	City	State	Zipcode	Name
2	1 Crossgates Mall Road	Albany	NY	12203	Apple Store Cross Gates
3	Duke Rd & Walden Ave	Buffalo	NY	14225	Apple Store Walden Galleria
4	630 Old Country Rd.	Garden City	NY	11530	Apple Store Roosevelt Field
5	160 Walt Whitman Rd.	Huntington Station	NY	11746	Apple Store Walt Whitman
6	9553 Carousel Center Drive	Syracuse	NY	13290	Apple Store Carousel
7	2655 Richmond Ave	Staten Island	NY	10314	Apple Store Staten Island
8	7979 Victor Road	Victor	NY	14564	Apple Store Eastview
9	1591 Palisades Center Drive	West Nyack	NY	10994	Apple Store Palisades
10	125 Westchester Ave.	White Plains	NY	10601	Apple Store The Westchester
11	103 Prince Street	New York	NY	10012	Apple Store SoHo

Address data in a spreadsheet ready to be geocoded into lat lon coordinates.

FILTERING A DATASET

Often when we find third party map data, we end up with way more that we need. For example, a county map of the US should contain all 3,142 counties and county-equivalent areas, however, you might only need one state's counties, or just a few select counties.

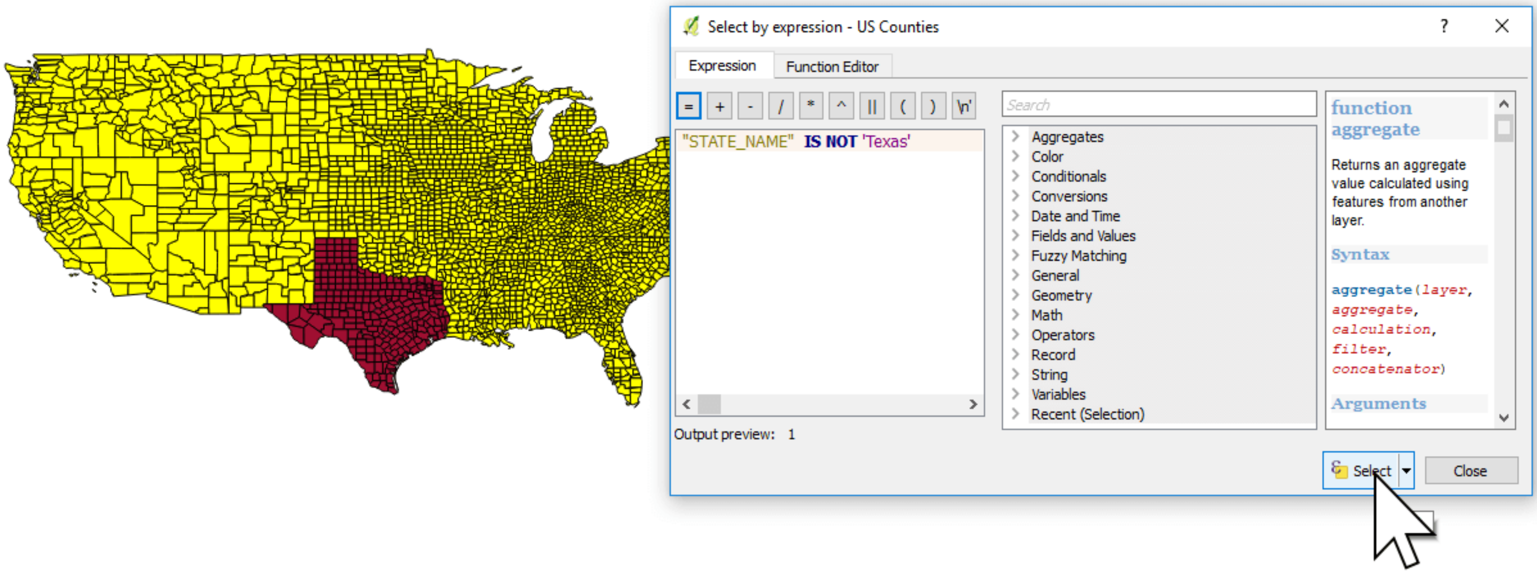
You could join your data with the whole dataset, but then you've got a of distracting data that will be rendered on the map.

Instead, trim the fat, and filter out the unwanted data with a desktop GIS so you're only dealing with your areas of interest. It's cleaner, and it's simpler than you might think.

The task here is to remove all of the features from the dataset that you aren't interested in. In a GIS this is done by selecting features using an expression or query. For example:

```
“STATE” IS NOT ‘Texas’
```

In QGIS this, expression would select all of the counties in the dataset which has a `STATE` attribute value other than Texas. Once selected, removing them is usually as simple as hitting the delete key.

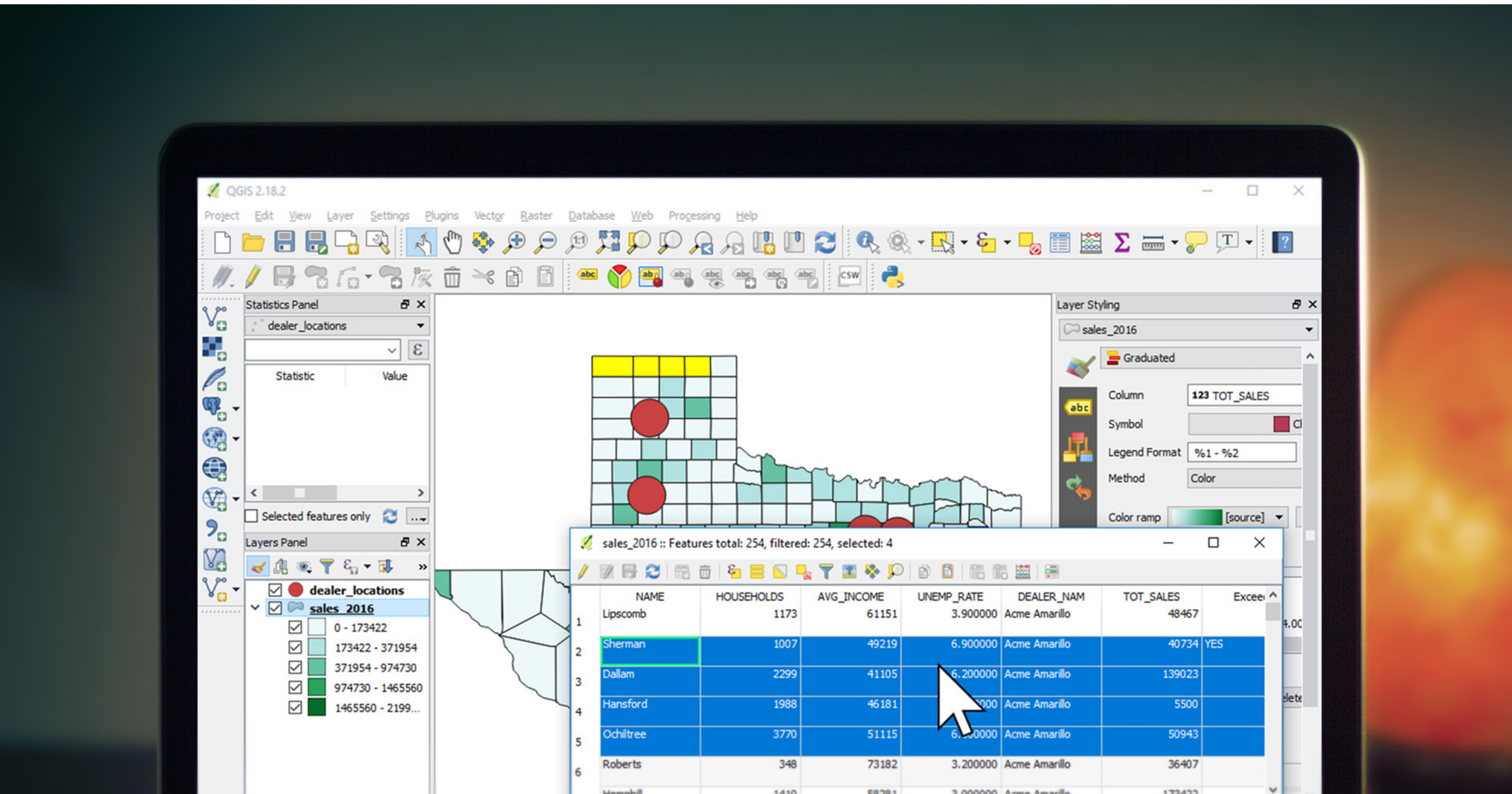


The Select by expression tool in QGIS lets you build complex expressions to target data for selection.

The above examples are just an introduction to the concepts. In the Tutorials chapter, we will be showing exactly how to perform these actions using the a free desktop GIS system called QGIS.

GIS Software

With so many GIS software solutions available on the market, how do we even begin choosing which one to use?



In this chapter we will give you the background knowledge that you need in order to answer that questions for yourself.

For those of you that have tried looking for GIS software for the first time, there's one thing that you will know for sure: you aren't short of choices. So I guess the first thing you want to know is which one is best? The answer is…

There is no answer.

It's best to think of GIS software like tools in a toolbox, you don't want to use a screwdriver to bang in a nail. Often GIS solutions will involve a "software stack".

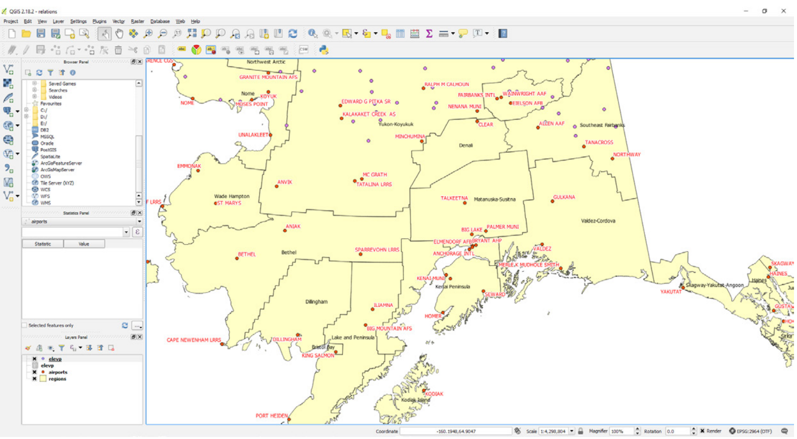
This means we use several tool in unison. For example, we use a desktop GIS system for creating and modifying our GIS data, then once the data is ready we upload it to a web GIS system in order to distribute the maps within our organization.

TYPES OF GIS SOFTWARE

Choosing the right GIS tool is really about understanding the problems that you are looking solve. Most GIS software falls into one of the four categories below

DESKTOP GIS

Desktop GIS platforms are the swiss army knives of GIS. They are used for creating, editing, visualizing, managing, and analyzing geographic data.



Desktop GIS systems are complex programs. As with tools such as Photoshop or Microsoft Excel, most users can figure out the basics pretty quickly, but becoming a true master can take years.

The QGIS Interface

The best way to go about learning a desktop GIS system is through a problem/solution approach. So rather than trying to learn everything it's capable of, it's best start with figuring out the parts that are applicable to the problems you want to solve.

As I mentioned in a previous chapter, for most GIS beginners those problems are:

- Creating visualizations, category, quantity, heat maps, etc
- Performing queries and visualizing the results
- Joining spreadsheets with map data
- Geocoding addresses
- Editing feature geometry
- Editing feature attribute data

Desktop GIS systems are capable of a whole lot more, but these topics are usually a good starting point.

Later in this chapter we will take a look at the most popular desktop GIS systems, ranging from free and open source solutions to expensive paid products.

WEB GIS

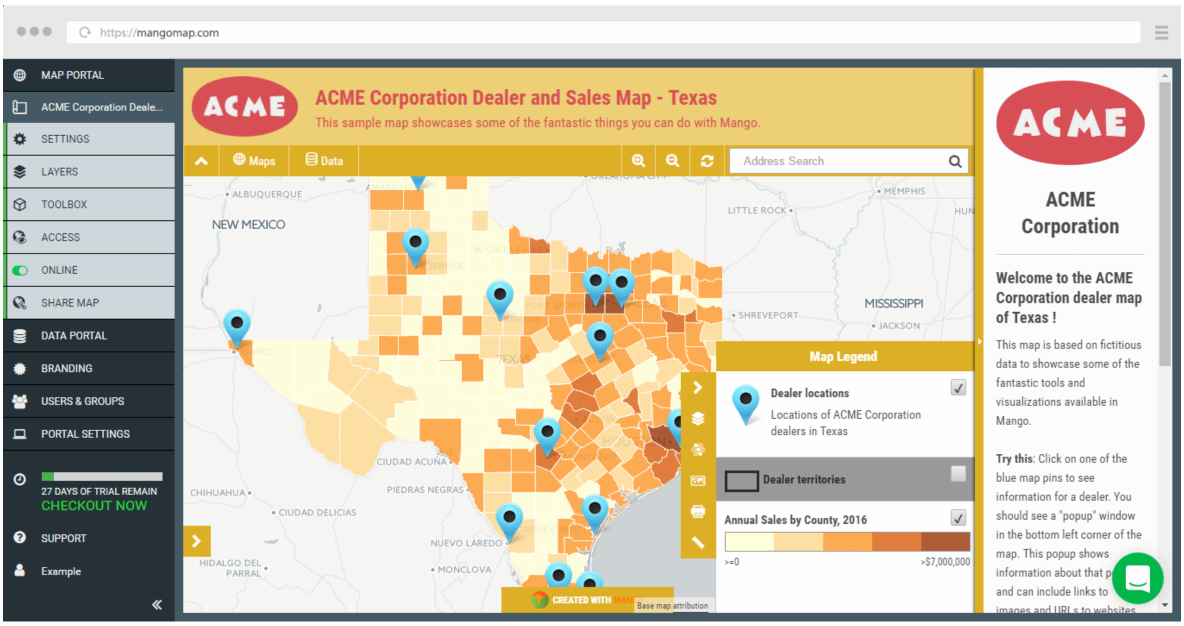
Like all forms of information distribution, the internet has completely revolutionized GIS.

In a time before web based GIS systems, when you wanted the answer to a location related question or to view a specific visualization, you would contact the GIS team and ask them to prepare a map for you.

Once the map was ready you would take a walk over to the GIS department to view the map on a workstation or collect a print out.

With the latest generation of super easy to use web GIS systems, all that has changed. Web GIS gives non-expert GIS users the ability to easily interact with map visualizations and answer their own questions without having to involve a GIS team.

What's more they can answer those questions from anywhere, anytime, using any web accessible device.



The Mango map editor interface

Web GIS systems used to be very expensive to develop, deploy and maintain. A typical deployment would require programmers, servers and system administrators.

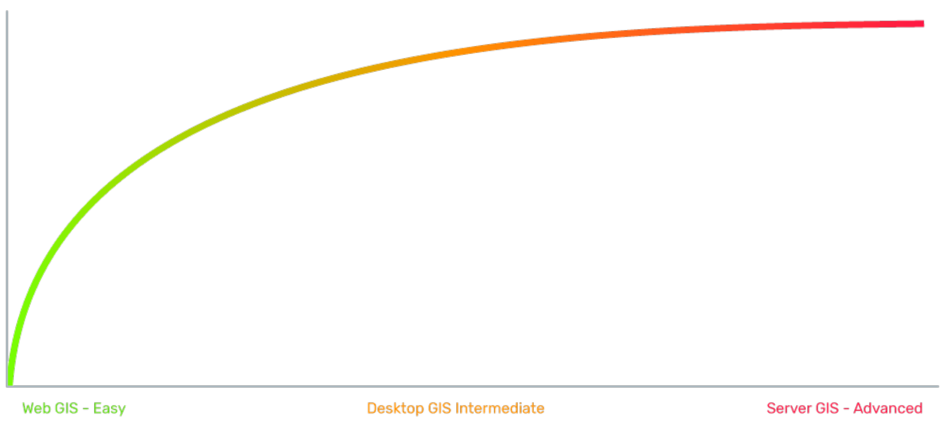
But now thanks to cloud based services such as Mango these barriers to entry have been removed.

Now web maps can be created for a very low cost, using a point and click interface in a matter of minutes using nothing more than a web browser!

The next chapter is devoted exclusively to web GIS. We will be going into much more detail regarding the benefits of web GIS.

SERVER GIS

If a web GIS has the lowest barrier to entry, then a server GIS definitely has the highest, with by far the steepest learning curve.



The steep learning curve of Server GIS

As you are new to GIS you likely won't need to concern yourself with server based GIS solutions. Generally they are only used for large scale mapping efforts and very specific niche tasks.

The most common server GIS system is a spatial database.

A spatial database is used as an alternative to the file formats discussed earlier. They are commonly used in cases where the dataset is extremely large, or in scenarios where numerous people need to access and edit the data at the same time, possibly from different locations.

SPECIALIZED GIS

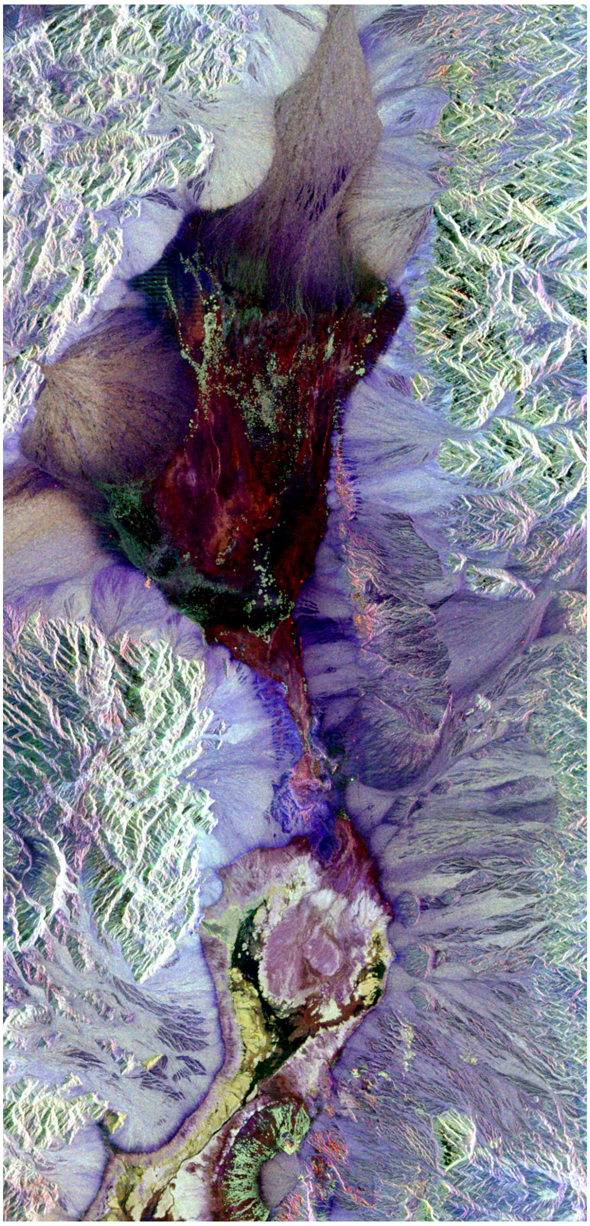
As with server GIS systems, specialized and niche GIS systems are not something you need to concern yourself with in the beginning.

These products tend to perform very specialized tasks. There are really too many use cases to list so I will just give you a couple of interesting examples just to show you how diverse this sector is.

IMAGE CLASSIFICATION

Image Classification is a specialized GIS application whereby pictures of the earth are turned into meaningful map data e.g. land use types or classes. These pictures or images are captured digitally and separated into different spectral bands including the visible spectrum (Red-Green-Blue) and others such as infra-red.

Based on the spectral information stored for each pixel, we can perform analysis known as classification to assign a pixel to a class based on its similarity to reference pixels. The process of "images to information" makes it easier for professionals and decision makers to understand patterns in the natural world as well as change over time.

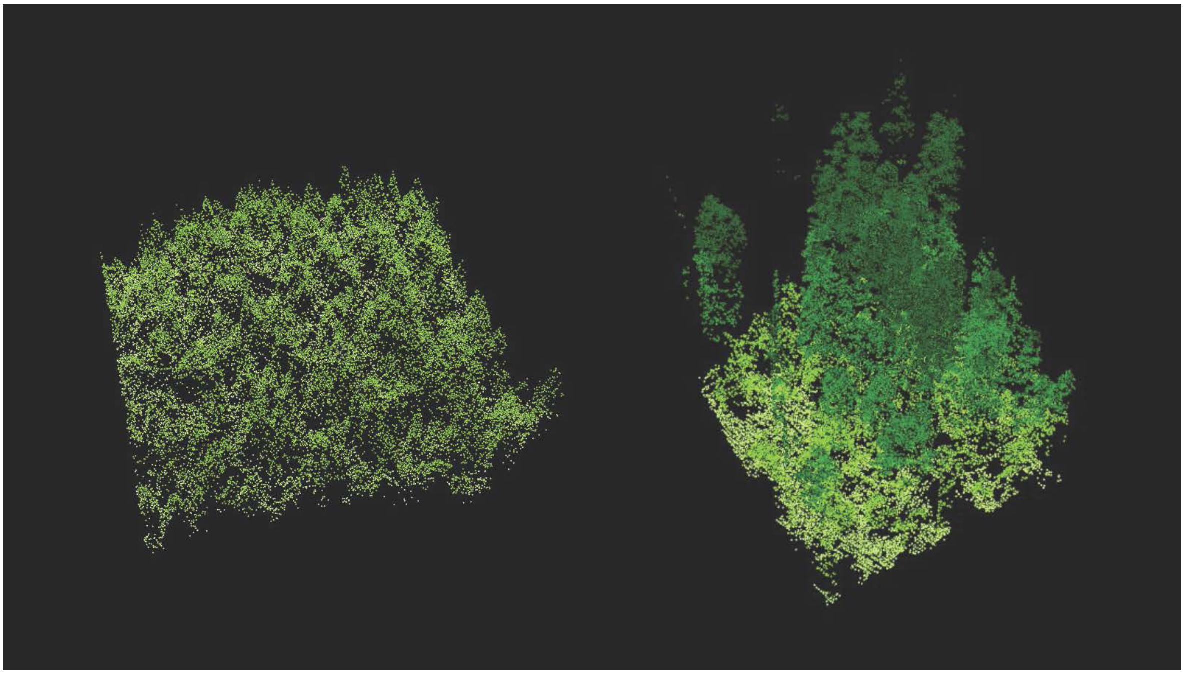


Synthetic aperture radar image of Death Valley colored using polarimetry.

By NASA/JPL - http://photojournal.jpl.nasa.gov/catalog/PIA01349, Public Domain, https://commons.wikimedia.org/w/index.php?curid=3730092

LIDAR

Lidar data is best understood as a “point cloud” of the world outside, whether that’s a tropical rainforest or a highway interchange. Modern Lidar systems are capable of modeling the real world with a very high point density with thousands of points per square meter.



Lidar imaging comparing old-growth forest (right) to a new plantation of trees (left).

By Sarah Frey, Oregon State University - Flickr, CC BY-SA 2.0, https://commons.wikimedia.org/w/index.php?curid=48412762

As you might expect, making sense of the is information is a monumental task requiring some serious processing power and sophisticated routines. The best of these are capable of being trained to automatically identify features based on size shape and also to identify ground level, with all above-ground features removed.

Getting Started with Desktop GIS

Now that we covered briefly the various types of GIS software available, the next few chapters will zero in on the types of GIS software that you, as a beginner you are most likely to need. Namely; desktop GIS and web GIS.

Later in this chapter we will take a look at the most popular GIS desktop platforms and what differentiates them, but before do that we are going to look at the features and functions that they all have in common before we move onto web GIS.

DESKTOP GIS: THE BASICS

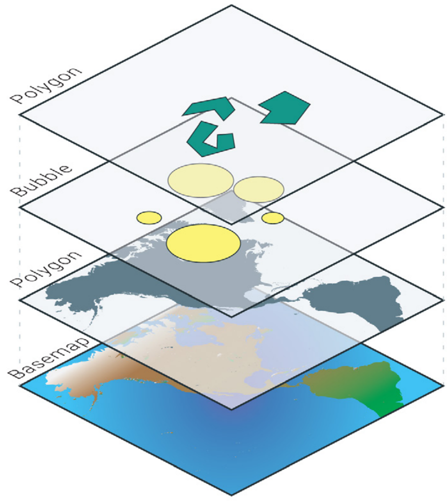
As a beginner your decision over which desktop GIS software to use likely won't be driven by functionality. All desktop GIS systems have the same core functionality which is usually more than enough for the requirements of new GIS users.

So let's take a look at what those commons functions are.

ADDING LAYERS

Most maps are made up of multiple layers stacked up on top of each other. In the legend the name of each layer will be displayed alongside a checkbox that can be used to hide and display the layer.

Also the legend will contain a graphic that let's us know what the colors, fills and strokes on the map represent.



Sometime those layers are all using different datasets, but it's also possible for multiple layers within the map to be using the same dataset.

For instance you might have one quantity layer that shows this year's sales by county and another showing last year's sales.

Both layers could be using the same dataset as it contains attribute columns for both this year and last year.

Stacking layers like this is interesting because it allows us to make comparisons. By putting this year's sale data above last year's in the stack, we can turn this year's sales layer on an off in order to make a quick comparison and visualize any changes.

STYLING LAYERS

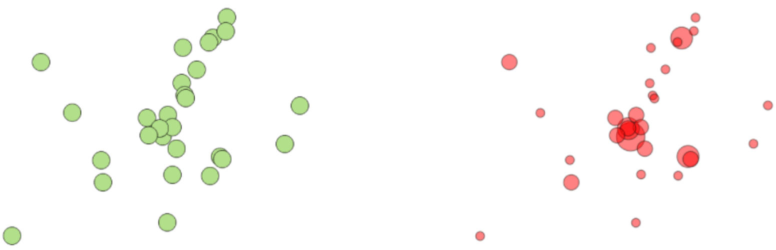
Desktop GIS systems usually give you very fine grained control over the appearance of the layer.

You can use category and quantity class breaks as we covered earlier, but you can also do a lot more. The three things that you can visually style are the stroke, fill and labels of a feature.

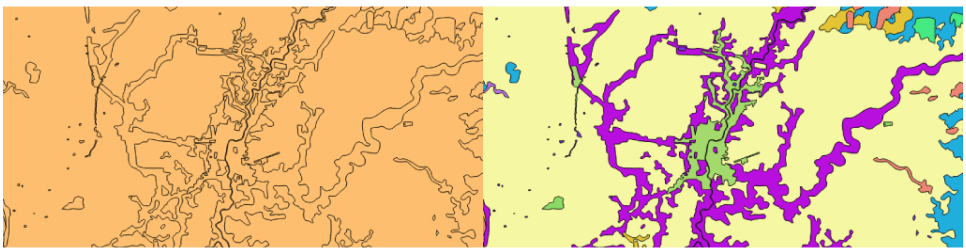
STROKE

The stroke is used for lines, or the outlines of points or polygons. You can control the color, thickness, transparency and style of the lines. You can also stack various line styles on top of each other to create some interesting composite styles.

Line stroke styles

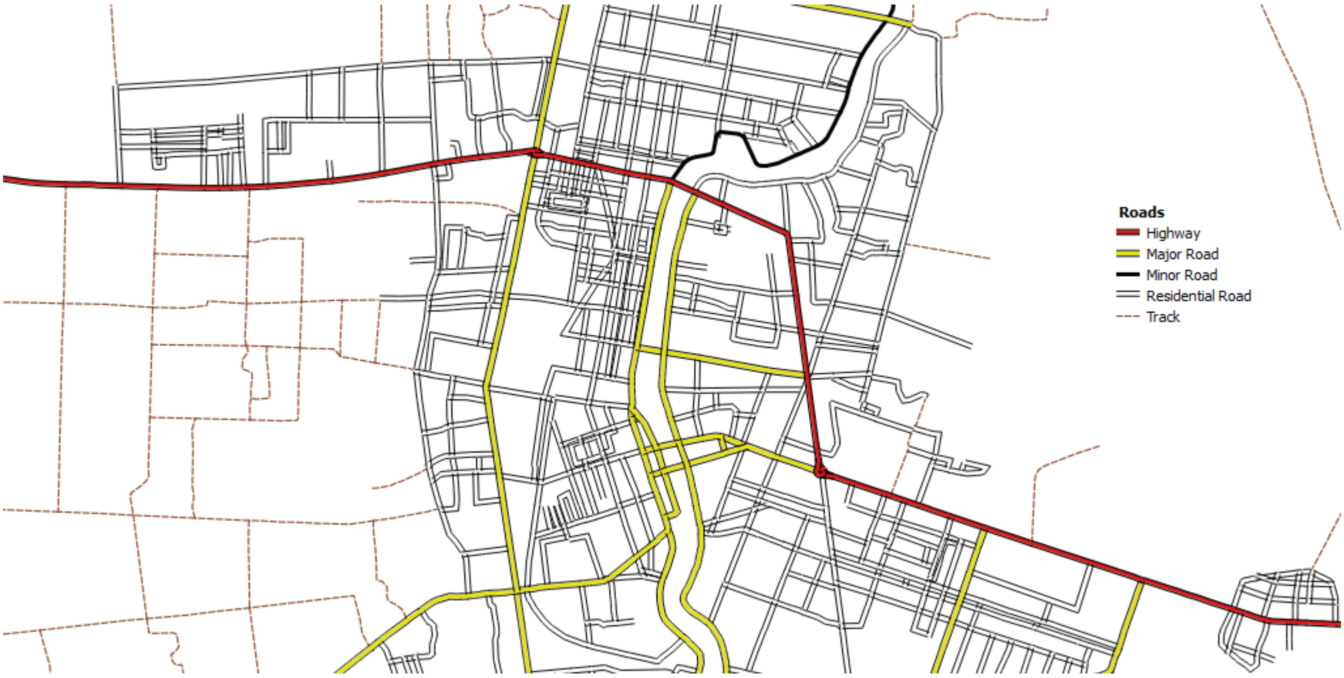


Point stroke styles



Polygon stroke styles

The strokes can also be modified based on the attribute data. For example if we had a layer called “Roads” and the data had an attribute column called `type` which is either `Highway`, `Major Road`, `Minor Road`, `Residential Road` or `Track` we could tell the GIS to render a different style based on the type value.



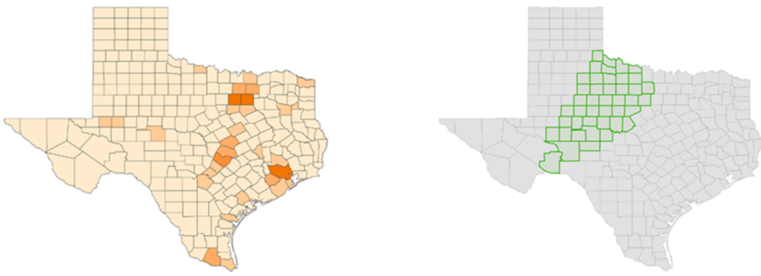
Varied stroke styles based on attribute data

FILL

A fill can be used for either a polygon or the center of a point. You can control the color, style and transparency of the fill and as with the stroke it can be changed based on an attribute value in the data.

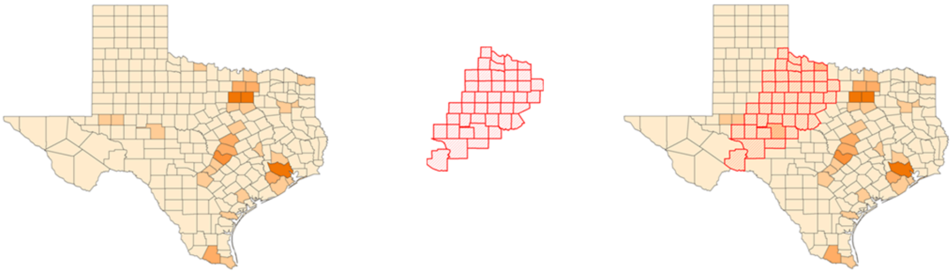
Fill styles are particularly interesting as they often allow you to communicate two pieces of information at the same time by stacking layers.

For example we might want to visualize our sales data by county for the previous year and also at the same time see which counties were targeted in our annual sales plan.



Isolating annual sales plan target areas to create a new layer

To achieve this we can have a standard quantity layer to represent the sales data. We can then add a second layer that only displays the target counties and gives them a cross-hatched fill.

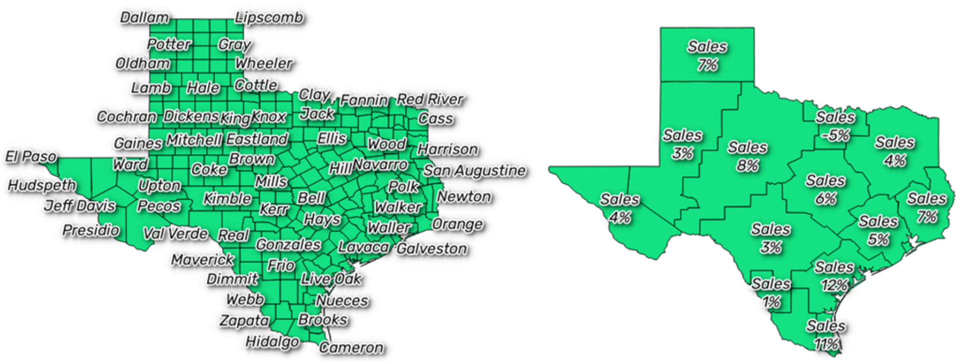


Layering annual sales data with annual sales plan target areas

Now without needing to turn layers on and off we can visualize our sales data and our target counties at the same time.

LABELS

Lines, points and polygons can all be labeled in a GIS. The label can be made up of one or more attributes from the feature that's being labeled.

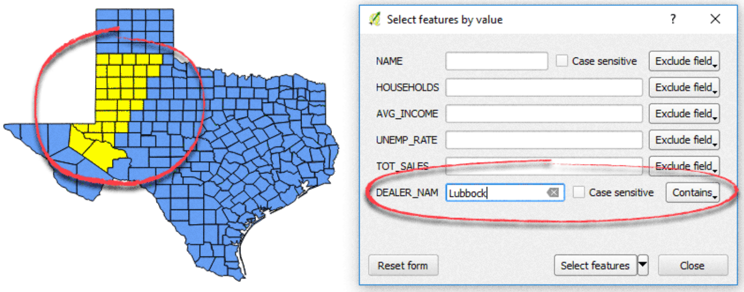


Feature labels should be relevant and clear

You can also modify the font, size, fill and outline of labels. An important consideration when styling labels is to make sure that they stand out from the map and can be easily read.

CREATING SELECTION SETS

A selection set is a subset of features within a layer. Selection sets are used to visualize a particular group of features or to perform bulk operations such as aggregating, editing or deleting. Selected features will be highlighted on the map.

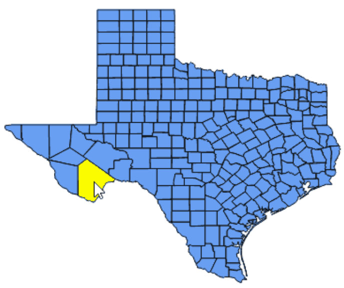


Creating a selection set by filtering values in QGIS

A selection set can be created in a number of different way, using tools within the GIS:

MAP POINT AND CLICK

This is the most straightforward. Just point and click on the features that you wish to add to the selection.

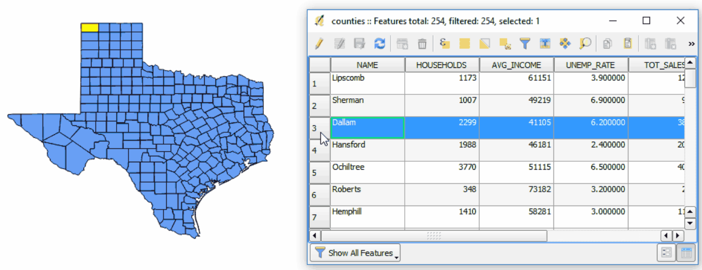


Clicking features to create a selection set in QGIS

TABLE POINT AND CLICK

Layers in your map have what is know as a table view. The table view just presents the same data that's displayed on the map, but as a table that looks similar to a spreadsheet.

From the table view, it's also possible to select single or multiple features.



Selecting table rows to create a selection set in QGIS

ATTRIBUTE QUERY

Sometimes referred to as an expression. You can select features by building rules, e.g;

```
Select counties where “sales > $1,000,000”
```

```
Select roads where “type IS ‘highway’”
```

Different desktop GIS systems have different ways of constructing these queries, with some using a syntax similar to Excel formulas, and others using a graphical query builder. Either way the concept is the same.

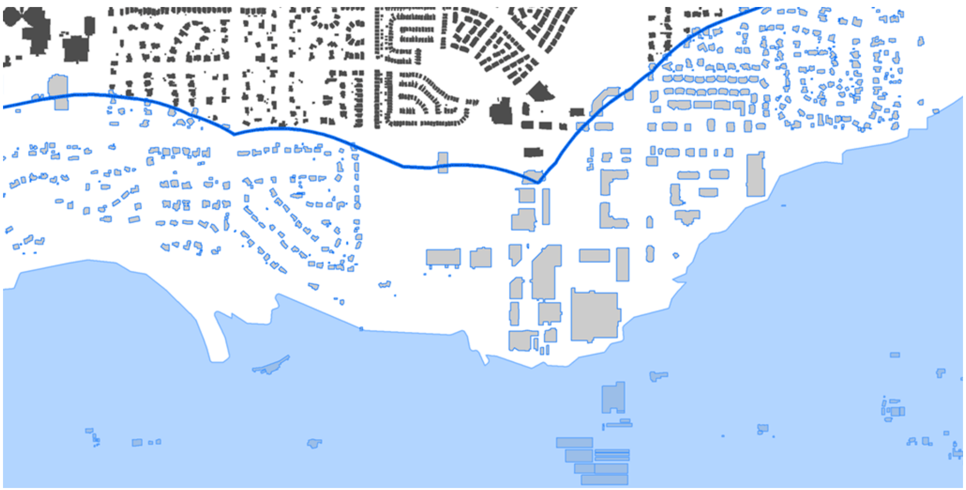
SPATIAL QUERIES

Now this is where things start to get really interesting, as this is functionality that you definitely can't perform in a spreadsheet program.

A spatial query allows you to create rules that relate to the location of the feature in relation to others.

Spatial queries can be used to answer some very complex questions, such as:

- Which distributors have overlapping sales territories?
- Which houses are within 500m of the 2010 flood area?



Flood query results highlighted in blue

What's more, these spatial queries can be combined with attribute queries. For example:

- Select counties that don't contain a dealership AND sales were above $1,000,000
- Select all properties within 10 miles of a store AND the property value is greater than $500k

These combined queries are extremely powerful. Within a few click we can get insights into location related patterns and trends that could take weeks to manually extract from a spreadsheet.

WORKING WITH SELECTION SETS

Once you have a selection set, what can you do with it?

- Extract the selected features to a new dataset
- Delete the features from the existing dataset
- Create a report that aggregates attributes from the selected features
- Bulk edit the attributes of the selection set
- Create a printout of the selection set
- Export the selected features to a spreadsheet

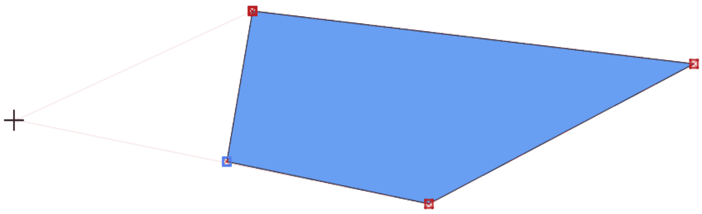
EDITING AND MANAGING DATA

A GIS isn't just somewhere to upload your data, it's also a place to create and maintain your data.

A GIS allows you to add, edit and delete features. Editing the attribute data works in much the same way as a spreadsheet. You can edit feature attribute by either clicking on features from the map view or selecting features in the table view.

EDITING FEATURES

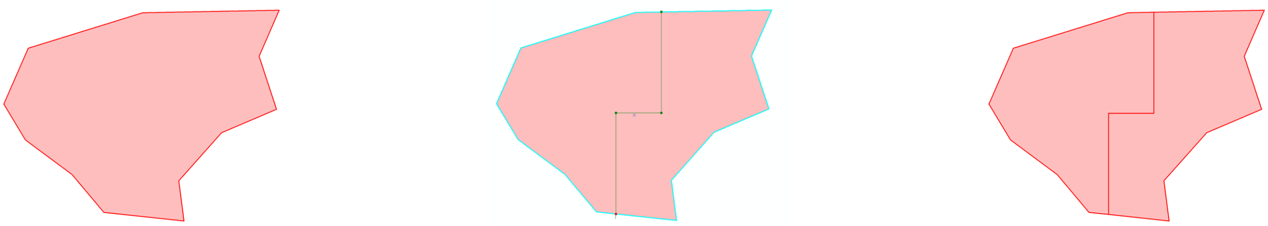
As with a spreadsheet, you can apply formatting rules, edit cells directly or even create whole new columns and populate them with values using formulas.



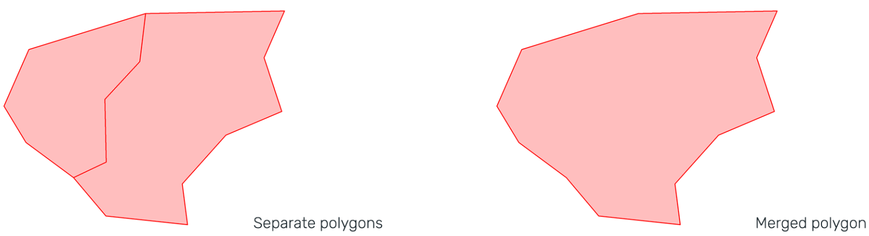
Editing a feature by dragging corner nodes

Unlike a spreadsheet however a GIS can be used to create edit and delete geometry. Points can have their positions changed and you can modify the shape of lines and polygons by clicking and dragging vertices.

Another common task is splitting and joining features within a layer. Joining features is usually just a case of selecting the adjoining features and then pressing a join button, while a split is achieved by drawing a line through an existing feature.



Splitting a feature by drawing a line



Merging features

There are lots of other tools within a GIS that can join and split data in a range of exotic ways, but they are outside of the scope of this introduction.

CREATING NEW FEATURES

To create features by hand we usually need a point of reference such as a satellite image that allows us to place a point or trace a line or polygon.

This process is known a digitization and is quite straight forward in a GIS.

Once we’ve added the geometry, the attribute data is usually added via an onscreen form.

Another way of adding geometry for a new feature is using a path that’s been recorded in a GPS and uploaded to your computer (usually in GPX format).

POPULAR DESKTOP GIS SOFTWARE

The two most used desktop GIS programs by far are ArcGIS and QGIS. Both offer similar functionality but there are some major differences in terms of how they are distributed and their surrounding ecosystems.

ARCGIS FOR DESKTOP

ArcGIS for Desktop is produced by Esri who have been the market leaders in the GIS software industry since the beginning, and ArcGIS for Desktop—their flagship product—is the most well known and widely used of all GIS systems.

ArcGIS Desktop is usually purchased from a local distributor. The cost varies by region, but at the time of publication the most basic ArcGIS license is $1,500 in the USA.

Additional modules will incur additional costs, as will a maintenance contract which is required if you wish to receive software updates.

Support is provided either by local Esri distributors or online by Esri themselves via support forums.

There are different tiers of support and the response times will vary based on which tier you have subscribed to.

ArcGIS for Desktop is only available for Windows computers. There is no Mac on Linux versions available.

QGIS

Previously known as Quantum GIS, QGIS is a free and open-source desktop GIS system that's been around since 2002.

QGIS is extremely popular, with a very large user base and a very active supporting community. It's also open source, which means it can be downloaded and installed for free without restriction.

Being open-source means that it's been developed by a community of developers rather than a company, and the license for the software ensures that it will always be free. Another advantage it has over ArcGIS for Desktop is that it works on Windows, Mac or Linux systems.

Support is provided by the very active and very supportive community. In addition paid support packages are available from companies such as Boundless [boundlessgeo.com].

QGIS is the default choice for new GIS users, especially when the price differential is factored in.

At Mango we are QGIS users and strongly recommend QGIS as the first choice for all new desktop GIS users.

Later we will be devoting an entire chapter to QGIS in which we will show you how it can be used to perform the basic GIS tasks that we covered earlier in the chapter.

Web GIS

The Internet has revolutionized GIS.

GIS is no longer too expensive, too complicated or too isolated. It's now easy to create engaging interactive web maps in minutes, and what's more, they can be shared with anyone, anywhere, on any device.

We always knew GIS was powerful, we all knew we should be using it, and we all had a list of go to excuses to justify our reasons for not acting:

- We can't afford GIS
- It's just too complicated to get set up
- Our users can't learn to use a GIS map
- It won't work because our team is distributed
- We work in the field, not a GIS lab

One by one, web GIS has systematically destroyed all of the old objections to GIS adoption. Now you can literally set up and share an easy to use, mobile ready web map and distribute it throughout your organization in an afternoon.

What's more, it will cost less than your office coffee bill!

The web has revolutionized every aspect of information distribution, be it news, video, the encyclopedia or how we communicate. In GIS, the shift is even more dramatic, a web GIS is about so much more than functionality: it's about the democratization of data.

IDEA: DEMOCRATIZATION OF DATA

Throughout ancient times purple was associated with royalty. The only way to produce this specific color was to extract it from a small mollusk found near the ancient Phoenician city of Tyre (situated in present-day Lebanon).

As many as 10,000 shellfish were required to obtain just a single gram of Tyrian purple. Due to its rarity and cost to produce, only affluent rulers from Egypt, Persia, and Rome could afford to wear it.

In the middle ages, Queen Elizabeth I of England decreed that only close relatives of her royal family could dress in this regal color.

In 1856, everything changed when an 18-year-old chemistry student, William Henry Perkin, stumbled across the first synthetic dye, mauve, while researching an anti-malaria drug for a professor. The young Englishman astutely patented the dye and spearheaded its mass production in the midst of the Industrial Age.

Suddenly, common people could afford to dress in a color that was previously reserved for only the rich and powerful. Not surprisingly Perkin became wealthy, retired at 36 to focus on research, and was subsequently knighted for his scientific contributions.

Data is the New Purple.

— Brent Dykes, The Age Of Data Democratization: How To Effectively Share Data Across Your Business

If you think desktop GIS is powerful, you haven't seen anything yet. Web GIS isn't about outperforming desktop GIS or even competing with it. It's about making the power of GIS available to everyone.

While watching a proficient GIS user visualizing map data in their desktop GIS system is pretty cool, seeing Susan in sales — without any prior training and using only her web browser — locate every household within her territory that isn't a client, in a zip code with an average household income above $250k, with a median age between 35 and 45, is nothing short of mind blowing!

Your job when building web GIS systems is to create something that's so focused and so refined that it's self explanatory. The user doesn't need training, they don't need to be incentivized to use it. The power and benefits are so obvious they will be wondering how they ever worked without it.

In this chapter we are going to show you the ingredients that make up a great web GIS system.

THE SCALPEL VS THE SWISS ARMY KNIFE

Early web maps tried their hardest to replicate the look, feel and functionality—read complexity—of desktop GIS systems. It reminds me of the famous Henry Ford quote:

> ***"If I had asked people what they wanted, they would have said faster horses."***

— Henry Ford

It took time before the GIS industry began to realise that what we needed wasn't more, it was in fact less.

What's funny is that this shift didn't come from the deeply entrenched GIS industry, instead it came from Google when they released Google Maps. Although Google Maps isn't a GIS in the traditional sense of the word, it is extremely simple and very easy to use.

When Google Maps burst onto the scene with a product that was so easy to use, that within weeks everyone from preschoolers to grandparents were navigating the world using nothing but a web browser, the entire GIS industry did a collective face palm and said "Of course…".

Since then, web GIS has had something of a renaissance. Everyone now realizes that the true power of web GIS is simplicity and availability.

So instead of web GIS systems that try to do everything, the now accepted best practice is to build "story maps".

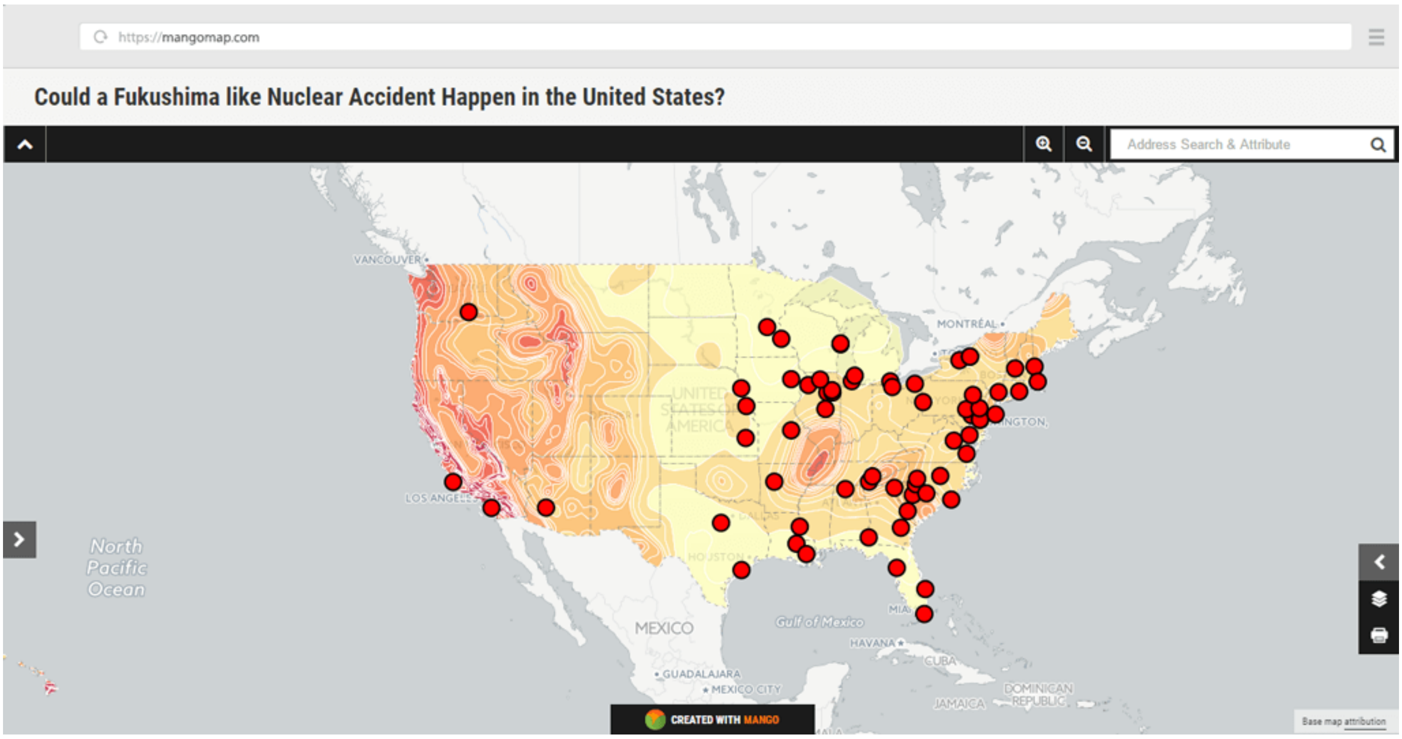
A story map is a map that's laser-focused on answering one particular data related question or workflow. But our users have lots of questions, I hear you say. Then build lots of maps!

STORY MAPS

If you are taking the time to produce a web map it means that you have data that is worth sharing and therefore contains a visual message that will be of use to others.

We use the term 'story map' to describe a map that is extremely focused and exists to tell the story of the data in the clearest way possible.

A truly great web map will leave the user with absolutely no doubt about the story that's being told through the data and they will leave feeling that they have learned something new through their interaction with the map.



Could a Fukushima like accident happen in the US?

The map above is a great example of a focused high impact map with a clear message.

As soon as the map is loaded the user is greeted with the question in large font at the top of the page "Could a Fukushima like Nuclear Accident Happen in the United States?", immediately they know what they are looking for in the map data that is presented before them.

The map itself uses rich colors for maximum visual impact and large red markers for the nuclear facilities.

For those of us used to stylising maps on the desktop with subdued pastel colors and cross hatch fills it can take a little getting used to these high impact color schemes but we need to remember that when making web maps it's not only the medium that's changed, it's also the audience.

The internet is the land of constant distractions and miniscule attention spans, your map has to compete with those distractions and engage the user from first second the page loads.

When we look at the legend we can see that it's extremely simple with only two layers, nuclear facilities and seismic hazard data.

We could have easily of added a point layer showing the location of each individual earthquake but would it have added to the story or just acted as an unnecessary distraction from the maps core story? When it comes to story maps, less is definitely more.

WILL WEB GIS REPLACE DESKTOP GIS?

The short answer is no — at least not anytime soon. The two technologies are complementary and have a symbiotic relationship. Desktop GIS is to web GIS what a word processor is to a blog.

A desktop GIS is fantastic at creating and manipulating GIS data, but useless when it comes to distribution and accessibility.

A popular workflow is to use a desktop GIS for the data preparation and heavy lifting. Then once the data is ready using the data to create a web GIS application in order to put the maps and data into that hands of those who need it.

Tutorials

Step-by-step guides to simple
but powerful GIS workflows.

When organizations decide to dive into web mapping, it's common that they won't have all map-ready data required to deploy maps. In most cases, you'll have data across a range of databases and spreadsheets, with place names, or addresses, or zip codes.

Unfortunately, none of these data resources are what we'd call map-ready data. In most cases, you will need to source and prepare data from a third party source, and merge it with your existing business data.

You're probably already collecting lots of valuable data. Many business tools and databases already contain a mountain of useful location information: Addresses, place names like state and county names, or even administrative codes like zip codes, or fips codes are all super useful for visualizing business data on maps.

Taking that location data, and merging it with map-ready spatial data can seem like a massive barrier to entry into web mapping and online GIS, but it really doesn't have to be.

We've put together this series of useful tutorials to help you really get the most out your business data by joining and combining it with existing map-ready data, and exploring the insights that web map visualizations offer in an interactive, easily shareable GIS web map.

The tutorials walk you through step-by-step, and require no previous knowledge of mapping or GIS.

They also can be achieved on any PC using free software. The desktop GIS client QGIS is our preferred GIS application at Mango. It's free, open-source, and has a plethora of community developed tools that can do amazing things with map data.

The tutorials cover the following useful workflows:

⟶ Create a Web Map from a Spreadsheet Containing Regions

⟶ Create a Web Map from Addresses in a Spreadsheet

⟶ How to Remove Unwanted Areas From Your Map Data

⟶ How to Turn a List of Locations into an Online Heat Map

Create a Web Map from a Spreadsheet Containing Regions

Spreadsheets can be fickle beasts. We've all use them, we all kind of know how to use them, sometimes we even love them - but are we really getting the maximum value out of them?

If you're not a tabular wizard who can massage the data into just the right chart, or the perfect subset of figures, it can all just look like so much noise. It's too easy to miss key insights and opportunities, and there's nothing worse than leaving money on the table when it comes to business.

If your organization has any physical presence, you're probably also collecting geographic data as part of your business intelligence, but extracting location based insights from static spreadsheets is a challenge.

Yes, you can create subsets of data by region or county, pivot sales data into regions, and binning data for charts that highlight sales trends, but trying to understand the physical geography of your data is often where we hit the wall. Lots of sheets, lots of figures, but not much actual insight.

UNCOVERING HIDDEN INSIGHTS

Much of this type of analysis is near impossible in a spreadsheet, and if you've spent time and money on collecting business intelligence, you know that extracting the right answers is a battle you can't ignore.

Thankfully, there's a simple way to combine all that data you've collected in spreadsheets with existing spatial data to create maps that reveal actionable insights for your organization.

Say you have a spreadsheet containing some interesting data about each county in the US. Visualizing that data on an interactive web map will not only aid visual exploration of the patterns in the data, but also can be layered and enhanced with other business data such as sales data, franchise territories, expansion opportunities, or even data that might help mitigate risks to your business.

No matter what type of data you have, web maps provide superior experience, providing a visual context for business data analysis, sales mapping, real estate or territory mapping, or demographics within and around target geographies.

Create Online Map Visualizations That Deliver Rapid Insights

We've designed this tutorial to help achieve an insightful interactive web map that can be shared with anyone, on any device.

Online heat maps like the one you'll create in this tutorial are super accessible for end users - great web maps will work on any device, and can be easily shared and explored - unlike a spreadsheet which, simply due to screen size and complexity, really only work on desktops.

Combining two data sources together is called a join. You can perform joins on just about any two sources of data where the spreadsheet data contains at least one matching column that's also found in the spatial data.

Creating a map of state names, county names, county codes, zip codes, region names… in fact, any administrative area can be combined with spreadsheet data into a new spatial file, ready to make brilliant interactive web maps.

All that's required is that our spreadsheet contains at least once column with fields that match data so we can join our spatial data with the business intelligence and extract our spatial intelligence.

```
Spatial data + Business intelligence = Spatial Intelligence
```

Joins can be done quickly and easily using free tools, and you can apply this technique to many types of spatial data that is freely available online.

In this tutorial, we're going to map U.S unemployment rate by county, which could serve as a useful reference map in all sorts of business scenarios from franchise territories, store locations, or job creation services to support strengthening communities with high unemployment.

In this tutorial, you'll learn the following:

- ⟶ `Understanding the data types`
- ⟶ `Loading spatial data in QGIS`
- ⟶ `Loading spreadsheet data in QGIS`
- ⟶ `Joining the spreadsheet with the spatial data`
- ⟶ `Exporting the joined dataset as a Shapefile`
- ⟶ `Mapping the dataset in an online GIS`

You won't need to know how to use QGIS to complete this tutorial and create a web map. We'll take you through everything you need to know, step by step. Online mapping is a great skill that can bring massive insights for your organization, so you'll find that once you've made your first map, you'll want to make more and uncover more hidden opportunities for your organization.

Tutorial resources

For this tutorial, you'll need just two things:

- ⟶ QGIS, which you can download for free at `qgis.org`
- ⟶ The sample data package we've prepared, which you can download here: `https://goo.gl/8Z2bYh`

Install QGIS and unzip the sample data package. If you don't have a compression utility to unzip the package, you can download WinRAR at `rarlab.com`.

For this tutorial, we're using QGIS 2.18. Earlier and later versions follow the same flow, but some items may be in different places in the interface.

1. Understanding the Data

The sample data contains five files:

```
data.csv counties.shp counties.dbf counties.prj counties.shx
```

`data.csv` is tabular data of U.S unemployment by county, and it contains two columns: “fips” (county code), and “rate”.

This CSV doesn’t contain any geographic coordinates so cannot be mapped on its own. This is the kind of data you might have in your own spreadsheets.

The remaining four files, `counties.[shp/dbf/prj/shx]` contain the county boundaries for the contiguous United States. Collectively, these files are known as a Shapefile. Shapefile is the de facto standard for sharing geographic data, and can be read by most mapping applications, including Mango.

You may come across other map data formats such as KML, Tab Files or GeoJSON, but Shapefile’s are the most common and we will focus on those for this tutorial.

SOURCING OTHER SPATIAL DATA

As discussed in chapter 3, Shapefiles with administrative boundaries for most countries and regions worldwide can be downloaded for free online, often via government data portals such as the Census Bureau. To find this Shapefile, we could just Google for “US county Shapefile” and we’ll find various sources for the data.

When sourcing spatial data to join with your data, it’s important to remember that your data needs to contain at least one matching region identifier. For example, a shapefile of census data might contain county `fips` codes, or it might contain county names. If your spreadsheet has county names, but the spatial file contains fips codes, you’ll need to create a new column in your spreadsheet and insert the correct fips code for each record.

Disclaimer: The unemployment data contained in the supplied CSV is **not guaranteed to be accurate**. Always ensure you validate any data you source online for its accuracy and recency.

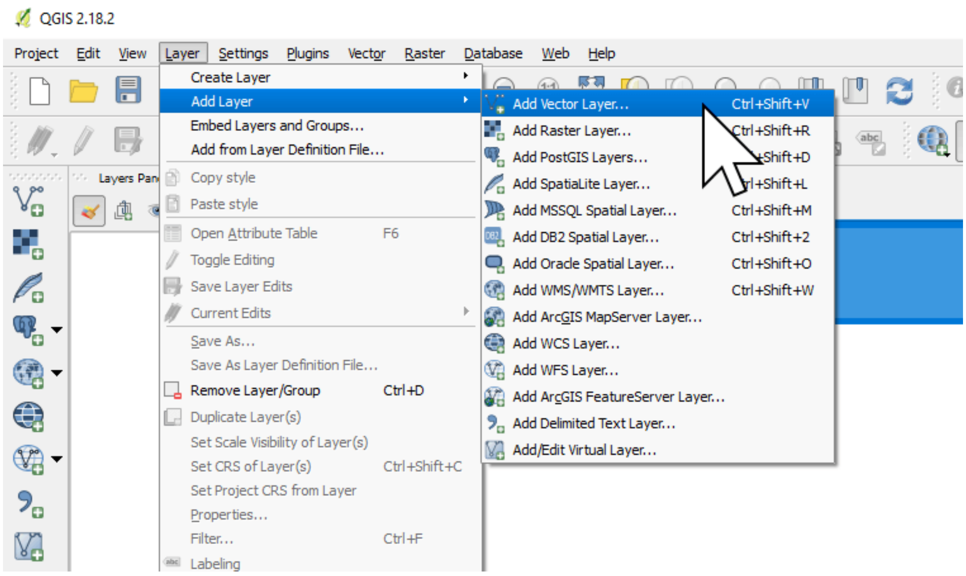
Pro Tip: Administrative boundaries can change, so it's always wise to check the recency of Shapefiles to ensure the data contains both the most current geographic information, and the most up-to-date ancillary data, such as census data, if that is required for your visualizations.

2. Loading spatial data into QGIS

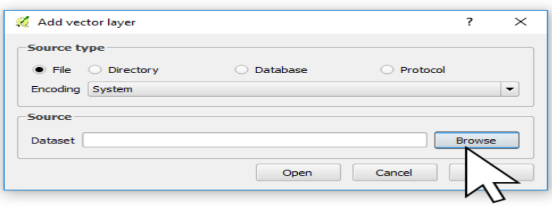
Now that we have our unemployment data and a Shapefile of U.S counties, we need to join these two files together. For this we are going to be using QGIS which is a very popular free and open source desktop GIS program that can be downloaded from here.

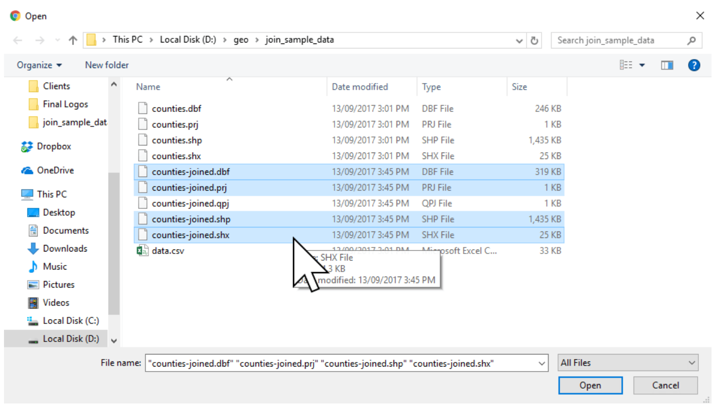
⟶ Open QGIS

⟶ From the menu bar choose Layer → Add Layer → Add Vector Layer.

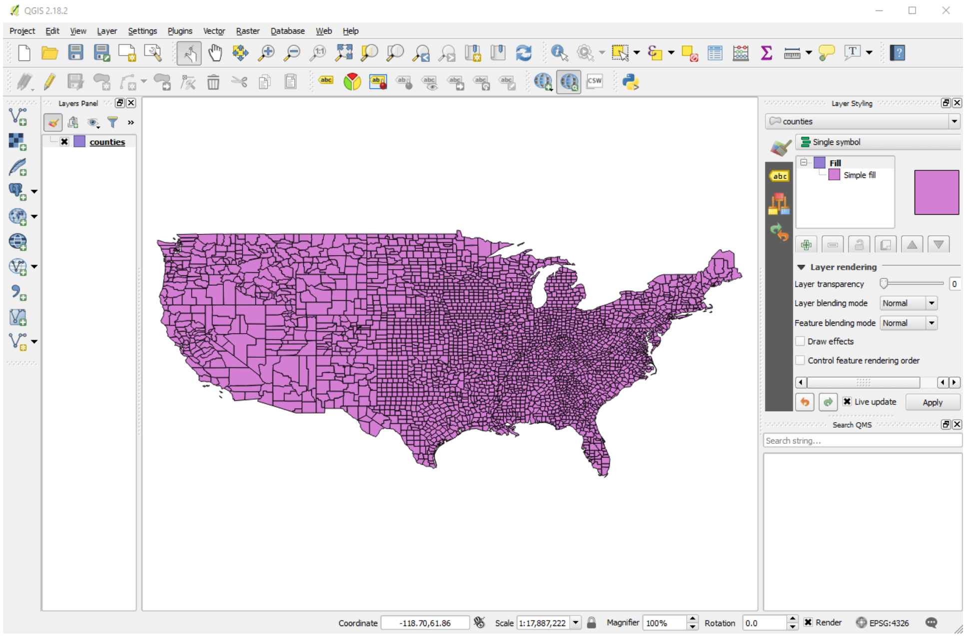


⟶ Navigate to the sample join data and select counties.shp.

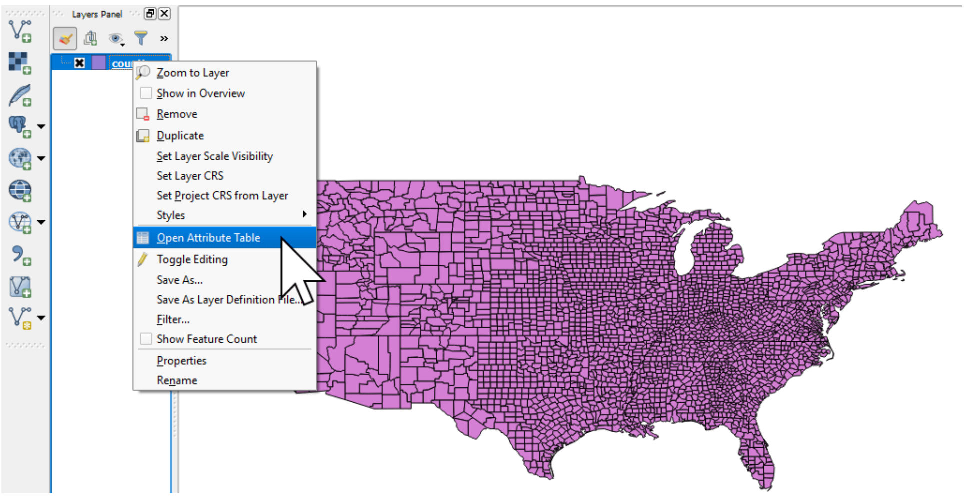




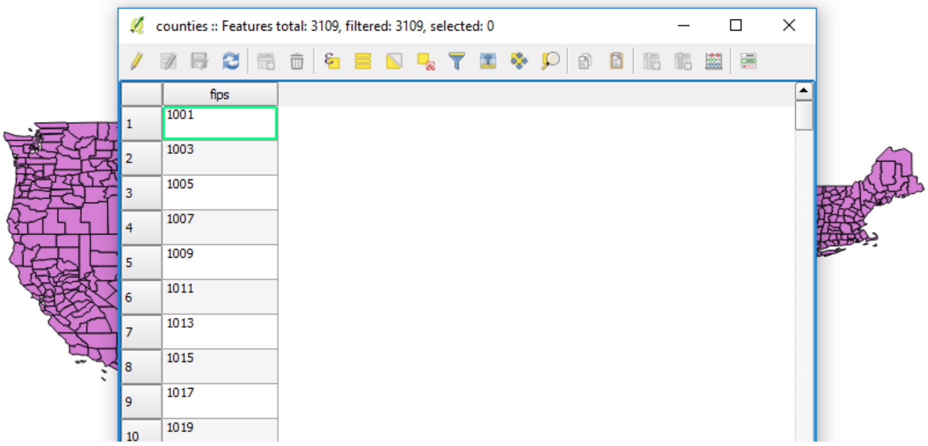
You will now see a map on screen showing all the counties of the contiguous US.



→ Right click on "counties" in the left hand layer panel and choose "Open Attribute Table".



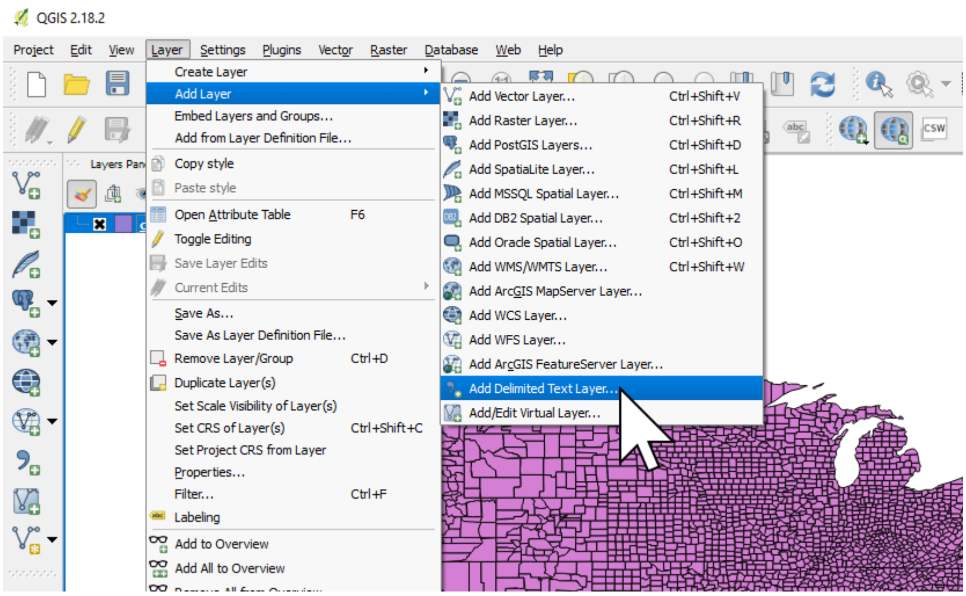
We will be able to see the information contained in the attribute table for the counties Shapefile. As you can see each county only has one piece of data and that's the fips code - a unique county ID.



3. Loading the spreadsheet data in QGIS

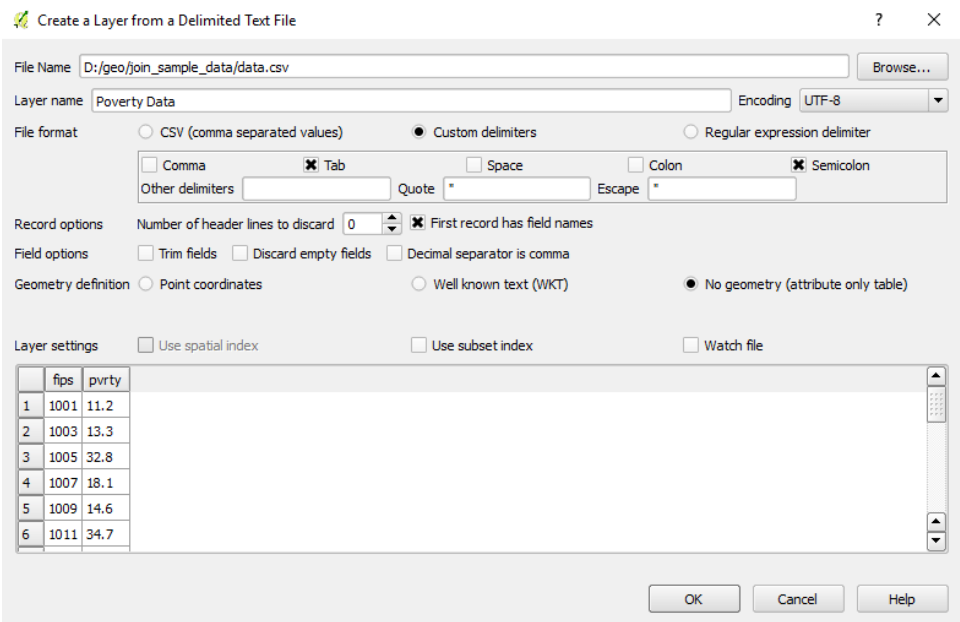
Next we need to import our CSV.

⟶ From the menu, select Layer → Add Layer → Add Delimited Text Layer



⟶ Click `Browse` and select the file `data.csv` from the downloaded sample data.

⟶ Set the Layer Name to `Poverty Data`, and set the File format to `Custom delimiters` and select `Tab`.



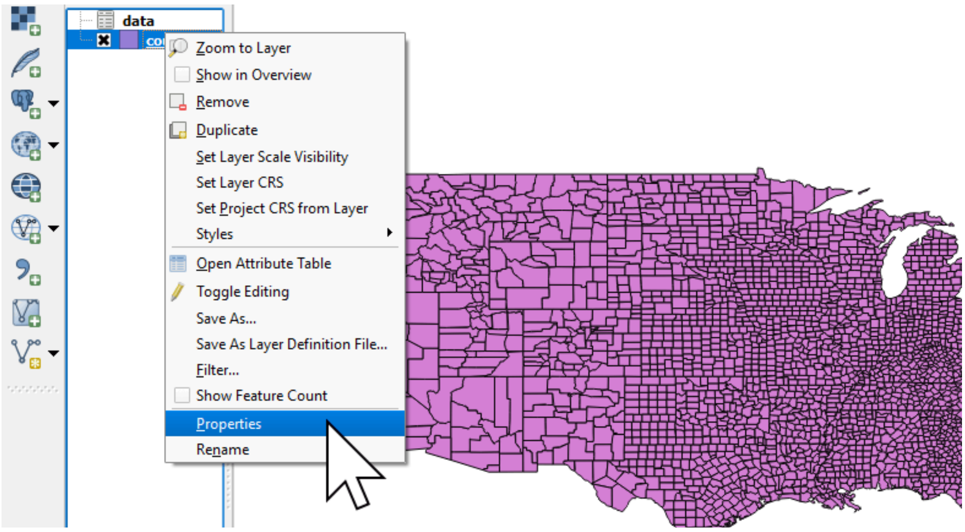
You should see a preview of the table with two columns: fips and pvrty.

⟶ Finally, set the Geometry definition option to `No Geometry` then press `OK`.

4. Joining the Shapefile and the Spreadsheet

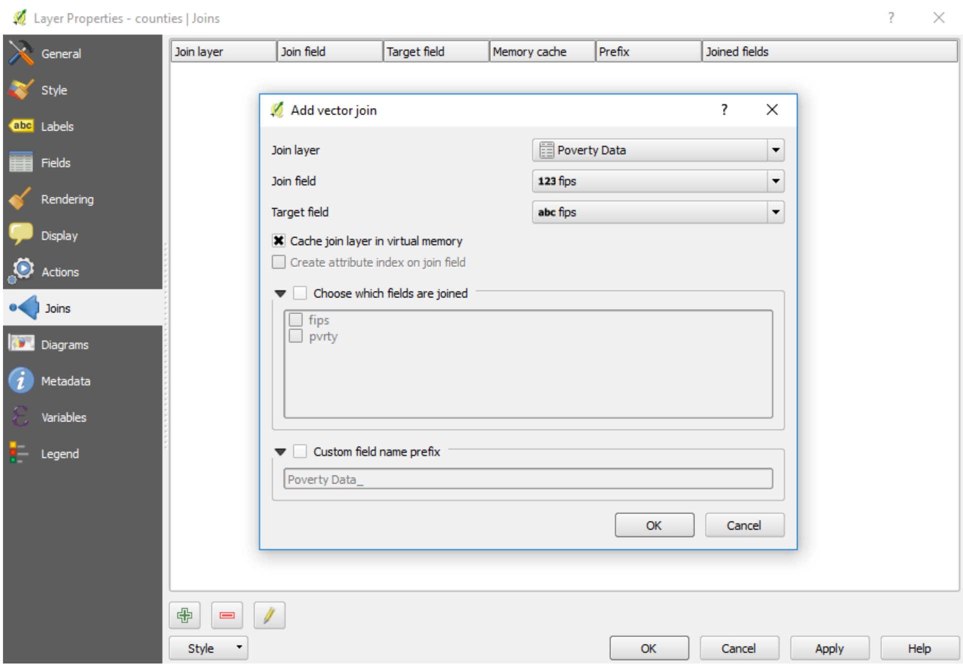
Now it's time to "join" the datasets together.

⟶ Right click on `"counties"` in the `Layers Panel` on the left hand side and choose `Properties`.



⟶ Go to the `Joins` tab and click the `green plus icon` at the bottom.

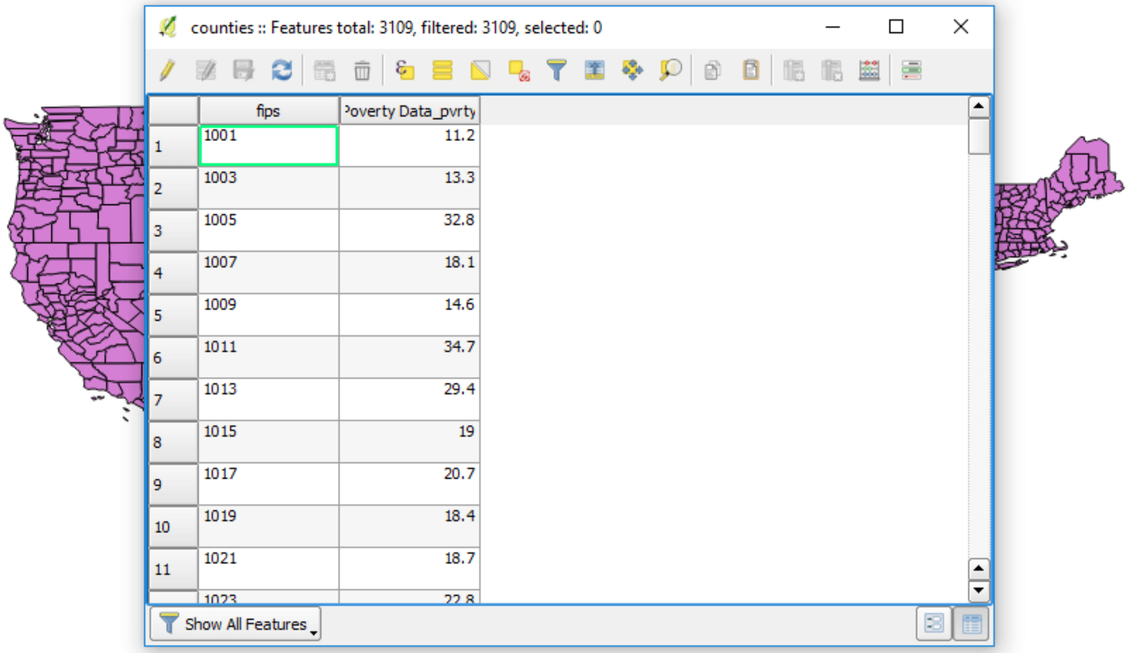
To join datasets, we need to use a property which is unique and present in both datasets, in this case it's the `"fips"` code column in the Shapefile and the `"fips"` code column in the CSV.



When we press OK, QGIS will match the records using the fips code and then add the additional columns from our CSV to the attribute table of the Shapefile.

Pro Tip: For this example, we're joining the records using the fips code, but you could just as easily join records using any column that has unique values, such as state name or zip code. Trying to join on columns where the values aren't all unique will cause data errors.

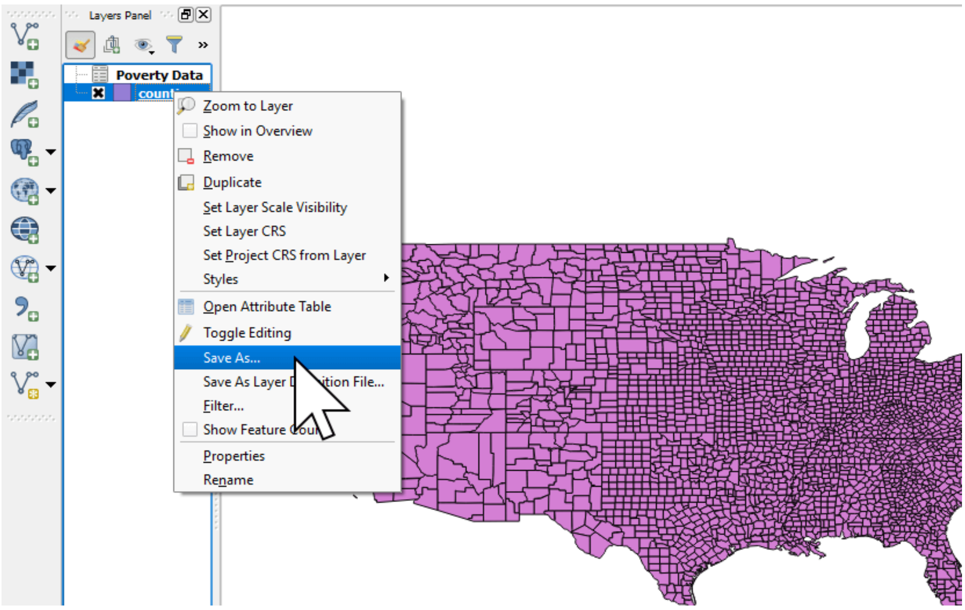
Once complete we can right click on counties in the left hand layer menu and select "View Attribute Table", you will now see that a new column has been added called "`data_pvrty`" that contains the unemployment rate for each county.

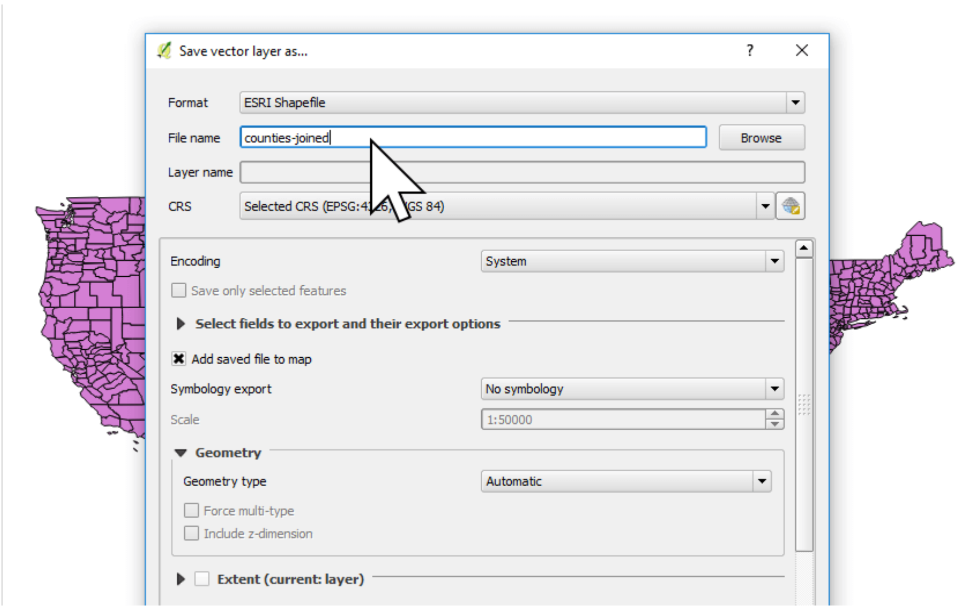


→ Now, close the attribute table.

5. Exporting the joined dataset as a Shapefile

→ Right click again on counties and select “Save As” to save the newly joined dataset. We now have a Shapefile that can be used in any mapping application to visualize the unemployment rate of counties.





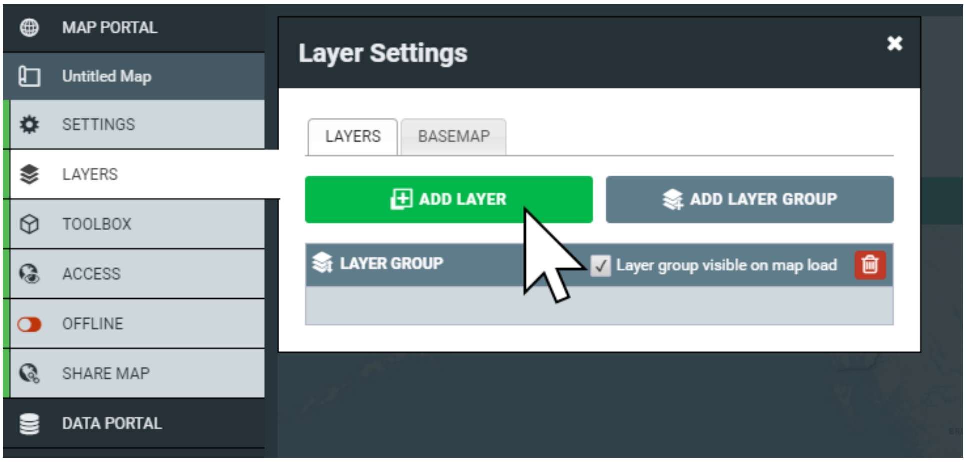
6. Creating our Heatmap

Now that you have your Shapefile let's upload it to Mango and take a look. If you don't yet have an account you can sign up for free at mangomap.com.

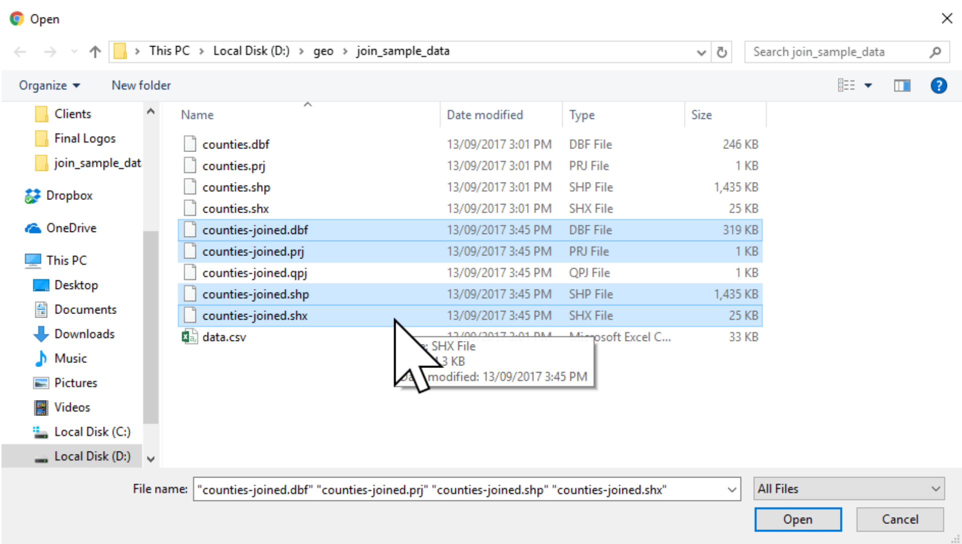
Once you have logged into your map portal complete the following steps:

⟶ Press the "CREATE NEW MAP" button the admin sidebar

⟶ Press "Add Layer → Upload Data"

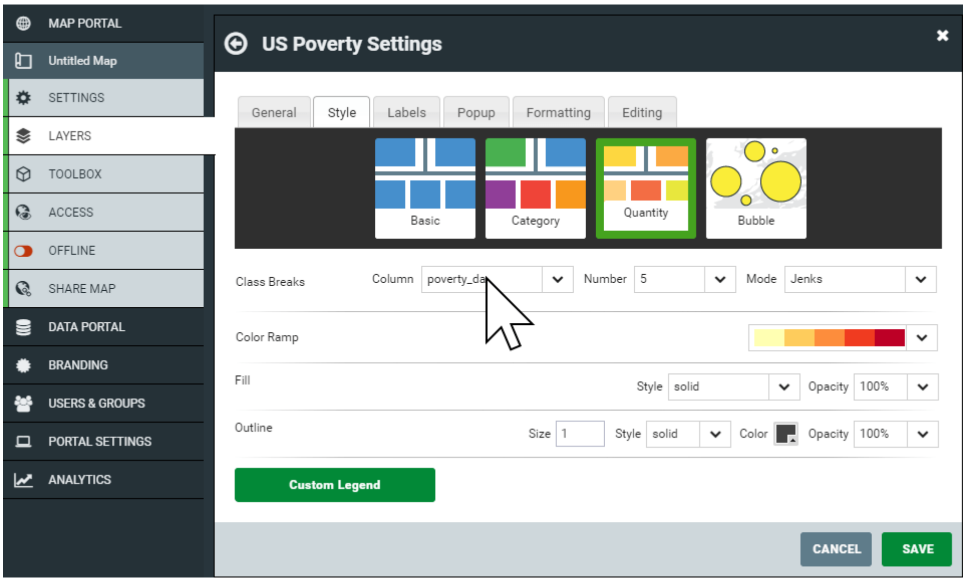


⟶ Select the four files of the Shapefile we made.

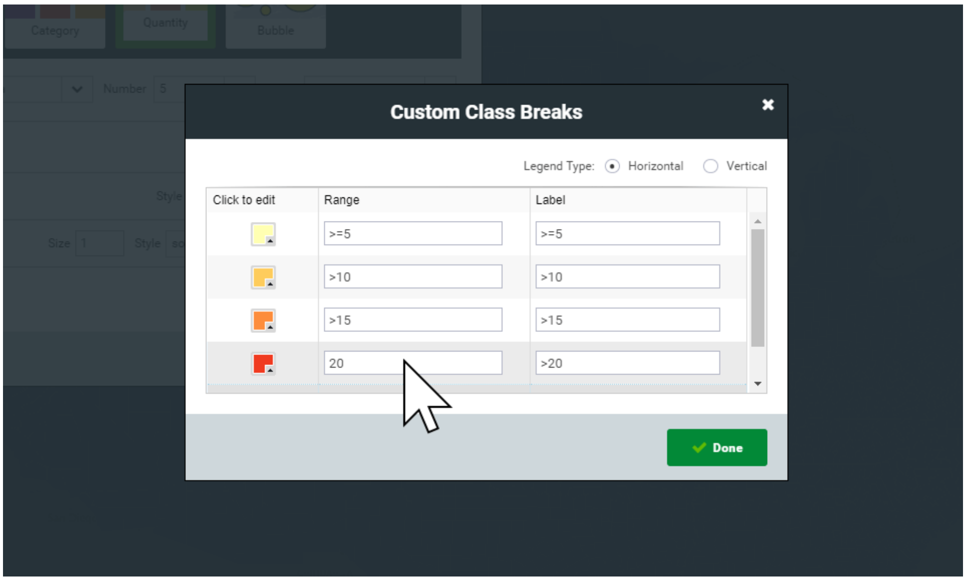


⟶ Once the upload finishes choose "Quantity" in the Layer Style panel.

⟶ In the Class Breaks row, select “data_pvrty” from the Column dropdown, and select “5” as the Number.



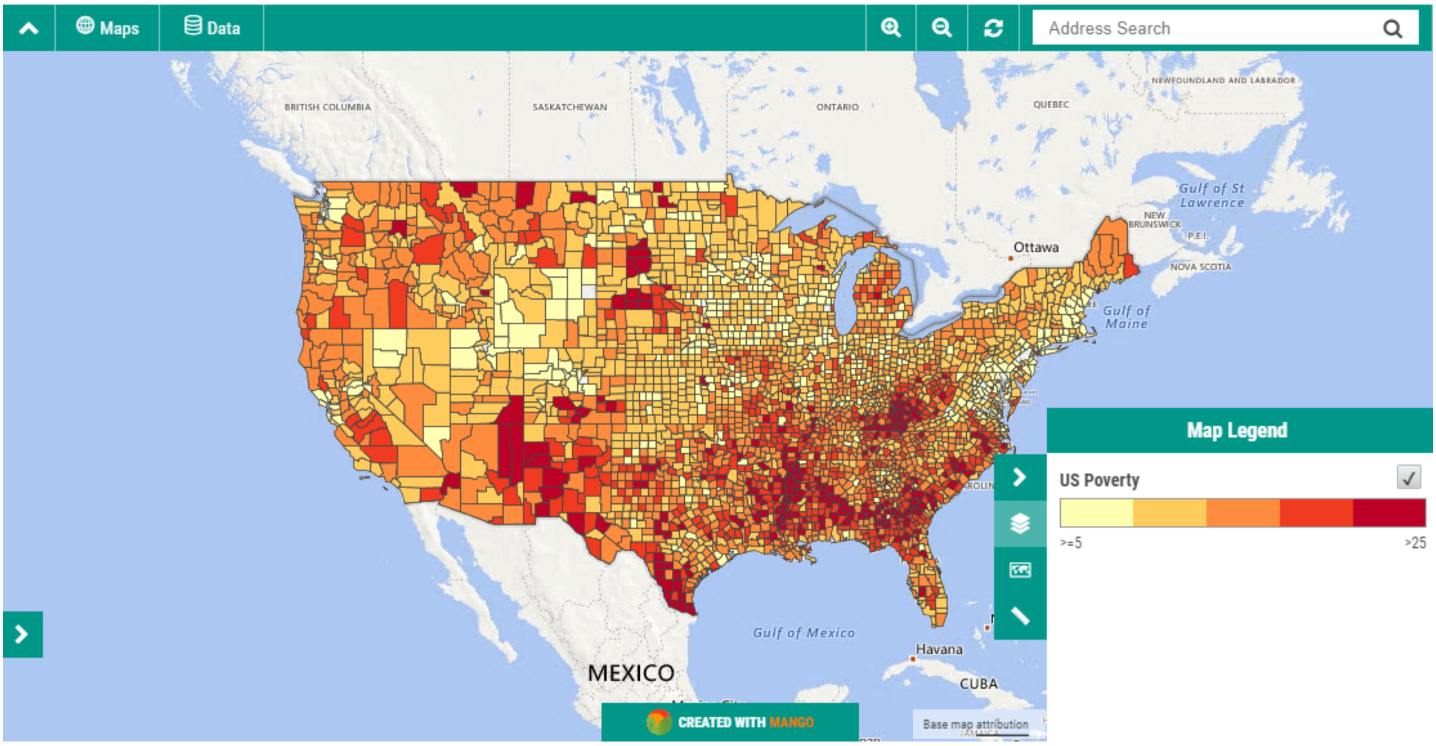
⟶ Press the green Custom Legend button at the bottom of the panel, and update the values in the Range column in the table to 5, 10, 15, 20 and 25. You’ll see that the labels will update accordingly.



⟶ Press Done, then hit Save on the Style panel.

⟶ Admire your quantity map

You will now be able to see your heat map showing U.S counties by unemployment. The darker the red, the higher the unemployment rate is in that county. The graduated colors allow us to quickly identify clusters of unemployment throughout the country on a very granular level.



All done! Easier than you expected, right?

Click on a county, and you'll see a popup of the attributes for that county. Depending on the data you've joined, you could transform unemployment and population figures into dynamic charts using the attributes of each county, or add links, videos, or even images. You can even create a calculator tool that allows users to input a number and perform calculations based on attribute values for each county.

Adjusting the visualization

Take some time to experiment with different values in the layer settings panel to see what happens to the map. Maps make it easy to draw conclusions based on bold visualizations, but you will find that minor changes to quantity settings will reveal vastly different visualizations.

Like music in a film can elevate underlying emotion or tension, cartographic design plays a large role in shaping the viewer's understanding of a map. Turn the sound down on that blockbuster, and it probably doesn't seem so dramatic.

So too, it's important to remember that maps can quite easily lead to conclusions or assumptions that might not be accurate on close investigation of the underlying data, and can reveal the map maker's bias towards a certain conclusion.

This is often desirable when creating focused story maps - after all, we usually create maps to tell a specific story, or lead viewers to a certain conclusion. Data truth is critical to your success, so extracting inaccurate conclusions won't help your business, so it's important to remain vigilant of inherent biases or trying to make a map that looks good, but fails to communicate the truth in the underlying data.

Now imagine other data you've got that could be overlaid on top of this unemployment map. Perhaps your store locations, your sales territories, or your customers. If you have addresses in a spreadsheet that you want to visualize on top of this map, head over to our tutorial on turning spreadsheet data into map ready coordinates.

Create a Web Map from Addresses in a Spreadsheet

Spreadsheets are a fantastic way to store and share data. It's ubiquitous, and can be analyzed and transformed into revealing charts and tables. However, there's only so much we can extract out of a spreadsheet for analysis while remaining concise.

Deep analysis via spreadsheets is too often obscured by the sheer amount of data. While digging into hundreds or thousands of columns of tabular data, it's too easy miss the forest for the trees.

Providing a thousand-foot view for users to drill down and extract insights often requires pretty serious excel skills and endless sheets of pivot table wizardry, collated content and graphed goodies, presenting the original data as various key takeaways.

Once you've created your cache of analyses, sharing it with colleagues is simple enough - until it's not.

Ever tried viewing a 20 sheet Excel file on a mobile? It's not fun, and it's not conducive to understanding any of the findings presented in the data, no matter how good those charts looked on desktop.

Worse still, if you're also handling geographic data in your spreadsheets, it's likely you'll be missing some of the most important findings. The insights that can be illuminated when spreadsheet data is mapped provides the bridge for that thousand-foot view: the forest that's been

in front of us the whole time, obscured by the minutiae.

WEB MAPS ARE SUPERIOR TO SPREADSHEETS

Web map visualizations (also known as Online GIS) and mapping spreadsheet data to uncover the meaning of your location data delivers powerful outcomes for anyone working with spreadsheets. Data and trend discovery, visualizing geographic hotspots and opportunities becomes plainly apparent once you have created a map from spreadsheet data.

Online maps are superior to spreadsheets in another fundamental way: accessibility. Web maps are made for mobile, and look and feel just as visually impactful on mobile or tablet as they do on desktop.

The key advantage of visualizing your data on maps, is that you can analyze your empire alongside other spatial data, like income or employment data from a national census, or understand your regional logistics needs by overlaying county and state spatial data that covers just about anything you can think of - flood and natural disasters, transport and rail networks.

No matter what type of data you have, web maps provide superior experience, provide spatial context and insights for business data analysis, sales mapping, real estate and territory mapping, and store performance - really, anything you've spent time and money on collecting in a spreadsheet; the possibilities are endless.

CREATING MAPS FROM SPREADSHEETS

There are a number of options for mapping spreadsheet data - some free, some premium. Most deliver adequate results, but all too often your data won't quite fit into the preset systems, so it's common to find records missing from your mapped data.

An obvious choice might be Google maps. You can make simple maps on free services like Google maps, but then Google's got all your data, and all your underlying data is exposed to the public when you need to share your maps. It also turns out that unless you want to pay them the big bucks—and we're talking proper big bucks and complex credit usage tables that obfuscate the actual price—you can't do much else with your data on Google maps besides just adding a few colored markers and maybe embedding it in your website.

That's fine for a map of your outlets, and maybe that's all you need. If so, head on over to Google's map maker and go wild.

Bye!

Are they gone? Ok, for those that want more—you and I, here's where it gets good.

SO HOW CAN I EXTRACT INSIGHTS FROM MY SPREADSHEET DATA?

Firstly, in order to display your address data on a map, we'll need to generate the geographic coordinates (latitude and longitude) of the location of each address in your spreadsheet.

We can do this through a process known as **geocoding**. Geocoding is the process of assigning geographic coordinates to a physical location.

In an ideal world, technology would just take care of all this for us, right? We've got self-driving cars, so why not self-mapping spreadsheets? One day, perhaps.

Unfortunately, spreadsheet applications aren't quite at Philip K. Dick-level self-consciousness just yet, so while we wait for that (terrifying) day, we need to prepare and process our own address lists into map-ready data ourselves.

THANKFULLY, IT'S REALLY QUITE SIMPLE (ROLL UP SLEEVES NOW).

It requires a few tools - free, open source tools that are readily available.

If you're able to create and manage your spreadsheets and databases of suppliers, stockists, customers, whatever you bread and butter is, then it won't stretch your abilities at all to learn how to turn all those valuable addresses into even more valuable coordinates.

Once you've got your coordinates, it's a piece of cake to then visualize them on a map, and extract those juicy insights that can help to energize and transform your business through spatial intelligence.

We've put together a foolproof tutorial that will walk you through all the steps required to convert your spreadsheet into a map by geocoding the addresses in your spreadsheet.

In this tutorial we'll cover the following:

- ⟶ `Installing QGIS`
- ⟶ `Installing MMQGIS`
- ⟶ `Converting your spreadsheet of addresses to CSV`
- ⟶ `Geocoding your addresses`
- ⟶ `Mapping your spreadsheet`

GATHER YOUR TOOLS

For this tutorial, you'll need two things before we begin:

- ⟶ `Your spreadsheet (or our example CSV) and a spread-sheet application (Excel, OpenOffice Calc, Google sheets, etc). Download it here:` `https://goo.gl/bcYGjG`

- ⟶ `Download QGIS - the best free open source desktop GIS. Download it at` `qgis.org`

If you don't already have QGIS installed, download it now. It's completely free, and is a great platform to continue learning about mapping and GIS.

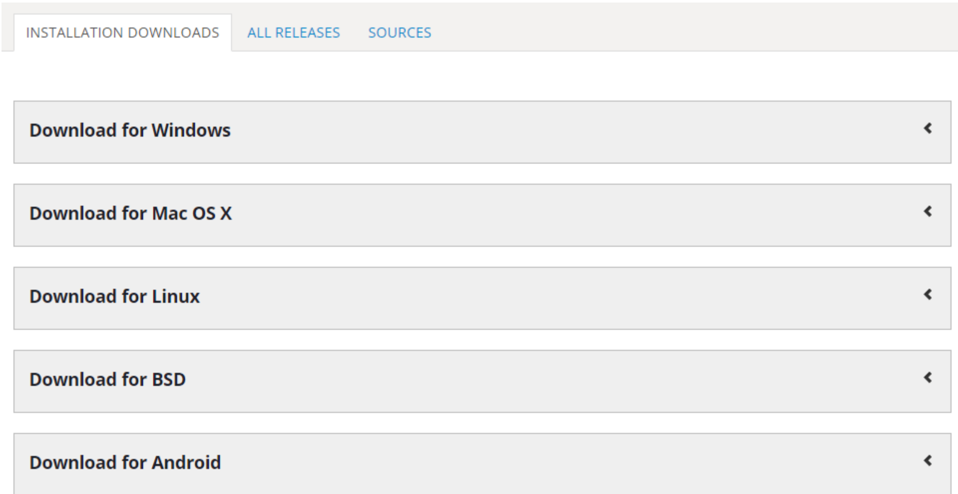
Don't worry, **you don't need to be a QGIS expert to complete this tutorial**. However, if you would like to learn more about it, then here is a great place to get started: `https://goo.gl/zi8qNb`

If you don't have any address data of your own, but would like to follow along, you can use the spreadsheet I'm using in this tutorial.

ℹ `For this tutorial, we're using QGIS 2.18, but it should` also work much the same on earlier and newer versions.

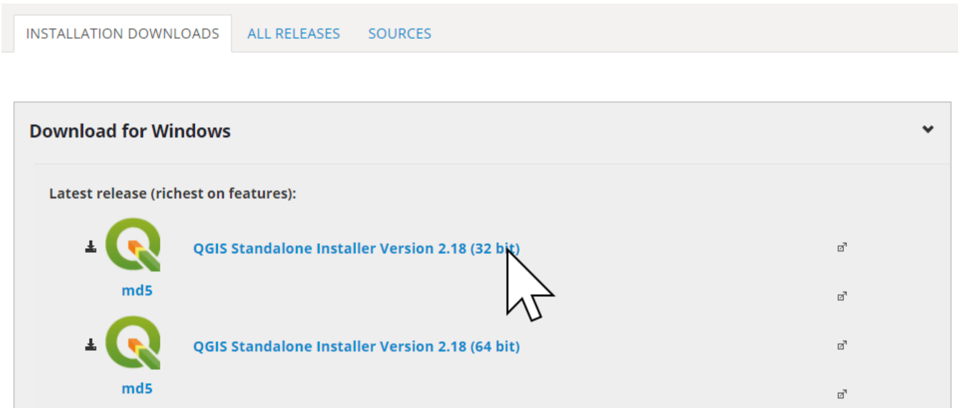
1. Installing QGIS

⟶ First things first, download and install QGIS at `qgis.org` if you haven't already.



Unlike a lot of desktop GIS, QGIS works on all major desktop operating systems, which makes it easily accessible to everyone.

⟶ Click on your operating system to see the available



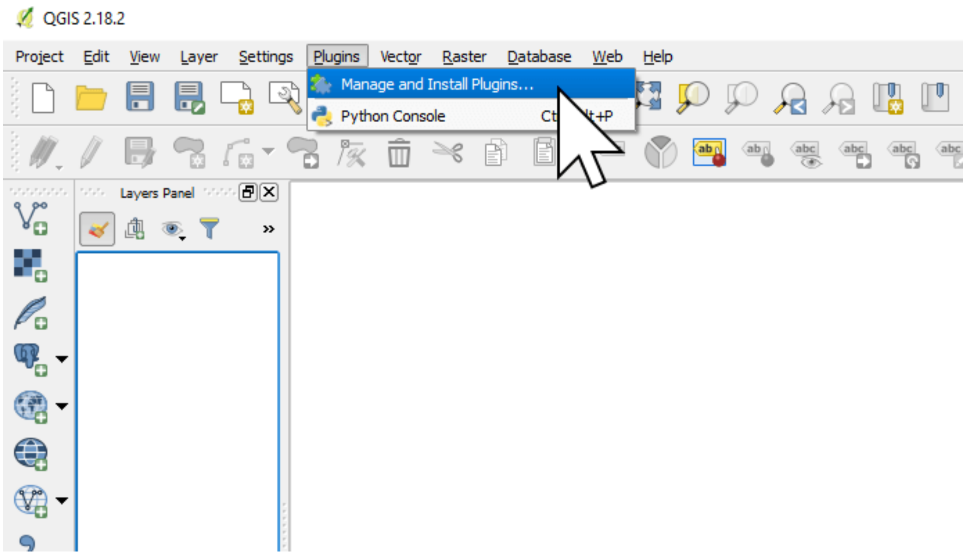
versions, and click on the link to start the download. In this tutorial, we'll be using version 2.18, but you can use any version that is compatible with your system.

Once the download has finished, install QGIS and move on to step #2.

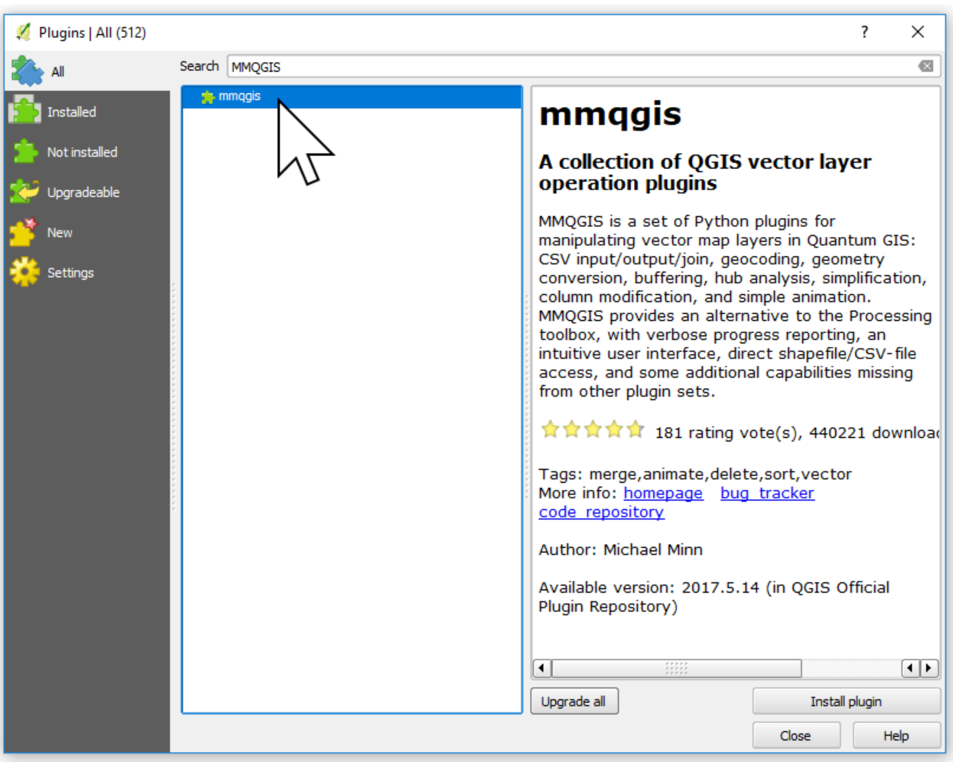
2. Install the MMQGIS Plugin

The first thing we need to do is install the MMQGIS plugin that adds geocoding capabilities to QGIS. We'll feed it the list of addresses, and it will output a list of shiny new geographic coordinates.

MMQGIS is included in the QGIS Plugin Repository so we can install it from directly within QGIS.

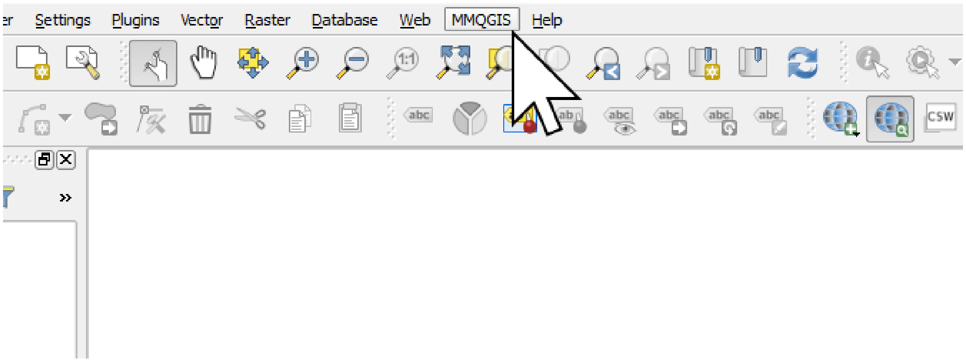


⟶ Open up the QGIS application to get started.



⟶ Choose "Plugins" from the menu bar and choose "Manage and Install plugins…"

⟶ Search for MMQGIS in the top search bar, select if from the results, and click "Install plugin".



⟶ Once the plugin is installed, close the Plugins window.

⟶ You should now see a new menu item called "MMQGIS" next to the Help menu.

3. Converting Your Spreadsheet of Addresses to CSV

Let's return to our spreadsheet for a moment. Before we can geocode the list of addresses, we'll need to check our spreadsheet to ensure we will get the best results from the geocoding.

Geocoding requires that the data is in CSV (comma separated values). A CSV file is essentially a spreadsheet, just without all the formatting and colors you'll find in Excel.

The example spreadsheet provided is already prepared as a CSV, so if you're using that you can jump straight to the next section.

Your spreadsheet with addresses must be broken into columns called "address", "city", "state" and "country". e.g.:

	A	B	C	D	E
1	Address	City	State	Zipcode	Name
2	1 Crossgates Mall Road	Albany	NY	12203	Apple Store Cross Gates
3	Duke Rd & Walden Ave	Buffalo	NY	14225	Apple Store Walden Galleria
4	630 Old Country Rd.	Garden City	NY	11530	Apple Store Roosevelt Field
5	160 Walt Whitman Rd.	Huntington Station	NY	11746	Apple Store Walt Whitman
6	9553 Carousel Center Drive	Syracuse	NY	13290	Apple Store Carousel
7	2655 Richmond Ave	Staten Island	NY	10314	Apple Store Staten Island
8	7979 Victor Road	Victor	NY	14564	Apple Store Eastview
9	1591 Palisades Center Drive	West Nyack	NY	10994	Apple Store Palisades
10	125 Westchester Ave.	White Plains	NY	10601	Apple Store The Westchester
11	103 Prince Street	New York	NY	10012	Apple Store SoHo

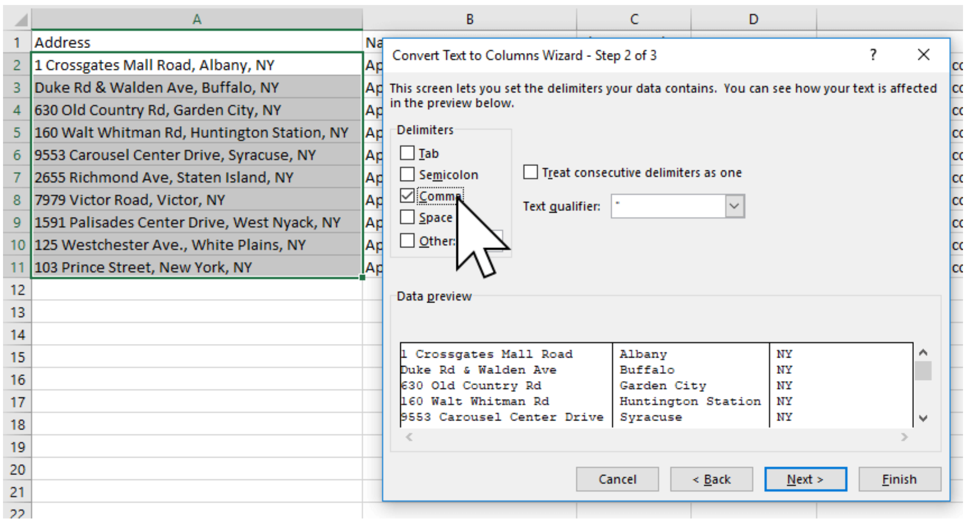
- Address: 630 Old Country Road
- City: Garden City
- State (or equivalent outside US): NY
- Country: USA

Most geocoders are quite smart, and will be able to work out that NY in

	A	B	C
1	Address	Name	Phone Number
2	1 Crossgates Mall Road, Albany, NY	Apple Store Cross Gates	(518) 869-3192
3	Duke Rd & Walden Ave, Buffalo, NY	Apple Store Walden Galleria	(716) 685-2762
4	630 Old Country Rd, Garden City, NY	Apple Store Roosevelt Field	(516) 248-3347
5	160 Walt Whitman Rd, Huntington Station, NY	Apple Store Walt Whitman	(631) 425-1563
6	9553 Carousel Center Drive, Syracuse, NY	Apple Store Carousel	(315) 422-8484
7	2655 Richmond Ave, Staten Island, NY	Apple Store Staten Island	(718) 477-4180
8	7979 Victor Road, Victor, NY	Apple Store Eastview	(585) 421-3030
9	1591 Palisades Center Drive, West Nyack, NY	Apple Store Palisades	(845) 353-6756
10	125 Westchester Ave., White Plains, NY	Apple Store The Westchester	(914) 428-1877
11	103 Prince Street, New York, NY	Apple Store SoHo	(212) 226-3126

the state column is New York, so don't worry if your data is in a slightly different format to that shown above.

If all of the address is contained in a single column then you will need to



split the address into the separate fields - address, city, state.

Addresses in a single column need to be separated into individual fields before geocoding.

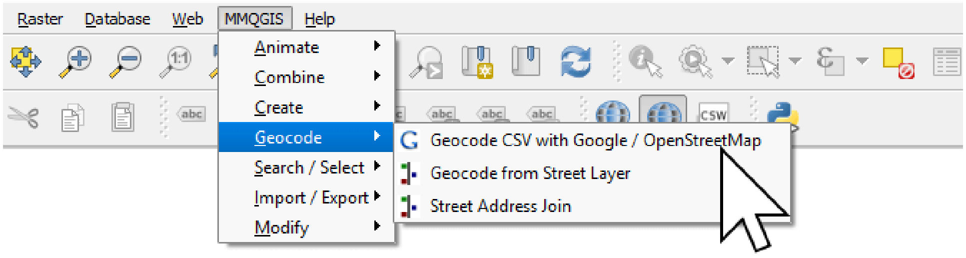
Excel's Convert Text to Columns Wizard lets you quickly split your address into individual columns.

If you're using your own data, you can convert your spreadsheet to a CSV file in Excel or any other spreadsheet program by opening the spreadsheet and pressing "Save As" then choosing "CSV (Comma delimited)" from the "Save as type" option.

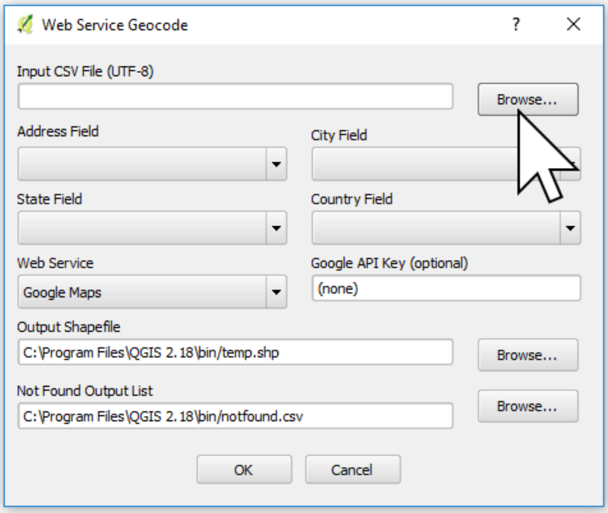
4. Geocoding The Addresses

Now that we have our data in the correct format it's time to geocode it in QGIS.

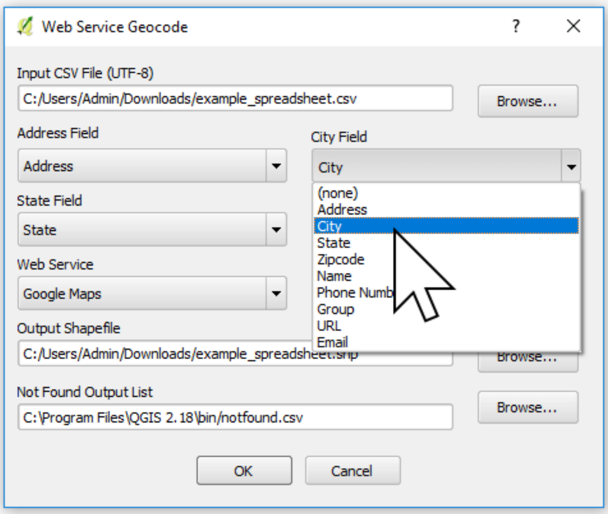
⟶ In the menu bar go to: MMQGIS → Geocode → Geocode CSV with Google / OpenStreetMap.



⟶ Click `Browse...` and select your CSV



⟶ Once you've selected your CSV file, we'll need to tell the Geocoder which columns represent the individual address components - the address, city, state and country.



MMQGIS will do it's best to detect the appropriate column names (and it's best is often very good), but if you use obscure column names or a different language and it doesn't get it all correct, just click on the dropdown menu and choose the correct column in your CSV.

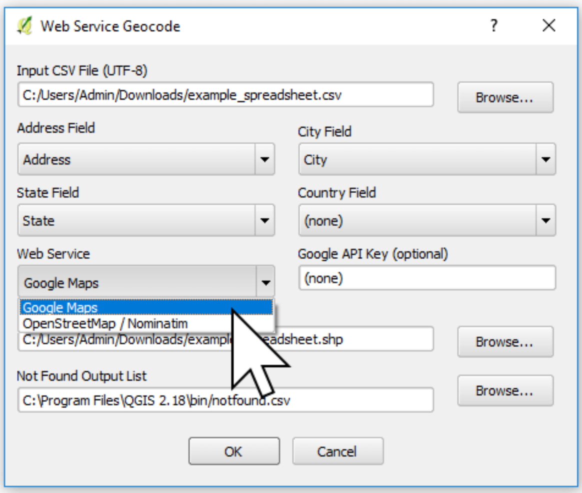
We also have a choice of web services to perform the geocoding. We can choose from **Google** or **OpenStreetMap/Nominatim**.

OpenStreetMap has no overall limit[1]. However, the Nomanitim geocoder is designed to support OSM mapping initiatives, and due to the community (and donation) driven nature of OSM, it's not a suitable choice for ongoing business use, without chipping in to the OSM intiative.

> OSM's Nominatim service is mainly there to power the search bar on openstreetmap.org. We are in principle happy for the public API to be used by external users for creative and unexpected uses. However, be aware that the service runs on donated servers and has a very limited capacity. We therefore ask you to limit your use and adhere to this usage policy.

OSM's Nominatim usage policy

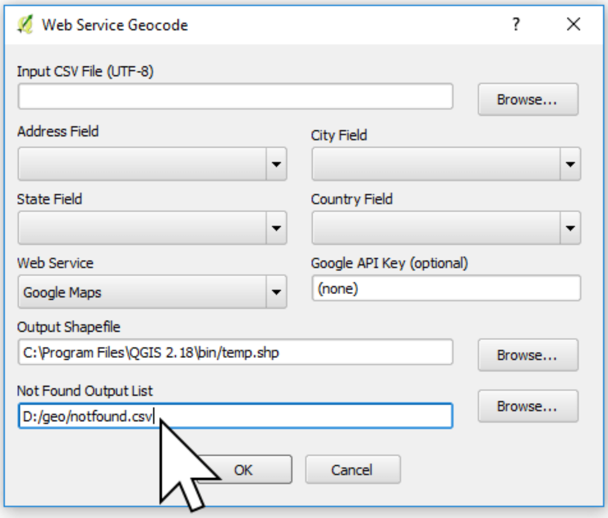
Google tends to provide the best results but only allows you to geocode 2500 records per day[2] without a premium API key.



Pro Tip: If your CSV has more than 2500 records and you wish to use Google you can break the spreadsheet into separate files each with 2500 records and process one each day. Once complete join the output files back together using MMQGIS's merge layer function.

1 https://operations.osmfoundation.org/policies/nominatim
2 https://developers.google.com/maps/documentation/geocoding/usage-limits

→ Now make sure the `Output Shapefile` path and the `Not Found Output List` path are valid, and you have permission to write to that location. To change the path, just click on the Browse button and choose a preferred location.



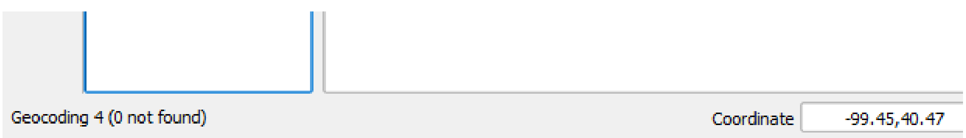
Geocoded results will be stored as a Shapefile which, despite it's singular name, is actually a collection of at least four files (.shp, .shx, .dbf, .prj). The Shapefile is the de facto standard for storing GIS data and nearly all mapping applications support it, including Mango.

A second file will be stored that contains a CSV of any records that could not be successfully geocoded. These records usually have an incomplete address, spelling mistakes, or are in remote areas not covered by the geocoder.

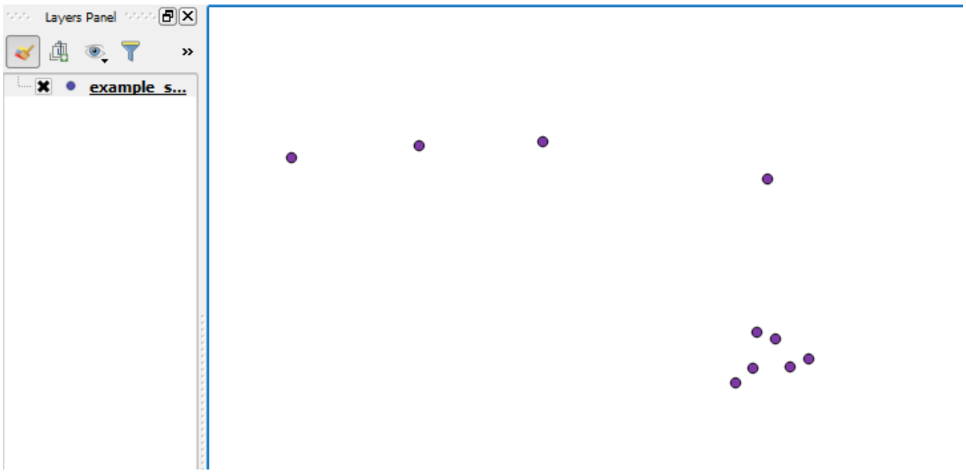
! Important: On Windows, the default output path for the Not Found Output List is often in a system folder that MMQGIS can't write to, and you will receive an error if you use that value. Update the Not Found Output List path to a path you have permissions to access (like My Documents or Downloads) before continuing.

→ We're all good to go. Hit `OK` to proceed with the Geocoding.

→ While processing, you'll see the progress in the bottom left corner of QGIS



⟶ When the geocoding is complete, your address points will appear in QGIS as circle markers.



If you don't have a basemap turned on in QGIS, you'll just see floating points against white - this is normal and means your geocoding is done and your geocoded coordinates are now in a shapefile in the `Output Location` you set at the previous step.

⟶ Go to the folder that contained your notfound.csv to see if any addresses weren't located. In this case, an empty CSV is a good result!

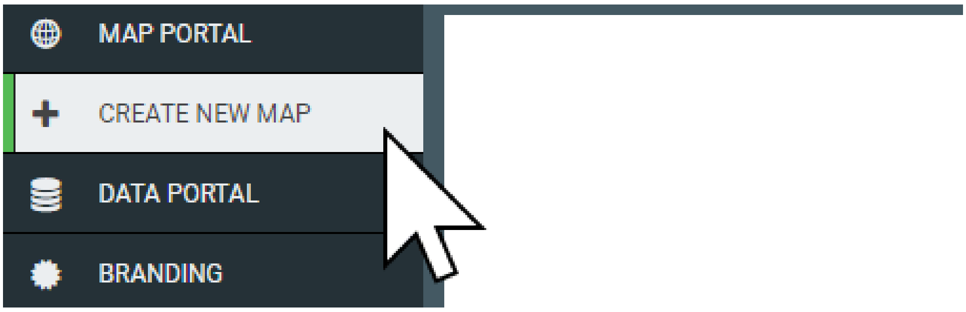
	A	B	C	D	E	F	G	H	I
1	Address	City	State	Zipcode	Name	Phone Nu	Group	URL	Email
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									

5. Mapping Your Addresses

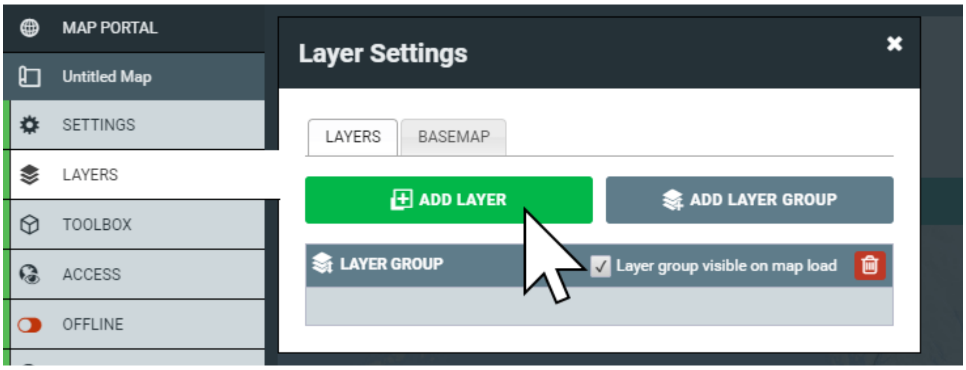
Now that you have your Shapefile, let's upload it to Mango and take a look. If you don't yet have an account you can sign up for a 30-day free trial at `mangomap.com/sign-up`.

Once you have signed in to your account, we're ready to make a map!

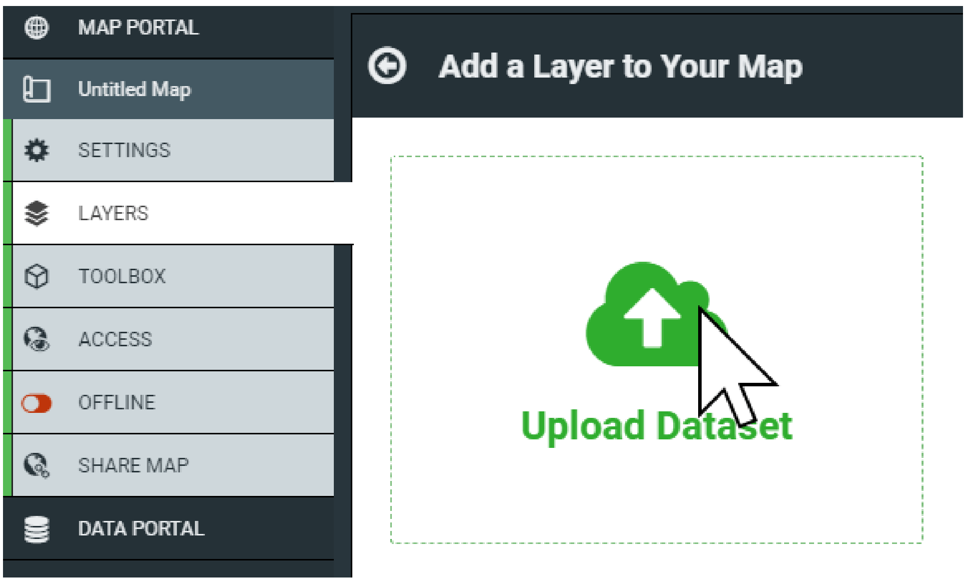
⟶ Press the `CREATE NEW MAP` button in your administration sidebar.



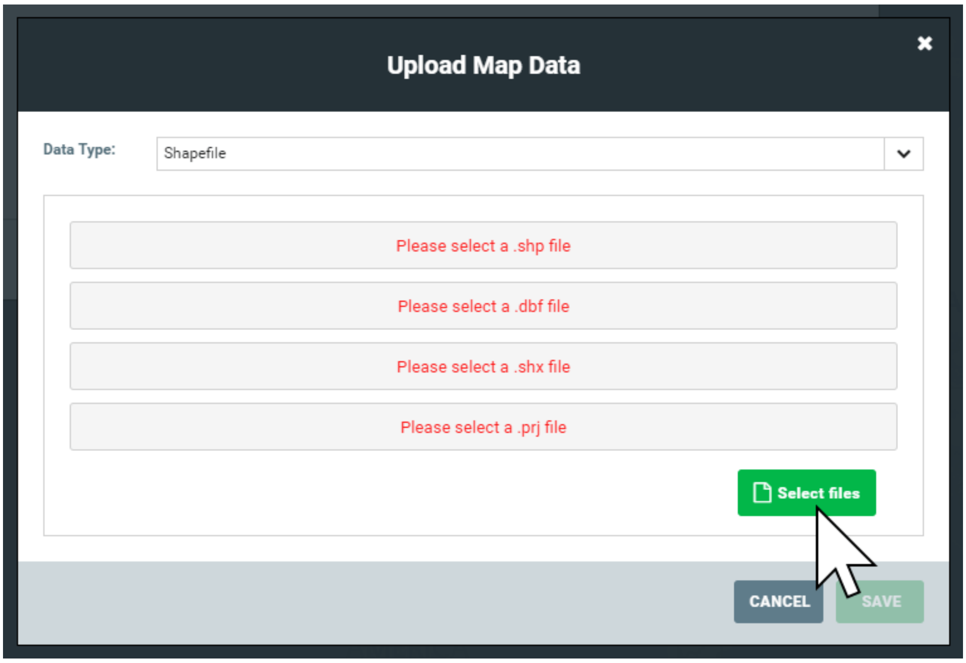
⟶ When your map is ready, click `LAYERS`, then click on the `Add Layer` button.



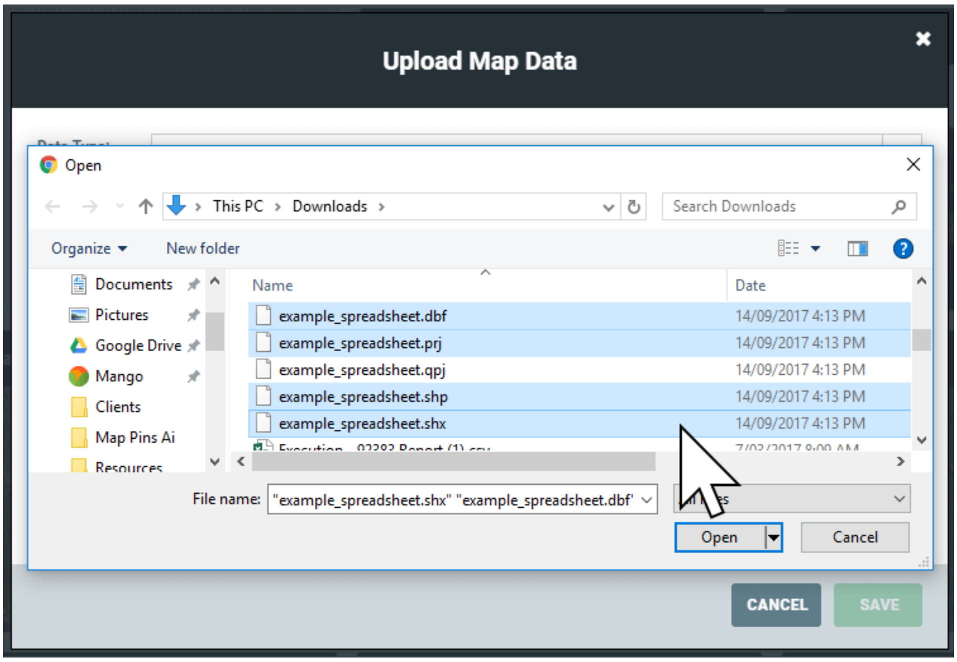
⟶ Now click on `Upload Dataset`



⟶ Here you can select the type of data you want to upload. We've got a Shapefile, so we don't need to change anything. Click on `Select files`

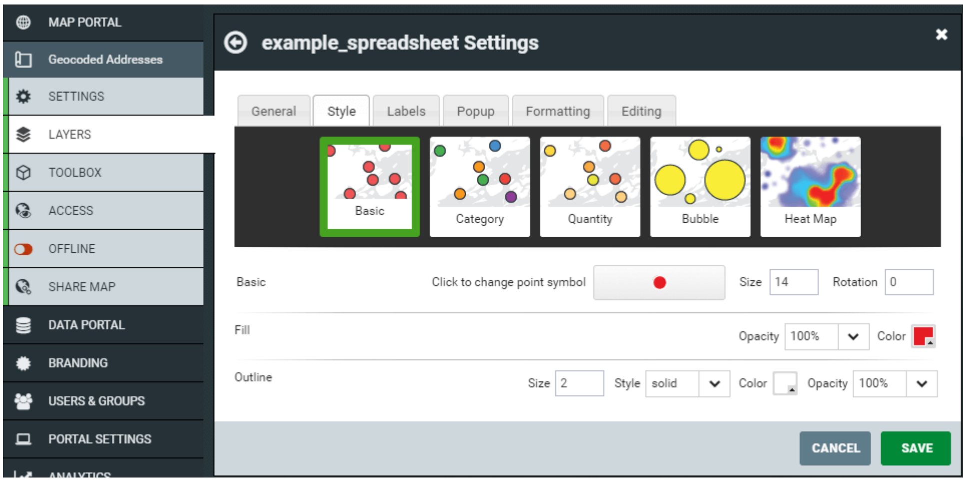


⟶ Navigate to the output location and select the four files of the Shapefile generated by the geocoder and upload them to Mango. To select multiple files at once, press and hold Ctrl and clicking each file (⌘ + click for Mac).

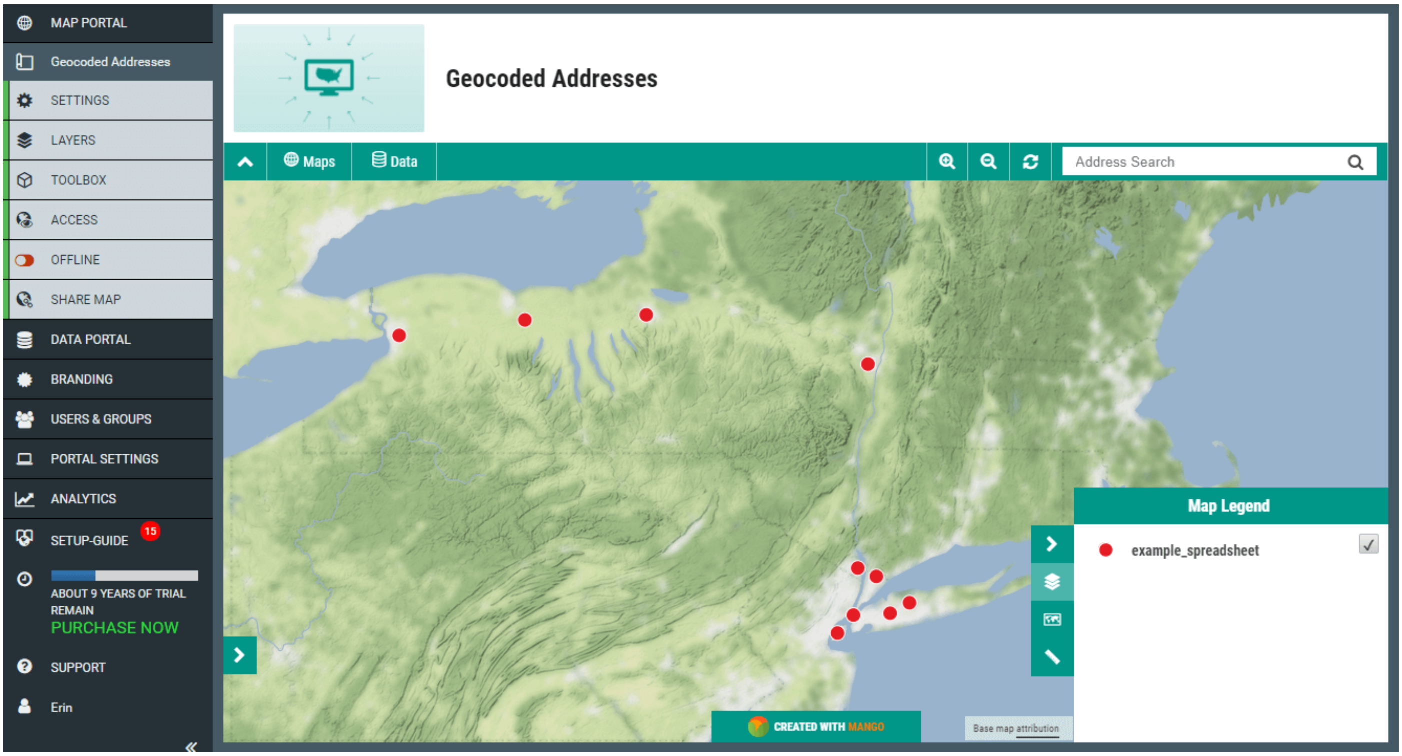


⟶ Once the upload is complete, press `SAVE` and Mango will process your data.

→ `When the processing is complete, you'll be shown the Layer Settings panel. Here, simply select a color for your points, and hit Save.`



→ `Admire your geocoded locations on your map`



It's that simple!

Now you've got your addresses on a web map, you can add supplementary map layers on top or below to provide context and extra information, or add tools like Mango's Proximity Analysis tool to perform spatial queries to find even deeper insights such as *Cities within 100 mi of a dealership with population above 50,000.*

How To Remove Unwanted Regions From Your Map Data

So you've sourced the perfect shapefile covering US Counties with demographic and population data, and some neat geographic information like the square mile area of land and water for each county. Now, you're ready to make a web map of sales data for Texas or add the data to your online GIS.

Problem is, like most map data sourced from third parties, this Shapefile contains features and information outside of your area of interest - in fact, it contains geometry and attributes for every single US county.

If you've ever downloaded a shapefile from the US Census Bureau's website, you'll find this is a common problem.

Short of making a strange map with lots of plain colored areas and then some nice shaded Texan counties, what can we do to trim the fat, so to speak, and filter out all the information we don't need while keeping all the good data we do need?

The task here is to remove all of the features from the dataset that you don't need. In a GIS this is done by selecting our desired features using an expression or query, and deleting the rest.

"Features? Expressions? Queries?!! I just want to map my sales data!"

I hear you.

Sounds like it's going to be complicated, I know, but I'll make this painless. You got this.

If you've found spatial data and are looking to filter out the unwanted data, I'm going to assume you know a little bit about mapping. After all, you found this tutorial, so you're on third base and the home stretch is in sight.

So: Expressions. An expression in GIS is much like a formula in Excel.

Just like we could select a range in Excel with a simple `=A1:A10`, in a GIS we can create an expression to select a subset of features. For example, we can isolate the unwanted areas in our US County Shapefile with a simple expression:

```
"STATE" IS NOT 'Texas'
```

In QGIS, this expression would create a selection of all counties in the dataset, except for those with a `STATE` attribute value of Texas. Once selected, removing them is as simple as hitting the delete key.

Let's walk through it. In this tutorial, we'll cover the following:

⟶ Importing the spatial data into QGIS

⟶ Creating a selection to delete information outside our area of interest

⟶ Exporting the filtered data as a Shapefile

The Data

For those that don't have any data, but want to play along at home, we've prepared a sample Shapefile of US Counties for you to work with. Download it here: `https://goo.gl/2neank`

The data is in Shapefile format, which as you learned in the GIS Data chapter, is the de facto standard for sharing map data or publishing your data to a web map or online GIS. If you have your own spatial data in a Shapefile, just substitute the field references in the tutorial for the unwanted data you have.

The Tools

We will be filtering the data using QGIS a desktop GIS. If you don't already have QGIS, download it now at `qgis.org`, and install it on your system. It's completely free, and is a great platform to continue learning about mapping and GIS.

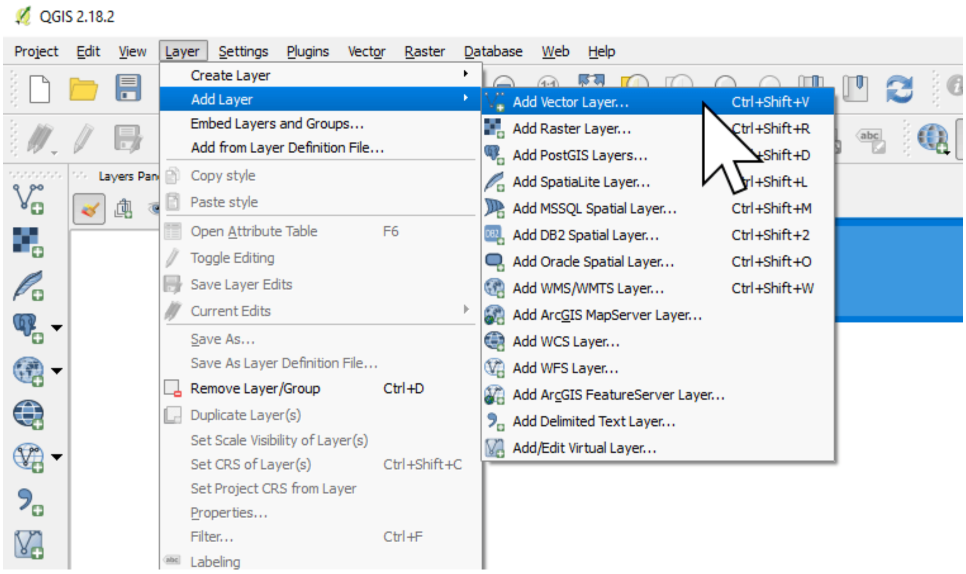
Don't worry, **you don't need to be a QGIS expert to complete this tutorial**. However, if you would like to learn more about it, then here is a great place to get started: `https://goo.gl/zi8qNb`

ℹ For this tutorial, we're using QGIS 2.18, but it should also work much the same on earlier and newer versions.

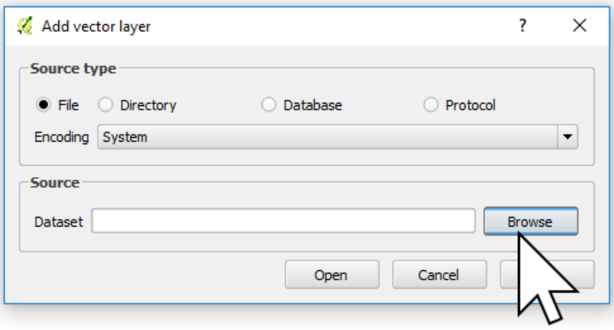
1. Import the spatial data to QGIS

⟶ Open QGIS

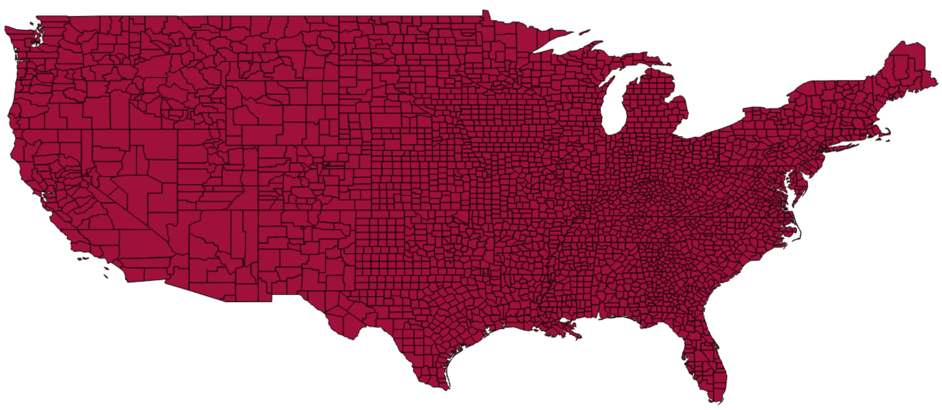
⟶ From the menu bar choose Layer → Add Layer → Add Vector Layer



⟶ Navigate to your shapefile and add the dataset by selecting the .shp component of the dataset.

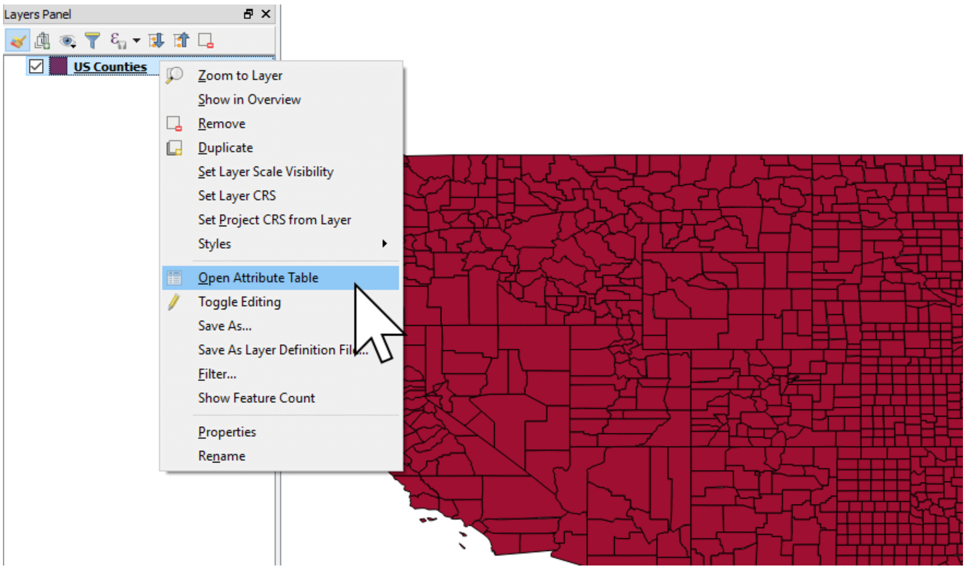


You'll see your data appear as a plain map of counties.

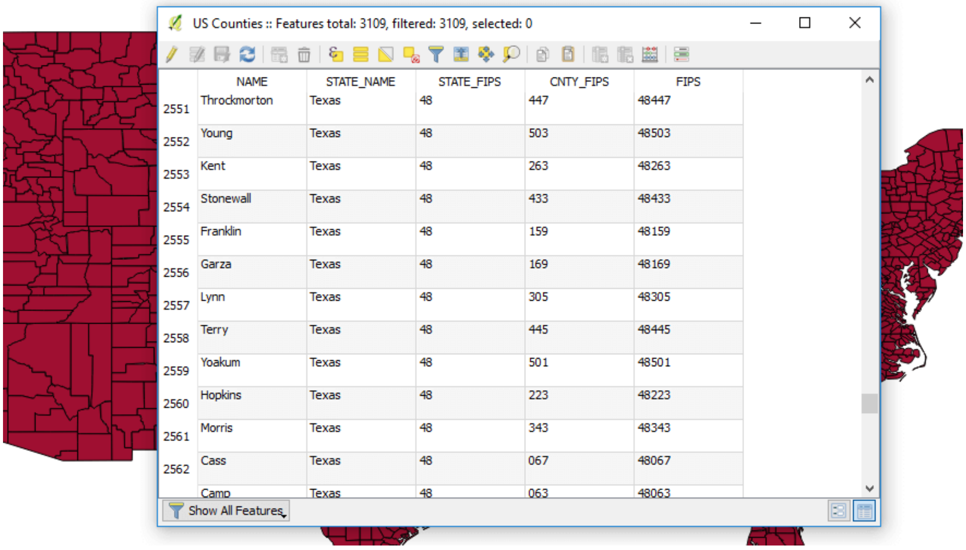


To understand what attributes are available to us to create an expression, let's open up the Attribute table and take a look at what's inside.

⟶ `Right click on the US Counties in the Layers Panel and select Open Attribute Table.`



Here we can see that there is series of columns, and lots of rows - 3,109 rows - one row for each County in the contiguous US. Each of these rows represents the county polygons you see on the map.



The columns shown are the attribute fields for each county:

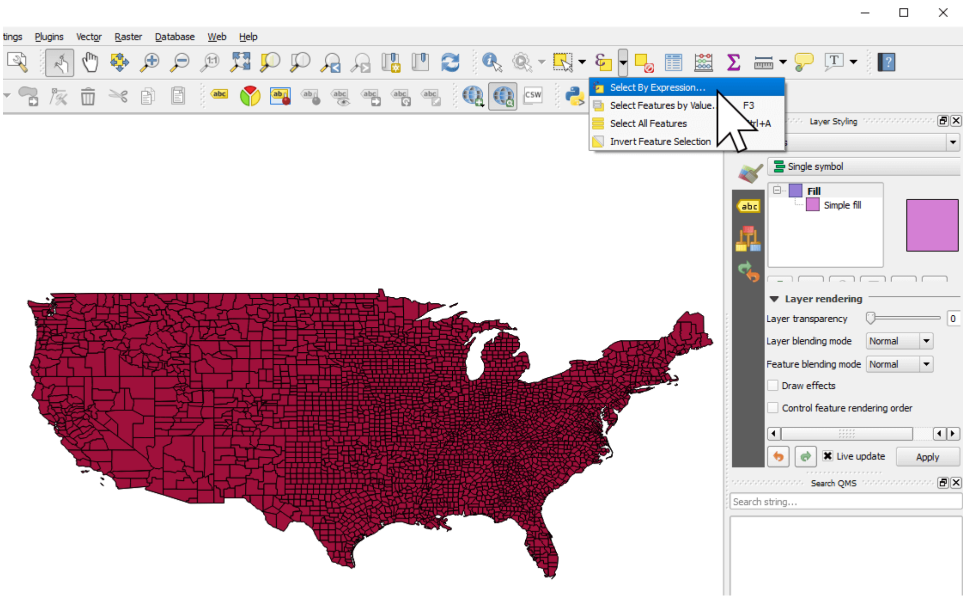
- `NAME` the county name;
- `STATE_NAME` the State the county is in;
- `STATE_FIPS` the adminstrative code for the state;
- `CNTY_FIPS` the administrative code for the county, and;
- `FIPS` which is simply state and fips code combined into one unique number.

To create our filter with an expression, we need one attribute that will be consistent across all the counties we want to keep. In this case, we have state names, so we can simply use “Texas” in our filter. If you only had a State FIPS code, you could use that instead.

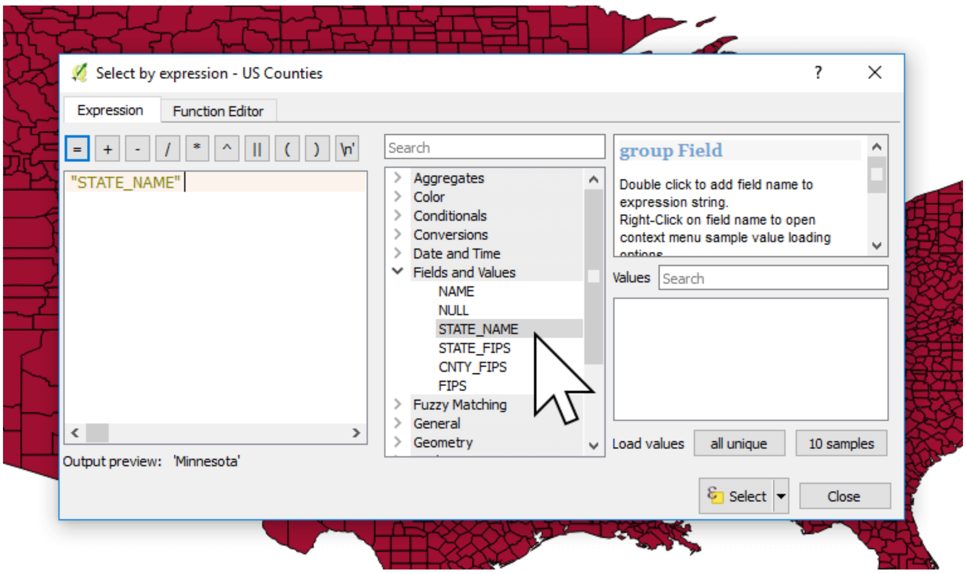
We’re now ready to create our Selection.

2. Creating a selection to delete information outside our area of interest

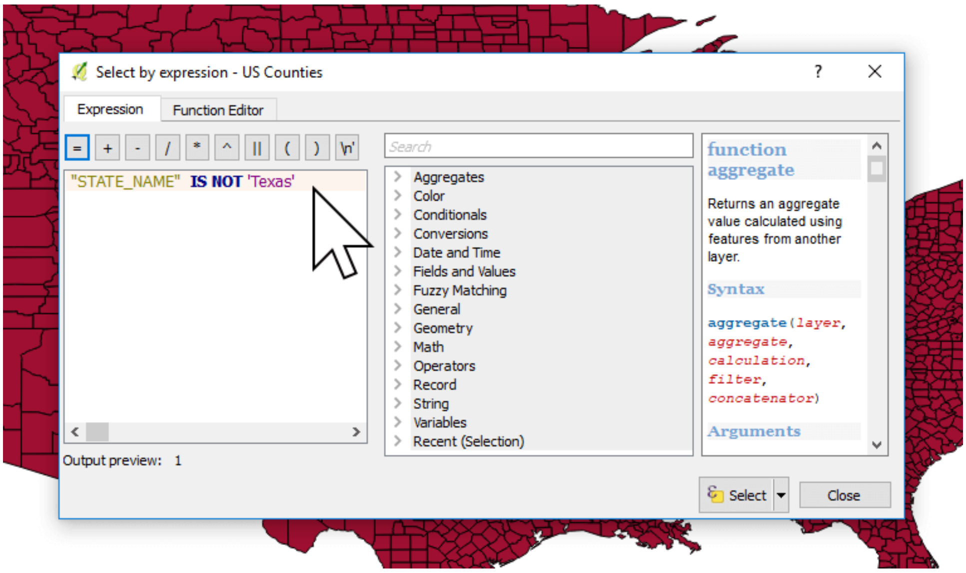
⟶ From the toolbar at the top, click on the expression icon and choose Select By Expression…



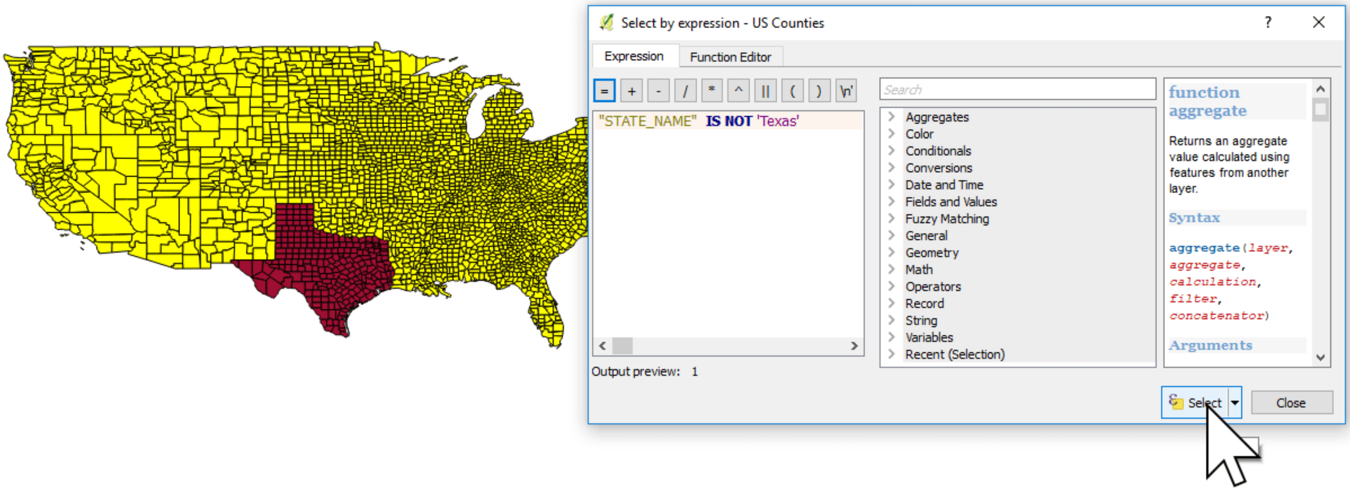
⟶ In the expressions list, click on `Fields and Values`, then double click on the field that contains our consisten attribute of Texas: `STATE_NAME` . You should then see `“STATE_NAME”` appear in the expres-sion editor on the left.



⟶ Now, click in the expression editor and type: IS NOT 'Texas'

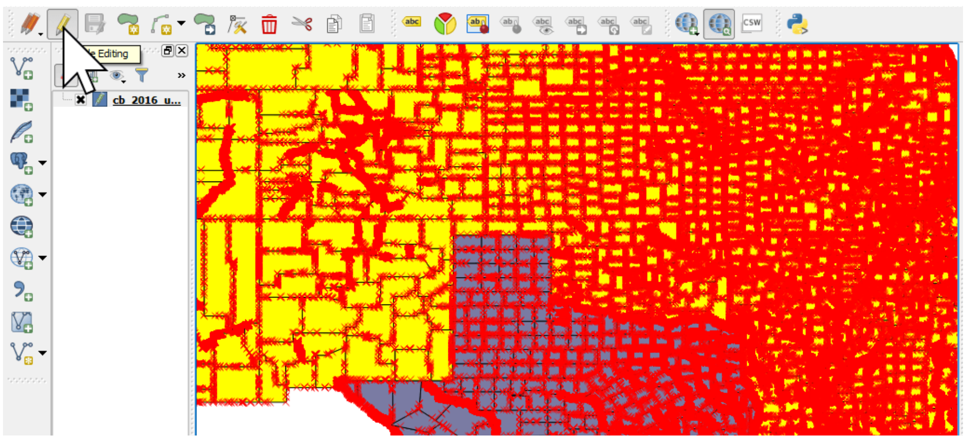


⟶ Click Select, and you'll see all counties in states that aren't Texas will be highlighted yellow.

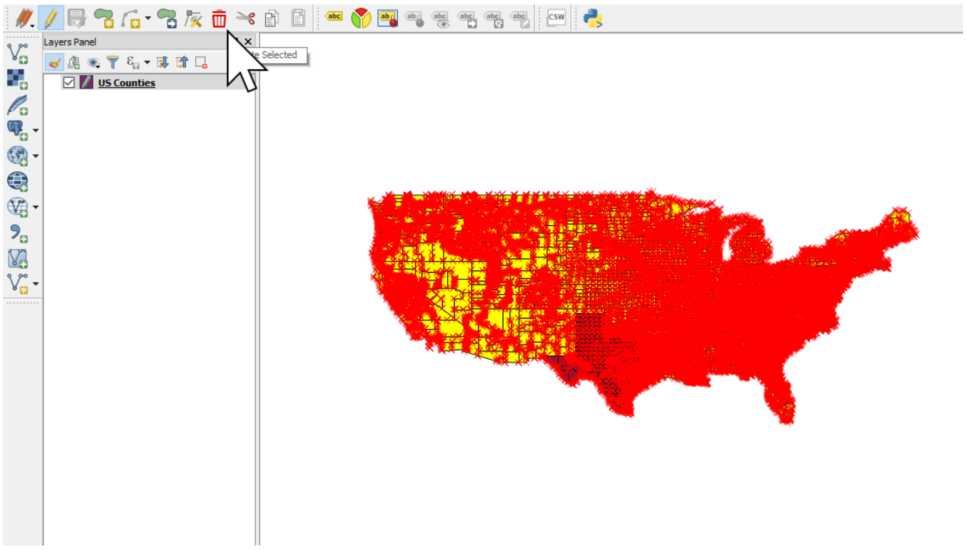


⟶ Close the Select by expression window.

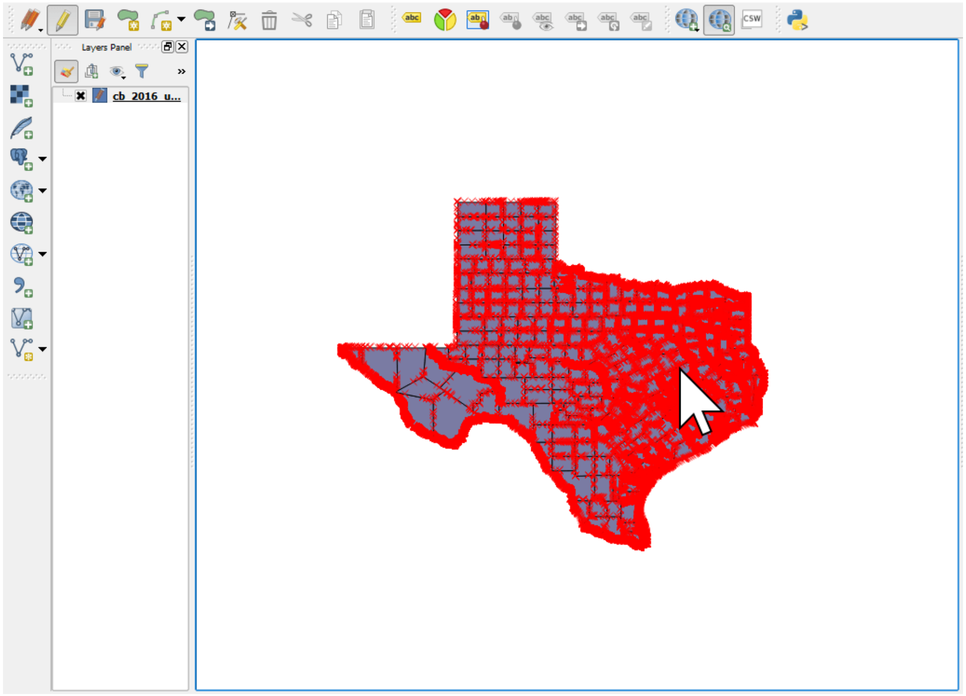
⟶ Now, toggle layer editing by clicking on the pen icon on the toolbar. In edit mode, you will see that all features will have red node markers.



⟶ Click on the red trash icon on the toolbar to delete the unwanted counties in the selection in one fell swoop.



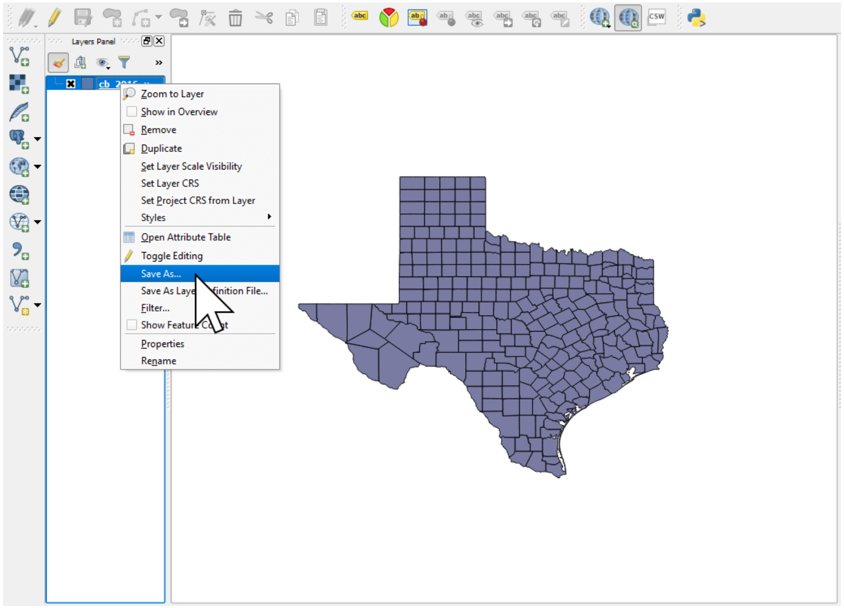
⟶ Now we should be left with only Texas counties. Nice.



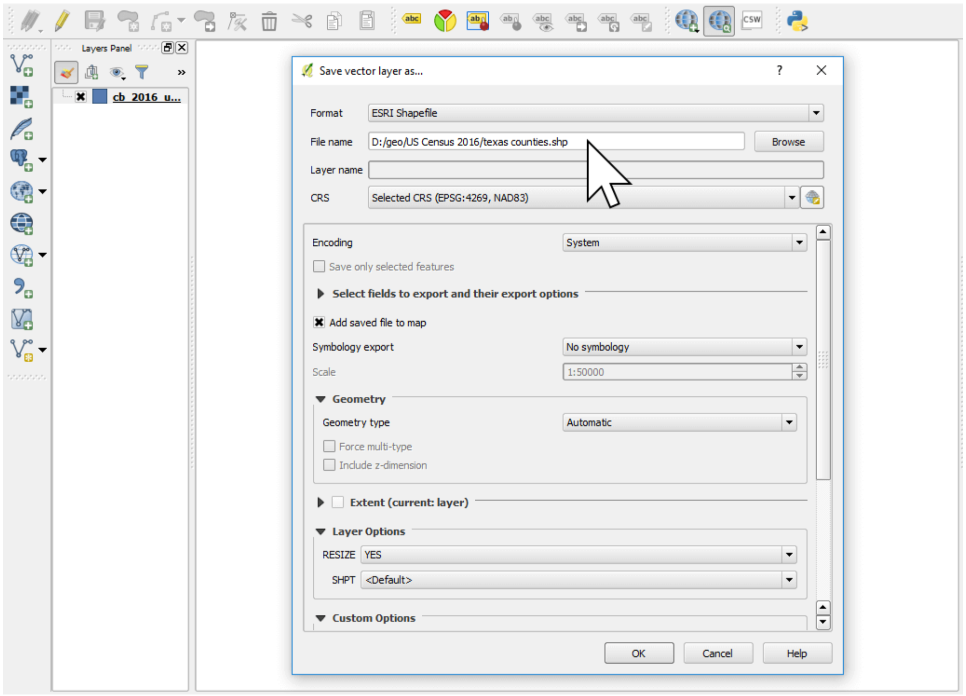
⟶ Exit edit mode by clicking on the pen icon again.

3. Export your data to Shapefile

⟶ Right click on your layer in the legend and select Save As...



⟶ Name your new dataset, and click OK.



And that's it!

You've removed unwanted data from your spatial file of US Counties, and now you have a shapefile containing only your area of interest. Good work!

HOW TO TURN A LIST OF LOCATIONS INTO AN ONLINE HEAT MAP

Creating great maps needs great data. Depending on what we're trying to visualize, we might find that we're missing a key attribute, or that the data we need is in a different dataset. This tutorial will show you some GIS magic to get a count of markers in a county and create a stunning quantity map for your online GIS.

Sometimes it just all comes together, and we've got everything we need to make a great map: we've got a spatial file with our points, and we've got a spatial file with our regions. Let's say points are all the colleges and universities across the US, the the polygons are all the counties in the US.

Adding them both to a web map will give us two layers. From the point markers we can see that some counties have lots of colleges, some have fewer. Great. Now, say we want to shade each county by the number of colleges. We'll just go to the layer style settings for the counties and select the points and… wait. The points aren't in the county layer, so there's no way to create a choropleth map showing the quantity of schools per County. So how can we do this?

To create a quantity web map, we need a numerical column in the County data that lists the schools per county. In a spreadsheet, this would be simple enough - just copy the entire count column from the college spreadsheet, and paste in the County spreadsheet.

But when dealing with spatial files, it's not quite that simple.

Lucky for you, we've written a step-by-step tutorial to run you through the whole process of merging data from one shapefile into another, so that we have the required quantity data to create an online choropleth map.

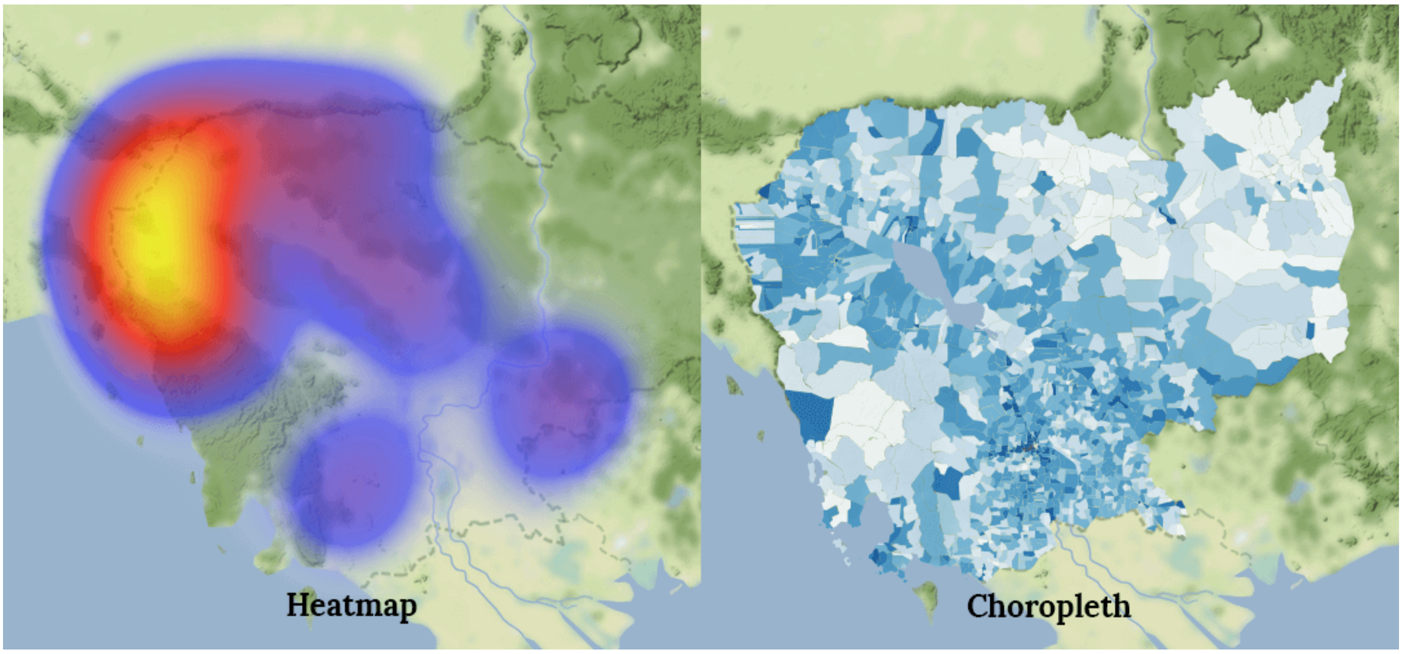
WHAT IS A CHOROPLETH MAP?

Sometimes choropleth maps are referred to as "heat maps". Now this isn't strictly true.

Renowned cartographer Gretchen Peterson[3], has a succinct definition:

> ***"A choropleth map shows a change across a geographic landscape within enumeration units such as countries, states, or watersheds.***
> ***A heat map shows a change across a geographic landscape as a rasterized dataset–conforming to an arbitrary, but usually small, grid size."***

In visual terms, the difference is quite clear:



What we'll be creating today is an online Choropleth map, sometimes referred to as a Quantity map.

3 www.gretchenpeterson.com

In this tutorial, we'll cover the following:

- Importing the spatial data into QGIS
- Calculate the number of points in each county
- Clean up the data
- Create an online quantity map of colleges per county

THE DATA

To create our quantity map, we're going to be using two datasets:

- The first is US colleges (point data) which contains the locations of all US schools.
- The second dataset is US counties (polygon data).

Download the sample data here: `https://goo.gl/K7S8sm`

The objective is to create a "heat map" that lets us see which counties have the highest number of colleges.

The data is in Shapefile format which is the de facto standard for sharing map data. Despite its singular name, it is in fact a collection of files – with a minimum of four key files: (.shp, .shx, .dbf, .prj).

THE TOOLS

We will be preparing the data using a popular open source GIS (geographic information system) program called QGIS. If you don't already have QGIS, download it now. It's completely free, and is a great platform to continue learning about mapping and GIS.

Don't worry, **you don't need to be a QGIS expert to complete this tutorial**. However, if you would like to learn more about it, then here is a great place to get started: `https://goo.gl/zi8qNb`

For this tutorial, we're using QGIS 2.18, but it should also work much the same on earlier and newer versions.

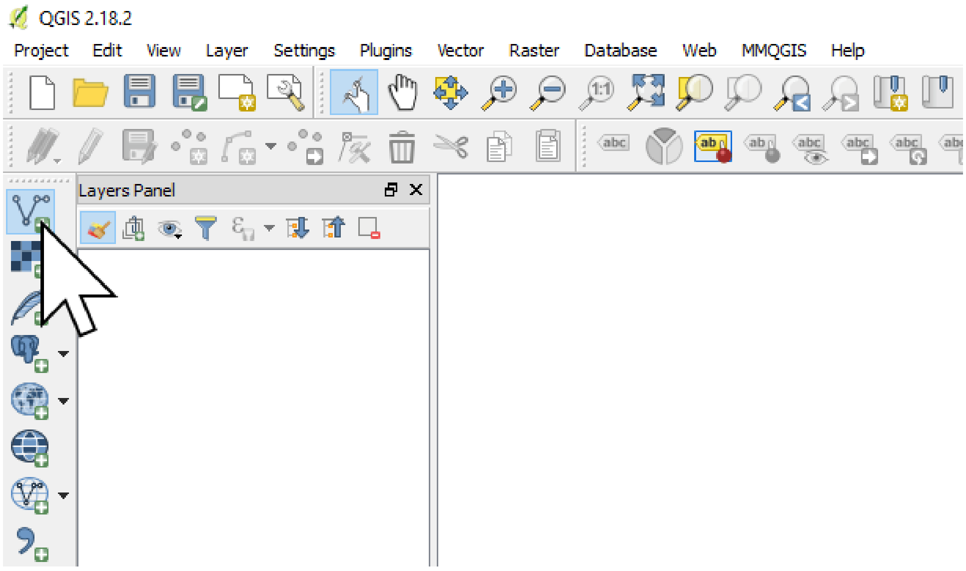
Let's get started!

1. Importing the Spatial Data into QGIS

First unzip the sample data you downloaded and save it in a local directory. You will see there are two Shapefiles: Colleges and Universities, and US Counties.

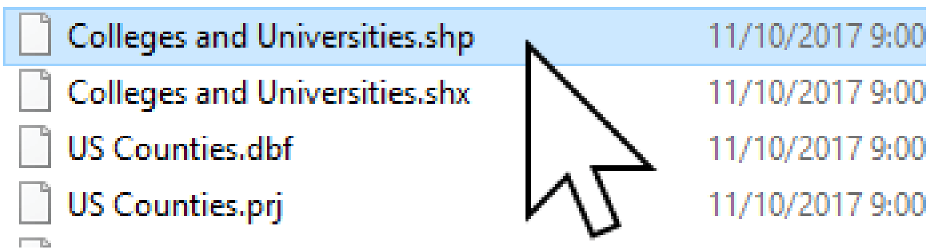
Now we need to open QGIS and complete the following steps:

⟶ From the left toolbar, click on the Add Vector layer icon.

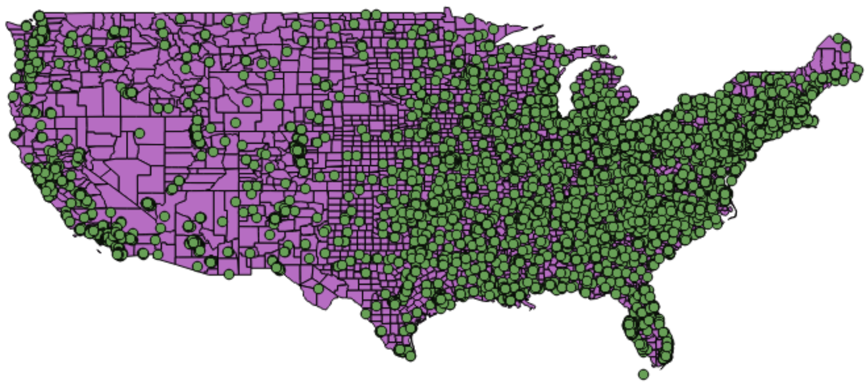


⟶ Press the Browse button and choose the US Counties.shp file and press Open.

⟶ 3. Repeat steps one and two, and open the Colleges and Universities.shp file.



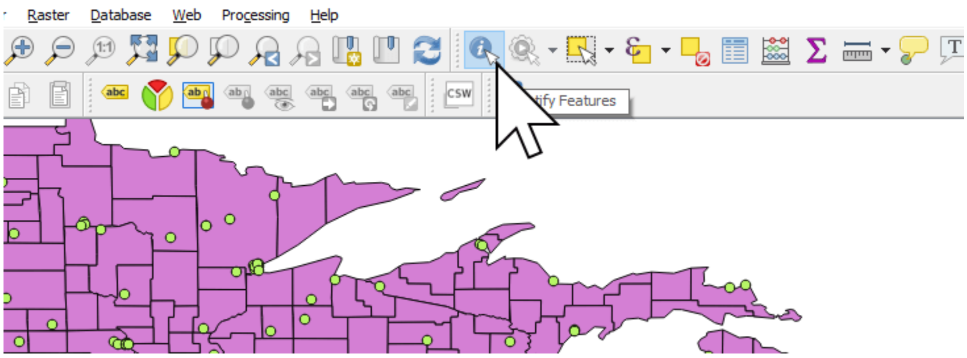
→ Now we can see the two datasets displayed on the map as layers.



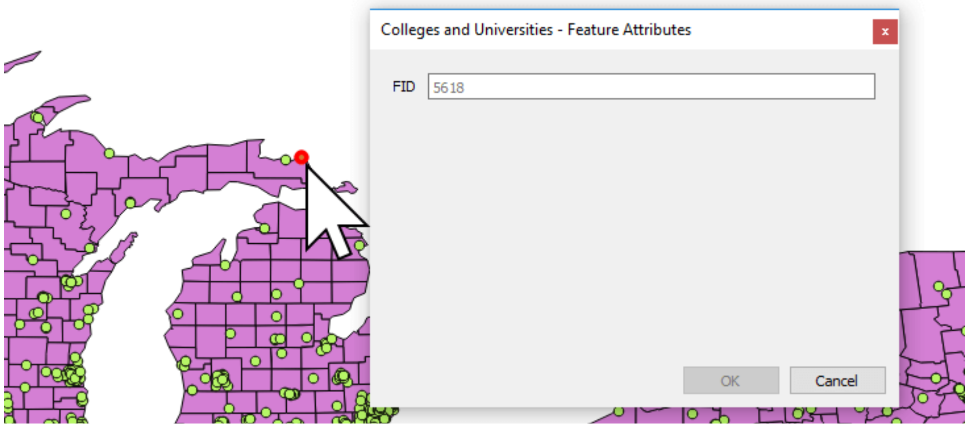
ℹ A Shapefile is made up of two main parts, the first is the geometry which you can now see on the map, the second is the attribute data which contains data about each feature.

Let's take a look at the attribute data for one of the Universities.

→ Activate the Identify Features tool, found on the top icon bar.



→ Click on a feature on the map to reveal the attributes.



We can see that the Universities data we're using doesn't contain any information besides an FID, a unique number assigned to each location. We're only interested in using the location, so we don't need any other data.

We're now going to use these location points and find out how many there are in each County.

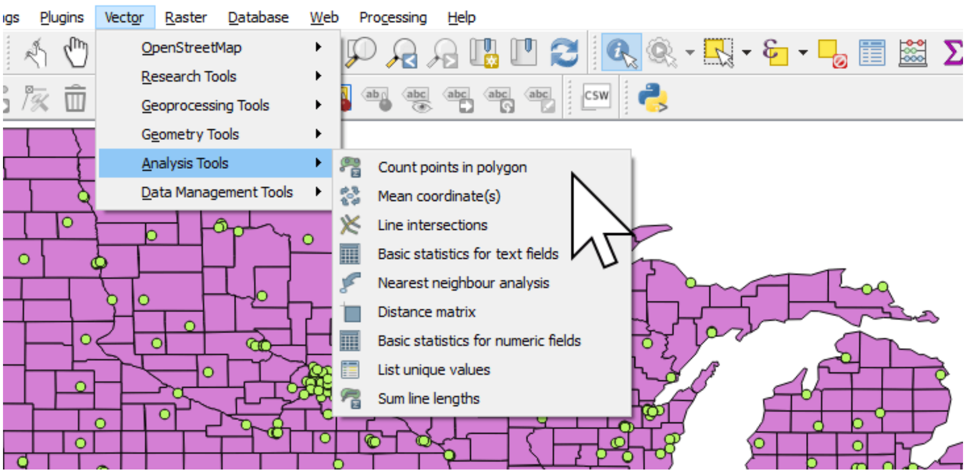
2. Counting the Points (colleges) Inside Each Polygon (Counties)

As you can see from the points on the map, there's a huge number of Colleges and Universities across the US. Some counties have dozens, some have none. Obviously, counting all of the points manually would be very hard work.

Thankfully, QGIS can do the hard work for us!

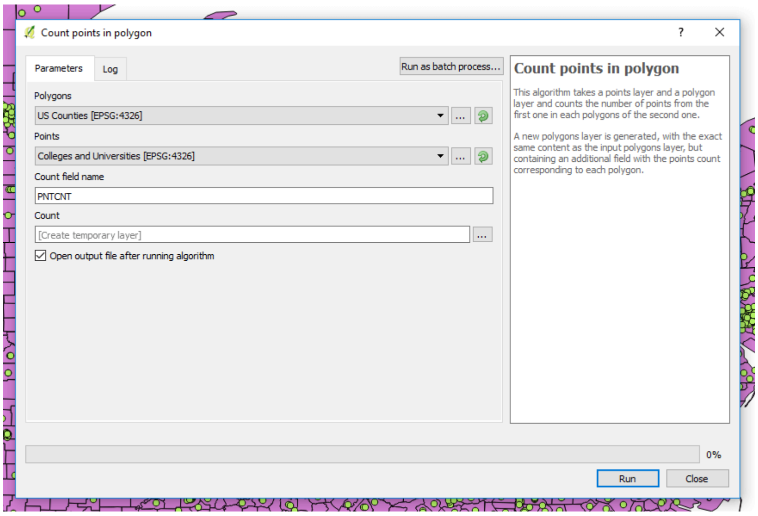
We're now going to use QGIS to count how many colleges are in each county, and add that number as an extra column called "PNTCNT" in the attribute data of each country.

⟶ In the main menu press **Vector → Analysis Tools → Count points in polygon**.

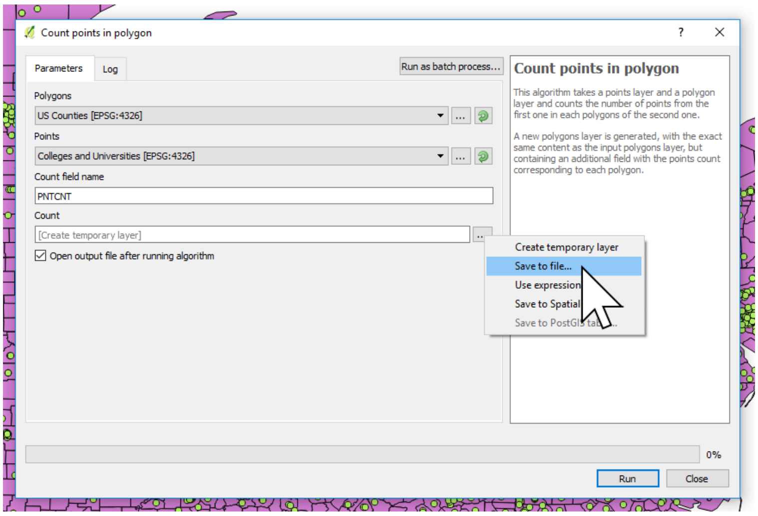


The Count points in polygon feature will attempt to automatically add our polygon and point layers. If you had more layers active in QGIS, you would need to select the correct layers.

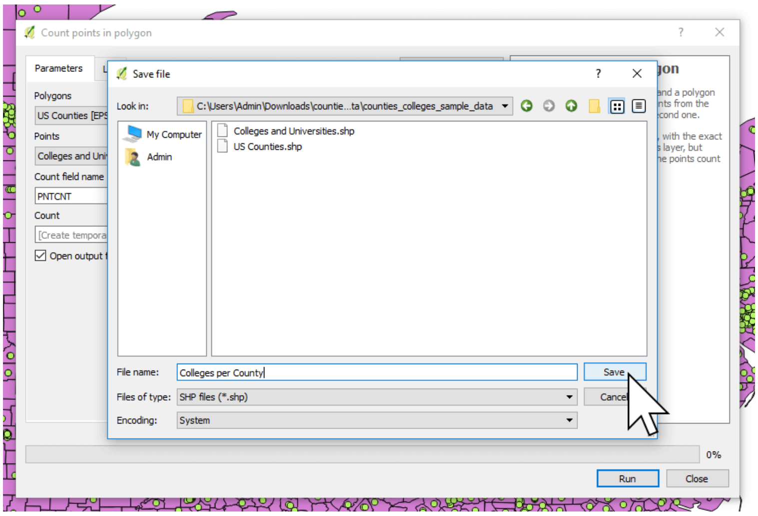
→ In the `Count field name` field, type `PNTCNT`.



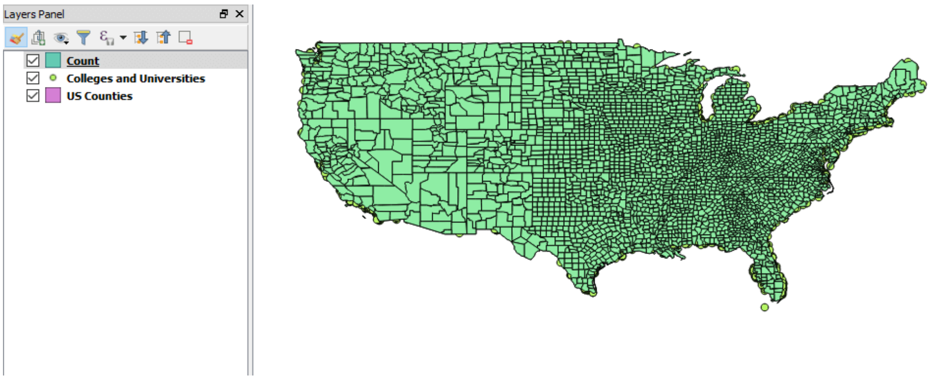
→ On the `Count` row, click the three dot icon `…` and select `Save to file...`



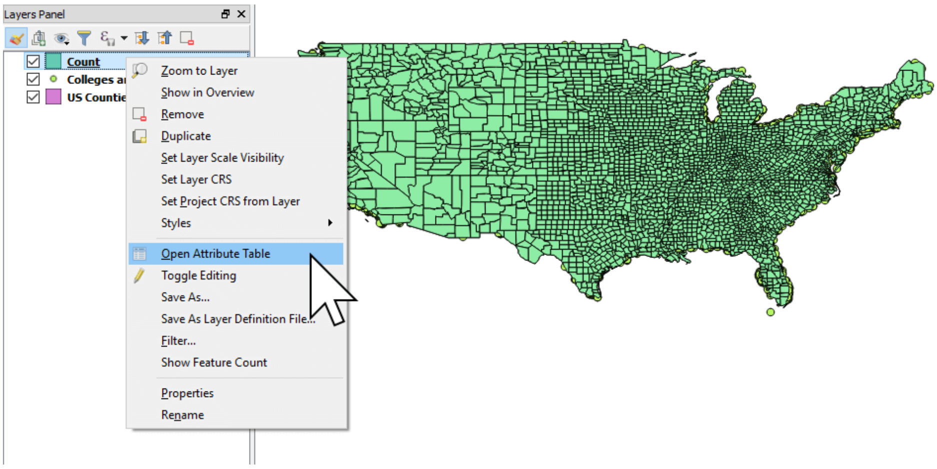
→ Browse to a location to where you want to save the Count results set the File name to Colleges per County. Click `Save`, then click `Run`.



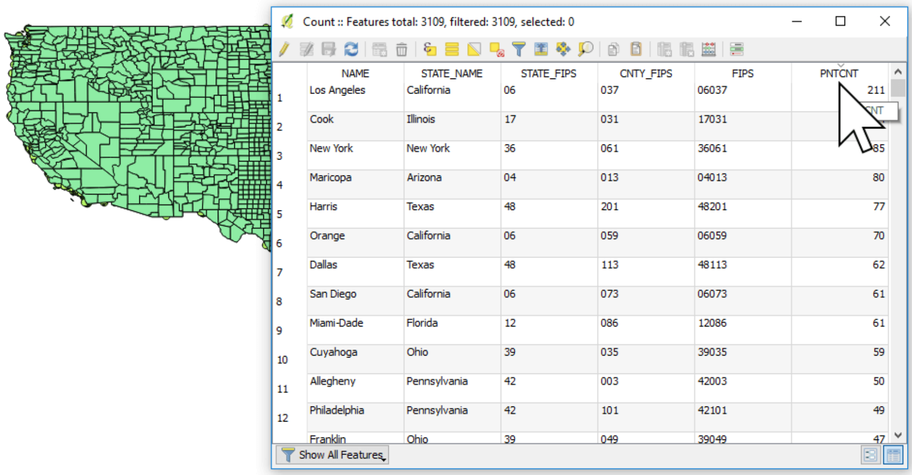
QGIS will calculate the number of points per county. When complete, a new map layer will appear on screen.



→ `In the left hand to Layers Panel, you can now right click on the new layer "Count", and select "Open Attribute Table".`



Here in the "`PNTCNT`" column we can see the number of schools in each county. Mission accomplished!

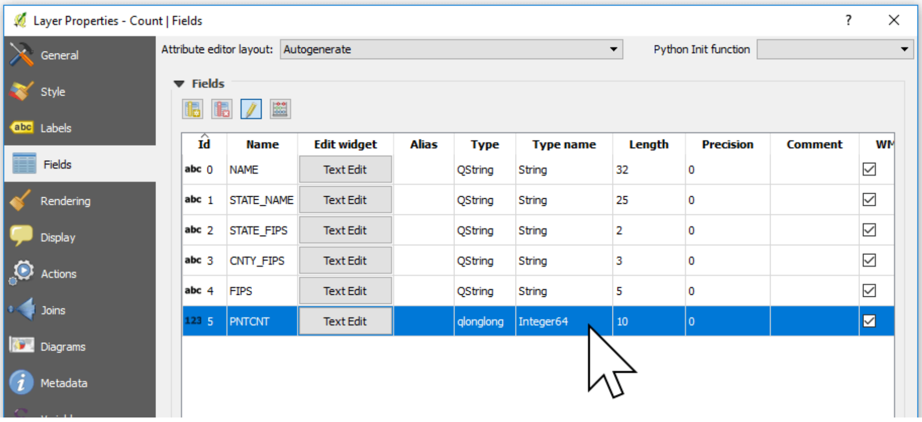


	NAME	STATE_NAME	STATE_FIPS	CNTY_FIPS	FIPS	PNTCNT
1	Los Angeles	California	06	037	06037	211
2	Cook	Illinois	17	031	17031	[illegible]
3	New York	New York	36	061	36061	85
4	Maricopa	Arizona	04	013	04013	80
5	Harris	Texas	48	201	48201	77
6	Orange	California	06	059	06059	70
7	Dallas	Texas	48	113	48113	62
8	San Diego	California	06	073	06073	61
9	Miami-Dade	Florida	12	086	12086	61
10	Cuyahoga	Ohio	39	035	39035	59
11	Allegheny	Pennsylvania	42	003	42003	50
12	Philadelphia	Pennsylvania	42	101	42101	49
	Franklin	Ohio	39	049	39049	47

3. Data Clean Up

We're almost done, but before we create our online quantity map, we'll need to adjust the `PNTCNT` column we just created.

If you open the Count layer properties, you find that "`PNTCNT`" is a 64 bit floating point integer, as indicated by the `Type` name `Integer64`.

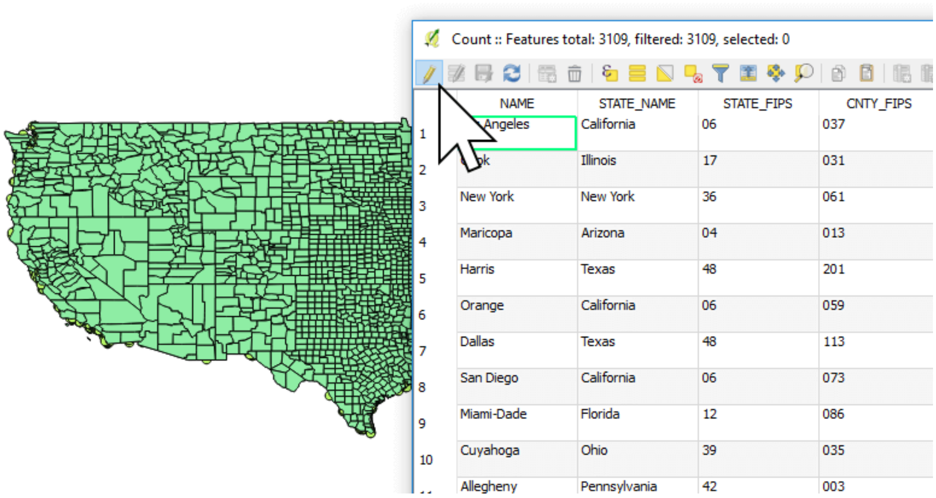


What we need for our web map is whole number, a plain old Integer.

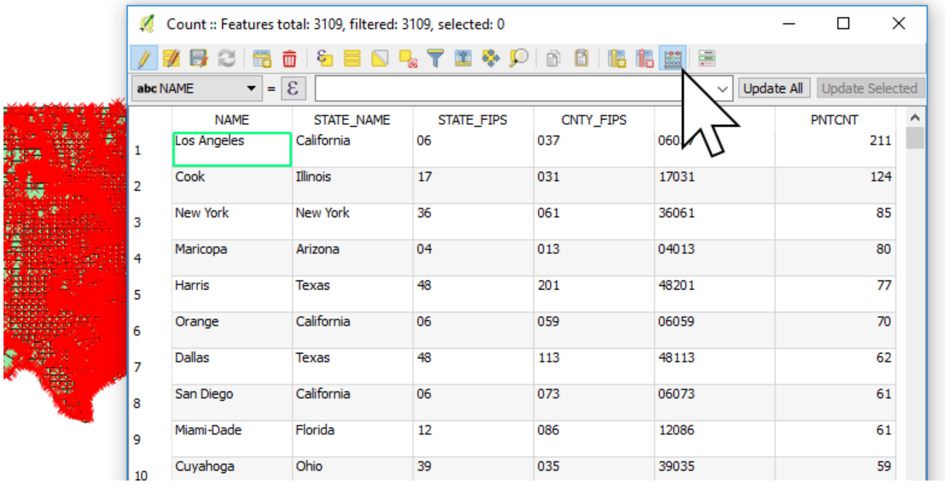
Let's change it now. To fix this we need to convert the field to whole number using the Field Calculator.

⟶ `Return to the attribute table by right click "Count" and choose "Open attribute table".`

⟶ `Press the Edit mode button - it looks like a pencil - from the top menu bar.`



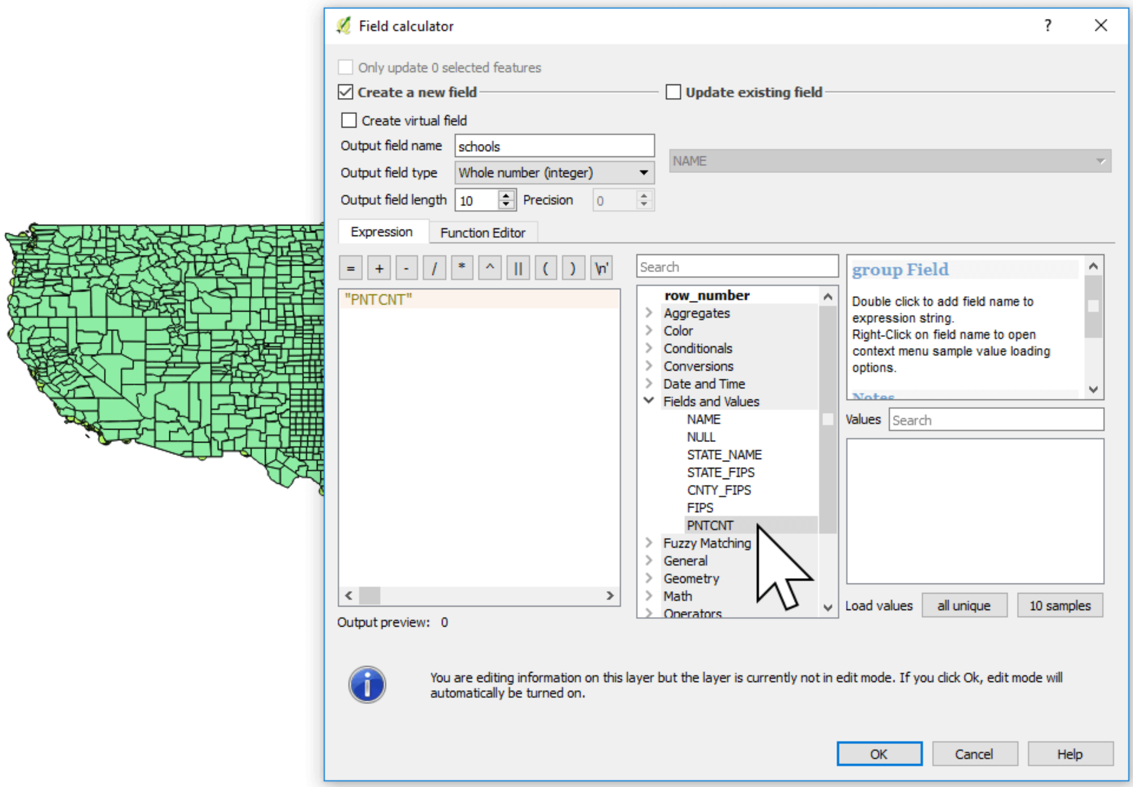
⟶ Now press the Open field calculator button - it looks like an abacus - from the top menu (Or press CTRL+ I).



⟶ In the Field calculator, change Output field name to schools.

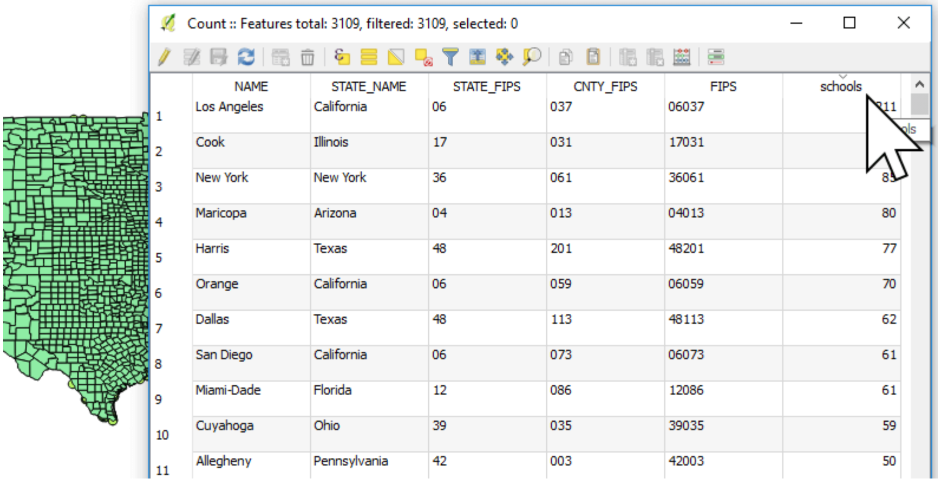
⟶ Make sure Output field type is “Integer (whole number)”.

⟶ Select Fields and values from the functions list, and double click on PNTCNT. You’ll see “PNTCNT” appear in the Expression window.

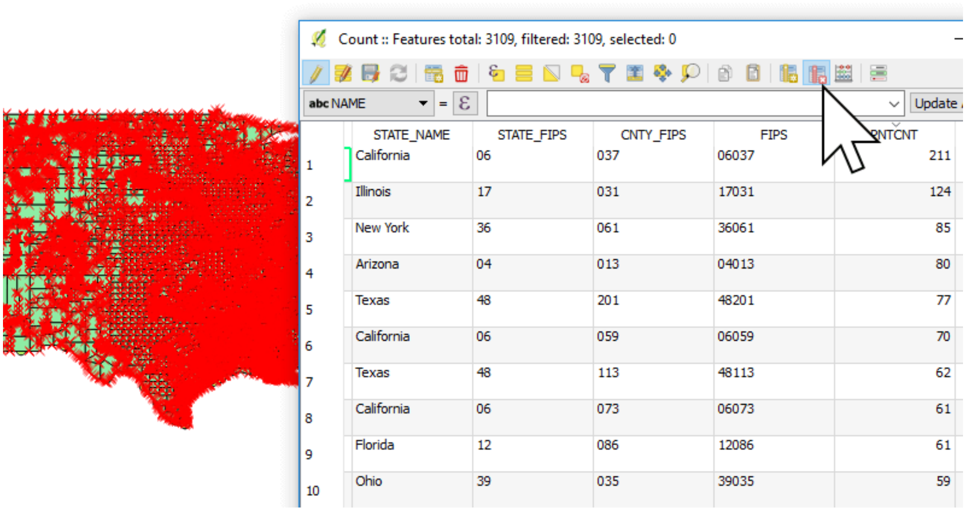


⟶ Press `OK`.

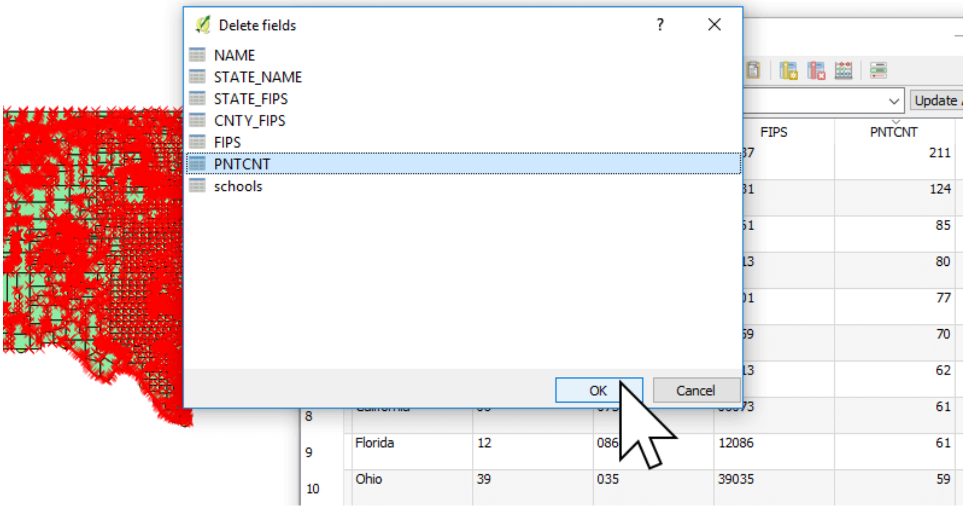
You will now see a new column called `schools`. The values might look the same as before, but now they're in the format we need.



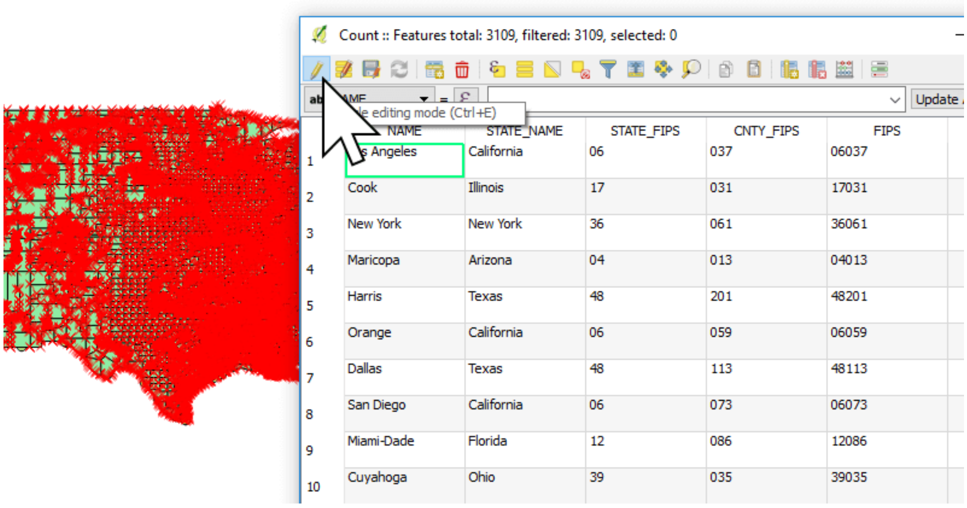
⟶ We can now delete the `PNTCNT` column by pressing the `"Delete columns"` button in the top menu bar – it's the red column icon with the 'x' – or by pressing `Ctrl + l`.



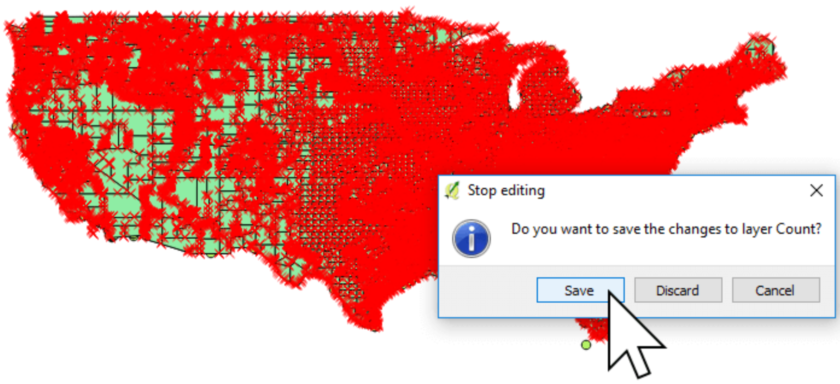
⟶ Now select `PNTCNT` and click `OK`.



→ We should finish up by pressing exiting editing mode by clicking the pencil icon, or pressing `Ctrl+E` on your keyboard.



→ QGIS will prompt to save changes. Click `Save`.



→ Exit QGIS and head over to Mango.

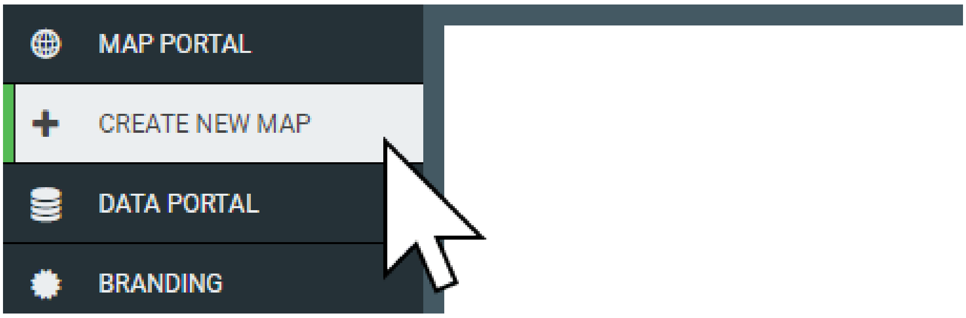
4. Creating Our Quantity Map in Mango

We're going to create a choropleth map. In Mango, we call this type of map a Quantity map for the sake of simplicity.

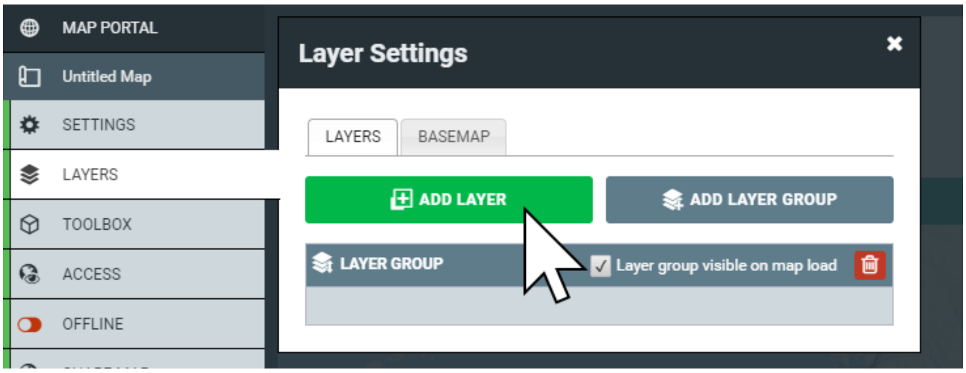
Now that you have your Shapefile, let's upload it to Mango and take a look. If you don't yet have an account, you can sign up for a 30-day free trial here at `mangomap.com/sign-up`.

Once you have signed in to your account, we're ready to make a map!

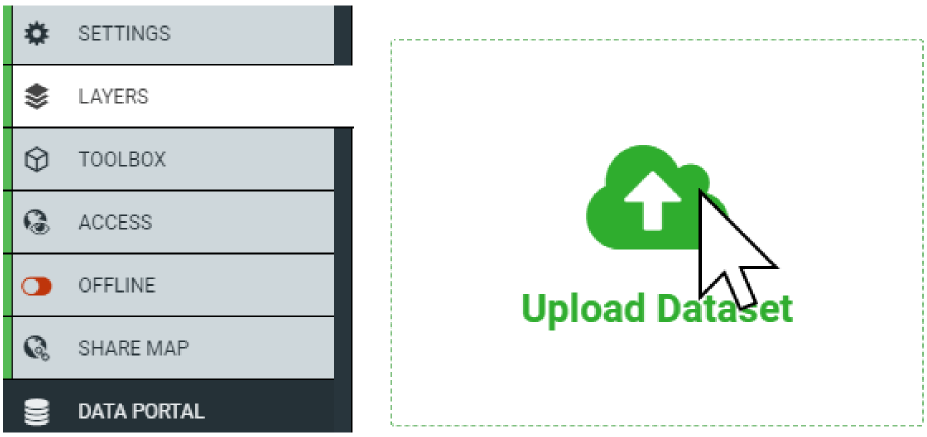
⟶ Press the "CREATE NEW MAP" button in your administration sidebar.



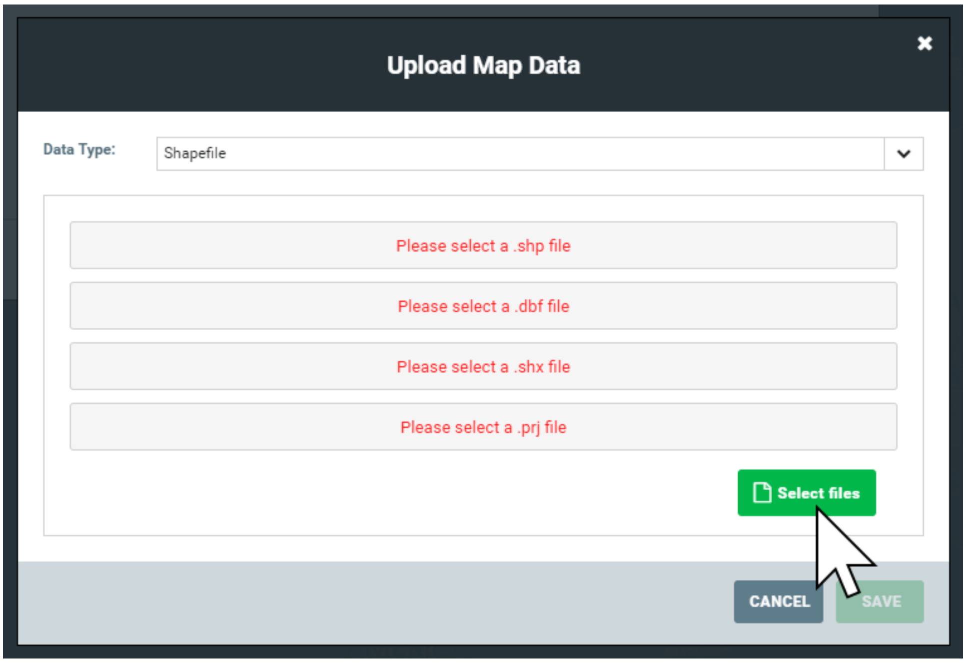
⟶ When your map is ready, click "LAYERS", then click on the "Add Layer" button.



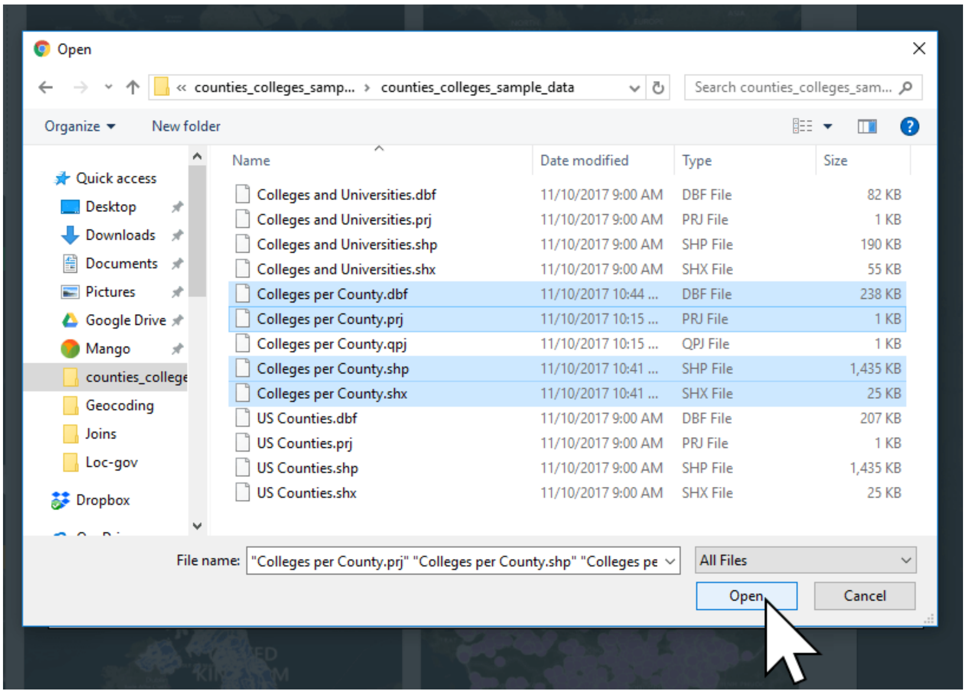
⟶ Now click on "Upload Dataset"



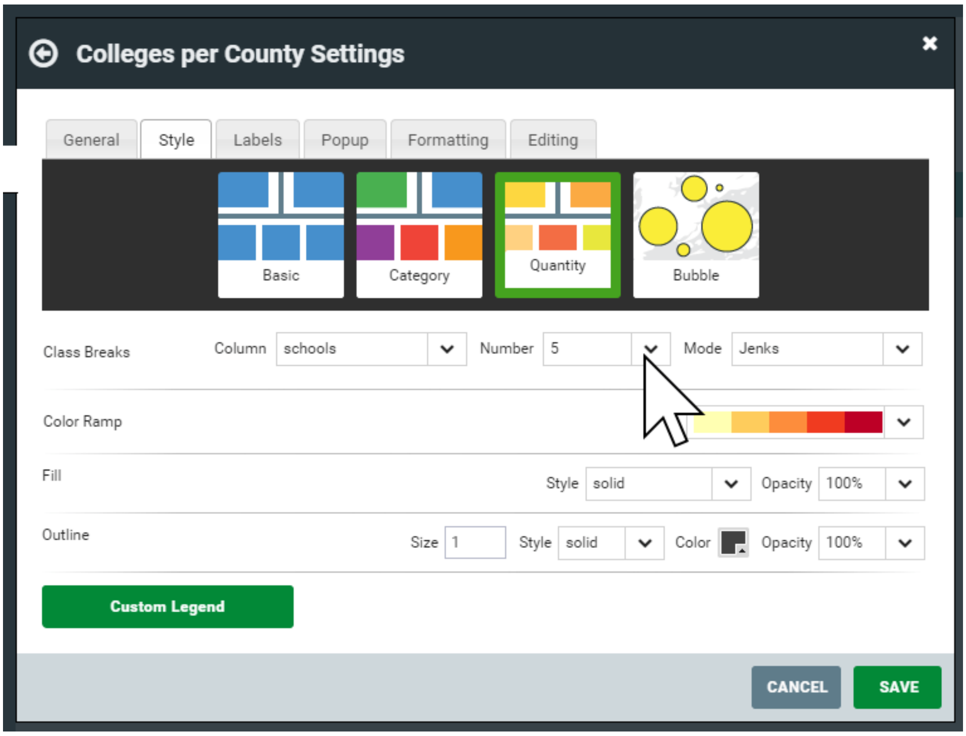
⟶ Here you can select the type of data you want to upload. We've got a Shapefile, so we don't need to change anything. Click on Select files.



⟶ Navigate to the output location of our Colleges per County Shapefile, and select the four files of the Shapefile generated by the geocoder and upload them to Mango. To select multiple files at once, press and hold Ctrl and clicking each file (⌘ + click for Mac).

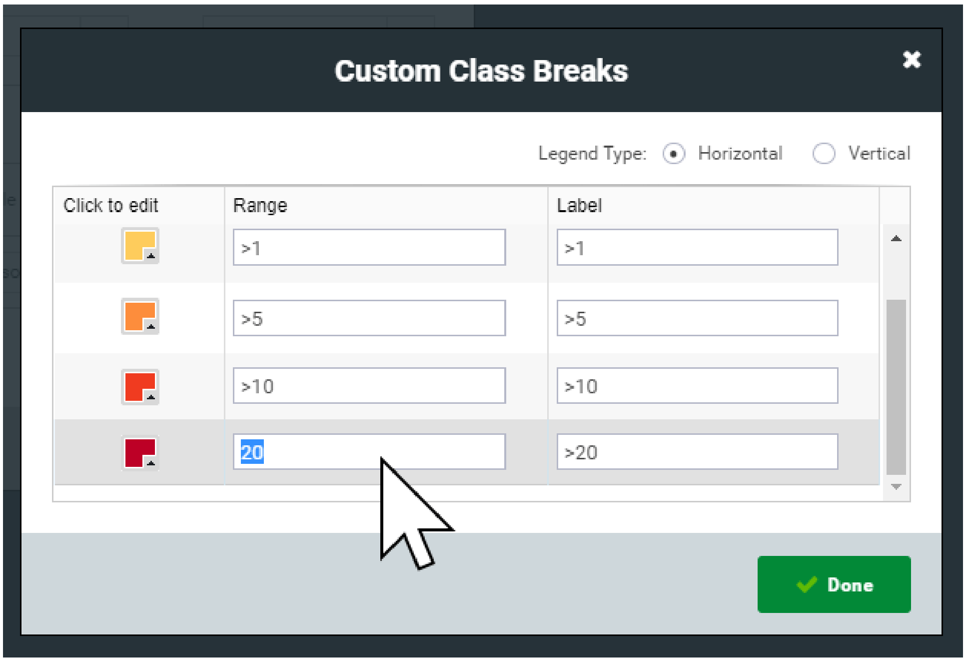


⟶ Once the upload finishes choose "Quantity" in the layer style panel.



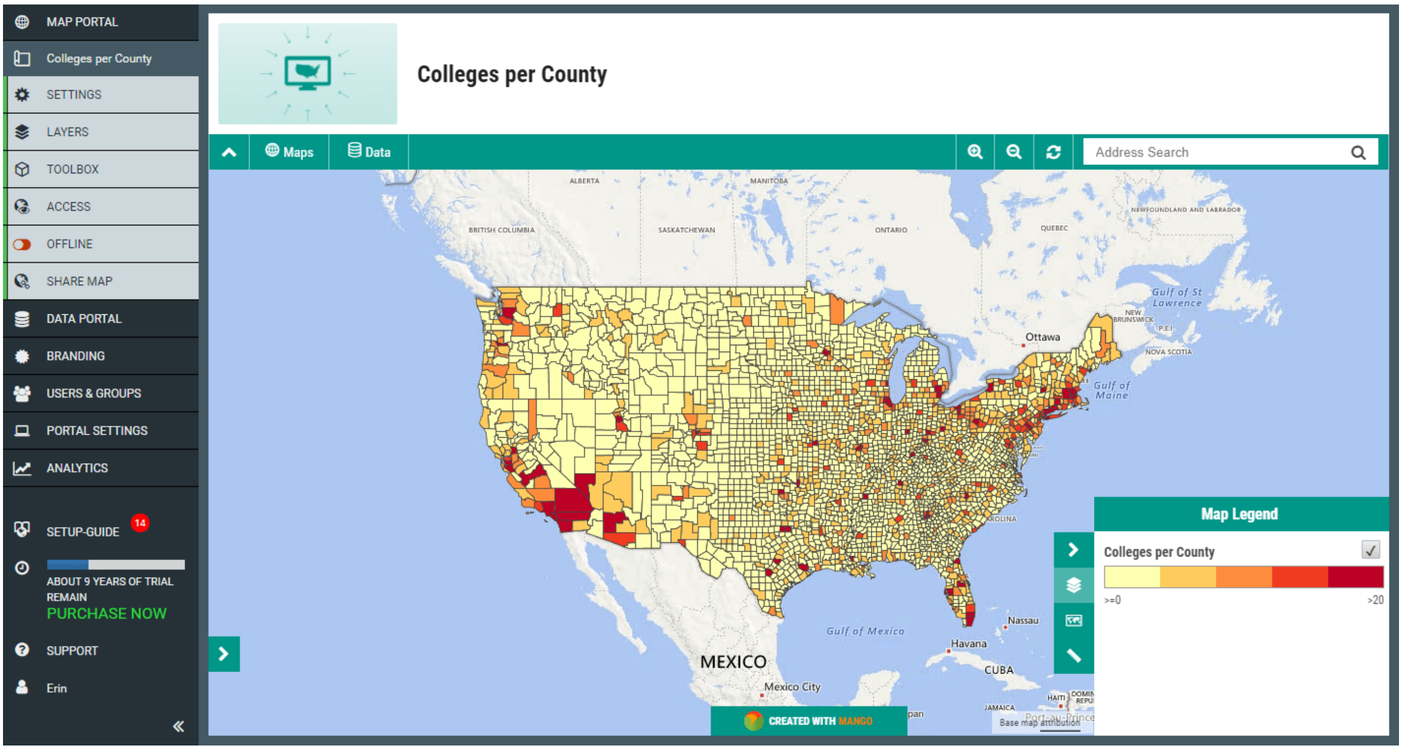
⟶ In the Class Breaks row, select Column "schools" and set Number to 5.

Click the green `Custom Legend` button at the bottom of the panel and change the Range column in the table to 0, 1, 5, 10, 20.



⟶ Press `Done`, the press `Save` on the Layer Settings panel.

You will now be able to see your quantity map.



Nice work!

It's a simple process, yet reveals a fascinating web map visualization. We can see large swathes of the US aren't services by any universities or colleges, instead, they are clustered around the larger population areas on the east and west coasts. Sometimes, quantity maps can serve to reinforce existing understanding, or to reveal new understanding about the physical geography of our society. This same process can be applied to many types of data for your business or organization.

Returning to the Layers panel, you can experiment with different custom Class Breaks, to see how minor changes in how the data is classified will result in big changes on the web map visualization.

Ready to get started?

Dive straight into GIS web mapping with a 30-day free trial of Enterprise plan on Mango.

Sign up at www.mangomap.com/sign-up

Questions?

If you have questions about Mango, we'd love to talk!

You can request a demo from one of our amazing team members. Just click the link below to book an appointment, and we will be in touch!

Talk to you soon.

Book an appointment at www.mangomap.com/contact

Mango

The Simple Online GIS

Make Amazing Interactive Web Maps That You and Your Users Will Love!