

Transcripts - Intro to Mapping and GIS Merging CSV data into shapefiles (QGIS)

Hello everybody! Welcome back to the hands-on portion QGIS, we just got done adding hospital dots to the map, the dots on the map representing every hospital in the United States, so we're going to go ahead and put that data away for a second and come back to it later and work more with this country shapefile that we worked in the first module.

What we're going to make in the next 2 videos is a choropleth map which is going to show fertility rates for every country and the United States, at least most of them. So the end result will be a map that shows areas with higher fertility rates with darker shades and those with lower fertility rates with lighter shades, which is very common you see choropleth maps all over the place.

The first step we're going to do is we're going to actually download fertility data, which is broken down by country by the World Bank and we'll go there and download that data. We're going to join that in with the country's shapes that we already have, and I'll explain what joining is in a second.

So that's the first step, so that's we're going to be working on in this video, and then the second video we will actually be coloring the map and so you'll see which areas have higher fertility rates and they'll pop out on the map.

So this is the first step, like I said, is going to the World Bank and downloading some data from there, the fertility data. So go ahead and open up that website. I'm going to do that here in this worldbank.org, go ahead and scroll down and there should be a little spot under this data panel for view more data.

Go ahead and open that up. Go ahead and start typing in the word "fertility" and the first thing that should pop up is "fertility rate, total", go ahead and click on that, and then there's this nifty little download button CSV, which we worked on in the last class, we're going to go work on another CSV here, so go ahead and click on download there.

Go ahead and go to the folder that that downloaded in. For me, it's the downloads folder and it's a zip drive. So go ahead and unzip that and once you unzip that you should see a series of files.

You can go ahead and ignore the metadata ones, those just tell you a little bit more about the spreadsheet. The one we're going to be working with that actually contains all the data is this API_SP... file. It's kind of a long ugly file name, but it's got the data in it, so that's what's important.

So go ahead and open that up if you have Excel or any other spreadsheet editing tool, go ahead and open it up with that. This step isn't necessarily required, you can work around it, but if you have Excel or you have some sort of way of editing a spreadsheet this would be super helpful

going forward. The only thing we're going to do is really remove the first few rows in this spreadsheet. So I have Excel open here, again you can do this in a million different programs, but as you can see here, we have the data source, we have this last updated.

So it's really not until row five that we actually have the data that we want and you can see these are all fertility rates, these are all the countries, we have this country code in here, that's really going to be super important later, and then we have all of the years starting with 1960 going all the way to 2016 fertility rates for every country.

So what we're just going to use for this class is just the 2016 data. So go ahead and just first off remove those first four rows because those are useless for us, and then go ahead and remove all of the 1960 to 2015 data. So you should only have 2016 data and go ahead and remove the 2017 column as well, which is at least as of right now is empty. So go ahead and just save that.

Your sheets should now look like this. We have the country, make sure that country code is in there again, that's super important, and then we have their fertility rate for 2016. So that's all we need to do in Excel or whatever you're using, so go ahead and close that.

Go ahead and open QGIS backup, and we're going to once again add this layer, just like we've been doing with the shapefile and the dots, we're going to be adding it in a very similar way, so you want to go up to add layer, go to add delimited text layer because this is a CSV.

So go ahead and click on the file name, go ahead and find where it's at. For me, it's in the downloads folder still kind of originally where it was at. I just edited that data and just saved over the file that was downloaded. So it's still in the downloads folder for me. So go ahead and open that up and you should see, once again, we have some pre-populated going on.

Again, it's guessing that it's a CSV, which it is, so that's great. This geometry definition is going to be a little bit different though. You remember last time in our spreadsheet we had a longitude and latitude fields, which again help those hospitals get on the map.

This time we actually don't have any geometry information inside the spreadsheet. It's only the country, the code, the rate and a couple other things, there's no geometry in there.

So go ahead and hit on no geometry, and go ahead and hit add. So you're not going to see anything added to the map and again that's because we don't have any geometry, it doesn't have any way of putting this on there.

If we had latitude longitude or if we had some sort of other way like if this was a shape file instead of a CSV it would show up on the map hopefully, but this is not, this is simply a spreadsheet with data in it, and the way we're actually going to show it on the map is we're actually going to take this data and we're going to merge it with the country shapefile.

So this process is called joining and sequel and the idea is you have your country's shapes which we already have on the map and we have some other data which is the fertility rates for all those countries and we're going to merge them together, so the shapefiles will now have the fertility rates included in them.

So I'll walk you through that as well. It's really important through this process that we have a column in each of the spreadsheets, in the shapefile, and also in the spreadsheet that matchup and we do in this case and I'll show you what I mean by that.

Okay, so go ahead and right-click on your country's shape file here and then go to open attribute table, which we've done before, we're going to do it again.

Again remember we looked at this before this is every country on the map that has shapes is in this spreadsheet. What you're going to want to particularly pay close attention to is this ADM0_A3 column.

This is a three digit code for each country, it's called the world development indicator, and this is really helpful because within our fertility data from the World Bank, there's also this world development indicator column, so it's easy for us to take this data and merge it with their fertility data because of these similar columns, and I'll show you kind of what that looks like.

But just know that that's in there you can kind of go up and down and see it, they're all very consistently 3-digit codes regardless of what country it is and how long the name of the country is, so that's really nice so it's uniform and every country has its own unique code, which is important.

So go ahead and close out of that. Then we're going to go ahead and right-click on the World Bank data that we added and go ahead and open attribute table on that as well.

And here's the data that we were just looking at, but your notice here that along with the fertility rate, which is over here on the far right which is this 2016 column, we also have a country code and again, it's a 3-digit code regardless of what country it is, it's a consistent 3-digit code.

Unfortunately this 3-digit code matches up with the 3-digit codes that are within that country shapefile, and that's what we're going to actually use to join this data. So the end result again will be this country's shape file with the fertility rates in it. So in order to join it, it's really easy to do in QGIS. So go ahead and close out of that.

Go ahead and right click on our country shapefile again. Go to properties, and then I put it out in the first module video, I didn't really explain much of it, but it's a joins panel. This is really important, so go ahead and click on that. This is where we're going to do the joining and then click this little plus button.

And this is where we're going to basically tell what field to join on, and when I say join again, that's sort of like a database term, but if it confuses you just think of merging, so tell it how to merge. So in this case the layer we're going to be joining is our CSV, which it's already pulled in there. Again, this is the CSV with the fertility rates in it.

So we've got the country shapefile fold up, we want to join in the World Bank data with the fertility rates. So the join layer is going to be this spreadsheet is a `api.csv`. The join field is going to be again the country code as I mentioned, that's where we're going to match on, and the Target Field is going to be what column within the shapefile that we're going to join on.

And in this case our country's shapefile again, as I remember I tell you, that 3-digit code is in the column called `ADM0_A3`. So the Target Field is going to be `ADM0_A3` now if it was the code was in a column named `code` or `country code` or something similar you'd want to make sure the Target Field was correctly selected here.

But in this instance that is the field that we want to join on. These are the two fields in our two data sets that have the similar code that we can actually line the data up with. So the rest of this stuff is optional, but you can kind of look through it, I wouldn't worry too much about any of this stuff.

The only thing you may want to change here is this custom field name prefix. This is basically once one joins the fertility data into the country shape file the data from the fertility shapefile or from the fertility CSV is going to be given some sort of prefix by default, you can just go ahead and remove that which I which I am.

I'm just going to go ahead and remove that all together, we don't need that and I'll show you kind of what I mean once I open up the final product, but for now just go ahead and delete that and hit OK. So now you should see this little setting up here join layer with the value, go ahead and hit OK again.

You're not going to see anything happen, but if you right click again on your country shapefile and go back to open attributes table and you go all the way to the right which I'm doing here, you should see some new columns in year. You may not have gotten this far to the right before, but I can assure you that these four columns over here are brand-new, you may recognize this 2016 column, this is from our fertility data remember? but it is now merged into our shapefile.

So every every country, you know, let's take United States as an example, the United States there in the shapefile and not only do we have the data about the United States from the original shape file, but we now have the fertility rate from the World Bank merged into it as well.

And the same is true for all of the countries that are within the shape file. So, that's all we're going to do in this video, but in the next video we're actually going to take those numbers and

make a colored map called a choropleth map based on those fertility rates. So just go ahead and leave it be and open up the next video and keep going.