Case Study 1-1: Converting, Streamlining, and Projecting Well Data for Merced County

Engineering 180
Summer Session A 2022
Created by Madeline Brown
Adapted from Dr. Erin Hestir, Dr. Joshua Viers, and Lorenzo Booth

Assigned: Thursday May 26, 2022
Due: Thursday, June 2, 2022

Skills: Creating Projects, Flat Files, Data Conversion, Projections, Problem Solving, Joins, Summary Statistics, Symbology

Case Study Scenario: You work at a local water utility as their GIS Specialist. The public has raised concern about well contamination, so your manager needs a map made for an upcoming meeting showing any contaminated wells within Merced County.

Your hydrology coworkers gave you two "flat" files that you know you can import into ArcGIS Pro. These files include:

- `gama_edf_merced_index_dms.csv`: The active ground monitoring well locations from the state Groundwater Ambient Monitoring and Assessment program (GAMA). [1]
- `gama_edf_merced_clean.csv`: contaminant information for wells in Merced County, from GAMA. [2]

You also have one Shapefile – collection of multiple files within a folder that all work together to display ONE dataset.

- `MercedCountyCAT`- a zipped folder containing Merced County boundaries in polygon format in California Teale Albers projection [3]

1. Convert from DMS to DD in Excel

a. First you need to convert the WELL locations from degree-minute-second (a notation that is convenient for humans, using paper maps) to decimal degrees (a more machine-readable format).
b. Use Excel to complete this conversion from DMS to DD. HINT: You can split data up into columns then "reassemble."

```
1° = 60' = 3600"
1' = (1/60.0)°
1" = (1/3600.0)°

So, 37° 17' 17.3608" N equals
37 + (17/60) + (17.3608/3600)
= 37.28815578 degrees
```

There are 6 s.f.'s in the seconds measurement, so the correct representation would be 37.2881

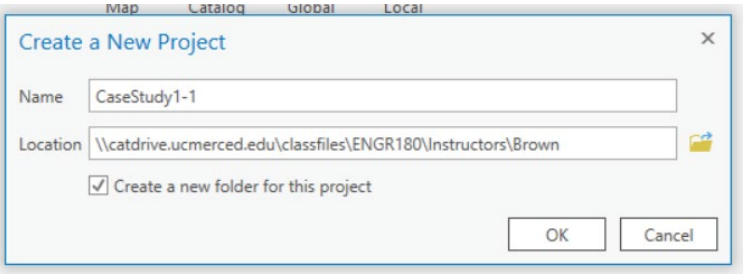
Refer to the box above for a note about the conversion. For this exercise, retain as many bits of precision as possible, but do think about what the appropriate number of significant figures would be.

c. Save the converted file as a new CSV with three columns: well name, longitude, and latitude (in decimal degrees). Name your CSV – `GAMAWellsDD`. We will use this data shortly. Send this file to yourself so you can download it to your Classfiles folder on the remote server.

2. Prep Remote Server and ArcGIS Pro

a. As always, log into Classfiles first!
b. Download the 3 datasets mentioned above to a folder in your Classfiles called `CaseStudy1Data`. Always bring over all project critical data first. It's frustrating to have to do mid-project! Unzip any zipped files, or else they won't show up in Pro. Be sure to download your `GAMAWellsDD.csv` too!
c. Open ArcGIS Pro and sign into your ArcGIS Account (top right corner, your regular DUO info)
d. Be sure to use "Your ArcGIS organization's URL" section. See Remote Server instructions if you need a refresher.
e. Select *New – Blank Templates – Map*
f. Save your project as `CaseStudy1-1` to your root (last name) Classfiles folder. ArcGIS Pro will automatically make a folder with the same name for you!
g. Every time you run a *Geoprocessing* tool, it will output to this location. Make sure the box for *Create a new folder* is checked.

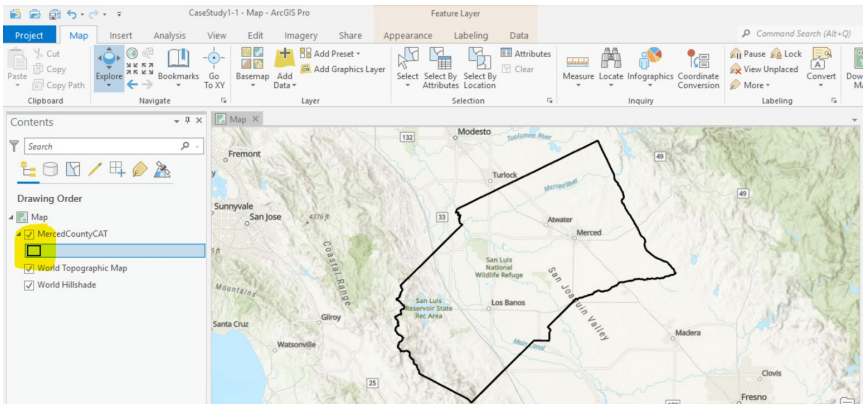
Important! Don't use spaces or special symbols when naming GIS related files!! EVER. Dashes and underscores are OK.



3. Opening shapefiles in ArcGIS Pro

ArcGIS Pro does not open files like most programs that you may be familiar with. To open a shapefile, first you must make a *Folder Connection* with the directory where your data reside.

a. Create a folder connection to the folder that holds the data you downloaded from CatCourses. You can create folder connections under the *Add Data* button, found in the *Map* Ribbon, or through the *Catalog* pane on the right of the screen by right clicking *Folders* and *Add Folder Connection*.
b. Drag and drop MercedCountyCAT, into your map window in ArcGIS Pro. Experiment with rearranging the order of the data sets in the *Table of Contents* (the left sidebar with your layers), and experiment with toggling the layer visibility (with the checkboxes in the *Table of Contents*).
c. A Merced County polygon should show up in your map frame. It may have different default symbology. Click on the small polygon in the *Table of Contents* and change the symbology to a solid black outline with no fill.



It is common to download data in a shapefile format or receive data as a shapefile from outside organizations. If you worked for SoCal Edison, you might receive a shapefile of transmission lines from PG & E. Shapefiles are low-tech and can be shared across GIS software. If you as a SoCal Edison employee edit the shapefile, it doesn't change anything on PG & E's side.

4. Confirming Projection Information

a. Right Click on your `MercedCountyCAT` layer and access the *Properties* (at the very bottom)
b. Go to *Source* → *Spatial Reference*

- Screenshot the *Spatial Reference* portion for submission. This confirms our data is in an appropriate California-specific projection.
- In a written component, identify the PROJECTION used and describe why it is appropriate for mapping contaminated wells in Merced County.

5. Using Basemaps in ArcGIS Pro

ArcGIS Pro allows you to display your data as an overlay on a variety of maps provided by Esri's web service. You can change your basemap under the *Basemap* button, found in the *Map* ribbon. Basemaps are useful for orienting yourself or providing context to data. Note: ArcGIS Pro may complain about different projections/coordinate systems when enabling different basemaps. Just make a note of it for now, the reason behind the warning will become clear after you learn more about the program.

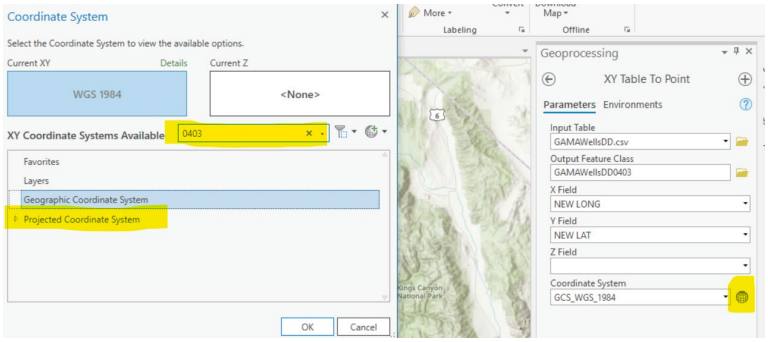
a. Change your basemap to a different one besides the default. Pan Around. Change the basemap back to Topographic.

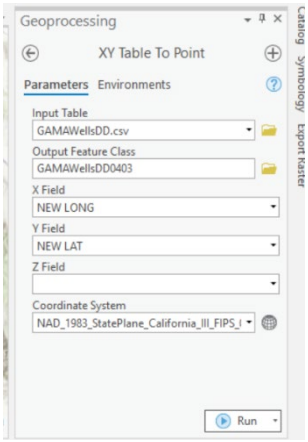
HEY! Have you saved your project lately? Just like with Microsoft Word or other software, it's good to save, save, and save often! It does not automatically save ANYTHING!

6. Import Well Data

a. Use the *Add Data* tool in the *Map* Ribbon to import our flat file of well locations in Merced County.
b. Select *XY Point Data*. This will open a *Geoprocessing* Pane where we'll set criteria.
c. Set your input table to `GAMAWellsDD` – the data your converted from DMS to DD earlier.
d. Name your output `GAMAWells0403`, set your X field and Y field.
e. For the projection information, recall that your coworker said it was in California State Plane 0403. This zone includes the Merced area, so that seemed trustworthy.
f. To find this projection, click the globe icon and search 0403
g. See how the drop-down arrow narrows down to help us find the right Projected Coordinate System?
h. Click through until you find NAD 1983 StatePlane California III FIPS 0403 (meters), then click *OK.*

i. With your input criteria set, click *Run*!





- This will take the "flat file" information and place it onto your map as geospatial data in a Feature Class format. More on that later.
- The layer will automatically be added to your *Table of Contents*. Right click on the layer, and *Zoom to Layer*…
- Take a screenshot of your newly imported points- you'll submit this in a side-by-side comparison after we fix them in the next few steps.
- Answer: Where are the well points showing up in your map display? Where should they be showing up based on the information from your coworker? Can you speculate on what is going on?

Continue on the Next Page.

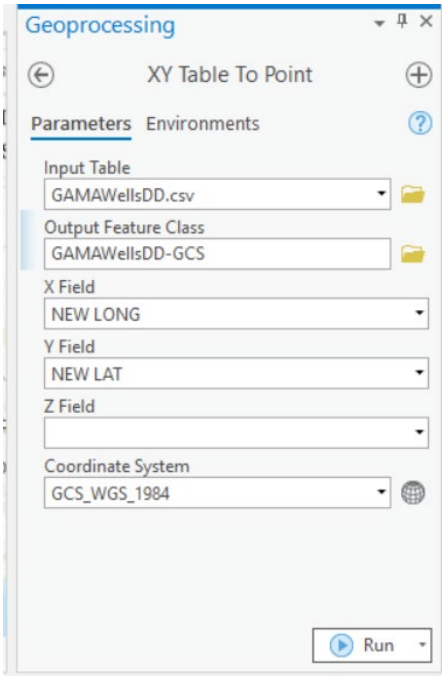
Page break for suspense!

This is a map projection issue. The software used the wrong Spatial Reference (Geoid, Ellipsoid, Datum, Projection, Projection Coordinate System…) when it placed the points.

Projection issues are common, but frustrating! You return to your coworker and ask them about the original coordinates. Ah! They remember that those were NOT projected at all. Those were Geographic Coordinates [Lecture 2 for refresher]. That means you can re-read them in unprojected, then project them to what we want.

7. Correcting the Projection Issue

a. Right click on `GAMAWellsDD0403` in your *Table of Contents* and *Remove*.
b. Right click on the Merced County layer and do *Zoom to Layer* to return us to our area of interest.
c. Re-import our flat file using *XY Point Data* (Steps 7a through 7c). Name your output file `GAMAWellsDDGCS`
d. This time for the projection, accept the default.
e. *Run*!



- Your data should now show up within Merced County!
- Do NOT delete/remove this layer from your Table of Contents.
- Screenshot your newly imported points within Merced County. Submit this screenshot side by side with your screenshot from section 6.

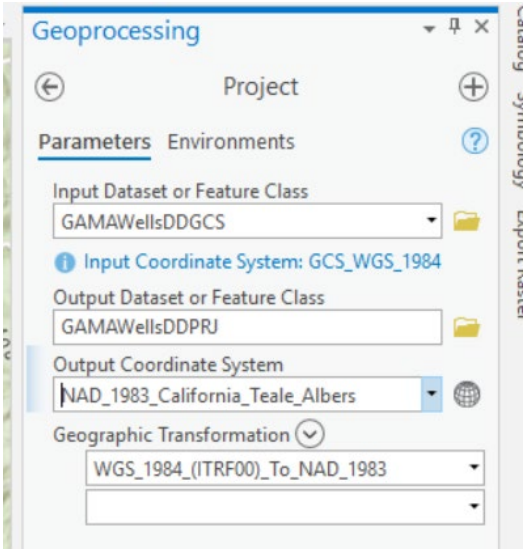
Why not try to adjust them somehow? There are ways to reproject data in ArcGIS Pro. As newbies in the GIS world, it is quicker and better for our mental health to just redo things! It can be very frustrating to try and undo/untangle issues. THIS IS AN IMPORTANT POINT TO REMEMBER! There may be frustrating times in these labs and in other GIS. Sometimes it is easier to just start over from the last place that things ran successfully! This is also a good life skill on occasion as well.

8. Projecting Data: GCS to PCS

We now have two layers – `MercedCountyCAT` IS projected because it was given to us in a shapefile with a preexisting applied Projected Coordinate System of California Teale Albers. The `GAMAWellsDDGCS` layer is unprojected because we read it in with the GCS_WGS_1984 Geographic Coordinate System.

a. Access the Projection information (like steps 6a-b) for GAMAWellsDDGCS
b. Screenshot the projection information. Save this for submission in a few steps.
c. Go to your *Analysis* ribbon and access *Tools*.
d. Search for the *Project* Geoprocessing tool. [pronunciation: PRO-ject]

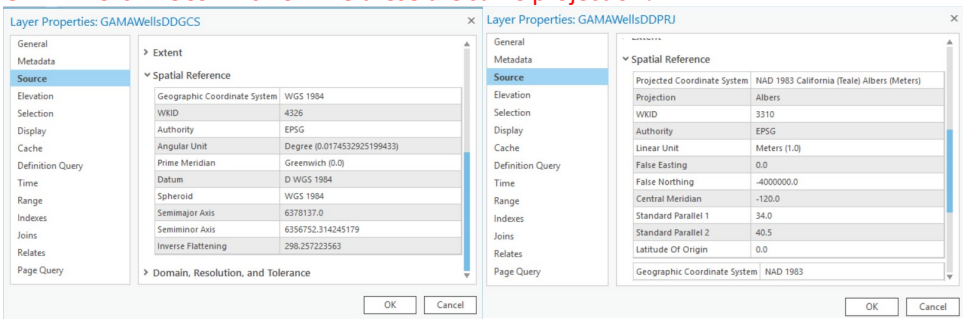
e. Using *Project* will take our data from a Geographic Coordinate System and project it (using the behind the scenes math) into a Projected Coordinate System.
f. Set your input to `GAMAWellsDDGCS`, output to `GAMAWellsDDPRJ`
g. For the Output Coordinate System - we *could* go search for the NAD 1983 California (Teale) Albers (Meters) projection how you did earlier, but there's an easier way!
h. Click the drop down under Output Coordinate System and select `MercedCountyCAT` – it will have the *Project* tool match the projection information of `MercedCountyCAT`!



We always want to have our data layers in the same projection. We want to "unwrap" the 3D earth and flatten it in the same way.

9. Confirming Projections

a. The newly projected Wells should show up automatically in your *Table of Contents*. Right click and *Zoom to Layer*.
b. Things may LOOK like they're in the same place, but the only way to confirm is through checking the metadata for the Spatial Reference information.
c. Screenshot and submit the Spatial Reference information for both GAMAWellsDDPRJ and GAMAWellsDDGCS. Answer: Are these the same projection?



d. Now right click and remove `GAMAWellsDDGCS`. Since we properly projected the data, we need to remove the unprojected data to prevent working on any old layers.

10. Joining Aspatial Table Data

Now that our Well locations are in the correct projection, we want to bring in some external data. In our case, we will JOIN data about well contamination. Remember that when we converted from DMS to DD, we saved the columns of Well Name, Longitude, and Latitude?

We can join tables if we have a "unique id" – a field where the names match PERFECTLY.

First, we want to take our external CSV and import it into our geodatabase. This will generate unique object identifiers (OIDs) for each observation.

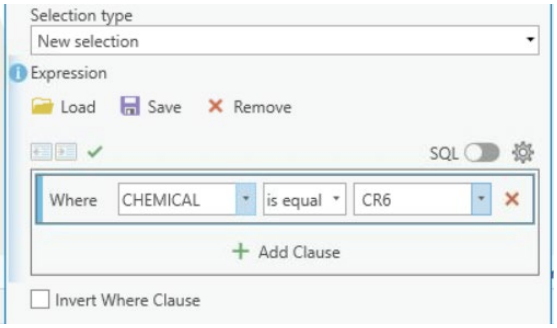
a. Open the *Catalog* sidebar and navigate to your project geodatabase.

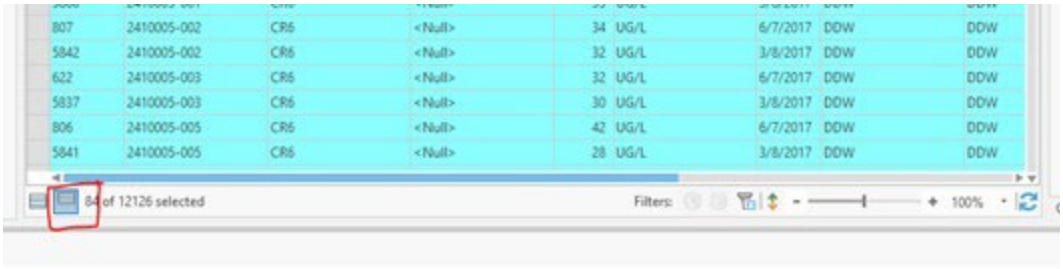
b. Right click the geodatabase and select *Import -> Table.*
c. This will open a *Table to Table* window in your Geoprocessing panel on the right.
d. Set your Parameters to import the gama_edf_merced_clean.csv.
e. Name your output GAMAWellsData. Click *Run*.

11. Operations on tabular data in ArcGIS

Let's start by summarizing the well information by computing mean, min and max concentrations for hexavalent chromium (CR6). Information can be found on the meanings of the chemical abbreviations and maximum concentration levels (MCLs) on the GeoTracker GAMA home page [4].

a. Right click on your new `GAMAWellsData` table and open it. [If it isn't in your *Table of Contents*, drag and drop it in from *Catalog*]
b. First let's select the entries matching the contaminant of interest from the well info table. Click on the *Select by Attributes* button in the ribbon toolbar up top.
c. We want to select all entries from our well table where CHEMICAL = CR6. Use +New Expression to isolate all of the Wells that have CR6 using the drop-down options.
Press *OK* when you are done.
d. In your attribute table, click on the little blue stacks at the bottom so that it switches our displayed rows to just those we selected with the query above.





e. Next, let's summarize the RESULT column by the WELL NAME column. Right click on the WELL NAME column header and navigate to *Summarize.* Since we switched to our selected query rows, it will summarize on just those.

This will open a Summary Statistics pane.

f. We want to retain the CHEMICAL and the WELL NAME field, so those are our Case Fields, and we want to compute the statistics of minimum, maximum, and average (mean) from the RESULT column, so those are our Statistics Field(s). Click *Run.* This adds a new table to your Standalone Tables.

12. Attribute joins in ArcGIS

Now, we are ready to join this information to our well location layer. Right click on your well location layer in the Table of Contents.

a. Right click on `GAMAWellsDDPRJ` and access *Joins and Relates -> Add Join…*
b. In the subsequent window, perform the join operation based on the WELL NAME field in both your layer and your summary statistic table. By selecting *"Keep all records"*, you will retain well

locations that do not have a matching summary in your summary table. ArcGIS Pro will represent these non-existent pairings with a null value.

c. Click *Run*.

13. Exploring Symbology

In our next lab we'll learn additional tasks based off work begun here. Also, we'll begin to create production quality maps. Screenshots, like we've submitted today, are *not* maps! Creating production quality maps is fun, and by doing it properly following cartographic conventions, you'll be able to effectively communicate with viewers, such as Merced County residents concerned with well contamination. But for now….

We want to symbolize the well locations based on levels of contaminants.

First, watch: *Visual Resource:* (106) Symbolizing Data in ArcGIS Pro - YouTube

Then, watch: this short video by Esri about normalizing data. Normalizing and Classifying Choropleth Maps - Esri Videos: GIS, Events, ArcGIS Products & Industries. Also visit Esri's Data Classification Methods for supplemental information as needed [5].

Briefly summarize what it means to normalize quantitative data. List the seven data classification methods. Then, describe how you could apply normalization to the CR6 measurements.

a. Try customizing the layer symbology based on the column corresponding… to average (mean) concentration for hexavalent chromium. Take a screenshot.
b. Customize your symbology based on minimum concentration for hexavalent chromium. Take a screenshot.
c. Customize your symbology based on maximum concentration for hexavalent chromium. Take a screenshot.

Submit above 3 screenshots of varied symbologies using the classification methods of your choosing.

Good work! You have reached the end of Lab 1 (Case Study 1-1).
MAKE SURE TO SAVE YOUR PROJECT WHEN YOU EXIT!
We will use this project for Case Study 1-2.

Deliverables: Submit all listed deliverables below in a SINGLE .pdf document on Catcourses.

Section 4
- Screenshot the Spatial Reference portion for submission. This confirms our data is in an appropriate California-specific projection.
- In a written component, identify the PROJECTION used and describe why it is appropriate for mapping contaminated wells in Merced County.

Sections 6/7
- Answer: Where are the well points showing up in your map display? Where should they be showing up based on the information from your coworker? Can you speculate on what is going on?
- Side by side screenshots of incorrectly projected (ocean) and unprojected well locations.

Section 9c
- Screenshot and submit the Spatial Reference information for both GAMAWellsDDPRJ and GAMAWellsDDGCS.

Section 13
- Briefly summarize what it means to normalize quantitative data. List the seven data classification methods. Then, describe how you could apply normalization to the CR6 measurements.
- Submit a screenshot of 3 side by side different symbologies

References

[1] caesar0301/awesome-public-datasets. “A topic-centric list of high-quality open datasets in public domains. By everyone, for everyone!” https://github.com/caesar0301/awesome-public-datasets

[2] jdorfman/awesome-json-datasets. “A curated list of awesome JSON datasets that don’t require authentication.” https://github.com/jdorfman/awesome-json-datasets

[3] Madeline Brown. Merced County Boundaries for ENGR180. Isolated from publicly available State of CA Counties layer via ArcGIS Online.

[4] State Water Resources Control Board. “GeoTracker GAMA: Download GAMA Data.” http://geotracker.waterboards.ca.gov/gama/datadownload

[5] Data Classification Methods. Esri. https://pro.arcgis.com/en/pro-app/2.8/help/mapping/layer-properties/data-classification-methods.htm