Assignment 3 write-up: Further Analysis
Project: Land Use Change in Kern County, CA
Michelle Williams

This assignment represents a change of direction from my initial idea to explore land uses correlated with bare soil in Kern County as potential areas for *Coccidioides* growth and transmission. *Coccidioides* is a genus of fungi endemic to arid and semi-arid regions of the Western Hemisphere and is responsible for annual outbreaks of pneumonia-like disease called Coccidioidomycosis, or Valley Fever. Infection occurs when humans inhale airborne fungal spores that arise from soil disturbance.

In early April, my lab group refined the research question to look at land use change in southeast CA from farmland to non-farmland as a potential risk factor for coccidioidomycosis. I decided to change the direction of this project so that my end product could be used in that analysis.

Using Kern 2016-2018 as a test case, I worked out a method that identifies areas of change using raster analysis. The resulting raster can then be converted back into vector-type spatial data to get the latitude and longitude at each 500 by 500 meter 'pixel' of change.

Successes:

I had to do a lot of online research to understand and implement all the steps for raster plotting and analysis in R, but it worked! There are still some things to figure out, but I am very proud of myself that I got the backbone of the workflow down. I was also happy that I could use the more simple data from the CA Department of Conservation instead of having to process several gigabytes worth of remote imagery data, which my personal laptop probably doesn't even have the computing power to do.

Challenges:

I originally tried to use Python, but found more resources for what I was trying to do in R. I also tried to do the analysis in a JuPyter Notebook in DataHub once I settled on coding in R but had trouble figuring out how to install the packages I needed for this analysis. As a workaround I ended up working in R Studio on my desktop to create the markdown file in my GitHub repository.

While the workflow and dataset is fine for the scope of this project, one of the major limitations of this analysis is that the spatial resolution is at 1km or coarser (the documentation for CA Dept of Conservation mentioned that the land use maps were at 4km spatial scale but rasterizing at 4km didn't capture all the detail of the polygons). Using 30 meter LandSat imagery data on vegetation indices and bare soil would give better spatial resolution for small-scale changes in land surface, but that requires a lot more time, analysis skill, and computing power than I currently have.

<u>Next Steps:</u>

I still need to figure out some details, like translating the NAD83 spatial information to a more universal CRS like WGS84, and learning how to customize maps with R's base plot() function and ggplot(). I would like to also be able to summarize the area that changed from farmland to other land uses, and the surface area at each cluster of 500x500 meter pixels (e.g. a 2-cell by 2-cell cluster of 500x500 cells would be 1 square km if I'm doing my math right). I will also look at LU change across other years since the changes 2016-2018 may be too sparse to analyze as risk factors for disease.