Narrative

Maxwell Reikosky

12/10/2020

1. Brief Substantive Background / Goal

I had two goals for this project; 1) Create a map that shows the extent of migration and displacement in countries with high climate vulnerability; 2) Use relevant datasets to make a 3D and/or moving model in the form of a gif/short video. I felt that particularly in the case of measuring bodies, a 3D model would be compelling (since bodies take up space, as opposed to color hue or numerical values). My original intent was to present these two concepts (migration and climate vulnerability) together in a single visual. I wanted to make the resulting visualization simple and aesthetically pleasing to highlight the dynamic, striving for a strong immediate impression over communicating a great deal of information. After a little research on the web, I found a 3D modelling package called rayshader that is designed to take datasets/ggplots and make rather attractive models (https://www.rayshader.com/). I decided to take a crack at applying my comp tools skills and experience to this new package and see what cool visualizations could result.

2. Collecting Data

Accurate and robust data on migration is hard to come by, mostly because such a large percentage is informal and goes unreported/analyzed. I found available migration datasets from the IOM and Princeton but they didn't contain comprehensive enough information. As a result, I decided to look at refugee populations instead, which tend to interact with clearer reporting mechanisms. I quickly found a downloadable World Bank dataset online with refugee populations by country throughout the world from 1960 - 2019. I then did pretty broad searches on climate vulnerability, not totally sure which aspect of vulnerability I wanted to look at. There are many datasets online so I wanted to find something that would give forecasts for the future to account for climate change. I ended up finding a bunch of downloadable datasets through an independent non-profit organization called DARA. They do some globally scaled work on measuring climate and I chose a climate vulnerability dataset that included projections to 2030. Since I was making a map of refugee population, I thought the column on climate mortality would be a solid complement, allowing the 3D model to show bodies in an alternative (negative) way. Both of these datasets were in excel format and appeared comprehensive enough to make a global refugee-climate vulnerability map.

3. Cleaning / Pre-Processing Data

This project required the cleaning, pre-processing, and combining of three datasets in total. The third dataset provided information that rendered a global map when run through ggplot. For this global map-making aspect of the project, I followed the step-by-step guide of a blogger (https://www.r-bloggers.com/2017/02/how-to-make-a-global-map-in-r-step-by-step/). This blogger showed me how to download the map dataset, process my other datasets of interest, cleanly merge the datasets, and then render a ggplot visualization. Part of this task required recoding many country names in my refugee population and climate vulnerability datasets to match the country names of the map world dataset. I basically had to look through especially the refugee population dataset to find sort of oddly phrased country names like "Dem Rep. of Congo, The" and manually replace them with "Democratic Republic of the Congo." Otherwise, the process of cleaning datasets was simple. I isolated the country names and 2019 refugee population columns in my refugee population dataset, as well as the country names and 2030 climate mortality columns in my climate vulnerability dataset, and sequentially joined these into my world map dataset. I did have to change the column year name for the

refugee population dataset since it originally read as "2019" which would not allow me to use the select function. I also shortened the column names on the climate vulnerability dataset, since they included non-stringed information in parentheses. After this, the data was ready to roll!

4. Analysis and visualization

After cleaning the data, the 2D ggplot went smooth. I made a map for each respective dataset with higher values represented by color density. I did notice that in both cases, 1-2 countries were magnitudes higher than the others, so the maps did not have much diversity. I then made additional maps for both datasets with these outlier countries removed, and this added more geographic spread of information. I figured I would make elevation in the 3D model equal one of the dataset values (ref pop or clim mort) but did not know which one would be the best base, so I made maps of both.

Using rayshader was a different adventure. I could at first not get the package to install but learned that it had to do with the common installation library and that I could only open the package properly as an administrator. I then followed the instructions near the bottom of this page (https://www.rayshader.com/) to drop my ggplot into the rayshader 3D plot function. Given the information on this page, I did not learn how to make a map where color represented one variable and elevation represented another. Instead, it looked like I could only make redundant information of color and elevation. After adjusting the width, height, angle of the model tilt, and angle/zoom of the camera snapshot, I found that having too much information on the model muddled things up. So, I stripped it down, removing the legend title, changing the color scheme, and shortening axis labels to highlight the stark visual effect of the data.Happy with the result, I tried to follow through on my original plan. First, I wanted to make a gif from my models. I found this guide to be on the right track (https://wcmbishop.github.io/rayshader-demo/) but sunk many hours without success. I think this stems from two problems with rayshader and my approach: 1) Very little documentation was available online about rayshader. There was maybe one stack overflow question on ggplot gifs, and it was not particularly helpful. Rayshader seems like new software that is used by niche R groups. So finding additional guidance was tough and I spent a lot of time winging it and experimenting. 2) The instructions I found seemed to only work with piped down elevation dataset strings. This is at least my conclusion because I tried to follow all of the steps but consistently was thwarted. I could not assign my ggplot based 3D model to a string, and without such a string, my work seemed incompatible with the rayshader gif making process. I believe there is a way to make it work, but I ran out of time and energy and wanted to ensure what I had was quality. My other goal was to make a label as seen again at rayshader.com. I think I ran into the same set of problems and so ditched the effort.

Okay finally (sorry this is so long) I could not get these models to successfully knit to pdf. BUT THEN ROCHELLE HELPED ME. There was little help available online. I found this (https://github.com/tylermorganwall/rayshader/issues/61) which I unsuccessfully tried. I received an error saying that LaTex could not knit the file. I followed the debugging instructions it gave me here (https://www.rdocumentation.org/packages/rgl/versions/0.100.54/topics/rglIds) and received the same error. It was at this point that I talked to Pete during his office hours. We messed around with the rgl package, but again were unsuccessful. He suggested I screenshot my printed 3D models and insert them directly into the markdown as pictures. Not the optimal outcome but I accomplished the fundamentals of my goal and feel I gained some solid independent experience while flexing my new R skills. For analysis, the models allowed for recognition of certain dynamics, such as Pakistan (and subsaharan Africa) having both high refugee population and projected climate mortality.

UPDATE: Rochelle suggested I knit to HTML instead of pdf. Though I tried this earlier on in the debugging process and it did not work, upon her suggestion, I applied recent debugging code (opening and closing the rgl device and widget) and got the 3D models to knit. For some reason my first model, however, still will not show up in the knit.

5. Future work

If I were to do this project again, I would perhaps use a different package since rayshader has scarce and difficult to parse public info. That said, its visualizations are eye-catching, smooth, and may be useful when I

have elevation datasets. I am most interested and engaged by the visualization component of coding. I want to delve into geospatial mapping, especially as I study migration and climate converge over time. I hope to work for policy think tanks in the future and feel the capacity to handle datasets and generate visuals will be useful. After getting a stronger hold of R, I hope to learn other general languages as well, perhaps Python, to expand my fluency in data visualization and even artistic virtual imaging.