Artur Genser
Project 1 Writeup (This is being turned in late because I got an extension to 5/17)

Context and Purpose

The basic question at hand is what is the relationship between right-wing politics and wealth in Poland? I chose to explore this because in 2015, the party Prawo i Sprawiedliwość (PiS), or Law and Justice in English, took power in Poland for the first time since the party's inception in 2001. It won, in part, with its support from the working class. My project intends to highlight this relationship with a set of visualizations.

Pre-Processing

Very minimal pre-processing was necessary. First I downloaded the election data from https://wybory2018.pkw.gov.pl/pl/index and then downloaded income data from a visualization I found at http://biqdata.wyborcza.pl/biqdata/7,159116,23750826,ocean-biedy-z-wyspami-bogactwa-tak-wyglada-mapa-dochodow-mieszkancow.html. I then used the identification codes for each commune to combine the data. Some manual combination was necessary because Warsaw is split into districts for the election and one other commune did not have election data (I think because they held elections at a different time). From there, I deleted the columns I felt were unnecessary and kept the ones I felt I might use. I then saved that as .csv file and used an online converter to convert it to a .json file because that is how the project scaffold imported the data and I decided to keep it simple.

How Does It Work?

The main visualization is a map of Poland. The size of the commune is encoded as the size of the dot, and wealth and support for PiS is encoded by a function in the LAB color space. The yellower the dot, the wealthier and less supportive, the bluer, the poorer and more supportive. The darker, the closer to being less support and poorer. I decided to make this the centerpiece because it not only shows the geographic trends, but also allows the viewer to see an overall trend among the population. The dot size helps discern which areas have more people. This has the unfortunate side effect that there is a lot of overlap. The overlap was a cost, but I felt it was worth the reader understanding the population density. Also, to mitigate this effect, I chose to place the dots in population order, so that large dots with potential to obscure were placed behind smaller dots. The color scheme was chosen because 1. blue and yellow are opposite colors as well as being the most colorblind friendly 2. the color is along the trend and allows the viewer to see which communes fit into the trend of being wealthy and not supportive or poor and more supportive. However, because so much data is being encoded (there are thousands of commune data points, each one encoding geographic location, population, income, and support for PiS) I included seven other, smaller visualizations to make this largest one more legible.

The first is a scatterplot. This gets rid of the geographic component and allows the viewer to easily see the formula behind the coloring. I struggled with weather or not to include a key for the main map, but including this graph acts as a key in way that adding a key feels redundant. I decided to include tick marks on the axis, but only labeled the y-axis. This is because while PiS support is very definitive, the data for income is abstract. It is being estimated by income tax collected, and then is also in a foreign currency, so knowing specific numbers is not very helpful.

Next are the six maps off to the side. First, I isolate the variables of income and support to make their individual trends more visible. Next, I combine the data into larger jurisdictions (counties and provinces) to highlight the larger regional trends. Lastly, I separate the largest 20 communes from the rest. This helps with the overlap problem a little, while also giving me the opportunity to label the largest Polish cities for an American audience that would be unfamiliar with Polish geography. Lastly, though not a visualization of data, I included a map of Poland in the upper right-hand corner to give some more geographic context for an American audience.

Alternatives and Arbitrariness

Plenty of alternative choices could have been made about coloring. I could have chosen a different pair of colors or a discrete color scheme instead of a gradient one. The axis of the colors could have also been different (blue and yellow moved by 90 degrees in either direction). I also experimented with different variations on how light or dark to make the colors or how sudden to make the gradient. Some of these decisions were reasoned with what made the data and trend I wanted show most visible, but others were simply a decision of what I felt looked best.

Other alternatives could have been to not encode population as this is not the trend being explored. Or one could have chosen to not encode geographic location at all, and only explored the data along the two variables of the trend. This would take away the geographic data that the reader can garner, but perhaps would allow the visualization to explore the relationship between the variables a bit more (maybe a regression, some more graphs, some generally more in depth data analysis).

The specific layout, where to put the maps and how big to make them, was almost entirely up to me. However, the maps and graph themselves were all data driven, though deciding how to make latitude and longitude into a flat map is somewhat arbitrary.