Brendan Sigale
Introduction to GIS: GEOG:1050:0A03
Final Project Report

A Geographic Analysis of the Relationship Between Poverty and Natural Disasters

Introduction

In 1979, the Federal Emergency Management Agency (FEMA) was founded by two Presidential Executive orders. In addition to their roles in managing emergencies and providing relief from natural disasters, they also began collecting data about all known natural disasters, both past and present. This has allowed natural disaster scientists to gain an insight into where natural disasters happen, why they might happen, and how it relates to other variables. This report will explore the relationship between median household income, percentage of a county population in poverty, and propensity to have a natural disaster. It will do this based on the total amount of disasters each county has seen, but will also analyze it on a yearly basis to determine trends over recent years to determine if there are any observable trends in who lives in which counties.

This report will be useful for a multitude of fields, such as economists, risk analysts, and natural disasters scientists. By understanding this relationship and any trends, people in these professions will be able to more accurately provide their customers, such as counties or states, with accurate information, advice, and rates to be ready when disaster does strike. In addition, those in regions prone to disaster would be able to analyze their financial situations and plan to relocate to regions where they are less likely to be struck by disaster to ensure more financial stability.

Data

The data used for this project comes from two sources. FEMA has all of their data publicly available, containing their most up to date information on all disasters in a convenient .csv file which contains where the disaster happened, when it happened, how long it lasted, and what type of disaster it was.

https://www.fema.gov/media-library/assets/documents/28318

The second piece of data comes from the US Census Bureau's Small Area Income and Poverty Index, which hosts a dataset containing over 75,000 rows of data containing the median income for each county by year from 2017 to 1989, excluding 1994, 1992, 1991, and 1990.

https://www.census.gov/data-tools/demo/saipe/saipe.html?s_appName=saipe&menu=map_proxy&map_yearSelector=2017&map_geoSelector=aa_c&map_yearSelector=2017&map_geoSelector=aa_c&s_year=2017,2016,2015,2014,2013,2012,2011,2010,2009,2008,2007,2006,2005,2004,2003,2002,2001,2000,1999,1998,1997,1996,1995,1993,1989&s_measures=aa_snc&menu=grid_proxy&s_inclUsTot=n&s_inclStTot=n

This data was used in ArcGIS along with a shapefile used in Lab 3 which contained the boundaries of all counties in the United States and projected Nthe NAD 1983 Zone 15N Coordinate System.

Analysis

All of this data was cleaned and processed using a combination of R, Microsoft Excel, ArcGIS, and an online GIF maker. These programs and the processes completed within them allowed for detailed analysis by creating summary tables, easily and automatically separating the data, and creating several bivariate chloropleth maps. These tools and methods all combined to allow for detailed data processing, analysis, and studying for this report.

Data processing began in R with similar steps taken for both sets of data. The poverty information came complete with, among many variables, the county FIPS code, years, median household income, and percent of the county population in poverty. These columns were extracted from the rest of the dataset for analysis. Two sets of data were gathered - the average percentage of the population in poverty and the average median income per county over all years in the dataset as well as individual files containing the information from all counties broken up by year.

A similar set of steps was used to process the FEMA disaster information. FEMA uses a unique location signified which is not consistent with easier to read FIPS codes. Instead, they log the state as well as the 3 digit county code prefixed by 99, with a whole new number made up for data which is not in a clear county. Thus, data had to be processed in R to combine the county codes with accurate state information. Using the 'cdlTools' package in R, the state names were converted into a new column. The last 3 digits of the FEMA region codes contain the county, so they were extracted using substring. The state and county codes were then pasted together to form a complete FIPS region code. The data was then cleaned to only have columns signifying the FIPS code, type, length, and year of the disaster. A sample of the code is shown below.

```
##Separated by Year
       path_out <- "./DataByYear"
       for(i in 1989:2017) {
         tmpdf <- subset(df, year==i)
         write.csv(tmpdf, file = paste(path_out, i, ".csv", sep = ''), )
       }

## Grouped by region, averages over all years in dataset.
       df <- df[complete.cases(df),]
       df <- group_by(df, region)

       dfSumm <- summarize(df,
                           avgPov = mean((pctPoverty)),
                           avgIncome = mean(medianIncome))

       write.csv(dfSumm, file="AveragesByCounty.csv")
```

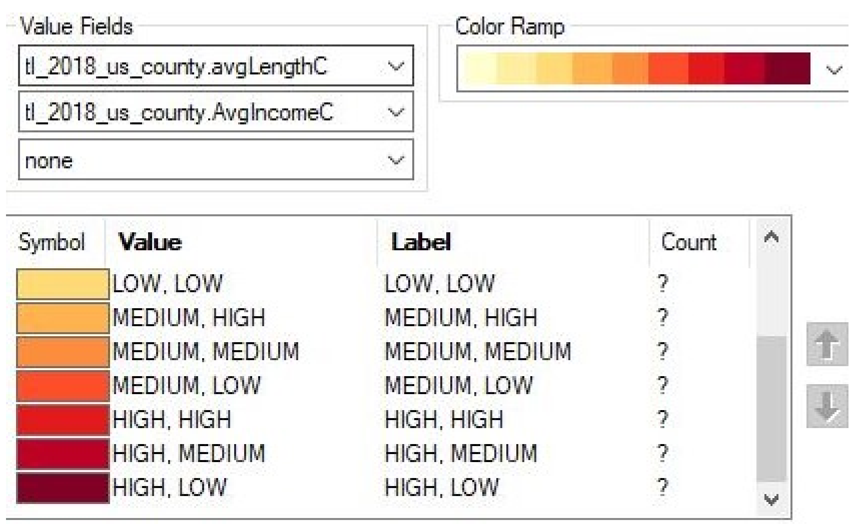
This data was then imported into ArcGIS. In order to make the necessary bivariate chloropleth maps, the data had to be joined together. All sets of data were joined on their County Codes, which in the county shapefile was called GeoID and in both the FEMA and poverty datasets is called region. Using this data, two styles of maps were created. The first type shows both the relationship between the length of disaster and median income or poverty rate in each county. The second style of map shows the relationship between the length of disaster and poverty rate per county per year.

After the data was imported and joined, it was classified based on the length of disaster and the poverty rate. The breaks of the classifications are as follows:

```
Length of Disasters
       "Disasters In 2010.csv.length" <2573
       "Disasters In 2010.csv.length" >= 2573 AND "Disasters In 2010.csv.length" <=4012
       "Disasters In 2010.csv.length" > 4012

Percentage of the Population in Poverty
       "DataByYear2010.csv.pctPoverty" <= 10
       "DataByYear2010.csv.pctPoverty" > 10 AND "DataByYear2010.csv.pctPoverty" <= 20
       "DataByYear2010.csv.pctPoverty" >20
```

A color ramp was chosen from the Colorbrewer Style set provided in early labs. The chosen color ramp ranges from a pale yellow to deep red. While this can occasionally be difficult to interpret for minute data, it was chosen to represent that even if there is a low poverty rate or if there are a low amount of disasters, there are still people in poverty and the county is at risk of a disaster. A sample of the color ramp is below (the entire ramp did not fit in the window).



These maps were then meshed together through an online gif maker to show trends in the data and the relationship, spaced at 4 seconds per year and 12 seconds on the slides showing the overall relationships at the end.

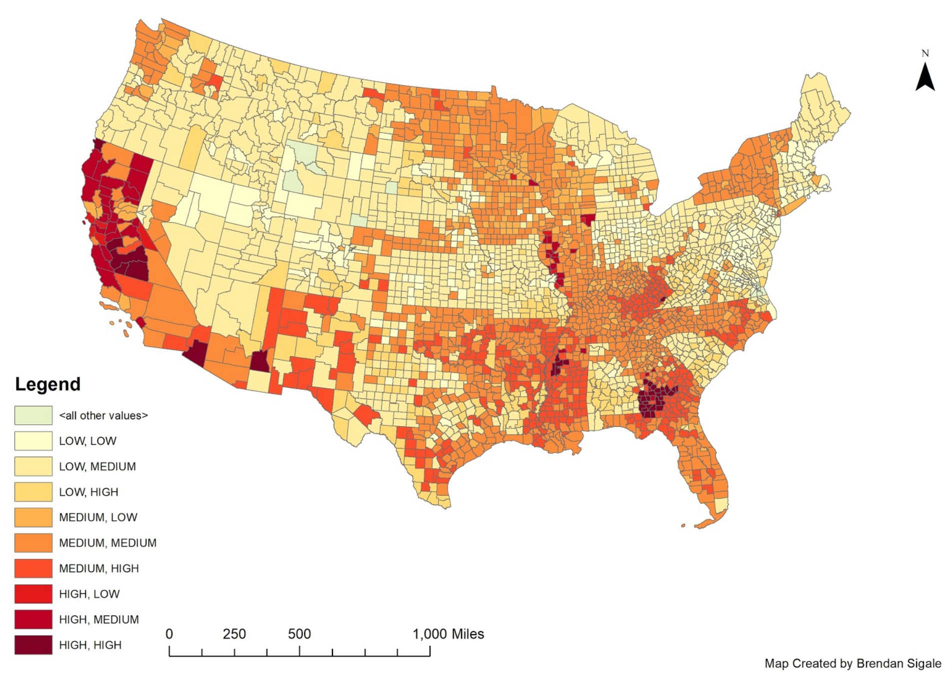
The legend for all maps is difficult to read due to the legend styling in ArcMAP. The order of variables is [LENGTH OF NATURAL DISASTER], [PERCENTAGE OF THE POPULATION IN POVERTY].

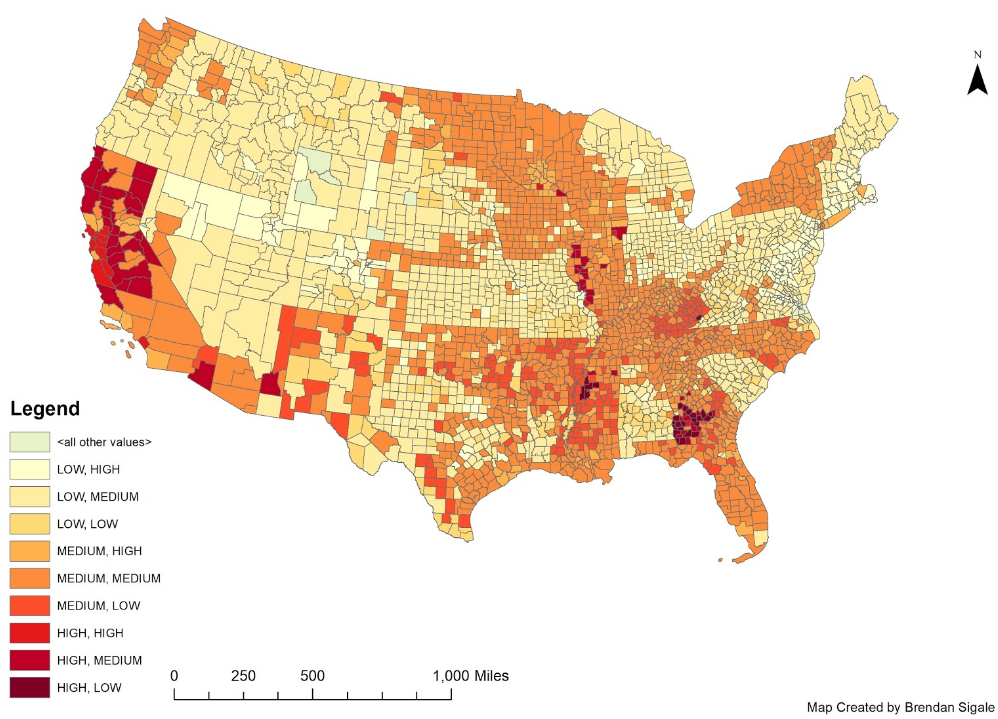
Results

This data speaks volumes about the relationship between economic factors and propensity for a region to experience a natural disaster. Between both of the normal maps as well as the time series data, scientists and economists will be able to glean better insight into what income levels tend to live in disaster-stricken areas.

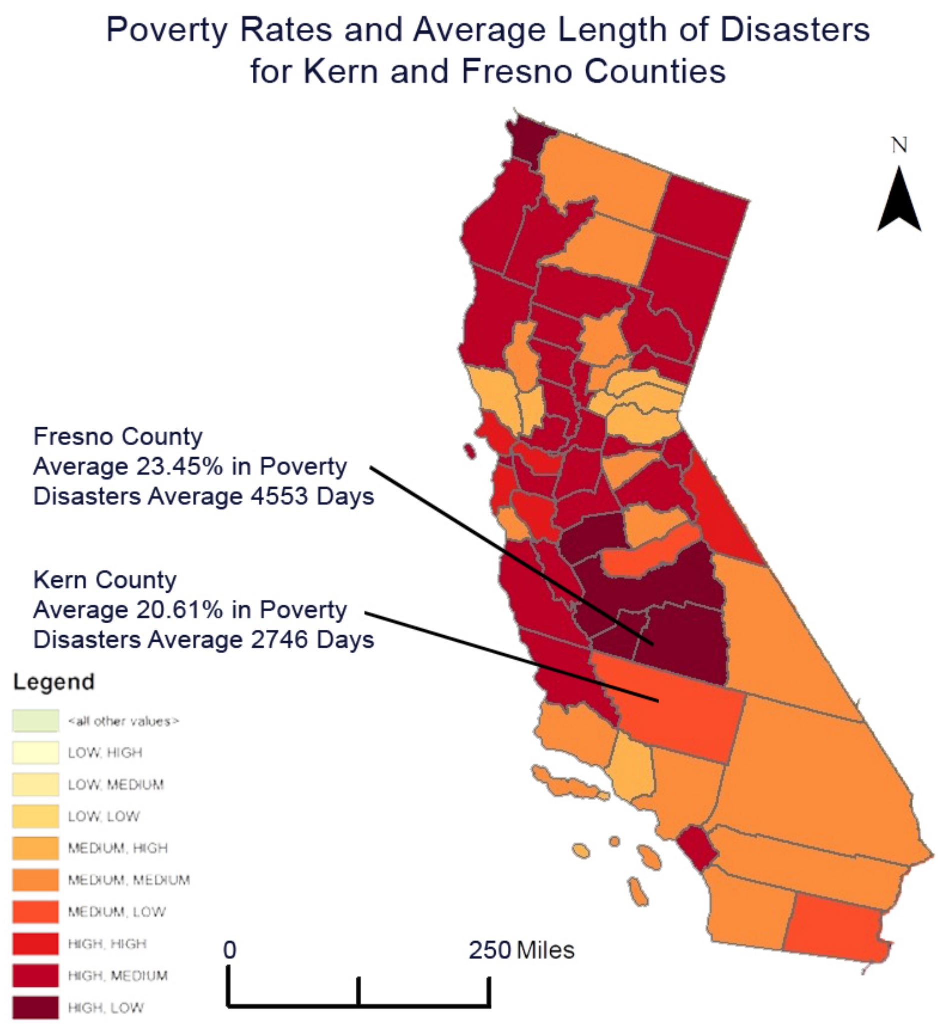
In the normal maps, we can see the relationship between the length of natural disasters, the percentage of the population in poverty, and annual income. The maps and data demonstrate how most disasters have a medium length and affect people with a median income. In addition, many of the counties which experience medium or high amounts of disasters are low or medium income. Counties which are high income or have low poverty rates are rarely found within areas which frequently experience natural disasters.

Relationship Between Length of Disaster and Poverty in the United States





California is an especially interesting state when comparing these variables. California experiences the most amount of disasters but also has a consistently high income across all counties. California experiences a disproportionately high amount of disasters compared to all other states, yet has an abnormally high average median income and a low rate of poverty. There are, however, hotspots of high poverty, such as Fresno County, which is informally known as the "meth capital of the world". Fresno County has the second highest poverty rate in the nation, only second to Kern county right below it. Kern, however, has experienced fewer disasters than Fresno, likely due to its surrounding topography. Kern is county (identified by Bakersfield) is surrounded by mountains on almost all sides, whereas Fresno county is open to the Pacific. Other than these counties, however, everyone in California is experiencing significantly more disasters than the rest of the country.



The time series maps are extremely insightful. They demonstrate a clear trend of the United States experiencing less natural disasters overall, but more poverty. The number of disasters was extremely high up until 2005, at which point there was a severe drop in the number of disasters per year. However, as the housing crisis began in 2008, the map shows the strong climb in poverty rates nationwide and their eventual slow recovery in recent years. While the GIF cannot be imported into this document, it is viewable at the link below.

https://imgur.com/a/63jKtqg

Using time series maps, the effects of enormous disasters such as Hurricane Katrina can be assessed. While Katrina does not show up as a high disaster length in the data for an unknown reason, the poverty data can be used to analyze the hurricane's effects. In the years following Katrina, the maps display a shifting in the population to be more impoverished, pushing some counties into the next category of poverty. Contrariwise, in the years following Hurricane Ike's landfall in 2008, most of the region stayed within its economic category. This lack of change compared to Louisiana indicates the higher economic stability found on the coast of Texas, likely due to there being no large cities along the Texas coast which would be affected by a large storm.

Conclusion

These maps answer a relatively complex geographic question: what is the relationship between income levels and where natural disasters happen? The question is simple on the surface, but when analyzed across each county in the United States, becomes much more complex. Through the maps created, it can be determined that disasters tend to strike a wide range of groups across the United States. However, there are clear, consistent patterns as to who lives in those regions - those who are in poverty. While this relationship is not mutually exclusive, there are great swaths of people who are impoverished who would likely live in an area less prone to disaster if given the choice. The cause of this is not known, nor is it certain that this 28-year range encompasses a representative portion of history. However, these maps do show the rise and fall of overall poverty levels as well as the changes in location and frequency of natural disasters.

I believe these maps will be useful to risk analysts and economists to determine trends in poverty across the United States and how natural disasters affect them. By using the time series maps especially, they can analyze the trends in these effects to plan for disasters, help consumers save money on insurance, and more effectively prepare for disasters.

Were this project to be completed again, I would likely choose a different way to summarize the data. Using the length of a disaster is not entirely representative of the severity of a disaster. Some regions, such as Louisiana, experienced what is arguably the worst storm in modern history, yet the FEMA data does not accurately represent this. I would also change the data processing method, as it inexplicably caused errors such as some 1996 simply having no financial data, meaning the year had to be skipped and decreased the accuracy of any trend analysis. This project would have been significantly easier had more time been spent learning how to use the model builder, as I could have then developed a model to automatically join, classify, and process the data. ArcMAP is not designed for running loops of data processing, but using the model builder would have made it easier, or even using a program such as R with the ggplot2 package to create and export the maps automatically.

I believe the skills I learned in this class helped greatly in my ability to put this project together. In addition to the ArcMAP skills learned during the lab, the lecture has taught me how to properly pose geographic questions and put together full geographic queries.