Aaron Marks
Geog 363-S2 – Final Report

Analysis of Natural Hazards in PA

Some may argue the common saying "Money is everything" being true, or not, but in some cases it can be a determining factor. One of these cases, for example, would be the cost of living determining where a person or a family decides to locate their residence, or perhaps a business. These expenses are quite in depth as you know but the one being pin pointed here involves insurance costs of property valuables like a building or a car. House insurance, as well as auto insurance, policies probably include various coverage options that determine which party, the insurer or customer, pays for the damages, and how much, caused by specific events. The problem that arises when a person is deciding where to live is that these insurance cost estimates are often not available or brought up upon purchasing of a building.

The goal of the analysis is to gather data on the occurrences and intensities of past recorded storms to generate a map representing predicted overall cost of living by county in Pennsylvania. While originally there was going to be more than just storm data included as cost factors these 'outside' contributors would just clutter the map and should be made into their own map. That is, if someone were interested in other areas of insurance coverage. As an example, this analysis would help a person pick out their coverage options dealing with weather related damages and/or accidents while a crime based map would be more detailed towards a theft coverage selection process. Additionally the project was aimed to be done using mostly methods covered in the course and within four weeks of lab sessions, about ten to twelve hours total.

With the original goal in mind during the beginning of the project the first step was to gather the recorded data available for anything related to affecting the price of in insurance. There is a lot of data out there and it is fairly easy to find, it's just a matter of finding what you want. Weather data found included Tornado touchdown points and lift off points, as well as their paths. Also there was data on recorded areas of hail, high speed winds, wind squalls, flooding, hurricane paths, and some others. Other data ranged from being natural such as earthquakes, volcanoes, fires and water quality to society based including crime rates covering theft, assault, murders and so on. Data used was pulled from PASDA, National Atlas and NOAA National Weather Service's Storm Prediction Center. All of the data found could have been used as there was a cost related to event had it occurred to the person selecting an area.

After adding all of the data to a database and importing a lot of it into ArcGIS it became apparent that there was in fact too much data, and therefore time involved, for the original idea of just a final map of a cost estimate. There would be so many classes involved that the map wouldn't serve as a general estimate anymore and would probably need to be at a finer scale than just the county level. While looking through the data it was easy to spot three or four division of data, weather, earth related and people related. With a person interest in the weather data that became the new aim for the project and other data was set aside and not worked with.

As with any project the first thing advisable to do is to setup a database in ArcCatalog to contain all of the data and files generated from it. This also allows for easier navigation for both the user and ArcGIS when 'pointing' to files. Importing data from various sources is bound to require the data to be projected similarly in order for the data's location to be significant. Most of the data came projected with NAD83_GEO, so others that weren't or didn't have any where then set to this projection. This projection info was easily obtained through the metadata visible in ArcCatalog and then easily edited in ArcGIS.

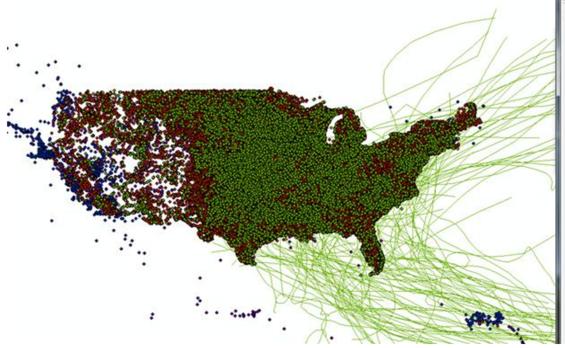
With the data available and formatted for use it was time to determine how much of it was used, regarding both time and area. Time wise I found that of the weather data used, as well as the other natural disasters that weren't included in the final map, all overlapped the time frame of 1955 to 2004 with some having data on both ends as far back as the 1560s. Each data layer was clipped to a common date range of 1955 to 2000 using the select by attribute function and using a selection query such as "YEAR" >= '1955' AND "YEAR" <= '2000'. The name of the variable for year varied but was easily identified. Once that was done it was still apparent that there was too much data and that was due to the availability of data for most of the United States and surrounding oceans. To consolidate the data a little further I decided to focus on just Pennsylvania. This was originally done by selection of attribute but data values such as a tornado that began in Maryland were labeled as Pennsylvania as they crossed the border and spent more time there than where it came from. To solve this, a selection by location was performed for each data set using the Pennsylvania county shapefile's extent. These past selections, both time and area, were then exported as new shapefiles to reduce the processing time of the large data set.

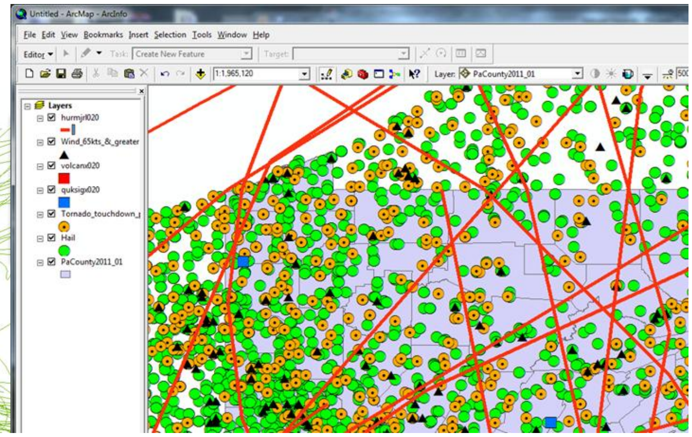
The next major part of the project before anything past visual inspection could be performed was to associate each data point with its respective county. This step involved using the joins feature provided in ArcGIS. As we know this allows for us to join data sets, or parts of a dataset, together to either fill in gaps or provide additional data to existing data. All of the weather data included a county FIPS code as well as a STATE field. With the county level in mind it would make sense to join the data by a field that occurs in both attribute tables, such as the FIPS code. The weather data however didn't contain full FIPS codes, such as 42029 for Chester County here in Pennsylvania, but instead only included the county FIPS of 029 and the State part of it, the 42, as 'PA' in a separate field 'STATE'. The county shapefile contained full FIPS codes on the other hand. Two methods came to mind about how to join the two. The first was to make a new field and generate a Full FIPS code for the weather data by adding the appropriate first two numbers based on a formula, or just remove the state part of the FIPS code from the county shapefile. I wasn't able to figure out how to do that though so the other way was then considered. The second way was to join by spatial location. The first procedure used involved adding county data to each point, but later realized I wanted to join the other way around. It was a learning experience even though I could have saved a lot of time had I tried the third method first. When combing the data spatially ArcGIS prompted me with some possible new fields to be generated during the process such as the sum, average, minimum, maximum and others for each existing field. This turned out to work and allow for a later step of classifying each county.

Before classifying each county I had in mind a different method in producing the map. The original idea in my mind was to have the county map as background reference map to allow someone to identify their location of interest by referencing the county borders. On top of this map was going to be the cost of living overlay, which would be an interpolation of zone coverage costs based on past occurrence costs. Various methods were performed from Inverse Distance weighting, Point Density Patterns to Kriging. Based on the results produced from these methods I further reduced the data set and decided to come up with an alternative form of representation to see which would work best.

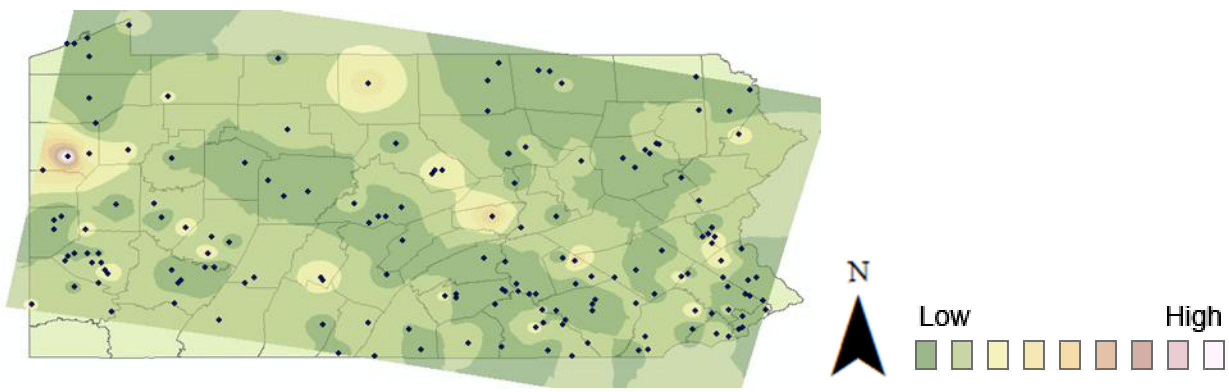
While realizing that maybe I was overthinking the process I went back to some of the basics dealing with some simple classification. This is when the final joins method actually occurred when I needed weather data associated with each county in order to classify it against the other counties. Using the joined data I was able to generate classification maps for the counties for each type of disaster using the generated averages and/or totals. With time running low the final step was to generate a final map combing everything into a total cost estimate map. Without actual costs being found the project took a little turn in only dealing with the risk involved, which can be assumed to correlate to costs. To combine the data was sort of a mystery given the depth of the data and varying cost and effects different natural disasters. The only thing I could think of was to reclassify a few of the data sets into two, originally three, new classes, splitting the risk into high and low and then overlapping the data. Overlapping three types of risks with high and low for each would result in a state map with six possible risk/cost rankings.

With all that said and done here are the results up to this point. These first two images show all of my data sets when initially loaded into ArcGIS just to get an idea of how much data there originally was.

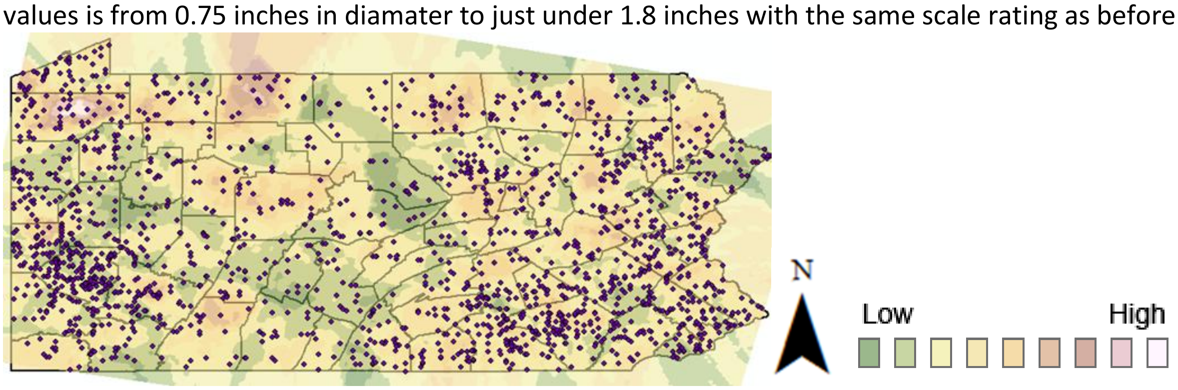




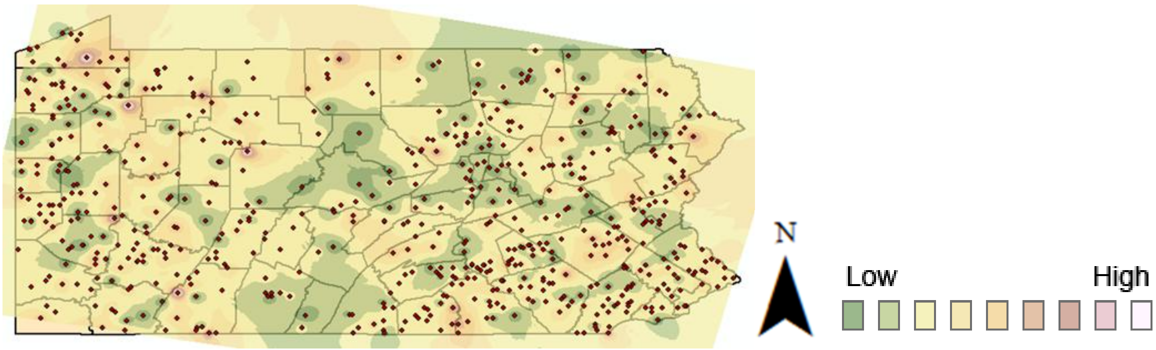
After various interpolation methods the following maps were generated. The first image shows the interpolation of Wind data, recorded areas of wind over 65 knots per hour, using Inverse distance weighting with a radius of 2500 meters and wind speed as the factor. The range of values here are from 65 kts to 130 kts with green on the low end up through red and then finally white being the highest.



For the next set of data I decided to go with a different method to experiment with the results. For Hail data, below, I decided to go with Kriging based on the size of the Hail recorded. The range of values is from 0.75 inches in diamater to just under 1.8 inches with the same scale rating as before.

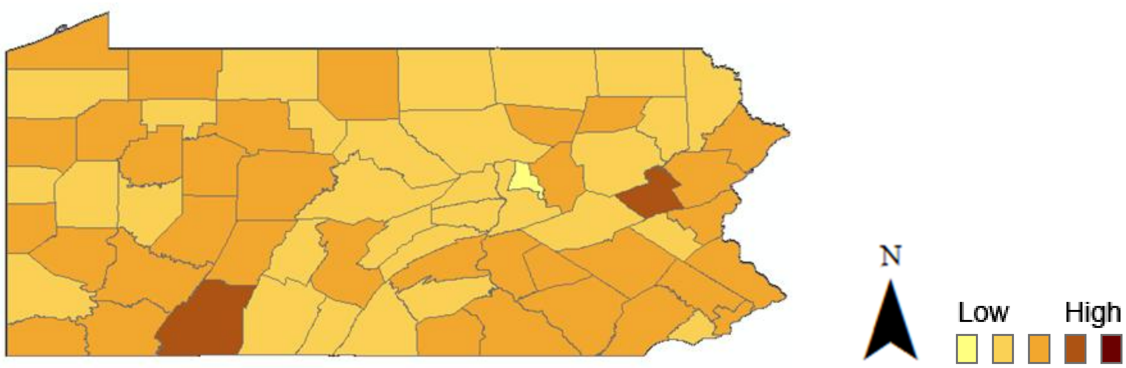


For the tornado data I used a couple of different methods and settings but ended up deciding that Inverse Distance Weighting produced results similar to the classification shown later. The range of values is from an Fscale of 0 to the maximum of an Fscale of 5, with the same scale rating as before.

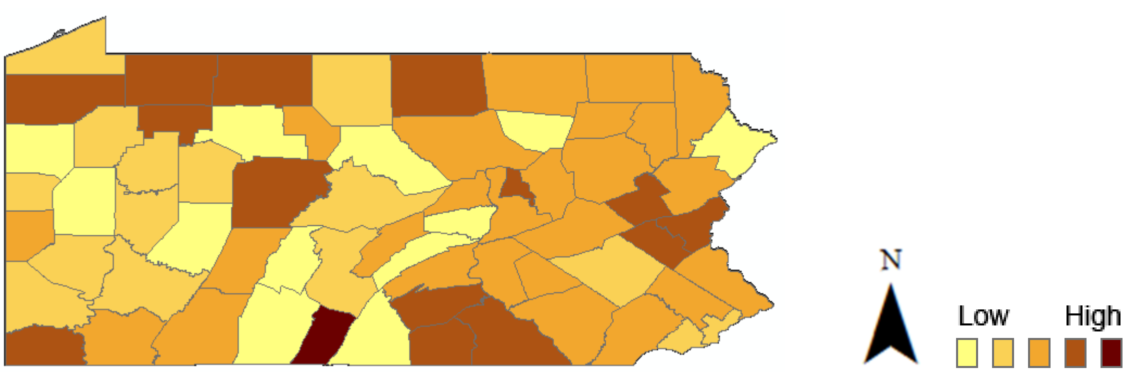


Combining these is quite difficult but may provide information at their indivudual level of detail if someone were to want it. For classifiying I wasn't sure what to include so I chose hurricane data over wind data as I would imagine the costs being greater for a hurricane than wind alone. The following classification maps take on a new color scheme but use the same gradient.

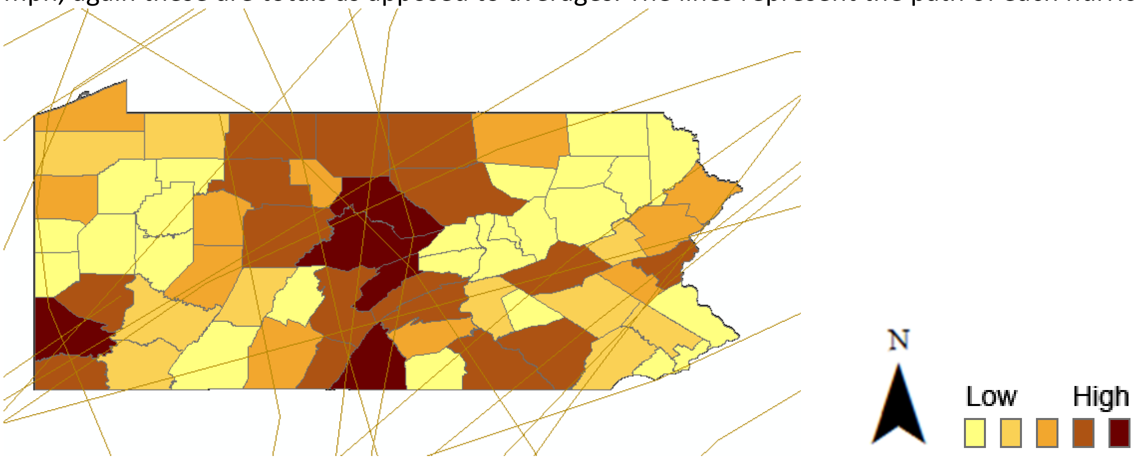
The first map below is of Tornado risk. I classified each county by the average Fscal to occur in the boundaries into four classes. Using natural breaks and the light to dark color scheme you can easily identify areas of various risk. The tornado color scale ranges from 0.33 to 2.0 Fscales.



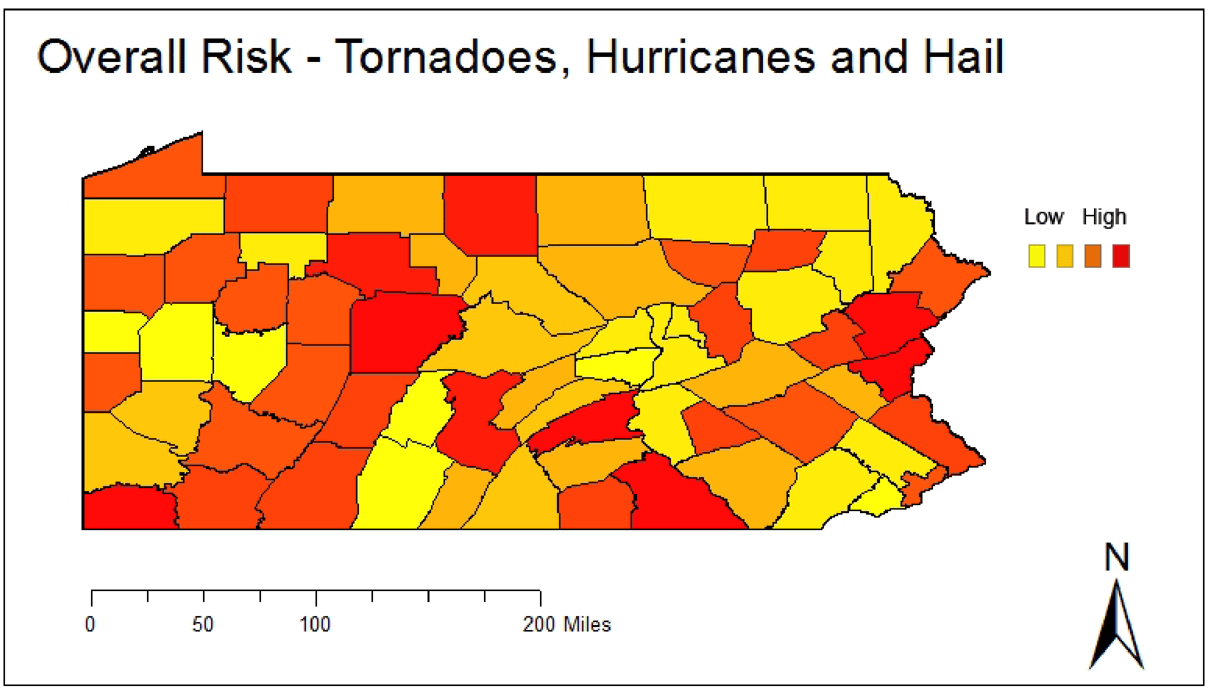
The next map is of Hail size by county. I classified each county by the average size of all the recorded hail in the boundaries into four classes. Using natural breaks and the light to dark color scheme again you can easily identify areas of various risk. The hail color scale ranges from 0.80 to 1.4 inches in diameter.



The last map is of hurricane data by county. I classified each county by the total wind speeds expeerienced in the boundaries into four classes. Using natural breaks and the light to dark color scheme again you can easily identify areas of various risk. The hurricane color scale ranges from 0 to 360 mph, again these are totals as apposed to averages. The lines represent the path of each hurricane



The final step was to combine these 3 classifications of various scales into an overall risk map. To do this as mentioned before I re classified each map into two classes, using natural breaks again. I assigned yellow to the low risk and red to the high risk classes. Overlaying the maps with 33% transparency on each produced the following map. Areas of low risk in all areas remained yellow while areas of high in areas and low in others would appear in two shades of orange depending on how many categories they fell under high in. Areas in red are clearly areas that were of high risk in all three categories.



With this final map generated and limted time to try to achieve my initial goal I ended it with that. While it seems simple as a final output and presentation the process was quite time consuming and a learning experience. The interpolation maps were also time consuming and not included in the 'final' single map but can be used for individual assessments with more detail than just the county level like the final map. Interpolation on data such as weather events outside of temperate or air pressure arent quite as useful but may be used to generate ‘hot spot’ maps. Areas where an event has occurred are more likely to have a repeat than an area that has never experienced such scales of intensity. Anything is possible however so the interpolation of the maps represent the extent of risk based on past events. Interpolation was also only producing decent results on data sets with a lot of points. Interpolation of earthquakes was near impossible and generated results that could not be useful as there wasn’t enough data to back it up.

All of these methods and findings are also backed by factors not considered such as land type, elevation, as well as the other data collected. The overall risk map only consideres three events that I chose to be considered important. While there was data values for croploss and damages it was only recorded in intgers of millions of dollars, meaning a value of 0 doesn’t necissarily mean there were no damages. Another area of error would be putting the three selected disasters to be of equal risk when combing them, I considered a small ball of hail to be on the same level as a low scale tornado, yet a tornado may cause more damage than the small hail. There were so many possibilities I just had to make an estimate yet still provide a usefull result.

With the result we can safely estimate the associated cost of living in certain counties based on the final classification. An area of red, as mentioned before, is prone to experience higher intensities of the three possible weather disasters while an area in yellow have only been exposed to the lighter side of things. Even with little knowledge of Pennsylvanias geography one could match up the areas of low risk with mountainous areas as apposed to flatter areas. This correlation is probably due to a flat area giving a storm the opportunity to become stronger than in a mountainous area. An area exposed to harsher conditions is going to have more damages and losses involved and therfore the cost of insurance coverage will most likely be more than other areas.

While performing the final project, without any specific absolute criteria other than using what I learned left the project open ended. This became evident when my final map wasn’t entirely what I was going for as everything isnt possible, especially with only what I know how to do thus far. Available data also becomes an issue and may require additional data collection to be carried out. Possible future operations within this project might include making a crime based risk map, which would be possible with mostly the one data set which included all sorts of crime data. Additionally the many factors that may control the occurance of specific data can be considered to help with understanding areas of risk. A final, more involved, future area of analysis would include predicting the weather taking in consideration of time intervals and observing the change between them. All of these areas of further analaysis would be of interest to me had the time and knowledge not been a limting factor. With those two factors present I would have also liked to have improved the ‘final map’ as I what I had in mind was going to have a little more to it. Overall though the important part behind the project is realizing and actually performing the work involved in creating what appear to be some of the most basic maps customized to some fairly basic needs.