Floods - Midterm Project

Ajay Krishnakumar

2023-11-06

Introduction

(Forgive the following grandiosity:)

Since time immemorial,the idea of floods and rising waters and the calamities they bring have captured the human imagination. Since time immemorial, that fear has been true. Civilization blossomed on the flood plains - of the Tigris and Euphrates, the Indus and the Nile, and ever has water been our making and our unmaking. With the looming specters of global warming, climate change, storms, unpredictability and environmental damage, understanding the effect of floods is increasingly important. Even more important, I would argue, is figuring out if these effects are visited disproportionately on the most vulnerable parts of our population.

I will not pretend that this analysis comes even close to such a lofty goal. But it provides some information. Certainly, it provided me with plenty. Do floods affect poor parts of the US more than wealthier parts? Are responses to floods and financial assistance distributed in an equitable manner? It's not good news, I'm afraid. But the railing against the state of affairs will come later. There are steps to be taken on the journey between now and then.

Data Cleaning

Data Cleaning largely consisted of improving the way the data was labelled so that it was easier to parse, fixing some obvious issues with the way the data was stored and then getting each dataset in a good shape for the joins that would assemble the large dataset I used for most of my analysis.

The final dataset consists of data from NOAA, FEMA data on disaster summaries and on assistance given, and finally census data about poverty in various counties across the US.

Beginning with the NOAA data: this was read in for 2020 and 2021 and the two data frames were joined after a few trivial steps. This was then filtered to include only Flood events. Immediately, a few things stood out: the way the dates are coded and the fact that they were coded twice and differently. This was fixed by dropping the less informative columns and using

Table 1: Floods in 2020 and 2021

State	Number Of Floods
KENTUCKY	968
WEST VIRGINIA	396
TENNESSEE	336
ARKANSAS	288
VIRGINIA	264
ALABAMA	250
OHIO	230
LOUISIANA	218
MISSOURI	214
ILLINOIS	200
GEORGIA	186
MISSISSIPPI	146

lubridate to sort the rest. The damages columns were converted to numeric and multiplied by a thousand or a million depending on the suffix attached to them in the data.

Next each of the two fema datasets was read into a dataframe,cleaned and combined. Then floods was left joined to this combined dataframe. Columns such as magnitude which didn't contain any information pertinent to floods were dropped.

Then census data on poverty for 2020 and 2021 were read in, cleaned, combined. The biggest thing here was getting the names of the columns to be something easily parse-able which was achieved through various lines of stringr code. The floods-fema data was left joined with this and voila.

Now we have three data frames - just the floods data from NOAA, the floods and FEMA data and the floods, FEMA and Census data combined. That means its time for some EDA.

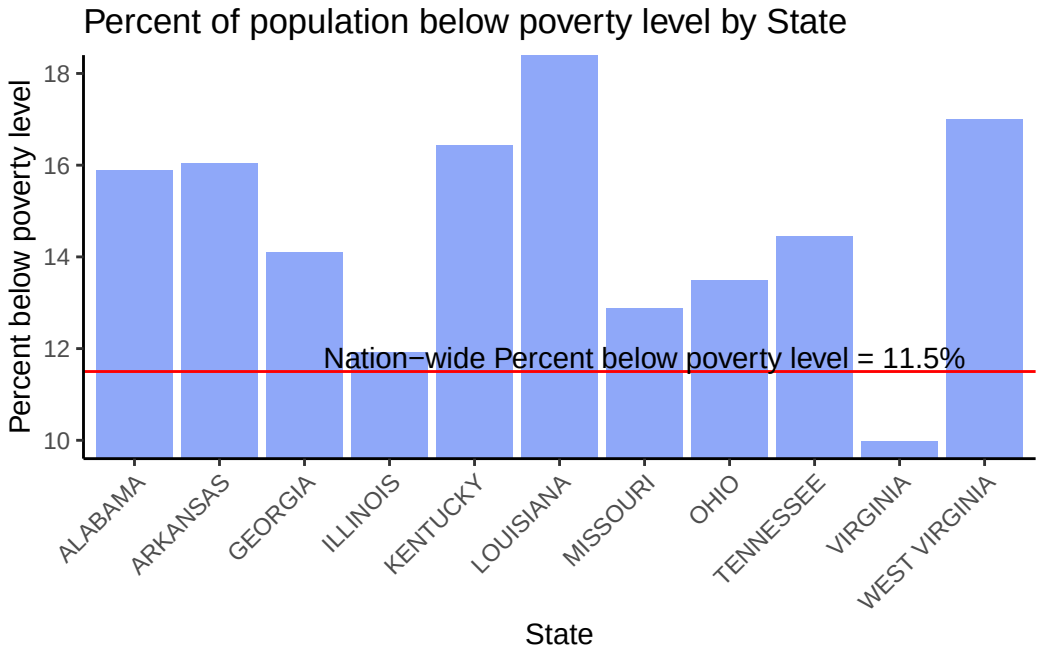
EDA

So we can start by just looking at the number of flood events in each state over the two years in question. This might point to something interesting or at the very least give some indication of where to focus analysis geographically. It'll help if we can identify trends in relationship between poverty and floods across states and if there's any obviously similar characteristics across states.

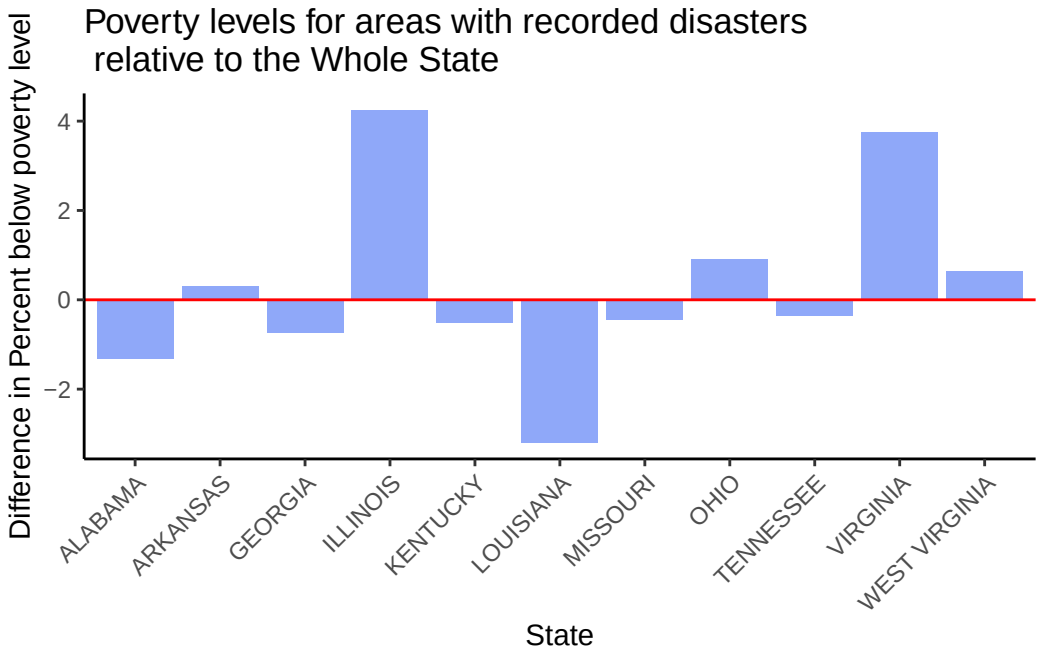
My focus, in this analysis, is to look at whether floods disproportionately affect poor areas and what relationship poverty has, if any, with FEMA claims. Table 1 Displays just the top 12 states by flood incidents.

EDA - Intial Scouting (Noah looking out from the Ark)

Let's start by visualizing how poverty varies across the states in question. We can limit ourselves to the top twelve most flooded states as indicated in Table 1.

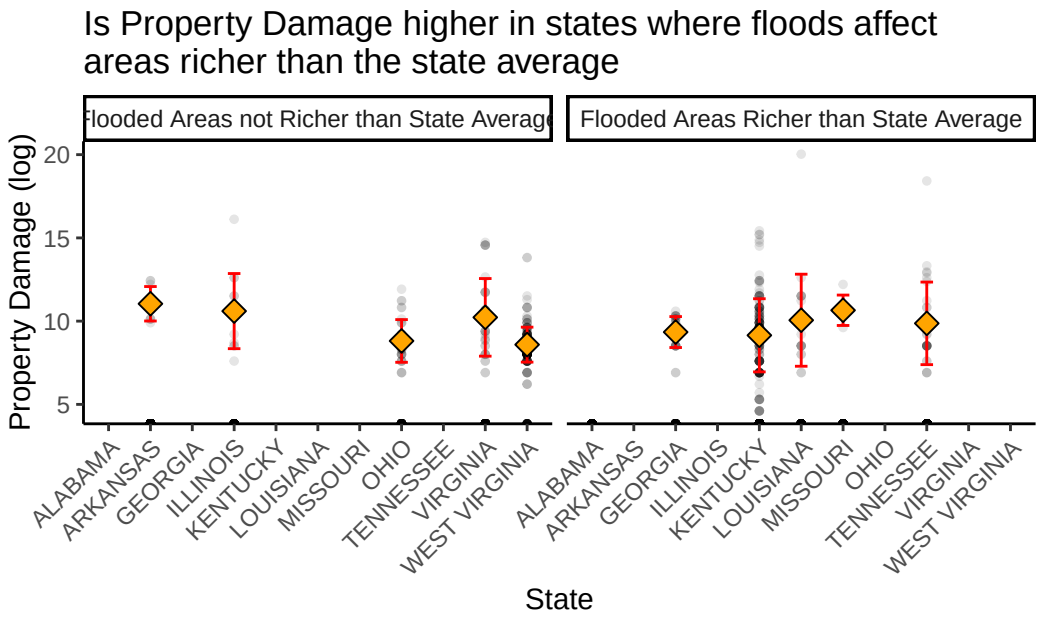


We can see that most of these states, barring Virginia, are above the US Percent of population below poverty level.(Congressional Research Service, Poverty in the United States 2021 and 2020). Let's see how that compares for only those counties with flood events in 2020 or 2021.

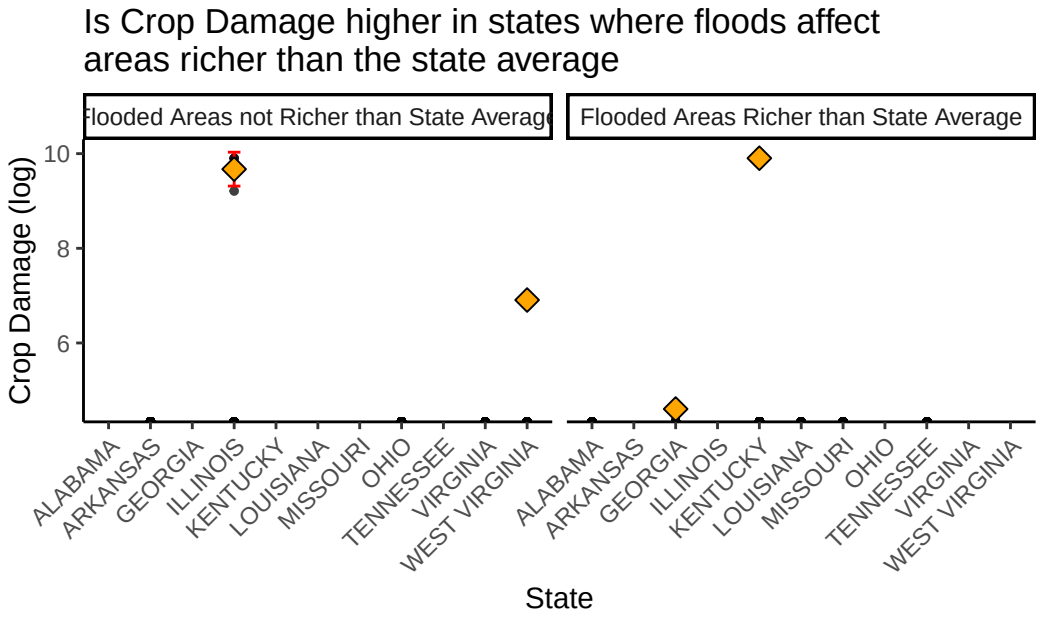


There are some states where the areas affected by floods are poorer on average than the rest of the state. Equivalently there are some where the flooded areas are less poor. But this graph is not in itself all that informative. There are also many differences between states, their livelihoods, the distribution of the population. It might make sense to narrow our EDA to some states that share not just apparently similar qualities from a socio-economic standpoint but also in the way floods have affected them in 2020-2021.

So: Are there any trends in deaths, property damage and crop damage between the two groups of states we've created above? Can this help us?



Orange diamonds indicate the mean log property damage, error bars are in red



Orange diamonds indicate the mean log Crop damage, error bars are in red

The crop damage graph is almost immediately indicative of which states are more rural/rely ostensibly more on farming. What's interesting is the difference in crop damage across states. Is the crop damage higher because the flooding was more sever/more prolonged or was it because the primary crop grown there is more susceptible to flood damage? It might be interesting, in a future analysis, with some agricultural data to evaluate that.

The deaths from floods in 2020-2021 were zero in most states which is great news. However it appears Tennessee really suffered, with close to 20 deaths from one flood alone. This wasn't really worth visualizing as a graph so I've mentioned this information here.

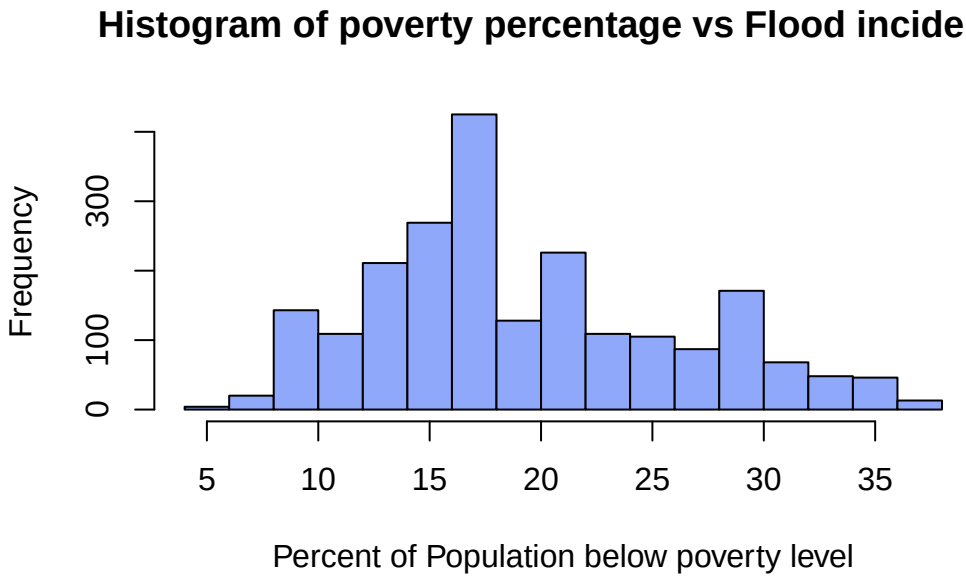
By far the more interesting graph is the one detailing property damage. West Virginia had a lot of floods but had lower property damage levels than Virginia and its mean property damage is the lowest of all these states.

Tennessee, West Virginia, Virginia, Louisiana,Ohio and Kentucky are the most interesting of this bunch. All have had a number of floods which span a large range of property damage values. Could this be perhaps because the floods have occurred across both rural and urban areas? Or is it perhaps because valuable infrastructure was damaged in the floods - roads, farming equipment and their like? It might make sense to narrow focus again to these states and look at differences in poverty levels across their counties to see what we can find.

EDA - Diluvial Details

So, what questions has our initial exploration left us with:

- Do poorer areas get affected worse by floods?
- Is property damage damage largely a function of damage to individual homes or to infrastructure?



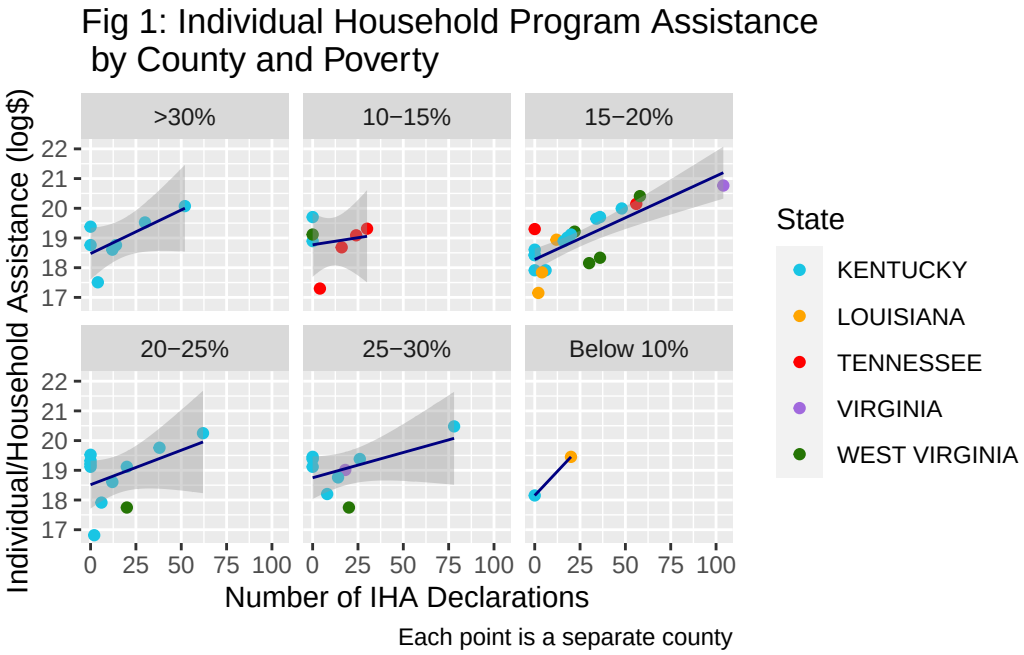
There doesn't seem to be a real trend showing that areas with a higher percentage of inhabitants below the poverty line are hit more frequently by floods. This doesn't mean they aren't hit harder.

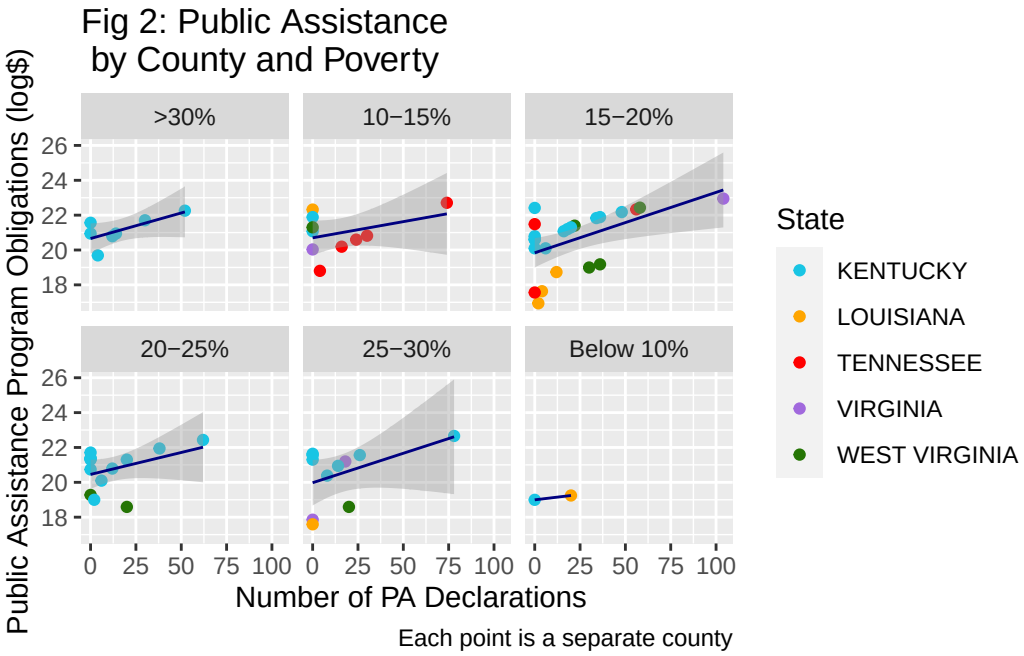
How would we go about seeing if that is the case?

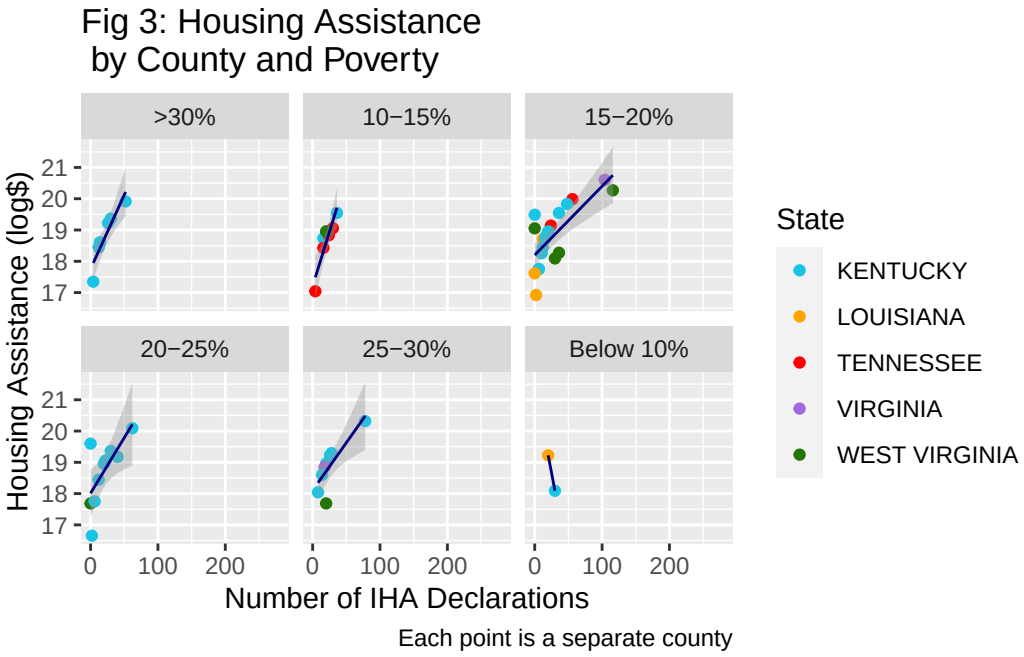
We can look at the fema information. How many requests for funding came in? How are funds distributed across various poverty levels to a) infrastructure b) individual assistance c) Housing Assistance.

So we'll look at the amount of money FEMA grants against the number of declarations of assistance to see if there's a trend. We'll make a plot for each of a, b and c above and see what conclusions can be drawn.

A brief caveat here is that while we have housing assistance amounts, we don't have data for the number of declarations for them because they're lumped into individual household assistance by FEMA. The general analysis below holds, however, and is - in my opinion - still valuable.







Let's break this down. Across all three of the above graphs, moving along the x axis, I would contend, is a measure of the severity of the flood. Moving along the y axis indicates the magnitude of financial assistance - the response. You'd hope there's a positive slope to the trend line I've inserted, and mercifully for humanity, this is the case basically in all of the above instances. The steeper the trend line, generally, the larger the assistance for an increase

in severity. Naively, one might expect that this rises as the percentage of population below the poverty level increases.

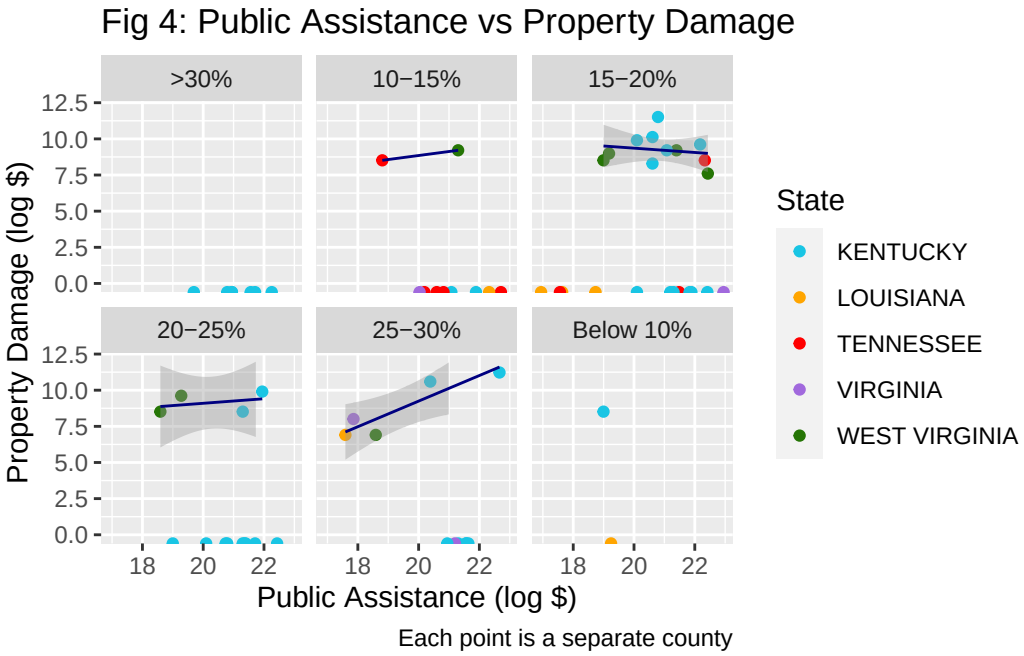
What is interesting is the near-complete absence of counties with flood events in the less than 10% below poverty line category. The other categories have more or less a similar number of floods, although its worth pointing up that the second lowest poverty level is also associated with having generally lower levels of Assistance Declarations from FEMA - i.e they have fewer high severity floods than poorer areas. Kentucky is the only state with counties affected by floods whose population is more than 30% people/households below the poverty line. And Tennessee doesn't have any counties affected by floods in 2020 and 2021 with greater than 20% of its population below the poverty line.

Starting with Figure 1: Barring the 10-15% below poverty level group, it would appear that the sensitivity of financial assistance to changes in severity are largely the same across all our poverty levels as far as Individual and Household Assistance is concerned. The 'spread of 'shadow' is indicative of the uncertainty in the trend so more data might certainly change this for the better or the worse. The cynic would be inclined to say that equal progression in financial assistance relative to severity across various poverty levels would be akin to a uniform tax rate, which is to say, regressive.

Figure 2: This one is largely more indicative of assistance to roads, government buildings, and rebuilding efforts of that ilk and therefore not particularly informative of financial assistance as far as poverty goes. On a base level though, if one were to make the contention that poorer areas have less infrastructure, we would expect to see significantly fewer requests for public assistance for poorer areas. Clearly this isn't true.

Figure 3: The magnitude of relative response remains the same and I point you to my earlier comment about what a cynic would be inclined to say of such a state of affairs. What is interesting here is the foreshortening of the lines. For high IHA declarations it would appear that Housing Assistance is not approved as much which seems incredibly problematic since one would expect it to be the case that more severe floods would necessitate more assistance repairing houses. This discrepancy might be explained by a prevalence of flood insurance and more data would be needed to confirm or debunk this idea. Something to look into.

So let's have a quick look at whether property damage is a function of damage to homes or infrastructure:



We can see from figure 4 that there generally isn't any relationship between property damage and Public Assistance. In fact the property damages are often 0(hence their peeking from the bottom of the graph) despite large public assistance.

The sheer number of 0's for declared property damage likely means we can't tell if housing assistance and property damage are related either.

Conclusions

Does it appear that floods disproportionately affect poorer communities in the US and amongst those states we've narrowed our analysis to?

It's hard to say given only two years of data. But we can certainly say that communities with lower levels of poverty are affected less often and less severely.

Moreover the federal government's response is generally distributed evenly relative to severity of the flood event and regardless of the poverty levels of the community affected. I would argue that the overall effect of this is to reinforce income and wealth divides, particularly in places where flood incidents occur frequently.

A few other points:

- Interestingly, it appears that landlocked US states are most affected by floods. I would've expected coastal states to flood more.

- We haven't been able to conclude whether property Damage is largely related to housing or infrastructure. This is a question that beggars answering and future analysis with relevant datasets might go some way to answering this.
- What affects crop damages in floods? Is it the severity of the flood? Or the sensitivity of the crop to floods? In a world of changing climates, it might be an important question to answer - which crops are most resistant to more and less water than they are normally accustomed.

I think it would also be useful to extend the analysis done above to:

- Other states. Including this above would have confused the insights and distracted from the findings - such as they are - but that's not to say that this isn't useful for later.
- Over time: How have flood incidents, severity and their impact on poorer communities changed over time? This is a really important question to answer and this EDA is barely a token acknowledgement of that.