Floods

Ajay Krishnakumar

2023-11-06

Introduction

#Data Cleaning

So immediately, a few things stand out: the way the dates are coded and the fact that they're coded twice and differently. What I'll shoot for here is to have dates as month, day and time and also as a single datetime

After dates, the damages columns need fixing.

```
[1] 13531
```

Let's join the floods data with the combined fema data:

We can also join this data with poverty data from the census. This data has to be cleaned first:

Now, in order to join the census data and the floods combined data, we need to reduce the geographic area name column to just the county. Furthermore, we will rename it to have the same name as in floods.

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.

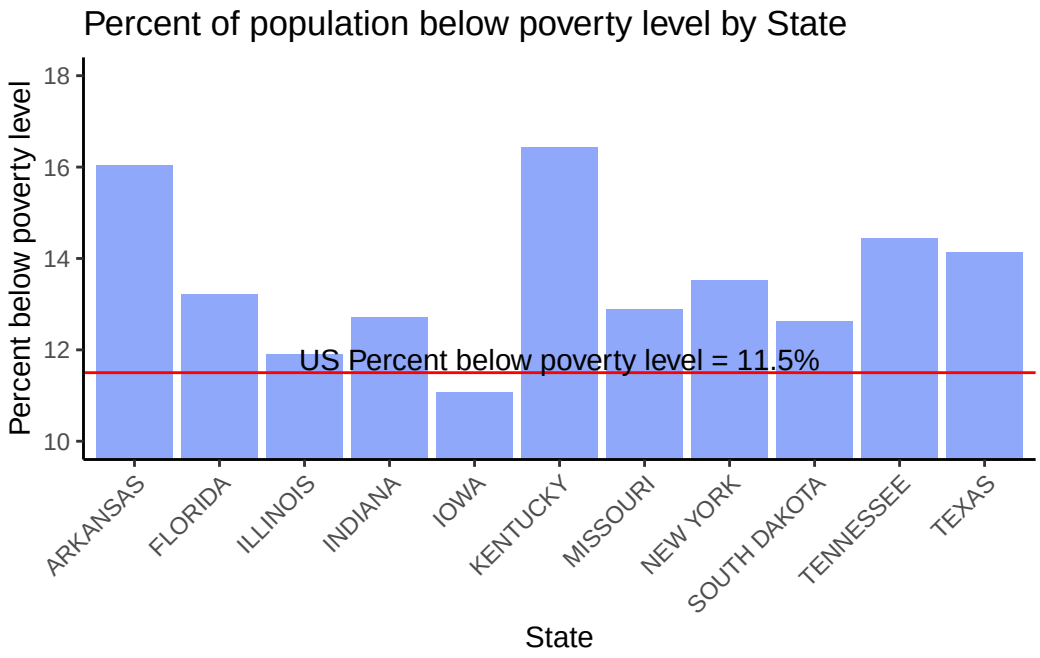
EDA

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 if we want to have that focus geographically.

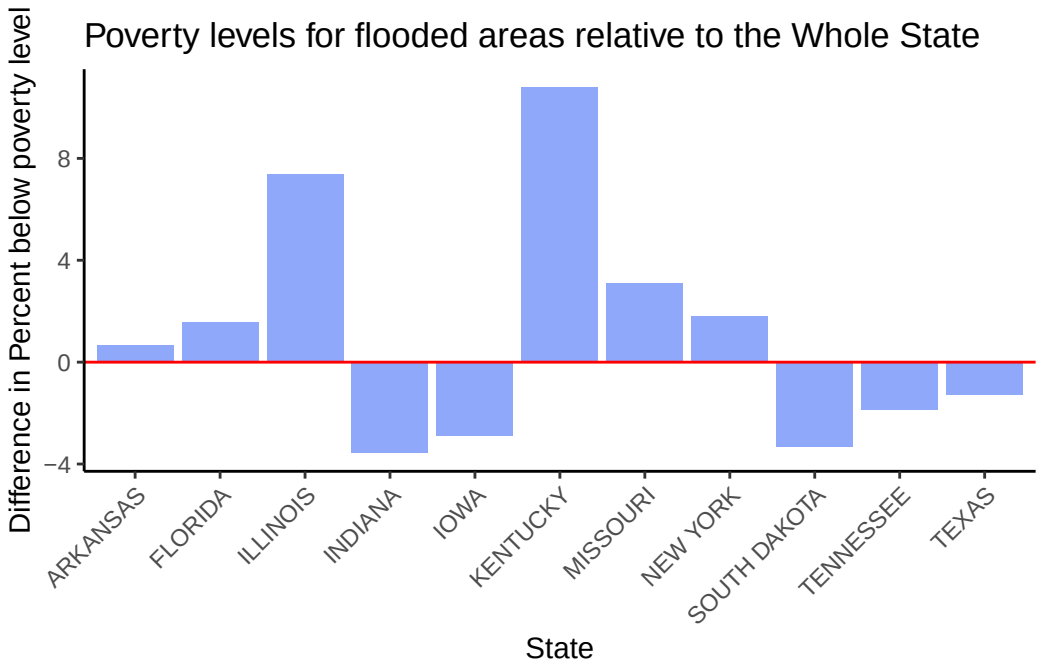
```
# A tibble: 44 x 2
# Groups:   State [44]
   State             n
   <chr>         <int>
 1 KENTUCKY       1610
 2 SOUTH DAKOTA    980
 3 TENNESSEE       752
 4 ARKANSAS        680
 5 MISSOURI        433
 6 ILLINOIS        424
 7 IOWA            362
 8 NEW YORK        356
 9 INDIANA         318
10 WEST VIRGINIA   268
# i 34 more rows
```

We have coastal states like New York and West Virginia represented but it would appear that the large majority of floods happen in non-coastal states. What's interesting here is that South Dakota has a large number of flood events. So do Kentucy, Arkansas, Tennessee and Missouri but those latter states are geographically close. South Dakota is a geographic outlier here. 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. It might be worth paying special attention to South Dakota later.

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 above.



We can see that most of these states, barring Iowa, 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.



We can see that there are a few states here where the flooded areas appear poorer than the state and others where the flooded areas aren't as poor as the rest of the state. This doesn't

necessarily mean that in these states the floods happen in areas with wealthier people. We can use the census data to look at flooding in areas where the income is high above the poverty level (Which is by no means equivalent to saying areas with high levels of income.)We can look at areas between 300% and 500% of the US poverty level as being comfortably above the poverty level but we'll come to that later. Let's first look at the states