## **U.S National Disasters over the last 64 years**

IST 652, 23 March 2020

## **Group Members**

Shawn Anderson - regional analysis, next steps

Yen Yung (Randy) Geszvain - data ingestion, transformation and preliminary data analysis

Brian Taylor - by state and top 5 state analysis, conclusions

Kaycee Williams - exploratory data analysis (first two questions)

## **Project Summary**

Team East Coast focused on the history of natural disasters using a dataset that provides every federal emergency and disaster from 1953 to 2017. Information includes the duration, location, and type of the emergency or disaster, as well as any assistance programs for individuals, public, and/or households. The team performed an analysis on several data questions and reported the results via visualizations with tables and graphs. The report will discuss the descriptive business questions, process of cleaning and explorations, and next steps.

## **Dataset**

The dataset is from Kaggle, a free online community for data scientists, at: <https://www.kaggle.com/fema/federal-disasters>

# **Business Questions**

The team focused on a descriptive analysis looking at the type of emergency/disaster by where it took place. After, a further deep dive was done on the findings for a better insight of why and/or what was causing the results.

Below is a list of questions the team looked at.

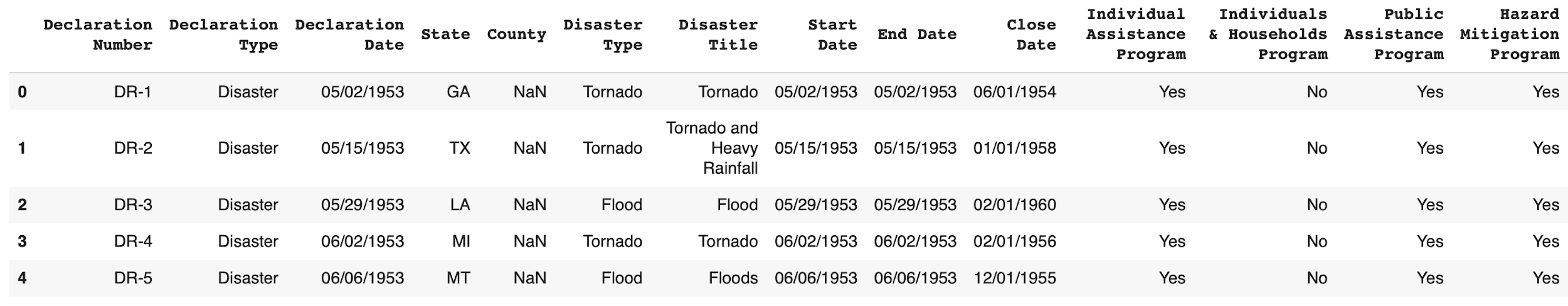
* What declaration type is most common?
* What state has the highest number of declarations?
* What are the major disasters in the top five disaster prone states?
* Which regions of the United States have had the greatest number of hazard events declarations over the last 64, 7 and 3-year periods?
  + Are there regions within the United States which have dramatically less counts?
    - If so, where are they?
  + Are there any visual correlations among these maps individually and/or in relation to each other?
    - If so, what are they?

# **Process**

The team processed the dataset through steps of importation, cleaning, and exploration to answer the business questions. Without doing this process the team would not be able to analyze the data properly.

**Data Importation**

We utilized the pandas package to import the csv file and transform the dataset into a dataframe. Below is the preview (head()) of the dataset. The dataset was pretty clean when we downloaded it from Kaggle. It includes many categorical labels such as declaration number, declaration type, etc.



**Data Cleaning**

We observed some attributes which needed transformation such as column headers, boolean type data, etc.

Column headers: Sometimes machine learning models don’t take spaces in the parameter. We updated the column header by running the script below. The manual update on column name let us proceed applying analysis or modeling without problems.

|  |
| --- |
| data.columns = ["DeclarationNumber","DeclarationType","DeclarationDate", "State","County","DisasterType","DisasterTitle","StartDate","EndDate","CloseDate","IndividualAssistanceProgram","IndividualsHouseholdsProgram","PublicAssistanceProgram","HazardMitigationProgram"] |

Boolean type: Then, we convert boolean data into 1 or 0.

|  |
| --- |
| data['IndividualAssistanceProgram'] = data['IndividualAssistanceProgram'].map({'Yes': 1, 'No': 0}) data['IndividualsHouseholdsProgram'] = data['IndividualsHouseholdsProgram'].map({'Yes': 1, 'No': 0}) data['PublicAssistanceProgram'] = data['PublicAssistanceProgram'].map({'Yes': 1, 'No': 0}) data['HazardMitigationProgram'] = data['HazardMitigationProgram'].map({'Yes': 1, 'No': 0}) |

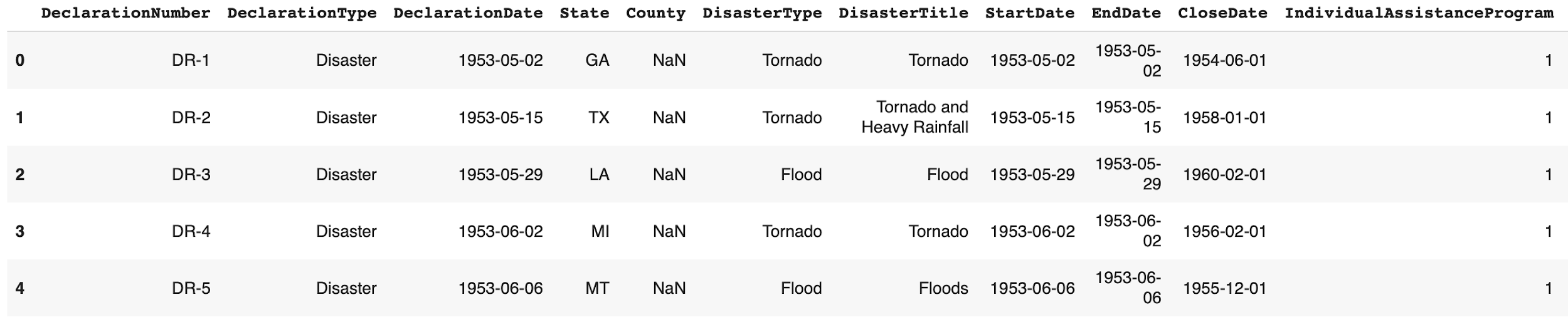
Column Type: We also transformed the column type on datetime and categorical data. By doing so, we can quickly feed the data into seaborn package and matplotlib package.

|  |
| --- |
| data['DeclarationDate'] = pd.to\_datetime(data['DeclarationDate']) data['StartDate'] = pd.to\_datetime(data['StartDate']) data['EndDate'] = pd.to\_datetime(data['EndDate']) data['CloseDate'] = pd.to\_datetime(data['CloseDate']) data['DeclarationNumber'] = data['DeclarationNumber'].astype('category') data['DeclarationType'] = data['DeclarationType'].astype('category') data['State'] = data['State'].astype('category') data['County'] = data['County'].astype('category') data['DisasterType'] = data['DisasterType'].astype('category') data['DisasterTitle'] = data['DisasterTitle'].astype('category') data['IndividualAssistanceProgram'] = data['IndividualAssistanceProgram'].astype('category') data['IndividualsHouseholdsProgram'] = data['IndividualsHouseholdsProgram'].astype('category') data['PublicAssistanceProgram'] = data['PublicAssistanceProgram'].astype('category') data['HazardMitigationProgram'] = data['HazardMitigationProgram'].astype('category') |

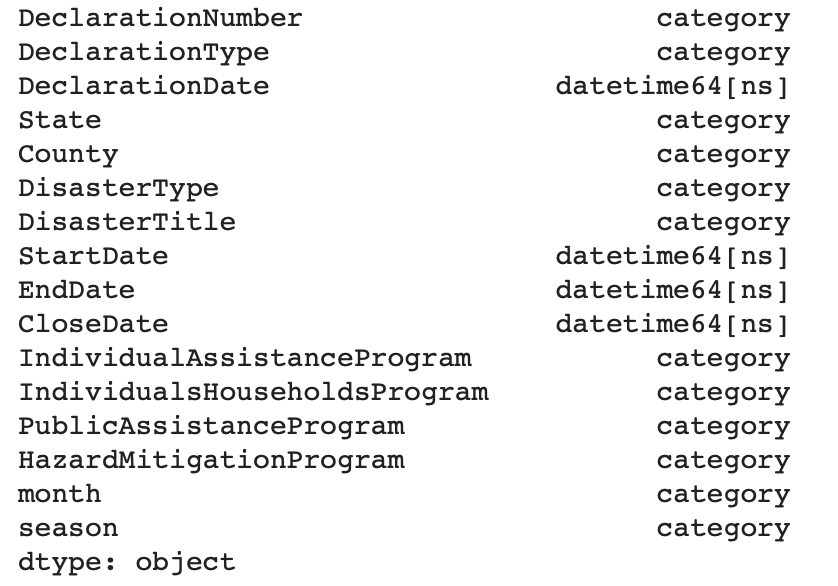
Appending more columns: for analysis purpose, we added month and season columns to further categorize and assist in future plotting.

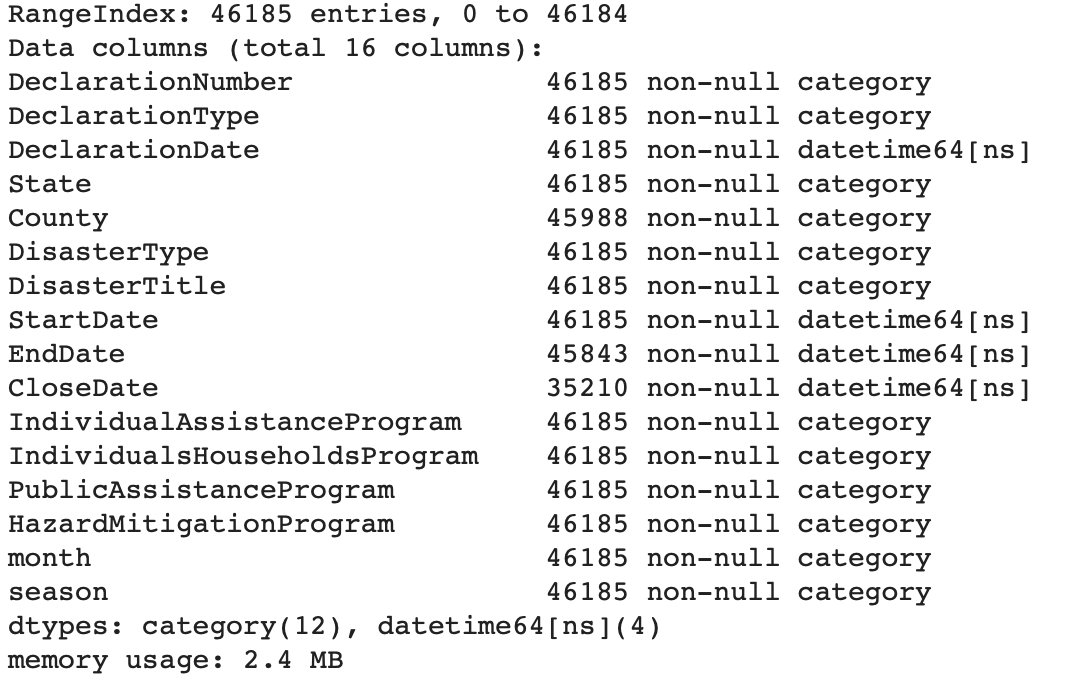
|  |
| --- |
| data['month'] = data['DeclarationDate'].dt.strftime('%m') data['season'] = (data['DeclarationDate'].dt.month%12 + 3)//3 #Month 12,1,2: Winter; 3,4,5: Spring; 6,7,8: Summer; 9,10,11: Fall data['month'] = data['month'].astype('category') data['season'] = data['season'].astype('category') |

After cleaning and transformation, here is the preview of the new dataframe.



**Data types and transformation**



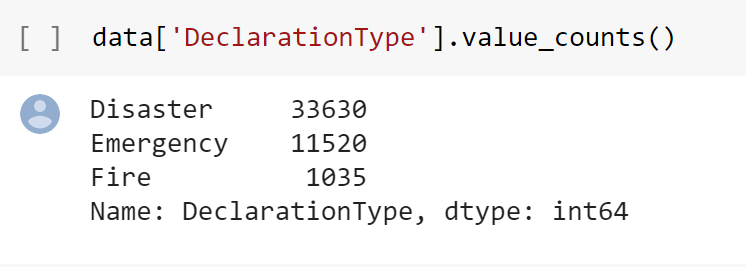
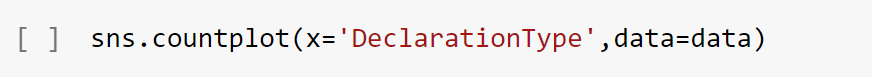


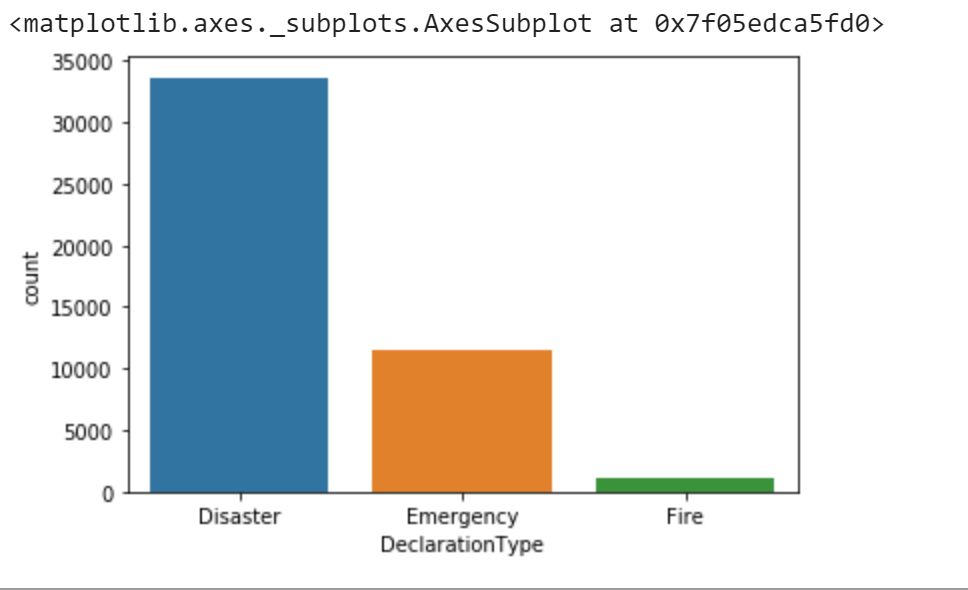
# **Data Analysis**

This section describes each of the business questions in greater detail, including an explanation of how the analysis was executed and an interpretation of the results. Multiple visuals and graphs are used to show results.

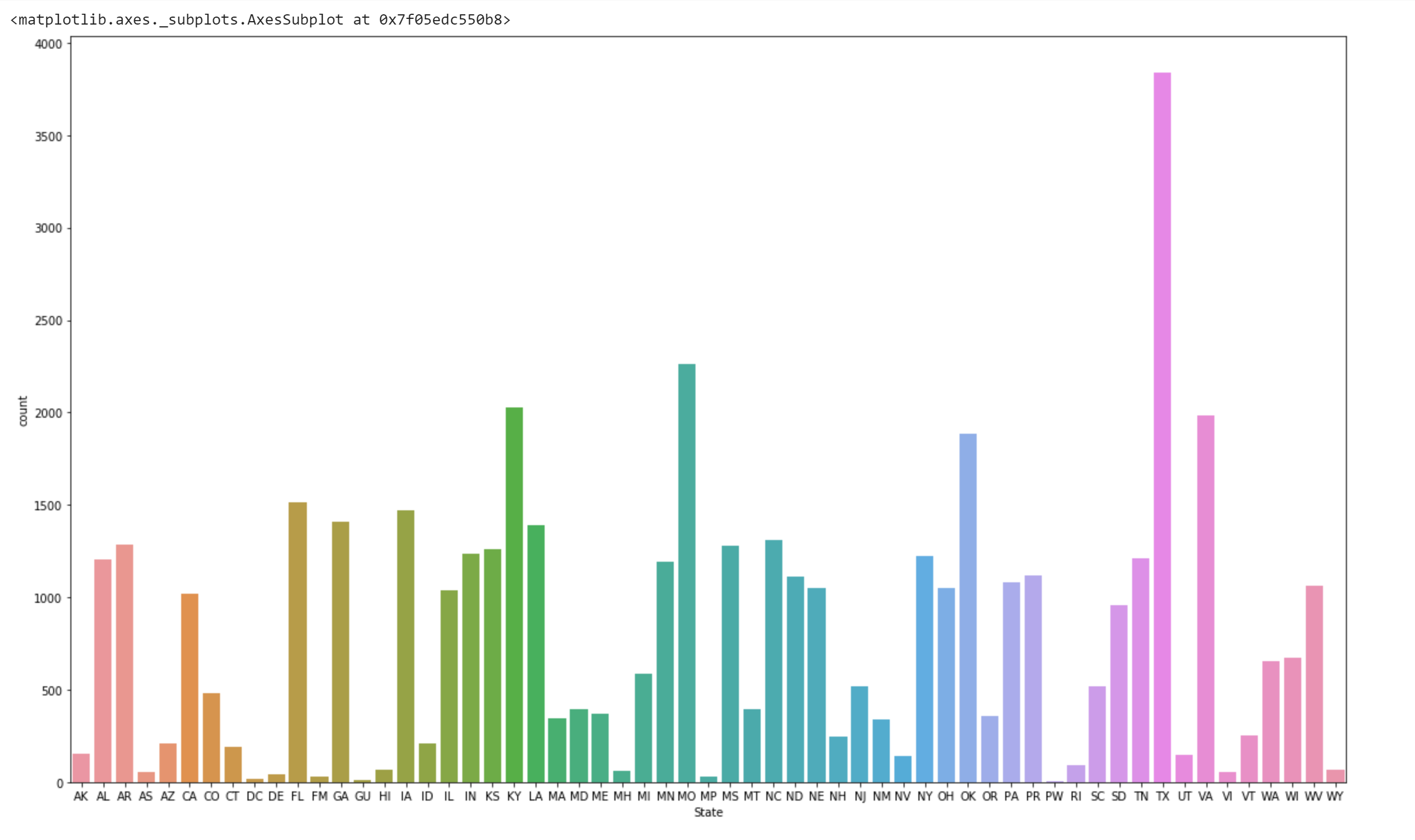
**What declaration type is most common?**

This question is a high level analysis of the dataset, specifically looking at the declaration types. Seeing what declaration type is most common helped us figure out what to focus on. To answer this data question, the following code was executed. Using the pandas frame created, we specifically looked at the declaration type with the results of counts of each. The results show that among disasters, emergencies, and fires, the most common was disasters at 33,630 instances.

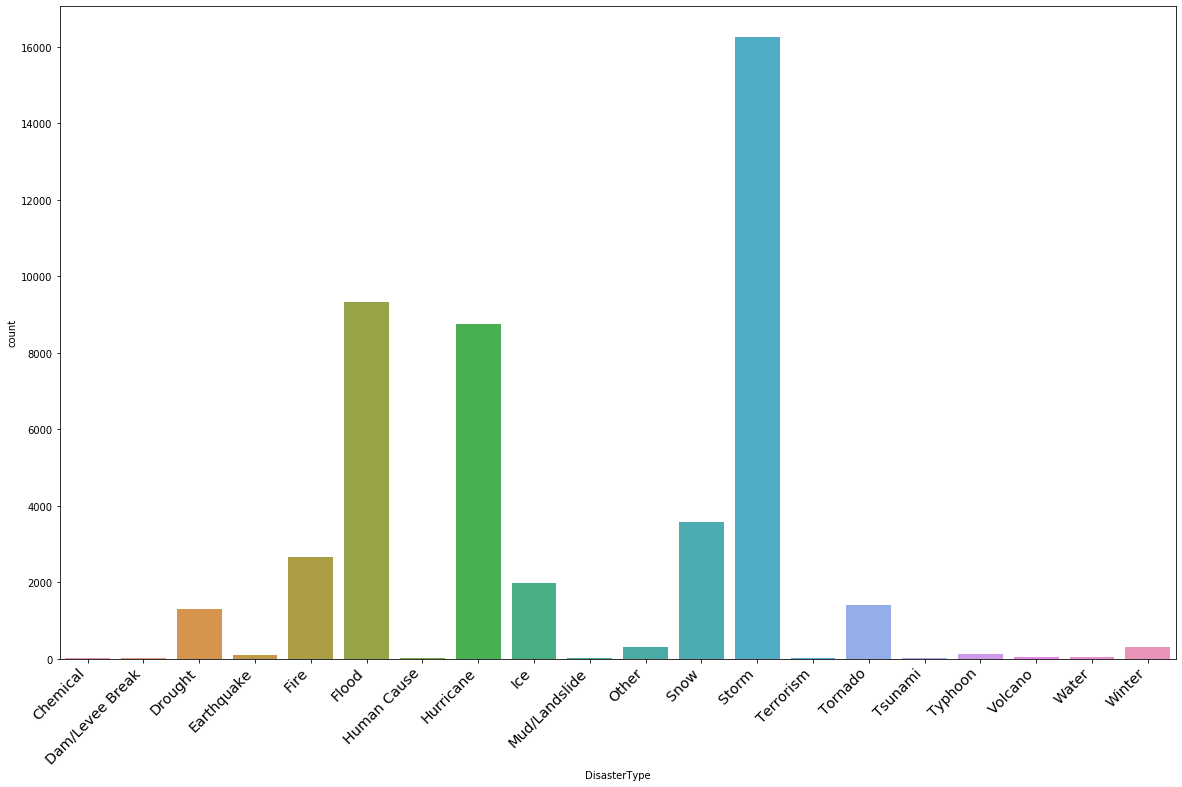
Additionally, the results are presented graphically using the below code.

Here is a graph showing the declaration types, disaster being the more frequent.

**What state has the highest number of declarations?**

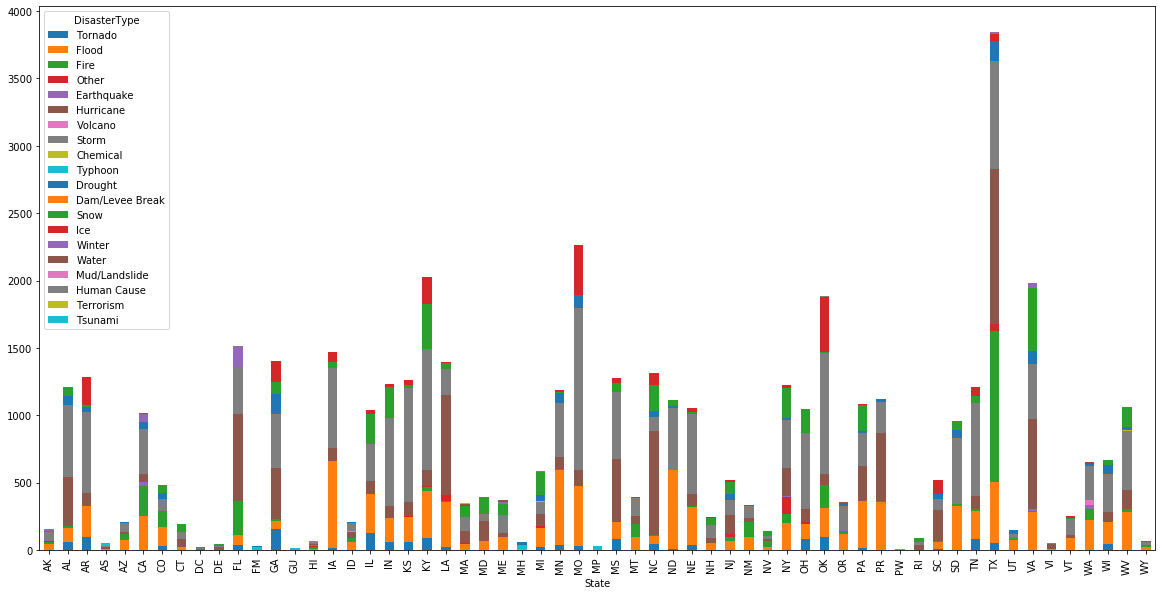
Below is a graph of the declaration counts by state. The state that has the most counts is Texas. This was surprising to us, since with the recent fires in California, we thought it would have more counts. 

**What are the major disasters in the top five disaster prone states?**

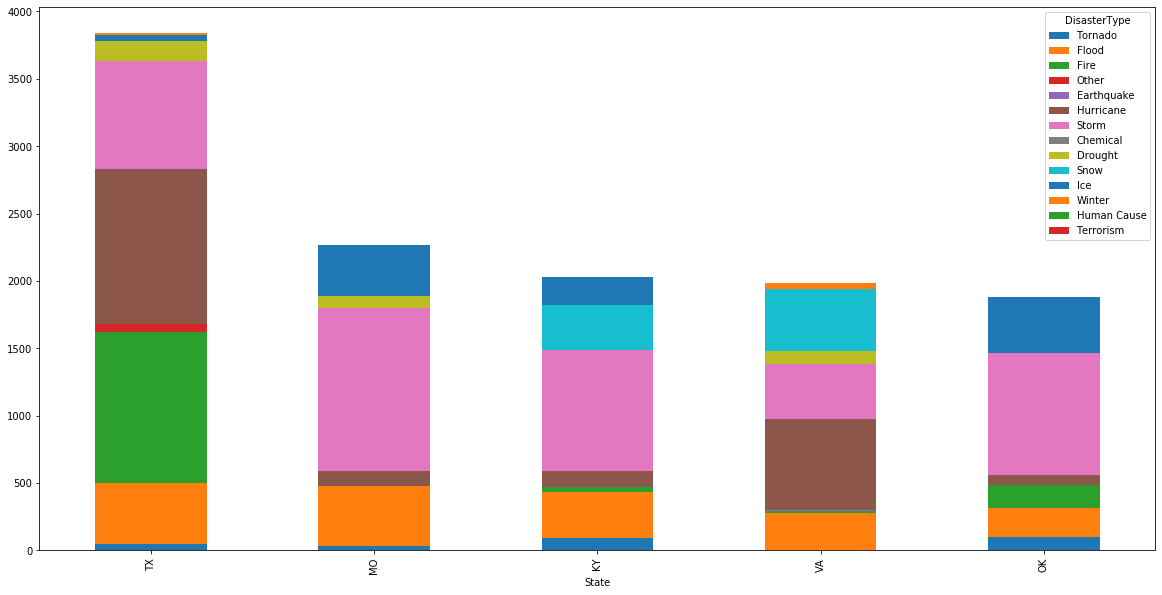


Storms, floods, hurricanes, snow, fire and ice lead the causes of federal disasters. What this indicates to a policy maker might be that severe and often unexpected meteorological events are the predominant causes of declared federal disasters. In areas where these phenomena are common, there is likely a need for federal assistance. This also indicates that the weather event is of greater magnitude than the state emergency responders can handle. It might be recommended that those states with a high frequency of a certain type of federal disaster invest in their response resources to those specific disasters.

Texas, Missouri, Kentucky, Virginia and Oklahoma are the top 5 disaster prone states over the last 64 years.



Those 5 states have significantly more disasters than the rest of the states and territories.



The types of disasters are different between those states. While “storms” contribute significantly, hurricanes greatly affect Texas and Virginia. This is expected due to their geographic location on the western coast of the Gulf of Mexico as well as the eastern seaboard. With the exception of Texas, snow and ice storms contribute significantly to the other four states’ disaster totals. This may be due to the latitude which they span, at the edge of snowy winters and hot summers, which makes the extreme seasons of summer and winter less predictable to prepare for.

**Which regions of the United States have had the greatest and least number of hazard event declarations over the last 64, 7 and 3-year periods (see Maps 1-3 below)?**

* + **If so, where are they?**

The Southern and Midwest regions have a long-term record for the greatest number of hazard event declarations - this is particularly true for the geographically second largest state, Texas. and along the longitude distance spanning Virginia, Kentucky and Missouri. In sharp contrast, the larger landmass and less densely populated Intermountain West region has maintained a consistently low event history in recent and long-term history spanning 64-years.

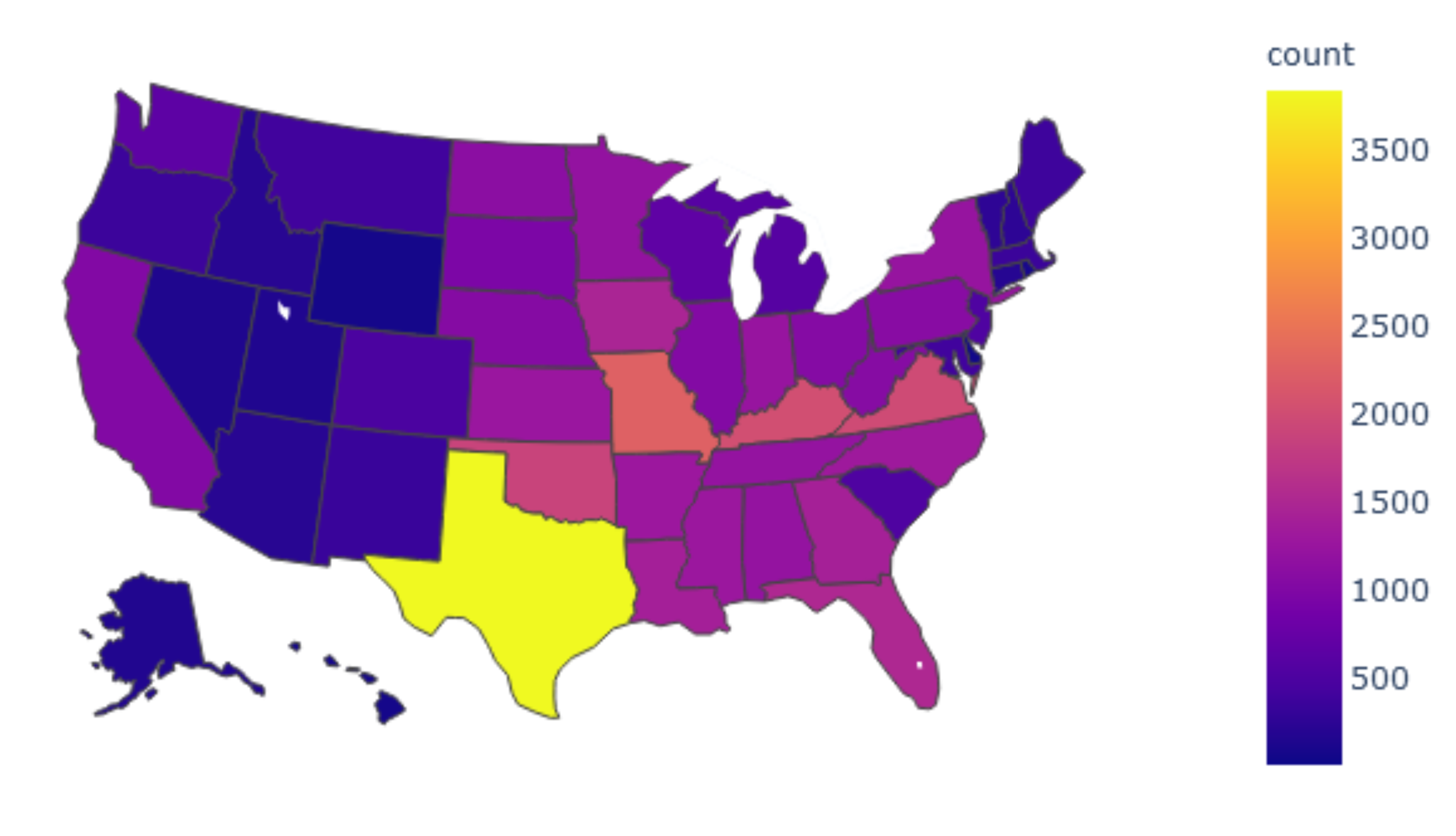
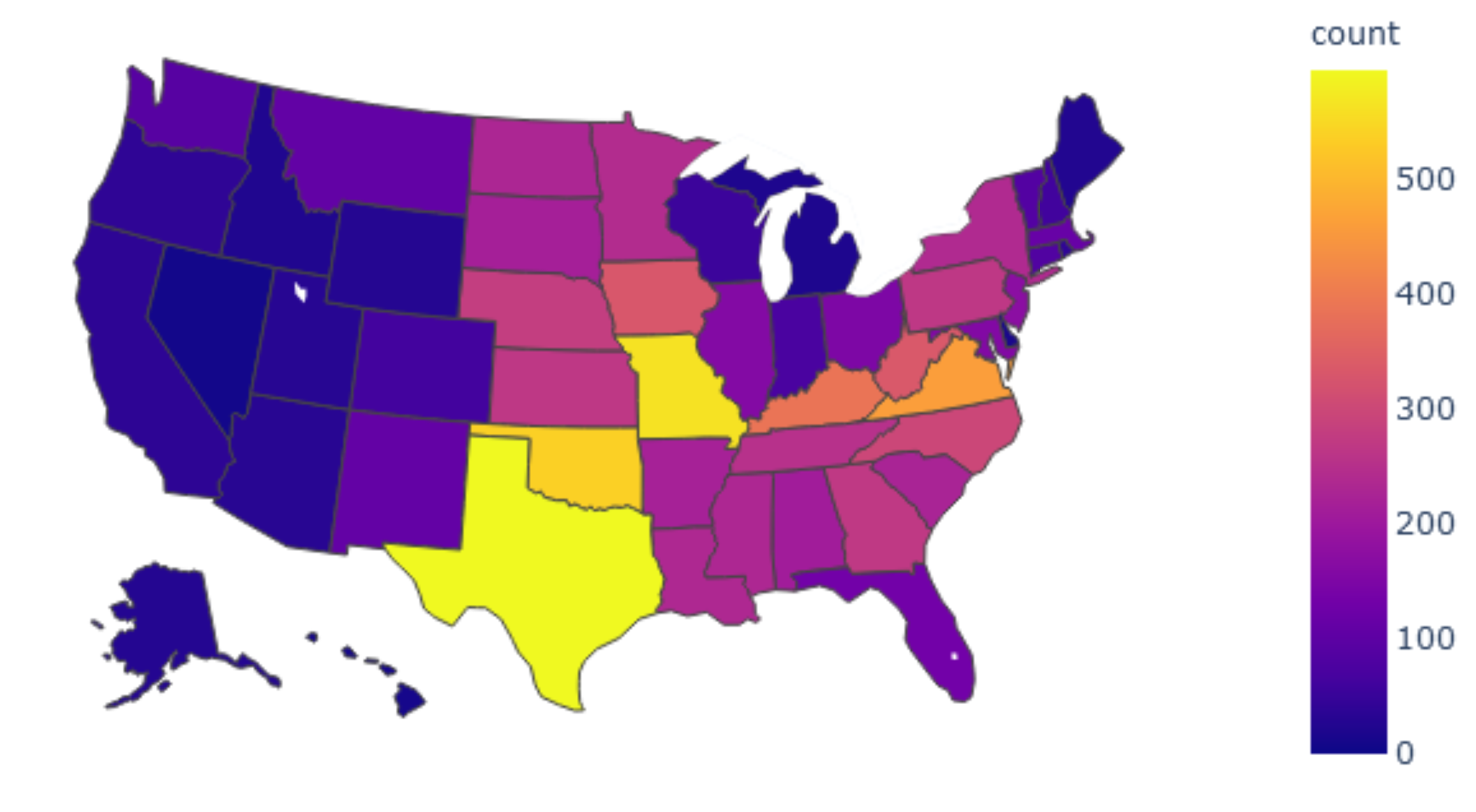
**Are there any interesting observations and/or visual correlations among these maps and/or other data either individually and/or in relation to each other?**

* **If so, what and where are they?**

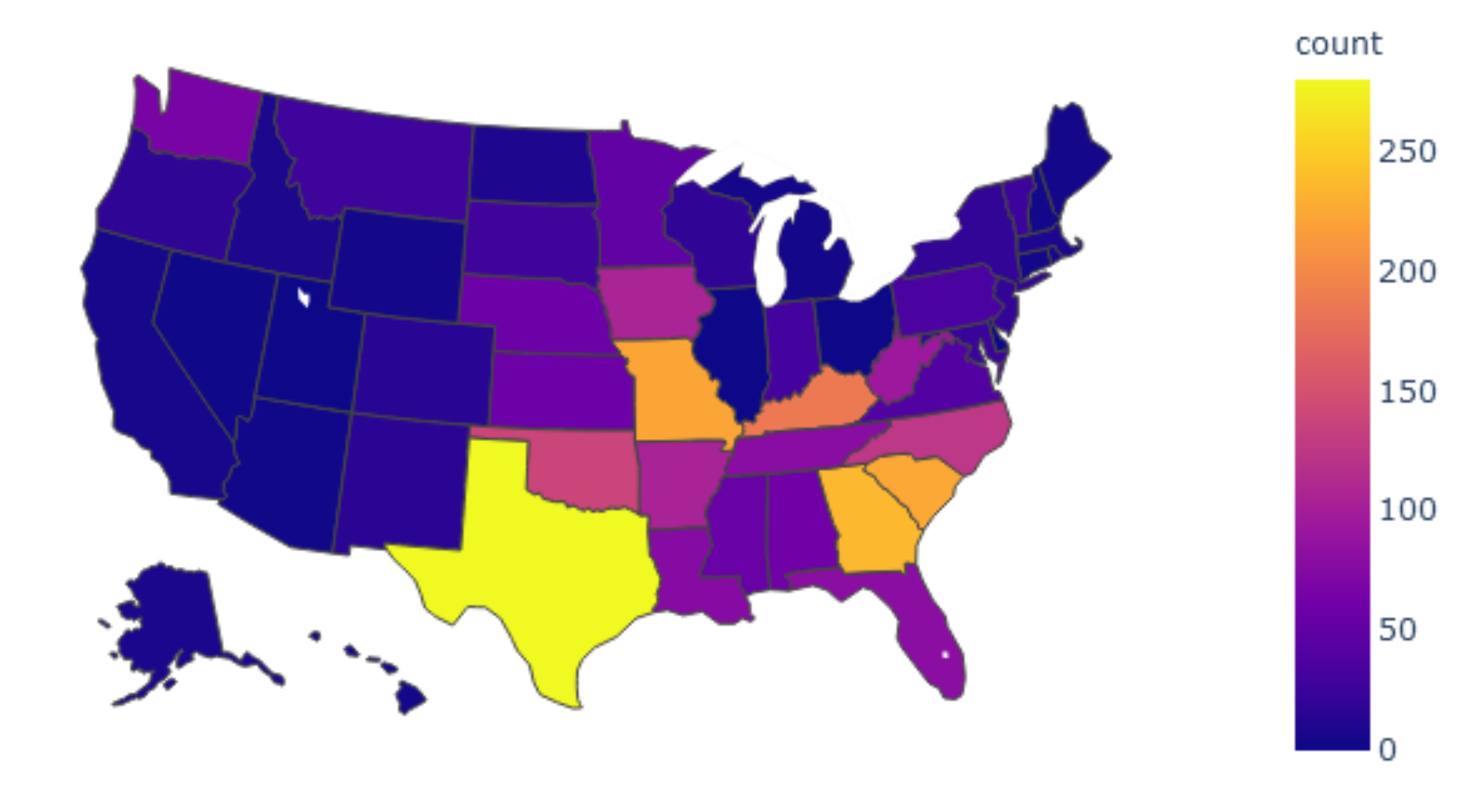
**Relative Landmass:** We noted that Texas has a substantially larger portion of declarations made by governing officials in addition to being the second largest state landmass, with Alaska being significantly larger than all others. In the case of Texas, and if accounting for landmass alone, one might improperly presume Alaska and most of the larger western states should have the greatest number of declarations. However, the data speaks contrary to this logic with Alaska and most of the larger western states generally encompassing the lowest 10% of per state declarations throughout the 64-year record. Additionally, the 7 and 3-year periods indicate ongoing recent stability.

**Population Density:** However, the geographically smaller and more densely populated states clearly show more declaration activity. Clearly, as it relates to the number of declarations per state, more is at play than landmass. Could population density be a primary variable? In that declarations are made by governing and other administrative bodies for the chief purpose of preserving life and property, we believe Next Steps should include testing state declarations per resident density.

**Annual Rainfall:** We observed a significant lack of declarations in the interior western states composing the Intermountain West. This portion of the country is famous for its uniquely arid high-desert mountain characteristics in addition to generally lower and more dispersed population levels. The latitudinal line dividing the nation in half, just above the western portion of Texas, is consistent with the walls of Rocky Mountain Region which deflect heavier moisture-rich western-moving weather patterns. Image 1, Annual State Precipitation Averages [<https://www.currentresults.com/Weather/US/average-annual-state-precipitation.php>] , shows a strong inverse color correlation - with the exception of Alaska which falls beyond the more temperate lower 48 zone. This stark correlation indicates the possibility of a very strong relationship between annual precipitation and the number of declarations per state.

**Map 1: 64-Year Total Hazard Events****Map 2: 7-Year Total Hazard Events**

**Map 3: 3-Year Total Hazard Events**



**Conclusions & Next Steps**

Understanding prior declaration histories and driving influences can assist government planners in establishing interstate support agreements and for federal planners to allocate budgets and staging facilities accordingly. The outcomes of our project provided helpful data conclusions and some interesting insights and hypotheses which we feel can produce clear value and offer additional research opportunities to obtain greater clarity on nationwide declaration patterns.

**West vs. East:** Throughout the 64-year history of federal disaster declarations in this data, the elevated arid ‘high-desert’ and less population-dense interior Rocky Mountain states (Intermountain West) and the extreme winter Alaska and Maine regions, have issued few state disaster declarations. This is in clear contrast to the eastern conditions which exist in the flatter, more humid and more densely populated Midwest, Southern and Eastern states.

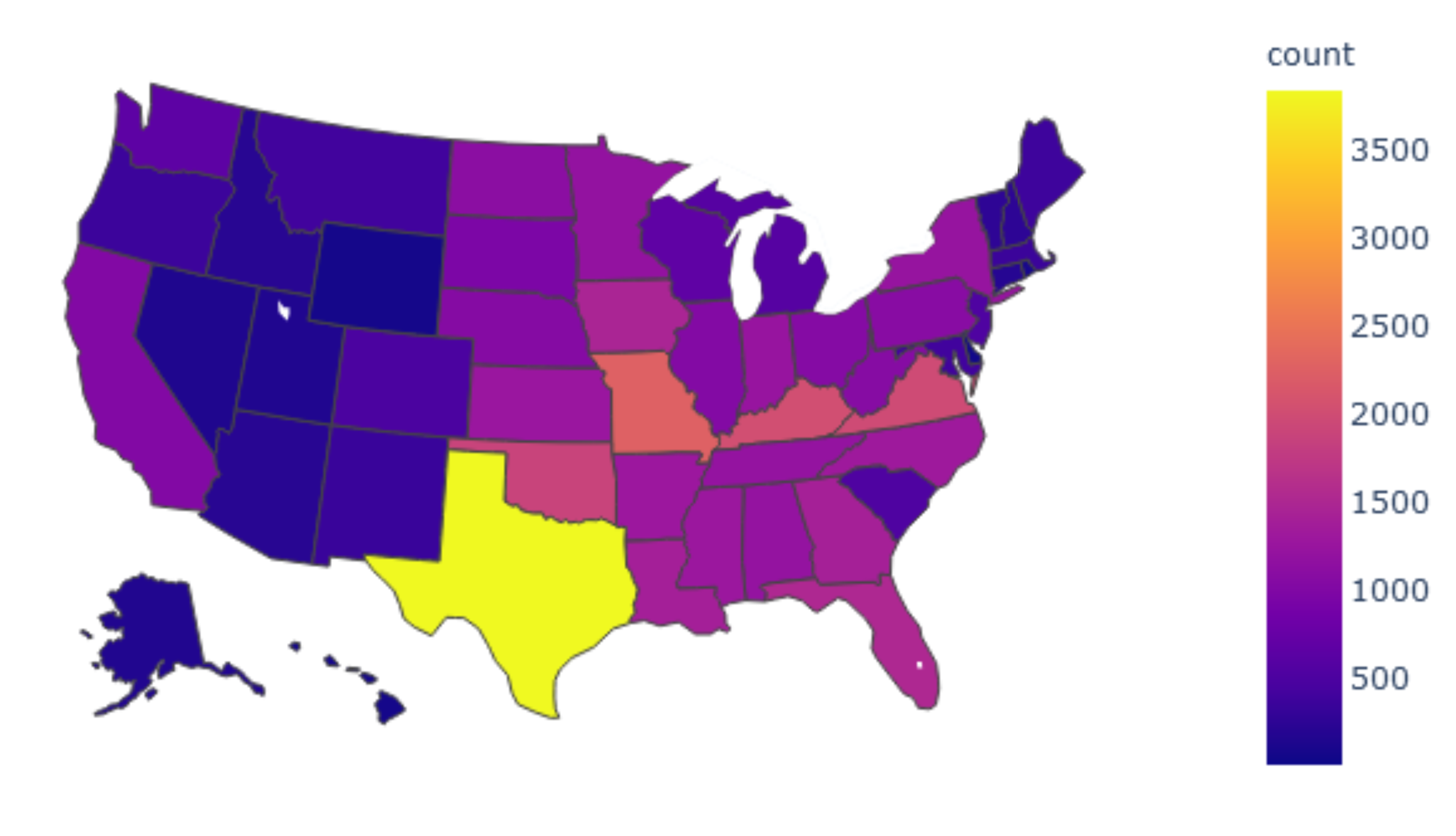
The complications for these regions east of the Intermountain West are further compounded with warm humid air updrafts from southern atlantic mixing with colder arctic systems moving down and eastward across the Midwest and Eastern regions. Unlike the larger western states, the areas east of the mountains are smaller (less landmass) and more densely populated, receive greater annual moisture and issue far more disaster declarations.

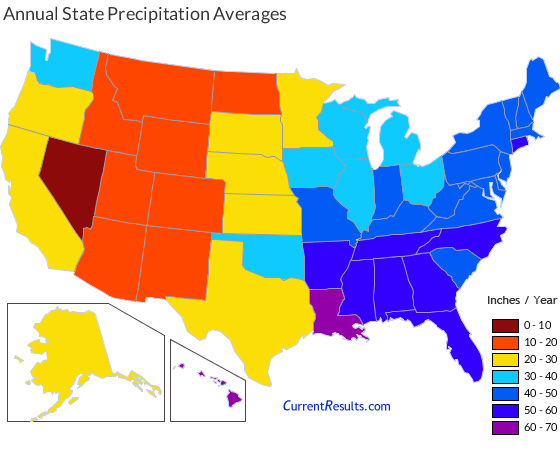
*Through our research and future research into the strong visual correlations we discovered between annual rainfall [Map 5 below] and population density [IMap 6 below], we believe a clear relationship may exist between the impact annual moisture in temperate zones and population density have on the total number of declarations issued for each state. We believe a substantial visual match between the number of declarations and annual rainfall will be discovered after reading in Map 5 data and inverting the color schema to match the U.S. maps we produced in this report. In like manner, we believe a population density map of the same schema, by state, will also provide a lesser but evidently similar heat pattern to the others.*

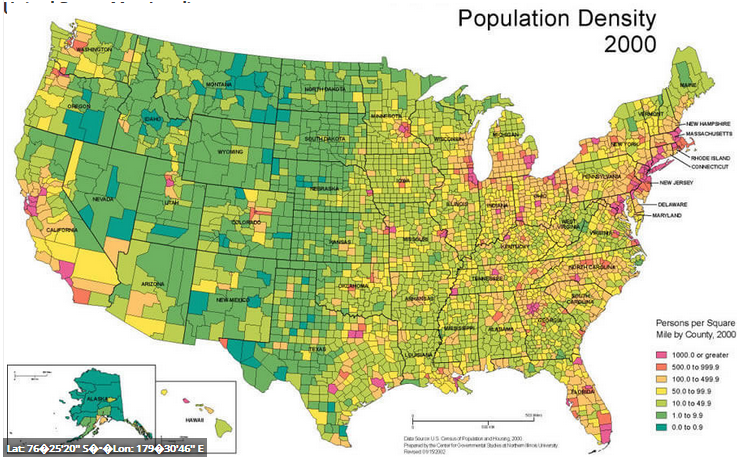
**Relative Position:** States located along the longitudinal midline of the more moisture and population rich continental Eastern United States (Missouri, Kentucky and Virginia), tend to issue more event declarations. These states suffer a greater variety of weather patterns to be continually prepared for, whereas their northern and southern neighbors have relatively consistent single-season weather risk-types to prepare for.

In consequence, it is reasonable that individuals and states with obviously limited resources - yet broader year-round risk types to prepare for [think wide and narrow preparations] - may necessitate more declarations due to a limited knowledge, skills, and budget capacity to prepare for the full breadth and depth of northern or southern sphere weather influences.

**Map 4: 64-Year Total Hazard Events**



**Map 5: Annual National Rainfall by State**

**Map 6: Population 2000 Density by State**

**Additional Next Steps**

As stated earlier, we observed correlations We believe this project would be well-suited for Associated Rule Mining (ARM) techniques and may provide value to the following related questions:

* Can ARM discover hidden relations between event types, periods and/or locations?
* Could ARM reveal a possible connection between droughts in one region creating changes to atmospheric conditions (e.g. pressure, particulates, non-standard ionization) which then influence other seemingly disconnected inter-region weather patterns?
* Is it consistently true, in all areas of the United States, that increased droughts lead to increased wildfires?
* Is it possible that increased wildfires lead to increased soil erosion which then lead to increased mudslides or increases long-term fire risks?
  + If so, do these increased mudslides occur equally in all climates? - or just the more temperate regions?

**Sources**

Dataset: Federal Disasters, <https://www.kaggle.com/fema/federal-disasters>

Map 5: Annual National Rainfall by State, <https://www.currentresults.com/Weather/US/average-annual-state-precipitation.php>

Map 6: Population 2000, Density by State, <https://www.worldmap1.com/united-states-population-density-map>