Examining the Correlation Between Demographics and the Destructive Effects of Weather Events in the US

Business Analytics for Managers (BYBG-7975-001)

Prof. RP Raghupathi

April 27th, 2020

Team Hyperion

Lu Chen

Liangrui Yu

Ryan Waffle

Wooyul Shim

**Matrix**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Hypothesis** | **Question** | **Analytic type** | **Variable(s)** | **Chart** | **Conclusion** |
| An increase in weather events from 2010 through 2017 have led to an increase in weather related injuries, death, and damage in the US. | Have weather events increased over the 8-year period? | Descriptive | Event\_ID, Year | Line Chart | Weather events have overall slightly decreased over the last 8 years. |
| As weather events have increased, have weather death, or damage increased over the 8-year period? | Predictive | Event\_ID, Injuries\_Indirect, Injuries\_Direct, Death\_Indirect, Death\_Direct, Property\_Damage, Crop\_Damage | Dual Lines | There is a positive correlation between weather events and total casualties. |
| Weather related damage differs greatly by region and by event type. | What is the distribution of damage caused by weather? | Descriptive | Count of Events, Property\_Damage, Crop\_Damage | Bar Chart | Flash floods caused the most damage, followed by flood, tornado, and hail. |
| Which weather event type causes the most Property/Corp damage? | Descriptive | Event Type, Property\_Damage, Crop\_Damage | Bubble Chart | Flash flood mostly led to property damage, and drought contributed most to crop damage. |
| Which state region has had the most weather-related damage? | Descriptive | State\_Name, Property\_Damage, Crop\_Damage | Map View | The Gulf Coast region had the most weather-related damage. |
| Which states have the lowest median household income? | Descriptive | State\_Name, Median Household Income | Horizontal Bar Chart | Mississippi had the lowest average median household income. |
| Which county has experienced the most weather-related damage? | Descriptive | County\_Name, Property\_Damage, Crop\_Damage | Heat Map | Galveston and Harris County, TX had the most amount of weather-related damage. |
| Casualties from weather events differ greatly by region and by event type. | What is the distribution of casualties and damage caused by weather for each month? | Descriptive | Month, Injuries\_Indirect, Injuries\_Direct, Death\_Indirect, Death\_Direct, Property Damage, Crop Damage | Boxplot | There is not a normal distribution for casualties and damage for months during 2010-2017. For Total Damage, August is a large outlier. |
| Which type of events cause most injuries/deaths over the 8 years? | Descriptive | Event type, Injuries\_Indirect, Injuries\_Direct, Death\_Indirect, Death\_Direct | Pie Chart | Tornado causes most injuries/deaths over the 8 years, which leads to 31.54% deaths and injuries; then is excessive heat and heat. |
| Which state region has had the most weather-related casualties? | Descriptive | State\_Name, Injuries\_Indirect, Injuries\_Direct, Death\_Indirect, Death\_Direct | Map View | The Midwest US had the most weather-related casualties. |
| Which states have the lowest rate of insured individuals? | Descriptive | State\_Name, Uninsured Rate | Tree Map | Texas, Alaska, Florida, Nevada, Georgia and Oklahoma have the highest average uninsured rate. |
| Which county has experienced the most weather-related casualties? | Descriptive | County\_Name, Injuries\_Indirect, Injuries\_Direct, Death\_Indirect, Death\_Direct | Heat Map | Counties in Midwest had the greatest number of casualties, specifically Jasper County, MO and Cleveland County, OK |
| An increase in weather events has had a positive correlation to an increased population change. | Which indicator of the severe level of the severe events influence the population change most? (Deaths directly/ crop damage/ property damage etc.) | Descriptive | State\_Name, Population\_Change, Injuries\_Indirect, Injuries\_Direct, Death\_Indirect, Death\_Direct, Property\_Damage, Crop\_Damage | Dual combination plot of line and area charts | The total number of direct deaths and injuries caused by the events influence the population change most. |
| Population change in which states suffer most from the severe weather events? | Descriptive | State\_Name, Population\_Change, Count of event | Dual combination plot of line and bar charts | Some middle states’ population change suffers most from the extreme events. |
| Which type of events influence the population change most? | Descriptive | Event type, Population change | Bar chart | Dust devil, debris blow, and lightning influence population change most. |
| Can we estimate the net change in migration with current net migration and weather events? | Predictive | Count of Events, State\_Name, Net\_Migration | Scatter plot + Extrapolation | There is a positive correlation between weather events and net migration. |
| Most states’ businesses are not easily influenced by the extreme weather events. | As the number of total businesses in a state increases, does crop damage from storms increase? | Predictive | Total Business, Crop\_Damage | Scatterplot | As total businesses increase, the amount of crop damage caused by storms also increases |
| As economic indicators worsen (high unemployment, low median household income, high poverty rate, low wage quotient), the rate of injury, death, and damage per event increase. | As unemployment rate \* poverty rate increases, do weather event casualties per event increase? | Predictive | Unemployment Rate, Poverty Rate, Injuries\_Indirect, Injuries\_Direct, Death\_Indirect, Death\_Direct | Scatterplot | There is a positive linear relationship between unemployment rate and weather event casualties, and a positive linear relationship between poverty rate and weather event casualties (significant) |
| As median household income \* wage quotient decreases, does weather event casualties per event increase? | Predictive | Median Household Income, Wage Quotient, Crop\_Damage, Property Damage | Scatterplot | There is a negative linear relationship between median household income and weather event damage (significant), and a negative linear relationship between wage quotient and weather event damage. |
| In recent years, people made better preparation for the extreme events to decrease damage. | How do casualties per event change in these years and what will the numbers be in the next few years? | Predictive | Casualties per event | Dynamic graph | Compared with the first two years, casualties in the next few years decreased a lot and are predicted to keep in a low level in the next two years. |

**Statistics:**

We will look to create a correlation matrix using all numeric demographic independent variables. Given the value of each variable and its correlation to the other variables, we can determine which demographics are most correlated with one another, and if we can expect any similar results for hypotheses using these variables. For example, if a strong positive correlation exists between unemployment rate and poverty rate, we can expect that an association between unemployment rate and storm deaths should be similar to an association between poverty rate and storm deaths.

**Machine Learning Model:**

We will look to run an Artificial Neural Network to determine if the count of weather events, weather event type, and total damages can predict whether or not a death or injury (a death or injury not directly caused by the weather event) can occur in any US county. Using the above variables (county, count events, event type, total damage) as the input variables, we will determine if they will predict if either a death will occur (1) or not (0) in a county over a period of time. We will use SPSS Modeler to build the ANN model and output the results.

**Number of Weather Events in the US per Year**

A. Figure 1 is a line chart which shows how many weather events happened in the US per year from 2010 to 2017.

B. This chart indicates a minor downtrend in the number of total weather events from 2010 to 2017 except for 2011, with a peak of 76,889 cases (2011 Tornado outbreak, with hundreds of confirmed tornados in a 4-day span).

C. The number of weather events may keep declining slowly or flatten in the future; however, careful examination of deaths and damage must be considered for policy recommendations.

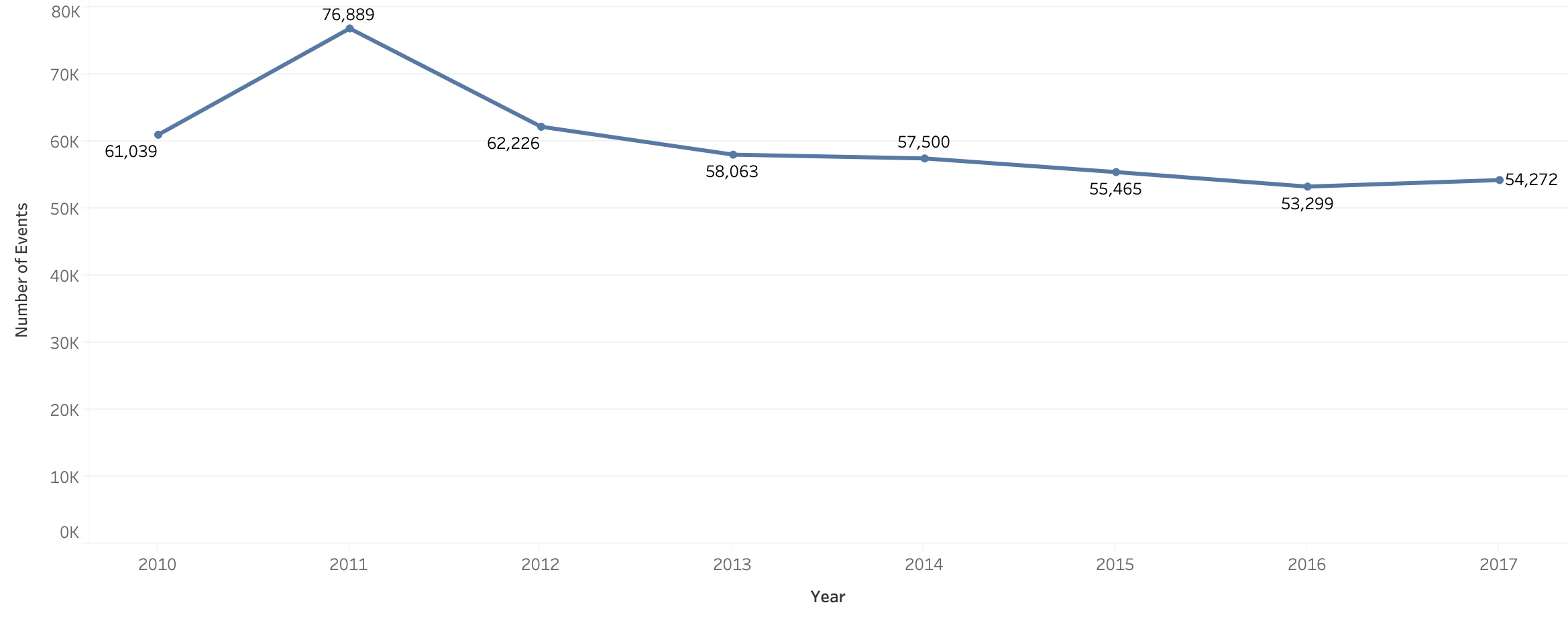
****

Figure 1: Line Chart of Number of Weather Events in the US from 2010-2017-Wooyul Shim

**Total Casualties and Total Damage caused by Weather Events in the US**

A. Figure 2 is a line chart that shows total casualties (injuries and deaths, both direct and indirect) and total damage (property damage and crop damage) in the US from 2010 to 2017.

B. The overall trends of total casualties and total number of weather events are quite similar. However, the total damage number sharply went up to more than 60 billion in 2017.

C.

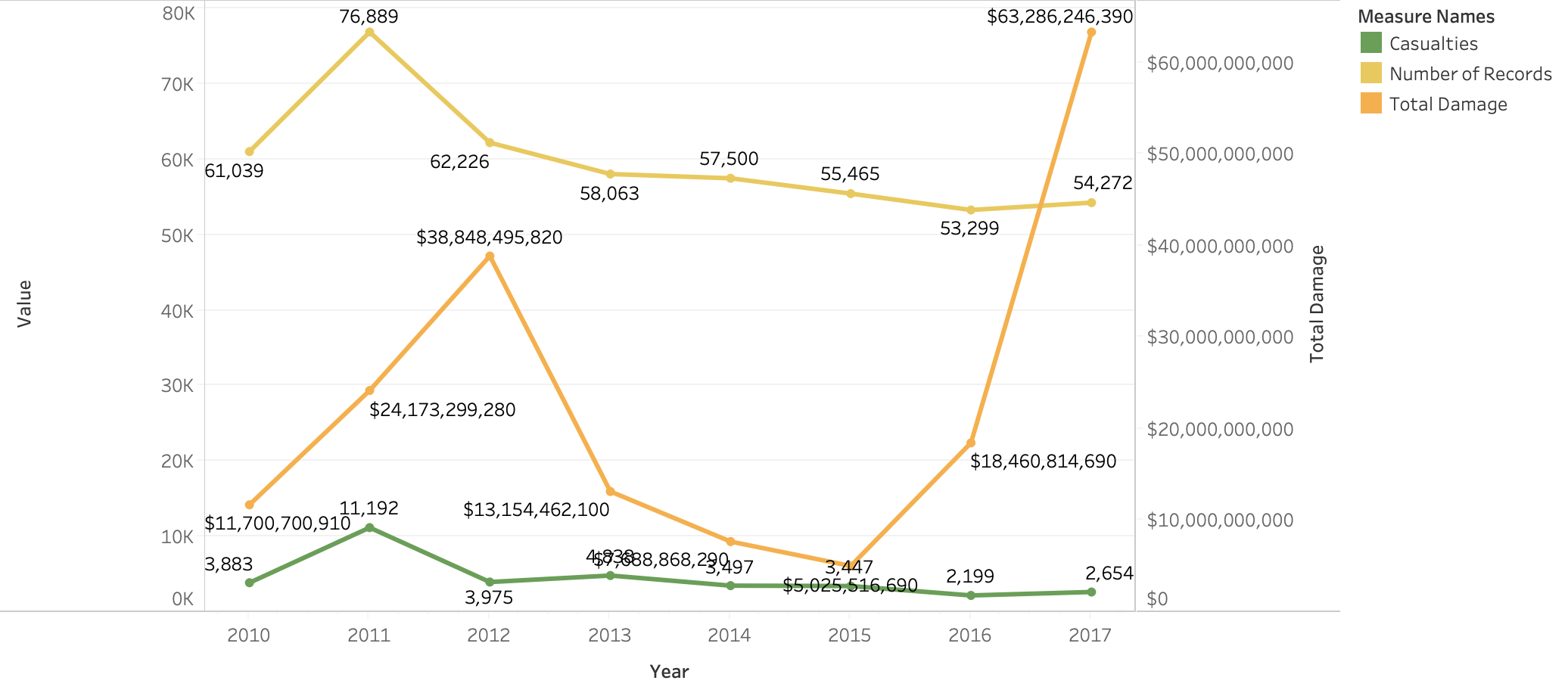
****

Figure 2: Line Chart for Weather Related Casualties and Damage in the US from 2010-2017- Lu Chen

**Total Damage of Each Type of Weather Event**

A. Figure 3 is a bar chart that shows total damage (property damage and crop damage) caused by each type of weather event, listed in descending order.

B. This chart indicates that flash floods result in the most damage, followed by flood, tornado, and hail.

C. This chart implies that most weather events do not result in significant damage; People should not be overreacting to weather events since about ⅔ types of weather events lead to little damage.

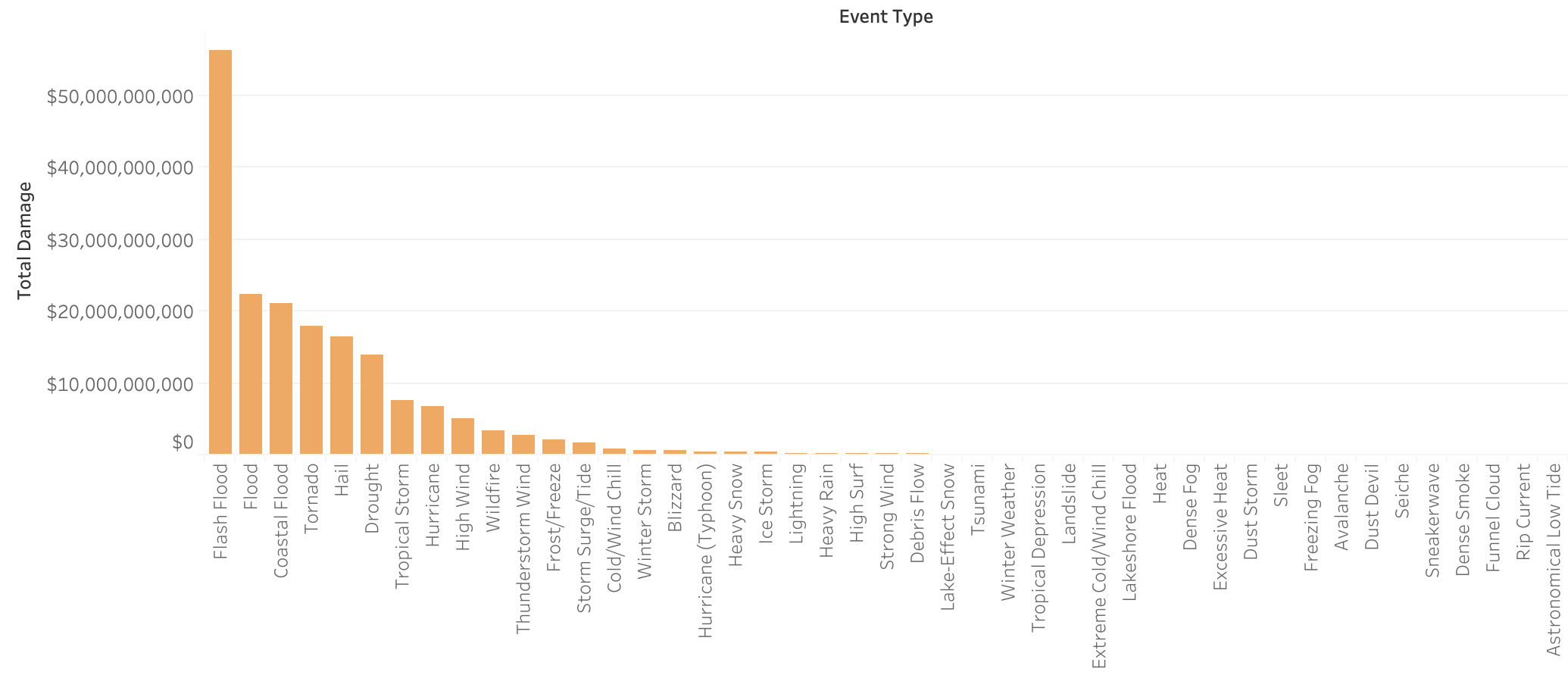
****

Figure 3: Bar Chart for Total Damage of Each Type of Weather Event - Lu Chen

**Property Damage and Crop Damage for Each Type of Weather Event**

A. Figure 4 is a drill down bubble chart showing which type of weather event contributes most to property and crop damage, respectively.

B. This chart indicates that flash flood mostly leads to property damage, and drought contributes most to crop damage.

C. This chart implies that governors and the federal government should be prepared to deal with severe outcomes of drought and floods, specifically in communities where these types of events are most common.

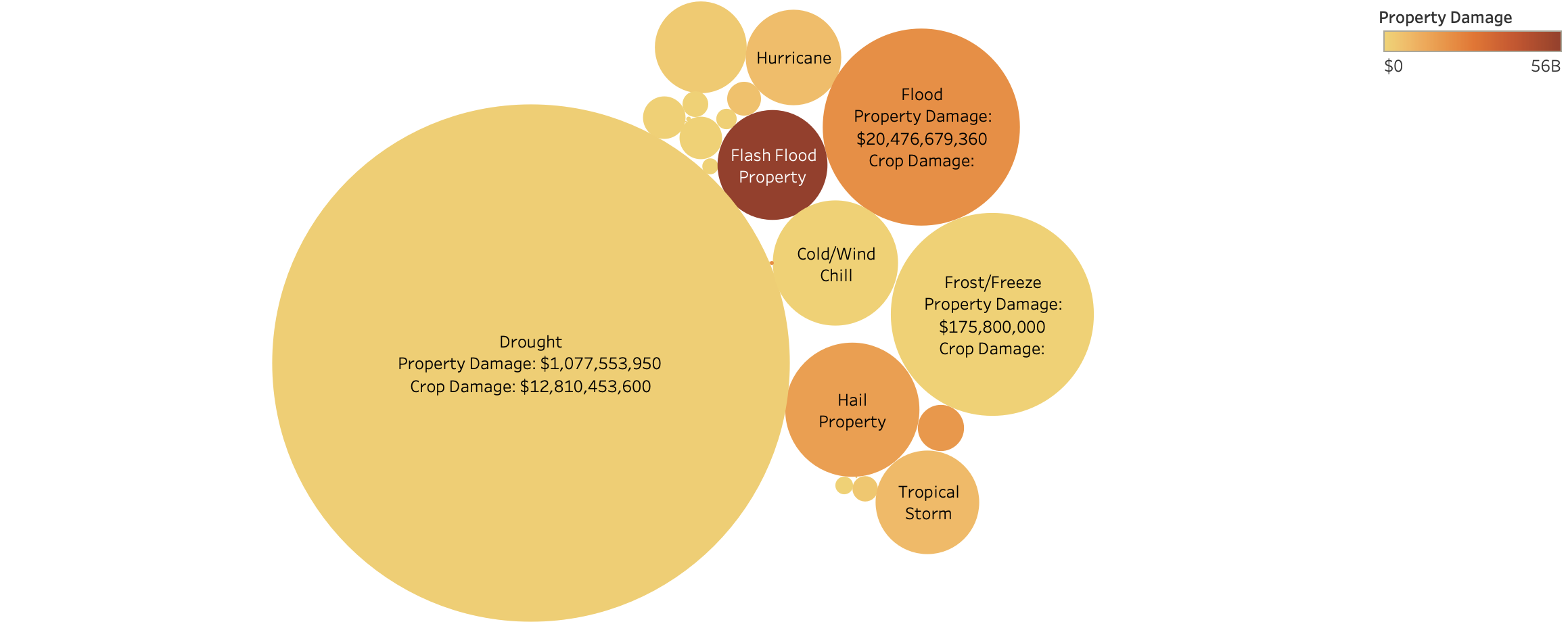
****

Figure 4: Bubble Chart for Drill down Damage for Each Type of Weather Event - Lu Chen

**Total Damage Caused by Weather Events per US Region**

A. Figure 5 is a map which shows the total damage (property and crop) per US Region (East Coast, Gulf Coast, Midwest, Rocky Mountain, West Coast, Alaska & Hawaii) in the US caused by any recorded weather event, from 2010 – 2017.

B. This map indicates that weather events in the US Gulf Coast have been the costliest in their damage, causing over $90 billion in damages during the 8-year span, while no region saw less than $123 million in weather related damages.

C. The implication from this chart is that gulf coast states are most at risk for damage and most in need of government funding for mitigation and recovery and in need of public planning for evacuations, shelters, and safety.

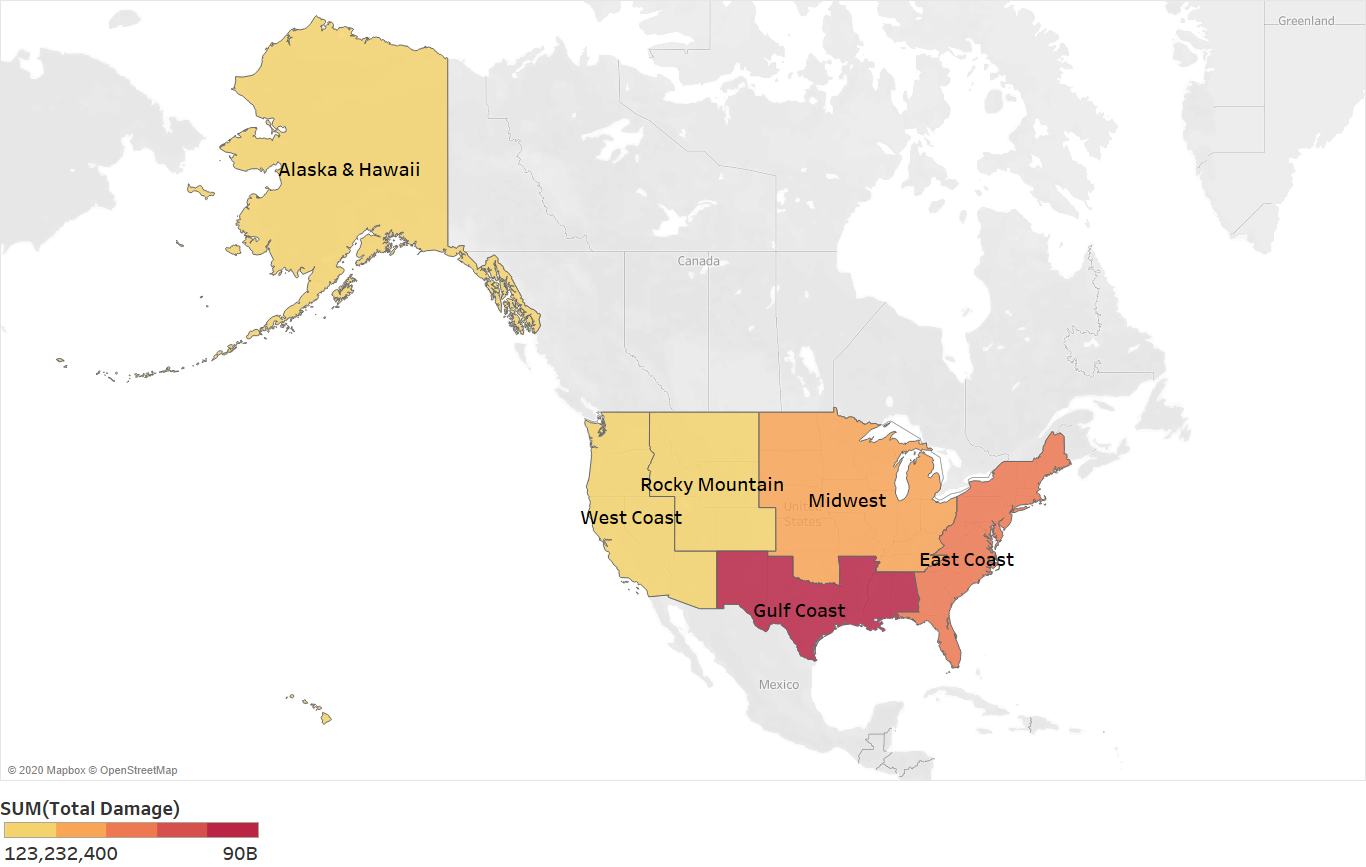


Figure 5: Map of total weather damage for 6 US regions from 2010 – 2017 - Wooyul Shim

**Average Median Household Income in US per State**

A. Figure 6 is a horizon bar chart showing the average median household income in each state in the US, listing in ascending order.

B. This chart indicates that Mississippi is the state with lowest average median household income.

C. Governors of states with low median household incomes, which appear mostly in the south and gulf coast regions of the US, should pay extra attention to reduce or mitigate damage and injury related to weather events.

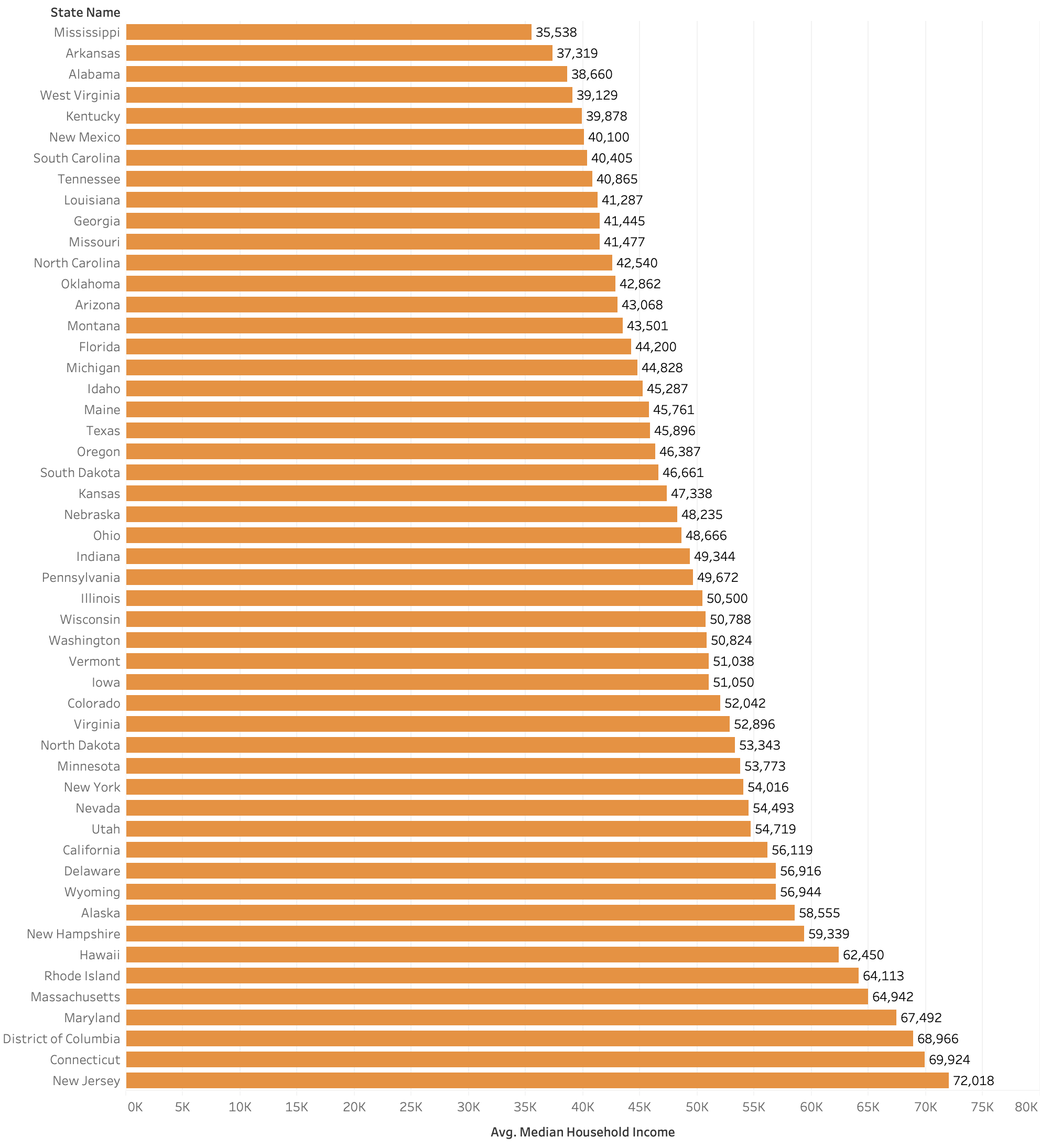
****

Figure 6: Horizon Bar Chart for Average Median Household Income per State- Lu Chen

**Total Damage Caused by Weather Events in US Gulf Coast**

A. Figure 7 is a drilled-down heat map which shows the total damage (property and crop) per county for the Gulf Coast US region caused by any recorded weather event, from 2010 – 2017.

B. This map indicates that the most amount of damage since 2010 from weather events occurred in coastal counties in this region.

C. The implication from this chart is that coast counties are at very high risk of damage, likely from hurricanes, tropical storms, and floods, indicating need for government funding of infrastructure and policy to respond to these threats, including levy building, property building codes, and shelters.

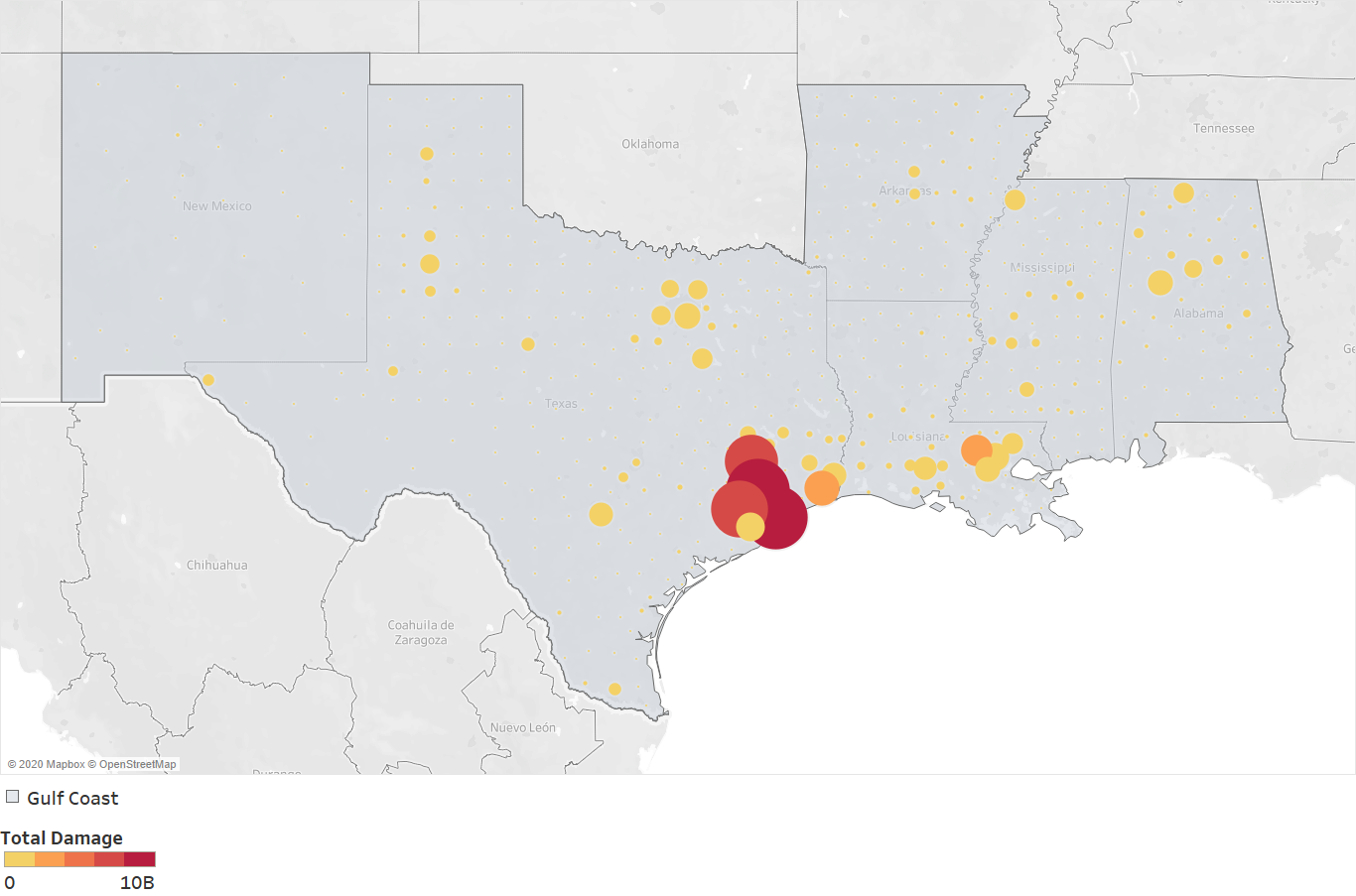


Figure 7: Heat Map of total weather damage for US Gulf Coast Region from 2010 – 2017 - Wooyul Shim

**Distribution of Damage and Casualties by Month**

A. Figure 8 is a bar chart that shows the monthly distribution for both weather-related damage and casualties.

B. This indicates that August and October are months with the most weather damage, while April and July saw the largest number of casualties.

C. The implication from this chart is that governments should increase funding and preparedness for these high impact months.

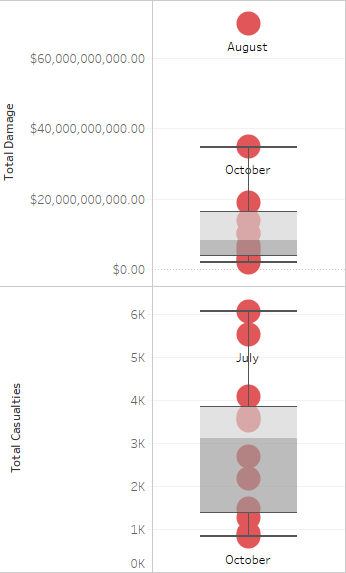
****

Figure 8: Boxplot of Damage and Casualties per Month -Wooyul Shim

**Comparison of deaths and injuries caused by different types of events**

A. Figure 9 is a pie chart which shows the count and percentage of total deaths and injuries caused by different types of weather events, from 2010 – 2017.

B. This map indicates that the most amount of deaths and injuries in total since 2010 was caused by Tornado, which leads to 31.54% deaths and injuries; then is excessive heat and heat.

C. The implication from this chart is that tornadoes, excessive heat, heat and winter weather threaten lives most. This indicates the need for forecasting and making preparation for these extreme events.

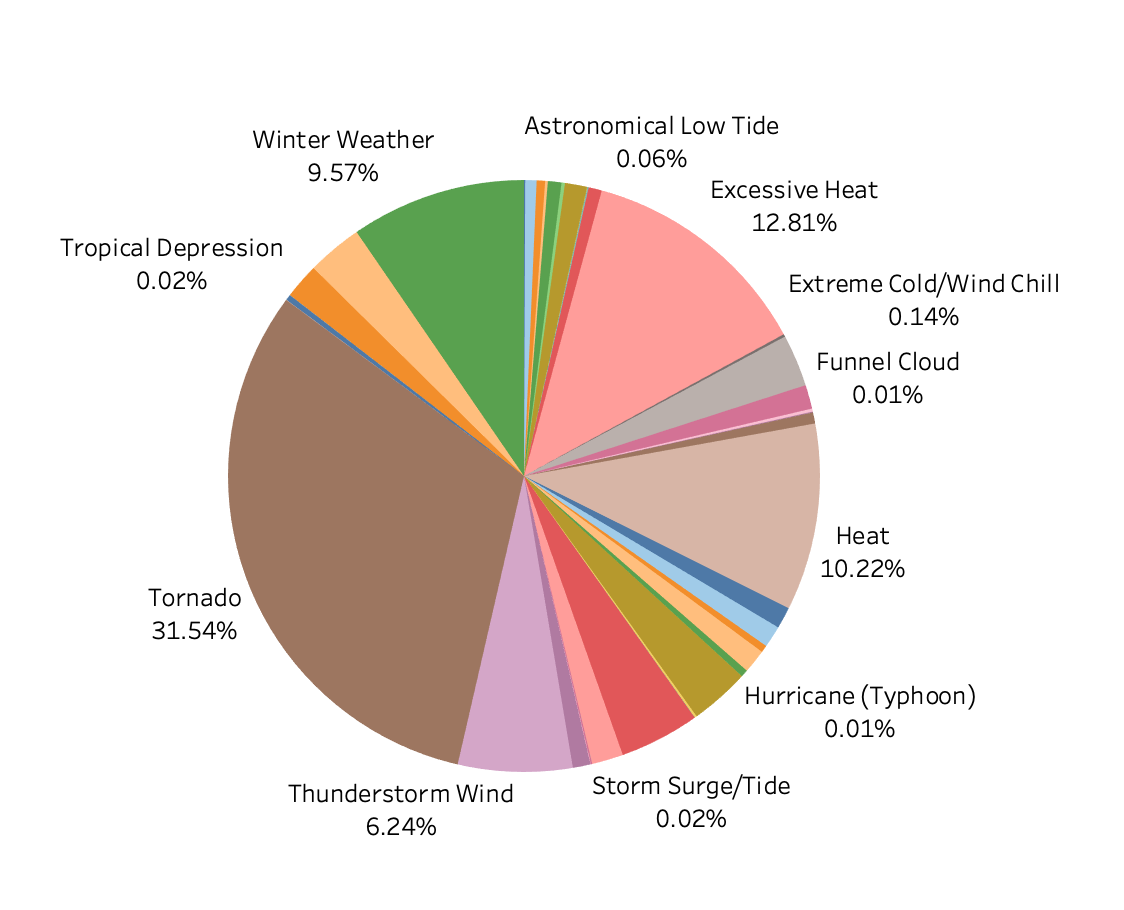
****

Figure 9: Pie Chart of count of Deaths and Injuries caused by different Events from 2010 – 2017 -Liangrui Yu

**Total Casualties by Weather Events per US Region**

A. Figure 10 is a map which shows the total casualties (injuries and deaths, both direct and indirect) per US Region (East Coast, Gulf Coast, Midwest, Rocky Mountain, West Coast, Alaska & Hawaii) in the US caused by any recorded weather event, from 2010 – 2017.

B. This map indicates that weather events in the Midwest have resulted in the greatest number of injuries and loss of life, with over 12,000 deaths over the 8-year period, while no region saw fewer than 100 injuries and deaths.

C. This chart suggests the significant need for hospitals and medical care funding in the Midwest as both direct and indirect casualties are common in the Midwest from weather events. An extended wait for an ambulance, lengthy travel time for medical care, or lack of medical insurance can result in a preventable casualty.

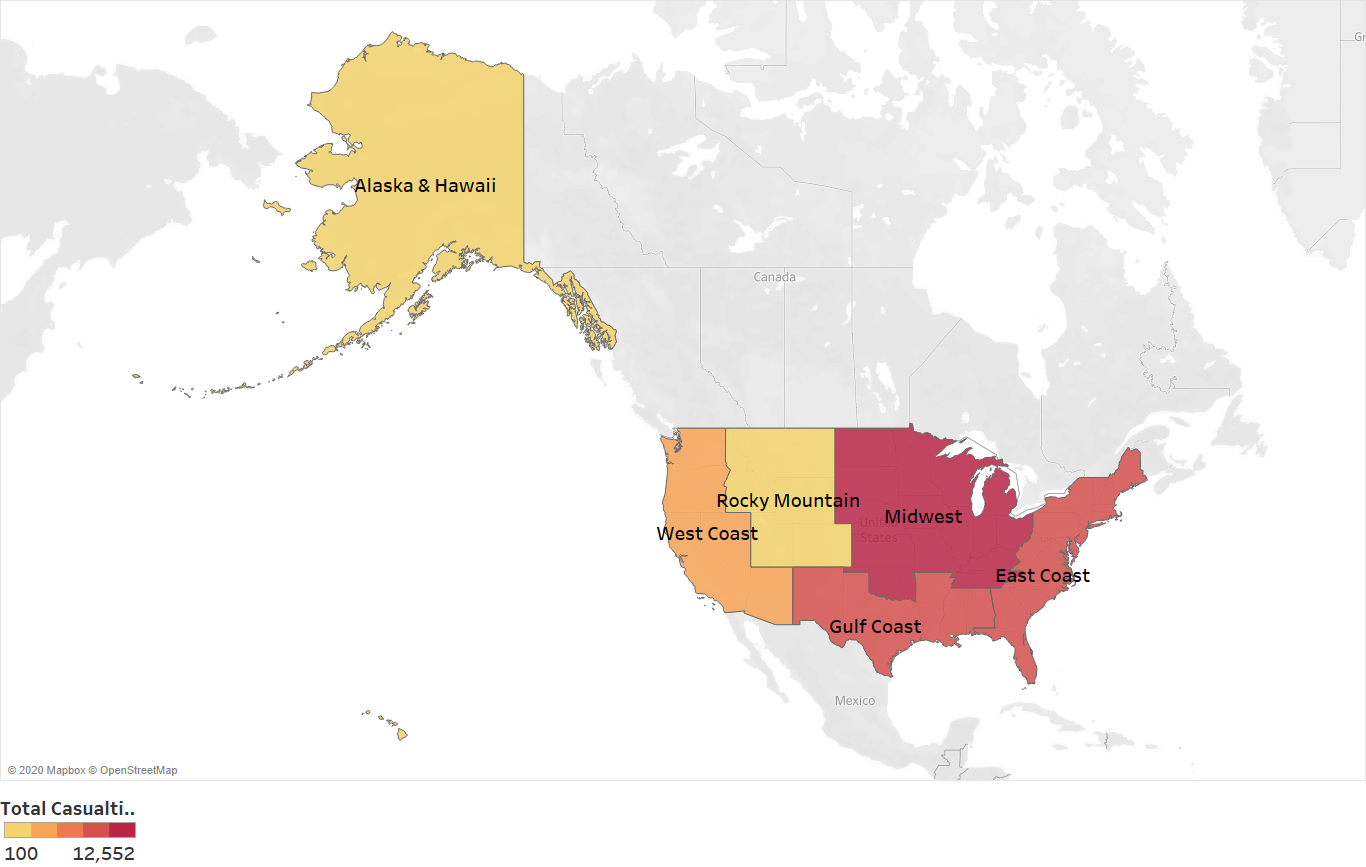


Figure 10: Map of total weather casualties for 6 US regions from 2010 – 2017 – Ryan Waffle

**Uninsured rate of different states**

A. Figure 11is a tree map which shows the average uninsured rate of different states from 2010 to 2017.

B. This map indicates that Texas, Alaska, Florida, Nevada, Georgia and Oklahoma have the highest average uninsured rate, while Massachusetts, District of Columbia, Hawaii, Vermont, Minnesota have the lowest average uninsured rate.

C. The implication from this chart is that typically, the southern states have higher uninsured rates than the northern states (except for Alaska and Hawaii), putting their residents at higher risk of indirect injury and death, due to a lack of access and means to pay for medical care.

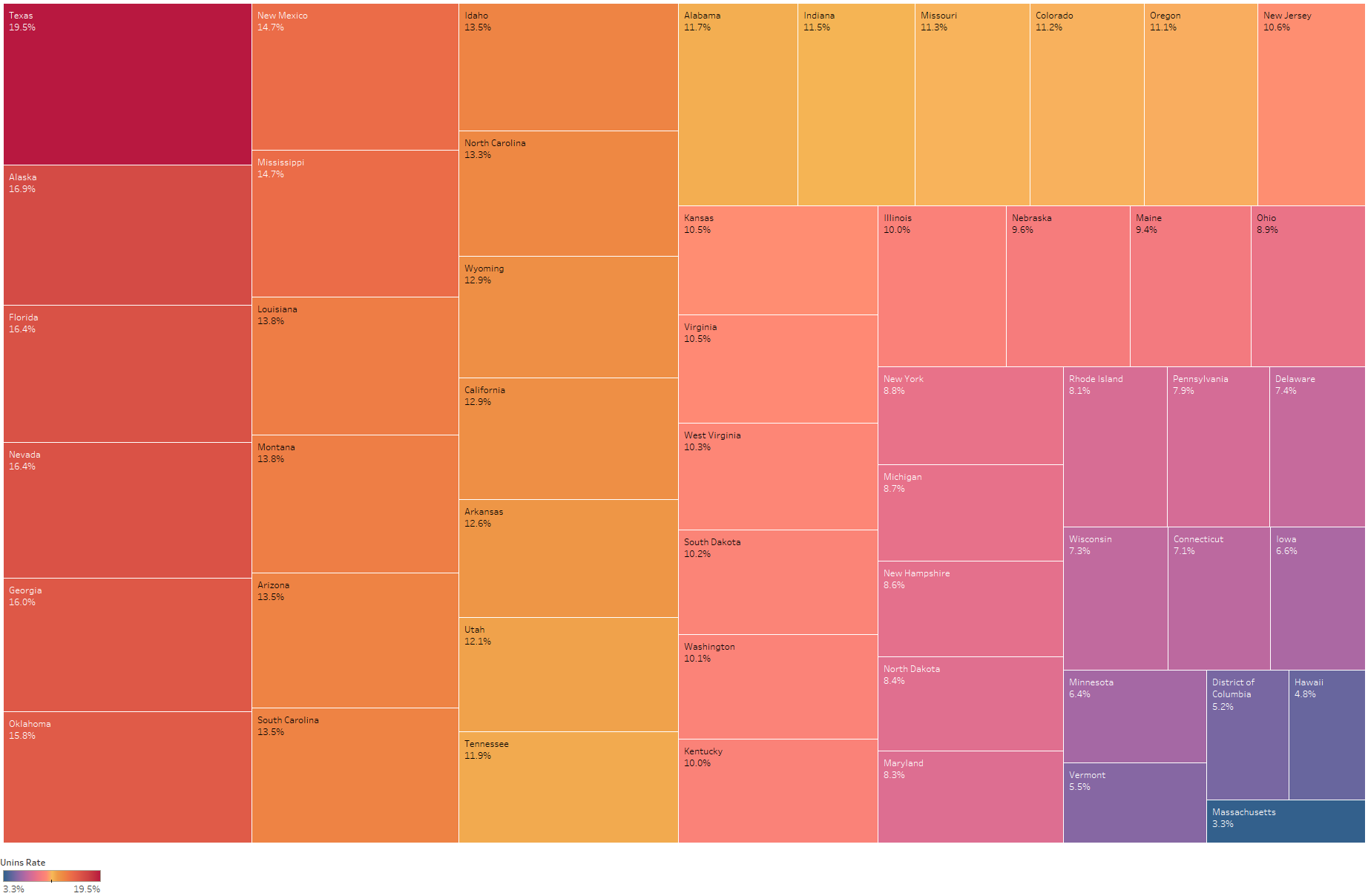
****

Figure 11: Tree Map of uninsured rate of different states-Liangrui Yu

**Total Casualties Caused by Weather Events in US Midwest region**

A. Figure 4 is a drilled-down heat map which shows the total casualties (property and crop) per county for the Midwest US region caused by any recorded weather event, from 2010 – 2017.

B. This map indicates that the most amount of damage since 2010 from weather events has occurred in counties in Kansas, Missouri and Oklahoma, which are a part of the ‘Tornado Alley’ region of the US.

C. The implication from this chart is the Midwest is highly susceptible to death and injury from certain events, most notably Tornados. This indicates the need for funding for warning systems, shelters, and increased funding for hospitals and health care overall.

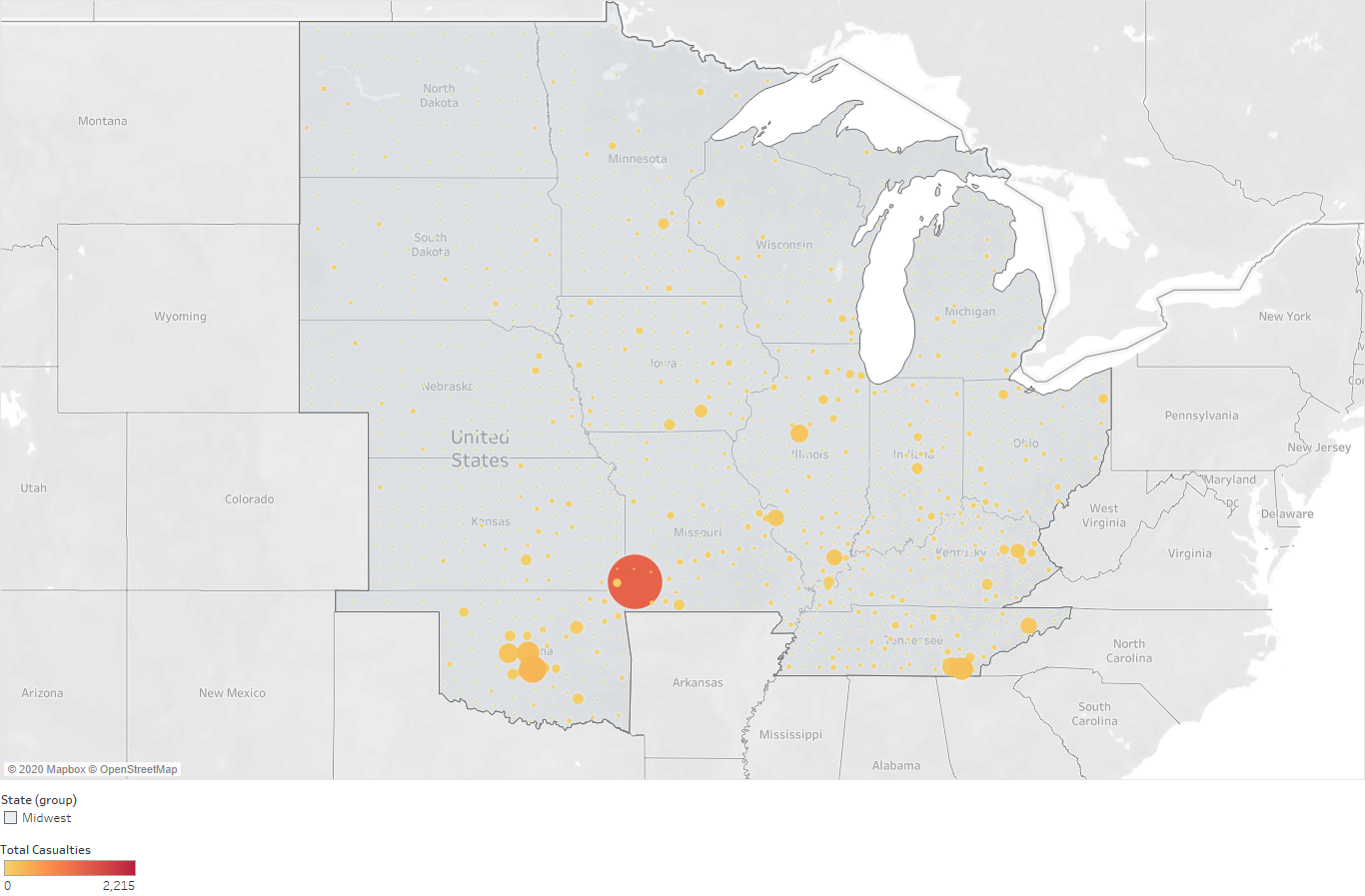


Figure 12: Heat Map of total weather-related casualties for US Midwest Region from 2010 – 2017 – Ryan Waffle

**Indicator caused by the extreme events that influence population change most**

A. Figure 13 is a dual combination plot of line and area chart which shows the different indicators (Crop damage, Direct deaths and injuries, indirect deaths and injuries) caused by the extreme events from 2010 to 2017.

B. This plot shows that among the indicators, the total number of direct deaths and injuries caused by the events influence the population change most.

C. The implication from this chart is that after the first two years, the crop damage and direct deaths and injuries declined significantly, which indicates that the society made great advance in preparing for the extreme events.

**A close up of a map

Description automatically generated**

Figure 13 - Dual Combination Plot of indicator that influence population change most -Liangrui Yu

**States’ population change suffers most from the extreme weather events**

A. Figure is a dual combination plot of line and bar chart which shows which states suffer most from the extreme weather events using the indicator of direct deaths and injuries from 2010 to 2017.

B. This plot shows that some middle states’ population change suffers most from the extreme events.

C. The implication from this chart is that compared with other states, middle states have a simpler economic mix, and the income of the residents rely less on the geographic. Moreover, most middle states are agricultural states and are more easily influenced by the extreme events.

**A picture containing text, map, table, filled

Description automatically generated**

Figure 14 - Dual Combination Plot of Population Changes in which states are influenced by deaths and injuries most - Liangrui Yu

**Events type that influence population change most**

A. Figure 15 is a bar chart which shows average population change that is affected by different types of events.

B. This chart shows that dust devil, debris blow, and lightning influence population change most.

C. The implication from this chart is that compared with some more severe events like tornadoes, events that are more common influence the population change more.

**A picture containing table, building, red, sitting

Description automatically generated**

Figure 15 - Bar Chart of Population Change influenced by Event Type-Liangrui Yu

**Number of Weather Events and Net Migration per State**

A. Figure 16 is a scatter plot showing the number of weather events and net migration in each state from 2010-2017.

B. This chart indicates that with the number of weather events increasing, the number of net migrations will increase too.

C. The implication from this chart is that there may be some other more significant factors that influence migration.

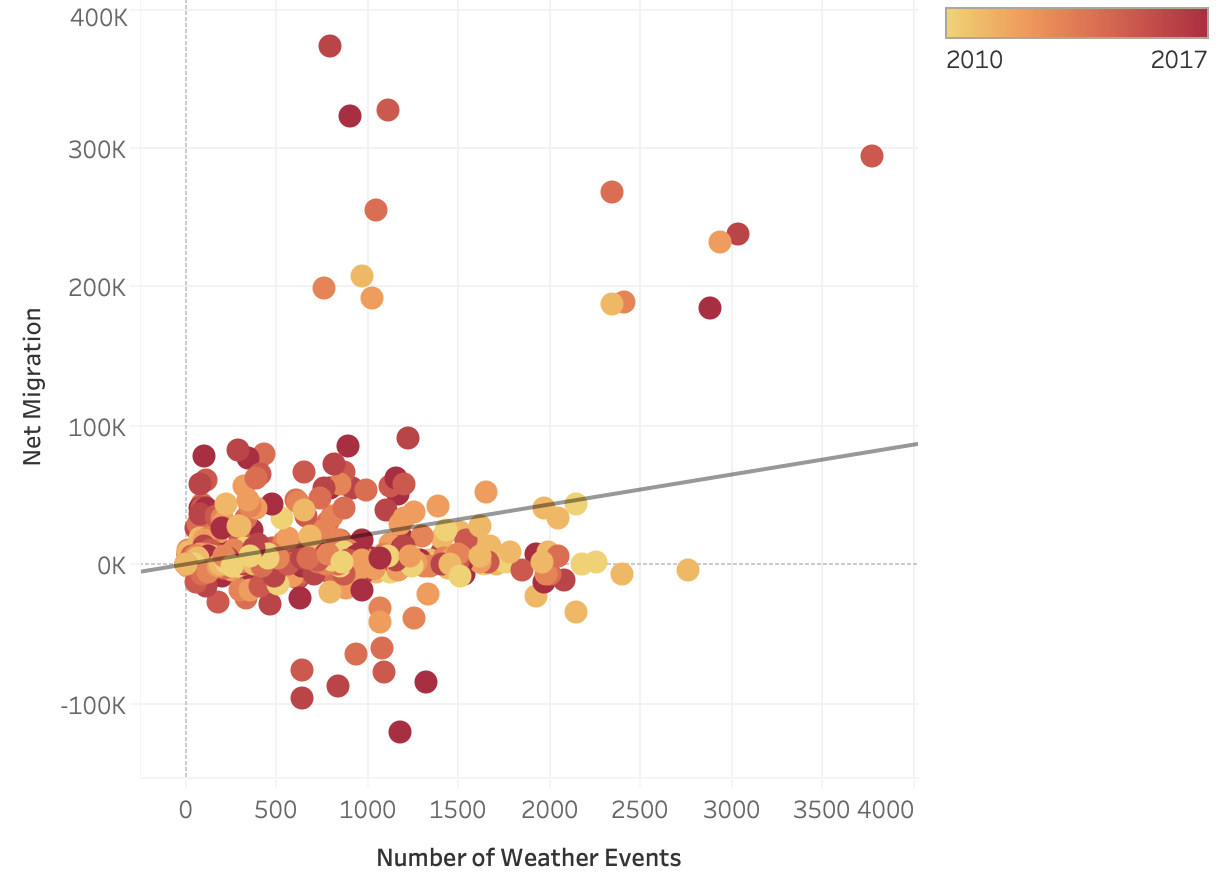
****

Figure 16: Scatter Plot of Number of weather Events and Net Migration per State from 2010-2017 - Lu Chen

**Scatter plot of Total Business vs Crop Damage**

A. Figure 17 is a scatter plot showing the total number of businesses per state vs the total weather-related crop damage per state.

B. This chart shows a positive relationship between total businesses and crop damage, i.e. as the total number of businesses increases, crop damage from storms increases.

C. The implication from this chart is that areas with a large number of businesses are at high risk of having crops damaged. This calls for a need to have businesses increase protection of their crops.

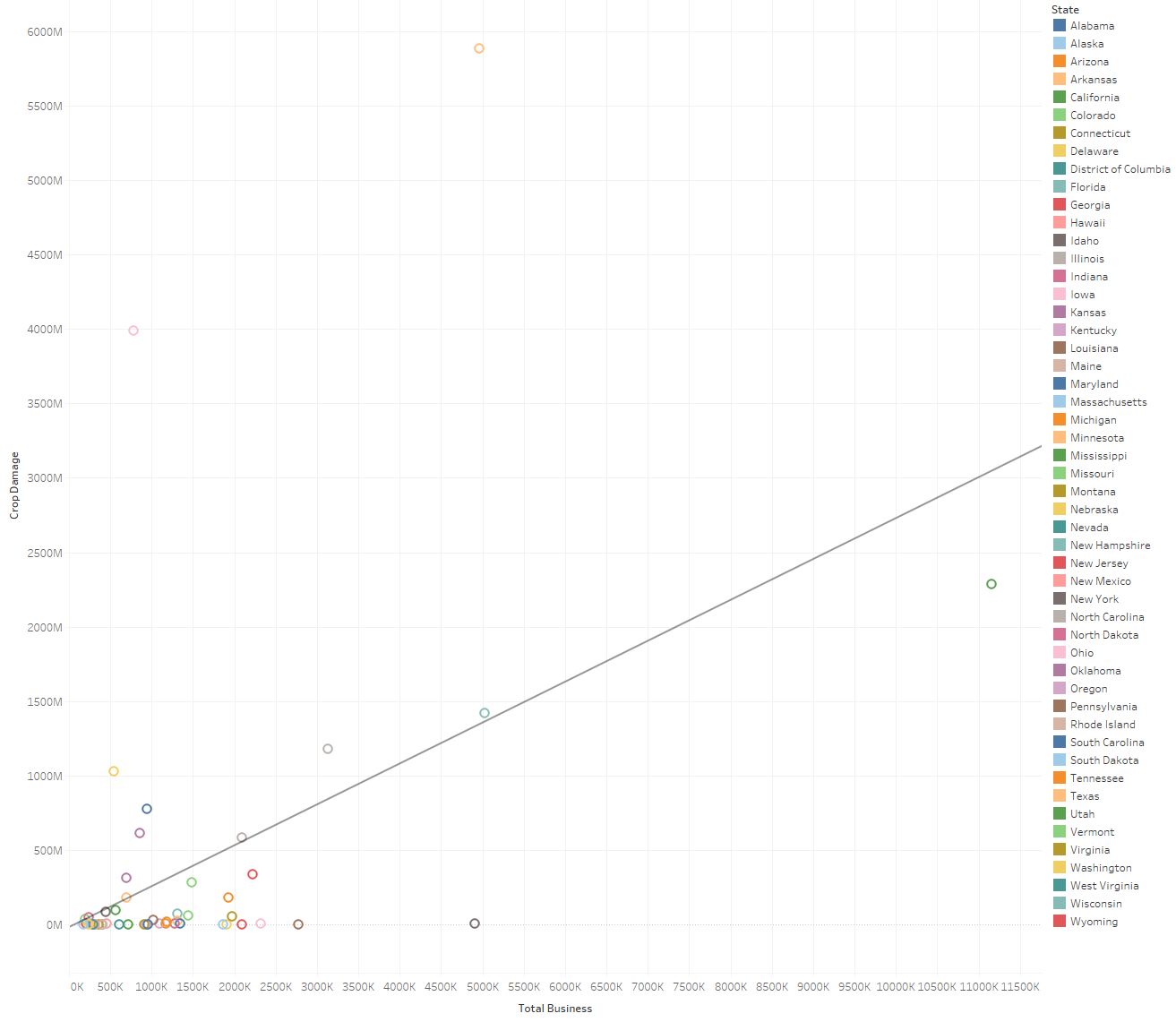
****

Figure 17: Scatter Plot of Number of weather Events and Net Migration per State from 2010-2017- Ryan Waffle

**Scatter Plot of Unemployment Rate and Poverty Rate vs Total Casualties**

A. Figure 18 is a scatter plot showing unemployment rate and poverty rate vs total weather-related casualties.

B. This chart shows a positive correlation between unemployment rate and weather-related casualties and poverty rate and weather-related casualties. The latter is statistically significant.

C. The implication from this chart is that areas with high unemployment and poverty rates are susceptible to weather related injuries and death, calling for local and state governments to provide funding for vulnerable populations.

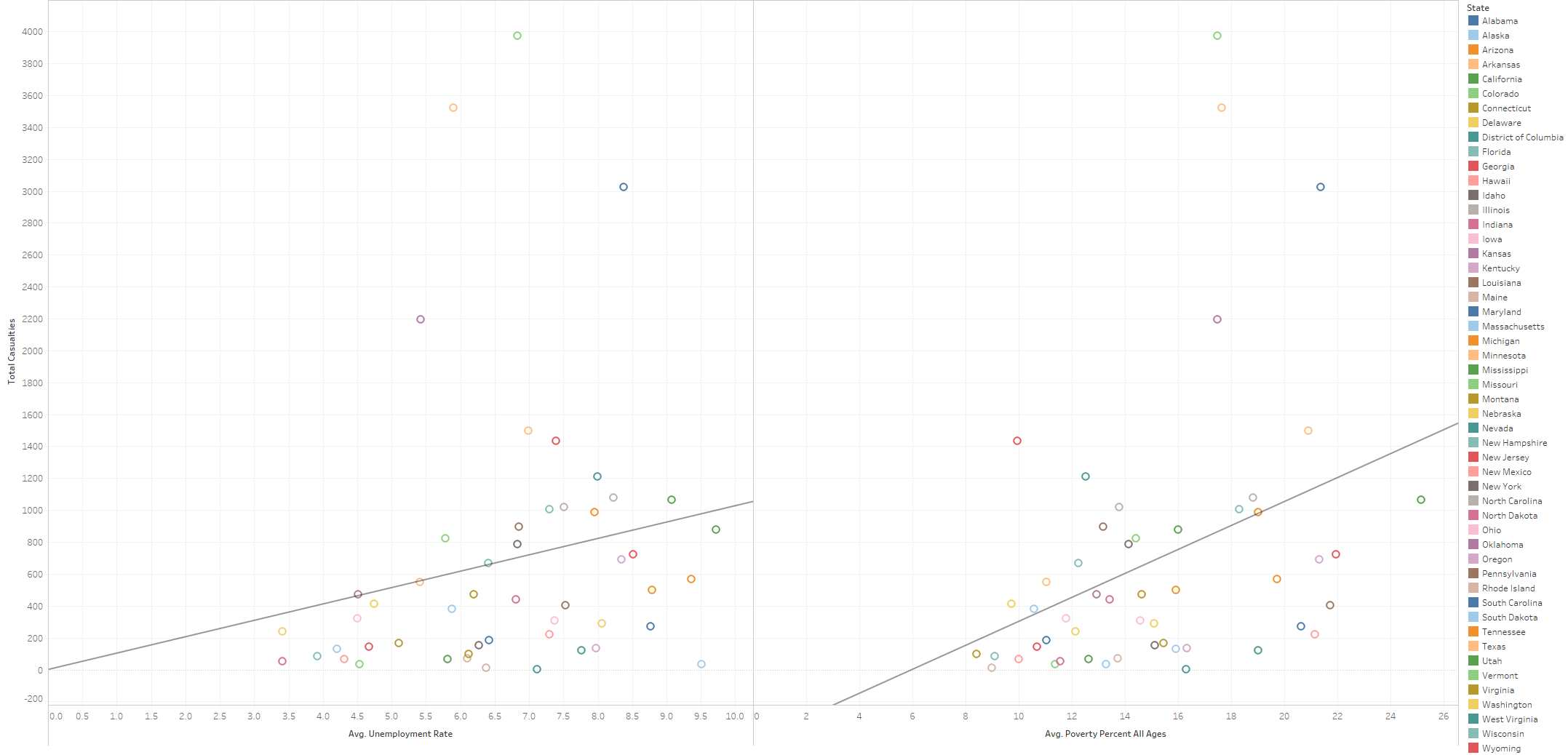
****

Figure 18: Scatter Plot of Number of weather Events and Net Migration per State from 2010-2017 - Ryan Waffle

**Scatter Plot of Median Household Income and Wage Quotient vs Total Weather Damages**

A. Figure 19 is a scatter plot showing median household income and wage quotient vs total weather-related damages.

B. This chart shows a negative correlation both independent variables and total weather-related damages, with the median household income vs total weather-related damages as statistically significant.

C. The implication from this chart is that poorer communities are more likely to experience a high cost of weather events. Poorer communities often provide fewer taxes for government expenditure, leading to inadequate procedures, shelters and overall infrastructure.

****

Figure 19: Scatter Plot of Number of weather Events and Net Migration per State from 2010-2017- Ryan Waffle

**Count of casualties per event every year and in the future two years**

A. Figure 20 is a dynamic graph which shows the count of casualties from 2010-2017 and also predicts the number in 2018 and 2019.

B. This chart indicates that since 2011, the casualties per event has decreased significantly. Also, the number will be about 0.0610 in the next two years.

C. The implication from this chart is that casualties per event have been kept steady down and are predicted to keep smooth in the next two years, which shows the ability to prevent casualties has increased a lot.

**A close up of a map

Description automatically generated**

Figure 20: Dynamic graph of casualties per event every year and in the future two years- Wooyul Shim

**Statistical Model**

**Demographic and Economic Independent Variable Correlation Matrix**

Hypothesis: There is a strong positive correlation between many of the demographic and economic independent variables, specifically between the Poverty, Unemployment, and Uninsured figures.

A. Figure 21 is a correlation matrix between all independent variables in regard to county demographic and economic data.

B. Figure 21 shows that there are strong positive correlations between many of the independent variables. The strongest correlations are between number employed, number unemployed, number insured, number uninsured, and poverty estimates.

C. Unsurprisingly, wealth demographics are often correlated to one another - areas with higher wages and wealth are likely to have lower unemployment, higher insurance rates etc. This would suggest that any further models may see some duplication from using these variables, and that correlated indicators likely will show similar associations to target variables.

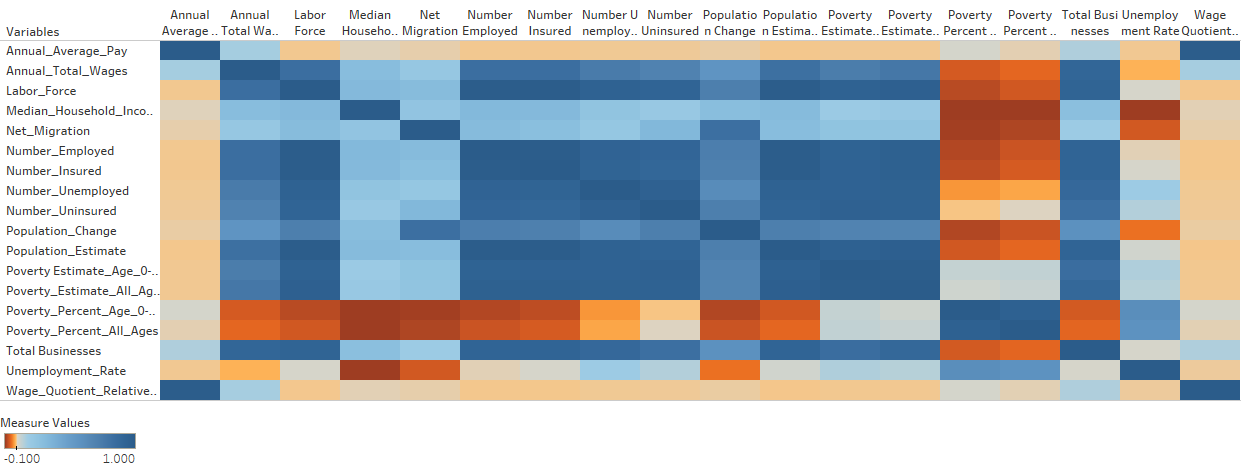
****

Figure 21: Correlation matrix between all independent variables

**Machine Learning Model**

**Artificial Neural Network - Using the count of weather events, their weather event type, and total damages per county to predict whether or not a county will experience a direct weather-related death or injury.**

Hypothesis: The count of weather events per type and the total damages for all weather events in any county can be used to predict whether or not a direct death or injury will occur in a county, with greater than 60% accuracy.

**Model**: Artificial Neural Network in SPSS Modeler

**Variables**: Count of Event ID for each Event Type, Sum of Injuries Direct, Sum of Deaths Direct, Sum of Crop Damage, Sum of Property Damage, Direct Casualty Occurred (1 = Yes, 0 = No) [Target Variable]

**Training vs Testing Split**: 70:30

A. Figures 22 and 23 are a confusion matrix and ROC plot for the machine learning model, respectively.

B. Figure 22 shows that the model had an accuracy of 75% in whether or not a death would occur in a county, based on the chosen variables. From Figure 22 and 23, we can see that the model had an area under curve of 0.832.

C. The model may be useful for counties and states to predict whether or not they may see direct death or injury from a weather-related event, based on upcoming weather forecasts. Such a model, if improved on a more granular level, could allow local governments to adequately prepare rescue or medical resources for an upcoming storm or other weather event.

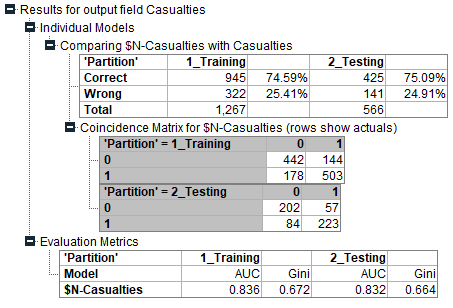


Figure 22: Confusion matrix for Artificial Neural Network model

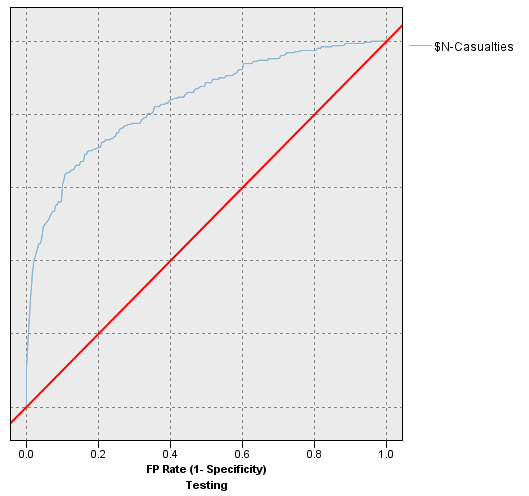


Figure 23: ROC Curve for Artificial Neural Network model