Visualisations of Tornado Data in the United States from 1950-2018

Abstract: Investigating how frequent and dangerous the tornadoes in the United States have been and their distribution throughout the country.

Student:

Nardiena Althafia Pratama

Student Number:

46223713

Course:

COSC3000 – Visualisation, Computer Graphics, and Data Analysis

Table of Contents

Introduction

Data

For this project, I have created visualisations of tornado occurrences in the United States of America from 1950 to 2018 inclusive. The data includes a variety of fields – the ones that are relevant to this project are listed below:

- Tornado number (om)
- Date and time of tornado occurrence (year, month, day, date, time)
- State of tornado occurrence
- F-scale or magnitude
- Injuries
- Fatalities
- Starting latitude and longitude of the tornado

The data was obtained from National Oceanic and Atmospheric Administration's (NOAA) National Weather Service Storm Prediction Center. The website's link is https://www.spc.noaa.gov/wcm/#data.

Project Aims

The aim of this project is to visualise the tornado occurrences in the U.S. using the dataset obtained above. The focus will be on:

- Mapping out the locations of each storm occurrence on the U.S. map
- Visualising the trends and patterns between casualty numbers and tornado frequencies over the years
- Visualising the trends and patterns between casualty numbers and tornado frequencies with respect to each state
- Highlighting trends on casualty numbers based on the magnitude

Project Purpose

The purpose of this project is to analyse the distribution patterns of the storms, and how frequent and dangerous the tornadoes have been.

The distribution patterns of the tornadoes can be helpful as it identifies where the hotspots of the storms are, as well as finding out how often a certain area could expect the storms to happen, if any.

It is also useful to know how frequent and dangerous these storms can be. This is because it helps create more awareness in terms of how people prepare for the storms. The data visualisations can help expose how unsafe each tornado magnitude is based on the casualty trends and patterns.

Methods

Dataset Source

The dataset used in this project can be found on the *Severe Weather Database Files* section of the Warning Coordination Meteorologist's (WCM) Page on the National Oceanic and Atmospheric Administration's (NOAA) National Weather Service Storm Prediction Center website, as shown in Figure 1.

The Storm Prediction Center (SPC) is defined to be "a government agency that is part of the National Centers for Environmental Prediction (NCEP), operating under the control of the National Weather Service (NWS), which in turn is part of the National Oceanic and Atmospheric Administration (NOAA) of the United States Department of Commerce (DoC)." [1]

The database files are separated by the category of the severe weather (tornado, hail, and damaging wind) and by year of occurrences, from 1950 to 2018, each on a separate CSV file. It is also possible to retrieve the CSV files for all occurrences of each severe weather event from 1950 to 2018 (combined), instead of downloading a CSV for each year. The database downloaded and used for this project is the file that has all the occurrences of tornadoes from 1950 to 2018 (the latter).

---

[1] "Storm Prediction Center." Wikipedia, Wikimedia Foundation, 27 Apr. 2020, en.wikipedia.org/wiki/Storm_Prediction_Center.

Severe Weather Database Files (1950-2017)

The tables below provide the links to comma separated value (.csv) files for tornadoes, hail, and damaging wind, as compiled from *NWS Storm Data*. Tornado reports exist back to 1950 while hail and damaging wind events date from 1955. The full hail and wind datasets are very large. To facilitate quicker downloads these data have been parsed by decade (1950s through 1990s), and half-decade, or less (from 2000 to the present). Tornado data are provided in raw csv file format below. Actual tornado tracks only (not including individual state segments) are provided in the "Actual_tornadoes.csv" file. Again, please read the format specification for the tornado data.

Please read this document carefully as it describes the format of the .csv files (especially important for tornadoes!). **Note 1**: The NWS changed criteria for severe hail from 0.75 inch minimum to 1.00 inch minimum in 2010. For legacy purposes, 0.75 inch hail reports will continue to be included in the annual hail csv files. Download the files and sort accordingly if needed. ***Note 2: See last two pages of the database specification document about the addition of the estimated F-scale column "fc", added to tornado.csv data in 2015.***

****** Changes made to the database in 2016 for the 1950-2015 U.S. tornado records *****

Estimated F-scale ratings were calculated for over 1800 tornadoes in the database previously rated with an F-scale entry of -9 (unknown). The scheme used to modify unknown F-scale to estimated F-scale is described in this document:

www.spc.noaa.gov/wcm/OneTor_F-scale-modifications.pdf

For all tornadoes modified in the database, a new information field was added to the database in 2016. The "fc" field (29th column of information in OneTor). "fc" is set to 1 for any tornado with a prior F-scale rating of -9. Thus, it is easy to filter previously unknown ratings, if desired. Please note! Tornadoes were likely underreported prior to 1953. Many tornado statistics are derived from 1953 (or later) to the present. Also, keep in mind that off-CONUS tornadoes are included in the latest version of the 1950-2015 tornado data. These events, while small in total count, include tornadoes in AK, HI, PR, and DC. It is best to filter results carefully when making queries on the tornado database.

It should also be noted that these data are used by the NWS for verification purposes and may not accurately reflect all storm events. Monetary loss information is highly suspect and should be used with caution, if at all. This article provides a good overview about the shortcomings in the NWS severe weather data provided here.

!!! IMPORTANT INFORMATION REGARDING THE SPC SEVERE WEATHER DATABASE !!!

- **Read the new change log for latest information/udpates.**

U.S. TORNADOES* (1950-2018) *Read format decscription document! Change Log Last Update: 30 September 2019
1950-2018_all_tornadoes.csv (6.9 mb) Raw database dump includes all state and continuing county segments.
1950-2018_actual_tornadoes.csv (6.7 mb) Single tracks. No state segments or continuing county info (e.g. sg="1").
2014-2015-onetor-dat.csv (2.9 mb) SPC Tornado Database with identifier to connect back to DAT data.

TORNADO	HAIL	DAMAGING WIND
2018_torn.csv (0.1 mb) **Updated: 30 Sep 2019**	2018_hail.csv (0.8 mb) **Updated: 30 Sep 2019**	2018_wind.csv (1.7 mb) **Updated: 30 Sep 2019**
2017_torn.csv (0.2 mb)	2017_hail.csv (1.2 mb)	2017_wind.csv (2.0 mb)
2016_torn.csv (0.1 mb)	2016_hail.csv (1.1 mb)	2016_wind.csv (1.9 mb)
2015_torn.csv (0.1 mb)	2015_hail.csv (0.9 mb)	2015_wind.csv (1.5 mb)
2014_torn.csv (0.1 mb)	2014_hail.csv (0.9 mb)	2014_wind.csv (2.8 mb)
2013_torn.csv (0.1 mb)	2013_hail.csv (1.0 mb)	2013_wind.csv (1.5 mb)
2012_torn.csv (0.1 mb)	2012_hail.csv (1.4 mb)	2012_wind.csv (1.7 mb)
2011_torn.csv (0.2 mb)	2011_hail.csv (2.0 mb)	2011_wind.csv (2.5 mb)
2010_torn.csv (0.1 mb)	2010_hail.csv (1.1 mb)	2010_wind.csv (1.6 mb)
2009_torn.csv (0.1 mb)	2009_hail.csv (1.4 mb)	2009_wind.csv (1.5 mb)
2008_torn.csv (0.1 mb)	2008_hail.csv (1.7 mb)	2008_wind.csv (1.7 mb)
2005-2007_torn.csv (0.3 mb) **Updated: 30 Sep 2019**	2005-2007_hail.csv (4 mb)	2005-2007_wind.csv (4 mb)
2000-2004_torn.csv (0.7 mb)	2000-2004_hail.csv (6 mb)	2000-2004_wind.csv (6 mb)
90-99_torn.csv (1.3 mb)	90-99_hail.csv (6 mb)	90-99_wind.csv (8 mb)
80-89_torn.csv (0.8 mb)	80-89_hail.csv (2.5 mb)	80-89_wind.csv (3.5 mb)
70-79_torn.csv (0.9 mb)	70-79_hail.csv (1 mb)	70-79_wind.csv (1.6 mb)
60-69_torn.csv (0.7 mb) **Updated: 30 Sep 2019**	60-69_hail.csv (0.67 mb)	60-69_wind.csv (0.90 mb)
50-59_torn.csv (0.6 mb)	55-59_hail.csv (0.20 mb)	55-59_wind.csv (0.28 mb)

All torn in one 1.5 mb zip file.
All hail in one 6.0 mb zip file.
All wind in one 7.5 mb zip file.

Map of State FIPS Numbers here.
Text list of County FIPS Numbers here.
Better list of County FIPS Numbers (as xls file) here.

National Weather Service • Since 1870

Figure 1: *The 1950-2018 archive of all tornadoes, hail storms, and damaging winds in the United States.*

Data Acquisition and Preparation

Excel Spreadsheet

The CSV file for all tornado occurrences in the United States from 1950 to 2018 (combined) is downloaded.

Once the data had been loaded into the spreadsheet, columns that weren't needed were removed, and some of the column titles had to be renamed so that they would be more comprehensible:

- *yr → year*
- *mo → month*
- *dy → day*
- *st → state*
- *mag → magnitude*
- *inj → injuries*
- *fat → fatalities*
- *slat → slatitude*
- *slon → slongitude*

The CSV file was then converted into an Excel Workbook, since saving it as a CSV may result in 'Possible Data Loss' in Microsoft Excel.

Tableau

The Excel Workbook is then imported as a Data Source in Tableau.

Data Visualisation: Tableau

Using the Tableau software, I visualized:

How Tornadoes Are Distributed Through the United States – Complex Behaviour Over Space

To visualize the tornado distribution, two distribution maps were made – one for revealing the density of tornadoes throughout the country and another to reveal the distribution based on magnitude.

For the density distribution map, the 'symbol maps' visualization type was chosen, using the starting longitude and latitude fields as columns and rows, respectively. The state and year fields were used as dimensions for the filters, which helps see the distribution based on those filters.

Similarly, for the magnitude distribution map, the starting longitude and latitude fields were used just like above, but two latitude fields are used on the rows option instead of just one – this creates a dual axis feature, which combines the magnitude distribution map, and the 'highlight magnitude' feature on top of it (using the 'Choose a Magnitude' feature on the right side of the map, the chosen magnitude will be highlighted in red on the map.

Number of Casualties and Tornado Frequency Over Time – Complex Behaviour Over Time and Multiple Dependent Variables

To visualise this, a new calculated field was created, which is the *Total Number of Casualties* field, where it is the sum of injuries and fatalities for a given row. The *fatalities*, *injuries*, *Total Number of Casualties*, and *Number of Records* fields (dependent variables) were placed on the row section, while the *year* field (independent variable) was placed on the column section. The *fatalities* and *injuries* fields were combined in one graph.

This created a visualisation that consists of three stacked line graphs, each dependent variable having a trend line. The top one for *fatalities* and *injuries* vs *year*, the middle one for *Total Number of Casualties* vs *year,* and the bottom one for *Number of Records (frequency of tornado)* vs *year.*

Number of Casualties and Tornado Frequency at Each State – Complex Behaviour Over Space

To visualise this, two state maps of the United States were made – one mapping out the number of tornadoes for each state, colour-coded so that the darker colour shows a higher number; and another mapping out the number of total casualties (*Total Number of Casualties* field) for each state, also colour-coded so that the darker colour shows a higher number.

Trends and/or Patterns on Casualties Based on Magnitude

To visualise this, the magnitude was set as the independent variable in the column section, whereas the *Total Number of Casualties, fatalities,* and *injuries* were set as the dependent variables in the row section. A scatter plot was then made based on these variables, with trend lines to help determine the trends in the graph.

Results

Using Tableau, six visualisations were produced to explore the main ideas. The interactive versions of these visualisations can be found here: https://tabsoft.co/35GcK6w

The dummy account credentials that can be used to view are:

Email: nardienapratama@yahoo.com

Password: dummyaccount1

Note: The maps will only show mainland United States, unless stated otherwise.

How Tornadoes Are Distributed Through the United States

A density distribution map of tornado occurrences in the United States is shown in figure 2. Using the map, it is possible to see areas in the country that have had higher densities of tornadoes (number of tornadoes per unit area) from 1950 up to 2018, or any time period in between.

The year filter is set so that the time period for the occurrences start in 1950 up to 2018. A pattern that is revealed is that the south-eastern area has the highest densities of tornadoes. The south-eastern area bordering the gulf (the Gulf Coast) and several areas towards the centre regions also show high densities, whereas the lowest tornado densities are mostly in the western area of mainland US.

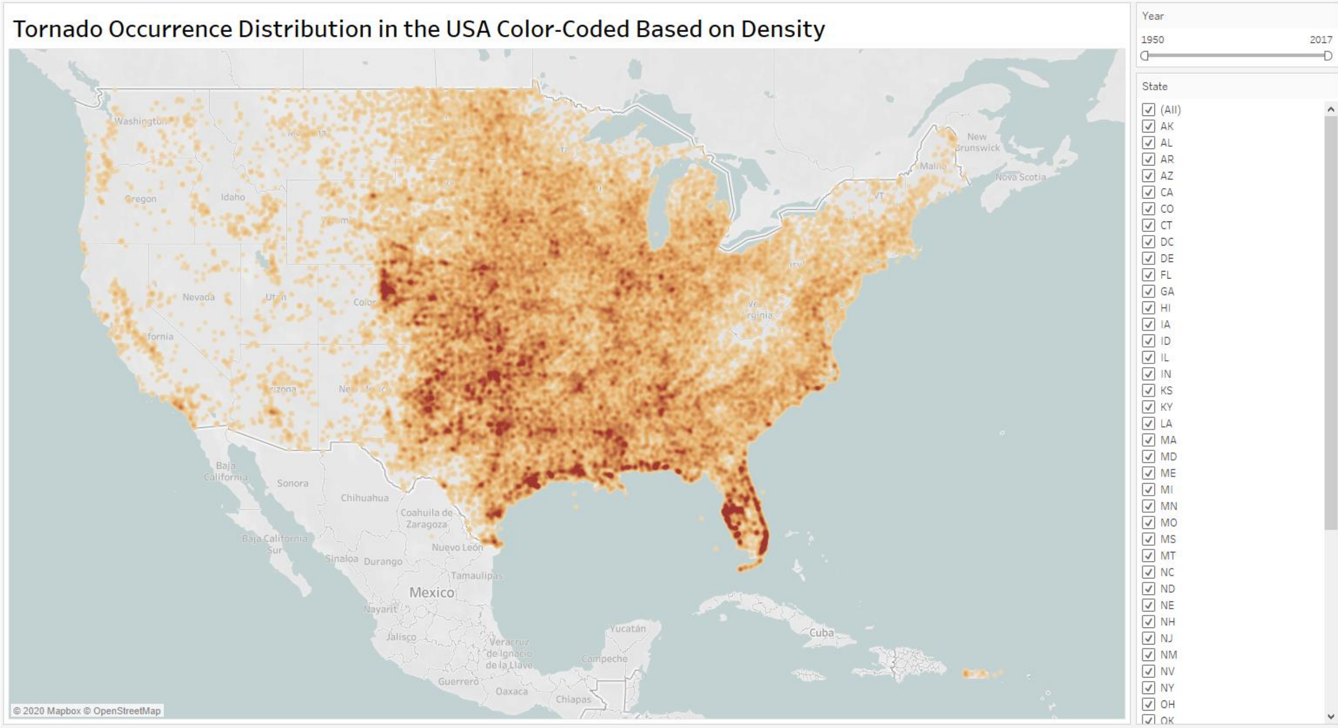


Figure 2: *A density distribution map showing the location of each tornado occurrence from 1950 to 2015 – darker areas signify higher densities of tornadoes.*

In figure 3 below, the same map is shown, but with a more transparent layer for the densities and a more prominent colour for the map background – so that the state areas are more identifiable. Although this map doesn't reveal the density patterns nearly as good as the previous one (due to it being more transparent), this one shows the state names to better identify where exactly the hotspots are.

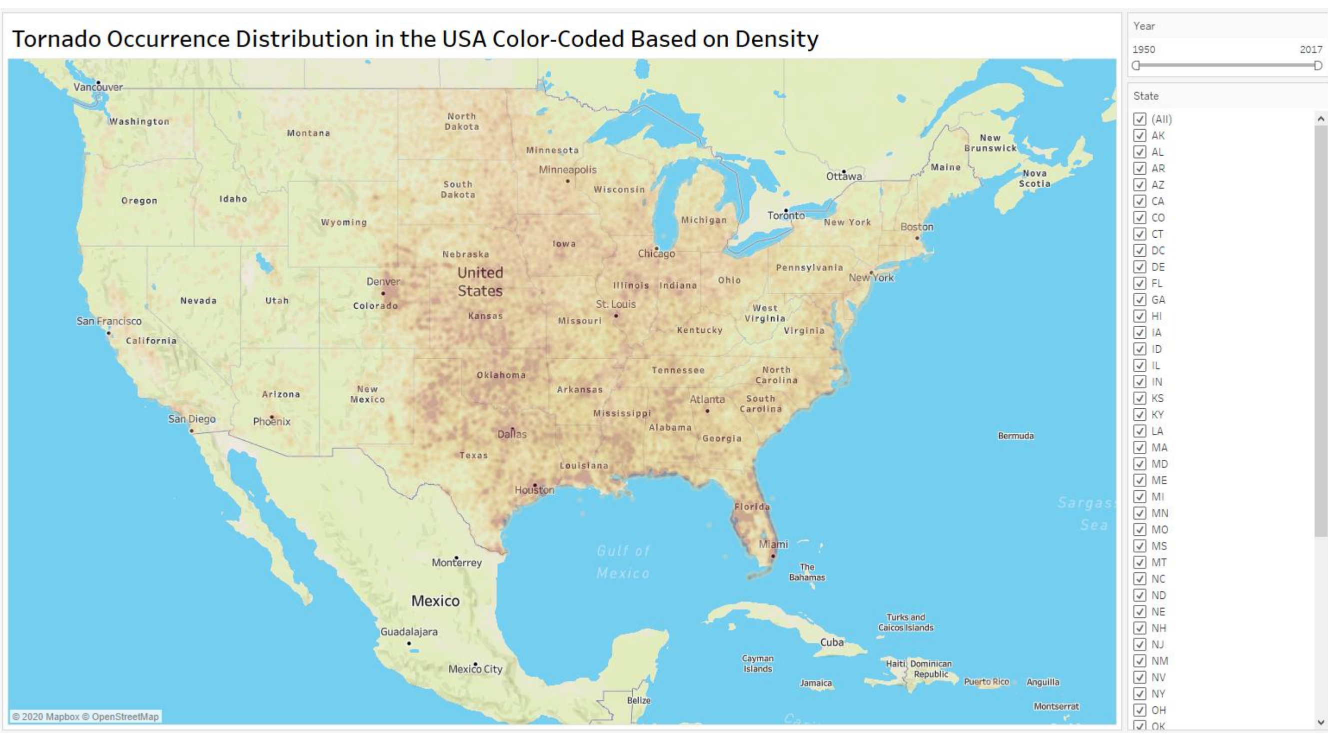


Figure 3*: A density distribution map showing the location of each tornado occurrence from 1950 to 2015 (the same as above), with slightly different configurations.*

As seen in figure 3, the hotspots are located mostly in the Florida state, followed by the southern area, which includes part of Texas and Louisiana, and some areas in the central states, such as Kansas, Oklahoma, and Colorado.

As shown from both figure 2 and 3, there seems to be very few to no occurrences of tornadoes in the western region of the United States, which include Oregon, Idaho, Washington, Nevada, California, among others.

Another aspect that can be explored when it comes to the tornadoes' distributions is how they are spread out in terms of magnitude, where 0 is the lowest, 5 is the highest, and -9 represents unknown magnitudes. This is shown in figure 4.

From the figure below, it is fairly obvious that the vast majority of the tornadoes have been of a magnitude of 0 to 1, with magnitudes 2 and 3 following behind. Most of the tornadoes with unknown magnitudes (yellow) are located towards the central-southern area, more specifically the Texas state area. However, based on this map alone, it is rather difficult to determine the distributions for some of the other magnitudes, such as magnitudes 4 and 5.

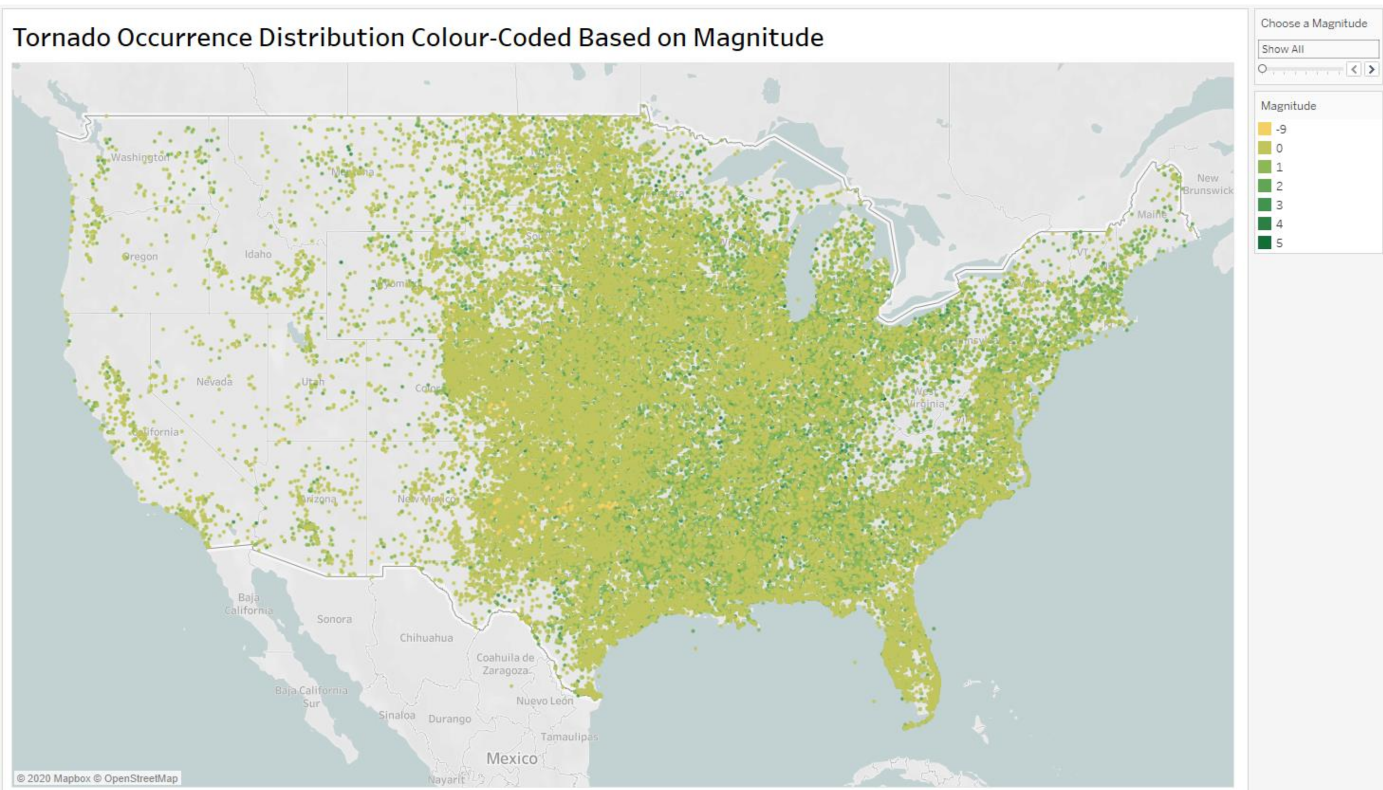


Figure 4: *A colour-coded distribution map based on magnitude for tornado occurrences from 1950 to 2018.*

In order to reveal the distribution patterns for magnitudes 4 and 5, the 'Choose a Magnitude' filter is set to magnitudes 4 and 5 in figures 5 and 6, respectively.

From figure 5 (below), there seems to be fewer Magnitude 4 tornadoes compared to the ones of Magnitude 0, and that they seem to be fairly spread out towards the eastern-central areas of the country.

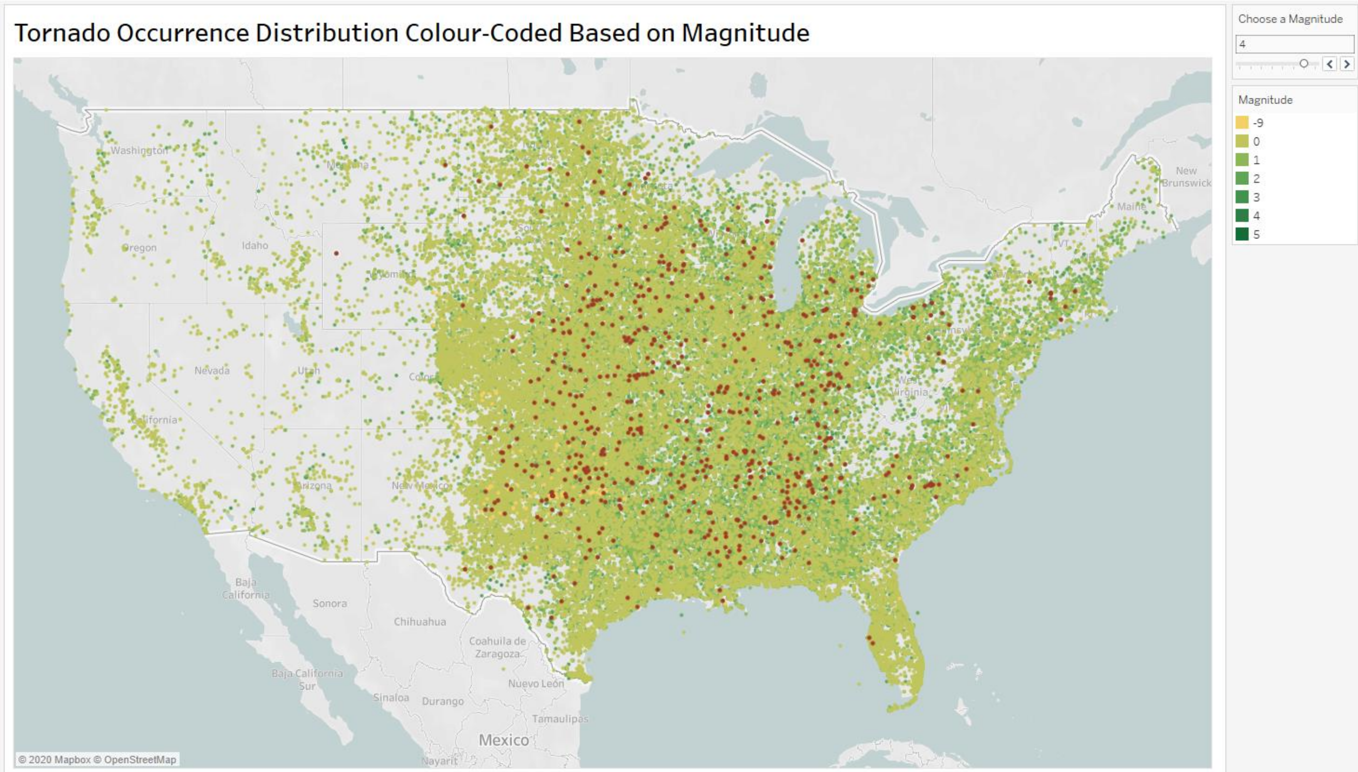


Figure 5: *A colour-coded distribution map based on magnitude for tornado occurrences from 1950 to 2018 (magnitude 4 is highlighted in brown).*

Figure 6 (below) shows the distribution for Magnitude 5 tornadoes. There seems to be no particular trend or pattern other than the storms are spread out, again, towards the eastern-central areas.

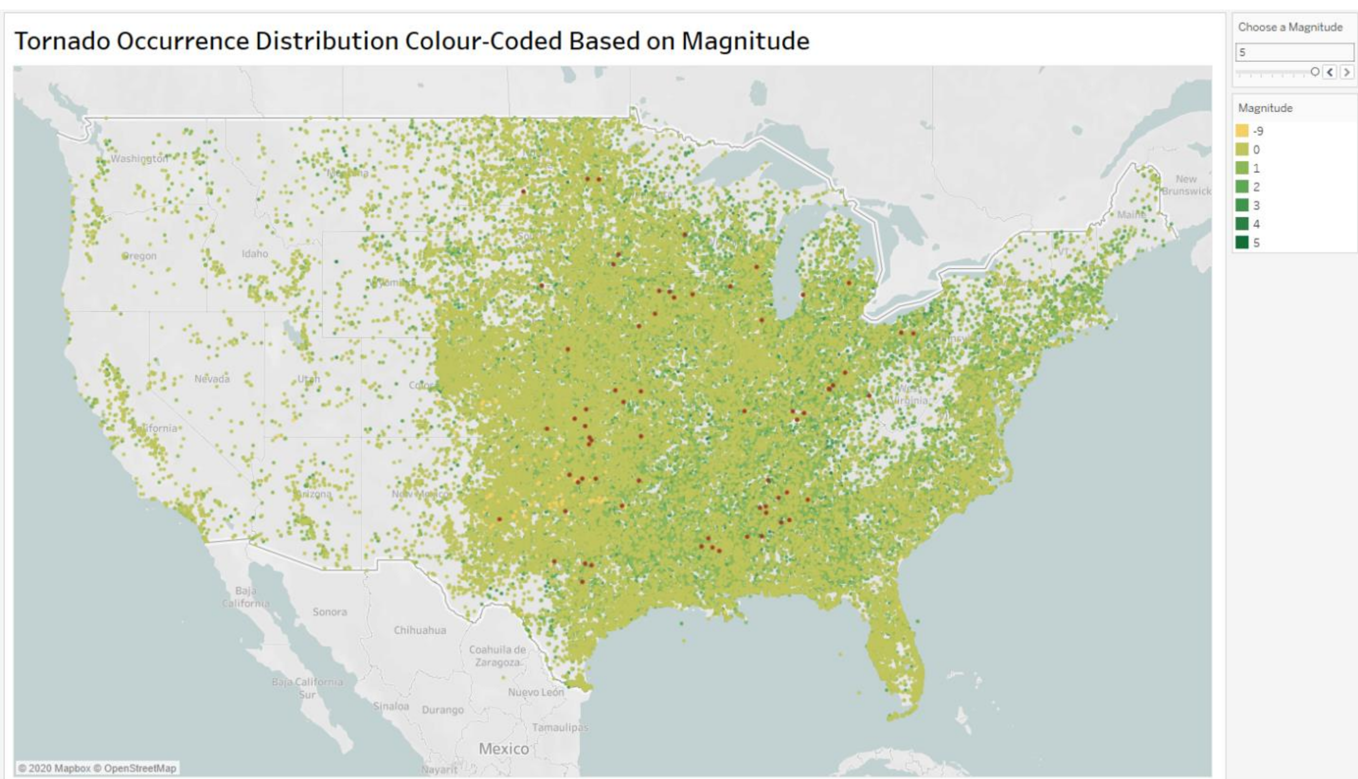


Figure 6: *A colour-coded distribution map based on magnitude for tornado occurrences from 1950 to 2018 (magnitude 5 is highlighted in brown).*

Based on figures 5 and 6, it can be seen that the presumably "more dangerous" tornadoes occur less frequently compared to the lower magnitude tornadoes, as shown in figure 4.

Number of Casualties and Tornado Frequency Over Time

In this section, patterns between casualties and tornado frequencies will be explored.

In figure 7 below, on the top line graph which shows the numbers of *fatalities* and *injuries* versus time, there doesn't seem to be a distinguishable relationship that can be deduced between the two variables and time. Although the numbers fluctuate from year to year, it shows several peaks in 1953, 1965, 1971, 1974, 1979, 1984, and 2011 – most of the peaks are until 1984; after that it only peaked again in 2011. The number of injuries (0-8000 range) seem to be consistently higher than the number of fatalities (0-600 range). This may be because it is more likely that a person was to get injured due to a tornado storm than to die, assuming that most people in tornado hotspots would likely have access to bunkers and other safety places.

The middle line graph shows the number of casualties (*fatalities + injuries)* versus time. Since the number of casualties is just a summation of the number of fatalities and injuries, it isn't surprising that there seems to be a similar trend to the previous graph in that there are several peaks until 1984, and another peak later in 2011.

The bottom line graph shows the number of tornado occurrences recorded every year from 1950 to 2018. Unlike the other two graphs, this one shows a noticeable relationship between the number of records and time. It seems that as the years increase, so does the number of tornado occurrences.

However, despite the increasing number of tornadoes over the years, the lack of positive correlation on the number of casualties seems odd. It should be expected that more tornadoes would result in more casualties, but it may not be the case here due to external factors – perhaps tornado occurrences back then weren't being recorded as accurately in certain places as they are now, hence the lower number of records back in the 1950s up to the late 1990s compared to the 2000s.

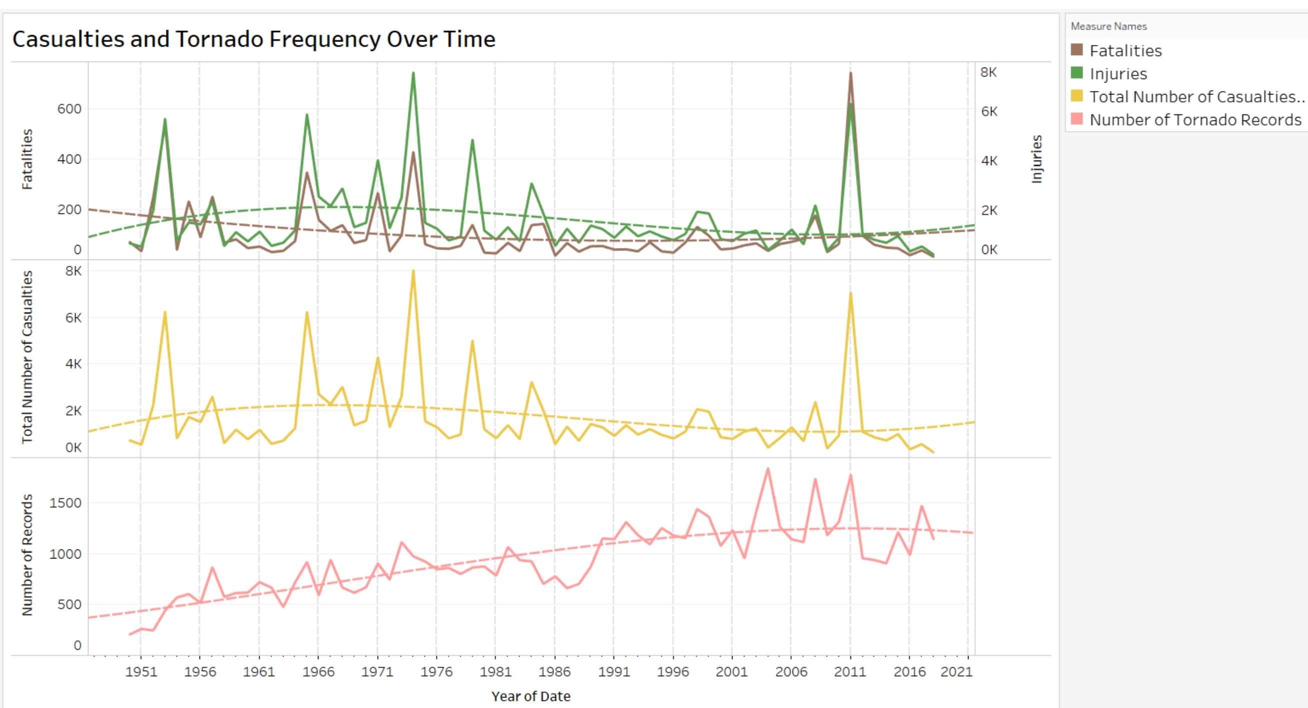


Figure 7: *Three stacked line graphs that show the correlations between number of fatalities, injuries, total casualties, and tornado occurrences versus time (in years).*

Number of Casualties and Tornado Frequency in Each State

Note: Alaska is included in the maps below as they contain the minimum values

This section will discuss patterns in the distributions of casualty numbers and tornado frequencies in each US state.

Figure 8 below is a map which shows each state's tornado frequency from 1950 to 2018 and it is colour-coded based on the frequency. As shown, the state with the highest tornado frequency is Texas, with 8,861 occurrences and the states with the lowest frequencies are Alaska, on the top left, and Washington, D.C., on the right side (*the labels aren't visible as the map is interactive and the user would need to hover over the state or zoom in closer to see the state name*).

The fact that Texas has the highest number of occurrences make sense considering that the tornado occurrence density in the several parts of that state was high (see figure 2). Although the tornado density was highest in Florida, the larger area of the Texas state should be considered, hence explaining the higher total number of storms compared to Florida. After Texas, several of the central and south-eastern states have the next highest number of occurrences

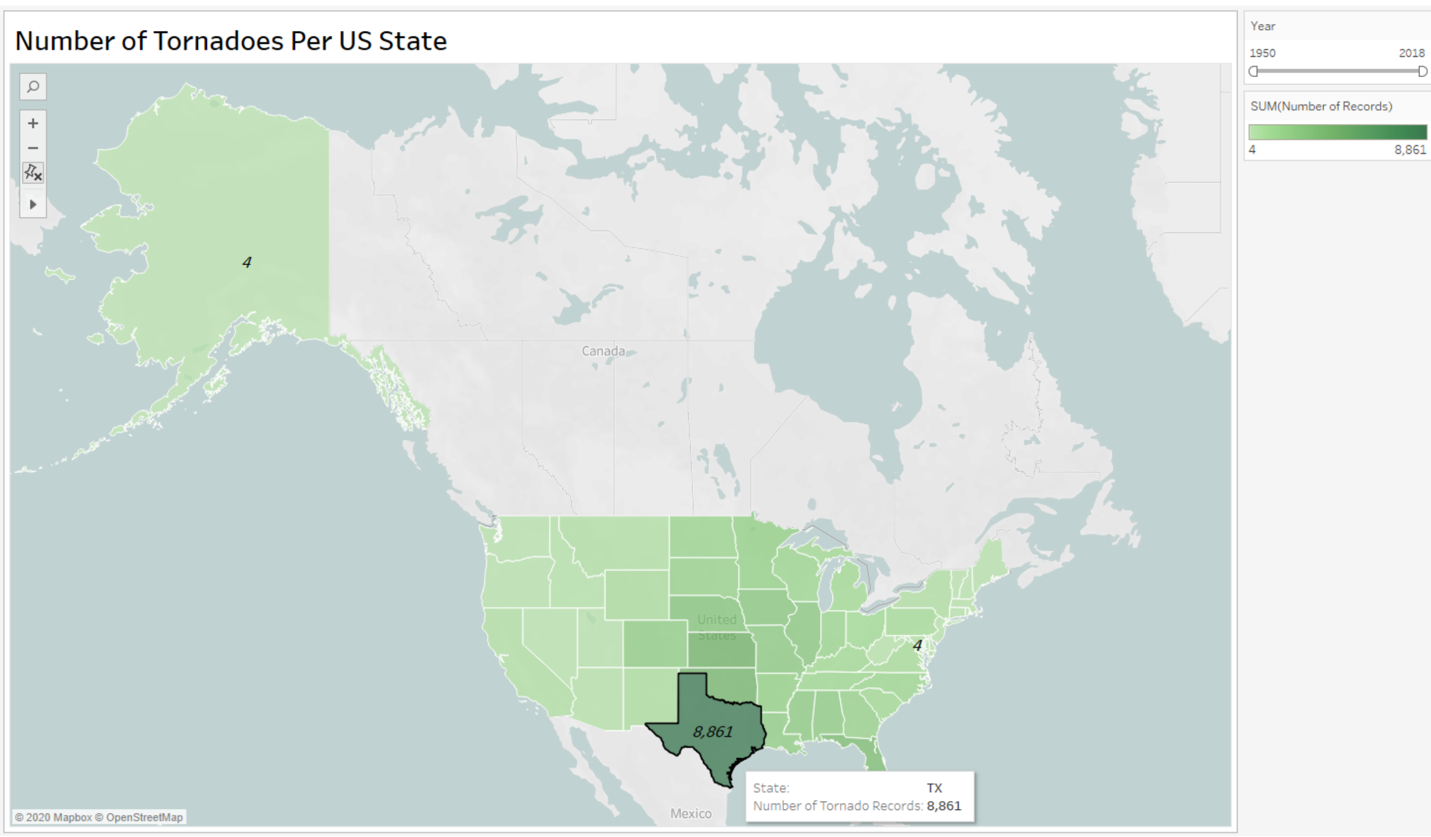


Figure 8: *A colour-coded U.S. state map representing number of tornado occurrences from 1950 to 2018 (based on the year filter on the right side).*

To compare and see whether the number of casualties and tornado occurrences per state have a correlation, figure 8 is compared to figure 9 below, which shows the number of tornado casualties for each state.

In figure 9, it can be seen that the highest number of tornado casualties is, again, in Texas, followed by some of its neighbouring states in the central and south-eastern regions, just like the trend in figure 8. The lowest number is in Alaska, as expected, as well. This makes sense since a higher number of storms would likely result in more injuries and fatalities and vice versa.

It may be possible that the trend in figures 8 and 9 here is unlike the trend for figure 7 (which shows an increase in tornado occurrences, yet no trend in casualty numbers) because perhaps the areas that may not have accurately detected their tornadoes back in the 1950s-1990s happen to be the ones that don't have the maximum or minimum casualty and tornado numbers here, and so it isn't as noticeable.

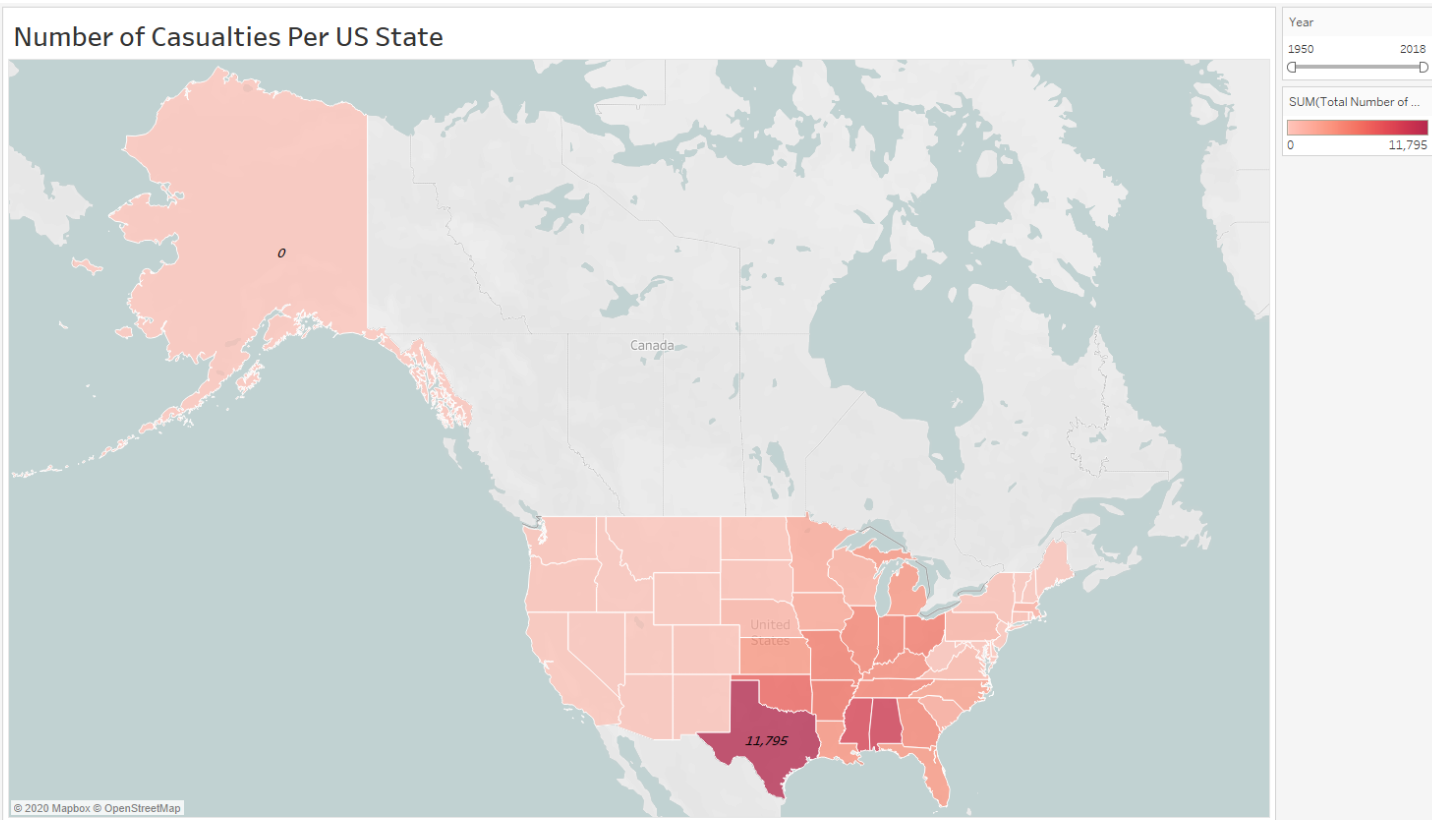


Figure 9: *A colour-coded U.S. state map representing number of casualties from 1950 to 2018 (based on the year filter on the right side).*

Trends and/or Patterns on Casualties Based on Magnitude

Note: Number of casualties for the unknown magnitude storms will be ignored for this section as it seems that there are no records not only for the storms' magnitudes but also their casualty numbers.

Figure 10 is a scatter plot which shows the correlation between the number of casualties (total number, fatalities, and injuries) versus the magnitude of the storm. As expected, the higher the magnitude of the tornado, the higher the number of casualties. However, this trend only applies until Magnitude 4, after which the number decreases for all three casualty categories (total number, fatalities, and injuries). Perhaps this may due to the fact that Magnitude 5 tornadoes, due to their higher "danger" level, usually result in better preparation and/or evacuation from the people in the predicted storm area, hence the lower numbers.

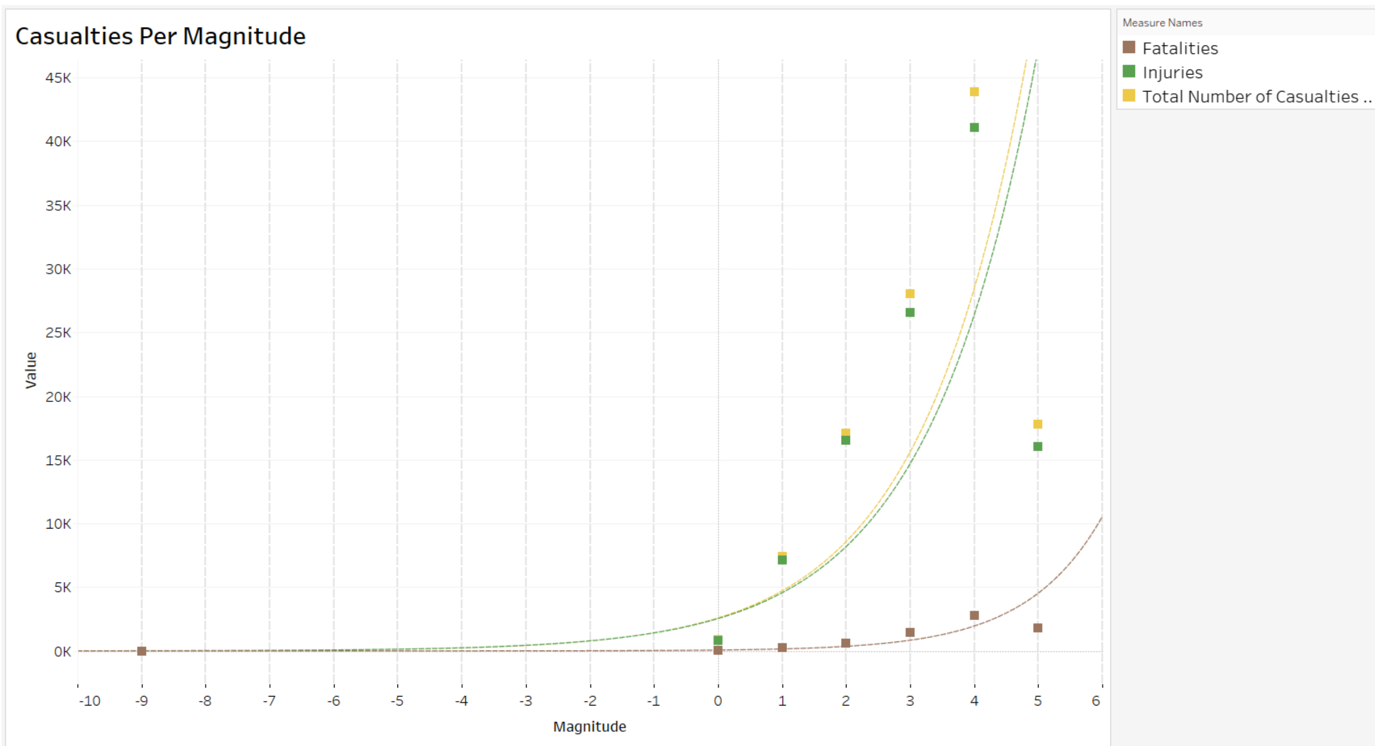


Figure 10: *A scatter plot of number of fatalities, injuries, and total casualties versus the magnitude of the tornado.*

Conclusion

Successes

All in all, this project has revealed the informative trends and patterns of the Tornado Occurrences data. Using the Tableau software, the data was successfully visualised in the form of distribution maps, state maps, line graphs and scatter plots. This helped identify trends in the tornado locations, correlations between casualties and tornado frequency, and between casualties and magnitudes. The visualisations also covered several of the methods and techniques discussed in the lectures, which were complex behaviour over space and time and including multiple dependent variables.

Findings

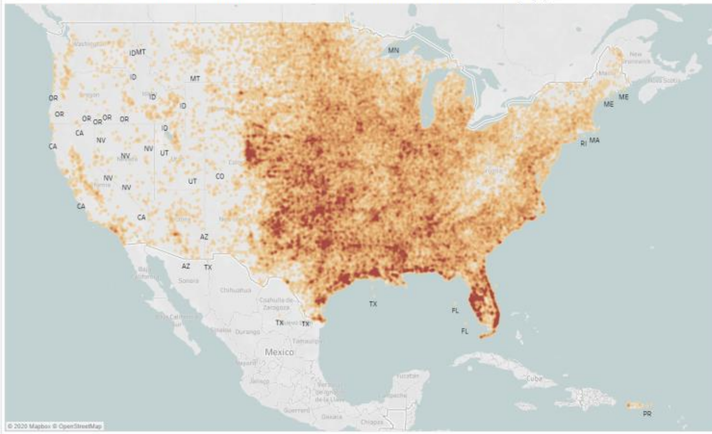
These are the patterns and trends found in the visualisations:

- Density of tornadoes from 1950 to 2018 is highest in the south-eastern, Gulf Area, and several central regions in mainland US
- Majority of tornadoes in mainland US were of the lower magnitudes 0 to 1
- Several peaks for total casualty numbers up to 1984; only peaked again in 2011
- Number of tornado occurrence records continues to increase from 1950 to 2018
- No correlation between the increasing tornado occurrences and number of total casualties – may be due to the aforementioned external factors
- States that have higher frequencies of tornadoes also have higher total numbers of casualties and vice versa
- Total casualty numbers increase as magnitude increases, but only up to magnitude 4

Areas for Improvement

One of the areas that could be improved is so that the state names in the density distribution map (figure 2) can be seen better – the state names were covered since the tornado location points were above some of the state names. When attempting to add labels, the states weren't centralised in each area like they are in the state maps, since the distribution maps aren't divided by states:

Tornado Occurrence Distribution in the USA Color-Coded Based on Density (2)



Moreover, there were several location points on the maps that were not plotted in the correct locations (some were out of bounds), perhaps due to wrong location points in the original dataset. To improve on this, the wrong location points should be removed to make the map more presentable. This should be possible to do on Tableau.

Since all the maps had a *year filter* feature to highlight the areas/location points for specific time periods, it could have visualised the maps better if it were in the form of a short animation – so the user wouldn't have to manually change the *years* to see each data set.

Another area for improvement would be to create visualisations that feature multiple independent variables, which could possible show trends of a dependent variable over two dimensions.

References

"Storm Prediction Center WCM Page." *Storm Prediction Center*, www.spc.noaa.gov/wcm/#data.

SPC Tornado, Hail, and Wind Database Format Specification (for .Csv Output). www.spc.noaa.gov/wcm/SPC_severe_database_description.pdf.

"Storm Prediction Center." *Wikipedia*, Wikimedia Foundation, 27 Apr. 2020, en.wikipedia.org/wiki/Storm_Prediction_Center.

"Gulf Coast of the United States." *Wikipedia*, Wikimedia Foundation, 2 May 2020, en.wikipedia.org/wiki/Gulf_Coast_of_the_United_States.