Elizabeth Hora; UEP-232: Final Project
12 December 2022

# Beyond Tornado Alley: The Threat Posed by Tornadoes in the Continental United States

## Introduction:

Severe weather events pose a risk to individuals across the United States and costs billions of dollars of damage each year to property. Out of all severe weather events, tornadoes have costed the United States nearly one billion dollars in damage to property including residential blocks, agriculture, and machinery per year since 1950 (Nouri et al., 2021). Tornadoes can develop with and without the presence of a supercell, a storm system characterized by severe thunderstorms and intense winds (Aguado and Burt, 2005). Characterizing the damages associated with and the intensity of tornadoes is challenging. The Fujita Scale was first developed based on a tornado outbreak from April 3$^{rd}$-4$^{th}$ in 1874 to estimate the wind speeds of tornadoes by developing groups F0-F5 that increase in severity (Schmidlin, 2016). However, scholars introduced a more nuanced Enhanced Fujita Scale in 2007 that uses 28 damage indicators and modified the bins from the original Fujita Scale to more accurately reflect the intensity of a tornado (Schmidlin, 2016). Understanding where tornadoes are more likely to occur as well as their intensity will help identify which areas are more at risk and where resources should be directed. The National Oceanic and Atmospheric Administration (NOAA) has published data on tornadoes from 1950 until 2021 ("Storm Prediction Center"). Techniques such as the Doppler Radar, storm chasing, and citizen journalism have all improved tracking tornadoes since 1950 as well as describing their impacts on communities in the wake of a tornado passing through. Tornadoes commonly occur in regions of the United States called 'Tornado Alley' (Nouri et al., 2021)(Moore and DeBoer, 2019). However, the area where tornadoes frequently occur could be altered due to climate change. As the population of the United States grows, there will be a higher density of people who are affected by tornadoes, meaning that this is a problem that cannot be ignored.

## Data and Data Sources:

The data downloaded for this project are sufficient to conduct my analysis in both ArcGIS Pro and R (shown in Table A1). The Tornado Paths data contain many columns documented in the codebook provided by NOAA ("Storm Prediction Center"). The State Fips Codes were from the U.S. Bureau of Labor Statistics and were used to help search states by name rather than by the state fips code ("Appendix D"). The County and State outlines are from the United States Census Bureau (data from 2018) and were needed to make the spatial images and calculations ("Cartographic Boundary Files"). These three components were necessary for the analysis done in ArcGIS Pro and R. However, the analysis in R had some additional calculations that would be challenging to do within ArcGIS Pro. The Inflation Consumer Price Index (CPI) values from the U.S. Bureau of Labor Statistics were used to adjust damages associated with tornadoes to inflation in terms of the dollar in January 2022 ("Databases, Tables & Calculators").

**Methods:**

This section will detail the different analyses carried out to assess how tornadoes have changed in the United States (shown in Table A1). Each part will explain which program was used to conduct the analysis. An inventory of packages needed in R is provided in Table A2.

Grouping by Decade: These spatial data (Tornado Paths) were challenging to analyze because they had a time component. The frequency of tornadoes has a periodic behavior (documented in Nouri et al., 2021), so comparing each year to each other will be too small of a timeframe to be able to appreciate general trends. I grouped the data by decade to better visualize trends over time. I did this by first creating a new column that changed the year to the decade (e.g., 1956 is changed to 1950). I then looped over the list of decades and saved a shapefile for each. This process can be done in ArcGIS Pro by calculating a new field and then selecting by attribute for each decade, but this is much more time consuming.

Normalizing the Tornado Counts: I made a map of the counts of tornadoes in the Continental United States at the county level normalized by the county area in $km^2$ to identify which areas of the United States are the most impacted by tornadoes. After projecting the data to an Albers Conic 1983 North America projection, I used the st_intersection() function in R to calculate the number of tornadoes that passed through each county. The normalizing by area is important because the sizes of counties in the United States are not the same (e.g., San Bernadino County, CA is 52,000 $km^2$ while Falls Church County, VA is 5.3 $km^2$). I found the total area of the county by adding the area of land and water ("Cartographic Boundary Files"). I looped over the list of decades and plotted the normalized number of tornadoes per county by $km^2$ using ggplot. In ArcGIS Pro, I made a plot for the 1950s and 2010s by using the corresponding decade shapefile I saved. For each decade, I first projected the shapefiles to Albers Conic 1983 North America. I then found the intersection between tornado paths and each county with a Spatial Join, creating a new column Join Count. I then created a new field and calculated the normalized tornado count by area per county by dividing the Join Count by the total area of the county (area of land and water) and then converted from square meters to square kilometers. I used graduated colors with natural breaks and imported my symbology from the 1950s to be used on the 2010s to maintain the same scale to better compare the two decades.

Calculating the Mean Center of Each Decade: Another way that I showed how tornadoes have changed over time is to plot the mean center of tornadoes weighted by the median width of the tornado. The goal of this was to replicate results found by a recent study by Cao et al., 2021 stating "over the latest 30-year warming period stronger tornado activities have made a statistically-significant geographic shift eastward, northeastward, and southeastward". I used the Mean Center tool within ArcGIS Pro for each decade shapefile, weighting by the width of the tornado, and included them on a single layout. Weighting by the width enables the center to be moved by larger tornadoes, rather than having small tornadoes count equally as large tornadoes (larger tornadoes are more of a threat to people and property than small tornadoes).

Assessing Damages: Assessing the cost with each tornado is important in understanding the risks tornadoes pose to the United States as a whole. There were several challenges in relating damages from the 1950s to the 2010s. The first issue was how NOAA encoded their estimated property damage (loss): from 1950-1995, loss was recorded in 8 bins with defined ranges; from 1996-2015, loss was recorded in millions of dollars; and from 2016-2021, loss was recorded in dollars. The other issue was that the loss was not adjusted for inflation. To convert the loss into

January 2022 dollars, I used the CPIs for each month from 1950-2022 on the U.S. Bureau of Labor Statistics website ("Databases, Tables & Calculators"). For the binned data in 1950-1995, I took the average of the lower and upper limit and then converted the dollars into my reference unit, using Equation 1. For the dollar amounts from 1996-2015, I first converted millions of dollars into dollars. I then adjusted the data from 1996-2021 using Equation 1. I then took these dollars adjusted for inflation across all years and classified them back into the same defined bins as the data were in 1950-1995.

$$(E1)\ Amount\ in\ January\ 2022\ Dollars = \frac{Cost\ of\ Tornado\ in\ Given\ Year}{CPI\ in\ Given\ Year} * CPI\ for\ January\ 2022$$

**Discussion:**

The distribution of tornadoes in the Continental United States has changed dramatically from the 1950s until the 2020s (Figures 1, A1). The Global Moran's I Spatial Autocorrelation for the 1950s was 64.83, and 56.05 for the 2010s. Both z-scores are very large, and each have an associated p-value of close to 0, which is statistically significant. Therefore, there is strong evidence to suggest that the distribution of tornadoes in the Continental United States is clustered. This result is consistent with the literature (Nouri et al., 2021)(Moore and DeBoer, 2019). As shown visually in Figure 1 and mathematically in Figure 2, there are more regions in the Continental United States that experience a tornado each year. Although not numerous, the Southwest (e.g., southern California, Nevada, Arizona, and New Mexico) has many counties that experienced no tornadoes in the 1950s, but then later experienced tornadoes in the 2010s. In between the 1950s and 2010s, there was no clear increase from decade to decade (Figure A1).

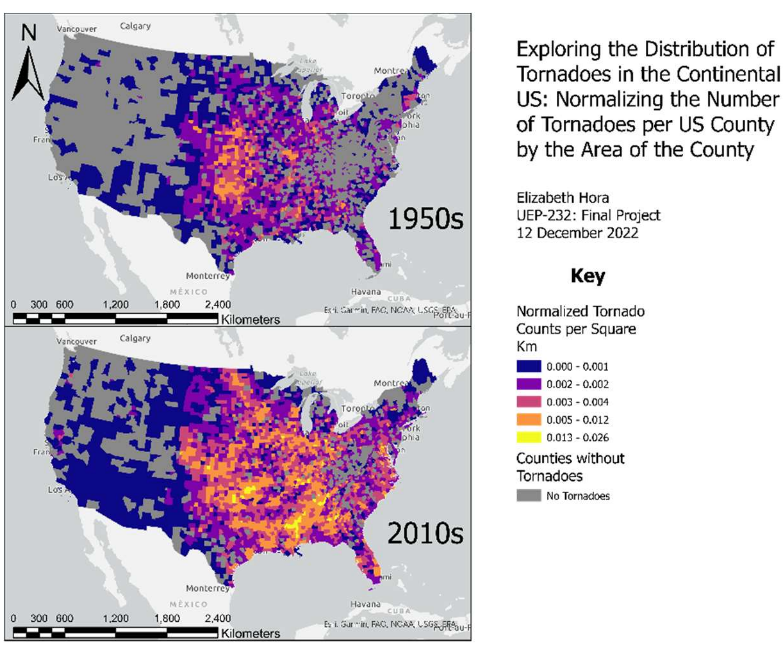


Figure 1: Each map corresponds to a different decade. The number of tornadoes per county were normalized by the area of the county (km$^2$). The same scale (natural breaks) was applied to both the normalized counts of tornadoes in the 1950s and in the 2010s, the most recent complete decade. Moreover, the number of counties that did not experience a single tornado (in grey) has decreased between the 1950s and the 2010s.

A concerning and emerging trend is illustrated in Figure 2, showing that there are more tornadoes occurring on the densely populated East Coast in the 2010s compared to the 1950s. Additionally, there are more urban areas in Illinois, Ohio, and Indiana where tornadoes are more likely to occur than in Kansas and Nebraska, where tornadoes used to occur more frequently (Figures 1, 2, A1). This raises the risk of fatalities or injuries as well as potentially increases the damages to property associated with each tornado. When examining the results of the Anselin Local Moran's Cluster and Outlier Analysis, there are more high-high clusters and low-high outliers in the 2010s compared to the 1950s (Figure 2). The fact that there are more high-high clusters is also concerning because that means that a high-risk county is not isolated, but rather there are more groups of counties that are high-risk.

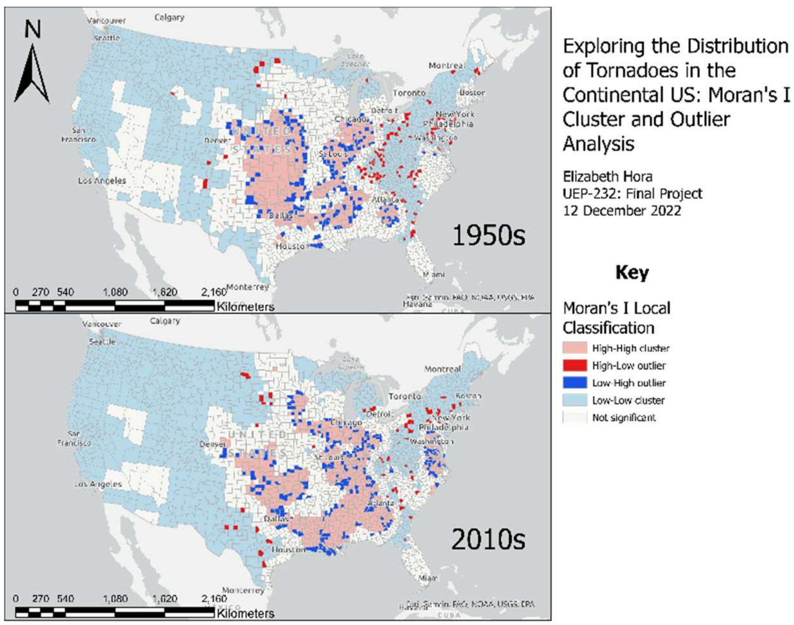


Figure 2: The Anselin Local Moran's Cluster and Outlier Analysis tests the attributes in the normalized tornado count per county field for localized outliers or clusters for all 3108 counties in the Continental United States. Compared to the 1950s, the 2010s has 140 more high-high clusters, 47 less high-low outliers, 53 more low-high outliers, 19 less low-low clusters, and 127 less not significant relationships. The Global Moran's I Spatial Autocorrelation (not shown) tests the attributes in the normalized tornado count per county field and establishes if their spatial distribution. Both the 1950s and 2010s have a z-score above 50 and a p-value that rounds to 0, meaning that there is statistically significant evidence to suggest that the distribution of tornadoes is clustered in each decade.

The mean centering technique applied in this case shows the general location tornadoes occur in the 1950s until 2021. In Figure 3, the mean center for each decade was weighted by the width of each tornado. Tornadoes that are wider will therefore contribute more to the mean center than smaller tornadoes. From the 1950s until the 2000s, there is natural variation in where the tornadoes occur, shown by slight changes in the position of each decade's mean center. However, from the 2010s onwards, the mean center starts to travel Eastward. This is generally consistent with the assertations made by [Cao et al., 2021](), where they split tornadoes into two groups (two 30-year periods of 1959–1988 and 1989–2018) and perform a t-test on the geographic centers, finding that tornadoes are now occurring more often in the Northeast. My results indicate that tornadoes are likely occurring more frequently Eastward, but not necessarily Northward.

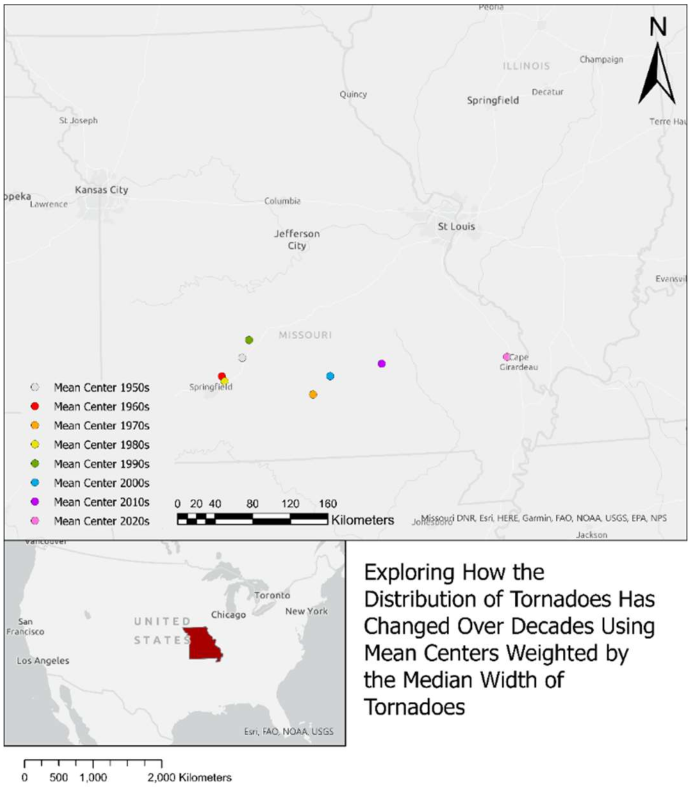


Figure 3: The mean centers of the tornado paths weighted by tornado width were taken for each decade and plotted on the same map. The mean centers from the 1950s until the 2000s appear to be relatively close to each other in Southern Missouri. However, from 2000 to 2010, the mean center starts to move Eastward, with the mean center for the 2020s (this decade is not yet complete, so this will be subject to change until 2030) moving the most Eastward.

Another component to tornadoes is the property damage caused (Figure 4). How often tornadoes strike a particular area and the typical scale of damage experienced are both important to insurance companies when picking rates to insure individuals. Since the damages recorded prior to 1996 are recorded differently than damages recorded in 1996 and later, these two periods are difficult to compare. In the years preceding 1996, there is a clear upward trend in the number and fraction of tornadoes that exceed $500,000 in damages when adjusted for inflation. In the years including and after 1996, there is a slight decrease in the fraction of tornadoes but about the same number of tornadoes that exceed $500,000 in damages when adjusted for inflation. One difference between the periods is the ability of storm detection. With the Doppler Radar, teams of meteorologists, and full-time storm chasers, there are more methods for predicting and detecting tornadoes. Therefore, in earlier time periods, there could be undercounting of minor tornadoes, which would result in a higher fraction of costlier storms.

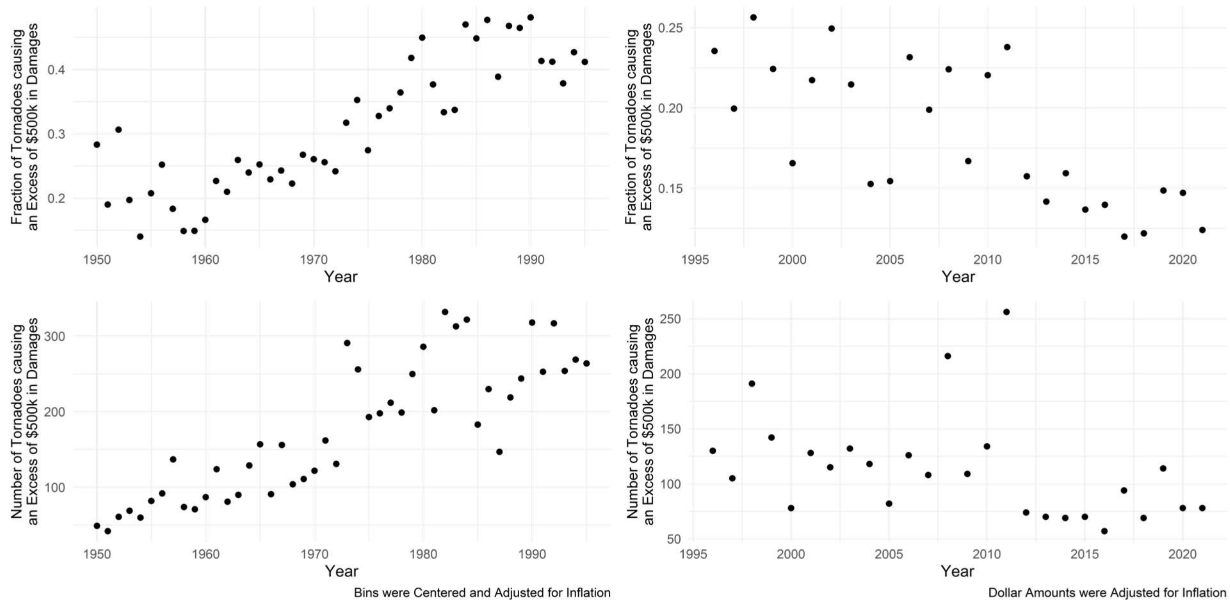


Figure 4: These plots show the fraction of tornadoes (top row) or number of tornadoes (bottom row) per year that cause an excess of $500,000. The period from 1950-1995 (left side) is separated from 1996-2021 (right side) due to how the damages were encoded and the math needed to adjust these values for inflation. When looking at 1950-1995, the fraction and number of costly tornadoes increases, while the fraction and number slightly decline for 1996-2021. It is challenging to compare tornadoes up to and before 1995 versus tornadoes from 1996 onwards.

**Conclusions:**

Tornadoes are going to be a threat posed to more people today than what they were when NOAA first started collecting data. Studying tornadoes is a dynamic field where the process of how tornadoes are formed with or without a present supercell storm is not entirely known (Aguado and Burt, 2005). There are many random phenomena in nature that contribute to tornado formation. Through the normalized tornado counts by county area (Figure 1), the Anselin Local Moran's Cluster and Outlier Analysis (Figure 2), and the mean center by decade (Figure 3), tornadoes exhibit a concerning trend of occurring more frequently on the densely populated East Coast in particular. The damages caused by tornadoes is also of great concern (Figure 4). Another way to analyze the threat of tornadoes is to compare the frequency of each type according to the Enhanced Fujita Score (Figures A2, A3). Stronger tornadoes are more likely to lead to fatalities. Roughly 6,000 people have been killed as a result of tornadoes from 1960 until 2021 ("Storm Prediction Center"). The effects of tornadoes will not be felt evenly: those who are economically disadvantaged and live in mobile homes make up the greatest percentage of fatalities by residence category despite being a minority of residents (Aguado and Burt, 2005). Climate change is predicted to bring more severe thunderstorms (Diffenbaugh et al., 2013). Such storms have the potential to form tornadoes. While climate change models will not exactly predict how often and how intense tornadoes will be, the threat to human life and property posed by tornadoes is more severe for more people than ever before.

**References:**

Aguado, E. and Burt, J.E., "Tornadoes." *Encyclopedia of World Climatology*. Dordrecht, Springer, 2005.

*Appendix D – USPS State Abbreviations and FIPS Codes*. U.S. Bureau of Labor Statistics, 2005, https://www.bls.gov/respondents/mwr/electronic-data-interchange/appendix-d-usps-state-abbreviations-and-fips-codes.htm. Accessed 9 September 2022.

Cao Z., et al., "Geographic Shift and Environment Change of U.S. Tornado Activities in a Warming Climate." *Atmosphere*, vol. 12, 5, 26 April 2021, https://doi.org/10.3390/atmos12050567.

*Cartographic Boundary Files – Shapefile*. United States Census Bureau, 2021, https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html. Accessed 9 September 2022.

*County Population Totals: 2010-2019*. United States Census Bureau, 2021, https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html. Accessed 29 November 2022.

*Databases, Tables & Calculators by Subject: CPI for All Urban Consumers (CPI-U)*. U.S. Bureau of Labor Statistics, 2022. https://data.bls.gov/timeseries/CUUR0000SA0. Accessed 26 November 2022.

Diffenbaugh, N.S., et al. "Robust increases in severe thunderstorm environments in response to greenhouse forcing." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, No. 41, 8 October 2013, doi:10.1073/pnas.1307758110.

Moore T.W., and DeBoer, T.A., "A review and analysis of possible changes to the climatology of tornadoes in the United States." *SAGE journals*, vol. 43, 3. 25 February 2019, https://doi.org/10.1177/0309133319829398.

Nouri N., et al., "Explaining the trends and variability in the United States tornado records using climate teleconnections and shifts in observational practices." *Scientific reports*, vol. 11, No. 1, 1741. 18 January 2021, doi:10.1038/s41598-021-81143-5.

Schmidlin, T.W., "Fujita Tornado Scale." *Encyclopedia of Natural Hazards.* Dordrecht, Springer, 2013.

*Storm Prediction Center*. NOAA, 2022, https://www.spc.noaa.gov/gis/svrgis/. Accessed 9 September 2022.

**Appendix:**

For the sake of transparency, my work in R and ArcGIS Pro can be accessed on GitHub.

| Data | Data Source | Data Type | Program Used |
|---|---|---|---|
| Tornado Paths | NOAA | Vector- Line | ArcGIS + R |
| State Fips Codes | U.S. Bureau of Labor Statistics | Non-Spatial Tabular | ArcGIS + R |
| County Outlines | U.S. Census Bureau | Vector- Polygon | ArcGIS + R |
| State Outlines | U.S. Census Bureau | Vector- Polygon | R |
| Inflation CPIs | U.S. Bureau of Labor Statistics | Non-Spatial Tabular | R |
| County Population | U.S. Census Bureau | Non-Spatial Tabular | R |

Table A1: This table is a summary of the data that I have downloaded for this project as well as which parts of the analysis the data were used for.

| R Package | Uses |
|---|---|
| gganimate | Animating a GIF- Counties, Animating a GIF- Fatalities |
| lubridate | Animating a GIF- Fatalities |
| patchwork | Assessing Monetary Damages |
| png | Normalizing the Tornado Counts, Assessing Monetary Damages |
| readxl | Grouping by Decade |
| sf | Grouping by Decade, Normalizing the Tornado Counts, Animating a GIF |
| tidyverse | Grouping by Decade, Normalizing the Tornado Counts, Animating a GIF, Assessing Monetary Damages |
| units | Normalizing the Tornado Counts |

Table A2: This table summarizes which R packages were needed for each part of the analysis (Uses column). There are some Uses that were not included in the methods section, but they are included in the Additional Uses section at the end of the Appendix. Tidyverse is an alternative to Base-R and is therefore used for every part of the analysis in R. Other packages are needed for certain parts of the analysis.

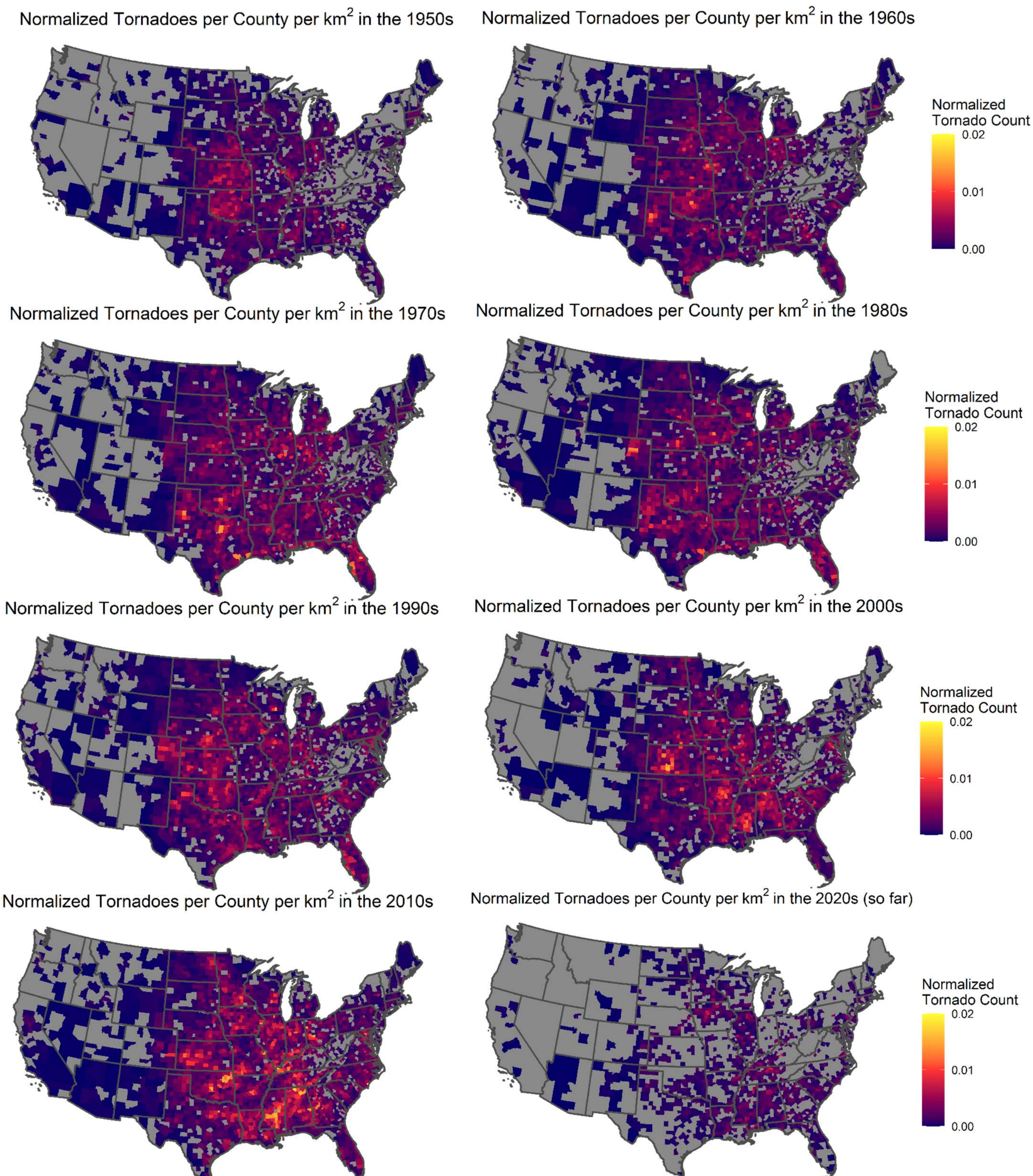


Figure A1: Each map corresponds to a different decade. The number of tornadoes per county were normalized by the area of the county ($km^2$). The same defined interval was applied to both the normalized counts of tornadoes in the 1950s until the 2020s, as is the case in Figure 1. The 2020s decade is incomplete and will change as more data are collected and reported by NOAA.

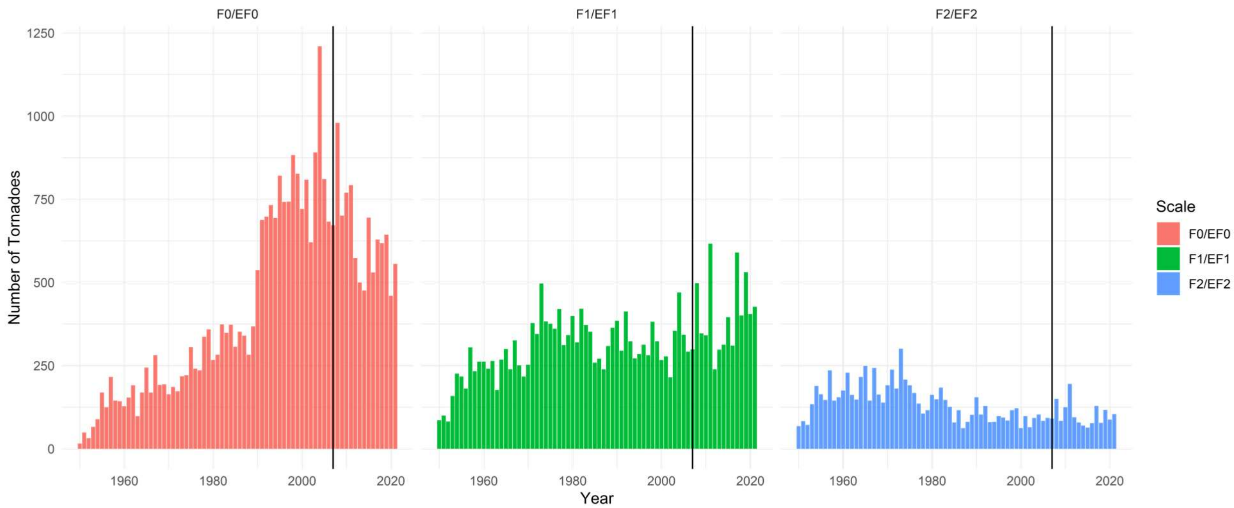


Figure A2: This plot shows the sum of tornadoes by rank, with tornadoes before 2007 classified using the Fujita Scale and tornadoes 2007 and onwards classified using the Enhanced Fujita Scale, separated by a vertical black line. There was a dramatic increase in F0 tornadoes in the 1990s, while the number of F1 and F2 tornadoes did not increase in the same way.

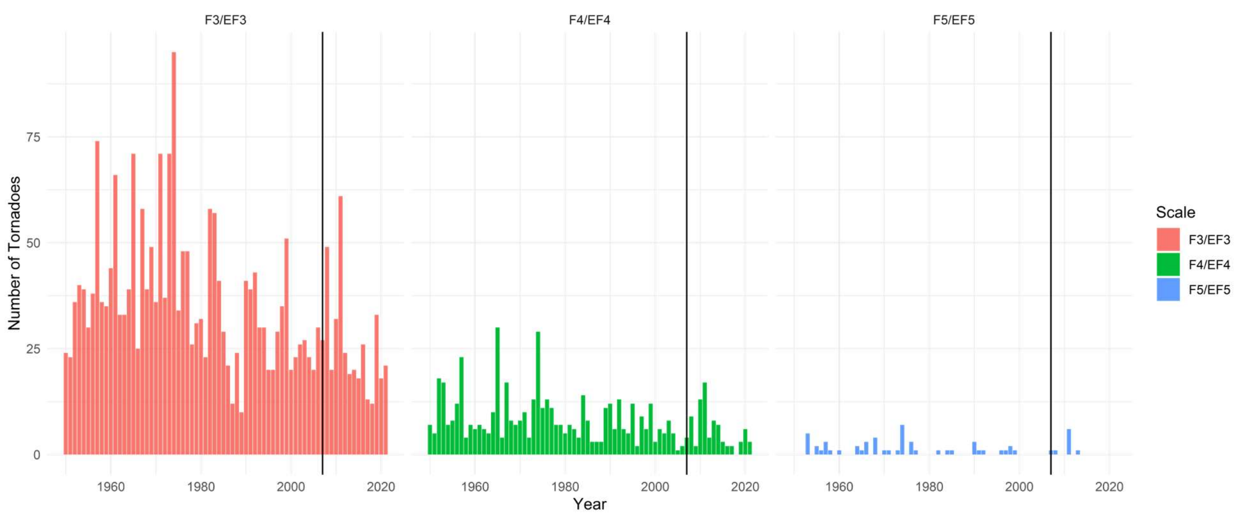


Figure A3: This plot shows the sum of tornadoes by rank, with tornadoes before 2007 classified using the Fujita Scale and tornadoes 2007 and onwards classified using the Enhanced Fujita Scale, separated by a vertical black line. Very powerful tornadoes (F5/EF5) are very rare in comparison to smaller, less powerful tornadoes. There does not seem to be an obvious increase in very powerful tornadoes as of yet.

**Additional Uses:**

Animating a GIF- Counties: I made a GIF in R that animated the paths of tornadoes per year scaled by median width of each tornado. The red lines are the tornadoes for a given year, with the grey lines being paths of all tornadoes- this gives the viewer a sense of where tornadoes rarely occur without watching the whole GIF to see which areas are denser than others. Each year was its own frame, and the GIF has an animation speed of 2 frames per second to allow time for the viewer to process each year. The goal of this GIF was to illustrate where tornadoes typically occur and that the number of tornadoes is not constant from year to year. The GIF is attached to this assignment (gif_1_counties.gif).

Animating a GIF- Fatalities: Another angle to analyze the impact of tornadoes is through fatalities, animated in R. These data are in a non-spatial tabular format (although the states represent a spatial concept). I chose a bar chart to best represent the number of fatalities per state, not normalized by population, for the 10 states with the greatest number of tornado fatalities. Each year was its own frame, and the GIF has an animation speed of 4 frames per second to allow the viewer to compare each state with the GIF moving at a reasonable pace. While I did not have the time nor the space to analyze fatalities, the makeup of the fatalities should be studied. It is surprising that a state like Alabama has the greatest number of tornado fatalities despite being the 24th most populated state in the Continental United States ("County Population Totals"). The GIF is attached to this assignment (gif_2_fatalities.gif).