MTH 463 - Math Modeling

Final Report: U.S. Tornado Analysis

I certify that this work is original and not a product of anyone's work but my own.

Adam Cunningham

Submitted: April 30th, 2020

Due by: April 30th, 2020

Contents

List of Figures

Abstract

Tornadoes are a common phenomenon within the United States, with more than a thousand occurring annually. The data set described in this report is a list of around 64,000 tornadoes that touched down in the U.S. between the years 1950 and 2018. Each tornado has specific details such as the date it happened, the states it traveled through, etc. To answer the question formulated during this analysis, this report will look deeper into the details of a tornado's direction in which it travels, and the amount of destruction a tornado leaves.

1 Introduction

The National Oceanic and Atmospheric Administration (NOAA), has a data set of historical tornadoes in the United States from 1950 to 2018. This data set contains 64,825 tornado entries. These entries each have 29 columns of information associated with them. There is documentation that covers the details of each of these attributes, this documentation is shown in the appendix of this report. For the research and analysis done in this report, only the columns detailing - the state, the starting latitude and longitude, the ending latitude and longitude, the tornado magnitude, the amount of injuries, the amount of fatalities, and the amount of property loss - will be used.

2 Methods and Procedures

2.1 Preliminary Questions and Analysis

The first thing to do with this data set was devise a plan for analysis. The following preliminary questions were asked before looking into this data:

1. What amount of damage does a tornado cause on average?
2. What states are mainly targeted by tornadoes?
3. How long does a tornado last on average?

Through these questions a flowchart was then created and to be followed during analysis and testing. Figure 1 shows the original plan. As shown, the next step is to import the data set into Mathematica for organization.

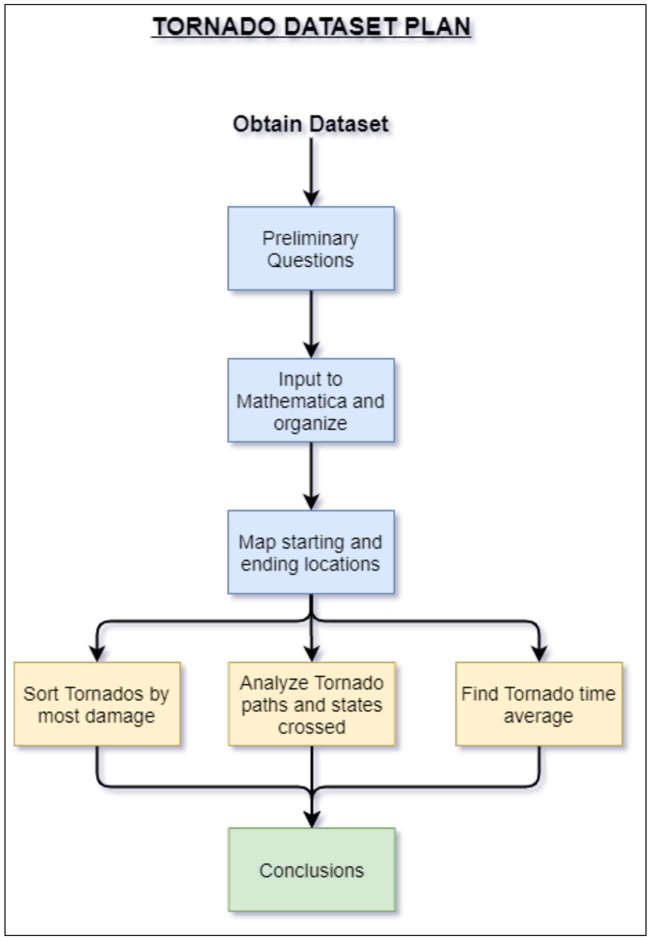


Figure 1: Preliminary Flowchart Plan

Upon going further and further into the analysis of the data set, the questions ended up changing to reflect what could actually be extracted from the data set itself, as well as, new questions that arose from visualizations. The new list of questions were:

1. Does tornado damage increase with tornado magnitude?
2. What states are mainly targeted by tornadoes?
3. Do tornadoes travel in a Northeast direction?
4. How far does a tornado travel on average?

To first move the data into Mathematica for use the Import command was used to import a data set with column headers. The resulting set is detailed below in Figure 2.

```
In[27]:= data = Import["C:\\MTH 463\\Dataset\\1950-2018alltornadoes.csv", "Dataset", HeaderLines → 1]
```

Out[27]=

om	yr	mo	dy	date	time	tz	st	stf	stn
1	1950	1	3	1950-01-03	11:00:00	3	MO	29	1
1	1950	1	3	1950-01-03	11:00:00	3	MO	29	1
1	1950	1	3	1950-01-03	11:10:00	3	IL	17	1
2	1950	1	3	1950-01-03	11:55:00	3	IL	17	2
3	1950	1	3	1950-01-03	16:00:00	3	OH	39	1
4	1950	1	13	1950-01-13	05:25:00	3	AR	5	1
5	1950	1	25	1950-01-25	19:30:00	3	MO	29	2
6	1950	1	25	1950-01-25	21:00:00	3	IL	17	3
7	1950	1	26	1950-01-26	18:00:00	3	TX	48	1
8	1950	2	11	1950-02-11	13:10:00	3	TX	48	2
9	1950	2	11	1950-02-11	13:50:00	3	TX	48	3
10	1950	2	11	1950-02-11	21:00:00	3	TX	48	4
11	1950	2	11	1950-02-11	23:55:00	3	TX	48	5
12	1950	2	12	1950-02-12	00:30:00	3	TX	48	6
13	1950	2	12	1950-02-12	01:15:00	3	TX	48	7
14	1950	2	12	1950-02-12	06:10:00	3	TX	48	8
15	1950	2	12	1950-02-12	11:57:00	3	TX	48	9
16	1950	2	12	1950-02-12	12:00:00	3	MS	28	1
17	1950	2	12	1950-02-12	12:00:00	3	MS	28	2
18	1950	2	12	1950-02-12	12:00:00	3	TX	48	10

rows 1–20 of 64825 columns 1–10 of 29

Unable to store data in notebook.

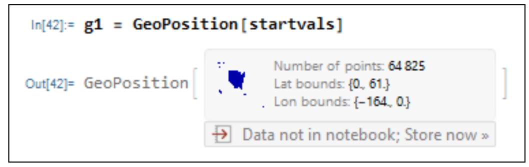
Figure 2: Initial Data Set Import

As stated before, the columns of particular interest are the ones regarding damage and geospatial positioning. The first portion of the data to be analyzed is the starting ending spatial data points. This is to answer one of the original questions: What states are mainly targeted by tornadoes? To do this a geo histogram was used with the map of the United States as the target.

First the data was extracted using the "slon" and "slat" headers of the original data set. These starting values were then turned into geo spatial data points using the GeoPosition function of Mathematica.

```
start = data[All, {"slat", "slon"}];
startvals = start[All, Values] // Normal
```

(a) Starting Lat and Lon Values



(b) GeoPosition Conversion

Figure 3: Starting Latitude and Longitude Values Extraction

These sets of values were then used to create a geo histogram of the united states. This geo histogram gave a much better visual to understand which states were effected the most by tornadoes. Below Figure 4 shows the histogram and tornado density. Here it can be clearly seen that states such as Oklahoma and Louisiana have the majority of tornadoes from this data set.

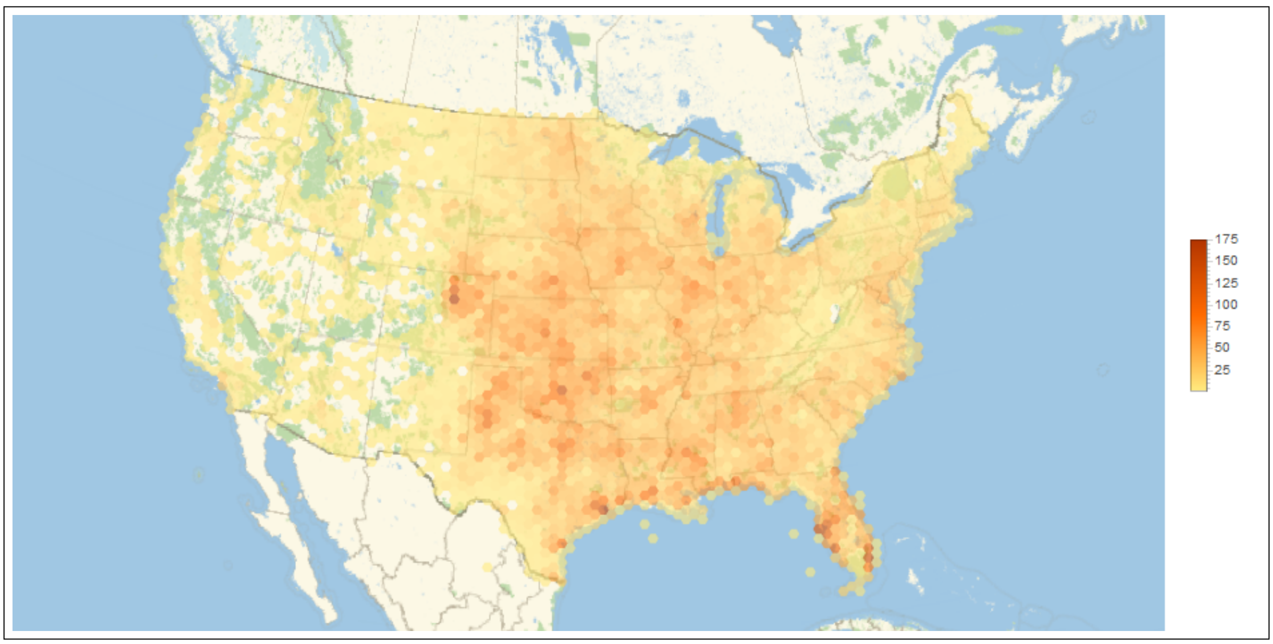


Figure 4: U.S. Tornado Geo Histogram

Using Tally and Sort functions in Mathematica, a list of the states with the most tornadoes can be found. From this list the top three states are Texas with 8,861 tornadoes, Kansas with 4,302 tornadoes, and Oklahoma with 3,921 tornadoes. For the further analysis of tornadoes in general, these three states were used.

```
ReverseSortBy[Tally[states], Last]
```

TX	8861
KS	4302
OK	3921

Figure 5: States With the Most Tornadoes

2.2 Texas

2.2.1 Data Extraction

Analysis will start with the state that contains the most tornadoes in the data set, Texas. The analysis will start with the direction and location of each tornado itself. This can be found since the data set contains the starting and ending latitude and longitudes of most tornadoes. A handful of tornadoes do not have an included ending spatial data point, these tornadoes will be ignore during analysis. Below, Figures 6a and 6b, shows the tornadoes related to Texas being extracted from the data set and stored in a variable texas. The txslat, txslon, txelat, and txelon variables are created to move each of the four points associated with each tornado out of a data set environment and into a vector environment. This will help for future calculations done on each of the four points. This same technique is also used when analyzing tornadoes in other states as will be shown.

```
Texas = data[Select[#st == "TX" && #elat ≠ 0 &], {"slat", "slon", "elat", "elon"}]
```

slat	slon	elat	elon
26.88	-98.12	26.88	-98.05
29.42	-95.25	29.52	-95.13
29.67	-95.05	29.83	-95.0
32.35	-95.2	32.42	-95.2
32.98	-94.63	33.0	-94.7
33.33	-94.42	33.45	-94.42
32.08	-98.35	32.1	-98.33
31.52	-96.55	31.57	-96.55
31.8	-94.2	31.88	-94.12
31.8	-94.2	31.8	-94.18
32.42	-99.5	32.42	-99.48
31.9	-98.6	31.73	-98.6
36.4	-100.8	36.42	-100.77
29.78	-98.83	29.67	-98.57
32.67	-101.88	32.65	-101.85
33.35	-96.92	33.3	-96.92
30.18	-96.4	30.02	-96.05
29.9	-96.4	29.72	-96.07
36.08	-102.53	36.08	-102.25
33.93	-98.68	33.93	-98.77

rows 1–20 of 4296

(a) All Tornadoes in Texas

```
txslat = Texas[All, 1] // Normal;
txslon = Texas[All, 2] // Normal;
txelat = Texas[All, 3] // Normal;
txelon = Texas[All, 4] // Normal;
```

(b) Specific Texas Latitudes and Longitudes

Figure 6: Texas Tornado Geo Spatial Path Data

These points where then plotted onto a map of Texas using the GeoPath function in Mathematica. The command to accomplish this is shown in Figure 7. The resulting output of this command is shown in Figure 40 in the results section of this report.

```
GeoGraphics[Table[{Red, GeoPath[{{txslat[[i]], txslon[[i]]}, {txelat[[i]], txelon[[i]]}}]},
    {i, Length[Texas]}], GeoRange → Texas, United States ADMINISTRATIVE DIVISION , Frame → True]
```

Figure 7: Command to Plot GeoPath of Each Tornado

2.2.2 Clustering

Looking at the results of the GeoPath plot of Texas' tornadoes, there appears to be a directional pattern. This pattern seems to be that each tornado travels in a Northeastern direction upon its start. To find the conclusion to this theory a use of clustering is necessary. Each tornado will be put into one of three different clusters, all of which relate to the slope of a tornado's path. A tornado can either have no slope, a positive slope, or a negative slope. To calculate the three clusters, a slope function was devised, below is that function.

```
slope[x1_, y1_, x2_, y2_] := If[x2 - x1 ≠ 0, (y2 - y1) / (x2 - x1), "undefined"]
```

Figure 8: Defined Slope Function

This function takes in two x-y points and calculates the slope of those two points. Since the slope uses division to function, it was necessary to implement an if statement to make sure a divide by zero error didn't occur. For any tornado path that doesn't have a slope, "undefined", or 0, will be output. From this slope function, three variables named txc1, txc2, and txc3 are created. Since these clusters are taken directly from the Texas data set variable, the actual vector values need to be extracted.

```
txc1 = Texas[Select[slope[#slat, #slon, #elat, #elon] === "undefined" || slope[#slat, #slon, #elat, #elon] === 0 &], Values] // Normal;
txc2 = Texas[Select[slope[#slat, #slon, #elat, #elon] > 0 &], Values] // Normal;
txc3 = Texas[Select[slope[#slat, #slon, #elat, #elon] < 0 &], Values] // Normal;
```

Figure 9: Texas Tornado Clusters

Using the same GeoPath method from earlier, each tornado path can be plotted on the Texas map. This time the positive slope, negative slope, and 0 slope tornadoes can be colored differently. For this interpretation, positive slope is blue, negative slope is red, and no slope is black. Figure 41 shows the result of this plot.

```
Show[GeoGraphics[Table[{Black, GeoPath[{{txc1[[i, 1]], txc1[[i, 2]]}, {txc1[[i, 3]], txc1[[i, 4]]}}]}, {i, Length[txc1]}], GeoRange → Texas, United States ADMINISTRATIVE DIVISION, Frame → True],
 GeoGraphics[Table[{Blue, GeoPath[{{txc2[[i, 1]], txc2[[i, 2]]}, {txc2[[i, 3]], txc2[[i, 4]]}}]}, {i, Length[txc2]}], GeoRange → Texas, United States ADMINISTRATIVE DIVISION, Frame → True],
 GeoGraphics[Table[{Red, GeoPath[{{txc3[[i, 1]], txc3[[i, 2]]}, {txc3[[i, 3]], txc3[[i, 4]]}}]}, {i, Length[txc3]}], GeoRange → Texas, United States ADMINISTRATIVE DIVISION, Frame → True]]
```

Figure 10: Method for Plotting Texas Tornado Clusters

2.2.3 Distributions

For the final portions of analysis, the distribution of tornado path length, and tornado path angle, were researched. The length of each tornado path can be found by using the Mathematica function GeoDistance. This function returns the distance between two geo spatial points. These lengths were found for each tornado in Texas.

```
txlength = Table[QuantityMagnitude[GeoDistance[{txslat[[i]], txslon[[i]]}, {txelat[[i]], txelon[[i]]}]], {i, Length[txslat]}];
```

Figure 11: Variable to Hold the Length of Each Tornado in Texas

For this calculation, the QuantityMagnitude function was also used. This is to help return values without units as the GeoDistance function returns values with miles units appended. Now that these lengths are isolated into a single vector, a histogram plot can be created using the log of txlength; this is shown in Figure 12. The histogram created by this line can be seen in Figure 42 in the results section.

```
Labeled[Show[Histogram[Log[txlength], ChartStyle → LightBlue, PlotLabel → "Texas Tornado Length Distribution"]],
 {"Log Tornado Length (Miles)", Rotate["Num Tornados", 90 Degree] }, {Bottom, Left}]
```

Figure 12: Command to Output Log Distribution of Lengths on a Histogram

The length distributions are all reported as log normal distributions due to observations and functions used during analysis. The original histogram for the Texas tornado path length txlength variable is as shown.

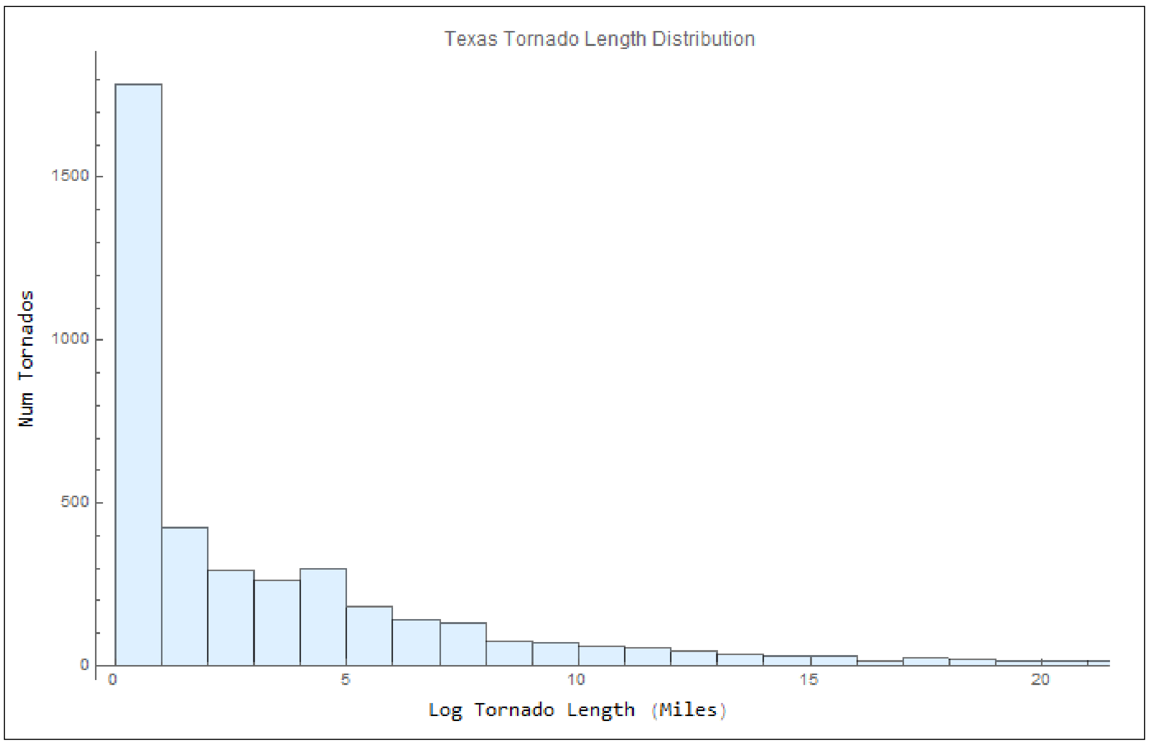


Figure 13: Heavily Skewed Histogram for Texas Tornado Lengths

This makes sense as there isn't a large volume of tornadoes that pass the 10 mile mark. The amount of skewness for this distribution is shown in Figure 14. A highly positive skewness shows that a log normal distribution is what will most likely fit this distribution of lengths. This is why the log command is used when plotting the distributions. On the other hand, the angles did not have as high a skewness, so a regular normal distribution was used.

```
Skewness[txlength]

7.18299
```

Figure 14: Tornado Length Distribution Skewness

Next the angles of each tornado were analyzed. To look into the angle distribution a variable txangle was declared using the GeoDirection function.

```
txangle = Table[GeoDirection[{txslat[[i]], txslon[[i]]}, {txelat[[i]], txelon[[i]]}], {i, Length[txslat]}];
```

Figure 15: Variable to Hold the Angle of Each Tornado in Texas

This function takes in two starting and ending geo spatial points and finds the azimuth of those two angles. Figure 16 below shows the azimuth unit circle. Just below that is the command to plot the histogram itself.

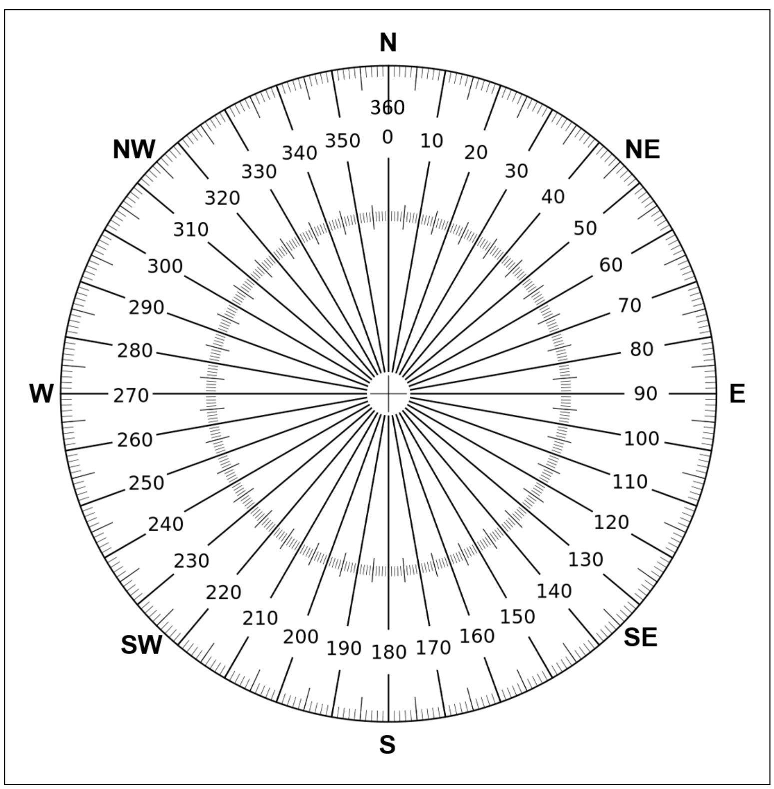


Figure 16: Azimuth Unit Circle

```
Labeled[Show[Histogram[txangle, ChartStyle → LightGreen, PlotLabel → "Texas Tornado Angle Distribution"]],
 {"Log Tornado Angle (Azimuth)", Rotate["Num Tornados", 90 Degree] }, {Bottom, Left}]
```

Figure 17: Command to Output Log Distribution of Angles on a Histogram

This original distribution, as discussed more in the discussions and conclusions section, had outliers due to the fact that some tornadoes don't have any slope or angle at all. This confused Mathematica causing it to output incorrect results. Firstly, to remove this cluster, the other two clusters were combined into a fourth cluster txc4 as shown below.

```
txc4 = Join[txc2, txc3];
```

Figure 18: Texas Tornado Cluster 4

This cluster was then used to produce the angle variable much like the txangle declared before, this is seen in Figure 19. Finally, shown in Figure 20, this new variable was input into the same command to produce a histogram.

```
txangle2 = Table[GeoDirection[{txc4[[i, 1]], txc4[[i, 2]]}, {txc4[[i, 3]], txc4[[i, 4]]}], {i, Length[txc4]}];
```

Figure 19: Declaration of txangle2 for Angle Distribution

```
Labeled[Show[Histogram[txangle2, ChartStyle → LightPurple, PlotLabel → "Texas Tornado Angle Distribution"]],
 {"Tornado Angle (Azimuth)", Rotate["Num Tornados" , 90 Degree]}, {Bottom, Left}]
```

Figure 20: Command to Output Log Distribution of Correct Angles on a Histogram

2.3 Kansas

2.3.1 Data Extraction

The data for Kansas tornadoes were taken from the original data set much like how the Texas tornadoes were taken. The values were stored in four separate variables kaslat, kaslon, kaelat, kaelon. This is shown below along with the command to graph the preliminary tornado paths without any clustering.

```
kaslat = Kansas[All, 1] // Normal;
kaslon = Kansas[All, 2] // Normal;
kaelat = Kansas[All, 3] // Normal;
kaelon = Kansas[All, 4] // Normal;
```

Figure 21: Variables to Store Kansas Starting and Ending Latitude and Longitudes

```
GeoGraphics[Table[{Red, GeoPath[{{kaslat[[i]], kaslon[[i]]}, {kaelat[[i]], kaelon[[i]]}}]},
  {i, Length[Kansas]}], GeoRange → Kansas, United States ADMINISTRATIVE DIVISION , Frame → True]
```

Figure 22: Command to Plot Paths of Tornadoes in Kansas

2.3.2 Clustering

Now that the tornado paths were plotted correctly, it was time once again to create clusters based on the slope of each path. This is to help confirm the Northeast direction hypothesis. The next command, shown in Figure 23, shows the three clusters created and the command to plot them on the same Kansas map.

```
kac1 = Kansas[Select[slope[#slat, #slon, #elat, #elon] === "undefined" || slope[#slat, #slon, #elat, #elon] === 0 &], Values] // Normal;
kac2 = Kansas[Select[slope[#slat, #slon, #elat, #elon] > 0 &], Values] // Normal;
kac3 = Kansas[Select[slope[#slat, #slon, #elat, #elon] < 0 &], Values] // Normal;

Show[GeoGraphics[Table[{Black, GeoPath[{{kac1[[i, 1]], kac1[[i, 2]]}, {kac1[[i, 3]], kac1[[i, 4]]}}]}, {i, Length[kac1]}], GeoRange → Kansas, United States ADMINISTRATIVE DIVISION , Frame → True],
 GeoGraphics[Table[{Blue, GeoPath[{{kac2[[i, 1]], kac2[[i, 2]]}, {kac2[[i, 3]], kac2[[i, 4]]}}]}, {i, Length[kac2]}], GeoRange → Kansas, United States ADMINISTRATIVE DIVISION , Frame → True],
 GeoGraphics[Table[{Red, GeoPath[{{kac3[[i, 1]], kac3[[i, 2]]}, {kac3[[i, 3]], kac3[[i, 4]]}}]}, {i, Length[kac3]}], GeoRange -> Kansas, United States ADMINISTRATIVE DIVISION , Frame → True]]
```

Figure 23: Created Clusters and Geo Path Plot

2.3.3 Distributions

Lastly for Kansas, the distribution of path lengths and path angles were analyzed. Both a kaangle variable and kalength variable were declared to store the tornado angles and tornado paths respectively.

```
kalength = Table[QuantityMagnitude[GeoDistance[{kaslat[[i]], kaslon[[i]]}, {kaelat[[i]], kaelon[[i]]}]], {i, Length[kaslat]}];
```

```
kaangle = Table[GeoDirection[{kaslat[[i]], kaslon[[i]]}, {kaelat[[i]], kaelon[[i]]}], {i, Length[kaslat]}];
```

Figure 24: Kaangle and Kalength Variable Declarations

The two variables were then used to create two histogram plots that highlight the distribution of each. As with the Texas angles, all the tornado paths with zero slope were removed using the technique shown in the Texas Distributions section. A new cluster kac4 was created to hold clusters 2 and 3, then that cluster was used to produce the kaangle2 variable.

```
Labeled[Show[Histogram[Log[kalength], ChartStyle → LightBlue, PlotLabel → "Kansas Tornado Length Distribution"]],
 {"Log Tornado Length (Miles)", Rotate["Num Tornados", 90 Degree] }, {Bottom, Left}]
```

```
Labeled[Show[Histogram[kaangle2, ChartStyle → LightPurple, PlotLabel → "Kansas Tornado Angle Distribution"]],
 {"Tornado Angle (Azimuth)", Rotate["Num Tornados" , 90 Degree]}, {Bottom, Left}]
```

Figure 25: Length and angle Histogram Commands

2.4 Oklahoma

2.4.1 Data Extraction

The starting and ending latitude and longitudes were first taken from the main data set to use for tornado path analysis in Oklahoma. These were stored in a oklahoma variable. Next okslat, okslon, okelat, and okelon were extracted into separate variables as shown below.

```
okslat = oklahoma[All, 1] // Normal;
okslon = oklahoma[All, 2] // Normal;
okelat = oklahoma[All, 3] // Normal;
okelon = oklahoma[All, 4] // Normal;
```

Figure 26: Variables to Store Oklahoma Starting and Ending Latitude and Longitudes

These variables were then used to create geo paths on a map of Oklahoma using this next command in Figure 27

```
GeoGraphics[Table[{Red, GeoPath[{{okslat[[i]], okslon[[i]]}, {okelat[[i]], okelon[[i]]}}]}, {i, Length[oklahoma]}], GeoRange -> oklahoma, Frame → True]
```

Figure 27: Oklahoma Geo Plot Command

2.4.2 Clustering

For the clustering of each of the tornado paths the slope command was once again used to separate the positive, negative, and zero slope paths. Each were stored in three cluster variables okc1, okc2, and okc3.

```
okc1 = oklahoma[Select[slope[#slat, #slon, #elat, #elon] === "undefined" || slope[#slat, #slon, #elat, #elon] === 0 &], Values] // Normal;
okc2 = oklahoma[Select[slope[#slat, #slon, #elat, #elon] > 0 &], Values] // Normal;
okc3 = oklahoma[Select[slope[#slat, #slon, #elat, #elon] < 0 &], Values] // Normal;
```

Figure 28: Oklahoma Geo Plot Clusters

These three clusters were used to plot a clustered geo path plot of Oklahoma tornadoes, below is the command that was used to do that.

```
Show[GeoGraphics[Table[{Black, GeoPath[{{okc1[[i, 1]], okc1[[i, 2]]}, {okc1[[i, 3]], okc1[[i, 4]]}}]}, {i, Length[okc1]}], GeoRange -> oklahoma, Frame → True],
 GeoGraphics[Table[{Blue, GeoPath[{{okc2[[i, 1]], okc2[[i, 2]]}, {okc2[[i, 3]], okc2[[i, 4]]}}]}, {i, Length[okc2]}], GeoRange -> oklahoma, Frame → True],
 GeoGraphics[Table[{Red, GeoPath[{{okc3[[i, 1]], okc3[[i, 2]]}, {okc3[[i, 3]], okc3[[i, 4]]}}]}, {i, Length[okc3]}], GeoRange -> oklahoma, Frame → True]]
```

Figure 29: Oklahoma Geo Plot Cluster Command

2.4.3 Distributions

For the last part of analysis, the distributions of tornado lengths and tornado angles were again looked into. A variable oklength was created, along with a variable okangle.

```
oklength = Table[QuantityMagnitude[GeoDistance[{okslat[[i]], okslon[[i]]}, {okelat[[i]], okelon[[i]]}]], {i, Length[okslat]}];
```

Figure 30: Oklahoma Tornado Length Variable

The okangle variable still includes the outliers from the tornadoes without slope, this was removed before continuing. The output is a new variable named okangle2 as shown in Figure 31 below.

```
okc4 = Join[okc2, okc3];

okangle2 = Table[GeoDirection[{okc4[[i, 1]], okc4[[i, 2]]}, {okc4[[i, 3]], okc4[[i, 4]]}], {i, Length[okc4]}];
```

Figure 31: Oklahoma Tornado Angle Variable

Both variables were then used to create corresponding histograms to display the information. The methods to produce these histograms are shown below in Figure 32.

```
Labeled[Show[Histogram[Log[oklength], ChartStyle → LightBlue, PlotLabel → "Oklahoma Tornado Length Distribution"]],
  {"Log Tornado Length (Miles)", Rotate["Num Tornados", 90 Degree] }, {Bottom, Left}]
```

```
Labeled[Show[Histogram[okangle2, ChartStyle → LightPurple, PlotLabel → "Oklahoma Tornado Angle Distribution"]],
 {"Tornado Angle (Azimuth)", Rotate["Num Tornados" , 90 Degree]}, {Bottom, Left}]
```

Figure 32: Length and Angle Histogram Commands

2.5 Damages

To answer the first question proposed in this report, four different categories were extracted from the main data set. These columns are tornado magnitude (mag), total injuries (inj), total fatalities (fat), and total property loss (loss). Looking into the documentation about each of the categories a specific problem presents itself. The values for property loss change as time goes on. From 1950 to 1996 the property loss is reported as a number 1-9 representing different dollar amounts. The amounts range from $50 to $5,000,000,000. After 1996 the estimated dollar amount was reported instead. This means that for tornadoes in 1950 to 1996 the number will need to be converted to the dollar amount. Lastly, the other column values for injuries, and fatalities are shown in regular integer numbers. To not skew the data during analysis, the dollar amount of property loss is then divided by 1,000,000. The full method in which this is done in Mathematica is shown below in Figure 33. This leads to an output of integer numbers that are more in line with the other categories.

```
damage = data[All, {"mag", "inj", "fat", "loss"}];
damage = damage[All, Values] // Normal;
Do[Which[damage[[i, 4]] < 1, damage[[i, 4]] *= 1000000,
              damage[[i, 4]] == 1, damage[[i, 4]] = 50,
              damage[[i, 4]] == 2, damage[[i, 4]] = 500,
              damage[[i, 4]] == 3, damage[[i, 4]] = 5000,
              damage[[i, 4]] == 4, damage[[i, 4]] = 50000,
              damage[[i, 4]] == 5, damage[[i, 4]] = 500000,
              damage[[i, 4]] == 6, damage[[i, 4]] = 5000000,
              damage[[i, 4]] == 7, damage[[i, 4]] = 50000000,
              damage[[i, 4]] == 8, damage[[i, 4]] = 500000000,
              damage[[i, 4]] == 9, damage[[i, 4]] = 5000000000], {i, Length[damage]}]
Table[damage[[i, 4]] /= 1000000., {i, Length[damage]}];
```

Figure 33: Extracting Damage Data and Converting Property Loss Values

The original data set provides a column that indicates what the magnitude of that specific tornado was. To look at the total damages per tornado magnitude, a new total variable was created to hold a summation of the three damage categories.

```
total = Take[damage, All, {2, 4}];
total = Total[total, {2}];
damage = Take[damage, All, 1];
SortBy[Tally[damage], First];
totaldamage = Table[Append[damage[[i]], total[[i]]], {i, Length[damage]}]
```

Figure 34: Combining Magnitude and Total Damages into Two Column Data

From this a flattened matrix can be created to give a finalized value for each magnitude. This is done through the Flatten command in Mathematica. The matrix is also sorted in ascending order. Figure 35 shows the command and the corresponding output. As can be seen, this matrix includes a row entry for values equal to a magnitude of -9. According to the documentation of this data set, -9 represents tornado magnitudes that are uncertain.

```
totaldamage = Flatten[{#[[1, 1]], Total[#[[All, 2]]]}] & /@ GatherBy[totaldamage, #[[1]] &];
totaldamage = SortBy[totaldamage, First]
{{-9, 0}, {0, 1630.43}, {1, 13259.}, {2, 41222.}, {3, 60064.1}, {4, 60479.5}, {5, 22428.4}}
```

Figure 35: Combining Magnitude and Total Damages into Two Column Data

These uncertain cases are removed before finally graphing the magnitude vs. damage totals.

```
totaldamage = DeleteCases[totaldamage, {x_, y_} /; x == -9]
Labeled[ListLinePlot[totaldamage, PlotStyle → {Thickness[0.005], Hue[0.14, 0.82, 1.]},
  PlotLabel → Style["Total Damage per Tornado Magnitude", 24]],
 {"Magnitude", Rotate["Total Damage", 90 Degree] }, {Bottom, Left}]
```

Figure 36: Removing Uncertain Cases and Plotting Data

Next, each of the specific categories were plotted against the increasing tornado magnitude. This was to give a better idea on the details of each category for each tornado magnitude. From using previous data extraction methods, the injuries were all extracted from the main data set, this is shown in Figure 37a below. To plot the data a list log plot was used. The command for this is shown below as well.

```
injuries = data[All, {"mag", "inj"}];
injuries = injuries[All, Values] // Normal
```

(a) Injury Data

```
Labeled[ListLogPlot[injuries, PlotStyle → PointSize[0.01],
  PlotLabel → Style["Injuries vs Magnitude", 24]],
 {"Magnitude", Rotate["Num Injuries", 90 Degree] }, {Bottom, Left}]
```

(b) Injury Data Plot Command

Figure 37: Injury Data Analysis

This method of analysis was also used to look at the amount of damages per tornado magnitude for both the fatalities and the property loss. Figure 38 shows the fatalities and Figure 39 shows the property loss.

```
fatalities = data[All, {"mag", "fat"}];
fatalities = fatalities[All, Values] // Normal
```

(a) Fatality Data

```
Labeled[ListLogPlot[fatalities, PlotStyle → PointSize[0.01],
  PlotLabel → Style["Fatalities vs Magnitude", 24]],
 {"Magnitude", Rotate["Num Fatalities", 90 Degree] }, {Bottom, Left}]
```

(b) Fatality Data Plot Command

Figure 38: Fatality Data Analysis

```
loss = data[All, {"mag", "loss"}];
loss = loss[All, Values] // Normal;
loss = DeleteCases[loss, {x_, y_} /; y == 0]
```

(a) Property Loss Data

```
Labeled[ListLogPlot[loss, PlotStyle → PointSize[0.01],
  PlotLabel → Style["Property Loss vs Magnitude", 24]],
 {"Magnitude", Rotate["Property Loss (millions)", 90 Degree] }, {Bottom, Left}]
```

(b) Property Loss Data Plot Command

Figure 39: Property Loss Data Analysis

For the property loss the same command as before was used to normalize the data to integer values. The uncertain cases, where property loss equals zero, were also removed.

3 Results

3.1 Texas

A preliminary plotting of Texas' tornado paths was done when analysis of Texas first started. The image below shows this plot output.

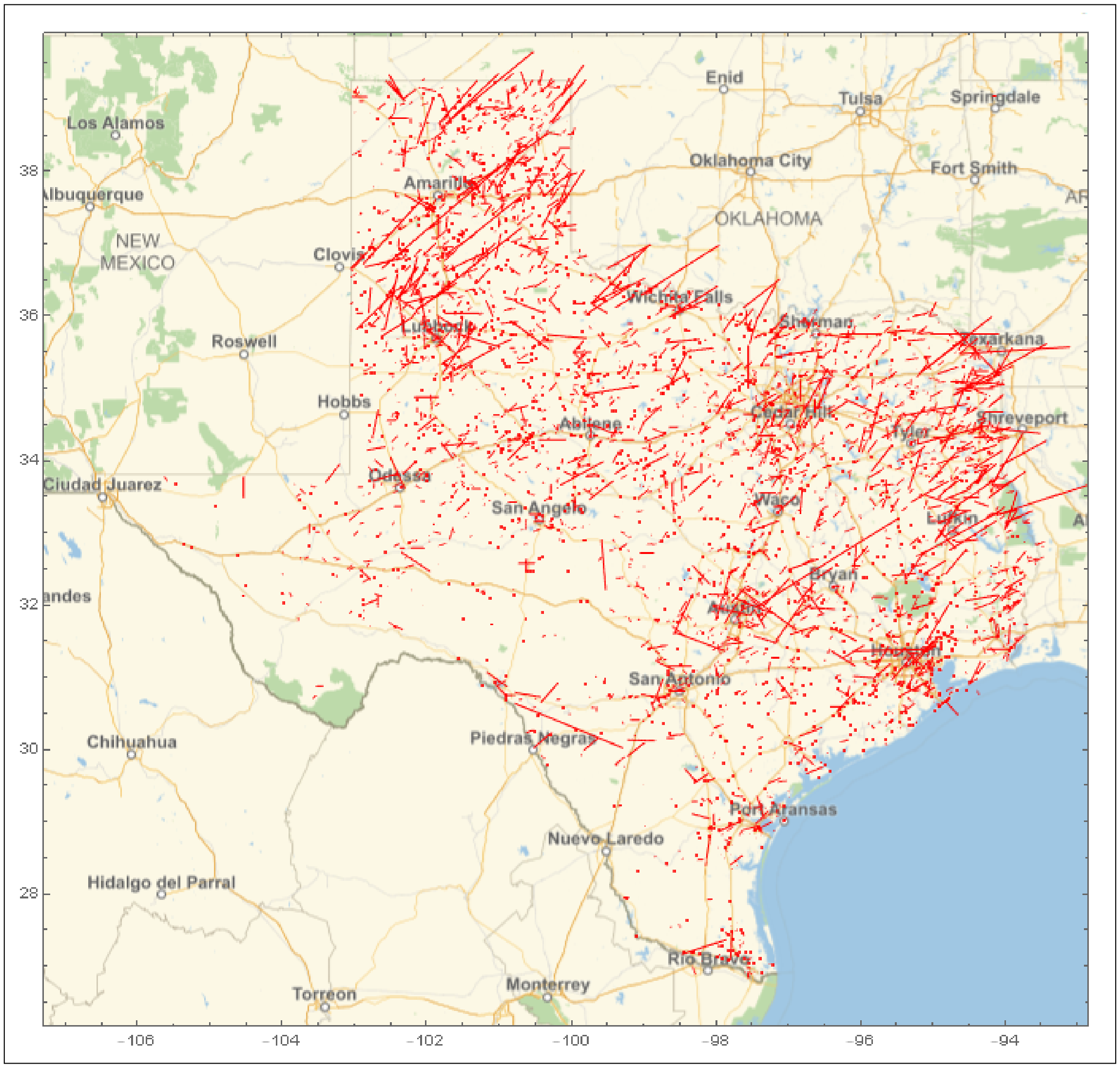


Figure 40: Texas Tornado Path Plot

These tornado paths were then separated into three different clusters based on their corresponding slopes. Each were colored different colors and then plotted onto the Texas map. Figure 41 shows the clusters.

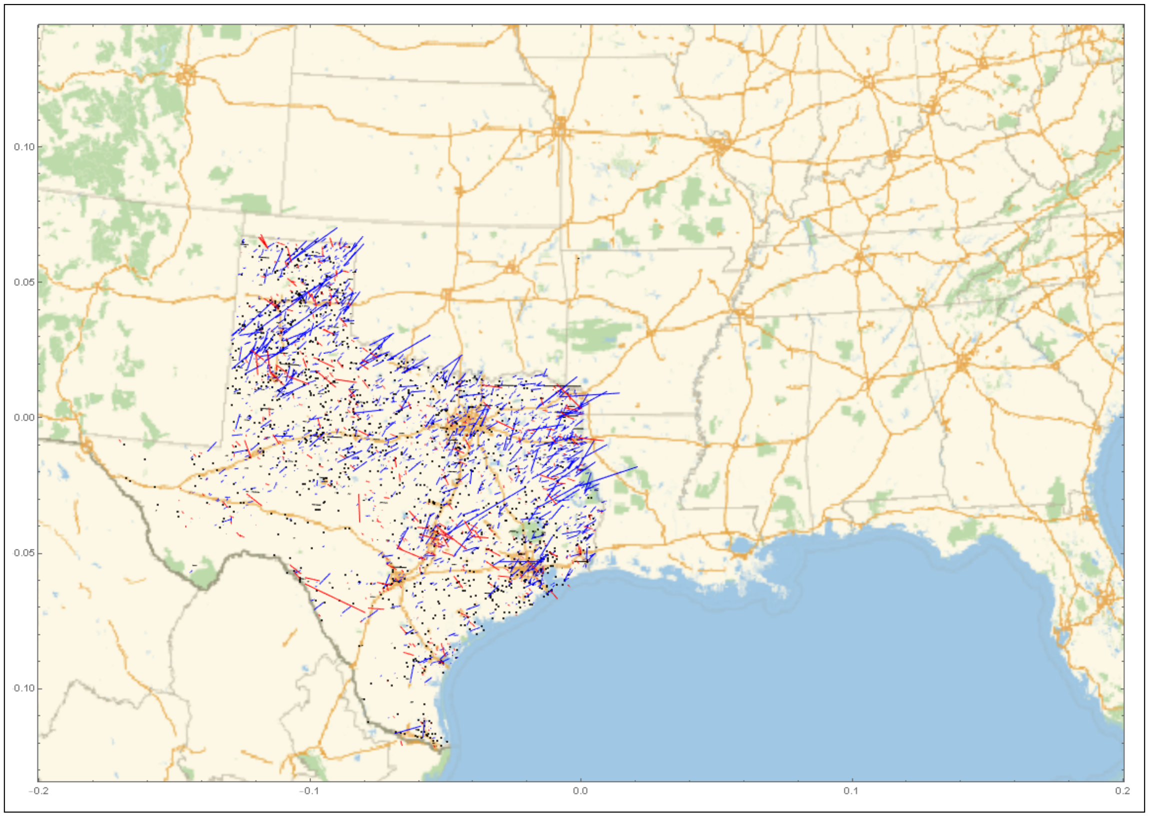


Figure 41: Texas Tornado Path Cluster Plot

Moving onto the distributions of the tornado path lengths and angles for each tornado in Texas, the histogram below shows the distribution of path lengths.

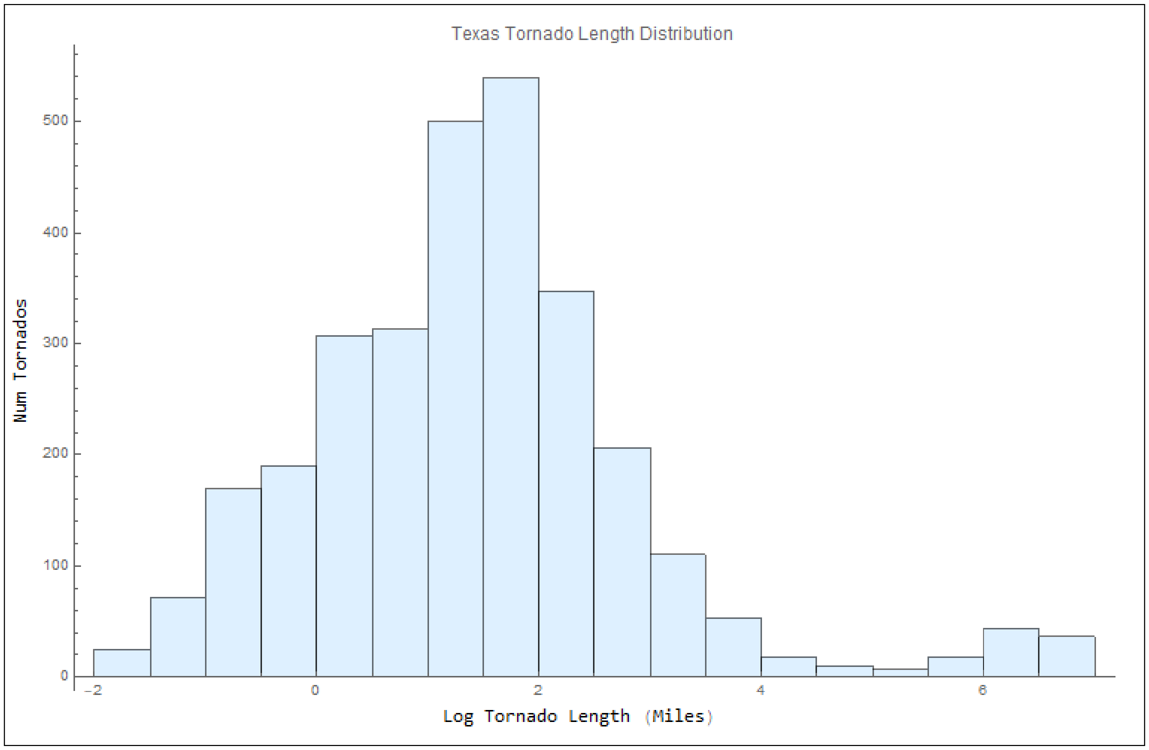


Figure 42: Texas Tornado Path Length Distribution

For the last part of analysis of Texas tornadoes, the distribution of angles of each tornado path was analyzed. There were outliers that needed to be removed from the original distribution, but upon removing them the output below was obtained.

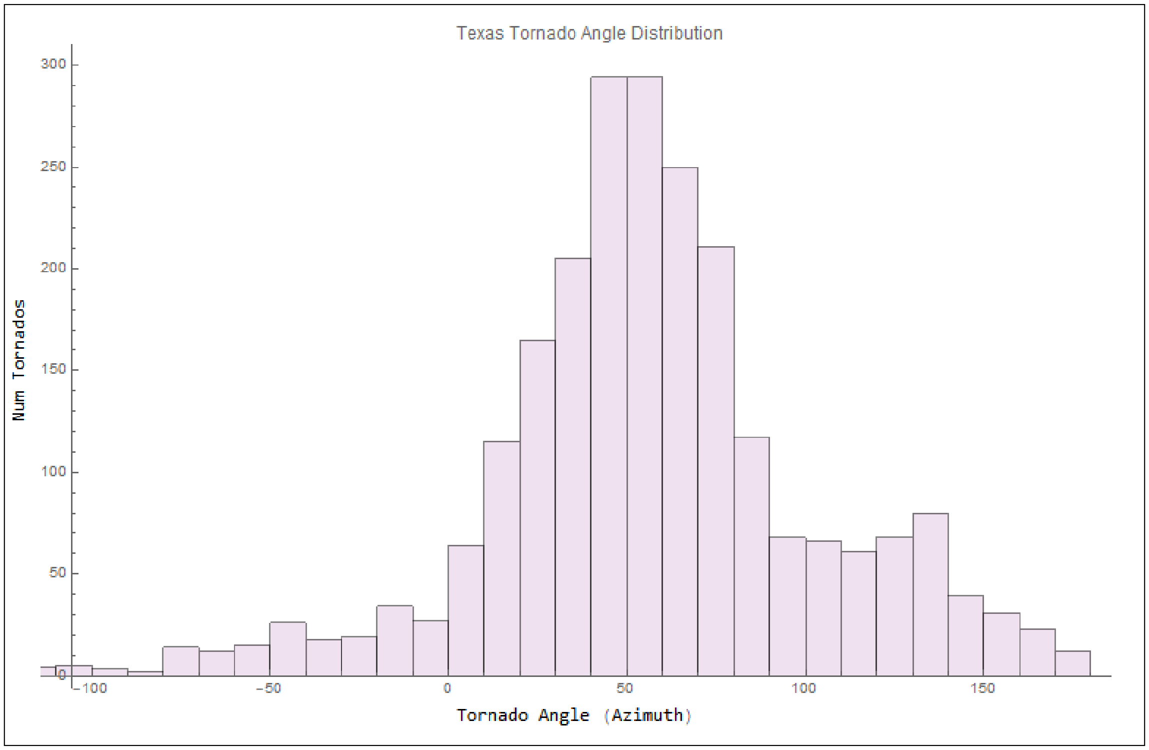


Figure 43: Texas Tornado Path Angle Distribution

3.2 Kansas

To first get a good understanding of how the tornadoes act in Kansas, a plot of the tornado paths was done through the command in Figure 22.

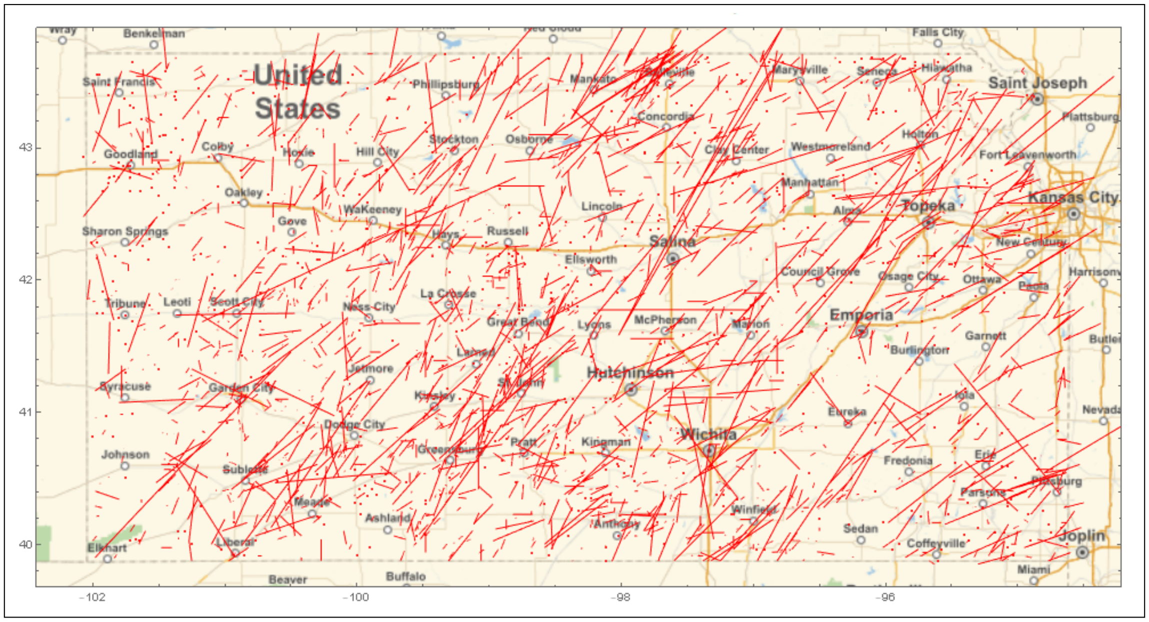


Figure 44: Kansas Tornado Path Plot

Once again three clusters based on tornado path slope were created. These were then plotted on the same map to check if the Northeast direction hypothesis upholds.

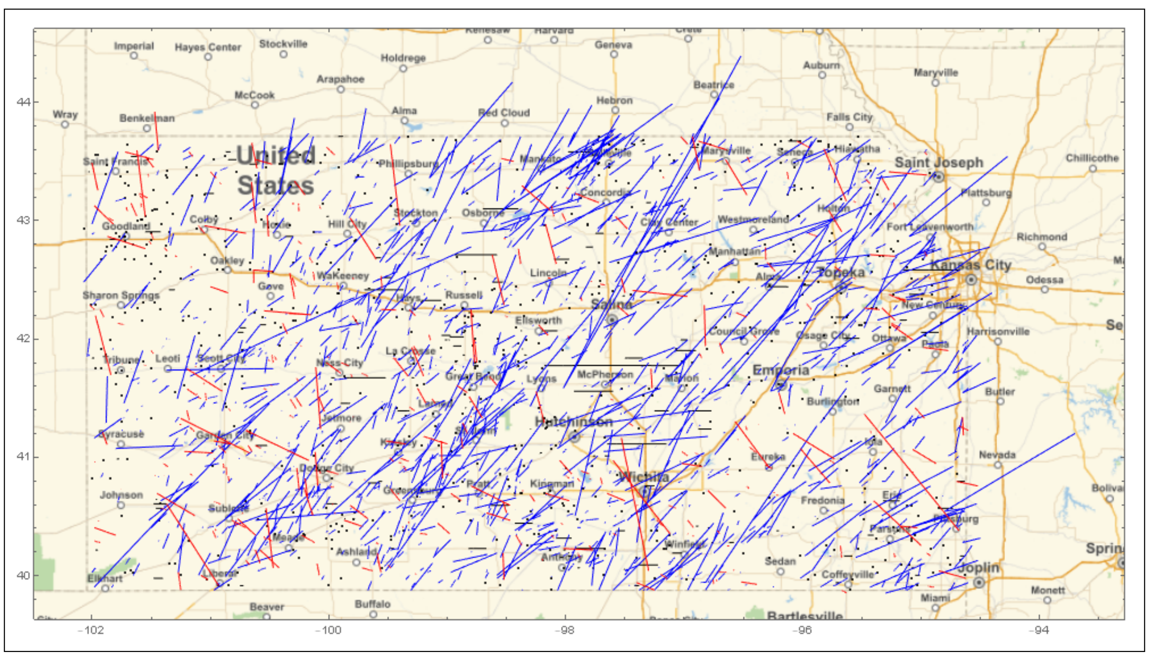


Figure 45: Kansas Tornado Path Cluster Plot

As for the distributions, both the angle and length of each tornado was looked at again. The commands in Figure 25 were both used to produce these corresponding outputs seen below.

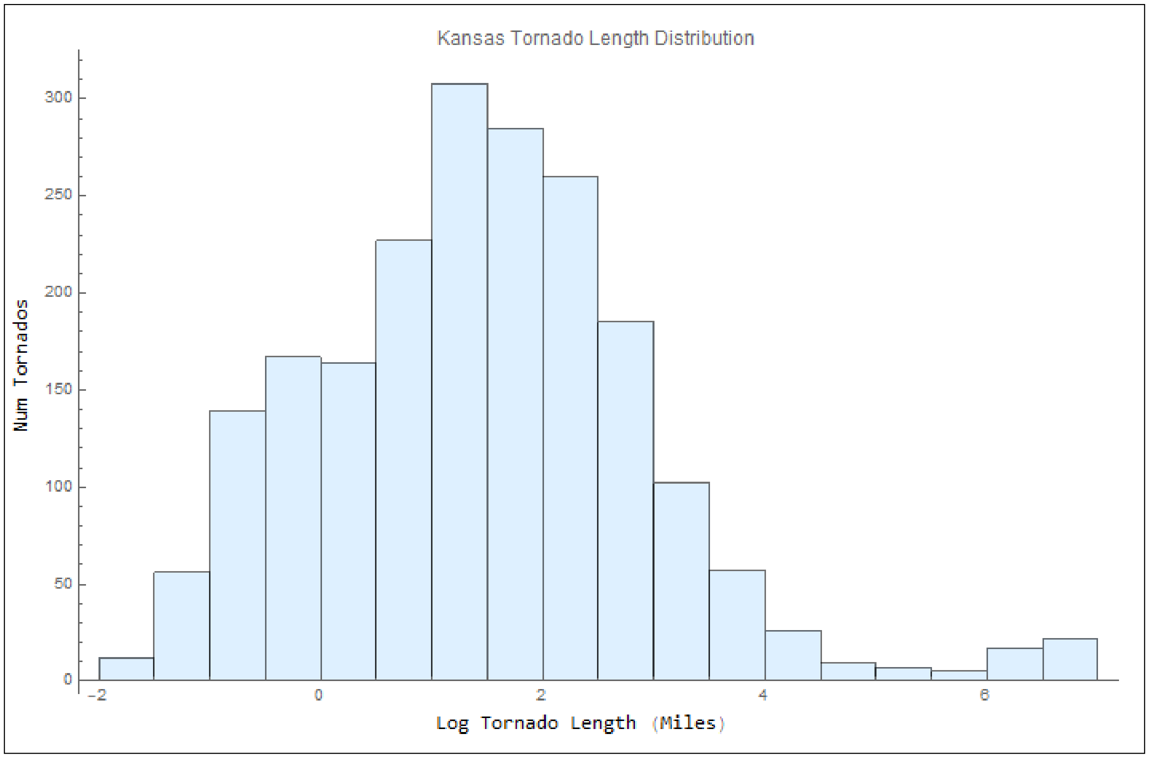


Figure 46: Kansas Tornado Length Histogram

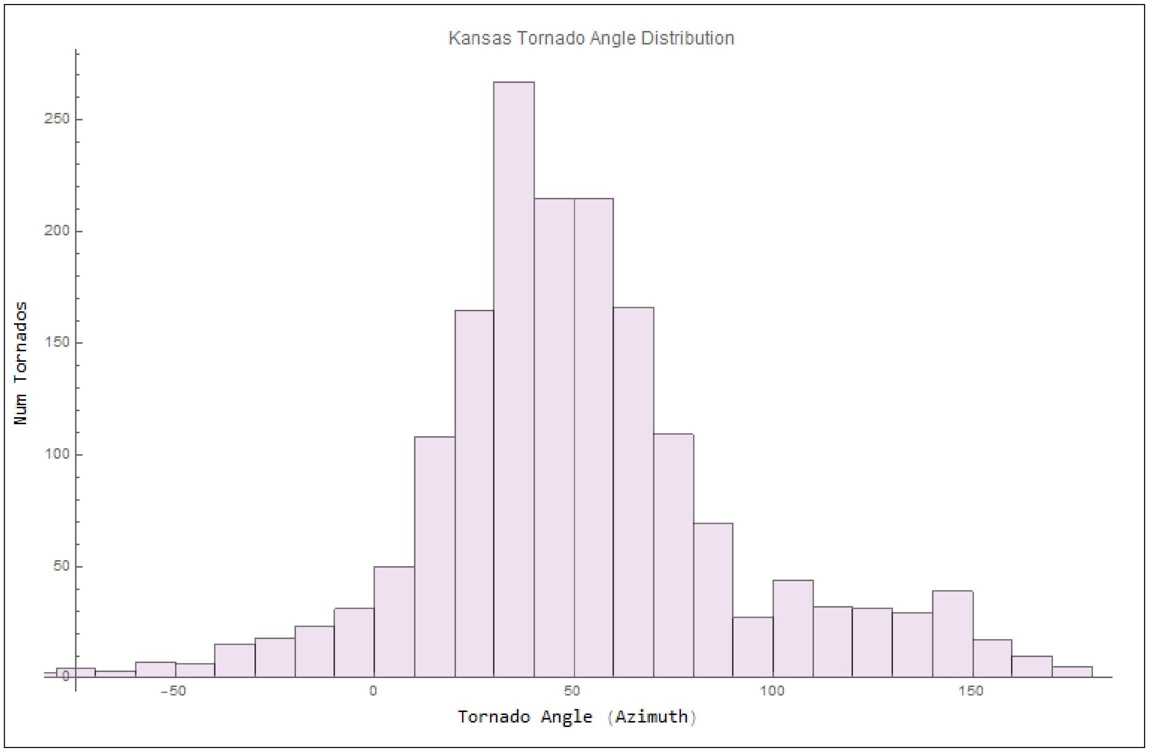


Figure 47: Kansas Tornado Angle Histogram

3.3 Oklahoma

For the third and final state Oklahoma was analyzed for its tornadoes. This was done the same was as Texas and Kansas. The next four figures show the original paths, the clustered paths, the length distribution, and the angle distribution of the Oklahoma tornadoes respectively.

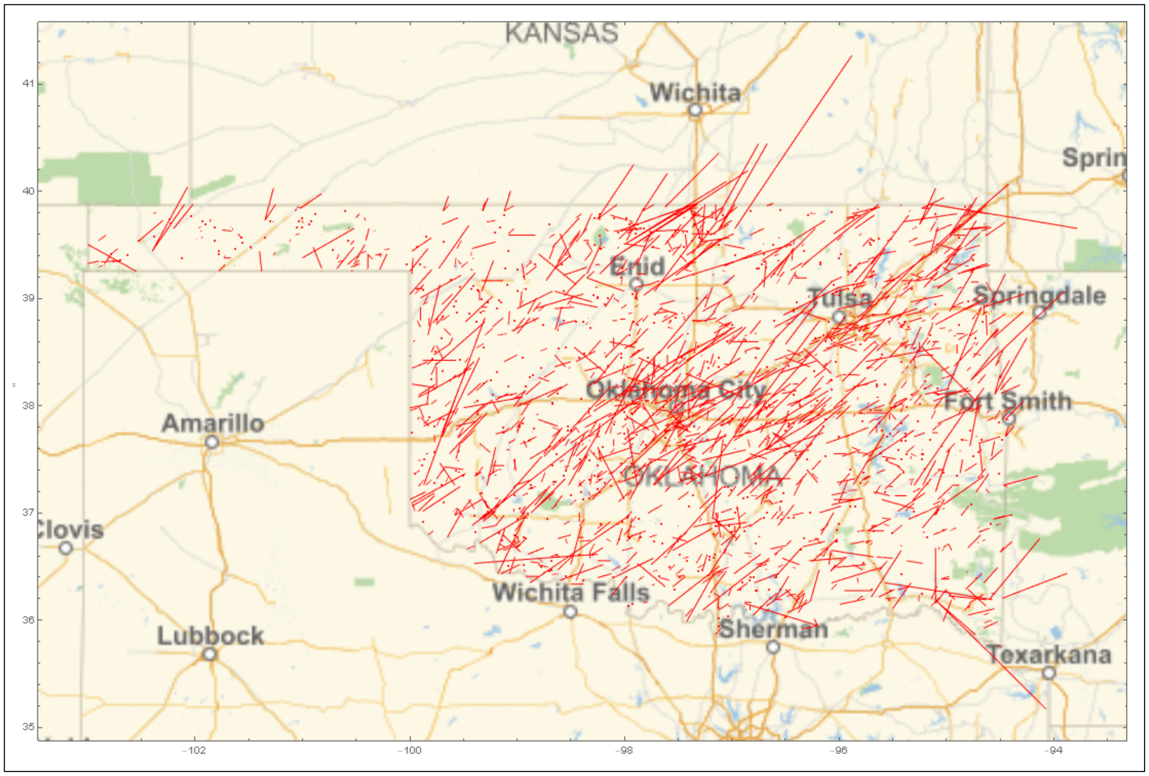


Figure 48: Oklahoma Tornado Path Plot

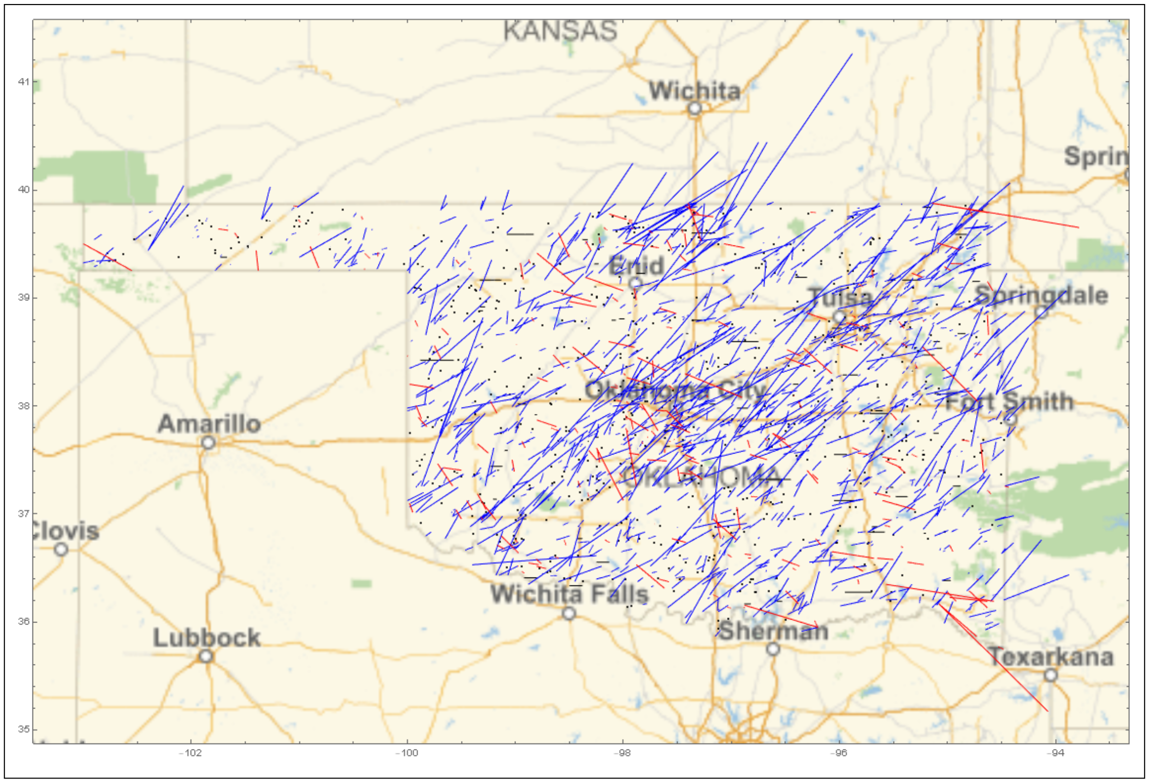


Figure 49: Oklahoma Tornado Path Cluster Plot

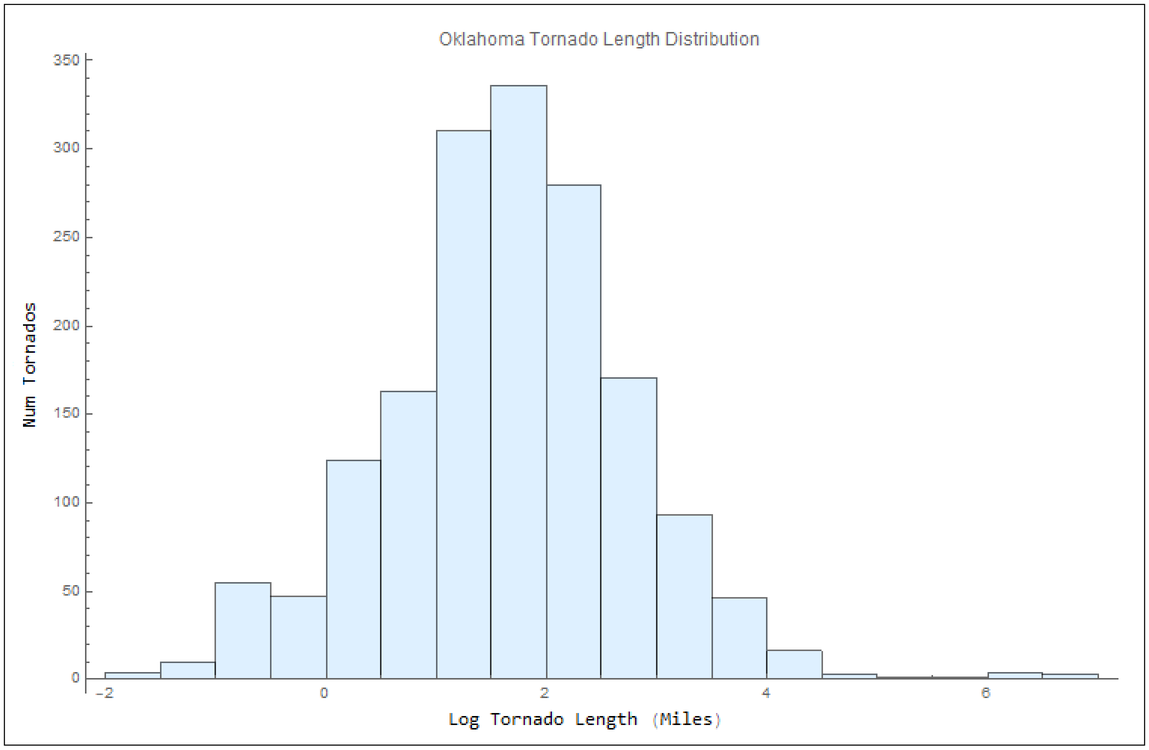


Figure 50: Oklahoma Tornado Length Histogram

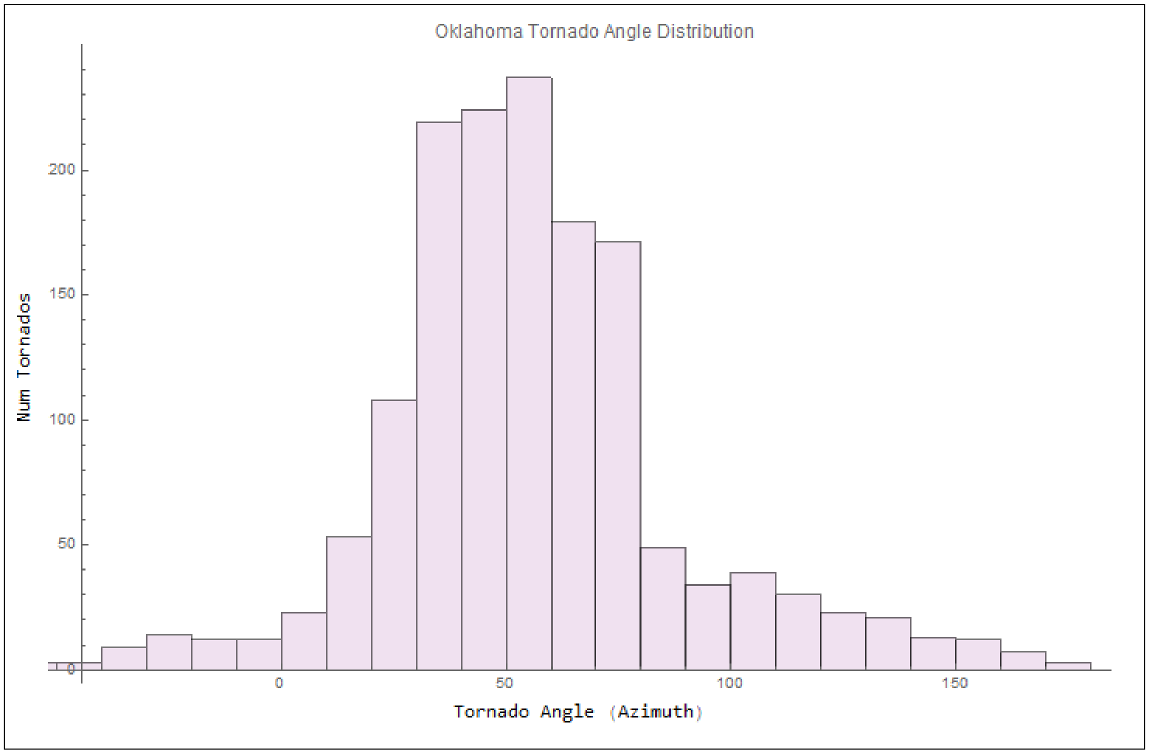


Figure 51: Oklahoma Tornado Angle Histogram

3.4 Damages

Analysis was done on the total damages that a tornado creates. For first analysis the total damage per tornado magnitude was extracted from the data set. This total damage was then plotted against each tornado magnitude. The resulting plot is shown below.

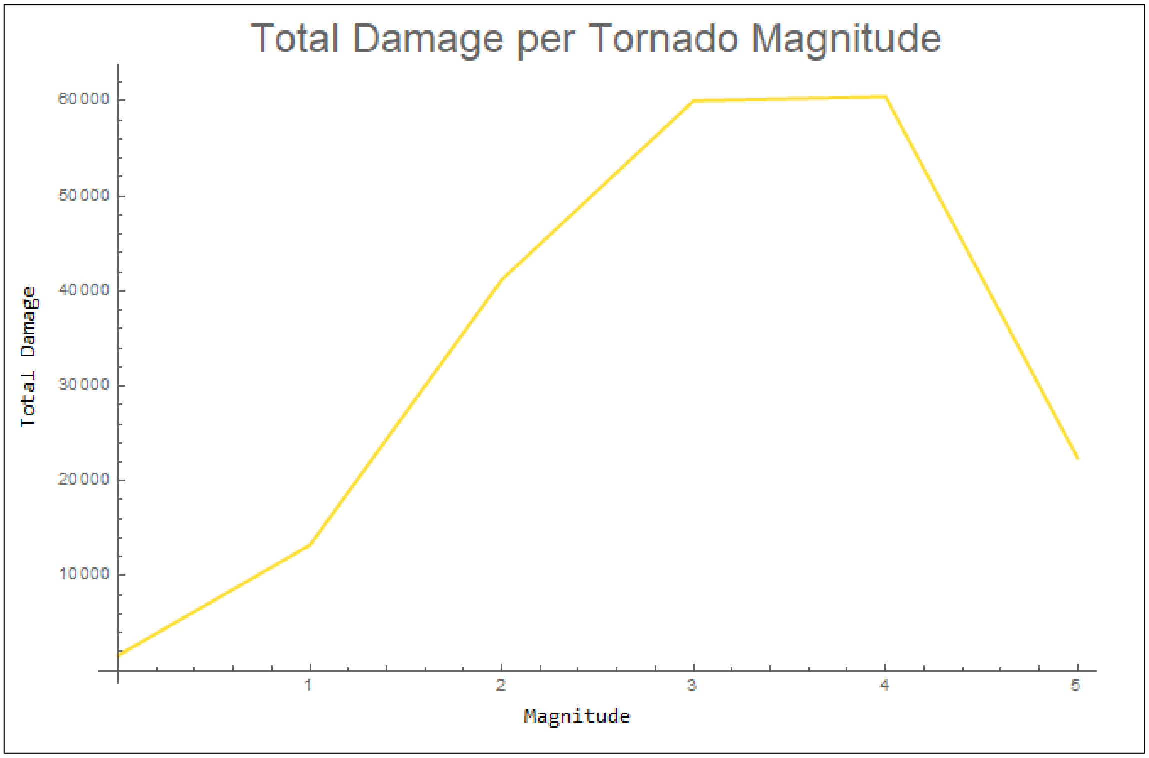


Figure 52: Total Damage per Tornado Magnitude Plot

Next each of the specific categories, injuries, fatalities, and property loss, were also potted using list log plots. The next three figures show the plots for each of the categories respectively.

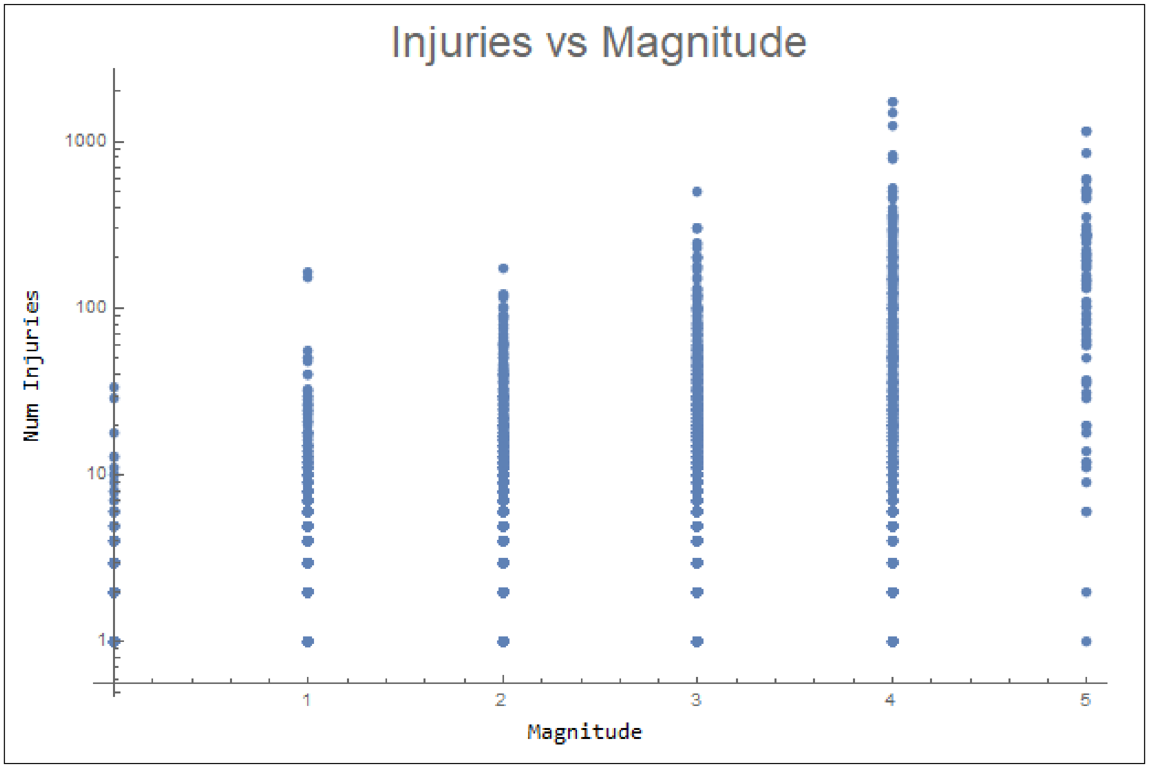


Figure 53: Amount of Injuries Per Tornado Magnitude

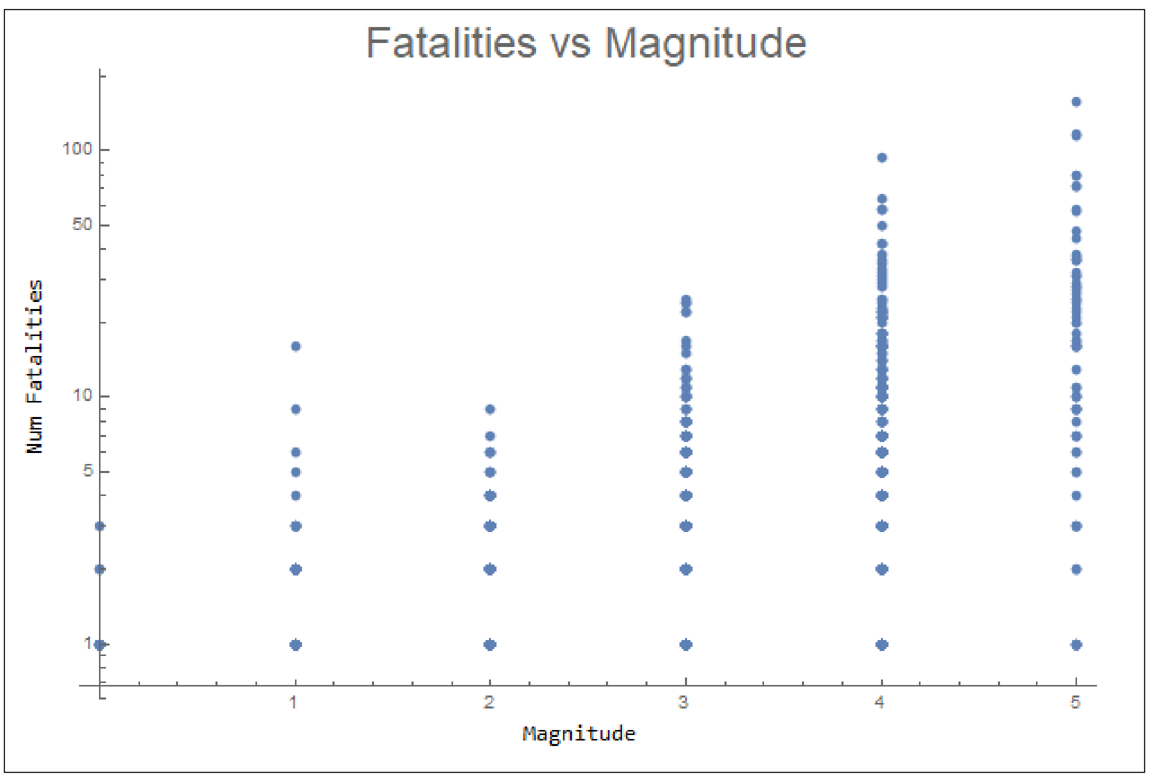


Figure 54: Amount of Fatalities Per Tornado Magnitude

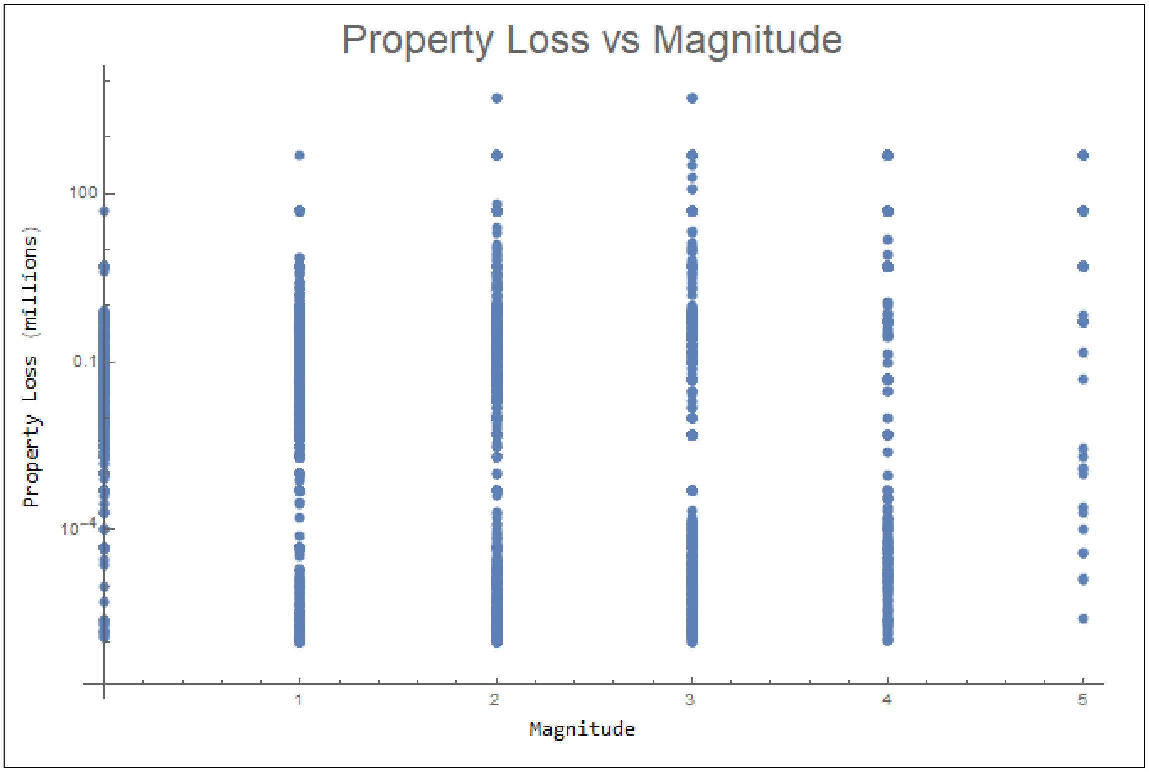


Figure 55: Amount of Property Loss Per Tornado Magnitude

4 Discussion and Conclusion

First, looking at the data set, is was determined what would be main areas of interest. Overall, the starting and ending latitudes and longitudes were widely used throughout the analysis. They were used to plot a geo histogram of the starting points. Through here it was more clear which states would have the most tornado volume. However, Florida and Louisiana were seen as the ones with the most, when later looking at the exact list of tornadoes per state, Texas, Kansas, and Oklahoma were picked as the three with the most tornadoes. Texas had a much higher amount of tornadoes compared to all the rest of the states, this is due to that fact that it is the largest state in the country and is located in the Central United States.

When first looking at the paths of each tornado in Texas, there was a seeming trend to the directions that the tornadoes were traveling. This trend looked as if a majority of tornadoes were traveling in the Northeast direction. When later trying to confirm this using a histogram of the azimuth measurements, it wasn't precisely clear. This is due to an outlier that drastically skewed the data. Figure 56 below shows there is a large amount of tornadoes that have a angle of around 200°.

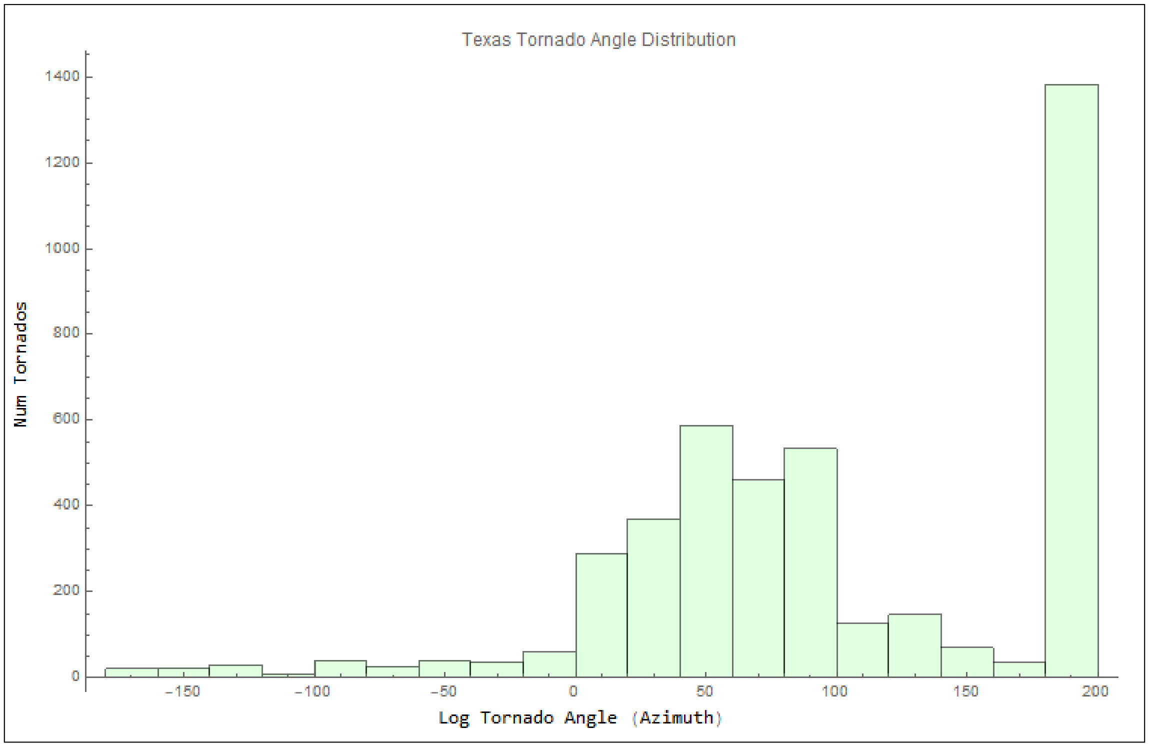


Figure 56: Texas Tornado Path Angle Distribution w/ Outliers

It was suspected that the cluster involving tornadoes that have no slope was causing this. Looking at the histogram, in Figure 57, for only the first cluster txc1, this indeed is the case. Ultimately, removing these clusters as described helped to produce the histogram displayed in Figure 43 in the results section. This new histogram now clearly shows that there is a Northeast trend for tornadoes that land in Texas, as most values reside around the 40°to 60°area. This change was also made to the angles of Kansas and Oklahoma.

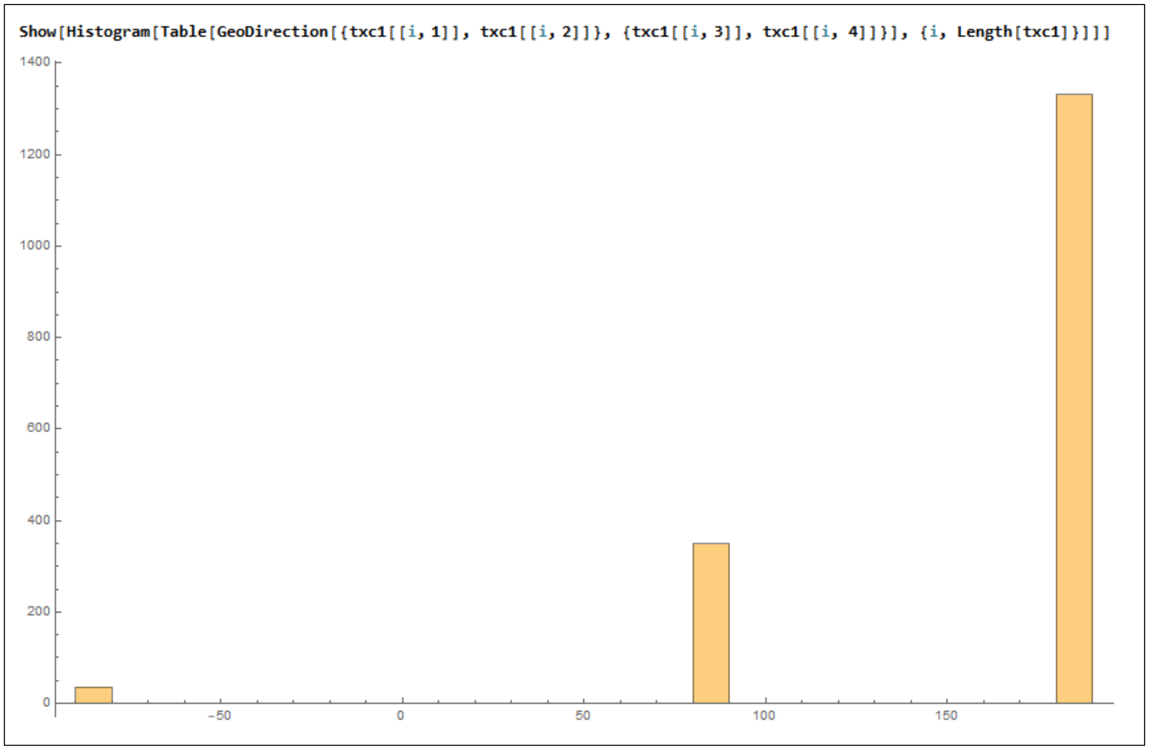


Figure 57: Texas No Slope Tornado Cluster Histogram

Finally, looking at the standard deviation and mean for the angles gives a more conclusive idea into which direction the tornadoes are definitively traveling. For Texas tornadoes the values are as reported below.

```
StandardDeviation[txangle2]
57.7145°
```

(a) Standard Deviation

```
Mean[txangle2]
50.6422°
```

(b) Mean

Figure 58: Further Analysis of Texas Tornado Angle Distribution

This analysis was also done for the angle distributions of Kansas and Oklahoma tornadoes, as seen in Figures 59 and 60. All the numbers shown give conclusive evidence that most tornadoes in any state travel in a Northeastern direction.

```
StandardDeviation[kaangle2]

49.0686°
```

(a) Standard Deviation

```
Mean[kaangle2]

47.3025°
```

(b) Mean

Figure 59: Further Analysis of Kansas Tornado Angle Distribution

```
StandardDeviation[okangle2]

36.8081°
```

(a) Standard Deviation

```
Mean[okangle2]

54.7921°
```

(b) Mean

Figure 60: Further Analysis of Oklahoma Tornado Angle Distribution

After all analysis was done on the angle distributions for the three states with the most tornadoes, it was clear that there was a common Northeast direction that each of the tornadoes usually took. Figures 43, 47, 51 all show a normal peak between 40°and 60°as previously predicted. Looking further into the kurtosis values associated with each state's angles in Figures 61, 62, and 63 below, it can be seen that each are relatively close to a normal distribution. The commands needed to obtain each of these graphs is also shown.

```
txangle2 = DeleteCases[txangle2, x_ /; x == 0];
txakurt = Kurtosis[txangle2];
QuantilePlot[txangle2,
 PlotLabel → "Texas Tornado Angle Distribution Quantile Plot",
 Epilog → {Text[Style[StringForm["Kurtosis: ``", txakurt], 18], {-1, 2}]}]
```

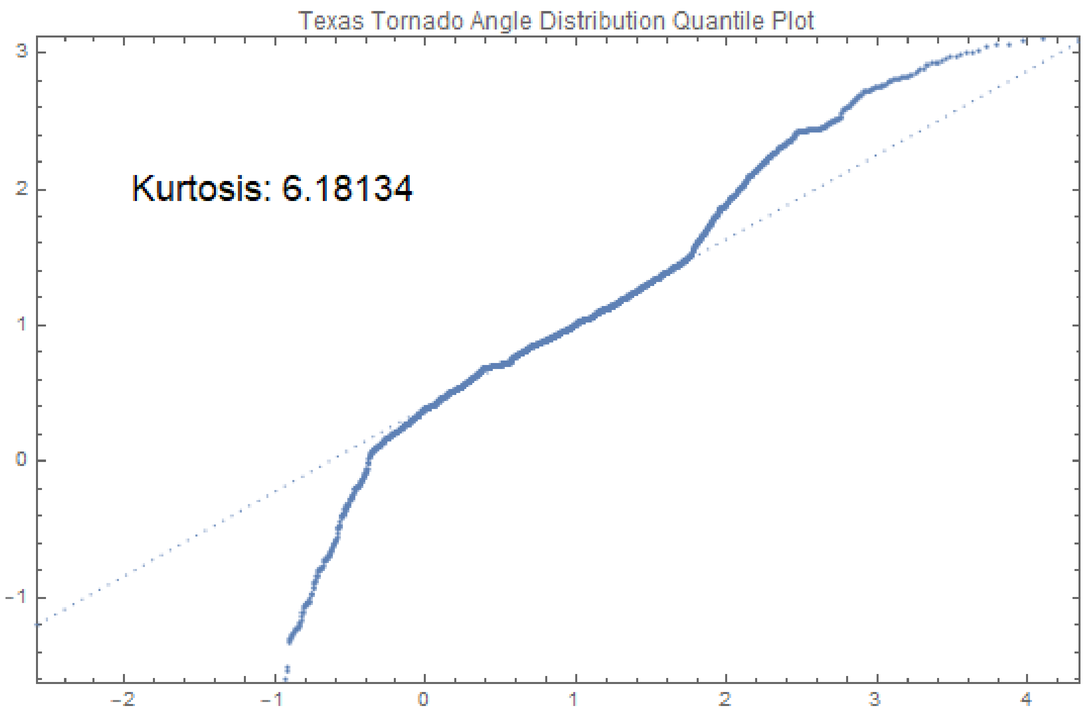


Figure 61: Texas Angle Distribution Kurtosis

```
kaangle2 = DeleteCases[kaangle2, x_ /; x == 0];
QuantilePlot[kaangle2,
 PlotLabel → "Kansas Tornado Angle Distribution Quantile Plot",
 Epilog → {Text[Style[StringForm["Kurtosis: ``", Kurtosis[kaangle2]], 18], {-1, 2}]}]
```

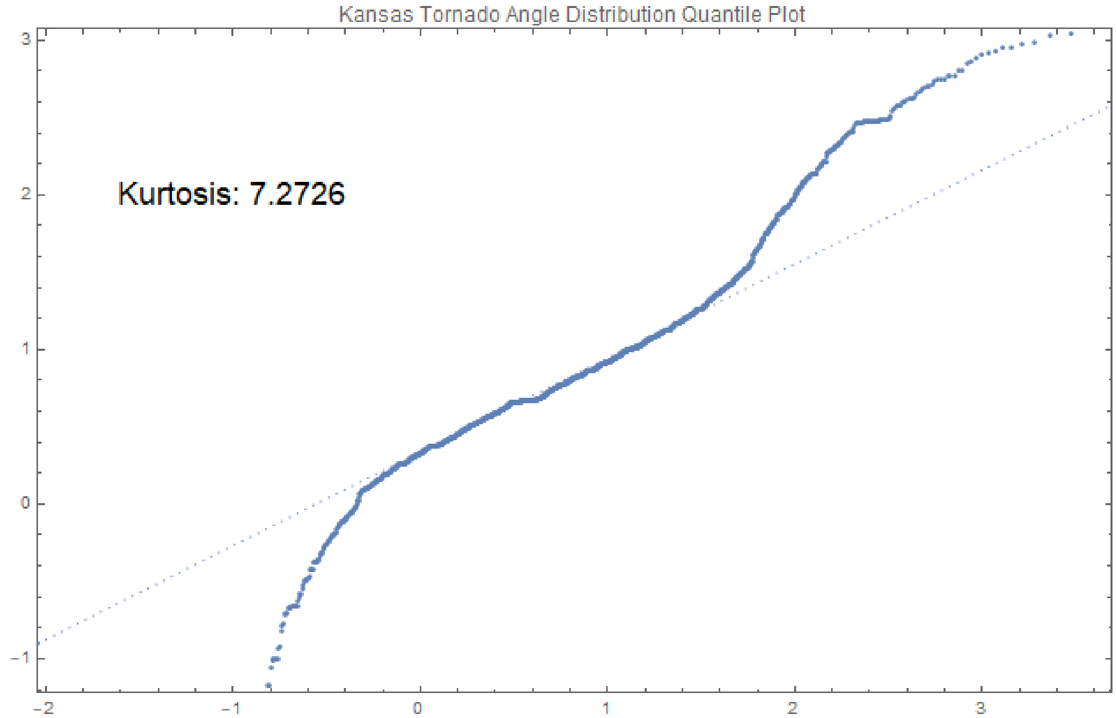


Figure 62: Kansas Angle Distribution Kurtosis

```
okangle2 = DeleteCases[okangle2, x_ /; x == 0];
QuantilePlot[okangle2,
 PlotLabel → "Oklahoma Tornado Angle Distribution Quantile Plot",
 Epilog → {Text[Style[StringForm["Kurtosis: ``", Kurtosis[okangle2]], 18], {0, 2}]}]
```

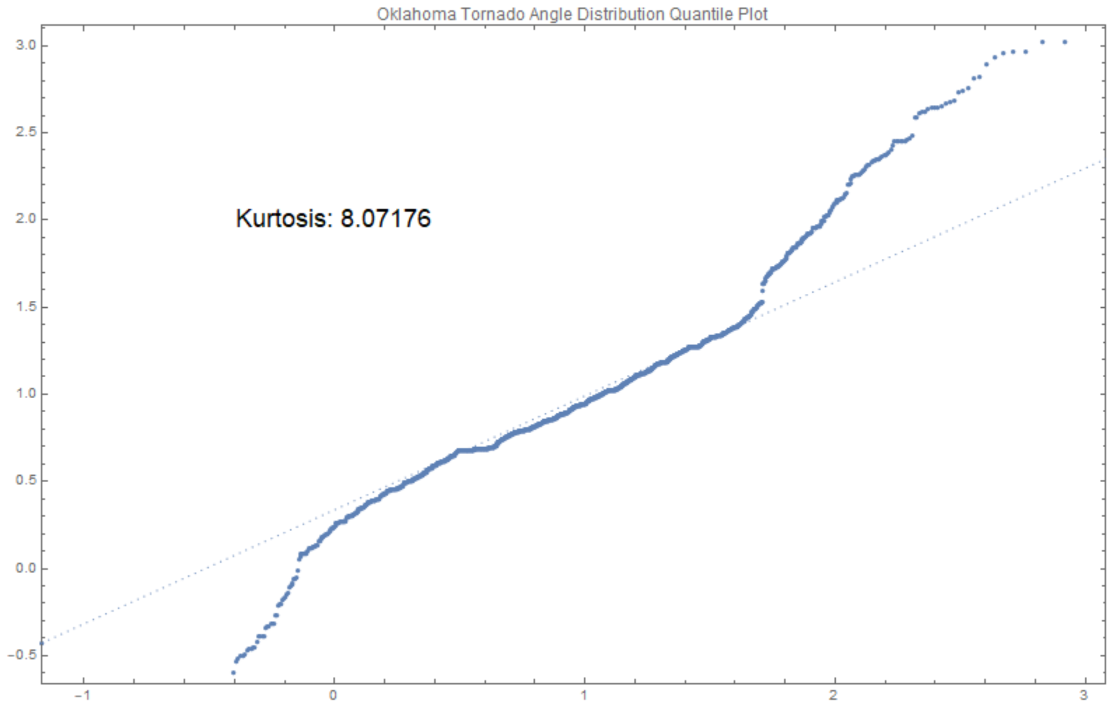


Figure 63: Oklahoma Angle Distribution Kurtosis

Looking at the tails of each quantile plot shows why each of the kurtosis values are so high compared to regular normal value of 3. Each quantile plot shows that the peak resembles a normal distribution closely, while the tails aren't as close. This would mean that the peak is very distance compared to the tails, and that each side of the peak would show a large drop off. Looking at the angle distribution results in the results section shows this to be the case.

To answer the question of how far does a tornado travel on average the length distributions were consulted. The current distributions for each state in the results section show a promising log normal distribution. More analysis into the kurtosis of each distribution shows this to be true.

```
txlength2 = DeleteCases[txlength, x_ /; x == 0];
kurtosis = Kurtosis[Log[txlength2]];
QuantilePlot[Log[txlength2],
 PlotLabel → "Texas Tornado Length Distribution Quantile Plot",
 Epilog → {Text[Style[StringForm["Kurtosis: ``", kurtosis], 18], {-1, 4}]}]
```

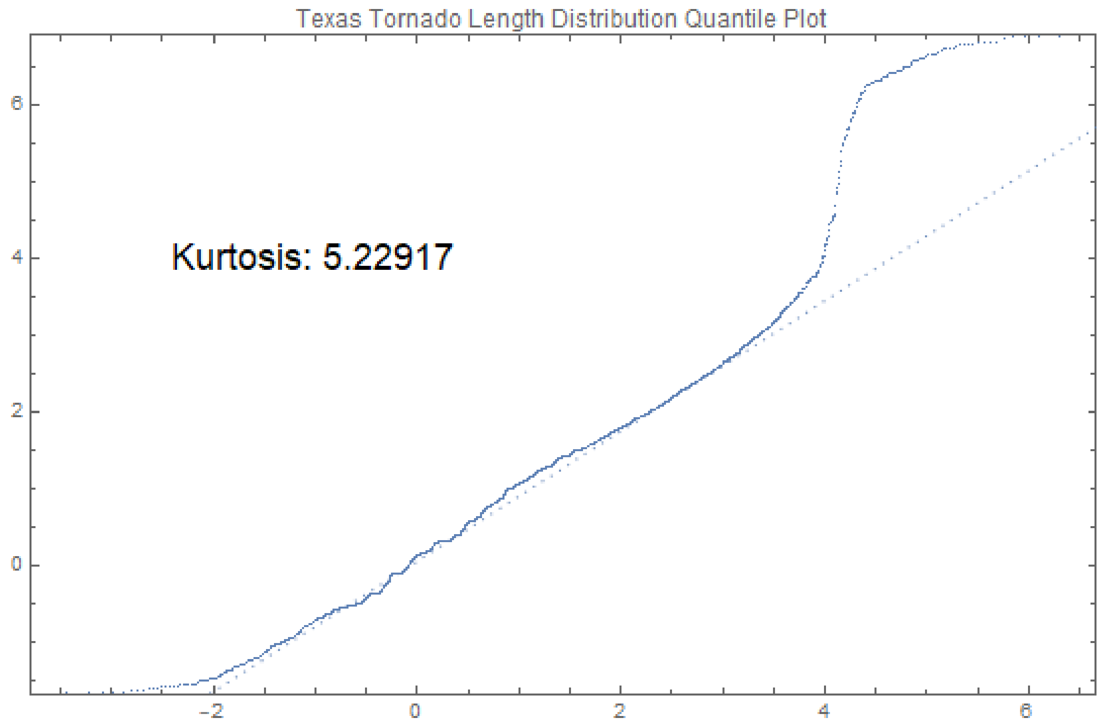


Figure 64: Texas Length Distribution Kurtosis

This Figure 64 above shows that the length distribution of tornadoes in Texas closely resembles a log normal distribution. Only the right tail deviates from the line, leading to the slightly high kurtosis value. Figures 65 and 66 also show the quantile plots of Kansas and Oklahoma tornado length distribution, both of which display even smaller kurtosis values.

```
kalength2 = DeleteCases[kalength, x_ /; x == 0];
QuantilePlot[Log[kalength2],
 PlotLabel → "Kansas Tornado Length Distribution Quantile Plot",
 Epilog → {Text[Style[StringForm["Kurtosis: ``", Kurtosis[Log[kalength2]]], 18], {-1, 4}]}]
```

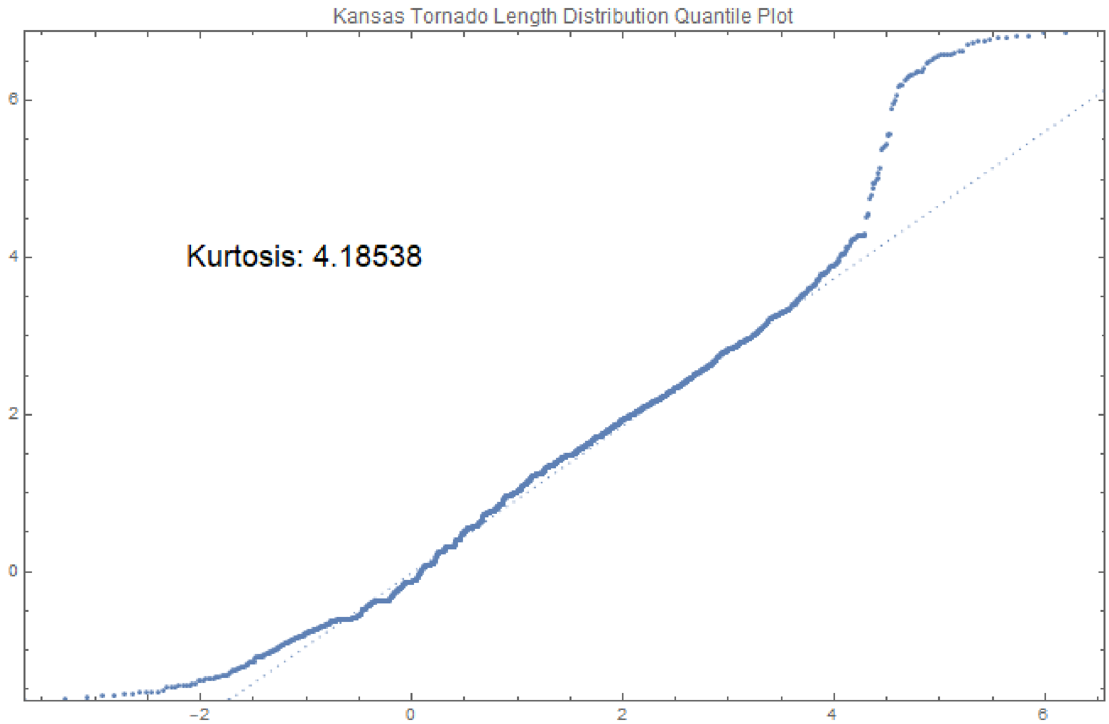


Figure 65: Kansas Length Distribution Kurtosis

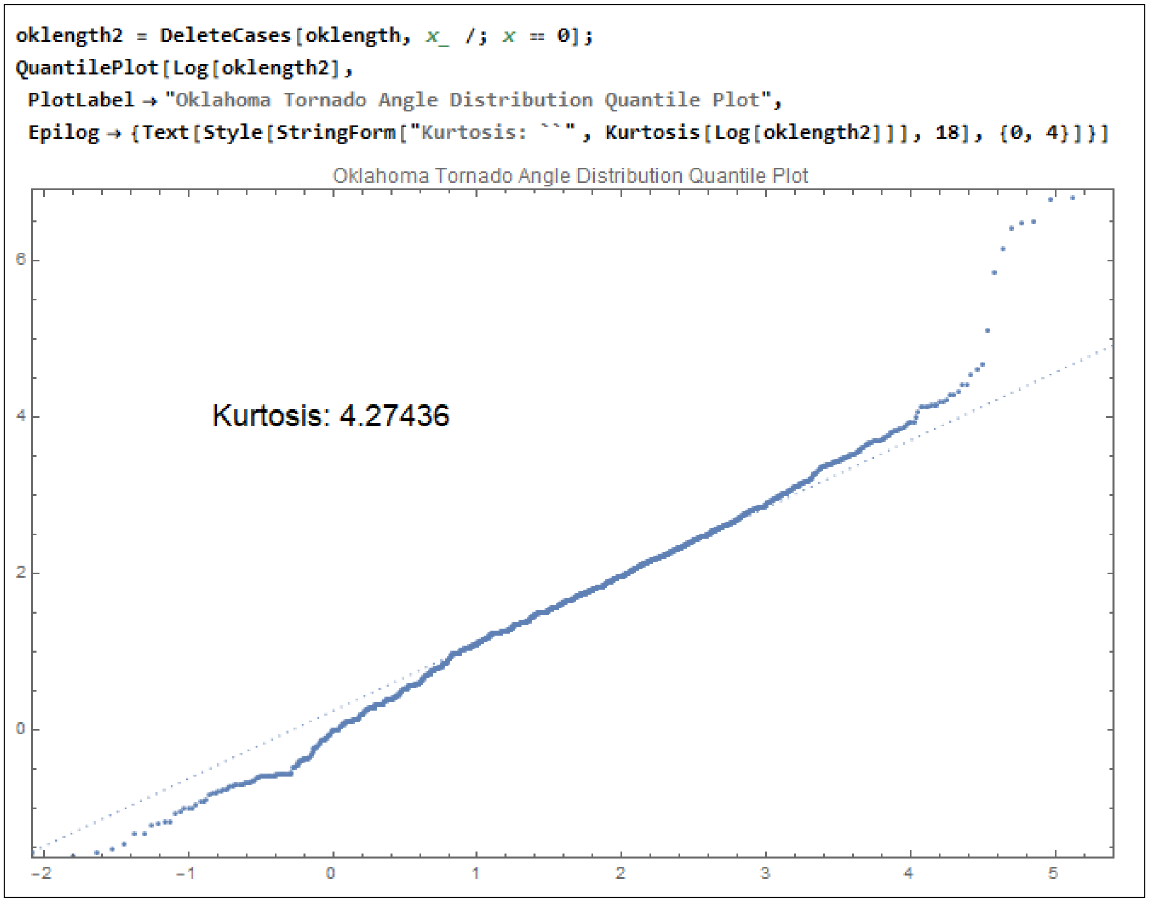


Figure 66: Oklahoma Length Distribution Kurtosis

Finally, taking the mean gives a better idea of what the average tornado length is in miles. This was done for each of the three states. For Texas and Kansas the mean length was almost double that of Oklahoma. This could be due to the fact that Oklahoma doesn't have as many tornadoes as Kansas or Texas. Taking the average of these three numbers give a final tornado length average of around 15.1 miles per tornado.

```
Mean[txlength]
18.834
```

(a) Texas Mean Length

```
Mean[kalength]
17.2678
```

(b) Kansas Mean Length

```
Mean[oklength]
9.15268
```

(c) Oklahoma Mean Length

Figure 67: State Tornado Length Means

The last portion of analysis was set out to answer question number 1: Does tornado damage increase with tornado magnitude? The results section shows very promising answers to this question. Overall, it is to be assumed that damage does in fact increase as tornado magnitude increases. This can be concluded from the plot in Figure 52. From a magnitude of 1 to 4 there is increasing damage across the board. However, there is a large drop off of damage for the highest tornado magnitude. This is due to the fact that there are vastly less tornadoes that reach this magnitude of 5. Looking at the exact amount of each tornado in each category this becomes more clear.

```
{{{-9}, 111}, {{0}, 29884}, {{1}, 21829}, {{2}, 9496}, {{3}, 2703}, {{4}, 714}, {{5}, 88}}
```

Figure 68: Tornado Amount Per Magnitude

As shown in Figure 68 above, there are only 88 tornadoes in the magnitude 5 category. There is also a small amount of magnitude 4 tornadoes compared to magnitude 3. However, according to the total damage plot, magnitude 4 tornadoes still do more damage than 3. This would suggest that each magnitude 4 tornado does a massive amount of damage.

As for each of the list log plots for injuries, fatalities, and property loss, there does seem to be an increase in each category with magnitude. This is especially apparent with the fatalities plot in Figure 53. It makes sense that the larger magnitude tornadoes contain more fatalities, as fatalities are a rare occurrence per tornado when compared with property loss and injuries.

Overall, the analysis performed with each of the three states, Texas, Kansas, and Oklahoma, answered questions 2, 3, and 4. The most compelling result from this analysis is the Northeast direction of the tornadoes. This question was not initially asked because the analyst did not know this could be a possibility. Upon further research it seemed like there was a specific direction that most tornadoes traveled in. As for the damage analysis, the results show a clear correlation with tornado magnitude, ultimately answering question number 1.

5 References

[1] NOAA/SVRGIS, “Usa - historical tornado tracks - spc noaa,” Dec 2016. [Online]. Available: https://oasishub.co/dataset/usa-tornado-historical-tracks-noaa

6 Appendices

6.1 Data Set Documentation

SPC Tornado, Hail, and Wind Database Format Specification (for .csv output)

Comma separated value (.csv) files available at http://www.spc.noaa.gov/wcm/#data

Please note! These files are an attempt to represent the data that is submitted to the *Storm Data* publication by National Weather Service field offices. Careful review of the data is conducted at the National Climate Data Center and the Storm Prediction Center. Nonetheless, some errors/discrepancies may exist. Please contact Gregory.Carbin@noaa.gov for additional information/clarification. (This document was last updated on April 7, 2010.)

Field No. - (MySQL torn field id), (hail field id), (wind field id)
Description

1-(om)
Tornado number - A count of tornadoes during the year: Prior to 2007, these numbers were assigned to the tornado as the information arrived in the NWS database. Since 2007, the numbers may have been assigned in sequential (temporal) order after event date/times are converted to CST. However, do not use "om" to count the sequence of tornadoes through the year as sometimes new entries have come in late, or corrections are made, and the data are not re-sequenced.

NOTE: Tornado segments that cross state borders and/or more than 4 counties will have same OM number. See information about fields 22-24, below.

2-(yr)
Year, 1950-2009

3-(mo)
Month, 1-12

4-(dy)
Day, 1-31

5-(date)
Date in yyyy-mm-dd format

6-(time)
Time in HH:MM:SS

7-(tz)
Time zone - All times, except for ?=uknown and 9=GMT, were converted to 3=CST. This should be accounted for when building queries for GMT summaries such as 12z-12z.

8-(st)
State - Two-letter postal abbreviation (PR=Puerto Rico, VI=Virgin Islands)

9-(stf)
State FIPS number (Note some Puerto Rico codes are incorrect.)

10-(stn)
State number - number of this tornado, in this state, in this year: May not be sequential in some years. Note: discontinued in 2008. This number can be calculated in a spreadsheet by sorting and after accounting for border crossing tornadoes and 4+ county segments.

11-(f), or (sz), or (mag)
F-scale (EF-scale after Jan. 2007): values -9, 0, 1, 2, 3, 4, 5 (-9=unknown). Or, hail size in inches. Or, wind speed in knots (1 knot=1.15 mph).

12-(inj)
Injuries - when summing for state totals use sn=1, not sg=1 (see below)

13-(fat)
Fatalities - when summing for state total use sn=1, not sg=1 (see below)

14-(loss)
Estimated property loss information - Prior to 1996 this is a categorization of tornado damage by dollar amount (0 or blank-unknown; 1<$50, 2=$50-$500, 3=$500-$5,000, 4=$5,000-$50,000; 5=$50,000-$500,000, 6=$500,000-$5,000,000, 7=$5,000,000-$50,000,000, 8=$50,000,000-$500,000,000, 9=$5000,000,000.) When summing for state total use sn=1, not sg=1 (see below). From 1996, this is tornado property damage in millions of dollars. Note: this may change to whole dollar amounts in the future. Entry of 0 does not mean $0.

15-(closs)
Estimated crop loss in millions of dollars (started in 2007). Entry of 0 does not mean $0.

16-(slat)
Starting latitude in decimal degrees

17-(slon)
Staring longitude in decimal degrees

18-(elat)
Ending latitude in decimal degrees

19-(elon)
Ending longitude in decimal degrees

20-(len)
Length in miles

21-(wid)
Width in yards

22-(ns), **23-(sn)**, **24-(sg)**
Understanding these fields is critical to counting state tornadoes, totaling state fatalities/losses. The tornado segment information can be thought of as follows:

22- ns=Number of States affected by this tornado: 1, 2, or 3.
23- sn=State Number: 1 or 0 (1=entire track info in this state).
24- sg=Tornado SeGment number: 1, 2, or -9 (1=entire track info).

1,1,1 = Entire record for the track of the tornado (unless all 4 fips codes are non-zero).
1,0,-9 = Continuing county fips code information only from 1,1,1 record, above (same om).
2,0,1 = A two-state tornado (st=state of touchdown, other fields summarize entire track).
2,1,2 = First state segment for a two-state (2,0,1) tornado (state same as above, same om).
2,1,2 = Second state segment for two-state (2,0,1) tornado (state tracked into, same om).
2,0,-9 = Continuing county fips for a 2,1,2 record that exceeds 4 counties (same om).
3,0,1 = A three-state tornado (st=state of touchdown, other fields summarize entire track).
3,1,2 = First state segment for a three-state (3,0,1) tornado (state same as 3,0,1, same om).
3,1,2 = Second state segment for three-state (3,0,1) tornado (2nd state tracked into, same om as 3,0,1 record).
3,1,2 = Third state segment for a three-state (3,0,1) tornado (3rd state tracked into, same om as the initial 3,0,1 record).

25-(f1) 1st County FIPS code
26-(f2) 2nd County FIPS code
27-(f3) 3rd County FIPS code
28-(f4) 4th County FIPS code - Additional counties will be included in sg=-9 records with same om number (see description above).

29-(mt) WIND ONLY
Magnitude-type is only used for wind data. EG=estimated gust, MG=measured gust, MS=measured sustained, ES=estimated sustained (started in 2006).

Tornado database file updated to add "fc" field for estimated F-scale rating in 2016. Valid for records altered between 1950-1982. See next page for methodology used.

29-(fc) fc=0 for unaltered (E)F-scale rating. fc=1 if previous rating was -9 (unknown).

Between 1953 and 1982, 1864 CONUS tornadoes were coded in the official database with an F-scale rating of -9 (unknown).

The table below explains how these tornado records were modified to provide an estimated F-scale rating. All changed records are identified in the database by the "fc" field (fc=1 if the F-scale was changed from -9 to another value, fc=0, all unchanged F-scales).

IF property loss is equal to:	THEN set F-scale equal to:	IF path length <=5 miles add:	IF path length >5 miles add:
0,1 (<$50)	0	0	+1
2,3 (up to $5K)	1	-1	+1
4,5 (up to $500K)	2	-1	+1
6,7 (up to $50M)	3	-1	+1
8,9 (up to $5B)*	4	-1	+1

** No F=-9 tornado records met the 8,9 property loss criteria.*

Using the table above on unknown F-scale rated tornado records that included property loss and path length information resulted in the following breakdown of tornadoes ranked by estimated F-scale (percent of total F=-9 records is shown):

F0: 1038 tornadoes (55.5%)
F1: 742 tornadoes (40.1%)
F2: 26 tornadoes (1.3%)
F3: 52 tornadoes (2.7%)
F4: 6 tornadoes (0.3%)
F5: None

1864 F=-9 records were changed/modified between 1953 and 1982.