DATA EXPLORATION AND VISUALIZATION

The analysis of the dataset has been performed in R, Tableau and TIBCO Spotfire.

The code and the output snippets have been shown for reference and reader's clear understanding.

Figure 2 shows the distribution of all tornado reports by month. There is a marked peak in activity during May with the main season running from April through July. Here we focus on this month three- period. Of the 59,947 tornado reports in this region over the period 1950–2014, 66.08% of them occurred in April, May, June or July. We have classified these months as storm seasons and the other months as no-storm seasons.

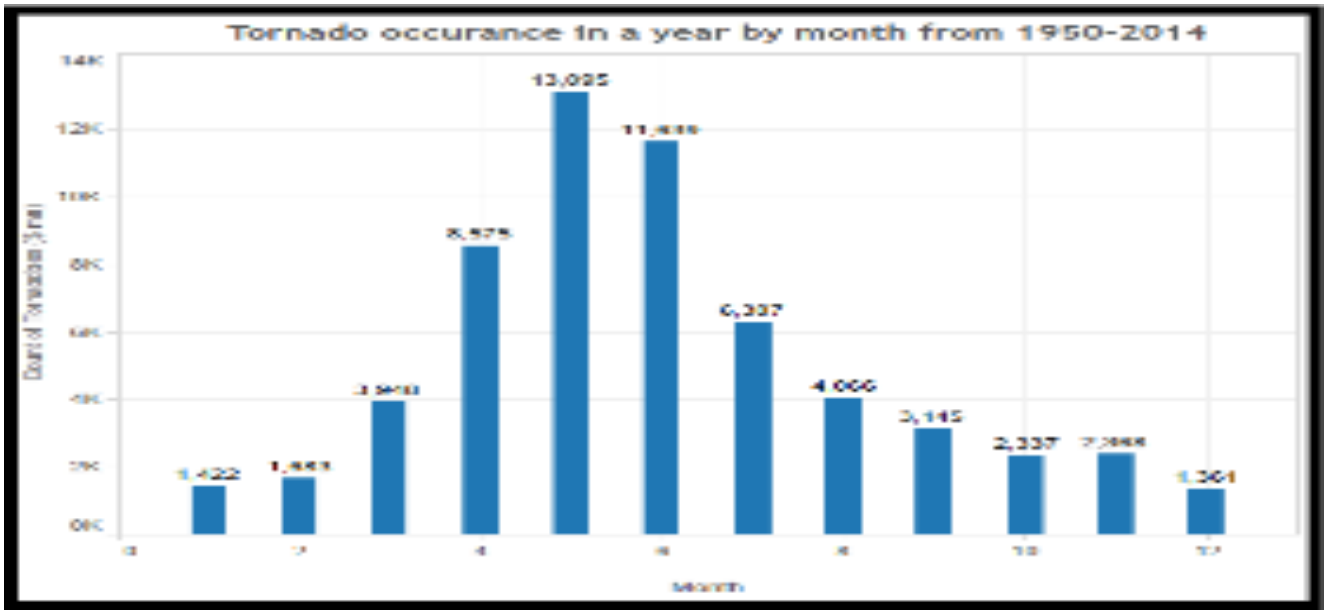

Figure 2.

Figure 3 shows the distribution of all tornado intensity reports by month. There is a marked peak in the intensity for the seasons of fall and spring. There is a significant dip in the intensity of the tornado in summer season.

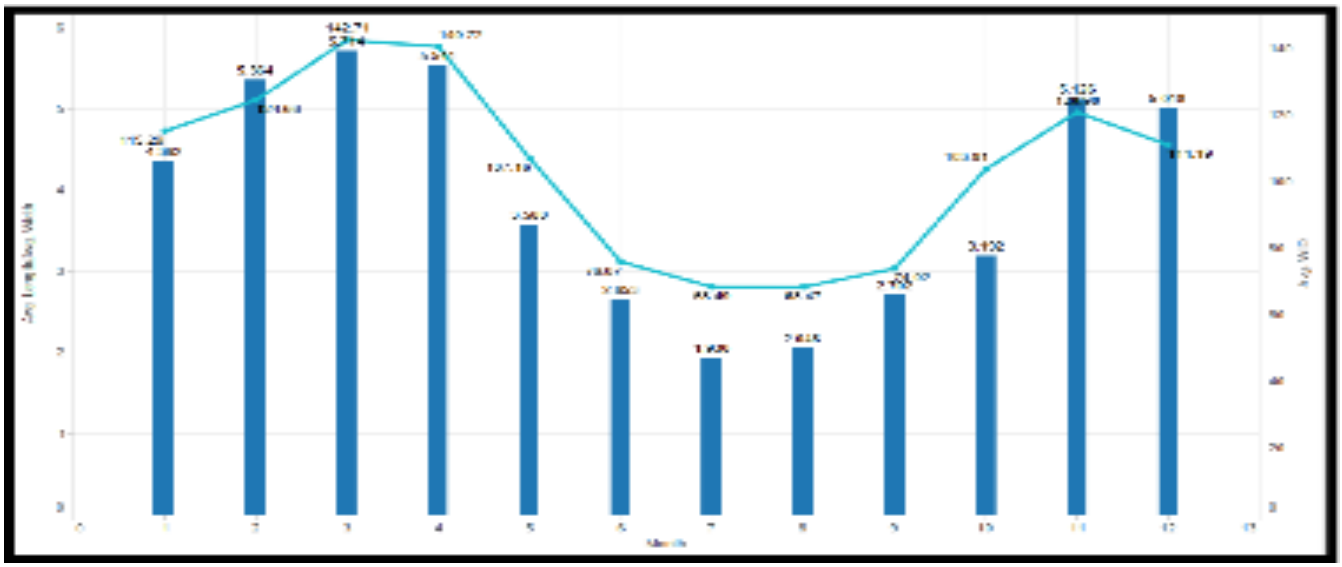

Figure 3.

Figure 4 shows the distribution of fatalities and injuries in the given US states during 1950 to 2014. The worst hit state being MA with an average count of 10,367 injuries and 637 fatalities. For some states such as DC, Porto Rico, Idaho the data is not available.

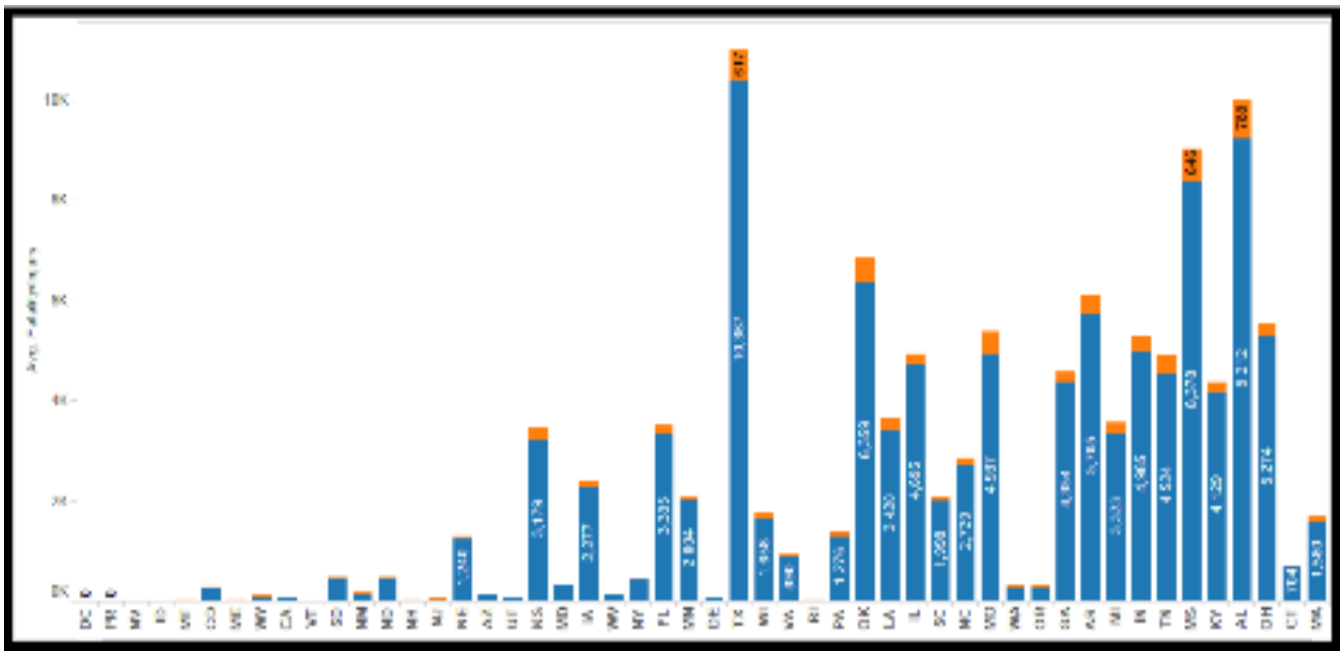

Figure 4.

Figure 5 shows the distribution of tornado count in the given US states over the period of 50 years. TX is the state hit by maximum tornadoes in the last 50 years with ~ 8200 tornadoes. It can be observed that the states which are hit by tornadoes badly, the per capita income has not been raised as much as compared to the other

states. The occurrence of natural calamities might also be contributing factor in lower rate of increase in the per capita.

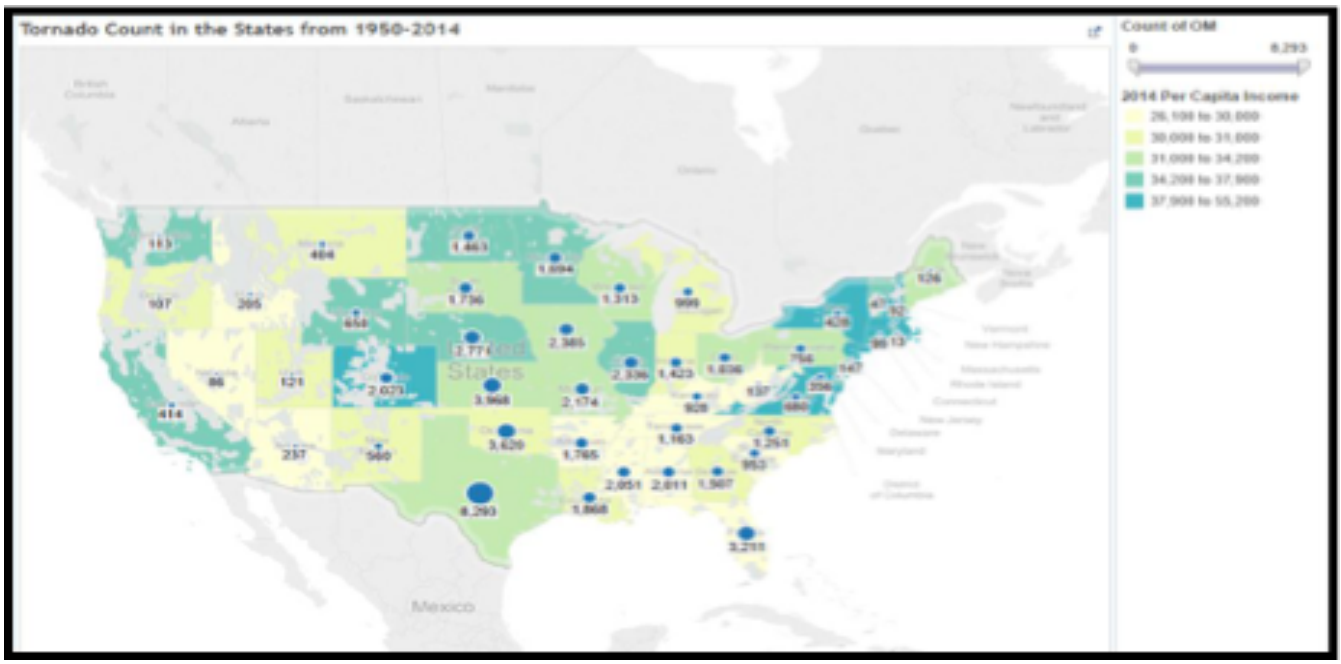

Figure 5.

Figure 6 shows the property loss incurred by various state in the United States of America in the last 50 years. It can be seen that the worst hit state has been the state of Texas and the maximum losses of ~ $13K have occurred in this state. There are state with very few losses as well such as WA, NV but it can be seen that even with very few tornado hits in these states the loss is comparatively higher than those states which are hit by tornadoes more often.

With the above observations we can infer that the most badly hit states are those in the central region thus making it prone to tornadoes.

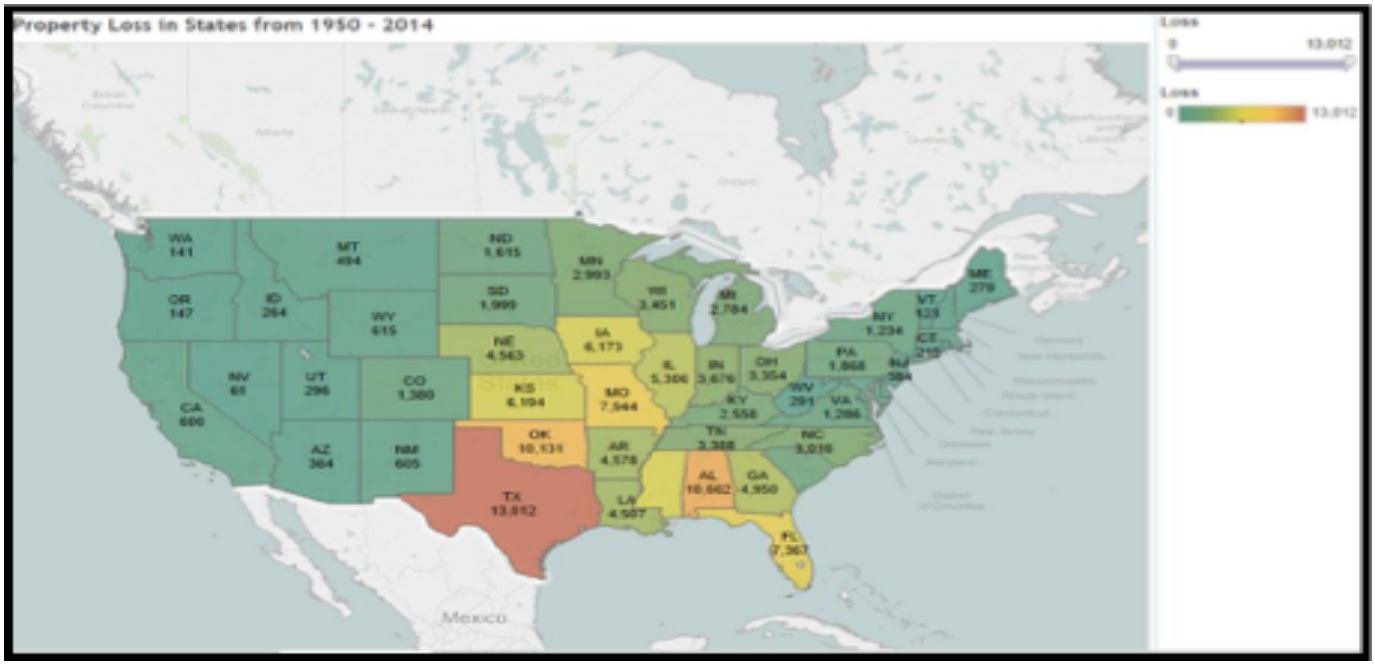

Figure 6.

The figure 7 below describes the loss trend present in the dataset. It can be seen that the crop loss has a lot of null values until 2005. A peek in the property and crop loss can be described by a major calamity in 2010 occurring once in several years.

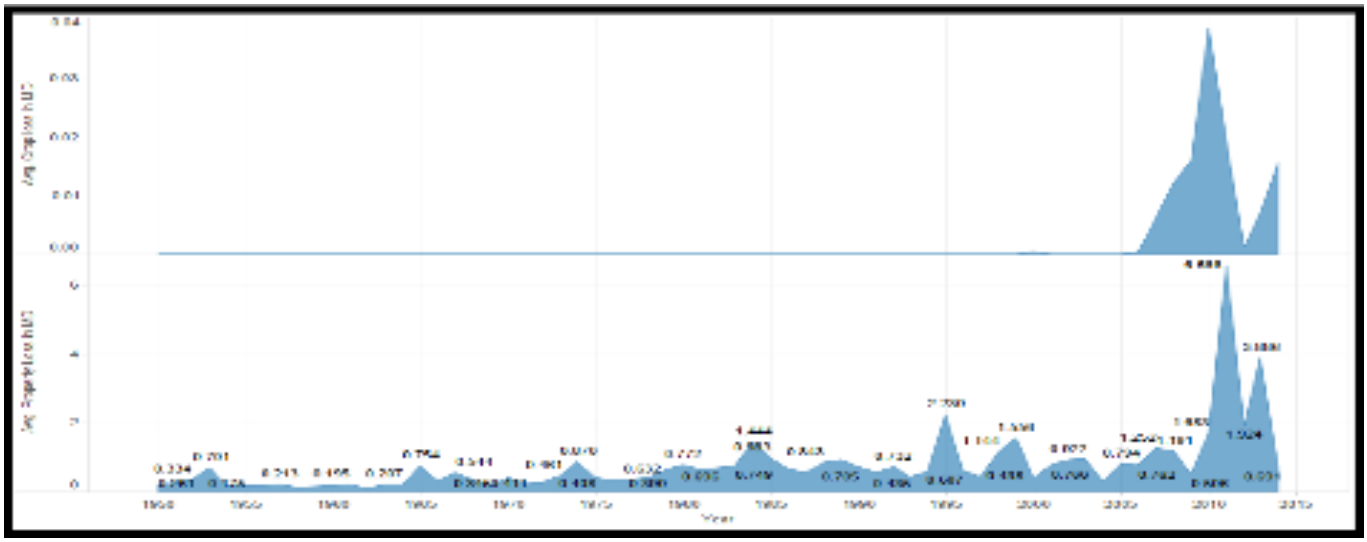

Figure 7.

Figure 8 above shows the distribution of property losses per state in last 50 years. This has been distributed in the order of state severity where the severity is defined as the amount of losses that the state has incurred due to the tornado activity in the

last 50 years of available data in the dataset. Thus we have classified states broadly in four state severity categories.

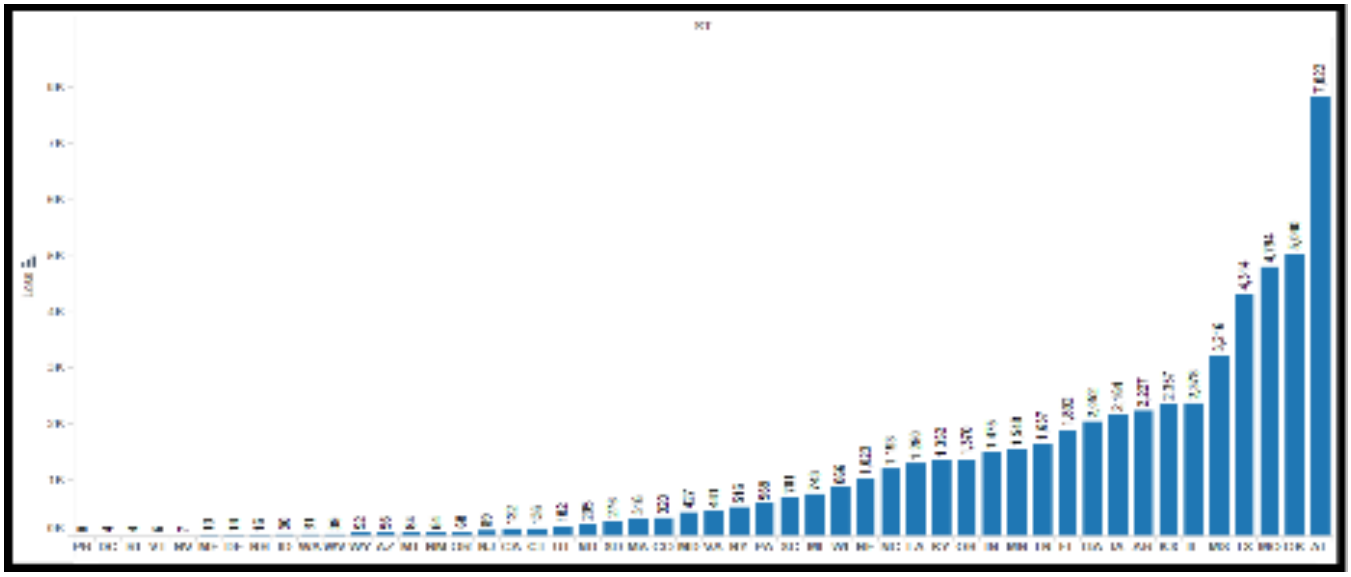

Figure 8.

Looking at our data, there was a significantly higher frequency of tornados in April through July, and significantly less in other months. This was a consistent pattern, which was similarly present since 1950. Thus, Storm Season is a binary variable indicating whether the tornado occurred during tornado season or not. The State variable was consolidated into four categories based on the percentage of all tornados that hit the state. For example, Texas contains roughly 9% of all tornados that occurred in the United States for our dataset. Given this, we categorized the states accordingly into the State Severity variable.

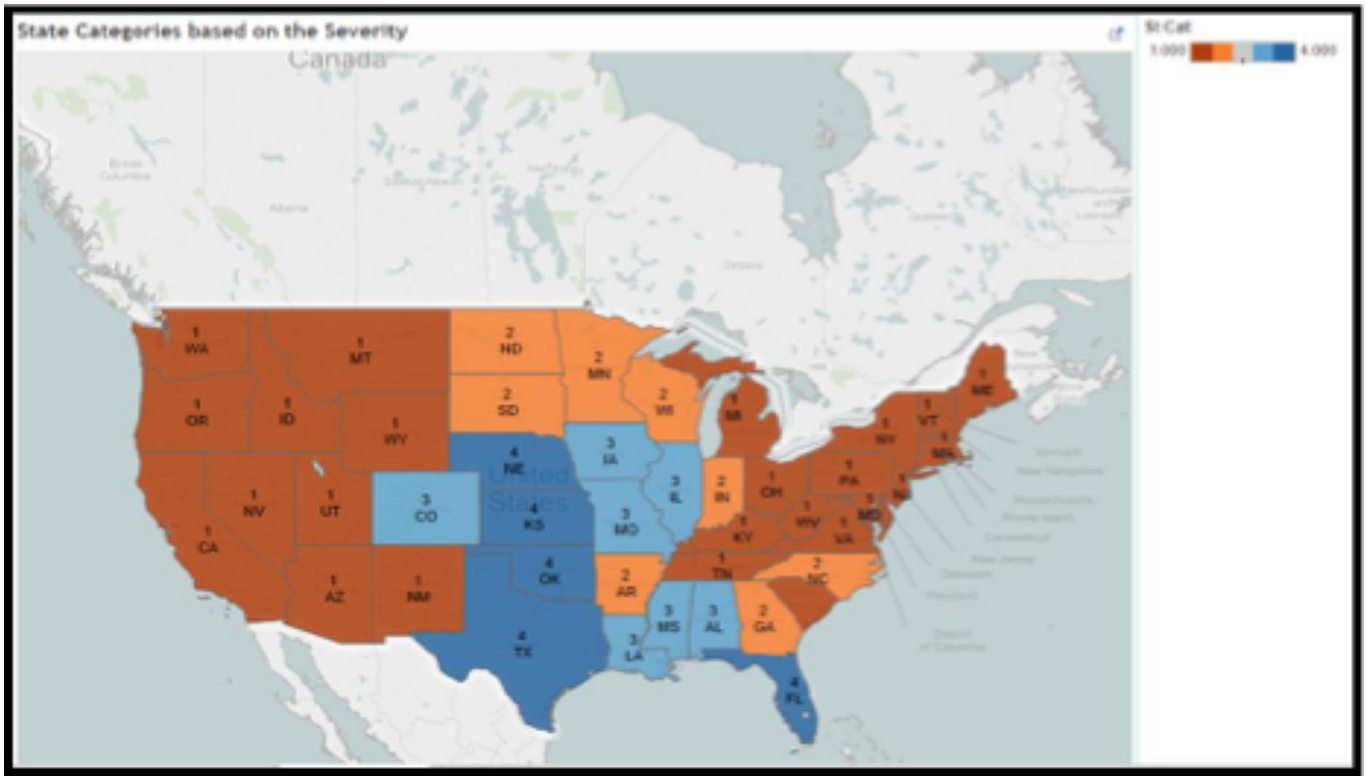

Figure 9.

The State variable was consolidated into four categories based on the percentage of all tornados that hit the state (see figure 8). For example, Texas contains roughly 13.8% of all tornados that occurred in the United States for our dataset. Given this, we categorized the states accordingly into the State Severity variable. Figure 9 shows a color-coded map of the categorized states. Another categorical predictor is the Fujita scale (F-Scale). This is based on an aerial survey of the affected area that gives a categorical estimate of the severity of damage. The scale is measured in increasing severity from zero to five.