Kennesaw State University

College of Computing and Software Engineering

Department of Computer Science

CS 4412 (01) Data Mining – (Fall 2018)

Project Report

Group 3

Alexander Urbanyak

Gene Vakhroushev

Jason Walters

November 14, 2018

Study of Storm Events and Tornadoes in the

United States from 1950 to 2018

# Abstract

The main objective of this study is to find efficient classification methods of data mining. Our goal is to take data provided by the National Oceanic and Atmospheric Administration (NOAA), formally the National Climatic Data Center (NCDC) and determine important information regarding tornadoes. Our team will use data collected in the United States over the past 68 years, from 1950 to 2018. We hope to discover the number of tornadoes in each city, damage caused by tornadoes in each city, number of tornadoes in each state, damage caused by tornadoes in each state, the number of tornadoes that started within each hour, and finally tornado damage within each hour.

# Introduction

This project is an example of a classification problem. The database contains information of different storm types, some attributes only apply to certain types of storms. The data contains information on various types of storms and includes injuries, deaths, and estimated property damage in the United States from 1950 to 2018. For our experiment we will be looking at only storm events classified as tornadoes. Our team will use High-Performance Computing Cluster (HPCC) Systems to analyze the data and HPCC Visualizer[[1]](#endnote-1) bundle to display the results.

Our purpose is to take data provided by the NOAA and find damage caused by tornadoes per hour, what time of day tornadoes most often occur, what states have the most tornadoes, and what city do most tornadoes happen. All damage estimates are in U.S. dollars over the last 68 years.

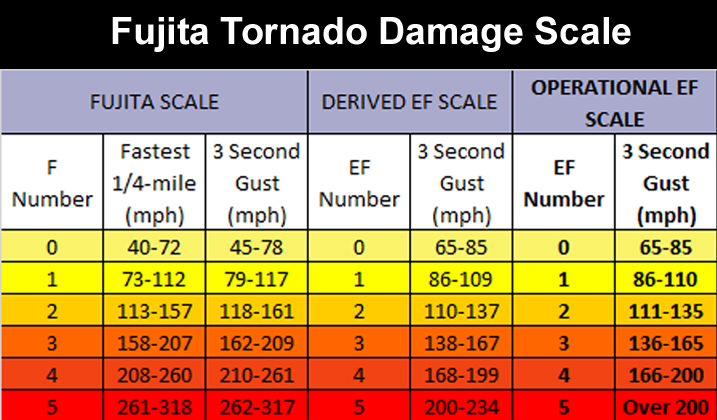


Figure 1. Fujita Scale for Tornadoes

According to the National Severe Storms Laboratory (NSSL) “A tornado is a narrow, violently rotating column of air that extends from the base of a thunderstorm to the ground. Because wind is invisible, it is hard to see a tornado unless it forms a condensation funnel made up of water droplets, dust and debris. Tornadoes are the most violent of all atmospheric storms.”[[2]](#endnote-2) Tornadoes are classified on a scale established by Dr. Fujita, a severe storms research scientist from the University of Chicago. The scale ranges from F0 to F5, where F0 is the weakest and F5 is classified as the strongest.

# Experiment

Our team took storm data provided by the National Oceanic and Atmospheric Administration (NOAA) in the United States from 1950 to 2018. We first retrieved a data set that included statistical data on all storms that occurred in the United States. The data was preprocessed by the NOAA. The original dataset contained 1,514,332 records of storm events in U.S. and U.S. territories. Two more tables were created from this dataset, one including records only for the tornado event type, which contained 68,587 records. Another table was created for all tornado records that had a specified starting location, which contained 33,531 records. Our preliminary purpose of this experiment was to find a method to calculate damage from a tornado based on starting location. This calculation would allow local, state, and federal governments a faster turnaround time to distribute aid and start other humanitarian actions.

# Results

All results are based on the data provided by the National Oceanic and Atmospheric Administration (NOAA), from 1950 to 2018. Figure 2 and 3 display results for all storm event types for all states and U.S. territories. Storm event types are tornadoes, flooding, strong wind, blizzard, winter storm, hail, extreme temperature, and many more. Figure 4 and 5 use only the data of tornado events with specified starting location. Figure 6 and 7 use only the data of tornado events even if the staring location of a tornado is not specified, the results were filtered to only contain valid states and U.S. territories. Figure 8 and 9 use the same data set as figures 6 and 7, the start time of tornadoes and damaged caused at a specific time were round down to the nearest hour.

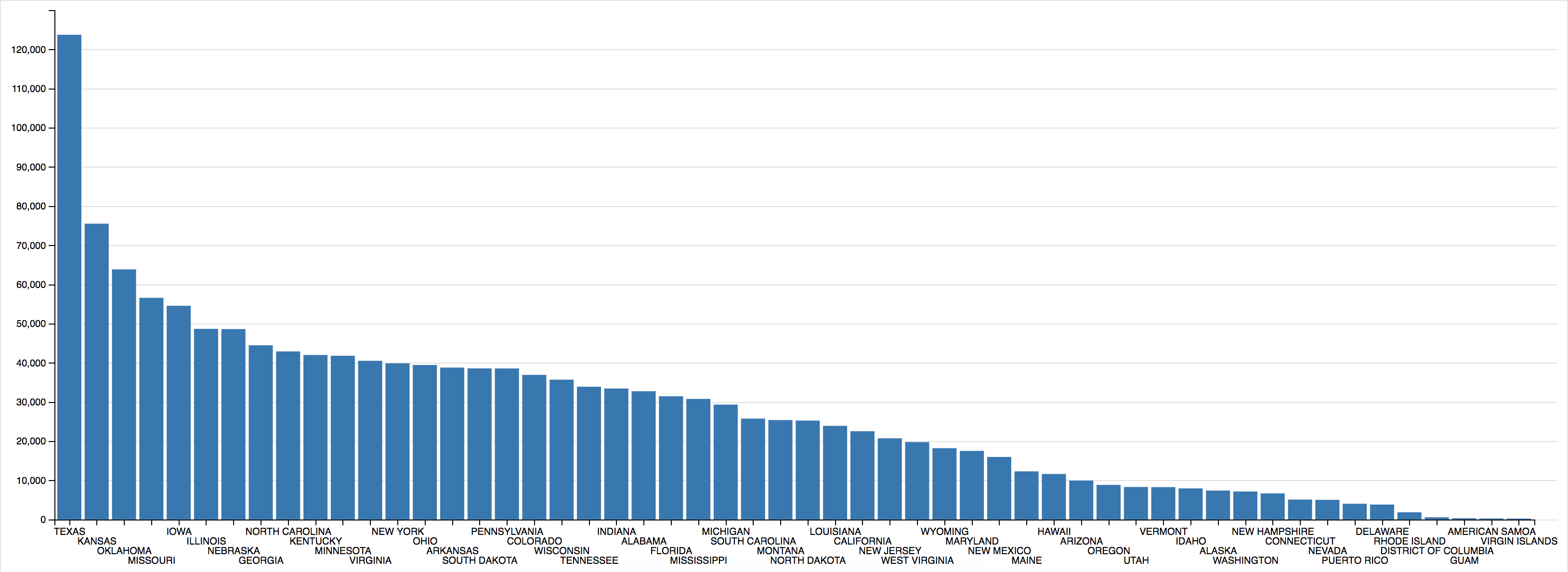


Figure 2. Number of storms for each state, inhabited U.S. territories, and the District of Columbia from 1950 to July 2018



Figure 3. Damage of all storms for each state, inhabited U.S. territories, and the District of Columbia from 1950 to July 2018

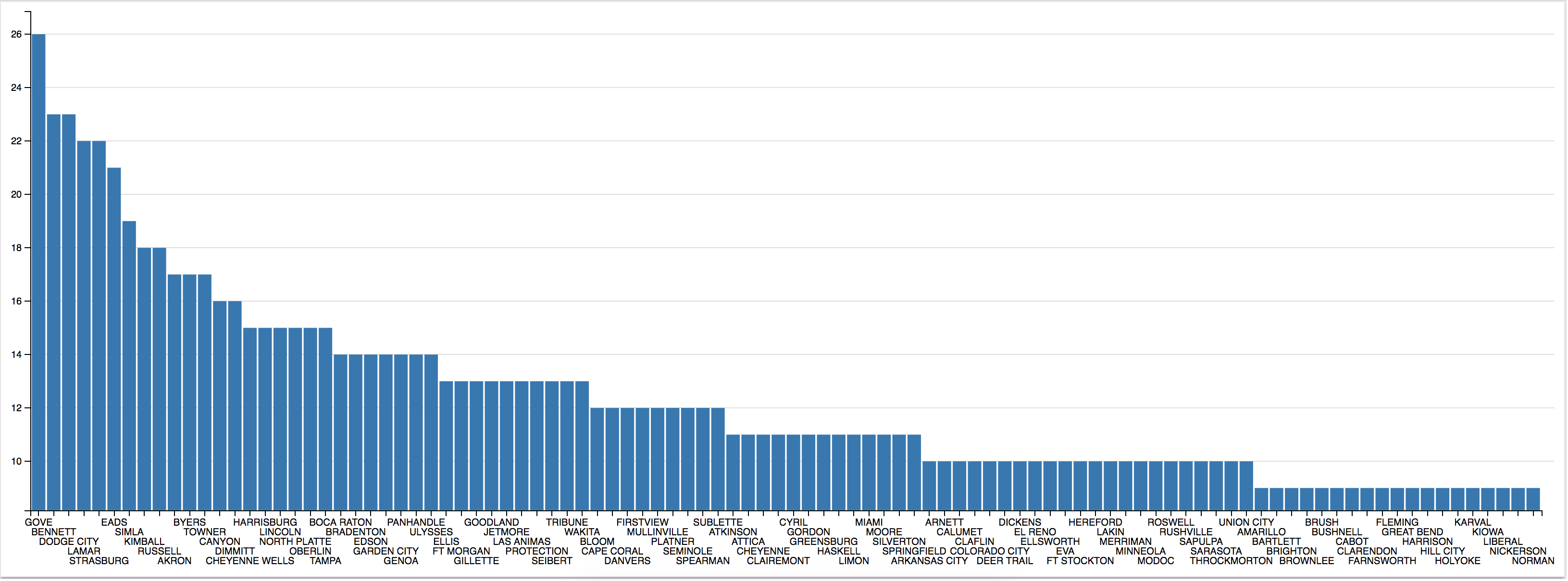


Figure 4. Number of tornadoes for each city. Top 100 U.S. cities with most occurrences from 1950 to July 2018

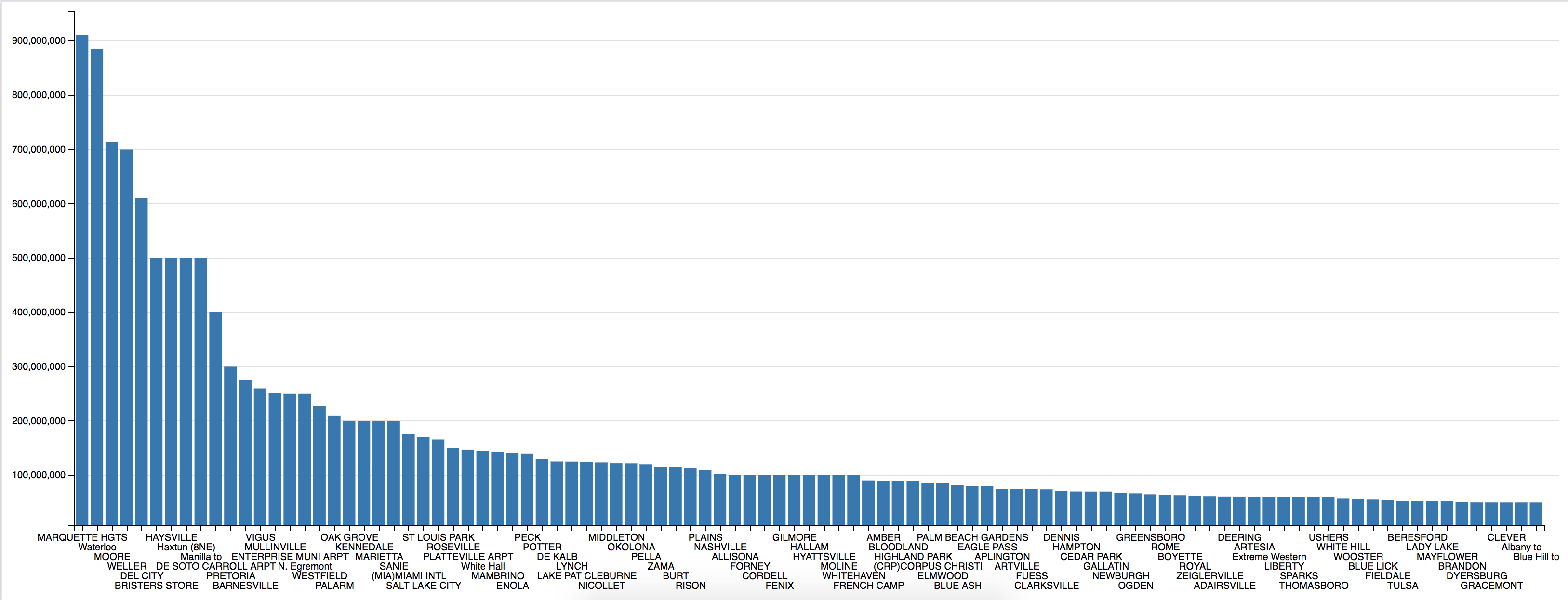


Figure 5. Tornado damage in USD for each city. Top 100 U.S. cities with most damage incurred from 1950 to July 2018

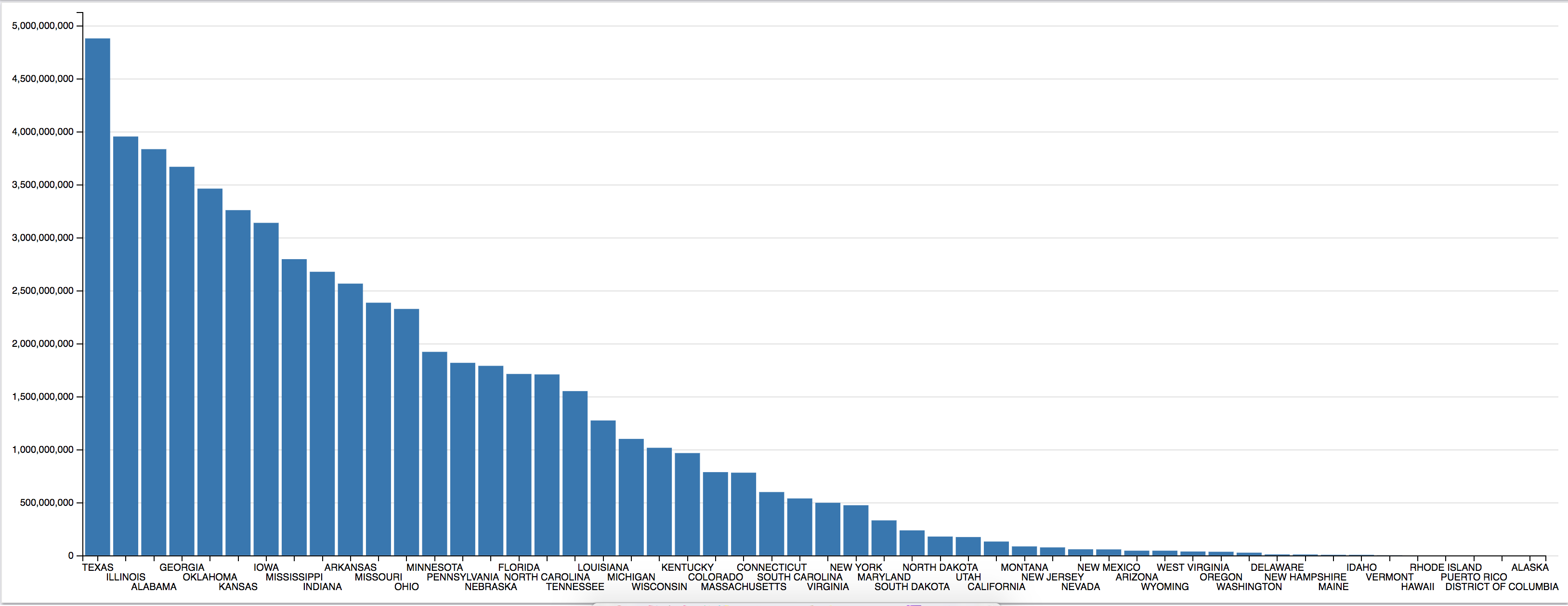


Figure 7. Tornado damage in USD for each state and U.S. territory from 1950 to July 2018

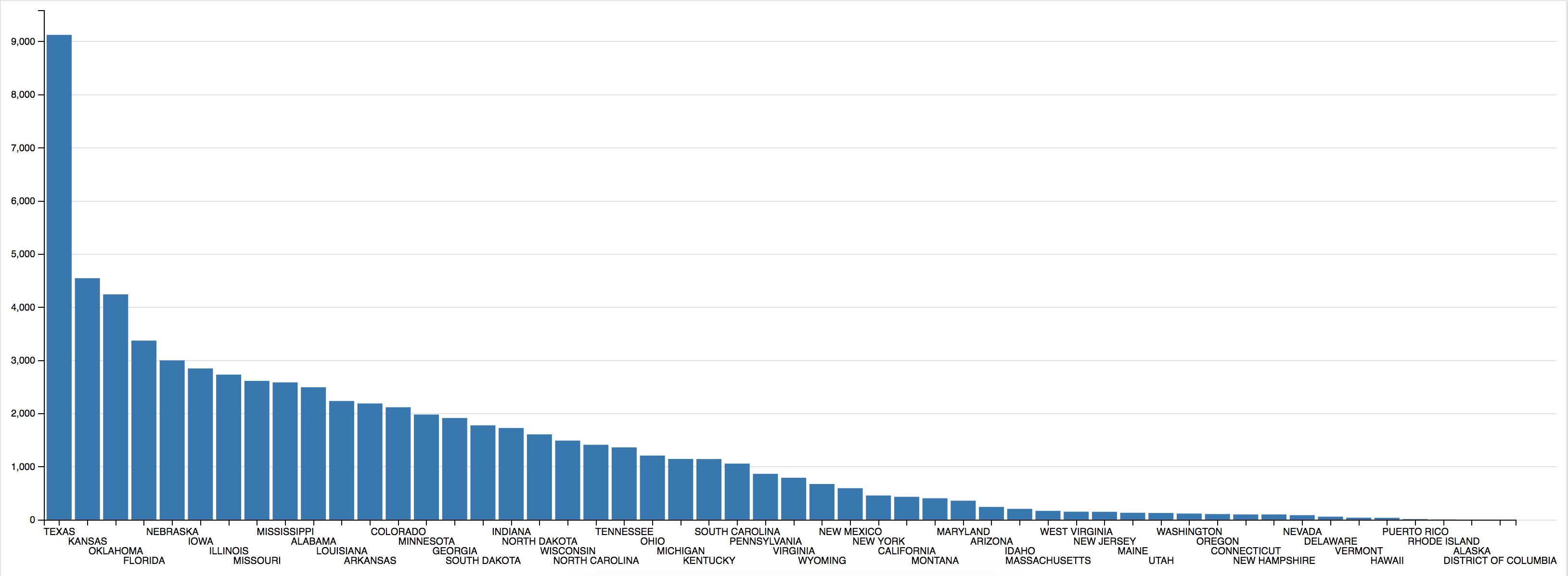


Figure 6. Number of tornadoes for each state and U.S. territory from 1950 to July 2018



Figure 8. Number of tornadoes that started within each hour for all states and U.S. territories from 1950 to July 2018 in military time

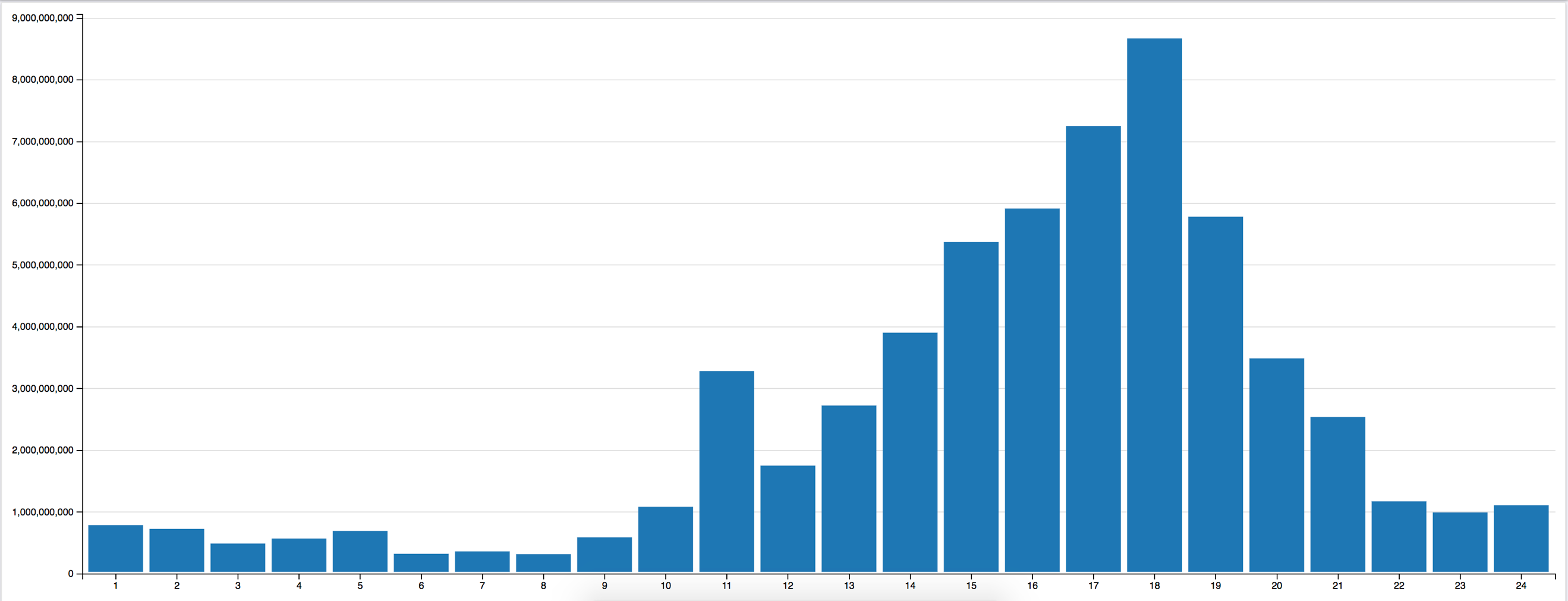


Figure 9. Tornado damage caused in USD within each hour for all states and U.S. territories from 1950 to July 2018 in military time

# Conclusion

According to the data the state of Georgia leads all states, inhabited U.S. territories, and the District of Columbia with the total damage caused by any storm type. From 1950 to 2018 Georgia had just under $60 billion in damage and Texas is next with $35 billion. However, when we refined the search and only looked at tornadoes Texas lead in occurrences with just over 9,000 and Georgia, which lead in all storms fell to 15th with just under 2,000. When looking at damage caused just by tornadoes Texas leads with just under 5 billion, followed by Illinois, Alabama, and Georgia. According to the data most tornadoes start at 1700 (5 P.M.) and the hour with the most damage is 1800 (6 P.M.). Begin/end times and locations, property damaged caused by the storms are all estimates, which means results presented are not 100% accurate but given the amount of data used the estimates provide reasonably accurate trends and statistics.

1. "Hpcc-systems/Visualizer." Hpcc-systems. GitHub. October 18, 2018. Accessed November 10, 2018. https://github.com/hpcc-systems/Visualizer. [↑](#endnote-ref-1)
2. Tornado Basics." NOAA National Severe Storms Laboratory. Accessed November 14, 2018. https://www.nssl.noaa.gov/education/svrwx101/tornadoes/. [↑](#endnote-ref-2)