Individual Project Task 3 - Final Report
Laura Zhang
4/29/2021

This project involves the Severe Weather Database Files (1950-2019) of the U.S. National Oceanic and Atmospheric Administration (NOAA).

Synopsis

Because of the coronavirus disease 2019 (COVID-19) pandemic, the preparation for the hurricane season becomes more important to protect people from COVID-19[1]. Under- or overrating a tornado can bring unnecessary injuries and property loss, or panic.

The Fujita-scale (F-scale) was implemented nationally for rating tornado intensity[2]. Wind speed categories were established[3], and for each range of wind speed, the damage is described from light (F0), moderate (F1) to incredible (F5). However, the scale left little room for strength of construction that might cause buildings to sustain greater damage at lower wind speeds[2]. According to the SPC Severe Database Description, since January 1, 2007, the F-scale was replaced by Enhanced Fujita Scale (EF-scale).

Same as the F-scale, EF-scale has six strength categories from EF0 to EF5, representing increasing degrees of damage. However, the new scale considers construction quality and standardizes various structures. Wind speed levels assigned to damage does differ between the two scales. A tornado with a rating of F5 on the F-scale may be classified as EF4 or even lower following the EF-scale[4,5].

To figure out if the change of tornado rating standard had an impact on the total number of tornadoes classified as a specific level, we use Severe Weather Database Files (1950-2019) to analysis tornado data of 1950-2006 (F-scale era) and 2007-2019 (EF-scale era) separately. We will test the hypothesis that EF-scale leads to a decrease in the number of tornadoes classified as EF3 and higher. Null hypothesis significance testing will be used, for it can help us decide whether we should support or reject the hypothesis.

Data Processing

Loading the Tornado Data from compressed comma separated file.

```
tornado_data <- read.csv(file="D:/1950-2019_all_tornadoes.csv",
header=T, sep=",")
head(tornado_data)
```

	tornado_number	year	month	day	date	time	time_zone	state	state_FIPS_number	state_number	F_scale	injuries
1	1	1950	1	3	1950-01-03	11:00:00	3	MO	29	1	3	3
2	1	1950	1	3	1950-01-03	11:00:00	3	MO	29	1	3	3
3	1	1950	1	3	1950-01-03	11:10:00	3	IL	17	1	3	0
4	2	1950	1	3	1950-01-03	11:55:00	3	IL	17	2	3	3
5	3	1950	1	3	1950-01-03	16:00:00	3	OH	39	1	1	1
6	4	1950	1	13	1950-01-13	05:25:00	3	AR	5	1	3	1

	fatalities	loss	crop_loss	starting_latitude	starting_longitude	ending_latitude	ending_longitude	length_in_miles
1	0	6	0	38.77	-90.22	38.83	-90.03	9.5
2	0	6	0	38.77	-90.22	38.82	-90.12	6.2
3	0	5	0	38.82	-90.12	38.83	-90.03	3.3
4	0	5	0	39.10	-89.30	39.12	-89.23	3.6
5	0	4	0	40.88	-84.58	0.00	0.00	0.1
6	1	3	0	34.40	-94.37	0.00	0.00	0.6

	width_in_yards	number_of_states_affected_by_this_tornado	state_number.1	tornado_segment_number	X1st_county_FIPS_code
1	150	2	0	1	0
2	150	2	1	2	189
3	100	2	1	2	119
4	130	1	1	1	135
5	10	1	1	1	161
6	17	1	1	1	113

	X2nd_county_FIPS_code	X3rd_county_FIPS_code	X4th_county_FIPS_code	F_scale_rating
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0

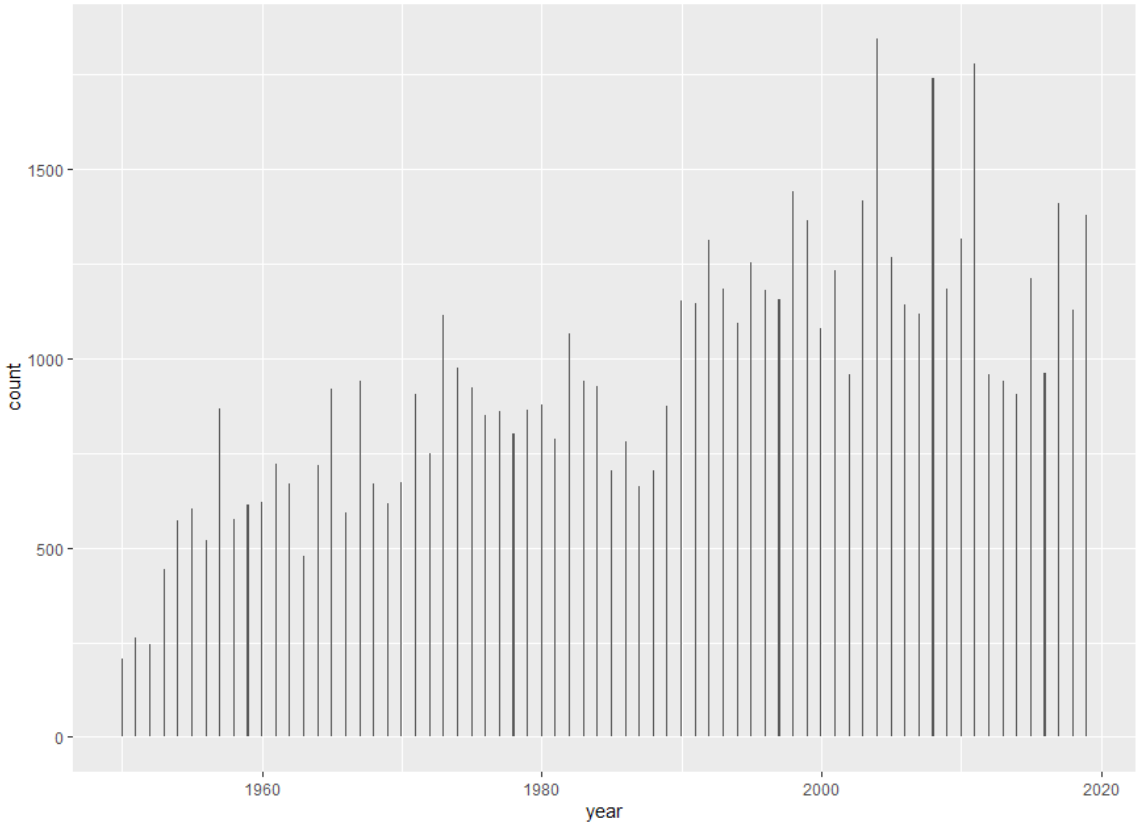
According to the SPC Severe Database Description, F-scale of -9 means unknown, that is, the intensity cannot be determined due to a lack of information. Rows whose F-scale equal -9 should be removed to filter data with available F-scale records.

```
known_F_scale <- subset(tornado_data, F_scale>=0)
```

Visualization

To test our hypothesis that EF-scale leads to a decrease in the number of tornadoes classified as EF3 and higher, we need to prove that the total number of tornadoes did not go down.

```
ggplot(known_F_scale, aes(x=year))+geom_histogram(binwidth=0.1)
```



Generally, total cases after 2007 were not less than years before 2007. That is, if there was, indeed, a decrease in the number of tornadoes classified as EF3 and higher, the reason should not be any decrease of total tornado cases.

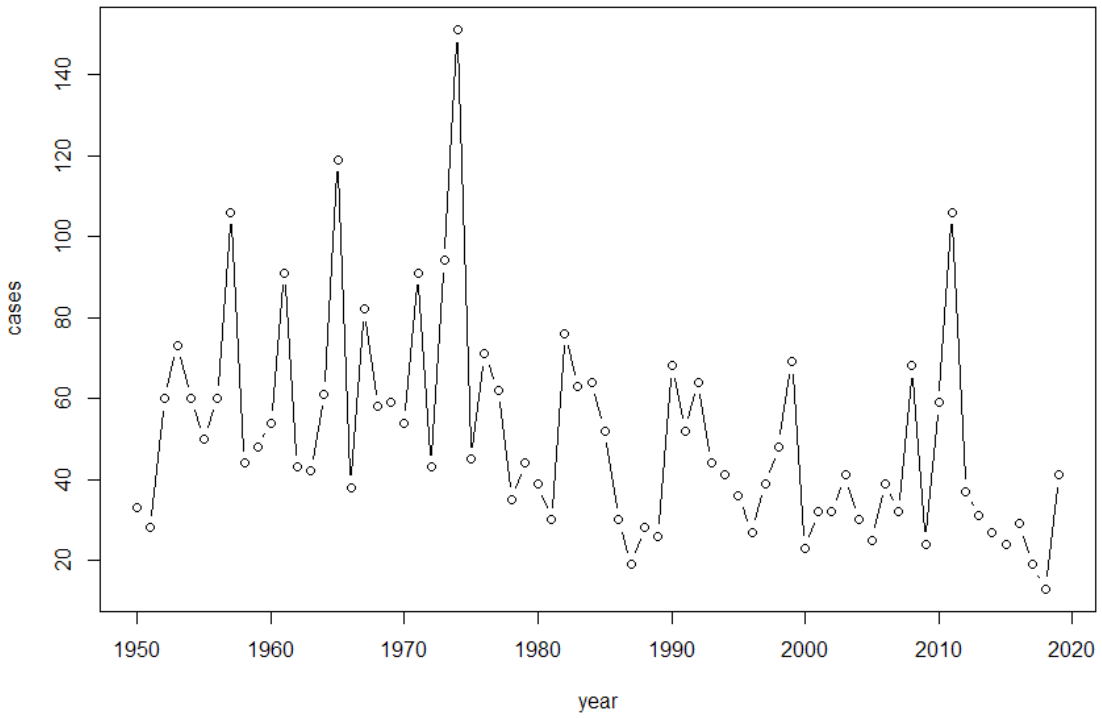
Next, we select data with F-scale equal to or greater than 3.

```
F_scale_3 <- subset(tornado_data, F_scale>=3)
```

Then get the total number of tornadoes with F-scale equal to or greater than 3 in each year. For example, in 2010, the number of tornadoes with F-scale equal to or greater than 3 should be:

```
sum(F_scale_3$year=="2010")
```

Push these numbers into a data frame. Plot the year-cases relationship.



There is a descending trend after 2007, but we need a strong evidence.

Hypothesis testing
Null hypothesis: The new EF-scale did not lead to a decrease in the number of tornadoes classified as EF3 and higher.
Alternative hypothesis: The new EF-scale lead to a decrease in the number of tornadoes classified as EF3 and higher.
Assumptions: The wind speed over the years has been relatively constant.
Test method: Null hypothesis significance testing.
Significance level: 0.05
Mean: 48.45
Null: 33.66
n: 13
Standard Deviation: 24.21
Standard Error: 6.71
T-score: 2.20
P-value: `2*(1-pt(2.201521, 12))` = 0.04800562
P-value is smaller than 0.05, the null hypothesis should be rejected. The new EF-scale did lead to a decrease in the number of tornadoes classified as EF3 and higher.

Discussion

There are limitations in this test, such as the value of n is too small (in fact, it has been only 14 years since EF-scale was implemented in the U.S., so we lack data indeed), F-scale or EF-scale of many cases are not recorded, etc.

In the future, we will focus on the assessment of EF-scale, and whether the new scale has brought changes to the analysis of tornadoes.

References

[1]https://www.cdc.gov/disasters/covid-19/docs/318076A_Hurricane-key-Messages-COVID-19_HTML.pdf

[2]https://en.wikipedia.org/wiki/Fujita_scale

[3]McDonald, J. R. (2001). T. Theodore Fujita: His contribution to tornado knowledge through damage documentation and the Fujita scale. Bulletin of the American Meteorological Society, 82(1), 63-72.

[4]https://en.wikipedia.org/wiki/Enhanced_Fujita_scale

[5]Doswell III, C. A., Brooks, H. E., & Dotzek, N. (2009). On the implementation of the enhanced Fujita scale in the USA. Atmospheric Research, 93(1-3), 554-563.