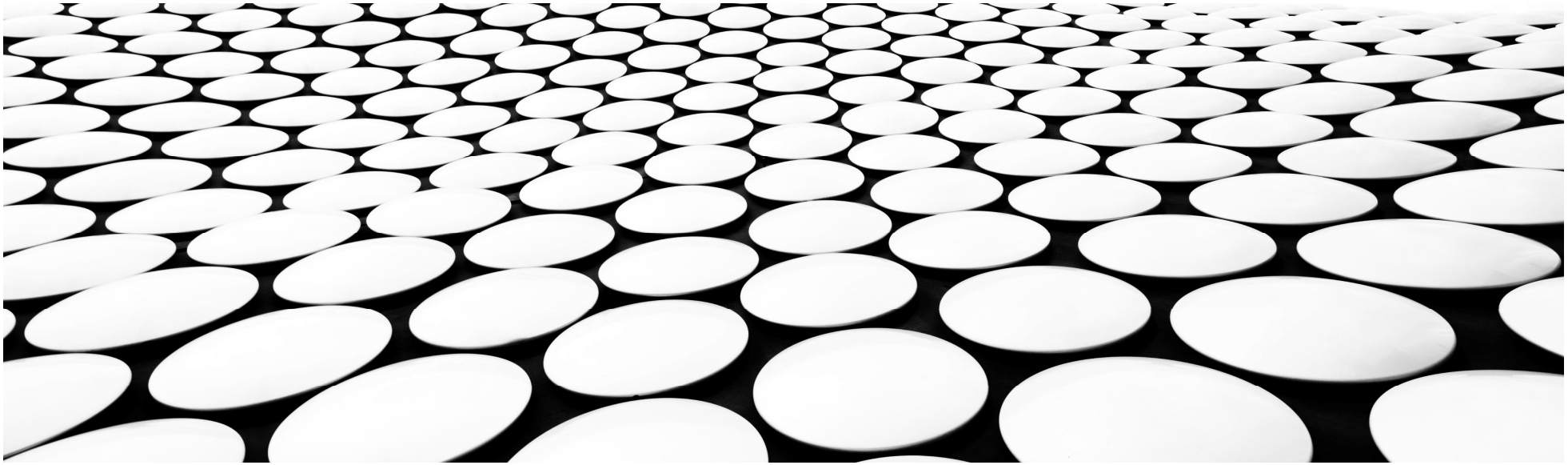
DSC 530 FINAL PROJECT –TORNADOES AMONG US!

ANTHONY WILSON



HAVE TORNADOES COST MORE IN PROPERTY DAMAGE OVER THE LAST TEN YEAR?

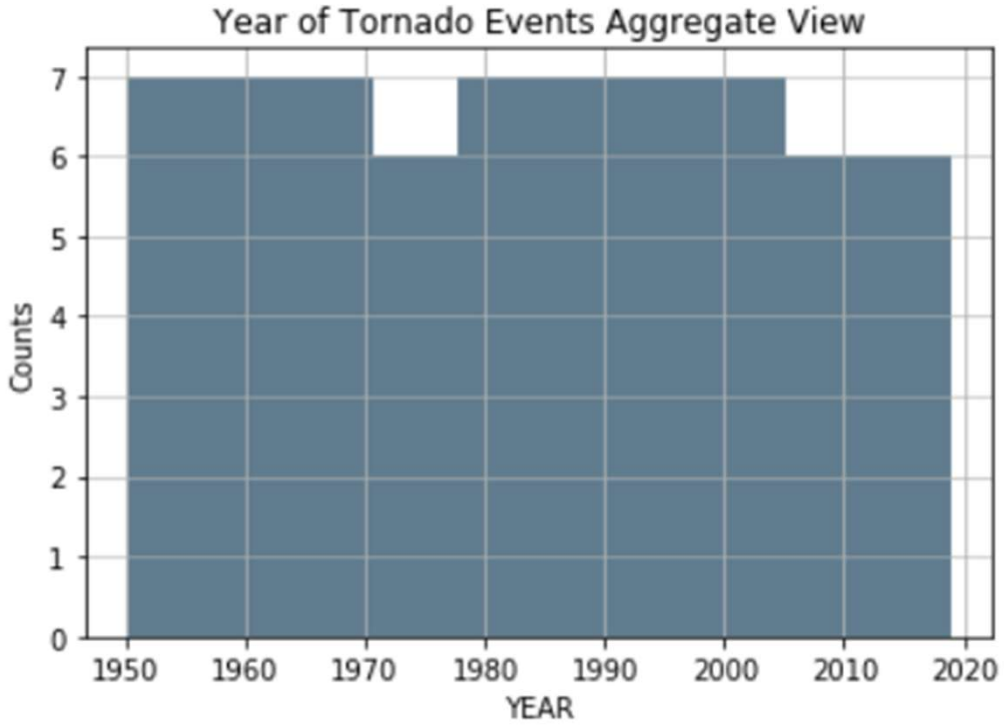
HYPOTHESIS/QUESTION

VARIABLES

- YEAR – Four-digit number for year the tornado happened (Ex. 2019)
- DAMAGE_PROPERTY – Estimated property damage in dollars
- DAMAGE_BUCKET – Binned estimation of Damaged Property
  - '<= 200000','<= 400000','<= 600000','<= 800000','<= 1000000','<= 1200000','<= 1400000','<= 1600000','<= 1800000','<= 2000000','<= 2200000','<= 2400000','<= 2600000'
- TOR_F_SCALE – Fujita scale that measures the strength of the tornado, based on the amount of damage caused by tornado
  - EF0 – Light Damage (40 – 72 mph),
  - EF1 – Moderate Damage (73 – 112 mph)
  - EF2 – Significant damage (113 – 157 mph)
  - EF3 – Severe Damage (158 – 206 mph)
  - EF4 – Devastating Damage (207 – 260 mph)
  - EF5 – Incredible Damage (261 – 318 mph)a scale of tornado storm severity
  - EFU/NA – Missing/unidentified
- TOR_LENGTH – Length of tornado in miles, when it touched the ground
- TOR_WIDTH – Width of the tornado in feet while on the ground

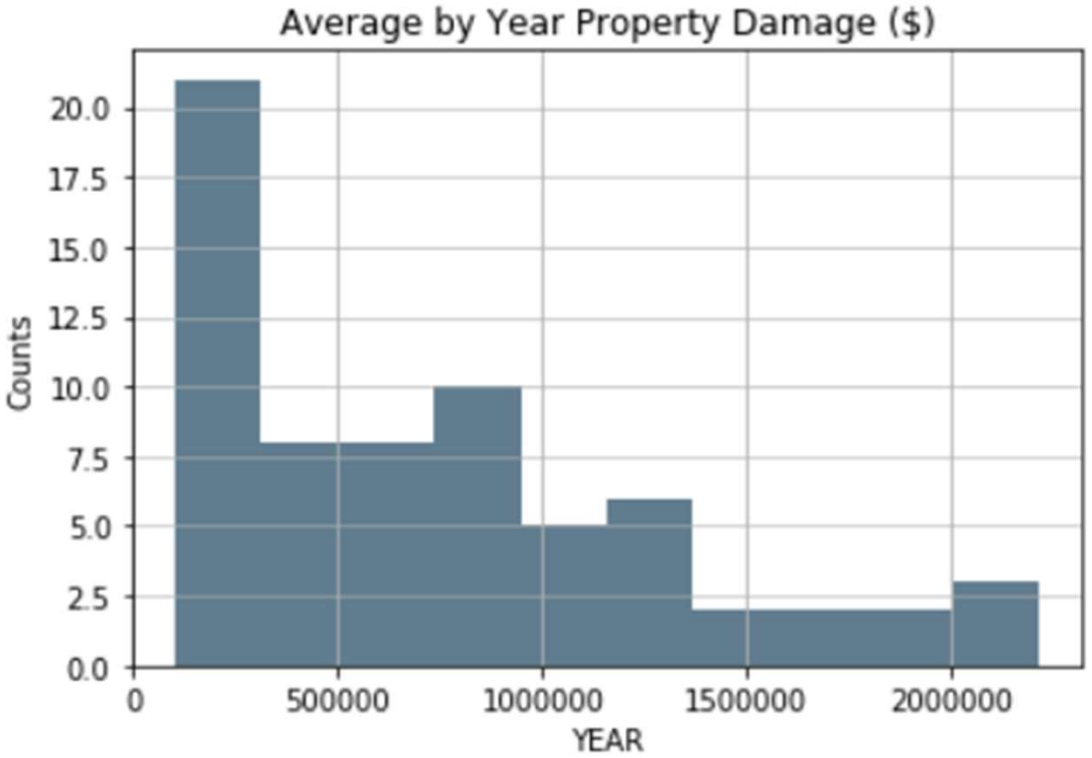
YEAR

- count: 67
- mean: 1983.835821
- std: 20.197207
- min: 1950
- 0.25: 1966.5
- 0.5: 1984
- 0.75: 2000.5
- max: 2019



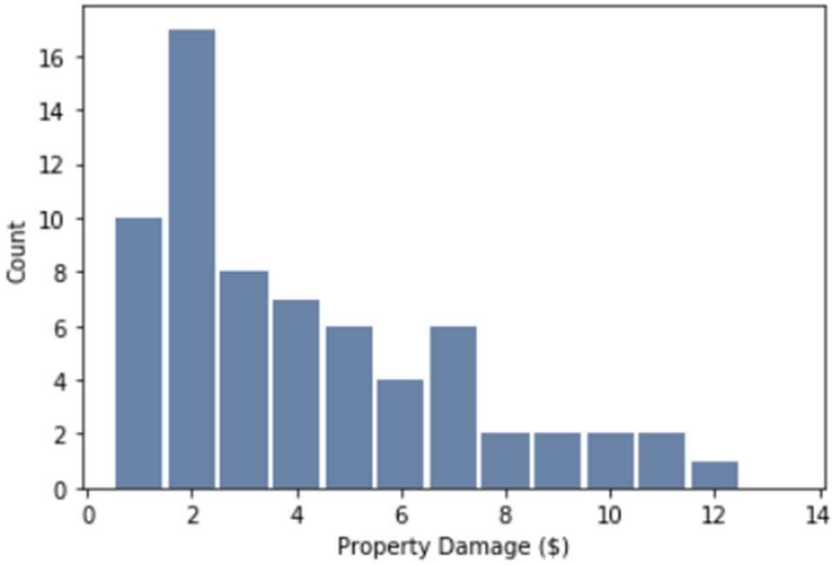
DAMAGE_PROPERTY

- count: 67
- mean: 740082.16
- std: 567117.99
- min: 99852.5
- 0.25: 242397.53
- 0.5: 589639.81
- 0.75: 1058827.41
- max: 2214829.58



DAMAGE_BUCKET

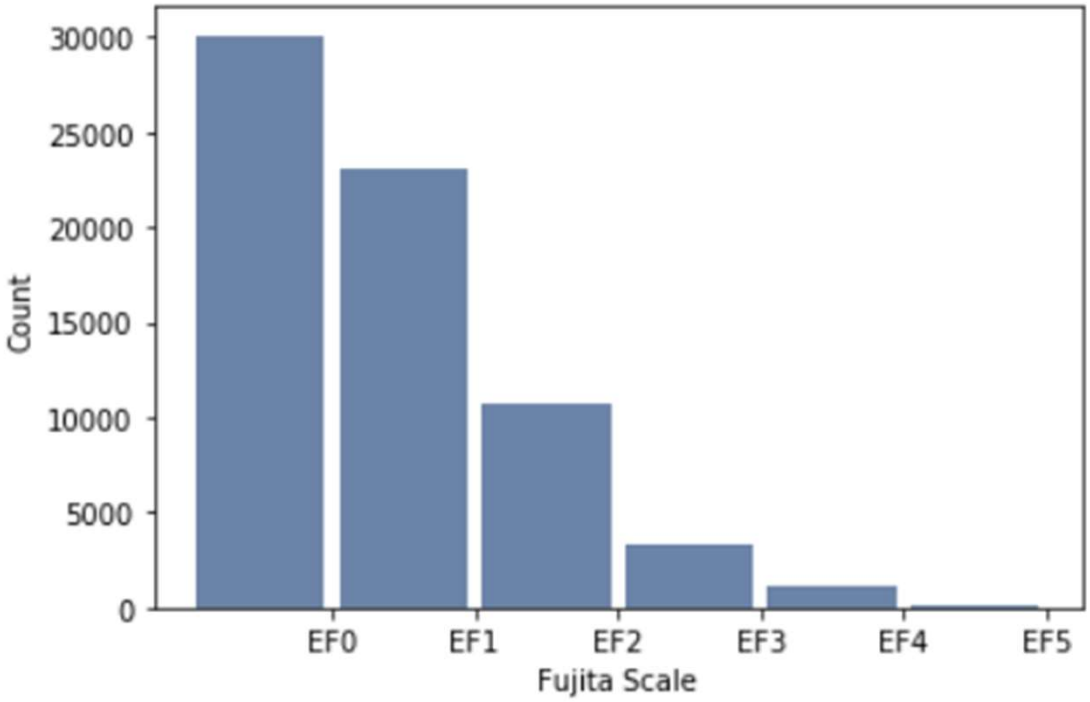
- count: 67
- unique: 12
- top: 2
- freq: 17



id:1 bucket: <= 200000
id:2 bucket: <= 400000
id:3 bucket: <= 600000
id:4 bucket: <= 800000
id:5 bucket: <= 1000000
id:6 bucket: <= 1200000
id:7 bucket: <= 1400000
id:8 bucket: <= 1600000
id:9 bucket: <= 1800000
id:10 bucket: <= 2000000
id:11 bucket: <= 2200000
id:12 bucket: <= 2400000
id:13 bucket: <= 2600000

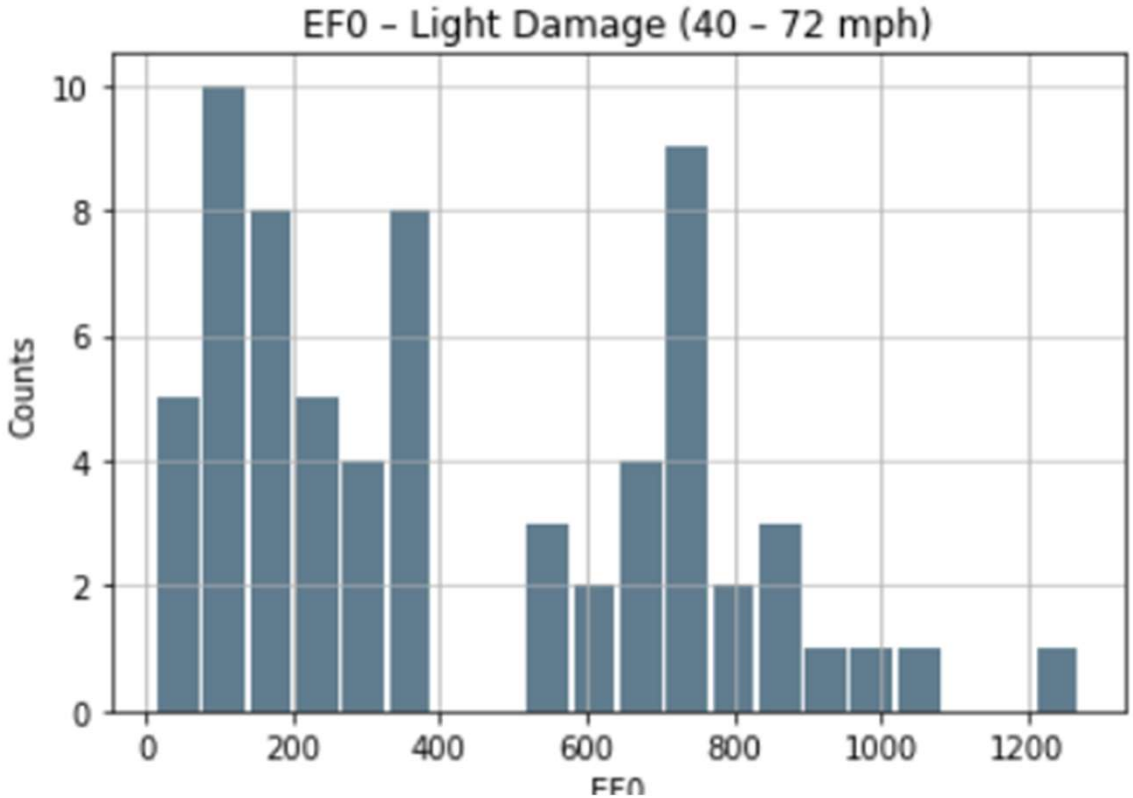
TOR_F_SCALE

- count: 68456
- unique: 6
- top: EF0
- freq: 30087



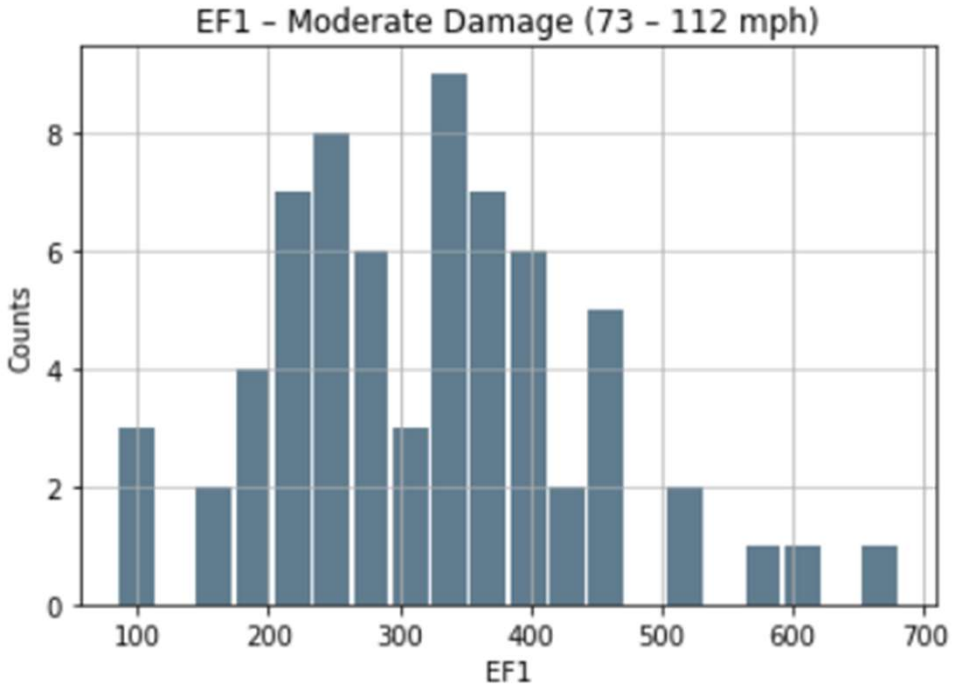
EF0

- count: 67
- mean: 424.93
- std: 311.62
- min: 13
- 0.25: 156
- 0.5: 343
- 0.75: 709.5
- max: 1271



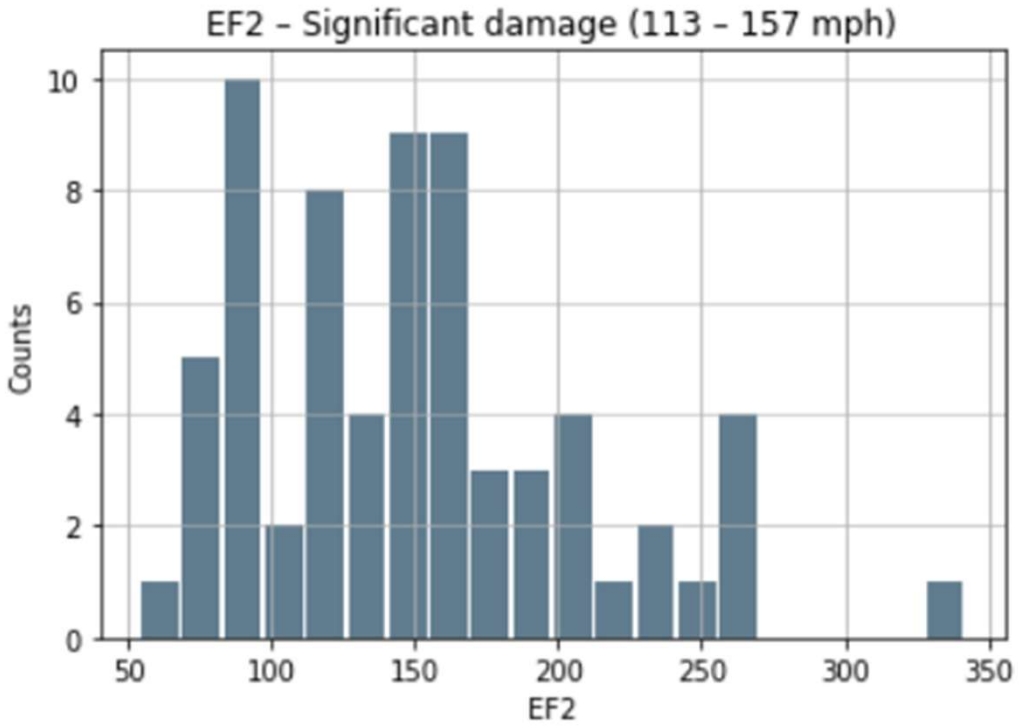
EF1

- count: 67
- mean: 320.37
- std: 118.63
- min: 84
- 0.25: 240.5
- 0.5: 324
- 0.75: 386
- max: 682



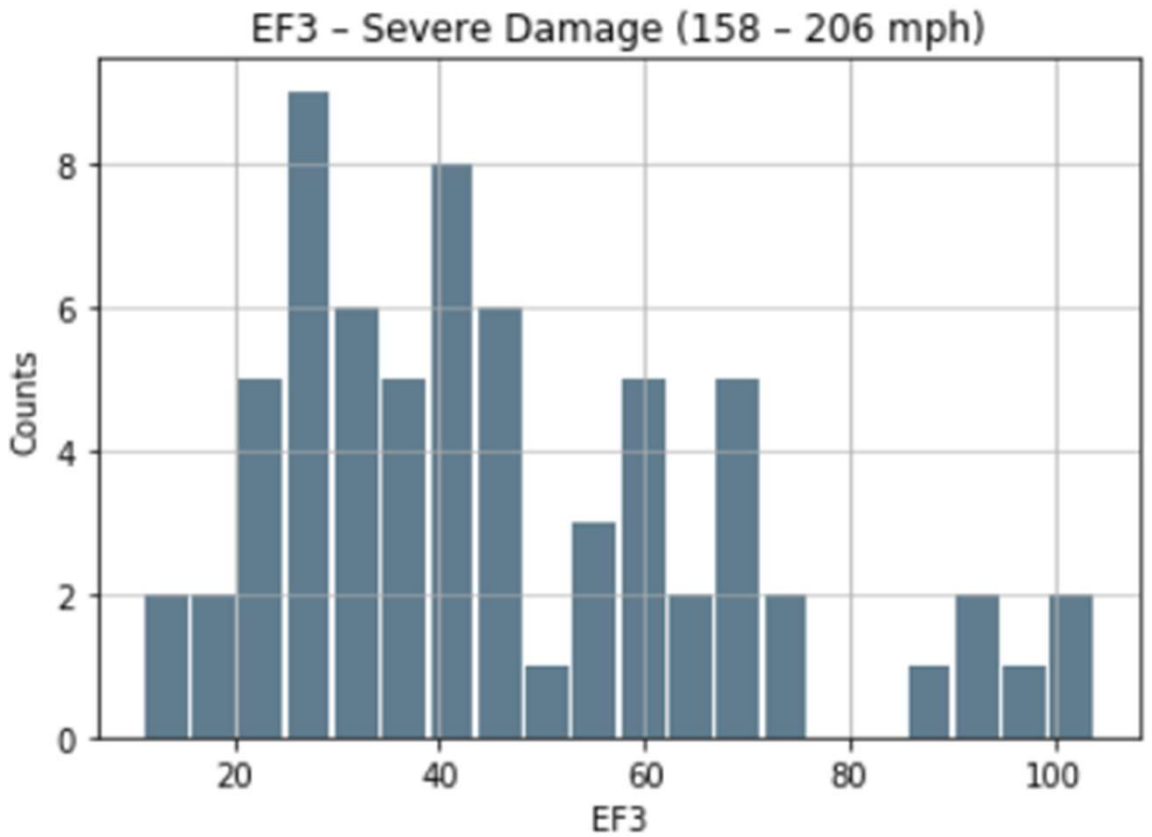
EF2

- count: 67
- mean: 151
- std: 58.27
- min: 54
- 0.25: 101.5
- 0.5: 151
- 0.75: 180.5
- max: 342



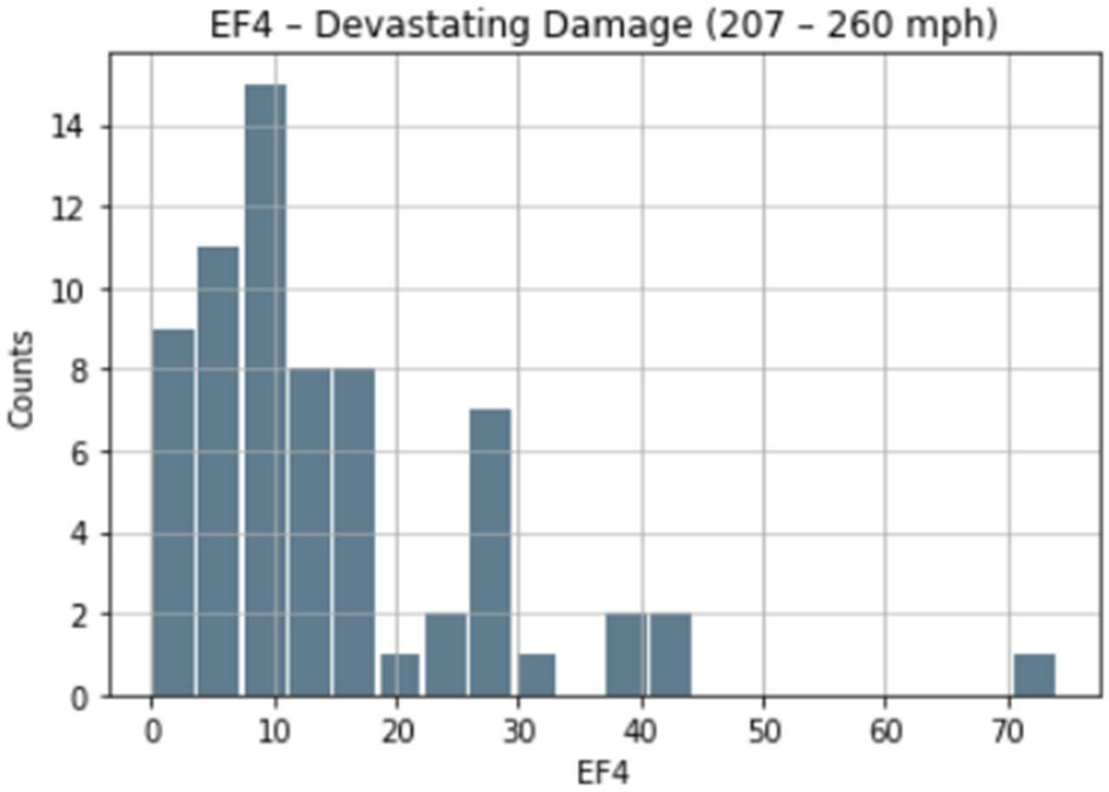
EF3

- count: 67
- mean: 46.46
- std: 22.17
- min: 11
- 0.25: 28.5
- 0.5: 41
- 0.75: 60.5
- max: 104



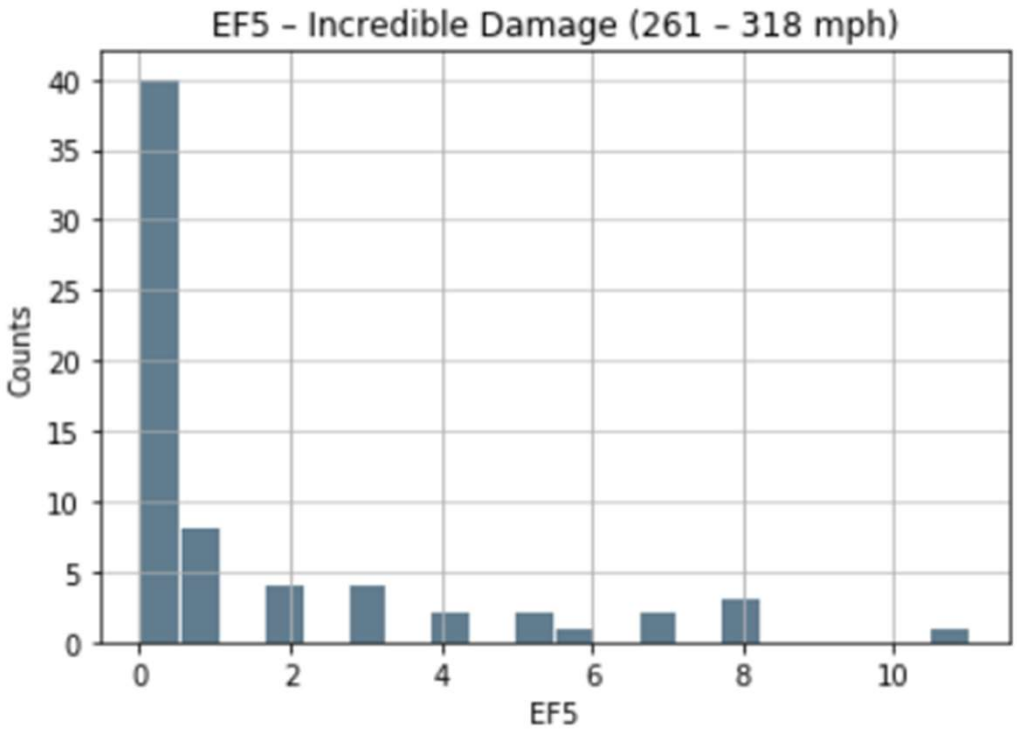
EF4

- count: 67
- mean: 14.63
- std: 12.71
- min: 0
- 0.25: 6.5
- 0.5: 11
- 0.75: 18
- max: 74



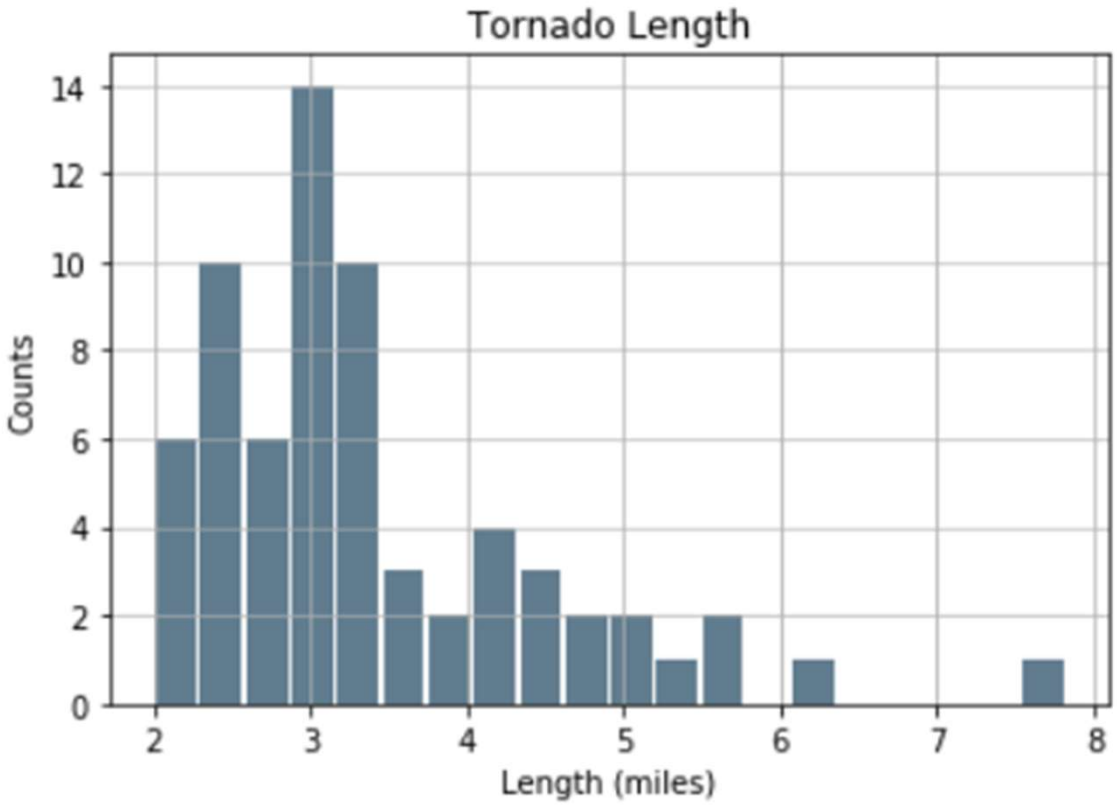
EF5

- count 67.00
- mean 1.51
- std 2.57
- min 0.00
- 25% 0.00
- 50% 0.00
- 75% 2.00
- max 11.00



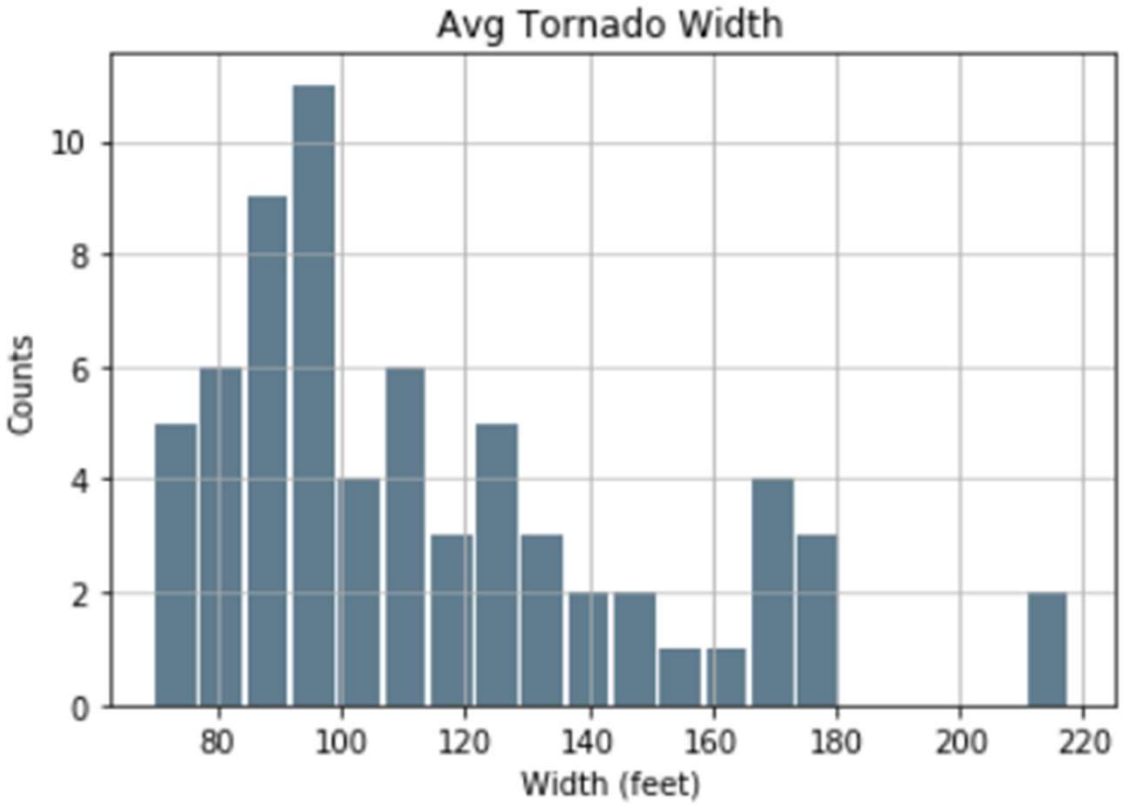
TOR_LENGTH

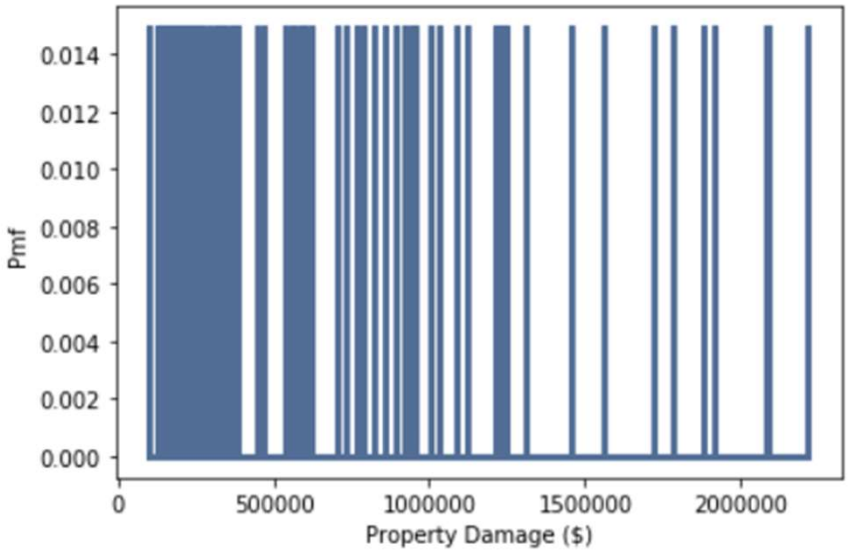
- count: 67
- mean: 3.39
- std: 1.11
- min: 1.99
- 0.25: 2.71
- 0.5: 3.06
- 0.75: 3.86
- max: 7.82



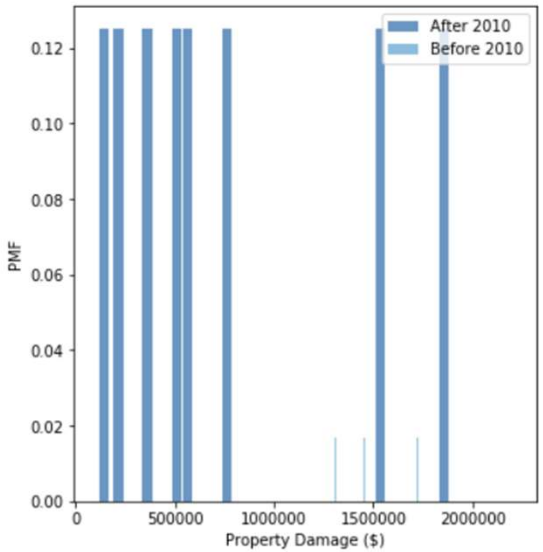
TOR_WIDTH

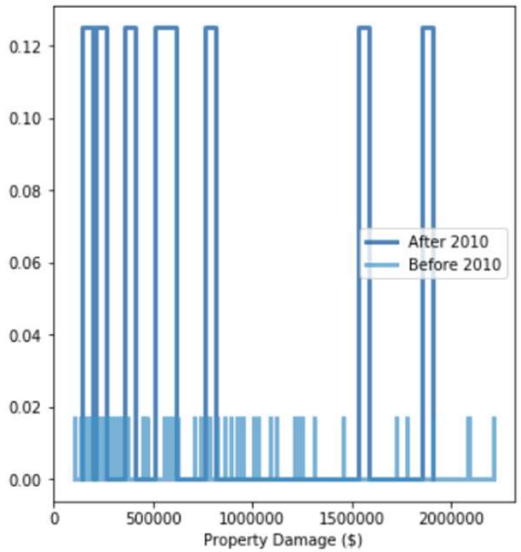
- count: 67
- mean: 114.37
- std: 34.99
- min: 69.52
- 0.25: 90.39
- 0.5: 104.18
- 0.75: 130.16
- max: 217.94





DAMAGE_PROPERTY PMF – ROUND 1



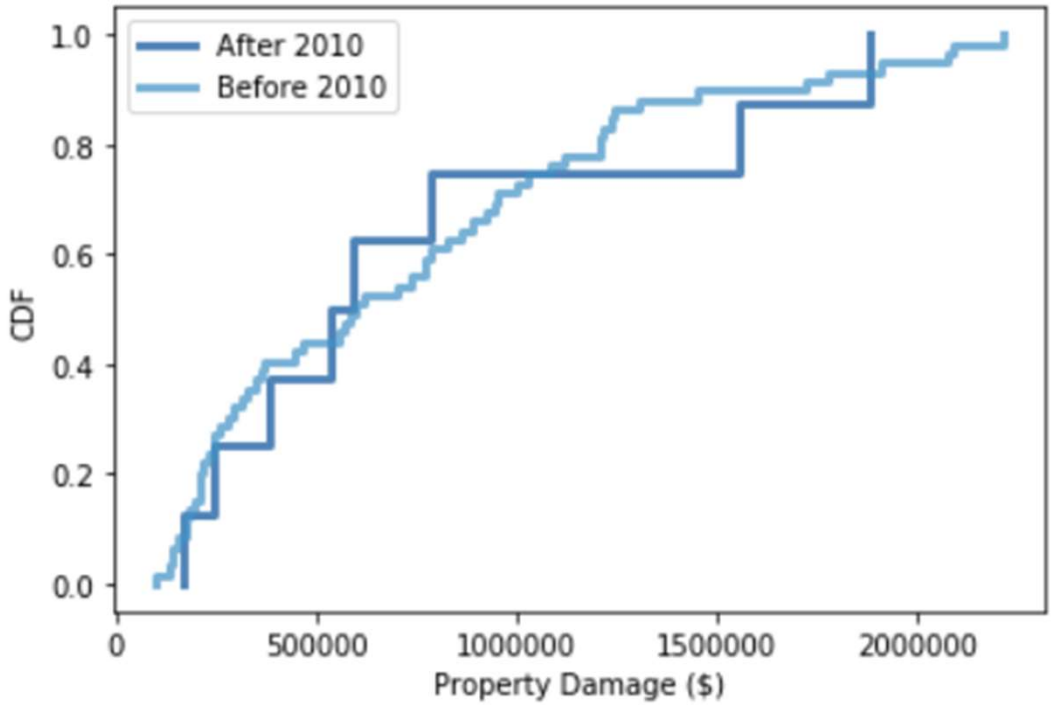


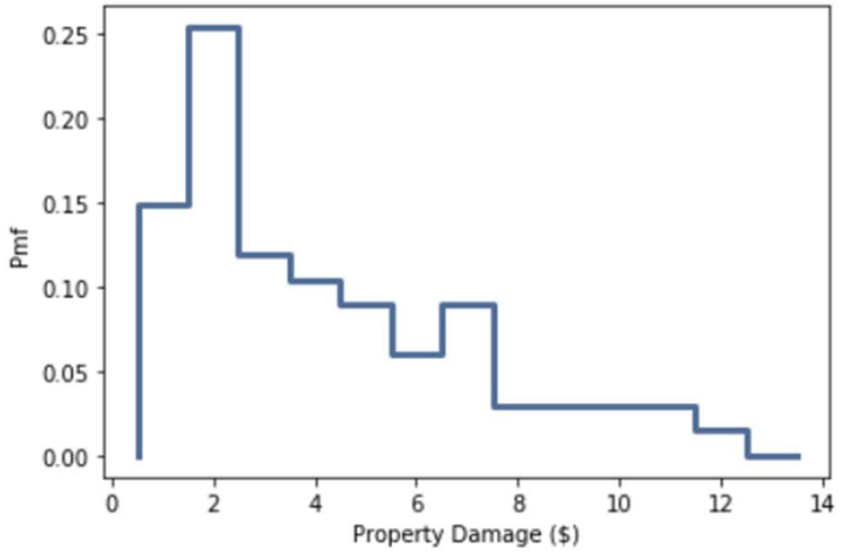
DAMAGE_PROPERTY PMF – ROUND 1

- Comparing all data prior to 2010 vs everything after 2010
- Graphs are difficult to read
- Seems like all values are similar in likelihood
- After has less values, which causes the likelihood to increase for each value

DAMAGE_PROPERTY CDF – ROUND 1

- After 2010 – more blocky due to less data
- Similar in pattern

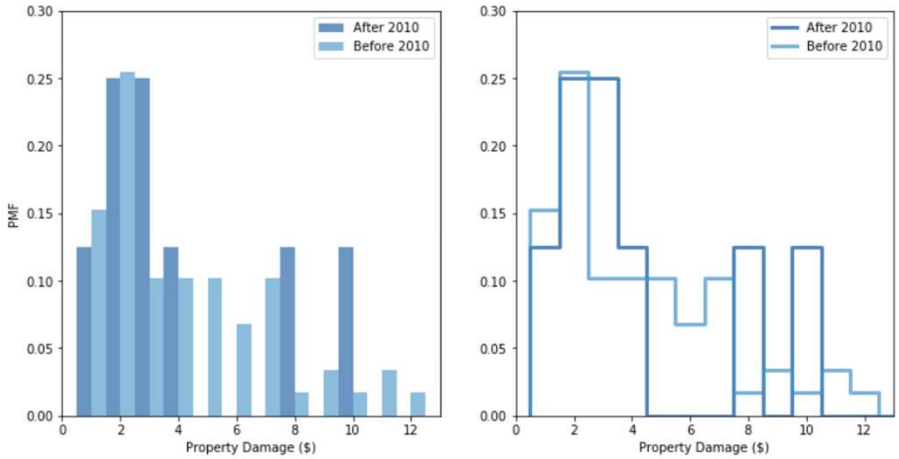


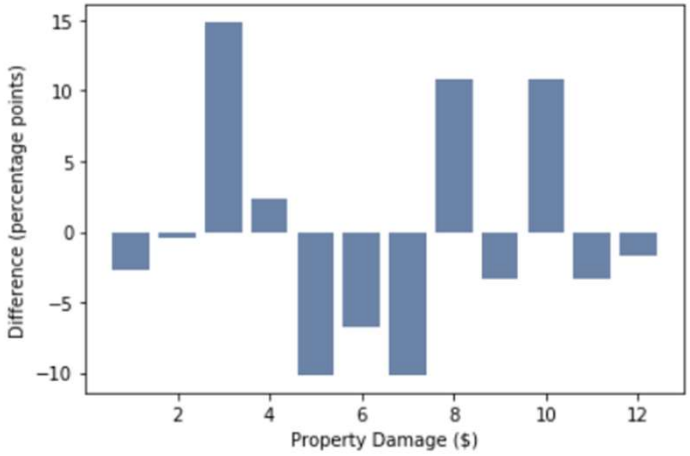


DAMAGE_BUCKET PMF – ROUND 2

DAMAGE_PROPERTY PMF – ROUND 2

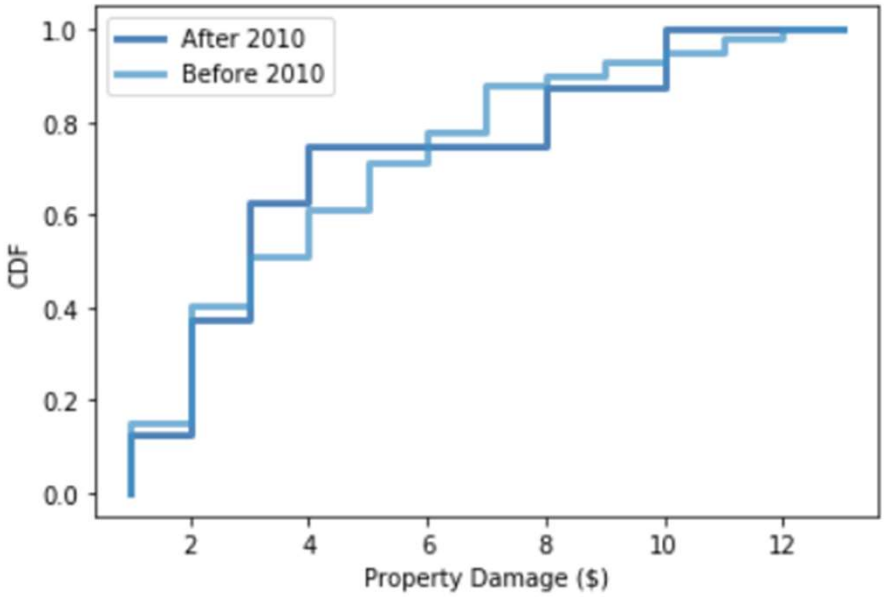
- Easier to compare data
- After has less values still





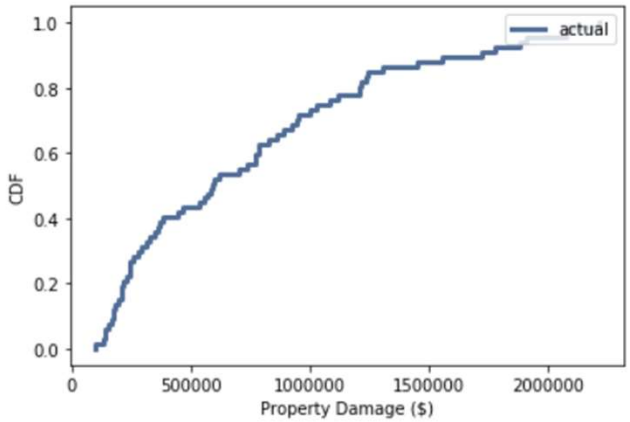
DAMAGE_PROPERTY CDF – ROUND 2

- “Before 2010” following same pattern with bigger steps
- “After 2010” less steps more smooth

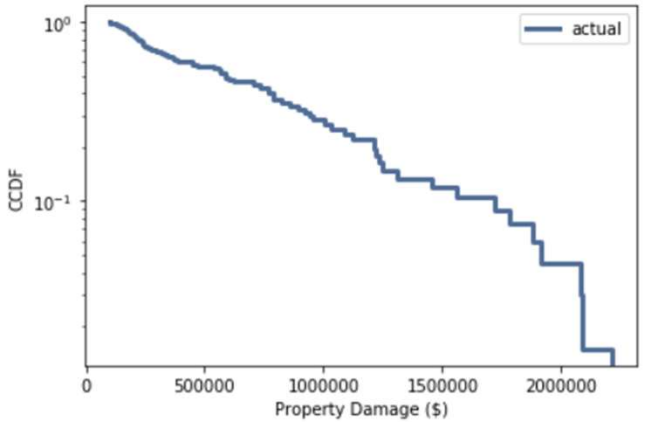


EXPONENTIAL DISTRIBUTION ON DAMAGED_PROPERTY

CDF



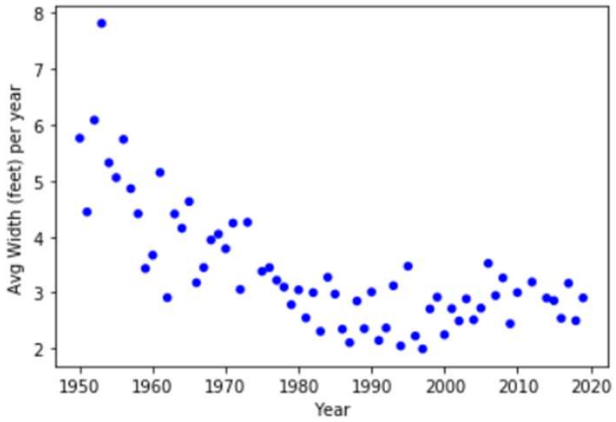
CCDF



0.008008888815100956

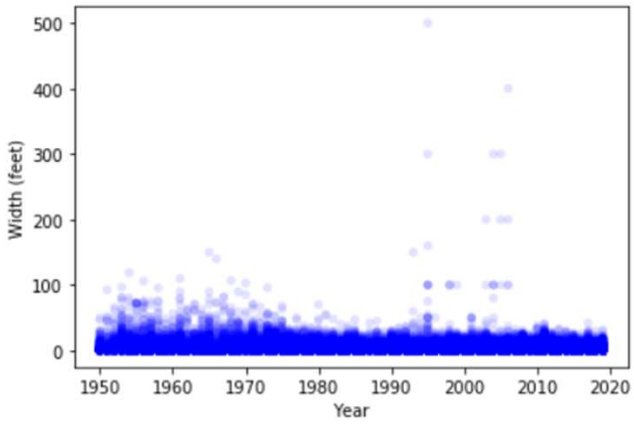
CORRELATION BETWEEN YEAR AND TOR_LENGTH

Avg . Tornado length per year



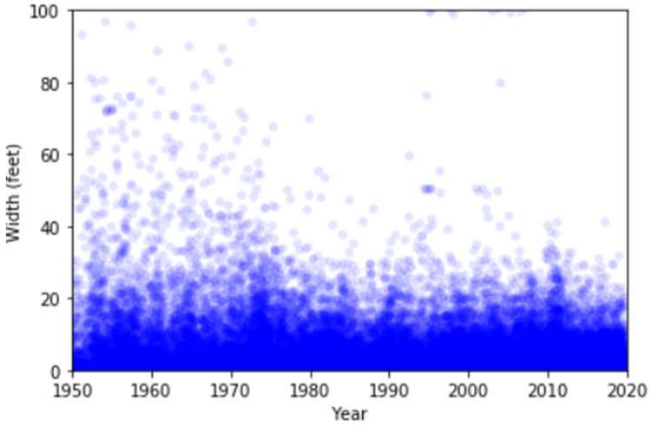
- Pearson correlation: -0.70892305
- Spearman: -0.698699018277596
- Transformed with log correlation: -0.7125315901211206

Raw Tornado length



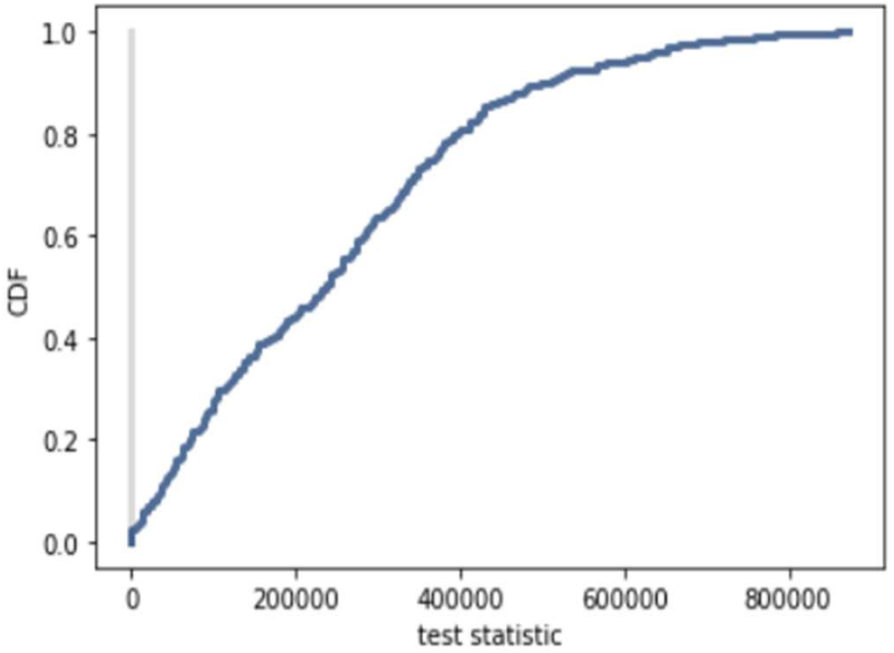
- Pearson correlation: -0.08890348
- Spearman: 0.10193580487255906
- Transformed with log correlation: -0.008008888815100956

Raw Tornado length zoom



$$H_0 : \mu_0 = \mu_1$$
$$H_A : \mu_0 \neq \mu_1$$

- Ho – mean property damage prior to 2010 = mean property damage after 2010
- Ha – they don't equal
- Two-sided tail p-value = 1



COMPARING THE MEANS

HYPOTHESIS TESTING

MULTI-LINEAR REGRESSION

- DAMAGE_PROPERTY_t ~ YEAR + C(TOR_F_SCALE)
- Not a good fit
- Potential multicollinearity
- F-statistic is high
- Low R-squared value (.045)
  - YEAR AND TOR_F_SCALE ONLY CONTRIBUTE 4.5% VARIATION ON DAMAGE_PROPERTY_t

OLS Regression Results

Dep. Variable:	DAMAGE_PROPERTY_t	**R-squared:**	0.045
Model:	OLS	**Adj. R-squared:**	0.045
Method:	Least Squares	**F-statistic:**	534.2
Date:	Sat, 29 Feb 2020	**Prob (F-statistic):**	0.00
Time:	17:29:25	**Log-Likelihood:**	-1.2430e+06
No. Observations:	68456	**AIC:**	2.486e+06
Df Residuals:	68449	**BIC:**	2.486e+06
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-9.378e+07	8.07e+06	-11.625	0.000	-1.1e+08	-7.8e+07
C(TOR_F_SCALE)[T.EF1]	5.804e+05	1.66e+05	3.505	0.000	2.56e+05	9.05e+05
C(TOR_F_SCALE)[T.EF2]	1.985e+06	2.19e+05	9.077	0.000	1.56e+06	2.41e+06
C(TOR_F_SCALE)[T.EF3]	5.97e+06	3.44e+05	17.362	0.000	5.3e+06	6.64e+06
C(TOR_F_SCALE)[T.EF4]	1.914e+07	5.75e+05	33.270	0.000	1.8e+07	2.03e+07
C(TOR_F_SCALE)[T.EF5]	6.965e+07	1.59e+06	43.809	0.000	6.65e+07	7.28e+07
YEAR	4.697e+04	4038.906	11.629	0.000	3.91e+04	5.49e+04

Omnibus:	258435.857	**Durbin-Watson:**	1.944
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	285418950166.893
Skew:	85.430	**Prob(JB):**	0.00
Kurtosis:	10004.794	**Cond. No.**	2.26e+05

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.26e+05. This might indicate that there are strong multicollinearity or other numerical problems.

CONCLUSION