

Questions 1 - 8: all students - Question 9-10: Steven - Questions 11-12: Bart - Questions: 13-14: Herlinde - Questions 15-16: Laurens - ODS report: Bart and Laurens

Introduction

Every single year human lives, ecosystems and properties are at risk due to wildfires in the state of North-Carolina (NC). Wildfires harm both air and water quality while efforts to fight wildfires are expensive and put emergency personnel in harms way. A better understanding of the dynamics behind destructive wildfires can aid prevention efforts. This report analyses the causes, timings and sizes of wildfires across different counties in NC. The data is obtained from the United States Forest Service (USFS) and spans across 1992 to 2015.

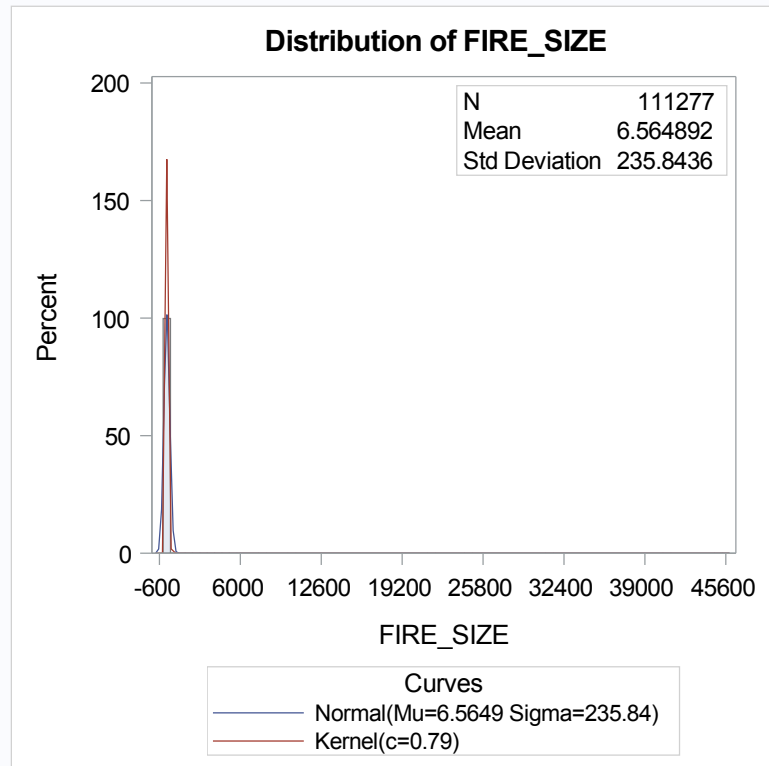
The analyses in this report have been performed on a dataset containing the following categorical variables: the county the fire occurred in, its fips code, the year the fire was discovered in and the official cause with a matching cause code. Additionally two categorical variables have been added: fire sizes binned into 7 categories and the months the wildfires have been discovered in. The continuous variables studied are the discovery date of the fire, the fire size and the coordinates of the fire. The only missing values in the dataset are for the variables county and the fips code. The county variable also contained several erroneous entries, all of which were correctable. Below the first 5 rows of the dataset are depicted as an example.

The first five observations from the dataset

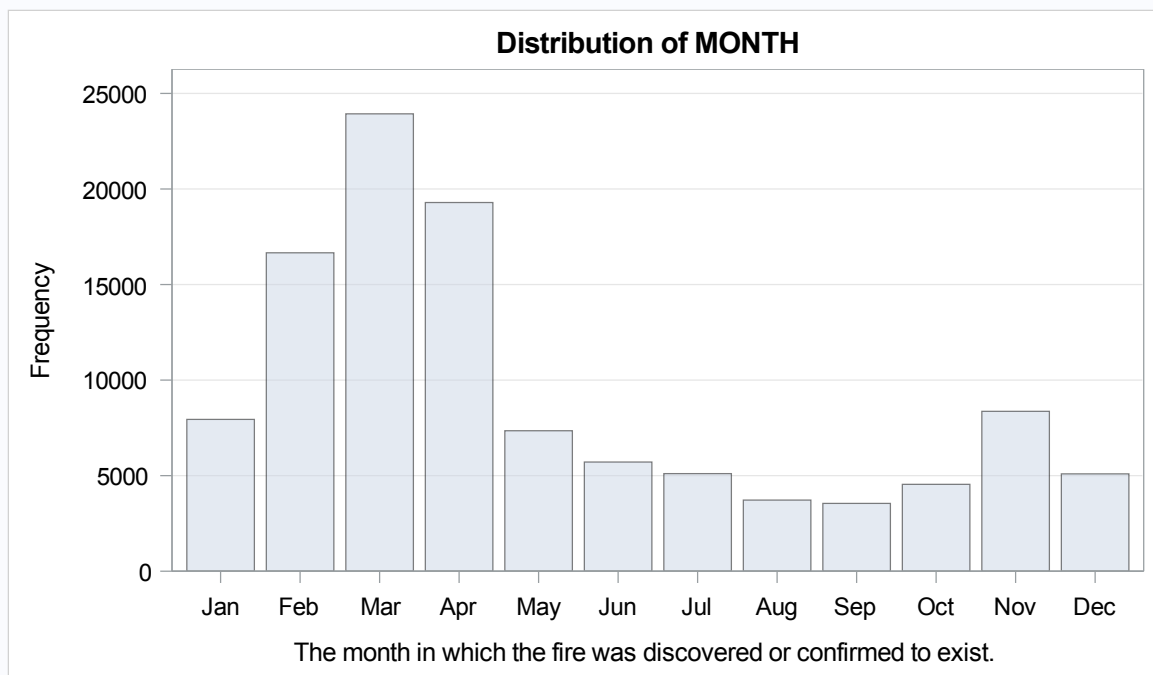
Obs	Year	Discovery date	Cause code	Cause	Fire size	Latitude	Longitude	State	County	Fips code	Month	Fire size class
1	2005	11MAR05	2	Equipment Use	0.6	35.23	-82.88	NC	Buncombe	21	Mar	B
2	2005	27JAN05	7	Arson	50.3	35.00	-83.35	NC	Macon	113	Jan	C
3	2005	06FEB05	7	Arson	0.1	35.93	-81.72	NC	Caldwell	27	Feb	A
4	2005	12FEB05	5	Debris Burning	125	36.00	-81.59	NC	Caldwell	27	Feb	D
5	2005	16APR05	5	Debris Burning	25	35.99	-81.85	NC	Avery	11	Apr	C

Number of Variable Levels				
Variable	Label	Levels	Missing Levels	Nonmissing Levels
STATE	Two-letter alphabetic code for the state in which the fire burned (or originated), based on the nominal designation in the fire report.	1	0	1
COUNTY	County, or equivalent, in which the fire burned (or originated), based on nominal designation in the fire report.	101	1	100
FIPS_CODE	Three-digit code from the Federal Information Process Standards (FIPS) publication 6-4 for representation of counties and equivalent entities.	101	1	100
FIRE_YEAR	Calendar year in which the fire was discovered or confirmed to exist.	24	0	24
STAT_CAUSE_CODE	Code for the (statistical) cause of the fire.	13	0	13
STAT_CAUSE_DESCR	Cause of the fire.	13	0	13

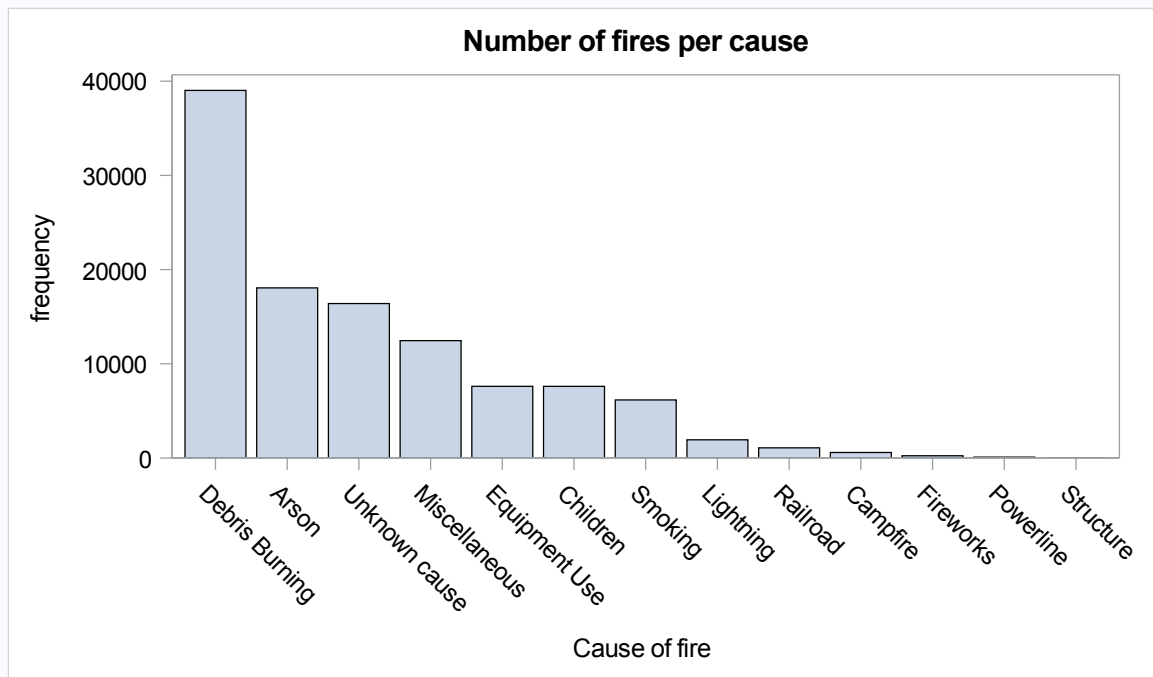
The table displays the number of levels for each categorical variable. The values for the state variable are all the same, namely 'NC' as a reference to the state of North-Carolina. Both the county and fips_code variables have 100 different levels and missing values. The data spans across 24 years and contains thirteen different categories for causes.



There do not appear to be any inconsistencies in any continuous variables. However, the size of the wildfires is skewed to the right, with a few unusually large wildfires while most wildfires do not spread far. The most extreme observation being a wildfire which, according to the data, burned 45294 acres of land. There's no reason to assume these extreme values are errors as they appear to be realistic.



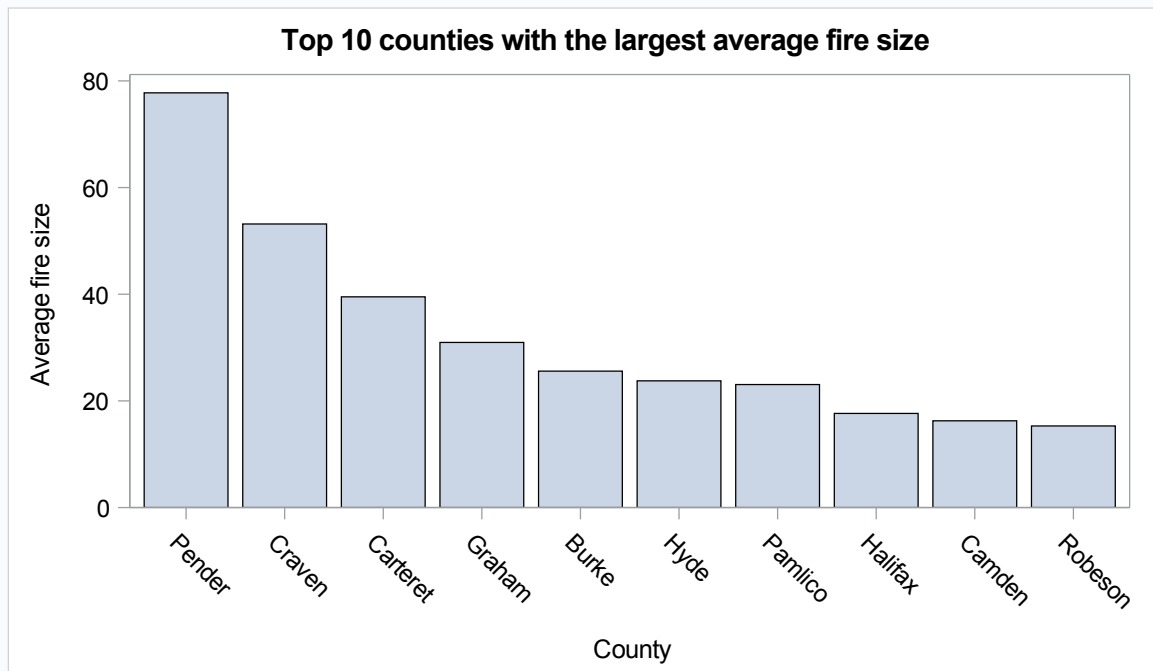
During late winter and early spring wildfires occur more often while summer is the calmest period.



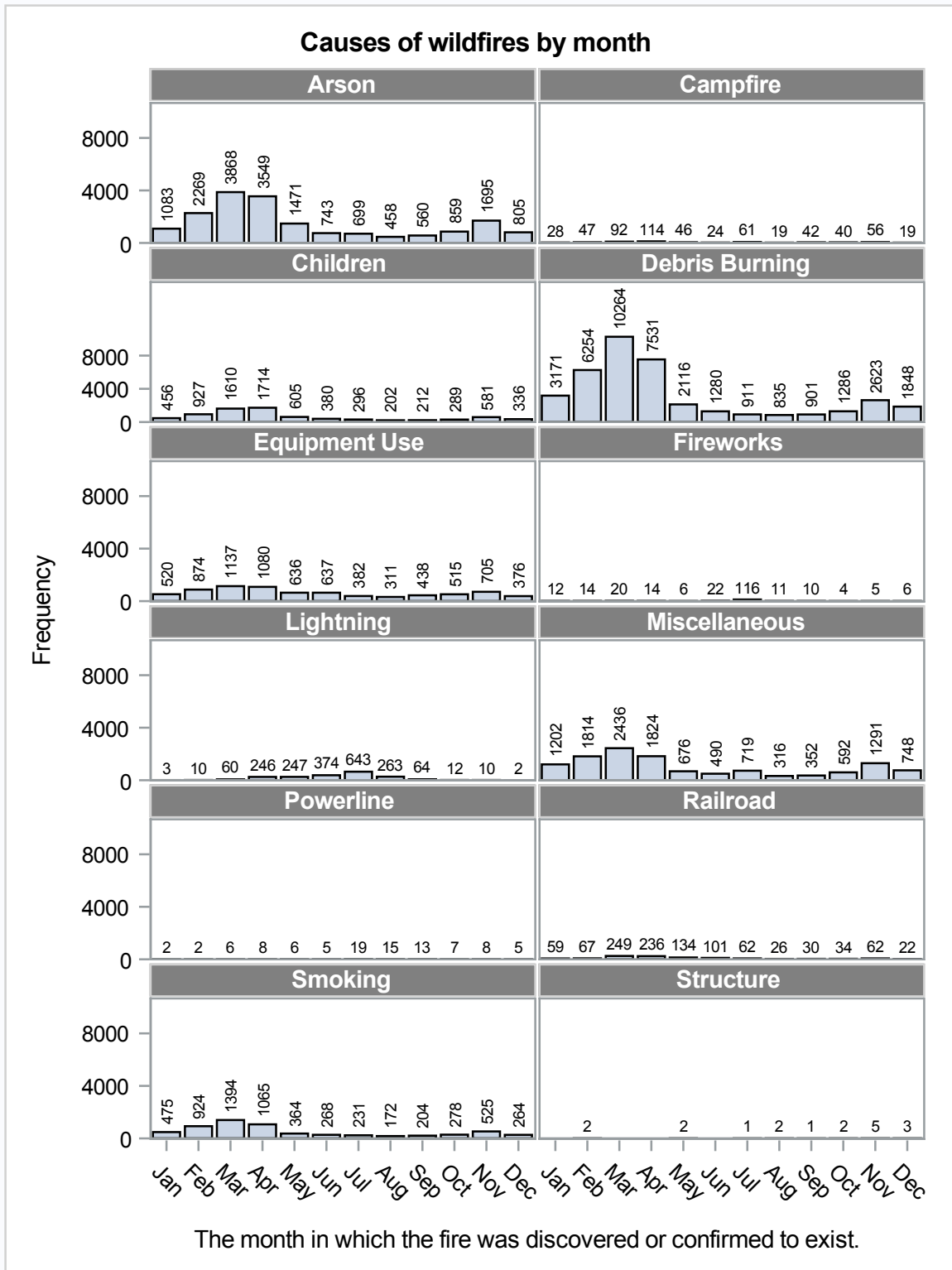
By far the most common cause of wildfires is debris burning, which accounts for more than one third of all wildfires. A distant second is Arson. The third most common official cause is equipment use, which has barely been documented more often compared to wildfires caused by children. However, the large amount of unknown causes leaves some room for ambiguity. The category 'miscellaneous', by assumption, consists of a multitude of causes that cannot be assigned to any other category and thus to hold data for numerous less common causes. The category of wildfires with an unknown cause is not taken into account.

Obs	COUNTY	_TYPE_	_FREQ_	n_obs	average	lower_CI	upper_CI
1	Pender	1	511	511	77.759041096	-42.45166728	197.96974947

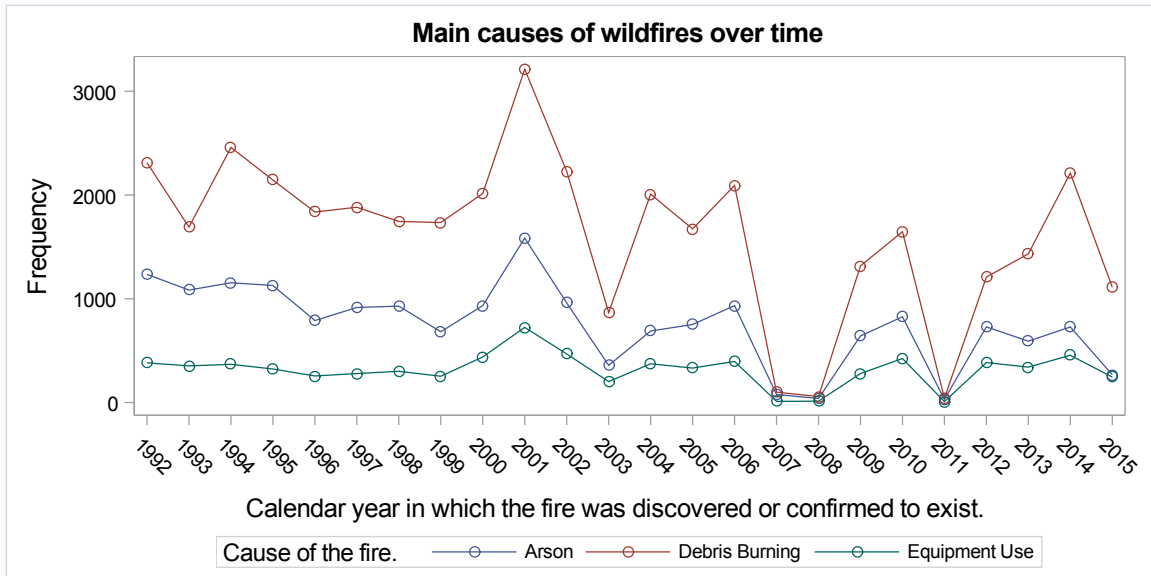
The county with historically the largest average fire size is Pender, where wildfires have on average burned 77.76 acres each.



The bar chart displays the 10 counties where historically the average fire size has been the largest. The county of Pender clearly towers above the others.



With the exception of lightning and fireworks, most causes follow the same trend. Namely a spike in late winter and early spring, while summer has the least occurrences of wildfires. Fireworks and lightning go against the trend as they more often cause wildfires during summer.



The amount of wildfires due to debris burning and arson has been on the decline while the amount of wildfires caused by equipment use seems to be quite steady over time.

	<i>Fire Size: Estimate of acres within the final perimeter of the fire</i>			
	Number of fires	Total size	Average size	Maximum size
Top 3 causes of fire				
Debris Burning	39020	135370.5	3.469	2738.000
Arson	18059	175660.9	9.727	24600.00
Equipment Use	7611	39568.33	5.199	2300.000

Eventhough debris burning causes more wildfires, those wildfires are on average a lot more contained compared to those caused by arson and equipment use. Wildfires caused by arson have torched on average 30 % more acres of land while debris burning accounts for more than double the amount of wildfires.

Goodness-of-Fit Tests for Normal Distribution

STAT_CAUSE_DESCR = Arson
Fitted Normal Distribution for logtransform

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.120094	Pr > D	<0.010
Cramer-von Mises	W-Sq	39.545589	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	313.607194	Pr > A-Sq	<0.005

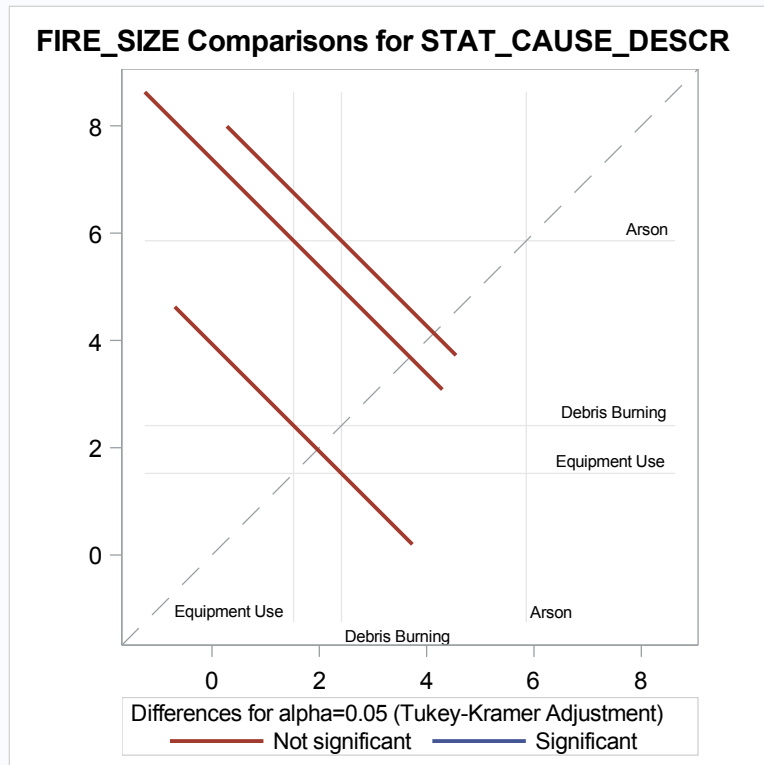
STAT_CAUSE_DESCR = Debris Burning
Fitted Normal Distribution for logtransform

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.101094	Pr > D	<0.010
Cramer-von Mises	W-Sq	66.984678	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	500.035621	Pr > A-Sq	<0.005

STAT_CAUSE_DESCR = Equipment Use
Fitted Normal Distribution for logtransform

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.178533	Pr > D	<0.010
Cramer-von Mises	W-Sq	46.068030	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	298.999145	Pr > A-Sq	<0.005

The log transformed fire size distributions are not normally distributed: visually no normal distribution can be seen when looking at the histograms or the QQ-plots. In addition, for the Goodness of Fit tests, the p-values are lower than 0.05 (the null hypothesis indicates a normal distribution while the alternative hypothesis indicates a distribution deviating from normality). Based on these p-values, the null hypothesis is rejected, leading to the conclusion that the distribution deviates from a normal distribution.



With a p-value of 0.88 (>0.05) there is no significant difference between the fire sizes caused by debris burning and equipment use in 2015. The same counts for the pairs of debris burning vs arson and equipment use vs arson with respective p-values of 0.14 and 0.16 (>0.05). The visualization above clearly shows that the fire sizes between debris burning and equipment use are fairly similar.

Conclusion

The analysis of the wild fire occurrence in the state of North Carolina reported in this paper, suggests that most fires occur in early spring while the fire frequency is the lowest during summer. Based on the data from 1992 until 2015, the major causes of these fires are burning debris, arson and equipment use. Each of the top 3 causes follow the same seasonal trend in terms of occurrence. In absolute numbers, burning debris causes most fires while fires caused by arson are on average larger in size. For fires registered in 2015, the average fire sizes originating from the top 3 causes were not significantly different as formally shown by anova test. Pender county has on average the largest wildfires, followed by Craven and Carteret county. This average is however influenced by outliers (extremely large fires) which cause the confidence interval to be very wide.