

Homework SAS

Wildfire analysis for the state North Carolina

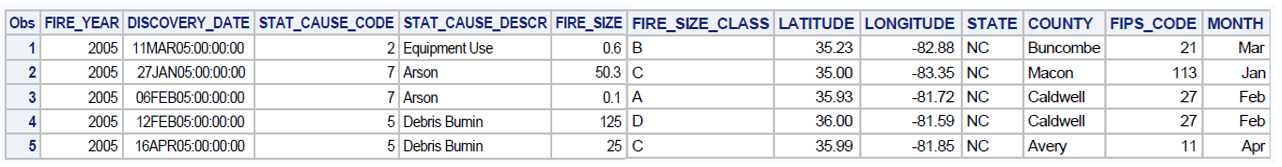
Herlinde Wynendaele - Laurens Van Paemel - Steven Vandeleene - Bart Moens

# Introduction

The report describes the analysis of wild fire occurrences in the state of North Carolina. The data is obtained from the United States Department of Agriculture Forest Services contains data from wildfires that occurred in the United States from 1992 to 2015. In the analysis we investigate the timings, causes and fire sizes for the different counties in North Carolina

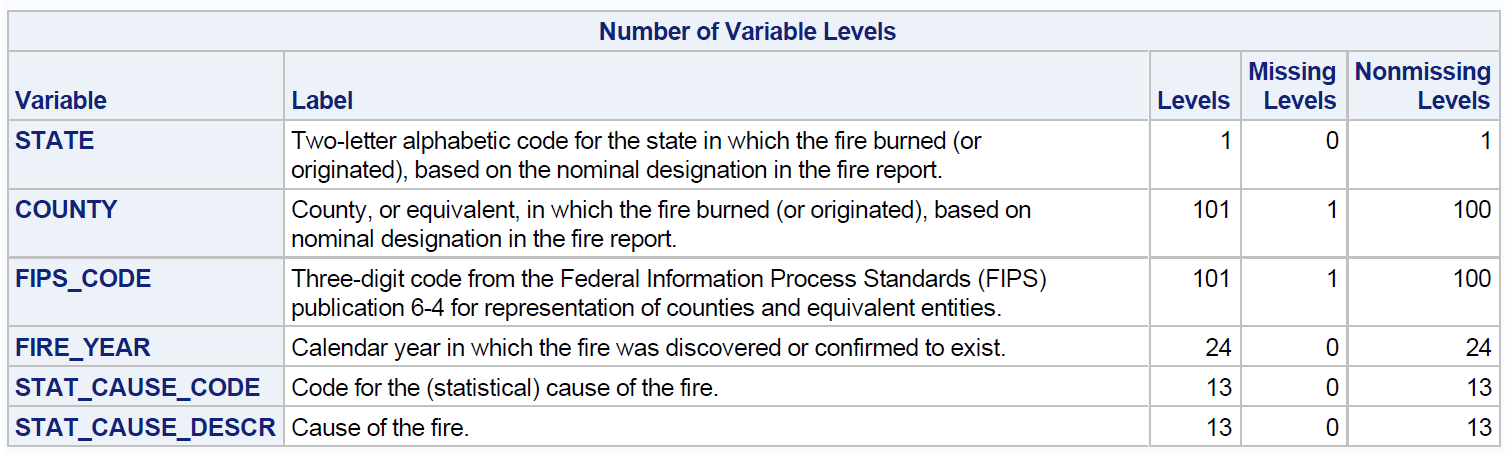
# Data Preparation

In the data preparation step all variables were formatted, labels with information were added, month and county translated to a text format and a class for fire size was added.

First 5 observations of the data set after the data preparation step

Analysis of missing values

Via the SAS Freq procedure the missing levels and values were investigated.



The details for all missing data for each variable can be retrieved in the SAS ODS output report.

Analysis of inconsistencies in the data

Via the SAS Univariate procedure inconsistencies and extreme observations in the data are investigated. Details can be found in the SAS program and the ODS output report.

# Descriptive Analysis of the data

## Analysis of wildfires by Month

|  |  |
| --- | --- |
| The data clearly indicates that wildfires more often occur in the months February, March and April. |  |

## Analysis of wildfires by cause

|  |  |
| --- | --- |
| Debris burning is identified as major cause of wildfires in the state of North Carolina. |  |

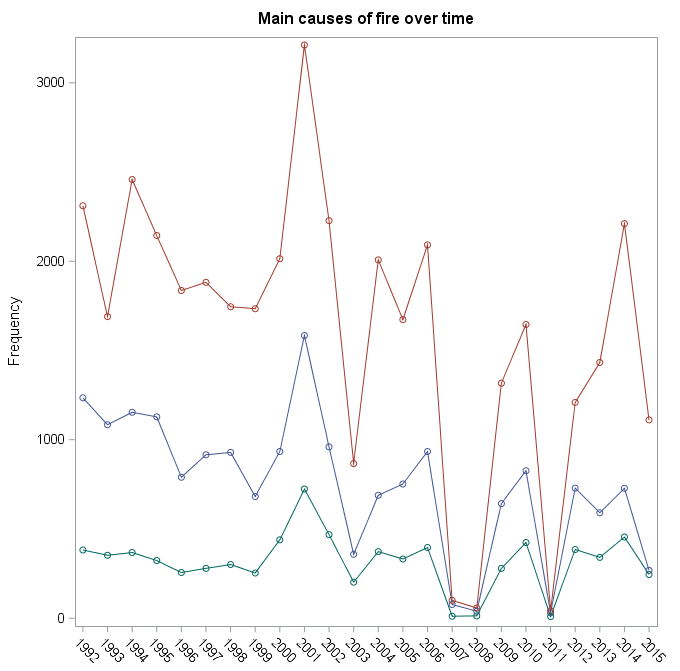
## Analysis of wildfires by county

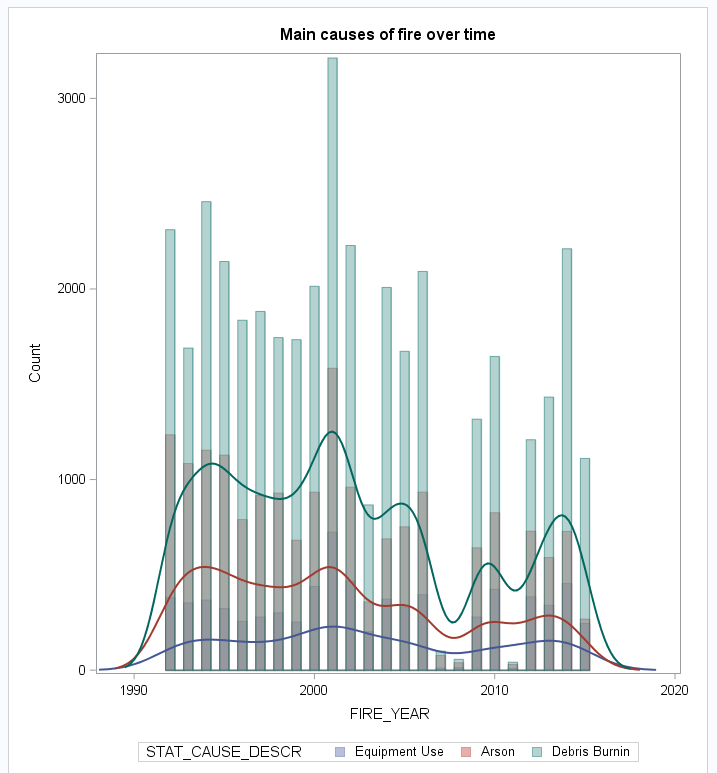
|  |  |
| --- | --- |
| The county Pender has in average the biggest wildfires. In addition to the average, the lower and upper confidence intervals were calculated. |  |

## Analysis of causes of fire versus the size by month

|  |  |
| --- | --- |
|  |  |
|  |  |
|  | The top 3 causes of fire size are considered to be Debris burning, Arson and Equipment use.  The category wildfires with an unknown cause are not taken into account. For the category miscellaneous assumption is made that this category holds data for wildfires that can not be assigned to any other category, but holds data for numerous different reasons. |

## Analysis of top 3 cause of wild fires over time





## Analysis of the distribution of log transformed fire size for the top 3 causes

|  |  |
| --- | --- |
|  |  |
|  | |

We conclude the log transformed fire size distributions are not normally distributed.

## Analysis of significant difference in fire size in the year 2015 between the 3 main causes

|  |  |
| --- | --- |
| The SAS GLM procedure is used to analyze any significant differences between the 3 main cause of wildfires. |  |