PCA Report on Forest Fires

Carmelo Emmanuele
Zachary Goodsell

April 9, 2019

1 Abstract

Forest fires are a major environmental issue, creating economical and ecological damage while endangering human lives. In order to save human lives fast detection is a key element for controlling such phenomenon. A factor for forest fires are meteorological condition, such as: temperature and wind are known to influence forest fires and several fire indexes, such as the forest Fire Weather Index (FWI). The way we obtained the data was by using a correlation plot and grouping the data based on the categories. Lastly, the aim is to predict the burned area of forest fires, in the northeast region of Portugal, by using meteorological data.

2 Introduction

2.1 Measures and Context

The data we decided to investigate has to due with forest fires and the many attributes that surround the event. This is interesting because with the data we can tell what contributes to the fires in the most active months. With this data of when the fires have started, one could plan ahead for staffing and resources that could prevent or reduce the severity of these fires. The data set shows different categories of why the forest fire started and with these numerical values can find ways to prevent the fire.

2.2 Background

The variables in the excel spreadsheet has different reasons on why the forest fire has started. For example, the Fine Fuel Moisture Code (FFMC) is a numeric rating of the moisture content of litter and other cured fine fuels. This code is an indicator of the relative ease of ignition and the flammability of fine fuel. Also, the Duff Moisture Code (DMC) is a numeric rating of the average moisture content of loosely compacted organic layers of moderate depth. This code gives an indication of fuel consumption in moderate duff layers and medium-size woody material. Then, the Drought Code (DC) is a numeric rating of the average moisture content of deep, compact organic layers. This code is a useful indicator of seasonal drought effects on forest fuels and the amount of smoldering in deep duff layers and large logs. Lastly, the Initial Spread Index (ISI) is a numeric rating of the expected rate of fire spread. It combines the effects of wind and the FFMC on rate of spread without the influence of variable quantities of fuel.

2.3 Questions and Reasons

The reason we're doing this is so agencies can more accurately predict when a large scale forest fire may breakout. With these predictions the agencies can more efficiently manage their staff and resources to where they need in order to prevent larger scale forest fires.

2.4 Explaining PCA Data

PCA is being used to help answer by creating correlations between the different attribute in the data. For example, if we can find a correlation of relative humidity and wind that give a better idea of a fire starting in certain months, one could more easily predict a fire breakout. The groupings in the data show how the fire started and during which month it occurred in.

3 Methods

3.1 Discussion

Our data set originally included 13 variable, which were X, Y, month, day, FFMC , DMC, DC, ISI, temp, RH, wind, rain, area. From this we removed X, Y, and day because we found that they were irrelevant. Next, we removed the data from the non-forest fire season because we found March, June, July, August, and September the most active months for the forest fire season. The most relevant fields we found were month, FFMC , DMC, DC, ISI, temp, RH, and wind. The area for most of the data was zero due to a lack of reporting and because of how this skewed our results we omitted it. We obtained the data from UCI machine learning repository found here.

3.2 Plotting Data

To use PCA on this data we set we used a correlation matrix graph to find the variables that were most closely correlated. With the correlation matrix we found that the following where most closely related - FFMC and ISI, DC and DMC, and temp, RH, and wind. With these correlated data sets we were able to construct PCA's for each data set.

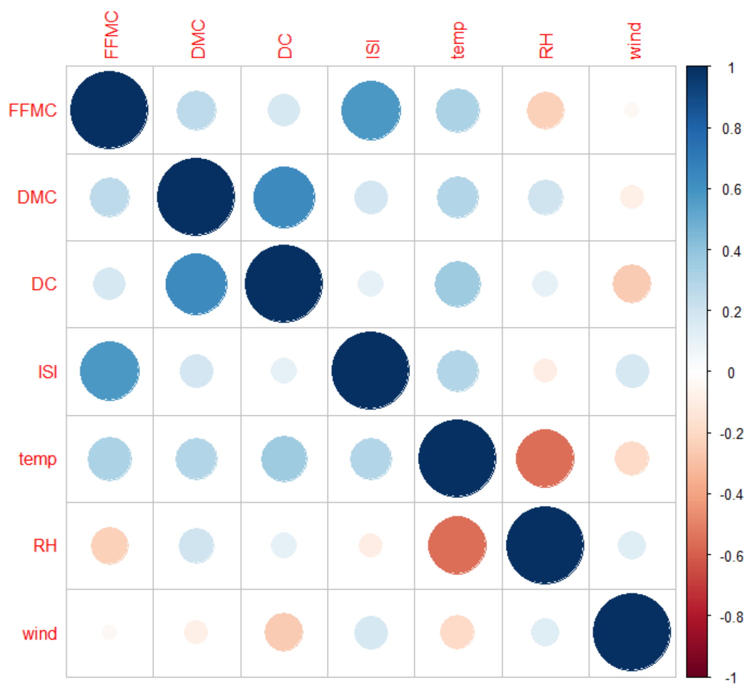

Figure 1: This is the correlation matrix of the forest fire data set

4 Results

4.1 Discussion

With temperature and relative humidity we were able to conclude the following. The months of June, July, August, and September has a higher relative humidity, but the higher temperatures in these months resulted in forest fires occurring. In March the temperatures are significantly lower but the very low humidity contributes to the forest fires occurring.

With Initial Spread Index and Fine Fuel Moisture Code we were able to conclude the following. In the months of March and September we see that FFMC is more of a contributing factor to forest fires occurring. In July and August we see that ISI is more of a contributing factor to forest fires occurring. For June none of these factors seem to affect when forest fires occur.

With Duff Moisture Code and Drought Code we were able to conclude the following. In the month of September it correlated with a low DMC. However, in the months of August, June, and July the numerical values for DC and DMC were average. Lastly, in the month of March the DC value was extremely low.

With temperature, relative humidity, and wind we were able to conclude the following. Winds is the new data values added to this compared to the first graph, just temperature and relative humidity. The only month that doesn't have an average amount of wind is March and this helps the occurrence of forest fires because it combats the low temperatures.

4.2 Graphs

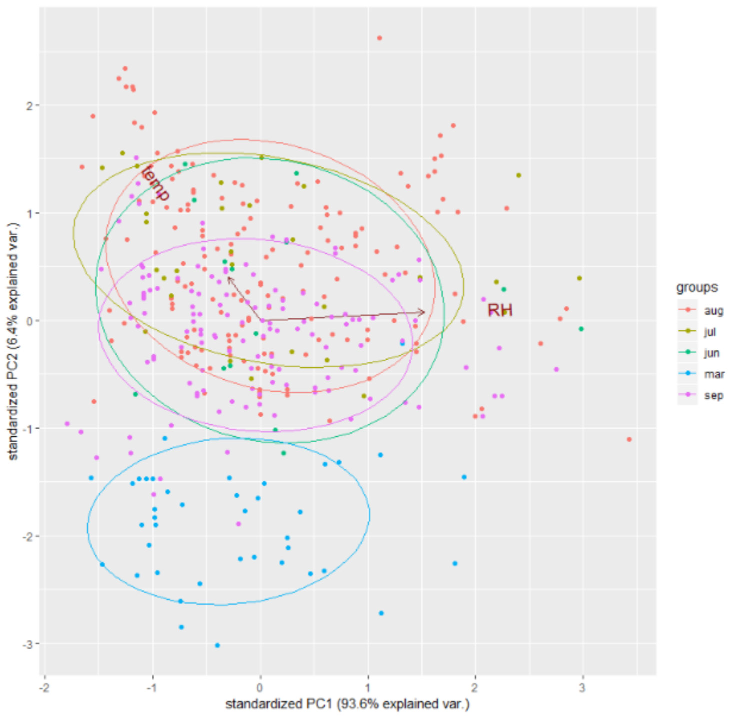

Figure 2: This grouping is based on temperature and humidity PCA's

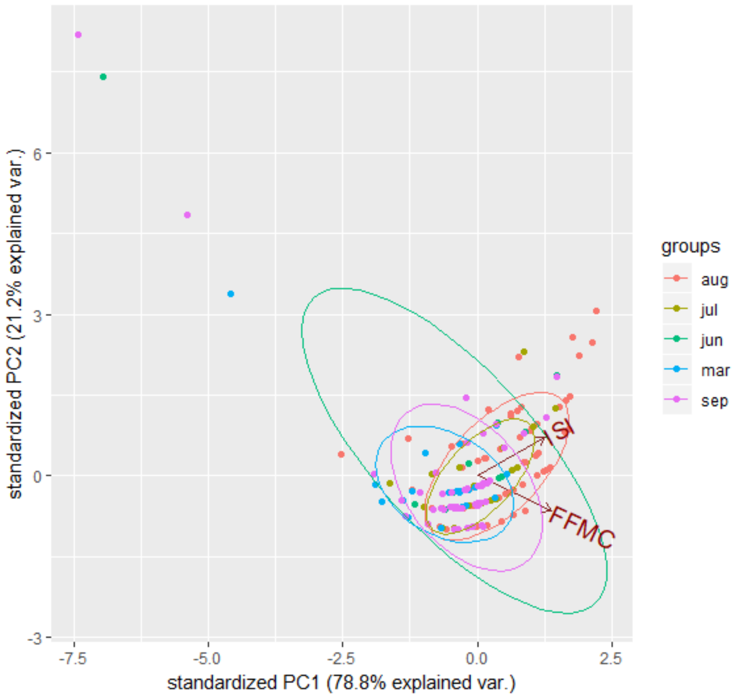

Figure 3: This grouping is based on Spread Index and Fine Fuel Moisture Code PCA's

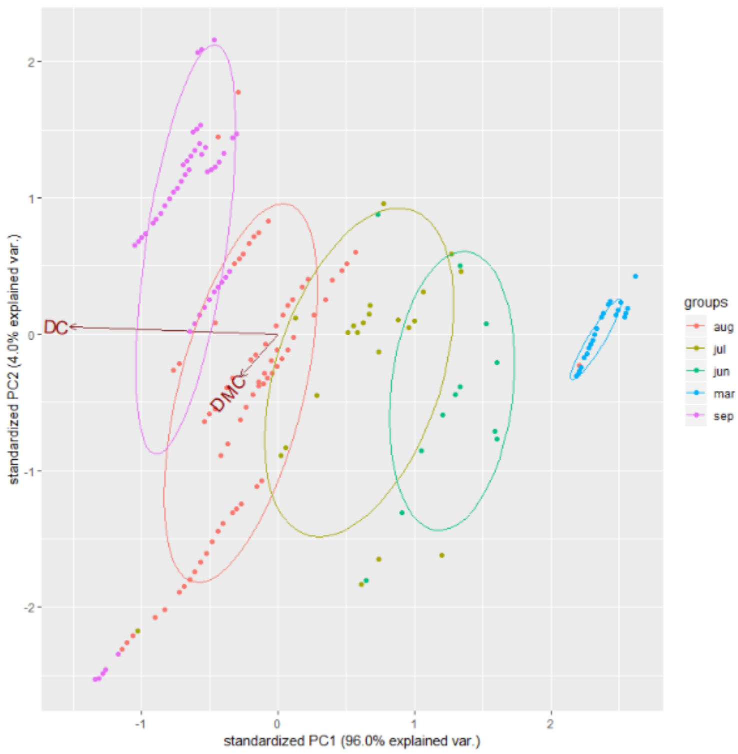

Figure 4: This grouping is based on Drought Code and Duff Moisture Code PCA's

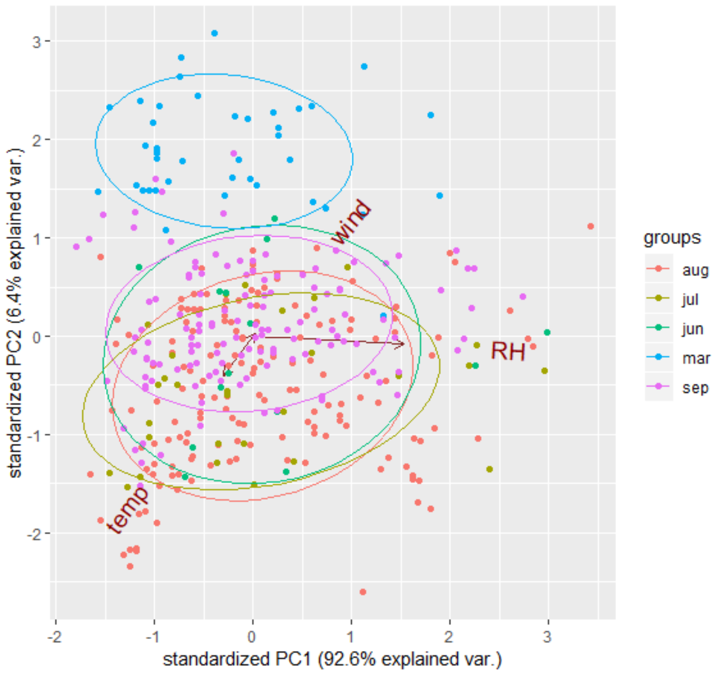

Figure 5: This grouping is based on the temperature, wind, and humidity PCA's

4.3 Limitations

Some limitations that we faced with data would be the area that each fire burned. In the original data set the area burned was described in hectares . However, with a lack of data collected, this was zero in most cases, which skews PCA's if this data is used. If this data included we could predict the size a forest fire could grow to based on the data set.

4.4 Generalize

From everything we collected there a few things that we can generalize. When Fine Fuel Moisture Code, Duff Moisture Code, Drought Code, and Relative Humidity are low this contributes to a better condition for forest fires. When Initial Spread Index, Temperature, and Wind are high this also contributes to a better condition for forest fires. Even if all of the conditions are in the favor of the fire, even a few can cause good forest fire conditions. For example in March, even when it's cold the high winds and low humidity create a condition that's perfect for forest fires.

5 Appendix

All data and code we used can be found here - `https://github.com/zl22good/Forest-Fire-PCA`