Cluster Report on Forest Fires

Carmelo Emmanuele
Zachary Goodsell

April 30, 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

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.

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 cluster's on this data we set we used a line graph to find the elbow in our data set on forest fires. With the cluster plot we found that the following months that had forest fires and 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 cluster plot for the data set.

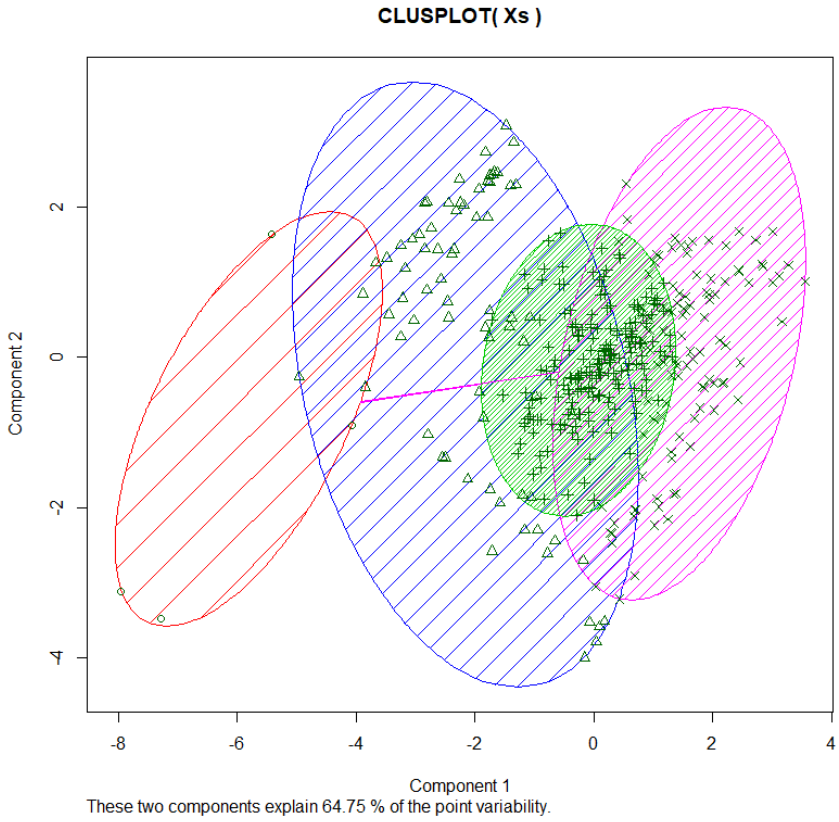

Figure 1: This is the cluster plot based on the months forest fires occurred

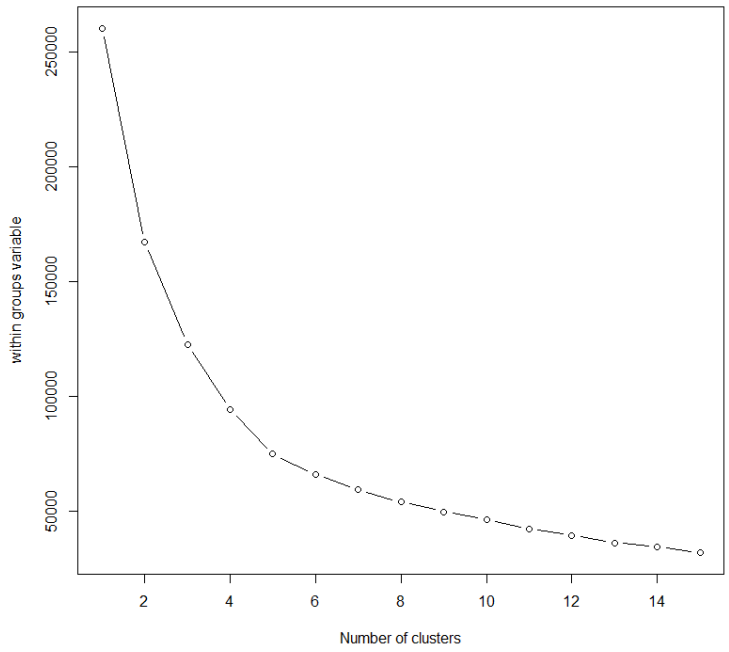

Figure 2: This figure shows the bend in our data set

3.3 Classification

The groupings is based on the cluster of what month the forest fire happened. Then, the individual shapes show the actual months that were recorded for a forest fire.

4 Results

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. All the month clusters successfully included its respective months.

5 Appendix

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