

# Assignment 1.3 Inforgraphic about Wildfire

Long Pham

2023-04-01

## Open and attach data source

Imported forestfires file using RStudio function. Attached the data from table for easier use.

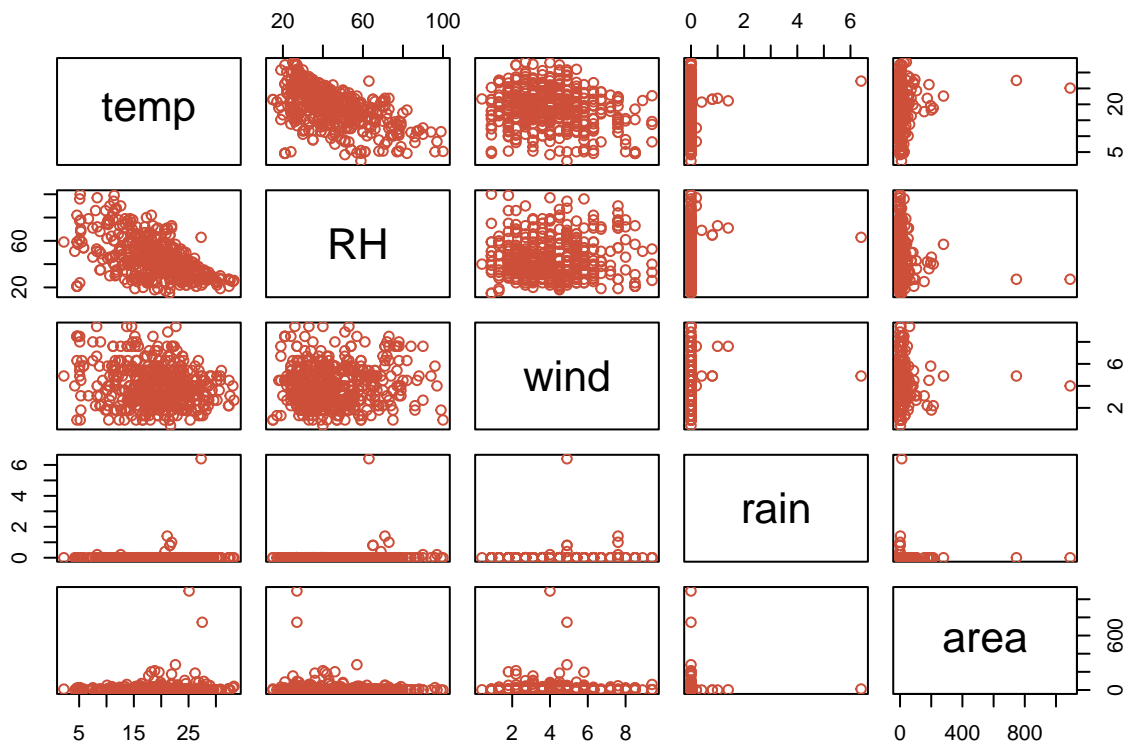
```
forestfires <-  
  read.csv("C:/Users/phaml/OneDrive/Langara/4800 Data Analysis and Stat Infer/Week 2/Week 2.2/forestfires.csv")  
attach(forestfires)
```

## Testing relevance between variables:

Temp, RH, wind, rain as independent variables Area as dependent variables

## Using scatter plot

```
plot(forestfires[9:13], col='tomato3')
```

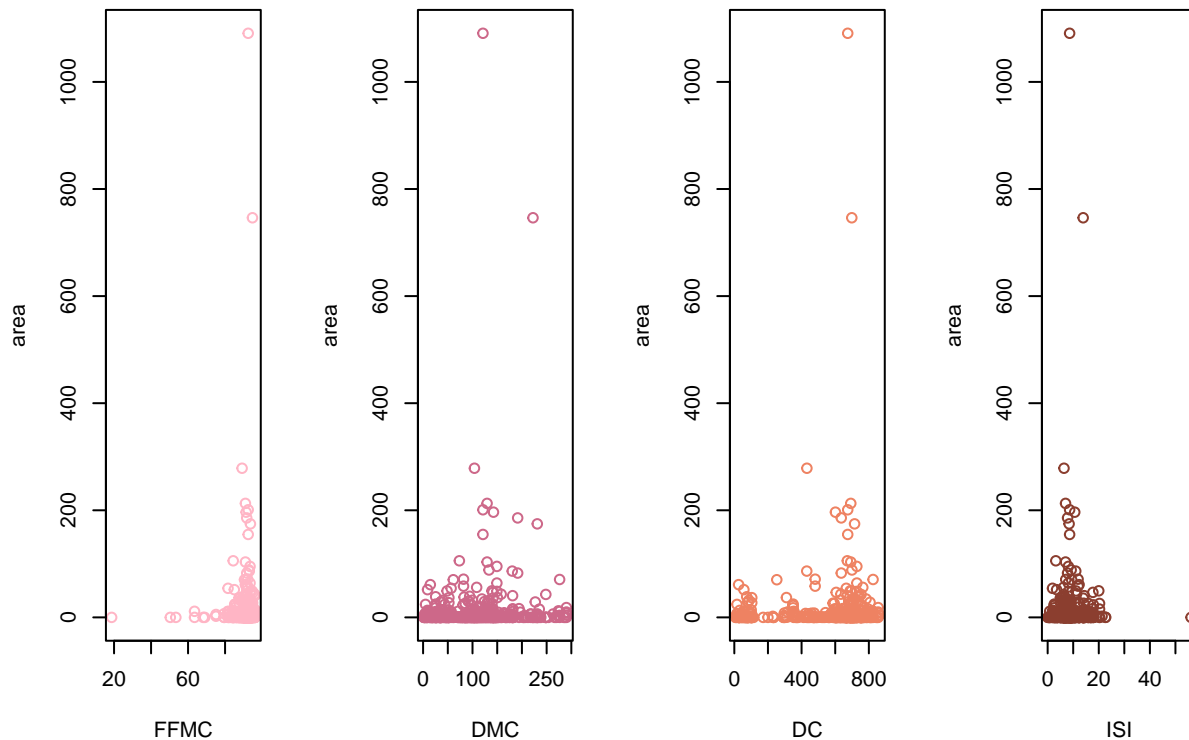


As we can see from the graph, the correlations between fire area regarding of other independent values are not strong. Hence, we can conclude that none of the factors singlehandedly affect the fire area.

## FWI indexes test

Next we test the correlations between indexes from Fire Weather Index model to see if any of this index has a strong relationship with the burned area or not.

```
par(mfrow=c(1,4))
plot(FFMC,area, col="pink1")
plot(DMC, area, col="palevioletred3")
plot(DC, area, col="salmon2")
plot(ISI, area, col="coral4")
```

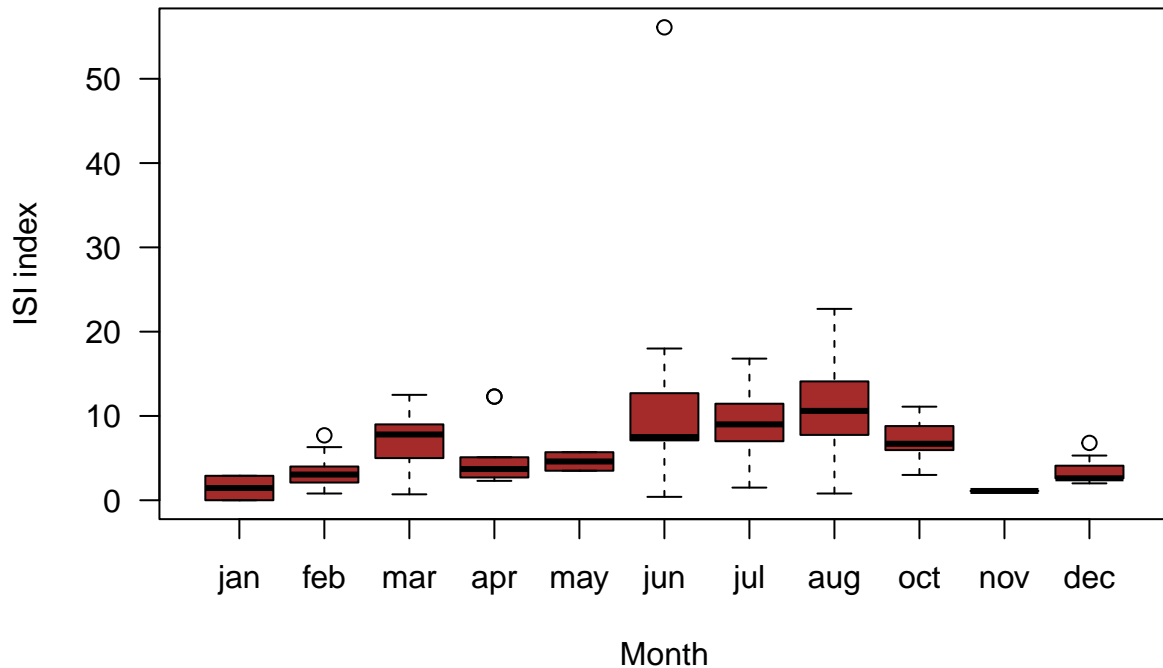


As we can see we have the same result with the previous test above. Therefore, no single variable affect directly to the burned area.

## Which month could spread the fire the fastest

Up next, we gonna use variable month to check the independent variable ISI. ISI is Initial Spread Index in the FWI model, represents how fast the fire get spread.

```
x1 <- factor(forestfires$month, levels=
             c("jan", "feb", "mar", "apr", "may", "jun", "jul", "aug", "oct", "nov", "dec"))
plot(ISI ~ x1, xlab="Month", ylab="ISI index", las=1, col="brown")
```



As we can see from the graph, the summer months including Jun, July and August have the most suitable condition for the fire to spread.

## Accepting result from the original reports of Paulo Cortez (A Data Mining Approach to Predict Forest Fires using Meteorological Data, 2007)

One of the result I want to use for the purpose of the info graphic is table 4 in his report. The table indicates 4 variables affect the forest fires are: Temperature (73.2%), Rain (18.0%), wind (4.7%) and humidity (4.1%).

## Conclusion

My own take from the analysis is that not a single natural factor is the main cause for forest fires. As the matter of fact, it is the combination of many factors together and would be proven with different analysis method (not included in here). However, we can see which month has the highest ISI - index indicates spreading speed of fire. Last but not least, the The sensitivity analysis from Paulo Cortez was mentioned to point out that even despite being a single cause to forest fire, temperature, rain, wind and humidity all play a role in wildfire.

## Reference

Cortez, P. (2007). A Data Mining Approach to Predict Forest Fires using Meteorological Data. University of Minho.