

Katherine Le

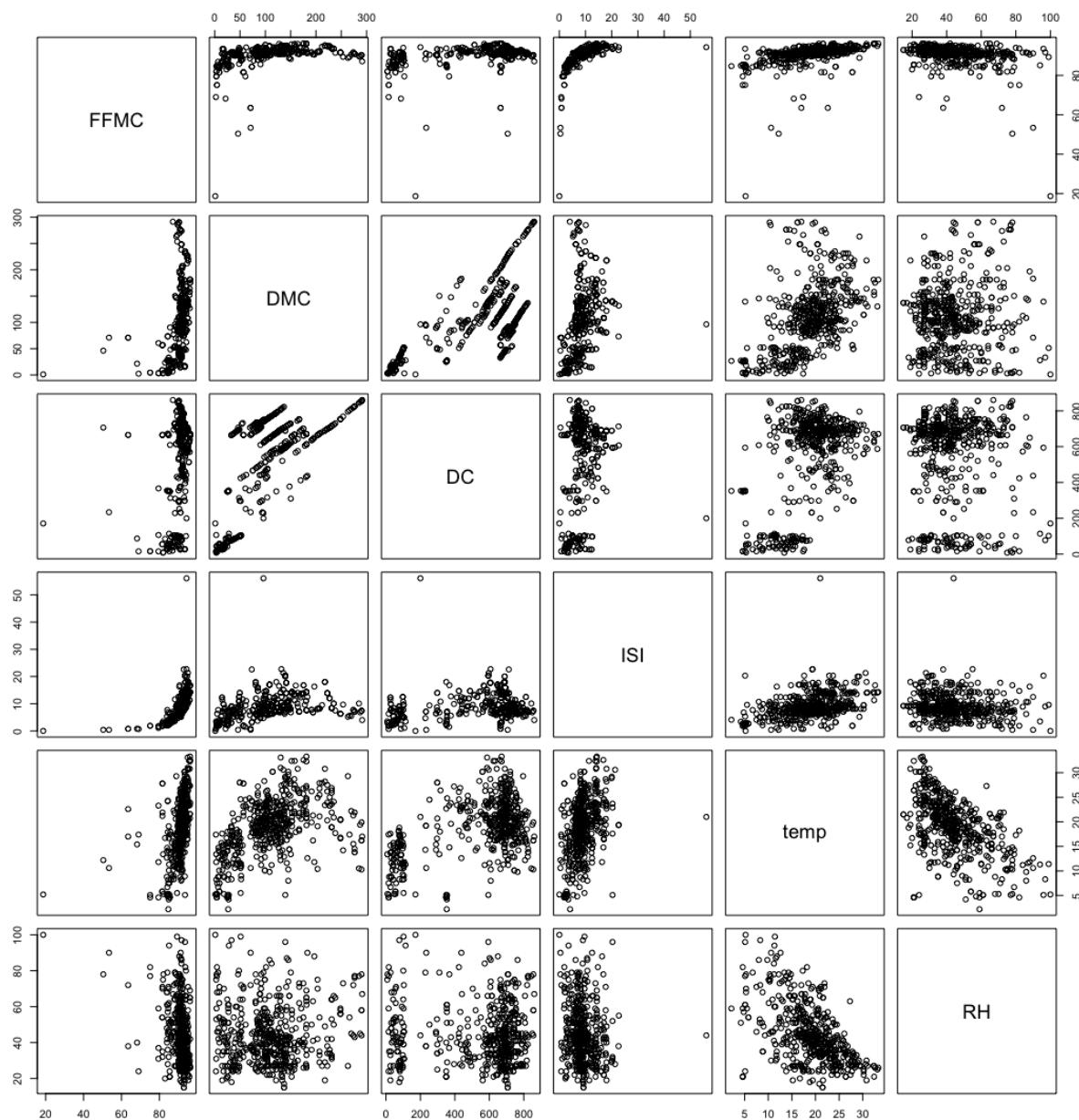
Professor Kevin Bailey

STA 2260

8 December 2024

Forest Fires in the Northeast Region of Portugal

This dataset from the UC Irvine Machine Learning Repository was used in linear regression models to predict forest fire damage in northeastern Portugal through various factors. After going through and plotting all the numerical data, the response, ISI, had some relationship with FFMC, DMC, DC, temp, and RH.



According to the information provided by the dataset and ClimateData.ca:

The FWI system is the Fire Weather Index (FWI) System.

FFMC index from the FWI system and ranged from 18.7 to 96.20 during the data collection period

Fine Fuel Moisture Code: "the dryness of fuels on the surface of the forest floor.

This layer is most likely to catch fire due to human activity."

DMC index ranging from 1.1 to 291.3

Duff Moisture Code: "fuel dryness just under the forest floor. Fuel in this layer is often ignited during lightning strikes."

DC index going from 7.9 to 860.6

Drought Code: "fuel dryness deeper in the forest floor. The dryness of this layer factors into how hard it is to put out deep burning wildfires."

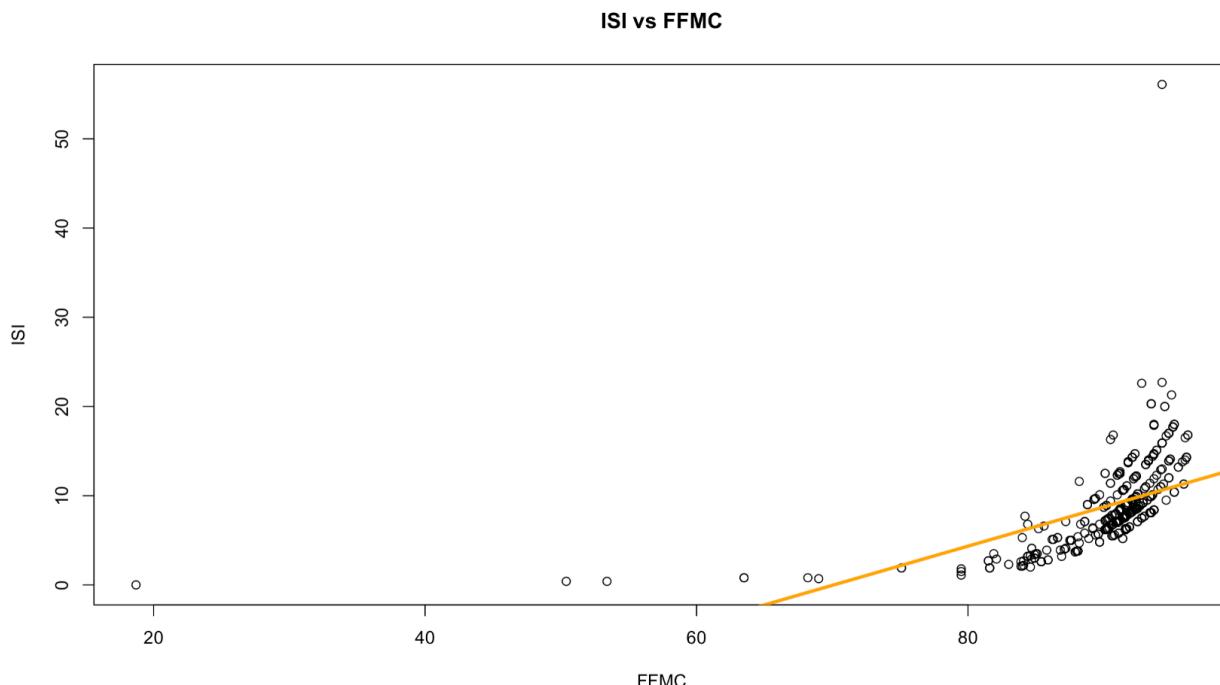
ISI index with a low of 0.0 and a high of 56.10

Initial Spread Index: "ISI = FFMC + Wind Speed. This index combines wind speed conditions with the surface fuel dryness (FFMC) to estimate how fast fires could spread."

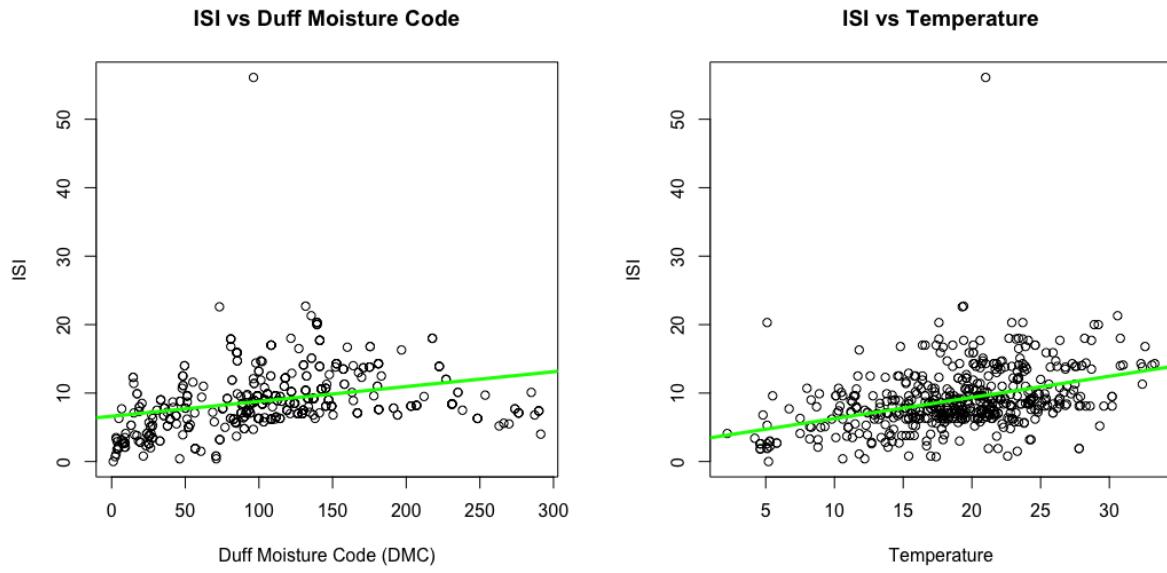
temp is the temperature in degrees Celsius: 2.2 to 33.30

RH stands for relative humidity going from 15% to 100%

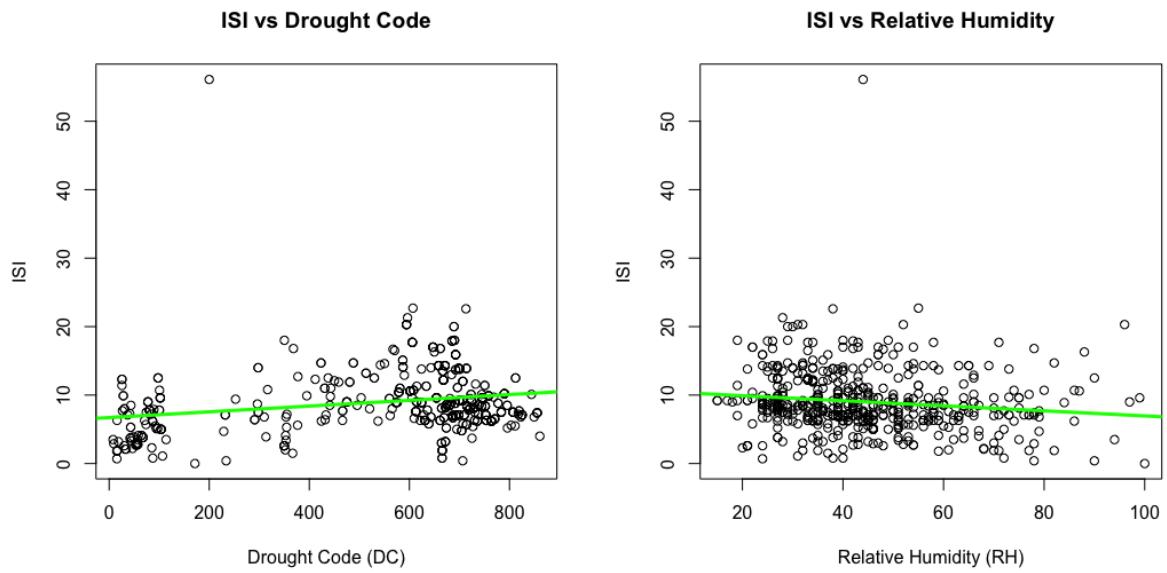
Model 1: From the Coefficient of Determination we see the highest R^2 of 0.2814 out of all the other response to predictors models and lowest AIC of 2869.112. This relationship gives us an equation of: $y = -30.79473 + 0.43926 * \text{FFMC}$



In comparison, ISI vs DMC or ISI vs Temperature could be used as visual indicators of a slight relationship. This is reflected in the R^2 of about 0.1.

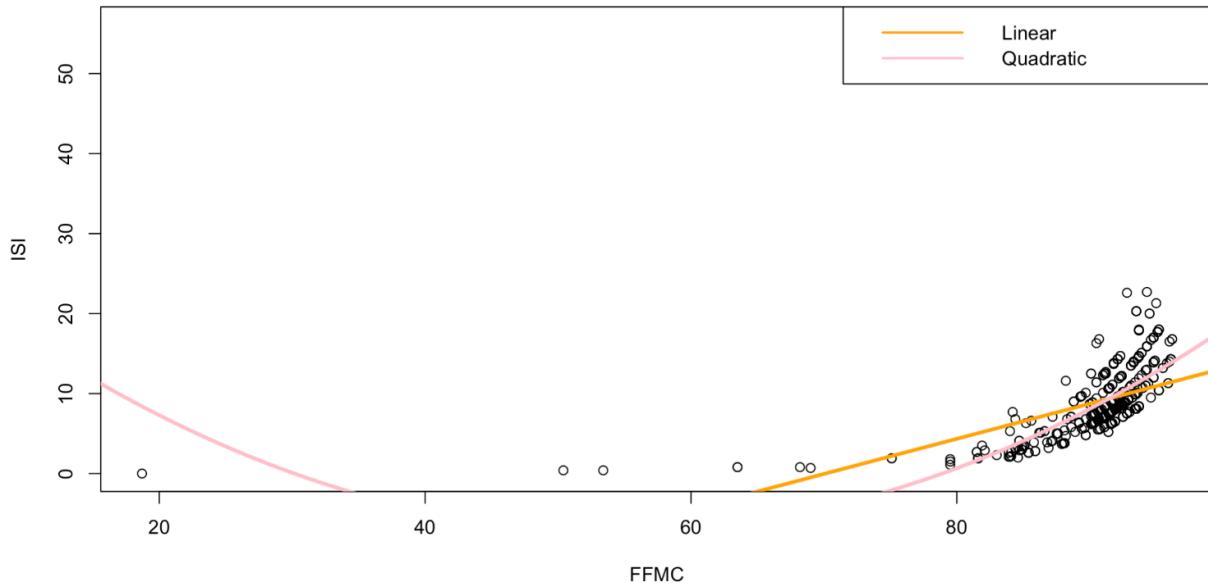


While ISI vs DC and ISI vs RH show little to negative relationship with their R^2 of less than 0.1:

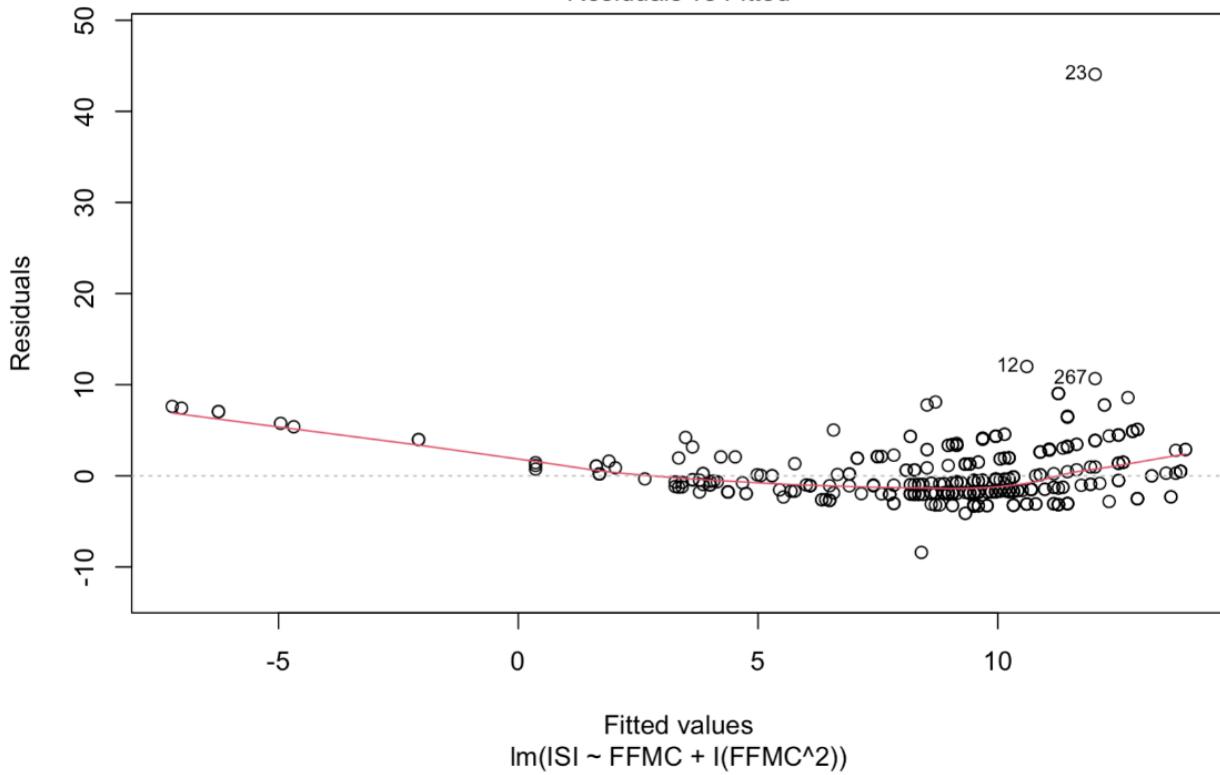


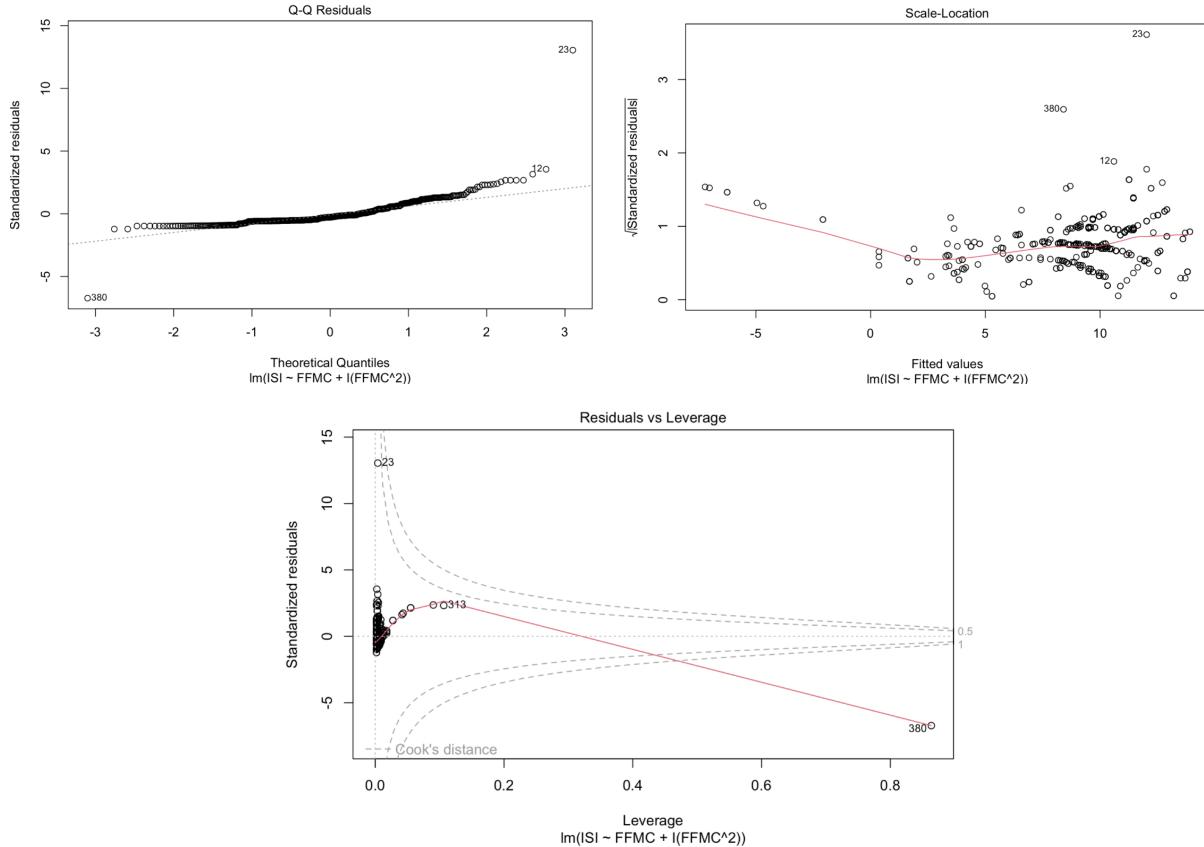
However, we see the linear line is not the best fit for the ISI vs FFMC graph. Calculating the equation for the line $ISI = 28.985850 - 1.328283*FFMC + 0.012179*FFMC^2$ gave us a larger and closer to 1 R^2 of 0.4488 and lower AIC of 2733.053.

ISI vs FFMC



Residuals vs Fitted





Model 2: When looking at ISI vs FFMC and DC, we see that DC is not important, as the R^2 remains close to the value it was when just looking at FFMC as the predictor.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	Signif. codes:
(Intercept)	-2.992e+01	2.855e+00	-10.478	<2e-16 ***	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
FFMC	4.229e-01	3.262e-02	12.965	<2e-16 ***	
DC	1.102e-03	7.258e-04	1.518	0.13	

Residual standard error: 3.86 on 514 degrees of freedom

Multiple R-squared: 0.286, Adjusted R-squared: 0.2832

F-statistic: 103 on 2 and 514 DF, p-value: < 2.2e-16

However, when squaring both predictors, DC and DC² p-value show they are just as important as the FFMC and FFMC². Their combined R² goes up to about 0.47 in comparison to FFMC and FFMC²'s R² of about 0.45.

```

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 2.632e+01 5.400e+00 4.873 1.47e-06 ***
FFMC       -1.275e+00 1.459e-01 -8.735 < 2e-16 ***
I(FFMC^2)   1.182e-02 1.010e-03 11.697 < 2e-16 ***
DC          1.136e-02 2.864e-03  3.965 8.39e-05 ***
I(DC^2)     -1.495e-05 3.312e-06 -4.514 7.90e-06 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.315 on 512 degrees of freedom
Multiple R-squared:  0.4755,    Adjusted R-squared:  0.4714 
F-statistic: 116 on 4 and 512 DF,  p-value: < 2.2e-16

```

In comparison, FFMC and Wind did not appear to be important factors together, however their squared values gave various p-values significance, but the best R^2 and lowest AIC of any calculated.

```

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 29.3336813 5.2353288 5.603 3.44e-08 ***
FFMC        -1.4100497 0.1405894 -10.030 < 2e-16 ***
I(FFMC^2)   0.0127472 0.0009522 13.388 < 2e-16 ***
wind         0.7698500 0.3249614  2.369  0.0182 *  
I(wind^2)   -0.0369941 0.0347816 -1.064  0.2880  
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.298 on 512 degrees of freedom
Multiple R-squared:  0.4809,    Adjusted R-squared:  0.4768 
F-statistic: 118.6 on 4 and 512 DF,  p-value: < 2.2e-16

> # Adjusted R-squared: 0.4768
> AIC(m2.ffdmc.wind, m2.ffdmc.wind.sq)
      df      AIC
m2.ffdmc.wind     4 2860.265
m2.ffdmc.wind.sq  6 2708.006
> |

```

Comparison: Considering the model 1 for $ISI = 28.985850 - 1.328283*FFMC + 0.012179*FFMC^2$ gave us an R^2 of 0.4488 and AIC of 2733.053 using a single predictor for a response and the model 2 using FFMC and wind gave us $ISI = 29.3336813 - 1.4100497*FFMC + 0.0127472*FFMC^2 + 0.7698500*wind - 0.0369941*wind^2$ where $R^2 = 0.4768$ and AIC = 2708.006, model 2 would be the better predictor of which predictors resulted in the greatest response of Initial Speed Index since it has similar but better R^2 and AIC values.

Works Cited

“UCI Machine Learning Repository.” *Archive.ics.uci.edu*,
archive.ics.uci.edu/dataset/162/forest+fires.

“Understanding Fire Weather and Climate Change Basics — ClimateData.ca.” *ClimateData.ca*,
16 Aug. 2024,
climatedata.ca/resource/understanding-fire-weather-and-climate-change-basics/.