Regression Analysis to Determine Area of Forest Fires Using Meteorological Data

Rosa Soriano August 3 2022

STA 4164-B001 University of Central Florida rosasoriano9884@knights.ucf.edu

Abstract

As forest fires pose as an environment hazard, action on predicting the natural disaster is constantly being researched. Our dataset provided is composed of 518 observations collected from Montesinho natural park in Portugal during different months and days of the week. To have effective fire rescue response, our data set is collected using data mining. The rise of technology allows to have access to data including Multiple Regression Model, Decision Trees, Random Forests, Neural Networks, and Support Vector Machines. For this analysis, we are going to use the aspects of the Linear Regression Analysis and Multiple Regression Analysis to predict the likelihood of a forest fire's occurrence based on a research study's predictor variables conducted on data mining. The Multiple Regression Model will give us a traditional review of how model is related to in a forest fire's performance.

Objectives

The focus will be on determining which variables are the best in predicting area a forest fire will cover. We are determining how strong these variables in our projected models. Our Multiple Regression Model will visually display all these answers. What else can we learn from applying descriptive statistics in this study? Is this case study resolved or needs further work?

Source of Data

This data comes from Kaggle which is an online community of those who work in data science and machine learning. This platform consists of databases, code, discussions, courses, and competitions. Our dataset specifically is compiled by "JACKSONKWONG" from the research article "A Data Mining Approach to Predict Forest Fires using Meteorological Data". The dataset originates from Manuel Rainha's work in specifically spatial, temporal and FWI data and Braganca Polytechnic Institute's work in the weather station database.

Environment of Study

The Regression Analysis will be conducted on Windows 10 Version 21H2 for x64-based Systems. R studio along with R coding will be used. This version is the latest R-4.2.1 for Windows. Updates are available for both Windows and R when this was created.

Data Exploration

The variables FFMC pertains to moisture content surface litter code index measurement. DMC is shallow organic layer code index measurement. DC is deep organic layer code index measurement. ISI is wind velocity code index measurement. Temperature is in Celsius. Relative Humidity is in percentage. Wind speed is km/h. Area is the portion of land burned in hectares.

Let us explore the frequencies of our dataset. These values have major value differences which gives the possibility of outliers. The averages of each variable are as follows: FFMC: 90.64468, DMC: 110.8732, DC: 547.94, ISI: 9.021663, temp:18.88917, RH: 44.2882, wind: 4.017602, rain: 0.021663, and area: 12.84729. The ranges of each variable are as follows: FFMC: 77.5, DMC: 290.2, DC: 852.7, ISI: 56.1, temp: 31.1, RH: 85, wind: 9, rain: 6.4, and area: 1090.84. The medians of each variable are as follows: FFMC: 91.6, DMC: 108.3, DC: 664.2, ISI: 8.4, temp: 19.3, RH: 42, wind: 4, rain: 0, and area: 0.52. The single modes of each variable are as follows: FFMC: 91.6, DMC: 99, DC: 745.3, ISI: 9.6 temp: 19.6, RH: 27, wind: 3.1, rain: 0, and area: 0. The multiple modes of each variable are as follows: FFMC: 91.6 and 92.1, DMC: 99, DC: 745.3, ISI: 9.6 temp: 17.4 and 19.6, RH: 27, wind: 2.2 and 3.1, rain: 0, and area: 0.

Boxplots (see Appendix A), histograms (see Appendix B), and scatterplots (see Appendix C) are presented for each variable in the dataset. Overall, our data has an abundance of data, however, it is generally loose. Our Adjusted R squares and Multiple R squares appear to be close to 0. Area likely to be covered in a forest fire is mostly dependent on higher temperatures, lower measurement of rain, and lower relative humidity. Higher levels of moisture content surface litter and shallow and deep organic layers are also noticeable. Fire velocity and wind speed are varied. The highest frequency in moisture content surface litter and shallow and deep organic layers are also noticeable. Lower frequencies occur with rain and fire velocity. Every other predictor variable has high frequency but varied. Area overall is not catastrophic. Positive correlations are noted in data relating to in temperature, moisture content surface litter, shallow layers, and deep organic layers. Negative correlations are noted in relative humidity, rain, and fire velocity.

A scatterplot matrix visually presents all the combinations of scatterplots for each variable.

The five linear regression assumptions include existence, independence, linearity, homoscedasticity, and normal distribution will need to be considered to have a valid result. All values exist and have finite values. All dependent values are resulted on its own result. All functions are linear. Variance of each value is the same. Dependent values have normal distributions on the linear regression lines.

Collinearity exists when the independent variables have an intense linear relationship. This is measured if Variance Inflation Factor (VIF) exceeds a value of 5 and 10. A value above 5 may need to be observed while 10 indicates a critical collinearity. This is highly dependent on the respective dataset. None of variables have high collinearity in all three models (see Appendix F).

Data Preparation

Although the dataset is abundant, we still have potential values that can be considered for the purpose of the study. This includes all combinations of x and y coordinates of the map as well as month and day of the week. This dataset, however, is claimed by the data compiler to have no missing attribute values.

The dataset is currently composed of many outliers. Our boxplots have many outliers that the graphs can give a false appearance to be scatterplots. Removing a few outliers will not simply improve the linear relationships.

Possible transformations can be data collected from this research paper, the dataset complier, and authors of the primary case study. Additional data exploration is internally transformed to search and compute values. The risk for human error still exists.

If someone wanted to use a dummy variable, it would be used as a categorical result of the area. That can mean if the area burned is critical or not, deadly or not, or catastrophic or not. This would represented as 0 not present or 1 as present. Since our primary dependent variable was quantitative, we did not need to present one at this time.

Modeling

There are three univariate models that have been developed (see Appendix D). The corresponding ANOVA tables for these models and variables are also presented (see Appendix E). Model #1 is a comprehensive model while Model #2 and Model #3 come from significant predictor variables.

Model #1 has Multiple R squared value of 0.016 and Adjusted R squared value of 0.000517. It has the highest Multiple R squared value for containing all the predictor variables. Model #2 has Multiple R squared value of 0.014 and Adjusted R squared value of 0.00432. Model #3 has Multiple R-squared value of 0.0105, Adjusted R-squared value of 0.00468. Both R squares appear to be close in value and not a major concern of fit to the regression lines. Multiple regression data. All models have high values of residuals due to plenty of outliers. Model #1 has residuals from -31.1 to -15.5 (1Q) to -9.4 (Median) to -0.8 (3Q) to 1069.0. Model #2 has residuals from -28.4 to -14.8(1Q) to -9.8 (Median) to -1.1(3Q) to 1069.6. Model #3 has residuals from -27.9 to -14.5(1Q) to -10.4(Median) to -2.4(3Q) to 1070.3. Relative Humidity, temperature, and rain have mostly great interactions according to the Multiple Regression ANOVA Table. Confounding variables can be oxygen and fuel levels that are associated with higher temperatures. Confounding variables can also be lack of proximity to a body of water for relative humidity and rain. DMC and ISI both have greater codex index measurements are also affected by confounding variables pertaining to the water cycle. The organic layer is involved with infiltration while the wind velocity is involved with evaporation.

My comprehensive model includes all variables which is Model #l. Model selection comes from variables with significant impact to the line that best fits the data points. Model #1 is area=2.49381-0.02331(FFMC)+0.07649(DMC)-0.00569(DC)-0.69844(ISI)+0.84795(temp)-0.19629(RH)+1.52710(wind)-2.54001(rain)+E. Model #2 is area=9.2882+0.06040(DMC)-0.5538(ISI)+0.9533(temp)-0.2193(RH)-1.6456(rain)+E. Model #3 is area=1.302-0.122(RH)-2.140(rain)+0.900(temp)+E. Each value is a beta value in our linear regression equation. We can determine significance by performing hypothesis testing using the F-test found in the Multiple regression ANOVA table. Each model can have a hypothesis test using t values to individually test for the significance of each slope.

Model diagnostics includes the process collecting data, compiling a dataset, determine the details of the descriptive statistics, finding regression lines, determine details of Linear Regression Analysis, determine details of Multiple Regression Analysis, and deciding on the best model. This process requires overlooking potential errors, discovering all possible data, and determining its application to the respective case study.

As the last part for our model diagnostics, we must determine the best fit model. The best model would have to have most these values closest to the regression line.

It has to meet certain criteria for Linear Regression as well. It looks as if Model #3 will be a good fit.

Discussion

Model #3 is reduced to three significant variables which again meets all five major linear regression assumptions including existence, independence, linearity, homoscedasticity, and normal distribution. These predictor variables have the most impact to determine how much land will be covered in a normal fire. My model performance is loosely accurate; however, more observations may be needed. The authors of the primary study yearn for more data and testing to further the status of the study.

Conclusion

Since what is presented in our regression analysis, we can determine that forest fires have a noticeable impact on the environment's destruction. Model #3 has the best fitted model with an Adjusted R-square value of 0.00468. Model #1 has all predictor variables with a Multiple R-square value at 0.016. Our correlations are not very strong, however, due to lack of additional research. RH, rain, and temperature are all statistically significant. DMC and ISI also important codes to consider when predicting most fires, usually not catastrophic. At the end of the matter, the author of the related study recommends more research to enhance the subject of the matter.

References (APA 7)

A data mining approach to predict forest fires using ... - uminho. (2007). Retrieved July 18, 2022, from http://www3.dsi.uminho.pt/pcortez/fires.pdf

JacksonKwong. (2022, June 3). *Forest fires*. Retrieved July 18, 2022, from https://www.kaggle.com/datasets/jacksonkwong/forest-fires

Kleinbaum, D. G. (2014). *Applied regression analysis and other multivariable methods*. Cengage Learning.

Appendix A

Boxplots

Boxplots for each predictor variable is presented. The boxplots are miniature with multiple outliers. The data is overall loose. Area likely to be covered in a forest fire is mostly dependent on higher temperatures, lower measurement of rain, and lower relative humidity. Higher levels of moisture content surface litter and shallow and deep organic layers are also noticeable. Fire velocity and wind speed are varied.

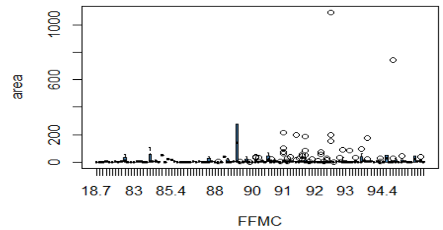


Figure 1A

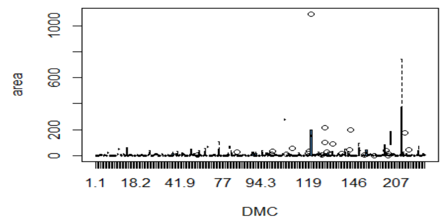


Figure 2A

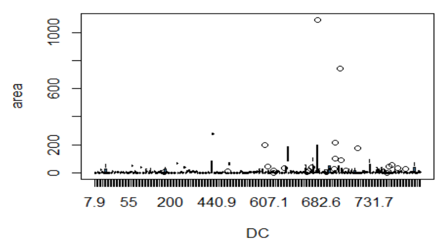


Figure 3A

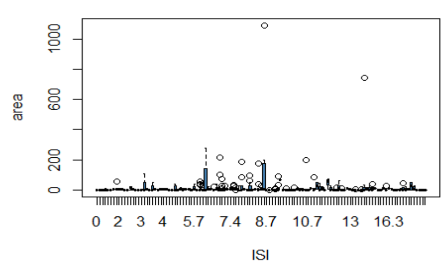


Figure 4A

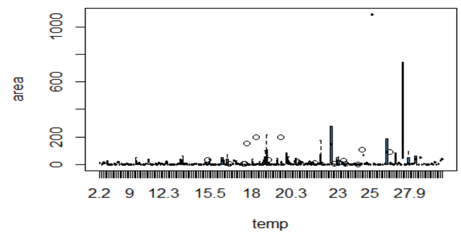


Figure 5A

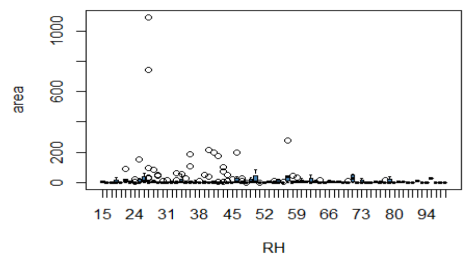


Figure 6A

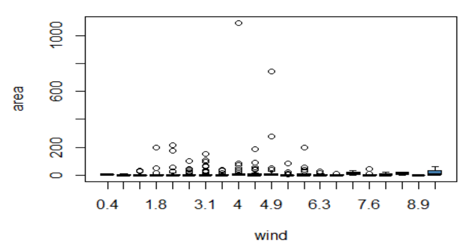


Figure 7A

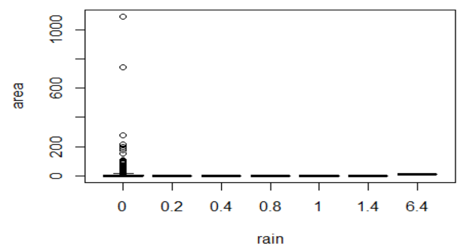
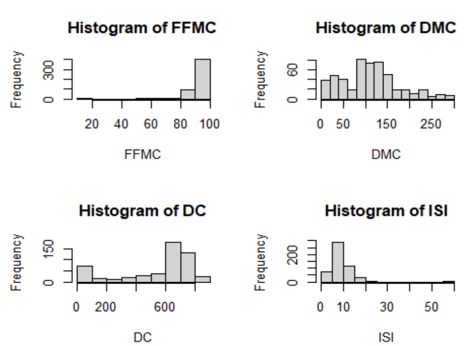


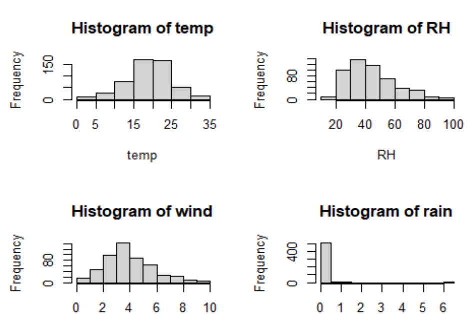
Figure 8A

Appendix B

Histograms

Histograms for each predictor variable is presented. The highest frequency in moisture content surface litter and shallow and deep organic layers are also noticeable. Lower frequencies occur with rain and fire velocity. Every other predictor variable has high frequency but varied. The data is overall loose. Area overall is not catastrophic.





Appendix C

Scatterplots

Scatterplots for each predictor variable is presented. Positive correlations are noted in data relating to in temperature, moisture content surface litter, shallow layers, and deep organic layers. Negative correlations are noted in relative humidity, rain and fire velocity. The data is overall loose. Scatterplot matrix visually presents all the combinations of scatterplots for each variable.

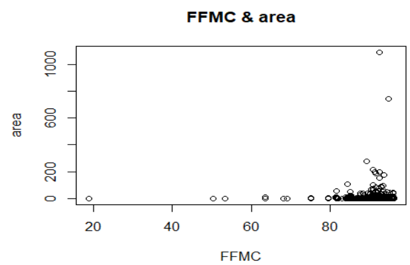


Figure 1C

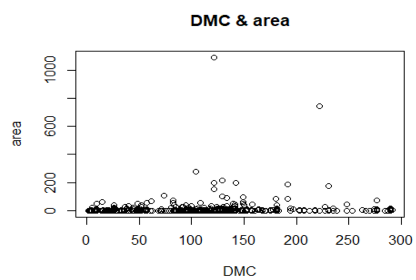


Figure 2C

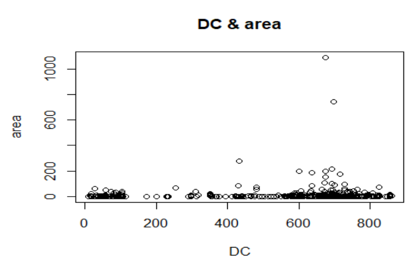


Figure 3C

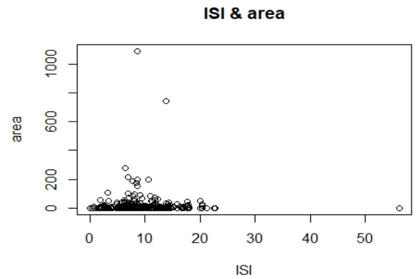


Figure 4C

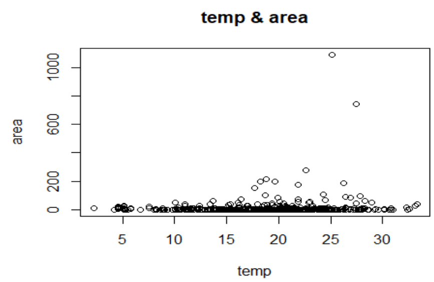


Figure 5C

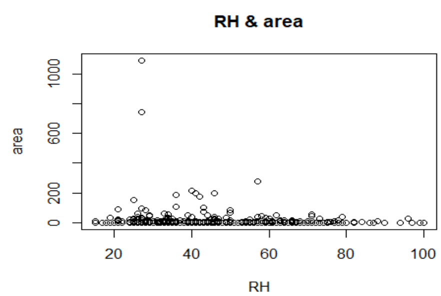


Figure 6C

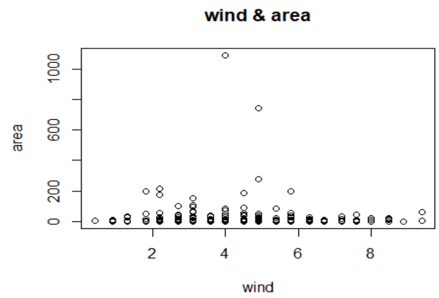


Figure 7C

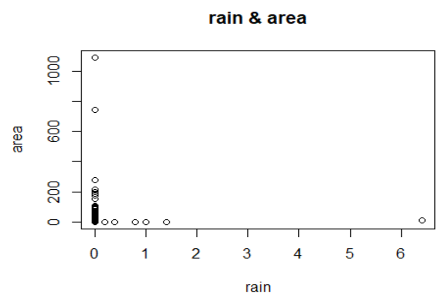


Figure 8C

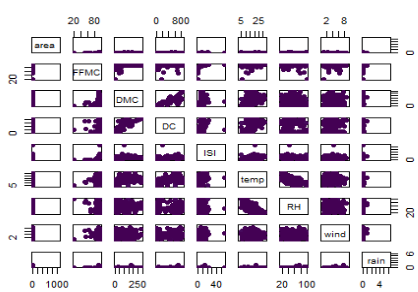


Figure 9C

Appendix D

Linear Regression Graphs

Linear Regression graphs consists on the three proposed models. Every model has a weak negative correlation with outliers. Most values are in accord with the fit line in Model #3. The data is overall loose.

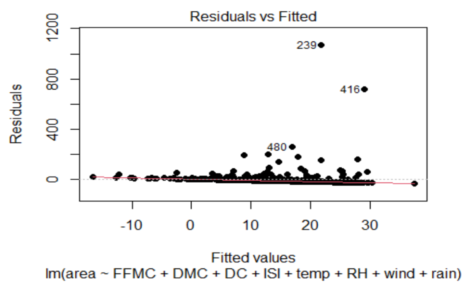


Figure 1D

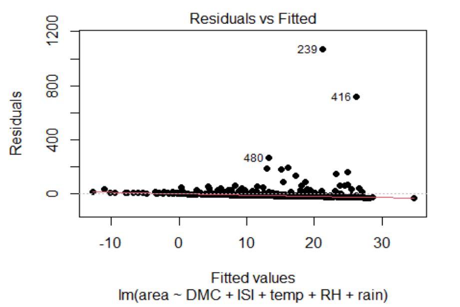


Figure 2D

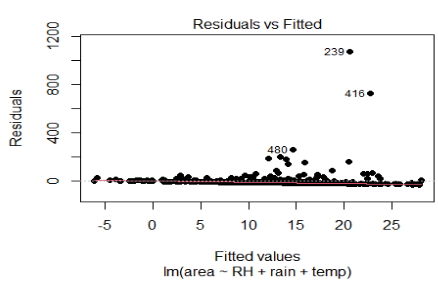


Figure 3D

Appendix E

ANOVA Tables

ANOVA tables of every predictor variable, Linear Regression model, and Multiple Regression model. This is useful for hypothesis testing as well as comparing of statistics. The data is overall loose.

```
## Analysis of Variance Table
##
## Response: area
##            Df  Sum Sq Mean Sq F value Pr(>F)
## FFMC        1    3366    3366    0.83   0.36
## Residuals 515 2087499    4053
```

```
summary(lm_1)
```

```
##
## Call:
## lm(formula = area ~ FFMC, data = Fires)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -15.4  -13.3  -11.8   -5.8 1077.1
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -29.091     46.109   -0.63     0.53
## FFMC           0.463      0.508    0.91     0.36
##
## Residual standard error: 63.7 on 515 degrees of freedom
## Multiple R-squared:  0.00161,    Adjusted R-squared:  -0.000329
## F-statistic: 0.83 on 1 and 515 DF,  p-value: 0.363
```

Figure 1E

```
## Analysis of Variance Table
##
## Response: area
```

```
##              Df  Sum Sq Mean Sq F value Pr(>F)
## DMC            1   11140   11140    2.76  0.097 .
## Residuals 515 2079724    4038
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(lm_2)
```

```
##
## Call:
## lm(formula = area ~ DMC, data = Fires)
##
## Residuals:
##     Min      1Q Median      3Q     Max
##   -25.8   -13.5   -10.1    -5.1 1077.3
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.8036     5.5915    0.86    0.391
## DMC            0.0725     0.0437    1.66    0.097 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 63.5 on 515 degrees of freedom
## Multiple R-squared:  0.00533,    Adjusted R-squared:  0.0034
## F-statistic: 2.76 on 1 and 515 DF,  p-value: 0.0973
```

Figure 2E

```
## Analysis of Variance Table
##
## Response: area
##              Df  Sum Sq Mean Sq F value Pr(>F)
## DC             1    5099    5099    1.26   0.26
## Residuals 515 2085766    4050
```

```
summary(lm_3)
```

```
##
## Call:
## lm(formula = area ~ DC, data = Fires)
```

```
##
## Residuals:
##     Min      1Q Median      3Q     Max
##   -16.7   -14.3  -10.9    -5.4 1076.4
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     5.9037     6.7918    0.87     0.39
## DC              0.0127     0.0113    1.12     0.26
##
## Residual standard error: 63.6 on 515 degrees of freedom
## Multiple R-squared:  0.00244,	Adjusted R-squared:  0.000502
## F-statistic: 1.26 on 1 and 515 DF,  p-value: 0.262
```

Figure 3E

```
## Analysis of Variance Table
##
## Response: area
##            Df  Sum Sq Mean Sq F value Pr(>F)
## ISI         1     143     143    0.04   0.85
## Residuals 515 2090722    4060
```

```
summary(lm_4)
```

```
##
## Call:
## lm(formula = area ~ ISI, data = Fires)
##
## Residuals:
##     Min      1Q Median      3Q     Max
##   -18.3   -12.8  -12.1    -6.2 1078.0
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     11.807      6.217    1.90    0.058 .
## ISI              0.115      0.615    0.19    0.851
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 63.7 on 515 degrees of freedom
```

```
## Multiple R-squared:  6.82e-05,   Adjusted R-squared:  -0.00187
## F-statistic: 0.0351 on 1 and 515 DF,  p-value: 0.851
```

Figure 4E

```
## Analysis of Variance Table
##
## Response: area
##            Df  Sum Sq Mean Sq F value Pr(>F)
## temp        1   20017   20017    4.98  0.026 *
## Residuals 515 2070848    4021
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(lm_5)
```

```
##
## Call:
## lm(formula = area ~ temp, data = Fires)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -27.3  -14.7  -10.4   -3.4 1071.3
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -7.414      9.500   -0.78    0.435
## temp           1.073      0.481    2.23    0.026 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 63.4 on 515 degrees of freedom
## Multiple R-squared:  0.00957,    Adjusted R-squared:  0.00765
## F-statistic: 4.98 on 1 and 515 DF,  p-value: 0.0261
```

Figure 5E

```
## Analysis of Variance Table
##
## Response: area
##            Df  Sum Sq Mean Sq F value Pr(>F)
## RH          1   11924   11924    2.95  0.086 .
```

```
## Residuals 515 2078940    4037
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(lm_6)
```

```
##
## Call:
## lm(formula = area ~ RH, data = Fires)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -21.5  -14.4  -10.6   -3.5 1072.9
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   25.895      8.089    3.20   0.0015 **
## RH            -0.295      0.171   -1.72   0.0863 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 63.5 on 515 degrees of freedom
## Multiple R-squared:  0.0057, Adjusted R-squared:  0.00377
## F-statistic: 2.95 on 1 and 515 DF,  p-value: 0.0863
```

Figure 6E

```
## Analysis of Variance Table
##
## Response: area
##            Df  Sum Sq Mean Sq F value Pr(>F)
## wind        1     317     317    0.08   0.78
## Residuals 515 2090547    4059
```

```
summary(lm_7)
```

```
##
## Call:
## lm(formula = area ~ wind, data = Fires)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -14.8  -12.8  -11.9   -6.1 1078.0
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   11.089      6.885    1.61     0.11
## wind           0.438      1.565    0.28     0.78
##
## Residual standard error: 63.7 on 515 degrees of freedom
## Multiple R-squared:  0.000152,   Adjusted R-squared:  -0.00179
## F-statistic: 0.0781 on 1 and 515 DF,  p-value: 0.78
```

Figure 7E

```
## Analysis of Variance Table
##
## Response: area
##            Df  Sum Sq Mean Sq F value Pr(>F)
## rain        1     113     113    0.03   0.87
## Residuals 515 2090751    4060
```

```
summary(lm_8)
```

```
##
## Call:
## lm(formula = area ~ rain, data = Fires)
##
## Residuals:
##     Min      1Q Median      3Q     Max
##   -12.9   -12.9  -12.2    -6.3 1078.0
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    12.88       2.81    4.58  5.7e-06 ***
## rain           -1.58       9.48   -0.17     0.87
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 63.7 on 515 degrees of freedom
## Multiple R-squared:  5.43e-05,   Adjusted R-squared:  -0.00189
## F-statistic: 0.0279 on 1 and 515 DF,  p-value: 0.867
```

Figure 8E

```
##
## Call:
## lm(formula = area ~ FFMC + DMC + DC + ISI + temp + RH + wind +
##     rain, data = Fires)
##
## Residuals:
##    Min      1Q Median      3Q     Max
##  -31.1   -15.5   -9.4    -0.8 1069.0
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.49381   62.04813    0.04     0.97
## FFMC        -0.02331    0.66080   -0.04     0.97
## DMC          0.07649    0.06678    1.15     0.25
## DC          -0.00569    0.01628   -0.35     0.73
## ISI         -0.69844    0.77192   -0.90     0.37
## temp         0.84795    0.78718    1.08     0.28
## RH          -0.19629    0.23666   -0.83     0.41
## wind         1.52710    1.67000    0.91     0.36
## rain        -2.54001    9.67578   -0.26     0.79
##
## Residual standard error: 63.6 on 508 degrees of freedom
## Multiple R-squared:  0.016,  Adjusted R-squared:  0.000517
## F-statistic: 1.03 on 8 and 508 DF,  p-value: 0.41
```

Figure 9E

```
##
## Call:
## lm(formula = area ~ DMC + ISI + temp + RH + rain, data = Fires)
##
## Residuals:
##    Min      1Q Median      3Q     Max
##  -28.4   -14.8   -9.8    -1.1 1069.6
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.2882    18.7807    0.49     0.62
## DMC           0.0640     0.0551    1.16     0.25
```

```
## ISI              -0.5538     0.6756   -0.82      0.41
## temp              0.5933     0.7381    0.80      0.42
## RH               -0.2193     0.2261   -0.97      0.33
## rain             -1.6456     9.6036   -0.17      0.86
##
## Residual standard error: 63.5 on 511 degrees of freedom
## Multiple R-squared:  0.014,  Adjusted R-squared:  0.00432
## F-statistic: 1.45 on 5 and 511 DF,  p-value: 0.205
```

Figure 10E

```
## Analysis of Variance Table
##
## Response: area
##            Df  Sum Sq Mean Sq F value Pr(>F)
## RH          1   11924   11924    2.96  0.086 .
## rain        1       0       0    0.00  0.997
## temp        1    9950    9950    2.47  0.117
## Residuals 513 2068990    4033
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(lm_best)
```

```
##
## Call:
## lm(formula = area ~ RH + rain + temp, data = Fires)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -27.9  -14.5  -10.4   -2.4 1070.3
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.302     17.632    0.07     0.94
## RH            -0.122      0.204   -0.60     0.55
## rain          -2.140      9.594   -0.22     0.82
## temp           0.900      0.573    1.57     0.12
##
## Residual standard error: 63.5 on 513 degrees of freedom
## Multiple R-squared:  0.0105, Adjusted R-squared:  0.00468
## F-statistic: 1.81 on 3 and 513 DF,  p-value: 0.145
```

Figure 11E

```
## Analysis of Variance Table
##
## Response: area
##            Df  Sum Sq Mean Sq F value Pr(>F)
## FFMC        1    3366    3366    0.83  0.362
## DMC         1    8139    8139    2.01  0.157
## DC          1       8       8    0.00  0.965
## ISI         1    1184    1184    0.29  0.589
## temp        1   13549   13549    3.35  0.068 .
## RH          1    3707    3707    0.92  0.339
## wind        1    3249    3249    0.80  0.371
## rain        1     279     279    0.07  0.793
## Residuals 508 2057384    4050
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(
  object = lm_sefd
)

## Analysis of Variance Table
##
## Response: area
##            Df  Sum Sq Mean Sq F value Pr(>F)
## RH          1   11924   11924    2.96  0.086 .
## rain        1       0       0    0.00  0.997
## temp        1    9950    9950    2.47  0.117
## Residuals 513 2068990    4033
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(
  object = lm_sefd2
)

## Analysis of Variance Table
##
## Response: area
##            Df  Sum Sq Mean Sq F value Pr(>F)
## DMC         1   11140   11140    2.76  0.097 .
```

```
## ISI          1     453     453    0.11  0.738
## temp         1   13396   13396    3.32  0.069 .
## RH           1    4104    4104    1.02  0.314
## rain         1     118     118    0.03  0.864
## Residuals 511 2061653    4035
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 12E

Appendix F

Variance Inflation Factors

Collinearity is determined to see if data has a strong linear relationship. A value above 5 may need to be observed while 10 indicates a critical collinearity. This is highly dependent on our dataset.

Variables <chr>	Tolerance <dbl>	VIF <dbl>
FFMC	0.5898817	1.695255
DMC	0.4290578	2.330688
DC	0.4811846	2.078205
ISI	0.6336101	1.578258
temp	0.3756719	2.661897
RH	0.5263189	1.899989
wind	0.8767242	1.140610
rain	0.9571203	1.044801

Figure 1F

Variables <chr>	Tolerance <dbl>	VIF <dbl>
DMC	0.6269472	1.595031
ISI	0.8240888	1.213462
temp	0.4256377	2.349416
RH	0.5744461	1.740807
rain	0.9678674	1.033199

Figure 2F

Variables <chr>	Tolerance <dbl>	VIF <dbl>
RH	0.7031642	1.422143
rain	0.9693975	1.031569
temp	0.7068016	1.414824

Figure 3F