LAB 1 - Assessing Voting Challenges

Tanav Thanjavuru, Simran Gill, Ozkan Meric Ozcan

2024-07-02

Contents

1 Introduction

Forest fires not only represent a significant threat towards human life and property, but also affect the ecosystem. The northeast region of Portugal in particular has a history of wildfires because more than a third is covered by forests. Adding in the factor of hot and dry summers can potentially increase the risk of wildfires[1]. Even though wildfires are a normal part of the renewal process of forests, the forests are getting drier and people are living closer to them which causes an increase in safety issues[2]. This is why it is important for us to understand the behavior of wildfires, in particular the speed in which fire spreads.

With this investigation, we can continue to discuss how wildfires affect the environment along with social-economics and explore opportunities to do our part to put in place appropriate measures of safety. Along with produce strategies to minimize the risk of wildfires. This research will interest be in the interest of Environmental Scientists and Ecologists along with Fire Management Agencies who goal of protecting an decreases any through towards human life along with the ecosystem. To create a starting point for this discussion, this research will investigate the following research question:

> *What is the relationship between Fine Fuel Moisture Code (FFMC) and Initial Spread Index (ISI) in understanding wildfire spread dynamics in the northeast region of Portugal?*

FFMC and ISI are key components of the Canadian Forest Fire Weather Index system which are used to assess fire behavior. To provide context, FFMC represents fuel moisture of forest litter fuel under the shade of a forest canopy and ISI is a numeric rating that estimates how quickly a fire will spread after it starts. Exploring the relationship between these two variables can provide valuable insights in what conditions cause rapid wildfire spread. To answer this research question we will perform a bivariate analysis be developing a regression models by iteratively applying variable transformation. Then evaluating both the statistical and practical significance of results.

2 Description of the Data Source

3 Data Wrangling

4 Operationalization

```
data <- read.csv("forestfires.csv")

# shape of the data
head(data)
```

```
##   X Y month day FFMC  DMC    DC  ISI temp RH wind rain area
## 1 7 5   mar fri 86.2 26.2  94.3  5.1  8.2 51  6.7  0.0    0
## 2 7 4   oct tue 90.6 35.4 669.1  6.7 18.0 33  0.9  0.0    0
## 3 7 4   oct sat 90.6 43.7 686.9  6.7 14.6 33  1.3  0.0    0
## 4 8 6   mar fri 91.7 33.3  77.5  9.0  8.3 97  4.0  0.2    0
## 5 8 6   mar sun 89.3 51.3 102.2  9.6 11.4 99  1.8  0.0    0
## 6 8 6   aug sun 92.3 85.3 488.0 14.7 22.2 29  5.4  0.0    0
```

---

[1] The Subday Times. "Portugal wildfires: what to expect if you're travelling in summer 2024" (2024).

[2] National Geographic. "How to live with mega-fires? Portugal's feral forests may hold the secret" (2019).

```r
# do any columns have null values
print(colSums(is.na(data)))
```

```
##       X       Y   month     day    FFMC     DMC      DC     ISI    temp      RH    wind    rain    area
##       0       0       0       0       0       0       0       0       0       0       0       0       0
```
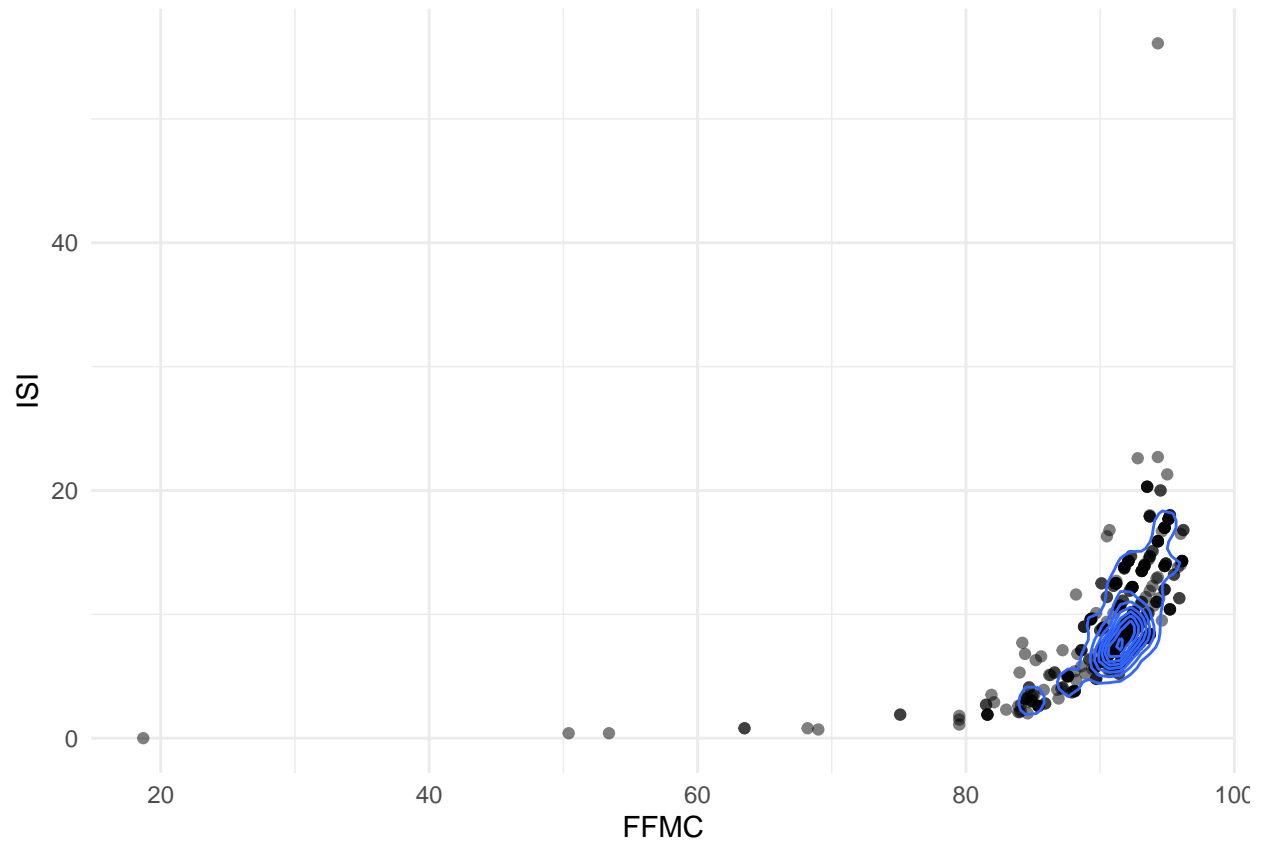
```r
vars <- c("FFMC", "DMC", "DC", "ISI", "temp", "RH", "wind", "rain", "area")
cor_matrix <- cor(data[vars], use = "complete.obs")

corrplot(cor_matrix, type = "upper", order = "hclust",
         tl.col = "black", tl.srt = 45)
```
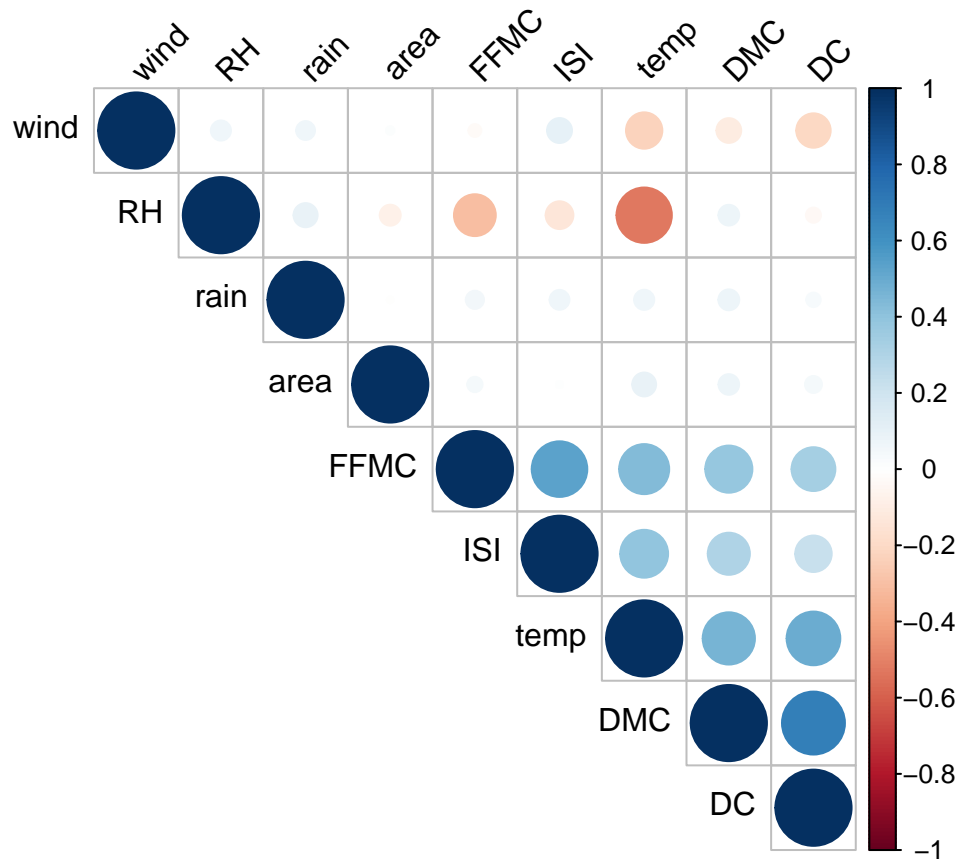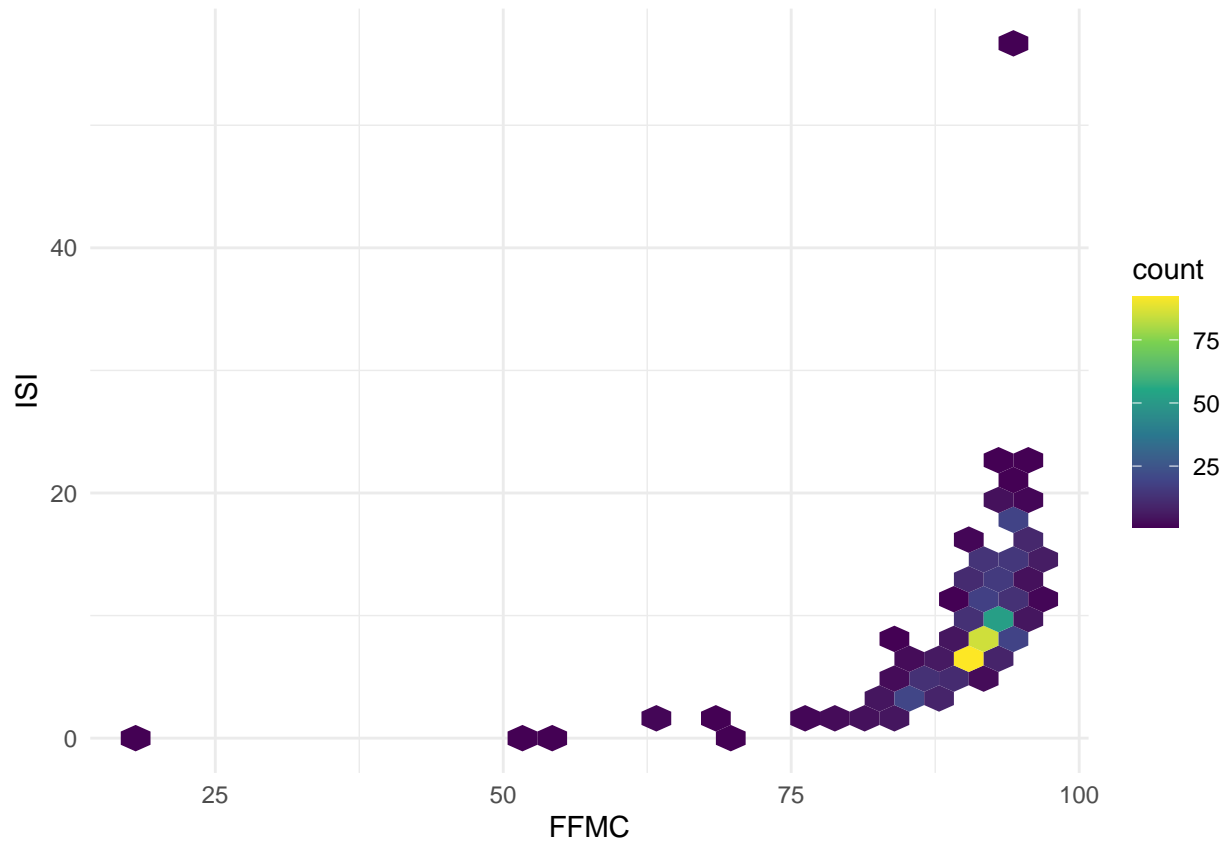


```r
ggplot(data, aes(x = FFMC, y = ISI)) +
  geom_point(alpha = 0.5) +    # scatter plot points with transparency
  geom_density_2d() +          # add 2D density contours
  theme_minimal()
```

```
ggplot(data, aes(x = FFMC, y = ISI)) +
  geom_hex() +
  scale_fill_viridis_c() +
  theme_minimal()
```

```r
# Detect Outlier in Column
detect_outlier <- function(x, threshold_multiplier = 4) {
  Q1 <- quantile(x, probs = 0.25)
  Q3 <- quantile(x, probs = 0.75)
  IQR <- Q3 - Q1
  outliers <- x > (Q3 + threshold_multiplier * IQR) | x < (Q1 - threshold_multiplier * IQR)
  outliers
}

# Remove Outlier Function
remove_outlier <- function(dataframe, columns = names(dataframe)) {
  for (col in columns) {
    outliers <- detect_outlier(dataframe[[col]])
    dataframe <- dataframe[!outliers, ]
  }
  dataframe
}


cleaned_data <- remove_outlier(data, c('FFMC', 'ISI'))

dim(cleaned_data)
```
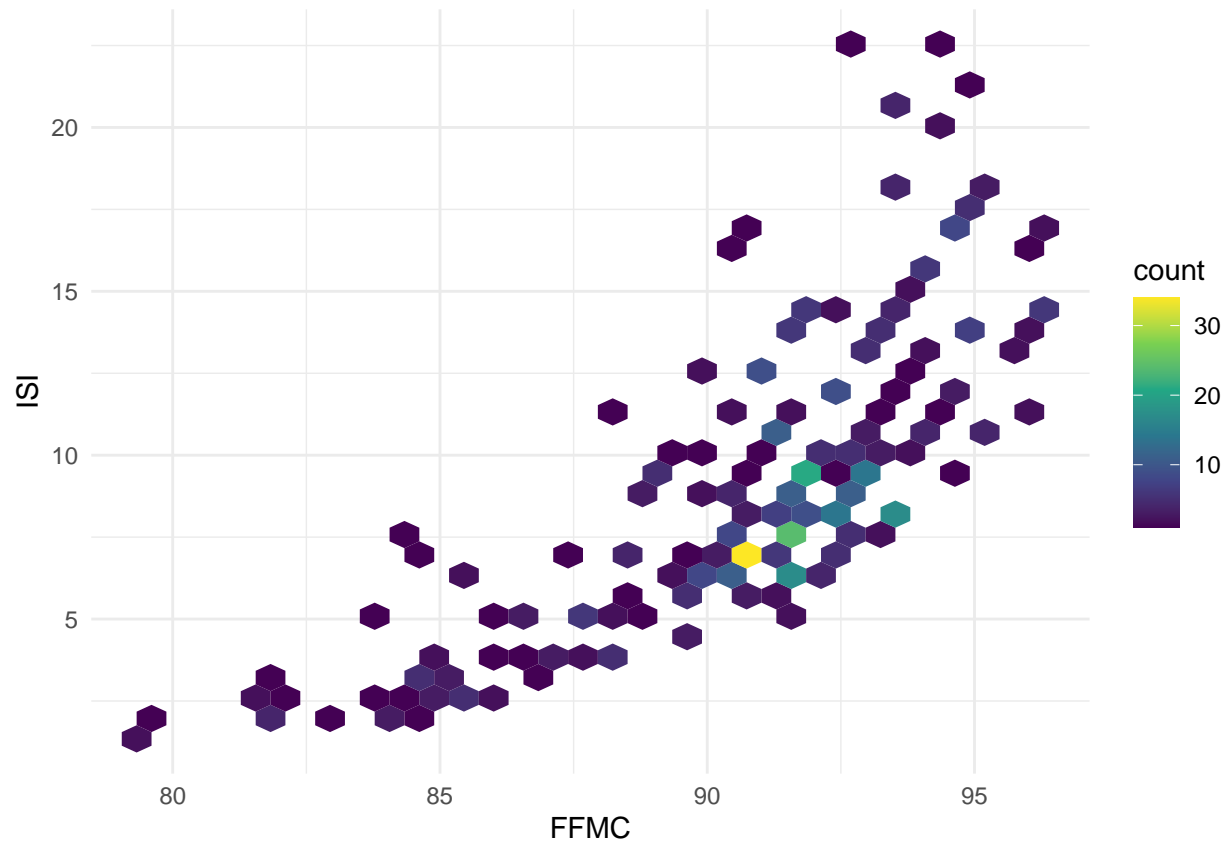
```
## [1] 507  13
```

```r
ggplot(cleaned_data, aes(x = FFMC, y = ISI)) +
  geom_hex() +
  scale_fill_viridis_c() +
  theme_minimal()
```


5 Model Specification

6 Model Assuptions

7 Model Results and Interpretation