Team 6 Project

2022-11-29

Guillermo Felce

What each variable means:

X - x-axis spatial coordinate within the Montesinho park map: 1 to 9

Y - y-axis spatial coordinate within the Montesinho park map: 2 to 9

month - month of the year: 'jan' to 'dec'

day - day of the week: 'mon' to 'sun'

FFMC - FFMC index from the FWI system: 18.7 to 96.20

DMC - DMC index from the FWI system: 1.1 to 291.3

DC - DC index from the FWI system: 7.9 to 860.6

ISI - ISI index from the FWI system: 0.0 to 56.10

temp - temperature in Celsius degrees: 2.2 to 33.30

RH - relative humidity in %: 15.0 to 100

wind - wind speed in km/h: 0.40 to 9.40

rain - outside rain in mm/m2 : 0.0 to 6.4

area - the burned area of the forest (in ha): 0.00 to 1090.84

Now, let's try running multiple visualizations and manipulating the variables to try to better understand what is going on and if we can extrapolate any hypotheses from them:

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
forestfires <- read.csv("forestfires.csv")

head(forestfires)
```

```
##   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
```

```
ggplot(forestfires, aes(x=month, y=area, fill=RH)) +
        geom_bar(stat="identity") + ggtitle("Burned Area of the Forest by Month")
```
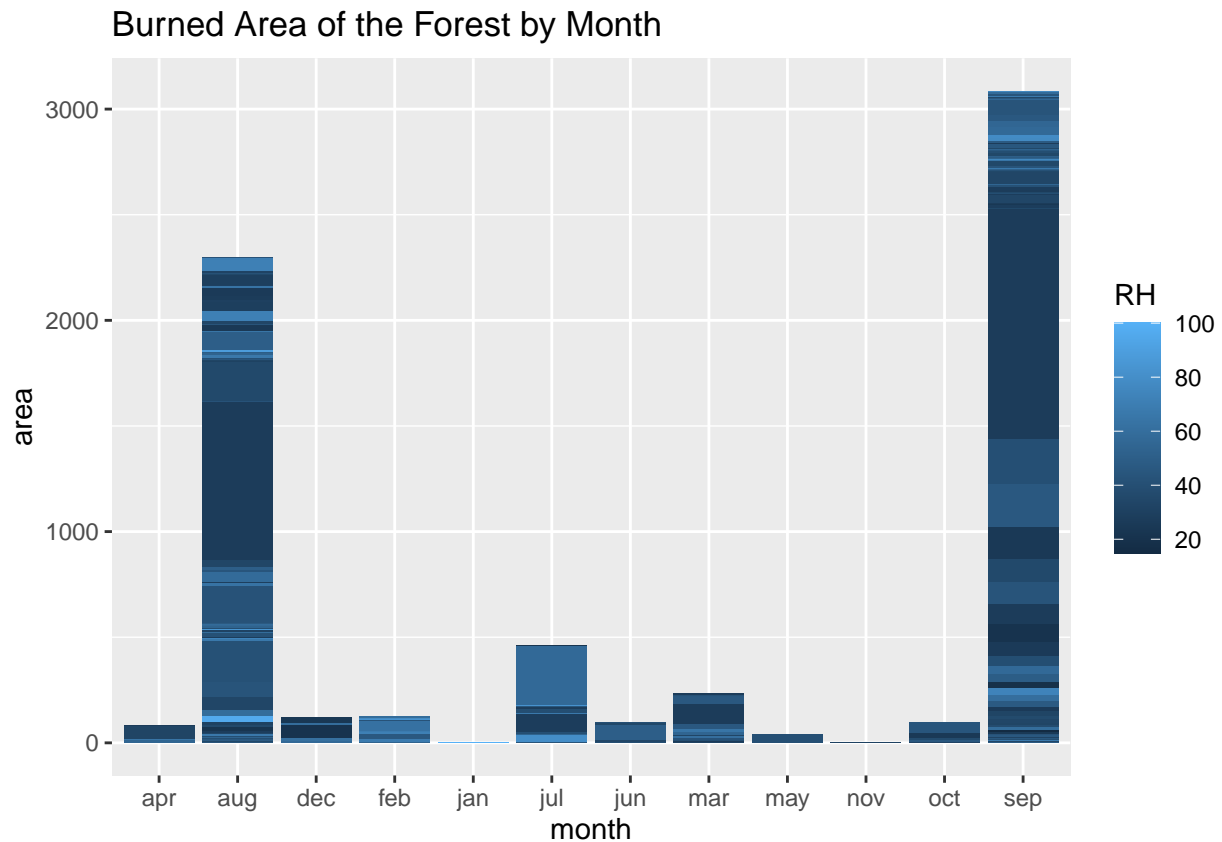


```
ggplot(forestfires, aes(x=month, y=area, fill=month)) + geom_point(color="purple") +
  ggtitle("Area vs. Month, 1 point = Individual Fire")
```

## Area vs. Month, 1 point = Individual Fire
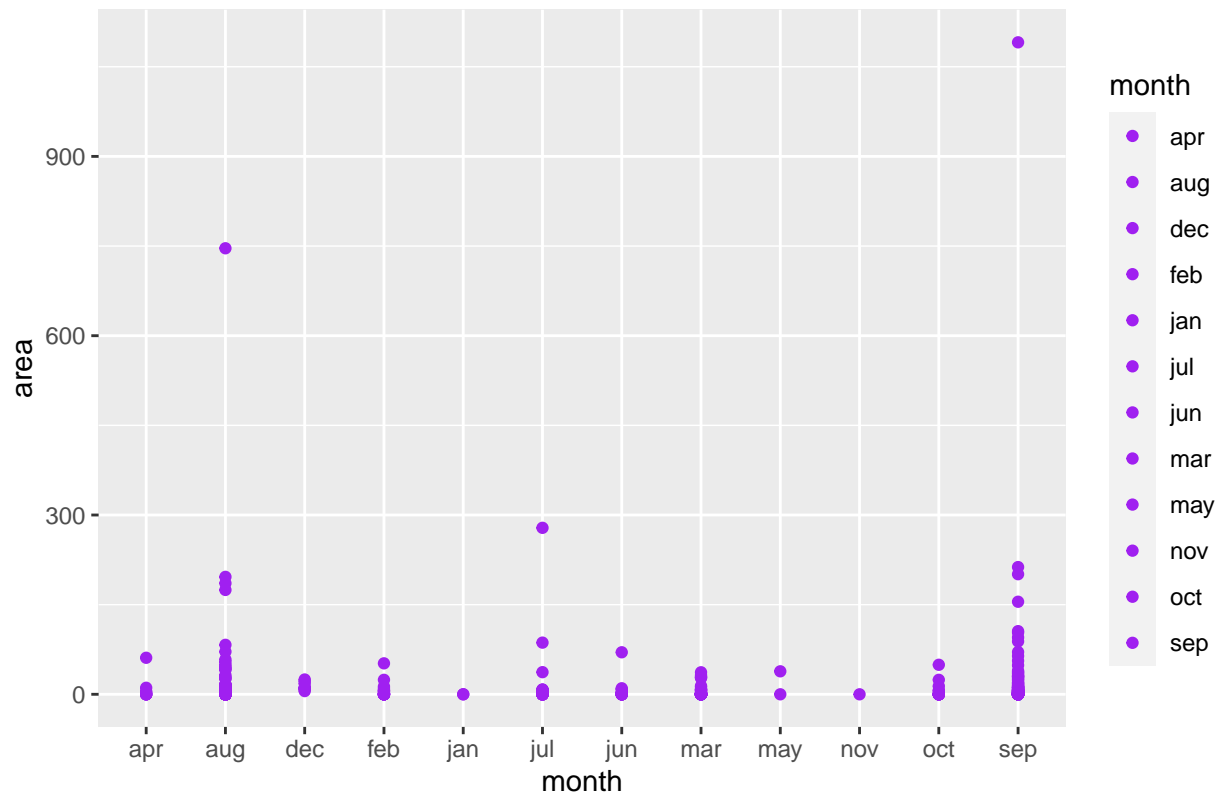


```
ggplot(forestfires, aes(x=month, y=area, color=RH, size=DC)) + geom_point(stat="identity") +
ggtitle("Forest Area Burned, Size = Drought Code and Color-coded by Humidity")
```

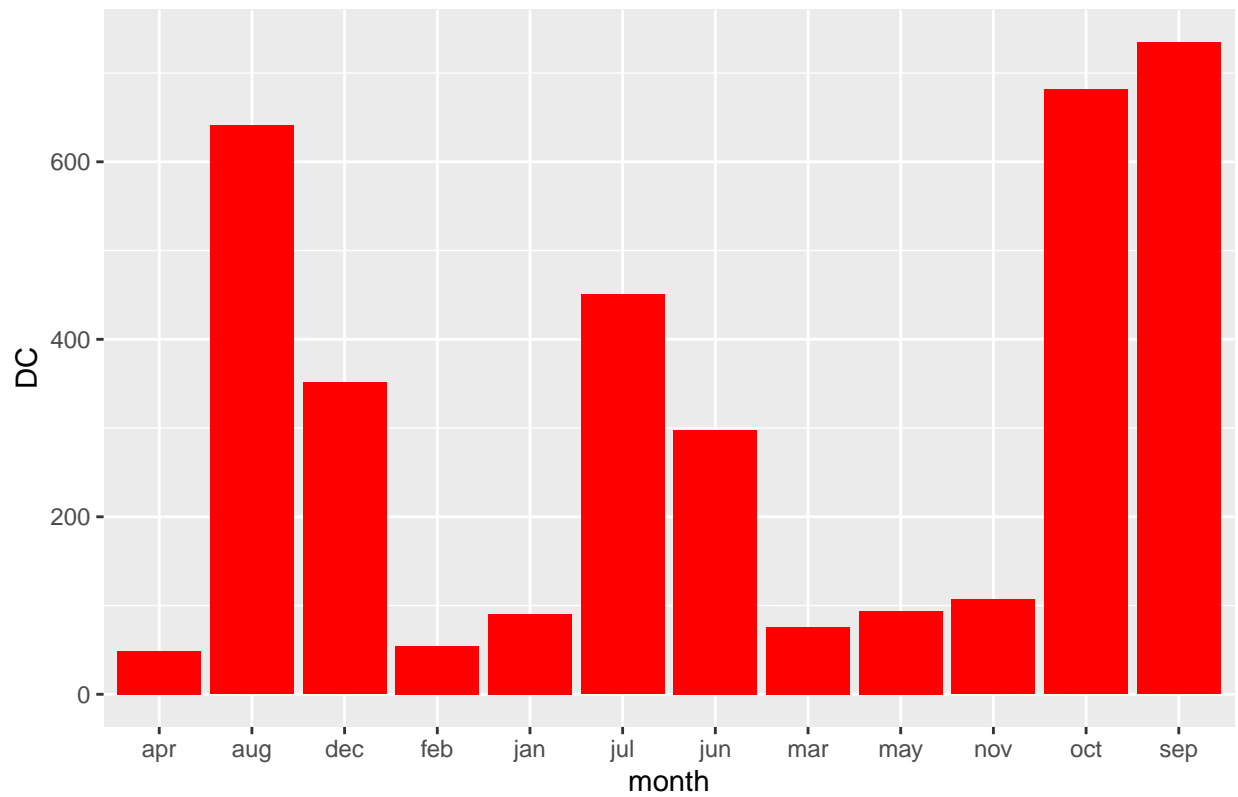Forest Area Burned, Size = Drought Code and Color−coded by Humidity
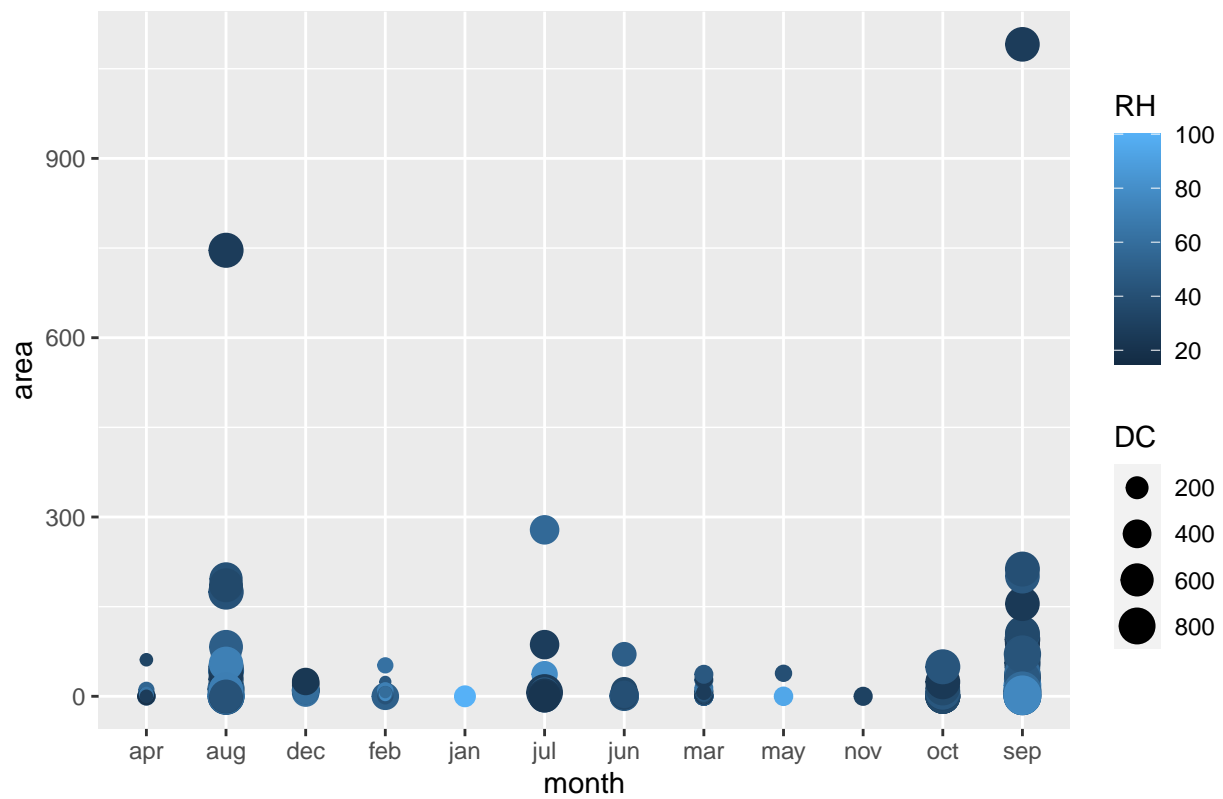
```
ggplot(forestfires, aes(x = month, y = temp, fill=month)) +
  geom_bar(stat = "summary", fun = "mean") + ggtitle("Average Temperature by Month")
```
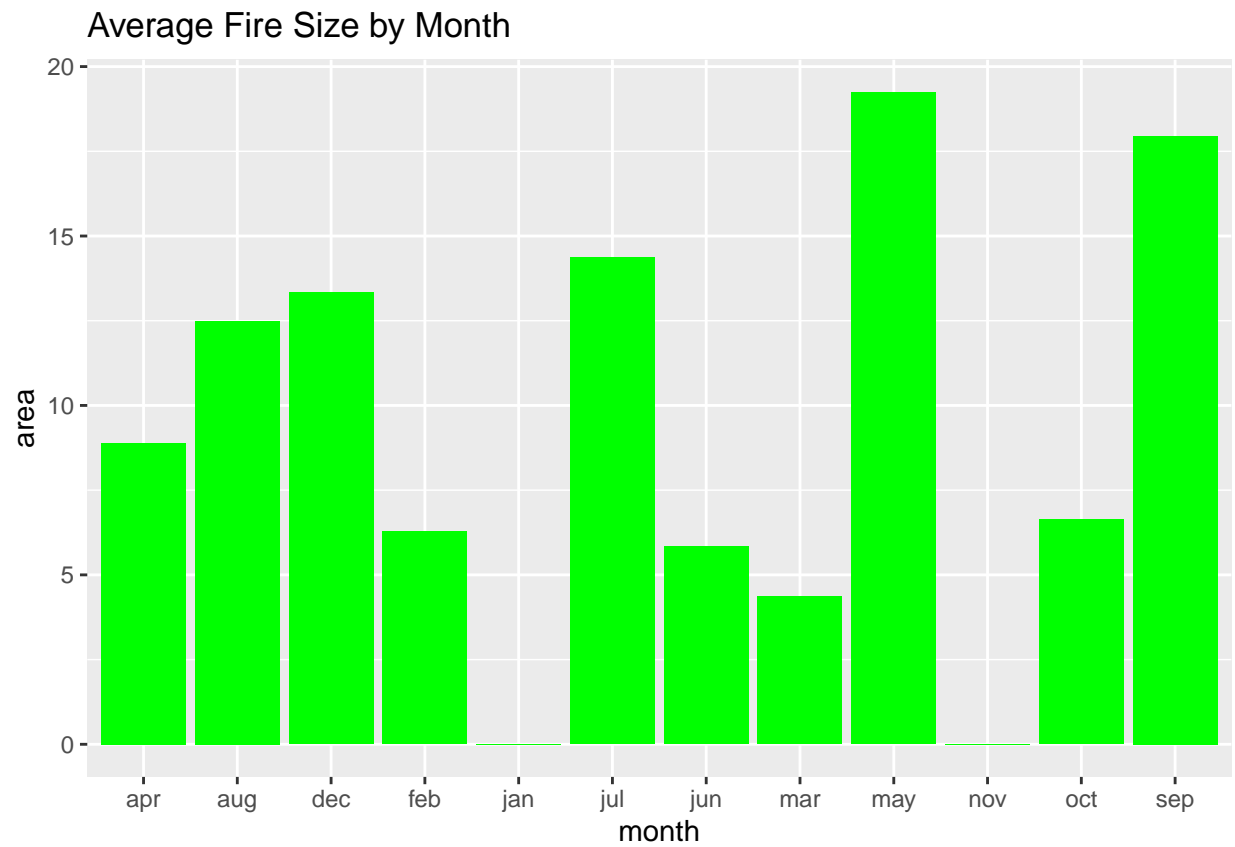
Average Temperature by Month
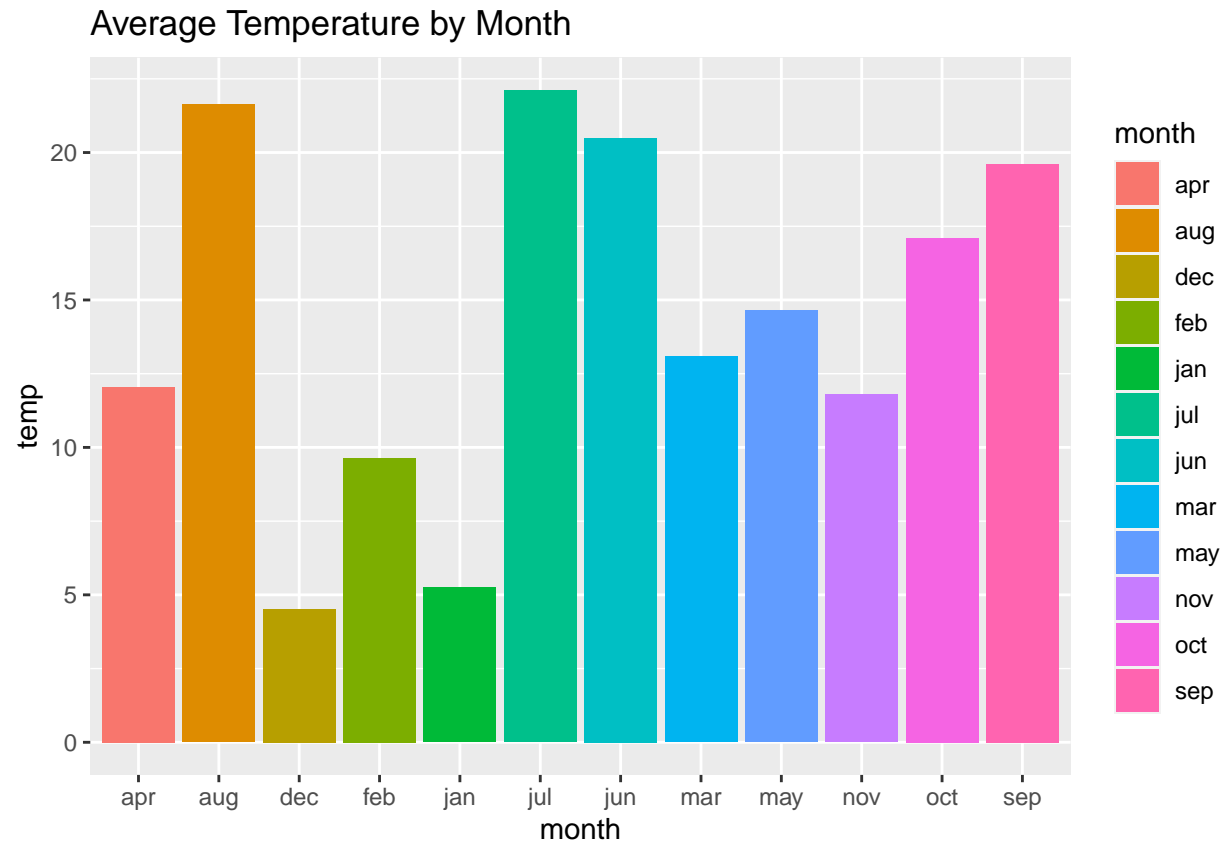

```r
ggplot(forestfires, aes(x = month, y = DC)) +
  geom_bar(stat = "summary", fun = "mean", fill="red") + ggtitle("Average Drought Code by Month")
```
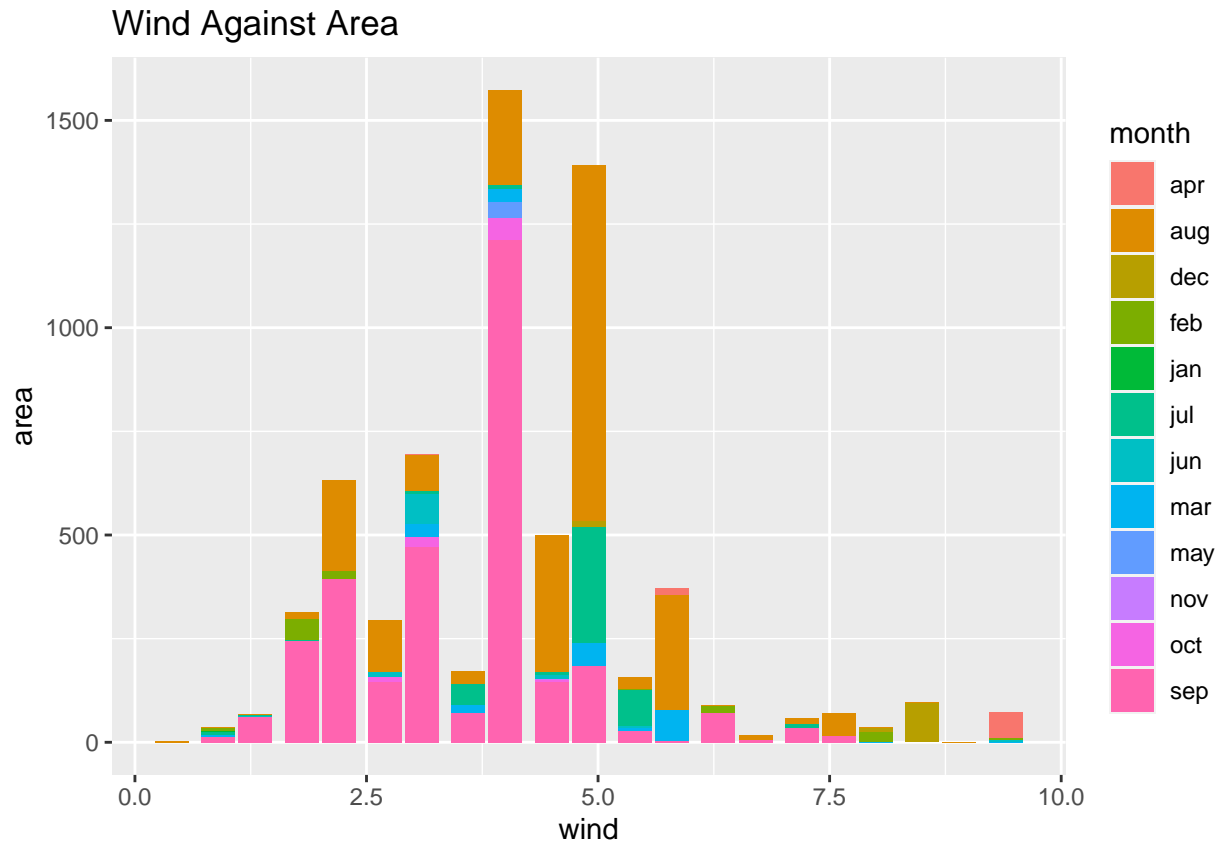
Average Drought Code by Month


```
ggplot(forestfires, aes(x = month, y = area)) +
  geom_bar(stat = "summary", fun = "mean", fill="green") + ggtitle("Average Fire Size by Month")
```

Average Fire Size by Month


```
ggplot(forestfires, aes(x = wind, y = area, fill=month)) +
  geom_bar(stat = "identity") + ggtitle("Wind Against Area")
```

Wind Against Area

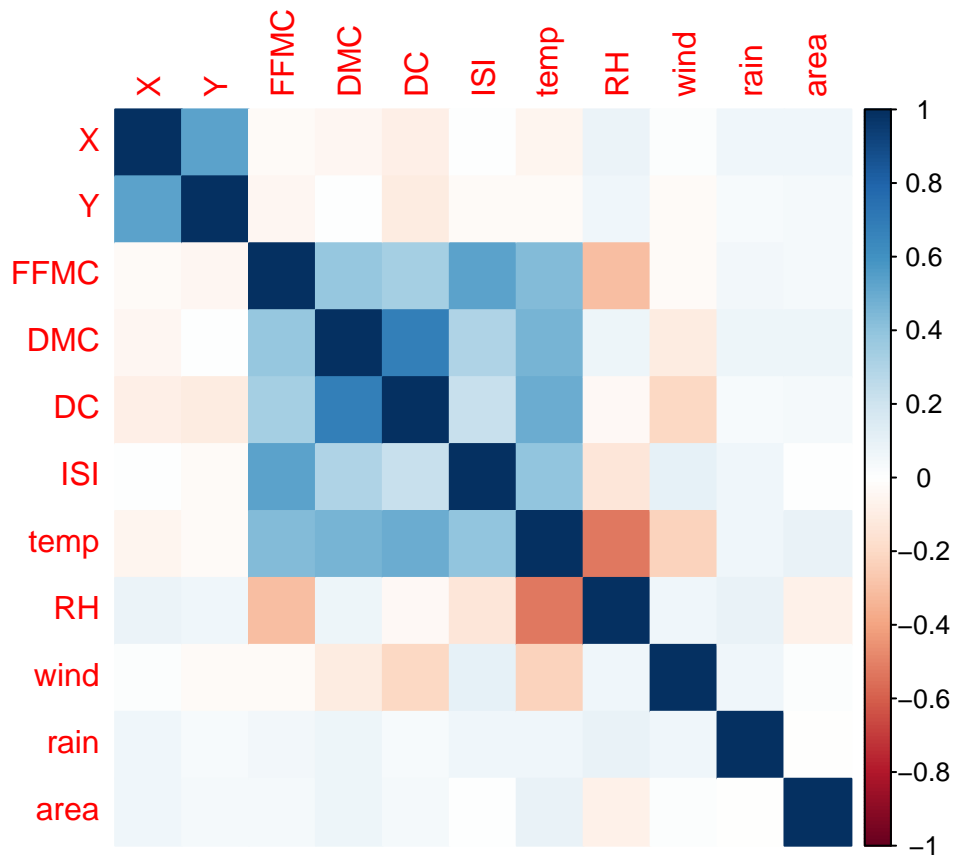Let's try finding the correlations between each of the variables, and then visualizing them in some correlation matrices:

```
library(corrplot)
```
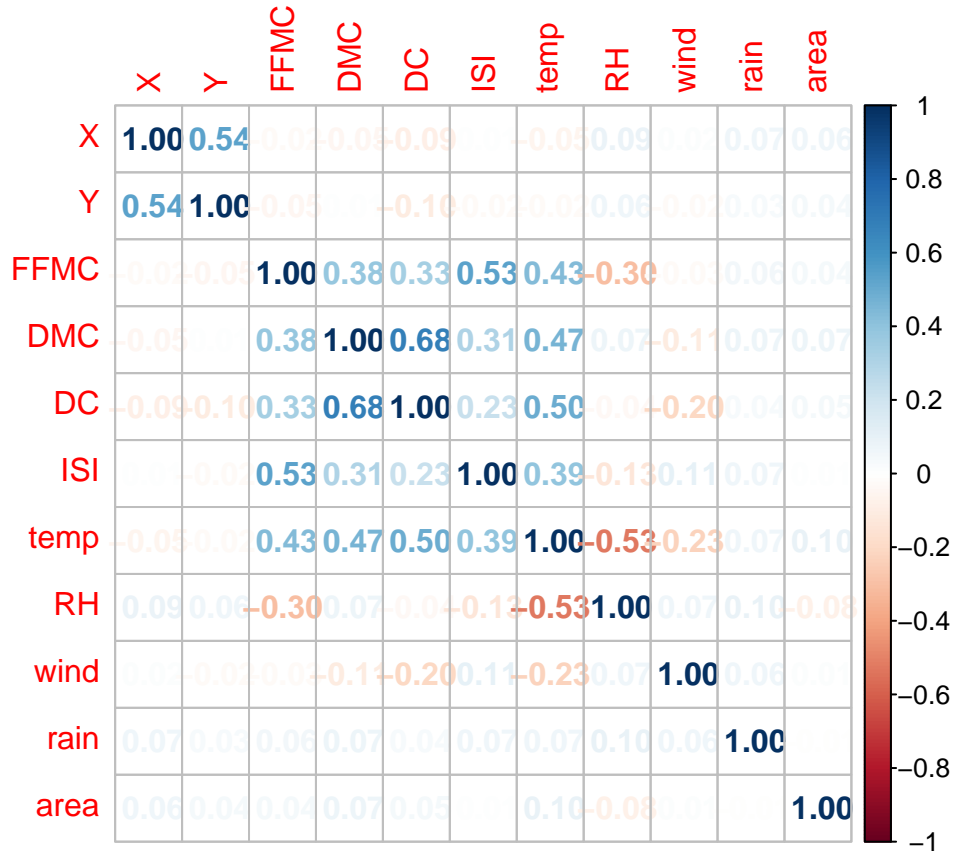
```
## corrplot 0.92 loaded
```

```
forest2 = subset(forestfires, select = -c(month,day) )

mydata.cor = cor(forest2)

corrplot(mydata.cor, method="color")
```

```
corrplot(mydata.cor, method="number")
```

It seems that DMC and DC have the highest correlation, at 0.68.