Analyzing Forest Fire Data

RZ

6/27/2021

Import the library and the setting working directory

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.6
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
setwd("~/Downloads/Dataquest/R")
fire <- read.csv('forestfires.csv')
print(colnames(fire))
```

```
##  [1] "X"     "Y"     "month" "day"   "FFMC"  "DMC"   "DC"    "ISI"   "temp"
## [10] "RH"    "wind"  "rain"  "area"
```

A single record indicates one fire record

Check the unique values of day and month

```
print(fire %>% pull(month) %>% unique())
```

```
##  [1] "mar" "oct" "aug" "sep" "apr" "jun" "jul" "feb" "jan" "dec" "may" "nov"
```

```
fire %>% pull(day) %>% unique()
```

```
## [1] "fri" "tue" "sat" "sun" "mon" "wed" "thu"
```

Convert the month and day to categorical variable

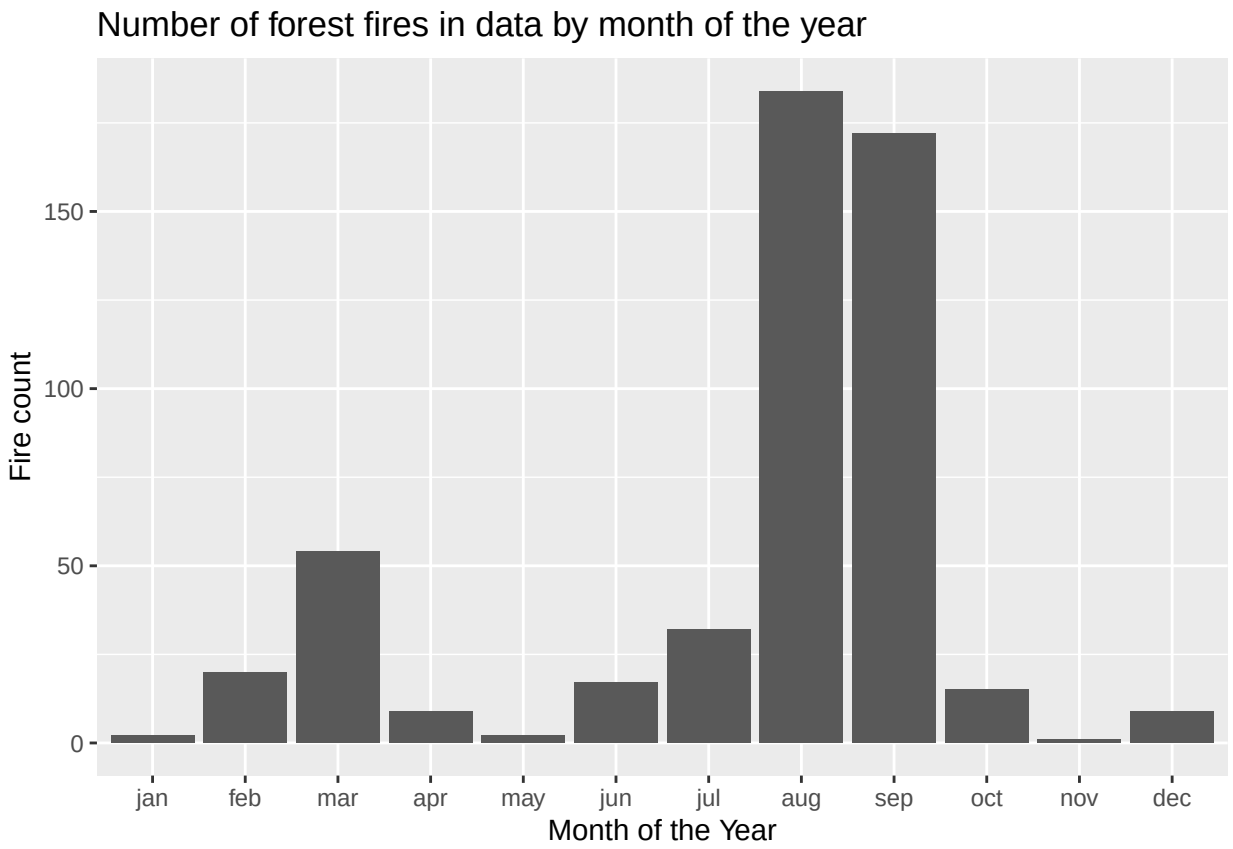
```
month_order <- c("jan", "feb", "mar",
                 "apr", "may", "jun",
                 "jul", "aug", "sep",
                 "oct", "nov", "dec")
dow_order <- c("sun", "mon", "tue", "wed", "thu", "fri", "sat")
fire <- fire %>% mutate(
  month = factor(month, levels = month_order),
  day = factor(day, levels = dow_order)
)
```

Check the most happen in which month and day

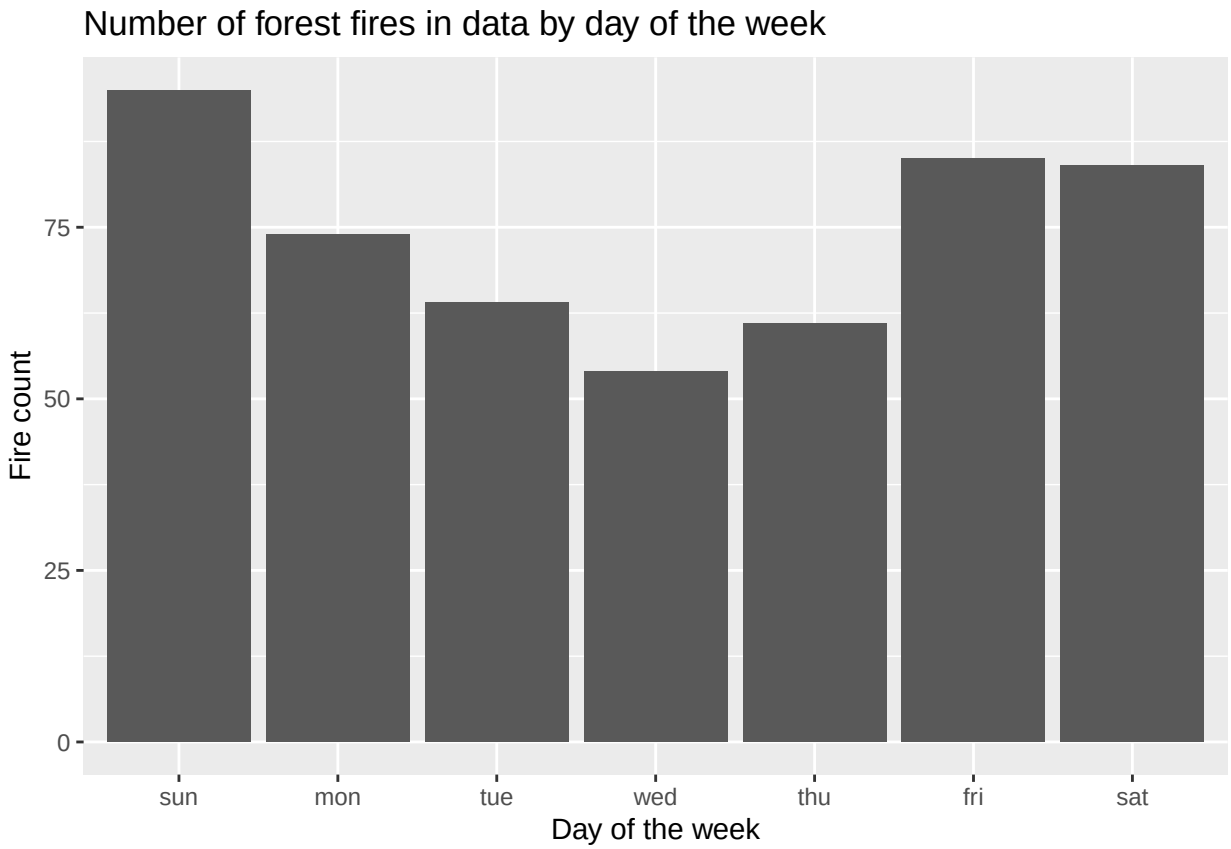
```
fire_month <- fire %>% group_by(month) %>% summarise(
  count_m = n()
)

fire_day <- fire %>% group_by(day) %>% summarise(
  count_d = n()
)
```

Visualize the counts by month and day

```
fire_month %>% ggplot(aes(x=month, y=count_m)) + geom_col() + labs(
  title = "Number of forest fires in data by month of the year",
  y = "Fire count",
  x = "Month of the Year"
  )
```

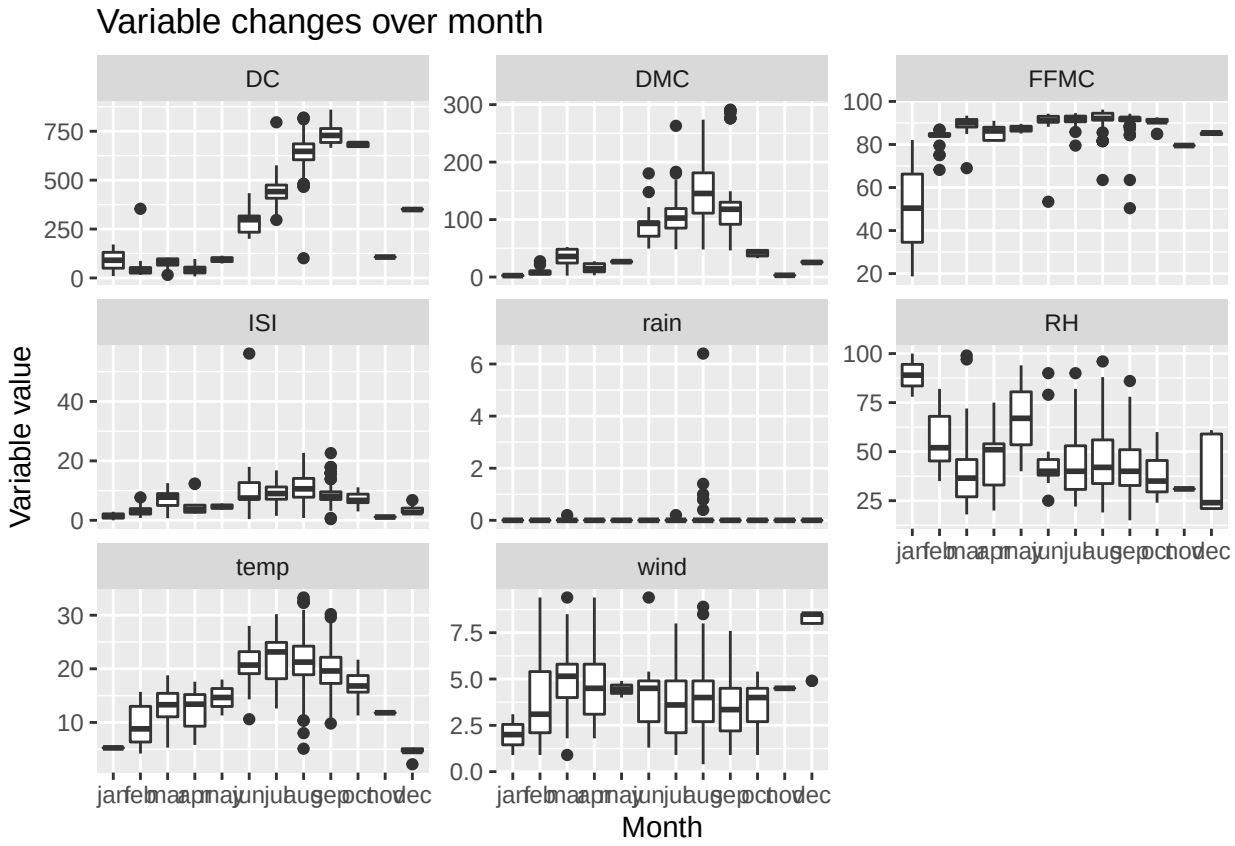



```
fire_day %>% ggplot(aes(x=day, y=count_d)) + geom_col() + labs(
  title = "Number of forest fires in data by day of the week",
  y = "Fire count",
  x = "Day of the week"
  )
```



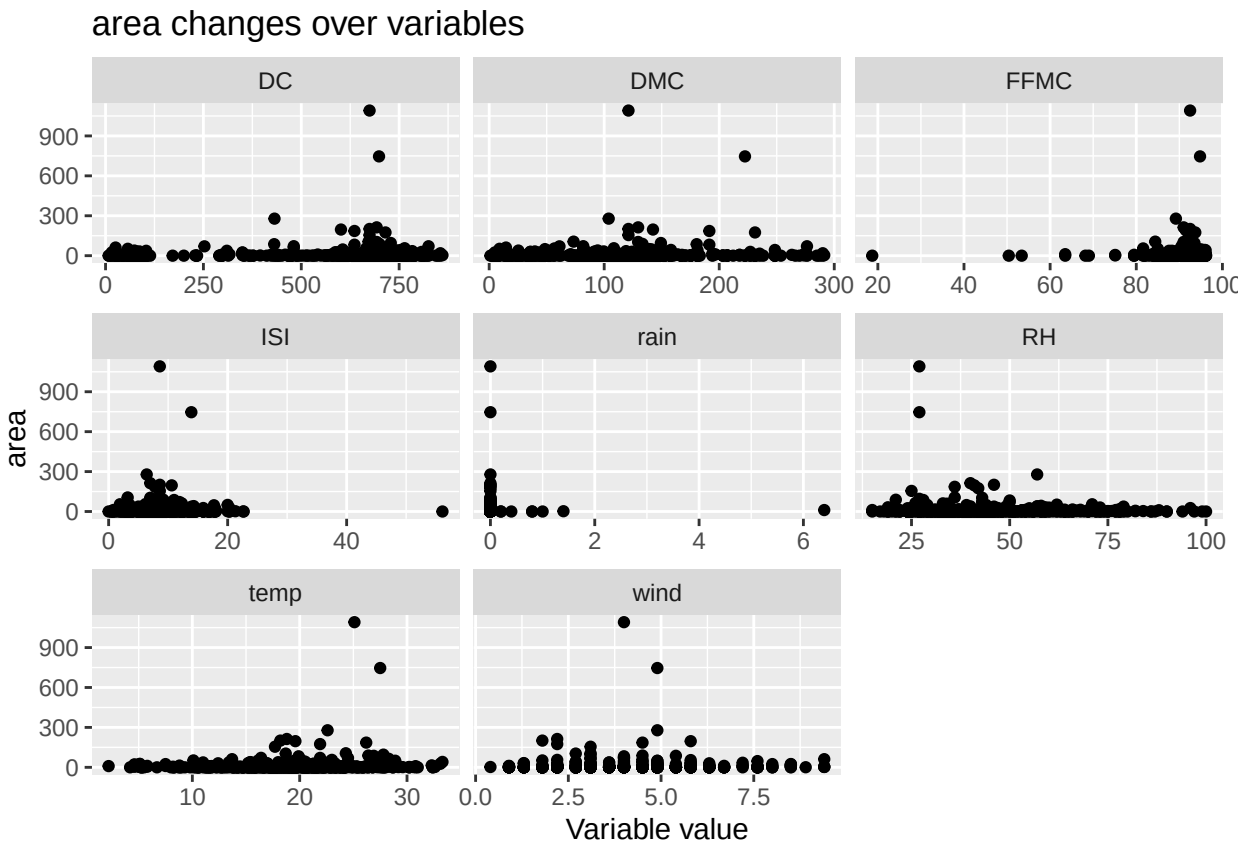
Check each variable relationship to month

```
fire_long <- fire %>%
  pivot_longer(
    cols = c("FFMC", "DMC", "DC",
             "ISI", "temp", "RH",
             "wind", "rain"),
    names_to = "data_col",
    values_to = "value"
  )

fire_long %>% ggplot(aes(x=month, y=value)) +geom_boxplot() + facet_wrap(vars(data_col), scale = 'free_
  title = "Variable changes over month",
  x = "Month",
  y = "Variable value")
```



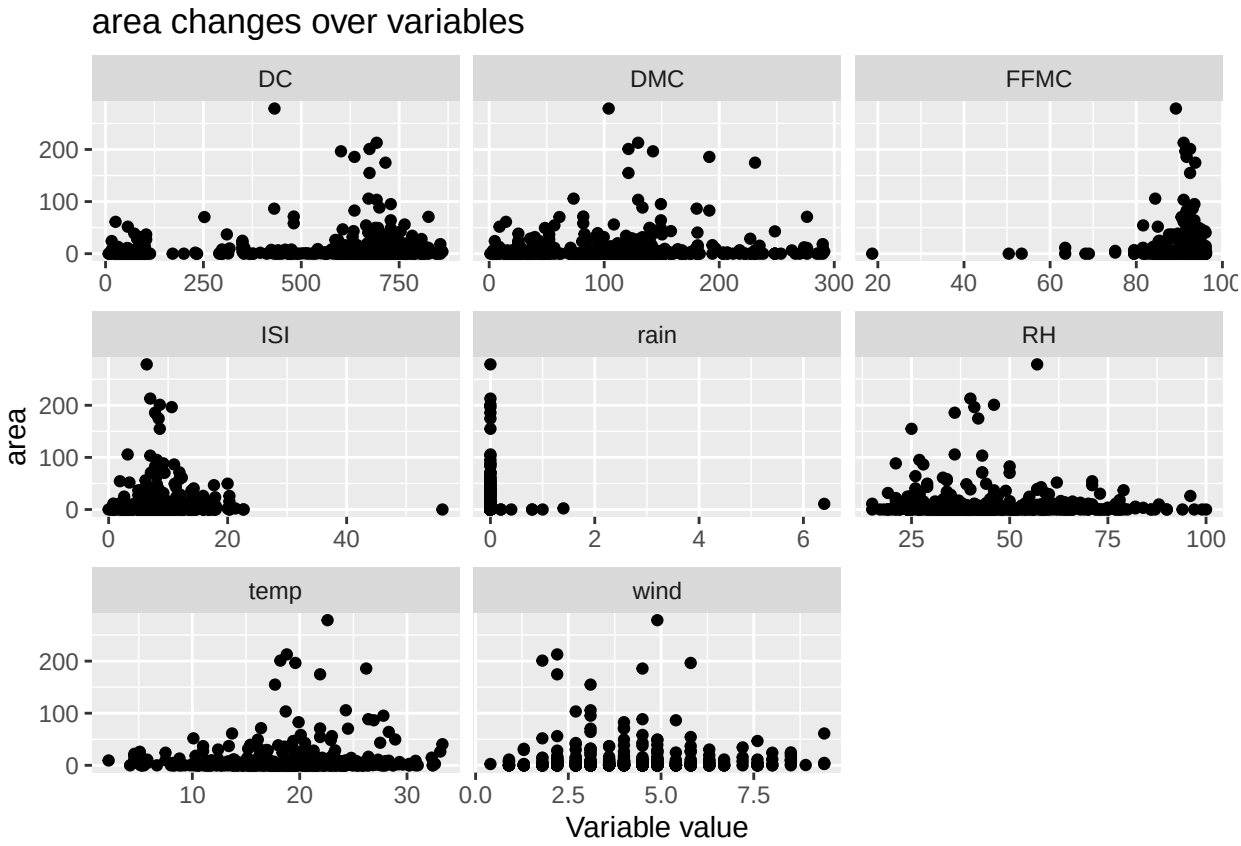
Check the relationship of the areas and each variable

```
fire_long %>% ggplot(aes(x=value, y=area)) +geom_point() + facet_wrap(vars(data_col), scale = 'free_x')
  title = "area changes over variables",
  x = "Variable value",
  y = "area")
```



remove the data with area smaller than 300 to remove outliers

```
fire_long %>% filter(area <300) %>% ggplot(aes(x=value, y=area)) +geom_point() + facet_wrap(vars(data_c
  title = "area changes over variables",
  x = "Variable value",
  y = "area")
```


area changes over variables
DC
DMC
FFMC
ISI
rain
RH
temp
wind
area
Variable value
0
100
200
300
0
250
500
750
20
40
60
80
0
2
4
6
25
50
75
100
10
20
30
0.0
2.5
5.0
7.5