

# Homework #1

Group 4

5/10/2020

## First, let's import the required libraries

```
library(ggplot2)
library(ggpubr)
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(RColorBrewer)
```

## Problem 1

- (a) Plot area vs.temp, area vs. month, area vs. DC, area vs. RH for January through December combined in one graph. Hint: Place area on Y axis and use 2x2 matrix to place the plots adjacent to each other.

```
# Import the Dataframe
forestfires <- data.frame(read.csv("./data/forestfires.csv"),
  stringsAsFactors = FALSE
)

# Convert the month column into factors and sort from Jan-Dec
forestfires$month <- factor(forestfires$month,
  levels = c(
    "jan", "feb", "mar",
    "apr", "may", "jun",
    "jul", "aug", "sep",
    "oct", "nov", "dec"
  )
)

# Create 4 scatter plots
p1 <- ggplot(forestfires, aes(temp, area)) +
  geom_point(color = "#d63447", alpha = 0.5) +
  ggtitle("Temp vs Area") +
  theme_classic()
```

```

p2 <- ggplot(forestfires, aes(month, area, color = month)) +
  geom_point() +
  scale_color_brewer(palette = "Set3") +
  theme_classic() +
  theme(legend.position = "none") +
  ggtitle("Month vs Area")
p3 <- ggplot(forestfires, aes(DC, area)) +
  geom_point(color = "#d63447", alpha = 0.5) +
  ggtitle("DC vs Area") +
  theme_classic()
p4 <- ggplot(forestfires, aes(RH, area)) +
  geom_point(color = "#d63447", alpha = 0.5) +
  ggtitle("RH vs Area") +
  theme_classic()

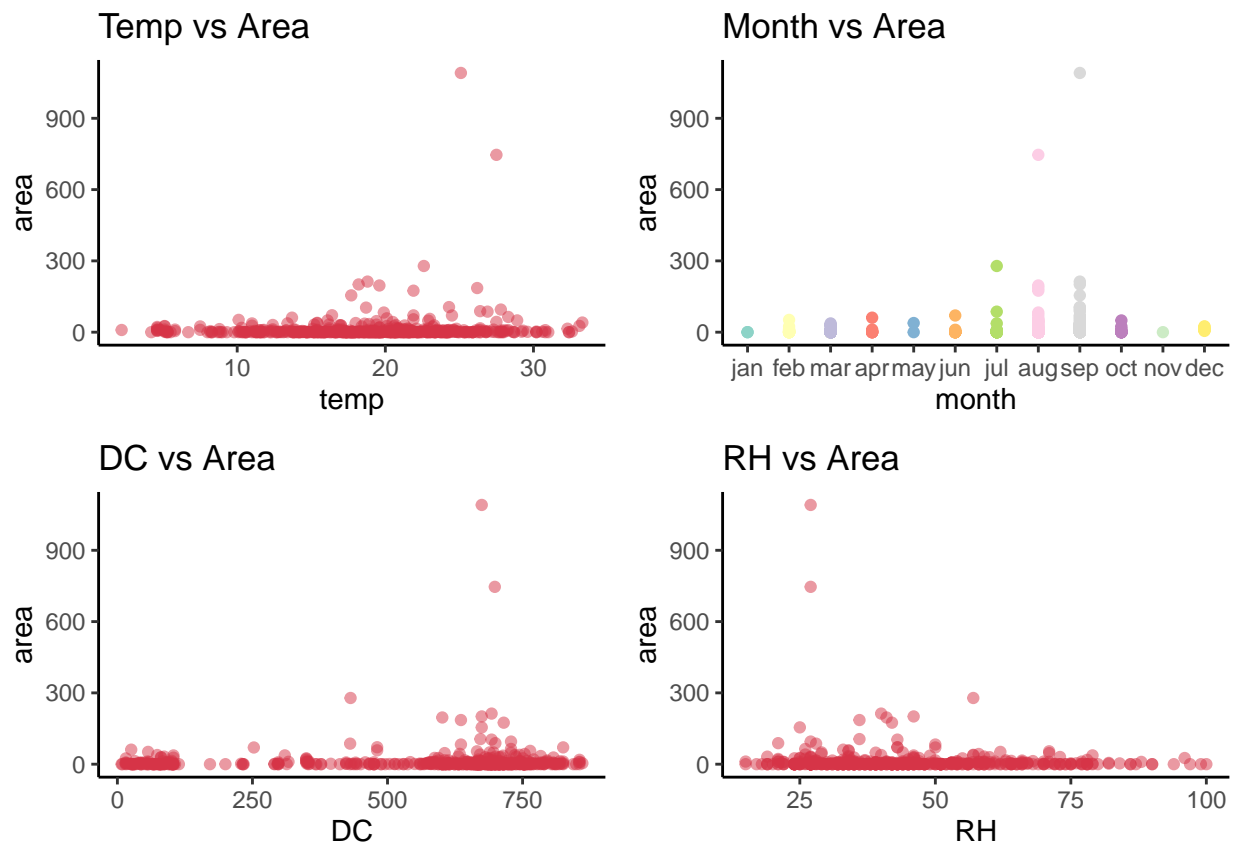
```

```

# Arrange plots P1-P4 into a 2x2 grid
fig <- ggarrange(p1, p2, p3, p4, ncol = 2, nrow = 2)

```

```
plot(fig)
```

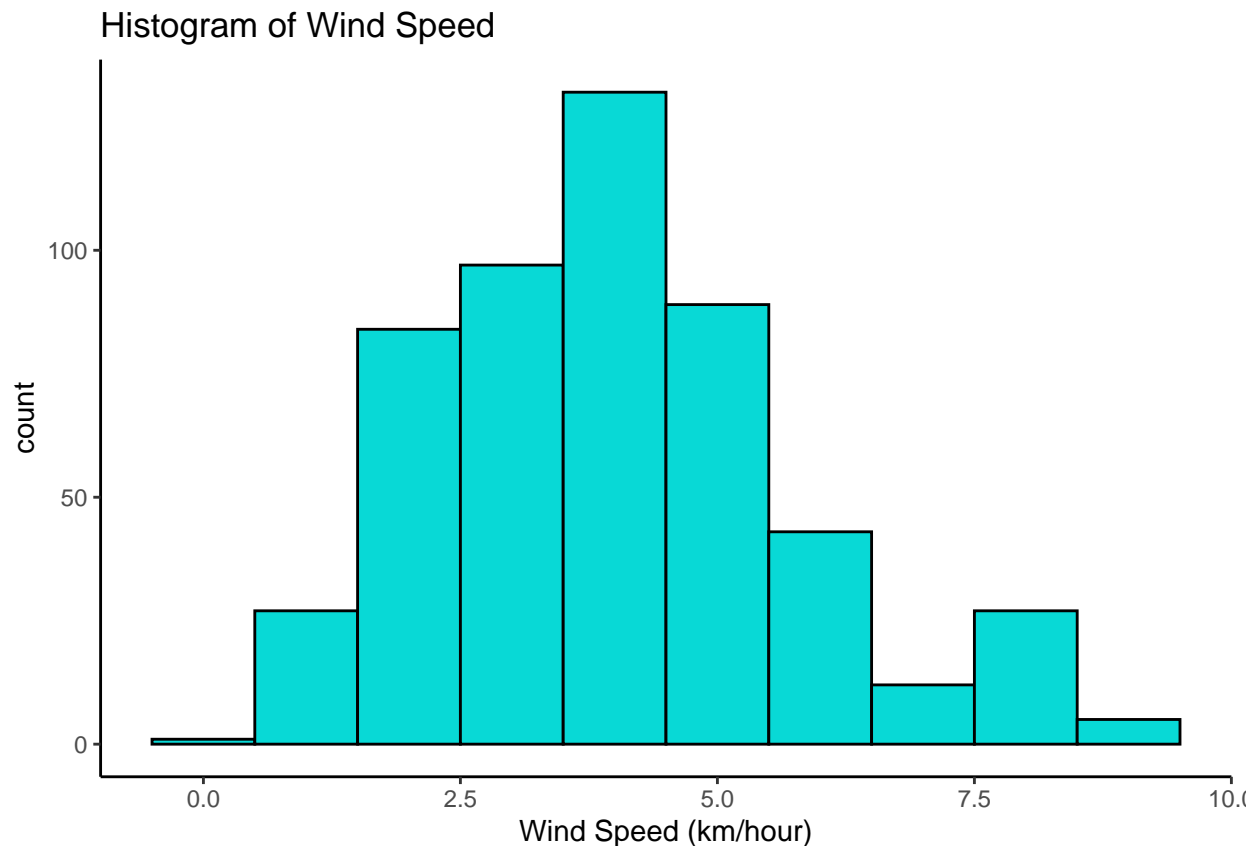


```
rm(list = c("p1", "p2", "p3", "p4", "fig"))
```

(b) Plot the histogram of wind speed (km/h).

```
# Create the Wind-Histogram
wind_hist <- ggplot(forestfires, aes(wind)) +
  geom_histogram(bins = 10, fill = "#08d9d6", color = "black") +
  theme_classic() +
  ggtitle("Histogram of Wind Speed") +
  labs(x = "Wind Speed (km/hour)")

# Plot
plot(wind_hist)
```



```
rm(wind_hist)
```

(c) Compute the summery statistics (min, 1Q, mean, median, 3Q, max,) of part b.

```
# Calculate the Quantiles
quantiles <- quantile(forestfires$wind)

# Print
cat("Minimum Wind Speed is :", quantiles[[1]], "\n")

## Minimum Wind Speed is : 0.4

cat("1st Quantile of Wind Speed is :", quantiles[[2]], "\n")

## 1st Quantile of Wind Speed is : 2.7

cat("Mean Wind Speed is :", mean(forestfires$wind), "\n")

## Mean Wind Speed is : 4.017602
```

```
cat("Median Wind Speed is :", quantiles[[3]], "\n")

## Median Wind Speed is : 4
cat("3rd Quartile of Wind Speed is :", quantiles[[4]], "\n")

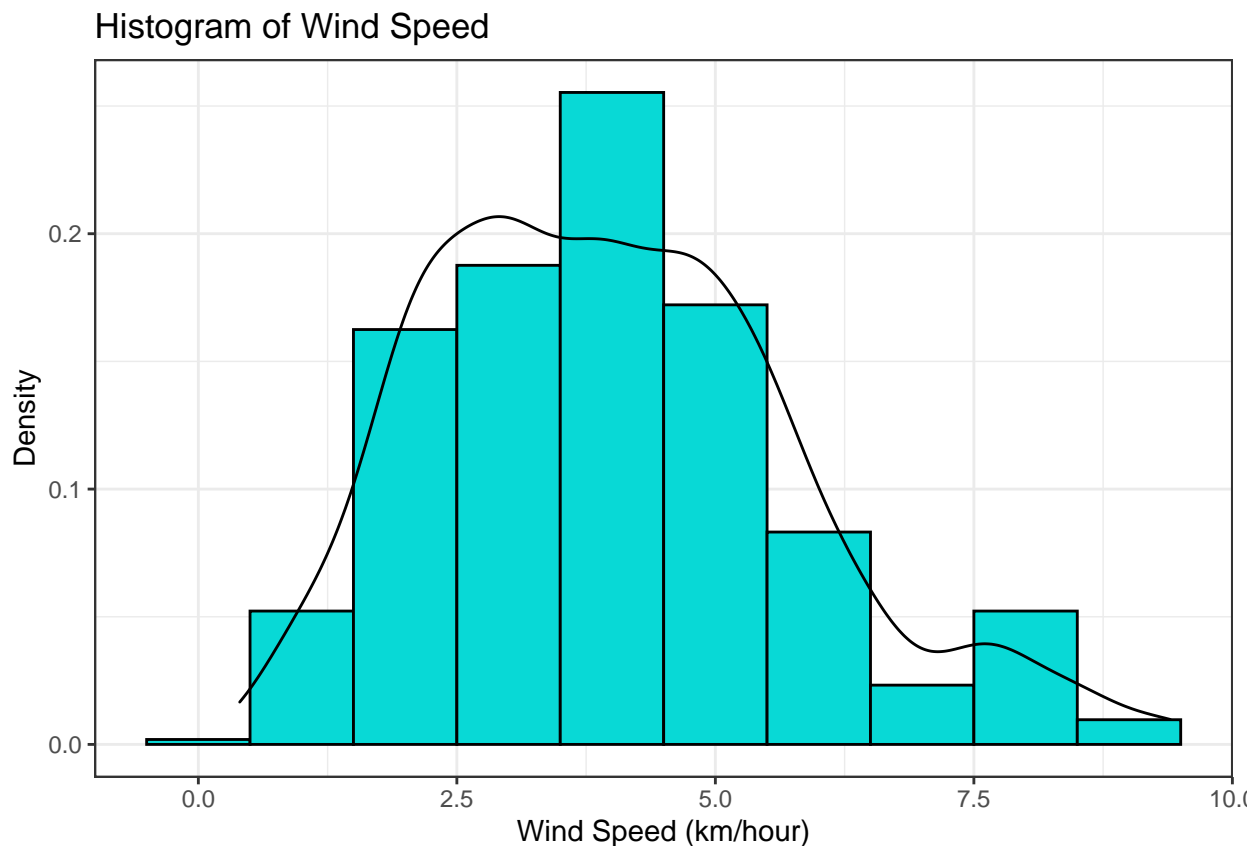
## 3rd Quartile of Wind Speed is : 4.9
cat("Maximum Wind Speed is :", quantiles[[5]], "\n")

## Maximum Wind Speed is : 9.4
rm(quantiles)
```

(d) Add a density line to the histogram in part b.

```
# Create the Histogram and Density Plot
wind_hist_plus_density <- ggplot(forestfires, aes(x = wind, y = ..density..)) +
  geom_histogram(bins = 10, colour = "black", fill = "#08d9d6") +
  geom_density(aes(y = ..density..), color = "black") +
  ylab("Density") +
  xlab("Wind Speed (km/hour)") +
  ggtitle("Histogram of Wind Speed") +
  theme_bw()

plot(wind_hist_plus_density)
```

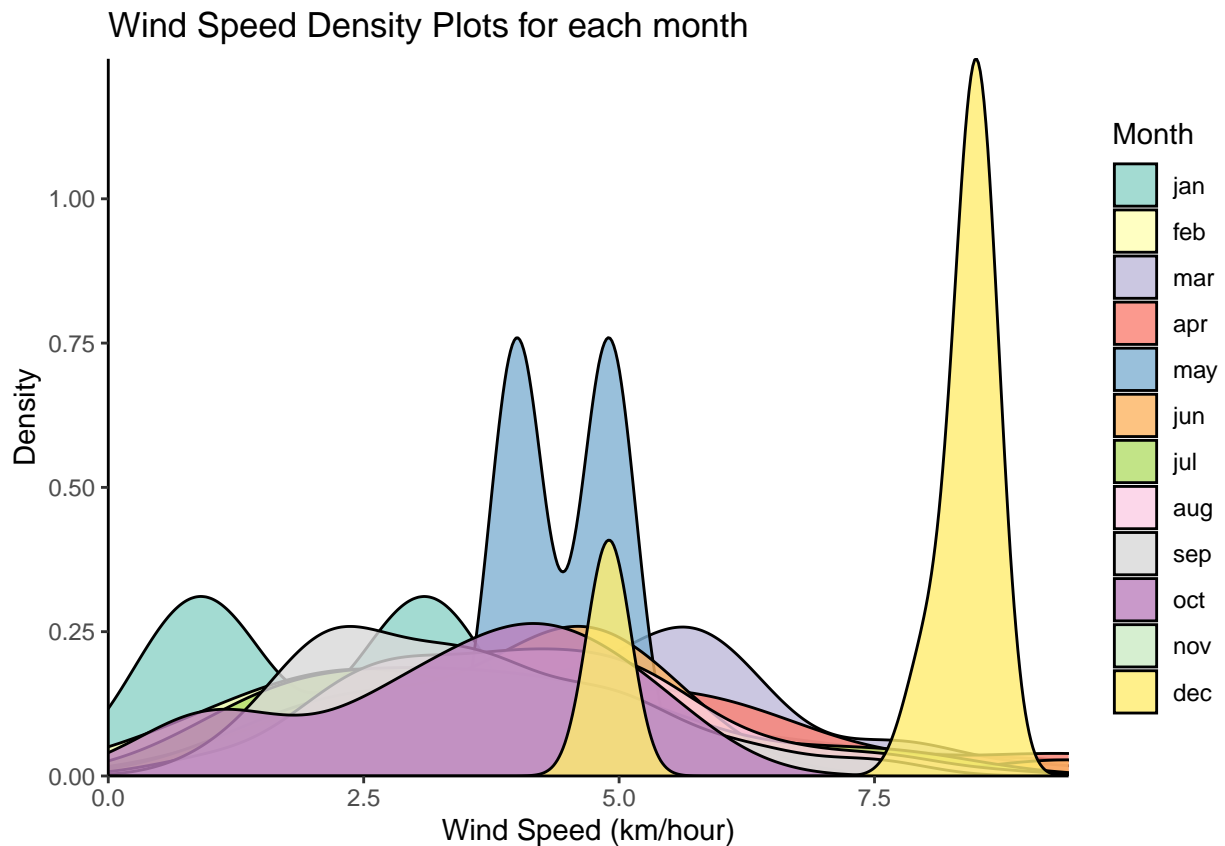


```
rm(wind_hist_plus_density)
```

- (e) Plot the wind speed density function of all months in one plot. Use different colors for different months in the graph to interpret your result clearly. [Hint: use `ggplot` + `geom_density` or `qplot(geom=density)`]

```
# Create month wise density plot
month_density_plot <- ggplot(forestfires, aes(
  x = wind,
  y = ..density..,
  fill = month
)) +
  geom_density(alpha = 0.8) +
  scale_fill_brewer(palette = "Set3") +
  theme_classic() +
  ggtitle("Wind Speed Density Plots for each month") +
  labs(x = "Wind Speed (km/hour)", y = "Density", fill = "Month") +
  scale_x_continuous(expand = c(0, 0), limits = c(0, NA)) +
  scale_y_continuous(expand = c(0, 0), limits = c(0, NA))

plot(month_density_plot)
```



```
rm(month_density_plot)
```