EXERCISE 4

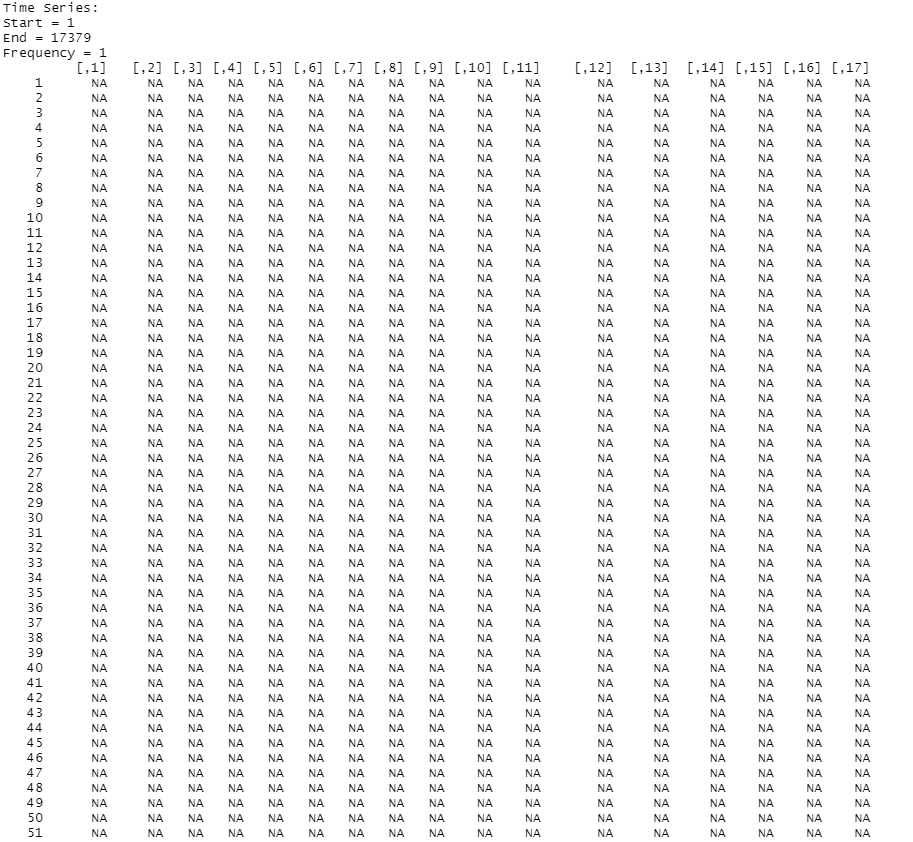
AIM: To explore various variable and row filters in R for cleaning data and applying various plot features in R on sample data sets and visualizing.

Dataset description:

Bike-sharing rental process is highly correlated to the environmental and seasonal settings. For instance, weather conditions, precipitation, day of week, season, hour of the day, etc. can affect the rental behaviors. The core data set is related to the two-year historical log corresponding to years 2011 and 2012 from Capital Bikeshare system. Bike sharing counts aggregated on hourly basis. Records: 17379 hours

# Filtering rows in R

filter(hour,hour$temp==0.24)

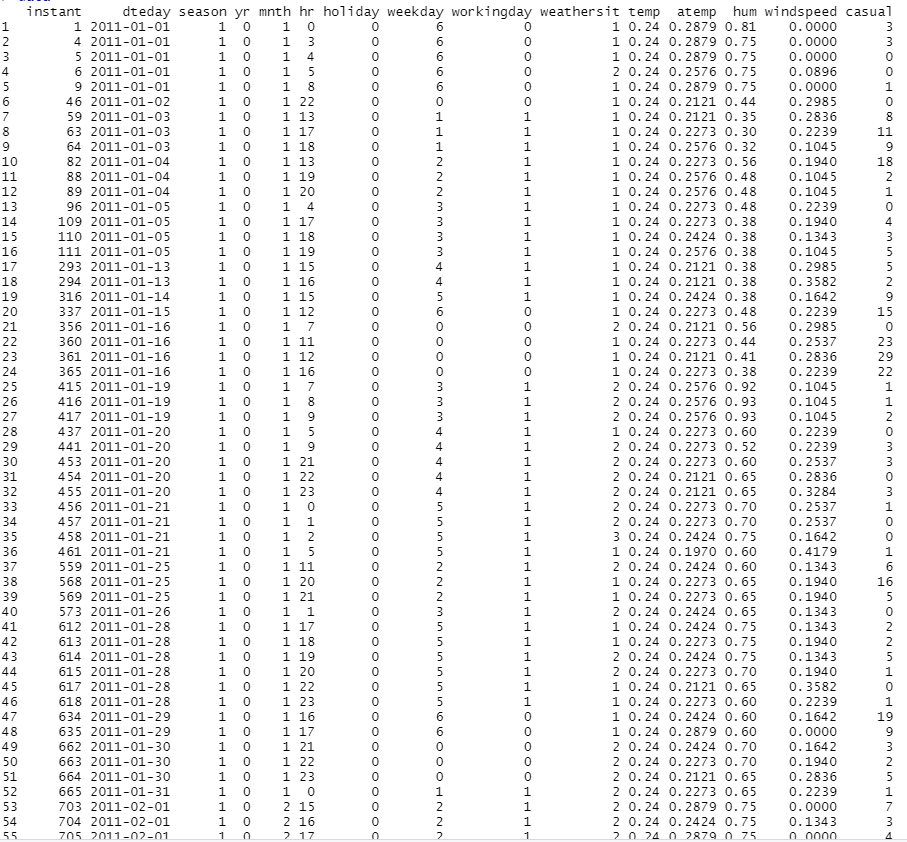


# Filtering using dplyr library

library(dplyr)

data=filter(hour,temp==0.24)

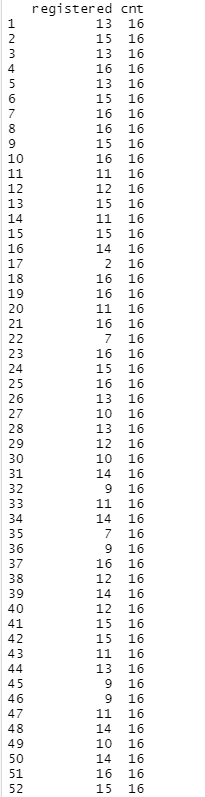
data



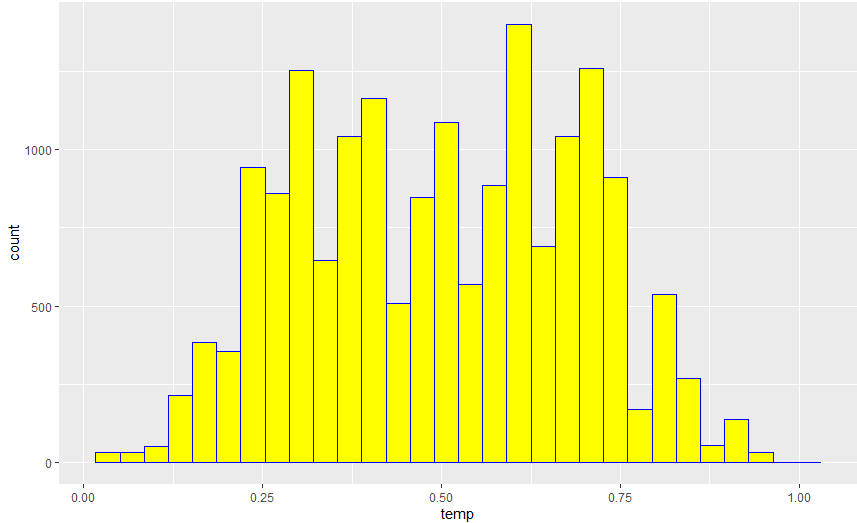
library(dplyr)

data1=filter(hour,cnt==16)

data1



ggplot(data=hour,aes(x=temp)) + geom\_histogram(fill='yellow',col='blue')



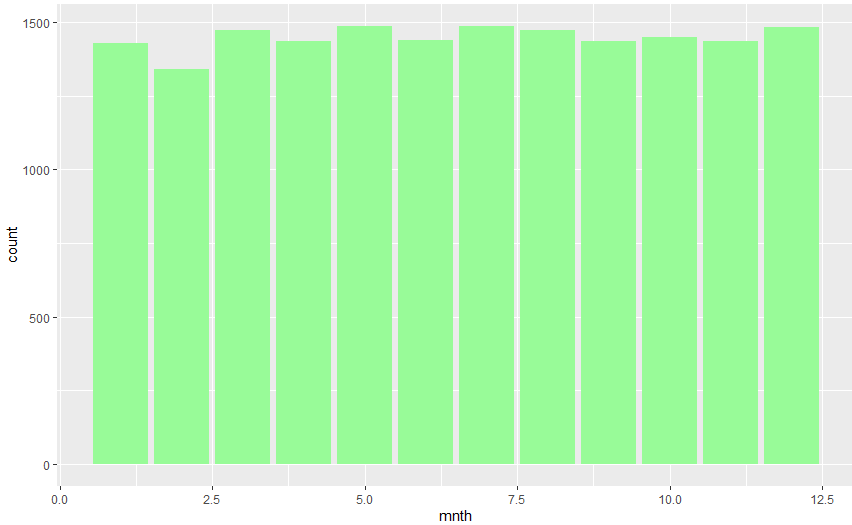
Insights:

Maximum temperature is 0.65 and it is about 1000 times.

CONCLUTION:

Highest temperature is about 0.6 and it occurs more than 1000 times

ggplot(data=hour,aes(x=mnth)) + geom\_bar(fill='palegreen')



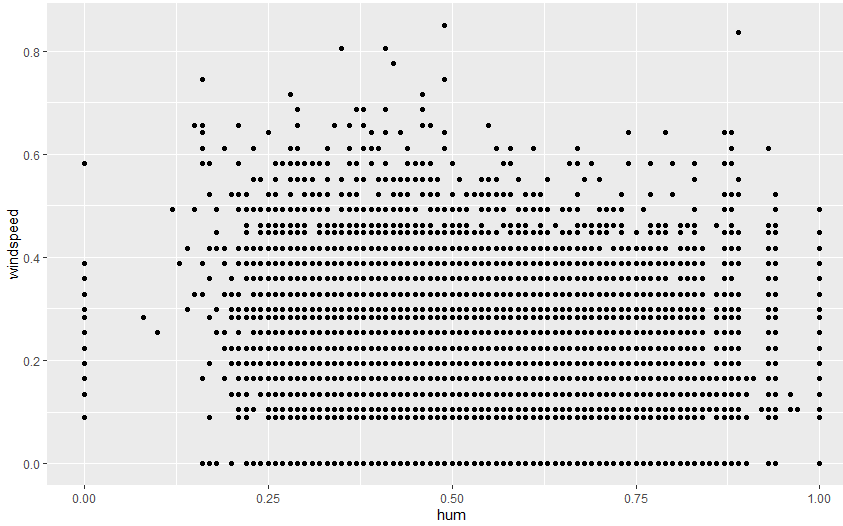
CONCLUTION:

Highest production is in the months march, may, July and December.

Insights:

Energy is produced more in the months march, may, July and December.

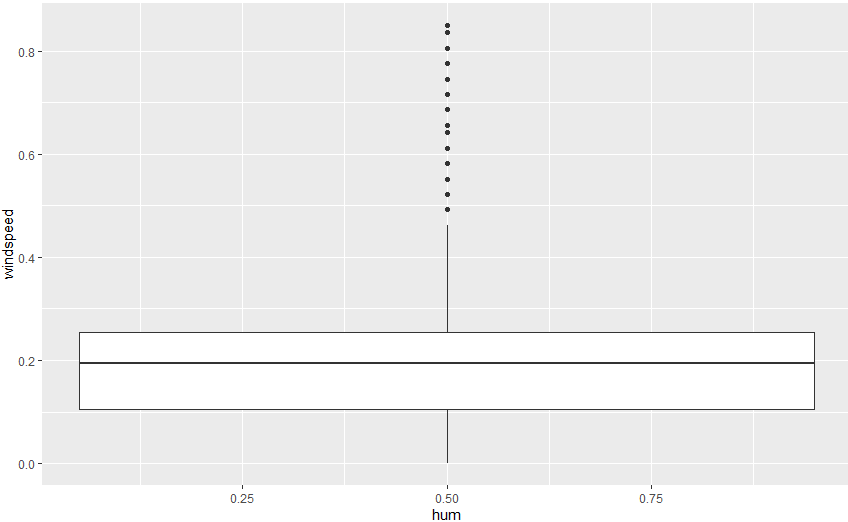
ggplot(data=hour,aes(x=hum,y=windspeed)) + geom\_point()



Insights:

Windspeed and humility are directly propotional.

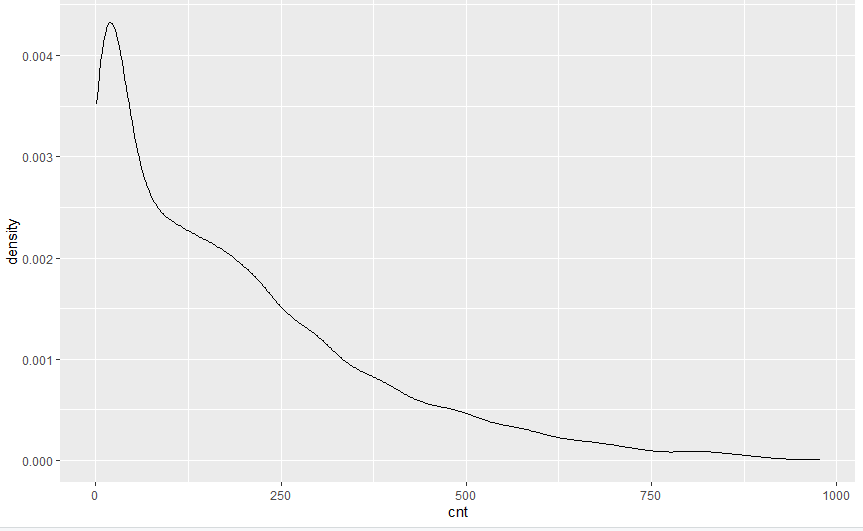
ggplot(data=hour,aes(x=hum,y=windspeed,fill=temp)) + geom\_boxplot()



Insights:

Outliers are present above the range of 0.5

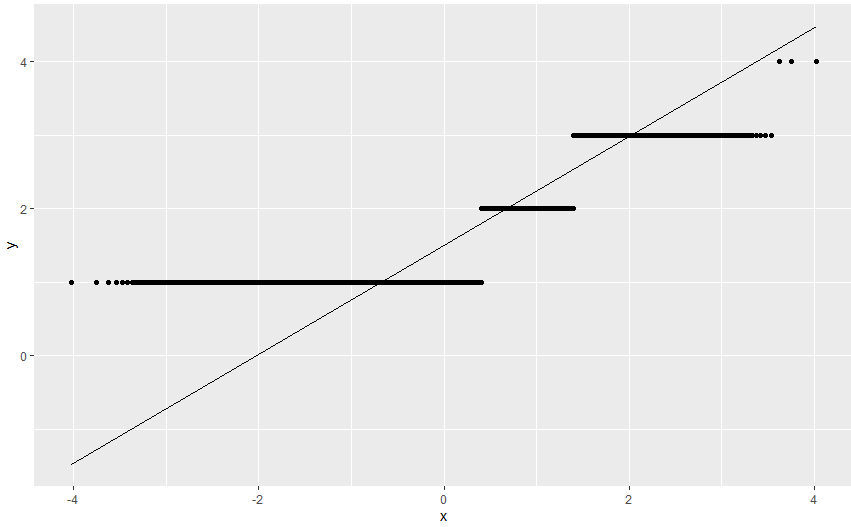
ggplot(data=hour,aes(x=year)) + geom\_density()



Insights:

As year passes production of energy decreases.

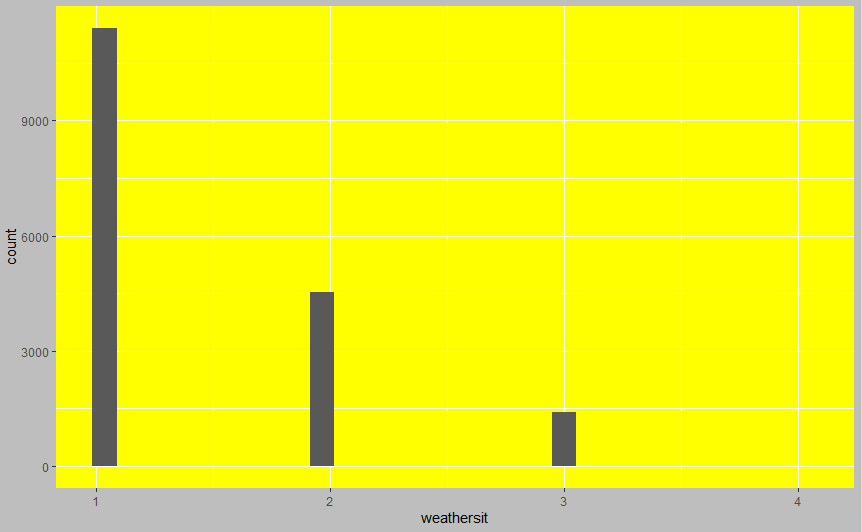
ggplot(data=hour,aes(sample=weathersit)) + geom\_qq() + geom\_qq\_line()



Insights:

This gives us the slope and intercept of the line connecting the points

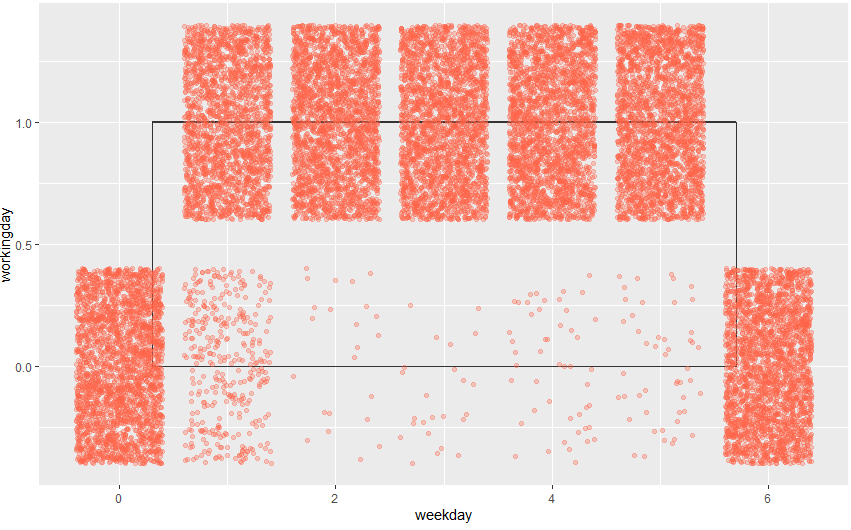
ggplot(data=hour,aes(x=weathersit)) + geom\_histogram() + theme(plot.background=element\_rect(fill='grey')) + theme(panel.background=element\_rect(fill='yellow'))



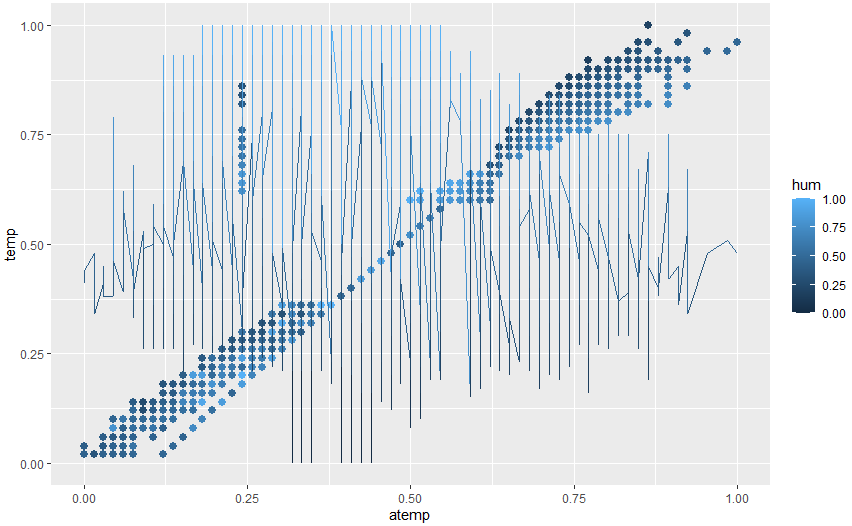
Insights:

I have assigned background colour for the histogram.

ggplot(data=hour,aes(x=weekday,y=workingday)) + geom\_boxplot(alpha=0) + geom\_jitter(alpha=0.3,colour='tomato')



ggplot(data=hour,aes(x=atemp,y=temp,colour=hum)) + geom\_point(size=2.5) + geom\_line(aes(y=hum))



Insights:

Here temperature is plotted against humility.

Result:

Thus exploration of various variable and row filters in R for cleaning data and applying various plot features in R on sample data sets and visualization were successfully executed.