

|  |  |
| --- | --- |
| **Student Name:** | Jayakarthi Boovendran |
| **Student ID**: | 19230487 |
| **Course:** | MBY – MSc Business Analytics |
| **Assignment Topic**: | Data Visualizations using R |

**Part – I**

Dataset: Marine Institute Buoy Wave Forecast

Description:

* Contains Wave climate forecasts data which includes significant wave height, mean wave period, mean wave direction, wave per unit crest length, peak period and energy period at Marine Institute weather and wave buoy locations
* Available @ [https://data.gov.ie/dataset/marine-institute-buoy-wave-forecast](https://data.gov.ie/dataset/marine-institute-buoy-wave-forecast/resource/1e9e1da0-2808-4ab5-8f00-7e22669424ce)

**Histograms**

**a) Histogram using R base graphics**

* **R Script**

data<-read.csv(url("http://erddap.marine.ie/erddap/tabledap/IMI-WaveBuoyForecast.csv?time,longitude,latitude,stationID,significant\_wave\_height,mean\_wave\_period,mean\_wave\_direction,wave\_power\_per\_unit\_crest\_length,peak\_period,energy\_period"))

data$significant\_wave\_height<-as.numeric(as.character(data$significant\_wave\_height))

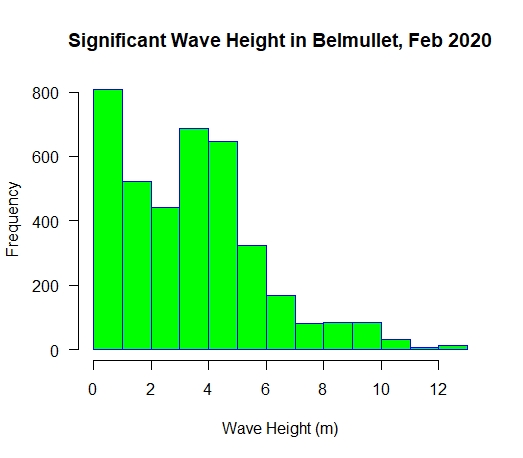
hist(data$significant\_wave\_height,

main="Significant Wave Height in Belmullet, Feb 2020",

xlab="Wave Height (m)", ylab="Frequency",

border="blue", col="green", las=1)

* **Output**



*Fig 1: Histogram for Significant Wave Height using Base R Graphics*

**b) Histogram using GGPLOT2**

* **R Script**

install.packages("ggplot2")

library("ggplot2")

ggplot(data=data, aes(data$significant\_wave\_height)) +

geom\_histogram(

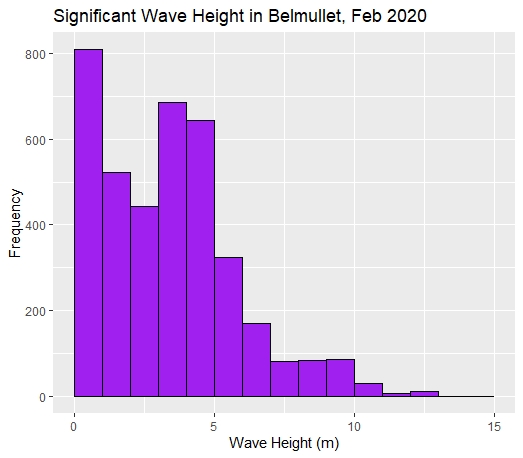
breaks=seq(0, 15, by = 1),

col='black', fill='purple')+

labs(title="Significant Wave Height in Belmullet, Feb 2020", x="Wave Height (m)", y="Frequency")+

xlim(c(0,15))

* **Output**



*Fig 2: Histogram for Significant Wave Height using GGPLOT2*

**Scatter Plots**

**a) Scatter plot using R base graphics**

* **R Script**

data<-read.csv(url("http://erddap.marine.ie/erddap/tabledap/IMI- WaveBuoyForecast.csv?time,longitude,latitude,stationID,significant\_wave\_height,mean\_wave\_period,mean\_wave\_direction,wave\_power\_per\_unit\_crest\_length,peak\_period,energy\_period"))

data$significant\_wave\_height<-as.numeric(as.character(data$significant\_wave\_height))

data$wave\_power\_per\_unit\_crest\_length<-as.numeric(as.character(data$wave\_power\_per\_unit\_crest\_length))

filtered\_data<-data[sample(nrow(data), 1000), ]

x <- filtered\_data$significant\_wave\_height

y <- filtered\_data$wave\_power\_per\_unit\_crest\_length

plot(x, y, main = "Wave Height vs Wave Power",

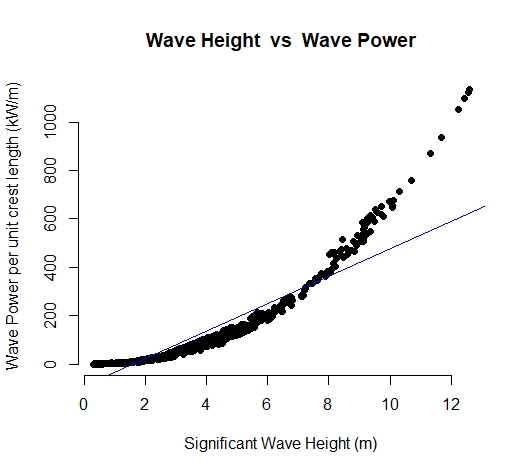
xlab = "Significant Wave Height (m)",

ylab = "Wave Power per unit crest length (kW/m)",

pch = 19, frame = FALSE)

abline(lm(y ~ x, data = mtcars), col = "blue")

* **Output**

****

*Fig 3: Scatter plot for Significant Wave Height and Wave Power using Base R Graphics*

**b) Scatter plot using GGPLOT2**

* **R Script**

install.packages("ggplot2")

library("ggplot2")

data<-read.csv(url("http://erddap.marine.ie/erddap/tabledap/IMI-WaveBuoyForecast.csv?time,longitude,latitude,stationID,significant\_wave\_height,mean\_wave\_period,mean\_wave\_direction,wave\_power\_per\_unit\_crest\_length,peak\_period,energy\_period"))

data$significant\_wave\_height<-as.numeric(as.character(data$significant\_wave\_height))

data$wave\_power\_per\_unit\_crest\_length<-as.numeric(as.character(data$wave\_power\_per\_unit\_crest\_length))

filtered\_data<-data[sample(nrow(data), 1000), ]

x <- filtered\_data$significant\_wave\_height

y <- filtered\_data$wave\_power\_per\_unit\_crest\_length

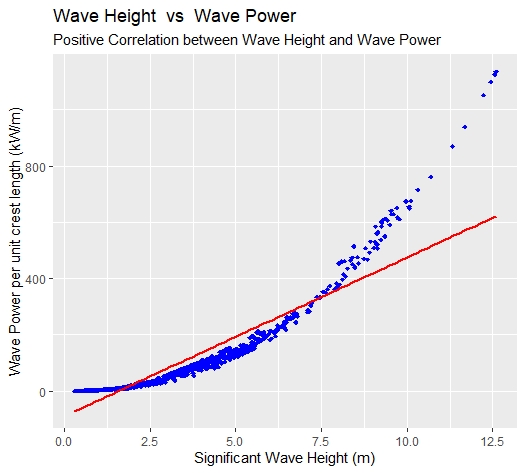
ggplot(filtered\_data, aes(x=filtered\_data$significant\_wave\_height, y=filtered\_data$wave\_power\_per\_unit\_crest\_length)) + geom\_point(shape=18, color="blue")+

geom\_smooth(method=lm, se=FALSE,color="red")+xlab("Significant Wave Height (m)")+

ylab('Wave Power per unit crest length (kW/m)')+

labs(title = "Wave Height vs Wave Power",subtitle='Positive Correlation between Wave Height and Wave Power')

* **Output**

****

*Fig 4: Scatter plot for Significant Wave Height and Wave Power using GGPLOT2*

**Part –II**

**a) Time Series plot for ‘Second Hand Apartment Prices’ dataset using R base graphics**

* **R Script**

library(lubridate)

require(graphics)

data<-read.csv("G:/MS5108 - Applied Customer Analysis/Assignments/2/Second hand Appartment Prices.csv")

data$YEAR<-year(as.Date(as.character(data$YEAR), format = "%Y"))

data$National<-as.numeric(gsub(",","",data$National))

data$Dublin<-as.numeric(gsub(",","",data$Dublin))

data$Cork<-as.numeric(gsub(",","",data$Cork))

data$Galway<-as.numeric(gsub(",","",data$Galway))

data$Limerick<-as.numeric(gsub(",","",data$Limerick))

data$Waterford<-as.numeric(gsub(",","",data$Waterford))

data$Other.Areas<-as.numeric(gsub(",","",data$Other.Areas))

color<-rainbow(ncol(data[2:8]))

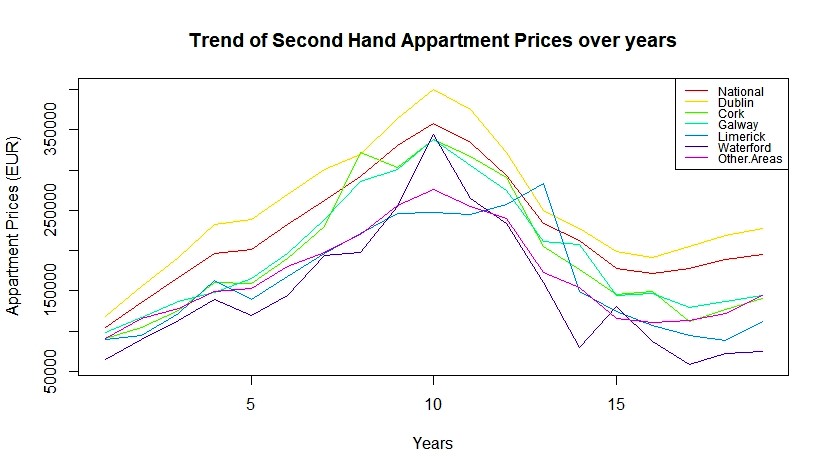
ts.plot(data[2:8],gpars=list(main="Trend of Second Hand Appartment Prices over years",xlab="Years", ylab="Appartment Prices (EUR)",lty=1,col=color,axes=FALSE))

axis(2) axis(1)

legend("topright", legend=colnames(data[2:8]), lty=1, col=color,cex = 0.75)

box()

Output



*Fig 5: Time Series plot for ‘Second Hand Apartment Prices’ using R base graphics*

**b) Time Series plot for ‘Second Hand Apartment Prices’ dataset using ‘dygraphs’**

* R Script

library(lubridate)

library(dygraphs)

library(xts)

AppartmentData<-read.csv("G:/MS5108 - Applied Customer Analysis/Assignments/2/Second hand Appartment Prices.csv")

AppartmentData$YEAR<-ymd(AppartmentData$YEAR, truncated = 2L)

data=xts( x=AppartmentData[2:8], order.by=AppartmentData$YEAR)

data$National<-as.numeric(gsub(",","",data$National))

data$Dublin<-as.numeric(gsub(",","",data$Dublin))

data$Cork<-as.numeric(gsub(",","",data$Cork))

data$Galway<-as.numeric(gsub(",","",data$Galway))

data$Limerick<-as.numeric(gsub(",","",data$Limerick))

data$Waterford<-as.numeric(gsub(",","",data$Waterford))

data$Other.Areas<-as.numeric(gsub(",","",data$Other.Areas))

dygraph(data,main = "Trend of Second Hand Appartment Prices over years")%>%

dyOptions(stackedGraph=FALSE,fillGraph=FALSE,fillAlpha=0.2,drawGrid=FALSE,axisLineWidth = 2)%>%

dyRangeSelector(strokeColor = "orange")%>%

dyAxis("y", label = "Appartment Price (EUR)") %>%

dyLegend(width=150)%>%

dySeries("Other.Areas", label = "Others",fillGraph = FALSE, color ='red',strokeWidth = 2) %>%

dySeries("National", fillGraph = FALSE, color = "blue",strokeWidth = 2) %>%

dySeries("Dublin", fillGraph = FALSE, color = "coral",strokeWidth = 2) %>%

dySeries("Cork", fillGraph = FALSE, color = "black",strokeWidth = 2, strokePattern = "dashed") %>% dySeries("Galway", fillGraph = FALSE, color = "deeppink",strokeWidth = 2) %>%

dySeries("Limerick", fillGraph = FALSE, color = "navy",strokeWidth = 2) %>%

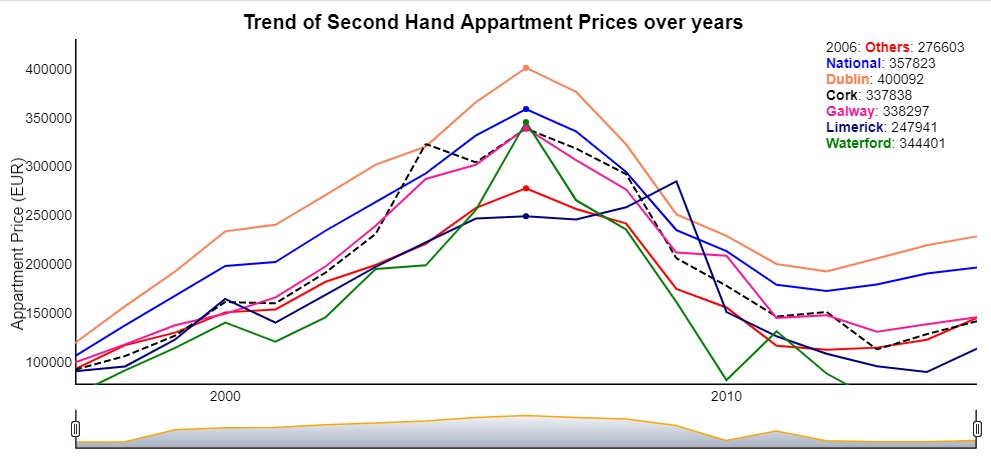
dySeries("Waterford", fillGraph = FALSE, color = "green",strokeWidth = 2)%>%

dyHighlight(highlightCircleSize = 5,

highlightSeriesBackgroundAlpha = 0.2,

hideOnMouseOut = FALSE,

highlightSeriesOpts = list(strokeWidth = 3))



*Fig 6: Time Series plot for ‘Second Hand Apartment Prices’ using dygraphs library*



*Fig 7: Time Series plot with ‘Cork’ highlighted on mouse hover;*

*(Jan 2006 – Jan 2010 selected in Range Selector)*

**Part –III**

**Towns of County Galway marked in a Map using R**

* **R Code**

install.packages('leaflet')

library('leaflet')

library(sp)

library(dplyr)

library(htmltools)

county\_data<-read.csv("G:/MS5108 - Applied Customer Analysis/Assignments/2/Irish Towns Co Ordinates.csv")

galway\_data<-filter(county\_data, county == "Galway")

galway\_data<-galway\_data[sample(nrow(galway\_data), 20), ]

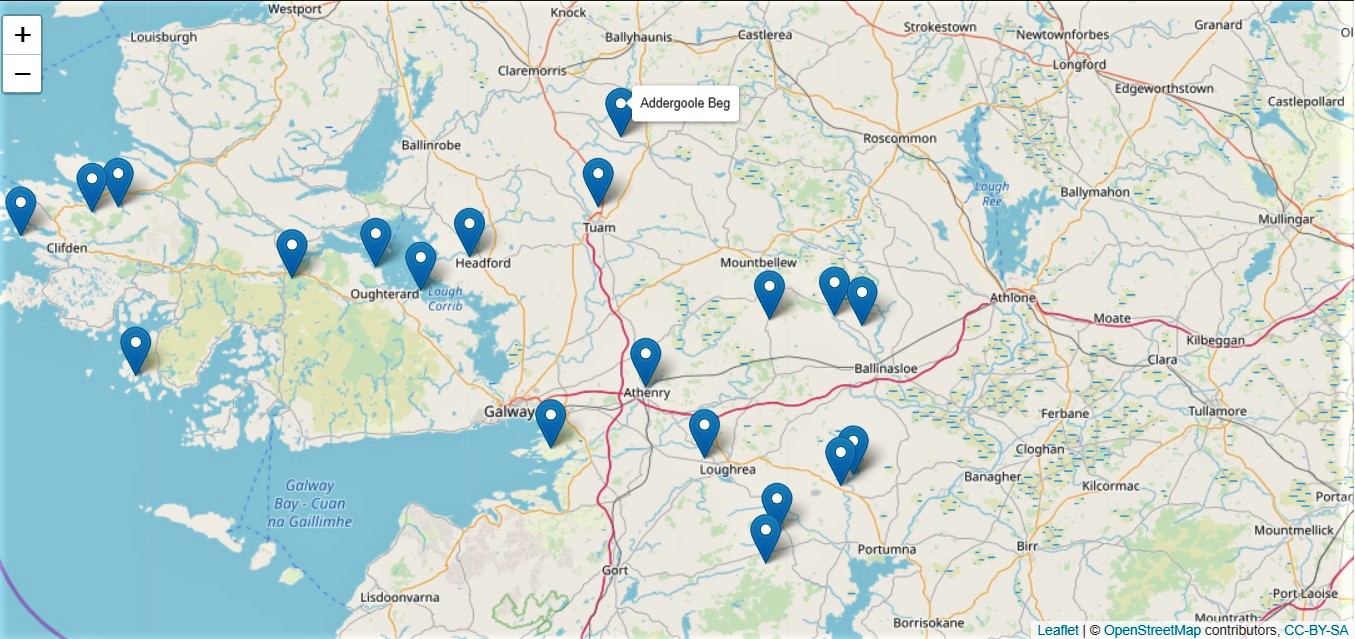
map<-leaflet() %>%

addTiles() %>%

addMarkers(data=galway\_data,lng =galway\_data$longitude ,lat =galway\_data$latitude,label = ~htmlEscape(galway\_data$name))

map

* **Output**

****

*Fig 8: Map showing 20 Towns from Co. Galway*

*(Town Names displayed as labels on mouse hover)*