Data Visualisation

Varun Dua

26/02/2020

## Part 1:

From the <https://data.gov.ie/> website, pick a dataset of your choice. In R Studio, represent any  
aspect of your dataset as a histogram using;  
a) The R base graphics, then  
b) GGPLOT2  
Using the same dataset, represent another aspect of the data as a scatter plot using;  
a) The R base graphics, then  
b) GGPLOT2 with colours

### Data:

<https://data.gov.ie/dataset/valentia-observatory-monthly-data>

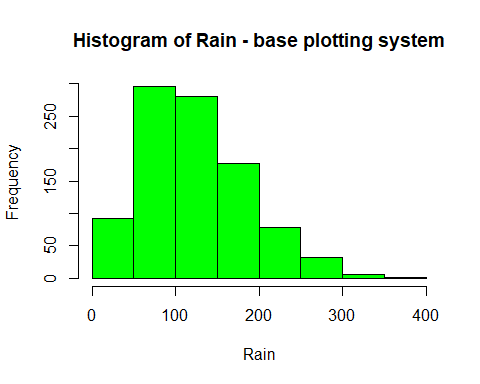
This dataset contains monthly elements measured at our synoptic station in Valentia Observatory, Co Kerry.The file is updated monthly. Values for each month include: Precipitation Amount, Mean Air Temperature, Maximum Air Temperature (C), Minimum Air Temperature, Mean Maximum Temperature, Mean Minimum Temperature, Grass Minimum Temperature, Mean Win Speed, Highest Gust, Sunshine duration.

### (A) Using Base Plotting System

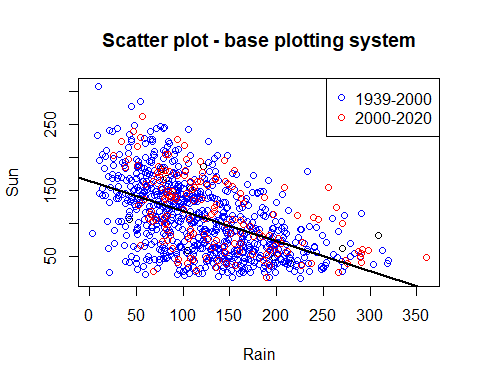
obs\_data <- read.csv("mly2275.csv", skip = 19, header = TRUE)  
head(obs\_data)

## year month meant maxtp mintp mnmax mnmin rain gmin wdsp maxgt sun  
## 1 1939 10 10.0 16.8 0.0 13.2 6.7 105.5 -4.4 9.9 55 140.0  
## 2 1939 11 10.3 14.0 4.0 12.4 8.1 251.9 0.5 13.9 60 36.4  
## 3 1939 12 6.0 13.6 -3.3 8.7 3.3 116.9 -7.8 10.6 59 65.8  
## 4 1940 1 5.8 12.3 -5.6 8.4 3.3 163.8 -8.4 11.6 51 76.3  
## 5 1940 2 7.9 12.8 -3.9 10.1 5.7 179.6 -5.8 13.0 55 67.2  
## 6 1940 3 8.5 13.6 -1.2 11.0 6.0 84.8 -3.9 12.0 52 119.0

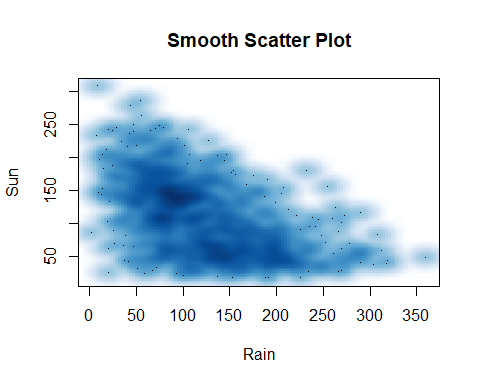
hist\_base <- hist(obs\_data$rain, main = "Histogram of Rain - base plotting system", xlab = "Rain", col = "Green")



with(obs\_data, plot(rain, sun, main = "Scatter plot - base plotting system",xlab = "Rain", ylab = "Sun", type = "p", pch = 1 ))  
with(subset(obs\_data, year<2000), points(rain,sun, col = "blue"))  
with(subset(obs\_data, year>2000), points(rain,sun, col = "red"))  
legend("topright", pch = 1, col = c("blue","red"), legend = c("1939-2000", "2000-2020"))  
  
model<- lm(sun ~ rain, obs\_data)  
abline(model, lwd=2)

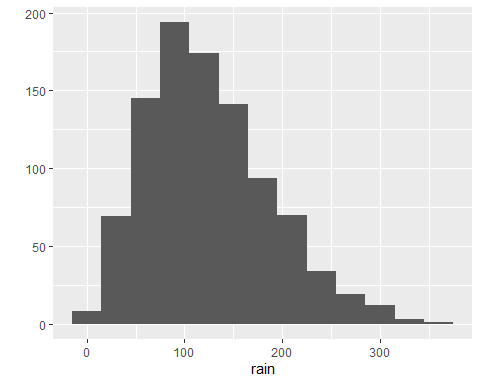


smoothScatter(obs\_data$rain, obs\_data$sun, xlab = "Rain", ylab = "Sun", main = "Smooth Scatter Plot")



### (B) Using ggplot2 Plotting System (function qplot)

library(ggplot2)  
qplot(rain, data = obs\_data, binwidth = 30)

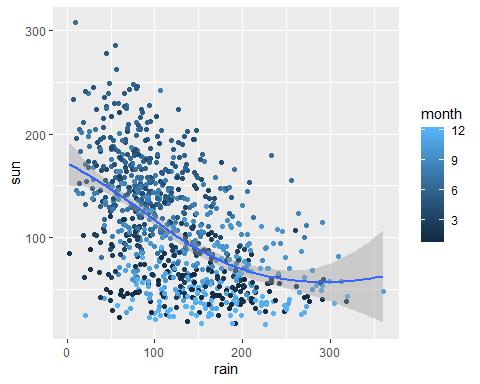


qplot(rain,sun, data = obs\_data, color = month, geom = c("point", "smooth"))

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 79 rows containing non-finite values (stat\_smooth).

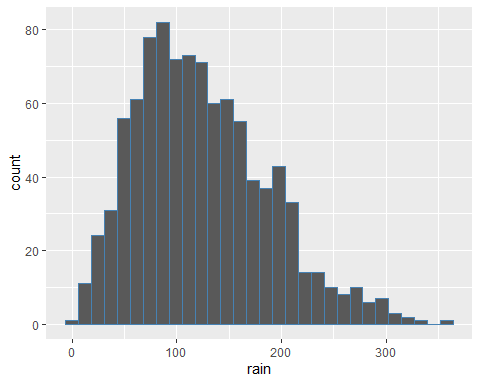
## Warning: Removed 79 rows containing missing values (geom\_point).



### (C) Using ggplot2 Plotting System (function ggplot)

ggplot(obs\_data, aes(rain)) + geom\_histogram(color = "steelblue")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

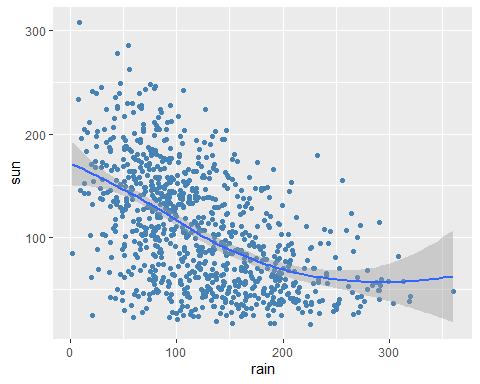


g<- ggplot(obs\_data, aes(rain,sun))  
g + geom\_point(color = "steelblue") + geom\_smooth()

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 79 rows containing non-finite values (stat\_smooth).

## Warning: Removed 79 rows containing missing values (geom\_point).



## Part2:

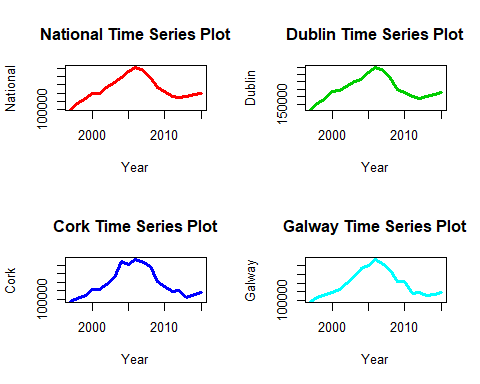
Using the “Second Hand Apartment Prices” dataset on Blackboard, perform a time series plot of  
price changes from year to year for all locations using;  
a) The R base graphics  
b) Any other R functions which will enable you to produce a more professional and higher  
quality graphic.

### (A) Using Base Plotting System

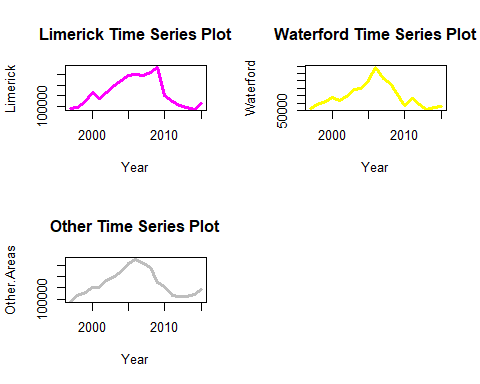
apt\_data <- read.csv("Second hand Appartment Prices.csv", skip = 1, header = TRUE)  
apt\_data

## X National Dublin Cork Galway Limerick Waterford Other.Areas  
## 1 1997 104,511 117,937 90,638 97,969 88,932 64,937 91,166  
## 2 1998 136,099 155,793 104,643 116,642 94,019 90,024 115,779  
## 3 1999 166,492 191,235 125,514 136,211 121,574 112,981 128,647  
## 4 2000 196,949 232,431 160,047 147,914 163,216 138,869 149,350  
## 5 2001 201,014 239,238 158,691 164,763 138,977 119,477 152,440  
## 6 2002 233,080 269,651 190,166 196,515 167,312 144,187 180,778  
## 7 2003 262,459 300,906 229,819 238,265 196,116 193,852 197,998  
## 8 2004 291,758 319,333 322,076 286,113 221,118 197,642 219,627  
## 9 2005 330,844 364,708 303,132 300,655 245,618 254,427 256,480  
## 10 2006 357,823 400,092 337,838 338,297 247,941 344,401 276,603  
## 11 2007 334,959 375,457 317,363 305,529 244,700 264,426 255,402  
## 12 2008 293,443 321,552 291,054 275,316 257,086 234,288 240,441  
## 13 2009 233,663 249,894 204,929 210,763 283,786 160,000 173,387  
## 14 2010 212,202 227,800 176,673 207,454 149,672 80,000 154,392  
## 15 2011 177,746 199,051 145,102 143,781 124,897 130,000 115,247  
## 16 2012 171,237 191,452 149,792 146,492 107,052 86,700 111,165  
## 17 2013 177,990 204,678 111,443 129,637 94,125 58,963 113,272  
## 18 2014 189,263 218,275 127,098 137,125 88,409 71,896 121,367  
## 19 2015 195,332 227,303 140,331 144,347 112,195 74,284 143,841

apt\_data$National<-as.numeric(gsub(",", "", apt\_data$National))  
apt\_data$Dublin<-as.numeric(gsub(",", "", apt\_data$Dublin))  
apt\_data$Cork<-as.numeric(gsub(",", "", apt\_data$Cork))  
apt\_data$Galway<-as.numeric(gsub(",", "", apt\_data$Galway))  
apt\_data$Limerick<-as.numeric(gsub(",", "", apt\_data$Limerick))  
apt\_data$Waterford<-as.numeric(gsub(",", "", apt\_data$Waterford))  
apt\_data$Other.Areas<-as.numeric(gsub(",", "", apt\_data$Other.Areas))  
  
par(mfrow =c(2,2))  
  
with(apt\_data, plot(X, National, type = 'l', col = 2, main = "National Time Series Plot", xlab = "Year", lwd = 3))  
with(apt\_data, plot(X, Dublin, type = 'l', col = 3, main = "Dublin Time Series Plot", xlab = "Year", lwd = 3))  
with(apt\_data, plot(X, Cork, type = 'l', col = 4, main = "Cork Time Series Plot", xlab = "Year", lwd = 3))   
with(apt\_data, plot(X, Galway, type = 'l', col = 5, main = "Galway Time Series Plot", xlab = "Year", lwd = 3))

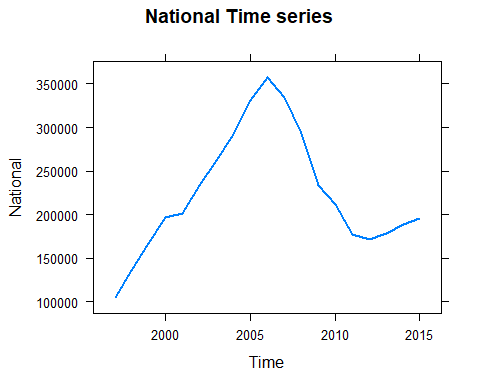


with(apt\_data, plot(X, Limerick, type = 'l', col = 6, main = "Limerick Time Series Plot", xlab = "Year", lwd = 3))  
with(apt\_data, plot(X, Waterford, type = 'l', col = 7, main = "Waterford Time Series Plot", xlab = "Year", lwd = 3))  
with(apt\_data, plot(X, Other.Areas, type = 'l', col = 8, main = "Other Time Series Plot", xlab = "Year", lwd = 3))

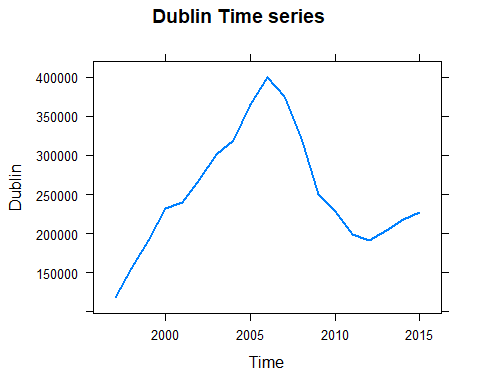


### (B) Using Lattice Plotting System

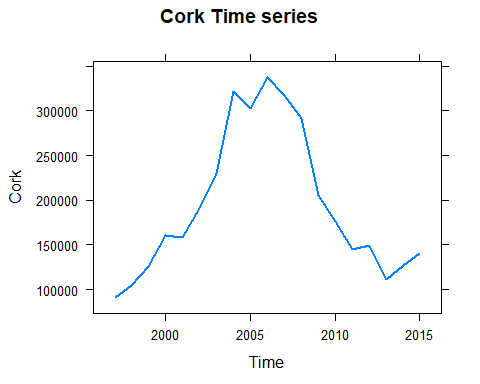
library(lattice)  
  
xyplot(National ~ X, data = apt\_data, type = 'l', lwd = 2, main = "National Time series", xlab = "Time")



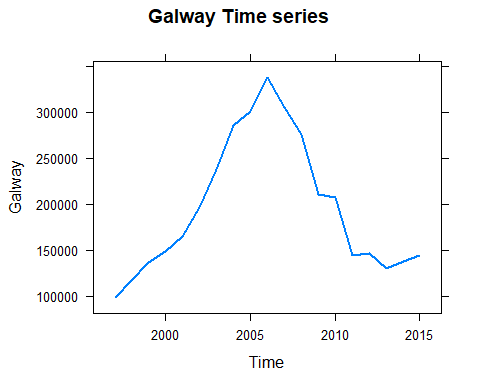
xyplot(Dublin ~ X, data = apt\_data, type = 'l', lwd = 2, main = "Dublin Time series", xlab = "Time")



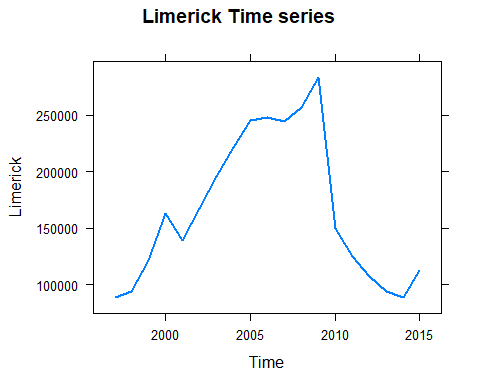
xyplot(Cork ~ X, data = apt\_data, type = 'l', lwd = 2, main = "Cork Time series", xlab = "Time")



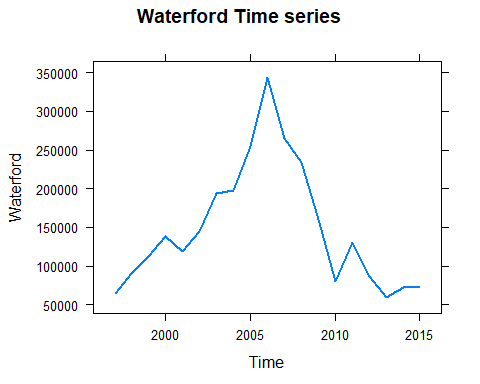
xyplot(Galway ~ X, data = apt\_data, type = 'l', lwd = 2, main = "Galway Time series", xlab = "Time")



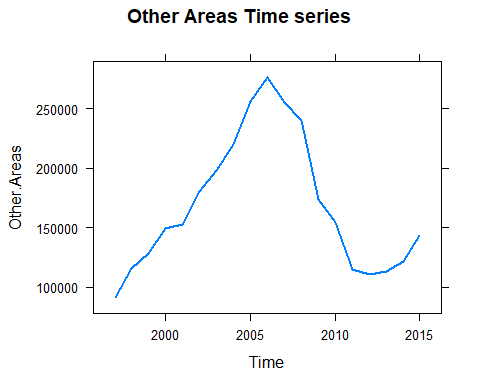
xyplot(Limerick ~ X, data = apt\_data, type = 'l', lwd = 2, main = "Limerick Time series", xlab = "Time")



xyplot(Waterford ~ X, data = apt\_data, type = 'l', lwd = 2, main = "Waterford Time series", xlab = "Time")

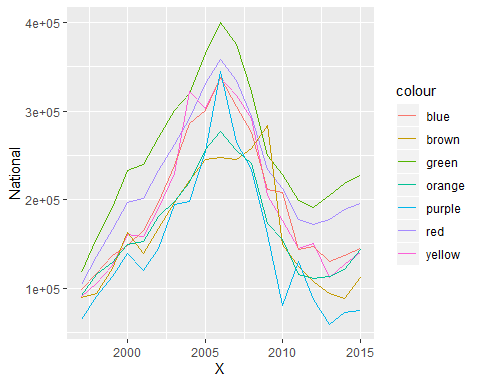


xyplot(Other.Areas ~ X, data = apt\_data, type = 'l', lwd = 2, main = "Other Areas Time series", xlab = "Time")



### (C) Using ggplot2 Plotting System

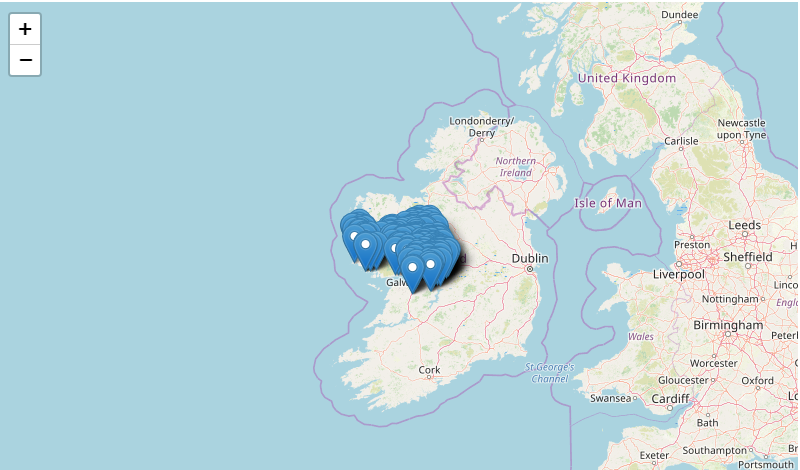
ggplot(apt\_data, aes(X)) + geom\_line(aes(X,National, color = "red")) + geom\_line(aes(X,Dublin, color = "green")) + geom\_line(aes(X,Cork, color = "yellow")) + geom\_line(aes(X,Galway, color = "blue")) + geom\_line(aes(X,Limerick, color = "brown")) + geom\_line(aes(X,Waterford, color = "purple")) + geom\_line(aes(X,Other.Areas, color = "orange"))



## Part3:

Selecting an Irish county of your choice, find the names and co-ordinates (easting, northing) of  
towns in that county. You must have at least 20 towns. Using your knowledge of R, construct a  
visual where the town names (or abbreviations of those town names) are placed on a map as per  
their co-ordinates. Make this visual as professional as possible.

library(leaflet)  
library(sp)  
map\_data<-read.csv("Irish Towns Co Ordinates.csv", header = TRUE)  
t<- map\_data$county == "Galway"  
galway\_data<-map\_data[t,]  
coords<-data.frame(galway\_data$longitude, galway\_data$latitude)  
sp<-SpatialPointsDataFrame(coords = coords, galway\_data)  
map<-leaflet() %>% addProviderTiles(providers$OpenStreetMap)   
map\_ireland<-map %>% addMarkers(data = sp, label = galway\_data$name)

  
  
qplot(easting, northing, data = galway\_data, geom = "text", label = name, size=0.25)

