library(tidyverse)

## -- Attaching packages ----------------------------------------------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.0 v purrr 0.3.3  
## v tibble 3.0.1 v dplyr 0.8.5  
## v tidyr 1.0.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

## -- Conflicts -------------------------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(ggplot2)  
library(lubridate)

##   
## Attaching package: 'lubridate'

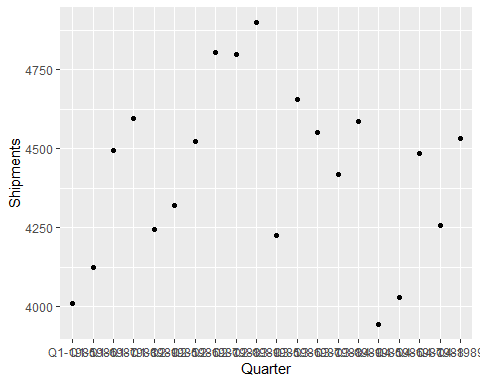
## The following object is masked from 'package:base':  
##   
## date

shipment <- read.csv(file = "C:/Users/UParekh/Desktop/Udit/ApplianceShipments.csv", header = TRUE)

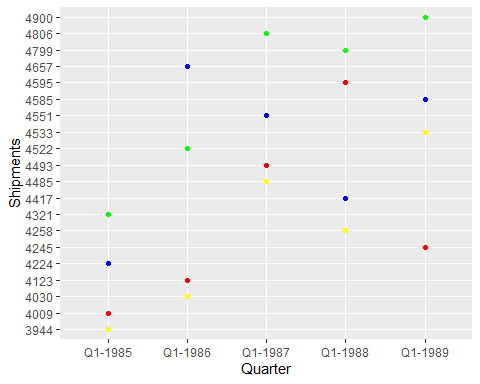
Problem one: Shipments of household appliances

#a  
#There was only one observation per quarter, a geom\_point is a better form of visualization  
  
ggplot(data = shipment, mapping = aes(x = Quarter, y = Shipments)) +  
 geom\_point() +  
 geom\_smooth(se = FALSE)

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



#c  
Quarter <- c('Q1-1985', 'Q1-1986', 'Q1-1987', 'Q1-1988', 'Q1-1989')  
Shipments <- c('4009', '4123', '4493', '4595', '4245')  
  
Q1 <- data.frame(Quarter, Shipments, stringsAsFactors = FALSE)  
  
Quarter <- c('Q2-1985', 'Q2-1986', 'Q2-1987', 'Q2-1988', 'Q2-1989')  
Shipments <- c('4321', '4522', '4806', '4799', '4900')  
  
Q2 <- data.frame(Quarter, Shipments, stringsAsFactors = FALSE)  
  
Quarter <- c('Q3-1985', 'Q3-1986', 'Q3-1987', 'Q3-1988', 'Q3-1989')  
Shipments <- c('4224', '4657', '4551', '4417', '4585')  
  
Q3 <- data.frame(Quarter, Shipments, stringsAsFactors = FALSE)  
  
Quarter <- c('Q4-1985', 'Q4-1986', 'Q4-1987', 'Q4-1988', 'Q4-1989')  
Shipments <- c('3944', '4030', '4485', '4258', '4533')  
  
Q4 <- data.frame(Quarter, Shipments, stringsAsFactors = FALSE)  
df<- data.frame(Q1,Q2,Q3,Q4)  
  
  
ggplot(df, aes(x=Quarter)) +   
 geom\_point(aes(y=Shipments), colour="red") +  
 geom\_point(aes(y=Shipments.1), colour="green") +  
 geom\_point(aes(y=Shipments.2), colour="blue") +  
 geom\_point(aes(y=Shipments.3), colour="yellow")

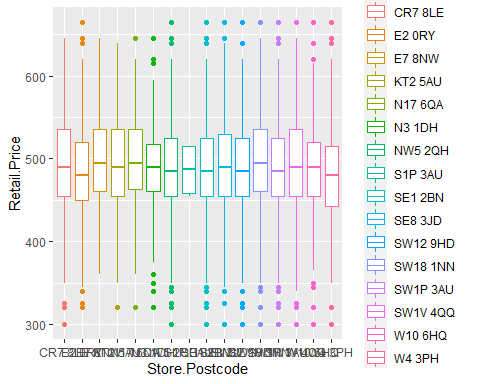


Problem 3: Laptop sales at a London computer chain

#a  
  
laptop <- read.csv(file = " C:/Users/UParekh/Desktop/Udit/LaptopSalesJanuary2008.csv", header = TRUE)  
  
# The lowest Average retail is of the store 'W4 3PH' with the average value being 481.0063  
# The highest Average retail is of the store 'N17 6QA' with the average value being 494.6341  
  
laptop%>%group\_by(Store.Postcode)%>%  
summarise(Avg\_retail = mean(Retail.Price,na.rm = TRUE))%>%  
arrange(Avg\_retail)

## # A tibble: 16 x 2  
## Store.Postcode Avg\_retail  
## <fct> <dbl>  
## 1 W4 3PH 481.  
## 2 E2 0RY 483.  
## 3 SW12 9HD 485.  
## 4 S1P 3AU 486.  
## 5 NW5 2QH 487.  
## 6 SE1 2BN 487.  
## 7 N3 1DH 487.  
## 8 SW1P 3AU 489.  
## 9 CR7 8LE 489.  
## 10 SW1V 4QQ 489.  
## 11 W10 6HQ 490.  
## 12 SE8 3JD 492.  
## 13 SW18 1NN 493.  
## 14 KT2 5AU 494.  
## 15 E7 8NW 494.  
## 16 N17 6QA 495.

#b  
   
ggplot(data = laptop, aes(x = Store.Postcode, y = Retail.Price, colour = Store.Postcode)) +  
 geom\_boxplot()



Problem 4: Laptop sales at a London computer chain

laptop\_sales <- read.csv(file = " C:/Users/UParekh/Desktop/Udit/LaptopSales.csv", header = TRUE)  
  
#a  
#i THe laptops are actually sold at the 'Retail.Price'  
  
  
#iii  
#The prices are the same at different stores. I took 'Processor.Speeds..GHz.' as a reference column and same processor computers cost the same at different stroes which led me to this conclusion  
  
laptop\_sales%>%group\_by(Processor.Speeds..GHz.)%>%  
summarise(Avg\_retail = mean(Retail.Price,na.rm = TRUE))%>%  
arrange(Processor.Speeds..GHz.)

## # A tibble: 3 x 2  
## Processor.Speeds..GHz. Avg\_retail  
## <dbl> <dbl>  
## 1 1.5 485.  
## 2 2 511.  
## 3 2.4 535.

#iv  
#The prices do change with the configuration of the computer. Although, as you can see there is no such relation between configuration and retail price  
  
laptop\_sales%>%group\_by(Configuration)%>%  
summarise(Avg\_retail = mean(Retail.Price,na.rm = TRUE))%>%  
arrange(Configuration)

## # A tibble: 864 x 2  
## Configuration Avg\_retail  
## <int> <dbl>  
## 1 1 338.  
## 2 2 289.  
## 3 3 377.  
## 4 4 329.  
## 5 5 392.  
## 6 6 348.  
## 7 7 456.  
## 8 8 406.  
## 9 9 316.  
## 10 10 268.  
## # ... with 854 more rows

#b  
#i The stores and customers are located in differet regions and they are mapped through the postal codes in this dataset  
  
#ii  
# The highest Average retail is of the store 'N17 6QA' with the average value being 522.9433  
laptop\_sales%>%group\_by(Store.Postcode)%>%  
summarise(Avg\_retail = mean(Retail.Price,na.rm = TRUE))%>%  
arrange(Avg\_retail)

## # A tibble: 16 x 2  
## Store.Postcode Avg\_retail  
## <fct> <dbl>  
## 1 E7 8NW 467.  
## 2 W4 3PH 469.  
## 3 SW1P 3AU 470.  
## 4 N3 1DH 471.  
## 5 CR7 8LE 472.  
## 6 S1P 3AU 512.  
## 7 SE1 2BN 520.  
## 8 W10 6HQ 520.  
## 9 E2 0RY 520.  
## 10 SW18 1NN 520.  
## 11 SW1V 4QQ 520.  
## 12 SE8 3JD 521.  
## 13 NW5 2QH 521.  
## 14 SW12 9HD 521.  
## 15 KT2 5AU 522.  
## 16 N17 6QA 523.

#iv  
cust\_travel <- data.frame(laptop\_sales$customer.X,laptop\_sales$customer.Y,laptop\_sales$store.X, laptop\_sales$store.Y)  
  
head(cust\_travel)

## laptop\_sales.customer.X laptop\_sales.customer.Y laptop\_sales.store.X  
## 1 532041 180995 534057  
## 2 529240 175537 528739  
## 3 533095 181047 535652  
## 4 529902 179641 534057  
## 5 531684 180948 528924  
## 6 529207 180969 528924  
## laptop\_sales.store.Y  
## 1 179682  
## 2 173080  
## 3 182961  
## 4 179682  
## 5 178440  
## 6 178440