Sales-Visualization.R

Arun Kumar Prasad

2021-07-24

```
r = getOption("repos")
r["CRAN"] = "http://cran.us.r-project.org"
options(repos = r)


SalesData = read.csv("D:/praneeta/praneeta/R/R case study 3 (Visualization)/SalesData.csv")
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(ggplot2)
install.packages("reshape2")
```

```
## Installing package into 'C:/Users/Arun Kumar Prasad/Documents/R/win-library/4.0'
## (as 'lib' is unspecified)

## package 'reshape2' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Arun Kumar Prasad\AppData\Local\Temp\Rtmp29Hv1G\downloaded_packages
```

```
library(reshape2)
install.packages("tidyr")
```

```
## Installing package into 'C:/Users/Arun Kumar Prasad/Documents/R/win-library/4.0'
## (as 'lib' is unspecified)

## package 'tidyr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Arun Kumar Prasad\AppData\Local\Temp\Rtmp29Hv1G\downloaded_packages
```

```
library(tidyr)
```

```
##
## Attaching package: 'tidyr'
```

```
## The following object is masked from 'package:reshape2':
##
##     smiths
```

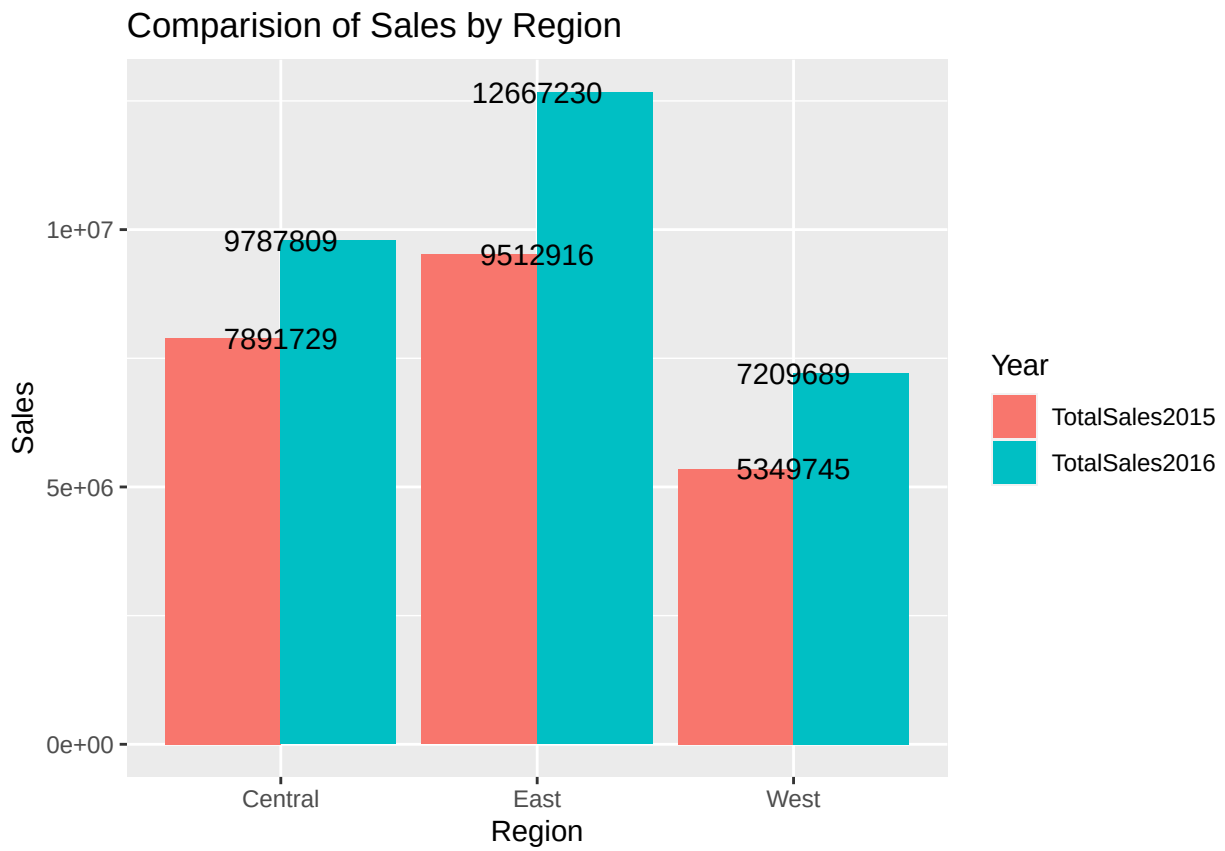
```
View(SalesData)
#question1
summ_data = SalesData%>%
  group_by(Region)%>%
  summarise(TotalSales2015 = sum(Sales2015),TotalSales2016 = sum(Sales2016))
summ_data
```

```
## # A tibble: 3 x 3
##   Region  TotalSales2015 TotalSales2016
##   <chr>            <dbl>          <dbl>
## 1 Central       7891729.       9787809.
## 2 East          9512916.      12667230.
## 3 West          5349745.       7209689.
```

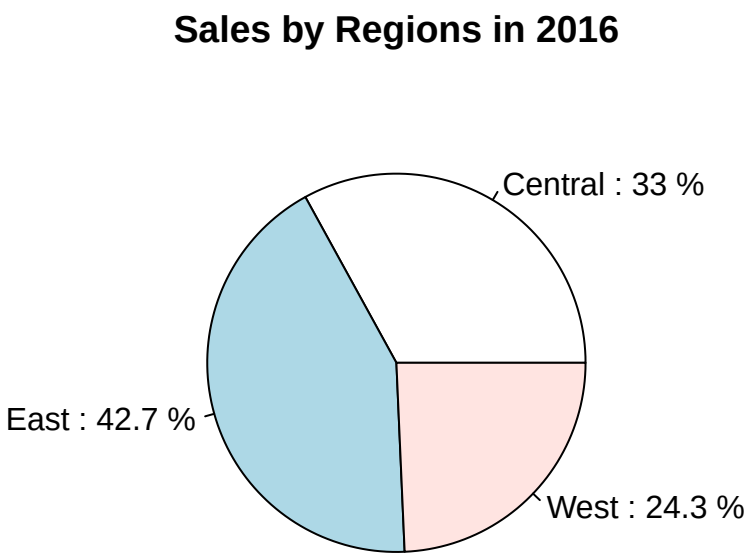
```
data_long = gather(summ_data,key = Year, value = Sales,2:3)
data_long
```

```
## # A tibble: 6 x 3
##   Region  Year               Sales
##   <chr>   <chr>              <dbl>
## 1 Central TotalSales2015  7891729.
## 2 East    TotalSales2015  9512916.
## 3 West    TotalSales2015  5349745.
## 4 Central TotalSales2016  9787809.
## 5 East    TotalSales2016 12667230.
## 6 West    TotalSales2016  7209689.
```

```
data_long$Sales <- round(data_long$Sales,0)
ggplot(data_long,aes(Region,Sales, fill = Year,label = Sales)) + geom_bar(stat = "identity",position =
  geom_text(size = 4) + xlab('Region') +ylab('Sales') + ggtitle("Comparision of Sales by Region")
```



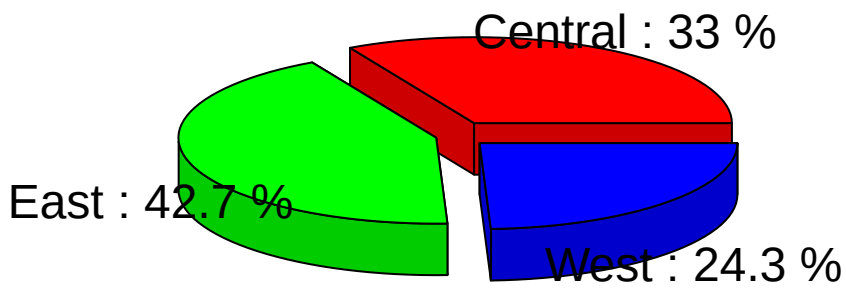

```
#question2
sales2016 <- SalesData%>%
  group_by(Region)%>%
  summarise(TotalSales2016 = sum(Sales2016))
value_percent <- sales2016$TotalSales2016/sum(sales2016$TotalSales2016)*100
pie(sales2016$TotalSales2016,labels = paste(sales2016$Region,":",round(value_percent,1),"%",sep = " "),
```




```
install.packages("plotrix")
```

```
## Installing package into 'C:/Users/Arun Kumar Prasad/Documents/R/win-library/4.0'
## (as 'lib' is unspecified)

## package 'plotrix' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Arun Kumar Prasad\AppData\Local\Temp\Rtmp29Hv1G\downloaded_packages
```

```
library(plotrix)
pie3D(sales2016$TotalSales2016,explode=0.1,labels = paste(sales2016$Region,":",round(value_percent,1),"
```



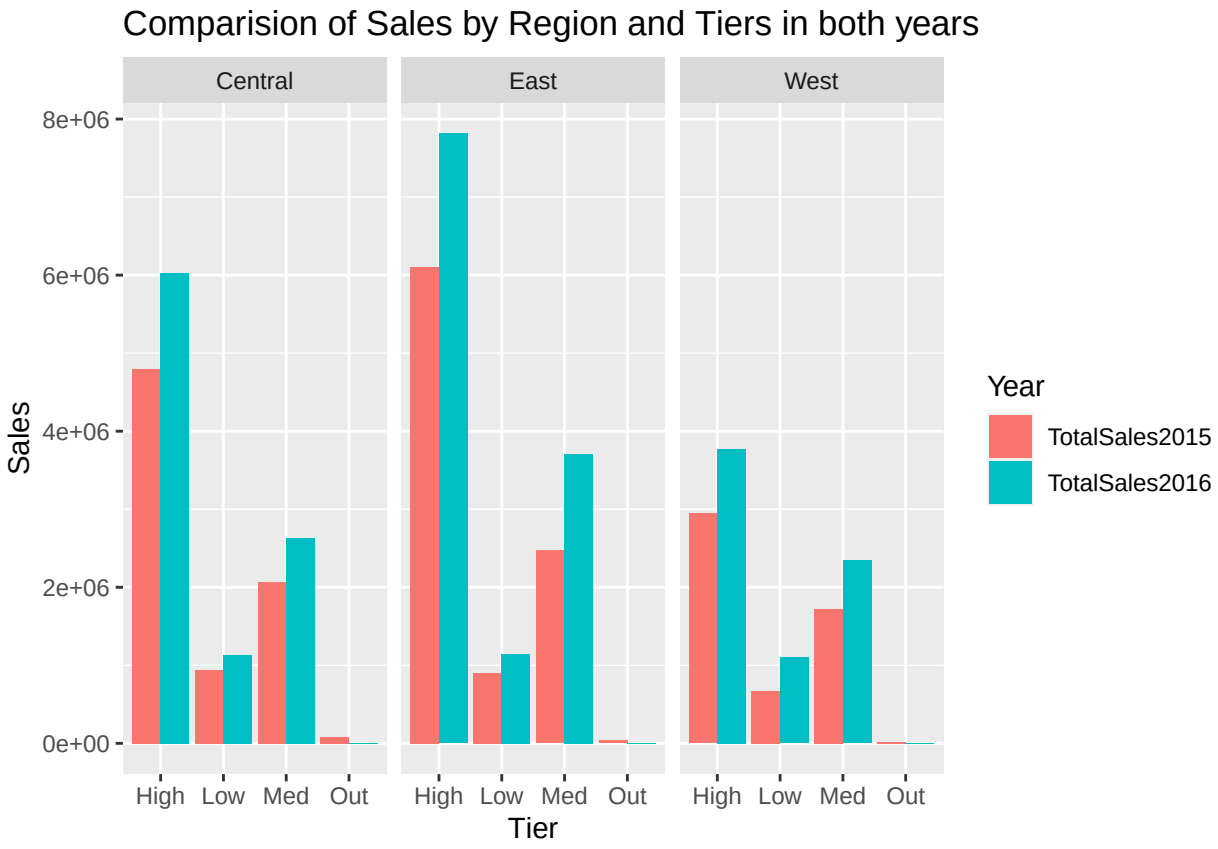

```
#question 3
rt_data <- SalesData%>%
  group_by(Region,Tier)%>%
  summarise(TotalSales2015 = sum(Sales2015),TotalSales2016 = sum(Sales2016))
```

```
## `summarise()` has grouped output by 'Region'. You can override using the `.groups` argument.
```

```
data_long1 <- gather(rt_data,key = Year,value = Sales,-c(Region,Tier))
data_long1
```

```
## # A tibble: 24 x 4
## # Groups:   Region [3]
##    Region  Tier  Year              Sales
##    <chr>   <chr> <chr>             <dbl>
##  1 Central High  TotalSales2015 4798698.
##  2 Central Low   TotalSales2015  943440.
##  3 Central Med   TotalSales2015 2068226.
##  4 Central Out   TotalSales2015   81365.
##  5 East    High  TotalSales2015 6102946.
##  6 East    Low   TotalSales2015  901666.
##  7 East    Med   TotalSales2015 2470998.
##  8 East    Out   TotalSales2015   37307.
##  9 West    High  TotalSales2015 2944789.
## 10 West    Low   TotalSales2015  671064.
## # ... with 14 more rows
```

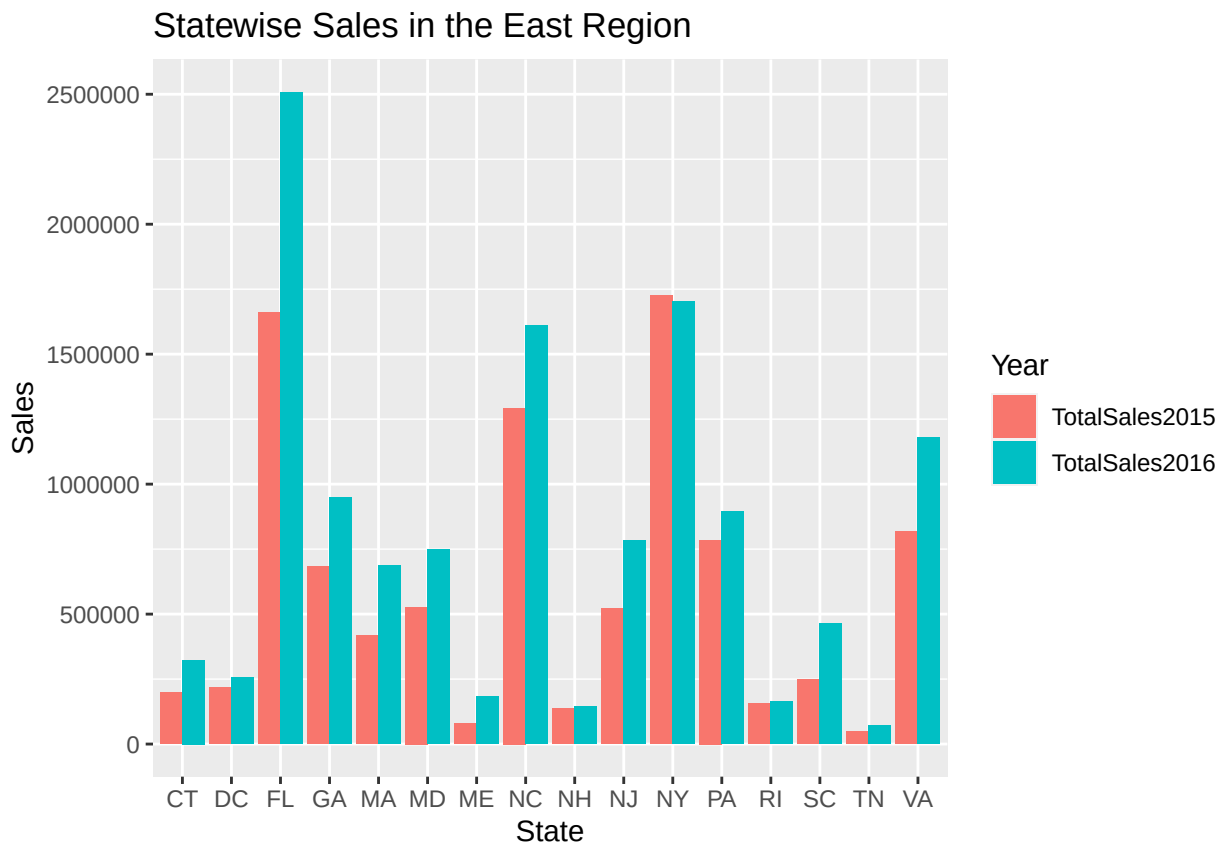
```
ggplot(data_long1,aes(Tier,Sales,fill = Year)) + geom_bar(stat = "identity", position = "dodge") + face
```



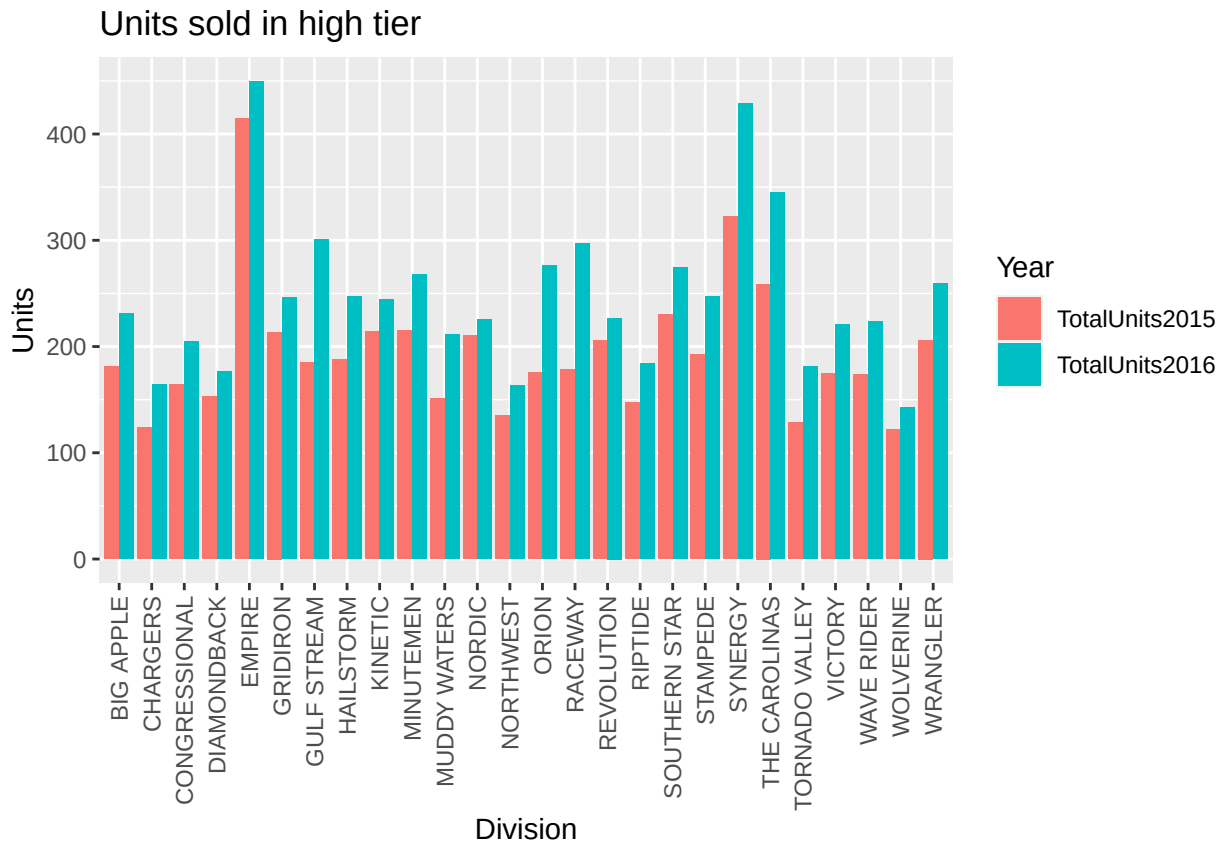

```
#question4
state_data <- SalesData %>%
  group_by(State)%>%
  filter(Region =="East")%>%
  summarise(TotalSales2015 = sum(Sales2015),TotalSales2016 = sum(Sales2016))

data_long2 <- gather(state_data,key=Year,value=Sales,-State)
data_long2
```

```
## # A tibble: 32 x 3
##    State Year              Sales
##    <chr> <chr>             <dbl>
##  1 CT    TotalSales2015  197203.
##  2 DC    TotalSales2015  216724.
##  3 FL    TotalSales2015 1660162
##  4 GA    TotalSales2015  681546.
##  5 MA    TotalSales2015  419279.
##  6 MD    TotalSales2015  527309.
##  7 ME    TotalSales2015   77426.
##  8 NC    TotalSales2015 1292802.
##  9 NH    TotalSales2015  136419.
## 10 NJ    TotalSales2015  520419.
## # ... with 22 more rows
```

```
ggplot(data_long2,aes(State,Sales,fill=Year)) + geom_bar(stat="Identity",position = "dodge") +ggtitle("
```




```
#question5
hightierdata <- SalesData %>%
  group_by(Division) %>%
  filter(Tier == "High") %>%
  summarise(TotalUnits2015 = sum(Units2015), TotalUnits2016 = sum(Units2016))

data_long3 <- gather(hightierdata,key=Year,value=Units,-Division)
ggplot(data_long3, aes(Division,Units,fill=Year)) + geom_bar(stat = "Identity", position="dodge") +ggti
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```



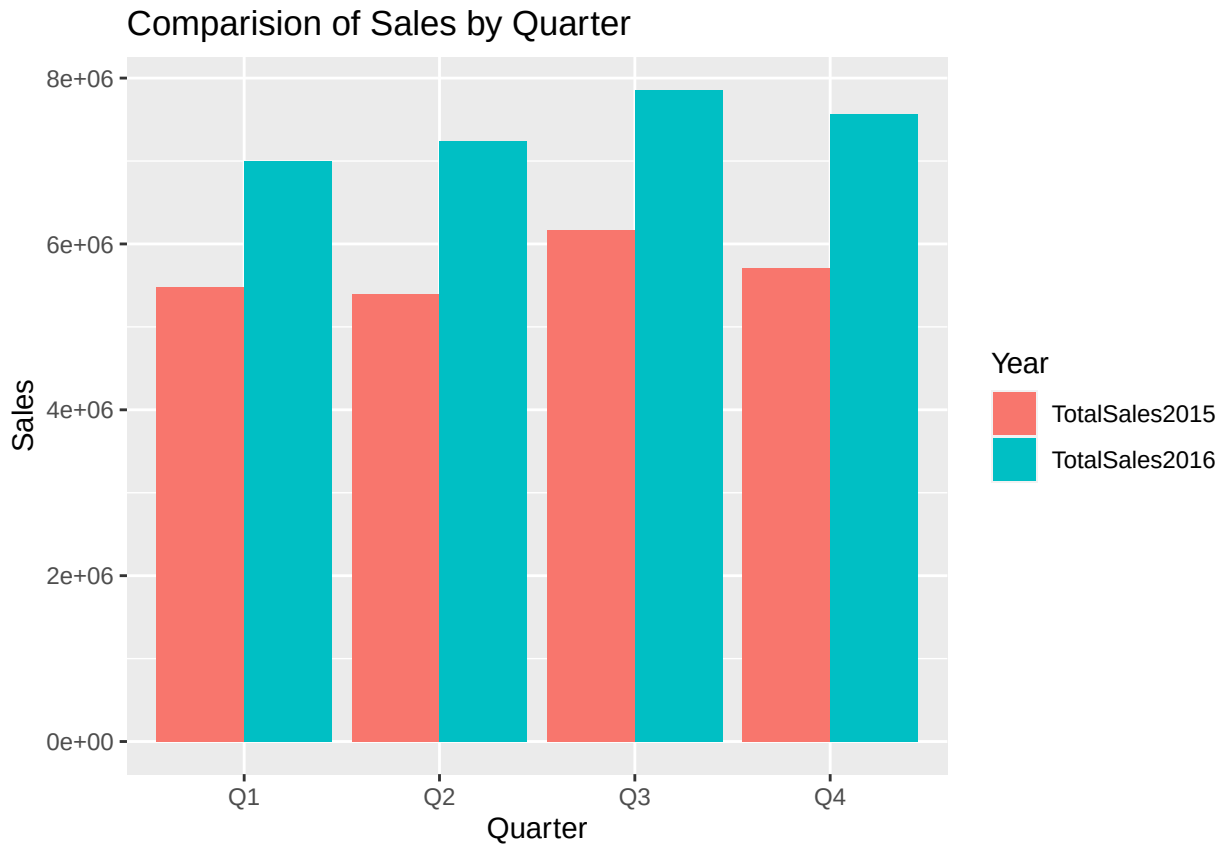

```
#question 6

SalesData$Qtr = case_when(
    SalesData$Month == "Jan" | SalesData$Month == "Feb" | SalesData$Month == "Mar" ~ "Q1",
    SalesData$Month == "Apr" | SalesData$Month == "May" | SalesData$Month == "Jun" ~ "Q2",
    SalesData$Month == "Jul" | SalesData$Month == "Aug" | SalesData$Month == "Sep" ~ "Q3",
    TRUE ~ "Q4"
  )

#question 7
qtr_data <- SalesData%>%
  group_by(Qtr)%>%
  summarise(TotalSales2015 = sum(Sales2015),TotalSales2016 = sum(Sales2016))
qtr_data
```

```
## # A tibble: 4 x 3
##   Qtr   TotalSales2015 TotalSales2016
##   <chr>          <dbl>          <dbl>
## 1 Q1          5485800.       6997953.
## 2 Q2          5390862.       7237361.
## 3 Q3          6164094.       7861546.
## 4 Q4          5713634.       7567868.
```

```
data_long4 = gather(qtr_data,key = Year, value = Sales,-Qtr)
data_long4
```

```
## # A tibble: 8 x 3
##   Qtr   Year              Sales
##   <chr> <chr>             <dbl>
## 1 Q1    TotalSales2015 5485800.
## 2 Q2    TotalSales2015 5390862.
## 3 Q3    TotalSales2015 6164094.
## 4 Q4    TotalSales2015 5713634.
## 5 Q1    TotalSales2016 6997953.
## 6 Q2    TotalSales2016 7237361.
## 7 Q3    TotalSales2016 7861546.
## 8 Q4    TotalSales2016 7567868.
```

```
ggplot(data_long4,aes(Qtr,Sales, fill = Year)) + geom_bar(stat = "identity",position = "dodge") + xlab(
  ggtitle("Comparision of Sales by Quarter")
```




```
#question 8
data_q1 = SalesData%>%group_by(Qtr,Tier)%>%filter(Qtr=="Q1")%>%summarise(TotalSales2015 = sum(Sales2015)
```

```
## `summarise()` has grouped output by 'Qtr'. You can override using the `.groups` argument.
```

```
data_q2 = SalesData%>%group_by(Qtr,Tier)%>%filter(Qtr=="Q2")%>%summarise(TotalSales2015 = sum(Sales2015)
```

```
## `summarise()` has grouped output by 'Qtr'. You can override using the `.groups` argument.
```

```r
data_q3 = SalesData%>%group_by(Qtr,Tier)%>%filter(Qtr=="Q3")%>%summarise(TotalSales2015 = sum(Sales2015)
```

```
## `summarise()` has grouped output by 'Qtr'. You can override using the `.groups` argument.
```

```r
data_q4 = SalesData%>%group_by(Qtr,Tier)%>%filter(Qtr=="Q4")%>%summarise(TotalSales2015 = sum(Sales2015)
```

```
## `summarise()` has grouped output by 'Qtr'. You can override using the `.groups` argument.
```

```r
par(mfrow=c(2,2))
par("mar")
```

```
## [1] 5.1 4.1 4.1 2.1
```

```r
par(mar=c(1,1,1,1))

value_percent1 <- data_q1$TotalSales2015/sum(data_q1$TotalSales2015)*100
pie(data_q1$TotalSales2015,labels = paste(data_q1$Tier,":",round(value_percent1,1),"%",sep = " "),main =
value_percent2 <- data_q2$TotalSales2015/sum(data_q2$TotalSales2015)*100
pie(data_q2$TotalSales2015,labels = paste(data_q2$Tier,":",round(value_percent2,1),"%",sep = " "),main =
value_percent3 <- data_q3$TotalSales2015/sum(data_q3$TotalSales2015)*100
pie(data_q3$TotalSales2015,labels = paste(data_q3$Tier,":",round(value_percent3,1),"%",sep = " "),main =
value_percent4 <- data_q4$TotalSales2015/sum(data_q4$TotalSales2015)*100
pie(data_q4$TotalSales2015,labels = paste(data_q4$Tier,":",round(value_percent4,1),"%",sep = " "),main =
```