Analysis of Sales Data from Pharm Company

Connect to Database

```
library(RMySQL)
```

```
## Loading required package: DBI
```

```
db_name_fh <- "sql3654492"
db_user_fh <- "sql3654492"
db_host_fh <- "sql3.freemysqlhosting.net"
db_pwd_fh <- "LfRIZIav7w"
db_port_fh <- 3306

mydb.fh <-  dbConnect(RMySQL::MySQL(), user = db_user_fh, password = db_pwd_fh,
                      dbname = db_name_fh, host = db_host_fh, port = db_port_fh)

dbcon <- mydb.fh
```

Question 1: Top five sales reps with the most sales broken down by year

```
sql <- "SELECT year FROM rep_facts GROUP BY year"
yearList <- dbGetQuery(dbcon, sql)

df <- data.frame(ranking = 1:5)

for (i in 1:nrow(yearList)){
  year <- yearList$year[i]

  sql <- paste("SELECT name, year, totalSales
                FROM rep_facts
                WHERE year = ", year, "
                GROUP BY name
                ORDER BY totalSales DESC
                LIMIT 5")

  result <- dbGetQuery(dbcon, sql)

  df <- cbind(df, paste(result$name, ":", as.character(result$totalSales)))
  colnames(df)[i+1] = year
}

print(df)
```

```
##   ranking                     2020                      2021
## 1       1 Aneeta Kappoorthy : 23092  Walison da Silva : 49496
## 2       2  Walison da Silva : 20424 Aneeta Kappoorthy : 41860
## 3       3    Lynette McRowe : 19412    Lynette McRowe : 30728
## 4       4      Prasad Patel : 14076      Prasad Patel : 14352
```

```
## 5       5      Helmut Schwab : 8924     Helmut Schwab : 11132
##                          2022
## 1  Walison da Silva : 26588
## 2    Lynette McRowe : 18676
## 3      Prasad Patel : 11592
## 4 Aneeta Kappoorthy : 11316
## 5      Helmut Schwab : 6900
```

Question 2: Total sold per year per region

```
sql <- "SELECT year FROM sales_facts GROUP BY year"
yearList <- dbGetQuery(dbcon, sql)
sql <- "SELECT territory FROM sales_facts GROUP BY territory"
regionList <- dbGetQuery(dbcon, sql)

df <- data.frame(region = regionList$territory)

for (i in 1:nrow(yearList)){
  year <- yearList$year[i]

  sql <- paste("SELECT territory, year, totalSales
                FROM sales_facts
                WHERE year = ", year, "
                GROUP BY territory
                ORDER BY territory")

  result <- dbGetQuery(dbcon, sql)

  df <- cbind(df, result$totalSales)
  colnames(df)[i+1] = year
}

print(df)
```

```
##          region    2020    2021    2022
## 1          East  648384 1181313  877213
## 2          EMEA 1314520 1654609 1301654
## 3 South America  764819 1805899  975368
## 4          West  814449 1406778  565130
```

Question 3:Total sold per quarter per year

```
sql <- 'SELECT CONCAT(year, " ", quarter) AS time, totalSales
        FROM sales_facts
        GROUP BY year, quarter
        ORDER BY year, FIELD(quarter, "first", "second", "third", "forth")'

df <- dbGetQuery(dbcon, sql)

x <- 1:nrow(df)

plot(x, df$totalSales, "o", xaxt='n', xlab = "quarter",
     ylab = "Total Sales(USD)", main = "Total sold per quarter per year")
```

```
axis(side = 1, at = x,labels = df$time)
```