Practicum II CS5200 (part-III)

Modi, Harshkumar

Fall 2023

```
# Load the required libraries package
library(RMySQL)
```

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

```
library(ggplot2)
library(scales)
```

```
# Connect to the MySQL database
mysql_db <- dbConnect(RMySQL::MySQL(),
                      host='salesdatabase.ccgfbvk5hpfc.us-east-1.rds.amazonaws.com',
                      port=3306,
                      username='admin',
                      password='CS5200Fall23!')

dbExecute(mysql_db, "USE salesdb;")
```

```
## [1] 0
```

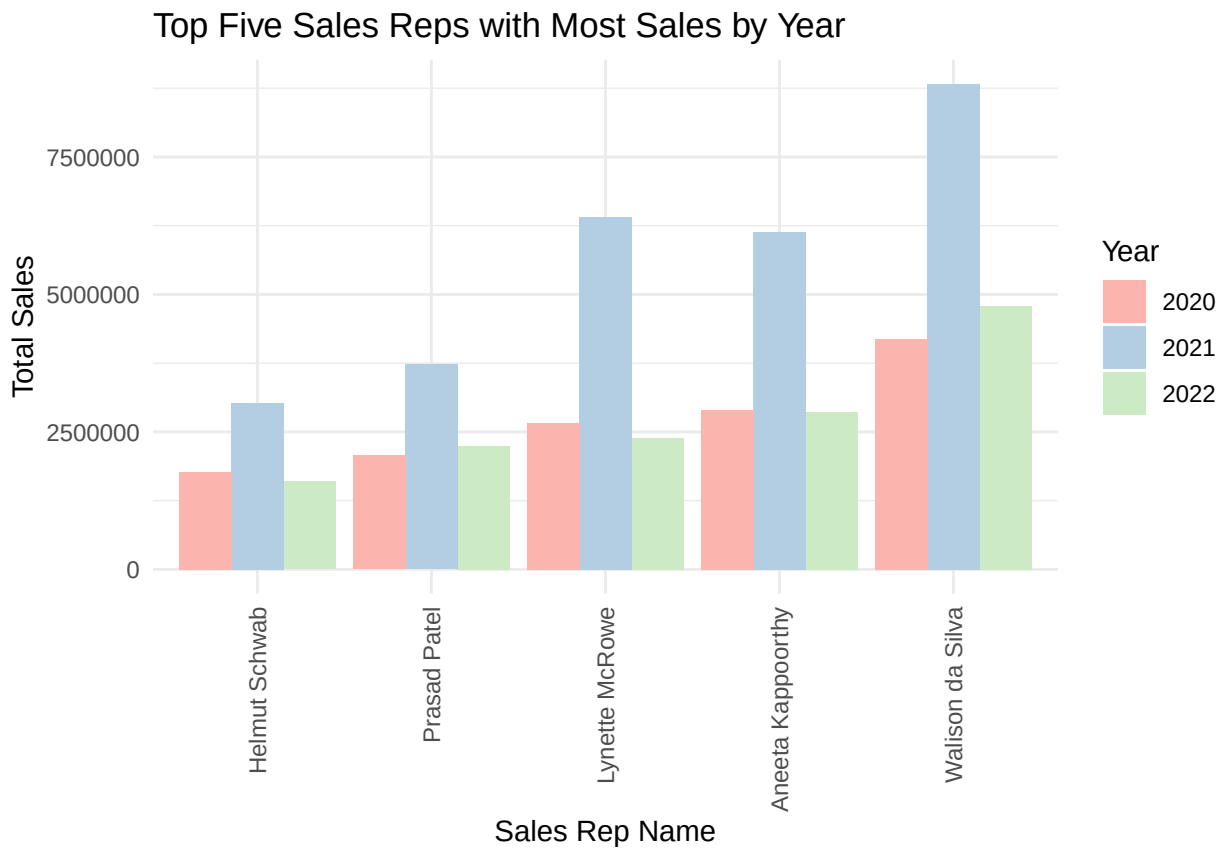
Analytical Query I: Top five sales reps with the most sales broken down by year.

```
query1 <- unique(dbGetQuery(mysql_db, "WITH RankedSales AS (
SELECT year, firstName, lastName, repID, SUM(repQuarterYearSold) as TotalSales,
RANK() OVER (PARTITION BY year ORDER BY SUM(repQuarterYearSold) DESC) as SalesRank
FROM reps_facts
GROUP BY year, repID, firstName, lastName)
SELECT year, firstName, lastName, repID, TotalSales
FROM RankedSales
WHERE SalesRank <= 5
ORDER BY year DESC, SalesRank;
"))
```

```
## Warning in .local(conn, statement, ...): Decimal MySQL column 4 imported as
## numeric
```

```
plot <- ggplot(query1, aes(x=reorder(paste(firstName, lastName, sep=' '), TotalSales), y=TotalSales, fi
  geom_bar(stat='identity', position='dodge') +
  theme_minimal() +
  scale_fill_brewer(palette="Pastel1", name="Year") +
  labs(title='Top Five Sales Reps with Most Sales by Year', x='Sales Rep Name', y='Total Sales') +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))

print(plot)
```



The plot above shows the top 5 sales representatives and their sales from each year. The trend we see here is an upward trend in terms of total sales for representatives

Analytical Query II: Total sold per year per region.

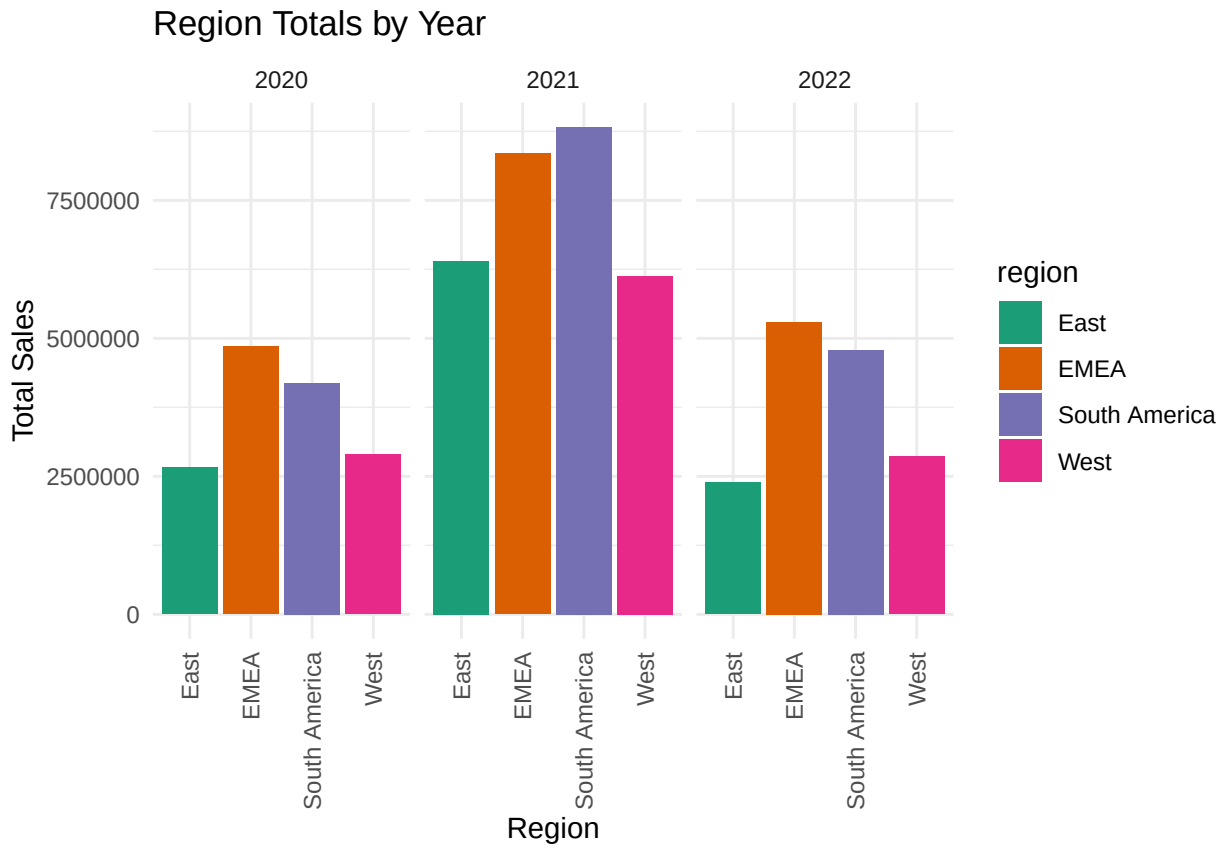
```
query2 <- unique(dbGetQuery(mysql_db, "SELECT year, region, SUM(regionTotal) AS Total
                            FROM sales_facts
                            GROUP BY year, region"))
```

```
## Warning in .local(conn, statement, ...): Decimal MySQL column 2 imported as
## numeric
```

```
plot_rotated_x <- ggplot(query2, aes(x=region, y=Total, fill=region)) +
  geom_bar(stat="identity") +
```

```
  facet_wrap(~year) + # Facets for each year
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  labs(title="Region Totals by Year", x="Region", y="Total Sales") +
  scale_fill_brewer(palette="Dark2")

print(plot_rotated_x)
```



The plot above shows the total sales region wise in each year. 2021 had the most number of sales amongst all of the data

Analytical Query III: Total sold per quarter per year.

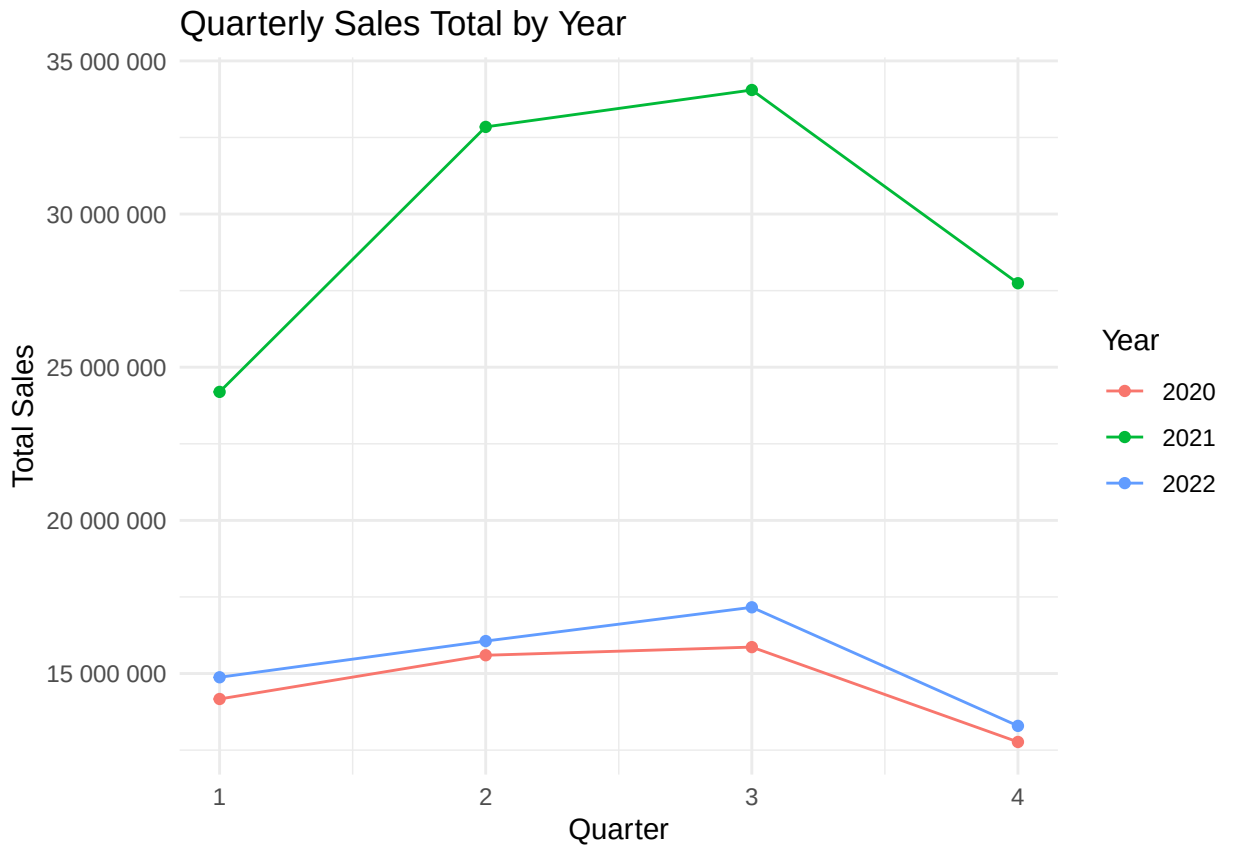
```
query3 <- unique(dbGetQuery(mysql_db, "SELECT quarter, year, SUM(quarterYearTotal) AS total
                            FROM sales_facts
                            GROUP BY quarter, year"))
```

```
## Warning in .local(conn, statement, ...): Decimal MySQL column 2 imported as
## numeric
```

```
query3$year <- factor(query3$year)

ggplot(query3, aes(x=quarter, y=total, group=year, color=year)) +
```

```
  geom_line() +
  geom_point() +
  theme_minimal() +
  labs(title="Quarterly Sales Total by Year",
       x="Quarter",
       y="Total Sales",
       color="Year") +
  scale_x_continuous(breaks=1:4) +
  scale_y_continuous(labels=label_number())
```



The plot above shows the total sales per quarter for all years. The trend of high sales in 2021 is evident once again.

```
dbDisconnect(mysql_db)
```

```
## [1] TRUE
```