

Assignment 4

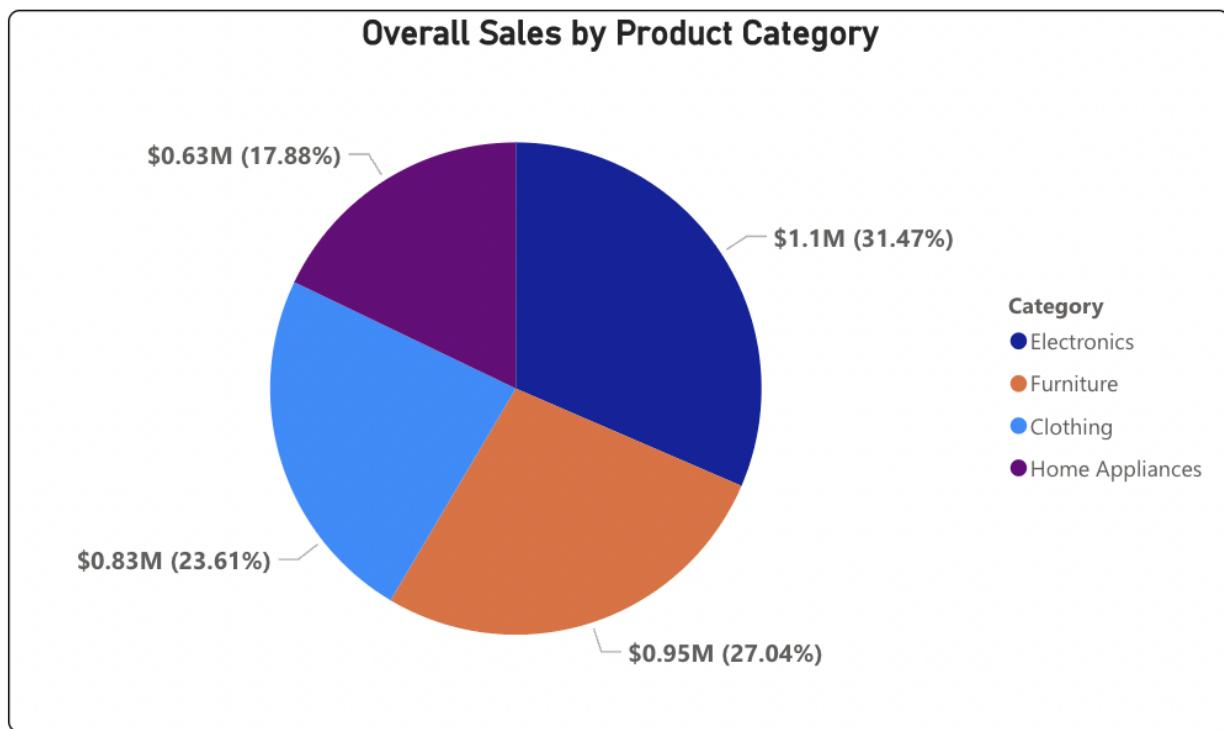
Submitted by Anitha Balachandran (016684486)

Q1. Chapter 10: Visualizing Proportions

"Visualizing Proportions" means creating a visualization that highlights the relative sizes of different parts that make up a whole.

Dataset used: Global Sales Inc. Sales Data

1. Pie Chart:



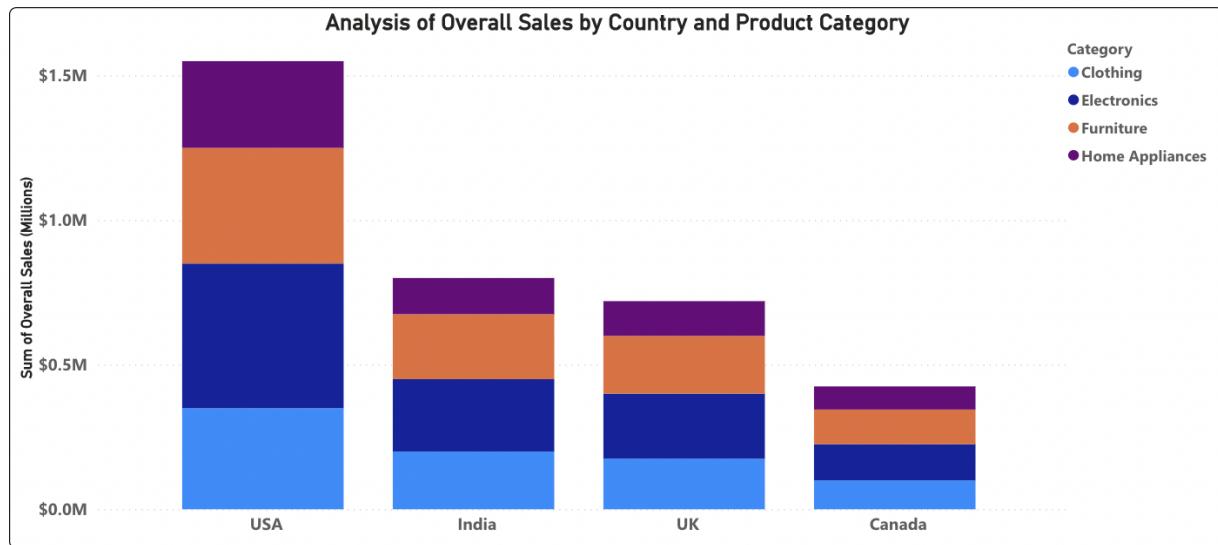
The above pie chart illustrates the distribution of overall sales across different product categories.

Observations:

- The chart indicates that Electronics is the leading product category with the highest sum of sales at \$1,100,000, followed by Furniture, Clothing, and Home Appliances.

- The chart shows that Electronics account for 31.47% of the total sum of overall sales.
- These insights provide valuable information to businesses looking to optimize their product mix and marketing strategies.

2. Stacked Bar Chart:

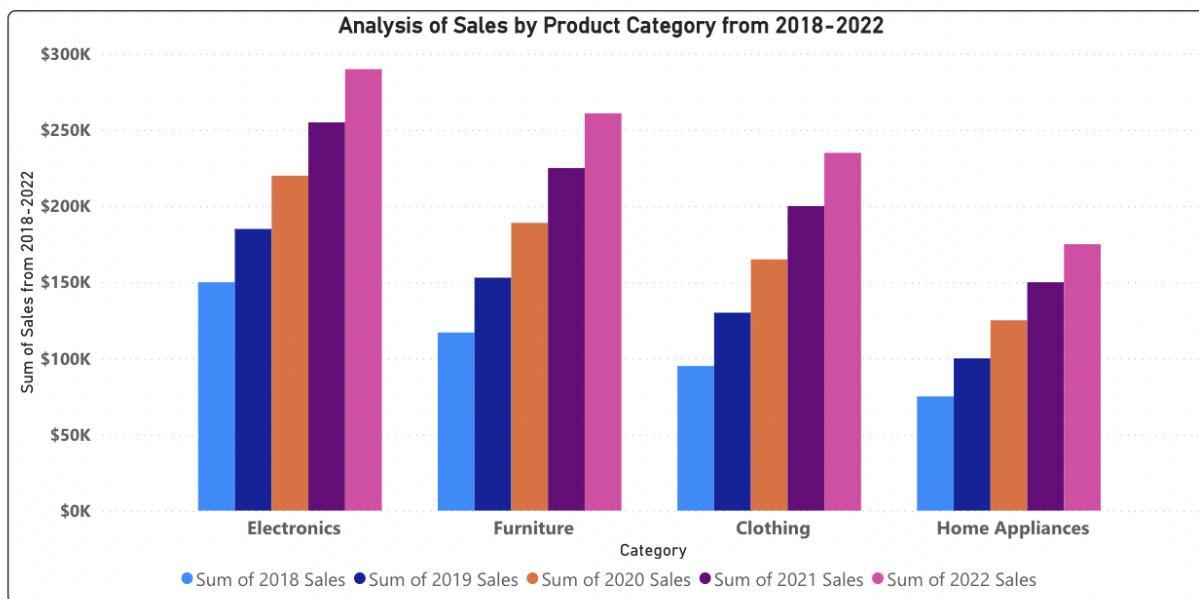


The above stacked bar chart illustrates the distribution of overall sales across different countries and product categories. The chart displays the total sum of sales for each country and the proportion of sales made up by each product category.

Observations:

- The chart shows that the USA has the highest total sum of overall sales at \$1,550,000, followed by India, UK, and Canada.
- The chart also highlights that Electronics was the top-selling product category, making up 14.31% of the total sum of overall sales across all countries, followed by Furniture, Clothing, and Home Appliances.
- The chart indicates that Canada has the least total sum of overall sales across all categories.
- These insights can help businesses in optimizing their sales strategy and targeting their markets in areas for better revenue generation.

3. Side by Side Bars:



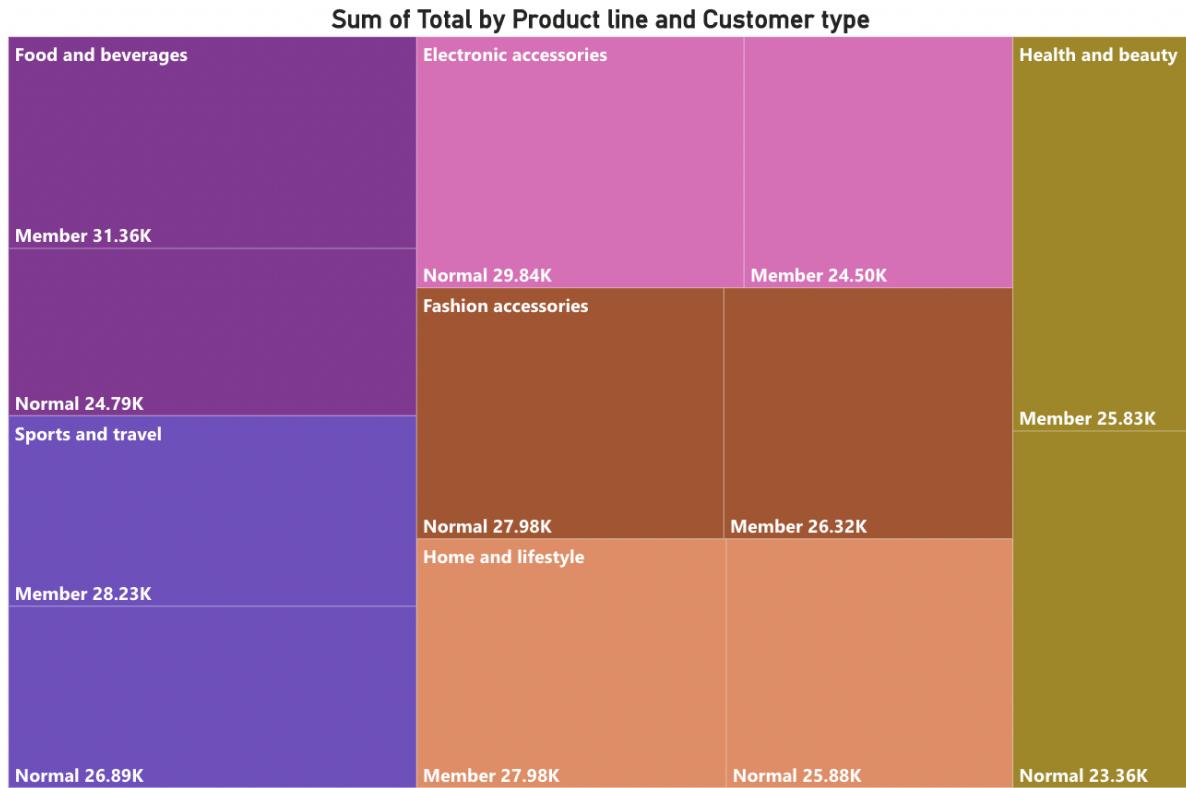
The above side by side bar chart illustrates the sales performance of different product categories from 2018-2022.

Observations:

- The chart indicates that sales have gradually increased for all categories over the five-year period and the business has experienced steady sales growth across all categories from 2018-2022.
- The chart shows that Electronics had the highest sum of sales in 2018 at \$150,000, which was 100.00% higher than Home Appliances, which had the lowest sum of sales at \$75,000.
- The chart also shows a positive correlation between the sum of sales in 2018 and the total sum of sales in 2022 indicating that the business has been able to sustain its sales growth over time.
- These insights provide valuable information to businesses looking to assess the sales performance of different product categories over time and identify areas for improvement.

Q2. Chapter 11. Visualizing Nested Proportions

1. Tree Map:



The above tree map shows the total sales by category and customer type. The size of each rectangle in the plot represents the total sales for a specific category and customer type, and the color and label of each rectangle indicate the corresponding category and customer type.

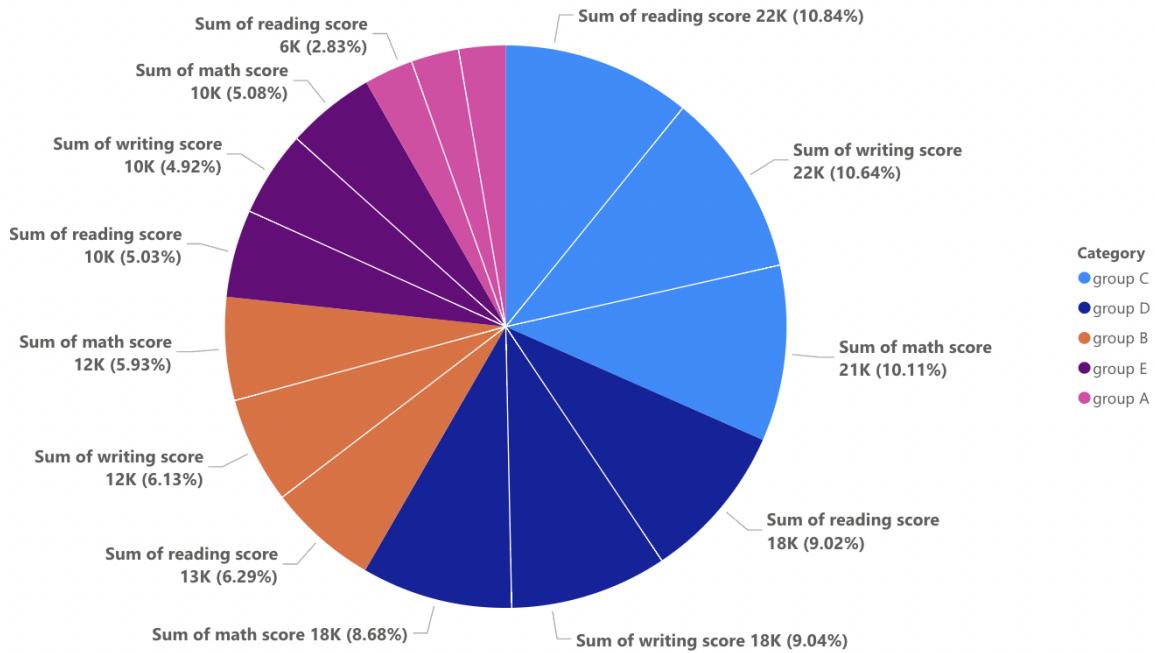
Observations:

- The biggest rectangle in the plot is indeed for the "Food and beverages" category and "Member" customer type, indicating that this group had the highest total sales of \$31.36k.
- The smallest vertical rectangles are for the "Health and Beauty" category and "Normal" customer type, indicating that this group had the lowest total sales of \$23.36k.
- For each category, the "Member" customer type generally had higher total sales than the "Normal" customer type.
- The "Fashion accessories" and "Electronic accessories" categories had similar total sales for both customer types of about \$54K.

- These insights can be helpful in identifying trends and patterns in sales data, which can inform business decisions and strategies.

2. Nested Pies:

Sum of reading score, Sum of writing score and Sum of math score by group category



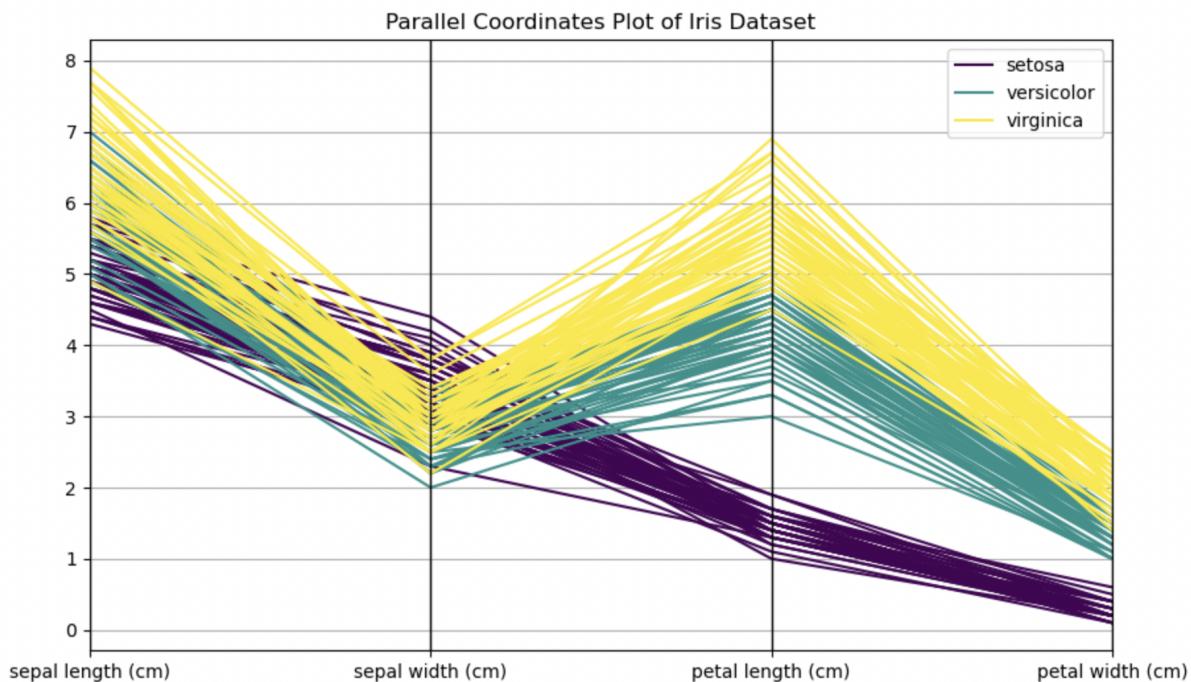
The above pie chart shows the distribution of the Sum of reading score, Sum of writing score, and Sum of math score by group category.

Observations:

- The charts reveal that group C has the highest Sum of reading score, Sum of writing score, and Sum of math score, followed by groups D, B, E, and A.
- The difference between the scores of different groups is relatively small for reading and writing, but larger for math .
- The charts suggest that there is some variation in the academic performance of students based on their group category, as measured by their Sum of reading score, Sum of writing score, and Sum of math score.
- These insights could be useful for identifying areas where educational interventions are needed to address any potential disparities in academic performance among different racial/ethnic groups.
- While it is possible to use inner and outer rings in a pie chart to display the distribution of scores by group category, this approach has a significant drawback

in that it can make it challenging to compare the sizes of the different segments accurately. Therefore, the above pie chart without rings described earlier can be more effective as it allows for more precise comparisons between the groups.

3. Parallel Set Plots:



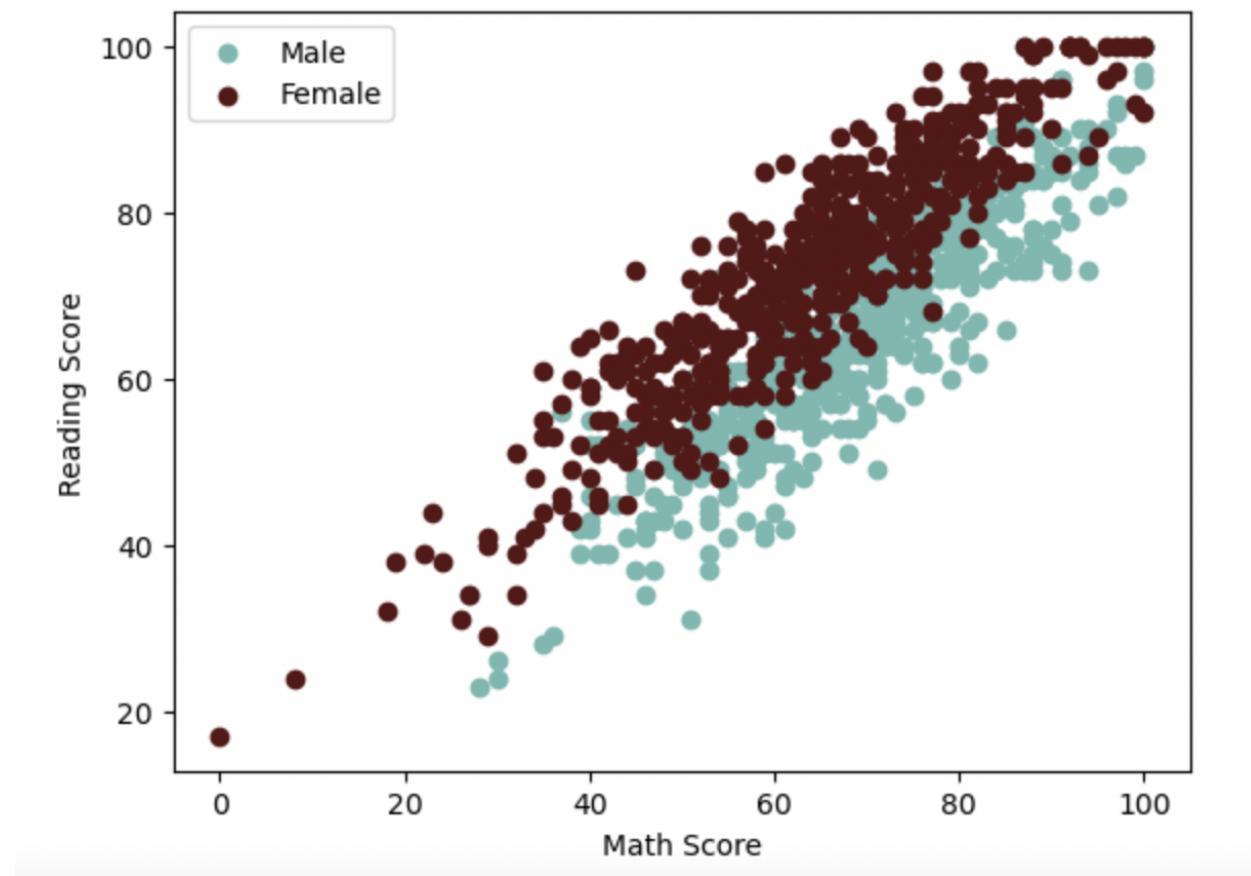
The above parallel coordinates plot shows the relationships between different features (sepal length, sepal width, petal length, and petal width) related to each other across the different iris species.

Observations:

- The setosa species is well separated from the other two species along most of the features, except for the sepal width where it overlaps with versicolor.
- The versicolor and virginica species show some overlap along the features, but can still be distinguished from each other based on their overall patterns.
- Petal length and petal width seem to be the most important features for distinguishing between the three species, as they show the greatest separation between them.
- Sepal length and sepal width show less separation between the species, and there is more overlap between them, particularly between versicolor and virginica.

Q3. Chapter 12. Visualizing Associations Among Two or More Quantitative Variables

1. Scatter Plots:

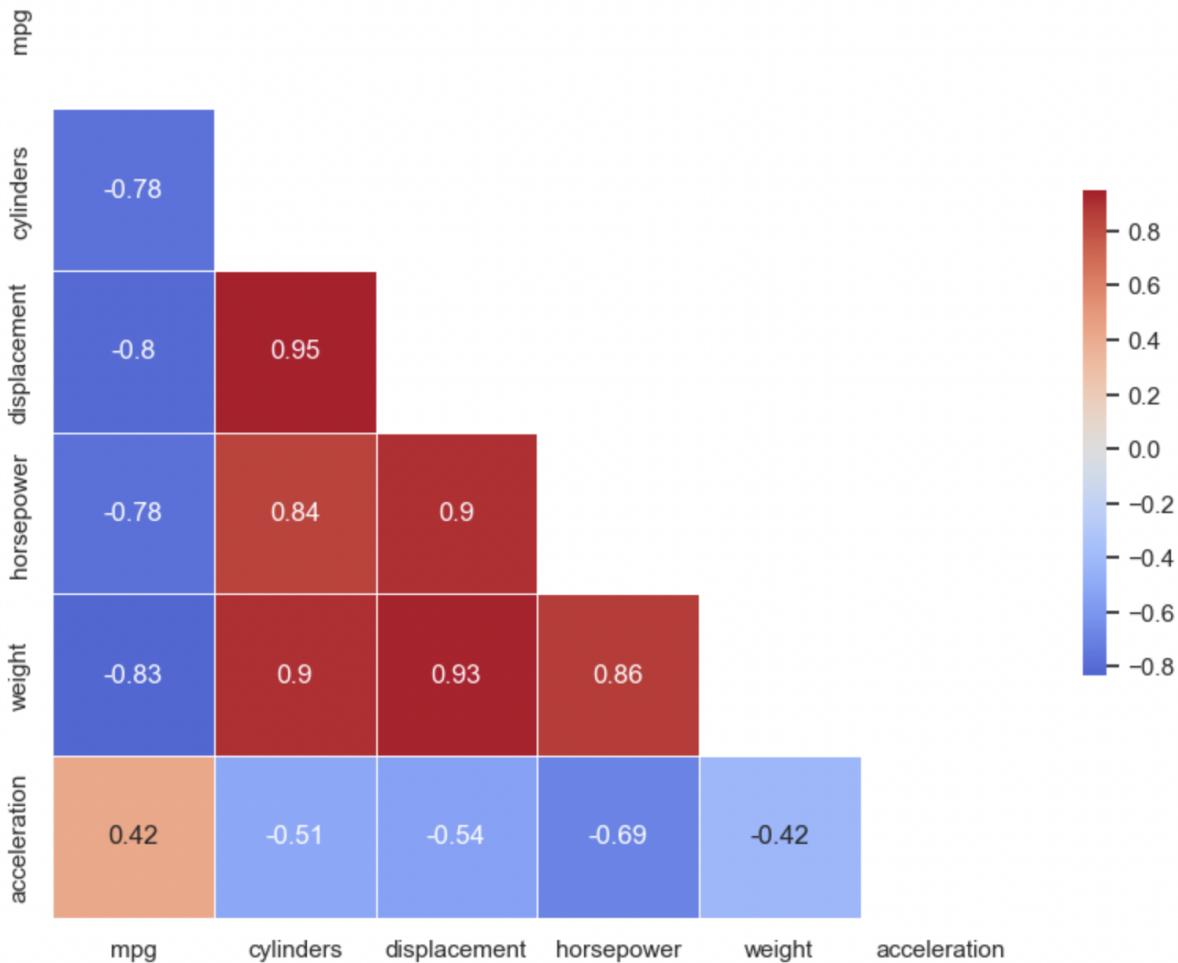


The above scatter plot shows the relationship between math score and reading score for male and female students separately.

Observations:

- There is a positive correlation between math score and reading score, indicating that students who perform well in math tend to perform well in reading as well.
- The scatter plot suggests that there are more female students with higher reading scores than male students. On the other hand, male students tend to perform slightly better in math than female students.
- There are some outliers where students perform well in one subject but not in the other, indicating that there may be other factors affecting academic performance besides gender.

2. Correlogram:

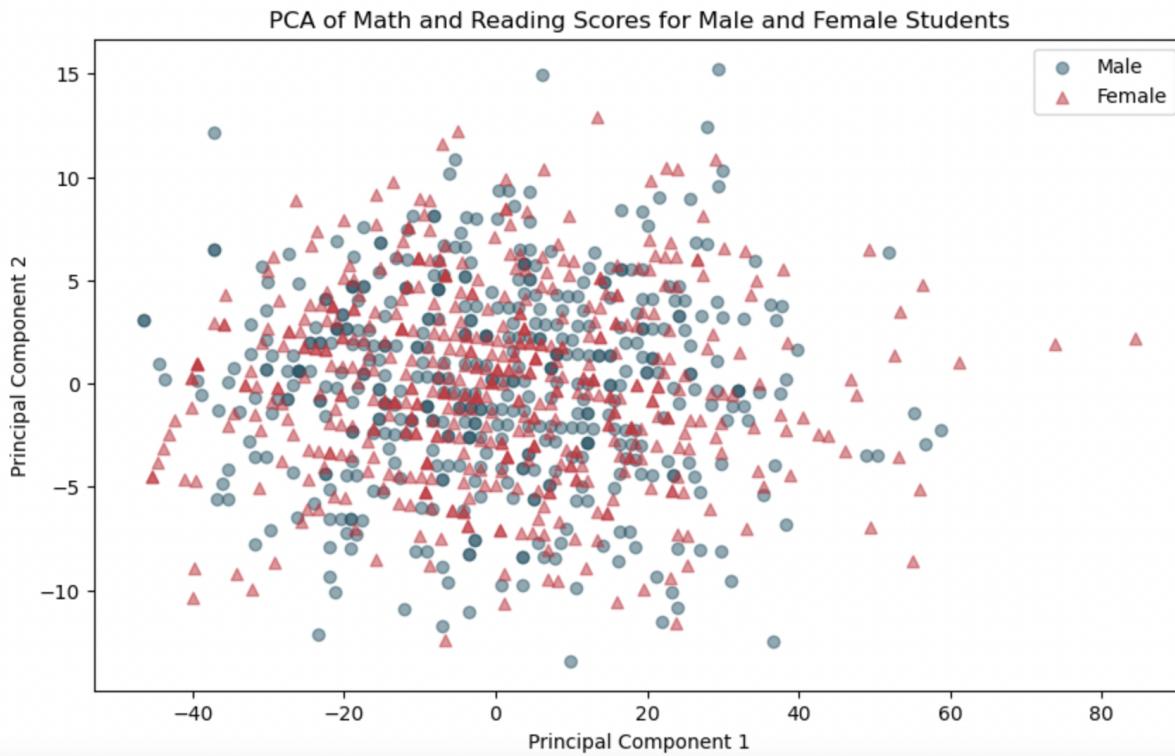


Observations:

- There is a strong negative correlation between mpg (miles per gallon) and the other variables such as cylinders, displacement, horsepower, and weight. This means that as these variables increase, mpg decreases, indicating that these variables are inversely proportional to the fuel efficiency of the car.
- The strongest positive correlation is between displacement and horsepower. This means that as engine displacement increases, the horsepower also tends to increase.
- Acceleration has a weak negative correlation with mpg, indicating that faster acceleration may lead to lower fuel efficiency, but this correlation is not very strong.

- Weight has the strongest correlation with mpg among the variables, which suggests that car manufacturers may need to focus on reducing the weight of cars in order to improve their fuel efficiency.

3. Dimension Reduction using PCA:



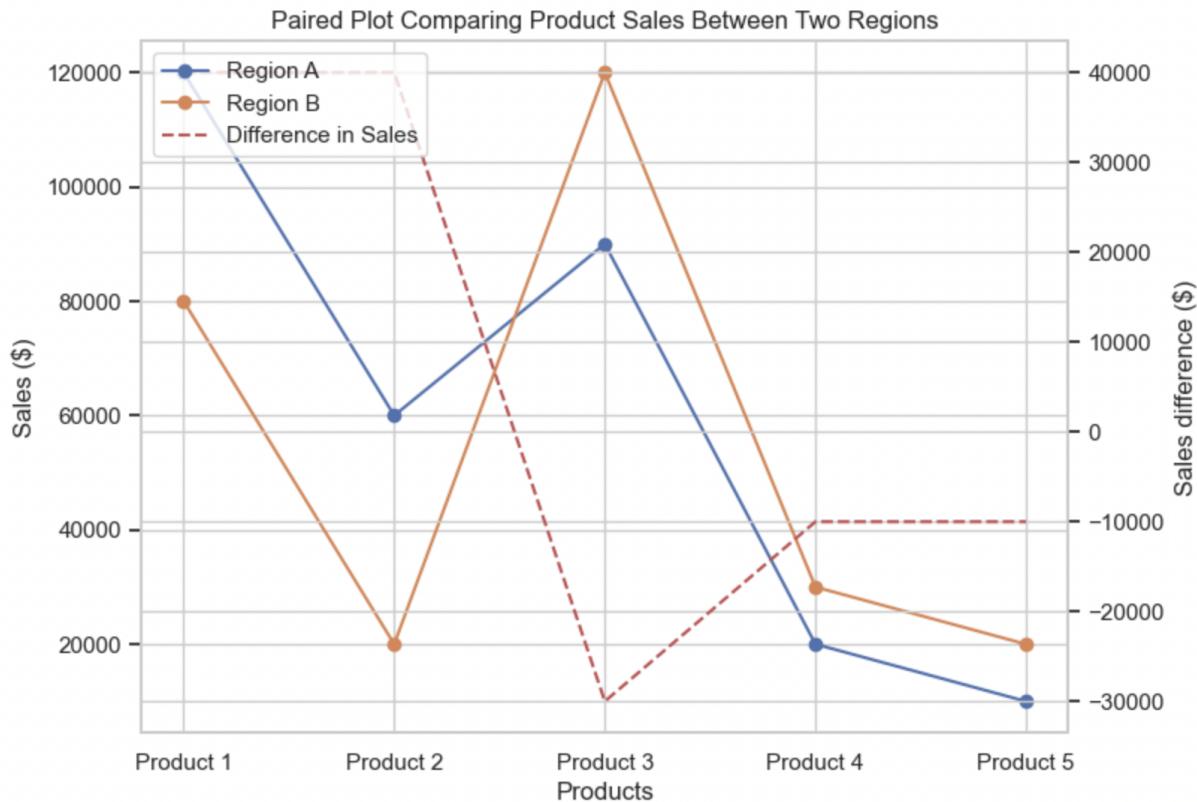
The above scatter plot shows the results of PCA analysis for math and reading scores of male and female students separately. The blue markers represent male students and the red markers represent female students.

Observations:

- The principal component 1 (PC1) explains most of the variation in both male and female data, whereas the principal component 2 (PC2) explains only a small portion of the remaining variation.
- The plot shows that there is a significant overlap between the male and female data in the PC1-PC2 space. However, there are still some notable differences between male and female students.
- The blue markers (male students) tend to have higher scores on PC1, while the red markers (female students) tend to have higher scores on PC2. This suggests that

there may be gender-related differences in the factors that influence math and reading scores.

4. Paired Plot:



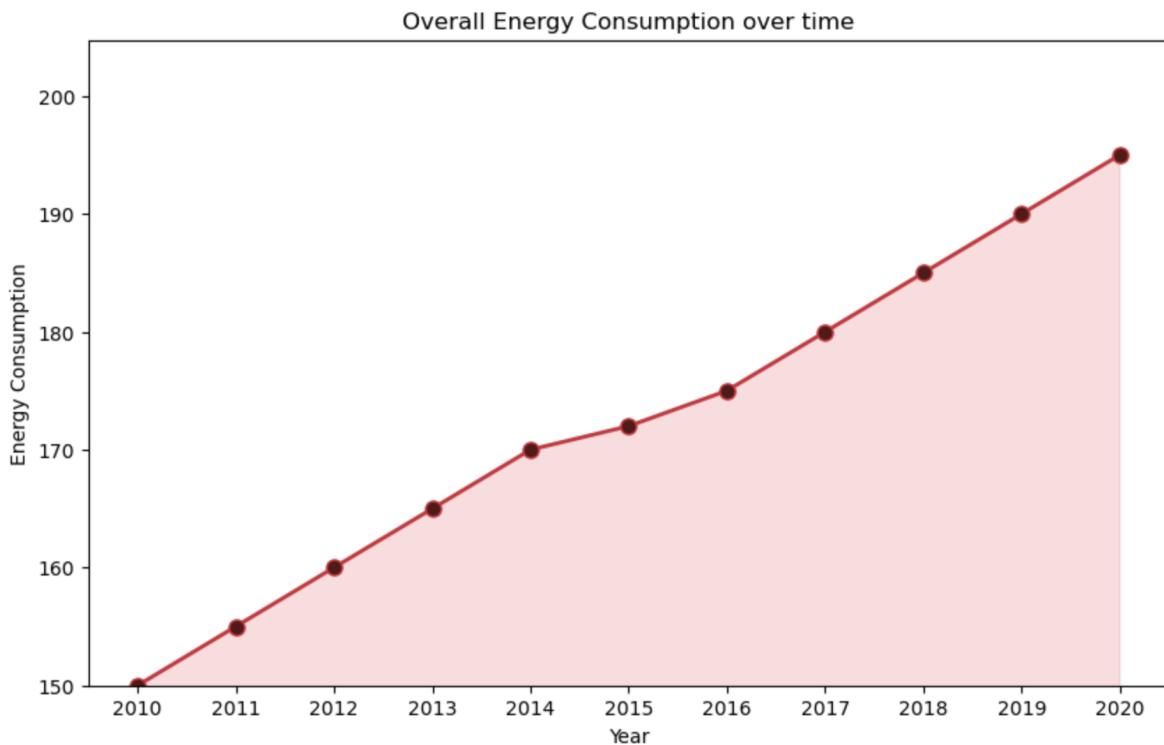
The above paired plot provides an easy way to compare the sales values of different products between the two regions and identify the differences in their rankings.

Observations:

- Region A has a higher overall sales value than Region B, with a difference of \$50,000.
- Product 1 and Product 3 are the top-selling products in both regions.
- In Region A, Product 2 has the third-highest sales value, while in Region B, it has the lowest sales value among all the products.
- Product 4 has the second-highest sales value in Region A, while in Region B, it has the fourth-highest sales value.
- Product 5 has the fourth-highest sales value in Region A, while in Region B, it has the second-highest sales value.
- The plot allows for easy identification of similarities and differences in product sales between the two regions.

Q4. Chapter 13. Visualizing Time Series and Other Functions of an Independent Variable

1. Individual Time Series:

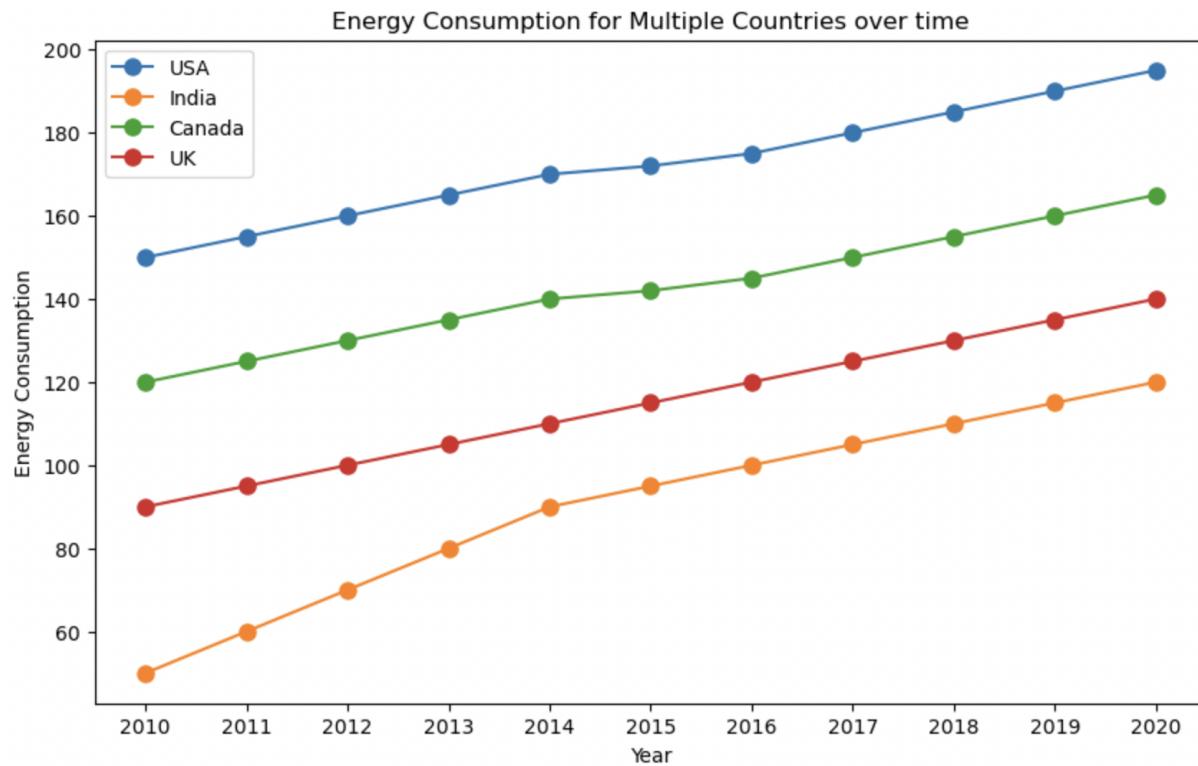


The above line plot shows the overall energy consumption from the year 2010 to 2020. The dots represent the individual data points for each year, and the line connects the dots to show the trend over time.

Observations:

- The data seems to show an upward trend in energy consumption over time, with some fluctuations around the trend.
- The energy consumption has steadily increased over the years, with a slight dip in the year 2015.
- The plot can be useful in identifying trends and patterns in energy consumption and can be used to make informed decisions for energy management.

2. Multiple Time series:



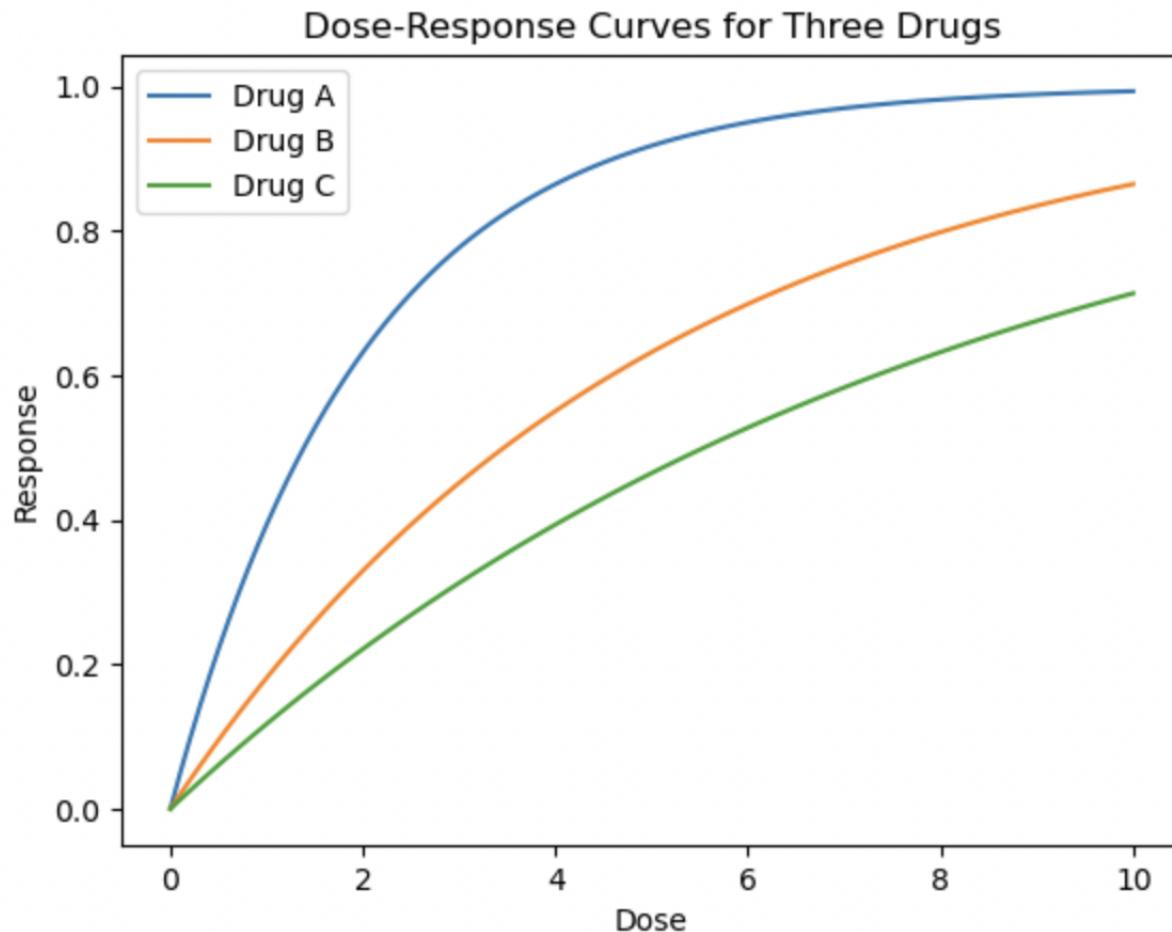
The above plot shows the overall energy consumption for four countries - USA, India, Canada, and UK from the year 2010 to 2020.

Observations:

- The energy consumption for the USA and Canada has been steadily increasing over the years.
- While for India, there has been a significant increase in energy consumption from 2014 onwards.
- The energy consumption for the UK has been relatively stable, with a slight increase in recent years.
- These insights can be useful for policymakers and energy companies to understand the changing energy consumption patterns across different countries and to develop strategies for managing energy demand and supply.

3. Dose-response curves:

Dose-response curves are commonly used in pharmacology to study the relationship between the dose of a drug and its therapeutic or toxic effects.



The above line plot shows the dose-response curves for three drugs - Drug A, Drug B, and Drug C and The graph shows the response of each drug (i.e., the number of patients cured) against increasing doses of the drug.

Observations:

- As the dose of each drug increases, the response (i.e., the number of patients cured) also increases.
- Drug A has the most potent response curve, with a steep slope that levels off quickly. This indicates that small increases in the dose of Drug A can have a significant impact on the number of patients cured.
- In contrast, Drug C has the weakest response curve, with a gentle slope that gradually levels off. This indicates that higher doses of Drug C may be required to achieve the same therapeutic effect as lower doses of Drug A or B.

=====XXXXXX=====

Note: Source code for few plots given above have been attached in .ipynb file