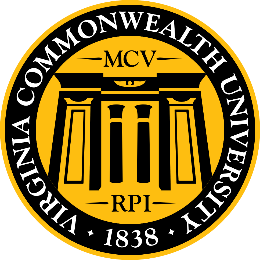
****

**VIRGINIA COMMONWEALTH UNIVERSITY**

**Statistical analysis and modelling (SCMA 632)**

# A5: Visualization - Perceptual Mapping for Business

**KATHRYN SHAJU**

**V01111327**

**Date of Submission: 15-07-2024**

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Title** | **Page No.** |
| **1.** | Introduction | **1** |
| **2.** | Objectives | **1** |
| **3.** | Business Significance | **2** |
| **4.** | Results & Interpretations | **3-5** |

**INTRODUCTION**

This study focuses on showcasing the data in a visual format. Here we have the taken the NSSO68 dataset and we have mainly identified and focused on displaying the total consumption of the state of Punjab in its different districts. For the visual perception of the data we have made use of tools such as histogram, bar plot and shapelift to plot on the maps.

A histogram is used to visualize the distribution of a continuous variable. It divides the data into bins and shows the frequency of data points in each bin.A continuous variable's distribution can be seen using a histogram. The data is separated into bins, and the frequency of data points within each bin is displayed.

A bar plot is a kind of visual aid that uses rectangular bars to represent categorical data, with each bar's length proportionate to the value it represents. Comparing the frequency, count, or other metrics (like sums or averages) of various categories is frequently done using bar graphs.

Plotting data on a map is a powerful visualization technique used to represent spatial data, allowing for geographic patterns and distributions to be easily identified and analysed. In our study we have made use of GeoJOSN file of the districts of Punjab which allow for geographical spatial data representation.

Mainly we have first plotted a Histogram to understand the distribution of the consumption across the districts, then we have plotted a bar plot to understand the frequency of consumption in each district and lastly we have plotted the data on the map of the state of Punjab with the districts to better visualize and understand the total consumption of the state as whole.

**OBJECTIVES**

1. To plot a histogram (to show the distribution of total consumption across different districts) and a barplot (To visualize consumption per district with district names) of the data in NSSO68 to indicate the consumption district-wise for the state of Punjab.
2. To plot total consumption on the Punjab state map using NSSO68.csv data

**BUSINESS SIGNIFICANCE**

The distribution of consumption levels among districts in Punjab can be seen visually by plotting a district-by-district histogram of total consumption. Businesses, legislators, and other stakeholders can benefit greatly from the business insights that this strategy can provide. Businesses can view the total distribution of consumption with the use of histograms. This facilitates the identification of patterns, such as the majority of districts' high, moderate, or low consumption. Districts with abnormally high or low consumption are known as outliers, and they are simple to find. These anomalies can need extra care or focused tactics. Histograms that represent historical consumption data can be used to forecast future trends in consumption. This helps with inventory planning and demand forecasts.

Making a bar plot of Punjab's district-by-district total consumption can aid in strategic decision-making and offer important commercial insights. Bar charts offer a simple means of comparing the overall consumption in various districts. It is easy to see which districts consume the most and which the least from this graphic comparison. Companies can pinpoint the major markets with the highest levels of consumption, enabling them to concentrate their marketing and sales endeavors on these regions.  
Businesses can improve the efficacy of their promotional efforts by creating district-specific targeted marketing campaigns based on consumption data.

Plotting Punjab's total consumption by district on a map can give companies, governmental organizations, and other stakeholders important information. Businesses can determine which districts have the highest and lowest levels of consumption by visualizing consumption statistics. Allocating resources and marketing campaigns can be guided by this information. Supply chains can be made more efficient by ensuring that products are delivered effectively through an understanding of consumption trends. Low-consumption areas might need fewer supplies, whereas high-consumption areas might need more regular replenishing.

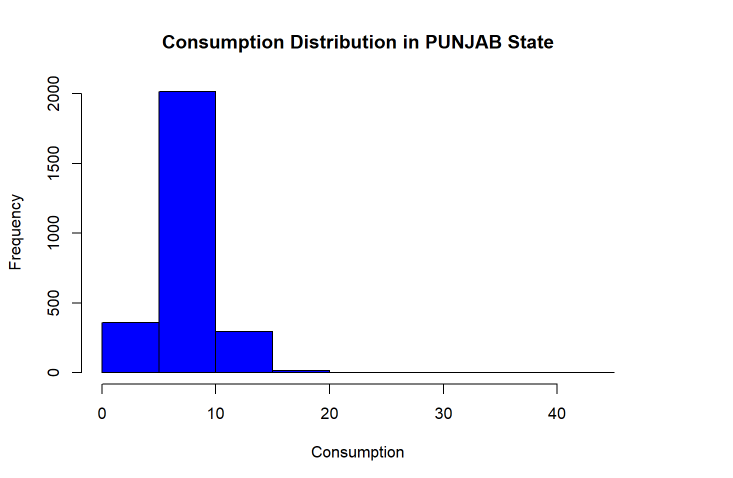
**RESULTS AND INTERPRETATIONS**

1. To plot a histogram (to show the distribution of total consumption across different districts) and a barplot (To visualize consumption per district with district names) of the data in NSSO68 to indicate the consumption district-wise for the state of Punjab.

Firstly we have set the directory and loaded the dataset ‘NSSO68.csv’. From here we have firstly filtered the data for the state of Punjab, once we have the data for Punjab, we have then moved on to cleaning the data, by checking for missing values and then we have subseted the data further to get the total consumption of the state of Punjab. Similarly we have also checked for outliers, and when found we have imputed with the mean of the variable that had an outlier.

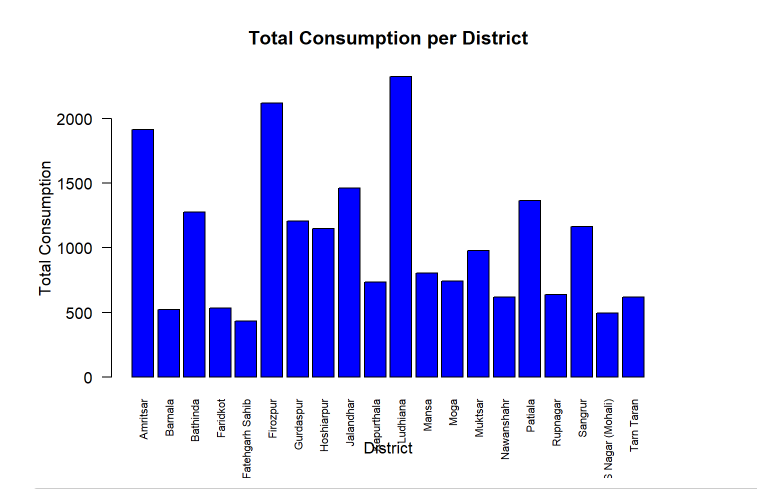
Now that we have filtered and cleaned the data, we have further mapped the data district wise and sector wise, also we have renamed the district codes with its respective district names, allowing us to identify the districts easily.

Following this we have plotted the histogram for the consumption distribution in state of Punjab.



From the graph above we observe that the frequency of consumption is high and the distribution is not very even, meaning its concentrated in some states more than others.

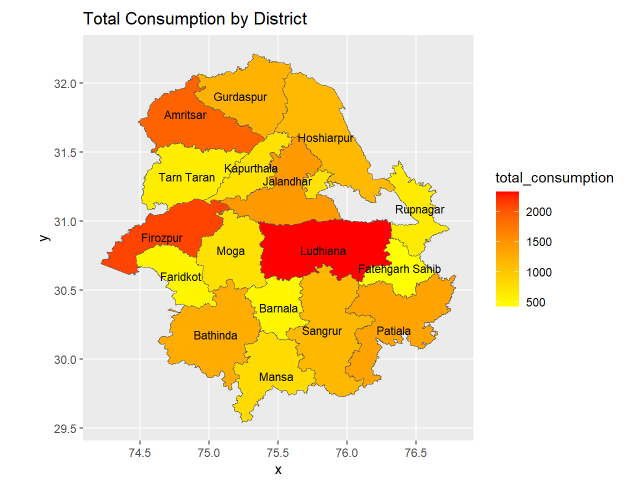
To better understand which districts in Punjab exactly have the most consumption we have created a bar plot of the same data. The results are as follows:



Here we can easily identify the districts with the most consumption, which as we can observe from above is the districts of Ludhiana, Firozpur and Amritsar. The bar plot allows us to visualize the frequency of consumption for individual districts thereby enhancing the accuracy of our decision making.

1. To plot total consumption on the Punjab state map using NSSO68.csv data

To plot on the map of Punjab , we have first loaded the shapelift GeoJOSN, named PUNJAB\_DISTRICTS. geojosn. This allows us to get the map of Punjab, with its districts. There on we use ggplot2 to code and plot the data on the map for total consumption. Here are the results:



Over here the darker orange represents the higher consumption districts and the lighter yellow showcases the lowest consumption districts. We observe that the darkest is the district of Ludhiana, followed by Firozpur and Amritsar, and the rest of the states have lighter and lighter shades comparative, this allows us to conclude these three districts to be the districts with higher consumption.

In this way we can study the consumption pattern in the state of Punjab through visual presentation tools, like Histogram, bar plots and by Plotting data on the map.