VIRGINIA COMMONWEALTH UNIVERSITY

Statistical analysis and modelling (SCMA 632)

A5: VISUALISATION – PERCEPTUAL MAPPING FOR BUSINESS

RIDDHI RUNGTA

V01107488

Date of Submission: 14-07-2024

CONTENTS

VISUALIZATION - PERCEPTUAL MAPPING FOR BUSINESS USING PYTHON

INTRODUCTION

The National Sample Survey Office (NSSO) conducts numerous surveys to gather socio-economic data essential for policy formulation, planning, and decision-making. The 68th round of NSSO data, contained in the dataset "NSSO68.csv," offers an in-depth view of consumption patterns across various districts in India. This dataset is crucial for understanding regional disparities in consumption, reflecting broader socio-economic conditions such as income levels, poverty rates, and lifestyle variations.

The dataset includes a wide range of consumption metrics, detailing expenditure on both essential and non-essential goods and services. By analyzing this data, researchers, policymakers, and businesses can gain valuable insights into consumer behavior and preferences at the district level. This detailed data is vital for identifying target markets, evaluating the effectiveness of social welfare programs, and creating localized strategies for economic development.

OBJECTIVES

- Draw a histogram of the data of the state assigned to indicate the consumption district-wise.
- Depict the consumption on the state map, showing consumption in each district for the state of Karnataka.

BUSINESS SIGNIFICANCE

Understanding consumption patterns is vital for businesses looking to enter or expand in the diverse Indian market. The NSSO68 dataset offers valuable insights into district-wise consumption, enabling companies to adapt their products and marketing strategies to the unique needs and preferences of various regions.

- **Market Penetration and Expansion:** Businesses can leverage the NSSO68 dataset to identify districts with high demand, optimizing their distribution channels for more effective market penetration and expansion strategies.
- **Product and Marketing Customization:** Companies can tailor their products and marketing efforts to meet the specific needs and preferences of different regions, boosting customer satisfaction and market share.
- **Policy Design and Resource Allocation:** Policymakers can use the data to address regional disparities, design targeted interventions, and improve resource allocation to stimulate economic activity in underserved areas.
- **Impact Monitoring and Evaluation:** Government agencies can utilize the dataset to monitor the effectiveness of existing policies and programs, enabling data-driven decisions to promote inclusive growth and sustainable development.
- **Investment and Business Planning:** Investors and entrepreneurs can analyze consumption data to pinpoint potential investment opportunities and create business plans that align with regional consumption trends and demands.

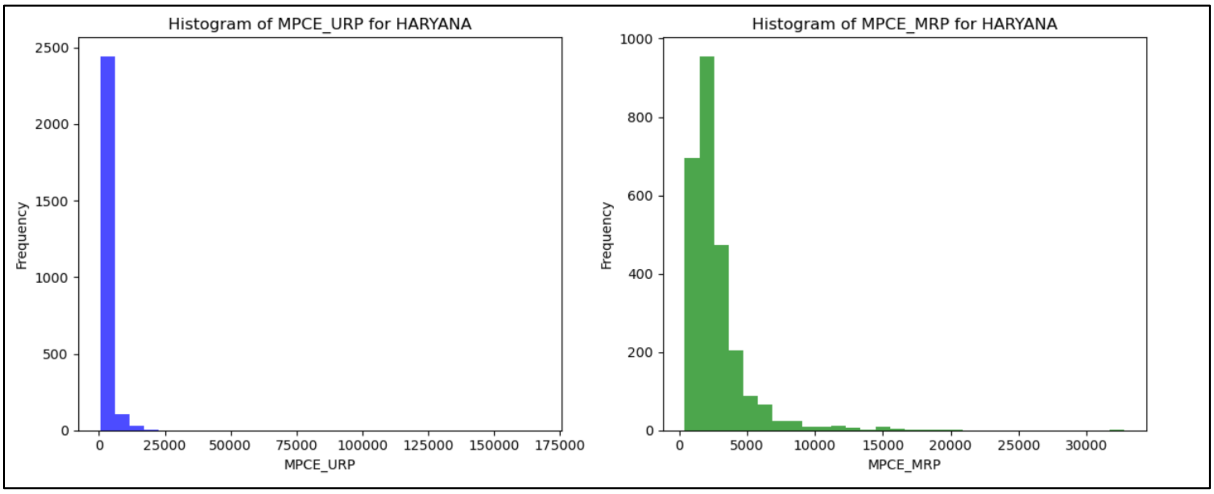
RESULTS AND INTERPRETATION

Objective 1: Draw a histogram of the data of the state assigned to indicate the consumption district-wise. [NSSO68.csv]

Code:

```
import pandas as pd
import matplotlib.pyplot as plt

# Load the data
file_path = 'NSSO68.csv'
data = pd.read_csv(file_path,low_memory=False)

# Inspect the first few rows of the dataset to understand its structure
print(data.head())

# Subset the data with specified columns and filter based on state_1
subset_data = data[['state', 'District', 'MPCE_URP', 'MPCE_MRP', 'state_1']]
filtered_data = subset_data[subset_data['state_1'] == 'HR']

import matplotlib.pyplot as plt

# Plot histograms for MPCE_URP and MPCE_MRP
plt.figure(figsize=(12, 5))

plt.subplot(1, 2, 1)
plt.hist(filtered_data['MPCE_URP'], bins=30, color='blue', alpha=0.7)
plt.title('Histogram of MPCE_URP for Haryana')
plt.xlabel('MPCE_URP')
plt.ylabel('Frequency')

plt.subplot(1, 2, 2)
plt.hist(filtered_data['MPCE_MRP'], bins=30, color='green', alpha=0.7)
plt.title('Histogram of MPCE_MRP for Haryana')
plt.xlabel('MPCE_MRP')
plt.ylabel('Frequency')

plt.tight_layout()
plt.show()
```

Result:



Interpretation:

The histograms generated by the code represent the distribution of Monthly Per Capita Expenditure (MPCE) for the state of Haryana, divided into two categories: MPCE based on Uniform Recall Period (URP) and MPCE based on Modified Recall Period (MRP).

Interpretation of the Histograms:

1. Histogram of MPCE_URP for Haryana:

- **X-axis**: Represents the Monthly Per Capita Expenditure (URP).
- **Y-axis**: Represents the frequency or count of observations within each expenditure range.
- **Distribution**: The histogram shows a highly skewed distribution with most values concentrated at the lower end of the expenditure spectrum.
- **Observation**:
  - Most households in Haryana have a low MPCE (URP), with expenditure values clustering below 25,000 INR.
  - There are very few households with higher MPCE (URP), indicating a large proportion of low-income households.

2. Histogram of MPCE_MRP for Haryana:

- **X-axis**: Represents the Monthly Per Capita Expenditure (MRP).
- **Y-axis**: Represents the frequency or count of observations within each expenditure range.
- **Distribution**: Like the MPCE_URP, the MPCE_MRP distribution is also right-skewed but appears to have a slightly broader spread.
- **Observation**:
  - Most households have MPCE (MRP) values below 5,000 INR.
  - A small number of households have higher MPCE (MRP), but the frequency decreases rapidly as the expenditure amount increases.

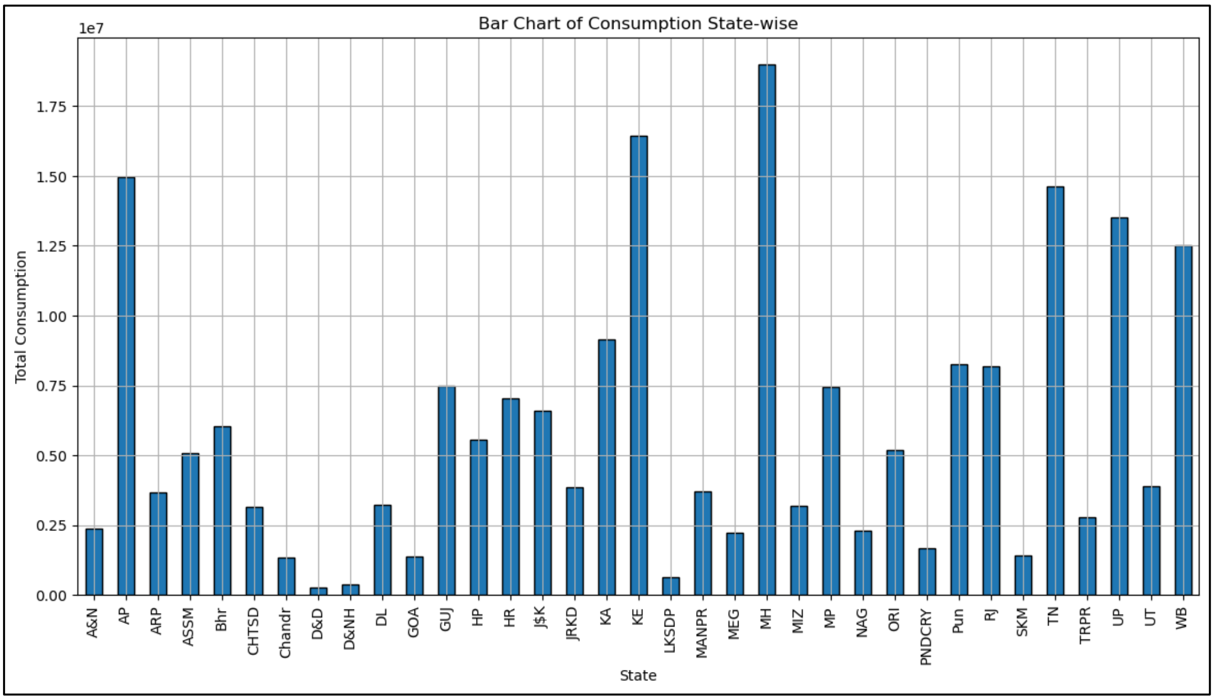
1. **Income Disparity**: The right-skewed nature of both histograms indicates significant income disparity within Haryana, with most households having low per capita expenditure and a few households with much higher expenditure.
2. **Consumption Patterns**:
    - **MPCE_URP**: Focuses on regular and frequently purchased items over a uniform recall period, likely showing basic and essential expenditures.
    - **MPCE_MRP**: Includes items purchased over varying recall periods, possibly capturing more occasional and less frequent purchases.
3. **Policy Implications**:
    - **Targeted Interventions**: The data suggests the need for targeted socio-economic interventions to uplift the lower-income groups, ensuring equitable growth and development.
    - **Resource Allocation**: Insights from the data can help in better allocation of resources and planning social welfare programs aimed at poverty alleviation and improving living standards.

The histograms of MPCE_URP and MPCE_MRP for Haryana provide valuable insights into the expenditure patterns of households in the state. The significant skewness towards lower expenditure values highlights the economic challenges faced by most of the population, emphasizing the need for targeted policy measures to address income inequality and promote inclusive growth.

Code:

```
# Assuming the relevant columns are named 'state_1' for state names and 'MPCE_URP' for consumption
state_column = 'state_1'  # Replace with the actual column name for states
consumption_column = 'MPCE_URP'  # Replace with the actual column name for consumption

# Group the data by state and sum the consumption
state_consumption = data.groupby(state_column)[consumption_column].sum()

# Plot the bar chart
plt.figure(figsize=(14, 7))
state_consumption.plot(kind='bar', edgecolor='black')
plt.title('Bar Chart of Consumption State-wise')
plt.xlabel('State')
plt.ylabel('Total Consumption')
plt.xticks(rotation=90)
plt.grid(True)
plt.show()
```

Result:



Interpretation:

Andhra Pradesh (AP), Maharashtra (MH), and Tamil Nadu (TN) exhibit the highest total consumption, highlighting significant market opportunities due to their elevated consumption levels. States like Gujarat (GUJ), Karnataka (KA), and West Bengal (WB) also demonstrate substantial consumption, indicating strong markets but slightly lower than the leading states. Conversely, states like Nagaland (NAG), Sikkim (SKM), and Goa (GOA) have the lowest total consumption, suggesting smaller markets or reduced demand.

- **Market Expansion:** Businesses should focus on expanding their operations and marketing initiatives in states with high consumption levels such as Andhra Pradesh, Maharashtra, and Tamil Nadu, as these states are likely to offer the highest returns on investment.
- **Resource Allocation:** Strategically allocate resources, including marketing budgets, sales teams, and distribution networks, to high-consumption states to maximize market penetration and sales.

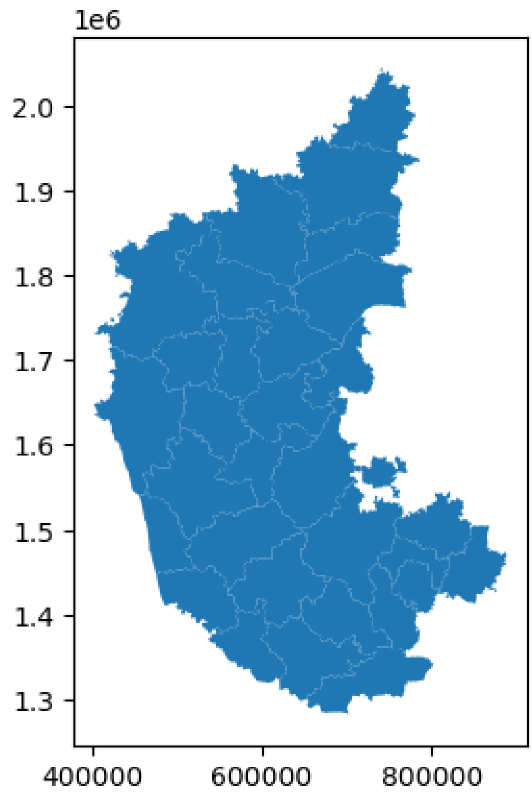
Objective 2: Depict the consumption on the state map, showing consumption in each district for the state of Karnataka. [NSSO68.csv]

Code:

```
#Import GeoPandas library
import geopandas as gpd

# Load the shapefile
gdf_districts = gpd.read_file('District.shp')

gdf_districts.head()

#plot the map
gdf_districts.plot()
```

Result:

	KGISDistri	LGD_Distri	KGISDist_1	BhuCodeDis	created_us	created_da	last_edite	last_edi_1	SHAPE_STAr	SHAPE_STLe	geometry
0	01	527	Belagavi	01	None	0000/00/00	SURESHBV1	2022-11-24	1.339772e+10	1.141488e+06	MULTIPOLYGON (((537523.31 1865366.861, 537555....
1	02	524	Bagalkot	02	None	0000/00/00	SURESHBV1	2022-09-08	6.561826e+09	6.682456e+05	POLYGON ((581917.898 1811433.959, 581946.875 1...
2	03	530	Vijayapura	03	None	0000/00/00	SURESHBV1	2022-11-24	1.050271e+10	7.032618e+05	POLYGON ((537523.31 1865366.861, 537516.168 18...
3	04	538	Kalburgi	04	None	0000/00/00	SURESHBV1	2022-11-09	1.097395e+10	9.181459e+05	MULTIPOLYGON (((680992.661 1951255.947, 681227...
4	05	529	Bidar	05	None	0000/00/00	SURESHBV1	2022-11-16	5.454415e+09	5.733925e+05	MULTIPOLYGON (((766506.612 1970249.909, 766481...



Interpretation:

Shapefiles are commonly used for mapping and spatial analysis due to their compatibility with various GIS software. They consist of at least three mandatory files: a main file (.shp) containing geometry data, an index file (.shx) for indexing, and a dBASE table (.dbf) that stores attribute data in tabular form. Additional optional files can also be included to provide more information.

The displayed shapefile data includes several key attributes for different districts: a district identifier (KGISDistri and KGISDist_1), an administrative code (LGD_Distri), a unique identifier (BhuCodeDis), and metadata about creation and editing dates and users. Additionally, it contains geometrical properties like area (SHAPE_STAr) and perimeter (SHAPE_STLe), and the geometric shapes (geometry) representing the district boundaries.

The plotted map provides a visual overview of the district boundaries within the state. This can be useful for spatial analysis, resource allocation, and regional planning.

The DataFrame provides detailed information about each district, including its name, identifiers, and geometric properties. This data is essential for any analysis involving geographic information systems (GIS).

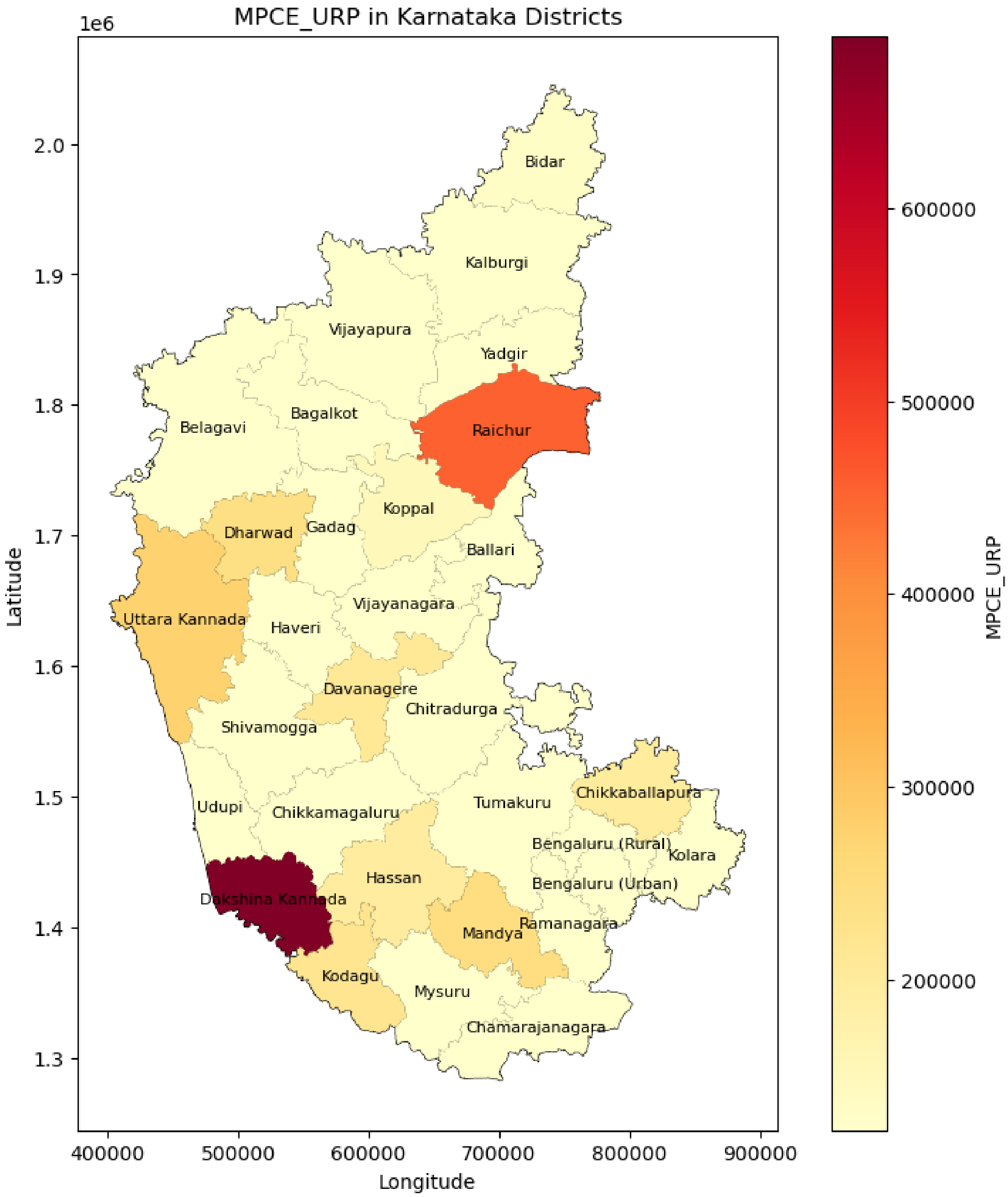
Code:

```python
import pandas as pd

# Load the NSSO68.csv file
nsso_data = pd.read_csv('NSSO68.csv',low_memory=False)

# Subset the data with specified columns and filter based on state_1
subset_data = nsso_data[['state', 'District', 'MPCE_URP', 'MPCE_MRP', 'state_1']]
filtered_data = subset_data[subset_data['state_1'] == 'KA']

# Load the district-codes.xlsx file
district_codes = pd.read_excel('district-codes.xlsx')

# Filter district codes for Karnataka
karnataka_districts = district_codes[district_codes['state name'] == 'Karnataka']

# Create a mapping from district codes to district names
district_mapping = dict(zip(karnataka_districts['dc'], karnataka_districts['district name']))

# Replace district codes in the filtered NSSO data with district names
filtered_data['District'] = filtered_data['District'].map(district_mapping)

# Sum all values district-wise
district_wise_sum = filtered_data.groupby('District').sum().reset_index()

# Exclude 'state_1' from the sum result
district_wise_sum = district_wise_sum[['District', 'MPCE_URP', 'MPCE_MRP']]

# Display the resulting DataFrame
district_wise_sum
```

```
df = district_wise_sum

# Merge shapefile with varibale related to Districts
#KGISDist_1 is our Distrcit Name in gdf_districts dataframe
#District is our Distrcit Name in df dataframe
gdf_merged = gdf_districts.merge(df, left_on='KGISDist_1', right_on='District', how='left')

import matplotlib.pyplot as plt
from matplotlib import colors
import numpy as np

# Define color scale
cmap = plt.cm.get_cmap('YlOrRd')  # Red to green colormap (reversed)
cmap.set_bad('white')  # Set NaN values to white
normalize = colors.Normalize(vmin=gdf_merged['MPCE_URP'].min(), vmax=gdf_merged['MPCE_URP'].max())

# Plot the map
fig, ax = plt.subplots(figsize=(10, 10))
gdf_districts.plot(ax=ax, facecolor='none', edgecolor='black', linewidth=0.8)  # Plot the district outlines

# Fill districts with color based on population values
infes_values = gdf_merged['MPCE_URP'].fillna(np.nanmin(gdf_merged['MPCE_URP']) - 1)  # Replace NaN values with a value lower than min
gdf_districts.plot(ax=ax, column=infes_values, cmap=cmap, linewidth=0, legend=False)

# Add district labels
for x, y, label in zip(gdf_merged.geometry.centroid.x, gdf_merged.geometry.centroid.y, gdf_merged['KGISDist_1']):
    ax.text(x, y, label, fontsize=8, ha='center', va='center')

# Set plot title and axis labels
ax.set_title('MPCE_URP in Karnataka Districts')
ax.set_xlabel('Longitude')
ax.set_ylabel('Latitude')

# Create and add colorbar
sm = plt.cm.ScalarMappable(cmap=cmap, norm=normalize)
sm.set_array([])
cbar = fig.colorbar(sm)
cbar.set_label('MPCE_URP')

# Show the plot
plt.show()
```

Result:

	District	MPCE_URP	MPCE_MRP
0	Bagalkot *	174637.12	188630.85
1	Bangalore	1971811.41	2097996.99
2	Bangalore Rural	147689.44	167381.05
3	Belgaum	359775.85	386126.91
4	Bellary	268798.33	307397.14
5	Bidar	132880.05	126224.45
6	Bijapur	216815.91	226354.48
7	Chamarajanagar *	223580.93	203602.53
8	Chikkaballapura	197961.96	191656.92
9	Chikmagalur	119842.93	137185.02
10	Chitradurga	121807.68	129759.68
11	Dakshina Kannada	689269.48	673994.71
12	Davanagere	210405.49	217903.53
13	Dharwad	240198.89	267251.33
14	Gadag *	112849.90	117375.49
15	Gulbarga	308941.08	309402.83
16	Hassan	195295.98	218821.13
17	Haveri *	151676.51	146201.05
18	Kodagu	219822.06	233656.29
19	Kolar	401164.07	412801.30
20	Koppal	149191.48	154493.29
21	Mandya	248067.76	257508.17



Interpretation:

The map represents the Monthly Per Capita Expenditure (MPCE) based on the Uniform Reference Period (URP) across various districts in Karnataka. MPCE is a key indicator of the economic well-being of households and reflects their consumption patterns. By analyzing the spatial distribution of MPCE, we can gain insights into economic disparities and consumption behavior within the state.

1. **High MPCE Districts:**

    - **Dakshina Kannada**: This district exhibits the highest level of MPCE, indicated by dark red shades on the map. High MPCE suggests that households in this district have higher consumption levels, which could be due to better economic conditions, higher income levels, or access to more resources and services.
    - **Raichur**: Represented by orange shades, Raichur also shows relatively high MPCE. This indicates that this district, like Dakshina Kannada, has a higher standard of living and better economic conditions compared to other districts.

2. **Low MPCE Districts:**

    - **Belagavi and Ballari**: These districts are shown in lighter yellow shades, indicating lower MPCE. This suggests that households in these districts have lower consumption levels, potentially due to lower income levels, limited access to resources, or other economic challenges.

The map's color gradient provides a visual representation of economic disparities across Karnataka. Districts with higher MPCE (dark shades) contrast with those having lower MPCE (lighter shades), highlighting regions with varying economic conditions. The significant variation in MPCE across districts underscores the presence of economic disparity within the state. This can have implications for social and economic policies aimed at reducing inequality.

1. **Identifying Economic Disparities:**

    The map serves as a crucial tool for visualizing and identifying economic disparities within Karnataka. By highlighting regions with high and low MPCE, it helps policymakers understand which areas require more attention and resources.

2. **Policy Interventions:**

    Policymakers can use this information to design targeted interventions aimed at uplifting economically backward regions. For example, districts with low MPCE might benefit from programs focused on job creation, improving access to education and healthcare, and enhancing infrastructure.

    Understanding the spatial distribution of consumption patterns aids in more effective resource allocation. Resources can be directed to regions that need them the most, ensuring a balanced and equitable development across the state.

3. **Regional Planning and Development:**

    The map can inform regional planning efforts by highlighting areas that require economic development. Planners can focus on creating opportunities for growth in lower MPCE districts, thereby promoting overall state development.

4. **Social and Economic Research:**

    Researchers can use the map to study the factors contributing to economic disparities within Karnataka. This could involve examining the relationship between MPCE and other socio-economic indicators like education levels, employment rates, and access to services.

The map depicting the Monthly Per Capita Expenditure (MPCE) based on the Uniform Reference Period (URP) in Karnataka is a valuable tool for visualizing economic disparities within the state. By identifying districts with high and low MPCE, it provides insights into regional consumption patterns and economic conditions. This information is crucial for policymakers, planners, and researchers to design targeted interventions, allocate resources effectively, and promote balanced regional development.

VISUALIZATION - PERCEPTUAL MAPPING FOR BUSINESS USING R

RESULTS AND INTERPRETATION

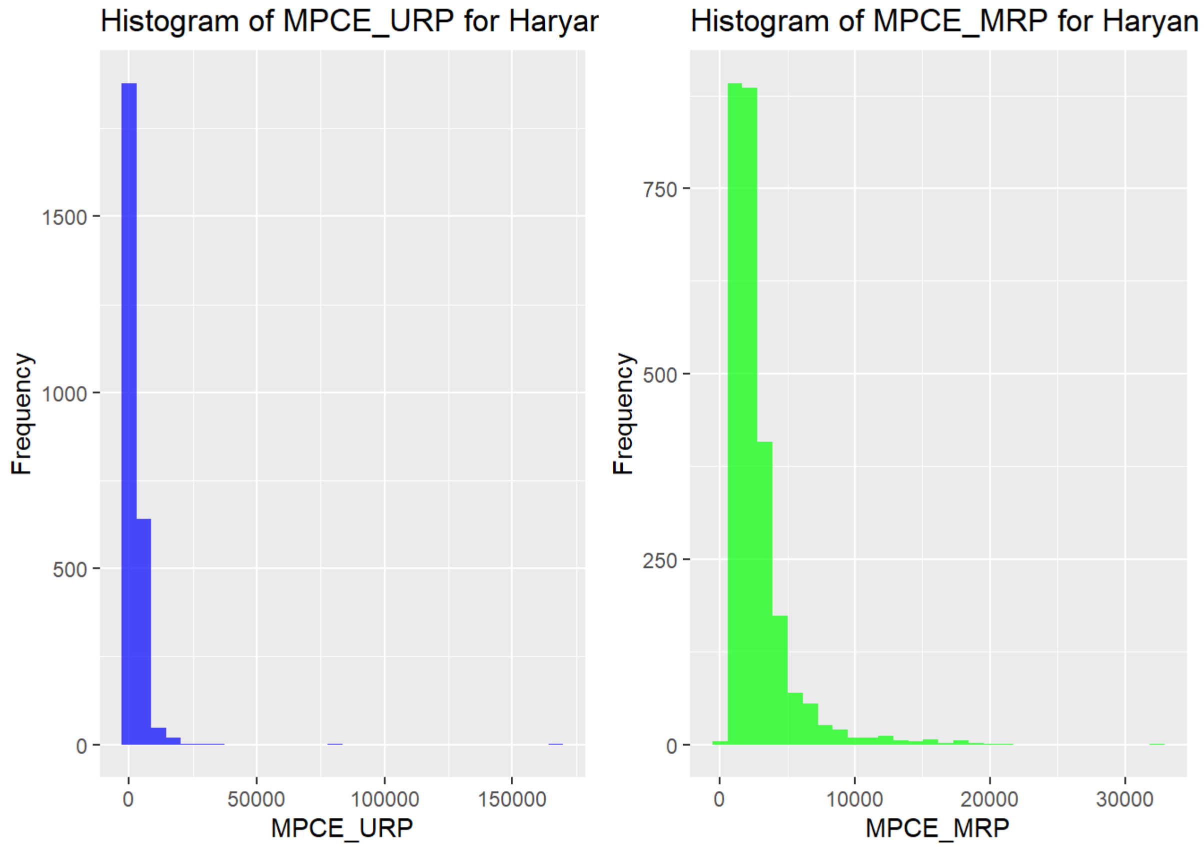
Objective 1. Draw a histogram of the data of the state assigned to indicate the consumption district-wise. [NSSO68.csv]

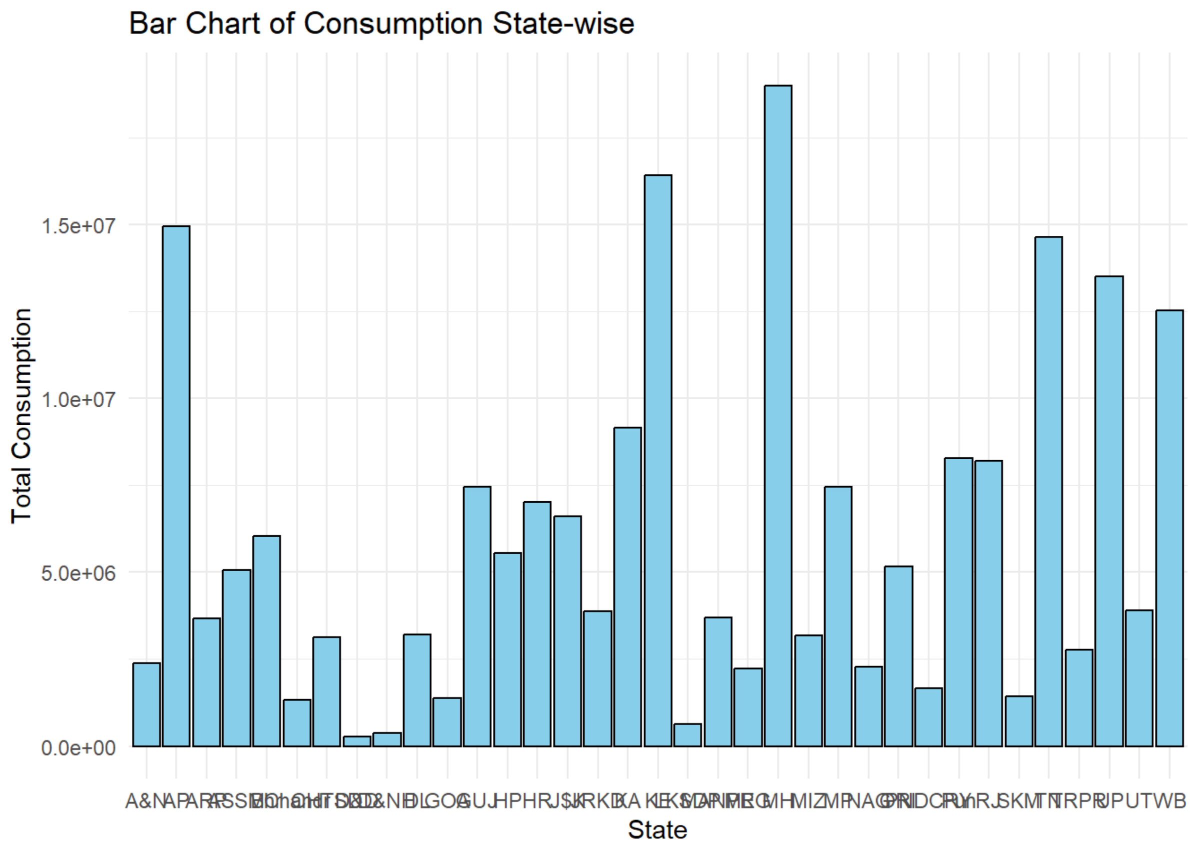
Code:
Draw a histogram of the data to indicate the consumption district-wise.

```
# Function to auto-install and load packages
install_and_load <- function(packages) {
  for (package in packages) {
    if (!require(package, character.only = TRUE)) {
      install.packages(package, dependencies = TRUE)
    }
    library(package, character.only = TRUE)
  }
}

# List of packages to install and load
packages <- c("readr", "ggplot2", "dplyr", "gridExtra")

# Call the function
install_and_load(packages)

# Load the data
file_path <- 'NSSO68.csv'
data <- read_csv(file_path)

# Inspect the first few rows of the dataset to understand its structure
print(head(data))

# Subset the data with specified columns and filter based on state_1
subset_data <- data %>%
  select(state, District, MPCE_URP, MPCE_MRP, state_1)

filtered_data <- subset_data %>%
  filter(state_1 == 'HR')

# Plot histograms for MPCE_URP and MPCE_MRP
p1 <- ggplot(filtered_data, aes(x = MPCE_URP)) +
  geom_histogram(bins = 30, fill = 'blue', alpha = 0.7) +
  ggtitle('Histogram of MPCE_URP for Haryana) +
  xlab('MPCE_URP') +
  ylab('Frequency')

p2 <- ggplot(filtered_data, aes(x = MPCE_MRP)) +
  geom_histogram(bins = 30, fill = 'green', alpha = 0.7) +
```

```
  ggtitle('Histogram of MPCE_MRP for Haryana) +
  xlab('MPCE_MRP') +
  ylab('Frequency')

# Arrange the plots side by side
grid.arrange(p1, p2, ncol = 2)

# Group the data by state and sum the consumption
state_consumption <- data %>%
  group_by(state_1) %>%
  summarise(total_consumption = sum(MPCE_URP))

# Plot the bar chart
ggplot(state_consumption, aes(x = state_1, y = total_consumption)) +
  geom_bar(stat = "identity", color = "black", fill = "skyblue") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(title = "Bar Chart of Consumption State-wise",
       x = "State",
       y = "Total Consumption") +
  theme_minimal()
```

Result:





Interpretation:

The histograms display the Monthly Per Capita Expenditure (MPCE) for Haryana using two different reference periods: the Uniform Reference Period (URP) and the Modified Reference Period (MRP). These visualizations help in understanding the distribution of household expenditures across the state.

Both histograms for MPCE_URP (blue) and MPCE_MRP (green) show a highly skewed distribution with a long right tail. This indicates that most households fall into the lower expenditure brackets, with only a few households reporting higher expenditures.

The histogram for MPCE_URP shows a very sharp decline in frequency as the expenditure increases. The majority of households have an MPCE below 25,000, with a significant concentration below 10,000. A small number of households have much higher expenditures, but these are rare, as indicated by the long tail extending towards higher MPCE values. Similar to the MPCE_URP, the MPCE_MRP histogram also indicates a concentration of households in the lower expenditure brackets, although it shows a relatively broader spread compared to MPCE_URP. The MRP method captures a wider range of expenditure values compared to the URP, but both methods highlight that the majority of the population has low monthly per capita expenditure.

The histograms for MPCE using URP and MRP data for Haryana reveal a highly skewed distribution, with most households falling into the lower expenditure brackets. This indicates significant economic disparities within the state, with a majority of the population experiencing low levels of consumption. The slightly broader spread in the MPCE_MRP histogram compared to MPCE_URP suggests that the MRP method captures a wider range of expenditure values. These insights highlight potential areas for targeted economic development and welfare programs to address the economic inequalities and improve the overall well-being of households in Haryana.

The bar chart illustrating total consumption state-wise reveals substantial variation in consumption across different states. States like Andhra Pradesh (AP) and Maharashtra (MH) exhibit notably high total consumption levels, suggesting either a larger population, higher per capita consumption, or a combination of both factors. In contrast, states such as Daman and Diu (D&D) and Dadra and Nagar Haveli (D&NH) show minimal total consumption, likely due to their smaller populations. The significant differences in total consumption across states highlight the diverse economic landscapes in India. States with higher consumption may benefit from strong economic activities and better living standards, while those with lower consumption may need more attention in terms of development policies and resource allocation.

Objective 2. Depict the consumption on the state map, showing consumption in each district for the state of Karnataka. [NSSO68.csv]

Code:

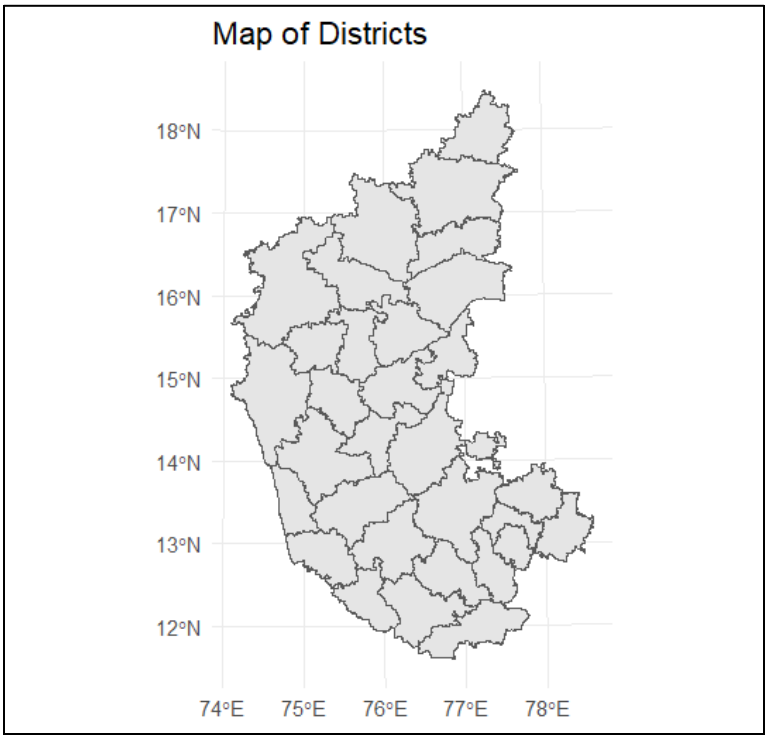
Depict the consumption on the state map, showing consumption in each district.

```
# Function to auto-install and load packages
install_and_load <- function(packages) {
  for (package in packages) {
    if (!require(package, character.only = TRUE)) {
      install.packages(package, dependencies = TRUE)
    }
    library(package, character.only = TRUE)
  }
}

# List of packages to install and load
packages <- c("sf", "ggplot2", "dplyr", "readr","readxl","janitor","tidyverse")

# Call the function
install_and_load(packages)

# Load the shapefile
gdf_districts <- st_read('District.shp')

# Display the first few rows
head(gdf_districts)

# Plot the map
ggplot() +
  geom_sf(data = gdf_districts) +
```

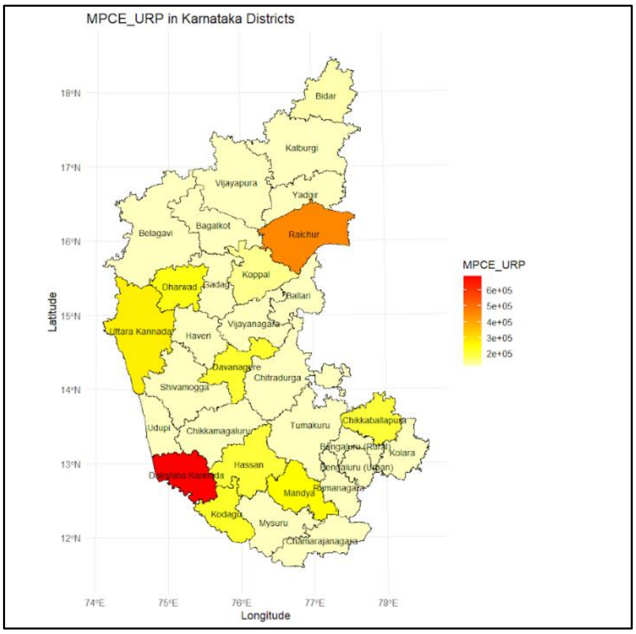
```
  labs(title = "Map of Districts") +
  theme_minimal()

# Load NSSO68.csv file
nsso_data <- read_csv('NSSO68.csv')

# Subset the data with specified columns and filter based on state_1 (assuming state_1 is
equivalent to state in R)
subset_data <- nsso_data %>%
  select(state_1, District, MPCE_URP, MPCE_MRP) %>%
  filter(state_1 == 'KA')  # Filter for Karnataka state

# Load district-codes.xlsx file
district_codes <- read_excel('district-codes.xlsx')
district_codes <- clean_names(district_codes)
names(district_codes)


# Filter district codes for Karnataka
karnataka_districts <- district_codes %>%
  filter(state_name == 'Karnataka')

# Create a mapping from district codes to district names
district_mapping <- karnataka_districts %>%
  select(dc, `district_name`) %>%
  deframe()

# Replace district codes in the filtered NSSO data with district names
subset_data <- subset_data %>%
  mutate(District = district_mapping[District])

# Sum all values district-wise
district_wise_sum <- subset_data %>%
  group_by(District) %>%
  summarise(MPCE_URP = sum(MPCE_URP), MPCE_MRP = sum(MPCE_MRP))

# Display the resulting DataFrame
print(district_wise_sum)

df <- district_wise_sum

# Merge the shapefile with the DataFrame
gdf_merged <- gdf_districts %>%
  left_join(df, by = c('KGISDist_1' = 'District'))


# Replace NaN values with a value lower than min for visualization purposes
gdf_merged$MPCE_URP[is.na(gdf_merged$MPCE_URP)] <-
min(gdf_merged$MPCE_URP, na.rm = TRUE) - 1
```

```
# Calculate the centroids
gdf_merged$centroid <- st_centroid(gdf_merged$geometry)

# Extract the coordinates of the centroids
gdf_merged$X <- st_coordinates(gdf_merged$centroid)[, "X"]
gdf_merged$Y <- st_coordinates(gdf_merged$centroid)[, "Y"]

# Plot using ggplot2
ggplot() +
  geom_sf(data = gdf_districts, color = "black", fill = "white") +
  geom_sf(data = gdf_merged, aes(fill = MPCE_URP), color = "black") +
  scale_fill_gradientn(colors = rev(heat.colors(10)), na.value = "white", name =
"MPCE_URP") +
  geom_text(data = gdf_merged, aes(x = X, y = Y, label = KGISDist_1), size = 3) +
  labs(title = "MPCE_URP in Karnataka Districts",
       x = "Longitude",
       y = "Latitude") +
  theme_minimal()
```

Result:



```
## # A tibble: 29 × 3
##    District          MPCE_URP MPCE_MRP
##    <chr>                <dbl>    <dbl>
##  1 Bagalkot *         174637.  188631.
##  2 Bangalore         1971811. 2097997.
##  3 Bangalore Rural    147689.  167381.
##  4 Belgaum            359776.  386127.
##  5 Bellary            268798.  307397.
##  6 Bidar              132880.  126224.
##  7 Bijapur            216816.  226354.
##  8 Chamarajanagar *   223581.  203603.
##  9 Chikkaballapura    197962.  191657.
## 10 Chikmagalur        119843.  137185.
## # ℹ 19 more rows
```



Interpretation:

A shapefile is a widely-used digital vector storage format for storing geometric locations and associated attribute information of geographic features. Developed by Esri, it can store data representing various geographic features, such as points, lines, and polygons, making it ideal for use in Geographic Information Systems (GIS). This dataset allows users to perform spatial analysis, create detailed maps, and make data-driven decisions based on geographic patterns and relationships. For instance, it can assist in urban planning, resource management, and environmental monitoring by visualizing and analyzing spatial distributions and relationships among different geographic features.

The map showcases the distribution of Monthly Per Capita Expenditure (MPCE) based on the Uniform Reference Period (URP) across the districts of Karnataka. The data, derived from the NSSO68 survey, has been visualized on the Karnataka state map, highlighting the variations in consumption across different districts. A color gradient from yellow to red represents increasing levels of MPCE, with yellow indicating lower expenditure and red indicating higher expenditure.

Yellow: Represents lower MPCE.

Red: Represents higher MPCE.

This gradient provides a visual cue to easily identify regions with varying levels of expenditure.

High MPCE Districts:

1. **Dakshina Kannada (Dark Red):**

    - This district has a high MPCE, indicating a higher average consumption level.
    - Possible reasons include higher income levels, better economic activities, and more developed infrastructure.

2. **Raichur (Orange):**

    - Exhibits high MPCE, similar to Dakshina Kannada.
    - Economic activities such as agriculture, power generation, and industrial presence might contribute to this.

Low MPCE Districts:

1. **Belagavi (Light Yellow):**

    - Shows lower MPCE, indicating lower average consumption.
    - Factors might include lower income levels, less industrialization, and fewer economic opportunities.

2. **Ballari (Light Yellow):**

    - Like Belagavi in terms of low MPCE.
    - Economic disparities and limited infrastructure development could be contributing factors.

```
# Subset the data with specified columns and filter based on state_1
subset_data = data[['state', 'District', 'MPCE_URP', 'MPCE_MRP', 'state_1']]
filtered_data = subset_data[subset_data['state_1'] == 'HR']
# Save the filtered data to a CSV file
filtered_data.to_csv('Haryana.csv', index=False)

print("Filtered data saved to 'Haryana.csv'")
```

```
Filtered data saved to 'Haryana.csv'
```

This step involved making a subset of the file NSS068.csv for Haryana state.