



VIRGINIA COMMONWEALTH UNIVERSITY

Statistical analysis and modelling (SCMA 632)

A4a- Visualization - Perceptual Mapping for Business

N V SATYANARAYAN

V01108247

Date of Submission: 15-07-2024

CONTENTS

Sl. No.	Title	Page No.
1.	Introduction	1
2.	Objectives	1
3.	Business Significance	1-2
4.	Results and analysis	2-20
5.	Conclusion	
6.	References	

***NOTE- PYTHON AND R CODES WTH RESULT ADDED IN GITHUB- [Satyanaldiga \(github.com\)](https://github.com/Satyanaldiga)**

Introduction

In the contemporary business environment, data visualization plays a crucial role in decision-making processes. One of the powerful tools in this domain is perceptual mapping, which helps businesses visualize the relative positioning of different entities (such as products, services, or districts) based on various attributes. This technique is particularly useful for identifying patterns, trends, and outliers in large datasets, enabling businesses to make informed strategic decisions.

This project focuses on the state of Madhya Pradesh, utilizing data from the National Sample Survey Office (NSSO) 68th round to analyze district-wise consumption patterns. By employing visualization techniques such as histograms and bar plots, we aim to uncover insights into the distribution of total consumption across different districts and highlight key consumption trends.

Objective

The primary objectives of this project are:

1. To visualize the distribution of total consumption across different districts in Madhya Pradesh using a histogram.
2. To create a bar plot that showcases the consumption per district, labeled with district names, to facilitate easy comparison.
3. To plot the "Meals at Home" data on a map of Madhya Pradesh, providing a geographical perspective on consumption patterns.

Business Significance

Understanding consumption patterns at the district level is vital for several reasons:

1. **Resource Allocation:** Businesses and government agencies can allocate resources more efficiently by identifying districts with high or low consumption.
2. **Market Strategy:** Companies can tailor their marketing and sales strategies based on consumption trends, ensuring that they target the right audience with appropriate products and services.
3. **Policy Formulation:** Policymakers can design targeted interventions and programs to address disparities in consumption, promoting balanced economic development across districts.
4. **Supply Chain Optimization:** By understanding regional consumption patterns, businesses can optimize their supply chains, reducing costs and improving service delivery.

District-Wise Consumption Patterns in Madhya Pradesh

CODES AND INTERPRETATION

PYTHON CODES

```
import pandas as pd
import geopandas as gpd
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import iqr
```

- **pandas (pd):**

Purpose: pandas is a powerful library for data manipulation and analysis. It provides data structures like DataFrames, which are used to store and manipulate tabular data.

Usage: This library will be used to load, process, and analyze the dataset (NSSO68.csv in this case).

- **geopandas (gpd):**

Purpose: geopandas extends pandas to allow spatial operations on geometric types. It enables working with geospatial data in a pandas-like manner.

Usage: This library will be used to handle and visualize geographical data, such as plotting data on a map of Madhya Pradesh.

- **matplotlib.pyplot (plt):**

Purpose: matplotlib is a plotting library for creating static, animated, and interactive visualizations in Python. pyplot is a module within matplotlib that provides a MATLAB-like interface for making plots.

Usage: This library will be used to create histograms, bar plots, and other visualizations to represent the data graphically.

- **numpy (np):**

Purpose: numpy is a fundamental library for numerical computing in Python. It provides support for arrays, matrices, and many mathematical functions to operate on these data structures.

Usage: This library will be used for numerical operations, such as calculating statistics or handling arrays of data.

- **scipy.stats (iqr):**

Purpose: scipy is a library used for scientific and technical computing. The stats module within scipy contains a large number of probability distributions and statistical functions.

Usage: The iqr function (Interquartile Range) from scipy.stats will be used to calculate the interquartile range, which is a measure of statistical dispersion. This can help in identifying the spread and outliers in the consumption data.

```
print(df.head())
```

	slno	grp	Round_Centre	FSU_number	Round	Schedule_Number	\
2091	2092	4.13E+31	1	41310	68	10	
2092	2093	4.13E+31	1	41310	68	10	
2093	2094	4.13E+31	1	41310	68	10	
2094	2095	4.13E+31	1	41310	68	10	
2095	2096	4.13E+31	1	41310	68	10	

	Sample	Sector	state	State_Region	...	pickle_v	sauce_jam_v	\
2091	1	2	23	231	...	0.0	0.0	
2092	1	2	23	231	...	0.0	0.0	
2093	1	2	23	231	...	0.0	0.0	
2094	1	2	23	231	...	0.0	0.0	
2095	1	2	23	231	...	0.0	0.0	

	Othrprocessed_v	Beveragestotal_v	foodtotal_v	foodtotal_q	state_1
2091	0.0	36.000000	552.822800	21.510180	MP
2092	0.0	16.666667	738.463667	26.233667	MP
2093	0.0	3.750000	337.001000	15.087706	MP
2094	0.0	7.142857	500.162857	20.257300	MP
2095	0.0	10.000000	483.254000	23.425425	MP

	Region	fruits_df_tt_v	fv_tot
2091	1	45.200000	60.200000
2092	1	27.666667	49.766667
2093	1	0.250000	11.500000
2094	1	22.285714	36.285714
2095	1	0.500000	33.000000

```
[5 rows x 384 columns]
```

The head() method displays the first five rows of the DataFrame, providing a quick look at the structure and content of the data.

```
print("Top Consuming Districts:")
```

```
print(district_summary.head(4))
```

```
Top Consuming Districts:
   District      total
0   Ujjain  1163.419995
1   Indore   916.912049
2   Bhind   880.732053
3  Tikamgarh  826.475000
```

In [25]:

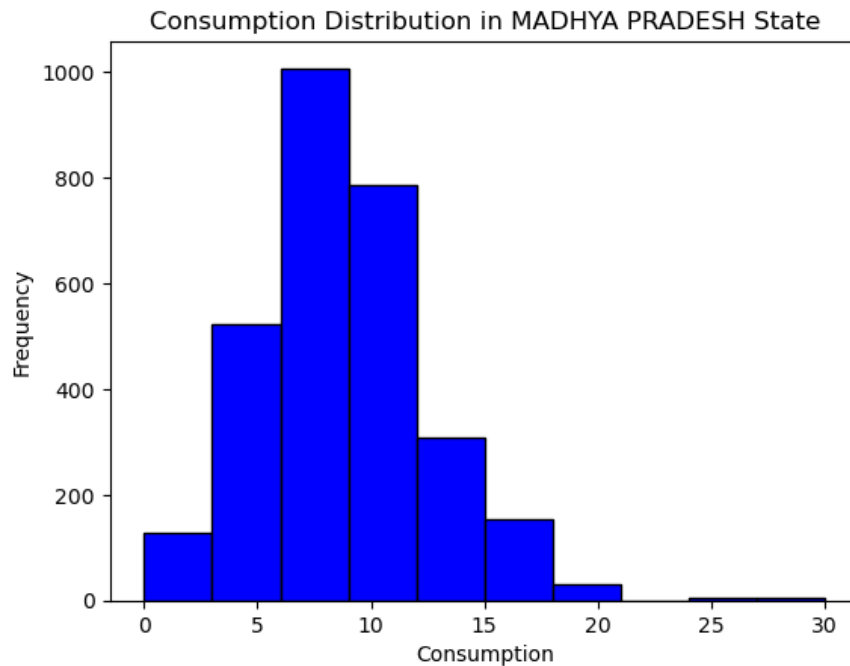
Interpretation:

1. **Top Consuming Districts:** The table lists the top four districts in terms of total consumption. Each row represents a district, and the corresponding total consumption value.
2. **District:**
 - **Ujjain:** This district has the highest total consumption, amounting to approximately 1163.42 units.
 - **Indore:** The second highest total consumption with about 916.91 units.
 - **Bhind:** The third highest total consumption, totaling around 880.73 units.
 - **Tikamgarh:** The fourth highest total consumption with approximately 826.48 units.
3. **Total Consumption:** The total column indicates the aggregate consumption value for each district. This could represent a sum of various consumption metrics such as food, beverages, and other goods.

Business Insights:

- **Resource Allocation:** Ujjain, Indore, Bhind, and Tikamgarh are key districts with high consumption. Businesses can focus marketing efforts and resource allocation in these areas to maximize sales.
- **Market Strategy:** These districts might have a higher demand for products and services. Tailoring marketing campaigns to these regions can help in capturing a larger market share.
- **Supply Chain Optimization:** Ensuring efficient supply chains and inventory management in these districts can prevent stockouts and meet the high demand efficiently.

Consumption Distribution in MADHYA PRADESH State



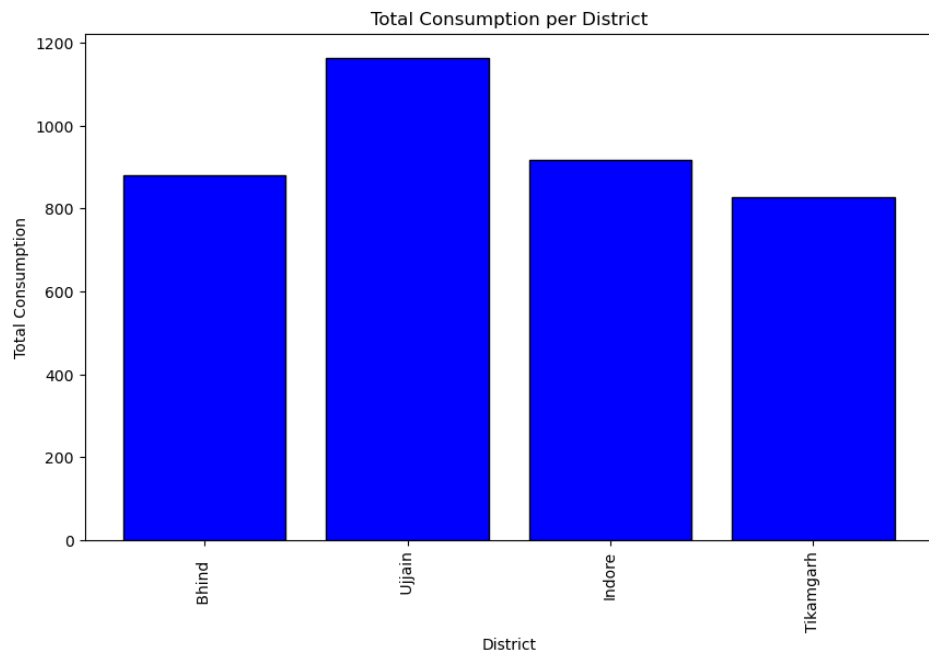
- X-Axis (Consumption):**
 - The x-axis represents the total consumption values. The consumption ranges from 0 to 30 units, divided into 10 bins (or intervals).
- Y-Axis (Frequency):**
 - The y-axis represents the frequency, or the number of districts, that fall within each consumption range.
- Distribution Shape:**
 - The histogram shows a right-skewed distribution, meaning that most of the districts have lower to moderate consumption values, with fewer districts having very high consumption.
- Peak Frequency:**
 - The highest bar is around the 5-10 consumption range, indicating that the most common consumption value among districts falls within this range. Approximately 1000 districts fall into this category.
- Lower Consumption:**
 - There are smaller bars towards the lower consumption range (0-5), indicating fewer districts have very low consumption values.
- Higher Consumption:**

- The bars gradually decrease as the consumption value increases, showing fewer districts with higher consumption values. Very few districts have consumption values above 20.

Business Insights:

- **Focus on Moderate Consumption:** Since the majority of districts have moderate consumption values, businesses should focus their marketing and distribution efforts on these areas to maximize reach and impact.
- **Identify High Consumption Districts:** The few districts with higher consumption values (above 15) could be key areas for premium product offerings and targeted marketing strategies.
- **Resource Allocation:** Understanding the distribution helps in better resource allocation, ensuring that products and services are available in sufficient quantities where demand is highest.

Total Consumption per District



1. **X-Axis (District):**
 - The x-axis lists the districts: Bhind, Ujjain, Indore, and Tikamgarh.
 - The district names are rotated vertically for better readability.
2. **Y-Axis (Total Consumption):**

- The y-axis indicates the total consumption values for each district.
- 3. **Bars:**
 - Each bar represents a district's total consumption.
 - The height of each bar corresponds to the total consumption value for that district.
- 4. **Total Consumption Values:**
 - **Bhind:** The total consumption is approximately 880 units.
 - **Ujjain:** The total consumption is the highest among the districts, around 1163 units.
 - **Indore:** The total consumption is around 917 units.
 - **Tikamgarh:** The total consumption is approximately 826 units.

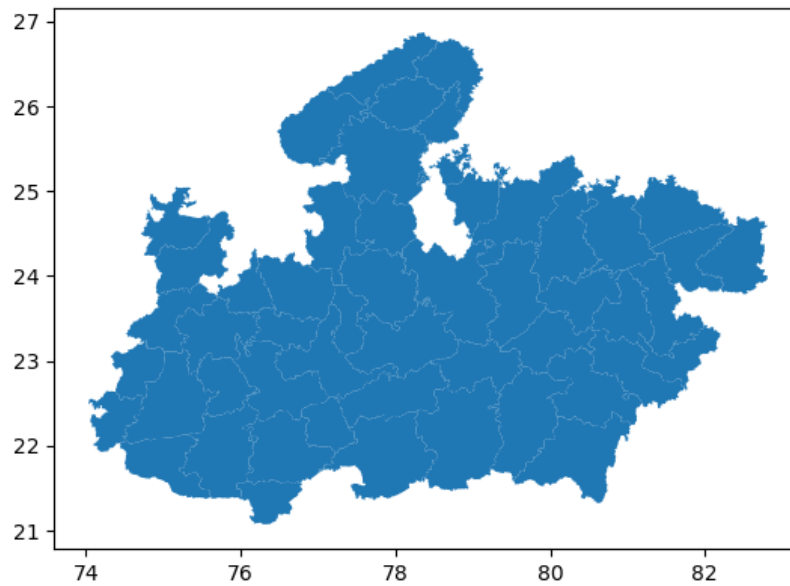
Key Insights:

1. **Ujjain:**
 - Ujjain has the highest total consumption, indicating it is the district with the greatest demand for goods/services among the listed districts.
2. **Indore:**
 - Indore follows Ujjain, with a significant consumption value, highlighting it as another key area for market activities.
3. **Bhind and Tikamgarh:**
 - Both Bhind and Tikamgarh have relatively lower total consumption values compared to Ujjain and Indore, but they still represent notable consumption levels.

Business Implications:

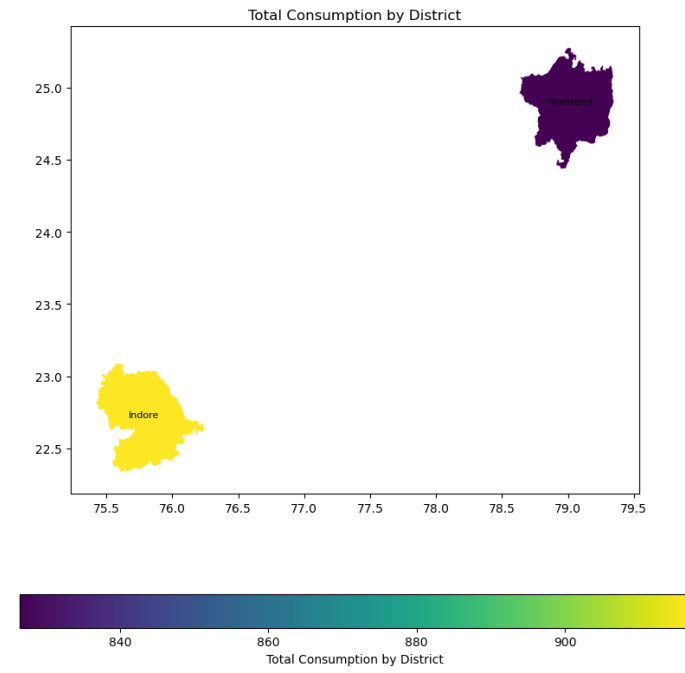
- **Resource Allocation:**
 - Allocate more resources and inventory to Ujjain and Indore to meet higher demand.
 - Plan marketing strategies targeting these high consumption districts to maximize sales.
- **Market Strategy:**
 - Tailor product offerings and promotions to the specific needs and preferences of consumers in Ujjain and Indore.
 - Consider developing strategies to boost consumption in Bhind and Tikamgarh to balance market presence across districts.

Plotting the Map



Plotting the Map: With GeoPandas, creating a basic map is as simple as calling the `plot()` function on the `GeoDataFrame`. This will generate a visual representation of Madhya Pradesh boundaries.

Total Consumption by District



- Geographical Position:**
 - Indore is located in the southwestern part of the state.
 - Tikamgarh is situated in the northeastern part of the state.
- Color Representation:**
 - The color gradient indicates the level of total consumption, ranging from dark purple (lower consumption) to yellow (higher consumption).
 - Indore is colored yellow, indicating it has a higher total consumption.
 - Tikamgarh is colored dark purple, indicating it has a lower total consumption.
- Consumption Values:**
 - The color bar at the bottom provides a scale for interpreting the colors.
 - Indore, represented in yellow, has a total consumption value around 900.
 - Tikamgarh, represented in dark purple, has a total consumption value around 840.
- Title and Labels:**
 - The title of the graph is "Total Consumption by District."
 - The x-axis and y-axis likely represent geographic coordinates (longitude and latitude, respectively).

This visualization highlights the disparity in total consumption between Indore and Tikamgarh, with Indore having significantly higher consumption levels.

RCODES

```
> print(dim(df))
```

```
[1] 4717 384
```

The command `print(dim(df))` returns the dimensions of a data frame `df`. The output `[1] 4717 384` indicates the following:

- **Rows:** There are 4,717 rows in the data frame.
- **Columns:** There are 384 columns in the data frame.

```
> print(colSums(is.na(MPData)))
      state_1      District      Region      Sector
           0           0           0           0
State_Region Meals_At_Home ricepds_v wheatpds_q
           0           26           0           0
  chicken_q    pulsep_q    wheatos_q No_of_Meals_per_day
           0           0           0           3
```

The command `print(colSums(is.na(MPData)))` calculates the number of missing (NA) values in each column of the data frame `MPData`. The output indicates how many missing values are present in specific columns. Here's the interpretation of the provided output:

- **Columns with No Missing Values:**

- state_1
- District
- Region
- Sector
- State_Region
- ricepds_v
- Wheatpds_q
- chicken_q
- pulsep_q
- wheatos_q

These columns have 0 missing values.

- **Columns with Missing Values:**

- Meals_At_Home: 26 missing values
- No_of_Meals_per_day: 3 missing values

```
print(unique(MPData$District))
[1] "Alirajpur" "Narsimhapur" "Dindori" "Anuppur" "Jabalpur"
[6] "Seoni" "Mandla" "Panna" "Katni" "Sidhi"
[11] "Satna" "Shahdol" "Damoh" "Umaria" "Rewa"
[16] "Sagar" "Singrauli" "Tikamgarh" "Gwalior" "Guna"
[21] "Bhind" "Shivpuri" "Datia" "Morena" "Chhatarpur"
```

```
[26] "Sheopur"      "Dewas"      "Jhabua"      "Shajapur"    NA
[31] "Rajgarh"      "Ujjain"      "Neemuch"      "Mandsaur"    "Ratlam"
[36] "Indore"       "Bhopal"      "Burhanpur"    "Chhindwara"  "Raisen"
[41] "Betul"        "Sehore"      "Hoshangabad"  "Harda"       "Balaghat"
[46] "Vidisha"      "East Nimar"  "Barwani"      "West Nimar"  "Dhar"
```

The command `print(unique(MPData$District))` lists all the unique district names present in the District column of the MPData data frame.

```
print(head(district_summary, 3))
# A tibble: 3 × 2
  District total
  <chr>      <dbl>
1 Ujjain    1163.
2 Indore     917.
3 Bhind      881.
> cat("Bottom 3 Consuming Districts:\n")
Bottom 3 Consuming Districts:
> print(tail(district_summary, 3))
# A tibble: 3 × 2
  District total
  <chr>      <dbl>
1 Mandla    151.
2 Anuppur   147.
3 Dindori   111.
```

Top 3 Consuming Districts:

- **Ujjain:** Total consumption value of 1163.
- **Indore:** Total consumption value of 917.
- **Bhind:** Total consumption value of 881.

Bottom 3 Consuming Districts:

- **Mandla:** Total consumption value of 151.
- **Anuppur:** Total consumption value of 147.
- **Dindori:** Total consumption value of 111.

This indicates that Ujjain, Indore, and Bhind are the top three districts in terms of total consumption, with Ujjain having the highest consumption. Conversely, Mandla, Anuppur, and Dindori are the bottom three districts, with Dindori having the lowest total consumption.

```
print(region_summary)
# A tibble: 6 × 2
  Region total
  <chr>      <dbl>
1 URBAN    6994.
2 URBAN    4845.
3 RURAL    4360.
4 URBAN    3684.
5 URBAN    2827.
6 URBAN    2411.
```

Region-wise Total Consumption:

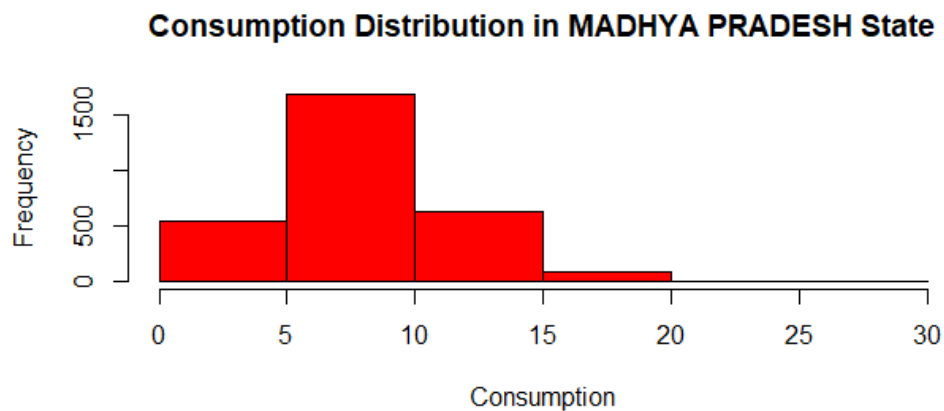
- **URBAN:** Total consumption values are 6994, 4845, 3684, 2827, and 2411.
- **RURAL:** Total consumption value is 4360.

This indicates that urban areas generally have higher total consumption values compared to rural areas. The highest total consumption is in an urban region with a value of 6994, while the lowest in the list is also in an urban region with a value of 2411. The only rural region listed has a total consumption value of 4360, which is lower than some urban regions but higher than others.

The summary suggests a pattern where urban regions tend to have higher consumption values compared to rural regions, reflecting potential differences in economic activity, lifestyle, and access to resources.

#A5)Visualization

#A)histogram to show the distribution of total consumption across different districts in madhya pradesh



Title:

- The title of the graph is "Consumption Distribution in MADHYA PRADESH State," indicating that the focus is on the distribution of consumption values across districts within the state.

X-axis (Consumption):

- The x-axis represents the consumption values, which range from 0 to 30.

Y-axis (Frequency):

- The y-axis represents the frequency of districts falling within specific consumption value ranges.

Distribution:

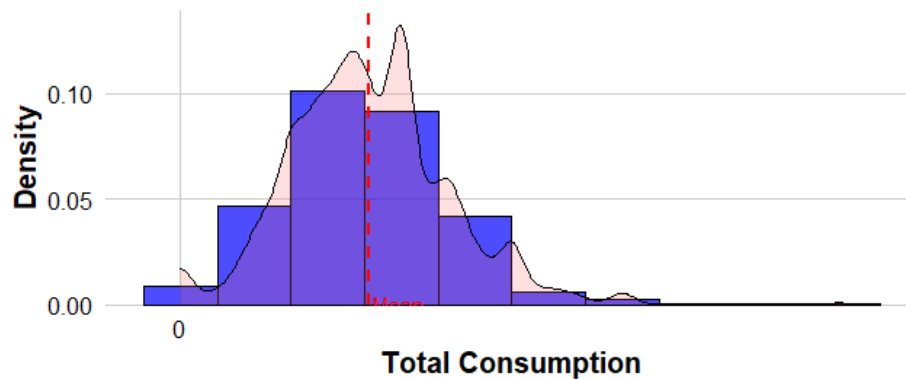
- The histogram bars indicate how many districts fall within each range of consumption values.
- The majority of districts have consumption values between 5 and 10, as indicated by the highest bar reaching a frequency of about 1500.
- Fewer districts have consumption values in the range of 0 to 5 and 10 to 15.
- Very few districts have consumption values greater than 15, as shown by the low frequency of bars beyond this range.

Summary:

- The distribution is right-skewed, with most districts having consumption values between 5 and 10, and fewer districts having higher consumption values.
- This suggests that while a majority of districts have moderate consumption, there are some districts with very high consumption, but they are relatively few.

Distribution of Total Consumption Across Districts in MADHYA PRADESH.

Distribution of Total Consumption Across Districts in MADHYA



X-axis (Total Consumption):

- The x-axis represents the total consumption values.

Y-axis (Density):

- The y-axis represents the density, indicating how frequently different total consumption values occur.

Histogram:

- The histogram is represented by the blue bars. It shows the frequency of districts within specific ranges of total consumption values.
- The height of the bars represents the number of districts with total consumption values in that range.

Density Plot:

- The black line represents the density plot, giving a smoothed estimate of the distribution of total consumption values.
- The density plot helps to visualize the overall shape of the distribution more clearly than the histogram alone.

Mean Line:

- The red dashed vertical line marks the mean of the total consumption values.
- This line helps to see where the average total consumption falls within the distribution.

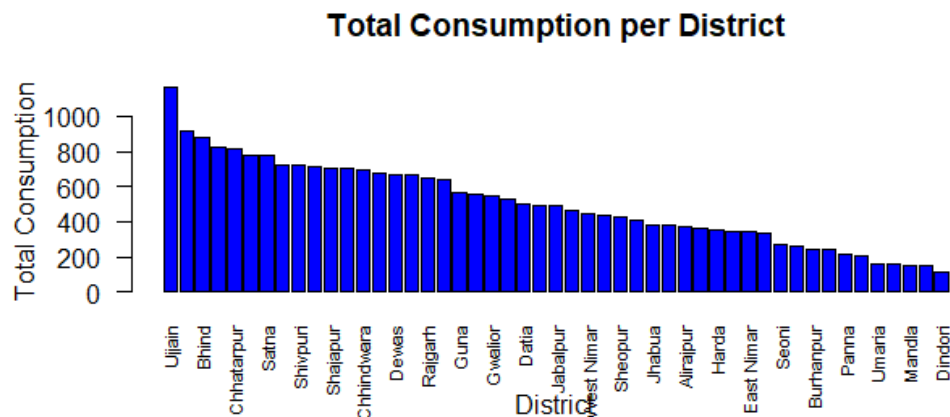
Distribution Shape:

- The distribution appears to be roughly symmetrical but slightly right-skewed, with a higher concentration of districts around the mean and tapering off towards higher consumption values.
- There are fewer districts with very high total consumption values, as indicated by the tapering of the density plot on the right side.

Summary:

- Most districts have total consumption values around the mean, with fewer districts having very low or very high consumption values.
- The combination of the histogram and density plot provides a comprehensive view of how total consumption is distributed among the districts in Madhya Pradesh.

Total Consumption per District



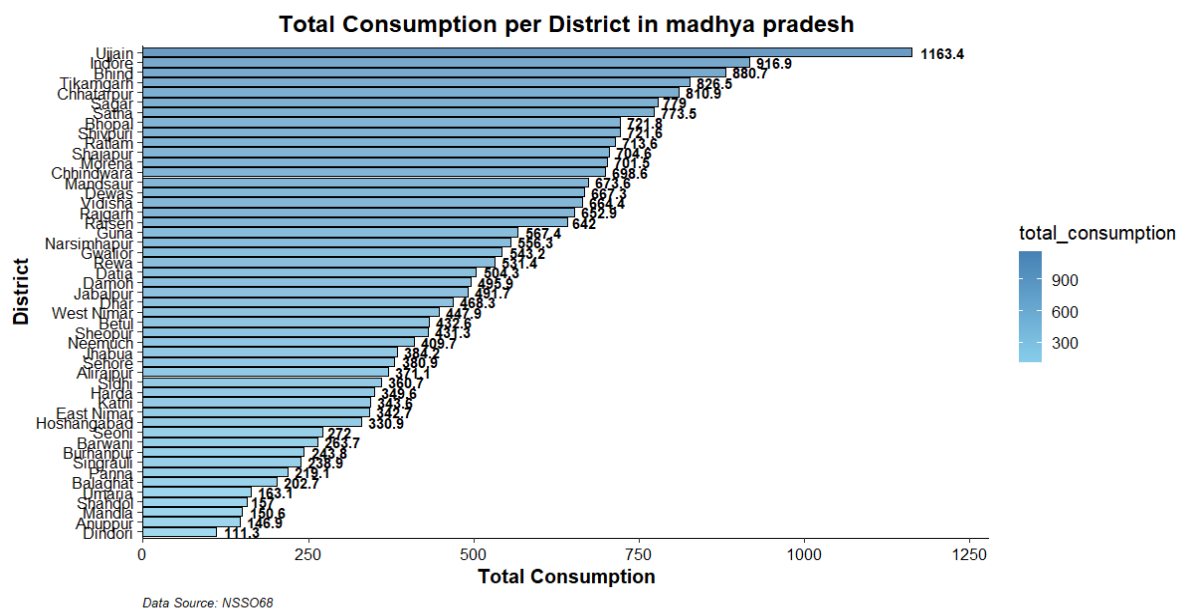
Key observations:

1. **Highest Consumption:** The district of Ujjain has the highest total consumption, exceeding 1000 units.
2. **Top Performers:** Following Ujjain, the districts of Bina, Chhatarpur, and Shahdol also show high levels of consumption, though none reach the same level as Ujjain.
3. **Moderate Consumption:** Districts such as Shivpuri, Shajapur, Chhindwara, Dewas, Rajgarh, and Guna show moderate levels of consumption.
4. **Lower Consumption:** Districts like Vidisha, Seoni, Jabalpur, Alirajpur, Harda, and several others exhibit lower total consumption.

- Lowest Consumption:** The districts with the lowest consumption are Mandla and Dindori.

The graph visually highlights the disparity in consumption levels across the districts, with Ujjain standing out significantly. The trend indicates a gradual decrease in consumption from left to right on the x-axis.

Total Consumption per District in Madhya Pradesh

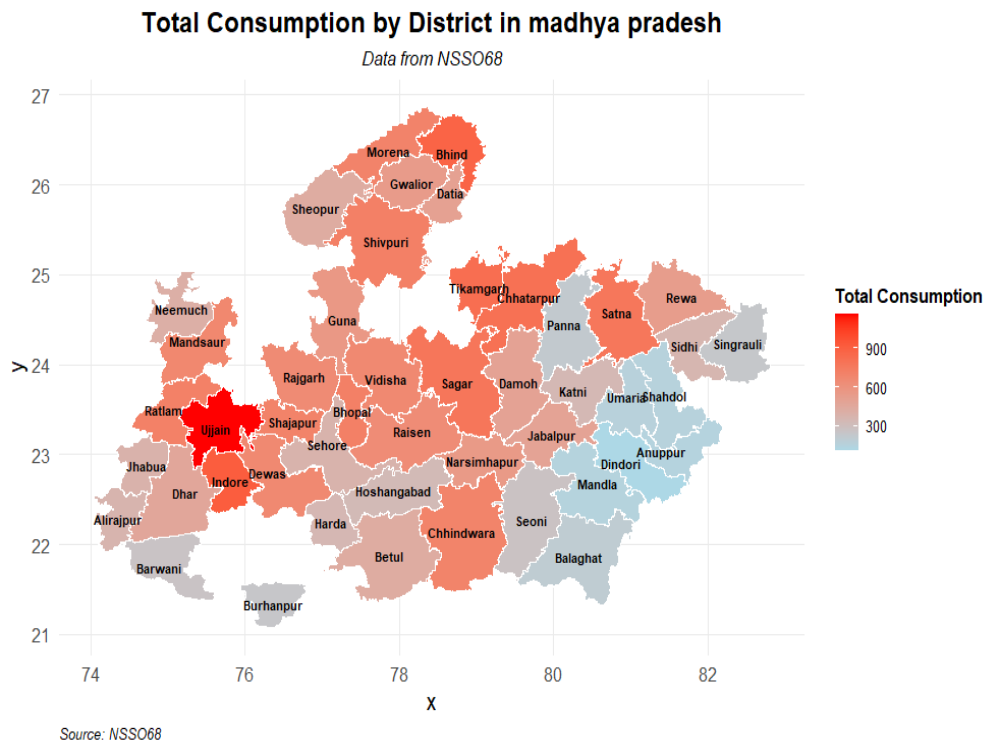


Key observations:

- Highest Consumption:**
 - Ujjain has the highest total consumption at 1163.4 units.
 - Indore follows closely with 916.9 units.
 - Other high-consumption districts include Bhopal (880.7 units), Bina (826.5 units), and Chhatarpur (810.9 units).
- Moderate Consumption:**
 - Districts such as Sagar (773.5 units), Ratlam (721.8 units), Shivpuri (704.6 units), and Raisen (642 units) show moderate levels of consumption.
 - Several other districts fall in this mid-range category, indicating a substantial but not the highest level of consumption.
- Lower Consumption:**
 - Districts such as Datia (534.2 units), Dindori (111.3 units), Barwani (219.9 units), and Anuppur (126.9 units) exhibit lower total consumption.

- A number of districts, including **Singrauli, Panna, and Balaghat**, fall toward the lower end of the spectrum, with consumption levels significantly below the top performers.

#B)Plot Total consumption and Meals at home on the madhya pradesh state map



This graph shows the "Total Consumption by District in Madhya Pradesh" on a geographical map. The map uses a color gradient to represent the levels of total consumption, with darker shades of red indicating higher consumption and lighter shades of blue indicating lower consumption..

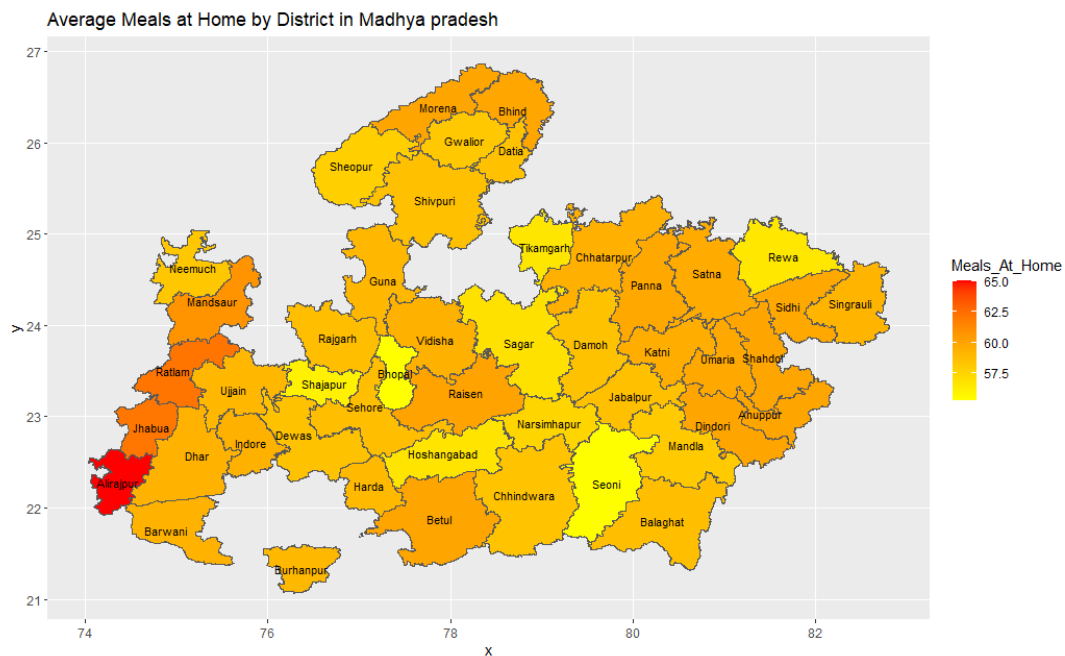
Key observations:

1. Highest Consumption:

- **Ujjain** is highlighted in the darkest red, indicating it has the highest total consumption.
- **Indore** also shows high consumption, represented by a dark red shade.

2. High Consumption Areas:

- Districts such as **Bhopal, Bina, Chhatarpur, Ratlam, and Sagar** are marked with darker shades of red, indicating high levels of consumption.
 - Central and western parts of Madhya Pradesh show higher consumption overall.
3. **Moderate Consumption Areas:**
- Districts like **Shivpuri, Morena, Raisen, Vidisha, and Tikamgarh** are in moderate shades of red and orange, showing mid-range consumption.
 - These districts are scattered across different parts of the state, indicating a more distributed pattern of consumption.
4. **Lower Consumption Areas:**
- Eastern districts like **Dindori, Mandla, Anuppur, and Balaghat** are shaded in blue, indicating lower total consumption.
 - Several districts in the northeastern and southern parts of Madhya Pradesh also show lower consumption levels.
5. **Geographical Patterns:**
- High consumption is concentrated in central and western regions.
 - Lower consumption is more prevalent in the eastern and southern regions.



1. **Color Scale:** The color scale on the right represents the average number of meals consumed at home. The scale ranges from a lighter yellow to dark red:
- **Light Yellow:** Represents districts where the average number of meals consumed at home is lower (around 57.5).

- **Orange to Dark Red:** Represents districts where the average number of meals consumed at home is higher (up to 65).
- 2. **Districts:** Each district in Madhya Pradesh is color-coded based on the average number of meals consumed at home.
 - **Alirajpur** and nearby districts in the southwest show the highest average number of meals at home, indicated by the darkest red color.
 - **Seoni** and **Hoshangabad** districts in the south-central part of the state show lower averages, indicated by the lighter yellow color.
- 3. **Geographical Distribution:** The map highlights regional differences:
 - **Southwest Districts:** Alirajpur, Jhabua, and Barwani have a higher average of meals at home.
 - **Central and Northern Districts:** Districts like Vidisha, Raisen, and Shivpuri have lower to moderate averages.
 - **Eastern Districts:** Rewa and surrounding areas show moderate averages.

This map can help identify areas where the average number of meals at home is higher or lower, which could be indicative of socio-economic factors, cultural practices, or other regional characteristics. For instance, areas with higher averages might have stronger home-cooking traditions or different economic conditions compared to areas with lower averages.