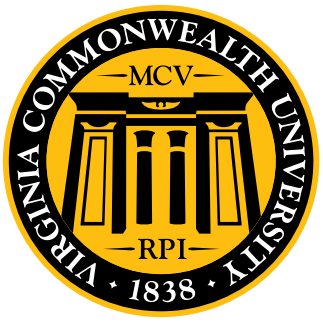


VIRGINIA COMMONWEALTH UNIVERSITY

Statistical analysis and modelling (SCMA 632)

A5.b: Visualisation – Perceptual Mapping for Business

JYOTHIS KANIYAMPARAMBIL THANKACHAN

V01110144

Date of Submission: 15-07-2024

CONTENTS

Introduction

Plotting the variable ‘nontotal_v’ on Karnataka state map

The study visually depicts the patterns of meat consumption in the state of Karnataka using data acquired from the National Sample Survey Office (NSSO). The aim is to create a graph of the variable 'nontotal_v,' which denotes the overall consumption of meat, on the map of Karnataka state using the NSSO68 dataset. This visualization seeks to illustrate the geographical dispersion and extent of meat consumption across different districts in Karnataka. The NSSO68 dataset provides comprehensive data on several consumption-related variables for both rural and urban sectors. Our goal is to identify and highlight places with both high and low levels of meat consumption by mapping the 'nontotal_v' variable. This analysis will provide valuable insights to policymakers and stakeholders, enabling them to acquire a more comprehensive understanding of regional consumption patterns in Karnataka. This would empower them to make educated decisions on targeted interventions and allocation of resources.

Objectives:

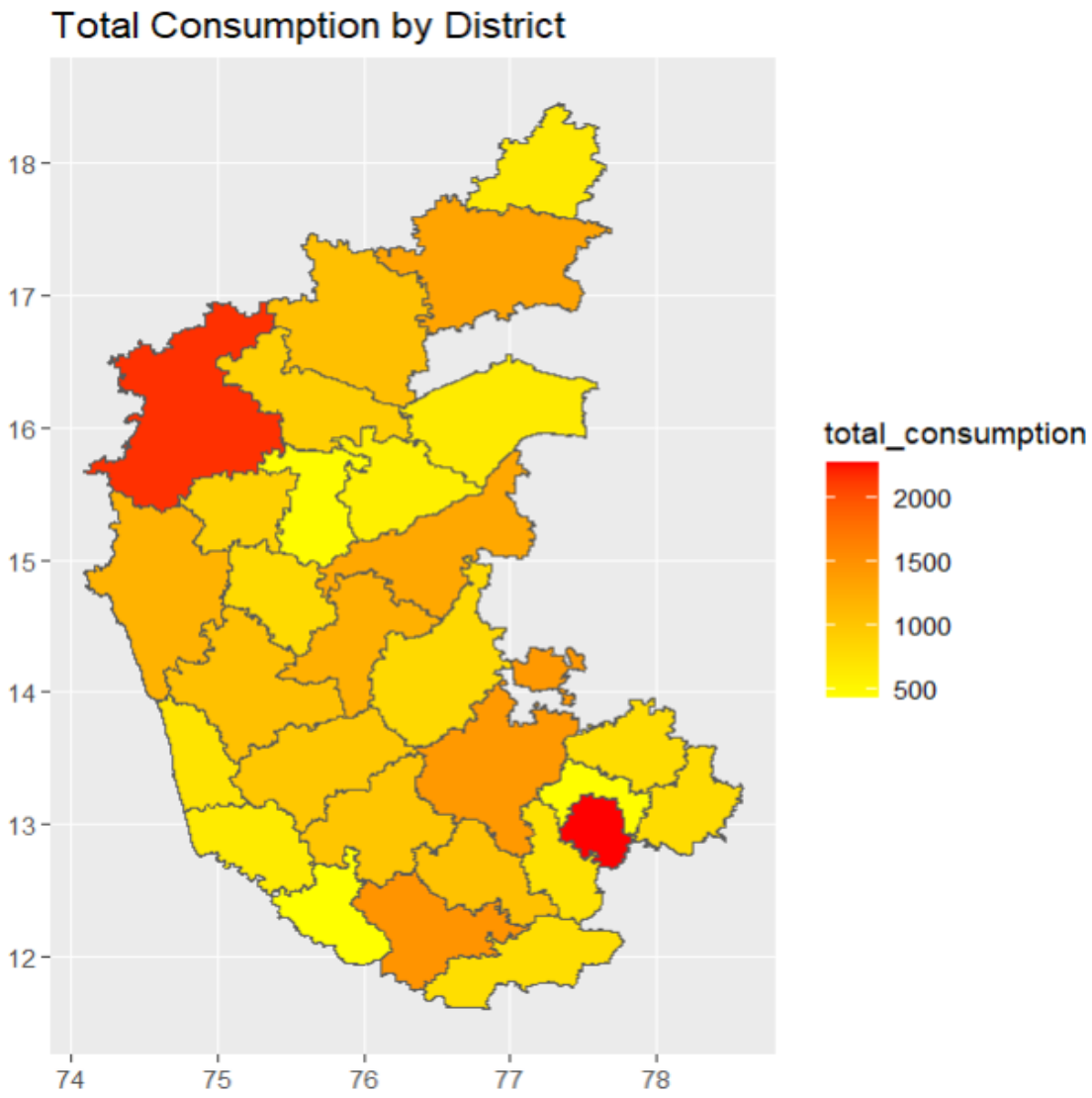
• Analyze and depict the patterns of meat consumption.
• Determine the specific patterns of consumption in different regions.

Business Significance: This has substantial business significance. It provides crucial data on consumption patterns in various regions, enabling companies to make strategic choices and deploy resources efficiently. Through the process of mapping meat consumption, companies can identify certain areas with high demand and improve their supply chains to achieve efficient distribution and reduce logistical costs. This data also enables market segmentation efforts, enabling businesses to tailor marketing strategies and product offerings based on district-specific preferences. Moreover, the map aids policymakers in assessing dietary trends and their impact on health, guiding initiatives to promote comprehensive nutrition and public health awareness. The geo map assignment enhances business planning by providing accurate information, enhances market competitiveness, and contributes to sustainable development and consumer welfare in Karnataka. Multidimensional Scaling (MDS) is a technique used to analyze and visualize the similarities or dissimilarities between objects or individuals.

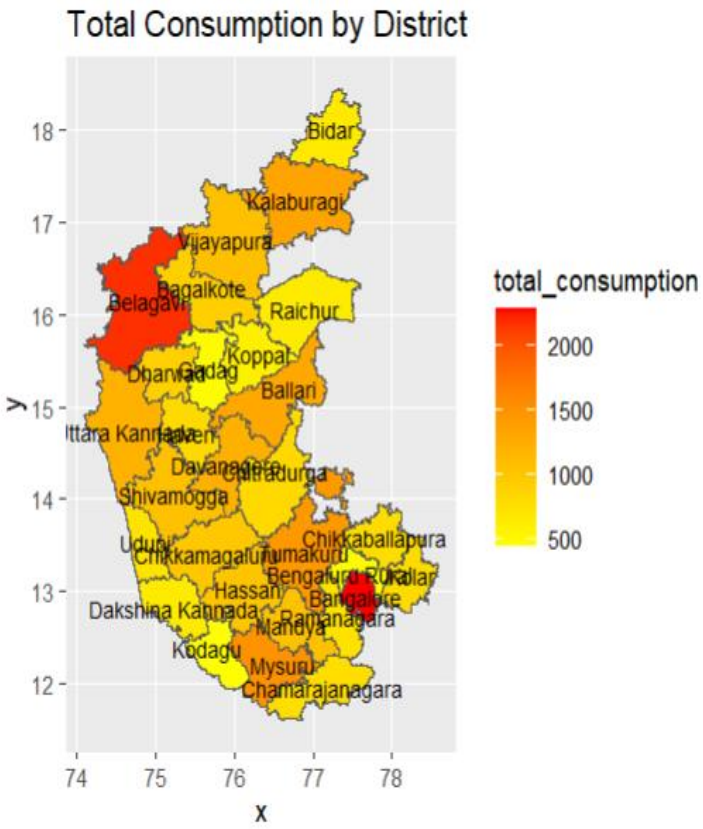
Results and Interpretation using R

Plot without labelling districts names

```
# Plot without labeling district names
ggplot(data_map_data) +
  geom_sf(aes(fill =total_consumption, geometry = geometry)) +
  scale_fill_gradient(low = "yellow", high = "red") +
  ggtitle("Total Consumption by District")
```



```r
# Plot with labelled district names
ggplot(data_map_data) +
  geom_sf(aes(fill = total_consumption, geometry = geometry)) +
  scale_fill_gradient(low = "yellow", high = "red") +
  ggtitle("Total Consumption by District") +
  geom_sf_text(aes(label = District, geometry = geometry), size = 3, color = "black")
```

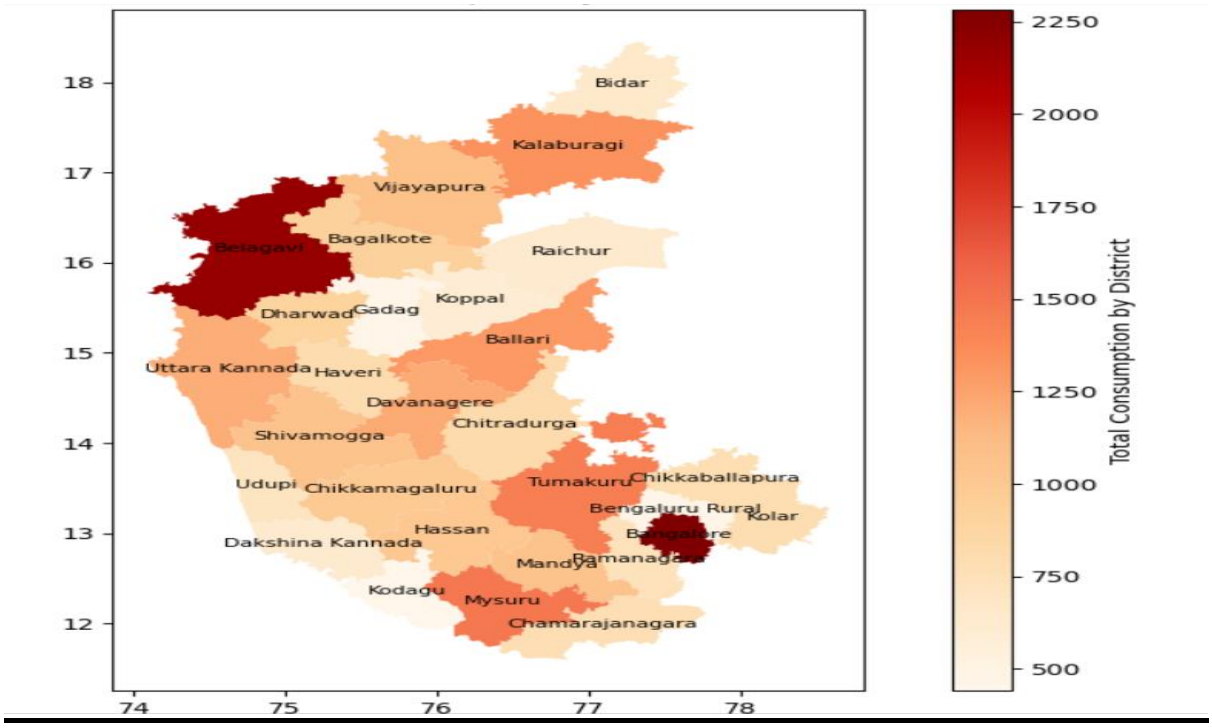


Interpretation:

The map displays the aggregate consumption each district in the state, using a color gradient from yellow to red to represent higher levels of consumption. The districts of Belagavi and Bengaluru Urban, marked in red, exhibit the highest overall consumption, surpassing 2000 units. Conversely, districts like as Kodagu and Chamarajanagara, which are highlighted in yellow, exhibit lower levels of consumption, approximately 500 units. This suggests notable geographical discrepancies in consumption patterns, with urban and densely populated regions displaying greater consumption in comparison to rural areas.

Results and Interpretation using Python

```
# Plot with labeled district names
fig, ax = plt.subplots(1, 1, figsize=(12, 8))
data_map_data.plot(column='total_consumption', cmap='OrRd', legend=True, ax=ax, legend_kwds={'label': "Total Consumption by District"})
data_map_data.apply(lambda x: ax.annotate(text=x['District'], xy=x.geometry.centroid.coords[0], ha='center', fontsize=8, color='black'), axis=1)
plt.title('Total Consumption by District')
plt.show()
```



Interpretation:

The map illustrates the aggregate consumption per district in Karnataka, India, where darker tones correspond to elevated consumption levels. Belagavi and Bengaluru are the districts that have the highest total consumption, as seen by the most intense red hues. Mysuru and Kalaburagi districts also exhibit comparatively high levels of consumption, albeit to a lesser degree. Conversely, districts such as Bidar, Koppal, and Chamarajanagara exhibit the least amount of overall consumption, as indicated by the palest hues on the map. This gradient visually illustrates the variation in consumption levels among different regions in Karnataka.

Recommendations

By analyzing the consumption patterns depicted on the map of Karnataka, it is possible to derive many suggestions to foster equitable and environmentally-friendly progress. In districts with high levels of consumption, such as Belagavi and Bengaluru, the main priority should be to enhance infrastructure and allocate resources more efficiently. Additionally, it is crucial to implement sustainability programs in order to effectively manage resources. To promote economic development and boost agriculture and rural entrepreneurship in low-consumption districts like Bidar, Koppal, and Chamarajanagara, it is advisable to make investments in relevant programs. Improve the infrastructure in rapidly expanding regions, promote economic diversification, and initiate public awareness campaigns regarding sustainable practices. Implement a surveillance and assessment framework to monitor consumption trends and the efficacy of policies, guaranteeing that decisions are based on data throughout the entire state.

R Codes

```
# Filtering for Karnataka

df_ka <- data %>%

  filter(state_1 == "KA")



# Sub-setting the data

ka_new <- df_ka %>%

  select(state_1, District, Region, Sector, State_Region, Meals_At_Home, ricepds_v, Wheatp
ds_q, chicken_q, pulsep_q, wheatos_q, No_of_Meals_per_day)



# Check for missing values in the subset

cat("Missing Values in Subset:\n")

print(colSums(is.na(ka_new)))



# Impute missing values with mean for specific columns

ka_new$Meals_At_Home <- impute_with_mean(ka_new$Meals_At_Home)



# Check for missing values after imputation

cat("Missing Values After Imputation:\n")

print(colSums(is.na(ka_new)))



# Finding outliers and removing them
```

```
outlier_columns <- c("ricepds_v", "chicken_q")

for (col in outlier_columns) {

  ka_new <- remove_outliers(ka_new, col)

}


# Summarize consumption

ka_new$total_consumption <- rowSums(ka_new[, c("ricepds_v", "Wheatpds_q", "chicken_q
", "pulsep_q", "wheatos_q")], na.rm = TRUE)


district_summary <- summarize_consumption("District")

cat("District Consumption Summary:\n")

print(district_summary)


# mapping districts so that meging of the tables will be easier

district_mapping <- c(

  "1" = "Belagavi",

  "2" = "Bagalkote",

  "3" = "Vijayapura",

  "4" = "Kalaburagi",

  "5" = "Bidar",

  "6" = "Raichur",

  "7" = "Koppal",
```

"8" = "Gadag",

"9" = "Dharwad",

"10" = "Uttara Kannada",

"11" = "Haveri",

"12" = "Ballari",

"13" = "Chitradurga",

"14" = "Davanagere",

"15" = "Shivamogga",

"16" = "Udupi",

"17" = "Chikkamagaluru",

"18" = "Tumakuru",

"19" = "Kolar",

"20" = "Bangalore",

"21" = "Bengaluru Rural",

"22" = "Mandya",

"23" = "Hassan",

"24" = "Dakshina Kannada",

"25" = "Kodagu",

"26" = "Mysuru",

"27" = "Chamarajanagara",

"28" = "Ramanagara",

```
  "29" = "Chikkaballapura"
)

ka_new$District <- as.character(ka_new$District)
ka_new$District <- district_mapping[ka_new$District]
#ka_new$District <- ifelse(ka_new$District %in% names(district_mapping), district_mapping[ka_new$District], ka_new$District)
View(ka_new)

# ka_consumption stores aggregate of total consumption district wise
ka_consumption <- aggregate(total_consumption ~ District, data = ka_new, sum)
View(ka_consumption)

#Plotting total consumption on the Karnataka state

Sys.setenv("SHkaE_RESTORE_SHX" = "YES")

data_map <- st_read("D:\\Assignments_SCMA632\\KARNATAKA_DISTRICTS.geojson")
View(data_map)

data_map <- data_map %>%
  rename(District = dtname)
```

```
# merging ka_consumption and data_map tables

data_map_data <- merge(ka_consumption,data_map,by = "District")

View(data_map_data)


# Plot without labeling district names

ggplot(data_map_data) +

  geom_sf(aes(fill =total_consumption, geometry = geometry)) +

  scale_fill_gradient(low = "yellow", high = "red") +

  ggtitle("Total Consumption by District")


# Plot with labelled district names

ggplot(data_map_data) +

  geom_sf(aes(fill = total_consumption, geometry = geometry)) +

  scale_fill_gradient(low = "yellow", high = "red") +

  ggtitle("Total Consumption by District") +

  geom_sf_text(aes(label = District, geometry = geometry), size = 3, color = "black")
```

Python Codes

```
# Filtering for Karnataka
df_ka = data[data['state_1'] == "KA"]

# Sub-setting the data
ka_new = df_ka[['state_1', 'District', 'Region', 'Sector', 'State_Region', 'Meals_At_Home', 'ricepds_v',
'Wheatpds_q', 'chicken_q', 'pulsep_q', 'wheatos_q', 'No_of_Meals_per_day']]
```

```
# Check for missing values in the subset
print("Missing Values in Subset:")
print(ka_new.isnull().sum())

# Impute missing values with mean for specific columns
ka_new['Meals_At_Home'].fillna(ka_new['Meals_At_Home'].mean(), inplace=True)

# Check for missing values after imputation
print("Missing Values After Imputation:")
print(ka_new.isnull().sum())

# Remove outliers
for col in outlier_columns:
    ka_new = remove_outliers(ka_new, col)
```

In [77]:

```
# Summarize consumption
ka_new['total_consumption'] = ka_new[['ricepds_v', 'Wheatpds_q', 'chicken_q', 'pulsep_q',
'wheatos_q']].sum(axis=1)

district_summary =
ka_new.groupby('District')['total_consumption'].sum().reset_index().sort_values(by='total_consumption',
ascending=False)
print("District Consumption Summary:")
print(district_summary)

# Mapping districts so that merging of the tables will be easier
district_mapping = {
    "1": "Belagavi", "2": "Bagalkote", "3": "Vijayapura", "4": "Kalaburagi", "5": "Bidar",
    "6": "Raichur", "7": "Koppal", "8": "Gadag", "9": "Dharwad", "10": "Uttara Kannada",
    "11": "Haveri", "12": "Ballari", "13": "Chitradurga", "14": "Davanagere", "15": "Shivamogga",
    "16": "Udupi", "17": "Chikkamagaluru", "18": "Tumakuru", "19": "Kolar", "20": "Bangalore",
    "21": "Bengaluru Rural", "22": "Mandya", "23": "Hassan", "24": "Dakshina Kannada",
    "25": "Kodagu", "26": "Mysuru", "27": "Chamarajanagara", "28": "Ramanagara", "29": "Chikkaballapura"
}

ka_new['District'] = ka_new['District'].astype(str).map(district_mapping).fillna(ka_new['District'])
print(ka_new)

# ka_consumption stores aggregate of total consumption district-wise
ka_consumption = ka_new.groupby('District')['total_consumption'].sum().reset_index()
print(ka_consumption)

# Load and plot Karnataka state map
data_map = gpd.read_file('D:\\Assignments_SCMA632\\KARNATAKA_DISTRICTS.geojson')

data_map = data_map.rename(columns={'dtname': 'District'})
print(data_map)

# Merging ka_consumption and data_map tables
data_map_data = data_map.merge(ka_consumption, on='District')
print(data_map_data)

# Plot with labeled district names
fig, ax = plt.subplots(1, 1, figsize=(12, 8))
data_map_data.plot(column='total_consumption', cmap='OrRd', legend=True, ax=ax,
legend_kwds={'label': "Total Consumption by District"})
data_map_data.apply(lambda x: ax.annotate(text=x['District'], xy=x.geometry.centroid.coords[0],
ha='center', fontsize=8, color='black'), axis=1)
plt.title('Total Consumption by District')  plt.show()
```

References

1. www.github.com