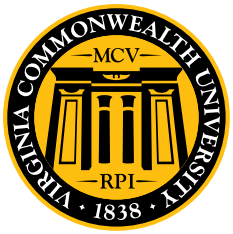


VIRGINIA COMMONWEALTH UNIVERSITY

Statistical analysis and modelling (SCMA 632)

A6: Visualization - Perceptual Mapping for Business

SHRINITHA SURESH KUMAR
V01151332

Date of Submission: 09-07-2025

Introduction

Food consumption patterns vary widely across regions and are influenced by factors such as income levels, cultural preferences, and access to resources. Understanding these patterns is crucial for policymakers, businesses, and researchers to make informed decisions about resource allocation, marketing strategies, and nutrition programs.

This assignment focuses on visualizing district-wise food consumption data in Tamil Nadu using statistical and spatial tools. By leveraging descriptive plots and geographic maps, we aim to gain deeper insights into the distribution and regional variations in food expenditure. Such visualizations not only help in identifying disparities but also enable targeted interventions and business strategies.

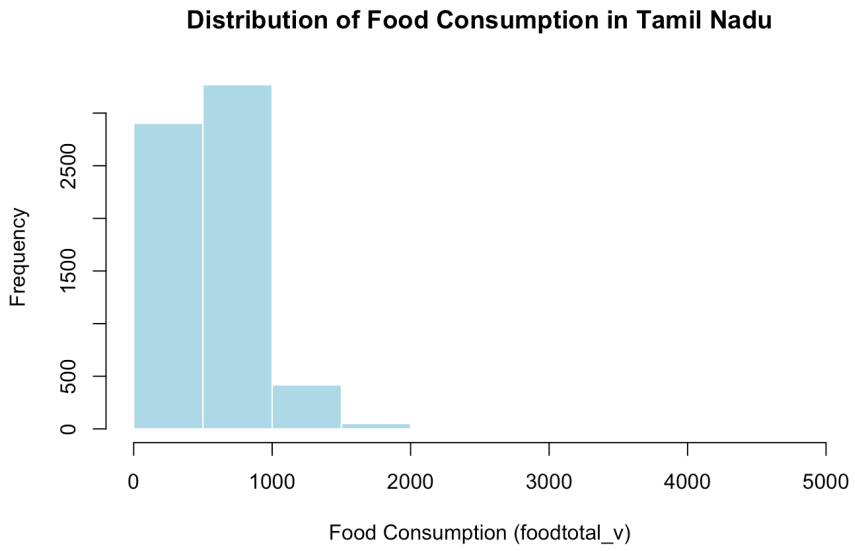
Objectives

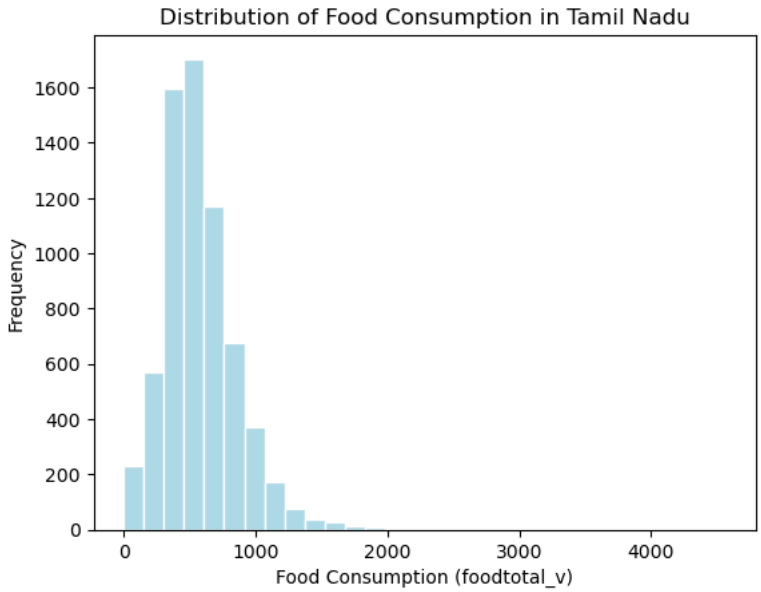
- To analyze and understand the distribution of household food consumption across districts in Tamil Nadu.
- To identify districts with higher and lower average food consumption levels.
- To visualize regional disparities using statistical plots (histogram and bar plot) and spatial representations (choropleth map).
- To derive actionable insights that can support policy formulation, supply chain planning, and localized marketing decisions.

Results and Interpretations

1. **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 Assignment A1 to indicate the consumption district-wise for the state assigned to you.**

Histogram:



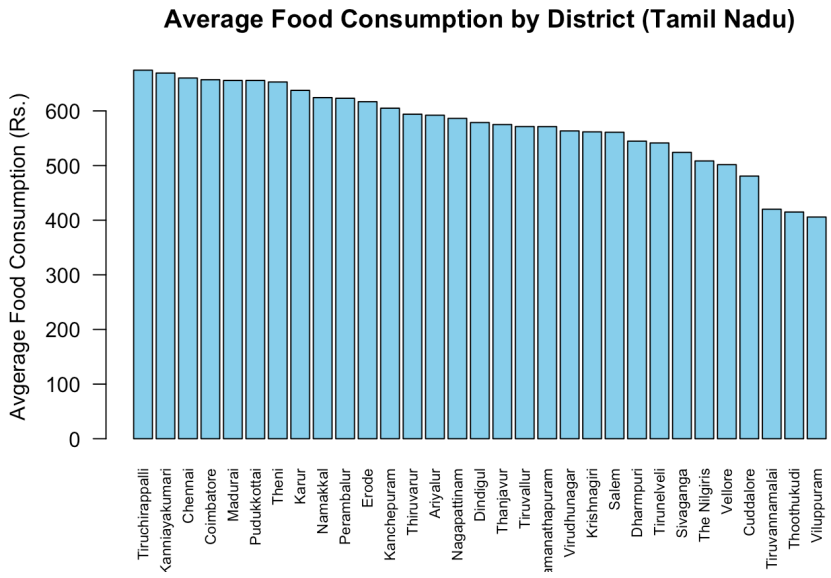


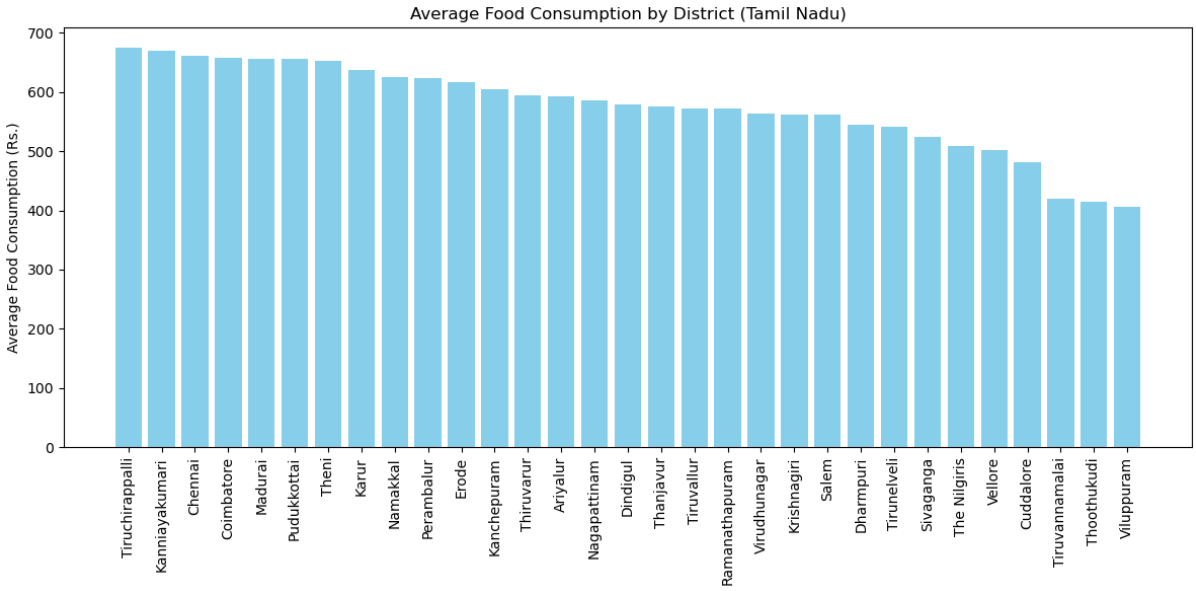
Interpretation:

The histogram provides an overview of how food consumption (measured as `foodtotal_v`) is distributed among households across Tamil Nadu. The shape of the histogram can help identify whether the consumption is skewed (e.g., toward lower or higher spending), whether there are any outliers, and the spread of the data.

In this case, most households tend to cluster around a certain consumption level, with fewer households reporting very high or very low consumption. A right-skewed distribution would suggest that while most households spend moderately, a few spend significantly more.

Barplot:





Interpretation:

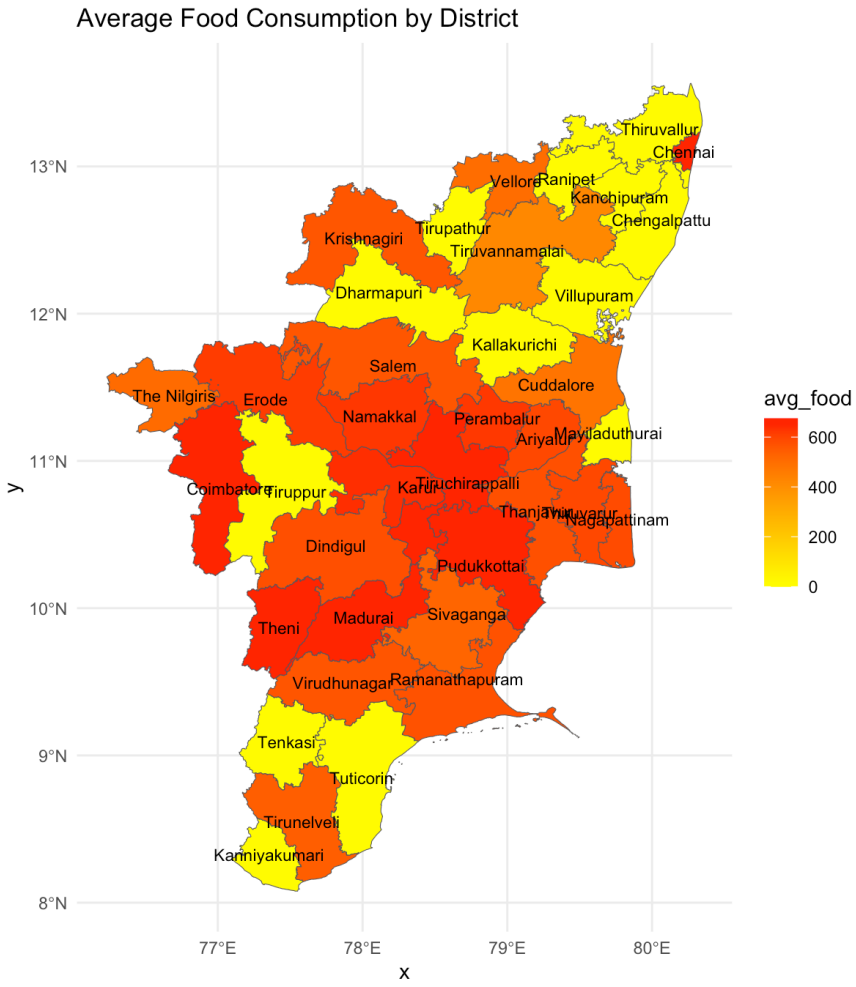
The bar plot compares the average food consumption across different districts in Tamil Nadu. By visualizing district-level differences, it becomes clear which districts have higher or lower spending on food.

Districts with higher average consumption may reflect higher income levels, better access to food markets, or different consumption patterns. Conversely, lower averages could indicate relatively lower income or different dietary habits.

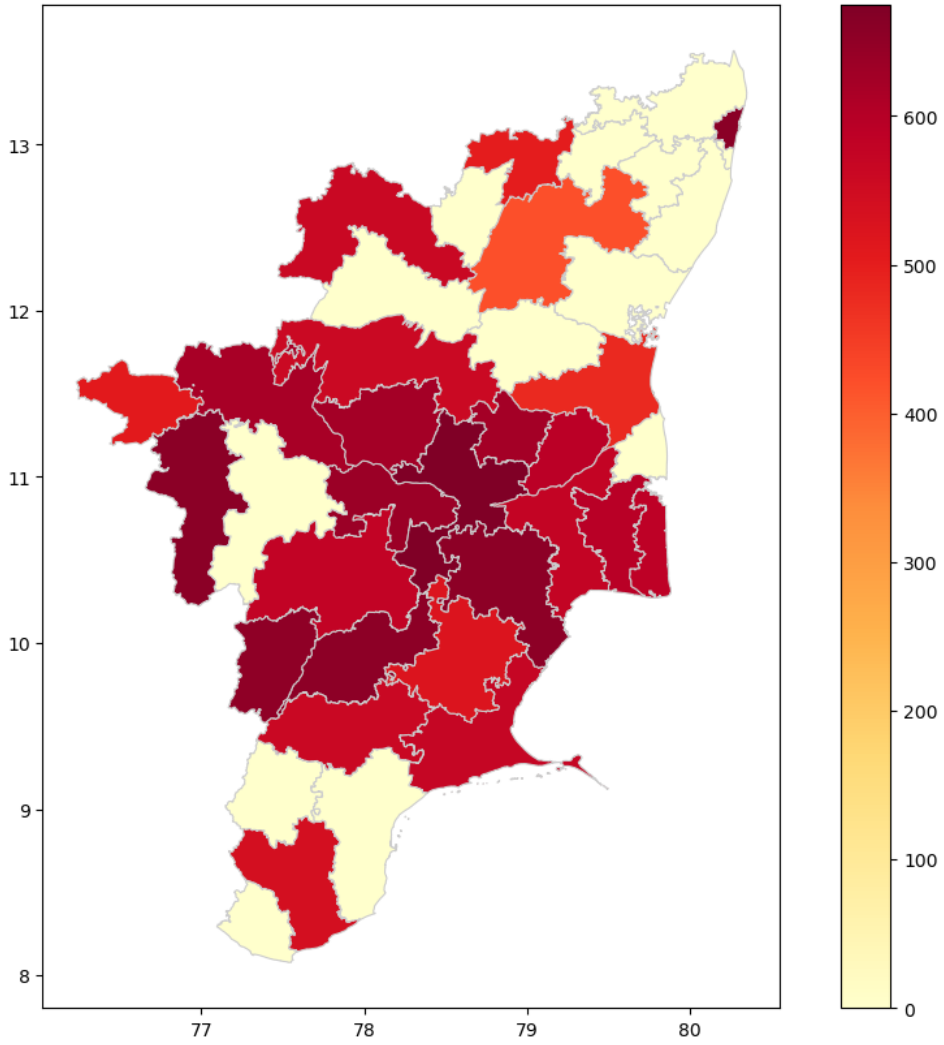
This visualization supports regional analysis and can guide policymakers or businesses in targeting nutrition programs, food retail expansion, or marketing strategies more effectively.

2. Plot {'any variable of your choice'} on the Karnataka (or the state assigned to you) state map using NSSO68.csv data

R:


Average Food Consumption by District
Thiruvallur
Chennai
Vellore
Ranipet
Kanchipuram
Chengalpattu
Krishnagiri
Tirupathur
Tiruvannamalai
Dharmapuri
Villupuram
Kallakurichi
Salem
Cuddalore
The Nilgiris
Erode
Namakkal
Perambalur
Ariyalur
Mayiladuthurai
Coimbatore
Tiruppur
Karur
Tiruchirappalli
Thanjavur
Tiruvarur
Nagapattinam
Dindigul
Pudukkottai
Theni
Madurai
Sivaganga
Virudhunagar
Ramanathapuram
Tenkasi
Tuticorin
Tirunelveli
Kanniyakumari
13°N
12°N
11°N
10°N
9°N
8°N
77°E
78°E
79°E
80°E
x
y
avg_food
600
400
200
0

Python:



The map offers a geographic perspective by showing how average food consumption varies across Tamil Nadu districts on a map. Districts are shaded according to their average consumption values, with higher values typically shown in darker or more intense colors.

This spatial view allows us to easily identify regional patterns. For example, coastal or urban districts might show higher consumption levels, while more rural or interior districts might show lower averages.

Such mapping is crucial for visual storytelling and for understanding regional inequalities or opportunities. It helps in designing location-specific policies, resource allocation, or business expansion plans.

PYTHON CODE:

```
pip install geopandas

#------------------------------#
# Step 1: Load and inspect data
#------------------------------#

import pandas as pd
import matplotlib.pyplot as plt
import geopandas as gpd

# Set working directory and read the data
df = pd.read_csv('/Users/shrinithask/Desktop/VCU/Stastical
analysis/Assignments/Data/NSSO68.csv')

# Inspect column names
print(df.columns)

# Check the 'foodtotal_v' column
print(df['foodtotal_v'].head())

#--------------------------------------#
# Step 2: Filter the data for Tamil Nadu
#--------------------------------------#

# Check how many tamilnadu rows are there
print((df['state_1'] == 'TN').sum())

# Filter rows where state_1 is 'TN'
tn = df[df['state_1'] == 'TN'].copy()
print(tn.shape)

# Histogram of food consumption in tamil nadu
plt.hist(tn['foodtotal_v'].dropna(), bins=30, color='lightblue', edgecolor='white')
plt.title("Distribution of Food Consumption in Tamil Nadu")
plt.xlabel("Food Consumption (foodtotal_v)")
plt.ylabel("Frequency")
plt.show()

#-------------------------------------------#
# Step 3: Group-wise summary at District level
#-------------------------------------------#
```

```
# Compute district-wise average food consumption
tn['District'] = tn['District'].astype(str).str.zfill(2)  # Pad district numbers to 2 digits
tn['DWCons'] = tn.groupby('District')['foodtotal_v'].transform('mean')

#----------------------------------------------------------#
# Step 4: Create mapping of District Codes to Names
#----------------------------------------------------------#

# The error is that we have 29 district codes (range 1-30) but 31 district names
# Solution: Make sure both lists have the same length by adjusting the range

district_map = pd.DataFrame({
   'DistrictCode': [f"{i:02d}" for i in range(1, 32)],  # Changed to range(1, 32) to create 31
codes
   'DistrictName': [
      "Tiruvallur", "Chennai", "Kanchepuram", "Vellore", "Dharmpuri", "Tiruvannamalai",
      "Viluppuram", "Salem", "Namakkal", "Erode", "The Nilgiris", "Coimbatore",
      "Dindigul", "Karur", "Tiruchirappalli", "Perambalur", "Ariyalur", "Cuddalore",
      "Nagapattinam", "Thiruvarur", "Thanjavur", "Pudukkottai", "Sivaganga", "Madurai",
      "Theni", "Virudhunagar", "Ramanathapuram", "Thoothukudi", "Tirunelveli",
      "Kanniayakumari", "Krishnagiri"
   ]
})

#----------------------------------------------------------#
# Step 5: Merge mapping into main data using District code
#----------------------------------------------------------#

# Create a DistrictCode column from District number
tn['DistrictCode'] = tn['District']

# Merge to get District names
tn = tn.merge(district_map, on='DistrictCode', how='left')

#-----------------------------------------------#
# Step 6: Summarize and Plot Bar Chart
#-----------------------------------------------#

# Create summary table: average food consumption by district
district_avg = tn.groupby('DistrictName')['foodtotal_v'].mean().reset_index()
district_avg = district_avg.sort_values(by='foodtotal_v', ascending=False)

# Barplot: average food consumption by district
plt.figure(figsize=(12, 6))
plt.bar(district_avg['DistrictName'], district_avg['foodtotal_v'], color='skyblue')
```

```
plt.xticks(rotation=90)
plt.title("Average Food Consumption by District (Tamil Nadu)")
plt.ylabel("Average Food Consumption (Rs.)")
plt.tight_layout()
plt.show()

#-----------------------------------------------#
# Step 7: Choropleth Map using GeoPandas
#-----------------------------------------------#

# Read GeoJSON file
data_map = gpd.read_file("/Users/shrinithask/Desktop/VCU/Stastical
analysis/Assignments/Data/TAMIL NADU_DISTRICTS.geojson")

# Check and rename the district column
data_map = data_map.rename(columns={'dtname': 'DistrictName'})

# Merge geo data with average food data
data_map_data = data_map.merge(district_avg, on='DistrictName', how='left')

# Fill missing values with 0
data_map_data['foodtotal_v'] = data_map_data['foodtotal_v'].fillna(0)

# Plot choropleth map
fig, ax = plt.subplots(1, 1, figsize=(12, 10))
data_map_data.plot(column='foodtotal_v',
                   cmap='YlOrRd',
                   linewidth=0.8,
                   ax=ax,
                   edgecolor='0.8',
                   legend=True)
```

R CODE:

```
#-------------------------------#
# Step 1: Load and inspect data ----
#-------------------------------#

# Set working directory to the location of your dataset
setwd('/Users/shrinithask/Desktop/VCU/Stastical analysis/Assignments/Data')

# Read the NSSO68 dataset
df <- read.csv('NSSO68.csv')
```

```
# Inspect column names
names(df)

# Check the 'foodtotal_v' column
head(df$foodtotal_v)



#---------------------------------------#
# Step 2: Filter the data for Karnataka ----
#---------------------------------------#

# Check how many Karnataka rows are there
sum(df$state_1 == 'TN')

# Filter rows where state_1 is 'TN'
tn <- df[df$state_1 == 'TN', ]
dim(tn)

# Histogram of food consumption in Karnataka
hist(ka$foodtotal_v,
     main = "Distribution of Food Consumption in Tamil Nadu",
     xlab = "Food Consumption (foodtotal_v)",
     col = "lightblue",
     border = "white")



#-------------------------------------------#
# Step 3: Group-wise summary at District level ----
#-------------------------------------------#

# Convert District column to factor (if not already)
tn$District <- as.factor(tn$District)

# Load dplyr for grouping
library(dplyr)

# Add district-wise average food consumption column
tn <- tn %>%
```

```
  group_by(District) %>%
  mutate(DWCons = mean(foodtotal_v, na.rm = TRUE)) %>%
  ungroup()


#-------------------------------------------------------#
# Step 4: Create mapping of District Codes to Names ----
#-------------------------------------------------------#

district_map <- data.frame(
  DistrictCode = sprintf("%02d", 1:31),  # Format as 01, 02, ..., 31
  DistrictName = c("Tiruvallur", "Chennai", "Kanchepuram", "Vellore",
"Dharmpuri", "Tiruvannamalai",
                   "Viluppuram", "Salem", "Namakkal", "Erode", "The Nilgiris",
"Coimbatore",
                   "Dindigul", "Karur", "Tiruchirappalli", "Perambalur", "Ariyalur",
"Cuddalore",
                   "Nagapattinam", "Thiruvarur", "Thanjavur", "Pudukkottai",
"Sivaganga", "Madurai",
                   "Theni", "Virudhunagar", "Ramanathapuram", "Thoothukudi",
"Tirunelveli",
                   "Kanniayakumari", "Krishnagiri"),
  stringsAsFactors = FALSE
)


#-------------------------------------------------------#
# Step 5: Merge mapping into main data using District code ----
#-------------------------------------------------------#

# Create a DistrictCode column from District number
tn <- tn %>%
  mutate(DistrictCode = sprintf("%02d", as.numeric(District)))  # Converts 1 to
'01', etc.

# Merge to get District names
tn <- tn %>%
  left_join(district_map, by = "DistrictCode")
```

```
#-----------------------------------------------#
# Step 6: Summarize and Plot Bar Chart ----
#-----------------------------------------------#

# Create summary table: average food consumption by district
district_avg <- tn %>%
  group_by(DistrictName) %>%
  summarise(avg_food = mean(foodtotal_v, na.rm = TRUE)) %>%
  arrange(desc(avg_food))  # Sort by consumption

# Barplot: average food consumption by district
barplot(height = district_avg$avg_food,
        names.arg = district_avg$DistrictName,
        las = 2,                # Rotate x-axis labels vertically
        col = "skyblue",
        main = "Average Food Consumption by District (Tamil Nadu)",
        ylab = "Avgerage Food Consumption (Rs.)",
        cex.names = 0.7)        # Adjust label size if too crowded


# Choropleth Maps
# Plot data on the map itself

# a variable of our choice
# geojson file or the shapefile

install.packages("sf")
library(ggplot2)
library(sf) # mapping
library(dplyr)

#Sys.setenv("SHAPE_RESTORE_SHX" = "YES")

data_map <- st_read("TAMIL NADU_DISTRICTS.geojson")
View(data_map)


# Step 1: Ensure district name column matches in both datasets
```

```
data_map <- data_map %>%
  rename(DistrictName = dtname)  # Rename if needed

# Step 2: Left join spatial data with data values
data_map_data <- data_map %>%
  left_join(district_avg, by = "DistrictName")  # Keeps all districts

# Step 3: Replace NA with 0 for missing data
data_map_data$avg_food[is.na(data_map_data$avg_food)] <- 0

# Step 4: Plot using ggplot2
ggplot(data_map_data) +
  geom_sf(aes(fill = avg_food, geometry = geometry)) +
  scale_fill_gradient(low = "yellow", high = "red") +
  ggtitle("Average Food Consumption by District") +
  theme_minimal() +
  geom_sf_text(aes(label = DistrictName), size = 3, color = "black")
```