

**VIRGINIA COMMONWEALTH UNIVERSITY**

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

**A5: Visualization - Perceptual Mapping for Business**

**NITESH REDDY**

**V01106546**

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

**INTRODUCTION**

**Analysis of Consumption Data in Haryana**

To analyze the consumption data from the NSSO, focusing on the top and bottom three consuming districts in Haryana, we'll clean and adjust the dataset. The data includes district-specific variances and consumption-related figures for both urban and rural areas. We'll be using R/Python for its robust capabilities in handling and analyzing large datasets.

**Goals:**

* **Summarize consumption statistics** by area and district.
* **Manage outliers** and address missing values.
* **Assess the significance of mean variances**.
* **Standardize district and sector names** for consistency.

**Objectives:**

1. **Plot a histogram** to show the distribution of total consumption across different districts.
2. **Create a bar plot** to visualize consumption per district, with district names clearly labelled, to highlight the consumption patterns district-wise in Haryana.

These analyses will provide valuable insights for policymakers and stakeholders, facilitating targeted initiatives and promoting balanced development across the state.

**RESULTS & INTERPRETATION**

**a) R**

**Code:**

library(ggplot2)

library(sf) # mapping

library(dplyr)

Sys.setenv("SHAPE\_RESTORE\_SHX" = "YES")

data\_map <- st\_read("D:\VCU Bootcamp\SCMA\Data\HARYANA\_DISTRICTS.geojson")

View(data\_map)

data\_map <- data\_map %>%

rename(District = dtname)

colnames(data\_map)

data\_map\_data <- merge(HR\_consumption,data\_map,by = "District")

View(data\_map\_data)

ggplot(data\_map\_data) +

geom\_sf(aes(fill =total\_consumption, geometry = geometry)) +

scale\_fill\_gradient(low = "yellow", high = "red") +

ggtitle("Total Consumption\_by\_District")

ggplot(data\_map\_data) +

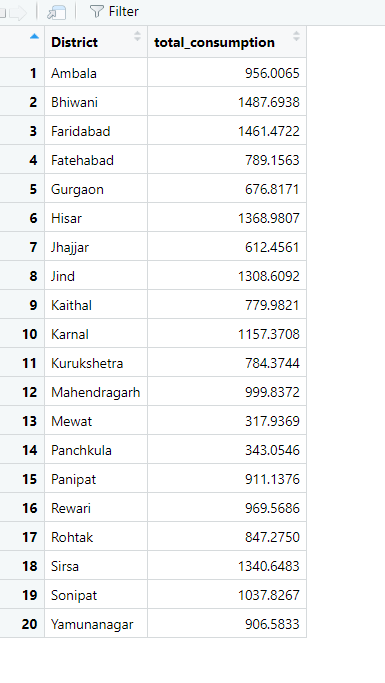
geom\_sf(aes(fill = total\_consumption, geometry = geometry)) +

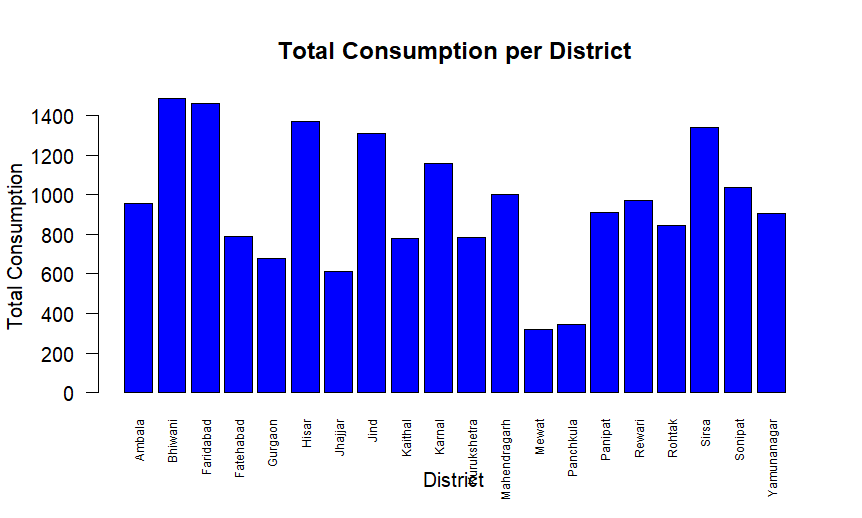
scale\_fill\_gradient(low = "yellow", high = "red") +

ggtitle("Total Consumption by District") +

geom\_sf\_text(aes(label = District, geometry = geometry), size = 3, color = "black")

**Result:**





**Interpretation:**

From the above dataset, we can see the districts with the highest total consumption are Faridabad, Bhiwani, and Sonipat. The districts with the lowest total consumption are Mewat, Mahendrakar, and Panchkula.

**B) Python**

**Code:**

hrnew['total\_consumption'] = hrnew[['ricepds\_v', 'Wheatpds\_q', 'chicken\_q', 'pulsep\_q', 'wheatos\_q']].sum(axis=1)

def summarize\_consumption(df, group\_col):

summary = df.groupby(group\_col)['total\_consumption'].sum().reset\_index().sort\_values(by='total\_consumption', ascending=False)

return summary

district\_summary = summarize\_consumption(hrnew, 'District')

region\_summary = summarize\_consumption(hrnew, 'Region')

print("Top Consuming Districts:")

print(district\_summary.head(4))

print("Region Consumption Summary:")

print(region\_summary)

Result:

Top Consuming Districts:

District total\_consumption

12 13 1487.693831

18 19 1461.472232

11 12 1368.980678

10 11 1340.648296

Region Consumption Summary:

Region total\_consumption

0 1 10792.293321

1 2 8264.494089

Code:

hrnew['District'] = hrnew['District'].astype(str).map(district\_mapping).fillna(hrnew['District'])

hrnew['Sector'] = hrnew['Sector'].astype(str).map(sector\_mapping).fillna(hrnew['Sector'])

plt.hist(hrnew['total\_consumption'], bins=10, color='blue', edgecolor='black')

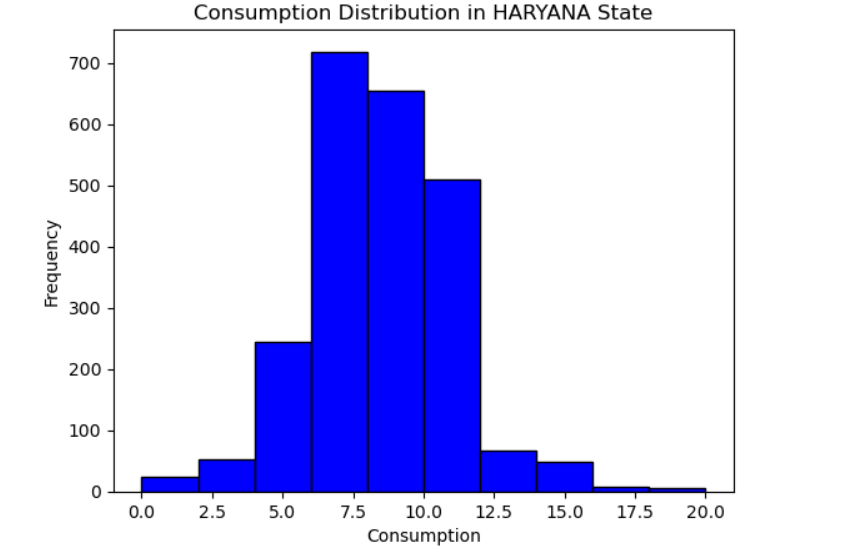
plt.xlabel('Consumption')

plt.ylabel('Frequency')

plt.title('Consumption Distribution in HARYANA State')

plt.show()

Result:



Code:

HR\_consumption = hrnew.groupby('District')['total\_consumption'].sum().reset\_index()

plt.figure(figsize=(10, 6))

plt.bar(HR\_consumption['District'], HR\_consumption['total\_consumption'], color='blue', edgecolor='black')

plt.xlabel('District')

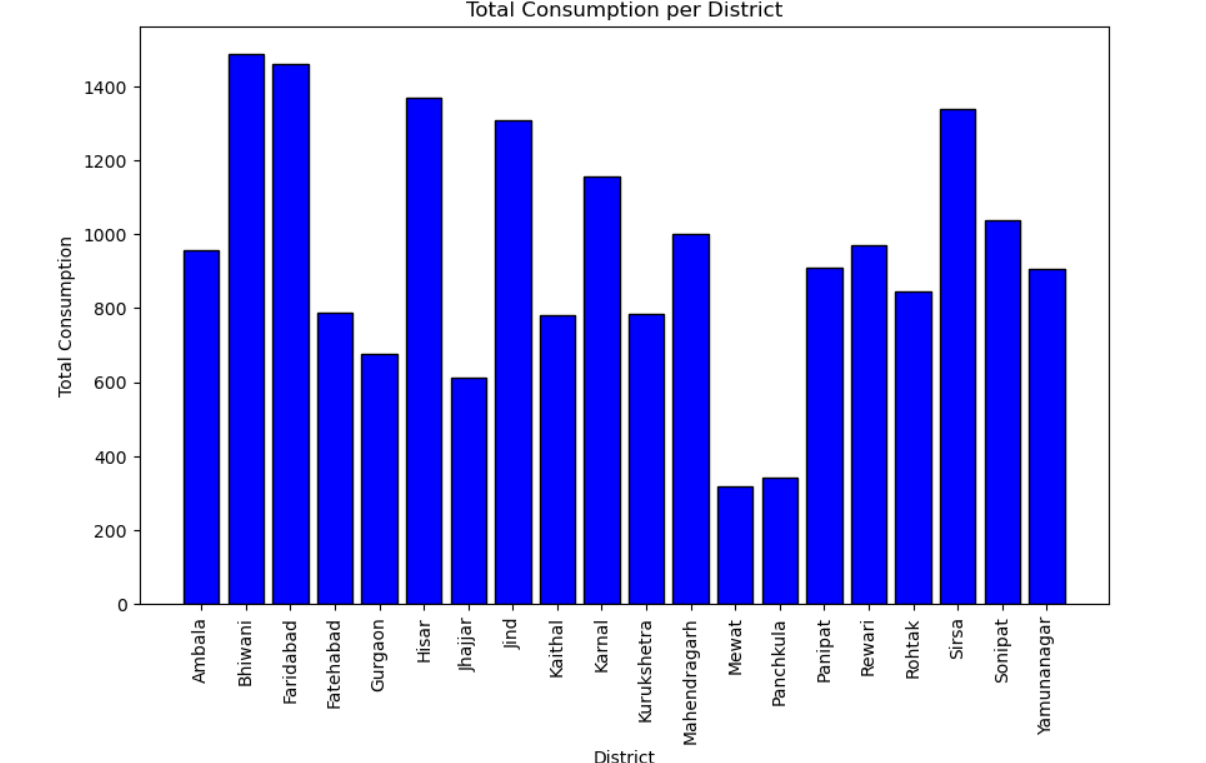
plt.ylabel('Total Consumption')

plt.title('Total Consumption per District')

plt.xticks(rotation=90)

plt.show()

Result:



Code:

fig, ax = plt.subplots(1, 1, figsize=(10, 10))

data\_map\_data.plot(column='total\_consumption', ax=ax, legend=True, legend\_kwds={'label': "Total Consumption by District",

'orientation': "horizontal"})

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()

Result:

