

```

# Set the working directory and verify it
setwd('C:\\Users\\anjel\\Downloads\\SCMA')
getwd()

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

# Load required libraries
libraries <- c("dplyr", "readr", "readxl", "tidyr", "ggplot2", "BSDA", "sf")
lapply(libraries, install_and_load)

# Reading the file into R
data <- read.csv("NSS068.csv")

# Filtering for MH
df <- data %>%
  filter(state_1 == "MH")

# Display dataset info
cat("Dataset Information:\n")
print(names(df))
print(head(df))
print(dim(df))

# Finding missing values
missing_info <- colSums(is.na(df))
cat("Missing Values Information:\n")
print(missing_info)

# Subsetting the data
mhnew <- df %>%
  select(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)

# Finding outliers and removing them
remove_outliers <- function(df, column_name) {
  Q1 <- quantile(df[[column_name]], 0.25, na.rm = TRUE)
  Q3 <- quantile(df[[column_name]], 0.75, na.rm = TRUE)
  IQR <- Q3 - Q1
  lower_threshold <- Q1 - (1.5 * IQR)
  upper_threshold <- Q3 + (1.5 * IQR)
  df <- subset(df, df[[column_name]] >= lower_threshold & df[[column_name]] <=
upper_threshold)
  return(df)
}

```

```

outlier_columns <- c("ricepds_v", "chicken_q")
for (col in outlier_columns) {
  mhnew <- remove_outliers(mhnew, col)
}

# Summarize consumption
mhnew$total_consumption <- rowSums(mhnew[, c("ricepds_v", "wheatpds_q",
"chicken_q", "pulsep_q", "wheatos_q")], na.rm = TRUE)

# Summarize and display top consuming districts and regions
summarize_consumption <- function(group_col) {
  summary <- mhnew %>%
    group_by(across(all_of(group_col))) %>%
    summarise(total = sum(total_consumption)) %>%
    arrange(desc(total))
  return(summary)
}

district_summary <- summarize_consumption("District")
region_summary <- summarize_consumption("Region")

cat("Top Consuming Districts:\n")
print(head(district_summary, 6))
cat("Region Consumption Summary:\n")
print(region_summary)

# Rename districts and sectors
district_mapping <- c("1" = "Nandurbar", "2" = "Dhule", "3" = "Jalgaon", "4" =
"Buldana", "5" = "Akola", "6" = "Washim", "7" = "Amravati",
  "8" = "Wardha", "9" = "Nagpur", "10" = "Bhandara", "11" =
"Gondiya", "12" = "Gadchiroli", "13" = "Chandrapur",
  "14" = "Yavatmal", "15" = "Nanded", "16" = "Hingoli", "17" =
"Parbhani", "18" = "Jalna", "19" = "Aurangabad",
  "20" = "Nashik", "21" = "Thane", "22" = "Mumbai (Suburban)",
"24" = "Raigarh", "25" = "Pune", "26" = "Ahmadnagar",
  "27" = "Bid", "28" = "Latur", "29" = "Osmanabad", "30" =
"Solapur", "31" = "Satara", "32" = "Ratnagiri",
  "33" = "Sindhudurg", "34" = "Kolhapur", "35" = "Sangli")
sector_mapping <- c("2" = "URBAN", "1" = "RURAL")

mhnew$District <- as.character(mhnew$District)
mhnew$Sector <- as.character(mhnew$Sector)
mhnew$District <- ifelse(mhnew$District %in% names(district_mapping),
district_mapping[mhnew$District], mhnew$District)
mhnew$Sector <- ifelse(mhnew$Sector %in% names(sector_mapping),
sector_mapping[mhnew$Sector], mhnew$Sector)

View (mhnew)

```

```

# Aggregate total consumption per district
MH_consumption <- aggregate(total_consumption ~ District, data = mhnew, sum)

# Data for histogram
consumption_values <- MH_consumption$total_consumption

# Histogram of total consumption values
hist(consumption_values, breaks = 10, col = 'blue', border = 'black',
      xlab = "Total Consumption", ylab = "Frequency", main = "Histogram of Total
Consumption per District")

View(MH_consumption)

# Additional Plot: Bar plot of total consumption per district
barplot(MH_consumption$total_consumption,
        names.arg = MH_consumption$District,
        las = 2, # Makes the district names vertical
        col = 'blue',
        border = 'black',
        xlab = "District",
        ylab = "Total Consumption",
        main = "Total Consumption per District",
        cex.names = 0.7) # Adjust the size of district names if needed

# Plot total consumption on the Maharashtra state map
data_map <- st_read("C:\\Users\\anjel\\Downloads\\MAHARASHTRA_DISTRICTS.geojson")

data_map <- data_map %>% rename(District = dtname)
data_map_data <- merge(MH_consumption, data_map, by = "District")

View(data_map)

# Plotting the map
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 = 2, color =
"black")

```