Global-Multidimensional-Poverty-Index-Data.r

elvir

2024-12-29

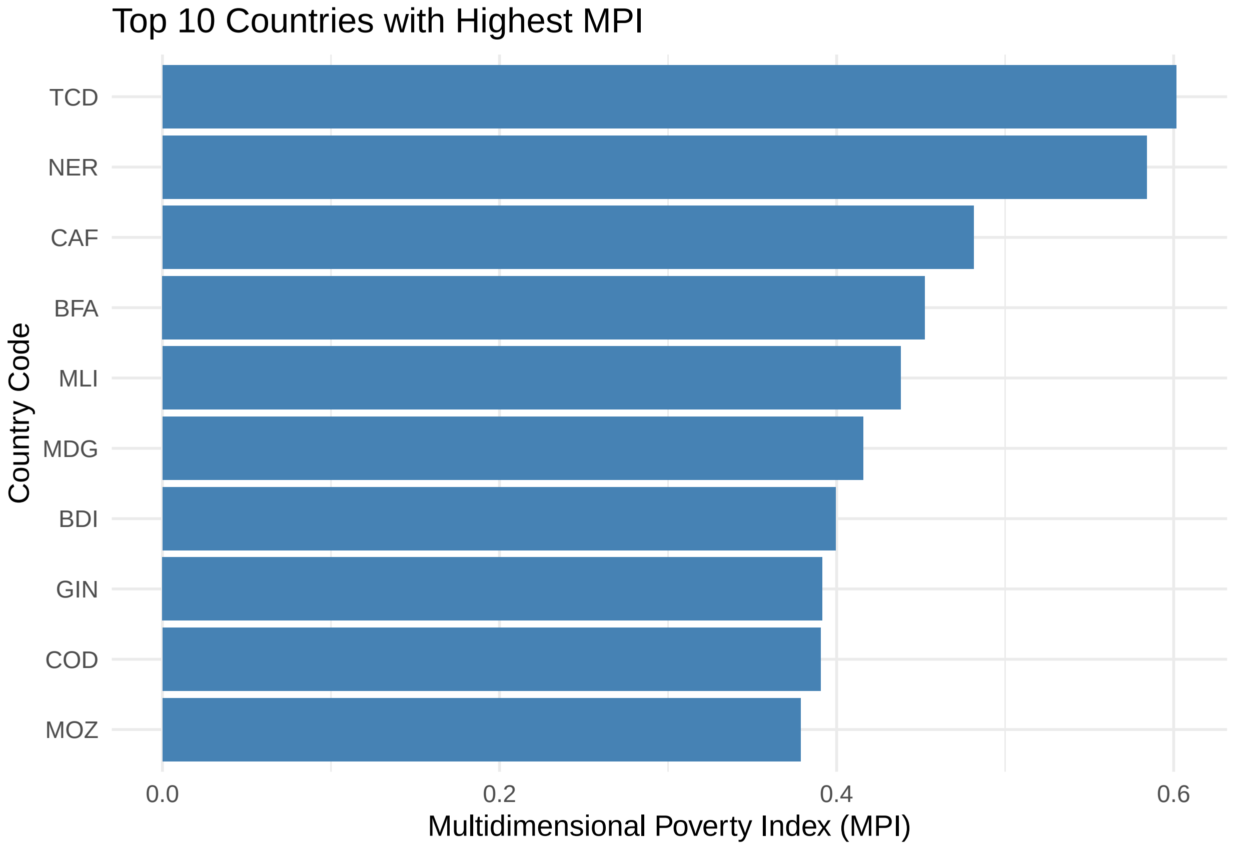
```r
# Load necessary libraries
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(ggplot2)
library(tidyr)
library(knitr)
library(scales)

# Load the dataset
data <- read.csv('hdx_hapi_poverty_rate_global.csv', stringsAsFactors = FALSE)

# Clean the dataset
data_cleaned <- data[-1, ]
colnames(data_cleaned) <- c('location_code', 'has_hrp', 'in_gho', 'provider_admin1_name', 'admin1_code'

# Convert numeric columns to appropriate data types
numeric_columns <- c('mpi', 'headcount_ratio', 'intensity_of_deprivation', 'vulnerable_to_poverty', 'in
data_cleaned[numeric_columns] <- lapply(data_cleaned[numeric_columns], as.numeric)

# Convert date columns to Date format
data_cleaned$reference_period_start <- as.Date(data_cleaned$reference_period_start)
data_cleaned$reference_period_end <- as.Date(data_cleaned$reference_period_end)

# Generate summary statistics
summary_statistics <- summary(data_cleaned[numeric_columns])

# Group data by location_code and calculate mean values
country_summary <- data_cleaned %>%
  group_by(location_code) %>%
  summarise(across(all_of(numeric_columns), mean, na.rm = TRUE))
```

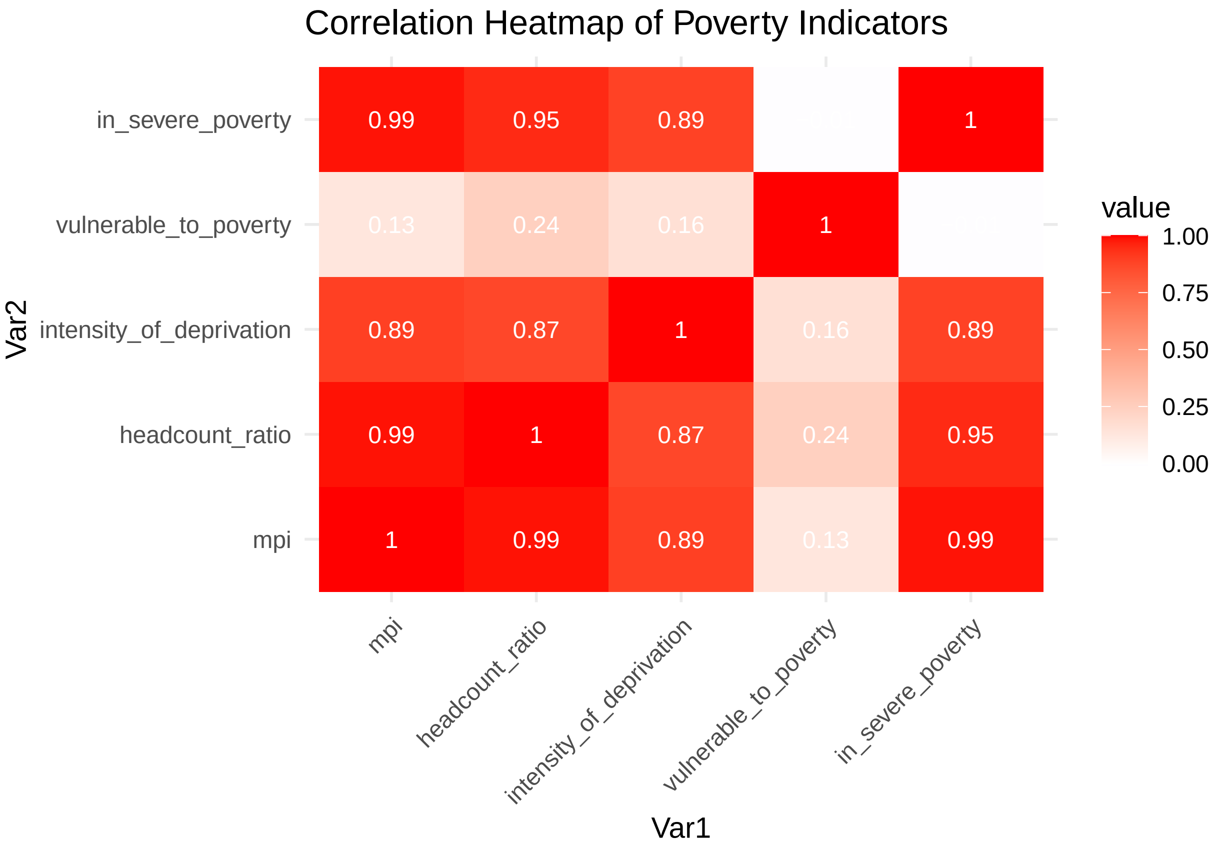
```
## Warning: There was 1 warning in `summarise()`.
## i In argument: `across(all_of(numeric_columns), mean, na.rm = TRUE)`.
## i In group 1: `location_code = "AFG"`.
## Caused by warning:
## ! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.
## Supply arguments directly to `.fns` through an anonymous function instead.
##
##   # Previously
##   across(a:b, mean, na.rm = TRUE)
##
##   # Now
##   across(a:b, \(x) mean(x, na.rm = TRUE))
```

```r
# Top 10 countries with highest MPI
top_10_mpi <- country_summary %>%
  arrange(desc(mpi)) %>%
  head(10)

# Visualization: Top 10 countries with highest MPI
ggplot(top_10_mpi, aes(x = reorder(location_code, mpi), y = mpi)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  coord_flip() +
  labs(title = "Top 10 Countries with Highest MPI",
       x = "Country Code",
       y = "Multidimensional Poverty Index (MPI)") +
  theme_minimal()
```

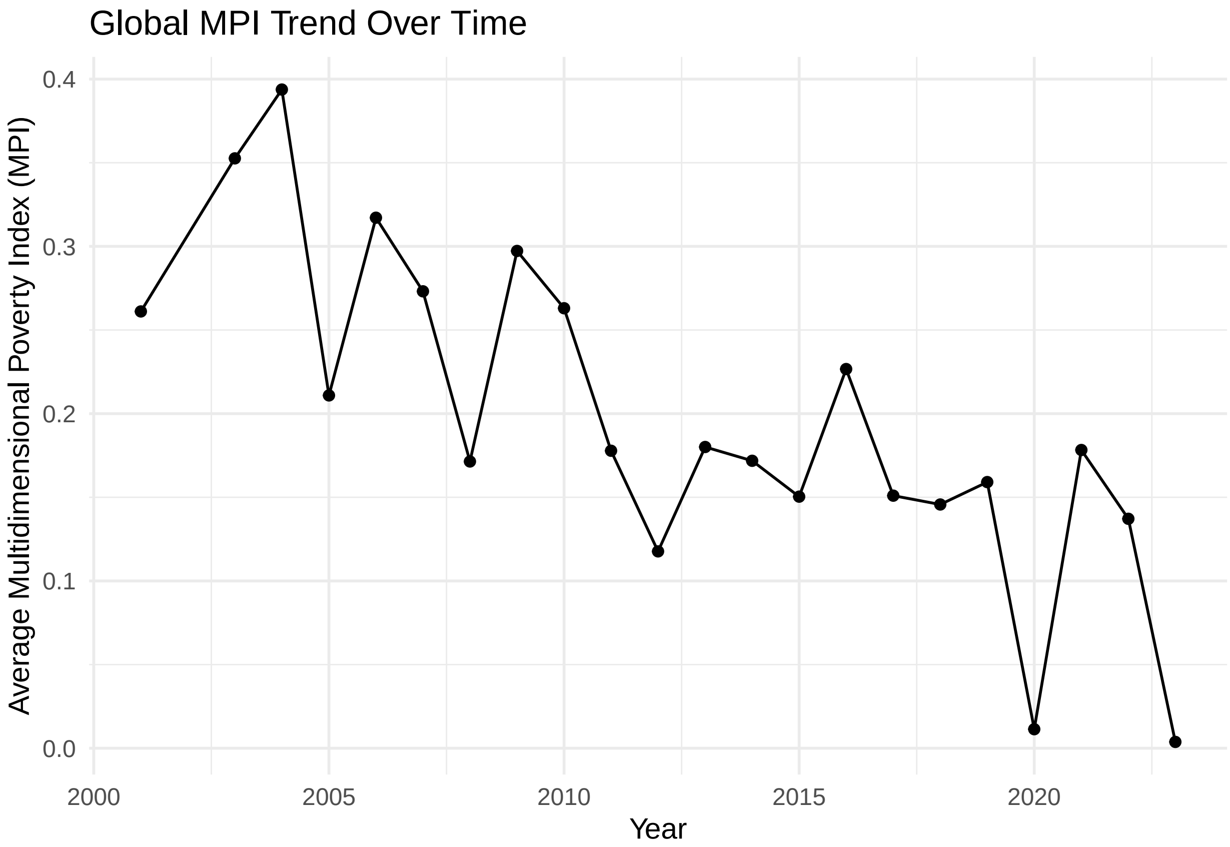


```r
# Correlation analysis
correlation_matrix <- cor(data_cleaned[numeric_columns], use = "complete.obs")
```

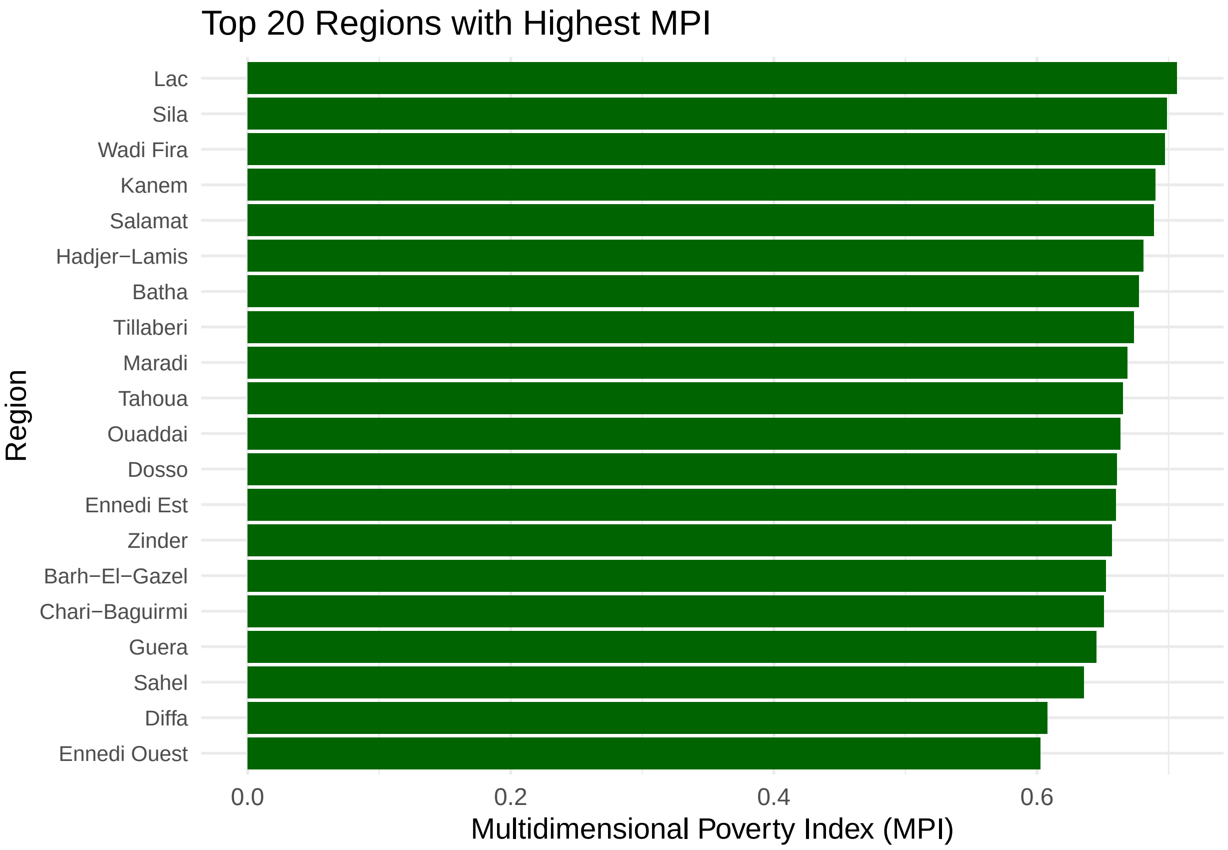
```
# Visualization: Correlation heatmap
ggplot(data = reshape2::melt(correlation_matrix), aes(x = Var1, y = Var2, fill = value)) +
  geom_tile() +
  geom_text(aes(label = round(value, 2)), color = "white", size = 3) +
  scale_fill_gradient2(low = "blue", high = "red", mid = "white", midpoint = 0) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Correlation Heatmap of Poverty Indicators")
```



```
# Time series analysis
time_series_data <- data_cleaned %>%
  group_by(reference_period_start) %>%
  summarise(across(all_of(numeric_columns), mean, na.rm = TRUE))

# Visualization: Time series of MPI
ggplot(time_series_data, aes(x = reference_period_start, y = mpi)) +
  geom_line() +
  geom_point() +
  labs(title = "Global MPI Trend Over Time",
       x = "Year",
       y = "Average Multidimensional Poverty Index (MPI)") +
  theme_minimal()
```



```r
# Regional analysis
regional_summary <- data_cleaned %>%
  group_by(admin1_name) %>%
  summarise(across(all_of(numeric_columns), mean, na.rm = TRUE)) %>%
  arrange(desc(mpi))

# Top 20 regions with highest MPI
top_20_regions <- head(regional_summary, 20)

# Visualization: Top 20 regions with highest MPI
ggplot(top_20_regions, aes(x = reorder(admin1_name, mpi), y = mpi)) +
  geom_bar(stat = "identity", fill = "darkgreen") +
  coord_flip() +
  labs(title = "Top 20 Regions with Highest MPI",
       x = "Region",
       y = "Multidimensional Poverty Index (MPI)") +
  theme_minimal() +
  theme(axis.text.y = element_text(size = 8))
```



```r
# Generate report
report <- capture.output({
  cat("Poverty Rate Analysis Report\n\n")

  cat("1. Summary Statistics\n")
  print(kable(summary_statistics))

  cat("\n2. Top 10 Countries with Highest MPI\n")
  print(kable(top_10_mpi))

  cat("\n3. Correlation Analysis\n")
  print(kable(correlation_matrix))

  cat("\n4. Top 20 Regions with Highest MPI\n")
  print(kable(top_20_regions))
})

# Save report to a text file
writeLines(report, "poverty_analysis_report.txt")

# Print confirmation
cat("Analysis complete. Report saved as 'poverty_analysis_report.txt'")
```

```
## Analysis complete. Report saved as 'poverty_analysis_report.txt'
```