

Data Formulation To fit a Logistic, Multinomial, Ordinal, Generalized Ordinal, Poisson

```
# Load packages
library(tidyverse)
library(stringr)

# Install the GGally package (if not already installed)
#install.packages("GGally")

# Load the GGally package
#library(GGally)

# Import data
data <- read_csv("Draft_data.csv")
```

Data Wrangling

```
# Data wrangle
# Get the column names
column_names <- colnames(data)
world_data <- data |> slice(1:217)
world_data <- world_data[, !(names(world_data)
                             %in% c("P", "Handwash", "EAB",
                                     "EAS", "EAP", "EAL"))]

# List of columns to exclude from conversion
exclude_columns <- c("Region", "Country", "CC")
```

```
# Use lapply to convert all other columns to numeric
world_data[, setdiff(names(world_data), exclude_columns)] <- lapply(
  world_data[, setdiff(names(world_data), exclude_columns)],
  function(x) as.numeric(as.character(x))
)
```

PM Squared

```
# Create the quadratic term for pm
world_data$PM_squared <- world_data$PM^2
```

Sanitation Squared

```
# Create the quadratic term for sanitation
world_data$Sanitation_Squared<- world_data$Sanitation^2
```

Undernourishment Squared

```
# Create the quadratic term for undernourishment
world_data$U_Squared<- world_data$U^2
```

Renewable energy Squared

```
# Create the quadratic term for renewable energy
world_data$RE_Squared<- world_data$RE^2
```

Tobacco Squared

```
# Create the quadratic term for tobacco
world_data$Tobacco_Squared<- world_data$Tobacco^2
```

Immunization squared

```
# Create the quadratic term for tobacco
world_data$IMM_Squared<- world_data$IMM^2
```

PM categorical

```
# Define the breaks and labels for the categories
breaks <- c(0.0, 10.0, 25.0, 50.0, Inf)
labels <- c("Low", "Moderate", "High", "Very High")

# Create a categorical variable based on the specified categories
# Create the categorical variable in your data frame
world_data <- world_data %>%
  mutate(pm_category = cut(PM, breaks = breaks, labels = labels, include.lowest = TRUE))
# Relevel the factor to make "Low" the reference level
world_data$pm_category <- relevel(world_data$pm_category, ref = "Low")
```

PM ordinal

```
# Define the levels and their order
levels_order <- c("Low", "Moderate", "High", "Very High")

# Convert the variable to a factor
world_data$pm_ordinal <- factor(world_data$pm_category,
                                levels = levels_order, ordered = TRUE)
```

Region categorized Data

```
# Assuming your dataset is named 'df' and the column is 'ColumnName'
world_data$Country <- str_to_upper(world_data$Country)
#world_data$Country

# Create a new column named "Region" in the dataset
world_data$Region <- NA
#colnames(world_data)

# List of East Asia and Pacific countries
east_asia_pacific_countries <- c("AMERICAN SAMOA", "AUSTRALIA",
                                "BRUNEI DARUSSALAM", "CAMBODIA",
                                "CHINA", "FIJI", "FRENCH POLYNESIA",
                                "GUAM", "HONG KONG SAR, CHINA",
                                "INDONESIA", "JAPAN", "KIRIBATI",
                                "KOREA, DEM. PEOPLE'S REP.",
                                "KOREA, REP.", "LAO PDR",
```

```

"MACAO SAR, CHINA", "MALAYSIA",
"MARSHALL ISLANDS", "MICRONESIA",
FED. STS.", "MONGOLIA", "MYANMAR",
"NAURU", "NEW CALEDONIA", "NEW ZEALAND",
"NORTHERN MARIANA ISLANDS", "PALAU",
"PAPUA NEW GUINEA", "PHILIPPINES",
"SAMOA", "SINGAPORE", "SOLOMON ISLANDS",
"THAILAND", "TIMOR-LESTE", "TONGA", "TUVALU",
"VANUATU", "VIET NAM")

# Assign the "East Asia and Pacific" region to the specified countries
world_data$Region[world_data$Country %in% east_asia_pacific_countries] <-
  "East Asia and Pacific"

# List of Europe and Central Asia countries
europe_central_asia_countries <- c("ALBANIA", "ANDORRA",
  "ARMENIA", "AUSTRIA",
  "AZERBAIJAN", "BELARUS",
  "BELGIUM", "BOSNIA AND HERZEGOVINA",
  "BULGARIA", "CHANNEL ISLANDS",
  "CROATIA", "CYPRUS", "CZECHIA",
  "DENMARK", "ESTONIA", "FAROE ISLANDS",
  "FINLAND", "FRANCE", "GEORGIA",
  "GERMANY", "GIBRALTAR", "GREECE",
  "GREENLAND", "HUNGARY", "ICELAND",
  "IRELAND", "ISLE OF MAN", "ITALY",
  "KAZAKHSTAN", "KOSOVO",
  "KYRGYZ REPUBLIC", "LATVIA",
  "LIECHTENSTEIN", "LITHUANIA",
  "LUXEMBOURG", "MOLDOVA",
  "MONACO", "MONTENEGRO",
  "NETHERLANDS", "NORTH MACEDONIA",
  "NORWAY", "POLAND", "PORTUGAL",
  "ROMANIA", "RUSSIAN FEDERATION",
  "SAN MARINO", "SERBIA", "SLOVAK REPUBLIC",
  "SLOVENIA", "SPAIN", "SWEDEN",
  "SWITZERLAND", "TAJIKISTAN", "TURKIYE",
  "TURKMENISTAN", "UKRAINE",
  "UNITED KINGDOM", "UZBEKISTAN")

# Assign the "Europe and Central Asia" region to the specified countries
world_data$Region[world_data$Country %in% europe_central_asia_countries] <-
  "Europe and Central Asia"

```

```

# List of Latin America and Caribbean countries
latin_america_caribbean_countries <- c(
  "ANTIGUA AND BARBUDA", "ARGENTINA", "ARUBA",
  "BAHAMAS, THE", "BARBADOS", "BELIZE", "BOLIVIA",
  "BRAZIL", "BRITISH VIRGIN ISLANDS",
  "CAYMAN ISLANDS", "CHILE", "COLOMBIA", "COSTA RICA",
  "CUBA", "CURACAO", "DOMINICA", "DOMINICAN REPUBLIC",
  "ECUADOR", "EL SALVADOR",
  "GRENADA", "GUATEMALA", "GUYANA", "HAITI", "HONDURAS",
  "JAMAICA", "MEXICO", "NICARAGUA", "PANAMA", "PARAGUAY",
  "PERU", "PUERTO RICO",
  "SINT MAARTEN (DUTCH PART)", "ST. KITTS AND NEVIS",
  "ST. LUCIA", "ST. MARTIN (FRENCH PART)",
  "ST. VINCENT AND THE GRENADINES",
  "SURINAME", "TRINIDAD AND TOBAGO",
  "TURKS AND CAICOS ISLANDS", "URUGUAY",
  "VENEZUELA, RB", "VIRGIN ISLANDS (U.S.)"
)
# Assign the "Latin America and Caribbean" region to the specified countries
world_data$Region[world_data$Country %in% latin_america_caribbean_countries] <-
  "Latin America and Caribbean"

# List of Middle East and North Africa countries
middle_east_north_africa_countries <- c(
  "ALGERIA", "BAHRAIN", "DJIBOUTI", "EGYPT, ARAB REP.",
  "IRAN, ISLAMIC REP.", "IRAQ", "ISRAEL", "JORDAN",
  "KUWAIT", "LEBANON",
  "LIBYA", "MALTA", "MOROCCO", "OMAN", "QATAR",
  "SAUDI ARABIA", "SYRIAN ARAB REPUBLIC", "TUNISIA",
  "UNITED ARAB EMIRATES",
  "WEST BANK AND GAZA", "YEMEN, REP."
)
# Assign the "Middle East and North Africa" region to the specified countries
world_data$Region[world_data$Country %in% middle_east_north_africa_countries] <-
  "Middle East and North Africa"

# List of North American countries
north_american_countries <- c("BERMUDA", "CANADA", "UNITED STATES")
# Assign the "North America" region to the specified countries
world_data$Region[world_data$Country %in% north_american_countries] <-

```

```

"North America"

# List of South Asian countries
south_asian_countries <- c("AFGHANISTAN", "BANGLADESH", "BHUTAN",
                           "INDIA", "MALDIVES", "NEPAL", "PAKISTAN", "SRI LANKA")
# Assign the "South Asia" region to the specified countries
world_data$Region[world_data$Country %in% south_asian_countries] <-
  "South Asia"

sub_saharan_africa_countries <- c(
  "ANGOLA", "BENIN", "BOTSWANA", "BURKINA FASO", "BURUNDI",
  "CABO VERDE", "CAMEROON", "CENTRAL AFRICAN REPUBLIC",
  "CHAD", "COMOROS", "CONGO, DEM. REP.", "CONGO, REP.",
  "COTE D'IVOIRE", "EQUATORIAL GUINEA", "ERITREA", "ESWATINI",
  "ETHIOPIA", "GABON", "GAMBIA, THE", "GHANA", "GUINEA",
  "GUINEA-BISSAU", "KENYA", "LESOTHO", "LIBERIA", "MADAGASCAR",
  "MALAWI", "MALI", "MAURITANIA", "MAURITIUS", "MOZAMBIQUE",
  "NAMIBIA", "NIGER", "NIGERIA", "RWANDA", "SAO TOME AND PRINCIPE",
  "SENEGAL", "SEYCHELLES", "SIERRA LEONE", "SOMALIA", "SOUTH AFRICA",
  "SOUTH SUDAN", "SUDAN", "TANZANIA", "TOGO", "UGANDA",
  "ZAMBIA", "ZIMBABWE"
)
# Assign the "Sub-Saharan Africa" region to the specified countries
world_data$Region[world_data$Country %in% sub_saharan_africa_countries] <-
  "Sub-Saharan Africa"

# Convert "Region" to a factor variable
world_data$Region <- as.factor(world_data$Region)
# Set "Sub-Saharan Africa" as the reference level
world_data$Region <- relevel(world_data$Region, ref = "Sub-Saharan Africa")

```

Mortality categorized data

```

# Assuming you have a data frame named 'your_data' with a 'mortality' variable
world_data <- world_data %>%
  mutate(Mortality_Category = ifelse(Mortality > 25, "High", "Low"))
# Convert "Mortality_Category" to a factor

```

```
world_data$Mortality_Category <- factor(world_data$Mortality_Category)
# Relevel the factor to make "Low" the reference level
world_data$Mortality_Category <- relevel(world_data$Mortality_Category, ref = "Low")
```

Mortality ordinal

```
# Define the levels and their order
levels_order <- c("Low", "High")

# Convert the variable to a factor
world_data$Mortality_Ordinal <-
  factor(world_data$Mortality_Category,
        levels = levels_order, ordered = TRUE)
```

Final Data

```
# Filter the data to keep only countries with GDP greater than 3400
world_data <- world_data %>% filter(GDP < 33500)

# Removes rows with missing values
world_data <- na.omit(world_data)

# Save your data frame to a CSV file
write.csv(world_data, "final_bst210_project_data.csv", row.names = FALSE)
```