

# Group 1 - Selected Databases and Indicators

*Master's in Business Analytics – Culminating Project*

**Week of: 2**

Joshua Cabal, Namrata Patil, Christopher Yeung, Celine Abrahamian

---

## Database 1: Food Prices for Nutrition

<https://databank.worldbank.org/source/food-prices-for-nutrition>

This database contains information about CoAHD, the “cost and affordability of a healthy diet.” It is a fairly new addition to the World Bank’s database collection, as the data only go back to 2017 and up to 2022. The indicators reflect the population’s access to healthy food and the relative cost of that food. We chose this database because food is not typically one of the major metrics of the health of a population (such as GDP), yet all of life revolves around food. We are interested in how the cost of healthy food might have positive and/or adverse effects on a population’s mortality, education, and governance, and vice versa. We will use all the indicators from this database in our analysis because they all involve food and food insecurity, or are necessary for context involving food and food insecurity (such as “Population”).

### Food Price Indicators

- Cost of an energy sufficient diet
- Cost of a nutrient adequate diet
- Cost of a healthy diet
- Cost of a healthy diet relative to the cost of sufficient energy from starchy staples
- Cost of fruits
- Cost of vegetables
- Cost of starchy staples
- Cost of animal-source foods
- Cost of legumes, nuts and seeds
- Cost of oils and fats
- Cost share for fruits in a least-cost healthy diet
- Cost share for vegetables in a least-cost healthy diet
- Cost share for starchy staples in a least-cost healthy diet
- Cost share for animal-sourced foods in a least-cost healthy diet
- Cost share for legumes, nuts and seeds in a least-cost healthy diet
- Cost share for oils and fats in a least-cost healthy diet
- Cost of fruits relative to the starchy staples in a least-cost healthy diet
- Cost of vegetables relative to the starchy staples in a least-cost healthy diet
- Cost of animal-sourced foods relative to the starchy staples in a least-cost healthy diet

- Cost of legumes, nuts and seeds relative to the starchy staples in a least-cost healthy diet
- Cost of oils and fats relative to the starchy staples in a least-cost healthy diet
- Affordability of an energy sufficient diet: ratio of cost to the food poverty line
- Affordability of a nutrient adequate diet: ratio of cost to the food poverty line
- Affordability of a healthy diet: ratio of cost to the food poverty line
- Affordability of an energy sufficient diet: ratio of cost to food expenditures
- Affordability of a nutrient adequate diet: ratio of cost to food expenditures
- Affordability of a healthy diet: ratio of cost to food expenditures
- Percent of the population who cannot afford sufficient calories
- Percent of the population who cannot afford nutrient adequacy
- Percent of the population who cannot afford a healthy diet
- Millions of people who cannot afford sufficient calories
- Millions of people who cannot afford nutrient adequacy
- Millions of people who cannot afford a healthy diet
- Population

## Database 2: Worldwide Governance Indicators

<https://databank.worldbank.org/source/worldwide-governance-indicators>

Worldwide Governance Indicators assess 6 key indicators of governance:

- Control of Corruption
- Government Effectiveness
- Political Stability and Absence of Violence/Terrorism
- Regulatory Quality
- Rule of Law
- Voice and Accountability

These indicators provide valuable insights into how governments perform and its impact on economic and social development, human rights, security and conflict, and public health. This dataset includes data for the USA from 1996 to 2023. We are interested in how these indicators might have positive and/or adverse effects on a population's mortality, education, and governance, and vice versa

## Governance Indicators

### Control of Corruption

1. Control of Corruption: Estimate
2. Control of Corruption: Number of Sources
3. Control of Corruption: Percentile Rank
4. Control of Corruption: Percentile Rank, Lower Bound of 90% Confidence Interval

5. Control of Corruption: Percentile Rank, Upper Bound of 90% Confidence Interval
6. Control of Corruption: Standard Error

### **Government Effectiveness**

1. Government Effectiveness: Estimate
2. Government Effectiveness: Number of Sources
3. Government Effectiveness: Percentile Rank
4. Government Effectiveness: Percentile Rank, Lower Bound of 90% Confidence Interval
5. Government Effectiveness: Percentile Rank, Upper Bound of 90% Confidence Interval
6. Government Effectiveness: Standard Error

### **Political Stability and Absence of Violence/Terrorism**

1. Political Stability and Absence of Violence/Terrorism: Estimate
2. Political Stability and Absence of Violence/Terrorism: Number of Sources
3. Political Stability and Absence of Violence/Terrorism: Percentile Rank
4. Political Stability and Absence of Violence/Terrorism: Percentile Rank, Lower Bound of 90% Confidence Interval
5. Political Stability and Absence of Violence/Terrorism: Percentile Rank, Upper Bound of 90% Confidence Interval
6. Political Stability and Absence of Violence/Terrorism: Standard Error

### **Regulatory Quality**

1. Regulatory Quality: Estimate
2. Regulatory Quality: Number of Sources
3. Regulatory Quality: Percentile Rank
4. Regulatory Quality: Percentile Rank, Lower Bound of 90% Confidence Interval
5. Regulatory Quality: Percentile Rank, Upper Bound of 90% Confidence Interval
6. Regulatory Quality: Standard Error

### **Rule of Law**

1. Rule of Law: Estimate
2. Rule of Law: Number of Sources
3. Rule of Law: Percentile Rank
4. Rule of Law: Percentile Rank, Lower Bound of 90% Confidence Interval
5. Rule of Law: Percentile Rank, Upper Bound of 90% Confidence Interval
6. Rule of Law: Standard Error

### **Voice and Accountability**

1. Voice and Accountability: Estimate
2. Voice and Accountability: Number of Sources
3. Voice and Accountability: Percentile Rank
4. Voice and Accountability: Percentile Rank, Lower Bound of 90% Confidence Interval

5. Voice and Accountability: Percentile Rank, Upper Bound of 90% Confidence Interval
6. Voice and Accountability: Standard Error

## Database 3: World Development Indicators

<https://databank.worldbank.org/source/world-development-indicators>

The World Development Indicators is the primary World Bank collection of development indicators, compiled from officially recognized international sources. It presents the most current and accurate global development data available, and includes national, regional and global estimates. We were primarily interested in two different series groups from this data set:

1. Education Indicators
2. Mortality Indicators

Without loss of generality, all datasets were created using the US as the selected country and all available years. Naturally, different countries will have different prevalent indicators, and therefore the research question we choose to develop will determine which indicator(s) are most important to analyze mortality.

### World Development Indicators - Education

The following datasets were created using the World Development Indicators database. They use the United States as the selected country and all the available years.

1. Education Access & Enrollment in the U.S. by Year
2. Education Spending & Government Investment in the U.S. by Year
3. Education Outcomes & Workforce Impact in the U.S. by Year
4. Education & Gender Disparities in the U.S. by Year

The education indicator measures key aspects of the U.S. education system, including access to schooling, government investment, and educational outcomes. It provides insight into the role of education in economic and social development as it highlights trends in enrollment rates, funding allocation, and workforce readiness.

### Education Indicators

#### Education Access & Attainment

1. Primary education, duration (years)
2. Secondary education, duration (years)
3. Educational attainment, at least Bachelor's degree, population 25+, total (%)
4. Educational attainment, at least completed primary, population 25+, total (%)
5. Educational attainment, at least completed upper secondary, population 25+, total (%)
6. Educational attainment, Doctoral or equivalent, population 25+, total (%)

### **Education Investment & Expenditure**

1. Government expenditure on education, total (% of GDP)
2. Government expenditure on education, total (% of government expenditure)
3. Expenditure on primary education (% of government expenditure on education)
4. Expenditure on secondary education (% of government expenditure on education)
5. Expenditure on tertiary education (% of government expenditure on education)
6. Adjusted savings: education expenditure (% of GNI)
7. Adjusted savings: education expenditure (current US\$)

### **Education & Workforce Readiness**

1. Labor force with advanced education (% of total working-age population with advanced education)
2. Labor force with basic education (% of total working-age population with basic education)
3. Labor force with intermediate education (% of total working-age population with intermediate education)
4. Unemployment with advanced education (% of total labor force with advanced education)
5. Unemployment with basic education (% of total labor force with basic education)
6. Unemployment with intermediate education (% of total labor force with intermediate education)
7. Share of youth not in education, employment, or training, total (% of youth population)

### **Education & Gender Disparities**

1. Primary education, pupils (% female)
2. Secondary education, pupils (% female)
3. Educational attainment, at least Bachelor's degree, population 25+, female (%)
4. Educational attainment, at least Master's degree, population 25+, female (%)
5. Educational attainment, Doctoral or equivalent, population 25+, female (%)
6. Trained teachers in primary education, female (% of female teachers)
7. Trained teachers in secondary education, female (% of female teachers)
8. Tertiary education, academic staff (% female)

### **Education Infrastructure & Teaching Staff**

- Preprimary education, duration (years)
- Primary education, teachers
- Secondary education, teachers
- Trained teachers in primary education (% of total teachers)
- Trained teachers in secondary education (% of total teachers)
- Trained teachers in upper secondary education (% of total teachers)

## World Development Indicators - Mortality

The following datasets were created using the World Development Indicators database. They use the United States as the selected country and all the available years.

- Risk Factors and Mortality in the US by Year
- US Nutrition and Health Data by Year
- US Emissions Data by Year
- Birth and Fertility Rates in US by Year
- Poverty and Environmental Indicators in US by Year

### Mortality Indicators

Many mortality indicators are available. Total mortality rates or rates amongst specific ages would likely be the target columns we could build models to predict.

- **HIV-Related Indicators:**
  - Adults (ages 15+) and children (ages 0-14) newly infected with HIV
  - Adults (ages 15-49) newly infected with HIV
  - Antiretroviral therapy coverage (% of people living with HIV)
  - Antiretroviral therapy coverage for PMTCT (% of pregnant women living with HIV)
  - Children (0-14) living with HIV
  - Children (ages 0-14) newly infected with HIV
  - Incidence of HIV, ages 15-24 (per 1,000 uninfected population ages 15-24)
  - Incidence of HIV, ages 15-49 (per 1,000 uninfected population ages 15-49)
  - Incidence of HIV, all (per 1,000 uninfected population)
  - Young people (ages 15-24) newly infected with HIV
  - Prevalence of HIV, female (% ages 15-24)
  - Prevalence of HIV, male (% ages 15-24)
  - Prevalence of HIV, total (% of population ages 15-49)
  - Women's share of population ages 15+ living with HIV (%)
- **Infectious Diseases & Other Health Conditions:**
  - Incidence of malaria (per 1,000 population at risk)
  - Incidence of tuberculosis (per 100,000 people)
  - Cause of death, by communicable diseases and maternal, prenatal and nutrition conditions (% of total)
  - Cause of death, by injury (% of total)
  - Cause of death, by non-communicable diseases (% of total)
  - Diabetes prevalence (% of population ages 20 to 79)
  - Female genital mutilation prevalence (%)
- **Mortality & Life Expectancy Indicators:**
  - Life expectancy at birth, female (years)
  - Life expectancy at birth, male (years)
  - Life expectancy at birth, total (years)
  - Mortality rate, adult, female (per 1,000 female adults)

- Mortality rate, adult, male (per 1,000 male adults)
- Mortality rate, infant (per 1,000 live births)
- Mortality rate, infant, female (per 1,000 live births)
- Mortality rate, infant, male (per 1,000 live births)
- Mortality rate, neonatal (per 1,000 live births)
- Mortality rate, under-5 (per 1,000 live births)
- Mortality rate, under-5, female (per 1,000 live births)
- Mortality rate, under-5, male (per 1,000 live births)
- Mortality caused by road traffic injury (per 100,000 population)
- Mortality from CVD, cancer, diabetes or CRD between exact ages 30 and 70 (%)
- Mortality from CVD, cancer, diabetes or CRD between exact ages 30 and 70, female (%)
- Mortality from CVD, cancer, diabetes or CRD between exact ages 30 and 70, male (%)
- Mortality rate attributed to household and ambient air pollution, age-standardized (per 100,000 population)
- Mortality rate attributed to household and ambient air pollution, age-standardized, female (per 100,000 female population)
- Mortality rate attributed to household and ambient air pollution, age-standardized, male (per 100,000 male population)
- Mortality rate attributed to unintentional poisoning (per 100,000 population)
- Mortality rate attributed to unintentional poisoning, female (per 100,000 female population)
- Mortality rate attributed to unintentional poisoning, male (per 100,000 male population)
- Mortality rate attributed to unsafe water, unsafe sanitation and lack of hygiene (per 100,000 population)
- Suicide mortality rate (per 100,000 population)
- Suicide mortality rate, female (per 100,000 female population)
- Suicide mortality rate, male (per 100,000 male population)
- Number of infant deaths
- Number of neonatal deaths
- Number of under-five deaths
- Number of deaths ages 5-9 years
- Number of deaths ages 10-14 years
- Number of deaths ages 15-19 years
- Number of deaths ages 20-24 years
- **Other Mortality and Risk Indicators:**
  - Mortality rate attributed to unintentional poisoning (per 100,000 population)
  - Probability of dying among adolescents ages 10-14 years (per 1,000)
  - Probability of dying among adolescents ages 15-19 years (per 1,000)
  - Probability of dying among children ages 5-9 years (per 1,000)
  - Probability of dying among youth ages 20-24 years (per 1,000)
  - Risk of catastrophic expenditure for surgical care (% of people at risk)

- Risk of impoverishing expenditure for surgical care (% of people at risk)
- Survival to age 65, female (% of cohort)
- Survival to age 65, male (% of cohort)
- **Environmental and Lifestyle Indicators:**
  - People practicing open defecation (% of population)
  - People practicing open defecation, rural (% of rural population)
  - People practicing open defecation, urban (% of urban population)
  - Prevalence of current tobacco use (% of adults)
  - Prevalence of current tobacco use, females (% of female adults)
  - Prevalence of current tobacco use, males (% of male adults)
  - Total alcohol consumption per capita (liters of pure alcohol, projected estimates, 15+ years of age)
  - Total alcohol consumption per capita, female (liters of pure alcohol, projected estimates, female 15+ years of age)
  - Total alcohol consumption per capita, male (liters of pure alcohol, projected estimates, male 15+ years of age)