



Global Nutrition and Health: Analyzing the Relationship Between Diet Composition and Disease Prevalence

Meme Habel

Yichen Qin and Ruth Seiple

Introduction

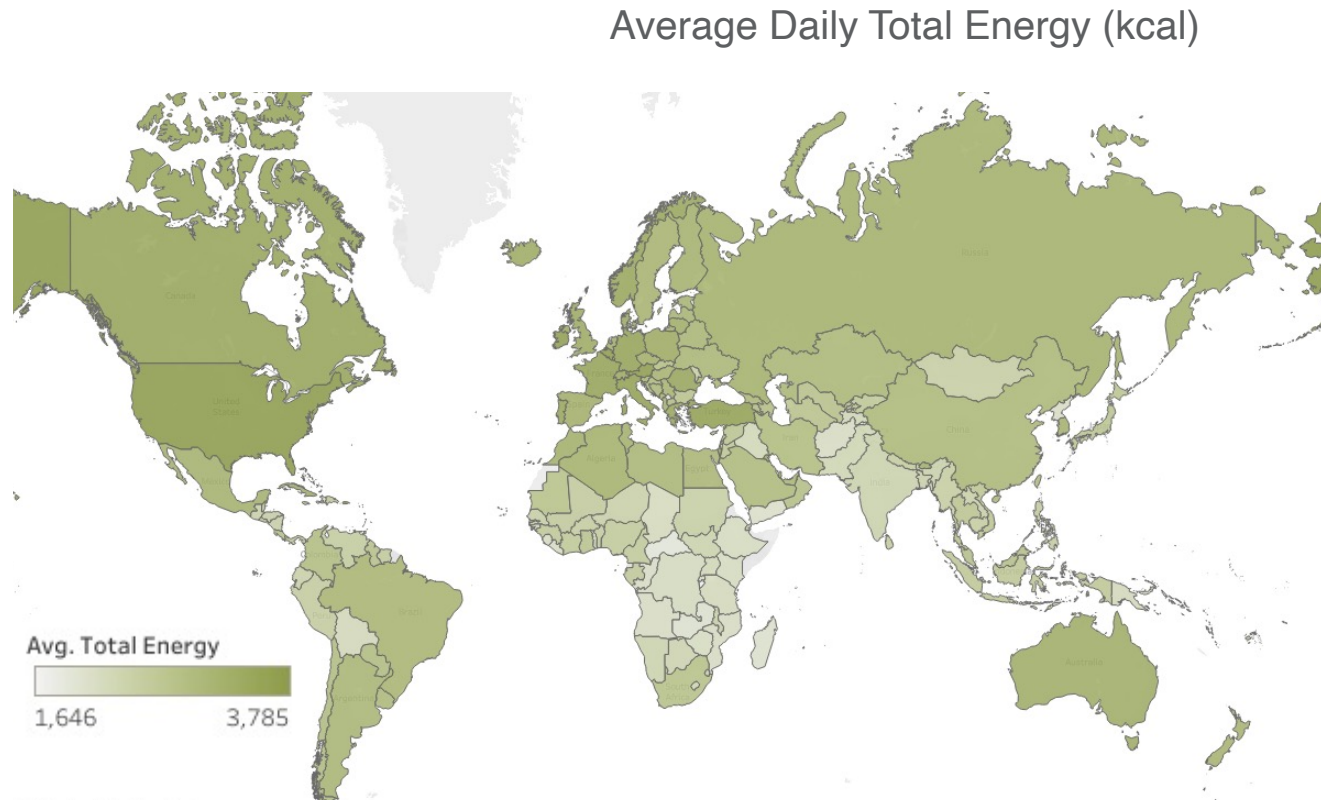


vs.



Problem Description

- Understand the impact of diet composition and lifestyle habits on disease prevalence.



Data Definitions

Macronutrient Composition:

- Location_id **primary key*
- Country
- Year
- total_energy (kcal)
- avg_total_fat (g)
- avg_total_protein (g)
- avg_total_carbohydrates (g)
- avg_energy_from_animal_products (kcal)

Data Definitions

Diseases:

- Location_id **primary key*
- Location_name (country)
- Cause_id
- Cause_name (disease name)
- Prevalence_Rate: Total cases per 100,000 population
- Death_Rate: Deaths per 100,000 population
- DALY_Rate: Disability adjusted life years (DALYs) per 100,000 population

Note: DALY = The sum of years lost due to premature death (YLLs) and years lived with disability (YLDs). One DALY equals one lost year of healthy life.

Data Definitions

Alcohol Consumption:

- Location_id **primary key*
- Year
- Beverage_Type (Beer, Wine, Spirits, Other alcoholic beverages, or All types)
- Alc_consumption_per_capita

Physical Activity:

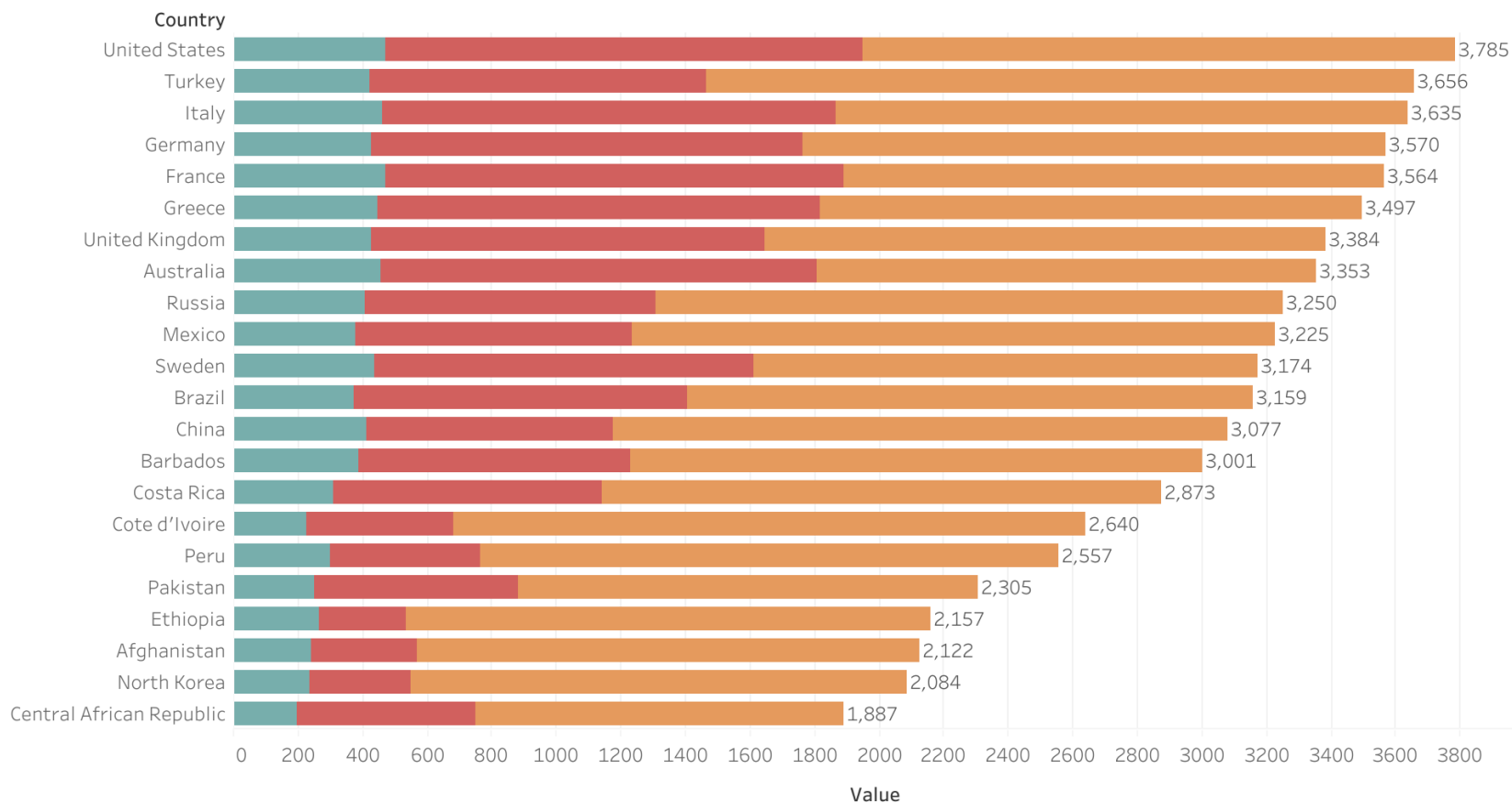
- Location_id **primary key*
- Year
- Percent_insufficient_phys_activity (Percent of population attaining less than 150 minutes of moderate-intensity physical activity per week, or less than 75 minutes of vigorous-intensity physical activity per week, or equivalent.)

Average Daily Energy Composition

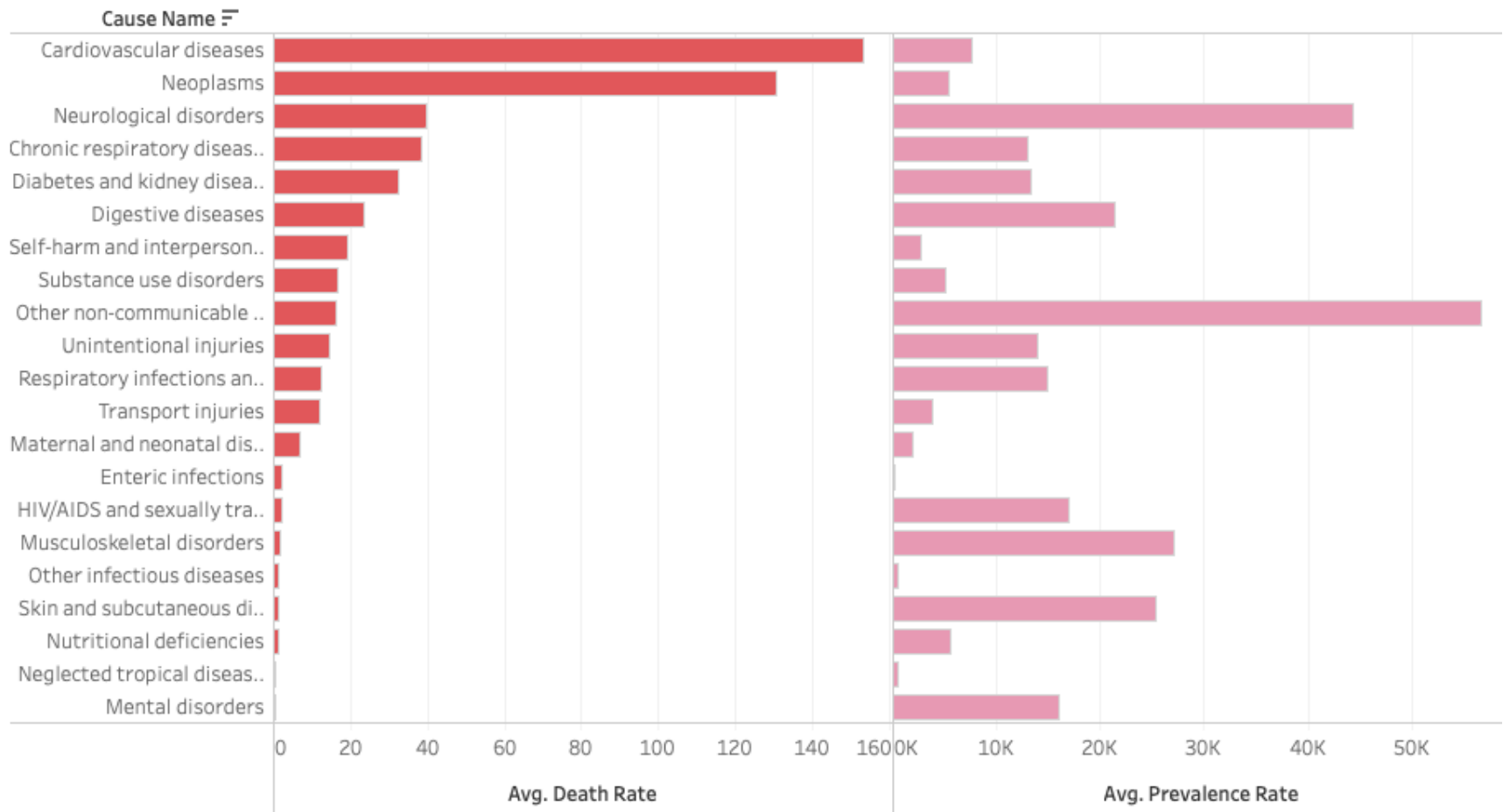
Breakdown of Energy Sources

Measure Names

- Avg. Total Energy From Carbohydrates
- Avg. Total Energy From Fat
- Avg. Total Energy From Protein



USA Disease Death and Prevalence Rates



Model Setup

K-Folds Cross-Validation

- $K = 10$

Min-Max Normalization

- Range: $[0, 1]$

Recursive Feature Elimination (RFE)

- Backwards feature selection algorithm

Evaluation Metrics:

- Mean Squared Prediction Error (MSPE)
- Mean Absolute Error (MAE)
- R-squared

Types of models:

- Multiple Linear Regression
- Random Forests

Cardiovascular Diseases

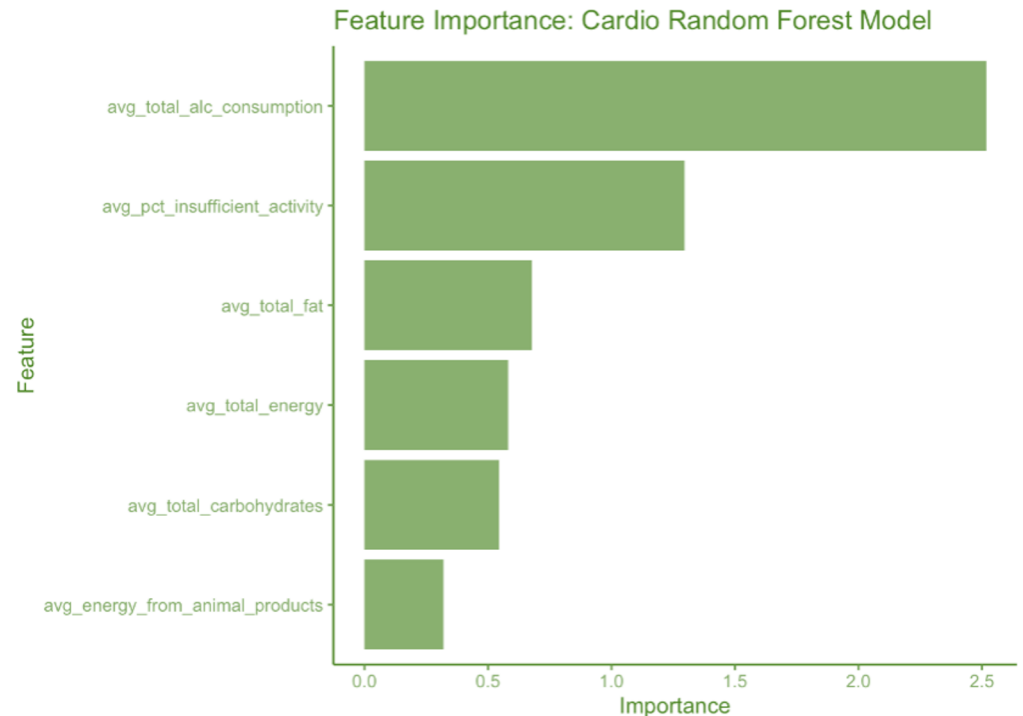
Target Variable: Average Disease Prevalence Rate

Linear Regression Model

- Predictors and Coefficients
 - avg_total_alc_consumption:
 - -182.302717***
 - avg_energy_from_animal_products:
 - 0.197004
- MSPE: 0.04210
- MAE: 0.1654
- R-Squared: 0.1594

Random Forest Model

- MSPE: 0.03643
- MAE: 0.1548
- R-Squared: 0.2781



Neoplasms

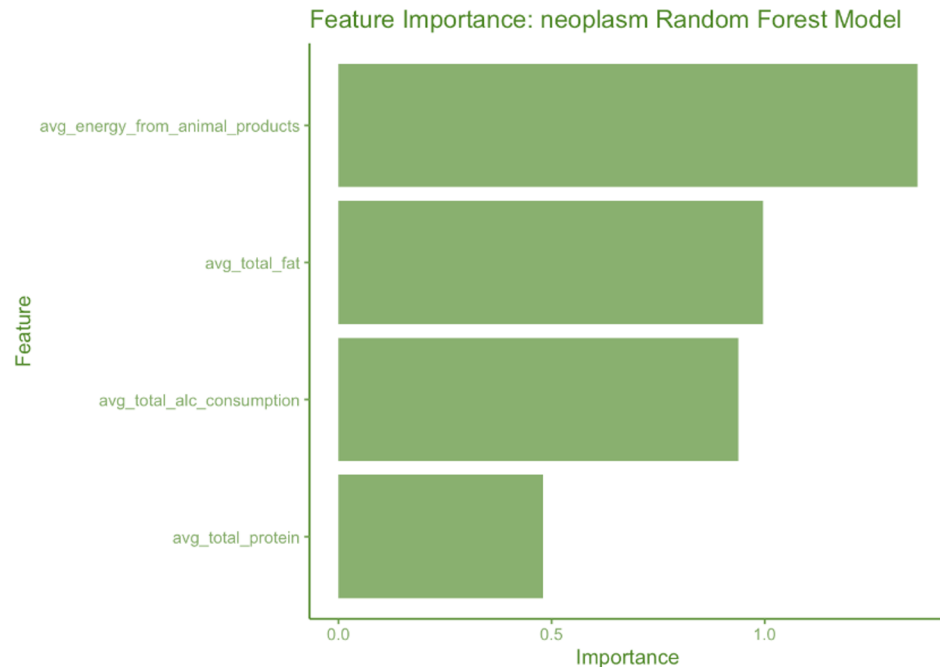
Target Variable: Average Disease Prevalence Rate

Linear Regression Model

- Predictors and Coefficients
 - avg_total_alc_consumption:
 - 14.85078***
 - avg_total_fat:
 - 92.39680***
- MSPE: 0.01164
- MAE: 0.07863
- R-Squared: 0.6411

Random Forest Model

- MSPE: 0.01203
- MAE: 0.07709
- R-Squared: 0.6287



Chronic Respiratory Diseases

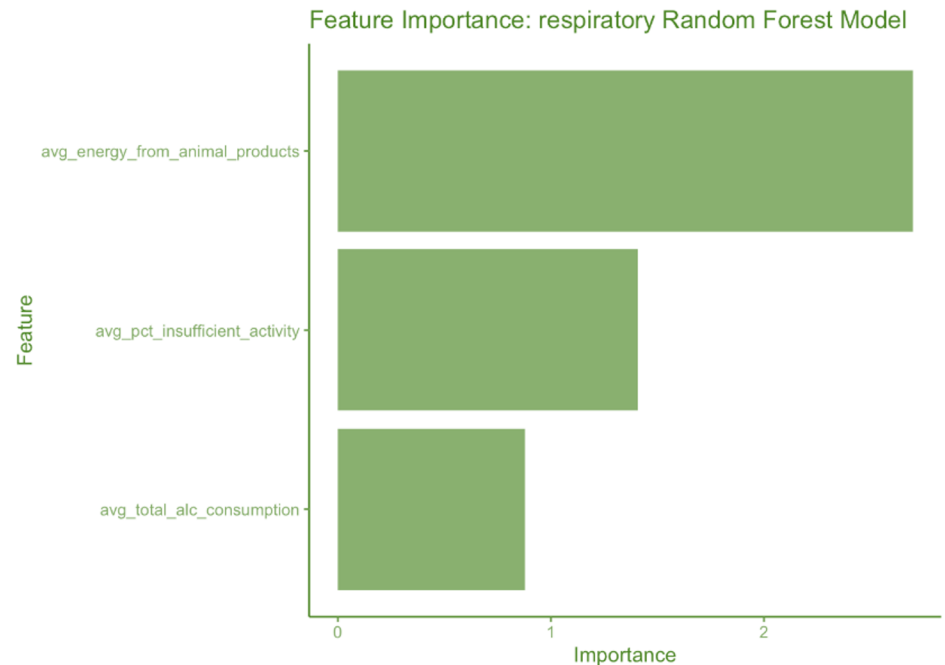
Target Variable: Average Disease Prevalence Rate

Linear Regression Model

- Predictors and Coefficients
 - avg_energy_from_animal_products:
 - 3.914034***
 - avg_pct_insufficient_activity:
 - 59.496601***
- MSPE: 0.02671
- MAE: 0.1276
- R-Squared: 0.3626

Random Forest Model

- MSPE: 0.02184
- MAE: 0.1167
- R-Squared: 0.4816



Conclusions

- Energy from animal products → higher risk of Cardiovascular and Chronic Respiratory Diseases
- Alcohol Consumption → higher risk for Neoplasms
- Insufficient Physical Activity → higher risk for Chronic Respiratory Diseases
- Large total daily fats consumption → higher risk for Neoplasms

References

- “Gho I by Category I Recorded Alcohol per Capita Consumption, from 2010 - Updated May 2021.” World Health Organization, World Health Organization, apps.who.int/gho/data/node.main.A1039?lang=en. Accessed 1 July 2024.
- Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2021 (GBD 2021). Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2024.
- “Prevalence of Insufficient Physical Activity among Adults Aged 18+ Years (Age-Standardized Estimate) (%)” World Health Organization, World Health Organization, [www.who.int/data/gho/data/indicators/indicator-details/GHO/prevalence-of-insufficient-physical-activity-among-adults-aged-18-years-\(age-standardized-estimate\)-\(-\)](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/prevalence-of-insufficient-physical-activity-among-adults-aged-18-years-(age-standardized-estimate)-(-)).
- Ritchie, Hannah, et al. “Diet Compositions.” Our World in Data, 28 Dec. 2023, ourworldindata.org/diet-compositions.