Assessing the Effects of Diet on Disease

Zhaokailu Gu, zg2382@columbia.edu
Huan Sun, hs3240@columbia.edu
Renfang Yang, ry2403@columbia.edu

Abstract

There's a saying that what we eat determines what we are. Intuitively, diet directly affects health. And we aim to test this hypothesis and study the effect of nutrition on disease in this project. To be more specific, our research questions include:

Static Images/ Geospatial data:

1) How's the undernutrition distribution across the world changes in relation to time?
2) How does the diet pattern vary for different countries, sex, and age groups?
3) Do people from certain countries, sex, or age groups more likely to have some diseases?
4) Did diet patterns change over time in country, sex, age group level?
5) Do some diseases surge or diminish over time?
6) Is there a correlation between the diet pattern and disease rate?

Text Analyses Visualizations:

1) How does the word count frequency for specific food involved in food/nutrition subreddits?
2) Conduct sentiment analysis to study public perceptions of specific food/nutrition based on discussions of related foods/nutrients in Reddit comments.

Network Visualizations:

1) The relationship between certain food imports and related disease rates changes.
2) Does the social network impact people's perception of healthy dieting?

By answering these questions, we hope to give some insights into medical research, health institutions, the public health sector, and relevant businesses.

Techniques:

- ggplot2
- Spatial data techniques
- Text analysis techniques
- Social Network visualization tools

Data Description:

Data Source:

- WHO: https://www.who.int/data/gho/info/gho-odata-api
  - Leading cause of death: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death
  - Life expectancy and healthy life expectancy https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-life-expectancy-and-healthy-life-expectancy
- Dietary: https://www.globaldietarydatabase.org/
  - Dietary factors include
    Foods:
    Fruits, Non-starchy vegetables, Potatoes, Other starchy vegetables, Beans and legumes, Nuts and seeds, Refined grains, Whole grains; Total processed meats, Unprocessed red meats, Total seafood, Eggs, Cheese, Yogurt including fermented milk;

    Beverages:
    Sugar-sweetened beverages, Fruit juices, Coffee, Tea, Total milk;

    Macronutrients:
    Total protein, Saturated fat, Monounsaturated fat, Total omega-6 fatty acids, Seafood omega-3 fatty acids, Plant omega-3 fatty acids

    Micronutrients:
    Dietary cholesterol, Dietary fiber, Added sugars, Calcium, Sodium, Iodine, Iron, Magnesium, Potassium, Vitamin A with supplements, Vitamin B1, Vitamin B2, Vitamin B3, Vitamin B6, Vitamin B9, Vitamin B12, Vitamin C, Vitamin D, Vitamin E, Zinc

  - Year
    1990, 1995, 2000, 2005, 2010, 2015, 2018.
  - Country
    185 countries
  - Age
    0-11mo., 12-23mo., 2-5, 6-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+
  - Sex
    Male and female
  - Residence
    Urban and rural (as defined by each survey's characteristics)

    - Education Level
      Low (0-6 years formal education), medium (6.01-12.0 years), and high (12.01+ years)
- Reddit: Using Reddit API or PushShift API to retrieve subreddit comments from the following subreddits.
  We plan to scrap
    - r/nutrition, r/HealthyFood
- Diabetes-related:
    - https://diabetesatlas.org/data/en/

Visualizations:

- Geographic distribution of different dietary factors at country level
    - Map with interactive options to select dietary factors, the year, and demographics.
- Dietary intake patterns in different groups including sex, age, etc.
    - Bar charts with interactive options to select year, country, and dietary factors.
- Time trend of undernutrition intake
    - Line charts with interactive options to select country, certain group, and dietary factors.
- Time trend of some diseases
    - Line charts with interactive options to select country, certain group, and dietary factors
- Correlation between certain diseases and undernutrition
    - Scatter plots with interactive options to select diseases and undernutrition
- Word Cloud of discussion on diet
- Social Network of people involved in the same topic
- Other meaningful chars as we get hands dirty

Potential Holdbacks:

- Despite our datasets appearing to be comprehensive now, they may turn out to have missing data or invalid entries as we dig in.
- Dealing with Reddit API may be tedious and time-consuming.
- We haven't learned text analysis or social network visualization yet, which may be a risk.
- Although data visualizations do shed light on our topic, they cannot demonstrate casualties.