You are what you eat

Following this well-known saying, science continues to search for, and the media eagerly reports on, miracle foods and diets thought to improve health. While a direct causative link between a particular food and health outcome requires clinical studies, this project aims to explore the link between diet and health worldwide, using country-level data. You will explore the national supply of over a hundred food items, and look for relationships with the incidence of non-communicable diseases. Do popular theories, such as "eat your vegetables," and concepts like the Mediterranean diet hold up at the global level? Is there a "super food," or combination of foods, that results in improved health?

Milestones

1. Project selection: Form teams of 2 or 3 and select a project from the provided list.
2. Literature study: Go through the following resources for background on the project and write a half to 1 page summary for each one
    a. Willett, W. (1994). Diet and health: what should we eat? Science, 264(5158), 532–537. http://www.prochange.com/media/diet_and_health.pdf.
    b. We're So Confused: The Problems With Food and Exercise Studies: http://www.nytimes.com/2016/08/11/upshot/were-so-confused-the-problems-with-food-and-exercise-studies.html
3. Data exploration and cleaning: The data sources for this project are food supply organized by country, from 1961-2013 for:
   Crops: http://faostat3.fao.org/download/FB/CC/E
   Livestock and fish: http://faostat3.fao.org/download/FB/CL/E
   as well incidence of non-communicable diseases (cancer, cardiovascular disease, and diabetes), also by country, for 2000 and 2012: http://apps.who.int/gho/data/node.main.A859?lang=en
    a. Decide on a suitable database to store the data, and on a computing resource to process the data.
    b. Read the data descriptions on each web site. Describe how the data was collected and what is being reported.
    c. Determine which countries are common to the disease and food supply data and create a combined table with all data.
    d. Write functions to sort and visualize the data for any given food type or disease
        i. Graphically, e.g. bar chart
        ii. In a map interface
    e. Check for correlations between each food item, or aggregates of items, and disease incidence.
4. Proposal: Propose methodologies and ideas to be implemented, tested and interpreted for your final project.
    a. Decide on a *performance metric* to evaluate the relation between diet and health.

b. *Feature extraction*: extract a basic set of features from the food supply data that you will use to predict the disease data.
c. Implement the following baseline techniques:
    i. *Null model*: disease incidence in a country is the same as the global average
    ii. *Linear regression*: predict health outcomes using the features you selected
d. Other project essentials:
    i. See if you can improve on the linear regression predictions.
    ii. Test specific hypotheses that you generate, e.g. “eating more vegetables is healthy.”
    iii. Search for relations between health and diet that have no hypothetical motivation. Is there a mystery “super food?”
    iv. State any caveats underlying your results, i.e. other factors that affect health besides diet, and how you would address them.