Wisdom Akpan
CS 314
Final Project

The question I sought to answer was initially what is the relation if any between produce availability and various diseases. In other words, does an apple a day really keep the doctor away? But due to difficulties encountering in my data scavenging phase I switched to an analysis of relationship between the per capita availability of various fruits and vegetables and the per capita per day intake of an assortment of vitamins and minerals. Or in other words, is this apple really a part of a balanced breakfast?

For this assignment data sets of produce and nutrient availability were downloaded in CSV file format from Data Planet's Statistical database.
My initial proposal for this project would have drawn data from the CDC for disease totals but the lack of usable file formats forced me to reconsider my project.
(http://statisticaldatasets.data-planet.com/dataplanet)

In this visualization dots in a time series for each of the ten nutrients chosen for this project were plotted and overlaid onto a bar chart that conveys the pounds available per person per year of a produce over the course of a series of years.

The data density of this vis is quite high. The height of the bars encodes the amount of produce available. The y value of the dots encodes the amount of the amount of nutrient per person per day. The scales for both produce and nutrient were represented on the right and left sides of the y axis respectively. The location of both bars and dots in the x direction encodes the year the samples were measured. Only one set of produce data can be visualized at a time so color encoding isn't necessary, but that isn't the case for the nutrients, so each nutrient is categorized by a color that is shared between the dots and the checkbox label that refers to that series of dots. Throughout the term height on a fixed axis was consistently emphasized as an effective way of encoding exact quantities and color was noted as a great way to convey category. As a result of following these design principles the various dot series can be easily differentiated and the values of various dots and bars can be compared and contrasted intuitively. Overlaying similar data is a good way to minimize the amount of space used on conveying information so it was implemented in the design of this vis.

Each dot and bar is a national average. Just like in previous projects the samples of the survey are far too large to reasonably represent so instead of downloading the set of samples aggregates for each year were used. Unlike past projects the aggregates weren't created accurately since Data Planet already provided their own. The variety of categories of data to compare has many permutations so in order to allow for users to investigate as their curiosity leads filtering of nutrient data sets was instated with color labeled checkboxes and a geometric zoom activated by mouse scroll and mouse drag. These features allow for users to focus on data they find most relevant.

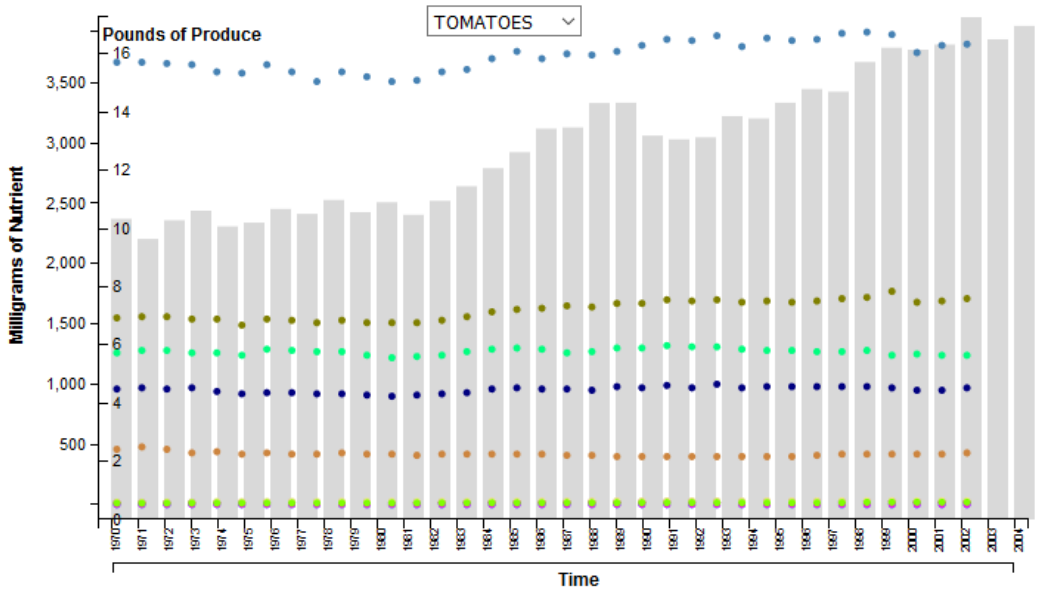
Other interactive elements include a drop-down selector to choose which fruit or vegetable you want to see graphed, a color highlight when a bar is moused over, and a feature that highlights all of the dots with the same year value and details the nutrient average and year of the dot specifically moused over in a popup caption. These interactive features were chosen because they follow the conventions of interactive websites I've encountered in the past.

Feedback received earlier in the design process greatly influenced the final visualization. During the critique, I was advised to include clarifying mouse over boxes, utilize geometric zoom/translational scrolling, and per capita units all of which were included in the final design. Other suggestions like comparing multiple sets of produce simultaneously, utilizing semantic zooming, and limiting total filters were considered and even tried but were eventually deemed luxuries due to the fact they demanded more time and expertise then I had available.

Semantic zoom, in particular, would have been really effective because it would adjust the scales and spacing of dots proportionally to the zoom. This means new trends could be unearthed for plot series where the values are similar relative to the default axes, and plots that were basically straight dotted lines could become more dynamic looking once you take closer look. As things stand you can still find those trends by mousing over a bunch of points in a row and comparing the individual values but this requires a lot of cognitive effort that a good visualization should bypass. If I had time I would certainly add that since it would contribute a lot to answering the initial questions.

In totality, even though it is not without areas for improvement, I think this visualization does a satisfactory job allowing for the user to explore the relationship between produce availability and nutrient intake. The channels used for encoding are intuitive and effective and the tools for exploration are multifaceted and useful.

Relationship between Produce Availibilty and Nutrition



☑ Iron ☑ Calcium ☑ Potassium ☑ Vitamin E ☑ Vitamin A ☑ Vitamin B6
☑ Dietary Fiber ☑ Cholesterol ☑ Phosphorous ☑ Niacin ☑ Sodium