# You're Gonna Die: Diet Edition

Mat Kelly, Sai Chaitanya Tirumerla, and Ibrahim Ben Mustafa
Department of Computer Science
Old Dominion University

Fig. 1. The final visualization allows a user to enter data in the left pane based on their personal and diet attributes to produce a plot showing the effects on each diet attribute and data about when he or she will perish based on estimates and an aggregation of many data sources.

**Abstract**—What you consume can kill you, or at least have an effect on how long and well you live. In this paper we put aside other attributes that have an effect on your lifespan and focus on diet. We created and will describe a form-based graphical visualization that allows a user to enter his traits and dietary habits to have an estimate generated for the time left to live and the quantity of time each diet attribute has on lessening this number.

## 1 INTRODUCTION

Selfish as humans are, we are primarily concerned with ourselves and aspects of the world that affect each of us. As one of two inevitabilities in the world[1], many are concerned with the question, "When am I going to die?", with the other question's answering being a consistent "April 15". As a simple date is often not convincing, many seek a way to provide justification for answers to this question that might seem arbitrarily derived. As documentation would be extensive, we wish focus on one aspect that contributes to eventual death, the effects of one's diet on mortality. In this paper we will describe a visualization that allows a user to specify a select set of consumption trends that they exhibit and from this, combine the data as an input to a system involving their demographic data to output a date of death with that visually describes the causal relationship between what one consumes and when one will perish.

## 2 RELATED WORK

The causal relationship between death and consumption (what one eats, though tuberculosis, too can be fatal) is one that is a naïve of the other factors of one's life that affect mortality though is not negligible like attempting to estimate based on accidental death or the probability of being affected by an Act of God. While digging too deeply into data about one's diet (e.g., the contributory effects of the inadequate consumption of a select set of amino acids causing muscle degradation and eventually death) is well beyond the scope of a

simple way to convey an estimate (amino acids mean statistically little compared to being struck by lightning), we still wish to correlate its effects. Other for-fun efforts have been done in this regard, particularly with the famed The Death Clock[2] website, which takes into account aspects of one's lifestyle like Date of Birth and combined it with lifestyle choices like whether one smokes and one's Body Mass Index. As Death Clock sets a good premise on an easy-to-use black boxed estimation system, we adopted this form-based input method to something with a greater plot appeal.

## 3 APPROACH/DESIGN

In information visualization, one is normally supplied data and derives a visualization from it. We took a different approach here, developed a visualization, examined the data, and adapted our visualization as we discovered more data or realized some data did not exist in an extractable manner (e.g., ambiguously and solely displayed in plots where value precision was difficult). We wished to enable the user to supply data in and easy-to-use form-based system, enabling it to be user-driven rather than restricted to pre-set values. Allowing for the exploration and providing interactive facilities are key facets of a user-driven rather than author-driven visualization [1].

We were limited in adopting Segel's principles yet strived to do so by seeding our efforts with an initial design that was extensible to these approaches. Figure 2 shows our initial bare-bones design while Figure 1 shows our end-product. Note in the initial design how we allowed for Segel's **annotations** though this idea was eventually scrapped for simplicity (at the detriment of clarity) due to data restrictions (Section 4).

Some aspects of our initial design *persisted* to the final product while others we *removed* and some new ones *added*. From the initial design, we imagined a user using the form on the left (persisted) to

---

- *Mat Kelly, Sai Chaitanya Tirumerla, and Ibrahim Ben Mustafa are graduate students of computer science at Old Dominion University, e-mail: {mkelly, stirumer, iben}@cs.odu.edu*
- *Live visualization:* `http://www.cs.odu.edu/~mkelly/semester/2013_spring/project1/`

---

[1] Per the quote, "The only things certain in life are death and taxes."

---

[2] `http://deathclock.com`

enter their age, ethnicity and ZIP code. The age and ethnicity would be used for a baseline mortality lifespan (persisted), which we later realized also varied by gender (added). The ZIP code would be used as an input for other location-based variances and would be restricted to Virginia for simplicity and mapped (removed). This proved difficult, as much of the data was based on national surveys or metrics.

The inputs below the user's attributes would provide a numerical input facility for users to enter data about their diet (persisted). We had not defined these in the initial design, which proved to be a major hurdle in the project. Once the user's attributes (and optionally, the user's diet attributes) had been entered, a plot would be created to display the degradation of their health over time (removed). The right edge of the X-axis would be the value obtained of average lifespan based on the user's attributes with the black line being the value derived from all user and diet attributes, red being if the diet attributes were exacerbated and the green line representing the potential lifespan if the diet attributes were to improve. This method for displaying the diet on the left pane was scrapped due to insufficient data being available and an alternative approach used.

The final user interface element was a "Lifestyle slider" atop the graph (removed). For one to move this hypothetical slider would be to represent a lifestyle change toward indulgence versus consciousness. Through this we wished to instill Segel's feature of **matching on content** via the aligning of the far extremes of the slider with the representative lines in the plot [1]. The gradient, we thought, would be intuitive, as the black "baseline" was gradated from each color with a representative label of "sustain" on the slider giving more insight into the plot lines. This feature, also, was removed for lack of data.

To replace the plot in our final product, we chose to adopt the "right-side equals lifespan" attribute of our initial design but chose instead to pivot each diet attribute from the right side. Though we were a little limited in the software used, the distance a line color span from the right Y-axis is its effects on lifespan, as gauged by the X-axis. An additional life is provided to serve as the cumulative sum of these values and a more descriptive message with the results of the left-pan parameters adorns the top of the plot to account for the plots lack of clarity. Each line color is represented in the key on the left of the plot, as is standard with plots with multiple lines. It should also be noticed that the meaningfulness of the Y-axis is naught in the final design whereas it provided the qualitative "Quality of Life" metric in the initial design. As this value was subjective and sparse in the data set, we abandoned it yet still (semi-)effectively applied one dimensional data to a line plot by using the Y-axis to distinguish lines that would otherwise overlap.
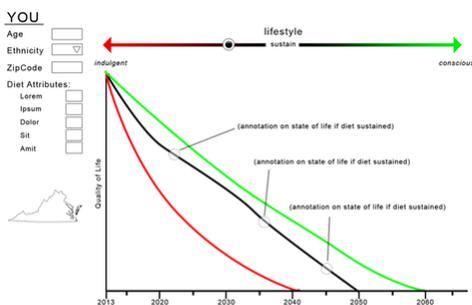
Fig. 2. Our initial design (contrast to our final design in Figure 1) was more ambitious but could not be supported by data discovered. Note the additional user interface elements that were removed as well as the lack of definition of the attributes we intended on eventually discovering.

A further feature we added was to allow the user to estimate values (e.g., sugar consumed from sweetened drinks per day) based on a collection of intuitive inputs through a easy-to-use interface (Figure 3). Unlike many dietary calculators, where this field might not be populated or populated inaccurately due to most people's lack of com-

prehension of the nutritional aspects of food consumed, this allows the user to get a more accurate input value, which would lead to a more accurate end-result
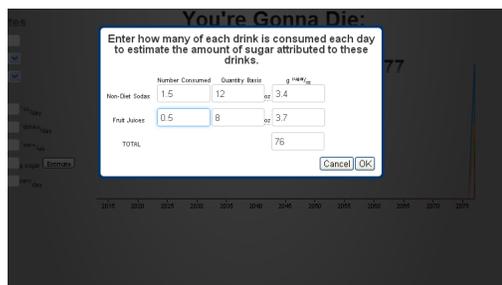
Fig. 3. A popup allows the user to estimate the amount of sugar consumed based on the quantity consumed of both fruit juices and non-diet sodas. Each field is cusomizable if a low-sugar version of the drink is used instead and the value of quantity of each class of drink's sugar content is pre-filled in the drink's respective row.

## 4 THE DATA

The data that we supposed would be available was not. The primary source of the data was the Center for Disease Control and Prevention (CDC)'s website. Determining what data would be useful required us first to determine whether the data was demographically based (which would be useful for the user attribute portion of the left pane) or casually based (e.g., diet causes diseases causes death), with the latter being extremely sparse. CDC provides a long list of tables[3], from which we utilized the following:

- Death rates for diseases of heart, by sex, race, Hispanic origin, and age: US 1950-2010 (Table 30)

- Death rates for Cancer by sex, age, race: US 1950-2010 (Table 32)

- Death rates for cerebrovascular diseases, by sex, race, Hispanic origin, and age: US 1950-2010 (Table 32)

- How the age can increase the possibility of having serious diseases: US 1980-2010 (Table 49)

Other insightful data that was interesting and relevant yet not useful from CDC were Leading Cause of Death for Men[4]/Women[5], the Nutritional Health and Nutritional Examination (NHANES III) Survey[6], and many others. We found many other non-useful sources that lead in dead-ends from the likes of Cancer.gov[7] and the National Institute of Health's (NIH) National Center for Biotechnology Information[8]. After a bit of frustration in finding dead-end sources, we organized our data and attempted to extract facts from those articles whose provided data was not tabular as well as those documents whose data extended beyond our scope. We found reputable sources from which were could infer facts like "One or more fried foods weekly causes an 18% increase in mortality." Identifying these for each of our diet attributes of:

1. Red Meat Consumption ($> 3oz$ daily causes increase of 15% in mortality)

---

[3] http://www.cdc.gov/nchs/hus/contents2011.htm
[4] http://www.cdc.gov/men/lcod/2009/index.htm
[5] http://www.cdc.gov/women/lcod/2009/index.htm
[6] http://www.cdc.gov/nchs/nhanes/nh3data.htm
[7] http://progressreport.cancer.gov/doc_detail.asp?pid=1&did=2007&chid=71&coid=731#trends
[8] e.g., http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2916052/#R13

2. Sugar Drinks (>11oz, 10%)

3. Fried Foods (> 1/week, 18%)

4. Alcohol (> 2 for men, > 1 for women, 5-7%)

5. Refined Grains (> 1/day, 10%)

allowed us to convert the inline data into something that was able to be plotted.

## 5 IMPLEMENTATION/TOOLS

The bulk of the implementation is done in JavaScript, HTML and CSS. Data cleaning was all done manually, which proved tedious but more accurate. jQuery is used extensively to interact with DOM elements. The jQuery[9] plugin "Foggy"[10] is used to provide an initial blurring effect on the graph prior to data being ready. Twitter Bootstrap[11] is used for a consistent user-interface and fancy text input field highlighting on focus. D3.js[12] is used for the graph display, which provided a surprisingly degree of inflexibility in the way we wanted to display the data (hence the superfluous lines). Adobe Photoshop[13] was also used for miscellaneous graphical manipulation.

## 6 CONCLUSIONS

The design we created is not entirely clear though "allows a user to observe through a visualization the effects of dietary consumption on morality", as we originally set out to design. The shoot-first-ask-questions-later approach of design and data acquisition (respectively) caused the project to require a lot more time and effort with little benefit besides satiating our initial desire (and desire alone) to imagine an implementation that would be both locally relevant and useful to many.

## 7 FINAL THOUGHTS

The project was an interesting yet frustrating endeavor in working as a group to create a visualization. Even with the assignment of one student per group who had taken the Information Visualization course, that did not prove to be a barrier in lieu of taking some must time to overcome the first behemoth of data acquisition and cleaning.

### 7.1 Contribution of Each Group Member

We turned circles for the longest time, first acquiring many sources of data only finding many were not useful after-the-fact (and way too close to the deadline). At crunch time, Chaitu and Ibrahim sorted through our large data collection to figure out what would be useful. Ibrahim performed further efforts of acquiring citable phrases for each attribute we had chosen. Chaitu worked on documentation and presentation content, largely derived from our initial design proposal. Mat created the initially proposed design, the final design and did all of the coding.

### 7.2 What Was Learned

What we learned was not so much in the realm of visualization as it was in trying to organize a project and work as a group. Our methods of finding data and organizing it to support the hypothesis we had was haphazard and had us chasing our tails thus restricting the time we had to code the end-product. The dependency of a good visualization requiring good data and vise-versa popped up many times, which led us to be frustrated that we could not get the project to progress.

## REFERENCES

[1] E. Segel and J. Heer. Narrative visualization: Telling stories with Data. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1139–1148, 2010.

---

[9] http://jquery.com/

[10] http://nbartlomiej.github.com/foggy/

[11] http://twitter.github.com/bootstrap/

[12] http://d3js.org/

[13] http://www.adobe.com/products/photoshop.html