Data Visualization for Medical Price Education and Transparency

Medical Price Transparency and Education Through Data Gathering and Visualization

Alex Sciuto

Jodi Forlizzi and Aisling Kelliher
Advisors

Data Visualization for Medical Price Education and Transparency

Medical Price Transparency and Education Through Data Gathering and Visualization

A thesis submitted to the School of Design, Carnegie Mellon University, for the degree of Master of Design in Interaction Design

Alex Sciuto *author* ⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯

Jodi Forlizzi *advisor* ⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯

Aisling Kelliher *advisor* ⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯

abstract

The health care system in the United States is changing rapidly. Individual patients are expected to become educated medical consumers making informed choices and paying for those choices. Many researchers and designers are studying how medical consumers understand their medical care, but there is an opportunity for meaningful design strategies using data visualization to help consumers understand how much they pay for their care. This thesis uses service and user-centered design methods and interactive data visualization to create systems that gather medical prices and display them back to users all with the goal of creating more educated medical consumers.

Introduction

MOTIVATION

Florence Nightingale, the mother of modern nursing, is widely known for the reforms she advocated in the second half of the nineteenth century. She led efforts to improve the care of wounded and sick soldiers in the Crimean War, and after the war she founded a nursing school in London for the training of skilled nurses. She is less well known for her visual thinking and excellence in the field of statistics. She realized that using visual means to communicate complex ideas would help "to affect thro' the Eyes what we fail to convey to the public through their word-proof ears." She used data graphics about sick and wounded in the Crimean War to convert incomprehensible statistics into meaningful information that moved people who viewed the graphics to change their actions.
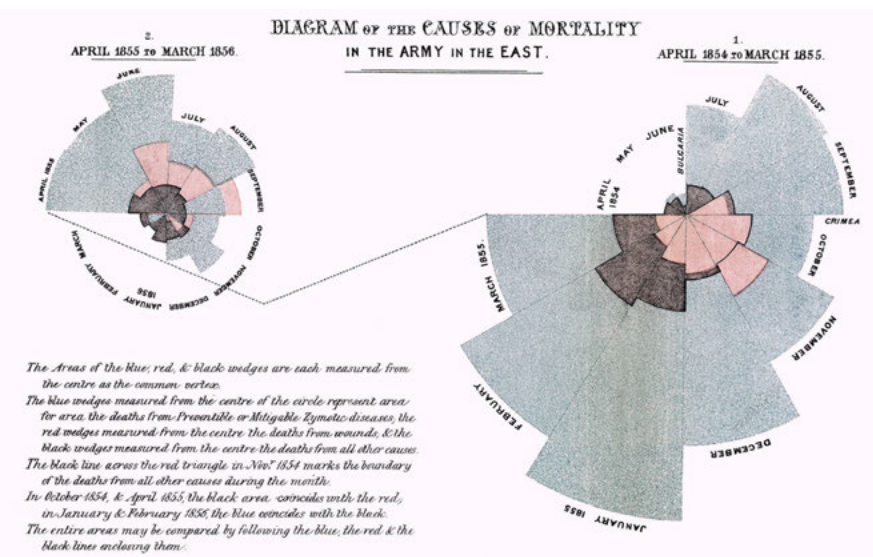

Florence Nightingale. "Diagram of the causes of mortality in the army in the East". 1857.

Gelman, Andrew and Anthony Unwin. "Infovis and Statistical Graphics: Different Goals, Different Looks"

2013. Journal of Computational and Graphical Statistics

This thesis project is motivated by two impulses: a deep love and respect for the communicative power of well-designed visual information and a belief that this power should be put to use to substantially and positively impact its users. I have chosen to focus on medical price visualizations because it is an area where well-designed information graphics can make a large, positive impact on medical consumers as they make medical choices. Also, while much work has been done on medical data visualizations, there is a lack of design work that attempts to make medical prices more understandable.

This thesis is important because it suggests an expanded role for information visualization by combining information design with service design and methods of crowdsourcing. This method could be applied to many other problems where data is decentralized, fragmented, and hard for customers to get.

THE PROBLEM

This thesis is also important because of the timeliness of its content. The Affordable Care Act of 2010 was designed to provide Americans with better health security by modifying how United States citizens purchase health care. The law's stated goal is to expand coverage, hold insurance companies accountable, lower health care costs, guarantee more choice, and enhance the quality of care for all Americans. As of May 2014, nearly 20 million people have gained health care with the help of the Affordable Care Act.

The Affordable Care Act has created a new complex ecology of health care providers, costs, and networks for receiving care. Overall, patients are bearing more financial risks. This was a vision of the ACA. People are now supposed to be more cautious about health care spending, but it is nearly impossible to search for and monitor costs. The end result is that people get bills that surprise and anger them making them feel more alienated from their medical professionals.

New tools and services are necessary to educate and engage in their new role as medical consumers. The projects in this thesis all attempt to bring participants into this new role.

OVERVIEW

This thesis is composed of three projects:

PGHMedicalPrices.us A website that solicits users to upload photos of their medical bills.

CareShare Service Concept A service design concept that lets individuals track their medical care expenses and make choices about where to go for care.

Pittsburgh Price Explorer An alternative method for gathering and disseminating medical price data.

CREATE　　AQUIRE　　PARSE　　MINE　　FILTER　　REPRESENT　　REFINE　　INTERACT

1 PGHMedicalPrices.us

2 CareShare Service Concept

3 Pittsburgh Price Explorer
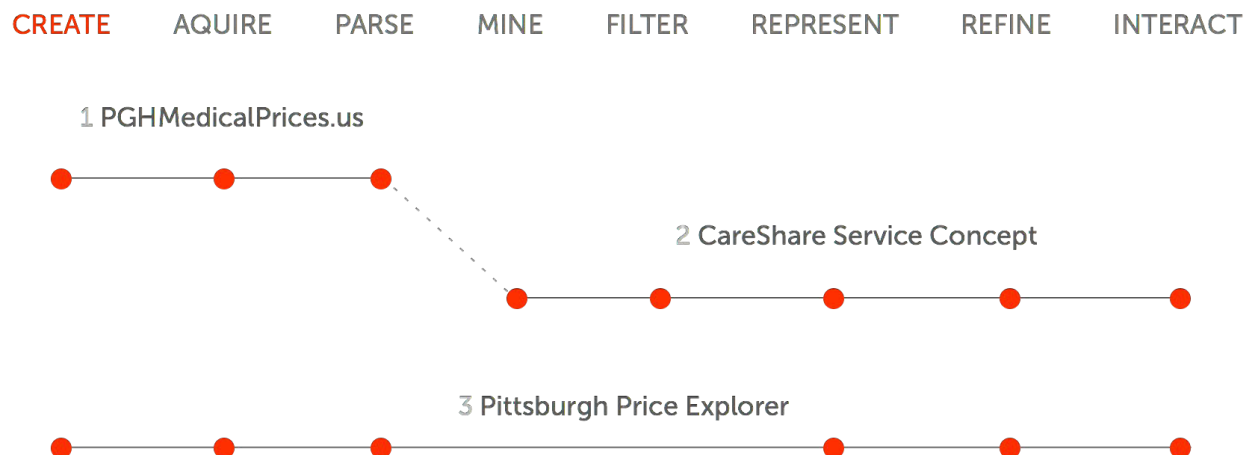
Ben Fry's phd thesis lays out a seven step process for creating data visualizations. In this diagram, Fry's framework is applied to the three projects comprising this thesis.

The thesis documentation is composed of four main parts:

Literature and Commercial Review A review of current literature in crowdsourcing, web-based data visualizations, and health care 2.0. This also includes an overview of current commercial products combining data visualization, crowdsourcing, and health care.

Process Themes An overview of the most important lessons and themes from this process.

Process Narrative An overview of the process of creating the final artifacts.

Design Solution Images and a description of the final three design artifacts.

The thesis makes five contributions:

Local Crowdsourcing Crowdsourcing is practice of using groups of anonymous workers to perform tasks computers cannot complete. Typically, these crowds are anonymous. This thesis suggest a use for local, context-specific workers.

Conversing with Data Exploring new ways to bring data visualization into the iterative design process.

Levels of Fidelity for Data Visualization Enumerating different levels of fidelity of design when creating data visualization.

Creating Data for Visualization Visualization has historically not been concerned where the raw data comes from. This thesis argues that visualizers should care where the data comes from so that they are not stuck only visualizing easy-to-access data.

Reflecting Data Back to the Users Successful data visualization should provide a mirror to let the users see themselves in a new ways.

See Shneiderman, Ben. "Inventing Discovery Tools: Combining Information Visualization with Data Mining." 2001 and Fry, Ben. "Computational Information Design." 2004. for two examples of visualizers giving overviews of the visualization process without considering where their data comes from.

Literature Review

In this section, I will give a brief overview of the three academic areas that are closely related to my thesis: crowdsourcing and distributed work, online information visualization, and health 2.0.

CROWDSOURCING

crowdsourcing and design

Crowdsourcing is the practice of enlisting many humans to perform simple tasks that when combined solve problems that a computer or individual person could not solve. Crowdsourcing goes by many names that highlight the multitude of uses crowdsourcing has: peer production, user-powered systems, user-generated content, collaborative systems, community systems, social systems, social search, social media, collective intelligence, wikinomics, crowd wisdom, smart mobs, mass collaboration, and human computation. Doan, Ramakrishnan, and Halevy explicitly define crowdsourcing broadly as any system as long as "it enlists a crowd of humans to help solve a problem defined by the system owners."

Doan, Anhai, Raghu Ramakrishnan, and Alon Y. Halevy. "Crowdsourcing systems on the World-Wide Web"

2011. ACM.org

The origins of crowdsourcing depend on how one chooses to define it. Architecture competitions and prizes for design work have been held for decades, and these projects that enlist many teams to create sets of proposed designs for a building or product can be considered crowdsourcing. An example of such an architecture competition is the design of the Sydney Opera House in 1957. Jørn Utzon won five thousand pounds in 1957 when his design was selected for the iconic hall. Usually crowdsourcing is considered a new method of bringing people together to solve a common problem, but it is much older than the Internet.

Murray, Peter. The Saga of the Sydney Opera House: The Dramatic Story of the Design and Construction of the Icon of Modern Australia

2004. London: Spon Press

Designers have employed crowdsourced systems in their practice in various ways. One popular use of crowdsourcing is to perform user studies. Crowdsourced user studies have several benefits. They are cheaper and easier to perform than traditional user studies. They can also reach a larger or more targeted audience than a recruitment-based study. Crowdsourced systems for design follow the model of asking a user a series of objective questions either about themselves or about a prototype they have been shown, then asks them a series of questions to understand their response to the prototype. Using this method, a designer can get feedback or iterative inspiration from a great variety of people the designer would normally not have access to.

Kittur, Chi, and Suh. "Crowdsourcing user studies with Mechanical Turk"
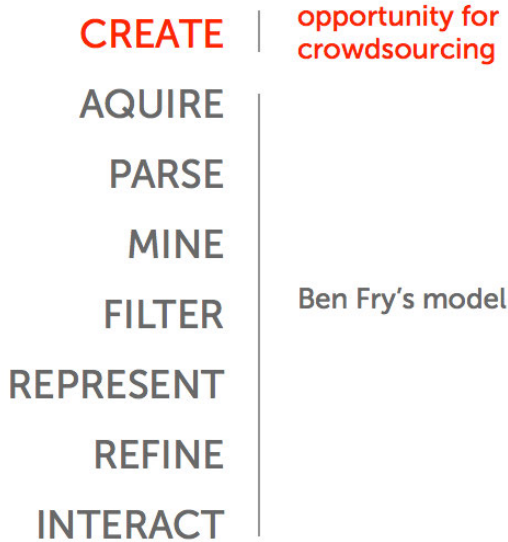
2008. ACM Digital Library

Others have incorporated a variety of crowd-sourced features into software systems. Soylent, a series of add-ons for Microsoft Word, offers editorial functions powered by crowdsourced workers. These functions include shortening paragraphs while maintaining meaning, proofreading grammar and syntax, and offloading arbitrary word processing such as formatting citations. Be My Eyes, a mobile app, connects crowdsourced

volunteers with blind people who need help with short, time-sensitive tasks. The app translates the blind person's phone's video feed into spoken conversation. Be My Eyes makes it transparent how the crowdsourced working is helping. Duolingo is another crowdsourced service for translating content into new languages. Unlike Be My Eyes, Duolingo offers crowdsourced workers an experience of learning a new language, using the output of that experience as the input of its crowdsourced system.

Bernstein, Michael S., Little, et al.. "Soylent: A Word Processor with a Crowd Inside "
2010. ACM Digital Library

Be My Eyes
2012.

Savage, Neil. "Gaining Wisdom from Crowds."
2012 Communications of the ACM

CREATE — opportunity for crowdsourcing
AQUIRE
PARSE
MINE
FILTER — Ben Fry's model
REPRESENT
REFINE
INTERACT

crowdsourcing for gathering data

Designers use crowdsourced systems to complete a variety of tasks that a computer alone could not achieve. Similarly, crowdsourcing offers visualizers a method of gathering difficult to obtain data. Crowdsourced systems have a number of attributes that make it attractive for visualization. Crowdsourced systems turn human beings into sensors that can do complex analysis of the environment around them. A crowdsourced system can ask complex questions like "Is the area around you crowded and if so, what is your location?" The human workers in a crowdsourced visualization system can also perform creative or difficult tasks to create the data. For example, a system may ask workers to draw a picture or find a difficult piece of data in their homes.

Two visualization practice areas that have used crowdsourced systems to power visualizations are news visualization and artistic visualization. Both of these practice areas offer visualizers flexibility to try new techniques and have quick project timelines. Two visualization projects that use crowdsourced data are the Aaron Koblin's "Ten Thousand Cents" project and an innovative opinion graphic in the New York Times on the death of Osama bin Laden.
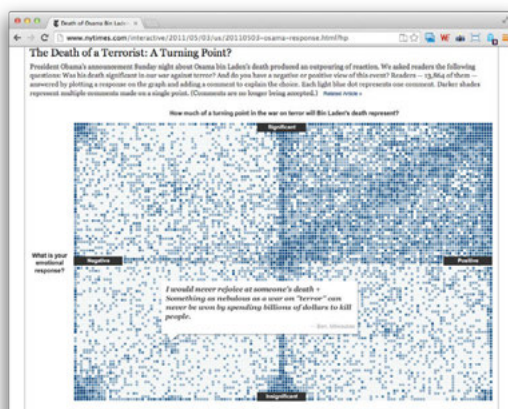
When Osama bin Laden was killed in 2011, the news caused a huge uproar in the United States. There were many discussions, particularly about the importance of the event and whether his death (as opposed to capture or imprisonment) was a good or bad thing. The New York Times graphic team produced "The Death of a Terrorist: A Turning Point?", a crowdsourced system that let readers mark their opinion along a two-dimensional grid and also see what others felt about bin Laden's death.

The visualization got thousands of responses, and the corresponding grid of responses brought out interesting emergent patterns of how people answered. People tended to align their answers to the center-lines or edges of the grid. The extreme points also got a large number of

selections. This implies that a simpler, three-by-three grid would have been just as effective, but the detailed grid made it more obvious just how many viewers gave their input.

Aaron Koblin's Ten Thousand Cents project is an early piece (2008) that used crowdsourced data to gather novel data for use in a visualization. Koblin paid 10,000 crowdsourced workers a small amount of money to draw a small part of an image of hundred dollar bill. The drawings were then combined into an interactive quilt.

Ten Thousand Cents uses crowdsourcing to gather data—the paths and colors generated by an online drawing program—that would never be gathered by other means. The piece also shows how the rich variety of responses still creates a recognizable dollar image.

localized crowdsourcing: an opportunity

Crowdsourced systems today do little to filter those who participate in a crowdsourced project. Amazon Mechanical Turk, the most common platform for crowdsourced programs, allows the designer to indicate desired skills the crowdsourced workers should have as well as the crowdsourced workers' country of origin. For designers who use Mechanical Turk, as long as these basic levels are met, every worker is considered interchangeable.

Good designers though are focused not on interchangeable and anonymous groups of users, but instead design for concrete situations, scenarios, and users. The need for quality feedback not just from skilled workers but from skilled workers grounded in a particular context is not being met by Amazon Mechanical Turk or any other crowdsourced system. This is in part because the desire for specific participants violates the primary strength of crowdsourced systems, their large number of users.

Stolterman, Erik. "The Nature of Design Practice and Implications for Interaction Design Research"

2008. International Journal of Design

There is an opportunity for combining the best of small-scale personally recruited design methods with crowdsourced automation. These hybrid systems would allow designers to recruit targeted people either online or offline and lead them into an automated process where their membership in the desired community would be confirmed. The selected users would them complete tasks.

This project attempts to use a form of this hybrid crowdsourced system. Instead of relying upon small payments to anonymous users, the study attempted to recruit local participants to PGHMedicalPrices.us where they could upload a photo of a medical bill. Participants were offered incentive price gift cards to a local chain of grocery stores.
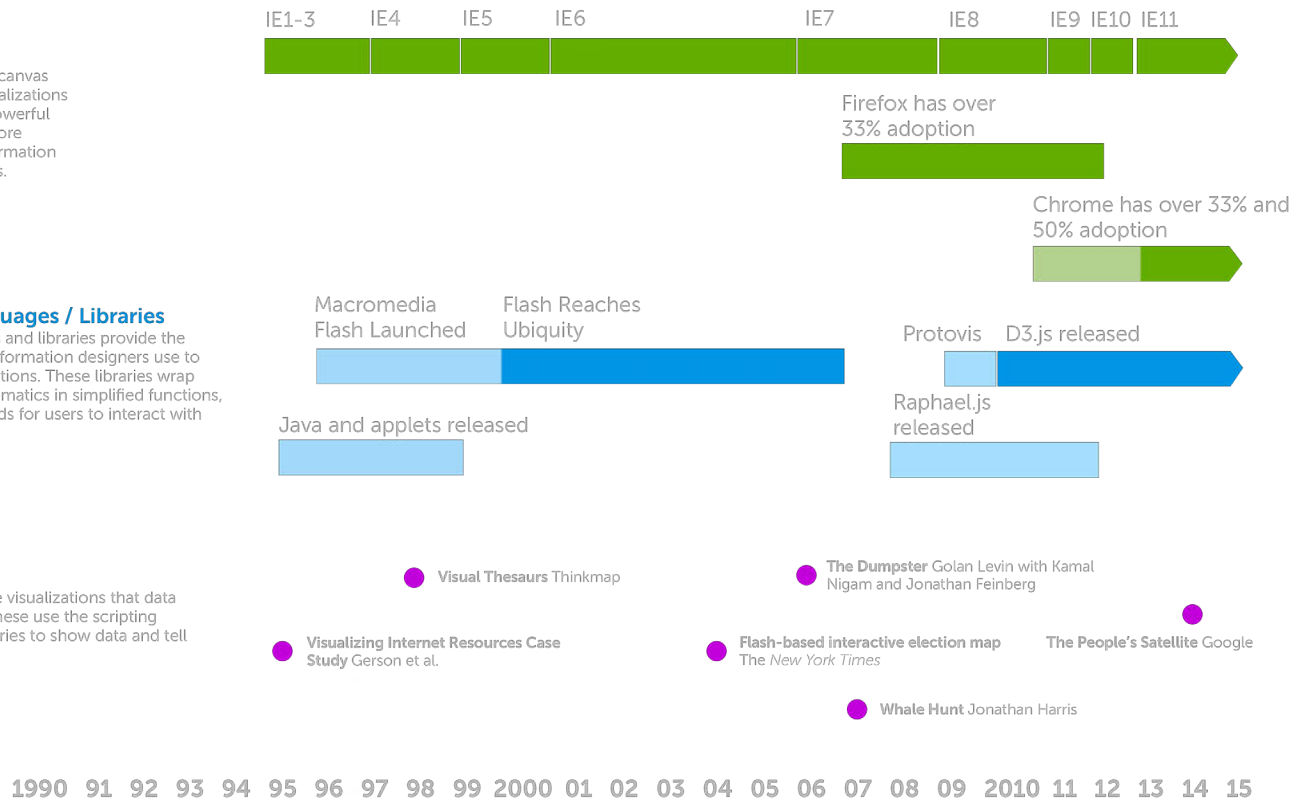
Browser

The browser is the canvas on which all visualizations take place. More powerful browsers enable more expressive and information information designs.

Firefox has over 33% adoption

Chrome has over 33% and 50% adoption

Scripting Languages / Libraries

Scripting languages and libraries provide the functionality that information designers use to make their visualizations. These libraries wrap complicated mathematics in simplified functions, and provide methods for users to interact with the visualization

Macromedia Flash Launched

Flash Reaches Ubiquity

Protovis · D3.js released

Raphael.js released

Java and applets released

Applications

Applications are the visualizations that data visualizers make. These use the scripting languages and libraries to show data and tell stories.

Visual Thesaurs Thinkmap

The Dumpster Golan Levin with Kamal Nigam and Jonathan Feinberg

Visualizing Internet Resources Case Study Gerson et al.

Flash-based interactive election map The *New York Times*

The People's Satellite Google

Whale Hunt Jonathan Harris

1990 91 92 93 94 95 96 97 98 99 2000 01 02 03 04 05 06 07 08 09 2010 11 12 13 14 15

Timeline of history of web-based visualization based on browser, visualization library, and noteworthy applications.

WEB-BASED DATA VISUALIZATION

Web-based information visualization has been a subject of research and practice since the early days of the world wide web. Tim Berners-Lee publicly debuted the world wide web in 1992, and by 1993, the first web browser, the Mosaic Browser by Marc Andreessen, gave web browsers a graphical user interface to access web content. The early world wide web focused on linking text and low-resolution images together in the form of hypertext documents. There would be many uncoordinated attempts to add graphical and interactive capabilities to web browsers necessary to enable web-based data visualizations. Eventually by the late 2000s, with the adoption of HTML5, CSS3, and javascript as a common standard for markup, styling, and interaction, web developers and programmers began creating visualization tools and using those tools to create richly interactive web-based visualizations.

early visualization experiments

Early experiments in visualization on the web focused on visualizing the web itself. Visualizers attempted to show the topology of the web in graphical form. Network graphs of servers, of web pages, of hypertext links were made. Visualizers also tried to map the web to physical geographies or organizational maps. Realizing that web page connections told something about the real world, by mapping the connections of web pages, visualizers could map other phenomena like the relationships between different departments at a university. Finally, early visualizers attempted to visualizer web user's cognitive models for the data they were viewing. This came in the form of interactive bookmarking applications, hot lists of useful websites, and other ways to divide content.

7
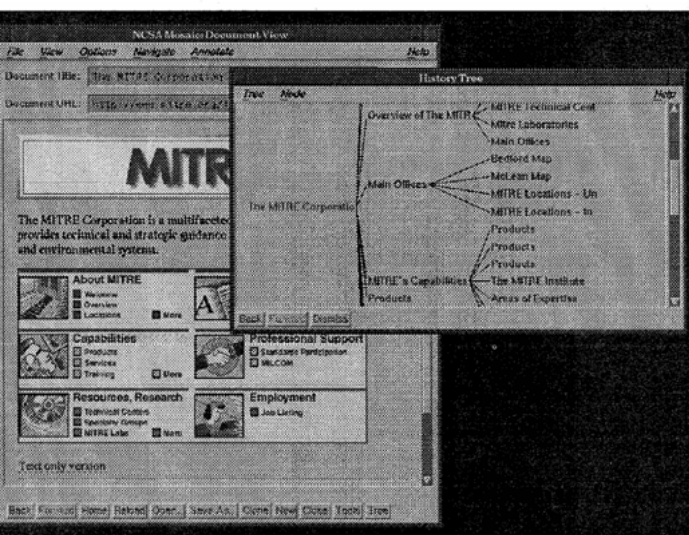
An example of this early kind of web visualization is Gerson et al's 1995 Visualizing Internet Resources Case Study. The case study included three different visualizations that were additions to the Mosaic Browser. The first visualization, "Hyperspace View," was intended to aid users in exploring the web without getting lost. Tree graphs were used to visualize the path the user traveled through and also to record alternative paths that were not taken. A second case study enabled browser users to interactively change the styling of links and where links pointed to make them more memorable and more personal. "For information resources to be effective and to enhance problem solving and analysis, they should allow each user to construct his or her own information space with links and associations." It's a radical response to the author-centric approach to the web. The final case study included analyzing gathered web content and correlating word pairs to highlight interesting content. The goal of this visualization was to make the already overwhelming amount of information available more easy to consume.

Gershon, N. et al. "Case study: visualizing Internet resources"

1995. Information Visualization Proceedings

Soon after, the power of web-based visualization to visualize datasets unrelated to the web itself was explored. Wood et al. created a 3D histogram visualizing environmental data. Wood et al. also gave four possible scenarios for how web-based visualizations could take place. They suggested visualization could happen as either an image or 3d file sent to the browser. They noted that this is the standard method of visualization circa 1996. They then suggest that the raw data could be sent over the web, and the user or their own software could be responsible for the visualization. Finally, Wood et al. suggest an interactive model where the visualization would be created by the producer of the data but the user could modify and change the visualization. It's interesting to see these different scenarios, because two of them became prevalent and two scenarios are not used. Today in the mid 2010s, images are still widely used to quickly convey visualizations over the web (see the figures in this thesis for example). The most interesting visualizations today use the fourth model Wood et al. suggested. Scenario three, where the user is responsible for the visualization has not been adopted. But some applications, like Tableau, Processing, or pre-made D3.js graphs templates foreshadow that user-controlled visualization may just have had to wait longer for its time to come.

Wood, J., Brodlie, K., Wright, H. " Visualization over the World Wide Web and its application to environmental data"

1996. Visualization Proceedings

These early experiments in web visualization were primarily concerned with the possibilities of what technology could deliver. When viewing the work that was created, it is clear that little attention was put towards design, storytelling, or insight generation.

uncoordinated technology

When the web was created, it focused on the exchange of text documents and embedded images. Almost immediately, web users tried to add multimedia content and advanced styling. In order to satisfy this demand for greater visual sophistication, browsers began implementing non-standard HTML markup like <blink>, <center>, <spacer>, <stroke>. As expectations for web bounded ahead of agreed upon specifications, different technologies provided additional features that information visualizers used to make visualizations.

Introduced in 1995, the Java applet gave developers direct access to the hardware processing of a computer through the web browser. Visualizers could write code in Java, then run the applet in the browser of any computer that had loaded the appropriate free Java software. Early visualizations used applets to access the full power of the computer while still existing on the web. Even today, users who wish to share Processing visualizations must use Java applets to share them through the web. The benefits of using applets were myriad in the early web. These applets bypassed the limits of HTML, enabled greater interactivity, and offered developers greater computing power.
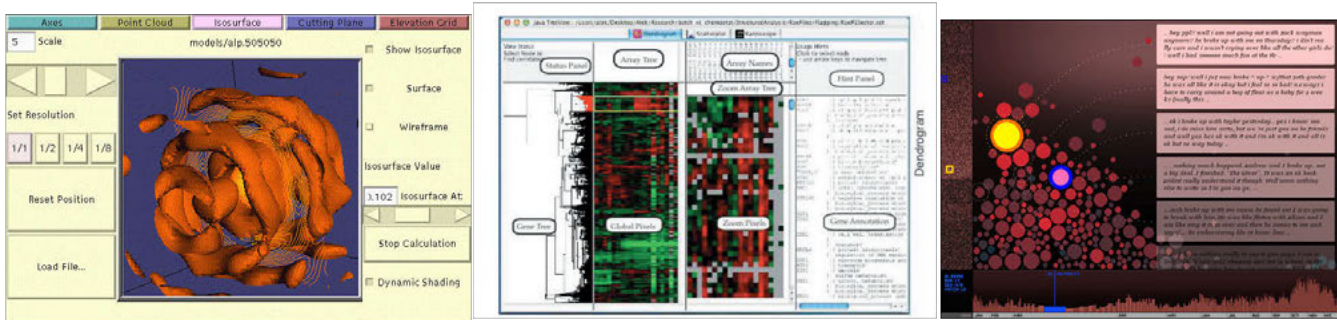
Three Java applets: VizWiz (1997), Java TreeView (2004), the Dumpster (2006)

Java applets were appealing, but they were not to be a lasting technology. For a variety of reasons, mainly the creators of Java's choice to sue Microsoft for its incomplete implementation of Java applets, some popular browsers didn't include applet support. In addition, Java had numerous security issues that made using applets dangerous. Browser makers responded to this danger by requiring users to jump through additional dialogs and warnings before running an applet. In the end, these drawbacks led to less and less use of Java applets. Today, outside of specialized and older websites, applets are not used.

Adobe Flash is a technology developed by Macromedia and later acquired by Adobe that added rich interaction and multimedia support to standard websites. Introduced in 1996, the Flash Player must be downloaded and installed by each individual user. This requirement meant that for a number of years, Flash adoption would grow until it reached ubiquity in the early 2000s. Flash offered similar features to Java applets, but Flash's creators focused on maintaining usability, security, and compatibility. A visualizer could know that if they created a visualization in Flash, it would appear correctly no matter the computer or browser used.

9
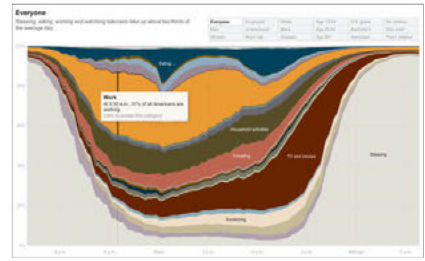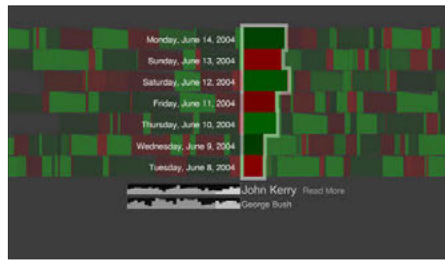
Three Flash apps: XML/SWF Charts (2003-2015), In The News (2004), How We Spend Our Day (2008)

Flash and Java applets were the two main formats used by early web-based visualizers. They circumvented the very limited design palette early web browsers offered. Because they were both non-standard, proprietary additions, their popularity eventually faded as open, standardized technologies caught up to the expectations of visualizers.

standards-based data visualization

Creating the standards that determine the capabilities of web content is a collaborative affair. The main stakeholders in the process, technology companies like Microsoft, Apple, Google, and nonprofit, open-source organizations like Mozilla and Opera, and individual engineers and developers, come together under the aegis of the World Wide Web Consortium, to develop agreed upon standards. Those standards are then taken by each organizations and built into the products they offer. As an example of the time frame for developing standards, HTML 4 was standardized in 1997 and updated in 2000. Work on what would become HTML5 began in 2004. the HTML5 standard was completed in 2014, ten years after beginning work. Throughout those ten years, browser developers incrementally added in support for HTML5 features as features were added and removed from the draft specification.

By 2010, web browsers had caught up enough to the rich features that Java applets and Flash offered, that developers and visualizers switched to developing standards-based visualizations. This standards-based visualization method is used for most visualizations today. These visualizations use a variety of technologies to deliver visualizations. Data is delivered over HTTP using JSON or XML data structures. SVG and Canvas are the browser-embedded graphics technologies that enabled visual expression, especially
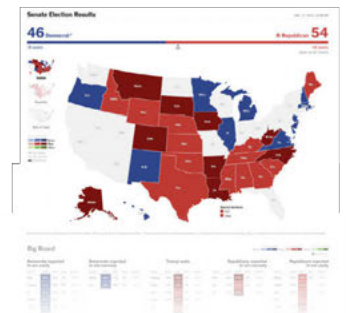
| 2000 | 2004 | 2008 | 2012 | 2014 |

Ten years of interactive electoral maps from the New York Times. In 2000, the map was a static image. In 2004 and 2008, the Times used Flash to make their map. In 2012 and 2014, the map was made using HTML5 web technologies.

for lines, shapes, and three dimensional rendering. Javascript has become expressive and standardized enough to translate received data into SVG and Canvas-based representations. CSS provided an interface to style elements on the screen. And HTML5 is the glue that keeps all these technologies together. Taken together, these technologies turn the web browser into a powerful, widespread, and easily accessible environment for visualizers.

Today visualizations using the web browser have three large benefits over using other technologies. Because web browsers are widely installed, a visualization created for a web browser can be run on millions of computers and mobile devices without any additional installation of software. Because web browsers are connected to the Internet, a single web-based visualization can be distributed to anyone with an Internet connection. Because the visualization lives on the Internet, the visualization can communicate back to the designer about what the user is doing with the visualization.

HEALTH 2.0

In the United States in the last decade and a half, the Internet has become widespread, routinely accessed, and accessed on a variety of devices. The ubiquity of the Internet in people's daily lives has affected many parts of how people live. How they access health care is one such area. The term Health 2.0 defines the set of practices, products, and services that have been created that use the Internet to connect patients to the medical system in new ways.

A large minority of Internet users today are taking advantage of such Internet-connected services. Twenty-seven percent of Internet users have tracked their weight, diet, exercise routine, or other health indicator online. Six percent of Internet users have posted questions or comments about health or medical issues to health websites. Eleven percent of social network users have posted questions, comments, or information to their social networks. These statistics show that large numbers of Internet users are using Health 2.0 technologies to record their health, research their illnesses, and connect with others based on their health conditions.

Fox, Susannah. "The Social Life of Health Information, 2011"
2011. Pew Research Center

Belt THVD, Engelen LJ, Berben SA, Schoonhoven L. "Definition of Health 2.0 and Medicine 2.0: A Systematic Review"
2010. JMIR.org

Boulos, Maged NK, et al. "How smartphones are changing the face of mobile and participatory health care: an overview, with example from eCAALYX"
2011. Biomedical Engineering Online

There are a number of ways that patients today are using the Internet to connect to the health care system. Patients use search engines to research ailments. They use online forums to connect with other patients and express their feelings. They use connected apps to track and communicate their health to others. They use review sites to learn about doctor quality. All these practices are nascent and new products and services are beginning to be developed that specifically serve the desires for people to connect and monitor their health.

All of these practices are marked by a number of characteristics. Belt et al. list themes of Health 2.0 medical literature:

> **Patients and Consumers** People who come into the medical system are both traditional patients who are taken care of by knowing doctors, but also consumers, engaged in their

own care and making decisions best for them. This theme emphasizes the empowerment of patients, a class of people usually considered to have little agency.

Web 2.0 Web 2.0 refers to the communication technologies that health care organizations have to connect people and share information. There is also an emphasis on co-opting or combining services that were not created for health care but could be used in health care.

Professionals or Caregivers Professionals refer to the groups of people in the health care industry who are not either doctors or patients. Caregivers refers to the family and friends of patients who help them during illness and recovery. Stakeholders like medical payers, health students, researchers, entrepreneurs, and government entities all play a role in Health 2.0.

Social Networking Social networks refer to online communities where people can virtually gather and share information.

Collaboration Collaboration refers to different stakeholders working together to achieve their goals. The most obvious collaboration is between doctor and patient, but the other stakeholders also have the ability to collaborate if given the chance.

Health Information Health information refers to the data that is produced by the medical system. This data can be used to communicate and achieve consensus if shared effectively.

All of these trends are changing the traditional service model that health care holds. In the traditional model, patients are treated by skilled doctors who have all the answers and follow the interests of the patient. There is no mediation between the doctor and patient, and concerns of price or ability to pay are not factored into the discussion. The prototypical example of this fast-fading type of care is the small-town family doctor who knows each patient by name, and if they are unable to pay, takes in-kind payment.

One doctor, writing on his personal blog, summarizes how doctors are turning away from this model of thinking:

> *I have spoken about the idea that physicians in my generation (Generation X) are a group that trained during an explosion of medical information. We are a group that is challenging the mental model of omniscient physician – we don't want to hold all the answers for our patients because we'll fail if we do. We want to learn something new from every patient, every colleague, and every industry, every day, so we can be good educators, too. Now we can, and we are.*

As the health care model changes, there are already consumer services sprouting up to empower and connect medical consumers. These products are not being co-opted for medical consumer purposes, but are expressly designed for tomorrow's medical consumer. In the next section, I'll cover some of these new products.

Deyton, Ted. "My Definition of Health 2.0 : The Transition to Personal, Participatory Health Care"

2008. tedeytan.com

COMMERCIAL PRODUCTS REVIEW

Each of the products below take user data, combines it with other data sources, then displays back to the user using the web browser. The goal of this review of commercial products is to show the range of services in the health care context that are using user data to provide value back to users. As you'll read, each service has their own particular value-proposition that makes them worthwhile for users to spend their time and money on.

user data
website surveys of treatments taken, side effects, and effectiveness

other data
list of medications
list of ailments
Wikipedia.org descriptions

curecrowd

effectiveness bar charts
average effectiveness of different treatments for that that illness

illness descriptions
taken from Wikipedia for context

curecrowd

CureCrowd is a user-generated database of treatments for illnesses. Users use the site to find illness they are experiencing, and CureCrowd tells what kinds of treatments other users have tried for that illness. CureCrowd also records what kind of side-effects each user experienced as well as an effectiveness rating. A survey is prominently featured on each page that solicit the user to enter in their information for that particular illness. CureCrowd adds a small level of context for the illness by including a description of each illness pulled from Wikipedia.org.

CureCrowd's greatest weakness is its lack of context for different treatments. For headaches, aspirin is listed as the most effective treatment, but how many aspirin, what time of the day the aspirin was taken, and other variables are not recorded. Because of this lack of context, CureCrowds data outputs—a series of bar charts—is sparse. The user is usually left wanting more information. But CureCrowd does a good job of quickly showing available treatments and ascertaining whether or not a treatment is worth considering.

CureCrowd offers a minimal approach to crowdsourcing health data, but there is much more that can be done to gather and display medical and price data.

Google Surveys data
surveys administered by Google to internet-users on what medications they take and effectiveness levels

user data
website surveys on what medications they take and effectiveness levels

pharmacist recommendations
list of recommendations from pharmacists for different medications

medication price data
cost of the medication as well as generics

I◉DINE

overall feeling bar charts and grid
Is the medication worth it? Does it work? Is it a hassle?

side effects rate bar chart
side effects ranked by prevalence

generic or brand name grid
enter your copay amount, and iodine tells whether you should get the generic or the brand name version.

side effects timeline
timeline showing when common side effects occur. Time partitioned into first few hours, first few days, first few weeks, and first few months.

iodine

Iodine is a service with a goal similar to CureCrowd, to educate consumers about the medications they take. But unlike CureCrowd, Iodine uses more data sources and provides more data outputs for the user to experience. This additional functionality gives the user a more complete view of a particular medication, but the directness of CureCrowd's value-proposition is lost. In addition to user surveys, Iodine used Google Surveys to quickly survey 100,000 people in the United States about the medication they use, its effectiveness, and any side-effects. Iodine also includes medication-specific tips from pharmacists, FDA information, as well as medication cost information.

Because of the use of Google Surveys, Iodine has a larger base of information that CureCrowd. CureCrowd has thirteen reviews for aspirin, while Iodine has 702 reviews. Also, because Google Surveys are paid, Iodine was able to ask more detailed information about effectiveness and side effects. Iodine offers more information about side effects and efficacy that CureCrowd. This greater level of detail and coverage enables the site to answer more questions about more treatments.

Iodine also uses additional datasets effectively to give additional information about treatments. Their pharmacists recommendations provide a subjective take on each drug, and the FDA information that is included gives details that come with prescription medication that many people usually throw away. Generic and brand-name prices add another facet that describe medication. By combining all these different sets of data together, Iodine gives the user a multi-faceted view of medication.

14

user genetic data
send a swab of your dna to service

▷

23andMe

▷

heritage circular
hierarchical chart
shows what percentage of DNA comes
from different geographic populations.

health status reports
table of common illnesses as well as
estimates of chance user will have them
based on their DNA. Service suspended by
the Federal Drug Administration on Nov. 22,
2013.

23 and me

23 and Me is a genetic sequencing service that for a fee will sequence a user's genetic information, then display ethnic heritage information as well as implications for current and future health. Health prognoses include responsiveness to certain drugs, likelihood of inheriting particular diseases, and information on physical traits like a flushing reaction to alcohol. When 23 and Me first launched, their pitch was about the novelty of getting to know more about a person's own genes, but the company quickly moved onto to advertising the predictive power that genetic information has. Because it has found such an interesting dataset and a way to massively market that dataset, 23 and Me has received a huge amount attention as well as the Food and Drug Administration halting the health implications portion of its service.

23 and Me features a different service model. Its users are both customers and products for other services it offers. As its privacy policy notes:

> *We may share aggregate information with third parties, which is any information that has been stripped of your Registration Information (e.g., your name and contact information) and aggregated with information of others so that you cannot reasonably be identified as an individual ("Aggregate Information"). This Aggregate Information is different from "individual-level" information. Individual-level Genetic Information or Self-Reported Information consists of data about a single individual's genotypes, diseases or other traits/characteristics information. For example, Aggregate Information may include a statement that "30% of our female users share a particular genetic trait," without providing any data or testing results specific to any individual user.*

With this language, it makes it clear that 23 and Me has a split-level plan for its data. On the one hand, it uses visualization and information design to make its consumer-experience pleasant and entertaining, but these visualizations are a step in their bigger goal of amassing a large set of genetic information that they can then mine to answer medical questions.

user activity data
steps, gps, heart rate data

user eating data
self-reporting journaling for food eaten

user weight data

▷

sleep and activity line graph
shows sleep and activity over the day, week, c month

▷

health dashboard grid
shows key indicators for the day

health scoreboards
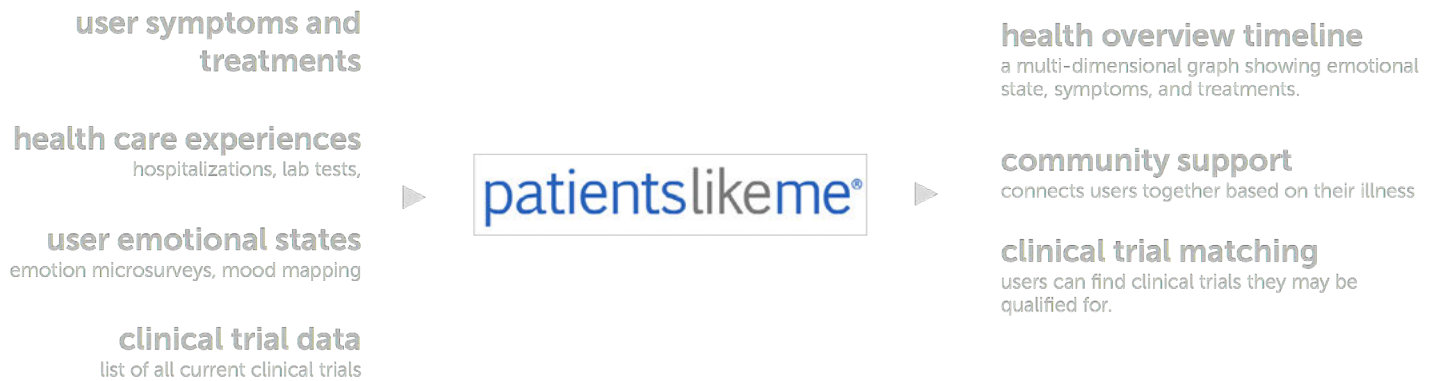compete with your friends to see who walks the most

fitbit

Fitbit is a company that sells wearable devices that record a variety of health information about a user. Health information includes steps walked, heart rate, gps location, activity level, and flights of stairs climbed. Users also have the option to add in additional data, such as their weight, daily eating patterns, and activity descriptions. Fitbit then shows that data back to its users in low-resolution indicators on the devices themselves, and through a web-based dashboard. The dashboard includes line graphs that show the history of sleep and activity, and graphical indicators to show progress throughout the day.

Fitbit's primary feature is that its wearable devices help users achieve a goal of walking 10,000 steps each day. To do this, even the least expensive Fitbit devices feature indicators to ambiently remind the user of their the progress towards their goal throughout the day. As the day progresses, lights on the device cumulatively light up until 10,000 steps are reached. At that point all the lights are reached. The 10,000 steps is a marketing-created goal that loosely aligns to recommendations for physical activity, but more importantly it provides a clear, achievable goal that makes the initial investment appear worthwhile.

Fitbit also differentiates itself from the other examples in this section by providing ways for users to share their data with other users. Users can connect with prior friends or find new friends, make groups, and share goals and stats. This sharing provides a social dynamic that encourages people to reach their goals. Fitbit's information is not very personal information, and the methods Fitbit provides for sharing limit the kinds of sharing that can take place. I believe that Fitbit will need to explore greater levels of sharing for continued success, but they will need to tread carefully to protect their users' sense of privacy.

patients like me

Patients Like Me is a comprehensive health tracking service. It was founded to enable users to share how they treat serious, long-term conditions. It originally focused on diseases like ALS (Lou Gehrig's Disease) and Parkinsons, but has since expanded to offer functionality to track the health history of any disease. Users enter all the information they receive themselves. In exchange for all this information, users receive a multi-variate graph of all their health statistics charted over time. When filled out, the chart presents a detailed view of an illness: its symptoms, treatments, medications, and how the user emotionally feels. Patients Like Me also shares the anonymized health data with research

user symptoms and treatments

health care experiences
hospitalizations, lab tests,

user emotional states
emotion microsurveys, mood mapping

clinical trial data
list of all current clinical trials

▶ patientslikeme® ▶

health overview timeline
a multi-dimensional graph showing emotional state, symptoms, and treatments.

community support
connects users together based on their illness

clinical trial matching
users can find clinical trials they may be qualified for.

partners, similar to 23 and Me. Patients Like Me goes so far as to have an Openness Policy as well as a Privacy Policy. The Openness Policy explains: "Currently, most health care data is inaccessible due to privacy regulations or proprietary tactics. As a result, research is slowed, and the development of breakthrough treatments takes decades. ... But it doesn't have to be that way. When you and thousands like you share your data, you open up the health care system."

Patients Like Me serves as a centralized experience that records all facets of an illness. It has taken advantage of the lack of control patients have over their own health data by giving them the chance to recreate their health records, piece-by-piece. Patients Like Me also allows users to record information that a medical record may discount: level comfort, happiness level, mood. It then uses all this information to give users visualizations and services that make the data-entry worthwhile. As the seriousness of the illness increases, the more likely a user will be willing to invest the time to control their medical health data and reap the benefits that Patients Like Me provides.

	Patient Investment	Data Crowdsourced	Incentivization	Grouped Data Uses
curecrowd	low	treatment efficacy	none	unclear
IODINE	low	medication efficacy and side effects	none	unclear
23andMe	medium	dna	novelty, insight into personal history	database of individual dna for research and development
fitbit	medium	personal movement	insight into personal habits, competition with friends	unclear
patientslikeme	high	emotional state and treatment details	access to research trials, community connections	database of individual medical histories for research and development
SAVE ON medical	low	medical procedure cost information	required to pay through the service	unclear
Design Opportunity Area	medium / high	medical procedure cost information	insight into medical history, owning and controlling medical data	???

competitive conclusion

This section has introduced five services that are gathering data from users and presenting it back to users using information design and data visualization. Each service though, has a different service model, a different set of target users, and different ways that users can interact with the service. With the exception of CureCrowd, these services have found an audience that is interested in exchanging data for insights. This bodes well for designers working in this field. There is much that can still be done with medical data, and users will try using it.

These services all avoid discussions of price and cost of health care. This is a large, though expected, oversight. Because these services rely on users to donate their information, and users have historically had no incentive to care about or track the cost of their health care, health care prices have not featured prominently in these services offerings. Because of the changing health care landscape in the United States, people will increasingly care about the cost of their health care. Providing services that enable users to see and reflect back on their health care prices and how it relates to their health outcomes is a promising.

Process Themes

This thesis has iterated and changed throughout its eight-month life. As data has become available or discovered to be unavailable or less prevalent than was expected, the thesis has had to morph to accommodate these changes. Originally, the expectation was that this thesis would be primarily about data visualization of available data. As time progressed, I felt a more interesting challenge was not visualizing data that is easily acquired but finding ways to acquire data that is not well-maintained or available. As I have come to the end of the thesis process, I have engaged in a gamut of skills one would expect from a professional data visualization specialist, such as transforming data, dealing with methods of representation, developing the visualization, and finding a good story to tell. I have also had to use skills one would more likely associate with a journalist or businessperson, such as cold calling hospitals, handing out fliers, and strategizing publicity.

CONVERSING WITH DATA

It is rare that the designer has the design all in her head in advance, and then merely translates it. Most of the time, she is in a kind of progressive relationship-as she goes along, she is making judgments. Sometimes, the designer's judgments have the intimacy of a conversational relationship[.]
Donald Schön
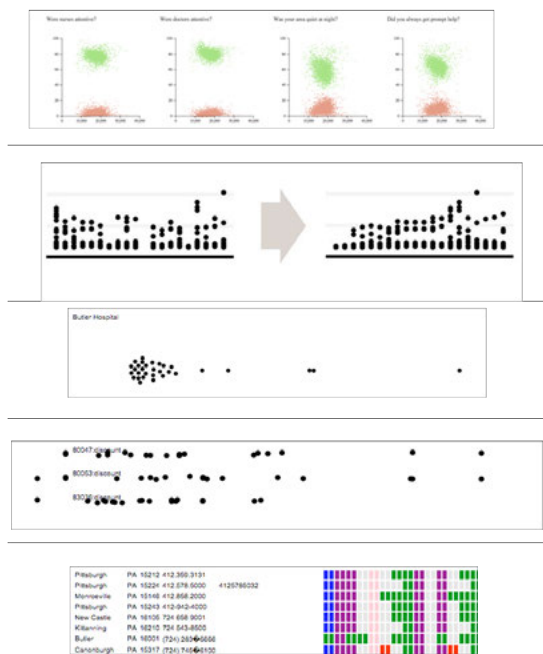
Examples of data sketches in order to prove or disprove hypotheses quickly.

A goal of this thesis project was to develop methods to engage in a conversation with the material of data. Having a conversation with a set of data is a challenge because data is abstract, decontextualized, and unsummarized. In order to overcome this, I attempted to create data representation as quickly as possible. These representations of the data gave me the material I needed to begin to a conversation with data.

This conversation consisted mostly of two questions. The first question was whether or not the data supported the story I was telling myself about the data. Constructed narratives are almost always much cleaner and clearer than the reality they are attempting to organize. By having a hypothesis about the world, I could mentally construct how I thought a representation of the data should look if my hypothesis were true. Creating a representation of the data checked my hypothesis. If the representation argued against my interpretation, I then changed my mental story.

In order for hypothesis testing to work, I needed only a basic sketch of the data. The most

successful sketches were thus created quickly and with minimum effort. These sketches had some common characteristics. They were non-interactive—any filtering or selecting of data was done by changing variables in the javascript code. They were not visually designed, but instead consisted mostly of black dots and labels added without regard to location or typography.

Looking back on some examples of this kind of data sketching, almost all the sketches were done to begin to understand distributions of values along an axis or to compare one set of values to another set.

The other question I asked was how best to represent it. These questions arose later during the process as I was crafting final visuals.

Example of the different levels of fidelity before arriving at a working prototype.

LEVELS OF FIDELITY

During the data sketching process, I arrived at three levels of fidelity before creating the final visualization: idea exploration using pen and paper, data exploration using rudimentary visualizations, and communication design using visual design programs.

idea exploration using pen and paper

During this stage, I drew ideas for visualizations using pen and paper. The sketches' primary purpose was to record brainstorming so that I could return to it at a later date. The sketches also allowed me to show my work to other designers and receive immediate feedback. Because I wasn't working directly with the data, the sketches had an off-the-cuff quality to them. A series of lines might represent a bar chart, while scribble marks would represent a user-interface element. Sketches had different levels of scope as well. Some sketches were just a series of boxes that marked out content areas, other sketches were working out the specific interface of a visualization, while other sketches were walk-throughs of what a user would experience over time. Working in this way, I could create full visualization ideas that I could later come back to.

data exploration using rudimentary visualizations

Sketching data only brings the designer so far. As noted above, data is both extremely precise, but also entirely ephemeral. This makes it hard to accurately draw what real data in graphical form will look like. In order to make the transition from idea to viewing the idea with real data, I used rudimentary graphs.

These graphs were rudimentary because they did not contain any colors, did not contain any contextual markers like axes or labels, and were non-interactive. The goal of the graphs was to see the core visualization and nothing else.

This process was very useful for quickly creating a test visualization and seeing what the visualization looked like with actual data. But it did require discipline to use. A rudimentary visualization could be the nugget of an interesting visualization but just missing a small key ingredient. On the other hand, a rudimentary visualization could be entirely off-base. Experience helped me classify the visualizations, but on reflection, I abandoned some visualizations too soon.

communication design

In order to explore what the final visualizations would look like without actually coding the visualization, I used the typical tools of a user interface or communication designer to mock up my ideas. Adobe illustrator has a series of graphing tools that can quickly output charts and graphs. Adobe Photoshop and Bohemian Coding Sketch both have tools for creating visual mockups.
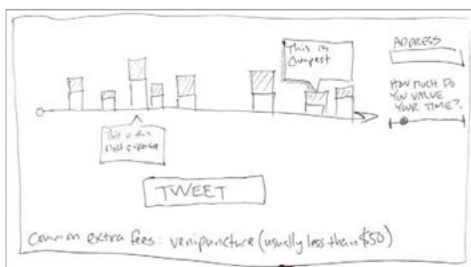
HAND-DRAWN SKETCH
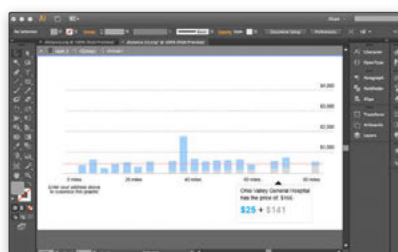
UI DESIGN

DEPLOYED PROTOTYPE

CREATING DATA FROM NOTHING

Data visualizers almost always work with data handily available. That data may need to be cleaned and organized, but the raw information is available to the visualizer. The profusion of open datasets, data services that offer databases, and competitions that provide a dataset for the visualizer to work on is evidence of this. Professional data visualizers either work in-house for an organization visualizing the organizations' data, work as a consultant brought in for specific projects, or work in media where they acquire data from fellow journalists or producers.

Two examples of returned calling sheets. These sheets were handed out to volunteers who helped call hospitals to gather price information. The sheets included phone numbers to call, procedure information to ask about, and a sample transcript to make the volunteer more comfortable. The most important part of the sheet was the bottom grid that clearly indicated what data the volunteer was supposed to gather.

These working methods severely handicap the data visualizer. It determines what they work on, and it determines the conclusions they reach. With the exception of Freedom of Information Act Requests, this way of working also guarantees that the visualizer must please the owners of the data, who often have an agenda and a story to tell. Promotional "infographics" and short animated videos abound that cherry-pick data points to tell fluffy, public relations stories. These disasters are all a combination of data visualization ineptitude combined with a strong agenda on the part of the commissioners of the graphics.

Health data is particularly hard to access. Raw health data is often harder to access. Patient health data is protected by a range of laws that hamper developers and visualizers. Price data is provided in bill form months after a procedure, and bills contain cryptic codes and multiple price amounts for what insurance pays, what the consumer pays, what the hospital charges, the discount negotiated, and other factors. The holders of health care data have little incentive for independent visualizers to experiment with the data and create visualizations.

This thesis project has attempted to circumvent these challenges by using crowdsourcing and commissioning volunteers to make calls. With so many people having cameras integrated into their smart phones, they are able to provide photos to a system without much work. We have also developed a system to break the manual work of calling hospitals into chunks of work done by volunteers. Because of this independence, we have been allowed to have an adversarial position with the health care industry that we could not have had were we working with an industry partner.

This thesis project has used two of many methods that an enterprising data visualizer can gather data themselves. There are other methods available as well. John Keefe, of WNYC's Data New Team works with projects that combine diy hacking to produce small objects. These objects could gather news and report back to the visualizer—homemade sensors. The web contains a huge amount of data in unstructured form, data, like the ratings of doctors, could be scraped and recombined. There are many possibilities to create data and not just visualize what has already been turned into a CSV file.

REFLECTING DATA BACK TO THE USER

Visual journalist Chad Skelton argues in a short talk that data visualizers often want to give their users powerful telescopes or microscopes to view data with. These kinds of visualizations either take a big and confusing topic and zoom in on a specific detail of data or give a broad overview of a dataset by aggregating and summarizing. But Skelton finds that his most popular visualizations are the ones that offer a mirror to the viewer allowing them to see themselves in the data.

In this thesis, I have been interested in ways to customize visualizations that allow the user to see themselves in the data more easily. By showing a user a reflection of themselves, the visualization becomes more shareable and interesting, but it also allows the viewer to more easily understand the data being presented because they have the baseline of themselves. Once they have oriented themselves in the data, they can begin to explore how the data changes based on changing their own characteristics. What if the user was older? Or younger? Or lived in a different city? These questions and customizations can give the user insight not only into how they relate to the data, but how other people relate to the data as well.

There are different types of user characteristics we can customize around. One type, usually simpler characteristics, can be manually inputted and changed by the user. These characteristics include age, gender, living location, wealth. As the user changes categories or values the visualization changes. Another type of characteristic are the digital materials users produce. By connecting a visualization to a user's Twitter or Facebook or Gmail account, the visualization can learn something about the user and customize accordingly. The user's digital data can also be visualized as a complement to the primary data set.

2014

AUG   SEPT   OCT   NOV   DEC   2015 JAN   FEB   MARCH   APRIL   MAY

Exploration of Available Data

Design,   Develop,   and publicize PghMedicalPrices.us

Interview Consumers   Ideate design solutions   Design Medical Bill Explorer

Crowdsource   Design Map   Develop and
Pittsburgh Prices   Interface   Publicize Maps

Process Timeline

INITIAL DATA EXPLORATIONS

The original expectation for the thesis was to visualize the medical data in United States Medicare's Hospital Compare Database. Using exploratory visualization methods, we found that the data Hospital Compare contained to be too abstract to be useful.
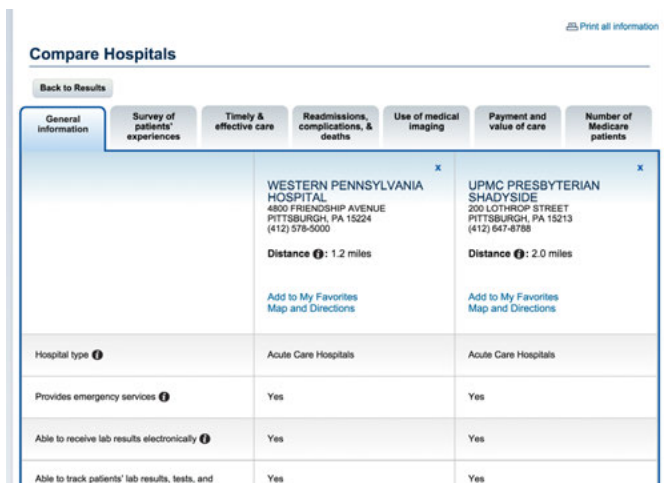
The hospital compare user interface.
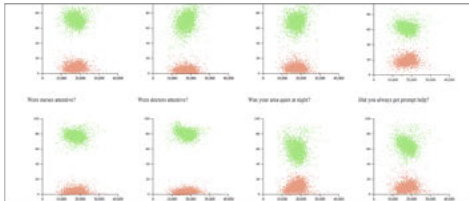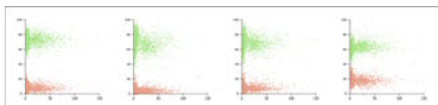
The Hospital Compare dataset is produced by US Medicare from data from Medicare-certified hospitals. There are over 4,000 of these certified hospitals in the United States. The data set includes a few different interesting categories. Hospital Compare runs a patient survey that measures some quality measure from the patient's point of view. In addition to quality indicators, the database included Medicare's average per patient payment made to each hospital.

All this data for so many hospitals made it appear to be a good first step towards visualizing the hospital landscape in the United States, but as we performed exploratory analysis of the data, we found the data to be too sparse to give a full portrait of a hospitals' quality. Also the Hospital Compare data did not include the many non-hospital providers such as private doctors' offices, minute clinics, and surgery centers.

Below is a summary of the most interesting visualizations we created based on the Hospital Compare data.

Data Explorations Using Hospital Compare Data

	Question Asked	Conclusion	Example Image
Mapping of Data	What is the geographical range of the Hospital Compare data?	Hospital Compare covers hospitals across the United States for a small range of procedures.	
Price/Quality Correlation	Does hospital quality correlate with average Medicare reimbursement?	There is not a strong correlation between reimbursement and quality.	
Distance/Quality Correlation	Does hospital quality correlate with distance to the nearest metropolitan area?	There is not a strong correlation between distance from metro area and quality.	
Local Hospital Comparison	Do local hospitals differ significantly in their Hospital Compare data?	Hospitals do not appear to the user to differ significantly in terms of quality or price.	

I did a few more visual tests with the data, and no matter how I changed the variables, Medicare's data didn't give insightful views into the data. A new data source was needed.

PGHMEDICALPRICES.US

I expanded the search for other data sources that could provide information around consumer medical price information.

In particular, I wanted to know how much procedures cost, where they are performed, and if the procedures had positive or negative medical outcomes. These questions can be objectively answered and would not rely upon medical consumer reviews or surveys. For the data to be useful, it needed to be at the individual medical provider level.

Doing a survey of the available data sources, there were no sources that provided this information. The U.S Department of Health and Human Services publishes the Medical Expenditure Panel Survey. "MEPS is the most complete source of data on the cost and use of health care and health insurance coverage," according to the databases website.

Unfortunately this dataset is available for research only and researchers are not allowed to publish hospital-level information. Being in Pennsylvania, another resource for medical data is the Pennsylvania Health Care Cost Containment Council. The PH4C publishes detailed medical billing information on a provider-level basis. PH4C's medical data is available for a fee, but the range of procedures covered left out the most common out-patient procedures. PH4C's data is also delayed, so up-to-date prices are never available.

Because no dataset was available that met my needs, I brainstormed ways to gather the desired data. I had focused my earlier explorations on what medical consumers want, so I now looked at the data available to medical consumers about their own care. Medical consumers are given many documents during their medical care, and most consumers dispose of them without paying much attention, except when it might be a bill. Medical bills could provide me with the data necessary to build a dataset of procedure prices and procedure volume, and perhaps if the consumer answered a short survey, I could also get a measure of outcomes.

Every medical procedure done in the United States has an associated Current Procedural Terminology (CPT) code associated with it. This code allows doctors, hospitals, medical billers, and insurers to speak a common language in order to accurately translate what medical practitioners actually do into bills for insurers and medical consumers. A common example, emergency room visits are coded on a sliding scale based on severity of symptoms and time spent. Emergency room Level 1, the least intensive visit, is coded 99201. Emergency Room Level 5, the most intense, is coded 99205.

74020	77057	45378
Abdominal x-rays, complete	Mammogram, screening	Colonoscopy (non-Medicare)
73610	80047	80061
Ankle x-rays, 3 or more views	Basic Metabolic Panel	Lipid panel
72141	93005	93306
Cervical spine MRI	EKG tracing only	Echocardiogram

Many medical bills from medical providers and explanation of benefits from insurers include the CPT code along with the price the hospital charged. If we could have medical consumers give us this information, we could create a live database of medical prices for throughout the United States. I looked at examples of crowdsourcing as possible solutions for gathering this data on a wide scale.

The first question I had to answer was what chunk of work I would expect my crowd workers to do. Ideally, the crowd would provide accurate CPT codes, descriptions and reviews. Unfortunately, medical bills are not designed for clarity and ease of reading. Insurers have some incentive for clear communication, but hospitals have little incentive

Screenshot of Google Form used to test medical consumers' abilities to enter in billing information.

to help their consumers understand the charges on their bill, especially when the consumer's insurer is covering the majority of the bill.

In order to test how well average medical consumers would be at understanding and translating their medical bill, I created a simple prototype powered by Google Forms. I gave the prototype to a half-dozen people who had medical bills, and I asked them to attempt to translate their bill into data. The results of the experiment showed that average people were not experienced enough to enter their own data. Some medical bills were missing CPT codes entirely, others had hospital-specific codes that were not of use unless a skilled medical biller interpreted them. Participants also reported that it was a lot of work for them to enter the information. Based on the feedback, I decided that asking our crowdsourced workers to interpret and enter their own information would be too high of a burden.

At this point, I began to look in earnest at other ways to gather the medical bill data without having to require people to fill our a complex form. I decided that a photo-based service would be the best idea. Smart phones that have cameras are ubiquitous, and mobile browsers today have the abilities to interface with those cameras.

As I began to build the system using traditional design and development practices, I also held six hour-long interviews with older, non-college-aged Pittsburgh residents. In the interviews, I asked them a range of questions about:

Prior medical experiences and subsequent bills.

- What they did with the bills they received.
- How they tracked their medical expenses.
- And I presented participants with wireframes of different potential visualizations.

From the conversations, I came away with some important lessons about my users:

- Users wanted to learn more about how much their medical care costs.
- Users receive and often keep medical bills, but they do not use the data. It is a backup
- Users feel deep anger towards the medical system at the moment they get a bill and afterwards.

Medical bills do not help assuage this anger. Bills are vague about what procedures are being billed, and confusing about what should be paid by insurance or by individuals.

These findings encouraged me to continue designing and producing PGHMedicalPrices. us. In later interviews, I was able to show interviewees working prototypes of the system. Responses to system were positive, and every person who saw the system said they would like to use it. In retrospect, I should have realized that their enthusiasm for the system during the interview would wear off if asked to use the system without a prior one-hour interview.

PITTSBURGH PRICE EXPLORER

After PGHMedicalPrices.us had been live for a month, I was worried that the site was not getting as much traffic as I had hoped based on the interviews. PGHMedicalPrices. us needed more time to gain popularity, but the academic schedule required that I keep moving forward.

I brainstormed other ways to gather medical price data from hospitals. I had read articles about people who had called hospitals asking for prices. What if I took that technique but focused on a single metropolitan area's hospitals and gathered data from a variety of procedures instead of just one? This was a worthwhile challenge. I decided to aim for 22 hospitals in Pittsburgh and aim to collect price data on forty common, outpatient procedures.

The next challenge was how to make all the phone calls. I did an initial round, and I found that hospitals were not prepared to answer questions regarding cost. Because they were not prepared, calls for this first round took on average twenty minutes per hospital once I got a person on the phone. The first round of calling ended up taking more than eight hours of phone calling time including receiving calls days later with the information I was asking for. I began a second round of calls, but many hospitals have a single person who answers these kinds of questions, and some became suspicious of multiple calls.

In order to get around this problem, I created a series of worksheets for volunteers to make calls. Each worksheet contained two or three hospitals and included two or three procedures. I tracked which volunteer got which worksheet to make sure only one person had a particular set of hospitals. At the bottom of the page was a grid that volunteers would fill out as they made the calls. They could then text me a photo of the sheet with the gathered information.

Five examples of returned forms. Some were returned using email, some were returned with a photo. Some people used the grid that was provided. Others made their own. Many people added additional comments and information.

Volunteers expressed surprise and bewilderment at the poor response from most hospitals. For them it was an educational experience that got many people indignant at the health care system and hospital in particular. They were surprised that hospitals offer discounts to the uninsured, and they were surprised that hospitals were not prepared to answer questions about medical prices. One participant reflected:

> *I had a very enlightening morning shopping around for MRI's and X-Rays. The main thing I learned was the discount you receive when uninsured, I wish I had known when I was uninsured for 6 months before I found full time employment after graduating college. I remember having to pay a bill in full without even knowing about breaks! I also learned you can access price information if you know the codes, I honestly did not believe a hospital would be open to sharing the price amounts. Next time I need something medical I am going to definitely shop around ... unless it's an emergency.*

This method was successful at getting data. As you can see from the grid below, I had hoped to gather data on the top forty procedures. I was able to get data on half of the procedures.


With the data gathered, I created the Pittsburgh Price Explorer website. The creation took place in four phases. I sketched ideas for what visualizations the site would include, I did preliminary UI designs to give myself a guide, I developed the site, and then I did some light user testing and optimization. The site is overall successful, but given another chance, I would include user testing earlier than I did.

Final Prototypes

PGHMEDICALPRICES.US

www.pghmedicalprices.us

PGHMedical prices is an web-based application that enables users to take photos of their bills and send the photos to the project team. The system allows submissions through text-message, mobile browser, or desktop browser. Users are prompted to find a medical bill, hide any personal information, and take a photo of the bill. They then have the option to enter a phone number or email address if they would like to be contacted about the project in the future.
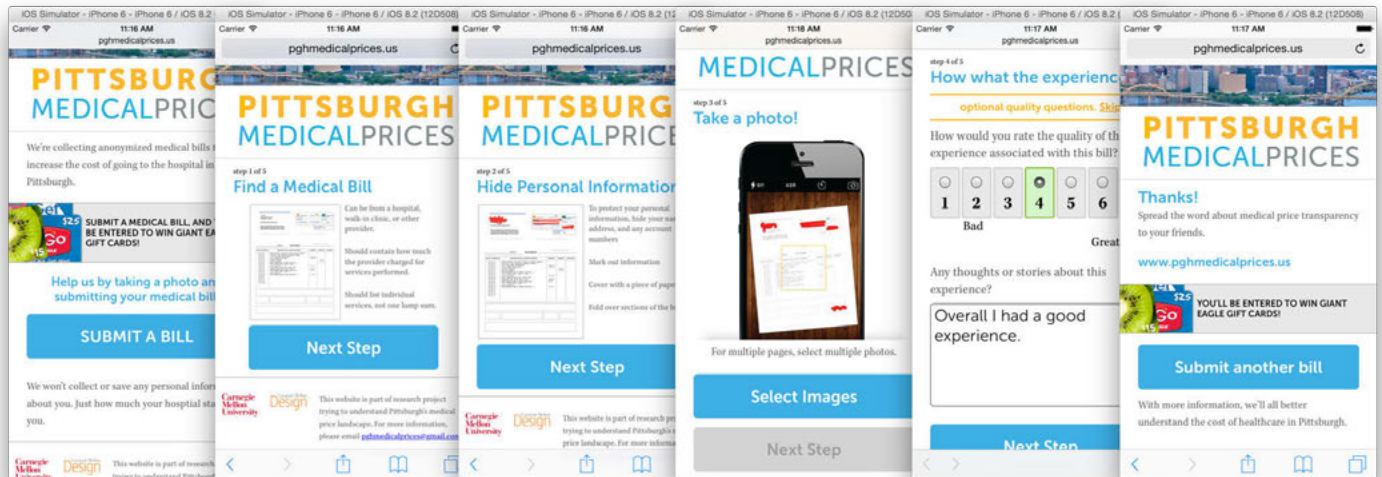
Screen shots of PGHMedicalPrices.us

The system was designed using a traditional design process. Sketches were created, then higher level UI designs. These UI designs were informally user-tested before the site was developed. After the site was developed, additional user tests were performed asking people to complete the process of uploading a bill. Changes to the UI were made based on these user tests. Most of the users tested were graduate students and older university staff members. They were able to complete the central task easily, though the user testing did not explore motivation or incentives for using this in real life.

The system has been rather unsuccessful at gathering medical bills. In total the system has gathered approximately fifty medical bills from a dozen users. Future work needs to be done to make the system more appealing to use with a greater reward for the user to upload bills. An entire service could be created around uploading bills, parsing the data, then reflecting the findings back to the user to tell them how much or little they are paying for their health care.

DISCOUNT VS FULL PRICE   HOSPITAL COMPARISON   OVERALL RANKING   DISTANCE AND PRICE
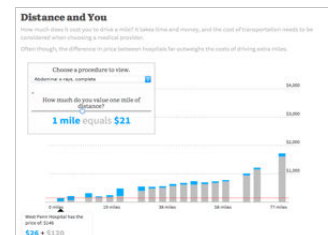
PITTSBURGH PRICE EXPLORER

http://pghmedicalprices.us/price-explorer/

Pittsburgh Price Explorer is a web-based visualization that shows four different visualizations to the user helping them to understand how the prices in Pittsburgh for medical care differs from hospital to hospital. The data is composed of price data from 20 hospitals from throughout the Pittsburgh region. Each hospital has full and discount price from twenty common procedures. The visualization begins with a one-question quiz to hook the viewer into caring about the data, then displays the data in different formats.

The visualization page was an opportunity to use some of the lessons from the literature review:

- The data was a custom-gathered dataset that volunteers worked to create. Without the work of these volunteers, the visualization would have been impossible to make. This was a surprisingly effective use of gathering a small, local crowd of people to help do research.

- The visualization begins with a one-question quiz in order to hook people into the experience. By asking for the user's opinion, the visualization becomes a conversation between user and system, and not just a one-way transfer of information.

- The visualization asks people to enter their address. With this information, the visualizations modify themselves to use the entered location as an axis point for distance measurements. This customization more easily allows people to see what hospitals are near to them, and gives them a better context around the data.

SERVICE DESIGN CONCEPT

In collaboration with Trisha Black, a CMU Master in Biotechnology, Innovation and Computation (MSBIC) student, we created a service design concept that would combine price and quality data about medical providers into an interface that would give doctors and patients concrete information to plan referrals and tests. Often, doctors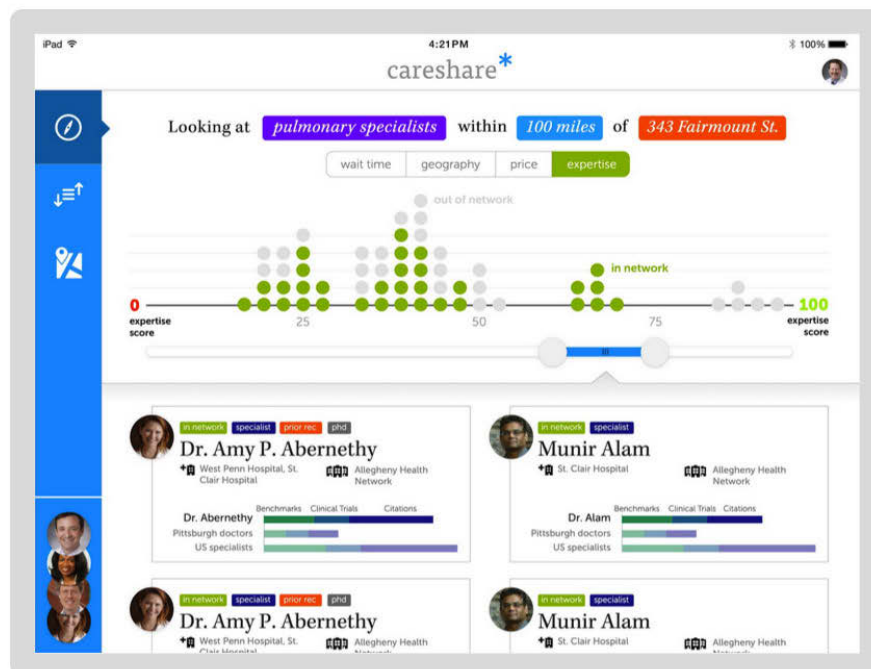 refer based on a variety of factors, such as personal familiarity with other doctors. Having objective data that doctor and patient can agree upon facilitates conversation by providing a common set of facts and data to build the

explore providers

list all providers

map of providers

saved doctors

overview label

chart showing providers based on expertise

detailed information on providers

conversation upon. The service would track medical costs and also communicate back to the doctor what procedures or tests had been performed, even if the secondary provider was not on the same network.

We envision the service beginning at the doctor's office. The doctor will use a tablet computer to draw a patient into a discussion of the subsequent medical procedures the doctor is suggesting. Using the application, the doctor or a nurse can find providers of the needed procedures. The medical provider saves the recommendations on the application. The recommendations are then saved for the user to access in their homes. When the user visits the secondary provider, the original doctor or nurse can see who the patient visited and get information about the visit.

By presenting price and quality data, the service inserts concerns about price and ability to pay into the medical conversation in an ambient way. The quality information about the medical providers also gives a ground truth that the doctor can either agree with or react to. This insertion of data thus solves two big problems with medical referrals, suggestions for treatments that are too expensive, and referrals based on friendship or other non-medical reasons.

Most of this thesis has been concerned with creating systems that are achievable in the short term. The system envisioned here is currently not possible because of misalignment of incentives in the health care industry. Medical providers who work for hospital systems are often required to refer their patients to within their networks, even if that may be a more expensive or low-quality option. Insurance companies who hold much of the price data that could power such a system would greatly anger hospitals by releasing this type of health data. Some insurance companies though are starting to release estimators similar to what CareShare is suggesting.

Conclusion
Changing People's Behavior

The health care industry is changing quickly, but people only intermittently access health care, and when they do they use old mental models that exclude considerations of price and quality. The most important conclusion from my research into health care is that designers need to create systems to jolt people into seeing the health care system in a new light. Take the mother of two children and show her how she has the power to make choices about her care based on more than gut feelings. Show middle-aged people that they can access health care and not have to worry about price because they know what they'll have to pay beforehand.

Crowdsourced systems and data visualization offer a way to jolt people by involving them not just in the viewing of gathered data, but by including them in the gathering process. By asking people to participate and gather data, they will look at their data in a new and more critical way. Once they've added the data, the system can show how their data fits in with the larger system. Like one of the crowdsourced participants said, "I had a very enlightening morning shopping around for MRI's and X-Rays [prices] ... Next time I need something medical I am going to definitely shop around[.]"

The more specific lessons of this thesis—data sketching, reflecting data in visualizations, generating hard-to-get data, and levels of fidelity in data-visualization—are all in pursuit of this goal of changing people's behavior.

I hope anyone who reads this documentation and views the projects is inspired to create data visualizations projects that tackle tough questions that don't have easily-accessed data sources available. Data visualization can do much more than help expert users understand complex data or provide public-relations visuals for selling ideas or products to the public.

Thanks

Thanks so much to my two advisors, Jodi Forlizzi and Aisling Kelliher. Thanks to all the volunteers who helped me gather data, and thanks to Carolyn for putting up with me.

Bibliography

Baker, Robert. "Watertracker: Structured Data, Integrated Response - Ushahidi." N.p., 25 June 2012. Web. 01 Nov. 2014. <http://www.ushahidi.com/2012/06/25/watertracker/>.

Belt, Tom H Van De, Lucien Jlpg Engelen, Sivera Aa Berben, and Lisette Schoonhoven. "Definition of Health 2.0 and Medicine 2.0: A Systematic Review." Journal of Medical Internet Research 12.2 (2010): n. pag. Web.

Bennet, John. "Bringing Design to Software Ch 9 - Schon." Bringing Design to Software. Addison-Wesley, n.d. Web. 22 Apr. 2015.

Bernstein, Michael S., Greg Little, Bjorn Hartmann, Robert C. Miller, Mark S. Ackerman, David R. Karger, David Crowell, and Katrina Panovich. "Soylent: A Word Processor with a Crowd Inside." UIST '10 (2010): 313-22. Print.

Blumenthal, David, and Sara R. Collins. "Health Care Coverage under the Affordable Care Act - A Progress Report — NEJM." New England Journal of Medicine. New England Journal of Medicine, 17 July 2014. Web. 22 Apr. 2015. <http://www.nejm.org/doi/full/10.1056/NEJMhpr1405667>.

Bos, Lodewijk, Andy Marsh, Denis Carroll, Sanjeev Gupta, and Mike Reese. "Patient 2.0 Empowerment." Proceedings of the 2008 International Conference on Semantic Web & Web Services (2008): 164-67. Web. 16 Jan. 2015.

Boulos, Maged, Steve Wheeler, Carlos Tavares, and Ray Jones. "How Smartphones Are Changing the Face of Mobile and Participatory Healthcare: An Overview, with Example from ECAALYX." BioMedical Engineering OnLine 10.1 (2011): 24. Web.

Bowen, Simon, Andy Dearden, Peter Wright, Daniel Wolstenholme, and Mark Cobb. "Participatory Healthcare Service Design and Innovation." Proceeding PDC '10 Proceedings of the 11th Biennial Participatory Design Conference (2010): 155-58. Web.

Doan, Anhai, Raghu Ramakrishnan, and Alon Y. Halevy. "Crowdsourcing Systems on the World-Wide Web." Communications of the ACM 54.4 (2011): 86. Web.

Eysenbach, Gunther. "Medicine 2.0: Social Networking, Collaboration, Participation, Apomediation, and Openness." Journal of Medical Internet Research 10.3 (2008): E22. Web.

Fox, Susannah. "The Social Life of Health Information, 2011." Pew Research Centers Internet American Life Project. Pew Research Center, 11 May 2011. Web. 12 Apr. 2015. <http://pewinternet.org/Reports/2011/Social Life of Health Info.aspx>.

Gelman, Andrew, and Antony Unwin. "Infovis and Statistical Graphics: Different Goals, Different Looks." Journal of Computational and Graphical Statistics 22.1 (2013): 2-28. Web.

Gershon, Nahum, Joshua LeVasseur, Joel Winstead, Jame Croall, Ari Pernick, and William Ruh. "Case Study Visualizing Internet Resources." Proceedings of the Proceedings on Information Visualization (1995): 122-28. Print.

Ginsburg, P. B. "Shopping For Price In Medical Care." Health Affairs 26.2 (2007): W208-216. Web.

Goodchild, Michael F., and J. Alan Glennon. "Crowdsourcing Geographic Information for Disaster Response: A Research Frontier." International Journal of Digital Earth 3.3 (2010): 231-41. Web.

Goodchild, Michael F. "Citizens as Sensors: The World of Volunteered Geography." GeoJournal 69.4 (2007): 211-21. Web.

Hibbard, Judith H., and Ellen Peters. "SICHCD: Data Presentation Approaches That Facilitate the Use of Information in Choice." Annual Review of Public Health 24.1 (2003): 413-33. Web.

Howe, Jeff. Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business. New York: Crown Business, 2008. Print.

Howe, Jeff. "The Rise of Crowdsourcing ." Wired 14.06: The Rise of Crowdsourcing. Wired Magazine, 01 June 2006. Web. 13 Nov. 2014.

Kittur, Aniket, Ed H. Chi, and Bongwon Suh. "Crowdsourcing User Studies with Mechanical Turk." Proceeding CHI '08 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2008): 453-56. Print.

Lease, Matthew, Jessica Hullman, Jeffrey P. Bingham, Bernstein Michael, Juho Kim, Walter S. Lasecki, Saeideh Bakhshi, Tanushree Mitra, and Robert C. Miller. "Mechanical Turk Is Not Anonymous." Social Science Research Network (2013): n. pag. 16 Mar. 2013. Web. 28 Apr. 2015. <http://ssrn.com/abstract=2228728>.

Mol, Annemarie. The Logic of Care: Health and the Problem of Patient Choice. London: Routledge, 2008. Print.

Murray, Peter. The Saga of the Sydney Opera House: The Dramatic Story of the Design and Construction of the Icon of Modern Australia. London: Spon, 2004. Print.

Norheim-Hagtun, Ida, and Patrick Meier. "Crowdsourcing for Crisis Mapping in Haiti." Innovations: Technology, Governance, Globalization 5.4 (2010): 81-89. Web.

Peters, Ellen, Louise Meilleur, and Mary Kate Tompkins. Numeracy and the Affordable Care Act: Opportunities and Challenges. Working paper. Columbus OH: Psychology Department, The Ohio State U, 2013. Print.

Peters, Ellen, Louise Meilleur, and Mary Kate Tompkins. "Numeracy and the affordable care act: Opportunities and challenges." Chapter prepared for the Roundtable on Health Literacy, Institute of Medicine (2013).

Rodgers, Simon. "Twitter Mapping: Foundations." Source. Open News, 10 Oct. 2014. Web. 13 Nov. 2014.

Shneiderman, B. "Inventing Discovery Tools: Combining Information Visualization with Data Mining." Information Visualization 1.1 (2002): 5-12. Web.

Stolterman, Erik. "The Nature of Design Practice and Implications for Interaction Design Research." International Journal of Design 2.1 (n.d.): 55-65. Print.

Swan, Melanie. "Crowdsourced Health Research Studies: An Important Emerging Complement to Clinical Trials in the Public Health Research Ecosystem." Journal of Medical Internet Research 14.2 (2012): E46. Web.

Tu, H. T., and J. H. May. "Self-Pay Markets In Health Care: Consumer Nirvana Or Caveat Emptor?" Health Affairs 26.2 (2007): W217-226. Web.

Tudor-Locke, Catrine, and David R. Bassett. "How Many Steps/Day Are Enough?" Sports Medicine 34.1 (2004): 1-8. Web.

Wood, Jason, Ken Brodlie, and Helen Wright. "Visualization Over The World Wide Web And Its Application To Environmental Data." Proceedings of the IEEE Visualization Conference 7 (1996): n. pag. Print.

Zook, Matthew, Mark Graham, Taylor Shelton, and Sean Gorman. "Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake." World Medical & Health Policy2.2 (2010): 7. Web.Introduction