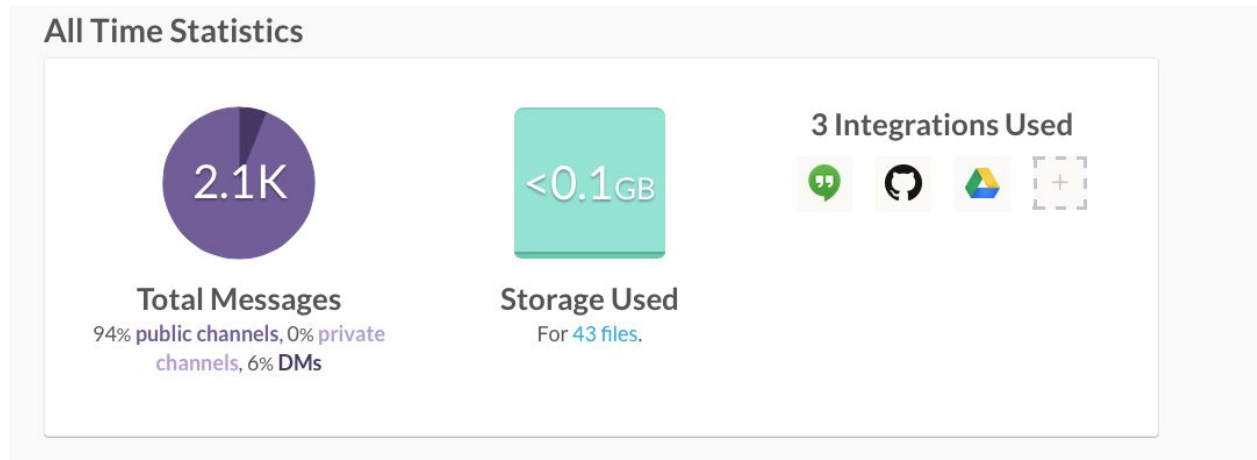**Project Process Book**

Team: Melodic Transport
Nathaniel Burbank
Jeremy Clark
Bruno Carriere
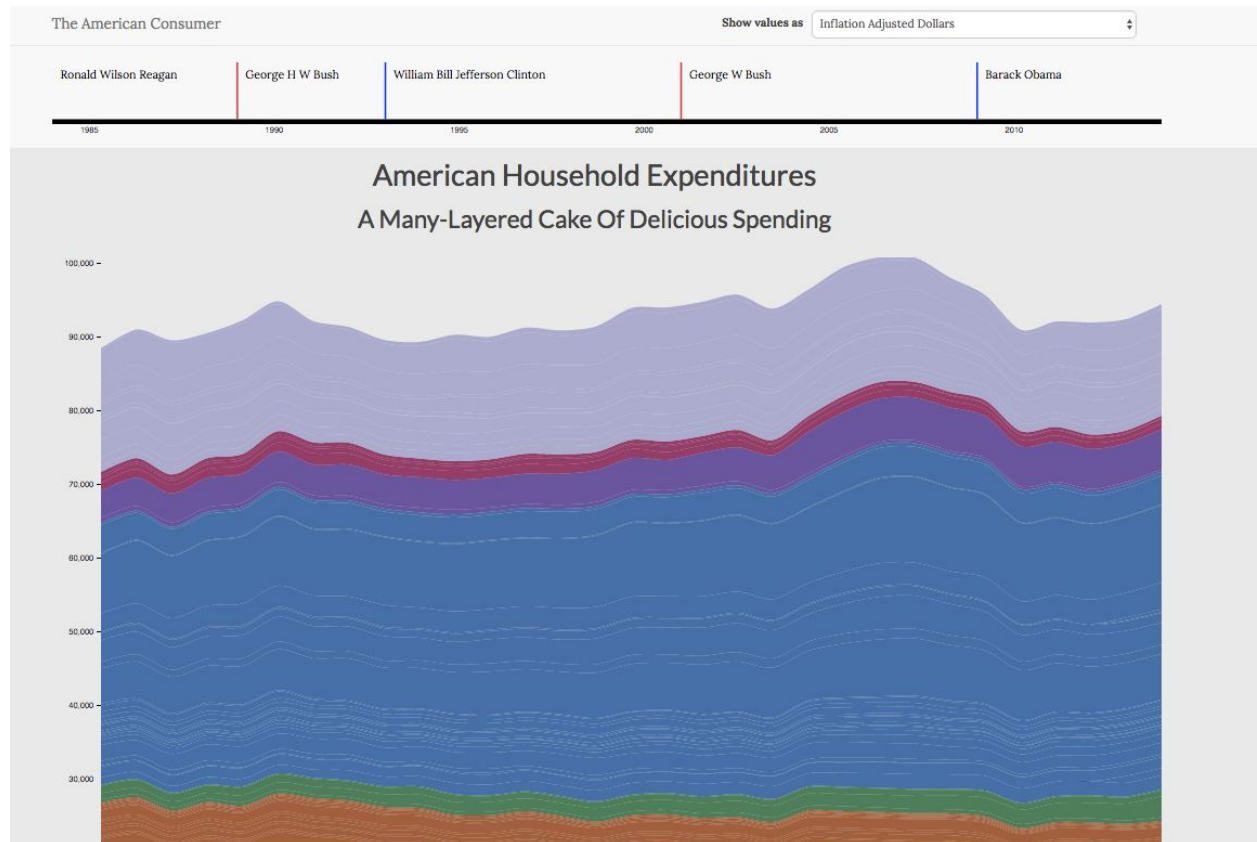
### May 2:  Final Project Submission  ###



We're finally done! In our final entry to our process book, I want to take a moment talk not about the project itself but the tools we used. We were able to work together using Slack, google hangouts and github to really come together as a team. Our project is great!

### April 25:  Project Prototype V2  ###

While we may not be quite as far along on the project as we had initially hoped we would be by this stage, things are finally coming together.(See the screenshots below.)  We have working and complete views of the timeline, area chart, radar chart, and tree visualizations. Moreover all four are connected via a central scrubber that allows you to select a range of years (for the area chart) and an end and start year for the other charts. As you move the slider, all of the charts animate allowing the viewer to observe changes over time. As we look toward our final project submission that's due a week from today, we're hoping to add more story elements that will direct viewers to the interesting changes in spending patterns at different periods in time. One of the things we're struggling with though is the manner in which country-wide averages masks some of the interesting changes in spending patterns and region by region differences.
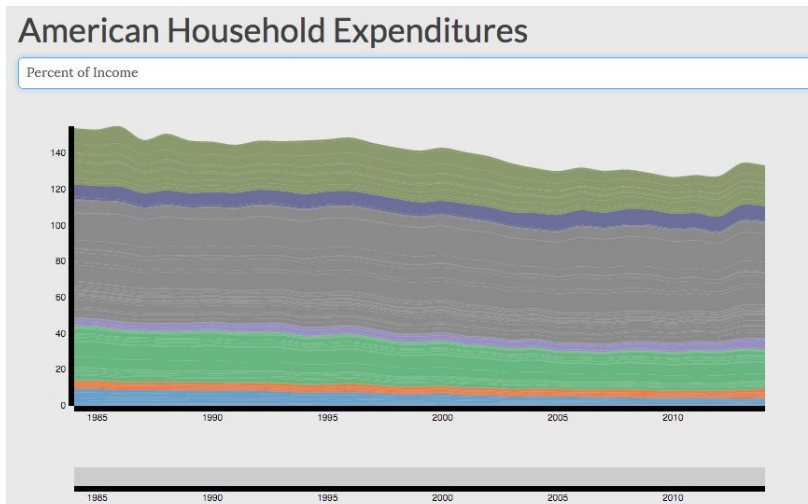
The American Consumer | Show values as | Inflation Adjusted Dollars

| Ronald Wilson Reagan | George H W Bush | William Bill Jefferson Clinton | George W Bush | Barack Obama |

1985          1990          1995          2000          2005          2010

## American Household Expenditures
### A Many-Layered Cake Of Delicious Spending

100,000 —
90,000 —
80,000 —
70,000 —
60,000 —
50,000 —
40,000 —
30,000 —

mographic Category

uintiles of income before taxes

mographic Characteristic

I Consumer Units

### April 18, 2016:  Project Prototype V1  ###

This week, we finally got down to brass tacks and actually started coding our project. We had our dataset, we had a lot of ideas, and then we just had to split up the work and take a crack at it. Although we're not quite as far along as we had initially hoped we would be at this stage, we have working versions of two out of our three main visualizations, and a template for our overall

web page design. The main focus for the following week will be finishing up the different views on our areachart and start-chart visualization, adding more details/events to our timeline controler, and also implementing our "innovative" tree visualization.

Things are coming along nicely…



The American Consumer
25 Years Of Spending Trends

Every year the Bureau of Labor Statistics produces The Consumer Expenditure Survey, a study that provides information on the buying habits of American consumers, including data on their expenditures, income, and consumer unit (families and single consumers) characteristics. The survey data are collected for the Bureau of Labor Statistics by the U.S. Census Bureau.

**American Household Expenditures**

Percent of Income

### April 11, 2016:  Receiving Expert Feedback  ###

Group 1 reviewed the second iteration of our project plan and gave us very well thought out feedback. On the one hand, the fact that they understood and communicated back to us most of the main themes we were hoping to communicate gave us confidence that our visualizations will be effective communication tools once actually implemented. However, they suggested that we should spend more time thinking about the interesting stories we can pull out of our data, which is a good insight. Additionally, they had some questions about our "tree" visualization, which is probably the most novel visualization we're planning on attempting. This feedback caused us to reflect on what we were really trying to communicate with that particular visualization. After hearing their feedback, as a team we spent a while debating exactly how we wanted to encode the different properties in our dataset within it, while still enabling the visualization to communicate some broad trends about the increased burden of certain expense categories.

*CS-171 Visualization Expert Evaluation*

*Reviewer: The Free Radicals (Online Studio 1, Group 1) Under review: Melodic Transport (Online Studio 1, Group 6)*

*General thoughts:*

*You project includes a lot of interesting data visualization and interactions which will definitely give the user a lot of options to explore the data and make their own conclusions. We think your project would be even stronger if you were able to improve the storytelling side. For example,*

could you pull out some interesting stories or results from your own exploration of the data? This would show the reader what is possible with the tool you have developed (tacit tutorial) and also stimulate them to think about other connections and trends that might be interesting for them to explore.

How are you dealing with inflation? Is the BLS dataset already inflation adjusted? If so it would be useful to show the index year for dollar prices so this is clear to the reader.
Severity rating: 0

The following comments correspond to the numbered items highlighted below:
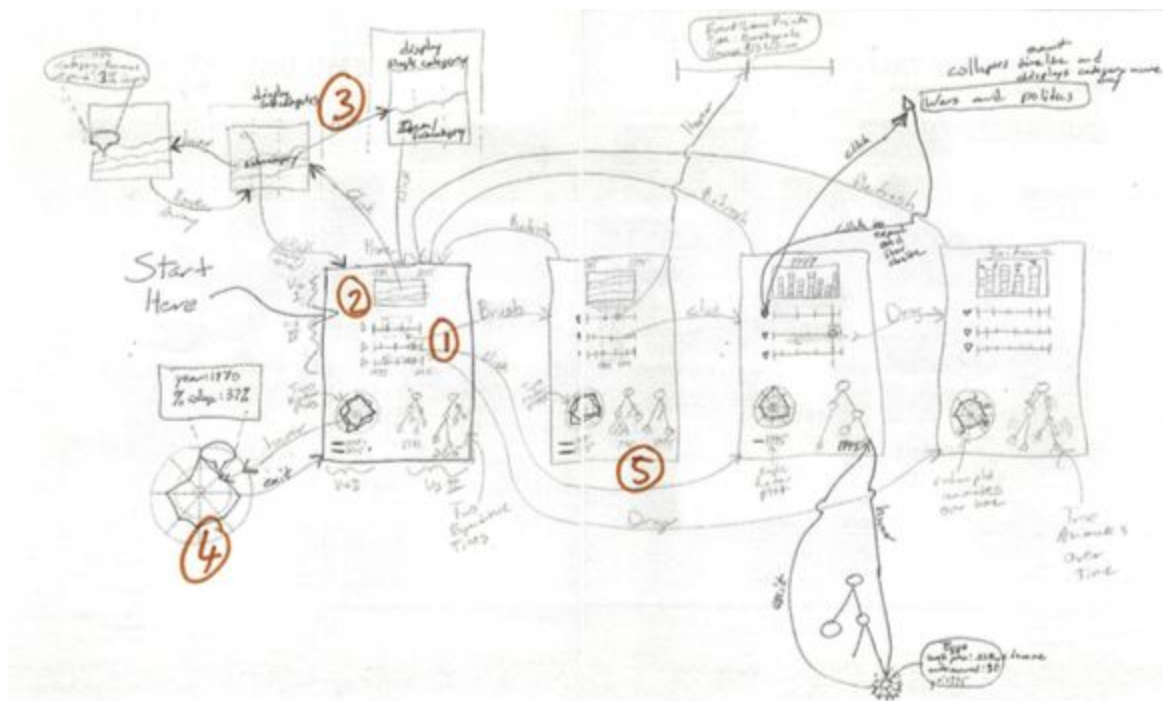


1. You have a lot of interactive elements which provide some excellent functionality. We think it would be useful to think about how you convey to the reader which elements are interactive, so they don't miss any of your cool features. Perhaps some icons for the charts and slider bars for the timelines. You may have already planned these and we just didn't pick up on them from the sketches.Severity rating: 1

2.  *How are you defining spending, by household or by individual?*
    - *If by household, how will you deal with households where the mother and father have different race or education level (thinking of the radar plot here)?*

    - *If by individual, how will disaggregate their spending from the family as a whole? Severity rating: 0*
3.  *We like how you can drill down into each category to move from the stacked layout to individual area chart. This is a great way to show the collective trend while also allowing the reader to estimate the trends more accurately with an isolated view of the data. Severity rating: 0*

4.  *The radar plot should be quite an effective way of showing how the breakdown of respondents has changed over time. Have you thought about including gender as one of your categories? If the breakdown of respondents is significantly different over time, have you considered how this might skew the results? Severity rating: 1*

5.  *The force directed tree diagram is the only part of the project that we had some trouble with. From our reading of the storyboard it seems like this is the only visualization where the reader can compare before and after costs side-by-side, however in our view this visual encoding has some limitations for quantitative comparisons. Firstly, the layout of the force directed tree will be different every time it is called, making a side-by-side comparison difficult, but you might be able to get around this by making the tree sticky . The bigger challenge will be for the reader to accurately compare the area of nodes as a way of calculating unit price of expenditure as % of income, especially when the eye must pass over several other node circles before making the comparison. Severity rating: 2*

### April 4, 2016:  Project Plan V2  ###

With feedback from our helpful TA's we set out to revise our project plan, taking account of the suggestion that we should "experiment with several ways of providing the historical events with the line graph/area chart." We created a revised and more detailed feature list (pasted below); Bruno built an awesome "data explorer" to help us understand the dataset; and we continued to refine our project plan:

**Project Plan V2**
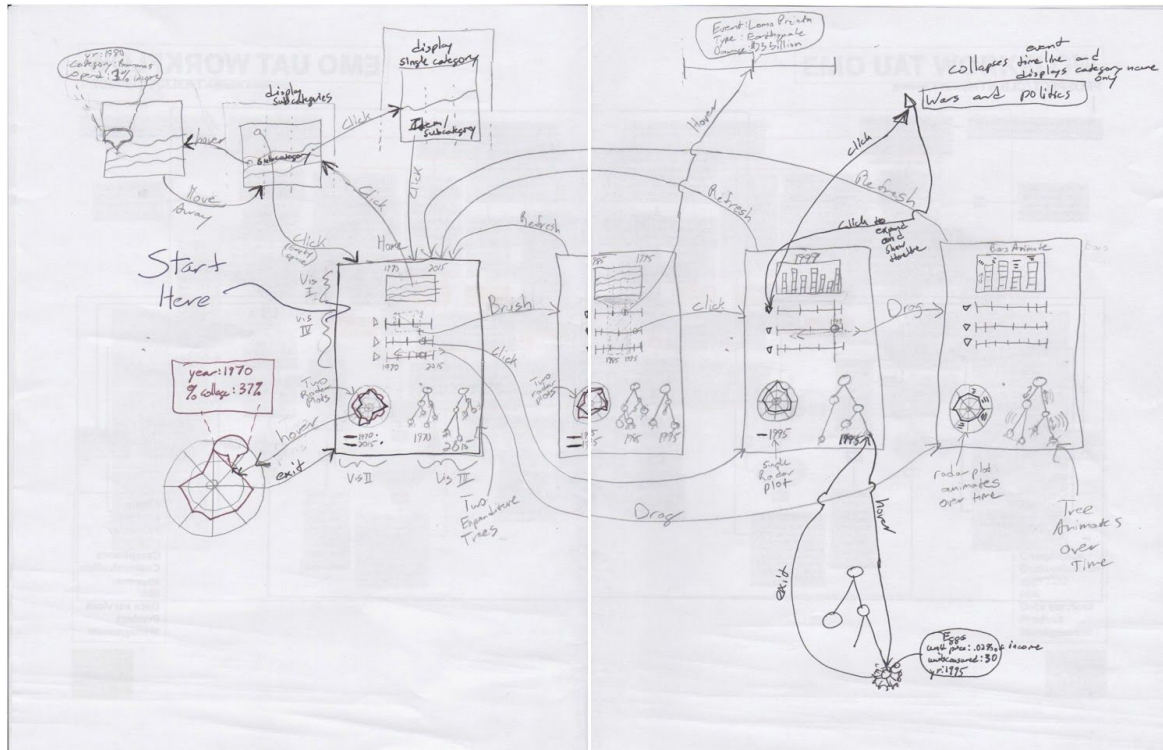
**Project Feature List**

**Stacked Area Chart:** The focal point of our project will be a stacked area chart that shows average spending over time across several different general categories such as food, transportation, etc. This chart will be scrubbable/searchable in several ways.

**Timelines/Control Panel:** A set of linked timelines below the chart will offer the main way that the user can adjust the area chart. These will allow the user to "brush" and select different periods of time to zoom in on, at which point the area chart will re-scale to show the selected time period. However, these timelines will not just act as controls for the area-chart display. Rather, they will also serve as a display mechanism for key historical events across different categories. One timeline might display economic recessions and bull-market periods, while another could display the political party in control of the white house, or key foreign events.  As the user brushes on any of the timeline-controls, all will update simultaneously in a linked fashion. This will enable the user to explore the different correlations and connections between consumer spending and historical events.

**Percentage/Aggregate dollar view toggle**: a toggle selector or button will enable the user to switch back and forth between viewing the data in raw dollar terms, and percentage-per-year terms, which emphasize changes in proportion overtime.

**Demographic Filters**: the user will also be able to filter the dataset by race, income, educational level, region, and perhaps other criteria, to visualize how spending changes vary across different socio-economic categories. We're envisioning a series of toggles or checkboxes off the side that would enable these filters to be turned on or off.

**Stacked Area Detail View:** Selecting and double clicking on a layer within the chart, the user will be able to zoom in on just that category, and then see a more detailed breakdown of sub categories. For example, if the user zoomed in on the transportation category, they might see a detailed breakdown of changes in American consumer spending habits on public transportation, automobile purchases, and gasoline over the given period of time.

**Stacked Bar Chart Single Year View:** If the user clicks on a single year (or a single historical event) the main view will switch from an area chart overtime to a barchart of categories of spending at a given year, enabling more detail to be displayed for that time slice.

**Historical event mouseover tooltip:** Mousing over a specific historical event will provide more details about that event.

**Secondary star chart visualization:** As a compliment to the area-chart visualization, we will also display a star chart visualization that demographic information about the expenditure survey respondents, with each point representing a different demographic. In the default view, two star chart area plots will be displayed (semi-translucently) on top of each other, representing the data from the beginning and end of the selected time period. As the user "drags" along the timeline sliders, the star chart may animate.

**Secondary Tree Graph visualization:** As a compliment to the area-chart visualization, we will also display a force-directed tree graph visualization that displays spending across the given categories. A root node represents income, the first branched nodes will represent categories, and the third-level nodes will represent expenditure items. There will be a repulsive force between nodes parent-child nodes based on what % of overall spending that node represents. The size of the expenditure nodes will represent the unit price of that expenditure as a % of income. There will then be a cluster of nodes around the expenditure that represents the total consumption of that particular item. In the end, here's the story the graph should tell:
- The more spread out, the more money was spent
- The bigger the clusters around the 'expenditure' nodes, the more consumption of that item

As the user "drags" along the timeline sliders, the graph should 'animate' to reflect over time consumption.

Additional Sketches:

### March 28, 2016:  Project Plan V1  ###

This is the first time when we really tried to put it all together with a coherent story, sketches of our proposed visualizations, and defined questions we were trying to answer. The centerpiece of this part was our screencast.

**Project Proposal V1**

Consumer spending accounts for 70% of the American economy. Yet we rarely step back and consider the ways in which our spending habits have changed, and why those changes have happened. Our goal with this project is to explore how American consumer spending trends have changed overtime, and provide insights into the underlying social trends, demographic forces, and historical events that have motivated these shifts.

Using time-series data from the annual Consumer Expenditure Survey published by the Bureau of Labor Statistics, we aim to create a series of visualizations that will first inform the user about the ways in which modern spending habits differ from those of prior generations. Some categories (such as housing and healthcare) have grown significantly in recent decades, while others (such as food) have become a much smaller part of American consumers' budgets, and we want to display these shifts. Our hope is that we'll have detailed enough data so that the user will also be able to zoom in on each of these categories to display more detailed information about changes in subcategories. For example, if the user zoomed in on the transportation category, they might see a detailed breakdown of changes in American consumer spending habits on public transportation, automobile purchases, and gasoline over the given period of time.

Once we've built a visualization that displays *how* spending habits have changed, we will also work to provide insights into *why* these shifts have occurred. Our second goal is to overlay the changes in spending trends with other demographic and historical changes that we believe are correlated with the changes in spending habits. For example, family size has decreased significantly over the past century, while simultaneously, the share of Americans who live in cities has increased, and access to higher education has been greatly expanded.  We believe each of these shifts might be apparent in the spending data, and we hope that by layering the different trends on top of one another the salient trends will become apparent.  Additionally, our visualizations can be enhanced by comparing overtime consumer spending trends with timelines of interesting events in recent American history (military actions, stock market values, major weather events/natural disasters, presidential party, major legislation, etc.).

As a third goal, we'd also like users to be able to filter the data by race, income, educational level, region, and perhaps other criteria, to visualize how spending changes vary across different socio-economic categories.

## Tasks:
- Become proficient with the Consumer Expenditure Survey dataset. While it's clear that there's a lot of data available at a very high level of detail, we're still working to understand the way it's encoded so we can extract what we need from it to form the basis of our visualizations.
- Do some basic exploratory data analysis in Excel or R to get a handle on the most dramatic changes in spending habits over the given time period that we're hoping to (1) display and (2) explain.
- Make a list of 5 other changes in demographics that we believe would be most correlated/explanatory with regard to the changes in spending habits.
- Make a list of the key historical events/dates that we believe are connected to the given changes in spending habits.
- Make sketches of both our main and secondary visualizations
- Prepare simplified dataset as csv file (or files) that contain only the data points we need for our visuzlations
- Build an initial version of the visualizations, and collect feedback.
- Revise, revise, revise.
- Rinse and repeat…

## Dataset

The bulk of the data for our project will be gathered from the Consumer Expenditure Survey published by the Bureau of Labor Statistics. There are other data sets available from the BLS that we are also considering such as Unemployment and Employment, Consumer Price Index and various price averages.

The CEX contains a large number of metrics. For each survey participant we have access to information such as:

- Age of reference person
- Composition of consumer units (families, children, etc)
- Education level
- Income level
- Region of residence

The data related to expenditures covers a wide range of purchases, including:

- Detailed food purchased for home

- Housing and shelter costs
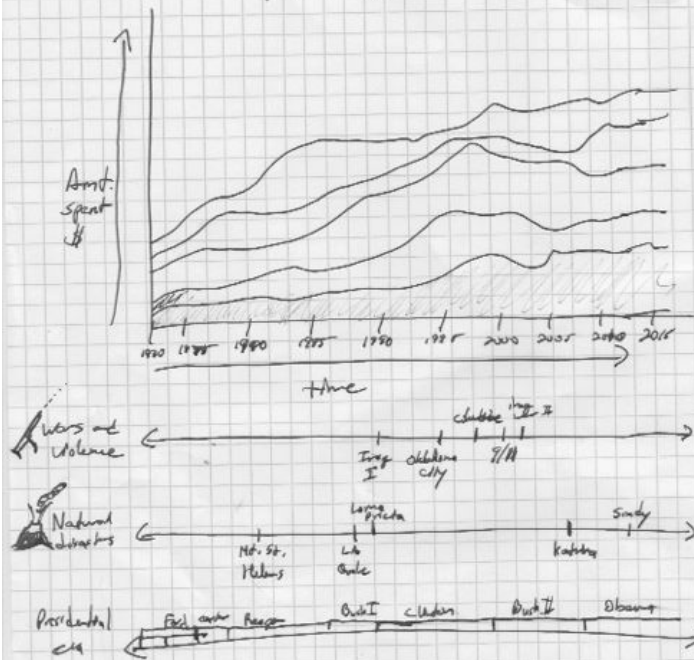- Apparel and services
- Health care spending
- Entertainment

In addition, we will supplement the data set with timeline information for various interesting and meaningful events in recent American history: Military actions, stock market values, natural disasters, presidential and party changes, major legislation, etc.

Supplemental Data

- [The Presidents Of the USA](#)
- [The International Disaster Database](#)
- [Timeline of United States History](#)
- [Timeline of United States Military operations](#)
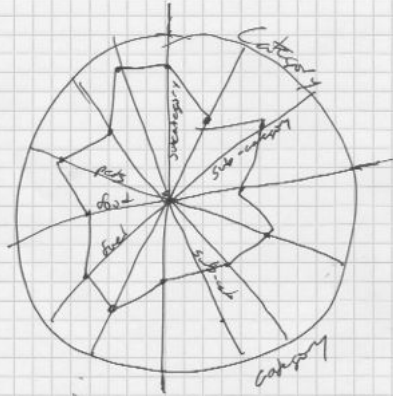- [Chronology of Significant Events (CA Dep. Of Finance)](#)

# Possible Visualizations

# Viz I



Amt.
spent
$

1920  1925  1930  1935  1950  1995  2000  2005  2010  2015

time

Wars and
violence

Iraq I   Oklahoma   9/11
City

Chablee II II

Natural
disasters

Mt. St.
Helens

Loma
Prieta

LA
Quake

Katrina

Sandy

Presidential
era

Ford  Carter  Reagan   Bush I  Clinton   Bush II   Obama

Brushing along historical
X-axis will zoom main stack chart, and add
indicators to the stack chart.

# Vis II



Category

1970 1975 1980 1985 1990 1995 2000 2005

As slider moves, sdar
plot animates. Sort of
a "flip book" effect

---

# Vis III

Item [ Fuel ▽ ]    ▨▨▨ = Consumption
                        over
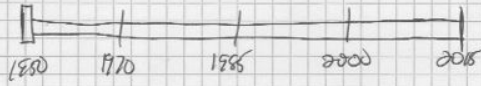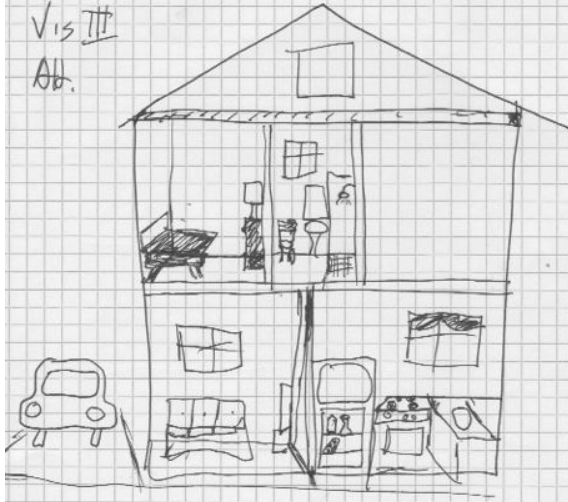                        time
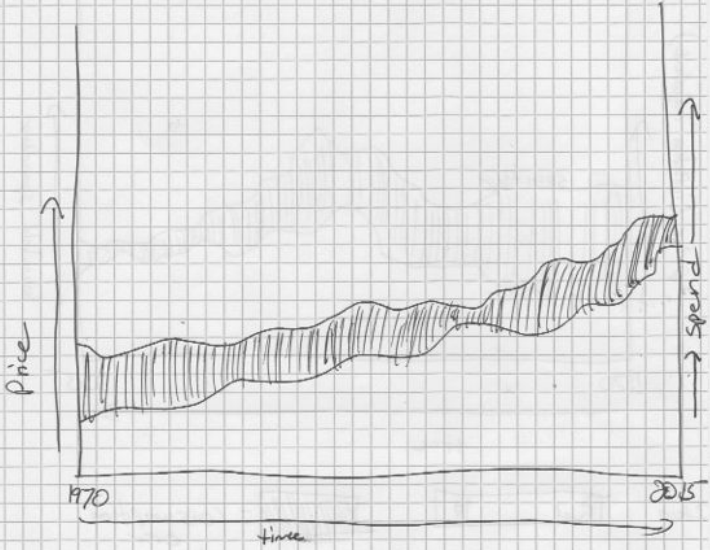
Price

Spend

1970                                          2015

Time

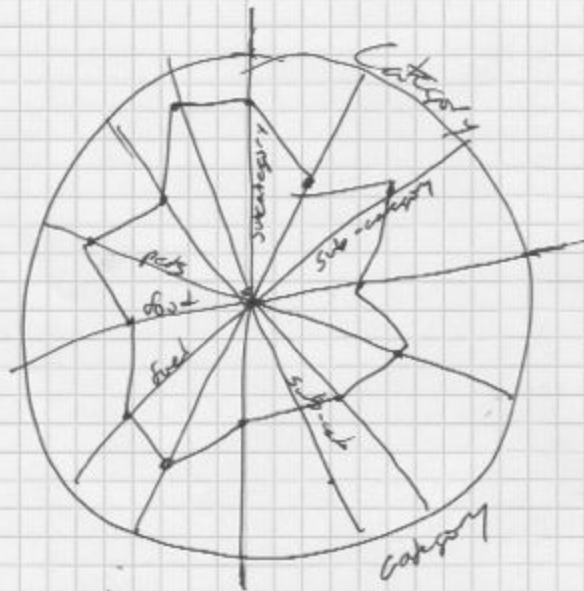As the slider moves year over year, the size of items in the house changes to reflect relative speed.

Slide →  relative speed.

|————|————|————|————|
1850   1970   1985   2000   2018

Vis III
Alt.



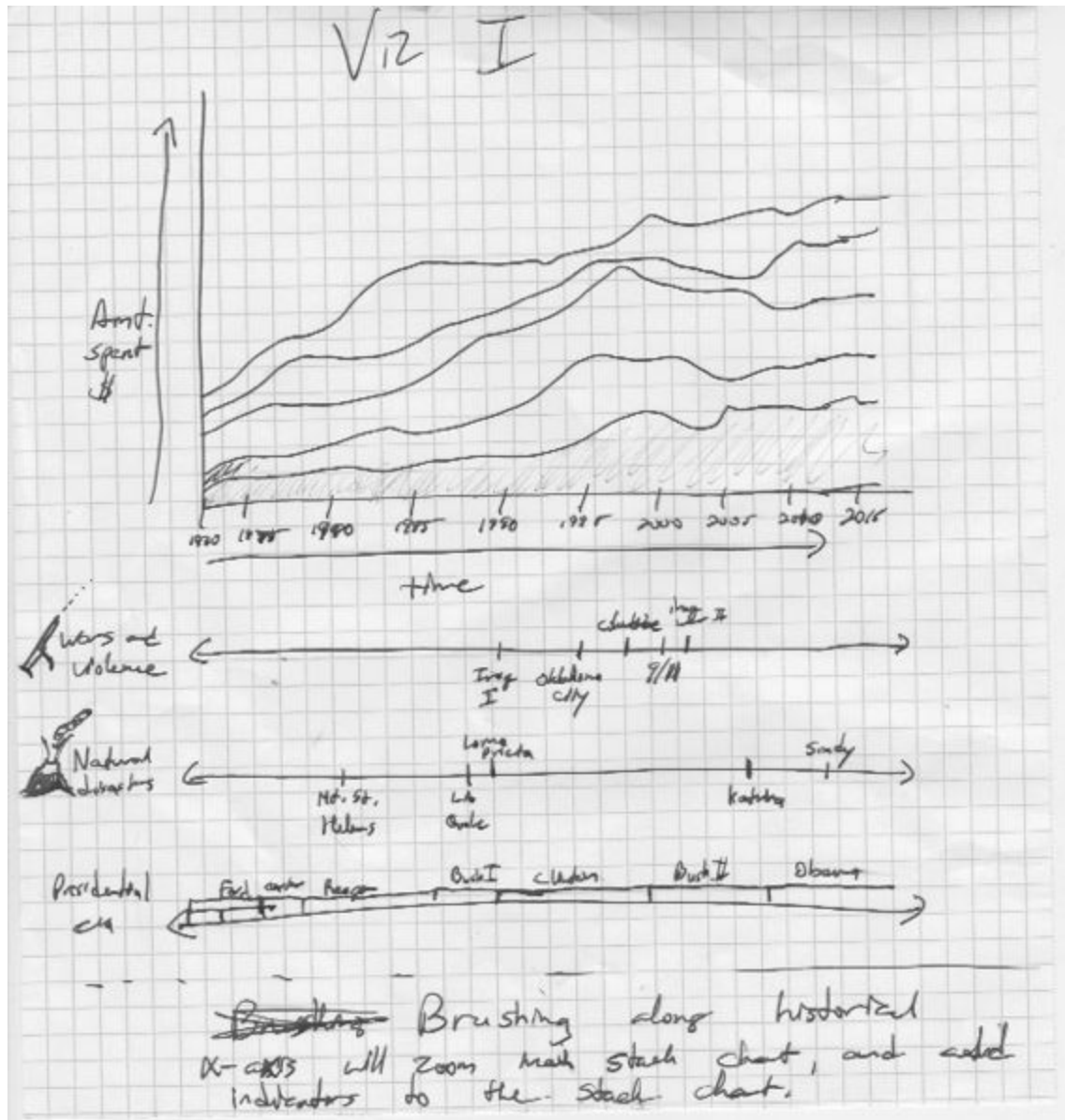As the slider moves year over year, the size of items in the house changes to reflect relative speed.

Slide →  relative speed.

|————|————|————|————|
1850   1970   1985   2000   2018

Vis III
Alt.

Vis II



1970   1975   1980   1985   1990   1995   2000   2005
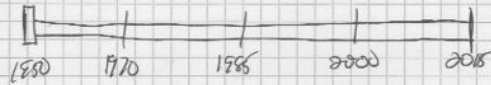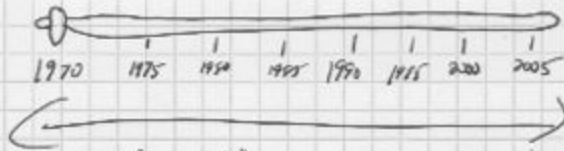
As slider moves, star
plot animates. Sort of
a "flip book" effect

### March 27, 2016:  Expanded Questions  ###

After writing our initial proposal, we expanded on it a bit and brainstormed about the different types of questions we might want to answer:

### Children / Family spending

* How much do children cost / how much do people spend on family?
*  How does it vary across regions and other dimensions?
* How has that changed over time and some guesses as to why?
* What are differences in spending between households with children and those without?
* Could also frame this as, as the number of Children per Household goes down, the amount we're spending per child goes up

### Contributions

* How do contributions to educational, political, religious organizations, and charities vary over dimensions, such as:
        * Time?
        * Region?
        * Household income?
        * Education Level?


### Educational expenses

Show the evolution of the cost of education across time.

* How have education costs changed over time?
* What percentage of household income is spent on education over various dimensions, such as?
        * Time?
        * Region?
        * Household income?
        * Education Level?


### "Market Basket", CPI

This is an interesting measure of the way prices of goods and services changes over time, such that it tracks
changes in cost of living.
* What is the correlation between changes in wages and changes in cost of living?


### Correlations to external events

* What is the relationship between US consumer spending over time and:
 * prevailing political party control?
 * prevailing stock market conditions?

* gas prices? (Specifically is the a discernible connection between gas prices and transportation costs)
 * other commodity prices?

* Are there discernable patterns in US consumer spending as a reaction to significant events, such as:
        * 9/11 and other military events?
        * Natural disasters (Kitrina, Tsunamis (south asia/japan), etc.)?
        * Stock market events (i.e. single-day/week crashes)?
        * Elections?

### Longitudinal spending changes over the last century

* The rise and fall of big box stores. For example, it would be cool if we had an animated map where you could see the spread of stores such as Walmart, Target, and Best buy throughout the 1990s, only to see them start retreating around 2008 or so. This combined with some data about aggregate spending in big-box stores vs. online and/or small retailers could be interesting visualization of this trend.

* The change in American spending patterns over the last century: Although the BLS survey has been collected annually in recent years, the original survey data goes all the way back to 1880 (collected every 10 years). The changes in spending patterns are most dramatic when you look over a long period of time. For example, in 1901, the average american spent 42% of their income on food. By 2003, that share had fallen to 13%. Spending on transportation (all forms) grew from basically nothing in 1901 to ~20% in 2003. This report (http://www.bls.gov/opub/uscs/home.htm) is the best summary of the types of stories we could tell with this data, and it's all based on the BLS survey data. There are probably more interesting long term stories that could be told about aggregate Americans -- i.e., the declining cost of food, the increasing cost of health care, the increase of other "non-essential expenditures" etc.


### March 22, 2016:  Initial Proposal  ###

After we found our dataset, our proposal came together relatively quickly:

Title: **The American consumer -- what do we spend our money on?**

Our proposal is to explore how American consumer spending trends have changed overtime. Using time-series data from the annual Consumer Expenditure Survey published by the Bureau of Labor Statistics, we aim to create a series of visualizations

that show how certain categories have grown (such as housing and healthcare) while others have become a smaller part of American consumer budgets. Our goal is that our visualizations will show high-level general trends while also allowing users to zoom-in on specific categories that are interesting. We'd also like users to be able to filter the data by race, income, educational level, region, and perhaps other criteria, to visualize how spending changes vary across different socio-economic categories. Additionally, visualizations can be enhanced by comparing overtime consumer spending trends with timelines of interesting events in recent American history (military actions, stock market values, major weather events/natural disasters, presidential party, major legislation, etc).

Miscellaneous Sketches: