Process document: An American Day

Ben McMorran and Francisco Sanchez

Overview and motivation

Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.

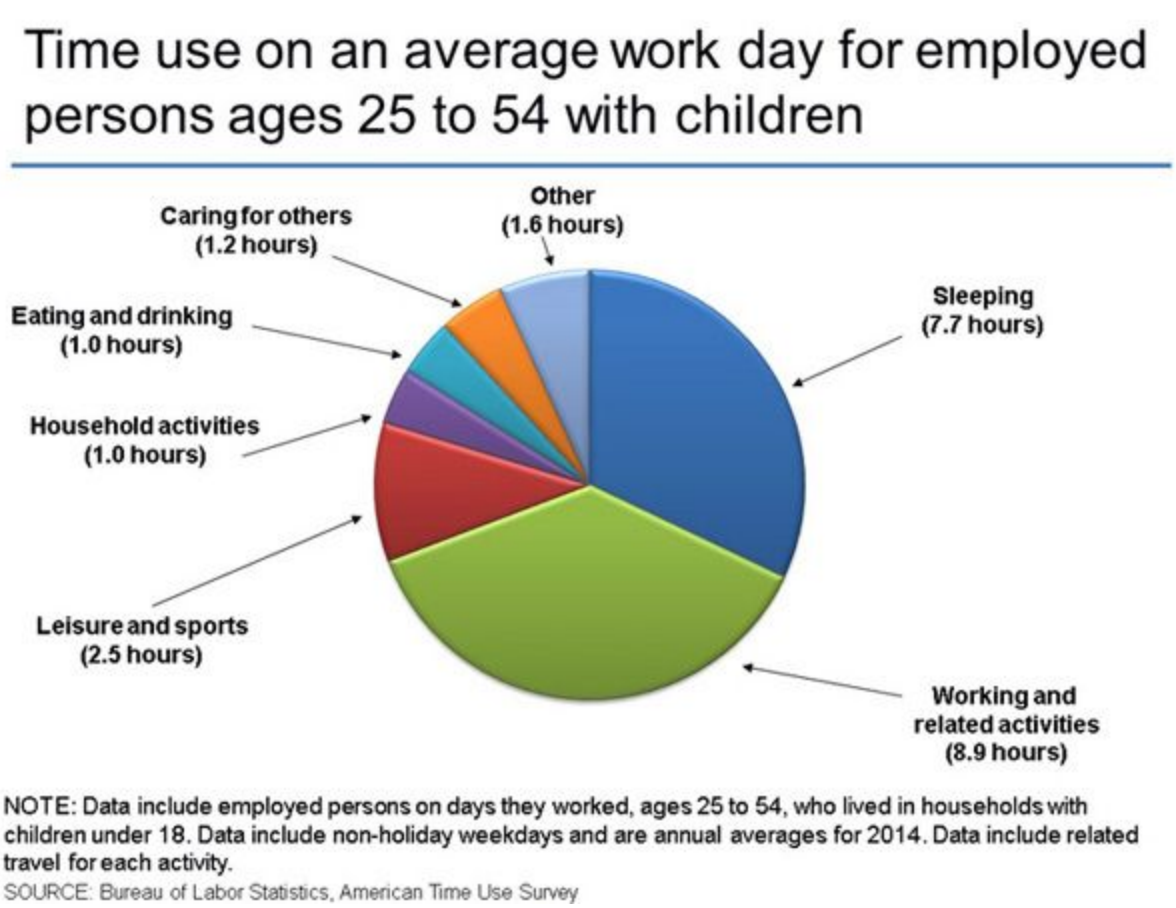
How does the average American spend their day? Using data collected by the Bureau of Labor Statistics (BLS) in the American Time Use Survey (ATUS), our project set out to explore the the daily habits that we all share. ATUS is a continuously run survey that asks respondents aged 15 and older to recall every activity they did in the last 24-hour period. Over 3,000 samples are collected each month, and this information is used in a wide variety of applications such as demographic research, quantifying the amount of non-paid time that people work, and measuring time spent on elder and child care.

Compared to these economic applications, our project is directed towards a much more casual, exploratory audience. We seek to create an interactive, fun, exploratory tool for users quickly gain a high level understanding of the ATUS data, see how they compare to the average, and look for insights in time use differences among demographics. While the ATUS data obviously works for comparing daily routines, the interview date combined with age of the respondent allows to to explore other timescales as well, such as a week, year, or even a lifetime. Our visualization should make it easy to jump between these timescales during exploration.

Related work

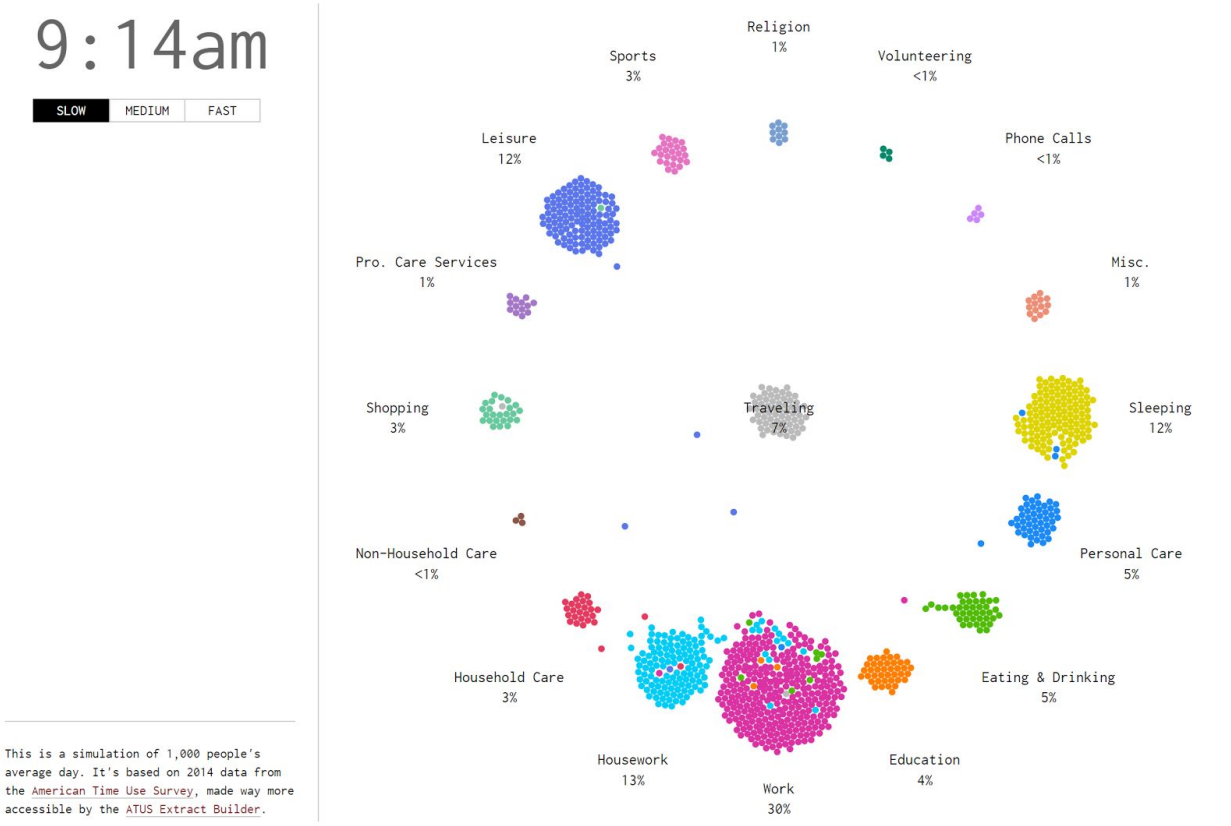
Anything that inspired you, such as a paper, a web site, visualizations we discussed in class, etc.

We are far from the first to visualize the ATUS data, and we drew inspiration from several sources. First, the BLS itself provides some simple (albeit non-interactive) visualizations of the data in the PDFs they publish on their website.



While these simple visualizations work to some degree for conveying very high level statistics, they are poor from a visualization perspective. Pie charts use an angular encoding, which is known to be perceptually inferior to position encoding. BLS also provides bar charts for some statistics, but here they suffer from being non-interactive and trapped inside a PDF. However, the BLS information can still provide useful guidance for finding insights in the data, and our tool should be able to visualize these same insights.

Another visualization comes from FlowingData. This visualization focuses much more on individuals that simple averages like the BLS. Each dot in the visualization represents a person moving about the activities of their day. Transitions are simulated by a Markov process trained on the original ATUS data.



The individualized aspect of this visualization adds a lot of visual interest, but it's difficult to gain deep insight. Time is mapped to actual time through animation, so it is difficult to compare relative percentages of individuals in different activities at different times of day. One thing this visualization does well is show the sequence of activities that an individual participates in throughout their day. It's easy for this to wash out in averages, but the sequence may provide important insights.

FlowingData has several other visualizations of the ATUS data focusing on different aspects like the most common activities broken down by age and sex and different times of day, when Americans leave for work by county, and histograms of how many hours Americans spend on different activities. Each of these visualizations do an excellent job of telling a focused, compelling story about the data while allowing the user to interact with the data to see how they compare.

Questions

What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?

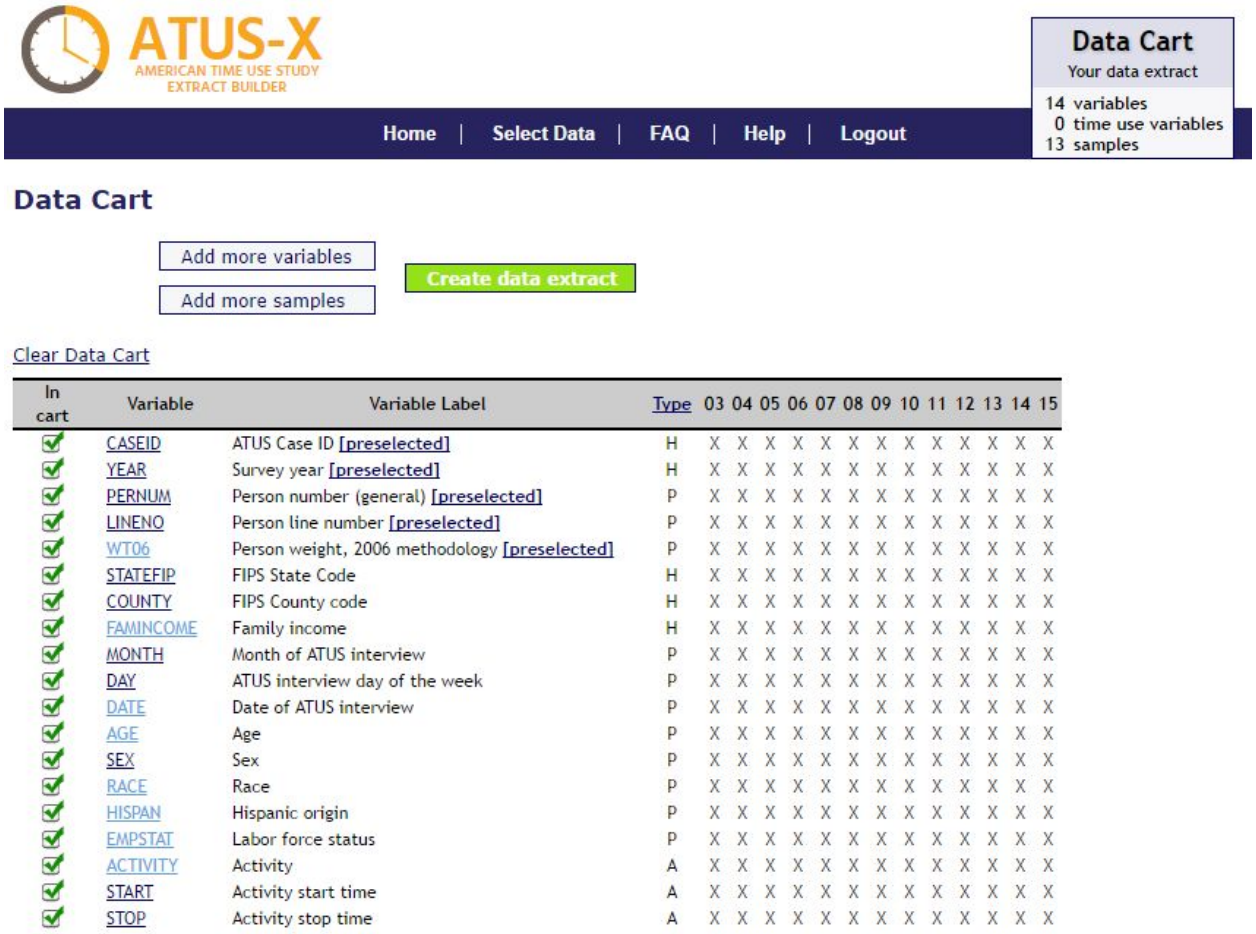
Our visualization is designed primarily as a tool to explore the answers to three main questions:

- Which activities do Americans do most commonly over different time periods (day, week, year, and lifetime)?
- How do common activities vary by location within the US?
- Are there significant differences in activity prevalence between different demographics (split on age, sex, race, employment, and income)?

Data

Source, scraping method, cleanup, etc.

As discussed above, our data originally came from the BLS. However, the raw BLS data is spread across several files and has slight methodology changes year-to-year, making it difficult to interpret. Instead we used the ATUS Extract Builder (ATUS-X). This tool allows users to select the fields of interest from the ATUS data and then download a subset of the original data. It is possible to select time use variables as well and have the system automatically determine average hours per day for these activities, but doing so produces a rectangular extract which removes the time of day information that we need for visualization. We instead opted for a hierarchal extract (shown below), which while more difficult to process, provides much richer information on individual activities.



In cart	Variable	Variable Label	Type	03	04	05	06	07	08	09	10	11	12	13	14	15
☑	CASEID	ATUS Case ID [preselected]	H	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	YEAR	Survey year [preselected]	H	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	PERNUM	Person number (general) [preselected]	P	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	LINENO	Person line number [preselected]	P	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	WT06	Person weight, 2006 methodology [preselected]	P	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	STATEFIP	FIPS State Code	H	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	COUNTY	FIPS County code	H	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	FAMINCOME	Family income	H	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	MONTH	Month of ATUS interview	P	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	DAY	ATUS interview day of the week	P	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	DATE	Date of ATUS interview	P	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	AGE	Age	P	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	SEX	Sex	P	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	RACE	Race	P	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	HISPAN	Hispanic origin	P	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	EMPSTAT	Labor force status	P	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	ACTIVITY	Activity	A	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	START	Activity start time	A	X	X	X	X	X	X	X	X	X	X	X	X	X
☑	STOP	Activity stop time	A	X	X	X	X	X	X	X	X	X	X	X	X	X

The raw data is a plain text file that looks like this:

```
120030100013280020030606037013
220030100013280020030100100106020030103815546 2.67215800006001011001000 2
320030100013280020031301240 4:00:0005:00:00
42003010001328002003
320030100013280020030102010 5:00:0005:30:00
42003010001328002003
320030100013280020030101010 5:30:0015:30:00
42003010001328002003
```

The first character encodes the record type, and subsequent character sequences encode a subset of the attributes shown above. We wrote a Python script that, given a configuration JSON file and the raw data file, creates four CSV files with aggregated activities and demographics. One CSV file is produced for each timescale of interest (day, week, year, and lifetime). Both the demographics and the activities of interest are set in the configuration file.

```
Minute,all:work,all:calls,all:social_relax_leisure,all:household_activities
…
0,0.01629,0.00029,0.01119,0.00498,…
1,0.01627,0.00029,0.01124,0.00500,…
2,0.01628,0.00028,0.01131,0.00489,…
```

These CSV files are suitable for final preprocessing in d3 before being displayed in the in visualization.

Exploratory data analysis

What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?

Because we were planning to create an exploratory analysis tool, we did less exploratory analysis in our initial planning that other projects may have. Instead we focused primarily on exploring how others had visualized the data (see related work). ATUS is a very rich data set, so we had to figure out which aspects in particular we wanted to highlight to avoid overwhelming the user.

None of the visualizations we found provided a lifetime view of the ATUS data, which we felt could provide useful insight into how people's habits change over time. To validate the feasibility of such a view, we used ATUS-X to quickly create an extract of respondent age versus time spent on education, leisure, and work. We used Excel to create a line graph of this data and found that clear, interesting trends were visible.

We also tried developing tools to find the medoid, or most central, respondent in a given demographic, with the thought that these values would represent a "typical" day. However, computing the medoid is an N^2 problem, so computing it for the entire data set is infeasible.

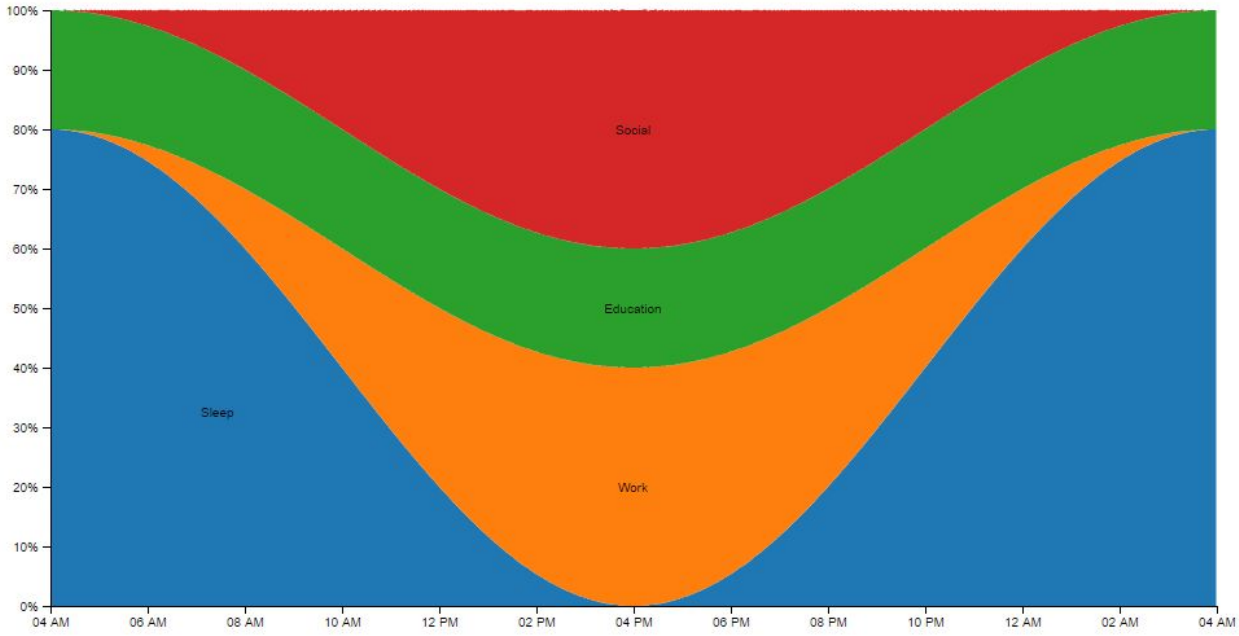
Finally, we looked at the distribution of activities to guide our choice of how to bin activities. Sleep, for example, is counted by the BLS under the category of personal care, but it is common enough that we felt it warranted its own bin. Other activities, like government services and calls, are so infrequent that they can be combined into a miscellaneous bin. Top level BLS categories were used for initial prototype design, but custom designed bins will be used in the final visualization.

Design evolution

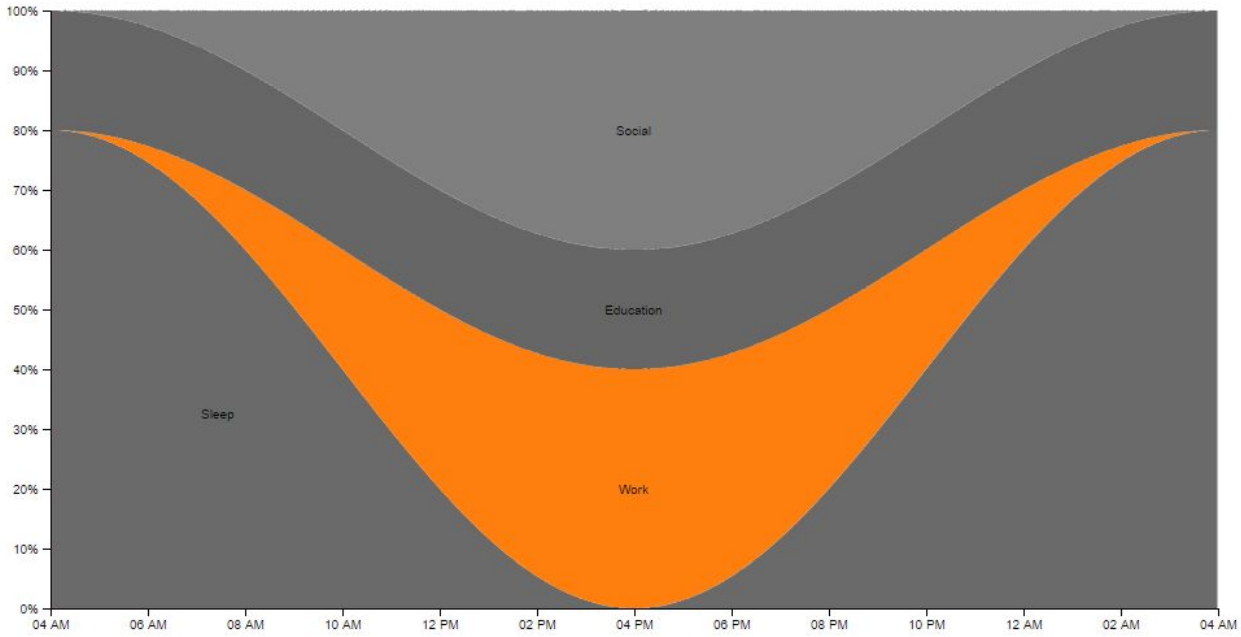
What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?

BRING IN SOME OF THE PROPOSAL IDEAS HERE SO THERE ARE MORE POTENTIAL DESIGNS

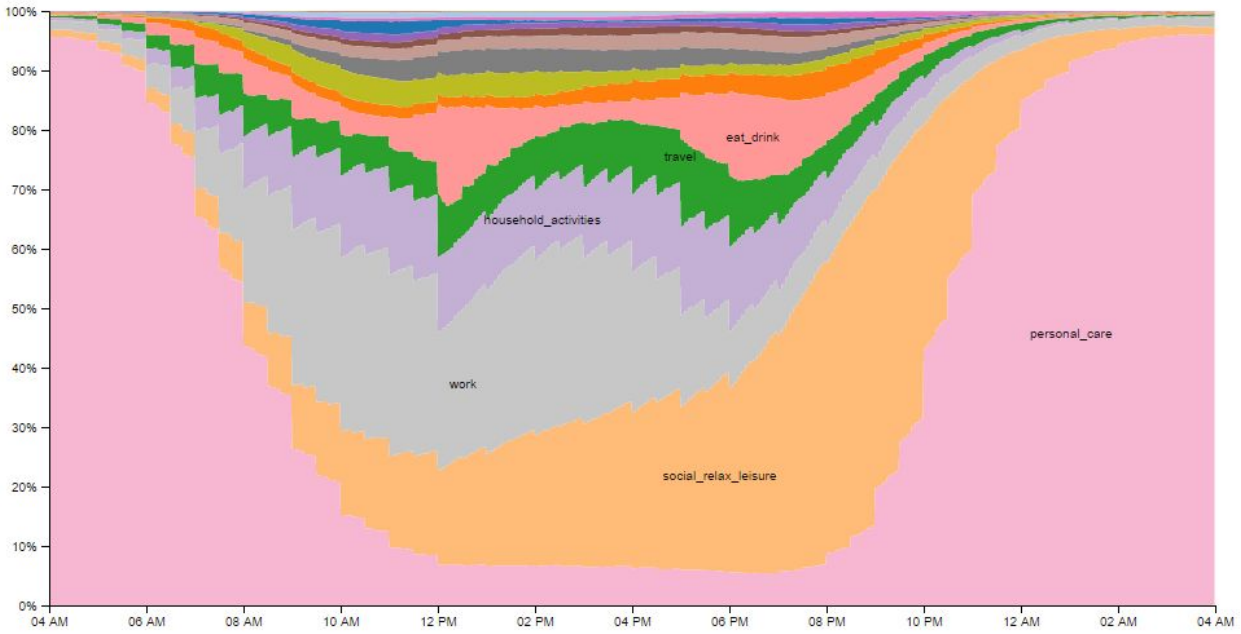
Most of our design ideas centered around a stacked area chart to provide a high level overview of the data, so we first developed a stacked area chart using synthetic data to ensure we understood how to build one in d3. We also experimented with approaches for automatically placing labels on the graph. In this screenshot labels are placed by finding the centroid of the largest area.



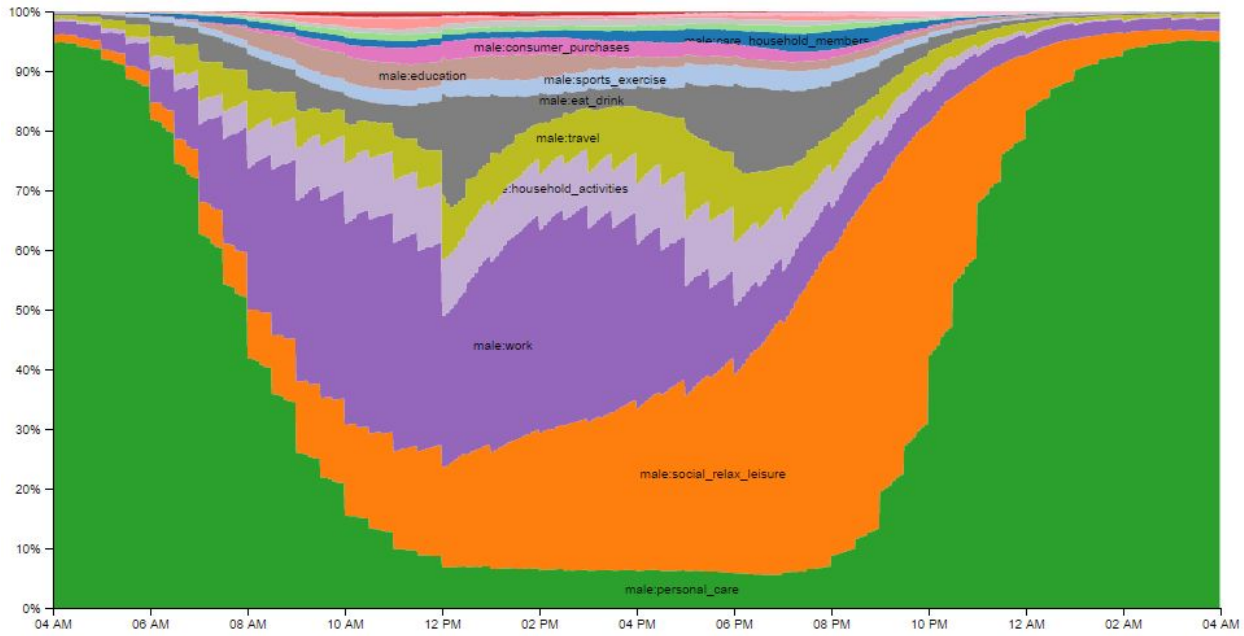
We also started experimenting with interaction techniques. This screenshot demonstrates desaturating the colors for every other activity when hovering over one activity. This technique makes use of preattentive visual pop out to make tracking the activity of interest easier.



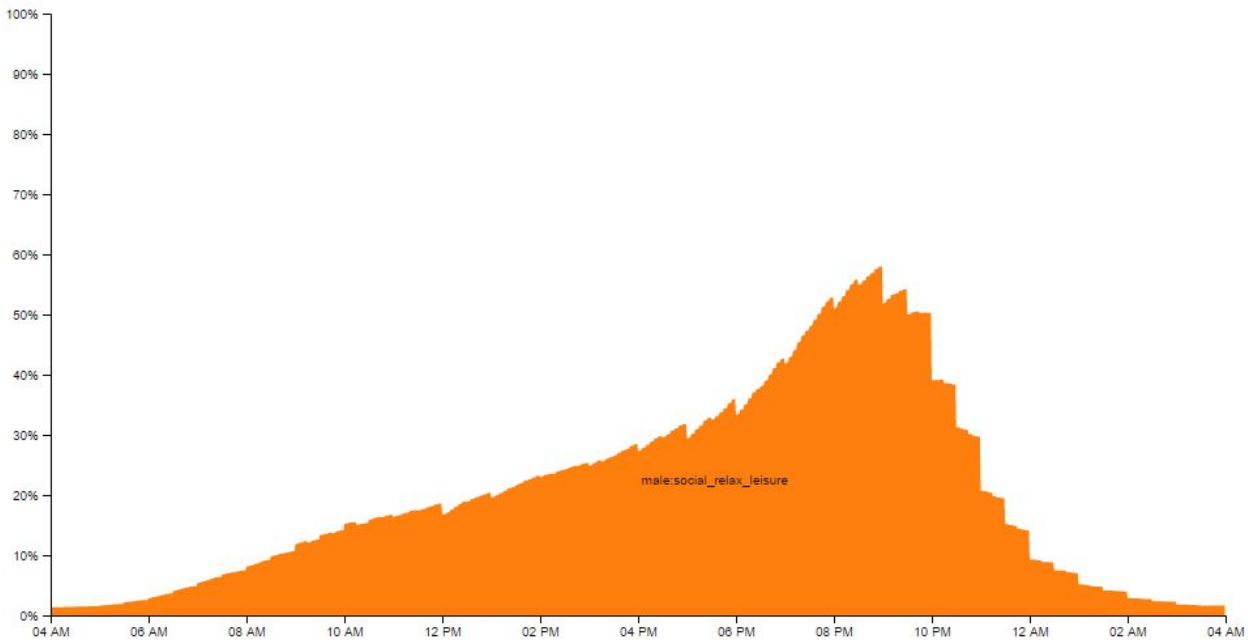
Confident in our ability to create the chart, we next tried incorporating real data into the visualization. This was a good test of our entire data processing pipeline. Already some interesting insights begin to emerge, like how people tend to get off work on the hour. There are also some problems with label placement beginning to show up. Finally, there are many thin, infrequent activities at the top of the visualization. Changing the binning of activity codes and providing alternative ways of visualizing these infrequent activities may help.



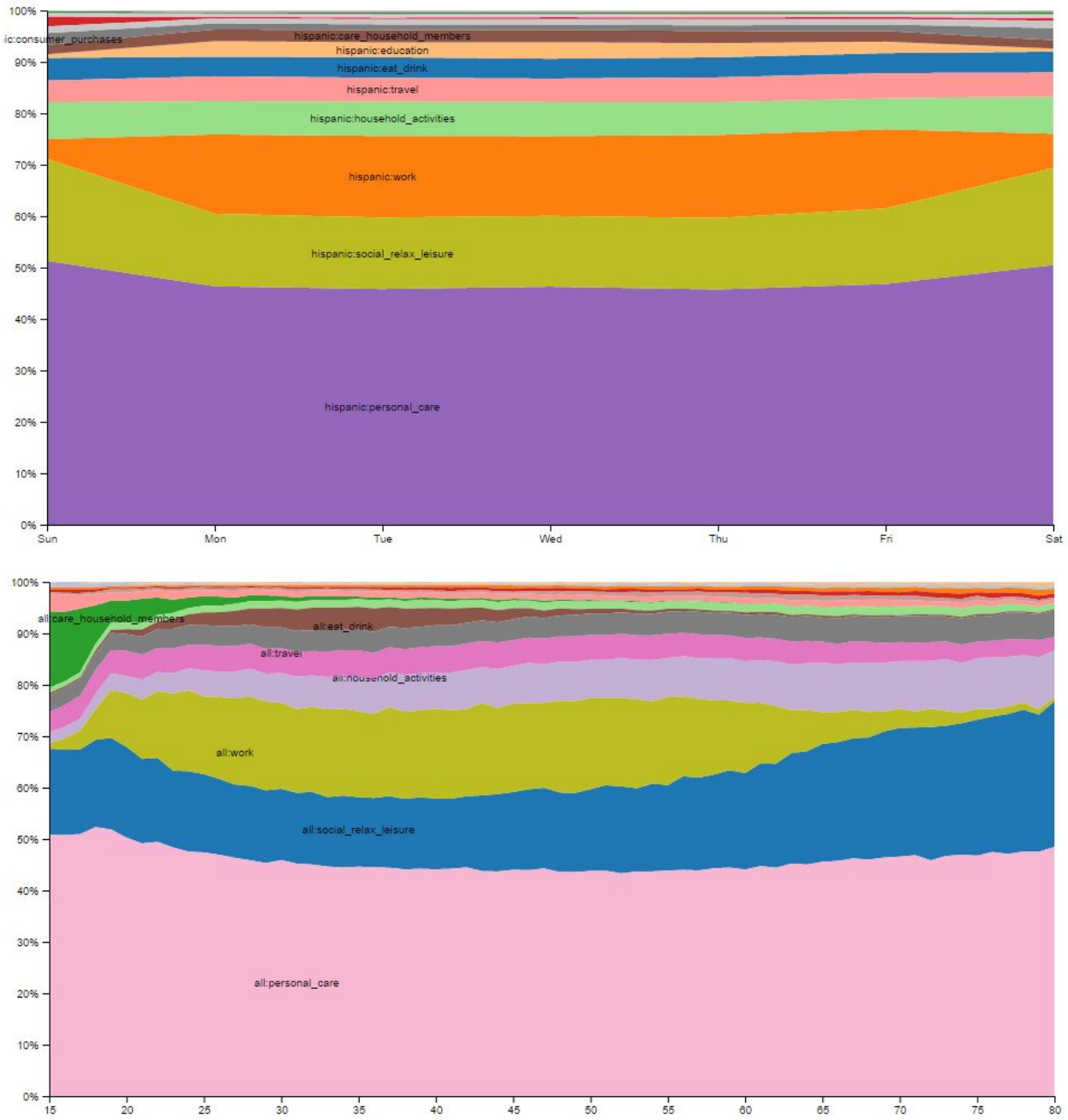
We also tried successfully visualizing the daily activity breakdown for a specific demographic (in this case male). Notice the slightly increased amount of work and decreased amount of household activities when compared to the average.



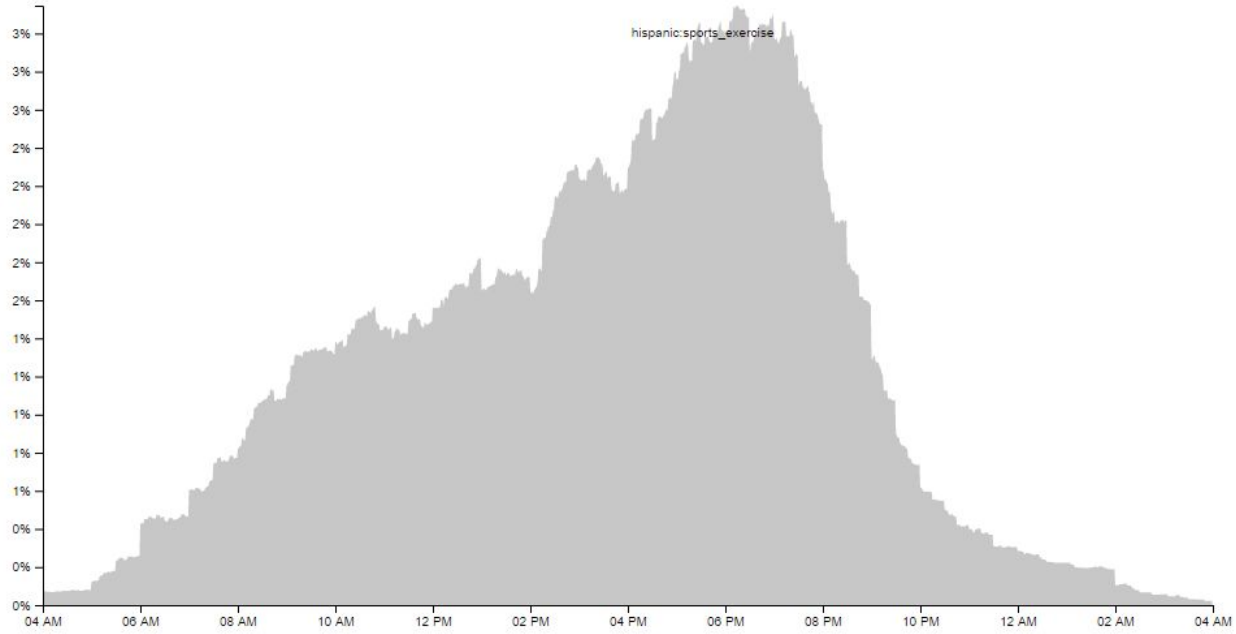
With a high-level summary visualization complete, we next turned to the visualization of individual activities. Stacked area charts do a good job of compactly conveying relative distributions over time, and they use a perceptually accurate length encoding. However, the visualization is less well suited for visualizing an individual activity over time because each time slice does not share a common baseline. We developed an animated transition that fades out all other activities, then slides the selected activity down to a common baseline when it is clicked. This new visualization provides better perceptual accuracy and the transition helps the user maintain context while the visualization is transformed. Here you can easily see how leisure activities increase in the evening as people come home from work.



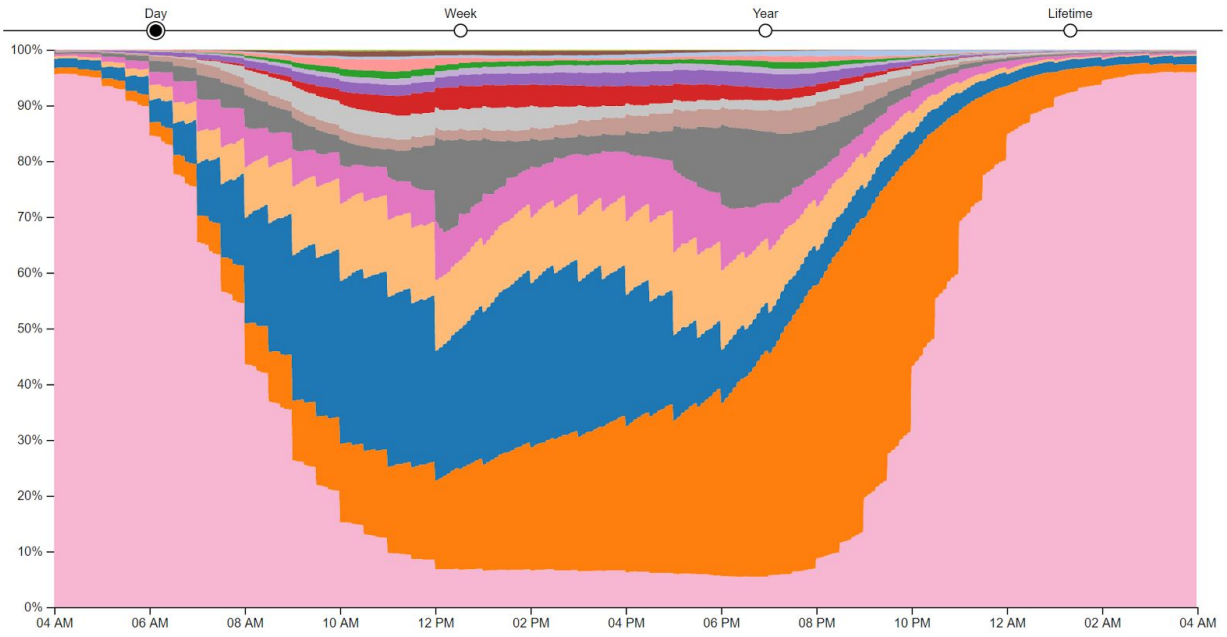
We also started developing the code necessary to support additional timescales. Week views and lifetime views are shown below.



One thing we noticed while developing the transitions between the individual activity view and complete view was that the resulting aligned baseline chart was often nearly unreadable for infrequent activities, like sports and exercise. To fix this, we adjusted the transition to align the baseline, then rescale the y axis to put the peak of the selected activity at the top of the chart. This two-step transition helps the user maintain context throughout the entire transition while improving the readability of the final chart.

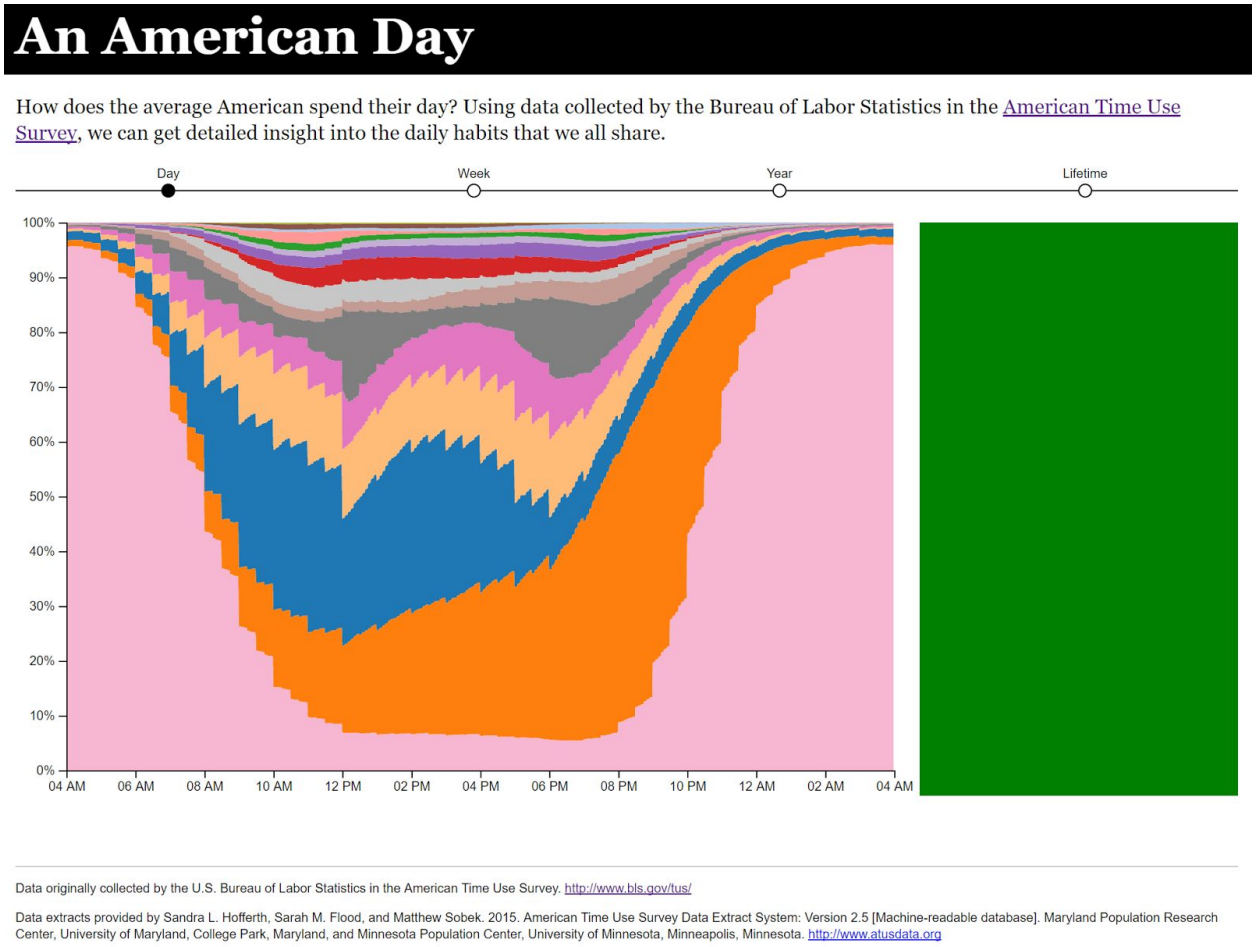


At this point we felt confident in the base view of our visualization, and started adding additional views to provide additional insight and context. The first thing we did was add a slider at the top of the visualization to select the timescale of interest. Transitions between timescales are animated. Placing this element at the top is appropriate because the timescale provides the context for every other element of the visualization.



Next we began drafting how the final website layout would look. The screenshot below shows a header, some draft introductory text, the timescale slider, stacked area chart, and a blocked out

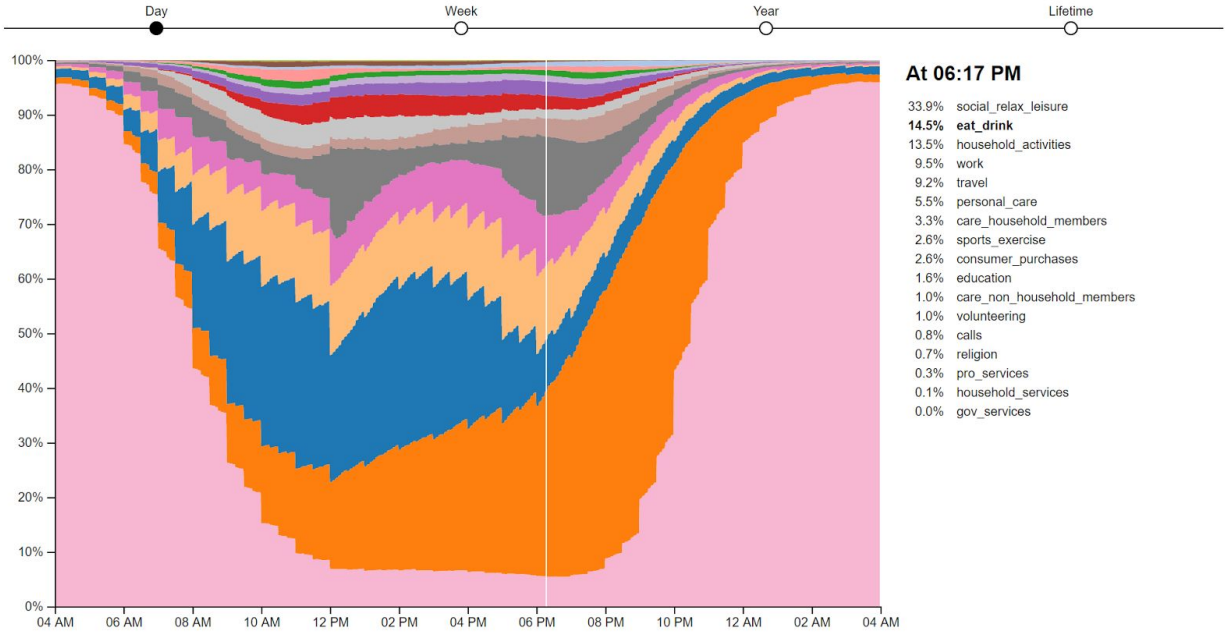
area for displaying detailed information when the user hovers over the area chart. Area labels are absent in this view because we are still determining the best way to place them.



The sidebar view shows exact percentage breakdowns for each activity either on average if the user is not hovering over the stacked area chart, or at the specific instant they are hovering over. A thin vertical cursor helps the user understand which moment in time they are currently selecting. The list of activities in the sidebar is automatically sorted in order from most frequent to least frequent. The interaction technique is designed to only provide detailed information when the user requests it to avoid overwhelming them. The activity that the user is hovering over is displayed in bold in the sidebar. This provides some context, but direct activity labeling is still needed.

An American Day

How does the average American spend their day? Using data collected by the Bureau of Labor Statistics in the American Time Use Survey, we can get detailed insight into the daily habits that we all share.



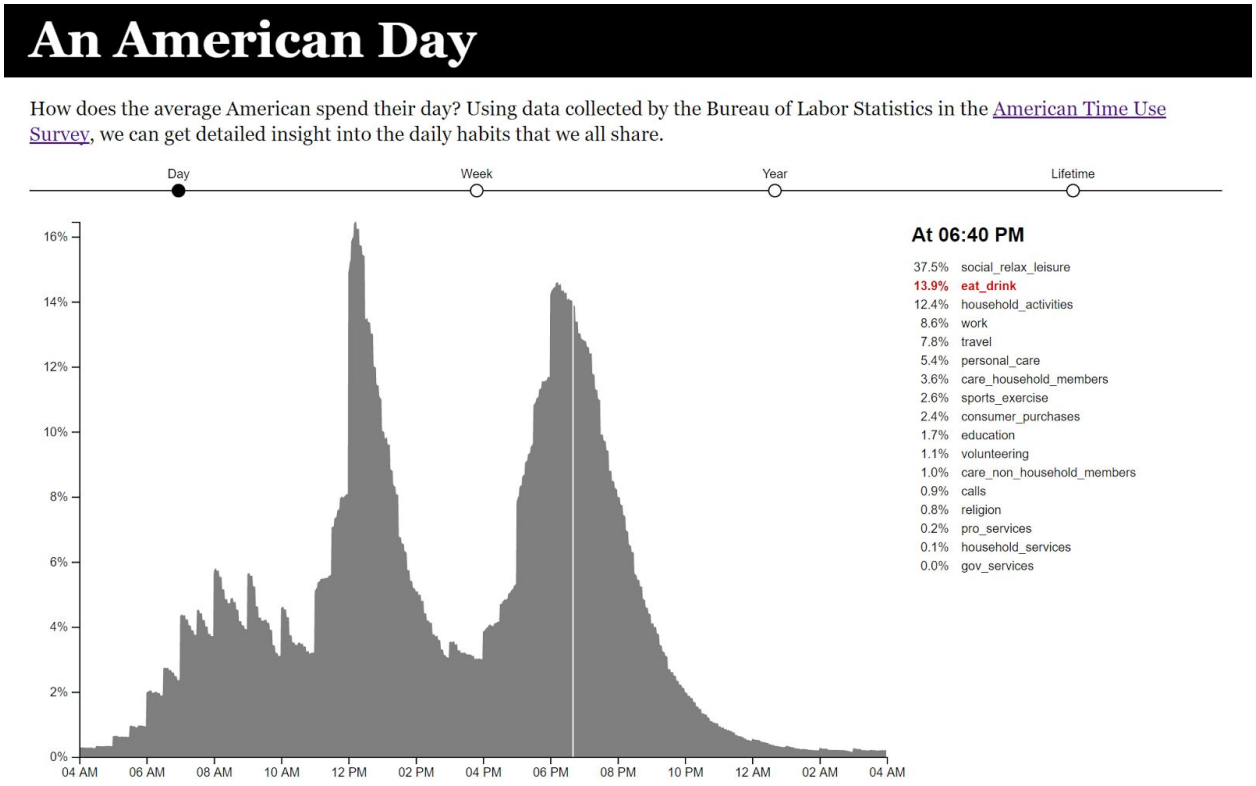
Data visualization by Ben McMorran and Francisco Sanchez.

Data originally collected by the U.S. Bureau of Labor Statistics in the American Time Use Survey. http://www.bls.gov/tus/

Data extracts provided by Sandra L. Hofferth, Sarah M. Flood, and Matthew Sobek. 2015. American Time Use Survey Data Extract System: Version 2.5 [Machine-readable database]. Maryland Population Research Center, University of Maryland, College Park, Maryland, and Minnesota Population Center, University of Minnesota, Minneapolis, Minnesota. http://www.atusdata.org

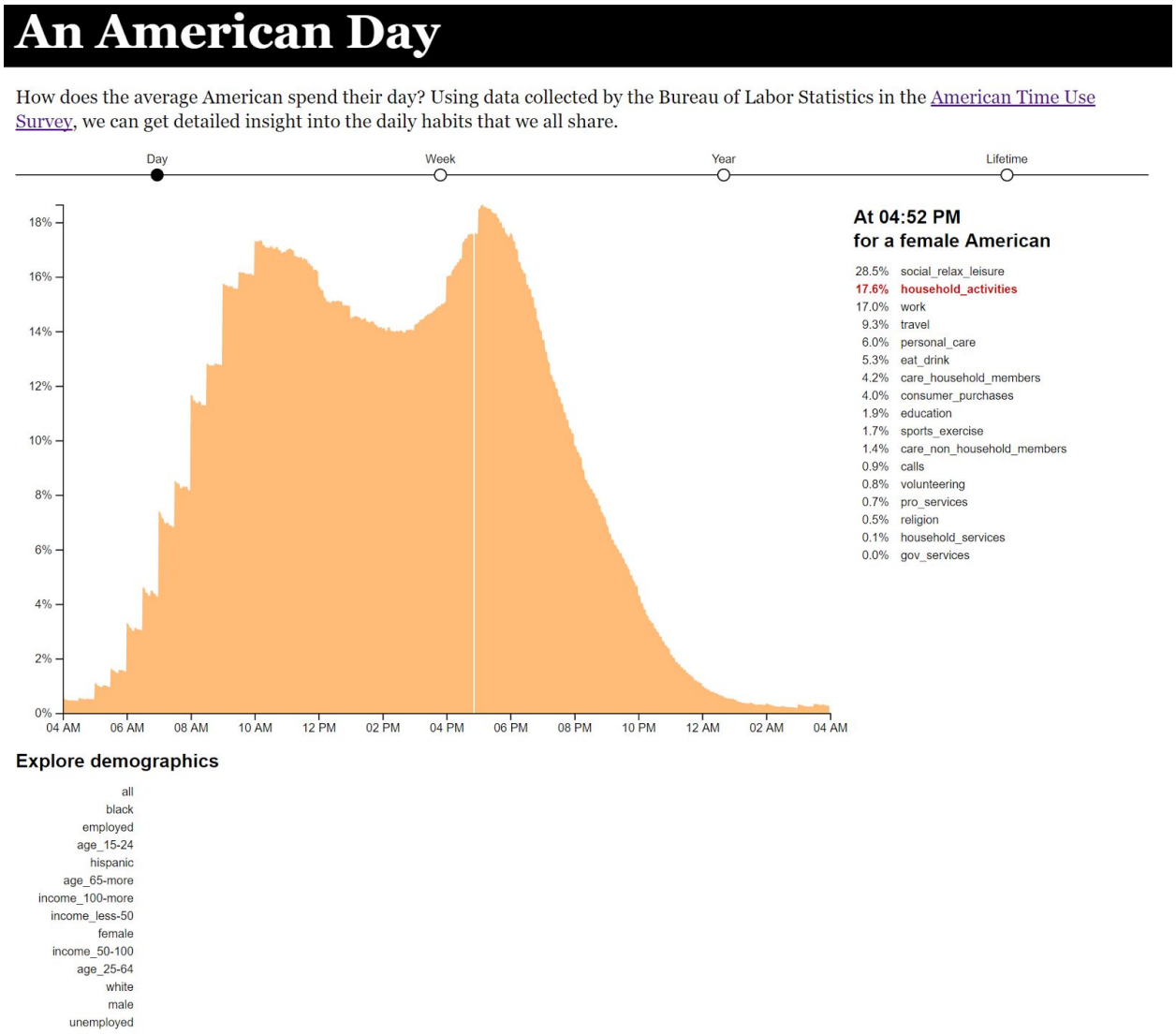
Favicon made by Freepik from www.flaticon.com. Used under CC 3.0 BY.

The sidebar displays the same information when an individual activity is selected. This helps provide global context even when viewing detailed information.



We next began building demographic information into the visualization. The list of demographics in the lower section of the visualization are clickable, and this action will trigger an animated transition in the area chart from the current demographic to the selected demographic. Animating this transition is important because it highlights the relative differences between demographics. Initially this animation both rescaled the y axis and animated the shape of the area in the same transition, but animation does a poor job of showing absolute differences in activity prevalence between demographics. Instead we settled on a two-step transition that first animates the shape on a fixed axis, then rescales the y axis to normalize the graph again.

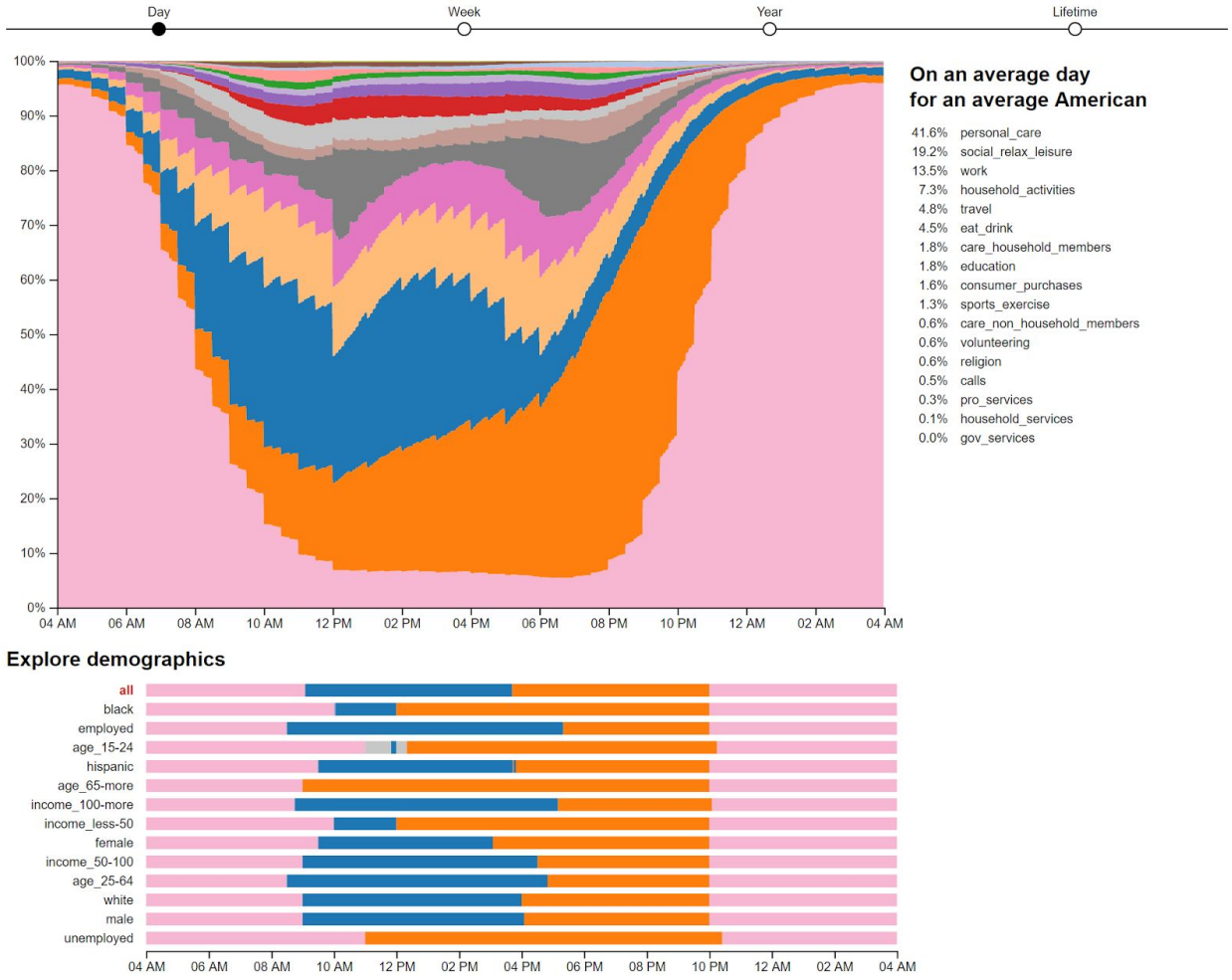
The sidebar text also updates to remind the user of the demographic they are currently exploring, in this case for females.



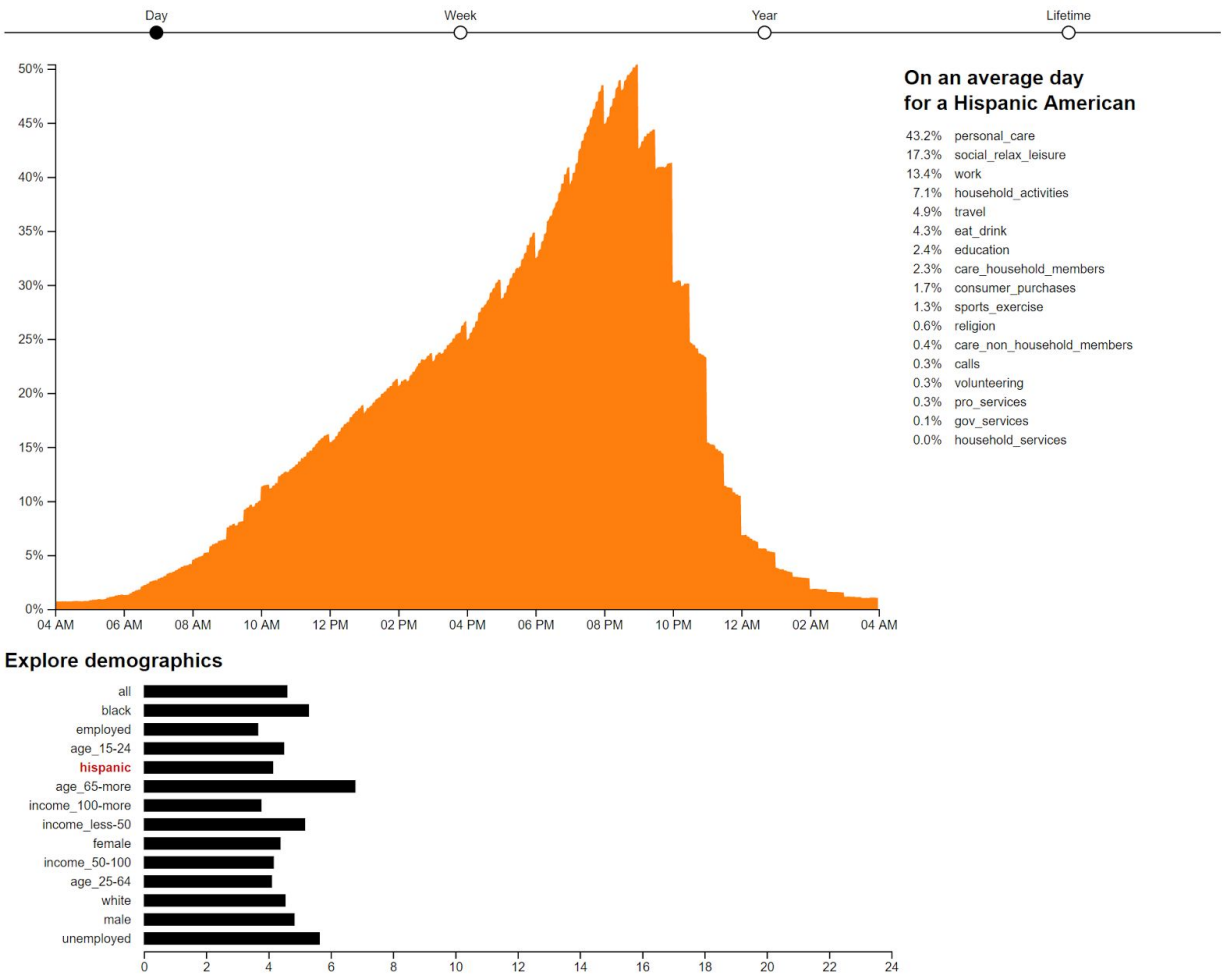
The screenshot below shows a first attempt at providing a summary visualization for each demographic in the demographics section. This visualizations shows the most common activity at each time of day for each demographic. While visually interesting, we are concerned that this visualization is potentially misleading because no one person in the dataset has a day like this. For example, it is unlikely that the average black American that works works only 2 hours per day, but it would be easy to draw such a conclusion from the visualization. Additionally, labels are needed to avoid depending entirely on color for activity encoding.

An American Day

How does the average American spend their day? Using data collected by the Bureau of Labor Statistics in the American Time Use Survey, we can get detailed insight into the daily habits that we all share.



The demographic view also adjusts to display average hours per day in each demographic when an individual activity is selected. This visualization provides context for the selected activity can suggest other demographics for the user to explore. In this example, we can see that older people tend to do leisure for significantly more hours per day than other demographics.



Implementation

Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.

COMPLETE ONCE VISUALIZATION IS FINALIZED

Evaluation

What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

COMPLETE ONCE VISUALIZATION IS FINALIZED