­­­­­­­­CS 573 Final Project Proposal

# Basic Info

Project title: An American Day

Team members

* Ben McMorran ([bjmcmorran@wpi.edu](mailto:bjmcmorran@wpi.edu), benmcmorran on Github)
* Francisco Sanchez ([fmsanchez@wpi.edu](mailto:fmsanchez@wpi.edu), friscis on Github)

Project repository: <https://github.com/benmcmorran/DataVisFinal>

# Background and Motivation

The reason the American Time Use Survey (ATUS) data was selected for this project was because this data consists of a large amount of data that can be approached and analyzed in countless different ways, which allows us to come up with our own unique approach to visualizing our analysis. Another reason, we selected this dataset was because when we searched for previous visualizations of this data most of them seemed to do a poor job at conveying the results of their analysis. As a result, we thought we could apply the skills and knowledge from this course to develop an interactive visualization that is versatile and concise. Some personal motivations for choosing this project include our own curiosity for how different demographics influence the ways people utilize their time, as well as being able to explore and see where our own time use fits best within the data.

# Project Objectives

We identified three main questions that our visualization will try to answer.

1. Which activities do Americans do most commonly throughout different time periods (day, week, year, etc.)?
2. How do common activities vary by location within the US?
3. Are there significant differences in activity prevalence between different classes (age, sex, race, education, and income)?

Answering these questions should give us a better understanding of how Americans spend their time and could reveal interesting differences between time use for different classes of Americans. For example, we may find gender differences in the amount of time spent on housework. These differences can help to quantify how far we are as a society from an egalitarian ideal.

Additionally, this project is designed to appeal as a casual exploratory tool to a wide audience. Everyone has experience with time use from their daily life, and they will likely find comparing their own experiences to the averages to be an interesting and fun exercise.

# Data

Original data comes from the Bureau of Labor Statistics (<http://www.bls.gov/tus/>) in the form of several linked tables for each year of the survey with information about activities, respondents, and households of respondents. However, this raw data is very detailed and can be difficult to work with, so instead we will be collecting our data using ATUS-X, the American Time Use Survey Extract Builder (<https://www.atusdata.org/atus/>). This online tool provides an easy interface for extracting relevant attributes from the large ATUS data set and summarizing activities at the daily level if desired.

# Data Processing

The data obtained from ATUS-X is already mostly clean, but some processing will still be required to extract the quantities of interest. ATUS-X data is provided at either the respondent level or the individual activity level. Both of these data sets are far too fine grained for our visualization needs. A preprocessing tool will be need to aggregate the individual activity data into distributions of activities by time of day, and to aggregate by respondent attributes like age, sex, and race. Additionally, the survey does not use uniform random sampling, but instead oversamples minorities and certain locations to ensure good statistics for these populations. A weighting factor is assigned to each respondent to account for this oversampling, and this weight will need to be incorporated into our preprocessing tool to ensure that overall averages accurately reflect the American population.

We are planning to implement the preprocessing as a Python script that can be run directly on the CSV files produced by ATUS-X and output CSV files suitable for consumption by d3. Python has good CSV support built-in and none of our preprocessing is statistically intensive enough to demand the use of a dedicated statistical tool like R. Both team members are already familiar with Python, making it a natural choice.

# Visualization Design

Blah blah blah

Top Tier Hierarchy – smaller set of groups, makes visualization cleaner and allows for easier color assignment

# Must-have Features

* Stacked area chart that displays the percentage of respondents doing a given activity throughout a day
* 24-hour breakdown of the activities each demographics is doing during each block of time
* Selectable demographics to compare how each demographics’ stacked area chart compares to the entire data set and other demographics
* Line graph for each activity, accessible by clicking the activity on the stacked area chart
* Average number of hours per day per demographic based one which activity line graph is showing
* Checkboxes for each demographic to add to the line graph for any activity
* US map that is gradient shaded based on which states (or counties) report the highest average number of hours

# Optional Features

* Selector for different time scales for the stacked area chart (week, year, and lifetime)
* Hover vertical cursor and expanded info side panel with exact data values for stack area chart
* Hover vertical cursor for the line graph with expanded info side panel and real-time updates to the US map
* User survey to input your own data and see what demographic your data fits best into

# Project Schedule

Our project schedule is broken down weekly. Dates indicate deliverables for that date.

* November 1: Project proposal completed.
* November 8: Complete preprocessing tools completed and non-interactive d3 implementation.
* November 15: Integrate real data with d3, add interactivity, tweak as needed.
* November 22: Collect feedback from friends and classmates on prototype.
* November 29: Integrate feedback, polish prototype visualization, collect more feedback in class.
* December 6: Integrate feedback, prepare final website.
* December 13: Polish process book, prepare final webcast, present in class.