This github project summary is broken down into three parts: Motivating Questions, Project Overview and Project Steps (Data Preparation, Website Preparation)

I) Motivating Questions:

Which issues has America shifted most on over any period between 2000 and 2022, and by how much? How do those issues change as different demographic cohorts are considered?

II) Project Overview:

Many [commentators](url) have observed significant shifts in America's views on issues such as abortion over the last two decades. This project aims to explore broader trends by identifying which issues have seen the most significant changes in public opinion. An interactive website has been developed to showcase these changes, allowing users to select different demographics and time periods for analysis. The data, sourced from the General Social Survey (GSS), benefits from rigorous collection methods: trained pollsters, compensated participants, in-person surveys, and sessions lasting over an hour, ensuring reliability and depth.

III) Project Steps:

a) Data Preparation

The foundation of this project is the GSS [dataset](url) where each row corresponds to a complete survey conducted with an individual, and each column represents a response or a demographic attribute. Initial data cleaning involved removing questions with low response rates or that were unrelated to beliefs (e.g. BALLOT, ballot type used in the interview). Subsequently, 16 data tables were created—one representing the entire population and the others for specific demographic groups like males and young people. These tables are stored in "..\own\_data\_objects\melted\_tables". A crucial function, compare\_years\_delta in toolbox.py, is used to generate the datatable seen on the website, using one of the melted tables, start year, and final year. This datatable is then sorted to highlight which issues have experienced the most substantial shifts in public opinion.

For the tooltips in column 1, or the values in column 2, I did the following: copied and pasted two pdfs ([GSS Codebook Index, and GSS Codebook Main Body](url)) into seperate text files, and used regex to create two dictionaries called labels and answers. “labels” provides the question for the label when hovering over column 1, and “answers”, which replaces the answer code (a number) with the actual answer seen in column 2 (e.g. GRASS, 1.0 is replaced by Should).

b) Website Presentation

Heroku was used for the hosting for the website, and Dashtable to create the datatable/ layout of the site.

Website that produces data table from gss data starting in 2000

How someone use the website: play around with it a bit. Spend one minute, hover over a couple questions.

* Remove bad variables like EVSTRAY
* Change widths based on screen size
* Five interesting findings playing around with the data
* Design Choices

GSS considered best survey in the country.

Decided not to do hyperlink, too much info. When show project to people, maybe include then

Ever wonder

Design Choices:

* Github page, do I hyperlink, add text to Hyperlink on website
* For David Shor, should I include a five things from the data
* GSS Data, where should I say it
* Get width of buttons programmatic with screen size
* Put some question (ever wonder) or more explanation
* Subcohort show for Whole Country, is just Whole Country clear or can I just say Country
  + Thoughts on “US Overview” (can show GPT suggestions)
* Do I put right after title in smaller letter: Select Cohort, Timeline
* Different formatting in middle column to make it pop more
* Draft email to Biden folks – should I put in examples of demographic and time period, or leave it out? By putting one example, can lead someone to be curious and see it

One Sentence:

* “Ever wonder on which issues America has shifted the most on? Find out below, selecting demographic and time period
* “Using the best survey in the country, the GSS, see which issues America has changed the most on given a time period and segment of the country.”
* “See which issues America has moved the most on – using the GSS.”

|  |
| --- |
|  |
| **Question** | **Answer** | **2000 Percentage** | **2022 Percentage** | **2022-2000 Delta** |
| GRASS | 1 | 33.5% | 70.2% | 36.6% |
| HOMOSEX | 4 | 28.8% | 61.2% | 32.4% |
| DIVLAW | 1 | 25.1% | 53.7% | 28.7% |
| PREMARSX | 4 | 41.8% | 69.1% |  |

Note:

* Some questions are only asked on certain years, e.g. HAPUNHAP (02, 08, 18, 22) – “happy with life?”
* Question year pairings under 850 positive responses, and some questions did not have enough respondents to show in a given year e.g. GRASS (2004) – should marijuana be legal.

Pre Notes:

* See which questions the American public has shifted most given the adjustable
* GSS Data Table: Sorted by questions America has shifted the most on given the adjustable parameters below

Shor email: “Just want to help Democracy’s survival.”

Code Process:

1. Df 🡪 each row is survey, columns are questions or stuff about individual
   1. Use YearQuestion Pairings
   2. Fix Age, Degree, Partyid fix
   3. Get new dataframe, proud of, put in database
2. Melted table 🡪 Each row has YEAR, none or param, Question, Answer, %, (Total Answers)
   1. Modify melted, e.g. 1000 answers
   2. Include params of interest
   3. Should have 16 melted params (1 generic, 15 params based)
   4. May not want Total Answers to show
   5. When click param, another button automatically should pop up and one is already selected
3. Final table 🡪
   1. Given options, spits out what want

**Fonts**

* **General Text and Headers**: The font family used throughout the website, including the table, is **'Arial', sans-serif**.

**Font Sizes**

* **Dropdowns, Input, and General Text**: There's no specific font size set in the CSS, which means it defaults to the browser's standard, which is usually around 16 pixels for body text.
* **Table Data and Headers**: Again, there's no explicit size set, so it defaults to the browser's standard.

**Colors**

* **Background Colors**:
  + **Page Background**: **#f4f4f9** (a very light gray)
  + **Table Odd Rows**: **rgba(0, 0, 0, 0.05)** (very light gray for alternating rows in the table)
  + **Table Hover**: **rgba(220, 220, 220, 0.8)** (light gray when hovering over a row)
  + **Table Headers**: **white**
* **Text Colors**:
  + **Main Heading (h1)**: **#333** (dark gray)
  + **General Text and Table Data**: There's no specific color set, so it defaults to the browser's standard black.

**Borders and Styling**

* **Table Borders**: **1px solid #dee2e6** (a light gray border around the table and cells)
* **Input Borders**: **1px solid #ccc** (light gray border around inputs)

This setup from the Dash application uses Bootstrap themes, which could also influence other stylings like button colors, margins, or paddings that aren't explicitly overridden in your custom CSS.

**Why**: PDF, Rstudio send. Less is more. Then in couple weeks, can hit him with powerpoint/ github repo. But I also believe honestly, that the world should have something like this. Like should be easily searchable, which issues have shifted most among the country, or among specific parts of the country.

**Shor email:** With email, hey created this thing I think is interesting. Anyone doing boring grunt work I could do for them? Anyway can volunteer? Or is there a textbook I could study, and if mastered, can have just a shot to volunteer? Think can bring lot of value.

Goal: Get ordered list of which surveyed topics have shifted the most since 2000 – 2022, and 2010 – 2022. Ideally can get them in a format that is ordered. Get all data points where saw minimum of X movement in last 20 years. Inspired by [abortion tweet](https://twitter.com/davidshor/status/1773133260119273945).

|  |  |  |  |
| --- | --- | --- | --- |
| Question | Answer | Delta 2000 to 2022 (%) | Delta 2010 to 2022 (%) |
| … | … | … | … |
| … | … | … | … |

Or just given two years (eg 2000, 2022) gets three columns.

**First need:**

|  |  |  |  |
| --- | --- | --- | --- |
| Question | Answer | Year | % of answers |
|  |  |  |  |
|  |  |  |  |

1. Determine all the unique question, answer, year pairings
2. Determine how many times each question, answer, year combination appears
3. Determine how many times each question, year combination appears
4. Divide 3) / 2), place value in row 1) with the unique question, answer, year
5. Create table with all pairings given dataframe

OR

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Question | Attribute (Politics, age, etc…) | Answer | Year | % of answers |
|  |  |  |  |  |
|  |  |  |  |  |

Maybe

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Question | Attribute (Politics, age, etc…) | Answer | Delta 2000 to 2022 (%) | Delta 2010 to 2022 (%) |
| … | … | … | … | … |
| … | … | … | … | … |

GSS axis can break down: age, degree, health, politics, race, sex

Dimensions (72390, 6691)

Get following, for each question and year:

* Breakdown in responses, bunch of tables. Rows are answers, years are question
* Then can query each table for an answer,