New Goal:

1. PDF couple example tables
2. RStudio interactive with parameters: start year, final year, up/down, # to see
   1. Clean df 🡪 the main params, and know question and years, filter with that
   2. Labels and Answers:
      1. Labels 🡪 codebook index, just to add ones I don’t have
      2. Answers 🡪 whole new update, codebook main one, just to do ones in labels
         1. Need to manually put in AGE, DEGREE, PARTYID
   3. Confirm/ get tables working where can do extraparameter + still work. Get 6 melted table
   4. Get website (labels, answers appear on hover)
   5. Deploy to Heroku

\*\* Can’t aggregate and then split

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

**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,