Eliciting People's First-Order Concerns: Text Analysis of Open-Ended Survey Questions

Beatrice Ferrario
(Harvard)

Stefanie Stantcheva
(Harvard)

January 2022

SOCIAL
ECONOMICS
LAB

Large-scale Social Economics Surveys and Open-ended Questions

Surveys are a key tool for understanding people's views

Multiple choice questions are the backbone of most surveys, but:

- they may prime respondents
- they may omit relevant options

Open-ended questions do not constrain nor prime respondents

- Valuable tool for eliciting first-order concerns!
- Need text analysis methods

Related Literature

Role of perceptions of inequality and mobility for tax policy: Gimpelson and Treisman (2018); Hauser and Norton (2017); Alesina, Stantcheva, and Teso (2018); .

Perceptions (and misperceptions) of tax rates: De Bartolome (1995); Gideon (2017); Ballard and Gupta (2018); Rees-Jones and Taubinsky (2019); Chetty, Friedman, and Saez (2013); and Feldman, Katuscak, and Kawano (2016).

Perceptions (and misperceptions) of tax systems: Stantcheva (2021); Fisman, Gladstone, Kuziemko, and Naidu (2020).

Perception of own ranking in income distribution & experiences: (Cruces, Perez-Truglia, and Tetaz (2013); Karadja, Mollerstrom, and Seim (2017); Roth and Wohlfart (2018); Hvidberg, Kreiner, and Stantcheva (2020).

Text analysis of non-survey data: Antweiler and Frank (2004); Baker, Bloom, and Davis (2016); Groseclose and Milyo (2005); Gentzkow and Shapiro (2010); Tesei, Durante and Pinotti (2018); Gentzkow, Kelly, and Taddy (2019).

Text analysis of survey data: Rare, but growing! Roberts et al. (2014); Brugidou (2003); Stantcheva (2020); Houde and Wekhof (2021);

This Paper

- Uses data from two surveys on income and estate taxes
  - 5,140 U.S. respondents aged 18-70
  - Broadly representative of the U.S. population
- Presents the application of text analysis methods to open-ended survey questions
- Shows key results about people’s first-order considerations on income and estate taxation
  - Focus on partisan divergences for the talk.

Using Open-Ended Survey Questions

What do Open-ended Questions Measure?

From broad & big picture to narrow & targeted open-ended questions

- *"When you think about federal personal income taxation, what are the main considerations that come to your mind?"*
- vs. *"What would be the effects on the U.S. economy if the federal personal income taxes were increased?"* or *"Which groups do you think would lose if the estate tax were increased?*

First-order concerns & considerations that matter to people and are top of mind

- "Gut reactions" vs Profound views

Text Analysis Methods for Open-Ended Questions

Preprocessing steps

- Parse the answers to reduce the number of distinct text elements: remove punctuation, excess spaces, numbers, misspelled words, and “stop words.”
- Lemmatize to group all inflected forms of a word (e.g., “policies” and “policy;” “were” and “be”)...
- Optional: Remove words that feature in the question or that are structurally part of answers (“believe,” “should,” “think”...)

Word Clouds

- Choose the size of word groups (i.e., “n-grams”). Given short answers, 1- or 2-grams is best.
- Font size for each n-gram is proportional to its frequency
- Pros:
  - First step in visualizing the data and scanning answers
- Cons:
  - Do not account for synonyms

Keyness Analysis

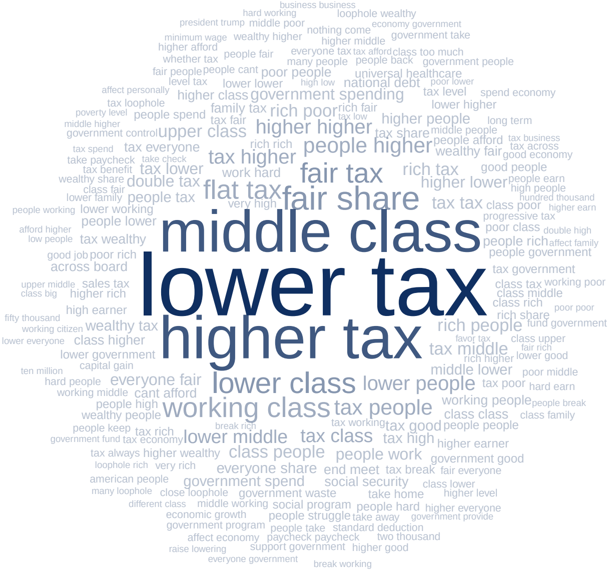
- Based on a relative frequency analysis
- Compare the use of n-grams between two groups (a reference and a target group)
- The keyness scores of an n-gram is based on the χ^2 test statistic for the null hypothesis that the propensity to use the n-gram is the same for the reference and target groups.
- Intuitively, words with high keyness scores are used relatively more frequently by the target group than the reference group.

Topic Analysis

- Topics are defined using keywords.
- Different approaches to extract topics and keywords.
  - from manual classification to semi-supervised and unsupervised algorithms
- Choice depends on the type of texts
- In our case, given length of answers and manageable sample size, topics are defined by sets of keywords we choose.
- Important to perform sensitivity checks

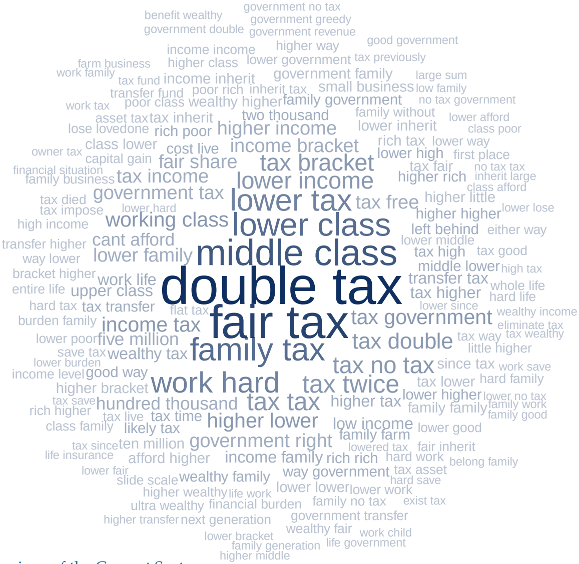
Application:
How Do People Think About Taxes?

What are you Main Considerations about the Income Tax System?



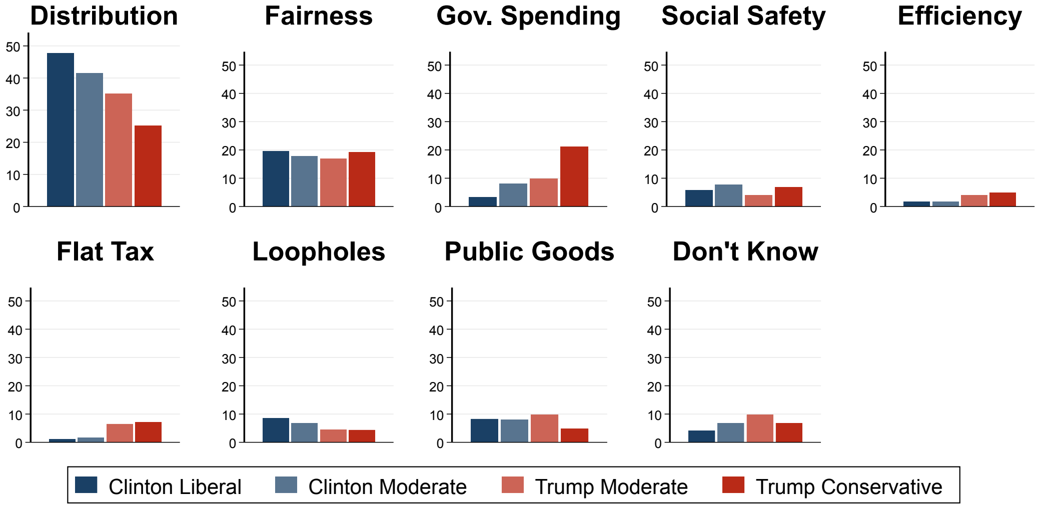
▸ Goals of a Good Tax System
▸ Shortcomings of the Current System

What are your Main Considerations about the U.S. Federal Estate Tax?



▸ Goals of a Good Tax System ▸ Shortcomings of the Current System

What are your Main Considerations about the Income Tax?
Relative Frequency of Topics by Political Views

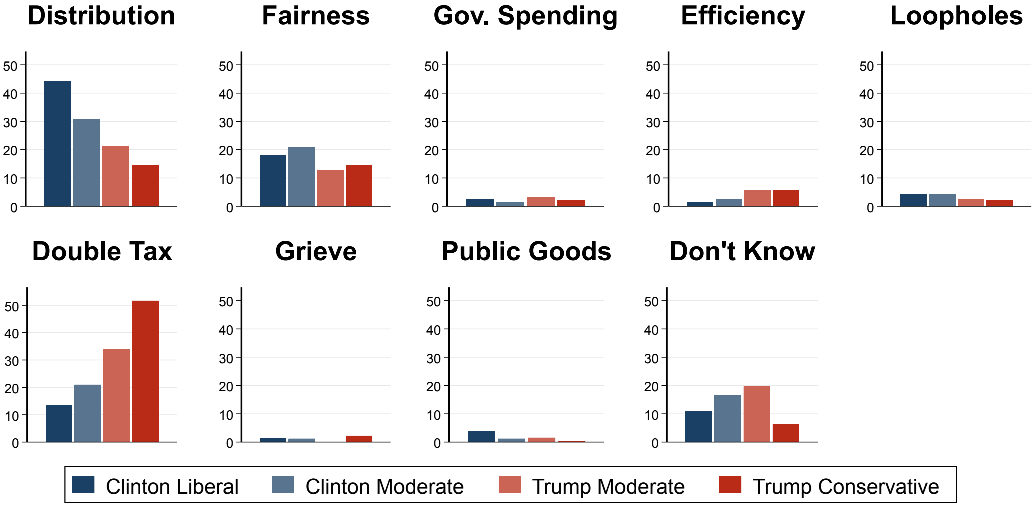


‣ Example of Answers by Topic ‣ Topics by Age Group

Topic Analysis: Keywords for the Main Topics Identified

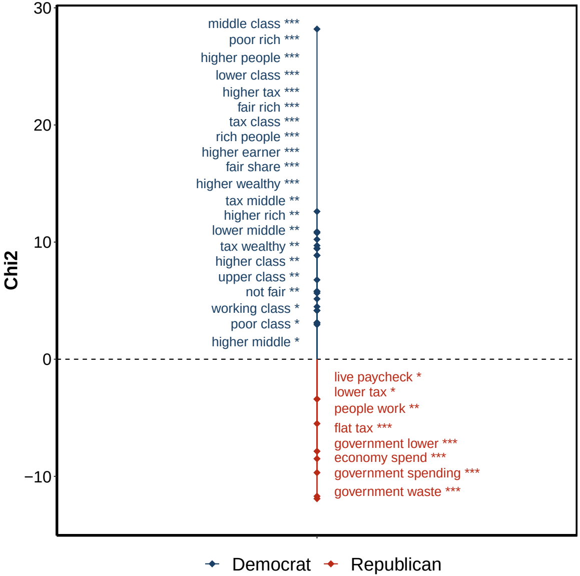
Distribution:	Middle class; working class; low income; wealthy; millionaire; rich; billionaire; corporations & pay/tax
Fairness:	Fair; unfair
Gov. Spending:	Government spending & high; government spending & cut; deficit; debt; government & waste; balance & budget; government & budget; government & control & spend
Social safety:	Social services; governmental services; governmental program & fund; governmental program & cover; help & poor; pay & poor; social program; poor work; live & paycheck; provide & family
Efficiency:	Hurt & economy; work hard; work less; work more; create & job; depress; negative/detrimental/destroy/damage & economy; competition; innovation; crea
Flat tax:	Flat tax
Loopholes:	Loopholes; lawyer; account; tax evasion; evade; avoid taxes
Public goods:	Infrastucture; education; healthcare
Don't know:	Not know; knowledgeable enough; idk; not sure; know enough; unsure
Double Tax:	Already taxed/paid; twice & tax/pay
Grieve:	Grieve; bury; funeral

What are Your Main Considerations About the Estate Tax?
Relative Frequency of Topics by Political Views

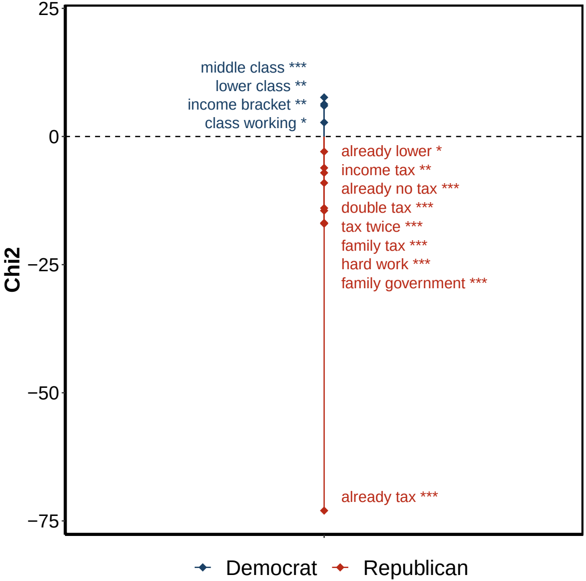


‣ Examples of Answers by Topic ‣ Topics by Age Group

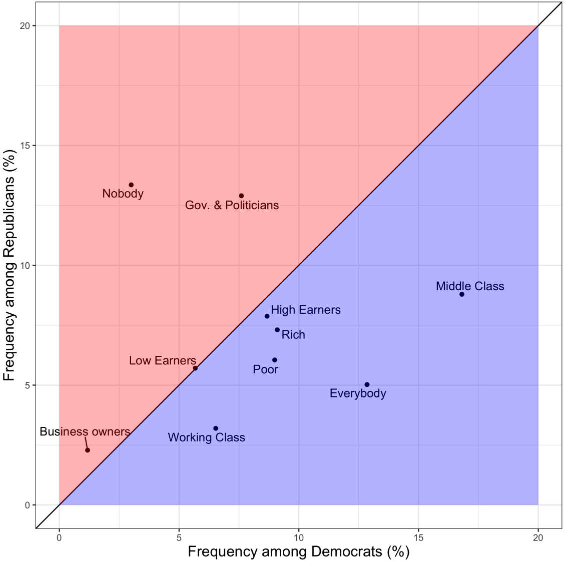
What are you Main Considerations about the Income Tax System?: Keywords by Political Views



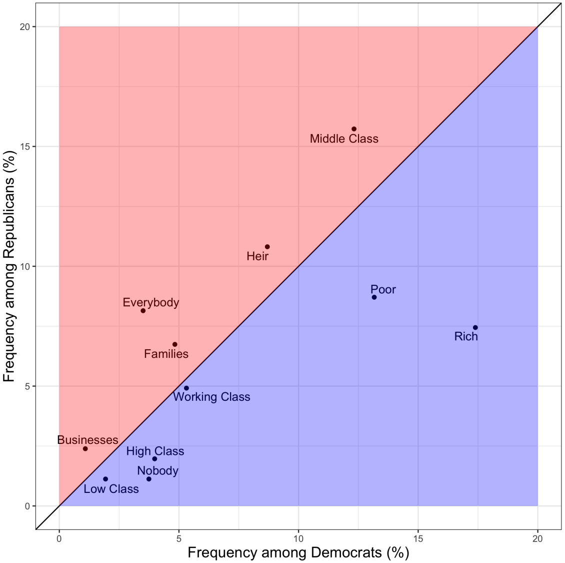
What are you Main Considerations about the U.S. Federal Estate Tax?: Keywords by Political Views



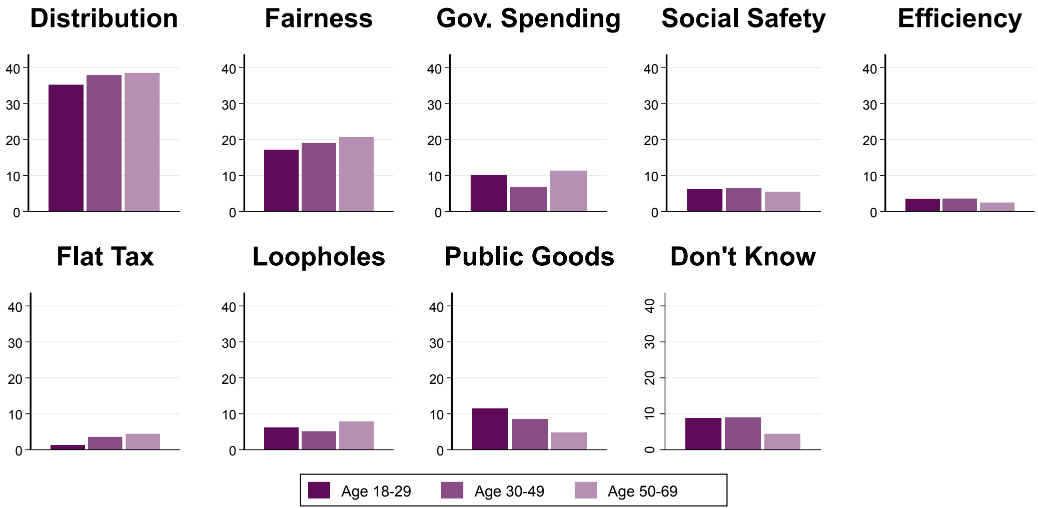
Distributional Effects of Income Tax Increase: Who Gains if Taxes on High-Earners were to be Increased?



Distributional Effects of Estate Tax Increase: Who Loses if the Estate Tax were to be Increased?

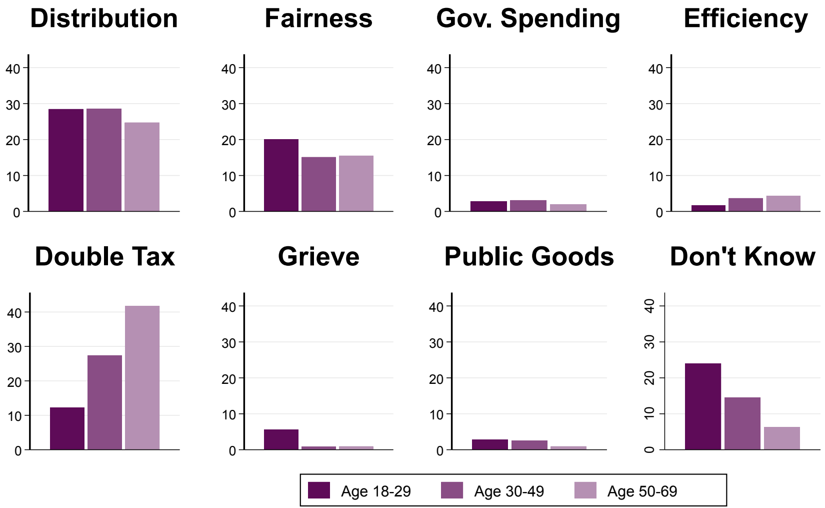


What are your Main Considerations about the Income Tax? Relative Frequency of Topics by Age Groups



‣ Back

What are Your Main Considerations About the Estate Tax?
Relative Frequency of Topics by Age Groups



Distribution

Fairness

Gov. Spending

Efficiency

Loopholes

Double Tax

Grieve

Public Goods

Don't Know

▸ Back

Conclusion

Open-ended survey questions offer the potential to elicit people's first-order considerations on policy issues.

By not constraining respondents to a given set of answer choices, they avoid priming them to think of otherwise non-salient options or omitting relevant options.

Can range from broader to more targeted.

Leverage recent & on-going advances in text analysis for visualization and quantitative study.

Can yield new insights across many areas of economics.

Appendix

Preprocessing Steps – Keyness & Wordclouds (I)

- Given an answer d_i :

1. Parse d_i : lower-case every word, remove punctuation, spaces in excess, numbers, misspelled words, common words that carry no intrinsic meaning (“stopwords”) such as “and,” “the,” “each,” “then.”
2. Lemmatizing remaining words, i.e. grouping together the inflected forms of a word so they can be analyzed as a single item.
   - Use Mechura’s (2016) English lemmatization list available from the lexicon package.
   - E.g., : “policies" becomes policy, “were" becomes “be". → reduces number of distinct textual elements

→ Output: $(\hat{d}_i)$

‣ Back

Preprocessing Steps – Keyness & Wordclouds (II)

3. Remove words coming from the question as well as extra words related to the structure of answer.
   - E.g., for the question, "What are your main considerations about the income tax system?": remove “main," “considerations," “income," “policy" from the answers, as well as “think," “believe," “should"...
4. Transform $\hat{d}_i$ into numerical vector c_i in which each element is a 2-gram, i.e. a 2-component expression of two words which were separated by 0 or 1 word in the original text. Group together 2-grams which correspond to the same inverted two words. Manually remove 2-grams which do not make sense and duplicated 2-grams (e.g., "tax tax").
   - E.g., take d_1 = “We should tax the wealthy more and tax the poor less.” After steps 1-2-3 becomes: $\hat{d}_i$ = “tax wealthy more tax poor less”. After step 4 becomes: [‘tax wealthy’ = 1, ‘tax more’ = 2, ‘wealthy more’ = 0 (because it is not grammatically coherent), ‘tax poor’ = 1, etc ...]

▸ Back

Preprocessing Steps – Topic analysis

- Given an answer d_i :

1 Parse d_i : lower-case every word, remove punctuation, spaces in excess, numbers, misspelled words, very common words that carry no intrinsic meaning (“stopwords”) such as and, the, each, then.

2 Reduce remaining words to common root (stemming)

  - Use Snowball stemming algorithm, by Porter (2001).
  - *policies* and *policy* become *polic* → reduces number of distinct textual elements

→ Output: ($\hat{d}_i$)

3 Transform $\hat{d}_i$ into numerical vector $\mathbf{c}_i$ in which each element is the count of a distinct textual token (either a word or an *n*-components expression, *n-gram*)

- F.e. take d_1 = *“We should tax the wealthy more and the poor less.”*
  - After steps 1-2 becomes: $\hat{d}_1$ = *“tax wealthi poor less”*
  - After step 3 becomes: [‘tax’ = 1, ‘wealthi’ = 1, ‘poor’ = 1, ‘less’ = 1, ‘tax wealthi’ = 1, ‘wealthi poor’ = 1, ‘poor less’ = 1, ‘house’ = 0, ...]

4 Generate topic dummy variables equal to 1 when an element of c_i matches a custom-made topic dictionary.

Keyness Score

Consider a given n-gram i. Let j be the group index, with $j = 0$ for the reference group and $j = 1$ for the target group.

Let $A_{i,j}$ be the observed number of occurrences of the n-gram i in group j and $A_{-i,j}$ the observed number of occurrences of all other n-grams (except the one we consider) in this group.

Let R_i be the total number of occurrences of n-gram i in both groups, C_j be the number of occurrences of all n-grams in group j, and N the overall number of occurrences of n-grams in both groups.

Compute $E_{i,j}$, the expected frequency of a given n-gram i in group j: $E_{i,j} = \frac{R_i \times C_j}{N}$ and $E_{-i,j}$ the expected frequency of all other n-grams in group j: $E_{-i,j} = \frac{R_{-i} \times C_j}{N}$

Keyness Score – cont.

The χ^2 test statistic is:

$$\chi^2 = (-1)^{\mathbb{1}\{E_{-i,1} > A_{-i,1}\}} \sum_{k \in \{-i,i\}} \sum_{j=0}^{1} \frac{(A_{k,j} - E_{k,j})^2}{E_{k,j}}$$

We compare this statistic to the distribution of a χ^2 distribution law with one degree of freedom (i.e., number of groups -1).

A negative χ^2 indicates that the word is significantly more frequent in the reference group. In absolute value terms, the null hypothesis is rejected at the 10% level when $|\chi^2| > 2.71$ (*), at the 5% level when $|\chi^2| > 3.84$ (**), and at the 1% level when $|\chi^2| > 6.63$ (***).

Main Considerations about the Income Tax? Example Answers by Topic

Distribution:

"That the rich and wealthy do not pay their fair share of taxes."

"Everyone, including the rich and corporations should pay their fair share."

"I would want working class and middle class people to get tax cuts and I'd be willing to pay more in taxes for that to happen."

Fairness:

"I have trouble with the concept of tax brackets that punish an individual for being successful";

"I believe Everyone should be taxed fairly and the most wealthy should not escape carrying their weight."

Gov. spending:

"Current tax rates being raised are a result of government mismanagement of funds and over spending without appropriate oversight. Taxes really can't effectively be lowered until government spending is properly controlled.";

"I am okay with raising personal income tax to reduce deficit but not for entitlement programs."

▸ Back

Main Considerations about the Income Tax? Example Answers by Topic (II)

Social safety net:

“What are the taxes going towards? I strongly believe in funding going towards education and infrastructure.”

“Cut government spending on social welfare programs for lower taxes and privatize most government services for lower taxes e.g. mail, law enforcement, parks, schools...”

Effiency:

“I am concerned about the push to raise taxes on persons with higher incomes. I do believe in trickle down economics and that government should pretty much keep their hands off.”

Flat Tax:

“We need a flat tax. Tax forms are complex.” ; “I think tax Rates are not fairly representative for most taxpayers. I support a flat tax rate for all except the totally disabled and indigent.”

Loopholes:

“I think the more you make, the more you should pay. We need to close the loopholes that are there to make sure that those who make more actually pay more.”

▸ Back

Main Considerations about the Estate Tax? Example Answers by Topic

Distribution:

"It can help keep the ultra wealthy accountable for their wealth."
"Passing wealth from one generation to the next contributes to wealth inequality. Federal estate tax should be much higher."

Fairness:

"I don't think there should be a federal estate tax because it's kind of unfair to have to pay taxes on money that already belongs to your family and has most likely had taxes paid on it already."

Gov. spending:

"I believe in smaller government, so all taxes should be lower. I actually think we should have a flat tax for income - period. Then estate taxes wouldn't even be an issue."

Public goods:

"I would like higher taxes to pay for more domestic spending such as education, healthcare, etc."

▸ Back

Main Considerations about the Estate Tax? Example Answers by Topic

Efficiency:

“Lower taxes mean I have more disposable income to spend therefore more products can be mad and more jobs created. I feel it is wrong to penalize people for increased wealth’

Loopholes:

“The wealthy don’t ususally pay these taxes, they find a loophole. Why should my children have to pay taxes on things I’ve already paid taxes on during my lifetime?”

Double taxation:

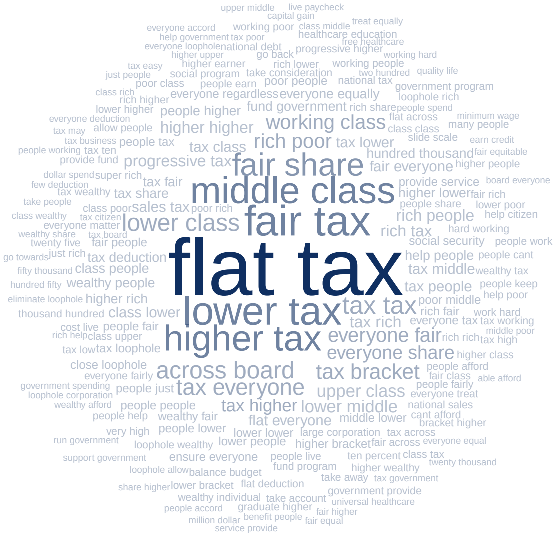
“I think it is ridiculous, you pay taxes twice.”

Grief:

“I don’t think we should have one at all. You’re taxing a family member for the death of their loved one? That’s messed up.”

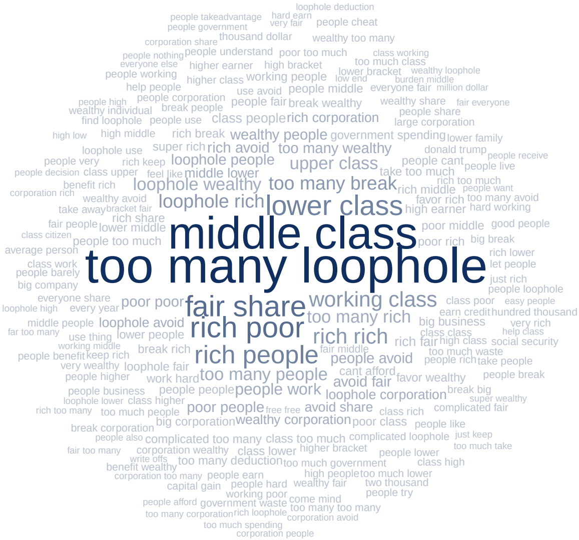
▸ Back

What would be the Goal of a Good Income Tax System?



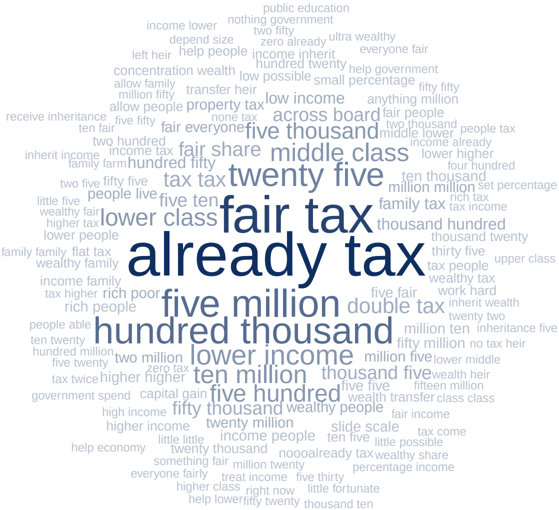
▸ Back

What are the Shortcomings of the Income Tax System?



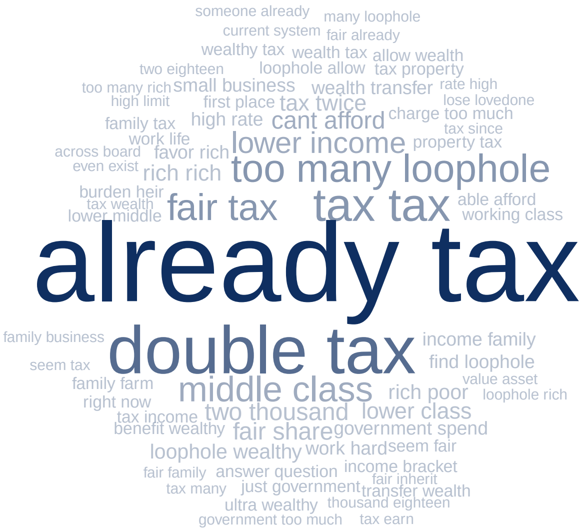
▸ Back

What would be the Goal of a Good Estate Tax?



‣ Back

What are the Shortcomings of the Federal Estate Tax?



▸ Back