Characteristics of an American Ideologue: Revisiting a Linear Relationship

Yousuf Farhan

1/20/2022

Abstract

The American General Social Survey 2021 sheds light on various societal benchmarks. This paper extrapolates on this data and demonstrates that anything close to a linear model cannot be formed between age, family income, level of educational attainment, political party affiliations and political views, providing a more focused yet holistic view on American political opinion. Some academic literature has established that a nigh-linear relationship is sufficiently accurate to make predictions about how individuals with certain characteristics are likely to vote. This paper uses this data to demonstrate otherwise. It revisits this phenomenon and refutes it entirely. It shows how a few selected socio-economic characteristics do not have a linear relationship with political party affliations or political views that is statistically significant, both individually and cumulatively. While a plethora of factors including but not limited to different forms of media and changes in educational curriculum may exist, this paper leaves these reasons up to speculation. However its findings strongly contrast scholarly literature on this topic which dates to a few decades ago to even a few years ago.

Introduction

The R language (R Core Team 2021) is used, in addition to a variety of packages and scholarly sources. Please refer to the Bibliography section or the "references.bib" file in the Git repository for more information. The relationship between economic and social classes with political views and party leanings has drawn the attention of many academics, statisticians and political scientists and commentators. Many have tried to claim that a linear relationship between the factors that define economic and social classes and political leanings and views exists (Dupont & Bateman, 2012; Burris, 1987). This paper will test whether this phenomenon still holds true. To statisticians, trends are very important. However, trends regarding social behaviour should be revisited frequently as societal opinion, economic conditions and contemporary problems change.

Data

How it was obtained

The entirety of the data used as input in this paper was obtained through the American General Social Survey 2021. The data is free to use for academic purposes. It was saved as a Stata file in the data subdirectory of the inputs folder, along with ancillary documents regarding how to read it.

Preparing the data

The data file consists of five-hundred and sixty-five columns. This paper uses four of these to account for the age, family income, political party affiliation and political views variables. Using the documentation presented in the "GSS 2021 Codebook R1b.pdf" file in the data subdirectory of the inputs directory, we can described the intricacies of how the data is represented.

In the dataset political party affiliation exists on a spectrum, with values ranging from 0, which shows that the individual strongly supports the Democrats, to 6, which shows that the individual strongly supports the Republicans. The values 1 to 2 represent an individual having moderate support and slightly leaning towards the Democrats respectively. Values 2 and 4 represent individuals that lean slightly towards Democrats and Republicans. The value of 3 represents individuals that do not lean towards either of these two major parties at all. Values 5 and 6 represent an individual being mildly penchant towards the Republicans and being strong Republican supporters respectively.

Furthermore, income is input by a respondent based on multiple ranges. For this reason, we will treat income range as categorical data. This variable represents household income. These household income ranges start from under $1000, represented by the integer values of 1, to $25,000, represented by the integer value of 12. There are ten incomes ranges inbetween, each with a $1,999 increment. A refusal to answer is represented by 13.

Education level ranges in years from 0, representing no formal education to 17, which represents five years of college education or higher.

Similar to political party affiliations, the political views of respondents are also assigned integer values ranging from 1 to 7. 1 and 2 represent that the individual is "Extremely Liberal" and "Liberal" respectively. The values 3, 4 and 5 represent "Slightly Liberal"," Moderate, Middle of the Road" and "Slightly Conservative" respectively. "Conservative" and "Extremely Conservative" are represented by the integers 6 and 7 respectively.

There were a total of 146 null values for the column which represented an individuals party affiliations. These rows were omitted.

Results

Lets take a look at the how respondents answered when asked about their political party affiliations.

Levels of Predesposition towards the Democratic and Republican Parties
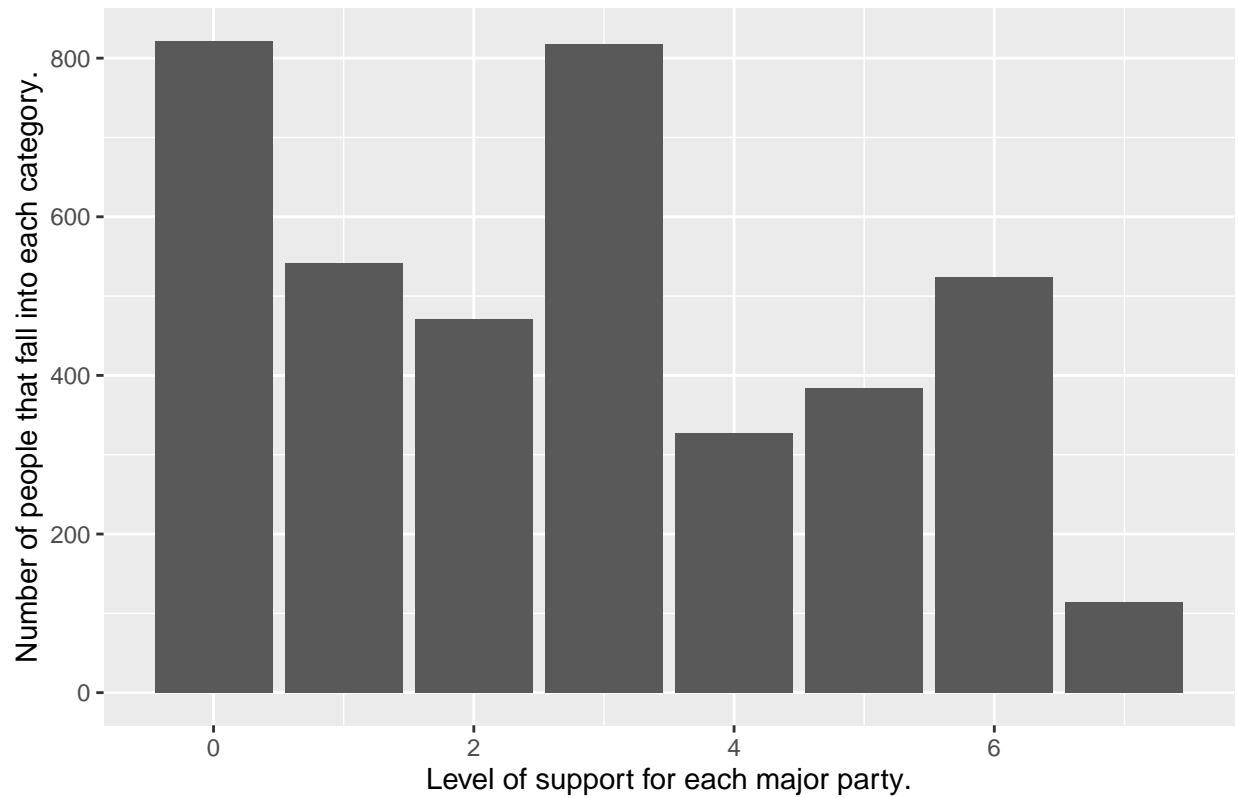
To make this simpler for analysis, lets use the categories described in the previous subsection to re-plot this.

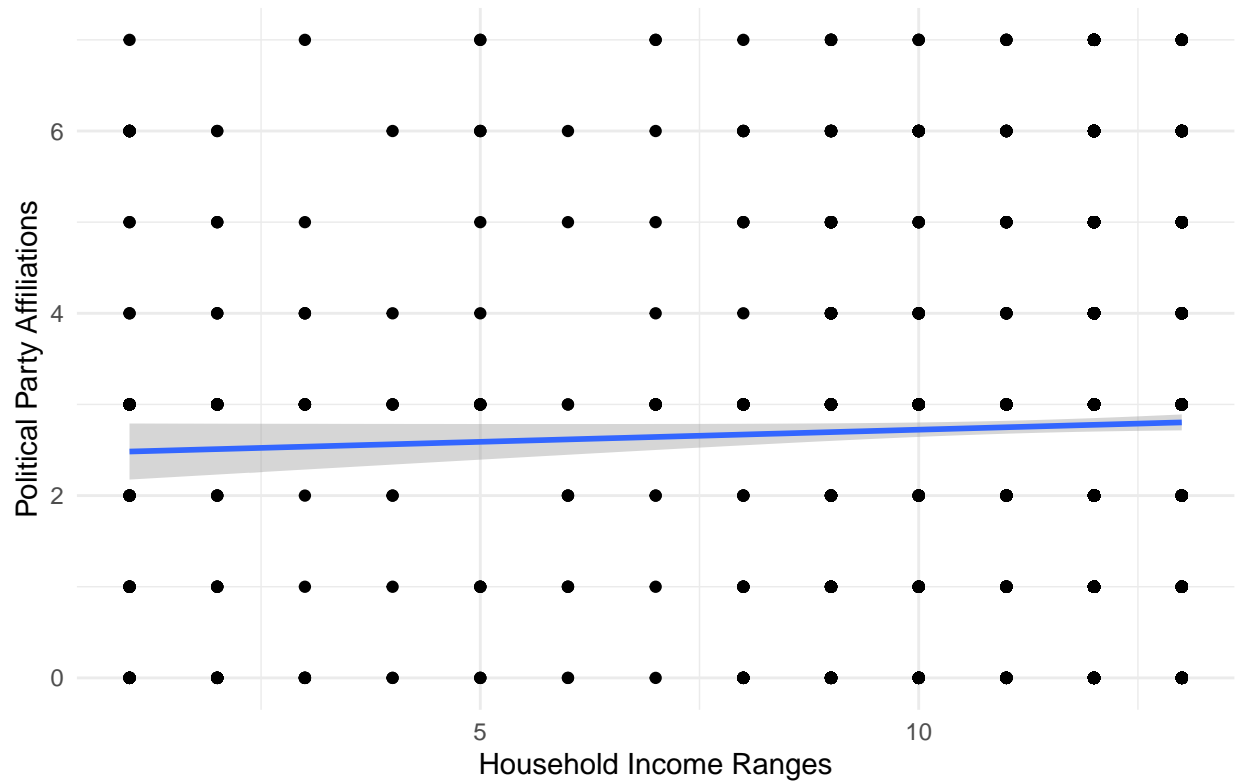## Levels of Predesposition towards the Democratic and Republican Parties
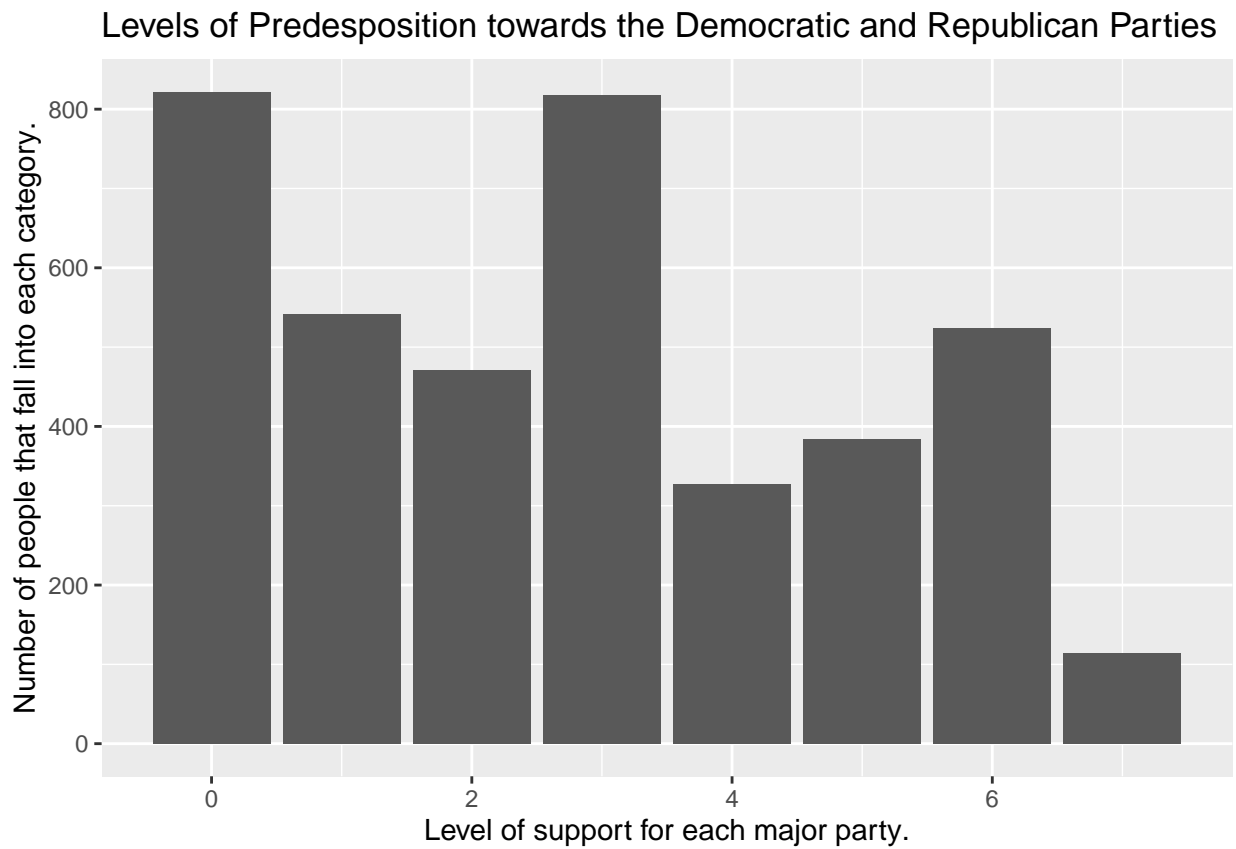
Since this recreation shows the cumulative number of respondents that identified with multiple levels of support for either parties in each category, the total number is much higher.

Can Simple Linear Models be created?

Due to dealing what categorical data, any attempt to create linear models would require assigning numeric values to categorical data. These values were defined in the Preparing the Data subsection in the Data section.

A Simple Linear Model between Income and Party Affliation
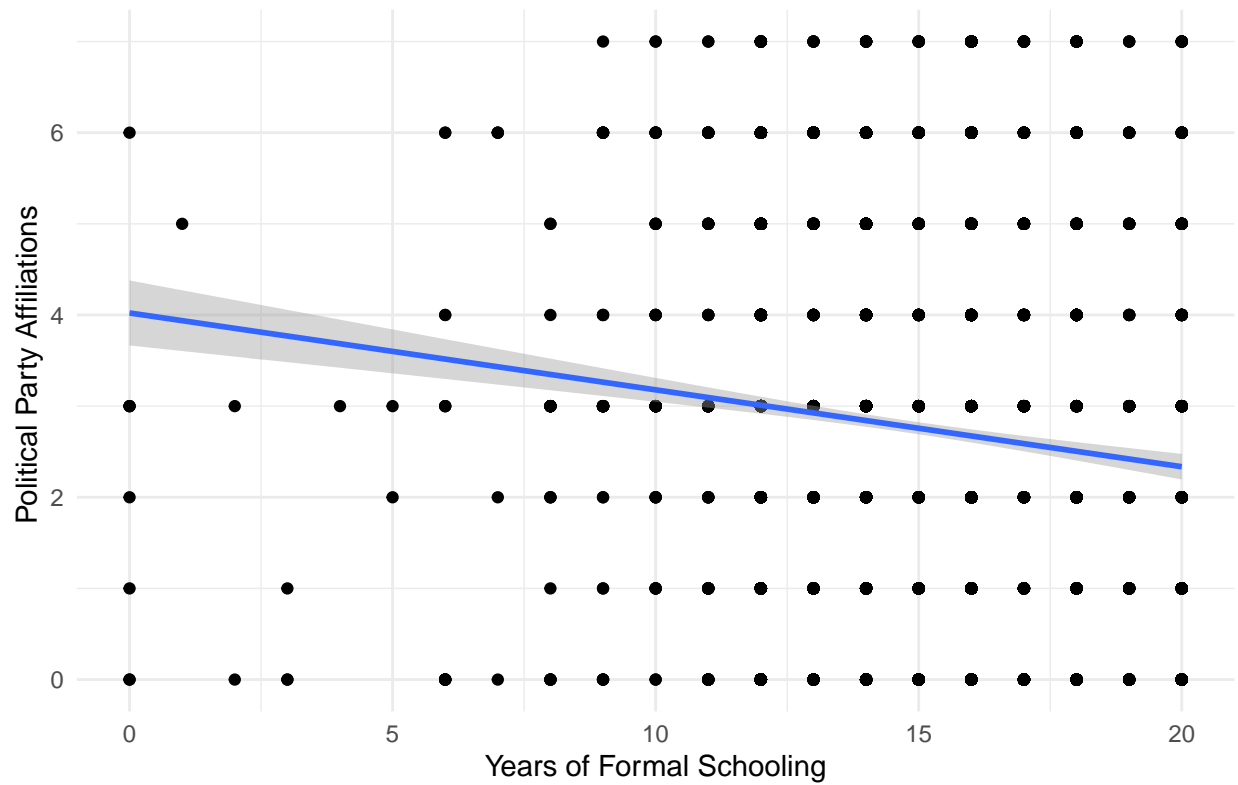
```
## 'geom_smooth()' using formula 'y ~ x'
```


3 is neutral.
The higher or lower the value if from 3, the stronger is the support for the Republicans or Democrats respectively.

A statistically insignificant relationship exists between income range and political party affiliation.

A Simple Linear Model between Level of Education and Political Party Affiliation

`geom_smooth()` using formula 'y ~ x'

3 is neutral.
The higher or lower the value if from 3, the stronger is the support for the Republicans or Democrats respectively.

A statistically insignificant relationship exists between level of education and political party affiliation. There is a very marginal correlation between more years of education and positive support for the Democrats.

A Simple Linear Model between Income Range and Political Views
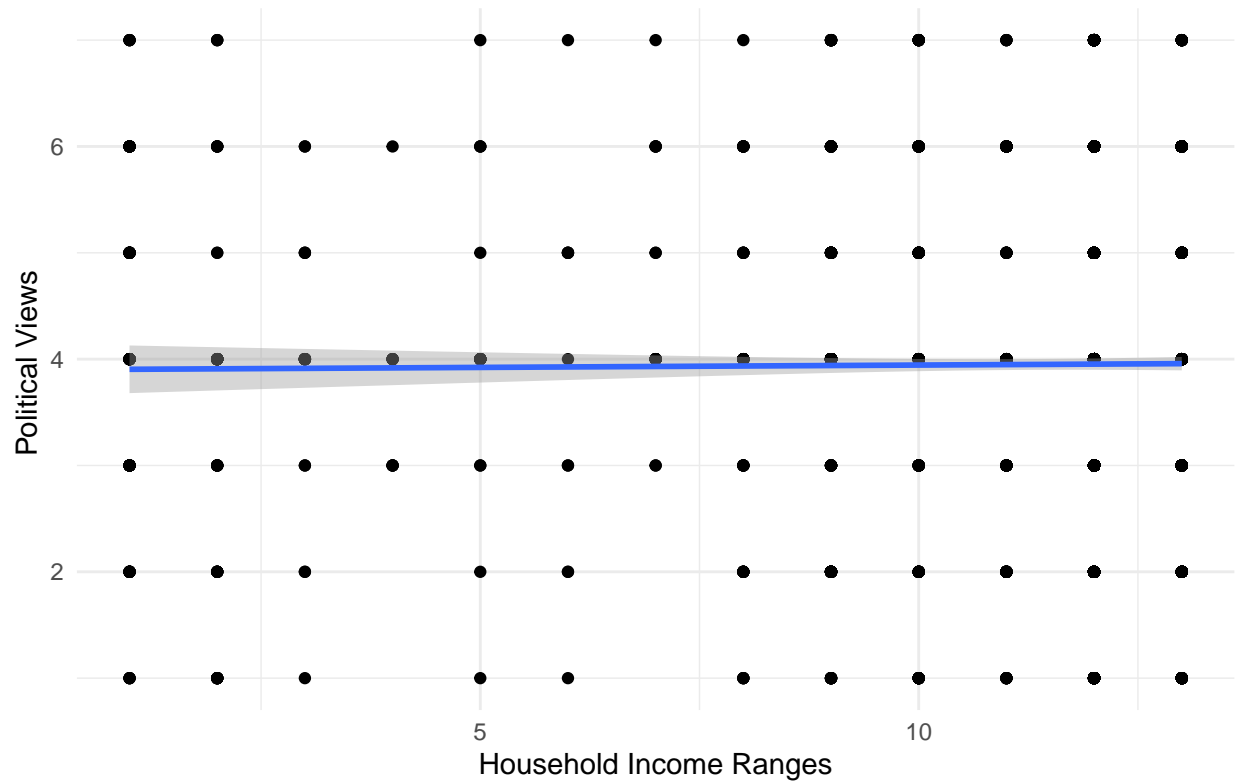
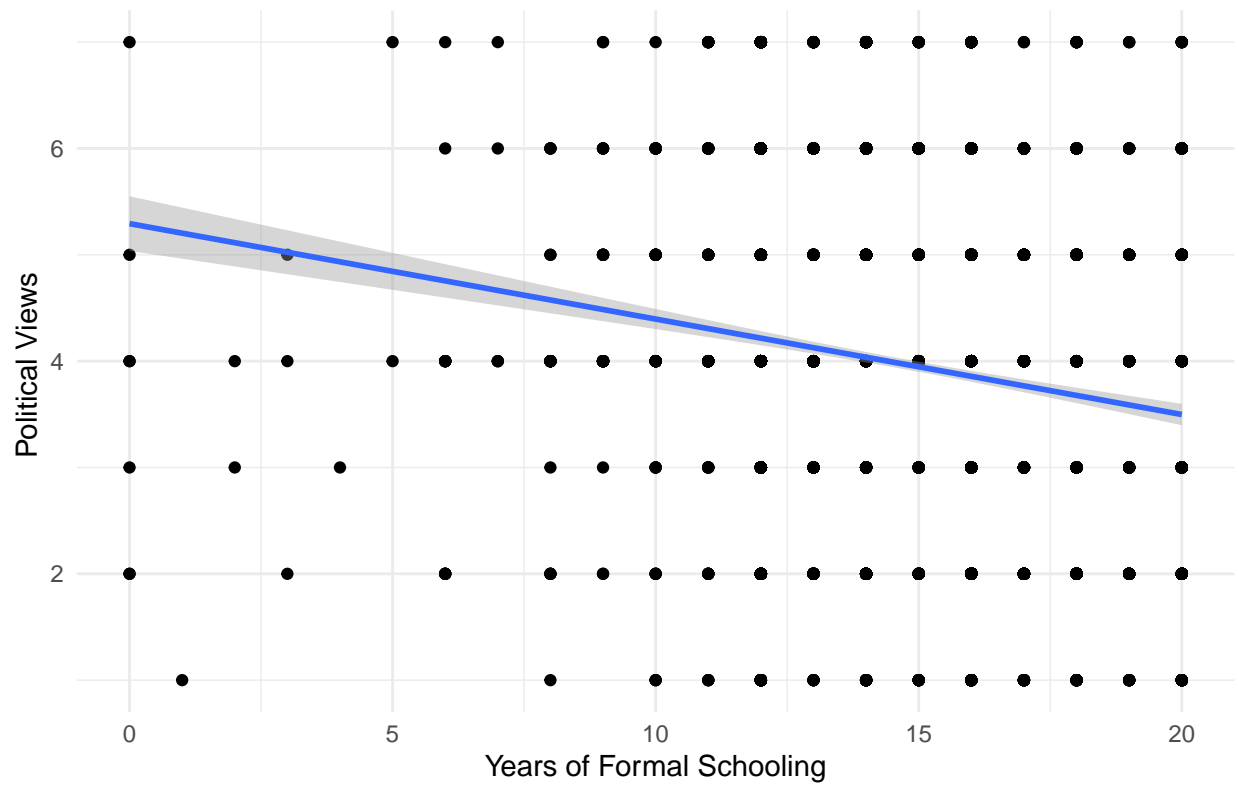## `geom_smooth()` using formula 'y ~ x'

4 is neutral.
The higher or lower the value if from 4, the stronger they identify as Liberal or Conservative respectively.

Income range and political views do not have a statistically significant relationship.

A Simple Linear Model between Years of Education and Political Views
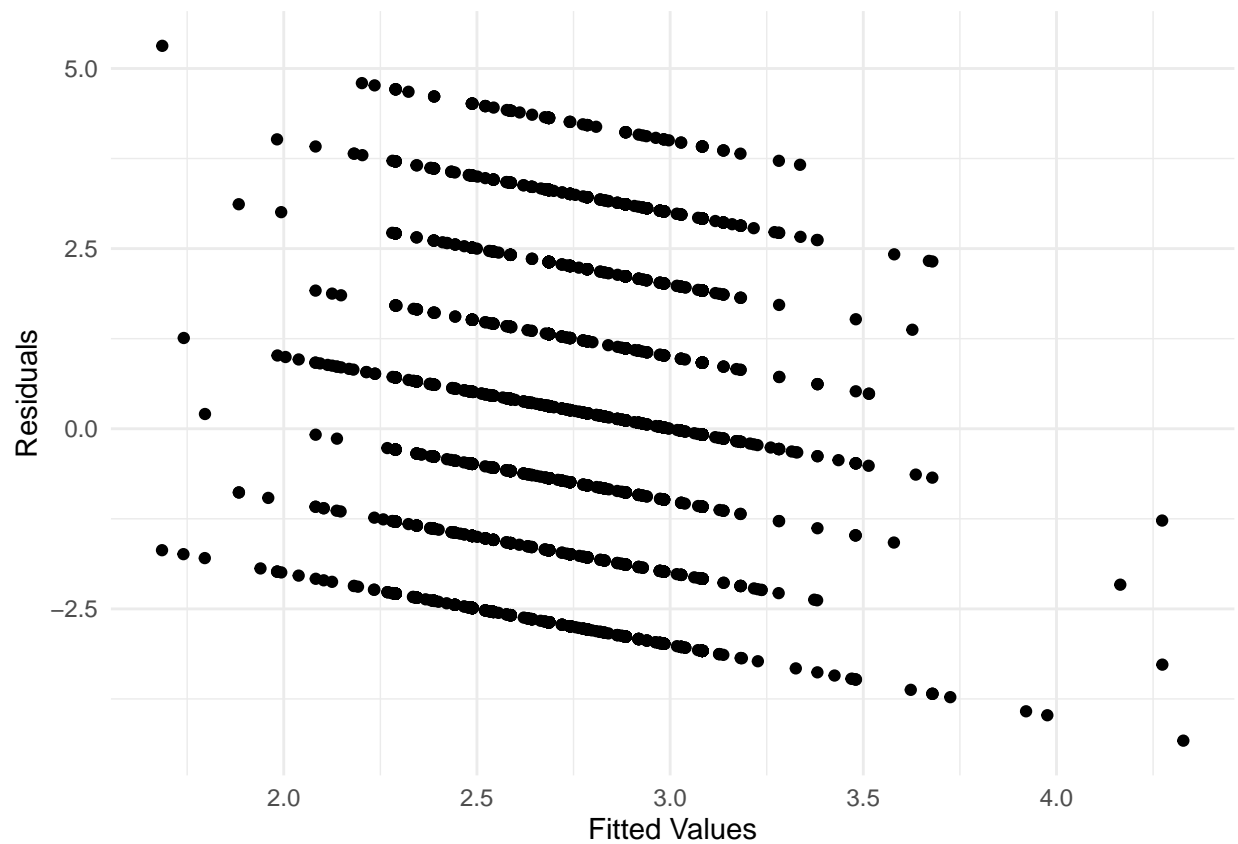
`geom_smooth()` using formula 'y ~ x'

4 is neutral.
The higher or lower the value if from 4, the stronger they identify as Liberal or Conservative respectively.

Years of formal education and political views do not have a statistically significant relationship.

Cumulative Impact of Socio-Economic Characteristics on Political Party Affiliations

As one can easily deduce from our previous findings, since neither years of formal schooling nor household income range had any statistically significant relationship with political party affiliations, they do not have such a relationship with political party affiliations cumulatively either.

Cumulative Impact of Socio-Economic Characteristics on Political Views
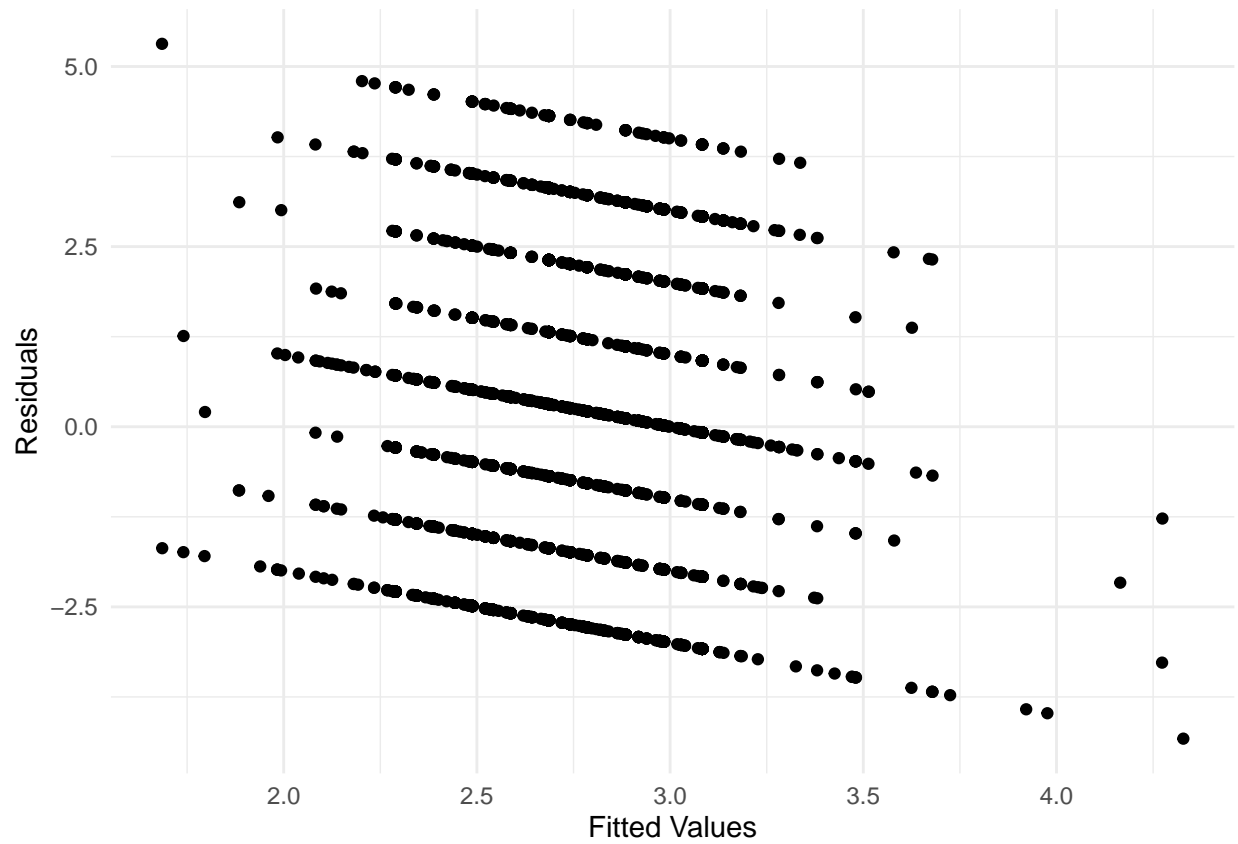
The cumulative relationship of years of formal schooling and household income ranges do not have a statistically significant relationship either.

Discussion

The General Social Survey 2021

Strengths

The primary strength of the General Social Survey 2021 is the sheer amount of variables used to give a very holistic perspective on American societal opinions and characteristics. This allows statisticians to observe if any trends exist, make predictions and test the different relationships that various characteristics have with various political opinions and leanings.

Shortcomings

Assigning numeric values to categorical data is very poorly done. There is a severe lack of consistency when a value is being determined to represent neutrality or strong opinions or affiliations towards one side. For instance, complete neutrality for political party affiliations is represent by an integer value of 3, but for political views by an integer value of 4. This makes it harder to perform cumulative computations. A vastly superior approach would be assigning a completely neutral stance a value of 0. This would have made the data very easy to analyze and perform complex computations on. Then, the distance from this value can be used to compute the numbers and intensity of deviations from it. Due to this, multiple linear regression could be only at a very elementary level, as many more levels of data preparation would be required.

Additional data should have also been gathered, such as whether repondents' political views and socioeconomic conditions have changed in the last five years, whether they have become more interested in politics in the last five years and what is their primary source of getting information about politics. This is all covered by the supplementary survey in the appendix.

Characteristics of an American Ideologue: Revisiting a Linear Relationship

What makes this paper unique?

The uniqueness of this paper can be attributed primarily to one important deduction made; how the lines between socio-economic characteristics and political party affiliations and political views in America have blurred significantly since academics and statisticians revisited this topic last. It paves the way for further analyses and discourse regarding what significant changes in both societal opinion, the media for political communication, party policies and the way political popularity is achieved has occurred very recently.

Shortcomings

While this paper demonstrated that the relationship between certain socio-economic characteristics and political views and party affiliations is now non-existent, it does not give any reasons for it. This is a topic that needs to be revisted in order to demonstrate causality. The supplementary survey (link in the Appendix) collects additional data to make better statistical inferences on what caused this change. Furthermore, due to important relevant data missing from the General Social Survey 2021 that the supplementary survey asks for, this paper does not go beyond making very rudimentary computations.

Bibliography

```
## Loading required namespace: bibtex


## [1] _GSS General Social Survey_. <URL: https://gss.norc.org/>.
##
## [2] D. P. Dupont and I. J. Bateman. "Political affiliation and
## willingness to pay: An examination of the nature of benefits and means
## of provision". In: _Ecological Economics_ 75 (2012), p. 43-51. DOI:
## 10.1016/j.ecolecon.2012.01.012.
##
## [3] D. P. Dupont and I. J. Bateman. "Political affiliation and
## willingness to pay: An examination of the nature of benefits and means
## of provision". In: _Ecological Economics_ 75 (2012), p. 43-51. DOI:
## 10.1016/j.ecolecon.2012.01.012.
##
## [4] R. Francois. _bibtex: Bibtex Parser_. R package version 0.4.2.3.
## 2020. <URL: https://github.com/romainfrancois/bibtex>.
##
## [5] K. Moon. _ggiraphExtra: Make Interactive ggplot2. Extension to
## ggplot2 and ggiraph_. R package version 0.3.0. 2020. <URL:
## https://github.com/cardiomoon/ggiraphExtra>.
##
## [6] R Core Team. _R: A Language and Environment for Statistical
## Computing_. R Foundation for Statistical Computing. Vienna, Austria,
## 2021. <URL: https://www.R-project.org/>.
##
## [7] H. Wickham, W. Chang, L. Henry, et al. _ggplot2: Create Elegant
## Data Visualisations Using the Grammar of Graphics_. R package version
## 3.3.5. 2021. <URL: https://CRAN.R-project.org/package=ggplot2>.
##
## [8] H. Wickham and E. Miller. _haven: Import and Export SPSS, Stata and
## SAS Files_. R package version 2.4.3. 2021. <URL:
## https://CRAN.R-project.org/package=haven>.
```

Appendix

Supplementary Survey

Link: "https://kq8fy5oqcd5.typeform.com/to/YCtp0F0N"

Git Repository

Github: "https://github.com/yfuoft/p3.git"

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.