Characteristics of an American Ideologue: 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.

Introduction

The R language (R Core Team 2021) is used, in addition to a variety of packages and scholarly sources. Please refer to the References 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. As statisticians, trends are 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.

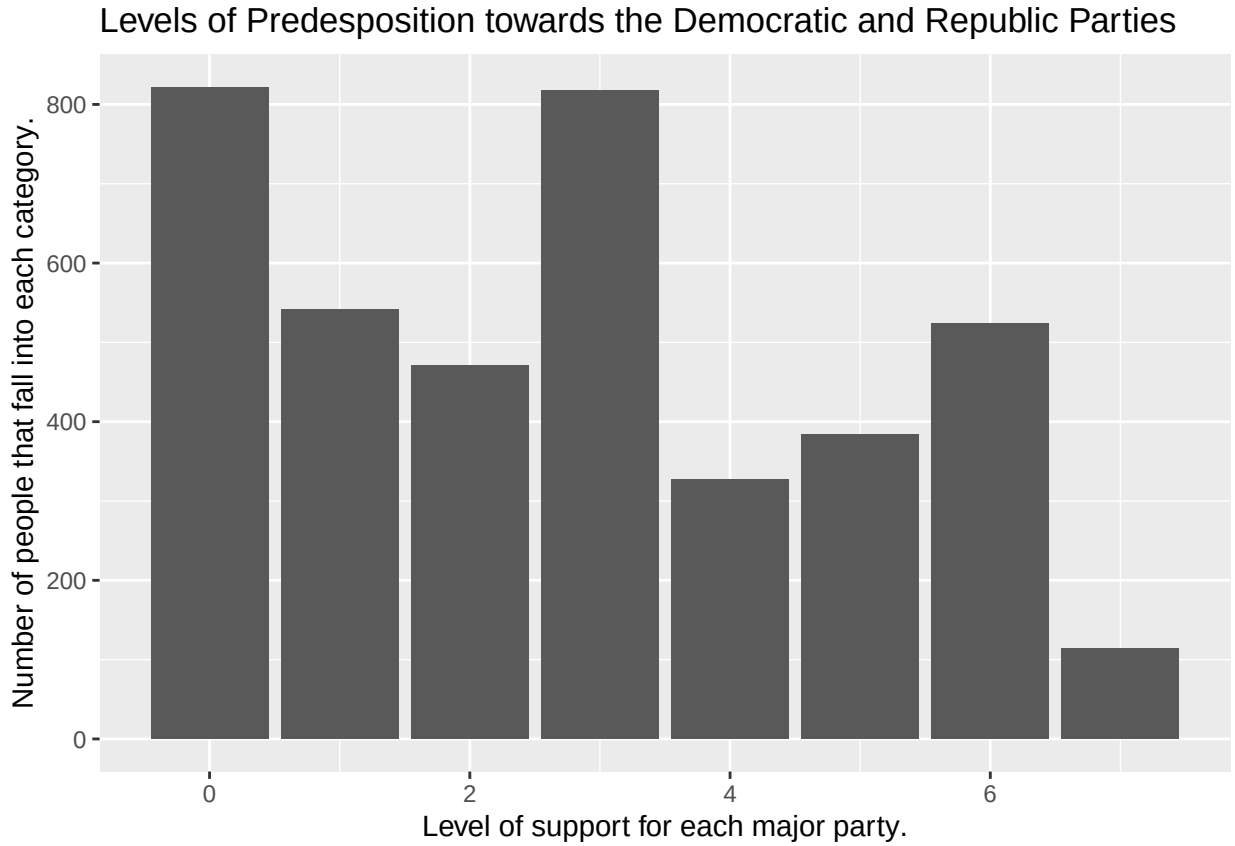
Education level ranges in years from 0, representing no formal education to 17, which represents five years of college education or higher.

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.

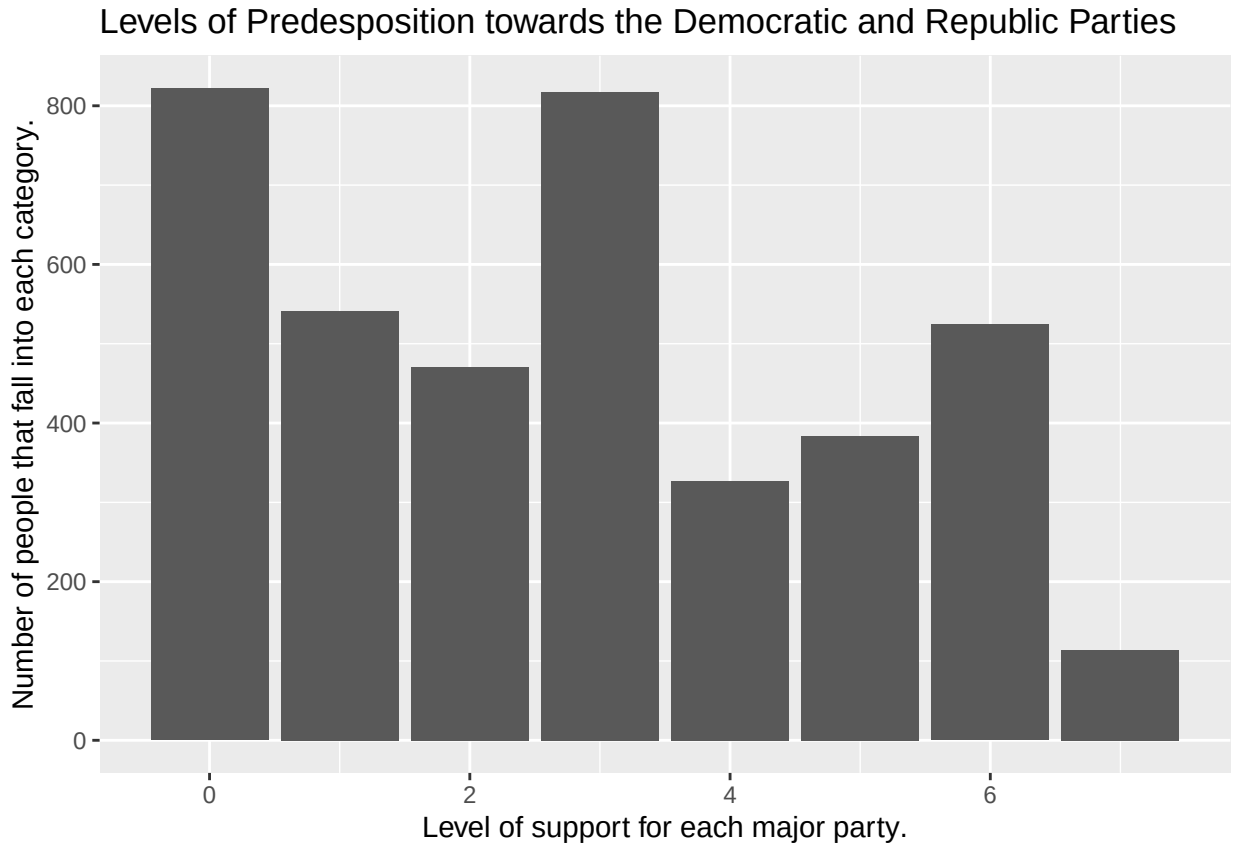
There were a total of 146 null values for the column which represented an individuals party affiliations. These rows were omitted.

Results

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



To make this simpler for analysis, lets use the categories described in the previous subsection to replot this.



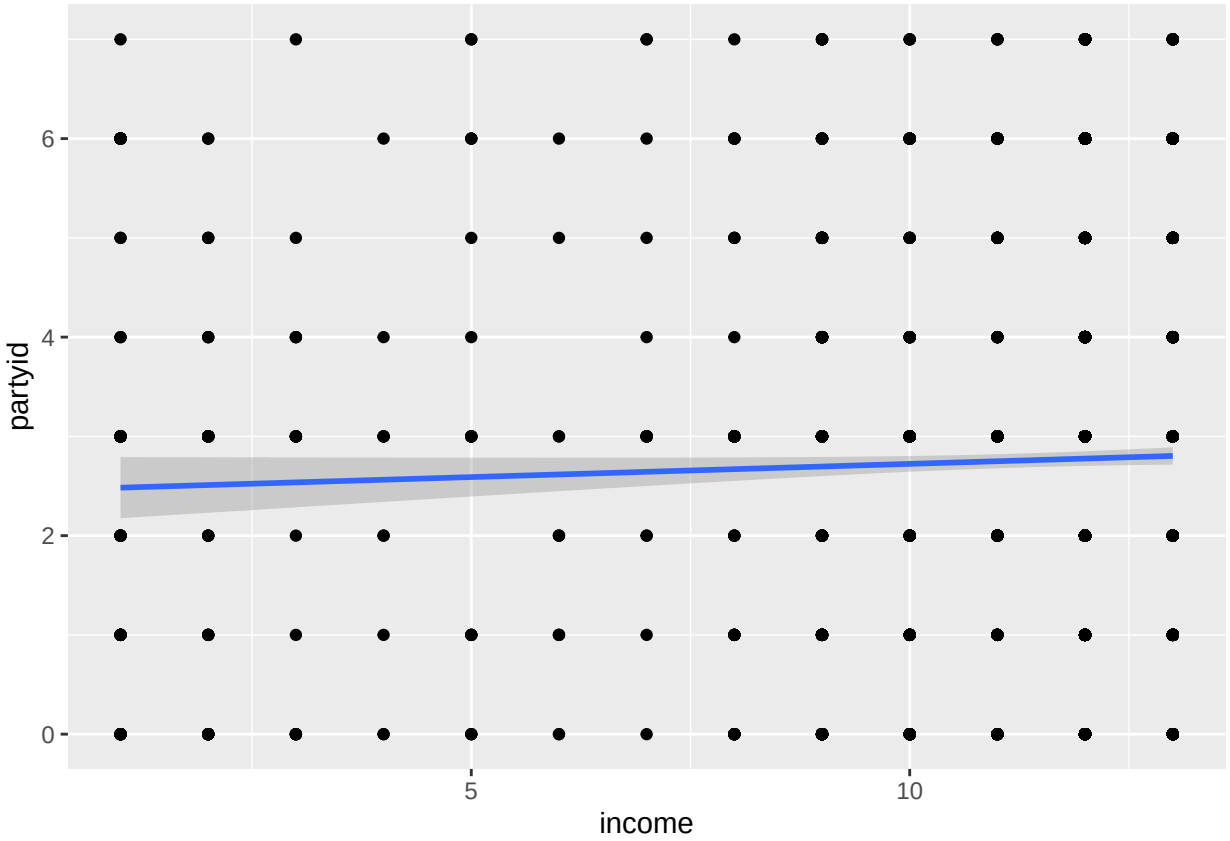
Since this recreation shows the cumulative number of respondents that identified with multiple levels of support for either parties in each category, the total number is much higher.

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.

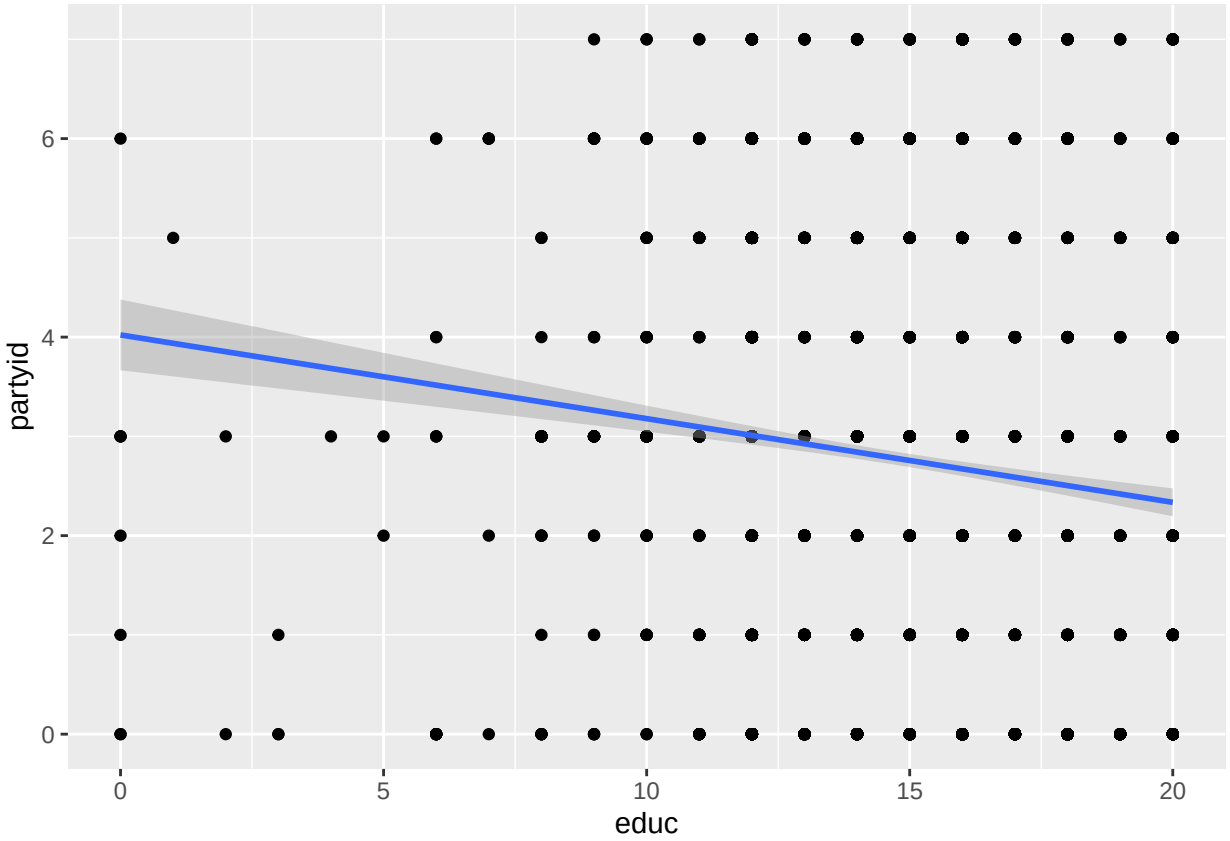
A Simple Linear Model between Income and Party Affliation

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



A statistically insignificant relationship exists between income range and political party affiliation. ### A Simple Linear Model between Level of Education and Political Party Affiliation

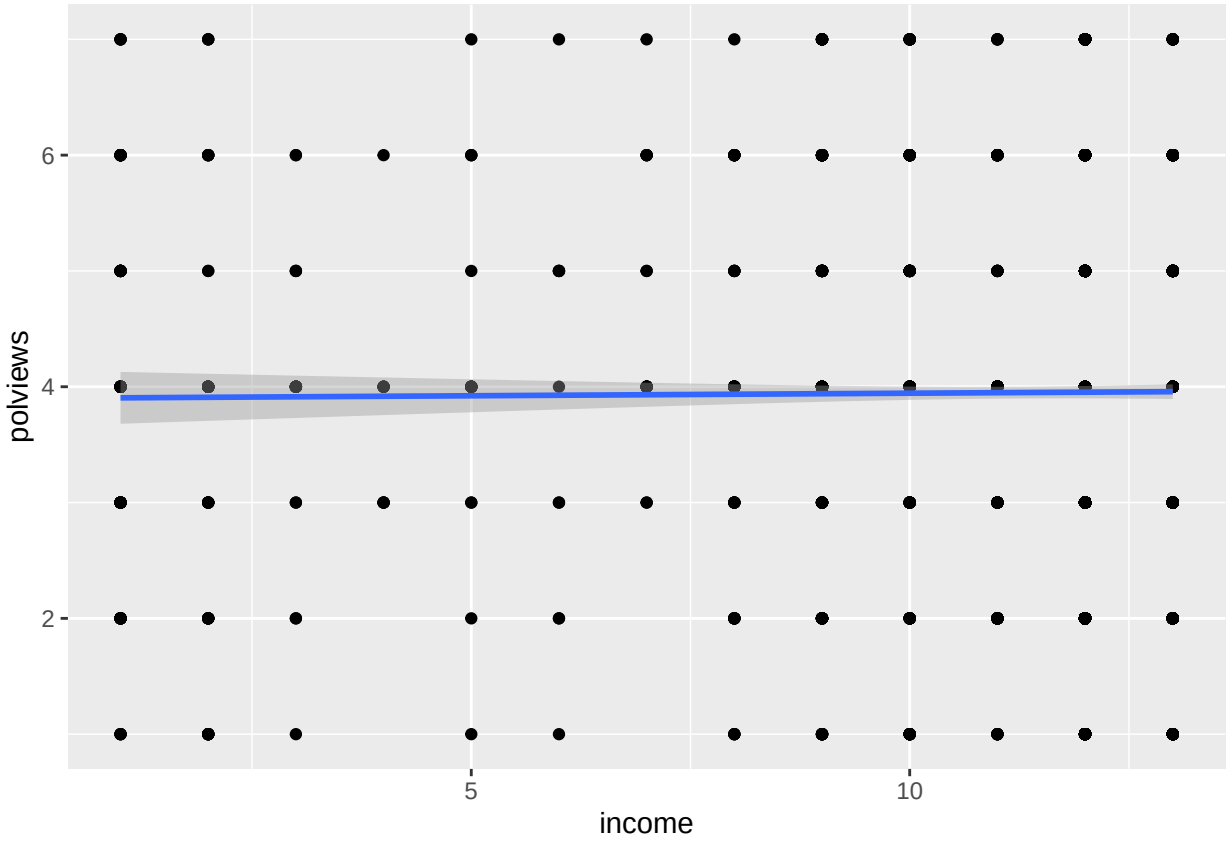
```
## `geom_smooth()` using formula 'y ~ x'
```



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

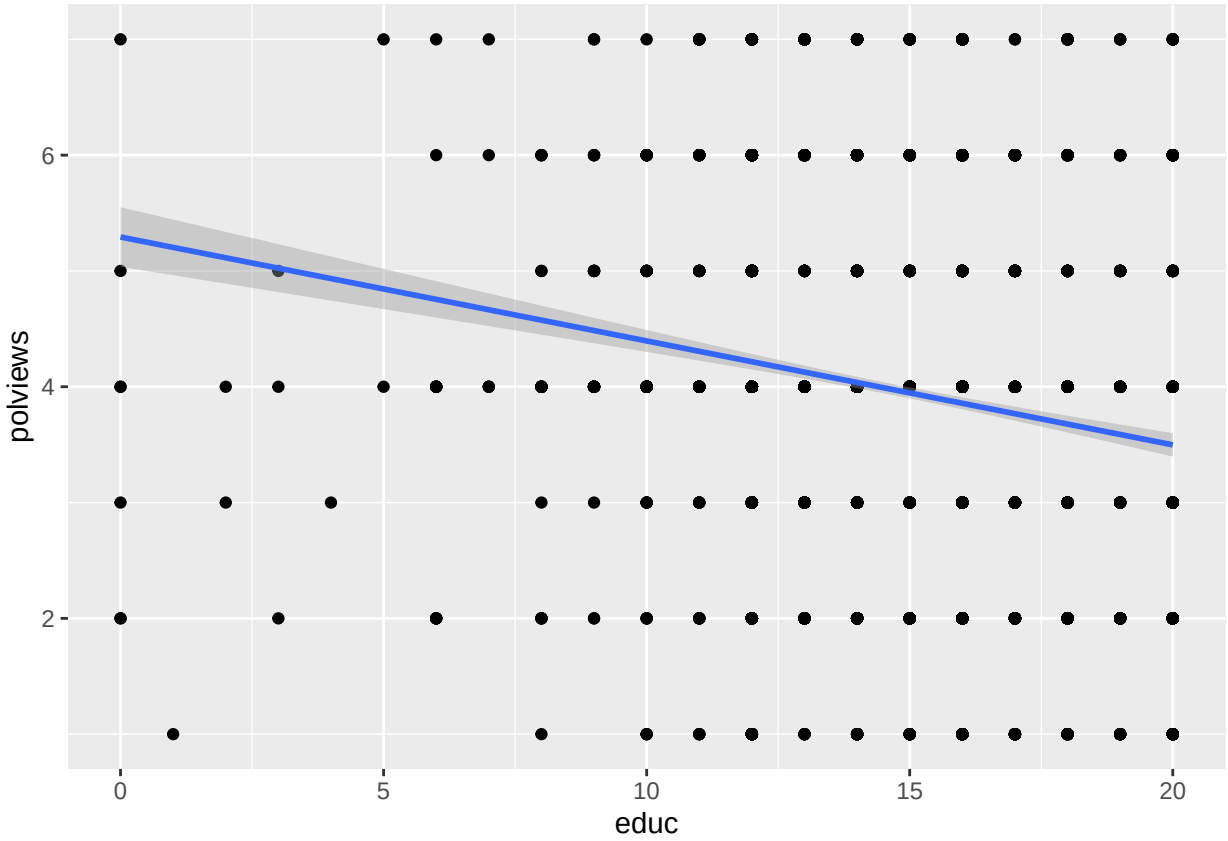
```
## `geom_smooth()` using formula 'y ~ x'
```



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'
```



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

References

```
## Loading required namespace: bibtex

## [1] 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.
##
## [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] 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/>.
```

Links

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/.