Trifecta of personality (Temp name)

Cody's Data Science Portfolio

Abstract

In under 250 words give a brief summary of the project so far.

1. Discovery and Data Preparation

Provide a one-sentence introduction with what is covered by the first deliverable along with a link to another html page that gives the details.

My motivation for this topic is I'm interested if there is a link between education, economic standing, and political standing.

My source for my initial data is Pew Research Center, (https://www.people-press.org/dataset/2015-governance-survey/), which conducted a phone survey of more that 6,000 adults and also draws on data from other surveys they've done back to 1994. The survey size was small and could've been better but they are also drawing inferences from previous surveys to better extrapolate from the data.

```
suppressMessages(library("tidyverse"))
suppressMessages(library("haven"))
# Add or change the strings to make a path to the file that you downloaded
path = file.path("Governance 2015.public.sav")
dataset = read_sav(path)


nrow(dataset)
```

```
## [1] 6004
```

```
#colnames(dataset)
```

The variables of the data set are mostly understandable and comes with a readme.txt and a topline.pdf that you can further look at when going to the link.

I will make a new tibble snippet to mess around with and take certain variables and see where that takes me

```
snippet <- dplyr::filter(dataset, sex == 1 & q44 == 2)
snippet <- dplyr::select(snippet, sex, q44, everything())

snippet
```

```
## # A tibble: 137 x 196
##        sex     q44 respid  sample attempt  fcall refusal int_date   ilang
##    <dbl+l> <dbl+l>  <dbl> <dbl+l>   <dbl>  <dbl> <dbl+l>    <dbl> <dbl+l>
##  1 1 [Mal~ 2 [A w~      2 1 [Lan~       7 150827 1 [Yes]   150903 1 [Eng~
```

```
##  2 1 [Mal~ 2 [A w~      8 1 [Lan~       1 150827 0 [No]    150827 1 [Eng~
##  3 1 [Mal~ 2 [A w~    321 1 [Lan~       5 150829 0 [No]    150829 1 [Eng~
##  4 1 [Mal~ 2 [A w~     68 1 [Lan~       1 150827 0 [No]    150827 1 [Eng~
##  5 1 [Mal~ 2 [A w~    171 1 [Lan~       1 150828 0 [No]    150828 1 [Eng~
##  6 1 [Mal~ 2 [A w~    247 1 [Lan~       2 150828 0 [No]    150828 1 [Eng~
##  7 1 [Mal~ 2 [A w~    145 1 [Lan~       1 150828 0 [No]    150828 1 [Eng~
##  8 1 [Mal~ 2 [A w~    399 1 [Lan~       3 150830 0 [No]    150830 1 [Eng~
##  9 1 [Mal~ 2 [A w~     26 1 [Lan~       1 150827 0 [No]    150827 1 [Eng~
## 10 1 [Mal~ 2 [A w~    102 1 [Lan~       2 150828 0 [No]    150828 1 [Eng~
## # ... with 127 more rows, and 187 more variables: cregion <dbl+lbl>,
## #   state <dbl+lbl>, density <dbl+lbl>, sstate <dbl>, susr <chr>, usr <chr>,
## #   scregion <dbl+lbl>, sdensity <dbl+lbl>, llitext0 <dbl>, llitext <dbl+lbl>,
## #   qintro2 <dbl+lbl>, phase <dbl>, formab <dbl+lbl>, form <dbl+lbl>,
## #   qs1 <dbl+lbl>, q1f1 <dbl+lbl>, q2f2a <dbl+lbl>, q3f2b <dbl+lbl>,
## #   qb4af1 <dbl+lbl>, qb4bf1 <dbl+lbl>, qb4cf1 <dbl+lbl>, qb4df1 <dbl+lbl>,
## #   qb4ef1 <dbl+lbl>, qb4ff1 <dbl+lbl>, qb4gf1 <dbl+lbl>, qb4hf2 <dbl+lbl>,
## #   qb4if2 <dbl+lbl>, qb4kf2 <dbl+lbl>, qb4lf2 <dbl+lbl>, qb4mf2 <dbl+lbl>,
## #   qb4nf2 <dbl+lbl>, q10 <dbl+lbl>, q11 <dbl+lbl>, q12 <dbl+lbl>,
## #   q14 <dbl+lbl>, q15 <dbl+lbl>, q23 <dbl+lbl>, q24f1 <dbl+lbl>,
## #   q25f2 <dbl+lbl>, qb27 <dbl+lbl>, qb28 <dbl+lbl>, q33a <dbl+lbl>,
## #   q33b <dbl+lbl>, qb34a <dbl+lbl>, qb34b <dbl+lbl>, qb34c <dbl+lbl>,
## #   qb34d <dbl+lbl>, qb34e <dbl+lbl>, qb34gf1 <dbl+lbl>, qb34hf1 <dbl+lbl>,
## #   qb34if2 <dbl+lbl>, qb34jf2 <dbl+lbl>, qb35 <dbl+lbl>, qb36 <dbl+lbl>,
## #   qb37 <dbl+lbl>, qb37os <chr>, qb38 <dbl+lbl>, qb39 <dbl+lbl>,
## #   q42a <dbl+lbl>, q42b <dbl+lbl>, q42c <dbl+lbl>, q42d <dbl+lbl>,
## #   q42f <dbl+lbl>, q42g <dbl+lbl>, q42h <dbl+lbl>, q42i <dbl+lbl>,
## #   q42j <dbl+lbl>, q42l <dbl+lbl>, q42m <dbl+lbl>, q43 <dbl+lbl>,
## #   qb45 <dbl+lbl>, qb46 <dbl+lbl>, int1 <dbl+lbl>, int2 <dbl+lbl>,
## #   int3m <dbl+lbl>, oftvote <dbl+lbl>, qb50a <dbl+lbl>, qb50b <dbl+lbl>,
## #   qb50e <dbl+lbl>, qb51a <dbl+lbl>, qb51b <dbl+lbl>, qb51e <dbl+lbl>,
## #   q62 <dbl+lbl>, q70a <dbl+lbl>, q70b <dbl+lbl>, q70d <dbl+lbl>,
## #   q70f <dbl+lbl>, q70i <dbl+lbl>, q70j <dbl+lbl>, q70kf1 <dbl+lbl>,
## #   q70lf1 <dbl+lbl>, q70mf1 <dbl+lbl>, q70nf2 <dbl+lbl>, q70of2 <dbl+lbl>,
## #   q70pf2 <dbl+lbl>, q70qf2 <dbl+lbl>, q71a <dbl+lbl>, q71b <dbl+lbl>,
## #   q71d <dbl+lbl>, q71f <dbl+lbl>, ...
```

```
nrow(snippet)
```

```
## [1] 137
```

So this snippet I've made is just showing some interesting insights you can get from this data. I chose a random question which is 44 and made it so I can only see males that chose the answer 'a worse place to live' If the filter() function was working like the dplyr.tidyverse.org reference said it works then I could give you the number of occurences of those 2 things happening at the same time.

```
Q.44   On balance, do you think having an increasing number of people of many different
races, ethnic groups and nationalities in the United States makes this country a better
place to live, **a worse place to live**, or doesn't make much difference either way?
```