Trifecta of Personality (Part 1)

Cody's Data Science Portfolio

Intro

1. Discovery and Data Preparation

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)


ndata <- nrow(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)
nmale <- nrow(snippet)
snippet <- dplyr::filter(snippet, q44 == 2)
snippet <- dplyr::select(snippet, sex, q44, everything())

nsnip <- nrow(snippet)

maleprop <- nmale/ndata
thisgroupprop <- nsnip/ndata
thisgroupofmalesprop <- nsnip/nmale
```

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'.

```
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?
```

This is the percentages of males in the total dataset

```
maleprop
```

```
## [1] 0.526982
```

This is the percentage of the males that answered B for question 44 out of the entire dataset

```
thisgroupprop
```

```
## [1] 0.02281812
```

This is the percentage of males that answered B for question 44 out of the number of males in the total dataset

```
thisgroupofmalesprop
```

```
## [1] 0.04329962
```