Final Project for WVS Dataset

Emma Jackson

2022-05-04

# Introduction

The *WVS* dataset contains information collected from the World Values Survey from 1995-1997 to ascertain people’s view on the government’s handling of poverty in their country. The dataset contains 6 variables and 5381 observations.

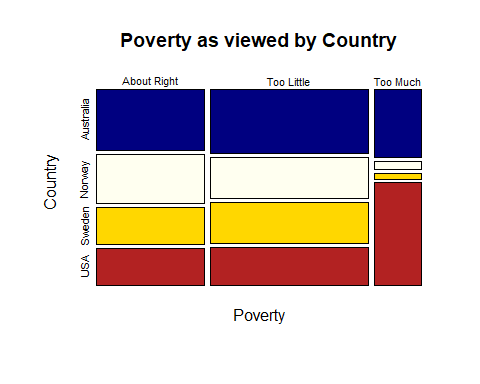
The variable of interest is poverty and has three levels and stores the user’s answer to the question, “Do you think that what the government is doing for people in poverty in this country is about the right amount, too much, or too little?”

The other variables used in this report are more information on the user, religion, degree, country, age, and gender. These variables will be examined in our analysis section to see if there is a relationship with poverty. Below is a summary of our dataset:

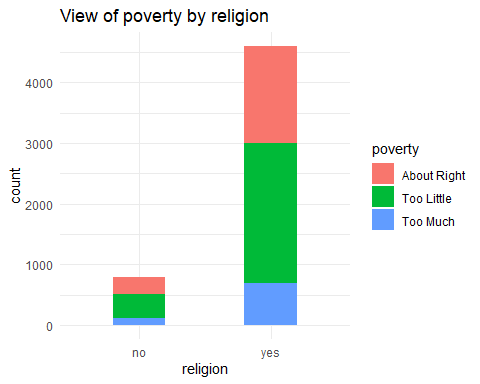
## poverty religion degree country age   
## About Right:1862 no : 786 no :4238 Australia:1874 Min. :18.00   
## Too Little :2708 yes:4595 yes:1143 Norway :1127 1st Qu.:31.00   
## Too Much : 811 Sweden :1003 Median :43.00   
## USA :1377 Mean :45.04   
## 3rd Qu.:58.00   
## Max. :92.00   
## gender ageCategories   
## female:2725 Baby Boomers:1428   
## male :2656 Gen X :1623   
## Gen Z : 727   
## Millennial :1603

# Visualizations

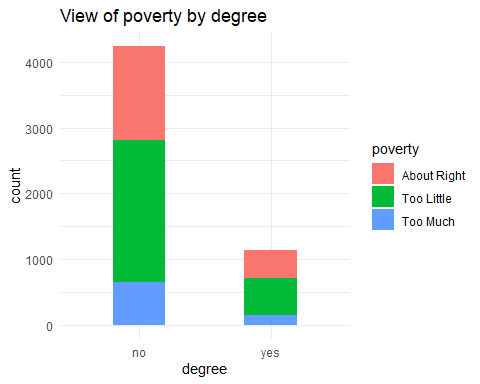
To determine the strongest relationship between poverty and the other variables, we first examine them visually to see if it is worth analyzing. This first graph depicts the relation between people’s view on if their government is doing enough for poverty in their country. From this visual, we see the greatest disparity between countries in the “Too Much” column.



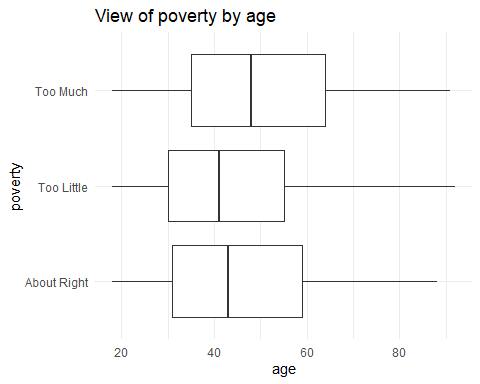
The following graph examines the relationship of religion and views on the government’s actions on poverty. This clearly shows the majority of people surveyed are religious.



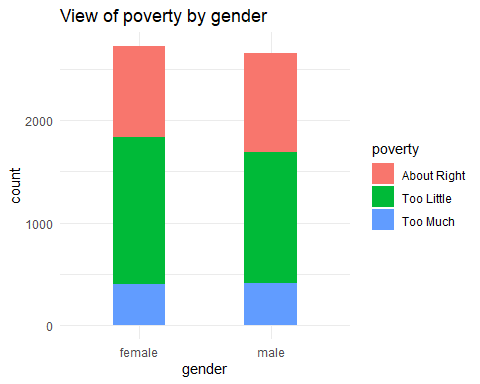
This graph shows the proportion of those who were surveyed that hold a degree and their views on the government’s involvement with poverty. Significantly more people do not hold degrees.



Using our only numerical variable, age, we create a boxplot to view age against poverty and see if there is any relation or observations we can spot. Examining this graph, we see that people who selected, “Too Much” tended to have a older mean age while “Too Little” tended to have a younger mean age.



Lastly, this graph looks at males vs. females that were surveyed. It appears that slightly more females were surveyed and it also looks like females may have been more inclined to view the government as doing “Too little” to handle poverty. This will be especially interesting to test in our analysis section next.



# Analysis

After visualizing several variables against our variable of interest, poverty, we will now transition into our analysis. Given we were unable to clearly see relationships in our graphs,this section is especially important. With this dataset, it is mostly made up of categorical variables, making our options for analysis limited.

To return to our original function for this report, we are trying to decide if our variables are independent from poverty. To determine this, we will use a Chi-Squared Test. This test will be repeated with several of our variables.

## Analysis of poverty and country variables

## Australia Norway Sweden USA  
## About Right 622 494 374 372  
## Too Little 952 597 604 555  
## Too Much 300 36 25 450

## Pearson's Chi-squared test  
##   
## data: table1  
## X-squared = 597.62, df = 6, p-value < 2.2e-16

This result shows a very low p-value, significantly less than 0.05. This means we reject that the two variables of country and poverty are independent. You can see from our table that very few people from Norway and Sweden answered, “Too Much.”

## Analysis of poverty and religion variables

##   
## no yes  
## About Right 272 1590  
## Too Little 393 2315  
## Too Much 121 690

##   
## Pearson's Chi-squared test  
##   
## data: table2  
## X-squared = 0.083005, df = 2, p-value = 0.9593

After doing a test for the independence of religion and poverty, with a large p-value of .9593 we can conclude they do not have a significant relationship.

## Analysis of poverty and degree variables

##   
## no yes  
## About Right 1426 436  
## Too Little 2155 553  
## Too Much 657 154

##   
## Pearson's Chi-squared test  
##   
## data: table3  
## X-squared = 8.8109, df = 2, p-value = 0.01221

The analysis of poverty and degree results in a p-value of .01221 and we reject that they are independent of one another.

## Analysis of poverty and gender variables

##   
## female male  
## About Right 892 970  
## Too Little 1435 1273  
## Too Much 398 413

##   
## Pearson's Chi-squared test  
##   
## data: table4  
## X-squared = 12.353, df = 2, p-value = 0.002077

The analysis of poverty and gender results in a p-value of .002077 and we reject that they are independent of one another.

## Analysis of poverty and age variables

##   
## Pearson's Chi-squared test  
##   
## data: table5  
## X-squared = 86.474, df = 6, p-value < 2.2e-16

To conduct a Chi-Squared test, we convert the numeric values in the age variable into four categories, Gen Z (age 18-25), Millennial (age 26-40), Gen X (age 40-56), and Baby Boomer (age 56 and above). The analysis of poverty and our age categories results in a very low p-value and we reject that they are independent of one another.

# Conclusion

The results of our analysis reveal that the variables country, degree, gender, and age may have a relation to our variable of interest, poverty, with country and age having the most significant effect on it.

Further research would need to be done into the dataset to determine the exact effect country and age have on poverty. The next recommended step would be to gather more numerical data on the topic to analyze if country and age could be used to predict an individual’s view on if their government is doing enough to handle the issue of poverty in their country.