

# Are the older American population against the growing trade agreement with Germany?

*Matt Moocarme*

*May 31, 2016*

The goal is to understand if how the opinion of Americans varies with age on the recent trade agreement with Germany. My hypothesis is that older Americans are less favourable to the recent trade agreements.

The data comes from the Pew research center, which can be found at '<http://www.pewglobal.org/category/datasets/?download=31944>'

In R, first I load in the data to the workspace:

```
library(foreign)
trade <- read.spss('data/Pew_trade.sav',
                  to.data.frame = T)

attr <- attributes(trade)
labels <- attr$variable.labels
```

If we look at the first question we can see that it is the most general and should be sufficient to test our hypothesis.

```
labels[labels = 'Q1']
```

```
##
```

```
## "Q1 What do you think about growing trade between [GERMANY: Germany / U.S.: the U.S.] and other coun
```

## Exploratory data plotting

Now we have the data we are going to tidy it up so that we are not working with any extraneous data. Separating the data via question, country and age. Since age is a continuous variable I am going to separate it into five separate categories.

```
library(dplyr)
library(tidyr)
library(ggplot2)
# Choose selected columns and rows
rel.data <- trade %>%
  filter(country == 'United States') %>%
  select(age, Q1)

# Add age bracket
trade.agecat <- rel.data %>%
  mutate(age.bracket = factor(ifelse(age<25, '18-24', ifelse(age < 40, '25-39', ifelse(age < 60, '40-59', '50-59'))))

# tidy data
trade.byage <- trade.agecat %>%
  gather(Question, Answer,-age,-age.bracket) %>%
```

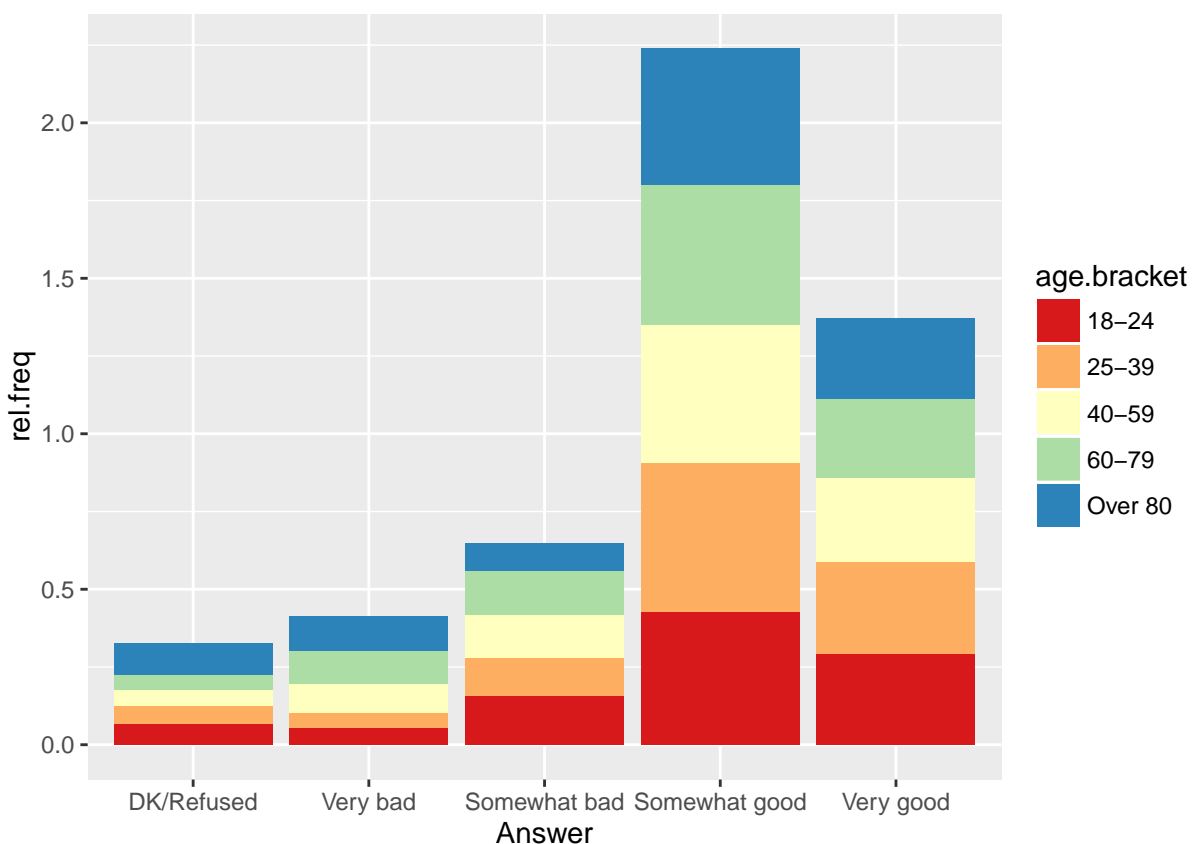
```
group_by(age.bracket, Question, Answer ) %>%
summarise(Total.counts = n()) %>%
mutate(rel.freq = Total.counts/sum(Total.counts))
```

Let me just organise the levels so they make a bit more visual sense.

```
trade.byage$Answer <- factor(trade.byage$Answer, levels = c( 'DK/Refused', 'Very bad', 'Somewhat bad',
```

Now we can plot the data with respect to the question answer, broken down by age bracket.

```
age.plot = ggplot(data = na.omit(trade.byage), aes(x = Answer, y = rel.freq, fill = age.bracket)) + geom_bar()
age.plot + scale_fill_brewer(palette = 'Spectral')
```



We can see that in general most americans seem to be in favour of the trade agreement with Germany

```
fav.trade <- rel.data %>%
group_by(Q1) %>%
summarise(total.counts = n()) %>%
mutate(percent = 100*total.counts/sum(total.counts)) %>%
filter(Q1 == 'Very good' | Q1 == 'Somewhat good')

sum(fav.trade[, 'percent'])
```

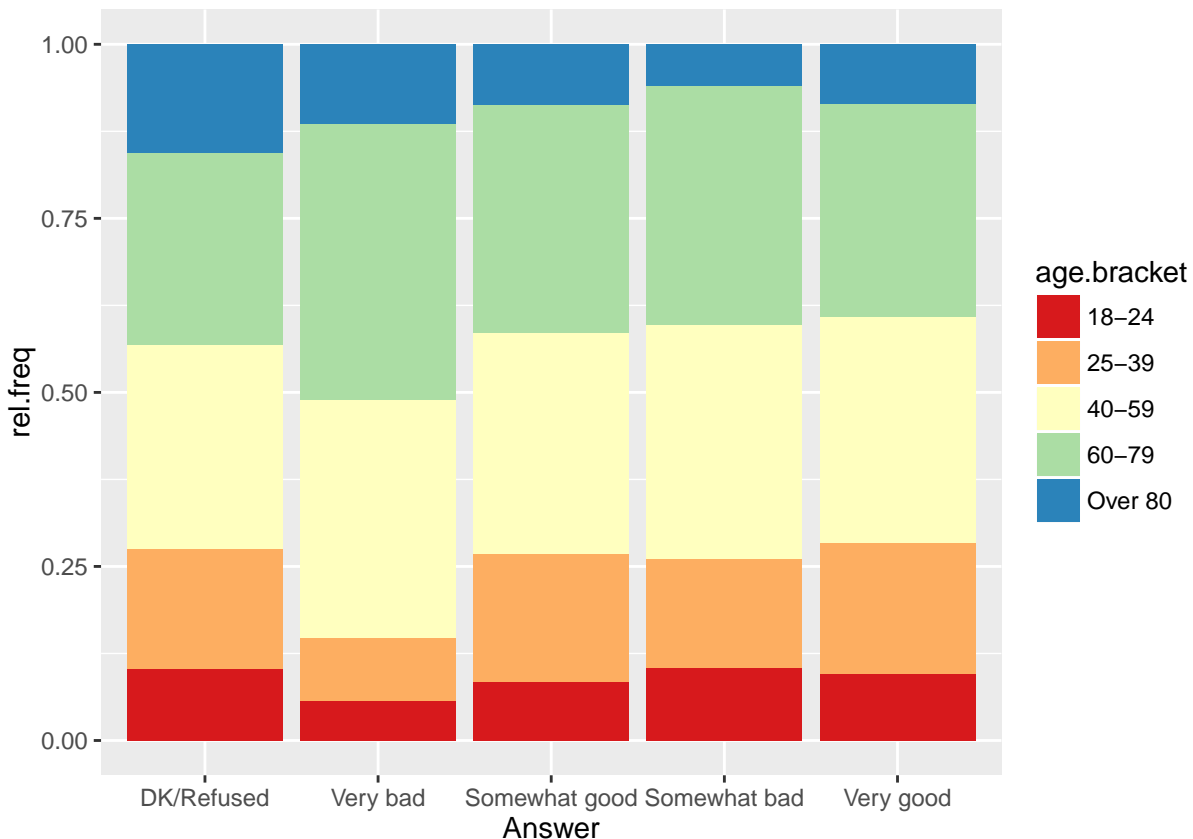
```
## [1] 72.05589
```

In fact we can see that almost three quarters of americans are in favour of the agreement.

If we zoom in on the question answers we can get a break down of ages, such that all the question answers will total 1.0, and if the height of the ‘Very bad’ bar plot for the colour representing the age bracket ‘18-24’ has height equal to 0.12, then 12% of the people that answered ‘Very bad’, were bewteen the ages of 18-24.

```
trade.byage <- trade.agecat %>%
  gather(Question, Answer,-age,-age.bracket) %>% group_by( Question, Answer, age.bracket ) %>% summaris
trade.byage$Answer <- factor(trade.byage$Answer, levels = c( 'DK/Refused', 'Very bad', 'Somewhat good',

age.plot = ggplot(data = na.omit(trade.byage), aes(x = Answer, y = rel.freq, fill = age.bracket)) + geom
age.plot + scale_fill_brewer(palette = 'Spectral')
```



Here we can see that there *may* be a slight correlation between age and survey answer in that as the answer varies from ‘Very bad’ to ‘Very good’ the share for the age brackets corresponding to those over 60 becomes less and less. If age were truly independent then we would expect an equal share for all question answers.

### Statistical hypothesis testing

Since it is hard to determine visually whether there is any dependence on the age with the question answer we can use statistical hypothesis testing in the form of a  $\chi^2$  test.

First we have to organise the data in the correct form. We will do it with the rows representing the question answers, and the columns representing the age bracket. Each cell will show the total frequency that that instance occurred.

```
trade.chisq.tb <- trade.agecat %>%
  mutate(Age18to24 = age.bracket == '18-24', Age25to39 = age.bracket == '25-39', Age40to59 = age.bracket == '40-59', Age60to79 = age.bracket == '60-79', Age80plus = age.bracket == '80plus')
  select(-age, -age.bracket) %>%
  group_by(Q1) %>%
  summarise_each(funs(sum))
```

```
trade.chisq.tb
```

```
## Source: local data frame [5 x 6]
```

```
##
```

	Q1	Age18to24	Age25to39	Age40to59	Age60to79	Age80plus
	(fctr)	(int)	(int)	(int)	(int)	(int)
## 1	Very good	26	51	88	83	23
## 2	Somewhat good	38	83	143	148	39
## 3	Somewhat bad	14	21	45	46	8
## 4	Very bad	5	8	30	35	10
## 5	DK/Refused	6	10	17	16	9

Now we test the null hypothesis that the answer to question 1 is independent of age with a 5% significance level with the Pearson's  $\chi^2$  test.

```
chisq.test(trade.chisq.tb[,2:6])
```

```
##
```

```
## Pearson's Chi-squared test
```

```
##
```

```
## data: trade.chisq.tb[, 2:6]
```

```
## X-squared = 13.586, df = 16, p-value = 0.6296
```

The p-value for this test is greater than 5% (0.05), therefore we cannot reject the null hypothesis that the question answer is independent of age for the question on whether Americans favour the growing trade agreement with Germany.