Subjective Well-Being Data Task

Guillermo Ortiz

4/4/2022

Executive Summary

Inequality has been a growing topic of discussion for economists and for society more broadly. Income inequality has received the bulk of the attention, but we may also care about inequality as is pertains to non-market goods such as ”how rewarding your life is” or ”your sense of security”. In this Data Task, I evaluate subjective measures of well-being and try to better understand them, their determinants and their implications for policy.

Loading Packages

```r
if(!require(tidyverse)) install.packages("tidyverse")
```

```
## Loading required package: tidyverse
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.6
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.1.2
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(tidyverse)

if(!require(rsample)) install.packages("rsample")
```

```
## Loading required package: rsample
```

```
## Warning: package 'rsample' was built under R version 4.1.3
```

```
library(rsample)
```

Question 1

a.) Load ratings.csv

```
ratings <- read.csv("ratings.csv")
```

b.) There are 1,056 unique workers and 17 unique aspects in the data set.

```
#unique respondents:
unique(ratings$worker) %>% length() #1,056 unique workers
```

```
## [1] 1056
```

```
#unique aspects:
unique(ratings$aspect) %>% length() #17 unique aspects
```

```
## [1] 17
```

c.) There are 211 respondents with more than 17 answers, i.e., who answered at least one aspect more than once. After including only the most recent rating for each worker-aspect, 237 observations were dropped.

```
count(ratings, worker) %>% filter(n > 17) %>% nrow() #211 respondents have more than 17 answers.
```

```
## [1] 211
```

```
#total observations:
nrow(ratings) #18,189 total observations
```

```
## [1] 18189
```

```
#unique workers times unique workers:
1056*17 #17,952
```

```
## [1] 17952
```

```
#need to drop:
18189-17952
```

```
## [1] 237
```

```
#drop:
ratings_u <- ratings %>% group_by(worker, aspect) %>% filter(time==max(time)) %>% ungroup()

#dropped:
nrow(ratings)-nrow(ratings_u)
```

```
## [1] 237
```

d.) Report min, 25th percentile, 50th percentile, 75th percentile, and maximum subjective riches value.

```
subjective_riches <- ratings %>% group_by(worker) %>% summarise(avg=mean(rating)) %>% .$avg
summary(subjective_riches)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.765  49.029  61.441  61.646  75.015 100.000
```

Question 2

a.) Load demographics.csv

```
demographics <- read.csv("demographics.csv")
```

b.) Number of rows of demographics dataset = 1,056. It is the same as the number of unique respondents from question 1.

```
nrow(demographics) #1,056
```

```
## [1] 1056
```

```
nrow(demographics) == unique(ratings$worker) %>% length()
```

```
## [1] TRUE
```

c.) Merge subjective riches data with demographics data.

```
demographics_m <- demographics %>% left_join(
  ratings %>% group_by(worker) %>%
    summarise(subjective_riches=mean(rating)),
  by="worker"
)
```

d.) Income is positively and significantly correlated with subjective riches data. An increase of 10,000 monetary units in total household income is associated with an increase of 0.92 points in the subjective riches measure.

```
summary(lm(subjective_riches~income, data=demographics_m))
```

```
## 
## Call:
## lm(formula = subjective_riches ~ income, data = demographics_m)
## 
## Residuals:
##    Min     1Q Median     3Q    Max 
## -57.97 -12.95   0.19  13.06  41.84 
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 5.630e+01  1.010e+00  55.734  < 2e-16 ***
## income      9.241e-05  1.437e-05   6.431 1.92e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 18.65 on 1054 degrees of freedom
## Multiple R-squared:  0.03775,    Adjusted R-squared:  0.03684 
## F-statistic: 41.35 on 1 and 1054 DF,  p-value: 1.922e-10
```

e.) After controlling for age, gender, education and race, income is still significantly correlated with subjective riches. However, the magnitude of the association has now marginally decreased. An increase of 10,000 monetary units in total household income now is associated with an increase of 0.87 points in the subjective riches measure. In addition, it seems that men and multiracial ethnicities have, on average, higher subjective riches scores than asian women (baseline case). Education levels and other ethnicities do not seem to have significance in explaining subjective riches scores.

```
summary(lm(subjective_riches~income+age+c(age^2)+male+education+race, data=demographics_m))
```

```
## 
## Call:
## lm(formula = subjective_riches ~ income + age + c(age^2) + male + 
##     education + race, data = demographics_m)
## 
## Residuals:
##    Min     1Q Median     3Q    Max 
## -58.53 -12.65  -0.25  12.68  45.60 
## 
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                      8.415e+01  1.958e+01   4.297 1.89e-05 ***
## income                           8.697e-05  1.523e-05   5.709 1.48e-08 ***
## age                             -3.529e-01  2.959e-01  -1.192   0.2333    
## c(age^2)                         4.009e-03  3.420e-03   1.172   0.2414    
## male                              2.570e+00  1.182e+00   2.175   0.0298 *
## educationBachelor's degree       -2.064e+01  1.866e+01  -1.106   0.2689
## educationDoctoral degree         -1.853e+01  1.919e+01  -0.966   0.3343
## educationGraduate degree         -2.174e+01  1.891e+01  -1.149   0.2507
## educationHigh school             -2.300e+01  1.874e+01  -1.227   0.2199
## educationLess than high school -1.610e+01  2.044e+01  -0.788   0.4310
## educationMaster's degree         -2.337e+01  1.877e+01  -1.245   0.2134
## educationSome college            -2.370e+01  1.865e+01  -1.271   0.2041
## raceBlack (non-Hispanic)         -1.381e+00  3.074e+00  -0.449   0.6533
## raceHispanic (any race)          -1.698e+00  3.296e+00  -0.515   0.6065
## raceMultiracial                   7.718e+00  4.410e+00   1.750   0.0804 .
## raceOther                         8.899e+00  8.023e+00   1.109   0.2676
## raceWhite (non-Hispanic)          6.307e-01  2.392e+00   0.264   0.7921
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.6 on 1039 degrees of freedom
## Multiple R-squared:  0.0568, Adjusted R-squared:  0.04227
## F-statistic:  3.91 on 16 and 1039 DF,  p-value: 3.334e-07
```

f.) Since more people in the household generally means a higher household income, I assume that including household size as a control would erode the income variable's significance in explaining subjective riches. This is because the income variable is measured as total household income instead of average household income. In addition, there may be other variables that better explains subjective riches measures not included in the dataset.

Question 3

a.) First, I would take the average of the subjective aspects related with health, both mental and physical. Those are: "you not feeling anxious", "your health", "your mental health" and "your physical fitness". I would call "health" to that indicator and merge it with the demographics dataset. Next, I would stratify the age variable by the nearest 10 and the income variable by using its quartiles. Finally, I would calculate the health indicator's average by each age group and give it a different color for each income strata. This would make the graph a little easier to understand.

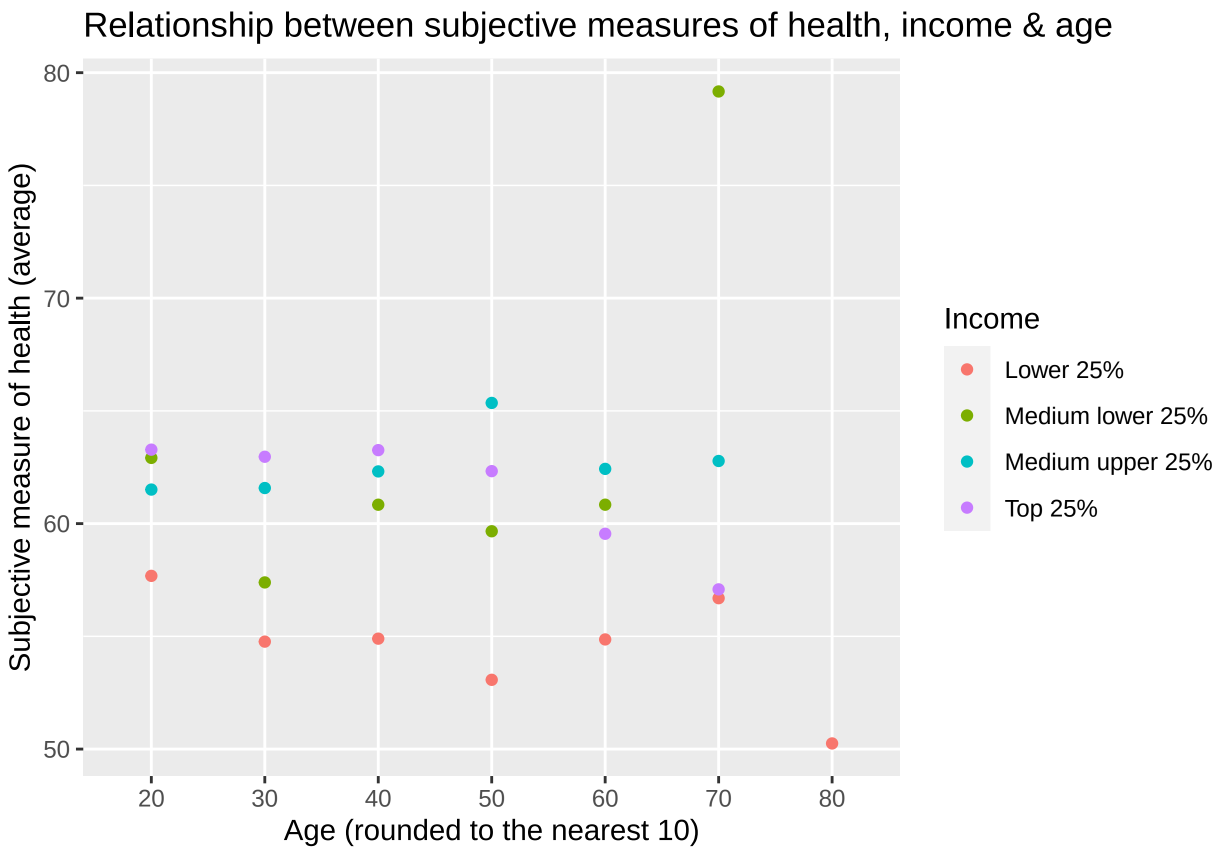
b.) Produce and save the scatterplot.

```
ratings_h <- ratings_u %>%
  filter(aspect%in%c("you not feeling anxious", "your health",
                     "your mental health", "your physical fitness")) %>%
  group_by(worker) %>%
  summarise(health=mean(rating)) %>% ungroup()

demographics_h <- left_join(demographics_m, ratings_h,
          by="worker")

demographics_h %>% mutate(age_strata=factor(round(age,-1)),
                          income_strata=make_strata(income)) %>%
  group_by(age_strata, income_strata) %>% summarise(avg_health=mean(health)) %>%
  ggplot(aes(age_strata,avg_health, col=income_strata)) + geom_point() +
  scale_color_discrete(name = "Income", labels = c("Lower 25%", "Medium lower 25%", "Medium upper 25%",
  ggtitle("Relationship between subjective measures of health, income & age") +
  xlab("Age (rounded to the nearest 10)") +
  ylab("Subjective measure of health (average)")
```

```
## `summarise()` has grouped output by 'age_strata'. You can override using the `.groups` argument.
```



c.) I believe that the analysis provided in this Data Task may have helped clarify which the determinants of well-being are and, maybe, which are not. Since the proxies used for well-being in this analysis (subjective ratings) depend on several highly volatile factors (attitude of the respondent, mood of that particular day, subjective comparisons among immediate peers), they may not be the best measures of inequality to steer public policy. In fact, these subjective measures of well-being may be very highly influenced by the subjects' enviroment, thus preventing them from properly diagnose themselves nor to assess their own well being on a broader sense. This measures are not completely impractical, however, since they may be telling something about the attitudes and aptitudes of the respondents, and their receptivity to inequality reducing policies.