Exam2

Kurata

6/27/2020

Q1 & 2

Q3

This is a cross-sectional data set as the summary for year below suggests that all data is from 2015, and the inequality_gini variable is recorded by country.

```
head(inequality_data)
```

```
##   iso2c country inequality_gini year
## 1    AL Albania            32.9 2015
## 2    AM Armenia            32.4 2015
## 3    AT Austria            30.5 2015
## 4    BY Belarús            25.6 2015
## 5    BE Belgium            27.7 2015
## 6    BZ  Belize              NA 2015
```

```
str(inequality_data)
```

```
## 'data.frame':    203 obs. of  4 variables:
##  $ iso2c          : chr  "AL" "AM" "AT" "BY" ...
##  $ country        : chr  "Albania" "Armenia" "Austria" "Belarús" ...
##  $ inequality_gini: num  32.9 32.4 30.5 25.6 27.7 NA 47.8 NA NA 46.7 ...
##  $ year           : num  2015 2015 2015 2015 2015 ...
```

```
summary(inequality_data$year)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2015    2015    2015    2015    2015    2015
```

Q4

```
##     iso2c country inequality_gini year
## 174    SE  Sweden            29.2 2015
```

```
##    iso2c country inequality_gini year
## 40    DK Denmark            28.2 2015
```

Q5

```
##    iso2c country inequality_gini year
## 13    BR  Brazil            51.9 2015
```

Q6

It is better to have lower inequality_gini scores.

Q7

```
##   iso2c country inequality_gini year
## 1    AL Albania            32.9 2015
## 2    AM Armenia            32.4 2015
## 3    AT Austria            30.5 2015
## 4    BY Belarús            25.6 2015
## 5    BE Belgium            27.7 2015
## 6    BZ  Belize              NA 2015
```

Q8

```
##   iso2c country inequality_gini year
## 1    AL Albania            32.9 2015
## 2    AM Armenia            32.4 2015
## 3    AT Austria            30.5 2015
## 4    BY Belarus            25.6 2015
## 5    BE Belgium            27.7 2015
## 6    BZ  Belize              NA 2015
```

Q9

```
##     iso2c        country inequality_gini year
## 161    SI       Slovenia            25.4 2015
## 190    UA        Ukraine            25.5 2015
## 4      BY        Belarus            25.6 2015
## 39     CZ Czech Republic            25.9 2015
## 92     XK         Kosovo            26.5 2015
```

Q10

```
avg = mean(inequality_data$inequality_gini, na.rm = TRUE)
```

Q11 create new variables

Q12 crosstab

```
##   high_inequality low_inequality.mean
## 1               0                  1
## 2               1                  0
## 3              NA                 NA
```

Q13 for loop

```
## [1] "World Bank"
## [1] "African Development Bank"
## [1] "Bill and Melinda Gates Foundation"
```

Q14

Government Expenditure on Education may correlate to inequality, since I assume less equal countries to have less accessible education.

Q15 import variables

Q16 rename variable

```
library(data.table)
setnames(education_expenditure,"SE.XPD.TOTL.GB.ZS","total_education_expenditure")
```

Q17 merge data sets

Q18 remove countries with NA

Q19 filter data

Q20 count

Q21 apply

```
##  [1] 30.4 30.5 31.7 31.8 31.8 32.3 32.4 32.7 32.8 32.9 33.2 33.5 33.8 34.0 34.0
## [16] 34.2 35.0 35.4 35.5 35.9 35.9 36.2 37.4 37.7 39.5 40.5 40.6 40.8 41.0 41.0
## [31] 41.5 42.4 43.1 43.4 44.4 46.0 47.8 48.4 49.6 51.1 51.9 57.1
```

Q22 label

Q23 save labels