542

March 17, 2020

1 Happiness, Inequality, and Mental Health Policy Preprocessing the Data

1.0.1 Bingyao Zou, Aurora Deng, Andrew Golden

---

2 Introduction

2.0.1 *Bacground*

Happiness is a feeling that comes over you when life is good. According to current studies, six key variables are found to support happiness: income; healthy life expectancy; social support; freedom; trust; and generosity. Of these six predictors of happiness, the most important is income, which at the national level is measured as the Gross Domestic Product (GDP) or economic growth. We admit that economic security drives greater contentment, and thus leads to higher happiness scores. In more developed countries, however, economic growth may not buy citizens as much gains in well-being as poor countries.

In this case, our group is assessing income in a different way—**Income Inequality & Economic Inequality**. We assume that if a country's economic growth is not equally distributed, it will lead to lower levels of happiness for its citizens. Moreover, a sense of fairness and trust is the cornerstone of community, and having satisfying social relationships in the community is essential to improving well-being.

Before exploring the relationship between happiness and inequality, we need to rule out some potential impacts, among which the impact our group is focusing on comes from governments' actions to improve happiness. It is well known that some countries provide **mental health support** in order to support their citizens, such as funding, coordination, legislation, the establishment of information systems, and the procurement and distribution of essential medicines. Given the huge influence of governmental efforts, our group decided to separate the countries with and without published mental health policies.

In general, our group's research topic is ***how inequality relates to reported happiness in countries with and without mental health programs.***

2.0.2 *Data and Variables*

Happiness Score by Country

- Description: The national happiness level based on respondents' rating of their own lives
- Source: 2019 United Nations World Happiness Report, using data collected by Gallup
- Data Type:
  - Numerical, from 0 to 100
  - Higher is better
- Limitation: Happiness is hard to quantify and measure, which can mean different things to different people, and across cultures. Since there is no worldwide standard for happiness, the research associated with it is considered to be less reliable.

GINI Index

- Description: The degree of inequality in the distribution of family income in a country
- Source: 2010 Central Intelligence Agency, The World Factbook
- Data Type:
  - Numerical, from 0 to 100
  - Higher indicates greater inequality

Mental Policies Summary

- Description: Whether a country is in action for mental health
- Source: 2017 World Health Organization (WHO)
- Data Type:
  - Binary, 0 or 1
  - 1 indicates the country has at least one mental policy, 0 indicates no policy
- Processing Detail: The original data lists the specific mental-health related public policies of each country. We define a new binary variable based on the summary so that we can see directly whether a country's government publishes their mental policy or not
- Limitation: Binary variable can be in only one of two categories — either yes or no, so we can only learn if a country has a related policy. However, among these policies, some are heavily invested while some are still in the planning stage. In order to better distinguish their specific impacts on the happiness score, it would be useful to suggest a ladder which can represent the scale of these policies.

3 Data Cleaning

Get the tables using pandas

[1]:
```
import pandas as pd
from urllib.request import urlopen
```

[2]:
```
```

```
gini=pd.read_pickle(urlopen("https://github.com/argolden/
 ↪computational-governance/raw/master/gini.pkl"),compression=None)
happiness=pd.read_pickle(urlopen("https://github.com/Zoubyyy37/PUBPOL542Group/
 ↪raw/master/2019happiness.pkl"),compression=None)
mentalhealth=pd.read_pickle(urlopen("https://github.com/auroraD-11/MyData/raw/
 ↪master/MH1.pkl"),compression=None)
```

Check/Rename column names

[3]:
```
gini.columns
```

[3]:
```
Index(['Rank', 'Country', 'GINI', 'Year'], dtype='object')
```

[4]:
```
happiness.columns
```

[4]:
```
Index(['countryorregion', 'scoreofhappiness'], dtype='object')
```

[5]:
```
mentalhealth.columns
```

[5]:
```
Index(['PUBLISHSTATES', 'Year', 'WHOregion', 'Country', 'Law',
       'GovExpenditures', 'PolicyPlan', 'LawEnactedYear', 'PPPublicYear'],
      dtype='object')
```

[6]:
```
#In order to have a common column "country"
happiness.columns = ['Country', 'scoreofhappiness']
```

The original mentalhealth data has been recorded for multiple years. Our group decided to keep the latest record and droped previous data.

[7]:
```
mentalhealth2 = mentalhealth[['Country','PolicyPlan','Year']]
```

[8]:
```
mentalhealth2.sort_values('Country')
```

[8]:

	Country	PolicyPlan	Year
286	Afghanistan	Yes	2016
152	Afghanistan	Yes	2014
110	Albania	Yes	2014
258	Albania	Yes	2016
57	Algeria	Yes	2014
..	...	...	...
160	Yemen	Yes	2014
223	Zambia	Yes	2016
55	Zambia	Yes	2014
46	Zimbabwe	Yes	2014
224	Zimbabwe	Yes	2016

[346 rows x 3 columns]

[9]:
```
mentalhealth3 = mentalhealth2.drop_duplicates(subset='Country', keep='first')
```

[10]:
```
mentalhealth4 = mentalhealth3[['Country','PolicyPlan']]
```

4 Data Integration

Merge gini data and happiness data and save the new data frame

[11]:
```
dirtymerge1=gini.merge(happiness,how='outer',indicator=True)
```

[12]:
```
#Request the countries where the happiness data frame found no match
dirtymerge1.loc[dirtymerge1['_merge']=='right_only',"Country"]
```

[12]:
```
157             Czech Republic
158       United Arab Emirates
159                      Qatar
160                    Bahrain
161          Trinidad & Tobago
162                     Kuwait
163                South Korea
164            Northern Cyprus
165                      Libya
166            North Macedonia
167                    Lebanon
168                Ivory Coast
169        Congo (Brazzaville)
170    Palestinian Territories
171                    Somalia
172                     Gambia
173                       Iraq
174           Congo (Kinshasa)
175                    Myanmar
176                  Swaziland
177                      Syria
178                Afghanistan
Name: Country, dtype: object
```

[13]:
```
#Request the countries where the gini data frame found no match
dirtymerge1.loc[dirtymerge1['_merge']=='left_only',"Country"]
```

[13]:
```
2        Micronesia, Federated States of
12                     Papua New Guinea
16                             Eswatini
17                          Gambia, The
19                Congo, Republic of the
```

```
40                                Guyana
48                                Angola
52     Congo, Democratic Republic of the
55                          Cote d'Ivoire
59                               Djibouti
75                               Maldives
89     Falkland Islands (Islas Malvinas)
92                            Korea, South
96                                   Macau
100                           West Bank
105                           Greenland
109                           Macedonia
121                          Timor-Leste
126                        European Union
127                Sao Tome and Principe
150                              Czechia
155                        Faroe Islands
156                               Jersey
Name: Country, dtype: object
```

[14]:
```
#Improve the merge result
replacements1={'Gambia, The': 'Gambia',
               'Congo (Brazzaville)': 'Congo, Republic of the',
               'Congo (Kinshasa)': 'Congo, Democratic Republic of the',
               'Ivory Coast': "Cote d'Ivoire",
               'Korea, South': 'South Korea',
               'North Macedonia': "Macedonia",
               'Czechia': 'Czech Republic'}
```

[15]:
```
happiness.Country.replace(replacements1,inplace=True)
```

[16]:
```
gini.Country.replace(replacements1,inplace=True)
```

[17]:
```
dirtyMerge2=gini.merge(happiness,left_on="Country",␣
 ↪right_on='Country',how='outer',indicator=True)
```

[18]:
```
dirtyMerge2
```

[18]:

	Rank	Country	GINI	Year	scoreofhappiness	\
0	1.0	Lesotho	63.2	1995	3.802	
1	2.0	South Africa	62.5	2013	4.722	
2	3.0	Micronesia, Federated States of	61.1	2013	NaN	
3	4.0	Haiti	60.8	2012	3.597	
4	5.0	Botswana	60.5	2009	3.488	
..	...	...	...	...	...	
167	NaN	Iraq	NaN	NaN	4.437	
168	NaN	Myanmar	NaN	NaN	4.360	

```
169   NaN                        Swaziland   NaN   NaN             4.212
170   NaN                            Syria   NaN   NaN             3.462
171   NaN                      Afghanistan   NaN   NaN             3.203

         _merge
0          both
1          both
2     left_only
3          both
4          both
..          ...
167  right_only
168  right_only
169  right_only
170  right_only
171  right_only

[172 rows x 6 columns]
```

Merge mental policy data and previously merged data and save the new data frame

[19]:
```
dirtyMerge3=dirtyMerge2.merge(mentalhealth4, on='Country')
```

Our group's Data Frame:

[20]:
```
df2=dirtyMerge3[['Country','GINI','scoreofhappiness','PolicyPlan']]
```

[21]:
```
#Drop the missiving value
dffinal=df2.dropna()
```

[22]:
```
dffinal
```

[22]:
```
          Country  GINI  scoreofhappiness PolicyPlan
0         Lesotho  63.2             3.802         No
1    South Africa  62.5             4.722        Yes
2           Haiti  60.8             3.597         No
3        Botswana  60.5             3.488        Yes
4         Namibia  59.7             4.639        Yes
..            ...   ...               ...        ...
124       Belgium  25.9             6.923        Yes
125       Ukraine  25.5             4.332         No
126        Sweden  24.9             7.343        Yes
127      Slovenia  24.4             6.118         No
128      Slovakia  23.7             6.198        Yes

[121 rows x 4 columns]
```

5 Saving File to Disk

For future use in R

[23]:
```
dffinal.to_pickle("dffinal.pkl")
```

[24]:
```
from rpy2.robjects import pandas2ri
pandas2ri.activate()

from rpy2.robjects.packages import importr

base = importr('base')
base.saveRDS(dffinal,file="dffinal.RDS")
```

```
---------------------------------------------------------------------------

ModuleNotFoundError                       Traceback (most recent call last)

<ipython-input-24-e702f883aa90> in <module>
----> 1 from rpy2.robjects import pandas2ri
      2 pandas2ri.activate()
      3
      4 from rpy2.robjects.packages import importr
      5


ModuleNotFoundError: No module named 'rpy2'
```