GDP and Future Orientation

Alex Huang

GDP and Future Orientation

Overview:

This project reproduce the findings of the article at http://www.nature.com/articles/srep00350. According to the findings, the GDP/capita of countries are positively correlated to how much their population searches for the next year, relative to how much it searches for the previous year. This ratio is dubbed the Future Orientation Index (FOI). So for example for the year 2017 the FOI can be calculated as: FOI = number of searches for the term "2018" / number of searches for the term "2016".

Installation

Install the required packages.

Run the following commands in your R console to install WDI and gtrendsR.

```
install.packages("WDI")
install.packages("devtools")
if (!require("devtools")) install.packages("devtools")
devtools::install_github("PMassicotte/gtrendsR")
```

World Bank Dat (WDI)

We will use WDI to load data on GDP/capita and number of internet users per country from the World Development Indicators.

Load the WDI package.

```r
require(WDI)
```

```
## Loading required package: WDI
```

```
## Loading required package: RJSONIO
```

Extract the needed data.

We need the Gross Domestic Produce (GDP) per capita corrected by the Purchase Power Parity (PPP). PPP is a way to compare GDP by accounting for cost of goods in the country rather than market exchange. The GDP per capita PPP data reflects more what citizens of the country can buy. In the WDI database this is referred to as NY.GDP.PCAP.PP.KD. (this indicator can be found using the WDIsearch function - see the WDI reference for details if you are interested in finding other data: https://cran.r-project.org/web/packages/WDI/WDI.pdf). We need data from all countries in the year 2016. Below is a map of countries with GDP per capita PPP as colour.

```r
table = WDI(indicator=c('NY.GDP.PCAP.PP.KD','IT.NET.BBND','SP.POP.TOTL'), country='all', start = 2016,
summary(table)
```

```
##     iso2c              country               year      NY.GDP.PCAP.PP.KD
##  Length:264         Length:264         Min.   :2016   Min.   :   647.9
##  Class :character   Class :character   1st Qu.:2016   1st Qu.:  3980.2
##  Mode  :character   Mode  :character   Median :2016   Median : 12045.1
```

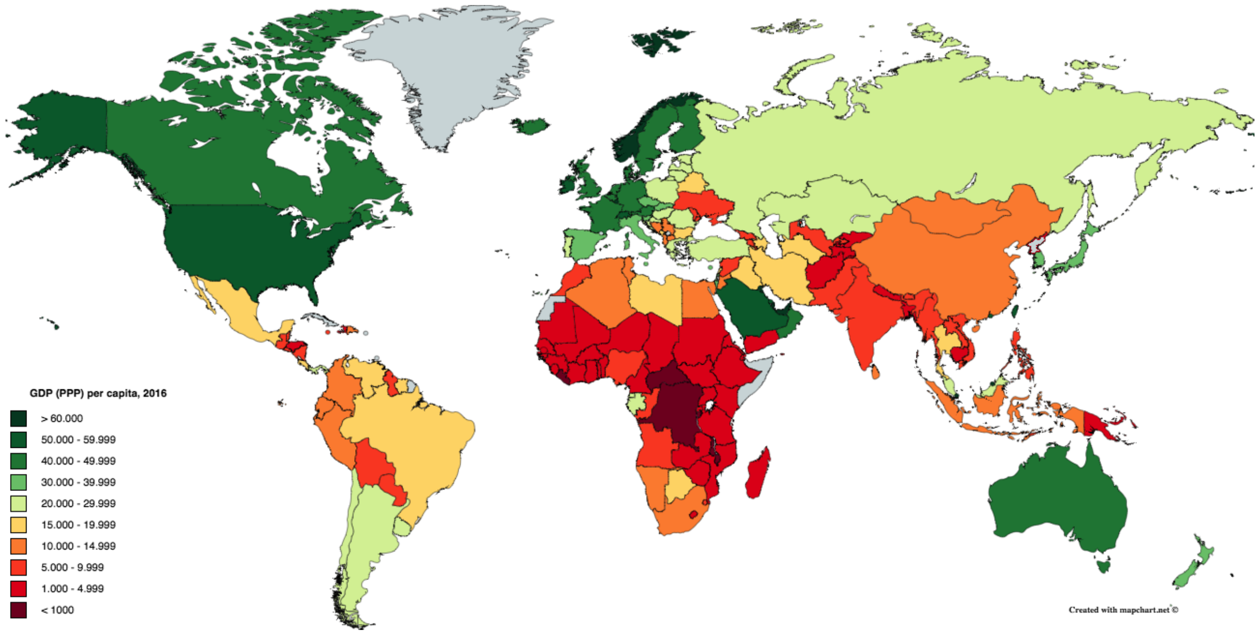


Figure 1: GDP Per Capita by Country in 2016

```
##                                        Mean   :2016   Mean   : 17396.4
##                                        3rd Qu.:2016   3rd Qu.: 23679.4
##                                        Max.   :2016   Max.   :118215.3
##                                                       NA's   :43
##   IT.NET.BBND          SP.POP.TOTL           iso3c
##  Min.   :       80    Min.   :1.110e+04   ABW    :  1
##  1st Qu.:    30430    1st Qu.:1.523e+06   AFG    :  1
##  Median :   573000    Median :1.011e+07   AGO    :  1
##  Mean   : 35342647    Mean   :3.009e+08   ALB    :  1
##  3rd Qu.:  5068589    3rd Qu.:5.582e+07   AND    :  1
##  Max.   :916669071    Max.   :7.442e+09   (Other):226
##  NA's   :22           NA's   :2           NA's   : 33
##                                             region           capital
##  Europe & Central Asia (all income levels)     :56                : 26
##  Sub-Saharan Africa (all income levels)        :47   Abu Dhabi   :  1
##  Latin America & Caribbean (all income levels) :41   Abuja       :  1
##  East Asia & Pacific (all income levels)       :35   Accra       :  1
##  Middle East & North Africa (all income levels):21   Addis ababa:  1
##  (Other)                                       :31   (Other)     :201
##  NA's                                          :33   NA's        : 33
##      longitude        latitude                      income
##           : 26             : 26   Lower middle income :54
##  -0.126236:  1   -0.229498:  1   Upper middle income :53
##  -0.20795 :  1   -1.27975 :  1   High income: nonOECD:38
##  -1.53395 :  1   -1.95325 :  1   Low income          :34
##  -10.7957 :  1   -11.6986 :  1   High income: OECD   :31
##  (Other)  :201   (Other)  :201   (Other)             :21
##  NA's     : 33   NA's     : 33   NA's                :33
##           lending
##  Aggregates     :20
##  Blend          :15
```

```
##  IBRD          :62
##  IDA           :64
##  Not classified:70
##  NA's          :33
##
```

The summary gave a nice overview of our data. Additionally, we can use the head() function to take a look at what the actual data in the table looks like

```r
head(table)
```

```
##   iso2c                                        country year
## 1    1A                                     Arab World 2016
## 2    1W                                          World 2016
## 3    4E   East Asia & Pacific (excluding high income) 2016
## 4    7E Europe & Central Asia (excluding high income) 2016
## 5    8S                                     South Asia 2016
## 6    AD                                        Andorra 2016
##   NY.GDP.PCAP.PP.KD IT.NET.BBND SP.POP.TOTL iso3c
## 1         15533.695    18419019   406452690   ARB
## 2         15023.509   916669071  7442135578   WLD
## 3         12841.957   352841430  2051431154  <NA>
## 4         18523.417    64913332   417424643  <NA>
## 5          5621.096    27671379  1766383450   SAS
## 6                NA       32490       77281   AND
##                                      region          capital longitude
## 1                                Aggregates
## 2                                Aggregates
## 3                                      <NA>             <NA>      <NA>
## 4                                      <NA>             <NA>      <NA>
## 5                                Aggregates
## 6 Europe & Central Asia (all income levels) Andorra la Vella    1.5218
##   latitude               income        lending
## 1                    Aggregates     Aggregates
## 2                    Aggregates     Aggregates
## 3     <NA>                 <NA>           <NA>
## 4     <NA>                 <NA>           <NA>
## 5                    Aggregates     Aggregates
## 6  42.5075 High income: nonOECD Not classified
```

Remove unwanted entries.

It seems like some of the entries are not countries, but regions! We can remove them by comparing whether the region field is equal to 'Aggregates'. And we added a column column called "FOI" to calculate the FOI value for the future.

```r
table = table[complete.cases(table),]
table = table[!(table$region=='Aggregates'),]
table = subset(table, select = c("iso2c", "country", "NY.GDP.PCAP.PP.KD", "IT.NET.BBND", "SP.POP.TOTL"))
table["INT.POP"]=table["IT.NET.BBND"]

table["IT.NET.USER.P3"]=NULL
table["SP.POP.TOTL"]=NULL

table$FOI = NA
```

Additionally, countries with less than 5 million inhabitants are removed. The article mentioned above did this to remove outliers

```
table=table[!(table$INT.POP<5000000),]
```

Google Trends

Set up gtrendsR

Log in to Google using your username and password. code not shown.

```
## Loading required package: devtools
```

```
## Downloading GitHub repo PMassicotte/gtrendsR@master
## from URL https://api.github.com/repos/PMassicotte/gtrendsR/zipball/master
```

```
## Installing gtrendsR
```

```
## "C:/PROGRA~1/R/R-34~1.2/bin/x64/R" --no-site-file --no-environ --no-save  \
##   --no-restore --quiet CMD INSTALL  \
##   "C:/Users/AlexH/AppData/Local/Temp/RtmpeGpY32/devtools3c603b06b88/PMassicotte-gtrendsR-1965b42"  \
##   --library="C:/Users/AlexH/Documents/R/win-library/3.4" --install-tests
```

```
##
```

```
## Loading required package: gtrendsR
```

Extract data from Google Trends.

For each country we need the FOI, which is the ratio between the volume of searches for "2015" and "2013". Note that with Google Trends we can query a maximum of 5 countries at a time, so we won't get all the data in one go. Rather it is worth making a for loop that goes through all the country codes.

Google Trends doesn't give absolute volumes, but relative ones. It always sets the largest data to 100 and scales the rest accordingly. However if you search for two things at the same time, or two countries at the same time, the results will have the correct proportion to each other. This means that to get the correct FOI, for each country code you need to extract data for "2015" and "2013" searches in the same go! In other words you should only have one call of the gtrends function in your for loop.

```
# start_date = as.Date("2016-01-01")
# end_date = as.Date("2017-01-01")
i=1
for(current_country in table[,"iso2c"])
  {
    print(current_country)

    result=gtrends(c("2015","2017"),geo=current_country, time="2016-01-01 2017-01-01")

    FOI=sum(result$interest_over_time$hits[result$interest_over_time$keyword=='2017'])/sum(result$intere
    print(FOI)
    table[i,'FOI']=FOI
    i=i+1
}
```

```
## [1] "AR"
## [1] 0.8840361
## [1] "AU"
## [1] 0.9797619
## [1] "BD"
## [1] 0.3114162
```

```
## [1] "BR"
## [1] 0.9150418
## [1] "CA"
## [1] 0.8068182
## [1] "CN"
## [1] 0.5279553
## [1] "CO"
## [1] 0.5804511
## [1] "DE"
## [1] 1.381967
## [1] "ES"
## [1] 0.7107001
## [1] "FR"
## [1] 1.057353
## [1] "GB"
## [1] 1.097842
## [1] "IN"
## [1] 0.4465355
## [1] "IT"
## [1] 0.9129815
## [1] "JP"
## [1] 1.182648
## [1] "KR"
## [1] 0.5159106
## [1] "MX"
## [1] 0.6945607
## [1] "NL"
## [1] 1.14526
## [1] "PH"
## [1] 0.3491228
## [1] "PL"
## [1] 0.5133333
## [1] "RU"
## [1] 0.4620107
## [1] "TH"
## [1] 0.3610315
## [1] "TR"
## [1] 0.7043597
## [1] "UA"
## [1] 0.3617437
## [1] "US"
## [1] 0.8503937
## [1] "VN"
## [1] 0.4262821
```

```
#table = table[complete.cases(table),]
```

Regress GDP per capita PPP on FOI and plot

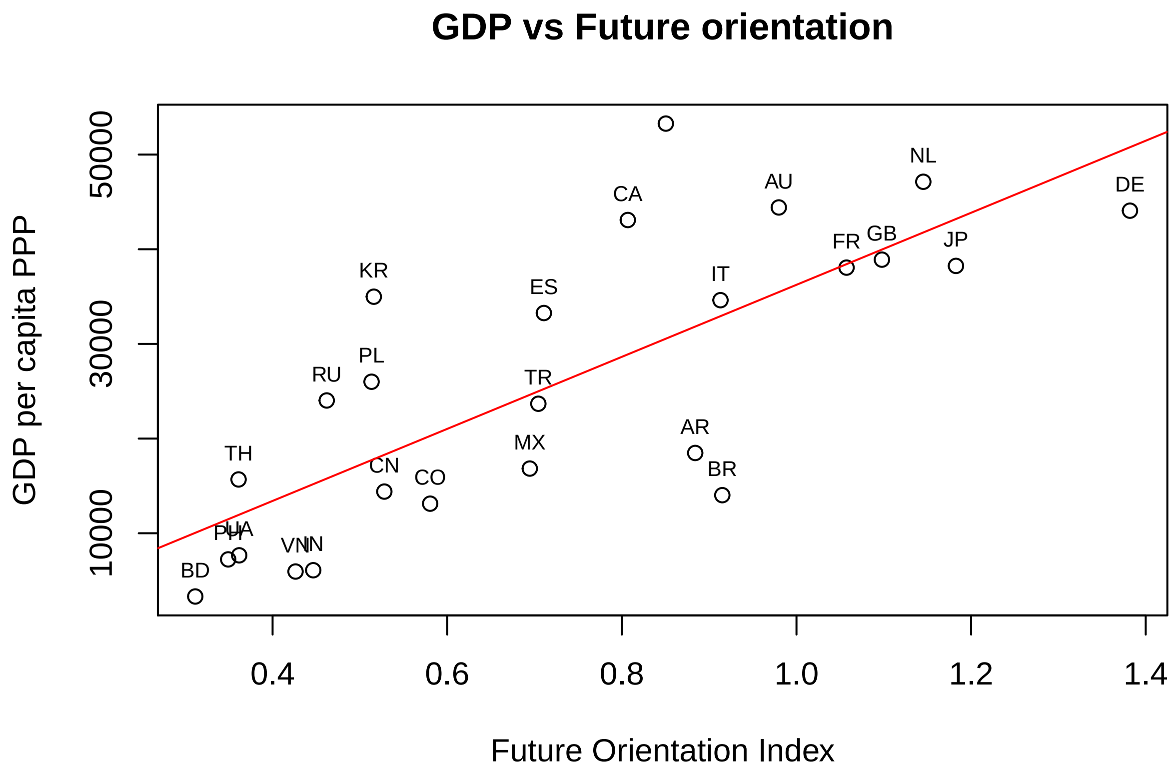
Now that we have the FOI index and GDP per capita, PPP value for each country, we can make a regression and plot the result.

```
reg = lm(NY.GDP.PCAP.PP.KD~FOI, data=table)
summary(reg)
```

```
##
```

```
## Call:
## lm(formula = NY.GDP.PCAP.PP.KD ~ FOI, data = table)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -18987  -6724  -1315   8026  22722
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1806       5296  -0.341    0.736
## FOI            38049       6740   5.645 9.55e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10030 on 23 degrees of freedom
## Multiple R-squared:  0.5808, Adjusted R-squared:  0.5626
## F-statistic: 31.87 on 1 and 23 DF,  p-value: 9.555e-06
```

```r
plot(table$FOI, table$NY.GDP.PCAP.PP.KD, main='GDP vs Future orientation', ylab='GDP per capita PPP', xl
text(table$FOI, table$NY.GDP.PCAP.PP.KD, labels=table$iso2c, cex= 0.7, pos=3)
abline(reg, col='red')
```



We can see in this plot that there is a positive correlation between FOI and GDP per capita PPP value. And the relationship is statistically significant. (high t value and low p value)

One might be quick to conclude that countries that look towards the future caused them to be richer. However determining causality is much more complicated, there are more than 1 possible explanation for the result.

An alternative explanation could be that richer countries have more necessities taken care of and have the ability to look forward.