Country GDP Growth v. Tourism Expenditure Growth

Vinicius Cid

Berk Kirikoglu

Jeffrey Li

Applied Quantitative Methods

Project One

AQM

Country GDP Growth v. Tourism Expenditure Growth

Introduction

For our first project, we have chosen to put our R Language skills to the test by demonstrating our skills in the areas of data mining, data analytics, and data visualization. We have undertaken the challenge to discover whether or not if there is an underlying relationship between a country's Gross Domestic Product (GDP) rate of change and its tourism expenditure rate of change. The purpose of this report is to outline the methodology used to approach this task and record the findings as a result of the task undertaken.

Methodology

Our methodology was chosen according to the recent R Language skills we have learned including but not limited to: RStudio developer tools, plyr package, dplyr package, ggplot2 package, regression analysis charts, ANOVA tables, and APIs.

The following methodology was used for this report:

- Sourcing data frames using the World Development Indicator (WDI) Application Programming Interface (API)
- Examining data frames functions for data integrity using RStudio functions
- Manipulating data to achieve relevant information using plyr and dplyr
- Constructing data visuals to display findings using ggplot2

Sourcing and Selection

The data was sourced from the WDI using their API. The two sets of data we used were country GDP's in current US$, millions and tourism expenditures in current US$. The data sourced from the raw data are from the following 6 countries: Greece, Jamaica, Canada, Hong Kong, and Kazakhstan. The countries selected are from what our team perceives to be a blend of countries whose GDP is heavily reliant on tourism expenditure as well as those which are not. This careful selection is to ensure that we are able to capture a clear relationship between tourism expenditure and GDP if there one is present.

Data Integrity

After extracting data from the 6 countries there are two sets of relevant data frames. One is for GDP and one is for tourism expenditure, however, further examination was conducted to inspect the data frame such as column names, missing data, and data dimensions. This step was to confirm the data conform to our requirements to ensure the integrity of our findings. The data was then plotted for a visual inspection and concluded fit for modelling.

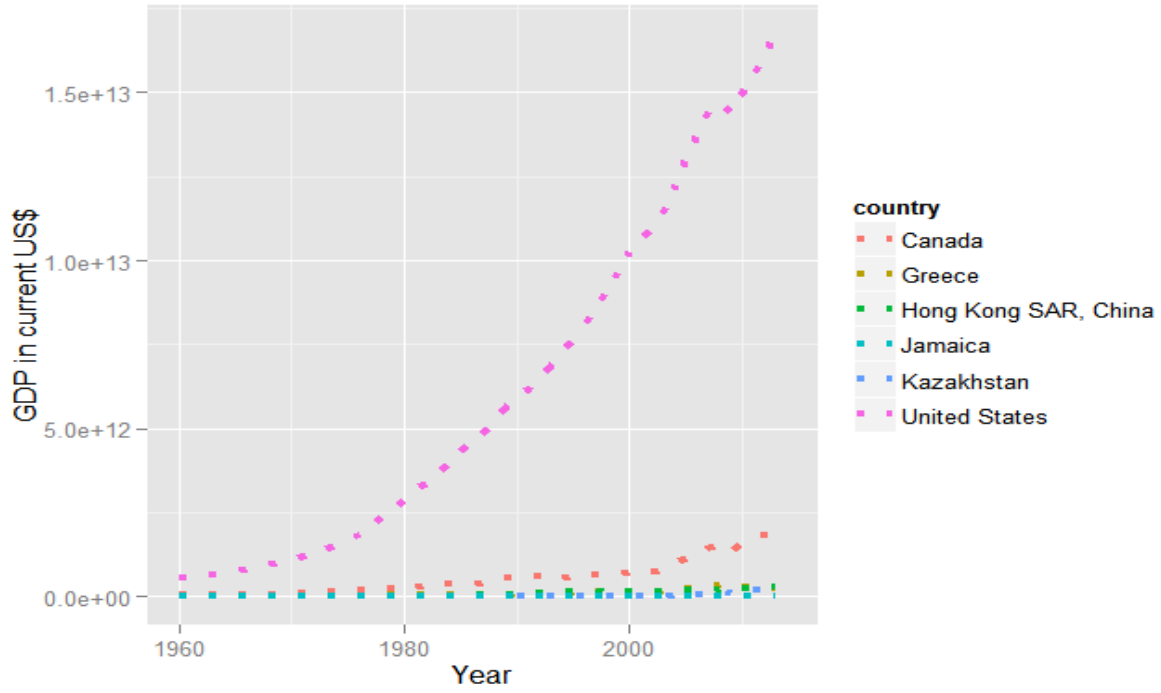

Figure 1. Plot of 6 selected countries' GDP against time in years

Data Manipulation

Data cleansing was the next step in the process. The data frames were further filtered for relevant information and a new variable was generated to determine the rate of change. The filtering enabled for both data frames to be from the same period of 1996 to 2011. As well, comparisons between GDP and tourism expenditure are now possible because of the matching units of measurement, namely the rates of change. The two data frames are easily comparable now that its dimensions and units of measurement are the same.

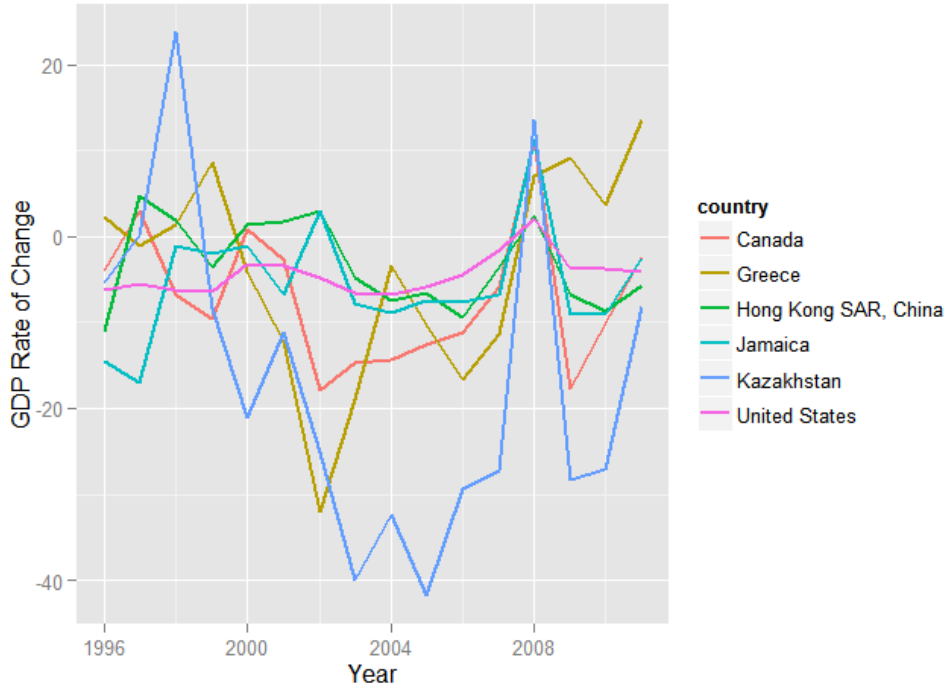

Figure 2. Plot of 6 countries' GDP rate of change against time in yeares

Data Visualization

The final step was to construct visuals to visually interpret the information and determine the findings. The first visual constructed was a scatterplot of the 6 countries' GDP rates of change and tourism expenditure rates of change. This enables us to see the general relationship.

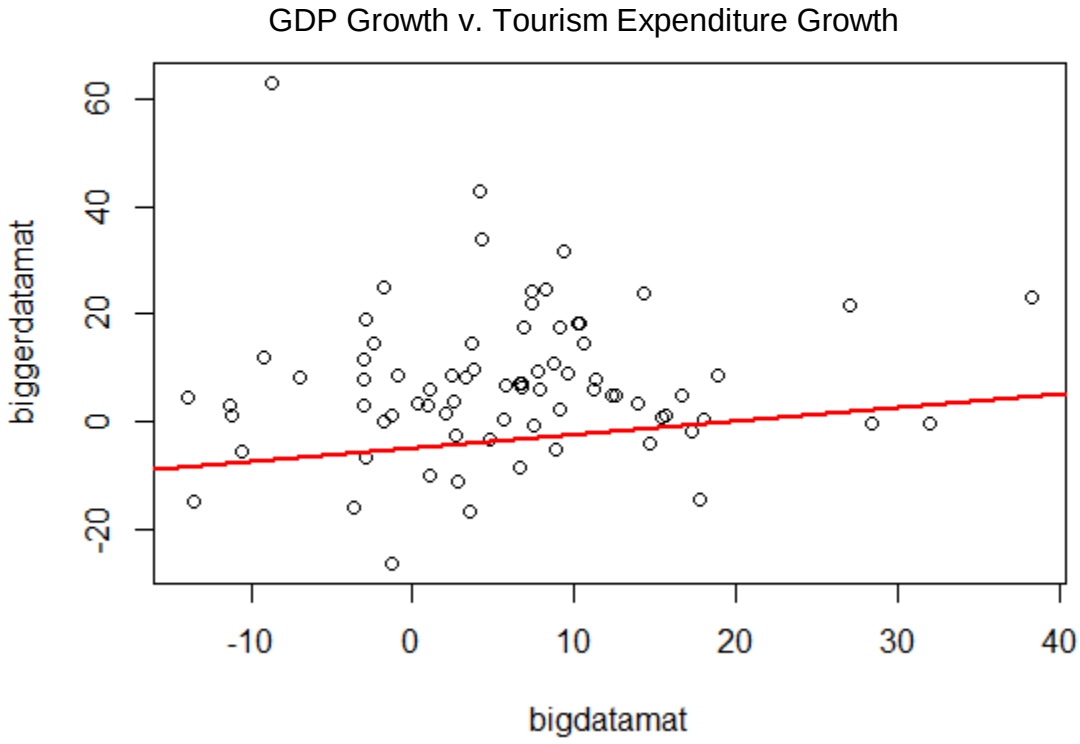

Figure 3. Scatterplot of GDP rate of change and tourism expenditure rate of change for all countries

Findings

As seen in the scatterplot between GDP and tourism expenditure rate of change (Figure 3), there is a moderate positive correlation between the two variables. Generally speaking, this means an increase in tourism expenditure will result in an increase in GDP. A best fit line is also included in the scatterplot to further illustrate this relationship.

R Source

```
#Load packages
library(WDI)
library(plyr)
library(dplyr)
library(ggplot2)

#Update data from WDI
new_cache = WDIcache()
WDIsearch('gdp', cache=new_cache)

#Search WDI for GDP data
WDIsearch("GDP, current US")

#Download data "GDP, current US$, millions"
dat = WDI(indicator='NY.GDP.MKTP.CD',
country=c('GR','JM','CA','HK','KZ','US'), start=1960, end=2014)

#Examine data
head(dat)
dat

#Columns: country, NY.GDP.NKTP.CD, year
#Rename column to GDP
colnames(dat)[colnames(dat)=="NY.GDP.MKTP.CD"] <- "GDP"

#Check data frame
head(dat)


#Kazakhstan missing data from before 1990
#Clean data frame from 1996 to 2011
#Added new column rate of change in GDP growth
```

```
bigdata <- dat %>%
  select(country, GDP, year) %>%
  group_by(country) %>%
  mutate(change = ((lag(GDP)-GDP)/GDP)*100) %>%
  filter(year > 1995, year < 2012)
bigdata
head(bigdata)

#Data for GDP Rate of Growth 1996 to 2011
#Year by year GDP rate of change by Country
answer <- ggplot(data = bigdata, aes(x = year, y = change, color =
country)) + geom_line(size = 1) + ylab ("GDP Rate of Change") + xlab
("Year")
answer

#Search WDI for Tourism Expenditure data
WDIsearch("Tourism")

#Download data "International tourism receipts (current US$)"
datt = WDI(indicator='ST.INT.RCPT.CD',
country=c('GR','JM','CA','HK','KZ','US'), start=1960, end=2014)

#Examine data
head(datt)
datt

#Only has data for 1995 onwards
#Rename column to Receipts
datt1 <- rename(datt, Receipts = ST.INT.RCPT.CD)
head(datt1)

#Clean data frame to from 1996 to 2011
#Added new column rate of change in Tourism Expenditures
```

```
biggerdata <- datt1 %>%
  select(country, Receipts, year) %>%
  group_by(country) %>%
  mutate(change = ((lag(Receipts)-Receipts)/Receipts)*100) %>%
  filter (year > 1995 & year < 2012)
head(biggerdata)

#Plot using ggplot
#Year by year Tourism Expenditure Rate of Change by Country
answer2 <- ggplot(data = biggerdata, aes(x = year, y = change, color =
country)) + geom_line(size = 1) + ylab("Tourism Money Change") +
xlab("Year")
answer2

#Formatting data frames into matrices
bigdatamat <- matrix(bigdata$change, nrow = 1, ncol = 96)
biggerdatamat <- matrix(biggerdata$change, nrow = 1, ncol = 96)

#Plotting GDP against Tourism Expenditure in scatterplot
plot(bigdatamat, biggerdatamat)
abline(lm(biggerdatamat ~ bigdatamat), col="red", size=1.5)

#Regressing GDP growth against Tourism Expenditure growth
#Residual plot
nicefit <- lm(bigdata$change ~ biggerdata$change)
nicefit
plot(nicefit)

summary(nicefit)
```