Assignment 2

Nivedita Mazumdar
School of Information Technology
Illinois State University
nmazumd@ilstu.edu

October 21, 2014

Abstract

This document analyses the Government Spending for North America

1 Introduction

Government spending refers to any expenditure made by local, regional, and national governments. In most countries, government spending makes up a significant portion of the GDP. Spending is accomplished in several major areas, including future investments, acquisitions, and transfer payments.Government spending or expenditure of country includes all government consumption, investment, and transfer payments.

Government spending or expenditure for a country is usually deeply scrutized.In general, the larger the government's deficit is, the less comfortable citizens feel about spending or investing money, so the economy tends to decline. The opposite also applies."Deficit" here refers to government funds used to reduce the difference between revenues and budget.

In this report we are comparing the government spending or expenditure for countries in North America i.e. **Canada,Mexico and United States of America**.Among all the North American countries, Canada is considered to be the freest economy.This means that Canada is most likely to survive and prosper in comparison to other countries during times of social and fiscal chaos. Two decades ago Canada suffered a deep recession and teetered on the brink of a debt crisis caused by rising government spending but with cut in spending and balanced budget,Canadian economy boomed. Mexico is ranked 3rd out of three countries in the North America region.Though Mexico's economy is rising,it was one of the Latin American nations most affected by the 2008 recession with its Gross Domestic Product contracting by more than 6%. USA is the 12th freeset country in the world.

2 Result Section

Below is the actual code executed to create the data table and the graph for our analysis.For this assignment, the required data has been fetched from Quandl as mentioned in the assignment.The assignment has 2 parts for analysis:

1. Analysis by change in Government Spending
2. Analysis by Level of Government Spending

2.1 Analysis by Change

For this part, we fetched the data for the three North American countries namely Canada,Mexico and USA through Quandl API and put them into different data frames.These data frames were then merged in pairs to get the final data frame containing data for all the three countries.This is done using merge command. Before starting with the code, it is necessary to install all the required packages.

```
> library(XML)
> library(ggplot2)
> library(reshape2)
> library(plyr)
> library(Quandl)
```

Get the tables on the page for analysis based on the change from Quandl. Read data for Canada from Quandl.For this we directly take the R code for dataset from Quandl and store it in a dataframe named "first" in R.Here the authcode is the unique userID for a user and `"ODA/CAN_GGX"` is the code for the dataset in Quandl.Now since the dataset contains many rows,we will select only first 10 rows for analysis.The command head is used to display the table.The resulting table shows the data by change for government spending for Canada for 10 consecutive years.

```
> first<-Quandl("ODA/CAN_GGX", trim_start="1980-12-31", trim_end="2019-12-31"
+ ,transformation ="diff", authcode="K5uN_2KMV3bSt35DHxxM")
> Canada<-first[1:10,]
> head(Canada,10)
```

```
         Date  Value
1  2019-12-31 40.298
2  2018-12-31 37.927
3  2017-12-31 36.019
4  2016-12-31 35.574
5  2015-12-31 30.948
6  2014-12-31 27.433
7  2013-12-31 20.166
8  2012-12-31  9.412
9  2011-12-31 19.882
10 2010-12-31 46.449
```

Similarly, we will read data for Mexico and USA from Quandl and store them in different dataframes.The resulting tables show the data by change for government spending for Mexico and USA.

```
> second<-Quandl("ODA/MEX_GGX", trim_start="1990-12-31", trim_end="2019-12-31"
+ , transformation="diff", authcode="K5uN_2KMV3bSt35DHxxM")
> Mexico<-second[1:10,]
> head(Mexico,10)
```

```
         Date   Value
1  2019-12-31 376.308
2  2018-12-31 388.613
3  2017-12-31 274.840
4  2016-12-31 156.228
5  2015-12-31 238.300
6  2014-12-31 325.475
7  2013-12-31 137.794
8  2012-12-31 415.035
9  2011-12-31 271.109
10 2010-12-31 258.523
```

```
> third<-Quandl("ODA/USA_GGX", trim_start="1980-12-31", trim_end="2019-12-31"
+ , transformation="diff", authcode="K5uN_2KMV3bSt35DHxxM")
> USA<-third[1:10,]
> head(USA,10)
```

```
         Date   Value
1  2019-12-31 381.048
2  2018-12-31 338.801
3  2017-12-31 307.934
4  2016-12-31 338.511
5  2015-12-31 309.512
6  2014-12-31 161.835
7  2013-12-31 102.269
8  2012-12-31  63.086
9  2011-12-31  46.755
10 2010-12-31 -37.553
```

As of now we have 3 dataframes, each for every country in North America but this not what we want.We want a dataset containing data for all the three countries.To get the combined dataframe, we will merge these dataframes 2 at a time to get the final dataset using merge command. Here merging Canada and Mexico to get Merge1 using merge command.The command for merge contains "by" parameter which contains the field used to match both the dataframes.Here this parameter is "Date" which is a common field in Canada and Mexico.

```
> merge1<-merge(x=Canada,y=Mexico,by=c("Date"))
> head(merge1)
```

```
        Date Value.x Value.y
1 2010-12-31  46.449 258.523
2 2011-12-31  19.882 271.109
3 2012-12-31   9.412 415.035
4 2013-12-31  20.166 137.794
5 2014-12-31  27.433 325.475
6 2015-12-31  30.948 238.300
```

Similarly,here merging Mexico and USA to get Merge2.Same as merge1,"by" parameter is "Date" which is a common field in Mexico and USA.

```
> merge2<-merge(x=Mexico,y=USA,by=c("Date"))
> head(merge2)

        Date Value.x Value.y
1 2010-12-31 258.523 -37.553
2 2011-12-31 271.109  46.755
3 2012-12-31 415.035  63.086
4 2013-12-31 137.794 102.269
5 2014-12-31 325.475 161.835
6 2015-12-31 238.300 309.512
```

Finally merging Merge1 and Merge2 to get a dataframe finalmerge containing data for North American countries.

```
> finalmerge<-merge(x=merge1,y=merge2,by=c("Date"))
> head(finalmerge)

        Date Value.x.x Value.y.x Value.x.y Value.y.y
1 2010-12-31    46.449   258.523   258.523   -37.553
2 2011-12-31    19.882   271.109   271.109    46.755
3 2012-12-31     9.412   415.035   415.035    63.086
4 2013-12-31    20.166   137.794   137.794   102.269
5 2014-12-31    27.433   325.475   325.475   161.835
6 2015-12-31    30.948   238.300   238.300   309.512
```

In the above displayed table we have four columns,one for Canada and USA each and two for Mexico.Hence we need to remove this extra unwanted column for Mexico.To do this we will use `as.data.frame()` command to remove the extra column(here it is 4th column). To refine the final table,we rename the columns for the final dataframe using rename command.For the rename command,install plyr package before using it.

```
> finalmerge<-rename(finalmerge,c("Value.x.x"="Canada","Value.y.x"="Mexico",
+ "Value.y.y"="USA"))
> finalmerge<-as.data.frame(finalmerge[,-4])
> head(finalmerge,10)
```

```
         Date Canada  Mexico     USA
1  2010-12-31 46.449 258.523 -37.553
2  2011-12-31 19.882 271.109  46.755
3  2012-12-31  9.412 415.035  63.086
4  2013-12-31 20.166 137.794 102.269
5  2014-12-31 27.433 325.475 161.835
6  2015-12-31 30.948 238.300 309.512
7  2016-12-31 35.574 156.228 338.511
8  2017-12-31 36.019 274.840 307.934
9  2018-12-31 37.927 388.613 338.801
10 2019-12-31 40.298 376.308 381.048
```

Now we create a new data frame df with the contents of finalmerge.Then df is replaced with the contents of Canada,Mexico and USA.For creating a graph we need our dataframe in long format but since current format of our dataframe is wide, we use melt function to convert wide format to long format.Since we donot want Date field in long format, hence we use id.vars=”Date”.

```
> df<-data.frame(finalmerge)
> df.m<-melt(df,id.vars=c("Date"))
> df<-data.frame(Canada,Mexico,USA)
```

Creating a bargraph to analyse the level of Government spending for North America using ggplot.Here we are using `position=position_dodge()` to display a bar for each country for the respective year.Here we have date on x-axis, value for change on y-axis and color for the bar outline. gglot2 package is required for creating a bar graph using ggplot.

```
> ggplot(df.m,aes(x=Date,y=value,fill=variable))+geom_bar(stat="identity"
+ ,position=position_dodge(),color="black")+
+ ggtitle( "Figure1:Graph showing analysis by Change")
```

Figure1:Graph showing analysis by Change

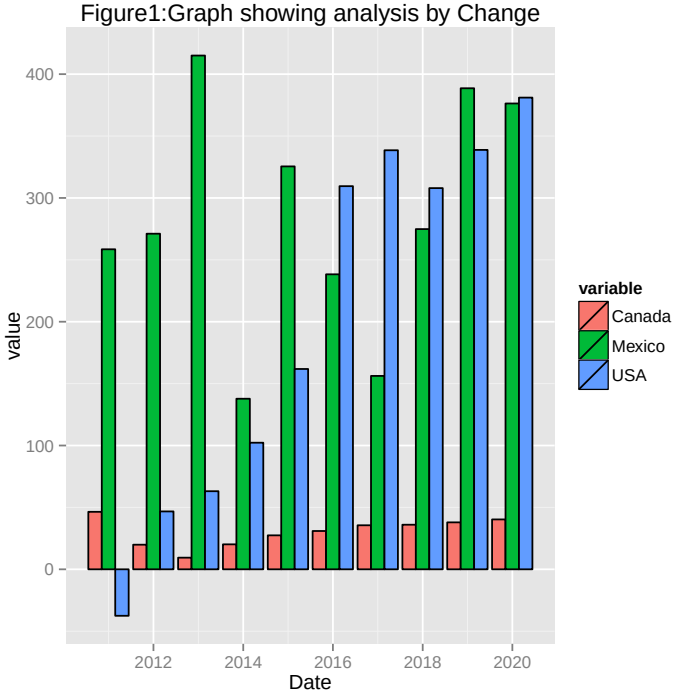

Figure1. shows a graph for government spending by change.The change in government spending for every country is not linear, for some years there is a decrease and for other years there is an increase in the government spending.For the year 2010 where there was a positive change or increase in the government spending for Canada and Mexico,USA had a negative change or decrease in the expenditure.This was the year of recession.

2.2 Analysis by Level

For the second part,we fetched the data from Quandl API for the countries and cleaned it to get the desired dataset.Similar steps are followed as mentioned in the part for analysis by change.

Get the table on the page for analysis based on the level in government spending.We have downloaded the csv file from Quandl API and then importing it into R using read command.

```
> bylevel<-read.csv(file="multisets.csv")
> bylevel<-rename(bylevel,c("ODA.USA_GGX...Value"="USA",
+  "ODA.CAN_GGX...Value"="Canada","ODA.MEX_GGX...Value"="Mexico"))
```

Cleaning the data to get only the first 10 rows for the table.Same steps are followed as followed for analysis by change.

```
> bylevel<-as.data.frame(bylevel[1:10,])
> head(bylevel,10)
```

```
         Date      USA   Canada   Mexico
1  31/12/2019 8229.420 1044.415 6131.039
2  31/12/2018 7848.372 1004.117 5754.731
3  31/12/2017 7509.571  966.190 5366.118
4  31/12/2016 7201.637  930.171 5091.278
5  31/12/2015 6863.126  894.597 4935.050
6  31/12/2014 6553.614  863.649 4696.750
7  31/12/2013 6391.779  836.216 4371.275
8  31/12/2012 6289.510  816.050 4233.481
9  31/12/2011 6226.424  806.638 3818.446
10 31/12/2010 6179.669  786.756 3547.337
```

```
> bylevel.new<-data.frame(bylevel)
> bylevel.m<-melt(bylevel.new)
```

Creating a bargraph to analyse the change in Government spending for North America using ggplot same as mentioned in analysis by change.

```
> ggplot(bylevel.m, aes(x=Date, y=value,fill=variable)) +geom_bar(stat="identity"
+ ,position=position_dodge(),color="black")+
+   ggtitle( "Figure2:Graph showing analysis by level")
```

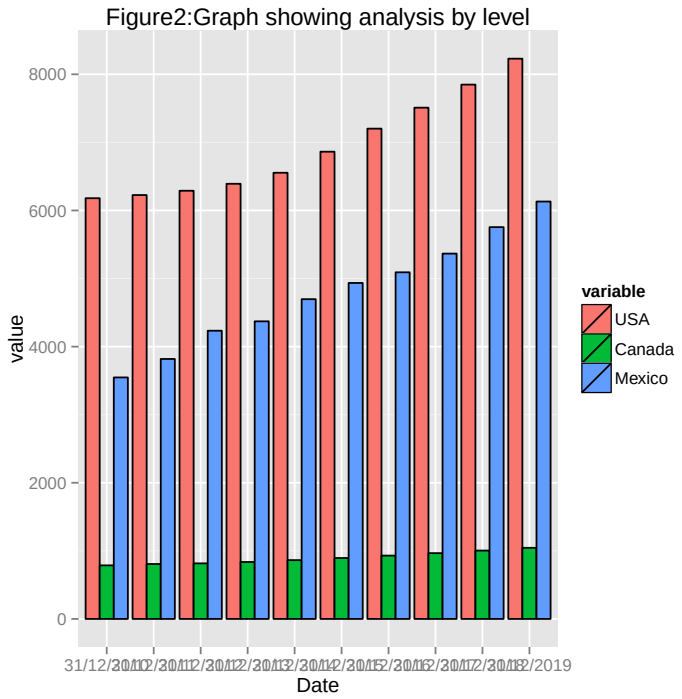

Figure2. shows a graph for Government spending by Level.It shows the difference in the level of spending by the government of a particular country.As

Indicated by the graph, the level of expenditure has increased for every country over the years.However,the government spending for Canada is much less compared to the other two countries.While the level of expenditure for Canada is in the range of 800 for the year 2014, USA and Mexico has it in the range more than 2000.

3 Data Section

Although we have explored the data in the code section, this section specifically explores every dataset used.Exploring data means to identify or viewing the internal structure of the data set.

- str() shows the number of variables,name of the variables,number of records and other details about the dataframe.
- summary() gives a summary of the dataframe.It shows the minimum and maximum value for every column or variable in the dataframe,the median etc..
- class() gives the data type of the variable, whether it is a dataframe or a number or a character.

3.1 Analysis by Change

For the first part i.e. Analysis by change, we have 3 data frames for each country and a final data frame containing the desired data.

```
> str(Canada)

'data.frame':        10 obs. of  2 variables:
 $ Date : Date, format: "2019-12-31" "2018-12-31" ...
 $ Value: num  40.3 37.9 36 35.6 30.9 ...

> summary(Canada)

      Date                Value
 Min.   :2010-12-31   Min.   : 9.412
 1st Qu.:2013-04-01   1st Qu.:21.983
 Median :2015-07-01   Median :33.261
 Mean   :2015-07-01   Mean   :30.411
 3rd Qu.:2017-09-30   3rd Qu.:37.450
 Max.   :2019-12-31   Max.   :46.449

> class(Canada)

[1] "data.frame"

> str(Mexico)
```

```
'data.frame':        10 obs. of  2 variables:
 $ Date : Date, format: "2019-12-31" "2018-12-31" ...
 $ Value: num  376 389 275 156 238 ...
```

```
> summary(Mexico)
```

```
      Date                Value
 Min.   :2010-12-31   Min.   :137.8
 1st Qu.:2013-04-01   1st Qu.:243.4
 Median :2015-07-01   Median :273.0
 Mean   :2015-07-01   Mean   :284.2
 3rd Qu.:2017-09-30   3rd Qu.:363.6
 Max.   :2019-12-31   Max.   :415.0
```

```
> class(Mexico)
```

```
[1] "data.frame"
```

```
> str(USA)
```

```
'data.frame':        10 obs. of  2 variables:
 $ Date : Date, format: "2019-12-31" "2018-12-31" ...
 $ Value: num  381 339 308 339 310 ...
```

```
> summary(USA)
```

```
      Date                Value
 Min.   :2010-12-31   Min.   :-37.55
 1st Qu.:2013-04-01   1st Qu.: 72.88
 Median :2015-07-01   Median :234.88
 Mean   :2015-07-01   Mean   :201.22
 3rd Qu.:2017-09-30   3rd Qu.:331.26
 Max.   :2019-12-31   Max.   :381.05
```

```
> class(USA)
```

```
[1] "data.frame"
```

```
> str(finalmerge)
```

```
'data.frame':        10 obs. of  4 variables:
 $ Date  : Date, format: "2010-12-31" "2011-12-31" ...
 $ Canada: num  46.45 19.88 9.41 20.17 27.43 ...
 $ Mexico: num  259 271 415 138 325 ...
 $ USA   : num  -37.6 46.8 63.1 102.3 161.8 ...
```

```
> summary(finalmerge)
```

```
      Date                Canada           Mexico           USA
 Min.   :2010-12-31   Min.   : 9.412   Min.   :137.8   Min.   :-37.55
 1st Qu.:2013-04-01   1st Qu.:21.983   1st Qu.:243.4   1st Qu.: 72.88
 Median :2015-07-01   Median :33.261   Median :273.0   Median :234.88
 Mean   :2015-07-01   Mean   :30.411   Mean   :284.2   Mean   :201.22
 3rd Qu.:2017-09-30   3rd Qu.:37.450   3rd Qu.:363.6   3rd Qu.:331.26
 Max.   :2019-12-31   Max.   :46.449   Max.   :415.0   Max.   :381.05
```

```
> class(finalmerge)

[1] "data.frame"

>
```

3.2 Analysis by Level

For the second part, we are executing the same functions only on one variable i.e. on bylevel.

```
> str(bylevel)

'data.frame':        10 obs. of  4 variables:
 $ Date  : Factor w/ 40 levels "31/12/1980","31/12/1981",..: 40 39 38 37 36 35 34 33 32 31
 $ USA   : num  8229 7848 7510 7202 6863 ...
 $ Canada: num  1044 1004 966 930 895 ...
 $ Mexico: num  6131 5755 5366 5091 4935 ...
```

```
> summary(bylevel)

         Date        USA           Canada           Mexico
 31/12/2010:1   Min.   :6180   Min.   : 786.8   Min.   :3547
 31/12/2011:1   1st Qu.:6315   1st Qu.: 821.1   1st Qu.:4268
 31/12/2012:1   Median :6708   Median : 879.1   Median :4816
 31/12/2013:1   Mean   :6929   Mean   : 894.9   Mean   :4795
 31/12/2014:1   3rd Qu.:7433   3rd Qu.: 957.2   3rd Qu.:5297
 31/12/2015:1   Max.   :8229   Max.   :1044.4   Max.   :6131
 (Other)   :4
```

```
> class(bylevel)

[1] "data.frame"
```