

Government Spending - North America

Mr. Ruturaj Jayawant
School of Information Technology
Illinois State University
rsjayaw@ilstu.edu

October 22, 2014

1 Introduction

This discussion would analyze the government spending in the countries of North America. For our discussion, we will be analyzing the following three countries in the North America: United States, Canada and Mexico. Government Spending, in simple words, can be termed as the inclusion of government consumption, investments and transfer payments. Every country prepares an annual budget which consists of the funds for various departments who would look for the well-being of the nation. This budget depends on the revenue growth of the country as well as its foreign relations.

2 Tutorial

2.1 Data Fetching and Data Cleaning (For Bar Graph by Level)

Firstly, we need to install certain packages so that we can use certain functions. For our analysis, we will require the following three packages: *Quandl*, *ggplot2*, and *reshape2*. After importing the above three libraries, we can use various functions such as *ggplot()*, *Quandl()*, *melt()*, etc.

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

The code chunk below would fetch the data with the help of Quandl API. Here **SpendingDataUSA**, **SpendingDataCanada** and **SpendingDataMexico** dataframes will be initialized to the values fetched from the Quandl API for USA, Canada and Mexico respectively. The **trim_start** and **trim_end** would define the range of the data. Lastly, **authcode** is optional, which would authenticate the user to the Quandl website. We can register on Quandl Website and obtain a unique **authcode**.

```

> SpendingDataUSA <- Quandl("ODA/USA_GGX",
+   trim_start = "1980-12-31", trim_end = "2019-12-31",
+   authcode = "pRKG3j-kt3Cs9qBer1sD")
> SpendingDataCanada <- Quandl("ODA/CAN_GGX",
+   trim_start = "1980-12-31", trim_end = "2019-12-31",
+   authcode = "pRKG3j-kt3Cs9qBer1sD")
> SpendingDataMexico <- Quandl("ODA/MEX_GGX",
+   trim_start = "1990-12-31", trim_end = "2019-12-31",
+   authcode = "pRKG3j-kt3Cs9qBer1sD")

```

The code chunk below demonstrates the use of *plyr()* command, which is used to change the name of the column in a dataframe. In this case, this command changes the column name from **Value** to **SpendingDataUSA** in the **SpendingDataUSA** dataframe. Similarly, the *plyr()* function can be used on the **SpendingDataCanada** and **SpendingDataMexico** dataframes. In the *plyr* command, we use `replace = c("Value" = "USA")` to specify which column's name would be altered. Also, `x = SpendingDataUSA` is used to indicate the dataframe for which the column name needs to be altered.

```

> SpendingDataUSA <- plyr :: rename(x = SpendingDataUSA,
+   replace=c("Value" = "USA"))
> SpendingDataCanada <- plyr :: rename(x = SpendingDataCanada,
+   replace=c("Value" = "Canada"))
> SpendingDataMexico <- plyr :: rename(x = SpendingDataMexico,
+   replace = c("Value" = "Mexico"))

```

Here, we intend to gather the government spending of USA, Canada and Mexico into a single dataframe **SpendingData**. Hence, we use the *merge()* function. We use the *merge()* function twice. During its first use, we bring the data from the **SpendingDataUSA** and **SpendingDataCanada** in the **SpendingData** dataframe. During the second use of *merge()*, we bring the data from the **SpendingData** and **SpendingDataMexico** in the **SpendingData** dataframe.

```

> SpendingData <- merge(SpendingDataUSA, SpendingDataCanada,
+   by.x = "Date", by.y = "Date", all.x = T, all.y = F)
> SpendingData <- merge(SpendingData, SpendingDataMexico,
+   by.x = "Date", by.y = "Date", all.x = T, all.y = F)

```

Now, we are completed with data cleaning. *write.csv()* function is used to write the data from the dataframe into a CSV file. Here, `rows.names` is initialized to `FALSE` so that the rows in the CSV file would not be indexed.

```

> write.csv(SpendingData, "SpendingDataNorthAmerica.csv",
+   row.names = FALSE)

```

2.2 Data

The **SpendingData** consists the data extracted with the help of Quandl API.

```
> SpendingData
```

	Date	USA	Canada	Mexico
1	1980-12-31	946.700	151.632	NA
2	1981-12-31	1072.000	176.785	NA
3	1982-12-31	1191.300	205.814	NA
4	1983-12-31	1299.500	224.011	NA
5	1984-12-31	1409.000	243.886	NA
6	1985-12-31	1541.900	266.068	NA
7	1986-12-31	1648.900	279.113	NA
8	1987-12-31	1741.400	296.263	NA
9	1988-12-31	1832.600	319.266	NA
10	1989-12-31	1968.500	333.214	NA
11	1990-12-31	2143.200	366.559	206.968
12	1991-12-31	2298.300	398.434	229.524
13	1992-12-31	2431.600	415.673	253.790
14	1993-12-31	2520.600	422.405	282.398
15	1994-12-31	2613.200	426.646	324.101
16	1995-12-31	2732.700	437.198	490.989
17	1996-12-31	2848.400	428.947	713.951
18	1997-12-31	2935.400	422.250	911.513
19	1998-12-31	3026.000	443.370	1025.293
20	1999-12-31	3177.600	460.855	1241.348
21	2000-12-31	3353.500	479.714	1351.317
22	2001-12-31	3593.670	504.105	1438.197
23	2002-12-31	3825.912	512.469	1583.758
24	2003-12-31	4074.760	542.680	1731.262
25	2004-12-31	4309.797	559.141	1764.827
26	2005-12-31	4611.184	600.532	2043.862
27	2006-12-31	4846.382	627.568	2376.790
28	2007-12-31	5168.492	659.564	2601.368
29	2008-12-31	5591.564	703.481	3140.707
30	2009-12-31	6217.222	740.307	3288.814
31	2010-12-31	6179.669	786.756	3547.337
32	2011-12-31	6226.424	806.638	3818.446
33	2012-12-31	6289.510	816.050	4233.481
34	2013-12-31	6391.779	836.216	4371.275
35	2014-12-31	6553.614	863.649	4696.750
36	2015-12-31	6863.126	894.597	4935.050
37	2016-12-31	7201.637	930.171	5091.278
38	2017-12-31	7509.571	966.190	5366.118
39	2018-12-31	7848.372	1004.117	5754.731
40	2019-12-31	8229.420	1044.415	6131.039

In the above dataframe, the first column **Date**, consists of the range of dates as specified in the Quandl API. The second column **USA**, consists the government spending by USA for the corresponding dates. The third column **Canada**, consists the government spending by Canada for the corresponding dates. The final column **Mexico**, consists the government spending by Mexico for the corresponding dates.

As we have the data cleaned, we will need to determine the class of the object of our dataframe **SpendingData**. This would be done with the help of *class()* function. The *class()* function is used in the following way:

```
> class(SpendingData)
```

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

Further, we use the *str()* function to view the internal structure of the objects in the **SpendingData** dataframe.

```
> str(SpendingData)
```

```
'data.frame':      40 obs. of  4 variables:
 $ Date   : Date, format: "1980-12-31" "1981-12-31" ...
 $ USA    : num  947 1072 1191 1300 1409 ...
 $ Canada: num  152 177 206 224 244 ...
 $ Mexico: num   NA NA NA NA NA NA NA NA NA ...
```

Lastly, we use the *summary()* function to understand the statistics of the data. The *summary()* function computes the mean, median, minimum and maximum values in each column.

```
> summary(SpendingData)
```

Date	USA	Canada	Mexico
Min. :1980-12-31	Min. : 946.7	Min. : 151.6	Min. : 207
1st Qu.:1990-09-30	1st Qu.:2099.5	1st Qu.: 358.2	1st Qu.: 940
Median :2000-07-01	Median :3265.6	Median : 470.3	Median :1904
Mean :2000-07-01	Mean :3906.6	Mean : 539.9	Mean :2498
3rd Qu.:2010-04-01	3rd Qu.:6189.1	3rd Qu.: 751.9	3rd Qu.:4130
Max. :2019-12-31	Max. :8229.4	Max. :1044.4	Max. :6131
			NA's :10

2.3 Reshaping the Data (For Bar Graph by Level)

The **SpendingData** dataframe consists the data in the wide format. To plot a bar graph of the **SpendingData** dataframe, we will require the dataframe in long format. Hence, to transform the data into long format, we use the `melt()`. Here, the columns can be merged with the help of `measure.vars = c("col1", "col2", "col3"...)`.

```
> data <- melt(SpendingData, id.vars = "Date",
+   measure.vars = c("USA", "Canada", "Mexico"))
> data <- plyr :: rename(x = data,
+   replace=c("variable" = "Country", "value" = "Money_Spent"))
```

Now the **data** dataframe is ready to be plotted and thus, Government Spending(By Level) graph can be drawn for USA, Canada and Mexico.

2.4 Data Fetching and Data Cleaning (For Bar Graph by Change)

Now, we will need to compute the values for the **ChangeUSA**, **ChangeCanada** and **ChangeMexico** to draw the graph for Government Spending(North America) - By Change. Firstly, we will create a new dataframe **data1** and initialize it to **SpendingData**. Then we add columns **ChangeUSA**, **ChangeCanada**, and **ChangeMexico** to the **data1** dataframe by initializing these columns to 0. Lastly, we compute the values of the **ChangeUSA** column by the following formula: **ChangeUSA[current] = USA[next] - USA[current]**.

This formula is written in the for loop ranging from 1 to 39. Similarly, we can compute the values for **ChangeCanada** and **ChangeMexico** columns.

```
> data1 <- SpendingData
> data1$ChangeUSA <- 0
> for (i in 1:39 )
+ {
+   data1$ChangeUSA[i] <- data1$USA[i+1] - data1$USA[i]
+ }
> data1$ChangeCanada <- 0
> for (i in 1:39 )
+ {
+   data1$ChangeCanada[i] <- data1$Canada[i+1] - data1$Canada[i]
+ }
> data1$ChangeMexico <- 0
> for (i in 1:39 )
+ {
+   data1$ChangeMexico[i] <- data1$Mexico[i+1] - data1$Mexico[i]
+ }
```

Lastly, we do not require the **USA**, **Canada**, and **Mexico** columns. Hence, we initialize these columns to **NULL**.

```
> data1$USA <- NULL
> data1$Canada <- NULL
> data1$Mexico <- NULL
```

Thus, we have the cleaned data in the **data1** dataframe.

2.5 Reshaping the Data (For Bar Graph by Change)

The **data1** dataframe consists the data in the wide format. To plot a bar graph of the **data1** dataframe, we will require the dataframe in long format. Hence, to transform the data into long format, we use the *melt()*. Here, the columns can be merged with the help of `measure.vars = c("col1", "col2", "col3"...)`.

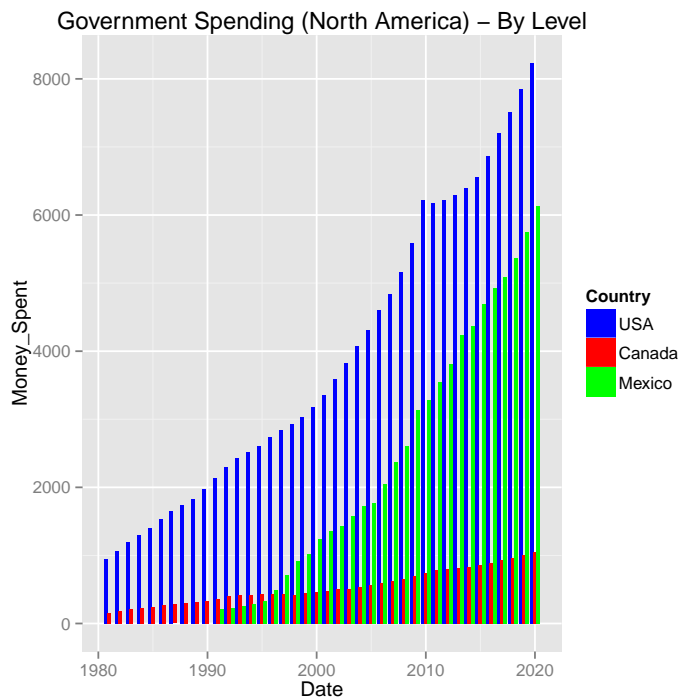
```
> data2 <- melt(data1, id.vars = "Date",
+   measure.vars = c("ChangeUSA", "ChangeCanada", "ChangeMexico"))
> data2 <- plyr :: rename(x = data2,
+   replace=c("variable" = "Country",
+   "value" = "Change_Money_Spent"))
```

Now the **data2** dataframe is ready to be plotted and thus, Government Spending(By Change) graph can be drawn for USA, Canada and Mexico.

3 Results

In the code chunk below, we have the Date along the X-axis and Money_Spent along the Y-axis. We fill the graph with the values in the Country column. `ggplot()` is used to draw a graph which takes arguments as dataframe and `aes()` which is used to define the axes of the graph. `ggtitle()` is used to add a title to the graph. Lastly, `scale_fill_manual()` is used to manually alter the colors used in the graph.

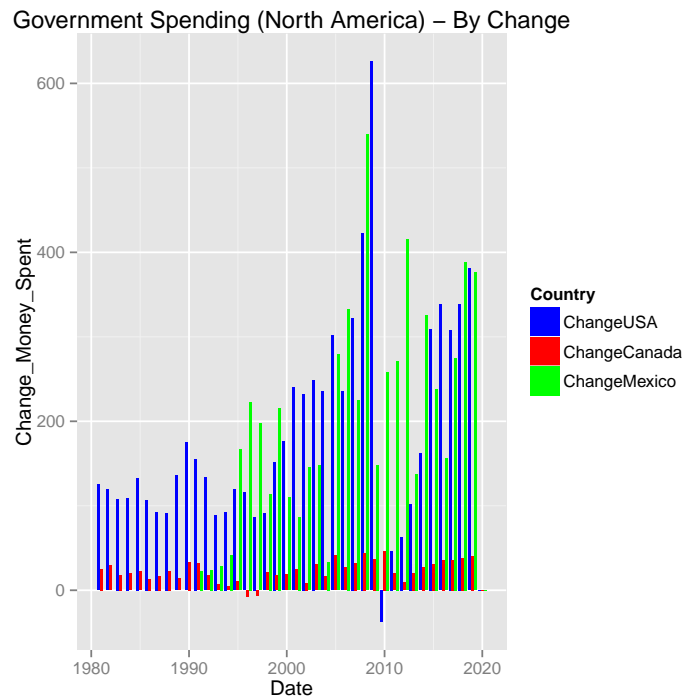
```
> ggplot(data = data, aes(x = Date, y = Money_Spent, fill = Country)) +  
+   geom_bar(stat = "identity", position = position_dodge()) +  
+   ggtitle("Government Spending (North America) - By Level") +  
+   scale_fill_manual(values=c("blue", "red", "green"))
```



The above graph shows the Government Spending in North America for USA, Canada and Mexico by level. In this graph, we notice that there has been a steady growth in government spending for the countries in North America since 1980.

In the code chunk below, we have the Date along the X-axis and Change_Money_Spent along the Y-axis. We fill the graph with the values in the Country column. `ggplot()` is used to draw a graph which takes arguments as dataframe and `aes()` which is used to define the axes of the graph. `ggtitle()` is used to add a title to the graph. Lastly, `scale_fill_manual()` is used to manually alter the colors used in the graph.

```
> ggplot(data = data2, aes(x = Date, y = Change_Money_Spent, fill = Country)) +
+   geom_bar(stat = "identity", position = position_dodge()) +
+   ggtitle("Government Spending (North America) - By Change") +
+   scale_fill_manual(values=c("blue", "red", "green"))
```



The graph above shows the Government Spending in North America for USA, Canada and Mexico by change. In this graph, we notice that there was a constant growth in the government spending in the North American countries. However, during 2009-10, it is seen that there has been a significant change in the government spending of USA.

4 Conclusion

We have analysed the government spending of the North American countries. Since 1990, there had been no significant change with the Mexican Government spending. However, in 2009, there was a significant rise in the government spending of USA. The reason for this will be credited to the Great Recession. Hence, the USA government implemented Stimulus Act, which intended to save and create jobs. The approximate cost of the Stimulus Package was around \$ 787 billion.

In Canada, during 1995-96, the Canadian government decided to reduce the economic budget for the country. This was achieved with cutting the nominal spending and public sector employment of the Canada. Hence, we notice a negative change for Canada in the year 1995 in the Government Spending (North America) - By Change graph.

5 References

1. http://en.wikipedia.org/wiki/American_Recovery_and_Reinvestment_Act_of_2009
2. <http://www.fraserinstitute.org/uploadedFiles/fraser-ca/Content/research-news/research/publications/BudgetBlueprint.pdf>
3. https://www.quandl.com/ODA/USA_GGX-United-States-General-Government-Total-Expenditure-USD-Billions
4. https://www.quandl.com/ODA/CAN_GGX-Canada-General-Government-Total-Expenditure-USD-Billions
5. https://www.quandl.com/ODA/MEX_GGX-Mexico-General-Government-Total-Expenditure-USD-Billions