## 

## 

## 

## 

## Colombia Foreign Assistance and World Bank Indicators Analysis

Authors: Jennifer Han, Emilio Ramos Monzalvo, Eric Brattin

Submission Date: March 18, 2020

Number of Pages: 31

## 

## 

## Background

##### **Domain Context**

Our team will be cleaning, analyzing, and then drawing conclusions regarding open source U.S government foreign assistance transaction data published online at [www.foreignassistance.gov](https://www.foreignassistance.gov/) by the U.S. Department of State. We will then connect FA.gov data to World Bank education data indicators (WBI) we refer to as outcomes using OECD DAC codes. The Organisation for Economic Co-operation and Development's Development Assistance Committee (DAC) has an official list of codes that relate to international development sectors they have developed along with its 36 member countries, which includes the US. The purpose of this master list of codes is for the development of official statistics and as a means to improve coordination in international policy.

FA data encompasses both budgetary data or data about requested funds for the future (appropriated funds) and financial data or obligated and spent data. For the purposes of our exercise we are looking solely at financial data both obligated and spent. Attempts to interlink these datasets do not suggest causation or correlation and are only intended to contribute to further progress towards accountability and transparency in international development. The foreign assistance fiscal years we will explore are divided and coupled into 3 dataset CSV files, of which our team will be required to clean to maintain efficiency given in total there are nearly 2 million rows of financial transaction data on FA.gov comprising 20 government agencies. Will will attempt to scale down these data sets to a combined 32,570 rows and 6 columns using the SQLDF feature in R Studio, which will limit our data to transactions that benefited Colombia.

Because the numerical measurements in the foreign assistance dataset are confined to dollars spent, we will also be attempting to draw larger conclusions based on World Bank Indicator data that tracks education indicators of development. We have chosen education indicators, but are not assuming aid directed towards security, humanitarian aid, or economic development doesn’t connect to potential progress in education sector development, given the complex, multi-dimensional nature of foreign aid. As we are unable to evaluate individual project performance data (PPR Data), we cannot conclude with authority as to why increases in aid in generic categories like "health" or "humanitarian assistance" would impact education outcomes. Although we see through our data analysis an unknown relationship developing between funding in specific programmatic areas and outcomes in another, the intricacies of these relationships will remain unexplored.

We have chosen Colombia as our country of choice based on the availability of data. We are intentionally avoiding connecting dollars spent in the security sector and social indicators such as crime rate and corruption, given the lack of specificity in aid data and to avoid the appearance of forcing conclusions we believe would be scientifically unproven and impossible to reproduce, given the project scope and time allotted.

We understand that effective foreign assistance is only partially dependent on the amount of money spent. Quality of programming, improved accountability, long-term direct commitments, local buy-in, flexibility of instruments (grants vs. loans) etc. that overtime consistently builds self-sufficient institutions in developing countries matters as well. As such, saying more children go to school in Colombia thanks to more funding ignores factors like quality, commitments, and program design. The data we have analyzed should continue to be used as an initial exploratory effort to help gauge how to build lasting institutions that consistently allow for outcomes to improve not because of a more self-reliant region, but because well-interpreted data helped all actors understand the importance of creative cooperation.

Our overall questions are how much aid was directed to Colombia from 2009-2019, followed by a breakdown based on sector, and a look at how educational outcomes in Colombia during this period were affected, if at all.

##### **Data Sets In Scope**

1. U.S Foreign Assistance Given to Colombia in Years 2009 - 2019 ([data here](https://www.foreignassistance.gov/downloads/TransactionData/FullDataset/ForeignAssistance-FullDataSet.zip))
2. World Bank Indicators for Colombia ([data here](http://api.worldbank.org/v2/en/country/COL?downloadformat=csv))

## 

## Data Cleansing

Foreign assistance data sets are massive files encompassing nearly 2 million rows. To make these files more manageable and more in-scope to our area of interest, we used SQLDF to SELECT only columns we are interested in using in our analysis such as Country, Agency, and Money Spent WHERE Country = Colombia, which constricted our view and made the task of analysis less cumbersome. RBIND was used to merge these data sets into one single point of interest.

See Code Snippets 1A to 1D.

## Data Overview Analysis

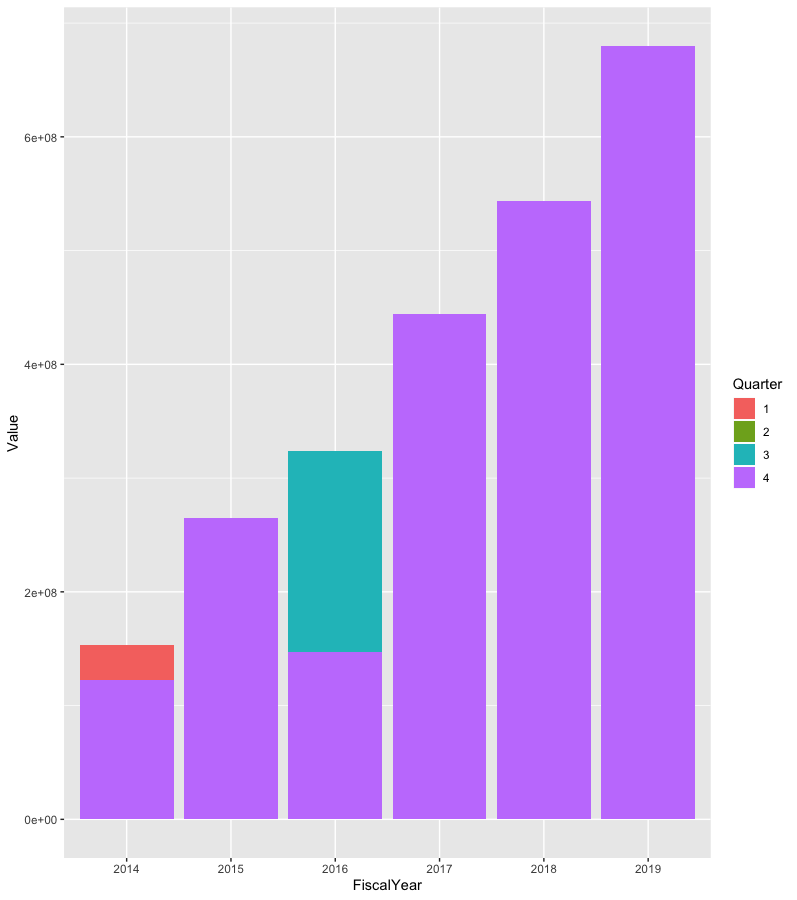
##### **Overall US Contribution to Colombia Over Time**

Over time, the US contribution to Colombia has changed drastically.We can easily see how the amount of money put into Colombia has increased.

See diagram below. See Code Snippet 2A.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| FiscalYear | Quarter1 | Quarter2 | Quarter3 | Quarter4 |
| 2014 | 153329487 | 96362915 | 100629677 | 122876094 |
| 2015 | 140247005 | 82125317 | 91148145 | 265034346 |
| 2016 | 158717162 | 131836846 | 323798533 | 147488797 |
| 2017 | 134701991 | 225690109 | 76204047 | 444152835 |
| 2018 | 208750189 | 144034969 | 222655805 | 543326807 |
| 2019 | 136596387 | 166499421 | 150147451 | 679876380 |

See diagram below. See Code Snippet 2B.



We can easily see how the contribution has gone up every year for the past 5 years. It is also important to note how most of the contribution is coming at the 4th fiscal quarter. This could be for a variety of reasons, but the most notable could be that the money is being given to Colombia once the US knows how much is left over at the end of the Fiscal Year. If we perform a linear model on the contribution over the years, we would need to aggregate the contribution per year as the number varies a lot over the quarters as shown below.

See diagram below. See Code Snippet 2H and 2I:

|  |  |
| --- | --- |
| Aggregate by Fiscal Quarter | Aggregate by Fiscal Year |
|  |  |
| Call:  lm(formula = Value ~ Time, data = dfAgg)  Residuals:  Min 1Q Median 3Q Max  -173433149 -78907593 -32042384 74762452 330709469  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 50577767 52931693 0.956 0.34969  Time 12441214 3704447 3.358 0.00284 \*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 125600000 on 22 degrees of freedom  Multiple R-squared: 0.3389, Adjusted R-squared: 0.3089  F-statistic: 11.28 on 1 and 22 DF, p-value: 0.002839 | Call:  lm(formula = Value ~ FiscalYear, data = dfAggYear)  Residuals:  Min 1Q Median 3Q Max  -220751136 -31033868 -3031250 56706561 164593670  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) -2.628e+11 2.076e+10 -12.66 4.87e-07 \*\*\*  FiscalYear 1.307e+08 1.031e+07 12.68 4.80e-07 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 108100000 on 9 degrees of freedom  Multiple R-squared: 0.947, Adjusted R-squared: 0.9411  F-statistic: 160.9 on 1 and 9 DF, p-value: 4.796e-07 |

With the information above, we easily see how the correlation of contribution over time is better represented by an aggregation of contribution over the Fiscal Year rather than separating it by Fiscal Quarter. The p-value returned from the linear model using the Fiscal Year is much smaller than the p-value from the model using the Fiscal Quarter as the predictor. This also proves that the second model makes more sense. When one looks at the Adjusted R-squared value of the second model we see how it is very close to 1 which means that the linear model very closely reflects the contribution to Colombia over time. Finally, the slope of the second linear model suggests that the contribution is increasing by $130,700,000 every year.

##### **Agencies Expenses Over Time**

There are a total of 15 US Agencies that contributed to Colombia in the past five years. Therefore instead of performing an individual analysis for each, we decided to focus on the top three Agencies who gave the most contributions to Colombia. Those agencies include International Development, Overseas Private Investment Corporation, and Department of State.

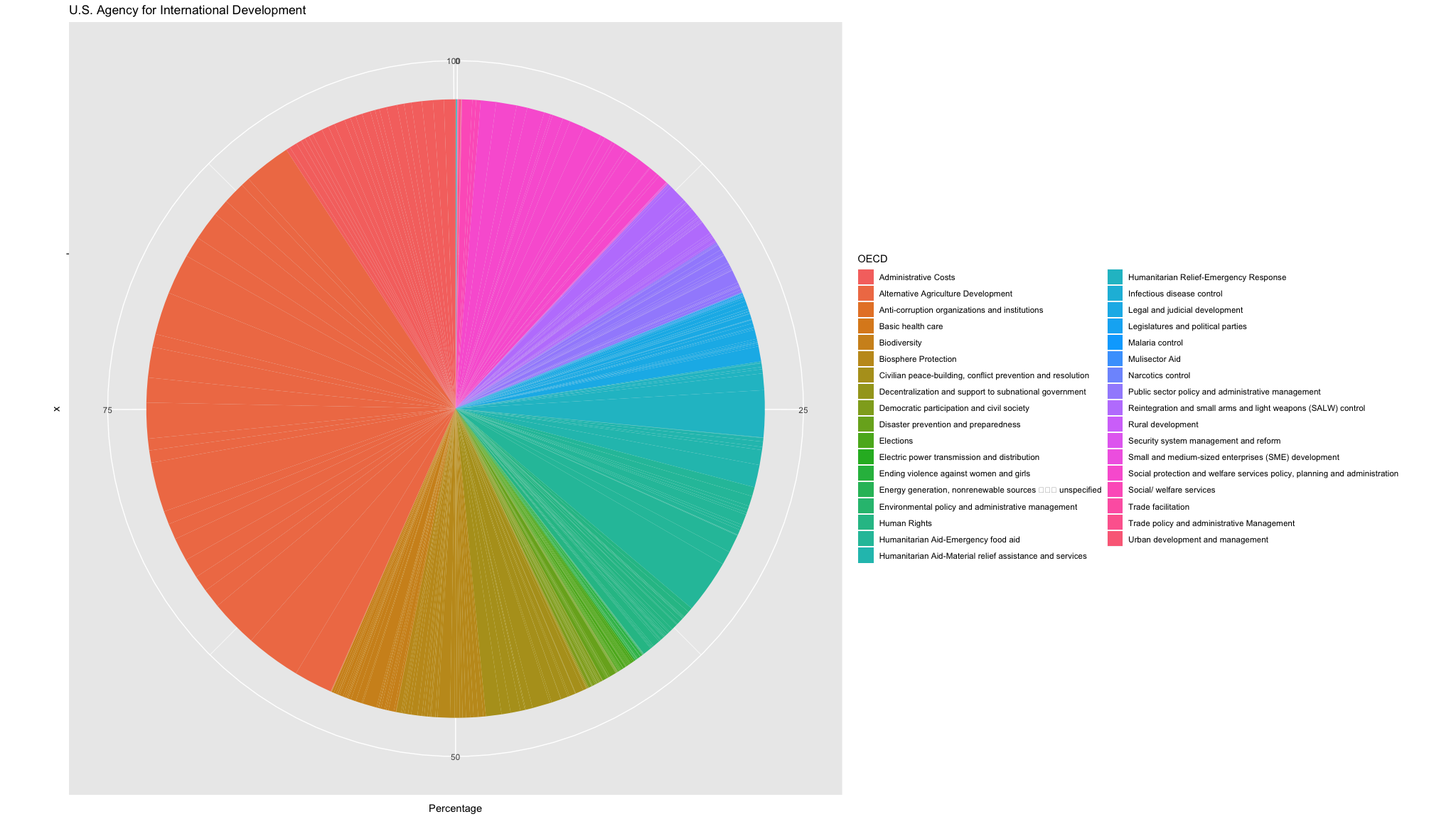
See diagram below. See Code Snippet 2E.

|  |  |
| --- | --- |
| Agency | Value |
| **U.S. Agency for International Development** | **2413172740** |
| **Overseas Private Investment Corporation** | **1174750000** |
| **Department of State** | **831931141.6** |
| Department of Defense | 284278530 |
| Department of Justice | 154001137.8 |
| Department of Labor | 31227597.8 |
| Peace Corps | 26136817.7 |
| Inter-American Foundation | 14919004.3 |
| U.S. Trade and Development Agency | 9923779.7 |
| Department of the Treasury | 4021602.7 |
| Department of the Interior | 892672.7 |
| Department of Homeland Security | 552047.7 |
| U.S. Department of Agriculture | 337644.3 |
| Department of Energy | 80000 |
| Department of Commerce | 6000 |

U.S. Agency for International Development

Out of these three agencies, the Agency for International Development contributed the most money to Colombia. Their main focus seemed to be the Alternative Agricultural Development as 34.23% of the contributions are going directly there. The other two significant contributions are going to Social protection and Administrative Costs with a contribution percentage of 10.65% and 9.23% respectively.

See diagram below. See Code Snippet 2F.

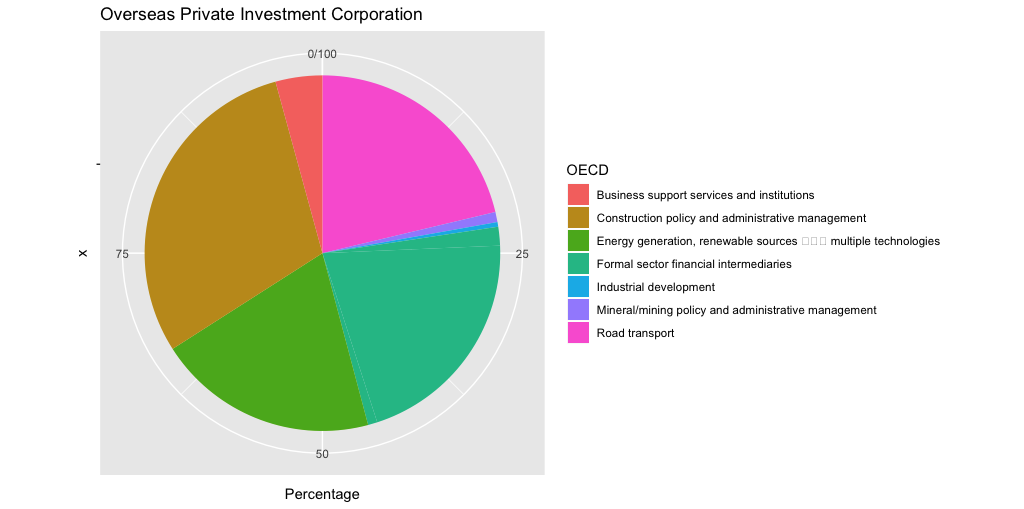


## 

Overseas Private Investment Corporations

The overseas private investment corporations include private sectors of the government that contribute to other companies in other countries.

See diagram below. See Code Snippet 2F.



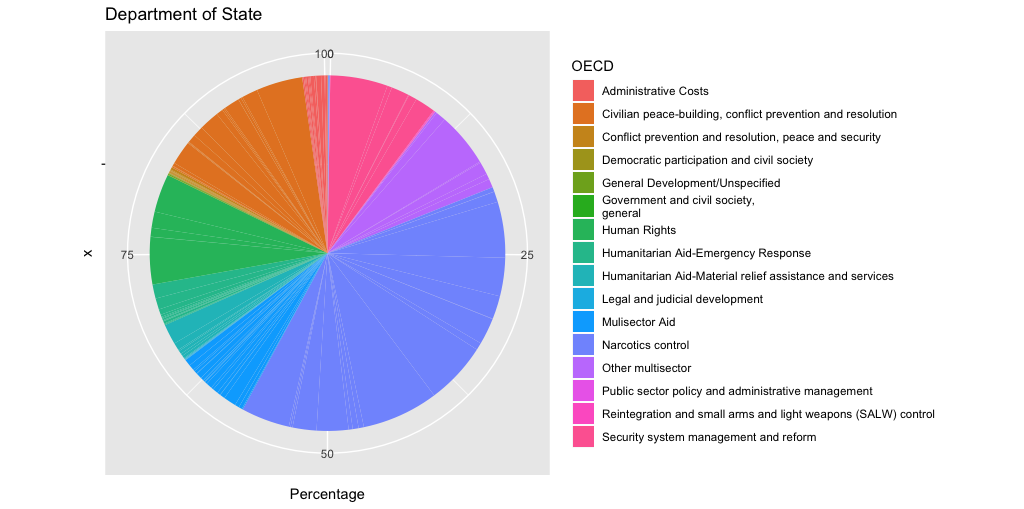


In this particular agency we see a more spread out contribution from the United States. There is no constant contribution to an individual OECD as the Agency for International Development contributes to Agricultural Development every year. The Private Investment Corporation has mostly picked out a few sectors in Colombia to improve business support, road transportation and energy generation. Even though there were a total of six contributions from the US to Colombia through the Private Investment Corporation, the total adds up to the second most significant Agency overall.

Department of State

The Department of State focuses on foreign policies. This agency focuses on protecting American interests overseas. Therefore, we expect a wide variety of contributions towards the different development programs across Colombia.

See diagram below. See Code Snippet 2F.



## 

## 

## 

## 

## 

## 

## 

## As stated above, this agency has a wide portfolio of contributions. The main end point of the contributions seems to be Narcotics Control with 39.02% of the total contributions. It is widely known that Colombia has a big market for Narcotics and it is not surprising that most of the Department of State’s contributions are ending up trying to control the distribution of Narcotics. The interesting part is that the contributions to Narcotics Control has only gone up after 2017 which indicates that the effect of Narcotics in the US has not gone down.

##### 

##### **Organization for Economic Cooperation and Development (OECD) Contributions**

The amount of types of organizations receiving contributions from the United States is overwhelming. The list ranges from Narcotics Control to Road Transportation and more. Out of all these Organizations for Economic Cooperation and Development (OECDs), the top three with the most contributions include Alternative Agricultural Development, Narcotics Control, and Construction policy and administrative management.

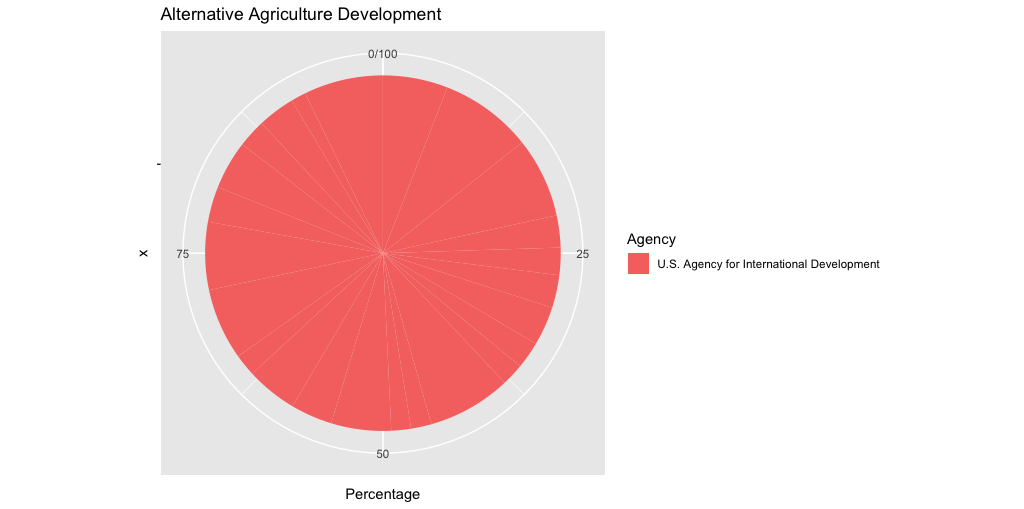
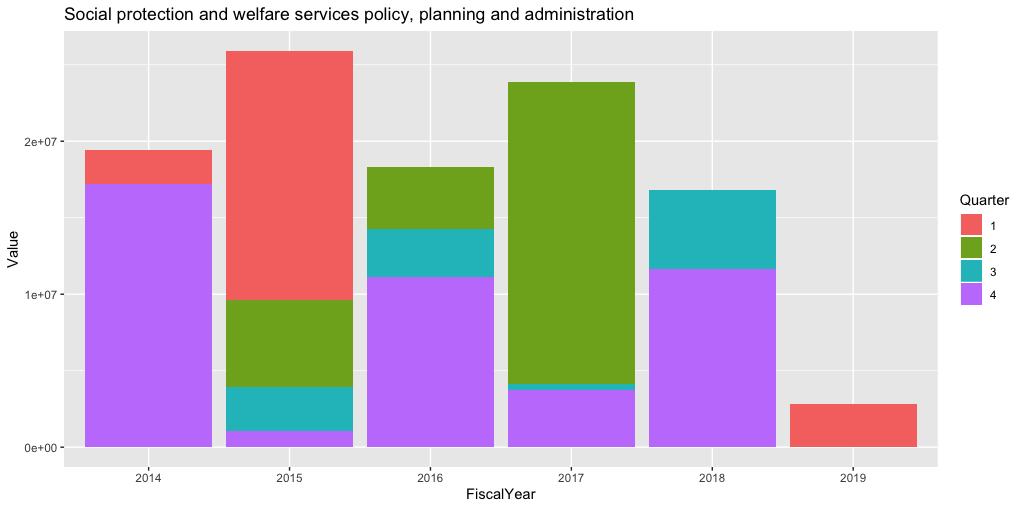
See diagram below. See Code Snippet 2C.

|  |  |
| --- | --- |
| OECD | Value |
| Alternative Agriculture Development | 825996233.9 |
| Narcotics control | 501593264.4 |
| Construction policy and administrative management | 350000000 |
| Formal sector financial intermediaries | 273000000 |
| Social protection and welfare services policy, planning and administration | 257043533.7 |
| Civilian peace-building, conflict prevention and resolution | 254276166.7 |
| Road transport | 250766400 |
| Administrative Costs | 242255380.3 |
| Energy generation, renewable sources /multiple technologies | 236344877.1 |
| Mulisector Aid | 210467959.2 |
| Humanitarian Aid-Emergency food aid | 171072384 |
| Human Rights | 166175302.9 |
| General Development/Unspecified | 114118194.2 |
| Biosphere Protection | 112282477.5 |
| Public sector policy and administrative management | 103614019.4 |
| Reintegration and small arms and light weapons (SALW) control | 95706242.42 |

Alternative Agricultural Development

This OECD can include the development of crops like corn and soybeans to import to the US from Colombia. The OECD tops all other ones in terms of contributions, but it seems that there is only one US agency with interest in this OECD. That agency being the Agency for International Development.

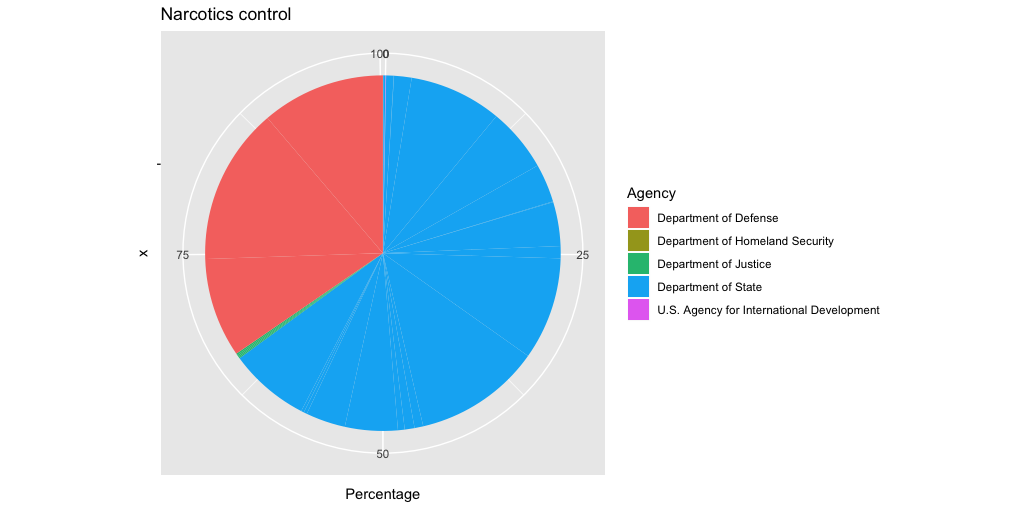
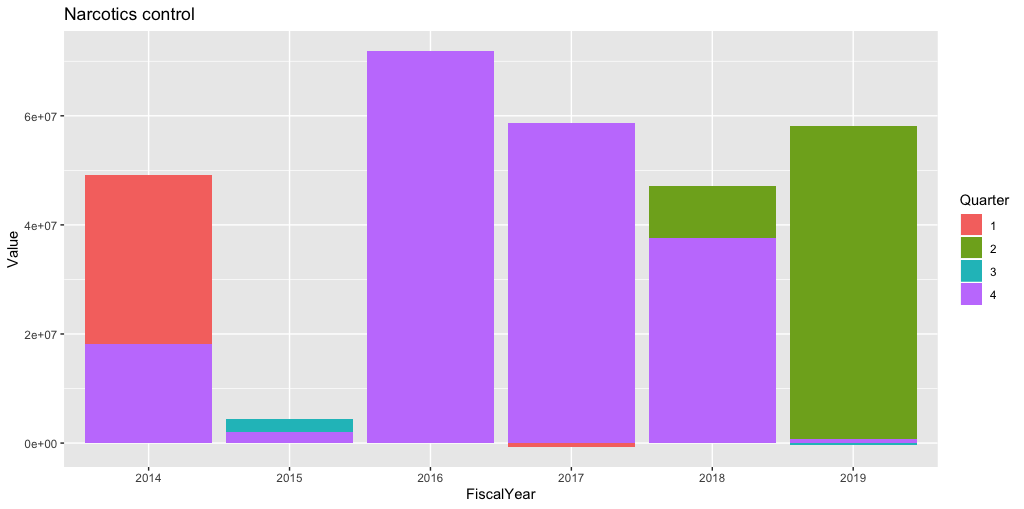
See diagram below. See Code Snippet 2D.



Narcotics Control

Narcotics Control is a popular choice for most agencies that have an interest in the protection of United States interests. That includes the Department of Defense, Homeland Security, State and Justice. All these grand agencies have made it their priority to stop the influx of drugs coming from Colombia with the Department of State having contributed more than 50% of the total contributions to the Narcotics Control Organization.

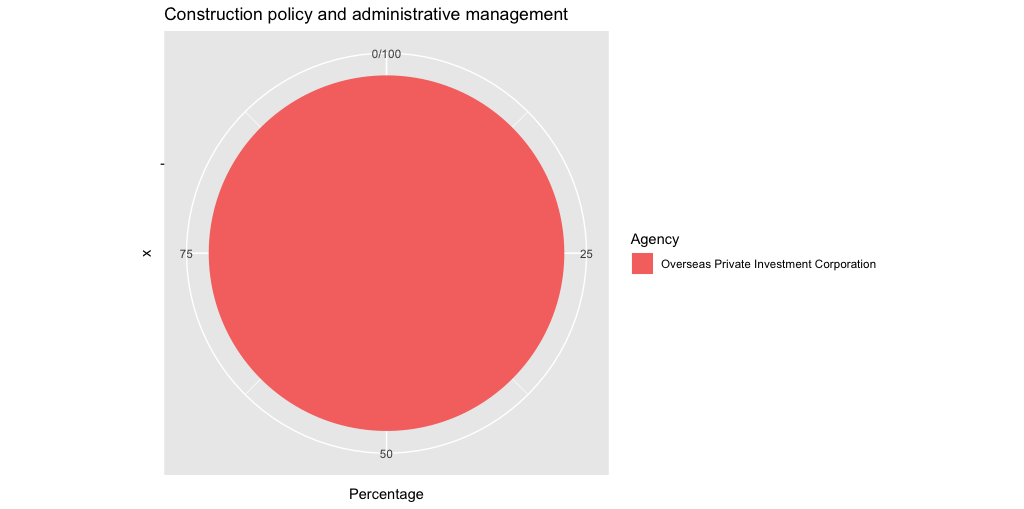
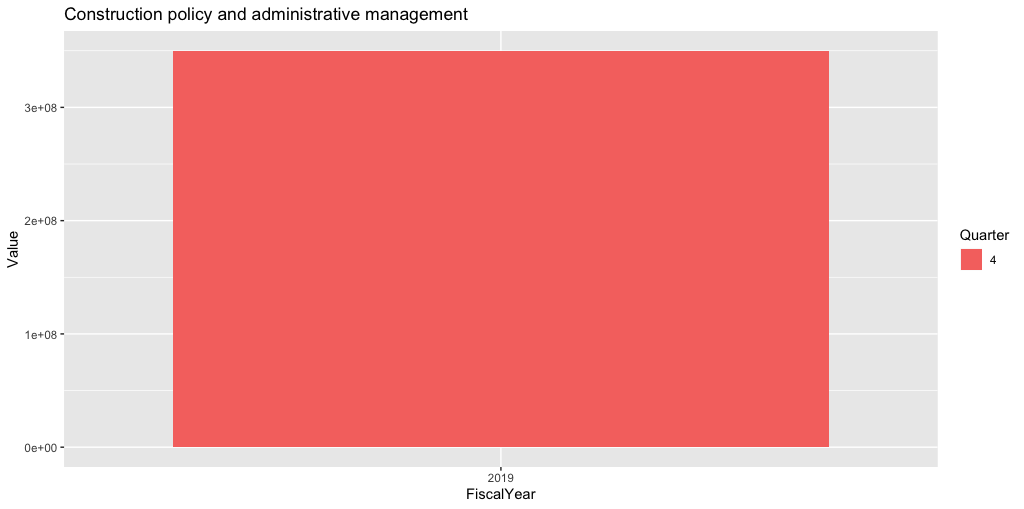
See diagrams below. See Code Snippet 2D.



Construction Policy and Administrative Management

This OECD concentration of the construction of policies in Colombia. This means that the US has certain interests in how Colombia might run the country and decides to interfere. Now this OECD only had one contribution from the Overseas Private Investment Corporation that is so significant that it made the OECD the third biggest OECD in the past five years.

See diagrams below. See Code Snippet 2D.



## 

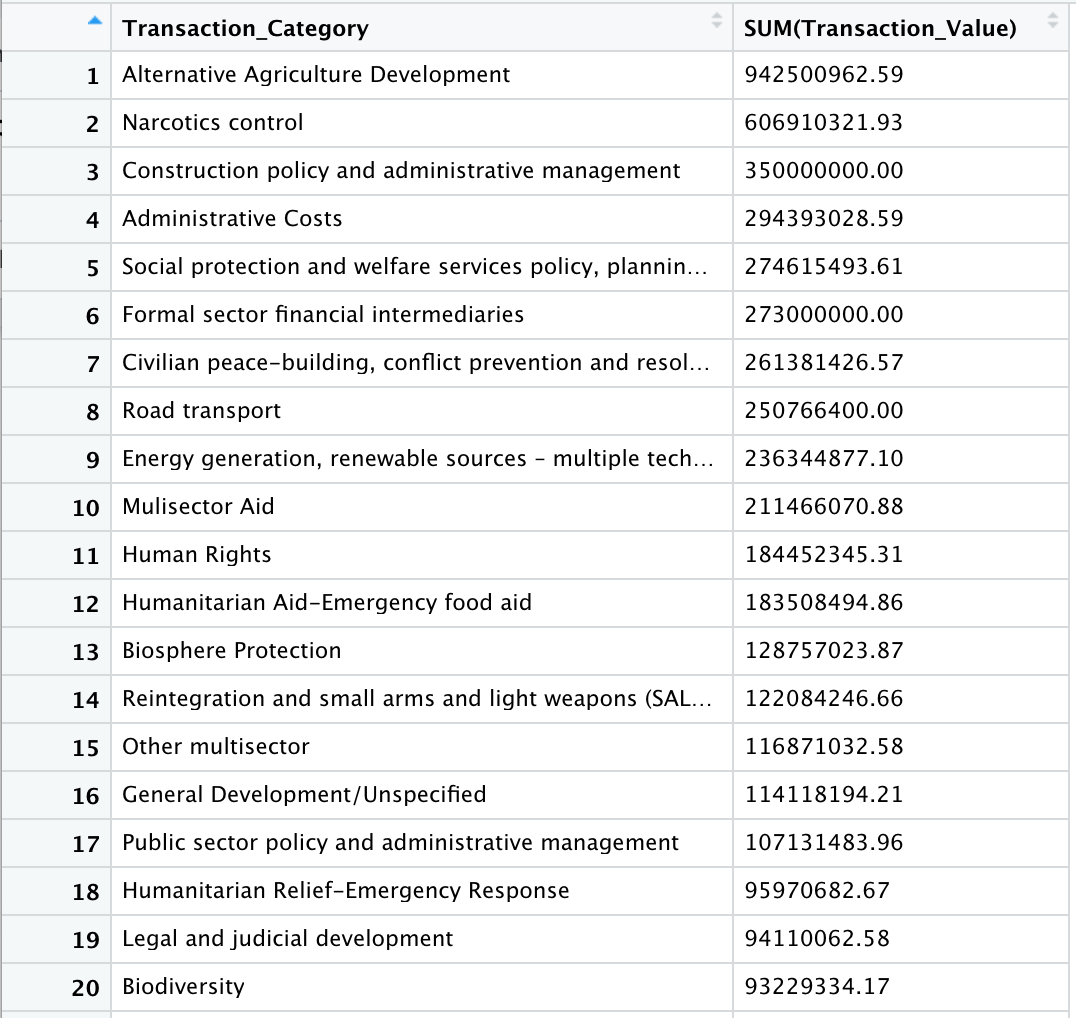
## Spikes in Narcotics Control and Education Indicators

##### **Effects of Narcotics Control Foreign Assistance Analysis**

In exploratory analysis, we discovered that the top 3 largest sums of total Colombia foreign assistance given has been towards Alternative Agriculture Development, Narcotics Control, and Construction Policy and Administrative Management transaction categories in various years over the time period of 2009 to 2019:

* Around $900 million ($942,599,862 precisely) total aid was given to Alternative Agriculture Development. This transaction category received the highest total sums of aid.
* Around $600 million ($606,910,321 precisely) total aid was given to Narcotics Control aid.
* Around $350 million ($350,000,000 precisely) total aid was given to Construction Policy and Administration Development Efforts in Colombia.

See diagram below. See Code Snippet 3A.

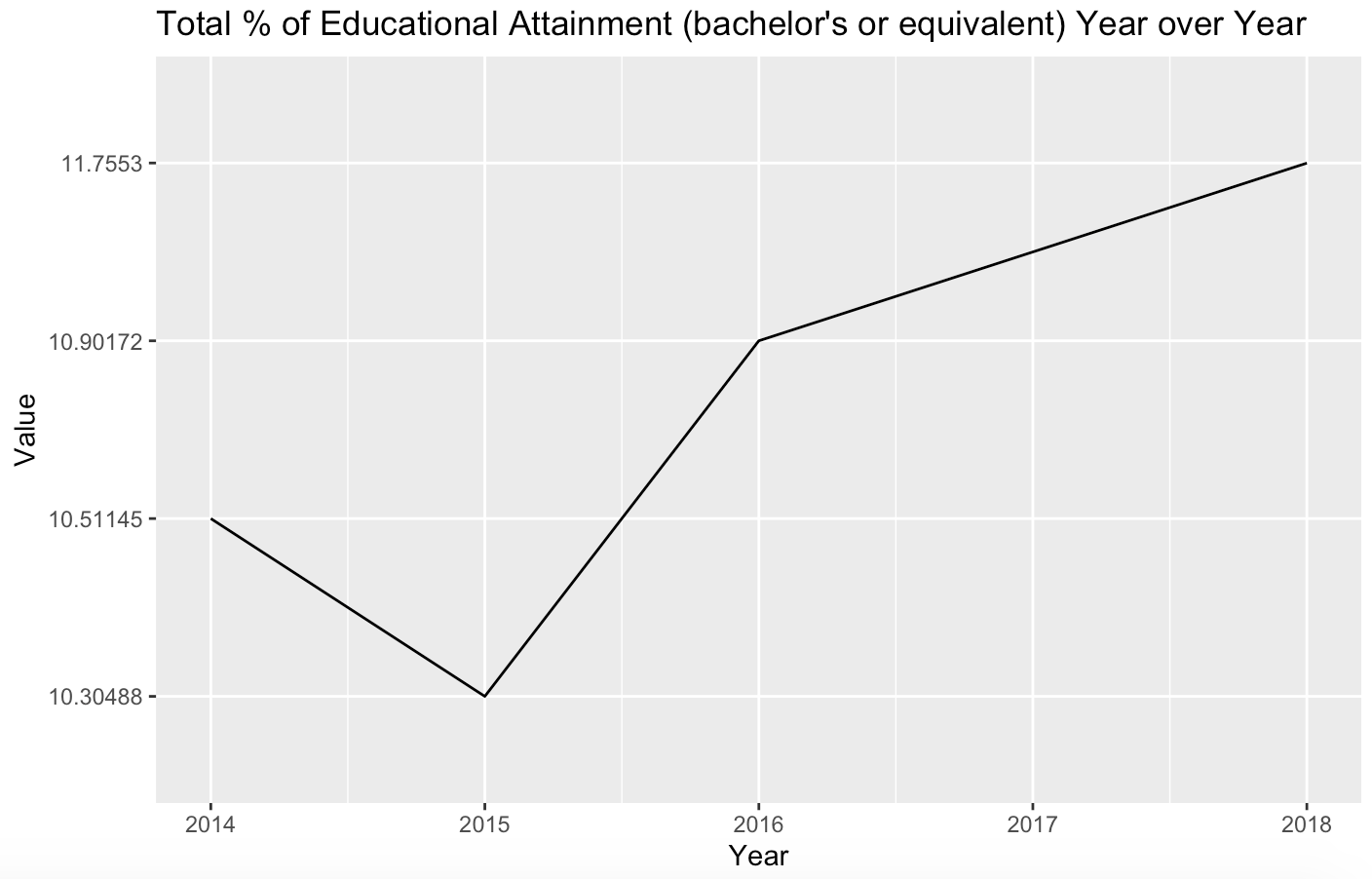
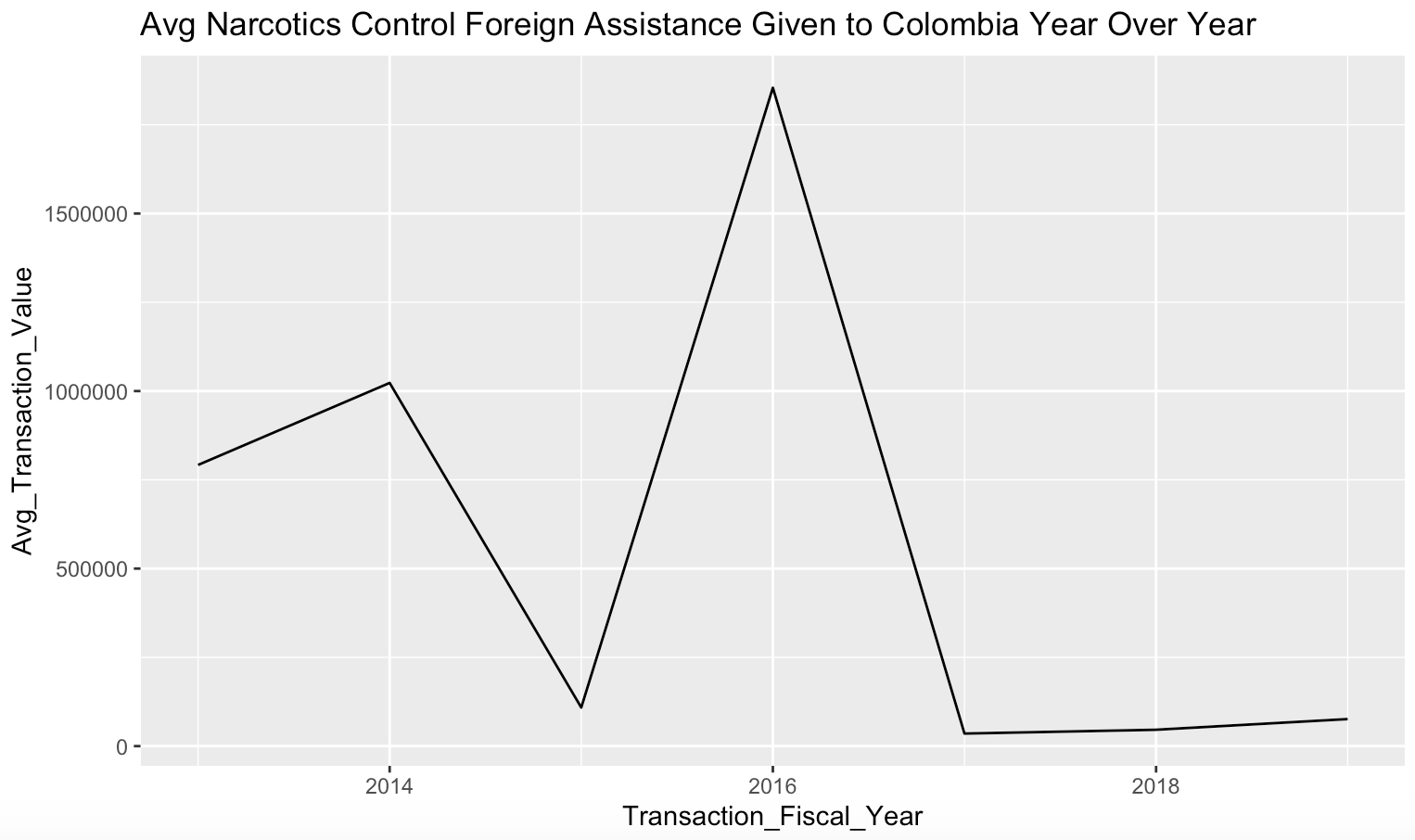


Since Narcotics Control aid was the second highest amount of foreign assistance given, we hypothesized that Narcotics Control efforts may influence education improvements in Colombia.

In our initial analysis of potential correlations, we discovered that there was, in fact, a correlation in the spike in Narcotics Control aid given in 2016, and improvements in the World Bank Indicator “Educational attainment, at least Bachelor's or equivalent, population 25+, total (%) (cumulative)” after the spike in aid was given in 2016.

The average quarterly amount of Narcotics Control foreign assistance given to Colombia spiked in 2016, where almost $2 million was given per quarter on average in 2016. This is up nearly $800k from the second highest amount of aid, given in 2014. How did this affect the World Bank Indicator for Educational Attainment (bachelor’s or equivalent)? The educational attainment percentage for this indicator increased from 10.9% to 11.7% from 2016 to 2018. This is a 7% increase of ~0.8 percentage points in the Educational Attainment (bachelor’s or equivalent) indicator. This correlation could suggest that the increase in foreign aid for Narcotics Control efforts given in 2016 resulted in the increase in the Educational Attainment (bachelor’s or equivalent) indicator in the 2 years after aid was given in 2016.

See diagrams below. See Code Snippet 3B.



Since we found a correlation between Narcotics Control and one of the WB indicators, Total % of educational attainment (bachelor’s or equivalent, we hypothesize that there may be more World Bank indicator improvements correlated with the 2016 spike in Narcotics Control foreign aid given to Colombia.

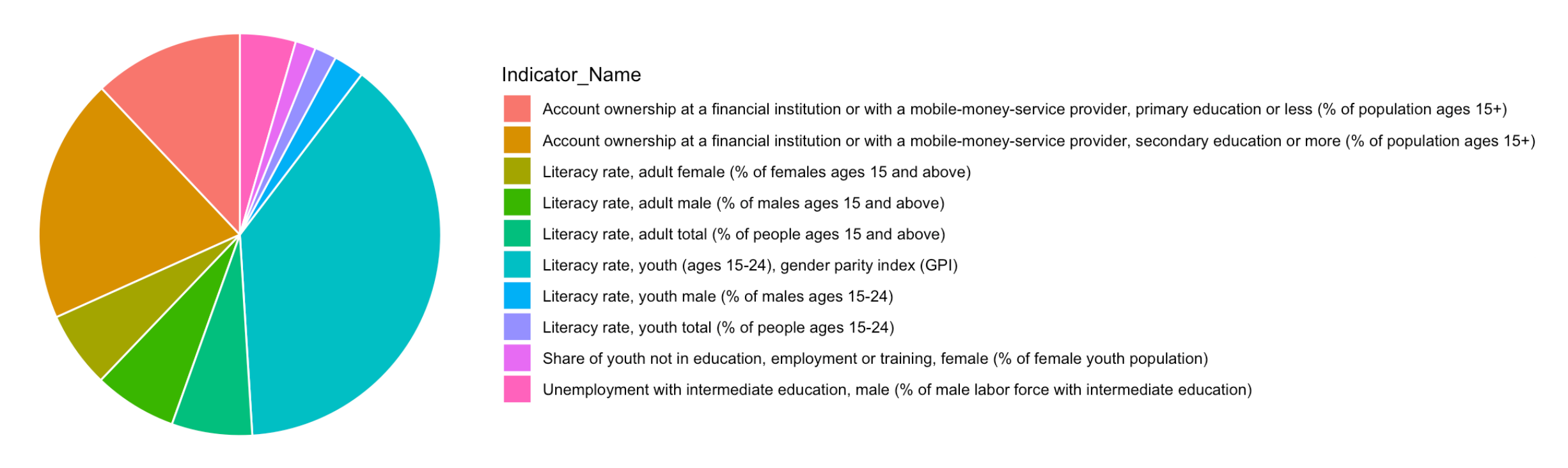
We discovered that there are 50 World Bank indicators that increased in value from 2016 to 2017. The chart below is only showing a subset of the top 20 World Bank indicators that had the largest change in value from 2016 to 2017. Some of these values require interpretation in the domain context, since increases in values for some indicators don’t necessarily mean improvements. It is dependent on what the indicator is. For example, when unemployment rates increase in value, this cannot be classified as an improvement for Colombia. 42 out of the 50 WB indicators that increased in value were positive improvements based on the domain context. Other indicators we chose to ignore are indicators that don’t seem to have a clear connection with Narcotics Control aid, indicators like “Government expenditure on education, total (% of GDP)”).

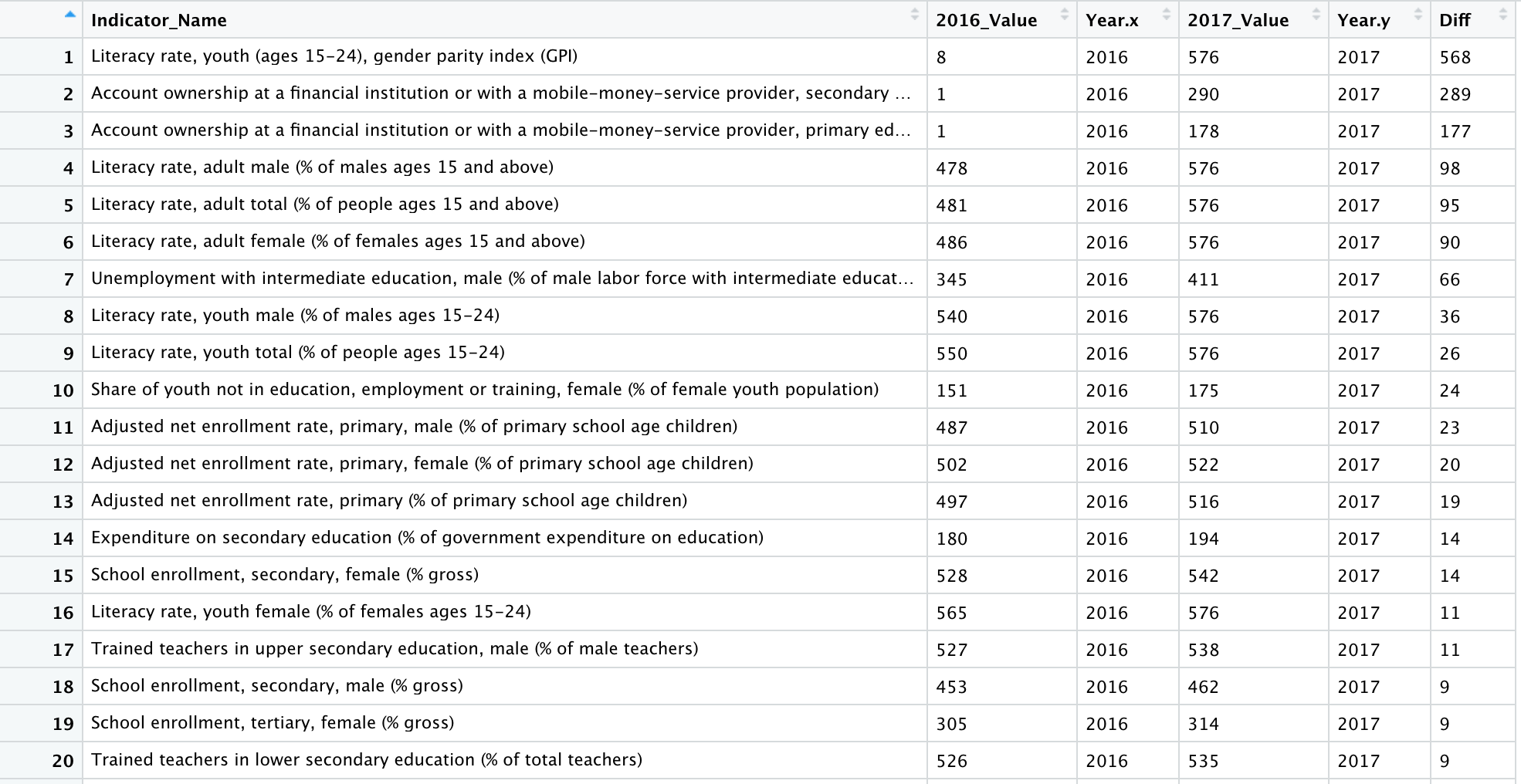
Thus, these were the top 24 indicators that had positive improvements from 2016 to 2017 and were correlated to the spike in 2016 Narcotic Control aid:

* Literacy rate, youth (ages 15-24), gender parity index (GPI)
* Literacy rate, adult male (% of males ages 15 and above)
* Literacy rate, adult total (% of people ages 15 and above)
* Literacy rate, adult female (% of females ages 15 and above)
* Literacy rate, youth male (% of males ages 15-24)
* Literacy rate, youth total (% of people ages 15-24)
* Adjusted net enrollment rate, primary, male (% of primary school age children)
* Adjusted net enrollment rate, primary, female (% of primary school age children)
* Adjusted net enrollment rate, primary (% of primary school age children)
* School enrollment, secondary, female (% gross)
* Literacy rate, youth female (% of females ages 15-24)
* School enrollment, secondary, male (% gross)
* School enrollment, tertiary, female (% gross)
* School enrollment, secondary (% gross)
* School enrollment, primary, private (% of total primary)
* School enrollment, secondary, male (% net)
* School enrollment, tertiary, male (% gross)

Since these 24 indicators all fall predominantly in 2 categories, literacy rates and school enrollment rates, this could suggest a relationship between the correlation of Narcotics Control aid and creating an environment where education success is enabled. This strong tie between drugs and education may help support the effects of controlling drug culture and creating healthy, safe environments for youth to develop and grow. There are likely many other factors contributing to the improvements in literacy rates and school enrollment rates which we explore further in the next section. However the correlations we have uncovered may provide new areas in studying the effects of Narcotics Control aid to explore through deeper analysis moving forward.

See diagrams below. See Code Snippet 3C.





##### 

##### **Deeper Exploration of Contributing Factors to the 2016-2017 Increase in Literacy and Enrollment Rates**

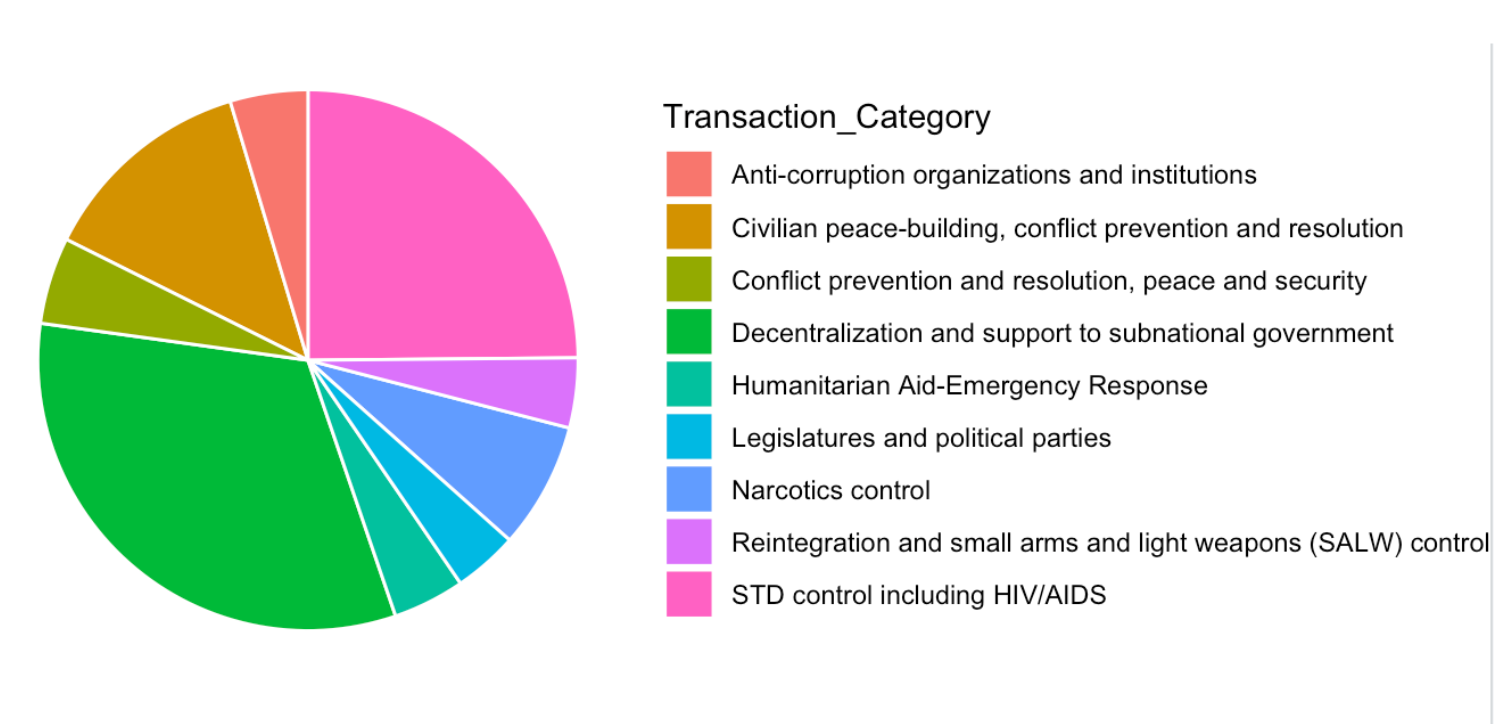
Our previous analysis begs the question: what were other categories for Foreign Assistance that worked in conjunction with the Narcotics Control aid given in 2016 that correlated with a spike in 2016 improvements in literacy rates and enrollment rates.

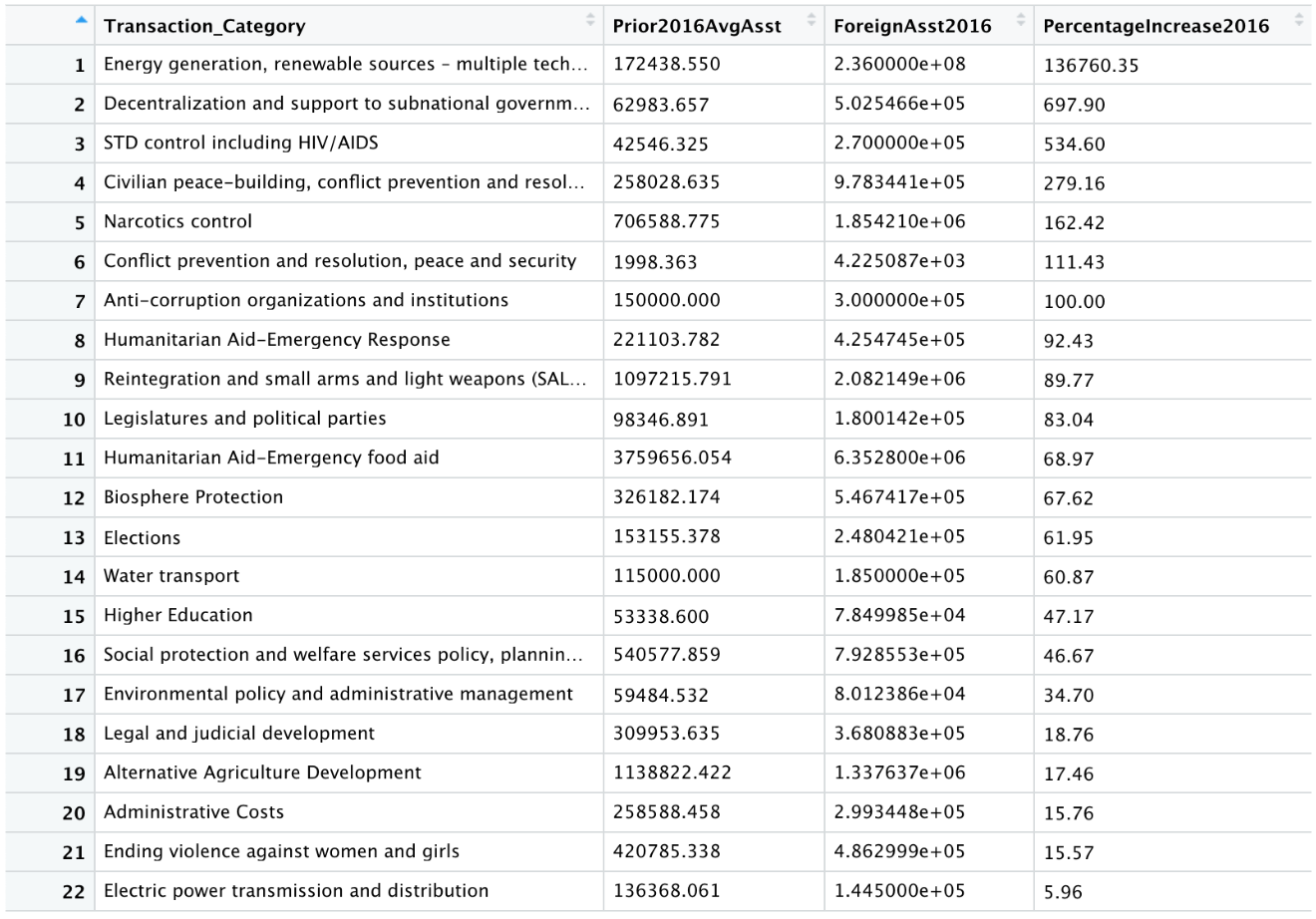
We found that the following Foreign Assistance categories saw over a 100% increase in 2016 aid given compared to the average yearly aid given pre-2016:

* Energy Generation, Renewable Sources aid saw a dramatic increase in percentage aid given in 2016, but we have filtered this transaction category out in the pie chart below given it is likely not strongly correlated with increases in literacy and enrollment rates.
* Decentralization and Support to Subnational Government aid had a 698% spike in aid given in 2016.
* STD Control Including HIV / AIDS aid saw a 534% spike in aid given in 2016.
* Civilian Peace-building, Conflict Prevention and Resolution aid saw a 297% spike in aid given in 2016.
* Narcotics Control aid saw a 162% spike in aid given in 2016.
* Conflict Prevention and Resolution, Peace and Security aid saw a 111% spike in aid given in 2016.
* Anti-corruption Organizations and Institutions aid saw a 100% spike in aid given in 2016.

We can see that the top spikes in 2016 aid, with the exception of the Energy Generation transaction category, were related to improving health, civilian peace, and criminal corruption. A The second highest amount of aid was given to supporting subnational governments, like state and local governments, often tied closely with the state and local education systems. What’s interesting is that there were 3 other prominent foreign aid transaction categories that were not on this list as potential correlating factors: Education Policy, Education Training, and Higher Education. These insights could suggest that investments in other areas of civilian life outside of direct education training and education policy can be strong influencers of education success.

See diagrams below. See Code Snippet 4A.





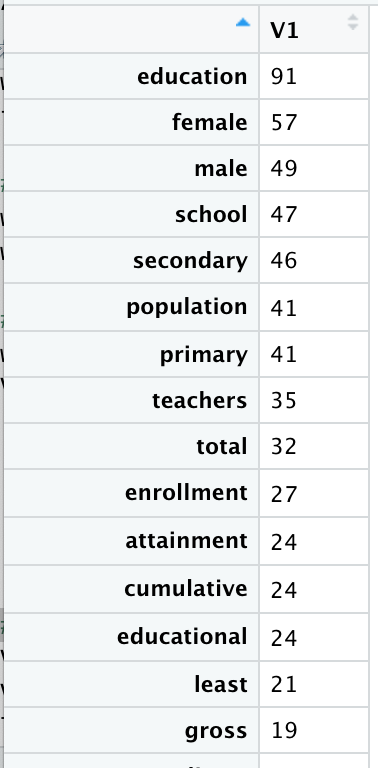
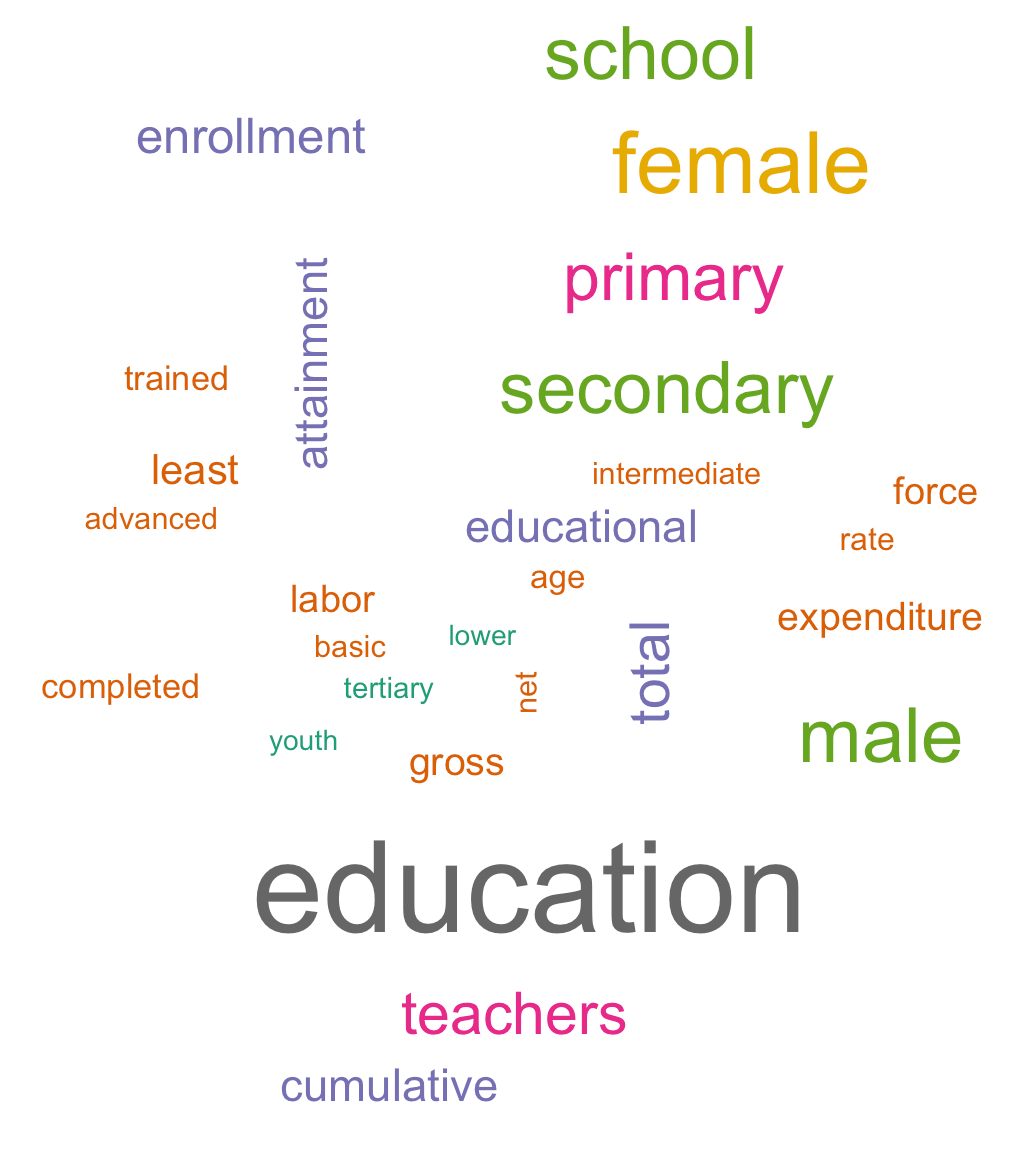
## 

## Text Mining Overview of Indicator Categories

The word cloud below shows the top 50 words that appeared at least 10 times in the list of unique World Bank indicators set that we used for this report’s analysis. With 675 rows of data in the World Bank Indicators data set, the word cloud gives us an overview of the common categories for the unique indicators in the data set:

* Education is a prominent category in the World Bank Indicators we have used in our analysis, and it is reflected through the text mining insights here. At the highest frequency, 91 unique indicators mention “education”. 47 mention “school”, 35 mention “teachers”, 27 mention “enrollment”, and 24 mention “ educational”. All of these are associated with education.
* The second 2 highest term frequencies in the unique indicators list are “female” and “male”, 57 and 49 respectively, so we can quickly see that Colombia’s World Bank indicators is tracking differences in gender in the indicators.
* “Secondary” and “primary” have 46 and 41 term frequencies respectively. This clues us into further breakdowns in the education population that the World Bank indicators are tracking for.

See diagram below. See Code Snippet 5A.

****

## Top Insights

**Overview Top Findings**

* A linear model better represents the contributions to Colombia over time per year rather than per quarter.
* Using the Fiscal Year model, we see an increase of $130,700,000 every year.
* Most contributions to OECDs come from one agency.
* An agency contributes to multiple OECDs.
* Narcotic Control is one of the only OECDs to receive multiple contributions from different agencies.
* The private investment coming from the US is very specific.

**Correlations Top Findings**

* The average quarterly amount of Narcotics Control foreign assistance given to Colombia spiked in 2016, where almost $2 million was given per quarter on average in 2016.
* The educational attainment percentage for this indicator increased from 10.9% to 11.7% from 2016 to 2018. This is a 7% increase of ~0.8 percentage points in the Educational Attainment (bachelor’s or equivalent) indicator.
* We discovered that there are 50 World Bank indicators that increased in value from 2016 to 2017.
* The Top 24 indicators that improved all fell predominantly in 2 categories, literacy rates and school enrollment rates.
* This could suggest a strong tie between drugs and education and may help support the effects of controlling drug culture and creating healthy, safe environments for youth to develop and grow.
* Top spikes in 2016 aid, with the exception of the Energy Generation transaction category, were related to improving health, civilian peace, and criminal corruption:
  + Decentralization and Support to Subnational Government aid = 698% spike
  + STD Control Including HIV / AIDS aid = 534% spike
  + Civilian Peace-building, Conflict Prevention and Resolution aid = 297%
  + Narcotics Control aid = 162% spike
  + Conflict Prevention and Resolution, Peace and Security aid = 111% spike
  + Anti-corruption Organizations and Institutions aid = 100% spike
* It is noteworthy that there were 3 other prominent foreign aid transaction categories that were not on this list as potential correlating factors: Education Policy, Education Training, and Higher Education.

## Code Appendix

##### **Section: Data Cleansing**

Code Snippet 1A:

FullFAData14\_15<-read.csv('ForeignAssistance-FullDataSet-2015-and-Before.csv')

FullFAData16\_17<-read.csv('ForeignAssistance-FullDataSet-2016-to-2017.csv')

FullFAData18\_19<-read.csv('ForeignAssistance-FullDataSet-2018-and-Later.csv')

colnames(FullFAData18\_19)[3]<- "Agency"

colnames(FullFAData18\_19)[4]<- "Bureau"

colnames(FullFAData18\_19)[13]<- "Country"

colnames(FullFAData18\_19)[21]<- "Money\_Spent"

colnames(FullFAData18\_19)[23]<- "Year"

colnames(FullFAData18\_19)[30]<- "DAC\_Code"

colnames(FullFAData16\_17)[3]<- "Agency"

colnames(FullFAData16\_17)[4]<- "Bureau"

colnames(FullFAData16\_17)[13]<- "Country"

colnames(FullFAData16\_17)[21]<- "Money\_Spent"

colnames(FullFAData16\_17)[23]<- "Year"

colnames(FullFAData16\_17)[30]<- "DAC\_Code"

colnames(FullFAData14\_15)[3]<- "Agency"

colnames(FullFAData14\_15)[4]<- "Bureau"

colnames(FullFAData14\_15)[13]<- "Country"

colnames(FullFAData14\_15)[21]<- "Money\_Spent"

colnames(FullFAData14\_15)[23]<- "Year"

colnames(FullFAData14\_15)[30]<- "DAC\_Code"

Code Snippet 1B:

COL18\_19Clean <-sqldf("SELECT Agency, Bureau,Country, Money\_Spent, Year, DAC\_Code FROM FullFAData18\_19 WHERE Country ='Colombia'")

COL16\_17Clean <-sqldf("SELECT Agency, Bureau,Country, Money\_Spent, Year, DAC\_Code FROM FullFAData16\_17 WHERE Country ='Colombia'")

COL14\_15Clean <-sqldf("SELECT Agency, Bureau,Country, Money\_Spent, Year, DAC\_Code FROM FullFAData14\_15 WHERE Country ='Colombia'")

Code Snippet 1C:

FullCOLSet<-rbind(COL18\_19Clean, COL16\_17Clean, COL14\_15Clean)

Code Snippet 1D:

WBIData<-read.csv('Colombia World Bank Indicators for Education RAW - API\_COL\_DS2\_en\_csv\_v2\_716391.csv')

**Section: Data Overview Analysis**

Code Snippet 2A:

df <- read.csv("ColombiaForeignAssistanceRAW.csv", header = TRUE)

colnames(df) <- c("Agency", "ReceivingLocation", "TransactionValue", "FiscalYear", "Quarter", "OECD", "AssistanceSectorCode")

Code Snippet 2B:

fAgg <- aggregate(.~df$FiscalYear+df$Quarter, df %>% select(TransactionValue), sum)

colnames(dfAgg) <- c("FiscalYear", "Quarters", "Value")

dfAgg <- dfAgg[dfAgg$FiscalYear >= 2014, ]

dfAgg$FiscalYear <- as.character(dfAgg$FiscalYear)

dfAgg$Quarter <- as.character(dfAgg$Quarter)

ggplot(dfAgg, aes(x = FiscalYear, y = Value, fill = Quarter)) +

geom\_col(position = "identity")

Code Snippet 2C:

df14 <- df[df$FiscalYear >= 2014,]

dfAggTotal <- aggregate(.~df14$OECD, df14 %>% select(TransactionValue), sum)

colnames(dfAggTotal) <- c("OECD", "Value")

View(dfAggTotal[order(-dfAggTotal$Value), ])

top5OECD = dfAggTotal[order(-dfAggTotal$Value), ]$OECD[c(1:5)]

Code Snippet 2D:

for (oecd in top5OECD) {

print(oecd)

df14 <- df[(df$FiscalYear >= 2014) & (df$OECD == oecd),]

dfAggAgency <- aggregate(.~df14$FiscalYear+df14$Quarter+df14$Agency, df14 %>% select(TransactionValue), sum)

colnames(dfAggAgency) <- c("FiscalYear", "Quarter", "Agency", "Value")

dfAggAgency$Percentage <- 100\*dfAggAgency$Value / sum(dfAggAgency$Value)

dfAgg <- aggregate(.~df14$FiscalYear+df14$Quarter, df14 %>% select(TransactionValue), sum)

colnames(dfAgg) <- c("FiscalYear", "Quarter", "Value")

dfAgg$FiscalYear <- as.character(dfAgg$FiscalYear)

dfAgg$Quarter <- as.character(dfAgg$Quarter)

pltVar <- ggplot(dfAgg, aes(x = FiscalYear, y = Value, fill = Quarter)) +

geom\_col(position = "identity") +

ggtitle(oecd)

bp<- ggplot(dfAggAgency, aes(x="", y=Percentage, fill=Agency))+

geom\_bar(width = 1, stat = "identity")

pie <- bp + coord\_polar("y", start=0) +

ggtitle(oecd)

print(pltVar)

print(pie)

}

Code Snippet 2E:

df14 <- df[df$FiscalYear >= 2014,]

dfAggTotal <- aggregate(.~df14$Agency, df14 %>% select(TransactionValue), sum)

colnames(dfAggTotal) <- c("Agency", "Value")

View(dfAggTotal[order(-dfAggTotal$Value), ])

top5Agency = dfAggTotal[order(-dfAggTotal$Value), ]$Agency[c(1:5)]

Code Snippet 2F:

for (agency in top5Agency) {

print(agency)

df14 <- df[(df$FiscalYear >= 2014) & (df$Agency == agency),]

dfAggOEDC <- aggregate(.~df14$FiscalYear+df14$Quarter+df14$OECD, df14 %>% select(TransactionValue), sum)

colnames(dfAggOEDC) <- c("FiscalYear", "Quarter", "OECD", "Value")

dfAggOEDC$Percentage <- 100\*dfAggOEDC$Value / sum(dfAggOEDC$Value)

dfAgg <- aggregate(.~df14$FiscalYear+df14$Quarter, df14 %>% select(TransactionValue), sum)

colnames(dfAgg) <- c("FiscalYear", "Quarter", "Value")

dfAgg$FiscalYear <- as.character(dfAgg$FiscalYear)

dfAgg$Quarter <- as.character(dfAgg$Quarter)

pltVar <- ggplot(dfAgg, aes(x = FiscalYear, y = Value, fill = Quarter)) +

geom\_col(position = "identity") +

ggtitle(agency)

bp<- ggplot(dfAggOEDC, aes(x="", y=Percentage, fill=OECD))+

geom\_bar(width = 1, stat = "identity")

pie <- bp + coord\_polar("y", start=0) +

ggtitle(agency)

print(pltVar)

print(pie)

}

Code Snippet 2G:

for (agency in top5Agency) {

print(agency)

df14 <- df[(df$FiscalYear >= 2014) & (df$Agency == agency),]

dfAgg <- aggregate(.~df14$FiscalYear+df14$Quarter+df14$OECD, df14 %>% select(TransactionValue), sum)

colnames(dfAgg) <- c("FiscalYear", "Quarter", "OECD", "Value")

dfAgg$FiscalQuarter <- paste(as.character(dfAgg$FiscalYear), as.character(dfAgg$Quarter))

pltVar <- ggplot(dfAgg, aes(x = FiscalQuarter, y = Value, fill = OECD)) +

geom\_col(position = "identity") +

ggtitle(paste(agency, " over time for each year"))

print(pltVar)

}

Code Snippet 2H:

dfAgg <- dfAgg[dfAgg$FiscalYear >= 2014, ]

dfAgg$FiscalYear <- as.character(dfAgg$FiscalYear)

dfAgg$Quarter <- as.character(dfAgg$Quarter)

dfAgg$Time <- seq(length(dfAgg$FiscalYear))

model <- lm(Value ~ Time, dfAgg)

summary(model)

Code Snippet 2I:

dfAggYear <- aggregate(.~df$FiscalYear, df %>% select(TransactionValue), sum)

colnames(dfAggYear) <- c("FiscalYear", "Value")

model2 <- lm(Value ~ FiscalYear, dfAggYear)

summary(model2)

##### **Section: Effects of Narcotics Control Foreign Assistance Analysis**

Code Snippet 3A:

sqldf("SELECT Transaction\_Category, SUM(Transaction\_Value) FROM ForeignAsst GROUP BY Transaction\_Category ORDER BY SUM(Transaction\_Value) DESC")

Code Snippet 3B:

# Filter the data set to be only the World Bank Indicator rows year over year for the Educational attainment bachelor's indicator

EduAttainBach <- WBIndicators[WBIndicators$Indicator\_Name=="Educational attainment, at least Bachelor's or equivalent, population 25+, total (%) (cumulative)",]

# Create a data viz to show the changes in this indicator from 2014 to 2018

g <- ggplot(EduAttainBach, aes(x=Year, y=Value, group=1)) + geom\_line() + ggtitle("Total % of Educational Attainment (bachelor's or equivalent) Year Over Year")

g

# Filter the data set for foreign assistance to be only the rows year over year given for the transcation category Narcotics Control

NarcoticsControl <- sqldf("SELECT Transaction\_Category, Transaction\_Fiscal\_Year, AVG(Transaction\_Value) AS Avg\_Transaction\_Value FROM ForeignAsst WHERE Transaction\_Category = 'Narcotics control' GROUP BY Transaction\_Category, Transaction\_Fiscal\_Year")

# Create a data viz to show the foreign assistance given to Narcotics control year over year

g <- ggplot(NarcoticsControl, aes(x=Transaction\_Fiscal\_Year, y=Avg\_Transaction\_Value)) + geom\_line() + ggtitle("Avg Narcotics Control Foreign Assistance Given to Colombia Year Over Year")

g

Code Snippet 3C:

# Filter the data set to be only the World Bank Indicator rows year over year for the Educational attainment bachelor's indicator

EduAttainBach <- WBIndi

# Which World Bank Indicators had a positive increase in value from 2016 to 2017

# Pull all of the WB Indicator values for 2016

df1 <- sqldf("SELECT \* FROM WBIndicators WHERE Year == 2016")

colnames(df1)[3] <- "2016\_Value"

# Pull all the WB Indicator values for 2017

df2 <- sqldf("SELECT \* FROM WBIndicators WHERE Year == 2017")

colnames(df2)[3] <- "2017\_Value"

# Merge all of the values into 1 data frame

IndicatorsIncreased2016to2017 <- merge(df1, df2, by="Indicator\_Name")

# Convert the 2016\_value and 2017\_value functions to numeric type

IndicatorsIncreased2016to2017$`2016\_Value` <- as.numeric(IndicatorsIncreased2016to2017$`2016\_Value`)

IndicatorsIncreased2016to2017$`2017\_Value` <- as.numeric(IndicatorsIncreased2016to2017$`2017\_Value`)

# Filter out any indicators where the 2017 value was less than the 2016 value

IndicatorsIncreased2016to2017$Diff <- (IndicatorsIncreased2016to2017$`2017\_Value`-IndicatorsIncreased2016to2017$`2016\_Value`)

IndicatorsIncreased2016to2017 <- IndicatorsIncreased2016to2017[IndicatorsIncreased2016to2017$Diff > 0,]

# Drop duplicate columns

IndicatorsIncreased2016to2017 <- IndicatorsIncreased2016to2017[,-2]

IndicatorsIncreased2016to2017 <- IndicatorsIncreased2016to2017[,-4]

# Reorder the rows by the Diff column

IndicatorsIncreased2016to2017 <- IndicatorsIncreased2016to2017[order(-IndicatorsIncreased2016to2017$Diff),]

# Reset the row numbers

rownames(IndicatorsIncreased2016to2017) <- NULL

# Create a pie chart to show the top 10 indicators increased from 2016 to 2017

TopIndicatorsIncreased2016to2017 <- ggplot(data = IndicatorsIncreased2016to2017[1:10,], aes(x="", y=Diff, fill=Indicator\_Name)) +

geom\_bar(stat="identity", width=1, color="white") +

coord\_polar("y", start=0) +

theme\_void()

TopIndicatorsIncreased2016to2017

##### **Section: Deeper Exploration of Contributing Factors to the 2016-2017 Increase in Literacy and Enrollment Rates**

Code Snippet 4A:

# Find the Avg Foreign Assistance given per Transaction Category prior to 2016

AvgForeignAsstPriorTo2016 <- sqldf("SELECT Transaction\_Category, AVG(Transaction\_Value) AS Prior2016AvgAsst

FROM ForeignAsst

WHERE Transaction\_Fiscal\_Year NOT IN (2016, 2017, 2018, 2019)

GROUP BY Transaction\_Category")

# Find the diff in foreign assistance given per Transaction Category for 2016 compared to the Avg in pre-2016 years

# Create a dataframe with just the 2016 foreign assistance given per Transaction Category

ForeignAsst2016 <- sqldf("SELECT Transaction\_Category, AVG(Transaction\_Value) AS ForeignAsst2016

FROM ForeignAsst

WHERE Transaction\_Fiscal\_Year = 2016

GROUP BY Transaction\_Category")

# Merge this data set with the AvgForeignAsstPriorTo2016 data set on Transaction\_Category to find the foreign assistance spikes in 2016

ForeignAsstSpikesIn2016 <- merge(AvgForeignAsstPriorTo2016, ForeignAsst2016, by="Transaction\_Category")

ForeignAsstSpikesIn2016$PercentageIncrease2016 <- round((((ForeignAsstSpikesIn2016$ForeignAsst2016 - ForeignAsstSpikesIn2016$Prior2016AvgAsst)/ForeignAsstSpikesIn2016$Prior2016AvgAsst) \* 100), digits = 2)

# Filter out all rows where there wasn't an increase in foreign assistance in 2016 compared to the avg of prior year

ForeignAsstSpikesIn2016 <- ForeignAsstSpikesIn2016[ForeignAsstSpikesIn2016$PercentageIncrease2016 > 0, ]

# Sort in descending order

ForeignAsstSpikesIn2016 <- ForeignAsstSpikesIn2016[order(-ForeignAsstSpikesIn2016$PercentageIncrease2016),]

rownames(ForeignAsstSpikesIn2016) <- NULL

View(ForeignAsstSpikesIn2016)

# Create a pie chart to show these findings for the Top 10 Foreign Assistance spikes in 2016

# Excludes the transaction category "Energy generation, renewable sources – multiple technologies" since this is likely unrelated to literacy rates and large in value that will skew the visualization

TopForeignAsstSpikesIn2016Graph <- ggplot(data = ForeignAsstSpikesIn2016[2:10,], aes(x="", y=PercentageIncrease2016, fill=Transaction\_Category)) +

geom\_bar(stat="identity", width=1, color="white") +

coord\_polar("y", start=0) +

theme\_void()

TopForeignAsstSpikesIn2016Graph

##### **Section: Text Mining Overview of Indicator Categories**

Code Snippet 5A:

# Text Mining

# Install packages

install.packages("tm")

library("tm")

install.packages("wordcloud")

library("wordcloud")

# Extract just the text from the Indicator\_Name column. Convert into a Corpus format.

wb <- WBIndicators$Indicator\_Name

# Remove duplicates and keep only unique World Bank Indicator Names

wb <- unique(wb)

wbVec <- VectorSource(wb)

wbCorpus <- Corpus(wbVec)

# Clean the corpus

# Convert to lowercase

wbCorpus <- tm\_map(wbCorpus, content\_transformer(tolower))

# Remove punctuation

wbCorpus <- tm\_map(wbCorpus, removePunctuation)

# Remove numbers

wbCorpus <- tm\_map(wbCorpus, removeNumbers)

# Remove stop words

wbCorpus <- tm\_map(wbCorpus, removeWords, stopwords("english"))

# Convert to a term document matrix

wbTDM <- TermDocumentMatrix(wbCorpus)

inspect(wbTDM)

# Calculate the frequency for each term in total across all documents

wbMatrix <- as.matrix(wbTDM)

wbFreq <- rowSums(wbMatrix)

# Sort with the highest frequency words first

wbFreq <- sort(wbFreq, decreasing=TRUE)

View(wbFreq)

# Create a word cloud for the World Bank indicator names

indicatorCloudDF <- data.frame(word=names(wbFreq), freq=wbFreq)

wordcloud(indicatorCloudDF$word, indicatorCloudDF$freq, colors=brewer.pal(8, "Dark2"), max.words=50, min.freq = 10)