Old files testing

Jose Luis Delgado Davara

5/30/2018

This file checks discrepancies between the two files GovData and GovDataIND of the 2017 ingestion.

```
library(dplyr)
library(stringr)
```

```
Data2017 = read.csv("../Ingestion Fall 2017 – Regina Files/GovData360.csv", sep = ";"
)
DataIND2017 = read.csv("../Ingestion Fall 2017 – Regina Files/GovData360IND.csv", sep
= ";")
```

Indicators in the values file not registered in the Indicators file

```
setdiff(unique(Data2017$Indicator_Id), unique(DataIND2017$Indicator_Id))
```

```
## character(0)
```

All indicators in the values file are registered in the indicators file.

Indicators in the IND file not registered in the values file

```
notmatched.indicators = setdiff(unique(DataIND2017$Indicator_Id), unique(Data2017$Ind
icator_Id))
print(notmatched.indicators)
```

```
##   [1] "BTI4"   "BTI57"  "BTI70"  "CPIA1"  "CPIA11" "CPIA12" "CPIA13"
##   [8] "CPIA14" "CPIA15" "CPIA16" "CPIA17" "CPIA18" "CPIA2"  "CPIA20"
##  [15] "CPIA4"  "CPIA5"  "CPIA6"  "CPIA8"  "CPIA9"  "DB1"    "DB10"
##  [22] "DB100"  "DB101"  "DB104"  "DB105"  "DB107"  "DB108"  "DB11"
##  [29] "DB12"   "DB13"   "DB14"   "DB15"   "DB16"   "DB17"   "DB18"
##  [36] "DB19"   "DB2"    "DB21"   "DB23"   "DB24"   "DB25"   "DB27"
##  [43] "DB32"   "DB34"   "DB35"   "DB37"   "DB38"   "DB4"    "DB40"
##  [50] "DB41"   "DB42"   "DB43"   "DB44"   "DB45"   "DB46"   "DB47"
##  [57] "DB48"   "DB49"   "DB51"   "DB52"   "DB53"   "DB54"   "DB55"
##  [64] "DB56"   "DB57"   "DB59"   "DB60"   "DB61"   "DB63"   "DB67"
##  [71] "DB68"   "DB70"   "DB72"   "DB73"   "DB74"   "DB75"   "DB76"
```

```
##  [78] "DB77"   "DB78"   "DB79"   "DB80"   "DB81"   "DB82"   "DB83"
##  [85] "DB84"   "DB86"   "DB87"   "DB89"   "DB91"   "DB92"   "DB93"
##  [92] "DB94"   "DB95"   "DB96"   "DB97"   "DB99"   "ES1"    "ES100"
##  [99] "ES101"  "ES102"  "ES103"  "ES104"  "ES105"  "ES106"  "ES107"
## [106] "ES108"  "ES109"  "ES110"  "ES111"  "ES112"  "ES113"  "ES114"
## [113] "ES115"  "ES116"  "ES117"  "ES118"  "ES12"   "ES120"  "ES121"
## [120] "ES122"  "ES13"   "ES14"   "ES15"   "ES16"   "ES17"   "ES18"
## [127] "ES19"   "ES2"    "ES20"   "ES21"   "ES23"   "ES24"   "ES25"
## [134] "ES26"   "ES27"   "ES28"   "ES29"   "ES3"    "ES30"   "ES31"
## [141] "ES32"   "ES33"   "ES34"   "ES35"   "ES36"   "ES37"   "ES38"
## [148] "ES39"   "ES4"    "ES40"   "ES41"   "ES42"   "ES43"   "ES44"
## [155] "ES45"   "ES46"   "ES47"   "ES48"   "ES49"   "ES5"    "ES50"
## [162] "ES51"   "ES52"   "ES53"   "ES54"   "ES55"   "ES56"   "ES57"
## [169] "ES58"   "ES59"   "ES6"    "ES60"   "ES61"   "ES62"   "ES63"
## [176] "ES64"   "ES65"   "ES66"   "ES67"   "ES68"   "ES69"   "ES7"
## [183] "ES70"   "ES71"   "ES72"   "ES73"   "ES74"   "ES75"   "ES76"
## [190] "ES77"   "ES78"   "ES79"   "ES8"    "ES80"   "ES81"   "ES82"
## [197] "ES83"   "ES84"   "ES85"   "ES86"   "ES87"   "ES88"   "ES89"
## [204] "ES9"    "ES90"   "ES91"   "ES92"   "ES93"   "ES94"   "ES95"
## [211] "ES96"   "ES97"   "ES98"   "ES99"   "GCI10"  "GCI100" "GCI101"
## [218] "GCI102" "GCI103" "GCI104" "GCI11"  "GCI12"  "GCI128" "GCI13"
## [225] "GCI130" "GCI131" "GCI132" "GCI133" "GCI134" "GCI135" "GCI15"
## [232] "GCI160" "GCI18"  "GCI28"  "GCI29"  "GCI30"  "GCI31"  "GCI32"
## [239] "GCI33"  "GCI34"  "GCI35"  "GCI36"  "GCI38"  "GCI39"  "GCI40"
## [246] "GCI41"  "GCI42"  "GCI43"  "GCI44"  "GCI45"  "GCI46"  "GCI47"
## [253] "GCI48"  "GCI52"  "GCI53"  "GCI54"  "GCI55"  "GCI56"  "GCI57"
## [260] "GCI58"  "GCI59"  "GCI60"  "GCI61"  "GCI68"  "GCI71"  "GCI73"
## [267] "GCI74"  "GCI75"  "GCI76"  "GCI77"  "GCI78"  "GCI79"  "GCI8"
## [274] "GCI80"  "GCI82"  "GCI83"  "GCI84"  "GCI85"  "GCI86"  "GCI87"
## [281] "GCI89"  "GCI9"   "GCI90"  "GCI91"  "GCI92"  "GCI93"  "GCI94"
## [288] "GCI95"  "GCI96"  "GCI97"  "GCI98"  "GCI99"  "SC1"    "SC10"
## [295] "SC11"   "SC12"   "SC13"   "SC14"   "SC15"   "SC16"   "SC17"
## [302] "SC18"   "SC19"   "SC2"    "SC20"   "SC21"   "SC22"   "SC23"
## [309] "SC24"   "SC25"   "SC26"   "SC27"   "SC28"   "SC29"   "SC3"
## [316] "SC4"    "SC5"    "SC6"    "SC7"    "SC8"    "SC9"    "WBL1"
## [323] "WBL10"  "WBL100" "WBL101" "WBL102" "WBL103" "WBL104" "WBL105"
## [330] "WBL106" "WBL107" "WBL108" "WBL109" "WBL11"  "WBL110" "WBL111"
## [337] "WBL112" "WBL113" "WBL114" "WBL115" "WBL116" "WBL117" "WBL118"
## [344] "WBL119" "WBL12"  "WBL120" "WBL121" "WBL122" "WBL123" "WBL124"
## [351] "WBL125" "WBL126" "WBL127" "WBL128" "WBL129" "WBL13"  "WBL130"
## [358] "WBL131" "WBL132" "WBL14"  "WBL15"  "WBL16"  "WBL17"  "WBL18"
## [365] "WBL19"  "WBL2"   "WBL20"  "WBL21"  "WBL22"  "WBL23"  "WBL24"
## [372] "WBL25"  "WBL26"  "WBL27"  "WBL28"  "WBL29"  "WBL3"   "WBL30"
## [379] "WBL31"  "WBL32"  "WBL33"  "WBL34"  "WBL35"  "WBL36"  "WBL37"
## [386] "WBL38"  "WBL39"  "WBL4"   "WBL40"  "WBL41"  "WBL42"  "WBL43"
## [393] "WBL44"  "WBL45"  "WBL46"  "WBL47"  "WBL48"  "WBL49"  "WBL5"
## [400] "WBL50"  "WBL51"  "WBL52"  "WBL53"  "WBL54"  "WBL55"  "WBL56"
```

```
## [407] "WBL57"  "WBL58"  "WBL59"  "WBL6"   "WBL60"  "WBL61"  "WBL62"
## [414] "WBL63"  "WBL64"  "WBL65"  "WBL66"  "WBL67"  "WBL68"  "WBL69"
## [421] "WBL7"   "WBL70"  "WBL71"  "WBL72"  "WBL73"  "WBL74"  "WBL75"
## [428] "WBL76"  "WBL77"  "WBL78"  "WBL79"  "WBL8"   "WBL80"  "WBL81"
## [435] "WBL82"  "WBL83"  "WBL84"  "WBL85"  "WBL86"  "WBL87"  "WBL88"
## [442] "WBL89"  "WBL9"   "WBL90"  "WBL91"  "WBL92"  "WBL93"  "WBL94"
## [449] "WBL95"  "WBL96"  "WBL97"  "WBL98"  "WBL99"  "WGI1"   "WGI2"
## [456] "WGI3"   "WGI4"   "WGI5"   "WGI6"
```

There are 459 indicators in the Fall 2017 ingestion that do not match in the two files. The INDICATORS files have more that the values file.

Data sources with inconsistencies

```
Sources = gsub("[[:digit:]]", "", notmatched.indicators)
print('List of data sources with inconsintencies and number of inconsistencies:')
```

```
## [1] "List of data sources with inconsintencies and number of inconsistencies:"
```

```
table(Sources)
```

```
## Sources
##  BTI CPIA   DB   ES  GCI   SC  WBL  WGI
##    3   16   77  118   78   29  132    6
```