

DATA606_Final_Project_RJM

RJM

2020-02-13

Introduction

The question chosen to answer was to figure out whether there is a relationship between a country's wealth (GDP output) and the type of industry that its economy is mainly engaged in. In other words, the aim was to find the impact of main industries in particular countries or regions on the economy and general well-being of its inhabitants. To approach it from a research angle, we used the Worldbank's datasets to find the main industries in the countries or regions of the world. The two main industry categories we used were knowledge-based and traditional. The knowledge-based industries was further segregated into manufacturing and services for the knowledge-based economies while agriculture and minerals were considered to be sub-categories for traditional industries.

The initial hypothesis is:

1. The knowledge-based economies are wealthier.
2. The traditional economies are mostly developing or emerging.

Data

We thought that the industry composition of GDP is an indicator that will tell us about the main industries in a certain country or region. For that purpose, we used Worldbank's website to collect data on the four categories namely agriculture, minerals, services, and manufacturing. The data had to be curated to fit it between the years of 1995-2018 as most of the countries had missing information for years before 1995. There was still some need to impute missing data for certain countries or regions but we figured after some experimentation that it is best to leave it blank to avoid weird fluctuations in the graph.

The sources and the links for our data are as follows:

1. GDP

<https://data.worldbank.org/indicator/NY.GDP.MKTP.CD>

2. GDP per Capita

<https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>

3. Agriculture, forestry, and fishing as % of GDP

<https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS>

4. Ores and Metals exports (% of merchandise exports) taken as a proxy to mineral production and exports

<https://data.worldbank.org/indicator/TX.VAL.MMTL.ZS.UN>

5. Service, value added (% of GDP)

<https://data.worldbank.org/indicator/NV.SRV.TOTL.ZS>

6. Manufacturing, value added (% of GDP)

<https://data.worldbank.org/indicator/NV.IND.MANF.ZS>

The code for the preparation of the data is as follows:

```
library(flexdashboard)
library(DT)
library(ggplot2)
library(readxl)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(tidyr)
library(stringr)
```

```
rm(list = ls())
```

```
#disable scientific notation, so that actual decimal values are imported instead of exponential factors
options(scipen = 999)
```

```
# Importing country Metadata dataset into R
```

```
download.file("https://github.com/rjmirza/DATA-606/raw/master/final_project/datasets/GDP.xls", "GDP.xls")
country_metadata_dataset <- read_excel("GDP.xls", col_names = TRUE, sheet = "Metadata - Countries")
```

```
# Importing GDP (1995-2018) by country dataset into R
```

```
gdp_dataset <- read_excel("GDP.xls", col_names = TRUE, sheet = "Data", skip = 3) %>%
  data.frame(., stringsAsFactors = F) %>%
  select(., 1,2,3,40:63)
```

```
# Importing GDP percapita (1995-2018) by country dataset into R
```

```
download.file("https://github.com/rjmirza/DATA-606/raw/master/final_project/datasets/GDP%20per%20Capita", "GDP_per_Capita.xls")
gdp_percapita_dataset <- read_excel("GDP_per_Capita.xls", col_names = TRUE, sheet = "Data", skip = 3) %>%
  data.frame(., stringsAsFactors = F) %>%
```

```

select(., 1,2,3,40:63)

# Importing Manufacturing GDP (1995-2018) percentage by country dataset into R
download.file("https://github.com/rjmirza/DATA-606/raw/master/final_project/datasets/Manufacturing.xls", "Manufacturing.xls")
gdp_manufacturing_dataset <- read_excel("Manufacturing.xls", col_names = TRUE, sheet = "Data", skip = 3) %>%
  data.frame(., stringsAsFactors = F) %>%
  select(., 1,2,3,40:63)

# Importing Agriculture GDP (1995-2018) percentage by country dataset into R
download.file("https://github.com/rjmirza/DATA-606/raw/master/final_project/datasets/Agriculture.xls", "Agriculture.xls")
gdp_agriculture_dataset <- read_excel("Agriculture.xls", col_names = TRUE, sheet = "Data", skip = 3) %>%
  data.frame(., stringsAsFactors = F) %>%
  select(., 1,2,3,40:63)

# Importing Service GDP (1995-2018) percentage by country dataset into R
download.file("https://github.com/rjmirza/DATA-606/raw/master/final_project/datasets/Service.xls", "Service.xls")
gdp_service_dataset <- read_excel("Service.xls", col_names = TRUE, sheet = "Data", skip = 3) %>%
  data.frame(., stringsAsFactors = F) %>%
  select(., 1,2,3,40:63)

# Importing Industries GDP (1995-2018) percentage by country dataset into R
download.file("https://github.com/rjmirza/DATA-606/raw/master/final_project/datasets/Industries.xls", "Industries.xls")
gdp_industries_dataset <- read_excel("Industries.xls", col_names = TRUE, sheet = "Data", skip = 3) %>%
  data.frame(., stringsAsFactors = F) %>%
  select(., 1,2,3,40:63)

# Importing Ores_Metals_Minerals GDP (1995-2018) percentage by country dataset into R
download.file("https://github.com/rjmirza/DATA-606/raw/master/final_project/datasets/Ores_Metals_Minerals.xls", "Ores_Metals_Minerals.xls")
gdp_ores_metals_minerals_dataset <- read_excel("Ores_Metals_Minerals.xls", col_names = TRUE, sheet = "Data", skip = 3) %>%
  data.frame(., stringsAsFactors = F) %>%
  select(., 1,2,3,40:63)

```

```

df1 <- gather(gdp_dataset, "year", "GDP", 4:27) %>% select(1, 4, 5)
df1$GDP <- df1$GDP/1000000
df2 <- gather(gdp_percapita_dataset, "year", "GDP Percapita", 4:27) %>% select(1, 4, 5)
df3 <- gather(gdp_industries_dataset, "year", "Industry Percent of GDP", 4:27) %>% select(1, 4, 5)
df4 <- gather(gdp_service_dataset, "year", "Services Percent of GDP", 4:27) %>% select(1, 4, 5)
df5 <- gather(gdp_agriculture_dataset, "year", "Agriculture Percent of GDP", 4:27) %>% select(1, 4, 5)
df6 <- gather(gdp_manufacturing_dataset, "year", "Manufacturing Percent of GDP", 4:27) %>% select(1, 4, 5)
df7 <- gather(gdp_ores_metals_minerals_dataset, "year", "Ores_Metals_Minerals Percent of GDP", 4:27) %>% select(1, 4, 5)
df <- merge(df1, df2, all.y = T)
df <- merge(df, df3, all.y = T)
df <- merge(df, df4, all.y = T)
df <- merge(df, df5, all.y = T)
df <- merge(df, df6, all.y = T)
df <- merge(df, df7, all.y = T)
df <- merge(country_metadata_dataset, df, by.x = "TableName", by.y = "Country.Name", all.y = T)
summary(df)

```

##	TableName	Country Code	Region	IncomeGroup
##	Length:6336	Length:6336	Length:6336	Length:6336
##	Class :character	Class :character	Class :character	Class :character
##	Mode :character	Mode :character	Mode :character	Mode :character

```
##
##
##
##
## SpecialNotes      year      GDP      GDP Percapita
## Length:6336      Length:6336      Min.   :    11      Min.   :   278.4
## Class :character  Class :character  1st Qu.:   4581      1st Qu.:  3053.2
## Mode  :character  Mode  :character  Median :   29549      Median :   8261.1
##                                     Mean  : 1764054      Mean   :  14919.8
##                                     3rd Qu.:  368923      3rd Qu.: 20406.8
##                                     Max.   :85804391      Max.   :139962.3
##                                     NA's   :423         NA's   :679
## Industry Percent of GDP Services Percent of GDP Agriculture Percent of GDP
## Min.   : 0.0034      Min.   : 9.727      Min.   : 0.0249
## 1st Qu.:19.8557      1st Qu.:44.972      1st Qu.: 3.1915
## Median :25.7390      Median :52.958      Median : 8.4866
## Mean   :27.3662      Mean   :53.203      Mean   :12.1676
## 3rd Qu.:32.3540      3rd Qu.:61.439      3rd Qu.:18.4360
## Max.   :87.7969      Max.   :96.465      Max.   :79.0424
## NA's   :871         NA's   :1174      NA's   :832
## Manufacturing Percent of GDP Ores_Metals_Minerals Percent of GDP
## Min.   : 0.000      Min.   : 0.000
## 1st Qu.: 8.126      1st Qu.: 1.153
## Median :12.695      Median : 3.018
## Mean   :13.188      Mean   : 7.339
## 3rd Qu.:16.725      3rd Qu.: 6.633
## Max.   :191.998      Max.   :86.540
## NA's   :1156      NA's   :1717
```

```
# removing characters from the year and converting the type to numeric
df$year <- str_extract(df$year, "[[:digit:]]+") %>%
  as.numeric(df$year)
```

```
incomegroup_df <- df %>%
  filter(., is.na(IncomeGroup)) %>%
  filter(., `Country Code` %in% c("EAR", "FCS", "HIC", "HPC", "LDC", "LIC", "LMC", "LMY", "LTE", "MIC", "PRE", "PS"))
  arrange(TableName, year)
```

```
economy_by_region_df <- df %>%
  filter(., is.na(IncomeGroup)) %>%
  filter(., TableName %in% c("East Asia & Pacific", "Europe & Central Asia", "Latin America & Caribbean",
  select(1,3,4,6:8,10:13) %>%
  arrange(TableName, year)
```

```
knowledge_traditinoal_dF <- df %>%
  filter(., !is.na(IncomeGroup)) %>%
  select(1,3,4,6,7,10:13) %>%
  mutate("Knowledge based Percent of GDP" = ifelse(is.na(`Services Percent of GDP`), 0, `Services Percent of GDP`),
         ifelse(is.na(`Manufacturing Percent of GDP`),0,`Manufacturing Percent of GDP`),
         "Traditinoal based Percent of GDP" = ifelse(is.na(`Agriculture Percent of GDP`),0,`Agriculture Percent of GDP`),
         ifelse(is.na(`Ores_Metals_Minerals Percent of GDP`),0,`Ores_Metals_Minerals Percent of GDP`))
  arrange(TableName, year)
```

```

country_gdp_mean_sd_dF <- knowledge_traditinoal_dF %>%
  group_by(TableName) %>%
  summarise("Country Mean GDP" = mean(GDP, na.rm=TRUE),
            "Country SD GDP" = sd(GDP, na.rm=TRUE)
            )

world_knowledge_gdp_percent_mean_dF <- knowledge_traditinoal_dF %>%
  group_by(year) %>%
  summarise("World Mean Knowledge GDP percent" = mean(`Knowledge based Percent of GDP`, na.rm=TRUE))

world_traditional_gdp_percent_mean_dF <- knowledge_traditinoal_dF %>%
  group_by(year) %>%
  summarise("World Mean Traditinoal GDP percent" = mean(`Traditinoal based Percent of GDP`, na.rm=TRUE))

knowledge_traditinoal_dF <- knowledge_traditinoal_dF %>%
  merge(., country_gdp_mean_sd_dF, by.x = "TableName", by.y = "TableName", all.y = T) %>%
  mutate("Country SD GDP in percent" = `Country SD GDP`/`Country Mean GDP`*100) %>%
  merge(., world_knowledge_gdp_percent_mean_dF, by.x = "year", by.y = "year", all.y = T) %>%
  merge(., world_traditional_gdp_percent_mean_dF, by.x = "year", by.y = "year", all.y = T) %>%
  select(1:5,10:16) %>%
  na_if(., 0) %>%
  arrange(TableName, year)

```

Exploratory data analysis

This is an observational study where we took existing data to support our hypothesis. We thought that the best way to compare the GDP output amongst different regions or type of economy (income-based differentiation) would be to use linear graphs with multiple lines representing each region or economic class.

Below are all the data produced in taular form after transformation to fit the purpose of this study:

```

### initial dataframe
DT::datatable(df, options = list(pageLength = 5))

```

PhantomJS not found. You can install it with webshot::install_phantomjs(). If it is installed, please

```

### incomegroup dataframe
DT::datatable(incomegroup_df, options = list(pageLength = 5))

```

```

### economy by region dataframe
DT::datatable(economy_by_region_df, options = list(pageLength = 5))

```

```

### knowledge and traditinoal GDP's dataframe
DT::datatable(knowledge_traditinoal_dF, options = list(pageLength = 5))

```

As it is apparent from the above, the graphical representation makes it much easier to decipher the results.

```

NIG <- length(unique(incomegroup_df[["TableName"]]))

valueBox(NIG, color = "primary")

```

13 ### Inference

The results in the forms of the graphs are produced below"

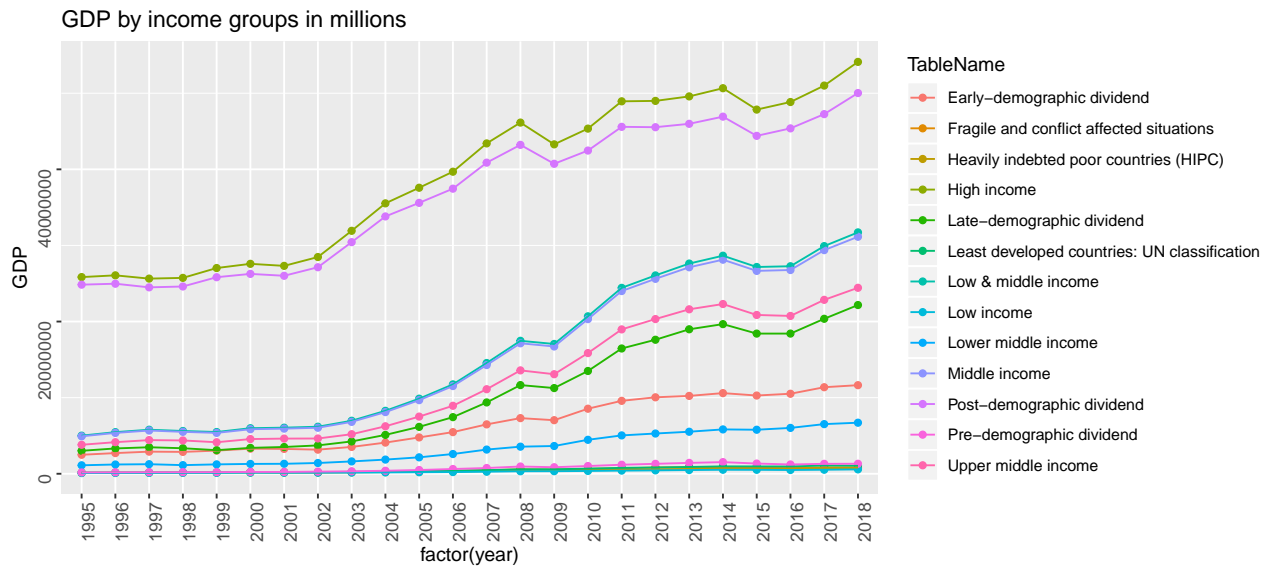
GDP

This the general introduction to give an idea on how the GDP output compares amongst the nations with different economic classes and demographic stages.

```
ggplot(incomegroup_df, aes(x=factor(year), colour=TableName, group = TableName)) +  
  geom_point(aes(y = `GDP`)) +  
  geom_line(aes(y = `GDP`)) +  
  theme(axis.text.x = element_text(size=10, angle=90)) +  
  theme(axis.text.y = element_text(size=10, angle=90)) +  
  labs(title = "GDP by income groups in millions")
```

Warning: Removed 5 rows containing missing values (geom_point).

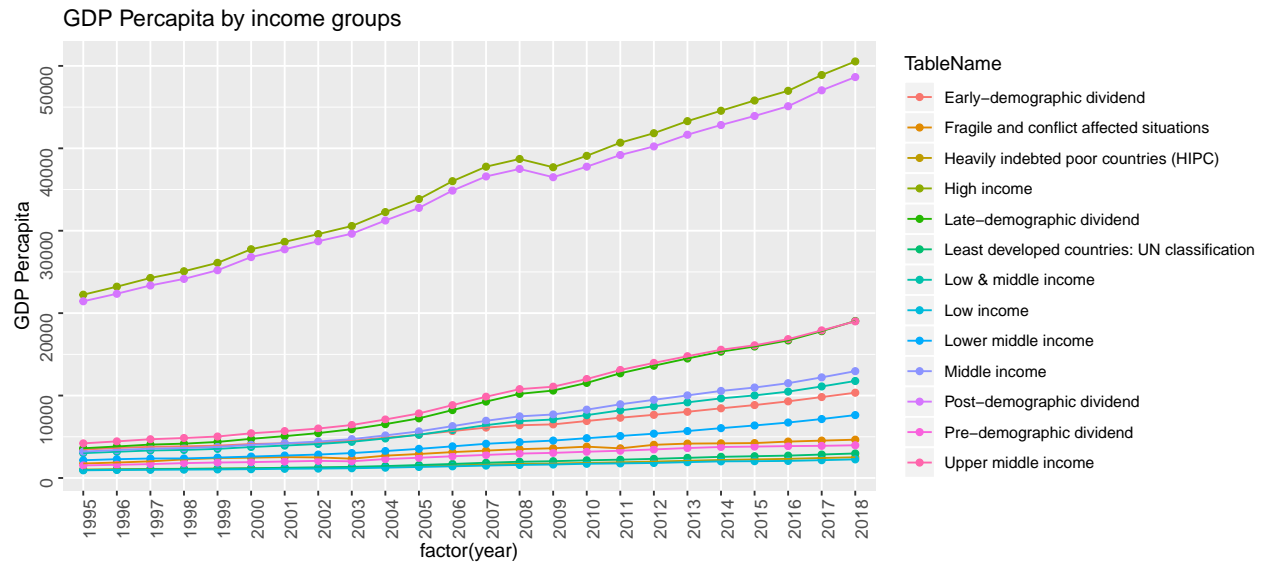
Warning: Removed 5 rows containing missing values (geom_path).



GDP Per Capita

GDP per Capita was produced to give an idea about regarding the impact of the population on the output.

```
ggplot(incomegroup_df, aes(x=factor(year), colour=TableName, group = TableName)) +  
  geom_point(aes(y = `GDP Percapita`)) +  
  geom_line(aes(y = `GDP Percapita`)) +  
  theme(axis.text.x = element_text(size=10, angle=90)) +  
  theme(axis.text.y = element_text(size=10, angle=90)) +  
  labs(title = "GDP Percapita by income groups")
```



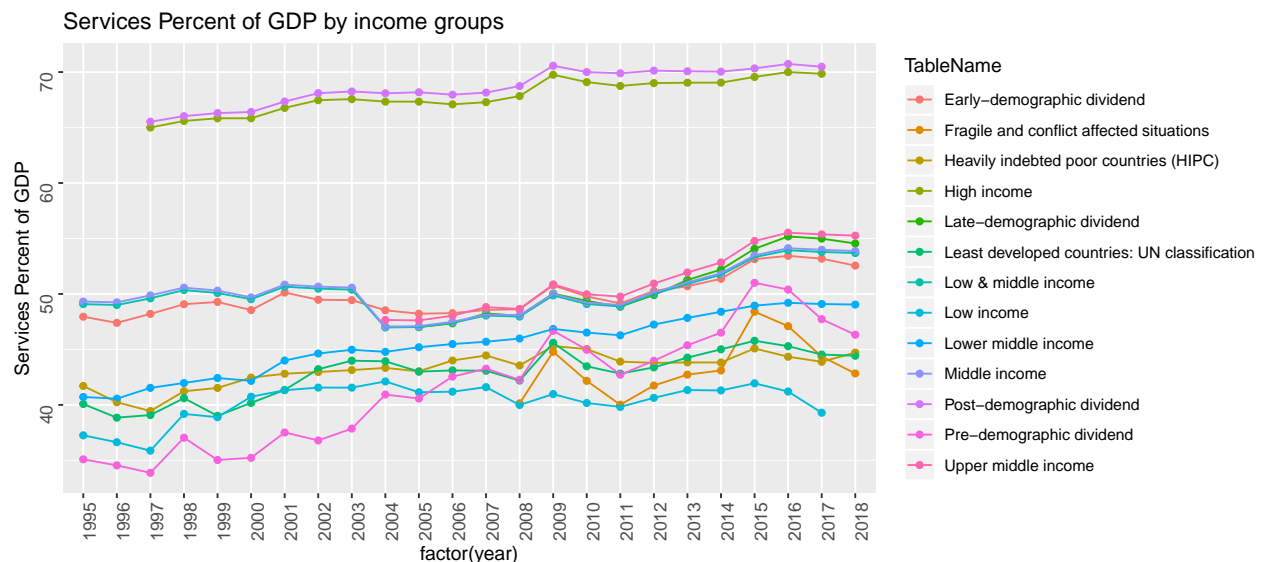
Services Percent of GDP

These graphs represent the portion of GDP produced by the activities related to the Services industry. We can see that the countries with high income are heavily involved in this industry and are also the highest in GDP output as indicated in the above graphs.

```
ggplot(incomegroup_df, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Services Percent of GDP`)) +
  geom_line(aes(y = `Services Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Services Percent of GDP by income groups")
```

Warning: Removed 38 rows containing missing values (geom_point).

Warning: Removed 38 rows containing missing values (geom_path).



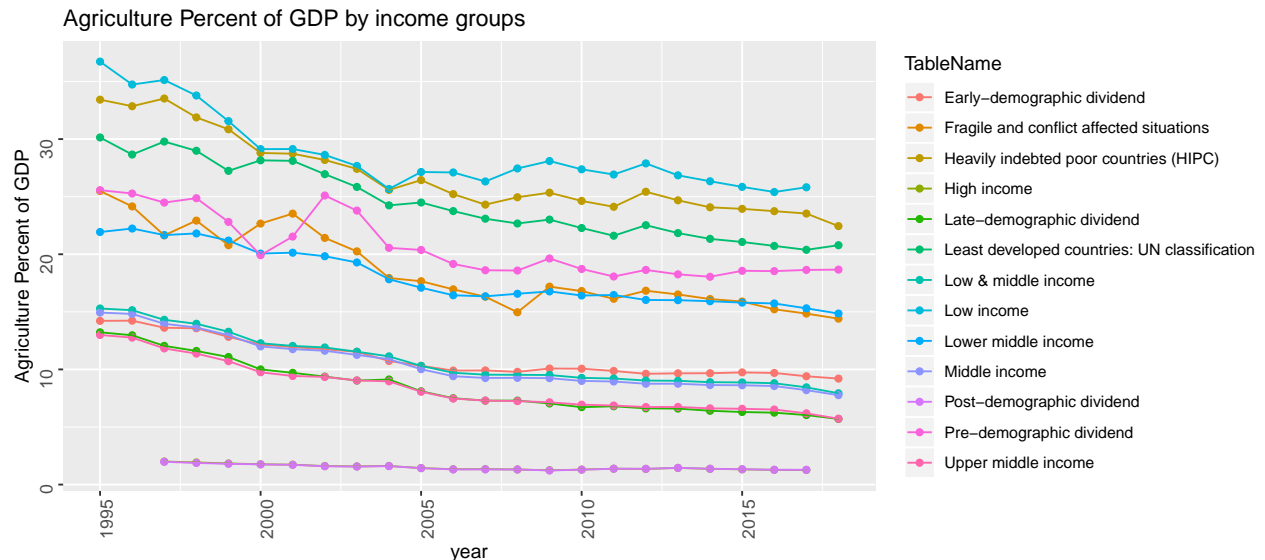
Agriculture Percent of GDP

We can see that the Agriculture industry's contribution to GDP are lowering over time for all types of economic classes. This is more starkly obvious for the countries with high income.

```
ggplot(incomegroup_df, aes(x=year, colour=TableName, group = TableName)) +  
  geom_point(aes(y = `Agriculture Percent of GDP`)) +  
  geom_line(aes(y = `Agriculture Percent of GDP`)) +  
  theme(axis.text.x = element_text(size=10, angle=90)) +  
  theme(axis.text.y = element_text(size=10, angle=90)) +  
  labs(title = "Agriculture Percent of GDP by income groups")
```

```
## Warning: Removed 7 rows containing missing values (geom_point).
```

```
## Warning: Removed 7 rows containing missing values (geom_path).
```

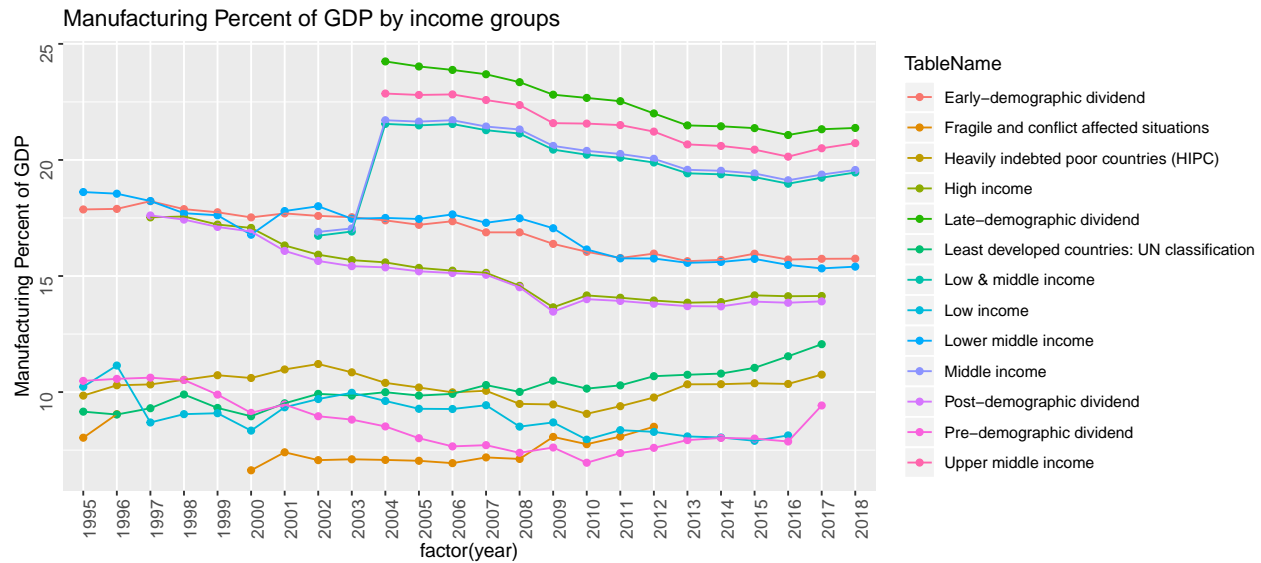


Manufacturing Percent of GDP

The graph below illustrated the impact of the manufacturing industry on the different countries' GDP output. It seems that the high income economies have been able to move their source of GDP output to other industries but manufacturing still plays a significant role. It is more dominant of an industry for the middle income countries (including upper and lower middle).

```
## Warning: Removed 52 rows containing missing values (geom_point).
```

```
## Warning: Removed 49 rows containing missing values (geom_path).
```

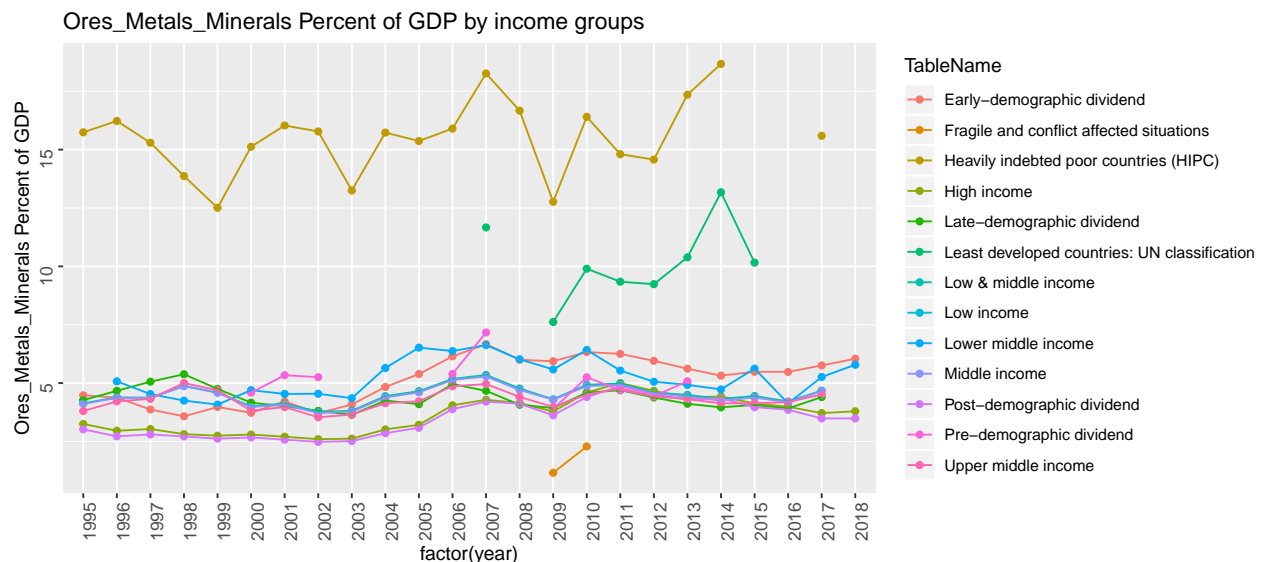
Ores_Metals_Minerals Percent of GDP

This category gave us the most significant challenge with the imputation of data. However, after some trial and error, we were able to satisfactorily use the data to create a graph where we could observe some trends. It seems that this traditional industry contributes mainly to the GDP of countries with high debt and it cannot be associated with economies that are better established and mature.

```
ggplot(incomegroup_df, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Ores_Metals_Minerals Percent of GDP`)) +
  geom_line(aes(y = `Ores_Metals_Minerals Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Ores_Metals_Minerals Percent of GDP by income groups")
```

Warning: Removed 84 rows containing missing values (geom_point).

Warning: Removed 77 rows containing missing values (geom_path).



Economy by region Charts

To further test our inferences, we decided to check the performance of these industries based on the regions. We already knew the economic conditions of each region and could see that our results matched with the outcomes based on the economic classes.

Number of Regions

```
NIG <- length(unique(economy_by_region_df[["TableName"]]))  
  
valueBox(NIG, color = "primary")
```

7

Service based economy

```
valueBox("North America", color = "info")
```

North America

```
### Manufacturing based economy  
valueBox("East Asia and Pacific", color = "info")
```

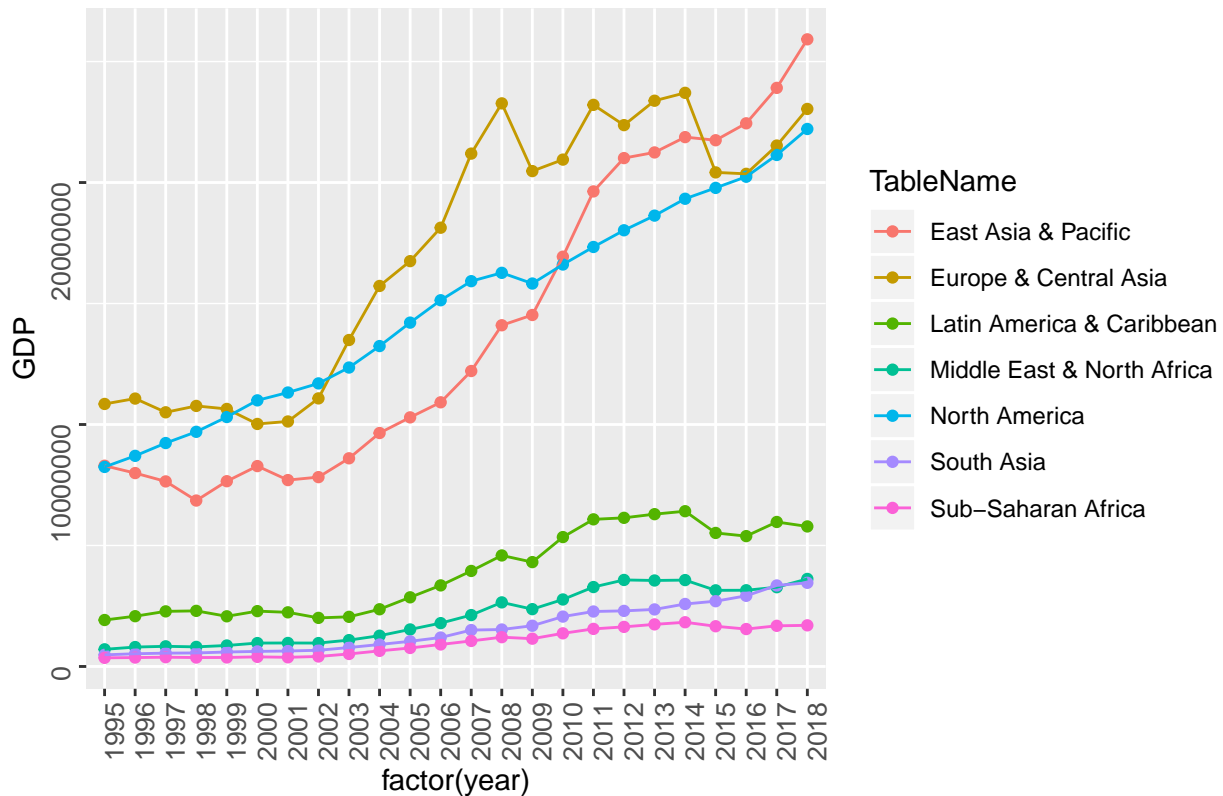
East Asia and Pacific

```
### Agriculture based economy  
valueBox("South Asia", color = "info")
```

South Asia

```
### GDP  
  
ggplot(economy_by_region_df, aes(x=factor(year), colour=TableName, group = TableName)) +  
  geom_point(aes(y = `GDP`)) +  
  geom_line(aes(y = `GDP`)) +  
  theme(axis.text.x = element_text(size=10, angle=90)) +  
  theme(axis.text.y = element_text(size=10, angle=90)) +  
  labs(title = "GDP by Region in millions")
```

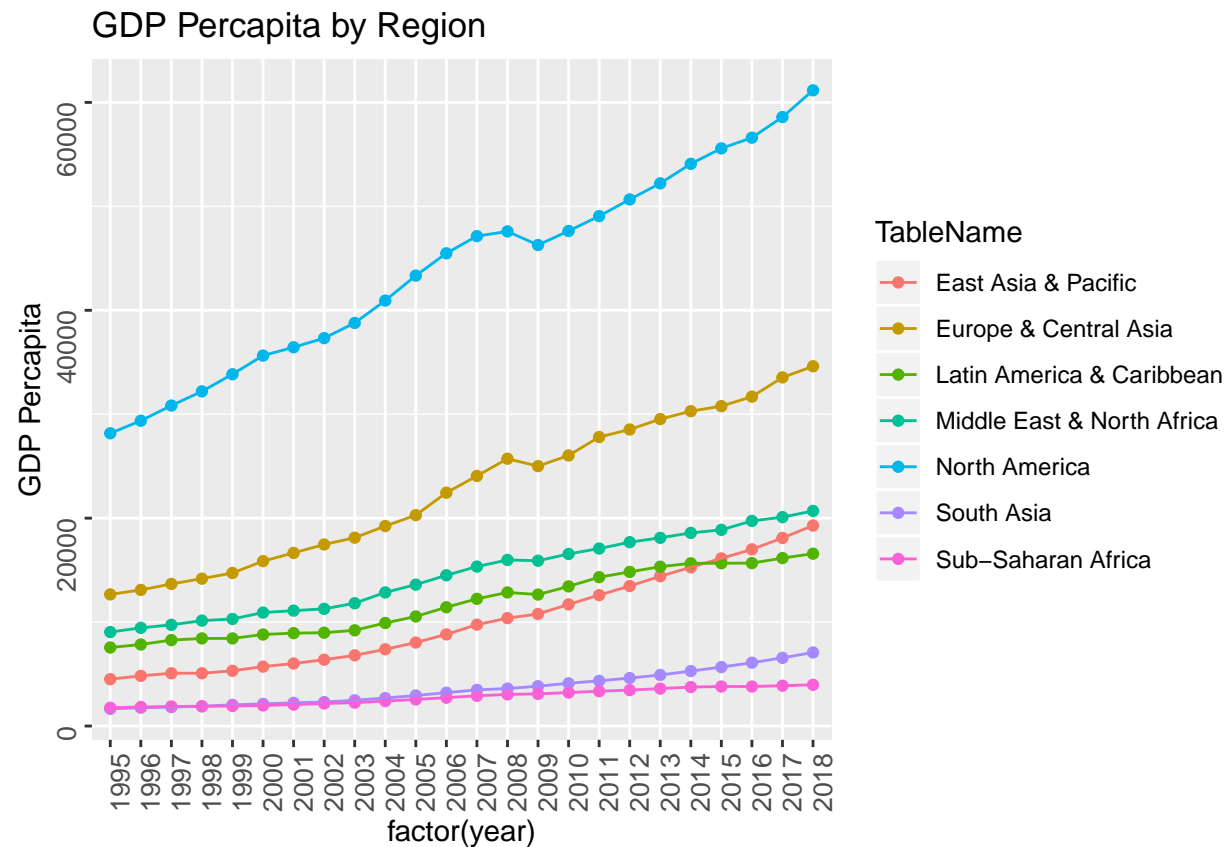
GDP by Region in millions



The above graph shows that the generally wealthy regions of the world have a higher GDP.

GDP Percapita

```
ggplot(economy_by_region_df, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `GDP Percapita`)) +
  geom_line(aes(y = `GDP Percapita`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "GDP Percapita by Region")
```



This graph further illustrates the GDP output and neutralizes the population advantage. Here we can see the East Asia & Pacific slipping down but still edging up quite nicely.

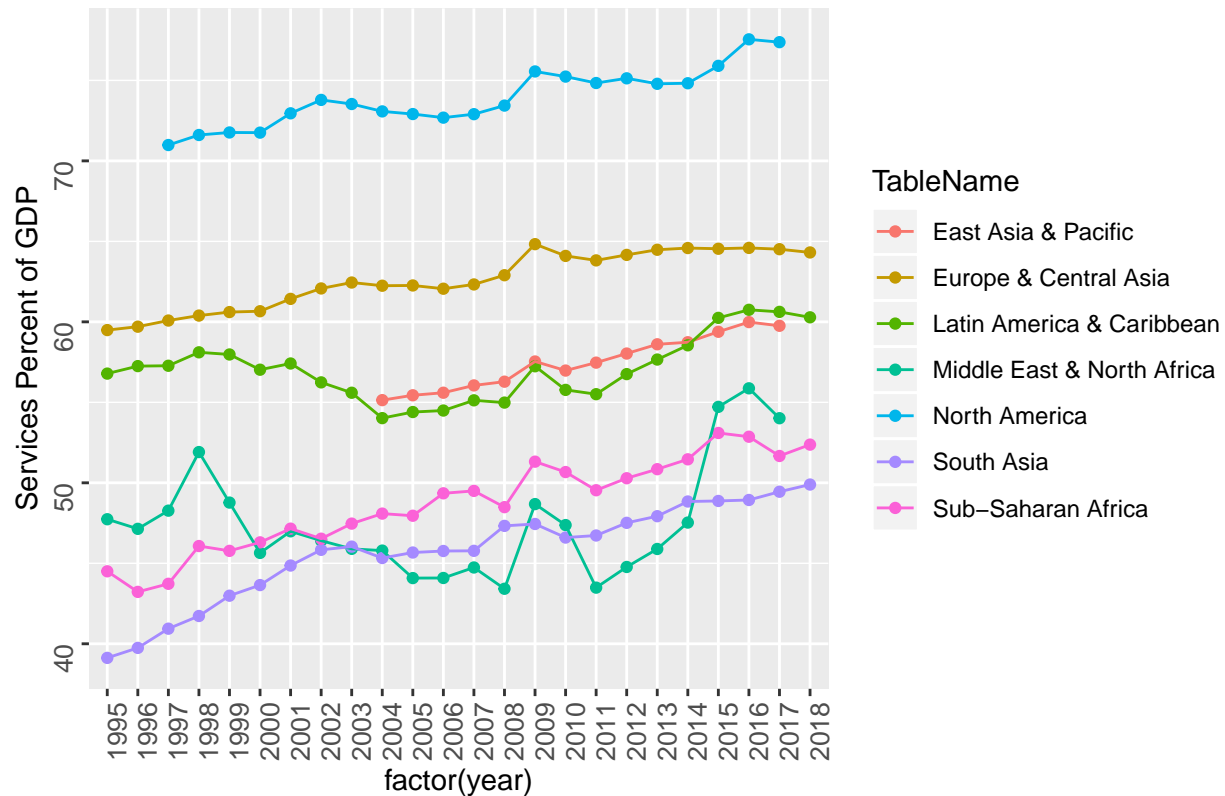
```
### Services Percent of GDP
```

```
ggplot(economy_by_region_df, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Services Percent of GDP`)) +
  geom_line(aes(y = `Services Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Services Percent of GDP by Region")
```

```
## Warning: Removed 14 rows containing missing values (geom_point).
```

```
## Warning: Removed 14 rows containing missing values (geom_path).
```

Services Percent of GDP by Region



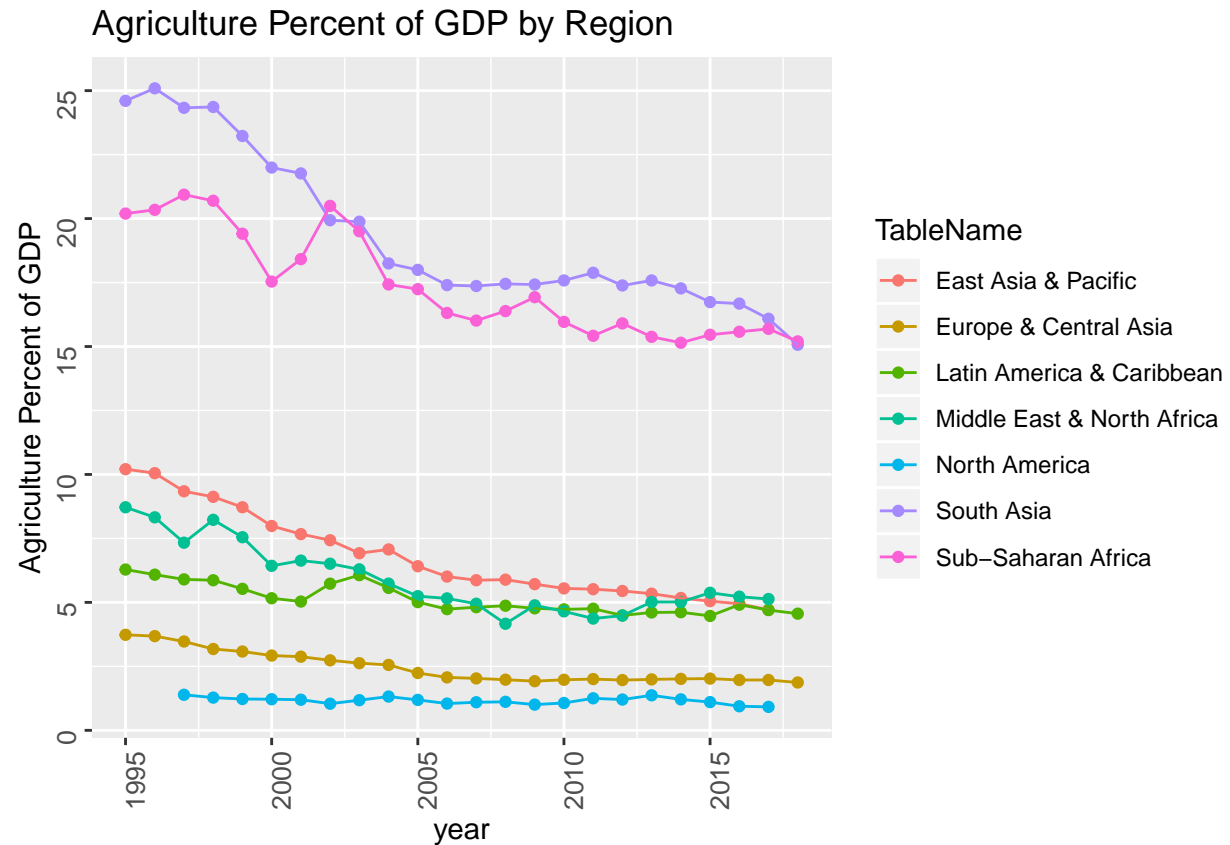
The services industry is dominant in the relatively wealthier regions of North America and Europe.

Agriculture Percent of GDP

```
ggplot(economy_by_region_df, aes(x=year, colour=TableName, group = TableName)) +
  geom_point(aes(y = `Agriculture Percent of GDP`)) +
  geom_line(aes(y = `Agriculture Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Agriculture Percent of GDP by Region")
```

Warning: Removed 5 rows containing missing values (geom_point).

Warning: Removed 5 rows containing missing values (geom_path).



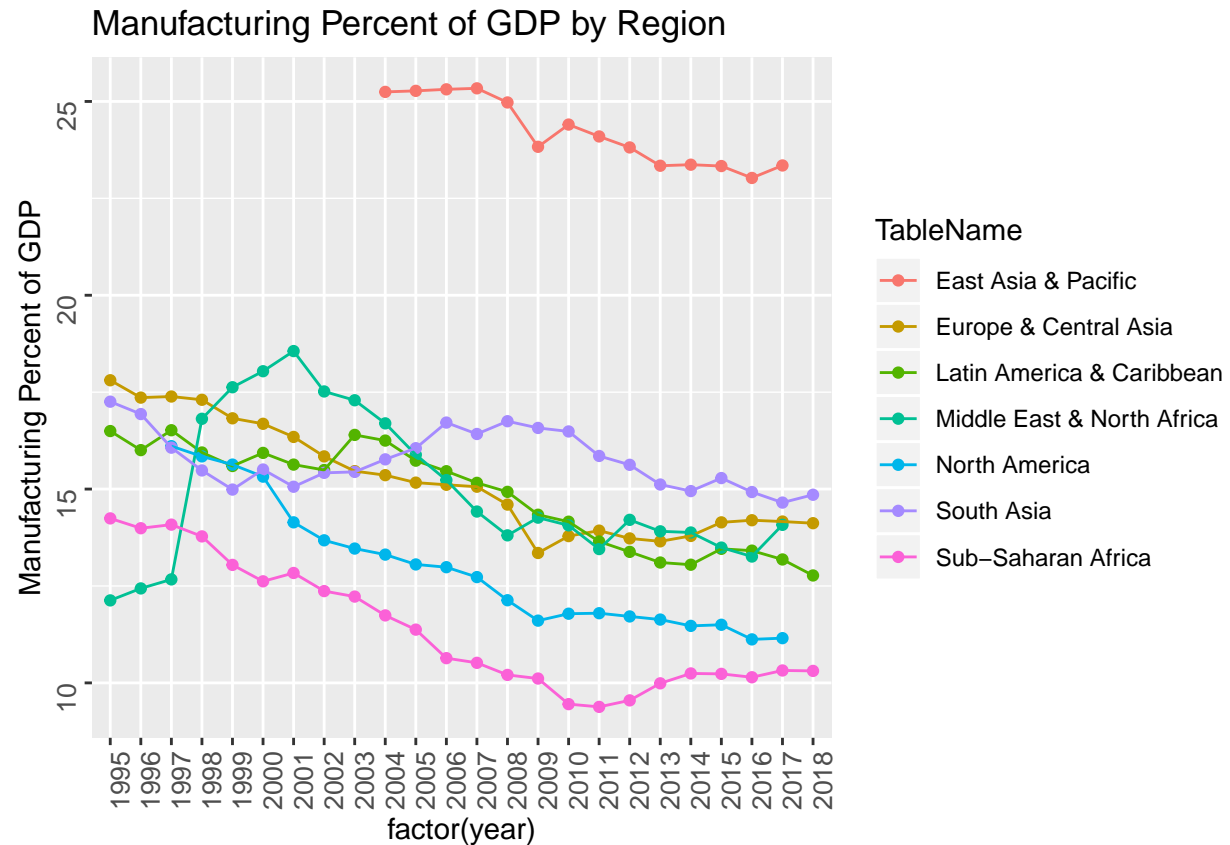
Agriculture industry's contribution to GDP is really small for wealthier regions but it is in a downward trend for all other regions as well.

Manufacturing Percent of GDP

```
ggplot(economy_by_region_df, aes(x=year, colour=TableName, group = TableName)) +
  geom_point(aes(y = `Manufacturing Percent of GDP`)) +
  geom_line(aes(y = `Manufacturing Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Manufacturing Percent of GDP by Region")
```

Warning: Removed 14 rows containing missing values (geom_point).

Warning: Removed 14 rows containing missing values (geom_path).



While manufacturing industry is still a significant part of the GDP for wealthier countries, it is really dominant in the regions associated with access to low-cost labour.

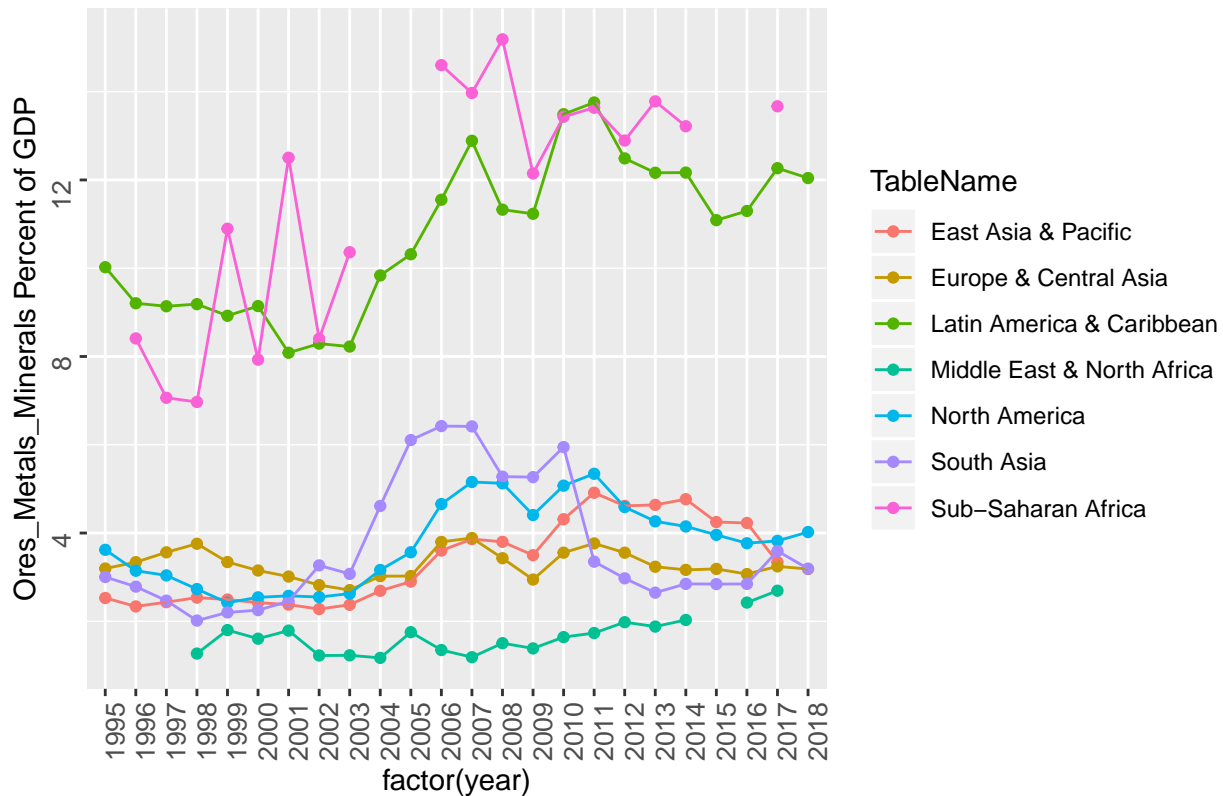
```
### Ores_Metals_minerals Percent of GDP
```

```
ggplot(economy_by_region_df, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Ores_Metals_Minerals Percent of GDP`)) +
  geom_line(aes(y = `Ores_Metals_Minerals Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Ores_Metals_minerals Percent of GDP by Region")
```

```
## Warning: Removed 12 rows containing missing values (geom_point).
```

```
## Warning: Removed 7 rows containing missing values (geom_path).
```

Ores_Metals_minerals Percent of GDP by Region



Data imputation remained a challenge with Minerals but still the can see the trends in the above graph where wealthier regions are moving away from the Minerals' production while the developing nations are still very heavily reliant on this industry. An anomaly was observed in regards to the Middle East and Sub-saharan Africa, it is because the data did not include the GDP contributions from Oil or Gas exports.

Knowledge vs traditinoal GDP

```
knowledge_traditinoal_SD_1to25percent_dF <- knowledge_traditinoal_dF %>%
  filter(.,`Country SD GDP in percent` <= 25) %>%
  arrange(TableName, year)

knowledge_traditinoal_SD_25to35percent_dF <- knowledge_traditinoal_dF %>%
  filter(.,`Country SD GDP in percent` > 25 & `Country SD GDP in percent` <=35) %>%
  arrange(TableName, year)

knowledge_traditinoal_SD_35to45percent_dF <- knowledge_traditinoal_dF %>%
  filter(.,`Country SD GDP in percent` > 35 & `Country SD GDP in percent` <=45) %>%
  arrange(TableName, year)

knowledge_traditinoal_SD_45to55percent_dF <- knowledge_traditinoal_dF %>%
  filter(.,`Country SD GDP in percent` > 45 & `Country SD GDP in percent` <=55) %>%
  arrange(TableName, year)

knowledge_traditinoal_SD_55to65percent_dF <- knowledge_traditinoal_dF %>%
  filter(.,`Country SD GDP in percent` > 55 & `Country SD GDP in percent` <=65) %>%
```



```

  arrange(TableName, year)

knowledge_traditinoal_SD_65to75percent_dF <- knowledge_traditinoal_dF %>%
  filter(`Country SD GDP in percent` > 65 & `Country SD GDP in percent` <=75) %>%
  arrange(TableName, year)

knowledge_traditinoal_SD_75to100percent_dF <- knowledge_traditinoal_dF %>%
  filter(`Country SD GDP in percent` > 75) %>%
  arrange(TableName, year)

SD_Groups <- c("knowledge_traditinoal_SD_1to25percent_dF", "knowledge_traditinoal_SD_25to35percent_dF", "I

```

Number of SD percent groups

```

NSDG <- length(SD_Groups)

valueBox(NSDG, color = "info")

```

7

```

### Recessions

valueBox("2001, 2008-2009", color = "info")

```

2001, 2008-2009

World chart

Below are the results based on countries around the world where the data is separated in two main groups namely knowledge-based (Services and Manufacturing) and traditional (Agriculture and Minerals) economies:

```

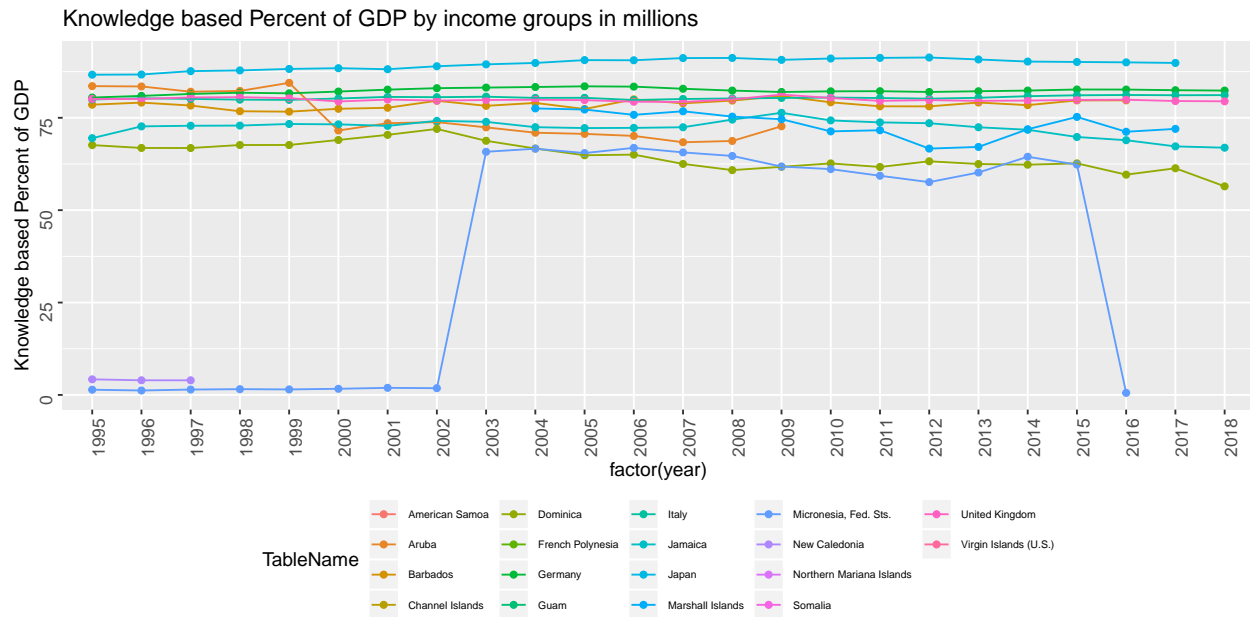
### SD upto 25 percent

ggplot(knowledge_traditinoal_SD_1to25percent_dF, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Knowledge based Percent of GDP`)) +
  geom_line(aes(y = `Knowledge based Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Knowledge based Percent of GDP by income groups in millions") +
  theme(legend.position = "bottom", legend.text = element_text(size=6), legend.margin = margin(t = 0, u

```

```
## Warning: Removed 213 rows containing missing values (geom_point).
```

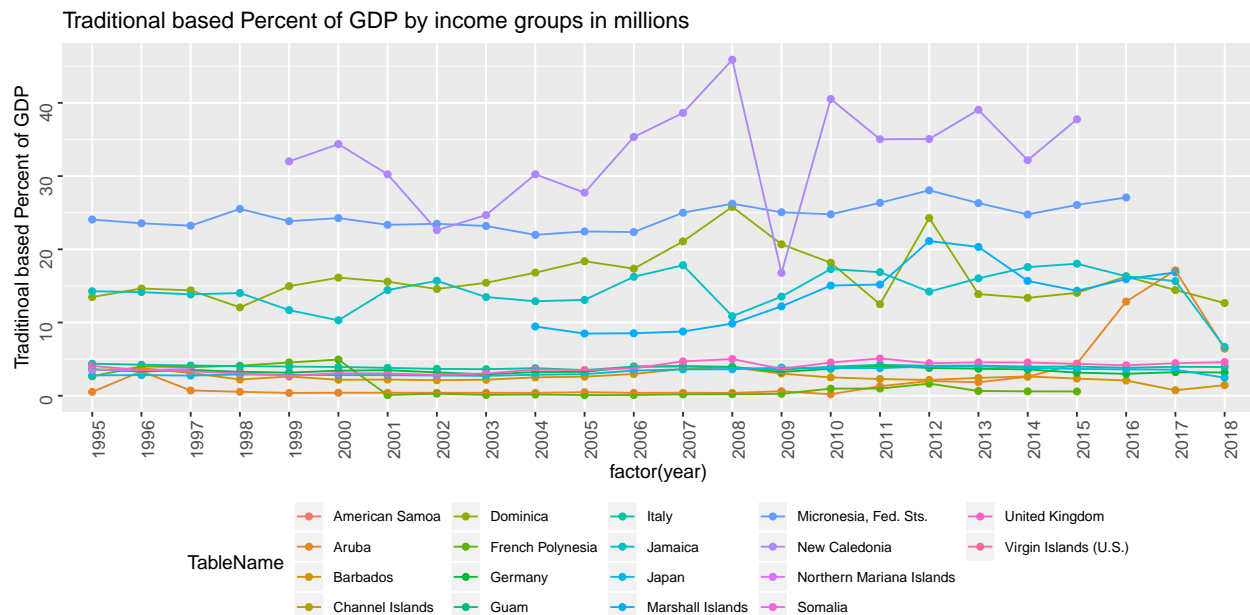
```
## Warning: Removed 213 rows containing missing values (geom_path).
```



```
ggplot(knowledge_traditinoal_SD_1to25percent_dF, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Traditinoal based Percent of GDP`)) +
  geom_line(aes(y = `Traditinoal based Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Traditional based Percent of GDP by income groups in millions") +
  theme(legend.position = "bottom", legend.text = element_text(size=8), legend.margin = margin(t = 0, u
```

Warning: Removed 163 rows containing missing values (geom_point).

Warning: Removed 162 rows containing missing values (geom_path).

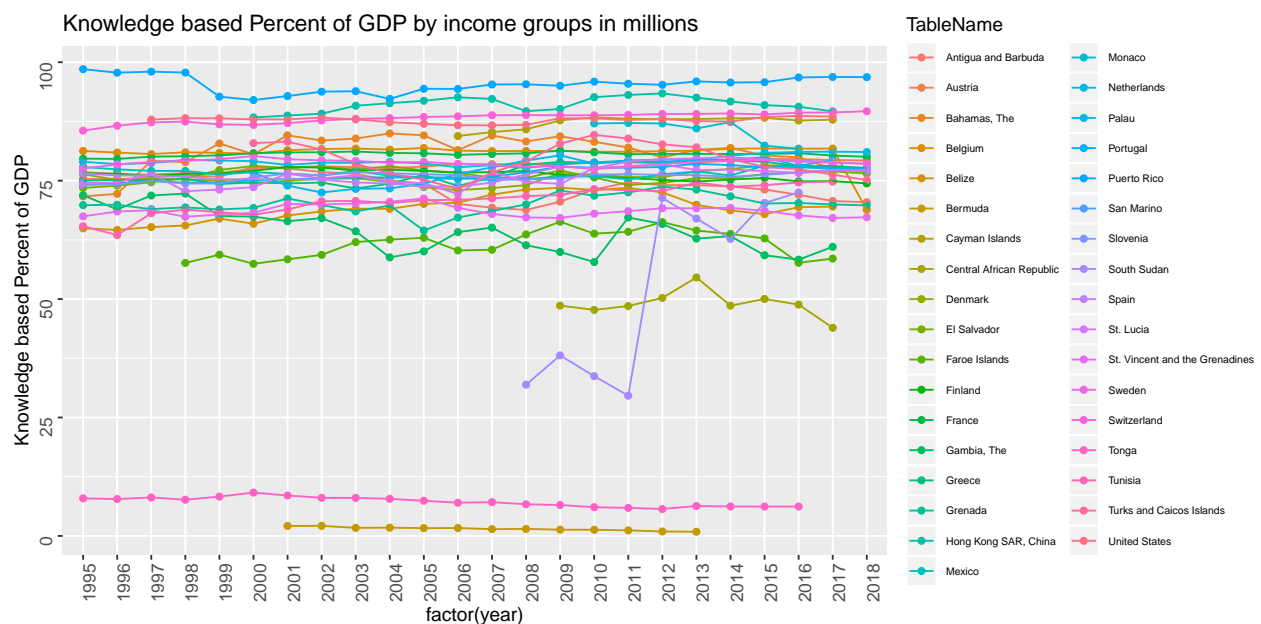


```
### SD >25 and upto 35 percent
```

```
ggplot(knowledge_traditinoal_SD_25to35percent_dF, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Knowledge based Percent of GDP`)) +
  geom_line(aes(y = `Knowledge based Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Knowledge based Percent of GDP by income groups in millions") +
  theme(legend.text = element_text(size=6), legend.margin = margin(t = 0, unit='cm'))
```

```
## Warning: Removed 124 rows containing missing values (geom_point).
```

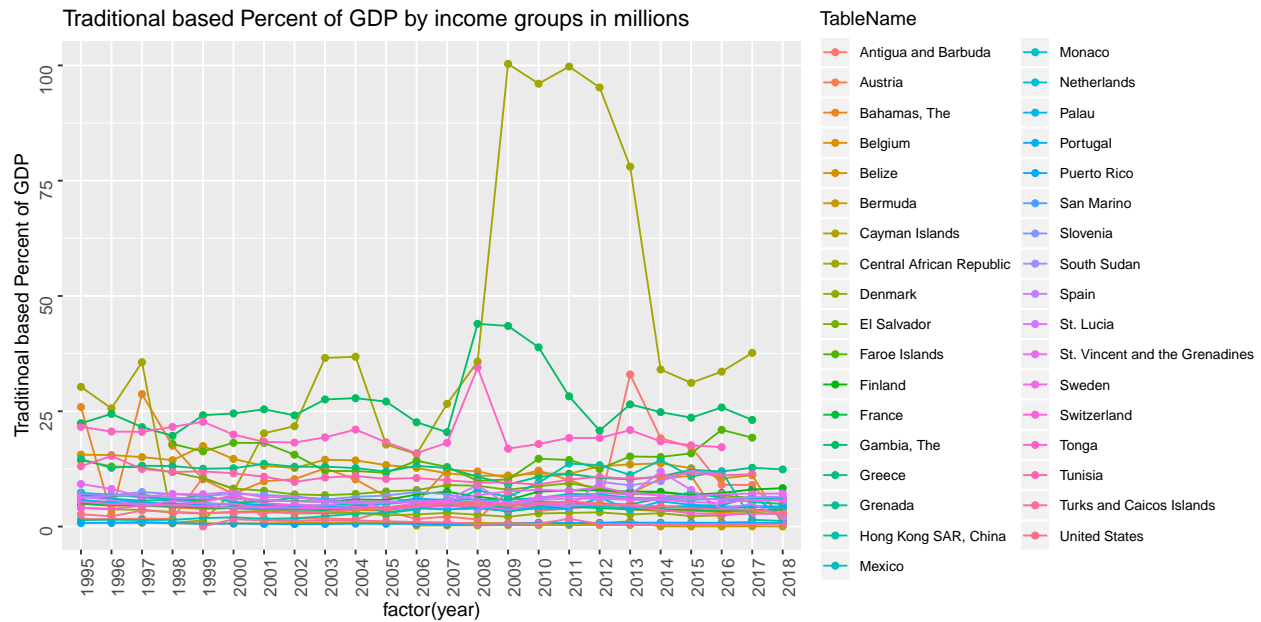
```
## Warning: Removed 124 rows containing missing values (geom_path).
```



```
ggplot(knowledge_traditinoal_SD_25to35percent_dF, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Traditinoal based Percent of GDP`)) +
  geom_line(aes(y = `Traditinoal based Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Traditional based Percent of GDP by income groups in millions") +
  theme(legend.text = element_text(size=8), legend.margin = margin(t = 0, unit='cm'))
```

```
## Warning: Removed 95 rows containing missing values (geom_point).
```

```
## Warning: Removed 95 rows containing missing values (geom_path).
```

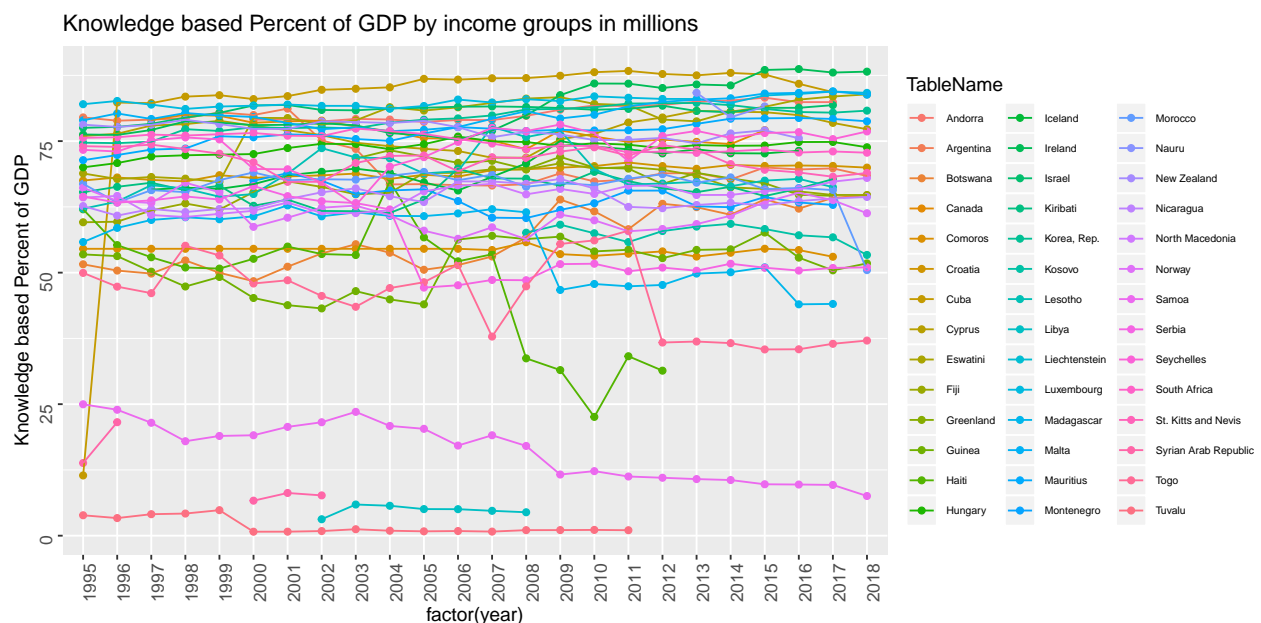


SD >35 upto 45 percent

```
ggplot(knowledge_traditinoal_SD_35to45percent_dF, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Knowledge based Percent of GDP`)) +
  geom_line(aes(y = `Knowledge based Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Knowledge based Percent of GDP by income groups in millions") +
  theme(legend.text = element_text(size=6), legend.margin = margin(t = 0, unit='cm'))
```

Warning: Removed 145 rows containing missing values (geom_point).

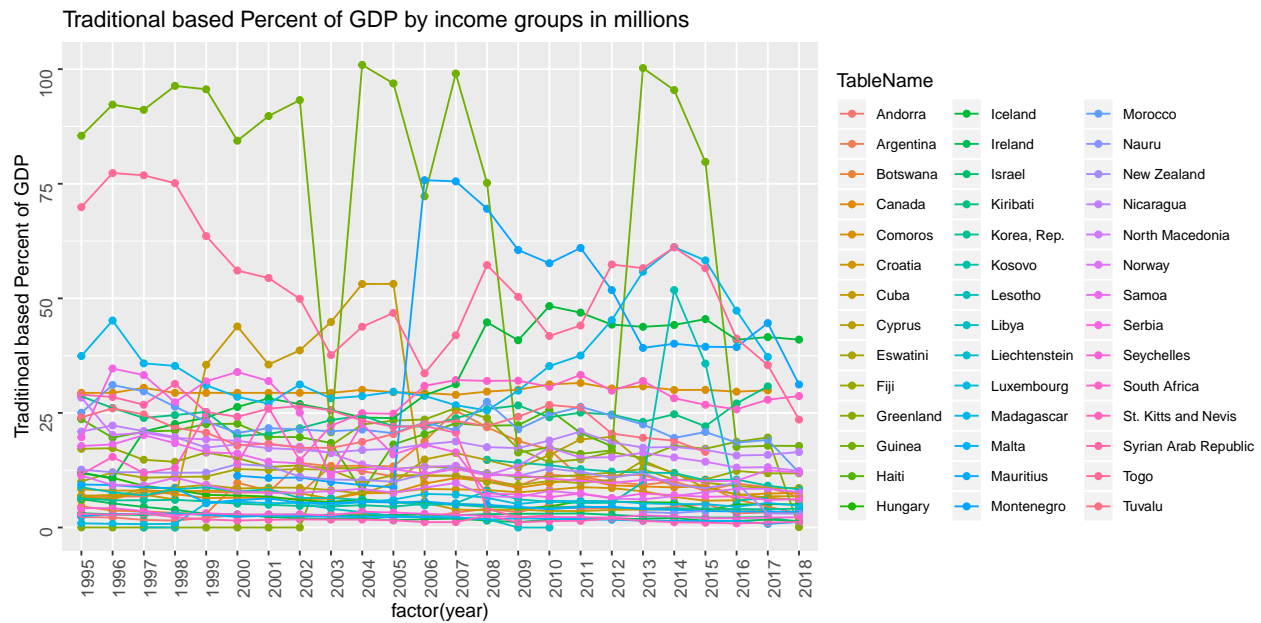
Warning: Removed 142 rows containing missing values (geom_path).



```
ggplot(knowledge_traditinoal_SD_35to45percent_dF, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Traditinoal based Percent of GDP`)) +
  geom_line(aes(y = `Traditinoal based Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Traditional based Percent of GDP by income groups in millions") +
  theme(legend.text = element_text(size=8), legend.margin = margin(t = 0, unit='cm'))
```

Warning: Removed 98 rows containing missing values (geom_point).

Warning: Removed 95 rows containing missing values (geom_path).



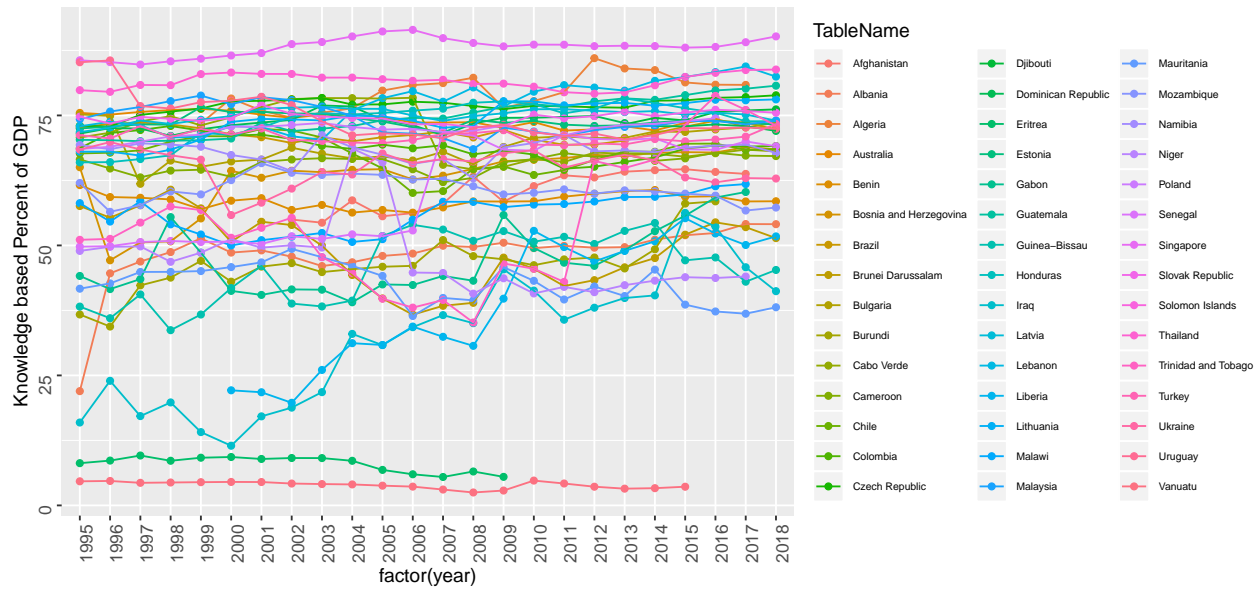
SD >45 upto 55 percent

```
ggplot(knowledge_traditinoal_SD_45to55percent_dF, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Knowledge based Percent of GDP`)) +
  geom_line(aes(y = `Knowledge based Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Knowledge based Percent of GDP by income groups in millions") +
  theme(legend.text = element_text(size=6), legend.margin = margin(t = 0, unit='cm'))
```

Warning: Removed 79 rows containing missing values (geom_point).

Warning: Removed 79 rows containing missing values (geom_path).

Knowledge based Percent of GDP by income groups in millions

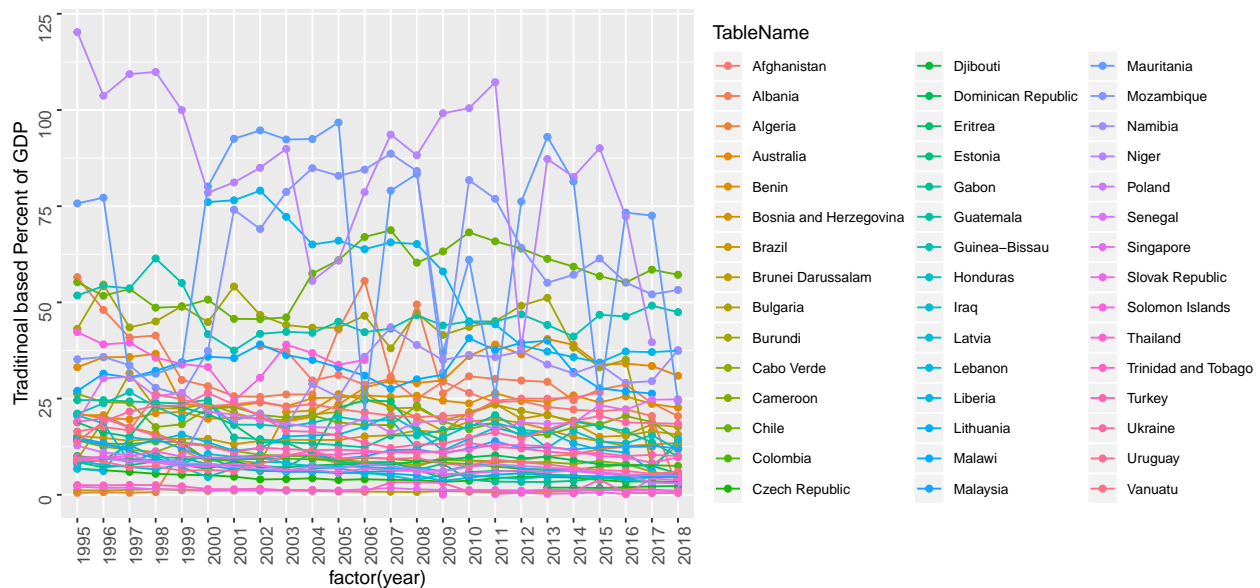


```
ggplot(knowledge_traditinoal_SD_45to55percent_dF, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Traditinoal based Percent of GDP`)) +
  geom_line(aes(y = `Traditinoal based Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Traditional based Percent of GDP by income groups in millions") +
  theme(legend.text = element_text(size=8), legend.margin = margin(t = 0, unit='cm'))
```

Warning: Removed 48 rows containing missing values (geom_point).

Warning: Removed 42 rows containing missing values (geom_path).

Traditional based Percent of GDP by income groups in millions

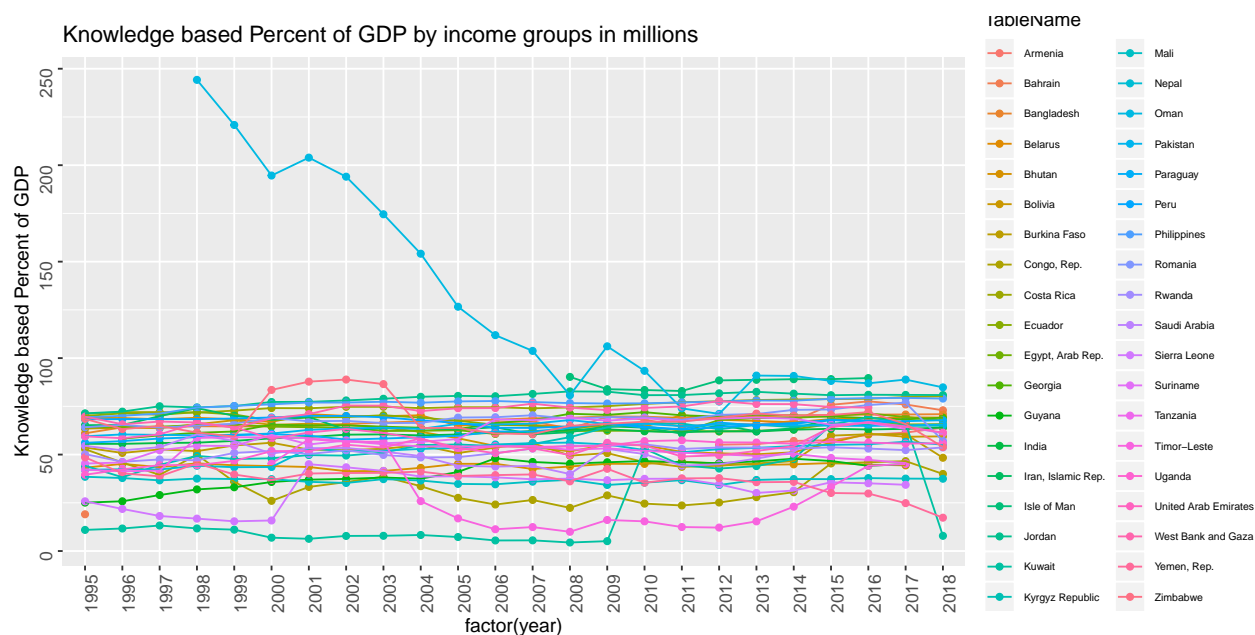


```
### SD >55 upto 65 percent
```

```
ggplot(knowledge_traditinoal_SD_55to65percent_dF, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Knowledge based Percent of GDP`)) +
  geom_line(aes(y = `Knowledge based Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Knowledge based Percent of GDP by income groups in millions") +
  theme(legend.text = element_text(size=6), legend.margin = margin(t = 0, unit='cm'))
```

```
## Warning: Removed 61 rows containing missing values (geom_point).
```

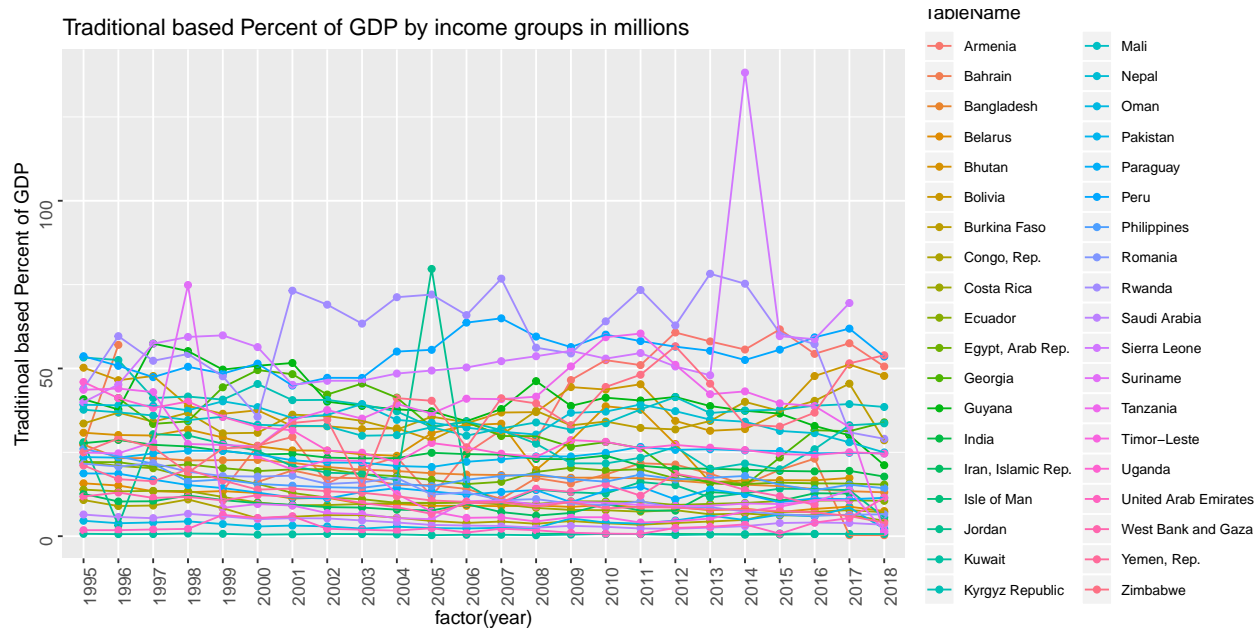
```
## Warning: Removed 51 rows containing missing values (geom_path).
```



```
ggplot(knowledge_traditinoal_SD_55to65percent_dF, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Traditinoal based Percent of GDP`)) +
  geom_line(aes(y = `Traditinoal based Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Traditional based Percent of GDP by income groups in millions") +
  theme(legend.text = element_text(size=8), legend.margin = margin(t = 0, unit='cm'))
```

```
## Warning: Removed 33 rows containing missing values (geom_point).
```

```
## Warning: Removed 29 rows containing missing values (geom_path).
```

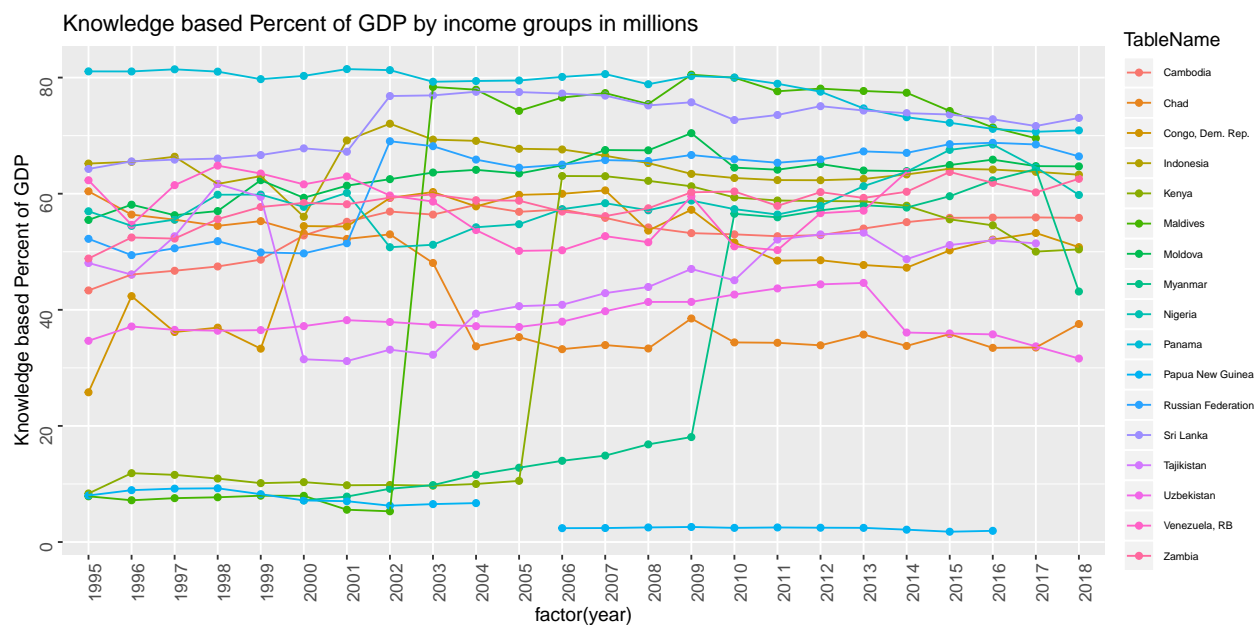



SD >65 upto 75 percent

```
ggplot(knowledge_traditinoal_SD_65to75percent_dF, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Knowledge based Percent of GDP`)) +
  geom_line(aes(y = `Knowledge based Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Knowledge based Percent of GDP by income groups in millions") +
  theme(legend.text = element_text(size=6), legend.margin = margin(t = 0, unit='cm'))
```

Warning: Removed 14 rows containing missing values (geom_point).

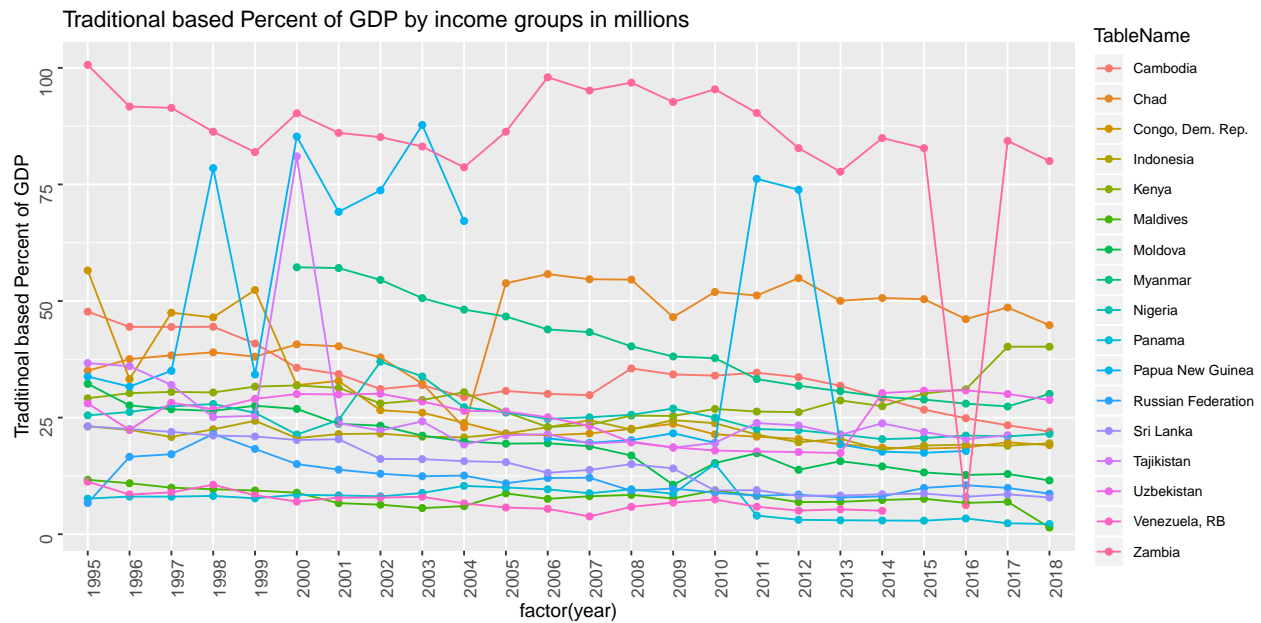
Warning: Removed 13 rows containing missing values (geom_path).




```
ggplot(knowledge_traditinoal_SD_65to75percent_dF, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Traditinoal based Percent of GDP`)) +
  geom_line(aes(y = `Traditinoal based Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Traditional based Percent of GDP by income groups in millions") +
  theme(legend.text = element_text(size=8), legend.margin = margin(t = 0, unit='cm'))
```

Warning: Removed 13 rows containing missing values (geom_point).

Warning: Removed 12 rows containing missing values (geom_path).

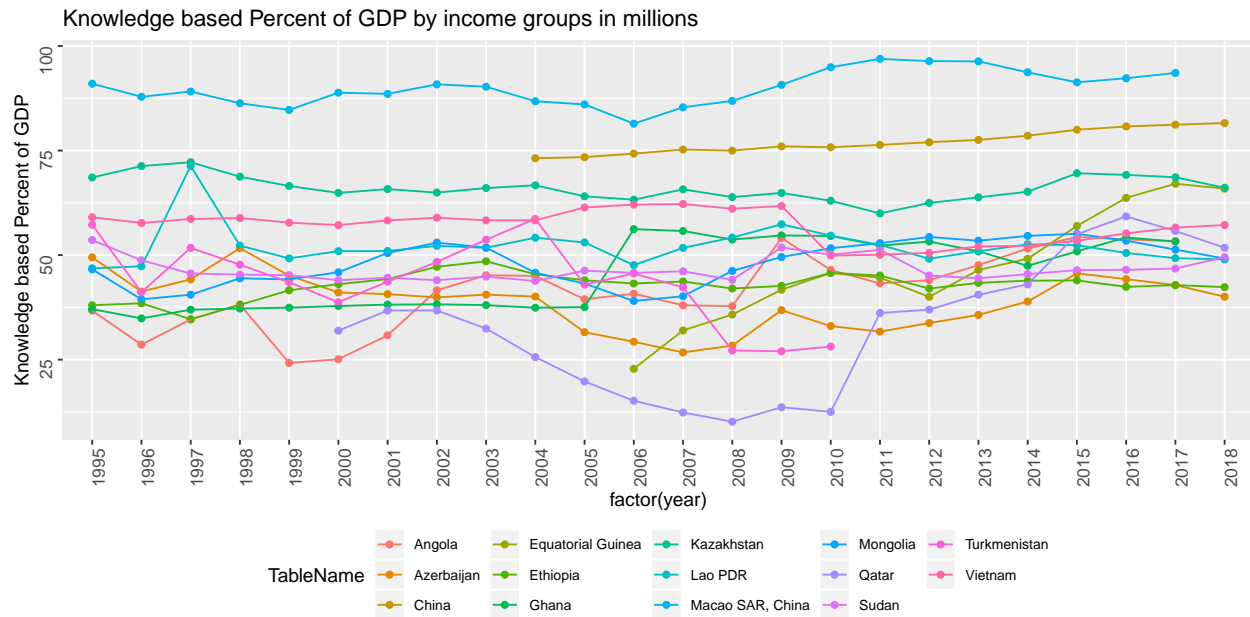


SD >75 percent

```
ggplot(knowledge_traditinoal_SD_75to100percent_dF, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Knowledge based Percent of GDP`)) +
  geom_line(aes(y = `Knowledge based Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Knowledge based Percent of GDP by income groups in millions") +
  theme(legend.position = "bottom", legend.text = element_text(size=8), legend.margin = margin(t = 0, unit='cm'))
```

Warning: Removed 36 rows containing missing values (geom_point).

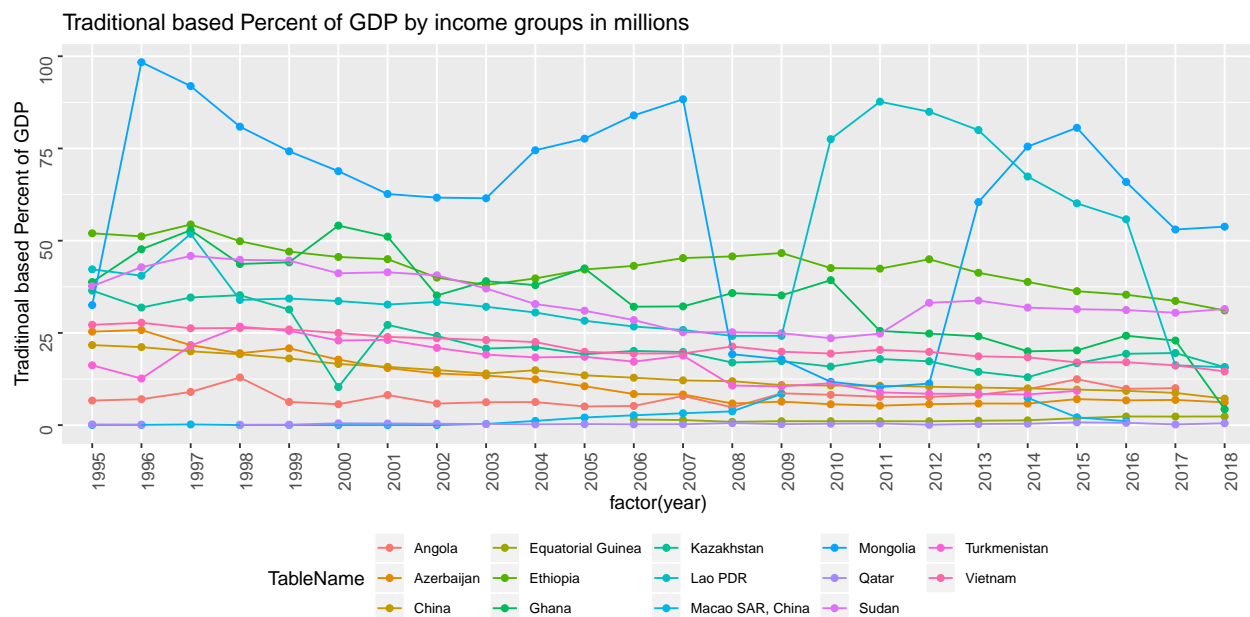
Warning: Removed 36 rows containing missing values (geom_path).



```
ggplot(knowledge_traditinoal_SD_75to100percent_dF, aes(x=factor(year), colour=TableName, group = TableName)) +
  geom_point(aes(y = `Traditinoal based Percent of GDP`)) +
  geom_line(aes(y = `Traditinoal based Percent of GDP`)) +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  theme(axis.text.y = element_text(size=10, angle=90)) +
  labs(title = "Traditional based Percent of GDP by income groups in millions") +
  theme(legend.position = "bottom", legend.text = element_text(size=8), legend.margin = margin(t = 0, u
```

Warning: Removed 22 rows containing missing values (geom_point).

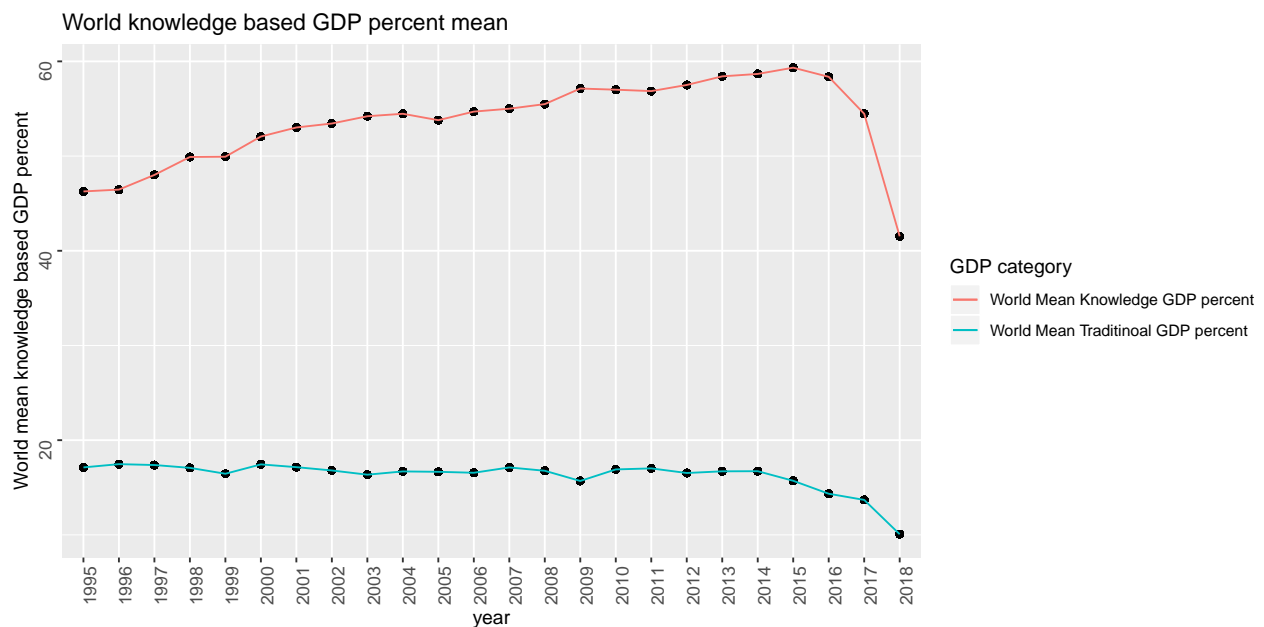
Warning: Removed 17 rows containing missing values (geom_path).



Conclusion

Below is the conclusive graph that illustrates the changes in the mean of the knowledge-based and traditional as a percentage of GDP over the years. We can see that the knowledge-based industries are dropping more rapidly. It is mainly due to the continuing drop of manufacturing to the GDP. The graph for the traditional insutries in more in line with what the initial understanding of these types of industries.

```
ggplot(knowledge_traditinoal_dF, aes(x=factor(year), group = 1)) +  
  geom_point(aes(y = `World Mean Knowledge GDP percent`)) +  
  geom_line(aes(y = `World Mean Knowledge GDP percent`, colour = "1")) +  
  geom_point(aes(y = `World Mean Traditinoal GDP percent`)) +  
  geom_line(aes(y = `World Mean Traditinoal GDP percent`, colour = "2")) +  
  theme(axis.text.x = element_text(size=10, angle=90)) +  
  theme(axis.text.y = element_text(size=10, angle=90)) +  
  labs(title = "World knowledge based GDP percent mean") +  
  xlab("year") +  
  ylab("World mean knowledge based GDP percent") +  
  scale_color_discrete(name = "GDP category", labels = c("World Mean Knowledge GDP percent", "World Mean Traditinoal GDP percent"))
```



References

1. GDP

<https://data.worldbank.org/indicator/NY.GDP.MKTP.CD>

2. GDP per Capita

<https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>

3. Agriculture, forestry, and fishing as % of GDP

<https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS>

4. Ores and Metals exports (% of merchandise exports) taken as a proxy to mineral production and exports

<https://data.worldbank.org/indicator/TX.VAL.MMTL.ZS.UN>

5. Service, value added (% of GDP)

<https://data.worldbank.org/indicator/NV.SRV.TOTL.ZS>

6. Manufacturing, value added (% of GDP)

<https://data.worldbank.org/indicator/NV.IND.MANF.ZS>

For more information on demographic dividend, please refer to the following link:

<https://www.imf.org/external/pubs/ft/fandd/2006/09/basics.htm>