Midterm project

Shicheng Wang

10/16/2019

Introduction

As the world pace into the information age and the worldwide education has access to upper level, there are more and more innovations created over recent 40 years. According to the world intellectual property organization (WIPO) said on 15th Oct,2019, the Asia filed two-thirds of global patent, trademark and industrial design applications last year, outperforming the rest of the world in technology development. In the meantime, the US remains the country most interested in expanding into new markets. Us residents filed the most overseas patents in 2018, followed by Japanese and German residents. For the general scale, the aim of this report is to futher explore global patent application status and research on relationship between GDP and technology standard.

Materials and methods

I exact the data from https://data.worldbank.org and concentrate on the science&technology sector. I take the patent applications(both resident and non-resident), scientific and technical journal articles, research and developmemnt expenditure and high-technology exports. Additionally, i decide to add the variable GDP (current US$) form the Economy & Growth sector into the dataset and plan to figure out how science and technology could influence the GDP.Futhermore, it is acknowledged that technology performance vary form country and year, the The following data analyse steps show below.

```
library(knitr)
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------------------------------------------------ t

## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## -- Conflicts ---------------------------------------------------------------------------------- tidyve
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows
```

Access data

Using the WDI package and function to get data from world bank database.

```
#install the package
library(WDI)
new_wdi_cache<-WDIcache()
```

```r
#find out the IDs for the indicators we are intersted in
WDIsearch("Patent applications.")
```

```
##      indicator     name
## [1,] "IP.PAT.RESD" "Patent applications, residents"
## [2,] "IP.PAT.NRES" "Patent applications, nonresidents"
```

```r
WDIsearch("Scientific and technical.")
```

```
##                                    indicator
##                             "IP.JRN.ARTC.SC"
##                                         name
## "Scientific and technical journal articles"
```

```r
WDIsearch("Research and development expenditure.")
```

```
##      indicator           name
## [1,] "GB.XPD.RSDV.GN.ZS" "Research and development expenditure (% of GNI)"
## [2,] "GB.XPD.RSDV.GD.ZS" "Research and development expenditure (% of GDP)"
```

```r
WDIsearch("High-technology exports.")
```

```
##      indicator
## [1,] "TX.VAL.TECH.MF.ZS"
## [2,] "TX.VAL.TECH.MANF.ZS"
## [3,] "TX.VAL.TECH.CD"
##      name
## [1,] "High-technology exports (% of manufactured exports)"
## [2,] "High-technology exports (% of manufactured exports)"
## [3,] "High-technology exports (current US$)"
```

```r
WDIsearch("gdp.*US\\$", cache = new_wdi_cache)
```

```
##       indicator
##  [1,] "EG.GDP.PUSE.KO.87"
##  [2,] "EG.GDP.PUSE.KO.KD"
##  [3,] "EU.EGY.USES.GDP"
##  [4,] "NV.SRV.DISC.CD"
##  [5,] "NY.GDP.DISC.CD"
##  [6,] "NY.GDP.FCST.KD.87"
##  [7,] "NY.GDP.MKTP.CD"
##  [8,] "NY.GDP.MKTP.CD.XD"
##  [9,] "NY.GDP.MKTP.KD"
## [10,] "NY.GDP.MKTP.KD.87"
## [11,] "NY.GDP.PCAP.CD"
## [12,] "NY.GDP.PCAP.KD"
## [13,] "NYGDPMKTPSACD"
## [14,] "NYGDPMKTPSAKD"
##       name
##  [1,] "GDP per unit of energy use (1987 US$ per kg of oil equivalent)"
##  [2,] "GDP per unit of energy use (2000 US$ per kg of oil equivalent)"
##  [3,] "GDP per unit of energy use (1987 US$ per kg of oil equivalent)"
##  [4,] "Discrepancy in GDP, value added (current US$)"
##  [5,] "Discrepancy in expenditure estimate of GDP (current US$)"
##  [6,] "GDP at factor cost (constant 1987 US$)"
##  [7,] "GDP (current US$)"
```

```
##  [8,] "GDP deflator, index (2000=100; US$ series)"
##  [9,] "GDP (constant 2010 US$)"
## [10,] "GDP at market prices (constant 1987 US$)"
## [11,] "GDP per capita (current US$)"
## [12,] "GDP per capita (constant 2010 US$)"
## [13,] "GDP,current US$,millions,seas. adj.,"
## [14,] "GDP,constant 2010 US$,millions,seas. adj.,"
```

```
#download the data
wdi_dat <- WDI(indicator =
c("IP.PAT.RESD", "IP.PAT.NRES", "GB.XPD.RSDV.GD.ZS","TX.VAL.TECH.CD","NY.GDP.MKTP.CD"), start = 1968, en
names(wdi_dat)
```

```
##  [1] "iso2c"            "country"          "year"
##  [4] "IP.PAT.RESD"      "IP.PAT.NRES"      "GB.XPD.RSDV.GD.ZS"
##  [7] "TX.VAL.TECH.CD"   "NY.GDP.MKTP.CD"   "iso3c"
## [10] "region"           "capital"          "longitude"
## [13] "latitude"         "income"           "lending"
```

```
head(wdi_dat)
```

```
##   iso2c    country year IP.PAT.RESD IP.PAT.NRES GB.XPD.RSDV.GD.ZS
## 1    1A Arab World 1999         809          NA                NA
## 2    1A Arab World 1989         549        1202                NA
## 3    1A Arab World 1990          NA          NA                NA
## 4    1A Arab World 1987          NA          NA                NA
## 5    1A Arab World 1988          NA          NA                NA
## 6    1A Arab World 1993         702        1815                NA
##   TX.VAL.TECH.CD NY.GDP.MKTP.CD iso3c     region capital longitude
## 1             NA   644050661585   ARB Aggregates
## 2             NA   322859532932   ARB Aggregates
## 3             NA   446901684421   ARB Aggregates
## 4             NA   313040206370   ARB Aggregates
## 5             NA   307889062413   ARB Aggregates
## 6             NA   477133962042   ARB Aggregates
##   latitude     income    lending
## 1          Aggregates Aggregates
## 2          Aggregates Aggregates
## 3          Aggregates Aggregates
## 4          Aggregates Aggregates
## 5          Aggregates Aggregates
## 6          Aggregates Aggregates
```

Data cleaning

```
#remove aggregated regional values.
wdi_dat <- subset(wdi_dat, region != "Aggregates")

#rename variables
names(wdi_dat)[which(names(wdi_dat) == "IP.PAT.RESD")] <- "apply_resident"
names(wdi_dat)[which(names(wdi_dat) == "IP.PAT.NRES")] <- "apply_no_resident"
names(wdi_dat)[which(names(wdi_dat) == "GB.XPD.RSDV.GD.ZS")] <- "expenditure_gdp"
names(wdi_dat)[which(names(wdi_dat) == "TX.VAL.TECH.CD")] <- "export"
names(wdi_dat)[which(names(wdi_dat) == "NY.GDP.MKTP.CD")] <- "GDP"
```

```
#clean NA data
da = na.omit(wdi_dat)
dim(da)
```

```
## [1] 711  15
```

Innovation standard worldwide

World economic table

```
#process data
da['apply'] = da$apply_resident+da$apply_no_resident
growth <- da %>% group_by(year) %>%
  summarize(apply=mean(apply),expenditure_gdp=mean(expenditure_gdp),
            export=mean(export),GDP=mean(GDP))

#variables transformation
growth$year <- growth$year
growth$apply <- growth$apply/1000
growth$expenditure_gdp <- growth$expenditure_gdp*10
growth$export <- growth$export*10
growth$GDP <- growth$GDP/100000000000
```

Produce a table

```
kable(growth, digits = 4, align = "c",booktabs=TRUE ,
caption = "Growth worldwide",col.names = c("year","apply", "expentiture","export","GDP" )) %>% kable_sty
```

Table 1: Growth worldwide

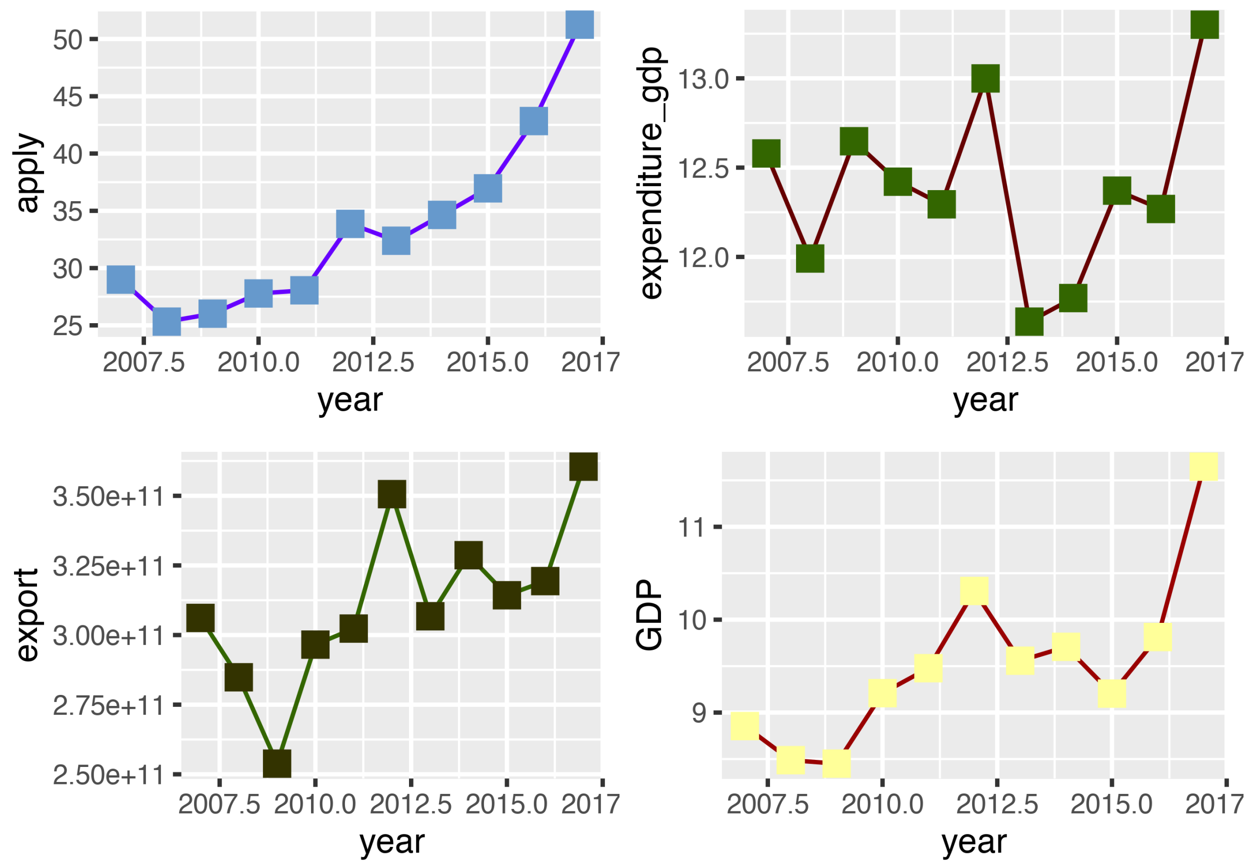
year	apply	expentiture	export	GDP
2007	28.9933	12.5793	306141616367	8.8546
2008	25.3250	11.9909	284827370061	8.4876
2009	26.0204	12.6476	253818444087	8.4502
2010	27.7847	12.4211	296558382104	9.2090
2011	28.0607	12.2944	302322180200	9.4777
2012	33.8594	12.9994	350601450821	10.3092
2013	32.3880	11.6374	306742623790	9.5583
2014	34.6429	11.7695	328670413233	9.7076
2015	36.9610	12.3710	314283416703	9.2041
2016	42.8358	12.2683	319369572092	9.8131
2017	51.2508	13.3005	360465266684	11.6473

The table above shows the average apply amounts/Research and development expenditure/high-technology export/GDP of the whole world, the next analysis based on this table.

Variables trend

```
#patent applications trend
p1 = ggplot(growth, aes(x=year, y=apply)) + geom_line(colour='#6600FF') +geom_point(size=4,colour='#669

#expentiture trend
p2 = ggplot(growth, aes(x=year, y=expenditure_gdp)) + geom_line(colour='#660000') +geom_point(size=4,co
```

```
#export trend
p3 = ggplot(growth, aes(x=year, y=export)) + geom_line(colour='#336600') +geom_point(size=4,colour='#33

#GDP trend
p4 = ggplot(growth, aes(x=year, y=GDP)) + geom_line(colour='#990000')+
  geom_point(size=4,colour='#FFFF99',shape=15)


gridExtra::grid.arrange(p1,p2,p3,p4,ncol = 2)
```



After processing data from 2007 to 2017, we can see the world GDP have been fluctrating in one decade. According to these plots, we can see some facts: During 2008-2009, there are apparently sharps in these variables, we all know the subprime crisis happened in America along with a worldwide depression. Accordingly, i guess this crisis could lead to these sharps of export/GDP/expenditure/apply status. During 2012-2017, world might develop steadily due to less wars and crisises.

Country's innovation level

This session's target is to figure out the relationships among existing variables.

How income/ effect country's innovation status

```
income <- da %>% group_by(income,year) %>% summarize(apply=mean(apply),expenditure_gdp=mean(expenditure_
# income and apply
```

```r
p5 <- ggplot(data=income, aes(x=year, y=apply,fill=income)) +
      geom_col(stat="identity",position = "dodge") + scale_x_continuous("year",breaks = seq(2007,2017,1
```

```
## Warning: Ignoring unknown parameters: stat
```

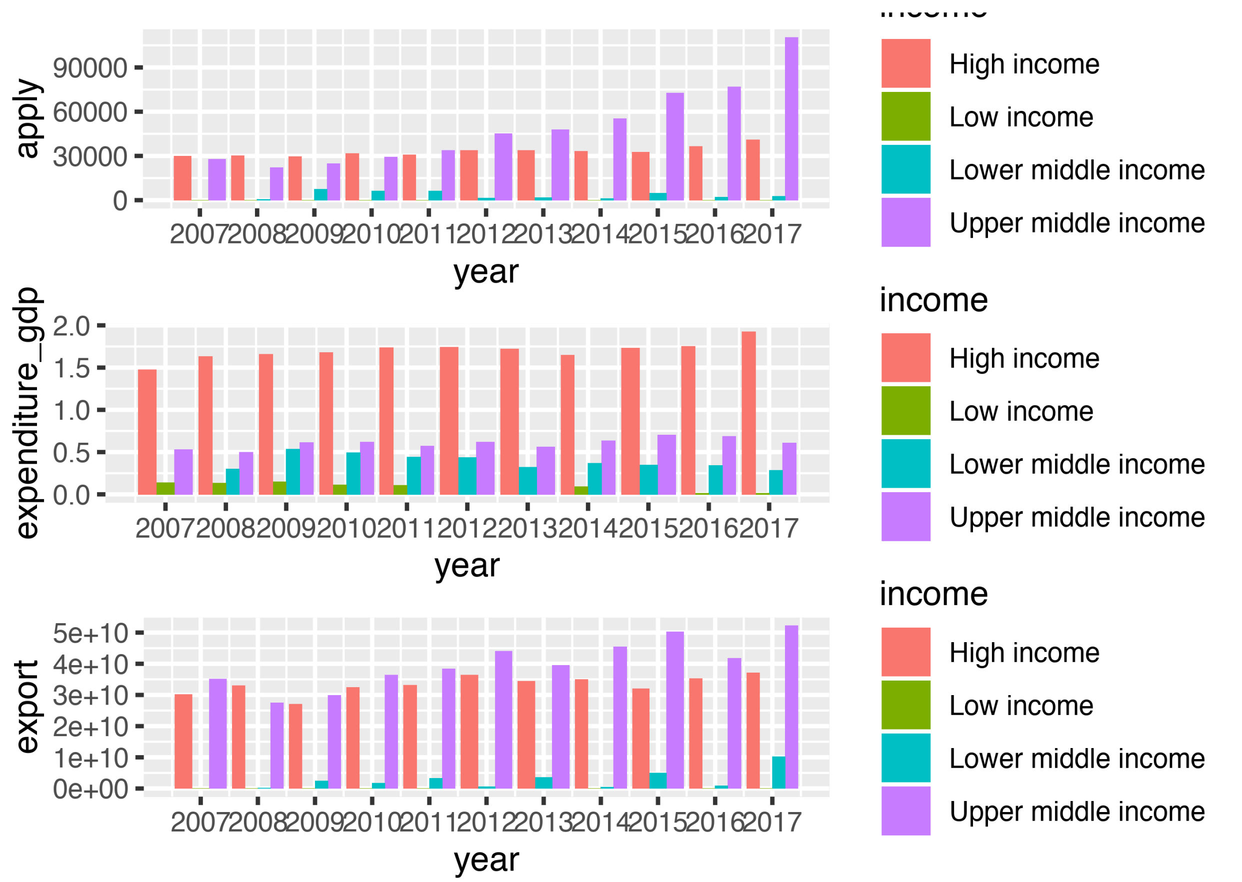
```r
# income and research and development expenditure (% of GDP)
p6 <- ggplot(data=income, aes(x=year, y=expenditure_gdp,fill=income)) +
      geom_col(stat="identity",position = "dodge") + scale_x_continuous("year",breaks = seq(2007,2017,1
```

```
## Warning: Ignoring unknown parameters: stat
```

```r
# income and high-technology exports (current US$)
p7 <- ggplot(data=income, aes(x=year, y=export,fill=income)) +
      geom_col(stat="identity",position = "dodge") + scale_x_continuous("year",breaks = seq(2007,2017,1
```

```
## Warning: Ignoring unknown parameters: stat
```

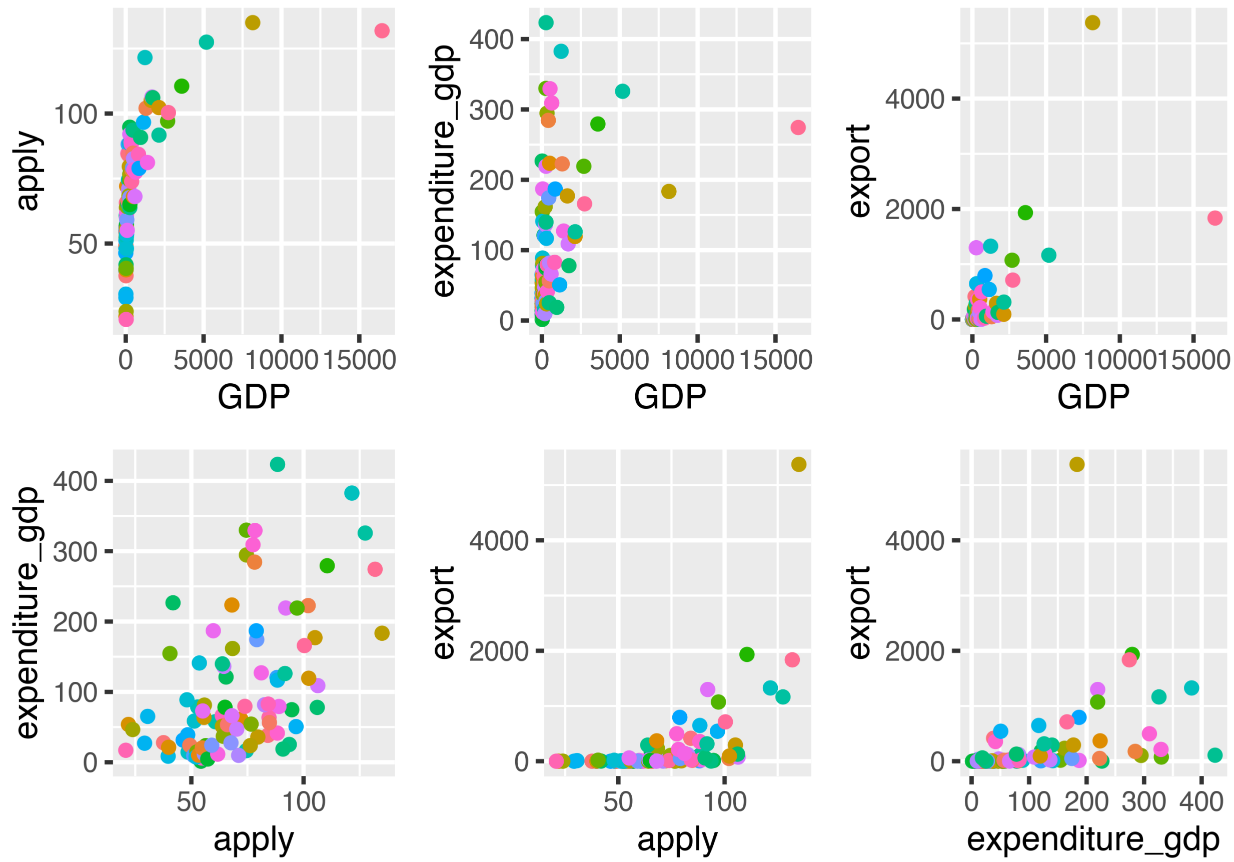
```r
gridExtra::grid.arrange(p5,p6,p7,ncol = 1)
```



From above three plots we can see the main contribution of innovation through 10 years are high-income group and upper middle income group. It can be ascribed to the reality that more income could bring with large research and education funds and correspondlingly these countries would have more suitable environment and more desire to study with meaningful topics. Low-income countries, by contrast, are not having good conditions to innovate.

Relationships between variables

```
#group and choose country
country <- da %>% group_by(country) %>%  summarize(apply=mean(apply),expenditure_gdp=mean(expenditure_g
#screening data by GDP
country <- country[order(country$GDP,decreasing=F),]
#variables transformation
country$apply <- 10*log(country$apply)
country$expenditure_gdp <- country$expenditure_gdp*100
country$export <- country$export/100000000
country$GDP <- country$GDP/1000000000

#GDP & apply
p7 <- ggplot(country, aes(x = GDP, y = apply, color = country)) +
geom_point() +theme(legend.position = 'none')
#GDP & expenditure
p8 <- ggplot(country, aes(x = GDP, y = expenditure_gdp, color = country)) +
geom_point() +theme(legend.position = 'none')
#GDP & export
p9 <- ggplot(country, aes(x = GDP, y = export, color = country)) +
geom_point() +theme(legend.position = 'none')
#apply & expenditure
p10 <- ggplot(country, aes(x = apply, y = expenditure_gdp, color = country)) +
geom_point() +theme(legend.position = 'none')
#apply & export
p11 <- ggplot(country, aes(x = apply, y = export, color = country)) +
geom_point() +theme(legend.position = 'none')
#expenditure & export
p12 <- ggplot(country, aes(x = expenditure_gdp, y = export, color = country)) +
geom_point() +theme(legend.position = 'none')

gridExtra::grid.arrange(p7,p8,p9,p10,p11,p12,ncol = 3)
```



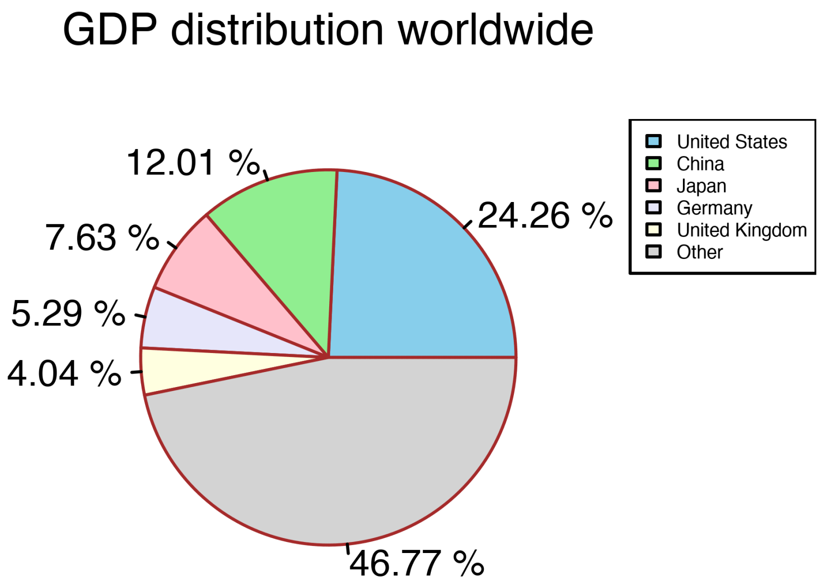
Due to there are mainly four variables in the dataset, so it could exist six relationships, so i make six pairs and present a pointplot respedctively. It is easily to perceive that there is no obvious relationships between apply number and expenditure. Meanwhile, the other pairs have obvious positive correlation. Additionally, expenditure and GDP relation plot might have two lines and i think it could be attributed to some other factors such as geographical location, climate, national policy, and the like.

GDP ranking of countries

Rank the data in order of the country's GDP and compare each line's GDP value with the 50% of total GDP of the world one by one. From the result, we can see the sum of the five largest countries countries all has already reached the 50% of world's GDP which is larger than the total GDP of the remaining countries.

Now produce a pieplot:

```r
a <- c(164.6,81.5,51.8,35.9,27.4,317.4)
b <- c("United States","China","Japan","Germany","United Kingdom","Other")
piepercent<- paste(round(100*a/sum(a), 2), "%")

pie(a,labels=piepercent,col= c("skyblue","lightgreen","pink","lavender","lightyellow","lightgrey"),
    radius = 0.85,border="brown",lty=6,main='GDP distribution worldwide')

legend("topright",b,cex=0.5,fill=c("skyblue","lightgreen","pink","lavender","lightyellow","lightgrey"))
```



From the pie plot we could see US, China, Japan, Germany, UK has relative high GDP which added uo to over 50% which more than the combined GDP of the rest of countries, so next analysis part concentrate on US and clearify some relationships of variables.

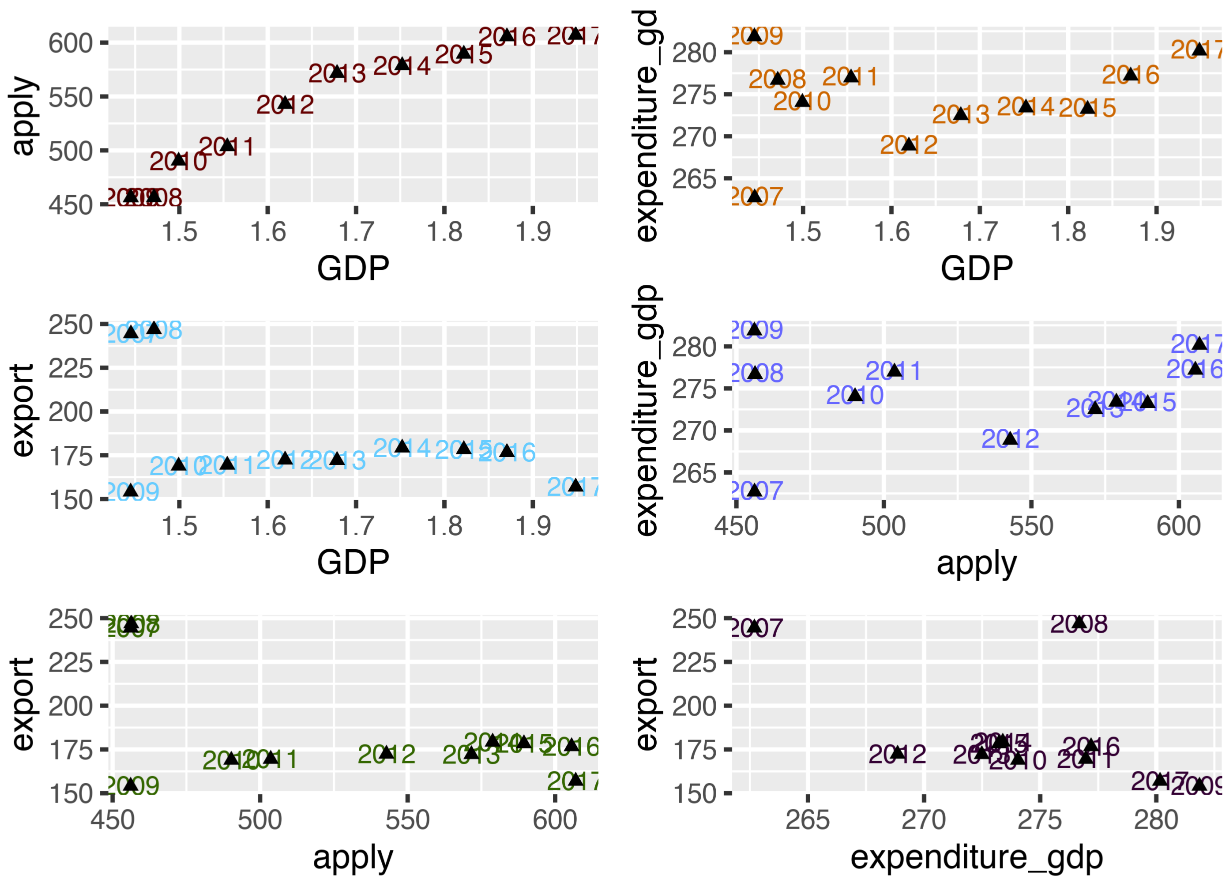
US

```
#creat a new dataframe
x <- filter(da, country == "United States") %>% arrange(year)
new <- x[,c(3,6,7,8,16)]

# transformation
new$year <- new$year
new$apply <- new$apply/1000
new$expenditure_gdp <- new$expenditure_gdp*100
new$export <- new$export/1000000000
new$GDP <- new$GDP/10000000000000

#US data chart
kable(new, digits = 4, align = "c",booktabs=TRUE ,
caption = "Us 10-year data",col.names = c("year", "expentiture","export","GDP","apply" )) %>% kable_styl
```

Table 2: Us 10-year data

year	expentiture	export	GDP	apply
2007	262.692	244.4796	1.4452	456.154
2008	276.683	246.8841	1.4713	456.321
2009	281.859	154.1085	1.4449	456.106
2010	274.046	168.9394	1.4992	490.226
2011	276.965	169.4642	1.5543	503.582
2012	268.859	172.3872	1.6197	542.815
2013	272.486	172.1454	1.6785	571.612
2014	273.389	179.2640	1.7522	578.802
2015	273.222	178.3495	1.8219	589.410
2016	277.191	176.6682	1.8707	605.571
2017	280.161	156.9371	1.9485	606.956

```r
#GDP & apply
u1 <- ggplot(new, aes(x = GDP, y = apply, color = year)) +
geom_text(aes(label = year), size=3, nudge_y = 0.2,colour='#660000') +
geom_point(colour='Black',shape=17)
#GDP & expenditure
u2 <- ggplot(new, aes(x = GDP, y = expenditure_gdp, color = year)) +
geom_text(aes(label = year), size=3, nudge_y = 0.2,colour='#CC6600') +
geom_point(colour='Black',shape=17)
#GDP & export
u3 <- ggplot(new, aes(x = GDP, y = export, color = year)) +
geom_text(aes(label = year), size=3, nudge_y = 0.2,colour='#66CCFF') +
geom_point(colour='Black',shape=17)
#apply & expenditure
u4 <- ggplot(new, aes(x = apply, y = expenditure_gdp, color = year)) +
geom_text(aes(label = year), size=3, nudge_y = 0.2,colour='#6666FF') +
geom_point(colour='Black',shape=17)
#apply & export
u5 <- ggplot(new, aes(x = apply, y = export, color = year)) +
geom_text(aes(label = year), size=3, nudge_y = 0.2,colour='#336600') +
geom_point(colour='Black',shape=17)
#expenditure & export
u6 <- ggplot(new, aes(x = expenditure_gdp, y = export, color = year)) +
geom_text(aes(label = year), size=3, nudge_y = 0.2,colour='#330033') +
geom_point(colour='Black',shape=17)

gridExtra::grid.arrange(u1,u2,u3,u4,u5,u6,ncol = 2)
```



In the first three charts, we can see as the GDP and year increasing, the export and expenditure correspondingly increase, however the export amount seems does not has the connection with GDP. In the second half, it does not appear any clear relationship.

conclusion

In general, this report have respectively conducted five surveys and researches on the both worldwide and US(a specific country) scale. According to these plots we could see the GDP could largely influence countries' technology development. However, the GDP polarized around the world, we can imagine some low income countries might not be in a good developing condition for a long time which could be a serious reality. What's more, the country's science and technology standard could fluctuates as the world economic situation change. How to deal with the crisis is an important question for every country.