DATA SCIENCE

Div: A		Batch: B2	
Name of the Students	Roll No.	Enrollment No.	
Shripad Govekar	41	12020200	
Mahesh Badade	39	12020264	
Rupesh Kadam	43	12020044	
Abhishek Patil	53	12020111	

Title: Using suitable data visualization tools (minimum 2) compare data of 5 appropriately contrasting countries

Statement: Data Visualization of capital of GDP and GDP per capita data of 5 appropriately contrasting countries using box plot, bar plot and pie chart.

Metadata:

Name of creator of data: United Nations Statistics Division, New York, National Accounts Statistics: Analysis of Main Aggregates (AMA) database
Set Name of author of document: Unknown
Title of document: SYB63_230_202009_GDP and GDP Per Capita
Location of file: data.un.org
Size of file: 1.40 MB
Format: csv (comma separated values)

Descriptions of column:
Region/Country/Area No: contains unique id given to each observation.
Region/Country/Area: Names of the countries.
Year: Year in which data was observed.
Series: type of data
Value: values as per series
Footnotes: Blank

R-Code:

```r
#1. Reference http: data.un.org. Data of GDP and GDP per capita
#2. Use suitable data visualization tools (minimum 2) to compare data of 5
appropriately contrasting countries.
#The assignment is to be implemented in R

data <- read.csv("SYB63_230_202009_GDP and GDP Per Capita.csv")
library("highcharter")
library(dplyr)
library(tidyverse)
library(ggplot2)
data <- data[-c(1),]

names(data)[1] <- "Region/Country/Area No"
names(data)[2] <- "Region/Country/Area"
names(data)[3] <- "Year"
names(data)[4] <- "Series"
names(data)[5] <- "Value"
names(data)[6] <- "Footnotes"
names(data)[7] <- "Source"


data[,c(3,5)] <- sapply(data[,c(3,5)], as.numeric)
summary(data)

#----------------------------GDP Growth of INDIA--------------------------
highchart() %>%
  hc_chart(type = "column") %>%
  hc_xAxis(
    title = list(text = "Annual Year"),
    alternateGridColor = "#FDFFD5",
    plotLines = list(
      list(
        label = list(text = "This is a plotLine"),
        color = "#FF0000",
        width = 2,
        value = 5.5
      )
    )
  ) %>%
  hc_title(text = "Data of GDP Growth in India") %>%
  hc_plotOptions(column = list(enableMouseTracking = TRUE)
  ) %>%
  hc_add_series(name="1985",data=data$Value[data$Series == "GDP in current
prices (millions of US dollars)" & data$`Region/Country/Area` == "India" &
data$Year == "1985"]) %>%
  hc_add_series(name="1995",data=data$Value[data$Series == "GDP in current
prices (millions of US dollars)" & data$`Region/Country/Area` == "India" &
```

```r
data$Year == "1995"]) %>%
  hc_add_series(name="2005",data=data$Value[data$Series == "GDP in current
prices (millions of US dollars)" & data$`Region/Country/Area` == "India" &
data$Year == "2005"]) %>%
  hc_add_series(name="2010",data=data$Value[data$Series == "GDP in current
prices (millions of US dollars)" & data$`Region/Country/Area` == "India" &
data$Year == "2010"]) %>%
  hc_add_series(name="2015",data=data$Value[data$Series == "GDP in current
prices (millions of US dollars)" & data$`Region/Country/Area` == "India" &
data$Year == "2015"]) %>%
  hc_add_series(name="2017",data=data$Value[data$Series == "GDP in current
prices (millions of US dollars)" & data$`Region/Country/Area` == "India" &
data$Year == "2017"]) %>%
  hc_add_series(name="2018",data=data$Value[data$Series == "GDP in current
prices (millions of US dollars)" & data$`Region/Country/Area` == "India" &
data$Year == "2018"])

#--------------------Data of GDP in 2018--------------------------
pc <- mean(data$Value[data$Series == "GDP in current prices (millions of US
dollars)" & data$`Region/Country/Area` == "China"])
print(pc)
highchart() %>%
  hc_chart(type = "column") %>%
  hc_xAxis(
    title = list(text = "5 appropriately contrasting countries"),
    alternateGridColor = "#FDFFD5",
    plotLines = list(
      list(
        label = list(text = "This is a plotLine"),
        color = "#FF0000",
        width = 2,
        value = 5.5
      )
    )
  ) %>%
  hc_title(text = "Data of GDP in 2018") %>%
  hc_plotOptions(column = list(enableMouseTracking = TRUE)
  ) %>%
  hc_add_series(name="India",data=tail(data$Value[data$Series == "GDP in
current prices (millions of US dollars)" & data$`Region/Country/Area` ==
"India"], n=1))%>%
  hc_add_series(name="Germany",data=tail(data$Value[data$Series == "GDP in
current prices (millions of US dollars)" & data$`Region/Country/Area` ==
"Germany"], n=1))%>%
  hc_add_series(name="Japan",data=tail(data$Value[data$Series == "GDP in
current prices (millions of US dollars)" & data$`Region/Country/Area` ==
"Japan"], n=1))%>%
  hc_add_series(name="China",data=tail(data$Value[data$Series == "GDP in
current prices (millions of US dollars)" & data$`Region/Country/Area` ==
"China"], n=1))%>%
  hc_add_series(name="USA",data=tail(data$Value[data$Series == "GDP in current
```

```r
prices (millions of US dollars)" & data$`Region/Country/Area` == "United States
of America"], n=1))

#------------------Data of GDP Growth rate in 2018--------------------
pc <- mean(data$Value[data$Series == "GDP real rates of growth (percent)" &
data$`Region/Country/Area` == "China"])
print(pc)
highchart() %>%
  hc_chart(type = "column") %>%
  hc_xAxis(
    title = list(text = "5 appropriately contrasting countries"),
    alternateGridColor = "#FDFFD5",
    plotLines = list(
      list(
        label = list(text = "This is a plotLine"),
        color = "#FF0000",
        width = 2,
        value = 5.5
      )
    )
  ) %>%
  hc_title(text = "Data of GDP Growth Rate in 2018") %>%
  hc_plotOptions(column = list(enableMouseTracking = TRUE)
  ) %>%
  hc_add_series(name="Japan",data=tail(data$Value[data$Series == "GDP real
rates of growth (percent)" & data$`Region/Country/Area` == "Japan"], n=1))%>%
  hc_add_series(name="Germany",data=tail(data$Value[data$Series == "GDP real
rates of growth (percent)" & data$`Region/Country/Area` == "Germany"], n=1))%>%
  hc_add_series(name="USA",data=tail(data$Value[data$Series == "GDP real rates
of growth (percent)" & data$`Region/Country/Area` == "United States of
America"], n=1))%>%
  hc_add_series(name="China",data=tail(data$Value[data$Series == "GDP real
rates of growth (percent)" & data$`Region/Country/Area` == "China"], n=1))%>%
  hc_add_series(name="India",data=tail(data$Value[data$Series == "GDP real
rates of growth (percent)" & data$`Region/Country/Area` == "India"], n=1))

#------------------Data of GDP Per Capita in 2018-----------------------
pc <- mean(data$Value[data$Series == "GDP real rates of growth (percent)" &
data$`Region/Country/Area` == "China"])
print(pc)
highchart() %>%
  hc_chart(type = "column") %>%
  hc_xAxis(
    title = list(text = "5 appropriately contrasting countries"),
    alternateGridColor = "#FDFFD5",
    plotLines = list(
      list(
        label = list(text = "This is a plotLine"),
        color = "#FF0000",
        width = 2,
        value = 5.5
```

```r
      )
    )
  ) %>%
  hc_title(text = "Data of GDP Per Capita in 2018") %>%
  hc_plotOptions(column = list(enableMouseTracking = TRUE)
  ) %>%
  hc_add_series(name="India",data=tail(data$Value[data$Series == "GDP per
capita (US dollars)" & data$`Region/Country/Area` == "India"], n=1))%>%
  hc_add_series(name="China",data=tail(data$Value[data$Series == "GDP per
capita (US dollars)" & data$`Region/Country/Area` == "China"], n=1))%>%
  hc_add_series(name="Japan",data=tail(data$Value[data$Series == "GDP per
capita (US dollars)" & data$`Region/Country/Area` == "Japan"], n=1))%>%
  hc_add_series(name="Germany",data=tail(data$Value[data$Series == "GDP per
capita (US dollars)" & data$`Region/Country/Area` == "Germany"], n=1))%>%
  hc_add_series(name="USA",data=tail(data$Value[data$Series == "GDP per capita
(US dollars)" & data$`Region/Country/Area` == "United States of America"],
n=1))

#sub-setting data for line plots
Region <- data$`Region/Country/Area`[data$`Region/Country/Area` == "India" |
data$`Region/Country/Area` == "China" | data$`Region/Country/Area` == "Germany"
| data$`Region/Country/Area` == "United States of America" |
data$`Region/Country/Area` == "Japan"]
Year <-data$Year[data$`Region/Country/Area` == "India" |
data$`Region/Country/Area` == "China" | data$`Region/Country/Area` == "Germany"
| data$`Region/Country/Area` == "United States of America" |
data$`Region/Country/Area` == "Japan"]
Series <-data$Series[data$`Region/Country/Area` == "India" |
data$`Region/Country/Area` == "China" | data$`Region/Country/Area` == "Germany"
| data$`Region/Country/Area` == "United States of America" |
data$`Region/Country/Area` == "Japan"]
Value <-data$Value[data$`Region/Country/Area` == "India" |
data$`Region/Country/Area` == "China" | data$`Region/Country/Area` == "Germany"
| data$`Region/Country/Area` == "United States of America" |
data$`Region/Country/Area` == "Japan"]
df <- data.frame(Region, Year, Series, Value)

#---------------------Data of GDP Growth Rate--------------------------
growth <- subset(df, data$Series == "GDP real rates of growth (percent)")
ggplot(data=growth, aes(x=Year, y=Value, group=Region)) +
  geom_line(aes( color=Region), size = 1)+
  geom_point(aes(color = Region, size = 0.1))+
  theme(legend.position = "top")+
  labs(y = "Growth Rate", title = "GDP Growth Rate", subtitle = "from 1985 to
2018")

#-----------------------GDP Per Capita---------------------------------
gdp_capita <- subset(df, data$Series == "GDP per capita (US dollars)")
ggplot(data=gdp_capita, aes(x=Year, y=Value, group=Region)) +
  geom_line(aes( color=Region), size = 1)+
  geom_point(aes(color = Region, size = 0.1))+
```

```
    theme(legend.position = "top")+
  labs(y = "GDP Per Capita", title = "GDP Per Capita (in US dollars)", subtitle
= "from 1985 to 2018")

#------------------GDP in constant 2010 Prices---------------------------
gdp_2010 <- subset(df, data$Series == "GDP in constant 2010 prices (millions of
US dollars)")
ggplot(data=gdp_2010, aes(x=Year, y=Value, group=Region)) +
  geom_line(aes( color=Region), size = 1)+
  geom_point(aes(color = Region, size = 0.1))+
  theme(legend.position = "top")+
  labs(y = "GDP in constant 2010 prices (millions of US dollars)", title = "GDP
in constant 2010 prices (millions of US dollars)", subtitle = "from 1985 to
2018")
```

Outputs:
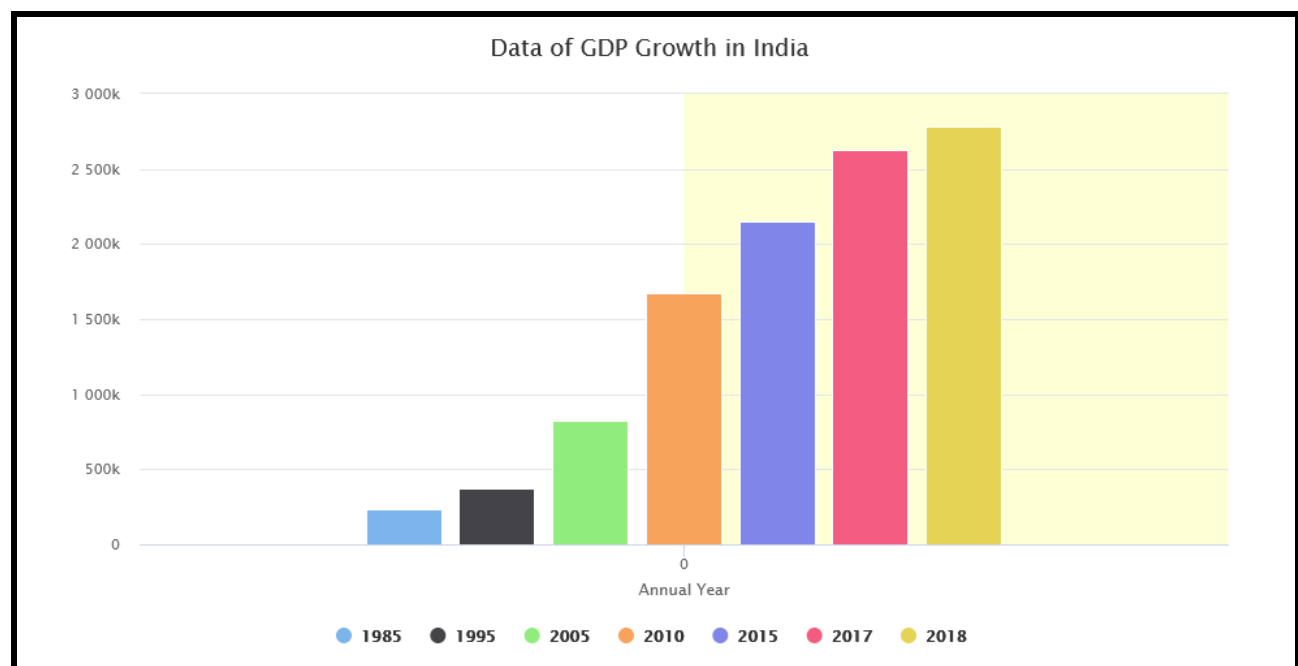
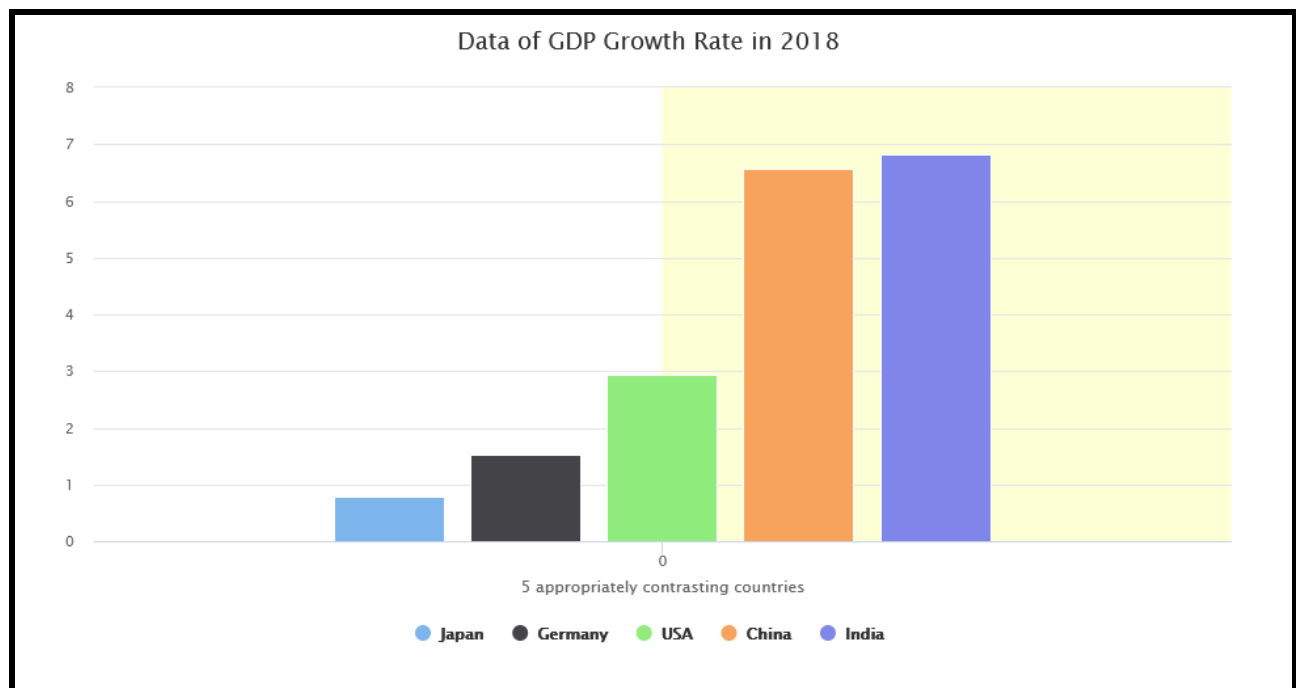**Fig 1: Data of GDP Growth in India**
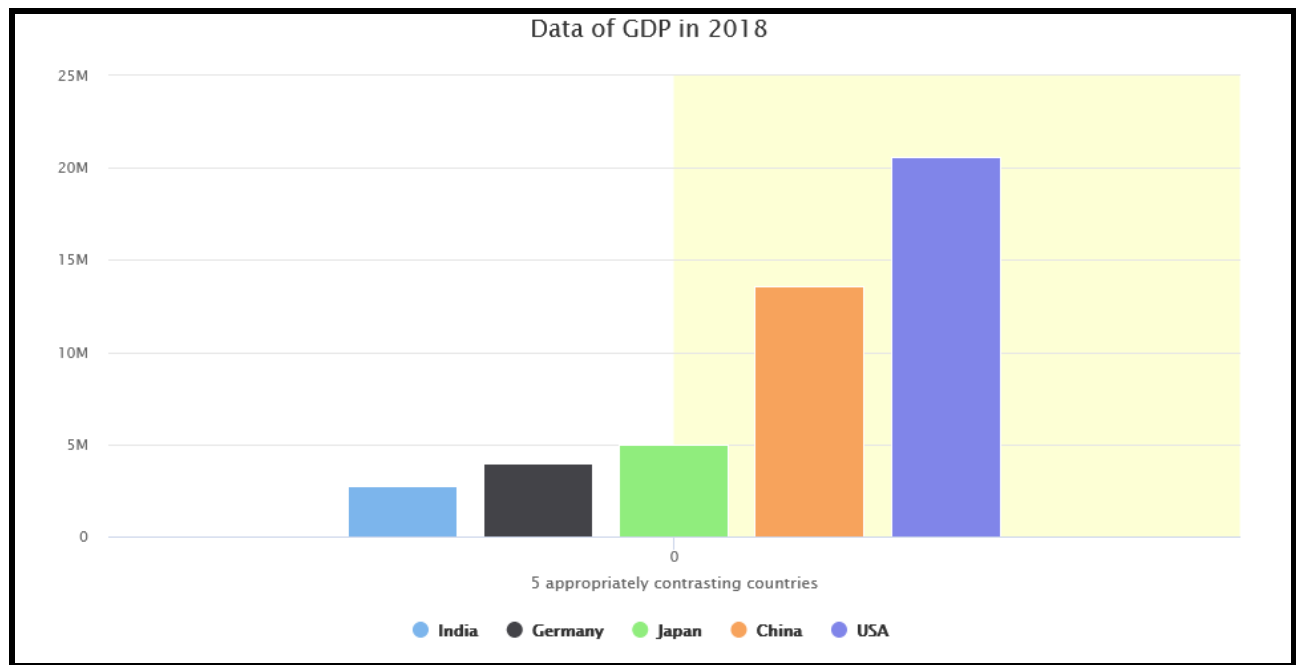
Fig 2: Data of GDP in 2018

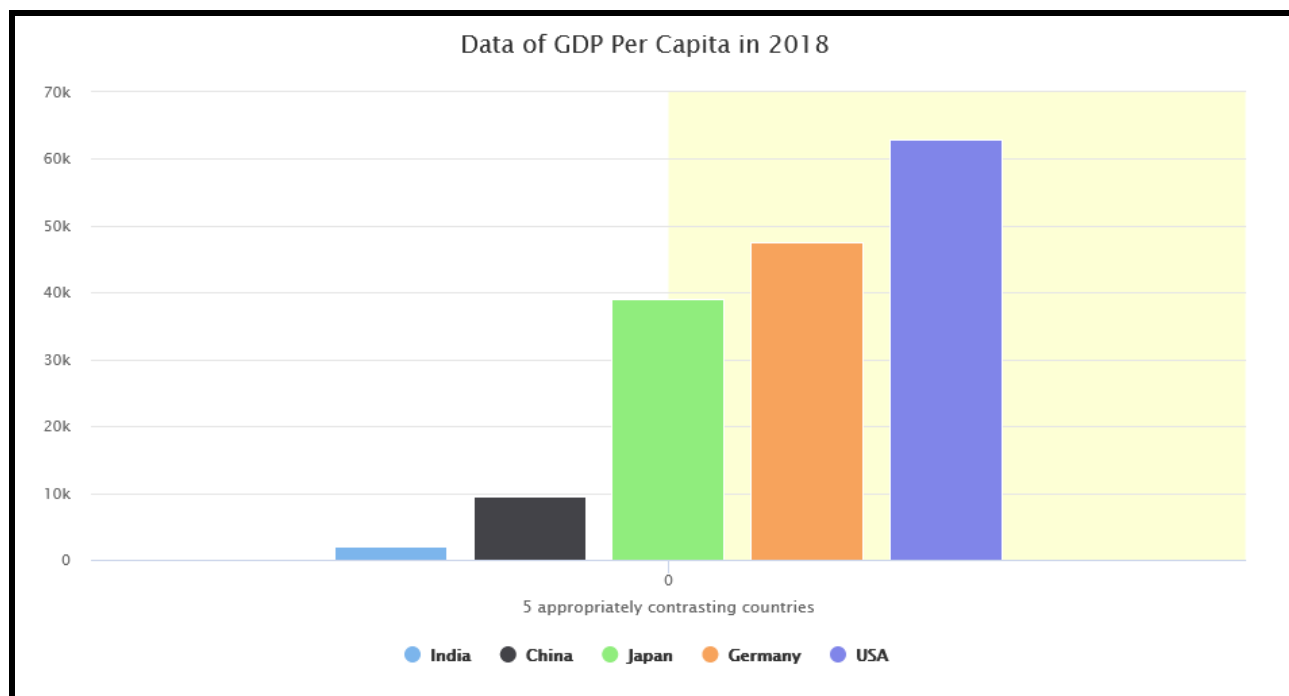**Fig 3: Data of GDP Growth Rate in 2018**

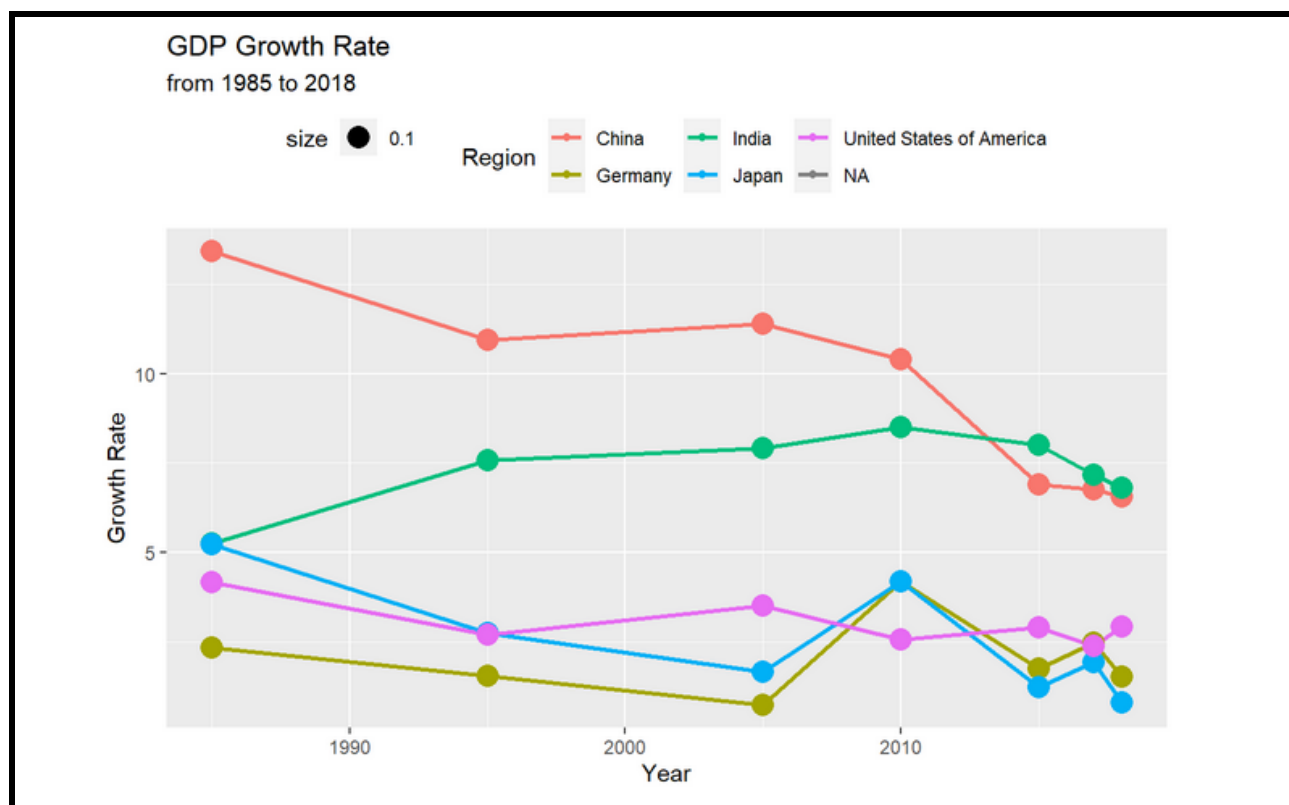**Fig 4: Data of GDP Per Capita in 2018**
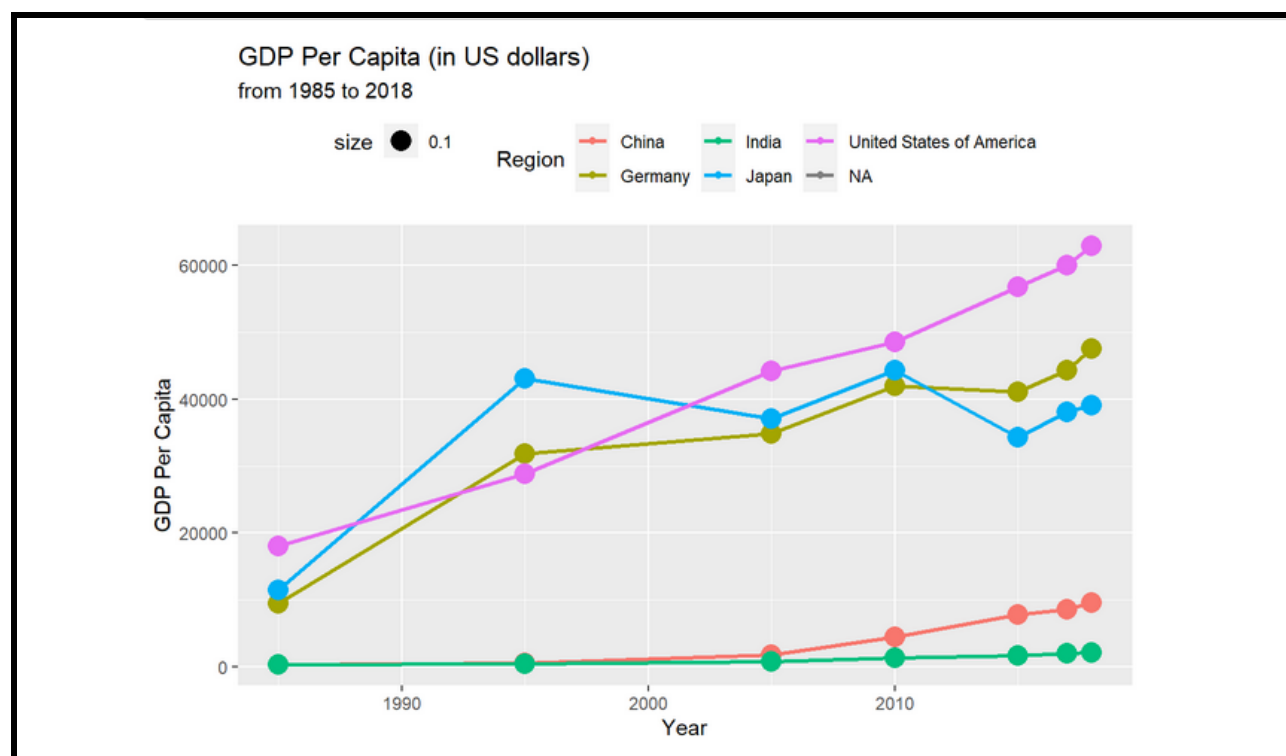

Fig 5: GDP Growth Rate

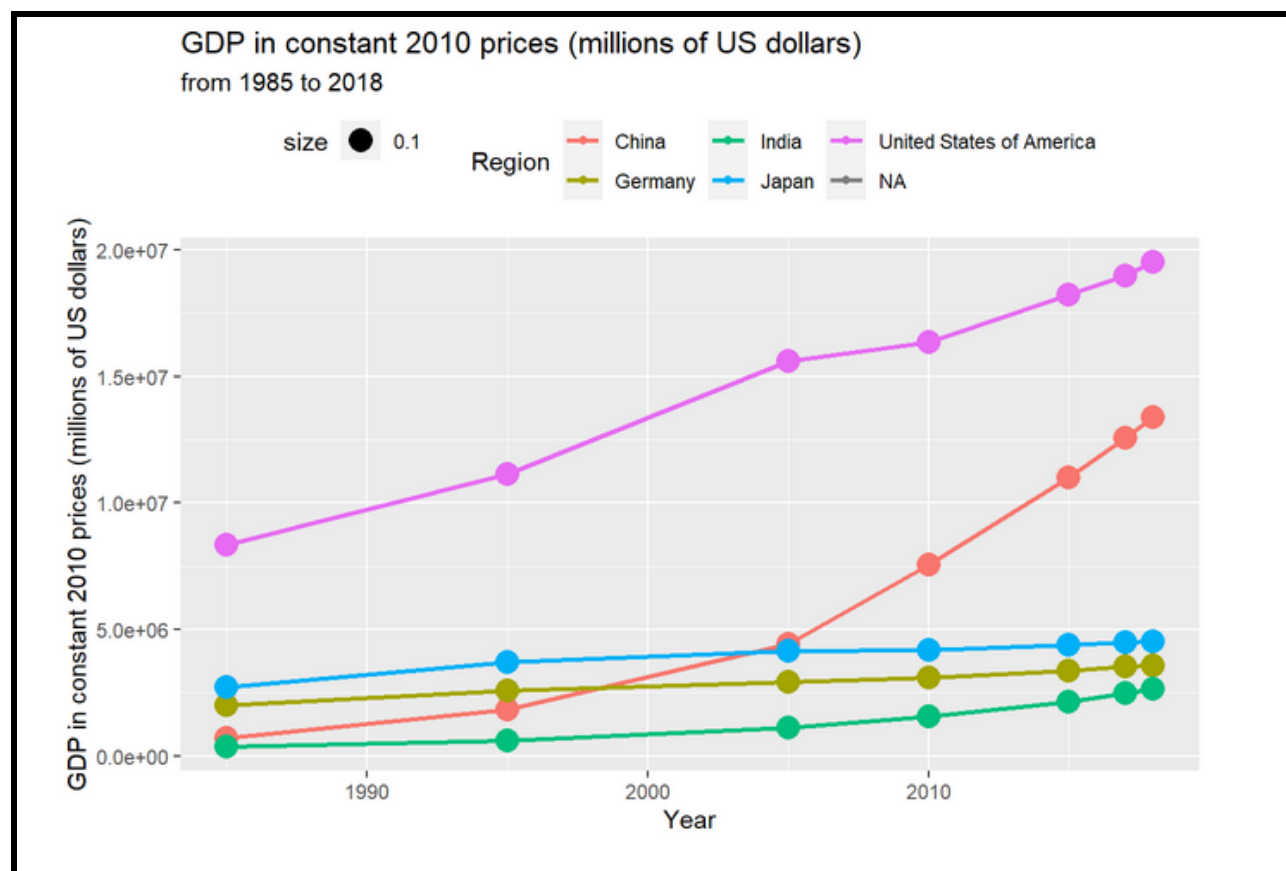**Fig 6: GDP Per Capita (in US Dollars)**

Fig 7: GDP in constant 2010 prices (millions of US dollars)

Conclusion:

- Top-5 countries according to Gross Domestic Product
- China and India are growing at very fast rate
- Although USA, Japan, and Germany are currently growing at slow rate most of their GDP growth lies in 20th Century
- China has made very remarkable growth over small period of time
- There is very greater need to increase GDP growth rate for India
- India is quite far from becoming a developed country.