

Indian economy in pixels

Abhishek Sharma

02/11/2022

Roll No: MDS202204

Visualization

Email: abhisheks@cmi.ac.in

Chennai Mathematical Institute

Abstract

In order to demonstrate the expansion of a certain sector of the Indian economy, this research combines data collected on more than 1000 firms over a two-decade period, including crucial firm-level variables like sales and total assets.

Data Description

This dataset includes financial information about firms from 2001 through 2022. We won't look at every type of company. We will instead focus on big businesses that aren't substantially influenced by the government. To this end, this dataset excludes:

1. businesses that are government-owned
2. firms that have either sales or total assets less than 10 crore
3. mining and construction firms
4. financial firms(Are excluded due to capital controls imposed on these firms by the government)
5. firms that are not listed on either the NSE or the BSE

In order to describe firm specific characteristics, we will use the following seven attributes from the dataset.

1. company name
2. incorporation year: The age and birth cohort of a firms are proxied by the year of incorporation
3. financial year:The year for which the company's financial information is recorded
4. sales: The revenue of the firm
5. total assets: Is a proxy for the size of the firm
6. retained profits/losses for the year
7. industry sector

Introduction

One of the economies in the world with the fastest growth rates is India. For the fiscal year 2022, India's GDP grew at an 8.7% annual pace. So it is reasonable to assume that businesses have been making more money every year. It is also reasonable to assume that this decade saw a greater number of businesses open than the previous one. Sales and total assets ought to have continued to rise annually. But are these assumptions accurate?

In this project, we will try to shed light on these assumptions using empirical insights.

Pictorial Display

```
general_wd='/home/abhishek/Documents/visualization/assignment2'  
library(testthat)  
library(assertive, warn.conflicts = FALSE)  
library(tidyr)
```

```
##  
## Attaching package: 'tidyr'
```

```
## The following object is masked from 'package:testthat':  
##  
## matches
```

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
##  
## The following object is masked from 'package:testthat':  
##  
## matches
```

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

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

```
library(ggplot2)  
library(grid)  
library(gridExtra)
```

```
##  
## Attaching package: 'gridExtra'
```

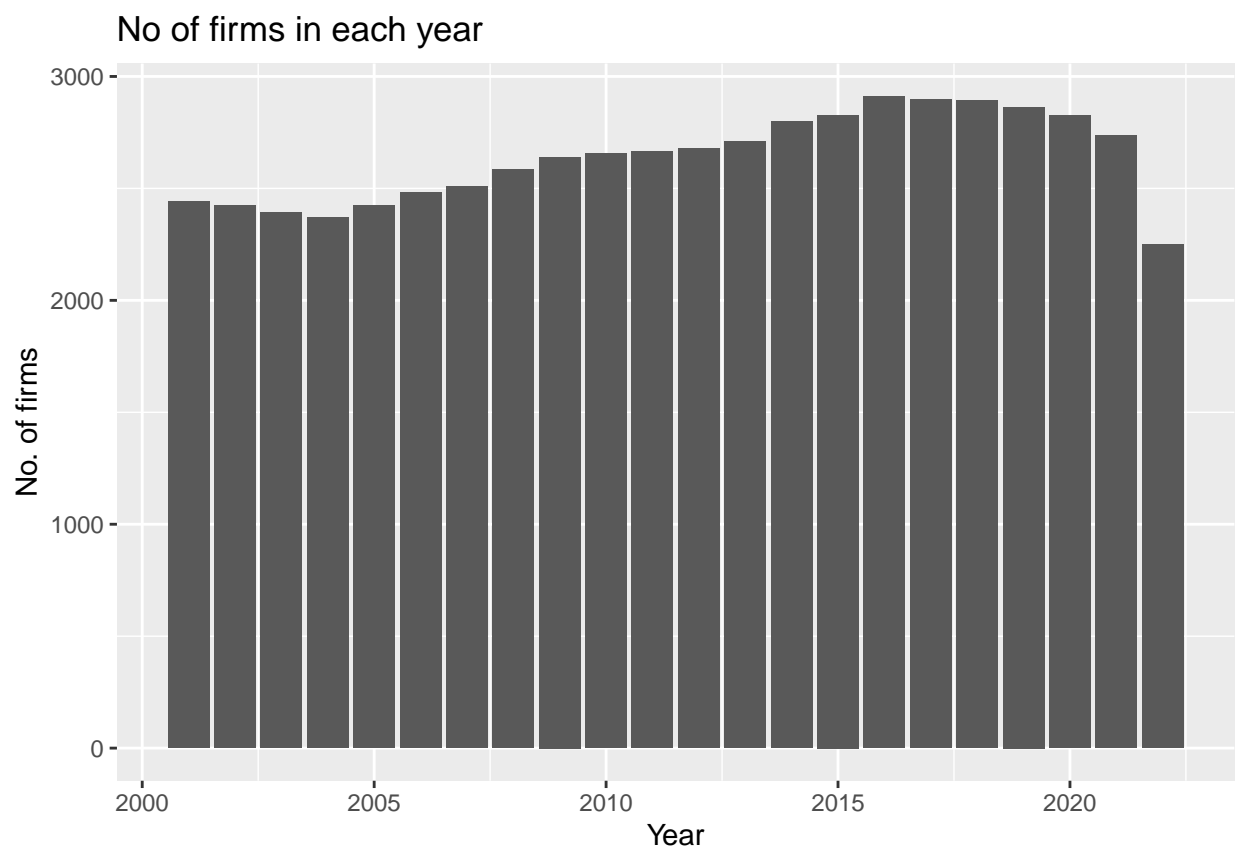
```
## The following object is masked from 'package:dplyr':  
##  
## combine
```

```
setwd(general_wd)  
ds = read.csv(file='dataset.csv',  
              header = TRUE )
```

```
##
```

According to our dataset, the number of firms in each year has remained almost constant.

```
df = ds %>% group_by(year) %>%  
  summarise(  
    no_of_firms = n_distinct(co_code)  
  )  
  
ggplot(data=df, aes(x=year, y=no_of_firms)) +  
  geom_bar(stat="identity")+  
  ggtitle( 'No of firms in each year')+  
  ylab('No. of firms')+  
  xlab('Year')
```

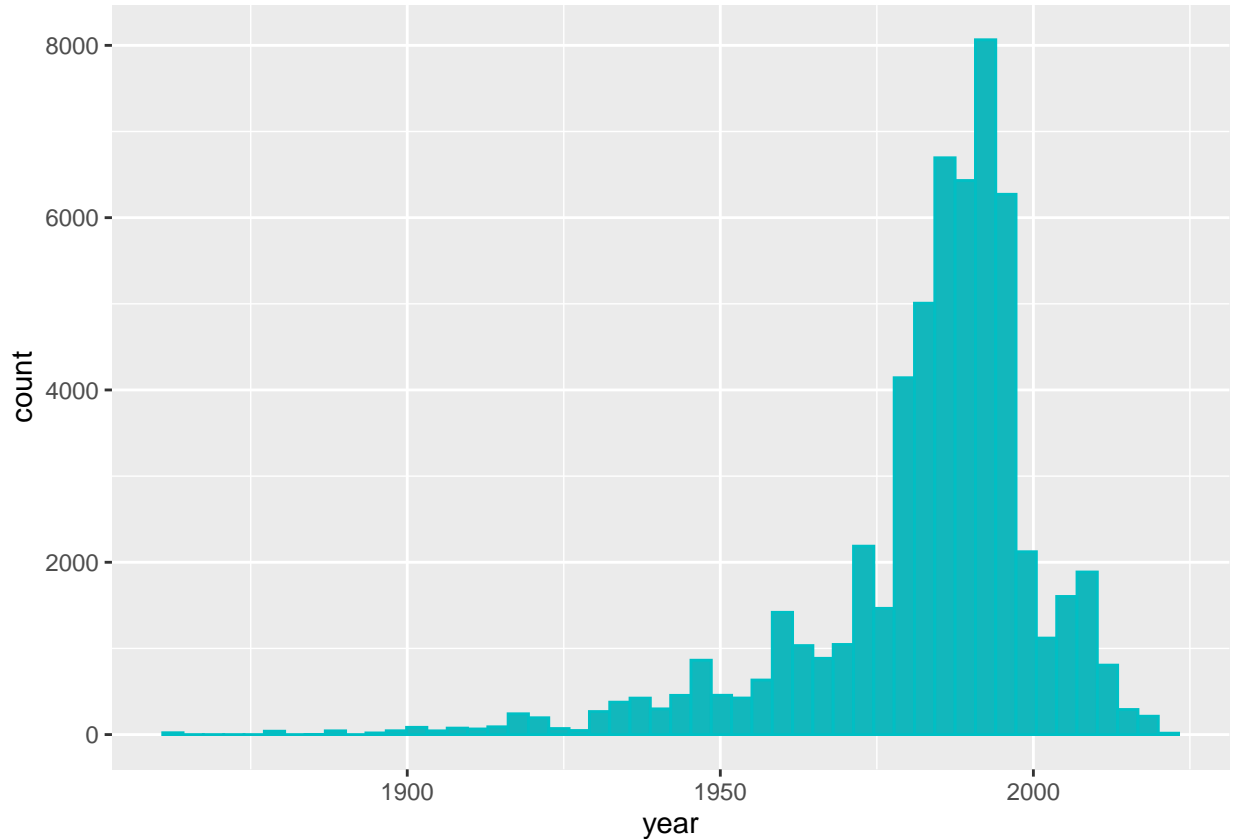


```
## bar chart
```

According to our dataset, most companies were founded between 1975 and 2000. Moreover, more companies were founded in the last decade than in the current decade.

```
x = ds$incorporation_year  
y = data.frame(x)
```

```
colnames(y)=c('year')
hist = ggplot(y, aes(x=year))+geom_histogram(bins=50, color="#00bfc4", fill = '#12b7bc')
## Most companies were started a couple year after independence.
print(hist)
```



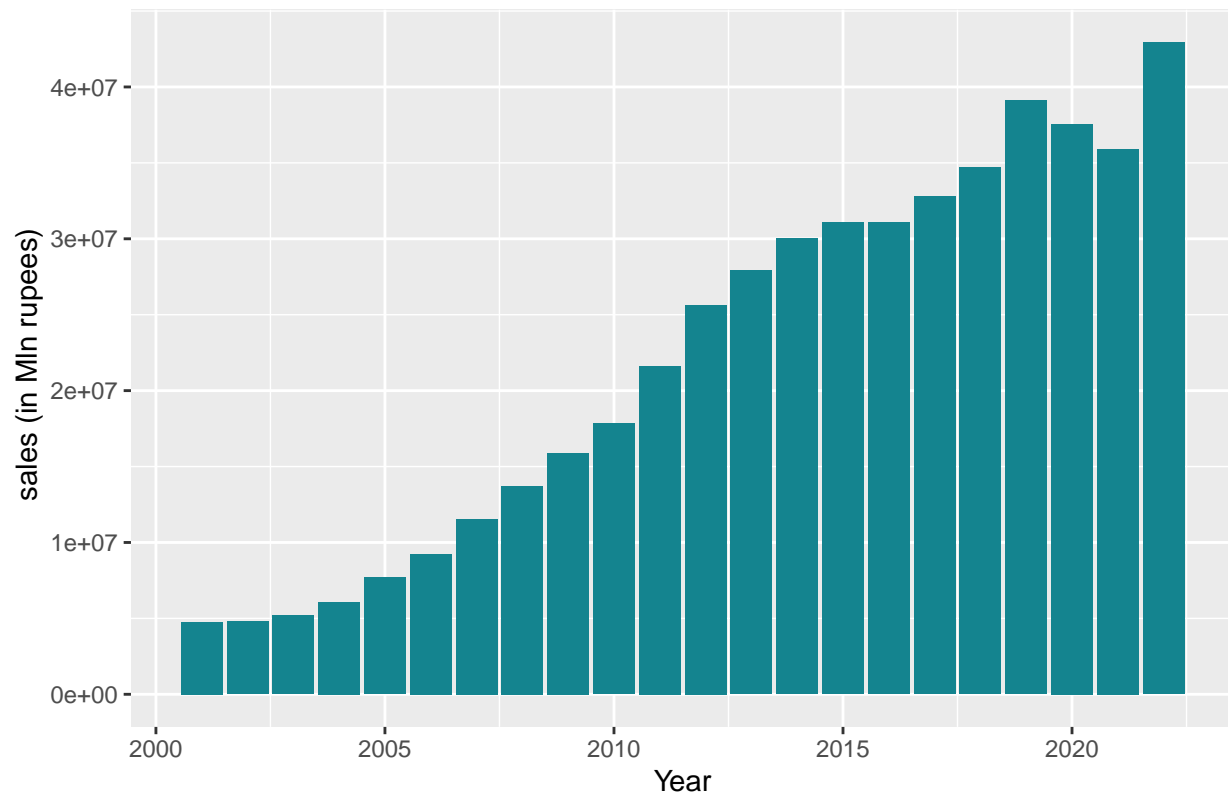
```
## most companies were incorporated between 1975 and 2000
## histogram
```

The sum total of sales of all firms in a given year has been increasing year over year. But the rate of increase has slowed down in recent years.

```
df = ds %>% group_by(year) %>%
  summarise( sum_sales = sum(sales),
             sum_total_assets = sum(total_assets),
             )

ggplot(data=df, aes(x=year, y=sum_sales)) +
  geom_bar(stat="identity", fill = '#14848f')+
  ggtitle( 'Sum Total of all sales per year')+
  ylab(' sales (in Mln rupees)')+
  xlab('Year')
```

Sum Total of all sales per year



```
## bar chart
```

Mean sales of firms per year has increased constantly over the years. But this is not the case for median sales of firms per year.

```
df = ds %>% group_by(year) %>%
  summarise( mean_sales = mean(sales),
             median_sales = median(sales),
             sum_sales = sum(sales),
             mean_total_assets = mean(total_assets),
             median_total_assets = median(total_assets)
            )

reg_mean = ggplot( df, aes(y=mean_sales, x=year) )+geom_line()+
  ggtitle( 'Mean Sales') +
  ylab('mean sales') +
  theme(panel.border = element_rect(colour = "black", fill=NA, size=0.5),
        plot.title = element_text(face="bold", hjust=0.5, size=15),
        axis.text.y=element_text(angle=90, hjust=0.5),
        )

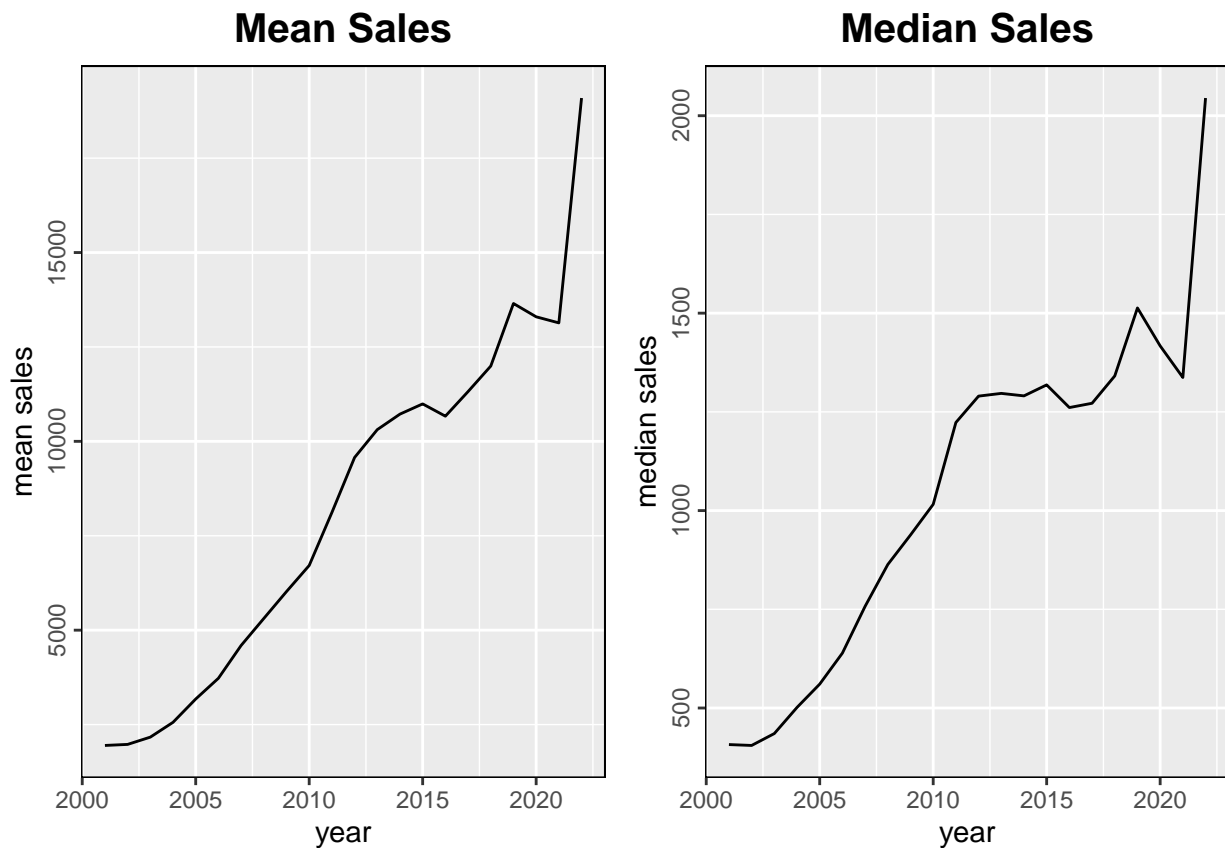
reg_median = ggplot( df, aes(y=median_sales, x=year) )+geom_line()+
  ggtitle( 'Median Sales') +
```

```

ylab('median sales')+
theme(panel.border = element_rect(colour = "black", fill=NA, size=0.5),
      plot.title = element_text(face="bold", hjust=0.5, size=15),
      axis.text.y=element_text(angle=90, hjust=0.5),
      )

grid.arrange(reg_mean, reg_median, ncol = 2)

```



```
## line plot
```

Mean total assets of firms per year has increased constantly over the years. But this is not the case for median total assests of firms per year.

```

df = ds %>% group_by(year) %>%
  summarise( mean_total_assets = mean(total_assets),
             median_total_assets = median(total_assets)
            )

reg_mean = ggplot( df, aes(y=mean_total_assets, x=year) )+geom_line()+
  ggtitle( 'Mean Total Assets')+
  ylab('mean total assets')+
  theme(panel.border = element_rect(colour = "black", fill=NA, size=0.5),

```

```

    plot.title = element_text(face="bold", hjust=0.5, size=15),
    axis.text.y=element_text(angle=90, hjust=0.5),

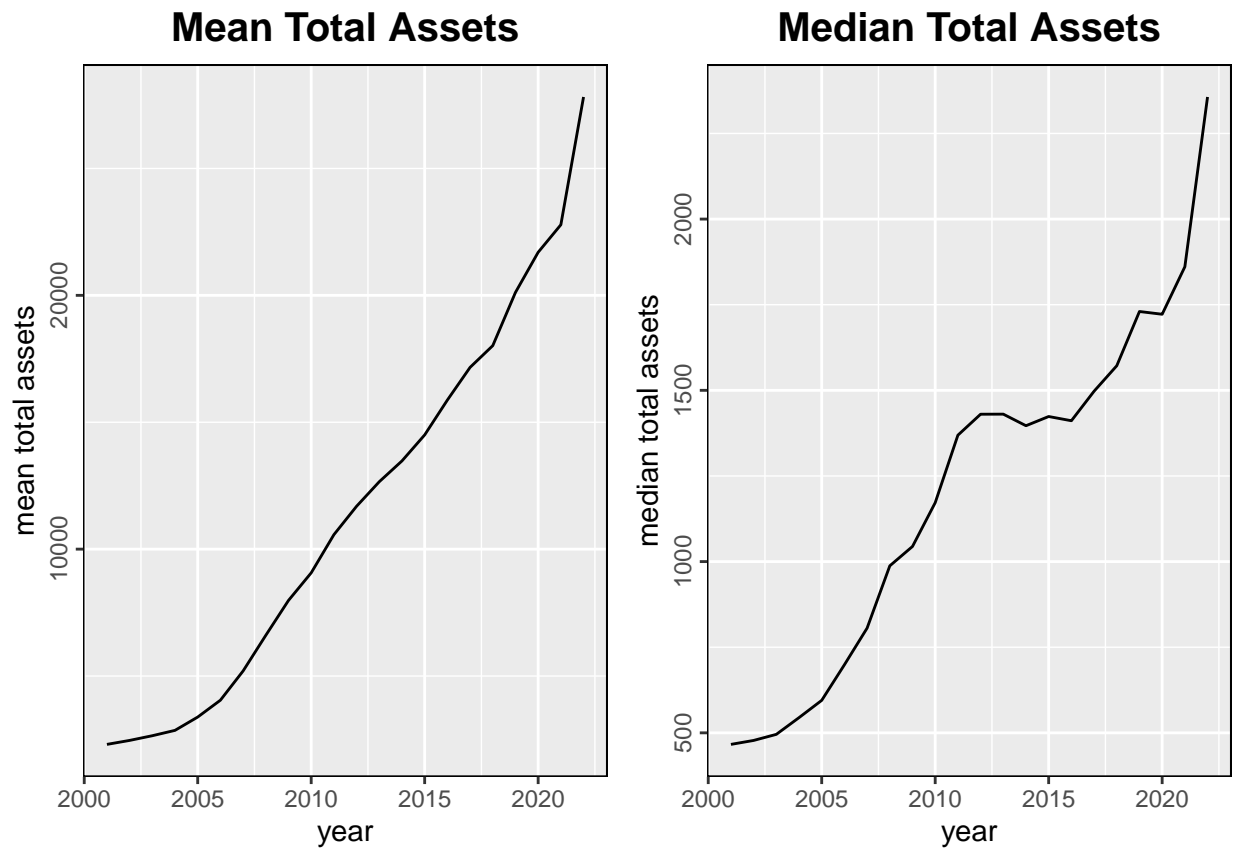
)

reg_median = ggplot( df, aes(y=median_total_assets, x=year) )+geom_line()+
  ggtitle( 'Median Total Assets')+
  ylab('median total assets')+
  theme(panel.border = element_rect(colour = "black", fill=NA, size=0.5),
    plot.title = element_text(face="bold", hjust=0.5, size=15),
    axis.text.y=element_text(angle=90, hjust=0.5),

  )

grid.arrange(reg_mean, reg_median, ncol = 2)

```



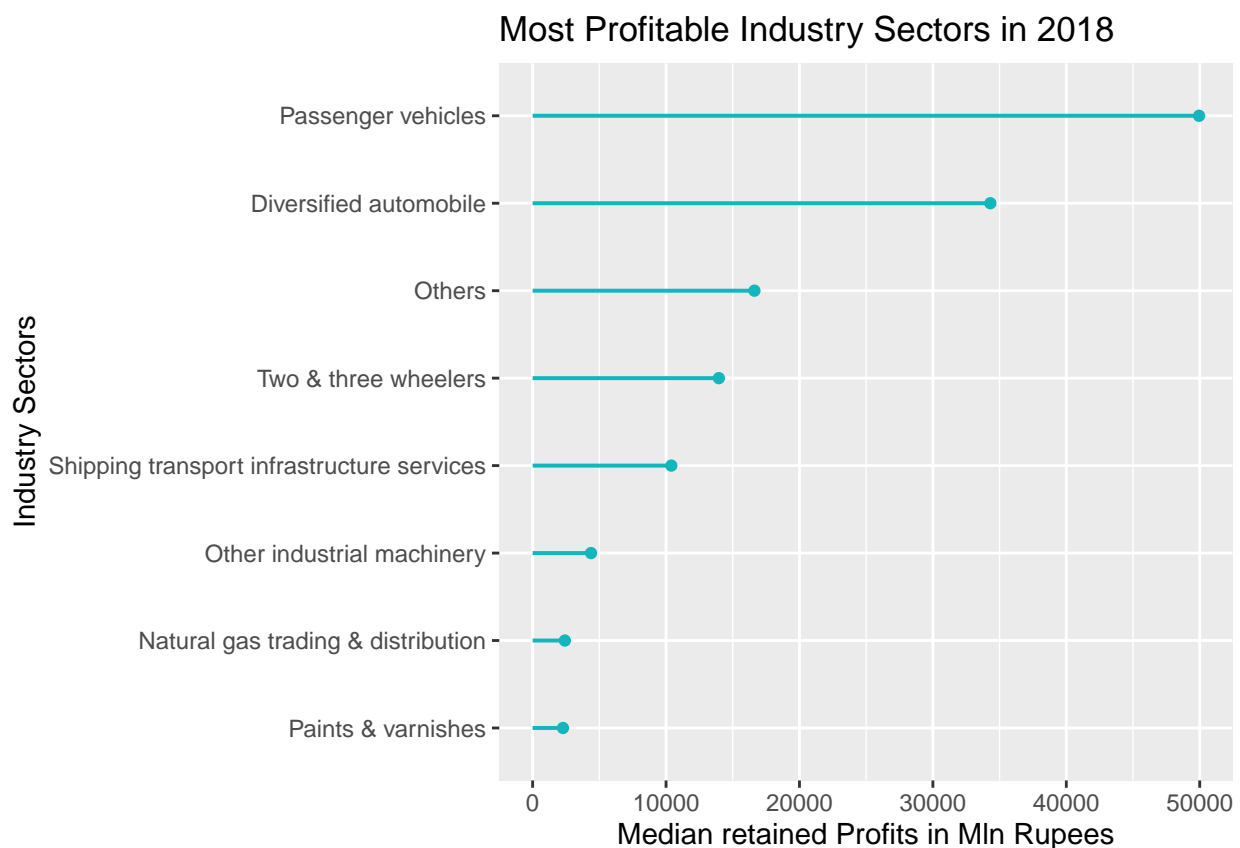
```
## line plot
```

The most profitable industry sectors are:


```

df = ds[(ds$year==2018),]
df = df %>% group_by(industry_type) %>%
  summarise(
    median_retained_profits = median(retained_profits)
  )
df = df[order(df$median_retained_profits, decreasing = TRUE),]
others = sum(df[-(1:5),]$median_retained_profits)
#View( df[(df$median_retained_profits>10000),] )
df = df[(1:7),]
df[8,] = list('Others', others)
##View(df)
name = df$industry_type
value = df$median_retained_profits
ggplot( df, aes(y=reorder(name,value), x=value) )+
  geom_point( size=1.5 , color = '#12b7bc') +
  geom_segment( aes(x=0, xend=value,y=name, yend=name), size = 0.7 , color = '#12b7bc') +
  ggtitle( 'Most Profitable Industry Sectors in 2018') +
  ylab('Industry Sectors') +
  xlab('Median retained Profits in Mln Rupees')

```

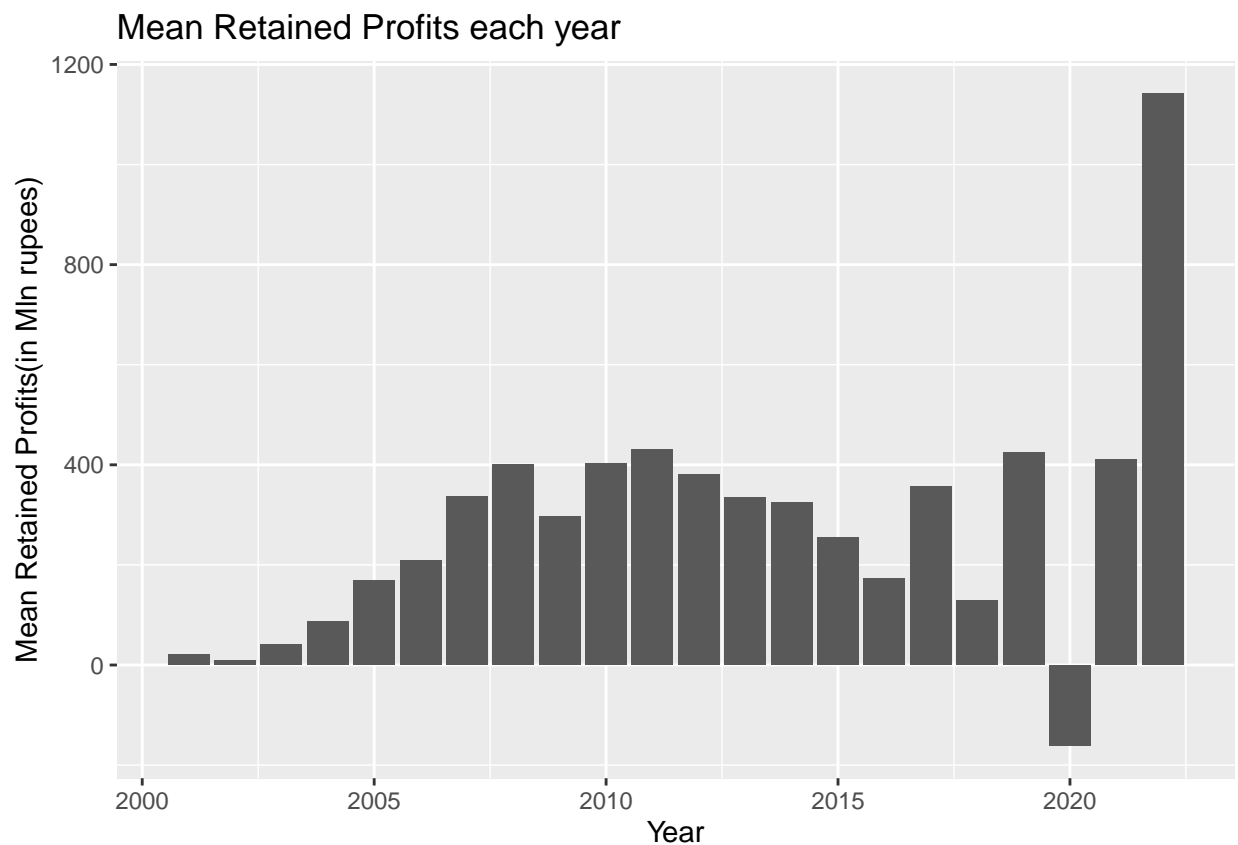


```
## Lolipop plot
```

Mean retained profits of firms each year saw a huge increase after the COVID pandemic.

```
df = ds %>% group_by(year) %>%
  summarise(
    mean_retained_profits = mean(retained_profits),
    median_retained_profits = median(retained_profits)
  )

ggplot(data=df, aes(x=year, y=mean_retained_profits)) +
  geom_bar(stat="identity")+
  ggtitle('Mean Retained Profits each year')+
  ylab('Mean Retained Profits(in Mln rupees)')+
  xlab('Year')
```



```
## bar chart
```

Analysis

The Above pictorial representation shows that -

assumptions can be wrong.

Conclusion

assumptions can be wrong.