Here we will be looking at the trends of FDI in India 2000 to 2017.

FDI support on different sectors over the years indicates changes in interest over time.

1. Firing up colab

```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

```
df = pd.read_csv('/content/FDI_in_India.csv')
df.head(3)
```

	Sector	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09
0	METALLURGICAL INDUSTRIES	22.69	14.14	36.61	8.11	200.38	149.13	169.94	1175.75	959.94
1	MINING	1.32	6.52	10.06	23.48	9.92	7.40	6.62	444.36	34.16

```
len(df)
```

```
63
```

```
df.info()
```

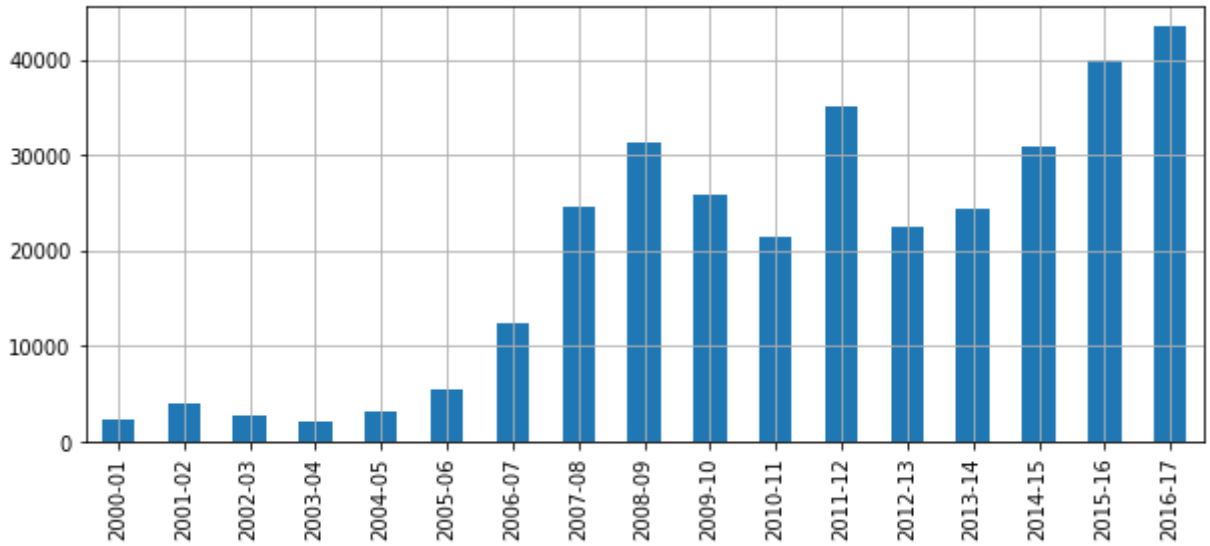
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 63 entries, 0 to 62
Data columns (total 18 columns):
 #   Column   Non-Null Count  Dtype  
---  ------   --------------  -----  
 0   Sector   63 non-null     object 
 1   2000-01  63 non-null     float64
 2   2001-02  63 non-null     float64
 3   2002-03  63 non-null     float64
 4   2003-04  63 non-null     float64
 5   2004-05  63 non-null     float64
 6   2005-06  63 non-null     float64
 7   2006-07  63 non-null     float64
 8   2007-08  63 non-null     float64
 9   2008-09  63 non-null     float64
 10  2009-10  63 non-null     float64
 11  2010-11  63 non-null     float64
 12  2011-12  63 non-null     float64
 13  2012-13  63 non-null     float64
 14  2013-14  63 non-null     float64
 15  2014-15  63 non-null     float64
 16  2015-16  63 non-null     float64
 17  2016-17  63 non-null     float64
dtypes: float64(17), object(1)
memory usage: 9.0+ KB
```

2. Setting the index and subsequent

```
df.set_index('Sector', inplace = True)
```

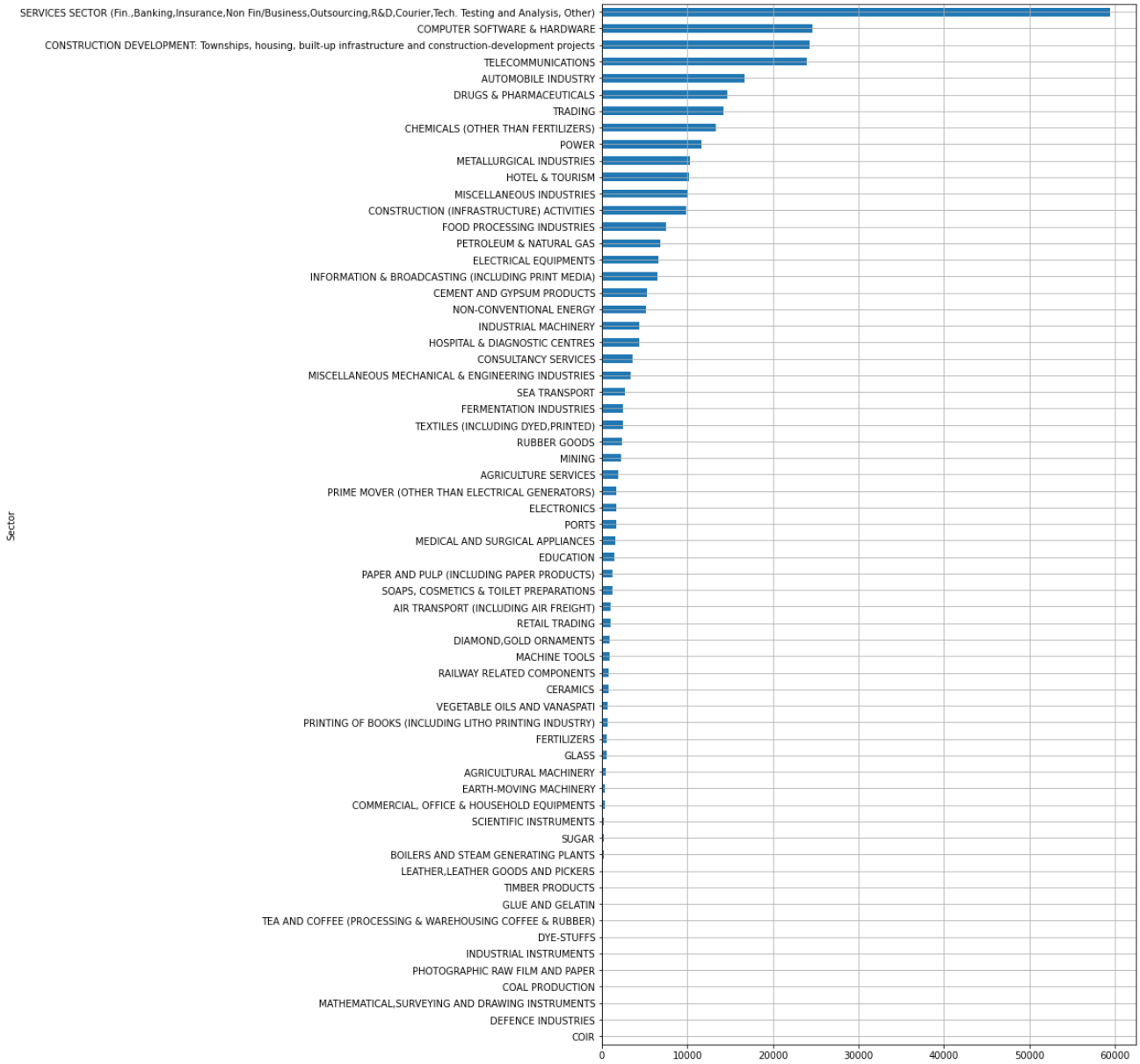
Year-wise total Investment

```
df.sum(axis=0).plot(kind='bar', figsize=(10,4))
plt.grid()
plt.show()
```



sector-wise total investment over the years

```
df.sum(axis=1).sort_values().plot(kind = 'barh', figsize=(10,20))
plt.grid()
plt.show()
```



We find that from 2007 onwards there has been an overall large increase in Investment levels.

Also, Services (Finance, Banking etc.) sector has had the max overall investment while Coir has had the least investment overall across the years.

finding trends

a. correlation

```
import numpy as np
```

```
new = df.transpose()
corrMatrix=new.corr()

corrMatrix.loc[:,:] =  np.tril(corrMatrix, k=-1) # borrowed from Karl D's answer

already_in = set()
result = []
for col in corrMatrix:
    perfect_corr = corrMatrix[col][corrMatrix[col] >= 0.9].index.tolist()
    if perfect_corr and col not in already_in:
        already_in.update(set(perfect_corr))
        perfect_corr.append(col)
        result.append(perfect_corr)

result
```

```
[['TELECOMMUNICATIONS',
  'TEXTILES (INCLUDING DYED,PRINTED)',
  'GLUE AND GELATIN',
  'ELECTRICAL EQUIPMENTS'],
 ['TRADING', 'AUTOMOBILE INDUSTRY'],
 ['SUGAR',
  'CONSTRUCTION (INFRASTRUCTURE) ACTIVITIES',
  'AIR TRANSPORT (INCLUDING AIR FREIGHT)'],
 ['RETAIL TRADING', 'SEA TRANSPORT'],
 ['SOAPS, COSMETICS & TOILET PREPARATIONS', 'INDUSTRIAL MACHINERY'],
 ['DEFENCE INDUSTRIES', 'MISCELLANEOUS MECHANICAL & ENGINEERING INDUSTRIES'],
 ['TEXTILES (INCLUDING DYED,PRINTED)', 'MEDICAL AND SURGICAL APPLIANCES'],
 ['GLASS', 'MATHEMATICAL,SURVEYING AND DRAWING INSTRUMENTS'],
 ['DIAMOND,GOLD ORNAMENTS', 'DYE-STUFFS'],
 ['FOOD PROCESSING INDUSTRIES', 'FERMENTATION INDUSTRIES']]
```

Above we set the correlation coefficient score to 0.9.

Now that we plot them to observe price trends considering each group as a separate segment wherein sectors have high correlation with each other.

```
X = list(df.columns)
#X.remove('Sector')
X
```

```
['2000-01',
 '2001-02',
 '2002-03',
 '2003-04',
 '2004-05',
 '2005-06',
 '2006-07',
 '2007-08',
 '2008-09',
 '2009-10',
 '2010-11',
 '2011-12',
 '2012-13',
 '2013-14',
 '2014-15',
```

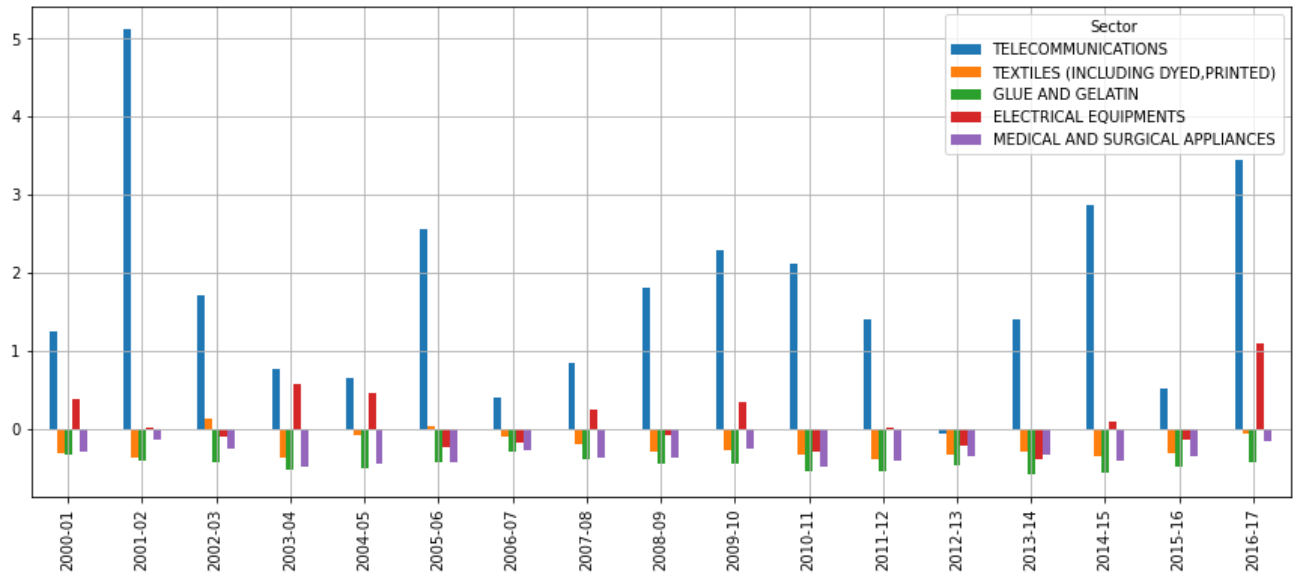
```
        '2015-16',
        '2016-17']
```

plotting graphs for each groups which are highly correlated

```
norm_df=(df-df.mean())/df.std()
norm_df.head(2)
```

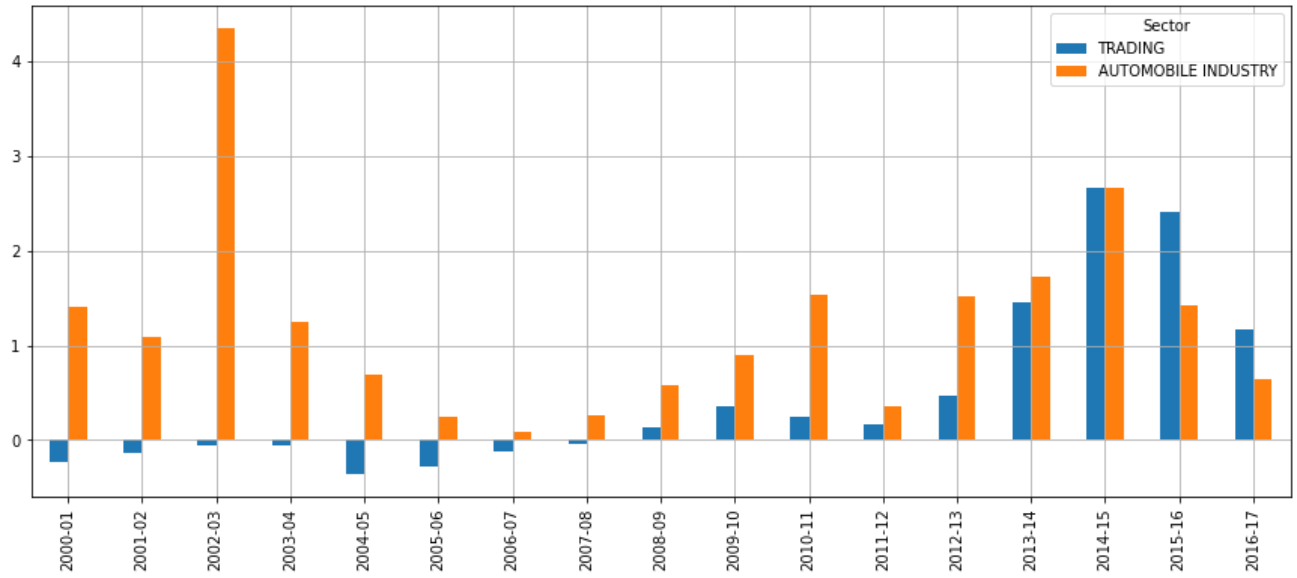
Sector	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006
METALLURGICAL INDUSTRIES	-0.134256	-0.315379	-0.072924	-0.393441	1.464559	0.296446	-0.041
MINING	-0.324673	-0.363644	-0.379483	-0.166255	-0.403888	-0.390107	-0.279

```
df_1 = norm_df.loc[['TELECOMMUNICATIONS',
  'TEXTILES (INCLUDING DYED,PRINTED)',
  'GLUE AND GELATIN',
  'ELECTRICAL EQUIPMENTS','MEDICAL AND SURGICAL APPLIANCES'], X]
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
#df_1
```



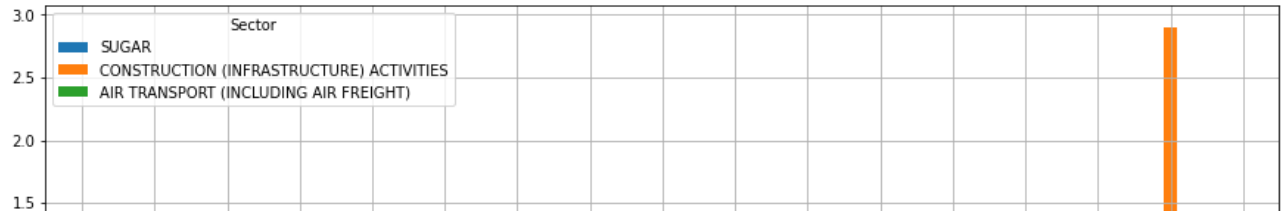
In the figure above, we see that while Telecom sector has had over average investment overall with Electrical equipments following. For others (glue, medical and textile) the investment has been below average throughout. In 2012-13 though investment in this category has been below average.

```
df_1 = norm_df.loc[['TRADING', 'AUTOMOBILE INDUSTRY'], X]
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
#df_1
```



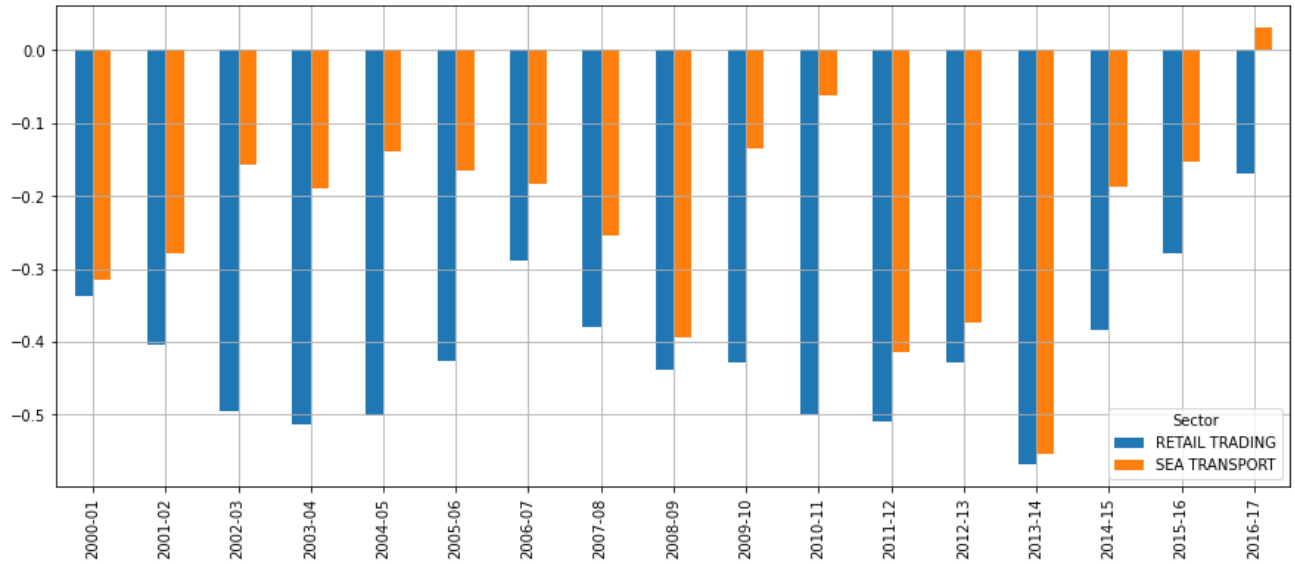
While Automobile has been above average but trading had a cyclic pattern with investment going to positive 2008 onwards.

```
df_1 = norm_df.loc[['SUGAR',
  'CONSTRUCTION (INFRASTRUCTURE) ACTIVITIES',
  'AIR TRANSPORT (INCLUDING AIR FREIGHT)'], X]
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
```

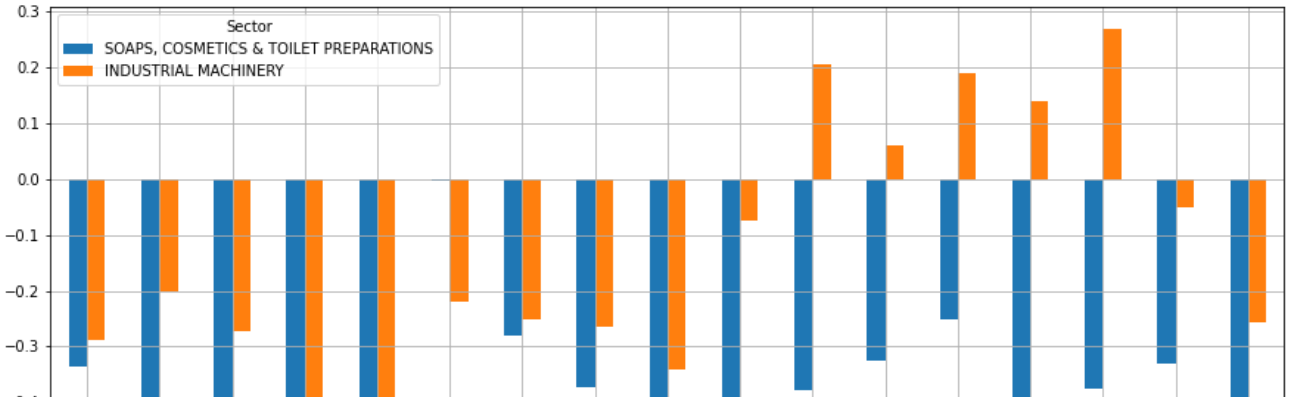


In this group we see that while overall this group has below average investment, construction saw an above average insvestment with highest in 2015.

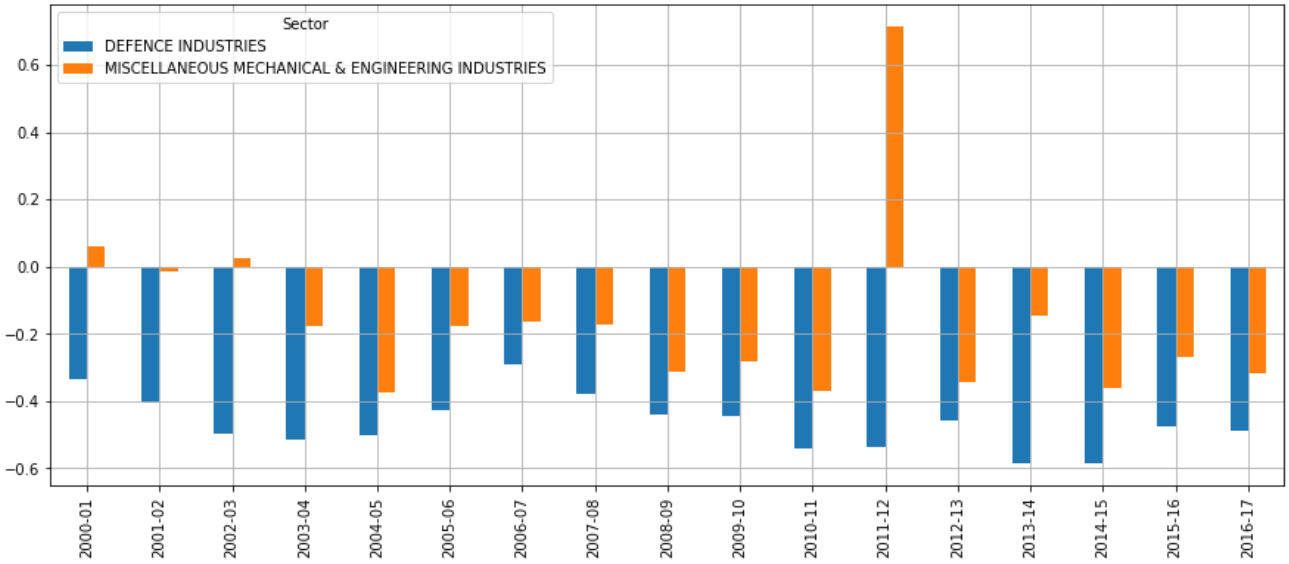
```
df_1 = norm_df.loc[['RETAIL TRADING', 'SEA TRANSPORT'], X]
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
```



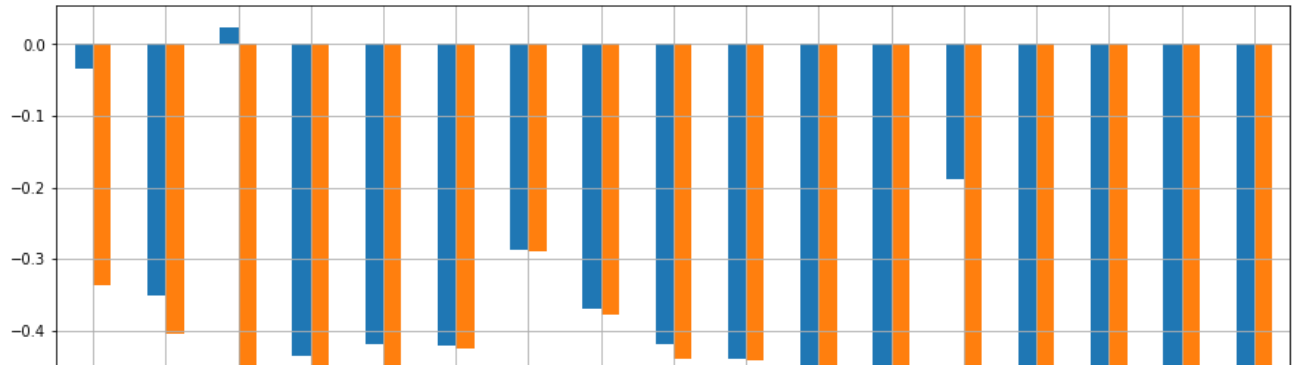

```
df_1 = norm_df.loc[['SOAPS, COSMETICS & TOILET PREPARATIONS', 'INDUSTRIAL MACHINERY'], X]
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
```



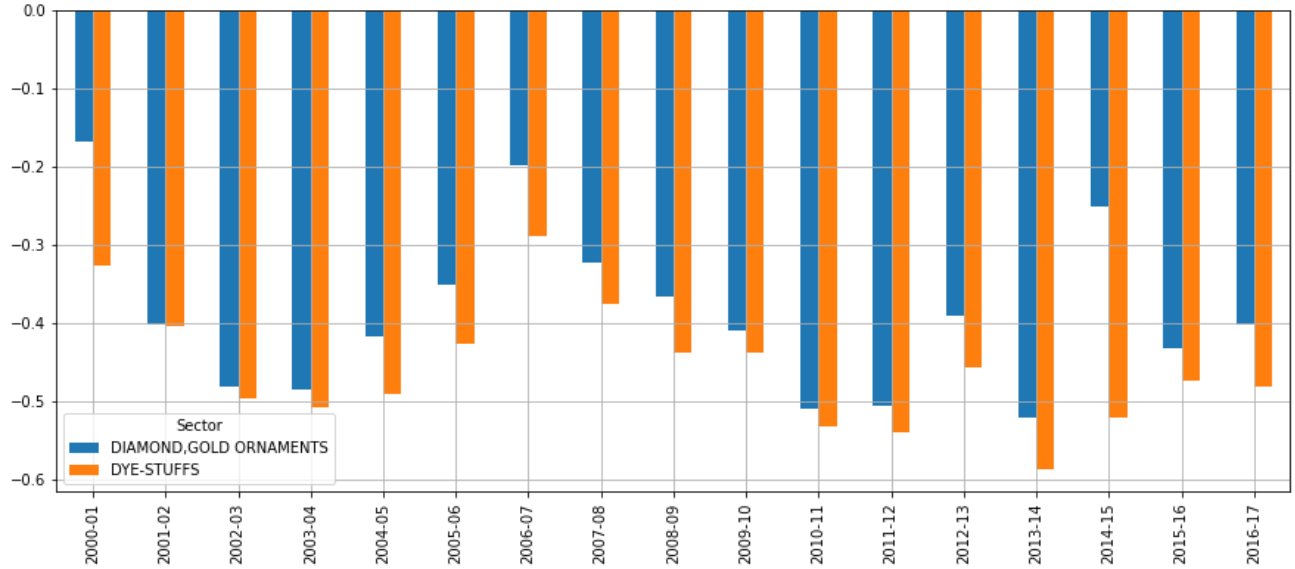

```
df_1 = norm_df.loc[['DEFENCE INDUSTRIES', 'MISCELLANEOUS MECHANICAL & ENGINEERING INDUSTRI
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
```




```
df_1 = norm_df.loc[['GLASS', 'MATHEMATICAL,SURVEYING AND DRAWING INSTRUMENTS'], X]
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
```

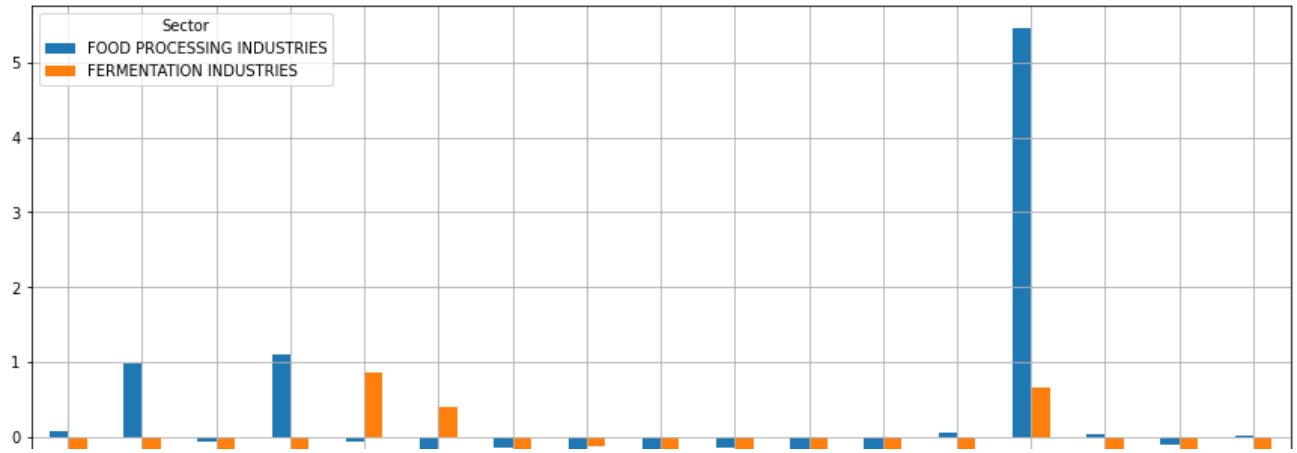



```
df_1 = norm_df.loc[['DIAMOND,GOLD ORNAMENTS', 'DYE-STUFFS'], X]
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
```



For the rest, we observe a cyclic pattern in investment. (Most below average).

```
df_1 = norm_df.loc[['FOOD PROCESSING INDUSTRIES', 'FERMENTATION INDUSTRIES'], X]
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
```



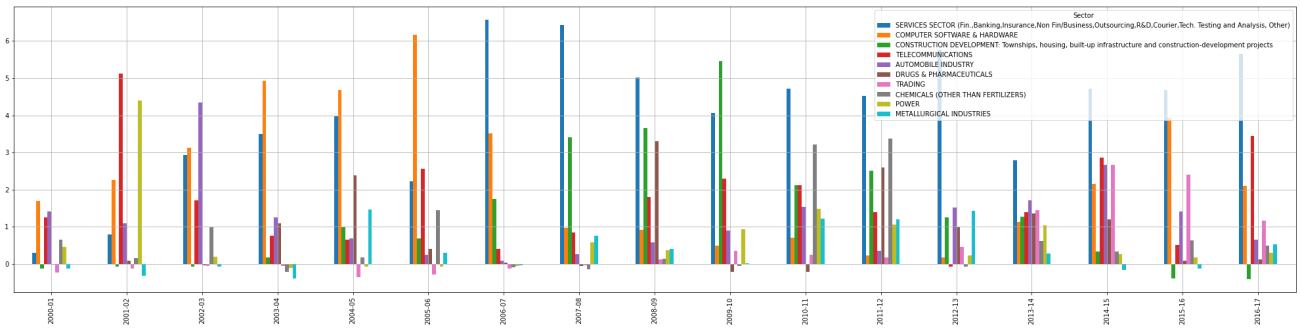
Top and bottom sectors invested in.

Top-most 10

```
df.sum(axis=1).nlargest(10)
```

```
Sector
SERVICES SECTOR (Fin.,Banking,Insurance,Non Fin/Business,Outsourcing,R&D,Courier,Tech
COMPUTER SOFTWARE & HARDWARE
CONSTRUCTION DEVELOPMENT: Townships, housing, built-up infrastructure and constructio
TELECOMMUNICATIONS
AUTOMOBILE INDUSTRY
DRUGS & PHARMACEUTICALS
TRADING
CHEMICALS (OTHER THAN FERTILIZERS)
POWER
METALLURGICAL INDUSTRIES
dtype: float64
```

```
df_1 = norm_df.loc[['SERVICES SECTOR (Fin.,Banking,Insurance,Non Fin/Business,Outsourcing,
'COMPUTER SOFTWARE & HARDWARE',
'CONSTRUCTION DEVELOPMENT: Townships, housing, built-up infrastructure and construction-de
'TELECOMMUNICATIONS',
'AUTOMOBILE INDUSTRY',
'DRUGS & PHARMACEUTICALS',
'TRADING',
'CHEMICALS (OTHER THAN FERTILIZERS)',
'POWER',
'METALLURGICAL INDUSTRIES'   ], X]
df_1.transpose().plot(kind = 'bar', figsize=(35,8))
plt.grid()
plt.show()
```



While computer sector was a prominent sector until 2006, 2006 onwards services sector took the top notch.

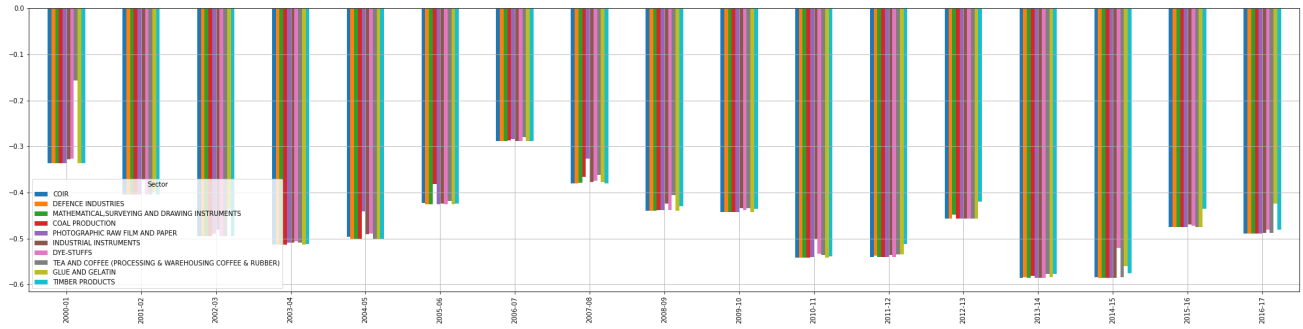
Bottom - 10

```
s = df.sum(axis=1).sort_values(ascending=True)
s.head(10)
```

```
Sector
COIR                                                              4.06
DEFENCE INDUSTRIES                                                5.12
MATHEMATICAL,SURVEYING AND DRAWING INSTRUMENTS                    7.98
COAL PRODUCTION                                                  27.74
PHOTOGRAPHIC RAW FILM AND PAPER                                  67.28
INDUSTRIAL INSTRUMENTS                                           76.12
DYE-STUFFS                                                       88.40
TEA AND COFFEE (PROCESSING & WAREHOUSING COFFEE & RUBBER)       111.22
GLUE AND GELATIN                                                128.39
TIMBER PRODUCTS                                                 157.68
dtype: float64
```

```
df_1 = norm_df.loc[['COIR',
'DEFENCE INDUSTRIES',
'MATHEMATICAL,SURVEYING AND DRAWING INSTRUMENTS',
'COAL PRODUCTION',
'PHOTOGRAPHIC RAW FILM AND PAPER',
'INDUSTRIAL INSTRUMENTS',
'DYE-STUFFS',
'TEA AND COFFEE (PROCESSING & WAREHOUSING COFFEE & RUBBER)',
'GLUE AND GELATIN',
'TIMBER PRODUCTS'], X]
df_1.transpose().plot(kind = 'bar', figsize=(35,8))
```

```
plt.grid()
plt.show()
```



Year wise Max and min Sectors.

```
df.idxmax()
```

```
2000-01                            MISCELLANEOUS INDUSTRIES
2001-02                                 TELECOMMUNICATIONS
2002-03                                AUTOMOBILE INDUSTRY
2003-04                       COMPUTER SOFTWARE & HARDWARE
2004-05                       COMPUTER SOFTWARE & HARDWARE
2005-06                       COMPUTER SOFTWARE & HARDWARE
2006-07    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
2007-08    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
2008-09    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
2009-10    CONSTRUCTION DEVELOPMENT: Townships, housing, ...
2010-11    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
2011-12    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
2012-13    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
2013-14                          FOOD PROCESSING INDUSTRIES
2014-15    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
2015-16    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
2016-17    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
dtype: object
```

```
df.idxmin()
```

```
2000-01                          NON-CONVENTIONAL ENERGY
```

```
2001-02                          NON-CONVENTIONAL ENERGY
2002-03                                  COAL PRODUCTION
2003-04    PRIME MOVER (OTHER THAN ELECTRICAL GENERATORS)
2004-05                                  COAL PRODUCTION
2005-06              BOILERS AND STEAM GENERATING PLANTS
2006-07                                           PORTS
2007-08                           SCIENTIFIC INSTRUMENTS
2008-09              BOILERS AND STEAM GENERATING PLANTS
2009-10                                  COAL PRODUCTION
2010-11                                  COAL PRODUCTION
2011-12                                  COAL PRODUCTION
2012-13                                  COAL PRODUCTION
2013-14    MATHEMATICAL,SURVEYING AND DRAWING INSTRUMENTS
2014-15                                  COAL PRODUCTION
2015-16                                  COAL PRODUCTION
2016-17                                  COAL PRODUCTION
dtype: object
```

The min values shows that while COIR is sector with total investment, in majority years COAL Production had least investment, while COIR didn't have the minimum investment in any year.