Foreign Direct investment Data

First of import the necessary python libraries

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

Read the Csv File

In [2]:
```
fdi=pd.read_csv('/content/FDI data.csv')
```

Find the Shape

In [3]:
```
fdi.shape
```

Out[3]:
```
(63, 18)
```

Find Rows

In [4]:
```
fdi.shape[0]
```

Out[4]:
```
63
```

Find Columns

In [5]:
```
fdi.shape[1]
```

Out[5]:
```
18
```

To get familiar with the data print first 5 and Last 5 records

In [6]:
```
fdi.head()
```

Out[6]:

	Sector	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007-08
0	METALLURGICAL INDUSTRIES	22.69	14.14	36.61	8.11	200.38	149.13	169.94	1175.75
1	MINING	1.32	6.52	10.06	23.48	9.92	7.40	6.62	444.36
2	POWER	89.42	757.44	59.11	27.09	43.37	72.69	157.15	988.68
3	NON-CONVENTIONAL ENERGY	0.00	0.00	1.70	4.14	1.27	1.35	2.44	58.82
4	COAL PRODUCTION	0.00	0.00	0.00	0.04	0.00	9.14	1.30	14.08

In [7]:
```
fdi.tail()
```

Out[7]:

	Sector	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007
58	PRINTING OF BOOKS (INCLUDING LITHO PRINTING IN...	0.00	0.00	6.30	0.00	0.06	9.90	20.04	35
59	COIR	0.00	0.00	0.00	0.00	0.47	0.59	0.04	0
60	CONSTRUCTION (INFRASTRUCTURE) ACTIVITIES	0.00	0.00	0.00	0.00	0.00	0.93	64.06	182
61	CONSTRUCTION DEVELOPMENT: Townships, housing, ...	24.33	51.75	36.10	47.04	152.06	228.71	1392.95	3887
62	MISCELLANEOUS INDUSTRIES	832.07	221.37	218.76	235.48	121.83	164.76	304.87	528

Find the info of the data or datatypes

In [8]:
```
fdi.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
```

In [9]:
```
fdi.dtypes
```

Out[9]:
```
Sector      object
2000-01    float64
2001-02    float64
2002-03    float64
2003-04    float64
2004-05    float64
2005-06    float64
2006-07    float64
2007-08    float64
2008-09    float64
2009-10    float64
2010-11    float64
2011-12    float64
2012-13    float64
2013-14    float64
2014-15    float64
2015-16    float64
2016-17    float64
dtype: object
```

Find the starting Index

In [10]:
```
fdi.index
```

Out[10]:
```
RangeIndex(start=0, stop=63, step=1)
```

Find column Names

In [11]:
```
fdi.columns
```

Out[11]:
```
Index(['Sector', '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'],
      dtype='object')
```

Find Missing or Null values

In [12]:
```
fdi.isnull().sum()
```

Out[12]:
```
Sector     0
2000-01    0
2001-02    0
2002-03    0
2003-04    0
2004-05    0
2005-06    0
2006-07    0
2007-08    0
2008-09    0
2009-10    0
2010-11    0
2011-12    0
2012-13    0
2013-14    0
2014-15    0
2015-16    0
2016-17    0
dtype: int64
```

We have no Null or Missing values

Descriptive Statistics

In [13]:
```
fdi.describe()
```

Out[13]:

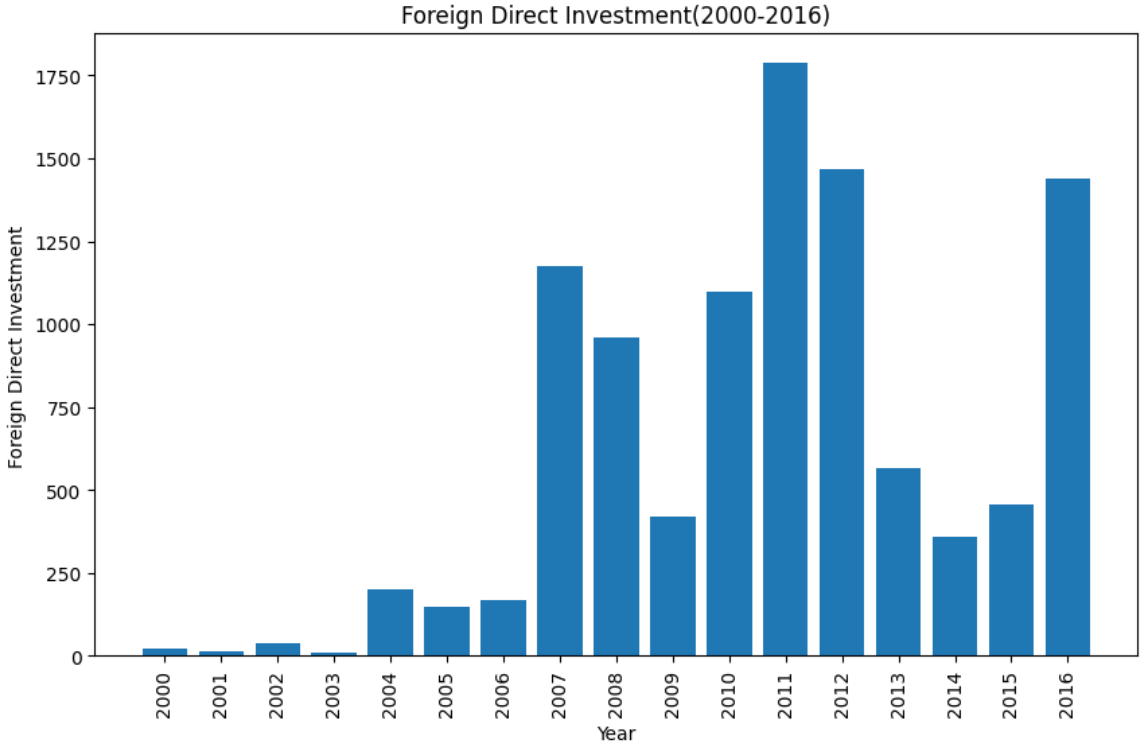
	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07
count	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000
mean	37.757302	63.931587	42.925714	34.727778	51.090317	87.932540	198.281905
std	112.227860	157.878737	86.606439	67.653735	101.934873	206.436967	686.783115
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.200000	0.215000	0.715000	1.230000	4.160000
50%	4.030000	5.070000	11.010000	6.370000	9.090000	22.620000	25.820000
75%	23.510000	44.830000	36.555000	38.660000	43.205000	63.855000	108.325000
max	832.070000	873.230000	419.960000	368.320000	527.900000	1359.970000	4713.780000

Just Rename the Date Columns for our convinience

In [14]:
```
fdi.rename(columns={'2000-01':'2000','2001-02':'2001','2002-03':'2002','2
```

FDI over the years (2000-2016) using bargraph

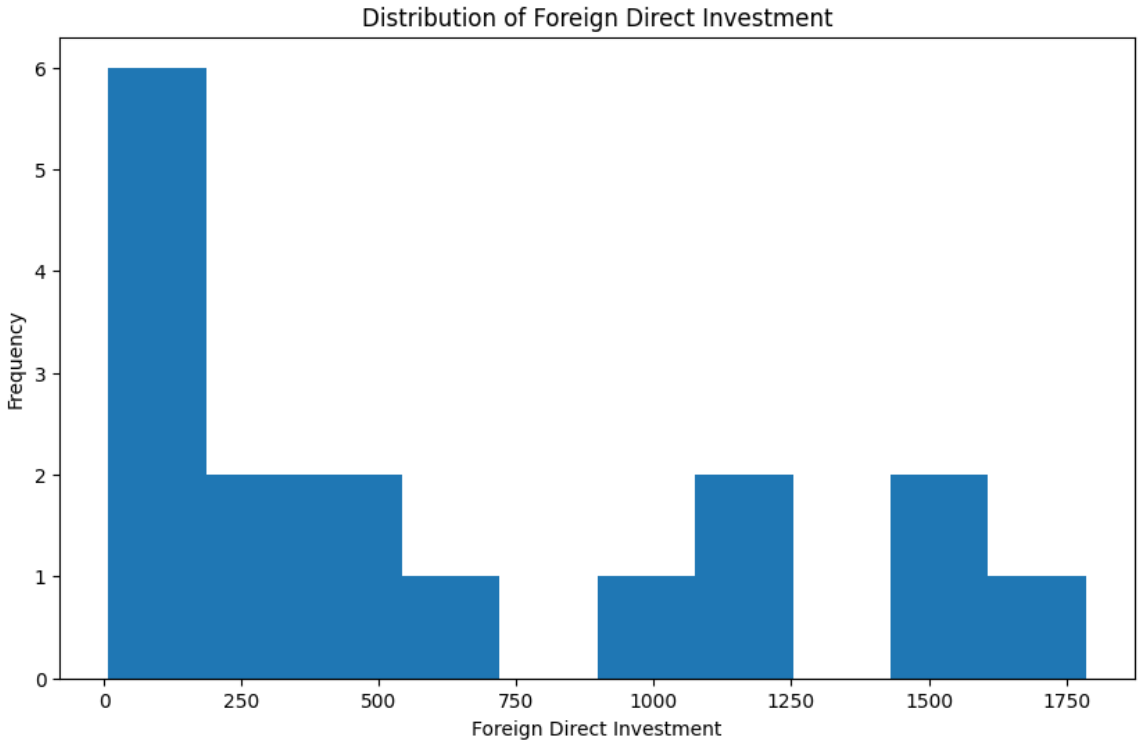
In [15]:
```
plt.figure(figsize=(10, 6))
plt.bar(fdi.columns[1:], fdi.iloc[0, 1:])
plt.xlabel('Year')
plt.ylabel('Foreign Direct Investment')
plt.title('Foreign Direct Investment(2000-2016)')
plt.xticks(rotation=90)
plt.show()
```



Observation:From the above Data we can see that 2011 has High FDI

Analyze the distribution of FDI values

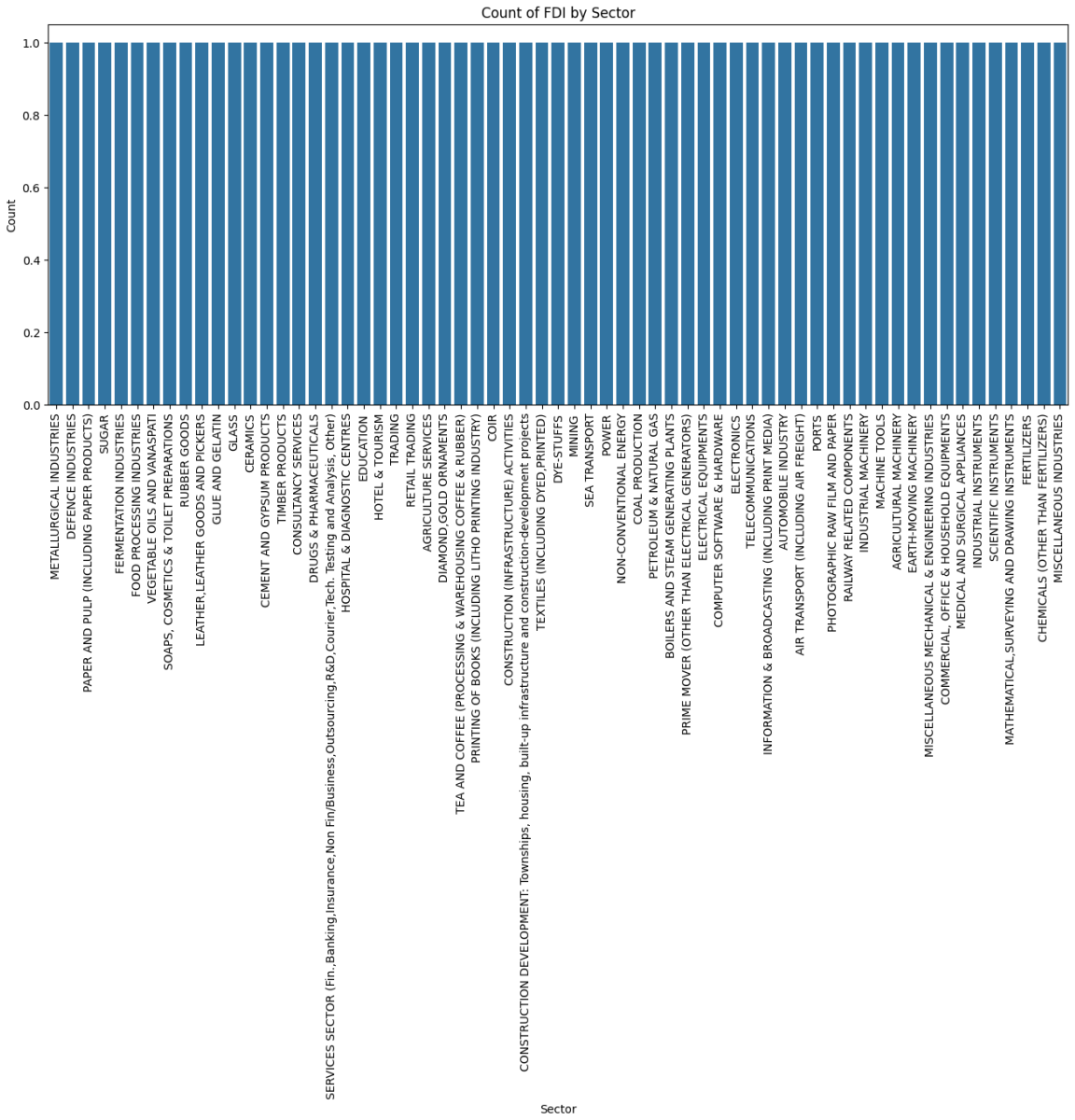
In [16]:
```
plt.figure(figsize=(10, 6))
plt.hist(fdi.iloc[0, 1:])
plt.xlabel('Foreign Direct Investment')
plt.ylabel('Frequency')
plt.title('Distribution of Foreign Direct Investment')
plt.show()
```



Observation: From the above graph we can see the Distribution of Foreign Direct Investment

Using Barplot plot Count of FDI by Sector

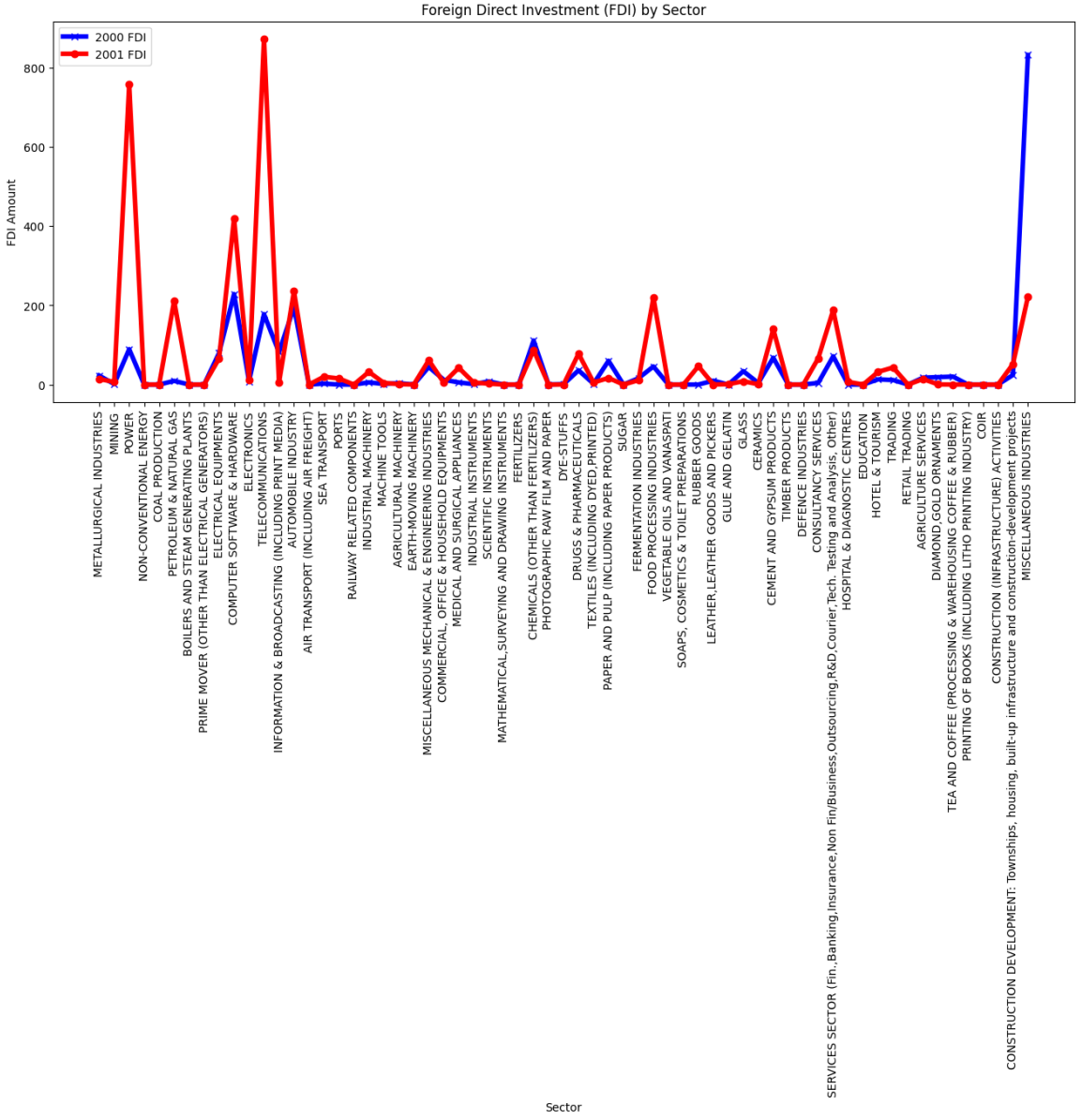
In [17]:
```
sector_counts = fdi['Sector'].value_counts()
plt.figure(figsize=(16, 6))
sns.barplot(x=sector_counts.index, y=sector_counts.values)
plt.xlabel('Sector')
plt.ylabel('Count')
plt.title('Count of FDI by Sector')
plt.xticks(rotation=90)
plt.show()
```



Observation: From the above plot we can see the count of Foreign Direct Investment by Sector

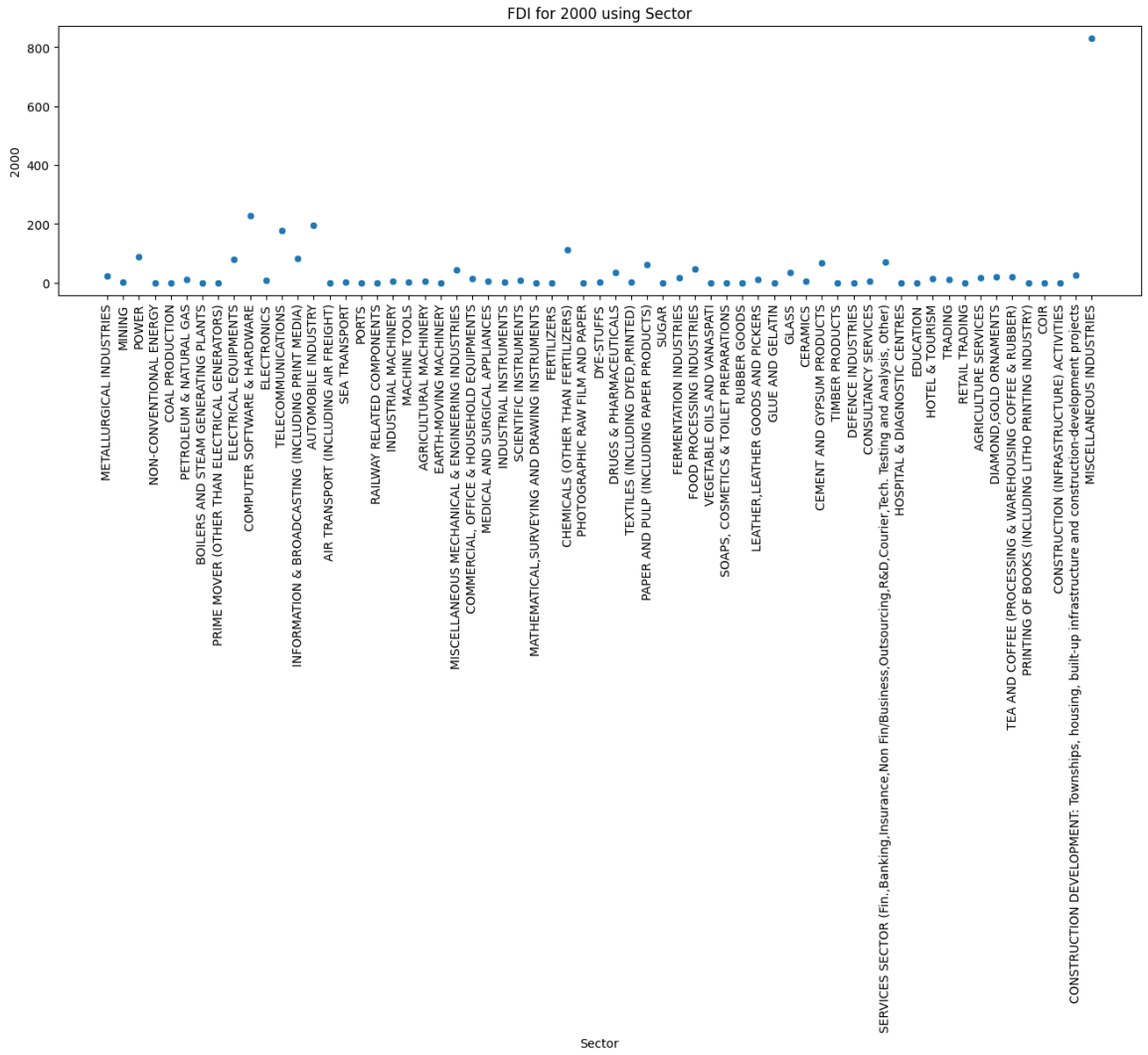
Multiplot for to find the FDI by sector on the year 2000 and 2001

In [18]:
```
plt.figure(figsize=(16, 6))
plt.plot(fdi['Sector'], fdi['2000'], 'b-x',lw=4, label='2000 FDI')
plt.plot(fdi['Sector'], fdi['2001'], 'r-o',lw=4, label='2001 FDI')
plt.xticks(rotation=90);
plt.title("Foreign Direct Investment (FDI) by Sector")
plt.xlabel("Sector")
plt.ylabel("FDI Amount")
plt.legend();
```



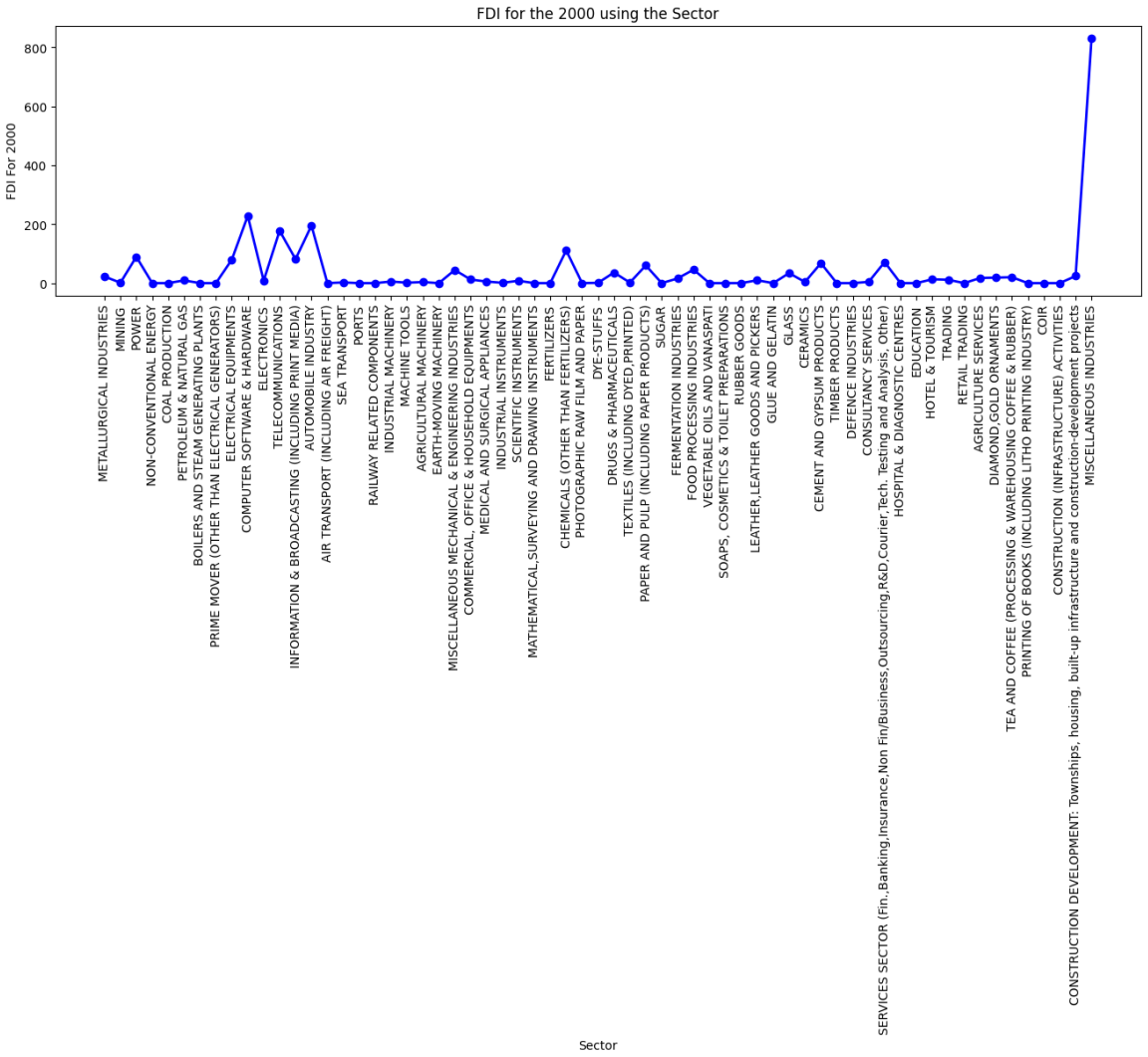
Observation: From the above plot we find the FDI For the year 2000 and 2001

In [19]:
```
plt.figure(figsize=(16, 4))
 sns.scatterplot(x='Sector',y='2000',data=fdi,label='Year 2000',legend=Fa
 plt.xticks(rotation=90)
 plt.title('FDI for 2000 using Sector')
 plt.show()
```



Find FDI for the year 2000 using Sector

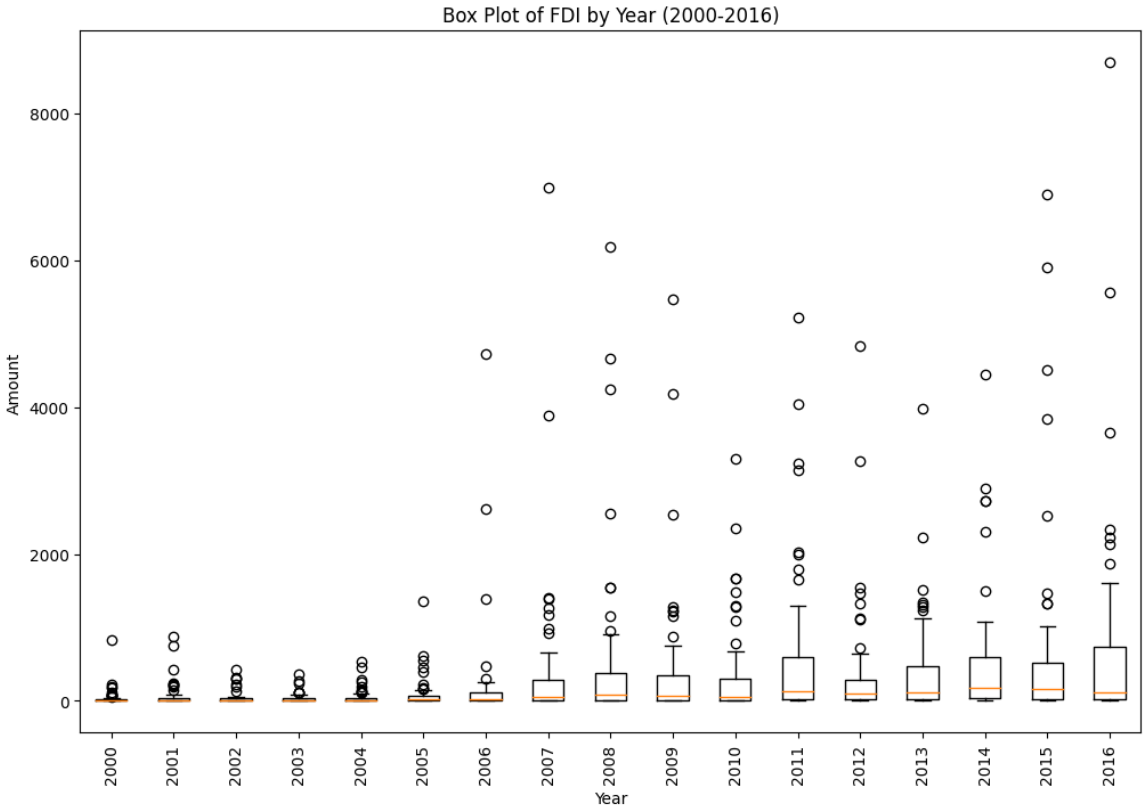
In [20]:
```
plt.figure(figsize=(16,4))
plt.plot(fdi['Sector'],fdi['2000'],'b-o',lw=2)
plt.xlabel('Sector')
plt.title('FDI for the 2000 using the Sector')
plt.ylabel('FDI For 2000')
plt.xticks(rotation=90)
plt.show()
```



In [21]:
```
plt.figure(figsize=(12,8))
boxplot = []
for col in fdi.columns[1:]:
    boxplot.append(fdi[col])

plt.boxplot(boxplot, labels=fdi.columns[1:])

plt.title("Box Plot of FDI by Year (2000-2016)")
plt.xlabel("Year")
plt.xticks(rotation=90)
plt.ylabel("Amount")
plt.show()
```



Observation From the above box plot we can see FDI for the 2000 - 2016