Read the pandas module

In [1]:

```
import pandas as pd
```

Load the excel data given in the assignment into a DataFrame

In [2]:

```
path='C:/Users/HP/Downloads/'
file='online_course.xlsx'
data=pd.read_excel(path+file)
```

Display the first 5 rows in the DataFrame

In [3]:

```
data.head()
```

Out[3]:

	Institution	Course Title	Course Subject	Participants (Course Content Accessed)	Certified	% Certified	% Grade Higher Than Zero	Total Course Hours (Thousands)	% Male	% Female	% Bachelor's Degree or Higher
0	MITx	Circuits and Electronics	Science, Technology, Engineering, and Mathematics	36105	3003	8.32	28.97	418.94	88.28	11.72	60.68
1	MITx	Introduction to Computer Science and Programming	Computer Science	62709	5783	9.22	39.50	884.04	83.50	16.50	63.04
2	MITx	Introduction to Solid State Chemistry	Science, Technology, Engineering, and Mathematics	16663	2082	12.49	34.89	227.55	70.32	29.68	58.76
3	HarvardX	Introduction to Computer Science	Computer Science	129400	1439	1.11	1.11	220.90	80.02	19.98	58.78
4	HarvardX	Health in Numbers: Quantitative Methods in Cli...	Government, Health, and Social Science	52521	5058	9.64	32.52	804.41	56.78	43.22	88.33

Which of the institutions has created more number of Courses?

In [4]:

```
data.Institution.max()
```

Out[4]:

```
'MITx'
```

Which subject has the maximum number of online courses?

In [5]:

```
l=list(data['Course Subject'])
m=0
r=l[0]
for i in l:
    f=l.count(i)
```

```
    f=f.count(i)
    if f>m:
        m=f
        r=i
print(r)
```

```
Humanities, History, Design, Religion, and Education
```

What are the Mean course hours?

In [6]:

```
data['Total Course Hours (Thousands)'].mean()
```

Out[6]:

```
94.9818275862069
```

What are the Median course hours?

In [7]:

```
data['Total Course Hours (Thousands)'].median()
```

Out[7]:

```
37.71
```

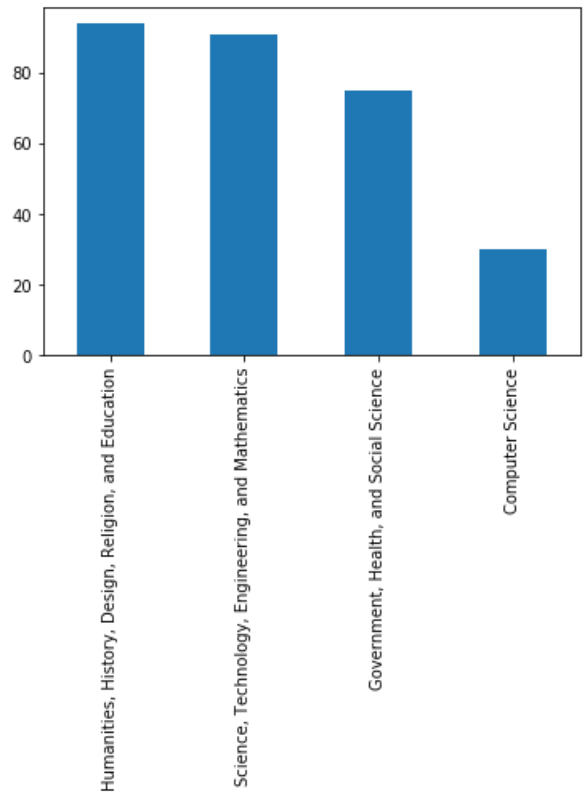
Create a bar plot of the course subjects

In [8]:

```
%matplotlib inline
data['Course Subject'].value_counts().plot.bar()
```

Out[8]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1a969453208>
```



What is the Highest % of Female Students across any course?

In [9]:

```
data['% Female'].max()
```

Out[9]:

74.76

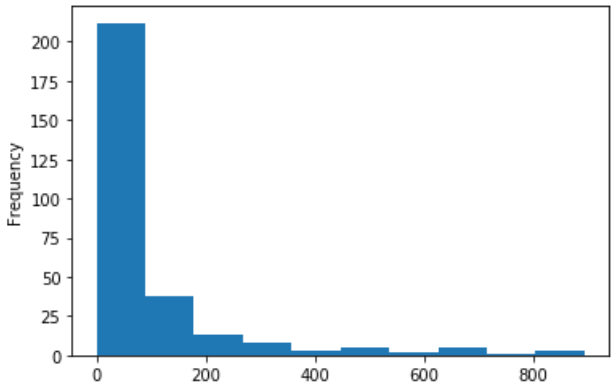
Create a Histogram of Course Hours to study the course trends.

In [10]:

```
data['Total Course Hours (Thousands)'].plot.hist()
```

Out[10]:

<matplotlib.axes._subplots.AxesSubplot at 0x1a96a059b08>



What is the lowest number of participants enrolled across any course?

In [11]:

```
data['Participants (Course Content Accessed)'].min()
```

Out[11]:

322

What is the lowest number of certified students across any course?

In [12]:

```
data['Certified'].min()
```

Out[12]:

0

What is the average % of Bachelor's Degree holders registered in any course?

In [15]:

```
s=data["% Bachelor's Degree or Higher"].sum()
n=data["% Bachelor's Degree or Higher"].count()
print(s/n)
```

72.07872413793103

What is the median of the column '% Grade Higher Than Zero'? What do you think this number means?

In [13]:

```
data['% Grade Higher Than Zero'].median()
```

Out[13]:

```
19.605
```

In [14]:

```
print("Median is the middle term in a series arranged in assending or descending order")
```

```
Median is the middle term in a series arranged in assending or descending order
```

In [ ]:

```

```