

In [16]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.cm as cm
```

In [2]:

```
credit_df = pd.read_csv("C:/Users/aishw/OneDrive/Documents/mukambike.csv")
credit_df
```

Out[2]:

	Unnamed: 0	Income	Limit	Ratings	Cards	Age	Education	Gender	Student	Married	Ethnicity	Balance
0	1	14.891	3606	283	2	34	11	Male	No	Yes	Caucasian	333
1	2	106.025	6645	483	3	82	15	Female	Yes	Yes	Asian	903
2	3	104.593	7075	514	4	71	11	Male	No	No	Asian	580
3	4	148.924	9504	681	3	36	11	Female	No	No	Asian	964
4	5	55.882	4897	357	2	68	16	Male	No	Yes	Caucasian	331
...
94	95	38.009	3075	245	3	45	15	Female	No	No	African American	0
95	96	14.084	855	120	5	46	17	Female	No	Yes	African American	0
96	97	14.312	5382	367	1	59	17	Male	Yes	No	Asian	1380
97	98	26.067	3388	266	4	74	17	Female	No	Yes	African American	155
98	99	36.295	2963	241	2	68	14	Female	Yes	No	African American	375

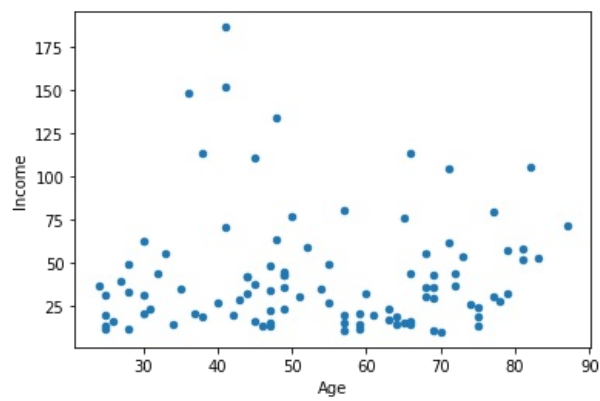
99 rows × 12 columns

In [3]:

```
credit_df.plot('Age', 'Income', kind='scatter', marker='o')
```

Out[3]:

<matplotlib.axes._subplots.AxesSubplot at 0x1b0249c5908>

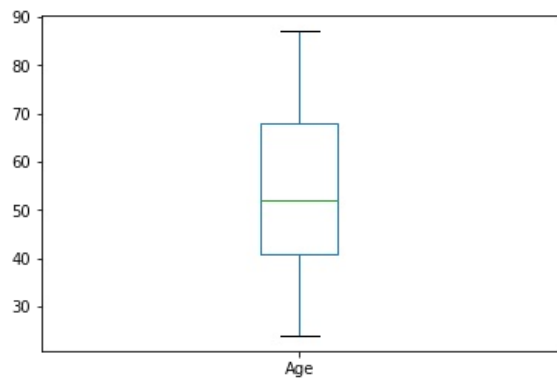


In [4]:

```
credit_df['Age'].plot(kind='box')
```

Out[4]:

<matplotlib.axes._subplots.AxesSubplot at 0x1b025120f88>

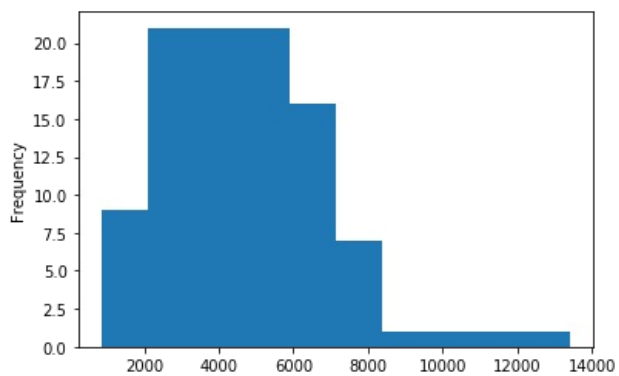


In [6]:

```
credit_df['Limit'].plot(kind='hist')
```

Out[6]:

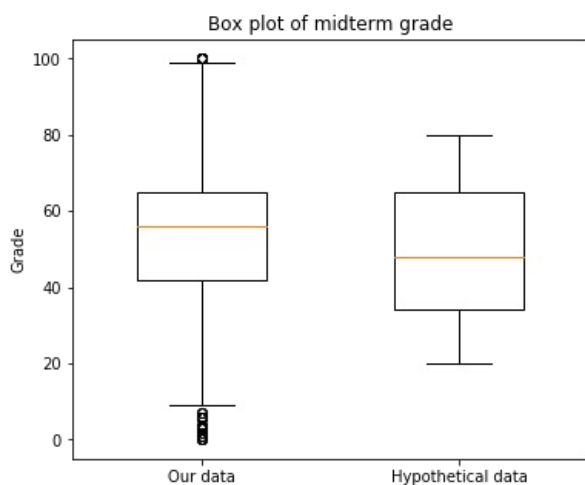
<matplotlib.axes._subplots.AxesSubplot at 0x1b025398d88>



In [7]:

```
import numpy as np
import matplotlib.pyplot as plt
np.random.seed(102)
grades = np.concatenate([[50,52,53,55,56,60,61,62,65,67]*20,
np.random.randint(0, 101, size=300)])
Q1 = np.percentile(grades , 25)
Q3 = np.percentile(grades , 75)
Q1,Q3 = np.percentile(grades , [25,75])
IQR = Q3 - Q1
ul = Q3+1.5*IQR
ll = Q1-1.5*IQR
outliers = grades[(grades > ul) | (grades < ll)]
print(outliers)
fig = plt.figure(figsize=(6,5))
hypo = np.random.randint(20, 81, size=500)
plt.boxplot([grades, hypo], widths=0.5)
plt.xticks([1,2],['Our data', 'Hypothetical data'])
plt.ylabel('Grade')
plt.title('Box plot of midterm grade')
plt.show()
```

[0 7 4 3 0 4 2 7 6 100 1 3 0 3 100 100 100 100
 4 0 3 6 6 6 100 7 6 100 100 6 3 6 1 6 0]



In [10]:

```
import numpy as np
data = [1, 2, 2, 2, 3, 1, 1, 15, 2, 2, 2, 3, 1, 1, 2]
mean = np.mean(data)
std = np.std(data)
print('mean of the dataset is', mean)
print('std. deviation is', std)
threshold = 3
outlier = []
for i in data:
    z = (i-mean)/std
    if z > threshold:
        outlier.append(i)
print('outlier in dataset of Z score is', outlier)
```

mean of the dataset is 2.6666666666666665
std. deviation is 3.3598941782277745
outlier in dataset of Z score is [15]

In [25]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import math
```

In [26]:

```
card_approval_df=pd.read_csv('C:/Users/aksha/OneDrive/Documents/mukambike.csv')
print(card_approval_df.head())
```

	Unnamed: 0	Income	Limit	Ratings	Cards	Age	Education	Gender	Student	\
0	1	14.891	3606	283	2	34	11	Male	No	
1	2	106.025	6645	483	3	82	15	Female	Yes	
2	3	104.593	7075	514	4	71	11	Male	No	
3	4	148.924	9504	681	3	36	11	Female	No	
4	5	55.882	4897	357	2	68	16	Male	No	

	Married	Ethnicity	Balance
0	Yes	Caucasian	333
1	Yes	Asian	903
2	No	Asian	580
3	No	Asian	964
4	Yes	Caucasian	331

In [27]:

```
print(card_approval_df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99 entries, 0 to 98
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Unnamed: 0      99 non-null    int64
1   Income          99 non-null    float64
2   Limit           99 non-null    int64
3   Ratings         99 non-null    int64
4   Cards           99 non-null    int64
5   Age             99 non-null    int64
6   Education       99 non-null    int64
7   Gender          99 non-null    object
8   Student         99 non-null    object
9   Married         99 non-null    object
10  Ethnicity       99 non-null    object
11  Balance         99 non-null    int64
dtypes: float64(1), int64(7), object(4)
memory usage: 9.4+ KB
None
```

In [28]:

```
card_approval_df[['Age', 'Limit', 'Education', 'Gender', 'Income']].describe()
```

Out[28]:

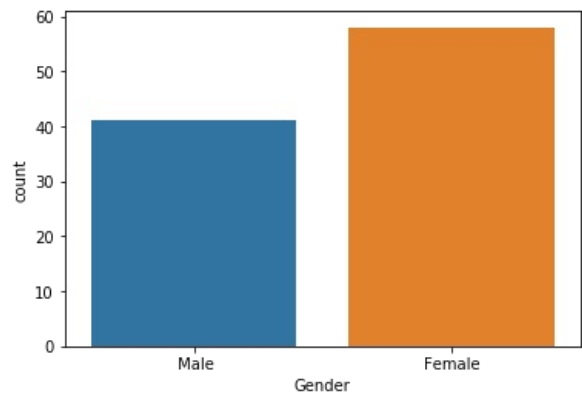
	Age	Limit	Education	Income
count	99.000000	99.000000	99.000000	99.000000
mean	53.585859	4782.161616	13.383838	42.817081
std	17.020402	2217.673117	3.102688	33.736446
min	24.000000	855.000000	5.000000	10.354000
25%	41.000000	3217.000000	11.000000	20.096000
50%	52.000000	4534.000000	14.000000	32.916000
75%	68.000000	6298.500000	16.000000	52.676000
max	87.000000	13414.000000	20.000000	186.634000

In [30]:

```
sns.countplot(card_approval_df.Gender)
```

Out[30]:

<matplotlib.axes._subplots.AxesSubplot at 0x1b027713708>

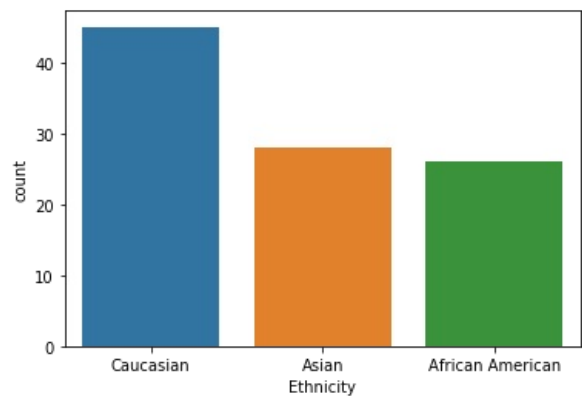


In [31]:

```
sns.countplot(card_approval_df.Ethnicity)
```

Out[31]:

<matplotlib.axes._subplots.AxesSubplot at 0x1b02774dc48>



In [32]:

```
card_approval_df[['Age', 'Limit', 'Education', 'Gender', 'Income']].corr()
```

Out[32]:

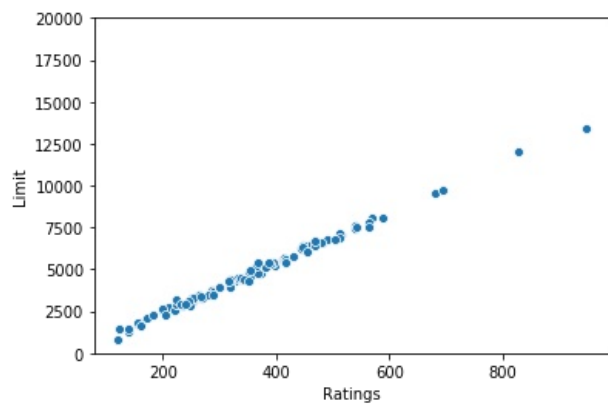
	Age	Limit	Education	Income
Age	1.000000	-0.095059	-0.007200	0.022481
Limit	-0.095059	1.000000	-0.091713	0.783661
Education	-0.007200	-0.091713	1.000000	-0.118052
Income	0.022481	0.783661	-0.118052	1.000000

In [42]:

```
sns.scatterplot(card_approval_df. Ratings ,card_approval_df.Limit)  
plt.ylim(0,20000)
```

Out[42]:

(0, 20000)

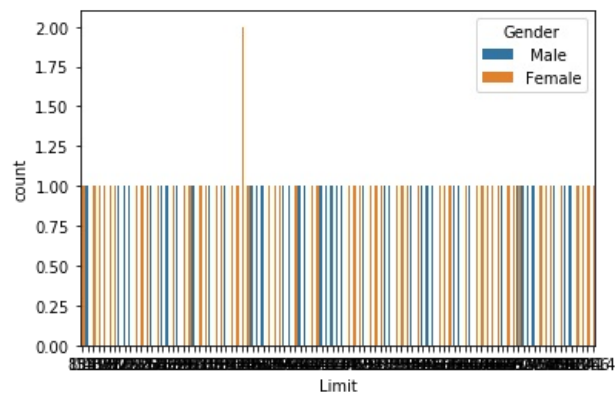


In [51]:

```
sns.countplot(data=card_approval_df,x='Limit',hue='Gender')
```

Out[51]:

<matplotlib.axes._subplots.AxesSubplot at 0x1b02830fe08>



In []: