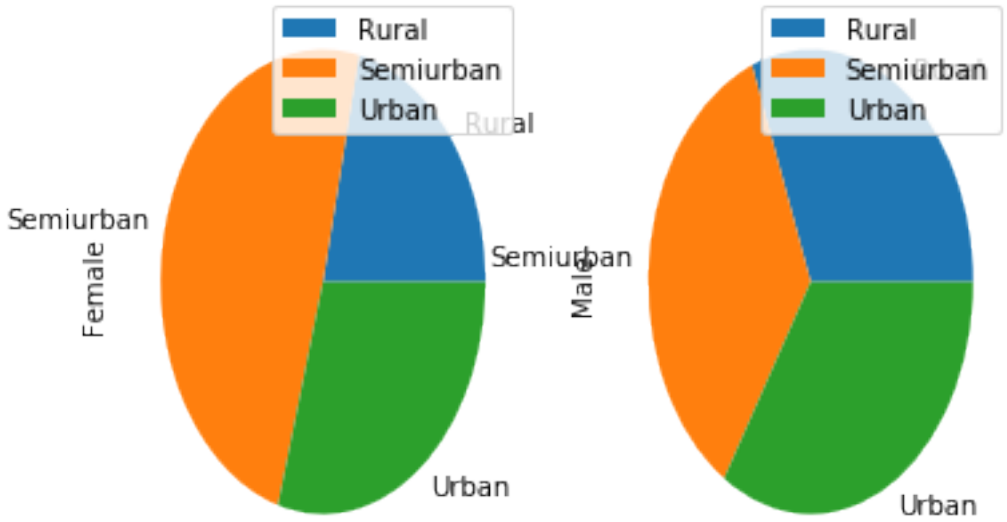
Train

December 4, 2017

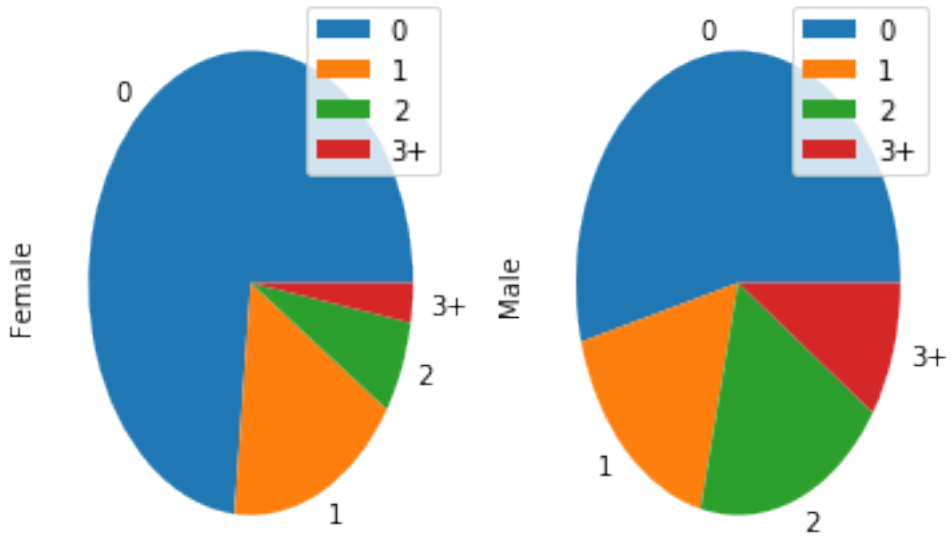
In [1]:
```
#Kan W. 6010163
#Term project for CS1201 Computer Programming I

import pandas as pd
import numpy as np
import matplotlib as plt
import pylab as py
%matplotlib inline
df = pd.read_csv('/Users/Kan/train.csv')
```

In [23]:
```
temp3 = pd.crosstab(df['Property_Area'], df['Gender'])
temp3.plot(kind='pie', subplots=True, grid=False)

#Applicants' property areas based on their gender.
```

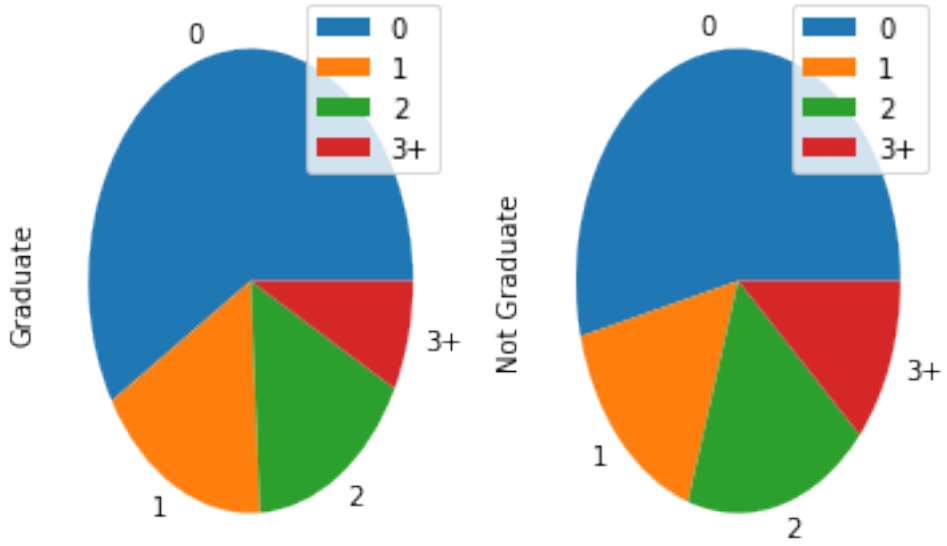
Out[23]:
```
array([<matplotlib.axes._subplots.AxesSubplot object at 0x000001352DEC3588>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000001352AF0C4A8>], dtype=ob
```



```
In [24]: temp3 = pd.crosstab(df['Dependents'], df['Gender'])
         temp3.plot(kind='pie', subplots=True, grid=False)

         #Applicants' dependents based on their gender.

Out[24]: array([<matplotlib.axes._subplots.AxesSubplot object at 0x000001352BD7E780>,
                <matplotlib.axes._subplots.AxesSubplot object at 0x000001352BD423C8>], dtype=ob
```

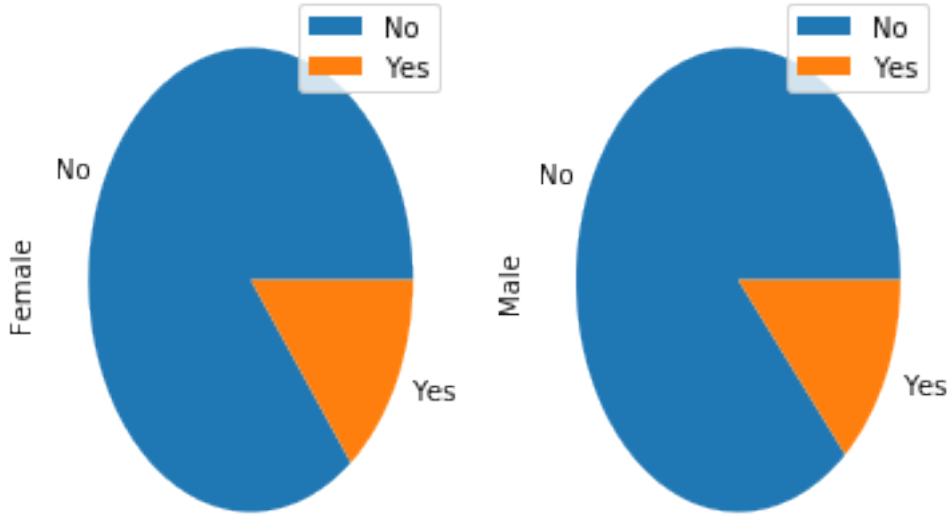



```
In [38]: temp3 = pd.crosstab(df['Dependents'], df['Education'])
         temp3.plot(kind='pie', subplots=True, grid=False)

         #Applicants' dependents based on their education level.

Out[38]: array([<matplotlib.axes._subplots.AxesSubplot object at 0x0000013546864F28>,
                <matplotlib.axes._subplots.AxesSubplot object at 0x0000013546914E48>], dtype=ob
```

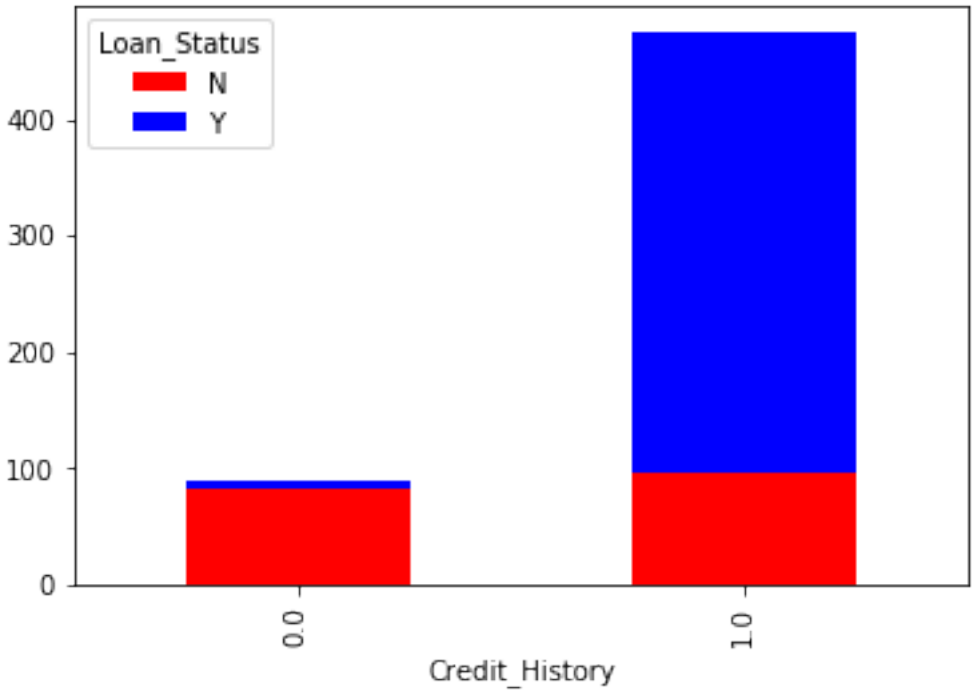


In [27]:
```
temp3 = pd.crosstab(df['Self_Employed'], df['Gender'])
temp3.plot(kind='pie', subplots=True, grid=False)

#Applicants' employment status based on their gender.
```

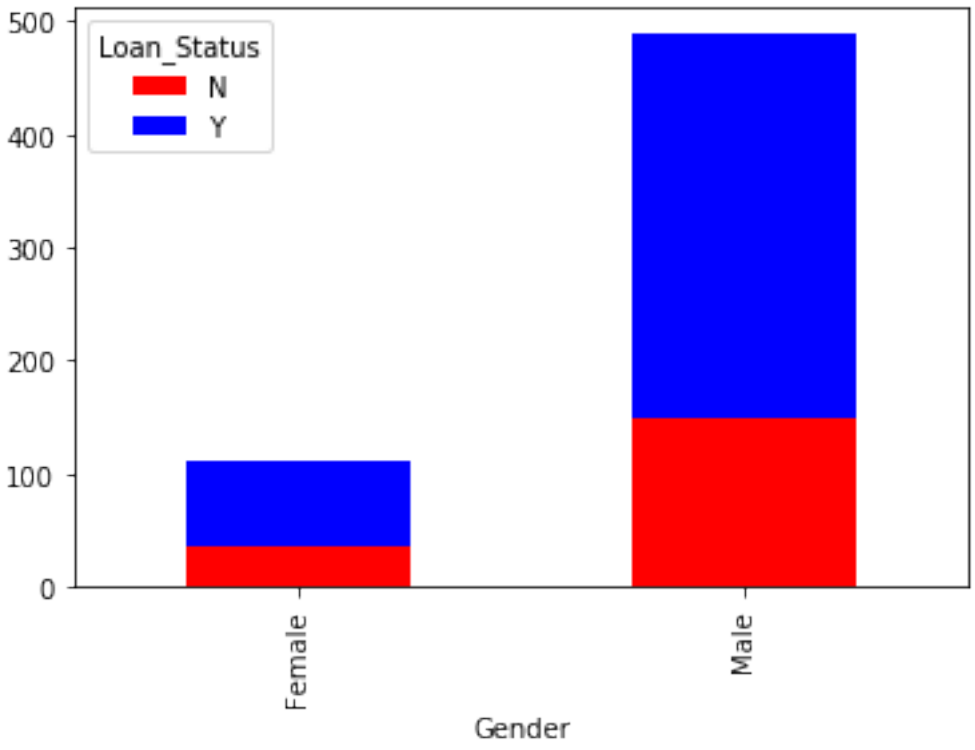
Out[27]:
```
array([<matplotlib.axes._subplots.AxesSubplot object at 0x000001352DD8D748>,
       <matplotlib.axes._subplots.AxesSubplot object at 0x000001352DD211D0>], dtype=ob
```



```
In [17]: temp3 = pd.crosstab(df['Credit_History'], df['Loan_Status'])
         temp3.plot(kind='bar', stacked=True, color=['red','blue'], grid=False)

         #Applicants' loan status based on their credit history.

Out[17]: <matplotlib.axes._subplots.AxesSubplot at 0x23fd26aad30>
```




```
In [4]: temp3 = pd.crosstab(df['Gender'], df['Loan_Status'])
        temp3.plot(kind='bar', stacked=True, color=['red','blue'], grid=False)

        #Applicants' loan status based on their gender.

Out[4]: <matplotlib.axes._subplots.AxesSubplot at 0x2d5849dd748>
```

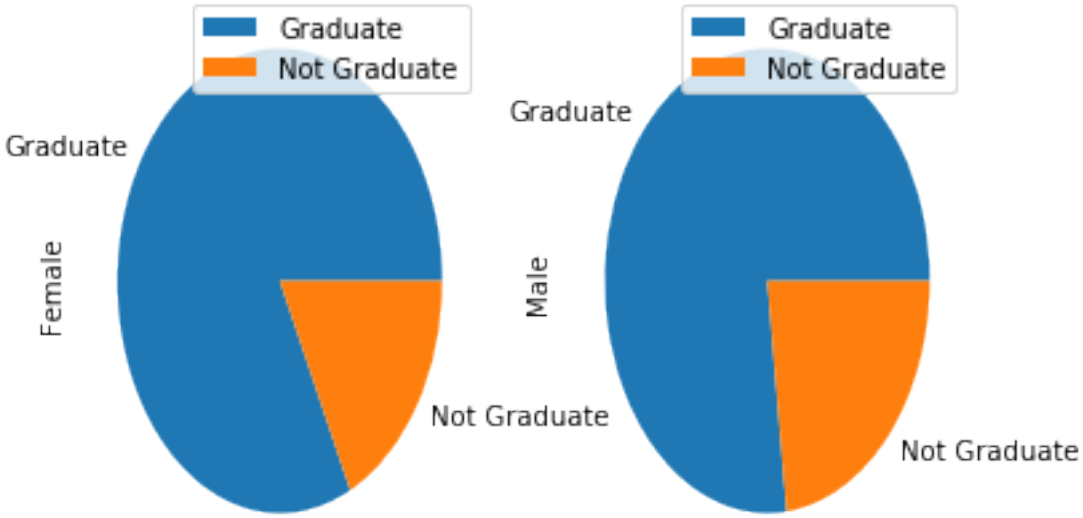



```
In [8]: df.apply(lambda x: sum(x.isnull()),axis=0)

        #Applicants' general information.
```

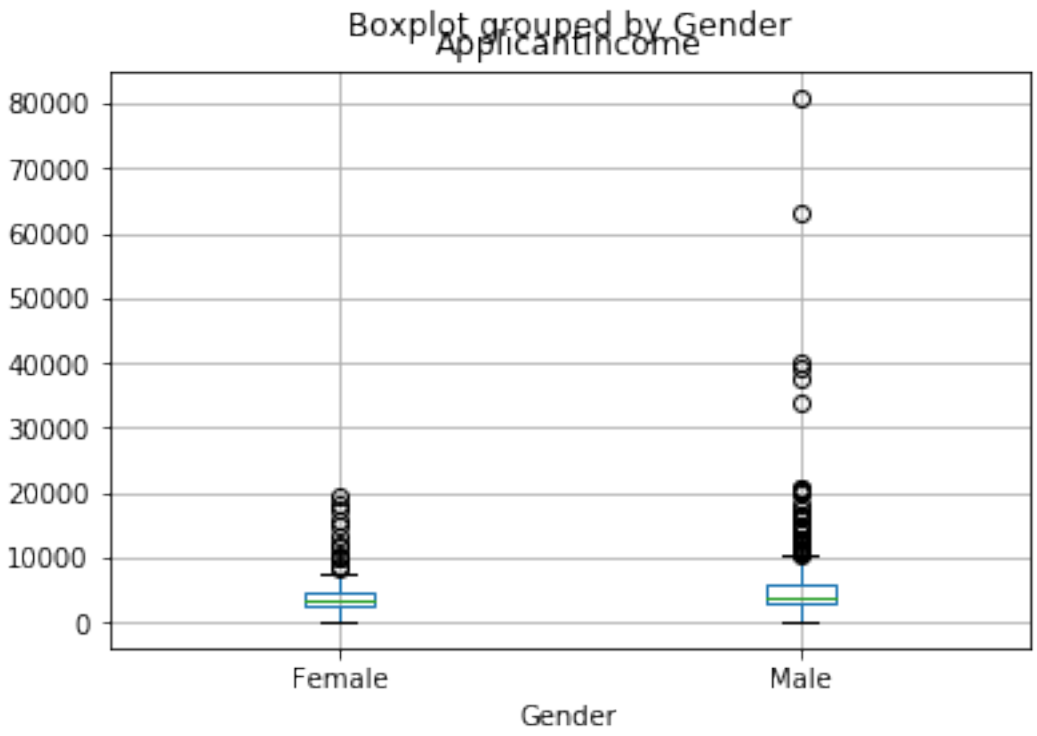
```
Out[8]: Loan_ID               0
        Gender               13
        Married               3
        Dependents           15
        Education             0
        Self_Employed        32
        ApplicantIncome       0
        CoapplicantIncome     0
        LoanAmount           22
        Loan_Amount_Term     14
        Credit_History       50
        Property_Area         0
        Loan_Status           0
        dtype: int64
```

```
In [34]: temp3 = pd.crosstab(df['Education'], df['Gender'])
         temp3.plot(kind='pie', subplots=True, grid=False)

         #Applicants' education level based on their gender.
```

```
Out[34]: array([<matplotlib.axes._subplots.AxesSubplot object at 0x000001352BB21748>,
                <matplotlib.axes._subplots.AxesSubplot object at 0x000001352C0AA438>], dtype=ob
```

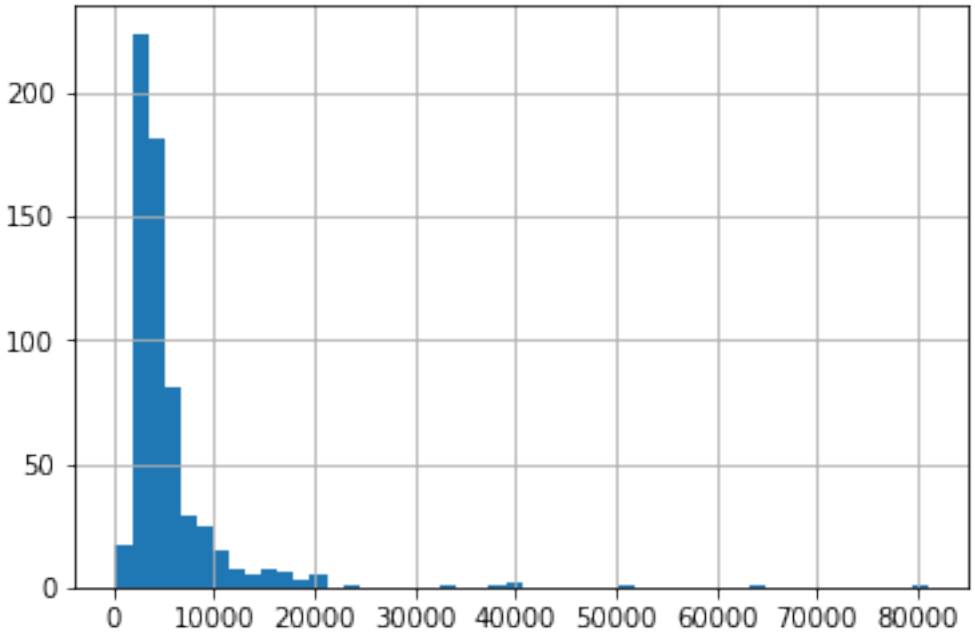



```
In [4]: df.boxplot(column='ApplicantIncome', by = 'Gender')

        #Applicants' income based on their gender.

Out[4]: <matplotlib.axes._subplots.AxesSubplot at 0x1856e0a77b8>
```



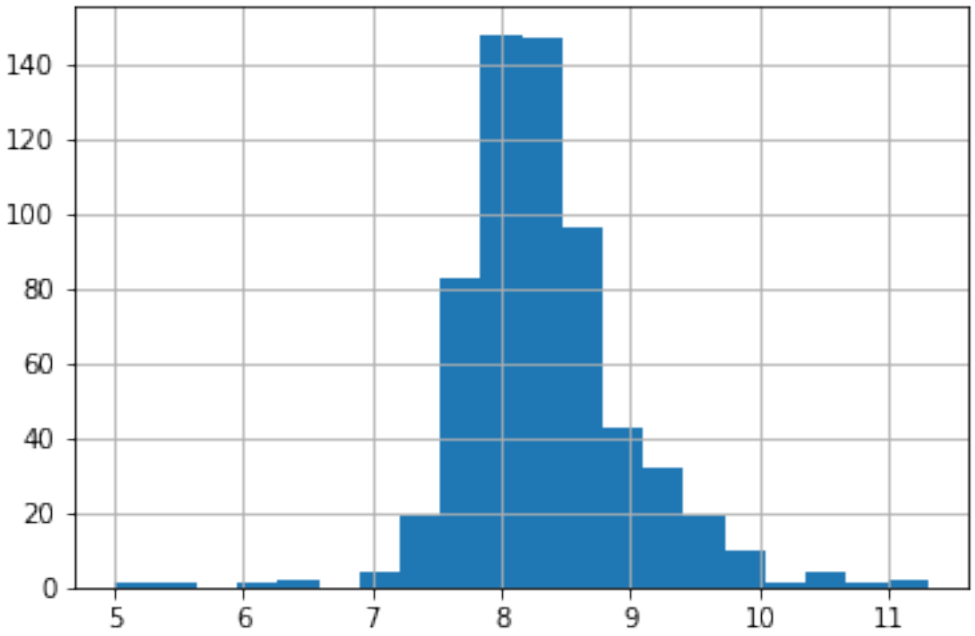

```
In [6]: df['ApplicantIncome'].hist(bins=50)

        #Histogram showing the applicants' income.
        #Most applicants' income are in the range of 0-5000.
```

```
Out[6]: <matplotlib.axes._subplots.AxesSubplot at 0x1f16c983f98>
```



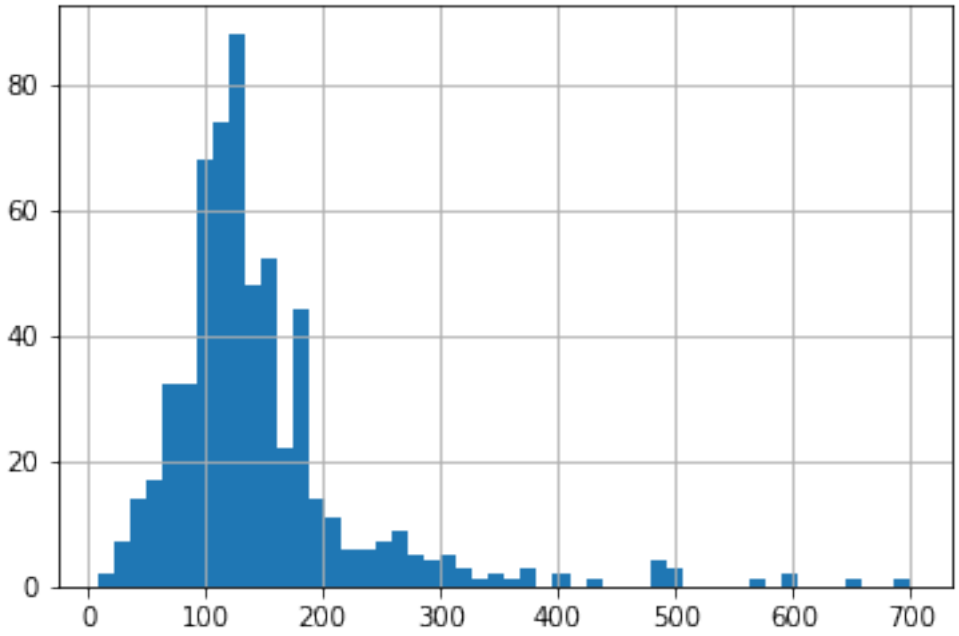
In [22]:
```
df['ApplicantIncome_Log'] = np.log(df['ApplicantIncome'])
df['ApplicantIncome_Log'].hist(bins=20)
```

Out[22]: <matplotlib.axes._subplots.AxesSubplot at 0x2d58641ba90>



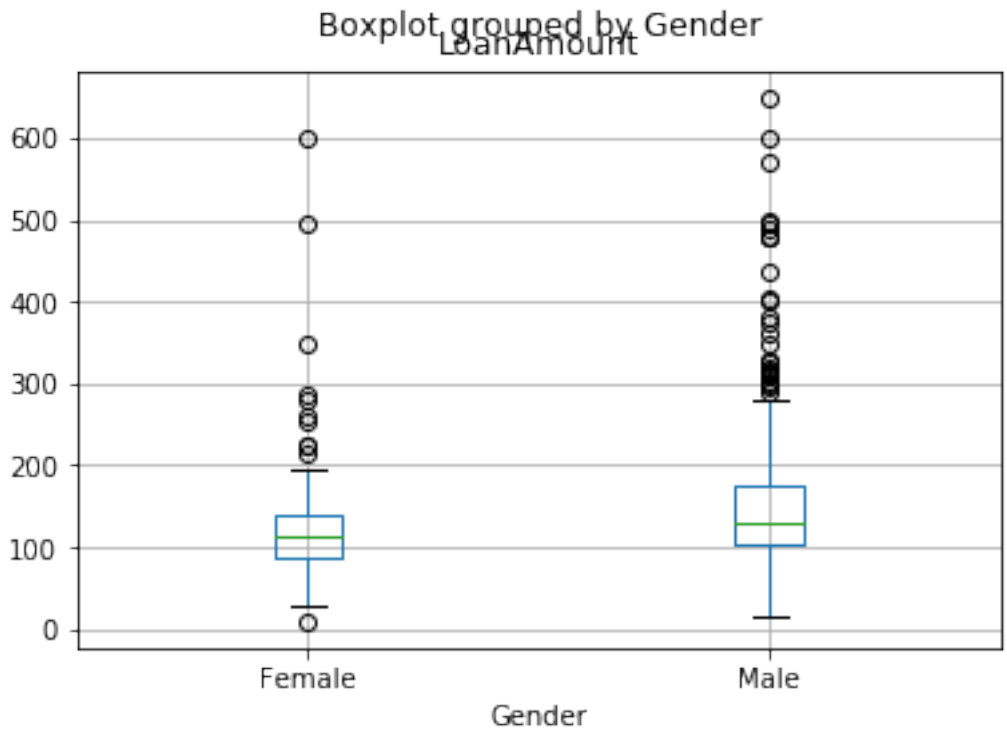
In [12]:
```
df['LoanAmount'].hist(bins=50)

#Histogram showing the applicants' loan amount. Most of them are between 100-200.
```

Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x23fd13059e8>



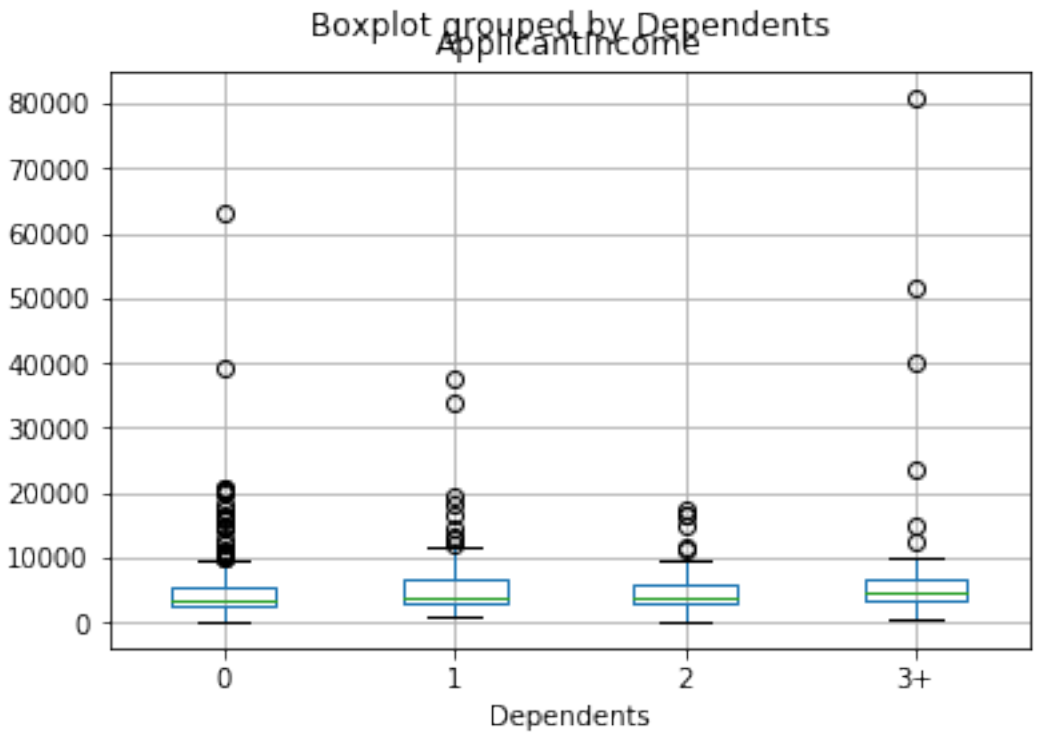
In [39]:
```
df.boxplot(column='LoanAmount', by = 'Gender')

#Loan amount based on applicants' gender.
```

Out[39]: <matplotlib.axes._subplots.AxesSubplot at 0x1cfbe890160>



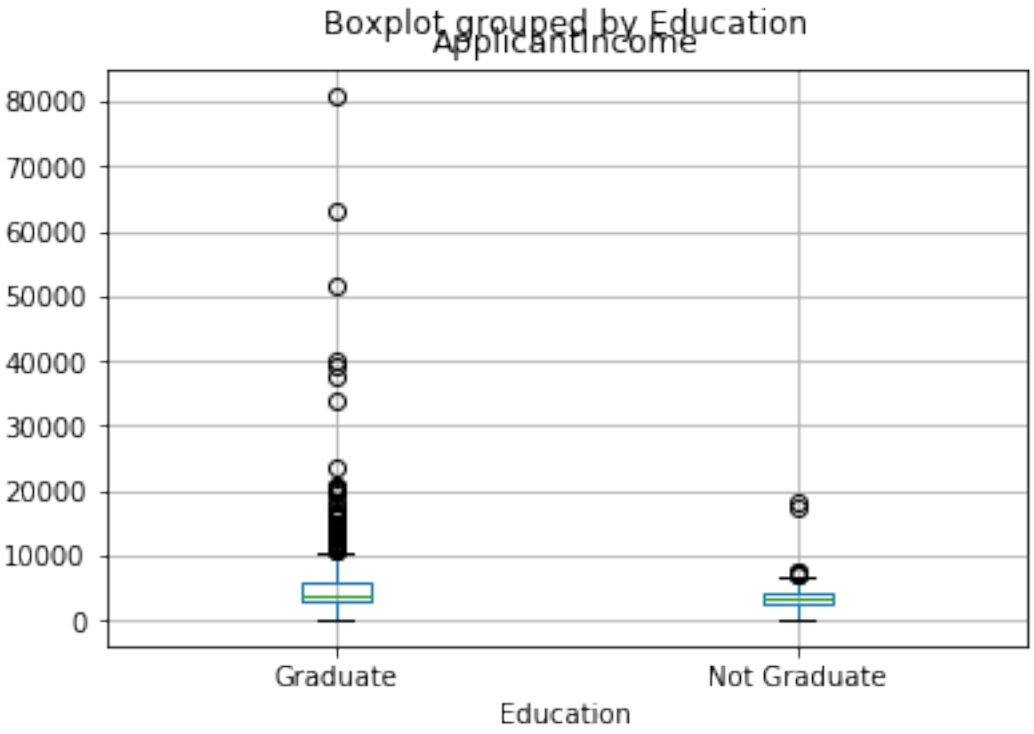

```
In [22]: df.boxplot(column='ApplicantIncome', by = 'Dependents')

         #Applicants' income based on their dependents.
```

```
Out[22]: <matplotlib.axes._subplots.AxesSubplot at 0x1cfb32c2e10>
```




```
In [11]: df.boxplot(column='ApplicantIncome', by = 'Education')

         #Applicants' income based on their education.
```

```
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x1baadb13780>
```


Boxplot grouped by Education
ApplicantIncome
80000
70000
60000
50000
40000
30000
20000
10000
0
Graduate
Not Graduate
Education