

Lab

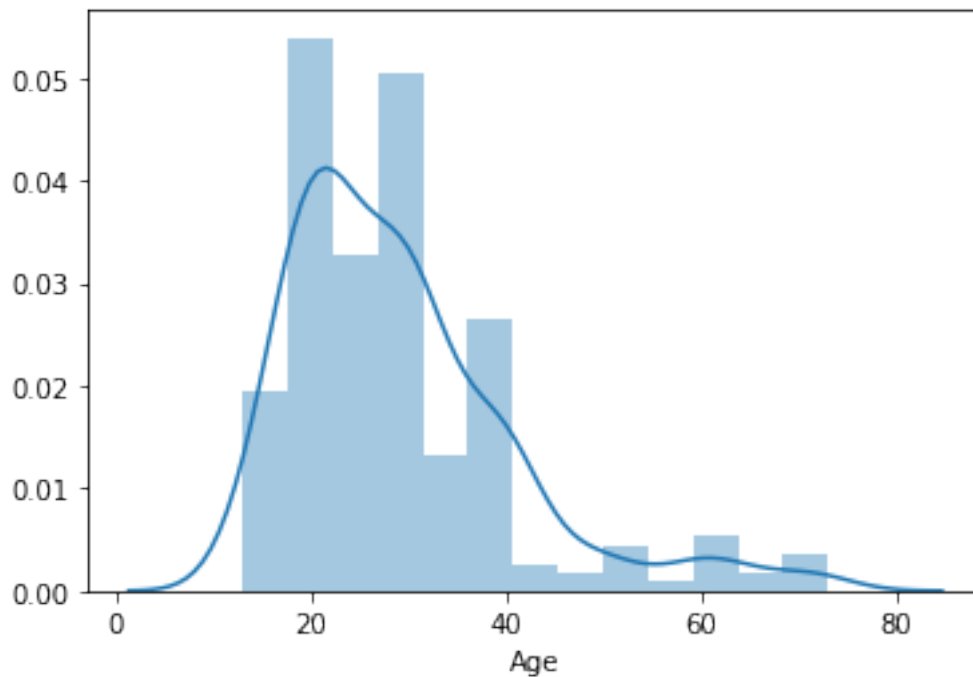
June 14, 2020

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

```
[14]: data = pd.read_csv("Data.csv")
data = data.rename(columns={0 : "Sex", 1 : "Age", 2 : "Choice"})
```

```
[15]: sns.distplot(data[data["Sex"] == "Female"]["Age"])
```

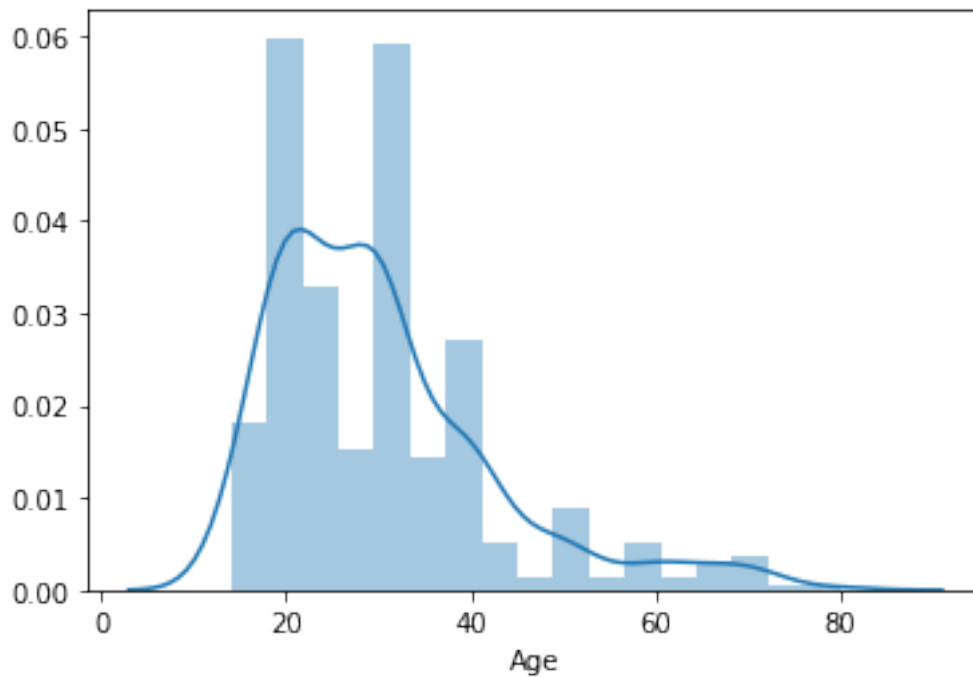
```
[15]: <matplotlib.axes._subplots.AxesSubplot at 0x7fc8d2208990>
```



```
[16]: data = data[data["Age"] >= 14]
slices = [19 + 5*i for i in range(5)]
```

```
data_Teen = data[19 >= data["Age"]]
sns.distplot(data['Age'])
data["Choice"]
```

```
[16]: 0      YES
      1      YES
      2      YES
      3      YES
      4      YES
      ...
      407    YES
      408    YES
      409    NO
      410    IDK
      411    IDK
      Name: Choice, Length: 409, dtype: object
```



```
[17]: def group(age):
      if age < 25:
          return '<25'
      elif age < 35:
          return '<35'
      elif age < 45:
          return '<45'
      else:
```

```
        return '>45'  
data['age_group'] = data['Age'].apply(group)
```

```
[18]: data['numeric_choice'] = data['Choice'].apply(lambda x: 1 if x == 'YES' else 0)
```

```
[19]: sns.barplot(x='age_group', y='numeric_choice', data=data[data['Choice']!='IDK'].  
        ↳sort_values('age_group'), hue='Sex')
```

```
[19]: <matplotlib.axes._subplots.AxesSubplot at 0x7fc8d20bc3d0>
```

