In [3]:

```
import numpy as np, pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

customers = pd.read_csv('Mall_Customers.csv')

customers.head()
```

Out[3]:

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

In [ ]:

```

```

In [5]:

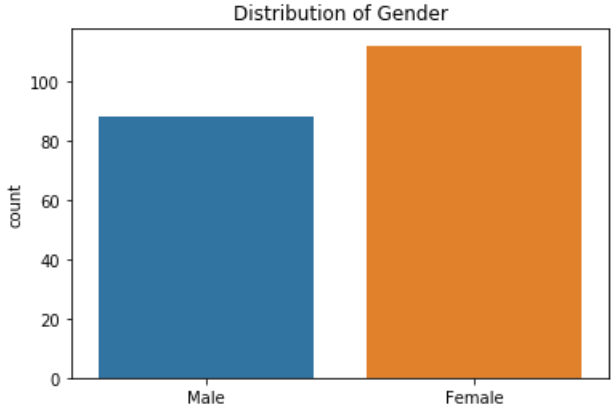
```
customers.describe()
```

Out[5]:

	CustomerID	Age	Annual Income (k$)	Spending Score (1-100)
count	200.000000	200.000000	200.000000	200.000000
mean	100.500000	38.850000	60.560000	50.200000
std	57.879185	13.969007	26.264721	25.823522
min	1.000000	18.000000	15.000000	1.000000
25%	50.750000	28.750000	41.500000	34.750000
50%	100.500000	36.000000	61.500000	50.000000
75%	150.250000	49.000000	78.000000	73.000000
max	200.000000	70.000000	137.000000	99.000000

In [7]:

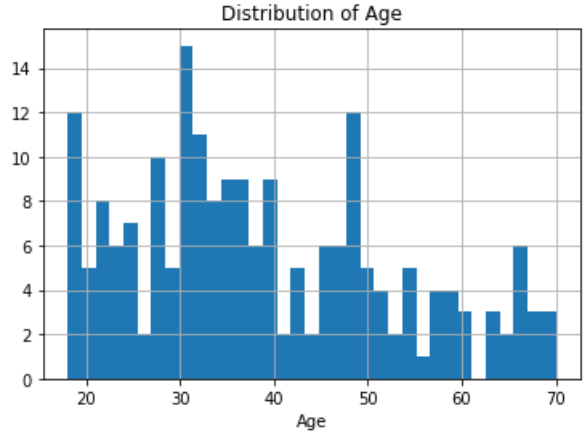
```
sns.countplot(x='Genre', data=customers);
plt.title('Distribution of Gender');
```



Genre

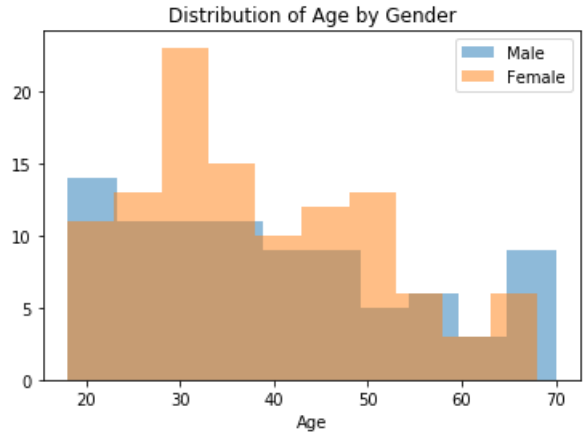
In [8]:

```
# Create a histogram of ages
customers.hist('Age', bins=35);
plt.title('Distribution of Age');
plt.xlabel('Age');
```



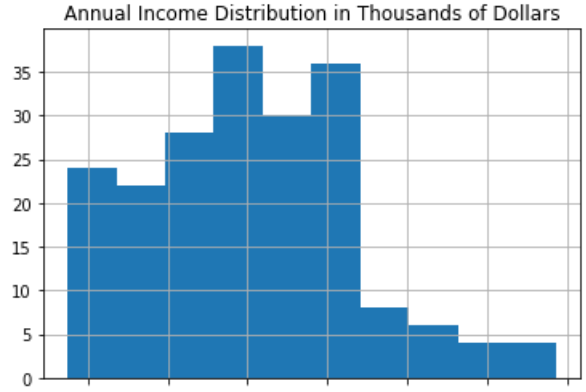
In [10]:

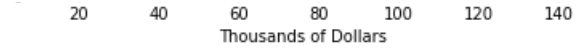
```
plt.hist('Age', data=customers[customers['Genre'] == 'Male'], alpha=0.5, label='Male');
plt.hist('Age', data=customers[customers['Genre'] == 'Female'], alpha=0.5, label='Female');
plt.title('Distribution of Age by Gender');
plt.xlabel('Age');
plt.legend();
```



In [11]:

```
customers.hist('Annual Income (k$)');
plt.title('Annual Income Distribution in Thousands of Dollars');
plt.xlabel('Thousands of Dollars');
```





In [13]:

```
plt.hist('Annual Income (k$)', data=customers[customers['Genre'] == 'Male'], alpha=0.5,
label='Male');
plt.hist('Annual Income (k$)', data=customers[customers['Genre'] == 'Female'], alpha=0.5,
label='Female');
plt.title('Distribution of Income by Gender');
plt.xlabel('Income (Thousands of Dollars)');
plt.legend();
```

In [14]:

```
# Create data sets by gender to save time in the future since gender seems to significantly impact
other variables
male_customers = customers[customers['Genre'] == 'Male']
female_customers = customers[customers['Genre'] == 'Female']

# Print the average spending score for men and women
print(male_customers['Spending Score (1-100)'].mean())
print(female_customers['Spending Score (1-100)'].mean())
```

```
48.51136363636363
51.526785714285715
```

In [16]:

```
sns.scatterplot('Age', 'Annual Income (k$)', hue='Genre', data=customers);
plt.title('Age to Income, Colored by Gender');
```

In [18]:

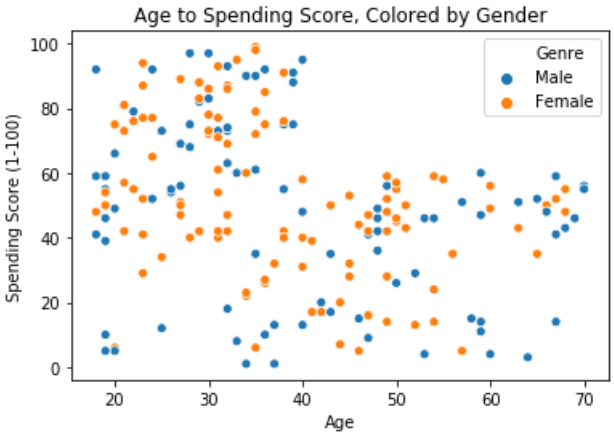
```
sns.heatmap(customers.corr(), annot=True)
```

Out[18]:

```
<matplotlib.axes._subplots.AxesSubplot at 0xe0dfd0>
```

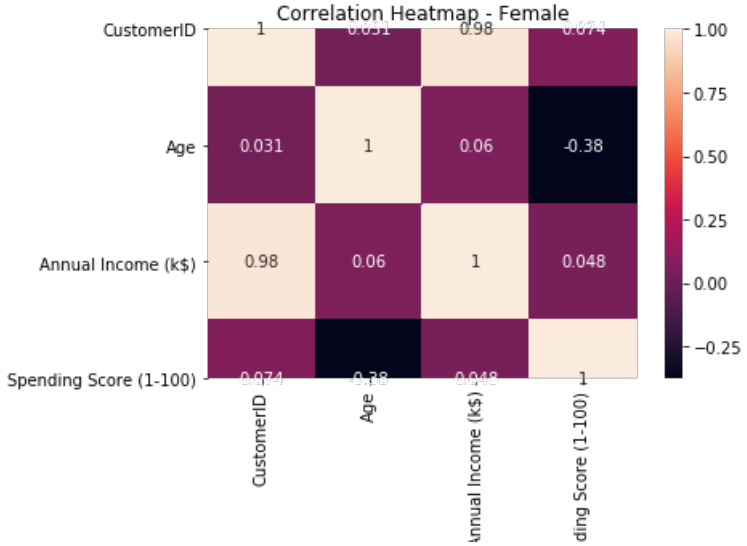
In [19]:

```
sns.scatterplot('Age', 'Spending Score (1-100)', hue='Genre', data=customers);
plt.title('Age to Spending Score, Colored by Gender');
```



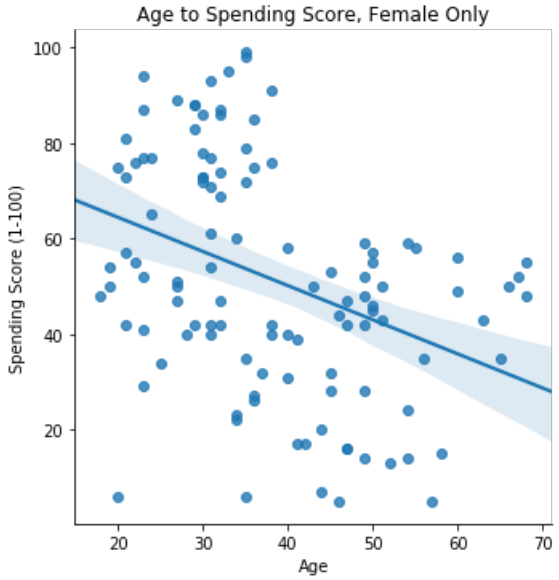
In [20]:

```
sns.heatmap(female_customers.corr(), annot=True);
plt.title('Correlation Heatmap - Female');
```



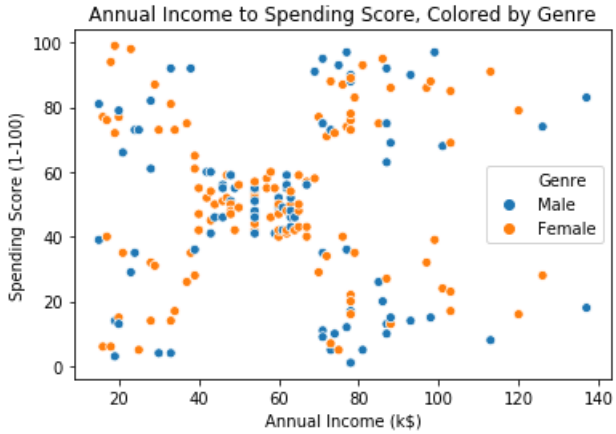
In [21]:

```
sns.lmplot('Age', 'Spending Score (1-100)', data=female_customers);
plt.title('Age to Spending Score, Female Only');
```



In [23]:

```
sns.scatterplot('Annual Income (k$)', 'Spending Score (1-100)', hue='Genre', data=customers);
plt.title('Annual Income to Spending Score, Colored by Genre');
```



In [ ]: