basic_visualizations

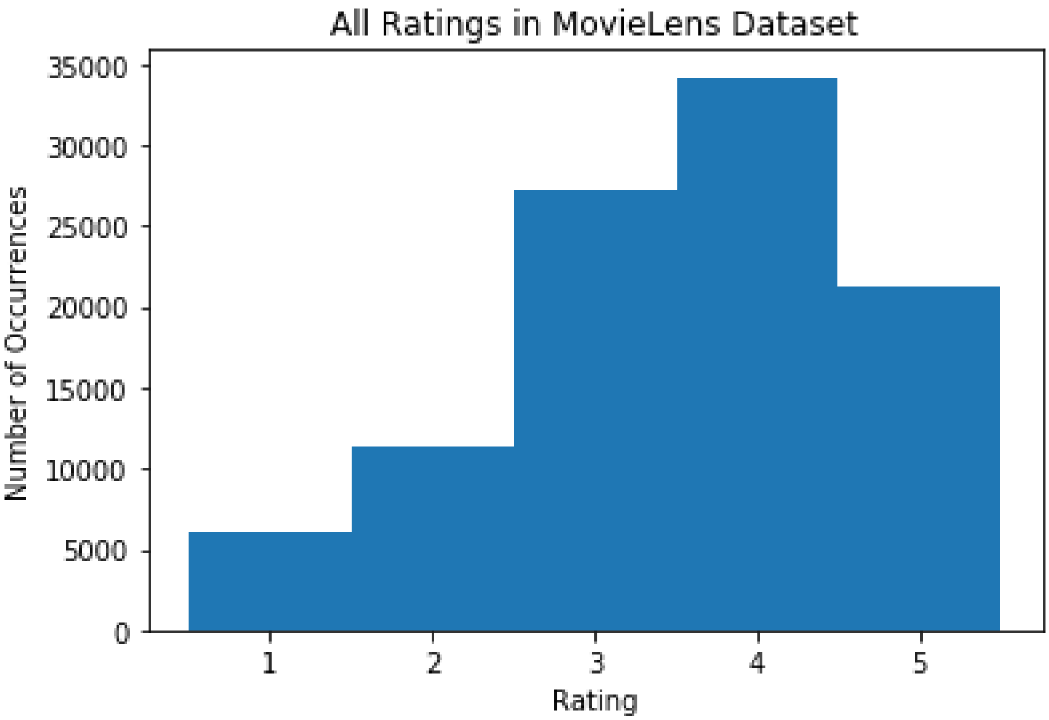
February 28, 2020

0.1 Basic Visualizations

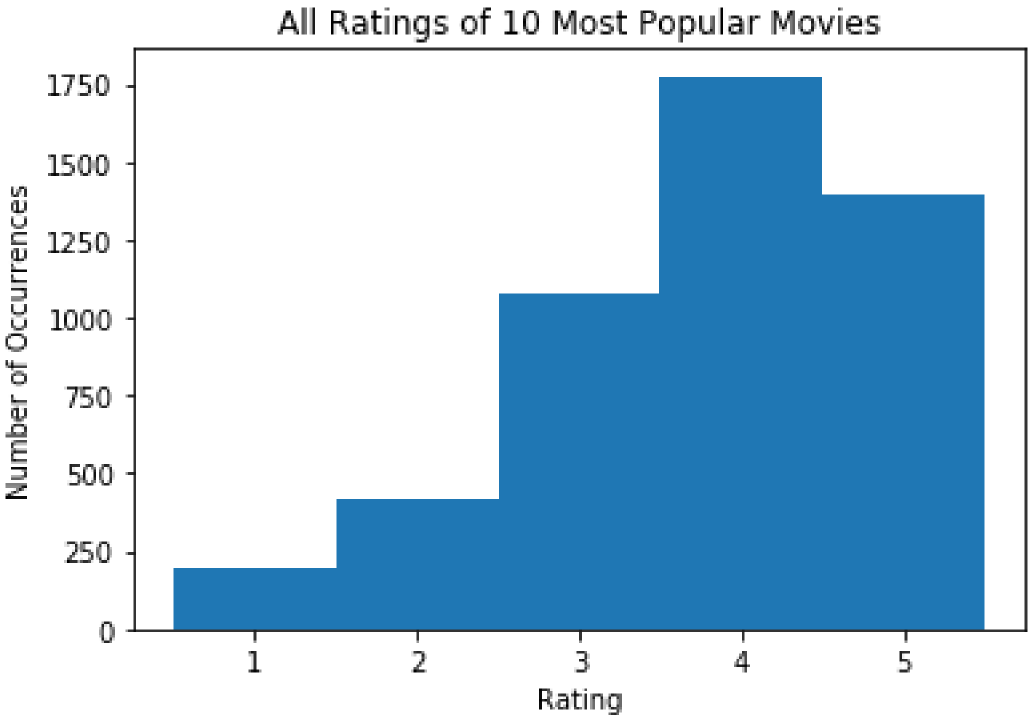
```
In [1]: import matplotlib.pyplot as plt
        import numpy as np
        import pandas as pd
```

```
In [2]: # load data from cleaned files
        movies = pd.read_csv('data/movies.csv')
        data = pd.read_csv('data/data.csv')
```

```
In [3]: # histogram of all ratings
        plt.hist(data['Rating'], bins=[0.5,1.5,2.5,3.5,4.5,5.5])
        plt.title('All Ratings in MovieLens Dataset')
        plt.xlabel('Rating')
        plt.ylabel('Number of Occurrences')
        plt.savefig('basic_1')
```

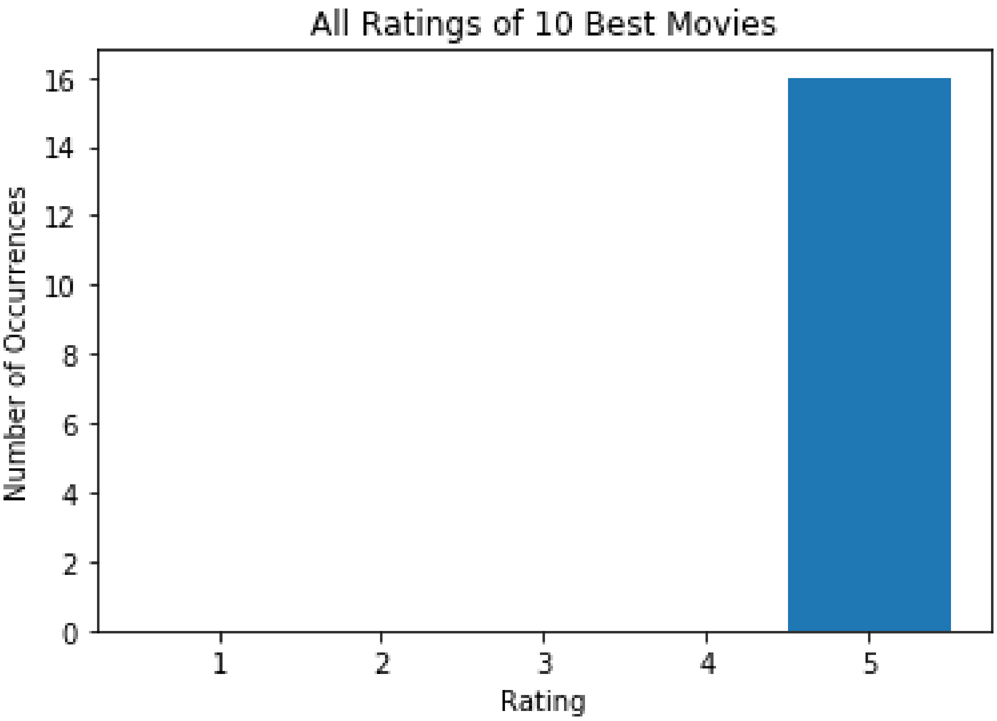


```
In [238]: # get IDs of 10 most popular movies
          pop = data['Movie'].value_counts().head(10).index
          all_ratings = []

          for mov in pop:
              counts = data['Rating'].loc[data['Movie'] == mov]
              all_ratings = np.concatenate((counts, all_ratings))

          # plot all ratings for 10 most popular movies
          plt.hist(all_ratings, bins=[0.5,1.5,2.5,3.5,4.5,5.5])
          plt.title('All Ratings of 10 Most Popular Movies')
          plt.xlabel('Rating')
          plt.ylabel('Number of Occurrences')
          plt.savefig('basic_2')
```



```
In [239]: # get top 10 highest-rated movies
          sorted_df = movies.sort_values(by='avg_rating', ascending=False)
          best = sorted_df.head(10)['ID']

          best_ratings = []
```

```
for idx in best:
    counts = data['Rating'].loc[data['Movie'] == idx]
    best_ratings = np.concatenate((counts, best_ratings))

# plot all ratings for 10 most popular movies
plt.hist(best_ratings, bins=[0.5,1.5,2.5,3.5,4.5,5.5])
plt.title('All Ratings of 10 Best Movies')
plt.xlabel('Rating')
plt.ylabel('Number of Occurrences')
plt.savefig('basic_3')
```



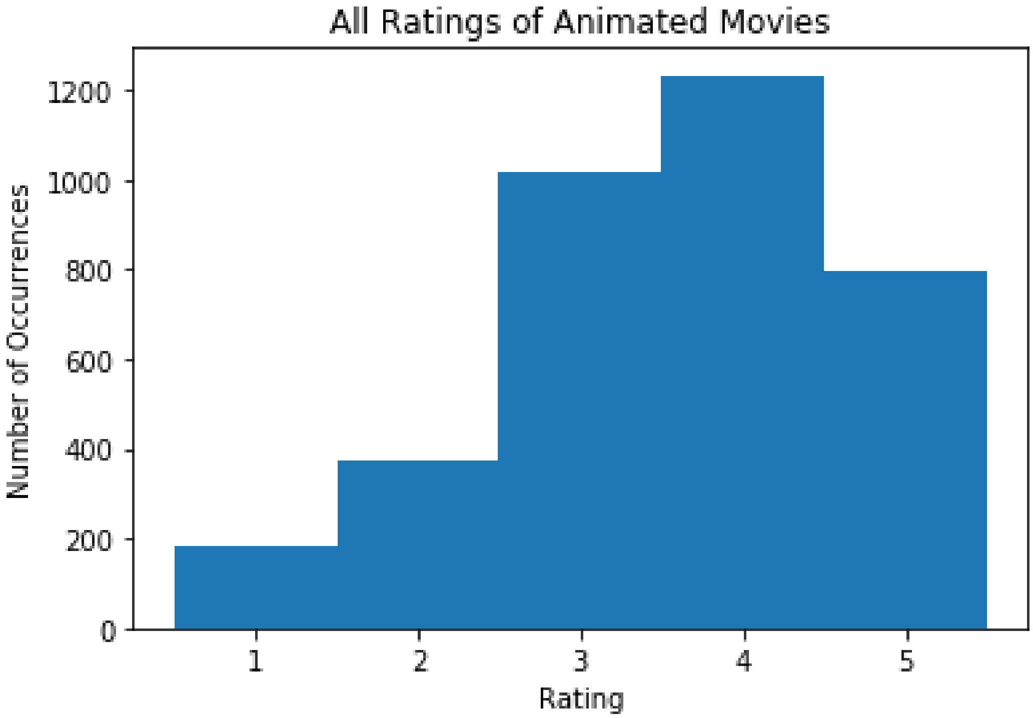
```
In [241]: # see top 10 best movies and verify their average ratings
          sorted_df.head(10)
```

```
Out[241]:         ID                                         Title  Unknown  \
          1588  1589                   Someone Else's America (1995)        0
          1179  1180                              Prefontaine (1997)        0
          1525  1526                            Aiqing wansui (1994)        0
          1282  1283                                 Star Kid (1997)        0
          1637  1638  Entertaining Angels: The Dorothy Day Story (1996)        0
          1112  1113                  They Made Me a Criminal (1939)        0
          1191  1192        Marlene Dietrich: Shadow and Light (1996)        0
          807    808                      Great Day in Harlem, A (1994)        0
          1456  1457              Saint of Fort Washington, The (1993)        0
```

```
1489  1490                          Santa with Muscles (1996)        0

      Action  Adventure  Animation  Childrens  Comedy  Crime  Documentary  \
1588       0          0          0          0       0      0            0
1179       0          0          0          0       0      0            0
1525       0          0          0          0       0      0            0
1282       0          1          0          1       0      0            0
1637       0          0          0          0       0      0            0
1112       0          0          0          0       0      1            0
1191       0          0          0          0       0      0            1
807        0          0          0          0       0      0            1
1456       0          0          0          0       0      0            0
1489       0          0          0          0       1      0            0

      ...  Musical  Mystery  Romance  Sci-Fi  Thriller  War  Western  \
1588  ...        0        0        0       0         0    0        0
1179  ...        0        0        0       0         0    0        0
1525  ...        0        0        0       0         0    0        0
1282  ...        0        0        0       1         0    0        0
1637  ...        0        0        0       0         0    0        0
1112  ...        0        0        0       0         0    0        0
1191  ...        0        0        0       0         0    0        0
807   ...        0        0        0       0         0    0        0
1456  ...        0        0        0       0         0    0        0
1489  ...        0        0        0       0         0    0        0

      num_ratings  tot_rating  avg_rating
1588            1           5         5.0
1179            3          15         5.0
1525            1           5         5.0
1282            3          15         5.0
1637            1           5         5.0
1112            1           5         5.0
1191            1           5         5.0
807             1           5         5.0
1456            2          10         5.0
1489            2          10         5.0

[10 rows x 24 columns]
```

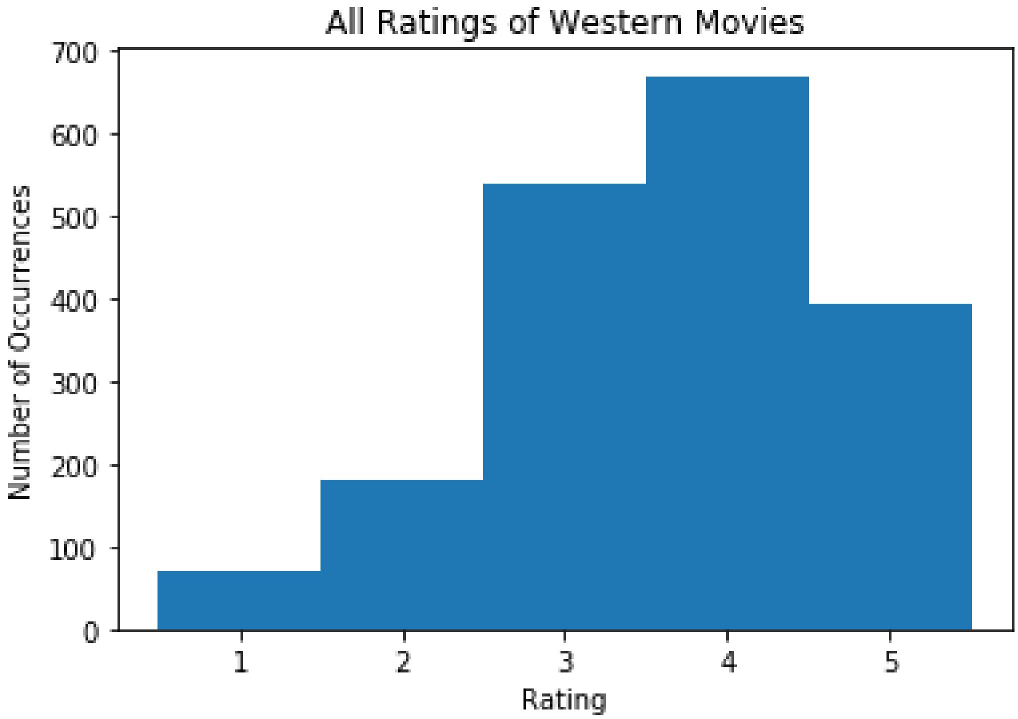
```
In [242]: # get ratings of all animated movies
          animated = movies[movies['Animation'] == 1]['ID']
          anim_ratings = []

          for idx in animated:
              counts = data['Rating'].loc[data['Movie'] == idx]
              anim_ratings = np.concatenate((counts, anim_ratings))
```

```
# plot all ratings for animated movies
plt.hist(anim_ratings, bins=[0.5,1.5,2.5,3.5,4.5,5.5])
plt.title('All Ratings of Animated Movies')
plt.xlabel('Rating')
plt.ylabel('Number of Occurrences')
plt.savefig('basic_4_1')
```

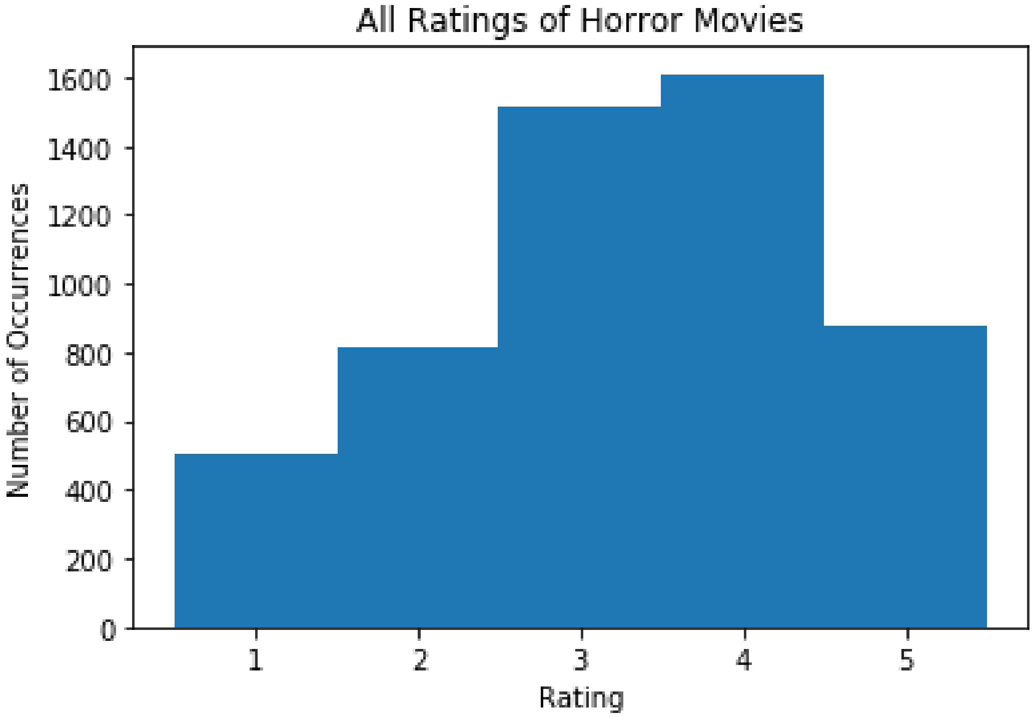


In [243]:

```
# get ratings of all Western movies
western = movies[movies['Western'] == 1]['ID']
west_ratings = []

for idx in western:
    counts = data['Rating'].loc[data['Movie'] == idx]
    west_ratings = np.concatenate((counts, west_ratings))

# plot all ratings for animated movies
plt.hist(west_ratings, bins=[0.5,1.5,2.5,3.5,4.5,5.5])
plt.title('All Ratings of Western Movies')
plt.xlabel('Rating')
plt.ylabel('Number of Occurrences')
plt.savefig('basic_4_2')
```



```
In [244]: # get ratings of all horror movies
          horror = movies[movies['Horror'] == 1]['ID']
          horr_ratings = []

          for idx in horror:
              counts = data['Rating'].loc[data['Movie'] == idx] # add 1 to account for 1-index
              horr_ratings = np.concatenate((counts, horr_ratings))

          # plot all ratings for animated movies
          plt.hist(horr_ratings, bins=[0.5,1.5,2.5,3.5,4.5,5.5])
          plt.title('All Ratings of Horror Movies')
          plt.xlabel('Rating')
          plt.ylabel('Number of Occurrences')
          plt.savefig('basic_4_3')
```


All Ratings of Horror Movies
Number of Occurrences
0
200
400
600
800
1000
1200
1400
1600
Rating
1
2
3
4
5