Altair Exercises

This notebook will explore multiple different visualizations in Altair.

Part 2

The next exercises will be reproducing the following data from FiveThirtyEight's [Creating the next Bechdel Test (https://projects.fivethirtyeight.com/next-bechdel/)]

In [1]:
```
import pandas as pd
import numpy as np
import altair as alt
```

In [2]:
```
# enable correct rendering
alt.renderers.enable('default')

# uses intermediate json files to speed things up
alt.data_transformers.enable('json')
```

Out[2]:
```
DataTransformerRegistry.enable('json')
```

In [3]:
```
# read all the tables
all_tests_df = pd.read_csv('../assets/nextBechdel_allTests.csv')
cast_gender = pd.read_csv('../assets/nextBechdel_castGender.csv')
top_2016 = pd.read_csv('../assets/top_2016.csv')

# set up the tables for use
actors_movies = top_2016.set_index('Movie').join(cast_gender.set_index
movies_order = top_2016.sort_values(by=['Rank'])['Movie'].tolist()
```

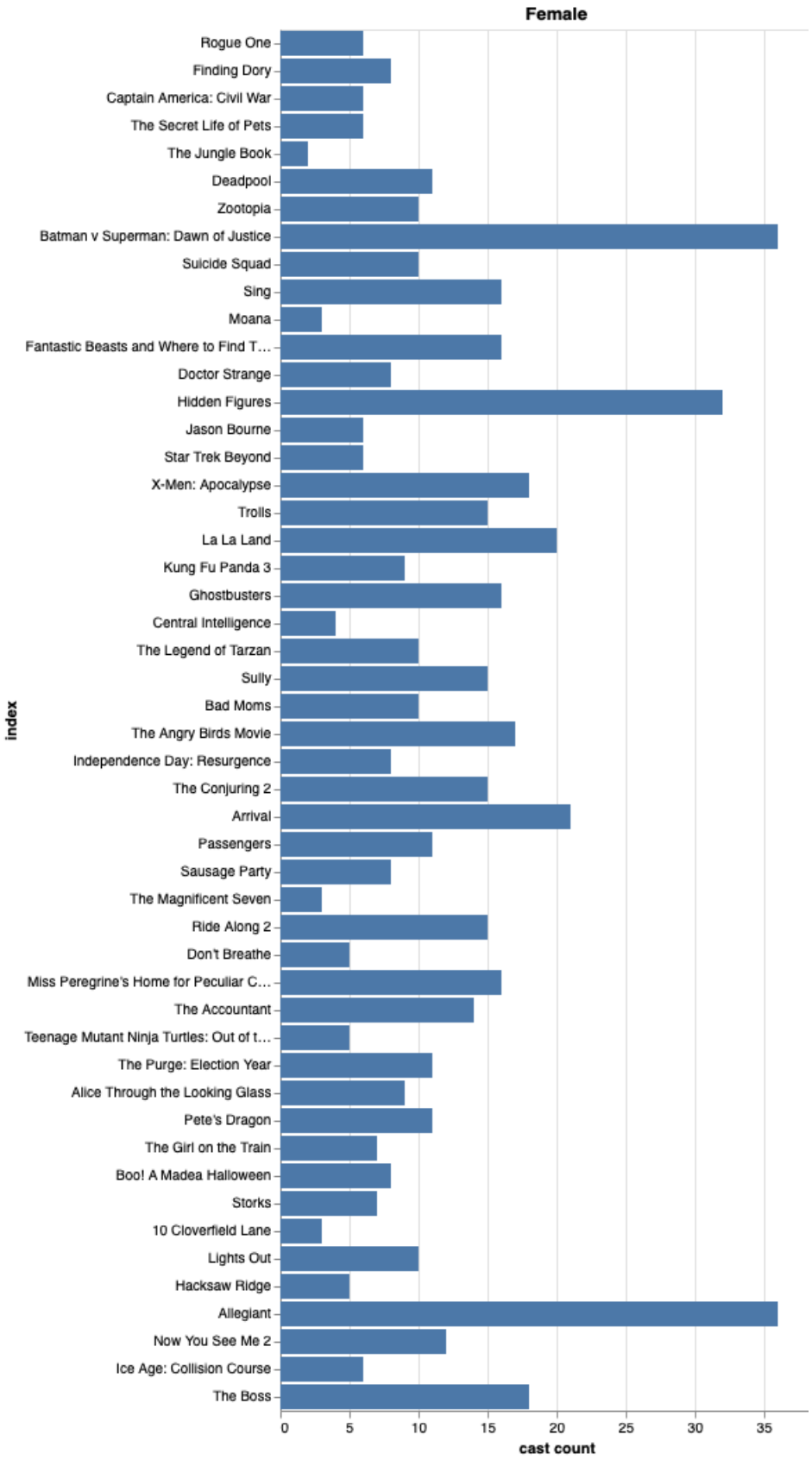
Variables Encoded

In [4]:
```
base = alt.Chart(actors_movies).transform_filter(
    (alt.datum.TYPE != 'Unknown') & (alt.datum.GENDER != 'Unknown') &
)

encoding = base.transform_filter(
    alt.datum.GENDER == 'Female'
).encode(
    y= alt.Y(
        'index:N',
        sort= movies_order
    ),
    x=alt.X('count(index):Q',
            title='cast count'),
)
# Encode bar mark and circle mark

bar = encoding.mark_bar().properties(title='Female')
```

```
circle = encoding.mark_circle().properties(title='Female')
```

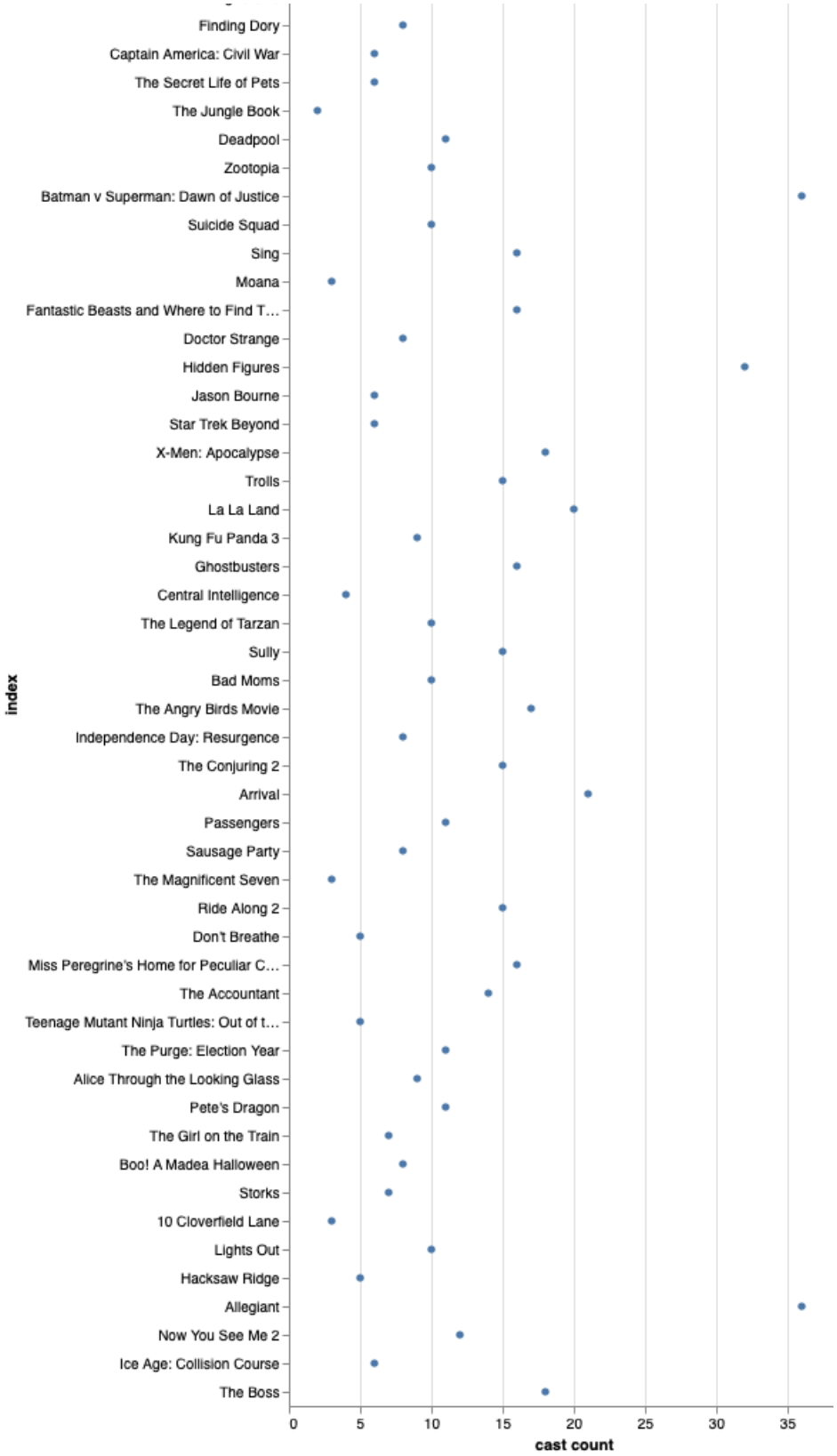
In [5]:
```
bar
```

Out[5]:



In [6]:
```
circle
```

Out[6]:



Increase Variables: Charting Actor/Actress Genders

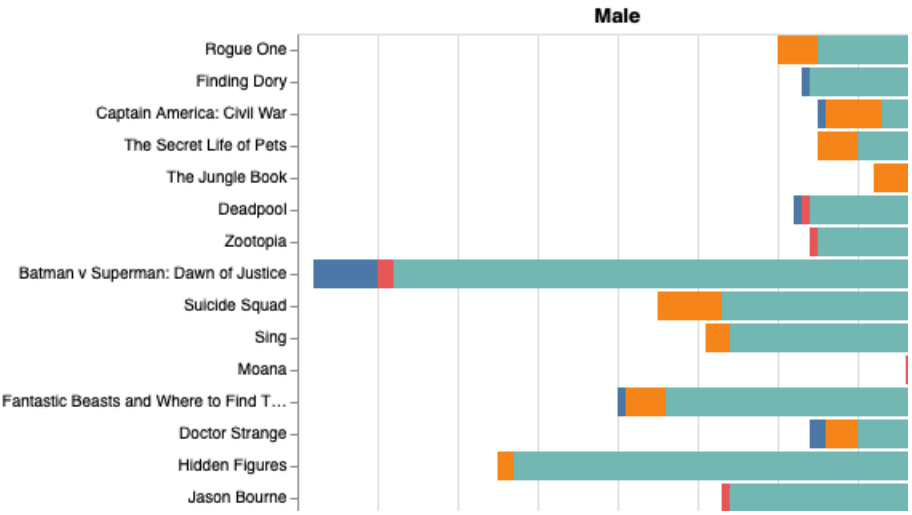
In [7]:
```
cast_gender.head()
```

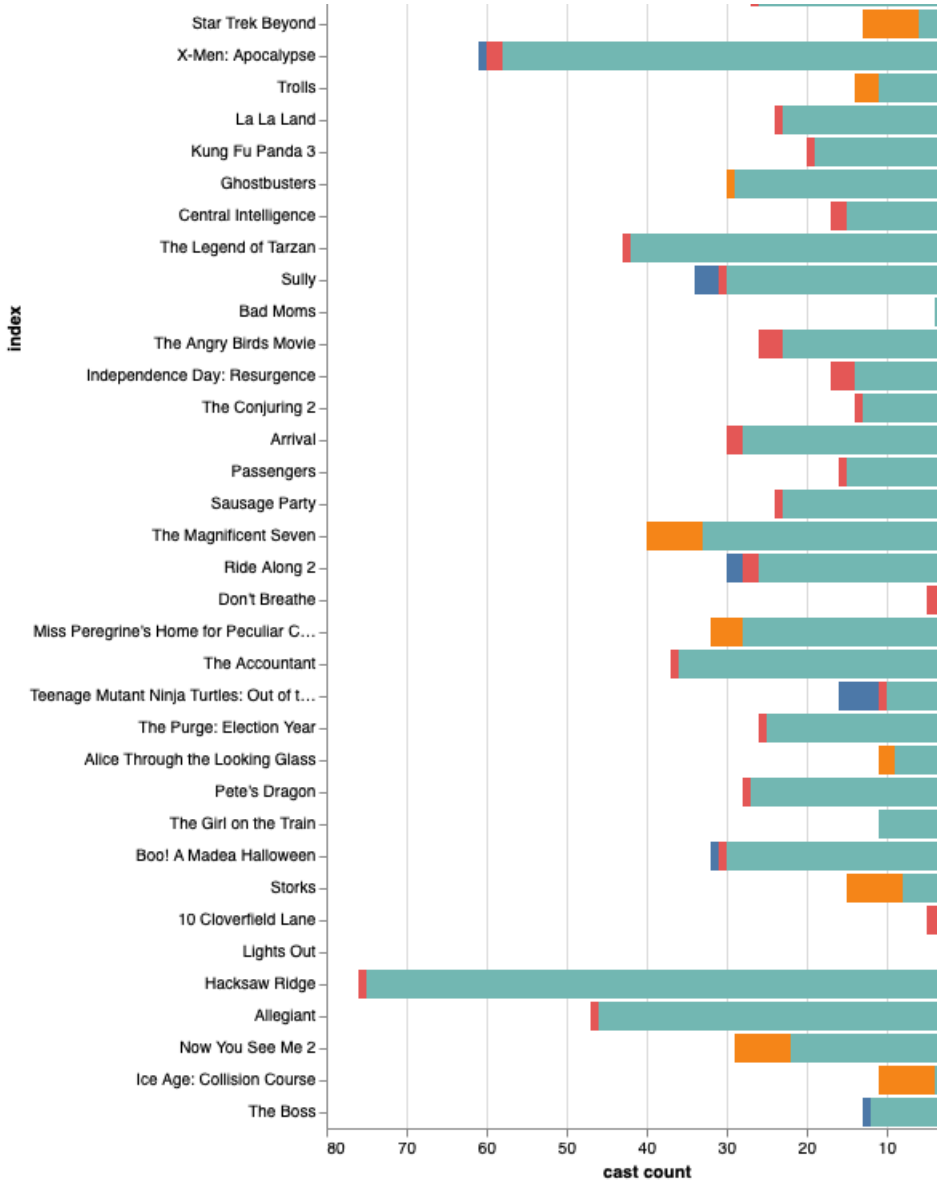
Out[7]:

	MOVIE	ACTOR	CHARACTER_NAME	TYPE	BILLING	GENDER
0	Boo! A Madea Halloween	Tyler Perry	Madea/Joe/Brian	Leading	1	Male
1	Boo! A Madea Halloween	Cassi Davis	Aunt Bam	Supporting	2	Female
2	Boo! A Madea Halloween	Patrice Lovely	Hattie	Supporting	3	Female
3	Boo! A Madea Halloween	Yousef Erakat	Jonathan	Supporting	4	Male
4	Boo! A Madea Halloween	Lexy Panterra	Leah	Supporting	5	Female

In [8]:
```
# Charting Female Actresses
f_encoding = base.encode(
        y= alt.Y(
            'index:N',
            sort= movies_order,
            axis=None
        ),
        x=alt.X('count(index):Q',
                title='cast count',),
        color=alt.Color('TYPE:N')
    )
female = f_encoding.mark_bar().properties(title='Female')

# Charting Male Actors
m_encoding = base.transform_filter(
    alt.datum.GENDER == 'Male'
).encode(
            y= alt.Y(
            'index:N',
            sort= movies_order
        ),
        x=alt.X('count(index):Q',
                sort='descending',
                title='cast count'),
        color=alt.Color('TYPE:N')
    ).mark_bar().properties(title='Male')
male = m_encoding.mark_bar().properties(title='Male')

# Middle Chart
middle = base.encode(
    y=alt.Y('Rank:O', axis=None),
    text=alt.Text('Rank:Q'),
    color=alt.Color('bechdel:N')
).mark_text().properties(width=20)


# Merge together the three charts, male, middle, female
male | middle | female
```

Out[8]:





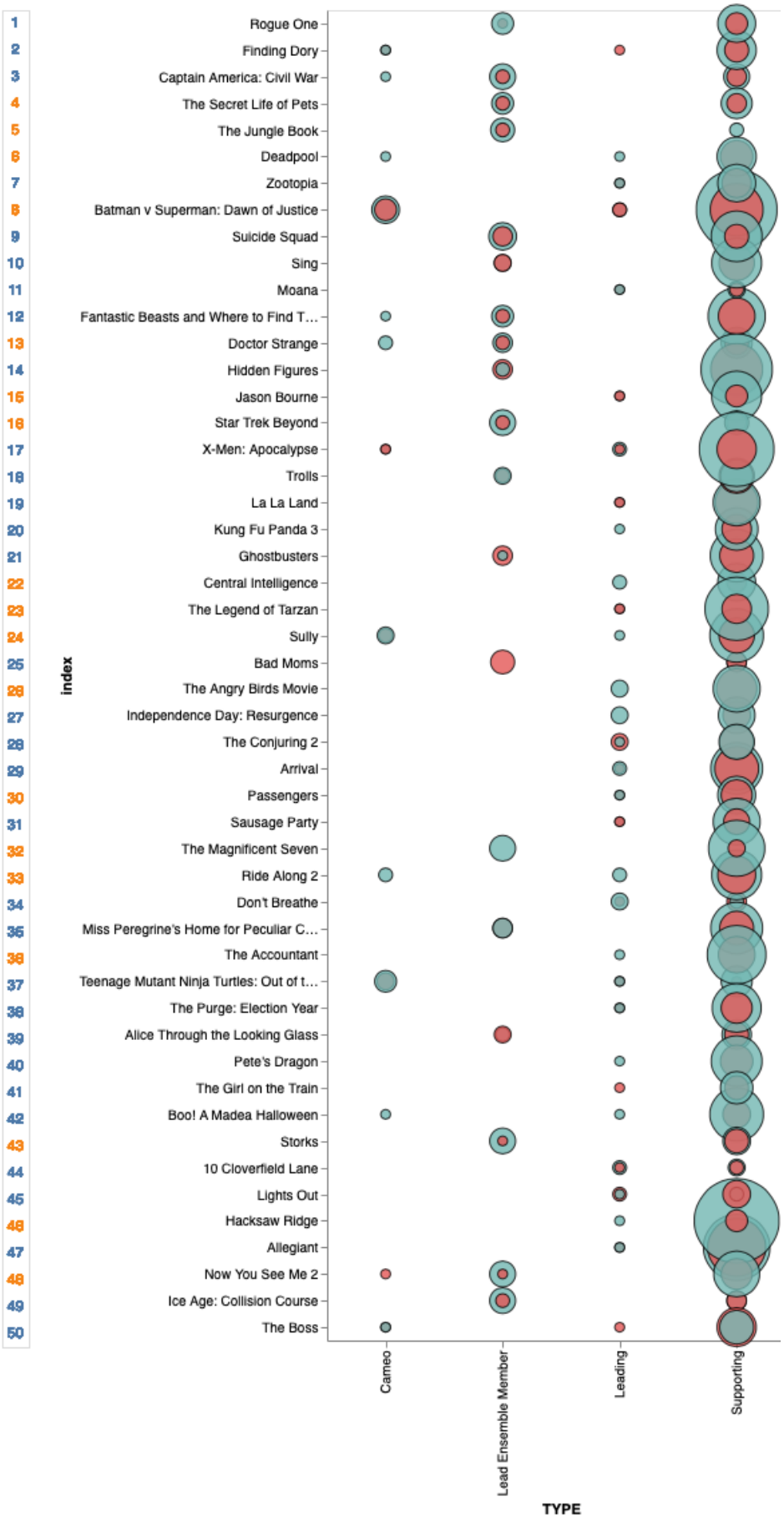
Alternative Encodings

In [9]:
```
def alternative_encoding_one():
    """
    return call to altair function for the new visualization
    """
    plot = base.mark_circle(
        opacity=0.8,
        stroke='black',
        strokeWidth=1
    ).encode(
        alt.Y('index:N',
              sort= movies_order),
        alt.X('TYPE:N'),
        alt.Size('count()',
                 scale=alt.Scale(range=[0,4000]),
                 legend=alt.Legend(symbolFillColor='white')
                ),
        color='GENDER:N'
     ).properties(
         width=350,
         height=995 # Had to adjust from 880 because otherwise "middle
                    # asked about it in Slack by Jakob Cronberg with r
     )

    return plot
```

```
    return plot
```

In [10]:
```
al_enc_one = alternative_encoding_one()
middle | al_enc_one
```

Out[10]:



Typically, this sort of visualization is not preferred due to the fact that humans have a difficult time interpreting area, particularly in circles. The ability to interpret "twice as big" with a circle is much more difficult than with a bar graph, etc.

Exercise adapted and modified from UMSI homework assignment for SIADS 522.