**Problem 4: Analyze Data of your Choice with Tools of your choice**

1. A short description of your dataset.

For this problem set, I have chosen to analyze the revenue fluctuations in the weekend box office by studio over the past 20 years. I will be using the Box Office Mojo site to scrape the data using Python and the Pattern Web API.

1. Your process for obtaining the data

I first spent time reviewing the requirements for the assignment and I wanted to choose a data set that would allow me to practice more complex data scraping from sites using the Pattern Web API. I did a search on the different sites providing with this data and found out Box Office Mojo. I then reviewed the data that would be available for scraping and refining on their site. I evaluated the different variables that I will be employing in the visualization to make sure I will be able to answer to my questions.

1. Refinements made to the dataset

Original data sets can be found at:

<http://www.boxofficemojo.com/alltime/weekends/?pagenum=m50&sort=opengross&p=.htm&order=DESC>

<http://www.boxofficemojo.com/alltime/weekends/?pagenum=m3040&sort=opengross&p=.htm&order=DESC>

<http://www.boxofficemojo.com/alltime/weekends/?pagenum=m2530&sort=opengross&p=.htm&order=DESC>

<http://www.boxofficemojo.com/alltime/weekends/?pagenum=m2025&sort=opengross&p=.htm&order=DESC>

<http://www.boxofficemojo.com/alltime/weekends/?pagenum=m1520&sort=opengross&p=.htm&order=DESC>

<http://www.boxofficemojo.com/alltime/weekends/?pagenum=m1015&sort=opengross&p=.htm&order=DESC>

I wrote a Python script employing the Pattern Web API (script included in my submission) to scrap the data from these sites and create a csv file ( included in my submission). I them used a spreadsheet program to open the whole file and copy the contents into the “Many Eyes”Paste the data section.

1. An explanation of the type of visualization you chose to use
2. Relationships or trends discovered in your visualization
3. Any screenshots relevant to your story