

**Title:** Visualizing Star Wars (or Star Wars Episode \$: Trip to the Data Mine!)

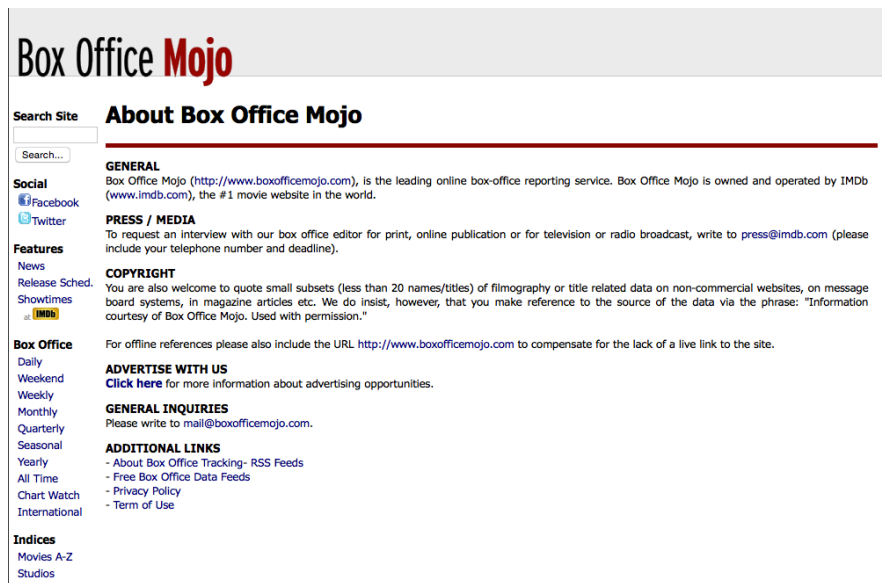
**Topic Covered:** Finding Credible Data, Visualizing Data, Mining Data

**Time Frame:** 45 min – 1 hour

**Description:** The scope of this workshop is to show the thought process that goes into visualizing credible data and give a quick overview of the entire infographic process.

**Session One:**

- Start with reminding students of discussion expectations and outlining the process and lesson objectives.
- Ask if anyone can think of any possible data trends in Star Wars movies over time. Discuss and guide group to the question of if they make more money over time.
- Turn on classroom projector and attached computer. Go to the website: [boxofficemojo.com](http://boxofficemojo.com) . Ask if this can be considered credible? How do you know? Direct them to the “About” page located at the bottom, where we find it is owned by [imdb.com](http://imdb.com) (as of 2016), which is considered by most to be a credible site, at least for public use data.



- Along the left side of the Box Office Mojo (BOM) website is an “Index by Franchise” option. This will lead to many popular movie franchises, but the one in question is Star Wars. Clicking on the franchise page will give a large amount of data, including total gross, opening, and opening date.

- Ask group to see if they can spot any trends in the data. Since The Force Awakens is the highest gross, does that mean it is the most successful? Ask if there is any reasoning as to why these numbers are not necessarily comparable. Lead group to the discussion of inflation and how these numbers are hard to compare because a dollar from 1977 is different than a dollar now. How could we adjust? While it is possible to use an inflation curve and multiply by the appropriate factor, BOM actual includes an adjustor at the top right corner, which will do all the work. Simply set it to the date of the first film (1977) and all numbers will be adjusted to the dollar value of that time.

Adjuster: Actuals

**Franchises**

---

### Star Wars

Total Grosses		Opening Weekends			
Rank	Title (click to view)	Studio	Gross / Theaters	Opening / Theaters	Date
1	<a href="#">Star Wars: The Force Awakens</a>	BV	\$936,390,462	4,134	12/18/15
2	<a href="#">Star Wars: Episode I - The Phantom Menace</a>	Fox	\$431,088,295	3,126	5/19/99
3	<a href="#">Star Wars: Episode III - Revenge of the Sith</a>	Fox	\$380,270,577	3,663	5/19/05
4	<a href="#">Star Wars</a>	Fox	\$307,263,857	1,750	5/25/77
5	<a href="#">Star Wars: Episode II - Attack of the Clones</a>	Fox	\$302,191,252	3,161	5/16/02
6	<a href="#">Return of the Jedi</a>	Fox	\$252,583,617	1,764	5/25/83
7	<a href="#">The Empire Strikes Back</a>	Fox	\$209,398,025	1,278	5/21/80
8	<a href="#">Star Wars (Special Edition) (Re-release)</a>	Fox	\$138,257,865	2,375	1/31/97
9	<a href="#">The Empire Strikes Back (Special Edition) (Re-release)</a>	Fox	\$67,597,694	2,322	2/21/97
10	<a href="#">Return of the Jedi (Special Edition) (Re-release)</a>	Fox	\$45,470,437	2,111	3/14/97
11	<a href="#">Star Wars: Episode I - The Phantom Menace</a>	Fox	\$43,456,382	2,655	2/10/12

#### UPCOMING RELEASES

Title (click to view)	Studio	Release Date
<a href="#">Rogue One: A Star Wars Story</a>	BV	12/16/15
<a href="#">Star Wars: Episode VIII</a>	BV	12/15/15
<a href="#">Untitled Han Solo Star Wars Anthology Film</a>	BV	5/25/16

Adjuster: 1977, \$2.23

### Franchises

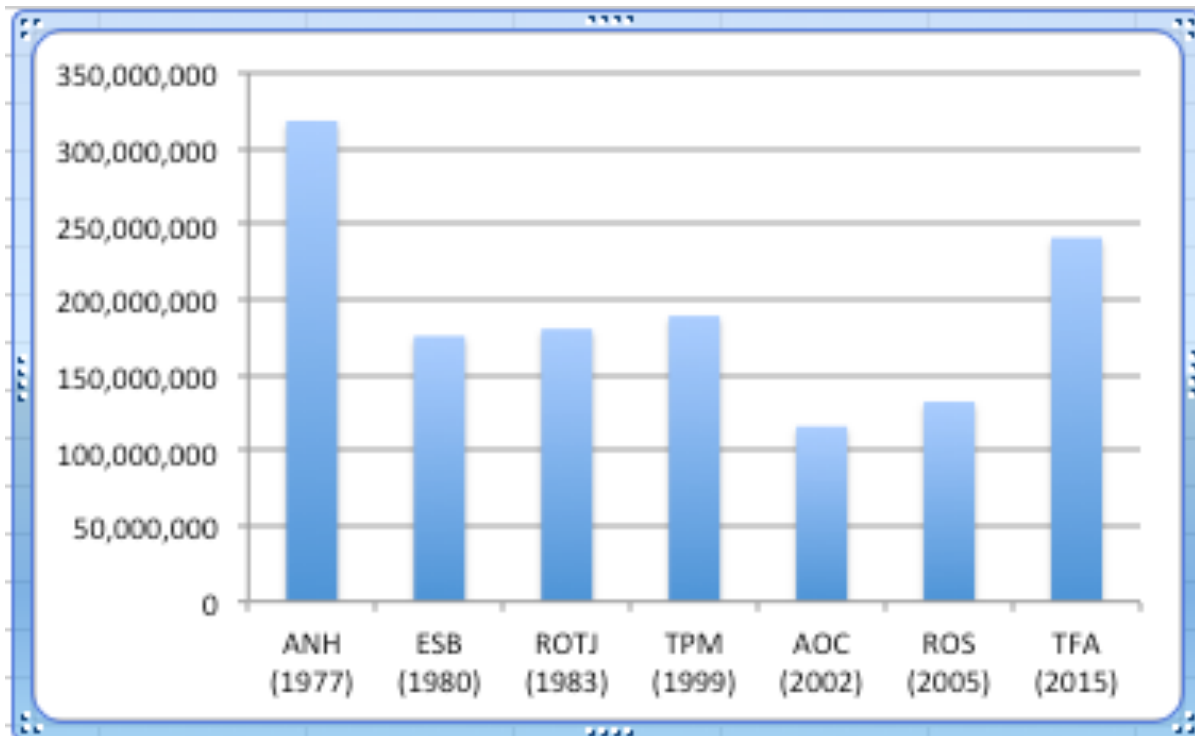
---

## Star Wars

Total Grosses		Opening Weekends		UPCOMING RELEASES		
Rank	Title (click to view)	Studio	Adj. Gross / Theaters	Opening / Theaters	Date	
1	<a href="#">Star Wars</a>	Fox	\$318,296,900 / 1,750	\$1,554,500 / 43	5/25/77	
2	<a href="#">Star Wars: The Force Awakens</a>	BV	\$241,037,000 / 4,134	\$63,559,300 / 4,134	12/18/15	
3	<a href="#">Star Wars: Episode I - The Phantom Menace</a>	Fox	\$189,161,500 / 3,126	\$28,454,900 / 2,970	5/19/99	
4	<a href="#">Return of the Jedi</a>	Fox	\$180,592,900 / 1,764	\$16,296,400 / 1,002	5/25/83	
5	<a href="#">The Empire Strikes Back</a>	Fox	\$175,862,800 / 1,278	\$4,070,800 / 126	5/21/80	
6	<a href="#">Star Wars: Episode III - Revenge of the Sith</a>	Fox	\$132,293,800 / 3,663	\$37,724,200 / 3,661	5/19/05	
7	<a href="#">Star Wars: Episode II - Attack of the Clones</a>	Fox	\$115,987,400 / 3,161	\$30,716,400 / 3,161	5/16/02	
8	<a href="#">Star Wars (Special Edition) (Re-release)</a>	Fox	\$67,171,000 / 2,375	\$17,444,800 / 2,104	1/31/97	
9	<a href="#">The Empire Strikes Back (Special Edition) (Re-release)</a>	Fox	\$32,841,600 / 2,322	\$10,676,800 / 2,111	2/21/97	
10	<a href="#">Return of the Jedi (Special Edition) (Re-release)</a>	Fox	\$22,091,300 / 2,111	\$7,916,000 / 2,111	3/14/97	
11	<a href="#">Star Wars: Episode I - The Phantom Menace (in 3D) (Re-release)</a>	Fox	\$12,234,400 / 2,655	\$6,326,800 / 2,655	2/10/12	
12	<a href="#">Star Wars (Re-issue) (Re-release)</a>	Fox	\$11,738,800 / 1,070	\$2,857,100 / 1,070	8/13/82	
13	<a href="#">Star Wars: The Clone Wars</a>	WB	\$10,920,700 / 3,452	\$4,538,000 / 3,452	8/15/08	
14	<a href="#">The Empire Strikes Back (Re-issue) (Re-release)</a>	Fox	\$10,070,100 / 1,006	\$2,995,700 / 992	11/19/82	

- From this data, we see that the original Star Wars actually made more money (as of April 2016), so there does not seem to be much of a trend over time. But how could we tell? It is often hard to find trends just by looking at numbers. How can we graph this data so as to visualize if there are any trends? Lead discussion toward a simple bar graph, with movies in sequential order at the x-axis and gross sales along the y axis. Talk about which entries to include (guide class towards original releases on not the re-releases). Remind them that making choices about which data to include and which data to exclude shapes the visualization, and that is part of visualizing data, and therefore it is not neutral.
- Launch Excel and adjust windows so that the data is half the screen, and the spreadsheet is on the other half. Have the group call out the appropriate years and numbers for each film. (Note that for the years, the only intention is for them to be in sequential order for a bar graph, there is no need to graph the years). A bar graph should show that there are no clear trends.

ANH (1977)	318,296,900
ESB (1980)	175,862,800
ROTJ (1983)	180,592,900
TPM (1999)	189,161,500
AOC (2002)	115,987,400
ROS (2005)	132,293,800
TFA (2015)	241,037,000



- Observant individuals may spot that the first film in a trilogy tends to do far better, and that the third grosses more than the second, but only slightly. Knowing this trend, how would we expect the next Star Wars movie to do in the box office? What could change this?
- Returning to the 1977-adjusted data, have the class pair up with their tablemate and write down 5 different data characteristics from the website that you DIDN'T explore, including opening day gross, total number of theaters, opening day theaters, re-release data, iMax theater releases, worldwide numbers, ticket prices adjusted for inflation, and so on.)

- Ask each pair to contribute one unexplored characteristic via post it, and then review. Point out that there is data that is not available from this site to tell a more complete story, like the quality of the movie. Ask where we might find some sort of measure of the quality (possibility of rotten tomatoes.) Is that a credible way to examine quality? Why or why not? Ask if anyone can think of more data to tell a complete story, and what type of site might be able to provide it?
- Wrap up the lesson with each pair writing down one question they're left with about Star Wars data on a post-it. Hang the post-its in the classroom.
- The workshop may end here or individuals may work with the data on their own, as in the notes below. They could even upload this data to an online infographic designer such as piktochart, infogr.am, etc. and make a quick visualization on what they found. This should give a brief overview of working with data, visualizing it, and finding stories from data.

**Notes:**

There are plenty of ways this could go much further (originals vs re releases, Number of theaters, amount of competition, etc.) It could go on for at least another hour or even week. The discussion should be curtailed toward the end so there is plenty of time to see the visualization and even revise the graph if time allows. Other data that could be graphed includes critical ratings, presence of characters, merchandising sales, number of explosions (external source required), or even digging into percentage of female characters, amount of CGI, total marketing budgets, etc.