



JMP Instructional data set 009

Disney and Pixar Animated Movies

JMP platforms and statistical methods

Graph Builder: Scatterplot; comparative dot plots
Formula Editor: Creating new variables

Problem statement

Animation has long been a major medium used in movies. Disney's "Snow White and the Seven Dwarfs" (1937) was the first full-length animated feature. This film kicked off what is now known as the Golden Age of Animation, with the creation of classic films such as "Pinocchio" (1940) and "Bambi" (1942), all done from hand-drawn techniques.

The 1980s and 1990s marked another renaissance for animation, driven by Disney's resurgence with hits like "The Little Mermaid" (1989) and "The Lion King" (1994), which combined traditional hand-drawn techniques with compelling storytelling and memorable music.



The advent of CGI (computer-generated imagery) revolutionized the creation of animated movies. Pixar's "Toy Story" (1995) was the first feature-length film created entirely with CGI. This technological leap allowed for greater creativity and realism in animation, leading to an explosion of visually stunning films and new studios.

Initially, animation was primarily seen as a medium for children. This perception has shifted dramatically, as animated films often address complex themes, layered storytelling, emotional depth, and sophisticated humor, allowing them to be appreciated by all ages.

Data on critic and audience ranking, Academy Awards, and financial metrics (budget and box office revenues) have been gathered to make comparisons, explore trends, and identify most/least popular and successful movies.

Data

IDS_009.jmp

Movie	Name of movie
Year	Year film was released (1937, ... , 2025)
Studio	Disney or Pixar
TomatoMeter	Average professional critic score from Rotten Tomatoes
Metacritic	Average professional critic score from Metacritic
Popcornmeter	Average audience score from Rotten Tomatoes
IMDb	Average audience score from IMDb
Academy Award Nomination	Number of Academy Award nominations
Academy Award Wins	Number of Academy Award wins
Budget	Cost to create film
Domestic Box Office Revenue	Revenue generated from movie theatres in U.S. and Canada
Int'l Box Office Revenue	Revenue generated from movie theatres outside of U.S. and Canada
Worldwide Box Office Revenue	Total Revenue generated from movie theatres worldwide
Profit Ratio	Total Revenue divided by Budget
Adjustment to 2025 Dollars	Factor to standardize cost and revenue variables to 2025 dollars

Exercises

The exercises consist of creating a variety of visualization. Some will provide a graphical means to compare the critic and audience scores to assess level of agreement and disagreement, identify most/least popular movies, see how these scores correlate with audience/critic ratings and Academy Award nominations/wins.

Note: You'll notice that not all the movies have complete budget and/or revenue numbers. It's important to keep this in mind when exploring the data as some of the graphs you'll create will only be displaying a subset of all the movies.

1. Create a scatterplot comparing the two critic review scores (Tomatometer vs Metacritic). Write 3-5 bullet point sentences summarizing some of the interesting features in the data revealed by this graph.

Hint: These are examples of the kinds of questions this graph can address.

- Are the critic scores generally in agreement or does one tend to score higher/lower than the other?
- Identify some movies where the scores are the most different (i.e., Tomatometer scores being much higher than or lower than Metacritic scores). Identify movies where the scores are exactly the same or very close. Does this agreement/disagreement tend to occur in just the high, mid, or low scores, or across the full range?
- Identify some of the top ranked movies from Tomatometer. Where do those movies rank with Metacritic? Are there any that are notably different?
- Identify some of the lowest ranked movies from Tomatometer. Where do those movies rank with Metacritic? Are there any that are notably different?

Instructions:

Use Graph > Graph Builder. Place 'Tomatometer' on the Y axis, and 'Metacritic' on the X. Remove the smooth function from the graph by deselecting the icon in the palette.  Add a line representing the exact same score from both sources by: a) right-click on the graph and choose Customize; b) select the '+' symbol to add script; c) from the Templates drop down, select Y Function; d) edit the script to read "Y Function(y=x, x)" and click OK.

2. Create a scatterplot comparing the two audience review scores (Popcornmeter vs IMDb). Write 3-5 bullet point sentences summarizing some of the interesting features in the data revealed by this graph.

Hint: These are examples of the kinds of questions this graph can address.

- Are the audience scores generally in agreement or does one tend to score higher/lower than the other?
- Identify some movies where the scores are the most different (i.e., Popcornmeter scores being much higher than or lower than IMDb scores). Identify movies where the scores are exactly the same or very close. Does this agreement/disagreement tend to occur in just the high, mid, or low scores, or across the full range?
- Identify some of the top ranked movies from Popcornmeter. Where do those movies rank with IMDb? Are there any that are notably different?
- Identify some of the lowest ranked movies from Popcornmeter. Where do those movies rank with IMDb? Are there any that are notably different?

Instructions:

Follow the same steps in Exercise 1 using 'Popcornmeter' on the Y axis and 'IMDb' on the X.

3. Create two new variables that are the average of the two sets of scores ('Avg Critic Score' and 'Avg Audience Score').

Instructions:

Highlight the 'Tomatometer' and 'Metacritic' columns in the data table. Right-click on the highlighted area and select New Formula Column > Combine > Average. Name the column 'Avg Critic Score'. Perform the same operation for the two audience scores. Name the column 'Avg Audience Score'.

4. Create a scatterplot comparing the 'Avg Critic Score' to 'Avg Audience Score'. Write 3-5 bullet point sentences summarizing some of the interesting features in the data revealed by this graph.
Hint: Similar questions outlined in Exercises 1 and 2 can be addressed. Also find a few of your favorite animated movies. Where do they rank in these two scores?

Instructions:

Follow the same steps in Exercises 1 and 2 using 'Avg Critic Score' on the Y axis and 'Avg Audience Score' on the X.

5. Create a graph that can display the 'Average Critic Score' or 'Average Audience Score' over time (from 1937 to 2025). Identify the top 10 critic ranked movies. What years are they from? Identify the top 10 audience ranked movies. What years are they from? Do you have an interpretation of these results? Why might this make comparing the audience scores across all the movies from 1937 to 2025 as a mean of ranking them not quite accurate?

Instructions:

Use Graph > Graph Builder. Place 'Avg Critic Score' on the Y axis and 'Year' on the X. Select Redo > Column Switcher under the red triangle. Select 'Avg Critic Score' as the Initial Column to Switch. Select that and 'Avg Audience Score' for the Replacement Columns.

6. Create a new variable that is the overall average of all four scores.

Instructions:

Highlight the 'Tomatometer', 'Metacritic', 'Popcornmeter', and 'IMDb' columns in the data table. Right-click on the highlighted area and select New Formula Column > Combine > Average. Name the column 'Overall Avg Score'.

7. Which score ('Tomatometer', 'Metacritic', 'Avg Critic Score', 'Popcornmeter', 'IMDb', 'Avg Audience Score', or 'Avg Overall Score' do you think is best to use to rank these movies? Explain why.

8. Create a scatterplot that can display the number of 'Academy Award Nominations' or 'Academy Award Wins' on the Y axis vs the 'Avg Critic Score' or 'Avg Audience Score' on the X axis. Describe the overall association between the number of 'Academy Award Nominations' and the two Scores ('Avg Critic Score' and 'Avg Audience Score'). Are there any movies that have a somewhat surprising results (e.g., at least one nomination despite having a very low Critic and/or Audience Score, or no nominations despite having a very high Critic and/or Audience score)? Perform the same analyses for the number of Academy Award Wins.

Instructions:

Use Graph > Graph Builder. Place 'Academy Award Nominations' on the Y axis and 'Avg Critic Score' on the X. Then choose Redo > Column Switcher under the red triangle. Select 'Academy Award Nominations' as the Initial Column to Switch. Select that and 'Academy Award Wins' for the Replacement Columns. Again choose Redo > Column Switcher under the red triangle. Select 'Avg Critic Score' as the Initial Column to Switch. Select that and 'Avg Audience Score' for the Replacement Columns. Now all four possible scatterplots can be viewed by selecting the desired columns.

9. Create comparative dot plots that display the ratings or award nominations/wins for the two studios from 1995 onwards. Are there any noticeable differences between the Pixar movies vs Disney movies?

Instructions:

Use Graph > Graph Builder. Place 'Tomatometer' on the Y axis and 'Studio' on the X. Select the Line of Fit from the graph palette. 

Choose to display the Means and Std Devs from the Line of Fit controls on the left. Choose Redo > Column Switcher under the red triangle. Select 'Tomatometer' as the Initial Column to Switch. Select that and all the other rankings and award variables for the Replacement Columns.

Choose Local Data Filter under the red triangle. Select 'Year' as the Initial Column to Switch. Select the years 1995 to 2025. The data has now been subset to only include the movies from that time period.

10. Create a graph that can display the Movie Budgets or Box Office Revenues throughout the entire time period. Describe any overall trends observed. Why is it not a good idea to compare financial metrics such as these across the full range of years?

Note: There are numerous missing values in these variables, so the graphs won't be a complete display of all the movies.

Instructions:

Use Graph > Graph Builder. Place 'Budget' on the Y axis, and 'Year' on the X axis. Choose Redo > Column Switcher under the red triangle. Select 'Budget' as the Initial Column to Switch. Select that and the three Box Office Revenue variables for the Replacement Columns.

11. Create a new set of Budget and Revenue variables that are adjusted based upon the value of the dollar in 2024?

Instructions:

Choose Cols > New Columns. Name the variable 'Budget (adj to 2024 \$)'. Select Formula in the Column Properties drop down menu.

Create the following formula which multiplies the Budget values by a factor adjusting to 2024 dollars.

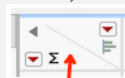
Budget • Adjustment to 2025 Dollars

Repeat this process to create three new variables for the Box Revenues.

12. Recreate the same graph from Exercise 10 using these new adjusted variables. Describe some interesting features/patterns observed in the data. Are there any movies that stand as resulting in very high Box Office Revenues relative to the others? What years were these movies released? What is a possible explanation for these movies performing so exceptionally well?
13. Highlight rows 80 through 87 in the data table. Return to the graph created in Exercise 12. Examine their Box Office Revenue. How do these revenues values compare to movies released before or after these 8?

Instructions:

Go to the data table and with the mouse, highlight rows 80-87. Now return to the graph and you'll see these movies highlighted. To un-highlight them, click on the lower left triangle region in the upper left corner of the data table.



14. Create a graph displaying the 'Profit Ratio' over time. Based upon the available movies, what years are the top 10 most profitable movies from? How do these top 10 movies compare to the others? A 'Profit Ratio' of 1 corresponds to a movie that generated Box Office Revenue equal to what it cost to make. Movie studios often consider a movie successful if the 'Profit Ratio' is at least 2.5. That is, the movie generated more than 2.5 times the Box Office Revenue as it cost to make. Identify the movies that did not achieve this ratio. Examine the 'Profit Ratio' for the movies from 2020 - 2023. Why did these movies perform so poorly regarding this metric? Are there any with a 'Profit Ratio' less than 1? Note: Due to the missing values in the data, only 71 of the 93 movies will be shown.

Instructions:

Use Graph > Graph Builder. Place 'Profit Ratio' on the Y axis, and 'Year' on the X axis. To compare the data to 2.5, double-click on the Y axis to open Axis Settings. Add a Reference Line at 2.5. Float your cursor at the top of the Y axis until the pointer turns into a hand pointing upwards. Click down and drag upward to rescale the Y axis to make it easier to compare each movie to 2.5.

15. Create a graph that can display the 'Overall Avg Rating', 'Total Box Office Revenue' or 'Profit Ratio' vs the 'Budget.' Does it appear that the more a studio spends on making the movie the higher the 'Overall Avg Rating' is? How about the 'Total Box Office Revenue' or 'Profit Ratio'?

Instructions:

Use Graph > Graph Builder. Place 'Budget' on the Y axis, and 'Year' on the X axis. Choose Redo > Column Switcher under the red triangle. Select 'Budget' as the Initial Column to Switch. Select that and the three Box Office Revenue variables for the Replacement Columns.

16. Consider how the ways audiences consume films have changed dramatically in recent decades compared to earlier eras. What important factors are missing from these data that make Box Office Revenue and Profit Ratio imperfect measures for accurately comparing the overall financial performance of films?