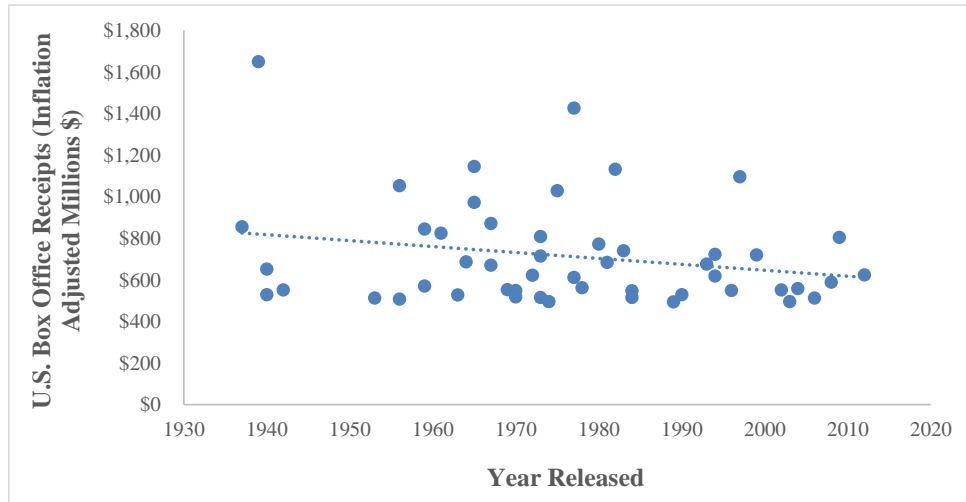


Chapter 3

Data Visualization

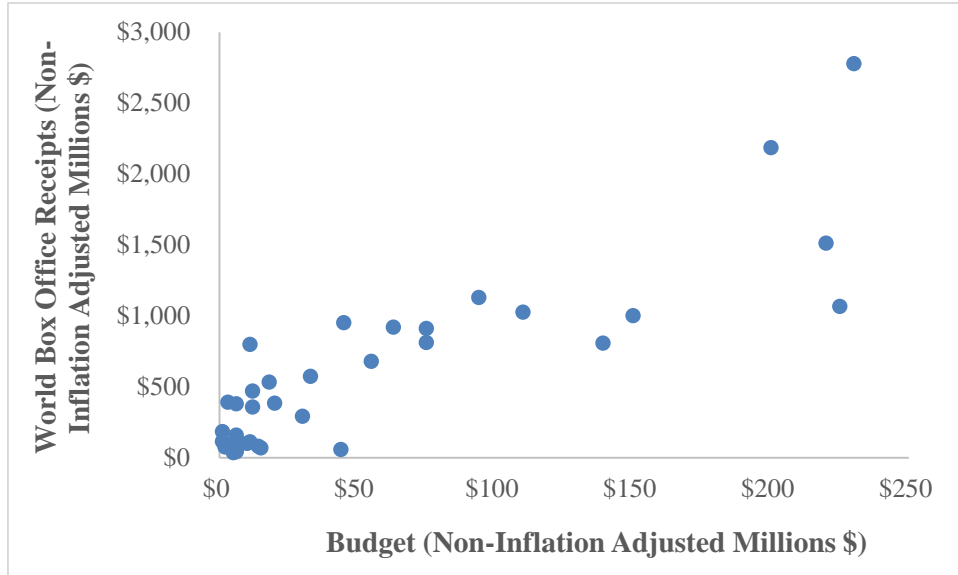
Case Problem: All-Time Movie Box Office Data

1. A scatter chart showing the relationship between year released and inflation-adjusted U.S. box office receipts is shown below.



This scatter chart indicates that box office receipts for the Top 50 movies have been relatively flat to slightly declining over time (adjusted for inflation). The value for *Gone with the Wind* is somewhat of an outlier here, but even if we remove *Gone with the Wind*, the trend is flat to slightly decreasing. Because this represents the receipts for top movies, this seems to indicate that either the total revenues of box office movies are not increasing over time or that there is greater competition for box-office receipts among movies so the top movies are not dominating the total box-office receipts.

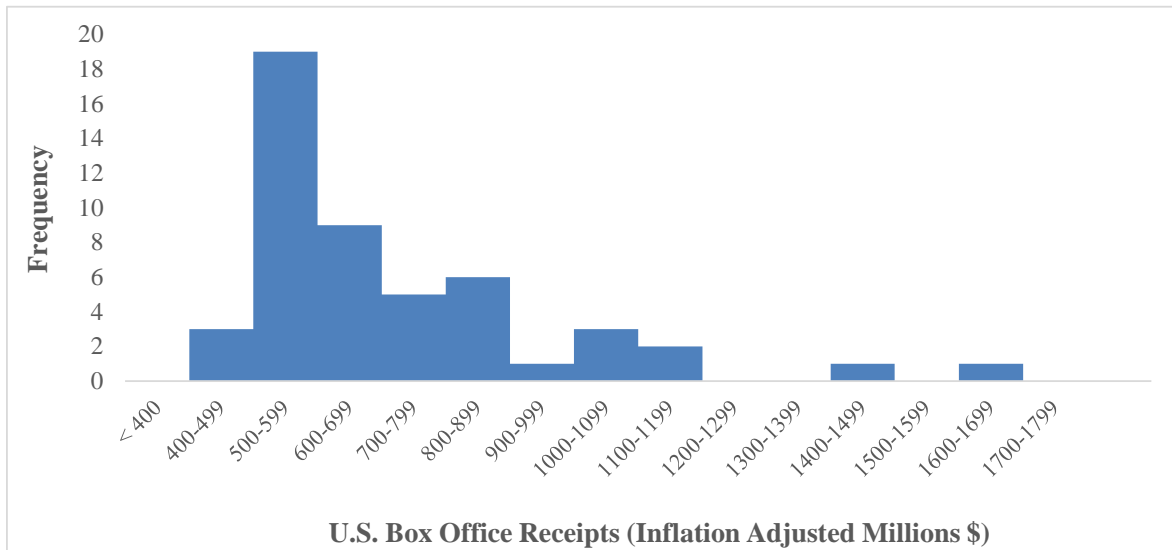
2. A scatter chart showing the relationship between budget and world box office receipts appears below.



This scatter chart seems to clearly show a positive relation between budget and world box office receipts. Top 50 movies with larger budgets appear to be correlated with higher box office receipts. Some of this relation may be driven by the fact that movie studios need to recoup their investment so movies with larger production budgets also receive greater marketing budgets and receive wider releases which lead to greater box office receipts, particularly worldwide.

3.

U.S. Box Office Receipts (Inflation Adjusted Millions \$)	Frequency	Percent Frequency
< 400	0	0%
400-499	3	6%
500-599	19	38%
600-699	9	18%
700-799	5	10%
800-899	6	12%
900-999	1	2%
1000-1099	3	6%
1100-1199	2	4%
1200-1299	0	0%
1300-1399	0	0%
1400-1499	1	2%
1500-1599	0	0%
1600-1699	1	2%
> 1700	0	0%
Total:	50	100%



The distribution of inflation-adjusted U.S. box office receipts is skewed to the right. Most of the Top 50 movies have inflation-adjusted box office receipts that are less than \$700 million dollars. However, there are a few movies that have much greater inflation-adjusted box office receipts. In particular, the movies *Gone With the Wind* (\$1.65 billion) and *Star Wars* (\$1.43 billion) appear to be outliers.

4. A PivotTable used to create the crosstabulation appears below.

Row Labels	PG	PG-13	R	UR	Grand Total
Action	1	2	3		6
Animated	8	1			9
Comedy		2	2		4
Drama	4	5	3	1	15
Horror			1		1
Musical	2				2
SciFi/Fantasy		6	6		12
Western	1				1
Grand Total	15	17	12	4	50

The crosstabulation shows that eight animated movies rated G are in the Top 50 movies. This is more than any other combination of genre and rating for a movie. SciFi/Fantasy movies also appear frequently in the Top 50 movies with six SciFi/Fantasy movies rated PG appearing in the Top 50 list and six rated PG-13.

Filtering the data to show only movies released in 1980 and later produces the PivotTable below.

Row Labels	PG	PG-13	R	Grand Total
Action	1	3	4	
Animated	1	1	2	
Comedy	1	1	2	
Drama		2	2	
SciFi/Fantasy	5	6	11	
Grand Total	1	8	11	21

The filtered PivotTable indicates that after 1980 the most common type of movie appearing in the Top 50 is PG-13 rated SciFi/Fantasy and only one G rated animated movie in the dataset was released after 1980. It appears that top box office receipt generating movies have changed from movies aimed at children to those aimed at older movie goers.

- The PivotTable below displays the average of inflation-adjusted U.S. box office receipts for each movie rating-genre pair.

Row Labels	PG	PG-13	R	UR	Grand Total
Action	506.6	856.4	664.1		701.9
Animated	696.6	557.4			681.2
Comedy		545.0	521.7		533.3
Drama	1016.2	612.3	895.8	622.8	519.9
Horror			808.7		808.7
Musical	915.3				915.3
SciFi/Fantasy		884.1	550.5		717.3
Western		553.0			553.0
Grand Total	798.3	722.3	665.2	618.7	519.9

This PivotTable shows that G-rated dramas generated the highest average inflation-adjusted box office receipts. Overall, musicals generated the highest average inflation-adjusted box office receipts, but this includes only a single data point. Comedies and westerns (also only a single data point) have the lowest average inflation-adjusted box office receipts. G and PG rated movies appear to have higher inflation-adjusted box office receipts in this data set than R rated and UR movies; this makes sense because G and PG movies can be seen by wider audiences.