**Outliers**

As we’ve seen with one-variable data, sometimes a value doesn’t fit the trend (like the number of Michael Jackson’s *Thriller* albums sold). The same can be true when we look for correlation between two variables. The values that don’t seem to fit the pattern are called **outliers.** Statisticians use a calculation to identify outliers, but we will just use our common sense to identify them.

1. If the data has an outlier, what do you think happens to your ability to make predictions from the data?

The following table lists the top movies in theaters on September 14, 2011. The data represents the number of weeks since the movie’s release and the total box office gross income.

|  |  |  |
| --- | --- | --- |
| **Movie** | **Weeks** | **Total Gross** |
| Contagion | 6 | $27,923,641 |
| The Help | 36 | $139,978,511 |
| Warrior | 6 | $6,706,702 |
| The Debt | 15 | $23,188,820 |
| Colombiana | 20 | $30,730,482 |
| Our Idiot Brother | 20 | $22,115,877 |
| Apollo 18 | 13 | $15,583,986 |
| Shark Night 3D | 13 | $15,315,710 |
| Crazy, Stupid, Love | 48 | $78,967,645 |
| Source: [the-numbers.com](http://www.the-numbers.com/charts/daily/2011/20110914.php) | | |

1. Examine the table and state which movie is possibly an outlier?
2. Make a scatter plot (by hand or with a calculator). Does the scatter plot help you determine which movie is possibly an outlier.
3. What might explain this outlier?
4. Do you think this outlier should be included in the data when trying to predict future movie sales?
5. The dashed line is the line of best fit with the outlier included, and the solid line is the line of best fit without the outlier.



Which line would be a better predictor of total gross box office sales for movies in general? Explain.

1. Based on your choice in question 6, predict the total gross sales for a movie 30 weeks after its release.
2. Based on your choice in question 6, predict how many weeks it will take for a film to have total gross sales of $50 million.