**Does the passing away of an actor have influence on the popularity of his works?**

**The dataset**

We’ve used a dataset from <https://snap.stanford.edu/data/web-Amazon.html>, containing the reviews from users from amazon from movies that they ordered through amazon.

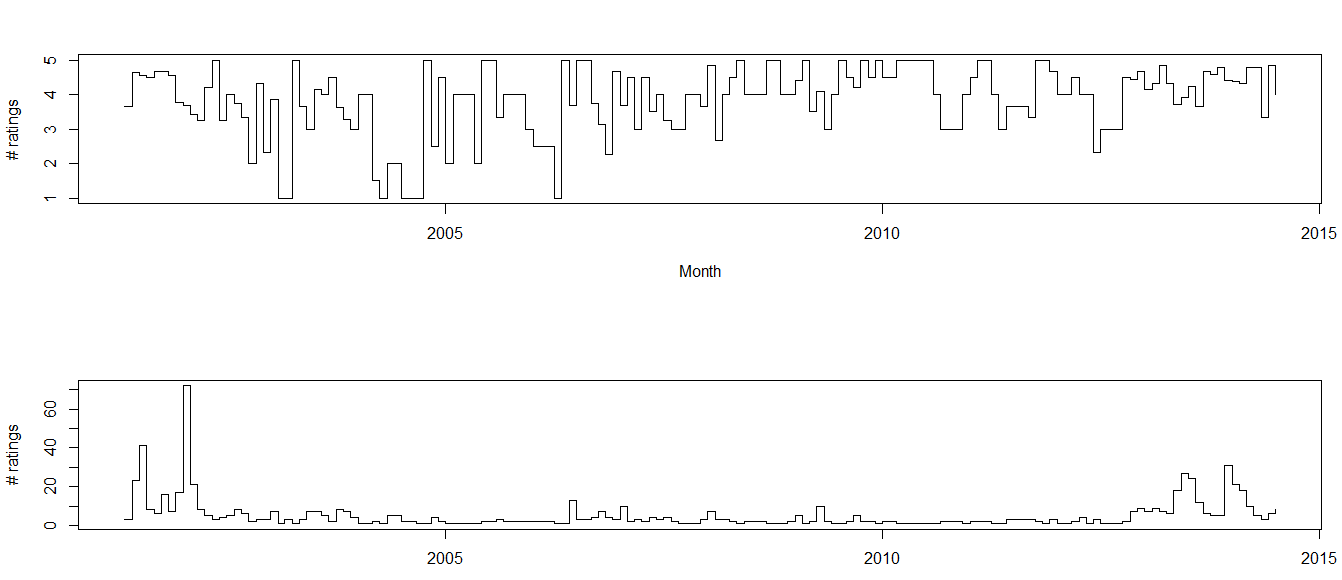
From the data we take the date of the rating, and the rating from 1 to 5.

Through the use of ‘Notepad++’ we were able to put the different movies of the same actor in the same file.

We have taken two or more different movies from three different actors that have passed away.

Through using Rstudio we have made it possible to plot these rating into a graph.

Here is an example of **Paul Walker**, the actor that passed away 30 November 2013:

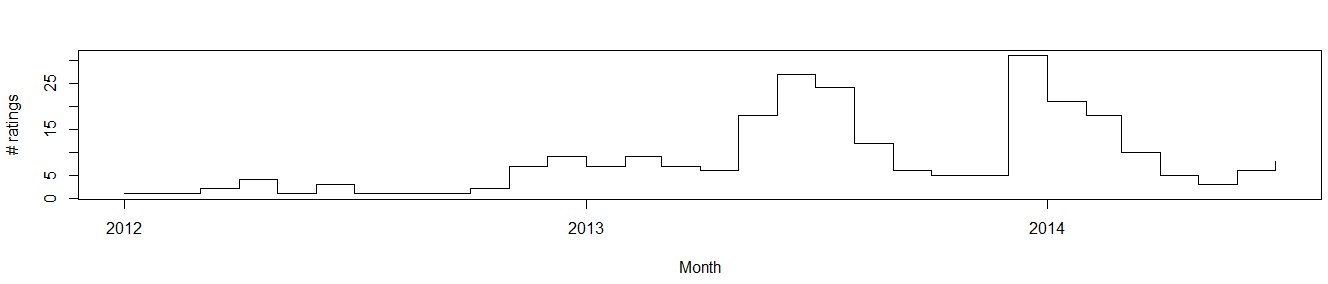


*The ratings of the movies Fast and Furious, 2 Fast 2 Furious, and Fast and Furious Tokyo Drift, all movies with the actor Paul walker.*

The upper graph is the ratings from 1 to 5 from the past years per month, the lower the amount of ratings per month from the past years. (going back to 2001).

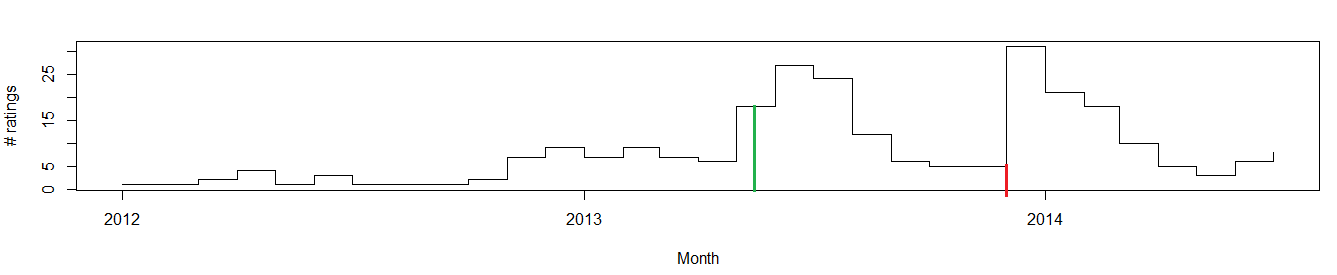
As one can see, there isn’t much too say about the ratings in the upper graph, they don’t really show a lot of consistency. So there isn’t much too say about the **height** of the rating. But there is something to say about the popularity. We see a rise at the beginning and end.

The rise at the end:



You can see a rise in the middle of 2013 and at the end of 2013.

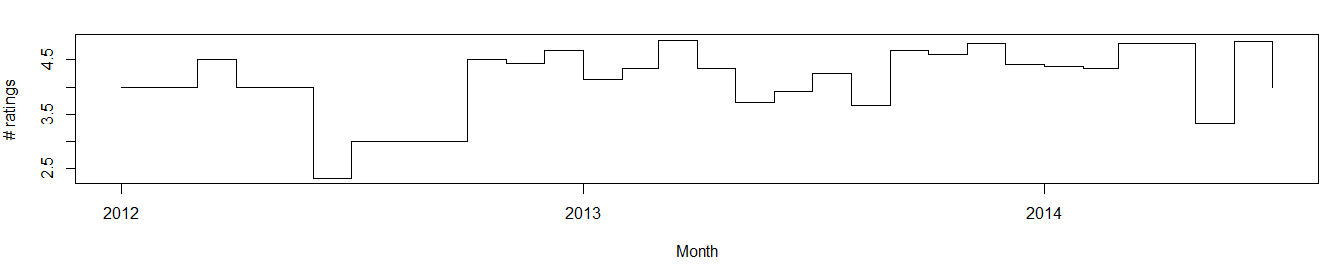
These rises can be explained:



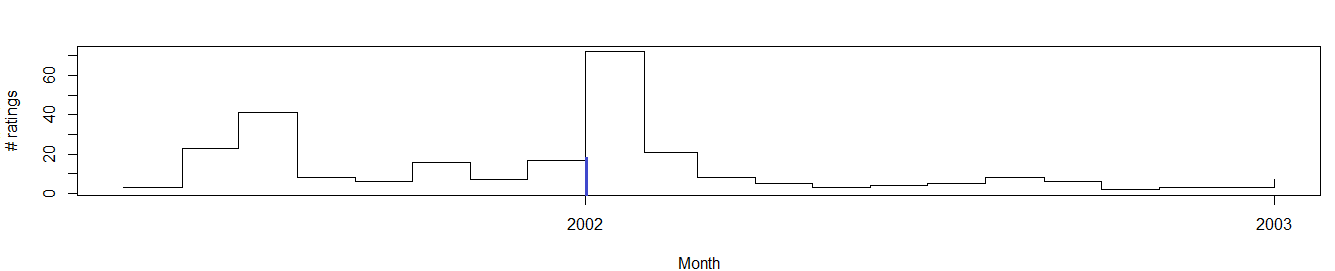
The green line indicates the release of the movie **Furious 6** on the 23nd of May 2013, a follow up movie of the previous movies, also starring Paul walker, casuing a rise in ratings

The red line indicates the passing away of Paul walker in November 2013, causing a rise in ratings of his works.

There is no real rise in the ratings itself:

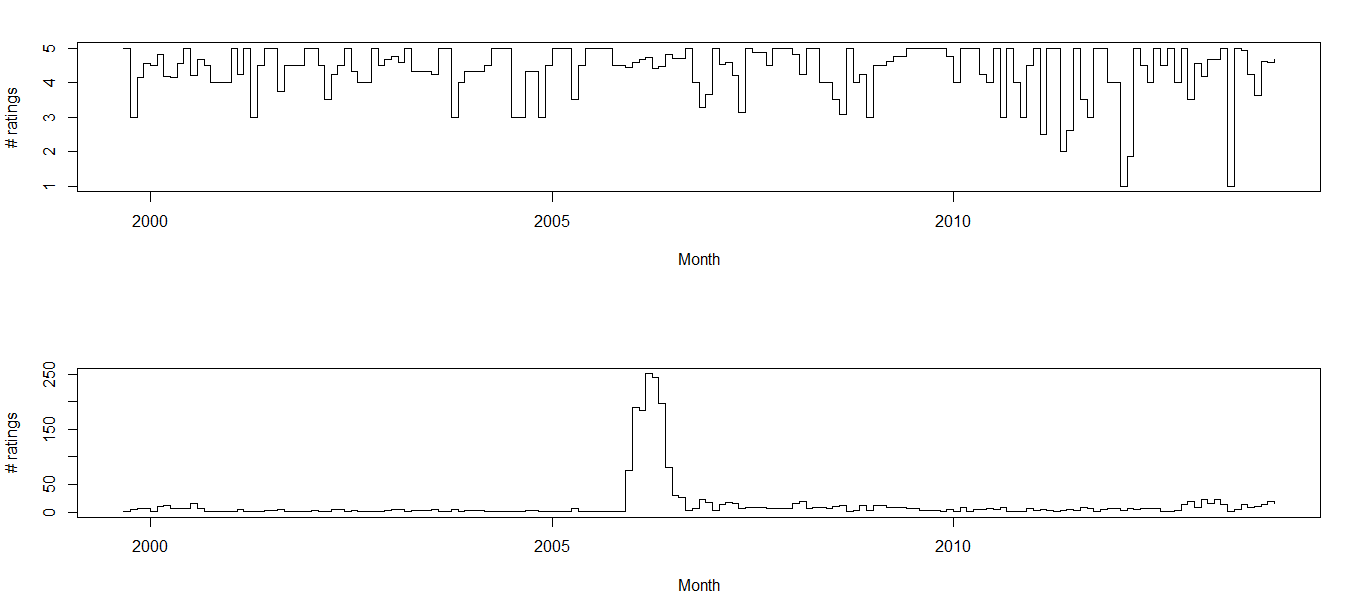


The bump at the beginning can also be explained:



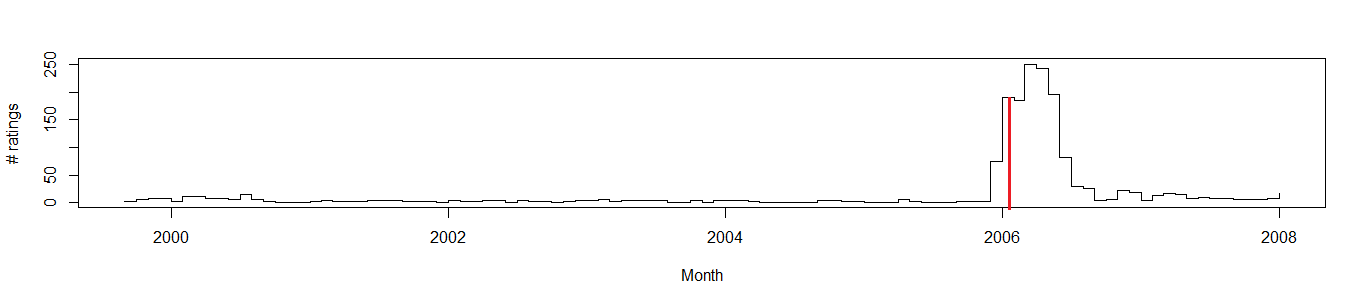
The blue line here indicates the release of the dvd fast and furious, on January the 2nd, causing a rise in popularity.

Here is another example of **Heath Ledger:**



As you can see again no real consistency in the ratings itself, but a clear bump in the amount of ratings.

Which also can be explained:



The red line is where brokeback mountain was released on DVD.

**Conclusion**

We have found out that the factor that influences the data the most is the release data. Which is logical. The death of an actor didn’t significantly change the ratings in all the movies that we used. Mostly in the example above, Fast and Furious. But this might be because of the many parts that this movie has. This would give the death of an actor the reason to watch the older parts. But for the other movies this didn’t really occur.

**Problems with our dataset**

We had too little data to do the research as big as we wanted.

The ratings are for the dvd products not the movie itself.

We should’ve had tested this for movies were no one did in order to compare.