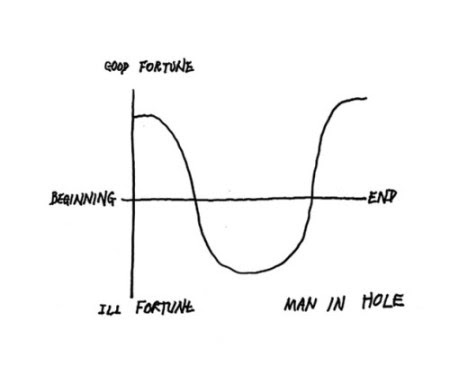
In 1965 the University of Chicago rejected Kurt Vonnegut’s college thesis, which claimed that all stories shared common structures, or “shapes”, including “Man in a Hole”, “Boy gets Girl” and “Cinderella”. Many years later the then already legendary Vonnegut gave a hilarious lecture on this idea – before continuing to read on please watch it here (about 4 minutes):

When you think about it the shape “Man in a Hole” (characters plunge into trouble and crawl out again) really is one of the most popular – even the Bible follows this simple script (see below)!



A colleague of mine, Professor Matthew Jockers from the University of Nebraska, has analyzed 50,000 novels and found out that Vonnegut was really up to something: there are indeed only half a dozen possible plots most novels follow.

You can read more about this project here: [The basic plots of fiction](http://www.dailymail.co.uk/news/article-2969919/The-basic-plots-fiction-Professor-analysed-40-000-novels-claims-just-SIX-possible-storylines.html)

Professor Jockers has written a whole book about this topic: “The Bestseller Code”. But what is even more mind-blowing than this unifying pattern of all stories is that you can do these analyses yourself – with any text of your choice! Professor Jockers made the syuzhet package publicly available on CRAN (“Syuzhet” is the Russian term for narrative construction).

A while ago I finished Dracula, the (grand-)father of all vampire and zombie stories. What a great novel that is! Admittedly it is a little slow-moving but the atmosphere is better than in any of the now popular TV series. Of course, I wanted to do an analysis of the publicly available Dracula text.

The following code should be mostly self-explanatory. First the original text (downloaded from Project Gutenberg: [Bram Stoker: Dracula](http://www.gutenberg.org/cache/epub/345/pg345.txt)) is broken down into separate sentences. After that the sentiment for each sentence is being evaluated and all the values smoothed out (by using some kind of specialized low pass filter). Finally the transformed values are plotted:

library(syuzhet)

dracula <- get\_text\_as\_string("data/pg345.txt")

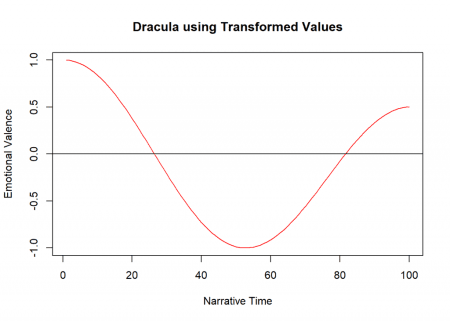
Dracula <- get\_sentences(dracula)

Dracula\_sent <- get\_sentiment(Dracula, method = "bing")

ft\_values <- get\_dct\_transform(Dracula\_sent, low\_pass\_size = 3, scale\_range = TRUE)

plot(ft\_values, type = "l", main = "Dracula using Transformed Values", xlab = "Narrative Time", ylab = "Emotional Valence", col = "red")

abline(h = 0)



In a way R has “read” the novel in no time and extracted the basic plot – pretty impressive, isn’t it! As you can see the story follows the “Man in a Hole”-script rather exemplary, which makes sense because at the beginning everything seems to be fine and well, then Dracula appears and, of course, bites several protagonists, but in the end they catch and kill him – everything is fine again.

**THE END**



…as a bonus, here is the plot that shows that the Bible also follows a simple “Man in a Hole” narrative (paradise, paradise lost, paradise regained). Fortunately, you can conveniently install the King James Bible as a package: <https://github.com/JohnCoene/sacred>

# devtools::install\_github("JohnCoene/sacred")

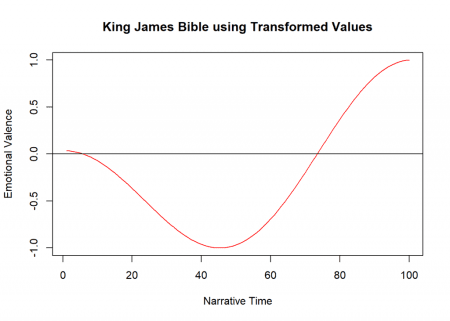
library(sacred)

KJV\_sent <- get\_sentiment(king\_james\_version$text, method = "bing")

ft\_values <- get\_dct\_transform(KJV\_sent, low\_pass\_size = 3, scale\_range = TRUE)

plot(ft\_values, type = "l", main = "King James Bible using Transformed Values", xlab = "Narrative Time", ylab = "Emotional Valence", col = "red")

abline(h = 0)



Simple stories often work best!