# Section 2:Read in the data

The following code assumes that the dataset spooky.csv lives in a data folder (and that we are inside a doc folder).

## Step 1: Using spooky

spooky<-read.csv('../data/spooky.csv',as.is=T)

### An overview of the data structure and content

Let’s first remind ourselves of the structure of the data.

dim<-dim(spooky)  
dim

## [1] 19579 3

head(spooky)

## id  
## 1 id26305  
## 2 id17569  
## 3 id11008  
## 4 id27763  
## 5 id12958  
## 6 id22965  
## text  
## 1 This process, however, afforded me no means of ascertaining the dimensions of my dungeon; as I might make its circuit, and return to the point whence I set out, without being aware of the fact; so perfectly uniform seemed the wall.  
## 2 It never once occurred to me that the fumbling might be a mere mistake.  
## 3 In his left hand was a gold snuff box, from which, as he capered down the hill, cutting all manner of fantastic steps, he took snuff incessantly with an air of the greatest possible self satisfaction.  
## 4 How lovely is spring As we looked from Windsor Terrace on the sixteen fertile counties spread beneath, speckled by happy cottages and wealthier towns, all looked as in former years, heart cheering and fair.  
## 5 Finding nothing else, not even gold, the Superintendent abandoned his attempts; but a perplexed look occasionally steals over his countenance as he sits thinking at his desk.  
## 6 A youth passed in solitude, my best years spent under your gentle and feminine fosterage, has so refined the groundwork of my character that I cannot overcome an intense distaste to the usual brutality exercised on board ship: I have never believed it to be necessary, and when I heard of a mariner equally noted for his kindliness of heart and the respect and obedience paid to him by his crew, I felt myself peculiarly fortunate in being able to secure his services.  
## author  
## 1 EAP  
## 2 HPL  
## 3 EAP  
## 4 MWS  
## 5 HPL  
## 6 MWS

summary(spooky)

## id text author   
## Length:19579 Length:19579 Length:19579   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character

sum(is.na(spooky))

## [1] 0

spooky$author<-as.factor(spooky$author)  
unique(spooky$author)

## [1] EAP HPL MWS  
## Levels: EAP HPL MWS

When we look into spooky data set, it is a 19579 rows and 3 columns dataset. Each row correspoding a unique id number, an excerpt of texts, and author name. Addtionally, there are no missing values. There are three authors, Like HPL is Lovecraft, MWS is Shelly, and EAP is Poe.