## Text Mining the Hitchhiker Trilogy

Greg Johnson

Let's read in Doug's first textbook.

```
#setwd("~/Documents/Analytics/R/DouglasAdams")
book1 <- readLines("data/HitchHiker.txt")</pre>
head(book1,20)
   [1] ""
##
##
    [2] ""
##
   [3] "
                                      Douglas Adams"
   [4] ""
   [5] "
##
                          The Hitch Hiker's Guide to the Galaxy"
       11 11
##
   [6]
##
   [7] "
                                          Book 1"
##
   [8] ""
   [9] ""
##
## [10]
## [11]
## [12]
                                                   for Jonny Brock and Clare Gorst"
## [13]
                                                   and all other
                                                                     Arlingtonians"
## [14]
                                                       tea, sympathy, and a sofa"
## [15] ""
## [16] "
              Far out in the uncharted backwaters of the unfashionable end of the"
## [17] "western spiral arm of the Galaxy lies a small unregarded yellow sun."
              Orbiting this at a distance of roughly ninety-two million miles is an"
## [18] "
## [19] "utterly insignificant little blue green planet whose apedescended life"
## [20] "forms are so amazingly primitive that they still think digital watches are"
```

Our data are in! But they're not tidyin the sense that we want one token per line. For a first look at our data, we will look at words.

3

## [1] 3