

CSCI 6444 : INTRODUCTION TO BIG DATA & ANALYTICS

Class Project 3 – Text Analytics in R

Professor: Stephen Kaisler

Yashwanth Raj Varadharajan

Separating chapters into individual files

We begin our program with installing required packages and importing them into our project library. Then we can set working directory and start working on the text analysis. Some of the main packages that we have used in our program are

tm - This package is used to provide a set of predefined sources which handle a directory, a vector interpreting each component as document, or data frame like structures.

readtext – This package is used to Import and Handle for Plain and Formatted Text Files. This package has a function called `readtext()` which provides an easy way to upload multiple separate files containing plain-text to a single dataframe at once.

wordcloud – This package is used to analyze texts and to quickly visualize the keywords as a word cloud.

quanteda - The package is designed to apply natural language processing to texts, from documents to final analysis.

syuzhet – This package is used for extracting sentiment and sentiment-derived plot arcs from text using a variety of sentiment dictionaries.

```
> #install.packages("tm")
> library(tm)
> #install.packages("readtext")
> library(readtext)
> #install.packages("wordcloud")
> library(wordcloud)
> #install.packages("quanteda")
> library(quanteda)
> #install.packages("syuzhet")
> library(syuzhet)
> #install.packages("text")
> library(text)
> #install.packages("textTinyR")
> library(textTinyR)
> library(dplyr)
> library(base)
> library(stringr)
> |
```

Once done with importing, we then set the working directory and import the file “TarzanOfTheApes.txt”. This dataset has contents of a book called Tarzan of the Apes by Edgar Rice Burroughs. It has a total of 28 Chapters but as per given guidelines, we are going to work only on the first 15 chapters.

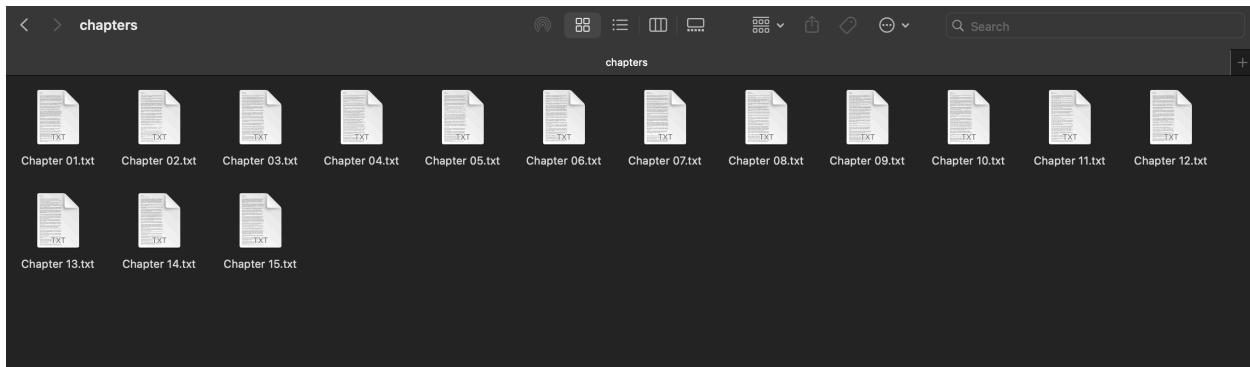
Now we start the separation process. We begin by defining indexes for the chapters. Then we separate the chapters by the logic that contents of a chapter file should start with the “Chapter N”, where N denotes the chapter number, and it ends with the position previous to “Chapter N+1”. Once we separate the file contents using the above logic, then we create a directory and then store all the chapters as separate files.

```
> # Defining all the indexes of the chapters that we are going to use
> # We are using chapters I to XV
> indx_ch1 <- which(book == "Chapter I", arr.ind=TRUE)
> indx_ch2 <- which(book == "Chapter II", arr.ind=TRUE)
> indx_ch3 <- which(book == "Chapter III", arr.ind=TRUE)
> indx_ch4 <- which(book == "Chapter IV", arr.ind=TRUE)
> indx_ch5 <- which(book == "Chapter V", arr.ind=TRUE)
> indx_ch6 <- which(book == "Chapter VI", arr.ind=TRUE)
> indx_ch7 <- which(book == "Chapter VII", arr.ind=TRUE)
> indx_ch8 <- which(book == "Chapter VIII", arr.ind=TRUE)
> indx_ch9 <- which(book == "Chapter IX", arr.ind=TRUE)
> indx_ch10 <- which(book == "Chapter X", arr.ind=TRUE)
> indx_ch11 <- which(book == "Chapter XI", arr.ind=TRUE)
> indx_ch12 <- which(book == "Chapter XII", arr.ind=TRUE)
> indx_ch13 <- which(book == "Chapter XIII", arr.ind=TRUE)
> indx_ch14 <- which(book == "Chapter XIV", arr.ind=TRUE)
> indx_ch15 <- which(book == "Chapter XV", arr.ind=TRUE)
> indx_ch16 <- which(book == "Chapter XVI", arr.ind=TRUE)
> book_chapter1 <- book[(indx_ch1+1:indx_ch2-1)]
> book_chapter2 <- book[(indx_ch2+1:indx_ch3-1)]
> book_chapter3 <- book[(indx_ch3+1:indx_ch4-1)]
> book_chapter4 <- book[(indx_ch4+1:indx_ch5-1)]
> book_chapter5 <- book[(indx_ch5+1:indx_ch6-1)]
> book_chapter6 <- book[(indx_ch6+1:indx_ch7-1)]
> book_chapter7 <- book[(indx_ch7+1:indx_ch8-1)]
> book_chapter8 <- book[(indx_ch8+1:indx_ch9-1)]
> book_chapter9 <- book[(indx_ch9+1:indx_ch10-1)]
> book_chapter10 <- book[(indx_ch10+1:indx_ch11-1)]
> book_chapter11 <- book[(indx_ch11+1:indx_ch12-1)]
> book_chapter12 <- book[(indx_ch12+1:indx_ch13-1)]
> book_chapter13 <- book[(indx_ch13+1:indx_ch14-1)]
> book_chapter14 <- book[(indx_ch14+1:indx_ch15-1)]
> book_chapter15 <- book[(indx_ch15+1:indx_ch16-1)]
> # Creating a directory to save all the chapters as individual files
> dir.create("chapters")
```

```
> write.table(book_chapter1, file = "chapters/Chapter 01.txt", sep = "\t", row.names=FALSE, col.names=FALSE, quote=FALSE)
> write.table(book_chapter2, file = "chapters/Chapter 02.txt", sep = "\t", row.names=FALSE, col.names=FALSE, quote=FALSE)
> write.table(book_chapter3, file = "chapters/Chapter 03.txt", sep = "\t", row.names=FALSE, col.names=FALSE, quote=FALSE)
> write.table(book_chapter4, file = "chapters/Chapter 04.txt", sep = "\t", row.names=FALSE, col.names=FALSE, quote=FALSE)
> write.table(book_chapter5, file = "chapters/Chapter 05.txt", sep = "\t", row.names=FALSE, col.names=FALSE, quote=FALSE)
> write.table(book_chapter6, file = "chapters/Chapter 06.txt", sep = "\t", row.names=FALSE, col.names=FALSE, quote=FALSE)
> write.table(book_chapter7, file = "chapters/Chapter 07.txt", sep = "\t", row.names=FALSE, col.names=FALSE, quote=FALSE)
> write.table(book_chapter8, file = "chapters/Chapter 08.txt", sep = "\t", row.names=FALSE, col.names=FALSE, quote=FALSE)
> write.table(book_chapter9, file = "chapters/Chapter 09.txt", sep = "\t", row.names=FALSE, col.names=FALSE, quote=FALSE)
> write.table(book_chapter10, file = "chapters/Chapter 10.txt", sep = "\t", row.names=FALSE, col.names=FALSE, quote=FALSE)
> write.table(book_chapter11, file = "chapters/Chapter 11.txt", sep = "\t", row.names=FALSE, col.names=FALSE, quote=FALSE)
> write.table(book_chapter12, file = "chapters/Chapter 12.txt", sep = "\t", row.names=FALSE, col.names=FALSE, quote=FALSE)
> write.table(book_chapter13, file = "chapters/Chapter 13.txt", sep = "\t", row.names=FALSE, col.names=FALSE, quote=FALSE)
> write.table(book_chapter14, file = "chapters/Chapter 14.txt", sep = "\t", row.names=FALSE, col.names=FALSE, quote=FALSE)
> write.table(book_chapter15, file = "chapters/Chapter 15.txt", sep = "\t", row.names=FALSE, col.names=FALSE, quote=FALSE)
```

We can infer that after running the above lines, each chapter in the main file has now been separated into 15 different files with each file containing one chapter.

Values	
book	Large character (11031 elements, 1.3 MB)
book_chapter1	chr [1:467] "Chapter I" "" "Out to Sea" "" "" "I had this story from one who had no business to...
book_chapter10	chr [1:3396] "Chapter X" "" "The Fear-Phantom" "" "" "From a lofty perch Tarzan viewed the vill...
book_chapter11	chr [1:3824] "Chapter XI" "" "\"King of the Apes\""" "" "" "It was not yet dark when he reached ...
book_chapter12	chr [1:4158] "Chapter XII" "" "Man's Reason" "" "" "There was one of the tribe of Tarzan who qu...
book_chapter13	Large character (4746 elements, 557.1 kB)
book_chapter14	Large character (5144 elements, 591.8 kB)
book_chapter15	Large character (5344 elements, 604.4 kB)
book_chapter2	chr [1:864] "Chapter II" "" "The Savage Home" "" "" "Nor did they have long to wait, for the ne...
book_chapter3	chr [1:1138] "Chapter III" "" "Life and Death" "" "" "Morning found them but little, if at all ...
book_chapter4	chr [1:1442] "Chapter IV" "" "The Apes" "" "" "In the forest of the table-land a mile back from...
book_chapter5	chr [1:1769] "Chapter V" "" "The White Ape" "" "" "Tenderly Kala nursed her little waif, wonder...
book_chapter6	chr [1:2043] "Chapter VI" "" "Jungle Battles" "" "" "The wanderings of the tribe brought them o...
book_chapter7	chr [1:2508] "Chapter VII" "" "The Light of Knowledge" "" "" "After what seemed an eternity to ...
book_chapter8	chr [1:2740] "Chapter VIII" "" "The Tree-top Hunter" "" "" "The morning after the Dum-Dum the t...
book_chapter9	chr [1:3180] "Chapter IX" "" "Man and Man" "" "" "Tarzan of the Apes lived on in his wild, jung...
filename	"TarzanOfTheApes.txt"



Corpus Analysis

A VCorpus is a collection of Text from of documents. A VCorpus is kept entirely in memory. VCorpus stands for Volatile Corpora. We have separated chapters and saved it as individual files. We will use corpus for text analysis. We will use a function called VCorpus. This will look into the specified location and make a corpus from the contents.

```
> # VCorpus
> TarzanData <- VCorpus(DirSource("/Users/yashwanth/Documents/GWU/Sem 2/Intro to Big Data/Project/Project 3/R Files/Dataset/chapters", ignore.case = TRUE, mode="text"))
> TarzanData
<<VCorpus>>
Metadata: corpus specific: 0, document level (indexed): 0
Content: documents: 15
> # Examining the Corpus
> str(TarzanData)
Classes 'VCorpus', 'Corpus' hidden list of 3
$ content:List of 15
```

Show Attributes		
Name	Type	Value
TarzanData	list [15] (S3: VCorpus, Corpus	List of length 15
Chapter 01.txt	list [2] (S3: PlainTextDocumer	List of length 2
Chapter 02.txt	list [2] (S3: PlainTextDocumer	List of length 2
Chapter 03.txt	list [2] (S3: PlainTextDocumer	List of length 2
Chapter 04.txt	list [2] (S3: PlainTextDocumer	List of length 2
Chapter 05.txt	list [2] (S3: PlainTextDocumer	List of length 2
Chapter 06.txt	list [2] (S3: PlainTextDocumer	List of length 2
Chapter 07.txt	list [2] (S3: PlainTextDocumer	List of length 2
Chapter 08.txt	list [2] (S3: PlainTextDocumer	List of length 2
Chapter 09.txt	list [2] (S3: PlainTextDocumer	List of length 2
Chapter 10.txt	list [2] (S3: PlainTextDocumer	List of length 2
Chapter 11.txt	list [2] (S3: PlainTextDocumer	List of length 2
Chapter 12.txt	list [2] (S3: PlainTextDocumer	List of length 2
Chapter 13.txt	list [2] (S3: PlainTextDocumer	List of length 2
Chapter 14.txt	list [2] (S3: PlainTextDocumer	List of length 2
Chapter 15.txt	list [2] (S3: PlainTextDocumer	List of length 2

As we can see the corpus is being created chapter wise. Then we use the str function to compactly display the internal structure of the corpus. Now we will perform analysis on the corpus. We will extract and examine the text from corpus. Then we will compute the Document Term Matrix. This is a mathematical matrix

that describes the frequency of terms that occur in a collection of documents. Here each row represents one document. each column represents one term.

```
> # Extract text from the corpus
> TarzanDataText <- TarzanData[[1]]
> # Examining the text
> TarzanDataText
<<PlainTextDocument>>
Metadata: 7
Content: chars: 469824
> # Document Term Matrix
> TarzanDataDTM <- DocumentTermMatrix(TarzanData)
> # Examining DTM
> TarzanDataDTM
<<DocumentTermMatrix (documents: 1, terms: 11699)>>
Non-/sparse entries: 11699/0
Sparsity : 0%
Maximal term length: 22
Weighting : term frequency (tf)
> str(TarzanDataDTM)
List of 6
$ i      : int [1:11699] 1 1 1 1 1 1 1 1 1 ...
$ j      : int [1:11699] 1 2 3 4 5 6 7 8 9 10 ...
$ v      : num [1:11699] 1 1 1 1 1 1 1 1 1 ...
$ nrow   : int 1
$ ncol   : int 11699
$ dimnames:List of 2
..$ Docs : chr "TarzanOfTheApes.txt"
..$ Terms: chr [1:11699] "_i_" "'ancient" "'appened" "'arf" ...
- attr(*, "class")= chr [1:2] "DocumentTermMatrix" "simple_triplet_matrix"
- attr(*, "weighting")= chr [1:2] "term frequency" "tf"
```

10 Longest Words in each chapter

Now that we have separated all the chapters file wise, we will find the 10 longest words in each chapter. To achieve this functionality, we are going to build a function. This function will take in corpus of each chapter and then return the 10 longest words in each chapter.

```

> # 10 longest words
> # Function to find the top ten longest words
> TenLongestWordsFun <- function(SampleCorpus) {
+   SampleDTM <- DocumentTermMatrix(SampleCorpus)
+   SampleMatrix <- as.matrix(SampleDTM)
+   SampleTerms <- colnames(SampleMatrix)
+   SampleSplitTerms <- unlist(strsplit(SampleTerms, " "))
+   SampleTermLengths <- nchar(SampleSplitTerms)
+   SampleSortedIndices <- order(SampleTermLengths, decreasing = TRUE)
+   SampleSortedTerms <- SampleSplitTerms[SampleSortedIndices]
+   SampleSortedTerms <- SampleSortedTerms[SampleSortedTerms != ""]
+   TenLongestWordsReturn <- SampleSortedTerms[1:10]
+   return(TenLongestWordsReturn)
+ }

```

The input given to this function is a corpus of each chapter. After receiving the corpus, we convert it into Document Term Matrix and get the terms. Then we split the terms based on spaces and get the lengths of terms and sort the terms by length in descending order. Now after removing empty terms, we will get top ten longest words.

Chapter	10 Longest Words
Chapter 1	feared--england's, thing--cowardice, battlefields--a, responsibility, thinking--there, investigation, man--mentally, ridiculousness, simultaneously, acknowledging
Chapter 2	children--picture, responsibilities, notwithstanding, self-interests, terror-stricken, acquaintances, consciousness, forgetfulness, matter-of-fact, nourishment--o
Chapter 3	playfellows--carefully, children--picture, minutely--strange, crouched--sabor, lioness--lashing, responsibilities, self-confidence, investigations, life--otherwise, notwithstanding
Chapter 4	playfellows--carefully, impossible--learning, one--twice--thrice, minutely--strange, crouched--sabor, e-l-e-p-h-a-n-t, lioness--lashing, self-confidence, investigations, life--otherwise

Chapter 5	playfellows--carefully, impossible--learning, one--twice--thrice, destruction--once, minutely--strange, crouched--sabor, e-l-e-p-h-a-n-t, lioness--lashing, self-confidence, depression--ah
Chapter 6	impossible--learning, one--twice--thrice, destruction--once, incomprehensible, minutely--strange, e-l-e-p-h-a-n-t, depression--ah, destructiveness, experimentation, greystoke--kept
Chapter 7	impossible--learning, one--twice--thrice, destruction--once, incomprehensible., munan-go-keewati,, e-l-e-p-h-a-n-t., personification,, remarkable--most, depression--ah,, destructiveness
Chapter 8	absent-mindedness, sailors--mutineers, destruction--once, incomprehensible, munan-go-keewati, personification, remarkable--most, stopped--beneath, comprehensible, depression--ah
Chapter 9	absent-mindedness, follow-the-leader, sailors--mutineers, silent--motionless, the--ah--gentleman, destruction--once, fifteenth-century, incomprehensible, munan-go-keewati, personification
Chapter 10	absent-mindedness, follow-the-leader, sailors--mutineers, silent--motionless, the--ah--gentleman, fifteenth-century, incomprehensible, munan-go-keewati, personification, remarkable--mos
Chapter 11	absent-mindedness, action--determined, admiration--watched, north-northeasterly, follow-the-leader, sailors--mutineers, silent--motionless, the--ah--gentleman, fifteenth-century, incomprehensible
Chapter 12	absent-mindedness, action--determined, admiration--watched, north-northeasterly, follow-the-leader, sailors--mutineers, silent--motionless, the--ah--gentleman, close-approaching, fifteenth-century
Chapter 13	absent-mindedness, action--determined, admiration--watched, grizzly--absolutely, north-northeasterly, english--evidently, follow-the-leader, sailors--mutineers, silent--motionless, the--ah--gentleman

Chapter 14	action--determined, admiration--watched, grizzly--absolutely, north-northeasterly, civilization--even, conventionalities, english--evidently, follow-the-leader, sailors--mutineers, silent--motionless
Chapter 15	action--determined, admiration--watched, grizzly--absolutely, north-northeasterly, civilization--even, conventionalities, english--evidently, follow-the-leader, self-control--more, silent--motionless

10 Longest Sentences in each chapter

Similarly we find the 10 longest words in each chapter. To achieve this functionality, we are going to build a function. This function will take in corpus of each chapter and then return the 10 longest sentences in each chapter.

```
> # 10 longest sentences
> # Function to find the top ten longest sentences
> TenLongestSentencecsFun <- function(SampleCorpus, ChapterName, LineEnd = "\n") {
+   SamplePlainText <- sapply(SampleCorpus, as.character)
+   SampleText <- paste(SamplePlainText, collapse = LineEnd)
+   SampleLines <- strsplit(SampleText, LineEnd)[[1]]
+   SampleLines <- SampleLines[!grepl(ChapterName, SampleLines, ignore.case = TRUE)]
+   SampleText <- paste(SampleLines, collapse = LineEnd)
+   SampleSentences <- unlist(strsplit(SampleText, "[.!?]\\s*"))
+   SampleSentences <- SampleSentences[SampleSentences != ""]
+   SampleSentencesLengths <- nchar(SampleSentences)
+   SampleSortedIndices <- order(SampleSentencesLengths, decreasing = TRUE)
+   SampleSortedSentences <- SampleSentences[SampleSortedIndices]
+   TenLongestSentencesReturn <- SampleSortedSentences[1:10]
+   return(TenLongestSentencesReturn)
+ }
```

The input given to this function is Corpus, chapter name and line end of each chapter. We first convert the corpus to plain text and combine lines into a single string. Now we split the lines using the specified line end and filter out lines containing chapter names. Then we combine remaining lines into a single string and split text into sentences using regex. Once that is done, we will remove empty sentences and get lengths of sentences. Now once we sort sentences by length in descending order we will get the top ten longest sentences.

Chapters	10 Longest Sentences
Chapter 1	<p>From the records of the Colonial Office and from the dead man's diary\nwe learn that a certain young English nobleman, whom we shall call John\nClayton, Lord Greystoke, was commissioned to make a peculiarly delicate\ninvestigation of conditions in a British West Coast African Colony from\nwhose simple native inhabitants another European power was known to be\nrecruiting soldiers for its native army, which it used solely for the\nforcible collection of rubber and ivory from the savage tribes along\nthe Congo and the Aruwimi Without\nwaiting to rise he whipped a revolver from his pocket, firing point\nblank at the great mountain of muscle towering before him; but, quick\nas he was, John Clayton was almost as quick, so that the bullet which\nwas intended for the sailor's heart lodged in the sailor's leg instead,\nfor Lord Greystoke had struck down the captain's arm as he had seen the\nweapon flash in the sun When my convivial host discovered that he had told me so much, and that\nI was prone to doubtfulness, his foolish pride assumed the task the old\nvintage had commenced, and so he unearthed written evidence in the form\nof musty manuscript, and dry official records of the British Colonial\nOffice to support many of the salient features of his remarkable\nnarrative Clayton asked no questions--he did not need to--and the following day,\nas the great lines of a British battleship grew out of the distant\nhorizon, he half determined to demand that he and Lady Alice be put\naboard her, for his fears were steadily increasing that nothing but\nharm could result from remaining on the lowering, sullen Fuwalda Late in the afternoon he saw her upper works fade\nbelow the far horizon, but not before he learned that which confirmed\nhis greatest fears, and caused him to curse the false pride which had\nrestrained him from seeking safety for his young wife a few short hours\nbefore, when safety was within reach--a safety which was now gone\nforever On the second day after the wounding of Black Michael, Clayton came on\ndeck just in time to see the limp body of one of the crew being carried\nbelow by four of his fellows while the first mate, a heavy belaying pin\nin his hand, stood glowering at the little party of sullen sailors So, whereas the captain might easily have been brought to regret his\nhasty speech had Clayton attempted to conciliate him, his temper was\nnow irrevocably set in the mold in which Clayton had left it, and the\nlast chance of their working together for their common good was gone I do not say the story is true, for I did not witness the happenings\nwhich it portrays, but the fact that in the telling of it to you I have\ntaken fictitious names for the principal characters quite sufficiently\nevidences the sincerity of my own belief that it MAY be true "\n\nThe captain had worked himself up to such a frenzy of rage that he was\nfairly purple of face, and he shrieked the last words at the top of his\nvoice, emphasizing his remarks by a loud thumping of the table with one\nhuge fist, and shaking the other in Clayton's face The Englishmen in Africa went even further, saying that these poor\nblacks were held in virtual slavery, since after their terms of\nenlistment expired their ignorance was imposed upon by their white\nofficers, and they were told that they had yet several years to serve</p>

Chapter 2

The brilliant birds and the little monkeys had become accustomed to\ntheir new acquaintances, and as they had evidently never seen human\nbeings before they presently, after their first fright had worn off,\napproached closer and closer, impelled by that strange curiosity which\ndominates the wild creatures of the forest and the jungle and the\nplain, so that within the first month several of the birds had gone so\nfar as even to accept morsels of food from the friendly hands of the\nClaytons Notwithstanding that both Clayton and his wife kept a sharp lookout\nthey saw nothing of larger animals, though on two occasions they had\nseen their little simian neighbors come screaming and chattering from\nthe near-by ridge, casting frightened glances back over their little\nshoulders, and evincing as plainly as though by speech that they were\nfleeing some terrible thing which lay concealed there Grown careless\nfrom months of continued safety, during which time he had seen no\ndangerous animals during the daylight hours, he had left his rifles and\nrevolvers all within the little cabin, and now that he saw the great\nape crashing through the underbrush directly toward him, and from a\nndirection which practically cut him off from escape, he felt a vague\nlittle shiver play up and down his spine The last entry in his diary was made the morning following her death,\nand there he recites the sad details in a matter-of-fact way that adds\nto the pathos of it; for it breathes a tired apathy born of long sorrow\nand hopelessness, which even this cruel blow could scarcely awake to\nfurther suffering:\n\nMy little son is crying for nourishment--O Alice, Alice, what shall I\ndo To this end Clayton selected four trees which formed a rectangle about\night feet square, and cutting long branches from other trees he\nconstructed a framework around them, about ten feet from the ground,\nfastening the ends of the branches securely to the trees by means of\nrope, a quantity of which Black Michael had furnished him from the hold\nof the Fuwalda High up among the branches of a mighty tree she hugged the shrieking\ninfant to her bosom, and soon the instinct that was as dominant in this\nfierce female as it had been in the breast of his tender and beautiful\nmother--the instinct of mother love--reached out to the tiny\nman-child's half-formed understanding, and he became quiet When the king ape released the limp form which had been John Clayton,\nLord Greystoke, he turned his attention toward the little cradle; but\nKala was there before him, and when he would have grasped the child she\nsnatched it herself, and before he could intercept her she had bolted\nthrough the door and taken refuge in a high tree No signs of habitation were visible, but that the land might easily\nsupport human life was evidenced by the abundant bird and animal life\nof which the watchers on the Fuwalda's deck caught occasional glimpses,\nas well as by the shimmer of a little river which emptied into the\nharbor, insuring fresh water in plenitude Clayton did not believe that Black Michael had the slightest intention\nof notifying the British government of their whereabouts, nor was he\nmany too sure but that some treachery was contemplated for the following\nday when they should be on shore with the sailors who would have to\naccompany them with their belongings During the long hours of darkness they caught but fitful snatches of\nsleep, for the night noises of a great jungle teeming with myriad\nanimal life kept their overwrought nerves on edge, so that a hundred\ntimes they were startled to wakefulness by piercing screams, or the\nstealthy moving of great bodies beneath them

Chapter 3

The brilliant birds and the little monkeys had become accustomed to\ntheir new acquaintances, and as they had evidently never seen human\nbeings before they presently, after their first fright had worn off,\napproached closer and closer, impelled by that strange curiosity which\ndominates the wild creatures of the forest and the jungle and the\nplain, so that within the first month several of the birds had gone so\nfar as even to accept morsels of food from the friendly hands of the\nClaytons He put the book back in the cupboard and closed the door, for he did\nnot wish anyone else to find and destroy his treasure, and as he went\nout into the gathering darkness he closed the great door of the cabin\nbehind him as it had been before he discovered the secret of its lock,\nbut before he left he had noticed the hunting knife lying where he had\nthrown it upon the floor, and this he picked up and took with him to\nshow to his fellows Grown careless\nfrom months of continued safety, during which time he had seen no\ndangerous animals during the daylight hours, he had left his rifles and\nrevolvers all within the little cabin, and now that he saw the great\nape crashing through the underbrush directly toward him, and from a\ndirection which practically cut him off from escape, he felt a vague\nlittle shiver play up and down his spine The last entry in his diary was made the morning following her death,\nand there he recites the sad details in a matter-of-fact way that adds\nto the pathos of it; for it breathes a tired apathy born of long sorrow\nand hopelessness, which even this cruel blow could scarcely awake to\nfurther suffering:\n\nMy little son is crying for nourishment--O Alice, Alice, what shall I\nSo intent was he upon this personal appraisement of his features that\nhe did not hear the parting of the tall grass behind him as a great\nbody pushed itself stealthily through the jungle; nor did his\ncompanion, the ape, hear either, for he was drinking and the noise of\nhis sucking lips and gurgles of satisfaction drowned the quiet approach\nof the intruder "\n\nBut when Kerchak spoke to her about it Kala threatened to run away from\nthe tribe if they did not leave her in peace with the child; and as\nthis is one of the inalienable rights of the jungle folk, if they be\ndissatisfied among their own people, they bothered her no more, for\nKala was a fine clean-limbed young female, and they did not wish to\nlose her A mile back in the forest the tribe had heard the fierce challenge of\nthe gorilla, and, as was his custom when any danger threatened, Kerchak\ncalled his people together, partly for mutual protection against a\ncommon enemy, since this gorilla might be but one of a party of\nseveral, and also to see that all members of the tribe were accounted\nfor It was in the next visit to the vicinity, following the adventure with\nold Sabor, that, as he approached the cabin, Tarzan noticed that from a\ndistance the door appeared to be an independent part of the wall in\nwhich it was set, and for the first time it occurred to him that this\nmight prove the means of entrance which had so long eluded him When Kerchak came to a halt a short distance from the cabin and\ndiscovered that he still held the rifle, he dropped it as he might have\ndropped a red hot iron, nor did he again attempt to recover it--the\nnoise was too much for his brute nerves; but he was now quite convinced\nthat the terrible stick was quite harmless by itself if left alone That her little Tarzan could destroy a great bull gorilla she knew to\nbe improbable, and so, as she neared the spot from which the sounds of\nthe struggle had come, she moved more warily and at last slowly and\nwith extreme caution she traversed the lowest branches, peering eagerly\ninto the moon-splashed blackness for a sign of the combatants

Chapter 4

From this primitive function has arisen, unquestionably, all the forms and ceremonials of modern church and state, for through all the countless ages, back beyond the uttermost ramparts of a dawning humanity our fierce, hairy forebears danced out the rites of the Dum-Dum to the sound of their earthen drums, beneath the bright light of a tropical moon in the depth of a mighty jungle which stands unchanged today as it stood on that long forgotten night in the dim, unthinkable vistas of the long dead past when our first shaggy ancestor swung from a swaying bough and dropped lightly upon the soft turf of the first meeting place Many moons ago, when he had been much smaller, he had desired the skin of Sabor, the lioness, or Numa, the lion, or Sheeta, the leopard to uncover his hairless body that he might no longer resemble hideous Histah, the snake; but now he was proud of his sleek skin for it betokened his descent from a mighty race, and the conflicting desires to go naked in prideful proof of his ancestry, or to conform to the customs of his own kind and wear hideous and uncomfortable apparel found first one and then the other in the ascendancy Squatting upon his haunches on the table top in the cabin his father had built--his smooth, brown, naked little body bent over the book which rested in his strong slender hands, and his great shock of long, black hair falling about his well-shaped head and bright, intelligent eyes--Tarzan of the apes, little primitive man, presented a picture filled, at once, with pathos and with promise--an allegorical figure of the primordial groping through the black night of ignorance toward the light of learning He put the book back in the cupboard and closed the door, for he did not wish anyone else to find and destroy his treasure, and as he went out into the gathering darkness he closed the great door of the cabin behind him as it had been before he discovered the secret of its lock, but before he left he had noticed the hunting knife lying where he had thrown it upon the floor, and this he picked up and took with him to show to his fellows He did not accomplish it in a day, or in a week, or in a month, or in a year; but slowly, very slowly, he learned after he had grasped the impossibilities which lay in those little bugs, so that by the time he was fifteen he knew the various combinations of letters which stood for every pictured figure in the little primer and in one or two of the picture books Tublat was his most consistent enemy, but it was through Tublat that, when he was about thirteen, the persecution of his enemies suddenly ceased and he was left severely alone, except on the occasions when one of them ran amuck in the throes of one of those strange, wild fits of insane rage which attacks the males of many of the fiercer animals of the jungle So intent was he upon this personal appraisement of his features that he did not hear the parting of the tall grass behind him as a great body pushed itself stealthily through the jungle; nor did his companion, the ape, hear either, for he was drinking and the noise of his sucking lips and gurgles of satisfaction drowned the quiet approach of the intruder "But when Kerchak spoke to her about it Kala threatened to run away from the tribe if they did not leave her in peace with the child; and as this is one of the inalienable rights of the jungle folk, if they be dissatisfied among their own people, they bothered her no more, for Kala was a fine clean-limbed young female, and they did not wish to lose her A mile back in the forest the tribe had heard the fierce challenge of the gorilla, and, as was his custom when any danger threatened, Kerchak called his people together, partly for mutual protection against a common enemy, since this gorilla might be but one of a party of several, and also to see that all members of the tribe were accounted for It was in the next visit to the vicinity, following the adventure with old Sabor, that, as he approached the cabin, Tarzan noticed that from a distance the door appeared to be an independent part of the wall in which it was set, and for the first time it occurred to him that this might prove the means of entrance which had so long eluded him

Chapter 5

From this primitive function has arisen, unquestionably, all the forms and ceremonials of modern church and state, for through all the countless ages, back beyond the uttermost ramparts of a dawning humanity our fierce, hairy forebears danced out the rites of the Dum-Dum to the sound of their earthen drums, beneath the bright light of a tropical moon in the depth of a mighty jungle which stands unchanged today as it stood on that long forgotten night in the dim, unthinkable vistas of the long dead past when our first shaggy ancestor swung from a swaying bough and dropped lightly upon the soft turf of the first meeting place Many moons ago, when he had been much smaller, he had desired the skin of Sabor, the lioness, or Numa, the lion, or Sheeta, the leopard to uncover his hairless body that he might no longer resemble hideous Histah, the snake; but now he was proud of his sleek skin for it betokened his descent from a mighty race, and the conflicting desires to go naked in prideful proof of his ancestry, or to conform to the customs of his own kind and wear hideous and uncomfortable apparel found first one and then the other in the ascendancy Squatting upon his haunches on the table top in the cabin his father had built--his smooth, brown, naked little body bent over the book which rested in his strong slender hands, and his great shock of long, black hair falling about his well-shaped head and bright, intelligent eyes--Tarzan of the apes, little primitive man, presented a picture filled, at once, with pathos and with promise--an allegorical figure of the primordial groping through the black night of ignorance toward the light of learning And then Lord Greystoke wiped his greasy fingers upon his naked thighs and took up the trail of Kulonga, the son of Mbonga, the king; while in far-off London another Lord Greystoke, the younger brother of the real Lord Greystoke's father, sent back his chops to the club's CHEF because they were underdone, and when he had finished his repast he dipped his finger-ends into a silver bowl of scented water and dried them upon a piece of snowy damask He put the book back in the cupboard and closed the door, for he did not wish anyone else to find and destroy his treasure, and as he went out into the gathering darkness he closed the great door of the cabin behind him as it had been before he discovered the secret of its lock, but before he left he had noticed the hunting knife lying where he had thrown it upon the floor, and this he picked up and took with him to show to his fellows That they more greatly feared an attack from the rear than whatever unknown enemies lurked in their advance was evidenced by the formation of the column; and such was the fact, for they were fleeing from the white man's soldiers who had so harassed them for rubber and ivory that they had turned upon their conquerors one day and massacred a white officer and a small detachment of his black troops He did not accomplish it in a day, or in a week, or in a month, or in a year; but slowly, very slowly, he learned after he had grasped the impossibilities which lay in those little bugs, so that by the time he was fifteen he knew the various combinations of letters which stood for every pictured figure in the little primer and in one or two of the picture books Tublat was his most consistent enemy, but it was through Tublat that, when he was about thirteen, the persecution of his enemies suddenly fits So intent was he upon this personal appraisement of his features that he did not hear the parting of the tall grass behind him as a great body pushed itself stealthily through the jungle; nor did his companion, the ape, hear either, for he was drinking and the noise of his sucking lips and gurgles of satisfaction drowned the quiet approach of the intruder "
But when Kerchak spoke to her about it Kala threatened to run away from the tribe if they did not leave her in peace with the child; and as this is one of the inalienable rights of the jungle folk, if they be dissatisfied among their own people, they bothered her no more, for Kala was a fine clean-limbed young female, and they did not wish to close her

Chapter 6

From this primitive function has arisen, unquestionably, all the forms and ceremonials of modern church and state, for through all the countless ages, back beyond the uttermost ramparts of a dawning humanity our fierce, hairy forebears danced out the rites of the Dum-Dum to the sound of their earthen drums, beneath the bright light of a tropical moon in the depth of a mighty jungle which stands unchanged today as it stood on that long forgotten night in the dim, unthinkable vistas of the long dead past when our first shaggy ancestor swung from a swaying bough and dropped lightly upon the soft turf of the first meeting place Many moons ago, when he had been much smaller, he had desired the skin of Sabor, the lioness, or Numa, the lion, or Sheeta, the leopard to uncover his hairless body that he might no longer resemble hideous Histah, the snake; but now he was proud of his sleek skin for it betokened his descent from a mighty race, and the conflicting desires to go naked in prideful proof of his ancestry, or to conform to the customs of his own kind and wear hideous and uncomfortable apparel found first one and then the other in the ascendancy Squatting upon his haunches on the table top in the cabin his father had built--his smooth, brown, naked little body bent over the book which rested in his strong slender hands, and his great shock of long, black hair falling about his well-shaped head and bright, intelligent eyes--Tarzan of the apes, little primitive man, presented a picture filled, at once, with pathos and with promise--an allegorical figure of the primordial groping through the black night of ignorance toward the light of learning And then Lord Greystoke wiped his greasy fingers upon his naked thighs and took up the trail of Kulonga, the son of Mbonga, the king; while in far-off London another Lord Greystoke, the younger brother of the real Lord Greystoke's father, sent back his chops to the club's CHEF because they were underdone, and when he had finished his repast he dipped his finger-ends into a silver bowl of scented water and dried them upon a piece of snowy damask He put the book back in the cupboard and closed the door, for he did not wish anyone else to find and destroy his treasure, and as he went out into the gathering darkness he closed the great door of the cabin behind him as it had been before he discovered the secret of its lock, but before he left he had noticed the hunting knife lying where he had thrown it upon the floor, and this he picked up and took with him to show to his fellows The finding of the still warm body of Kulonga--on the very verge of their fields and within easy earshot of the village--knifed and stripped at the door of his father's home, was in itself insufficiently mysterious, but these last awesome discoveries within the village, within the dead Kulonga's own hut, filled their hearts with dismay, and conjured in their poor brains only the most frightful of superstitious explanations That they more greatly feared an attack from the rear than whatever unknown enemies lurked in their advance was evidenced by the formation of the column; and such was the fact, for they were fleeing from the white man's soldiers who had so harassed them for a week, or in a month, or in a year; but slowly, very slowly, he learned after he had grasped the impossibilities which lay in those little bugs, so that by the time he was fifteen he knew the various combinations of letters which stood for every pictured figure in the little primer and in one or two of the picture books Tublat was his most consistent enemy, but it was through Tublat that, when he was about thirteen, the persecution of his enemies suddenly ceased and he was left severely alone, except on the occasions when one of them ran amuck in the throes of one of those strange, wild fits of insane rage which attacks the males of many of the fiercer animals of the jungle A mile back in the forest the tribe had heard the fierce challenge of the gorilla, and, as was his custom when any danger threatened, Kerchak called his people together, partly for mutual protection against a common enemy, since this gorilla might be but one of a party of several, and also to see that all members of the tribe were accounted for

Chapter 7

From this primitive function has arisen, unquestionably, all the forms and ceremonials of modern church and state, for through all the countless ages, back beyond the uttermost ramparts of a dawning humanity our fierce, hairy forebears danced out the rites of the Dum-Dum to the sound of their earthen drums, beneath the bright light of a tropical moon in the depth of a mighty jungle which stands unchanged today as it stood on that long forgotten night in the dim, unthinkable vistas of the long dead past when our first shaggy ancestor swung from a swaying bough and dropped lightly upon the soft turf of the first meeting place Many moons ago, when he had been much smaller, he had desired the skin of Sabor, the lioness, or Numa, the lion, or Sheeta, the leopard to uncover his hairless body that he might no longer resemble hideous Histah, the snake; but now he was proud of his sleek skin for it betokened his descent from a mighty race, and the conflicting desires to go naked in prideful proof of his ancestry, or to conform to the customs of his own kind and wear hideous and uncomfortable apparel found first one and then the other in the ascendancy Squatting upon his haunches on the table top in the cabin his father had built--his smooth, brown, naked little body bent over the book which rested in his strong slender hands, and his great shock of long, black hair falling about his well-shaped head and bright, intelligent eyes--Tarzan of the apes, little primitive man, presented a picture filled, at once, with pathos and with promise--an allegorical figure of the primordial groping through the black night of ignorance toward the light of learning Chapter XII
Man's Reason
There was one of the tribe of Tarzan who questioned his authority, and that was Terkoz, the son of Tublat, but he so feared the keen knife and the deadly arrows of his new lord that he confined the manifestation of his objections to petty disobediences and by some sudden stroke of treachery, and so he was ever on his guard against surprise And then Lord Greystoke wiped his greasy fingers upon his naked thighs and took up the trail of Kulonga, the son of Mbonga, the king; while in far-off London another Lord Greystoke, the younger brother of the real Lord Greystoke's father, sent back his chops to the club's CHEF because they were underdone, and when he had finished his repast he dipped his finger-ends into a silver bowl of scented water and dried them upon a piece of snowy damask The finding of the still warm body of last awesome discoveries within the village, within the dead Kulonga's own hut, filled their hearts with dismay, and conjured in their poor brains only the most frightful of superstitious explanations But when, finally, he realized that his antagonist was fastened to him where his teeth and fists alike were useless against him, Terkoz hurled himself about upon the ground so violently that Tarzan could but cling desperately to the leaping, turning, twisting body, and ere he had struck a blow the knife was hurled from his hand by a heavy impact against the earth, and Tarzan found himself defenseless That they more greatly feared an attack from the rear than whatever unknown enemies lurked in their advance was evidenced by the formation of the column; and such was the fact, for they were fleeing from the white man's soldiers who had so harassed them for rubber and ivory that they had turned upon their conquerors one day and massacred a white officer and a small detachment of his black troops But when one day a member of the tribe discovered the blacks in great numbers on the banks of a little stream that had been their watering place for generations, and in the act of clearing a space in the jungle and erecting many huts, the apes would remain no longer; and so Tarzan led them inland for many marches to a spot as yet undefiled by the foot of a human being He did not accomplish it in a day, or in a week, or in a month, or in a year; but slowly, very slowly, he learned after he had grasped the impossibilities which lay in those little bugs, so that by the time he was fifteen he knew the various combinations of letters which stood for every pictured figure in the little primer and in one or two of the picture books

Chapter 8

Many moons ago, when he had been much smaller, he had desired the skin\nof Sabor, the lioness, or Numa, the lion, or Sheeta, the leopard to\ncover his hairless body that he might no longer resemble hideous\nHistah, the snake; but now he was proud of his sleek skin for it\nbetokened his descent from a mighty race, and the conflicting desires\ninto go naked in prideful proof of his ancestry, or to conform to the\ncustoms of his own kind and wear hideous and uncomfortable apparel\nfound first one and then the other in the ascendency Chapter XII\n\nMan's Reason\n\nThere was one of the tribe of Tarzan who questioned his authority, and\nthat was Terkoz, the son of Tublat, but he so feared the keen knife and\nthe deadly arrows of his new lord that he confined the manifestation of\nhis objections to petty disobediences and irritating mannerisms; Tarzan\nknew, however, that he but waited his opportunity to wrest the kingship\nfrom him by some sudden stroke of treachery, and so he was ever on his\nnguard against surprise And then Lord Greystoke wiped his greasy fingers upon his naked thighs\nand took up the trail of Kulonga, the son of Mbonga, the king; while in\nfar-off London another Lord Greystoke, the younger brother of the real\nLord Greystoke's father, sent back his chops to the club's CHEF because\nthey were underdone, and when he had finished his repast he dipped his\nfinger-ends into a silver bowl of scented water and dried them upon a\npiece of snowy damask The finding of the still warm body of Kulonga--on the very\nverge of their fields and within easy earshot of the village--knifed\nand stripped at the door of his father's home, was in itself\ninsufficiently mysterious, but these last awesome discoveries within the\nvillage, within the dead Kulonga's own hut, filled their hearts with\nndismay, and conjured in their poor brains only the most frightful of\nnsuperstitious explanations But when, finally, he realized that his antagonist was fastened to him\nwhere his teeth and fists alike were useless against him, Terkoz hurled\nhimself about upon the ground so violently that Tarzan could but cling\ndesperately to the leaping, turning, twisting body, and ere he had\nstruck a blow the knife was hurled from his hand by a heavy impact\nagainst the earth, and Tarzan found himself defenseless That they more greatly feared an attack from the rear than whatever\nunknown enemies lurked in their advance was evidenced by the formation\nof the column; and such was the fact, for they were fleeing from the\nwhite man's soldiers who had so harassed them for rubber and ivory that\nthey had turned upon their conquerors one day and massacred a white\nofficer and a small detachment of his black troops For a moment the lioness remained quiet with intense gaze directed upon\nthe flitting Esmeralda, whose goal appeared to be the cupboard, into\nwhich she attempted to propel her huge bulk; but as the shelves were\nbut nine or ten inches apart, she only succeeded in getting her head\nin; whereupon, with a final screech, which paled the jungle noises into\ninsignificance, she fainted once again The Englishman, finally concluding that he was a prisoner, saw no\nalternative open but to accompany his captor, and thus they traveled\nslowly through the jungle while the sable mantle of the impenetrable\nforest night fell about them, and the stealthy footfalls of padded paws\nmingled with the breaking of twigs and the wild calls of the savage\nlife that Clayton felt closing in upon him What were the thoughts of his strange captor or guide Clayton could\nonly vaguely conjecture; but that he had heard the shot, and was in\nsome manner affected by it was quite evident, for he quickened his pace\nso appreciably that Clayton, stumbling blindly in his wake, was down a\ndozen times in as many minutes in a vain effort to keep pace with him,\nand soon was left hopelessly behind The afternoon was drawing to a close, and Clayton, disheartened and\ndiscouraged, was in a terrible quandary as to the proper course to\npursue; whether to keep on in search of Professor Porter, at the almost\ncertain risk of his own death in the jungle by night, or to return to\nthe cabin where he might at least serve to protect Jane from the perils\nwhich confronted her on all sides

Chapter 9

Man's Reason

There was one of the tribe of Tarzan who questioned his authority, and that was Terkoz, the son of Tublat, but he so feared the keen knife and the deadly arrows of his new lord that he confined the manifestation of his objections to petty disobediences and irritating mannerisms; Tarzan knew, however, that he but waited his opportunity to wrest the kingship from him by some sudden stroke of treachery, and so he was ever on his guard against surprise. And then Lord Greystoke wiped his greasy fingers upon his naked thighs and took up the trail of Kulonga, the son of Mbonga, the king; while in far-off London another Lord Greystoke, the younger brother of the real Lord Greystoke's father, sent back his chops to the club's CHEF because they were underdone, and when he had finished his repast he dipped his finger-ends into a silver bowl of scented water and dried them upon a piece of snowy damask. The finding of the still warm body of Kulonga--on the very verge of their fields and within easy earshot of the village--knifed and stripped at the door of his father's home, was in itself insufficiently mysterious, but these last awesome discoveries within the village, within the dead Kulonga's own hut, filled their hearts with dismay, and conjured in their poor brains only the most frightful of superstitious explanations. But when, finally, he realized that his antagonist was fastened to him where his teeth and fists alike were useless against him, Terkoz hurled himself about upon the ground so violently that Tarzan could but cling desperately to the leaping, turning, twisting body, and ere he had struck a blow the knife was hurled from his hand by a heavy impact against the earth, and Tarzan found himself defenseless. That they more greatly feared an attack from the rear than whatever unknown enemies lurked in their advance was evidenced by the formation of the column; and such was the fact, for they were fleeing from the white man's soldiers who had so harassed them for rubber and ivory that they had turned upon their conquerors one day and massacred a white officer and a small detachment of his black troops. For a moment the lioness remained quiet with intense gaze directed upon the flitting Esmeralda, whose goal appeared to be the cupboard, into which she attempted to propel her huge bulk; but as the shelves were but nine or ten inches apart, she only succeeded in getting her head in; whereupon, with a final screech, which paled the jungle noises into insignificance, she fainted once again. The Englishman, finally concluding that he was a prisoner, saw no alternative open but to accompany his captor, and thus they traveled slowly through the jungle while the sable mantle of the impenetrable forest night fell about them, and the stealthy footfalls of padded paws mingled with the breaking of twigs and the wild calls of the savage life that Clayton felt closing in upon him. What were the thoughts of his strange captor or guide Clayton could only vaguely conjecture; but that he had heard the shot, and was in some manner affected by it was quite evident, for he quickened his pace so appreciably that Clayton, stumbling blindly in his wake, was down a dozen times in as many minutes in a vain effort to keep pace with him, and soon was left hopelessly behind. The afternoon was drawing to a close, and Clayton, disheartened and discouraged, was in a terrible quandary as to the proper course to pursue; whether to keep on in search of Professor Porter, at the almost uncertain risk of his own death in the jungle by night, or to return to the cabin where he might at least serve to protect Jane from the perils which confronted her on all sides. He was trying to tell the stupid white man to plunge his poisoned arrows into Sabor's back and sides, and to reach the savage heart with the long, thin hunting knife that hung at Tarzan's hip; but the man would not understand, and Tarzan did not dare release his hold to do the things himself, for he knew that the puny white man never could hold mighty Sabor alone, for an instant.

Chapter 10

Man's Reason

There was one of the tribe of Tarzan who questioned his authority, and that was Terkoz, the son of Tublat, but he so feared the keen knife and the deadly arrows of his new lord that he confined the manifestation of his objections to petty disobediences and irritating mannerisms; Tarzan knew, however, that he but waited his opportunity to wrest the kingship from him by some sudden stroke of treachery, and so he was ever on his guard against surprise. The finding of the still warm body of Kulonga--on the very verge of their fields and within easy earshot of the village--knifed and stripped at the door of his father's home, was in itself sufficiently mysterious, but these last awesome discoveries within the village, within the dead Kulonga's own hut, filled their hearts with dismay, and conjured in their poor brains only the most frightful of superstitious explanations. But when, finally, he realized that his antagonist was fastened to him where his teeth and fists alike were useless against him, Terkoz hurled himself about upon the ground so violently that Tarzan could but cling desperately to the leaping, turning, twisting body, and ere he had struck a blow the knife was hurled from his hand by a heavy impact against the earth, and Tarzan found himself defenseless. For a moment the lioness remained quiet with intense gaze directed upon the flitting Esmeralda, whose goal appeared to be the cupboard, into which she attempted to propel her huge bulk; but as the shelves were but nine or ten inches apart, she only succeeded in getting her head in; whereupon, with a final screech, which paled the jungle noises into insignificance, she fainted once again. The Englishman, finally concluding that he was a prisoner, saw no alternative open but to accompany his captor, and thus they traveled slowly through the jungle while the sable mantle of the impenetrable forest night fell about them, and the stealthy footfalls of padded paws mingled with the breaking of twigs and the wild calls of the savage life that Clayton felt closing in upon him. What were the thoughts of his strange captor or guide Clayton could only vaguely conjecture; but that he had heard the shot, and was in some manner affected by it was quite evident, for he quickened his pace so appreciably that Clayton, stumbling blindly in his wake, was down a dozen times in as many minutes in a vain effort to keep pace with him, and soon was left hopelessly behind. The afternoon was drawing to a close, and Clayton, disheartened and discouraged, was in a terrible quandary as to the proper course to pursue; whether to keep on in search of Professor Porter, at the almost uncertain risk of his own death in the jungle by night, or to return to the cabin where he might at least serve to protect Jane from the perils which confronted her on all sides. He was trying to tell the stupid white man to plunge his poisoned arrows into Sabor's back and sides, and to reach the savage heart with the long, thin hunting knife that hung at Tarzan's hip; but the man would not understand, and Tarzan did not dare release his hold to do the things himself, for he knew that the puny white man never could hold mighty Sabor alone, for an instant. It seems that an old bookworm who has a book and curio shop in Baltimore discovered between the leaves of a very old Spanish manuscript a letter written in 1550 detailing the adventures of a crew of mutineers of a Spanish galleon bound from Spain to South America with a vast treasure of "doubloons" and "pieces of eight," I suppose, for they certainly sound weird and piracy. But when one day a member of the tribe discovered the blacks in great numbers on the banks of a little stream that had been their watering place for generations, and in the act of clearing a space in the jungle and erecting many huts, the apes would remain no longer; and so Tarzan led them inland for many marches to a spot as yet undefiled by the foot of a human being.

Chapter 11

Man's Reason
There was one of the tribe of Tarzan who questioned his authority, and that was Terkoz, the son of Tublat, but he so feared the keen knife and the deadly arrows of his new lord that he confined the manifestation of his objections to petty disobediences and irritating mannerisms; Tarzan knew, however, that he but waited his opportunity to wrest the kingship from him by some sudden stroke of treachery, and so he was ever on his guard against surprise. But when, finally, he realized that his antagonist was fastened to him nowhere his teeth and fists alike were useless against him, Terkoz hurled himself about upon the ground so violently that Tarzan could but cling desperately to the leaping, turning, twisting body, and ere he had struck a blow the knife was hurled from his hand by a heavy impact against the earth, and Tarzan found himself defenseless. For a moment the lioness remained quiet with intense gaze directed upon the flitting Esmeralda, whose goal appeared to be the cupboard, into which she attempted to propel her huge bulk; but as the shelves were but nine or ten inches apart, she only succeeded in getting her head in; whereupon, with a final screech, which paled the jungle noises into insignificance, she fainted once again. The Englishman, finally concluding that he was a prisoner, saw no alternative open but to accompany his captor, and thus they traveled slowly through the jungle while the sable mantle of the impenetrable forest night fell about them, and the stealthy footfalls of padded paws mingled with the breaking of twigs and the wild calls of the savage life that Clayton felt closing in upon him. What were the thoughts of his strange captor or guide Clayton could only vaguely conjecture; but that he had heard the shot, and was in some manner affected by it was quite evident, for he quickened his pace so appreciably that Clayton, stumbling blindly in his wake, was down a dozen times in as many minutes in a vain effort to keep pace with him, and soon was left hopelessly behind. The afternoon was drawing to a close, and Clayton, disheartened and discouraged, was in a terrible quandary as to the proper course to pursue; whether to keep on in search of Professor Porter, at the almost uncertain risk of his own death in the jungle by night, or to return to the cabin where he might at least serve to protect Jane from the perils which confronted her on all sides. He was trying to tell the stupid white man to plunge his poisoned arrows into Sabor's back and sides, and to reach the savage heart with the long, thin hunting knife that hung at Tarzan's hip; but the man would not understand, and Tarzan did not dare release his hold to do the things himself, for he knew that the puny white man never could hold mighty Sabor alone, for an instant. It seems that an old bookworm who has a book and curio shop in Baltimore discovered between the leaves of a very old Spanish manuscript a letter written in 1550 detailing the adventures of a crew of mutineers of a Spanish galleon bound from Spain to South America with a vast treasure of "doubloons" and "pieces of eight," I suppose, for they certainly sound weird and piraty. But when one day a member of the tribe discovered the blacks in great numbers on the banks of a little stream that had been their watering place for generations, and in the act of clearing a space in the jungle and erecting many huts, the apes would remain no longer; and so Tarzan led them inland for many marches to a spot as yet undefiled by the foot of a human being. Tarzan had seen the consternation depicted upon the faces of the little group as they witnessed the departure of the Arrow; so, as the ship was a wonderful novelty to him in addition, he determined to hasten out to the point of land at the north of the harbor's mouth and obtain a nearer view of the boat, as well as to learn, if possible, the direction of its flight.

Chapter 12

Man's Reason
There was one of the tribe of Tarzan who questioned his authority, and that was Terkoz, the son of Tublat, but he so feared the keen knife and the deadly arrows of his new lord that he confined the manifestation of his objections to petty disobediences and irritating mannerisms; Tarzan knew, however, that he but waited his opportunity to wrest the kingship from him by some sudden stroke of treachery, and so he was ever on his guard against surprise. But when, finally, he realized that his antagonist was fastened to him nowhere his teeth and fists alike were useless against him, Terkoz hurled himself about upon the ground so violently that Tarzan could but cling desperately to the leaping, turning, twisting body, and ere he had struck a blow the knife was hurled from his hand by a heavy impact against the earth, and Tarzan found himself defenseless. For a moment the lioness remained quiet with intense gaze directed upon the flitting Esmeralda, whose goal appeared to be the cupboard, into which she attempted to propel her huge bulk; but as the shelves were but nine or ten inches apart, she only succeeded in getting her head in; whereupon, with a final screech, which paled the jungle noises into insignificance, she fainted once again. The Englishman, finally concluding that he was a prisoner, saw no alternative open but to accompany his captor, and thus they traveled slowly through the jungle while the sable mantle of the impenetrable forest night fell about them, and the stealthy footfalls of padded paws mingled with the breaking of twigs and the wild calls of the savage life that Clayton felt closing in upon him. What were the thoughts of his strange captor or guide Clayton could only vaguely conjecture; but that he had heard the shot, and was in some manner affected by it was quite evident, for he quickened his pace so appreciably that Clayton, stumbling blindly in his wake, was down a dozen times in as many minutes in a vain effort to keep pace with him, and soon was left hopelessly behind. The afternoon was drawing to a close, and Clayton, disheartened and discouraged, was in a terrible quandary as to the proper course to pursue; whether to keep on in search of Professor Porter, at the almost uncertain risk of his own death in the jungle by night, or to return to the cabin where he might at least serve to protect Jane from the perils which confronted her on all sides. He was trying to tell the stupid white man to plunge his poisoned arrows into Sabor's back and sides, and to reach the savage heart with the long, thin hunting knife that hung at Tarzan's hip; but the man would not understand, and Tarzan did not dare release his hold to do the things himself, for he knew that the puny white man never could hold mighty Sabor alone, for an instant. It seems that an old bookworm who has a book and curio shop in Baltimore discovered between the leaves of a very old Spanish manuscript a letter written in 1550 detailing the adventures of a crew of mutineers of a Spanish galleon bound from Spain to South America with a vast treasure of "doubloons" and "pieces of eight," I suppose, for they certainly sound weird and piracy. But when one day a member of the tribe discovered the blacks in great numbers on the banks of a little stream that had been their watering place for generations, and in the act of clearing a space in the jungle and erecting many huts, the apes would remain no longer; and so Tarzan led them inland for many marches to a spot as yet undefiled by the foot of a human being. Tarzan had seen the consternation depicted upon the faces of the little group as they witnessed the departure of the Arrow; so, as the ship was a wonderful novelty to him in addition, he determined to hasten out to the point of land at the north of the harbor's mouth and obtain a nearer view of the boat, as well as to learn, if possible, the direction of its flight.

Chapter 13

An ordinary white man such as any of you--pardon me, I did not\mean just that--rather, a white man above the ordinary in physique and\intelligence could never, I grant you, have lived a year alone and\nnaked in this tropical jungle; but this man not only surpasses the\naverage white man in strength and agility, but as far transcends our\ntrained athletes and 'strong men' as they surpass a day-old babe; and\nhis courage and ferocity in battle are those of the wild beast Clayton could not but realize that the Frenchman felt his grief the\nmore keenly because D'Arnot's sacrifice had been so futile, since Jane\nhad been rescued before D'Arnot had fallen into the hands of the\nsavages, and again because the service in which he had lost his life\nhad been outside his duty and for strangers and aliens; but when he\nspoke of it to Lieutenant Charpentier, the latter shook his head For a moment the lioness remained quiet with intense gaze directed upon\nthe flitting Esmeralda, whose goal appeared to be the cupboard, into\nwhich she attempted to propel her huge bulk; but as the shelves were\nbut nine or ten inches apart, she only succeeded in getting her head\nin; whereupon, with a final screech, which paled the jungle noises into\ninsignificance, she fainted once again The Englishman, finally concluding that he was a prisoner, saw no\nalternative open but to accompany his captor, and thus they traveled\nslowly through the jungle while the sable mantle of the impenetrable\nforest night fell about them, and the stealthy footfalls of padded paws\nmingled with the breaking of twigs and the wild calls of the savage\nlife that Clayton felt closing in upon him What were the thoughts of his strange captor or guide Clayton could\nonly vaguely conjecture; but that he had heard the shot, and was in\nsome manner affected by it was quite evident, for he quickened his pace\nso appreciably that Clayton, stumbling blindly in his wake, was down a\nndozen times in as many minutes in a vain effort to keep pace with him,\nand soon was left hopelessly behind A single\nshot before they emerged from the jungle had announced to those in camp\nas well as on the ship that the expedition had been too late--for it\nhad been prearranged that when they came within a mile or two of camp\nnone shot was to be fired to denote failure, or three for success, while\ntwo would have indicated that they had found no sign of either D'Arnot\nnor his black captors The afternoon was drawing to a close, and Clayton, disheartened and\ndiscouraged, was in a terrible quandary as to the proper course to\npursue; whether to keep on in search of Professor Porter, at the almost\nuncertain risk of his own death in the jungle by night, or to return to\nthe cabin where he might at least serve to protect Jane from the perils\nwhich confronted her on all sides He was trying to tell the stupid white man to plunge his poisoned\nnarrows into Sabor's back and sides, and to reach the savage heart with\nthe long, thin hunting knife that hung at Tarzan's hip; but the man\nwould not understand, and Tarzan did not dare release his hold to do\nthe things himself, for he knew that the puny white man never could\nhold mighty Sabor alone, for an instant It seems that an old bookworm who has a book and curio shop in\nBaltimore discovered between the leaves of a very old Spanish\nmanuscript a letter written in 1550 detailing the adventures of a crew\nof mutineers of a Spanish galleon bound from Spain to South America\nwith a vast treasure of "doubloons" and "pieces of eight," I suppose,\nfor they certainly sound weird and piraty Tarzan had seen the consternation depicted upon the faces of the little\ngroup as they witnessed the departure of the Arrow; so, as the ship was\na wonderful novelty to him in addition, he determined to hasten out to\nthe point of land at the north of the harbor's mouth and obtain a\nnearer view of the boat, as well as to learn, if possible, the\ndirection of its flight

Chapter 14

An ordinary white man such as any of you--pardon me, I did not\mean just that--rather, a white man above the ordinary in physique and\intelligence could never, I grant you, have lived a year alone and\nnaked in this tropical jungle; but this man not only surpasses the\naverage white man in strength and agility, but as far transcends our\ntrained athletes and 'strong men' as they surpass a day-old babe; and\nhis courage and ferocity in battle are those of the wild beast Clayton could not but realize that the Frenchman felt his grief the\nmore keenly because D'Arnot's sacrifice had been so futile, since Jane\nhad been rescued before D'Arnot had fallen into the hands of the\nsavages, and again because the service in which he had lost his life\nhad been outside his duty and for strangers and aliens; but when he\nspoke of it to Lieutenant Charpentier, the latter shook his head For a moment the lioness remained quiet with intense gaze directed upon\nthe flitting Esmeralda, whose goal appeared to be the cupboard, into\nwhich she attempted to propel her huge bulk; but as the shelves were\nbut nine or ten inches apart, she only succeeded in getting her head\nin; whereupon, with a final screech, which paled the jungle noises into\ninsignificance, she fainted once again It was not the roaring and growling of the big beasts\nthat affected me so much as it was the stealthy noises--the ones that\nyou heard suddenly close by and then listened vainly for a repetition\nof--the unaccountable sounds as of a great body moving almost\nnoiselessly, and the knowledge that you didn't KNOW how close it was,\nor whether it were creeping closer after you ceased to hear it The Englishman, finally concluding that he was a prisoner, saw no\nalternative open but to accompany his captor, and thus they traveled\nslowly through the jungle while the sable mantle of the impenetrable\nforest night fell about them, and the stealthy footfalls of padded paws\nmingled with the breaking of twigs and the wild calls of the savage\nlife that Clayton felt closing in upon him What were the thoughts of his strange captor or guide Clayton could\nonly vaguely conjecture; but that he had heard the shot, and was in\nsome manner affected by it was quite evident, for he quickened his pace\nso appreciably that Clayton, stumbling blindly in his wake, was down a\nnd dozen times in as many minutes in a vain effort to keep pace with him,\nand soon was left hopelessly behind A single\nshot before they emerged from the jungle had announced to those in camp\nas well as on the ship that the expedition had been too late--for it\nhad been prearranged that when they came within a mile or two of camp\none shot was to be fired to denote failure, or three for success, while\ntwo would have indicated that they had found no sign of either D'Arnot\nnor his black captors The afternoon was drawing to a close, and Clayton, disheartened and\ndiscouraged, was in a terrible quandary as to the proper course to\npursue; whether to keep on in search of Professor Porter, at the almost\nuncertain risk of his own death in the jungle by night, or to return to\nthe cabin where he might at least serve to protect Jane from the perils\nwhich confronted her on all sides He was trying to tell the stupid white man to plunge his\npoisoned\nnarrows into Sabor's back and sides, and to reach the savage heart with\nthe long, thin hunting knife that hung at Tarzan's hip; but the man\nwould not understand, and Tarzan did not dare release his hold to do\nthe things himself, for he knew that the puny white man never could\nhold mighty Sabor alone, for an instant It seems that an old bookworm who has a book and curio shop in\nBaltimore discovered between the leaves of a very old Spanish\nmanuscript a letter written in 1550 detailing the adventures of a crew\nof mutineers of a Spanish galleon bound from Spain to South America\nwith a vast treasure of "doubloons" and "pieces of eight," I suppose,\nfor they certainly sound weird and piracy

Chapter 15

An ordinary white man such as any of you--pardon me, I did not\mean just that--rather, a white man above the ordinary in physique and\intelligence could never, I grant you, have lived a year alone and\nnaked in this tropical jungle; but this man not only surpasses the\naverage white man in strength and agility, but as far transcends our\ntrained athletes and 'strong men' as they surpass a day-old babe; and\nhis courage and ferocity in battle are those of the wild beast Clayton could not but realize that the Frenchman felt his grief the\nmore keenly because D'Arnot's sacrifice had been so futile, since Jane\nhad been rescued before D'Arnot had fallen into the hands of the\nsavages, and again because the service in which he had lost his life\nhad been outside his duty and for strangers and aliens; but when he\nspoke of it to Lieutenant Charpentier, the latter shook his head It was not the roaring and growling of the big beasts\nthat affected me so much as it was the stealthy noises--the ones that\nyou heard suddenly close by and then listened vainly for a repetition\nof--the unaccountable sounds as of a great body moving almost\nnoiselessly, and the knowledge that you didn't KNOW how close it was,\nor whether it were creeping closer after you ceased to hear it What were the thoughts of his strange captor or guide Clayton could\nonly vaguely conjecture; but that he had heard the shot, and was in\nsome manner affected by it was quite evident, for he quickened his pace\nso appreciably that Clayton, stumbling blindly in his wake, was down a\nndozen times in as many minutes in a vain effort to keep pace with him,\nand soon was left hopelessly behind A single\nshot before they emerged from the jungle had announced to those in camp\nas well as on the ship that the expedition had been too late--for it\nhad been prearranged that when they came within a mile or two of camp\none shot was to be fired to denote failure, or three for success, while\nntwo would have indicated that they had found no sign of either D'Arnot\nnor his black captors He was trying to tell the stupid white man to plunge his poisoned\nnarrows into Sabor's back and sides, and to reach the savage heart with\nthe long, thin hunting knife that hung at Tarzan's hip; but the man\nwould not understand, and Tarzan did not dare release his hold to do\nthe things himself, for he knew that the puny white man never could\nhold mighty Sabor alone, for an instant Gradually he became accustomed to the strange noises and the odd ways\nof civilization, so that presently none might know that two short\nmonths before, this handsome Frenchman in immaculate white ducks,\nwho\nlaughed and chatted with the gayest of them, had been swinging naked\nthrough primeval forests to pounce upon some unwary victim, which, raw,\nwas to fill his savage belly Tarzan had seen the consternation depicted upon the faces of the little\ngroup as they witnessed the departure of the Arrow; so, as the ship was\na wonderful novelty to him in addition, he determined to hasten out to\nthe point of land at the north of the harbor's mouth and obtain a\nnearer view of the boat, as well as to learn, if possible, the\ndirection of its flight When those who could had recovered, the entire story had been told to\nthe French commander; but the men were too ignorant to be able to tell\nhim at just what point on the coast the professor and his party had\nbeen marooned, so the cruiser had steamed slowly along within sight of\nland, firing occasional signal guns and scanning every inch of the\nbeach with glasses Philander; \"permit me to\nsuggest that doubtless the Moors who were conquered in the fifteenth\ncentury will continue in that most regrettable condition for the time\nbeing at least, even though we postpone discussion of that world\ncalamity until we may attain the enchanting view of yon FELIS CARNIVORA\nwhich distance proverbially is credited with lending

Corpus Cleaning - Data Wrangling

Now we begin the process of corpus cleaning. We first use function removeNumPunct to remove numbers and punctuation from the input text. This function takes a character vector x as input and uses the gsub function to replace any characters that are not alphabetic ([:alpha:]) or whitespace ([:space:]) with an empty string "". Then we use the tm_map function from the tm package to apply the removeNumPunct function to each document in the TarzanData corpus. It removes numbers and punctuation from each document and stores the result in TarzanDatacl. Now tm_map is again used to transform the text in TarzanDatacl to lowercase using the tolower function. The result is stored in TarzanDataLow. Then we lastly inspect the contents of TarzanDataLow.

```
> # Function to remove numbers and punctuation
> removeNumPunct <- function(x) gsub("[^[:alpha:][:space:]]*", "", x)
> removeNumPunct
function(x) gsub("[^[:alpha:][:space:]]*", "", x)
> # Removing numbers and punctuation
> TarzanDatacl <- tm::tm_map(TarzanData, content_transformer(removeNumPunct))
> TarzanDatacl
<<VCorpus>>
Metadata: corpus specific: 0, document level (indexed): 0
Content: documents: 1
> str(TarzanDatacl)
Classes 'VCorpus', 'Corpus' hidden list of 3
$ content:List of 1
..$ :List of 2
... ..$ content: chr [1:11031] "" "Tarzan of the Apes" "" ...
... ..$ meta   :List of 7
... ... .$ author      : chr(0)
... ... .$ datetimestamp: POSIXlt[1:1], format: "2024-05-03 15:13:02"
... ... .$ description  : chr(0)
... ... .$ heading     : chr(0)
... ... .$ id          : chr "TarzanOfTheApes.txt"
... ... .$ language    : chr "en"
... ... .$ origin      : chr(0)
... ... .- attr(*, "class")= chr "TextDocumentMeta"
... ... .- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
$ meta   : list()
..- attr(*, "class")= chr "CorpusMeta"
$ dmeta  :'data.frame':       1 obs. of  0 variables
```

```

> # Converting into lower case
> TarzanDataLow <- tm_map(TarzanDatacl, tm::content_transformer(tolower))
> TarzanDataLow
<<VCorpus>>
Metadata: corpus specific: 0, document level (indexed): 0
Content: documents: 1
> str(TarzanDataLow)
Classes 'VCorpus', 'Corpus' hidden list of 3
$ content:List of 1
..$ :List of 2
...$ content: chr [1:11031] "" "tarzan of the apes" "" ...
...$ meta :List of 7
...$ author : chr(0)
...$ timestamp: POSIXlt[1:1], format: "2024-05-03 15:30:22"
...$ description : chr(0)
...$ heading : chr(0)
...$ id : chr "TarzanOfTheApes.txt"
...$ language : chr "en"
...$ origin : chr(0)
...- attr(*, "class")= chr "TextDocumentMeta"
...- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
$ meta : list()
..$ :attr(*, "class")= chr "CorpusMeta"
$ dmeta :'data.frame': 1 obs. of 0 variables
> inspect(TarzanDataLow)
<<VCorpus>>
Metadata: corpus specific: 0, document level (indexed): 0
Content: documents: 1

[[1]]
<<PlainTextDocument>>
Metadata: 7
Content: chars: 456013

```

Now we are going to further preprocess the text data stored in the TarzanDataLow corpus and create a Document-Term Matrix (DTM) representation. First, the DocumentTermMatrix function is applied to TarzanDataLow to generate the DTM, which represents the frequency of terms (words) in each document of the corpus. Next, myStopwords is defined as a vector containing English stopwords, commonly occurring words that are often not useful for analysis. These stopwords are then removed from each document in the TarzanDataLow corpus using the tm_map function along with tm::removeWords, resulting in TarzanDataStop, a new preprocessed corpus with stopwords removed. Finally, tm::inspect is used to view the first document in the TarzanDataStop corpus, allowing for inspection of the preprocessing results, such as stopwords removal. This series of operations prepares the text data for further analysis by creating a DTM and removing common stopwords.

```

> # DTM for the corpus with no punc, num and lower case
> TarzanDataDTM <- DocumentTermMatrix(TarzanDataLow)
> TarzanDataDTM
<<DocumentTermMatrix (documents: 1, terms: 7704)>>
Non-/sparse entries: 7704/0
Sparsity : 0%
Maximal term length: 20
Weighting : term frequency (tf)
> str(TarzanDataDTM)
List of 6
 $ i      : int [1:7704] 1 1 1 1 1 1 1 1 1 ...
 $ j      : int [1:7704] 1 2 3 4 5 6 7 8 9 10 ...
 $ v      : num [1:7704] 3 3 1 1 2 1 2 1 3 15 ...
 $ nrow   : int 1
 $ ncol   : int 7704
 $ dimnames:List of 2
   ..$ Docs : chr "TarzanOfTheApes.txt"
   ..$ Terms: chr [1:7704] "abandon" "abandoned" "abandoning" "abashed" ...
- attr(*, "class")= chr [1:2] "DocumentTermMatrix" "simple_triplet_matrix"
- attr(*, "weighting")= chr [1:2] "term frequency" "tf"
> myStopwords <- c(tm::stopwords("english"))
> myStopwords
[1] "i"        "me"       "my"       "myself"    "we"       "our"      "ours"
[8] "ourselves" "you"     "your"     "yours"     "yourself" "yourselves" "he"
[15] "him"      "his"      "himself"  "she"       "her"      "hers"     "herself"
[22] "it"       "its"      "itself"   "they"     "them"     "their"    "theirs"
[29] "themselves" "what"   "which"    "who"      "whom"     "this"     "that"
[36] "these"    "those"   "am"       "is"       "are"      "was"      "were"
[43] "be"        "been"    "being"    "have"     "has"      "had"      "having"
[50] "do"        "does"    "did"      "doing"    "would"    "should"   "could"
[57] "ought"    "i'm"     "you're"   "he's"     "she's"    "it's"     "we're"
[64] "they're"  "i've"    "you've"   "we've"    "they've"  "i'd"      "you'd"
[71] "he'd"      "she'd"   "we'd"     "they'd"   "i'll"     "you'll"   "he'll"
[78] "she'll"    "we'll"   "they'll"  "isn't"    "aren't"   "wasn't"   "weren't"
[85] "hasn't"   "haven't" "hadn't"   "doesn't" "don't"    "didn't"   "won't"
[92] "wouldn't" "shan't"  "shouldn't" "can't"    "cannot"   "couldn't" "mustn't"
[99] "let's"     "that's"  "who's"    "what's"   "here's"   "there's"  "when's"
[106] "where's"  "why's"   "how's"   "a"        "an"       "the"      "and"
[113] "but"      "if"      "or"      "because"  "as"       "until"    "while"
[120] "of"       "at"      "by"      "for"      "with"     "about"    "against"
[127] "between"  "into"    "through" "during"   "before"   "after"    "above"
[134] "below"    "to"      "from"    "up"       "down"     "in"       "out"
[141] "on"       "off"     "over"    "under"   "again"    "further"  "then"
[148] "once"     "here"    "there"   "when"    "where"   "why"      "how"
[155] "all"      "any"     "both"    "each"    "few"     "more"     "most"
[162] "other"    "some"    "such"    "no"      "nor"     "not"      "only"
[169] "own"      "same"    "so"      "than"    "too"     "very"

```

The TermDocumentMatrix function from the tm package is applied to TarzanDataStop, resulting in TarzanDataStopDTM, which represents the frequency of terms (words) in each document of the corpus. Next, the findFreqTerms function

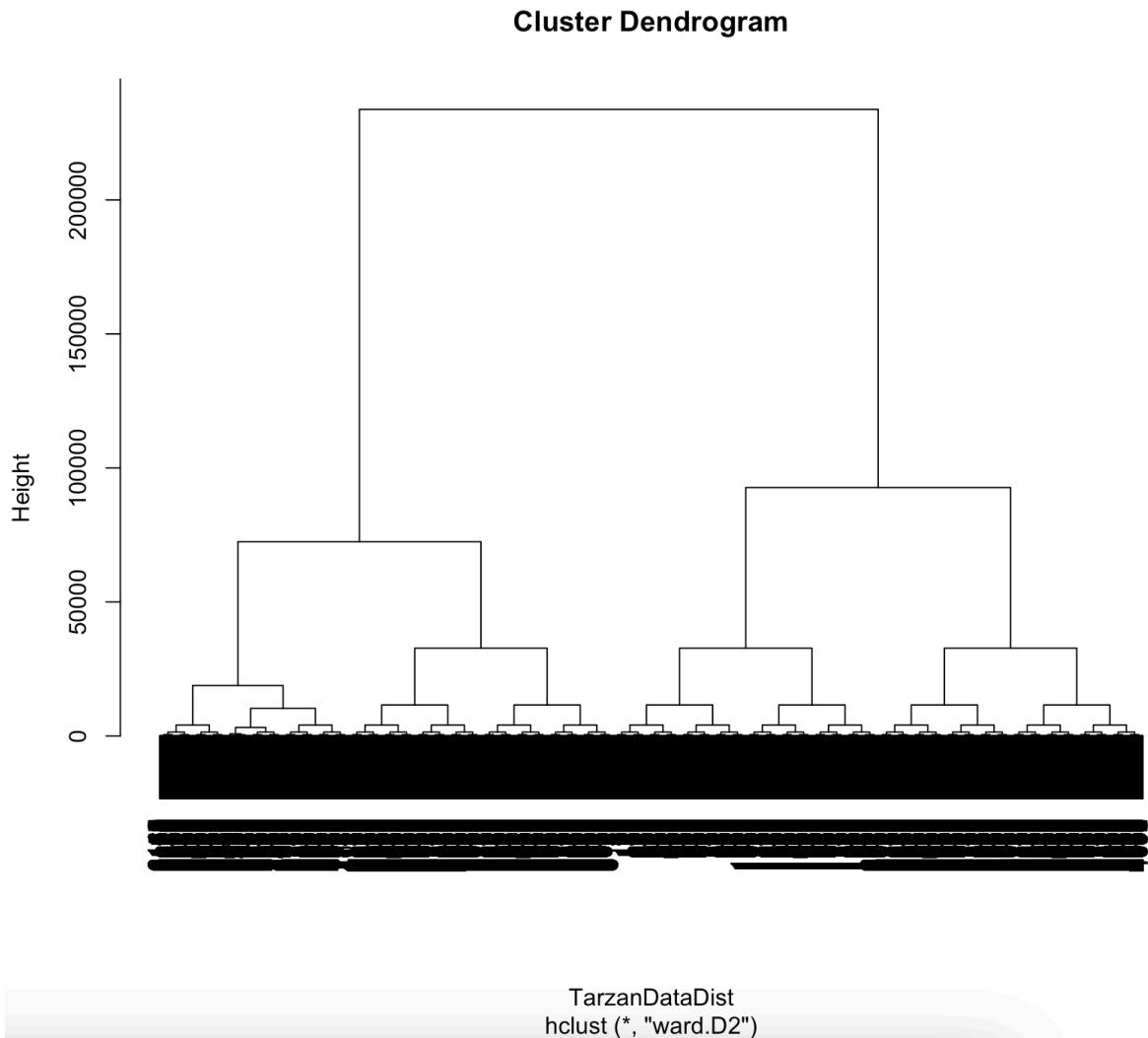
is used to identify terms that occur at least 5 times in the TarzanDataStopDTM. These terms are stored in the freqTerms vector. Following that, the length of the third term in freqTerms is calculated using the nchar function, and the term itself is printed. Then, the term frequency of the first document in the TarzanDataStop corpus is computed using the termFreq function, resulting in TarzanDatatf. Finally, inspect is used to view the TarzanDataStopDTM, providing a detailed inspection of the term-document matrix, allowing for further analysis and interpretation of the text data.

```
> # DTM for the corpus with removed stopwords
> TarzanDataStopDTM <- tm::TermDocumentMatrix(TarzanDataStop)
> TarzanDataStopDTM
<<TermDocumentMatrix (terms: 7605, documents: 1)>>
Non-/sparse entries: 7605/0
Sparsity      : 0%
Maximal term length: 20
Weighting      : term frequency (tf)
> # Terms that occur atleast 5 times
> freqTerms <- tm::findFreqTerms(TarzanDataStopDTM, lowfreq = 5)
> freqTerms
 [1] "able"        "aboard"       "absolute"     "accompanied"   "accompany"    "accomplished"
 [7] "accord"       "account"      "accustomed"   "across"        "act"          "add"
[13] "added"        "addition"     "admit"        "advance"       "advanced"     "advancing"
[19] "advantage"    "adventure"    "afraid"       "africa"        "african"      "afternoon"
[25] "ages"         "agility"       "ago"          "agony"         "ahead"        "aid"
[31] "aint"         "air"          "alarm"        "alice"         "alive"        "almost"
[37] "alone"         "along"        "aloud"        "already"      "also"         "always"
[43] "america"      "american"     "ammunition"   "among"         "amphitheater" "anchor"
[49] "ancient"      "anger"        "angry"        "animal"       "animals"      "announced"
[55] "another"      "answer"       "answered"     "answering"    "antagonist"   "anthropoid"
[61] "anthropoids"  "anxious"      "anything"     "apart"        "ape"          "apeman"
[67] "apes"         "apparent"     "appeared"     "apprehension" "approach"     "approached"
[73] "approaching" "archer"       "archimedes"   "arena"        "arm"          "armed"
> tm::inspect(TarzanDataStopDTM)
<<TermDocumentMatrix (terms: 7605, documents: 1)>>
Non-/sparse entries: 7605/0
Sparsity      : 0%
Maximal term length: 20
Weighting      : term frequency (tf)
Sample        :
  Docs
Terms      TarzanOfTheApes.txt
clayton    268
darnot     186
great      237
jungle     223
little     290
man        264
now        180
one        304
tarzan     571
upon       428
```

Dendrogram

We are now going to construct a dendrogram to visualize the hierarchical clustering of the documents based on their textual similarity. First, the term-document matrix stored in TarzanDataStopDTM is converted into a data frame TarzanDatadf, which is then used to calculate the distance between the documents using the dist function, resulting in TarzanDataDist. Subsequently, hierarchical clustering is performed on TarzanDataDist using the hclust function with the method specified as "ward.D2", which minimizes the within-cluster variance. The resulting hierarchical clustering tree is stored in TarzanDataDG. Finally, the dendrogram is plotted using the plot function, although it's noted that the plot may be cluttered, suggesting that further adjustments or alternative visualization methods may be needed for clearer interpretation of the clustering results.

```
> # Dendrogram
> TarzanDatadf <- as.data.frame(TarzanDataStopDTM[[1]])
> TarzanDataDist <- dist(TarzanDatadf)
> TarzanDataDG <- hclust(TarzanDataDist, method = "ward.D2")
> str(TarzanDataDG)
List of 7
 $ merge      : int [1:7604, 1:2] -1 -3 -5 -7 -9 -11 -13 -15 -17 -19 ...
 $ height     : num [1:7604] 1 1 1 1 1 1 1 1 1 ...
 $ order      : int [1:7605] 6145 6146 6147 6148 6149 6150 6151 6152 6153 6154 ...
 $ labels     : NULL
 $ method     : chr "ward.D2"
 $ call       : language hclust(d = TarzanDataDist, method = "ward.D2")
 $ dist.method: chr "euclidean"
 - attr(*, "class")= chr "hclust"
> # Plot not helpful as it is cluttered
> plot(TarzanDataDG)
```



But there are simply too many words to make out anything. A better way to get a good visualization is a word cloud, which in this case generates larger words when they are more frequent.

Word Cloud

We use word cloud visualization of the term frequencies in the TarzanDatatf object is created. First, the variable words is assigned the names of the term frequency vector TarzanDatatf, which contains the terms (words) and their corresponding frequencies. Then, a color palette pal is generated using the brewer.pal function from the RColorBrewer package, specifying a sequential palette of green shades with 9 colors. Next, the wordcloud function is applied to words and TarzanDatatf,

generating the word cloud visualization. The colors parameter is set to use the generated color palette, excluding the first four colors, ensuring that the word cloud is displayed with a pleasing green color scheme. The resulting word cloud, stored in TarzanDatawc, visually represents the frequencies of terms in the text data, with larger and more prominent words indicating higher frequencies.

```
> # Word Cloud
> words <- names(TarzanDatatf)
> words
 [1] "abandon"      "abandoned"     "abandoning"    "abashed"      "abated"
 [6] "abatis"        "abduction"     "aberration"   "ability"       "able"
[11] "aboard"        "abode"         "abound"        "aboutlord"    "abroad"
[16] "abruptly"      "absence"       "absent"        "absentmindedness" "absinthe"
[21] "absolute"      "absolutely"    "absorbed"      "absorbing"    "abstract"
[26] "abstruse"      "abundance"     "abundant"      "abuses"        "abyssmal"
[31] "accede"         "accentuate"   "accentuated"  "accept"        "accepted"
[36] "accession"      "accident"      "accidental"   "accidentally" "accidents"
[41] "accommodate"    "accompanied"  "accompany"    "accomplish"   "accomplished"
[46] "accomplishment" "accord"        "accorded"      "according"    "account"
[51] "accounted"      "accouterments" "accuracy"      "accurately"   "accused"
[56] "accustom"       "accustomed"   "achievement"  "aching"        "acknowledging"
[61] "acquaint"       "acquaintance" "acquaintances" "acquainted"   "acquiesced"
[66] "acquirement"    "across"        "act"          "acted"        "action"
[71] "actiondetermined" "actions"      "active"        "acts"         "acuteness"
[76] "add"            "added"        "adding"        "addition"    "additional"
[81] "addressed"      "addressing"   "adds"          "adequate"    "adjacent"
[86] "adjusted"       "adjusting"    "admiral"       "admiration"  "admirationwatched"
[91] "admired"        "admit"        "admitted"     "admitting"   "admonish"
[96] "admonished"     "admonition"  "adopted"       "adroitly"     "adult"
[101] "adults"         "advance"      "advanced"      "advancing"   "advantage"
[106] "advent"         "adventure"   "adventureboth" "adventures"   "adventurous"
[111] "adverbs"        "adversary"   "advised"       "aerial"       "aerie"
```

We will create two word clouds using different color palettes. First, a color palette pal is created using the brewer.pal function with the "BuGn" palette from the RColorBrewer package, containing 9 shades of green. The structure of pal is examined using str. Then, the wordcloud function is used to create a word cloud visualization using words and TarzanDatatf, with the colors parameter set to pal excluding the first four colors, resulting in TarzanDatawc. Next, a different color palette pal2 is generated using the "Spectral" palette, also containing 9 colors. Another word cloud visualization is created using pal2, resulting in TarzanDatawc. Finally, the first 10 content items of the TarzanDatacl corpus are retrieved and printed to inspect the text data stored within TarzanDataText.

forest think waters warl
ate gleamed call harbor growls extended sie sufficiently baffled jar **thought** undergrowth
manuscript writhing stumbled without michaels silk claytons coast joy
ends females wailing hide place quarters climbed one
rewarded speak wailed diamond room likeness work carefully
wildly beings revel act wives effectively sabor hut sight vision
halt twisting ignorant lofty done rapidity intense fronds swayed could bales attempted
dried forward new witnessed speaks **professor** mistress
golden six result hearing express frightful untouched recalled seething screams lead palisade moonlight
tops result distinctly replaced shrill friend impenetrable loan brother showing shame nobleman animals powerful responded
tittle king engaged gasped needs chase fit hastened figures animals
cot golden result sank fullgrown height thirteen barred became dieu mingled
equally precious king silent brother showing shame nobleman animals powerful responded
instinct told engaged gasped needs chase fit hastened figures animals powerful responded
common developed despair species verdure thirteen barred became dieu mingled
sick enemies towering numerous meat red glimpses person add stranger melted
upward opening discover grass slip shortly betokened wondering snowy din narrated
frequency fetch accept bird whenever meat red glimpses person add stranger melted
common knife mark escapeforth state cat **kala** old camp fascinating idea
realization acquaint familiarity seems nursing risen can protector spear beginning
fact ask name **many** early merest characteristics ink garments building moments odd boat started land bury
frequency sore impossible scalp taut flat happened lip signs moaning point marry boat awaiting clothes valley
throng sun within awake seconds pressed eye meager stand store tublats talon keen pots bid talked
ended made latter shock argument buying accorded clothed city points lie size wanted points
trip worm worked cold wondrous given let afternoon tree floor talon keen pots bid talked
determine cut amphitheater falling attempts paw tightly late wrest turf strolled muscles flowers fairly tin
shaft bolgan amorphous script may sorrow stepped thats natives dig habitation failed clay future owed
caught victims samuel danger delight vanished sobbed title spain characters pile bits gave
solely bravely drink prey completely slivers case renew likely regained accord patience
hilt sorry tossed won dog history different anyone toy comrade michael emergency
ribs get pointed help guns noise inert families things roar
suddenly birth examination noise inert families things post
strange o

Quanteda

Next, we can use a different library (Quanteda) to get term frequencies and other information. The quanteda package is employed to perform text analysis on the TarzanDataText corpus. Initially, the text content of the first 10 documents within the corpus is tokenized using the tokens function, resulting in TarzanDataTokens, a list of tokens representing each document's text. The structure of TarzanDataTokens is displayed using str. Following this, a document-feature matrix (DFM) is created from TarzanDataTokens using the dfm function, producing TarzanDataDFM, which represents the frequency of tokens in each document. The structure of TarzanDataDFM is shown using str. Next, the document frequency (the number of documents in which each feature/token appears) is calculated from TarzanDataDFM using the docfreq function, resulting in TarzanDataDocFreq, which is displayed using str and printed. Finally, the weights of the features in TarzanDataDFM are adjusted using dfm_weight, resulting in TarzanDataWeights, which is also displayed using str. These operations illustrate the initial steps of text analysis using quanteda, including tokenization, creating a document-feature matrix, calculating document frequency, and weighting the features.

```
> # quanteda
> TarzanDataTokens <- quanteda::tokens(TarzanDataText$content[1:10])
> str(TarzanDataTokens)
List of 10
$ text1 : chr(0)
$ text2 : chr [1:4] "Tarzan" "of" "the" "Apes"
$ text3 : chr(0)
$ text4 : chr(0)
$ text5 : chr "By"
$ text6 : chr(0)
$ text7 : chr [1:3] "Edgar" "Rice" "Burroughs"
$ text8 : chr(0)
$ text9 : chr(0)
> TarzanDataDFM <- quanteda::dfm(TarzanDataTokens)
> str(TarzanDataDFM)
Formal class 'dfm' [package "quanteda"] with 8 slots
..@ docvars :data.frame:   10 obs. of  3 variables:
... ..$ docname_ : chr [1:10] "text1" "text2" "text3" "text4" ...
... ..$ docid_  : Factor w/ 10 levels "text1","text2",..: 1 2 3 4 5 6 7 8 9 10
... ..$ segid_  : int [1:10] 1 1 1 1 1 1 1 1 1 1
..@ meta    :List of 3
... ..$ system:List of 5
```

```

> TarzanDataDocFreq <- quanteda::docfreq(TarzanDataDFM)
> str(TarzanDataDocFreq)
Named int [1:8] 1 1 1 1 1 1 1 1
- attr(*, "names")= chr [1:8] "tarzan" "of" "the" "apes" ...
> TarzanDataDocFreq
  tarzan      of      the      apes      by      edgar      rice burroughs
  1          1          1          1          1          1          1          1
> TarzanDataWeights <- quanteda::dfm_weight(TarzanDataDFM)
> str(TarzanDataWeights)
Formal class 'dfm' [package "quanteda"] with 8 slots
..@ docvars : 'data.frame':   10 obs. of  3 variables:
...$ docname_ : chr [1:10] "text1" "text2" "text3" "text4" ...
...$ docid_   : Factor w/ 10 levels "text1","text2",...: 1 2 3 4 5 6 7 8 9 10
...$ segid_   : int [1:10] 1 1 1 1 1 1 1 1 1 1
> TarzanDataWeights
Document-feature matrix of: 10 documents, 8 features (90.00% sparse) and 0 docvars.
  features
docs  tarzan of the apes by edgar rice burroughs
text1    0  0  0  0  0  0  0  0
text2    1  1  1  1  0  0  0  0
text3    0  0  0  0  0  0  0  0
text4    0  0  0  0  0  0  0  0
text5    0  0  0  0  1  0  0  0
text6    0  0  0  0  0  0  0  0
[ reached max_ndoc ... 4 more documents ]

```

The term frequency-inverse document frequency (tf-idf) scores are computed based on the document-feature matrix TarzanDataDFM. The dfm_tfidf function from the quanteda package is utilized, specifying the term frequency (tf) scheme as "count" and the document frequency (df) scheme as "inverse". This function computes the tf-idf scores for each term in the matrix, which represent the importance of each term in the context of the corpus by considering both the frequency of occurrence in a document and the rarity of occurrence across the corpus. The resulting tf-idf weighted document-feature matrix is stored in TarzanDataTFIDF, the structure of which is displayed using str. This computation of tf-idf scores enables more nuanced analysis by accounting for both term frequency and document frequency across the corpus.

```

> # Computing the tf-idf score
> TarzanDataTFIDF <- quanteda::dfm_tfidf(TarzanDataDFM, scheme_tf = "count", scheme_df="inverse")
> str(TarzanDataTFIDF)
Formal class 'dfm' [package "quanteda"] with 8 slots
..@ docvars :data.frame':   10 obs. of  3 variables:
.. ..$ docname_ : chr [1:10] "text1" "text2" "text3" "text4" ...
.. ..$ docid_  : Factor w/ 10 levels "text1","text2",...: 1 2 3 4 5 6 7 8 9 10
.. ..$ segid_  : int [1:10] 1 1 1 1 1 1 1 1 1 1
..@ meta    :List of 3
.. ..$ system:List of 5
.. .. ..$ package-version:Classes 'package_version', 'numeric_version' hidden list of 1
.. .. .. ..$ : int [1:3] 4 0 2
.. .. ..$ r-version  :Classes 'R_system_version', 'package_version', 'numeric_version' hidden list of 1
.. .. .. ..$ : int [1:3] 4 4 0
.. .. ..$ system   : Named chr [1:3] "Darwin" "arm64" "yashwanth"
.. .. .. ..- attr(*, "names")= chr [1:3] "sysname" "machine" "user"
.. .. ..$ directory : chr "/Users/yashwanth/Documents/GWU/Sem 2/Intro to Big Data/Project/Project 3/R Files/Dataset"
.. .. ..$ created   : Date[1:1], format: "2024-05-03"
.. ..$ object:List of 10
.. .. ..$ unit      : chr "documents"
.. .. ..$ what      : chr "word"
.. .. ..$ tokenizer : chr "tokenize_word4"
.. .. ..$ ngram     : int 1
.. .. ..$ skip      : int 0
.. .. ..$ concatenator: chr "_"
.. .. ..$ weight_tf  :List of 3
.. .. .. ..$ scheme: chr "count"
.. .. .. ..$ base  : NULL
.. .. .. ..$ k    : NULL
.. .. ..$ weight_df  :List of 2
.. .. .. ..$ scheme: chr "inverse"
.. .. .. ..$ base  : num 10
.. .. ..$ smooth    : num 0
.. .. ..$ summary   :List of 2
.. .. .. ..$ hash: chr(0)
.. .. .. ..$ data: NULL
.. .. ..$ user     : list()
..@ i       : int [1:8] 1 1 1 1 4 6 6 6
..@ p       : int [1:9] 0 1 2 3 4 5 6 7 8
..@ Dim    : int [1:2] 10 8

```

Syuzhet

The sentiment analysis is conducted using the "syuzhet" method on the text data stored in the TS object. The `get_sentiment` function, presumably from the `syuzhet` package, is applied to `TS` to compute the sentiment scores for the text. The resulting sentiment scores, which represent the emotional polarity of the text, are stored in the `TSSentiment` object. This analysis allows for the quantification of sentiment in the text data, providing insights into the overall emotional tone conveyed by the text. The specific interpretation of the sentiment scores will depend on the methodology and scale used by the "syuzhet" method.

The sentiment of this work can be examined by the library `Syuzhet`, which is crucial for deciphering its underlying meaning. Import the library first, then divide it into sentences.

```
> # Sentiment using syuzhet method
> TSSentiment <- get_sentiment(TS, "syuzhet")
> TSSentiment
[1] 0.90 -0.25 1.10 2.40 0.10 2.60 1.25 1.15 -3.15 0.55 0.00 3.50 2.00 2.95 1.10 2.75 -0.35 -0.75
[19] 1.00 0.00 0.50 0.00 0.00 0.55 0.80 0.00 -2.25 -0.25 -0.25 0.40 -0.50 0.20 -0.15 0.00 -0.50 -1.00
[37] -0.75 -2.15 -0.50 -0.25 -1.05 -1.50 -1.85 -1.00 -1.20 1.00 1.00 -0.50 0.20 -1.05 -0.30 0.10 -0.85 -2.25
[55] -0.50 -0.40 -1.15 -1.75 0.25 1.10 -1.00 1.65 0.00 -0.25 0.50 1.05 -0.10 1.15 0.00 0.00 0.00 -0.80
[73] 0.80 -1.00 0.00 -1.00 0.40 -1.00 -0.50 0.00 0.05 0.00 0.40 0.75 -0.75 -1.00 -0.90 0.15 0.00 0.05
[91] 0.50 -1.00 -3.90 1.25 0.00 -0.25 -0.75 0.75 -0.50 -0.35 -0.25 0.00 -0.55 0.40 -0.60 0.80 0.80 -1.75
[109] -0.50 0.00 -2.65 1.00 0.00 -1.65 1.15 -0.10 -1.45 0.25 1.60 -0.10 0.25 0.85 0.50 -0.50 -0.90 -0.90
[127] -0.75 -0.50 0.80 -0.55 0.00 0.00 1.15 -2.65 0.80 0.00 0.60 0.50 -0.50 0.00 1.30 -0.50 0.80 0.80
[145] 0.00 -1.40 -0.75 -3.00 0.50 -0.25 -0.55 -1.00 -0.50 0.40 -0.40 -0.50 -0.20 -0.85 -0.40 -2.80 -3.20 -0.50
[163] -2.00 -1.00 -0.60 -2.45 0.00 0.00 0.00 -0.75 -0.75 0.50 2.45 0.00 0.00 0.90 -0.50 -2.10 -0.75 0.55
[181] -1.05 0.00 0.00 -0.75 0.55 -2.15 0.65 -0.75 -0.25 1.40 0.00 0.35 -1.50 -2.00 0.75 0.50 0.80 -0.10
[199] 0.75 2.55 -0.50 -2.25 -4.10 0.50 2.95 0.10 0.40 0.80 1.00 0.55 1.35 0.50 0.80 -1.00 -0.85 -0.75
```

Now we conduct sentiment analysis using the "bing" method. The `get_sentiment` function, likely from a sentiment analysis package such as `syuzhet` or `textdata`, is employed with the "bing" method to compute sentiment scores for the text. The "bing" method typically utilizes the Bing Liu lexicon, which contains lists of positive and negative words, to determine the sentiment of the text. The resulting sentiment scores, which indicate the presence of positive or negative sentiment in the text, are stored in the `TSBing` object. This analysis provides insight into the emotional polarity of the text based on the presence of words with positive or negative connotations.

```
> # Sentiment using bing method
> TSBing <- get_sentiment(TS, "bing")
> TSBing
[1] -3 -2 1 1 0 3 -2 3 -3 1 0 5 3 2 0 3 -2 -1 0 1 1 0 0 -1 0 0 -3 -1 -1 0 -1 0 0 0 -1 -1 -1
[38] -3 -1 0 -2 -2 0 -2 0 0 -2 1 -1 0 -1 0 -3 -1 -1 -2 -2 0 1 -1 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0
[75] 0 0 0 -1 0 1 0 0 1 0 -1 -1 1 -1 0 0 -1 -4 1 0 -1 -1 1 -1 -1 1 0 -1 0 0 1 1 -1 0 -1 0 0 1 1 -1 -1 0 -2
[112] 0 0 -1 1 1 -3 -1 1 -1 1 -1 1 -1 2 0 -1 0 0 -1 0 0 1 -3 1 0 1 1 -1 0 1 0 0 0 0 1 -1 -2 -2 1 -1 0 0 -1 -2 -4
[149] 2 0 -1 -2 -1 -1 0 -1 -1 0 0 -2 -1 -1 -3 -1 0 -4 0 0 0 -1 -1 0 0 0 0 1 -1 -2 -2 1 -1 0 0 -1 1
[186] -2 0 -1 0 1 0 -1 -3 -1 1 2 1 1 1 3 -1 -4 -6 1 2 0 2 0 1 2 3 1 1 -1 -1 0 0 1 -1 -3 -1 0
[223] 2 0 0 1 -2 -1 1 1 -3 -1 -1 1 0 3 0 0 0 1 1 -2 1 0 1 -2 1 0 2 2 -2 1 2 1 0 2 1 0 2 1 0 2
[260] 1 1 -1 4 -1 1 1 0 0 1 -1 0 -2 -2 -1 2 1 0 0 -2 1 5 -1 1 2 0 0 0 0 0 0 0 0 2 -1 1 2 -4
[297] 1 2 -1 -2 -1 -2 2 0 -4 -2 0 -1 0 -5 0 -1 -2 1 -3 0 -1 0 0 -1 -1 -3 -2 1 1 -3 -1 -2 2 -1 0 -2
[334] 0 -1 -4 4 -3 0 3 0 0 0 0 0 2 -1 -2 -2 0 1 -1 0 0 0 -4 -9 -1 -1 -2 -1 -1 -2 -2 -3 -3 0 -5 -1 -2
[371] 0 -1 0 1 -3 1 0 -2 4 1 0 -2 0 -2 0 0 -1 -1 1 -1 0 -5 -4 -2 0 -2 -2 0 -2 -2 0 -1 -1 -3 4 1 -2
[408] 0 -1 -4 -4 0 -1 -1 -1 0 -1 1 -2 0 -1 0 -2 0 -3 -2 0 0 0 1 0 -1 -4 -1 -2 -1 2 0 0 -3 3 -1 -1 0
[445] -1 -1 1 1 1 -1 1 -1 0 0 0 3 2 0 1 -1 5 2 0 -1 1 -2 -1 -3 1 0 -2 -1 0 0 -1 3 0 1 1 0 1
[482] -1 -1 -1 1 0 1 2 2 0 1 -2 1 -7 1 -1 -3 -1 1 1 0 2 -1 -2 -2 -3 -1 -3 2 -1 1 -1 0 -2 0 1 -1 0
[519] 2 1 -2 0 1 -1 0 -1 2 -2 0 -2 1 -1 -1 1 -1 0 0 -1 0 -3 1 0 2 0 2 0 1 0 0 2 1 0 -1 0
[556] 1 -3 0 1 -2 0 -1 -3 -3 0 -2 0 1 0 -1 0 0 0 0 0 -5 0 0 -1 -1 1 1 -1 -2 0 0 -1 -2 1 2 1
[593] -1 -1 -2 -3 1 1 -2 -3 -2 0 -1 2 -6 1 3 -1 -1 -4 -3 -1 -2 -1 0 -3 0 -2 -1 -4 0 0 -2 3 2 2 0
[630] 0 -3 4 -2 0 0 0 -1 1 1 3 -7 -1 -1 0 -1 -1 0 -1 -3 -1 0 3 0 -1 -2 0 0 0 -1 3 0 2 0 -2 1 0
[667] 0 2 3 0 -1 -3 1 -3 3 -2 -1 0 0 -7 1 2 0 0 0 -2 -2 0 0 -2 0 1 0 0 -1 0 -1 -3 -1 -1 -3
[704] -4 -1 -1 0 -1 -2 -3 -2 1 -4 -3 -1 -2 1 0 -1 0 1 -2 2 -1 -1 1 0 -6 -1 0 0 1 1 -3 -1 -2 -1 0 2 -2
[741] 0 0 0 0 0 1 0 0 -2 -5 -1 0 0 0 1 1 -2 -3 0 -2 0 0 -2 0 1 -1 0 -6 -2 -1 1 1 -1 -1 1 1 1
[778] -1 -1 1 -3 0 -1 -1 1 1 0 0 -1 -4 0 -1 -3 -4 -2 -4 0 0 1 0 -2 -1 0 0 -2 -1 0 -1 1 -1 0 2 0 -1
[815] 1 0 2 0 2 -1 0 1 1 0 -1 -1 0 -2 2 -1 1 -1 -2 -1 -2 0 -1 0 2 -1 1 -1 -2 1 0 0 0 0 0 2 -1
[852] 0 0 0 -1 0 -1 0 -1 -1 -2 0 0 1 1 0 -5 -2 0 -2 -2 2 -2 0 -3 -1 1 -2 0 0 0 0 -1 -4 -2 0 0 0
[889] 0 0 0 1 0 2 0 -1 0 -3 -3 -1 1 -1 -2 -2 -1 -1 -3 -1 -2 -1 2 -1 1 2 2 -4 1 -2 0 -2 -2 -1 1 1
[926] 0 0 0 -2 1 0 -1 1 1 -1 0 1 0 -1 2 0 -2 1 0 -2 -1 0 -3 -1 -2 0 -1 1 0 0 0 -1 0 -1 -1 1
[963] -4 -3 0 1 0 -1 0 -2 -1 0 1 1 2 0 0 1 -2 0 -1 -1 1 -3 0 0 0 -1 1 0 2 -1 -1 -1 -1 1 0 -2
[1000] 0
[ reached getOption("max.print") -- omitted 3312 entries ]
```

For the "syuzhet" method, the `get_sentiment_dictionary()` function is used to obtain the sentiment dictionary. This dictionary likely contains lists of words categorized as positive, negative, or neutral, along with their associated sentiment scores.

Similarly, for the "bing" method, another sentiment dictionary is retrieved using the same function `get_sentiment_dictionary()`. This dictionary is specific to the "bing" method and is tailored for sentiment analysis using the Bing Liu lexicon, which consists of lists of positive and negative words.

These sentiment dictionaries are essential resources for sentiment analysis algorithms, as they provide the basis for determining the sentiment of text data by matching words from the text against entries in the dictionary and assigning sentiment scores accordingly.

> `TSDictionary`

	word	value
1	abandon	-0.75
2	abandoned	-0.50
3	abandoner	-0.25
4	abandonment	-0.25
5	abandons	-1.00
6	abducted	-1.00
7	abduction	-0.50
8	abductions	-1.00

> `TSDictionaryBing`

	word	value
1	abandon	-0.75
2	abandoned	-0.50
3	abandoner	-0.25
4	abandonment	-0.25
5	abandons	-1.00
6	abducted	-1.00
7	abduction	-0.50
8	abductions	-1.00

For the "syuzhet" method, the sentiment scores stored in TSSentiment are summed using the sum function, resulting in TSSum. This value represents the overall emotional valence of the text based on the sentiment scores calculated by the "syuzhet" method. Similarly, for the "bing" method, the sentiment scores stored in TSBing are summed using the sum function, yielding TSbingSum. This value indicates the overall emotional valence of the text according to the sentiment scores computed by the "bing" method. These overall emotional valence scores provide a quantitative measure of the emotional tone conveyed by the text data, with positive values indicating a positive emotional valence, negative values indicating a negative emotional valence, and values close to zero suggesting a more neutral emotional tone.

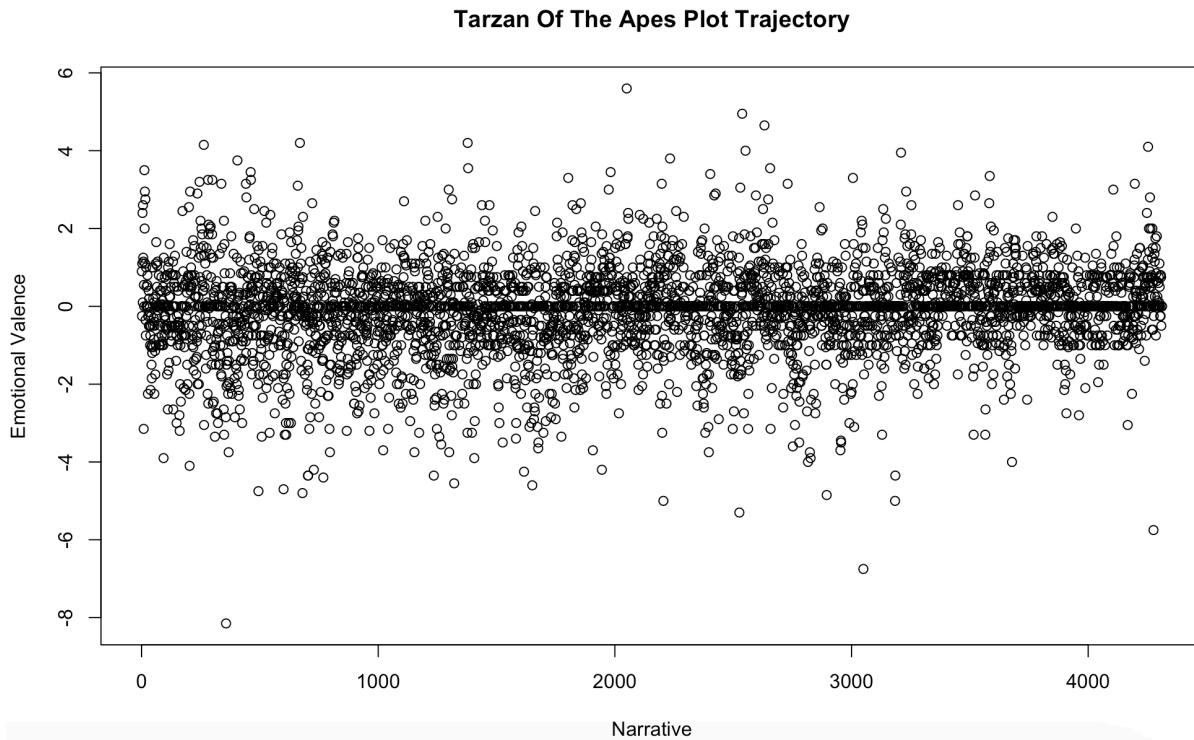
```
> # Overall emotional Valence in text for syuzhet
> TSSum <- sum(TSSentiment)
> TSSum
[1] -469.75
> # Overall emotional Valence in text for Bing
> TSbingSum <- sum(TSBing)
> TSbingSum
[1] -1183
```

For the "syuzhet" method, the mean of the sentiment scores stored in TSSentiment is computed using the mean function, resulting in TSmean. This value represents the average emotional valence of the text based on the sentiment scores calculated by the "syuzhet" method. Similarly, for the "bing" method, the mean of the sentiment scores stored in TSBing is calculated using the mean function, yielding TSbingMean. This value indicates the average emotional valence of the text according to the sentiment scores computed by the "bing" method. These mean emotional valence scores provide insights into the overall emotional tone of the text data, with positive values indicating a predominantly positive emotional tone, negative values suggesting a predominantly negative emotional tone, and values close to zero indicating a more neutral emotional expression.

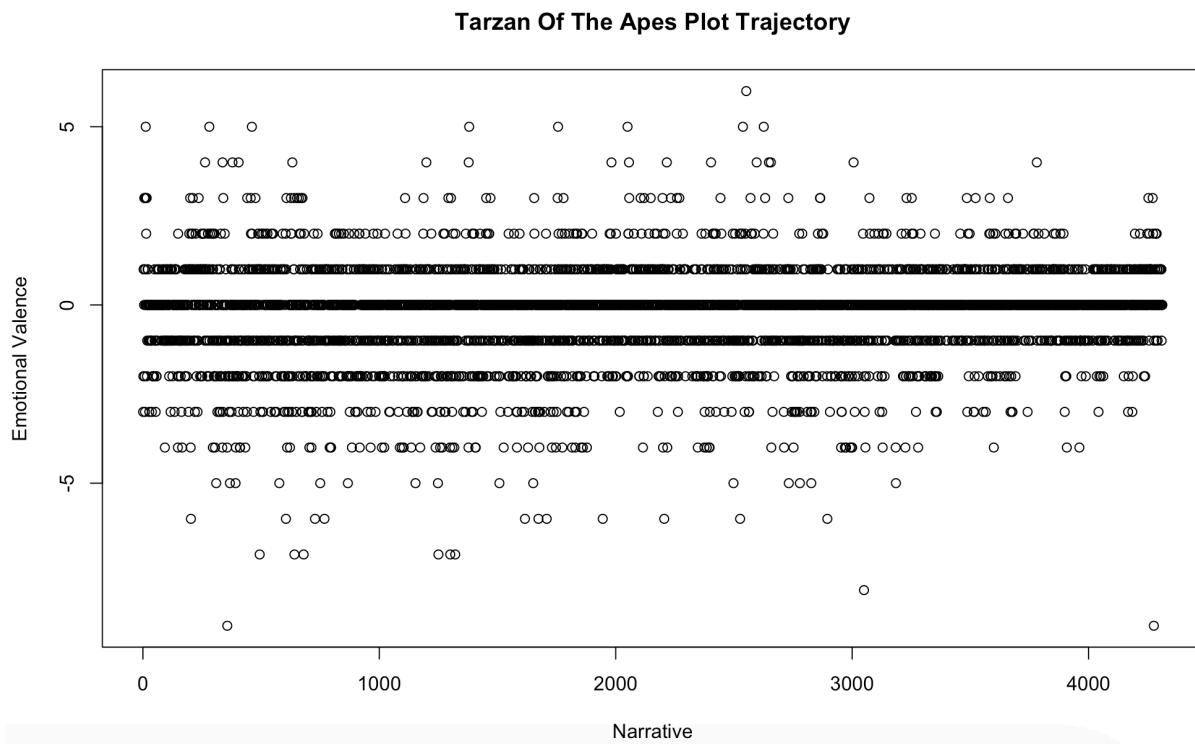
```
> # Mean of the emotional valence for syuzhet
> TSmean <- mean(TSSentiment)
> TSmean
[1] -0.1089402
> # Mean of the emotional valence for Bing
> TSbingMean <- mean(TSBing)
> TSbingMean
[1] -0.2743506
```

We are going to generate two plots to visualize the emotional trajectory of the text data using sentiment scores obtained from the "syuzhet" and "bing" methods.

For the "syuzhet" method, the plot function is used to create a plot of the sentiment scores stored in TSSentiment. The main title of the plot is set as "Tarzan Of The Apes Plot Trajectory", with the x-axis labeled as "Narrative" and the y-axis labeled as "Emotional Valence". This plot displays the emotional trajectory of the text data based on the sentiment scores calculated by the "syuzhet" method.



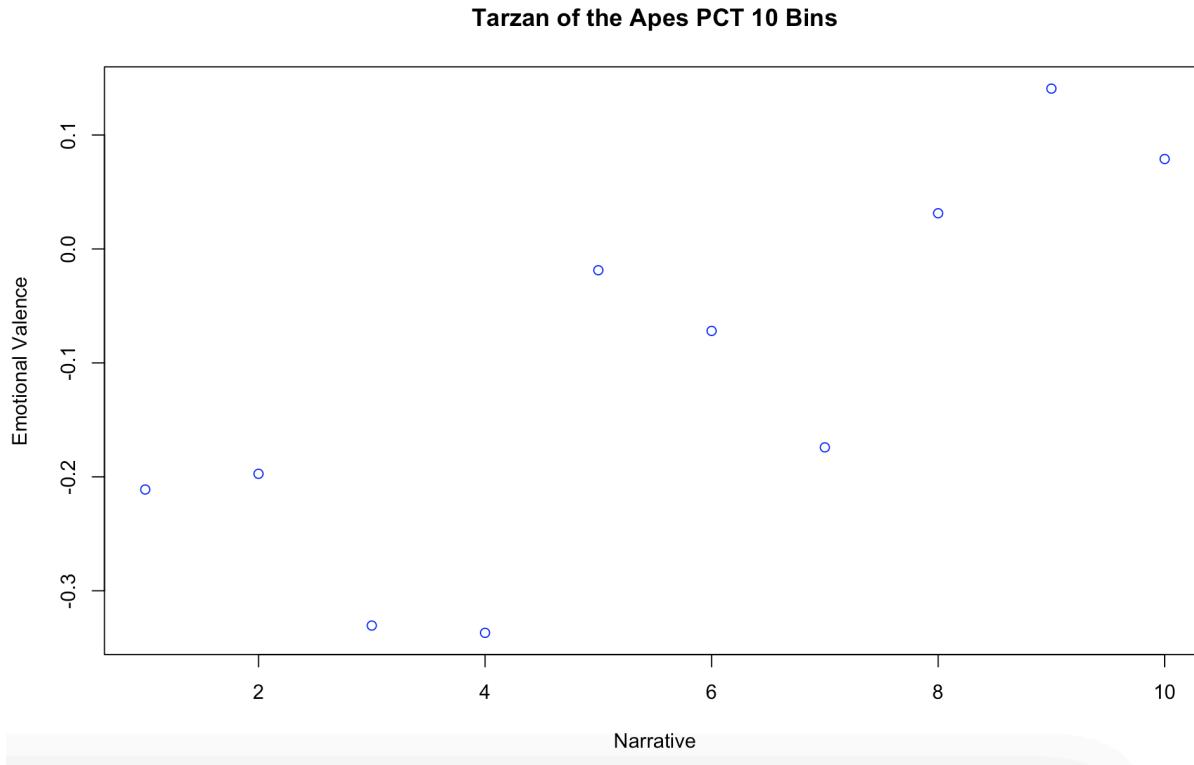
Similarly, for the "bing" method, another plot is generated using the plot function with the sentiment scores stored in TSBing. The main title, x-axis label, and y-axis label are set the same as in the "syuzhet" plot. This plot visualizes the emotional trajectory of the text data based on the sentiment scores computed by the "bing" method.



The sentiment scores obtained from the "syuzhet" method, stored in TSSentiment, are divided into 10 bins using the get_percentage_values function. The resulting percentages of sentiment values within each bin are stored in TSSentimentPctValue. The structure of TSSentimentPctValue is examined to understand its format.

```
> # PCT Value = 10 bins
> TSSentimentPctValue <- get_percentage_values(TSSentiment,bins = 10)
> structure(TSSentimentPctValue)
  1      2      3      4      5      6      7      8      9      10 
-0.21111111 -0.19733179 -0.33051044 -0.33689095 -0.01867749 -0.07192575 -0.17412993  0.03132251  0.14071926  0.07893519
> # PCT 10 Bins
> plot(TSSentimentPctValue, main = "Tarzan of the Apes PCT 10 Bins", xlab = "Narrative", ylab = "Emotional Valence", col="blue")
```

After calculating the percentage values in 10 bins, a plot is generated using the plot function to visualize the distribution of sentiment scores across the narrative. The main title of the plot is set as "Tarzan of the Apes PCT 10 Bins", with the x-axis labeled as "Narrative" and the y-axis labeled as "Emotional Valence". Additionally, the color of the plot is set to blue using the col parameter for better visualization.

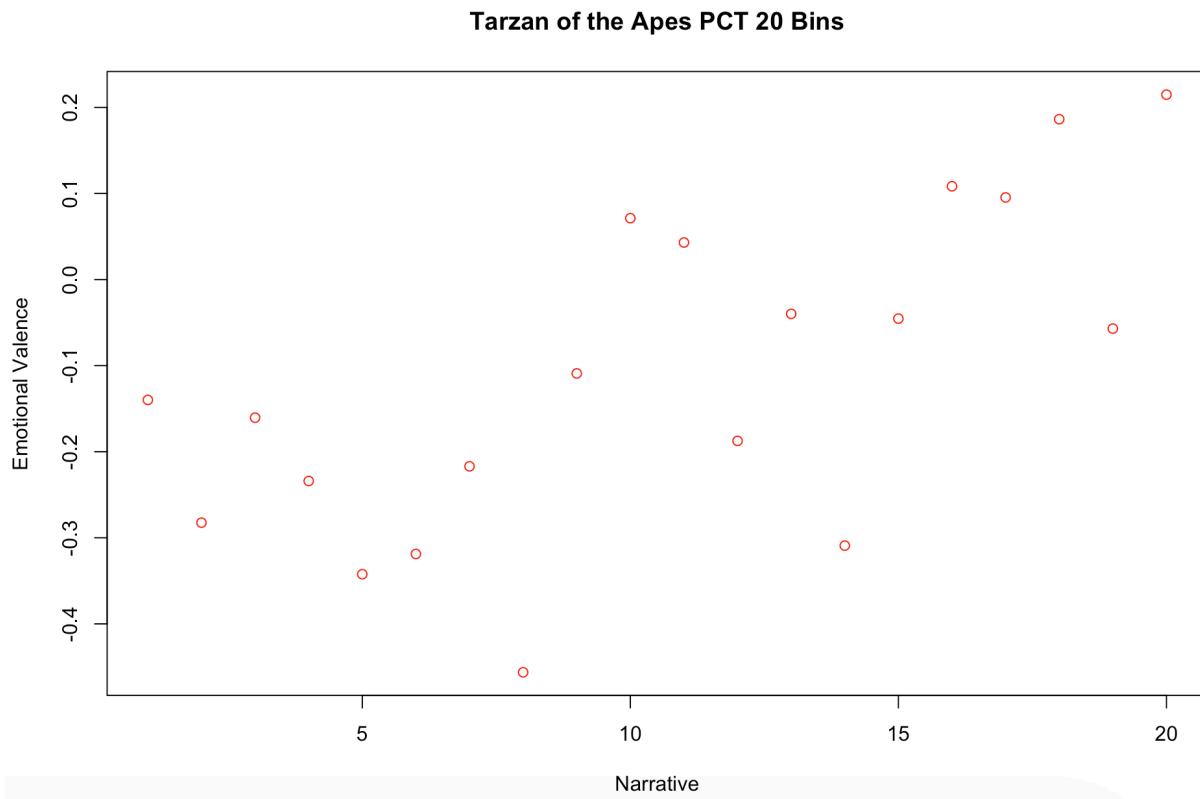


The sentiment scores obtained from the "syuzhet" method, stored in TSSentiment, are divided into 20 bins using the get_percentage_values function. Using more bins (20 instead of 10) allows for a finer-grained analysis of the sentiment distribution. The resulting percentages of sentiment values within each bin are stored in TSSentimentPctValue. The structure of TSSentimentPctValue is examined to understand its format.

```
> # PCT Value = 20 bins
> TSSentimentPctValue <- get_percentage_values(TSSentiment, bins = 20)
> structure(TSSentimentPctValue)
   1      2      3      4      5      6      7      8      9
-0.13981481 -0.28240741 -0.16046512 -0.23402778 -0.34232558 -0.31875000 -0.21697674 -0.45625000 -0.10906977
   10     11     12     13     14     15     16     17     18
  0.07129630  0.04305556 -0.18744186 -0.03981481 -0.30906977 -0.04537037  0.10837209  0.09537037  0.18627907
   19     20
 -0.05694444  0.21481481
> # PCT 20 Bins
> plot(TSSentimentPctValue, main = "Tarzan of the Apes PCT 20 Bins", xlab = "Narrative", ylab = "Emotional Valence", col = "red")
```

After calculating the percentage values in 20 bins, a plot is generated using the plot function to visualize the distribution of sentiment scores across the narrative. The main title of the plot is set as "Tarzan of the Apes PCT 20 Bins", with the x-axis

labeled as "Narrative" and the y-axis labeled as "Emotional Valence". Additionally, the color of the plot is set to red using the col parameter for better visualization.



Discussion:

We discussed about word clouds, sentiment analysis, and preprocessing as useful methods for text analysis. Word clouds offered a far more useful representation of word frequency than only examining document term frequency tables. Just as fascinating is the study of sentiment. As we've shown, while various methods of assessing emotion may yield radically divergent results, they all still highlight significant aspects of the text. Furthermore, to understand the text, it is beneficial to take into account the linear structure of the feeling across time rather than concentrating on a single global mean or sum value. The structure was agreed upon by all methods, notwithstanding their inability to agree on the exact emotion values.

The text must be properly preprocessed as well. The results were clearer when punctuation was removed, words were lowercased, and stemming was applied. In

particular, stemming proved useful. The word cloud produced after stemming seemed to be a better representation of the text since it better conveyed the meaning of the words. Furthermore, we found that certain visualization styles can be difficult or less useful for text analysis, including the cluster dendrogram, for which we were unable to remove enough words to create a plot that was usable. This at least illustrated the intricacy of large paragraphs in a very sparse manner.

Methods Used:

Here we just restate the important methods we used for analyzing the text from each library. This is not comprehensive, but you can see all the methods used throughout this write-up anyways.

tm

- VCorpus
- tm_map
- DocumentTermMatrix
- findFreqTerms

quanteda

- docfreq
- dfm_weight
- dfm_tfidf
- tokens_wordstem
- textplot_wordcloud

syuzhet

- get_sentences
- get_sentiment
- get_sentiment_dictionary

wordcloud

- wordcloud