

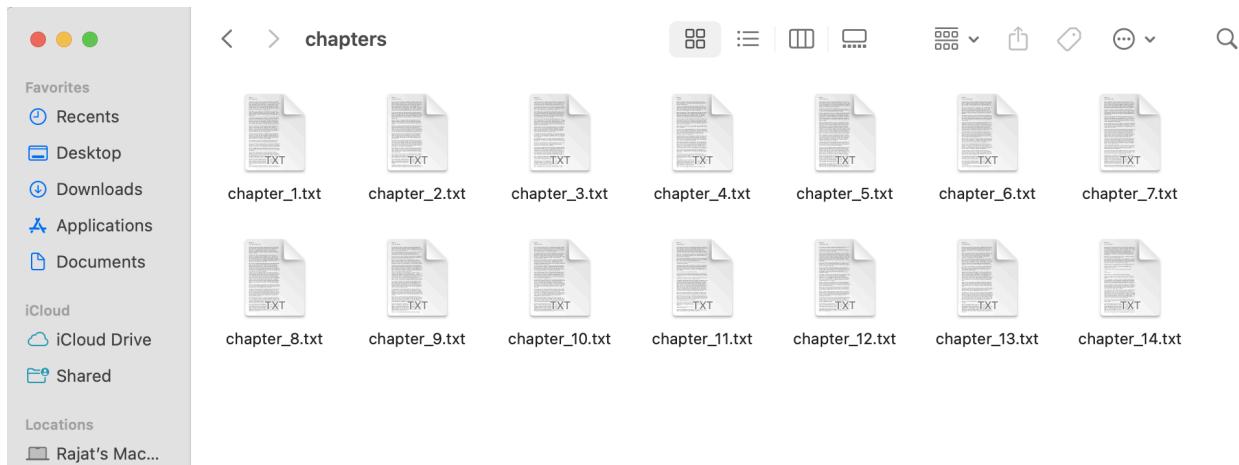
CSCI 6444: Big Data and Analytics

Class Project-3

Introduction to Text Analytics

1. i. Separating chapters into individual text files

```
> book <- readLines("APrincessOfMars.txt")
> chapter_starts <- grep("^CHAPTER [IVXLCDM]+", book)
> dir.create("chapters", showWarnings = FALSE)
> num_chapters_to_write <- 14
> for (i in seq_len(num_chapters_to_write))
+ {
+   start <- chapter_starts[i]
+
+   if (i < length(chapter_starts)) {
+     end <- chapter_starts[i + 1] - 1
+   }
+   else {
+     end <- length(book)
+   }
+
+   # Extract the chapter text
+   chapter_text <- book[start:end]
+
+   # Write the chapter to a file
+   chapter_filename <- paste0("chapters/chapter_", i, ".txt")
+   writeLines(chapter_text, chapter_filename)
+ }
```



We begin by reading the text file and splitting it into separate files for each chapter (I to IV). This makes it easier to analyze each chapter individually.

ii. Creating a VCorpus

```

> princess<-VCorpus(DirSource(".",ignore.case = TRUE,mode = "text"))
> str(princess)
Classes 'VCorpus', 'Corpus'  hidden list of 3
$ content:List of 14
..$ :List of 2
...$ content: chr [1:268] "CHAPTER I" "" "ON THE ARIZONA HILLS" ""
...$ meta  :List of 7
...$ author      : chr(0)
...$ timestamp: POSIXlt[1:1], format: "2025-04-26 19:58:16"
...$ description : chr(0)
...$ heading     : chr(0)
...$ id          : chr "chapter_1.txt"
...$ language    : chr "en"
...$ origin      : chr(0)
...- attr(*, "class")= chr "TextDocumentMeta"
...- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
..$ :List of 2
...$ content: chr [1:359] "CHAPTER X" "" "CHAMPION AND CHIEF" ""
...$ meta  :List of 7
...$ author      : chr(0)
...$ timestamp: POSIXlt[1:1], format: "2025-04-26 19:58:16"
...$ description : chr(0)
...$ heading     : chr(0)
...$ id          : chr "chapter_10.txt"
...$ language    : chr "en"
...$ origin      : chr(0)
...- attr(*, "class")= chr "TextDocumentMeta"
...- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"

```

```

> princess
<<VCorpus>>
Metadata: corpus specific: 0, document level (indexed): 0
Content: documents: 14

```

VCorpus() creates a volatile corpus, this reads all text files in the current folder and stores them as a text corpus named '*princess*' for further text analysis.

iii. Extracting text from the VCorpus

```

> ptext<-princess[[1]]
> ptext
<<PlainTextDocument>>
Metadata: 7
Content: chars: 13971
> ptext[1]
$content
[1] "CHAPTER I"
[2] ""
[3] "ON THE ARIZONA HILLS"
[4] ""
[5] ""
[6] "I am a very old man; how old I do not know. Possibly I am a hundred,"
[7] "possibly more; but I cannot tell because I have never aged as other"
[8] "men, nor do I remember any childhood. So far as I can recollect I have"
[9] "always been a man, a man of about thirty. I appear today as I did"
[10] "forty years and more ago, and yet I feel that I cannot go on living"
[11] "forever; that some day I shall die the real death from which there is"
[12] "no resurrection. I do not know why I should fear death, I who have"
[13] "died twice and am still alive; but yet I have the same horror of it as"
[14] "you who have never died, and it is because of this terror of death, I"
[15] "believe, that I am so convinced of my mortality."
[16] ""

```

We extract the first chapter from the corpus '*princess*' and access its first line of text using *ptext[1]*.

a. 10 longest words and 10 longest sentences in each chapter.

Note: The extracted sentences are too long to be displayed neatly in a table, to make them more readable and preserve the full text, we have included screenshots of the extracted analysis for each chapter.

```

> chapter_dir <- "chapters"
> chapter_files <- list.files(chapter_dir, full.names = TRUE)[1:14]
> dir.create("chapter_analysis", showWarnings = FALSE)
> for (i in seq_along(chapter_files)) {
+
+   chapter_text <- readLines(chapter_files[i])
+
+   chapter_full_text <- paste(chapter_text, collapse = " ")
+
+   words <- unlist(strsplit(chapter_full_text, "\\W+"))
+
+   words <- words[words != ""]
+
+   longest_words <- words[order(nchar(words), decreasing = TRUE)][1:10]
+
+   sentences <- unlist(strsplit(chapter_full_text, "(?<=[.!?])\\s+", perl=TRUE))
+
+   longest_sentences <- sentences[order(nchar(sentences), decreasing = TRUE)][1:10]
+
+   output_filename <- paste0("chapter_analysis/chapter_", i, "_analysis.txt")
+
+   cat("Top 10 Longest Words:\n", file = output_filename)
+   cat(paste(longest_words, collapse = "\n"), file = output_filename, append = TRUE)
+
+   cat("\n\nTop 10 Longest Sentences:\n", file = output_filename, append = TRUE)
+   cat(paste(longest_sentences, collapse = "\n\n"), file = output_filename, append = TRUE)
+
+ }

```

This script reads the first 14 chapter files, extracts the top 10 longest words and sentences from each, and saves the analysis to separate text files in a new '*chapter_analysis*' folder.

i. Chapter-1

```

Top 10 Longest Words:
subconsciously
characteristic
understanding
comparatively
sensitiveness
consternation
perpendicular
resuscitation
comparatively
resurrection

Top 10 Longest Sentences:
However, I am not prone to sensitiveness, and the following of a sense of duty, wherever it may lead, has always been a kind of fetish with me throughout my life; which may account for the honors bestowed upon me by three republics and the decorations and friendships of an old and powerful emperor and several lesser kings, in whose service my sword has been red many a time.

The fact that it is difficult to aim anything but imprecisions accurately by moonlight, that they were upset by the sudden and unexpected manner of my advent, and that I was a rather rapidly moving target saved me from the various deadly projectiles of the enemy and permitted me to reach the shadows of the surrounding peaks before an orderly pursuit could be organized.

Since we had entered the territory we had not seen a hostile Indian, and we had, therefore, become careless in the extreme, and were wont to ridicule the stories we had heard of the great numbers of these vicious marauders that were supposed to haunt the trails, taking their toll in lives and torture of every white party which fell into their merciless clutches.

In this instance I was, of course, positive that Powell was the center of attraction, but whether I thought or acted first I do not know, but within an instant from the moment the scene broke upon my view I had whipped out my revolvers and was charging down upon the entire army of warriors, shooting rapidly, and whooping at the top of my lungs.

The morning of Powell's departure was, like nearly all Arizona mornings, clear and beautiful; I could see him and his little pack animals picking their way down the mountainside toward the valley, and all during the morning I would catch occasional glimpses of them as they topped a hog back or came out upon a level plateau.

My horse was traveling practically unguided as I knew that I had probably less knowledge of the exact location of the trail to the pass than he, and thus it happened that he entered a defile which led to the summit of the range and not to the pass which I had hoped would carry me to the valley and to safety.

I do not believe that I am made of the stuff which constitutes heroes, because, in all of the hundreds of instances that my voluntary acts have placed me face to face with death, I cannot recall a single one where any alternative step to that I took occurred to me until many hours later.

I was positive now that the trailers were Apaches and that they wished to capture Powell alive for the fiendish pleasure of the torture, so I urged my horse onward at a most dangerous pace, hoping against hope that I would catch up with the red rascals before they attacked him.

I soon became so drowsy that I could scarcely resist the strong desire to throw myself on the floor of the cave for a few moments' rest, but I knew that this would never do, as it would mean certain death at the hands of my red friends, who might be upon me at any moment.

I know that the average human mind will not believe what it cannot grasp, and so I do not purpose being pilloried by the public, the pulpit, and the press, and held up as a colossal liar when I am but telling the simple truths which some day science will substantiate.

```

ii. Chapter-2

Top 10 Longest Words:
responsibilities
characteristics
companionship
manifestation
ludicrousness
precipitately
authoritative
understanding
theoretically
companionship

Top 10 Longest Sentences:

I understand that you belittle all sentiments of generosity and kindness, but I do not, and I can convince your most doughty warrior that these characteristics are not incompatible with an ability to fight." "Ordinarily I am not given to long speeches, nor ever before had I descended to bombast, but I had guessed at the keynote which would strike an answering chord in the breasts of the green Martians, nor was I wrong, for my harangue evidently deeply impressed them, and their attitude toward me thereafter was still further respectful.

Where your country?" "Yes, Dejah Thoris, I too am a prisoner; my name is John Carter, and I claim Virginia, one of the United States of America, Earth, as my home; but why I am permitted to wear arms I do not know, nor was I aware that my regalia was that of a chieftain." We were interrupted at this juncture by the approach of one of the warriors, bearing arms, accoutrements and ornaments, and in a flash one of her questions was answered and a puzzle cleared up for me.

What words of moment were to have fallen from his lips were never spoken, as just then a young warrior, evidently sensing the trend of thought among the older men, leaped down from the steps of the rostrum, and striking the frail captive a powerful blow across the face, which felled her to the floor, placed his foot upon her prostrate form and turning toward the assembled council broke into peals of horrid, mirthless laughter.

Numerous brilliantly colored and strangely formed wild flowers dotted the ravines and from the summit of the first hill I saw still other hills stretching off toward the north, and rising, one range above another, until lost in mountains of quite respectable dimensions; though I afterward found that only a few peaks on all Mars exceed four thousand feet in height; the suggestion of magnitude was merely relative.

I saw that the body of my dead antagonist had been stripped, and I read in the menacing yet respectful attitude of the warrior who had brought me these trophies of the kill the same demeanor as that evinced by the other who had brought me my original equipment, and now for the first time I realized that my blow, on the occasion of my first battle in the audience chamber had resulted in the death of my adversary.

Realizing that I was a somewhat favored character, and also convinced that the warriors did not know of my proficiency in their language, as I had plead with Sola to keep this a secret on the grounds that I did not wish to be forced to talk with the men until I had perfectly mastered the Martian tongue, I chanced an attempt to enter the audience chamber and listen to the proceedings.

Tars Tarkas himself seemed pleased with my reply, but his only comment was more or less enigmatical—"And I think I know Tal Halus, Jeddak of Thark." I now turned my attention to Dejah Thoris, and assisting her to her feet I turned with her toward the exit, ignoring her hovering guardian harpies as well as the inquiring glances of the chieftains.

The reason for the whole attitude displayed toward me was now apparent; I had won my spurs, so to speak, and in the crude justice, which always marks Martian dealings, and which, among other things, has caused me to call her the planet of paradoxes, I was accorded the honors due a conqueror; the trappings and the position of the man I killed.

I could not resist the ludicrousness of the spectacle, and holding my sides I rocked back and forth in the first laughter which had passed my lips in many days; the first, in fact, since the morning Powell had left camp when his horse, long unused, had precipitately and unexpectedly bucked him off head foremost into a pot of frijoles.

Springing upward, I struck him full in the face as he turned at my warning cry and then as he drew his short-sword I drew mine and sprang up again upon his breast, hooking one leg over the butt of his pistol and grasping one of his huge tusks with my left hand while I delivered blow after blow upon his enormous chest.

iii. Chapter-3

Top 10 Longest Words:
accoutrements
circumstances
questioning
eavesdropping
comparatively
internarying
irretrievably
architectural
compositions
conversation

Top 10 Longest Sentences:

"And wherefore, then, would your prisoner escape should you leave her, unless it was to follow you and crave your protection, and ask your pardon for the cruel thoughts she has harbored against you these past few days?" "You are right," I answered, "there is no escape for either of us unless we go together." "I heard your challenge to the creature you call Tars Tarkas, and I think I understand your position among these people, but what I cannot fathom is your statement that you are not of Barsom." "In the name of my first ancestor, then," she continued, "where may you be from?"

During the ages of hardships and incessant warring between their own various races, as well as with the green men, and before they had fitted themselves to the changed conditions, much of the high civilization and many of the arts of the fair-haired Martians had become lost; but the red race of today has reached a point where it feels that it has made up in new discoveries and in a more practical civilization for all that lies irretrievably buried with the ancient Barsomians, beneath the countless intervening ages.

"Because, John Carter," she replied, "nearly every planet and star having atmospheric conditions at all approaching those of Barsom, shows forms of animal life almost identical with you and me; and, further, Earth men, almost without exception, cover their bodies with strange, unsightly pieces of cloth, and their heads with hideous contraptions the purpose of which we have been unable to conceive; while you, when found by the Tharkan warriors, were entirely undressed and unadorned."

A similar wave of feeling seemed to stir her; she drew away from me with a sigh, and with her earnest, beautiful face turned up to mine, she whispered: "I believe you, John Carter; I do not know what a 'gentleman' is, nor have I ever heard before of Virginians but on Barsom no man lies; if he does not wish to speak the truth he is silent."

"The fact that you wore no ornaments is a strong proof of your un-Barsomian origin, while the absence of grotesque coverings might cause a doubt as to your earthliness." I then narrated the details of my departure from the Earth, explaining that my body there lay fully clothed in all the, to her, strange garments of mundane dwellers.

I can readily perceive that you are not of the Barsom of today; you are like us, yet different—but why should I trouble my poor head with such a problem, when my heart tells me that I believe because I wish to believe!" It was good logic, good, earthly, feminine logic, and if it satisfied her I certainly could pick no flaws in it.

The shores of the ancient seas were dotted with just such cities, and lesser ones, in diminishing numbers, were to be found converging toward the center of the oceans, as the people had found it necessary to follow the receding waters until necessity had forced upon them their ultimate salvation, the so-called Martian canals.

These three great divisions of the higher Martians had been forced into a mighty alliance as the drying up of the Martian seas had compelled them to seek the comparatively few and always diminishing fertile areas, and to defend themselves, under new conditions of life, against the wild hordes of green men.

They would kill you horribly anywhere upon the surface of Barsom if that were true; tell me it is not!" Her eyes were filled with a strange, weird light; her voice was pleading, and her little hands, reached up upon my breast, were pressed against me as though to wring a denial from my very heart.

These ancient Martians had been a highly cultivated and literary race, but during the vicissitudes of those trying centuries of readjustment to new conditions, not only did their advancement and production cease entirely, but practically all their archives, records, and literature were lost.

iv. Chapter-4

Top 10 Longest Words:
personification
humanitarianism
characteristics
responsibility
responsibility
granddaughter
investigation
communication
corresponding
conversation

Top 10 Longest Sentences:

Well, as you have said, your ways are not our ways, but you can fight well enough to do about as you please, and so, if you wish to give your woman to a captive, it is your own affair; but as a chieftain you should have those to serve you, and in accordance with our customs you may select any or all the females from the retinues of the chieftains whose metal you now wear." I thanked him, but assured him that I could get along very nicely without assistance except in the matter of preparing food, and so he promised to send women to me for this purpose and also for the care of my arms and the manufacture of my ammunition, which he said would be necessary.

While the court was entirely overgrown with the yellow, moss-like vegetation which blankets practically the entire surface of Mars, yet numerous fountains, statuary, benches, and pergola-like contraptions bore witness to the beauty which the court must have presented in bygone times, when graced by the fair-haired, laughing people whom stern and unalterable cosmic laws had driven not only from their homes, but from all except the vague legends of their descendants.

I finally chose a front room on the third floor, because this brought me nearer to Dejah Thoris, whose apartment was on the second floor of the adjoining building, and it flashed upon me that I could rig up some means of communication whereby she might signal me in case she needed either my services or my protection.

"But," he continued, in his fierce guttural tones, "if you run off with the red girl it is I who shall have to account to Tal Halus; it is I who shall have to face Tars Tarkas, and either demonstrate my right to command, or the metal from my dead carcass will go to a better man, for such is the custom of the Tharks."

In theory it may sound well, as is often the case with theories, but the results of ages of this unnatural practice, coupled with the community interest in the offspring being held paramount to that of the mother, is shown in the cold, cruel creatures, and their gloomy, loveless, mirthless existence.

The women and children of a man's retinue may be likened to a military unit for which he is responsible in various ways, as in matters of instruction, discipline, sustenance, and the exigencies of their continual roamings and their unending strife with other communities and with the red Martians.

The windows of the back rooms overlooked an enormous court, which formed the center of the square made by the buildings which faced the four contiguous streets, and which was now given over to the quartering of the various animals belonging to the warriors occupying the adjoining buildings.

CHAPTER XIII A PRISONER WITH POWER As I entered and saluted, Lorguas Ptolemi signaled me to advance, and, fixing his great, hideous eyes upon me, addressed me thus: "You have been with us a few days, yet during that time you have by your prowess won a high position among us.

One could easily picture the gorgeous foliage of the luxuriant Martian vegetation which once filled this scene with life and color; the graceful figures of the beautiful women, the straight and handsome men; the happy frolicking children—all sunlight, happiness and peace.

It was difficult to realize that they had gone; down through ages of darkness, cruelty, and ignorance, until their hereditary instincts of culture and humanitarianism had risen ascendant once more in the final composite race which now is dominant upon Mars.

v. Chapter-5

Top 10 Longest Words:
notwithstanding
responsibility
irresponsible
investigating
comparatively
conclusion
hieroglyphics
sufficiently
instructions
satisfactory

Top 10 Longest Sentences:
"I presume it is the better part of wisdom that we bow to our fate with as good grace as possible, Delah Thoris; but I hope, nevertheless, that I may be present the next time that any Martian, green, red, pink, or violet, has the temerity to even so much as frown on you, my princess." Delah Thoris caught her breath at my last words, and gazed upon me with dilated eyes and quickening breath, and then, with an odd little laugh, which brought roguish dimples to the corners of her mouth, she shook her head and cried: "What a child!"

"I knew that even though you became a member of the community you would not cease to be my friend; 'A warrior may change his metal, but not his heart,' as the saying is upon Barsom." "I think they have been trying to keep us apart," she continued, "For whenever you have been off duty one of the older women of Tars ~~Tarkas~~' retinue has always arranged to trump up some excuse to get Sola and me out of sight.

CHAPTER XIII LOVE-MAKING ON MARS Following the battle with the air ships, the community remained within the city for several days, abandoning the homeward march until they could feel reasonably assured that the ships would not return; for to be caught on the open plains with a cavalcade of chariots and children was far from the desire of even so warlike a people as the green Martians.

In Captain Carter's manuscript it is mentioned always by the name used in the written language of Helium and is spelled in hieroglyphics which it would be difficult and useless to reproduce.] While I was much interested in Delah Thoris' explanation of this wonderful adjunct to Martian warfare, I was more concerned by the immediate problem of their treatment of her.

Let us pity them, my chieftain, for even though we die at their hands we can afford them pity, since we are greater than they and they know it." Had I known the significance of those words "my chieftain," as applied by a red Martian woman to a man, I should have had the surprise of my life, but I did not know at that time, nor for many months thereafter.

And when you learn, John Carter, and if I be dead, as likely I shall be ere the further moon has circled Barsom another twelve times, remember that I listened and that I--smiled." It was all Greek to me, but the more I begged her to explain the more positive became her denials of my request, and, so, in very hopelessness, I desisted.

Only a few days since you, yourself, told me that these great brutes, by the uncertainty of their tempers, often were the means of turning victory into defeat, since, at a crucial moment, they might elect to unseat and rend their riders." "Show me how you accomplish these results," was Tars ~~Tarkas~~' only rejoinder.

If the thoats did not respond with sufficient celerity to the telepathic instructions of their riders they were dealt a terrific blow between the ears with the butt of a pistol, and if they showed fight this treatment was continued until the brutes either were subdued, or had unseated their riders.

"Sarkoja told Sola that you had become a true Thark," she said, "and that I would now see no more of you than of any of the other warriors." "Sarkoja is a liar of the first magnitude," I replied, "notwithstanding the proud claim of the ~~Tharks~~ to absolute verity." Delah Thoris laughed.

They would follow me like dogs, rubbing their great snouts against my body in awkward evidence of affection, and respond to my every command with an alacrity and docility which caused the Martian warriors to ascribe to me the possession of some earthly power unknown on Mars.

vi. Chapter-6

Top 10 Longest Words:
disconsolately
responsibility
inconsistency
circumstances
granddaughter
consideration
helplessness
Helplessness
conciliation
Illustration

Top 10 Longest Sentences:
"The man of Barsom," she finally ventured, "does not ask personal questions of women, except his mother, and the woman he has fought for and won." "But I have fought--" I started, and then I wished my tongue had been cut from my mouth; for she turned even as I caught myself and ceased, and drawing my silks from her shoulder she held them out to me, and without a word, and with head held high, she moved with the carriage of the queen she was toward the plaza and the doorway of her quarters.

Another thing I saw, too, which almost lost my life for me then and there, for it took my mind for the fraction of an instant entirely from my antagonist; for, as Delah Thoris struck the tiny mirror from her hand, Sarkoja, her face livid with hatred and baffled rage, whipped out her dagger and aimed a terrific blow at Delah Thoris; and then Sola, our dear and faithful Sola, sprang between them; the last I saw was the great knife descending upon her shielding breast.

"She says you have angered her, and that is all she will say, except that she is the daughter of a jed and the granddaughter of a jeddak and she has been humiliated by a creature who could not plow the teeth of her grandmother's ~~sorak~~." I pondered over this report for some time, finally asking, "What might a ~~sorak~~ be, Sola?" "A little animal about as big as my hand, which the red Martian women keep to play with," explained Sola.

I do not know why it is that I should always be happy and contented when you, John Carter, a stranger, are with me; yet at such times it seems that I am safe and that, with you, I shall soon return to my father's court and feel his strong arms about me and my mother's tears and kisses on my cheek." "Do people kiss, then, upon Barsom?" I asked, when she had explained the word she used, in answer to my inquiry as to its meaning.

"Were you to give me your word that neither you nor Delah Thoris would attempt to escape until we have safely reached the court of Tal Haju you might have the key and throw the chains into the river Iss." "It was better that you held the key, Tars ~~Tarkas~~," I replied. He smiled, and said no more, but that night as we were making camp I saw him unfasten Delah Thoris' fetters himself.

I believed that from the bottom of my heart, from the depth of my soul on that night in Korad as I sat cross-legged upon my silks while the nearer moon of Barsom raced through the western sky toward the horizon, and lighted up the gold and marble, and jeweled mosaics of my world-old chamber, and I believe it today as I sit at my desk in the little study overlooking the Hudson.

Finally Zad, realizing that he was tiring more than I, evidently decided to close in and end the battle in a final blaze of glory for himself; just as he rushed me a blinding flash of light struck full in my eyes, so that I could not see his approach and could only leap blindly to one side in an effort to escape the mighty blade that it seemed I could already feel in my vitals.

I had escaped it for all the years I had roamed the five continents and their encircling seas; in spite of beautiful women and urging opportunity; in spite of a half-desire for love and a constant search for my ideal, it had remained for me to fall furiously and hopelessly in love with a creature from another world, of a species similar possibly, yet not identical with mine.

I did not need a manual of green Martian etiquette to know what reply to make, for, in fact, I was so wild with anger that I could scarcely refrain from drawing my pistol and shooting him down for the brute he was; but I stood waiting with drawn long-sword, and my only choice was to draw my own and meet him in fair fight with his choice of weapons or a lesser one.

CHAPTER XIV A DUEL TO THE DEATH My first impulse was to tell her of my love, and then I thought of the helplessness of her position wherein I alone could lighten the burdens of her captivity, and protect her in my poor way against the thousands of hereditary enemies she must face upon our arrival at Thark.

vii. Chapter-7

Top 10 Longest Words:
contemplation
metamorphosis
particularly
predicaments
interruption
overruled
bewildered
surroundings
unfathomable
crystallized

Top 10 Longest Sentences:
Few western wonders are more inspiring than the beauties of an Arizona moonlit landscape; the silvered mountains in the distance, the strange lights and shadows upon hog back and arroyo, and the grotesque details of the stiff, yet beautiful cacti form a picture at once enchanting and inspiring; as though one were catching for the first time a glimpse of some dead and forgotten world, so different is it from the aspect of any other spot upon our earth.

I reasoned with myself that I had lain helpless for many hours within the cave, yet nothing had molested me, and my better judgment, when permitted the direction of clear and logical reasoning, convinced me that the noises I had heard must have resulted from purely natural and harmless causes; probably the conformation of the cave was such that a slight breeze had caused the sounds I heard.

To be held paralyzed, with one's back toward some horrible and unknown danger from the very sound of which the ferocious Apache warriors turn in wild stampede, as a flock of sheep would madly flee from a pack of wolves, seems to me the last word in fearsome predicaments for a man who had ever been used to fighting for his life with all the energy of a powerful physique.

Fear is a relative term and so I can only measure my feelings at that time by what I had experienced in previous positions of danger and by those that I have passed through since; but I can say without shame that if the sensations I endured during the next few minutes were fear, then may God help the coward, for cowardice is of a surely its own punishment.

Late in the afternoon my horse, which had been standing with dragging rein before the cave, started slowly down the trail, evidently in search of food and water, and I was left alone with my mysterious unknown companion

From then until possibly midnight all was silence, the silence of the dead; then, suddenly, the awful moan of the morning broke upon my startled ears, and there came again from the black shadows the sound of a moving thing, and a faint rustling as of dead leaves.

My only alternative seemed to lie in flight and my decision was crystallized by a recurrence of the rustling sound from the thing which now seemed, in the darkness of the cave and to my distorted imagination, to be creeping stealthily upon me.

There also came to my nostrils a faintly pungent odor, and I could only assume that I had been overcome by some poisonous gas, but why I should retain my mental faculties and yet be unable to move I could not fathom.

CHAPTER II THE ESCAPE OF THE DEAD A sense of delicious dreaminess overcame me, my muscles relaxed, and I was on the point of giving way to my desire to sleep when the sound of approaching horses reached my ears.

I had not long to wait before a stealthy sound apprised me of their nearness, and then a war-bonnetted, paint-streaked face was thrust cautiously around the shoulder of the cliff, and savage eyes looked into mine.

viii. Chapter-8

Top 10 Longest Words:
characteristics
irregularities
characteristic
consequently
independently
accouterments
noiselessness
gesticulating
comparatively
interminable

Top 10 Longest Sentences:

He sat his mount as we sit a horse, grasping the animal's barrel with his lower limbs, while the hands of his two right arms held his immense spear low at the side of his mount; his two left arms were outstretched laterally to help preserve his balance, the thing he rode having neither bridle or reins of any description for guidance.

Their eyes were set at the extreme sides of their heads a trifle above the center and protruded in such a manner that they could be directed either forward or back and also independently of each other, thus permitting this queer animal to look in any direction, or in two directions at once, without the necessity of turning the head.

The respite my unexpected agility had given me permitted me to formulate plans for the immediate future and to note more closely the appearance of the warriors, for I could not disassociate these people in my mind from those other warriors who, only the day before, had been pursuing me.

The result is that they are infinitely less agile and less powerful, in proportion to their weight, than an Earth man, and I doubt that were one of them suddenly to be transported to Earth he could lift his own weight from the ground; in fact, I am convinced that he could not do so.

Coming, as they did, over the soft and soundless moss, which covers practically the entire surface of Mars with the exception of the frozen areas at the poles and the scattered cultivated districts, they might have captured me easily, but their intentions were far more sinister.

But the little sound caused me to turn, and there upon me, not ten feet from my breast, was the point of that huge spear, a spear forty feet long, tipped with gleaming metal, and held low at the side of a mounted replica of the little devils I had been watching.

Instead of progressing in a same and dignified manner, my attempts to walk resulted in a variety of hops which took me clear of the ground a couple of feet at each step and landed me sprawling upon my face or back at the end of each second or third hop.

It towered ten feet at the shoulder; had four legs on either side; a broad flat tail, larger at the tip than at the root, and which it held straight out behind while running; a gaping mouth which split its head from its snout to its long, massive neck.

Behind this first charging demon trailed nineteen others, similar in all respects, but, as I learned later, bearing individual characteristics peculiar to themselves; precisely as no two of us are identical although we are all cast in a similar mold.

My effort was crowned with a success which appalled me no less than it seemed to surprise the Martian warriors, for it carried me fully thirty feet into the air and landed me a hundred feet from my pursuers and on the opposite side of the enclosure.

ix. Chapter-9

Top 10 Longest Words:
circumstances
consideration
manifestation
consideration
therapeutics
scintillated
introduction
instructions
observation
immediately

Top 10 Longest Sentences:

I saw no signs of extreme age among them, nor is there any appreciable difference in their appearance from the age of maturity, about forty, until, at about the age of one thousand years, they go voluntarily upon their last strange pilgrimage down the river Iss, which leads no living Martian knows whither and from whose bosom no Martian has ever returned, or would be allowed to live did he return after once embarking upon its cold, dark waters.

They first repeated the word "sak" a number of times, and then Tars Tarkas made several jumps, repeating the same word before each leap; then, turning to me, he said, "sak!" I saw what they were after, and gathering myself together I "sacked" with such marvelous success that I cleared a good hundred and fifty feet; nor did I, this time, lose my equilibrium, but landed squarely upon my feet without falling.

What struck me as most remarkable about this assemblage and the hall in which they were congregated was the fact that the creatures were entirely out of proportion to the desks, chairs, and other furnishings; these being of a size adapted to human beings such as I, whereas the great bulks of the Martians could scarcely have squeezed into the chairs, nor was there room beneath the desks for their long legs.

My exhibition had been witnessed by several hundred lesser Martians, and they immediately broke into demands for a repetition, which the chieftain then ordered me to make; but I was both hungry and thirsty, and determined on the spot that my only method of salvation was to demand the consideration from these creatures which they evidently would not voluntarily accord.

The room was well lighted by a number of large windows and was beautifully decorated with mural paintings and mosaics, but upon all there seemed to rest that indefinable touch of the finger of antiquity which convinced me that the architects and builders of these wondrous creations had nothing in common with the crude half-brutes which now occupied them.

Owing to the waning resources of the planet it evidently became necessary to counteract the increasing longevity which their remarkable skill in therapeutics and surgery produced, and so human life has come to be considered but lightly on Mars, as is evidenced by their dangerous sports and the almost continual warfare between the various communities.

Evidently, then, there were other denizens on Mars than the wild and grotesque creatures into whose hands I had fallen, but the evidences of extreme antiquity which showed all around me indicated that these buildings might have belonged to some long-extinct and forgotten race in the dim antiquity of Mars.

As he banged me down upon my feet his face was bent close to mine and I did the only thing a gentleman might do under the circumstances of brutality, boorishness, and lack of consideration for a stranger's rights; I swung my fist squarely to his jaw and he went down like a felled ox.

Toward the center of the city was a large plaza, and upon this and in the buildings immediately surrounding it were camped some nine or ten hundred creatures of the same breed as my captors, for such I now considered them despite the suave manner in which I had been trapped.

Had the men been strangers, and therefore unable to exchange names, they would have silently exchanged ornaments, had their missions been peaceful—otherwise they would have exchanged shots, or have fought out their introduction with some other of their various weapons.

x. Chapter-10

Top 10 Longest Words:
characteristics
representation
ministrations
uncomfortable
intensity
intelligence
straightaway
considerable
monstrosity
voluntarily

Top 10 Longest Sentences:

The nights are either brilliantly illuminated or very dark, for if neither of the two moons of Mars happen to be in the sky almost total darkness results, since the lack of atmosphere, or, rather, the very thin atmosphere, fails to diffuse the starlight to any great extent; on the other hand, if both of the moons are in the heavens at night the surface of the ground is brightly illuminated.

This last device produces an intensely brilliant far-reaching white light, but as the natural oil which it requires can only be obtained by mining in one of several widely separated and remote localities it is seldom used by these creatures whose only thought is for today, and whose hatred for manual labor has kept them in a semi-barbaric state for countless ages.

I could not but wonder what this ferocious-looking monstrosity might do when left alone in such close proximity to such a relatively tender morsel of meat; but my fears were groundless, as the beast, after surveying me intently for a moment, crossed the room to the only exit which led to the street, and lay down full length across the threshold.

And it is well that nature has so graciously and abundantly lighted the Martian night, for the green men of Mars, being a nomadic race without high intellectual development, have but crude means for artificial lighting; depending principally upon torches, a kind of candle, and a peculiar oil lamp which generates a gas and burns without a wick.

It came, as I later discovered, not from an animal, as there is only one mammal on Mars and that one very rare indeed, but from a large plant which grows practically without water, but seems to distill its plentiful supply of milk from the products of the soil, the moisture of the air, and the rays of the sun.

The work had evidently been wrought by a master hand, so subtle the atmosphere, so perfect the technique; yet nowhere was there a representation of a living animal, either human or brute, by which I could guess at the likeness of these other and perhaps extinct denizens of Mars.

Both of Mars' moons are vastly nearer her than is our moon to Earth; the nearer moon being but about five thousand miles distant, while the further is but little more than fourteen thousand miles away, against the nearly one-quarter million miles which separate us from our moon.

The nearer moon of Mars makes a complete revolution around the planet in a little over seven and one-half hours, so that she may be seen hurtling through the sky like some huge meteor two or three times each night, revealing all her phases during each transit of the heavens.

Across the threshold lay stretched the sleepless guardian brute, just as I had last seen him on the preceding day; apparently he had not moved a muscle; his eyes were fairly glued upon me, and I fell to wondering just what might befall me should I endeavor to escape.

This girl alone, among all the green Martians with whom I came in contact, disclosed characteristics of sympathy, kindness, and affection; her ministrations to my bodily wants were unfailing, and her solicitous care saved me from much suffering and many hardships.

xi. Chapter-11

Top 10 Longest Words:
overwhelmingly
accomplishing
gesticulated
intimidating
transcending
commencement
executioner
momentarily
perceptibly
momentarily

Top 10 Longest Sentences:

My beast had an advantage in his first hold, having sunk his mighty fangs far into the breast of his adversary; but the great arms and paws of the ape, backed by muscles far transcending those of the Martian men I had seen, had locked the throat of my guardian and slowly were choking out his life, and bending back his head and neck upon his body, where I momentarily expected the former to fall limp at the end of a broken neck.

I am ever willing to stand and fight when the odds are not too overwhelmingly against me, but in this instance I perceived neither glory nor profit in pitting my relatively puny strength against the iron muscles and brutal ferocity of this enraged denizen of an unknown world; in fact, the only outcome of such an encounter, so far as I might be concerned, seemed sudden death.

I had at least two friends on Mars; a young woman who watched over me with motherly solicitude, and a dumb brute which, as I later came to know, held in its poor ugly carcass more love, more loyalty, more gratitude than could have been found in the entire five million green Martians who rove the deserted cities and dead sea bottoms of Mars.

Suddenly I came to myself and, with that strange instinct which seems ever to prompt me to my duty, I seized the cudgel, which had fallen to the floor at the commencement of the battle, and swinging it with all the power of my earthly arms I crashed it full upon the head of the ape, crushing his skull as though it had been an eggshell.

Evidently devoid of all the finer sentiments of friendship, love, or affection, these people fairly worship physical prowess and bravery, and nothing is too good for the object of their adoration as long as he maintains his position by repeated examples of his skill, strength, and courage.

I was standing near the window and I knew that once in the street I might gain the plaza and safety before the creature could overtake me; at least there was a chance for safety in flight, against almost certain death should I remain and fight however desperately.

With a shriek of fear the ape which held me leaped through the open window, but its mate closed in a terrific death struggle with my preserver, which was nothing less than my faithful watch-thing; I cannot bring myself to call so hideous a creature a dog.

CHAPTER VI. A FIGHT THAT WON FRIENDS The thing, which more nearly resembled our earthly men than it did the Martians I had seen, held me pinioned to the ground with one huge foot, while it jabbered and gesticulated at some answering creature behind me.

I glimpsed him just before he reached the doorway and the sight of him, now roaring as he perceived his lifeless fellow stretched upon the floor, and frothing at the mouth, in the extremity of his rage, filled me, I must confess, with dire forebodings.

They seemed to be deep in argument, and finally one of them addressed me, but remembering my ignorance of his language turned back to Tars Tarkas, who, with a word and gesture, gave some command to the fellow and turned to follow us from the room.

xii. Chapter-12

Top 10 Longest Words:
representative
conversations
unnecessarily
intentionally
circumstances
intellectual
conversation
satisfaction
appropriated
humanitarian

Top 10 Longest Sentences:

Between these walls the little Martians scampered, wild as deer; being permitted to run the full length of the aisle, where they were captured one at a time by the women and older children; the last in the line capturing the first little one to reach the end of the gauntlet, her opposite in the line capturing the second, and so on until all the little fellows had left the enclosure and been appropriated by some youth or female.

CHAPTER VII. CHILD-RAISING ON MARS After a breakfast, which was an exact replica of the meal of the preceding day and an index of practically every meal which followed while I was with the green men of Mars, Sola escorted me to the plaza, where I found the entire community engaged in watching or helping at the harnessing of huge mastodonian animals to great three-wheeled chariots.

As I came to a halt before him, Tars Tarkas pointed over the incubator and said, "Sak." I saw that he wanted me to repeat my performance of yesterday for the edification of Lorgus Ptomei, and, as I must confess that my prowess gave me no little satisfaction, I responded quickly, leaping entirely over the parked chariots on the far side of the incubator.

I do not mean that the adult Martians are unnecessarily or intentionally cruel to the young, but theirs is a hard and pitiless struggle for existence upon a dying planet, the natural resources of which have dwindled to a point where the support of each additional life means an added tax upon the community into which it is thrown.

Entirely unknown to their mothers, who, in turn, would have difficulty in pointing out the fathers with any degree of accuracy, they are the common children of the community, and their education devolves upon the females who chance to capture them as they leave the incubator.

They were not wanted, as their offspring might inherit and transmit the tendency to prolonged incubation, and thus upset the system which has maintained for ages and which permits the adult Martians to figure the proper time for return to the incubators, almost to an hour.

Every one but myself—men, women, and children—were heavily armed, and at the tail of each chariot trotted a Martian hound, my own beast following closely behind ours; in fact, the faithful creature never left me voluntarily during the entire ten years I spent on Mars.

It is the universal language of Mars, through the medium of which the higher and lower animals of this world of paradoxes are able to communicate to a greater or less extent, depending upon the intellectual sphere of the species and the development of the individual.

As I later learned, they had been to the subterranean vaults in which the eggs were kept and had transported them to the incubator, which they had then walled up for another five years, and which, in all probability, would not be visited again during that period.

Sola's duties were now doubled, as she was compelled to care for the young Martian as well as for me, but neither one of us required much attention, and as we were both about equally advanced in Martian education, Sola took it upon herself to train us together.

xiii. Chapter-13

Top 10 Longest Words:
reinforcements
simultaneously
circumstances
southeasterly
requisitioned
southwesterly
unquestionably
hallucination
majestically
irresistible

Top 10 Longest Sentences:

For example, a proportion of them, always the best marksmen, direct their fire entirely upon the wireless finding and sighting apparatus of the big guns of an attacking naval force; another detail attends to the smaller guns in the same way; others pick off the gunners; still others the officers; while certain other quotas concentrate their attention upon the other members of the crew, upon the upper works, and upon the steering gear and propellers.

Instantly the scene changed as by magic; the foremost vessel swung broadside toward us, and bringing her guns into play returned our fire, at the same time moving parallel to our front for a short distance and then turning back with the evident intention of completing a great circle which would bring her up to position once more opposite our firing line; the other vessels followed in her wake, each one opening upon us as she swung into position.

As Sola and I entered the plaza a sight met my eyes which filled my whole being with a great surge of mingled hope, fear, exultation, and depression, and yet most dominant was a subtle sense of relief and happiness; for just as we neared the throng of Martians I caught a glimpse of the prisoner from the battle craft who was being roughly dragged into a nearby building by a couple of green Martian females.

Sola and I had entered a building upon the front of the city, in fact, the same one in which I had had my encounter with the apes, and, wishing to see what had caused the sudden retreat, I mounted to an upper floor and peered from the window out over the valley and the hills beyond; and there I saw the cause of their sudden scurrying to cover.

I could not fathom the seeming hallucination, nor could I free myself from it; but somewhere in the innermost recesses of my soul I felt a strange yearning toward these unknown foemen, and a mighty hope surged through me that the fleet would return and demand a reckoning from the green warriors who had so ruthlessly and wantonly attacked it.

Whether they had discovered us or simply were looking at the deserted city I could not say, but in any event they received a rude reception, for suddenly and without warning the green Martian warriors fired a terrific volley from the windows of the buildings facing the little valley across which the great ships were so peacefully advancing.

This operation required several hours, during which time a number of the chariots were requisitioned to transport the loot, which consisted in arms, ammunition, silks, furs, jewels, strangely carved stone vessels, and a quantity of solid foods and liquids, including many casks of water, the first I had seen since my advent upon Mars.

The sight was awe-inspiring in the extreme as one contemplated this mighty floating funeral pyre, drifting unguided and unmanned through the lonely wastes of the Martian heavens; a derelict of death and destruction, typifying the life story of these strange and ferocious creatures into whose unfriendly hands fate had carried it.

As the craft neared the building, and just before she struck, the Martian warriors swarmed upon her from the windows, and with their great spears eased the shock of the collision, and in a few moments they had thrown out grappling hooks and the big boat was being hauled to ground by their fellows below.

It had never been given me to see such deadly accuracy of aim, and it seemed as though a little figure on one of the craft dropped at the explosion of each bullet, while the banners and upper works dissolved in spurts of flame as the irresistible projectiles of our warriors moved through them.

xiv. Chapter-14

Top 10 Longest Words:
responsibilities
expressionless
unintelligible
administration
accordances
opportunities
possibilities
satisfactory
manufactured
intelligence

Top 10 Longest Sentences:
Say what you please to Tars Tarkas, he can mete out no worse fate to me than a continuation of the horrible existence we are forced to lead in this life." This wild outbreak on the part of Sola so greatly surprised and shocked the other women, that, after a few words of general reprimand, they all lapsed into silence and were soon asleep.

It will not be well for you to permit Tars Tarkas to learn that you hold such degenerate sentiments, as I doubt that he would care to entrust such as you with the grave responsibilities of maternity." "I see nothing wrong with my expression of interest in this red woman," retorted Sola.

With this added incentive I nearly drove Sola distracted from my education and within a few more days I had mastered the Martian tongue sufficiently well to enable me to carry on a passable conversation and to fully understand practically all that I heard.

Customs have been handed down by ages of repetition, but the punishment for ignoring a custom is a matter for individual treatment by a jury of the culprit's peers, and I may say that justice seldom misses fire, but seems rather to rule in inverse ratio to the ascendancy of law.

The training of myself and the young Martians was conducted solely by the women, who not only attend to the education of the young in the arts of individual defense and offense, but are also the artisans who produce every manufactured article wrought by the green Martians.

I could not but note the unnecessary harshness and brutality with which her guards treated her; so different from the almost maternal kindness which Sola manifested toward me, and the respectful attitude of the few green Martians who took the trouble to notice me at all.

After they had retired for the night it was customary for the adults to carry on a desultory conversation for a short time before lapsing into sleep, and now that I could understand their language I was always a keen listener, although I never proffered any remarks myself.

I knew that she was fond of me, and now that I had discovered that she hated cruelty and barbarity I was confident that I could depend upon her to aid me and the girl captive to escape, provided of course that such a thing was within the range of possibilities.

They live at peace with all their fellows, except when duty calls upon them to make war, while we are at peace with none; forever warring among our own kind as well as upon the red men, and even in our own communities the individuals fight amongst themselves.

Oh, it is one continual, awful period of bloodshed from the time we break the shell until we gladly embrace the bosom of the river of mystery, the dark and ancient Iss which carries us to an unknown, but at least no more frightful and terrible existence!

Looking at the longest words and sentences from each chapter shows how varied the vocabulary is throughout the text. Each chapter brings its own mix of lengthy words, which reflects how the story's themes and style keep evolving as the book progresses.

2. i. Document Term Matrix (DTM)

```

> princessDTM<-DocumentTermMatrix(princess)
> princessDTM
<<DocumentTermMatrix (documents: 14, terms: 6007)>>
Non-/sparse entries: 12104/71994
Sparsity : 86%
Maximal term length: 19
Weighting : term frequency (tf)
> inspect(princessDTM)
<<DocumentTermMatrix (documents: 14, terms: 6007)>>
Non-/sparse entries: 12104/71994
Sparsity : 86%
Maximal term length: 19
Weighting : term frequency (tf)
Sample :
          Terms
Docs      and but for had that the upon was which with
chapter_1.txt 90 13 20 25 50 190 12 36 20 25
chapter_10.txt 113 18 34 41 50 193 18 54 15 26
chapter_11.txt 82 14 9 29 37 119 13 18 13 23
chapter_12.txt 58 9 16 13 19 131 8 16 18 20
chapter_13.txt 63 11 17 29 39 131 15 21 11 29
chapter_14.txt 115 17 41 35 43 171 18 49 9 33
chapter_3.txt 92 15 18 19 26 166 20 36 34 20
chapter_4.txt 68 21 10 22 19 148 10 26 16 20
chapter_7.txt 64 9 18 16 12 166 7 17 22 12
chapter_8.txt 81 7 15 24 11 184 22 32 17 11
> str(princessDTM)
List of 6
$ i      : int [1:12104] 1 1 1 1 1 1 1 1 1 ...
$ j      : int [1:12104] 67 69 70 75 77 100 109 116 120 125 ...
$ v      : num [1:12104] 1 1 1 1 7 1 1 4 1 1 ...
$ nrow   : int 14
$ ncol   : int 6007
$ dimnames:List of 2
..$ Docs : chr [1:14] "chapter_1.txt" "chapter_10.txt" "chapter_11.txt" "chapter_12.txt" ...
..$ Terms: chr [1:6007] "'gentleman'" "'til" "\\\"and" "\\\"as" ...
- attr(*, "class")= chr [1:2] "DocumentTermMatrix" "simple_triplet_matrix"
- attr(*, "weighting")= chr [1:2] "term frequency" "tf"

```

We create a Document-Term Matrix (DTM) for the full text, which organizes the frequency of each word across the document. Inspecting the DTM shows us the number of sparse entries (71994), sparsity (86%), and also the most common words in each of the chapters.

ii. Term Document Matrix (TDM)

```

> princessTDM<-TermDocumentMatrix(princess)
> princessTDM
<<TermDocumentMatrix (terms: 6007, documents: 14)>>
Non-/sparse entries: 12104/71994
Sparsity           : 86%
Maximal term length: 19
Weighting          : term frequency (tf)

```

We generate a Term-Document Matrix (TDM), which maps each word to the documents it appears in, helping us analyze word patterns across the text. Similar to DTM's, it also gives us an idea about the sparsity in the text and the longest term length (19).

iii. Making it a data frame

```

> princessdf<-data.frame(ptext[1])
> princessdf[1]
                                         content
1                               CHAPTER I
2
3                               ON THE ARIZONA HILLS
4
5
6   I am a very old man; how old I do not know. Possibly I am a hundred,
7   possibly more; but I cannot tell because I have never aged as other
8   men, nor do I remember any childhood. So far as I can recollect I have
9   always been a man, a man of about thirty. I appear today as I did
10  forty years and more ago, and yet I feel that I cannot go on living
11  forever; that some day I shall die the real death from which there is
12  no resurrection. I do not know why I should fear death, I who have
13  died twice and am still alive; but yet I have the same horror of it as
14  you who have never died, and it is because of this terror of death, I
15  believe, that I am so convinced of my mortality.

```

We create a data frame containing the text from the first chapter, which allows for easier manipulation and analysis of the content.

iv. Function to remove numbers and punctuation and generating a clean corpus

```

> removeNumPunct<-function(x) gsub("[^[:alpha:][:space:]]*", "", x)
> removeNumPunct
function(x) gsub("[^[:alpha:][:space:]]*", "", x)

```

```

> princesscl<-tm::tm_map(princess, content_transformer(removeNumPunct))
> princesscl
<<VCorpus>>
Metadata: corpus specific: 0, document level (indexed): 0
Content: documents: 14
> str(princesscl)
Classes 'VCorpus', 'Corpus' hidden list of 3
$ content:List of 14
..$ :List of 2
... ..$ content: chr [1:268] "CHAPTER I" "" "ON THE ARIZONA HILLS" "" ...
... ..$ meta :List of 7
... ...$ author : chr(0)
... ...$ timestamp: POSIXlt[1:1], format: "2025-04-26 19:58:16"
... ...$ description : chr(0)
... ...$ heading : chr(0)
... ...$ id : chr "chapter_1.txt"
... ...$ language : chr "en"
... ...$ origin : chr(0)
... ...- attr(*, "class")= chr "TextDocumentMeta"
... ...- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
..$ :List of 2
... ..$ content: chr [1:359] "CHAPTER X" "" "CHAMPION AND CHIEF" "" ...
... ..$ meta :List of 7
... ...$ author : chr(0)
... ...$ timestamp: POSIXlt[1:1], format: "2025-04-26 19:58:16"
... ...$ description : chr(0)
... ...$ heading : chr(0)
... ...$ id : chr "chapter_10.txt"
... ...$ language : chr "en"
... ...$ origin : chr(0)
... ...- attr(*, "class")= chr "TextDocumentMeta"
... ...- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"

> inspect(princesscl)
<<VCorpus>>
Metadata: corpus specific: 0, document level (indexed): 0
Content: documents: 14

[[1]]
<<PlainTextDocument>>
Metadata: 7
Content: chars: 13720

[[2]]
<<PlainTextDocument>>
Metadata: 7
Content: chars: 18678

[[3]]
<<PlainTextDocument>>
Metadata: 7
Content: chars: 12822

[[4]]
<<PlainTextDocument>>
Metadata: 7
Content: chars: 10232

```

We define a function to remove numbers and punctuation from the text. Then, we apply this function to clean the content of the text corpus, resulting in a version with only alphabetic characters and spaces, this removes irrelevant characters that add no value to the text.

v. Lowercase transformation

```
> princesslow<-tm_map(princesscl, tm::content_transformer(tolower))
> princesslow
<<VCorpus>>
Metadata: corpus specific: 0, document level (indexed): 0
Content: documents: 14
> str(princesslow)
Classes 'VCorpus', 'Corpus' hidden list of 3
$ content:List of 14
..$ :List of 2
...$ content: chr [1:268] "chapter i" "" "on the arizona hills" "" ...
...$ meta :List of 7
...$ author : chr(0)
...$ timestamp: POSIXlt[1:1], format: "2025-04-26 19:58:16"
...$ description : chr(0)
...$ heading : chr(0)
...$ id : chr "chapter_1.txt"
...$ language : chr "en"
...$ origin : chr(0)
...- attr(*, "class")= chr "TextDocumentMeta"
...- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
..$ :List of 2
...$ content: chr [1:359] "chapter x" "" "champion and chief" "" ...
...$ meta :List of 7
...$ author : chr(0)
...$ timestamp: POSIXlt[1:1], format: "2025-04-26 19:58:16"
...$ description : chr(0)
...$ heading : chr(0)
...$ id : chr "chapter_10.txt"
...$ language : chr "en"
...$ origin : chr(0)
...- attr(*, "class")= chr "TextDocumentMeta"
...- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"

> inspect(princesslow)
<<VCorpus>>
Metadata: corpus specific: 0, document level (indexed): 0
Content: documents: 14

[[1]]
<<PlainTextDocument>>
Metadata: 7
Content: chars: 13720

[[2]]
<<PlainTextDocument>>
Metadata: 7
Content: chars: 18678

[[3]]
<<PlainTextDocument>>
Metadata: 7
Content: chars: 12822

[[4]]
<<PlainTextDocument>>
Metadata: 7
Content: chars: 10232
```

We convert the text to lowercase to standardize the text, making it easier for further analysis by removing case sensitivity.

vi. Document Term Matrix (DTM) for cleaned corpus

```
> princesslowDTM<-DocumentTermMatrix(princesslow)
> princesslowDTM
<<DocumentTermMatrix (documents: 14, terms: 4491)>>
Non-/sparse entries: 10695/52179
Sparsity : 83%
Maximal term length: 17
Weighting : term frequency (tf)
> str(princesslowDTM)
List of 6
$ i      : int [1:10695] 1 1 1 1 1 1 1 1 ...
$ j      : int [1:10695] 3 5 25 33 39 41 45 74 76 87 ...
$ v      : num [1:10695] 1 7 1 1 4 1 1 1 3 ...
$ nrow   : int 14
$ ncol   : int 4491
$ dimnames:List of 2
..$ Docs : chr [1:14] "chapter_1.txt" "chapter_10.txt" "chapter_11.txt" "chapter_12.txt" ...
..$ Terms: chr [1:4491] "abandoning" "ability" "able" "ablest" ...
- attr(*, "class")= chr [1:2] "DocumentTermMatrix" "simple_triplet_matrix"
- attr(*, "weighting")= chr [1:2] "term frequency" "tf"
> inspect(princesslowDTM)
<<DocumentTermMatrix (documents: 14, terms: 4491)>>
Non-/sparse entries: 10695/52179
Sparsity : 83%
Maximal term length: 17
Weighting : term frequency (tf)
Sample :
          Terms
Docs      and but for had her that the was which with
chapter_1.txt 93 13 20 27 0 50 190 38 20 25
chapter_10.txt 117 20 34 41 31 50 193 54 16 26
chapter_11.txt 89 14 10 29 32 38 122 19 13 23
chapter_12.txt 61 10 16 13 3 20 131 17 18 20
chapter_13.txt 65 11 19 29 24 40 131 21 11 30
chapter_14.txt 118 20 43 35 38 45 172 50 10 34
chapter_3.txt 94 17 18 19 0 26 166 37 34 20
chapter_4.txt 70 21 10 22 6 19 148 26 17 20
chapter_7.txt 67 9 18 16 5 12 166 18 24 12
chapter_8.txt 84 7 15 24 36 11 184 32 17 11
```

We create a Document-Term Matrix (DTM) for the lowercase version of the text, enabling the analysis of word frequency and distribution across the document and individual chapters; the cleaning so far reduces the sparsity by 1%.

vii. Matrix representation of DTM

```
> as.matrix(princesslowDTM)
      Terms
Docs    abandoning ability able ablest about above abreast abruptly absence absent absolute absolutely abundantly abysmal
chapter_1.txt      0      0      1      0      7      0      0      0      0      0      0      0      0      0      0      0      0      0
      Terms
Docs    accidentally accompanied accompany accomplish accomplished accomplishing accord accordance accorded according
chapter_1.txt      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0
      Terms
Docs    account accounted accounting accounts accouterments accoutrements accumulate accuracy accurately accusations
chapter_1.txt      1      0      0      0      0      0      0      0      0      0      0      1      0      0
      Terms
Docs    accustomed acid acquainted acquired across act acted action actions activities acts actual adapted add added adding
chapter_1.txt      0      0      0      0      4      0      1      0      0      0      1      0      0      0      0      0      0      0
      Terms
Docs    addition additional address addressed addressing adjoined adjoining adjunct adjusting administration admission admit
chapter_1.txt      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0      0
      Terms
Docs    ado adopt adopted adoration adult adults advance advanced advancement advancing advantage advent adventure
chapter_1.txt      0      0      0      0      0      0      0      0      0      0      0      1      0      0      1      0      0
      Terms
Docs    adventures adversary advice advised advising affair affect affection affections afford african after afternoon
chapter_1.txt      1      0      0      0      0      0      0      0      0      0      0      0      0      0      0      3      1
      Terms
Docs    afterward afterwards again against age aged ageold ages agile agility ago agonies agreeable agreed ahead aid aim
chapter_1.txt      0      0      0      5      0      1      0      0      0      0      1      0      0      1      2      1      1
      Terms
Docs    aimed air airtight aisle alacrity albinos alien alighted alighting alike alive all allegiance alliance allowed
chapter_1.txt      0      0      0      0      0      0      0      0      0      0      2      4      0      0      0      0      0
      Terms
Docs    allowing alloy almost alone along aloud alover already also alternative although altogether aluminum always america
chapter_1.txt      0      0      1      1      0      0      0      1      0      1      0      0      0      0      0      3      0
      Terms
Docs    amity ammunition among amongst amounted amusement anaesthesia ancestor ancestors ancestry ancient and andi anger
chapter_1.txt      0      0      1      0      0      0      0      0      0      0      0      0      0      0      0      93      0      0
      Terms
Docs    angered angrily animal animals ankle anklet annihilate annihilation announcing annoy anomalies another answer
chapter_1.txt      0      0      0      2      0      0      0      0      0      0      0      0      0      0      0      1      0
      Terms
Docs    answered answering antagonist antecedents antelope antennae antennaelike antics antiquity anxious anxiously any
chapter_1.txt      0      0      0      0      1      0      0      0      0      0      0      0      0      0      0      0      4
```

This code converts the Document-Term Matrix (DTM) of the lowercase text into a regular matrix format, allowing us to view the term frequency data.

viii. Removing stop words and TDM of resulting corpus

```
> mystopwords<-c(tm::stopwords("english"))
> mystopwords
[1] "i"        "me"       "my"       "myself"    "we"        "our"      "ours"      "ourselves" "you"
[10] "your"     "yours"    "yourself"  "yourselves" "he"        "him"      "his"       "himself"   "she"
[19] "her"       "hers"     "herself"   "herselves"  "it"        "its"      "itself"    "they"      "them"
[28] "theirs"    "themselves" "what"     "which"     "who"       "whom"     "whom"     "this"      "that"
[37] "those"     "am"        "is"        "are"       "was"       "were"     "were"     "be"        "been"
[46] "have"      "has"       "had"       "having"    "do"        "does"     "does"     "did"       "doing"
[55] "should"    "could"    "ought"    "i'm"      "you're"   "he's"     "he's"     "she's"    "it's"
[64] "they're"   "i've"     "you've"   "we've"    "they've"  "i'd"      "you'd"   "he'd"     "she'd"
[73] "we'd"      "they'd"   "i'll"     "you'll"   "he'll"    "she'll"   "we'll"    "they'll"  "isn't"
[82] "aren't"    "wasn't"   "weren't"  "hasn't"   "haven't" "hadn't"  "doesn't" "don't"    "didn't"
[91] "won't"     "wouldn't" "shan't"   "shouldn't" "can't"    "cannot"  "couldn't" "mustn't"  "let's"
[100] "that's"   "who's"    "what's"   "here's"   "there's"  "when's"  "where's"  "why's"   "how's"
[109] "a"         "an"        "the"      "and"      "but"      "if"       "or"       "because"  "as"
[118] "until"    "while"    "of"       "at"       "by"       "for"     "with"     "about"    "against"
[127] "between"  "into"     "through"  "during"   "before"   "after"    "above"    "below"    "to"
[136] "from"     "up"        "down"    "in"       "out"      "on"       "off"      "over"     "under"
[145] "again"    "further"  "then"     "once"     "here"     "there"   "when"    "where"    "why"
[154] "how"      "all"       "any"      "both"     "each"     "few"      "more"    "most"     "other"
[163] "some"     "such"     "no"       "nor"     "not"      "only"    "own"     "same"     "so"
[172] "than"     "too"      "very"
```

```

> princessstop<-tm::tm_map(princesslow,tm::removeWords,mystopwords)
> tm::inspect(princessstop[[1]])
<<PlainTextDocument>>
Metadata: 7
Content: chars: 9820

```

chapter

arizona hills

```

old man old know possibly hundred
possibly tell never aged
men remember childhood far can recollect
always man man thirty appear today
forty years ago yet feel go living
forever day shall die real death
resurrection know fear death
died twice still alive yet horror
never died terror death
believe convinced mortality

```

```

conviction determined write
story interesting periods life death
explain phenomena can set words
ordinary soldier fortune chronicle strange events
befell ten years dead body lay undiscovered
arizona cave

```

```

> princessstopTDM<-tm::TermDocumentMatrix(princessstop)
> princessstopTDM
<<TermDocumentMatrix (terms: 4393, documents: 14)>>
Non-/sparse entries: 9661/51841
Sparsity : 84%
Maximal term length: 17
Weighting : term frequency (tf)

```

```

> tm::inspect(princessstopTDM)
<<TermDocumentMatrix (terms: 4393, documents: 14)>>
Non-/sparse entries: 9661/51841
Sparsity : 84%
Maximal term length: 17
Weighting : term frequency (tf)
Sample :
    Docs
Terms   chapter_1.txt chapter_10.txt chapter_11.txt chapter_12.txt chapter_13.txt chapter_14.txt chapter_3.txt chapter_4.txt
dejah      0          5          17          3          10          20          0          0
first       3          6          3          1          5          2          5          6
great       1          7          5          2          9          4          3          3
little      4          4          4          1          6          8          11         3
martian     0          13         3          2          9          5          6          9
one         3          19         1          10         7          11         7          5
sola        0          6          11         2          6          10         0          2
thoris      0          5          17         3          10         20         0          0
toward      3          9          1          1          2          3          10         6
upon        12         19         13         8          15         18         20         11

```

This code removes common stopwords (like "the", "and", etc.) from the lowercase version of the text using a list of stopwords from the library 'tm'. It then creates a Term Document Matrix (TDM) of the

filtered text to analyze the term frequency across the chapters. The resulting TDM excludes the most common words, focusing on more meaningful terms for analysis.

ix. Frequent terms, length of a term, and term frequency

```
> freqterms<-tm::findFreqTerms(princessstopTDM,lowfreq = 5)
> freqterms
[1] "ability"      "accompanied"   "across"       "act"        "actions"      "added"       "addressed"    "adjoining"
[9] "advanced"     "advantage"     "affection"    "afterward"   "age"         "ages"        "ahead"       "aid"
[17] "air"          "almost"       "alone"        "along"      "also"        "always"      "among"       "ancient"
[25] "animal"       "animals"      "another"     "answered"    "answering"    "arizona"     "arm"        "armed"
[33] "appeared"     "approach"     "approached"   "approaching" "assume"      "attack"      "attempt"     "attention"
[41] "around"        "ask"          "asked"        "back"        "balance"     "barsoom"     "battle"      "bearing"
[49] "audience"     "away"          "back"        "believe"    "beneath"     "beneath"     "beside"     "best"
[57] "beautiful"    "became"       "become"      "behind"     "blood"       "blow"        "bodies"      "body"
[65] "better"        "beyond"       "black"       "bring"       "broke"       "brought"     "brute"       "brutes"
[73] "bottom"        "break"        "breast"      "camp"        "can"         "captive"     "care"        "cares"
[81] "building"     "buildings"    "call"        "carry"       "carter"      "case"        "catch"       "caught"
[89] "carefully"    "carried"      "carry"       "cease"       "ceased"      "center"      "chamber"     "chapter"
[97] "caused"        "cautiously"   "cavalcade"   "cave"        "child"       "children"    "circumstances" "city"
[105] "chariot"     "chariots"     "chieftain"   "chieftains"  "cold"        "color"       "come"        "command"
[113] "claim"        "clear"        "cliff"       "close"      "cold"        "color"       "come"        "command"

> freqterms<-tm::findFreqTerms(princessstopTDM,lowfreq = 10)
> freqterms
[1] "across"       "ages"        "air"         "almost"      "alone"       "also"        "always"      "among"       "animal"
[10] "animals"      "another"     "answered"    "arm"        "arms"        "around"     "asked"       "attempt"     "attention"
[19] "audience"     "away"        "back"        "barsoom"    "battle"      "beast"       "beautiful"   "behind"      "believe"
[28] "better"        "blow"        "body"        "breast"     "brought"     "brute"       "building"    "buildings"   "came"
[37] "can"          "captive"     "carter"     "cave"        "center"     "chamber"     "chapter"     "chariots"    "chieftain"
[46] "chieftains"   "city"        "close"       "cold"        "colon"      "come"        "common"     "community"   "continued"
[55] "convinced"    "course"      "craft"       "creature"   "creatures"   "customs"     "dark"        "day"         "days"
[64] "dead"          "death"       "dejah"       "determined" "direction"   "discovered"  "ears"        "earth"       "earthly"
[73] "eggs"          "either"      "enclosure"  "end"        "enormous"   "enough"      "entered"    "entire"      "entirely"
[82] "escape"        "even"        "ever"        "every"      "evidently"   "except"      "explained"  "extreme"     "eyes"
[91] "face"          "fact"        "far"        "fear"        "feet"        "fell"        "fellow"     "felt"        "ferocity"
[100] "fight"
[ reached 'max' / getOption("max.print") -- omitted 204 entries ]

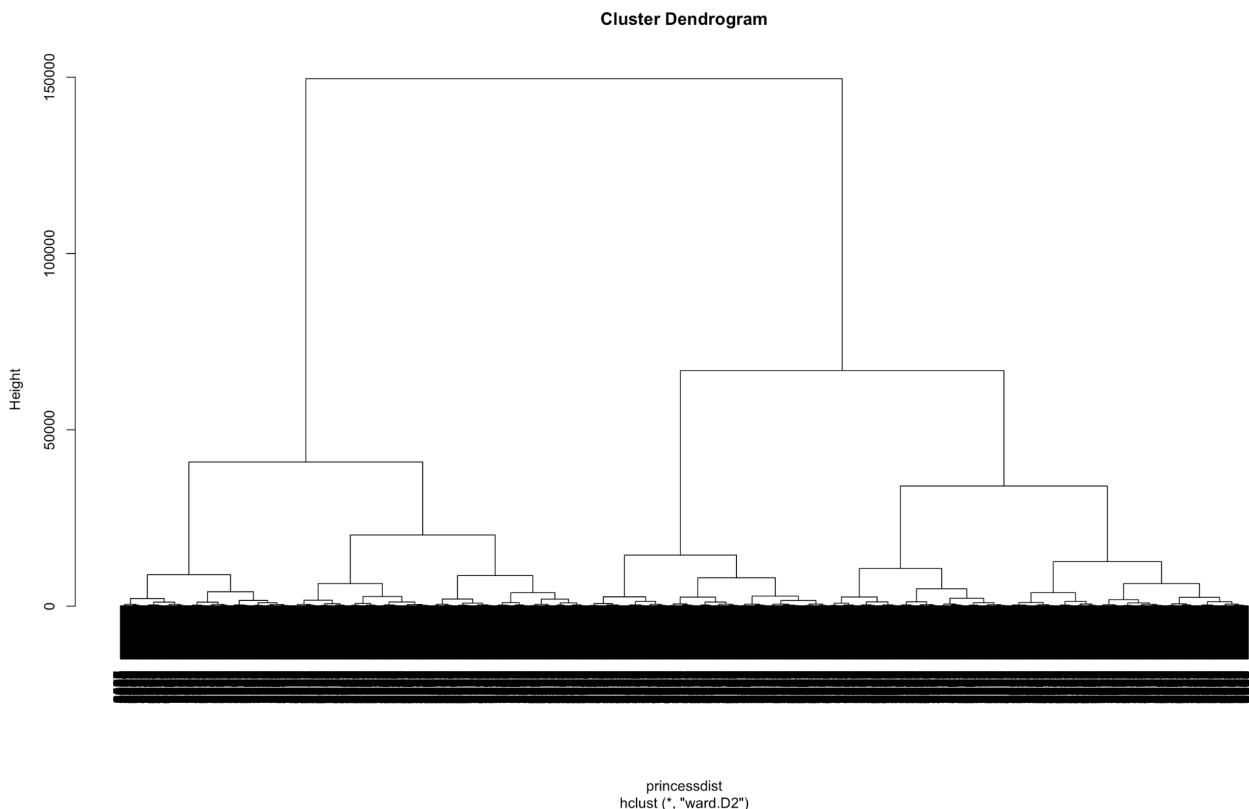
> nchar(freqterms[7])
[1] 9
> freqterms[7]
[1] "addressed"

> princessf<-tm::termFreq(princessstop[[1]])
> princessf
  able account accurately across acted acts advent
  1      1           1       4      1      1      1      1
afternoon aged ago agreed ahead aid aim
  1      1           1       1      2      1      1
  almost alone already alternative always among animals
  1      1           1       1      3      1      2
antelope anything apaches apartments appear apprehension arizona
  1      1           3       1      1      1      5
  armed arming army arose around arrows assure
  1      1           3       1      3      3      1
  attempt attention attraction attributed average await back
  1      1           1       1      1      1      3
  balls bathed beast beautiful became become befell
  1      1           1       1      1      2      2
  belt belts best bestowed better bidding body
  1      1           1       1      3      1      5
bottom bows braves brief bright brisk bristling
  1      1           1       1      1      1      1
  broke brought burros came camp can canteen
  1      1           1       3      3      3      1
captain captains capture carbine careless carry carter
  2      1           1       1      1      1      2
```

The above code identifies terms that appear at least 5 or 10 times in the text, checks the length of the 7th term, and calculates term frequency for the first document after removing stopwords.

x. Dendrogram, and eliminating sparse words for better readability

```
> princessdfnew<-as.data.frame(princessstopTDM[[1]])  
> princessdist<-dist(princessdfnew)  
> princessDG<-hclust(princessdist,method="ward.D2")  
> str(princessDG)  
List of 7  
 $ merge      : int [1:9660, 1:2] -1 -6796 -7897 -2 -7361 -4 -1817 -2533 -3144 -3847 ...  
 $ height     : num [1:9660] 0 0 0 0 0 0 0 0 0 0 ...  
 $ order      : int [1:9661] 8597 6165 72 5407 1890 5408 7946 6166 4818 3931 ...  
 $ labels     : NULL  
 $ method     : chr "ward.D2"  
 $ call       : language hclust(d = princessdist, method = "ward.D2")  
 $ dist.method: chr "euclidean"  
- attr(*, "class")= chr "hclust"  
> plot(princessDG)
```

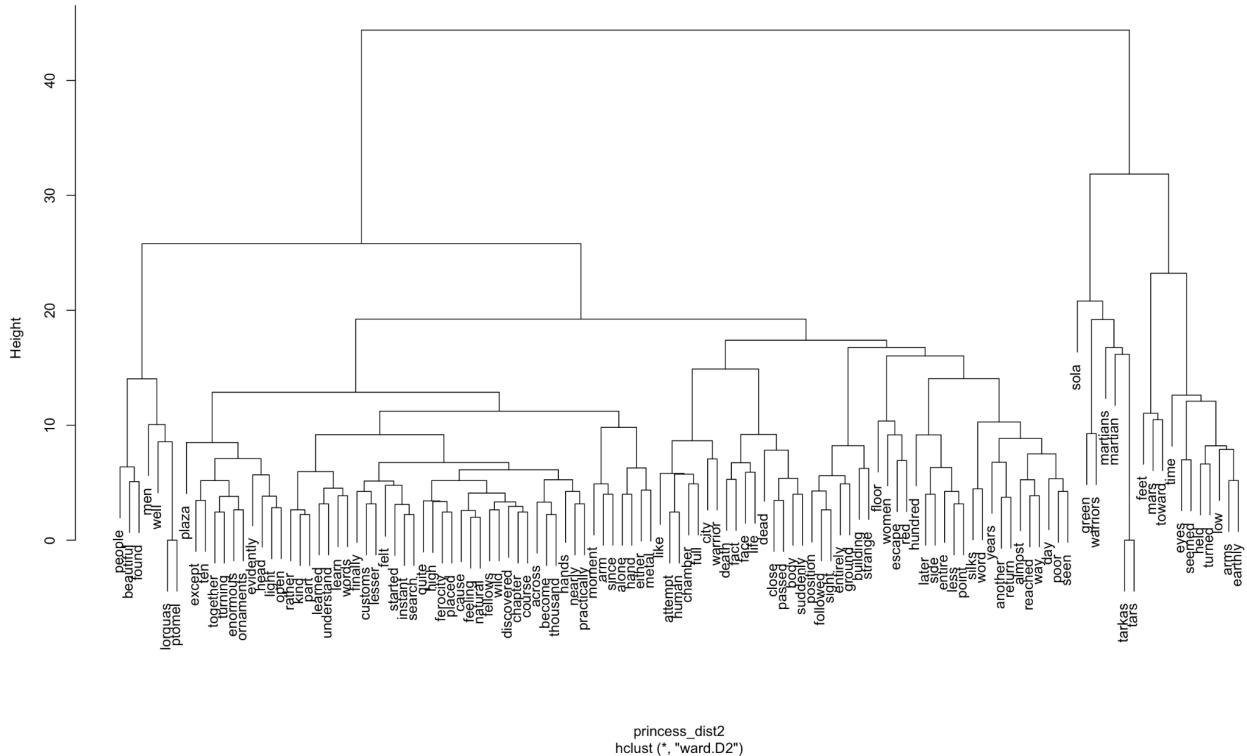


```

> custom_stopwords
[1] "a"          "an"         "the"        "and"        "or"         "but"        "if"         "then"       "than"
[10] "so"         "thus"       "yet"        "still"      "also"       "very"       "just"       "even"       "only"
[19] "ever"       "never"     "never"      "in"         "on"         "at"         "by"         "for"        "with"
[28] "from"       "of"         "out"        "up"         "down"      "over"       "under"      "about"      "around"
[37] "before"     "after"     "between"    "among"      "through"   "into"       "onto"       "upon"       "without"
[46] "within"     "he"         "she"        "it"         "they"      "we"         "you"       "i"          "me"
[55] "him"         "her"       "us"         "them"      "my"         "your"      "his"        "their"     "our"
[64] "its"         "mine"      "yours"     "ours"      "theirs"    "can"        "could"     "may"        "might"
[73] "shall"       "should"    "will"       "would"     "must"      "now"        "soon"       "then"       "today"
[82] "tonight"     "tomorrow"  "yesterday"  "again"      "always"    "sometimes"  "often"      "never"      "ever"
[91] "one"          "two"        "three"     "few"        "many"      "much"       "more"       "most"       "some"
[100] "any"          "none"       "all"        "several"   "each"      "every"      "both"       "say"        "said"
[109] "says"        "go"         "goes"       "went"      "come"      "comes"      "came"       "make"       "makes"
[118] "made"        "get"        "gets"       "got"        "know"      "knows"     "knew"       "see"        "sees"
[127] "saw"          "think"     "thinks"    "thought"   "take"      "takes"      "took"       "give"       "gives"
[136] "gave"        "look"       "looks"     "looked"    "good"      "bad"        "better"     "best"       "great"
[145] "small"       "large"     "little"    "long"      "short"    "old"        "new"        "young"     "first"
[154] "last"         "next"      "thou"      "thee"      "thy"       "hath"      "doth"       "art"        "o"
[163] "ye"           "thus"      "hence"     "whence"    "wherefore" "king"      "queen"     "prince"     "princess"
[172] "lord"         "lady"      "master"    "miss"      "sir"       "mr"        "mrs"        "ms"         "child"
[181] "children"    "boy"       "girl"      "man"       "woman"    "indeed"    "therefore" "meanwhile" "perhaps"
[190] "however"     "besides"   "whose"     "where"    "when"      "while"     "because"   "although"  "though"
[199] "whether"      "away"      "back"      "forward"   "near"      "fan"       "left"       "right"     "behind"
[208] "beside"       "began"     "continued" "replied"   "answered" "cried"     "called"     "asked"     "thing"
[217] "things"       "something" "anything"  "everything" "nothing"
> princessstopnew <- tm_map(princessstop, removeWords, custom_stopwords)
> princessnewTDM <- TermDocumentMatrix(princessstopnew)
> princessnewTDM_sparse <- removeSparseTerms(princessnewTDM, sparse = 0.5)
> princess_mat <- as.matrix(princessnewTDM_sparse)
> princess_df2 <- as.data.frame(princess_mat)
> princess_dist2 <- dist(princess_df2)
> princess_DG2 <- hclust(princess_dist2, method = "ward.D2")
> plot(princess_DG2, main = "Dendrogram after stopword and sparse term removal")

```

Dendrogram after stopword and sparse term removal



This creates a dendrogram for chapter analysis by first removing additional custom stopwords, calculating distances, applying sparse term removal, and calculating term frequency for the processed

text. The dendrogram visualizes the hierarchical clustering of terms after cleaning and reducing data sparsity.

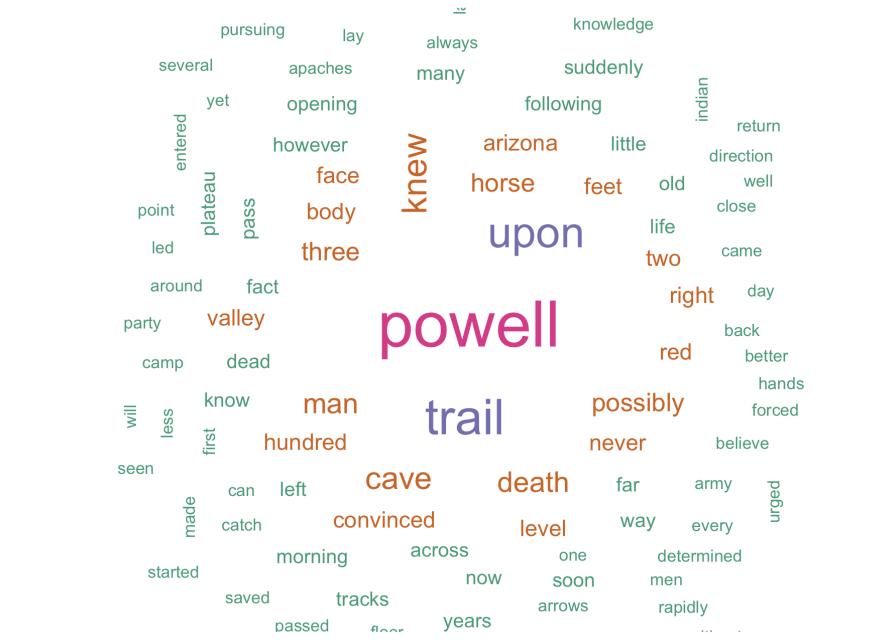
xi. Wordcloud

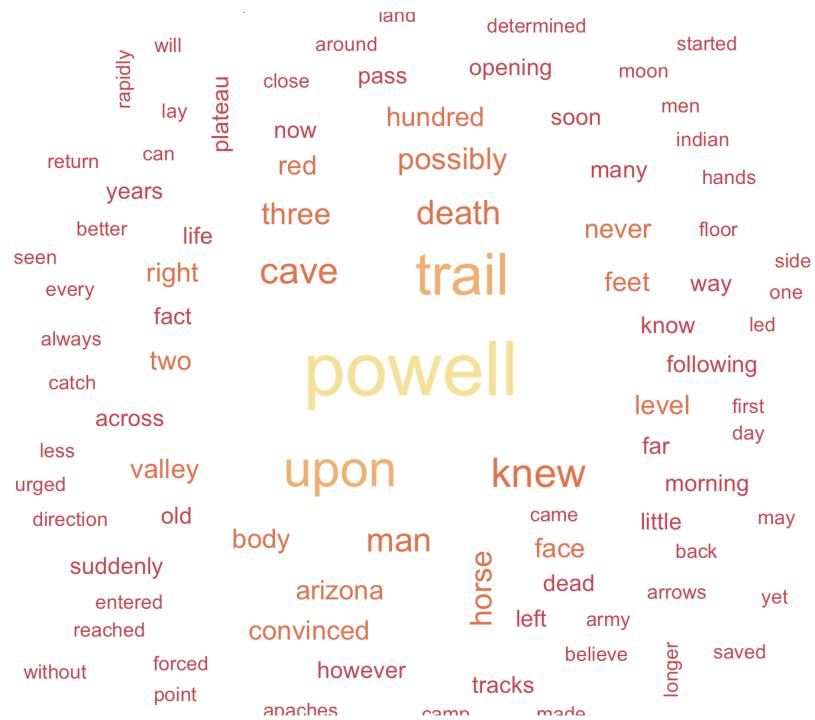
```

> words<-names(princesstf)
> words
[1] "able"          "account"        "accurately"      "across"         "acted"          "acts"           "advent"
[8] "adventures"    "afternoon"       "aged"           "ago"            "agreed"         "ahead"          "aid"
[15] "aim"           "alive"          "almost"          "alone"          "already"        "alternative"   "always"
[22] "among"          "animals"         "another"         "antelope"       "anything"       "apaches"        "apartments"
[29] "appear"         "apprehension"    "arizona"        "arm"            "armed"          "arming"         "army"
[36] "arose"          "around"          "arrows"          "assure"         "attacked"       "attempt"        "attention"
[43] "attraction"     "attributed"      "average"         "await"          "back"           "backward"       "balls"
[50] "bathed"         "beast"          "beautiful"      "became"         "become"         "befell"         "believe"
[57] "belt"           "belts"          "best"           "bestowed"       "better"         "bidding"        "body"
[64] "borne"          "bottom"          "bows"            "braves"         "brief"          "bright"         "brisk"
[71] "bristling"      "broad"          "broke"          "brought"        "burros"         "came"           "camp"
[78] "can"            "canteen"         "carter"         "captain"        "captains"       "capture"        "carbine"
[85] "careless"        "carry"          "carter"         "cartridge"      "cartridges"     "casually"        "catch"
[92] "catching"        "cavalry"         "cave"           "center"         "certain"        "chamber"        "chances"
[99] "chapter"         "characteristic" "charging"       "chase"          "childhood"      "chronicle"      "civil"
[106] "civilization"   "claim"          "clear"          "cliff"          "close"          "clustered"      "clutches"
[113] "colossal"        "colt"            "come"           "commenced"      "commission"     "company"        "comparatively"
[120] "confederate"     "conjured"        "consternation"  "constituted"     "constitutes"     "continue"        "continued"
[127] "continuing"      "continuously"    "conviction"     "convince"        "convinced"      "country"        "course"
[134] "cowardice"        "creeping"        "crept"          "crude"          "cunning"         "customary"      "cut"
[141] "dangerous"       "dangers"         "dark"           "darkness"       "dash"           "dawn"           "day"
> pal <- brewer.pal(8, "Dark2")
> str(pal)
chr [1:8] "#1B9E77" "#D95F02" "#7570B3" "#E7298A" "#66A61E" "#E6AB02" "#A6761D" "#666666"
> princessWC<-wordcloud(words,wordsfreq,colors = pal[1:4],random.order = FALSE)

> pal <- brewer.pal(9, "Spectral")
> princessWC2<-wordcloud(words,wordsfreq,colors = pal[1:4],random.order = FALSE)
> str(princessWC2)
NULL

```





The code above generates wordclouds using the frequency of the terms from the processed text. The word 'powell' appears as the largest, followed by 'trail', 'cave', and 'death', indicating how often the term is seen across the text.

xii. Tokenization

```
> princesstokens<-quanteda::tokens(princesstext$content)
> str(princesstokens)
List of 268
 $ text1 : chr [1:2] "CHAPTER" "I"
 $ text2 : chr(0)
 $ text3 : chr [1:4] "ON" "THE" "ARIZONA" "HILLS"
 $ text4 : chr(0)
 $ text5 : chr(0)
 $ text6 : chr [1:17] "I" "am" "a" "very" ...
 $ text7 : chr [1:13] "possibly" "more" "but" "I" ...
 $ text8 : chr [1:15] "men" "nor" "do" "I" ...
 $ text9 : chr [1:15] "always" "been" "a" "man" ...
 $ text10: chr [1:15] "forty" "years" "and" "more" ...
 $ text11: chr [1:14] "forever" "that" "some" "day" ...
 $ text12: chr [1:14] "no" "resurrection" "I" "do" ...
 $ text13: chr [1:16] "died" "twice" "and" "am" ...
 $ text14: chr [1:15] "you" "who" "have" "never" ...
 $ text15: chr [1:9] "believe" "that" "I" "am" ...
 $ text16: chr(0)
 $ text17: chr [1:12] "And" "because" "of" "this" ...
 $ text18: chr [1:14] "story" "of" "the" "interesting" ...
 $ text19: chr [1:14] "explain" "the" "phenomena" "I" ...
 $ text20: chr [1:11] "ordinary" "soldier" "of" "fortune" ...
 $ text21: chr [1:14] "befell" "me" "during" "the" ...
 $ text22: chr [1:2] "Arizona" "cave"
 $ text23: chr(0)
```

We tokenize the text content of the ‘princess’ object using the `quanteda::tokens()` function, which splits the text into individual tokens, including words and other elements. The result, ‘`princesstokens`’, is a list of tokenized text.

xii. DFM of tokenized corpus

```
> princessdfm<-quanteda::dfm(princesstokens)
> str(princessdfm)
Formal class 'dfm' [package "quanteda"] with 8 slots
..@ docvars : 'data.frame':    268 obs. of  3 variables:
... ..$ docname_ : chr [1:268] "text1" "text2" "text3" "text4" ...
... ..$ docid_   : Factor w/ 268 levels "text1","text2",...: 1 2 3 4 5 6 7 8 9 10 ...
... ..$ segid_   : int [1:268] 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 2 0
... ... ..$ r-version      :Classes 'R_system_version', 'package_version', 'numeric_version' hidden list of 1
... ... ... .$ : int [1:3] 4 5 0
... ... ..$ system        : Named chr [1:3] "Darwin" "arm64" "rajat"
... ... ... - attr(*, "names")= chr [1:3] "sysname" "machine" "user"
... ... ..$ directory      : chr "/Users/rajat/Desktop/2/BD/P3/Text"
... ... ..$ created        : Date[1:1], format: "2025-04-26"
..@ 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 5
... ... ..$ scheme  : chr "unary"
... ... ..$ base   : NULL
... ... ..$ c    : NULL
... ... ..$ smoothing: NULL
... ... ..$ threshold: NULL
... ... ..$ smooth   : num 0
... ... ..$ summary  :List of 2
... ... ..$ hash    : chr(0)
... ... ..$ data    : NULL
... ..$ user    : list()
..@ i      : int [1:2452] 0 0 5 6 7 8 9 10 11 12 ...
..@ p      : int [1:889] 0 1 99 114 252 257 258 267 320 322 ...
```

The above code creates a document-feature matrix (DFM) using the ‘`quanteda`’ package, transforming the tokenized text into a structured matrix format.

xiii. Frequency of terms in DFM

```
> princessdocfreq<-quanteda::docfreq(princessdfm)
> str(princessdocfreq)
Named int [1:888] 1 98 15 138 5 1 9 53 2 3 ...
- attr(*, "names")= chr [1:888] "chapter" "i" "on" "the" ...
> princessdocfreq
```

chapter	i	on	the	arizona	hills	am	a
1	98	15	138	5	1	9	53
very	old	man	how	do	not	know	possibly
2	3	6	1	7	17	4	6
hundred	more	but	cannot	tell	because	have	never
5	5	12	5	1	4	14	5
aged	as	other	men	nor	remember	any	childhood
1	19	4	3	2	1	4	1
so	far	can	recollect	always	been	of	about
9	4	3	1	3	6	91	7
thirty	appear	today	did	forty	years	and	ago
2	1	1	3	2	4	74	1
yet	feel	that	go	living	forever	some	day
3	2	48	2	1	1	8	3
shall	die	real	death	from	which	there	is
2	1	1	7	9	19	4	8
no	resurrection	why	should	fear	who	died	twice
6	1	1	2	1	4	2	1
still	alive	same	horror	it	you	this	terror
1	2	3	1	15	2	15	1
believe	convinced	my	mortality	conviction	determined	to	write
3	5	44	1	1	3	54	1
down	story	interesting	periods	life	explain	phenomena	only
10	2	1	1	4	1	1	4
set	here	in	words	an	ordinary	soldier	fortune
2	1	31	1	14	1	1	1
chronicle	strange	events	befell	me	during	ten	dead
2	1	1	2	26	3	2	4
body	lay	undiscovered	cave	told	mortal	see	manuscript
5	3	1	8	1	1	2	1
until	after	passed	over	for	eternity	average	human
6	3	3	5	20	1	1	1
mind	will	what	grasp	purpose	being	pilloried	by
2	3	2	1	1	2	1	11
public	pulpit	press	held	up	colossal	liar	when
1	1	1	1	8	1	1	6

We then calculate the document frequency of terms in the '*princessdfm*' using the *quanteda::docfreq()* function. The document frequency represents how often each term appears across documents. The result, '*princessdocfreq*', is a vector showing the document frequency for each term in the corpus.

xiv. Assigning weights to words

```
> princessweights<-quanteda::dfm_weight(princessdfm)
> str(princessweights)
Formal class 'dfm' [package "quanteda"] with 8 slots
..@ docvars :'data.frame':   268 obs. of  3 variables:
.. ..$ docname_ : chr [1:268] "text1" "text2" "text3" "text4" ...
.. ..$ docid_  : Factor w/ 268 levels "text1","text2",...: 1 2 3 4 5 6 7 8 9 10 ...
.. ..$ segid_  : int [1:268] 1 1 1 1 1 1 1 1 1 1 ...
..@ meta    :List of 3
.. ..$ system:List of 5
.. .. .$. package-version:Classes 'package_version', 'numeric_version' hidden list of 1
.. .. .$. r-version   :Classes 'R_system_version', 'package_version', 'numeric_version' hidden list of 1
.. .. .$. $ : int [1:3] 4 2 0
.. .. .$. directory  : chr "/Users/rajat/Desktop/2/BD/P3/Text"
.. .. .$. created    : Date[1:1], format: "2025-04-26"
.. ..$. object:List of 10
.. .. .$. 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 5
.. .. .$. scheme   : chr "unary"
.. .. .$. base     : NULL
.. .. .$. c         : NULL
.. .. .$. smoothing: NULL
.. .. .$. threshold: NULL
.. .. .$. smooth    : num 0
.. .. .$. summary   :List of 2
.. .. .$. hash: chr(0)
.. .. .$. data: NULL
.. ..$. user : list()
..@ i      : int [1:2452] 0 0 5 6 7 8 9 10 11 12 ...
..@ p      : int [1:889] 0 1 99 114 252 257 258 267 320 322 ...

> princessweights
Document-feature matrix of: 268 documents, 888 features (98.97% sparse) and 0 docvars.
  features
docs chapter i on the arizona hills am a very old
text1 1 1 0 0 0 0 0 0 0
text2 0 0 0 0 0 0 0 0 0
text3 0 0 1 1 1 1 0 0 0
text4 0 0 0 0 0 0 0 0 0
text5 0 0 0 0 0 0 0 0 0
text6 0 3 0 0 0 0 2 2 1 2
[ reached max_ndoc ... 262 more documents, reached max_nfeat ... 878 more features ]
```

The code above applies term weighting to the DFM using the `quanteda::dfm_weight()` function. The resulting '`princessweights`' object contains the weighted term frequencies for each document.

xv. Assigning 'tfidf' scores

```
> princesstfidf<-quanteda::dfm_tfidf(princessdfm,scheme_tf = "count",scheme_df = "inverse")
> str(princesstfidf)
Formal class 'dfm' [package "quanteda"] with 8 slots
..@ docvars :data.frame': 268 obs. of 3 variables:
.. ..$ docname_ : chr [1:268] "text1" "text2" "text3" "text4" ...
.. ..$ docid_ : Factor w/ 268 levels "text1","text2",...: 1 2 3 4 5 6 7 8 9 10 ...
.. ..$ segid_ : int [1:268] 1 1 1 1 1 1 1 1 1 1 ...
..@ meta :List of 3
.. ..$ system:List of 5
.. .. .$. package-version:Classes 'package_version', 'numeric_version' hidden list of 1
.. .. .$. r-version :Classes 'R_system_version', 'package_version', 'numeric_version' hidden list of 1
.. .. .$. system : Named chr [1:3] "Darwin" "arm64" "rajat"
.. .. .-. attr(*, "names")= chr [1:3] "sysname" "machine" "user"
.. .. $. directory : chr "/Users/rajat/Desktop/2/BD/P3/Text"
.. .. $. created : Date[1:1], format: "2025-04-26"
.. .. $. object:List of 10
.. .. .$. 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
.. .. $. user : list()
..@ i : int [1:2452] 0 0 5 6 7 8 9 10 11 12 ...
..@ p : int [1:889] 0 1 99 114 252 257 258 267 320 322 ...
..@ Dim : int [1:2] 268 888
..@ Dimnames:List of 2
.. ..$. docs : chr [1:268] "text1" "text2" "text3" "text4" ...
```

```
> princesstfidf
```

Document-feature matrix of: 268 documents, 888 features (98.97% sparse) and 0 docvars.

```
features
docs    chapter      i      on      the arizona    hills      am      a      very      old
text1 2.428135 0.4369087 0        0        0        0        0        0        0        0
text2 0          0        0        0        0        0        0        0        0        0
text3 0          0        1.252044 0.2882557 1.729165 2.428135 0        0        0        0
text4 0          0        0        0        0        0        0        0        0        0
text5 0          0        0        0        0        0        0        0        0        0
text6 0          1.3107262 0        0        0        0        2.947785 1.407718 2.127105 3.902027
[ reached max_ndoc ... 262 more documents, reached max_nfeat ... 878 more features ]
```

We compute the Term Frequency-Inverse Document Frequency (TF-IDF) for the DFM using the `quanteda::dfm_tfidf()` function. TF-IDF is a statistical measure used to evaluate the importance of a word in a document relative to the entire corpus. The function uses "count" for term frequency and "inverse" for document frequency. The result, '`princesstfidf`', is a matrix where each term's importance is represented by its TF-IDF score

xvi. Extracting the text as a data frame and reading the file as one large string

```
> princesstextdf<-as.data.frame(princesstext$content)
> princesstextdf
princesstext$content
1 CHAPTER I
2
3 ON THE ARIZONA HILLS
4
5
6     I am a very old man how old I do not know Possibly I am a hundred
7     possibly more but I cannot tell because I have never aged as other
8     men nor do I remember any childhood So far as I can recollect I have
9     always been a man a man of about thirty I appear today as I did
10    forty years and more ago and yet I feel that I cannot go on living
11    forever that some day I shall die the real death from which there is
12    no resurrection I do not know why I should fear death I who have
13    died twice and am still alive but yet I have the same horror of it as
14    you who have never died and it is because of this terror of death I
15        believe that I am so convinced of my mortality
16
17    And because of this conviction I have determined to write down the
18    story of the interesting periods of my life and of my death I cannot
19        explain the phenomena I can only set down here in the words of an
20        ordinary soldier of fortune a chronicle of the strange events that
21    befell me during the ten years that my dead body lay undiscovered in an
22                                Arizona cave
```

```
> princessasstring<-get_text_as_string("APrincessOfMars.txt")
> princessasstring
CHAPTER I ON THE ARIZONA HILLS I am a very old man; how old I do not know. Possibly I am a hundred, possibly more; but I cannot tell because I have never aged as other men, nor do I remember any childhood. So far as I can recollect I have always been a man, a man of about thirty. I appear today as I did forty years and more ago, and yet I feel that I cannot go on living forever; that some day I shall die the real death from which there is no resurrection. I do not know why I should fear death, I who have died twice and am still alive; but yet I have the same horror of it as you who have never died, and it is because of this terror of death, I believe, that I am so convinced of my mortality. And because of this conviction I have determined to write down the story of the interesting periods of my life and of my death. I cannot explain the phenomena; I can only set down here in the words of an ordinary soldier of fortune a chronicle of the strange events that befell me during the ten years that my dead body lay undiscovered in an Arizona cave. I have never told this story, nor shall mortal man see this manuscript until after I have passed over for eternity. I know that the average human mind will not believe what it cannot grasp, and so I do not purpose being pilloried by the public, the pulpit, and the press, and held up as a colossal liar when I am but telling the simple truths which some day science will substantiate. Possibly the suggestions which I gained upon Mars, and the knowledge which I can set down in this chronicle, will aid in an earlier understanding of the mysteries of our sister planet; mysteries to you, but no longer mysteries to me. My name is John Carter; I am better known as Captain Jack Carter of Virginia. At the close of the Civil War I found myself possessed of several hundred thousand dollars (Confederate) and a captain's commission in the cavalry arm of an army which no longer existed; the servant of a sta
```

The code converts the '*princesstext\$content*' into a data frame and then loads the content of the text as a string using the *get_text_as_string()* function.

xvii. get_sentences() to parse individual sentences

```
> ps<-get_sentences(princessasstring)
> ps
[1] "CHAPTER I ON THE ARIZONA HILLS I am a very old man; how old I do not know."
[2] "Possibly I am a hundred, possibly more; but I cannot tell because I have never aged as other men, nor do I remember any childhood."
[3] "So far as I can recollect I have always been a man, a man of about thirty."
[4] "I appear today as I did forty years and more ago, and yet I feel that I cannot go on living forever; that some day I shall die the real death from which there is no resurrection."
[5] "I do not know why I should fear death, I who have died twice and am still alive; but yet I have the same horror of it as you who have never died, and it is because of this terror of death, I believe, that I am so convinced of my mortality."
[6] "And because of this conviction I have determined to write down the story of the interesting periods of my life and of my death."
[7] "I cannot explain the phenomena; I can only set down here in the words of an ordinary soldier of fortune a chronicle of the strange events that befell me during the ten years that my dead body lay undiscovered in an Arizona cave."
[8] "I have never told this story, nor shall mortal man see this manuscript until after I have passed over for eternity."
[9] "I know that the average human mind will not believe what it cannot grasp, and so I do not purpose being pilloried by the public, the pulpits, and the press, and held up as a colossal liar when I am but telling the simple truths which some day science will substantiate."
[10] "Possibly the suggestions which I gained upon Mars, and the knowledge which I can set down in this chronicle, will aid in an earlier understanding of the mysteries of our sister planet; mysteries to you, but no longer mysteries to me."
```

```
> str(ps)
chr [1:2321] "CHAPTER I ON THE ARIZONA HILLS I am a very old man; how old I do not know." ...
```

We extract the sentences from the ‘princessasstring’ text using the `get_sentences()` function and store it in the ‘ps’ variable.

xviii. get_sentiments() with ‘syuzhet’ and ‘bing’

```
> pssyuzhet<-get_sentiment(ps, "syuzhet")
> pssyuzhet
[1] 0.00 0.60 0.00 -1.00 -3.40 0.00 -0.90 -0.25 -0.30 3.20 0.80 -0.10 -1.05 -0.65 0.95 0.65 2.00 2.60 0.50 0.80 1.40
[22] 0.00 0.80 0.50 -4.95 1.65 0.65 1.00 0.00 -0.75 -0.25 2.85 1.50 -0.85 0.70 -1.50 -0.50 -0.75 0.80 -1.50 1.05 1.20
[43] 1.80 0.05 -1.00 -1.25 0.25 0.70 -2.00 0.25 -1.35 -0.65 -1.25 -0.25 1.30 1.00 -0.35 0.20 -0.60 0.30 1.00 1.20
[64] 0.80 0.80 0.00 -1.00 -0.40 -0.45 3.70 -0.60 1.20 -0.75 -0.95 1.25 0.05 -0.85 0.60 -1.50 -0.50 0.00 -0.40 -0.90 0.40
[85] -0.20 1.55 -1.15 -0.15 0.00 -0.50 0.65 0.00 -2.05 -1.60 -1.25 -0.50 -2.75 -4.70 -1.05 -0.50
[ reached 'max' / getOption("max.print") -- omitted 2221 entries ]
```

```
> psbing<-get_sentiment(ps,"bing")
> psbing
[1] 0 0 0 -2 -4 0 -2 0 -1 1 1 0 1 0 2 1 2 2 0 2 3 0 0 -1 -8 1 0 0 0 0 -1 2 3 -1 0 -2 -1 0 -1 -2 -1 0 1
[44] -1 -1 -1 2 1 -2 1 -2 0 -3 0 -3 0 1 -2 0 -1 1 1 1 0 1 0 -2 -1 0 2 -2 0 -3 -1 0 -2 -1 1 -2 -1 -1 0 -3 0 -1 2
[87] -1 -3 0 -1 0 0 -4 -1 -1 1 -3 -8 -2 -1
[ reached 'max' / getOption("max.print") -- omitted 2221 entries ]
```

We calculate the sentiment of the sentences stored in ‘ps’ using two different sentiment analysis methods: ‘syuzhet’ and ‘bing’, the output shows varying sentiment scores for both the methods.

```
xix. get_sentiment_dictionary()
```

```
> psdictionary<-get_sentiment_dictionary()  
> psdictionary
```

	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
9	aberrant	-0.60
10	aberration	-0.80
11	abhor	-0.50
12	abhorred	-1.00
13	abhorrent	-0.50
14	abhors	-1.00
15	abilities	0.60
16	ability	0.50
17	abject	-1.00
18	ablaze	-0.25
19	abnormal	-0.50
20	aboard	0.25

```
> psdictionarybing<-get_sentiment_dictionary("bing")
```

```
> psdictionarybing
```

	word	value
1	a+	1
2	abound	1
3	abounds	1
4	abundance	1
5	abundant	1
6	accessable	1
7	accessible	1
8	acclaim	1
9	acclaimed	1
10	acclamation	1
11	accolade	1
12	accolades	1
13	accommodative	1
14	accommodative	1
15	accomplish	1
16	accomplished	1
17	accomplishment	1
18	accomplishments	1
19	accurate	1
20	accurately	1

The code retrieves sentiment dictionaries for sentiment analysis. The outputs provide the words and their associated sentiment scores from the respective dictionaries. Both methods have varying sentiment scores for the same words.

xx. Sum of sentiment vector to get overall emotional valence

```
> pssum<-sum(pssyuzhet)
> pssum
[1] 63.3
> psbingsum<-sum(psbing)
> psbingsum
[1] -393
```

This code calculates the total sentiment scores for the text, the sum of the sentiment scores from the 'syuzhet' method, results in a positive value of 63.3, indicating an overall positive sentiment. The sum of the sentiment scores from the 'bing' method, results in a negative value of -393, indicating an overall negative sentiment according to the 'bing' dictionary. These contrasting results suggest differing sentiment interpretations based on the methods used.

xxi. Mean and summary() of sentiment vector

```
> psmean<-mean(pssyuzhet)
> psmean
[1] 0.02727273
> psbingmean<-mean(psbing)
> psbingmean
[1] -0.1693236

> summary(pssyuzhet)
   Min. 1st Qu. Median      Mean 3rd Qu.      Max.
-7.10000 -0.70000  0.00000  0.02727  0.75000  8.00000
> summary(psbing)
   Min. 1st Qu. Median      Mean 3rd Qu.      Max.
-9.00000 -1.00000  0.00000 -0.1693  1.0000  8.0000
```

This code computes the average sentiment score for each method and provides summaries for the sentiment scores, the average sentiment score from the 'syuzhet' method is 0.0272, suggesting a slightly positive overall sentiment. The average sentiment score from the 'bing' method is -0.169,

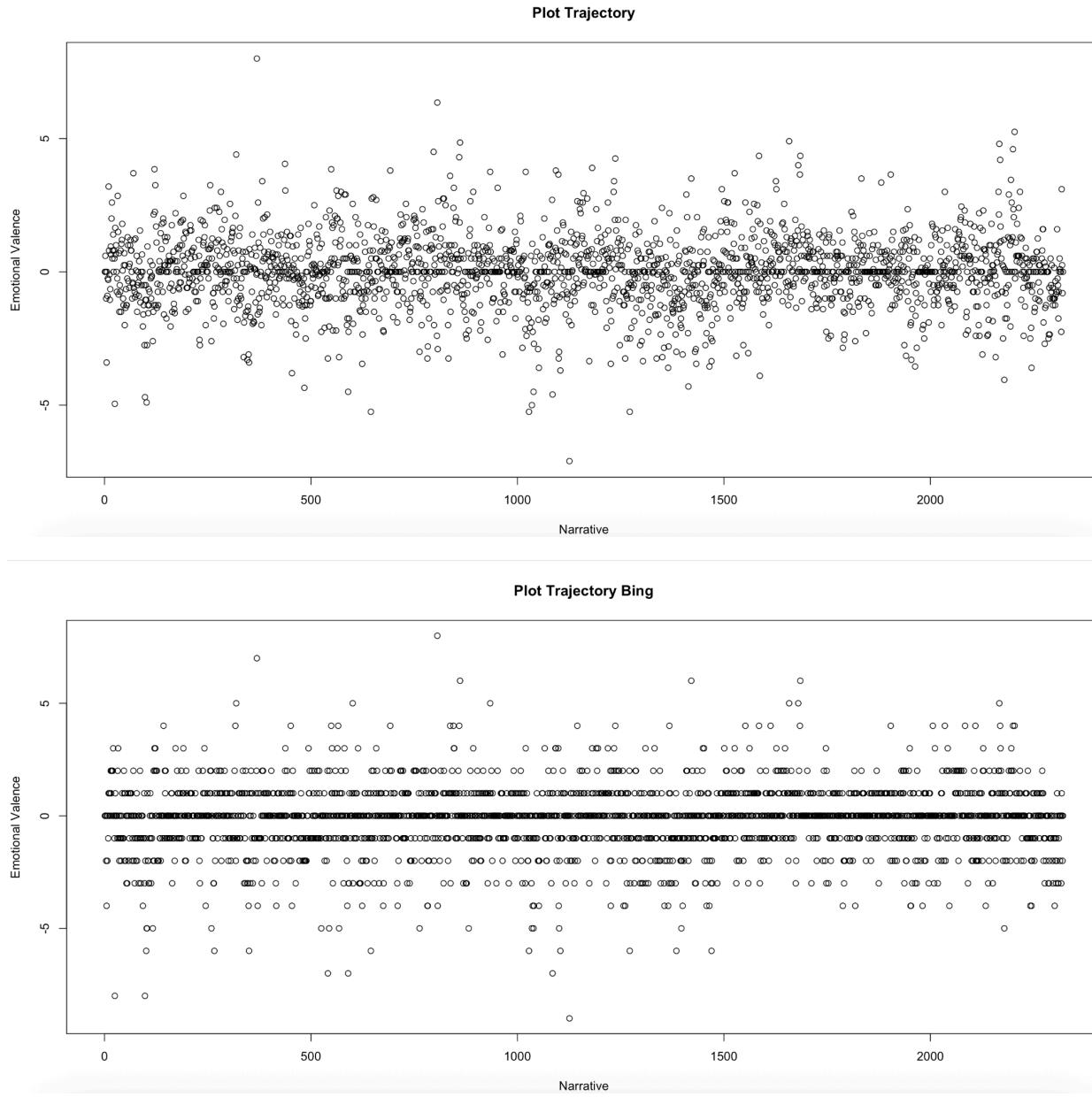
suggesting a slightly negative overall sentiment. The summary provides statistical details (minimum, first quartile, median, mean, third quartile, maximum) for the sentiment scores.

xxii. get_nrc_sentiment()

```
> pssyuzhetnrc<-get_nrc_sentiment(ps)
> pssyuzhetnrc
   anger anticipation disgust fear joy sadness surprise trust negative positive
1      0            0     0  0    0      0     0    0      0     0
2      0            0     0  0    1      0     0    0      0     1
3      0            0     0  0    0      0     0    0      0     0
4      1            1     1  2    0      2     1    1    2      1
5      4            2     2  5    1      3     2    2    5      1
6      1            1     1  1    0      1     1    0    2      2
7      2            1     0  0    1      1     2    3    1      4
8      0            0     0  0    0      0     0    0      1     0
9      0            1     1  0    0      0     0    1    1      3
10     0            0     0  0    0      0     0    2    0      4
[ reached 'max' / getOption("max.print") -- omitted 2311 rows ]
> pssumrnrc<-sum(pssyuzhetnrc)
> pssumrnrc
[1] 15869
> psmeanrnrc<-colMeans(pssyuzhetnrc)
> psmeanrnrc
   anger anticipation disgust fear joy sadness surprise trust negative positive
0.5290823  0.6079276  0.3403705  0.7376131  0.5299440  0.5170185  0.3321844  0.7475226  1.1309780  1.3644981
> summary(pssyuzhetnrc)
   anger anticipation disgust fear joy sadness surprise trust negative positive
Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000
1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000
Median :0.0000 Median :0.0000 Median :0.0000 Median :0.0000 Median :0.0000 Median :0.000 Median :0.0000
Mean   :0.5291 Mean   :0.6079 Mean   :0.3404 Mean   :0.7376 Mean   :0.5299 Mean   :0.517 Mean   :0.3322
3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000
Max.  :7.0000 Max.  :5.0000 Max.  :5.0000 Max.  :8.0000 Max.  :6.0000 Max.  :8.000 Max.  :5.0000
   trust negative positive
Min. :0.0000 Min. :0.000 Min. :0.000
1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:0.000
Median :0.0000 Median :1.000 Median :1.000
Mean   :0.7475 Mean   :1.131 Mean   :1.364
3rd Qu.:1.0000 3rd Qu.:2.000 3rd Qu.:2.000
Max.  :7.0000 Max.  :9.000 Max.  :10.000
```

This code performs sentiment analysis using the NRC sentiment lexicon. Fear, trust, and anticipation are the dominant emotions in the text based on the NRC sentiment analysis. Negative and positive sentiments are balanced, indicating a mix of contrasting emotions across the text.

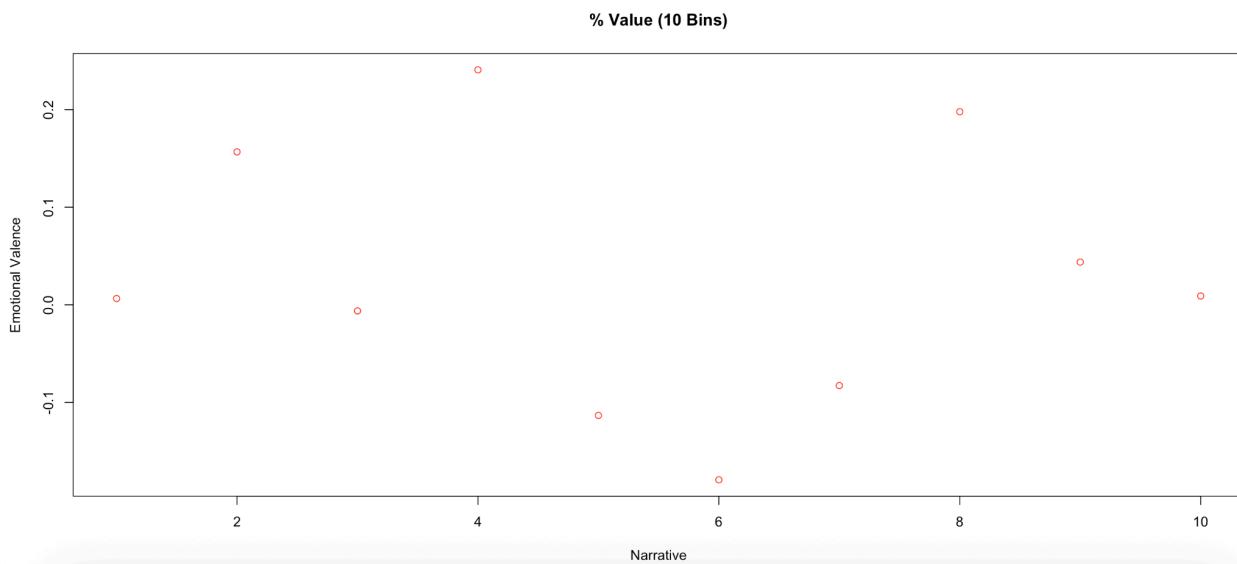
xxiii. Plot of Narrative v/s. Emotional Valence for 'syuzhet' and 'bing' methods



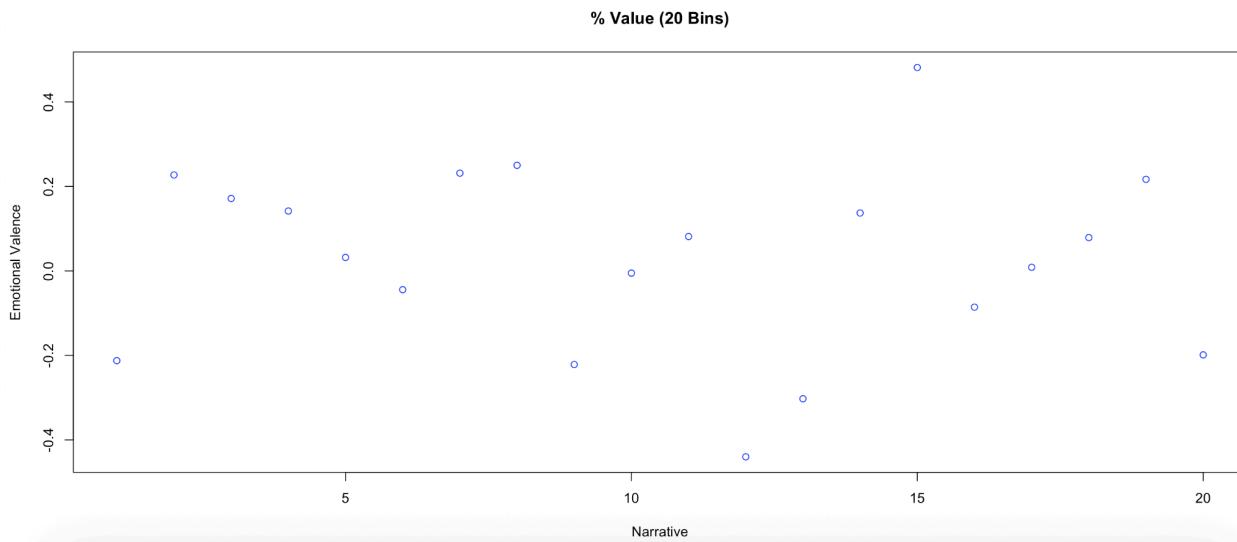
The code generates trajectory plots for sentiment analysis using the 'syuzhet' and 'bing' sentiment dictionaries. Both plots offer a visual representation of how emotions evolve throughout the narrative, they help to compare the emotional trajectory and variance between the two sentiment analysis approaches.

xxiv. %-Value Plot of Narrative v/s. Emotional Valence with 10 and 20 bins

```
> pssentimentpctvalue<-get_percentage_values(pssyuzhet,bins=10)
> structure(pssentimentpctvalue)
 1   2   3   4   5   6   7   8   9   10 
0.006437768 0.156681034 -0.006250000 0.240732759 -0.113362069 -0.179310345 -0.082758621 0.197844828 0.043750000 0.009051724
> plot(pssentimentpctvalue,main="% Value (10 Bins)",xlab="Narrative",ylab="Emotional Valence",col="red")
```



```
> pssentimentpctvalue<-get_percentage_values(pssyuzhet,bins=20)
> structure(pssentimentpctvalue)
 1   2   3   4   5   6   7   8   9   10 
-0.212393162 0.227155172 0.171551724 0.141810345 0.031896552 -0.044396552 0.231465517 0.250000000 -0.221551724 -0.005172414
 11  12  13  14  15  16  17  18  19  20 
0.081465517 -0.440086207 -0.302586207 0.137068966 0.481465517 -0.085775862 0.008620690 0.078879310 0.216810345 -0.198706897
> plot(pssentimentpctvalue,main="% Value (20 Bins)",xlab="Narrative",ylab="Emotional Valence",col="blue")
```



The code generates percentage value plots to visualize the distribution of emotional valence across the narrative using the syuzhet sentiment scores. Both plots demonstrate how the emotional valence fluctuates across the narrative, with the number of bins affecting the level of detail. The sentiment trajectory goes up and down, highlighting emotional shifts throughout the text.

xxv. Exploring 'topicmodels' and 'textmineR'

```

> princess_quanteda <- corpus(princess)
> texts=corpus_reshape(princess_quanteda, to = "paragraphs")
> tokens_texts <- tokens(texts, remove_punct = TRUE)
> tokens_stemmed <- tokens_wordstem(tokens_texts)
> par_dtm <- dfm(tokens_stemmed)
> par_dtm <- dfm_remove(par_dtm, stopwords("english"))
> par_dtm <- dfm_trim(par_dtm)
> par_dtm <- convert(par_dtm, to = "topicmodels")
> set.seed(1)
> lda_model <- topicmodels::LDA(par_dtm, method="Gibbs", k = 5)
> terms(lda_model, 5)
      Topic 1     Topic 2     Topic 3     Topic 4     Topic 5
[1,] "upon"      "sola"      "martian"   "onli"      "littl"
[2,] "thori"     "warrior"   "one"       "toward"    "tar"
[3,] "dejah"     "befor"     "earth"     "first"     "day"
[4,] "mar"        "great"     "green"     "turn"      "face"
[5,] "chieftain" "now"       "two"       "back"      "hand"

> chapter_dir <- "chapters"
> chapter_files <- list.files(chapter_dir, full.names = TRUE)[1:14]
> texts <- sapply(chapter_files, function(x) paste(readLines(x), collapse = " "))
> dtm <- CreateDtm(
+   doc_vec = texts,
+   doc_names = paste0("chapter_", seq_along(texts)),
+   ngram_window = c(1, 1),
+   stopword_vec = tm::stopwords("english"))
> lda_model <- FitLdaModel(
+   dtm = dtm,
+   k = 5,
+   iterations = 500,
+   burnin = 200,
+   alpha = 0.1,
+   beta = 0.05)
> top_terms <- GetTopTerms(lda_model$phi, M = 5)
> print(top_terms)
      t_1      t_2      t_3      t_4      t_5
[1,] "thoris"   "man"     "upon"    "upon"    "one"
[2,] "dejah"    "knew"    "warriors" "toward"  "martian"
[3,] "upon"     "cave"    "eyes"    "feet"    "martians"
[4,] "sola"     "powell"  "seemed"  "mars"    "little"
[5,] "tarkas"   "face"    "great"   "first"   "green"
> lsa_model <- FitLsaModel(dtm = dtm,
+                           k = 5)
> lsa_top_terms <- GetTopTerms(phi = lsa_model$phi, M = 5)
> print(lsa_top_terms)
      t_1      t_2      t_3      t_4      t_5
[1,] "anaesthesia" "dejah"   "martian" "powell"  "warriors"
[2,] "apache"      "thoris"  "one"     "now"     "upon"
[3,] "apprised"    "sola"    "prisoner" "prisoner" "craft"
[4,] "arroyo"       "sarkoja" "martians" "woola"   "fire"
[5,] "attracts"    "people"  "woola"   "may"    "swung"

```

This code applies topic modeling techniques to analyze the text data. The analysis identifies recurring topics across chapters and provides insights into the most prominent terms associated with those topics. The top terms vary across all three models (LDA, LSA, and Gibbs sampling), indicating that

different approaches yield distinct, but complementary, insights into the text's underlying themes and topics.

3. Discussion

In this project, we gained insights into the various aspects of text analytics. Working with large text datasets, such as the chapters from *A Princess of Mars*, provided us with a hands-on understanding of how to preprocess and clean text data to make it suitable for analysis. We started by breaking down the text into manageable chunks, then performed several preprocessing steps, including tokenization, stopword removal, and stemming. This helped us understand the importance of cleaning text before applying any analysis or modeling. It also highlighted the necessity of transforming raw text into a structured format, such as Document-Term Matrices (DTM) or Term-Document Matrices (TDM), which are essential for extracting meaningful insights from unstructured data.

Another takeaway from the project was the exploration of various text mining techniques, such as sentiment analysis, topic modeling, and frequency analysis. Sentiment analysis allowed us to dive deeper into the emotional tone of the text, providing insights into the narrative's overall mood and emotional trajectory. Topic modeling, with techniques like Latent Dirichlet Allocation (LDA), helped us uncover themes within the text, allowing us to observe how topics evolved across chapters. The project gave us experience in visualizing these results through word clouds, dendograms, and sentiment plots, which helped us understand how to communicate insights effectively. The process of experimenting with different models and interpreting their outputs was key in learning how to derive insights from text data and appreciate the complexities involved in working with unstructured information.