

CS4907/CS6444 Big Data and Analytics

Class project #3 Report

Spring 2025

Group 4

Tianfang Fang, Rohith Vijayakumar

Text Analytics in R

Data Set: *The Princess of Mars* by Edgar Rice Burroughs,

chapters I – XIV

Data Set: *The Princess of Mars* by Edgar Rice Burroughs, use chapters I – XIV
(about 60 pages)

The problem is to process a large document and analyze it.

1. Create a VCorpus after you separate the chapters into individual files.

a. Prior to removing the punctuation, find the 10 longest words and 10 longest sentences in each chapter. Prepare a table of this data as well as showing these items.

Firstly we splitted the document into respective chapters, and saved them in separate documents in a folder.

```
library(tokenizers)
library(stringi)
library(wordcloud)
library(syuzhet)
library(dplyr)
library(tm)
library(quanteda)

# Load text data
textData <- readLines('/Users/tianfangfang/Documents/Introduction to Big Data and Analytics_CSCI_6444/Project_3/APrincessOfMars.txt')

# Chapter Titles EXACTLY as in the text (all caps)
chpTitles <- c("CHAPTER I", "CHAPTER II", "CHAPTER III", "CHAPTER IV", "CHAPTER V",
             "CHAPTER VI", "CHAPTER VII", "CHAPTER VIII", "CHAPTER IX", "CHAPTER X",
             "CHAPTER XI", "CHAPTER XII", "CHAPTER XIII", "CHAPTER XIV", "CHAPTER XV",
             "CHAPTER XVI")

# Get positions of chapters
chapter_pos <- sapply(chpTitles, function(x) {
  pos <- which(textData == x)
  if(length(pos) == 0) {
    stop(paste("Chapter title not found:", x))
  }
  return(pos)
})

# Now loop through chapters
for (i in 1:(length(chapter_pos)-1)) {
  chapter_content <- textData[(chapter_pos[i] + 1):(chapter_pos[i + 1] - 1)]
  writeLines(chapter_content, paste0('/Users/tianfangfang/Documents/Introduction to Big Data and Analytics_CSCI_6444/Project_3/Chapters/Chapter', i, '.txt'))
}

# Chapter XV
chapter15_content <- textData[(chapter_pos[15] + 1):(chapter_pos[16] - 1)]
writeLines(chapter15_content, "/Users/tianfangfang/Documents/Introduction to Big Data and Analytics_CSCI_6444/Project_3/Chapters/Chapter15.txt")

# Chapter XVI (last one)
chapter16_content <- textData[(chapter_pos[16] + 1):length(textData)]
writeLines(chapter16_content, "/Users/tianfangfang/Documents/Introduction to Big Data and Analytics_CSCI_6444/Project_3/Chapters/Chapter16.txt")
```

The 16 chapters have been extracted from the given data set.

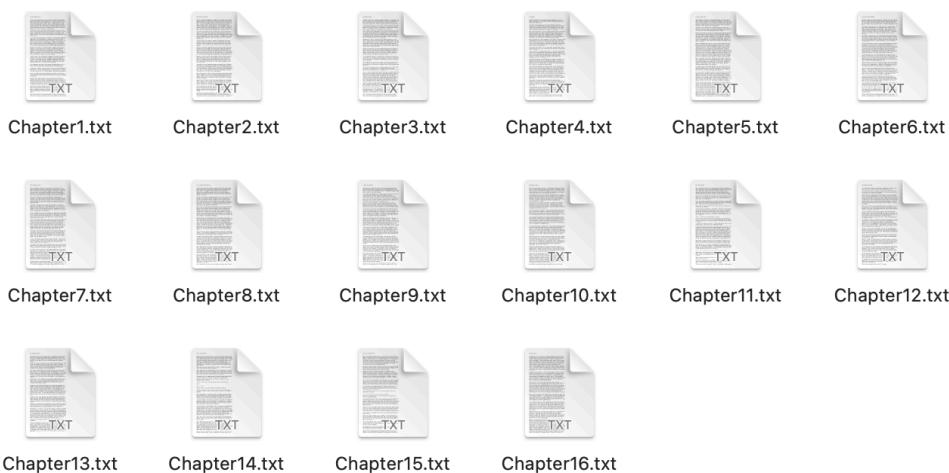


Table for the 10 longest words and 10 longest sentences found in each chapter

Chapter	10 longest words in decreasing order	10 longest sentences in decreasing order
1	"subconsciously" "characteristic" "understanding" "comparatively" "sensitiveness" "consternation" "perpendicular" "resuscitation" "resurrection" "undiscovered"	<p>[1] "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."</p> <p>[2] "The fact that it is difficult to aim anything but imprecations 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."</p> <p>[3] "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."</p> <p>[4] "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."</p> <p>[5] "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."</p> <p>[6] "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."</p> <p>[7] "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."</p> <p>[8] "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."</p> <p>[9] "I soon became so drowsy that I could scarcely resist the strong</p>

		<p>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."</p> <p>[10] "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."</p>
2	<p>"contemplation"</p> <p>"metamorphosis"</p> <p>"particularly"</p> <p>"predicaments"</p> <p>"interruption"</p> <p>"overstrained"</p> <p>"bewilderment"</p> <p>"surroundings"</p> <p>"unfathomable"</p> <p>"crystallized"</p>	<p>[1] "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."</p> <p>[2] "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."</p> <p>[3] "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."</p> <p>[4] "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."</p> <p>[5] "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 and the dead body of my friend, which lay just within my range of vision upon the ledge where I had placed it in the early morning."</p> <p>[6] "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."</p> <p>[7] "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."</p> <p>[8] "There also came to my nostrils a faintly pungent odor, and I could only assume that I had been overcome by some poisonous gas,</p>

		<p>but why I should retain my mental faculties and yet be unable to move I could not fathom."</p> <p>[9] "I had not long to wait before a stealthy sound apprised me of their nearness, and then a war-bonneted, paint-streaked face was thrust cautiously around the shoulder of the cliff, and savage eyes looked into mine."</p> <p>[10] "My longing was beyond the power of opposition; I closed my eyes, stretched out my arms toward the god of my vocation and felt myself drawn with the suddenness of thought through the trackless immensity of space."</p>
3	<p>"characteristics" "irregularities" "characteristic" "consciousness" "independently" "accouterments" "noiselessness" "gesticulating" "comparatively" "interminable"</p>	<p>[1] "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."</p> <p>[2] "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."</p> <p>[3] "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."</p> <p>[4] "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."</p> <p>[5] "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."</p> <p>[6] "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."</p> <p>[7] "Instead of progressing in a sane 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."</p> <p>[8] "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."</p> <p>[9] "Behind this first charging demon trailed nineteen others, similar in all respects, but, as I learned later, bearing individual characteristics</p>

		<p>peculiar to themselves; precisely as no two of us are identical although we are all cast in a similar mold."</p> <p>[10] "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."</p>
4	<p>"circumstances"</p> <p>"consideration"</p> <p>"manifestation"</p> <p>"therapeutics"</p> <p>"scintillated"</p> <p>"introduction"</p> <p>"instructions"</p> <p>"observation"</p> <p>"immediately"</p> <p>"surrounding"</p>	<p>[1] "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."</p> <p>[2] "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."</p> <p>[3] "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."</p> <p>[4] "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."</p> <p>[5] "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."</p> <p>[6] "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."</p> <p>[7] "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."</p> <p>[8] "Toward the center of the city was a large plaza, and upon this</p>

		<p>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."</p> <p>[9] "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."</p> <p>[10] "The other nine hundred and seventy-nine die violent deaths in duels, in hunting, in aviation and in war; but perhaps by far the greatest death loss comes during the age of childhood, when vast numbers of the little Martians fall victims to the great white apes of Mars."</p>
5	<ul style="list-style-type: none"> "characteristics" "representation" "ministrations" "uncomfortable" "intellectual" "intelligence" "straightaway" "considerable" "monstrosity" "voluntarily" 	<p>[1] "The nights are either brilliantly illumined 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."</p> <p>[2] "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."</p> <p>[3] "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."</p> <p>[4] "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."</p> <p>[5] "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."</p> <p>[6] "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."</p> <p>[7] "Both of Mars' moons are vastly nearer her than is our moon to</p>

		<p>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."</p> <p>[8] "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."</p> <p>[9] "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."</p> <p>[10] "This girl alone, among all the green Martians with whom I came in contact, disclosed characteristics of sympathy, kindliness, and affection; her ministrations to my bodily wants were unfailing, and her solicitous care saved me from much suffering and many hardships."</p>
6	<ul style="list-style-type: none"> "overwhelmingly" "accomplishing" "gesticulated" "intermediary" "transcending" "commencement" "executioner" "momentarily" "perceptibly" "forebodings" 	<p>[1] "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."</p> <p>[2] "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."</p> <p>[3] "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 roved the deserted cities and dead sea bottoms of Mars."</p> <p>[4] "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."</p> <p>[5] "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."</p> <p>[6] "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</p>

		<p>overtake me; at least there was a chance for safety in flight, against almost certain death should I remain and fight however desperately."</p> <p>[7] "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."</p> <p>[8] "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."</p> <p>[9] "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."</p> <p>[10] "The warrior whose gun I had struck up looked enquiringly at Tars Tarkas, but the latter signed that I be left to my own devices, and so we returned to the plaza with my great beast following close at heel, and Sola grasping me tightly by the arm."</p>
7	<p>"representative"</p> <p>"conversations"</p> <p>"unnecessarily"</p> <p>"intentionally"</p> <p>"circumstances"</p> <p>"intellectual"</p> <p>"conversation"</p> <p>"satisfaction"</p> <p>"appropriated"</p> <p>"humanitarian"</p>	<p>[1] "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."</p> <p>[2] "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."</p> <p>[3] "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."</p> <p>[4] "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."</p> <p>[5] "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."</p> <p>[6] "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</p>

		<p>left me voluntarily during the entire ten years I spent on Mars."</p> <p>[7] "I saw that he wanted me to repeat my performance of yesterday for the edification of Lorquas 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."</p> <p>[8] "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."</p> <p>[9] "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."</p> <p>[10] "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."</p>
8	<p>"reinforcements"</p> <p>"simultaneously"</p> <p>"circumstances"</p> <p>"southeasterly"</p> <p>"requisitioned"</p> <p>"southwesterly"</p> <p>"unaccountably"</p> <p>"hallucination"</p> <p>"majestically"</p> <p>"irresistible"</p>	<p>[1] "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."</p> <p>[2] "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."</p> <p>[3] "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."</p> <p>[4] "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."</p> <p>[5] "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</p>

		<p>wantonly attacked it."</p> <p>[6] "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."</p> <p>[7] "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."</p> <p>[8] "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."</p> <p>[9] "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."</p> <p>[10] "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 mowed through them."</p>
9	<p>"responsibilities"</p> <p>"expressionless"</p> <p>"unintelligible"</p> <p>"administration"</p> <p>"accouterments"</p> <p>"importunities"</p> <p>"possibilities"</p> <p>"satisfactory"</p> <p>"manufactured"</p> <p>"intelligence"</p>	<p>[1] "With this added incentive I nearly drove Sola distracted by my opportunities to hasten on 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."</p> <p>[2] "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."</p> <p>[3] "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."</p> <p>[4] "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."</p> <p>[5] "After they had retired for the night it was customary for the adults to carry on a desultory conversation for a short time before</p>

		<p>lapsing into sleep, and now that I could understand their language I was always a keen listener, although I never proffered any remarks myself."</p> <p>[6] "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."</p> <p>[7] "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."</p> <p>[8] "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!"</p> <p>[9] "I did not see the prisoner again for several days subsequent to our first encounter, and then only to catch a fleeting glimpse of her as she was being conducted to the great audience chamber where I had had my first meeting with Lorquas Ptomei."</p> <p>[10] "I did not even know that there were any better conditions to escape to, but I was more than willing to take my chances among people fashioned after my own mold rather than to remain longer among the hideous and bloodthirsty green men of Mars."</p>
10	<ul style="list-style-type: none"> "responsibilities" "characteristics" "companionship" "manifestation" "ludicrousness" "precipitately" "authoritative" "understanding" "theoretically" "granddaughter" 	<p>[1] "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."</p> <p>[2] "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."</p> <p>[3] "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."</p> <p>[4] "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</p>

		<p>the audience chamber and listen to the proceedings."</p> <p>[5] "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."</p> <p>[6] "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 headforemost into a pot of frijoles."</p> <p>[7] "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."</p> <p>[8] "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."</p> <p>[9] "I was convinced that the brute loved me; I had seen more evidences of affection in him than in any other Martian animal, man or beast, and I was sure that gratitude for the acts that had twice saved his life would more than outweigh his loyalty to the duty imposed upon him by cruel and loveless masters."</p> <p>[10] "The act was so womanly, so earthly womanly, that though it stung my pride it also warmed my heart with a feeling of companionship; it was good to know that someone else on Mars beside myself had human instincts of a civilized order, even though the manifestation of them was so painful and mortifying."</p>
11	<p>"accouterments"</p> <p>"circumstances"</p> <p>"questioningly"</p> <p>"eavesdropping"</p> <p>"comparatively"</p> <p>"intermarrying"</p> <p>"irretrievably"</p> <p>"architecture"</p> <p>"compositions"</p> <p>"conversation"</p>	<p>[1] "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 Barsoomians, beneath the countless intervening ages."</p> <p>[2] "\"Because, John Carter,\" she replied, \"nearly every planet and star having atmospheric conditions at all approaching those of Barsoom, 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 Tharkian warriors, were entirely undisfigured and unadorned."</p> <p>[3] "A similar wave of feeling seemed to stir her; she drew away</p>

		<p>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 Virginia; but on Barsoom no man lies; if he does not wish to speak the truth he is silent."</p> <p>[4] "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."</p> <p>[5] "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."</p> <p>[6] "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."</p> <p>[7] "Only in the valley Dor, where the river Iss empties into the lost sea of Korus, is there supposed to be a different language spoken, and, except in the legends of our ancestors, there is no record of a Barsoomian returning up the river Iss, from the shores of Korus in the valley of Dor."</p> <p>[8] "I soon found Sola and explained to her that I wished her to guard Dejah Thoris as she had guarded me; that I wished her to find other quarters where they would not be molested by Sarkoja, and I finally informed her that I myself would take up my quarters among the men."</p> <p>[9] "When I questioned her closely on this seeming familiarity with earthly things she laughed, and cried out: \"Why, every school boy on Barsoom knows the geography, and much concerning the fauna and flora, as well as the history of your planet fully as well as of his own."</p> <p>[10] "We decided that she must have been eavesdropping, but as we could recall nothing of importance that had passed between us we dismissed the matter as of little consequence, merely promising ourselves to be warned to the utmost caution in the future."</p>
12	<p>"personification"</p> <p>"humanitarianism"</p> <p>"characteristics"</p> <p>"responsibility"</p> <p>"granddaughter"</p> <p>"investigation"</p> <p>"communication"</p> <p>"corresponding"</p> <p>"conversation"</p> <p>"headquarters"</p>	<p>[1] "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."</p> <p>[2] "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</p>

		<p>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.\\"</p> <p>[3] "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."</p> <p>[4] "\"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 Hajus; 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."</p> <p>[5] "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."</p> <p>[6] "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."</p> <p>[7] "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."</p> <p>[8] "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."</p> <p>[9] "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."</p> <p>[10] "A PRISONER WITH POWER As I entered and saluted, Lorquas Ptomei 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."</p>
13	"notwithstanding" "responsibility" "irresponsible" "investigating" "comparatively" "companionship"	<p>[1] "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."</p>

	<p>"hieroglyphics" "sufficiently" "instructions" "satisfactory"</p>	<p>[2] "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 Dejah Thoris' explanation of this wonderful adjunct to Martian warfare, I was more concerned by the immediate problem of their treatment of her."</p> <p>[3] "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."</p> <p>[4] "\\"I presume it is the better part of wisdom that we bow to our fate with as good grace as possible, Dejah 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.\\""</p> <p>[5] "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."</p> <p>[6] "Day had now given away to night and as we wandered along the great avenue lighted by the two moons of Barsoom, and with Earth looking down upon us out of her luminous green eye, it seemed that we were alone in the universe, and I, at least, was content that it should be so."</p> <p>[7] "That she shared my sentiments in this respect I was positive, for on my approach the look of pitiful hopelessness left her sweet countenance to be replaced by a smile of joyful welcome, as she placed her little right hand upon my left shoulder in true red Martian salute."</p> <p>[8] "Dejah 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!\"</p> <p>[9] "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.\\""</p> <p>[10] "As my arm rested for an instant upon her I felt a thrill pass through every fiber of my being such as contact with no other mortal had even produced; and it seemed to me that she had leaned slightly toward me, but of that I was not sure."</p>
14	<p>"disconsolately" "responsibility" "inconsistency" "circumstances" "granddaughter" "consideration"</p>	<p>[1] "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 Dejah 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 Dejah Thoris; and then Sola, our dear and faithful Sola, sprang between them; the last I</p>

"helplessness"
"conciliation"
"Illustration"
"humiliations"

saw was the great knife descending upon her shielding breast."

[2] "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 Barsoom 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."

[3] "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."

[4] "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."

[5] "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 he 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."

[6] "\"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."

[7] "We rushed each other furiously time after time, 'til suddenly, feeling the sharp point of his sword at my breast in a thrust I could neither parry nor escape, I threw myself upon him with outstretched sword and with all the weight of my body, determined that I would not die alone if I could prevent it."

[8] "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.\\""

[9] "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."

[10] "As a matter of fact, the new-laid egg is but little larger than an ordinary goose egg, and as it does not commence to grow until subjected to the light of the sun the chieftains have little difficulty in transporting several hundreds of them at one time from the storage

		vaults to the incubators."
15	<p>"responsibilities" "phantasmagoria" "responsibility" "consciousness" "instantaneous" "granddaughter" "monosyllables" "interspersed" "unnoticeable" "indefinitely"</p>	<p>[1] "We made a most imposing and awe-inspiring spectacle as we strung out across the yellow landscape; the two hundred and fifty ornate and brightly colored chariots, preceded by an advance guard of some two hundred mounted warriors and chieftains riding five abreast and one hundred yards apart, and followed by a like number in the same formation, with a score or more of flankers on either side; the fifty extra mastodons, or heavy draught animals, known as zitidars, and the five or six hundred extra throats of the warriors running loose within the hollow square formed by the surrounding warriors."</p> <p>[2] ""She taught me rapidly the language and customs of my kind, and one night she told me the story I have told to you up to this point, impressing upon me the necessity for absolute secrecy and the great caution I must exercise after she had placed me with the other young Tharks to permit no one to guess that I was further advanced in education than they, nor by any sign to divulge in the presence of others my affection for her, or my knowledge of my parentage; and then drawing me close to her she whispered in my ear the name of my father."</p> <p>[3] "From that moment on he was the cruellest of the cruel, and I am awaiting the day when he shall win the goal of his ambition, and feel the carcass of Tal Hajus beneath his foot, for I am as sure that he but waits the opportunity to wreak a terrible vengeance, and that his great love is as strong in his breast as when it first transfigured him nearly forty years ago, as I am that we sit here upon the edge of a world-old ocean while sensible people sleep, John Carter.""</p> <p>[4] ""With final imprecations, Sarkoja hastened away to Tal Hajus to report her discovery, and while she was gone my mother, wrapping me in the silks and furs of her night coverings, so that I was scarcely noticeable, descended to the streets and ran wildly away toward the outskirts of the city, in the direction which led to the far south, out toward the man whose protection she might not claim, but on whose face she wished to look once more before she died."</p> <p>[5] "Our animals had been two days without drink, nor had they had water for nearly two months, not since shortly after leaving Thark; but, as Tars Tarkas explained to me, they require but little and can live almost indefinitely upon the moss which covers Barsoom, and which, he told me, holds in its tiny stems sufficient moisture to meet the limited demands of the animals."</p> <p>[6] "She was also less cold and cruel than most green Martian women, and caring little for their society, she often roamed the deserted avenues of Thark alone, or went and sat among the wild flowers that deck the nearby hills, thinking thoughts and wishing wishes which I believe I alone among Tharkan women today may understand, for am I not the child of my mother?"</p> <p>[7] "As soon as they were through with me I hastened to the chariot of Dejah Thoris, where I found my poor Sola with her chest swathed in bandages, but apparently little the worse for her encounter with Sarkoja, whose dagger it seemed had struck the edge of one of Sola's</p>

		<p>metal breast ornaments and, thus deflected, had inflicted but a slight flesh wound."</p> <p>[8] "His love for my mother had never diminished, and his own ambition in life was to reach a point where he might wrest the metal from Tal Hajus himself, and thus, as ruler of the Tharks, be free to claim her as his own, as well as, by the might of his power, protect the child which otherwise would be quickly dispatched should the truth become known."</p> <p>[9] "But one day the chance was lost forever, in so far as it could come in time to save his loved ones, for he was ordered away upon a long expedition to the ice-clad south, to make war upon the natives there and despoil them of their furs, for such is the manner of the green Barsoomian; he does not labor for what he can wrest in battle from others."</p> <p>[10] "The gleaming metal and jewels of the gorgeous ornaments of the men and women, duplicated in the trappings of the zitidars and thoats, and interspersed with the flashing colors of magnificent silks and furs and feathers, lent a barbaric splendor to the caravan which would have turned an East Indian potentate green with envy."</p>
16	"personification" "trustworthiness" "distinguishable" "unconsciousness" "longitudinally" "characteristic" "insignificance" "scintillations" "accommodations" "identification"	<p>[1] "Here, of course, I found the beasts of the warriors who quartered in the adjacent buildings, and the warriors themselves I might expect to meet within if I entered; but, fortunately for me, I had another and safer method of reaching the upper story where Dejah Thoris should be found, and, after first determining as nearly as possible which of the buildings she occupied, for I had never observed them before from the court side, I took advantage of my relatively great strength and agility and sprang upward until I grasped the sill of a second-story window which I thought to be in the rear of her apartment."</p> <p>[2] "Had nothing further than my own safety or pleasure been at stake no argument could have prevailed upon me to turn away the one creature upon Barsoom that had never failed in a demonstration of affection and loyalty; but as I would willingly have offered my life in the service of her in search of whom I was about to challenge the unknown dangers of this, to me, mysterious city, I could not permit even Woola's life to threaten the success of my venture, much less his momentary happiness, for I doubted not he soon would forget me."</p> <p>[3] "My quarters would have been suitable for housing the greatest of earthly emperors, but to these queer creatures nothing about a building appealed to them but its size and the enormity of its chambers; the larger the building, the more desirable; and so Tal Hajus occupied what must have been an enormous public building, the largest in the city, but entirely unfitted for residence purposes; the next largest was reserved for Lorquas Ptomei, the next for the jed of a lesser rank, and so on to the bottom of the list of five jeds."</p> <p>[4] "Ah, gladly would I give up my life a thousand times could I only hear them once again; but I could not then give even a second to the rapture of her sweet embrace, and pressing my lips to hers for the first time, I picked her up bodily and tossed her to her seat behind Sola again, commanding the latter in peremptory tones to hold her there by</p>

force, and then, slapping the thoat upon the flank, I saw them borne away; Dejah Thoris struggling to the last to free herself from Sola's grasp."

[5] "I can see her shining in the sky through the little window by my desk, and tonight she seems calling to me again as she has not called before since that long dead night, and I think I can see, across that awful abyss of space, a beautiful black-haired woman standing in the garden of a palace, and at her side is a little boy who puts his arm around her as she points into the sky toward the planet Earth, while at their feet is a huge and hideous creature with a heart of gold."

[6] "When we approached these points a warrior would be sent far ahead with a powerful field glass, and if no great body of red Martian troops was in sight we would advance as close as possible without chance of being seen and then camp until dark, when we would slowly approach the cultivated tract, and, locating one of the numerous, broad highways which cross these areas at regular intervals, creep silently and stealthily across to the arid lands upon the other side."

[7] "His father, Than Kosis, Jeddak of Zodanga, has made her voluntary marriage to his son the price of peace between our countries, but Tardos Mors will not accede to the demands and has sent word that he and his people would rather look upon the dead face of their princess than see her wed to any than her own choice, and that personally he would prefer being engulfed in the ashes of a lost and burning Helium to joining the metal of his house with that of Than Kosis."

[8] "I directed Sola to proceed with Dejah Thoris along one of the less frequented avenues to the southern boundary of the city, where I would overtake them with the thoats as quickly as possible; then, leaving them to gather what food, silks, and furs we were to need, I slipped quietly to the rear of the first floor, and entered the courtyard, where our animals were moving restlessly about, as was their habit, before settling down for the night."

[9] "That she had dropped into the last coma which would end in death, if the air supply remained unreplenished, I well knew, and so, throwing caution to the winds, I flung overboard everything but the engine and compass, even to my ornaments, and lying on my belly along the deck with one hand on the steering wheel and the other pushing the speed lever to its last notch I split the thin air of dying Mars with the speed of a meteor."

[10] "Without waiting for a further invitation I bolted up the winding runway which led to the second floor, and entering a great chamber at the front of the building was greeted by the frenzied Woola, who threw his great carcass upon me, nearly hurling me to the floor; the poor old fellow was so glad to see me that I thought he would devour me, his head split from ear to ear, showing his three rows of tusks in his hobgoblin smile."

b. Read Introduction to Text Analytics.docx. Follow the Rubric!! You will be graded according to the Rubric!!

First, we need to set our directory. I have called it “APrincess” and the text document is “APrincess.txt” contained therein.



Install the package “tm” which contains some of the methods we need.

Installing Essential R Packages for Text Mining and Sentiment Analysis

```
1 install.packages("tm")
2 install.packages("tokenizers")
3 install.packages("wordcloud")
4 install.packages("syuzhet")
5
91:1 (Top Level) ▾
Console Background Jobs ✘
R 4.4.2 · ~/Documents/Introduction to Big Data and Analytics_CSCI_6444/Project_3/Chapters/ ↵

The downloaded binary packages are in
  /var/folders/lb/j5d5pc257vd0tg32xf15zy5h0000gn/T//Rtmp8tqXhi/downloaded_packages
> library(tokenizers)
> library(stringi)
> library(wordcloud)
Loading required package: RColorBrewer
> library(syuzhet)
> library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':
  filter, lag

The following objects are masked from 'package:base':
  intersect, setdiff, setequal, union

> library(tm)
Loading required package: NLP
```

A VCorpus is a collection of documents. In our case, we will have just one document. A VCorpus is kept entirely in memory. VCorpus stands for Volatile Corpora.

The “.” means use the current directory, which we set by setwd(...). It will all .txt files into the VCorpus. We ignore case so that we don’t have to worry about and lower case.

```

> TextObj <- VCorpus(DirSource(".", ignore.case=FALSE, mode="text"))
> TextObj
<<VCorpus>>
Metadata: corpus specific: 0, document level (indexed): 0
Content: documents: 16
>

```

If we print Boiling, we get:

The creation of a text corpus in R using the `tm` package. The `VCorpus` function reads 16 text documents from the current directory into a corpus object named `TextObj`.

```

> str(TextObj)
Classes 'VCorpus', 'Corpus'  hidden list of 3
 $ content:List of 16
   ..$ :List of 2
   ...$ content: chr [1:267] "" "ON THE ARIZONA HILLS" "" ...
   ...$ meta   :List of 7
     ...$ author      : chr(0)
     ...$ timestamp: POSIXlt[1:1], format: "2025-04-20 23:58:07"
     ...$ description : chr(0)
     ...$ heading     : chr(0)
     ...$ id          : chr "Chapter1.txt"
     ...$ language    : chr "en"
     ...$ origin      : chr(0)
     ...- attr(*, "class")= chr "TextDocumentMeta"
     ...- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
   ..$ :List of 2
   ...$ content: chr [1:358] "" "CHAMPION AND CHIEF" "" ...
   ...$ meta   :List of 7
     ...$ author      : chr(0)
     ...$ timestamp: POSIXlt[1:1], format: "2025-04-20 23:58:07"
     ...$ description : chr(0)
     ...$ heading     : chr(0)
     ...$ id          : chr "Chapter10.txt"
     ...$ language    : chr "en"
     ...$ origin      : chr(0)
     ...- attr(*, "class")= chr "TextDocumentMeta"
     ...- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"

```

The `str(TextObj)` output reveals the internal structure of a `VCorpus` object, showing that it contains 16 plain text documents. Each document includes content and metadata like filename (`id`), language, and timestamp.

```
> TextFromCorpus <- TextObj[[1]]  
> TextFromCorpus  
<<PlainTextDocument>>  
Metadata: 7  
Content: chars: 13962
```

The output shows that `TextFromCorpus` is the first document extracted from the corpus, represented as a `PlainTextDocument`. It contains 13,962 characters and includes 7 metadata fields.

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

The output displays the content of the first document in the corpus, showing lines of text from the story titled "ON THE ARIZONA HILLS". It contains a personal narrative from an old man reflecting on life, death, and his mysterious past.

Creating VCorpus for the data

The screenshot shows an RStudio interface. In the top-left corner, there is a code editor window with several lines of R code. Lines 83 through 89 are visible, creating a VCorpus object from a directory of files. Line 91 shows the command to print the object to the console. Below the code editor is a tab bar with 'Console' and 'Background Jobs'. The 'Console' tab is active, displaying the output of the str(TextObj) command. The output shows that TextObj is a 'VCorpus' object containing 16 chapters, each represented as a 'TextDocument' with various metadata fields like author, timestamp, and language.

```
83 #----- create a VCorpus -----#
84 setwd('/Users/rohithyv/Documents/Introduction to Big Data and Analytics_CSCI_6444/Project_3/Chapters')
85 getwd()
86 TextObj <- VCorpus(DirSource("."), ignore.case=FALSE, mode="text")
87 str(TextObj)
88
89
91:1 (Top Level) ▾
```

Console Background Jobs

R 4.4.2 · ~/Documents/Introduction to Big Data and Analytics_CSCI_6444/Project_3/Chapters/ ↗

```
> str(TextObj)
Classes 'VCorpus', 'Corpus' hidden list of 3
$ content:List of 16
..$ :List of 2
... ..$ content: chr [1:267] "" "ON THE ARIZONA HILLS" "" ""
... ..$ meta   :List of 7
... ...$ author      : chr(0)
... ...$ timestamp: POSIXlt[1:1], format: "2025-04-20 23:58:07"
... ...$ description : chr(0)
... ...$ heading     : chr(0)
... ...$ id         : chr "Chapter1.txt"
... ...$ language    : chr "en"
... ...$ origin      : chr(0)
... ...- attr(*, "class")= chr "TextDocumentMeta"
... ...- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
..$ :List of 2
... ..$ content: chr [1:358] "" "CHAMPION AND CHIEF" "" ""
... ..$ meta   :List of 7
... ...$ author      : chr(0)
... ...$ timestamp: POSIXlt[1:1], format: "2025-04-20 23:58:07"
... ...$ description : chr(0)
... ...$ heading     : chr(0)
... ...$ id         : chr "Chapter10.txt"
... ...$ language    : chr "en"
... ...$ origin      : chr(0)
... ...- attr(*, "class")= chr "TextDocumentMeta"
... ...- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
..$ :List of 2
... ..$ content: chr [1:258] "" "WITH DEJAH THORIS" "" ""
... ..$ meta   :List of 7
... ...$ author      : chr(0)
... ...$ timestamp: POSIXlt[1:1], format: "2025-04-20 23:58:07"
... ...$ description : chr(0)
... ...$ heading     : chr(0)
... ...$ id         : chr "Chapter11.txt"
... ...$ language    : chr "en"
... ...$ origin      : chr(0)
... ...- attr(*, "class")= chr "TextDocumentMeta"
... ...- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
```

This screenshot shows the structure of a VCorpus created from 16 chapter files using R. Each text document (e.g., "Chapter1.txt", "Chapter10.txt", "Chapter11.txt") contains content and associated metadata, such as timestamp and language.

Compute DTM (Document Term Matrix) and TDM (Term Document Matrix)

```
> # compute the DTM and TDM -----
> TextObjDTM <- DocumentTermMatrix(TextObj)
> inspect(TextObjDTM)
<<DocumentTermMatrix (documents: 16, terms: 9281)>>
Non-/sparse entries: 18907/129589
Sparsity : 87%
Maximal term length: 19
Weighting : term frequency (tf)
Sample :
Terms
Docs      and for from had his that the upon was with
Chapter1.txt 90 20 9 25 14 50 190 12 36 25
Chapter10.txt 113 34 10 41 42 50 193 18 54 26
Chapter11.txt 82 9 10 29 3 37 119 13 18 23
Chapter13.txt 63 17 6 29 8 39 131 15 21 29
Chapter14.txt 115 41 22 35 20 43 171 18 49 33
Chapter15.txt 108 36 34 25 20 40 215 26 37 18
Chapter16.txt 1051 227 220 301 216 338 2260 213 381 265
Chapter3.txt 92 18 14 19 31 26 166 20 36 20
Chapter4.txt 68 10 12 22 10 19 148 10 26 20
Chapter8.txt 81 15 21 24 3 11 184 22 32 11
```

This output shows a **DocumentTermMatrix** created from the corpus, containing 16 documents and 9,281 unique terms. The matrix displays term frequencies for common words like "and", "the", and "was" across sample chapters, highlighting Chapter16.txt as having significantly higher term counts.

```
> str(TextObjDTM)
List of 6
 $ i      : int [1:18907] 1 1 1 1 1 1 1 1 1 ...
 $ j      : int [1:18907] 126 130 131 139 142 188 197 208 212 218 ...
 $ v      : num [1:18907] 1 1 1 1 7 1 1 4 1 1 ...
 $ nrow   : int 16
 $ ncol   : int 9281
 $ dimnames:List of 2
   ..$ Docs : chr [1:16] "Chapter1.txt" "Chapter10.txt" "Chapter11.txt" "Chapter12.txt" ...
   ..$ Terms: chr [1:9281] "'gentleman'" "'til" "\'tis" "\'after" ...
 - attr(*, "class")= chr [1:2] "DocumentTermMatrix" "simple_triplet_matrix"
 - attr(*, "weighting")= chr [1:2] "term frequency" "tf"
```

Above Figure show computing DTM

The **DocumentTermMatrix** reveals word usage patterns across the 16 chapter files, with a sparsity of 87%, indicating many zero entries. It uses term frequency weighting, and the sample shows how often specific words (e.g., "the", "and", "was") appear in each chapter.

Docs = 16, which represents the number of chapters/documents in the corpus.

Terms = 9281, which represents the number of unique words (terms) across all documents.

Term-Document Matrix Summary for Text Corpus

```
> TextObjTDM <- TermDocumentMatrix(TextObj)
> TextObjTDM
<<TermDocumentMatrix (terms: 9281, documents: 16)>>
Non-/sparse entries: 18907/129589
Sparsity          : 87%
Maximal term length: 19
Weighting         : term frequency (tf)
```

Above Figure show computing TDM

Documents = 16 represents the number of chapters

Terms = 9281 represents the number of unique words

Non-/Sparse entries = 18907 non-zero entries and 129589 - 18907 = 110682 zero entries

Sparsity = 87% of the matrix is empty

Maximal Term Length = 19 (length of the longest term)

This output shows a **TermDocumentMatrix** (TDM) with **9,281 unique terms** across **16 documents**, using term frequency as the weighting method. The matrix has a sparsity of **87%**, indicating many zero entries due to rare terms in documents.

Converting VCorpus to Dataframe

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

This output shows the conversion of a text document from the corpus into a data frame, where each line of text becomes a row under the `content` column. It allows for easier manipulation, analysis, or visualization of the text data in tabular format.

Inspection of a Single PlainTextDocument Extracted from Corpus

```
> TextFromCorpus
<<PlainTextDocument>>
Metadata: 7
Content: chars: 13962
```

`TextFromCorpus` is a `PlainTextDocument` object containing 13,962 characters. It also includes 7 metadata fields describing the document's attributes.

Corpus Cleansing - Data wrangling

Removing numbers and punctuation from the text data and creating clean corpus

```
> #-----Corpus Cleansing - Data Wrangling -----#
> removeNumPunct <- function(x) {
+   gsub("[^[:alpha:][:space:]]+", "", x)
+ }
> textCl <- tm::tm_map(TextObj, content_transformer(removeNumPunct))
> textCl
<<VCorpus>>
Metadata: corpus specific: 0, document level (indexed): 0
Content: documents: 16
> str(textCl)
Classes 'VCorpus', 'Corpus' hidden list of 3
$ content:List of 16
..$ :List of 2
... ...$ content: chr [1:267] "" "ON THE ARIZONA HILLS" "" ...
... ...$ meta    :List of 7
... ... .$.author      : chr(0)
... ... .$.timestamp: POSIXlt[1:1], format: "2025-04-20 23:58:07"
... ... .$.description: chr(0)
... ... .$.heading     : chr(0)
... ... .$.id         : chr "Chapter1.txt"
... ... .$.language    : chr "en"
... ... .$.origin      : chr(0)
... ... -. attr(*, "class")= chr "TextDocumentMeta"
... ... -. attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
```

This output shows corpus cleansing using a custom removeNumPunct function to eliminate numbers and punctuation from the text. The cleaned corpus textCl still contains 16 documents, now transformed to include only alphabetic characters and spaces.

Converting all the characters to lowercase if any

```
> # -----converts all characters to lowercase if any -----#
> TextLow <- tm_map(textCl, tm::content_transformer(tolower))
> str(TextLow)
Classes 'VCorpus', 'Corpus' hidden list of 3
$ content:List of 16
..$ :List of 2
... ...$ content: chr [1:267] "" "on the arizona hills" "" ...
... ...$ meta   :List of 7
... ... ..$ author      : chr(0)
... ... ..$ timestamp: POSIXlt[1:1], format: "2025-04-20 23:58:07"
... ... ..$ description : chr(0)
... ... ..$ heading     : chr(0)
... ... ..$ id         : chr "Chapter1.txt"
... ... ..$ language    : chr "en"
... ... ..$ origin      : chr(0)
... ... ..- attr(*, "class")= chr "TextDocumentMeta"
... ...- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
..$ :List of 2
... ...$ content: chr [1:358] "" "champion and chief" "" ...
... ...$ meta   :List of 7
... ... ..$ author      : chr(0)
... ... ..$ timestamp: POSIXlt[1:1], format: "2025-04-20 23:58:07"
... ... ..$ description : chr(0)
... ... ..$ heading     : chr(0)
... ... ..$ id         : chr "Chapter10.txt"
... ... ..$ language    : chr "en"
... ... ..$ origin      : chr(0)
... ... ..- attr(*, "class")= chr "TextDocumentMeta"
... ...- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
```

This step converts all text in the corpus to lowercase using `tm_map()` with the `tolower` function. The updated corpus `TextLow` retains 16 documents, and the content now reflects uniform lowercase formatting (e.g., “on the arizona hills”).

Overview of Text Corpus Structure Using `inspect(TextLow)` in R

```
> inspect(TextLow)
<<VCorpus>>
Metadata: corpus specific: 0, document level (indexed): 0
Content: documents: 16

[[1]]
<<PlainTextDocument>>
Metadata: 7
Content: chars: 13711

[[2]]
<<PlainTextDocument>>
Metadata: 7
Content: chars: 18669

[[3]]
<<PlainTextDocument>>
Metadata: 7
Content: chars: 12812

[[4]]
<<PlainTextDocument>>
Metadata: 7
Content: chars: 10221

[[5]]
<<PlainTextDocument>>
Metadata: 7
Content: chars: 12200
```

The output shows the `TextLow` corpus after transformation, containing 16 documents with their respective character counts. Each document is a `PlainTextDocument` with 7 metadata fields and varying lengths, such as 13,711 and 18,669 characters.

Extracting and Accessing a Chapter 1 Document from a Text Corpus

```
> TextLowChp1 <- TextLow[[1]]
> TextLowChp1[1]
$content
[1] ""
[2] "on the arizona hills"
[3] ""
[4] ""
[5] "i am a very old man how old i do not know possibly i am a hundred"
[6] "possibly more but i cannot tell because i have never aged as other"
[7] "men nor do i remember any childhood so far as i can recollect i have"
[8] "always been a man a man of about thirty i appear today as i did"
[9] "forty years and more ago and yet i feel that i cannot go on living"
[10] "forever that some day i shall die the real death from which there is"
[11] "no resurrection i do not know why i should fear death i who have"
[12] "died twice and am still alive but yet i have the same horror of it as"
[13] "you who have never died and it is because of this terror of death i"
[14] "believe that i am so convinced of my mortality"
[15] ""
[16] "and because of this conviction i have determined to write down the"
[17] "story of the interesting periods of my life and of my death i cannot"
[18] "explain the phenomena i can only set down here in the words of an"
[19] "ordinary soldier of fortune a chronicle of the strange events that"
[20] "befell me during the ten years that my dead body lay undiscovered in an"
[21] "arizona cave"
[22] ""
[23] "i have never told this story nor shall mortal man see this manuscript"
[24] "until after i have passed over for eternity i know that the average"
[25] "human mind will not believe what it cannot grasp and so i do not"
[26] "purpose being pilloried by the public the pulpit and the press and"
```

This output displays the content of the first cleaned and lowercased chapter from the corpus `TextLow`, stored in `TextLowChp1`. All text has been converted to lowercase with punctuation removed, making it ready for further natural language processing tasks.

Computing DTM (Document Term Matrix) for the above cleaned data

```
> #---- compute DTM using TextLow -----#
> TextLow_DTM <- DocumentTermMatrix(TextLow)
> str(TextLow_DTM)
List of 6
$ i      : int [1:15835] 1 1 1 1 1 1 1 1 1 ...
$ j      : int [1:15835] 5 8 45 53 62 64 69 106 108 ...
$ v      : num [1:15835] 1 7 1 1 4 1 1 1 1 3 ...
$ nrow   : int 16
$ ncol   : int 6499
$ dimnames:List of 2
..$ Docs : chr [1:16] "Chapter1.txt" "Chapter10.txt" "Chapter11.txt" "Chapter12.txt" ...
..$ Terms: chr [1:6499] "abandoning" "abash" "abashed" "ability" ...
- attr(*, "class")= chr [1:2] "DocumentTermMatrix" "simple_triplet_matrix"
- attr(*, "weighting")= chr [1:2] "term frequency" "tf"
```

The image shows the structure of a DocumentTermMatrix (TextLow_DTM) created using the tm package in R. It consists of 16 documents and 6499 unique terms, stored as a sparse matrix with term frequency ("tf") weighting.

```
> inspect(TextLow_DTM)
<<DocumentTermMatrix (documents: 16, terms: 6499)>>
Non-/sparse entries: 15835/88149
Sparsity          : 85%
Maximal term length: 17
Weighting          : term frequency (tf)
Sample             :
Terms
Docs      and but for had that the upon was which with
Chapter1.txt 93 13 20 27 50 190 12 38 20 25
Chapter10.txt 117 20 34 41 50 193 19 54 16 26
Chapter11.txt 89 14 10 29 38 122 13 19 13 23
Chapter13.txt 65 11 19 29 40 131 15 21 11 30
Chapter14.txt 118 20 43 35 45 172 18 50 10 34
Chapter15.txt 116 31 36 25 41 216 26 37 22 19
Chapter16.txt 1093 213 230 302 347 2274 214 386 190 266
Chapter3.txt 94 17 18 19 26 166 20 37 34 20
Chapter4.txt 70 21 10 22 19 148 11 26 17 20
Chapter8.txt 84 7 15 24 11 184 22 32 17 11
```

This output displays a sample from the DocumentTermMatrix showing term frequencies for 16 documents across common words like "and", "the", and "was". It reveals high sparsity (85%) and highlights Chapter16.txt as having significantly higher term counts compared to others.

	abandoning	abash	abashed	ability	able	ablest	aboard	about	above	abreast	abruptly	absence	absences	absent	absolute	absolutely	absorbed	abundantly	abuse	aby
Chapter1.txt	0	0	0	0	1	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0
Chapter10.txt	0	0	0	2	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0
Chapter11.txt	0	0	0	0	0	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0
Chapter12.txt	0	0	0	0	0	0	1	0	2	1	0	0	0	0	0	1	1	0	0	0
Chapter13.txt	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0
Chapter14.txt	0	0	0	1	1	0	0	4	1	0	0	0	0	0	0	0	0	0	0	0
Chapter15.txt	0	0	0	0	0	0	0	2	0	1	0	1	1	0	2	0	0	0	0	2
Chapter16.txt	0	1	1	5	7	0	1	45	26	2	0	3	0	0	2	3	1	0	0	0
Chapter2.txt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Chapter3.txt	0	0	0	0	0	0	0	0	8	3	0	0	0	0	0	0	1	0	0	0
Chapter4.txt	0	0	0	2	0	0	0	10	1	0	1	0	0	0	0	0	0	0	0	0
Chapter5.txt	0	0	0	1	1	0	0	5	0	0	0	0	0	0	0	0	0	0	1	0
Chapter6.txt	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
Chapter7.txt	0	0	0	0	1	0	0	8	0	1	0	0	0	0	0	0	0	0	0	0
Chapter8.txt	0	0	0	0	0	0	0	0	2	3	0	0	0	0	0	0	0	0	0	0
Chapter9.txt	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0

This table shows a detailed view of the Document-Term Matrix with frequency counts of selected terms (e.g., "abandoning", "about", "above") across multiple chapters. Chapter16.txt again stands out with significantly higher word counts, especially for "about" (45) and "above" (26).

Removing Common Stop words results in 9812 words

```
> #----Removing Stop Words ----#
> Swords <- c(tm::stopwords("english"))
> Swords
[1] "i"          "me"          "my"          "myself"       "we"          "our"         "ours"
[8] "ourselves" "you"         "your"        "yours"        "yourself"    "yourselves" "he"
[15] "him"        "his"         "himself"     "she"          "her"         "hers"        "herself"
[22] "it"          "its"         "itself"      "they"         "them"        "their"      "theirs"
[29] "themselves" "what"        "which"       "who"          "whom"        "this"        "that"
[36] "these"       "those"       "am"          "is"           "are"         "was"         "were"
[43] "be"          "been"        "being"       "have"        "has"         "had"         "having"
[50] "do"          "does"        "did"         "doing"       "would"      "should"     "could"
[57] "ought"       "i'm"         "you're"      "he's"         "she's"       "it's"        "we're"
[64] "they're"     "i've"        "you've"      "we've"       "they've"    "i'd"         "you'd"
[71] "he'd"        "she'd"       "we'd"        "they'd"      "i'll"        "you'll"     "he'll"
[78] "she'll"      "we'll"       "they'll"     "isn't"       "aren't"      "wasn't"     "weren't"
[85] "hasn't"      "haven't"    "hadn't"      "doesn't"    "don't"       "didn't"     "won't"
[92] "wouldn't"   "shan't"     "shouldn't"   "can't"       "cannot"     "couldn't"   "mustn't"
[99] "let's"        "that's"      "who's"       "what's"      "here's"     "there's"    "when's"
[106] "where's"    "why's"      "how's"       "a"           "an"         "the"        "and"
[113] "but"         "if"          "or"          "because"    "as"         "until"      "while"
[120] "of"          "at"          "by"          "for"         "with"       "about"      "against"
[127] "between"    "into"       "through"    "during"     "before"     "after"      "above"
[134] "below"       "to"          "from"       "up"          "down"       "in"         "out"
[141] "on"          "off"         "over"       "under"      "again"      "further"    "then"
[148] "once"        "here"       "there"      "when"       "where"      "why"        "how"
[155] "all"         "any"         "both"        "each"       "few"        "more"       "most"
[162] "other"       "some"       "such"        "no"          "nor"        "not"        "only"
[169] "own"         "same"       "so"          "than"       "too"        "very"
```

This output shows a list of English stop words retrieved using tm::stopwords("english") in R. These common words (e.g., "the", "is", "and", "of") are typically removed during text preprocessing to focus on more meaningful terms.

Output of Document After Stopword Removal Using `tm_map`

```
> TextlowSwordsRemoved <- tm::tm_map(TextLow, tm::removeWords, Swords)
> tm::inspect(TextlowSwordsRemoved[[1]])
<<PlainTextDocument>>
Metadata: 7
Content: chars: 9812

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

never told story shall mortal man see manuscript
passed eternity know average
human mind will believe grasp
purpose pilloried public pulpit press
held colossal liar telling simple truths
day science will substantiate possibly suggestions
gained upon mars knowledge can set
chronicle will aid earlier understanding mysteries
sister planet mysteries longer mysteries
```

This output displays the content of a text document after stop words have been removed using the `tm_map()` function in R. The resulting text is more focused, retaining only meaningful words like "death", "mysteries", and "knowledge" for improved analysis.

Removing more custom stopwords results in 9648 words

```
> customWords <- c("isnt", "mustnt", "shes", "cant", "well", "hed", "quite",
+                  "youve", "theyre", "whatever", "wasnt", "couldnt", "youll",
+                  "now", "ive", "theres", "never", "theres", "weve", "shant",
+                  "rather", "theyll", "id", "even", "may", "might", "lets", "shed",
+                  "hadnt", "doesnt", "shell", "whats", "youre", "whens", "youll",
+                  "arent", "wheres", "might", "cant", "couldnt", "ill", "were",
+                  "hes", "its", "will", "too", "upon", "how", "nothing", "heres",
+                  "thats", "whos", "just", "much", "hows", "one", "whys", "able",
+                  "wont", "wouldnt", "were", "whens", "whats", "theyve")
> #Removing custom Stop words additionally
> TextlowCSwordsRemoved <- tm::tm_map(TextlowSwordsRemoved, tm::removeWords, customWords)
> tm::inspect(TextlowCSwordsRemoved[[1]])
```

<<PlainTextDocument>>

Metadata: 7

Content: chars: 9648

arizona hills

```
old man old know possibly hundred
possibly tell 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
      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

told story shall mortal man see manuscript
      passed eternity know average
human mind believe grasp
purpose pilloried public pulpit press
held colossal liar telling simple truths
      day science substantiate possibly suggestions
      gained mars knowledge can set
chronicle aid earlier understanding mysteries
      sister planet mysteries longer mysteries
```

This output shows the document after removing both standard and custom stop words using `tm::removeWords()` in R. Additional filtering has made the text more concise by eliminating user-defined irrelevant terms, enhancing focus on significant content.

Corpus After Removing Standard and Custom Stop Words

```
> TextlowCSwordsRemoved
<<VCorpus>>
Metadata: corpus specific: 0, document level (indexed): 0
Content: documents: 16
```

This output confirms that TextlowCSwordsRemoved is a VCorpus containing 16 documents. All standard and user-defined stop words have been successfully removed, preparing the corpus for further text mining.

Term-Document Matrix After Stop Word Removal

```
> # Creating TDM and applying findfrequencyTerms freq = 5,25 and 50
> TextWithNoSW_TDM <- tm::TermDocumentMatrix(TextlowCSwordsRemoved)
> TextWithNoSW_TDM
<<TermDocumentMatrix (terms: 6381, documents: 16)>>
Non-/sparse entries: 14430/87666
Sparsity           : 86%
Maximal term length: 17
Weighting          : term frequency (tf)
```

This output displays a TermDocumentMatrix with 6381 terms across 16 documents, created after removing standard and custom stop words. The matrix has 86% sparsity and uses term frequency (tf) as the weighting scheme, ready for frequency-based analysis.

Inspecting TDM of text data post removing stop words again

```
> # Inspecting TDM of text data post removing stop words again
> tm::inspect(TextWithNoSW_TDM)
<<TermDocumentMatrix (terms: 6381, documents: 16)>>
Non-/sparse entries: 14430/87666
Sparsity : 86%
Maximal term length: 17
Weighting : term frequency (tf)
Sample :
    Docs
Terms   Chapter1.txt Chapter10.txt Chapter11.txt Chapter13.txt Chapter14.txt Chapter15.txt
dejah      0          5          17         10         20          7
great      1          7          5          9          4          9
green      0          6          2          4          5          8
helium     0          3          1          1          0          0
little     4          4          4          6          8          5
martian    0          13         3          9          5          4
sola       0          6          11         6          10         10
thoris     0          5          17         10         20          7
toward     3          9          1          2          3          3
warriors   2          4          1          7          3          4
    Docs
Terms   Chapter16.txt Chapter3.txt Chapter4.txt Chapter8.txt
dejah    113          0          0          0
great    100          3          3          6
green    50           1          2          8
helium   98           0          0          0
little   52           11         3          4
martian  29           6          9          7
```

This table displays term frequencies from the cleaned TermDocumentMatrix, showing how often specific words like "dejah", "martian", and "thoriss" appear in each chapter. Chapter16.txt contains the highest frequencies, indicating it is rich in key content terms.

Frequent Terms Extracted from Cleaned Text

```
> textFreq5 <- tm::findFreqTerms(TextWithNoSW_TDM, lowfreq = 5)
> textFreq25 <- tm::findFreqTerms(TextWithNoSW_TDM, lowfreq = 25)
> textFreq25
[1] "across"      "advanced"     "air"          "almost"       "alone"        "also"
[7] "always"      "among"       "animals"      "another"      "arms"         "asked"
[13] "away"        "back"        "barsoom"     "battle"       "beautiful"    "behind"
[19] "believe"     "body"        "brought"     "building"    "buildings"   "came"
[25] "can"         "carter"      "cave"        "chamber"     "chieftain"   "chieftains"
[31] "city"        "close"       "come"        "community"  "craft"        "darkness"
[37] "day"         "days"        "dead"        "death"       "dejah"       "direction"
[43] "discovered"  "earth"       "earthly"     "either"      "end"         "entered"
[49] "entire"      "entirely"    "escape"      "ever"        "every"       "evidently"
[55] "except"      "eyes"        "face"        "fact"        "far"         "fear"
[61] "feet"        "fell"        "fight"       "fighting"   "finally"     "first"
[67] "five"        "floor"       "followed"   "found"       "four"        "full"
[73] "given"        "great"       "green"       "ground"     "hajus"       "hand"
[79] "hands"        "head"        "heard"       "heart"      "held"        "helium"
[85] "high"         "hills"       "however"    "huge"        "hundred"    "incubator"
[91] "jeddak"       "john"        "kan"         "kantos"    "knew"        "know"
[97] "kosis"        "large"       "last"        "later"      "lay"         "learned"
[103] "left"         "less"        "life"        "light"      "like"        "little"
[109] "long"         "lorquas"    "lost"        "love"       "low"         "made"
[115] "make"         "man"         "many"       "mars"        "martian"    "martians"
[121] "meet"         "men"         "metal"      "mighty"     "mine"       "moment"
[127] "morning"     "must"        "name"       "nearly"     "new"        "night"
[133] "old"          "open"        "others"     "palace"    "passed"     "people"
[139] "place"        "plaza"       "point"      "poor"        "princess"   "prisoner"
[145] "ptomel"       "quarters"   "quickly"    "reach"      "reached"    "red"
[151] "replied"      "return"     "returned"  "roof"       "room"       "said"
[157] "sarkoja"     "saw"         "see"        "seemed"    "seen"        "several"
[163] "shall"        "short"      "side"       "sight"      "silks"       "since"
[169] "small"        "sola"        "soon"       "still"      "stood"      "strange"
[175] "take"         "tal"         "tarkas"    "tars"       "ten"        "thark"
[181] "tharks"       "thing"      "thoats"    "thoris"    "though"    "thought"
```

This output shows words appearing at least 25 times (textFreq25) in the Term-Document Matrix after removing stop words. Terms like "dejah", "great", "martian", and "sola" are among the most prominent, highlighting recurring themes in the text corpus.

High-Frequency Terms (≥ 50 Occurrences) in Text Corpus

```
> textFreq50 <- tm::findFreqTerms(TextWithNoSW_TDM, lowfreq = 50)
> textFreq50
[1] "among"    "another"   "back"      "barsoom"   "behind"    "body"      "building"  "came"
[9] "can"       "carter"    "city"      "day"       "days"      "dead"      "death"     "dejah"
[17] "ever"     "eyes"      "face"      "far"       "feet"      "first"     "floor"     "found"
[25] "great"    "green"     "ground"   "hajus"     "helium"    "hundred"   "jeddak"    "john"
[33] "know"     "left"      "life"      "little"    "long"      "made"      "man"       "mars"
[41] "martian"  "martians"  "men"       "must"      "night"     "people"    "reached"   "red"
[49] "said"     "saw"       "see"       "seemed"    "side"      "sola"      "soon"      "strange"
[57] "tal"       "tarkas"    "tars"      "thoris"    "though"    "three"     "time"     "toward"
[65] "turned"   "two"       "warriors" "way"       "within"    "without"   "women"    "zodanga"
> |
```

This output lists the most prominent terms that occur at least 50 times in the cleaned text corpus, such as "dejah", "martian", "helium", and "thoriss". These keywords are central to the document's theme and may guide topic modeling or keyword extraction.

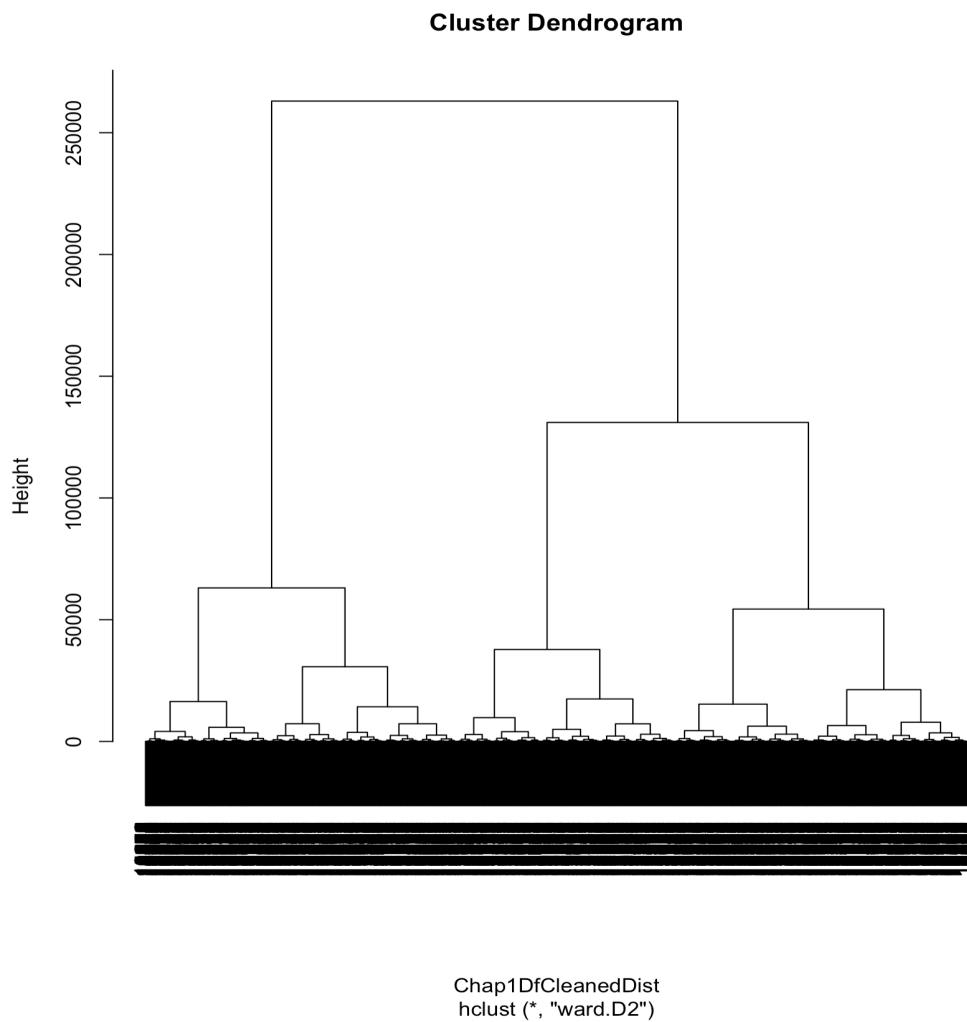
Hierarchical Clustering Structure for Chapter1 Data

```
> # Create Dendrogram
> Chap1DfCleaned <- as.data.frame(TextWithNoSW_TDM[[1]])
> Chap1DfCleanedDist <- dist(Chap1DfCleaned)
> Chap1Df_DG <- hclust(Chap1DfCleanedDist, method = "ward.D2")
> str(Chap1Df_DG)
List of 7
 $ merge      : int [1:14429, 1:2] -1 -5587 -12180 -3 -1786 -2487 -3086 -3772 -4684 -5595 ...
 $ height     : num [1:14429] 0 0 0 0 0 0 0 0 0 ...
 $ order      : int [1:14430] 10939 9623 5552 2475 3072 10940 10941 14427 13970 13324 ...
 $ labels     : NULL
 $ method     : chr "ward.D2"
 $ call        : language hclust(d = Chap1DfCleanedDist, method = "ward.D2")
 $ dist.method: chr "euclidean"
 - attr(*, "class")= chr "hclust"
```

This output represents the internal structure of a dendrogram created using hierarchical clustering (hclust) on cleaned term-document data from Chapter1. It uses the Ward.D2 method with Euclidean distance to cluster similar terms based on frequency.

Plotting Word Frequency Data Frame

```
> plot(Chap1Df_DG)
```



Dendrogram of Terms from Chapter1 Using Ward's Method

This dendrogram visually represents the hierarchical clustering of terms from Chapter1 based on their frequency similarity. The vertical axis shows the height (dissimilarity), with closely linked terms forming lower branches, while more distinct clusters merge higher up.

Not really helpful.

Let's try to eliminate some more sparse words:

```
173 #---Removing more words to reduce sparsity---#
174
175 moreWordsToRemove <- c("found", "next", "toward", "back", "far", "almost", "quickly",
176           "last", "little", "away", "though", "like", "day", "within",
177           "many", "thought", "come", "second", "new", "thing", "mighty",
178           "behind", "yet", "still", "konw", "comes", "another", "knew",
179           "saw", "ever", "first", "instant", "among", "two", "seen")
180 TextlowCSwordsRemoved_new <- tm::tm_map(TextlowCSwordsRemoved, tm::removeWords, moreWordsToRemove)
181 str(TextlowCSwordsRemoved_new)
```

Above plot is not helpful, hence let's delete some more words

Removing Additional Words to Improve Text Corpus Sparsity

```
> #---Removing more words to reduce sparsity---#
>
> moreWordsToRemove <- c("found", "next", "toward", "back", "far", "almost", "quickly",
+           "last", "little", "away", "though", "like", "day", "within",
+           "many", "thought", "come", "second", "new", "thing", "mighty",
+           "behind", "yet", "still", "konw", "comes", "another", "knew",
+           "saw", "ever", "first", "instant", "among", "two", "seen")
> TextlowCSwordsRemoved_new <- tm::tm_map(TextlowCSwordsRemoved, tm::removeWords, moreWordsToRemove)
> str(TextlowCSwordsRemoved_new)
Classes 'VCorpus', 'Corpus' hidden list of 3
$ content:List of 16
..$ :List of 2
... ..$ content: chr [1:267] "" " arizona hills" "" ...
... ..$ meta :List of 7
...   ..$ author      : chr(0)
...   ..$ timestamp: POSIXlt[1:1], format: "2025-04-26 18:52:39"
...   ..$ description : chr(0)
...   ..$ heading     : chr(0)
...   ..$ id          : chr "Chapter1.txt"
...   ..$ language    : chr "en"
...   ..$ origin      : chr(0)
...   ...- attr(*, "class")= chr "TextDocumentMeta"
...   ...- attr(*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"
```

Extra filler words like "found," "next," "toward," etc., are removed from the corpus using `tm_map` to further clean the text. This step reduces noise and improves the quality of subsequent text mining and analysis.

Let's create TDM for the above data and a Dendogram

Creating Term-Document Matrix and Hierarchical Clustering

```
> #---Creating TDM and new Dendogram----#
> TextlowCSwordsRemoved_newTDM <- tm::TermDocumentMatrix(TextlowCSwordsRemoved_new)
> TextlowCSwordsRemoved_newTDM
<<TermDocumentMatrix (terms: 6347, documents: 16)>>
Non-/sparse entries: 14035/87517
Sparsity : 86%
Maximal term length: 17
Weighting : term frequency (tf)
> Chap1DfCleaned_new <- as.data.frame(TextlowCSwordsRemoved_newTDM[[1]])
> Chap1DfCleanedDist_new <- dist(Chap1DfCleaned_new)
> Chap1DF_newDG <- hclust(Chap1DfCleanedDist_new, method = "ward.D2")
> str(Chap1DF_newDG)
List of 7
$ merge      : int [1:14034, 1:2] -1 -5402 -11873 -3 -1732 -2406 -2985 -3645 -4527 -5410 ...
$ height     : num [1:14034] 0 0 0 0 0 0 0 0 0 ...
$ order       : int [1:14035] 10675 9405 5367 2395 2972 10676 10677 14033 13595 12970 ...
$ labels     : NULL
$ method     : chr "ward.D2"
$ call        : language hclust(d = Chap1DfCleanedDist_new, method = "ward.D2")
$ dist.method: chr "euclidean"
- attr(*, "class")= chr "hclust"
> plot(Chap1DF_newDG)
```

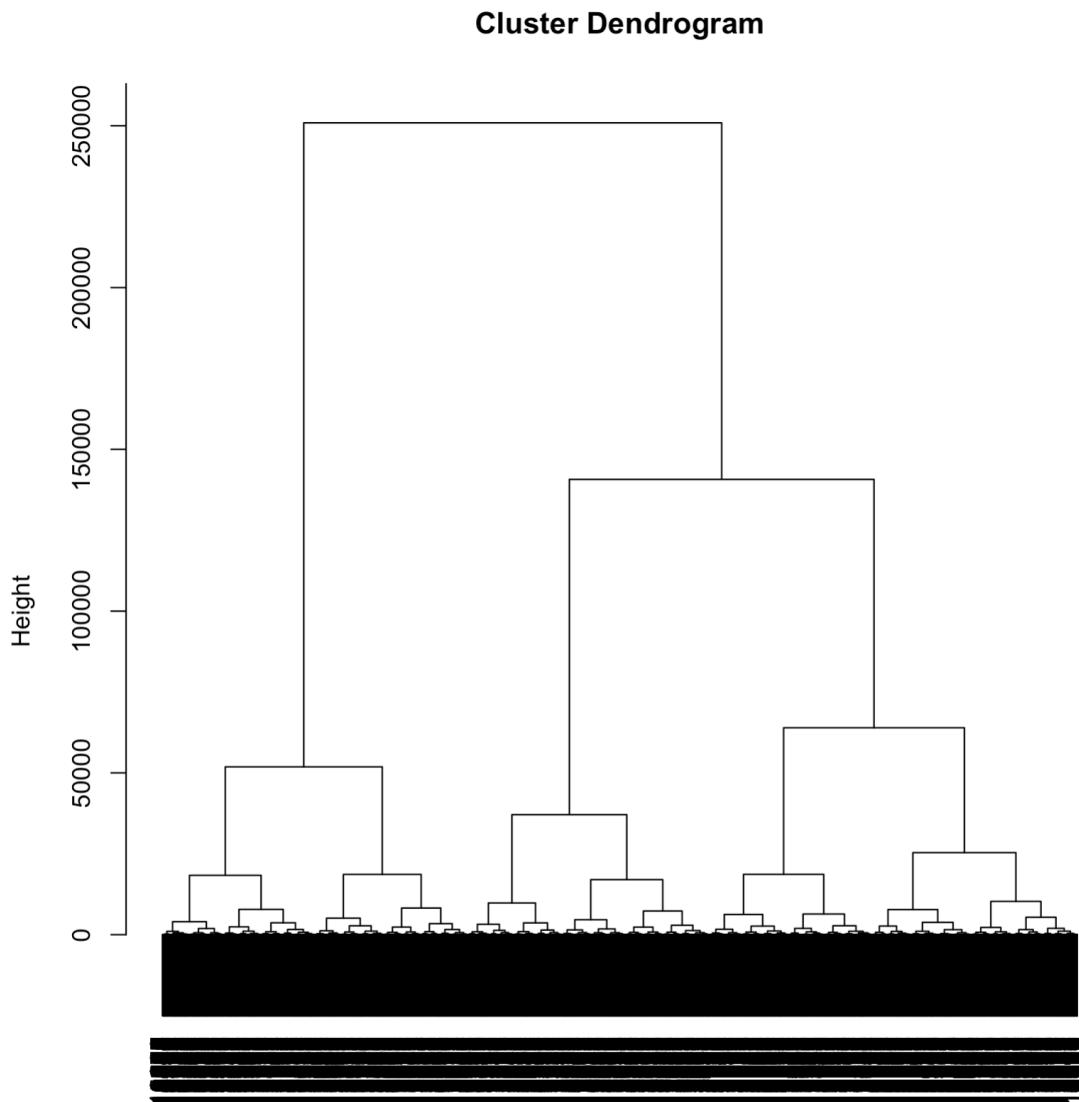
A Term-Document Matrix (TDM) was created to quantify term frequencies, followed by hierarchical clustering using the Ward.D2 method to generate a dendrogram structure for text grouping analysis.

Word Cloud

Set of words from Chapter 1

```
> #----Creating word cloud-----#
>
> Chap1TermFreq <- tm::termFreq(TextLowCSwordsRemoved_new[[1]])
> chap1Words <- names(Chap1TermFreq)
> chap1Words
[1] "account"      "accurately"    "across"        "acted"       "acts"
[6] "advent"        "adventures"   "afternoon"     "aged"        "ago"
[11] "agreed"       "ahead"        "aid"          "aim"         "alive"
[16] "alone"         "already"      "alternative"  "always"      "animals"
[21] "antelope"     "anything"    "apaches"      "apartments"  "appear"
[26] "apprehension" "arizona"      "arm"          "armed"       "arming"
[31] "army"          "arose"        "around"       "arrows"      "assure"
[36] "attacked"      "attempt"      "attention"    "attraction"  "attributed"
[41] "average"       "await"        "backward"     "balls"       "bathed"
[46] "beast"         "beautiful"   "became"       "become"      "befell"
[51] "believe"       "belt"         "belts"        "best"        "bestowed"
[56] "better"         "bidding"      "body"         "borne"       "bottom"
[61] "bows"           "braves"       "brief"        "bright"      "brisk"
[66] "bristling"     "broad"        "broke"        "brought"    "burros"
[71] "came"           "camp"         "can"          "canteen"     "canter"
[76] "captain"       "captains"    "capture"      "carbine"     "careless"
[81] "carry"          "carter"       "cartridge"   "cartridges"  "casually"
[86] "catch"          "catching"    "cavalry"      "cave"        "center"
[91] "certain"        "chamber"     "chances"      "characteristic" "charging"
[96] "chase"          "childhood"   "chronicle"   "civil"       "civilization"
[101] "claim"          "clear"        "cliff"        "close"       "clustered"
[106] "clutches"       "colossal"    "colt"         "commenced"   "commission"
[111] "company"        "comparatively" "confederate"  "conjured"    "consternation"
[116] "constituted"    "constitutes"  "continue"     "continued"   "continuing"
[121] "continuously"  "conviction"  "convince"     "convinced"   "country"
[126] "course"          "cowardice"   "creeping"    "crept"       "crude"
[131] "cunning"        "customary"   "cut"          "dangerous"  "dangers"
```

Cluster Dendrogram of Text Data Using Ward's Method



The hierarchical clustering dendrogram visualizes how documents group based on text similarity, with major divisions indicating distinct content clusters at higher heights.

Creating a color Palette and word cloud using words and their frequencies

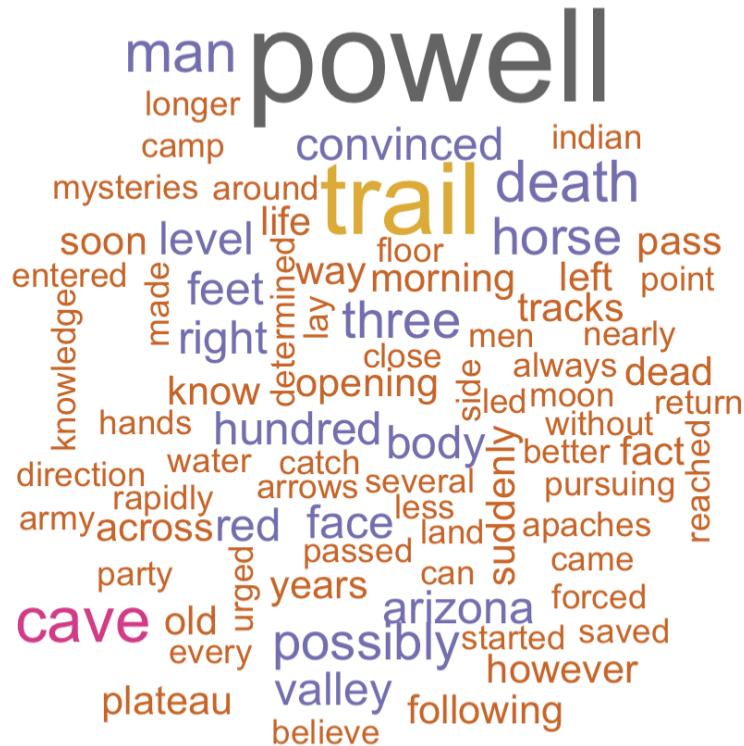
```
> # Try 1: Using Dark2 – Best for white backgrounds, high contrast  
> cp1 <- brewer.pal(8, "Dark2")  
> str(cp1)  
chr [1:8] "#1B9E77" "#D95F02" "#7570B3" "#E7298A" "#66A61E" "#E6AB02" "#A6761D" "#666666"
```

Section 1 – Action:

Selected the "Dark2" color palette using `brewer.pal` to ensure high contrast on a white background.

Section 2 – Result:

Generated a word cloud (`chap1WC1`) with customized dark-themed colors for better visual impact.

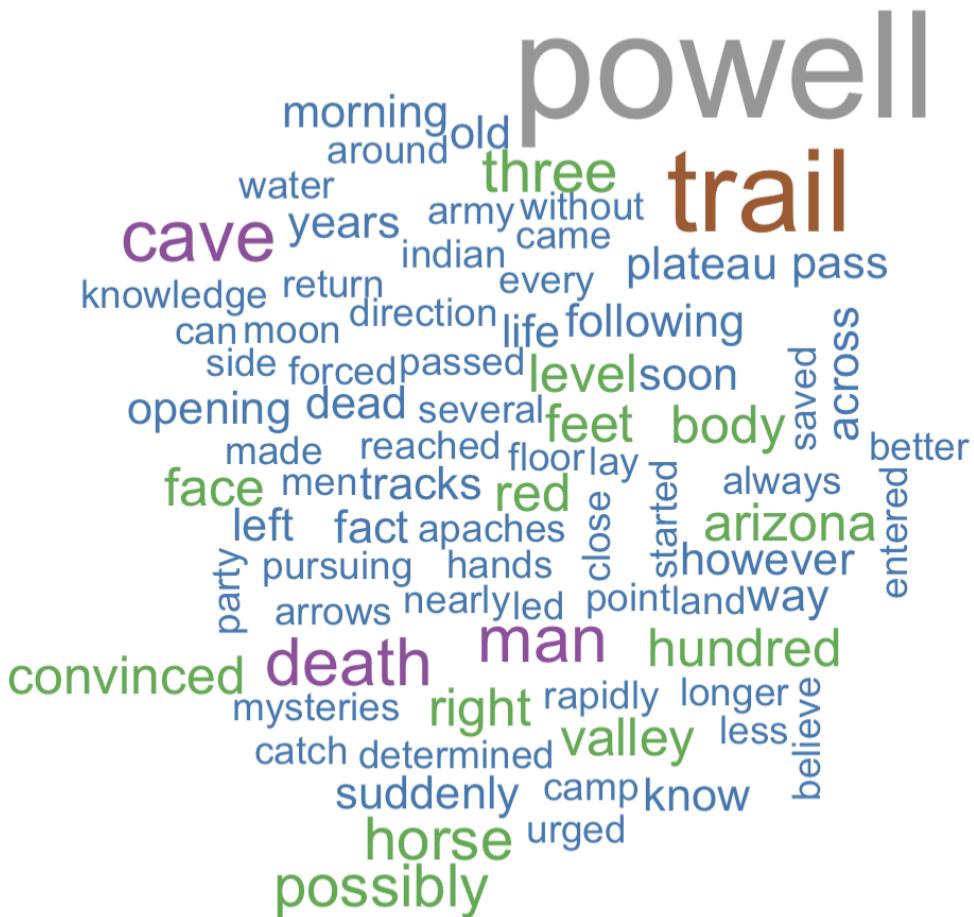


```

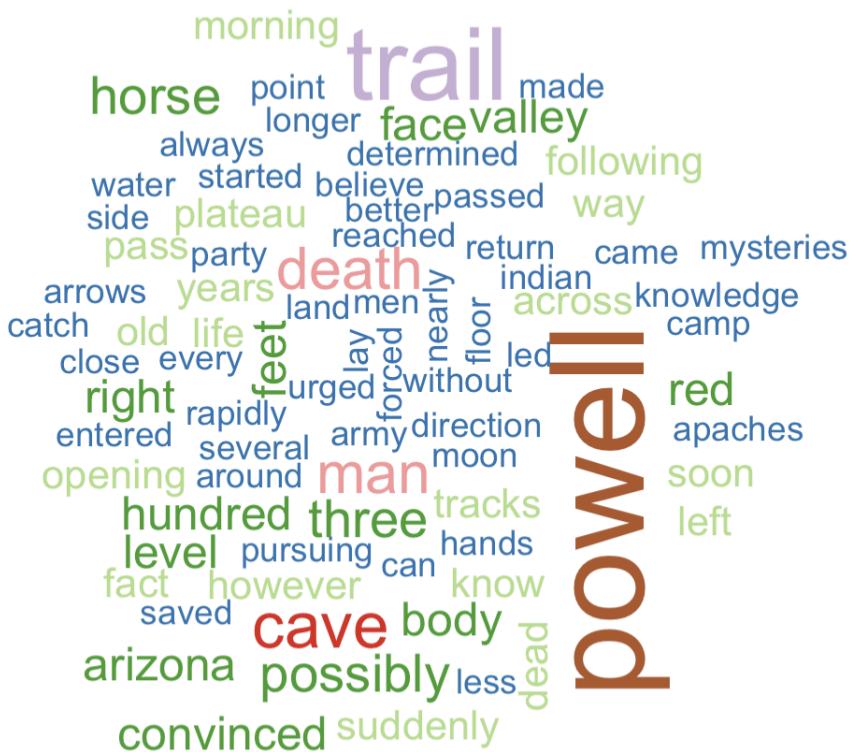
> # Try 2: Using Set1 - Bold and vibrant colors
> cp2 <- brewer.pal(20, "Set1")
Warning message:
In brewer.pal(20, "Set1") :
  n too large, allowed maximum for palette Set1 is 9
Returning the palette you asked for with that many colors

> str(cp2)
chr [1:9] "#E41A1C" "#377EB8" "#4DAF4A" "#984EA3" "#FF7F00" "#FFFF33" "#A65628" "#F781BF" ...
> chap1WC2 <- wordcloud(chap1Words, Chap1TermFreq, colors = cp2)

```



```
> cp3 <- brewer.pal(12, "Paired")
> str(cp3)
chr [1:12] "#A6CEE3" "#1F78B4" "#B2DF8A" "#33A02C" "#FB9A99" "#E31A1C" "#FDBF6F" ...
> chap1WC3 <- wordcloud(chap1Words, Chap1TermFreq, colors = cp3)
```



Exploring functions in quanteda package

First 10 lines of textCL (textCL is the Text obj post removing punctuation and Numbers)

```
> Chap1TextCL <- textCL[[1]]  
> Chap1TextCL$content[1:10]  
[1] ""  
[2] "ON THE ARIZONA HILLS"  
[3] ""  
[4] ""  
[5] "I am a very old man how old I do not know Possibly I am a hundred"  
[6] "possibly more but I cannot tell because I have never aged as other"  
[7] "men nor do I remember any childhood So far as I can recollect I have"  
[8] "always been a man a man of about thirty I appear today as I did"  
[9] "forty years and more ago and yet I feel that I cannot go on living"  
[10] "forever that some day I shall die the real death from which there is"
```

Before cleaning, the text corpus (`Chap1TextCL`) contains multiple blank lines. These gaps need to be removed to ensure better structure and flow for text preprocessing and analysis.

Removing Blank lines from textCL

```
> #---Removing Blank Lines----  
> rmBlanks <- function(x) {  
+   lines <- unlist(strsplit(as.character(x), "\n"))  
+   non_blank_lines <- lines[nchar(trimws(lines)) > 0]  
+   return(paste(non_blank_lines))  
+ }  
> textCLNoBlank <- tm_map(textCL, content_transformer(rmBlanks))  
>  
> textCLNoBlank_Chap1 <- textCLNoBlank[[1]]  
> textCLNoBlank_Chap1$content[1:10]  
[1] "ON THE ARIZONA HILLS"  
[2] "I am a very old man how old I do not know Possibly I am a hundred"  
[3] "possibly more but I cannot tell because I have never aged as other"  
[4] "men nor do I remember any childhood So far as I can recollect I have"  
[5] "always been a man a man of about thirty I appear today as I did"  
[6] "forty years and more ago and yet I feel that I cannot go on living"  
[7] "forever that some day I shall die the real death from which there is"  
[8] "no resurrection I do not know why I should fear death I who have"  
[9] "died twice and am still alive but yet I have the same horror of it as"  
[10] "you who have never died and it is because of this terror of death I"
```

A custom function `rmBlanks` was created to remove blank lines from the text corpus. After applying it with `tm_map`, the cleaned document shows continuous, non-empty lines, improving text consistency for further analysis.

Tokenization of the data

```
> #----Tokenization----#
> textClNoBlank_Chap1_Tokens <- quanteda::tokens(textClNoBlank_Chap1$content)
> str(textClNoBlank_Chap1_Tokens)
List of 223
$ text1 : chr [1:4] "ON" "THE" "ARIZONA" "HILLS"
$ text2 : chr [1:17] "I" "am" "a" "very" ...
$ text3 : chr [1:13] "possibly" "more" "but" "I" ...
$ text4 : chr [1:15] "men" "nor" "do" "I" ...
$ text5 : chr [1:15] "always" "been" "a" "man" ...
$ text6 : chr [1:15] "forty" "years" "and" "more" ...
$ text7 : chr [1:14] "forever" "that" "some" "day" ...
$ text8 : chr [1:14] "no" "resurrection" "I" "do" ...
$ text9 : chr [1:16] "died" "twice" "and" "am" ...
$ text10 : chr [1:15] "you" "who" "have" "never" ...
$ text11 : chr [1:9] "believe" "that" "I" "am" ...
$ text12 : chr [1:12] "And" "because" "of" "this" ...
$ text13 : chr [1:14] "story" "of" "the" "interesting" ...
$ text14 : chr [1:14] "explain" "the" "phenomena" "I" ...
$ text15 : chr [1:11] "ordinary" "soldier" "of" "fortune" ...
$ text16 : chr [1:14] "befell" "me" "during" "the" ...
$ text17 : chr [1:2] "Arizona" "cave"
$ text18 : chr [1:13] "I" "have" "never" "told" ...
$ text19 : chr [1:13] "until" "after" "I" "have" ...
$ text20 : chr [1:14] "human" "mind" "will" "not" ...
$ text21 : chr [1:12] "purpose" "being" "pilloried" "by" ...
$ text22 : chr [1:14] "held" "up" "as" "a" ...
$ text23 : chr [1:9] "which" "some" "day" "science" ...
$ text24 : chr [1:14] "which" "I" "gained" "upon" ...
$ text25 : chr [1:11] "this" "chronicle" "will" "aid" ...
$ text26 : chr [1:13] "of" "our" "sister" "planet" ...
$ text27 : chr [1:14] "My" "name" "is" "John" ...
$ text28 : chr [1:13] "Virginia" "At" "the" "close" ...
$ text29 : chr [1:8] "several" "hundred" "thousand" "dollars" ...
$ text30 : chr [1:13] "commission" "in" "the" "cavalry" ...
$ text31 : chr [1:13] "servant" "of" "a" "state" ...
$ text32 : chr [1:10] "Masterless" "penniless" "and" "with" ...
$ text33 : chr [1:13] "gone" "I" "determined" "to" ...
$ text34 : chr [1:9] "retrieve" "my" "fallen" "fortunes" ...
$ text35 : chr [1:11] "I" "spent" "nearly" "a" ...
```

The cleaned text was tokenized using `quanteda::tokens`, breaking it into individual words or small phrases. This step is crucial for preparing the data for frequency analysis, word cloud creation, and further text mining tasks.

Constructing a sparse document-feature matrix using quanteda::dfm()

```
> #----Constructing DFM----#
> textClNoBlank_Chap1_DFM <- quanteda::dfm(textClNoBlank_Chap1_Tokens)
> str(textClNoBlank_Chap1_DFM)
Formal class 'dfm' [package "quanteda"] with 8 slots
..@ docvars :'data.frame': 223 obs. of 3 variables:
...$ docname_ : chr [1:223] "text1" "text2" "text3" "text4" ...
...$ docid_ : Factor w/ 223 levels "text1","text2",..: 1 2 3 4 5 6 7 8 9 10 ...
...$ segid_ : int [1:223] 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 4 2
... . $ system : Named chr [1:3] "Darwin" "arm64" "rohithyv"
... . . - attr(*, "names")= chr [1:3] "sysname" "machine" "user"
... . $ directory : chr "/Users/rohithyv/Documents/Introduction to Big Data and Analytics_CSCI_6444/Project_3/Chapters"
... . $ 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
```

The tokenized text was successfully converted into a Document-Feature Matrix (DFM) using quanteda::dfm. This matrix organizes documents and their corresponding word features, enabling structured quantitative text analysis such as word frequency exploration and machine learning modeling.

Computing frequency of words within DFM

```
> #----Computing Frequency of words in DFM----#
> textClNoBlank_Chap1_DFM_Freq <- quanteda::docfreq(textClNoBlank_Chap1_DFM)
> str(textClNoBlank_Chap1_DFM_Freq)
Named int [1:887] 15 138 5 1 97 9 53 2 3 6 ...
- attr(*, "names")= chr [1:887] "on" "the" "arizona" "hills" ...
> textClNoBlank_Chap1_DFM_Freq
```

on	the	arizona	hills	i	am	a	very
15	138	5	1	97	9	53	2
old	man	how	do	not	know	possibly	hundred
3	6	1	7	17	4	6	5
more	but	cannot	tell	because	have	never	aged
5	12	5	1	4	14	5	1
as	other	men	nor	remember	any	childhood	so
19	4	3	2	1	4	1	9
far	can	recollect	always	been	of	about	thirty
4	3	1	3	6	91	7	2
appear	today	did	forty	years	and	ago	yet
1	1	3	2	4	74	1	3
feel	that	go	living	forever	some	day	shall
2	48	2	1	1	8	3	2
die	real	death	from	which	there	is	no
1	1	7	9	19	4	8	6
resurrection	why	should	fear	who	died	twice	still
1	1	2	1	4	2	1	1
alive	same	horror	it	you	this	terror	believe
2	3	1	15	2	15	1	3
convinced	my	mortality	conviction	determined	to	write	down
5	44	1	1	3	54	1	10
story	interesting	periods	life	explain	phenomena	only	set
2	1	1	4	1	1	4	2
here	in	words	an	ordinary	soldier	fortune	chronicle
1	31	1	14	1	1	1	2
strange	events	befell	me	during	ten	dead	body
1	1	2	26	3	2	4	5
lay	undiscovered	cave	told	mortal	see	manuscript	until
3	1	8	1	1	2	1	6
after	passed	over	for	eternity	average	human	mind
3	3	5	20	1	1	1	2
will	what	grasp	purpose	being	pilloried	by	public

Word frequencies were computed from the Document-Feature Matrix (DFM) using quanteda::docfreq. This output shows how often each word appears across documents, providing key insights for feature selection, word cloud creation, and further text analysis.

Assigning Weights to Terms in the Document-Feature Matrix (DFM)

```
> #---Assigning weights to the words----#
> textClNoBlank_Chap1_Wt <- quanteda::dfm_weight(textClNoBlank_Chap1_DFM)
> str(textClNoBlank_Chap1_Wt)
Formal class 'dfm' [package "quanteda"] with 8 slots
 ..@ docvars :'data.frame':   223 obs. of  3 variables:
 ...$ docname_ : chr [1:223] "text1" "text2" "text3" "text4" ...
 ...$ docid_   : Factor w/ 223 levels "text1","text2",...: 1 2 3 4 5 6 7 8 9 10 ...
 ...$ segid_   : int [1:223] 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/rohithyv/Documents/Introduction to Big Data and Analytics_CSCI_6444/Project_3/Chapters"
 ...$ 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
```

Weights were applied to the Document-Feature Matrix using `quanteda::dfm_weight`, allowing the model to adjust the influence of each term based on document or frequency-based schemes, preparing the data for more refined text analysis.

Weighting the Document-Feature Matrix (DFM)

```

> textClNoBlank_Chap1_Wt
Document-feature matrix of: 223 documents, 887 features (98.76% sparse) and 0 docvars.
    features
docs      on the arizona hills i am a very old man
text1    1   1       1   1 0   0 0     0   0   0
text2    0   0       0   0 3   2 2     1   2   1
text3    0   0       0   0 2   0 0     0   0   0
text4    0   0       0   0 3   0 0     0   0   0
text5    0   0       0   0 2   0 2     0   0   2
text6    1   0       0   0 2   0 0     0   0   0
[ reached max_ndoc ... 217 more documents, reached max_nfeat ... 877 more features ]

```

The DFM was enhanced by applying weights with `quanteda::dfm_weight`, enabling better control over the importance of words during further text analysis tasks like classification, clustering, or topic modeling.

Computing TF-IDF score on the DFM

```
> #---Computing TF-IDF scores----#
> textClNoBlank_Chap1_TFIDF <- quanteda::dfm_tfidf(textClNoBlank_Chap1_DFM, scheme_tf = "count", scheme_df = "inverse")
> str(textClNoBlank_Chap1_TFIDF)
Formal class 'dfm' [package "quanteda"] with 8 slots
..@ docvars :data.frame': 223 obs. of 3 variables:
.. ..$ docname_ : chr [1:223] "text1" "text2" "text3" "text4" ...
.. ..$ docid_ : Factor w/ 223 levels "text1","text2",...: 1 2 3 4 5 6 7 8 9 10 ...
.. ..$ segid_ : int [1:223] 1 1 1 1 1 1 1 1 1 ...
..@ meta   :List of 5
.. ..$ 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 4 2
.. .. .$. system    : Named chr [1:3] "Darwin" "arm64" "rohithyv"
.. .. .$. - attr(*, "names")= chr [1:3] "sysname" "machine" "user"
.. .. .$. directory : chr "/Users/rohithyv/Documents/Introduction to Big Data and Analytics_CSCI_6444/Project_3/Chapters"
.. .. .$. 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
```

TF-IDF scores were computed using `quanteda::dfm_tfidf`, adjusting the raw word frequencies to highlight terms that are important within specific documents but less frequent across the entire corpus. This enhances the ability to identify distinctive words for deeper text analysis.

```
> textClNoBlank_Chap1_TFIDF
Document-feature matrix of: 223 documents, 887 features (98.76% sparse) and 0 docvars.
  features
docs   on      the arizona hills      i      am      a      very      old      man
text1 1.172214 0.2084258 1.649335 2.348305 0      0      0      0      0      0
text2 0      0      0      1.0845994 2.788125 1.248058 2.047275 3.742367 1.570154
text3 0      0      0      0.7230663 0      0      0      0      0      0
text4 0      0      0      1.0845994 0      0      0      0      0      0
text5 0      0      0      0.7230663 0      1.248058 0      0      0      3.140307
text6 1.172214 0      0      0.7230663 0      0      0      0      0      0
[ reached max_ndoc ... 217 more documents, reached max_nfeat ... 877 more features ]
```

TF-IDF scores were calculated to emphasize important terms that are frequent within a document but rare across all documents. This transformation improves feature selection for text mining and classification tasks.

TF-IDF is a statistical score which represents the importance of word to a document in respective corpus and

Term Frequency (TF) provides frequency of a word in the respective document

Inverse Document Frequency (IDF) quantifies the significance of a term by reducing the influence of terms that appear frequently across documents and amplifying the importance of terms that are less common

Exploring functions of syuzhet package

Preparing Text Data Frame for Syuzhet Sentiment Analysis

```
> #----Exploring functions of syuzhet -----#
>
> TextObjChap1_DF <- as.data.frame(TextObj[[1]]$content)
> TextObjChap1_DF
TextObj[[1]]$content
1
2
3
4
5      I am a very old man; how old I do not know. Possibly I am a hundred,
6      possibly more; but I cannot tell because I have never aged as other
7      men, nor do I remember any childhood. So far as I can recollect I have
8      always been a man, a man of about thirty. I appear today as I did
9      forty years and more ago, and yet I feel that I cannot go on living
10     forever; that some day I shall die the real death from which there is
11     no resurrection. I do not know why I should fear death, I who have
12     died twice and am still alive; but yet I have the same horror of it as
13     you who have never died, and it is because of this terror of death, I
14             believe, that I am so convinced of my mortality.
15
16      And because of this conviction I have determined to write down the
17      story of the interesting periods of my life and of my death. I cannot
18      explain the phenomena; I can only set down here in the words of an
19      ordinary soldier of fortune a chronicle of the strange events that
20      befell me during the ten years that my dead body lay undiscovered in an
21                      Arizona cave.
22
23      I have never told this story, nor shall mortal man see this manuscript
24      until after I have passed over for eternity. I know that the average
25      human mind will not believe what it cannot grasp, and so I do not
26      purpose being pilloried by the public, the pulpit, and the press, and
27      held up as a colossal liar when I am but telling the simple truths
28      which some day science will substantiate. Possibly the suggestions
29      which I gained upon Mars, and the knowledge which I can set down in
30      this chronicle, will aid in an earlier understanding of the mysteries
31      of our sister planet; mysteries to you, but no longer mysteries to me.
32
```

The full chapter text was organized into a data frame structure to facilitate line-by-line sentiment extraction using the syuzhet package. This setup is essential for detailed emotional trajectory analysis.

Reading Chapter1.txt as string

Loading Full Chapter Text as a Single String

```
> Chap1_string <- get_text_as_string("Chapter1.txt")
> Chap1_string
```

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 state which had vanished with the hopes of the South. Masterless, penniless, and with my only means of livelihood, fighting, gone, I determined to work my way to the southwest and attempt to retrieve my fallen fortunes in a search for gold. I spent nearly a year prospecting in company with another Confederate officer, Captain James K. Powell of Richmond. We were extremely fortunate, for late in the winter of 1865, after many hardships and privations, we located the most remarkable gold-bearing quartz vein that our wildest dreams had ever pictured. Powell, who was a mining engineer by education, stated that we had uncovered over a million dollars worth of ore in a trifle over three months. As our equipment was crude in the extreme we decided that one of us must return to civilization, purchase the necessary machinery and return with a sufficient force of men properly to work the mine. As Powell was familiar with the country, as well as with the mechanical requirements of mining we determined that it would be best

The complete text from "Chapter1.txt" was read into R as a single string using `get_text_as_string`. This format enables easy manipulation, sentiment analysis, or text segmentation in one continuous operation.

Extracting sentences from the above string

```
> Chap1_sentences <- get_sentences(Chap1_string)
> str(Chap1_sentences)
chr [1:77] "ON THE ARIZONA HILLS I am a very old man; how old I do not know." ...
> Chap1_sentences
[1] "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 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."
[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."
[11] "My name is John Carter; I am better known as Captain Jack Carter of Virginia."
[12] "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 state which had vanished with the hopes of the South."
[13] "Masterless, penniless, and with my only means of livelihood, fighting, gone, I determined to work my way to the southwest and attempt to retrieve my fallen fortunes in a search for gold."
[14] "I spent nearly a year prospecting in company with another Confederate officer, Captain James K. Powell of Richmond."
[15] "We were extremely fortunate, for late in the winter of 1865, after many hardships and privations, we located the most remarkable gold-bearing quartz vein that our wildest dreams had ever pictured."
```

The full chapter text was split into 77 individual sentences using `get_sentences`, setting up the data for fine-grained sentiment scoring and emotional arc visualization.

Comparison of sentiment analysis between syuzhet and bing

Calculating Sentence-Level Sentiment Scores Using Syuzhet and Bing Methods

```
> #---Sentiment---#
> syuzhetSentiChap1 <- get_sentiment(Chap1_sentences, "syuzhet")
> syuzhetSentiChap1
[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
[17]  2.00  2.60  0.50  0.80  1.40  0.00  0.80  0.50 -4.95  1.65  0.65  1.00  0.00 -0.75 -0.25  2.85
[33]  1.50 -0.85  0.70 -1.50 -0.50 -0.75  0.80 -1.50  1.05  1.20  1.80  0.05 -1.00 -1.25  0.25  0.70
[49] -2.00  0.25 -1.35 -0.65 -1.25 -1.25 -0.25  1.30  1.00 -0.35  0.20 -0.60  0.30  1.00  1.20  0.80
[65]  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
> bingSentiChap1 <- get_sentiment(Chap1_sentences, "bing")
> bingSentiChap1
[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
[34] -1  0 -2 -1  0 -1 -2 -1  0  1 -1 -1 -1  2  1 -2  1 -2  0 -3  0 -3  0  1 -2  0 -1  1  1  0  1  0
[67] -2 -1  0  2 -2  0 -3 -1  0 -2 -1
```

Sentence-wise sentiment scores were computed using `get_sentiment` with both "syuzhet" (continuous values) and "bing" (discrete scores) methods, enabling detailed emotional analysis across the chapter.

Sentiment Dictionary of scores syuzhet Vs bing

Syuzhet Dictionary

Retrieving Syuzhet Sentiment Dictionary

```
> syuzhetSentiChap1Dict <- get_sentiment_dictionary()
> syuzhetSentiChap1Dict
      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
21     abolish -0.50
22    abominable -0.50
23   abominably -1.00
24     abominate -1.00
25   abomination -0.50
26      abort -0.50
27    aborted -0.80
```

This table displays the beginning of the sentiment dictionary extracted using the `get_sentiment_dictionary()` function from the `syuzhet` R package. Each word is associated with a numeric sentiment score: positive values indicate positive sentiment, and negative values indicate negative sentiment, supporting sentiment analysis tasks.

Bing Dictionary

Sentiment Dictionary Output Using "Bing" Lexicon

```
> bingSentiChap1Dict <- get_sentiment_dictionary("bing")
> bingSentiChap1Dict
```

	word	value
1	a+	1
2	abound	1
3	abounds	1
4	abundance	1
5	abundant	1
6	accessible	1
7	accessible	1
8	acclaim	1
9	acclaimed	1
10	acclamation	1
11	accolade	1
12	accolades	1
13	accommodative	1
14	accomodative	1
15	accomplish	1
16	accomplished	1
17	accomplishment	1
18	accomplishments	1
19	accurate	1
20	accurately	1
21	achievable	1
22	achievement	1
23	achievements	1
24	achievable	1
25	acumen	1
26	adaptable	1
27	adaptive	1
28	adequate	1
29	adjustable	1

This table shows the sentiment dictionary retrieved using the `get_sentiment_dictionary("bing")` function in R. Each word is associated with a sentiment value of 1, indicating positive sentiment according to the Bing Liu Lexicon, commonly used for binary (positive/negative) sentiment classification.

Summation of sentiment vector syuzhet Vs bing

Sentiment Score Summation Using Syuzhet and Bing Lexicons

```
> syuzhetSentiChap1Sum <- sum(syuzhetSentiChap1)
> syuzhetSentiChap1Sum
[1] 8.4
> bingSentiChap1Sum <- sum(bingSentiChap1)
> bingSentiChap1Sum
[1] -22
```

This output shows the total sentiment scores computed from two different dictionaries. The **syuzhet** dictionary results in a positive sentiment sum of **8.4**, while the **bing** dictionary produces a negative sentiment sum of **-22**, highlighting differences in sentiment evaluation across lexicons.

Mean of Sentiments syuzhet Vs bing

Mean Sentiment Score Calculation Using Syuzhet and Bing Lexicons

```
> #####Mean of the sentiments#####
> syuzhetSentiChap1Mean <- mean(syuzhetSentiChap1)
> syuzhetSentiChap1Mean
[1] 0.1090909
> bingSentiChap1Mean <- mean(bingSentiChap1)
> bingSentiChap1Mean
[1] -0.2857143
```

This output presents the mean sentiment scores derived from two different sentiment dictionaries. The **syuzhet** dictionary results in a slightly positive mean sentiment (**0.1091**), while the **bing** dictionary indicates an overall negative mean sentiment (**-0.2857**), illustrating how different lexicons can lead to contrasting interpretations of emotional tone.

Summary of sentiment syuzhet Vs bing

Summary Statistics of Sentiment Scores Using Syuzhet and Bing Lexicons

```
> #####Summary of sentiments#####
> summary(syuzhetSentiChap1)
   Min. 1st Qu. Median Mean 3rd Qu. Max.
-4.9500 -0.7500  0.0000  0.1091  0.8000  3.7000
> summary(bingSentiChap1)
   Min. 1st Qu. Median Mean 3rd Qu. Max.
-8.0000 -1.0000  0.0000 -0.2857  1.0000  3.0000
```

This output provides a statistical summary (minimum, 1st quartile, median, mean, 3rd quartile, and maximum) of the sentiment scores from the **syuzhet** and **bing** dictionaries. It highlights differences in the distribution of sentiment values, with **syuzhet** showing a slight positive mean and **bing** reflecting a negative mean, indicating contrasting sentiment interpretations across the two methods.

NRC (National Research Council) Sentiment Analysis

NRC Sentiment Analysis Score Matrix for Chapter 1 Sentences

```
> #-----NRC Sentiment Analysis----#
> Chap1NRC_SentiScores <- get_nrc_sentiment(Chap1_sentences)
> Chap1NRC_SentiScores
```

	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
11	0	0	1	0	0	0	0	0	1	1
12	1	0	1	3	1	1	1	4	4	3
13	1	1	0	0	0	0	0	0	1	2
14	0	0	0	0	0	0	0	2	1	3
15	0	0	0	0	1	1	1	1	1	3
16	0	0	0	0	0	0	0	0	1	1
17	1	0	1	1	0	0	0	1	2	1
18	0	0	0	0	0	0	1	1	0	2
19	0	1	0	0	0	0	0	1	0	1
20	0	3	0	1	2	0	1	2	0	3
21	0	0	1	0	1	1	2	1	2	2
22	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	1	0	1	1	0	1
24	0	2	0	1	0	1	1	2	2	1
25	4	3	4	4	1	3	0	1	7	2
26	1	0	0	1	0	0	0	1	3	3
27	0	2	1	1	1	0	2	2	0	2

This table displays the results of NRC sentiment analysis performed on individual sentences. Each row represents a sentence, and columns represent different emotions (anger, joy, trust, etc.) and polarities (positive, negative), with counts indicating the intensity or presence of each sentiment in the sentence. The analysis captures a broad emotional spectrum beyond just positive and negative sentiments.

NRC sentiment scores for each sentence

Total Sentiment Counts from NRC Analysis for Chapter 1

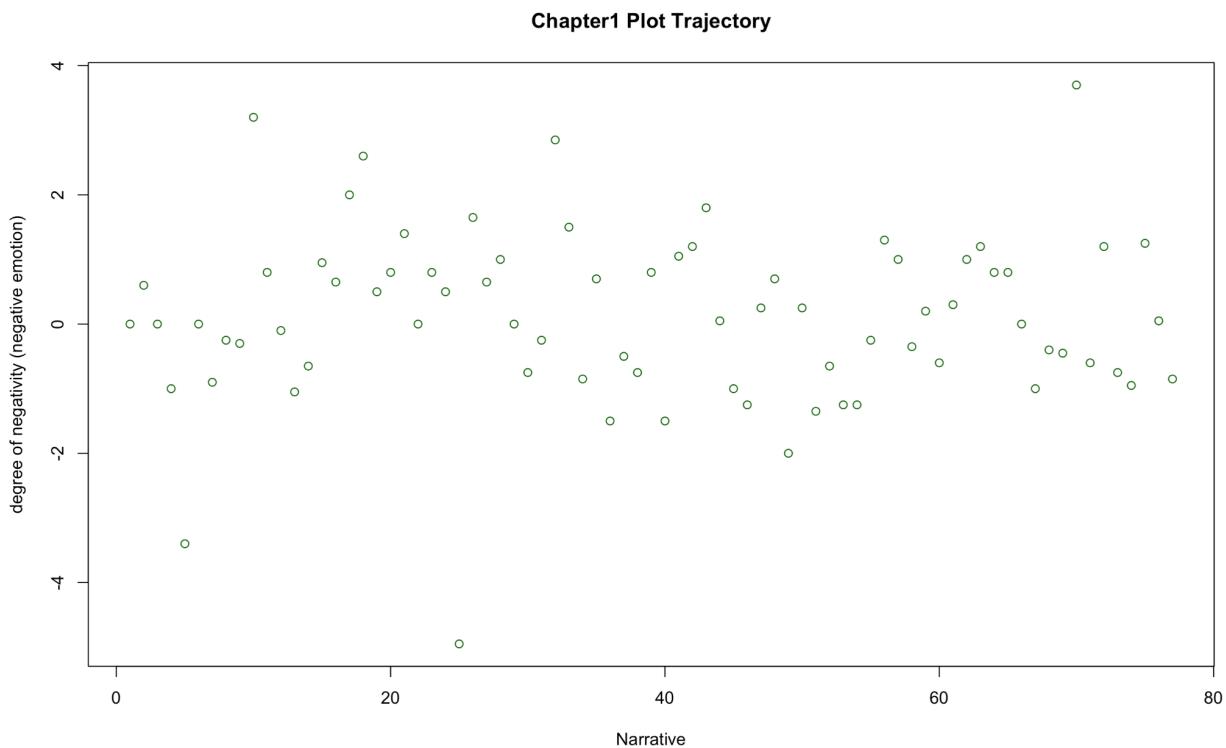
```
> Chap1_Counts <- colSums(Chap1NRC_SentiScores)
> Chap1_Counts
  anger anticipation      disgust      fear      joy      sadness      surprise
    38           57          31         69        37         37        47
  trust      negative      positive
    79          100         113
```

This output shows the overall counts of each sentiment and emotion category based on the NRC lexicon analysis of Chapter 1. Positive sentiment (113) appears most frequently, followed by negative sentiment (100) and trust (79), providing a summarized emotional profile of the text.

Plotting Narrative vs degree of negativity (negative emotion) for Chapter1.txt using syuzhet sentiment dictionary

Sentiment Trajectory Plot for Chapter 1

```
> #----Plotting Narrative vs degree of negativity (negative emotion)---#
> plot(syuzhetSentiChap1, main="Chapter1 Plot Trajectory",
+       xlab = "Narrative", ylab = "degree of negativity (negative emotion)", col = "darkgreen")
```

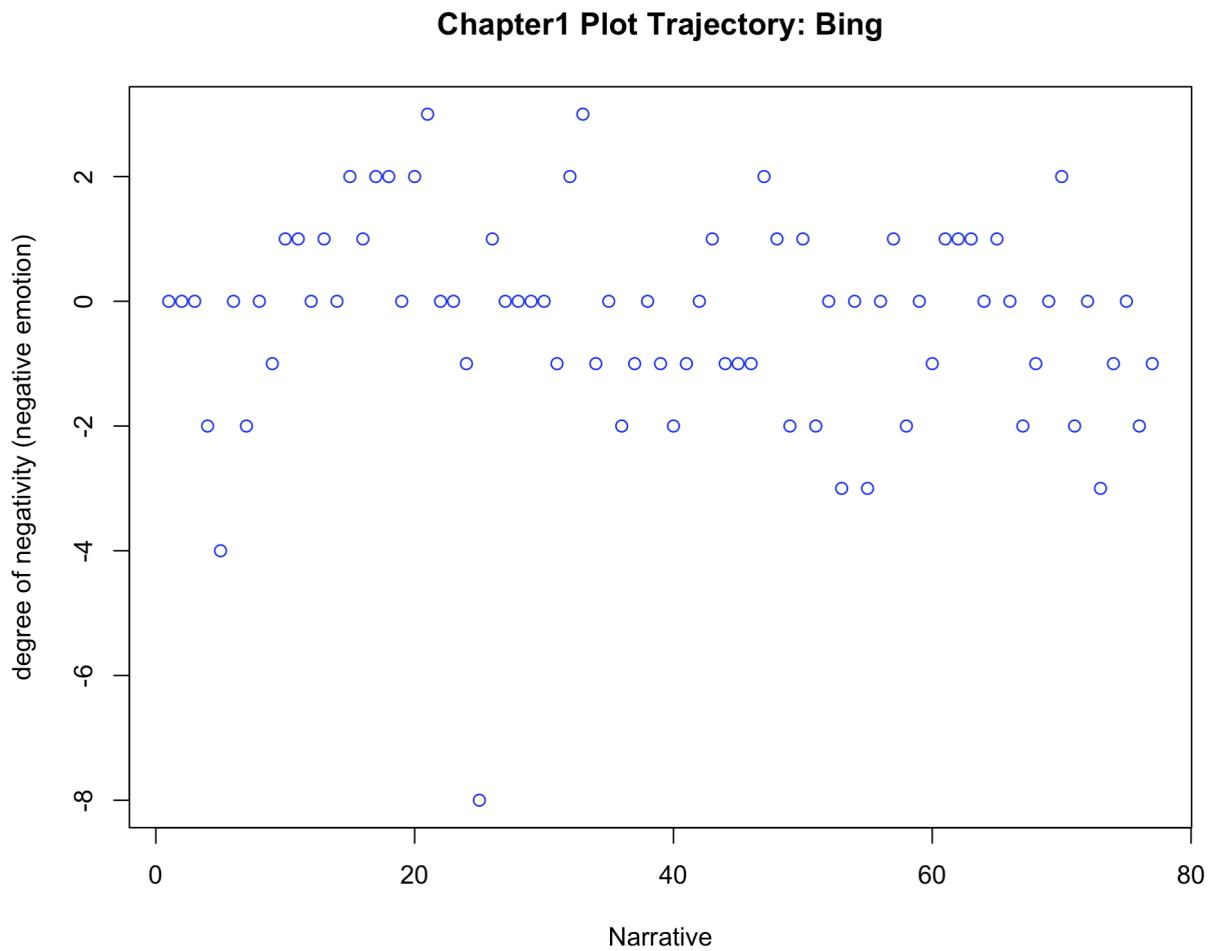


This scatter plot visualizes the emotional trajectory of Chapter 1, with each point representing the degree of negativity or positive sentiment across the narrative flow. The x-axis shows the sequence of the narrative, while the y-axis reflects the sentiment intensity, where negative values indicate stronger negative emotions and positive values indicate more positive emotions.

Plotting Narrative vs degree of negativity (negative emotion) for Chapter1.txt using bing sentiment dictionary

Sentiment Trajectory Plot for Chapter 1 Using Bing Lexicon

```
> plot(bingSentiChap1, main="Chapter1 Plot Trajectory: Bing",
+       xlab = "Narrative", ylab = "degree of negativity (negative emotion)", col = "blue")
```



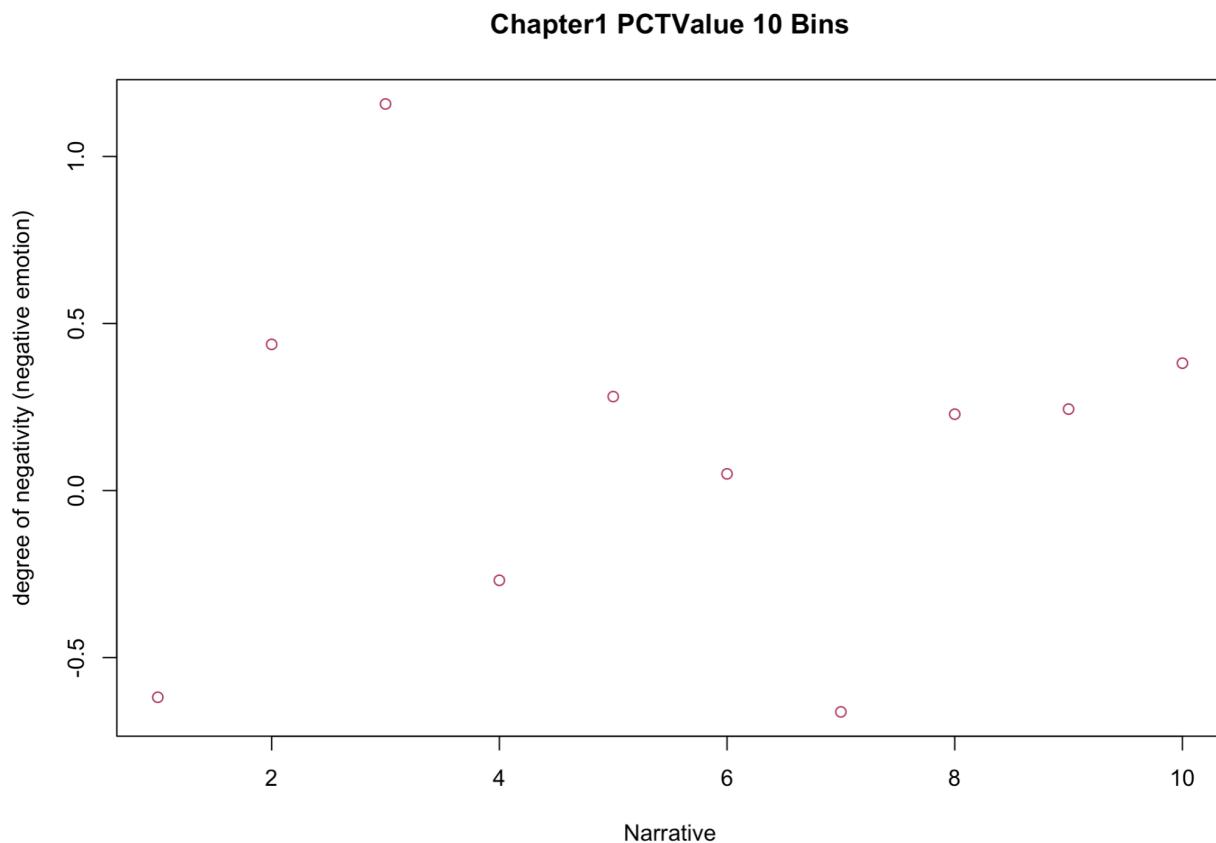
This plot illustrates the sentiment flow across Chapter 1 using the Bing sentiment lexicon. Each point represents a segment of the narrative, with the y-axis indicating the degree of negativity (negative values) or positivity (positive values). The plot highlights fluctuations in emotional tone, emphasizing sections with strong negative or slightly positive sentiment.

get_percentage_values function on syuzhet sentiment data with bins=10

```
> syuzhetSentiChap1_PCT10 <- get_percentage_values(syuzhetSentiChap1, bins = 10)
> syuzhetSentiChap1_PCT10
  1      2      3      4      5      6      7      8
-0.6187500  0.4375000  1.1571429 -0.2687500  0.2812500  0.0500000 -0.6625000  0.2285714
  9      10
  0.2437500  0.3812500
> structure(syuzhetSentiChap1_PCT10)
  1      2      3      4      5      6      7      8
-0.6187500  0.4375000  1.1571429 -0.2687500  0.2812500  0.0500000 -0.6625000  0.2285714
  9      10
  0.2437500  0.3812500
> plot(syuzhetSentiChap1_PCT10, main="Chapter1 PCTValue 10 Bins",
+       xlab = "Narrative", ylab = "degree of negativity (negative emotion)", col = "maroon")
```

Divides the chapter into 10 parts, calculates average sentiment per bin, and plots overall emotional progression using maroon color.

Sentiment Progression Plot for Chapter 1 (10 Bins Division)



This plot summarizes the emotional trend of Chapter 1 by dividing the narrative into 10 equal parts, displaying the average degree of negativity or positivity in each bin.

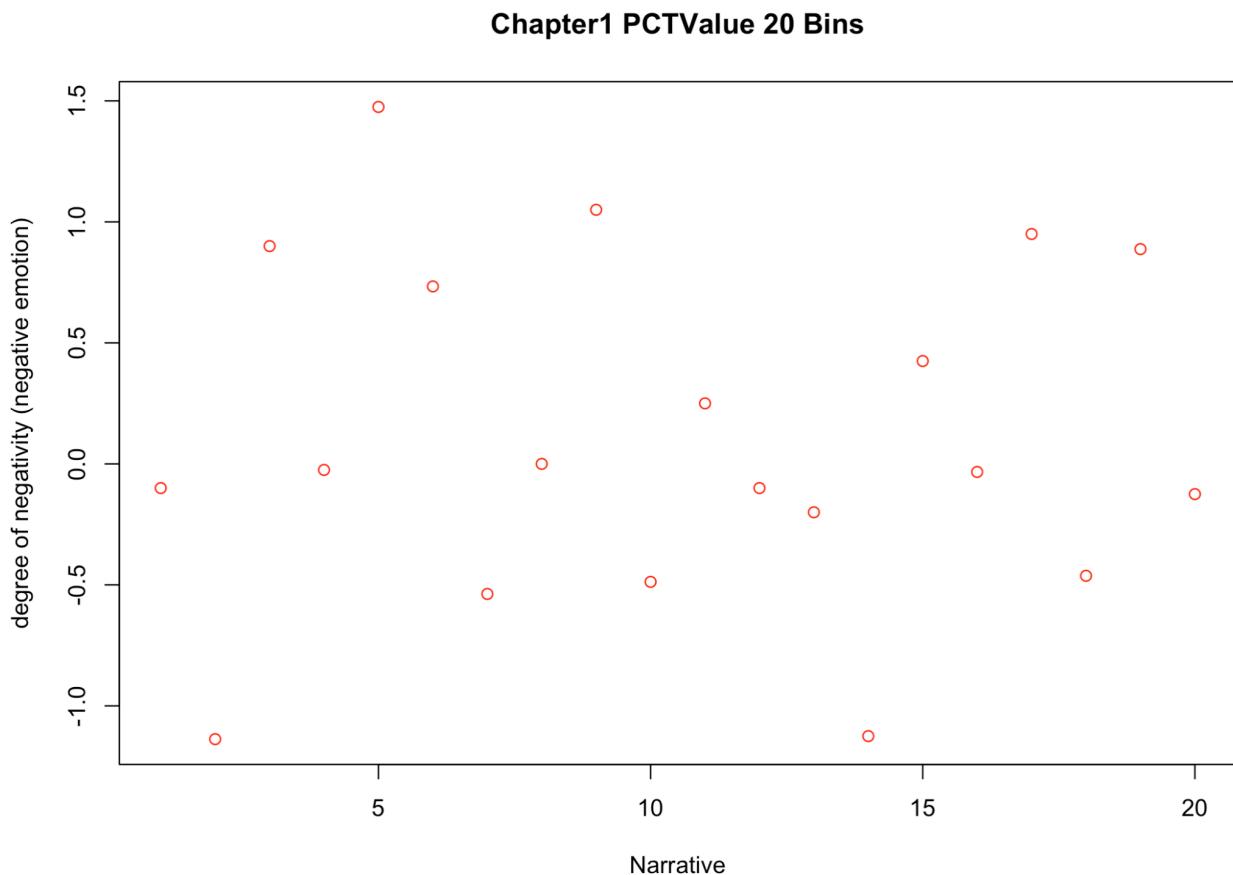
get_percentage_values function on syuzhet sentiment data with bins=20

Sentiment Progression Plot for Chapter 1 (20 Bins Division)

```
> syuzhetSentiChap1_PCT20 <- get_percentage_values(syuzhetSentiChap1, bins = 20)
> structure(syuzhetSentiChap1_PCT20)
  1      2      3      4      5      6      7
-0.10000000 -1.13750000  0.90000000 -0.02500000  1.47500000  0.73333333 -0.53750000
  8      9     10     11     12     13     14
  0.00000000  1.05000000 -0.48750000  0.25000000 -0.10000000 -0.20000000 -1.12500000
  15     16     17     18     19     20
  0.42500000 -0.03333333  0.95000000 -0.46250000  0.88750000 -0.12500000
> plot(syuzhetSentiChap1_PCT20, main="Chapter1 PCTValue 20 Bins",
+       xlab = "Narrative", ylab = "degree of negativity (negative emotion)", col = "red")
```

The narrative is divided into 20 parts, calculating the average sentiment for each segment to capture finer emotional shifts. The plot uses red color to highlight the degree of negativity or positivity across the chapter more precisely.

Chapter 1 Sentiment Trajectory Using 20 Narrative Bins



This plot shows the emotional flow of Chapter 1 divided into 20 parts, with each point representing the average degree of negativity or positivity in that segment, providing a finer-grained view of sentiment shifts throughout the narrative.

What we learned from the project?

Through this project, we gained **hands-on experience in performing end-to-end text analytics** on a large dataset — *The Princess of Mars* (Chapters I–XIV) — using R and multiple text mining packages.

1. Data Preparation and Cleaning:

We learned how to split a large text into chapters, create a VCorpus, and cleanse the text by removing punctuation, numbers, stop words, and custom-defined irrelevant terms. We standardized text by converting it to lowercase, resulting in a cleaner dataset ready for analysis.

2. Text Representation (DTM/TDM):

We constructed Document-Term Matrices (DTM) and Term-Document Matrices (TDM) to quantitatively represent the corpus. This taught us how to handle sparsity, analyze term frequency, and prepare the text for advanced modeling.

3. Exploring Text Features:

Extracting the **10 longest words** and **10 longest sentences** in each chapter helped us understand linguistic richness, sentence complexity, and vocabulary patterns across the narrative.

4. Word Frequency and Visualization:

We identified high-frequency words (e.g., "dejah", "martian") and visualized them using **word clouds**, providing intuitive insights into the most dominant themes of the novel.

5. Clustering and Dendograms:

Hierarchical clustering of terms using **Ward's method** revealed term relationships and hidden structures within the text. However, we also learned that sparse data can make dendograms less informative unless additional noise removal is performed.

6. Sentiment Analysis:

Using the **syuzhet**, **bing**, and **NRC** sentiment dictionaries, we computed sentiment scores:

- Syuzhet showed a slightly positive mean sentiment.
- Bing showed a negative overall sentiment.
- NRC revealed the richness of emotions like trust, anger, joy, and sadness. We understood that different lexicons can yield different interpretations of the same text.

7. Emotional Trajectory Visualization:

We visualized sentiment progressions by plotting sentiment scores across the narrative timeline (both raw and binned into 10 and 20 segments). These plots demonstrated how emotional tones rise and fall within the story, offering a deeper narrative analysis.

8. Advanced Text Mining (TF-IDF and quanteda):

We practiced more advanced techniques like **TF-IDF weighting** and **tokenization** using the **quanteda** package, preparing our text data for future machine learning or classification tasks.

Conclusion:

This project significantly improved our skills in **text preprocessing**, **feature extraction**, **sentiment analysis**, **visualization**, and **interpretation of narrative structures**. It also taught us the challenges of handling large textual datasets, such as sparsity, noise removal, and choosing appropriate lexicons for analysis. Overall, it was a complete workflow experience from raw text to actionable insights using R.