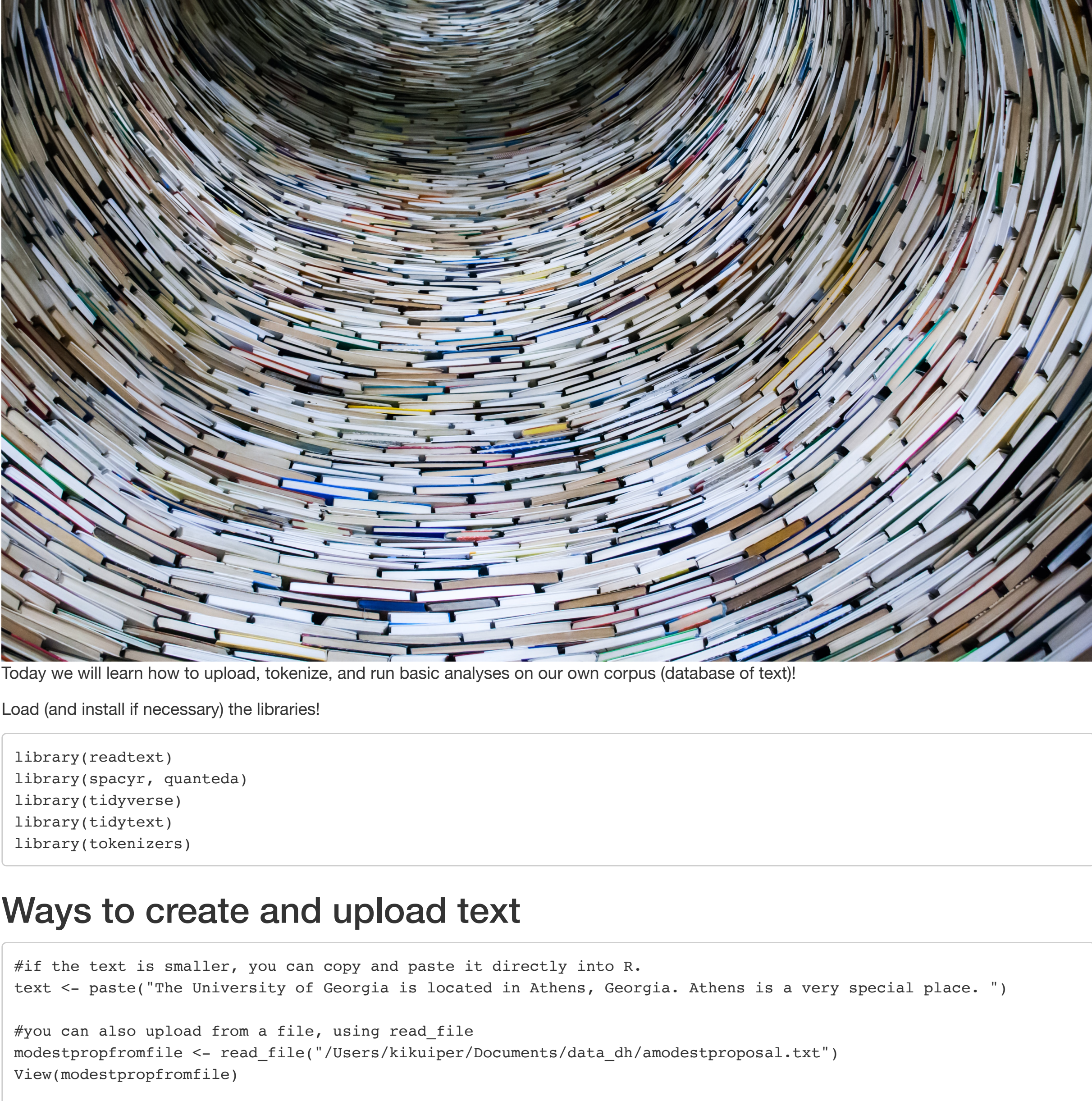


Introduction to Text Analysis in R

Before we get started, be sure that you have R, and R Studio downloaded and installed.



Today we will learn how to load, tokenize, and run basic analyses on our own corpus (database of text!)
Load (and install if necessary) the libraries!

```
library(readtext)
library(spacyr, quiet=TRUE)
library(tidyverse)
library(tidytext)
library(tokenizers)
```

Ways to create and upload text

```
#If the text is smaller, you can copy and paste it directly into R.
text <- paste("The University of Georgia is located in Athens, Georgia. Athens is a very special place. ")
```

```
#You can also upload from a file, using read_file
modestpropfromfile <- read_file("~/Users/kkuiper/Documents/data_dh/amodestproposal.txt")
View(modestpropfromfile)
```

```
#the text we will analyse today!
modestprop <- paste("A Modest Proposal
```

For preventing the children of poor people in Ireland, from being a burden on their parents or country, and for making them beneficial to the publick.

by Dr. Jonathan Swift

1729

It is a melancholy object to those, who walk through this great town, or travel in the country, when they see the streets, the roads, and cabin-doors crowded with beggars of the female sex, followed by three, four, or six children, all in rags, and importing every passenger for an alms. These mothers, instead of being able to work for their honest livelihood, are forced to employ all their time in strolling to beg sustenance for their helpless infants who, as they grow up, either turn thieves for want of work, or leave their dear native country, to fight for the Pretender in Spain, or sell themselves to the Barbadoes.

I think it is agreed by all parties, that this prodigious number of children in the arms, or on the backs, or at the heels of their mothers, and frequently of their fathers, is in the present deplorable state of the kingdom, a very great additional grievance; and therefore whoever could find out a fair, cheap and easy method of making these children sound and useful members of the commonwealth, would deserve no well of the publick, as to have his statue set up for a preserver of the nation.

But my intention is very far from being confined to provide only for the children of professed beggars: it is of a much greater extent, and shall take in the whole number of infants at a certain age, who are born of parents in effect as little able to support them, as those who demand our charity in the streets.

As to my own part, having turned my thoughts for many years upon this important subject, and maturely weighed the several schemes of our projectors, I have always found them grossly mistaken in their computation. It is true, for a child just dropt from its dam, may be supported by its milk, for a solar year, with little other nourishment: at most not above the value of two shillings, which the mother may certainly get, or the value in scraps, by her lawful occupation of begging; and it is exactly at one year old that I propose to provide for them in such a manner, as, instead of being a charge upon their parents, or the parish, or wanting food and raiment for the rest of their lives, they shall, on the contrary, contribute to the feeding, and partly to the clothing of many thousands.

There is likewise another great advantage in my scheme, that it will prevent those voluntary abortions, and that horrid practice of women murdering their bastard children, alas! too frequent among us, sacrificing the poor innocent babes, I doubt, more to avoid the expense than the shame, which would move tears and pity in the most savage and inhuman breast.

The number of souls in this kingdom being usually reckoned one million and a half, of these I calculate there may be about two hundred thousand couple, whose wives are breeders; from which number I subtract thirty thousand couple, who are able to maintain their own children, and (although I apprehend there cannot be so many under the present distresses of the kingdom) but this being granted, there will remain a hundred and seventy thousand breeders. I again subtract fifty thousand, for those women who miscarry, or whose children die by accident or disease within the year. There only remain a hundred and twenty thousand children of poor parents annually born. The question therefore is, how this number shall be reared and provided for? which, as I have already said, under the present situation of affairs, is utterly impossible by all the methods hitherto proposed. For we can neither employ them in handicraft or agriculture; they neither build houses, (I mean in the country) nor cultivate land: they can very seldom pick up a livelihood by stealing till they arrive at six years old; except where they are or towards parts, although I confess they learn the rudiments much earlier; during which time they can however be properly looked upon only as probationers; as I have been informed by a principal gentleman in the county of Cavan, who protested to me, that he never knew above one or two instances under the age of six, even in a part of the kingdom so renowned for the quickest proficiency in that art.

I am assured by our merchants, that a boy or a girl, before twelve years old, is no saleable commodity, and even when they come to this age, they will not yield above three pounds, or three pounds and half a crown at most, on the exchange; which cannot turn to account either to the parents or kingdom, the charge of nutriment and rags having been at least four times that value.

I shall now therefore humbly propose my own thoughts, which I hope will not be liable to the least objection.

I have been assured by a very knowing American of my acquaintance in London, that a young healthy child well nursed, is, at a year old, a most delicious nourishing and wholesome food, whether stewed, roasted, baked, or boiled; and I make no doubt that it will equally serve in a fricassee, or a ragoust.

I do therefore humbly offer it to publick consideration, that of the hundred and twenty thousand children, already computed, twenty thousand may be reserved for breed, whereof only one fourth part to be males; which is more than we allow to sheep, black cattle, or swine, and my reason is, that these children are seldom the fruits of marriage, a circumstance not much regarded by our savages, therefore, one male will be sufficient to serve four females. That the remaining hundred thousand may, at a year old, be offered in sale to the persons of quality and fortune, through the kingdom, always advising the mother to let them suck plentifully in the last month, so as to render them plump, and fat for a good table. A child will make two dishes at an entertainment for friends, and when the family dines alone, the fore or hind quarter will make a reasonable dish, and seasoned with a little pepper or salt, will be very good boiled on the fourth day, especially in winter.

I have reckoned upon a medium, that a child just born will weigh 12 pounds, and in a solar year, if tolerably nursed, encrease to 28 pounds.

I grant this food will be somewhat dear, and therefore very proper for landlords, who, as they have already devoured most of the parents, seem to have the best title to the children.

Infant's flesh will be in season throughout the year, but more plentiful in March, and a little before and after; for we are told by a grave author, an eminent french physician, that fish being a prolific dyet, there are more children born in Roman Catholick countries about nine months after Lent, than at any other season; therefore, reckoning a year after Lent, the markets will be more glutted than usual, because the number of Popish infants, is at least three to one in this kingdom, and therefore it will have one other collateral advantage, by lessening the number of Papists among us.

I have already computed the charge of nursing a beggar's child (in which list I reckon all cottagers, labourers, and four-fifths of the farmers) to be about two shillings per annum, rags included; and I believe no gentleman would refuse to give ten shillings for the carcass of a good fat child, which, as I have said, will make four dishes of excellent nutritive meat, when he hath only some particular whim, or his own family to dine with him. Thus the squire will learn to be a good landlord, and grow popular among his tenants, the mother will have eight shillings neat profit, and be fit for work till she produces another child.

Those who are more thrifty (as I must confess the times require) may flay the carcass; the skin of which, artificially dressed, will make admirable gloves for ladies, and summer boots for fine gentlemen.

As to our City of Dublin, shambles may be appointed for this purpose, in the most convenient parts of it, and butchers we may be assured will not be wanting; although I rather recommend buying the children alive, and dressing them hot from the knife, as we do roasting pigs.

A very worthy person, a true lover of his country, and whose virtues I highly esteem, was lately pleased in discoursing on this matter, to offer a refinement upon my scheme. He said, that many gentlemen of this kingdom, having of late destroyed their deer, he conceived that the want of venison might be well supplied by the bodies of young lads and maidens, not exceeding fourteen years of age, nor under twelve; so great a number of both sexes in every county being now ready to starve for want of work and services; and these to be disposed of by their parents if alive, or otherwise by their nearest relations. But with due deference to so excellent a friend, and so deserving a patriot, I cannot be altogether in his sentiments; for as to the males, my American acquaintance assured me from frequent experience, that their flesh was generally tough and lean, like that of our schoolboys, by continual exercise, and their taste disagreeable, and to fatten them would not answer the charge. Then as to the females, it would, I think, with humble submission, be a loss to the publick, because they soon would become breeders themselves; and besides, it is not improbable that some scrupulous people might be apt to censure such a practice, (although indeed very unjustly) as a little bordering upon cruelty, which, I confess, hath always been with me the strongest objection against any project, how well soever intended.

But in order to justify my friend, he confessed, that this expedient was put into his head by the famous Psalmannaor, a native of the island Formosa, who came from thence to London, above twenty years ago, and in conversation told my friend, that in his country, when any young person happened to be put to death, the executioner sold the carcass to persons of quality, as a prime dainty; and that, in his time, the body of a plump girl of fifteen, who was crucified for an attempt to poison the Emperor, was sold to his Imperial majesty's prime minister of state, and other great mandarins of the court in joints from the gibbet, at four hundred crowns. Neither indeed can I deny, that if the same use were made of several plump young girls in this town, who without one single groat to their fortunes, cannot stir abroad without a chair, and appear at a playhouse and assemblies in foreign fineries which they never will pay for, the kingdom would not be the worse.

Some persons of a desponding spirit are in great concern about that vast number of poor people who are aged, diseased, or maimed; and I have been desired to employ my thoughts what course may be taken, to ease the nation of so grievous an incumbrance. But I am not in the least pain upon that matter, because it is very well known, that they are every day dying, and rotting, by cold and famine, and filth, and vermin, as fast as can be reasonably expected. And as to the young labourers, they are now in almost as hopeful a condition. They cannot get work, and consequently pine away from want of nourishment, to a degree, that if at any time they are accidentally hired to common labour, they have not strength to perform it, and thus the country and themselves are happily delivered from the evils to come.

I have too long digressed, and therefore shall return to my subject. I think the advantages by the proposal which I have made are obvious and many, as well as of the highest importance.

For first, as I have already observed, it would greatly lessen the number of Papists, with whom we are yearly overrun, being the principal breeders of the nation, as well as our most dangerous enemies, and who stay at home upon purpose with a design to deliver the kingdom to the Pretender, hoping to take their advantage by the absence of so many good Protestants, who have chosen rather to leave their country, than stay at home and pay tithes against their conscience to an episcopal curate.

Secondly, The poorer tenants will have something valuable of their own, which by law may be made liable to a distress, and help to pay their landlord's rent, their corn and cattle being already seized, and money a thing unknown.

Thirdly, Whereas the maintenance of a hundred thousand children, from two years old, and upwards, cannot be computed at less than ten shillings a piece per annum, the nation's stock will be thereby increased fifty thousand pounds per annum, besides the profit of a new dish, introduced to the tables of all gentlemen of fortune in the kingdom, who have any refinement in taste. And the money will circulate among our selves, the goods being entirely of our own growth and manufacture.

Fourthly, The constant breeders, besides the gain of eight shillings sterling per annum by the sale of their children, will be rid of the charge of maintaining them after the first year.

Fifthly, This food would likewise bring great custom to taverns, where the vintners will certainly be so prudent as to procure the best receipts for dressing it to perfection; and consequently have their houses frequented by all the fine gentlemen, who justly value themselves upon their knowledge in good eating; and a skilful cook, who understands how to oblige his guests, will contrive to make it as expensive as they please.

Sixthly, This would be a great inducement to marriage, which all wise nations have either encouraged by rewards, or enforced by laws and penalties. It would encourage the care and tenderness of mothers towards their children, when they were sure of a settlement for life to the poor babes, provided in some sort by the publick, to their annual profit instead of expense. We should soon see an honest emulation among the married women, which of them could bring the fattest child to the market. Men would become as fond of their tenants, as of their flocks, and their pregnancy, as they are now of their mares in foal, their cows in calf, or sows when they are ready to farrow; nor offer to beat or kick them (as is too frequent a practice) for fear of a miscarriage.

Many other advantages might be enumerated. For instance, the addition of some thousand carcasses in our exportation of barrel'd beef: the propagation of swine's flesh, and improvement in the art of making good bacon, so much wanted among us by the great destruction of pigs, too frequent at our tables; which are no way comparable in taste or magnificence to a well grown, fat yearling child, which roasted whole will make a considerable figure at a Lord Mayor's feast, or any other publick entertainment. But this, and many others, I omit, being stuporous of brevity.

Supposing that one thousand families in this city, would be constant customers for infants flesh, besides others who might have it at merry meetings, particularly wedding and christenings, I compute that Dublin would take off annually about twenty thousand carcasses; and the rest of the kingdom (where probably they will be sold somewhat cheaper) the remaining eighty thousand.

I can think of no one objection, that will possibly be raised against this proposal, unless it should be urged, that the number of people will be thereby much lessened in the kingdom. This I freely own, and was indeed one principal design in offering it to the world. I desire the reader will observe, that I calculate my remedy for this one individual Kingdom of Ireland, and for no other that ever was, is, or, I think, ever can be upon Earth. Therefore let no man talk to me of other expedients: Of taxing our absentees at five shillings a pound: Of using neither clothes, nor household furniture, except what is of our own growth and manufacture: Of utterly rejecting the materials and instruments that promote foreign luxury: Of curing the expensiveness of pride, vanity, idleness, and gaming in our women: Of introducing a vein of parsimony, prudence and temperance: Of learning to love our country, wherein we differ even from Laplanders, and the inhabitants of Topinamoo; Of quitting our animosities and factions, nor acting any longer like the Jews, who were murdering one another at the very moment their city was taken: Of being a little cautious not to sell our country and consciences for nothing: Of teaching landlords to have at least one degree of mercy towards their tenants. Lastly, of putting a spirit of honesty, industry, and skill into our shopkeepers, who, if a resolution could now be taken to buy only our native goods, would immediately unite to cheat and exact upon us in the price, the measure, and the goodness, nor could ever yet be brought to make one fair proposal of just dealing, though often and earnestly invited to it.

Therefore I repeat, let no man talk to me of these and the like expedients, till he hath at least some glimpse of hope, that there will ever be some hearty and sincere attempt to put them into practice.

But, as to myself, having been wearied out for many years with offering vain, idle, visionary thoughts, and at length utterly despairing of success, I fortunately fell upon this proposal, which, as it is wholly new, so it hath something solid and real, of no expense and little trouble, full in our own power, and whereby we can incur no danger in disabling England. For this kind of commodity will not bear exportation; and flesh being of too tender a consistence, to admit a long continuance in salt, although perhaps I could name a country, which would be glad to eat up our whole nation without it.

After all, I am not so violently bent upon my own opinion, as to reject any offer, proposed by wise men, which shall be found equally innocent, cheap, easy, and effectual. But before something of that kind shall be advanced in contradiction to my scheme, and offering a better, I desire the author or authors will be pleased naturally to consider two points. First, As things now stand, how they will be able to find food and raiment for a hundred thousand useless mouths and backs. And secondly, There being a round million of creatures in humane figure throughout this kingdom, whose whole subsistence put into a common stock, would leave them in debt two million of pounds sterling, adding those who are beggars by profession, to the bulk of farmers, cottagers and labourers, with their wives and children, who are beggars in effect; I desire those politicians who dislike my overture, and may perhaps be so bold to attempt an answer, that they will first ask the parents of these mortals, whether they would not at this day think it a great happiness to have been sold for food at a year old, in the manner I prescribe, and thereby have avoided such a perpetual scene of misfortunes, as they have since gone through, by the oppression of landlords, the impossibility of paying rent without money, the want of common sustenance, with neither house nor clothes to cover them from the inclemencies of the weather, and the most inevitable prospect of intailing the like, or greater miseries, upon their breed for ever.

I profess in the sincerity of my heart, that I have not the least personal interest in endeavouring to promote this necessary work, having no other motive than the publick good of my country, by advancing our trade, providing for infants, relieving the poor, and giving some pleasure to the rich. I have no children, by which I can propose to get a single penny; the youngest being nine years old, and my wife past child-bearing.")

```
mtokensfromfile <- tokenize_words(modestpropfromfile)
#tokens from file
listmpfromfile <- tibble(word = mtokensfromfile[[1]])
View(listmpfromfile)
```

```
#using the tokenize function from tokenizers package
#there are several options, depending on what you need
#you can tokenize by paragraphs, sentences, lines, or words
mp_paragraphs <- tokenize_paragraphs(modestprop)
```

```
#inspect the results
print(mp_paragraphs)
mp_sentences <- tokenize_sentences(modestprop)
```

```
#inspect the results
print(mp_sentences)
```

```
#tokenize words
mp_tokens <- tokenize_words(modestprop)
```

```
#get the length
length(mp_tokens)
#what's up with this output?
```

```
#items
mp_stems <- tokenize_word_stems(modestprop)
print(mp_stems)
```

```
#Note that the output of length(mp_tokens) is 1. This is because mp_tokens is a list object with one entry for each text document.
#in order to view the results for this first list object we need to use the syntax [[1]]
mp_tokens[[1]]
```

```
#tidying up our data object to better analyze and access our data
#tibble is a function used to create and store data in a simple type of data frames in R
modestproptibble <- tibble(word = mp_tokens[[1]])
View(modestproptibble)
```

```
#count the words
modestpropcount <- modestproptibble %>%
  count(word, sort = TRUE)
View(modestpropcount)
```

```
#plot the most frequent tokens:
modestproptibble %>%
  count(word, sort = TRUE) %>%
  filter(n > 3) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(n, word)) +
  geom_col() +
  labs(y = NULL)
```

```
#output is messy! let's fix it
modestproptibble %>%
  count(word, sort = TRUE) %>%
  filter(n > 20) %>% #adjust number here to whatever you want!
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(n, word)) +
  geom_col() +
  labs(y = NULL)
```

```
Multword units, lexical bundles, and ngrams.
```

```
mp_bigrams <- tibble(text = modestprop) %>%
  unnest_tokens(bigram, text, token = "ngram", n = 2)
```

```
#show by frequency
mp_bigrams %>%
  count(bigram, sort = TRUE)
```

```
mp_trigrams <- tibble(text = modestprop) %>%
  unnest_tokens(trigram, text, token = "ngrams", n = 3)
```

```
#by frequency
mp_trigrams %>%
  count(trigram, sort = TRUE)
```

```
mp_four_word_ngrams <- tibble(text = modestprop) %>%
  unnest_tokens(ngram, text, token = "ngrams", n = 3)
```

```
mp_four_word_ngrams %>%
  count(ngram, sort = TRUE)
```

Part of speech tagging with udpipe.

Udpipe is an incredibly useful library for text analysis. It supports pos-tagging for many different languages, including Spanish, Urdu, Arabic, Hebrew, Russian, Italian, Dutch, English, etc.

```
library(udpipe)
teststr1 <- data.frame(doc_id = c("doc0", "doc1"),
  text = c("Udpipe is great for part of speech tagging.", "Go dawgs!"))
View(teststr1)
teststr2 <- udpipe(teststr1, "english")
```

```
#apply it to modest proposal
modestmp <- udpipe(modestprop, "english")
View(tagged_mp)
```

```
tagged_mp %>% count(tagged_mp$lemma, sort = TRUE) #get all the lemmas sorted
View(tagged_mp)
```

```
tagged_mp2 <- data.frame(word=tagged_mp$token, pos=tagged_mp$upos)
View(tagged_mp2)
tagged_mp2 %>% count(pos, sort = TRUE)
```

```
tagged_mp2filtered <- tagged_mp2 %>%
  filter(str_detect(pos, pattern='[!punct:]'))
```

```
View(tagged_mp2filtered)
tagged_mp2filtered %>% count(pos, sort = TRUE) %>%
  mutate(pos = reorder(pos, n)) %>%
  ggplot(aes(n, pos)) +
  geom_col() +
  ggtitle("Title") +
  theme_light()
```

Next week, we will learn about the quanteda package options for analysis and about different types of sentiment analysis and topic modeling in R

Check out these tutorials for more information or any of the works cited below!

[UDPipe Natural Language Processing Annotation](#)

[Tidyverse tutorial](#)

[Tokenizers package tutorial](#)

Works Cited

Arnold,Taylor, and Tilton-Lauren. 2017. [Basic Text Processing in R](#). The Programming Historian.

Benoit, K, Obeng, A, Watanabe, K, Kohel, Matsuo, A, Nutty, Paul, and Muller, Stefan. 2020. [readtext R package](#).

Benoit, K, and Matsuo, Akikata. 2020. [A Guide to Using spacyr](#).

Benoit K, Watanabe K, Wang H, Nutty P, Obeng A, Müller S, Matsuo A. 2018. [quanteda: An R package for the quantitative analysis of textual data](#).

Journal of Open Source Software, 3(30), 774. doi: 10.21105/joss.00774

Feinerer, Ingo. 2020. Introduction to the R Package Text Mining in R.

Mullen, Lincoln. 2018. [Introduction to the tokenizers package](#).

Mullen, Lincoln, Keyes, Os, Selivanov, Dmitriy, Arnold, Jeffrey, Kenneth, Benoit. 2018. [tokenizers R package](#).

Slige, Julia, and David Robinson. 2016. [Text Mining with R: A Tidy Approach](#).

Slige, Julia, and David Robinson. 2020. [Text Mining with R: A Tidy Approach](#).

Wickham, Hadley et al. 2019. [Welcome to the tidyverse](#).

Wickham, Jan. 2020. [UDPipe Natural Language Processing Annotation](#).