Chapter 4: Text Mining British Concepts in the 1833 Debates About the Abolition of Slavery

The British Abolition of Slavery in 1833, alongside the abolition of the slave trade by British naval vessels in 1806, were once celebrated as the sign that Britain was a nation committed to moral progress around the world. That claim has been significantly modified by historical research in recent years, as researchers have underscored that the abolition of slavery failed to end many forms of slavery; it wooed slave owners through instituting a prolonged form of slavery called "apprenticeships." In India and the rest of Asia, slavery meanwhile persisted unreformed.

Studying the interplay between white male politicians as they debated the fate of millions of enslaved persons from the West Indies to South Africa is in many ways an out-of-date exercise. Today, we recognize the degree to which enslaved people across British empire struggled to liberate themselves. A persistent tradition of uprisings and riots marked their ongoing resistance. The life of ideas and leaders — even women leaders like Queen Nannie — are important to this story.

We will find only an indirect trace of how the freedom struggles of the Caribbean changed laws and ideas in Britain. The personages and uprisings that rocked the West Indies were rarely named in Westminster. The direct records of those events are held in pamphlets and documents in the Caribbean itself, slowly being digitalized through other means.

And yet, those freedom struggles half a world a way did matter in many respects to British life. They circulated among evangelical readers; they forced philosophers of capitalism to reckon with the way that capitalism had been instantiated as misery in many places where it was enacted. British debates about slavery came to a head in the halls of parliament after 1832, when an evangelical middle class got the vote after decades of political struggles of their own. In 1833, the elected representatives of the evangelical middle class committed themselves to talking about slavery as a moral ill and bringing it to an end across empire.

One question that might allow us to unpack the impact of abolition is how British understandings of empire and capitalism were changed in the course of those debates — in a year when both parliament and the newspapers were filled with a detailed discussion of the suffering caused by the slave system. One way of approaching the question of what happened in 1833 is to look for the ideas that were up for debate. Were any of the fundamental concepts of the enlightenment challenged or redefined in the course of these arguments?

The Importance of Background Knowledge

Before we begin to think about the contribution that text mining can make to the problem, we first need to be able to summarize the understanding and debates of the historians who have studied these questions. The paragraphs above summarize some of the different perspectives on the slavery question taken by historians from the recent and more distant past. In short, we need some background knowledge. But how does an analyst get there, if they don't first have a series of prepared paragraphs to introduce them to a historical problem?

"Background knowledge" in a subject field refers to the kind of knowledge that one acquires by reading the major textbooks and secondary monographs from the field of History that cover the historical time period in question. This is extensive knowledge about which sources have been consulted, what scholars have debated, whether the actions of kings or parliaments or working-class people mattered at this period of time and why.

Remarks about the importance of reading background information may be experienced as off-putting by some undergraduate students of data, who want to be assured that in a course on text mining that they will acquire the tools to make wonderful discoveries. But just as students of statistics must study more than an introductory course in statistics to make a novel contribution, so too students of text mining must enroll in more than one course to do really important work. Reading a wikipedia entry cannot substitute for the kind of background knowledge that a student would acquire in the course of taking an introductory course on the History of Victorian England. Nor can Text Mining for Historical Analysis make up for courses in the undergraduate study of History, although this book's companion volume, The Dangerous Art of Text Mining, includes many arguments about the specific kind of knowledge of sources, obfuscations, agency, and experience that historical study makes available. Producing insight worthy of the attention of multiple fields is no simple game. It cannot be produced at the touch of a button or the application of a new algorithm to new data; it requires adjusting the algorithm, rethinking the questions, examining the data, and iterating through the work until something truly surprising has come to light.

A shortcut which Guldi recommends in The Dangerous Art of Text Mining for those who want to achieve results before enrolling in a specialist course on the historical background material is to engage in an ongoing collaboration with a professor, librarian, or graduate student already familiar with the background material, allowing that scholar's sense of validation and surprise to guide the work with data. In 'hybrid' work, where computer scientists and humanists work on the same team to produce new knowledge, collaborating over multiple iterations and investigations, there can be a division of labor around the research. In general, a high caliber of work can only be accomplished by teams where at least one valued member is as serious about history as they are about algorithmic study.

Not every historian needs access to every kind of background knowledge. If a course on Victorian Britain offers the ideal background for text mining the parliamentary debates, a course on American history offers a much better background to working with the Congressional Record. Scholars who have studied Latin American, African, or Asian history will necessarily have the background to work with other datasets. The point here is that understanding a time and a place represents a commitment of reading dozens of books. No single Wikipedia article or textbook will substitute for this work, and only this background work really equips an analyst to ask the questions about consensus and surprise that outfit an analyst for understanding when their data work is validating what scholars already know and when their data work has advanced to the discovery of something truly surprising.

For the purpose of our question about the abolition debatees in parliament, we recommend extra background reading in the major sources we consulted to summarize the debates above:

- David Ryden, West Indian Slavery and British Abolition, 1783-1807 (Cambridge: Cambridge University Press, 2005)
- Adam Hochschild, Bury the Chains: Prophets and Rebels in the Fight to Free an Empire (Boston: Houghton Mifflen, 2006).
- Christopher Leslie Brown, Moral Capital: Foundations of British Abolitionism (Chapel Hill: University of North Carolina Press, 2006)
- Richard Huzzey, Freedom Burning: Salvery and Empire in Victorian Britain (Ithaca: Cornell University Press, 2012).

- Claudius K. Fergus, Revolutionary Emancipation, Slavery and Abolitionism in the British West Indies (Baton Rouge, Louisiana State University Press, 2013).
- Robert Burroughs and Richard Huzzey, eds., The Suppression of the Atlantic Slave Trade (Manchester: Manchester University Press, 2015)
- Paula Dumas, Proslavery Britain: Fighting for Slavery in an Era of Abolition (New York: Palgrave Macmillan, 2016).
- Padraic X. Scanlan, Slave Empire: How Slavery Built Modern Britain (Boston: Little, Brown, 2020).
- Tom Zoellner, Island on Fire: The Revolt that Ended Slavery in the British Empire (Cambridge: Harvard University Press, 2022).

Note that we have attempted to gather titles from the past two decades, with perspectives ranging from an overview of the abolition movement to the biography of the abolitionists, to the pro-slavery lobby that argued against abolition and for apprenticeships, to the larger question of how imbricated slavery was within the British economy, to the Jamaican revolt of 1831 that forced Britain to end slavery.

Engagement with historical secondary sources is one of the markers of serious Digital History work which is frequently absent in publications from Computer Science or Data Science, with the result that it is sometimes unclear what the original contribution of a given article is to the historical literature. We believe that setting a higher standard of innovation in scholarship — by demanding findings that are literally surprising to the field – is a healthy practice, raising the bar for computational investigations of the human past.

The analyst beginning research on a problem of this kind does not necessarily have to read all of these books with equal attentiveness. But it is useful to spend a few hours reading at least the introductions to these books (or more recent contributions to the debate) to see what the authors agree and disagree about, which themes or ideas they find most interesting.

In the sections of text that follow, as we pursue the question more deeply, we will return to secondary sources, seeking out more specialized texts, for instance biographies of the individuals whose words we count. Pay attention to the footnotes and the texts that we cite to understand how the need for background knowledge evolves with research.

Finding the Research Question that the Tools of Text Mining Can Address

First, we need to think about the fit between the tools of text mining and the issue of the abolition of slavery. As explained above, we don't have the source base in Hansard to contribute to burning questions about the role of enslaved persons in pursuing their own freedom. If all we have is Hansard, we need to retailor our question to the dataset we have. With Hansard, we can count mentions of slavery and the way that slavery is addressed by the spakers who were in power at the time when slavery was nominally abolished throughout British empire. Can we write a research question that might light to insights on the basis of word count?

One place to start is with the speakers in parliament who contributed the most to the 1833 debates about slavery. As we saw in a previous chapter, one place where text mining excels is simply counting the number of words contributed to a given issue by individual speakers. Based on word counts, we can expect to accurately identify the speakers in parliament who contributed the most to the slavery question.

We can then move to asking about the language they shared to the language that makes each of their contributions unique. Another place where text mining excels is counting the words from each individual speaker, finding the words they shared and the words that were most distinctive of each speaker.

Why would we want to understand the words that made the principal speakers different? Because it may tell us something important about what they disagreed about. An understanding of the poles of debate can lead us to a detailed case about the tensions that emerged in parliament around abolition.

From this question we may be able to navigate to an original contribution about the role of arguments in parliament, perhaps contributing information that other historians have not been able to pull from their research.

Top Speakers in the Debates About Slavery

In a previous chapter, we looked into the top speakers in parliament in a given year. In this exercise, we want to begin by finding just those speakers who spoke the most in debates about slavery. We can use `str_detect()` to look for the word "slavery" in the titles of debate. Then we could the top speakers.

If you are new to code, take a moment to notice how changing only one line in the code below produces a different angle of analysis. The code below is otherwise very similar to the code in the previous chapters. One of the great pleasures of coding is the facility with which a few skills can quickly become powerful in the hands of an analyst who has many questions.

```r
# load packages
library("tidyverse")
library("tidytext")
library("lubridate")
library("hansardr")
library("ggrepel")

# load data
# load data
data("hansard_1830")
data("debate_metadata_1830")

# merge the Hansard speech text with metadata (e.g., debate title and speech date)
# create a new column 'year' by extracting the year from the full speech date
# find just the debates about slavery and who spoke them
slavery_1830 <- hansard_1830 %>%  # Create a new dataset called 'debates_1830'
  left_join(debate_metadata_1830, by = "sentence_id") %>% # Join with the debate metadata
  filter(str_detect(tolower(debate), "slavery")) %>% # filter title for 'slavery'
  mutate(year = year(speechdate)) # Extract and add the year from the 'speechdate' column

slavery_debates_1830 <- slavery_1830 %>%
  left_join(speaker_metadata_1830) %>% # join datasets
  select(sentence_id, speaker, suggested_speaker, text, year) %>% # retain relevant columns
  filter(suggested_speaker != "") # remove any lines with missing speaker information

# break up the text into words
slavery_words_1830 <- slavery_debates_1830 %>% # create a new dataset
    unnest_tokens(word, text) # break up the text into words
```

```r
# count the words per speaker
words_per_speaker_slavery_1830 <- slavery_words_1830 %>% # create a new dataset
  group_by(speaker) %>% # group the data by speaker and year
  summarize(words_per_speaker = n()) %>% # create a column with total words per speaker
  arrange(desc(words_per_speaker)) # arrange the words in descending order

# find the top speakers
top_slavery_speakers <- words_per_speaker_slavery_1830 %>% # create a new dataset
  ungroup() %>% # remove the previous data grouping
  arrange(desc(words_per_speaker)) %>% # arrange in descending order
  slice(1:20) # take only the top n speakers

# calculates the number of words each speaker said per year in the 1830 slavery dataset
# group the text data by speaker and year, count the words,
# and arrange the results in descending order
words_per_speaker_per_year_slavery_1830 <- slavery_words_1830 %>% # create a new dataset
  group_by(speaker, year) %>% # group by speaker and year
  summarize(words_pspy = n()) %>% # count words and store in new column
  arrange(desc(words_pspy)) # sort in descending order of word count

# create a summary of the number of words each speaker contributed per year
# in the slavery debates of 1830
# the resulting dataset helps identify the most vocal speakers
# and their yearly distribution of speech.
words_per_speaker_per_year_slavery_1830 <- slavery_words_1830 %>% # create a new dataset
  group_by(speaker, year) %>% # group by speaker and year
  summarize(words_pspy = n()) %>% # count words per speaker per year
  arrange(desc(words_pspy)) # sort from highest to lowest word count

# find the words per speaker per year for just the top speakers
top_words_pspy_slavery_1830 <- words_per_speaker_per_year_slavery_1830 %>%
  filter(speaker %in% unique(top_slavery_speakers$speaker)) # retain only relevant rows

# visualize who spoke about slavery in 1830, how much they said, and when they spoke.
# each point represents a speaker's total words for a given year
# the x-axis shows years (rounded to whole numbers), and the y-axis shows word counts.
ggplot(top_words_pspy_slavery_1830, # make a graph
       aes(x = year, # x axis = year
           y = words_pspy, # y axis = word count
           label = speaker)) + # use speaker names for labels
  geom_text_repel(size = 2) + # repel overlapping text labels
  geom_point() + # plot each point
  ggtitle("Who spoke about slavery, how much, and when?") + # plot title
  ylab("Number of words spoken") + # y axis label
  scale_x_continuous(breaks = # whole number x axis
                       function(x)
                         unique(floor(pretty(seq(min(x),
```
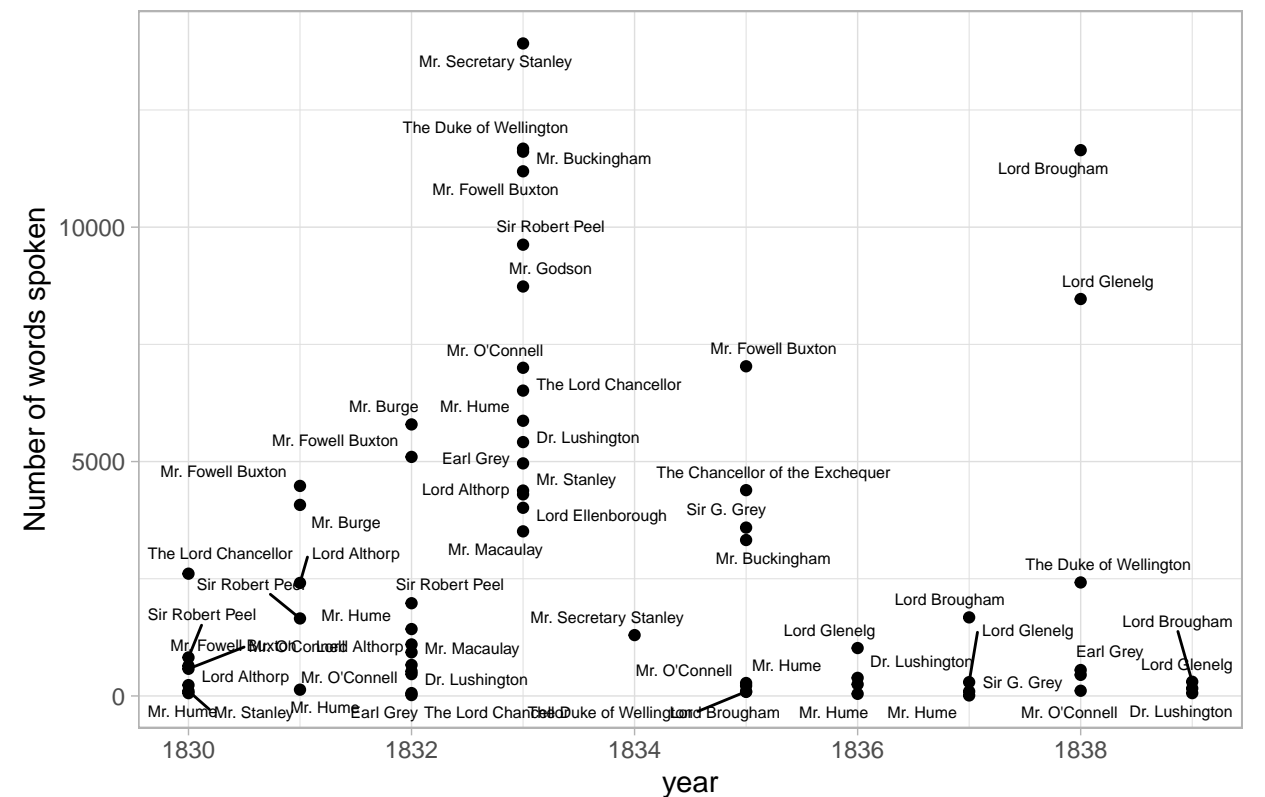
```
                                                             (max(x) + 1) * 1))))) +
  theme_light() # light theme styling
```

Warning: ggrepel: 3 unlabeled data points (too many overlaps). Consider
increasing max.overlaps

Who spoke about slavery, how much, and when?

You may notice that we have produced a different sort of visualization than the bar charts of the previous chapter. The ggplot library of graphics is a highly flexible language for producing many visualizations from the same data. Rather than the command for a bar chart – `geom_col()` – we have used a command for a dot plot – `geom_point()`. With this plot we also introduce another library, `ggrepel`, which gives us tools for enhancing the readability of our plot by making sure that labels do not overlap.

What does this chart of the top speakers about slavery tell us? The chart gives us some suggestions about where to start looking. It tells us that the five speakers who contributed the most to the debate in 1830 spoke twice as much as the speakers in other debates of the decade. The list of these speakers gives us a good place to start looking at data. The main speakers in 1833 were Edward Stanley, the Duke of Wellington, James Silk Buckingham, Fowell Burton, and Robert Peel. We also see the evidence of a second wave of debates, when issues of indentured servitude came to parliament's attention in 1838. Then, Lord Brougham and Lord Gleneig were the main contenders.

Wellington, Peel, Stanley, and Buckingham form a distinct set. Wellington and Stanley had the most in common. Wellington was a former and future prime minister and the hero of Waterloo. Stanley, the Earl of Derby, a future prime minister, was currently in transition from his role as Chief Secretary for Ireland to Secretary of War and the Colonies. His career was associated with the abolition of slavery, military rule and the establishment of schools for Ireland, and the expansion of the vote to the middle class. Both came from noble families; both were Whigs, although Wellington was three decades older.

Exploring the contributions of the top speakers and the tensions between them is one area in which the tools of text mining excel.

Using Our Skills to Explore the Historical Debates About the Abolition of Slavery

Understanding how to group data by speaker and year gives us the possibility of closely studying the contribution of different speakers. For instance, suppose that we wanted to read all of the debates relating to Slavery Abolition Act of 1833, which outlawed slavery in many British colonies, including the Caribbean and South Africa, although not across much of India, where debt slavery would persist for decades, and where the caste system would exclude many people from participation in the economy through the twenty-first century.

There are some new commands below that are used for demonstration purposes to show how one might navigate our newly joined dataset. Don't feel worried if you can't read every line. You will be learning how to aggregate data by speaker very soon.

```
## # A tibble: 6 x 2
##   speaker                words_per_speaker
##   <chr>                             <int>
## 1 Mr. Secretary Stanley             13918
## 2 The Duke of Wellington            11674
## 3 Mr. Buckingham                    11610
## 4 Mr. Fowell Buxton                 11191
## 5 Sir Robert Peel                    9626
## 6 Mr. Godson                         8734
```

Counting how many words were spoken in the 6,000 some sentences about the abolition of British slavery is our first reminder of the bias of this data set: the speakers are white men, many of them of a class invested in slavery. This collection of words is not suitable for investigating narratives that depict lived experiences of dehumanization, recorded from the words of enslaved people.

Nevertheless, distant reading allows us to investigate the dynamics of politics in the British parliamentary system that made the official abolition of slavery possible. We can take a snapshot at the speeches of the most active speakers – Buckingham, Stanley, Wellington, and Peel – by comparing their most-invoked words in these debates.

In the following code, we provide a list of speakers and use the `%in%` operator to filter for any of these speakers.

```r
library(knitr)
library(kableExtra)

# make a list of the names of the top speakers
pattern1 = c("Mr. Fowell Buxton", "Mr. Buckingham", "Mr. Stanley",
             "The Duke of Wellington", "Sir Robert Peel")

# search for the words said by the speakers in pattern1
top_slavery_speakers_1833 <- slavery_debates_1833 %>%
  filter(speaker %in% pattern1) # keep rows containing a speaker's name

top_slavery_speakers_1833 %>%
  head() %>%
  mutate(text = str_trunc(text, 120)) %>%   # optional: shorten long lines
  kable("latex", booktabs = TRUE) %>%
  kable_styling(full_width = FALSE, position = "left") %>%
  column_spec(2, width = "4cm")   # wrap the "text" column
```

sentence_id	speaker	suggested_speaker	text
S3V0016P0_13594	Mr. Fowell Buxton	thomas_buxton_2122	replied, that no man was more conscious than
S3V0016P0_13595	Mr. Fowell Buxton	thomas_buxton_2122	First, if Ministers were prepared with a plan fo
S3V0016P0_13596	Mr. Fowell Buxton	thomas_buxton_2122	He might be thought very obstinate, and he wa
S3V0016P0_13597	Mr. Fowell Buxton	thomas_buxton_2122	Until June next there was not a vacant day in t
S3V0016P0_13598	Mr. Fowell Buxton	thomas_buxton_2122	To postpone his Motion, therefore, would in eff
S3V0016P0_13599	Mr. Fowell Buxton	thomas_buxton_2122	Now, he was convinced, that it was absolutely i

Next, let's break the text into words and clean it up, removing numbers and stop words. We're lemmatizing each word into its word stem so that we can count singular and plural of the same word as one.

```r
library(textstem)

# clean up the list of words
filtered_words_per_top_slavery_speaker_1833 <- top_slavery_speakers_1833 %>%
  unnest_tokens(word, text) %>% # break up into words
  filter(!str_detect(word,"^\\s*[0-9]*\\s*$")) %>% # remove all numbers
  anti_join(stop_words, by = "word") %>% # remove stopwords
  mutate(word = lemmatize_words(word)) # lemmatize the word

# inspect the data
head(filtered_words_per_top_slavery_speaker_1833)
```

```
##         sentence_id          speaker suggested_speaker year      word
##              <char>           <char>            <char> <num>    <char>
## 1: S3V0016P0_13594 Mr. Fowell Buxton thomas_buxton_2122  1833     reply
## 2: S3V0016P0_13594 Mr. Fowell Buxton thomas_buxton_2122  1833 conscious
## 3: S3V0016P0_13594 Mr. Fowell Buxton thomas_buxton_2122  1833   subject
```

```
## 4: S3V0016P0_13594 Mr. Fowell Buxton thomas_buxton_2122  1833  introduce
## 5: S3V0016P0_13594 Mr. Fowell Buxton thomas_buxton_2122  1833 government
## 6: S3V0016P0_13594 Mr. Fowell Buxton thomas_buxton_2122  1833 relinquish
```

Next, let's count the words by how many times each speaker says them.

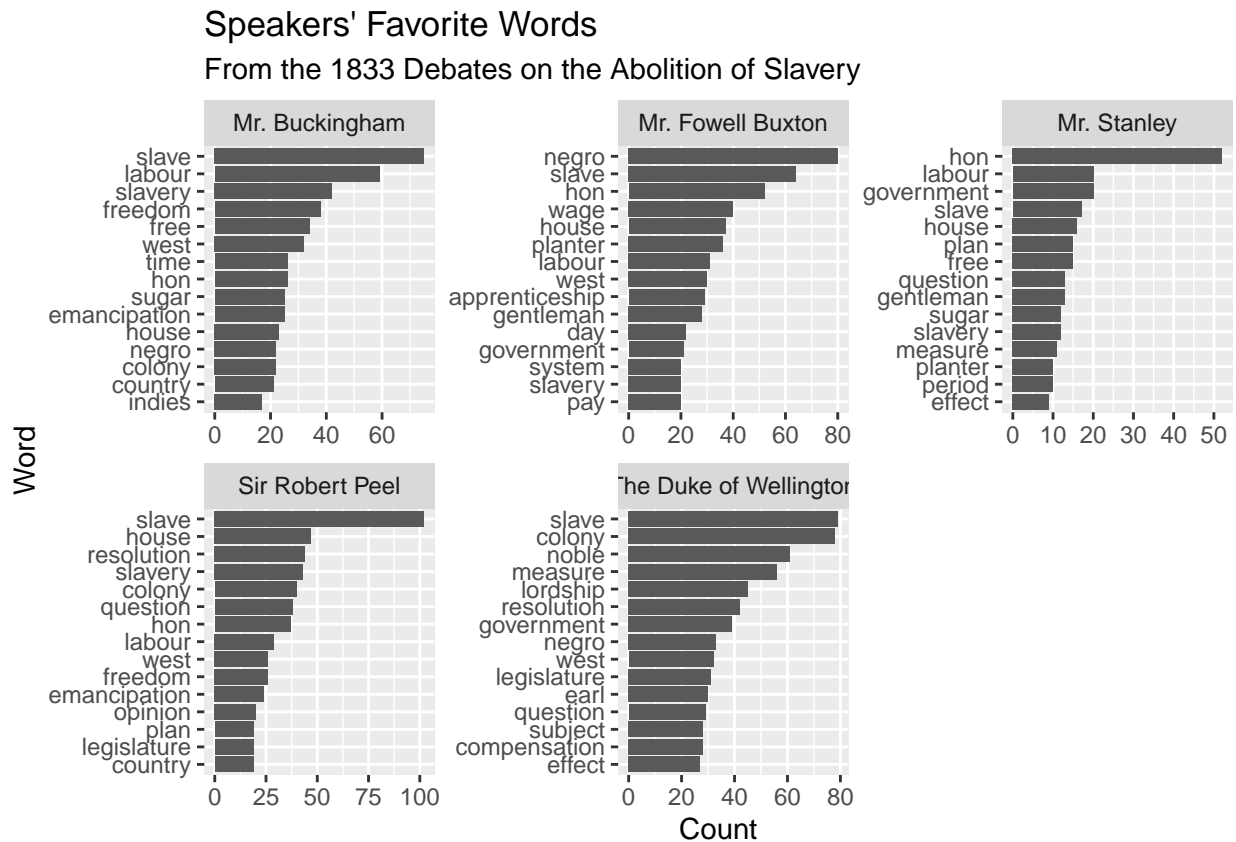For our visualization, we will look at only the words spoken most frequently.

```
top_words_per_abolition_speaker <- filtered_words_per_top_slavery_speaker_1833 %>%
  group_by(speaker, word) %>% # group the data by speaker
  summarize(n = n()) %>% # count how many times each word is spoken by each speaker
  arrange(desc(n)) %>% # arrange in descending order
  slice(1:15)  # find the 15 words that each speaker says the most frequently

kable(top_words_per_abolition_speaker)
```

speaker	word	n
Mr. Buckingham	slave	75
Mr. Buckingham	labour	59
Mr. Buckingham	slavery	42
Mr. Buckingham	freedom	38
Mr. Buckingham	free	34
Mr. Buckingham	west	32
Mr. Buckingham	hon	26
Mr. Buckingham	time	26
Mr. Buckingham	emancipation	25
Mr. Buckingham	sugar	25
Mr. Buckingham	house	23
Mr. Buckingham	colony	22
Mr. Buckingham	negro	22
Mr. Buckingham	country	21
Mr. Buckingham	indies	17
Mr. Fowell Buxton	negro	80
Mr. Fowell Buxton	slave	64
Mr. Fowell Buxton	hon	52
Mr. Fowell Buxton	wage	40
Mr. Fowell Buxton	house	37
Mr. Fowell Buxton	planter	36
Mr. Fowell Buxton	labour	31
Mr. Fowell Buxton	west	30
Mr. Fowell Buxton	apprenticeship	29
Mr. Fowell Buxton	gentleman	28
Mr. Fowell Buxton	day	22
Mr. Fowell Buxton	government	21
Mr. Fowell Buxton	pay	20
Mr. Fowell Buxton	slavery	20
Mr. Fowell Buxton	system	20

speaker	word	n
Mr. Stanley	hon	52
Mr. Stanley	government	20
Mr. Stanley	labour	20
Mr. Stanley	slave	17
Mr. Stanley	house	16
Mr. Stanley	free	15
Mr. Stanley	plan	15
Mr. Stanley	gentleman	13
Mr. Stanley	question	13
Mr. Stanley	slavery	12
Mr. Stanley	sugar	12
Mr. Stanley	measure	11
Mr. Stanley	period	10
Mr. Stanley	planter	10
Mr. Stanley	effect	9
Sir Robert Peel	slave	102
Sir Robert Peel	house	47
Sir Robert Peel	resolution	44
Sir Robert Peel	slavery	43
Sir Robert Peel	colony	40
Sir Robert Peel	question	38
Sir Robert Peel	hon	37
Sir Robert Peel	labour	29
Sir Robert Peel	freedom	26
Sir Robert Peel	west	26
Sir Robert Peel	emancipation	24
Sir Robert Peel	opinion	20
Sir Robert Peel	country	19
Sir Robert Peel	legislature	19
Sir Robert Peel	plan	19
The Duke of Wellington	slave	79
The Duke of Wellington	colony	78
The Duke of Wellington	noble	61
The Duke of Wellington	measure	56
The Duke of Wellington	lordship	45
The Duke of Wellington	resolution	42
The Duke of Wellington	government	39
The Duke of Wellington	negro	33
The Duke of Wellington	west	32
The Duke of Wellington	legislature	31
The Duke of Wellington	earl	30
The Duke of Wellington	question	29
The Duke of Wellington	compensation	28
The Duke of Wellington	subject	28
The Duke of Wellington	effect	27

Now visualize:

Speakers' Favorite Words
From the 1833 Debates on the Abolition of Slavery

From Concepts to ideas

From Concepts to ideas

What can we learn from this visualization about the history of slavery? How might we use the tools of text mining to learn something new about this important historical moment?

What shows up clearly in the counts of most frequently-used words is a series of keywords that motivated debate. The fact that all five speakers invoke the terms "slavery," "freedom," and "labour" suggests that they are operating as what historians call "concepts" – the theoretical constellations of ideas that produced much of the modern world. Terms such as "nation," "republic," "democracy," and "development" were argued over and redefined over the course of decades and centuries, until the terms eventually stabilized.

Historians believe that identifying the debates over these words is key to understanding the abstract forces that remade society, or the "conceptual changes" by which modern understanding was produced. Peter de Bolla and his coauthors defined concepts as " 'cultural entities' held by many agents collectively, that is held by a 'culture'." (de Bolla et. al). They argue that concepts allow cultures to feel their way through increasingly abstract modes of ideation, because concepts "provide the scaffolding that enables one to understand whatever is the object of attention and thought." In the era of the debates over the abolition of slavery, the concept "freedom" was doing just such important work, allowing ordinary British people and the politicians who represented them to reckon with certain aspects of the abhorrent economic reality that made Britain

so rich in the age of empire: the fact that Britain's wealth was largely due to the plunder of wealth from foreign lands and especially the theft of time from the millions of enslaved humans who harvested sugar, tea, cotton, and the other raw materials upon which empire's trade was based. Talking about "freedom" formed the basis for reckoning with capitalism and empire as systems, for holding them account to a higher set of principles. That work was at its infancy in the era of the British laws that outlawed the Atlantic slave trade and slavery in some parts of British empire, but the work of defining and promoting freedom continues to this day as the specific cases and principles that comprise freedom continue to be brought into consciousness through the work of social movements, politics, and ideation.

Historically speaking, as concepts arise, they tend to be contested. At the beginning, there was no single definition of "science" or "nation" to which all parties subscribed; these terms had to be debated and argued over until they became abstract ideas that could be invoked in support of other debates. Corpus linguistics gives us useful clues about how to identify potential concepts: as a rule of thumb, the words spoken about the most in any debates are those that are most contested. In the debates of 1833, we can tell that the words "slavery," "freedom," and "labour" may fit such a description. Often the words spoken most frequently are words whose definition is up for grabs; speakers who disagree with each other will talk at great length to assert their own definition of one of these key concepts. The two speakers who mention freedom the most in 1833 are Buckingham and Peel, and we might expect that they represent two poles of disagreement about what freedom is or should be.

One part of unpacking the life of concepts is to understand the contestation of concepts by individuals as they formulate ideas and arguments. In this usage, we follow de Bolla and his collaborators call the individual uses of concepts by writers and speakers "ideas" (de Bolla et al Ch 2.1.2). Therefore, for the following discussion, let us use "concept" to reference that shared terminology of values, especially "freedom," while reserving "idea" for any other terms we might use to understand how individual speakers make their cases. When people in the past argued about how to define terms like "freedom," they used many examples. Words such as "wages," "apprenticeship," and "regulation" were used in the slavery debates to signal how freedom was to be achieved.

We already have enough material to hypothesize about some major differences that marked the speakers out in their construction of the problems of freedom and slavery. Buckingham and Buxton, representing the abolition lobby, are more inclined to talk about the "negro" – a term whose usage has changed, but that in its early nineteenth-century context, was a reference to a color – black – via the Latin *nigrum*. Perhaps they were inclined to reference the individual experiences of enslaved persons – although this is a conjecture, something we'd need to check in the text. Buckingham also uses words for abstract values such as "enjoyment' and"life," where Buxton is more willing to talk about facts on the ground in "Jamaica," for instance the "insurrection."

Problems of imperial governance concern the other speakers, whose personal background was associated less with middle-class abolition movements than with the military and colonial order of British empire. Peel speaks of "colonies," "compensation," and "proprietors;" if we have read some background to the debates, we might guess that Peel is wrapped up with the technical questions of whether the imperial parliament in Westminster has the right to tell the colonial government in Jamaica, and what will be the fallout if the Westminster parliament tells the white planters in Jamaica that their property is suddenly invalid. Stanley's eyes are on markets – he speaks of "sugar," the major project produced by slave labor, "averages" (a hint that he may be speaking in technical terms about economic production, although we would need to read more to understand how), and "compensation" and "colonies," like Peel. Wellington's eyes are also on imperial issues, drawing attention (perhaps) to the implications of debates over slavery for India (although again more reading would be necessary to understand how India is being invoked).

Our many parenthetical remarks in the paragraph are an important reminder that we are merely at the

stage of hypothesis generation. We do not have enough material from this list of keywords to interpret the debates. We may begin to identify words and speakers for further analysis.

Even in a distant reading, we can detect that Robert Peel and the Duke of Wellington were more interested in problems of "compensation" for the owners who benefited from the system of enslavement than in the details of life in the West Indies – an issue of political expediency, but also a reflection of the ethical orientation of parliament in the 1830s.

The Most Distinctive Words of the Top Speakers About Slavery

If we want to see the words that make each speaker distinct as opposed to their top words, we will need a different tool than raw word count. We might hypothesize, for instance, that Buckingham talks about "freedom" more than the other speakers in the group, because "freedom" is ranked as the third word for Buckingham but the eleventh for Buxton, the fifth for Peel, and somewhere lower still for Wellington and Stanley. But the truth is that we don't know how statistically important these differences are without further comparison. Thinking about the problem with greater statistical richness means asking how many times Buckingham uttered "freedom" relative to the number of words Buckingham spoke in the slavery debates overall; we need to know not just the full count of one word, but how many individual words he used and how frequently he said each, relative to the other speakers. In short, statistics become important when we want to compare individuals.

What we need is a "distinctiveness measure," a statistical tool that allows us to rank an index of how "distinctive" a word is of each speaker. Further on in this book, we will go into the technicalities of measures of distinction in greater detail, the varieties of algorithms that can be used for this purpose, and how they work. For now, it is enough to know that "distinctiveness" is a useful tool for understanding which words each speakers says more relative to the other speakers.

Distinctiveness can help us to answer the question: what were the words that Buckingham said that other speakers did not say? What were the words that Peel uttered that other speakers never used? If raw word count helps us to understand which concepts speakers may be arguing over, distinctiveness helps us to interpret which keywords those speakers used to support their argument.

Using the function `bind_tf_idf()` we can apply an index of how distinctive each word is of each speaker. The tf-idf function will create an index for each term-speaker pair – a number which, if high, tells us that the word was very distinctive of each speaker, and if low, tell us that many of the speakers used the same word.

The `bind_tf_idf()` function takes three arguments. The first two are the names of the facets of the data which are to be assessed for how distinctive they are in relationship to each other – in this case, `word` and `speaker`. The third argument, `n`, the count of words per speaker.

```
# calculate the measure of how distinctive each keyword is of each speaker
tfidf_per_top_slavery_speaker_1833 <- top_words_per_abolition_speaker %>%
  bind_tf_idf(word, speaker, n) %>%
  arrange(desc(tf_idf))

# inspect the data
kable(tfidf_per_top_slavery_speaker_1833)
```

speaker	word	n	tf	idf	tf_idf
The Duke of Wellington	noble	61	0.0956113	1.6094379	0.1538804
Mr. Fowell Buxton	wage	40	0.0754717	1.6094379	0.1214670
The Duke of Wellington	lordship	45	0.0705329	1.6094379	0.1135183
Mr. Fowell Buxton	apprenticeship	29	0.0547170	1.6094379	0.0880636
Mr. Buckingham	time	26	0.0533881	1.6094379	0.0859248
The Duke of Wellington	measure	56	0.0877743	0.9162907	0.0804268
Mr. Fowell Buxton	negro	80	0.1509434	0.5108256	0.0771058
The Duke of Wellington	earl	30	0.0470219	1.6094379	0.0756789
Sir Robert Peel	resolution	44	0.0825516	0.9162907	0.0756413
Mr. Buckingham	freedom	38	0.0780287	0.9162907	0.0714970
The Duke of Wellington	compensation	28	0.0438871	1.6094379	0.0706336
The Duke of Wellington	subject	28	0.0438871	1.6094379	0.0706336
Mr. Fowell Buxton	day	22	0.0415094	1.6094379	0.0668069
Mr. Stanley	period	10	0.0408163	1.6094379	0.0656913
Mr. Buckingham	free	34	0.0698152	0.9162907	0.0639710
The Duke of Wellington	colony	78	0.1222571	0.5108256	0.0624520
Mr. Fowell Buxton	planter	36	0.0679245	0.9162907	0.0622386
Mr. Fowell Buxton	pay	20	0.0377358	1.6094379	0.0607335
Mr. Fowell Buxton	system	20	0.0377358	1.6094379	0.0607335
Sir Robert Peel	opinion	20	0.0375235	1.6094379	0.0603917
The Duke of Wellington	resolution	42	0.0658307	0.9162907	0.0603201
Mr. Buckingham	indies	17	0.0349076	1.6094379	0.0561816
Mr. Stanley	free	15	0.0612245	0.9162907	0.0560994
Mr. Stanley	plan	15	0.0612245	0.9162907	0.0560994
Mr. Stanley	gentleman	13	0.0530612	0.9162907	0.0486195
Mr. Fowell Buxton	gentleman	28	0.0528302	0.9162907	0.0484078
Mr. Stanley	hon	52	0.2122449	0.2231436	0.0473611
Mr. Buckingham	emancipation	25	0.0513347	0.9162907	0.0470375
Mr. Buckingham	sugar	25	0.0513347	0.9162907	0.0470375
Mr. Stanley	sugar	12	0.0489796	0.9162907	0.0448795
Sir Robert Peel	freedom	26	0.0487805	0.9162907	0.0446971
The Duke of Wellington	legislature	31	0.0485893	0.9162907	0.0445220
Mr. Stanley	government	20	0.0816327	0.5108256	0.0417001
Sir Robert Peel	emancipation	24	0.0450281	0.9162907	0.0412589
Mr. Stanley	measure	11	0.0448980	0.9162907	0.0411396
Mr. Buckingham	country	21	0.0431211	0.9162907	0.0395115
The Duke of Wellington	effect	27	0.0423197	0.9162907	0.0387772
Sir Robert Peel	colony	40	0.0750469	0.5108256	0.0383359
Mr. Stanley	planter	10	0.0408163	0.9162907	0.0373996
Sir Robert Peel	question	38	0.0712946	0.5108256	0.0364191
Mr. Stanley	effect	9	0.0367347	0.9162907	0.0336597
Sir Robert Peel	country	19	0.0356473	0.9162907	0.0326633
Sir Robert Peel	legislature	19	0.0356473	0.9162907	0.0326633
Sir Robert Peel	plan	19	0.0356473	0.9162907	0.0326633
The Duke of Wellington	government	39	0.0611285	0.5108256	0.0312260
Mr. Stanley	question	13	0.0530612	0.5108256	0.0271050

speaker	word	n	tf	idf	tf_idf
Mr. Buckingham	labour	59	0.1211499	0.2231436	0.0270338
The Duke of Wellington	negro	33	0.0517241	0.5108256	0.0264220
The Duke of Wellington	question	29	0.0454545	0.5108256	0.0232193
Mr. Buckingham	colony	22	0.0451745	0.5108256	0.0230763
Mr. Buckingham	negro	22	0.0451745	0.5108256	0.0230763
Mr. Fowell Buxton	hon	52	0.0981132	0.2231436	0.0218933
Mr. Fowell Buxton	government	21	0.0396226	0.5108256	0.0202403
Sir Robert Peel	house	47	0.0881801	0.2231436	0.0196768
Mr. Buckingham	slavery	42	0.0862423	0.2231436	0.0192444
Mr. Stanley	labour	20	0.0816327	0.2231436	0.0182158
Sir Robert Peel	slavery	43	0.0806754	0.2231436	0.0180022
Mr. Fowell Buxton	house	37	0.0698113	0.2231436	0.0155779
Sir Robert Peel	hon	37	0.0694184	0.2231436	0.0154903
Mr. Buckingham	west	32	0.0657084	0.2231436	0.0146624
Mr. Stanley	house	16	0.0653061	0.2231436	0.0145726
Mr. Fowell Buxton	labour	31	0.0584906	0.2231436	0.0130518
Mr. Fowell Buxton	west	30	0.0566038	0.2231436	0.0126308
Sir Robert Peel	labour	29	0.0544090	0.2231436	0.0121410
Mr. Buckingham	hon	26	0.0533881	0.2231436	0.0119132
The Duke of Wellington	west	32	0.0501567	0.2231436	0.0111922
Mr. Stanley	slavery	12	0.0489796	0.2231436	0.0109295
Sir Robert Peel	west	26	0.0487805	0.2231436	0.0108851
Mr. Buckingham	house	23	0.0472279	0.2231436	0.0105386
Mr. Fowell Buxton	slavery	20	0.0377358	0.2231436	0.0084205
Mr. Buckingham	slave	75	0.1540041	0.0000000	0.0000000
Mr. Fowell Buxton	slave	64	0.1207547	0.0000000	0.0000000
Mr. Stanley	slave	17	0.0693878	0.0000000	0.0000000
Sir Robert Peel	slave	102	0.1913696	0.0000000	0.0000000
The Duke of Wellington	slave	79	0.1238245	0.0000000	0.0000000

Note that the `bind_tf_idf()` function creates three new columns – `tf`, `idf`, and `tf_idf`. The one we really care about is the last, `tf_idf`, an index of how distinctive each word is of each speaker. A very high `tf_idf` marks words that are said very frequently by one speaker and never by any other speakers. A very low tf_idf marks words that are rarely said by anyone or said by everyone with nearly the same frequency. Looking for the high `tf_idf` words for each speaker will tell us which words are most unique to them – a useful clue to the distinctive arguments, ideas, and referents that each speaker contributed to the debate.

Now that we have an index of how distinctive each word is per speaker, we can use this distinctiveness tool to find which words are most distinctive of each speaker.

```
# find the words most distinctive of each speaker
most_distinctive_words_per_abolition_speaker <- tfidf_per_top_slavery_speaker_1833 %>%
  group_by(speaker) %>% # group by speaker
  arrange(desc(tf_idf)) %>% # arrange in descending order of tf_idf, the index of how distinctive each
  slice(1:15) %>% # find the top n words, ranked by tf_idf
  ungroup() %>% # ungroup the words
```

```
  mutate(index = row_number()) %>% # create a new column, "index," which tells what the #1-4 word is fo
  mutate(tf_idf = round(tf_idf, 4)) %>% # round the tf_idf score to only 4 decimal places
  select(index, word, tf_idf, n, speaker) # drop all columns except these

# create a series of smaller datasets reflecting the top speakers
speaker1 <- most_distinctive_words_per_abolition_speaker %>%
  filter(speaker == pattern1[1]) %>%
  select(word, tf_idf)

speaker2 <- most_distinctive_words_per_abolition_speaker %>%
  filter(speaker == pattern1[2]) %>%
  select(word, tf_idf)

speaker3 <- most_distinctive_words_per_abolition_speaker %>%
  filter(speaker == pattern1[3])%>%
  select(word, tf_idf)

speaker4 <- most_distinctive_words_per_abolition_speaker %>%
  filter(speaker == pattern1[4])%>%
  select(word, tf_idf)

speaker5 <- most_distinctive_words_per_abolition_speaker %>%
  filter(speaker == pattern1[5])%>%
  select(word, tf_idf)
```

Next, let's create a series of small tables displaying each speaker, their most distinctive words, and the `tf_idf` score of that word-speaker combination.

We could create a bar graph, of course, but because `tf_idf` measures represent an abstract concept – distinctiveness – visualizing this data risks confusing some possible readers, who might mistake the bar graph of `tf_idf` for a count of frequency. Analysts often have to make choices about how to represent their data so as to generate the least confusion about the meaning of any statistical measure. For abstract indexes like `tf_idf`, generated in the course of our research, we recommend offering simple tables along with prose that explains the significance of the numbers.

```
kable(speaker1, caption = unique(most_distinctive_words_per_abolition_speaker$speaker)[1])
```

Table 3: Mr. Buckingham

word	tf_idf
wage	0.1215
apprenticeship	0.0881
negro	0.0771
day	0.0668
planter	0.0622
pay	0.0607

16

word	tf_idf
system	0.0607
gentleman	0.0484
hon	0.0219
government	0.0202
house	0.0156
labour	0.0131
west	0.0126
slavery	0.0084
slave	0.0000

```
kable(speaker2, caption = unique(most_distinctive_words_per_abolition_speaker$speaker)[2])
```

Table 4: Mr. Fowell Buxton

word	tf_idf
time	0.0859
freedom	0.0715
free	0.0640
indies	0.0562
emancipation	0.0470
sugar	0.0470
country	0.0395
labour	0.0270
colony	0.0231
negro	0.0231
slavery	0.0192
west	0.0147
hon	0.0119
house	0.0105
slave	0.0000

```
kable(speaker3, caption = unique(most_distinctive_words_per_abolition_speaker$speaker)[3])
```

Table 5: Mr. Stanley

word	tf_idf
period	0.0657
free	0.0561
plan	0.0561
gentleman	0.0486
hon	0.0474
sugar	0.0449

word	tf_idf
government	0.0417
measure	0.0411
planter	0.0374
effect	0.0337
question	0.0271
labour	0.0182
house	0.0146
slavery	0.0109
slave	0.0000

```
kable(speaker4, caption = unique(most_distinctive_words_per_abolition_speaker$speaker)[4])
```

Table 6: Sir Robert Peel

word	tf_idf
noble	0.1539
lordship	0.1135
measure	0.0804
earl	0.0757
compensation	0.0706
subject	0.0706
colony	0.0625
resolution	0.0603
legislature	0.0445
effect	0.0388
government	0.0312
negro	0.0264
question	0.0232
west	0.0112
slave	0.0000

```
kable(speaker5, caption = unique(most_distinctive_words_per_abolition_speaker$speaker)[5])
```

Table 7: The Duke of Wellington

word	tf_idf
resolution	0.0756
opinion	0.0604
freedom	0.0447
emancipation	0.0413
colony	0.0383
question	0.0364

word	tf_idf
country	0.0327
legislature	0.0327
plan	0.0327
house	0.0197
slavery	0.0180
hon	0.0155
labour	0.0121
west	0.0109
slave	0.0000

The first thing that our measure of distinctiveness helps to clarify is what is not distinctive. The words "freedom," "labour," and "slavery" are not particularly distinctive of any speaker – all five speakers use these words a great deal, and no one uses any of the words so much more often than the others as to merit an index of the words as distinctive.

Instead, we have a new set of words that might qualify as "ideas" in the sense that they signify the unique contribution of each of the five speakers on the larger debate about the meanings of the concepts "freedom" and "slavery." Words such as "enjoyment" (for Buckingham), "wage" (for Buxton), "imperial" (for Peel), and "petition" (favored by Wellington) give a sense of the varying priorities of individuals.

It's worth making a list of the words that intrigue us, about which we might want to know more. It seems, for example, that each of the speakers cultivated their own set of examples and references – places whose anecdotes they used as examples for thinking through the consequences of continued slavery, the "apprenticeship" system of continued slavery, and immediate abolition. One route through the research might be to stop here and examine those terms and their references, looking up one at a time:

Buxton: Hottentots, Africans, Scotchman Stanley: Venezuela, mine Peel: Burke, Guadeloupe Wellington: Colombia

We may make many such hypotheses in the course of reading. It is always worth writing them down; we may always choose to return to them later. One valid use of text mining is exactly as a source for new hypotheses about the text and how they matter.

Before we get too excited about any of these words, however, we also need to think about the usefulness of word count and distinctiveness together. One of the tricky things about working with tf-idf is that it's a measure of distinctiveness, not a measure of significance. Which is to say that a highly-ranked word from any of these lists may be so distinctive because it was only spoken twice, and both times by one person. Does that make the word worthy of note? Not necessarily – it might just be irrelevant.

We can check the significance of our distinctiveness measure by another kind of representation of the same data, one that brings both distinctiveness and count into the same graphic.

```
# create a chart
most_distinctive_words_per_abolition_speaker %>%
  mutate(word = reorder_within(word, n, speaker)) %>% # rearrange the data by n and speaker
  ggplot(aes( # create a chart
    x = word,  # word on the x axis
    y = n, # n on the y axis
```

```
    fill = factor(tf_idf) )) + # the color of the bars in the bar chart will reflect tf_idf.
    # we must tell the computer to treat tf_idf as a "factor' because the the numbers are enumerated ra
  geom_col() + # create a bar chart
  facet_wrap(~speaker, scales = "free") + # create a series of mini-charts, breaking the data up by spe
  coord_flip() + # switch the information on x and y axes (this is a shortcut to make the 'word' labels
  scale_x_reordered() + # keep the 'word' axis ordered by n and speaker
  scale_fill_grey() + # use a gray scale (if this line is removed, a color chart will be produced)
  guides(fill = "none") + # don't show a legend for the colors of the bar chart (if removed, a color le
  scale_y_continuous(breaks = function(x)
    unique(floor(pretty(seq(min(x), (max(x) + 1) * 1))))) + # use whole numbers, not  decimals, for the
  labs(title = "Speakers\' Most Distinctive Words", # label the visualization with a title
       subtitle = "From the Debates in the 1830s on the Abolition of Slavery; pale means more distincti
       x = "Word", # x axis label
       y = "Count") # y axis label
```

Speakers' Most Distinctive Words

From the Debates in the 1830s on the Abolition of Slavery; pale means more dis
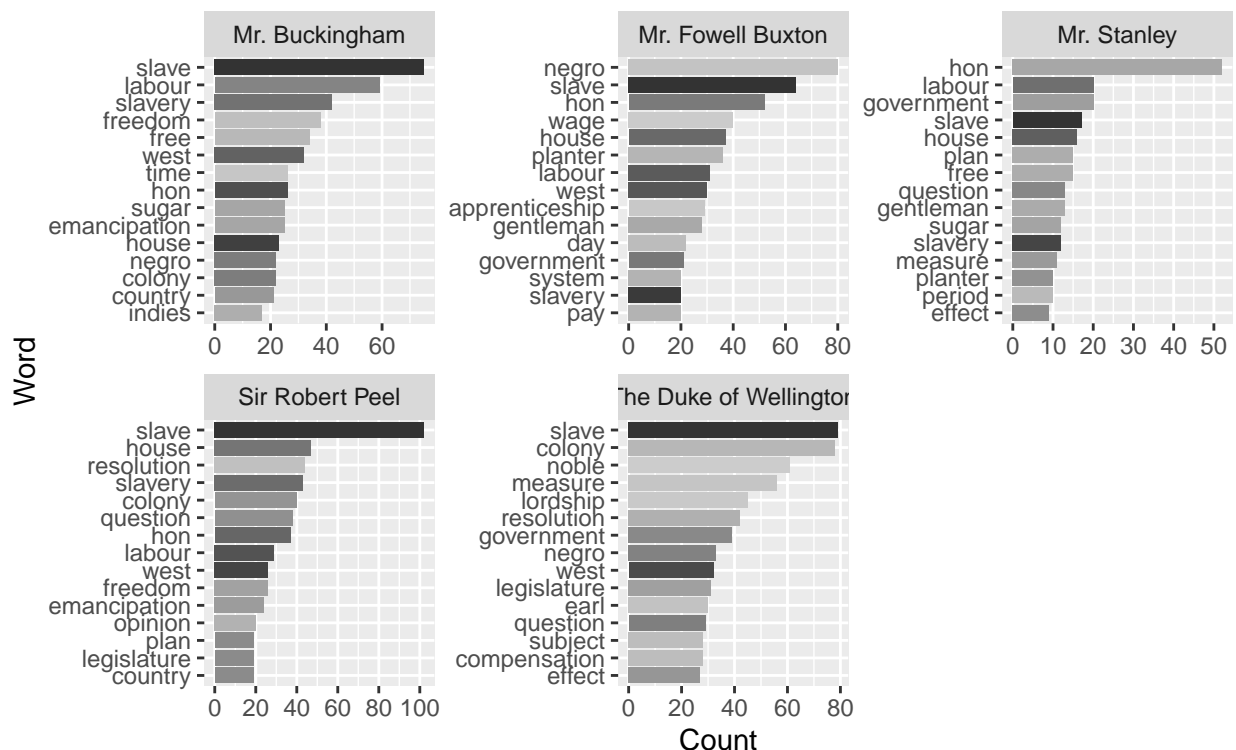
As opposed to the list, our bar chart shows both tf-idf distinctiveness (represented as paleness of the bar)
with count on the x axis. This visualization can guide us as to whether we want to investigate a particular
word because of its distinctiveness – or whether that word might be a poor marker of significance. For
instance, most of Stanley's distinctive words are revealed to have a wordcount of only 2, which is too few
to be significant; it simply doesn't make sense that "rum," "mine," and "barrel" merit investigation when

Stanley spoke them only twice in his voluminous speech-giving – that is, unless we already had a previous interest in the rum trade and its intersection with slavery. Statistically distinctive these words might be; worthy of further investigation they are not (at least not necessarily).

We can use wordcount with distinctiveness to identify a set of words that we do indeed wish to pursue further. We can see that the words "wage," "rear," "hottentots," and "flog" are, for Fowell Buxton, both distinctive and numerous, as are "ultimate," "qualify," "imperial," and "king's" For Robert Peel; nearly all of Stanley's words are distinctive, while Mr. Buckingham's most distinctive word, "thirdly," appears to be a mere rhetorical tick.

Nevertheless, there is room for caveats about settling for these words in particular. From the point of view of text mining, where we look to the number of words as significant, these results are unsatisfying. There are too many small results, where the count is smaller than 5. If we want to write an analysis of each speakers' referents, this might be a beginning, but we should not rest content that we have the best words for examinig the speakers yet. Let's rerun the code, this time adding instructions to only show us words that appear above a certain threshold. To do this, we will add just a single line of code. You can play with the number in the filter to see how the results change as we look for more and less distinctive words.

```r
most_distinctive_words_per_abolition_speaker2 <- tfidf_per_top_slavery_speaker_1833 %>%
  group_by(speaker) %>%
  filter(n > 5) %>% # <---------------- this is the only thing that's new
  arrange(desc(tf_idf)) %>%
  mutate(rank_tf_idf = row_number()) %>%
  slice(1:15) %>%
  ungroup() %>%
  mutate(word = reorder_within(word, n, speaker))

most_distinctive_words_per_abolition_speaker
```

```
## # A tibble: 75 x 5
##     index word         tf_idf     n speaker
##     <int> <chr>         <dbl> <int> <chr>
## 1       1 time         0.0859    26 Mr. Buckingham
## 2       2 freedom      0.0715    38 Mr. Buckingham
## 3       3 free         0.064     34 Mr. Buckingham
## 4       4 indies       0.0562    17 Mr. Buckingham
## 5       5 emancipation 0.047     25 Mr. Buckingham
## 6       6 sugar        0.047     25 Mr. Buckingham
## 7       7 country      0.0395    21 Mr. Buckingham
## 8       8 labour       0.027     59 Mr. Buckingham
## 9       9 colony       0.0231    22 Mr. Buckingham
## 10     10 negro        0.0231    22 Mr. Buckingham
## # i 65 more rows
```

```r
ggplot(data = most_distinctive_words_per_abolition_speaker2,
       aes(x = word, y = n, fill = factor(tf_idf) )) +
  geom_col() +
  scale_fill_grey() +
```

```
facet_wrap(~speaker, scales = "free") +
coord_flip() +
scale_x_reordered() +
guides(fill = "none") +
 scale_y_continuous(breaks = function(x)
   unique(floor(pretty(seq(min(x), (max(x) + 1) * 1))))) + # use whole numbers, not decimals, for the
labs(title = "Speakers\' Most Distinctive Words with N > 5",
     subtitle = "From the Debates in the 1830s on the Abolition of Slavery; pale means more distincti
     x = "Word",
     y = "Count")
```


Speakers' Most Distinctive Words with N > 5
From the Debates in the 1830s on the Abolition of Slavery; pale means more dis
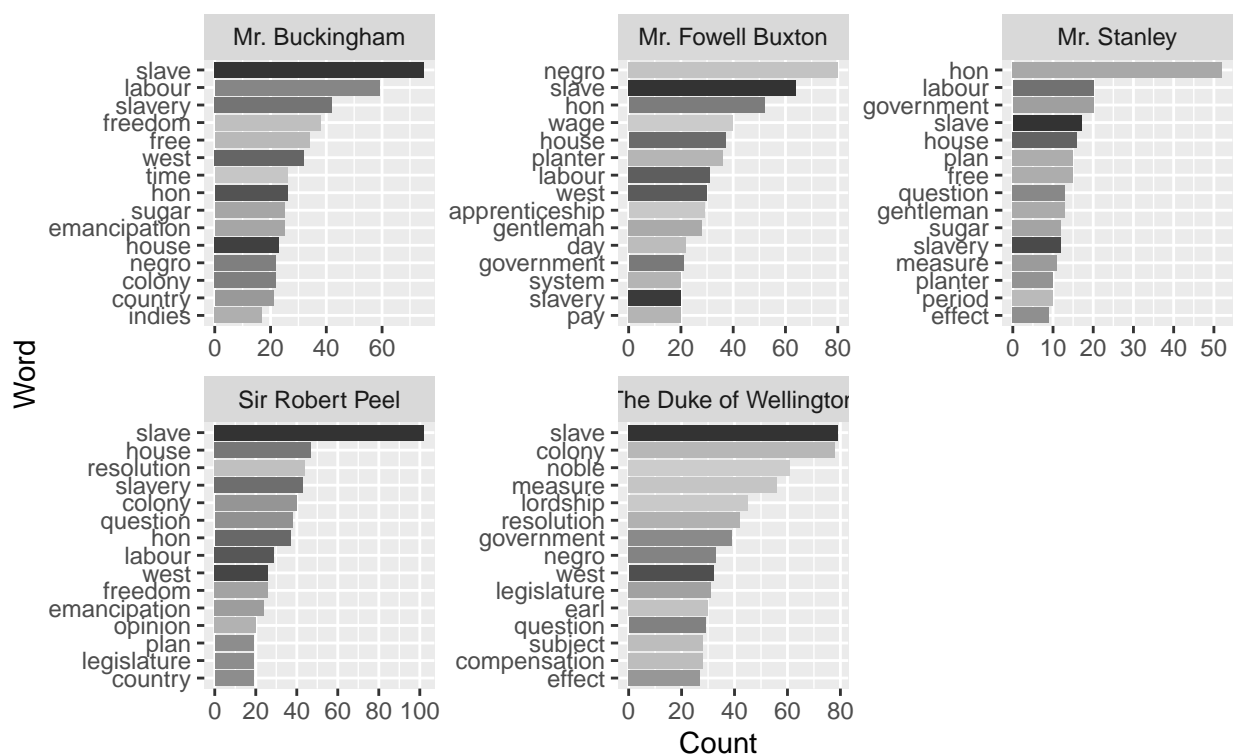
With this new view, it is less clear that Mr. Buckingham wanted for distinctive words. His use of "enjoyment," "wage," "reward," and "grow" are notable.

At this point, we may want to choose one speaker to inspect in greater detail. Our choice of which speaker to investigate might be motivated by distant reading or not. We are not compelled to investigate Edward Stanley because he spoke more than anyone else. We have used wordcount to guide me to five possible candidates. Nor are we forced to choose which candidate from these five to investigate further on the basis of which words they spoke or which words the computer finds distinctive. We can also make decisions based on outside reading, and good historians usually do.

Suppose we choose to investigate Mr. Buckingham. He is, after all, the outlier – a new arrival in parliament, elected by the evangelical middle class who obtained a presence in parliament only after 1832. We can use this information to investigate certain aspects of his speech. We might decide to grow curious about what he had to say about "enjoyment," "wage," "bonage," and "america," to take a smattering of his most distinctive words for this view.

What might we learn if we looked at Mr. Buckingham's words in greater detail? The visualizations above limit us to a small number of words per speaker; but we might learn a great deal more by looking at Mr. Buckingham's favorite words, whether by count of by tf-idf. In fact, computation is cheap, so let's do both.

```r
mr_buckingham_slavery_debates <- tfidf_per_top_slavery_speaker_1833 %>%
  filter(speaker == "Mr. Buckingham") %>%
  arrange(desc(n)) %>%
  slice(1:40)

mr_buckingham_slavery_debates
```

```
## # A tibble: 15 x 6
## # Groups:   speaker [1]
##    speaker        word             n     tf    idf tf_idf
##    <chr>          <chr>        <int>  <dbl>  <dbl>  <dbl>
##  1 Mr. Buckingham slave           75 0.154  0      0
##  2 Mr. Buckingham labour          59 0.121  0.223  0.0270
##  3 Mr. Buckingham slavery         42 0.0862 0.223  0.0192
##  4 Mr. Buckingham freedom         38 0.0780 0.916  0.0715
##  5 Mr. Buckingham free            34 0.0698 0.916  0.0640
##  6 Mr. Buckingham west            32 0.0657 0.223  0.0147
##  7 Mr. Buckingham time            26 0.0534 1.61   0.0859
##  8 Mr. Buckingham hon             26 0.0534 0.223  0.0119
##  9 Mr. Buckingham emancipation    25 0.0513 0.916  0.0470
## 10 Mr. Buckingham sugar           25 0.0513 0.916  0.0470
## 11 Mr. Buckingham house           23 0.0472 0.223  0.0105
## 12 Mr. Buckingham colony          22 0.0452 0.511  0.0231
## 13 Mr. Buckingham negro           22 0.0452 0.511  0.0231
## 14 Mr. Buckingham country         21 0.0431 0.916  0.0395
## 15 Mr. Buckingham indies          17 0.0349 1.61   0.0562
```

```r
ggplot(data = mr_buckingham_slavery_debates,
       aes(x = reorder(word, n), # the x axis will show words ordered by their count
           y = n),
           fill = factor(tf_idf)) + # the y axis will show n
  scale_fill_grey() +
  scale_y_continuous(breaks = function(x)
    unique(floor(pretty(seq(min(x), (max(x) + 1) * 1))))) + # use whole numbers, not decimals, for the
  geom_col() + # create a bar graph
  coord_flip() + # switches the x and y axis for a more elegant presentation
  labs(title = "Mr Buckingham\'s Most Frequenlty Spoken Words", # this is another way of labeling title
```

```
      subtitle = "From the 1833 Debates on the Abolition of Slavery",
      x = "Word",
      y = "Count")
```

Mr Buckingham's Most Frequenlty Spoken Words
From the 1833 Debates on the Abolition of Slavery
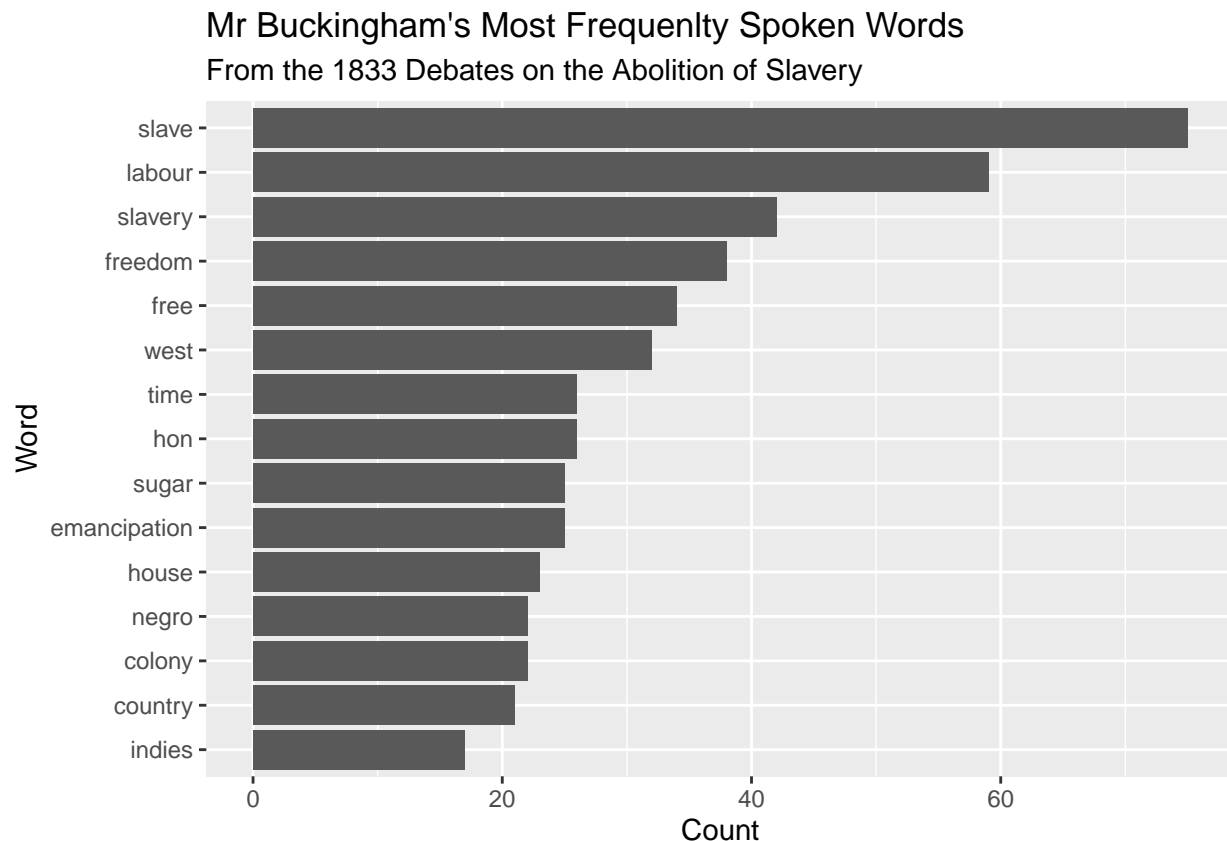
It is interesting that "labour" and "freedom" are used alongside "slavery" and "slave" as Buckingham"s favorite words. Also pronounced are his invocation of "rights," "wages," and "people," suggesting that Buckingham was interested in drawing equivalencies between the rights of working men in Britain and the rights of enslaved people in the West Indies.

But none of these words are particularly distinctive of Mr. Buckingham, at least in terms of their statistical profile. It's not clear that he used "labour" and "freedom" any more than other speakers. In fact, his top 50 words have a distinctiveness score of zero.

If we want to see the words that Buckingham used more than other speakers, we can return to the tf-idf scores that we generated a moment ago. In fact, all we have to do is change a single line of code to produce a visualization showing the top forty most distinctive words and their counts.

```
mr_buckingham_slavery_debates2 <- tfidf_per_top_slavery_speaker_1833 %>%
  filter(speaker == "Mr. Buckingham") %>%
  arrange(desc(tf_idf)) %>% # <------- the only line that changed
  slice(1:40)
```

```
ggplot(data = mr_buckingham_slavery_debates2,
       aes(x = reorder(word, n), # the x axis will show words ordered by their count
           y = n),
           fill = factor(tf_idf)) + # the y axis will show n
  scale_fill_grey() +
  scale_y_continuous(breaks = function(x)
    unique(floor(pretty(seq(min(x), (max(x) + 1) * 1))))) + # use whole numbers for the x axis
  geom_col() + # create a bar graph
  coord_flip() + # switches the x and y axis for a more elegant presentation
  labs(title = "Mr Buckingham\'s Most Distinctive Words",
       subtitle = "From the 1833 Debate on the Abolition of Slavery",
       x = "Word",
       y = "Count")
```


Mr Buckingham's Most Distinctive Words
From the 1833 Debate on the Abolition of Slavery
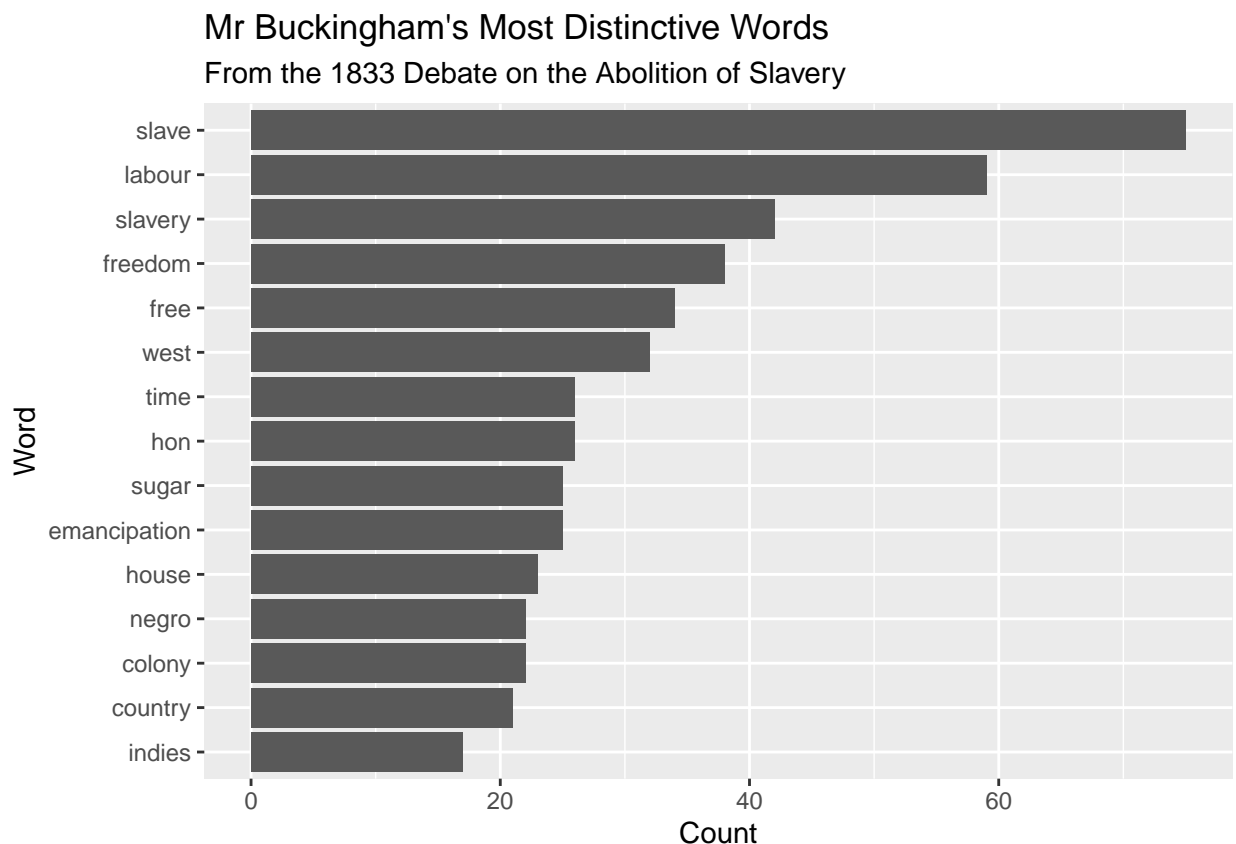
"Wages" emerges as a more distinctive word of Buckingham's that he said quite frequently – 12 times in all. It is not his most distinctive word; to find that, we can very slightly alter the code above to find the word he uses more than anyone else – but the result may be disappointing.

```r
mr_buckingham_most_distinctive_word <- tfidf_per_top_slavery_speaker_1833 %>%
  filter(speaker == "Mr. Buckingham") %>%
  arrange(desc(tf_idf)) %>%
  slice(1) # <------- the only line that changed

mr_buckingham_most_distinctive_word
```

```
## # A tibble: 1 x 6
## # Groups:   speaker [1]
##   speaker        word      n    tf   idf tf_idf
##   <chr>          <chr> <int> <dbl> <dbl>  <dbl>
## 1 Mr. Buckingham time     26 0.0534  1.61 0.0859
```

"Thirdly" is merely a rhetorical habit.

More interesting are substantive words that (perhaps) indicate Buckingham's values: wages enjoyment love universal independent devoted

– as well as the vices he (perhaps) abhors: servitude vindictiveness possessions ignorance grind[ing] destructive groundless cruel burthen

The "perhaps" is crucial. We are noting possible avenues for investigation for ourselves; this is not yet a proven list of findings from digital history. We should not be too quick to leap to conclusions about how Buckingham used these words without reading directly from his speeches.

```r
library(kableExtra)

# finding just the sentences where Buckingham uses the word "wages"
buckingham_sentences <- slavery_debates_1833 %>%
  filter(speaker == "Mr. Buckingham",
         str_detect(text, "wages"))

buckingham_sentences %>%
  head() %>%
  mutate(text = str_trunc(text, 120)) %>%   # optional: shorten long lines
  kable("latex", booktabs = TRUE) %>%
  kable_styling(full_width = FALSE, position = "left") %>%
  column_spec(2, width = "4cm")    # wrap the "text" column
```

sentence_id	speaker	suggested_speaker	text
S3V0018P0_6628	Mr. Buckingham	james_buckingham_3161	As long as slavery was continued, call it by w
S3V0018P0_8943	Mr. Buckingham	james_buckingham_3161	But all that was intended by the freedom of
S3V0018P0_8984	Mr. Buckingham	james_buckingham_3161	But, provided the labourers were free to choo
S3V0018P0_8987	Mr. Buckingham	james_buckingham_3161	There were many of these employed at Sheffi
S3V0018P0_8989	Mr. Buckingham	james_buckingham_3161	Let the sugar-cultivators of the colonies be a
S3V0018P0_9027	Mr. Buckingham	james_buckingham_3161	The negroes made each nearly two dollars a-d

Here they are as plaintext:

```
buckingham_sentences %>%
  head() %>%
  mutate(text = str_wrap(text, 80)) %>%
  select(text) %>%
  kable("latex", booktabs = TRUE, longtable = TRUE, escape = FALSE) %>%
  column_spec(1, width = "14cm")
```

text
As long as slavery was continued, call it by what name they might, whether apprenticeship or servitude, or by any other term, as long as forced subjection to an individual master, without power of removal, or of improving wages, remained, so long would there be danger of insurrection: for so long would there be powerful motives to rebel.
But all that was intended by the freedom of the negro, in the present instance, was to give him the liberty to take his labour, the only property he had, to the best market, to select his own employer, to negociate for his own wages, to earn his own bread, and to enjoy the fruits of his labour unmolested.
But, provided the labourers were free to choose whichever they preferred, it would happen that the timid and the indolent would choose the easiest and the pleasantest, and the number of applicants for this would reduce the wages to a low scale; while the boldest and the most reckless would choose the dangerous and the difficult, not for the danger and the difficulties, but for the high rewards attached to their performance; for the very destructiveness of their nature would narrow the circle of competitors, and the wages would accordingly be high.
There were many of these employed at Sheffield (the town he had the honor to represent), and he believed that the wages of this class of artizans was so high, as that with three or four days' labour in the week, as much might be earned as at any of the less destructive occupations in six.
Let the sugar-cultivators of the colonies be as free to choose their occupations as the steel-grinders of Sheffield, and there would be no just ground of complaint: high wages would follow dangerous and difficult employments, and low wages safe and easy trades.
The negroes made each nearly two dollars a-day; and paying one to their master, had still ample wages for themselves: but had they not been allowed to receive these wages, or had they been stimulated only by the whip, they would not have accomplished in a fortnight, what they here executed, in the best as well as most expeditious manner, in the short period of less than four days.

We see in these sentences that Buckingham is using "wages" in concert with the concepts of "freedom," "labour," and "property" to recapitulate a Lockean argument about how property depends on the work of men to enclose land, and how the duty fo the state is to uphold contracts that will protect exchanges built upon this system, upon which capitalism rests. A string of sentences follow about the specificities of wages. But the details are hard to interpret, in part because we are missing important context.

But before continuing further, we should observe that the sentences in our dataset are not necessarily from a single speech or even debate. If we look at their sentence_id numbers, we will see that the numbers are not contiguous. These sentences came in the train of other sentences which are missing in this dataset. To

interpret any individual sentence correctly, a data analyst must be aware of the importance of understanding that what we are seeing here is a series of discrete statements made on different days.

```
buckingham_sentences$sentence_id
```

```
## [1] "S3V0018P0_6628"  "S3V0018P0_8943"  "S3V0018P0_8984"  "S3V0018P0_8987"
## [5] "S3V0018P0_8989"  "S3V0018P0_9027"  "S3V0019P0_14756"
```

Sentences in Hansard are literally numbered by the order in which they appear in print. Some of the sentences numbers above occurred near to each other and may be from the same speech, but no sentences are actually back-to-back in their original context. There is information missing.

Missing information is the result of our strategy of identifying words like "slavery" and "wages" that appear within the same sentence. For the purpose of picking up on general patterns, this manner of counting and generalizing is useful. But if we abstract sentences from their context, it is almost impossible to tell what is going on.

Fortunately, we are not constrained by filtering for words. We can always go back to the original context of the speech or debate in which a sentence appears. We can use our list of `sentence_id` numbers to call up the entire speeches of which each sentence is a part (further down we will explain more about the technicalities of what happens in this section).

If we were working on a paper analyzing these speeches, we would likely want to read Buckingham's speeches about slavery in their entirety; the same would be true if our research project concerned the corporate reports of a company or a series of lawsuits. The process of reading and understanding cannot be replaced by counting words and analyzing them. Students of data science who pursue wordcount without background reading inevitably arrive at weak, useless or nonsensical results.

For the moment, let us proceed by reading a limited sample of Buckingham's speeches about slavery – just those speeches in which he invoked the word "wages."

The code for obtaining the speeches is given below; the full speech is printed in the Appendix.

```
# moving from a list of sentences to the full speeches for context
buck_speech_ids <- buckingham_sentences %>%
  select(sentence_id)

buck_speeches <- buck_speech_ids %>%
  left_join(hansard_1830, by = "sentence_id")

buck_speeches %>%
  head() %>%
  mutate(text = str_trunc(text, 120)) %>%   # optional: shorten long lines
  kable("latex",
        booktabs = TRUE,
        col.names=NULL,
        caption="Speeches by Buckingham about Slavery that invoke the word 'wages'") %>%
  kable_styling(full_width = FALSE, position = "left") %>%
  column_spec(2, width = "4cm")    # wrap the "text" column
```

Table 9: Speeches by Buckingham about Slavery that invoke the word 'wages'

S3V0018P0_6628	As long as slavery was continued, call it by what name they might, whether apprenticeship or servitude, or by any oth...
S3V0018P0_8943	But all that was intended by the freedom of the negro, in the present instance, was to give him the liberty to take h...
S3V0018P0_8984	But, provided the labourers were free to choose whichever they preferred, it would happen that the timid and the indo...
S3V0018P0_8987	There were many of these employed at Sheffield (the town he had the honor to represent), and he believed that the wag...
S3V0018P0_8989	Let the sugar-cultivators of the colonies be as free to choose their occupations as the steel-grinders of Sheffield, ...
S3V0018P0_9027	The negroes made each nearly two dollars a-day; and paying one to their master, had still ample wages for themselves:...

With missing sentences added back in, the speeches in their entirety are much easier to understand.

Mr. Buckingham began his long speech by introducing the three constituencies who demanded the abolition of slavery. Some are motivated by religion, others by claims about freedom, justice, and the British Constitution; and still others on moral or economic grounds, which Buckingham summarized as "policy alone."

Reviewing the call from religious circles for immediate abolition, and from policy and constitutional circles for delay, we see Buckingham invoking "wages" in the midst of a detailed case for immediate abolition. He challenges the proposals in the 1834 legislation to abolish slavery that called for a period of "apprenticeship," when enslaved people would be nominally freed but still forced to work without pay. The bill that finally passed would require apprenticeship, and historians have since argued that this unwillingness to end slavery represents a failure of the British state to commit to freedom. In Buckingham's arguments, we see the politician as a man ahead of his time – arguing that "apprenticeship" or "servitude" was an undisguised synonym for slavery, and denouncing any act that extended the period of enslavement.

We also see, in the second speech, we see Buckingham extraoplating from his experience of working with forced labor on naval vessels to the question of apprenticeship. We see him calling up another one of his more distinctive keywords – "America" – to make the case that the independent United States provided a historical example of the success of freeing an expensive colony to self-government.

Buckingham's commitments to endorse the humanity and rights of enslaved humans are abundantly clear. We see him making a detailed argument against the category of race itself as the basis for political or economic assumptions, especially assumptions about inferiority. He makes a case for dark-skinned people freely associating and participating in the economies of the Mediterranean, invoking ancient Egypt and the biblical Queen of Sheba too argue that dark-skinned Africans and their interracial descendents were destined for economic success.

We see Buckingham arguing that the free market naturally rewards laborers who choose dangerous work. Buckingham later generalizes about the rewards of a free market from his experience in Sheffield observing steel grinders and the high wages they commanded. Buckingham is arguing that only the freedom of wages – not apprenticeship – will induce formerly enslaved men to take on challenging work.

In this reading, the word "wages" is a category that allows Buckingham to navigate between his concerns and those of other policy-makers. Buckingham's invocation of wages suggest that he believes that capitalism has a utopian side in the freedoms it offers to workers. His anecdotes about Sheffield's steel workers and British naval vesslels suggest that he. . .

But we are not done. We can use the distinctive words in our list to investigate other aspects of the speech.

```r
buckingham_sentences_2 <- slavery_debates_1830 %>%
  left_join(speaker_metadata_1830) %>%
  filter(speaker == "Mr. Buckingham",
         str_detect(text, "enjoyment"))
```

```r
buckingham_sentences_2 %>%
  head() %>%
  mutate(text = str_wrap(text, 80)) %>%
  select(text) %>%
  kable("latex", booktabs = TRUE, longtable = TRUE, escape = FALSE) %>%
  column_spec(1, width = "14cm")
```

text
But when freedom had been granted—when the yoke had been taken off—when every man might seek his own employer, and fix his own terms of reward; when the blacks were elevated to the same enjoyment of equal rights with the whites, what was there to rebel for?
All men disliked to labour more than was necessary to obtain for them the enjoyments.of life; beyond this, they desired leisure, or at least the entire direction and control over the employment of their time.
And, as all slaves hitherto made free had bettered their condition from the moment of their freedom being attained, there was no good reason for doubting but that all the slaves in future to be emancipated, would run the same career of improvement, some faster and some slower than others, but all at least rising above that lowest point in the scale of existence, which now marks them the next link in creation to the beasts of the field, but which, being broken, they would rise, like other rational beings, to the enjoyment of all the privileges and all the virtues of manhood and humanity.
The negroes being released from their present degraded and depressed condition, would become subject to new motives, animated by new hopes, and cheered by new enjoyments. The developement of every new mental faculty would expand the desire for further intellectual attainment; and thus the now dormant powers of the negro mind would be brought out into progressively increasing exercise, till they became fitted for the highest enjoyment of all social and domestic pleasures.
He would begin, then, with the objections to immediate emancipation, which were raised by the noble Lord, the member for Stirling (Lord Dalmeny), which embraced the three following assertions:—That the slaves were too ignorant to be admitted at once to the enjoyment of freedom.

If the example of "wages" furnished sentences difficult to understand without broader context, the sentences about "enjoyment" give an example where Buckingham's case is clear without any further effort peering into the text. His utopian vision of capitalism becomes clear in these passages, which posit the conjecture that the freedom to work on a wage market gives labourers access to "enjoyment," and that enjoyment and its natural incentives therefore offers a reason why the slaves should be freed.

```
buckingham_sentences_3 <- slavery_debates_1830 %>%
  left_join(speaker_metadata_1830) %>%
  filter(speaker == "Mr. Buckingham",
         str_detect(text, "love"))
```

```
buckingham_sentences_3 %>%
  head() %>%
  mutate(text = str_wrap(text, 80)) %>%
  select(text) %>%
  kable("latex", booktabs = TRUE, longtable = TRUE, escape = FALSE) %>%
  column_spec(1, width = "14cm")
```

text
It could not be denied, that the love of ease was as common to the negro race as to every other.
If it were intended to confer at once upon the negro the enjoyment of political rights, the trust of the elective franchise, or elevation to judicial or other stations of civil or political authority, he confessed, that with ail his love of liberty, and advocacy of extended rights, he should pause before he assented to such a measure.
Baronet, that the love of The reference here was to the alliance of Solomon with one of the daughters of Pharaoh, whose language in the Canticles is expressive of her colour and her race.
Why, invariably this: that in both cases, the desire of gain was so much more powerful than the love of repose, that neither party slopped short in their exertions when their necessities were satisfied, but all pursued the same career of accumulation, adding as much as possible to that which they already pos- sessed; love of accumulation generally becoming more and more intense in proportion to the amount of the property possessed.

We might also elect to investigate "love," one of the words that appeared highly ranked through tf-idf but relatively sparse. We investigated the word because we felt that it might represent one of a set of virtues to which Buckingham looked – alongside "universal," "desire," and "religion."

This is a dead end. When we consult the text, we do not find an endorsement of "love" as a universal virtue that we hypothesized when we looked at the word list. Instead, we find the use of phrases such as "love of ease," "love of liberty," and "love of repose," as Buckingham invokes conjectures about the motivations of humankind, endorsing none of them. It would be foolhardy to make any kind of a positive argument about statements so diverse as these. We could try to unify them under one statement, but we would be falsifying evidence. In a research paper we will have no use for these examples, and we will pass over the word "love" without remark.

```r
buckingham_sentences_4 <- slavery_debates_1830 %>%
  left_join(speaker_metadata_1830) %>%
  filter(speaker == "Mr. Buckingham",
         str_detect(text, "universal"))
```

```r
buckingham_sentences_4 %>%
  head() %>%
  mutate(text = str_wrap(text, 80)) %>%
  select(text) %>%
  kable("latex", booktabs = TRUE, longtable = TRUE, escape = FALSE) %>%
  column_spec(1, width = "14cm")
```

text
The motives which had led to the almost universal demand throughout this country for the abolition of slavery were three-fold.

> The rule was universal, that men never rebelled because freedom was granted to them; and that the only danger of insurrection lay in a denial of rights which were justly due.
> It must be either, first, to satisfy the claims of abstract, right and justice, without referring to any other consideration; or, secondly, to meet the almost universal wishes of the people of England; or, thirdly, to effect the freedom and improvement of the slave population; or, fourthly, to benefit the West-India proprietors.
> Secondly, —As to its compliance with the almost universal wishes of the English people; scarcely anything could be more remote from it than this Bill.

We find in these sentences Buckingham invoking "universal" in the sense of an Enlightenment mandate, whether an expression of consensus or natural laws. Democratic demand for the abolition of slavery is an expression of a "universal" sentiment (sentences 1, 3 and 4). "Universal" laws govern behavior, including the fact that rebellion is more scarce in free nations (sentence 2). Universal truths, revealed through democratic processes, give the basis for legislation, according to Buckingham.

Presented with these particular sentences, having read some of Buckingham's arguments at greater length, we may feel confident that we understand how he has used the words. As we are making a wider argument about the use of words, buttressed by several examples, we do not necessarily need to follow every single sentence back to its original context.

```
buckingham_sentences_5 <- slavery_debates_1830 %>%
  left_join(speaker_metadata_1830) %>%
  filter(speaker == "Mr. Buckingham",
         str_detect(text, "desire"))
```

```
buckingham_sentences_5 %>%
  head() %>%
  mutate(text = str_wrap(text, 80)) %>%
  select(text) %>%
  kable("latex", booktabs = TRUE, longtable = TRUE, escape = FALSE) %>%
  column_spec(1, width = "14cm")
```

text
All men disliked to labour more than was necessary to obtain for them the enjoyments of life; beyond this, they desired leisure, or at least the entire direction and control over the employment of their time.
The developement of every new mental faculty would expand the desire for further intellectual attainment; and thus the now dormant powers of the negro mind would be brought out into progressively increasing exercise, till they became fitted for the highest enjoyment of all social and domestic pleasures.

> With increased intelligence, augmented wealth would be acquired; new desires would require new materials for satisfaction; the further developement of the resources of their own industry would furnish the means of payment or exchange; and the demand which would thus be created for British manufactures of every sort and kind, would be the most ample, as well as the most satisfactory, repayment of any temporary sacrifice which we might now be called upon to make, to carry this great measure of immediate emancipation into effect.
>
> If free labour were more productive than slave labour (as by the evidence read he had abundantly shown), it must then follow, that both masters and slaves would be enriched thereby—and the desire for increased enjoyments naturally following, we should no longer be engaged in sending out to the West Indies the miserable and scanty supply of salt herrings for the negroes' food, a few shirts and caps for their raiment, and an occasional addition to the implements and machinery of husbandry and manufacture for their use; but the freed men, having accumulated means of purchase and payment, increased supplies of necessaries, comforts, and luxuries of every kind, would be required; and all the arts and elegances of life would progressively be substituted for the rude materials now in use among them.
>
> But these were often preferred to more healthy and more pacific spots; because, in these the chances of gain, honour, promotion, and prize-money, were increased; and he (Mr. Buckingham) had himself heard drunk as a toast, in the gun-room mess of a man-of-war in the West Indies, ""a destructive war and a sickly season;"" the proposer justifying his wish, by the observation, that promotion was the desire of all, that this could not be quickened without vacancies, and the consolation of all was, that when these vacancies were occasioned by the two causes named, all parties were satisfied, as those who lived obtained speedy promotion, and those who died did not require any.
>
> Why, how many persons were there in England, who were able, without labour, to procure all they desired, from having other sources of income, which rendered labour on their parts unnecessary: and would it be borne that such persons should be forced to labour for their subsistence, when they could obtain all they needed without such occupation?

Buckingham invokes "desire" as the basis for understanding. Buckingham invokes universal desire in the spirit of Locke, Mandeville, and Smith as forces that incline men to leisure and pleasure and thus incentivize work (sentences 1, 2, 3, 6, 7). Buckingham suggests that former enslaved persons could eventually follow the American model, growing wealthy enough to become consumers who would buy finished products from British manufactures. A process of mutual enrichment would be the consequence (sentence 4).

The sole sentence we can't understand without broader context is Sentence #5. Let's inspect the sentence to obtain the sentence ID number.

```
buckingham_sentences_5[5]
```

```
##        sentence_id        speaker     suggested_speaker
##           <char>          <char>                <char>
## 1: S3V0018P0_8993 Mr. Buckingham james_buckingham_3161
##
##
## 1: But these were often preferred to more healthy and more pacific spots; because, in these the chan
```

```
##      year ambiguous fuzzy_matched ignored
##     <num>     <int>         <int>   <int>
## 1:  1833         0             0       0
```

The id of Sentence #5 is S3V0018P0_8993. Since we understand the gist of Buckingham's arguments about slavery, we might first try looking at the sentences immediately preceding this passage. The sentences are numbered chronologically, so all we have to do is to pull a few sentence numbers starting with S3V0018P0_8990.

```
missing_context <- hansard_1830 %>%
  filter(sentence_id %in% c("S3V0018P0_8990", "S3V0018P0_8991", "S3V0018P0_8992", "S3V0018P0_8993"))

missing_context %>%
  head() %>%
  mutate(text = str_wrap(text, 80)) %>%
  select(text) %>%
  kable("latex", booktabs = TRUE, longtable = TRUE, escape = FALSE) %>%
  column_spec(1, width = "14cm")
```

text
The source of discontent was, that the negroes were compelled to labour excessively, by coercion and terror of the whip, and were badly fed and badly clothed, though they laboured in crop-time eighteen hours out of the twenty-four: while the grinders of Sheffield were not compelled to labour, and, with eight hours' work per day throughout the whole week, could secure their being well fed, well clothed, and have some surplus left for enjoyment besides.
The army, it was well known, was, in war time especially, a service abundantly destructive of human life; the navy still more so, for, in addition to the risks of battle, the risks of shipwreck must be incurred.
Certain stations of service in both were also more dangerous than others, from climate and other causes.
But these were often preferred to more healthy and more pacific spots; because, in these the chances of gain, honour, promotion, and prize-money, were increased; and he (Mr. Buckingham) had himself heard drunk as a toast, in the gun-room mess of a man-of-war in the West Indies, ""a destructive war and a sickly season;"" the proposer justifying his wish, by the observation, that promotion was the desire of all, that this could not be quickened without vacancies, and the consolation of all was, that when these vacancies were occasioned by the two causes named, all parties were satisfied, as those who lived obtained speedy promotion, and those who died did not require any.

In Sentence #5, Buckingham is talking about how workers will pursue dangerous and difficult jobs when motivated by desire ("gain, honour, promotion, and prize-money.") For this reason, Buckingham concluded that "promotion was the desire of all," and that, if allowed to pursue them, the prospect of wages would incentivize formerly enslaved people as it had incentivized British navvvies to work hard in the pursuit of earnings.

Pulling Together an Analysis

In the foregoing exercises, we have investigated many dimensions of text use in the debates over slavery. We have come up with hypotheses and tested some (if not all) of them. We have read words in context, returning to the text for more sentences when we needed more information. The material that we now have is enough to make an argument about Buckingham's use of words to make an argument about capitalism that supported the immediate abolition of slavery.

We have applied our personal judgment in deciding when word lists are enough by themselves, when a few sentences can inform our understanding, and when we need to return to the speeches themselves for more context. Cultivating the judgment to know when to trust data, when to look for more context, and when to leave out an example entirely is one of the most important skills of the digital historian – and it is a skill almost never tught in departments of Comptuer Science or Data Science. Understanding which course to pursue is a matter of training, instinct, broad reading, and a willingness to engage all the tools. The interpreter who uses word count skilfully has no automatic answer for how to interpret a given speaker or given text; the top words are no sure guide, nor is tf-idf, nor the sentences linked to them. Only by knowing when to move from short sentences to longer speeches, reading the speeches in context, and examining the words thoughtfully is an adequate opinion of the text formed.

How much of the work should I show in this case? The answer may be surprising. None of them.

The process of testing hypothesess laid out here is considered part of the research. Most of the visualizations and discussion would not form part of a term paper, let alone a published paper. They are like the wood shavings that fill a workshop; they are evidence of the labor that has gone into making an ultimate artifact at which we aim, which is surprising knowledge about the past. At the end of the day, the shavings are swept from the workshop floor to be discarded or composted. They are not to be treasured or preserved. Just so with much of the intermediate work of reading, counting, and modeling that goes into the making of an insightful historical analysis. There is no shortcut through it. No language model gives us a substitute for thinking about words and their meaning.

The graphs we made heretofore give us context that would otherwise be invisible. If not for text mining, we would have little idea that Buckingham and Buxton both spoke about wages disproportionately to Stanley and Peel. We might select other terms through close reading that catch our eye, and we might write about how the speakers used these words to support an argument. In any case, that is the task ahead of us — to show how speakers used their words to make a political case. But without text mining, the basis for the selection of words would be hopelessly colored by our individual prejudices. We can make a fundamentally more objective case for which words we focus on when we approach the problem through text mining.

To make an argument about Buckingham's use of words, what we need is (1) a list of the words we chose to investigate and a statement of the role that text mining played in selecting them, and (2) a close reading of the words in context that proves that we read Buckingham's speeches.

For the purpose, we do not need to show the reader the output of our tf-idf or word count process. Instead, a list of words will suffice. In the prose, we would spend a sentence or two describing the process we went through of applying tf-idf with and without a threshold to inspect Buckingham's distinctive words when compared with four other top speakers in the debates. If we want the list formatted nicely, we might write some code like this:

```r
# concatenate a list of the words I'm interested in
my_words <- c("wage", "enjoyment", "love", "universal", "desire")
```

Table 15: Some of Mr. Buckingham's Most Distinctive Words in the Debates About Slavery

word	n	tf_idf

```r
# find the tfidf scores for my words by looking in previous data
my_words_w_tfidf <- tfidf_per_top_slavery_speaker_1833 %>%
  filter(speaker == "Mr. Buckingham", word %in% my_words) %>% # find just words in my list spoken by Mr
  ungroup() %>% # necessary because the previous dataset was grouped
  select(word, n, tf_idf) %>% # retain three columns
  mutate(tf_idf = round(tf_idf, 4)) # round tf-idf to four decimal places

# make a table
kable(
  my_words_w_tfidf,
  format = "latex",
  booktabs = TRUE,
  caption = "Some of Mr. Buckingham's Most Distinctive Words in the Debates About Slavery"
)
```

As to the substance of the argument, we have most of the interpretive work above in the sentences that move, word by word, through a synopsis of the arguments Buckingham made about slavery, through his use of the words "desire" and "universal" to sketch out the enlightenment system of perfection that he believed made immediate abolition both inevitable and desirable, and finally through his identification of "wages" as the crucial mechanism that made it possible for men's willingness to endure dangerous work to be rewarded and thus workers to be incentivized to learn and apply themselves – something slavery could not provide. We might cut and paste into a new series of paragraphs our piecemeal analyses above of words in context. We might reshape the resulting paragraph with topic and concluding sentences into a series of arguments supported by evidence. The resulting argument, based on the actual words of the primary text, will be – above all – highly persuasive, because it is composed solely of arguments and pieces of evidence, each argument corresponding to a series of pieces of evidence.

At the risk of repetition, let us review this key point: text mining is not meant to save us from reading. It is meant to help us read better. The positive result of text mining is a precise understanding of the words used to make particular arguments by historical actors, alongside a statistical understanding of which words mark out individuals from their peers. The best outcome of engaging in a text mining project should be judged not by the fanciness of the graph that we produce but by the seriousness with which the analyst engages original texts and their language.

To pursue a question seriously means being willing to walk into blind alleys in search of some historical surprise. The frustrating fact of our work so far is that it has not produced anything astonishing. That advocates of the abolition of slavery invoked what they believed to be the utopian aspects of capitalism as an argument has been noted by previous generations of historians. That Buckingham had a specific vocabulary with which to engage capitalism – enjoyment, wages, the universal, and desire being its keywords – may be a new finding, but it's hardly headline news.

The work we have performed thus far gives us the material for a 3-page paper that demonstrates competence but not insight.

To move forward towards insight, we would want to return to our hypotheses and perform more comparisons. Whether the goal is a five- or ten-page paper or a corporate report, we will need to demonstrate that our method can discover something fundamentally new.

Next Steps: From Competence to Insight

Fortunately, we have several hypotheses ready to hand. We could begin by testing Buckingham's use of the word "wages" against Buxton's use of the same term – which Buxton used an astonishing 40 times in the course of the slavery debates. Did they speak about wages in the same way? Did they make the same argument? But if our findings about wages challenge little that historians already know about utopian faith in capitalism at the beginning o the nineteenth century, this looks like an unpromising direction.

Alternatively, we might entertain our hypotheses about the other key speakers. We might analyze the distinctive words of the other speakers – Stanley, Peel and Wellington, whose speeches offer such a different perspective on the debates. For each subject being analyzed, we would use the same code, making minor adjustments to allow us to pull the sentences of each speaker and inspect them in greater detail.

One place to start is Peel's invocation of the word "assembly," a word that in this context refers to the local government of the West Indies, including Jamaica. These island assemblies were unique among the British colonies in having the kind of independence that the North American colonies desired but were denied. They had an independent legislature of their own, which had operated with relative independence since the seventeenth century. One question raised by the abolition of slavery was whether these assemblies would continue to legislate in the West Indies, and whether it might become possible for formerly enslaved people to have a vote or stand for representation in them – that is, whether it might be possible for persons of color to control their own political destiny.

Most of the speakers didn't use the word "assembly" or used it rarely. But for Peel, "assembly" was a distinctive word, used 11 times in 1833 alone.

```r
peel_sentences <- slavery_1830 %>% # start with the text of speeches
  left_join(speaker_metadata_1830) %>%  # merge with speaker information
  filter(speaker == "Sir Robert Peel",  # keep just the sentences by Peel
         str_detect(text, "Assembl|assembl")) # searching for both capitalized, uncapitalized, singular
```

```r
peel_sentences %>%
  head() %>%
  mutate(text = str_wrap(text, 80)) %>%
  select(text) %>%
  kable("latex", booktabs = TRUE, longtable = TRUE, escape = FALSE) %>%
  column_spec(1, width = "14cm")
```

text
But when the Ministers said that the Colonial Assemblies must carry the measures of emancipation into effect, he could not understand how the hon. and learned Gentleman could make a speech so full of vituperation, nor how such a speech could promote the object he had in view.

> He did not mean to vindicate the conduct of the Colonial Assemblies; he was not satisfied with the course they had pursued; but if it was true, that the time was not come when we could interfere, it was not prudent to indulge in such language, and to censure so heavily those who must continue the masters and legislators of the slaves.
>
> The noble Lord said in substance, ""I feel entire confidence that the document, which the House has not seen, but which is to be sent to the West Indies, and which the Colonial Assemblies are to be told they must adopt, under the penalty of having the Sugar Duties increased—
>
> Every Member must recollect the Resolutions of 1823; and he apprehended that the country stood at present in this position as to the colonies:—the present Ministry, since their accession to office, had framed certain Orders in Council, which they determined to enforce in the Crown colonies; and they had signified to those colonies which have independent legislative assemblies, that those Orders in Council must be accepted by them, without qualification, on the penalty of not being allowed to benefit by the reduction of the sugar duties about to be proposed.
>
> He was confident that they would not look to the achievement of any triumph over the West-Indian assemblies.
>
> Mr. Burke stated, that he had no confidence whatever in the Colonial Assemblies; he asserted the competence of Parliament to legislate on these subjects, and contended that the question of the abolition could only be decided by the Imperial Legislature.

Here again we must read the sentences in their original speeches to make sense of it.

```r
# moving from a list of sentences to the full speeches for context
library("dplyr")
library("kableExtra")

# Trim each source to just the fields you need
peel_base <- peel_sentences %>%
  select(sentence_id)

hansard_trim <- hansard_1830 %>%
  select(sentence_id, text)

debate_meta_trim <- debate_metadata_1830 %>%
  select(sentence_id, speechdate)

speaker_meta_trim <- speaker_metadata_1830 %>%
  select(sentence_id, speaker)

file_meta_trim <- file_metadata_1830 %>%
  select(sentence_id, speech_id)


peel_speeches <- peel_base %>%
  left_join(hansard_trim,  by = "sentence_id") %>%
  left_join(speaker_meta_trim,  by = "sentence_id") %>%
```

```
  left_join(file_meta_trim, by = "sentence_id") %>%
  left_join(debate_meta_trim, by = "sentence_id") %>%
  group_by(speech_id, speaker, speechdate) %>%
  summarise(speech = paste(text, collapse = " "), .groups = "drop") %>%
  mutate(speech = paste0(speechdate, ": ", toupper(speaker), " ", speech)) %>%
  select(speech)

kable(peel_speeches,
      format = "latex",
      booktabs = TRUE,
      col.names = "",
      caption = "Three Speeches by Buckingham about Slavery That Invoke the Word 'Wages'") %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

Table 17: Three Speeches by Buckingham about Slavery That Invoke the Word 'Wages'

1831-04-15: SIR ROBERT PEEL But when the Ministers said that the Colonial Assemblies must carry the measures of er
1832-05-24: SIR ROBERT PEEL Every Member must recollect the Resolutions of 1823; and he apprehended that the cou
1833-06-03: SIR ROBERT PEEL He was confident that they would not look to the achievement of any triumph over the V

In the three speeches above, we find Peel invoking the colonial assemblies of the West Indies in order to defend the principle of independent democracies being entitled to make their own laws, unhindered by empire – even when the laws contained principles so odious as that of the enslavement of human beings.

In the first speech, April 15, 1833, Peel noted his dissatisfaction with the Colonial Assemblies but urged moderation and warned against making hasty resolutions that could harm the interests of slaves, humanity, and planters.

In the second speech, May 24, 1833, Peel pointed out the danger of making hasty regulations from afar without understanding the local conditions in the West Indies, especially in light of recent insurrections in Jamaica. He warned against inciting slaves to resistance through "incautious language and false hopes," only to then suppress them with military force, which would be unjust and exacerbate the situation.

Cautioning against haste, Peel invoked the assemblies, suggesting that it was a desire to "triumph over" the planters in their assemblies that made advocates of abolition impatient. Against haste, he advised a slow and deliberate plan to ensure order – the plan for apprenticeships, which would continue slavery under another name for seven further years.

In the third speech, June 3, 1833, Peel again urged his listeners not to interfere in the affairs of colonial assemblies, "but to lay the foundation of future prosperity and tranquillity in those countries."

He raised questions about whether slaves were "fit for freedom." The crowning point in Peel's argument was a lesson about history. Peel shocked his listeners by alluding to the liberation of the slaves in the French colony of St. Domingo in 1794, which had been followed riots which Peel described as "atrocities," which resulted in the leveling of inhabitations and the enslavement of whites by black pirates. In his account, Peel argued that the violence was levied against white bodies by black ones, and that the same specter of black violence awaited Englishmen in the West Indies if they moved towards liberation too quickly.

The existence of slavery thus became the basis for arguing that the West Indies could have no possible future as an independent government:

> "When a state of slavery is that upon which we are to work, the very means which lead to liberty must partake of compulsion. The minds of men being crippled with that restraint, can do nothing for themselves; every thing must be done for them. The regulations can owe little to consent."

When we compare Buckingham's arguments about the "universal" with Peel's arguments about "assembly," what we see is a standoff between two versions of the enlightenment, both full of faith for hope in humanity and the power of reason, but which worked against each other in a dramatic standoff.

The first version – Buckingham's – defined freedom in economic terms. It found slavery abhorrent and declared that it must be immediately desired. Buckingham's contention that the laws of unfreedom, once righted, would turned former slaves into productive subjects was premised upon faith in the laws of the free market and the ordered workings of desire, which envisioned formerly enslaved persons becoming free and going on to pursue dangerous and rewarding work alongside the British working class.

The second version – Peel's – defined freedom as the ability of educated individuals to elect their own leaders. It defined the basis for democracy as sacroscanct. At the same time, it held up a standard for being fit to govern. Only educated, peaceful people could be allowed to elect their leaders. This bar to entry – the defining of the conditions under which liberal individuals would be fit to govern – was one of the major points that would delimit democracy for most of the nineteenth and twentieth-century experiments with democracy. Again and again, points of this kind would be invoked to artificially limit the freedoms of individuals (like the enslaved inhabitants of the West Indies). And the argument was composed of a slight-of-hand: on the one hand, the endorsement of democracy (and the colonial assembles) as sacrosanct places of law-making whose powers must not be challenged by an imperial authority. On the other hand, the fear of unruly individuals, driven by race-based terror, backed by anecdotes about black pirates. The upshot of this argument was the denial that formerly enslaved persons could be educated for granted independence, who looked to the coercive power of parliament alone for saftey.

Both Buckingham's and Peel's arguments about freedom were premised on rationality, humanity, and the idea that the duty of parliament was to promote peace around the world. But it was the latter – the argument that democracy could only succeed where coercion was applied to keep unruly forces at bay – that won the day.

Here, at last, an argument wider than mere personality, rhetoric, or affectation is starting to emerge. It is likely that if we pursued further looking into the top speakers and their favored words, we would be able to identify still further complcations and aspects of the enlightenment's encounter with slavery in 1833. The process would be the same as it was above: using code to move from words to sentences and sentences to their context, then reading, summarizing, and interpreting how words were shaped by speakers into arguments and positions that encapsulate the differences of world view at work in 1833 and the terrible consequences that flowed from the failures of that moment.

What makes this version of the argument "insightful". in comparison to the others? Put simply, an argument should reflect back on all possible levels of analysis – from the individual in their context to the global, political, and economic dimensions of the historical debate. We turn to historical analysis, as Quentin Skinner puts it, to act "as a kind of archaeologist, bringing buried intellectual treasure back to the surface, dusting it down and enabling us to consider what we think of it." (Skinner, Liberty Before Liberalism (1998), p. 112). In order to have an informed opinion about the positions of Buckingham and Peel around slavery – and therefore in order to assess the ways in which the liberties of enslaved persons were postponed again and

again – it is useful to have statistical insights into the keywords "universal" and "assemblies" upon which Buckingham and Peel staked their claims about freedom. We have not done the work when we arrive at lists of keywords, however. The keywords must be read and understood in context. The arguments must be summarized. Their implications must be considered. Above all, we must strive to grasp at an explanation that helps us to understand why the abolition of slavery was so politically difficult for the seemingly-well intentioned men who gathered in Westminster in 1833 – and what we can learn about reasoned claims about government from that date. If the keywords do not unlock some insights about that moment, they do us no good.

The claim of this book has been that the application of historical reasoning with text mining is useful not only for professional studies of history but also for the many applied areas where understanding the reaction to some crisis is useful – a theme imprtant in business management, journalism, and politics. The crisis we wish to interpret when text mining corporate reports or journalistic stories might not be the postponed abolition of slavery in British Empire. But it will still be useful for us to understand the significance of individual points of view – for instance, memos written by separate executives or originating in different wings of the government – and how they instrumentalize particular keywords as anchors for broader concepts, be those concepts "freedom" and "slavery" or "earnings" and "employment."

Did the Concept of "Freedom" Change After the Debates of 1833?

The reason why the top speakers in any debate matter is the hypothesis that their words may have changed the minds of many listeners. Spoken aloud before an audience of other members of the press, reprinted in the magazines and newspapers of the nation, the words spoken in parliament mattered.

We can ask this question two ways: first, by asking about the debates around slavery before and after 1833, and secondly by asking the same question about all of Hansard for the same dates.

```
slavery_debates_before_1833 <- slavery_1830 %>%
  filter(year < 1833)

freedom_collocates_before_1833 <- slavery_debates_before_1833 %>%
  mutate(when = "1830-32") %>%
  filter(str_detect(tolower(text), "labour")) %>%
  unnest_tokens(word, text) %>%
 # anti_join(stop_words) %>%
 # anti_join(custom_stop_words) %>%
 group_by(when, word) %>%
  reframe(n = n()) %>%
  arrange(desc(n))

top_freedom_collocates_before_1833 <- freedom_collocates_before_1833 %>%
  top_n(30)
```

```
## Selecting by n
```

```
top_freedom_collocates_before_1833
```

```
## # A tibble: 30 x 3
##    when    word       n
##    <chr>   <chr>  <int>
##  1 1830-32 the      225
##  2 1830-32 of       136
##  3 1830-32 to        88
##  4 1830-32 and       72
##  5 1830-32 that      56
##  6 1830-32 labour    50
##  7 1830-32 in        48
##  8 1830-32 was       41
##  9 1830-32 would     36
## 10 1830-32 be        33
## # i 20 more rows
```

```r
slavery_debates_after_1833 <- slavery_1830 %>%
  filter(year > 1833)

freedom_collocates_after_1833 <- slavery_debates_after_1833 %>%
  mutate(when = "1834-36") %>%
  filter(str_detect(tolower(text), "free")) %>%
  unnest_tokens(word, text) %>%
 # anti_join(stop_words) %>%
 # anti_join(custom_stop_words) %>%
 group_by(when, word) %>%
  reframe(n = n()) %>%
  arrange(desc(n))

top_freedom_collocates_after_1833 <- freedom_collocates_before_1833 %>%
  top_n(30)
```

```
## Selecting by n
```

```r
top_freedom_collocates_after_1833
```

```
## # A tibble: 30 x 3
##    when    word       n
##    <chr>   <chr>  <int>
##  1 1830-32 the      225
##  2 1830-32 of       136
##  3 1830-32 to        88
##  4 1830-32 and       72
##  5 1830-32 that      56
##  6 1830-32 labour    50
##  7 1830-32 in        48
##  8 1830-32 was       41
##  9 1830-32 would     36
```

```
## 10 1830-32 be        33
## # i 20 more rows
```

Let's use tf-idf to find the words that are most distinctive of each period – before and after 1833 – in the debates about slavery.

```r
freedom_collocates_before_1833 <- slavery_debates_before_1833 %>%
  mutate(when = "1830-32") %>%
  filter(str_detect(tolower(text), "free")) %>%
  unnest_tokens(word, text) %>%
 # anti_join(stop_words) %>%
 # anti_join(custom_stop_words) %>%
 group_by(when, word) %>%
  reframe(n = n()) %>%
  arrange(desc(n))

freedom_collocates_after_1833 <- slavery_debates_after_1833 %>%
  mutate(when = "1834-36") %>%
  filter(str_detect(tolower(text), "free")) %>%
  unnest_tokens(word, text) %>%
 # anti_join(stop_words) %>%
 # anti_join(custom_stop_words) %>%
 group_by(when, word) %>%
  reframe(n = n()) %>%
  arrange(desc(n))


tf_idf_freedom_collocates <- bind_rows(freedom_collocates_after_1833,
                                       freedom_collocates_before_1833) %>%
  bind_tf_idf(word, when, n) %>%
  arrange(-tf_idf) %>%
  select(when, word, n, tf_idf)

tf_idf_freedom_collocates_before <- tf_idf_freedom_collocates %>%
  filter(when == "1830-32") %>%
  select(-when) %>%
  top_n(15)
```

```
## Selecting by tf_idf
```

```r
tf_idf_freedom_collocates_after <- tf_idf_freedom_collocates %>%
  filter(when == "1834-36") %>%
  select(-when) %>%
  top_n(15)
```

```
## Selecting by tf_idf
```

```
kable(tf_idf_freedom_collocates_before, caption = "1830-32")
```

Table 18: 1830-32

word	n	tf_idf
civilization	7	0.0010459
colour	7	0.0010459
indies	7	0.0010459
let	7	0.0010459
domingo	6	0.0008965
numbers	6	0.0008965
whites	6	0.0008965
acquired	5	0.0007471
public	5	0.0007471
1	4	0.0005977
admitted	4	0.0005977
desire	4	0.0005977
extinction	4	0.0005977
freehold	4	0.0005977
sugar	4	0.0005977
us	4	0.0005977
use	4	0.0005977

```
kable(tf_idf_freedom_collocates_after, caption = "1834-36")
```

Table 19: 1834-36

word	n	tf_idf
august	10	0.0014961
become	9	0.0013465
1st	8	0.0011969
antigua	8	0.0011969
apprentice	8	0.0011969
apprentices	8	0.0011969
contract	8	0.0011969
apprenticeship	7	0.0010473
intermediate	7	0.0010473
breach	6	0.0008977
compact	6	0.0008977
magistrates	6	0.0008977
money	6	0.0008977
none	6	0.0008977
done	5	0.0007481
entire	5	0.0007481

word	n	tf_idf
faith	5	0.0007481

Let's look to see if freedom's meaning changed across everything.

```
hansard_before_1833 <- hansard_1830 %>%
  left_join(debate_metadata_1830, by = "sentence_id") %>%
  mutate(year = year(speechdate)) %>%
  filter(year < 1833) %>%
  bind_rows(hansard_1820)

hansard_after_1833 <- hansard_1830 %>%
  left_join(debate_metadata_1830, by = "sentence_id") %>%
  mutate(year = year(speechdate)) %>%
  filter(year > 1833)

freedom_collocates_before_1833 <- hansard_before_1833 %>%
  mutate(when = "before 1833") %>%
  filter(str_detect(tolower(text), "freedom")) %>%
  unnest_tokens(word, text) %>%
 # anti_join(stop_words) %>%
 # anti_join(custom_stop_words) %>%
 group_by(when, word) %>%
  reframe(n = n()) %>%
  arrange(desc(n))


freedom_collocates_after_1833 <- hansard_after_1833 %>%
  mutate(when = "after 1833") %>%
  filter(str_detect(tolower(text), "freedom")) %>%
  unnest_tokens(word, text) %>%
 # anti_join(stop_words) %>%
 # anti_join(custom_stop_words) %>%
 group_by(when, word) %>%
  reframe(n = n()) %>%
  arrange(desc(n))


tf_idf_freedom_collocates <- bind_rows(freedom_collocates_after_1833,
                                       freedom_collocates_before_1833) %>%
  bind_tf_idf(word, when, n) %>%
  arrange(-tf_idf) %>%
  select(when, word, n, tf_idf)

tf_idf_freedom_collocates_before <- tf_idf_freedom_collocates %>%
  filter(when == "before 1833") %>%
```

```
  select(-when) %>%
  #filter(n>5) %>%
  top_n(15, wt = tf_idf)

tf_idf_freedom_collocates_after <- tf_idf_freedom_collocates %>%
  filter(when == "after 1833") %>%
  select(-when) %>%
  #filter(n>10) %>%
  top_n(15, wt = tf_idf)
```

```
kable(tf_idf_freedom_collocates_before, caption = "1820-32")
```

Table 20: 1820-32

word	n	tf_idf
rye	20	0.0001642
resident	17	0.0001396
dodson	13	0.0001068
restrictive	12	0.0000985
sovereigns	12	0.0000985
jurats	11	0.0000903
naples	11	0.0000903
cinque	10	0.0000821
illustrious	10	0.0000821
respectfully	10	0.0000821
usage	10	0.0000821
allegiance	9	0.0000739
attorney	9	0.0000739
beer	9	0.0000739
continental	9	0.0000739
degrees	9	0.0000739
marquis	9	0.0000739

```
kable(tf_idf_freedom_collocates_after, caption = "1834-39")
```

Table 21: 1834-39

word	n	tf_idf
intermediate	13	0.0001482
carlow	12	0.0001368
unqualified	10	0.0001140
antigua	8	0.0000912
hume	8	0.0000912
carlos	7	0.0000798

47

word	n	tf_idf
raphael	7	0.0000798
recognised	7	0.0000798
sabbath	7	0.0000798
transaction	7	0.0000798
despicable	6	0.0000684
landlords	6	0.0000684
o'connell	6	0.0000684
secrecy	6	0.0000684
terminate	6	0.0000684