Chapter 3: Investigating Speakers and Change Over Time Using Grouped Data

In previous chapters, we investigated how breaking up strings of text into individual words or multi-word phrases can give us insight into a corpus. We did things like count the number of words within an entire data frame. But historians often want to profile not merely the collective use of language, but also the role of individual speakers and how their langauge changed over time. Or historians might want to count by other variables, like by the date or by the individual debate.

In this chapter, we will learn how to count by "grouping" the data, allowing us to take other fields, like speaker name, year, and debate title, into account. This chapter will use these groupings to gain insight into the disourse of speakers in Parliament. But in order to perform a critical analysis of speakers, we must first ask ourselves: why study the langauge of individual speakers in Parliament when 19th-century British Parliament was mostly a small group of elite white men and not people who reflected the British population?

One answer is that we can model the language of parliament to better understand the workings of institutions, rather than "great men" of history. We can better understand the dynamics of power that imposed itself on 19th-century British society. The language of Parliamentarians was not merely communication, but also an instrument of power through which these individuals imposed an imagined order upon Britian–an order that reflected the biases, fears, and desires of elite white men–many of whose fathers served in Parliament, and who themselves inhereted their place in Parliament through familial status.

Thus, when we analyze speakers we better understand which people were rewarded with power, and how this power was expressed through langauge. We get a glimps of how Parliament existed as a space where the power relations of a small, elite group were actualized as the legitimate order of British society–a power that extended outwards to the societies women and its poor.

Counting by Group to Find the Wordiest Speakers

Our first excersize is finding the most "wordy" Parliamentarians from 1830–that is, Parliamentarians who spoke the most words. Often, the Parliamentarians who spoke the most did so becuase fellow Parliamentarians trusted them to express the wants of their own social group.

To analyze speakers in Hansard, we will first need to import a new category of data from hansardr: "speaker_metadata."

```
library(hansardr)

data("speaker_metadata_1830")
```

```
head(speaker_metadata_1830)
```

```
##        sentence_id    speaker suggested_speaker ambiguous fuzzy_matched ignored
## 1 S3V0010P0_11508 Mr. Croker  john_croker_1539         0             0       0
## 2 S3V0010P0_11509 Mr. Croker  john_croker_1539         0             0       0
## 3 S3V0010P0_11510 Mr. Croker  john_croker_1539         0              0        0
## 4 S3V0010P0_11511 Mr. Croker  john_croker_1539         0              0        0
## 5 S3V0010P0_11512 Mr. Croker  john_croker_1539         0              0        0
## 6 S3V0010P0_11513 Mr. Croker  john_croker_1539         0              0        0
```

For now, we will just go over three relevant fields from `speaker_metadata`.

Each sentence from the Hansard corpus is assigned a unique ID, as represented by the "sentence_id" field. The speaker who stated a sentence is assigned the same ID. This allows us to join the data from speaker_metadata to other data from hansardr, such as the debate text.

The other two important fields we will address are the "speaker" and the "suggested_speaker" fields.

The "speaker" field contains the speaker name as it was originally transcribed within the Hansard corpus. The resulting field contains inconsistencies and a lack of standardization that can make analysis of speakers difficult. A single speaker may have been recorded by different permutations of his first, middle and last name(s) (for example, "William Gladstone" may be transcribed as "William E. Gladstone", "W. Gladstone", or "Mr. Gladstone," to name just a few). In other cases, a speaker may have been called by a rotating office title (like "Prime Minister"). It is also the case that different speakers are recorded by the same name. For instance, two people named Sir Robert Peel served in Parliament during the 1820s and both are evoked by the same spelling of the name. To make analysis of speaker names more challenging yet, during the digitization processes, optical character recognition (OCR) errors were introduced into the speaker names. Common OCR error include interpreting a lower case "L" as an "i," or the lower case letter "a" as an "o."

To address the hurdles around analysis of speakers, we added a "suggested_speaker" field. This field contains our suggestion for the true identity of the speaker. The unique speaker within the corpus are assigned standardized name and number combinations. While multiple speakers during the same period shared the same name (like how "Mr. W. Gladstone" could refer to William Ewart Gladstone or his son William Henry Gladstone) only William Gladstone is assigned number 3104.

In the following code we will make our data set smaller and easier to work with by just selecting the "speaker," "suggested_speaker," and "text" fields before tokenizing the data into individual words.

```r
library(tidytext)
library(tidyverse)
library(lubridate)

data("hansard_1830")

words_1830 <- hansard_1830 %>%
  left_join(speaker_metadata_1830) %>%
  select(speaker, suggested_speaker, text) %>%
  unnest_tokens(word, text)
```

```r
head(words_1830)
```

```
##                  speaker suggested_speaker   word
## 1 The Duke of Buccleugh walter_scott_6566   rose
## 2 The Duke of Buccleugh walter_scott_6566     my
## 3 The Duke of Buccleugh walter_scott_6566  lords
## 4 The Duke of Buccleugh walter_scott_6566     in
## 5 The Duke of Buccleugh walter_scott_6566 rising
## 6 The Duke of Buccleugh walter_scott_6566     to
```

With a dataframe containing fields for “speaker”, “suggested_speaker”, and “word,” we are now in a position to perform a new kind of analysis to see the number of words each speaker contributed to Parliament using two new functions: `group_by()` and `summarize()`. We will use `group_by()` to organize the data by group before using summarize for our count. By using `group_by()` we can count the number of times a word was stated by a speaker, or within a year, and so forth. `group_by()` can be passed multiple arguments that refer to fields in a data set. To group by speaker will pass “speaker” to the `group_by()` function. Another useful function along with `group_by()` is `summarize()`. The `summarize()` function can be used with any statistical transformation, for instance `mean()` or `max()`. Here, we will use `summarize()` with a function to count, which is `n()`.

In the following code the argument `words_spoken = n()` tells `summarize()` to create a new column with the number of words that reflect each unique speaker

```
words_per_speaker_1830 <- words_1830 %>%
  group_by(speaker) %>%
  summarize(words_spoken = n()) %>%
  arrange(desc(words_spoken))
```

```
## # A tibble: 6 x 2
##   speaker                         words_spoken
##   <chr>                                  <int>
## 1 Mr. Hume                              856835
## 2 Mr. O'Connell                         823600
## 3 Sir Robert Peel                       822825
## 4 Lord John Russell                     782439
## 5 The Chancellor of the Exchequer       620830
## 6 Lord Brougham                         596096
```

Most students of British history will be familiar with the names of the speakers on this list, which include two prime ministers as well as one noted Irish orator and nationalist, Daniel O’Connell. The careful analyst will be a little concerned about the speaker as “The Chancellor of the Exchequer,” however, which refers not to an individual but a title that was handed off to multiple individuals over the century.

The Chancellor of the Exchequer is the head financial officer of the United Kingdom. He oversees the work of the Treasury. In many cases, the Chancellor of the Exchequer was a preliminary step to becoming Prime Minister. We would expect most of the Chancellor of the Exchequer’s speeches in parliament to reflect a predominant concern with taxation and spending. However, it is also possible that different individuals who held this post prioritized different causes. We can test our hypothesis on the 1830s by comparing the top words spoken by different individuals who occupied this post.

In the code below we create a custom stop words list to eliminate the most frequent words said by all members of parliament. We have already addressed the importance of removing stop words as a means to see words that are more meaningful for an analysis. In previous chapters we used an existing stop words list from TidyText. The hansardr library also provides its own stop words list. It may be the case, however, that an analyst wants to curate their own stop words list that caters to their specific data set or research question. In the following code we create a stop words list by assigning a list of words to a tibble (a tidyverse-style data frame).

```
custom_stop_words = tibble(word = c("hon", "speaker", "house", "question", "lord", "bill", "committee",
"proposition", "persons", "principles","service", "found", "propositions", "office", "matter", "statemen
"paid", "increase", "moved", "means", "considerable", "supply", "intention", "debt", "received", "expens
"sale", "times", "amounted", "deficiency", "laws", "til", "agreed", "alluded", "mentioned", "meant", "ho
"sense", "arrangement", "address", "sufficient", "local", "clergy", "relief", "necessity", "application
```

We can now filter and clean our data and explore the langauge of speakers called Chancellor of the Exchequer. In the following code we first remove instances of John Spencer, who spoke so few words as Chancellor of the Exchequer that he contributes little to our project. Then we eliminate numbers, stop words, and custom stop words before counting each speaker's words with `group_by()` and `summarize()`.

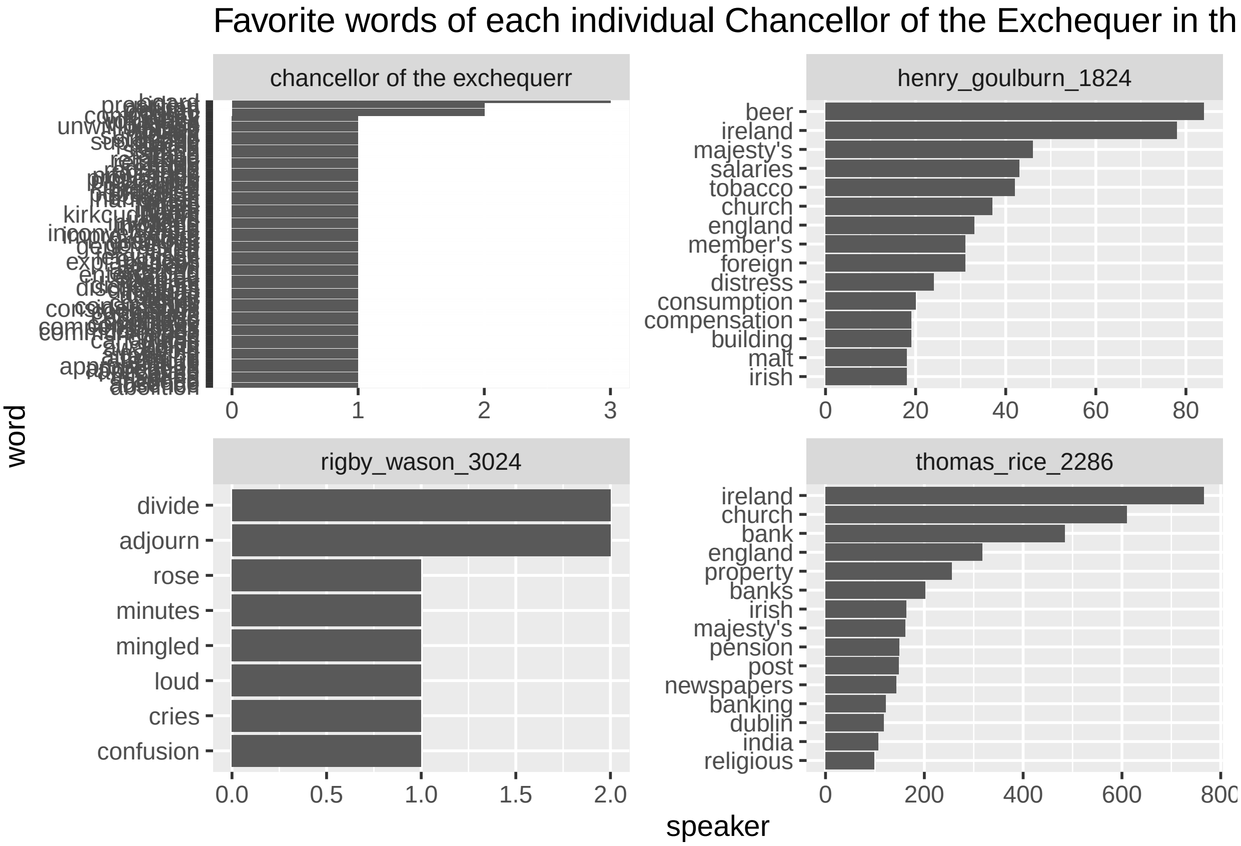
```
chancellor_of_the_exchequeur <- words_1830 %>%
  filter(str_detect(speaker, "Exchequer"),
         str_detect(word, "[a-z]"),
         suggested_speaker != "john_spencer_1234",
         suggested_speaker != "") %>%
   anti_join(stop_words) %>%
   anti_join(custom_stop_words) %>%
   group_by(suggested_speaker, word) %>%
   summarize(n = n()) %>%
   top_n(15) %>%
   ungroup()
```

```
head(chancellor_of_the_exchequeur)
```

```
## # A tibble: 6 x 3
##   suggested_speaker             word          n
##   <chr>                         <chr>     <int>
## 1 chancellor of the exchequerr abolition     1
## 2 chancellor of the exchequerr absence       1
## 3 chancellor of the exchequerr accede        1
## 4 chancellor of the exchequerr added         1
## 5 chancellor of the exchequerr affecting     1
## 6 chancellor of the exchequerr alien         1
```

```
chancellor_of_the_exchequeur <- chancellor_of_the_exchequeur %>%
  mutate(word = reorder_within(word, n, suggested_speaker))

ggplot(data = chancellor_of_the_exchequeur,
       aes(x = word, y = n)) +
  geom_col() +
  scale_x_reordered() +
  coord_flip() +
  facet_wrap(~suggested_speaker, scales = "free") +
  labs(x = "word", y = "speaker") +
  ggtitle("Favorite words of each individual Chancellor of the Exchequer in the 1830s")
```



The results show a division of interests. Goulburn is interested in talking about taxation on beer, tobacco, malt, and consumption; Peel is interested in addressing questions about how taxes relate to the church and its tithes, as well as political issues about how dissenters, Protestants and Catholics. Peel also thinks about commodities, but his commodities are different than Goulburn's; they include malt, land, and barley. Rice takes the issues in a different direction, reflecting on Dublin, India, banks, newspapers, and pensions. He shares with Peel a concern about property or land taxes.

How new is this information to readers of British history? We have long known that issues of what was taxed and how are a basic matter for politics of class, empire and nation; everyone wants someone else to pay for the government.

The research modeled in this textbook has a pedagogical function, and for that reason, we often present in these chapters examples that could be improved, and the above chart is one of those. Analysts should note several problems that might complicate our analysis and that call for further work. One of these is that a frequent word does not necessarily equate to importance: the most frequently-mentioned terms are typically the terms most subject to debate. Robert Peel, for instance, does not name Ireland's church tithes because he was interested in protecting the church in Ireland; to the contrary, he wished to abolish the system of tithes. Nor does the word list give us material for interpreting the rest of Peel's politics; the word list does nothing to tell the uninformed analyst about the movement for the abolition of taxes on grain, which forms a background to interpreting Peel's mentions of barley. The word list tells us little about the presence in 1830s Ireland of two churches, Roman Catholic and the Church of England. Without further reading, we might never known that Goulburn was famous for arguing for reductions on taxation. The analyst who comes to this data without a background in British history would do well to read in secondary sources before attempting to interpret the graph.

After some basic background research, however, the word lists can help produce insight, cluing the reader into differences such as Goulburn's relative willingness to talk about "distress," Peel's interest in "education," or Rice's interest in the post office. With word lists of this sort, we retrieve important clues about the salient differences contributed to political debates by individual ministers. A researcher interested in the careers of Chancellors of the Exchequer might start here before reading more deeply.

Another issue is the custom stop words list. For some researchers, the list in `custom_stop_words` may unnecessarily eliminate words whose usage they might want to track. Some researchers will want to identify words for rhetorical statements such as "hoped" or "feeling"; others may be intrigued about how different speakers refer to "information" or "knowledge." In general, I have opted for a deep list which includes every general allusion to governance, institutions, discovery, and communication. The beauty of custom lists is that each researcher can easily adjust the words to match their interests Analysts should understand how carefully they must tailor a custom stop words list to their project, and how much of a difference additions and subtractions make. Using a custom stop word list is very much an artisinal skill of research; it involves much the same skills of judgment, awareness, curiosity, and sensitivity to background issues from theory and secondary sources as does careful reading of primary sources.

One complication of lists is that to produce functional results, a custom stop words list must often be long; the hundreds of performative, rhetorical, and governmental words listed below is not a perfect list, and it would not work for every exercise. In later chapters, we will investigate approaches such as differential measurement which allow the reader to skip over custom stop words lists to obtain information about what distinguishes one speaker from another or one year from the next. No process is perfect, however, and analysts must often resort to some kind of custom filter to get closer to information that is useful for their project. There is no silver bullet for creating an absolutely accurate tool where text-mining research always produces useful information; rather, iterative inquiry, paired with background reading and in-depth reading of primary sources, is the basis upon which all insight is ultimately made.

Another issue has to do with the quality fo the data in the "speaker" field, which corresponds exactly to what is on the page in the nineteenth-century printed records of parliament. There is no guarantee that this method will produce a perfect analysis; we're going by who was lableled Chancellor of the Exchequer in Hansard, and there are no guarantees that the printed books label all speakers by their title. This is because the "speaker" field in the hansardr data set lists the names of the speakers as they are described in the published version of the debates. Wherever possible, we have annotated this information about speaker with static id's for most individual speakers, using the Parliamentary Identifiers used in parliament to this day. The field "suggested_speaker" uses these specific id numbers and gives a more accurate picture of the top speakers.

With more research, we could use the disambiguated speaker and date fields to revise the output above, creating a more accurate analysis that gives all the speeches by the individuals in question while they held the post of Chancellor of the Exchequer. Most of the time, we will want to use the suggested_speaker field to find and track speakers.

Notice that the lines of code below differ from the lines of code above only in that the argument of the function `group_by()` has been changed from "speaker" to "suggested_speaker:"

```
words_per_speaker_1830 <- words_1830 %>%
  filter(suggested_speaker != "") %>%
  group_by(suggested_speaker) %>%
  summarize(words_spoken = n()) %>%
  arrange(desc(words_spoken))
```

```
## # A tibble: 6 x 2
##   suggested_speaker    words_spoken
##   <chr>                       <int>
## 1 robert_peel_1664          1224576
## 2 henry_brougham_1679       1180672
## 3 joseph_hume_1712           875951
## 4 daniel_oconnell_2552       808577
## 5 john_russell_1885          802418
## 6 thomas_rice_2286           647551
```

In this view, we see the first and last names of individual speakers as well as their ID number from parliament (that is, Robert Peel is id 1664; 1664 is not, in this case, a date).

Working With Dates

Knowing who spoke the most in parliament is useful. But very often insight comes not from examining how counts change over time. What if we want to know not merely who spoke the most, but who was the top speaker for each year in parliament? Our `hansard_1830` data frame lists a date in the "speechdate" column, formatted as a four-digit year, two-digit month, and two digit day. To easily track speeches by year, we can add a new "field" – that is, a column – listing just the year of each speech.

Very frequently, when working with information about dates, we need to extract the month, day, or year from a data set. The lubridate package makes it easy to extract this information. We will use the `year()` function from lubridate with `mutate()` to create a new column that has just the extracted year from the `speechdate` field.

```
data("debate_metadata_1830")

words_1830 <- hansard_1830 %>%
  left_join(speaker_metadata_1830) %>%
  select(speaker, suggested_speaker, text) %>%
  unnest_tokens(word, text)

hansard_1830 <- hansard_1830 %>%
  left_join(debate_metadata_1830) %>%
  mutate(year = year(speechdate))

words_1830 <- hansard_1830 %>%
  left_join(speaker_metadata_1830) %>%
  select(speaker, suggested_speaker, text, year) %>%
  unnest_tokens(word, text)
```

```
##                  speaker suggested_speaker year   word
## 1 The Duke of Buccleugh walter_scott_6566 1830   rose
## 2 The Duke of Buccleugh walter_scott_6566 1830     my
## 3 The Duke of Buccleugh walter_scott_6566 1830  lords
## 4 The Duke of Buccleugh walter_scott_6566 1830     in
## 5 The Duke of Buccleugh walter_scott_6566 1830 rising
## 6 The Duke of Buccleugh walter_scott_6566 1830     to
```

Notice that the resulting dataset has a new field, “year.”

Now that we have a field for the year of each speech, we can return to “grouping” data to execute a faceted count where we count words by speaker and year. When we used the command `group_by()` above, we used it with one argument. But `group_by()` can take multiple arguments, allowing the analyst to perform grouped counts that involve multiple dimensions of the dataset. Next, we will use `group_by()` on two data fields at the same time – the “speaker” and “year” columns of as the arguments of `group_by()`, as in `group_by(speaker, year)`.

```
words_per_speaker_1830 <- words_1830 %>%
  filter(suggested_speaker != "") %>%
  group_by(speaker, year) %>%
  summarize(words_spoken = n()) %>%
  arrange(desc(words_spoken))
```

```
## # A tibble: 6 x 3
## # Groups:   speaker [4]
##   speaker              year words_spoken
##   <chr>               <dbl>        <int>
## 1 Lord Brougham        1839       200436
## 2 Lord Althorp         1831       180573
## 3 Lord Brougham        1838       175070
## 4 The Lord Chancellor  1831       161729
## 5 Mr. O'Connell        1834       158617
## 6 Lord Althorp         1833       154948
```

Notice that the result of grouping and counting is a data set with three columns: “speaker,” “year,” and “words spoken.” The last column is the count for each speaker per year – a field that makes it possible to analyze change over time.

When we use the `group_by()` command, the fields we use for grouping are preserved in the output. Two columns are the names of the facets we used for grouping – “speaker” and “year” – while the last column, “words_spoken,” is the count of how many words are recorded for each speaker per year.

Top Speakers in Debates About Slavery

What if we want to find 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 as before.

If you are new to code, take a moment to notice how changing only *one line* in the code below produces a totally different angle of analysis. The code below is otherwise exactly the same as the code in the sections above. 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.

```
library(ggrepel)

slavery_1830 <- hansard_1830 %>%
```

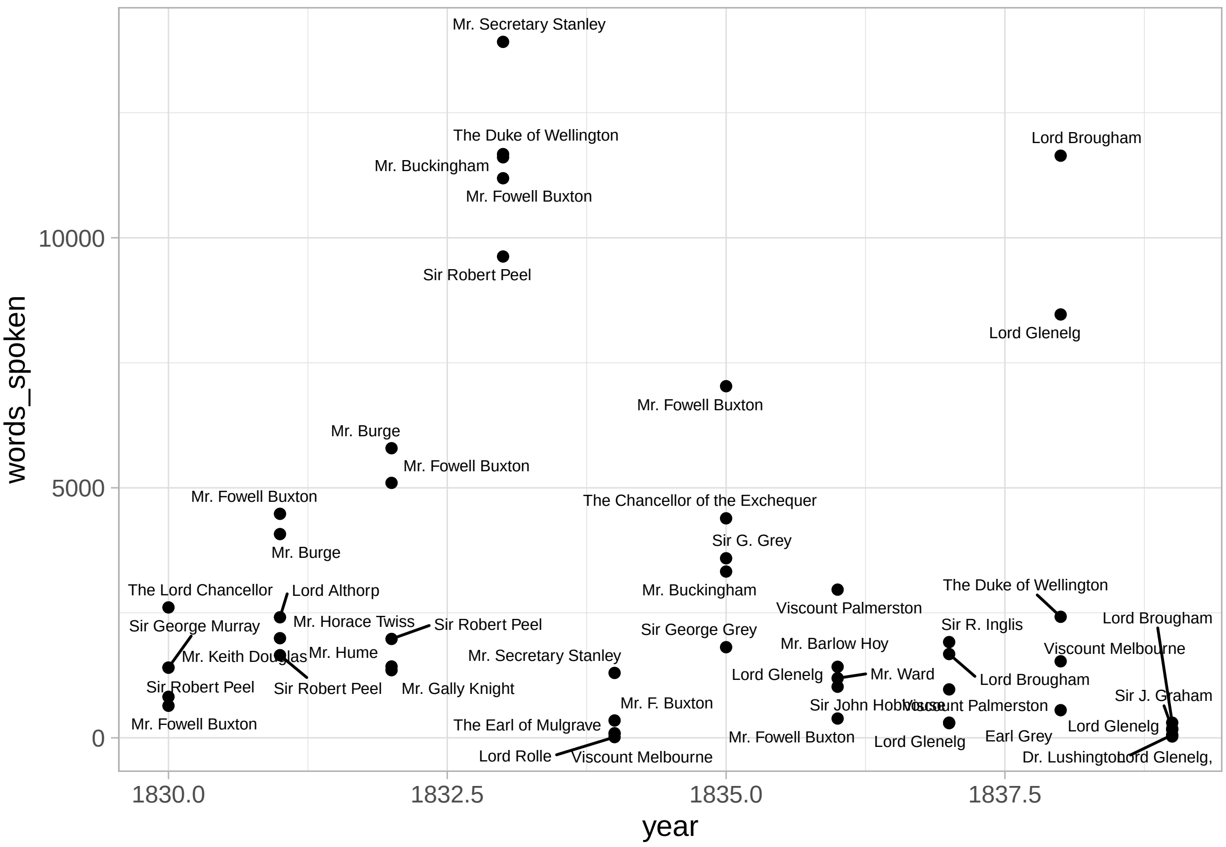
```r
  filter(str_detect(tolower(debate), "slavery")) %>%
  left_join(debate_metadata_1830) %>%
  mutate(year = year(speechdate))

slavery_words_1830 <- slavery_1830 %>%
  left_join(speaker_metadata_1830) %>%
  select(speaker, suggested_speaker, text, year) %>%
  unnest_tokens(word, text)

words_per_speaker_slavery_1830 <- slavery_words_1830 %>%
  filter(suggested_speaker != "") %>%
  group_by(speaker, year) %>%
  summarize(words_spoken = n()) %>%
  arrange(desc(words_spoken))

top_slavery_speakers <- words_per_speaker_slavery_1830 %>%
  group_by(year) %>%
  arrange(desc(words_spoken)) %>%
  slice(1:5)

ggplot(top_slavery_speakers,
       aes(x = year,
           y = words_spoken,
           label = speaker)) +
  geom_text_repel(size = 2) +
  geom_point() +
  theme_light()
```



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? For one thing, it dramatizes the fact of the debates around 1833 – when Britain voted to abolish the slave trade in British empire. This event was once lauded as proof of Britain's liberalizing mission around the globe, but more recently, historians of British empire have emphasized that the theoretical abolition of slavery in British empire did not result in the abolition of slavery in mahy parts of India and Africa under British control, nor did it minimize the punitive system of indentured servitude, by which many subjects of British empire were sent into conditions of forced labor abroad. Were we interested in considering the contribution of the major speakers to this debate, we would have their names: Mr. Secretary Stanley, the Duke of Wellington, Mr. Buckingham, Mr. Fowell Burton, and the Earl of Ripon. 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. In the next section, we will explore the contributions of these actors.

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: 20 x 2
##    speaker                   n
##    <chr>                 <int>
##  1 Mr. Buckingham        11610
##  2 Mr. Secretary Stanley 11368
##  3 The Duke of Wellington 10691
##  4 Sir Robert Peel        9626
##  5 The Earl of Ripon      9523
##  6 Mr. Fowell Buxton      9150
##  7 Mr. Godson             8734
##  8 Mr. O'Connell          6946
##  9 Mr. Hume               5843
## 10 The Lord Chancellor    5795
## 11 Dr. Lushington         5415
## 12 Earl Grey              4805
## 13 Lord Suffield          4467
## 14 Mr. Stanley            4319
## 15 Lord Ellenborough      4002
## 16 Lord Sandon            3846
## 17 Lord Althorp           3833
## 18 Mr. Macaulay           3252
## 19 Lord Howick            2995
## 20 Mr. Baring             2971
```

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.

```
pattern1 = c("Mr. Buckingham", "Mr. Stanley", "The Duke of Wellington", "Sir Robert Peel")
```

```
abolition_speakers <- slavery_debates %>%
  left_join(speaker_metadata_1830) %>%
  filter(speaker %in% pattern1)
```

```
head(abolition_speakers)
```

```
##     sentence_id
## 1 S3V0017P0_725
## 2 S3V0017P0_726
## 3 S3V0017P0_727
## 4 S3V0017P0_728
## 5 S3V0017P0_729
## 6 S3V0017P0_730
##
## 1                                         presented a Petition from Magistrates, Bankers, Merchants,
## 2
## 3 The latter was signed by 2, 468 persons, and was well entitled to attention, no less on account of
## 4
## 5
## 6                                                                   The petitioners referred to the 
##   speechdate                              debate year                speaker
## 1 1833-05-02 ABOLITION OF SLAVERY. -PETITIONS. ] 1833 The Duke of Wellington
## 2 1833-05-02 ABOLITION OF SLAVERY. -PETITIONS. ] 1833 The Duke of Wellington
## 3 1833-05-02 ABOLITION OF SLAVERY. -PETITIONS. ] 1833 The Duke of Wellington
## 4 1833-05-02 ABOLITION OF SLAVERY. -PETITIONS. ] 1833 The Duke of Wellington
## 5 1833-05-02 ABOLITION OF SLAVERY. -PETITIONS. ] 1833 The Duke of Wellington
## 6 1833-05-02 ABOLITION OF SLAVERY. -PETITIONS. ] 1833 The Duke of Wellington
##       suggested_speaker ambiguous fuzzy_matched ignored
## 1 arthur_wellesley_1299         0             0       0
## 2 arthur_wellesley_1299         0             0       0
## 3 arthur_wellesley_1299         0             0       0
## 4 arthur_wellesley_1299         0             0       0
## 5 arthur_wellesley_1299         0             0       0
## 6 arthur_wellesley_1299         0             0       0
```

```
abolition_speakers <- abolition_speakers %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  group_by(speaker, word) %>%
  summarize(n = n())
```

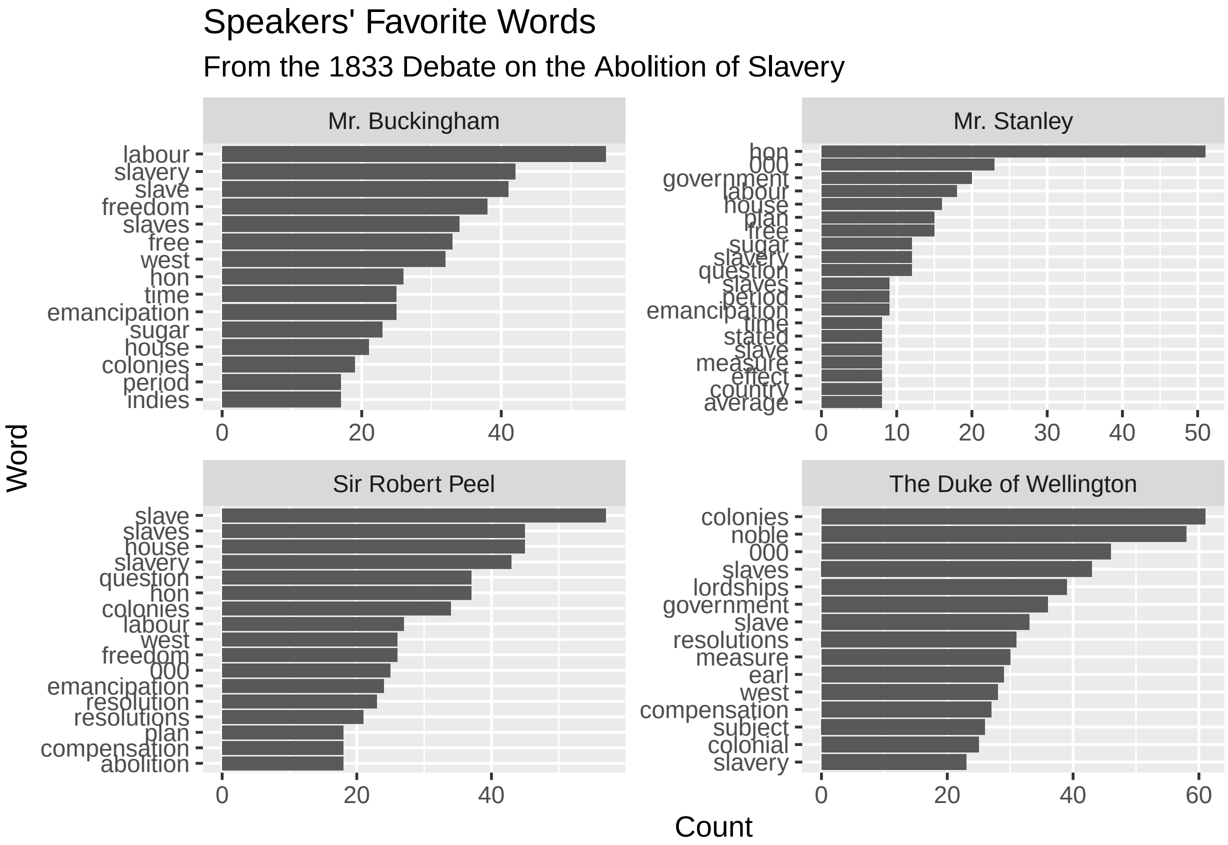
```
head(abolition_speakers)
```

```
## # A tibble: 6 x 3
## # Groups:   speaker [1]
##   speaker        word      n
##   <chr>          <chr> <int>
## 1 Mr. Buckingham 000      15
## 2 Mr. Buckingham 125       1
## 3 Mr. Buckingham 14         1
## 4 Mr. Buckingham 1793       1
## 5 Mr. Buckingham 18         1
## 6 Mr. Buckingham 2          1
```

For our visualization, we can use just the top 15 owrds

```
abolition_speakers <- abolition_speakers %>%
  arrange(desc(n)) %>%
  top_n(15) %>%
  ungroup() %>%
  mutate(word = reorder_within(word, n, speaker))
```

Now visualize:



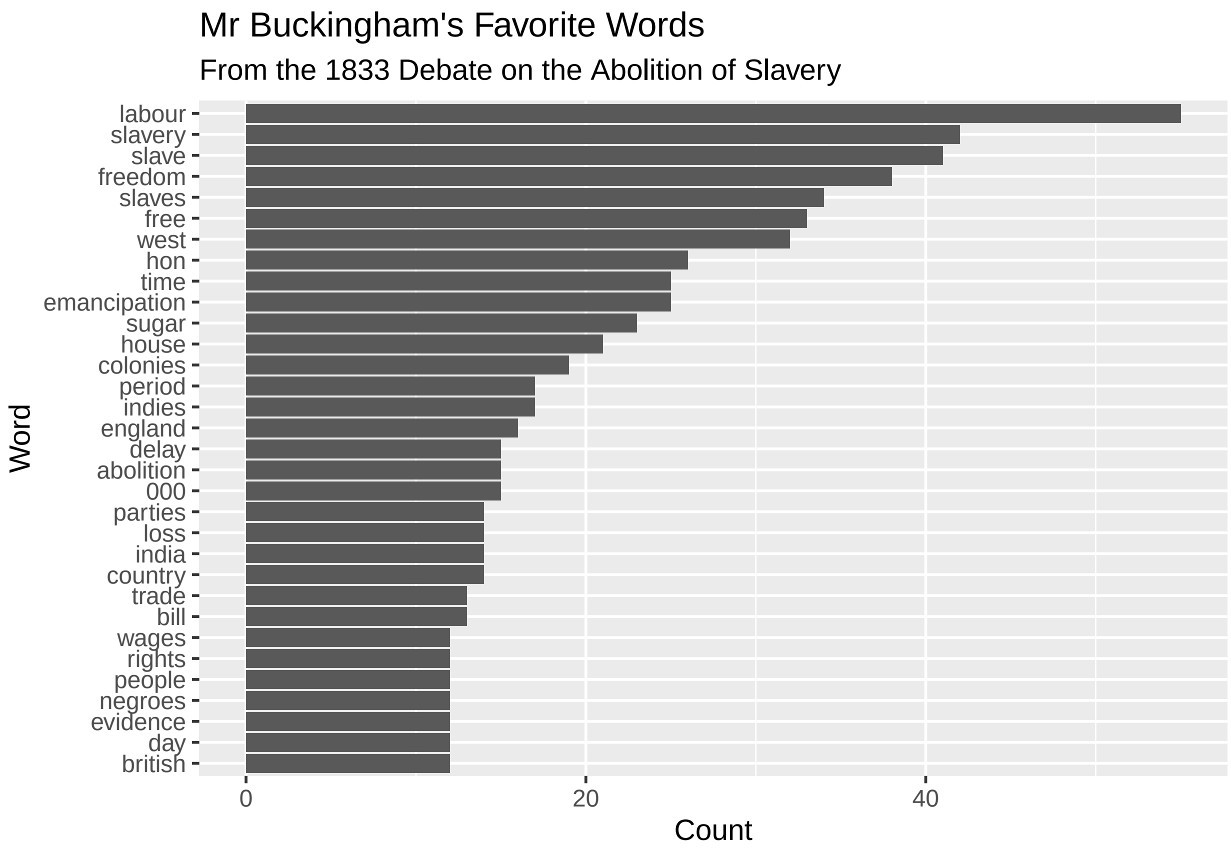
Unsurprisingly, many of the top speakers are less interested in the issue of slavery than in problems of governance. Even in a distant reading, we can detect that Robert Peel and the Duke of Wellington were more interested in problems of "compensation" for the owners who benefited from the system of 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 one exception to this trend, in the four speakers examined, appears to be James Silk Buckingham, the reformer who toured sites of slavery round the world, collecting anecdotes from the slaves themselves about their experiences of cruelty.

We can look into Mr. Buckingham's language in greater detail.

```
mr_buckingham_slavery_debates <- slavery_debates %>%
  left_join(speaker_metadata_1830) %>%
  filter(speaker == "Mr. Buckingham") %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  group_by(speaker, word) %>%
  summarize(n = n()) %>%
  arrange(desc(n)) %>%
  top_n(30)

ggplot(data = mr_buckingham_slavery_debates,
       aes(x = reorder(word, n), y = n)) +
  geom_col() +
  coord_flip() +
  labs(title = "Mr Buckingham\'s Favorite Words",
       subtitle = "From the 1833 Debate on the Abolition of Slavery",
       x = "Word",
       y = "Count")
```



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 int he West Indies.

We should not be too quick to leap to conclusions about how Buckingham used these words without reading directly from his speeches.

```
buckingham_sentences <- slavery_debates %>%
  left_join(speaker_metadata_1830) %>%
  filter(speaker == "Mr. Buckingham") %>%
  filter(str_detect(text, "labour"))
```

```
buckingham_sentences$text[1:10]
```

```
##  [1] "But, for the present, he would content himself with saying, that the two principal features to
##  [2] "This demand, however, was opposed by the Ministers, as well as by the West Indians, on various
##  [3] "As to the indolence of the slaves, and their incapacity or unwillingness to labour for their ow
##  [4] "All men disliked to labour more than was necessary to obtain for them the enjoyments of life; b
##  [5] "The other work was one entitled \"\"Wages, of the Whip, \"\" drawn up by Mr. Conder, a well kno
##  [6] "There was not the slightest ground, therefore, for the assumption, that, if liberated, the negr
##  [7] "Members be ignorant of the fact, that in our own immense empire of the East Indies, any quanti
##  [8] "If, instead of the unjust preference which had hitherto been given to the produce of slave-labo
##  [9] "If loss should actually accrue for the first few years, from the change from a system of slave
## [10] "member for Middlesex, that a much larger body of evidence than he seemed to contemplate, could
```

We show here only the first ten sentences from Buckingham"s speeches on the Abolition of Slavery Bill. We can see from these sentences the many ways in which he invokes "labour." Not all of them are representations about his ideas; in the first several sentences he ventriloquizes the sentiments of slave-owners who believe that their production will grind to a halt unless they employ extreme cruelty. We see Buckingham arguing against that position and for the validity of and productivity of free labor.

If we were working on a paper analyzing these speeches, we would likely want to read Buckingham's speeches in their entirety. We can do this by loading "file_metadata" to get information on which speech a sentence belongs.

```
data("file_metadata_1830")
```

```
head(file_metadata_1830)
```

```
##   sentence_id speech_id debate_id src_file_id      src_image src_column
## 1 S2V0022P0_0     51883      6335   S2V0022P0 S2V0022P0I0030          4
## 2 S2V0022P0_1     51883      6335   S2V0022P0 S2V0022P0I0030          4
## 3 S2V0022P0_2     51883      6335   S2V0022P0 S2V0022P0I0030          4
## 4 S2V0022P0_3     51883      6335   S2V0022P0 S2V0022P0I0030          4
## 5 S2V0022P0_4     51883      6335   S2V0022P0 S2V0022P0I0030          4
## 6 S2V0022P0_5     51883      6335   S2V0022P0 S2V0022P0I0030          4
```

`file_metadata_1830` is a data frame with fields that either reference the original Hansard transcripts (such as `src_column`, for the column of the sentence) or were added to provide insight into the content and structure of the debates (such as `debate_id`, which tells us which sentences belonging to an entire debate).

It can be joined with `buckingham_sentences` to give us a way to easy access an entire speech by Mr. Buckingham, as well as an entire debate.

```
buckingham_sentences <- left_join(buckingham_sentences, file_metadata_1830)

buckingham_sentences <- buckingham_sentences %>%
  select(sentence_id, speech_id, debate_id, speechdate, debate, text)
```

```
head(buckingham_sentences)
```

```
##      sentence_id speech_id debate_id speechdate
## 1 S3V0018P0_6610     87635      9789 1833-05-31
## 2 S3V0018P0_6625     87635      9789 1833-05-31
## 3 S3V0018P0_6643     87635      9789 1833-05-31
## 4 S3V0018P0_6646     87635      9789 1833-05-31
## 5 S3V0018P0_6657     87635      9789 1833-05-31
## 6 S3V0018P0_6658     87635      9789 1833-05-31
##                                         debate
## 1 MINISTERIAL PLAN FOR THE ABOLITION OF SLAVERY. ]
## 2 MINISTERIAL PLAN FOR THE ABOLITION OF SLAVERY. ]
## 3 MINISTERIAL PLAN FOR THE ABOLITION OF SLAVERY. ]
## 4 MINISTERIAL PLAN FOR THE ABOLITION OF SLAVERY. ]
## 5 MINISTERIAL PLAN FOR THE ABOLITION OF SLAVERY. ]
## 6 MINISTERIAL PLAN FOR THE ABOLITION OF SLAVERY. ]
##
## 1
## 2
## 3
## 4
## 5
## 6 There was not the slightest ground, therefore, for the assumption, that, if liberated, the negroes
```

Here we filter for a speech ID to view his sentences belonging to a single speech.

```
entire_speech <- buckingham_sentences %>%
  filter(speech_id == 89834) %>%
  select("text")
```

```
print(entire_speech$text)
```

```
## [1] "If the restraint were intended to prepare the slave for freedom, then he should say that a singl
## [2] "But for the artisans and artificers, who in large numbers maintained themselves by their own lab
## [3] "He was one of those who contended that no compensation should be paid till loss could be proved
## [4] "But for the cultivators, he believed no compensation would ever be required, as it had been prov
```

We can do something similar to see all the speeches belonging to a specific debate based on the "debate_id" field.

```
entire_debate <- buckingham_sentences %>%
  filter(debate_id == 9789) %>%
  select("text")
```

```
print(entire_debate$text)
```

```
## [1] "But, for the present, he would content himself with saying, that the two principal features to w
## [2] "This demand, however, was opposed by the Ministers, as well as by the West Indians, on various g
## [3] "As to the indolence of the slaves, and their incapacity or unwillingness to labour for their own
## [4] "All men disliked to labour more than was necessary to obtain for them the enjoyments of life; be
## [5] "The other work was one entitled \"\"Wages, of the Whip, \"\" drawn up by Mr. Conder, a well know
## [6] "There was not the slightest ground, therefore, for the assumption, that, if liberated, the negro
## [7] "Members be ignorant of the fact, that in our own immense empire of the East Indies, any quantity
## [8] "If, instead of the unjust preference which had hitherto been given to the produce of slave-labou
## [9] "If loss should actually accrue for the first few years, from the change from a system of slave l
```

Reading a sentence within the context of an entire speech as a string of speeches gives us a much richer sense of the back-and-forth of parliamentary speeches, as well as the positions taken by each speaker. We recommend that the analyst of history alternate between "close reading" of these speeches in context alongside the "distant reading" of words and phrases charted over time.

Women in 19th-century Parliament

For the most part, an analysis of speakers in Parliament will take the researcher down the road of the many elite white men who assumed powerful positions or passed honorary titles to their sons. Only 12 named women spoke in Parliament throughout the 19th-century. One woman who spoke the most was Mrs. Walrand from the 1824 debate, “Motion Respecting the Trial and Condemnation of Missionary Smith at Demerara.” Exploring her role in this debate can give us insight and speculation into the criteria that had to be set in place for a woman to appear as a speaker and instrument to the power of Parliament.

“Motion Respecting the Trial and Condemnation of Missionary Smith at Demerara” reviews the trail, conviction, and death of John Smith, an Methodist missionary from England assigned to the British slave colony of Demerara in South America. Mr. Smith’s trial accused him of assisting a slave rebellion that took place in Demerara in August of 1823. MP Brougham brought the topic of Mr. Smith’s trial before Parliament to argue that the legal procedings convicting Mr. Smith were unjust and strayed from the guidance of Britain to the leaders of its Demerara colony. One could read his testimony as an early foundation for Britian’s movement towards the Slavery Abolition Act of 1833, but his concerns–as well as the testimonies of other speakers–are nonetheless intertwined within a biased legal institution.

Mrs. Walrand spoke as a witness to the slave rebellion and served as a character witness condemning Mr. Smith. According to her testimony, she saw first hand the brutal violence of the rebells, having been a “defenceless lady” fired at by these “savages.”

```
data("hansard_1820")
data("speaker_metadata_1820")
data("file_metadata_1820")

mrs_walrand_speaker_metadata <- speaker_metadata_1820 %>%
  filter(str_detect(speaker, regex('Mrs(.*)Walrand', ignore_case = T)))

mrs_walrand_sentences <- left_join(mrs_walrand_speaker_metadata, file_metadata_1820, by = "sentence_id"
  left_join(., hansard_1820, by = "sentence_id") %>%
  select(debate_id, text)
```

```
head(mrs_walrand_sentences)
```

```
##   debate_id
## 1      5142
## 2      5142
## 3      5142
## 4      5142
## 5      5142
## 6      5142
##
## 1
## 2
## 3 I entreated the guard, in the name of every principle of humanity, just to let me send to Golden G
## 4
## 5
## 6
```

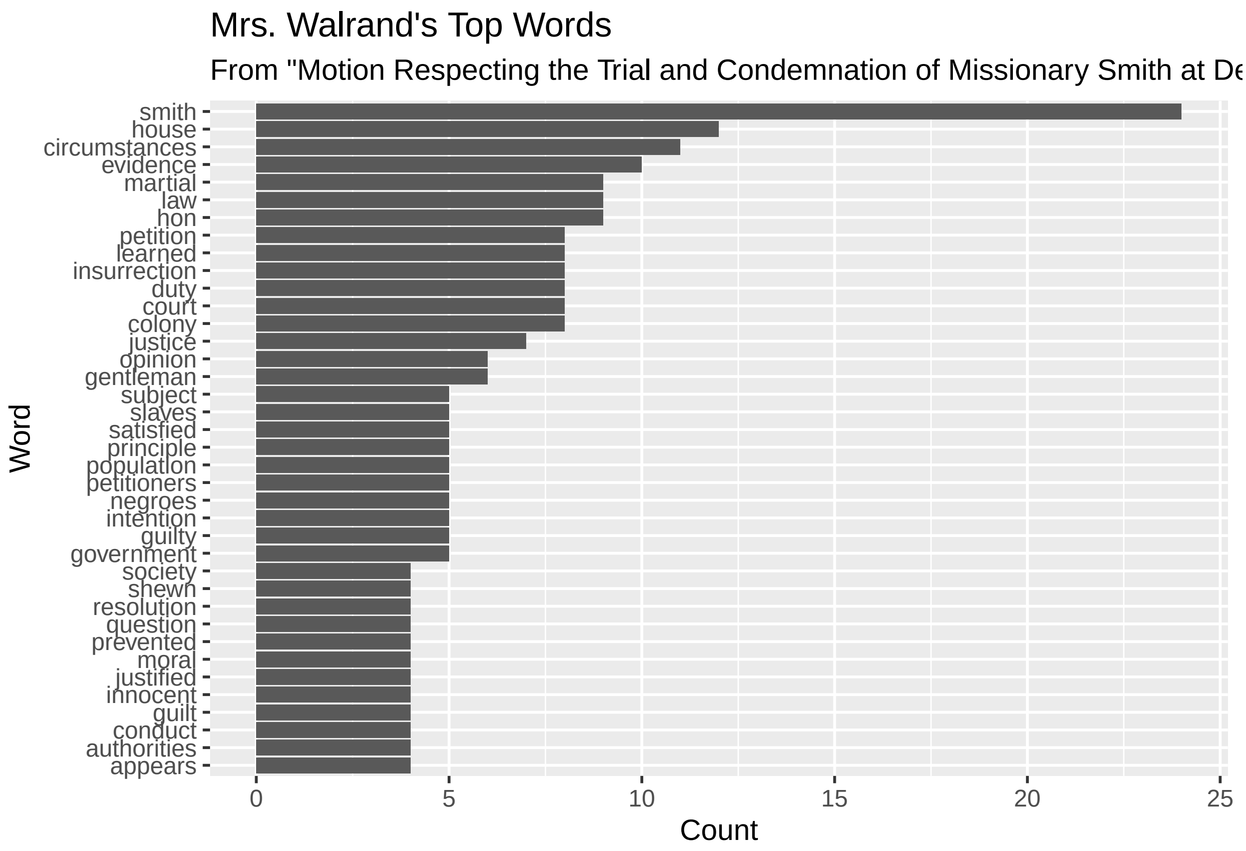
```
entire_debate <- hansard_1820 %>%
  left_join(file_metadata_1820, hansard_1820, by = "sentence_id") %>%
  filter(debate_id == 5142) %>%
  select("text")
```

```
head(entire_debate$text, 20)
```

```
##  [1] "rose, and addressed the House to the following effect:-"
##  [2] "; I confess that, in bringing before this House the question on which I now rise to address yo
##  [3] "I cannot conceal from myself, that, even in quarters where one would least have expected it, a
##  [4] "Many persons who have, upon all other occasions, been remarkable for their manly hostility to a
##  [5] "Nay, they would fain fasten upon any excuse to get rid of the subject. \"\""
##  [6] "What signifies inquiring, \"\" say they, \"\"into a transaction which has occurred in a differe
##  [7] "As if distance or climate made any difference in an outrage upon law or justice."
##  [8] "One would have rather expected that the very idea of that distance; the circumstance of the eve
##  [9] "Then, says another, too indolent to inquire, but prompt enough to decide, \"\"It is true there
## [10] "* From the edition published by Hatchard and Son, with the sanction of the London Missionary So
## [11] "ject;"
## [12] "but then every body knows how those petitions are procured, by what descriptions of persons th
## [13] "And, after all, it is merely about a poor missionary!\"\""
## [14] "It is the first time that I have to learn that the weakness of the sufferer; his unprotected si
## [15] "But, it is not enough that he was a Missionary; to make the subject still more unpalatable, for
## [16] "I hasten to this objection, with a view at once to dispose of it."
## [17] "Suppose Mr. Smith had been a methodist; what then?"
## [18] "Does his connexion with that class of religious people, because, on some points essential in th
## [19] "Are British subjects to be treated more or less favourably in courts of law; are they to have a
## [20] "Had he belonged to the society of the methodists, and been employed by the members of that comm
```

```
mrs_walrand_top_words <- left_join(mrs_walrand_speaker_metadata, hansard_1820) %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  filter(is.na(as.numeric(word))) %>%
  group_by(speaker, word) %>%
  summarize(n = n()) %>%
  arrange(desc(n)) %>%
  top_n(30)

ggplot(data = mrs_walrand_top_words,
       aes(x = reorder(word, n), y = n)) +
  geom_col() +
  coord_flip() +
  labs(title = "Mrs. Walrand\'s Top Words",
       subtitle = "From \"Motion Respecting the Trial and Condemnation of Missionary Smith at Demerara\
       x = "Word",
       y = "Count")
```



Exercises

1) Use the techniques shown above to move between a distant reading of the top words stated by Mrs. Walrand or Louisa Demont in all of her speeches to the full debate. (As a reminder, you can view all of a speaker's speeches by filtering for just that speaker.) How does transitioning between the different modes of reading text provide insight into this specific debate? How does this illuminate the role of women in 19th-century Parliament?

2) As we discussed in the beginning of this chapter, analyzing the individual or collective speeches of men in Parliament can give insight into the workings of the Parliament as an institution of power. We can explore the roles of elite white men as they exerted their will–and the will of their peers–through the Parliamentary institution. We can also explore the roles of women in this space and examine how women differ in their contributions, as they were prohibited from becoming members of Parliament. Further, the woman's presence had to be mediated by one of the MPs, as they could not be there without being selected or approved. In this way, women were treated as instruments of power in Parliament.

We can further explore how the role of women manifested in Parliament. First, count the total number of words women spoke in Parliament. To do this, refer to the table below, which visualizes the women in Parliament and the year in which they spoke. Can you find a pattern across the kinds of topics women are brought to Parliament to speak on?

Name	Year	Name	Year
Mrs. (Mary Ann) Clark	1809	Mrs. Disraeli	1856
Mrs. Bridgeman	1809	Ms. Cunninghame Graham	1887
Miss Mary Ann Taylor	1809	Mrs. Dillon	1902
Louisa Demont	1820	Mrs. M'Govern	1902
Franchette Martigner	1820	Mrs. Mitchel-Thomson	1907
Mrs. Walrand	1824	Ms. Cave	1907

Note: The way in which their names are recorded in this table might not reflect the way in which their names appear in the debates. This is because the names have been consolidated into one representation

3) Adjust the above code to search for a different office title. What can you infer about the different MPs who held this title? Can we support our inferences through close readings of their speeches?

To get you started, here are a few office positions: - Chancellor of the Exchequer - Prime Minister - Attorney General - Lord Chancellor

4) Visualization N shows the most verbose MPs for every year of the decade 1830. The counts are based on the "speaker" column–that is, the name of the speaker as it was originally transcribed in Hansard. Adjust the code so that instead you visualize the "suggested_speaker" column. This column contains the names of the speakers after they went through a "disambiguation" processes, that, where possible, assigned speakers with a standardized name and ID number. How are these two visualizations different? What are some pros and cons about visualizing the original speaker names and the disambiguated speaker names?

5) Above, we explored the contributions of four major speakers about slavery. But the names that we explored in Figure N don't match those that we saw were most important to the two peaks of debate in 1833 and 1838, shown in Figure M. Use the code to generate Figure N with an the list of names we extracted from Figure M: Mr. Secretary Stanley, the Duke of Wellington, Mr. Buckingham, Mr. Fowell Burton, the Earl of Ripon, Lord Brougham, and Lord Gleneig. What do you learn from this exercise? How are the results different than those in Figure N?