CTA Final

2024-04-23

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

library("ggplot2")

## Warning: package 'ggplot2' was built under R version 4.3.2

library("plyr")  
library("gdata")

##   
## Attaching package: 'gdata'

## The following object is masked from 'package:stats':  
##   
## nobs

## The following object is masked from 'package:utils':  
##   
## object.size

## The following object is masked from 'package:base':  
##   
## startsWith

library("stringr")  
library("data.table")

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:gdata':  
##   
## first, last

library("stringi")   
library("dplyr")

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':  
##   
## between, first, last

## The following objects are masked from 'package:gdata':  
##   
## combine, first, last, starts\_with

## The following objects are masked from 'package:plyr':  
##   
## arrange, count, desc, failwith, id, mutate, rename, summarise,  
## summarize

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

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

library("tidytext")   
library("kableExtra")

## Warning: package 'kableExtra' was built under R version 4.3.2

##   
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':  
##   
## group\_rows

library("babynames")  
library("ggthemes")  
library("patchwork")

##Download the Data  
data = fread("/Users/tyqtyq/Desktop/dataverse\_files/uk\_data.csv", encoding="UTF-8")  
##There are 958925 obs(speeches) will be study

#Check the colnames of the data set  
colnames(data)

## [1] "id\_speech" "id\_mp"   
## [3] "period" "last\_name"   
## [5] "first\_name" "date"   
## [7] "pm\_questions" "queen\_debate\_day1"   
## [9] "queen\_debate\_others" "m\_questions"   
## [11] "u\_questions" "other\_debate"   
## [13] "leader" "prime\_minister"   
## [15] "senior\_minister" "shadow"   
## [17] "cabinet" "chair"   
## [19] "government" "female"   
## [21] "age" "electoral\_cycle"   
## [23] "party" "linear\_trend"   
## [25] "words" "text"   
## [27] "emotive\_count" "neutral\_count"   
## [29] "emotive\_rhetoric" "emotive\_rhetoric\_log"   
## [31] "emotive\_words" "top\_topic"   
## [33] "anew\_rescaled" "emotive\_rhetoric\_liwc"   
## [35] "positive\_count" "negative\_count"   
## [37] "emotive\_positive" "emotive\_negative"   
## [39] "emotive\_count\_250\_8" "neutral\_count\_250\_8"   
## [41] "emotive\_rhetoric\_250\_8" "emotive\_count\_300\_10"   
## [43] "neutral\_count\_300\_10" "emotive\_rhetoric\_300\_10"  
## [45] "emotive\_count\_a1" "neutral\_count\_a1"   
## [47] "emotive\_rhetoric\_a1" "emotive\_count\_a2"   
## [49] "neutral\_count\_a2" "emotive\_rhetoric\_a2"

##Data Preparation: (1)Creat the Time vriable for plotting trens   
data$time= NA  
data$time[data$date>=as.Date("2001-01-01") & data$date<=as.Date("2001-06-30")] = "01/1"  
data$time[data$date>=as.Date("2001-07-01") & data$date<=as.Date("2001-12-31")] = "01/2"  
data$time[data$date>=as.Date("2002-01-01") & data$date<=as.Date("2002-06-30")] = "02/1"  
data$time[data$date>=as.Date("2002-07-01") & data$date<=as.Date("2002-12-31")] = "02/2"  
data$time[data$date>=as.Date("2003-01-01") & data$date<=as.Date("2003-06-30")] = "03/1"  
data$time[data$date>=as.Date("2003-07-01") & data$date<=as.Date("2003-12-31")] = "03/2"  
data$time[data$date>=as.Date("2004-01-01") & data$date<=as.Date("2004-06-30")] = "04/1"  
data$time[data$date>=as.Date("2004-07-01") & data$date<=as.Date("2004-12-31")] = "04/2"  
data$time[data$date>=as.Date("2005-01-01") & data$date<=as.Date("2005-06-30")] = "05/1"  
data$time[data$date>=as.Date("2005-07-01") & data$date<=as.Date("2005-12-31")] = "05/2"  
data$time[data$date>=as.Date("2006-01-01") & data$date<=as.Date("2006-06-30")] = "06/1"  
data$time[data$date>=as.Date("2006-07-01") & data$date<=as.Date("2006-12-31")] = "06/2"  
data$time[data$date>=as.Date("2007-01-01") & data$date<=as.Date("2007-06-30")] = "07/1"  
data$time[data$date>=as.Date("2007-07-01") & data$date<=as.Date("2007-12-31")] = "07/2"  
data$time[data$date>=as.Date("2008-01-01") & data$date<=as.Date("2008-06-30")] = "08/1"  
data$time[data$date>=as.Date("2008-07-01") & data$date<=as.Date("2008-12-31")] = "08/2"  
data$time[data$date>=as.Date("2009-01-01") & data$date<=as.Date("2009-06-30")] = "09/1"  
data$time[data$date>=as.Date("2009-07-01") & data$date<=as.Date("2009-12-31")] = "09/2"  
data$time[data$date>=as.Date("2010-01-01") & data$date<=as.Date("2010-06-30")] = "10/1"  
data$time[data$date>=as.Date("2010-07-01") & data$date<=as.Date("2010-12-31")] = "10/2"  
data$time[data$date>=as.Date("2011-01-01") & data$date<=as.Date("2011-06-30")] = "11/1"  
data$time[data$date>=as.Date("2011-07-01") & data$date<=as.Date("2011-12-31")] = "11/2"  
data$time[data$date>=as.Date("2012-01-01") & data$date<=as.Date("2012-06-30")] = "12/1"  
data$time[data$date>=as.Date("2012-07-01") & data$date<=as.Date("2012-12-31")] = "12/2"  
data$time[data$date>=as.Date("2013-01-01") & data$date<=as.Date("2013-06-30")] = "13/1"  
data$time[data$date>=as.Date("2013-07-01") & data$date<=as.Date("2013-12-31")] = "13/2"  
data$time[data$date>=as.Date("2014-01-01") & data$date<=as.Date("2014-06-30")] = "14/1"  
data$time[data$date>=as.Date("2014-07-01") & data$date<=as.Date("2014-12-31")] = "14/2"  
data$time[data$date>=as.Date("2015-01-01") & data$date<=as.Date("2015-06-30")] = "15/1"  
data$time[data$date>=as.Date("2015-07-01") & data$date<=as.Date("2015-12-31")] = "15/2"  
data$time[data$date>=as.Date("2016-01-01") & data$date<=as.Date("2016-06-30")] = "16/1"  
data$time[data$date>=as.Date("2016-07-01") & data$date<=as.Date("2016-12-31")] = "16/2"  
data$time[data$date>=as.Date("2017-01-01") & data$date<=as.Date("2017-06-30")] = "17/1"  
data$time[data$date>=as.Date("2017-07-01") & data$date<=as.Date("2017-12-31")] = "17/2"  
data$time[data$date>=as.Date("2018-01-01") & data$date<=as.Date("2018-06-30")] = "18/1"  
data$time[data$date>=as.Date("2018-07-01") & data$date<=as.Date("2018-12-31")] = "18/2"  
data$time[data$date>=as.Date("2019-01-01") & data$date<=as.Date("2019-06-30")] = "19/1"  
data$time[data$date>=as.Date("2019-07-01") & data$date<=as.Date("2019-12-31")] = "19/2"  
  
data$time2 = data$time  
data$time2 = str\_replace(data$time2, "/", "\_")  
  
data$stage = 0  
data$stage[data$m\_questions==1]= 1  
data$stage[data$u\_questions==1]= 2  
data$stage[data$queen\_debate\_others==1]= 3  
data$stage[data$queen\_debate\_day1==1]= 4  
data$stage[data$pm\_questions==1]= 5

##Data Preparation: (2) get simplified dataset and select the variables which are related to the research questions  
 data\_simplified<- data %>%  
 select(period, time, female, party, text, emotive\_count, neutral\_count, emotive\_rhetoric, emotive\_positive, emotive\_negative)  
##Check the new dataset  
head(data\_simplified)

## period time female party  
## 1: 2001-2005 01/1 0 Labour  
## 2: 2001-2005 01/1 0 Conservative  
## 3: 2001-2005 01/1 0 Labour  
## 4: 2001-2005 01/1 0 Conservative  
## 5: 2001-2005 01/1 0 Liberal Democrats  
## 6: 2001-2005 01/1 0 Ulster Unionist Party  
## text  
## 1: Under the provisions of Standing Order No 1A I am now required to ascertain whether Mr Michael J Martin is willing to be chosen as Speaker  
## 2: I beg to move That Mr Michael J Martin do take the Chair of this House as Speaker Mr Dalyell I begin by congratulating you on your accession as Father of the House Given your commitment to the House and your knowledge of how to use it effectively it gives all of us pleasure to see you supervising our proceedings today I too wish you a full and speedy recovery I also hope that you will continue to be as irreverent to those on your Front Bench as your predecessor as Father of the House was to mine As we meet this afternoon there are two important vacancies in our political institutions and I am happy to make a nomination for one of them In commending the right hon Member for Glasgow Springburn Mr Martin as Speaker I congratulate him on his re election to Parliament Although his election was contested I suspect that he did not experience the aggravation that the rest of us suffered in getting here I mention in passing that we in North West Hampshire saw a new form of co operation between the other two parties the Labour party ran a low key campaign to let the Liberal Democrats have a clear shot at me and the Liberal Democrats ran a low key campaign to give Labour a shot I am happy to say that my majority increased to more than 12 000 There are two reasons why I move the motion the first of which is continuity I do not believe that our Speaker should be put into play at the beginning of every Parliament I can do no better than quote the second report of the Select Committee on Procedure published in February In our view it would be undesirable in these circumstances for a multi candidate ballot to take place automatically As we have seen since the middle of the nineteenth century there has been a strong presumption that a Speaker once elected by the House is not subsequently challenged If it were to become accepted that a change in the composition of the House following a General Election were as a matter of course to lead to a change in the occupancy of the Chair we believe there are grave dangers that the office itself would be destabilised and in danger of becoming politicised Equally however we believe it is important that the House should not be denied the right to change its Speaker however unlikely it may be that that right will be exercised I agree with that sentiment as did the House when it agreed to change our procedures Although the Speaker should be validated there should be a presumption against challenging the incumbent Continuity is not the only reason If we were starting from scratch I believe that the House would choose the right hon Member for Springburn as its Speaker today His long service in the House and deep roots in the Back Benches his work on the Chairmen s Panel and domestic Committees his experience as Deputy Speaker his genial and approachable manner underpinned by a deep affection and commitment to the House all those qualities strike a chord with the House That commitment was confirmed in his acceptance speech last October and reinforced in the statement that we have just heard which was greatly welcomed All Speakers develop their own style and we saw the right hon Member for Springburn develop his in the previous Parliament a more approachable informal style of Speaker mixing with Members in the Tea Room and elsewhere no wig or silk stockings rather like the Scandinavian approach to the monarchy but so far without the bicycle A key role lies ahead for Mr Speaker in this Parliament There is a strong view that the House should reassert the accountability of the Executive to Parliament A growing number of hon Members want the House to be more relevant to the concerns of those whom we represent more effective in what it does and a better link between Government and governed The right hon Member for Springburn is well qualified to supervise and manage that debate during this Parliament I believe that he is entitled to support from the whole House in that role The House will understand why I was unable to propose the right hon Member for Springburn last time but I have no hesitation in proposing him today Question put and agreed to Resolved That Mr Michael J Martin do take the Chair of this House as Speaker Whereupon Mr Tam Dalyell left the Chair and Mr Michael J Martin was taken out of his place and conducted to the Chair by Sir George Young and Ann Keen  
## 3: Mr Speaker Elect it is a great pleasure to congratulate you on your return to the Chair of the House I think I speak for all Members when I say it is also a great pleasure for all of us that we have completed your election for a second time in rather shorter order than the six hours that it took us the first time May I add my congratulations to my neighbour my hon Friend the Member for Linlithgow Mr Dalyell on becoming Father of the House No hon Member has shown greater affection for the traditions of the House or carried out his duties as a Member with greater diligence than my hon Friend As a fellow West Lothian Member of Parliament perhaps I can explain on his behalf to the House that his injury is a tribute to the diligence with which he pursues his constituency office as it was incurred with an excess of enthusiasm at a meeting of the Linlithgow football team It therefore was an asset not a liability in the recent election Mr Speaker Elect there have been only 155 Speakers before you in the many centuries of your office When Speaker Yelverton in 1597 was asked to describe the necessary qualities of a Speaker he replied A man big and comely his carriage majestical his nature haughty and his purse plentiful Fortunately the characteristics of the Speaker have changed over the subsequent four centuries For myself I am relieved that we no longer expect our Speaker to be haughty just as you must be relieved that the House does not expect your purse to be plentiful Over the past seven months you have shown all the necessary qualities of a modern Speaker We could have expected no less from a Speaker who has brought to the Chair of the Chamber one of the longest records of service on the Chairmen s Panel You have been fair but firm you have turned aside confrontation with humour and you have got the better of those unwise enough to challenge your authority Many an awkward moment has been defused with your trademark catch phrase It s no nice Since you came to office your voice has become familiar in households across our country You spoke for all your Scottish compatriots in the Chamber when you magnificently brushed aside an impertinent question from the BBC on your accent with the retort I don t have an accent other people have an accent Yours is of course an accent which would normally lend authority to one of the traditional roles of the Speaker the selection of a Scotch whisky for the Speaker s brand For you though as a teetotaller that selection presented some obvious difficulties Your solution delighted the Members whom you invited on to an all party committee to carry out extensive research and tasting on your behalf Not always believing that we get the press that we deserve is of course a frequent experience of Members of this House even if the rest of the country thinks that we get the press that we richly deserve If ever in this Parliament the Lobby dares hint at criticism of our Speaker it can only be because as Chairman of the Administration Committee you banned Lobby correspondents from the Terrace of the Commons unless personally supervised You thereby liberated a grateful House to relax in privacy on the Terrace in these summer months Both you and I are now of that age when we share that puzzling sensation after each election that Members of Parliament seem to be getting younger than we remember To an even younger Chamber than before your commitment that families will be remembered in the proceedings of the House is particularly welcome That commitment comes of course from the affection and importance that you attach to your own family Those who know you well know also that no words of congratulations to you would be complete unless they also expressed appreciation of Mary your wife for her support to you in your role here and in your constituency Mr Speaker Elect on first taking the Chair you said that the Speaker had a clear duty to every side of the House especially to the Back Benchers It is because you have served the whole House with impartiality that the whole House has today returned you to the Chair unanimously Your colleagues and your friends congratulate you on it  
## 4: On behalf of all my right hon and hon Friends on the Opposition Benches I offer our sincerest congratulations to you Mr Speaker Elect on your re election Many colleagues on both sides of the House will share my relief that it was accompanied by less controversy and took dramatically less time than your election in October Your unanimous re election underlines the important constitutional fact that you are now very much Speaker of the whole of the House of Commons I add the Opposition s congratulations to those expressed to the Father of the House the hon Member for Linlithgow Mr Dalyell who as a highly active Back Bencher and champion of so many causes over decades is in every sense fully qualified for the post It is also a great pleasure to be the first to congratulate the right hon Member for Livingston Mr Cook on his appointment as Leader of the House of Commons We learned from the newspapers that apparently he has been unable to express all his real views as Foreign Secretary over the past four years My hon Friends are much looking forward to asking him for all his real views as Leader of the House during a succession of Question Times and statements over coming months On these occasions Mr Speaker Elect a great deal is usually made of the distinctive parliamentary existence of the holder of your great office that results from the need to cut yourself off from previous party affiliations You have indeed cut yourself off from party affiliations and as has been mentioned served this House impartially It has always been a tradition of Speakers that they do not visit the Tea Room and the bars as part of that procedure You have created a new tradition by being available and visiting the Tea Room and bars but not having a drink there That may keep you happy although it is incomprehensible to the rest of us but we are delighted that you have amended the traditions in that respect On a more serious point you are the custodian of the rules privileges and traditions of this House As the many newly elected Members will soon learn we all look to you as the independent champion of all parties in the House and of the rights of all hon Members As you know I have always made a particular point about protecting the rights of the Opposition Front Bench come to think of it though the rights of Back Benchers need protecting too especially those who have not spoken from the Back Benches for a long time I know that you will protect their rights in this Parliament There are also those occasions of high drama Mr Speaker Elect when votes are tied and it falls to you to use your casting vote although I admit that it is not immediately obvious that that will be necessary in a large number of instances You never know we will be working on it from this side of the House I hope that the fact that the Government have such a majority will not deter hon Members on both sides of the House from doing the job that they have been sent here to do to hold the Government to account Like so many past and present Members you and I care passionately about the House of Commons and the standing in which it is held throughout the country That is why I for one deeply regret the diminution of its importance and reputation which has accelerated although it did not begin in recent years I cannot be alone in thinking that that decline contributed in some ways to the disconnection between the public and Parliament that was highlighted by the lamentably low voter turnout in the general election last week Last October I expressed the hope that you would robustly resist all attempts to downgrade marginalise or bypass the House of Commons I repeat that today Few things would give me and I hope hon Members from all parties greater satisfaction in politics than to see this House restored to the centre of our national life The great issues that will come before us during the next few years some of which may be of supreme importance to the way in which our country is governed should be fully scrutinised and debated first and foremost in this Chamber On a personal note just as you are about to resume your duties I am planning to relinquish mine as Leader of the Opposition although I shall carry them out for a few weeks more I should like to thank you for the courtesy and kindness that you have shown me behind the scenes since you took the Chair of this House last October I am sure that your advice will continue to be of enormous benefit to hon Members throughout the House Again Mr Speaker Elect I congratulate you and wish you well for what undoubtedly promises to be an important and demanding period in our parliamentary history  
## 5: Mr Speaker Elect may I entirely associate my right hon and hon Friends with the personal and unanimous congratulations that the House has recorded on your re election as Speaker this afternoon We wish you all possible success for the forthcoming Parliament You will remember something about which we have joked privately since your first election At that time I recalled our first encounter when I was a schoolboy and you were on a picket line in your then capacity as a NUPE National Union of Public Employees regional official The unanimous endorsement of the House must mean that we have witnessed the ultimate triumph of NUPE man this afternoon We congratulate you on that I should like to take this opportunity also to congratulate the new Father of the House whose courtesy and advice to all of us especially those who have entered the House over the years as new Members has been unstinting and much appreciated It is a great happiness to see him occupy that position on behalf of us all but for me that happiness is tinged with one wee regret I shall dearly miss his distinguished predecessor Sir Edward Heath at Prime Minister s questions I do not have any ambition as Lib Dem leader to write a diary and publish it The book that I want to write and publish one day is the off the record commentary of Sir Edward Heath between 3 o clock and 3 30 every Wednesday afternoon on the parliamentary Conservative party and the questions that its members put to the Prime Minister of the day I shall miss the sensation when Sir Edward was notably upset or distressed by the viewpoints given by one of his own He had a tendency to expel a great sigh of frustration which tended to move me significantly along the Bench I shall miss those noises We must all share a sense of disappointment if not foreboding about the lack of engagement of so many of our fellow citizens in the election That being so I hope that early steps will be taken to examine our procedures and practices in this place to ensure that they are as relevant and comprehensible as possible to those outside There is a danger that all of us in all parties will be engulfed if we are not careful Given the balance of the outcome of the election I hope that you Mr Speaker Elect your senior officials and the new Leader of the House will give fresh consideration to the procedures of the House to ensure that they are made more flexible in some ways the House of Lords may provide instruction for us with regard to the rights of all parties especially those in opposition We look forward to contributing constructively to such discussions at the earliest opportunity In the meantime we wish you well Sir  
## 6: Mr Speaker Elect on my behalf and that of my colleagues I offer sincere congratulations on your being re elected as Speaker of this place We can vouch for your approachability the care with which you have listened to the issues that Members have brought to you and the way in which you have responded to them We look forward with confidence to that same care and consideration being offered in the months and years to come I congratulate the Father of the House on his conduct of today s proceedings It is not necessary for me to emphasise the way in which he has built a reputation over the years as a Back Bencher who has held Governments of all colours to account That reminds us that that is the prime function that many of us have in the House I am sure Mr Speaker Elect that with your consideration for Back Benchers and for the primary role of the House you will always be ready to assist us in ensuring that the Government are held to account It is commonplace these days to regret the decline of the standing of the House but we would all do well to remember that its standing depends upon each and every one of us in the contribution that we make to the House I am sure that we shall all be conscious of that responsibility However we rely upon you Mr Speaker Elect when it comes to issues where government may perhaps wish to cut corners or to do things elsewhere to insist upon the primacy of the House I shall make a cautionary comment on the remarks of the right hon Member for Ross Skye and Inverness West Mr Kennedy the leader of the Liberal Democrats We want to see the procedures of the House become more effective and clearly and easily understood However virtually every time that the reform of our procedures is broached the result is that the life of government is made easier That is not our objective  
## emotive\_count neutral\_count emotive\_rhetoric emotive\_positive  
## 1: 0 1 -6.666667 0.000000  
## 2: 52 33 5.000000 8.684211  
## 3: 55 29 7.471264 9.770115  
## 4: 61 25 9.137056 10.913706  
## 5: 45 14 13.478261 10.869565  
## 6: 22 6 10.457516 9.803922  
## emotive\_negative  
## 1: 0.0000000  
## 2: 0.7894737  
## 3: 0.5747126  
## 4: 0.2538071  
## 5: 2.6086957  
## 6: 0.6535948

#Data Organization: #obs in the dataset with a value of "emotive rhetoric" greater than 0, i.e., the text used more emotional words, were filtered out to form a new dataset.  
data\_emotive <- data\_simplified %>%   
 select(period, time, female, party, text, emotive\_count, neutral\_count, emotive\_rhetoric, emotive\_positive, emotive\_negative) %>%   
 filter(emotive\_rhetoric>0)

###Data Organization:obs in the dataset with a value of "emotional rhetoric" less than 0, i.e., the text used more neutral words, were filtered out to form a new dataset.  
data\_neutral <- data\_simplified %>%   
 select(period, time, female, party, text, emotive\_count, neutral\_count, emotive\_rhetoric) %>%   
 filter(emotive\_rhetoric<=0)

##Term Frequency of Female-related words  
## Word Frequency of Female-related Words in all data  
#Since the complete dataset is too large, I first sample it  
data\_samp <- data\_simplified %>%  
 sample\_n(10000)

##Tidy the text  
tokens\_all <- data\_samp %>%  
 unnest\_tokens(word, text)  
head(tokens\_all)

## period time female party emotive\_count neutral\_count emotive\_rhetoric  
## 1: 2001-2005 01/2 0 Labour 2 2 0  
## 2: 2001-2005 01/2 0 Labour 2 2 0  
## 3: 2001-2005 01/2 0 Labour 2 2 0  
## 4: 2001-2005 01/2 0 Labour 2 2 0  
## 5: 2001-2005 01/2 0 Labour 2 2 0  
## 6: 2001-2005 01/2 0 Labour 2 2 0  
## emotive\_positive emotive\_negative word  
## 1: 4.347826 0 i  
## 2: 4.347826 0 am  
## 3: 4.347826 0 struggling  
## 4: 4.347826 0 to  
## 5: 4.347826 0 understand  
## 6: 4.347826 0 the

## remove all stop words（the original dataset has removed all apostrophes)  
tokens\_all<- tokens\_all %>%  
 filter(!word %in% stop\_words$word)

##Assign a value of "1" to each obs for easy counting of quantities.  
tokens\_all <- tokens\_all %>%  
 mutate(obs=1)

##Word Frequency Analysis: Calculating the number of female-related words in all texts(sampled)  
tokens\_all$womword <- as.integer(grepl("women|woman|female|girl|feminist|feminism|feminine|gender|harassment|sexism|sexist|lady|madam|womenlike|ladylike", x = tokens\_all$word))

## Word Frequency of Female-related Words in Texts Using Strongly Emotive Rhetoric  
##Tidy the text  
tokens\_emo <- data\_emotive %>%  
 unnest\_tokens(word, text)  
head(tokens\_emo)

## period time female party emotive\_count neutral\_count  
## 1: 2001-2005 01/1 0 Conservative 52 33  
## 2: 2001-2005 01/1 0 Conservative 52 33  
## 3: 2001-2005 01/1 0 Conservative 52 33  
## 4: 2001-2005 01/1 0 Conservative 52 33  
## 5: 2001-2005 01/1 0 Conservative 52 33  
## 6: 2001-2005 01/1 0 Conservative 52 33  
## emotive\_rhetoric emotive\_positive emotive\_negative word  
## 1: 5 8.684211 0.7894737 i  
## 2: 5 8.684211 0.7894737 beg  
## 3: 5 8.684211 0.7894737 to  
## 4: 5 8.684211 0.7894737 move  
## 5: 5 8.684211 0.7894737 that  
## 6: 5 8.684211 0.7894737 mr

## remove all stop words（the original dataset has removed all apostrophes)  
tokens\_emo<- tokens\_emo %>%  
 filter(!word %in% stop\_words$word)

##Check the dataset  
head(tokens\_emo)

## period time female party emotive\_count neutral\_count  
## 1: 2001-2005 01/1 0 Conservative 52 33  
## 2: 2001-2005 01/1 0 Conservative 52 33  
## 3: 2001-2005 01/1 0 Conservative 52 33  
## 4: 2001-2005 01/1 0 Conservative 52 33  
## 5: 2001-2005 01/1 0 Conservative 52 33  
## 6: 2001-2005 01/1 0 Conservative 52 33  
## emotive\_rhetoric emotive\_positive emotive\_negative word  
## 1: 5 8.684211 0.7894737 beg  
## 2: 5 8.684211 0.7894737 move  
## 3: 5 8.684211 0.7894737 michael  
## 4: 5 8.684211 0.7894737 martin  
## 5: 5 8.684211 0.7894737 chair  
## 6: 5 8.684211 0.7894737 house

##Get the number of women-related words in all speeches through "filter" function  
 all\_women\_counts <-tokens\_all %>%   
 filter(womword==1)  
 ##Check the new dataset (especially the number of obs)  
 head(all\_women\_counts)

## period time female party emotive\_count neutral\_count  
## 1: 2017-2019 19/2 0 Conservative 18 1  
## 2: 2015-2017 16/1 1 Scottish National Party 27 35  
## 3: 2010-2015 14/2 0 Conservative 1 3  
## 4: 2010-2015 13/1 0 Conservative 1 2  
## 5: 2017-2019 17/2 1 Conservative 58 46  
## 6: 2017-2019 17/2 1 Conservative 58 46  
## emotive\_rhetoric emotive\_positive emotive\_negative word obs womword  
## 1: 40.476190 16.666667 11.9047619 lady 1 1  
## 2: -2.797203 3.496503 0.6993007 madam 1 1  
## 3: -3.448276 1.724138 0.0000000 lady 1 1  
## 4: -3.846154 3.846154 0.0000000 lady 1 1  
## 5: 2.247191 4.868914 0.7490637 lady 1 1  
## 6: 2.247191 4.868914 0.7490637 lady 1 1

##There are 120404 obs in the new dataset

##Assign a value of 1 to each observation for easy counting of quantities.  
tokens\_emo <- tokens\_emo %>%  
 mutate(obs=1)

##Word Frequency Analysis: Calculating the number of female-related words in texts with high levels of emotive rhetoric  
tokens\_emo$womword <- as.integer(grepl("women|woman|female|girl|feminist|feminism|feminine|gender|harassment|sexism|sexist|lady|madam|womenlike|ladylike", x = tokens\_emo$word))

##Check the dataset  
head(tokens\_emo)

## period time female party emotive\_count neutral\_count  
## 1: 2001-2005 01/1 0 Conservative 52 33  
## 2: 2001-2005 01/1 0 Conservative 52 33  
## 3: 2001-2005 01/1 0 Conservative 52 33  
## 4: 2001-2005 01/1 0 Conservative 52 33  
## 5: 2001-2005 01/1 0 Conservative 52 33  
## 6: 2001-2005 01/1 0 Conservative 52 33  
## emotive\_rhetoric emotive\_positive emotive\_negative word obs womword  
## 1: 5 8.684211 0.7894737 beg 1 0  
## 2: 5 8.684211 0.7894737 move 1 0  
## 3: 5 8.684211 0.7894737 michael 1 0  
## 4: 5 8.684211 0.7894737 martin 1 0  
## 5: 5 8.684211 0.7894737 chair 1 0  
## 6: 5 8.684211 0.7894737 house 1 0

##Get the number of women-related words in these texts with stronger emotive rhetoric through "filter" function  
 emo\_women\_counts <-tokens\_emo %>%   
 filter(womword==1)  
 ##Chect the obs of new dataset (especially the number of obs)  
 head(emo\_women\_counts)

## period time female party emotive\_count neutral\_count emotive\_rhetoric  
## 1: 2001-2005 01/1 0 Labour 165 73 8.803828  
## 2: 2001-2005 01/1 0 Labour 165 73 8.803828  
## 3: 2001-2005 01/1 0 Labour 165 73 8.803828  
## 4: 2001-2005 01/1 0 Labour 89 54 5.185185  
## 5: 2001-2005 01/1 0 Labour 89 54 5.185185  
## 6: 2001-2005 01/1 0 Labour 89 54 5.185185  
## emotive\_positive emotive\_negative word obs womword  
## 1: 8.325359 1.2440191 women 1 1  
## 2: 8.325359 1.2440191 women 1 1  
## 3: 8.325359 1.2440191 woman 1 1  
## 4: 8.296296 0.5925926 women 1 1  
## 5: 8.296296 0.5925926 women 1 1  
## 6: 8.296296 0.5925926 lady 1 1

##There are 120404 obs in the new dataset

## Word Frequency of Female-related Words in Texts Using More Neutral Words  
##Tidy the text  
tokens\_neu <- data\_neutral %>%  
 unnest\_tokens(word, text)  
##Check the dataset  
head(tokens\_neu)

## period time female party emotive\_count neutral\_count emotive\_rhetoric  
## 1: 2001-2005 01/1 0 Labour 0 1 -6.666667  
## 2: 2001-2005 01/1 0 Labour 0 1 -6.666667  
## 3: 2001-2005 01/1 0 Labour 0 1 -6.666667  
## 4: 2001-2005 01/1 0 Labour 0 1 -6.666667  
## 5: 2001-2005 01/1 0 Labour 0 1 -6.666667  
## 6: 2001-2005 01/1 0 Labour 0 1 -6.666667  
## word  
## 1: under  
## 2: the  
## 3: provisions  
## 4: of  
## 5: standing  
## 6: order

## remove all stop words（the original dataset has removed all apostrophes)  
tokens\_neu<- tokens\_neu %>%  
 filter(!word %in% stop\_words$word)

##Assign a value of 1 to each observation for easy counting of quantities.  
tokens\_neu <- tokens\_neu %>%  
 mutate(obs=1)

##Word Frequency Analysis: Calculating the number of female-related words in texts with equal or more neutral words  
tokens\_neu$womword <- as.integer(grepl("women|woman|female|girl|feminist|feminism|feminine|gender|harassment|sexism|sexist|lady|madam|womenlike|ladylike", x = tokens\_neu$word))

##Check the dataset  
head(tokens\_neu)

## period time female party emotive\_count neutral\_count emotive\_rhetoric  
## 1: 2001-2005 01/1 0 Labour 0 1 -6.666667  
## 2: 2001-2005 01/1 0 Labour 0 1 -6.666667  
## 3: 2001-2005 01/1 0 Labour 0 1 -6.666667  
## 4: 2001-2005 01/1 0 Labour 0 1 -6.666667  
## 5: 2001-2005 01/1 0 Labour 0 1 -6.666667  
## 6: 2001-2005 01/1 0 Labour 0 1 -6.666667  
## word obs womword  
## 1: provisions 1 0  
## 2: standing 1 0  
## 3: 1a 1 0  
## 4: required 1 0  
## 5: ascertain 1 0  
## 6: michael 1 0

##Get the number of women-related words in these texts with equal or more neutral words through "filter" function  
 neu\_women\_counts <-tokens\_neu %>%   
 filter(womword==1)  
 ##Chect the obs of new dataset (especially the number of obs)  
 head(neu\_women\_counts)

## period time female party emotive\_count neutral\_count  
## 1: 2001-2005 01/1 0 Conservative 86 113  
## 2: 2001-2005 01/1 0 Conservative 29 30  
## 3: 2001-2005 01/1 0 Labour 120 165  
## 4: 2001-2005 01/1 0 Liberal Democrats 154 162  
## 5: 2001-2005 01/1 0 Labour 34 49  
## 6: 2001-2005 01/1 0 Conservative 43 56  
## emotive\_rhetoric word obs womword  
## 1: -2.4953789 women 1 1  
## 2: -0.2832861 lady 1 1  
## 3: -2.8481013 women 1 1  
## 4: -0.5069708 women 1 1  
## 5: -3.4562212 lady 1 1  
## 6: -2.5341131 lady 1 1

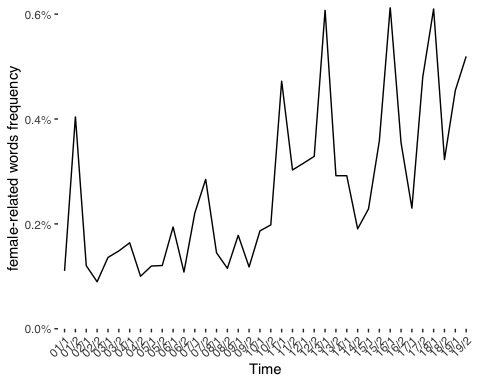
##There are 53371 obs in the new dataset

##Time Trends  
##Exploring Temporal Trends in Female-Related Words in all sampled data  
women\_counts\_time<- tokens\_all %>%  
 group\_by(time) %>%  
 mutate(time\_total = sum(obs)) %>% ## "n" in the original data is closure, so I decided to sum obs.  
 filter(womword==1) %>%  
 summarise(sum\_wom = sum(obs),  
 time\_total= min(time\_total))

head(women\_counts\_time)

## # A tibble: 6 × 3  
## time sum\_wom time\_total  
## <chr> <dbl> <dbl>  
## 1 01/1 2 1812  
## 2 01/2 48 11885  
## 3 02/1 25 20784  
## 4 02/2 13 14504  
## 5 03/1 33 24266  
## 6 03/2 15 10107

##plot the results to show the trend of time  
ggplot(women\_counts\_time, aes(time, sum\_wom / time\_total, group=1)) +  
 geom\_line() +  
 xlab("Time") +  
 ylab("female-related words frequency") +  
 scale\_y\_continuous(labels = scales::percent\_format(),  
 expand = c(0, 0), limits = c(0, NA)) +  
 theme\_tufte(base\_family = "Helvetica") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) # Rotate x-axis labels by 45 degrees

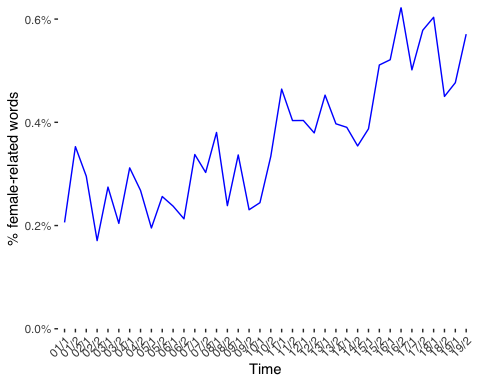


##Exploring Temporal Trends in Female-Related Vocabulary in Texts with Strong Levels of Emotive Rhetoric  
tokens\_emo\_time <- tokens\_emo %>%  
 group\_by(time) %>%  
 mutate(time\_total = sum(obs)) %>% ## "n" in the original data is closure, so I decided to sum obs.  
 filter(womword==1) %>%  
 summarise(sum\_wom = sum(obs),  
 time\_total= min(time\_total))

head(tokens\_emo\_time)

## # A tibble: 6 × 3  
## time sum\_wom time\_total  
## <chr> <dbl> <dbl>  
## 1 01/1 224 108723  
## 2 01/2 2541 720184  
## 3 02/1 2888 977096  
## 4 02/2 1107 648175  
## 5 03/1 2610 950789  
## 6 03/2 1256 615212

##plot the results to show the trend of time  
ggplot(tokens\_emo\_time, aes(time, sum\_wom / time\_total, group=1)) +  
 geom\_line(color = "blue") +  
 xlab("Time") +  
 ylab("% female-related words") +  
 scale\_y\_continuous(labels = scales::percent\_format(),  
 expand = c(0, 0), limits = c(0, NA)) +  
 theme\_tufte(base\_family = "Helvetica") +  
theme(axis.text.x = element\_text(angle = 45, hjust = 1)) # Rotate x-axis labels by 45 degrees

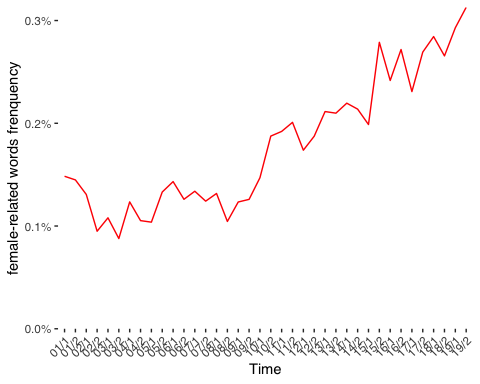


##Exploring Temporal Trends in Female-Related Vocabulary in Texts with Emotive Rhetoric Scores less than 0  
tokens\_neu\_time <- tokens\_neu %>%  
 group\_by(time) %>%  
 mutate(time\_total = sum(obs)) %>%  
 filter(womword==1) %>%  
 summarise(sum\_wom = sum(obs),  
 time\_total= min(time\_total))

##Check the dataset  
head(tokens\_neu\_time)

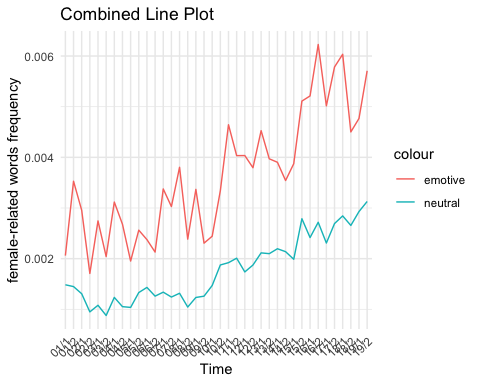
## # A tibble: 6 × 3  
## time sum\_wom time\_total  
## <chr> <dbl> <dbl>  
## 1 01/1 83 55920  
## 2 01/2 948 654350  
## 3 02/1 1349 1031759  
## 4 02/2 646 680489  
## 5 03/1 1216 1126207  
## 6 03/2 617 702776

##plot the results to show the trend of time  
ggplot(tokens\_neu\_time, aes(time, sum\_wom / time\_total, group=1)) +  
 geom\_line(color = "red") +  
 xlab("Time") +  
 ylab("female-related words frenquency") +  
 scale\_y\_continuous(labels = scales::percent\_format(),  
 expand = c(0, 0), limits = c(0, NA)) +  
 theme\_tufte(base\_family = "Helvetica") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) # Rotate x-axis labels by 45 degrees



##Combine two plots into one for the easier comparison  
combined\_plot <- ggplot() +  
 geom\_line(data = tokens\_emo\_time, aes(time, sum\_wom / time\_total, group = 1, color = "emotive")) +  
 geom\_line(data = tokens\_neu\_time, aes(time, sum\_wom / time\_total, group = 1, color = "neutral")) +  
 labs(title = "Combined Line Plot", x = "Time", y = "female-related words frequency")+  
 theme\_minimal()+  
theme(axis.text.x = element\_text(angle = 45, hjust = 1))

##Check the plot  
print(combined\_plot)

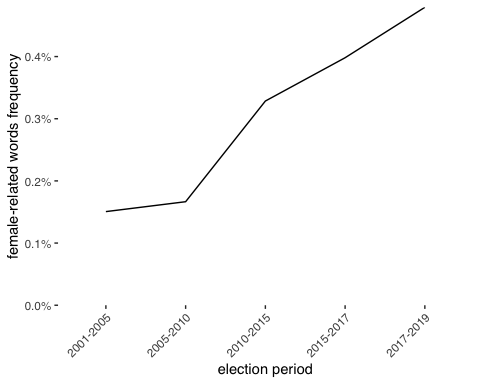


##Trends across Five Election Periods  
##Exploring Five Election Trends of Female-Related words in all sampled data  
women\_counts\_period<- tokens\_all %>%  
 group\_by(period) %>%  
 mutate(period\_total = sum(obs)) %>% ## "n" in the original data is closure, so I decided to sum obs.  
 filter(womword==1) %>%  
 summarise(sum\_wom = sum(obs),  
 period\_total= min(period\_total))

##Check the Data  
head(women\_counts\_period)

## # A tibble: 5 × 3  
## period sum\_wom period\_total  
## <chr> <dbl> <dbl>  
## 1 2001-2005 200 132786  
## 2 2005-2010 275 165137  
## 3 2010-2015 576 175360  
## 4 2015-2017 307 77110  
## 5 2017-2019 430 89727

##plot the results to show the trend of period  
ggplot(women\_counts\_period, aes(period, sum\_wom / period\_total, group=1)) +  
 geom\_line() +  
 xlab("election period") +  
 ylab("female-related words frequency") +  
 scale\_y\_continuous(labels = scales::percent\_format(),  
 expand = c(0, 0), limits = c(0, NA)) +  
 theme\_tufte(base\_family = "Helvetica") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) # Rotate x-axis labels by 45 degrees

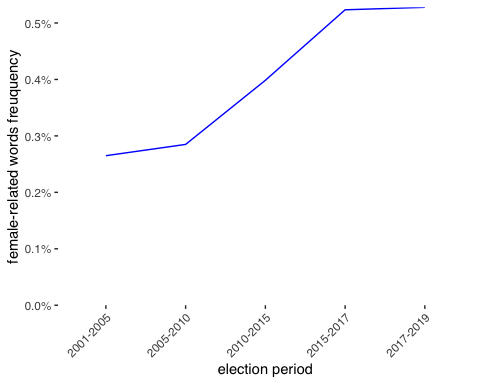


#Exploring Five Election Trends of Female-Related words in texts with high level of emotive rhetoric  
emo\_counts\_period<- tokens\_emo %>%  
 group\_by(period) %>%  
 mutate(period\_total = sum(obs)) %>% ## "n" in the original data is closure, so I decided to sum obs.  
 filter(womword==1) %>%  
 summarise(sum\_wom = sum(obs),  
 period\_total= min(period\_total))

##Check the Data  
head(emo\_counts\_period)

## # A tibble: 5 × 3  
## period sum\_wom period\_total  
## <chr> <dbl> <dbl>  
## 1 2001-2005 16331 6164292  
## 2 2005-2010 21909 7689555  
## 3 2010-2015 33968 8523533  
## 4 2015-2017 20947 4002135  
## 5 2017-2019 27249 5162697

##plot the results to show the trend of period  
ggplot(emo\_counts\_period, aes(period, sum\_wom / period\_total, group=1)) +  
 geom\_line(color = "blue") +  
 xlab("election period") +  
 ylab("female-related words freuquency") +  
 scale\_y\_continuous(labels = scales::percent\_format(),  
 expand = c(0, 0), limits = c(0, NA)) +  
 theme\_tufte(base\_family = "Helvetica") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) # Rotate x-axis labels by 45 degrees

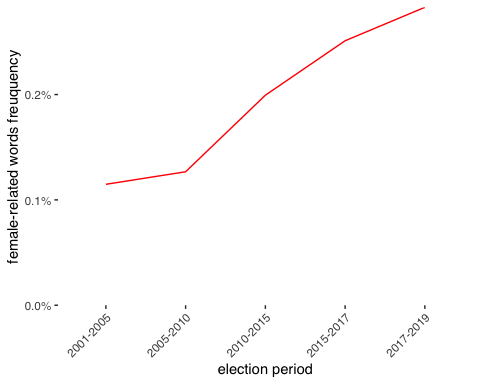


###Exploring Five Election Trends of Female-Related words in texts with more neutral words  
neu\_counts\_period <- tokens\_neu %>%  
 group\_by(period) %>%  
 mutate(period\_total = sum(obs)) %>% ## "n" in the original data is closure, so I decided to sum obs.  
 filter(womword==1) %>%  
 summarise(sum\_wom = sum(obs),  
 period\_total= min(period\_total))

head(neu\_counts\_period)

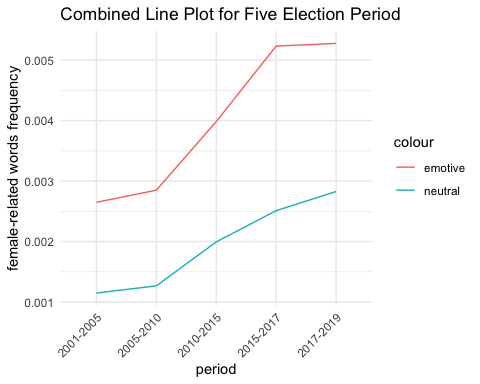
## # A tibble: 5 × 3  
## period sum\_wom period\_total  
## <chr> <dbl> <dbl>  
## 1 2001-2005 7661 6673172  
## 2 2005-2010 10615 8377331  
## 3 2010-2015 16895 8473723  
## 4 2015-2017 7728 3077927  
## 5 2017-2019 10472 3702604

##plot the results to show the trend of period  
ggplot(neu\_counts\_period, aes(period, sum\_wom / period\_total, group=1)) +  
 geom\_line(color = "red") +  
 xlab("election period") +  
 ylab("female-related words freuquency") +  
 scale\_y\_continuous(labels = scales::percent\_format(),  
 expand = c(0, 0), limits = c(0, NA)) +  
 theme\_tufte(base\_family = "Helvetica") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) # Rotate x-axis labels by 45 degrees



##Combine two plots into one for the easier comparison  
combined\_plot\_period <- ggplot() +  
 geom\_line(data = emo\_counts\_period, aes(period, sum\_wom / period\_total, group = 1, color = "emotive")) +  
 geom\_line(data = neu\_counts\_period, aes(period, sum\_wom / period\_total, group = 1, color = "neutral")) +  
 labs(title = "Combined Line Plot for Five Election Period", x = "period", y = "female-related words frequency")+  
 theme\_minimal()+  
theme(axis.text.x = element\_text(angle = 45, hjust = 1))

##Check the plot  
print(combined\_plot\_period)

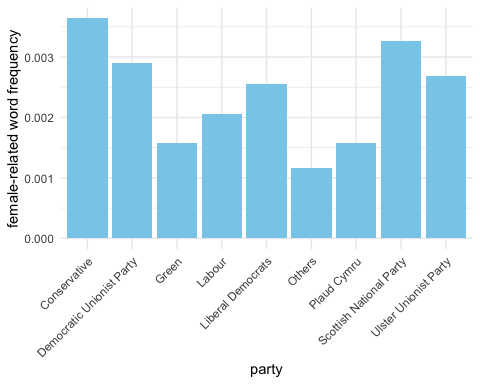


## Match female-related word frequency in all sampled texts and different party   
party\_counts <- tokens\_all %>%  
 group\_by(party) %>%  
 mutate(party\_total = sum(obs)) %>%  
 filter(womword==1) %>%  
 summarise(sum\_wom = sum(obs),  
 party\_total= min(party\_total))

##Check the dataset  
head(party\_counts)

## # A tibble: 6 × 3  
## party sum\_wom party\_total  
## <chr> <dbl> <dbl>  
## 1 Conservative 960 263680  
## 2 Democratic Unionist Party 21 7235  
## 3 Green 1 634  
## 4 Labour 574 278347  
## 5 Liberal Democrats 150 58725  
## 6 Others 8 6917

##plot the results  
ggplot(party\_counts, aes(x = party, y = sum\_wom / party\_total)) +  
 geom\_bar(stat = "identity", fill = "skyblue") + # Create bars  
 labs(x = "party", y = "female-related word frequency") + # Labels  
 theme\_minimal() + # Minimal theme  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) # Rotate x-axis labels by 45 degrees

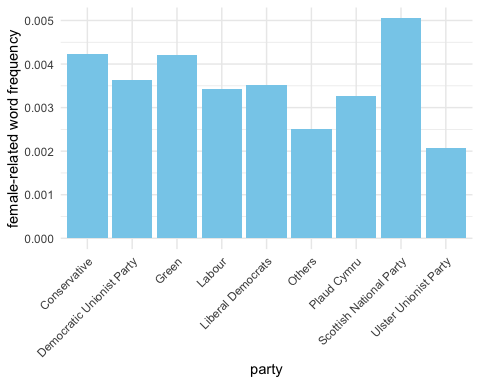


## Match female-related word frequency in texts which have higer level of emotive rhetoric and party in UK  
emo\_party\_counts <- tokens\_emo %>%  
 group\_by(party) %>%  
 mutate(party\_total = sum(obs)) %>%  
 filter(womword==1) %>%  
 summarise(sum\_wom = sum(obs),  
 party\_total= min(party\_total))

head(emo\_party\_counts)

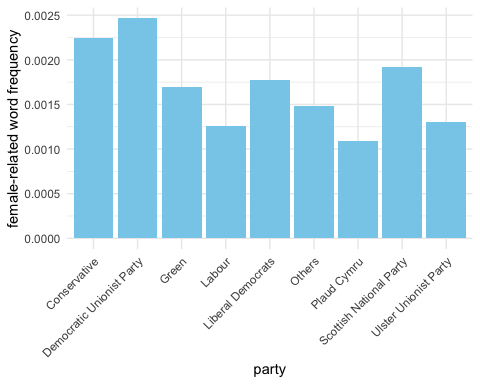
## # A tibble: 6 × 3  
## party sum\_wom party\_total  
## <chr> <dbl> <dbl>  
## 1 Conservative 55780 13157338  
## 2 Democratic Unionist Party 1748 482498  
## 3 Green 261 62110  
## 4 Labour 47588 13864732  
## 5 Liberal Democrats 8284 2355660  
## 6 Others 915 364897

##plot the results  
ggplot(emo\_party\_counts, aes(x = party, y = sum\_wom / party\_total)) +  
 geom\_bar(stat = "identity", fill = "skyblue") + # Create bars  
 labs(x = "party", y = "female-related word frequency") + # Labels  
 theme\_minimal() + # Minimal theme  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) # Rotate x-axis labels by 45 degrees party\_total= min(party\_total))



## Match female-related word frequency in texts which have more neutral words and different party  
neu\_party\_counts <- tokens\_neu %>%  
 group\_by(party) %>%  
 mutate(party\_total = sum(obs)) %>%  
 filter(womword==1) %>%  
 summarise(sum\_wom = sum(obs),  
 party\_total= min(party\_total))

##plot the results  
ggplot(neu\_party\_counts, aes(x = party, y = sum\_wom / party\_total)) +  
 geom\_bar(stat = "identity", fill = "skyblue") + # Create bars  
 labs(x = "party", y = "female-related word frequency") + # Labels  
 theme\_minimal() + # Minimal theme  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) # Rotate x-axis labels by 45 degrees



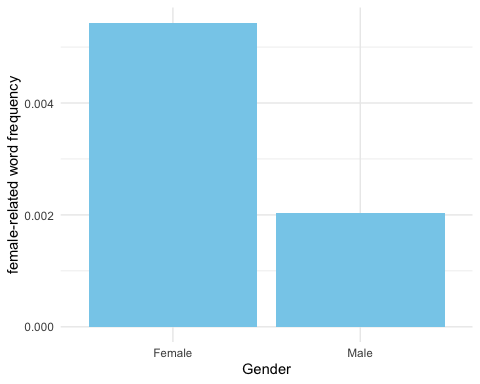
## Match female-related word frequency in all sampled texts rhetoric and gender of speakers  
gender\_counts <- tokens\_all %>%  
 group\_by(female) %>%  
 mutate(gender\_total = sum(obs)) %>%  
 filter(womword==1) %>%  
 summarise(sum\_wom = sum(obs),gender\_total= min(gender\_total))

##Make the continuous variable “female”, from “0” and “1” to the categorical variables “male” and “female”  
gender\_counts <- gender\_counts %>%   
 mutate(Gender = c("Male", "Female"),  
 female = c(0, 1))

#Check the dataset  
head(gender\_counts)

## # A tibble: 2 × 4  
## female sum\_wom gender\_total Gender  
## <dbl> <dbl> <dbl> <chr>   
## 1 0 1012 497435 Male   
## 2 1 776 142685 Female

##plot the results  
ggplot(gender\_counts, aes(x = Gender, y = sum\_wom / gender\_total)) +  
 geom\_bar(stat = "identity", fill = "skyblue") + # Create bars  
 labs(x = "Gender", y = "female-related word frequency") + # Labels  
 theme\_minimal()



## Match female-related word frequency in texts wich have higer level of emotive rhetoric and gender of speakers  
emo\_gender\_counts <- tokens\_emo %>%  
 group\_by(female) %>%  
 mutate(gender\_total = sum(obs)) %>%  
 filter(womword==1) %>%  
 summarise(sum\_wom = sum(obs),gender\_total= min(gender\_total))

head(emo\_gender\_counts)

## # A tibble: 2 × 3  
## female sum\_wom gender\_total  
## <int> <dbl> <dbl>  
## 1 0 63526 23586722  
## 2 1 56878 7955490

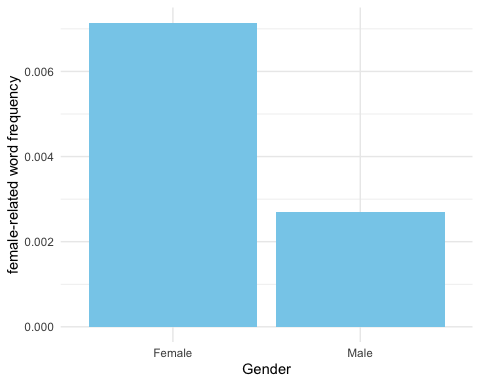
## "0" means male and "1" means female

##Make the continuous variable “female”, from “0” and “1” to the categorical variables “male” and “female”.  
emo\_gender\_counts <- emo\_gender\_counts %>%   
 mutate(Gender = c("Male", "Female"),  
 female = c(0, 1))

##Check the dataset  
head(emo\_gender\_counts)

## # A tibble: 2 × 4  
## female sum\_wom gender\_total Gender  
## <dbl> <dbl> <dbl> <chr>   
## 1 0 63526 23586722 Male   
## 2 1 56878 7955490 Female

##plot the results  
ggplot(emo\_gender\_counts, aes(x = Gender, y = sum\_wom / gender\_total)) +  
 geom\_bar(stat = "identity", fill = "skyblue") + # Create bars  
 labs(x = "Gender", y = "female-related word frequency") + # Labels  
 theme\_minimal()



## Match female-related word frequency in texts which have emotive rhetoric scores less than 0 and the gender of speakers  
neu\_gender\_counts <- tokens\_neu %>%  
 group\_by(female) %>%  
 mutate(gender\_total = sum(obs)) %>%  
 filter(womword==1) %>%  
 summarise(sum\_wom = sum(obs),gender\_total= min(gender\_total))

##Make the continuous variable “female”, from “0” and “1” to the categorical variables “male” and “female”.  
neu\_gender\_counts <- neu\_gender\_counts %>%   
 mutate(Gender = c("Male", "Female"),  
 female = c(0, 1))

##Check the dataset  
head(neu\_gender\_counts)

## # A tibble: 2 × 4  
## female sum\_wom gender\_total Gender  
## <dbl> <dbl> <dbl> <chr>   
## 1 0 39322 24839008 Male   
## 2 1 14049 5465749 Female

##plot the results  
ggplot(neu\_gender\_counts, aes(x = Gender, y = sum\_wom / gender\_total)) +  
 geom\_bar(stat = "identity", fill = "skyblue") + # Create bars  
 labs(x = "Gender", y = "female-related word frequency") + # Labels  
 theme\_minimal()

