**########################## tidytext sentiment analysis ###############################**

**library(tidytext)**

**library(janeaustenr)**

**library(dplyr)**

**library(stringr)**

**library(tidyr)**

**library(ggplot2)**

**########################## three sentiment dictionaries ##############################**

**bing<- get\_sentiments("bing")**

**bing**

**bing[1:20,]**

**nrc<- get\_sentiments("nrc")**

**nrc**

**nrc[1:20,]**

**nrc\_joy<- get\_sentiments("nrc") %>%**

 **filter(sentiment == "joy")**

**nrc\_joy**

**nrc\_joy[1:20,]**

**afinn<- get\_sentiments("afinn")**

**afinn**

**afinn[1:20,]**

**########################## three sentiment dictionaries ##############################**

**###################### austen\_books from library(janeaustenr) ########################**

**## tidy\_books: creating a table with a long column of words to be used for sentiment analysis**

**## includes the name of the book, the line number of the book, and the chapter of the book**

**tidy\_books <- austen\_books() %>%**

 **group\_by(book) %>%**

 **mutate(linenumber = row\_number(),**

 **chapter = cumsum(str\_detect(text, regex("^chapter [\\divxlc]",**

 **ignore\_case = TRUE)))) %>%**

 **ungroup() %>%**

 **unnest\_tokens(word, text)**

**dim(tidy\_books)**

**tidy\_books[1:20,]**

**tidy\_books[725001:725029,]**

**dim(tidy\_books)**

**table(tidy\_books[,1])**

**## considering a subset of tidy\_books (Emma)**

**emma=tidy\_books %>%**

 **filter(book == "Emma")**

**emma**

**## obtaining (bing) sentiment information for the book emma**

**emma\_sentiment=emma %>%**

 **inner\_join(get\_sentiments("bing")) %>%**

 **count(book, sentiment) %>%**

 **spread(sentiment, n, fill = 0) %>%**

 **mutate(sentiment = positive - negative)**

**emma\_sentiment ## frequencies of (bing) sentiment words**

**## obtaining (nrc) sentiment information for the book emma**

**emma\_sentiment=emma %>%**

 **inner\_join(nrc) %>%**

 **count(book, sentiment) %>%**

 **spread(sentiment, n, fill = 0) %>%**

 **mutate(sentiment = positive - negative)**

**emma\_sentiment ## frequencies of (nrc) sentiment words**

**## obtaining (bing) sentiment information for all six jane austen books**

**austen\_sentiment=tidy\_books %>%**

 **inner\_join(get\_sentiments("bing")) %>%**

 **count(book, sentiment) %>%**

 **spread(sentiment, n, fill = 0) %>%**

 **mutate(sentiment = positive - negative)**

**austen\_sentiment ## frequencies of (bing) sentiment words**

**## obtaining (bing) sentiment information for all six jane austen books**

**## here with the proportion of positive sentiment words**

**austen\_sentiment=tidy\_books %>%**

 **inner\_join(get\_sentiments("bing")) %>%**

 **count(book, sentiment) %>%**

 **spread(sentiment, n, fill = 0) %>%**

 **mutate(sentiment = 100\*positive/(positive+negative))**

**austen\_sentiment ## frequencies of (bing) sentiment words**

**## obtaining (nrc) sentiment information for all six jane austen books**

**austen\_sentiment=tidy\_books %>%**

 **inner\_join(get\_sentiments("nrc")) %>%**

 **count(book, sentiment) %>%**

 **spread(sentiment, n, fill = 0) %>%**

 **mutate(sentiment = positive - negative)**

**austen\_sentiment ## frequencies of (nrc\_joy) sentiment words**

**## using sentiment dictionary (bing) for all six jane austen books**

**## plotting results for blocks of 80 lines**

**jane\_austen\_sentiment <- tidy\_books %>%**

 **inner\_join(get\_sentiments("bing")) %>%**

 **count(book, index = linenumber %/% 80, sentiment) %>%**

 **spread(sentiment, n, fill = 0) %>%**

 **mutate(sentiment = positive - negative)**

**jane\_austen\_sentiment**

**dim(jane\_austen\_sentiment)**

**## for blocks of 80 text lines, determines the number of negative and positive sentiment word hits**

**## calculates the difference: positive minus negative word counts**

**## there are 920 such blocks**

**## sequence plotting of the difference (positive sentiment if above the line)**

**ggplot(jane\_austen\_sentiment, aes(index, sentiment, fill = book)) +**

 **geom\_col(show.legend = FALSE) +**

 **facet\_wrap(~book, ncol = 2, scales = "free\_x")**

**## using sentiment dictionary (nrc) for all six jane austen books**

**## plotting results for blocks of 80 lines**

**jane\_austen\_sentiment <- tidy\_books %>%**

 **inner\_join(get\_sentiments("nrc")) %>%**

 **count(book, index = linenumber %/% 80, sentiment) %>%**

 **spread(sentiment, n, fill = 0) %>%**

 **mutate(sentiment = positive - negative)**

**jane\_austen\_sentiment**

**dim(jane\_austen\_sentiment)**

**## for blocks of 80 text lines, determines the number of negative and positive sentiment word hits**

**## calculates the difference: positive minus negative word counts**

**## there are 920 such blocks**

**## sequence plotting of the difference (positive sentiment if above the line)**

**ggplot(jane\_austen\_sentiment, aes(index, sentiment, fill = book)) +**

 **geom\_col(show.legend = FALSE) +**

 **facet\_wrap(~book, ncol = 2, scales = "free\_x")**

**## using sentiment dictionary (nrc, with joyful terms) for all six jane austen books**

**## plotting results for blocks of 80 lines**

**jane\_austen\_sentiment = tidy\_books %>%**

 **inner\_join(get\_sentiments("nrc")%>%**

 **filter(sentiment =="joy")) %>%**

 **count(book, index = linenumber %/% 80, sentiment) %>%**

 **spread(sentiment, n, fill = 0)**

**jane\_austen\_sentiment**

**dim(jane\_austen\_sentiment)**

**## for blocks of 80 text lines, determines the number of negative and positive sentiment word hits**

**## there are 920 such blocks**

**## sequence plotting of the difference (number of joyful words)**

**ggplot(jane\_austen\_sentiment, aes(index, joy, fill = book)) +**

 **geom\_col(show.legend = FALSE) +**

 **facet\_wrap(~book, ncol = 2, scales = "free\_x")**

**###################### austen\_books from library(janeaustenr) ########################**

**################################ sentiment analysis of letter 1 #################################**

**## loading the first letter**

**carr="Dear Sir. The 13 of April I left Lexington for this country; at which time there had nothing certain concerning the government of this district been received from Congress. However I supposed that by the time I arrived here, the measures adopted by the general government would be known but this was not the case. On my arrival I found the people very much agitated concerning various reports in circulation here at that time, particularly in regard to this district being annexed to the Indiana territory and the regulations which Congress might adopt relative to slavery. A native of Virginia, and the second American lawyer to arrive in StLouis. Many were apprehensive that slavery would not only be prohibited, but the more ignorant were fearful least those already in their possession would also be manumitted. I discern from the Law, or that part of it which relates to this district, that Congress has been silent on the subject although it has been permitted in the territory of Orleans under certain restrictions. The French are reserved on the subject of the late change, but if I am not much mistaken many are scarcely disaffected towards it. I have the pleasure however to inform you that the disaffected part of the citizens are the most illiterate and ignorant, and that class of individuals which are generally kept in profound ignorance under Monarchical governments. I am ready to hide my face from my injured unsettled country where truth compels me to inform you that the French here are much better citizens than the Americans. I am ready to blush when I say that the French in this country according to numbers are much more pleased with the change than the Americans. But those vile wretches those Americans who have escaped to this country I believe to have been villains and fugitives from the justice of their country. Independent of the original turpitude of their characters, not being acquainted with the administration of justice and unwilling to make use of the same means to obtain their rights which the French did they became disgusted with the legal authority, and to revenge themselves of their adversaries gave a loose to the most malicious and evil passions which disgrace human nature. You have no idea how extremely unpopular particularly (with the Americans) some of the former commandants are. Several I could mention who will remain here a single day after the new system of government takes place at the very peril of their lives. The corruption of these Spanish Officers has been astonishing indeed. Their thirst for money has not been less than the most mercenary despot of Tripoli or Morocco, nor has it been less fed in proportion to the wealth of the parties. The most glaring injustice has been committed with impunity under sanction of the laws and the litigant who was able to pay the largest bribes to the commandant generally succeeded. They are not only threatened with numerous actions for civil injuries but the most horrid punishments are denounced against them. But I think it will be extremely difficult handling these Petty Kings in consequence of the undefined limits of the power with which they were vested. This certainly will or ought to be the place at which the Governor and Judges should sit for the purpose of legislation: it is also the only place which completely answers the requisitions of the act of Congress, for the sessions of the Supreme Court. If this takes place I shall certainly make this my place of residence. Some time since the people of StLouis had a convention which truly in a republican manner elected a committee of seven persons, to whom they gave it in charge to draw up and adopt such resolutions as might be most conducive to the interest of this District. This committee drew up an address which was sent to various parts advising the people to imitate their example: and after such resolutions were adopted, that there should be one person (to be chosen by a majority of votes) for every 700 inhabitants; who should convene at such place the Government should sit and hold themselves in readiness to offer him any information he might require. After having witnessed this Republican conduct and the peaceable quiet demeanor of the French in this country, how can it be argued that the Louisianians are not prepared for the reception of a Republican government? I trust sir we shall by our tranquil deportment manifest ourselves not only sensible of the blessings of liberty, but ultimately prove ourselves the best citizens in the United States. I hope sir you will favor me with a letter as soon as convenient and accept assurances of respect from dear Sir. Your most humble servant."**

**carr=tolower(carr)**

**carr=gsub("[.]","", ignore.case = TRUE,carr)**

**carr=gsub("[,]","", ignore.case = TRUE,carr)**

**carr=gsub("[;]","", ignore.case = TRUE,carr)**

**carr**

**## using tidytext**

**carrtb <- tibble(text = carr)**

**carrtb**

**tidy\_car <- carrtb %>%**

 **unnest\_tokens(word, text)**

**tidy\_car**

**tidy\_car %>%**

 **inner\_join(get\_sentiments("bing")) %>%**

 **count(sentiment) %>%**

 **spread(sentiment, n, fill = 0) %>%**

 **mutate(sentiment = positive / (positive + negative))**

**## using the document term matrix**

**library(tm) ## for creating the corpus and the document term matrix**

**corpus <- VCorpus(VectorSource(carr),readerControl = list(reader = readPlain))**

**corpus1 <- tm\_map(corpus, stripWhitespace)**

**corpus2 <- tm\_map(corpus1, content\_transformer(tolower))**

**corpus3 <- tm\_map(corpus2, removePunctuation)**

**corpus4 <- tm\_map(corpus3, removeNumbers)**

**corpus5 <- tm\_map(corpus4, removeWords, stopwords("english"))**

**corpcarr.dtm <- DocumentTermMatrix(corpus5,control=list(stemming=FALSE))**

**dim(corpcarr.dtm)**

**lettercarr=tidy(corpcarr.dtm)**

**lettercarr**

**colnames(lettercarr)**

**colnames(lettercarr)[colnames(lettercarr)=="term"]="word"**

**colnames(lettercarr)**

**lettercarr**

**## obtaining (bing) sentiment information for lettercarr**

**lettercarr %>%**

 **inner\_join(get\_sentiments("bing")) %>%**

 **count(document,wt=count,sentiment) %>%**

 **spread(sentiment, n, fill = 0) %>%**

 **mutate(sentiment = 100\*positive/(positive + negative))**

**################################ sentiment analysis of letter 1 #################################**

**################################ sentiment analysis of letter 2 #################################**

**## loading the second letter**

**secwar="Sir. The President of the United States having appointed you an Agent of Indian Affairs for upper Louisiana. You will consider it your duty to use all prudent and possible means in your power for conciliating and establishing the friendship and good will of the several Indian Nations and tribes residing in that part of Louisiana which lies to the Northward and Westward of the river Arkansas. You will be particularly attentive to our friends the Osage nation. You will endeavor to prevent all hostile acts or dispositions in the several nations or tribes, toward any of the white Citizens or inhabitants of the United States or towards any of their red brethren. You will endeavor to cultivate a pacific and friendly disposition generally among all the red people who may be placed under your care and super intendance and inculcate upon their minds the propriety and usefulness of a gradual introduction of the arts of civilization, especially the arts of agriculture and domestic manufactures, for which purpose you will from time to time furnish such of them with ploughs, hoes axes, spinning wheels, looms and other necessary materials, as will engage to make use of them, you will procure a suitable character as a blacksmith for the Osage Indians, and furnish him with proper tools and such quantities of iron and steel as may from time to time be found necessary you will take the earliest opportunity for healing the breach between the Osage nation and the party under big track, and endeavor to prevail on the latter with his partisans to return to the nation and to live in harmony. You will take the necessary measures for obtaining permission of the Big track and his party for the safe passage of any party which may be sent by the President of the United States to explore the sources of the Arkansas river and the interior country generally bordering on the waters of the Red River, the Arkansas and the South western branches of the Missouri. You will also prepare the minds of any other nations or tribes, for receiving as friends any such exploring parties, and for furnishing guides and provisions to such parties when requested so to do; for which they will receive reasonable compensation. You will in no case allow any of the Chiefs or others to set out on a visit to the President of the United States until they shall first have received through you a passport from the President or from the Department of War. You will expend no more money than shall in your judgment be really necessary. You will keep regular accounts of your expenditures and transmit the same to this Office at the end of each quarter of the year with the proper vouchers for Settlement. You will draw bills on this Department for such sums of money as may be necessary for the objects of your agency, taking care at all times to accompany your bills with letters of advice, stating the objects for which the money is required. Your compensation for your services will be 1500 dollars a year. The necessary interpreters will receive one dollar per day. Noel Mongrain and Colik Bolan will receive 365 dollars a year and no other emolument as long as they shall respectively conduct themselves faithfully and be continued in service. You will displace either of them and appoint others in their places whenever you shall judge the public interest requires it. You will correspond with Governor Harrison on all subjects relating to your Agency on which you may need his advice and you will consider his orders as binding on you in all cases, when not incompatible with these instructions or such others as you may hereafter receive from this Department."**

**secwar=tolower(secwar)**

**secwar=gsub("[.]","", ignore.case = TRUE,secwar)**

**secwar=gsub("[,]","", ignore.case = TRUE,secwar)**

**secwar=gsub("[;]","", ignore.case = TRUE,secwar)**

**secwar**

**## using tidytext**

**secwartb <- tibble(text = secwar)**

**secwartb**

**tidy\_secwar <- secwartb %>%**

 **unnest\_tokens(word, text)**

**tidy\_secwar**

**tidy\_secwar %>%**

 **inner\_join(get\_sentiments("bing")) %>%**

 **count(sentiment) %>%**

 **spread(sentiment, n, fill = 0) %>%**

 **mutate(sentiment = positive / (positive + negative))**

**## using the document term matrix**

**library(tm) ## for creating the corpus and the document term matrix**

**corpus <- VCorpus(VectorSource(secwar),readerControl = list(reader = readPlain))**

**corpus1 <- tm\_map(corpus, stripWhitespace)**

**corpus2 <- tm\_map(corpus1, content\_transformer(tolower))**

**corpus3 <- tm\_map(corpus2, removePunctuation)**

**corpus4 <- tm\_map(corpus3, removeNumbers)**

**corpus5 <- tm\_map(corpus4, removeWords, stopwords("english"))**

**corpsecwar.dtm <- DocumentTermMatrix(corpus5,control=list(stemming=FALSE))**

**dim(corpsecwar.dtm)**

**lettersecwar=tidy(corpsecwar.dtm)**

**lettersecwar**

**colnames(lettersecwar)**

**colnames(lettersecwar)[colnames(lettersecwar)=="term"]="word"**

**colnames(lettersecwar)**

**lettersecwar**

**## obtaining (bing) sentiment information for lettersecwar**

**lettersecwar %>%**

 **inner\_join(get\_sentiments("bing")) %>%**

 **count(document,wt=count,sentiment) %>%**

 **spread(sentiment, n, fill = 0) %>%**

 **mutate(sentiment = 100\*positive/(positive + negative))**

**################################ sentiment analysis of letter 2 #################################**

**############################## sentiment analysis of both letters ###############################**

**comb=c(carr,secwar)**

**## using tidytext**

**doc=c(1,2)**

**combtb=tibble(doc,text=comb)**

**combtb**

**tidy\_comb <- combtb %>%**

 **unnest\_tokens(word,text)**

**tidy\_comb**

**tidy\_comb %>%**

 **inner\_join(get\_sentiments("bing")) %>%**

 **count(doc,sentiment) %>%**

 **spread(sentiment, n, fill = 0) %>%**

 **mutate(sentiment = 100\*positive/(positive + negative))**

**## using the document term matrix**

**library(tm) ## for creating the corpus and the document term matrix**

**corpus <- VCorpus(VectorSource(comb),readerControl = list(reader = readPlain))**

**corpus1 <- tm\_map(corpus, stripWhitespace)**

**corpus2 <- tm\_map(corpus1, content\_transformer(tolower))**

**corpus3 <- tm\_map(corpus2, removePunctuation)**

**corpus4 <- tm\_map(corpus3, removeNumbers)**

**corpus5 <- tm\_map(corpus4, removeWords, stopwords("english"))**

**corpcomb.dtm <- DocumentTermMatrix(corpus5,control=list(stemming=FALSE))**

**dim(corpcomb.dtm)**

**lettercomb=tidy(corpcomb.dtm)**

**lettercomb**

**colnames(lettercomb)**

**colnames(lettercomb)[colnames(lettercomb)=="term"]="word"**

**colnames(lettercomb)**

**lettercomb**

**## obtaining (bing) sentiment information for lettercomb**

**lettercomb %>%**

 **inner\_join(get\_sentiments("bing") )%>%**

 **count(document,wt=count,sentiment) %>%**

 **spread(sentiment, n, fill = 0) %>%**

 **mutate(sentiment = 100\*positive/(positive + negative))**

**## COMMENT: The results of both approaches will be the same**

**## COMMENT: There may be small differences due to different treatments of special characters**

**############################## sentiment analysis of both letters ###############################**