1/1/2017

Predict 453

Autism Text Mining

David Pazmino

****

**Introduction**

This document will detail the study of a corpus of 6 dsi’s. Each dsi was selected based on the key words autism and neurodiversity.

**DSI’s**

**DSI 1**

|  |  |
| --- | --- |
| Metatag Type |  |
| Title | What is Neurodiversity? |
| Source | Psychology Today |
| Author(s) | John Elder Robison |
| Published/Uploaded | Oct 07, 2013 |
| URL | [https://www.psychologytoday.com/blog/my-life-aspergers/201310/what-is-neurodiversity](https://www.psychologytoday.com/blog/my-life-aspergers/201310/what-is-neurodiversity ) |
| Total Word Count | 1802 |

**DSI 2**

|  |  |
| --- | --- |
| Metatag Type |  |
| Title | Neurodiversity |
| Source | American Institute for Learning and Development |
| Author(s) | Thomas Armstrong |
| Published/Uploaded | 2011 |
| URL | <http://www.institute4learning.com/resources/articles/neurodiversity/> |
| Total Word Count | 1618 |

**DSI 3**

|  |  |
| --- | --- |
| Metatag Type |  |
| Title | Neurodiversity Rewires Conventional Thinking About Brains |
| Source | WIRED.com |
| Author(s) | Steve Silberman |
| Published/Uploaded | April 16, 2013 |
| URL | [https://www.wired.com/2013/04/neurodiversity/](https://www.wired.com/2013/04/neurodiversity/ ) |
| Total Word Count | 732 |

**DSI 4**

|  |  |
| --- | --- |
| Metatag Type |  |
| Title | Neurodiversity: Some Basic Terms & Definitions |
| Source | neurocosmopolitanism.com |
| Author(s) |  |
| Published/Uploaded | Sep 27, 2014 |
| URL | http://neurocosmopolitanism.com/neurodiversity-some-basic-terms-definitions/ |
| Total Word Count | 1791 |

**DSI 5**

|  |  |
| --- | --- |
| Metatag Type |  |
| Title | Neurodiversity as a Competitive Advantage |
| Source | HBR |
| Author(s) | Robert Austin, Gary Pisano |
| Published/Uploaded | June, 2017 |
| URL | [https://hbr.org/2017/05/neurodiversity-as-a-competitive-advantage](https://hbr.org/2017/05/neurodiversity-as-a-competitive-advantage ) |
| Total Word Count | 4144 |

**DSI 6**

|  |  |
| --- | --- |
| Metatag Type |  |
| Title | The Autism Rights Movement |
| Source | New York |
| Author(s) | [Andrew](http://www.cnn.com/profiles/tal-kopan) Solomon |
| Published/Uploaded | May 25, 2008 |
| URL | [http://nymag.com/news/features/47225/#](http://nymag.com/news/features/47225/) |
| Total Word Count | 5966 |

**EDA**

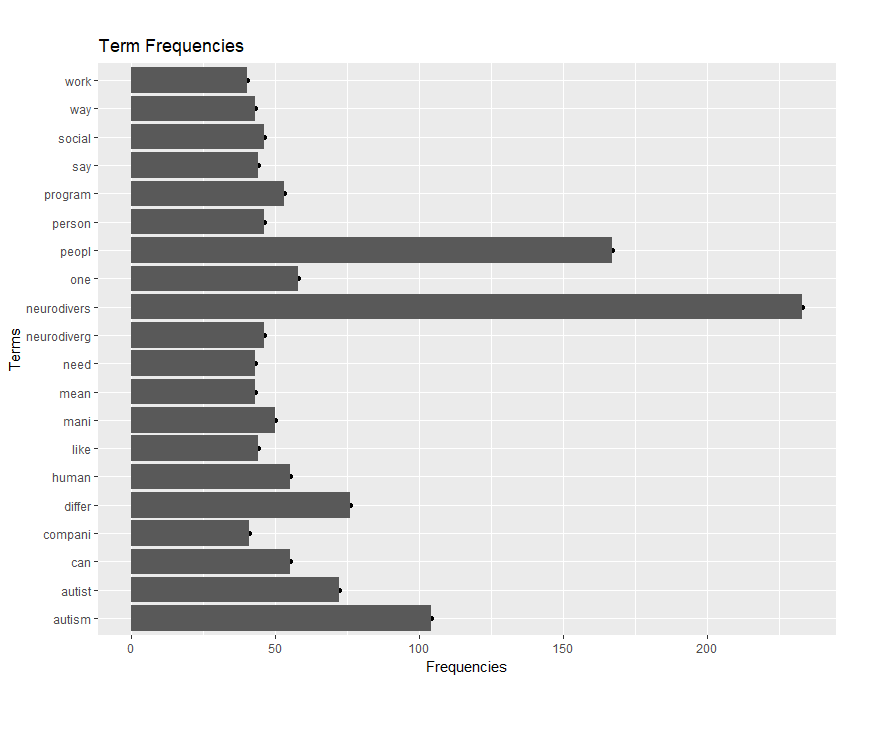
Let’s take a look at the corpus to see if we can see if anything that stands out. Below is the word cloud of our corpus. We can see that the most common words used are neuordiverse, autism and people. The plot below is the machine’s interpretaion of our words. I stemmed the words and removed the basic stop words but the EDA is pretty much the work of the computer algorithms.

****

***Code***

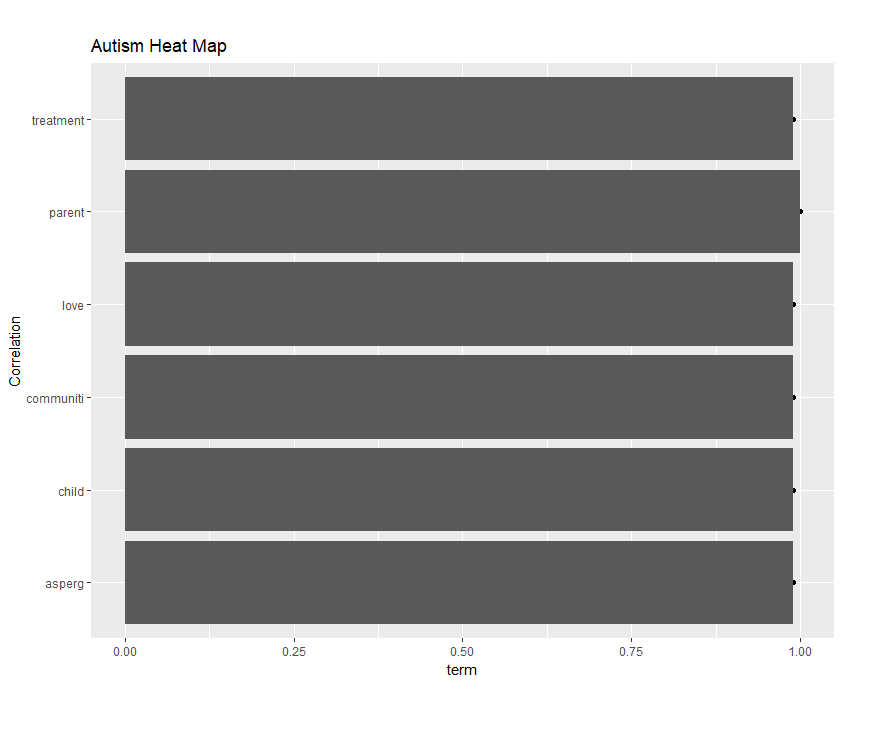
|  |
| --- |
| # Convert the text to lower case  docs <- tm\_map(docs, content\_transformer(tolower))  # Remove numbers  docs <- tm\_map(docs, removeNumbers)  # Remove english common stopwords  docs <- tm\_map(docs, removeWords, stopwords("english"))  # Remove punctuations  docs <- tm\_map(docs, removePunctuation)  # Eliminate extra white spaces  docs <- tm\_map(docs, stripWhitespace)  # Text stemming (reduces words to their root form)  library("SnowballC")  docs <- tm\_map(docs, stemDocument)  # Remove additional stopwords  docs <- tm\_map(docs, removeWords, c("clintonemailcom", "stategov", "hrod"))  library(tm) #load text mining library  setwd('E:/final') #sets R's working directory to near where my files are  a <-Corpus(DirSource("./"), readerControl = list(language="en")) #specifies the exact folder where my text file(s) is for analysis with tm.  summary(a) #check what went in  a <- tm\_map(a, removeNumbers)  a <- tm\_map(a, removePunctuation)  a <- tm\_map(a , stripWhitespace)  a <- tm\_map(a, tolower)  a <- tm\_map(a, removeWords, stopwords("english")) # this stopword file is at C:\Users\[username]\Documents\R\win-library\2.13\tm\stopwords  a <- tm\_map(a, removeWords, c("â€œ", "â€œ"))  a <- tm\_map(a, stemDocument, language = "english")  dtm <- TermDocumentMatrix(a)  m <- as.matrix(dtm)  v <- sort(rowSums(m),decreasing=TRUE)  d <- data.frame(word = names(v),freq=v)  head(d, 15)  library("wordcloud")  library("RColorBrewer")  par(bg="grey30")  wordcloud(d$word, d$freq, col=terrain.colors(length(d$word), alpha=0.9), random.order=FALSE, rot.per=0.3 ) |

Another way to view the data is through a histogram plot of term frequencies. We can see in the plot below that the word “neurodiversity” stands out followed by the word “people”. Autism is in third place. These frequencies are to be expected and show that our dsi’s do contain the material we are looking for.

****

***Code***

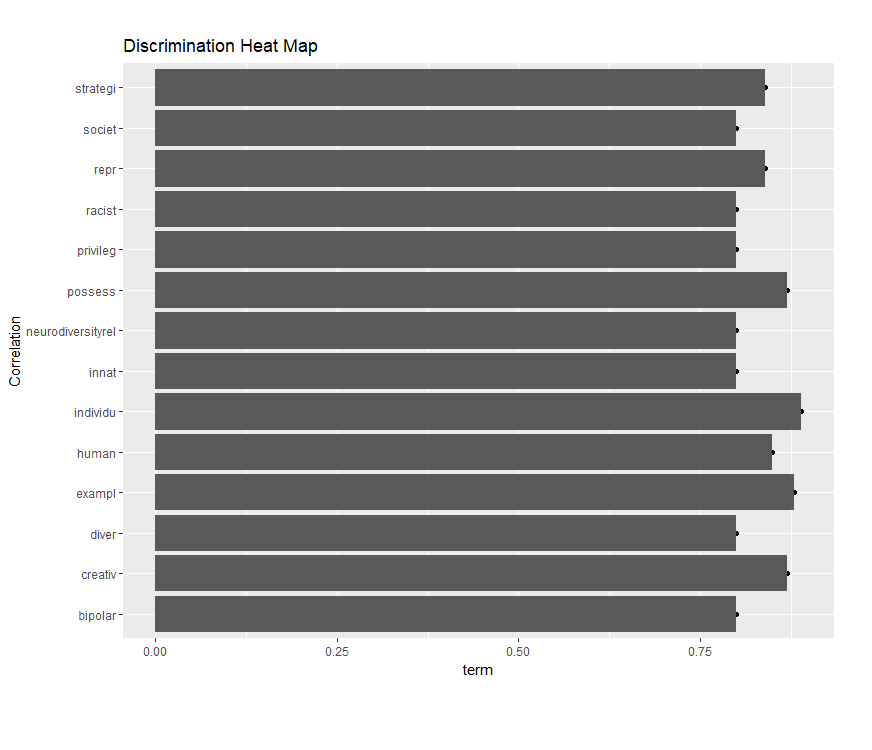
|  |
| --- |
| library("ggplot2")  df\_1 <- head(d, 20)  qplot(df\_1$word, df\_1$freq, main="Term Frequencies", xlab="Terms", ylab="Frequencies") + coord\_flip() + geom\_bar(stat="identity") +theme(plot.margin=unit(c(10,10,20,2),"mm")) |

****

I decided to look for word correlations with autism. I wanted to make sure our corpus makes sense in that we are talking about autism. Looking at the above heat map, we can see that our corpus does deal with autism. We do see some interesting but pratical correlations. “Autism” and treatment are correlated which makes sense. We also see that Autism and Love are correlated. One key sentiment I do see in the whole corpus is that neurodiversity is seen as positive.

***Code***

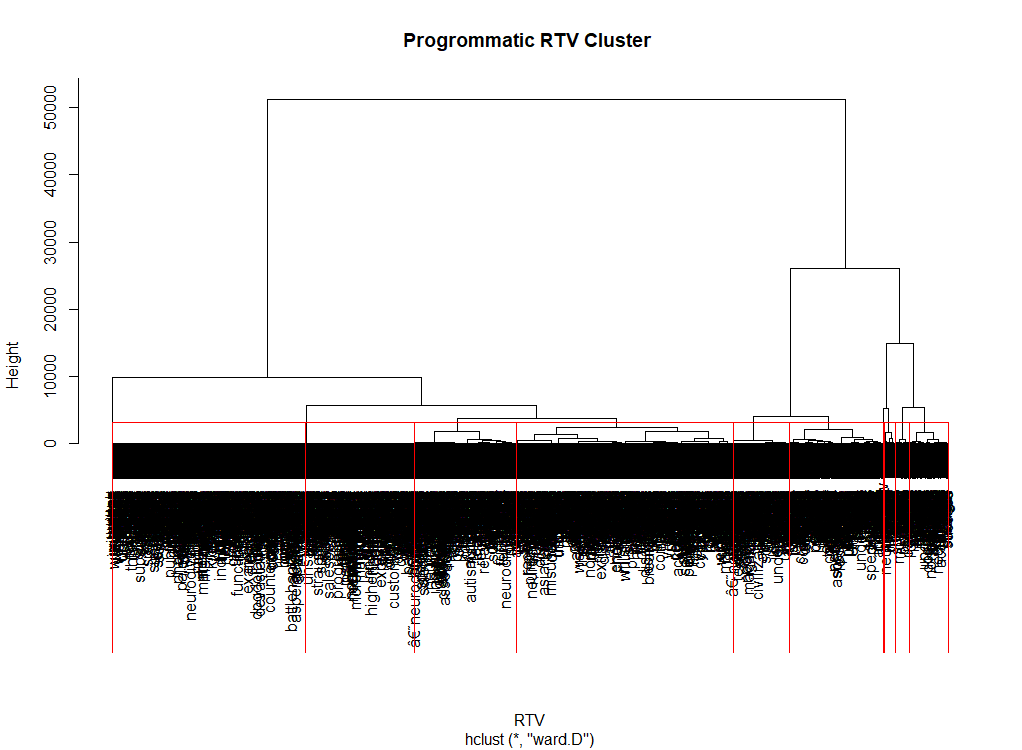
|  |
| --- |
| autism\_assoc <- findAssocs(dtm, "autism", .99)  qplot(names(autism\_assoc$autism), autism\_assoc$autism, xlab="Correlation", ylab="term", main="Autism Heat Map") +coord\_flip() + geom\_bar(stat="identity") +theme(plot.margin=unit(c(10,10,20,2),"mm")) |

****

I checked to see if there were any correlations with the word discrmination. Discrmination has become a key part to to this study and I did not know this when I first created this heat map. We do see that “discrimination” is correlated with “racist” which is what we would expect. But we also see the word “discrimination” is correlated to “neurodiversity”, a fact that will become extremely important later on.

***Code***

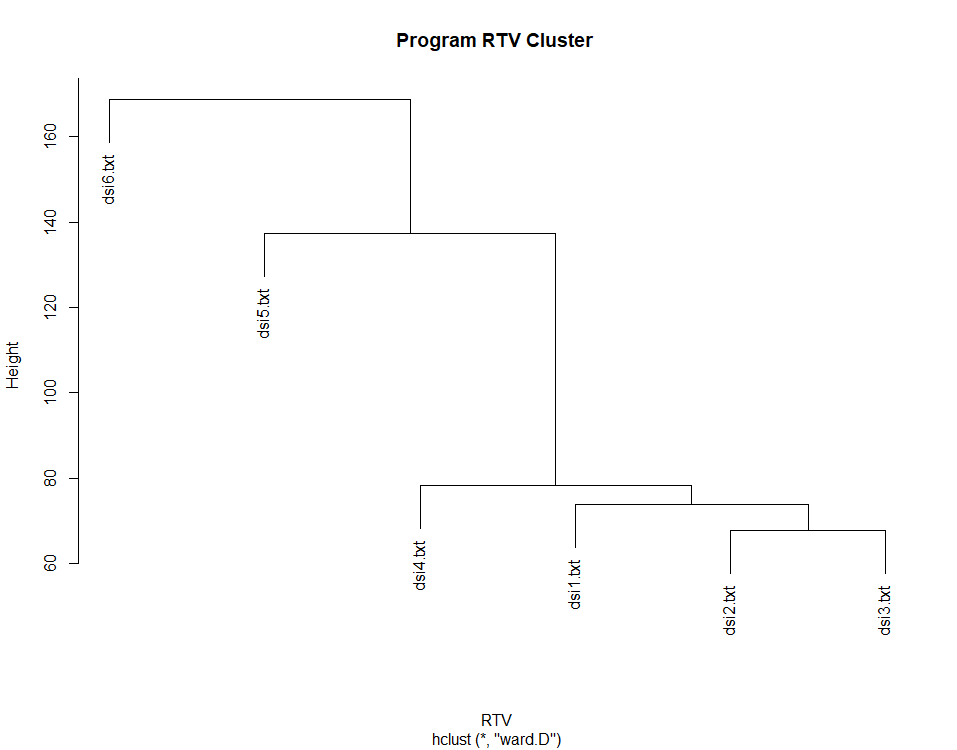
|  |
| --- |
| discrimin\_assoc <- findAssocs(dtm, "discrimin", .8)  qplot(names(discrimin\_assoc$discrimin), discrimin\_assoc$discrimin, xlab="Correlation", ylab="term", main="Discrimination Heat Map") +coord\_flip() + geom\_bar(stat="identity") +theme(plot.margin=unit(c(10,10,20,2),"mm")) |

****

I tried to let the program create clusters for us with the words we have. The above plot is not Intuitive but gives us an idea that there are some clusters in the data. There seems to be five clusters with an outlier cluster.

***Code***

|  |
| --- |
| # Ward Hierarchical Clustering  dd <- dist(scale(m), method = "euclidean")  d <- dist(dd, method = "euclidean") # distance matrix  fit <- hclust(d, method="ward")  plot(fit, main="Progrommatic RTV Cluster ", xlab="RTV") # display dendogram  groups <- cutree(fit, k=10) # cut tree into 5 clusters  # draw dendogram with red borders around the 5 clusters  rect.hclust(fit, k=10, border="red") |

****

I then created clusters based on the DSI’s. What is interesting is that dsi 6 and dsi 5 are outliers. These are the dsi’s with the largest word counts. There is a possibility that tf-idf was not considered in this cluster or it just may be the case that the two dsi’s are different. If you look at dsi 1 through 3, you can see that those dsi’s deal with definitions of neurodiversity while 5 and 6 deal with the application of neurodiversity in a social context. I believe this fact is interesting.

***Code***

|  |
| --- |
| # Ward Hierarchical Clustering  terms <-DocumentTermMatrix(a ,control = list(weighting = function(x) weightTfIdf(x, normalize = FALSE)))  dd <- dist(scale(terms), method = "euclidean")  d <- dist(dd, method = "euclidean") # distance matrix  fit <- hclust(d, method="ward")  plot(fit, main="Program RTV Cluster ", xlab="RTV") # display dendogram  groups <- cutree(fit, k=10) # cut tree into 5 clusters  # draw dendogram with red borders around the 5 clusters  rect.hclust(fit, k=10, border="red") |

**Summary**

We can tell that the articles in our corpus deal with autism. We do see some key words more prominent than others such as “neurodiversity” and “autism”. We also see some complex meaning here such as discrimnation and positive attribute of the austim concept. We can see that the program is able to differentiate at a higher level on what the articles are about by separating dsi 5 and 6 from 1,2 and 3. The issue is that I am not looking for this broad categorizaiton of the articles. I want to find a deeper meaning to the articles and the only way to do this was through some manual effort.

**Manual Work**

The below is my manual work.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 16800 | words |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  | dsi1 | dsi2 | dsi3 | dsi4 | dsi5 | dsi6 |  | tf | idf | tf-idf |
| age | 5 | 0 | 1 | 22 | 32 | 20 |  | 0.004762 | 0.146128 | 0.000696 |
| Parent | 2 | 1 | 0 | 0 | 2 | 29 |  | 0.002024 | 0.243038 | 0.000492 |
| people | 15 | 4 | 1 | 22 | 31 | 49 |  | 0.007262 | 0.066947 | 0.000486 |
| man | 11 | 8 | 1 | 19 | 40 | 33 |  | 0.006667 | 0.066947 | 0.000446 |
| Diversity | 13 | 6 | 1 | 42 | 19 | 30 |  | 0.006607 | 0.066947 | 0.000442 |
| Neurodiversity | 13 | 6 | 1 | 40 | 18 | 29 |  | 0.006369 | 0.066947 | 0.000426 |
| autism | 6 | 3 | 1 | 5 | 11 | 71 |  | 0.005774 | 0.066947 | 0.000387 |
| Person | 4 | 3 | 0 | 8 | 5 | 22 |  | 0.0025 | 0.146128 | 0.000365 |
| benefit | 2 | 0 | 0 | 0 | 9 | 0 |  | 0.000655 | 0.544068 | 0.000356 |
| act | 7 | 3 | 1 | 12 | 13 | 46 |  | 0.004881 | 0.066947 | 0.000327 |
| Disability | 4 | 4 | 0 | 0 | 0 | 6 |  | 0.000833 | 0.367977 | 0.000307 |
| time | 7 | 5 | 0 | 2 | 8 | 13 |  | 0.002083 | 0.146128 | 0.000304 |
| Cure | 7 | 0 | 0 | 1 | 1 | 12 |  | 0.00125 | 0.243038 | 0.000304 |
| need | 5 | 5 | 0 | 2 | 14 | 8 |  | 0.002024 | 0.146128 | 0.000296 |
| thing | 6 | 1 | 0 | 6 | 3 | 17 |  | 0.001964 | 0.146128 | 0.000287 |

For each DSI, I created the tf-idf calculations. My goal here was to create an RTV with the highest tf-idf scores. I took the top 50 scored words from each DSI and added them to a master DSI. I then eliminated the duplicates and took out words I felt did not make sense. I then created an equivalence class for each group of words.

The below is my final RTV. The RTV contains 98 words and noun phrases. Each element is characterized by how many word counts exist in each dsi.

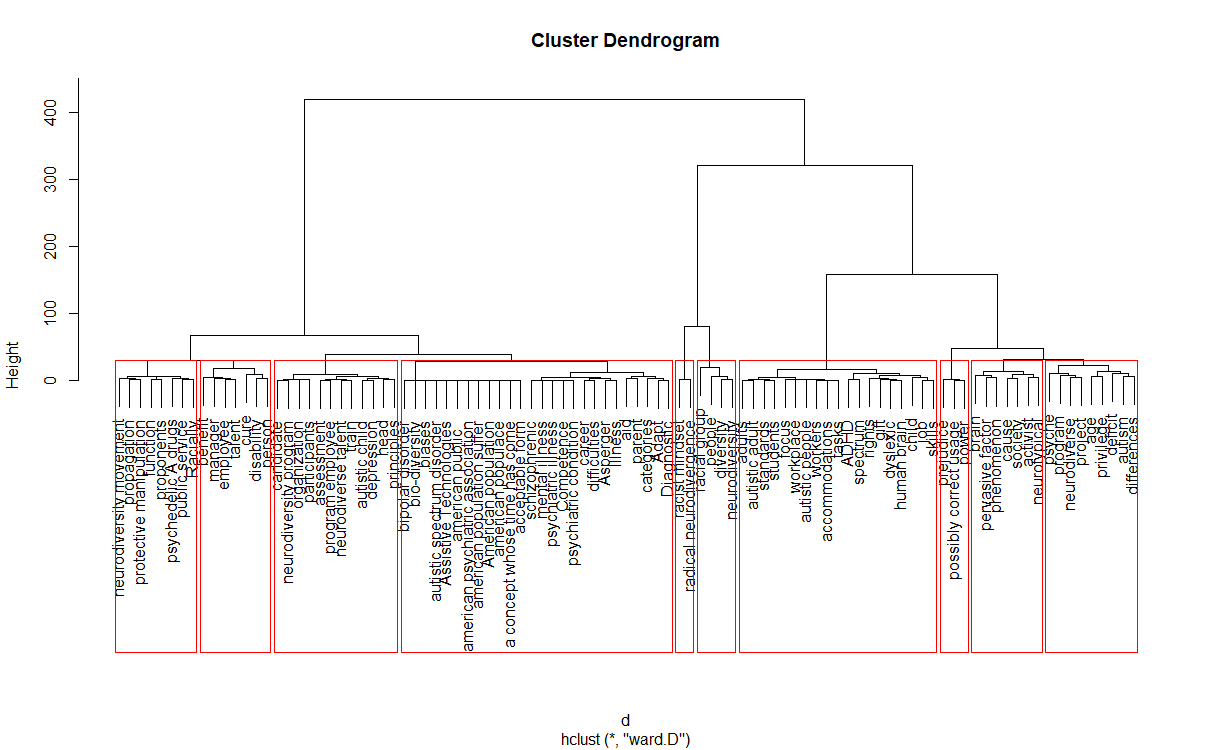
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| EC | RTV | dsi1 | dsi2 | dsi3 | dsi4 | dsi5 | dsi6 |
| peopleSupport | Activist | 0 | 0 | 1 | 8 | 0 | 57 |
|  | Gift | 1 | 2 | 1 | 0 | 0 | 3 |
|  | proponents | 2 | 0 | 0 | 13 | 0 | 0 |
|  | Skills | 0 | 1 | 1 | 0 | 10 | 1 |
|  | privilege | 0 | 4 | 1 | 22 | 31 | 49 |
| neurodiverseMeaning | neurodiverse talent | 0 | 0 | 0 | 0 | 7 | 0 |
|  | neurodiversity movement | 0 | 1 | 0 | 10 | 0 | 5 |
|  | phenomenon | 0 | 3 | 1 | 18 | 15 | 23 |
|  | categories | 0 | 4 | 0 | 0 | 2 | 0 |
|  | bio-diversity | 0 | 1 | 0 | 0 | 0 | 0 |
|  | a concept whose time has come | 0 | 1 | 0 | 0 | 0 | 0 |
| prejudice | prejudice | 0 | 6 | 1 | 42 | 19 | 30 |
|  | propagation | 0 | 0 | 0 | 11 | 0 | 0 |
| rights | Rights | 2 | 1 | 1 | 3 | 0 | 6 |
|  | Power | 0 | 6 | 1 | 40 | 18 | 29 |
|  | Talent | 0 | 0 | 0 | 0 | 20 | 1 |
| age | Adults | 0 | 0 | 1 | 0 | 0 | 2 |
|  | Child | 0 | 0 | 1 | 1 | 1 | 34 |
|  | Age | 5 | 0 | 1 | 22 | 32 | 20 |
| autism | autistic adult | 0 | 0 | 1 | 0 | 0 | 2 |
|  | autistic child | 0 | 0 | 0 | 0 | 0 | 10 |
|  | Autism | 6 | 3 | 1 | 5 | 11 | 71 |
|  | autistic people | 0 | 0 | 1 | 0 | 5 | 10 |
|  | autistic spectrum disorder | 0 | 1 | 0 | 0 | 0 | 0 |
| neuroDiverseperson | neurodiverse | 3 | 2 | 0 | 25 | 27 | 4 |
|  | neurodiversity | 13 | 6 | 1 | 40 | 18 | 29 |
|  | ADHD | 1 | 4 | 1 | 0 | 2 | 0 |
|  | dyslexic | 0 | 2 | 1 | 5 | 0 | 1 |
|  | bipolar disorder | 0 | 1 | 0 | 0 | 0 | 0 |
|  | Asperger | 0 | 2 | 0 | 0 | 1 | 11 |
|  | neurotypical | 1 | 0 | 1 | 22 | 4 | 4 |
|  | schizophrenic | 0 | 2 | 0 | 1 | 0 | 0 |
|  | spectrum | 1 | 4 | 1 | 5 | 3 | 17 |
|  | radical neurodivergence | 0 | 12 | 1 | 52 | 51 | 189 |
|  | depression | 0 | 0 | 0 | 0 | 0 | 7 |
| stigma | possibly correct usage | 0 | 6 | 1 | 40 | 18 | 29 |
|  | acceptable form | 0 | 1 | 0 | 0 | 0 | 0 |
| discrimination | Biases | 0 | 1 | 0 | 0 | 0 | 0 |
|  | differences | 9 | 2 | 1 | 0 | 2 | 1 |
|  | diversity | 13 | 6 | 1 | 42 | 19 | 30 |
|  | racial group | 0 | 10 | 1 | 34 | 41 | 103 |
|  | Racially | 0 | 0 | 0 | 17 | 0 | 0 |
|  | racist mindset | 0 | 13 | 1 | 43 | 50 | 196 |
| work | Career | 0 | 2 | 0 | 0 | 4 | 0 |
|  | disability | 4 | 4 | 0 | 0 | 0 | 6 |
|  | accommodations | 0 | 0 | 1 | 0 | 4 | 1 |
|  | employee | 0 | 0 | 0 | 0 | 22 | 0 |
|  | Adapt | 0 | 3 | 0 | 0 | 0 | 0 |
|  | workers | 0 | 0 | 1 | 0 | 5 | 0 |
|  | workplace | 1 | 0 | 1 | 0 | 3 | 0 |
|  | Job | 0 | 1 | 1 | 0 | 11 | 0 |
|  | manager | 0 | 0 | 0 | 0 | 17 | 0 |
|  | Train | 0 | 0 | 0 | 0 | 7 | 4 |
| illness | Illness | 0 | 2 | 0 | 0 | 0 | 9 |
|  | Cure | 7 | 0 | 0 | 1 | 1 | 12 |
|  | Cause | 5 | 0 | 1 | 6 | 11 | 14 |
|  | mental illness | 0 | 2 | 0 | 0 | 0 | 1 |
|  | pervasive factor | 0 | 5 | 1 | 15 | 6 | 38 |
|  | protective manipulation | 0 | 0 | 0 | 13 | 0 | 0 |
| psychiatric | psychiatric illness | 0 | 2 | 0 | 0 | 0 | 0 |
|  | psychiatric condition | 0 | 2 | 0 | 0 | 0 | 0 |
|  | psychedelic drugs | 0 | 1 | 0 | 21 | 0 | 0 |
|  | Diagnostic | 0 | 3 | 0 | 0 | 0 | 0 |
| society | Society | 6 | 1 | 1 | 2 | 0 | 3 |
|  | public service | 0 | 0 | 0 | 15 | 0 | 0 |
| deficits | Deficit | 0 | 1 | 2 | 0 | 1 | 4 |
|  | difficulties | 0 | 2 | 0 | 0 | 4 | 0 |
|  | Competence | 0 | 2 | 0 | 0 | 0 | 0 |
| help | Aid | 1 | 2 | 0 | 0 | 3 | 19 |
|  | assessment | 0 | 0 | 0 | 0 | 6 | 0 |
|  | Benefit | 3 | 0 | 0 | 0 | 16 | 0 |
|  | candidate | 0 | 0 | 0 | 0 | 9 | 0 |
|  | function | 3 | 1 | 0 | 10 | 0 | 7 |
|  | support | 4 | 0 | 0 | 2 | 8 | 4 |
|  | Assistive Technologies | 0 | 1 | 0 | 0 | 0 | 0 |
| population | american populace | 0 | 1 | 0 | 0 | 0 | 0 |
|  | American population | 0 | 1 | 0 | 0 | 0 | 0 |
|  | american population suffer | 0 | 1 | 0 | 0 | 0 | 0 |
|  | american public | 0 | 1 | 0 | 0 | 0 | 0 |
| organization | american psychiatric association | 0 | 1 | 0 | 0 | 0 | 0 |
|  | organization | 0 | 0 | 0 | 2 | 12 | 3 |
| head | Brain | 1 | 7 | 1 | 6 | 0 | 4 |
|  | Head | 0 | 0 | 0 | 0 | 4 | 6 |
|  | human brain | 0 | 2 | 1 | 2 | 0 | 0 |
|  | Psyche | 0 | 1 | 0 | 34 | 0 | 0 |
| people | Parent | 2 | 1 | 0 | 0 | 2 | 29 |
|  | People | 15 | 4 | 1 | 22 | 31 | 49 |
|  | Person | 4 | 3 | 0 | 8 | 5 | 22 |
|  | students | 0 | 0 | 1 | 4 | 1 | 0 |
| studies | participants | 0 | 0 | 0 | 0 | 11 | 0 |
|  | program | 0 | 3 | 0 | 0 | 34 | 3 |
|  | program employee | 0 | 0 | 0 | 0 | 6 | 0 |
|  | principles | 0 | 0 | 0 | 4 | 4 | 13 |
|  | project | 0 | 2 | 0 | 36 | 27 | 4 |
|  | standards | 0 | 0 | 1 | 3 | 0 | 0 |
|  | neurodiversity program | 0 | 0 | 0 | 0 | 10 | 0 |

I then took my RTV and truncated everything down to the Equivalence classes. The below is my final result.

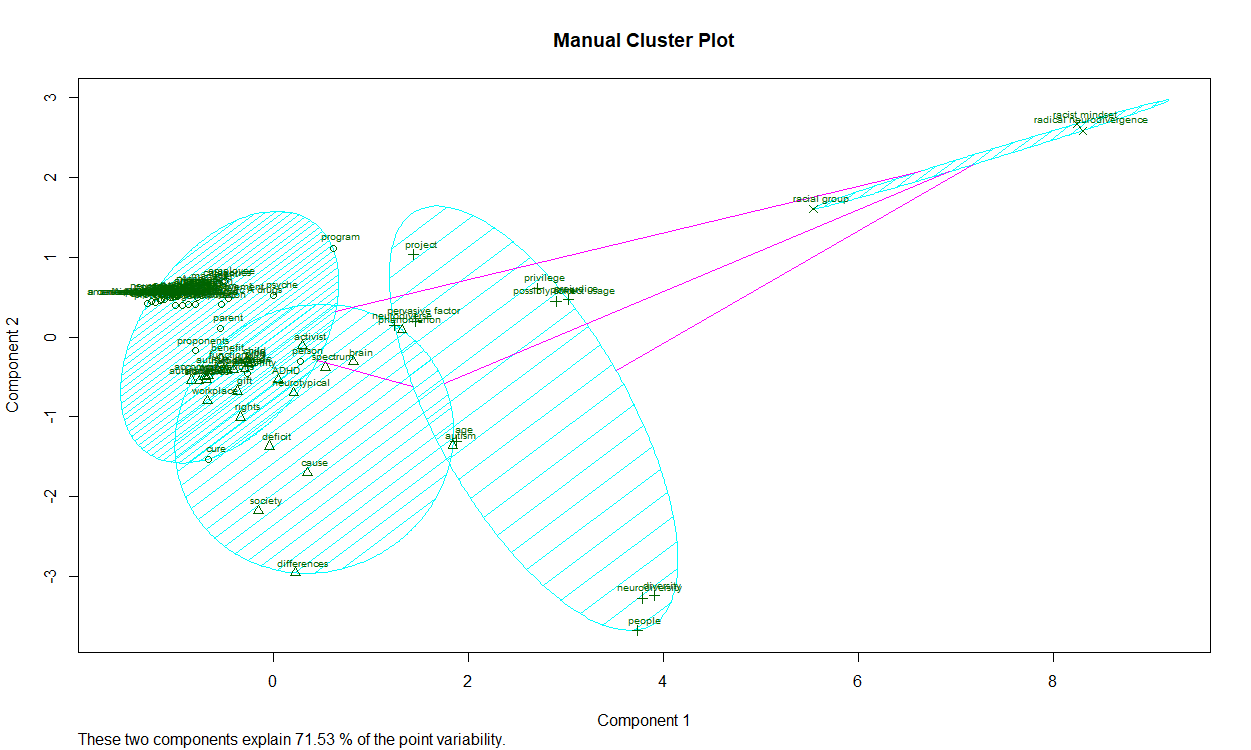
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | dsi1 | dsi2 | dsi3 | dsi4 | dsi5 | dsi6 |
| peopleSupport | 3 | 7 | 4 | 43 | 41 | 110 |
| neurodiverseMeaning | 0 | 10 | 1 | 28 | 24 | 28 |
| prejudice | 0 | 6 | 1 | 53 | 19 | 30 |
| rights | 2 | 7 | 2 | 43 | 38 | 36 |
| age | 5 | 0 | 3 | 23 | 33 | 56 |
| autism | 6 | 4 | 3 | 5 | 16 | 93 |
| neuroDiverseperson | 19 | 35 | 6 | 150 | 106 | 262 |
| stigma | 0 | 7 | 1 | 40 | 18 | 29 |
| discrimination | 22 | 32 | 4 | 136 | 112 | 330 |
| work | 5 | 10 | 4 | 0 | 73 | 11 |
| illness | 12 | 9 | 2 | 35 | 18 | 74 |
| pyschiatric | 0 | 8 | 0 | 21 | 0 | 0 |
| society | 6 | 1 | 1 | 17 | 0 | 3 |
| deficits | 0 | 5 | 2 | 0 | 5 | 4 |
| help | 11 | 4 | 0 | 12 | 42 | 30 |
| population | 0 | 4 | 0 | 0 | 0 | 0 |
| organization | 0 | 1 | 0 | 2 | 12 | 3 |
| head | 1 | 10 | 2 | 42 | 4 | 10 |
| people | 21 | 8 | 2 | 34 | 39 | 100 |
| studies | 0 | 5 | 1 | 43 | 92 | 20 |

**EDA of RTV**

The first thing I did with the RTV was to create a dendrogram to see if there were any clusters in the data. I can see six clusters so far using a hiearchical means cluster.

***Code***

|  |
| --- |
| library(HistDAWass)  #read data  data=read.csv('E:/data.csv', header = TRUE)  rownames(data) <- data$ï..  data$ï..<- NULL  #data <- t(data)  dd <- dist(scale(data), method = "euclidean")  #dd<-t(dd)  hc <- hclust(dd, method = "ward.D2")  plot(hc)  # Ward Hierarchical Clustering  d <- dist(dd, method = "euclidean") # distance matrix  fit <- hclust(d, method="ward")  plot(fit) # display dendogram  groups <- cutree(fit, k=6) # cut tree into 5 clusters  # draw dendogram with red borders around the 5 clusters  rect.hclust(fit, k=10, border="red") |

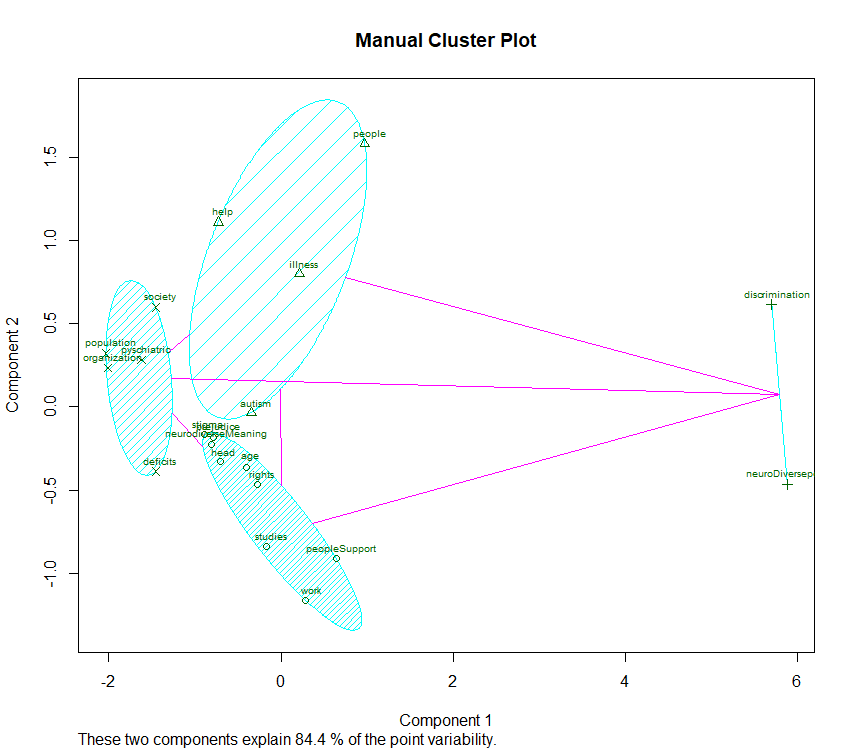
****

I also used a k-means alogrithm to split the RTV data into four clusters. What catches the eye immediately is the outlying cluster to the upper right. This cluster seems to be a cluster that denotes racism when it comes to autism. Even in our hiearchical cluster, we see a similar categorization of racism. We have a cluster of the words, “racist mindset, radical neuro divergence, racial group, people, diversity and neuro diversity”.

As for our other three clusters, we see some overlap but there tends to be a general tendency of society, accomodations and defintion. We can see more when we use the EC list.

***Code***

|  |
| --- |
| library(cluster)  new\_cluster <- pam(dd, 4)  plot(new\_cluster)  clusplot(new\_cluster, shade = TRUE, labels=3, cex.txt = 0.6, main="Manual Cluster Plot") |



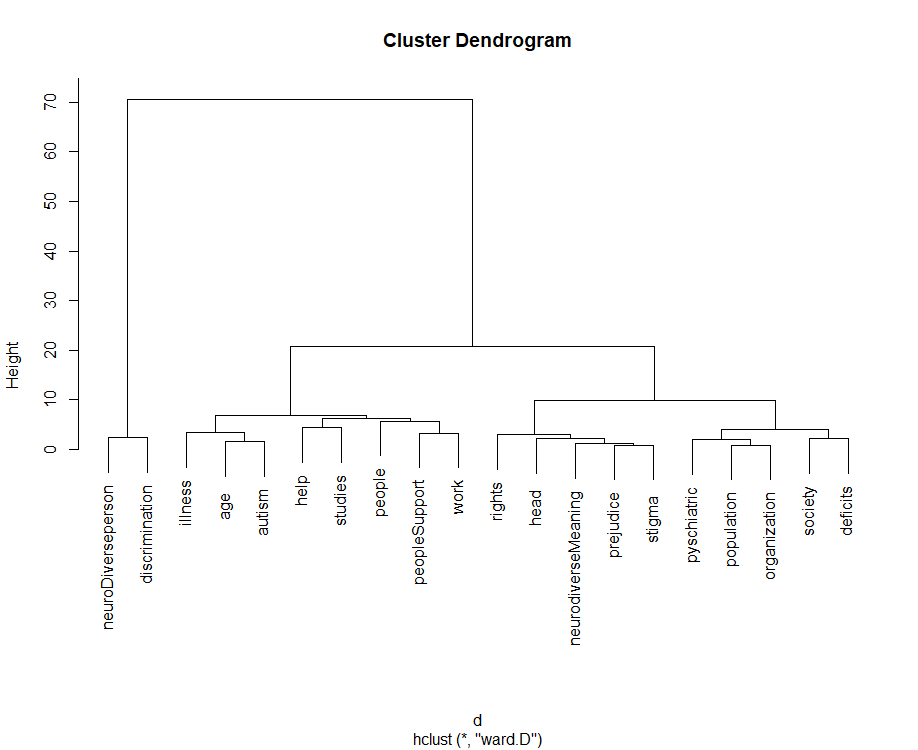
Our EC clusters show the discrmination pattern with discrimination and neuro divergence. We also see a “society” cluster to the far left. The words here are “poplulation” and “deficits”. This cluster represents society and how we view ourselves. In this corpus, society is defining who we are and who cannot remain in society and how this relates to autism.

The cluster on the top represents the definition of autism. The questions asked here are is autism an illness and if so how do we help.

The final cluster is accomodation and support for those with Autism. This support can be through studies, accomodations at work and rights of a neurodiverse indvidual.

***Code***

|  |
| --- |
| library(cluster)  new\_cluster <- pam(dd, 4)  plot(new\_cluster)  clusplot(new\_cluster, shade = TRUE, labels=3, cex.txt = 0.6, main="Manual Cluster Plot") |

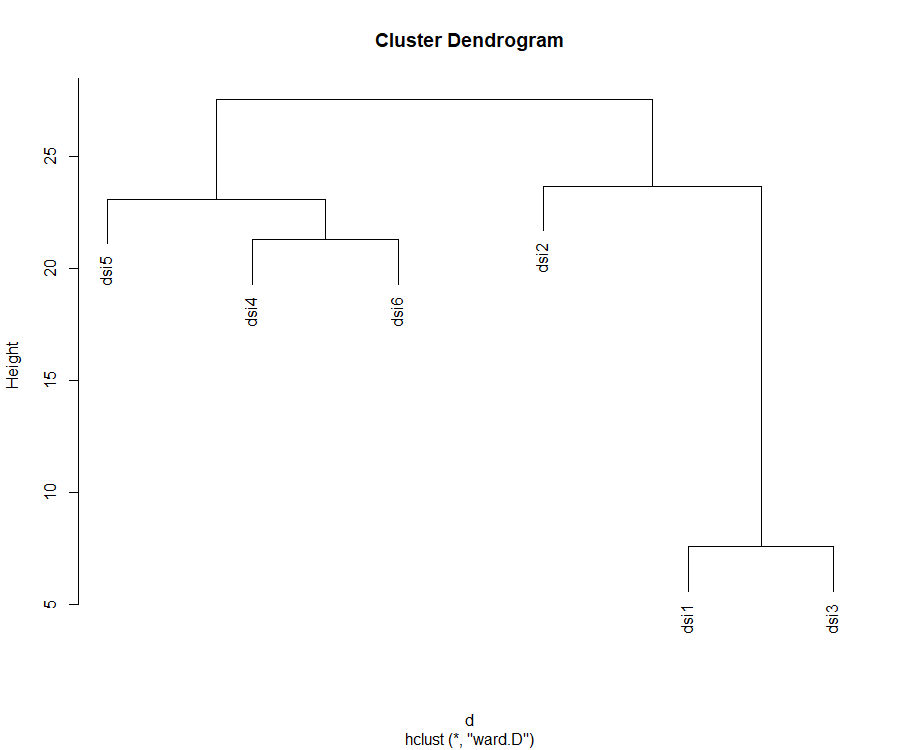


We can also see the same patterns with our hiearchical dendrogram. We see the discrimination cluster but we also see a cluster that contains the words “rights, head, neurodiversyMeaning, prejudice and stigma”. This cluster seems to represent how the meaning of neurodiversity is a controversial topic when it comes to rights.

***Code***

|  |
| --- |
| par(mfcol=c(1,1))  library(HistDAWass)  #read data  data=read.csv('E:/data1.csv', header = TRUE)  rownames(data) <- data$ï..  data$ï..<- NULL  #data <- t(data)  dd <- dist(scale(data), method = "euclidean")  dd<-t(dd)  hc <- hclust(dd, method = "ward.D2")  plot(hc)  # Ward Hierarchical Clustering  d <- dist(dd, method = "euclidean") # distance matrix  fit <- hclust(d, method="ward")  plot(fit) # display dendogram |

Let’s look at our dendrogram for our RTV.

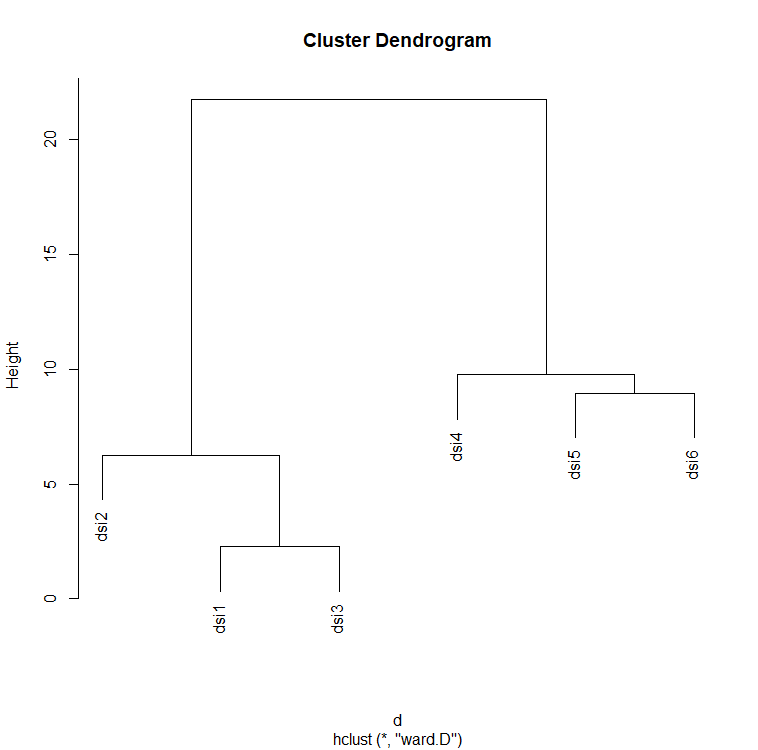


We can see that we have three clusters. 4,5 and 6 are clustered which makes sense. 2 seems to an outlier which might be an issue. There must be something about 2 that doesn’t match 1 and 3.

***Code***

|  |
| --- |
| library(HistDAWass)  #read data  data=read.csv('E:/data.csv', header = TRUE)  rownames(data) <- data$ï..  data$ï..<- NULL  data <- t(data)  dd <- dist(scale(data), method = "euclidean")  #dd<-t(dd)  hc <- hclust(dd, method = "ward.D2")  plot(hc)  # Ward Hierarchical Clustering  d <- dist(dd, method = "euclidean") # distance matrix  fit <- hclust(d, method="ward")  plot(fit) # display dendogram  groups <- cutree(fit, k=6) # cut tree into 5 clusters  # draw dendogram with red borders around the 5 clusters  rect.hclust(fit, k=10, border="red") |

I then tried to see how our EC DSI clusters look. We can see two clusters now. DSI 4,5 and 6 are combined into one cluster which makes sense. 5 and 6 represent how neurodiversity can be effective in our society. DSI 4 veers toward that meaning by emphasizing people’s misconceptions about neurodiverstiy.

****

***Code***

|  |
| --- |
| par(mfcol=c(1,1))  library(HistDAWass)  #read data  data=read.csv('E:/data1.csv', header = TRUE)  rownames(data) <- data$ï..  data$ï..<- NULL  data <- t(data)  dd <- dist(scale(data), method = "euclidean")  dd<-t(dd)  hc <- hclust(dd, method = "ward.D2")  plot(hc)  # Ward Hierarchical Clustering  d <- dist(dd, method = "euclidean") # distance matrix  fit <- hclust(d, method="ward")  plot(fit) # display dendogram |

**Final Taxonomy**

Using the clusters that we have I believe that the taxonomy is as follows.

|  |  |  |  |
| --- | --- | --- | --- |
| People Discrimination |  |  |  |
|  | neuroDiverseperson | neurodiverse |  |
|  |  | neurodiversity |  |
|  |  | ADHD |  |
|  |  | dyslexic |  |
|  |  | bipolar disorder |  |
|  |  | Asperger |  |
|  |  | neurotypical |  |
|  |  | schizophrenic |  |
|  |  | spectrum |  |
|  |  | radical neurodivergence |  |
|  |  | depression |  |
|  |  |  |  |
|  |  |  |  |
|  | discrimination | biases |  |
|  |  | differences |  |
|  |  | diversity |  |
|  |  | racial group |  |
|  |  | Racially |  |
|  |  | racist mindset |  |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Autism Definition |  |  |  |
|  | autism | autistic adult |  |
|  |  | autistic child |  |
|  |  | autism |  |
|  |  | autistic people |  |
|  |  | autistic spectrum disorder |  |
|  |  |  |  |
|  | illness | illness |  |
|  |  | cure |  |
|  |  | cause |  |
|  |  | mental illness |  |
|  |  | pervasive factor |  |
|  |  | protective manipulation |  |
|  |  |  |  |
|  | help | aid |  |
|  |  | assessment |  |
|  |  | benefit |  |
|  |  | candidate |  |
|  |  | function |  |
|  |  | support |  |
|  |  | Assistive Technologies |  |
|  |  |  |  |
|  | people | parent |  |
|  |  | people |  |
|  |  | person |  |
|  |  | students |  |
|  |  |  |  |

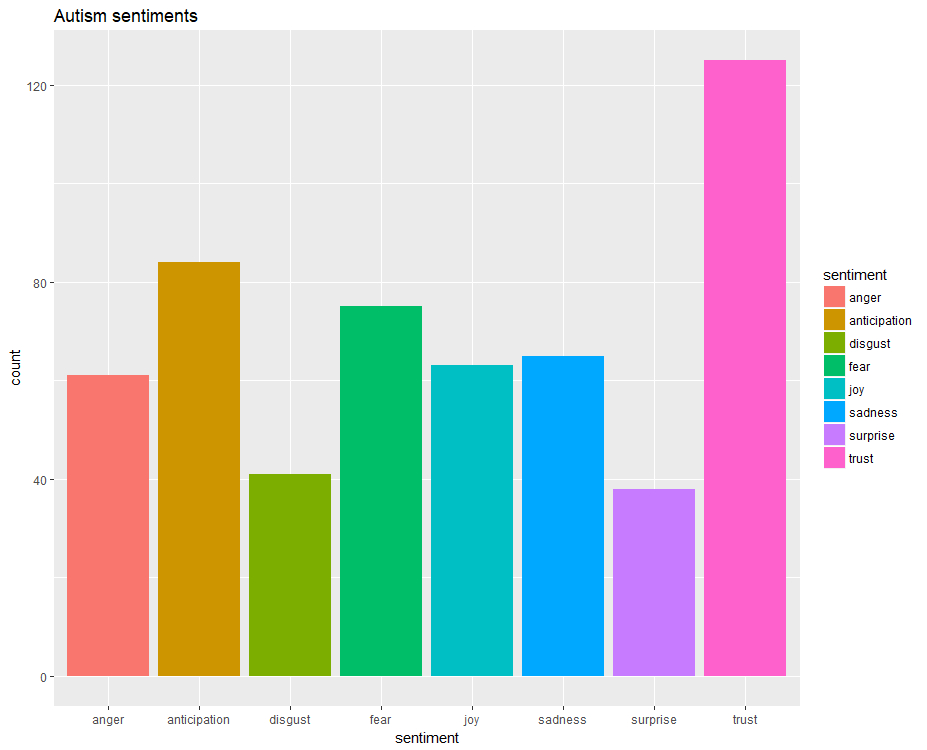
|  |  |  |  |
| --- | --- | --- | --- |
| Society |  |  |  |
|  | deficits | deficit |  |
|  |  | difficulties |  |
|  |  | Competence |  |
|  |  |  |  |
|  | population | american populace |  |
|  |  | American population |  |
|  |  | american population suffer |  |
|  |  | american public |  |
|  |  |  |  |
|  | organization | american psychiatric association |  |
|  |  | organization |  |
|  |  |  |  |
|  | pyschiatric | psychiatric illness |  |
|  |  | psychiatric condition |  |
|  |  | psychedelic  drugs |  |
|  |  | Diagnostic |  |
|  |  |  |  |
|  | society | society |  |
|  |  | public service |  |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Accomodations |  |  |  |
|  | peopleSupport | activist |  |
|  |  | gift |  |
|  |  | proponents |  |
|  |  | skills |  |
|  |  | privilege |  |
|  |  |  |  |
|  | neurodiverseMeaning | neurodiverse talent |  |
|  |  | neurodiversity movement |  |
|  |  | phenomenon |  |
|  |  | categories |  |
|  |  | bio-diversity |  |
|  |  | a concept whose time has come |  |
|  |  |  |  |
|  | prejudice | prejudice |  |
|  |  | propagation |  |
|  |  |  |  |
|  | rights | rights |  |
|  |  | power |  |
|  |  | talent |  |
|  |  |  |  |
|  | age | adults |  |
|  |  | child |  |
|  |  | age |  |
|  |  |  |  |
|  | work | career |  |
|  |  | disability |  |
|  |  | accommodations |  |
|  |  | employee |  |
|  |  | Adapt |  |
|  |  | workers |  |
|  |  | workplace |  |
|  |  | job |  |
|  |  | manager |  |
|  |  | train |  |
|  |  |  |  |
|  | head | brain |  |
|  |  | head |  |
|  |  | human brain |  |
|  |  | psyche |  |
|  |  |  |  |
|  | studies | participants |  |
|  |  | program |  |
|  |  | program employee |  |
|  |  | principles |  |
|  |  | project |  |
|  |  | standards |  |
|  |  | neurodiversity program |  |
|  |  |  |  |
|  | stigma | possibly correct usage |  |
|  |  | acceptable form |  |
|  |  |  |  |

**Conclusion**

Overall, I feel that the 6 dsi’s show a positive sentiment when it comes to autism. The articles state how people with autism fit into society and issues that arise to achieve this fitness. One of the key conclusions is that autism is not an illness (a belief from individuals) but an issue that should be accommodated and explored.

Below is a sentiment algorithm run on the words in the corpus. We can see that trust is the key sentiment.

****

***Code***

|  |
| --- |
| sentd<-get\_nrc\_sentiment(as.vector(d$word))  td<-data.frame(t(sentd))  td  td\_new <- data.frame(rowSums(td[2:2699]))  #The function rowSums computes column sums across rows for each level of a grouping variable.  #Transformation and  cleaning  names(td\_new)[1] <- "count"  td\_new <- cbind("sentiment" = rownames(td\_new), td\_new)  rownames(td\_new) <- NULL  td\_new2<-td\_new[1:8,]  #Visualisation  library("ggplot2")  qplot(sentiment, data=td\_new2, weight=count, geom="bar",fill=sentiment)+ggtitle("Autism sentiments") |