Ciencias Sociales
Universidad de la República
URUGUAY

R aplicado al análisis cualitativo

Basado en el análisis de la investigación:
"¿Y después qué?
Una mirada a los procesos de reinserción de personas recientemente liberadas del sistema penitenciario uruguayo."

Macarena Campiglia 4.949.670-9
Febrero 2023

Curso de Educación Permanente
Docentes:
Elina Gomez

Índice:

Introducción

El presente trabajo se enmarca en la entrega final del curso "R aplicado al análisis cualitativo" dictado en la unidad de Educación Permanente de la Universidad de la República. El mismo, dictado por la Mag. Elina Goméz, durante el mes de diciembre del 2022, pretende profundizar en el análisis de fuentes documentales a través de R, En el siguiente trabajo buscaremos presentar el análisis de una serie de entrevistas en profundidad realizadas en el marco de la investigación*¿Y después qué? Una mirada a los procesos de reinserción de las personas liberadas del sistema penitenciario uruguayo*. A través de este trabajo realizamos una codificación manual de los textos que surgieron de las entrevistas realizadas en la investigación, mediante la plataforma RQDA. Para poder presentar datos relevantes a este estudio se requirieron dos acciones fundamentales: una correcta limpieza y homogeneización de los datos así como el uso de técnicas de minería de datos. Como se verá a continuación.

Análisis documental mediante plataforma RQDA[1]

Librerías necesarias:

library(RQDA)

Ejecutar

RQDA()

- Visualización de algunos datos relevantes obtenidos

La función summaryCodings() nos devuelve una lista con el número de entrevistas vinculadas a cada código que creamos en RQDA y con summaryCodings(byFile= TRUE) podemos ver por entrevista el número de códigos asociados a la misma.

summaryCodings()

Ejemplo

```
Number of files associated with each code.

                    Adicciones                                  Formación académica
                             5                                                    7
                Alfabetización                         Institucionalidad académica
                             3                                                    2
                    Autoridades                                    Masculinización
                             3                                                    1
       Condiciones de encierro                                  Politicas Públicas
                             5                                                    6
       Condiciones de vivienda                                          Protocolos
                             1                                                   14
                Contexto social                               Recursos económicos
                             6                                                    8
Coordinación interinstitucional                                 Recursos humanos
                             6                                                   16
   Cuidado del personal técnico                                       Reincidencia
                             3                                                    8
         Educación sin castigo                                              Salud
                             4                                                    6
Efectos cognitivos del encierro                                  Situación de calle
                             2                                                    8
               Exclusión social                                       Trabajo digno
                             6                                                   14
                      Familia
                             3
```

[1] [^1*Archivo utilizado:* Final.rdqa (el mismo se conformó de seis entrevistas transcriptas en formato .txt)]

summaryCodings(byFile= TRUE)

Ejemplo

```
$Entrevista001AliciaAlvarez

                    Adicciones
                             1
                 Alfabetización
                             7
          Condiciones de encierro
                             6
                Contexto social
                             7
Coordinación interinstitucional
                             3
            Educación sin castigo
                             2
Efectos cognitivos del encierro
                             3
                Exclusión social
                             3
             Formación académica
                             3
   Institucionalidad académica
                             2
           Recursos económicos
                             1
              Recursos humanos
                             1
                  Reincidencia
                             2
            Situación de calle
                             2
                Trabajo digno
                             4
```

La tabla con codificaciones nos permite visualizar la cantidad categorías presentes en cada una de las entrevistas realizadas

tabla_cods=getCodingTable()

Ejemplo

	rowid	cid	fid	codename	filename	index1	index2	CodingLength
1	2	9	1	Coordinación interinstitucional	Entrevista001AliciaAlvarez	2031	2193	162
2	3	6	1	Trabajo digno	Entrevista001AliciaAlvarez	5420	5848	428
3	4	6	1	Trabajo digno	Entrevista001AliciaAlvarez	6515	6715	200
4	5	4	1	Institucionalidad académica	Entrevista001AliciaAlvarez	7745	8307	562
5	6	10	1	Recursos económicos	Entrevista001AliciaAlvarez	11354	11896	542
6	7	11	1	Recursos humanos	Entrevista001AliciaAlvarez	11897	12716	819
7	8	4	1	Institucionalidad académica	Entrevista001AliciaAlvarez	13128	13678	550
8	9	3	1	Formación académica	Entrevista001AliciaAlvarez	15001	15444	443
9	10	15	1	Contexto social	Entrevista001AliciaAlvarez	15696	16712	1016
10	11	15	1	Contexto social	Entrevista001AliciaAlvarez	18096	18525	429
11	12	16	1	Exclusión social	Entrevista001AliciaAlvarez	18813	19636	823
12	13	16	1	Exclusión social	Entrevista001AliciaAlvarez	19999	20720	721
13	14	5	1	Situación de calle	Entrevista001AliciaAlvarez	21105	21483	378
14	15	5	1	Situación de calle	Entrevista001AliciaAlvarez	22384	22778	394
15	16	17	1	Adicciones	Entrevista001AliciaAlvarez	22779	24248	1469
16	17	3	1	Formación académica	Entrevista001AliciaAlvarez	24422	24944	522
17	18	3	1	Formación académica	Entrevista001AliciaAlvarez	25282	25798	516
18	19	1	1	Alfabetización	Entrevista001AliciaAlvarez	26575	27136	561
19	20	15	1	Contexto social	Entrevista001AliciaAlvarez	25905	26573	668

Limpieza de textos

Instalaciones requeridas

- install.packages("pdftools") *para cargar archivos de texto en pdf*
- install.packages("tm") *para minería de texto*
- install.packages("SnowballC") *para minería de texto*
- install.packages("wordcloud2") *paquete para generar la nube de palabras*

librerías requeridas

```
library("pdftools")
library("tm")
library("SnowballC")
library("wordcloud2")
library("udpipe")
library("ggplot2")
library("readtext")
library("corpus")
library(“dplyr”)
```

Los datos obtenidos de las entrevistas presentan muchos caracteres, espacios y palabras que no sirven para realizar un análisis preciso de los datos recabados. Es por eso que a la hora de trabajar con textos planos como son las entrevistas que realizamos, así como otras bases de datos, utilizaremos las llamadas técnicas de “limpieza”.

- Limpieza de base de datos PDF`

Cargamos nuestra base de datos:

texto<-pdf_text("C:/Users/Usuario/OneDrive/Escritorio/EntregaR/Entrevistasrecopiladasanalisisquanteda.pdf")

Debido a que en el español se utilizan tildes y otros signos de puntuación, usamos la función _iconv()_ para quitarlos.

```
texto \<- iconv(texto, "UTF-8")
texto \<- iconv(texto, "latin1")
texto = iconv(texto, to="ASCII//TRANSLIT")
```

Utilizando la función Corpus(), indicamos la fuente de nuestro texto

docs \<- Corpus(VectorSource(texto))

Luego limpiamos de caracteres especiales:

toSpace \<- content_transformer(function (x , pattern) gsub(pattern, " ", x))

docs \<- tm_map(docs, toSpace,"\r\n")

docs \<- tm_map(docs, toSpace,"/")

docs \<- tm_map(docs, toSpace,"\@")

Convertimos el texto a minúsculas:

docs \<- tm_map(docs, content_transformer(tolower))

Quitamos puntuación

docs \<- tm_map(docs, removePunctuation)

Quitamos los números

docs \<- tm_map(docs, removeNumbers)

Quitamos las palabras comunes en español

docs \<- tm_map(docs, removeWords, stopwords("spanish"))

Y por último Quitamos palabras comunes que consideramos no tenían relación con lo analizado

docs \<- tm_map (docs, removeWords, c("sale","digo","decir","digamos","claro","dia", "seria","general","cosa","menos","podes","algun","risas","pasa","anos","puede", "claro","asi","tipo","vez","ver","habia","parte","alguna","siempre","situacion","capaz", "dos","trabajando","aca","mas","tambien","hacer","entonces","si","bueno","creo","despu es","hace","bueno","eh,","estan","veces","ser","ahora","muchas","bien,","ta,","si,","ahi", "bueno","persona","eh","bien","ta","parece","vos","cada","momento","van","cosas", "tener","todas"))

- Limpieza de base de datos TXT`

Cargamos nuestra base de datos

base_final<-read.delim2("C:/Users/Usuario/OneDrive/Escritorio/EntregaR/baseentrevistadoentrevista.txt",sep="\t",header = T)

Convertimos a DFM nuestra base de datos y al mismo tiempo realizamos su limpieza

text_df_pre \<- dfm(tokens(base_final\$text, remove_punct = TRUE,
remove_numbers = TRUE),
tolower = TRUE,
remove_numbers = TRUE, remove = c(stopwords("spanish"))) %\>%
dfm_remove(min_nchar=3) %\>%
dfm_trim(min_termfreq = 20, min_docfreq = 2)

Y eso nos devolverá un documento con todas las palabras de más de tres letras, que se repitan más de veinte veces y que se repitan en por lo menos dos entrevistas listo para trabajar..

Minería de texto

En esta sección buscaremos visualizar algunas de las técnicas de minería de textos que pueden servir tanto para análisis como para visualización de datos surgidos a partir de textos de formas dinámicas y atractivas.

Primeramente construimos una matriz term-document a partir de nuestro documento _docs_ al que le hicimos previamente una "limpieza" como se vió anteriormente.

```
mtd \<- TermDocumentMatrix(docs)
m \<- as.matrix(mtd)
v \<- sort(rowSums(m),decreasing=TRUE)
d \<- data.frame(word = names(v),freq=v)
```

Que podremos utilizar para dos herramientas de análisis:

- 1) Frecuencia de las 20 palabras más nombradas

```
palabras \<- docs %\>%
TermDocumentMatrix() %\>%
as.matrix() %\>%
rowSums() %\>%
sort(decreasing = TRUE)

palabras %\>%
head(20)
```

personas	trabajo	libertad	gente	dinali	trabajar	carcel	realidad	sistema
52	50	48	43	36	29	27	25	22
tiempo	proceso	reinsercion	familia	hoy	carceles	poblacion	privacion	salir
22	19	18	17	17	16	16	15	15
laboral	pensar							
15	15							

- 2) Nube de palabras

Para poner en juego distintos colores, tamaños y cantidad de palabras utilizadas, presentamos tres nubes distintas para ver distintos tipos de visualizaciones:

wordcloud2(data=d,10, size = 0.65, shape = "cloud", color = 'random-light',backgroundColor = 'light')

wordcloud2(data = d,5, size = 0.6, shape = "cloud", color="random-light", backgroundColor = 'grey')

wordcloud2(data=d,24, size = 0.65, shape = "triangle" ,color = 'random-dark',backgroundColor = "pink")

- 3) Co-ocurrencia de códigos

correr librería

library(ggplot2)

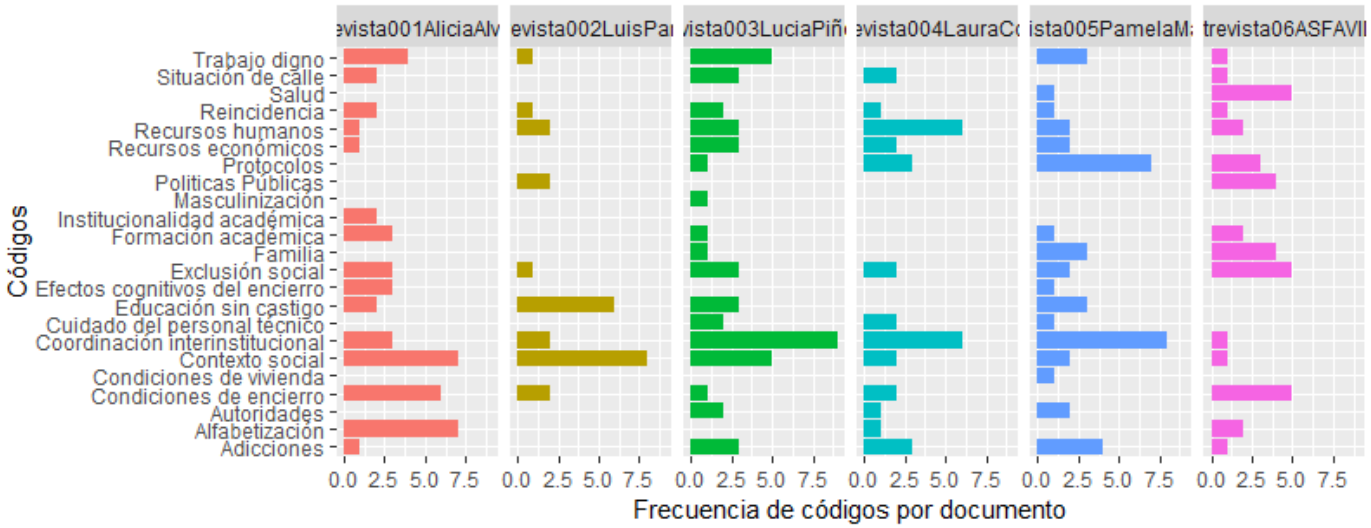
Otra de las herramientas del uso del RQDA, es la posibilidad de crear tablas de co-ocurrencia. Para este caso buscamos observar según cada entrevistas la frecuencia de aparición de los códigos generados, al mismo que logramos comparar la aparición entre los distintos entrevistados.

- Para esto usamos:

cods = getCodingTable()[,4:5]

Y luego podremos obtener una tabla a partir de la siguiente función.

```
ggplot(cods, aes(codename, fill=filename)) + geom_bar(stat="count") +
  facet_grid(~filename) + coord_flip() +
  theme(legend.position = "none") +
  ylab("Frecuencia de codigos por documento") + xlab("Codigos")
```


evista001AliciaAlv
evista002LuisPa
vista003LuciaPiñ
evista004LauraC
ista005PamelaM
trevista06ASFAVII
Trabajo digno
Situación de calle
Salud
Reincidencia
Recursos humanos
Recursos económicos
Protocolos
Políticas Públicas
Masculinización
Institucionalidad académica
Formación académica
Familia
Exclusión social
Efectos cognitivos del encierro
Educación sin castigo
Cuidado del personal técnico
Coordinación interinstitucional
Contexto social
Condiciones de vivienda
Condiciones de encierro
Autoridades
Alfabetización
Adicciones
Códigos
0.0 2.5 5.0 7.5
Frecuencia de códigos por documento