

# Trabajo Final- Análisis de textos de murgas 2023, Iael Klaczko

Curso R aplicado al análisis cualitativo, FCS

## Objetivos del trabajo

- Analizar los textos de las murgas de este año
- Visualizar las diferencias entre murgas que llegan a la última etapa de la competencia y las que no
- Visualizar diferencias entre murgas que tienen integrantes mujeres y murgas que no

## Datos

Los datos a usar son las letras de las murgas que participan del carnaval 2023, como no se encuentran disponibles en ningún lado descargué todas las transcripciones de los espectáculos de cada murga colgados en Youtube. Esto implica que puede haber palabras mal escritas o que no son tales, ya que surgen de un algoritmo que transforma voz en texto automáticamente en Youtube.

Con las letras descargadas armé una tabla que contiene además de las letras, el nombre de la murga, la máxima etapa alcanzada (son tres etapas, ninguna quedó afuera en la primera, por lo que la variable toma dos valores: “segunda” y “liguilla”), y la cantidad de mujeres que integran cada murga.

## Nubes de palabras



##	vamos	gente	bien	hacer	vida	ahora	bueno	siempre	solo	tiempo
##	206	201	199	169	163	161	159	155	136	131
##	puede	hace	quiero	mundo	nunca	verdad	años	cosas	dice	nadie
##	122	122	122	113	102	101	99	98	92	92

Las palabras que más aparecen entre todas las letras son 'vamos', 'gente', 'bien'.

Si miramos la nube de palabras de acuerdo a la máxima etapa alcanzada, las palabras que más se repiten en el caso de las murgas que pasaron a al ligullia son 'público', 'gente' palabras que refieren a grupos de personas. Mientras que las que más se repiten entre los textos de las murgas que no pasaron a la ligullia son 'feliz', 'vergüenza', palabras que tienen que ver con sentimientos.

liguilla

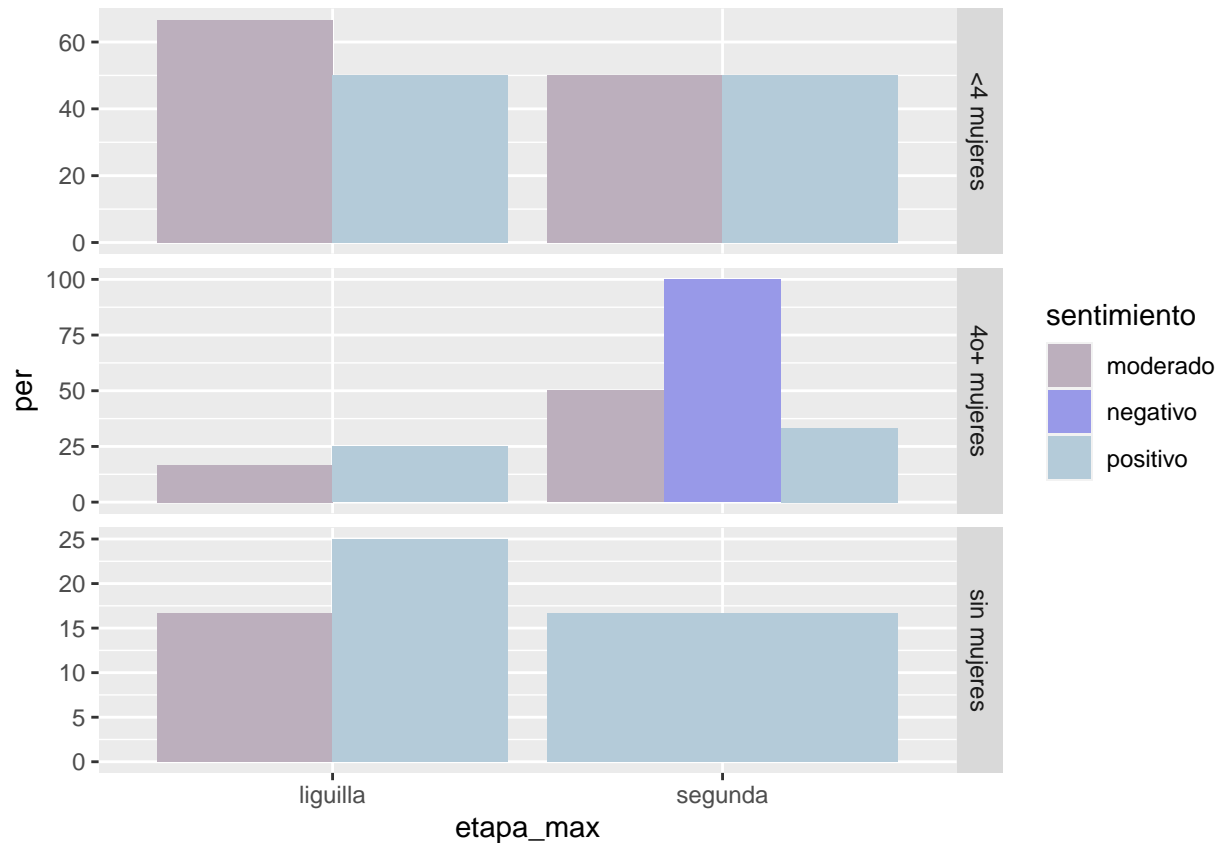


segunda

Al analizar la nube de palabras según murgas con distinta cantidaad de integrantes mujeres encontramos:

Es interesante hacer un análisis de sentimientos para ver si el texto de las murgas es mayoritariamente positivo, negativo o neutral.

Como la categorización en positivo/negativo sólo devolvería un texto como negativo, definí una categorización en tres, tomando como criterio el promedio de puntaje, aquellos textos que están por debajo del promedio son categorizados como 'moderado', los que están por arriba como 'positivo'.



La mayoría de las murgas sin mujeres o con pocas (<4 mujeres) clasifican a la liguilla, sus textos son mayoritariamente positivos, los que no lo son, son moderados. En el caso de las murgas con mayor cantidad de mujeres la mayoría de los textos son moderados.

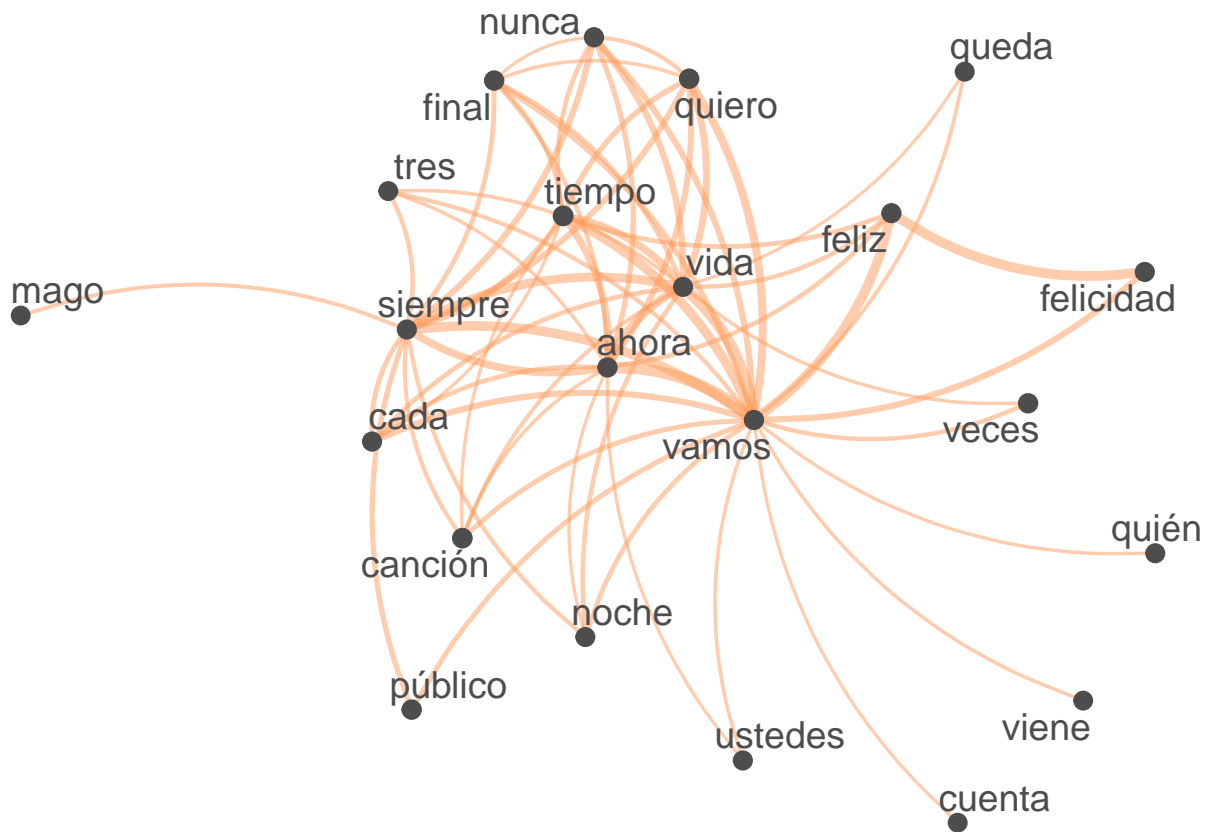
Es interesante ver también ver cómo se asocian las palabras utilizadas por cada murga con palabras clave que tienen que ver con la actualidad de nuestro país y sobretodo lo que ocurrió durante el último año, que es en lo que se basan las murgas para armar sus espectáculos. Voy a buscar asociaciones con las palabras: 'presidente', 'violencia', 'inseguridad', 'corrupción' y 'género'.

##	feature1	feature2	correlation
## 1	rompe	presidente	0.8301988
## 2	regresar	presidente	0.7375456
## 3	próximo	presidente	0.7340908
## 4	despertó	presidente	0.6981640
## 5	unido	presidente	0.6981640
## 6	contentos	presidente	0.6981640
## 7	compartiendo	presidente	0.6981640
## 8	pierde	presidente	0.6889541
## 9	pibes	presidente	0.6697781
## 10	juegan	presidente	0.6621887
## 11	meterse	presidente	0.6451157
## 12	contigo	presidente	0.6397351
## 13	mensajera	presidente	0.6244884
## 14	señales	presidente	0.6244884
## 15	vegetariano	presidente	0.6244884
## 16	sacó	presidente	0.6244884
## 17	invito	presidente	0.6244884
## 18	separé	presidente	0.6244884
## 19	retazo	presidente	0.6244884

Presidente aparece asociado en los textos de las murgas a las palabras 'rompe', 'regresar' y 'próximo'. Violencia aparece asociada con 'pensamos', 'empiezo' y 'tablero'. Inseguridad se asocia con 'cámara', 'ilegal' y 'desafío'. Corrupción se asocia con 'enero', 'penal' y 'europeo'. Por último, género se asocia con 'cifra', 'sospechar', 'cositas'.

Es bastante elocuente que presidente se asocie a palabras que dan la idea de recambio a futuro, violencia con la necesidad de pensar y empezar, inseguridad con cámaras y desafío, género con cifras, como si sólo se tratara de números, con sospechas y con una palabra que se considera inocua y habitante como 'cositas'.

Para finalizar analizo el contexto de las palabras mediante redes de co-ocurrencia, ya que suena razonable no sólo mirar palabras aisladas sino también junto con qué otras palabras se utilizan.



La palabra 'vamos' se encuentra en el centro de la red, siendo la palabra que más conexión tiene con otras, principalmente con 'vida', 'ahora' y 'tiempo'.