**Montevideo, 12 de Agosto de 2023**

**Facultad de Ciencias Sociales**

**UdelaR**

**Recuperación y análisis de texto con R Educación Permanente FCS Consigna de Trabajo Final 2023**

Consigna de Trabajo Final

Soc. Gervasio Riveiro Martínez

**Introducción y fundamentación del trabajo**

Para realizar la entrega de la consigna final del curso de Recuperación y análisis de texto con R de Educación Permanente. Opté por trabajar como tema la presencia de la cuestión Indígena en Uruguay en el discurso del orden parlamentario.

Los motivos que me llevaron a elegir este tema en principio se conectan con estudios de compañeros cercanos de la Facultad de Humanidades, realizando trabajos desde diversas aristas. Considero interesante también, el aporte sociológico para profundizar en una mirada sobre los discursos que ocupan en la agenda estatal y ciudadana este tema.

La cuestión indígena en Uruguay es un tema de relevancia histórica, cultural, sociológica y política. La situación de los charrúas, una de las poblaciones indígenas originarias de Uruguay, ha generado debates y reflexiones sobre la identidad nacional y los derechos humanos. La consideración de un punto de vista sociológico y político sobre esta temática es esencial para comprender cómo las dinámicas sociales y políticas han influido en la percepción y el tratamiento de las identidades y derechos de los pueblos indígenas en el país.

Parte de los nuevos discursos académicos que han circulado en la sociedad Uruguaya plantean la idea de la invisibilización de la presencia indígena en Uruguay. Siendo una cuestión históricamente marginada en la narrativa oficial. A pesar de las adversidades, las comunidades indígenas en Uruguay resisten conformando una conciencia colectiva y reivindicadora en cara a los derechos humanos.

En el caso de la sociología, podemos ver un potencial como herramienta que ayude a mostrar la profundidad de los relatos alternativos a los hegemónicos, sobre la historia y sobre la construcción de identidades en Uruguay. Dando lugar a posibles análisis que incluyan la perspectiva indígena histórica, pero también en el hoy y el presente en la resistencia de sus comunidades.

A nivel de antecedentes solamente mencionare autores como Mónica Michelena, Gonzalo Figueredo y Mónica Sans, entre otros. Nos presentan distinto material que trabaja la cuestión indígena y las transformaciones en las formas de abordaje de esta cuestión desde la academia, pero también desde el ámbito político e histórico en sus relatos.

**Selección de muestra y casos**

Un primer acercamiento trae consigo desde este curso, la posibilidad de trabajar los discursos parlamentarios presentes en los diarios de sesión, siendo materia prima privilegiada para explorar cómo estas cuestiones han sido abordadas y debatidas en el ámbito político Uruguayo, reflejando las tensiones y transformaciones en la sociedad.

Como muestra se presenta una base de datos que conforma 44 sesiones parlamentarias en su formato de diario de sesión, de las mismas se extraen los discursos sistematizados de los parlamentarios declarantes a través de la técnica de scraping aplicada en la web del parlamento. Esta herramienta fue desarrollada por Nicolás Schmidt, politólogo del departamento de ciencias políticas de UdelaR. Todas estas sesiones tienen como temporalidad desde la fecha actual hasta principios de marzo del año Dos mil veinte.

En cuanto a la selección de las sesiones, se realizó una búsqueda en los diarios de sesión en donde se menciona la palabra indígena y charrúa. De la base de datos construida se tomó la variable speech, refiriendo a las intervenciones parlamentarias. Haciendo un conteo podemos encontrar qué palabras y en qué contexto se asocian con las palabras que elegimos para centralizar el trabajo.

**Sobre el código y hallazgos**

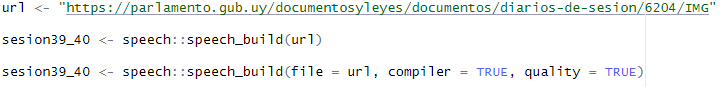
Se instalaron las siguientes librerías:

* Rvest
* Dplyr
* Quanteda
* Readtext
* Stringr
* ggplot2
* quanteda.textstats
* quanteda.textplots

Dentro del código hay instrucciones y explicaciones de las librerías utilizadas, y sobre la utilidad de las funciones aplicadas. Se destaca la utilización de SPEECH y PUY para la construcción de las bases de datos.

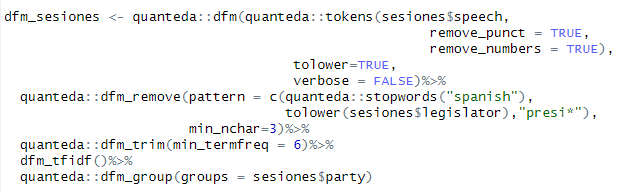


Luego se aplicó una función de speech para obtener el scraping de cada diario de sesión



Es importante aclarar que cada sesión tiene como formato **sesión** **Nº de sesión\_Nº de diario.**

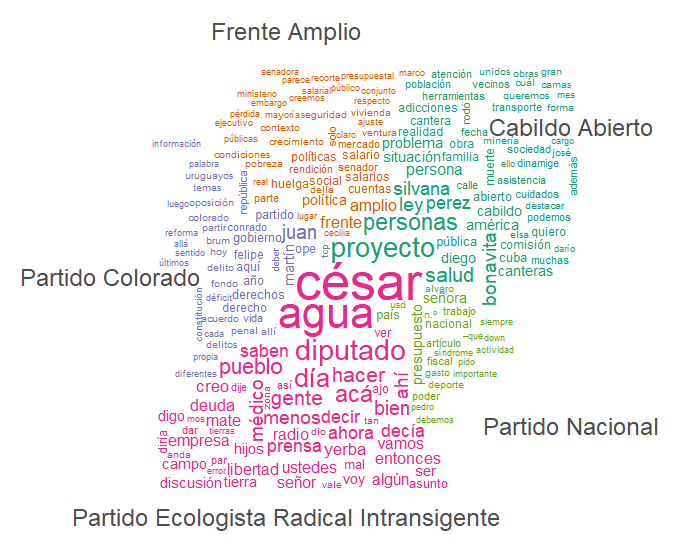
Una vez conjugada toda la base de datos con todos los diarios de sesión sistematizados, se hace una limpieza de los datos obteniendo un data frame quitando palabras que no son relevantes.

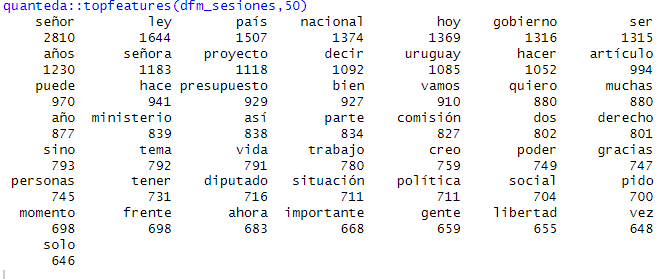


sacopalabras= c(quanteda::stopwords("spanish"),tolower(sesiones$legislator),"presidente", "señor", "vos", "al menos", "creo", "senadora", "senador", "ahora", "nº", "diputado")

Luego aplicamos funciones para visualizar datos. Una de ellas nos da una nube de conceptos que se repiten y que pueden ser mayormente asociados a distintos partidos políticos.

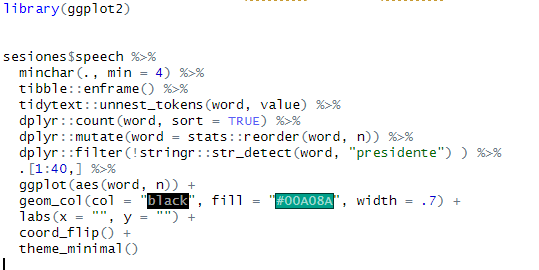




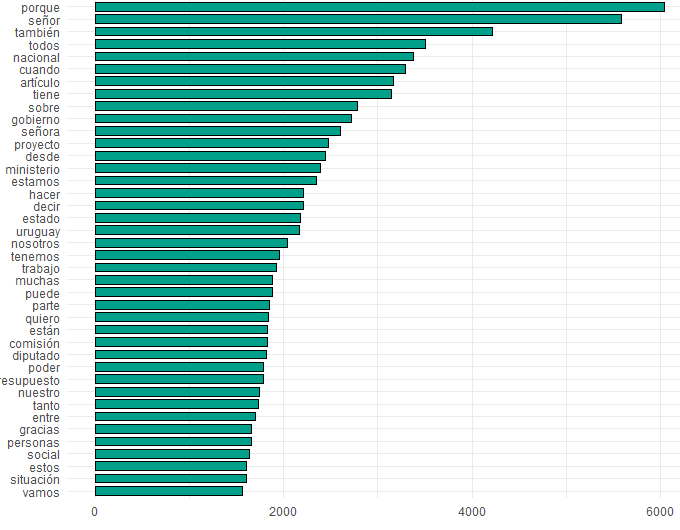


Entendiendo que se puede mejorar mucho la depuración de datos, podemos encontrar algunas palabras relevantes como derecho, vida, proyecto, libertad, dentro de las 50 primeras palabras.

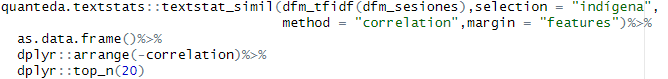
Con ggplot aplicamos una función para visualizar la presencia de las palabras en gráfico de barras



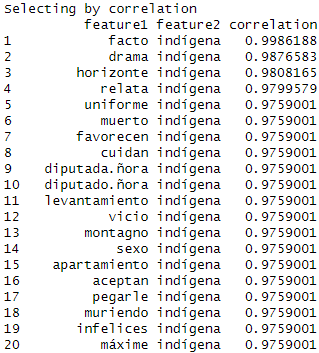
Obteniendo como resultado este gráfico:



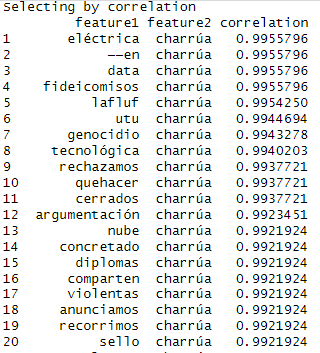
Con la librería quanteda, pudimos hacer una correlación de la palabra indígena y charrúa a través de esta función:



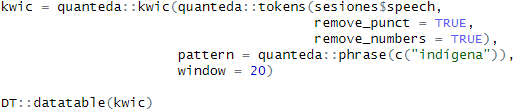
Aquí se muestran las 20 primeras palabras que se correlacionan con mayor intensidad en el caso de **Indígena.**

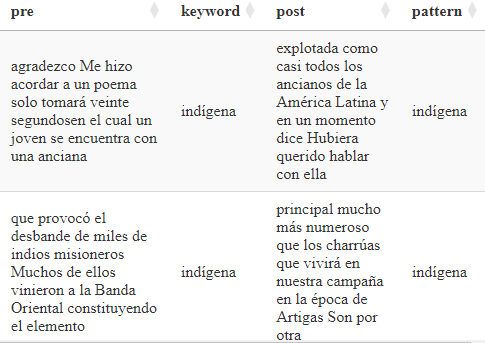


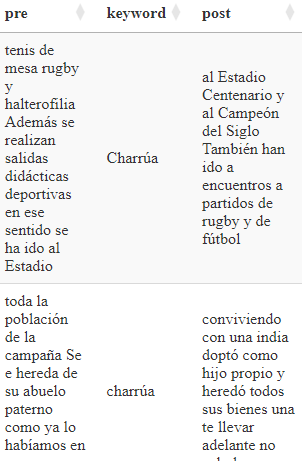
Aquí se muestran las 20 primeras palabras que se correlacionan con mayor intensidad en el caso de **Charrúa.**



Con la siguiente función presentamos el contexto en el que fueron dichas estas palabras







Por último quedan guardadas en formato Excel y Rdata las bases de datos utilizadas. Aquí dejo las funciones que utilicé para las mismas, y la función para combinar dichas bases.

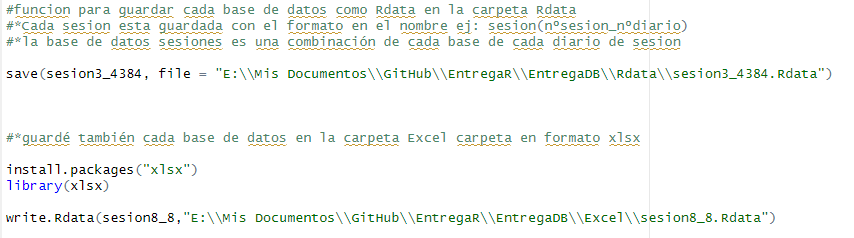
#Combine todas las bases de datos en una sola

sesiones <- rbind(sesion1\_4382, sesion11\_4336, sesion13\_4397, sesion15\_15, sesion15\_4274, sesion16\_17, sesion18\_18, sesion19\_19, sesion19\_4344, sesion2\_4261, sesion21\_21, sesion22\_4406, sesion24\_4408, sesion25\_4409, sesion27\_4470, sesion28\_4412, sesion28\_4471, sesion29\_29, sesion29\_30, sesion29\_4288, sesion3\_3, sesion3\_4328, sesion3\_4384, sesion31\_31, sesion31\_4415,

sesion36\_37, sesion36\_4295, sesion38\_4442, sesion39\_40, sesion4\_4388, sesion44\_4303, sesion48\_4307, sesion49\_4443,

sesion51\_4376, sesion52\_4377, sesion54\_4379, sesion58\_4317, sesion6\_4449, sesion6\_6, sesion7\_4391, sesion7\_4450, sesion7\_7,

sesion8\_4328, sesion8\_4333, sesion8\_8)



Para concluir, el trabajo de procesamiento de los datos, la elección del tema a trabajar y las técnicas que utilicé, hicieron que el trabajo fuese de mi agrado. Considero que es una línea de investigación sumamente interesante, y donde las ciencias sociales tenemos mucho para aportar desde este tipo de análisis.