La prueba técnica consiste en la carga de dos bases de datos (compartidas), y ejecutar comandos de análisis de datos con lenguaje de R (Posit). Hay una base de datos de nombre **Prestadores.xlsx**, en donde se presume es la lista de personas naturales, tanto naturales como jurídicas que fungen como prestadores ante el ADRES. De igual forma se tiene una segunda base de datos de nombre **Municipios.xlsx**, en donde se detalla información sociodemográfica de cada municipio de Colombia.

Se propone analizar la presencia de cada prestador detallado en la base de datos **Prestadores.xlsx** y cruzarlo con la base de datos **Municipios.xlsx**, esto a través de un ejercicio de unión de tablas a través de las clases *Departamento* y *Municipio* de cada registro.

Se crea una base de datos con SQLite, en donde los archivos .xslx mencionados anteriormente son cargados como tablas. Acto seguido se crea una conexión desde la IDE de RStudio y se toma cada tabla como objeto tipo *tibble*, y por medio de las librerías *dplyr* se hacen pipelines de transformación, y con la librería ggplot se hacen resúmenes descriptivos según la información sociodemográfica de cada prestador, agrupados por región y por municipio.

El dataframe obtenido del archivo de prestadores, tiene un tamaño de 60946 registros con 30 campos. De esta información se observa que es imperativo realizar un proceso de estandarización de los campos: nombre del departamento, nombre del municipio, nombre del prestador, y nombre de la razón social asociada al prestador.

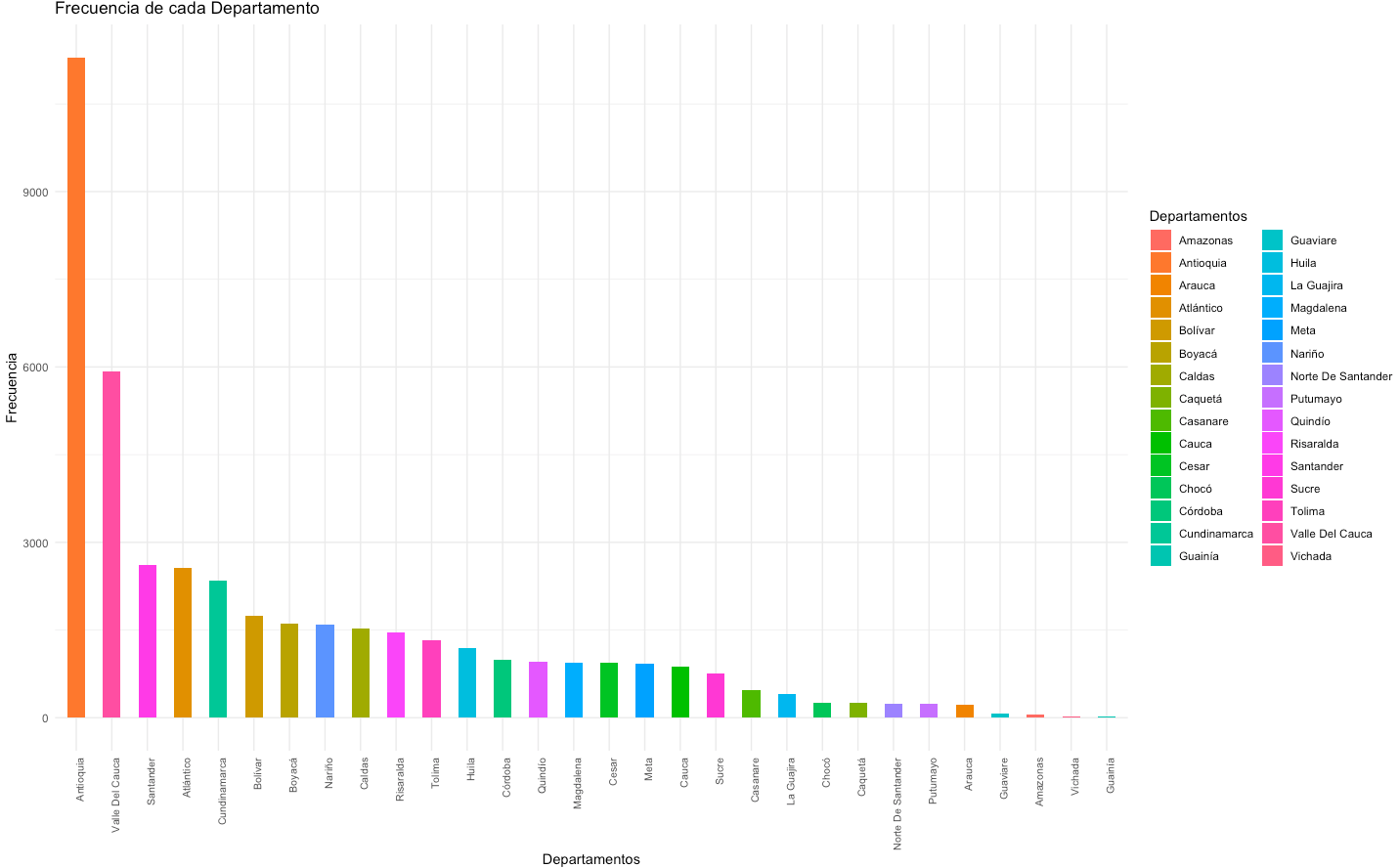
De forma equivalente la pena mencionar que el dataframe derivado del archivo de municipios, requiere atención en la estandarización de sus datos. Éste dataframe tiene un tamaño de 1128 registros con 8 campos. De los cuales los campos nombre del departamento, y nombre de municipios se deben tratar en aras de tenerlo normalizados.

Como referentes auxiliares de los campos nombre de departamento y de municipio, se descarga de la fuente de datos oficial del DANE, el nombre de cada departamento y municipio.

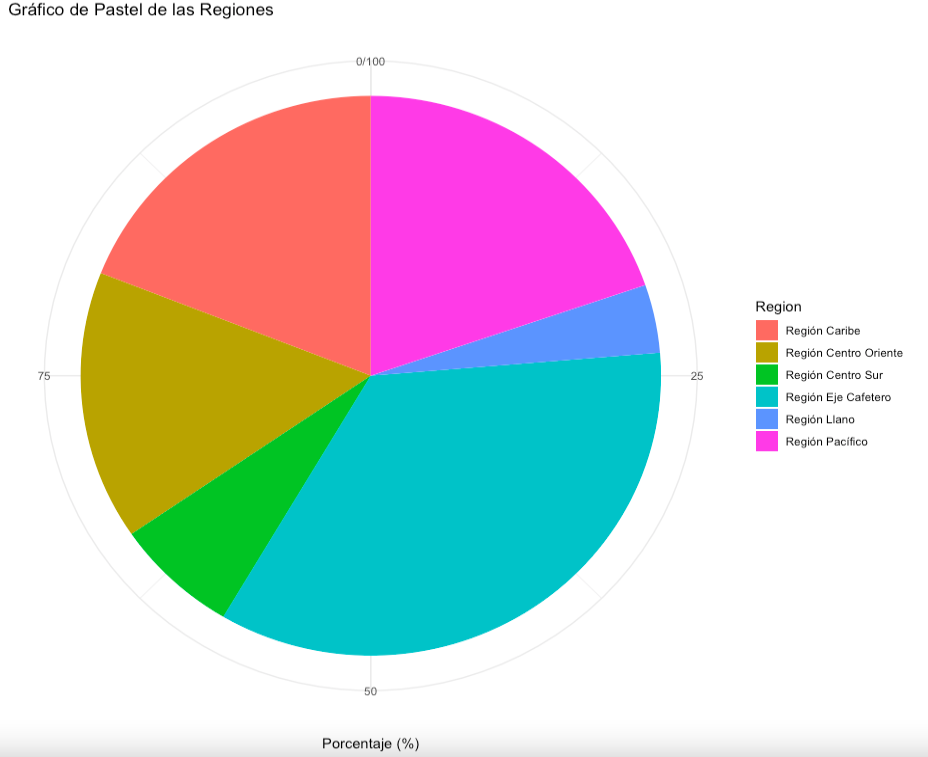
Se realizó el siguiente esquema de transformación de datos: remoción de caracteres especiales > Imputación de nombre de los departamentos (encontrados) como mal ubicados según algunos municipios > Transformación hacia letras minúsculas > Eliminación de espacios al inicio y al final de cada celda > Normalización de los nombres de cada municipio y departamento usando la referencia auxiliar (distancia de Levenshtein mínima) > Transformación de los campos hacia ‘Camel Case’ > Eliminación de duplicados.

Se procede a realizar la unión de los dataframes con bases en las entradas de nombre de municipio y departamento. En teoría el nuevo dataset debería poseer 60946 registros (como el dataset de prestadores), sin embargo, debido a la no colocación adecuada de los nombres de municipio y departamento, persisten errores. Es decir, se deben aplicar estrategias adicionales de corrección de datos. El nuevo dataset

Se utiliza el paquete de ggplot para exhibir algunos hallazgos sociodemográficos de los prestadores registrados.



**Figura 1.** Gráfico de barras de los prestadores en cada departamento.



**Figura 2.** Gráfico de pastel para los prestadores en cada región.