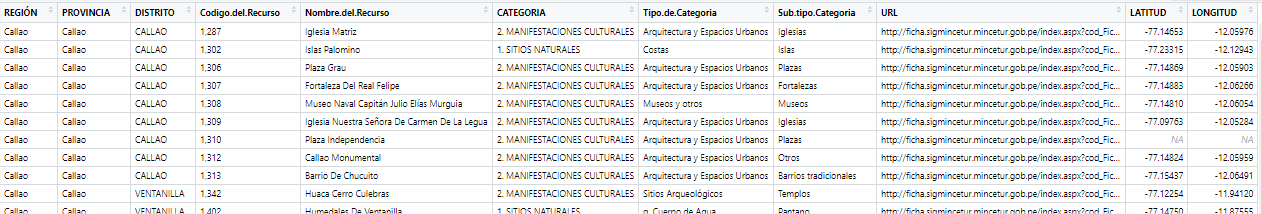
**TALLER 2**

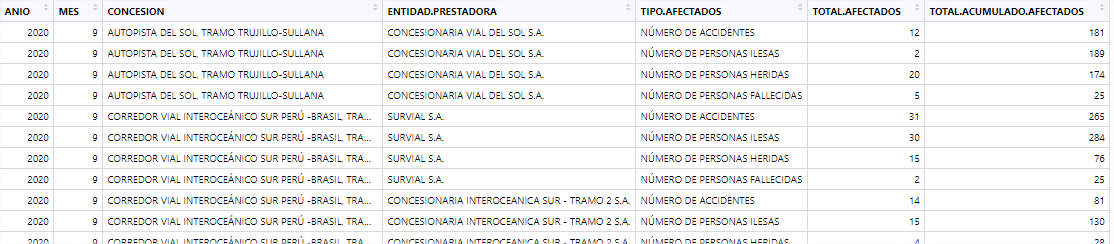
Indicaciones generales

1. Creen una carpeta de trabajo que tenga el primer apellido de cada integrante, por ejemplo, ROJAS\_LOPEZ\_MIRANDA (el grupo es el mismo que el del Taller 1).
2. Coloquen en esta carpeta todos los archivos de datos que corresponden al taller.
3. Creen un proyecto en esta carpeta.
4. Creen un script que contenga el código que permite realizar la lectura de cada conjunto de datos. **Sean ordenad@s.**
5. A lo largo de su trabajo debe utilizar todas las siguientes funciones: read.table o read.delim, read.csv o read\_csv, read.csv2 o read\_csv2, read\_xls, read\_xlsx. Es decir se debe aprecia el uso de por lo menos 5 funciones distintas.
6. El archivo datos\_01.txt contiene una lista de recursos turísticos de la Provincia Constitucional del Callao, los cuales han sido extraídos de la Plataforma Nacional de Datos Abiertos (<https://www.datosabiertos.gob.pe/dataset/inventario-de-recursos-tur%C3%ADsticos/resource/eae059d9-1730-44a5-a268-6516a03a8743>) Lea este archivo en RStudio y almacénelo en datos\_01A (tal vez le sea necesario utilizar el argumento encoding = ‘UTF-8’). Debe obtener un resultado como el que se muestra a continuación:



(la figura solo muestra una parte de los datos)

1. Extraiga la columna Región y explique su estructura (función str)
2. Extraiga el valor de décima latitud y justifique si es double o integer
3. Extraiga el elemento ubicado en el cruce de la cuarta fila **y** séptima columna
4. El archivo datos\_02.txt contiene indicadores mensuales acerca de accidentes de tránsito en el país durante el mes de setiembre del 2020, los cuales han sido extraídos de la Plataforma Nacional de Datos Abiertos (<https://www.datosabiertos.gob.pe/dataset/indicadores-mensuales-de-accidentes-de-trafico-ene-2019-dic-2019-organismo-supervisor-de-0#{}>). Lea este archivo en RStudio y almacénelo en datos\_02A. Debe obtener un resultado como el que se muestra a continuación:



(la figura solo muestra una parte de los datos)

1. Extraiga la segunda columna y explique su estructura (función str)
2. Extraiga los elementos 3 y 4 de la decimosegunda fila
3. Extraiga los elementos de la columna ENTIDAD.PRESTADORA y almacénelos en un objeto de nombre E. Luego, ejecute **table(E)** y explique los resultados.
4. El archivo datos\_03.csv contiene cifras acerca del avance del COVID-19 en los países de Sudamérica, los cuales han sido extraídos del portal web de Ojo Público (puedes leer más acerca del tema en <https://ojo-publico.com/1670/cifras-de-la-pandemia-covid-19-se-extiende-por-sudamerica>). Lea este archivo en RStudio y almacénelo en datos\_03A. Debe obtener un resultado como el que se muestra a continuación:

Tabla

Descripción generada automáticamente

1. Extraiga las cifras de Argentina e indique cuántos pacientes se han recuperado en dicho país
2. Extraiga la columna Positivos y sume las cantidades que hay en esta a fin de indicar el total de casos confirmados en todo Sudamérica
3. Extraiga la cantidad de casos positivos en Uruguay y los recuperados. Luego reste estas cantidades para obtener la cantidad de casos activos (Activos=Positivos – Recuperados)
4. El archivo datos\_04.csv contiene cifras acerca del avance del COVID-19 en los países de Sudamérica, los cuales han sido extraídos del portal web de Ojo Público (puedes leer más acerca del tema en <https://ojo-publico.com/1670/cifras-de-la-pandemia-covid-19-se-extiende-por-sudamerica>). Lea este archivo en RStudio, considerando solo 5 filas de datos. Asimismo, asigne los nombres de columnas que se observan en la imagen. Almacene el objeto en datos\_04A. Debe obtener un resultado como el que se muestra a continuación:

Interfaz de usuario gráfica, Tabla

Descripción generada automáticamente

1. Extraiga la columna Fallecidos e indique si es un vector atómico (use código R)
2. Extraiga la columna Recuperados y sume las cantidades que hay en esta a fin de dividir entre 5 y calcular el promedio de pacientes recuperados por país
3. Cree un vector atómico que contenga el dato de la primera fila y columna, y también el que se ubica en la última fila y columna.
4. El archivo datos\_05.csv contiene cifras acerca del avance del COVID-19 en los países de Sudamérica, los cuales han sido extraídos del portal web de Ojo Público (puedes leer más acerca del tema en <https://ojo-publico.com/1670/cifras-de-la-pandemia-covid-19-se-extiende-por-sudamerica>). Lea este archivo en RStudio, asignando los nombres de columnas que se observan en la imagen. Almacene el objeto en datos\_05A. Debe obtener un resultado como el que se muestra a continuación:

Tabla

Descripción generada automáticamente

1. Extraiga la columna Positivos y almacénela en un vector que se llame P. Haga lo mismo con la columna Fallecidos y almacénela en Fa.
2. Divida Fa/P y almacene el resultado en un vector de nombre LETALIDAD.
3. Obtenga el máximo valor de la columna Activos. Use la función **max**
4. El archivo datos\_06.csv contiene cifras acerca del avance del COVID-19 en los países de Sudamérica, los cuales han sido extraídos del portal web de Ojo Público (puedes leer más acerca del tema en <https://ojo-publico.com/1670/cifras-de-la-pandemia-covid-19-se-extiende-por-sudamerica>). Lea este archivo en RStudio, dejando las columnas con los nombres que se muestran en la figura. Puede que necesite dar alguna indicación acerca de la codificación de los caracteres. Almacene el objeto en datos\_06A. Debe obtener un resultado como el que se muestra a continuación:

Tabla

Descripción generada automáticamente

1. Extraiga la columna X7 e indique si es un vector atómico, lista o data frame.
2. Extraiga los datos que están almacenados en el cruce de las filas 3 y 5 con las columnas 4 y 6.
3. Extraiga la columna X1 y aplíquele la función **length**. ¿Qué significa el resultado que arroja?
4. El archivo datos\_07.xls contiene cifras acerca del avance del COVID-19 en los países de Sudamérica, los cuales han sido extraídos del portal web de Ojo Público (puedes leer más acerca del tema en <https://ojo-publico.com/1670/cifras-de-la-pandemia-covid-19-se-extiende-por-sudamerica>). Lea este archivo en RStudio y almacénelo en datos\_07A. Debe obtener un resultado como el que se muestra a continuación:

Tabla

Descripción generada automáticamente

1. Extraiga la columna de Pruebasrealizadas y verifique si es de tipo numérica
2. Seleccione todas las filas menos la 12 y 13 y todas las columnas menos la de Recuperados.
3. Extraiga la columna Enfermos y aplíquele la función **sort**. ¿Qué significa el resultado que arroja?
4. La hoja Datos del archivo datos\_08.xlsx contiene cifras acerca del avance del COVID-19 en los países de Sudamérica, los cuales han sido extraídos del portal web de Ojo Público (puedes leer más acerca del tema en <https://ojo-publico.com/1670/cifras-de-la-pandemia-covid-19-se-extiende-por-sudamerica>). Lea estos datos en RStudio y almacénelo en datos\_08A. Debe obtener un resultado como el que se muestra a continuación:

Tabla

Descripción generada automáticamente

1. Extraiga la tercera fila y verifique si es un vector. Si no es un vector, ¿qué estructura de datos es?
2. Extraiga la columna Fallecidos y sume las cantidades que hay en esta a fin de dividir entre 13 y calcular el promedio de pacientes fallecidos por país
3. Extraiga la columna Recuperados y aplíquele la función **median**. ¿Qué significa el resultado que arroja?
4. El archivo datos\_09.xlsx contiene cifras acerca del avance del COVID-19 en los países de Sudamérica, los cuales han sido extraídos del portal web de Ojo Público (puedes leer más acerca del tema en <https://ojo-publico.com/1670/cifras-de-la-pandemia-covid-19-se-extiende-por-sudamerica>). Lea estos datos en RStudio, considerando solo 7 filas y los nombres de columnas que se muestran en la imagen. Almacene el objeto en datos\_09A. Debe obtener un resultado como el que se muestra a continuación:

Tabla

Descripción generada automáticamente

1. Verifique si este objeto es un data frame. Asimismo, verifique si es un tibble (sugerencia: use is\_tibble)
2. Extraiga la columna País y verifique si es un vector atómico
3. Extraiga la columna Pruebas y verifique si es de tipo lógico
4. La hoja Datos del archivo datos\_10.xlsx contiene cifras acerca del avance del COVID-19 en los países de Sudamérica, los cuales han sido extraídos del portal web de Ojo Público (puedes leer más acerca del tema en <https://ojo-publico.com/1670/cifras-de-la-pandemia-covid-19-se-extiende-por-sudamerica>). Lea estos datos en RStudio y almacénelos en datos\_10A. Debe obtener un resultado como el que se muestra a continuación:

Tabla

Descripción generada automáticamente

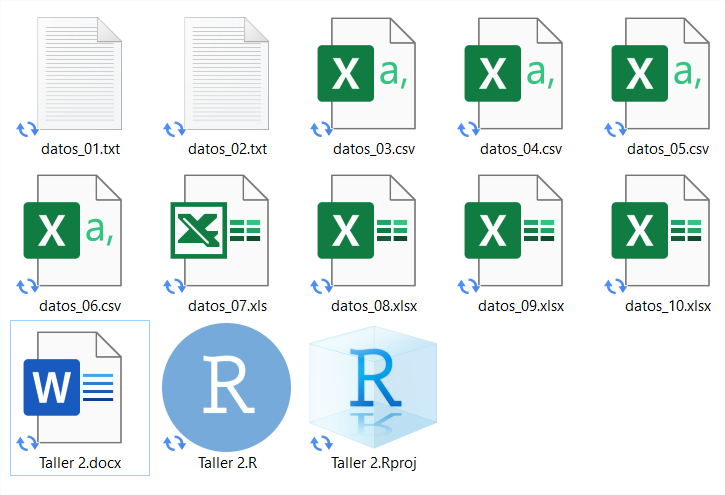
1. Aplique la función is.na al objeto datos\_10A y almacene el resultado en un objeto de nombre **NO** ¿Cuál es el resultado y para qué sirve esta función?
2. Sume los valores de **NO** e interprete el resultado.
3. Extraiga la columna Pruebasrealizadas y aplique la función **sum**. ¿Qué resultado arroja y por qué?

¿Cómo entregar el trabajo?

Una vez que termine sus códigos, verifique que en su carpeta se encuentre:

* Los 10 archivos de datos
* El archivo de códigos R
* El proyecto Rstudio

Ejemplo:



Luego, comprima la carpeta y súbala en el enlace de **Taller 2** en el aula virtual