

Actividad Individual 12

Esta actividad tiene que quedar disponible en su repositorio de GitHub con el resto de las actividades y tareas del curso. Asegurate que tanto Mauro como yo seamos colaboradoras de tu proyecto `Tareas_STAT_NT` creado hace dos semanas. Recordar seleccionar en las opciones de proyecto, codificación de código UTF-8. Recordar que para que tengas la última versión de tu repositorio en GitHub debes hacer pull a tu repositorio para tener la copia localmente en tu computadora y no generar inconsistencias y antes de terminar subir tus cambios con `commit` y `push`.

Ejercicio 1

Utilizamos una base de datos extraída del **Estudio Longitudinal de Bienestar en el Uruguay** (IECON), el cual consiste en un relevamiento longitudinal representativo de los niños que concurren al sistema de educación primaria pública.

Se recoge información referente a múltiples dimensiones del bienestar de los niños en la muestra así como también del resto de los integrantes de sus hogares, entre los que se destacan:

- situación nutricional.
- habilidades cognitivas y no cognitivas.
- imaginación; razonamiento y sentimientos.
- actividades de ocio e interacción social.
- logros educativos.
- ingresos.
- calidad de la vivienda y bienestar subjetivo, entre otros.

La información es relevada en Olas, la primera fue en 2004 a 3000 niños del primer grado de escuela. En las distintas olas se los sigue en el tiempo a los mismos niños (por eso es longitudinal).

- De **personas**, con información referente al niño y personas del hogar donde reside.
- De **hogares**, con información más enfocada al niño, sus capacidades de relacionamiento, emociones, amigos, etc.

Vamos a usar datos de la tercer ola (2012). La base de personas contiene información sobre las personas del hogar del niño que salió en la muestra. Contiene la variable `nform` (número de formulario) que permite identificar al **hogar** (es una *key*). Para cada hogar existe la variable `nper` (número de persona) que le asigna un número a cada **persona del hogar**.

La base de hogares contiene información de los hogares.

Tiene la variable `nform` (número de formulario) que permite identificar al hogar y la variable `nper` (número de persona encuestada) que permite identificar a la persona del hogar que fue encuestada.

1. Los datos están en `EVA_Personas_OLA_3.csv`, usá la función `read_csv` del paquete `readr` para leer los datos.
2. Usando `select()` seleccioná estas variables:
 - Número de formulario `nform`
 - Número de persona `nper`
 - Sexo `f12`
 - Edad `f11`
 - Situación conyugal `f2`

- Relacion de parentezco con el jefe/a **f10**
 - Nivel educativo que cursa o cursó **f4**
 - Ocupación principal / lo que es en OP/ Tareas. **g1, g4, g5**
 - Jubilaciones pensiones **h3**
 - Transferencia de otros hogares, del país y externos al país **h7 y h8**
 - Monto asignación familiar **h9a**
 - Beneficio en especia **h10**
3. Los nombres de las variables nos confunden, se los cambiamos con la función **rename()** del paquete **dplyr**. Guardalo en un nuevo objeto **personas_rename**
- **sit.conyugal:** **f2**
 - **nivel.educ:** **f4**
 - **parent.jefe:** **f10**
 - **edad:** **f11**
 - **sexo:** **f12**
 - **dedicacion.p:** **g1**
 - **tipo.p:** **g4**
 - **tareas.p:** **g5**
 - **jub.pen:** **h3**
 - **transf.pais:** **h7**
 - **transf.ext:** **h8**
 - **afam:** **h9a**
 - **benef.esp:** **h10**
4. Recodificá la variable **dedicacion.p** en **personas_rename** usando **case_match** guardá los datos en el objeto **personas_recode** como sigue
- 1 trabaja
 - 2 desocupado
 - 3 estudia
 - 4 jubilado
 - 5 pensionista
 - 6 t.hogar
 - 7 nada
5. Realizar un gráfico de barras de la variable **dedicacion.p** en del objeto **personas_recode** ordenado por frecuencia filtrando la categoría **NA**. Agregando nombre apropiados en los ejes.
6. ¿Cómo podemos ver la distribución de las edades de los que no hacen nada?
7. Recodificá edad en tramos [0-3), [3, 6), [6,12), [12, 15), [15, 18) , [18,21), [21,30) y [30, +) guardalo en **personas_recode_ed** e identificá en que tramo se da la mayor frecuencia
8. Tenemos información sobre los jefes de hogar para cada hogar correspondiente a los niños de la muestra. Nos interesa descubrir si hay una diferencia por sexo, es decir si hay predominancia de hogares con jefe de hogar hombre o mujer.
- Filtren la variable **parent.jefe** para los casos que toma valor 1 que corresponde a los jefes de hogar.
 - Agrupen por sexo.
 - Contamos cuantos casos hay en cada categoría (9=NS/NC).
 - Creamos una variable que indique el porcentaje.
9. Nos gustaría tener la variable **departamento** que indica el departamento al cuál pertenece el hogar de la muestra.

Tenemos esta información en la **base de Hogares** y una variable de macheo que es **nform**, la cuál indica el **número de formulario** y vale lo mismo para todos los miembros del hogar.

Nos interesa mantener todas las observaciones de la base de personas, y “pegarle” el departamento al cuál corresponde la función. Para eso usar `left_join`.

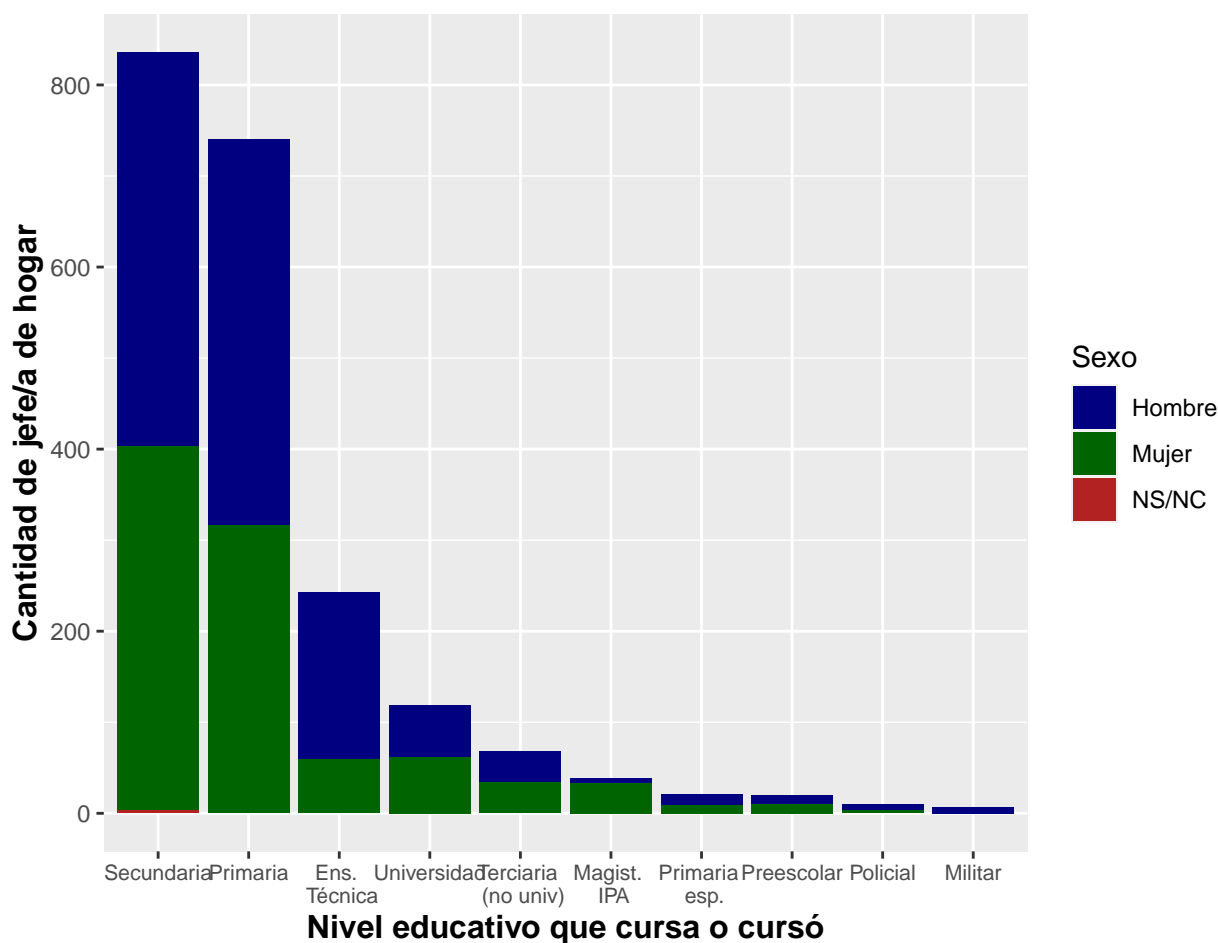
La base de `hogares` está en `EVA Hogares_01a_3.csv`

10. La variable departamento está codificada según INE. Queremos cambiar dicha codificación por el nombre del departamento.

`forcats` contiene funciones que nos permiten trabajar con factores.

Usamos la función `fct_recode` de la librería `forcats` para codificar. Montevideo 1, Artigas 2, Canelones 3, Colonia 5, Florida 8, Paysandu 11, Rivera 13 y Soriano 17

11. Utilizando funciones de `dplyr`, reportá una tabla que tenga la información de la proporción de Jefes/as según sexo para cada departamento, el valor 1 de la variable `parent.jefe` corresponde al jefe/a de hogar. La tabla debe contener cuatro columnas (`dept_cod`, `sexo`, `conteo` y `prop.`). Guardá el objeto generado con nombre `tabla`.
12. Hace un gráfico que muestre la información anterior
13. Replicá el siguiente gráfico de barras del nivel educativo del jefe/a de hogar según sexo.



Pasos para el gráfico:

- Filtrar a los jefes que responden la encuesta y los que tienen `nivel.educ` diferente de 99.
- Hacer el gráfico de barras apilados usando `sexo` para ello
- Ordenar las barras según frecuencia con `forcats`

- Usando **scales** cambiale las etiquetas a la variable **nivel.educ**
- usando **theme** el título de los ejes en negrita, en tamaño 12 y el tamaño del texto en los ejes es 8.
- El color debe ser definido con **scale** para Hombres (navyblue), Mujeres (darkgreen) y NS/NC (firebrick)