

Prueba extra 2021

NOMBRE:

19/6/2021

Explicativo sobre la prueba

Compromiso de honestidad del estudiante:

Esta prueba es de carácter *INDIVIDUAL*, cada estudiante debe completarla y no interactuar con el resto de los estudiantes ni otras personas mientras dure la prueba. La evaluación es con material, lo que implica que pueden utilizar todo lo que hemos visto en clase, los libros del curso, las presentaciones y las soluciones de actividades y tareas. Sin embargo *NO ESTÁ PERMITIDO* utilizar como material las respuestas, comentarios o sugerencias de otras personas que no sean docentes del curso (consultar a natydasilva@gmail.com), cualquier apartamiento de esto invalidará la prueba .

Entrega de la Prueba

Debes completar tu nombre en el YAML del archivo donde dice `author: "NOMBRE: "`. Los archivos y la información necesaria para desarrollar la prueba se encuentran en Eva en la pestaña Prueba_Extra. La revisión debe quedar en tu repositorio PRIVADO de GitHub en una carpeta que se llame Prueba_Extra con el resto de las actividades y tareas del curso. Parte de los puntos de la prueba consisten en que tu revisión sea completamente reproducible cuando clonemos tu repositorio. Asegurate de subir los archivos necesarios para la reproducibilidad (los datos van a ser necesarios). A su vez asegurate que tu repositorio de GitHub esté bien organizado.

La prueba dura 4 horas (10:00 a 14:00 hs), a las 14 deben enviarme el archivo .pdf y el .Rmd a natydasilva@gmail.com y por favor recordame tu usuario de GitHub para que sea más sencillo encontrar tu repositorio con la prueba.

Recordar que para que tengas la última versión de tu repositorio debes hacer `pull` a tu repositorio para no generar inconsistencias y antes de terminar subir tus cambios con `commit` y `push`. Hay puntos parciales por lo que si algunas de las respuestas de código te queda incompleta agregá en el entorno de código `eval=FALSE` para que no se evalúe el mismo y evitar posibles errores.

Ejercicio 1 (80 puntos)

Explicativo sobre los datos para el Ejercicio 1

Los datos que vamos a utilizar son del catálogo de datos abiertos (<https://catalogodatos.gub.uy>). Específicamente utilizaremos los datos de la Encuesta de Conocimientos, Actitudes y Prácticas de Ciudadanía Digital en Uruguay realizada por AGESIC. En la carpeta Revisión de la semana 10 encontrarás dos archivos `basededatosobcuiuca2013.csv` que contiene los datos de la encuesta y el archivo `metadatoscap2013.csv` que contiene los metadatos de dicha encuesta (descripción de variables, codificación, preguntas, etc).

(1 punto) *Pregunta 1*

Leer los datos `basededatosobscuicap2013.csv` usando la función `read_delim()` Para que la vida te sonría para arrancar la prueba, te sopló la primer respuesta para leer los datos (solo tenés que cambiar DATOS).

```
datadigital<- read_delim(file = "DATOS", delim = ";",
                        escape_double = FALSE,
                        locale = locale(encoding = "IBM437"), trim_ws = TRUE)
```

(2 puntos) *Pregunta 2*

Vamos a trabajar con un subconjunto de los datos. Generá un nuevo objeto llamado `digital_red` que contenga las variables desde `P1V15Departamento` hasta `P24V40` (en el archivo `metadatoscap2013.csv` podés encontrar el nombre de todas las variables junto con su descripción). El conjunto de datos reducidos (`digital_red`) debe tener 1006 filas y 30 columnas.

De ahora en más trabajaremos con el conjunto de datos reducidos `digital_red`

(7 puntos) *Pregunta 3*

Usando funciones del paquete `ggplot2`, hacé un gráfico de barras que contenga la cantidad de personas que utilizó alguna vez internet (`P11V25Utilizóvezinternet`). Cambiá los ticks del eje `x` donde dice 1 debe decir Si y donde dice 2 debe decir No y en el resto de los ticks nada. Para esta parte debes notar que la variable `P11V25Utilizóvezinternet` es continua. Cambiá la etiqueta del eje `y` a `Cantidad` y la del eje `x` a `Usó internet`.

El gráfico debe quedarte como sigue y en vez de “Gráfico a replicar” (`caption`) debes agregar un título que describa la figura y un comentario sobre la misma.

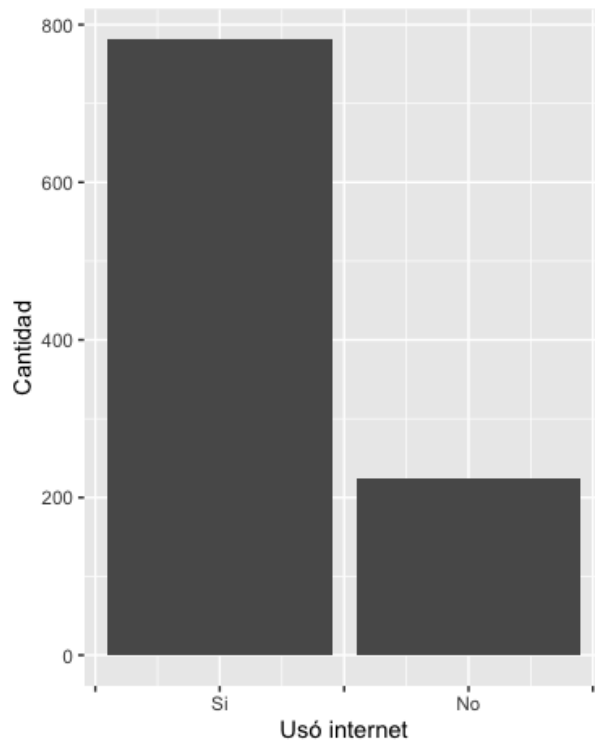


Figure 1: Gráfico a replicar

(7 puntos) Pregunta 4

Esta pregunta tiene tres partes:

1. Renombrá la variable `P11V25Utilizóvezinternet` como `usoint`
2. Transformá la variable `usoint` en una variable de tipo factor
3. Recodificá la variable `usoint` para que 1 sea Si, 2 sea No y 99 sea NA
4. Luego de todos los pasos anteriores repetí el gráfico de la Pregunta 3

(7 puntos) Pregunta 5

Usando funciones de `dplyr` respondé. ¿Qué proporción de personas utilizó alguna vez internet?

(10 puntos) Pregunta 6

Usando funciones de `dplyr` respondé. ¿Qué proporción de personas que utilizaron alguna vez internet son de Montevideo?

(10 puntos) Pregunta 7

Utilizando `rename_with` de `dplyr` cambiá el nombre de todas las variables que contienen `P14V30Lugaresutilizóinternet_` por `lugar_int`, por ejemplo la variable que se llama `P14V30Lugaresutilizóinternet_M4` se debe llamar `lugar_int_M4` y de forma similar el resto. Pista usá `gsub`

(12 puntos) *Pregunta 8*

Usando `ggplot2` realiza una visualización apropiada a tu elección para ver la distribución de las edades de las personas que nunca usaron internet según sexo. Debes incluir etiquetas apropiadas para los ejes y título. Describe brevemente que ves en el gráfico y comentá si hay o no diferencias en la edad entre hombres y mujeres.

A su vez usando funciones de `dplyr` respondé ¿Cuál es el promedio de edad de las personas que nunca usaron internet para cada sexo?

(14 puntos) *Pregunta 9*

Usando `ggplot2` replicá el siguiente gráfico para visualizar la distribución de la variable `P8V22Hogarconexiónainternet` por departamento. Incluí un título apropiado para el gráfico y un breve comentario de lo que se puede observar en el mismo. La codificación de la variable `P1V15Departamento` está en los metadatos por simplicidad miralos en la siguiente tabla:

| Código | Departamento |
|--------|----------------|
| 1 | Artigas |
| 2 | Canelones |
| 3 | Cerro Largo |
| 4 | Colonia |
| 5 | Durazno |
| 6 | Flores |
| 7 | Florida |
| 8 | Lavalleja |
| 9 | Maldonado |
| 10 | Montevideo |
| 11 | Paysandú |
| 12 | Río Negro |
| 13 | Rivera |
| 14 | Rocha |
| 15 | Salto |
| 16 | San José |
| 17 | Soriano |
| 18 | Tacuarembó |
| 19 | Treinta y Tres |

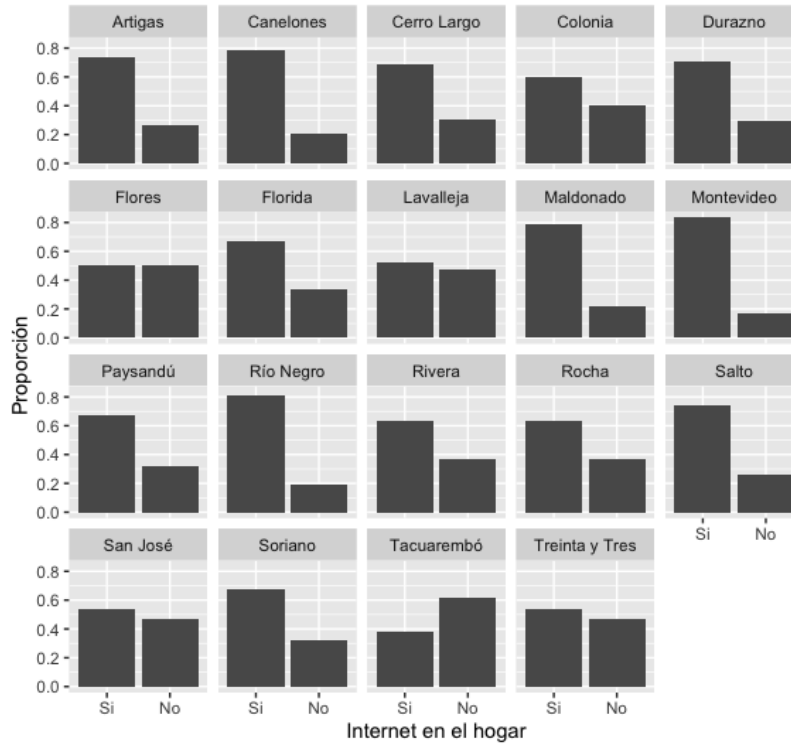


Figure 2: Gráfico a replicar

(10 puntos) *Pregunta 10*

Queremos saber con que frecuencia las personas usan internet para trabajar. Para esto:

- Se utilizó la variable `P22V38Trabajar` y dado que la misma contiene muchos valores `#¡NULO!` y `99`, se filtraron los valores diferentes a éstos.
- Posteriormente se realizó un gráfico de barras de `P22V38Trabajar` (filtrada) como el que aparece a continuación.

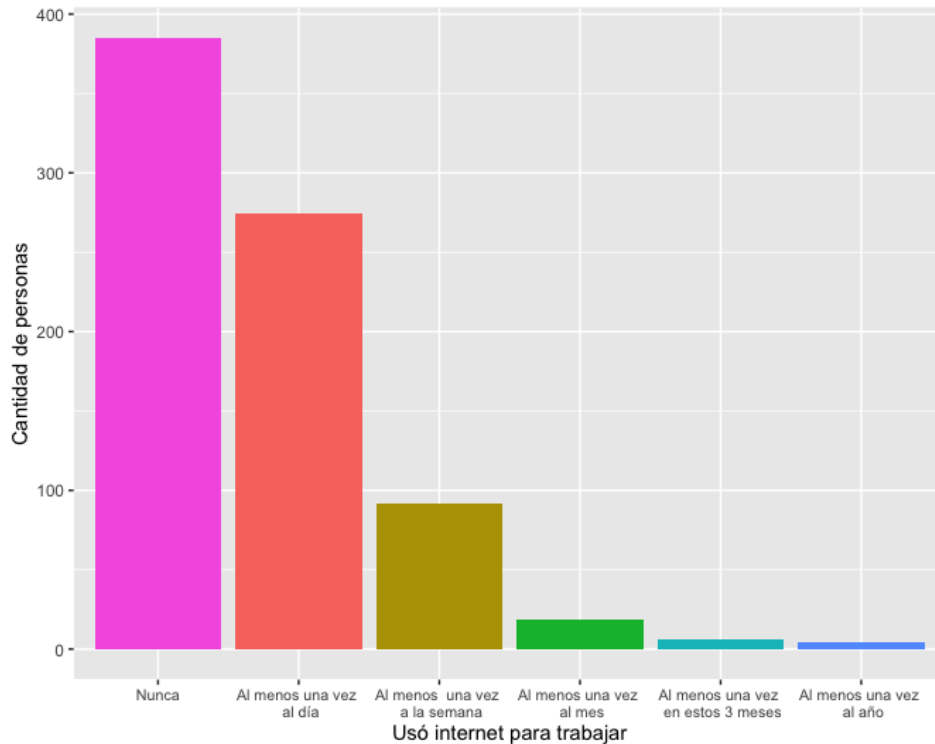


Figure 3: Gráfico de barras con la cantidad de personas según frecuencia de uso de internet para trabajar.

En base a esta figura respondí las siguientes preguntas:

1. ¿Qué argumento del gráfico de barras, en `aes()`, me permite colorear las barras como se ven en la Figura 3? (Pista, aquí los colores usados son por defecto).
2. Incluí la línea de código que corresponde a definir el tamaño del texto del eje x en 8.
3. Incluí la línea de código que define las etiquetas que se muestran en el eje x (Pista, observá que usando `\n` se genera un salto de renglón para que las etiquetas queden como se ve en el gráfico).

Ejercicio 2 (20 Puntos)

1. Indique cuál de las siguientes opciones NO corresponde a partes de un gráfico estadístico.

Seleccione una:

- a) Datos, mapeo estético y sistema de coordenadas
- b) Datos, mapeo estético y escalas
- c) Datos, mapeo estético y modelo estadístico
- d) Datos, mapeo estético y temas

2. Indica cuál de las opciones describe mejor lo que hace la función

```
func1 <-function(a){
  for(i in 1:a){
    b <- i^2
  }
  print(b)}
}
```

- a. Calcula un número al cuadrado
 - b. Calcula una secuencia de números al cuadrado
 - c. Calcula el producto de a por 2
 - d. Calcula una secuencia de números al cuadrado para valores positivos de a
3. Hacé una función llamada `simulaplot` que tenga como argumentos `n` tamaño de muestra, `sigma` el desvío y `forma`. Esta función simula datos para distintos tamaños de muestra con distinta media y desvío de una distribución normal usando `rnorm` y simula la misma cantidad de observaciones n de una distribución con `rgamma` cambiando el parámetro de forma. Como resultado de esta función se hace un diagrama de dispersión con x los datos simulados normales y y los datos gamma. A su vez si el argumento es `sigma` es negativo la función no debe ser evaluada y debe imprimir el mensaje “sigma debe ser positivo”.