# Revisión

*NOMBRE: Daniel Czarnievicz*

*24/5/2019*

### Explicativo sobre la prueba

Por favor completá tu nombre en el preámbulo del archivo donde dice `author: "NOMBRE: "`. El examen es individual y cualquier apartamiento de esto invalidará la prueba. Puede consultar el libro del curso durante la revisión http://r4ds.had.co.nz así como el libro de ggplot2 pero no consultar otras fuentes de información.

Los archivos y la información necesaria para desarrollar la prueba se encuentran en Eva en la Semana 10.

La revisión debe quedar en tu repositorio PRIVADO de GitHub en una carpeta que se llame Revisión con el resto de las actividades y tareas del curso. Parte de los puntos de la prueba consisten en que tu revisión sea reproducible y tu repositorio de GitHub esté bien organizado.
Además una vez finalizada la prueba debes mandarme el archivo pdf y Rmd a natalia@iesta.edu.uy .

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`.

### Explicativo sobre los datos

Los datos que vamos a utilizar en la prueba son los que trabajó Lucía en la clase de repaso de `tidyverse`. Estos datos son extraídos del Estudio Longitudinal de Bienestar en el Uruguay llevado a cabo por el Instituto de Economía (iecon), el cual consiste en un relevamiento longitudinal representativo de los niños que concurren al sistema de educación primaria pública.

La información es relevada en Olas en este caso vamos a usar datos de la tercer ola (2012) que contiene bases de personas, con información referente al niño y personas del hogar donde reside. Los meta datos con información sobre las variables se encuentra en el archivo ola3_meta.csv.

## Preguntas

1. Usando la función `read_csv` del paquete `readr` cargá la base de datos `ola_3.csv` que se encuentra disponible en el EVA y a estos datos nombralos `personas`.

```
personas <- readr::read_csv(file = "ola_3.csv")
```

2. Renombrá la variable `dpto_cod` como `depto`.

```
personas <- rename(personas, depto = dpto_cod)
```

3. La variable `sexo` tiene tres valores, recodificala para que el 1 sea `M` el 2 sea `F` y 9 sea `NS/NC` (no sabe). Guardá los nuevos datos en `personas_reco`.

```
personas_reco <- mutate(personas, sexo = if_else(sexo == 1, "M", if_else(sexo == 2, "F", "NS/NC")))
```

4. Usando funciones de `dplyr` respondé ¿Cuál es la proporción de personas según sexo?

```
personas_reco %>%
   group_by(sexo) %>%
   tally() %>%
   mutate(prop = n / sum(n)) %>%
```

```
  select(-n) %>%
  knitr::kable(digits = 2, caption = "Proporción de personas en la base según sexo",
               col.names = c("Sexo", "Proporción"))
```

Table 1: Proporción de personas en la base según sexo

| Sexo | Proporción |
|---|---|
| F | 0.52 |
| M | 0.48 |
| NS/NC | 0.00 |

5. Utilizando funciones de `dplyr`, reportá una tabla (con `xtable`) 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 (Departamento, Sexo, Conteo y Proporción). Guardá el objeto generado con nombre `tabla`.

La tabla debe ser igual a la siguiente:

| Departamento | Sexo | Conteo | Proporción |
|---|---|---|---|
| Artigas | F | 49 | 0.24 |
| Artigas | M | 151 | 0.76 |
| Canelones | F | 88 | 0.42 |
| Canelones | M | 122 | 0.58 |
| Colonia | F | 30 | 0.32 |
| Colonia | M | 65 | 0.68 |
| Florida | F | 73 | 0.59 |
| Florida | M | 50 | 0.41 |
| Montevideo | F | 391 | 0.45 |
| Montevideo | M | 474 | 0.55 |
| Paysandu | F | 109 | 0.41 |
| Paysandu | M | 159 | 0.59 |
| Rivera | F | 161 | 0.59 |
| Rivera | M | 107 | 0.39 |
| Rivera | NS/NC | 3 | 0.01 |
| Soriano | F | 33 | 0.38 |
| Soriano | M | 55 | 0.62 |

Figure 1: Tabla a replicar

```r
tabla <- personas_reco %>%
   mutate(parent.jefe = if_else(parent.jefe == 1, 1, 0)) %>%
   filter(parent.jefe == 1) %>%
   group_by(depto, sexo) %>%
   count(parent.jefe) %>%
   select(-parent.jefe) %>%
   group_by(depto) %>%
   mutate(prop = n / sum(n)) %>%
   rename(`Departamento` = depto, `Sexo` = sexo, `Conteo` = n, `Proporción` = prop)
print(xtable::xtable(x = tabla, caption = "Proporción de Jefes de hogar por departamento según sexo", di
      comment = FALSE)
```

| | Departamento | Sexo | Conteo | Proporción |
|---|---|---|---|---|
| 1 | Artigas | F | 49 | 0.24 |
| 2 | Artigas | M | 151 | 0.76 |
| 3 | Canelones | F | 88 | 0.42 |
| 4 | Canelones | M | 122 | 0.58 |
| 5 | Colonia | F | 30 | 0.32 |
| 6 | Colonia | M | 65 | 0.68 |
| 7 | Florida | F | 73 | 0.59 |
| 8 | Florida | M | 50 | 0.41 |
| 9 | Montevideo | F | 391 | 0.45 |
| 10 | Montevideo | M | 474 | 0.55 |
| 11 | Paysandu | F | 109 | 0.41 |
| 12 | Paysandu | M | 159 | 0.59 |
| 13 | Rivera | F | 161 | 0.59 |
| 14 | Rivera | M | 107 | 0.39 |
| 15 | Rivera | NS/NC | 3 | 0.01 |
| 16 | Soriano | F | 33 | 0.38 |
| 17 | Soriano | M | 55 | 0.62 |

Table 2: Proporción de Jefes de hogar por departamento según sexo

6. ¿Como podrías mostrar en una visualización la información de la tabla anterior para comparar la proporción de hombres y mujeres por departamento? Recordá poner nombre apropiados a los ejes y subtitulo (caption) que contenga el nombre de la figura y que información se muestra en la misma. Haz un comentario sobre lo que se observa.

```r
personas_reco %>%
   filter(parent.jefe == 1) %>%
   ggplot() +
   geom_bar(aes(fct_infreq(depto), y = ..count.. / sum(..count..), fill = as.factor(sexo)),
            position = "fill") +
   labs(y = "Proporción de jefes de hogar", x = "Departamento", fill = "Sexo") +
   coord_flip()
```

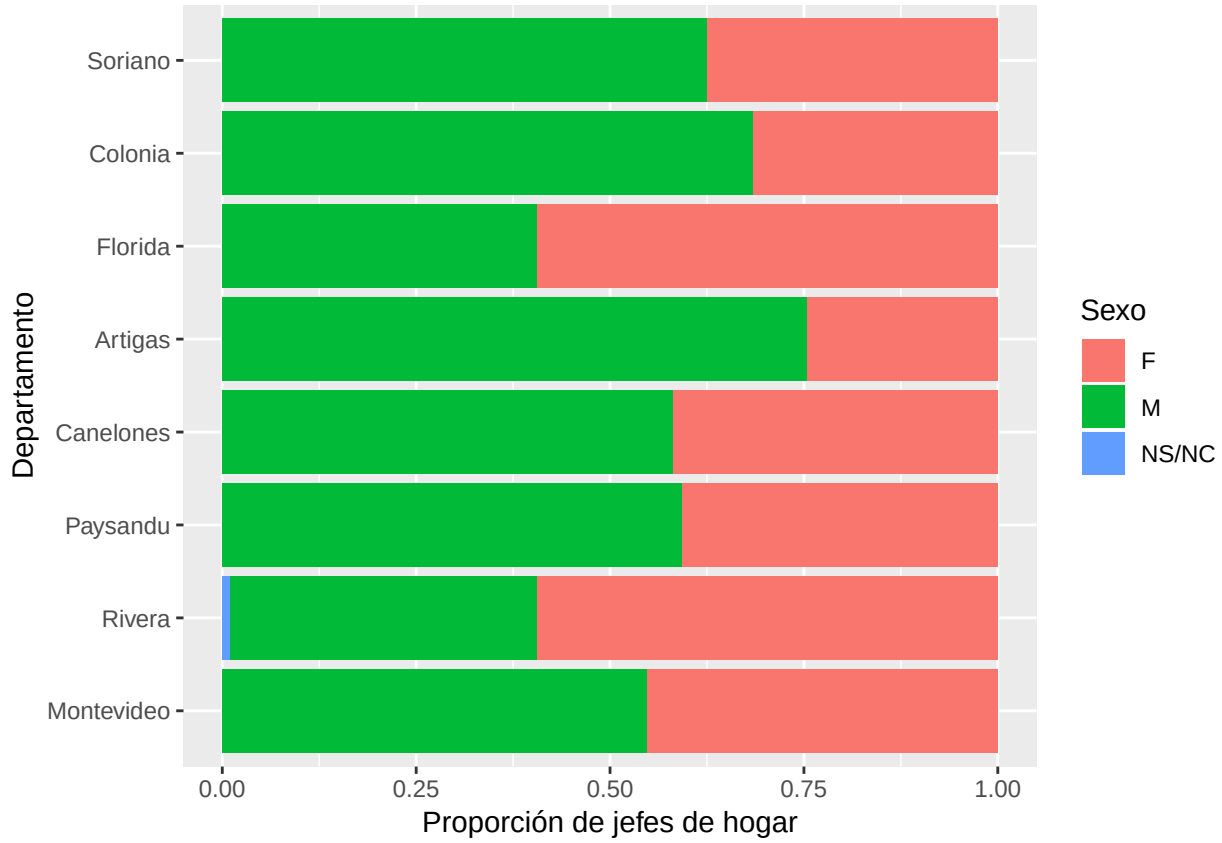


Figure 2: Gráfico de barras apiladas de la proporción de jefes de hogar según sexo y departamento. Se observa que para la mayoría de los departamentos en la muestra, hay una mayor proporción de jefes varones.

7. Usa la función `spread` de manera que en el objeto `tabla` generado en el punto anterior queden los departamentos como filas, el sexo como columnas (3 columnas: M, F, NS/NC) y en las celdas los valores de la variable `proporción`. ¿Obtenemos de esa manera un data set ordenado? ¿Por qué?

```
select(tabla, -Conteo) %>%
   spread(key = Sexo, value = `Proporción`)
```

```
# A tibble: 8 x 4
# Groups:   Departamento [8]
  Departamento     F     M `NS/NC`
  <chr>        <dbl> <dbl>   <dbl>
1 Artigas      0.245 0.755 NA
2 Canelones    0.419 0.581 NA
3 Colonia      0.316 0.684 NA
4 Florida      0.593 0.407 NA
5 Montevideo   0.452 0.548 NA
6 Paysandu     0.407 0.593 NA
7 Rivera       0.594 0.395  0.0111
8 Soriano      0.375 0.625 NA
```

No es tidy data dado que no sigue los principios de la misma: una fila por oberservación, una columna por variable, una celda por valor. En este caso, la variable `sexo` está distribuida en tres columnas.

8. Seleccioná las variables `depto`, `sexo` `nivel.educ` y `sit.conyugal`. Usando `mutate_if` para transformar las variables de tipo `integer` a tipo `factor`

```
select(personas_reco, depto, sexo, nivel.educ, sit.conyugal) %>%
   mutate_if(.predicate = is.double, .funs = as.factor)
```

```
# A tibble: 10,447 x 4
   depto      sexo  nivel.educ sit.conyugal
   <chr>      <chr> <fct>      <fct>
 1 Montevideo F     4          1
 2 Montevideo M     2          1
 3 Montevideo F     2          3
 4 Montevideo F     4          3
 5 Montevideo F     2          1
 6 Montevideo M     <NA>       1
 7 Montevideo F     5          1
 8 Montevideo F     3          2
 9 Montevideo M     3          1
10 Montevideo M     3          2
# ... with 10,437 more rows
```

9. Replique el siguiente gráfico realizado usando solo información de jefe/a de hogar (valor 1 de `parent.jefe`) para la situación conyugal (`sit.conyugal`). Agregue un subtitulo adecuado al gráfico y comente algo interesante del mismo.

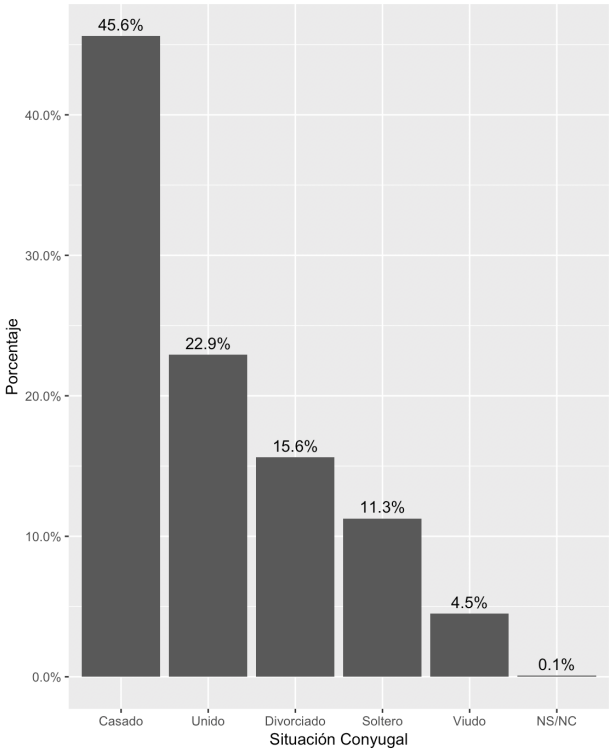


Figure 3: Gráfico a replicar

```r
personas_reco %>%
   filter(parent.jefe == 1) %>%
   mutate(sit.conyugal = as.factor(sit.conyugal),
          sit.conyugal = fct_recode(sit.conyugal,
                                    "Soltero" = "1",
                                    "Unido" = "2",
                                    "Casado" = "3",
                                    "Divorciado" = "4",
                                    "Viudo" = "5",
                                    "NS/NC" = "9")) %>%
   ggplot(aes(x = fct_infreq(sit.conyugal))) +
   geom_bar(aes(y = ..count../sum(..count..))) +
   geom_text(aes(label = scales::percent(..prop..), y = ..prop.., group = 1),
             stat= "count", vjust = -.5) +
   labs(x = "Situación Conyugal", y = "Porcentaje") +
   scale_y_continuous(labels = scales::percent)
```

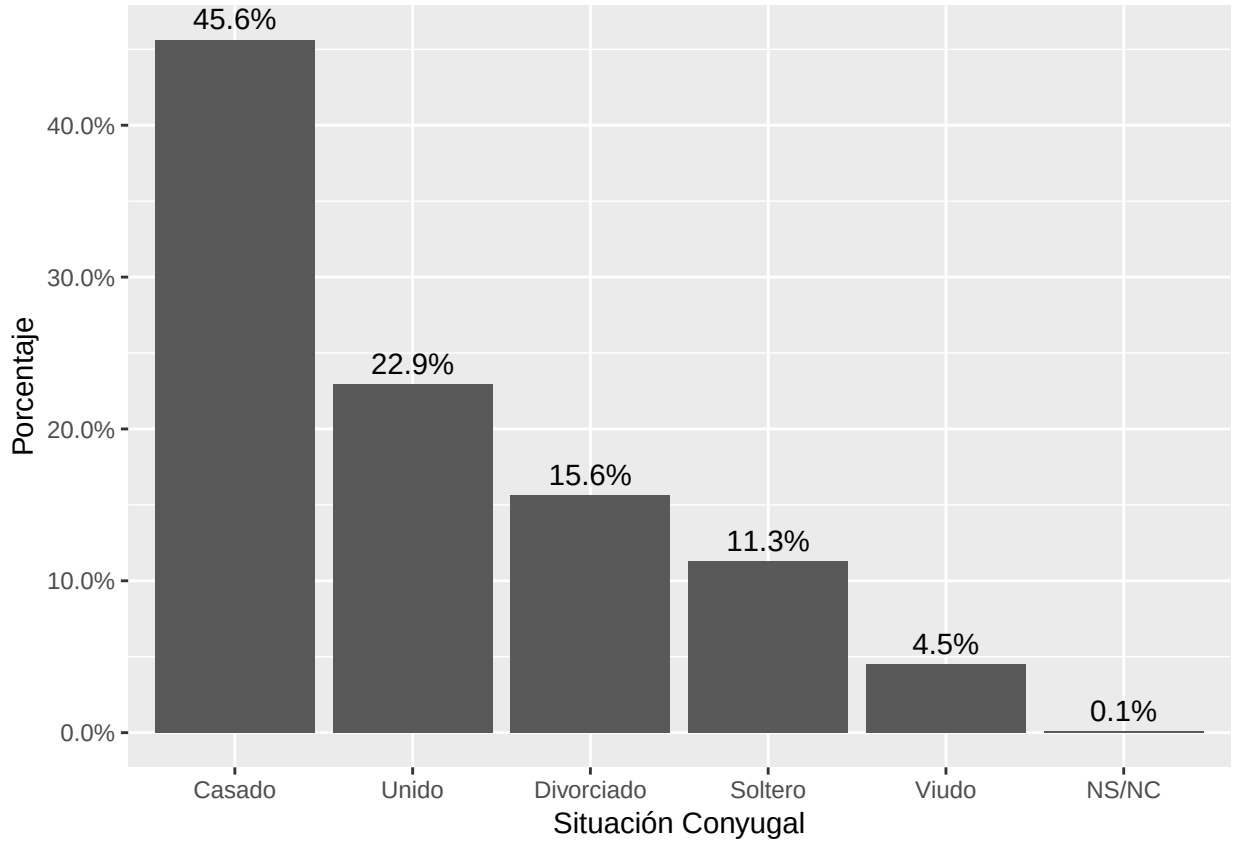


Figure 4: Gráfico de barras de la proporción de personas según situación conyugal para los jefes de hogares. Se observar que la mayoría de ellos/as viven en hogares constituidos, ya sean casados/as o unidos/as.

10. Replique el siguiente gráfico realizado usando solo información de jefe/a de hogar (valor 1 de parent.jefe) para la situación conyugal (`sit.conyugal`) y sexo. Agregue un subtitulo adecuado al gráfico y comente algo interesante del mismo.

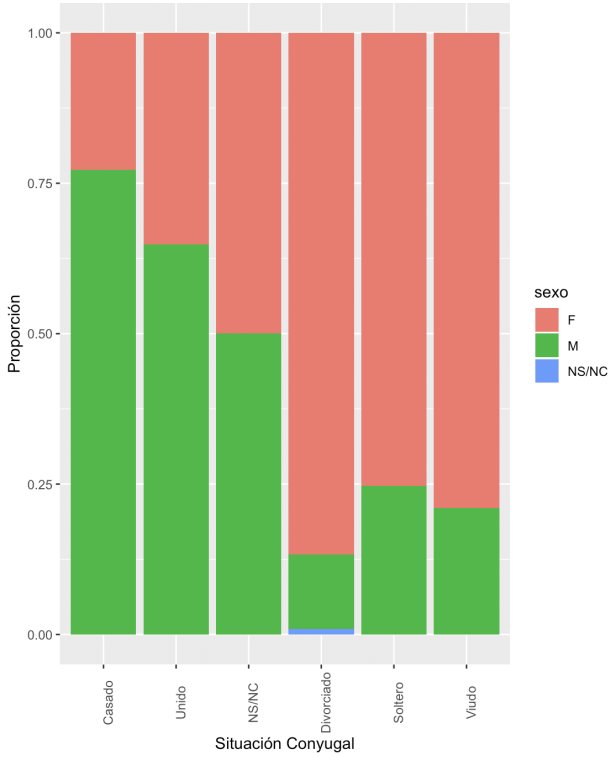


Figure 5: Gráfico a replicar

```
personas_reco %>%
   filter(parent.jefe == 1) %>%
   mutate(sit.conyugal = as.factor(sit.conyugal),
          sit.conyugal = fct_recode(sit.conyugal,
                                    "Soltero" = "1",
                                    "Unido" = "2",
                                    "Casado" = "3",
                                    "Divorciado" = "4",
                                    "Viudo" = "5",
                                    "NS/NC" = "9"),
          sit.conyugal = fct_relevel(sit.conyugal, "Casado", "Unido", "NS/NC", "Divorciado", "Soltero",
   ggplot() +
   geom_bar(aes(sit.conyugal, y = ..count.. / sum(..count..), fill = sexo), position = "fill") +
   labs(x = "Situación Conyugal", y = "Proporción") +
   theme(axis.text.x = element_text(angle = 90))
```

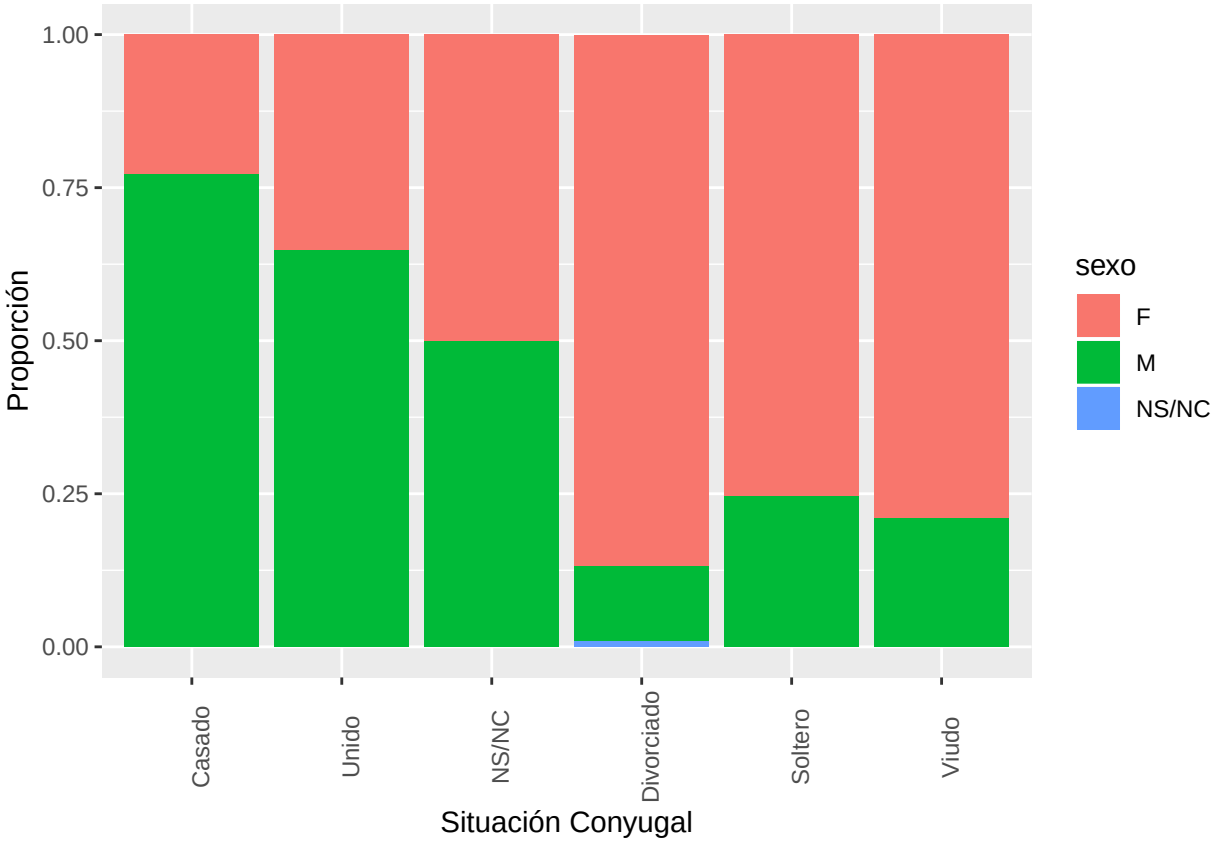


Figure 6: Gráfico de barras apiladas de la proporción de jefes de hogar según su situación conyugal. En los hogares constituidos, casados o unidos, el jefe de hogar suele ser varón, mientras que la situación inversa se observa en los hogares monoparentales.