

Material Didáctico Sistematizado.

# Introducción al estudio de la estructura social a partir de R.

José Javier Rodríguez de la Fuente.

Cita:

José Javier Rodríguez de la Fuente (2023). *Introducción al estudio de la estructura social a partir de R*. Material Didáctico Sistematizado.

Dirección estable: <https://www.aacademica.org/joserodriguez/112>

ARK: <https://n2t.net/ark:/13683/pq7B/Bdk>



Esta obra está bajo una licencia de Creative Commons.  
Para ver una copia de esta licencia, visite  
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

*Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.*

# INTRODUCCIÓN AL ESTUDIO DE LA ESTRUCTURA SOCIAL A PARTIR DE R

III Jornadas de Sociología - Universidad Nacional de Mar del  
Plata

José Rodríguez de la Fuente

Instituto de Investigaciones Gino Germani (UBA-CONICET)

3/15/23



# PROGRAMA DEL TALLER

1. ¿Qué es la estructura social?
2. ¿Por qué usar **R**?
3. Introducción (breve) a **R**
4. Trabajando con encuestas de hogares: EPH
5. Construyendo variables: las clases sociales
6. Explorando resultados. Tablas y gráficos
7. Dudas y consultas

# BIBLIOGRAFÍA RECOMENDADA

- Sacco, Nicolás; José Rodríguez de la Fuente y Sofía Jaime (2022). Libro de Cocina para el Análisis de las Clases Sociales en Argentina.
- Tiscornia, P. y Weksler, G. Introducción a R para Ciencias Sociales. Aplicación práctica en la EPH.

# 1. ¿QUÉ ES LA ESTRUCTURA SOCIAL?

# ¿QUÉ ES LA ESTRUCTURAL SOCIAL? (1)

- **Carácter temporal:** relaciones estables de la sociedad, regularidades de sus elementos constituyentes.
- **Carácter plural:** designa los rasgos de grupos y sociedades, en otras palabras, que excede a los aspectos individuales.
- Puede abordarse desde diversas dimensiones: económica, social, política, etc.

# ¿QUÉ ES LA ESTRUCTURAL SOCIAL? (2)

- La **estadística** es una disciplina central para el estudio de la estructura social
  - Da cuenta de la variabilidad de la vida social humana
  - Permite dar cuenta de la regularidades poblacionales (junto con la teoría)
- Importancia en cómo los datos son *recogidos y analizados*



# ¿ESTRUCTURA SOCIAL O ESTRUCTURA DE CLASES?

Al hablar de clases hablamos de la estructura socio-económica (aunque a veces se usen cómo sinónimos)

- Existen distintos enfoques:
  - (neo)marxista
  - (neo)weberiano
  - Funcionalista
  - Bourdiano
  - Estadístico / pragmático

Todos deben poder ser operacionalizados



## 2. ¿POR QUÉ USAR R?

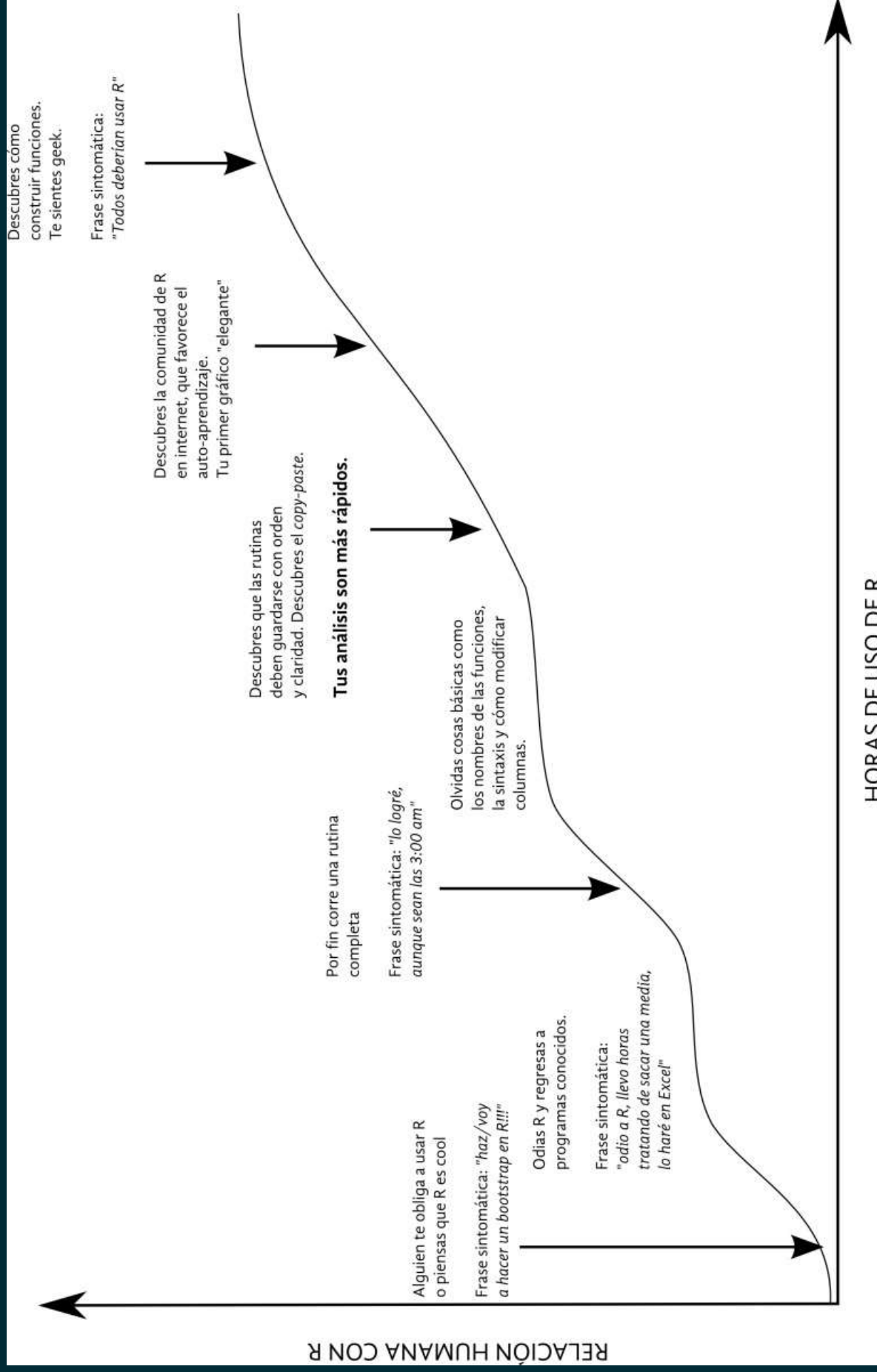
# VENTAJAS

- Código abierto: amplia cantidad de paquetes
- Es gratuito
- Gran comunidad de usuarios (foros)
- Uso extendido en la ciencia (sociales) y ámbito profesional
- Trabaja con la memoria virtual
- Capacidad de armar tableros, aplicaciones, documentos, etc.

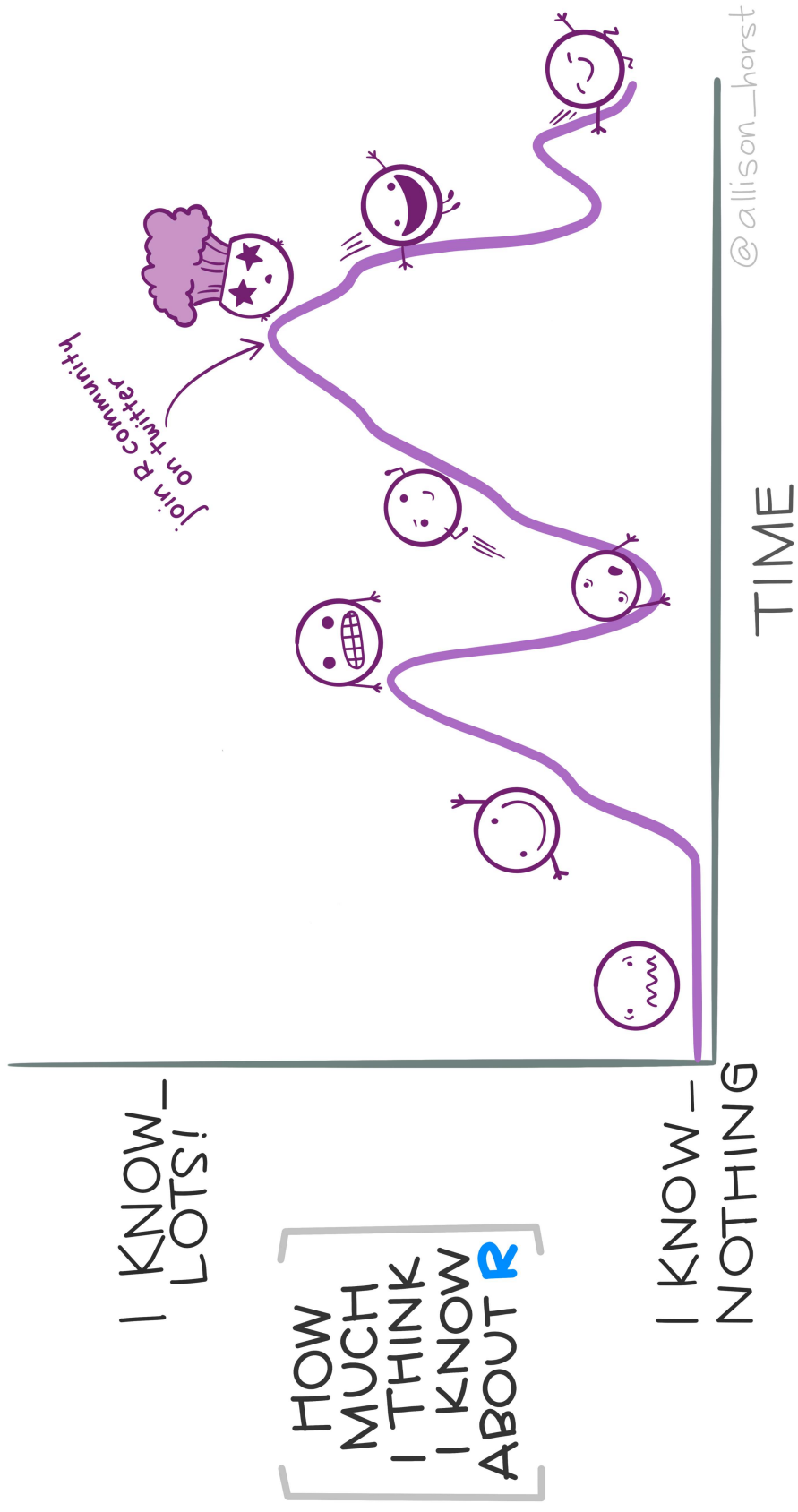
# DESVENTAJAS (1)

- Las órdenes deben darse por *sintaxis (no hay mucho point and click)*
- Hacer cosas *sencillas y rápidas* puede llevar más tiempo que en otros programas
- La curva de aprendizaje es lenta

# DESVENTAJAS (2)



# DESVENTAJAS (3)



# COMPARACIONES CON OTROS PROGRAMAS

Software	Interfaz	Curva de aprendizaje	Manipulación de data	Análisis estadístico	Gráficos	Especialidades
<i>SPSS</i>	Menus & Sintaxis	Gradual	Moderada	Alcance moderado Baja versatilidad	Buenos	Tablas personalizadas, ANOVA & Análisis multivariado
<i>Stata</i>	Menus & Sintaxis	Moderado	Fuerte	Amplio alcance Media versatilidad	Buenos	Análisis de panel, Análisis de encuesta & Imputación múltiple
<i>R</i>	Sintaxis	Empinada	Muy fuerte	Muy amplio alcance Alta versatilidad	Excelentes	Paquetes de gráficos, Web Scraping, Machine Learning & Predictive Modeling

Fuente: <https://sites.google.com/a/nyu.edu/statistical-software-guide/summary>

# EJEMPLOS DE USOS DE R EN EL ÁMBITO PÚBLICO Y ACADEMICO

- Tableros y reportes del SINTA
- Tablero interactivo de indicadores sobre VIH, ITS, hepatitis virales y tuberculosis
- Sacco, Nicolás; José Rodríguez de la Fuente y Sofía Jaime (2022). Libro de Cocina para el Análisis de las Clases Sociales en Argentina.



# 3. INTRODUCCIÓN (BREVE) A R

# DESCARGA E INSTALACIÓN

1. Debemos instalar **R** desde **CRAN** (the comprehensive R archive network)
2. Debemos instalar **RStudio**
  - Es un ambiente de desarrollo integrado (IDE) para programar en **R**
3. Recomendación de actualizar ambos programas periódicamente

# PARTES DE RSTUDIO

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for downloading and installing R. The code includes comments in Spanish and instructions on how to install R from the CRAN network or directly from the RStudio website. The code is as follows:

```
151  
152 ## Descarga e instalación  
153  
154 ::: {.incremental}  
155  
156 1. Debemos instalar R desde [**CRAN**] (the comprehensive R archive  
network)](https://cran.r-project.org/)  
157  
158 2. Debemos instalar [RStudio](http://www.rstudio.com/download)  
159 + Es un ambiente de desarrollo integrado (IDE) para programar en R  
160  
161 3. Recomendación de actualizar ambos al último ambiente  
162  
163 :::  
164  
165 ## Partes de RStudio  
166 # Partes de RStudio
```
- Console:** Shows the output of the R code, including the R version (R 4.2.1) and the RStudio logo. It also displays the RStudio welcome message and instructions on how to cite R or R packages in publications.

```
R 4.2.1 ~~/Repositorios_git/curso_estructura_social_R/  
RStudio  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.  
  
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.  
  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.  
>
```
- Environment:** Shows the current environment is empty.
- Files:** Lists files in the current directory, including .gitignore, .Rhistory, curso\_estructura\_social\_R.Rproj, custom.scss, images, presentacion\_files, presentacion.html, and presentacion.qmd.

## Ambiente de trabajo

## Scripto sintaxis

## Archivos del proyecto



# ¿CÓMO INSTALAR Y HACER FUNCIONAR LOS PAQUETES?

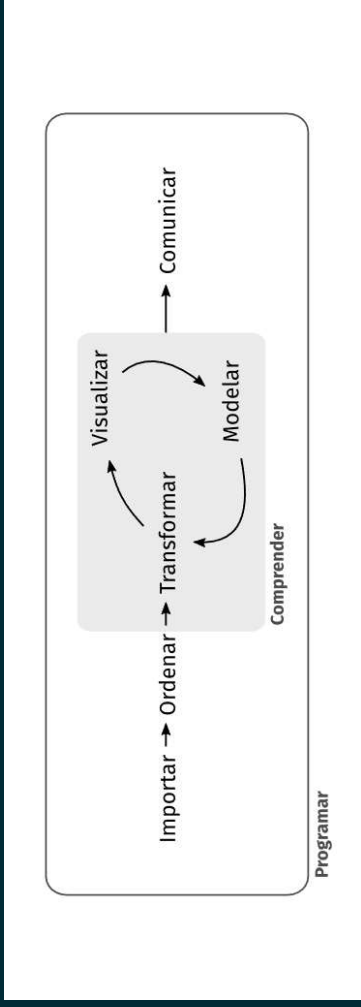
- Los paquetes se instalan via descarga de internet
- La operación se hace directamente desde la consola o script de **RStudio**

```
1 install.packages ("tidyverse")
```

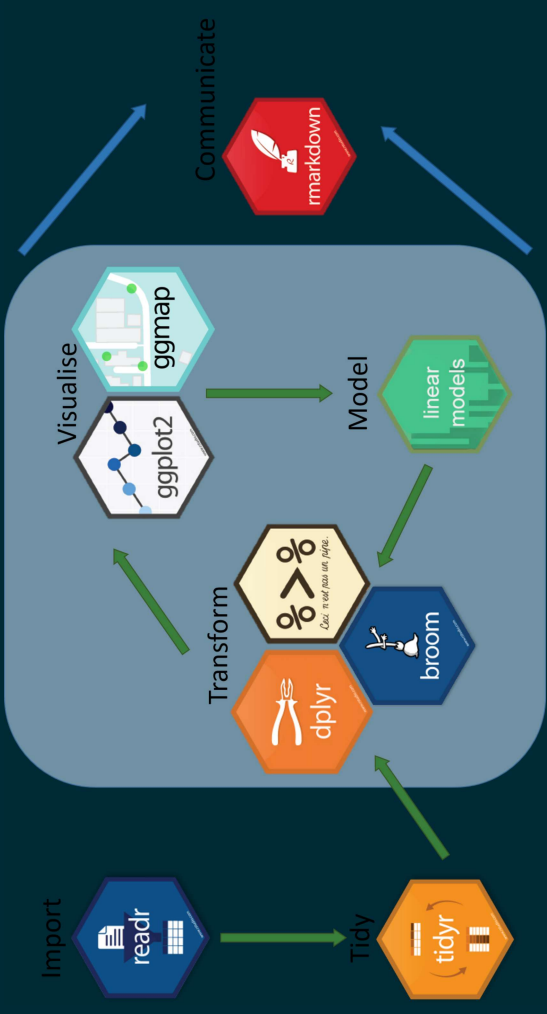
- Siempre que necesitemos usar un paquete debemos *activarlo*

```
1 library (tidyverse)
```

# FLUJO DE TRABAJO: DINÁMICA



Fuente: Wickham and Grolemund, 2016

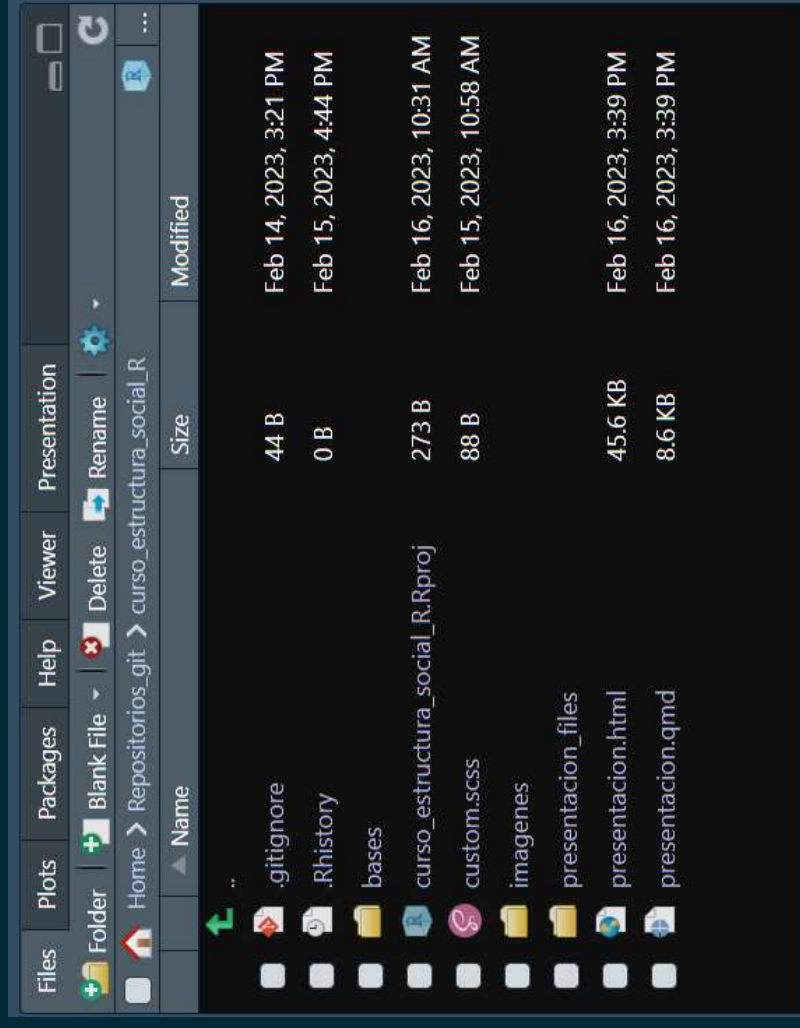


Fuente:

<https://teachdatascience.com/tidyverse/>

# FLUJO DE TRABAJO: PROYECTOS

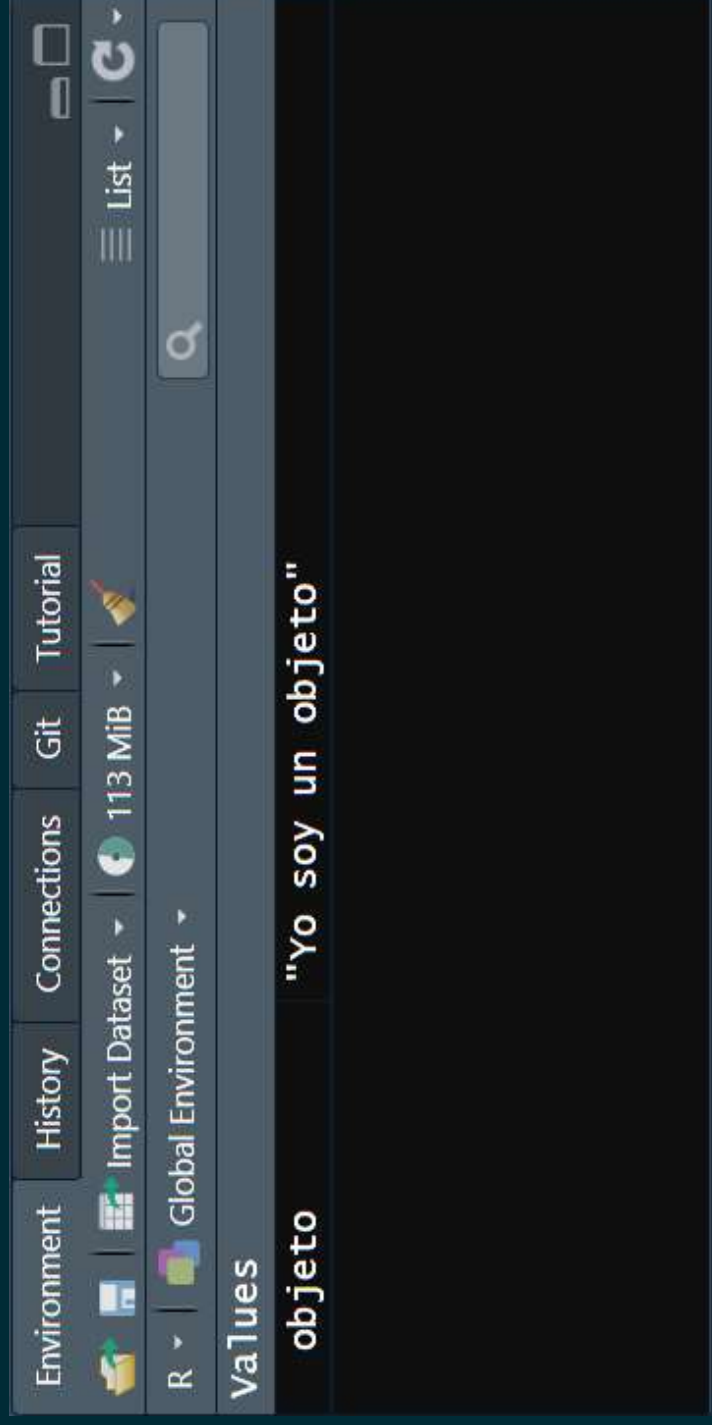
- Siempre es recomendable trabajar dentro de un proyecto
- Nos aseguraremos que todos los archivos y carpetas necesarios estén siempre en un lugar único
- Trabajaremos con rutas relativas.



# FLUJO DE TRABAJO: OBJETOS (1)

Todo lo que creamos en R es un objeto. Y en esto R es totalmente distinto a lo que conocíamos.

```
1 objeto <- "Yo soy un objeto"
```





# FLUJO DE TRABAJO: OBJETOS (2)

Podemos crear objetos a partir de:

## Números

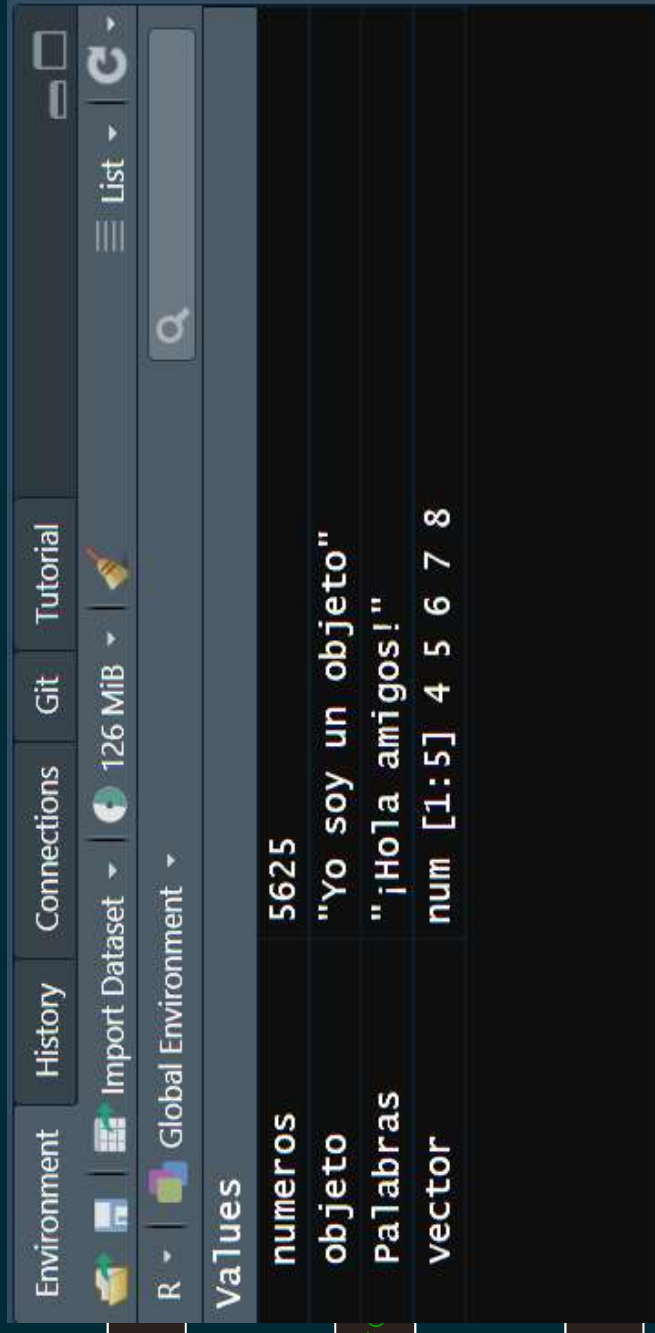
```
1 numeros <- 5625
```

## Palabras

```
1 Palabras <- "¡Hola amigos!"
```

## Vectores

```
1 vector <- c(4, 5, 6, 7, 8)
```



Listas, *data frames*, gráficos, funciones, mapas y muchas cosas más.

# FLUJO DE TRABAJO: ELEMENTOS DEL SCRIPT (CARACTERES ESPECIALES)

- De asignación (`< - o =`): le asignan un valor a un objeto.
- De anotación (`#`): permite escribir anotación en el script que no son leídas como código
- Pipa (`%>%`): permite concatenar una secuencia de acciones sobre un objeto
  - Pertenece al paquete `magrittr` que está en `tidyverse`

# FLUJO DE TRABAJO: ELEMENTOS DEL SCRIPT (OPERADORES)

## Operadores aritméticos

- + Suma
- - Resta
- \* Multiplicación
- / División

## Operadores relacionales

- > Mayor
- >= Mayor o igual
- < Menor
- <= Menor o igual
- == Igual
- != Distinto

## Operadores lógicos

- & y
- | ó

# FLUJO DE TRABAJO: OBJETOS (3)

Existe una *clase* especial de vector denominado **factor**

- Suelen ser utilizados como variables categóricas (nominales u ordinales)
- Tienen un conjunto fijo y conocido de valores que puede asumir
- Se suelen construir a partir de vectores de cadena

```
1 clases_sociales <- c("alta", "media", "baja", "baja", "media", "media", "ba
2 clases_factor <- factor(clases_sociales, levels = c("baja", "media", "alta"
3
4 table(clases_factor)
```

```
clases_factor
baja media alta
3      3      1
```

# PRINCIPALES FUNCIONES DE `dp1yr`










- **Filter**: filtrado de filas o casos
- **Select**: selección de columnas o variables
- **Arrange**: ordena los casos de una variable(s)
- **Count**: cuenta casos
- **Mutate**: crea nuevas columnas o variables
- **Group\_by / summarize**: agregan casos por variable

# **4. TRABAJANDO CON ENCUESTAS DE HOGARES: EPH.**

# LA ENCUESTA PERMANENTE DE HOGARES (I)

- Programa nacional de producción sistemática y permanente de indicadores sociales que lleva a cabo el INDEC.
- Comenzó a aplicarse desde 1973, bajo una modalidad (puntual) que implicaba un doble relevamiento anual (mayo y octubre).
- Actualmente cubre a 31 aglomerados del país y se realiza bajo una modalidad continua.

# LA ENCUESTA PERMANENTE DE HOGARES (II)

- Las principales dimensiones relevadas son:
  - situación laboral 
  - características demográficas básicas   
  - migración 
  - características habitacionales 
  - educación 
  - ingresos  
- Presenta un diseño de tipo panel (2-2-2).


























# ¿POR QUÉ USAR LA EPH PARA ESTUDIAR LA ESTRUCTURA DE CLASES?

- Medición precisa de **variables laborales**
  - Ocupación, rama de actividad, sector de actividad, categoría ocupacional, tamaño del establecimiento, etc.
- Relevamiento **continuo y generalizado**
  - Información desde 1974
  - Incorporación de aglomerados con el paso del tiempo
- Relevamiento de **otras dimensiones** relevantes

# LAS BASES DE MICRODATOS

Dependiendo del período que necesitemos analizar, encontraremos que las bases a descargar se encuentran en distintos formatos de archivo.

- 1974-2002        [foreign](#)
- 2003-2015         [haven](#)
- 2016-2022         [R base](#)

Vamos a probar descargar e importar la base más reciente

# IMPORTANDO LOS MICRODATOS (I)

- Descargar desde el sitio web de INDEC la última base disponible.
- Guardar el archivo en la carpeta **bases** del proyecto.
- Importar la base con la función **read.csv2**

```
1 eph_03_22 <- read.csv2("bases/usu_individual_T322.txt", header = T)
```

¿Cuántos registros y variables tenemos?

```
1 nrow(eph_03_22)
```

```
[1] 49232
```

```
1 ncol(eph_03_22)
```

```
[1] 177
```

# IMPORTANDO LOS MICRODATOS (II)

CODUSU <chr>	ANO4 <int>	TRIMESTRE <int>
TQRMNORVQHJNLTCDEGGFB00788741	2022	3
TQRMNOQSWHKNKRCDEGJBF00789123	2022	3
TQRMNOQSWHKNKRCDEGJBF00789123	2022	3
TQRMNOTSQHLLMLCDEIAD00703213	2022	3
TQRMNOTSQHLLMLCDEIAD00703213	2022	3
TQRMNOSWVHKMKOCDEHPJB00790947	2022	3
TQRMNOSWVHKMKOCDEHPJB00790947	2022	3
TQRMNOSWVHKMKOCDEHPJB00790947	2022	3
TQRMNOSWVHKMKOCDEHPJB00790947	2022	3
TQRMNOSWVHKMKOCDEHPJB00790947	2022	3

1-10 of 10,000 rows | 1-3 of 177 columns

Previous 1 2 3 4 5 6 1000ext

# EL PAQUETE **eph**(I)

- Creado por usuaries de la EPH:  
<https://holatam.github.io/eph/>
- Permite:
  - descargar fácilmente las bases de datos
  - etiquetar variables y categorías
  - organizar pools de datos en panel
  - estimar tasas de pobreza

# EL PAQUETE **eph**: FUNCIONES BÁSICAS

Vamos a activar el paquete **eph** y a descargar la base de individuos del 4to trimestre de 2019 utilizando la función **get\_microdata**.

```
1 library(eph)
2 eph_ind_419 <- get_microdata(year = 2019, trimester = 4, type = "individual")
```

Podemos etiquetar las variables y categorías automáticamente con la función **organize\_labels**.

```
1 eph_ind_419 <- organize_labels(eph_ind_419, type = 'individual')
```

Observamos las etiquetas de la variable **sexo**

```
1 attributes(eph_ind_419$CH04)

$label
[1] "Sexo"

$class
[1] "labelled" "integer"

$labels
Varon Mujer
 1      2
```

# PRIMEROS PASOS CON LA EPH: UNIÓN DE BASES (I)

- Es frecuente que en el estudio de la estructura social debamos combinar el uso de información **individual** y de **hogares**
- Para ello debemos unir las bases
- Debemos identificar a las variables **CODUSU** y **NRO\_HOGAR** para poder realizar la unión entre ambas bases
- Siempre es recomendable unir únicamente las variables que necesitamos

Pero primero... debemos descargar la base de hogares del 4to trimestre de 2019:

```
1 eph_hog_419 <- get_microdata(year = 2019, trimester = 4, type = "hogar",  
2 vars = c("CODUSU", "NRO_HOGAR", "II7", "III", "IX_TOT"))
```

# PRIMEROS PASOS CON LA EPH: UNIÓN DE BASES (II)

Ahora si estamos en condiciones de unir ambas bases

Para ello vamos a utilizar la función `left_join` del paquete `dp1yr` que se encuentra en el `tidyverse`

```
1 library(tidyverse)
2
3 eph_419 <- eph_ind_419 %>%
4 left_join(eph_hog_419, b
```

CODUSU <fct>	ANO4 <int>	TRIMESTRE <int>
TQRMNOSTVHKOKMCDEFOCD00641714	2019	4
TQRMNOSTVHKOKMCDEFOCD00641714	2019	4
TQRMNOSTVHKOKMCDEFOCD00641714	2019	4
TQRMNORQPHMNLPCDEFOCD00641575	2019	4
TQRMNORQPHMNLPCDEFOCD00641575	2019	4
TQRMNOQQVHJLLNCDEFOCD00629598	2019	4
TQRMNOQQVHJLLNCDEFOCD00629598	2019	4
TQRMNOQPPHKOLOCDEFOCD00618307	2019	4
TQRMNOQPPHKOLOCDEFOCD00618307	2019	4
TQRMNOQPPHKOLOCDEFOCD00618307	2019	4

1-10 of 10,000 rows | 1-3 of 180 colu... [Previous](#) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [1000Next](#)



# PRIMEROS PASOS CON LA EPH: FILTRADO

- Casi siempre es necesario seleccionar una porción de nuestra muestra
- Se aplican filtros por región, período, sexo, edad, situación laboral
- Para ello utilizaremos la función **filter** del **tidyverse**

Seleccionaremos el aglomerado de Mar del Plata - Batán (34) y observamos la cantidad de casos con **nrow**

```
1 eph_419_mdp <- eph_419 %>%  
2   filter(AGLOMERADO == 34)  
3  
4 nrow(eph_419_mdp)
```

[1] 1098

# PONDERACIÓN DE LA MUESTRA

- Debido a que estamos trabajando con una muestra probabilística y representativa de la población es necesario **ponderar** los datos
- Esto permite
  - Recalibrar la muestra salvando errores de sobre o subrepresentación
  - Expandir el número efectivo de casos al número total poblacional
- Esto debe hacerse cada vez que realizamos un procesamiento (no es como SPSS) y dependerá de que operación debamos realizar (suma, promedio, tabulado, regresión, etc.)
- Existencia de distintos ponderadores: PONDERA, PONDIO, PONDII, PONDIII

Probemos calcular el total poblacional de Mar del Plata - Batán

```
1 eph_419_mdp >%  
2 tally (PONDERA)
```

```
# A tibble: 1 × 1
```

```
  n  
1 641039
```

# **5. CONSTRUYENDO VARIABLES: LAS CLASES SOCIALES**

# LAS CARACTERÍSTICAS SOCIO-LABORALES

- Los esquemas de clases sociales son *operacionalizaciones* de las teorías de clases.
- Los elementos centrales para su construcción son las *variables*

Principales variables socio-ocupacionales de la EPH

Variable	Descripción	Etiqueta	Categorías
Condición de actividad	Población activa o inactiva	ESTADO	Ocupado / Desocupado / Inactivo / Menor de 10 años
Categoría ocupacional	Estatus jerárquico en el empleo	CAT_OCUP	Patrón / Cuenta propia / Empleado / Trabajador familiar
Ocupación	Código de ocupación	PP04D_COD (ocupación actual) / PP11D_COD (ocupación anterior)	Ver Clasificador Nacional de Ocupaciones (CNO)
Rama de actividad	Sector industrial de la ocupación	PP04B_COD (ocupación actual) / PP11B_COD (ocupación anterior)	Ver Clasificador de Actividades Económicas para Encuestas Socio-demográficas del Mercosur (CAES - MERCOSUR)
Tamaño del establecimiento	Tamaño del establecimiento del trabajo	PP04C / PP04c_99	Numérica

Note:

Fuente: elaboración propia

# CONDICIÓN DE ACTIVIDAD

Permite distinguir entre personas económicamente activas (PEA) y no económicamente activas (PNEA)

```
1 table (ep_h_419$ESTADO)
```

```
0      1      2      3      4
42 24441 2031 23530 8475
```

# OCUPACIÓN (1)

- INDEC utiliza el Clasificador Nacional de Ocupaciones
- Este se compone de 5 dígitos

CARÁCTER OCUPACIONAL 10 categorías que actúan como los grandes grupos ocupacionales	JERARQUÍA OCUPACIONAL cuatro categorías	TECNOLOGÍA OCUPACIONAL tres categorías	CALIFICACIÓN OCUPACIONAL cuatro categorías
0 Ocupaciones de dirección	0 Ocupaciones de dirección	1 Ocupaciones sin operación de maquinaria, equipos y sistemas	1 Ocupaciones de calificación profesional
1 Ocupaciones de gestión administrativa, de planificación, control de gestión y jurídico-legal	1 Ocupaciones de cuenta propia	2 Ocupaciones con operación de maquinaria, equipos o sistemas electromecánicos	2 Ocupaciones de calificación técnica
2 Ocupaciones de gestión presupuestaria, contable y financiera	2 Ocupaciones de jefatura	3 Ocupaciones con operación de equipos o sistemas informatizados	3 Ocupaciones de calificación operativa
3 Ocupaciones de comercialización, transporte, almacenaje y telecomunicaciones	3 Ocupaciones de ejecución directa		4 Ocupaciones no calificadas
4 Ocupaciones de servicios sociales básicos			
5 Ocupaciones de servicios varios			
6 Ocupaciones agropecuarias, forestales, de la pesca y de la caza			
7 Ocupaciones de la producción extractiva, energética, de construcción e infraestructura			
8 Ocupaciones de la producción artesanal, industrial y de reparación de bienes de consumo			
9 Ocupaciones auxiliares de la producción de bienes y de la prestación de servicios			

# OCUPACIÓN (II)

1 **table** (eph\_419\$PP04D\_COD)

1	1001	10111	10112	10113	10123	10131	10132	10133	10201	10202	10203	10311
6	8	1	24	4	1	8	19	4	18	72	16	3
10312	10313	10314	10321	10322	10323	10331	10332	10333	10339	11131	11201	11202
80	108	43	1	8	8	41	205	1680	6	99	2	7
11311	11312	11322	11331	11332	11333	11339	2001	20111	20112	20113	20123	20131
2	2	1	116	8	14	1	7	2	4	7	2	59
20132	20133	20201	20202	20203	20311	20312	20313	20323	20331	20332	20333	20339
8	1	8	14	1	8	1	109	1	138	164	483	1
3001	30111	30112	30113	30119	30122	30123	30131	30132	30133	30139	30201	30202
119	2	20	1028	5	4	16	6	54	208	3	1	13
30203	30311	30312	30313	30314	30319	30322	30323	30331	30332	30333	30339	31113
40	3	20	239	680	14	3	14	13	106	493	10	40
31122	31123	31131	31133	31203	31312	31313	31314	31319	31322	31323	31332	31333
1	5	1	13	1	3	34	21	5	1	18	1	21

# CATEGORÍA OCUPACIONAL

Permite discriminar a la población en función de su posicionamiento en las relaciones de producción y a partir del origen de sus ingresos

```
1 table (eph_419$CAT_OCUP)
```

```
0      1      2      3      4      9
32503 1012  5714 19127  162    1
```



# DE LAS CLASES EN EL PAPEL A LAS CLASES EN LA COMPUTADORA

- La *operacionalización* implica el pasaje de conceptos teórico-abstractos a indicadores empíricos.
- Estas clases-construtos no son clases equiparables a las que podemos percibir y comprender como existentes en la realidad social, aunque presenten probabilidades de serlo.
- El proceso de *operacionalización* implica una abstracción y un recorte realizado sobre la realidad.
- Confusión de la noción de **clase** con otros conceptos: estratos, grupos, sectores, fracciones.
- Hay varias propuestas teóricas de la estratificación que fueron operacionalizadas: EGP, Wright, SIOPS, SEI, ISEI, Germani, **Torrado**, Portes, etc.

# EL ESQUEMA DE CLASES DE TORRADO

Clase social	Estratos sociales	Estratos sociales
Clase alta	-	Directores de empresa
Clase media	Autónomos	Profesionales en función específica
		Propietarios de pequeñas empresas
	Asalariados	Pequeños productores autónomos
		Profesionales en función específica
Clase obrera	Autónomos	Cuadros técnicos y asimilados
		Empleados administrativos y vendedores
	Asalariados	Trabajadores especializados autónomos
		Obreros calificados
	Marginales	Obreros no calificados
		Peones autónomos
		Empleados domésticos

Fuente: elaboración propia en base a Torrado (1998)

## Sistema clasificatorio de la clase social

# PASO 1: PUNTO DE PARTIDA

## 2. Categoría de ocupación

Grupo de ocupación	Empleadores		Asalariados			Cuenta propia y familiar sin remuneración
	Sector privado	Sector público	Sector privado	Sector público	Servicio doméstico	
	Más de 5 ocupados	Hasta 5 ocupados	Más de 5 ocupados	Hasta 5 ocupados		
<b>1. Ocupación</b> 1. Empresarios, directores de empresas y funcionarios públicos superiores	1.1	1.2	1.1*			5.2
2. Propietarios de establecimientos	3	4.1.1	4.2			
3. Profesionales en función específica	2.1.1	2.1.2	2.3.1	2.3.2	2.4	2.2
4. Técnicos, docentes y supervisores			4.1.1	4.1.2	4.2	5.2
5. Empleados y vendedores	3	5.1	6.1.1	6.1.2	6.2	7
6. Trabajadores especializados			8.1.1	8.1.2	8.2	
7. Trabajadores no especializados	10	10	9.1.1	9.1.2	9.2	10
8. Empleados domésticos	11	11	11	11	11	11
9. Sin especificar	12	12	12	12	12	12

\* Modificación respecto a la propuesta de Torrado. Fuente: elaboración propia en base a Torrado (1998)

# Clasificador de la Condición Socio-ocupacional (CSO)

# PASO 2: OCUPACIÓN (I)

- Debemos preparar las variables para poder trabajarlas en la operacionalización.
- En primer lugar empezaremos por agregar ceros a los registros que tienen menos de 5 dígitos, utilizando la función `str_pad` del **tidyverse**.
- Luego vamos a separar los 5 dígitos del **CNO**. Para ello utilizaremos la función `str_sub` del **tidyverse**.
- Luego convertiremos esas variables en formato numérico.

```
1 eph_419$cno <- str_pad(eph_419$PP04D_COD, 5, pad = "0")
2
3 eph_419$cno12 <- str_sub(eph_419$cno, 1, 2) #señala que parta de la posición
4 eph_419$cno3 <- str_sub(eph_419$cno, 3, 3) #señala que parta de la posición
5 eph_419$cno4 <- str_sub(eph_419$cno, 4, 4)
6 eph_419$cno5 <- str_sub(eph_419$cno, 5, 5)
7
8 eph_419$cno12 <- as.numeric(eph_419$cno12)
9 eph_419$cno3 <- as.numeric(eph_419$cno3)
10 eph_419$cno4 <- as.numeric(eph_419$cno4)
11 eph_419$cno5 <- as.numeric(eph_419$cno5)
```

## PASO 2: OCUPACIÓN (II)

- Crearemos una variable que permita agrupar a las ocupaciones del mismo modo que lo hizo Torrado.
- Para esto generaremos la variable **grupo de ocupación (GO)**, utilizando los dos primeros dígitos del CNO.
- A través de la función `case_when` del **tidyverse** podremos asignarle un valor a la variable en la medida que se cumplan determinadas condiciones

```
1 eph_419 <- eph_419 %>%
2   mutate(go = case_when(((cno12 >= 0 & cno12 <=4) | (cno12 == 7) ~ 1,
3
4     (cno12 >= 5 & cno12 <=6) ~ 2,
5
6     (cno5 == 1) & ((cno12 >= 10 & cno12 <= 20) |
7     (cno12 == 32) | (cno12 >= 34 & cno12 <= 40) |
8     (cno12 >= 42 & cno12 <= 47) | (cno12 >= 49 &
```

# PASO 3: SECTOR DE ACTIVIDAD Y TAMAÑO DEL ESTABLECIMIENTO

- Vamos a recodificar dichas variables utilizando la función `recode` del paquete `car`.
- El sector de actividad (`pp04a`) quedará recodificado en 1) sector *privado* y 2) sector *público*.
- El tamaño del establecimiento (`pp04c` y `pp04c99`) quedará recodificado en 1) 5 o menos ocupados y 2) Más de 5 ocupados.

```
1 eph_419 <- eph_419 %>%
2   mutate(sector_act = car::recode(PP04A, "1=2; 2:3=1"))
3
4 eph_419 <- eph_419 %>%
5   mutate(tamaño = case_when(
6     (PP04C > 0 & PP04C <= 5) | (PP04C == 99 & PP04C
7     (PP04C > 5 & PP04C < 99) | (PP04C == 99 & PP04C
8     PP04C == 0 | PP04C99 == 0 ~ NA_real_
9     ))
```

# PASO 4: CREACIÓN DEL CSO DESAGREGADO (I)

Con las variables necesarias ya creadas y/o recodificadas, podemos empezar a *llenar los casilleros* del cuadro anteriormente presentado.

Asignación de códigos al CSO desagregado

N° sintaxis	N° real	CSO desagregado	CSO agregado
1	1.1	DIREC (ER - SPR. TE>5)	1
2	1.2	DIREC (AS - SPR. TE>5)	
3	2.1.1	PROF (ER-SPR. >5)	
4	2.1.2	PROF (ER-SPR. <=5)	
5	2.2	PROF (CP)	2
6	2.3.1	PROF (AS-SPR. >5)	
7	2.3.2	PROF (AS-SPR. <=5)	
8	2.4	PROF (AS-SPU)	
9	3	PPE	3
10	4.1.1	TECN (AS-SPR. TE>5)	
11	4.1.2	TECN (AS-SPR. TE<=5)	4
12	4.2	TECN (AS-SPU)	
13	5.1	PPA (ER - SPR. TE <=5)	

# PASO 4: CREACIÓN DEL CSO DESAGREGADO (II)

Crearemos una variable que se llame `cso_desag`

```
1 eph_419 <- eph_419 %>%
2   mutate(cso_desag = case_when (go==1 & CAT_OCUP==1 & tamaño==2 ~ 1,
3   go==1 & CAT_OCUP==1 & tamaño==1 ~ 13,
4   go==1 & CAT_OCUP==3 & sector_act==1 & tamaño
5   go==1 & CAT_OCUP==3 & sector_act==1 & is.na(
6   go==1 & CAT_OCUP==3 & sector_act==1 & tamaño
7   go==1 & CAT_OCUP==3 & sector_act==2 ~ 1,
8   go==1 & CAT_OCUP==2 ~ 14,
9   go==1 & CAT_OCUP==4 ~ 14,
10  go==1 & CAT_OCUP==NA ~ 14,
11
12  go==2 & CAT_OCUP==1 & tamaño==2 ~ 9,
13  go==2 & CAT_OCUP==1 & tamaño==1 ~ 13,
14  go==2 & CAT_OCUP==3 & sector_act==1 & tamaño
15  go==2 & CAT_OCUP==3 & sector_act==1 & is.na(
16  go==2 & CAT_OCUP==3 & sector_act==1 & tamaño
17  go==2 & CAT_OCUP==3 & sector_act==2 ~ 12,
18  go==2 & CAT_OCUP==2 ~ 14,
```



# PASO 5: CREACIÓN DEL CSO AGREGADO

Crearemos una variable que se llame `cso_agg`

```
1 eph_419 <- eph_419 %>%  
2   mutate(cso_agg = case_when(  
3     cso_desag <=2 ~ 1,  
4     cso_desag >=3 & cso_desag <=8 ~ 2,  
5     cso_desag ==9 ~ 3,  
6     cso_desag >=10 & cso_desag <=12 ~ 4,  
7     cso_desag >=13 & cso_desag <=14 ~ 5,  
8     cso_desag >=15 & cso_desag <=17 ~ 6,  
9     cso_desag ==18 ~ 7,  
10    cso_desag >=19 & cso_desag <=21 ~ 8,  
11    cso_desag >=22 & cso_desag <=24 ~ 9,  
12    cso_desag ==25 ~ 10,  
13    cso_desag ==26 ~ 11,  
14    cso_desag ==27 | is.na(cso_desag) ~ NA_real_))
```

# PASO 6: EL ESQUEMA FINAL (I)

¡Ya estamos en condiciones de elaborar el sistema de clases que la autora propone!

Con la función `factor` crearemos una variable factor etiquetando el esquema de clases.

```
1 eph_419 <- eph_419 %>%
2   mutate(clase6 = case_when(cso_agg == 1 ~ 1,
3
4     (cso_agg == 2 & CAT_OCU
5     (cso_agg == 2 & CAT_OCU
6     cso_agg == 7 ~ 4,
7     cso_agg == 8 | cso_agg
8     cso_agg == 10 | cso_agg
9     eph_419$clase6_factor <- factor(eph_419$clase6, labels = c("Clase alta",
10    "Clase media - a
11    "Clase media -as
12    "Clase obrera -
13    "Clase obrera -
14    "Clase obrera -
```

# DISTRIBUCIONES DE FRECUENCIAS (dp1yr)

- Con la variable ya construida tenemos dos formas (entre muchas) para realizar una distribución de frecuencia con los datos ponderados.
- En primer lugar mostraremos el ejemplo utilizando funciones del paquete **dp1yr** del **tidyverse**.

```
1 eph_419 %>%
2   filter(!is.na(clase6_factor)) %>% #quitamos los casos perdidos del análisis
3   group_by(clase6_factor) %>% #agrupamos los datos por clase
4   tally(PONDERA) %>% #hacemos un conteo ponderado
5   mutate (porcentaje = round((n/sum(n))*100, digits = 2)) #calculamos el porcentaje
```

```
# A tibble: 6 × 3
  clase6_factor      n porcentaje
<fct>          <int>      <dbl>
1 Clase alta      134342      1.11
2 Clase media - autónoma 1908959     15.8
3 Clase media -asalariada 4497102     37.2
4 Clase obrera - autónoma 1298505     10.7
5 Clase obrera - asalariada 3503745     29.0
6 Clase obrera - trabajadores marginales 760754      6.29
```

# DISTRIBUCIONES DE FRECUENCIAS (summarytools)

También podemos utilizar la función `freq` del paquete `summarytools`, que debemos instalarlo previamente.

```
1 library(summarytools)
2
3 freq(eph_419$class6_factor, weights = eph_419$PONDERA, cumul = FALSE, repor
```

```
Weighted Frequencies
eph_419$class6_factor
Type: Factor
Weights: PONDERA
```

```
-----
Freq          %
-----
Clase alta          134342.00      1.11
Clase media - autónoma 1908959.00     15.77
Clase media -asalariada 4497102.00     37.16
Clase obrera - autónoma 1298505.00     10.73
Clase obrera - asalariada 3503745.00     28.95
Clase obrera - trabajadores marginales 760754.00      6.29
Total             12103407.00    100.00
```

# **6. EXPLORANDO RESULTADOS. TABLAS Y GRÁFICOS**

# CLASES E INGRESOS (I)

- Tomando como variable dependiente los ingresos totales individuales (P47T), calcularemos el promedio por clase social.
- Para ello utilizaremos algunas funciones del paquete `dp1yr` y de `R base` como `weighted.mean`.

```
1 eph_419 %>%
2   filter(!is.na(clase6_factor)) %>% # Quitamos los casos que no tienen cla
3   group_by(clase6_factor) %>% # Agrupamos por clase
4   summarise(media = weighted.mean(P47T, w = PONDII)) # Calculamos el promedi

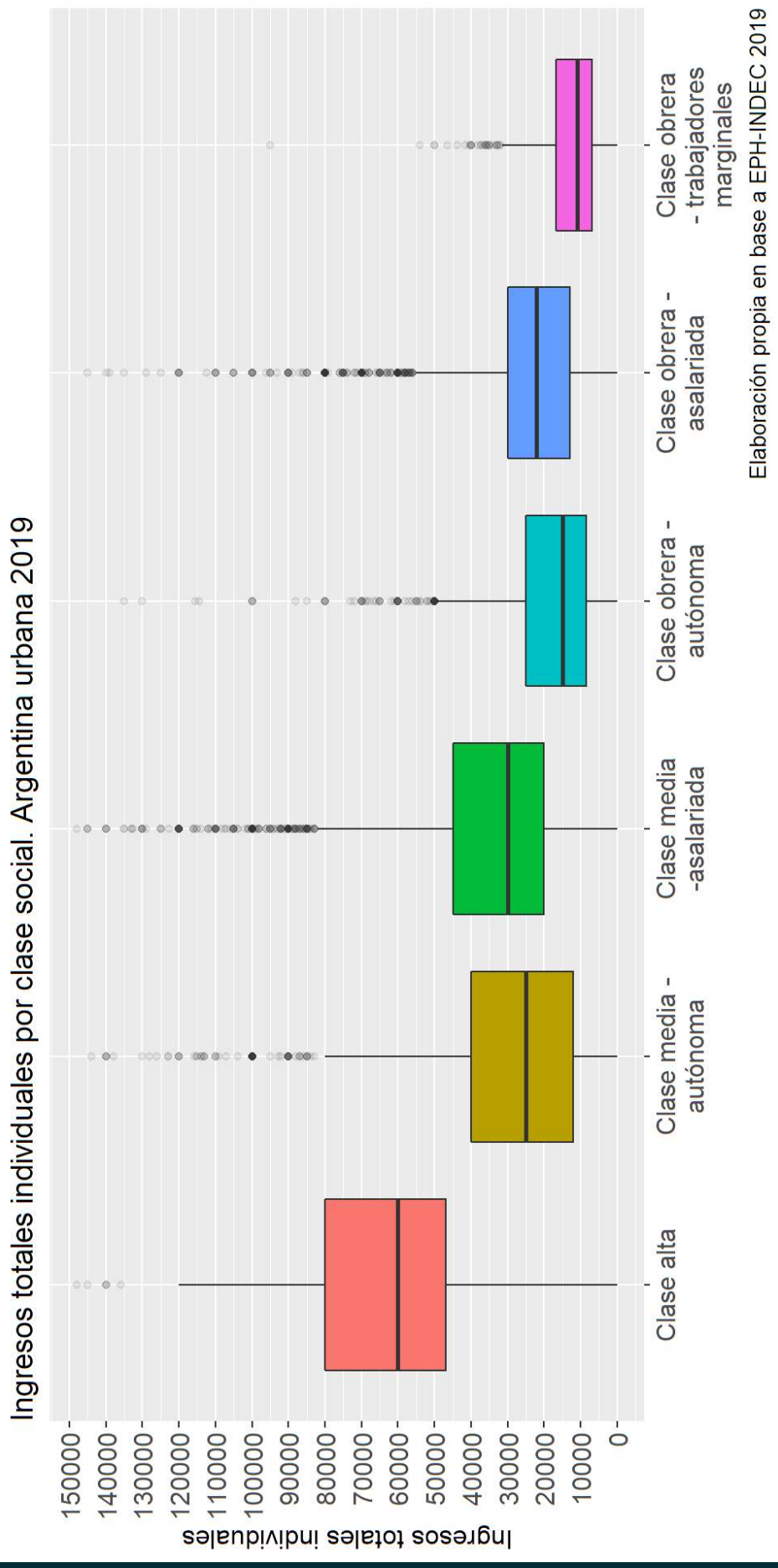
# A tibble: 6 x 2
  clase6_factor      media
  <fct>
1 Clase alta          76779.
2 Clase media - autónoma 33651.
3 Clase media -asalariada 37616.
4 Clase obrera - autónoma 19315.
5 Clase obrera - asalariada 24375.
6 Clase obrera - trabajadores marginales 12578.
```

# CLASES E INGRESOS (II)

- Otra forma de observar las desigualdades es a través de gráficos
- En este caso, a través del paquete **ggplot2** del **tidyverse**, construiremos paso a paso un diagrama de cajas y bigotes (*boxplot*).

```
1 eph_419 %>%
2   filter(!is.na(clase6_factor), P47T < 150000) %>%
3   ggplot(aes(x=clase6_factor, y=P47T, fill=clase6_factor, weight = PONDII))
4   geom_boxplot(outlier.alpha = 0.1, show.legend = FALSE) +
5   labs(title = "Ingresos totales individuales por clase social. Argentina u
6     y = "Ingresos totales individuales",
7     caption = "Elaboración propia en base a EPH-INDEC 2019") +
8   theme(plot.caption = element_text(size = 10),
9         plot.title = element_text(size = 14),
10        axis.title.x = element_blank(),
11        axis.title.y = element_text(size = 12),
12        axis.text.x = element_text(size = 12),
13        axis.text.y = element_text(size = 12)) +
14   scale_x_discrete(labels = function(x) str_wrap(x, width = 20)) +
15   scale_y_continuous(breaks=seq(0, 150000, 10000))
```

# CLASES E INGRESOS (II)





# LAS CLASES EN EL TIEMPO (I)

1. Podemos analizar como ha evolucionado la estructura de clases en su composición
2. Para ello, vamos a descargar las bases de la EPH del tercer trimestre entre 2018 y 2022.

```
1 eph_1822 <- get_microdata(year = 2018:2022, trimester = 3, type = "individual") # Descargamos e
2
3 table(eph_1822$ANO4) # Exploramos los casos por año
```

```
2018 2019 2020 2021 2022
56879 57229 41685 48668 49232
```

3. Luego aplicaremos el script para la creación del sistema de clases sociales a esta nueva base.
4. A continuación mostraremos los resultados en una tabla cruzada y en un gráfico de barras apiladas.

# LAS CLASES EN EL TIEMPO (III)

- Podemos construir la tabla usando las funciones del paquete `dp1yr`.
- La función `spread` permite pasar a un formato ancho de tabla.

```
1 eph_1822 %>%
2 filter(!is.na(clase6_factor)) %>% #Filtramos a los que no tienen clase
3 group_by(ANO4, clase6_factor) %>% #Agrupamos por año y clase
4 tally(wt=PONDERA) %>% #Calculamos la frecuencias ponderadas
5 group_by(ANO4) %>% #Agrupamos por año
6 mutate(percent = round((n/sum(n))*100, digits = 2)) %>% #Calculamos el po
7 select(clase6_factor, ANO4, percent) %>% #Seleccionamos las columnas con
8 spread(ANO4, value = percent) #Cambiamos a formato ancho de tabla
```

```
# A tibble: 6 × 6
  clase6_factor
<fct>
1 Clase alta
2 Clase media - autónoma
3 Clase media -asalariada
4 Clase obrera - autónoma
5 Clase obrera - asalariada
6 Clase obrera - trabajadores marginales
```

	`2018`	`2019`	`2020`	`2021`	`2022`
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
0.87	0.92	1.03	0.99	0.91	
14.8	15.7	14.8	15.8	15.8	
36.4	36.3	38.8	39	36.2	
10.1	10.8	12.7	10.8	11.0	
31.0	30.0	26.9	27.6	30.7	
6.79	6.35	5.78	5.79	5.39	

# LAS CLASES EN EL TIEMPO (IV)

Otra opción de presentación es la elaboración de un gráfico de barras. Para ello recurriremos al paquete **ggplot2** y a la función **geom\_col**.

```
1 eph_1822 %>%
2 filter(!is.na(clase6_factor)) %>% #Aplicamos los mismos pasos que para el cuadro
3 group_by(ANO4, clase6_factor) %>%
4 tally(wt=PONDERA) %>%
5 group_by(ANO4) %>%
6 mutate(percent = round((n/sum(n))*100, digits = 2)) %>%
7 ggplot(mapping = aes(x = ANO4, y = percent, fill = clase6_factor)) + #Indicamos qué variables
8 geom_col() + #Usaremos columnas
9 geom_text(aes(label = percent), position = position_stack(.5), size = 3) + #Agregamos los por
10 scale_fill_discrete(labels = function(x) str_wrap(x, width = 25)) +
11 labs(title = "Evolución de la estructura de clases sociales",
12       subtitle = "Población ocupada. Argentina urbana 2018-2022.",
13       caption = "Fuente: Elaboración propia en base a EPH-INDEC 2018-2022.",
14       fill = "Clase social") +
15 theme(legend.title = element_text(size = 10, face = "bold"),
16       legend.text = element_text(size = 11),
17       axis.title.x = element_blank(),
18       axis.title.y = element_blank(),
19       axis.text.y = element_text(size = 10),
20       axis.text.x = element_text(size = 10),
21       plot.title = element_text(size = 14, face = "bold"),
22       plot.subtitle = element_text(size = 12),
23       plot.caption = element_text(size = 10, hjust = 1)) +
```

# LAS CLASES EN EL TIEMPO (IV)

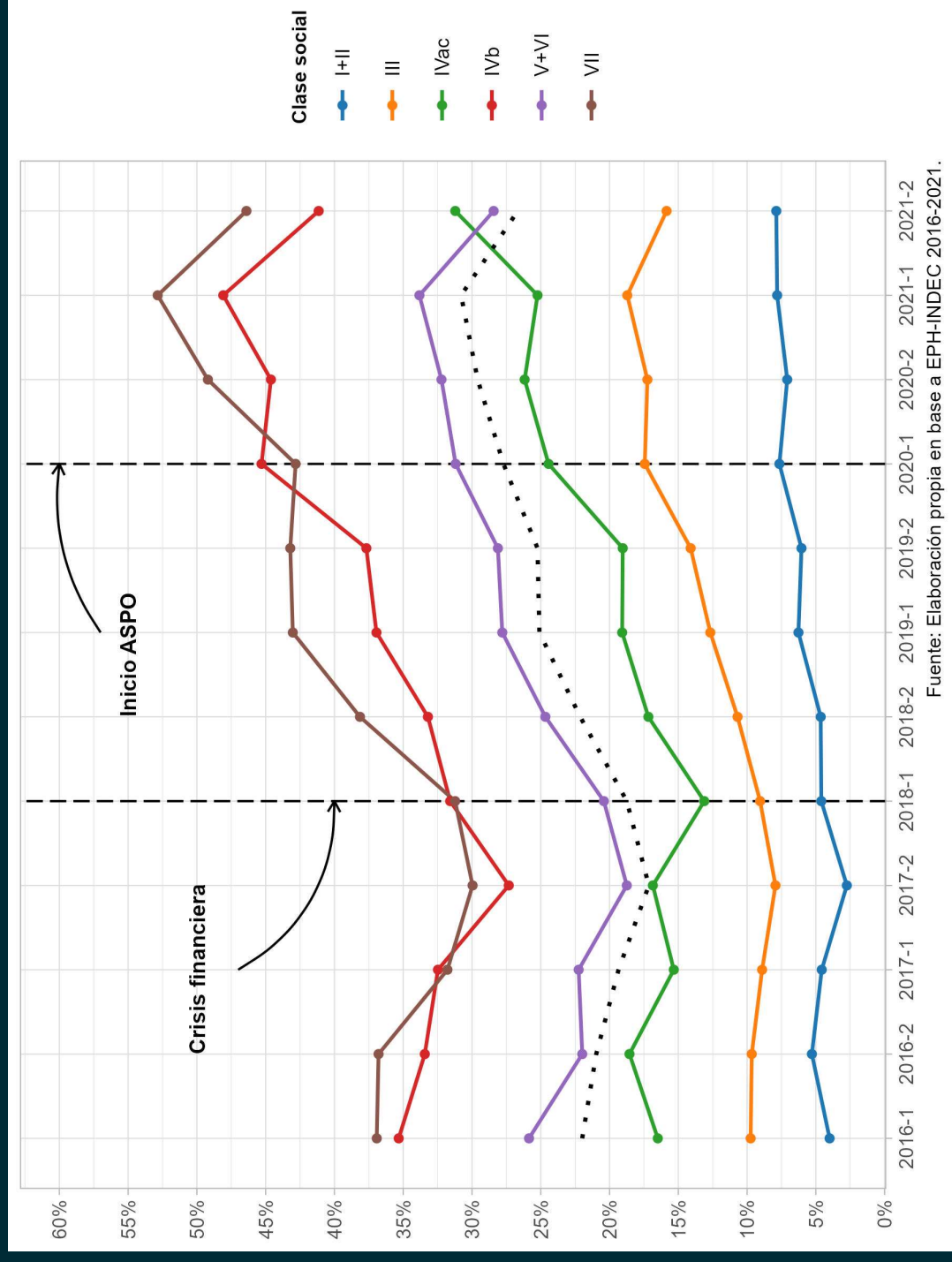
## Evolución de la estructura de clases sociales

Población ocupada. Argentina urbana 2018-2022.



Fuente: Elaboración propia en base a EPH-INDEC 2018-2022.

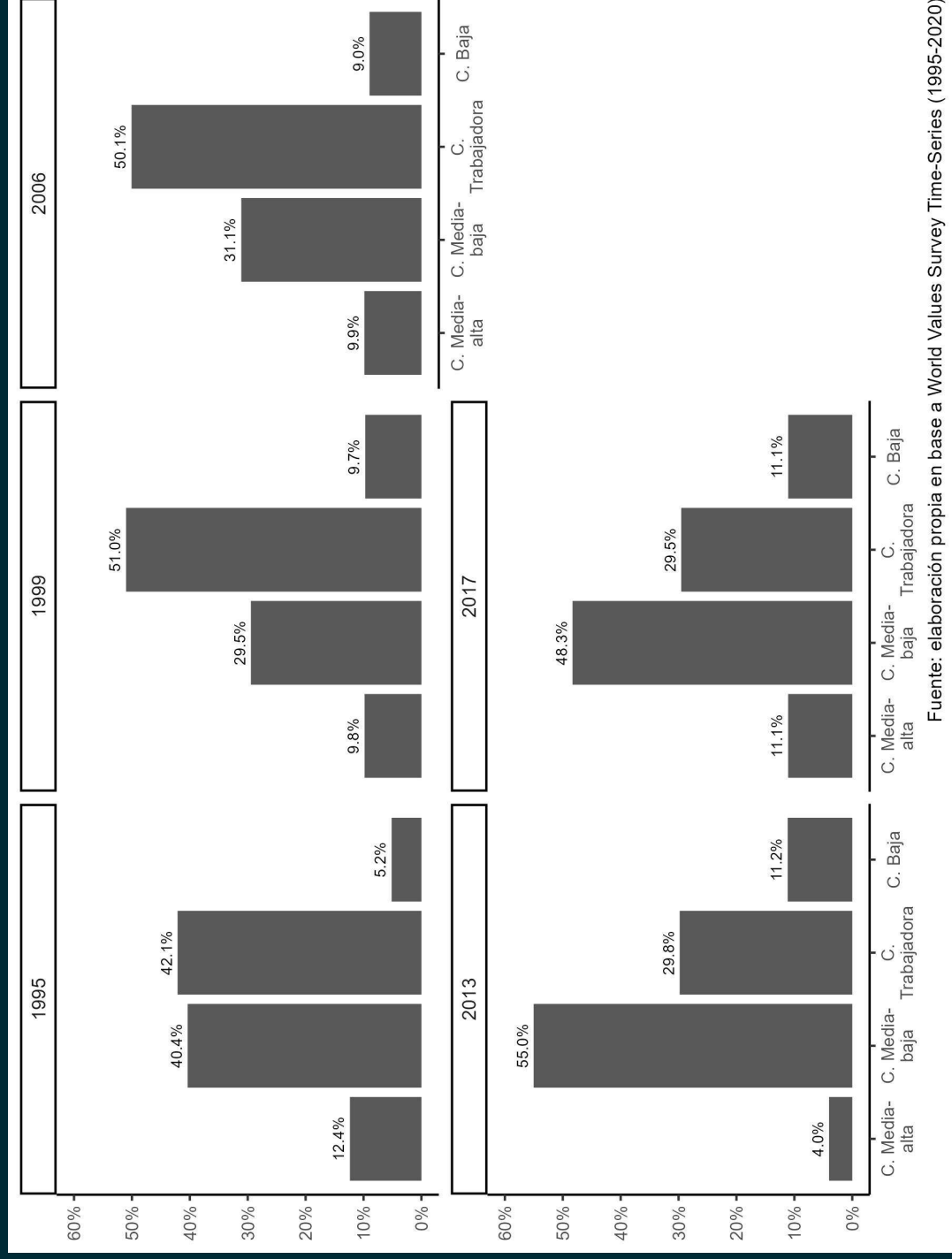
# OTROS EJEMPLOS DE GRÁFICOS (I)



Evolución de la pobreza por clase social. Argentina Urbana 2016-2021



# OTROS EJEMPLOS DE GRÁFICOS (II)

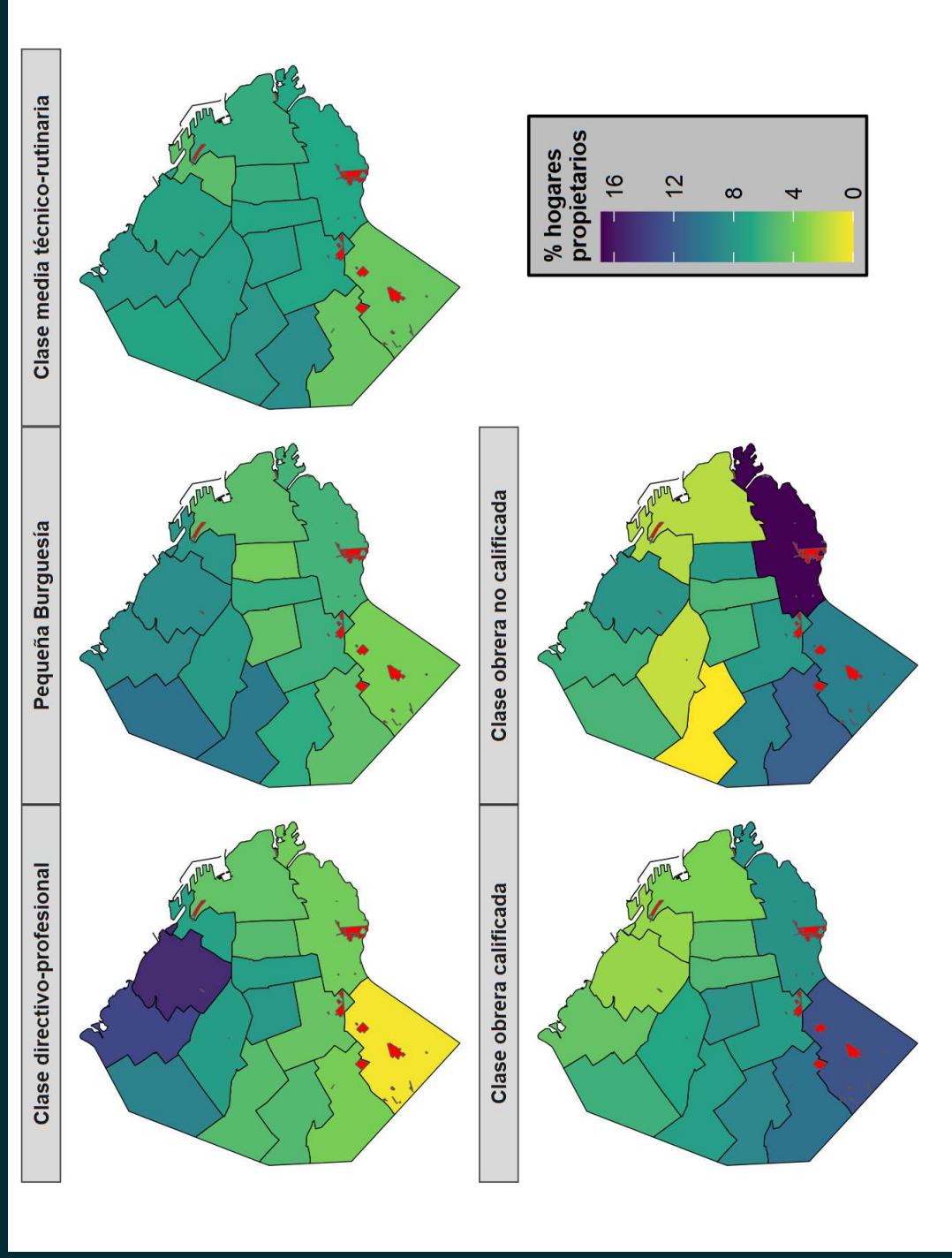


Percepción de clase social en Argentina. 1995-2017





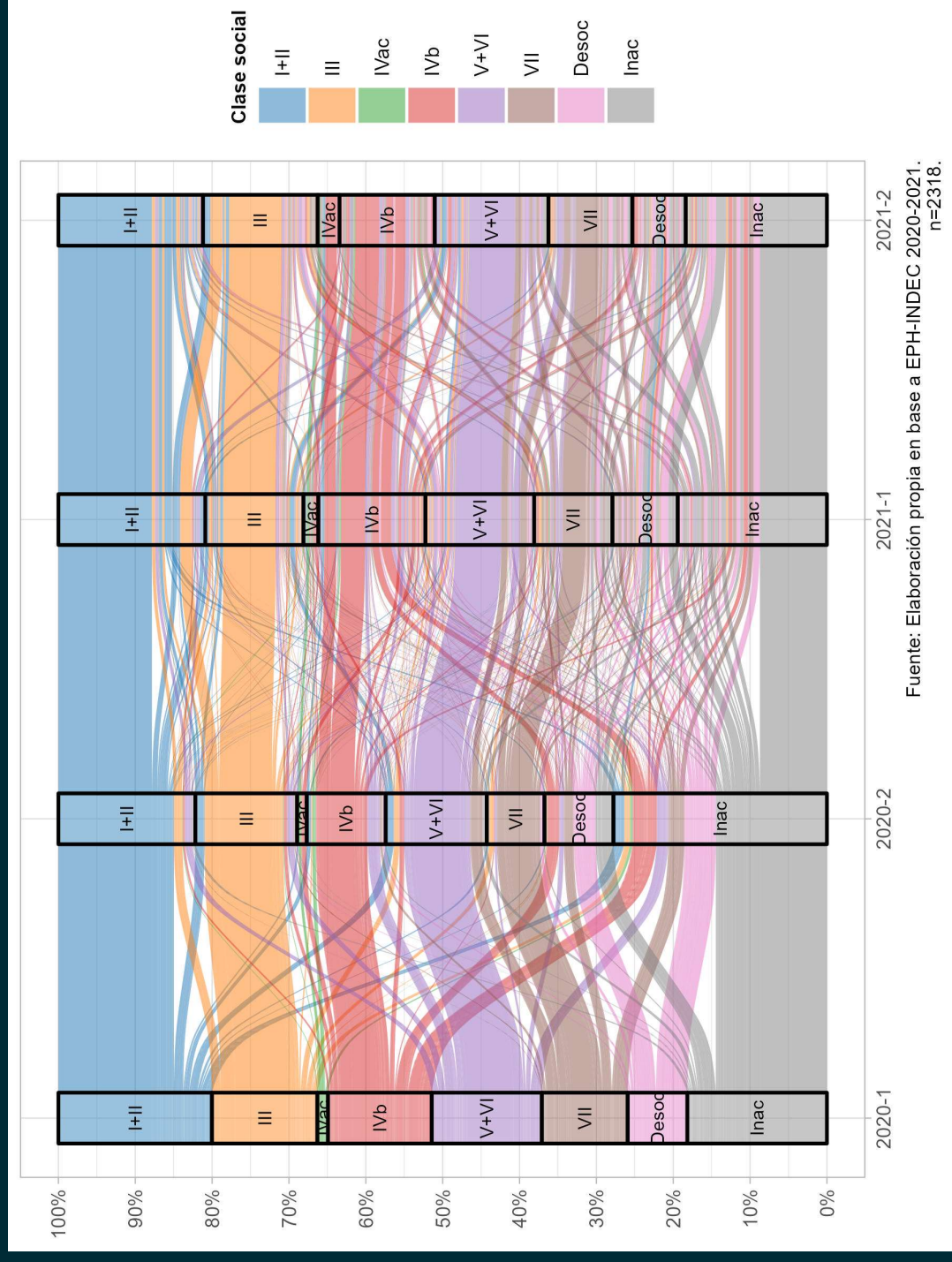
# OTROS EJEMPLOS DE GRÁFICOS (III)



Porcentaje de hogares propietarios de la vivienda según clase social. CABA 2015.



# OTROS EJEMPLOS DE GRÁFICOS (IV)



Transiciones de clase. Argentina urbana, 1er y 2do trimestre 2020 - 2021. Población mayor de 18 años.



# 7. DUDAS Y CONSULTAS

