Clase 2: tidyverse

Métodos Cuantitativos II

Sebastián Muñoz-Tapia

Antropología UAH

2023-03-15

FORMALIDADES DEL CURSO

Elementos de horario de entrada y participación

- ASISTENCIA:
  - mínimo 70% de actividades lectivas. En total son 26 clases considerando separadamente los dos bloques (viernes 08:30 a 09:50 y viernes 10:00 a 11:20). Se requiere 18 de cada bloque, aceptando 8 inasistencias.
  - La lista por bloque se retirará a: 35 m. al comienzo del primer bloque (9:05) y 15 m. al comienzo del segundo bloque. Posterior a eso no se reconocerá la asistencia.

Flexibilidad (limitada)

- Personas con **9 o 10** inasistencias podrán a optar a una prueba recuperativa si tienen un promedio mayor a **5,5** en evaluaciones individuales. La prueba será individual e incluye todos los contenidos del curso. Con nota superior a **4,0** podrán aprobar el curso en términos de asistencia.
- En el caso específico de todos los trabajos, por **cada día de atraso** en una entrega se descontarán **0,25** puntos de la nota final.
- Horario de atención estudiantes/ 48 Horas antes: **Miércoles desde 16:00**

Clases online desde 15 de Mayo

- hacer prueba en TEAMS: 24 de Marzo

Evaluaciones

- *Avance 1*: Entrega del formulario corregido Entrega del formulario corregido considerando sugerencias y observaciones de compañerxs, para posterior aplicación / Grupal/ **5%**
- *Evaluación Individual 1*: Prueba presencial individual: aspectos básicos de programación en R y tidyverse /Individual/ **20%**
- *Presentación de textos*: Discusión sobre análisis de datos: Grupal/ 2 fechas/ **10%**
- *Avance 2*: Presupuesto en Excel;Procesamiento (limpieza, transformaciones y recodificaciones); análisis (distribución de frecuencias, tablas de contingencia) /Grupal/ **10%**
- *Evaluación Individual 2*: Prueba presencial individual manipulación de bases y estadística descriptiva Individual/ **20%**
- *Asistencia, participación en clases y talleres*: Individual/ **15%**
- *Trabajo Final*: Incorpora trabajo de campo realizado, presupuesto, procesamientos estadísticos más relevantes y gráficos. Se exponen resultados / Grupal/ **20%**

Textos a presentar

- Becker, H. (2018). Datos, pruebas e ideas. Por qué los científicos sociales deberían tomárselos más en serio y aprender de sus errores. Siglo XXI (19-40; 63-87)
- Best, J. (2004). Uso y abuso de las estadísticas. La distorsión en la percepción de los problemas sociales y políticos. Cuatro Vientos. (1-62)
- D'Ignazio, C., & Klein, L. (2020). Data feminism. En Information, Communication & Society (Vol. 24, Número 13). The MIT Press. https://doi.org/10.1080/1369118x.2020.1836249 (1-48)
- Sevilla Moroder, J. (2005). Gramática de las gráficas. Pistas para mejorar las representaciones de datos. Universidad Pública de Navarra (11-50)
- Sosa-Escudero, W. (2019). Big data. Breve manual para conocer la ciencia de datos que ya invadió nuestras vidas. Siglo XXI. (11-45) ; 67-87)

vector, data.frame, matrix

Vector

2
3
4
5
1

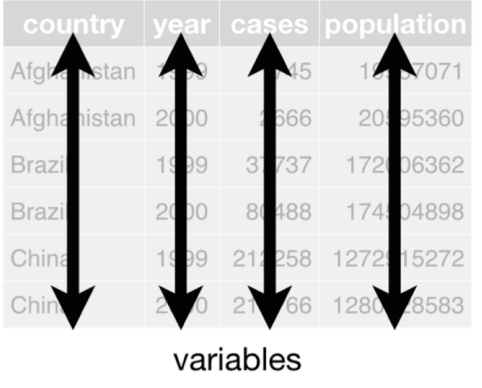
Data frame

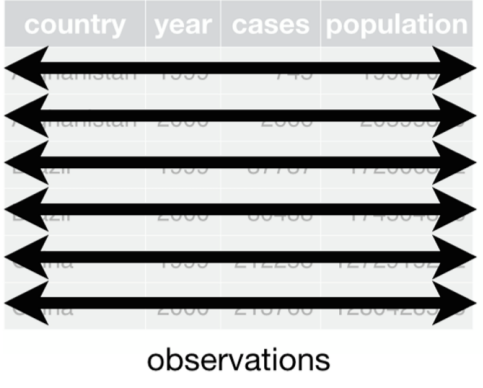
2	a	0
3	b	3
4	c	7
5	v	3
1	f	6

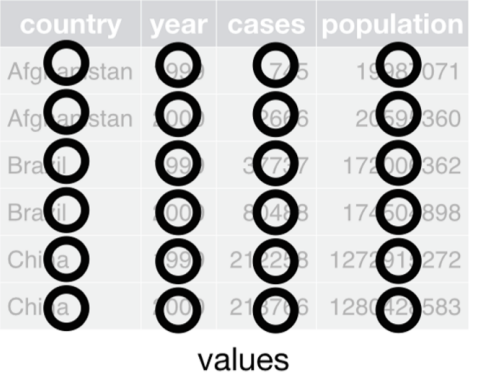
Matrix

2	8	0
3	6	3
4	5	7
5	4	3
1	3	6

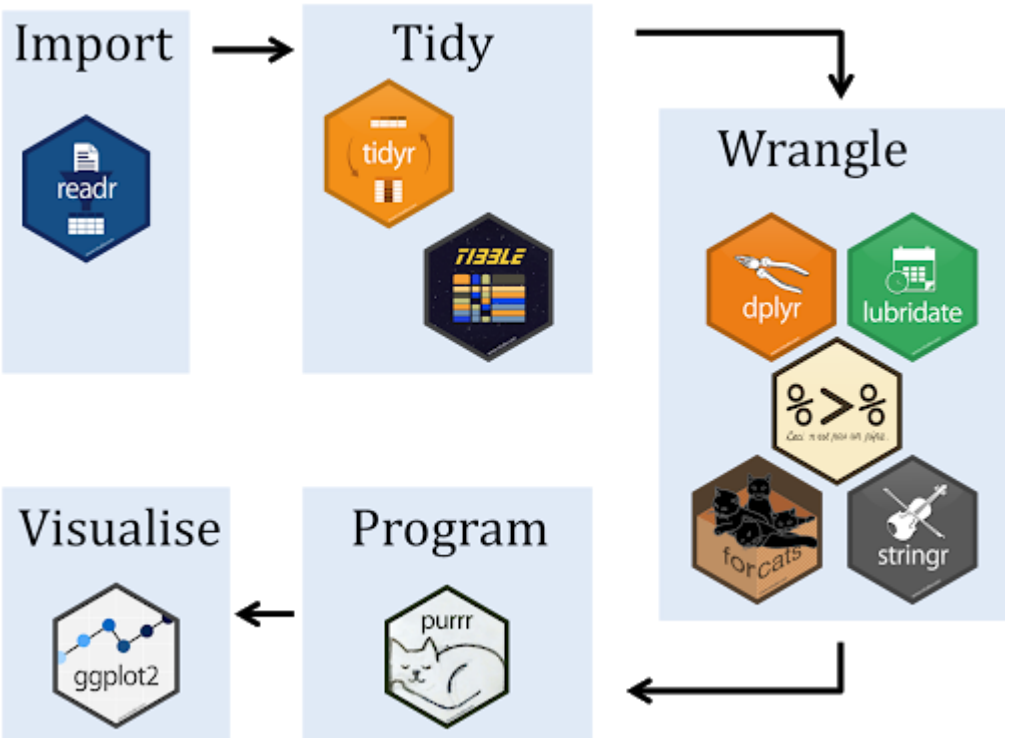
dentro de un data frame







tidyverse

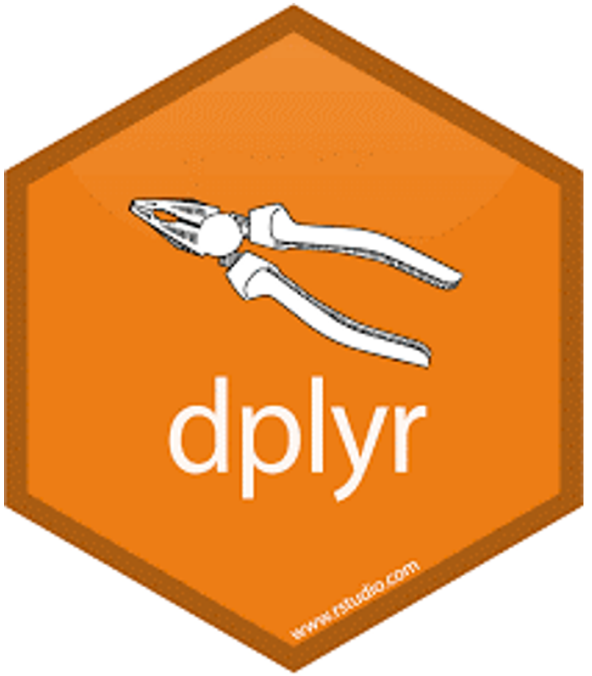

Import
readr
Tidy
tidyr
tibble
Wrangle
dplyr
lubridate
%>%
forcats
stringr
Program
purrr
Visualise
ggplot2

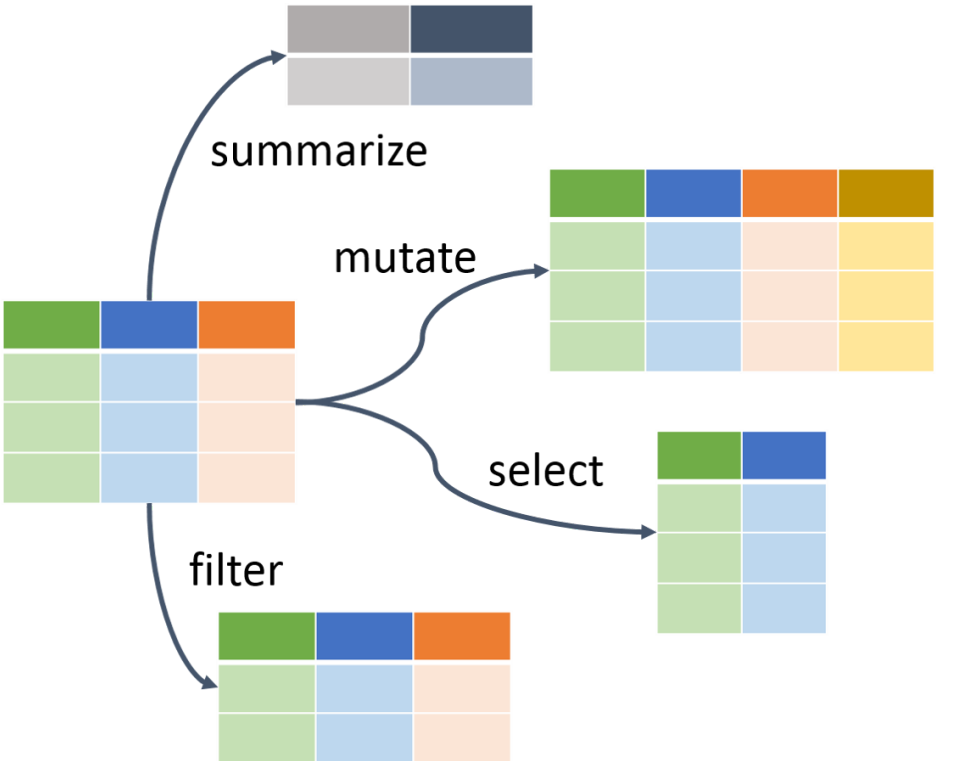
el pipe %>%

A QUICK INTRODUCTION TO DPLYR

```
dataframe %>%
    filter(…) %>%
    select(…) %>%
    mutate(…) %>%
    arrange(…) %>%
    summarise(…)
```

dplyr





select ()

Subset Variables (Columns)

df

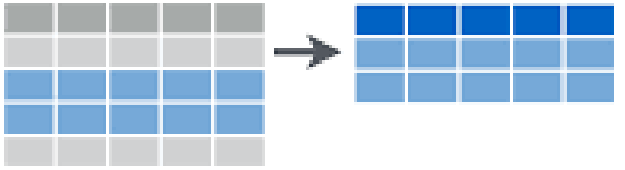
color	value
blue	1
black	2
blue	3
blue	4
black	5

value
1
2
3
4
5

```
select(df, -color)
```

filter ()

Subset Observations (Rows)



df

color	value
blue	1
black	2
blue	3
blue	4
black	5

color	value
blue	1
blue	4

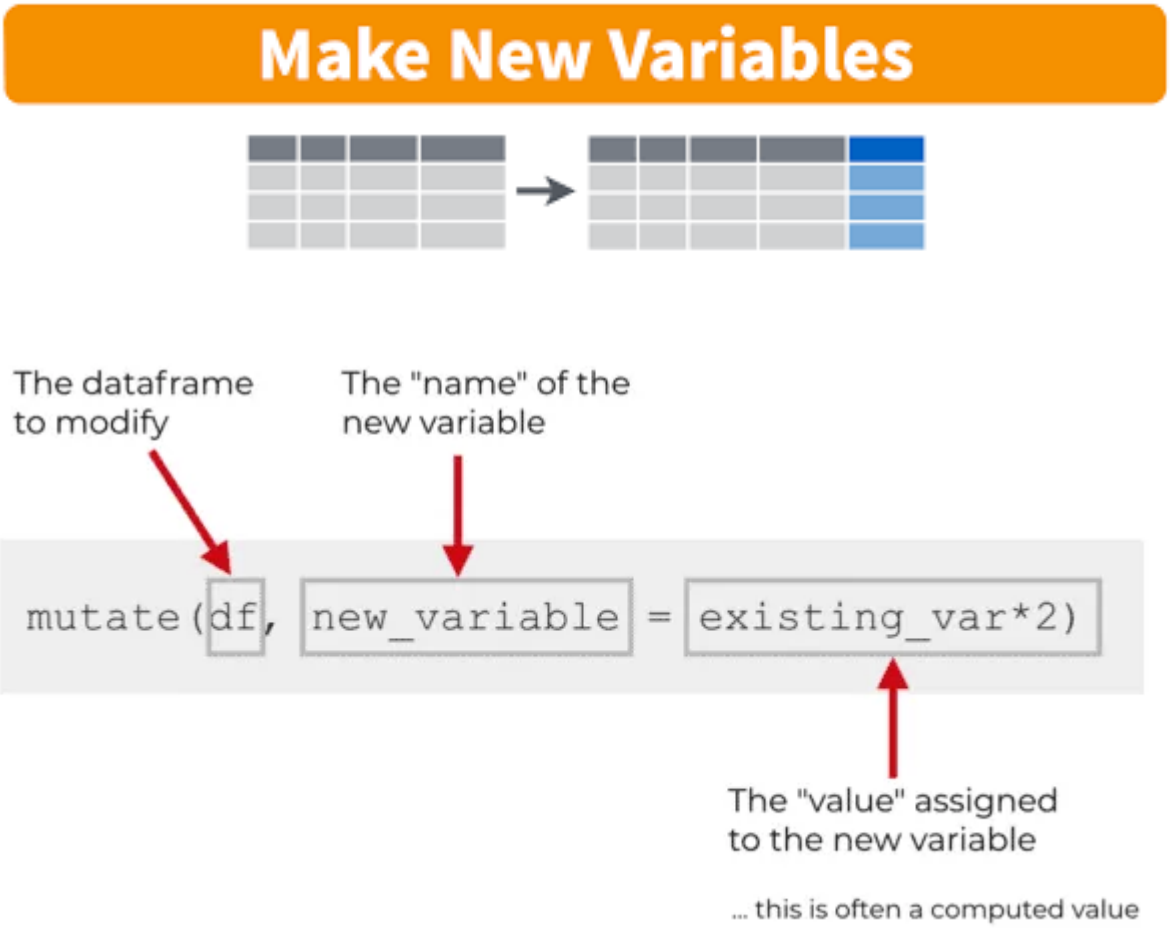
```
filter(df, value %in% c(1, 4))
```

Operadores

Operadores					
Aritméticos		Comparativos		Lógicos	
`+`	adición	`<`	menor que	`! x`	NO lógico
`-`	substracción	`>`	mayor que	`x & y`	Y lógico
`*`	multiplicación	`<=`	menor o igual que	`x && y`	id.
`/`	división	`>=`	mayor o igual que	`x | y`	O lógico
`^`	potencia	`==`	igual	`x || y`	id.
`%%`	módulo	`!=`	diferente de	`xor(x, y)`	O exclusivo
`%/%`	división de enteros				

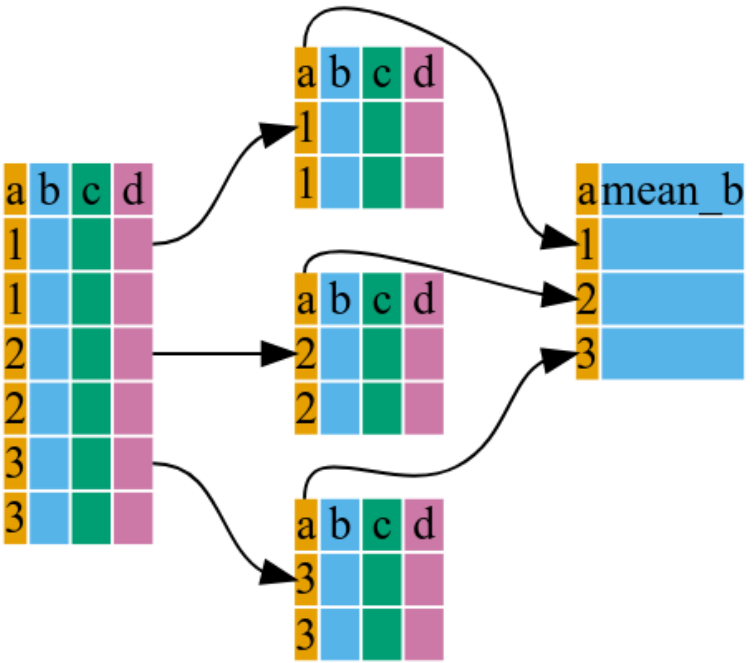
- %in%: lo que está dentro de...

Mutate

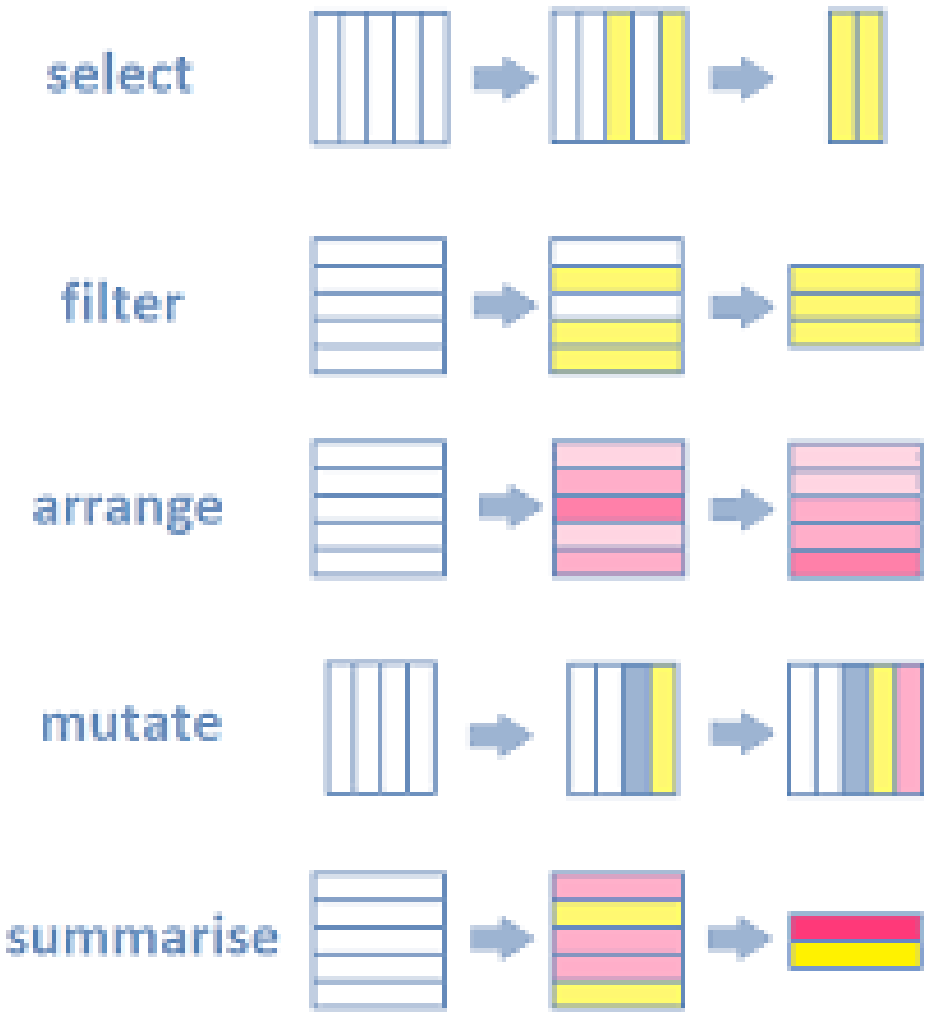

Make New Variables
The dataframe to modify
The "name" of the new variable
mutate(df, new_variable = existing_var*2)
The "value" assigned to the new variable
... this is often a computed value

Group_by + summarize

```
gapminder %>%
group_by(a) %>%
summarize(mean_b=mean(b))
```



Resumen



Entonces...

- Si en mi en la base de datos de mi encuesta, quisiera:
  - trabajar sólo con las mujeres ¿qué función utilizaría?
  - trabajar sólo con las variables que me interesan por ser el foco de mi grupo (e.g. solo las de "política" o "lectura")
  - hacer una tabla de la media y la mediana de la edad por sexo.
  - recodificar los ingresos en 3 grupos: altos, medios, bajos