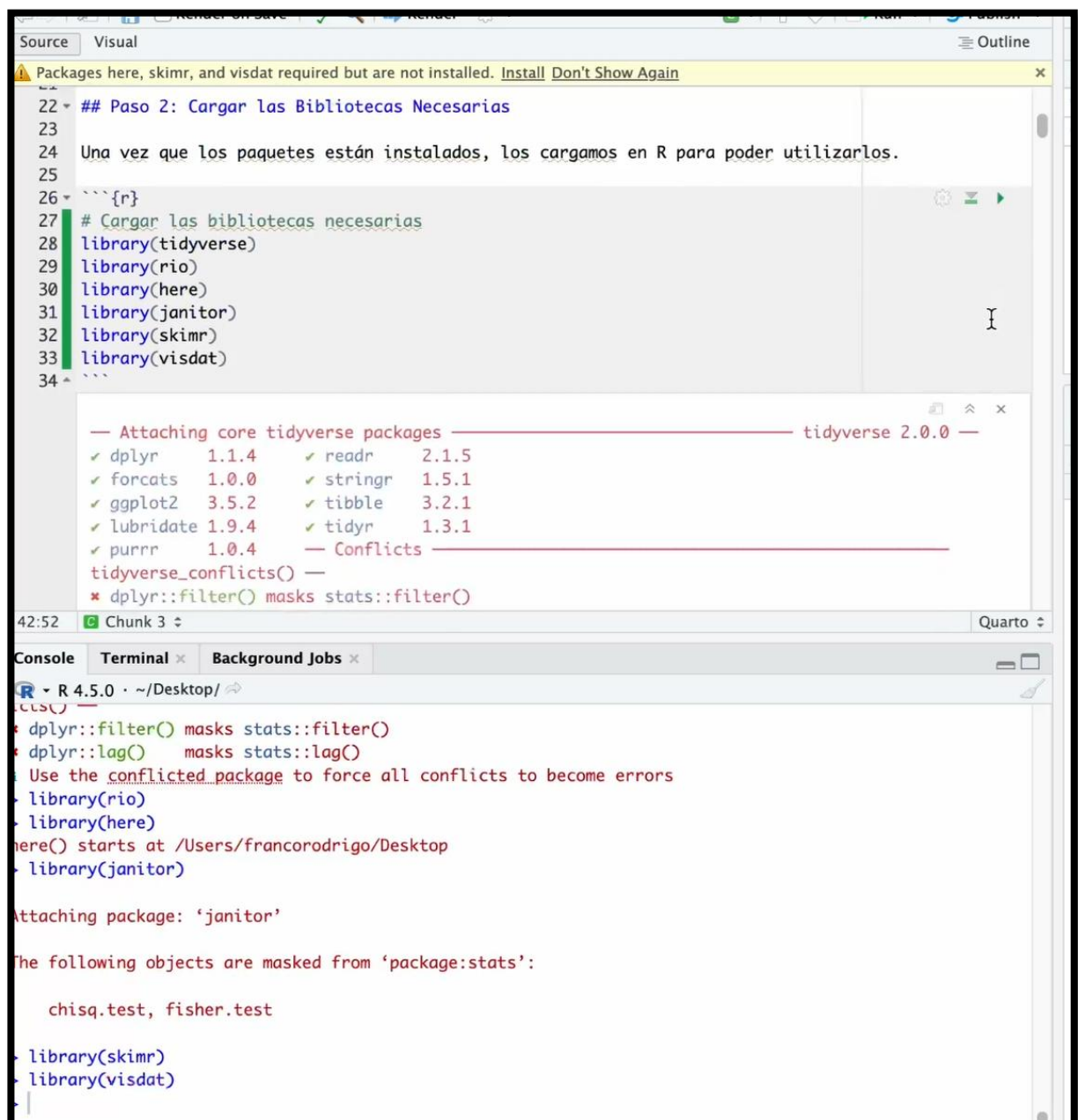


DESARROLLO -ACTIVIDAD 2:

Integrantes: (EQUIPO 7)

- Llancari Nivin Meyli
- Alejo Huaman, Melissa
- Vera Fonseca, July
- Ccuro Minaya, Lucia
- Flores Diaz, Christian



```
Source Visual Outline
⚠ Packages here, skimr, and visdat required but are not installed. Install Don't Show Again x
22 ## Paso 2: Cargar las Bibliotecas Necesarias
23
24 Una vez que los paquetes están instalados, los cargamos en R para poder utilizarlos.
25
26 ```{r}
27 # Cargar las bibliotecas necesarias
28 library(tidyverse)
29 library(rio)
30 library(here)
31 library(janitor)
32 library(skimr)
33 library(visdat)
34 ```

— Attaching core tidyverse packages — tidyverse 2.0.0 —
✓ dplyr 1.1.4 ✓ readr 2.1.5
✓ forcats 1.0.0 ✓ stringr 1.5.1
✓ ggplot2 3.5.2 ✓ tibble 3.2.1
✓ lubridate 1.9.4 ✓ tidyr 1.3.1
✓ purrr 1.0.4 — Conflicts —
tidyverse_conflicts() —
* dplyr::filter() masks stats::filter()

42:52 Chunk 3 ↕ Quarto ↕

Console Terminal x Background Jobs x
R - R 4.5.0 · ~/Desktop/
(RStudio)
* dplyr::filter() masks stats::filter()
* dplyr::lag() masks stats::lag()
Use the conflicted package to force all conflicts to become errors
library(rio)
library(here)
here() starts at /Users/francorodrigo/Desktop
library(janitor)

Attaching package: 'janitor'

The following objects are masked from 'package:stats':

chisq.test, fisher.test

library(skimr)
library(visdat)
```

Source Visual Outline

⚠ Packages here, skimr, and visdat required but are not installed. [Install](#) [Don't Show Again](#)

```
35
36 ## Paso 3: Importar el Dataset
37
38 En este paso, importamos el dataset desde el archivo .csv proporcionado y lo cargamos
39 correctamente en R.
40 ```{r}
41 # Importar el dataset desde el escritorio
42 data_covid_diabetes <- import('/Users/francorodrigo/Desktop/covid_19_diabetes.csv')
43
44 # Verificar las primeras filas del dataset
45 head(data_covid_diabetes)
46 ```
```

	Derivation.cohort <int>	pac_fue_hospital <chr>	duraci_hospita_diaz <int>	desenla_fallecido <chr>
1	1	Sí	15	No
2	1	Sí	14	Sí
3	1	Sí	11	No
4	1	Sí	1	No

42:52 Chunk 3 ▾ Quarto ▾

Console Terminal Background Jobs

```
R 4.5.0 - ~/Desktop/
dplyr::filter() masks stats::filter()
dplyr::lag() masks stats::lag()
Use the conflicted package to force all conflicts to become errors
library(rio)
library(here)
here() starts at /Users/francorodrigo/Desktop
library(janitor)

Attaching package: 'janitor'

The following objects are masked from 'package:stats':
  chisq.test, fisher.test

library(skimr)
library(visdat)
```

Source Visual Outline

⚠ Packages here, skimr, and visdat required but are not installed. [Install](#) [Don't Show Again](#)

```
$ Glucosa
: chr 0 203 103 102 132 0 124 273 0 0 ...
$ Glucosa_menor_60_o_mayor_500
: chr "No" "No" "No" "No" ...
$ AST_medido
: chr "Sí" "Sí" "Sí" "Sí" ...
$ AST
: int 44 21 68 23 68 67 153 106 282 493 ...
$ AST_mayor_40
: chr "Sí" "No" "Sí" "No" ...
$ ALT_medido
: chr "Sí" "Sí" "Sí" "Sí" ...
$ ALT
: num 17 15 51 19 151 38 91 32 128 474 ...
$ ALT_mayor_40
: chr "No" "No" "Sí" "No" ...
$ Glóbulos_blanco_medidos
: chr "Sí" "Sí" "Sí" "Sí" ...
$ Glóbulos_blanco_menor_1.8_o_mayor_4.8
: num 3.7 10 8.5 7.1 10.4 7.7 11.6 6.5 12.7 16.5 ...
$ Linfocitos_medidos
: chr "No" "Sí" "Sí" "Sí" ...
$ Linfocitos
: num 0.5 0.5 1.4 2.9 0.9 0.9 1.4 0.5 2.5 0.9 ...
$ Linfocitos_menor_1
: chr "Sí" "Sí" "No" "No" ...
$ IL6_medida
: chr "No" "No" "No" "Sí" ...
$ IL6
: num 0 0 0 4 0 ...
$ IL6_mayor_150
: chr "No" "No" "No" "No" ...
$ Ferritina_medida
: chr "Sí" "No" "No" "Sí" ...
$ Ferritina
: num 816 0 0 87 1519 ...
$ Ferritina_mayor_300
: chr "Sí" "No" "No" "No" ...
$ Proteína_C_reactiva_medida
: chr "Sí" "Sí" "Sí" "Sí" ...
$ Proteína_C_reactiva
: num 18.4 5.5 11.8 0.8 11.8 8.9 9.8 16.2 29.1 7.3
...
```

42:52 Chunk 3 ▾ Quarto ▾

Source Visual Outline

⚠ Packages here, skimr, and visdat required but are not installed. [Install](#) [Don't Show Again](#)

```
69 data_covid_diabetes <- clean_names(data_covid_diabetes)
70
71 # Verificar los nuevos nombres de las columnas
72 colnames(data_covid_diabetes)
73
```

[1] "derivation_cohort"	"pac_fue_hospital"
[3] "duraci_hospita_diaz"	"desenla_fallecido"
[5] "edad"	"severidad"
[7] "raza_negra"	"raza_blanca"
[9] "asiatico"	"latino"
[11] "infacto_mioc"	"enfer_vascu_periferica"
[13] "falla_card_conges"	"enfer_cardiovascular"
[15] "demencia"	"enf_pulmonar_obstruc_cronica"
[17] "diabetes_mell_complicada"	"diabetes_mell_simple"
[19] "enfer_renal"	"all_cns"
[21] "pure_cns"	"accidente_cerebrovascular"
[23] "convulsion"	"sincope_previo"
[25] "otra_afeccion_neurolologica_previa"	"otra_lesion_cerebral"
[27] "edad_2"	"puntuacion_edad"
[29] "saturacion_o2_medida"	"saturacion_o2"
[31] "saturacion_o2_menor_94"	"temperatura_medida"
[33] "temperatura"	"temperatura_mayor_38"

12:52 Chunk 3 Quarto

Console Terminal Background Jobs

```
R - R 4.5.0 - ~/Desktop/
> source_medida
55] "sodio_menor_139_o_mayor_154" "glucosa_medida"
57] "glucosa" "glucosa_menor_60_o_mayor_500"
59] "ast_medido" "ast"
61] "ast_mayor_40" "alt_medido"
63] "alt" "alt_mayor_40"
65] "globulos_blanco_medidos" "globulos_blanco"
67] "globulos_blanco_menor_1_8_o_mayor_4_8" "linfocitos_medidos"
69] "linfocitos" "linfocitos_menor_1"
71] "il6_medida" "il6"
73] "il6_mayor_150" "ferritina_medida"
75] "ferritina" "ferritina_mayor_300"
77] "proteina_c_reactiva_medida" "proteina_c_reactiva"
79] "proteina_c_reactiva_mayor_10" "procalcitonina_medida"
81] "procalcitonina" "procalcitonina_mayor_0_1"
83] "troponina_medida" "troponina"
85] "troponina_mayor_0_1"
```


Source Visual Outline

⚠ Packages here, skimr, and visdat required but are not installed. [Install](#) [Don't Show Again](#)

```
1st Qu.:0.01000  class :character
Median :0.01000  Mode  :character
Mean   :0.06096
3rd Qu.:0.02000
Max.   :7.81000
```

90

91 **## Paso 7: Transformación de los Datos**

92

93 Transformamos algunas variables, por ejemplo, convirtiendo las variables binarias (Sí/No) a factores.

94

95 ```{r}

96 # Convertir las variables binarias a factores

97 data_cleaned\$pac_fue_hospital <- factor(data_cleaned\$pac_fue_hospital, levels = c("Sí", "No"))

98 data_cleaned\$desenla_fallecido <- factor(data_cleaned\$desenla_fallecido, levels = c("Sí", "No"))

99

100 # Verificar los datos transformados

101 summary(data_cleaned)

102 ```

103

104 **## Paso 8: Exportación de los Datos Limpiados**

105

42:52 Chunk 3 Quarto

Console Terminal Background Jobs

R 4.5.0 ~/Desktop/

```
mean : 9.652
3rd Qu.:15.400
Max. :60.600

procalcitonina_medida procalcitonina procalcitonina_mayor_0_1 troponina_medida
Length:686           Min. : 0.000 Length:686           Length:686
Class :character     1st Qu.: 0.000 Class :character     Class :character
Mode :character      Median : 0.100 Mode :character      Mode :character
Mean : 1.461
3rd Qu.: 0.300
Max. :50.000

troponina troponina_mayor_0_1
Min. :0.00000 Length:686
1st Qu.:0.01000 Class :character
Median :0.01000 Mode :character
Mean :0.06096
3rd Qu.:0.02000
Max. :7.81000
```

Source Visual Outline

⚠ Packages here, skimr, and visdat required but are not installed. [Install](#) [Don't Show Again](#)

Mode :character	Median : 0.100	Mode :character	Mode :character
	Mean : 1.461		
	3rd Qu.: 0.300		
	Max. :50.000		
troponina	troponina_mayor_0_1		
Min. :0.00000	Length:686		
1st Qu.:0.01000	Class :character		
Median :0.01000	Mode :character		
Mean :0.06096			
3rd Qu.:0.02000			
Max. :7.81000			

```
103
104 ## Paso 8: Exportación de los Datos Limpiados
105
106 Finalmente, exportamos el dataset limpio y transformado a un nuevo archivo .csv.
107
108 ```{r}
109 # Exportar los datos limpios a un nuevo archivo .csv
110 export(data_cleaned, 'data/cleaned_covid_19_diabetes.csv')
111 ```
112
113
```

12:52 Chunk 3 Quarto

Console Terminal Background Jobs

R 4.5.0 ~/Desktop/

procalcitonina_medida	procalcitonina	procalcitonina_mayor_0_1	troponina_medida
Length:686	Min. : 0.000	Length:686	Length:686
Class :character	1st Qu.: 0.000	Class :character	Class :character
Mode :character	Median : 0.100	Mode :character	Mode :character
	Mean : 1.461		
	3rd Qu.: 0.300		
	Max. :50.000		
troponina	troponina_mayor_0_1		
Min. :0.00000	Length:686		
1st Qu.:0.01000	Class :character		
Median :0.01000	Mode :character		
Mean :0.06096			
3rd Qu.:0.02000			
Max. :7.81000			

```
# Exportar los datos limpios a un nuevo archivo .csv
export(data_cleaned, 'data/cleaned_covid_19_diabetes.csv')
```