



Práctica 2 | Procesamiento de Datos

Tomas Sebastian Hinojosa Meza

Instituto Tecnológico de Pabellón de Arteaga.

Ingeniería del Conocimiento

Profesor: Eduardo Flores Gallegos.

Con base en la base de datos proporcionada responda las siguientes preguntas:

¿Cuál es la edad media de mujeres que enfermaron de covid?

La media es: **37.97**

Comandos: `Edad <- dplyr::filter(d_covid, SEXO == 1)`
`> summary(Edad)`


```
> Edad <- dplyr::filter(d_covid, SEXO == 1)
> summary(Edad)
```

FECHA_ACTUALIZACION	ID_REGISTRO	ORIGEN	SECTOR
Min. :2023-01-24	Length:3748297	Min. :1.000	Min. : 1.000
1st Qu.:2023-01-24	Class :character	1st Qu.:2.000	1st Qu.: 4.000
Median :2023-01-24	Mode :character	Median :2.000	Median : 9.000
Mean :2023-01-24		Mean :1.767	Mean : 7.852
3rd Qu.:2023-01-24		3rd Qu.:2.000	3rd Qu.:12.000
Max. :2023-01-24		Max. :2.000	Max. :99.000

ENTIDAD_UM	SEXO	ENTIDAD_NAC	ENTIDAD_RES
Length:3748297	Min. :1	Length:3748297	Length:3748297
Class :character	1st Qu.:1	Class :character	Class :character
Mode :character	Median :1	Mode :character	Mode :character
	Mean :1		
	3rd Qu.:1		
	Max. :1		

MUNICIPIO_RES	TIPO_PACIENTE	FECHA_INGRESO	FECHA_SINTOMAS
Length:3748297	Min. :1.000	Min. :2022-01-01	Min. :2022-01-01
Class :character	1st Qu.:1.000	1st Qu.:2022-01-23	1st Qu.:2022-01-20
Mode :character	Median :1.000	Median :2022-05-21	Median :2022-05-19
	Mean :1.036	Mean :2022-05-13	Mean :2022-05-11
	3rd Qu.:1.000	3rd Qu.:2022-07-22	3rd Qu.:2022-07-20
	Max. :2.000	Max. :2023-01-23	Max. :2023-01-23

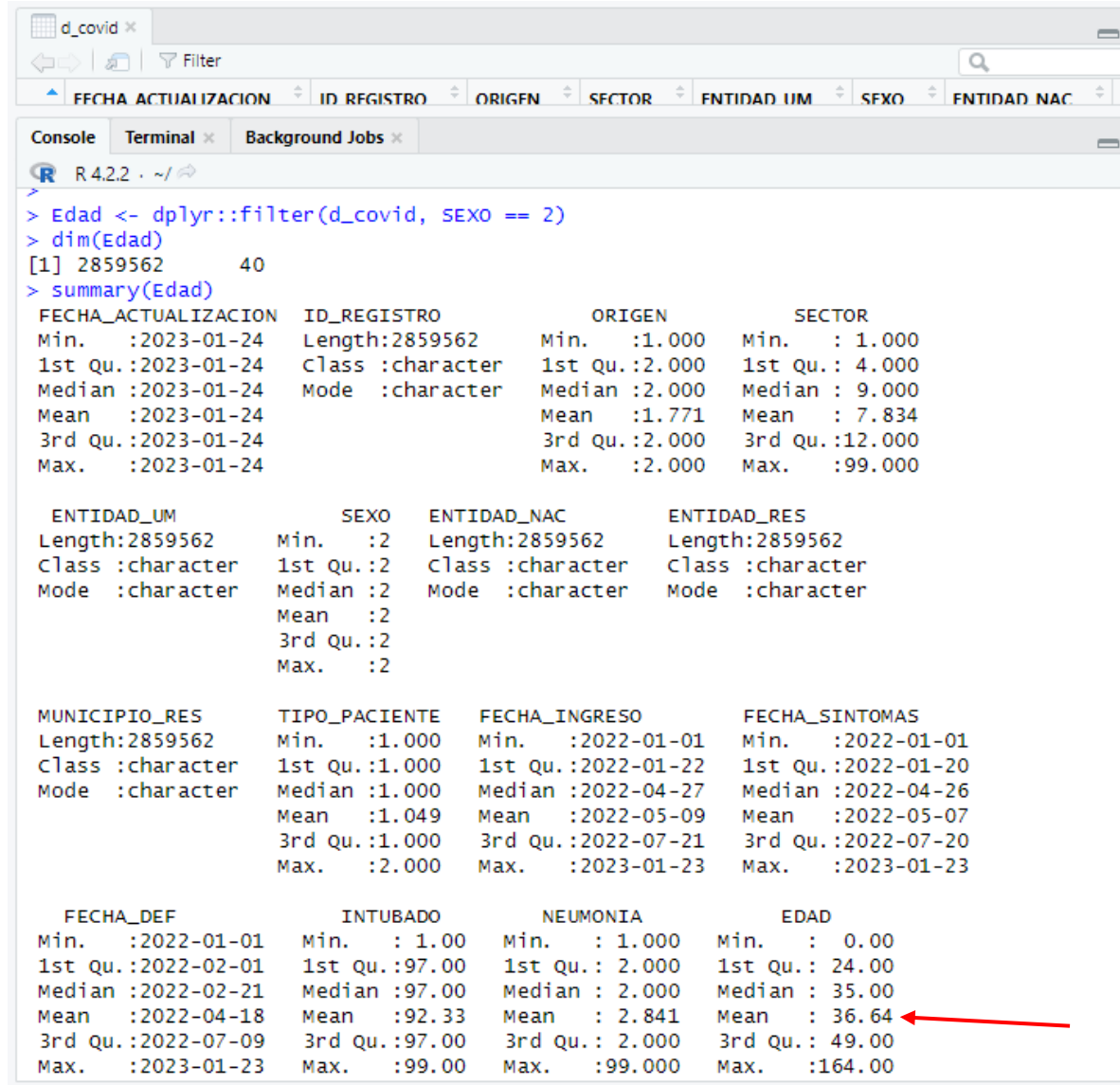
FECHA_DEF	INTUBADO	NEUMONIA	EDAD
Min. :2022-01-01	Min. : 1.00	Min. : 1.000	Min. : 0.00
1st Qu.:2022-02-01	1st Qu.:97.00	1st Qu.: 2.000	1st Qu.: 26.00
Median :2022-02-25	Median :97.00	Median : 2.000	Median : 36.00
Mean :2022-04-24	Mean :93.58	Mean : 3.049	Mean : 37.97
3rd Qu.:2022-07-16	3rd Qu.:97.00	3rd Qu.: 2.000	3rd Qu.: 49.00
Max. :2023-01-23	Max. :99.00	Max. :99.000	Max. :266.00
NA's :3731166			



¿Cuál es la edad media de hombres que enfermaron de covid?

La media es: 36.64

Comandos: `Edad <- dplyr::filter(d_covid, SEXO == 1)`
> `summary(Edad)`



The screenshot shows the RStudio interface with a console window. The console displays the following commands and output:

```
> Edad <- dplyr::filter(d_covid, SEXO == 2)
> dim(Edad)
[1] 2859562      40
> summary(Edad)
```

FECHA_ACTUALIZACION	ID_REGISTRO	ORIGEN	SECTOR
Min. :2023-01-24	Length:2859562	Min. :1.000	Min. :1.000
1st Qu.:2023-01-24	Class :character	1st Qu.:2.000	1st Qu.:4.000
Median :2023-01-24	Mode :character	Median :2.000	Median :9.000
Mean :2023-01-24		Mean :1.771	Mean :7.834
3rd Qu.:2023-01-24		3rd Qu.:2.000	3rd Qu.:12.000
Max. :2023-01-24		Max. :2.000	Max. :99.000

ENTIDAD_UM	SEXO	ENTIDAD_NAC	ENTIDAD_RES
Length:2859562	Min. :2	Length:2859562	Length:2859562
Class :character	1st Qu.:2	Class :character	Class :character
Mode :character	Median :2	Mode :character	Mode :character
	Mean :2		
	3rd Qu.:2		
	Max. :2		

MUNICIPIO_RES	TIPO_PACIENTE	FECHA_INGRESO	FECHA_SINTOMAS
Length:2859562	Min. :1.000	Min. :2022-01-01	Min. :2022-01-01
Class :character	1st Qu.:1.000	1st Qu.:2022-01-22	1st Qu.:2022-01-20
Mode :character	Median :1.000	Median :2022-04-27	Median :2022-04-26
	Mean :1.049	Mean :2022-05-09	Mean :2022-05-07
	3rd Qu.:1.000	3rd Qu.:2022-07-21	3rd Qu.:2022-07-20
	Max. :2.000	Max. :2023-01-23	Max. :2023-01-23

FECHA_DEF	INTUBADO	NEUMONIA	EDAD
Min. :2022-01-01	Min. :1.00	Min. :1.000	Min. :0.00
1st Qu.:2022-02-01	1st Qu.:97.00	1st Qu.:2.000	1st Qu.:24.00
Median :2022-02-21	Median :97.00	Median :2.000	Median :35.00
Mean :2022-04-18	Mean :92.33	Mean :2.841	Mean :36.64
3rd Qu.:2022-07-09	3rd Qu.:97.00	3rd Qu.:2.000	3rd Qu.:49.00
Max. :2023-01-23	Max. :99.00	Max. :99.000	Max. :164.00

A red arrow points to the value 36.64 in the 'Mean' row of the 'EDAD' column.

¿Cuántos hombres enfermaron en el municipio de Pabellón de Arteaga?

El numero de hombres que enfermaron en el municipio de Pabellón de Arteaga fue:

569

```
Edad <- dplyr::filter(d_covid, SEXO == 2 & as.character(ENTIDAD_RES) == "01" &
> as.character(MUNICIPIO_RES) == "006")
> dim(Edad)
[1] 569 40
```

The screenshot displays the RStudio interface. The top pane shows a data table with columns: ORIGEN, SECTOR, ENTIDAD_UM, SEXO, ENTIDAD_NAC, ENTIDAD_RES, MUNICIPIO_RES, TIPO_PACIENTE, FECHA_INGRESO, and FECHA_SINTOM. The bottom pane shows the R console with the following code and output:

```
R 4.2.2 ~ /
> Edad <- dplyr::filter(d_covid, SEXO == 1 & ENTIDAD_RES == 01 & MUNICIPIO_RES == 006)
>
> dim(Edad)
[1] 0 40
> Edad <- dplyr::filter(d_covid, SEXO == 2 & as.character(ENTIDAD_RES) == "01" & as.character(MUNICIPIO_RES) == "006")
> dim(Edad)
[1] 569 40
>
```

¿Cuál fue el municipio de Aguascalientes que tuvo la mayor cantidad de mujeres enfermas de covid?

El municipio de Aguascalientes

A	B	C
CLAVE_MUNICIPIO	MUNICIPIO	CLAVE_ENTIDAD
001	AGUASCALIENTES	01
002	ASIENTOS	01
003	CALVILLO	01
004	COSÍO	01
005	JESÚS MARÍA	01
006	PABELLÓN DE ARTEAGA	01
007	RINCÓN DE ROMOS	01
008	SAN JOSÉ DE GRACIA	01
009	TEPEZALÁ	01
010	EL LLANO	01
011	SAN FRANCISCO DE LOS ROMO	01

```
Source
Console Terminal x Background Jobs x
R 4.2.2 . ~/
> Edad <- dplyr::filter(d_covid, SEXO == 1 & as.character(ENTIDAD_RES) == "01" & as.character(MUNICIPIO_RES) == "001")
> dim(Edad)
[1] 49600 40
> Edad <- dplyr::filter(d_covid, SEXO == 1 & as.character(ENTIDAD_RES) == "01" & as.character(MUNICIPIO_RES) == "002")
> dim(Edad)
[1] 655 40
> Edad <- dplyr::filter(d_covid, SEXO == 1 & as.character(ENTIDAD_RES) == "01" & as.character(MUNICIPIO_RES) == "003")
> dim(Edad)
[1] 479 40
> Edad <- dplyr::filter(d_covid, SEXO == 1 & as.character(ENTIDAD_RES) == "01" & as.character(MUNICIPIO_RES) == "004")
> dim(Edad)
[1] 128 40
> Edad <- dplyr::filter(d_covid, SEXO == 1 & as.character(ENTIDAD_RES) == "01" & as.character(MUNICIPIO_RES) == "005")
> dim(Edad)
[1] 805 40
> Edad <- dplyr::filter(d_covid, SEXO == 1 & as.character(ENTIDAD_RES) == "01" & as.character(MUNICIPIO_RES) == "006")
> dim(Edad)
[1] 689 40
> Edad <- dplyr::filter(d_covid, SEXO == 1 & as.character(ENTIDAD_RES) == "01" & as.character(MUNICIPIO_RES) == "007")
> dim(Edad)
[1] 663 40
> Edad <- dplyr::filter(d_covid, SEXO == 1 & as.character(ENTIDAD_RES) == "01" & as.character(MUNICIPIO_RES) == "008")
> dim(Edad)
[1] 67 40
> Edad <- dplyr::filter(d_covid, SEXO == 1 & as.character(ENTIDAD_RES) == "01" & as.character(MUNICIPIO_RES) == "009")
> dim(Edad)
[1] 202 40
> Edad <- dplyr::filter(d_covid, SEXO == 1 & as.character(ENTIDAD_RES) == "01" & as.character(MUNICIPIO_RES) == "010")
> dim(Edad)
[1] 204 40
> Edad <- dplyr::filter(d_covid, SEXO == 1 & as.character(ENTIDAD_RES) == "01" & as.character(MUNICIPIO_RES) == "011")
> dim(Edad)
[1] 229 40
>
```

¿Cuántos pacientes ambulatorios hubo en el país?

La cantidad de pacientes ambulatorios que hubo en el país es: **6,332,217**

```
> Edad <- dplyr::filter(d_covid, TIPO_PACIENTE == 1)
> dim(Edad)
```

SEXO	ENTIDAD_NAC	ENTIDAD_RES	MUNICIPIO_RES	TIPO_PACIENTE	FECHA_INGRESO	FECHA_SINTOMAS	FECHA_DEF	INTUBADO
2	27	23	008	1	2022-07-22	2022-07-20	NA	
2	09	09	007	1	2022-03-09	2022-03-09	NA	
1	06	07	012	1	2022-08-23	2022-08-21	NA	
2	24	24	028	1	2022-07-07	2022-07-01	NA	
2	15	09	016	1	2022-08-02	2022-07-31	NA	
1	09	09	002	1	2022-03-06	2022-03-06	NA	
1	26	26	030	1	2022-01-11	2022-01-08	NA	
1	16	09	009	1	2022-04-26	2022-04-22	NA	
2	19	19	046	2	2022-01-01	2022-01-01	NA	
1	09	09	013	1	2022-07-28	2022-07-25	NA	
2	09	09	004	1	2022-01-09	2022-01-09	NA	
1	22	15	057	1	2022-02-21	2022-02-16	NA	

Showing 3 to 15 of 6,607,859 entries, 40 total columns

```

R 4.2.2 . ~/
> Edad <- dplyr::filter(d_covid, TIPO_PACIENTE == 1)
> dim(Edad)
[1] 6332217      40
>

```

Grafique con ggplot una gráfica con los pacientes de cada sector.

```

Source
Console Terminal Background Jobs
R 4.2.2 . ~/
> library(dplyr)

Attaching package: 'dplyr'

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

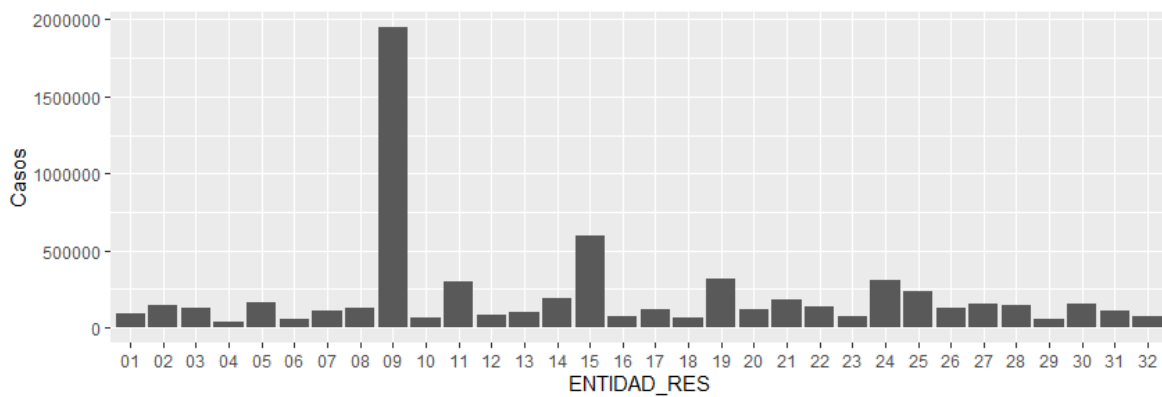
    filter, lag

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

    intersect, setdiff, setequal, union

> datos_por_estado <- d_covid %>%
+   group_by(ENTIDAD_RES) %>%
+   summarize(Casos = n())
> view(datos_por_estado)
>
> ggplot(datos_por_estado, aes(x = ENTIDAD_RES, y = Casos)) +
+   geom_bar(stat = "identity") + scale_fill_gradient(low="red", high="yellow") + theme_dark()
Error in ggplot(datos_por_estado, aes(x = ENTIDAD_RES, y = Casos)) :
  could not find function "ggplot"
> library(ggplot)
Error in library(ggplot) : there is no package called 'ggplot'
> library(ggplot2)
> ggplot(datos_por_estado, aes(x = ENTIDAD_RES, y = Casos)) +
+   geom_bar(stat = "identity") + scale_fill_gradient(low="red", high="yellow") + theme_dark()
> ggplot(datos_por_estado, aes(x = ENTIDAD_RES, y = Casos)) +
+   geom_bar(stat = "identity") + scale_fill_gradient(low="red", high="yellow") + theme_dark()
> ggplot(datos_por_estado, aes(x = ENTIDAD_RES, y = Casos)) +
+   geom_bar(stat = "identity") + scale_fill_gradient(low="red", high="yellow")
> ggplot(datos_por_estado, aes(x = ENTIDAD_RES, y = Casos)) +
+   geom_bar(stat = "identity") n + scale_fill_brewer(palette = "Blues")

```



¿Cuál es la edad máxima en la que se presenta una persona infectada?

La edad máxima es: **266?**

d_covid x

Filter

SEXO	ENTIDAD_NAC	ENTIDAD_RES	MUNICIPIO_RES	TIPO_PACIENTE	FECHA_INGRESO	FECHA_SINTOMAS	FECHA_DEF	INTUBADO
2	27	23	008	1	2022-07-22	2022-07-20	NA	
2	09	09	007	1	2022-03-09	2022-03-09	NA	
1	06	07	012	1	2022-08-23	2022-08-21	NA	
2	24	24	028	1	2022-07-07	2022-07-01	NA	
2	15	09	016	1	2022-08-02	2022-07-31	NA	
1	09	09	002	1	2022-03-06	2022-03-06	NA	
1	26	26	030	1	2022-01-11	2022-01-08	NA	
1	16	09	009	1	2022-04-26	2022-04-22	NA	
2	19	19	046	2	2022-01-01	2022-01-01	NA	
1	09	09	013	1	2022-07-28	2022-07-25	NA	
2	09	09	004	1	2022-01-09	2022-01-09	NA	
1	22	15	057	1	2022-02-21	2022-02-16	NA	

Showing 3 to 15 of 6,607,859 entries, 40 total columns

Console

Terminal x

Background Jobs x

R 4.2.2 . ~/

```

> Edad<- d_covid$EDAD
> max(d_covid$EDAD)
[1] 266
>
> summary(Edad)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.00   25.00   36.00   37.39   49.00  266.00
> |

```

Subir link de github con el reporte de la práctica. Poner comandos utilizados y respuestas a las preguntas. En la pregunta de graficar se calificará diseño.