

SEPTIEMBRE 2022 | UTN FRSR.

# TRABAJO PRÁCTICO ESTADÍSTICA N°3

Análisis de centro de Idiomas

## INTEGRANTES:

Fernando Gregoratti  
Mariano Farias  
Sandra Ferreyra  
Agustín Pérez Pesce  
Franco Sebastián Genre  
Mario Cristian Sanchez  
Braian Troncoso  
Exequiel Marín  
Santiago Mendoza  
David Ezequiel Sanchez  
Sebastián Galván



# BASE DE DATOS

Nuestra base de datos corresponde al de un centro de idiomas.

En este los alumnos cursan diferentes módulos por trimestre, a su vez todas las carreras, cada uno de los idiomas, se dividen en 5 años donde los alumnos cursan los mismos módulos cada año solo con diferente nivel de contenidos, aumentando el conocimiento y proponiendo nuevos desafíos.

Las notas trimestrales son notas numéricas entre 1 y 10, cada alumno tendrá 3 notas por modulo, y se considerara al módulo aprobado si el promedio de las notas del alumno en el módulo es mayor o igual a 6. Cada evaluación trimestral abarca los contenidos dados en el trimestre anterior.

## ESTRUCTURA BASE DE DATOS

<b>Id</b>	<b>Id del alumno</b>
<b>Nombre</b>	<b>Nombre del alumno</b>
<b>Apellido</b>	<b>Apellido del alumno</b>
<b>Modulo</b>	<b>Modulo que curso el alumno</b>
<b>Curso</b>	<b>Curso al que pertenece el módulo del alumno</b>
<b>Nota 1</b>	<b>Nota del primer trimestre</b>
<b>Nota 2</b>	<b>Nota del segundo trimestre</b>
<b>Nota 3</b>	<b>Nota del tercer trimestre</b>

## EJERCICIOS

**1 - Generar en Excel, una base de datos que tenga al menos 50 registros o más, como mínimo 5 campos, con datos numéricos y alfanuméricos.**

	A	B	C	D	E	F	G	H
1	id	nombre	apellido	modulo	curso	nota1	nota2	nota3
2	1	Daniel	Mendez	Comprension	3	7	7	6
3	2	Andrés	Diaz	Escritura	2	6	5	7
4	3	Javier	Andrade	Expresion Of	5	8	9	6
5	4	Pedro	Quiroga	Lectura	3	6	6	5
6	5	Jacinto	Rios	Comprension	4	9	8	8
7	6	Alicia	Andrade	Escritura	4	5	9	5
8	7	Mónica	De Leon	Expresion Of	1	4	5	8
9	8	Bartolomé	Carrasco	Lectura	3	8	6	7
10	9	Patricio	Hernandes	Comprension	2	2	7	6
11	10	Rodrigo	Troncoso	Escritura	4	9	6	5
12	11	Gustavo	García	Expresion Of	3	10	7	9
13	12	Wilson	Reinoso	Lectura	4	9	8	9
14	13	Mateo	Geronimo	Comprension	1	4	5	8
15	14	Jhonatan	Guissoni	Escritura	3	8	6	6
16	15	Martin	Sanchez	Expresion Of	2	7	4	6
17	16	Ariel	Estamati	Lectura	1	6	9	7
18	17	Damian	Guerrero	Comprension	4	5	9	7
19	18	Jesús	Benitez	Escritura	2	6	10	5
20	19	Andrea	Castro	Expresion Of	3	7	6	4
21	20	Antonio	Reyes	Lectura	2	7	7	7
22	21	Julia	Vargas	Comprension	4	9	9	9
23	22	Belén	Michaux	Escritura	1	8	9	10
24	23	Celeste	Jimenez	Expresion Of	3	10	8	10
25	24	Edgardo	Fernandez	Lectura	5	10	6	7
26	25	Filomeno	Montemani	Comprension	2	7	6	6

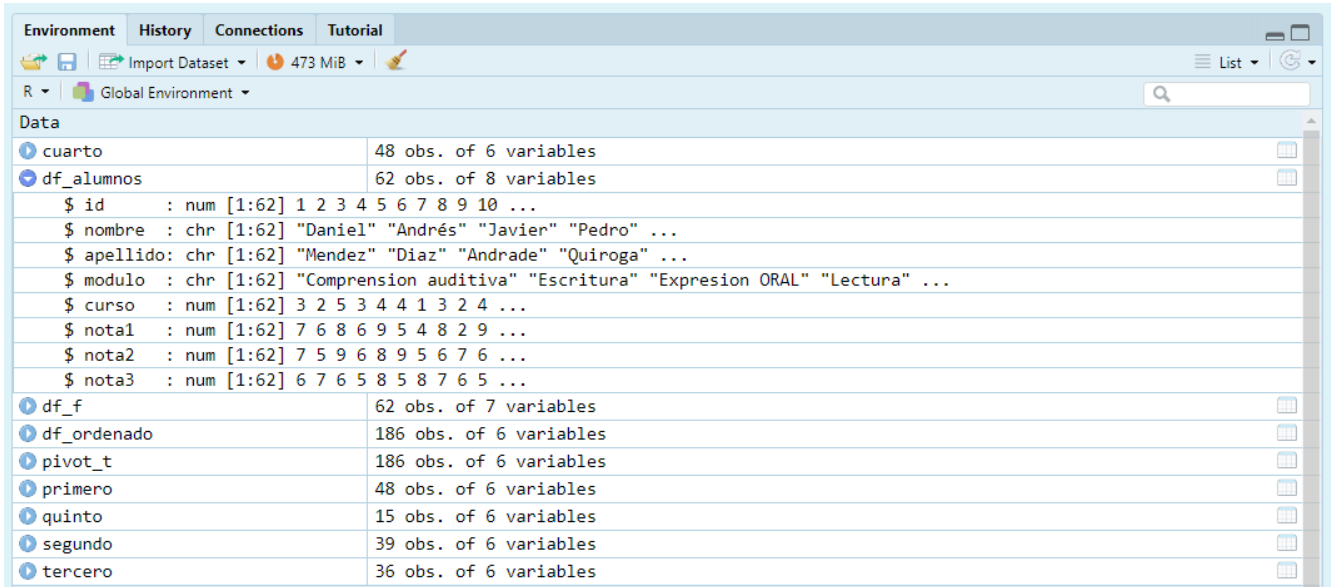
Figura 1.1 - Base de datos en Excel

## 2- Generar luego en R-ESTUDIO, un script que muestre dicho archivo (en la solapa de script) captura de pantalla.

df_alumnos x									
Filter									
	id	nombre	apellido	modulo	curso	nota1	nota2	nota3	
1	1	Daniel	Mendez	Comprension auditiva	3	7	7	6	
2	2	Andrés	Diaz	Escritura	2	6	5	7	
3	3	Javier	Andrade	Expresion ORAL	5	8	9	6	
4	4	Pedro	Quiroga	Lectura	3	6	6	5	
5	5	Jacinto	Rios	Comprension auditiva	4	9	8	8	
6	6	Alicia	Andrade	Escritura	4	5	9	5	
7	7	Mónica	De Leon	Expresion ORAL	1	4	5	8	
8	8	Bartolomé	Carrasco	Lectura	3	8	6	7	
9	9	Patricio	Hernandes	Comprension auditiva	2	2	7	6	
10	10	Rodrigo	Troncoso	Escritura	4	9	6	5	
11	11	Gustavo	García	Expresion ORAL	3	10	7	9	
12	12	Wilson	Reinoso	Lectura	4	9	8	9	
13	13	Mateo	Geronimo	Comprension auditiva	1	4	5	8	
14	14	Jhonatan	Guissoni	Escritura	3	8	6	6	
15	15	Martin	Sanchez	Expresion ORAL	2	7	4	6	
16	16	Ariel	Estamati	Lectura	1	6	9	7	
17	17	Damian	Guerrero	Comprension auditiva	4	5	9	7	
18	18	Jesús	Benitez	Escritura	2	6	10	5	
19	19	Andrea	Castro	Expresion ORAL	3	7	6	4	
20	20	Antonio	Reyes	Lectura	2	7	7	7	
21	21	Julia	Vargas	Comprension auditiva	4	9	9	9	
22	22	Belén	Michaux	Escritura	1	8	9	10	
23	23	Celeste	Jimenez	Expresion ORAL	3	10	8	10	
24	24	Edgardo	Fernandez	Lectura	5	10	6	7	

Figura 2.1 - Data Frame en R

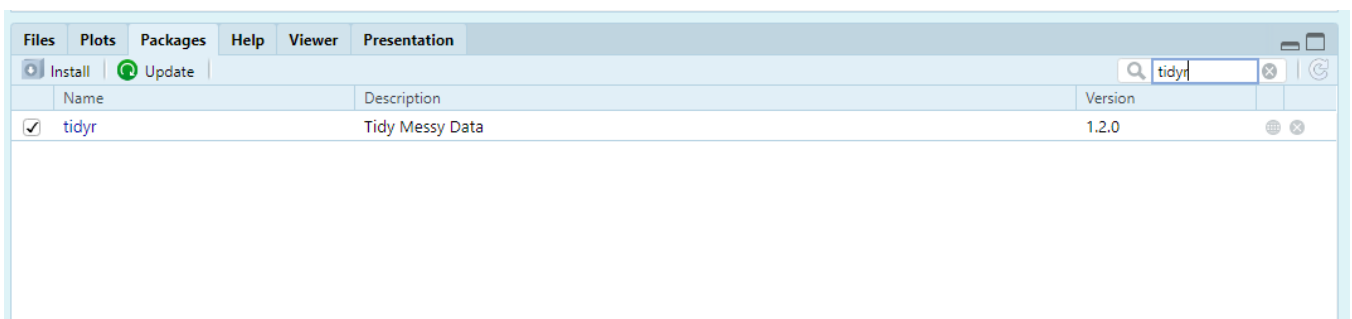
### 3 -En la solapa de variables y observaciones mostrar las variables del script (captura de pantalla)



Object	Details
cuarto	48 obs. of 6 variables
df_alumnos	62 obs. of 8 variables
\$ id	num [1:62] 1 2 3 4 5 6 7 8 9 10 ...
\$ nombre	chr [1:62] "Daniel" "Andrés" "Javier" "Pedro" ...
\$ apellido	chr [1:62] "Mendez" "Diaz" "Andrade" "Quiroga" ...
\$ modulo	chr [1:62] "Comprension auditiva" "Escritura" "Expresion ORAL" "Lectura" ...
\$ curso	num [1:62] 3 2 5 3 4 4 1 3 2 4 ...
\$ nota1	num [1:62] 7 6 8 6 9 5 4 8 2 9 ...
\$ nota2	num [1:62] 7 5 9 6 8 9 5 6 7 6 ...
\$ nota3	num [1:62] 6 7 6 5 8 5 8 7 6 5 ...
df_f	62 obs. of 7 variables
df_ordenado	186 obs. of 6 variables
pivot_t	186 obs. of 6 variables
primero	48 obs. of 6 variables
quinto	15 obs. of 6 variables
segundo	39 obs. of 6 variables
tercero	36 obs. of 6 variables

Figura 3.1 - Variables en R

### 4 -Instalar en la solapa inferior derecha las librerías tidyr, readxl, ggplot2, dplyr.



Package	Description	Version
<input checked="" type="checkbox"/> tidyr	Tidy Messy Data	1.2.0

Figura 4.1 - Tidyr instalado en RStudio

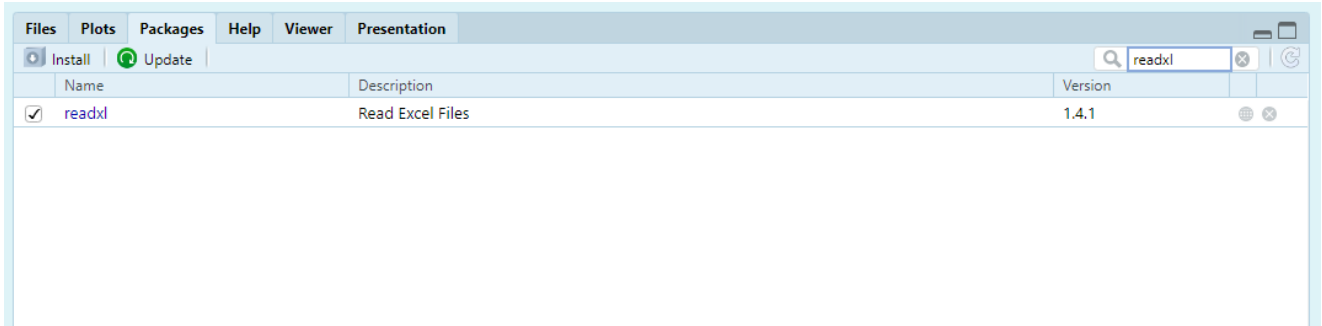


Figura 4.2 - Readxl instalado en RStudio

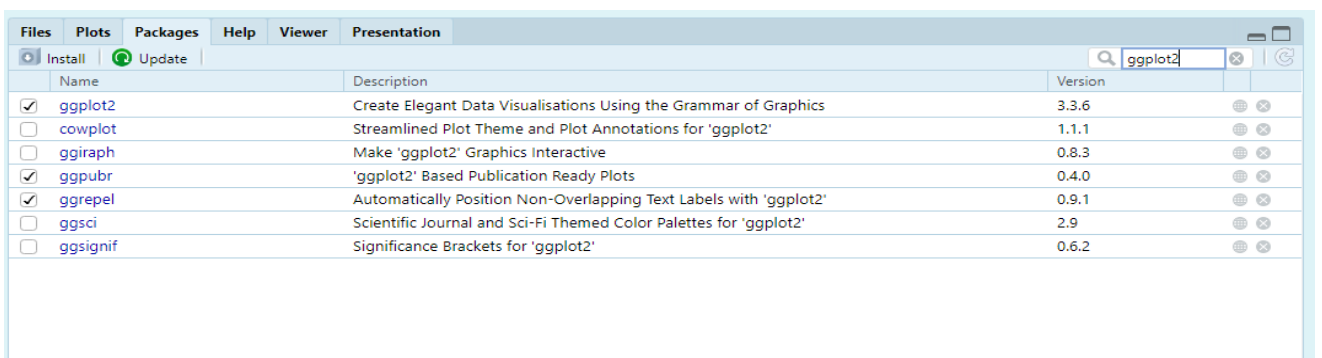


Figura 4.3 - GGplot instalado en RStudio

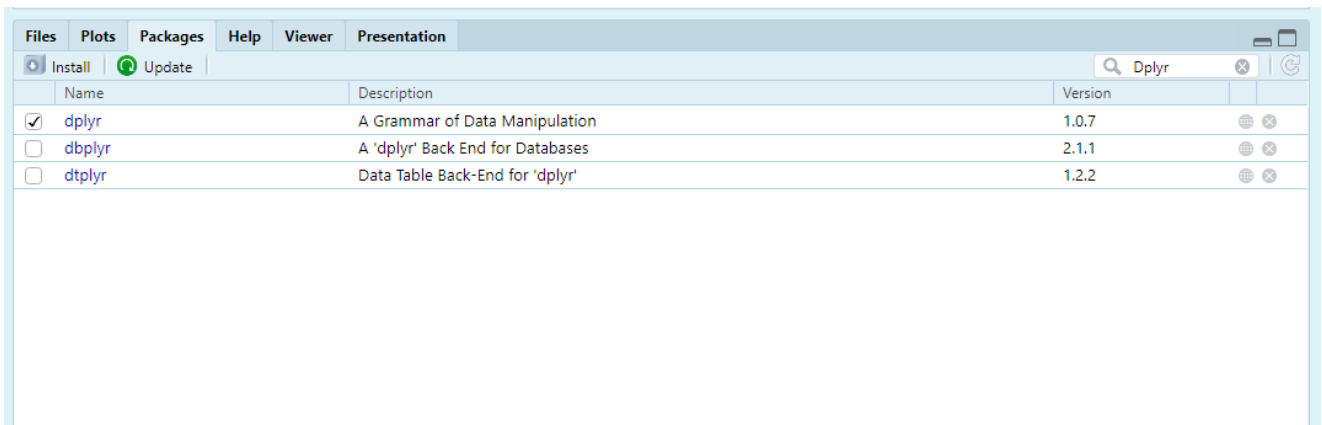


Figura 4.4 - Dplyr instalado en RStudio

**5- Con dos de las columnas de la tabla seleccionada generar un gráfico usando ggplot2 y mostrar gráficos de puntos, barra, lineal y fraccionado. (captura de pantalla).**

Gráfico de puntos: columna de notas y sus frecuencias relativas.

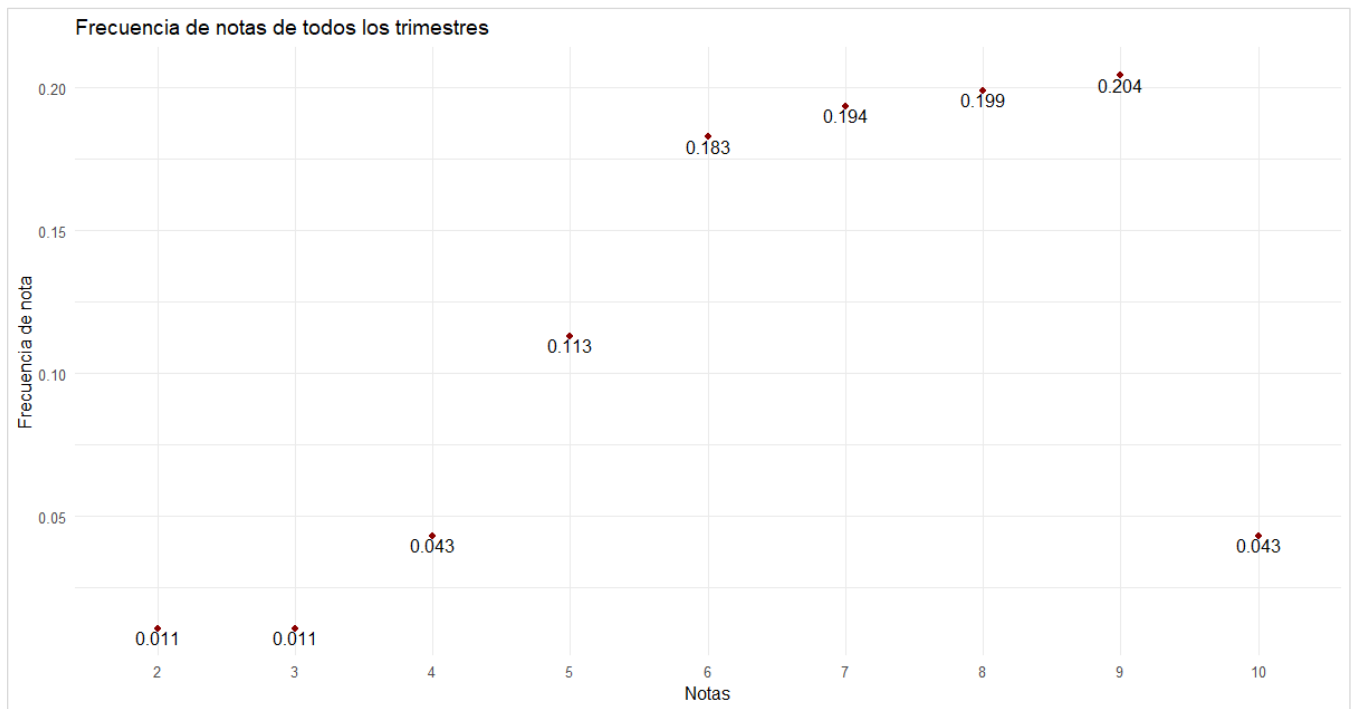


Figura 5.1 - Scatter plot

Gráfico de barras: Promedios agrupados por curso de cada módulo.

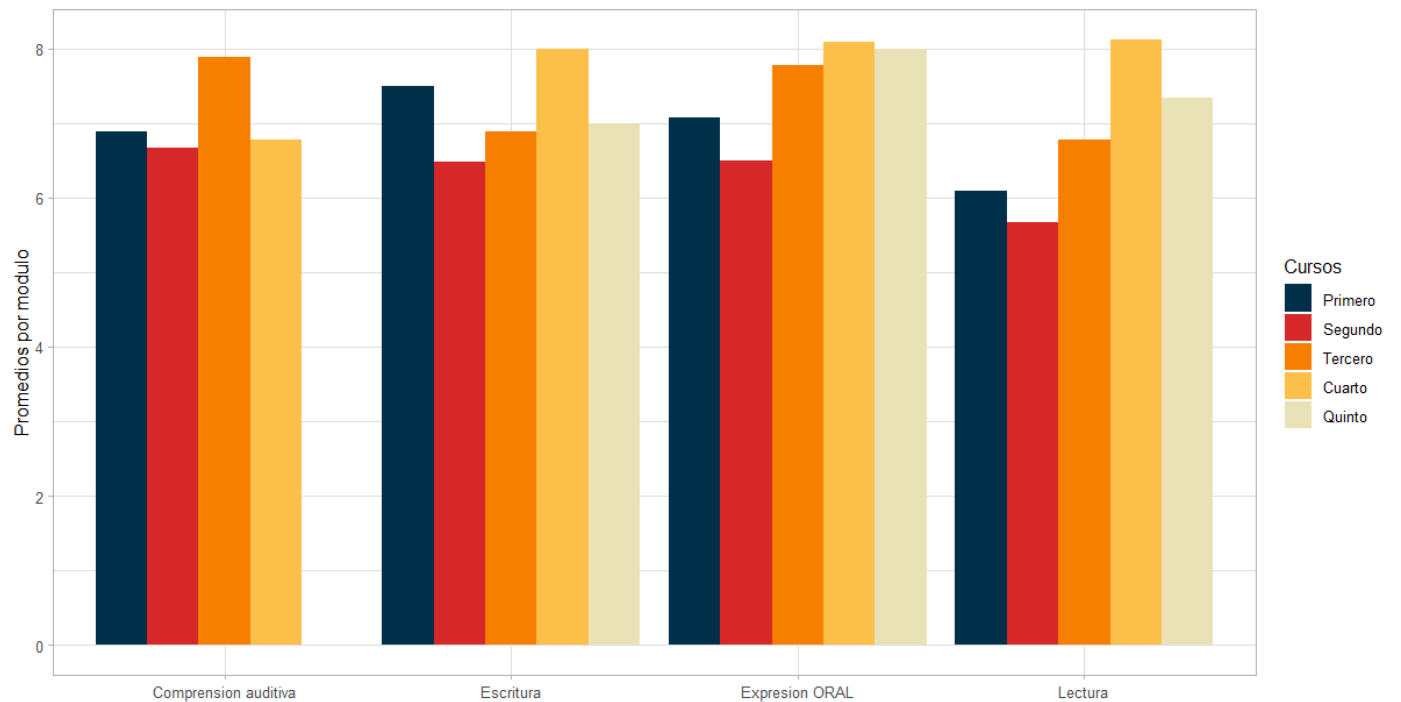


Figura 5.2 – Barplot

Gráficos de líneas: Notas trimestrales de cada alumno agrupados por curso y modulo

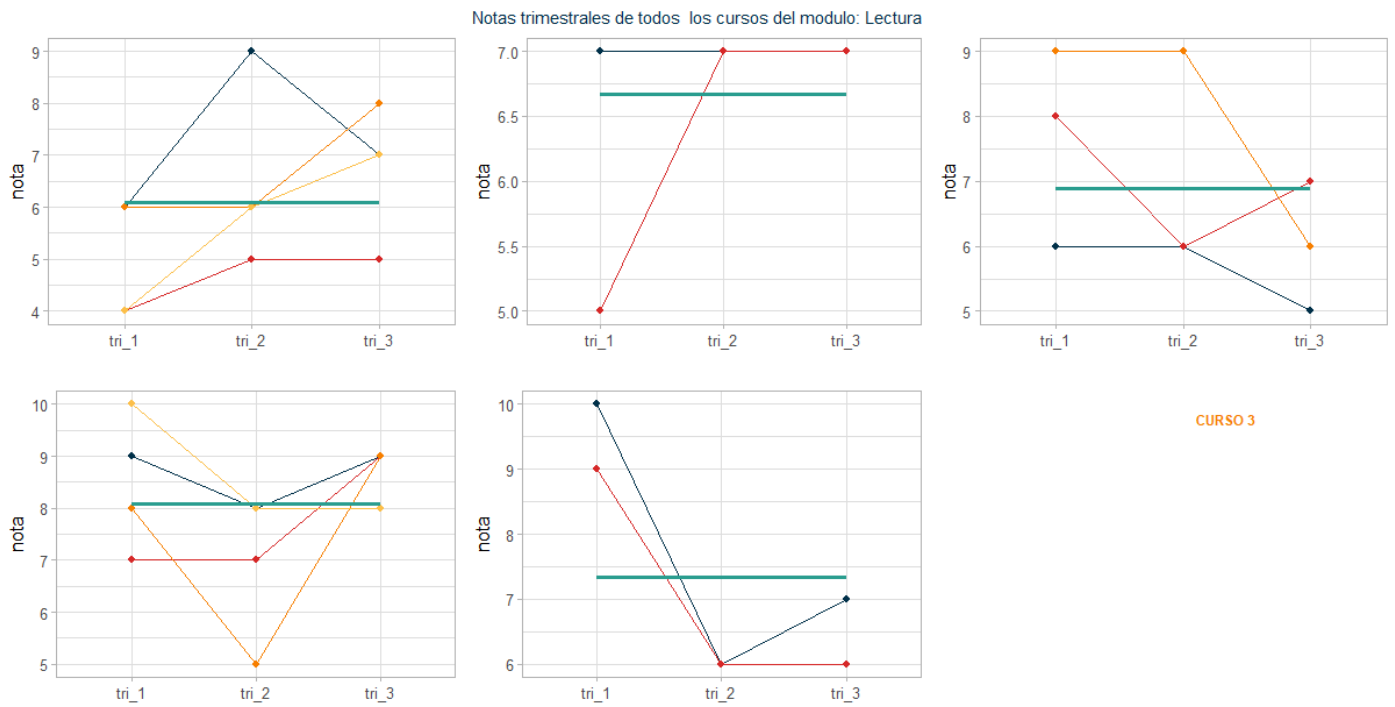


Figura 5.3 - Line plot

Gráfico de fraccionado: Cantidad de alumnos por curso

Alumnos por curso

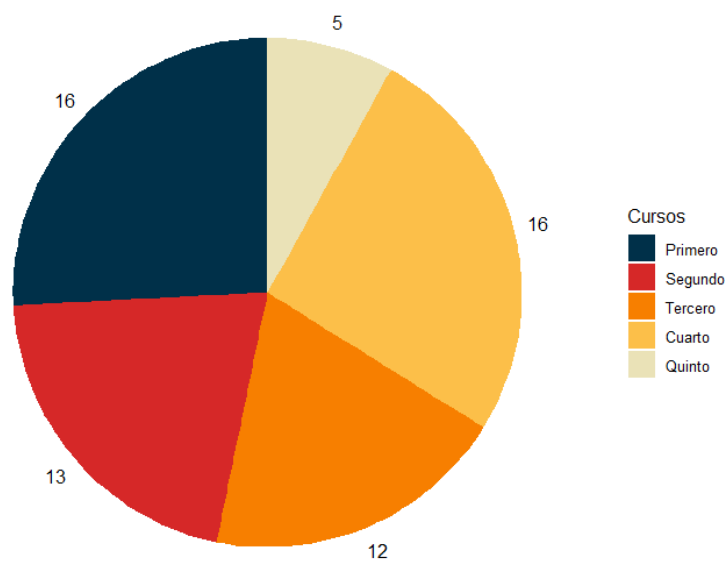
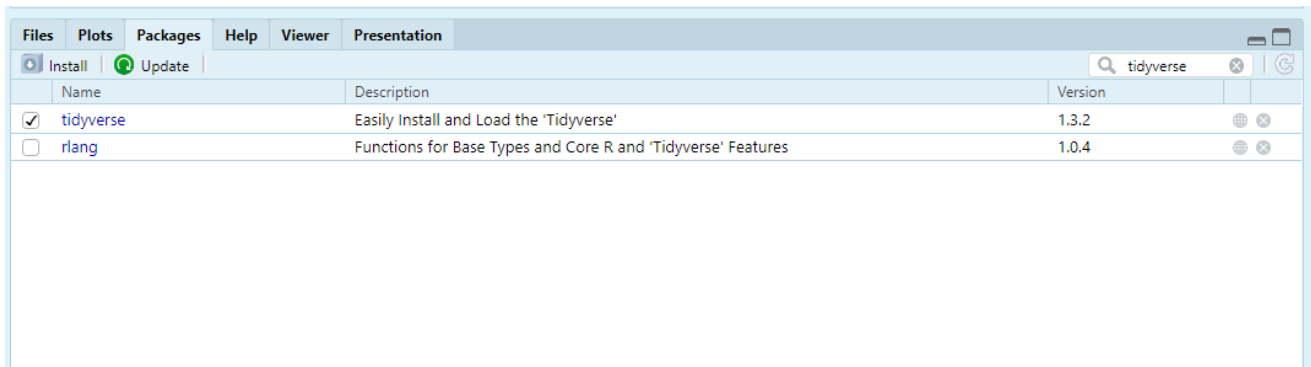


Figura 5.4 - Pie plot



## 6 - Instalar una nueva librería que se llama tidyverse (captura de pantalla)

Figura 6.1 - Tidyverse instalado en RStudio



## 7- Usando esta librería calcular la media, mediana y desviación estándar. (en la solapa de consola) captura de pantalla del uso de la librería) y mostrar resultados.

### # MEDIDAS DE CONCENTRACIÓN Y DISPERSIÓN

```
notas <- df_ordenado$nota
notas_ordenadas <- sort(notas)

media <- mean(notas)
mediana <- median(notas_ordenadas)
desvio <- sd(notas)

media
mediana
desvio
```

Figura 7.1 -  
Script R – Código

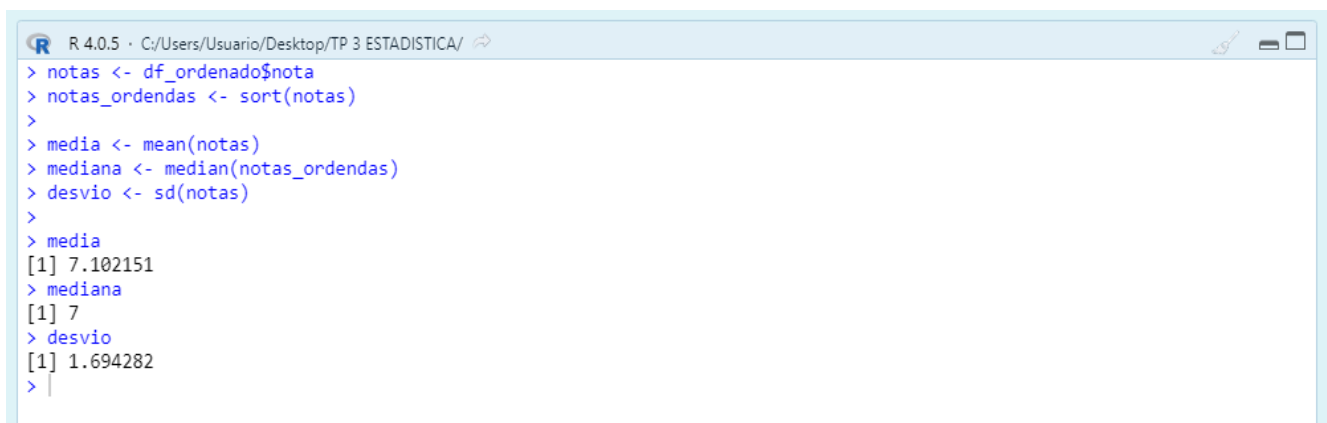


Figura 7.2 - Consola RStudio medidas calculadas