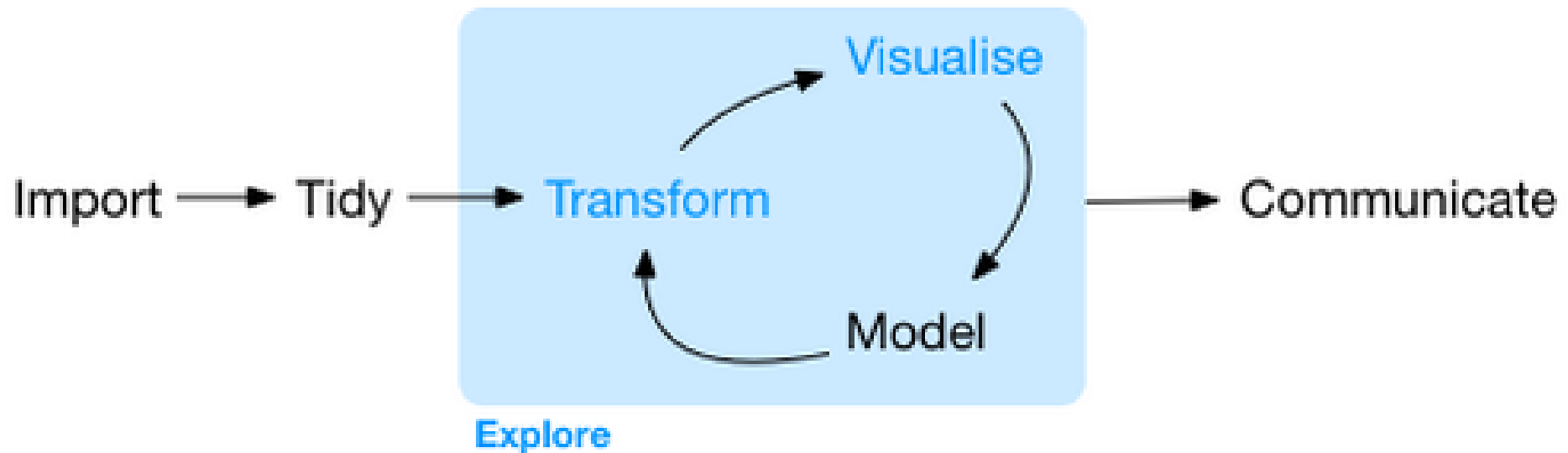


Sección 0

Análisis, interpretación y Visualización de datos

Exploración de datos



¿Cómo importar datos en R?

CSV, EXCEL, SPSS,
SAS y STATA

File > Import Dataset

csv, R-base

```
read.csv()
```

Librería readr

```
read_csv()
```

Datos tabulares utilizando readr

```
# Instalar el paquete  
install.packages («readr»)  
  
# Cargar el paquete  
library («readr»)
```

- Delimitados por caracteres:

```
read_delim(), read_csv(), read_tsv()
```

- De anchura fija:

```
read_fwf() y read_table()
```

CSV

- `read_csv()`: para leer archivos con coma (“,”) como separador
- `read_csv2()`: para leer archivos con punto y coma (“;”) como separador
- `read_tsv()`: para leer archivos con tabulador (“\t”) como separador
- `read_delim(,sep = ' | ')`: para leer archivos con separador distintos como puede ser el símbolo ‘|’

Función write_csv()

- Exportar dataframes

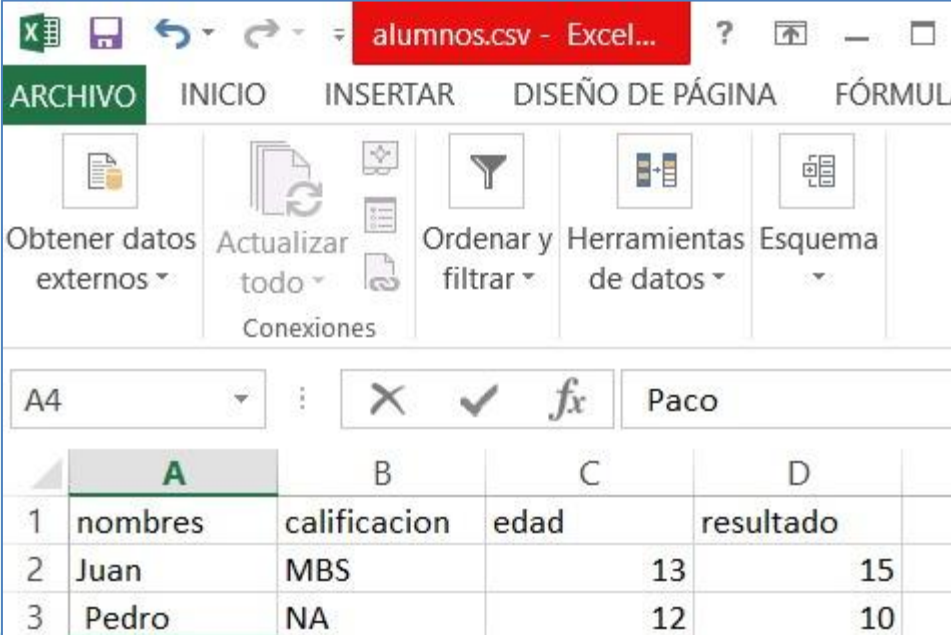
```
write_csv(x, file, col_names)
```

- x: data frame
- file: directorio donde guardar archivo
- col_names: Nombres de las variables (por defecto se encuentran en la primera fila del archivo, col_names = TRUE).

Ejemplo write_csv()

```
alumnos<-data.frame(nombres=c("Juan"," Pedro","  
Paco","Ana"," Florencia","Maria "," Cecilia", "Fernanda  
"),calificacion=c("MBS",NA,"B","B","MBS","MBS","D",NA),edad=  
c(13,12,NA,10,13,10,12,NA),resultado=c(15,10,9,2,11,15,5,8))
```

```
write_csv(alumnos,file =  
"C:/Users/AC/Desktop/alumnos.csv")
```



	A	B	C	D
1	nombres	calificacion	edad	resultado
2	Juan	MBS	13	15
3	Pedro	NA	12	10

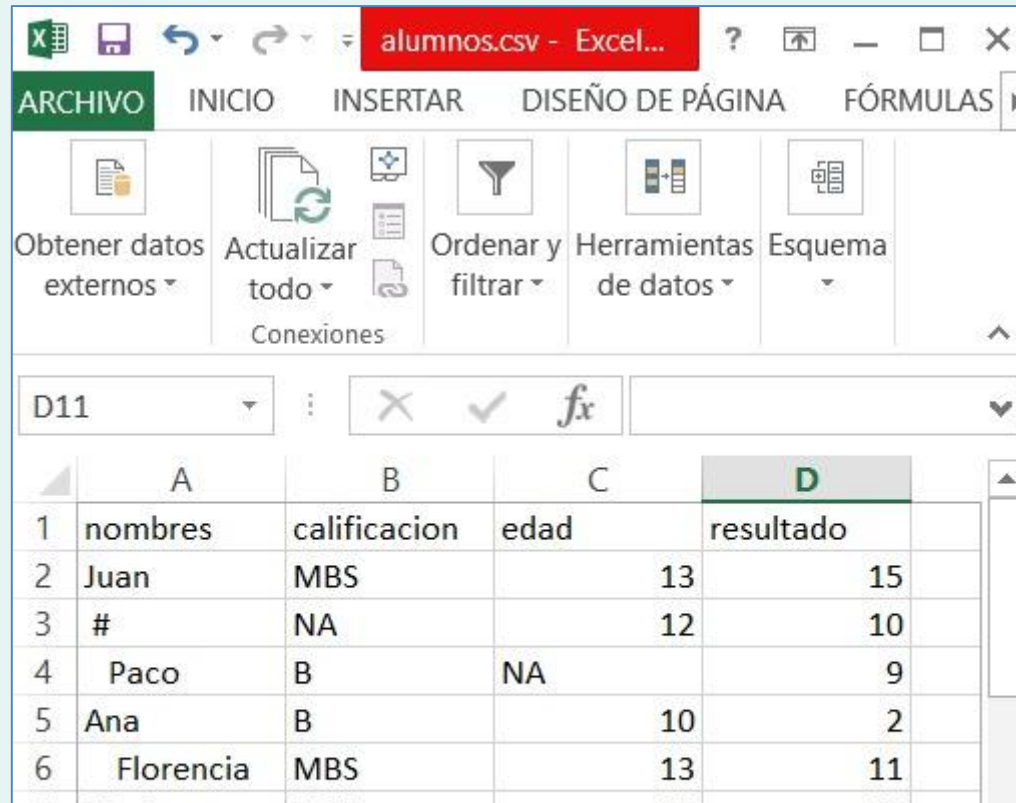
Función read_csv()

- Importar dataframes

```
read_csv(file,col_names,skip,na)
```

- file: dirección donde buscar archivo url o path
- col_names: Nombres de las variables
- skip: Número de fila desde donde comenzar a importar (por defecto, se importan todas)

Ejemplo read_csv()



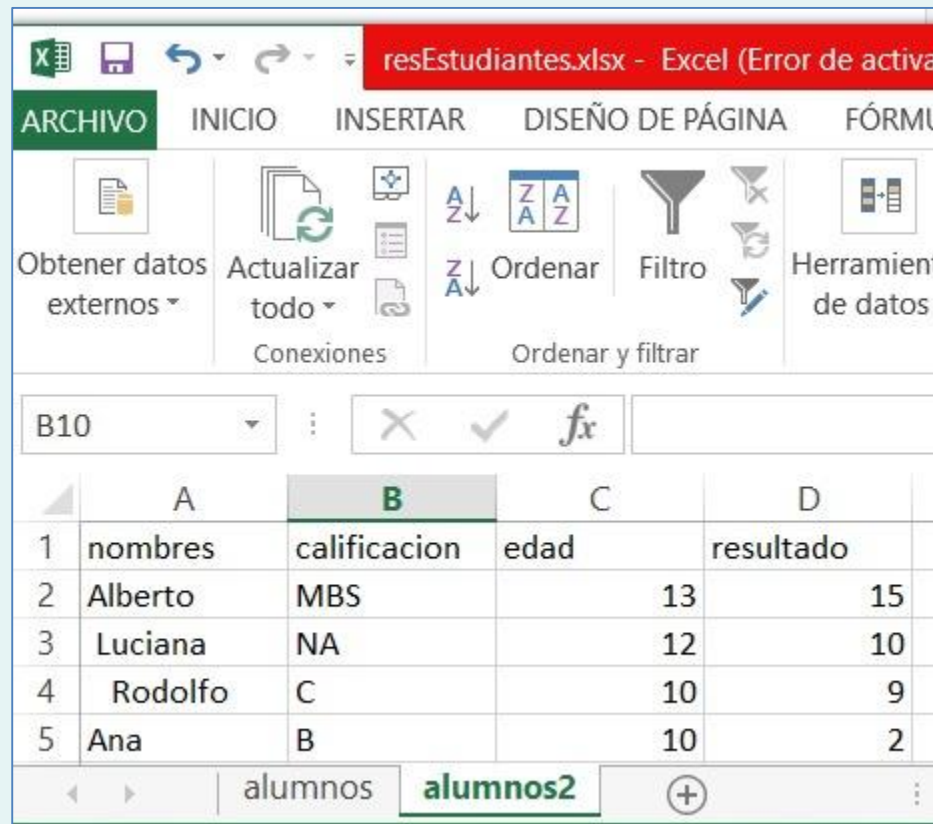
The screenshot shows the Microsoft Excel interface with the file 'alumnos.csv' open. The ribbon includes 'ARCHIVO', 'INICIO', 'INSERTAR', 'DISEÑO DE PÁGINA', and 'FÓRMULAS'. The 'FÓRMULAS' ribbon is active, showing options like 'Obtener datos externos', 'Actualizar todo', 'Ordenar y filtrar', 'Herramientas de datos', and 'Esquema'. The active cell is D11. The data table is as follows:

	A	B	C	D
1	nombres	calificacion	edad	resultado
2	Juan	MBS	13	15
3	#	NA	12	10
4	Paco	B	NA	9
5	Ana	B	10	2
6	Florencia	MBS	13	11

```
> read.csv(file = "C:/Users/AC/Desktop/alumnos.csv", na="#", strip.white = TRUE, sep=";")
  nombres calificacion edad resultado
1   Juan          MBS   13         15
2  <NA>          NA   12         10
3   Paco           B   NA          9
4   Ana           B   10          2
5 Florencia       MBS   13         11
6   Maria       MBS   10         15
7  <NA>           D   12          5
8 Fernanda       NA   NA          8
```

Otras funciones de lectura - ejemplo Excel

```
read_excel("C:/Users/AC/Desktop/resEstudiantes.xlsx",  
sheet = "alumnos2")
```



	A	B	C	D
1	nombres	calificacion	edad	resultado
2	Alberto	MBS	13	15
3	Luciana	NA	12	10
4	Rodolfo	C	10	9
5	Ana	B	10	2

Otras funciones de lectura - ejemplo Excel

```
[[1]]
# A tibble: 8 x 4
  nombres calificacion edad resultado
  <chr>    <chr>      <chr>    <dbl>
1 Juan    MBS          13         15
2 #       NA          12         10
3 Paco    B           NA          9
4 Ana     B           10          2
5 Florencia MBS       13         11
6 Maria   MBS       10         15
7 #       D           12          5
8 Fernanda NA          NA          8
```

```
[[2]]
# A tibble: 5 x 4
  nombres calificacion edad resultado
  <chr>    <chr>      <dbl>    <dbl>
1 Alberto MBS          13         15
2 Luciana NA          12         10
3 Rodolfo C           10          9
4 Ana     B           NA          NA
5 Florencia D           NA          NA
```

```
lapply(excel_sheets(
"C:/Users/AC/Desktop/res
Estudiantes.xlsx"),
read_excel,
path =
"C:/Users/AC/Desktop/res
Estudiantes.xlsx"))
```

```
> alumnos1<-resultados[[1]]
> str(alumnos1)
tibble [8 x 4] (S3: tbl_df/tbl/data.frame)
 $ nombres      : chr [1:8] "Juan" "# " "Paco"
 $ calificacion: chr [1:8] "MBS" "NA" "B"
 $ edad         : chr [1:8] "13" "12" "NA"
 $ resultado    : num [1:8] 15 10 9 2 11 15
```

Tipo de datos

- **Categóricos** (Sexo [Femenino, masculino], Regiones [Este, Oeste, Norte,...])
 - Ordinales por ej.:
 - Número de cajas de cigarros que una persona fuma en el día:
0,1,2,etc...
 - Variable picante: fuerte medio y bajo
 - No ordinales
- **Numéricos** (Población, índice de asesinatos, alturas,...)
 - Discretos or ej.:
 - Cantidad de cigarros que una persona fuma en el día:
0,1,2,3,...,36
 - Continuos por ej.:
 - Medida de altura de personas