

Ejemplos básicos de manipulación de datos en *Mathematica*

Importando un archivo de Excel

Primero un ejemplo muy sencillo, leyendo datos de una hoja de cálculo. La siguiente imagen muestra el contenido del archivo `fisicamate.xlsx` que tengo en mi computadora:



The screenshot shows the Microsoft Excel interface with the file `fisicamate.xlsx` open. The spreadsheet contains the following data:

Clave	Gpo	Pob	Cupo	Cap	CRN	Materia	Profesor	Día	Horario	Salón
DEPARTAMENTO DE FISICA Y MATEMATICAS										
Semestre Enero - Mayo 2011										
Grupos por profesor										
MA2005	1	21	30	42	15286	Matemáticas para el diseño	Arturo Preciado R	M V	11:30 a 13:00	A5 104
MA2006	1	30	30	30	15287	Sistemas aleatorios	Arturo Preciado R	M V	10:00 a 11:30	CDT1 04L02
MA1012	6	29	31	36	15257	Matemáticas I	Carlos Daniel Prado P	M V	7:00 a 8:30	A5 210
MA1014	3	21	25	35	15266	Matemáticas II	Carlos Daniel Prado P	LM	7:00 a 8:30	A4 111
MA1014	55	20	0	35	15269	Matemáticas II	Carlos Daniel Prado P	M V Mi	8:30 a 10:00 9:00 a 11:00	A4 202
MA1014	9	34	35	40	15273	Matemáticas II	Carlos Daniel Prado P	LM	11:30 a 13:00	A4 312
MA2002	1	35	35	42	15280	Matemáticas para ing III	Carlos Daniel Prado P	M V	10:00 a 11:30	A5 303
F1002	6	22	28	35	14964	Física I	Elena G Cabra I V	LM	7:00 a 8:30	A5 110
F1002	2	30	28	35	14960	Física I	Elena G Cabra I V	M V	11:30 a 13:00	A5 304

El siguiente comando de *Mathematica* muestra los documentos de Excel en la carpeta “Mis documentos” de mi computadora:

```
FileNames["*.xlsx"]  
  
{ejemplo.xlsx, fisicamate.xlsx}
```

El siguiente comando carga en esta sesión de *Mathematica* el contenido del documento `fisicamate.xlsx`, y ese contenido queda guardado en la variable que yo le llamé “libro”

```
libro = Import["fisicamate.xlsx"];
```

El siguiente comando muestra el séptimo renglón de la primera hoja del libro de Excel que fue importado:

```
Part[libro, 1, 7]  
  
{CD1000, 1., 30., 14721., Estadística administrativa I,  
Manuel González S, M V, 7:00 a 8:30, A4 207, }
```

El siguiente comando muestra el quinto campo del séptimo renglón de la primera hoja del libro de Excel que fue importado:

```
Part[libro, 1, 7, 5]  
  
Estadística administrativa I
```

Continúa en la siguiente página

Se cuenta cuantas veces se repite cada valor del quinto campo, en este ejemplo eso corresponde a cuantos grupos hay de cada materia:

```

agrupadas =
  Tally[Part[libro, 1, 7 ;; All], Function[{r1, r2}, Part[r1, 5] == Part[r2, 5]]];
contadas = Table[{Part[agrupadas, n, 1, 5], Part[agrupadas, n, 2]},
  {n, Length[agrupadas]}]

{{ Estadística administrativa I, 3}, { Estadística administrativa II, 3},
 { Estadística multivariante, 2}, { Estadística ciencias comportam, 1},
 { Introducción a la física, 2}, { Física I, 8}, { Física II, 6},
 { Electricidad y magnetismo, 5}, { Métodos numéricos en ing, 4},
 { Introducción a las matemáticas, 3}, { Probabilidad y estadística, 4},
 { Matemáticas, 2}, { Estadística investig cienc soc, 3},
 { Matemáticas para el diseño, 1}, { Matemáticas para el diseño I, 2},
 { Matemáticas I, 11}, { Matemáticas II, 13},
 { Ecuaciones diferenciales, 6}, { Matemáticas para ing III, 3},
 { Estadística II, 1}, { Matemáticas para el diseño II, 1},
 { Sistemas aleatorios, 1}, { Matemáticas para economía II, 1},
 { Matemáticas avanzadas, 2}, { Matemáticas discretas, 2}}

```

Continúa en la siguiente página

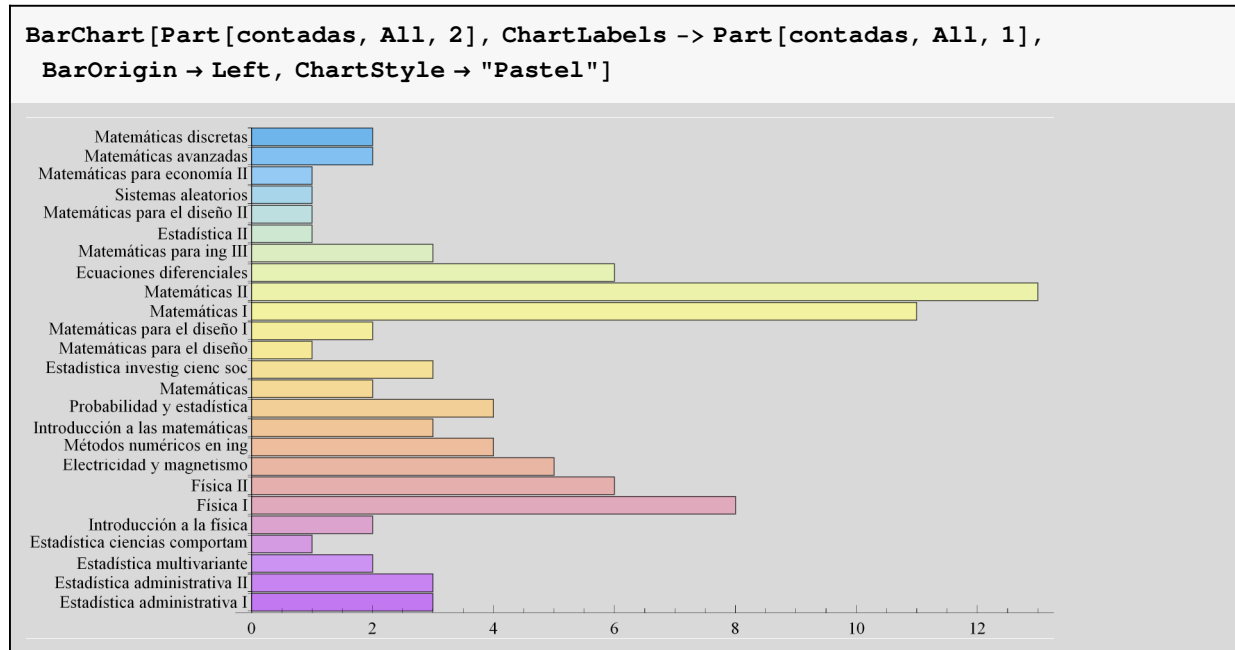
La misma información, mostrada en forma tabular

Grid[contadas, Dividers → All]

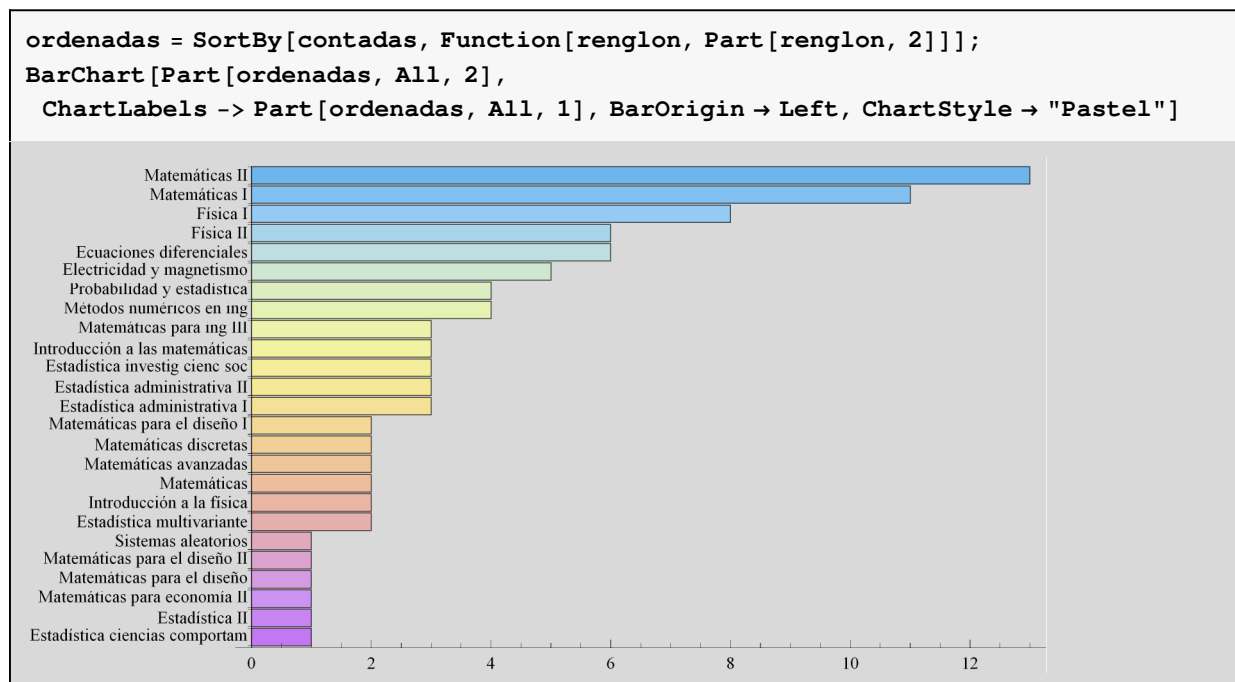
Estadística administrativa I	3
Estadística administrativa II	3
Estadística multivariante	2
Estadística ciencias comportam	1
Introducción a la física	2
Física I	8
Física II	6
Electricidad y magnetismo	5
Métodos numéricos en ing	4
Introducción a las matemáticas	3
Probabilidad y estadística	4
Matemáticas	2
Estadística investig cienc soc	3
Matemáticas para el diseño	1
Matemáticas para el diseño I	2
Matemáticas I	11
Matemáticas II	13
Ecuaciones diferenciales	6
Matemáticas para ing III	3
Estadística II	1
Matemáticas para el diseño II	1
Sistemas aleatorios	1
Matemáticas para economía II	1
Matemáticas avanzadas	2
Matemáticas discretas	2

Continúa en la siguiente página

La misma información, mostrada en forma gráfica



La misma información gráfica, ordenada de menor a mayor

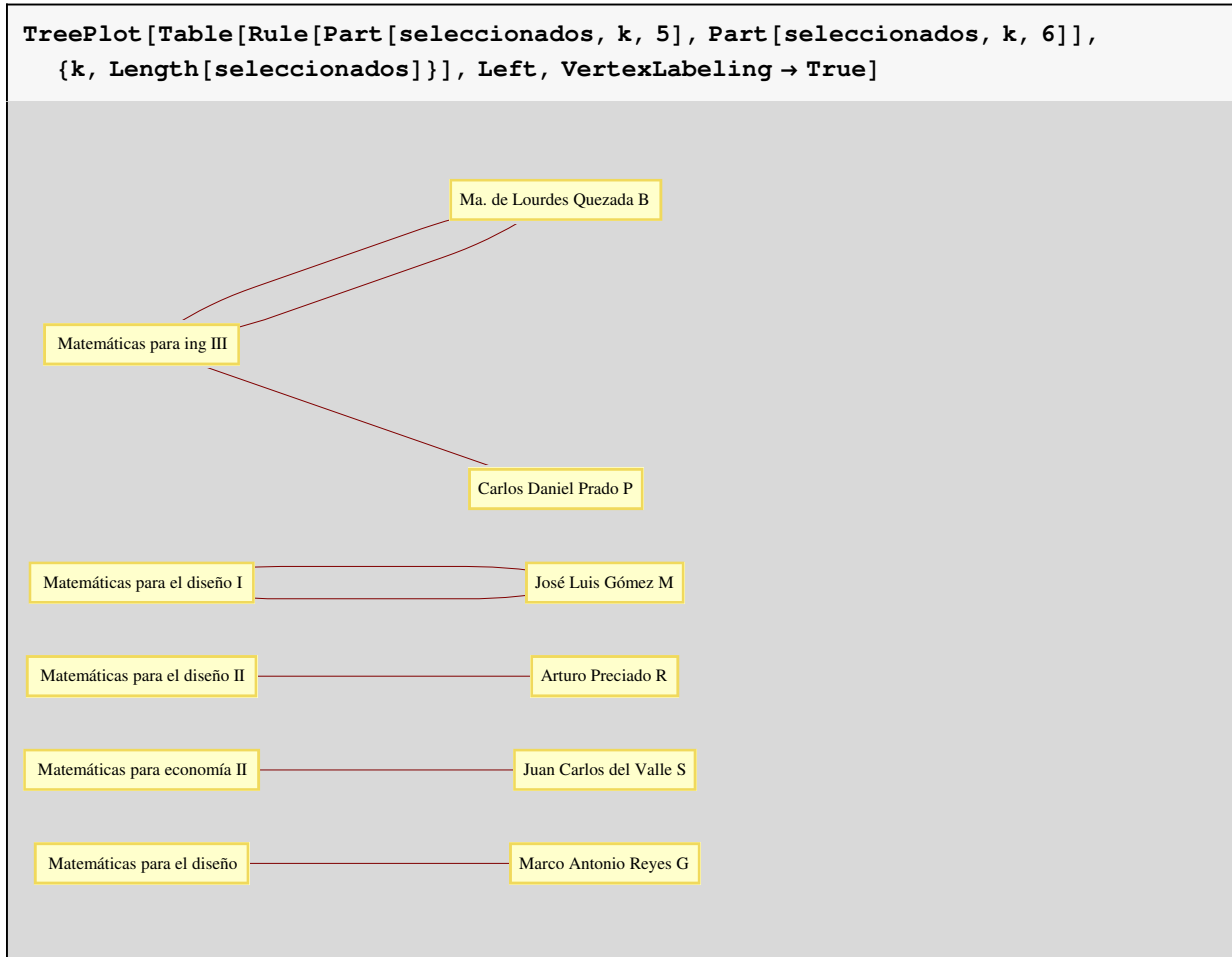


El siguiente comando selecciona, de todos los renglones de la primera hoja, sólo aquellos cuyo quinto campo incluye la palabra “para”

```
seleccionados = Select[Part[libro, 1],
  Function[renglon, StringMatchQ[Part[renglon, 5], ___ ~~ "para" ~~ ___]]]
{{MA1009, 2., 31., 15246., Matemáticas para el diseño,
  Marco Antonio Reyes G, M V , 11:30 a 13:00, A6 308, },
 {MA1011, 1., 21., 15248., Matemáticas para el diseño I,
  José Luis Gómez M, L M , 13:00 a 14:30, A4 310, },
 {MA1011, 2., 22., 15249., Matemáticas para el diseño I,
  José Luis Gómez M, M V , 14:30 a 16:00, A4 308, },
 {MA2002, 1., 35., 15280., Matemáticas para ing III,
  Carlos Daniel Prado P, M V , 10:00 a 11:30, A5 303, },
 {MA2002, 4., 23., 15283., Matemáticas para ing III,
  Ma. de Lourdes Quezada B, L M , 7:00 a 8:30, A4 201, },
 {MA2002, 5., 23., 15284., Matemáticas para ing III,
  Ma. de Lourdes Quezada B, L M , 8:30 a 10:00, A5 208, },
 {MA2005, 1., 21., 15286., Matemáticas para el diseño II,
  Arturo Preciado R, M V , 11:30 a 13:00, A5 104, },
 {MA3001, 1., 8., 15288., Matemáticas para economía II,
  Juan Carlos del Valle S, M V , 8:30 a 10:00, A4 111, }}
```

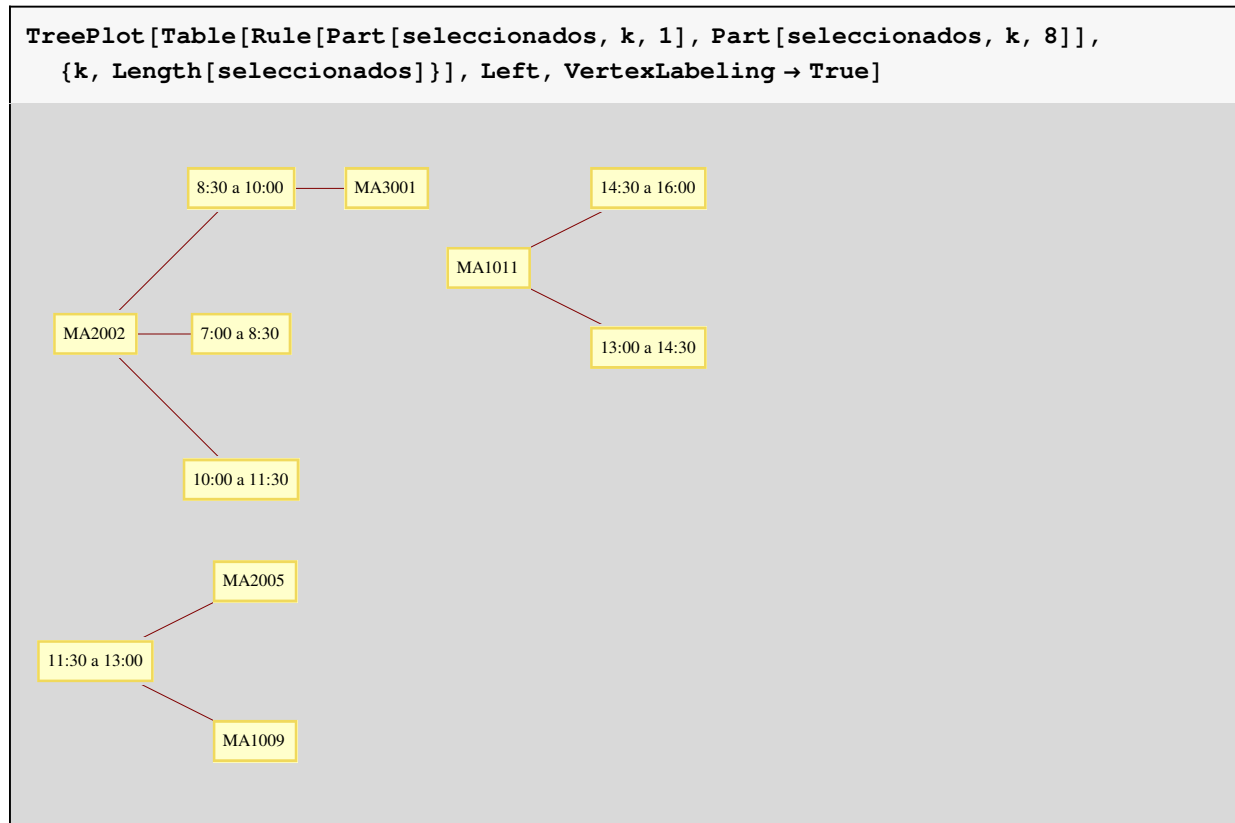
Continúa en la siguiente página

El siguiente comando de *Mathematica* produce una representación gráfica de las conexiones entre profesores y materias, para los renglones que fueron seleccionados en el comando anterior.



Continúa en la siguiente página

Para los mismos renglones que fueron seleccionados, ahora se muestran las conexiones entre horarios y claves de materia



Exportamos a un nuevo archivo de Excel los renglones que fueron seleccionados más arriba

```
Export["seleccion.xlsx", seleccionados]
```

seleccion.xlsx

Conectividad a bases de datos en servidor SQL

Los comandos anteriores trabajaron sobre datos que se importaron de una hoja de cálculo. *Mathematica* puede importar datos de diversos tipos de bases de datos. Por otro lado, los datos podrían estar en un servidor SQL. *Mathematica* tiene comandos que le permiten conectarse al servidor SQL, leer registros de la base de datos para hacer cálculos con ellos, como se muestra en la siguiente liga:

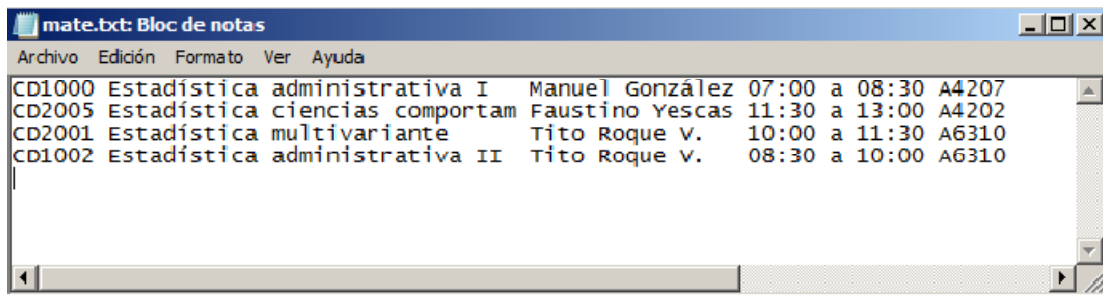
<http://reference.wolfram.com/mathematica/DatabaseLink/tutorial/DatabaseConnections.html>

Cuando una base de datos SQL ha sido leída en *Mathematica*, se pueden

hacer cálculos y operaciones sobre los registros exactamente de la misma forma como se hacen operaciones sobre los renglones de una hoja de cálculo importada.

Importando un archivo de texto

La siguiente imagen muestra el contenido del archivo mate.txt que tengo en mi computadora:



El siguiente comando de *Mathematica* muestra los documentos de texto en la carpeta “Mis documentos” de mi computadora:

```
FileNames["*.txt"]
```

```
{mate.txt}
```

El siguiente comando carga en esta sesión de *Mathematica* el contenido del documento fisicamate.xlsx, y ese contenido queda guardado en la variable que yo le llamé “texto”

```
texto = ReadList["mate.txt", String, RecordLists -> True]
```

```
{{CD1000 Estadística administrativa I Manuel González 07:00 a 08:30 A4207},
 {CD2005 Estadística ciencias comportam Faustino Yescas 11:30 a 13:00 A4202 },
 {CD2001 Estadística multivariante Tito Roque V. 10:00 a 11:30 A6310},
 {CD1002 Estadística administrativa II Tito Roque V. 08:30 a 10:00 A6310}}
```

El siguiente comando muestra el tercer renglón del documento de texto

```
Part[texto, 3]
```

```
{CD2001 Estadística multivariante Tito Roque V. 10:00 a 11:30 A6310}
```

Sin embargo en este momento tenemos el siguiente problema: el

documento de texto es “plano” y no podemos, por el momento, obtener el segundo campo del tercer renglón como se hizo antes con el archivo de Excel. *Mathematica* responde con un mensaje de error:

```
Part[texto, 3, 2]
```

```
Part::partw: Part 2 of {CD2001 Estadística multivariante Tito Roque V. 10:00 a 11:30 A6310} does not exist. >>
```

```
{ {CD1000 Estadística administrativa I Manuel González 07:00 a 08:30 A4207},
  {CD2005 Estadística ciencias
    comportam Faustino Yescas 11:30 a 13:00 A4202 },
  {CD2001 Estadística multivariante Tito Roque V. 10:00 a 11:30 A6310},
  {CD1002 Estadística administrativa
    II Tito Roque V. 08:30 a 10:00 A6310} }[[3, 2]]
```

En este ejemplo la solución es muy sencilla, porque cada campo cierta cantidad de letras, entonces puedo generar una lista con la misma estructura que si hubiera sido importada desde un archivo de Excel. Esa lista es guardada en la variable “campos”:

```
campos = Table[
  StringTake[Part[texto, j, 1], {{1, 6}, {7, 38}, {39, 54}, {55, 68}, {69, 73}}],
  {j, Length[texto]}]
```

```
{ {CD1000, Estadística administrativa I ,
  Manuel González , 07:00 a 08:30 , A4207}, {CD2005,
  Estadística ciencias comportam , Faustino Yescas , 11:30 a 13:00 , A4202},
  {CD2001, Estadística multivariante , Tito Roque V. ,
  10:00 a 11:30 , A6310}, {CD1002, Estadística administrativa II ,
  Tito Roque V. , 08:30 a 10:00 , A6310} }
```

Ahora sí, podemos obtener el segundo campo del tercer renglón con la mismos comandos que se hizo más arriba para el ejemplo que fue importado de Excel

```
Part[campos, 3, 2]
```

```
Estadística multivariante
```

El siguiente comando selecciona, de todos los renglones del texto, sólo aquellos cuyo segundo campo incluye la palabra “administrativa”

```
selec = Select[campos, Function[renglon,
  StringMatchQ[Part[renglon, 2], ___ ~~ "administrativa" ~~ ___]]]
{{CD1000, Estadística administrativa I ,
  Manuel González , 07:00 a 08:30 , A4207}, {CD1002,
  Estadística administrativa II , Tito Roque V. , 08:30 a 10:00 , A6310}}
```

Los renglones que fueron seleccionados se muestran en una “Grid” con divisiones:

```
Grid[selec, Dividers → All]
```

CD1000	Estadística administrativa I	Manuel González	07:00 a 08:30	A4207
CD1002	Estadística administrativa II	Tito Roque V.	08:30 a 10:00	A6310

Juntando datos con un campo codificado de forma distinta

Aquí se definen algunos datos donde “F” significa femenino y “M” masculino

```
data1 = {
  {"Alicia", "F"},
  {"Pedro", "M"},
  {"Cecilia", "F"}
}
{{Alicia, F}, {Pedro, M}, {Cecilia, F}}
```

Aquí se definen otros datos donde “01” significa femenino y “02” masculino

```
data2 = {
  {"Raúl", "02"},
  {"Luis", "02"},
  {"Ana", "01"}
}
{{Raúl, 02}, {Luis, 02}, {Ana, 01}}
```

En el segundo conjunto de datos se reemplaza “01” con “F” y “02” con “M”

```
data2corregido = ReplaceAll[data2, {"02" → "M", "01" → "F"}]
```

```
{{Raúl, M}, {Luis, M}, {Ana, F}}
```

Se juntan los datos

```
datos = Join[data1, data2corregido]
```

```
{{Alicia, F}, {Pedro, M}, {Cecilia, F}, {Raúl, M}, {Luis, M}, {Ana, F}}
```

Se muestran en formato de Grid

```
Grid[datos, Dividers → All]
```

Alicia	F
Pedro	M
Cecilia	F
Raúl	M
Luis	M
Ana	F

Se cuentan las mujeres

```
Count[datos, {_, "F"}]
```

```
3
```

Fin del documento

```
$Version
```

```
9.0 for Microsoft Windows (64-bit) (November 20, 2012)
```

```
DateString[]
```

```
Tue 7 May 2013 22:58:31
```