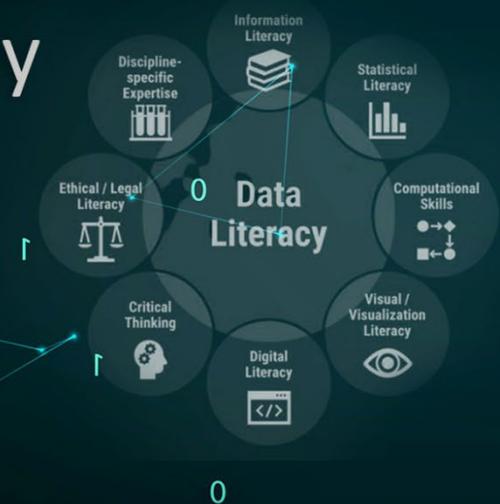




Alfabetización de Datos

Data Literacy



Prácticas del curso

Curso: Alfabetización de datos

Cursos de perfeccionamiento profesional – receso académico 2024

Organizado por: **Dirección de investigación**

Universidad Tecnológica de Panamá

Autor: Danny Murillo González

Centro de Investigación, Desarrollo e Innovación en Tecnologías de la Información y las Comunicaciones - CIDITIC

<https://orcid.org/0000-0003-0297-7213>

<https://scholar.google.com/citations?user=YNx08l0AAAAJ&hl=es>

Febrero 2024



Esta obra está bajo licencia internacional Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0.

Contenido

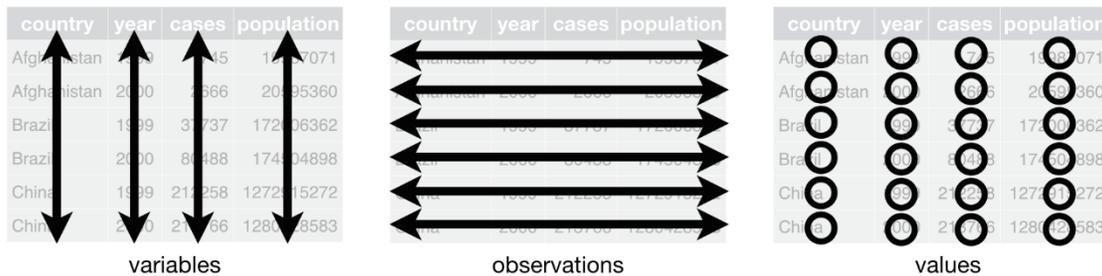
Práctica 1: Estructura de datos Tidy data	3
Práctica 2: Leer y Estructurar datos	7
Práctica 3: Uso de la Función BuscarX.....	8
Práctica 4: Tablas dinámicas	12
Práctica 3: Visualización de datos.....	17

Práctica 1: Estructura de datos | Tidy data

Herramienta: MS Excel

Tidy data (datos ordenados) : son simplemente datos organizados de una determinada manera donde el conjunto de datos deben cumplir con las siguientes tres reglas interconectadas:

1. Cada variable está contenida en una columna.
2. Cada observación se encuentra en una fila.
3. Cada valor (observación/variable) corresponde a una celda.



Ejemplos de tablas no tidy

Tabla 1: Salario de 3 personas:

year	Nicolas	Carla	Angela
2016	100	400	200
2017	500	600	700
2018	200	250	900

Tabla 2: Salario de 3 personas (wide format)

names	Salario_2016	Salario_2017	Salario_2018
Nicolas	100	500	200
Carla	400	600	250
Angela	200	700	900

Aunque en la tabla 1 y 2 se pueden entender perfectamente los datos, debemos analizar que en la tabla 1, no es común esa estructura, debido a que comúnmente colocamos los individuos en las filas y las variables o atributos en columnas. Al modificar la estructura como se ve en la tabla 2, podemos hacernos dos preguntas, en caso de seguir incrementando los años, ¿hacia dónde crecerá la tabla?

La respuesta es, crecerá a la derecha y este tipo de tabla se le conoce como **data wide o datos anchos o formato ancho**

La segunda pregunta es ¿es salario de 2016 al 2018 una variable?

La respuesta es, NO. Se debe a que 2016 -2018 realmente son valores que pudieran pertenecer a una variable año, por lo que es recomendable trabajar en las tres reglas de tidydata para transformar los datos a formato ancho o data long.

Tabla 3: Salario de 3 personas (long format)

names	year	salario
Nicolas	2016	100
Carla	2016	400
Angela	2016	200
Nicolas	2017	500
Carla	2017	600
Angela	2017	700
Nicolas	2018	200
Carla	2018	250
Angela	2018	900

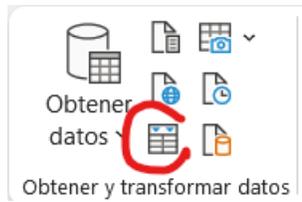
Practica en MS Excel

1. Utilizar **datos-practica-visualización.xlsx**
2. Vaya a la Hoja **country-data-wide**

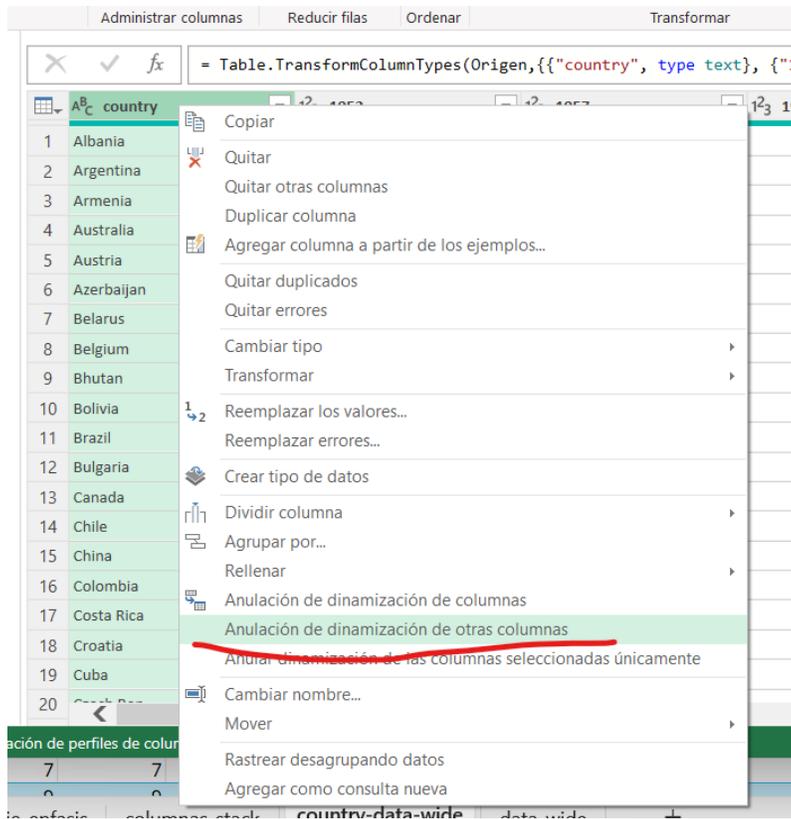
Esta tabla contiene 96 filas y 10 columnas

country	1952	1957	1962	1967	1972	1977	1982	1987	1992
Albania	-9	-9	-9	-9	-9	-9	-9	-9	5
Argentina	-9	-1	-1	-9	-9	-9	-8	8	7
Armenia	-9	-7	-7	-7	-7	-7	-7	-7	7
Australia	10	10	10	10	10	10	10	10	10
Austria	10	10	10	10	10	10	10	10	10
Azerbaijan	-9	-7	-7	-7	-7	-7	-7	-7	1
Belarus	-9	-7	-7	-7	-7	-7	-7	-7	7
Belgium	10	10	10	10	10	10	10	10	10
Bhutan	-10	-10	-10	-10	-10	-10	-10	-10	-10
Bolivia	-4	-3	-3	-4	-7	-7	8	9	9
Brazil	5	5	5	-9	-9	-4	-3	7	8
Bulgaria	-7	-7	-7	-7	-7	-7	-7	-7	8
Canada	10	10	10	10	10	10	10	10	10
Chile	2	5	5	6	6	-7	-7	-6	8
China	-8	-8	-8	-9	-8	-7	-7	-7	-7
Colombia	-5	7	7	7	7	8	8	8	9
Costa Rica	10	10	10	10	10	10	10	10	10

3. Seleccionar toda la tabla
4. Ir a la pestaña DATOS en Excel
5. Seleccione en la opción OBTENER Y TRANSFORMAR DATOS , el icono que indica DE UNA TABLA O RANGO



6. Se abrirá Power Query
7. Seleccione la columna country
8. Click al botón derecho y en el menú seleccione ANULACIÓN DE DINAMIZACIÓN DE OTRAS COLUMNAS



9. Se modificará la tabla de wide a long
10. Cerrar y cargar.
11. Se creará una nueva hoja con el nombre original de la tabla, renómbrala (opcional).

Si observa la nueva tabla ahora hay solo tres columnas con 864 filas.

La columna **country** a repetido los datos, y las columnas que contenían los años ahora son datos de la columna **Atributo**. Todos los valores de las columnas de los años (1952, 1957, 1962, 1967, 1972, 1977, 1982, 1987, 1992), ahora pertenecen a la columna **Valor**.

Lo que acaba de hacer MS Excel a través de Power Query es transformar la tabla (data wide) a una tabla ordenada (data long).

country	Atributo	Valor
Albania	1952	-9
Albania	1957	-9
Albania	1962	-9
Albania	1967	-9
Albania	1972	-9
Albania	1977	-9
Albania	1982	-9
Albania	1987	-9
Albania	1992	5
Argentina	1952	-9
Argentina	1957	-1
Argentina	1962	-1
Argentina	1967	-9
Argentina	1972	-9

Práctica 2: Leer y Estructurar datos

Herramienta: MS excel

Los documentos en formato XLSX tienen las características de poder almacenar los datos en hojas separadas, que su vez pueden servir para hacer cálculo entre ellas, pero si queremos hacer uso de estas hojas en formato CSV, no es posible, por el formato CSV no permite hojas en sus datos.

En esta práctica haremos uso de la función BUSCARX para buscar datos de una hoja en otra hoja, guardar la información en nuevas columnas y luego exportar los datos a CSV con la información necesaria.

Práctica cargar archivos .CSV

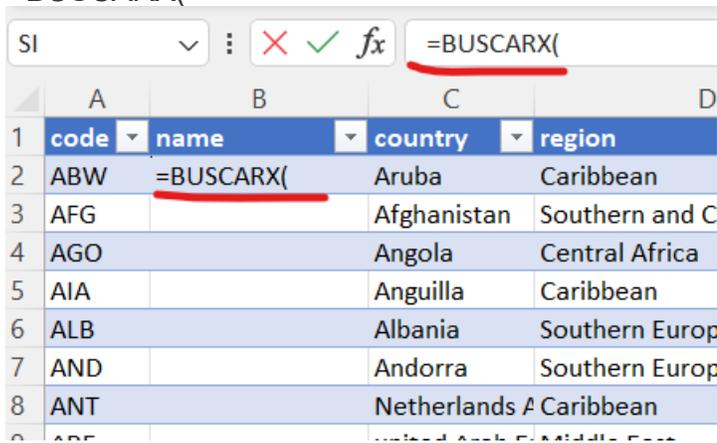
1. Abrir archivo **datos_practica-excel-CSV.xlsx**
2. **Integrar archivos CSV**
 - a. country-data-CODE-CSV.csv
 - b. country-data-continente-CSV
3. Vaya al menú **DATOS > De un archivo > de Texto / CSV**, seleccione el archivo **country-data-continente-CSV**
4. Renombre la hoja como continente.
5. Se mostrará la ventana de cargar datos, seleccione el botón CARGAR.
6. Nuevamente Vaya al Menú **DATOS > De un archivo > de Texto / CSV**, seleccione el archivo **country-data-CODE-CSV.csv**
7. Seleccione el botón CARGAR.
8. Renombre la hoja como CODE.

Práctica 3: Uso de la Función BuscarX

La función BUSCARX tiene como objetivo hacer búsqueda de datos en diferentes hojas, según el valor de búsqueda, para integrar el valor en la tabla principal. Esta función tiene tres parámetros:

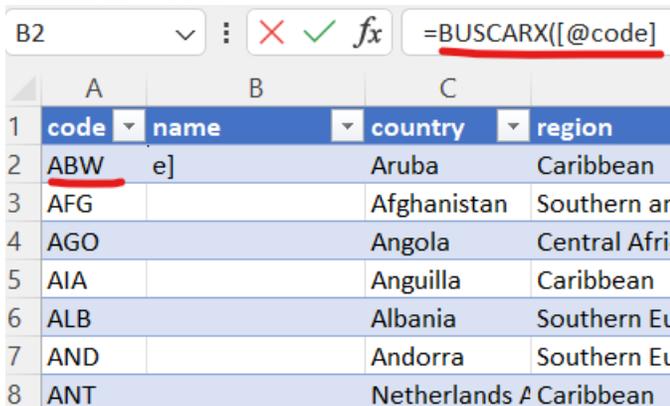
BUSCARX(valor buscado, matriz buscada, matriz devuelta o resultado)

1. Archivo **datos_practica-excel-CSV.xlsx**
2. En la hoja **country-data-CSV-format**, cree una nueva columna antes de la columna **name** y renómbrela como **country**.
3. En la primera fila de la columna **country** colocaremos:
=BUSCARX(



	A	B	C	D
1	code	name	country	region
2	ABW	=BUSCARX(Aruba	Caribbean
3	AFG		Afghanistan	Southern and C
4	AGO		Angola	Central Africa
5	AIA		Anguilla	Caribbean
6	ALB		Albania	Southern Europ
7	AND		Andorra	Southern Europ
8	ANT		Netherlands / Caribbean	

4. Se nos solicitará colocar el valor a buscar, seleccionar primer valor de la columna **code**.



	A	B	C	
1	code	name	country	region
2	ABW	=BUSCARX([@code]	Aruba	Caribbean
3	AFG		Afghanistan	Southern ar
4	AGO		Angola	Central Afri
5	AIA		Anguilla	Caribbean
6	ALB		Albania	Southern Eu
7	AND		Andorra	Southern Eu
8	ANT		Netherlands / Caribbean	

5. Al seleccionar el primer valor la celda se mostrará la siguiente formula, colocar punto y coma al final para colocar el siguiente parámetro
=BUSCARX([@code];

6. El siguiente parámetro está relacionado, en que hoja/ tabla debe buscar el valor seleccionado en la columna **code**. Para ellos vamos a colocar el cursor después de la función escrita como:

=BUSCARX([@code];

7. Nos ubicamos en la hoja **CODE** y seleccionamos todos los datos de la columna **CODE**.

Ahora en la barra de formula aparecerá:

=BUSCARX([@code];country_data_CODE_CSV[code]

The screenshot shows an Excel spreadsheet with the following data in columns A and B:

	A	B	C	D
215	BIH	Bosnia and Herzegovina		
216	BHS	Bahamas		
217	BHR	Bahrain		
218	BGR	Bulgaria		
219	BGD	Bangladesh		
220	BFA	Burkina Faso		
221	BEN	Benin		
222	BEL	Belgium		
223	BDI	Burundi		
224	AZE	Azerbaijan		
225	AuT	Austria		
226	AuS	Australia		
227	ATG	Antigua and Barbuda		
228	ATF	French Southern territories		
229	ATA	Antarctica		
230	ASM	American Samoa		
231	ARM	Armenia		
232	ARG	Argentina		
233	ARE	united Arab Emirates		
234	ANT	Netherlands Antilles		
235	AND	Andorra		
236	ALB	Albania		
237	AIA	Anguilla		
238	AGO	Angola		
239	AFG	Afghanistan		
240	ABW	Aruba		
241				

The formula bar at the top shows: `=BUSCARX([@code];country_data_CODE_CSV[code]`

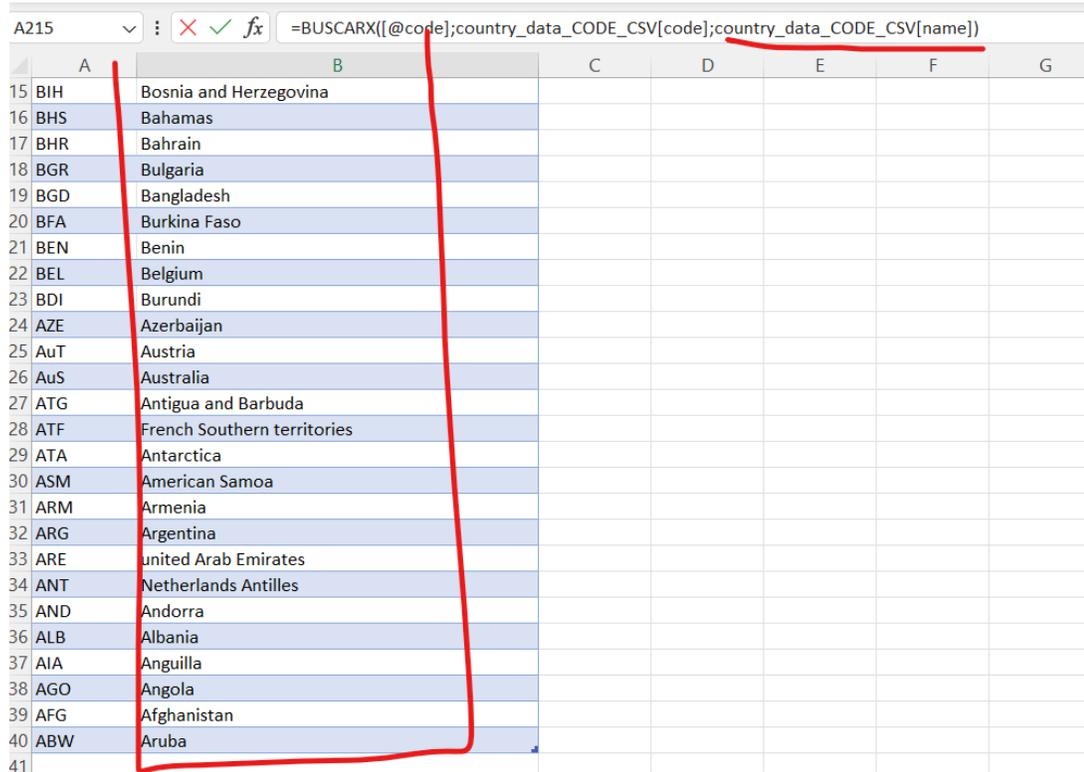
The bottom sheet tab is labeled `code`, which is circled in red.

8. Después de la fórmula, colocamos nuevamente punto y coma; se nos pida ahora seleccionar la columna que va a darnos los datos de

resultados.

Para ellos seleccionamos todos los datos la columna **name** de la misma hoja **code**. Al final de la fórmula debemos colocar el paréntesis de cierre y la fórmula quedará de la siguiente forma:

=BUSCARX([@code];country_data_CODE_CSV[code];country_data_CODE_CSV[name])



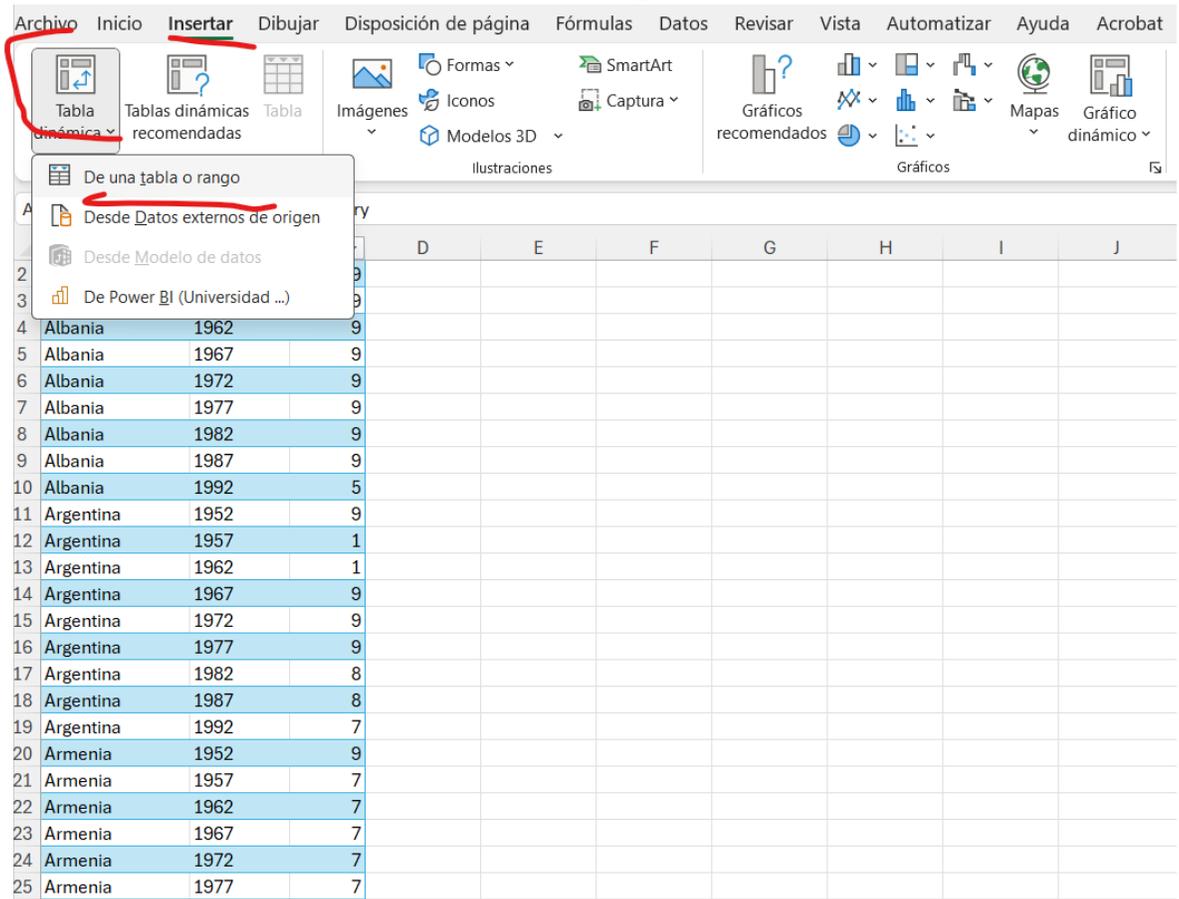
	A	B	C	D	E	F	G
15	BIH	Bosnia and Herzegovina					
16	BHS	Bahamas					
17	BHR	Bahrain					
18	BGR	Bulgaria					
19	BGD	Bangladesh					
20	BFA	Burkina Faso					
21	BEN	Benin					
22	BEL	Belgium					
23	BDI	Burundi					
24	AZE	Azerbaijan					
25	AuT	Austria					
26	AuS	Australia					
27	ATG	Antigua and Barbuda					
28	ATF	French Southern territories					
29	ATA	Antarctica					
30	ASM	American Samoa					
31	ARM	Armenia					
32	ARG	Argentina					
33	ARE	United Arab Emirates					
34	ANT	Netherlands Antilles					
35	AND	Andorra					
36	ALB	Albania					
37	AIA	Anguilla					
38	AGO	Angola					
39	AFG	Afghanistan					
40	ABW	Aruba					
41							

Práctica 4: Tablas dinámicas

La tabla dinámica es una de las funciones más potentes e intimidantes de Microsoft Excel. Potente porque puede ayudarte a resumir y dar sentido a grandes conjuntos de datos.

Además de servir como herramientas para resumir datos a través de variables categóricas, también es posible mostrar si una tabla no está ordenada.

1. Archivo datos **_practica-excel-CSV.xlsx**
2. Ir a la hoja **dem_score**
3. Ir al menú INSERTAR
4. Seleccione TABLA DINÁMICA > De una tabla a Rango



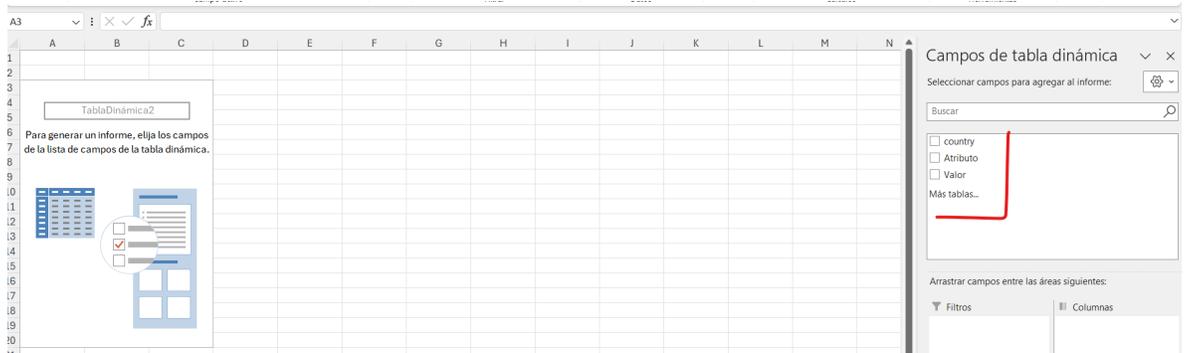
The screenshot shows the Microsoft Excel interface with the 'Insertar' ribbon selected. The 'Tabla Dinámica' button is highlighted with a red circle. A dropdown menu is open, showing the following options:

- De una tabla o rango
- Desde Datos externos de origen
- Desde Modelo de datos
- De Power BI (Universidad ...)

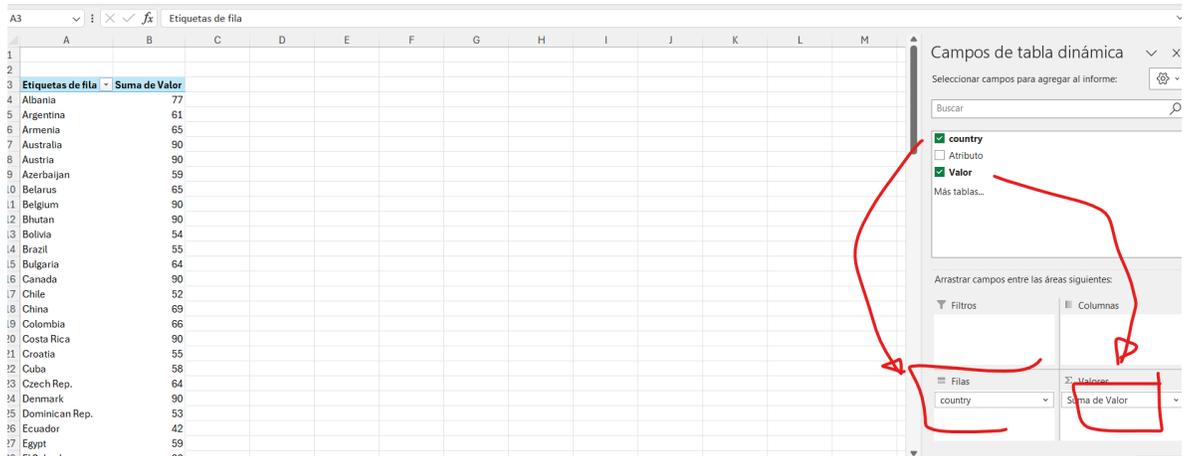
The background shows a data table with the following content:

	D	E	F	G	H	I	J
4	Albania	1962	9				
5	Albania	1967	9				
6	Albania	1972	9				
7	Albania	1977	9				
8	Albania	1982	9				
9	Albania	1987	9				
10	Albania	1992	5				
11	Argentina	1952	9				
12	Argentina	1957	1				
13	Argentina	1962	1				
14	Argentina	1967	9				
15	Argentina	1972	9				
16	Argentina	1977	9				
17	Argentina	1982	8				
18	Argentina	1987	8				
19	Argentina	1992	7				
20	Armenia	1952	9				
21	Armenia	1957	7				
22	Armenia	1962	7				
23	Armenia	1967	7				
24	Armenia	1972	7				
25	Armenia	1977	7				

5. Se mostrará la siguiente pantalla con los campos (variables) que contiene la tabla



6. Podemos con esta estructura conocer ¿Cuál es el país con mayor cantidad de ventas? Arrastramos el valor de **country** (variable categórica) al espacio de filas y arrastramos el campo **Valor** al espacio que dice **valores**. Se generara una tabla con los valores de la tabla ordenada por el campo **country**.



7. Si le damos click con el botón derecho donde aparece los datos de la columna Suma de valores, podemos ordenar los datos de mayor a menor.

B10 65

	A	B	C	D	E	F	G
1							
2							
3	Etiquetas de fila	Suma de Valor					
4	Albania	77					
5	Argentina	61					
6	Armenia	65					
7	Australia						
8	Austria						
9	Azerbaijan						
10	Belarus						
11	Belgium						
12	Bhutan						
13	Bolivia						
14	Brazil						
15	Bulgaria						
16	Canada						
17	Chile						
18	China						
19	Colombia						
20	Costa Rica						
21	Croatia						
22	Cuba						
23	Czech Rep.						
24	Denmark						
25	Dominican Rep.						
26	Ecuador						
27	Event						

Aptos N: 11

Ordenar

- Ordenar de menor a mayor
- Ordenar de mayor a menor
- Más opciones de ordenación...

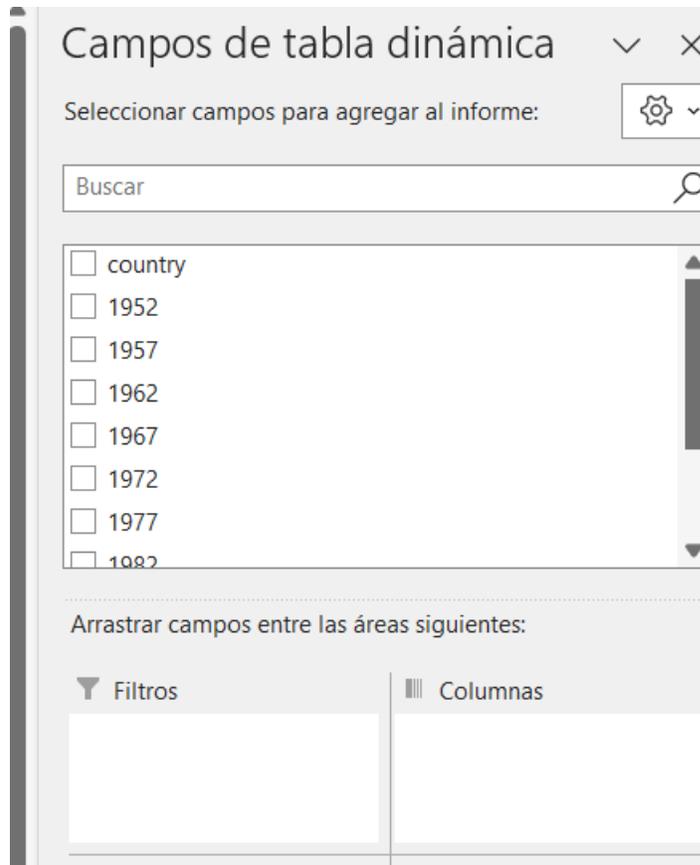
8. Con esto podemos ver los países con mayor venta en todos los años

Etiquetas de fila ▼	Suma de Valor
(en blanco)	6360
Norway	90
Italy	90
Ireland	90
Austria	90
Netherlands	90
Belgium	90
Sweden	90
Bhutan	90
United Kingdom	90
Canada	90
Australia	90
Japan	90
Costa Rica	90
New Zealand	90
Denmark	90
Saudi Arabia	90
Finland	90
Switzerland	90
Germany	90
United States	90

Ejemplo 2 con los datos originales sin ser tabla ordenada.

Si utilizamos ahora la tabla **country-data-wide** y hacemos el mismo procedimiento.

1. Ir al menú INSERTAR
2. Seleccione TABLA DINÁMICA > De una tabla a Rango
3. Se mostrará el conjunto de datos de la siguiente forma, por lo que solo es posible saber las venta de los países por cada año (columnas) , por lo que podríamos decir que esta es una tabla desordenada o no Tidy.



Práctica 3: Visualización de datos

Herramienta: DataWrapper

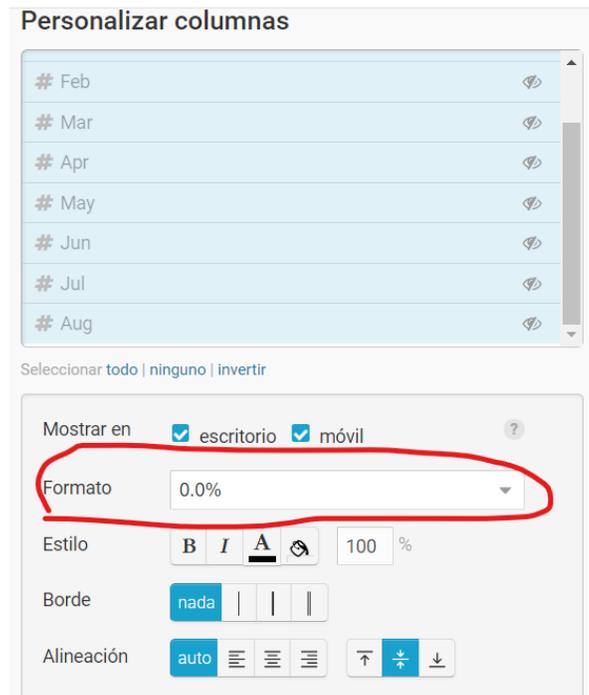
Es una herramienta de visualización de datos online creada para mostrar tablas, mapas y gráficos que pueden ser exportados como imagen o compartidos como enlace web.

<https://www.datawrapper.de/>

<https://app.datawrapper.de/chart/aXLjX/upload>

Práctica 1: Tabla

1. Cargar los datos del archivo
2. Seleccione la tabla color **tabla-color**
3. Seleccione en la pestaña TIPOS DE GRÁFICOS seleccione TABLA
4. En la pestaña MEJORAR hacer los siguientes cambios:
 - a. Clasificar tabla por columna: JAN
 - b. Clasificar tabla por dirección: ASCENDENTE
5. Personalizar Columna:
 - a. Selecciones los nombre de los meses de l año como se muestra en la figura y seleccione en FORMATO: 0.0%



6. En mapa de calor, seleccione ACTIVAR MAPA DE CALOR

Mapa de calor

Seleccionar paleta



Tipo

pasos

continuo



Interpolación

linear



Rango

2.4
MÍN.

4.95
CENTRO

14.7
MÁX.



Ocultar los valores en las columnas del mapa de calor

Mostrar leyenda de colores

Tu visualización está usando colores, así que te recomendamos que añadas una leyenda para explicar lo que significan los colores.

Desactivar mapa de calor

Para desactivar el mapa de calor de las columnas individuales, selecciónalas y personalízalas arriba.

7. Personalizar celdas : Seleccione las fila US y coloque en STYLE : negrita, haga lo mismo con la fila CANADA, para hacer énfasis en las filas con mayor valor.

Personalizar filas

Seleccionar fila **US [5]**

Style **B** **A** 100 %

Top border default

Bottom border default

Mover fila a comienzo de tabla

Ignorar formato del número

copy style reset

8. En la pestaña ANOTAR modificar:
- Titulo: Tasa de desempleo en países seleccionados
 - Descripción: Enero-agosto de 2020, ordenados por tasa de desempleo en enero.
 - Fuente de datos : OECD
9. Publicar e integrar
- Permite generar una URL del gráfico el cual podemos compartir.
 - Permite generar un código SCRIPT para insertar en una página web.
 - Permite descargar una imagen en PNG del gráfico.

Imagen del Tabla de mapa de calor generado

Unemployment rate in selected countries

January-August 2020, sorted by the unemployment rate in January.

country	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Japan	2.4%	2.4%	2.5%	2.6%	2.9%	2.8%	2.9%	3.0%
Netherlands	3.0%	2.9%	2.9%	3.4%	3.6%	4.3%	4.5%	4.6%
Germany	3.4%	3.6%	3.8%	4.0%	4.2%	4.3%	4.4%	4.4%
Mexico	3.6%	3.6%	3.2%	4.8%	4.3%	5.4%	5.2%	5.0%
US	3.6%	3.5%	4.4%	14.7%	13.3%	11.1%	10.2%	8.4%
South Korea	4.0%	3.3%	3.8%	3.8%	4.5%	4.3%	4.2%	3.2%
Denmark	4.9%	4.9%	4.8%	4.9%	5.5%	6.0%	6.3%	6.1%
Belgium	5.1%	5.0%	5.0%	5.1%	5.0%	5.0%	5.0%	5.1%
Australia	5.3%	5.1%	5.2%	6.4%	7.1%	7.4%	7.5%	6.8%
Canada	5.5%	5.6%	7.8%	13.0%	13.7%	12.3%	10.9%	10.2%
Finland	6.8%	6.9%	7.0%	7.3%	7.5%	7.8%	8.0%	8.1%

Source: OECD • Created with Datawrapper

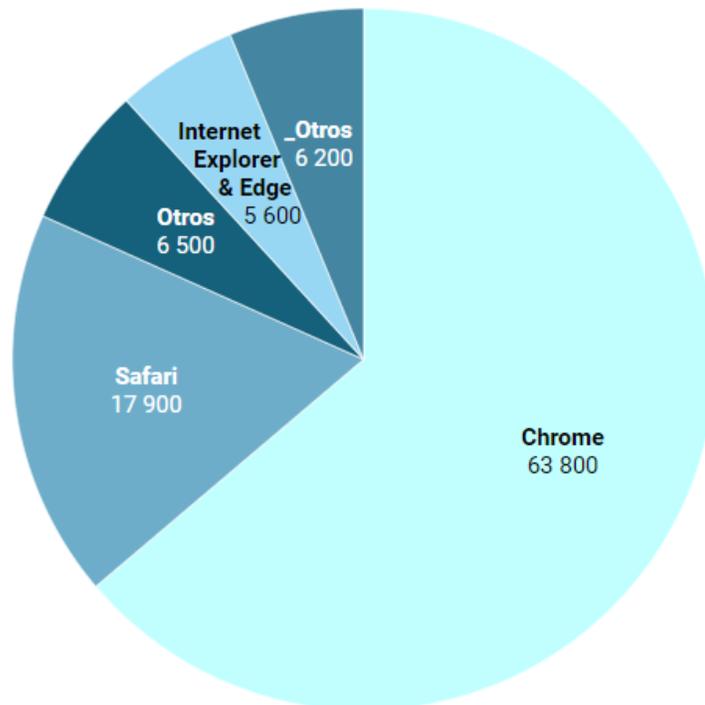
Práctica 2: Gráfico circular

1. Cargar los datos del archivo
2. Seleccione la tabla color **pie-énfasis**
3. Seleccione en la pestaña TIPOS DE GRÁFICOS seleccione **Gráfico de círculo**

Comentarios:

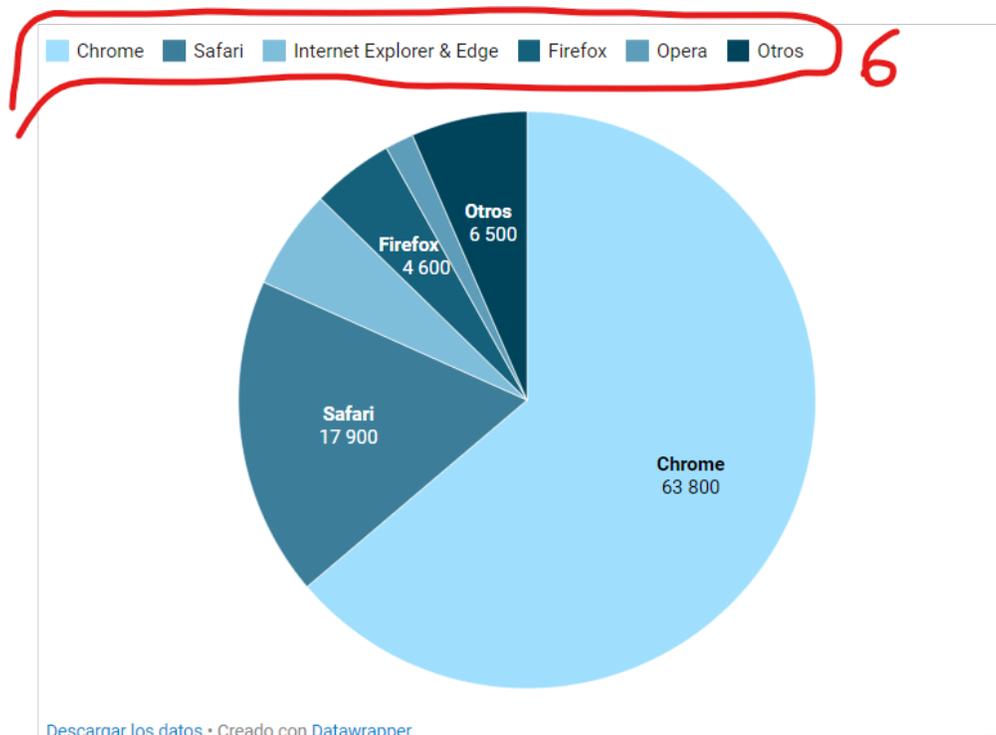
- Gráfico de Pastel generado con tonos de un solo color predeterminado.
- Solo se visualizan cinco partes en el gráfico (hay seis datos).
- Los datos se ordenaron de mayor a menor.
- El gráfico genero los colores en cada slide de acuerdo a su criterio.

Chrome Safari Otros Internet Explorer & Edge _Otros



[Descargar los datos](#) • Creado con [Datawrapper](#)

4. En la pestaña MEJORAR, sección AGRUPACIÓN, hacer los siguientes cambios:
 - a. Número máximo de fragmentos: 6

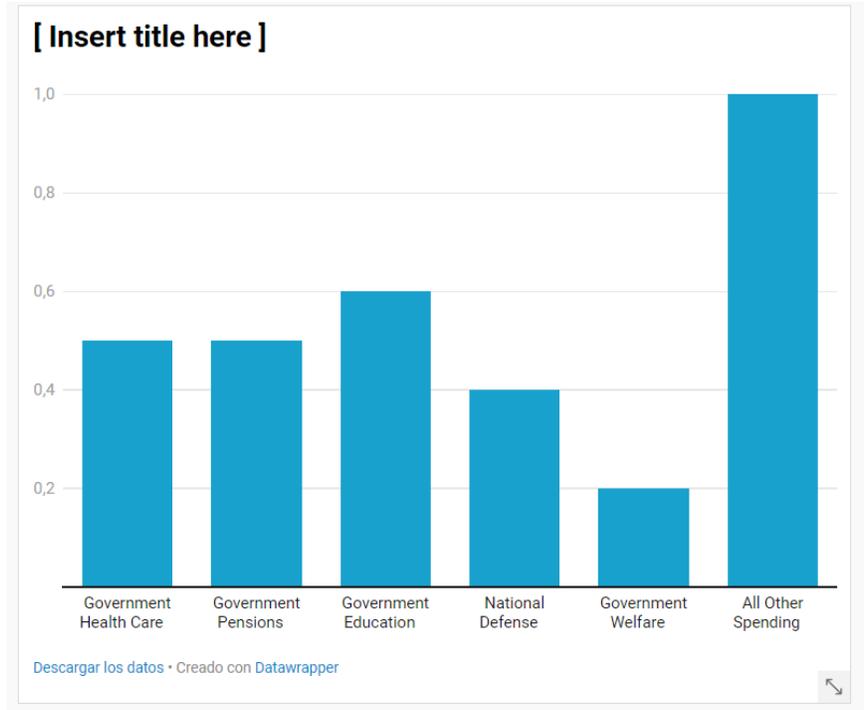


Se quiere hacer énfasis en el slide de mayor tamaño, el cual se puede hacer cambiando el color de lo slide no relevantes a colores neutros.

5. En la pestaña MEJORAR, **Fragmentos del círculo**, hacer los siguientes cambios:
 - a. Tamaño del gráfico: 60% (a criterio)
 - b. Ordenar por: Mantener el orden
 - c. Color de fragmento: personalizar colores
 - i. Chrome: #4b8cb7
 - ii. Safari: #c4c4c4
 - iii. Edge: #cecece
 - iv. Firefox: #e0e0e0
 - v. Opera: #f0f0f0
 - vi. Otros: #ffffff
6. En la pestaña MEJORAR, **Etiquetas**, hacer los siguientes cambios:
 - a. Seleccione, convertir valores a porcentaje
 - b. Seleccione, etiquetas externas
7. En la pestaña ANOTAR modificar:
 - a. Título: Uso del navegador Chrome por usuarios
 - b. Descripción: Cuota de mercado de navegadores web, octubre de 2020. Este informe se generó el 31/10/2020 en función del tráfico del mes pasado a todos los sitios web que utilizan las estadísticas web gratuitas de W3Counter.
 - c. Fuente de datos : OECD
8. Puede modificar el tipo de gráfico a Gráfico de anillo
9. En la pestaña MEJORAR, **Etiquetas**, hacer los siguientes cambios:
 - a. Seleccione, mostrar valor dentro del anillo.

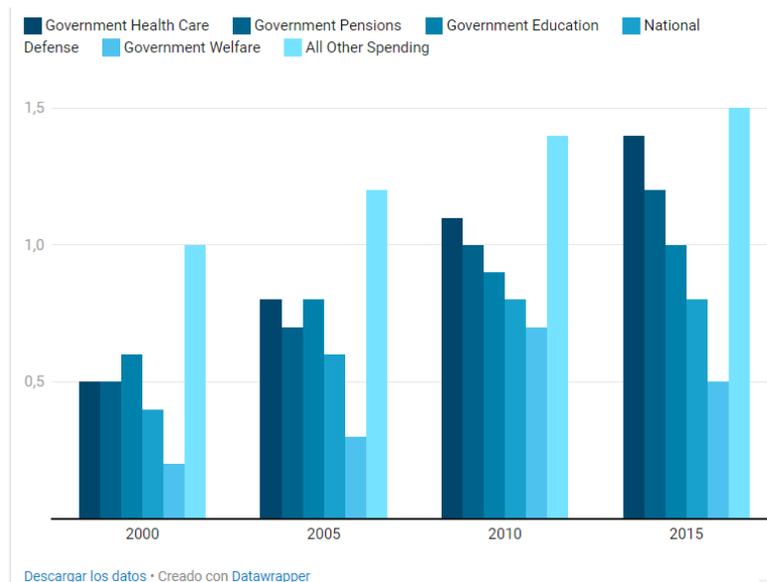
Práctica 3: Columnas

1. Cargar los datos del archivo
2. Seleccione la tabla color **columnas-stack**
3. Seleccione en la pestaña TIPOS DE GRÁFICOS seleccione **Gráfico de columnas**



Comentarios:

- Las columnas en la tabla representan columnas en el gráfico, pero DATAWRAPPER solo interpreta los datos de la primera columna.
4. Seleccione en la pestaña TIPOS DE GRÁFICOS seleccione **Columnas agrupadas**



Comentarios:

- Las columnas en la tabla representan columnas en el gráfico, que ahora se muestran en el gráfico de columnas agrupados, sin embargo el objetivo es mostrar un gráfico que pueda comparar los años por columnas.

5. Seleccione en la pestaña TIPOS DE GRÁFICOS seleccione **Columnas apiladas**
6. En la pestaña MEJORAR, **Eje vertical**, hacer los siguientes cambios:
 - a. Formato numérico: 1,000[00]
 - b. Formato de Fecha: 2015-2016
 - c. Color de columna: personalizar colores
 1. All Other Spending: #f0f0f0
 - d. Espacio entre columnas: 35%
 - e. Seleccione, conectar columnas con áreas
 - f. Formato numérico: igual que el eje vertical.
7. En la pestaña MEJORAR, **Etiquetas**, hacer los siguientes cambios:
 - a. Margen de etiqueta: 25%
 - b. Seleccione, mostrar valores
8. En la pestaña ANOTAR modificar:
 - a. Título: 15 años de gasto del gobierno estadounidense
 - b. Descripción: En billones de dólares estadounidenses, de 2000 a 2015.
 - c. Fuente de datos : OECD

