



Islas Canarias
Del 15 al 19 de noviembre de 2021



APLICACIÓN DE CONSULTA DE DATOS DEL MARCO INPUT-OUTPUT DE GALICIA EN R-SHINY

Esther Calvo Ocampo

Instituto Galego de Estatística (IGE)

esther.calvo@ige.eu

Raúl Jácome Rodríguez

Instituto Galego de Estatística (IGE)

raul.jacome@ige.eu

Antonio Albo Díaz

Instituto Galego de Estatística (IGE)

antonio.albo@ige.eu

M^a Trinidad Teijeiro Campo

Instituto Galego de Estatística (IGE)

trinidad.teijeiro@ige.eu

Marta Suárez Morao

Instituto Galego de Estatística (IGE)

marta.suarez@ige.eu

Carmen Vilas Cruz

Instituto Galego de Estatística (IGE)

carmen.vilas@ige.eu

Introducción

El Marco Input-Output de Galicia (MIOGAL) constituye una herramienta compleja pero fundamental para describir los flujos económicos de nuestra región. Es el retrato más detallado de una economía pero, quizás por este grado de detalle, resulta difícil a los usuarios extraer siquiera un bajo porcentaje de la información que contiene.

Habitualmente ha dominado su vertiente analítica, a través de las múltiples aplicaciones del análisis input output que, en su gran mayoría, toman de base la tabla input-output o tabla simétrica relegando a las tablas de origen y destino a un papel secundario. Éstas, no sólo sirven para derivar la matriz simétrica sino que son de gran relevancia para dar coherencia interna al entramado de información macroeconómica de nuestro sistema de cuentas económicas. Pero además, ofrecen una información rica en detalles, estructurada, comparable con otras economías y con otros períodos.

La presentación de un modo claro y comprensible de los datos difundidos es uno de los desafíos de los productores de estadística oficial, como recoge el Código de conducta de las estadísticas europeas.

En esta ponencia describimos el desarrollo de la aplicación de consulta diseñada en R-Shiny, que tiene por objetivo ofrecer al usuario una visión amigable de la información contenida en unas matrices de datos que, por su tamaño y por la metodología con que se construyen, requieren cierta especialización y conocimientos previos para su uso e interpretación.

Este desarrollo sigue la línea de trabajo marcada en el IGE de mejorar, mediante aplicaciones gráficas de consulta, la difusión de nuestras operaciones y actividades estadísticas.

Objetivos

Las tablas de origen y destino (TOD) son matrices rectangulares que describen los valores de las operaciones de bienes y servicios de una economía, desagregadas por productos (filas) y ramas de actividad (columnas).

La tabla de origen ofrece una imagen detallada de la oferta de bienes y servicios por producto y por tipo de oferente, distinguiendo en el caso de la oferta interior la producción por ramas, y en el caso de la oferta importada los bienes y servicios adquiridos a unidades no residentes.

La tabla de destino muestra los usos de la oferta de bienes y servicios. Estos usos pueden dividirse en consumos intermedios y consumo final, bien sea de los hogares, de las administraciones públicas o de las instituciones sin fines de lucro al servicio de los hogares (gasto en consumo final), para inversión (formación bruta de capital) o bien para su venta al exterior (exportaciones). Además, y este es un aspecto que muestra la vinculación con las cuentas anuales, la tabla de destino refleja los diferentes componentes del valor añadido por rama de actividad en la economía.

Existen dos tipos de entidades entre ambas tablas o, dicho de otro modo, la combinación de ambas tablas ofrece:

- Desde la perspectiva de la rama de actividad una imagen del volumen de producción total por rama, los bienes y servicios producidos; los consumos necesarios para alcanzar dicha producción y el detalle de éstos. La producción de cada rama debe cubrir los insumos, es decir, la producción ha de ser igual a los consumos intermedios y el valor añadido bruto, cuyos componentes (perspectiva rentas) son la remuneración de asalariados, el excedente de explotación y los otros impuestos netos sobre la producción y las importaciones.
- Desde la perspectiva de los productos nos permite conocer la composición de la oferta de cada producto presente en la economía gallega y los destinos (o demanda de los mismos). Para cada producto se cumple que la producción más las importaciones (que son la oferta total de la economía) es igual a la suma de consumos intermedios, gasto en consumo final y formación bruta de capital.

Las TOD publicadas contienen información de 72 ramas y 110 productos. Se difunden en hojas de cálculo (Excel y Open office) para su descarga completa. El objetivo de nuestra aplicación en R-Shiny es mostrar adecuadamente esta doble perspectiva que resumen las entidades contables que se cumplen en las TOD. La aplicación permitirá la

consulta de una rama de actividad o bien de un producto para así conocer o bien la estructura productiva de una determinada actividad (perspectiva rama) o bien el equilibrio oferta-demanda de cada uno de los productos (bienes y servicios) de la economía.

Tabla 1. Estructura simplificada de la tabla de origen del Marco Input-Output de Galicia

OFERTA		Ramas de actividad			Resto del mundo	Total
		R01	...	R97		
Productos	01A	Producción por producto y por rama de actividad			Importaciones por producto	Oferta total por producto
	...					
	97					
Total		Producción total por rama de actividad			Importaciones totales de la economía	Oferta total de la economía

Tabla 2. Estructura simplificada de la tabla de destino del Marco Input-Output de Galicia

DESTINO		Ramas de actividad			Gasto en consumo final	Formación bruta de capital	Resto del mundo	Total
		R01	...	R97				
Productos	01A	Consumos intermedios por producto y rama de actividad			Gasto en consumo final por producto	Formación bruta de capital por producto	Exportaciones por producto	Empleos totales por producto
	...							
	97							
Componentes del VAB		VAB por componente y por rama de actividad						
Total		Insumos totales por rama de actividad						

Metodología y resultados

Con el fin de sintetizar y organizar la cantidad de información que ofrece el Marco Input-Output, se desarrolló en R Shiny una aplicación que permite consultar la información por dos vías: por rama de actividad y por producto.

Información por producto

En el caso de los productos, se trató de condensar la información en forma de tablas y gráficos de forma que se muestre el equilibrio oferta-demanda. Con este fin, se incorporaron en esta parte un tipo de gráficos que permiten estudiar flujos y que

hicieron posible representar los flujos de la oferta y demanda de cada producto, los diagramas de Sankey.

Un diagrama de Sankey representa flujos, es decir, conexiones ponderadas que van de un nodo a otro.

En el contexto del Marco Input-Output, el diagrama de Sankey se utilizó para visualizar los flujos de la oferta y de la demanda del producto. Los nodos fuente son en este caso las ramas de actividad productoras del bien o servicio (producto) considerado, incluyendo también los nodos correspondientes a las importaciones del resto de España, importaciones del resto de la Unión Europea e importaciones del Resto del Mundo.

A su vez, toda esta oferta del producto considerado se agrupa en dos nodos centrales, que diferencian la procedencia de la oferta del producto (importación e interior).

Por último, los nodos de destino están formados por las ramas de actividad como demandantes del producto seleccionado así como por los restantes componentes de la demanda (gasto en consumo final, formación bruta de capital, exportaciones con destino al resto de España, exportaciones con destino al resto de la Unión Europea, exportaciones con destino al Resto del Mundo).

La librería que se utilizó para implementar este tipo de gráficos fue networkD3, htmlwidget que crea automáticamente gráficos interactivos.

Los datos de entrada se pueden almacenar en 2 formatos diferentes:

- data frame con las conexiones (3 columnas)
- Matriz de incidencia (matriz cuadrada)

En este caso se utilizó como formato de los datos de entrada un data frame, el cual enumera una por una la conexiones y está formado por 3 columnas:

- source: nodo de origen
- target: nodo de destino
- valor del flujo

Se construyó este data frame a partir de la matriz de origen, la matriz de destino de la producción interior a precios básicos y la matriz de destino de las importaciones a precios básicos. La primera nos facilitará la información relativa a la oferta y las dos segundas relativas a la demanda.

De la matriz de origen se obtiene para cada producto la oferta desagregada por ramas y por los diferentes tipos de importaciones. Esta información será la que se utilice para construir la primera parte del gráfico:

- nodos de origen: ramas de actividad de origen e importaciones.
- nodos de destino: son dos; IMPORTACIÓN e INTERIOR.
- flujos: representan la procedencia de la oferta según los valores de la matriz de origen.

Tabla 3. Flujos de la oferta para el producto 17A

cod_source	name_source	ord_source	name_target	ord_target	valor
R16_O	Industria da madeira e da cortiza, agás mobles; cestaría e espartaría	13	INTERIOR	74	854
R17_O	Industria do papel	14	INTERIOR	74	232631
IMP_ESP	Importacións do resto de España	71	IMPORTACIÓN	75	55862
IMP_UE	Importacións do resto da UE	72	IMPORTACIÓN	75	103380
IMP_RM	Importacions do resto do mundo	73	IMPORTACIÓN	75	2256

En la Tabla 3 se muestra la información procedente de la matriz de origen en el caso del producto 17A.

Los campos `cod_source` y `name_source` son los correspondientes al código y descripción de los nodos de origen y el campo `name_target` es el correspondiente a los nodos de destino. El campo `valor` es el que se utiliza para representar los flujos.

Por otro lado, la matriz de destino de la producción interior a precios básicos y la matriz de destino de las importaciones a precios básicos nos proporcionan información sobre los flujos con origen en la producción interior o en la importación y con destino los distintos componentes de la demanda.

En la Tabla 4 se muestra la información procedente de la matriz de producción interior a precios básicos para el producto 17A.

Tabla 4. Flujo de demanda de la producción interior para el producto 17A

name_source	ord_source	cod_target	name_target	ord_target	valor
INTERIOR	74	R10B_DI	Procesamento e conservación de peixes, crustáceos e moluscos	82	604
INTERIOR	74	R10C_DI	Fabricación de produtos lácteos	83	458
INTERIOR	74	R10E_DI	Outras industrias alimentarias	85	75
INTERIOR	74	R11_DI	Fabricación de bebidas	86	301
INTERIOR	74	R13_DI	Industria téxtil	87	2
INTERIOR	74	R16_DI	Industria da madeira e da cortiza, agás mobles;	89	7658

			cestaría e espartaría		
INTERIOR	74	R18_DI	Artes gráficas e reprodución de soportes gravados	91	1097
INTERIOR	74	R20_21_DI	Industria química e fabricación de produtos farmacéuticos	93	873
INTERIOR	74	R22_DI	Fabricación de produtos de caucho e plásticos	94	711
INTERIOR	74	R23_DI	Fabricación doutros produtos minerais non metálicos	95	965
INTERIOR	74	R24_DI	Metalurxia; fabricación de produtos de ferro, aceiro e ferroalixes	96	4158
INTERIOR	74	R25_DI	Fabricación de produtos metálicos, agás maquinaria e equipamento	97	547
INTERIOR	74	R29_DI	Fabricación de vehículos de motor, remolques e semirremolques	101	1178
INTERIOR	74	R45_DI	Venda e reparación de vehículos de motor	110	503
INTERIOR	74	R46_DI	Comercio polo xunto e intermediarios do comercio, agás de vehículos de motor	111	182
INTERIOR	74	R47_DI	Comercio polo miúdo, salvo de vehículos de motor e motocicletas	112	389
INTERIOR	74	R56_DI	Servizos de comidas e bebidas	118	1558
INTERIOR	74	R58_DI	Edición	119	226
INTERIOR	74	R59_60_DI	Actividades cinematográficas de video e televisión, gravación de son e edición musical; actividades de programación e emisión de radio e televisión	120	1
INTERIOR	74	R84_DI	Administración pública e defensa; Seguridade Social obrigatoria	136	772

INTERIOR	74	R85M_DI	Educación de mercado	137	5
INTERIOR	74	R85NM_DI	Educación de non mercado	138	406
INTERIOR	74	R86_88M_DI	Actividades sanitarias e de servizos sociais de mercado	139	7
INTERIOR	74	FBK	Formación bruta de capital	148	106
INTERIOR	74	EXP_ESP	Exportacións ao resto de España	149	23346
INTERIOR	74	EXP_UE	Exportacións ao resto da UE	150	175686
INTERIOR	74	EXP_RM	Exportacións ao resto do mundo	151	11671

En este caso hay un único nodo origen “INTERIOR” y la tabla recoge los flujos existentes de producción interior a distintos componentes de la demanda final para el caso del producto 17A, que es el que se está considerando.

En la Tabla 5 se muestra la información procedente de la matriz de importaciones a precios básicos para el producto 17A.

Tabla 5. Flujo de demanda de las importaciones para el producto 17A

name_source	ord_source	cod_target	name_target	ord_target	valor
IMPORTACIÓN	75	R10B_DI	Procesamento e conservación de peixes, crustáceos e moluscos	82	1034
IMPORTACIÓN	75	R10C_DI	Fabricación de produtos lácteos	83	784
IMPORTACIÓN	75	R10E_DI	Outras industrias alimentarias	85	130
IMPORTACIÓN	75	R11_DI	Fabricación de bebidas	86	515
IMPORTACIÓN	75	R13_DI	Industria téxtil	87	36
IMPORTACIÓN	75	R16_DI	Industria da madeira e da cortiza, agás mobles; cestería e espartaría	89	38772
IMPORTACIÓN	75	R17_DI	Industria do papel	90	55914
IMPORTACIÓN	75	R18_DI	Artes gráficas e reprodución de soportes gravados	91	38300
IMPORTACIÓN	75	R20_21_DI	Industria química e fabricación de produtos	93	1493

			farmacéuticos		
IMPORTACIÓN	75	R22_DI	Fabricación de produtos de caucho e plásticos	94	1217
IMPORTACIÓN	75	R24_DI	Metalurxia; fabricación de produtos de ferro, aceiro e ferroalixes	96	7112
IMPORTACIÓN	75	R25_DI	Fabricación de produtos metálicos, agás maquinaria e equipamento	97	935
IMPORTACIÓN	75	R27_DI	Fabricación de material e equipamento eléctrico	99	53
IMPORTACIÓN	75	R29_DI	Fabricación de vehículos de motor, remolques e semirremolques	101	2016
IMPORTACIÓN	75	R45_DI	Venda e reparación de vehículos de motor	110	1174
IMPORTACIÓN	75	R46_DI	Comercio polo xunto e intermediarios do comercio, agás de vehículos de motor	111	424
IMPORTACIÓN	75	R47_DI	Comercio polo miúdo, salvo de vehículos de motor e motocicletas	112	909
IMPORTACIÓN	75	R58_DI	Edición	119	8498
IMPORTACIÓN	75	FBK	Formación bruta de capital	148	2182

En este caso se muestran todos los flujos de destino de la producción importada del producto 17A. En este caso hay un único nodo origen “IMPORTACIÓN”.

La Tabla 3, la Tabla 4 y la Tabla 5 contienen toda la información necesaria para la elaboración del gráfico con el paquete networkD3.

En primer lugar construimos un data frame con las conexiones (links), que estará formado por 3 columnas: source, target, value y que se corresponden en este caso con el nodo fuente (cod_source en la primera tabla y name_source en las otras dos), el nodo destino (name_target en la primera tabla y cod_target en las otras dos) y la intensidad de cada flujo, que viene dada en este caso por el valor.

A partir de este data frame es necesario crear también otro data frame de nodos (nodes), en el que se listan todos los nodos del diagrama.

Otro aspecto importante a tener en cuenta al utilizar la librería networkDs es que las conexiones deben proporcionarse mediante id, no mediante el nombre que aparece en el

data frame con las conexiones, por lo que previamente a la realización del gráfico será necesario crear un campo id para source (IDsource) y otro id para target (IDtarget).

Se añadió además una columna group tanto al data frame links como al nodes que permite utilizar un color distinto para cada grupo de flujos o para cada grupo de nodos.

En este caso se optó por utilizar un único color para todos los nodos por lo que consideramos un único grupo y dos colores en el caso de los flujos para distinguir el flujo con procedencia en la producción interior del flujo con procedencia en la importación.

La sintaxis que se utilizó para la construcción del diagrama de flujos fue la siguiente:

sankeyNetwork (Links = links, Nodes = nodes,

Source = "IDsource", Target = "IDtarget",

Value = "value", NodeID = "name",

colourScale=my_color, LinkGroup="group", NodeGroup="group",

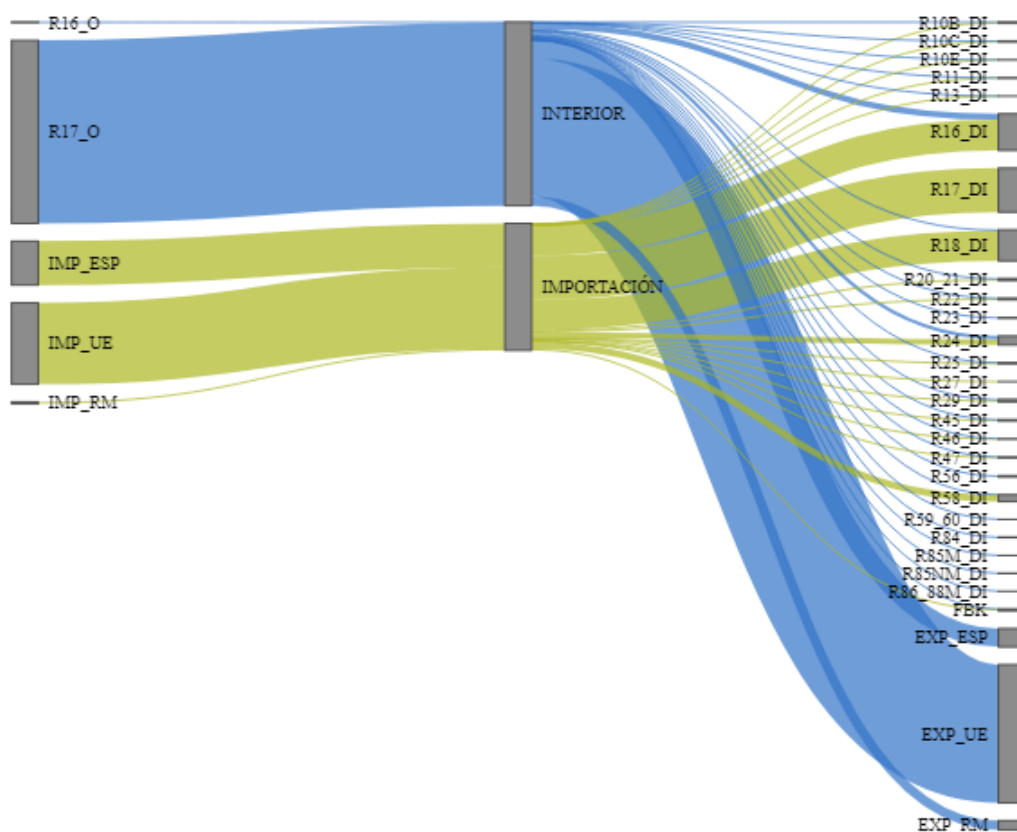
fontSize = 10,

nodePadding = 10, sinksRight=TRUE,iteration=0)

En el Gráfico 1 se puede observar el diagrama de Sankey obtenido para el caso del producto 17A.

El paquete networkD3 crea automáticamente gráficos interactivos, lo que permite en la visualización web arrastrar nodos y colocar el cursor sobre los enlaces para obtener más información.

Gráfico 1. Flujo de oferta y demanda del producto 17A Pasta de papel, papel y cartón



Fonte: IGE. Marco Input-Output de Galicia 2016

Información por rama de actividad

En cuanto a la información por rama de actividad; en la aplicación que se desarrolló se incluyen tablas y gráficos para esa rama de las cuentas de producción y explotación, de las principales producciones, de los principales consumos intermedios, del empleo así como de algunos ratios derivados de las cuentas de producción y explotación y del empleo de la rama.

Como novedad en este apartado es necesario destacar la utilización de los gráficos mapas de árboles jerárquicos o treemaps.

Este tipo de gráficos permiten la comparación de cantidades y la visualización de patrones de alguna estructura jerárquica mediante un uso eficiente del espacio.

Los treemaps son uno de los tipos de visualización más empleados en las presentaciones de Business Intelligence (BI); no obstante no deben utilizarse cuando el objetivo es realizar comparaciones exactas entre los componentes del conjunto de datos. Se utilizan para analizar cómo se divide el todo e identificar rápidamente las componentes más grandes y más pequeñas.

Estos gráficos están compuestos de rectángulos anidados cuyas áreas son proporcionales a los datos que representan.

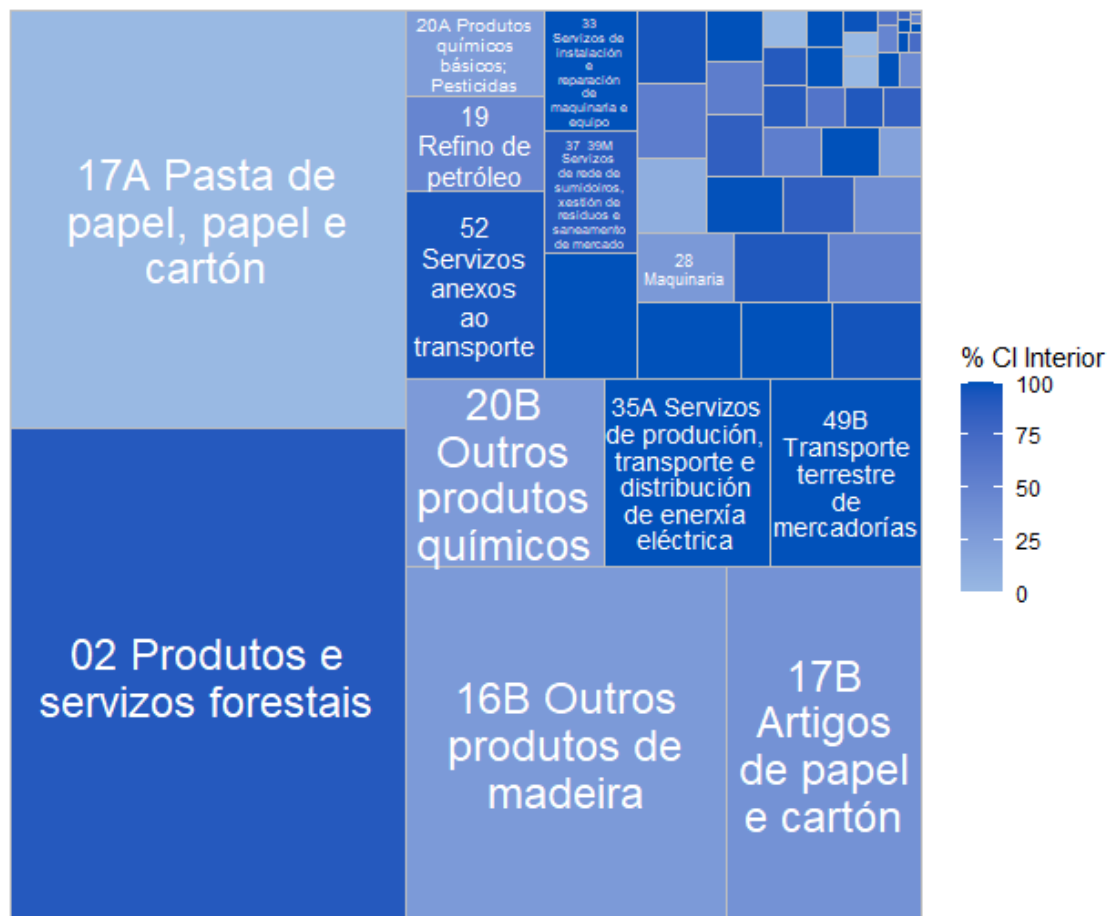
Cuando el conjunto de datos está estructurado en un orden jerárquico (con raíces, ramas y nodos) los rectángulos más grandes indican la raíz o las ramas principales del árbol de datos y se subdividen en rectángulos más pequeños que indican los tamaños de los nodos que tiene cada rama del árbol.

Además, un uso inteligente del color nos permite utilizar los treemaps para representar un segundo valor categórico o cuantitativo, no representado por el área del rectángulo.

En el contexto del Marco Input-Output, los gráficos treemap se utilizaron para visualizar los consumos intermedios de cada rama de actividad. En este caso, el tamaño de los rectángulos está asociado con el peso del consumo intermedio de cada producto en la rama de actividad y, por otra parte, el color se utilizó para diferenciar los productos según el porcentaje de consumo intermedio interior; es decir, se representan para cada rama de actividad dos variables: el peso del consumo intermedio por producto y el porcentaje de consumo intermedio interior de cada producto.

En el Gráfico 2 se puede visualizar el treemap que se obtuvo para el caso de la rama 17 Industria del papel.

Gráfico 2. Principales consumos intermedios de la rama



Para implementar este tipo de gráficos se utilizó el paquete `treemapify`, que permite crear treemaps en `ggplot2`. Es necesario utilizar `geom_treemap()` y especificar dentro de `aes` las variables.

En este caso, en `aes` se especificaron las dos variables a representar: en `area` el peso del consumo intermedio de un producto en la rama de actividad y en `fill` el porcentaje de consumo intermedio interior del producto. La sintaxis que se utilizó fue la siguiente:

```
ggplot (df, aes (area = Peso, fill = `"% CI Interior`,label = Codigo_completo)) +  
  
geom_treemap() +  
  
geom_treemap_text (colour = "white", place = "centre", grow = FALSE, reflow  
=TRUE, min.size = 5) +  
  
scale_fill_gradient (low = '#99b9e3', high = '#0051ba')
```

donde `df` es un data frame formado por 3 columnas: producto, peso de ese producto en el consumo intermedio de la rama y porcentaje del consumo intermedio de ese producto que es interior. La información contenida en el data frame es relativa en este caso a una rama de actividad.

El algoritmo que `geom_treemap()` utiliza por defecto para colocar los rectángulos por defecto es el “squarified”, que se basa en una estrategia que busca que cada bloque sea lo más cuadrado posible para facilitar la comparación entre ellos. En este algoritmo, la colocación de los rectángulos procede de una esquina, colocando los rectángulos en filas o columnas hasta que se colocan todos.

Conclusión

El Marco Input-Output de Galicia (MIOGAL) constituye una herramienta compleja pero fundamental para describir los flujos económicos de nuestra región. Tiene una doble vertiente: descriptiva y analítica. Siempre ha primado la analítica, como base del análisis input-output, pero no es despreciable y debe promocionarse la vertiente descriptiva.

Cada vez son más frecuentes las situaciones en las que nos encontramos con grandes cantidades de datos y cada vez se hace también más patente la necesidad de apoyarse en distintas herramientas que permitan facilitar la comprensión de la información. Se hace necesario favorecer la visualización de grandes cantidades de datos de manera sencilla y de un solo vistazo.

La incorporación en la difusión del Marco Input-Output de los gráficos Sankey y de los treemap no deja de ser un nuevo paso en la búsqueda de herramientas que ayuden a una mejor comunicación y comprensión de los resultados estadísticos.

No obstante, este tipo de visualizaciones tampoco son siempre fáciles de leer por lo que a veces es necesario evaluar si un gráfico de barras o un gráfico circular cuenta lo mismo aunque visualmente sea menos intenso.

Referencias

Veiguela, N. (2016). R-Shiny: Una herramienta para mejorar la difusión de las operaciones del sistema de cuentas económicas de Galicia. XIX Jornadas de Estadística de las Comunidades Autónomas. Madrid

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>

Winston Chang, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert and Barbara Borges (2021). shiny: Web Application Framework for R. R package version 1.6.0. <https://CRAN.R-project.org/package=shiny>

David Wilkins (2021). treemapify: Draw Treemaps in 'ggplot2'. R package version 2.5.5. <https://CRAN.R-project.org/package=treemapify>

J.J. Allaire, Christopher Gandrud, Kenton Russell and CJ Yetman (2017). networkD3: D3 JavaScript Network Graphs from R. R package version 0.4. <https://CRAN.R-project.org/package=networkD3>

IGE : Marco input-output de Galicia.

https://www.ige.eu/web/mostrar_actividade_estadistica.jsp?idioma=gl&codigo=0307007003

IGE: Aplicación gráfica de consulta do Marco Input-Output de Galicia 2016. <http://www.ige.eu/Shiny/MIOGAL/>