

## **VISUALIZACIÓN DE DATOS EN LAS ESTADÍSTICAS OFICIALES**

**Enara Galbete Ahechu**



**EUSKAL ESTATISTIKA ERAKUNDEA**  
**INSTITUTO VASCO DE ESTADISTICA**

Donostia - San Sebastián, 1  
01010 VITORIA - GASTEIZ  
Tel.: 945 01 75 00  
Fax.: 945 01 75 01  
E-mail: [eustat@eustat.es](mailto:eustat@eustat.es)  
[www.eustat.es](http://www.eustat.es)

---

# Presentación

Eustat tiene entre sus objetivos la divulgación de la información estadística de forma clara y comprensible. Para ello, está realizando un esfuerzo en la investigación en torno a la visualización de datos estadísticos con el fin de que puedan ser analizados y manejados de forma dinámica e interactiva por los usuarios.

En 2010 organizó el XXIII Seminario Internacional de Estadística bajo el lema "Examinando datos con gráficos interactivos dinámicos" y convocó una beca de 2 años de duración en formación e investigación de visualización de datos estadísticos.

Esta publicación da a conocer el trabajo de investigación realizado en la beca y ofrece herramientas útiles para aquellos usuarios interesados en conocer y emplear diversas formas de visualización de datos.

Este documento consta de tres apartados diferentes. En el primero de ellos, se habla sobre ciertos conceptos y criterios a tener en cuenta antes de hacer un gráfico. En el segundo apartado, se presenta el catálogo de gráficos estáticos y dinámicos. Para finalizar, en el tercer apartado se explican los procedimientos y las aplicaciones de los gráficos seleccionados para su publicación en la página web de Eustat y de los elegidos para los procesos de producción.

Vitoria-Gasteiz, Enero de 2013.

Josu Iradi Arrieta

Director General de EUSTAT

# Índice

PRESENTACIÓN .....	2
ÍNDICE .....	3
INTRODUCCIÓN.....	4
METODOLOGÍA .....	5
DIFERENTES TIPOS DE DATOS.....	5
CONCEPTUALIZACIÓN.....	6
POBLACIÓN.....	6
ESPACIO DE OBSERVACIÓN: .....	7
VARIABLES.....	9
CRITERIOS A TENER EN CUENTA ANTES DE REPRESENTAR UN GRÁFICO .....	9
CLASIFICACIÓN DE LOS GRÁFICOS .....	12
DATOS UNIVARIANTES.....	12
DATOS BIVARIANTES .....	14
DATOS DE MULTIVARIANTES .....	16
DATOS TEMPORALES .....	20
DATOS ESPACIALES .....	22
APLICACIONES .....	23
GRÁFICOS PARA PRODUCCIÓN .....	23
1. Dotplot o diagrama de puntos .....	23
2. Gráfico compuesto de diagramas de barras .....	25
3. Representación gráfica del análisis de correspondencias.....	26
4. Regresión lineal por tramos, por sectores. ....	28
5. Gráfico Scatterplot Matrix .....	30
GRÁFICOS PARA DIFUSIÓN .....	32
1. Gráfico Motion Chart.....	34
2. Gráfico de tipo Time Line.....	40
3. Visualización mediante Google Earth.....	43
CONCLUSIONES.....	48
BIBLIOGRAFÍA.....	49

# Introducción

Los datos recogidos en este Cuaderno Técnico son fruto del trabajo realizado en la beca sobre Visualización de Datos concedida en 2010 por el Instituto Vasco de Estadística para la formación e investigación en el campo de las metodologías estadístico-matemáticas.

Se convocó una beca de 2 años de duración para la formación e investigación en el campo de las metodologías estadístico-matemáticas de la producción estadística de tipo oficial, concretamente en torno a la visualización de datos estadísticos.

Este documento está dividido en los siguientes apartados:

En el primer capítulo, además de la introducción, se recoge también la exposición de los objetivos establecidos para llevar a cabo este cuaderno técnico.

En el segundo capítulo, se define una serie de criterios y conceptos a tener en cuenta antes de realizar un gráfico partiendo de un conjunto de datos.

En el tercer capítulo, se presenta una clasificación de los gráficos diseñados para reflejar diferentes tipos de datos.

El capítulo cuarto versa sobre los gráficos automatizados para su uso por el EUSTAT. Unos son gráficos estáticos para la producción; otros, en cambio, son gráficos dinámicos e interactivos diseñados para su difusión.

Por último, se exponen una serie de conclusiones relacionadas con los métodos y técnicas de visualización de datos.

Desde aquí, me gustaría agradecer el trabajo realizado por todos los miembros del Área de Metodología, Innovación e I+D de Eustat y por todos los miembros del grupo de visualización, también los consejos y la ayuda prestada por Yosu Yurramendi Mendizabal del Departamento de Ciencias Computacionales e Inteligencia Artificial de la Facultad de Informática de la UPV y, en general, la amabilidad de todos los trabajadores de Eustat.

**PALABRAS CLAVE:** gráficos estáticos, gráficos dinámicos, gráficos interactivos.

## Metodología

EUSTAT dispone de multitud de datos de diversa naturaleza y procedencia. Existen diferentes formas de analizar y presentar los datos, por ejemplo el uso de gráficos. Así pues, el objetivo principal de la visualización de datos no es otro que el análisis, síntesis y representación de datos a simple vista, utilizando para ello gráficos de diversa dificultad y tipología.

A cada tipo de datos se le asigna un tipo de visualización, que brinda la oportunidad de extraer la información de forma gráfica y bastante intuitiva dependiendo de la dificultad que entrañen los datos y el gráfico.

### Diferentes tipos de datos

EUSTAT dispone de dos tipos de datos: los que figuran en la web, disponibles en formato PC-AXIS; y los archivos de microdatos.

Antes de ser representados de manera gráfica, los datos deben ser clasificados en base a estos dos criterios:

#### 1. CARÁCTER DE LOS DATOS:

Dependiendo de su carácter, es decir, del número de variables que tengan, los datos pueden ser clasificados en 5 grupos distintos:

- i. **Datos univariantes.**
- ii. **Datos bivariantes.**
- iii. **Datos multivariantes.** Datos con tres o más variables
- iv. **Datos temporales.** Son secuencias de valores, observaciones o datos, medidos en momentos temporales concretos, ordenados cronológicamente y normalmente uniformemente espaciados entre sí.
- v. **Datos espaciales.** Variables relacionadas con ubicaciones espaciales concretas.

#### 2. PROCEDENCIA DE LOS DATOS:

En base a su procedencia, los datos pueden dividirse en dos tipos:

- i. **Tablas de frecuencia.** Se asigna a cada dato su correspondiente frecuencia.
- ii. **Microdatos.** A cada individuo le corresponde una relación de variables.

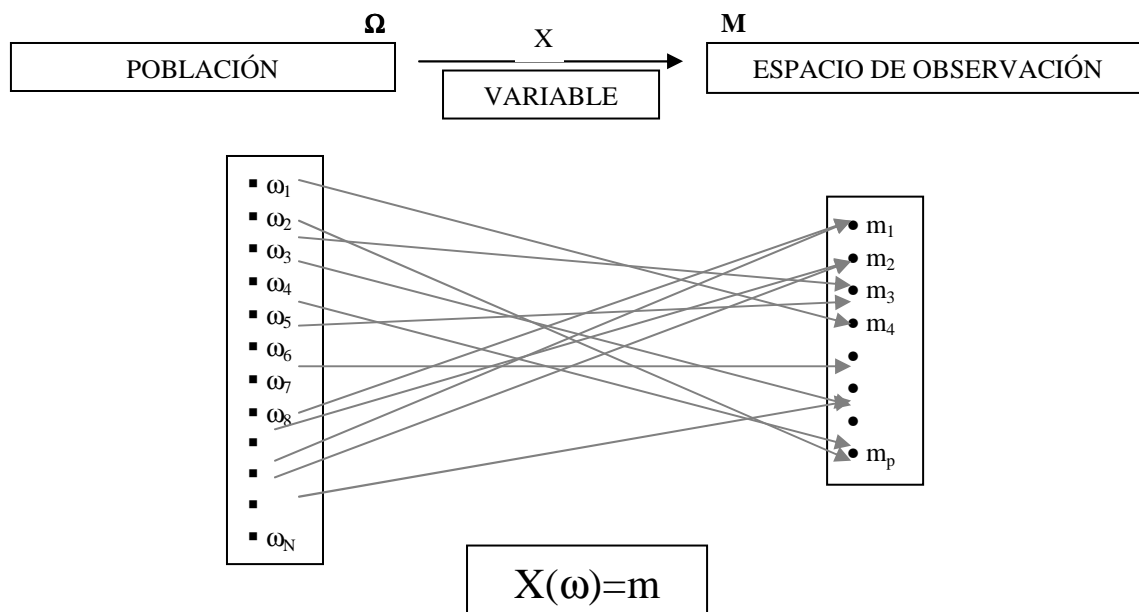
## Conceptualización

Son tres los conceptos básicos: **población**, espacio de **observación** y **variable**.

**Población:** El conjunto de todos los elementos a analizar estadísticamente.

**Espacio de observación:** El conjunto de todos los posibles resultados de un experimento realizado de manera aleatoria.

**Variable:** La aplicación existente entre la población y el espacio de observación.



Tanto la población como el espacio de observación pueden corresponder a diferentes tipos, aspecto este que ha de ser tenido en cuenta a la hora de construir un gráfico, puesto que también de ellos depende el carácter de las variables.

## Población

Existen diferentes tipos de **población** según sea su **carácter**, **tamaño** y **estructura**.

### 1. Clasificación según **carácter**:

- i. **Padrón o registros.** Se denomina padrón o registros a la numeración de los individuos que conforman una población estadística. Consisten, fundamentalmente, en medir la cantidad total de individuos de una población estadística, mediante el empleo de diferentes técnicas de enumeración o utilizando información administrativa.
- ii. **Muestra.** Es un subconjunto de la población, obtenido a partir de técnicas de muestreo y que es representativo del total de la población.

### 2. Clasificación según **tamaño**:

- i. **Discreta.** Dícese de la población que cuenta con una cantidad de elementos finita o infinita contable.
    - **Finita.** Dícese de la población que posee una cantidad de elementos finita.
    - **Infinita.** Dícese de la población que cuenta con una cantidad de elementos infinita contable.
  - ii. **Continua.** Población cuya variable aleatoria mide una característica continua.
3. Clasificación según **estructura**:

Los elementos de la población, las unidades del conjunto, pueden estar relacionadas o no. Según el tipo de relación la población tendrá distinta estructura.

- i. **Sin estructura.** Se da cuando los elementos de la población no están relacionados. Este es el caso más común en la estadística clásica. (Por ejemplo, *al hacer una muestra se supone que los elementos, es decir las personas, no tienen ninguna relación entre sí*).
- ii. **Estructura secuencial.** Cuando está definido un orden total entre los elementos de la población. Se da cuando en un proceso a un elemento concreto se le toman ciertas medidas y cuando mediante las medidas se le causa algo. (Por ejemplo, *terapia realizada paso por paso a un paciente, pasos del método de estudio de un alumno...*).
- iii. **Estructura lineal.** El caso más habitual es el de una serie temporal, las unidades van unas tras otras, y además, en este caso se define una unidad.
- iv. **Mapa.** En dos dimensiones (o en más) un grafo o mapa define una estructura (el vértice es la población, y las aristas definen la relación).
- v. **Plano.** En dos dimensiones, y cuando hay definida una unidad la población se sitúa en el plano  $R^2$ . (Por ejemplo, *los píxeles de una foto forman un grafo y además hay definida una medida entre los píxeles,  $R^2$* ).

## Espacio de observación:

Existen diferentes tipos de **espacios de observación** según sea su carácter, tamaño y estructura.

1. Clasificación según **carácter**:
  - i. **Cuantitativo.** Espacio de observación que consta solamente de sucesos básicos cuantitativos.

- ii. **Cualitativo.** Espacio de observación que consta solamente de sucesos básicos cualitativos.

2. Clasificación según **tamaño**:

- i. **Discreto.** Dícese de aquel espacio de observación que cuenta con una cantidad finita o infinita contable de sucesos básicos.
- **Finito.** Espacio de observación que cuenta con una cantidad finita de observaciones.
  - **Infinita.** Espacio de observación que cuenta con una cantidad infinita contable de observaciones.
- ii. **Continua.** Dícese de aquel espacio de observación que cuenta con una cantidad de sucesos básicos infinita e incontable.

3. Clasificación según **estructura**:

- i. **Nominal.** Espacio de observación compuesto por observaciones que carecen de orden entre sí y no obedecen a una jerarquía.
- ii. **Ordinal.** Espacio de observación compuesto por observaciones que pueden tomar diferentes valores ordenados, procedentes de una progresión natural, una secuencia o una escala establecida con anterioridad.
- **Total.** Cuando el espacio de observación es un conjunto totalmente ordenado. Es decir, cuando se puede comparar entre sí cualquier par de elementos del conjunto.
  - **Parcial.** Cuando el espacio de observación es un conjunto parcialmente ordenado. Es decir, cuando en varios pares de elementos del conjunto uno de los elementos sigue al otro. Pero no todos los pares de elementos del conjunto tienen que estar relacionados.
- iii. **Numeral.** Espacio de observación compuesto por observaciones numerales.

<b>Carácter</b>	Cualitativo	Cuantitativo
<b>Estructura</b>		
Nominal	√	
Orden parcial	√	
Orden total	√	√
Numeral	*	√

**Tamaño finito**

\*  $\{m_1, m_2\} \leftrightarrow \{0, 1\}$



## Variables

El carácter de una **variable** se establece en base al carácter, el tamaño y la estructura de su **espacio de observación**.

Según su espacio de observación, podemos diferenciar los siguientes tipos de variables.

1. **Cualitativas.** Son aquellas variables que reflejan diferentes cualidades, características o modalidades. A cada modalidad reflejada se la denomina atributo o categoría, y la medición consiste en clasificar los citados atributos.
  - i. **Nominales.** Variables que toman valores que no guardan entre sí ningún tipo de orden o jerarquía (por ejemplo, *lugar de nacimiento*).
  - ii. **Ordinales.** Aquellas que pueden tomar diferentes valores ordenados de una progresión natural, una secuencia, o una escala fijada con anterioridad (por ejemplo, *variable con modalidades: pequeña, mediana, grande*).
2. **Cuantitativas** Aquellas que pueden ser expresadas mediante números (por ejemplo, *edad, tamaño de la unidad familiar, superficie de la vivienda, etc.*). Se pueden cuantificar los resultados de los experimentos.
  - i. **Discretas.** Solamente toman valores enteros (por ejemplo, *edad, tamaño de la unidad familiar, etc.*).
  - ii. **Continuas.** Pueden tomar cualquier valor comprendido dentro de un intervalo, por lo que pueden tener decimales (por ejemplo, *superficie de la vivienda*).

## Criterios a tener en cuenta antes de representar un gráfico

Para que los gráficos sean "buenos" y que de su interpretación se pueda obtener la mayor información posible, es conveniente tener en cuenta una serie de criterios. W.S. Cleveland, por ejemplo, recogió lo siguiente en su libro "The Elements of Graphing Data" (1985):

- Vista o perspectiva limpia:

Datos: hay que desprestigiar todo lo que no sea necesario.

Utilizar elementos gráficos claros y manifiestos para mostrar los datos.

Utilizar un par de escalas rectas por cada variable. El espacio de datos será el interior del ángulo recto formado por las rectas de las escalas. Establecer las marcas de las escalas fuera del espacio de datos.

No desordenar el espacio de datos.

No poner demasiadas marcas en las escalas.

Utilizar una recta de referencia cada vez que haya un valor que debe atravesar el gráfico, pero sin mezclarla con los datos.

No mezclar las etiquetas de los datos en el espacio de datos cuantitativos.

Evitar el uso de notas, llaves y señales en el espacio de datos. Colocar las llaves y las señales fuera del espacio de datos, y las notas en la leyenda o en el texto.

Los signos dibujados, en caso de superponerse, deben ser claramente visibles.

Los conjuntos de datos, en caso de superponerse, deben poder distinguirse a simple vista.

En caso de reducción o reproducción del gráfico, no debe perderse la claridad de representación.

- Comprensión clara.

Mostrar de manera gráfica las principales conclusiones. Hacer leyendas informativas comprensibles.

Explicar de forma clara el significado de aquellas barras de error que reflejan la variabilidad de los datos (desviación estándar, error estándar, margen de confianza).

Al dibujar los logaritmos de una variable, las etiquetas de las escalas deben coincidir con las etiquetas de las marcas.

Corregir los errores de los gráficos.

Cuidar la limpieza.

- Escalas.

Seleccionar el espacio entre marcas de tal manera que abarque el intervalo de todos o de casi todos los datos.

Mantener las reducciones de la escala y seleccionarla de tal manera que comprenda todo el conjunto de datos.

En ocasiones, para mostrar dos escalas diferentes suele resultar de utilidad emplear dos rectas de escala para una misma variable.

Seleccionar las escalas adecuadas cuando haya que comparar dos gráficos entre sí.

No hay que empeñarse en mostrar siempre el cero en aquellas escalas que representan una magnitud.

Utilizar la escala logarítmica cuando sea importante comprender la variación porcentual del factor multiplicador o corregir el sesgo de grandes valores.

El hecho de mostrar los datos en una escala logarítmica puede ayudar a diferenciar mejor los datos.

Cortar la escala solamente cuando sea necesario. En caso de que el corte sea inevitable, utilizar un corte total de escala. No unir los valores numéricos situados a ambos lados del corte.

- Estrategia general:

Se puede introducir mucha información cuantitativa en un espacio reducido.

Dibujar gráficos puede convertirse en un proceso iterativo o experimental.

Cuando sea necesario, hacer el gráfico de los datos más de una vez.

Muchos gráficos útiles precisan de un análisis exhaustivo y responsable.

## Clasificación de los gráficos

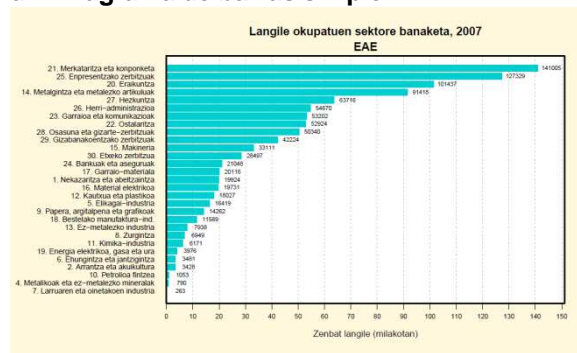
La siguiente clasificación obedece al carácter y procedencia de los datos:

### Datos univariantes

- **Tablas:**

- i. **Diagrama de barras**

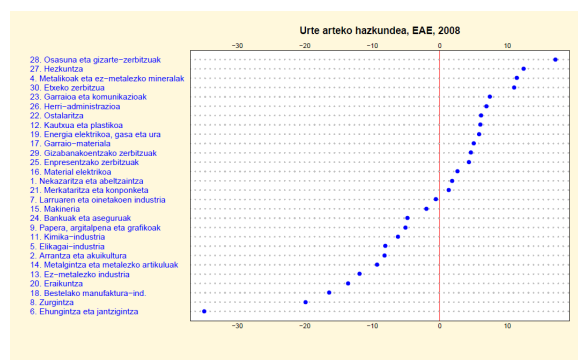
- a. **Diagrama de barras simple**



Se utiliza cuando se quieren mostrar las frecuencias de las modalidades de una variable. Dichas frecuencias se representan mediante barras paralelas, una sobre otra y ordenadas de mayor a menor tamaño. A fin de facilitar

su interpretación, se añade a la derecha de cada barra la frecuencia de la modalidad que representa, así como varias rectas verticales. Se puede cambiar el color de las barras y las escalas utilizadas.

- ii. **Diagrama de puntos o dotplot.**

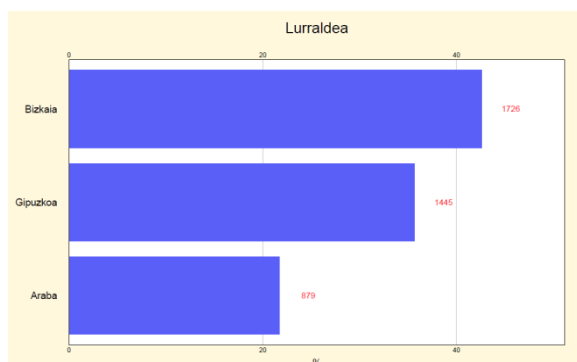


Se utiliza cuando se quieren representar las variaciones acontecidas en las diferentes modalidades de una variable. Según su variación, las modalidades se ordenan de mayor a menor y sus valores se representan mediante puntos (se utiliza un punto, en lugar de dibujar una barra). La línea roja dibujada en el 0 permite ver con rapidez si la

variación es positiva o negativa. Además, las rectas discontinuas verticales facilitan la interpretación.

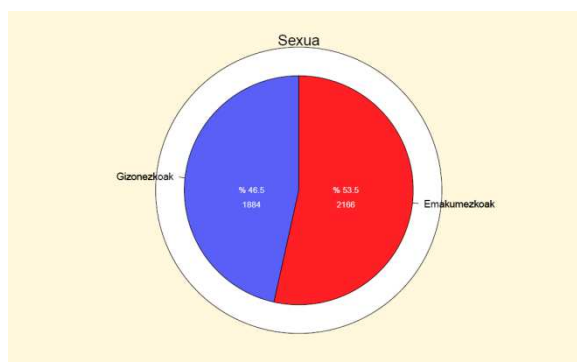
- **Microdatos:**

- i. **Diagrama de barras simple.**



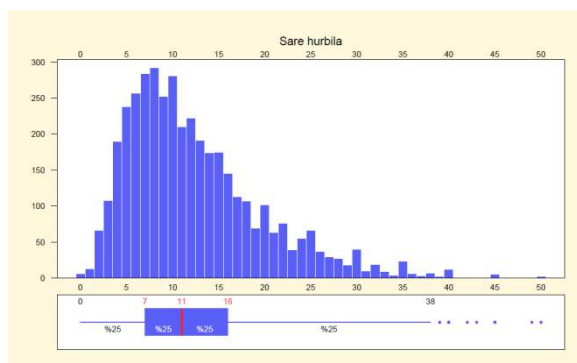
Es similar al diagrama de barras utilizado en las tablas. En este caso, cuando las modalidades tienen un orden propio, este se mantiene y, en caso de que no lo tengan, se ordenan de mayor a menor. Cuando la variable es continua, se eliminan los espacios entre las barras.

- ii. **Diagrama de sectores.**



Se utiliza para representar variables discretas que constan de pocas modalidades. Se asigna un color a cada sector (y, por tanto, a cada modalidad), especificando su correspondiente frecuencia y porcentaje.

- iii. **Histograma y diagrama de caja.**



Se utilizan para representar variables cuantitativas continuas. Se dibuja un diagrama de caja horizontal, en el que una línea roja gruesa representa la mediana. También en rojo figuran los datos relativos a los cuartiles y, para facilitar su interpretación, se añade información acerca del significado de la caja y de los

whiskers o ramas. Resulta adecuado para visualizar la presencia de valores atípicos y para analizar la distribución y simetría de las variables.

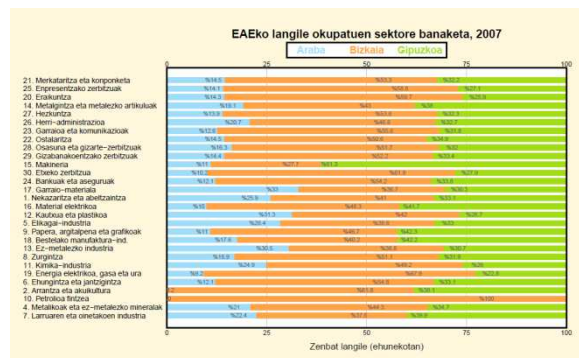
Sobre el diagrama de barras, se dibuja en color azul el histograma correspondiente a los datos, que nos ayudará a hacernos una idea sobre la distribución de la variable objeto de análisis.

## Datos bivariantes

- Tablas:

- i. Diagrama de barras

- a. Diagrama de barras porcentual o de frecuencias relativas.

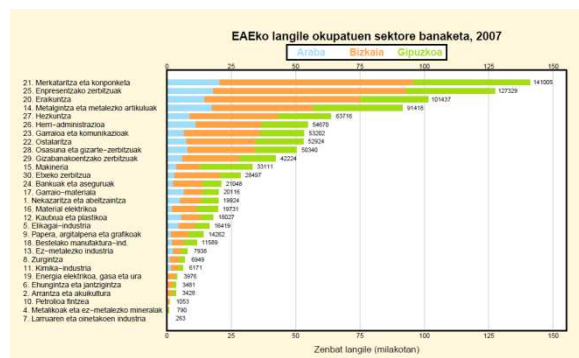


Se utiliza para representar las frecuencias de las modalidades de una variable cualitativa, respecto a los tres territorios históricos. Primero, se ordenan las modalidades de mayor a menor, conforme a su frecuencia en la C.A. de Euzkadi. Respetando

dicho orden, se calculan los porcentajes de los tres territorios y se dibujan a modo de barras, de manera contigua y en diferentes colores, completando todas ellas la barra que representa el 100% de la frecuencia total. Como ocurre con los diagramas de barras simples, a estos también se les añaden rectas verticales que facilitan su interpretación.

Es posible, en vez de utilizar territorios históricos, utilizar una variable cualitativa con cuatro modalidades siempre que una de ellas se pueda desglosar en las demás.

- b. Diagrama de barras categorizado o diagrama de barras de frecuencias absolutas.

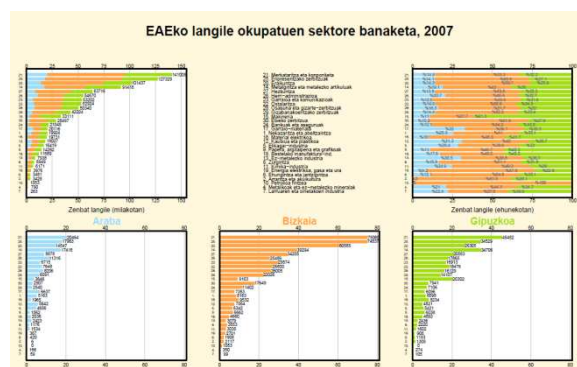


Se utiliza, también, para representar las frecuencias de las modalidades de una variable cualitativa, respecto a los tres territorios históricos. No obstante, en este caso el tamaño de las barras es proporcional la frecuencia que tiene la modalidad correspondiente en la

C.A. de Euzkadi. Tal y como ocurría en el caso anterior, los datos se ordenan según la frecuencia que tiene la modalidad correspondiente en la CAE, y se crea un diagrama de barras simple. Después, cada barra se divide en segmentos proporcionales a la frecuencia de los tres territorios, aplicando a cada uno de ellos el color que le corresponda. Al igual que sucedía en los gráficos analizados con anterioridad, se añaden líneas verticales que facilitan la interpretación.

También es posible, en vez de utilizar territorios históricos, utilizar una variable con cuatro modalidades siempre que una de ellas se pueda desglosar en las demás.

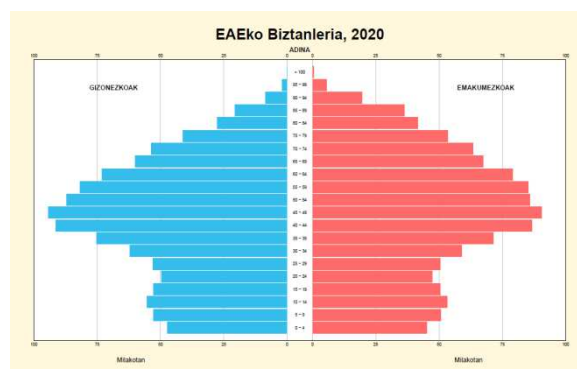
### c. Gráfico compuesto de diagramas de barras.



Se trata de una representación gráfica compuesta por un diagrama de barras categorizado y un diagrama de barras porcentual de la C.A. de Euskadi y de diagramas de barras simples de cada territorio histórico.

## ii. Pirámide.

### a. Pirámide simple.

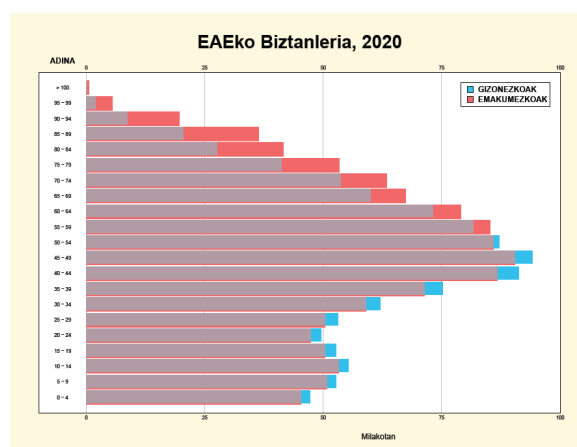


Gráficamente se trata de un doble histograma de frecuencias. Las barras del doble histograma se disponen horizontalmente, es decir, sobre la línea de las abscisas, y entre los dos histogramas se indican los grupos de edad. Los datos correspondientes a los hombres se

representan en azul, en la parte izquierda del doble histograma y los correspondientes a las mujeres en rojo y en la parte derecha. Se añaden rectas verticales para facilitar la interpretación. Se puede cambiar tanto las escalas como los colores.

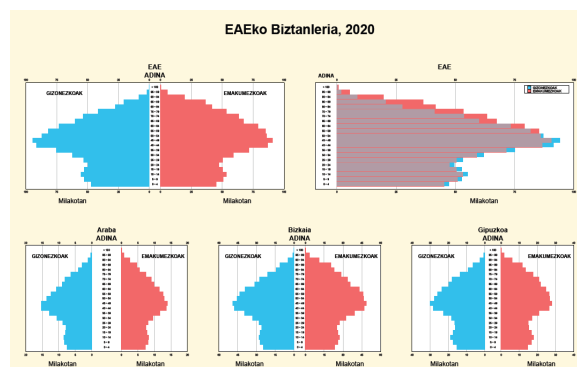
Aunque es un gráfico utilizado tradicionalmente para temas de población, también se puede utilizar para temas económicos, por ejemplo, para representar saldos comerciales de importación y exportación.

### b. Pirámide alternativa.



Es similar a la pirámide simple, pero en este caso las barras correspondientes a los dos sexos se colocan una sobre la otra, facilitando así su comparación de manera sencilla. Al igual que en el anterior se añaden rectas verticales y también se puede utilizar para temas económicos.

### c. Gráfico compuesto de pirámides.



Se trata de una representación gráfica formada por una pirámide simple y una pirámide alternativa de la C.A. de Euskadi y por pirámides simples de cada territorio histórico.

- Microdatos:

#### i. Regresión lineal por tramos.

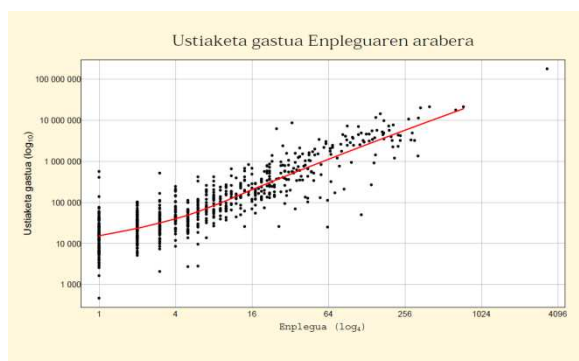


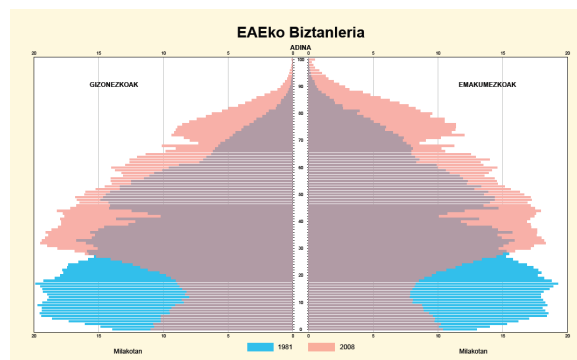
Gráfico utilizado para analizar si la relación existente entre dos variables cuantitativas es lineal por tramos. Cada dato se representa en el gráfico mediante un punto y se ajusta una recta de regresión por tramos (marcada en rojo). Si los datos son muy asimétricos, hay que modificarlos con antelación (por ejemplo, con logaritmos).

## Datos de multivariantes

- Tablas

#### i. Pirámides

##### a. Bipirámide.

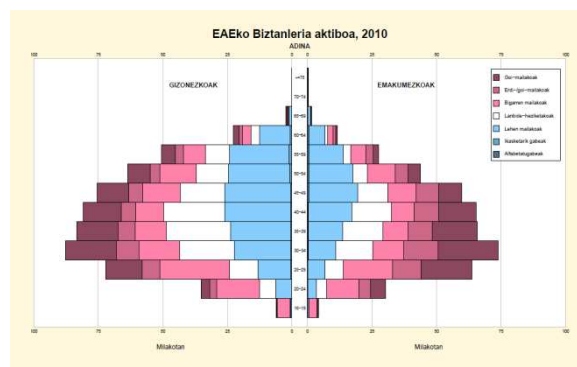


Representa, mediante dos pirámides, los datos de las poblaciones correspondientes a dos años, según la edad y el sexo. En concreto, dibuja la pirámide correspondiente a un año sobre la correspondiente al otro, cada una de un color.

El representar una pirámide sobre la otra y la transparencia de los colores, permiten comparar las poblaciones de dos años distintos, para cada grupo de edad y para cada sexo.



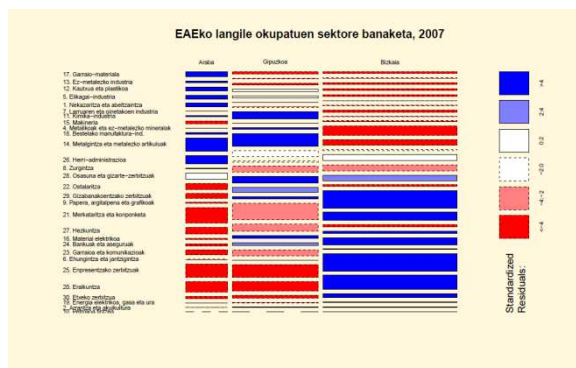
## b. Pirámide categórica.



Es similar a una pirámide simple, para datos multivariantes. En este caso, se representa la población para cada grupo de edad y cada sexo, pero según las modalidades de una variable cualitativa. Precisamente, para cada grupo de edad y cada sexo se representa la

barra que representa el tamaño de la población, dividiéndola en barras proporcionales a cada modalidad de la variable cualitativa y pintando cada una de un color. Al igual que sucede en los gráficos analizados con anterioridad, se añaden líneas rectas verticales que facilitan la interpretación.

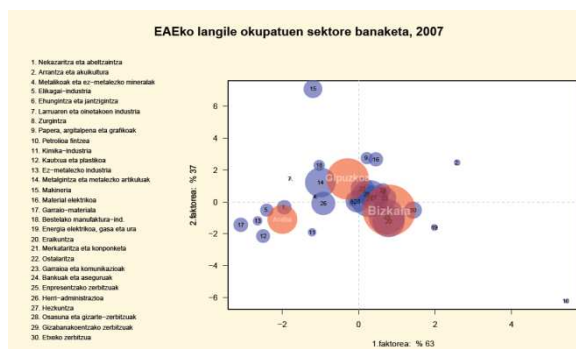
## ii. Diagrama de mosaico bicategorico.



Se emplea para analizar la relación existente entre dos variables cualitativas. La superficie de cada rectángulo representa la frecuencia de la modalidad de aquella variable que representa. La altura, en cambio, refleja la frecuencia de la modalidad de la otra variable. Los colores de los rectángulos, en cambio, reflejan el valor y el signo de

los restos. Es decir, el color viene a reflejar si las frecuencias de las celdas están por encima o por debajo de las frecuencias esperadas según el modelo proporcionado.

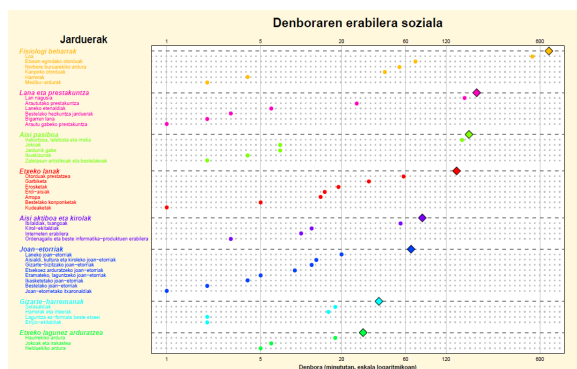
## iii. Representación gráfica del análisis de correspondencia.



El análisis de correspondencia es un método gráfico para analizar las relaciones existentes entre las variables categóricas de una tabla de contingencias. La cercanía (o lejanía) de un punto respecto a uno o los dos ejes muestra la similitud o distancia (positiva o negativa) para con la media de las modalidades

definidas en los ejes. Así, los puntos tenderán a estar más cerca (o lejos) de aquellas modalidades con las que más (o menos) se asemejen. El tamaño de las bolas, en cambio, indica la frecuencia. Del mismo modo, se utilizan dos colores, uno para la modalidad de las filas y otro para la de las columnas.

#### iv. Gráfico dotplot de Cleveland.



Originalmente, se trata de un diagrama de puntos que tiene múltiples variables, donde se observan el orden parcial y total de las variables. Se utiliza cuando la variable cualitativa a representar consta de varias modalidades y sub-modalidades. Por un lado, se completa el diagrama de puntos correspondiente a las modalidades; por otro, se

hace otro diagrama de puntos correspondiente a las sub-modalidades de cada modalidad.

Cuando la frecuencia de una modalidad sea muy superior a la del resto, puede que haya que utilizar una escala logarítmica para poder ver con claridad las diferencias respecto al resto de modalidades.

#### • Microdatos

##### i. Regresión lineal por tramos, por sectores.

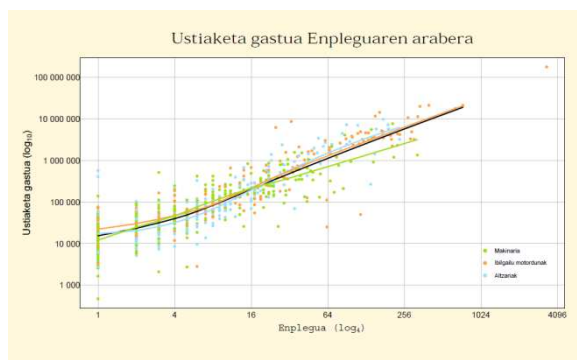
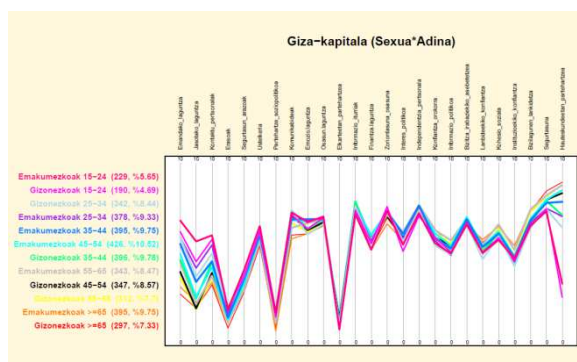


Gráfico utilizado para analizar si la relación existente entre dos variables cualitativas es lineal por tramos, según las modalidades de una tercera variable cualitativa. Cada dato se representa en el gráfico mediante un punto y posteriormente, se ajustan una recta de regresión por tramos general (en negro) y una recta de regresión por

tramos para cada modalidad de la variable cualitativa (cada una en su correspondiente color).

En ocasiones, por ejemplo, cuando las variables son muy asimétricas, resulta conveniente transformarlas antes de comenzar (por ejemplo, tomando logaritmos)

##### ii. Gráfico de coordenadas paralelas.

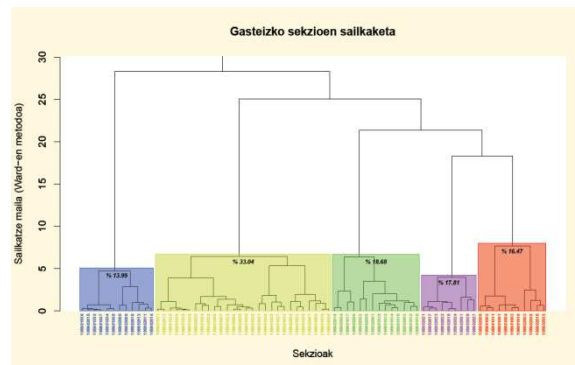


El objetivo del gráfico de coordenadas paralelas es representar conjuntos de datos de varias variables. Cada variable es representada como una escala vertical paralela a todas las demás. Para representar las observaciones, se dibuja en cada eje su punto correspondiente, utilizando un

color para cada modalidad. Luego, se unen todos los puntos de un mismo color con

líneas rectas. A la hora de representar un punto en el eje, se tiene en cuenta el valor que toma la observación en aquella variable representada en el eje.

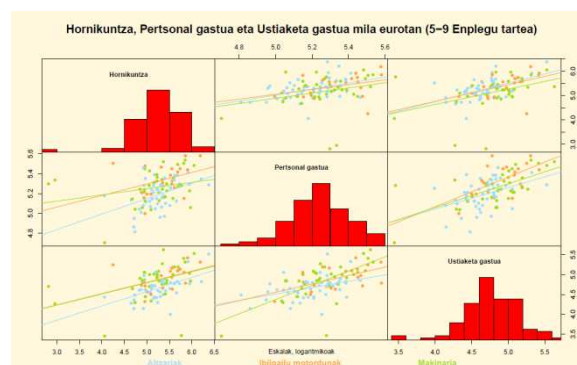
### iii. Dendrograma.



Se trata de una representación gráfica o diagrama en forma de árbol (dendro = árbol) que organiza los datos en sub-categorías. Estas, a su vez, se van subdividiendo hasta llegar al nivel de especificidad deseado. El objetivo es ordenar los datos por grupos, de tal manera que los datos de un mismo clúster o grupo

sean lo más parecidos entre sí y lo más diferentes posible respecto a los de otros clústeres. Esto puede ser útil para exponer las estructuras y uniones existentes entre los datos, aspectos que aun siendo desconocidos en un principio pueden resultar de utilidad. Para facilitar su comprensión, se da a cada clúster un color diferente y se añade el porcentaje de individuos que posee.

### iv. Gráfico Scatterplot Matrix.



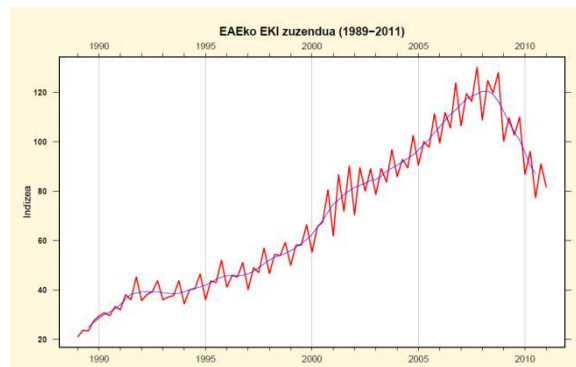
Se basa en una matriz de  $n \times n$  dimensiones (donde  $3 \leq n \leq 5$ ) formada por diagramas bivariantes. En las diagonales se representan histogramas, y en el resto de celdas diagramas de regresión lineal por tramos. Observando las diagonales se puede analizar la distribución y simetría de las variables y en el resto de celdas se pueden observar

diferentes rectas de regresión por tramos creadas al cruzar las diferentes variables.

## Datos temporales

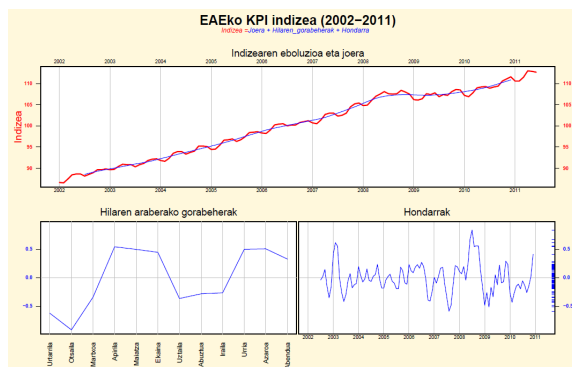
- Tablas

- i. Gráficos de índices.



Se trata de un gráfico que compara el índice (en rojo) con un índice desestacionalizado al que se le ha eliminado el efecto calendario (en azul). Comparando ambos índices se puede observar los crecimientos y decrecimientos que el índice objeto de análisis ha tenido en ciertas épocas del año.

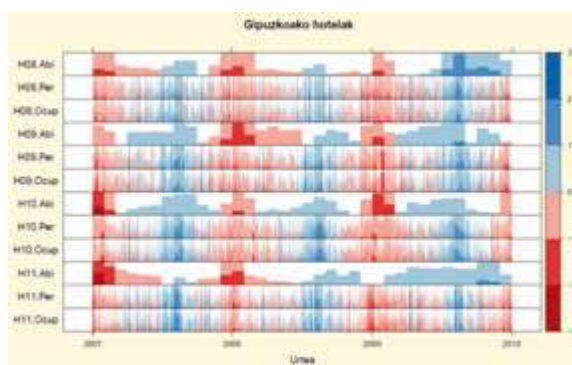
- ii. Gráficos de índices. Ciclos y tendencias.



En este gráfico se comparan también el índice y el índice desestacionalizado, a diferencia de que en este caso, al proporcionarse los datos por meses, se añaden un gráfico mensual y otro residual.

- Microdatos

- i. Gráficos Horizon.



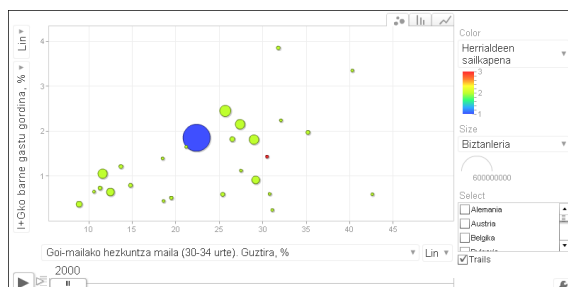
Los gráficos Horizon permiten visualizar al mismo tiempo y en un espacio limitado multitud de series temporales. En definitiva, se basan fundamentalmente en la representación de multitud de gráficos de líneas, utilizando colores para ahorrar espacio y visualizar todas las series al mismo tiempo. Concretamente, se

utiliza el rojo para reflejar valores negativos y el azul para los positivos. Los gráficos Horizon presentan una serie de ventajas ya que permiten, entre otras:

- Observar comportamientos especiales y tendencias que predominan.
- Analizar las series de manera independiente unas respecto a otras.
- Realizar comparaciones entre series.

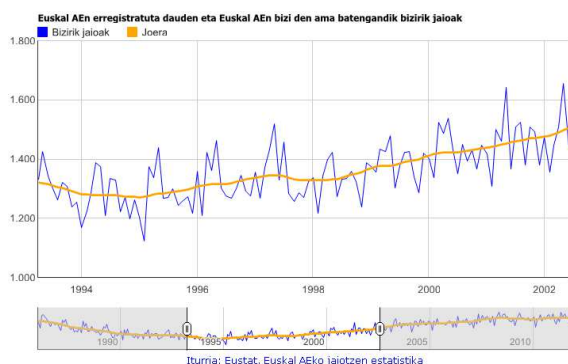
- Observar los cambios con la suficiente precisión como para percatarse de si conviene analizarlos en profundidad.

## ii. Gráficos Gapminder Motion Chart.



El Motion Chart es un gráfico dinámico que brinda la oportunidad de analizar los datos en cinco dimensiones diferentes. El usuario puede escoger qué cinco variables desea analizar. Deberá elegir, por ejemplo: las variables de los ejes X e Y; la variable que medirá el tiempo (la que viene en el archivo inicial); y las variables que determinarán el tamaño y color de las burbujas. Analizando los datos a lo largo del tiempo, este gráfico da la oportunidad de identificar tendencias y modelos que no se pueden apreciar en otros gráficos tradicionales, además de facilitar la interpretación y ofrecer diversas herramientas para jugar con él.

## iii. Gráficos Time Line.



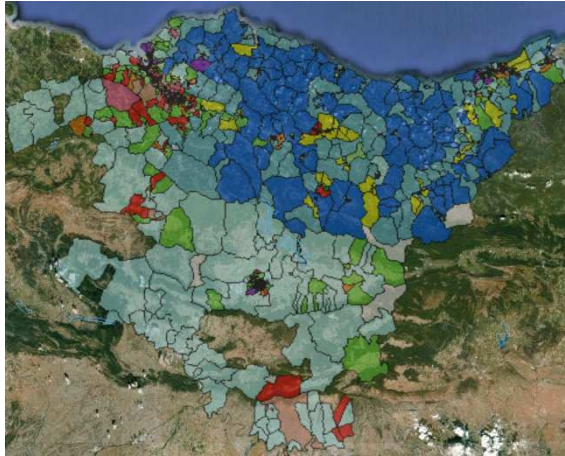
Se basan en un gráfico interactivo para visualizar series temporales. Colocando el cursor sobre cada una de las series que se muestran, aparecerá un cuadro en el que se detalla el valor y la fecha correspondiente a ese punto de la serie. La estructura que figura bajo el diagrama de líneas permite centrarse en un espacio temporal en concreto,

ajustando automáticamente la escala del eje Y al espacio seleccionado.



## Datos espaciales

- Tablas
- Microdatos
- i. Visualización mediante Google Earth.



Tras analizar los datos espaciales y crear un archivo kml, Google Earth ofrece una herramienta útil y sencilla para visualizar la capa confeccionada previamente. En este caso, analizando la tipología de las secciones censales de la C.A. de Euskadi (en adelante, CAE) se han creado 12 tipos y, después, el archivo kml. Se han coloreado las diferentes secciones censales de la CAE conforme a esos 12 tipos. La

transparencia de los colores permite, a medida que nos vamos acercando a una dirección, ver las calles, edificios, parques, etc. que se encuentran bajo la capa. Por otro lado, al pulsar sobre el mapa aparecerá un panel informativo que recoge ciertos datos sobre la zona (nombre de la localidad; territorio; sección censal; tipo; porcentaje de menores de 15 años, de mayores de 65, de vascoparlantes; de los que tienen titulación superior.) así como un link que nos deriva a la página web de EUSTAT.

## Aplicaciones

Una vez presentado el catálogo de gráficos, se seleccionan los más adecuados y útiles. Unos se automatizarán y se mostrarán en la página web; otros, en cambio, se utilizarán dentro de Eustat como herramienta de trabajo.

Estos gráficos se han implementado en el lenguaje de programación R.

### Gráficos para producción

Se han automatizado los R Scripts de ciertos gráficos, estandarizando las entradas para que el Grupo de Estadísticas Económicas de Eustat los pueda utilizar como herramienta de trabajo. Para utilizar estos gráficos basta con cargar y ejecutar las correspondientes funciones. Estos son los gráficos automatizados:

1. **Dotplot o diagrama de puntos.**
2. **Gráfico compuesto de diagramas de barras.**
3. **Representación gráfica del análisis de correspondencia.**
4. **Regresión lineal parcial por tramos.**
5. **Gráfico Scatterplot Matrix.**

#### 1. Dotplot o diagrama de puntos

**FUNCIÓN PARA CREAR UN DOTPLOT O DIAGRAMA DE PUNTOS en R: 'PuntuDiagrama()'**

##### DESCRIPCIÓN

La función 'PuntuDiagrama()' es una adaptación de la función 'dotchart()' de la biblioteca 'graphics' propia de R.

Es apropiado para visualizar las variaciones de las modalidades de la variable. Ordena las modalidades alfabéticamente o en base a su variación, y las representa mediante puntos. Dibuja una recta vertical de color rojo en el valor 0 para que se pueda apreciar con claridad el signo de la variación.

##### USO

**PuntuDiagrama(datuak, main=FALSE, ordenalfabet=FALSE, cex=0.8)**

...

## ARGUMENTOS

datuak:	Valores numéricos y sus etiquetas.
main:	Será FALSE por defecto, es decir, no escribirá el título del gráfico. Para escribir el título, señalar main="título deseado, entre comillas".
ordenalfabet:	Será FALSE por defecto, es decir, las etiquetas que definen los puntos no se ordenarán de manera alfabética. Si reza TRUE, las etiquetas aparecerán ordenadas de manera alfabética.
cex:	Parámetro para concretar el tamaño de los puntos y de las letras. Por defecto, tiene un valor de 0.8.

## DETALLES

Los diagramas de puntos o dotplots son los perfectos sustitutos de los diagramas de barras, sobre todo, cuando se quieren analizar las variaciones (positivas/negativas).

## OBSERVACIONES

Antes de utilizar esta función, tanto la función como el archivo rda que queremos representar deben estar guardados en la misma carpeta y, posteriormente, debemos especificar en R la dirección de la misma. Lo podemos hacer de dos formas: pulsando los iconos "Archivo → Cambiar dir... → (carpeta correspondiente)" en R; o bien introduciendo el comando `setwd("dirección completa de la ubicación de la carpeta entre comillas")`.

Una vez hecho esto, debemos cargar el archivo que deseamos representar introduciendo `load("nombre y extensión del archivo, entre comillas")` así como la función que queremos utilizar mediante la orden `source("nombre de la función, entre comillas")`.

## ESTRUCTURA DEL ARCHIVO DE ENTRADA

El archivo de entrada constará de dos columnas:

- En la primera columna, figurarán las etiquetas o categorías (pueden ser nombres, números, o nombres y números).
- En la segunda columna, se detallarán los valores (incrementos/descensos).

## EJEMPLO

```
setwd('C:/Produktzioako grafikoak/Puntu_diagrama')
source('PuntuDiagrama')
load('RGrafico1_1.rda')
PuntuDiagrama(datuak=RGrafico1_1, main="HAZKUNDEA")
```



## 2. Gráfico compuesto de diagramas de barras

### FUNCIÓN PARA CREAR UN GRÁFICO COMPUESTO DE DIAGRAMAS DE BARRAS EN R: 'MarraDiagramak()'

#### DESCRIPCIÓN

Este gráfico está compuesto por diferentes diagramas de barras: diagrama de barras categorizado, diagrama de barras porcentual, y diagramas de barras simples. La función 'MarraDiagramak()' engloba dentro de sí otras tres funciones: 'marradiagrama()', 'marradiagrama100()', y 'MarraDiagramak.gr()'. (Para más información, consultar las notas).

#### USO

**MarraDiagramak**(datuak, titulua="")

...

#### ARGUMENTOS

datuak:	Valores numéricos y sus etiquetas.
titulua:	Por defecto, no se escribirá el título del gráfico. Para asignarle un título señalar titulua="título deseado, entre comillas".

#### DETALLES

Esta visualización de diversos diagramas de barras nos brinda la oportunidad de comparar los datos de la CAE para cada uno de los tres territorios históricos. Por un lado, se puede comparar la distribución de los datos brutos y, por otro, la distribución de las frecuencias absolutas y relativas.

Resulta de utilidad a la hora de representar las frecuencias de las modalidades de una variable cualitativa respecto a las modalidades de otra, siempre y cuando esta última tenga cuatro modalidades y una de ellas pueda ser desglosada en el resto.

#### OBSERVACIONES

Antes de utilizar esta función, tanto la función como el archivo rda que queremos representar deben estar guardados en la misma carpeta y, posteriormente, debemos especificar en R la dirección de la misma. Lo podemos hacer de dos formas: pulsando los iconos "Archivo → Cambiar dir... → (carpeta correspondiente)" en R; o bien introduciendo el comando `setwd("dirección completa de la ubicación de la carpeta, entre comillas")`.

Una vez hecho esto, debemos cargar el archivo que deseamos representar indicando `load("nombre y extensión del archivo, entre comillas")` así como la función que queremos utilizar mediante la orden `source("nombre de la función, entre comillas")`.

Esta función utiliza, a su vez, otras tres funciones que son, a saber:

- La función 'marradiagrama()', que crea diagramas de barras simples y diagramas de barras de frecuencias absolutas, resultando ser un uso especial de la función 'barplot()' propia de R.

- La función 'marradiagrama100()' es la que crea el diagrama de barras porcentual o el diagrama de barras de frecuencias relativas. Esta función es, pues, un uso especial de la función 'barplot()' propia de R.
- La función 'MarraDiagramak.gr()' resulta ser una composición especial de las funciones 'marradiagrama()' y 'marradiagrama 100()'

La función 'MarraDiagramak' utiliza las funciones 'marradiagrama()' y 'marradiagrama100()' para crear diagramas de barras simples y diagramas de barras de frecuencias absolutas y relativas. Además, modifica la matriz de datos de entrada para utilizar la función 'MarraDiagramak.gr()'.

#### ESTRUCTURA DEL ARCHIVO DE ENTRADA:

El archivo de entrada constará de tres columnas:

- En la primera columna, figurarán las etiquetas o categorías (pueden ser nombres, números, o nombres y números).
- En la segunda, se concretarán los territorios históricos (se pueden introducir los nombres de los territorios o sus códigos).
- En la tercera columna, se detallarán los valores (incrementos/descensos).

#### **EJEMPLO**

```
setwd('C:/Produktzioako grafikoak/Barra_diagramak')
source('MarraDiagramak')
load('RGrafico2_1.rda')
datuak <- RGrafico2_1[-c(1,1+nrow(datuak)/3,1+2*nrow(datuak)/3), ]
MarraDiagramak(datuak, titulu="Langile-banaketa / Distribución personal")
```

### **3. Representación gráfica del análisis de correspondencias.**

#### **FUNCIÓN PARA REPRESENTAR EN R EL ANÁLISIS DE CORRESPONDENCIA: 'korrespondentziak()'**

##### **DESCRIPCIÓN**

Función para representar el análisis de correspondencia simple entre dos variables cualitativas.

Consiste en modificar y utilizar la función 'ca()' existente en la biblioteca o en el paquete 'ca'. (Para más información, consultar las notas).

##### **USO**

**Korrespondentziak**(datuak, main="izenburua", zeinux=0, zeinuy=0)

...

## ARGUMENTOS

datuak:	Archivo que se desea representar.
main:	Título. Será NULL por defecto. Para asignar un título al gráfico señalar main="título deseado, entre comillas".
zeinux:	El parámetro que controla el sentido de las abscisas. Será 0 por defecto, es decir, sin giro en el eje X. Para girar 180°, hay que darle el valor 1.
zeinuy:	El parámetro que controla el sentido de las ordenadas. Será 0 por defecto, es decir, sin giro en el eje Y. Para girar 180°, hay que darle el valor 1.

## DETALLES

El análisis de correspondencia es un método gráfico para analizar las relaciones entre las variables categóricas de una tabla de contingencia. El análisis de correspondencia es la generalización de una representación gráfica conocida, es decir, del diagrama de dispersión.

La proximidad (o distancia) de los puntos respecto a uno de los ejes, representa la proximidad o distancia (positiva o negativa) respecto de la media de las modalidades que lo definen. Así, los puntos tenderán a estar más cerca (o lejos) de las modalidades con las que tengan más (o menos) semejanza.

Por otro lado, el tamaño de las bolas representa su frecuencia,

## OBSERVACIONES

Antes de aplicar esta función, tanto la función como el archivo rda que queremos representar deben estar guardados en la misma carpeta y, posteriormente, debemos especificar en R la dirección de la misma. Lo podemos hacer de dos formas: pulsando los iconos "Archivo → Cambiar dir... → (carpeta correspondiente)" en R; o bien introduciendo el comando `setwd("dirección completa de la ubicación de la carpeta, entre comillas")`.

Una vez hecho esto, debemos cargar el archivo que deseamos representar indicando `load("nombre y extensión del archivo, entre comillas")` así como la función que queremos utilizar mediante la orden `source("nombre de la función, entre comillas")`.

Esta función utiliza, a su vez, las siguientes funciones:

- La función 'ca()' modifica la función 'ca()' del paquete 'ca' que efectúa un análisis simple de correspondencia. Da opción a controlar la dirección de los ejes de abscisas y ordenadas, es decir si se quieren rotar 180° o no.
- La función 'korresp.gr()' utiliza los resultados de la función 'ca()', mostrando las correspondencias en dos dimensiones.

## ESTRUCTURA DEL ARCHIVO DE ENTRADA:

El archivo de entrada constará de tres columnas:

- En la primera columna, figurarán las etiquetas o categorías (pueden ser nombres, números, o nombres y números).
- En la segunda, se concretarán los territorios históricos (se pueden introducir los nombres de los territorios o sus códigos).
- En la tercera se recogerán los valores.

### EJEMPLO

```
setwd('C:/Produkcziarako grafikoak/Korrespondentzia')
source('korrespondentziak')
load('RGrafico3_1.rda')
datuak<-RGrafico3_1[-c(1,nrow(datuak)/3+1,2*nrow(datuak)/3+1), ]
korrespondentziak(datuak, main="Korrespondentziak")
```

## 4. Regresión lineal por tramos, por sectores.

### FUNCIÓN PARA CONFECCIONAR Y REPRESENTAR LA REGRESIÓN LINEAL POR TRAMOS, POR SECTORES EN R: 'Erregresioa()'

#### DESCRIPCIÓN

Función para analizar si la relación existente entre dos variables cualitativas es lineal por tramos, según las modalidades de una tercera variable cualitativa.

La función 'Regresión()' engloba dentro de sí otras tres funciones: 'bitxiazein()', 'transf()', y 'LoessErregresio()'. (Para más información, consultar las notas).

#### USO

**Erregresioa**(datuak, titulua=NULL, azpititulua=NULL, logbektorea=c(0,0))

...

#### ARGUMENTOS

datuak:	Archivo que se desea representar.
titulua:	Título. Será NULL por defecto. Para asignar un título al gráfico señalar titulua="título deseado, entre comillas".
azpititulua:	Subtítulo. Este también será NULL por defecto. Para asignar un subtítulo señalar azpititulua="subtítulo deseado, entre comillas".
logbektorea:	Vector para hacer cálculos logarítmicos cuando se vaya a utilizar una escala logarítmica. De no ser que se señale lo contrario, no se harán cálculos logarítmicos (será c(0,0) por defecto). 1 corresponde a la escala logarítmica; y 0 a la escala existente.

## DETALLES

La regresión local (LOESS o LOWESS en inglés) es un nuevo método de construcción basado en métodos clásicos como la regresión lineal o la no lineal. Estos nuevos métodos de regresión están diseñados para casos en los que los procedimientos clásicos resultan inadecuados o insuficientes.

El objetivo de la regresión local es crear una función que describa, punto por punto, la parte determinista de las variaciones de los datos. Para ello, ajusta modelos sencillos sobre subgrupos locales, combinando la regresión lineal de los mínimos cuadrados y la regresión no lineal.

## OBSERVACIONES

Antes de aplicar esta función, tanto la función como el archivo rda que queremos representar deben estar guardados en la misma carpeta y, posteriormente, debemos especificar en R la dirección de la misma. Lo podemos hacer de dos formas: pulsando los iconos "Archivo → Cambiar dir... → (carpeta correspondiente)" en R; o bien introduciendo el comando `setwd("dirección completa de la ubicación de la carpeta, entre comillas")`.

Una vez hecho esto, debemos cargar el archivo que deseamos representar indicando `load("nombre y extensión del archivo, entre comillas")` así como la función que queremos utilizar mediante la orden `source("nombre de la función, entre comillas")`.

Esta función utiliza, a su vez, otras tres funciones:

- La función 'bitxiazein()' es la función para delatar los valores atípicos, estableciendo para ello un criterio para tomar por atípicos los valores. De hecho, se tomarán por valores extraños aquellos menores que el mínimo de los valores superiores a  $(q1 - zenbath * (q3 - q1))$  y aquellos mayores que el máximo de los valores inferiores a  $(q3 + zenbath * (q3 - q1))$ . El parámetro zenbath por defecto tiene un valor de 2,75; y q1 y q3 hacen referencia al primer y tercer cuartil respectivamente.
- La función 'transf()' nos permite buscar la mejor transformación de datos. Para ello, calcula el estadístico del test de normalidad de Shapiro-Wilks. Así, aquel que posea el mayor valor será el que proporcione la mejor transformación.
- La función 'LoessErregresio()' sirve para valorar si es necesario o no realizar la translación y utilizar la escala logarítmica. Posteriormente, si existen puntos del mismo valor los mueve para evitar que se superpongan. Después, dibuja los puntos, cada uno en su correspondiente color, y ajusta las regresiones locales, totales y parciales (sin tener en cuenta los valores atípicos). Por último, identifica y etiqueta los puntos que se encuentran alejados del modelo principal. Esta función utiliza, a su vez, la función 'loess()' de la biblioteca 'graphics' propia de R. Función esta última que sirve para calcular la regresión local.

## ESTRUCTURA DEL ARCHIVO DE ENTRADA:

El archivo de entrada constará de cuatro columnas:

- En la primera columna, se colocan los identificadores de los puntos (es decir, los que utilizaremos para identificar valores extraños, valores atípicos, etc.)

- En la segunda columna, figurarán las etiquetas o categorías (pueden ser nombres, números, o nombres y números).
- En la tercera columna, pondremos los valores correspondientes a la variable que irá en el eje X.
- En la cuarta columna, pondremos los valores correspondientes a la variable que irá en el eje Y.

### EJEMPLO

```
Setwd("C:/Produktzioarako grafikoa/ Zatikako erregresio lineala, sektoreka")
Source("Erregresioa")
load("RGrafico4_1.rda")
datuak <- RGrafico4_2
Erregresioa(datuak, titulua="Balioak Enpleguaren arabera", Azpititulua="Eskala
logaritmikoa", logbektorea=c(1,1))
```

## 5. Gráfico Scatterplot Matrix

### FUNCIÓN PARA REPRESENTAR EN R UN GRÁFICO SCATTERPLOT MATRIX CON HISTOGRAMAS DIAGONALES: 'binakakoa()'

#### DESCRIPCIÓN

Se trata de la representación gráfica que nos da la oportunidad de representar  $n$  ( $3 \leq n \leq 5$ ) variables cuantitativas.

La función 'binakakoa()' utiliza a su vez otras cuatro funciones: 'transf()', 'hist.puntuak()', 'panel.puntuak()', y 'binakako.pairs()'. (Para más información, consultar las notas).

#### USO

```
binakakoa(datuak, titulua=NULL, azpititulua=NULL, logbektorea=rep(0,ncol(datuak)-2)
...
```

#### ARGUMENTOS

datuak:	Archivo que se desea representar.
titulua:	Título. Será NULL por defecto, es decir, sin título. En caso de querer asignar un título, indicar titulua="título deseado, entre comillas".
azpititulua:	Subtítulo. Este también será NULL por defecto. Para poner un subtítulo, indicar azpititulua="subtítulo deseado, entre comillas".
logbektorea:	Vector para hacer cálculos logarítmicos. De no ser que se señale lo contrario, no se harán cálculos (por defecto, será rep(0,ncol(datos)-2)). 1 corresponde a la escala logarítmica; y 0 a la escala propia de la variable.

## DETALLES

El gráfico llamado Scatterplot explica cómo es la relación existente entre dos variables. Es decir, puede mostrar la forma y tendencias de la distribución, si la relación existente entre las variables es fuerte o débil, puede mostrar también los clústeres o modelos, y evidenciar valores extraños o atípicos.

El histograma, en cambio, expresa la densidad relativa de la distribución de la muestra. La cantidad de datos existente en cada uno de los intervalos de igual anchura resultantes de la partición del conjunto de datos inicial se expresa mediante barras, cuya altura será proporcional a su frecuencia.

La matriz plot está compuesta por celdas que albergan diagramas en su interior. Cada celda de diagrama de la matriz posee un diagrama que expresa la relación existente entre las variables de las filas y columnas de celdas. Por lo tanto, una matriz plot muestra las relaciones existentes entre todos los posibles pares de variables del conjunto de datos.

Todos los gráficos que figuran en las celdas de una matriz plot no tienen por qué ser iguales. En este caso, hemos utilizado matrices Scatterplot, excepto en las diagonales, donde se han representado histogramas para recrear una idea sobre la distribución de las variables.

## OBSERVACIONES

Antes de aplicar esta función, tanto la función como el archivo rda que queremos representar deben estar guardados en la misma carpeta y, posteriormente, debemos especificar en R la dirección de la misma. Lo podemos hacer de dos formas: pulsando los iconos "Archivo → Cambiar dir... → (carpeta correspondiente)" en R; o bien introduciendo el comando `setwd("dirección completa de la ubicación de la carpeta, entre comillas")`.

Una vez hecho esto, debemos cargar el archivo que deseamos representar indicando `load("nombre y extensión del archivo, entre comillas")` así como la función que queremos utilizar mediante la orden `source("nombre de la función, entre comillas")`.

Esta función utiliza, a su vez, otras tres funciones:

- La función '`transf()`', que delata los elementos atípicos.
- La función '`panel.hist()`', que representa los histogramas situados en la diagonal del gráfico.
- La función '`panel.puntuak()`', que representa los diagramas de dispersión y las regresiones del resto de celdas (a excepción de las situadas en la diagonal).
- La función '`binakako.pairs()`', que resulta ser una adaptación de la función '`pairs()`' de la biblioteca 'graphics' propia de R. Dicha función, a su vez, se encarga de representar histogramas en la diagonal de la matriz, y diagramas de dispersión y regresiones en el resto de celdas.

## ESTRUCTURA DEL ARCHIVO DE ENTRADA:

El archivo de entrada tendrá entre cuatro (para  $n=3$ ) y seis (para  $n=5$ ) columnas, dependiendo del número de variables a analizar:

- En la primera columna, figurarán las etiquetas o categorías (pueden ser nombres, números, o nombres y números).
- A partir de la segunda columna, se concretarán las variables a analizar (entre 3 y 5).

### EJEMPLO

```
setwd("C:/Produktzioarako grafikoa/ Scatterplot matrix grafikoa")
source("binakakoa")
load("RGrafico5_1.rda")
datuak <- RGrafico5_1
binakakoa(datuak, titulu="PB_sf, VAB_cf, C_P, eta EN_E (adin-geruzka)",
azpititulu="PB_sf, VAB_cf, C_P, eta EN_E logaritmo-eskalan", logbektorea=c(1,1,1,1))
```

## Gráficos para difusión

Se han creado 2 gráficos dinámico-interactivos y una visualización mediante Google Earth para publicar en la página web de Eustat. Para utilizar estos gráficos, basta con cargar y ejecutar las correspondientes funciones. Estos son los gráficos automatizados:

1. **Gráfico Motion Chart.**
2. **Gráfico del tipo Time Line.**
3. **Visualización mediante Google Earth.**

## HERRAMIENTA PARA LA VISUALIZACIÓN DE GRÁFICOS MOTION CHART Y TIME LINE. GOOGLE VISUALIZATION API

La IPA, Interfaz de Programación de Aplicaciones, o API (**A**pplication **P**rogramming **I**nterface en inglés) es un conjunto de funciones y procedimientos (o métodos, en el caso de programaciones dirigidas a objetos) que ofrece una biblioteca concreta para ser utilizada por otro software.

Uno de los principales objetivos del API es ofrecer un conjunto de funciones de uso general, por ejemplo, para dibujar ventanas o iconos de pantalla. Así, los programadores hacen uso de las ventajas que ofrece el API y evitan, tener que programar desde el principio.

El llamado Google Visualization API es un API JavaScript diseñado para facilitar a los usuarios el uso de la tecnología avanzada de visualización. Además, es fácil y rápido de integrar con nuevas visualizaciones. De hecho, es una interfaz común para introducir datos en la red y para obtener datos para las visualizaciones.

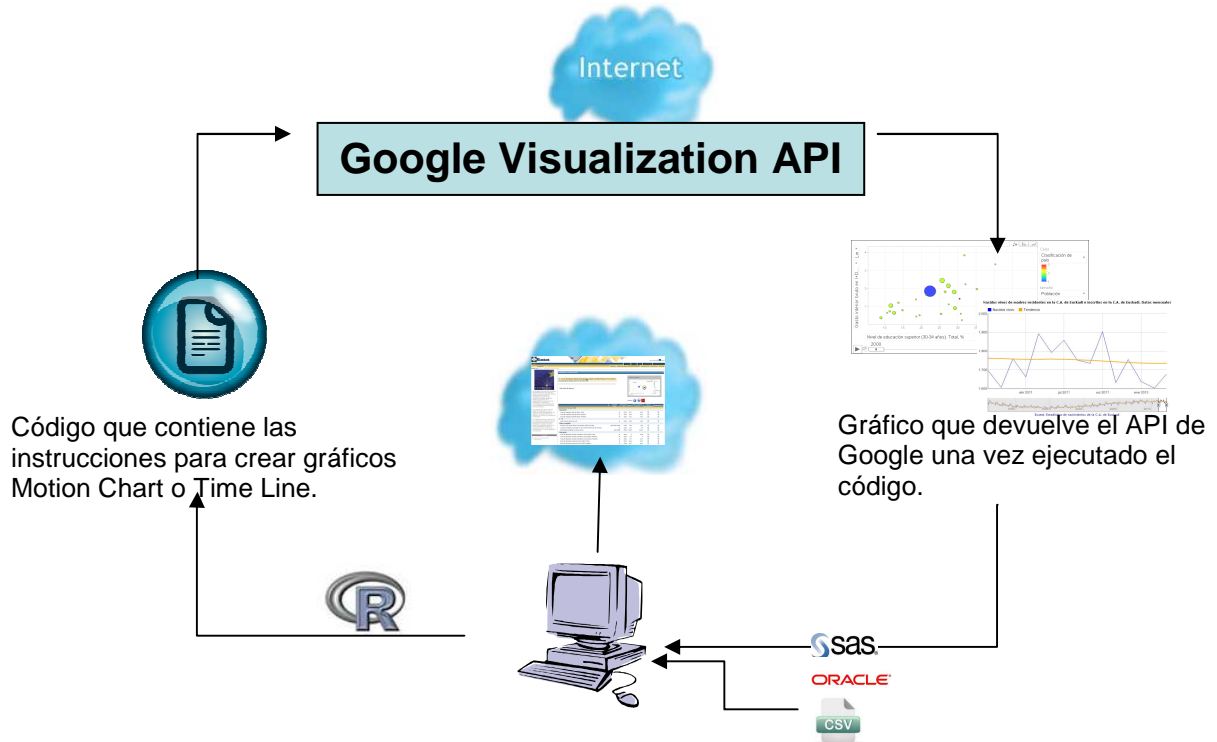
Se basa en dos elementos claves: la facilidad y la ubicuidad. Si tras seleccionar cualquier gráfico de los que figuran en la galería del Google Visualization API



programamos el código para crear el gráfico (respetando, claro está, las condiciones y estructura requeridas por la herramienta) y llamamos al API, éste se ejecutará on-line y mandará de vuelta el gráfico seleccionado.

Para poder visualizar el gráfico, será indispensable que el ordenador tenga instalado el Flash Player.

A continuación, pasamos a detallar el esquema del procedimiento a seguir para crear el gráfico seleccionado (o cualquier otro del API):



## HERRAMIENTA PARA VISUALIZAR EL ANÁLISIS DE LAS TIPOLOGÍAS DE LAS SECCIONES CENSALES.GOOGLE EARTH

Google Earth es un programa informático similar a un Sistema de Información Geográfica (GIS Geographic Information System en inglés). Permite visualizar imágenes del planeta, combinando imágenes tomadas por los satélites durante los últimos tres años, mapas y una base de datos muy buena.

El buscador de Google Earth puede ser utilizado para multitud de cosas, como por ejemplo:

- Buscar un país, una ciudad o una dirección en concreta de una ciudad.
- Buscar mediante sus coordenadas cualquier territorio del Planeta Tierra.
- Marcar las rutas existentes entre dos ciudades de un mismo país, o entre dos direcciones de una misma ciudad.
- Marcar la ruta más adecuada entre dos ciudades o países diferentes.

Dispone, también, de varias "Layers" o capas de información, que pueden ser activadas o desactivadas de manera independiente y que permiten visualizar diferentes tipos de datos del espacio geográfico.

Gracias a estas capas, los usuarios podrán, entre otras:

- Ver los nombres y las calles de una ciudad.
- Conocer la ubicación de centros escolares, hospitales, hoteles, restaurantes, parques, lugares de interés, etc.
- Visualizar fronteras, carreteras y vías de ferrocarril.
- Ver volcanes, epicentros de terremotos, lagos, lagunas y ríos, etc.
- Identificar la localización de los lugares de mayor transcendencia histórica y cultural.
- Visualizar en tres dimensiones colinas, montañas y otras altiplanicies.

## 1. Gráfico Motion Chart

### 1. FUNCIÓN EN R QUE PREPARA UN ARCHIVO CON LA ESTRUCTURA ADECUADA PARA CREAR GRÁFICOS DE TIPO MOTION CHART, PARTIENDO DEL FORMATO DE DATOS DE LOS INDICADORES DE EUSTAT: 'Elkartu\_gapminder()'

#### DESCRIPCIÓN

Tomando los archivos csv de los indicadores (una tabla por cada indicador), esta función devuelve el archivo necesario para utilizar una adaptación de la función 'gvisMotionChart()' del paquete 'googleVis' de R.

#### USO

**Elkartu\_gapminder**(lista1, fitxategia\_izena, euskara=TRUE, populazioa)

#### ARGUMENTOS

- |                   |  |
|-------------------|--|
| lista1:           | Es un archivo txt, en el que se concretarán, separados por espacios, los nombres de los archivos csv (sin extensión) correspondientes a los indicadores que queremos utilizar, separados por espacios. Deben obedecer el siguiente orden: primero el indicador del eje X; después, el del eje Y; a continuación, el indicador de población (en caso de que haya); y, por último, el resto de indicadores. Para finalizar, concretaremos el número de indicadores o archivos utilizados (número). |
| fitxategia_izena: | Nombre del archivo txt de salida.  |

gaztelania,	Idioma de los archivos de entrada y salida. Siempre habrá que
euskara, ingelesa:	elegir uno y sólo uno. Asignaremos TRUE al idioma seleccionado.
populazioa:	Saldrá FALSE por defecto. En caso de que haya un indicador que especifique la población, asignarle TRUE.

## DETALLES

El paquete 'googleVis' de R proporciona una herramienta fácil y útil para crear gráficos al estilo de los Motion Chart del API de Google. Para crear gráficos de este estilo, se ha creado la función 'Motion\_chart()', partiendo de las funciones 'gvisMotionChart()' y 'gvis()' del paquete 'googleVis'. Para poder ejecutar función 'Motion\_chart()' tal y como veremos en el siguiente apartado, el archivo de entrada debe tener una estructura especial, en concreto, la estructura del archivo de salida que se obtiene, precisamente, al ejecutar la función 'Elkartu\_gapminder()'.

El archivo de entrada de la función Motion\_chart (y el archivo de salida de la función 'Elkartu \_gapminder()') debe tener la siguiente estructura:

- 1ª columna: Columna de tipo carácter en la que deben figurar los nombres de las entidades (de los países, en este caso).
- 2ª columna: En esta columna se detallarán, en orden, los años objeto de análisis.
- 3ª columna: La variable a representar en el eje X.
- 4ª columna: La variable a representar en el eje Y.
- 5ª columna: Variable que determinará el color de las burbujas.
- 6ª columna: Variable que determinará el tamaño de las burbujas.
- A partir de la 7ª columna: Resto de variables a analizar.

NOTA: En el caso de que existan valores perdidos en la 6ª columna, será la siguiente columna sin valores perdidos la que fije el tamaño de las burbujas. Por lo tanto, aquellos archivos que contengan poblaciones no deben poseer valores perdidos.

La función 'Elkartu \_gapminder()' devuelve un archivo txt con la estructura arriba mencionada. Al crear el fichero, aparece un editor en R, y pulsando sobre las variables da opción a cambiar los nombres, guardándose de manera automática los cambios realizados.

## OBSERVACIONES

Primero, hay que guardar en una misma carpeta los archivos de los indicadores que se quieren utilizar. En caso de que entre los archivos de los indicadores no se encuentre el archivo que contiene la población, la carpeta debe contener también los archivos que recogen la población de la UE 27 (descargado desde la página web de Eurostat) y la de la CAE (descargado desde la página web de Eustat). Todos los archivos deben estar en formato csv.

Antes de ejecutar la función, debemos especificar en R el directorio en el que se encuentran los ficheros de los indicadores. Lo podemos hacer de dos formas: pulsando

los iconos "Archivo → Cambiar dir... → (carpeta correspondiente)" en R; o bien introduciendo el comando `setwd("dirección completa de la ubicación de la carpeta, entre comillas")`. El archivo que se creará tras ejecutar la función se guardará también en esa dirección.

### ESTRUCTURA DE LOS ARCHIVOS DE ENTRADA:

Los archivos de cada indicador deben tener la siguiente estructura (excepto los de la población, siempre que no sea uno de los indicadores):

- En la primera columna, irán los nombres de todos los países (no dejar la primera celda en blanco, ponerle un título). Tenemos que tener en cuenta que los nombres de los países deben estar escritos de la misma forma en todos los archivos (los acentos tal y como se encuentran ahora, sin paréntesis, etc.).
- En el resto de las columnas aparecerán los años. Los títulos de dichas columnas serán los años.

Además, en todos los archivos de inicio hay que utilizar el mismo símbolo para expresar los valores perdidos (el símbolo ".") y para los valores decimales (el símbolo ",").

### **EJEMPLO**

**Elkartu\_gapminder**(lista1="indicadores.txt", fitxategia\_izena="Euskadi\_en\_EU27", gaztelania=FALSE, euskara=FALSE, ingelesa=FALSE, populazioa=TRUE)

## **2. FUNCIÓN EN R PARA GENERAR EL CÓDIGO PARA CREAR UN GRÁFICO DEL TIPO MOTION CHART: 'Motion\_chart()'**

### **DESCRIPCIÓN**

Función que crea el código para construir el gráfico dinámico llamado Motion Chart, que permite analizar los datos en cinco dimensiones distintas. Dicho código dado en formato de texto hace referencia al Google Visualization API y puede ser introducido en una página web. Utiliza el archivo obtenido tras ejecutar la función 'Elkartu\_gapminder()' a modo de archivo de entrada.

La función 'Motion\_chart()' alberga en su interior otras cuatro funciones: 'my.gvis()', 'my.gvisMotionChart()', 'my.gvis\_web()' y 'my.gvisMotionChart\_web()'. (Para más información, consultar las notas).

### **USO**

**Motion\_chart**(lista1, fitxategia\_izena, izenb\_ind, euskara=TRUE, irteerako\_fitxategia, populazioa=FALSE, chartid, grafikoa=FALSE, kodea=FALSE)

...

### **ARGUMENTOS**

**lista1:** Es un archivo txt, en el que se concretarán, separados por espacios, los nombres de los archivos csv (sin extensión) correspondientes a los indicadores que queremos utilizar. Debe

	tratarse del mismo txt utilizado en el argumento 'lista1' de la función 'Elkartu_gapminder()'. fitxategia_izena:	Nombre del archivo a visualizar. Debe tratarse del mismo nombre que hemos dado al argumento 'fitxategia_izena' de la función 'Elkartu_gapminder()'.
	izenb_ind:	En caso de tratarse de NULL, dará a las variables de los indicadores los nombres que figuran en el archivo asignado al parámetro 'lista1'. De lo contrario, se podrá concretar un vector de caracteres con los nombres que queremos dar a los indicadores.
	gaztelania, euskara, ingelesa:	Idioma de los archivos de entrada y salida. Siempre habrá que elegir uno y sólo uno. Asignaremos TRUE al idioma seleccionado.
	irteera_fitxategia:	Nombre que tomará el archivo de salida.
	populazioa:	Será FALSE por defecto. En caso de que haya un indicador de población, asignarle TRUE.
	chartid:	Un nombre para identificar el gráfico.
	grafikoa, kodea:	Siempre habrá que elegir uno y sólo uno de entre los dos. Para ver el gráfico en la red: gráfico=TRUE. Para crear solamente el código: código=TRUE.

## DETALLES

Podemos decir que se trata de un gráfico fruto de la extensión de un diagrama de dispersión para el caso multivariante. De hecho, una vez seleccionadas las variables de los ejes X e Y, se podrán seleccionar, por ejemplo, el tamaño y color de las burbujas, o mostrar la evolución que las burbujas han tenido a lo largo del tiempo. Esas cinco variables se encuentran bajo elección del usuario, por lo que además de dinámico resulta también interactivo.

Estos son, en concreto, los aspectos que se pueden seleccionar:

- La variable del eje X.
- La variable del eje Y.
- La variable que medirá el tiempo. A decir verdad, esta última no se puede elegir, ya que coincidirá con la variable que mide el tiempo en el archivo de entrada que será, normalmente, el año.
- La variable que determinará el color de las burbujas.
- La variable que determinará el tamaño de las burbujas.

Gracias a la opción para analizar los datos a lo largo del tiempo, este gráfico permite identificar modelos y tendencias que no pueden verse en otros gráficos tradicionales.

Además de permitirnos seleccionar los datos, podemos verlos en tres gráficos distintos:



Gráfico tipo Motion Chart. (Ya citado).



Diagrama de barras dinámico. Añade a un diagrama de barras tradicional una barra dinámica que mide el tiempo. Así, puede analizarse la evolución de las barras a lo largo del tiempo.



Diagrama de líneas estático. Expresa de manera estática y mediante una línea la evolución que cada burbuja ha tenido a lo largo del tiempo.

Ofrece, además, otras herramientas de utilidad como, por ejemplo:

*Cambio de la frecuencia claridad las solución.*



*escala.* Permite elegir entre escala lineal y logarítmica. En ocasiones, cuando de una variable es mucho mayor a la del resto, no se pueden observar con diferencias entre variables. En estos casos, la escala logarítmica puede ser la

*Rastros.* Al utilizar la  opción *rastros* podemos observar las tendencias que las variables han tenido a lo largo del tiempo.



*Velocidad.* Podemos elegir la velocidad de reproducción del gráfico.

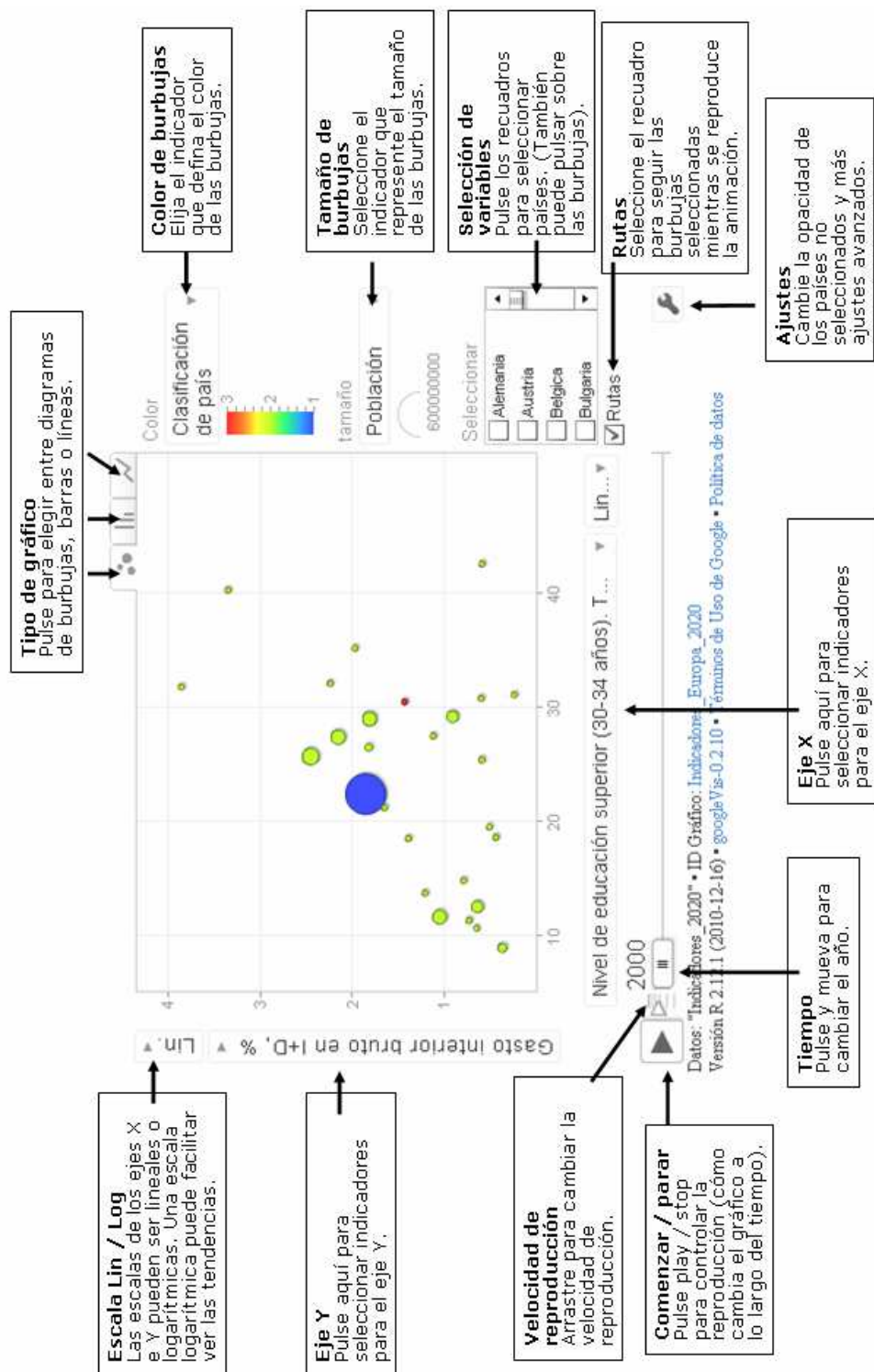


*Zoom.* Permite acercar la imagen y centrarnos en una sección concreta del gráfico.



*Panel de configuración.* Mediante el control móvil, permite ajustar la claridad de las burbujas (o barras, o líneas, en su caso) no seleccionadas.

A continuación, se muestra una imagen que explica de manera gráfica la estructura y funcionamiento del gráfico:



Adaptado de  
[www.gapminder.org](http://www.gapminder.org)

## OBSERVACIONES

Para poder aplicar esta función, debe estar instalada en R la biblioteca 'googleVis'. Si no está instalada, debemos seguir los siguientes pasos: Paquetes → Instalar paquete(s) → “seleccionar un CRAN mirror” → “seleccionar paquete googleVis”. Una vez instalada, no será necesario realizar de nuevo todo ese procedimiento.

En primer lugar, hay que ejecutar la función 'Elkartu\_gapminder()'. Después, habrá que asignar a los argumentos 'lista1' y 'fitxategia\_izena' el mismo nombre que figura en la función 'Elkartu\_gapminder()'.

También en ésta ocasión, deben estar en la misma carpeta los archivos que se quieren utilizar y todos los programas de R. Además, tal y como ocurría con la función anterior, R deberá apuntar a ese directorio.

### ESTRUCTURA DEL ARCHIVO DE ENTRADA:

La estructura del archivo será la misma del archivo obtenido al ejecutar la función 'Elkartu\_gapminder()'.

## EJEMPLO

```
Motion_chart(lista1="indicadores.txt", fitxategia_izena="Euskadi_en_EU27",
izenb_ind=nombres_indicadores_castellano, euskara=TRUE,
irteerako_fitxategia="codigo_web", populazioa=TRUE, chartid="Euskadi_en_la_UE_27",
kodea=TRUE)
```

## 2. Gráfico de tipo Time Line.

### FUNCIÓN EN R PARA GENERAR EL CÓDIGO PARA CREAR UN GRÁFICO DEL TIPO TIME LINE: 'TimeLine()'

#### DESCRIPCIÓN

Se trata de una función que, tomando un archivo de serie temporal y utilizando los gráficos y estructuras llamadas “control” del Google Visualization API, genera el código para crear un gráfico del tipo Time Line.

Este gráfico está formado por un diagrama de líneas y una estructura llamada “control”. El control móvil nos permite centrarnos en un espacio de tiempo concreto.

#### USO

```
TimeLine(fitxategia_izena, chart.type, chartid, controlid, chart.options, legend, series,
control.options, control.state, euskara=TRUE)
```



**ARGUMENTOS**

fitxategia_izena:	Nombre del archivo de entrada en formato csv.
chart_type:	Tipo de gráfico que deseamos representar. En este caso, se trata de un "LineChart" (diagrama de líneas), pero este mismo programa funcionaría también para 'AreaChart', 'ComboChart' o 'ScatterChart' del Google Visualization API.
chartid:	Un nombre para identificar el gráfico.
controlid:	Un nombre para identificar la estructura móvil llamada control.
chart.options:	Todas las opciones que puede tener el gráfico. Tanto el gráfico 'LineChart' como el resto, ofrecen multitud de opciones de configuración. Podemos elegir, entre otros, el color y la anchura de las series; el aspecto y el tamaño de los ejes X e Y; el tipo y tamaño de texto; o, incluso, la altura y anchura del gráfico. Hemos seleccionado algunos de esos aspectos al definir la función; otros, en cambio, los proporciona la herramienta por defecto.
control.options:	Todas las opciones que puede tener la estructura móvil llamada control. En esta ocasión también se han seleccionado algunas de las opciones de todas las existentes; el resto son las que proporciona la herramienta por defecto.
control.state:	Se trata del espacio de tiempo en el que queremos que en un principio se centre la estructura llamada control (y, por ende, el gráfico). Será un vector de dos componentes. El primero de ellos establecerá la fecha de inicio; el segundo, en cambio, la fecha de finalización. La fecha debe ir entrecomillada y en formato "aaaa-mm-dd". Por defecto, el gráfico se centrará en todo el intervalo de tiempo.
azpitulua:	Subtítulo del gráfico.
gaztelania, euskara, ingelesa:	Idioma de los archivos de entrada y salida. Siempre habrá que elegir sólo uno. Asignaremos TRUE al idioma seleccionado.

**DETALLES**

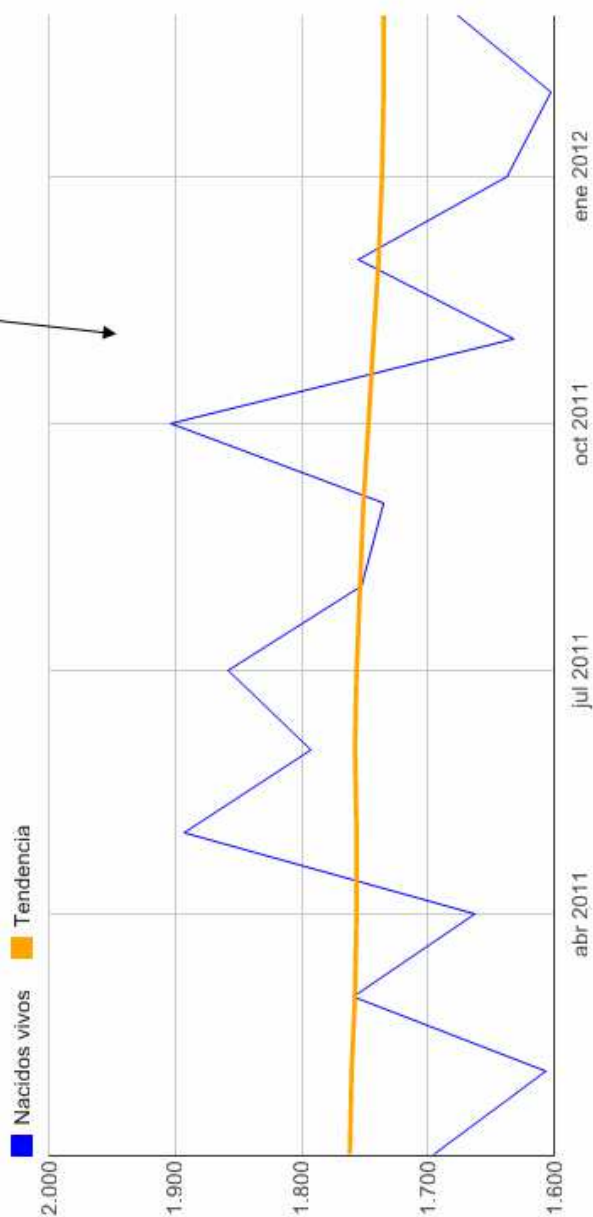
El gráfico TimeLine es un diagrama de líneas interactivo, que puede representar más de una serie temporal al mismo tiempo. Al acercar el ratón a la línea que representa una de las series temporales, aparecerá un cuadro de diálogo con información sobre el valor que tiene la serie en la fecha seleccionada.

Bajo el gráfico lineal aparece una herramienta que permite centrarse en un espacio temporal en concreto. Además, al centrarnos en un período concreto, el gráfico se re-escala automáticamente.

A continuación, se muestra una imagen que explica de manera gráfica la estructura y funcionamiento del gráfico:

**Gráfico.**  
Gráfico de líneas. Dependiendo de la estructura "control" muestra parte de la serie o la serie completa.

Nacidos vivos de madres residentes en la C.A. de Euskadi e inscritos en la C.A. de Euskadi. Datos mensuales



Eustat. Estadística de nacimientos de la C.A. de Euskadi

**Estructura móvil "control".**  
Permite centrarse en el intervalo de tiempo deseado. Al mover esta estructura el gráfico se re-escala automáticamente.

## OBSERVACIONES

Antes de ejecutar la función, debemos señalar en R el directorio en el que se encuentra el archivo csv que vamos a utilizar. Lo podemos hacer de dos formas: pulsando los iconos "Archivo → Cambiar dir... → (carpeta correspondiente)" en R; o bien introduciendo el comando `setwd("dirección completa de la ubicación de la carpeta, entre comillas")`. El archivo que se creará tras ejecutar la función se guardará también en esa dirección.

Una vez hecho esto, debemos cargar el archivo que vamos a utilizar indicando `load("nombre del archivo, entre comillas")` así como la función que queremos utilizar mediante la orden `source("nombre de la función, entre comillas")`.

## ESTRUCTURA DEL ARCHIVO DE ENTRADA:

El archivo deberá tener la siguiente estructura:

- En la primera columna aparecerán las fechas en formato "dd/mm/aaaa" (no hay que dejar la primera celda vacía, ponerle un título).
- En la segunda columna, introduciremos los datos de la serie que queremos colorear de azul, que en este caso se trata de los datos brutos de la serie de nacimientos (no hay que dejar la primera celda vacía, ponerle un título).
- En la tercera columna, introduciremos los datos de la serie que queremos colorear en naranja, que en este caso se trata de la tendencia de la serie de nacimientos (no hay que dejar la primera celda vacía, ponerle un título).

## EJEMPLO

```
TimeLine(fitxategia_izena="IPI_euskara", chart.type="LineChart", chartid= "chart_id",
controlid= "control_id", chart.options= list(chartArea=list(height="80%", width="90%"),
legend=list(position="top"), series=list(seriea1=list(color='blue', lineWidth=1),
seriea2=list(color='orange', lineWidth=1))), control.options= list(filterColumnIndex=0,
ui=list(chartType="LineChart", chartOptions=list(chartArea=list(width="90%",
height="50%"), series=list(seriea1=list(color="blue"), seriea2=list(color="orange"))),
minRangeSize=2592000000)), control.state=list("2011-01-01", "2011-12-01"),
euskara=TRUE)
```

## 3. Visualización mediante Google Earth

Al igual que en 1991, 1996 y 2001, en 2012 también se ha realizado la clasificación de las tipologías de las secciones censales de la CAE. El objeto de este análisis es clasificar las secciones censales en base a las características que comparten, uniendo secciones que comparten características comunes. Así, las secciones de un mismo tipo serán lo más homogéneas posible entre sí, y lo más diferentes posible al resto.

Este tipo de análisis resulta de gran utilidad en diversas áreas tales como: el diseño de muestreos estratificados, análisis relacionados con la planificación urbana, conocimiento exhaustivo de la situación socio-económica de la sociedad.

## **ANÁLISIS DE LAS SECCIONES CENSALES. METODOLOGÍA**

Para empezar, se han seleccionado de la base de datos de las estadísticas socio-demográficas (BSD) las principales viviendas familiares y los habitantes que en ellas residen. Después, se han seleccionado las variables a analizar y la mayoría (a excepción de la renta) se ha dividido en varias modalidades (no existe una cantidad máxima de divisiones, puesto que el Análisis Factorial Múltiple utilizado no limita el número de modalidades).

La mayoría de las variables seleccionadas reflejan las principales características de los individuos, tales como el sexo, la edad, el nivel educativo, la renta, la relación con el mundo laboral, la situación sociolingüística, y el lugar de nacimiento. Además de los aquí detallados, se suman al final los datos relativos a la vivienda (tamaño y año de construcción) y a la familia.

Una vez seleccionadas las variables objeto de análisis, se asignará a cada sección censal la frecuencia que tiene en cada modalidad, en valores relativos, siendo 100 el resultado de la suma de todas las modalidades de cada variable (a excepción de la renta). A pesar de no entrar dentro del análisis, la renta familiar también se ha clasificado por modalidades, puesto que resulta de ayuda a la hora de describir los tipos.

Partiendo de la matriz de datos creada mediante estas pautas, pero sin tener en cuenta las rentas y la lengua materna, se ha realizado el Análisis Factorial Múltiple y se han guardado los primeros 7 factores, puesto que representan el 70,56% de la varianza inicial (de la cantidad de información inicial). Posteriormente, se ha realizado una clasificación jerárquica con 7 factores. De esta forma, se han obtenido 15 secciones, y se ha aplicado un proceso de estabilidad. Pensando que eran demasiadas secciones, se han anexionado algunas de ellas obedeciendo a criterios de proximidad y lejanía, pasando a ser 12 las secciones. Por último, se ha tomado el centro de esas 12 secciones y se han utilizado como semilla para crear una nueva partición mediante el proceso kmeans o núées dynamiques.

## **LA FUNCIÓN EN R PARA VISUALIZAR EL ANÁLISIS DE TIPOLOGÍAS DE LA CAE MEDIANTE GOOGLE EARTH: 'KmlSortu()'**

### **DESCRIPCIÓN**

Esta función crea, partiendo de un archivo shp y adaptando la función '*kmlPolygon()*' de la biblioteca '*maptools*' de R, un archivo visualizable mediante el programa Google Earth.

Google Earth es una herramienta gratuita, sencilla y conocida que permite la visualización y el análisis de datos espaciales a simple vista y de manera sencilla.

### **USO**

**KmlSortu**(n\_tipol, fitxategia, gaztelania=TRUE, alpha)

### **ARGUMENTOS**

n\_tipol: Cantidad de tipos obtenidos tras realizar el análisis de las tipologías.

fitxategia:	Archivo shp que contiene toda la información necesaria.
gaztelania, euskara, ingelesa:	Idioma de la visualización que se va a crear. Siempre habrá que elegir uno y sólo uno. Asignaremos TRUE al idioma seleccionado.
alpha:	Parámetro que fija la transparencia de los colores que se utilizarán para colorear las secciones de Google Earth. Será 120 por defecto.
legenda:	Será FALSE por defecto, y no se mostrara la leyenda. Cuando sea TRUE, se añadirá leyenda con la descripción de las tipologías en el lado izquierdo de la pantalla.

## DETALLES

Al utilizar el mapa coloreado que clasifica las secciones censales de la CAE en diferentes tipos, el usuario podrá utilizar todas las opciones que ofrece la herramienta Google Earth.

A parte de las que ofrece la herramienta, el archivo kml de los tipos creados dispone de otras opciones.

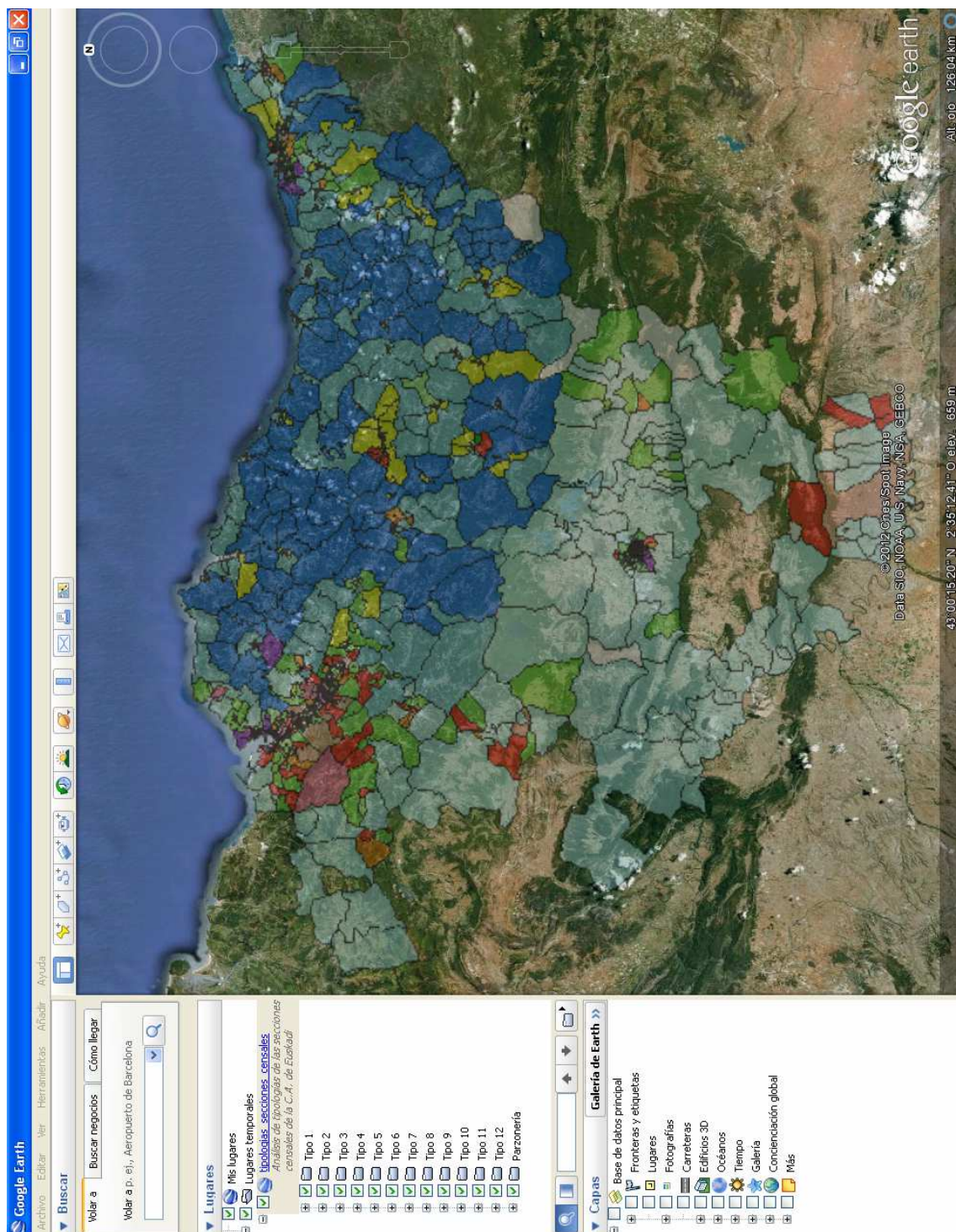
Por ejemplo, pulsando el cuadro desplegable de la izquierda el usuario podrá seleccionar solamente los tipos que le interesan. Por lo tanto, solo se colorearán en el mapa las secciones censales correspondientes a los tipos seleccionados, y al abrir en el cuadro desplegable dichos tipos, aparecerá la lista de las secciones clasificadas en ellos

En la medida en que nos aproximemos a una dirección, observaremos que la capa se vuelve cada vez más transparente, permitiendo ver las calles, las construcciones, los parques, etc. situados bajo la misma.

Pulsando sobre el mapa, aparece un cuadro de diálogo que contiene información sobre la sección censal (nombre del municipio; distrito; sección; tipo al que pertenece; porcentaje de habitantes menores de 16 años, mayores de 65 años, vascoparlantes y titulados superiores; link a la documentación del análisis de tipologías de la CAE de la página web de Eustat).

Bajo éstas líneas hemos recogido una imagen que muestra el aspecto que tiene la visualización del análisis de tipologías de las secciones censales de la CAE en Google Earth.





## OBSERVACIONES

En este caso, no es necesario especificar el directorio en el que se encuentra guardado el archivo shp antes de cargar la función ya que puede especificarse posteriormente, antes de ejecutar la función.

Podremos especificar la ubicación del directorio de las siguientes dos formas: pulsando los iconos "Archivo → Cambiar dir... → (carpeta correspondiente)" en R; o bien introduciendo el comando `setwd("dirección completa de la ubicación de la carpeta entre comillas")`. El archivo kml que se creará tras ejecutar la función se guardará también en esa dirección.

## ESTRUCTURA DEL ARCHIVO DE ENTRADA:

Para poder crear el archivo que colorea las secciones censales de la CAE según sus diferentes tipos o secciones, se precisa el archivo shp llamado shapefile.

El shapefile es un formato vectorial de almacenamiento digital, en el que se guarda la ubicación de los elementos geográficos y de sus atributos. Está formado por varios archivos informáticos. Concretamente, estos son los complementos del shp que se han utilizado aquí:

- .shp. Es un archivo que contiene las entidades geométricas de los objetos.
- .shx. Se trata de un archivo que contiene el índice de las entidades geométricas.
- .dbf. Base de datos que contiene información sobre los atributos de los objetos, en formato dBASE. Este archivo tiene la siguiente estructura:
  - Código de la sección, tanto el agregado como el desagregado (provincia, municipio, territorio y sección).
  - Cantidad total de individuos de la sección.
  - Tipo al que corresponde la sección.
  - Nombre del municipio en el que se encuentra la sección.
  - Variables que figurarán en el cuadro de diálogo: porcentaje de habitantes mayores de 16 años, mayores de 64 años, vascoparlantes, y titulados superiores.

NOTA: Se ha utilizado el archivo shp que contiene el seccionado de 2009, parzoneras incluidas.

## EJEMPLO

**KmlSortu**(n\_tipol=12, fitxategia="Seccion\_regionT.shp", gaztelania=TRUE, alpha=150)

## Conclusiones

La visualización de datos permite ver modelos, estructuras de datos, y tendencias imperceptibles a simple vista a través de tablas o ficheros. Gracias a los gráficos, podemos apreciar los modelos complejos que se ajustan a los datos y evaluar su utilidad. Precisamente, la visualización y síntesis de los datos refuerza su utilidad, y convierte los datos estadísticos en conocimiento.

Para aprovechar las ventajas que ofrece la visualización de datos, es necesario aplicar una metodología concreta y tener en cuenta unos principios básicos, especificados en el segundo capítulo.

Al representar un gráfico, se codifica la información cuantitativa y la categórica, normalmente, en base a la posición, al tamaño, a los signos y a los colores. Al analizar el gráfico, decodificamos esa información a simple vista, por lo que si el proceso de decodificación resulta efectivo, diremos que el método gráfico ha sido bueno.

Un método de visualización que conlleva a un proceso de decodificación a simple vista inefectivo puede hacer que no se detecten las principales características de los datos o que distorsione la percepción de la información.

En ocasiones, analizar los datos mediante herramientas de visualización permite evitar sorpresas y malentendidos que pueden surgir en la recopilación de datos.

El objetivo del trabajo era asignar los gráficos que mejor se ajusten a los diferentes tipos de datos de los que dispone Eustat. En ocasiones, ha sido difícil confeccionar un R-script general y parametrizado, al estar muy unido a las características de los datos del archivo.

A veces, hay que realizar ciertos cambios en el archivo antes de ejecutar los scripts, y el entorno de programación R ofrece una herramienta flexible para poder efectuar dichos cambios.

En el caso de los gráficos del API de Visualización de Google que ya están añadidos en la página web de Eustat, se pretende dejar a mano los códigos que crean dichos gráficos, para que los usuarios puedan añadirlos a sus blogs y páginas web.

La tecnología actual ha permitido hacer lo que hasta hace poco parecía imposible. Aún así, las estadísticas oficiales todavía están iniciándose en las posibilidades que ofrecen estas técnicas dentro de los ámbitos de la difusión y análisis de datos.

El trabajo realizado hasta la fecha en el ámbito de la visualización, y el que todavía está por hacer, tendrá consecuencias e influirá directamente en la utilización de la estadística oficial y en todos los usuarios.



## Bibliografía

- [1] YURRAMENDI, Y. (2011)

*Estatistika ofizialak sintetizatzeko eta ikustarazteko metodo grafikoen eta teknologia berrien azterketa.* Konputazio Zientziak eta Adimen Artifiziala saila.

- [2] CLEVELAND, W.S (1985)

*The Elements of Graphing Data.* Monterey, CA: Wadsworth.

- [3] YOUNG, F.W.; VALERO-MORA, P.M. AND FRIENDLY, M. (2006)

*Visual Statistics. Seeing Data with Dynamic Interactive Graphics.* Wiley, New Jersey.

- [4] GALBETE, E.; ADIN, A.; ARAMENDI, J.; IZTUETA, A. AND YURRAMENDI, J. (2012)

*Análisis de tipologías de las secciones censales de la CAE. Visualización en Google Earth.* IX Congreso Vasco de Sociología y Ciencia Política.

- [5] ESCOFIER B. AND PAGÈS, J. (1992)

*Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación.* Euskal Herriko Unibertsitatea, Argitalpen zerbitzua.

- [6] LÊ, S., JOSSE, J. AND HUSSON, F.. (Marzo de 2008)

*FactoMineR: An R Package for Multivariate Analysis.* Journal of Statistical Software, Volume 25, Issue 1.

- [7] EUSTAT (1996)

*Análisis de tipologías de las secciones censales de la C.A. de Euskadi.* Vitoria-Gasteiz: Instituto Vasco de Estadística.

- [8] EUSTAT (2001)

*Análisis de tipologías de las secciones censales de la C.A. de Euskadi.* Vitoria-Gasteiz: Instituto Vasco de Estadística.

- [9] GREENACRE (2008)

*La práctica de análisis de correspondencias.* Vitoria-Gasteiz: Instituto Vasco de Estadística. Fundación BBVA, Bilbao. Rubes Editorial.

- [10] **R Graph Gallery:** <http://gallery.r-enthusiasts.com/>
- [11] **R-Project:** <http://www.r-project.org/>
- [12] **FactoMineR:** <http://cran.r-project.org/web/packages/FactoMineR/index.html>
- [13] **Google Visualization API:** <https://developers.google.com/chart/>
- [14] **API. Wikipedia:**  
[http://es.wikipedia.org/wiki/Interfaz\\_de\\_programaci%C3%B3n\\_de\\_aplicaciones](http://es.wikipedia.org/wiki/Interfaz_de_programaci%C3%B3n_de_aplicaciones)
- [15] **Google Earth. Eduteca:** <http://www.eduteka.org/GoogleEarth.php>