

UNIVERSIDAD DE VALLADOLID

DISEÑO Y EVALUACIÓN DE SISTEMAS INTERACTIVOS

Análisis y Visualización de Datos MedioAmbientales de Castilla y León

Sergio García Prado

Seguimiento del trabajo en:
github.com/garciparedes/DatosMedioambientalesCYL

December 9, 2015

Contents

1	Introducción	2
1.1	Motivación:	2
1.2	Objetivos:	2
1.3	Estructura:	2
2	Análisis de los datos	2
2.1	Descripción	2
2.2	Adquisición	2
2.3	Estructura de los datos	2
2.4	Calidad	3
2.5	Selección de datos	4
2.6	Transformaciones	4
3	Planificación de la información	5
3.1	Propósito de la visualización	5
3.2	Factores que afectan al proyecto	5
4	Diseño	5
4.1	Introducción	5
4.2	Visualización de la relación entre producción y consumo	5
4.3	Visualización de producción	5
4.4	Visualización de consumo	5
5	Implementación	6
5.1	Bocetos:	6
6	Conclusiones	6
6.1	Conclusiones:	6
6.2	Aspectos a Resaltar	6
6.3	Cuestiones no tenidas en cuenta	6

1 Introducción

1.1 Motivación:

La visualización que se va a realizar consiste en la representación gráfica de datos medioambientales de la Comunidad Autónoma de Castilla y León. El objetivo de la práctica es descubrir aspectos poco obvios o destacar algo importante del conjunto de datos asignados.

1.2 Objetivos:

La representación gráfica que se va a realizar pretende mostrar el impacto que ha tenido el cierre de la Central Nuclear de Santa María de Garoña (Burgos) en el año 2012 y representar cómo ha solventado la región ese déficit energético (si es que lo ha hecho). También se pretende analizar si este suceso está ha afectado al consumo energético en la zona.

1.3 Estructura:

2 Análisis de los datos

2.1 Descripción

La Junta de Castilla y León suministra conjuntos de datos referentes a la comunidad autónoma. Para esta práctica se nos han asignado los datos medioambientales.

2.2 Adquisición

Los datos han sido obtenidos a partir de la iniciativa de Datos Abiertos de la Junta de Castilla y León.

El concepto "datos abiertos" es una filosofía y práctica que persigue que determinados tipos de datos estén disponibles de forma libre para todo el mundo, sin restricciones de derechos de autor, de patentes o de otros mecanismos de control. Tiene una ética similar a otros movimientos y comunidades abiertos, como el software libre, el código abierto y el acceso libre.

Los datos han sido obtenidos a partir del siguiente enlace: www.datosabiertos.jcyl.es/...

Una vez hemos obtenido los datos procederemos a realizar el análisis de los mismos.

2.3 Estructura de los datos

Los datos que vamos a manejar se encuentran en fichero con formato CSV (un tipo de documento en formato abierto sencillo para representar datos en forma de tabla, en las que las columnas se separan por comas (o punto y coma) y las filas por saltos de línea.), el cual a pesar de tener una estructura eficiente para ser manejado por máquinas, complica la lectura y extracción de información para los humanos. Además la cantidad de información contenida en este grande (más de 5000 elementos).

A pesar de estas dificultades, fijándonos en la cabecera del fichero, la cual nos da información de la estructura de los mismo, podemos conocer los campos nos encontraremos en cada una de las filas que contiene:

- **Indicador** Contiene información sobre a lo que se refiere el dato. Representa una variable de tipo cualitativo nominal. Tras realizar una extracción de datos, hemos obtenido que hay 49 indicadores distintos. Debido a que la cantidad es numerosa habrá que tratar de seleccionar los que más se adecuen a lo que queremos mostrar debido a que mostrar demasiada información de manera inadecuada dificulta la visión de lo que se pretende representar.
- **Provincia** Provincia a la que está referida la información. Este campo es de tipo cualitativo nominal. Se refiere a la segmentación territorial de la comunidad autónoma. Como sabemos, Castilla y León está dividida en 9 provincias. Conviene señalar que hay indicadores que no poseen información para todas las provincias, como es el caso del "Tráfico aéreo de pasajeros", que suponemos que restringe la información a las provincias que disponen de aeropuerto ya que el resto tendrían valor 0 y no tiene sentido indicarlo. Este hecho es algo a tener en cuenta cuando representemos la información.
- **Fecha Validez** Fecha hasta la cual el dato se consideraba como actual. Representa una variable cuantitativa debido a que es una variable temporal. A pesar de ello toma únicamente valores discretos (años). Esto es algo a tener en consideración cuando determinemos la manera en que será representado.
- **Valor** Valor que toma el dato respecto de su unidad. Representa una variable cuantitativa de tipo continuo. Este campo es el que contiene la "información valiosa", ya que es lo que se utilizará como medida para compararla con los demás. Se representa en forma de número racional positivo.
- **Unidad** Este campo nos proporciona la unidad de medida del dato, lo cual es algo muy importante a la hora de mostrar información ya que siempre se deben señalar tanto la unidad de medida como la escala a la que corresponde la información representada.
- **Frecuencia** Este campo representa la periodicidad con la que se ha obtenido el dato. La mayoría de los datos se han obtenido anualmente, aunque también existen algunos que se han obtenido bianualmente, como es el caso de los "Espacios Naturales".

2.4 Calidad

A mi juicio la calidad de los datos es buena a pesar de que existen algunos puntos mejorables.

Me llama la atención la columna de "meses" que, entiendo que está ahí porque es necesario especificar la unidad de medida de la frecuencia, pero creo que se debería haber indicado en la cabecera del fichero. Además también he encontrado valores atípicos en algunos datos, posiblemente producidos por un error en las mediciones o al transcribirlo al documento.

2.5 Selección de datos

Ahora determinaremos los datos que se utilizarán para la visualización. Como se dijo al principio del informe, el objetivo de la visualización será visualizar cómo ha repercutido el cierre de la central nuclear en la región. Para ello nos vamos a apoyar en los siguientes indicadores:

- Producción de energía con carbón
- Producción de energía eólica
- Producción de energía hidráulica
- Producción de energía nuclear
- Producción de energía primaria
- Producción energía solar en Castilla y León
- Consumo de energía final
- Consumo de energía del sector del transporte
- Consumo de energía del sector
- Consumo doméstico de electricidad
- Consumo doméstico de gas natural
- Consumo doméstico de G.L.P.
- Consumo doméstico de productos petrolíferos

De cada uno de estos indicadores se pretenden utilizar tanto la información de todas las provincias como fechas de validez.

2.6 Transformaciones

Para simplificar la visualización realizaremos varias transformaciones en los datos:

Fusionaremos estos indicadores en uno nuevo que denominaremos "Producción de energía final" cuyo valor será el sumatorio de todos los métodos de producción (Carbón, Eólica, Hidráulica, Nuclear, Primaria y Solar). La unidad de medida de todos ellos es "Toneladas Equivalentes de Petróleo" por lo cual no habrá que realizar ningún cambio de unidad para poder fusionarlos.

También fusionaremos los indicadores relacionados con el consumo doméstico en uno nuevo que denominaremos "Consumo Doméstico" cuyo valor será el sumatorio de estos (Electricidad, Gas Natural, G.L.P. y Productos Petrolíferos). En este caso la información no tiene la misma unidad de medida para todos ellos, por lo que habrá que adaptar cada uno de ellos a la misma (Toneladas Equivalentes de Petróleo) para luego crear el nuevo indicador a partir de estos. El motivo de convertirlos en Toneladas Equivalentes de Petróleo nos servirá para luego compararlos con la producción.

3 Planificación de la información

3.1 Propósito de la visualización

El proyecto de visualización se va a enfocar en el sector energético. El objetivo de la visualización es mostrar la relación entre la producción y el consumo energético de la comunidad autónoma. Además de mostrar la proporción tanto de los diferentes métodos de producción como de las formas de consumo y cómo estas han variado durante el tiempo.

También se pretende obtener conclusiones acerca de si el déficit producido por el cierre de la central nuclear de Santa María de Garoña (Burgos) en la producción energética de la comunidad autónoma ha sido compensado con el incremento de la producción mediante otros métodos alternativos. También queremos analizar si este déficit ha afectado al consumo energético, que se prevee que no ya que probablemente se habrá importado de comunidades colindantes.

3.2 Factores que afectan al proyecto

El consumo energético no está únicamente influenciado por la producción energética ya que depende muchos otros factores como la situación económica, el estado atmosférico, la actividad industrial, etc.

4 Diseño

Se proponen varias alternativas para la visualización.

4.1 Alternativa 1

4.1.1 Introducción

La visualización que se pretende implementar tiene como objetivo ser un "widget" que se acoplará a una página web para poder visualizar la relación entre producción y consumo, así como el desglose de estas dos variables para conocer la procedencia de cada una de ellas. La estructura de componente será de 3 pestañas (Relación entre Producción y Consumo, Producción y Consumo). Se ha escogido esta estrategia ya que mostrar toda la información a la vez además de ser menos atractiva visualmente, puede dificultar el entendimiento de la información. A continuación procederemos a detallar cada una de las pestañas.

4.1.2 Visualización de la relación entre producción y consumo

Esta visualización será la que englobará toda la información. Para ello se utilizarán los indicadores "Consumo de energía final" y "Producción de energía final" (anteriormente hemos descrito cómo obtenerlo). La visualización se basará en dos gráficos:

- Map Char

4.1.3 Visualización de producción

4.1.4 Visualización de consumo

Para simplificar la visualización se agregará un mapa de Castilla y Leon el cual represente la diferencia por provincias y así sea sencillo distinguir cuales son las que más cantidad aportan, tanto de producción como de consumo.

Para ello la visualización contará con 3 "secciones" referidas a la producción, al consumo y a la relación entre estas dos variables respectivamente.

Las visualizaciones de producción y consumo serán similares. Por lo tanto se explicará el gráfico de producción y al de consumo se aplicará la misma idea con sus respectivas variables.

Se pretende que contenga el mapa citado anteriormente junto con un "Stacked Area Chart" con el que se mostrará la aportación de cada una de las formas la producción respecto del tiempo. Para la producción serán: Producción de energía con carbón, eólica, hidráulica, nuclear, primaria y solar. Con este tipo de gráfico se puede visualizar las variaciones que estas han tenido durante el tiempo, además de la aportación que hace cada una de ellas a la producción total de energía.

Respecto de la visualización de la relación entre producción y consumo hay varios factores a tener en cuenta.

5 Implementación

5.1 Bocetos:

6 Conclusiones

6.1 Conclusiones:

6.2 Aspectos a Resaltar

6.3 Cuestiones no tenidas en cuenta