

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 y Transformaciones	3
3	Planificación de la información	5
3.1	Propósito de la visualización:	5
3.2	Factores que afectan al proyecto:	6
3.3	Enfoque:	6
4	Diseño	6
4.1	Forma de representación:	6
4.2	Decisiones de presentación:	6
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 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 cómo la región solventado ese déficit energético en la actualidad (si es que lo ha hecho). También se pretende analizar si este suceso está correlacionado con el 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 tenemos 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 CSV, 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 datos contenida en este es elevada (más de 5000 elementos).

A pesar de estas dificultades si que podemos ver la cabecera del fichero, la cual nos da información de la estructura de los mismo y qué campos nos encontraremos en cada las entradas 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 muy 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 mostrar.
- **Provincia** Provincia a la que está referida el dato. 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 destacar que hay indicadores que no poseen información para todas las provincias, como es el caso del "Tráfico aéreo de pasajeros", que suponemos restringe la información a las provincias que disponen de aeropuerto. Este hecho es algo a tener en cuenta cuando representemos los datos.
- **Fecha Validez** Fecha hasta la cual el dato se toma como válido. Representa una variable cuantitativa. La causa es que es una variable temporal. A pesar de ello toma únicamente valores discretos (años) Esto es algo a tener en cuenta cuando elijamos la manera en que lo representaremos.
- **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" del dato, ya que es la unidad que se utilizará como medida. 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 que se suministra información es muy importante señalar tanto la unidad de medida como la escala a la que corresponde la representación.
- **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 que se especifique 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 y Transformaciones

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 a 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 pretende utilizar los datos de todas las provincias así como fechas. Cabe destacar que fusionaremos los siguientes indicadores para simplificar la visualización:

- 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

Fusionaremos estos indicadores en uno nuevo llamado Producción energética cuyo valor será el sumatorio de estos. 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 combinarlos en el nuevo indicador.

- Consumo doméstico de electricidad
- Consumo doméstico de gas natural
- Consumo doméstico de G.L.P.
- Consumo doméstico de productos petrolíferos

Fusionaremos estos indicadores en uno nuevo llamado Consumo Doméstico cuyo valor será el sumatorio de estos. En este caso los valores no tienen la misma unidad de medida, por lo que habrá que adaptar cada uno de ellos a la misma para luego crear el nuevo indicador a partir de estos.

3 Planificación de la información

3.1 Propósito de la visualización:

El proyecto de visualización se va a enfocar en el sector energético. Se pretende sacar conclusiones sobre cómo ha afrontado el territorio el cambio hacia nuevos métodos de producción energética con respecto al paso del tiempo, en concreto con el cierre de la central nuclear de Santa María de Garoña (Burgos).

Se pretende relacionar el consumo y la producción energética durante el tiempo dividiendo la información por provincias. Para simplificar la visualización se agregará un mapa de Castilla y León el cual represente la diferencia por provincias y así sea sencillo distinguir cuáles 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 de 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.

3.2 Factores que afectan al proyecto:

3.3 Enfoque:

4 Diseño

4.1 Forma de representación:

4.2 Decisiones de presentación:

5 Implementación

5.1 Bocetos:

6 Conclusiones

6.1 Conclusiones:

6.2 Aspectos a Resaltar:

6.3 Cuestiones no tenidas en cuenta: