

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 3, 2015

Contents

1	Introducción	2
2	Análisis de los datos	2
2.1	Estructura de los datos	2
3	Planificación de la información	3
4	Diseño	3
5	Implementación	3
6	Conclusiones	3

1 Introducción

El análisis consiste en la obtención de aspectos poco obvios que se ocultan en los datos debido a la gran cantidad que manejaremos. Los datos han sido obtenidos a partir de la iniciativa de Datos Abiertos de la Junta de Castilla y León, concretamente estos se refieren a Datos Medioambientales.

2 Análisis de los datos

2.1 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 del mismo. Además gracias a la herramienta D3.js se ha obtenido información acerca de ellas. El fichero tiene las siguientes entradas:

- **Indicador** Contiene información sobre a lo que se refiere el dato.
- **Provincia** Provincia a la que está referida el dato. Como sabemos la Comunidad Autónoma de Castilla y León está dividida en 9 provincias, por lo tanto a falta de comprobar que los datos se refieran únicamente a esta CA ya podemos deducir que esta entrada se refiere a información categórica. 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. Este campo se limita a mostrar el año de validez. Aunque esto se pueda pensar como categorías no es la manera correcta, ya que el tiempo es una variable continua. 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. 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.
- **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".

- 3 Planificación de la información
- 4 Diseño
- 5 Implementación
- 6 Conclusiones