



GEOBIG

Gestión de grandes volúmenes de datos abiertos

Alejandro.guinea@Geograma.com

- Introducción
- Objetivo y Fases
- Identificación
- Preproceso y carga
- Transformación
- Análisis
- Conclusiones y próximos pasos

- Disponibles datos de alto valor
- Ya procesados
- La transformación digital requiere procesos simples y maduros
- Series de datos

Objetivo: Organizar los datos para poder aplicar algoritmos de indexación y clasificación

Fases:

1. La localización de los conjuntos de datos geográficos.
2. La transformación para su uso inmediato.
3. La clasificación y resolución de cuestiones semánticas.
4. La monitorización y actualización de los datos.

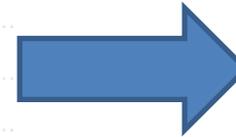
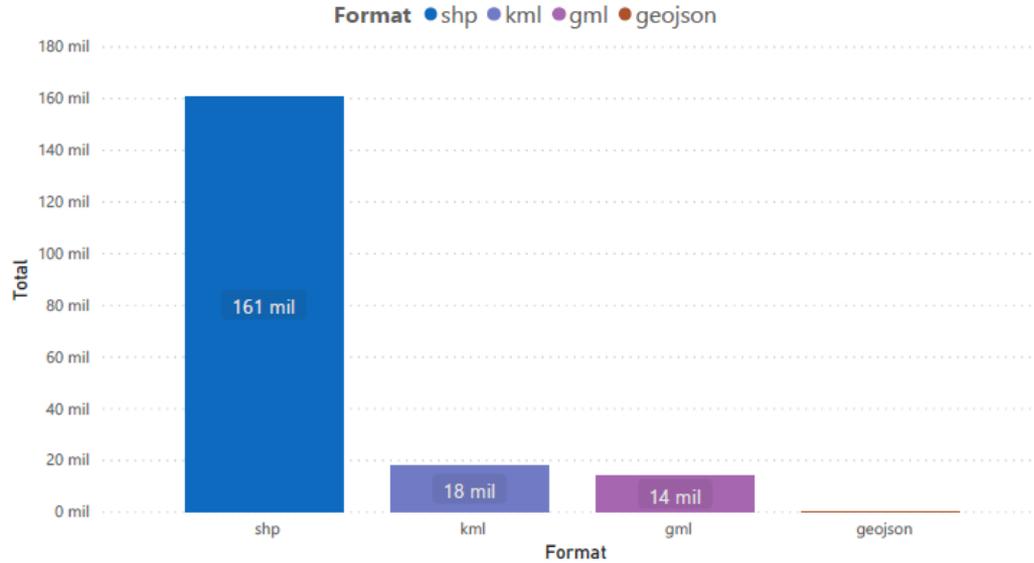
- Encontrar los datos
- Descargarlos
 - Servidor HTTP.
 - Servidor FTP.
 - Aplicación web del tipo “añadir a la cesta”.
 - Interfaz gráfico con mapas para seleccionar la información.
 - Envío por correo electrónico.
 - Descarga fichero a fichero.
- Autenticación
- Ver el interior

- Revisión del contenido de los conjuntos de datos, para asegurarse de que contienen información geográfica procesable.
- Confirmación de la disponibilidad de los datos, y de la correcta descompresión de los mismos.
- Captura de metadatos adicionales disponibles en la URL
- Comprobación y registro de volumen y fecha.

40.000
conjuntos de datos

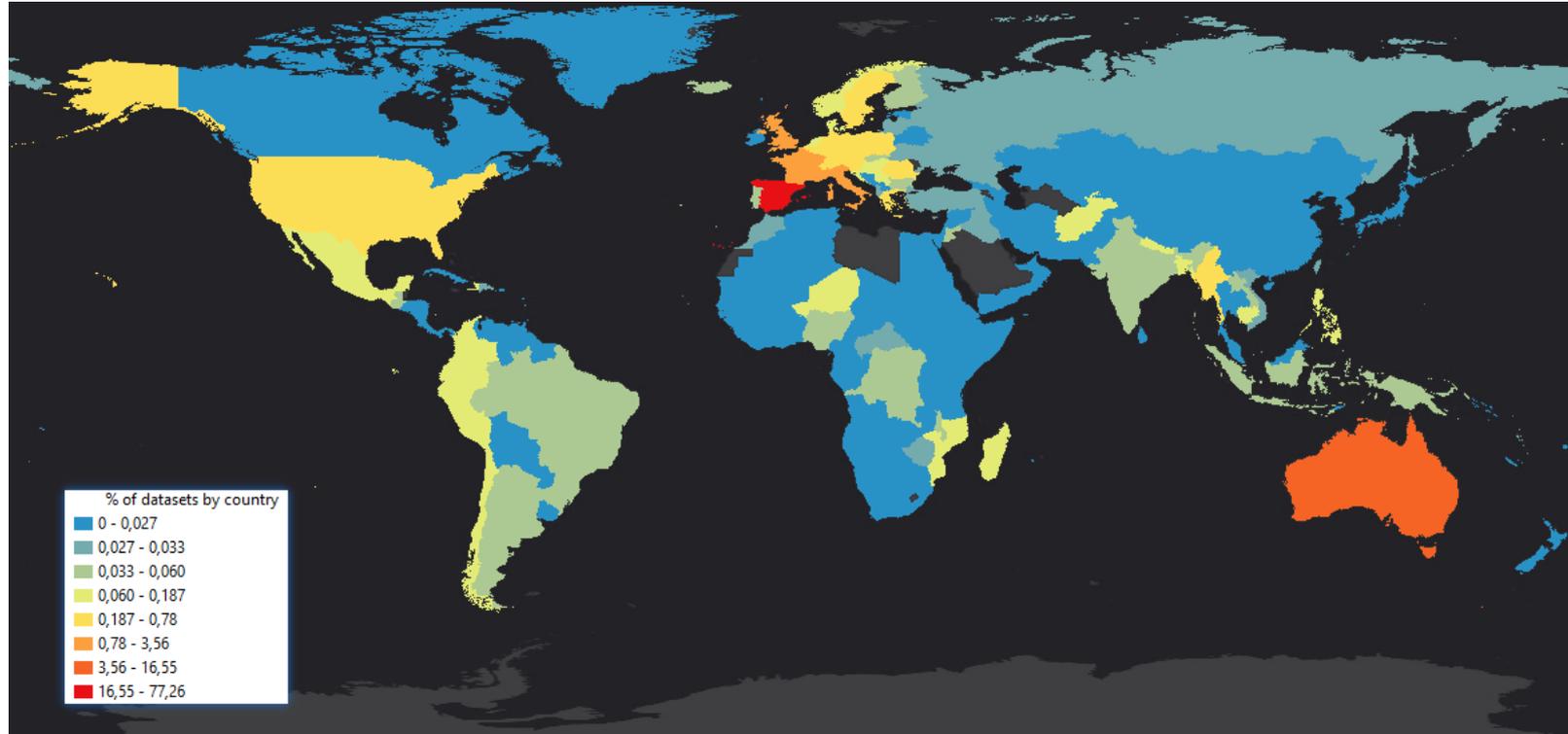
Procesos paralelos

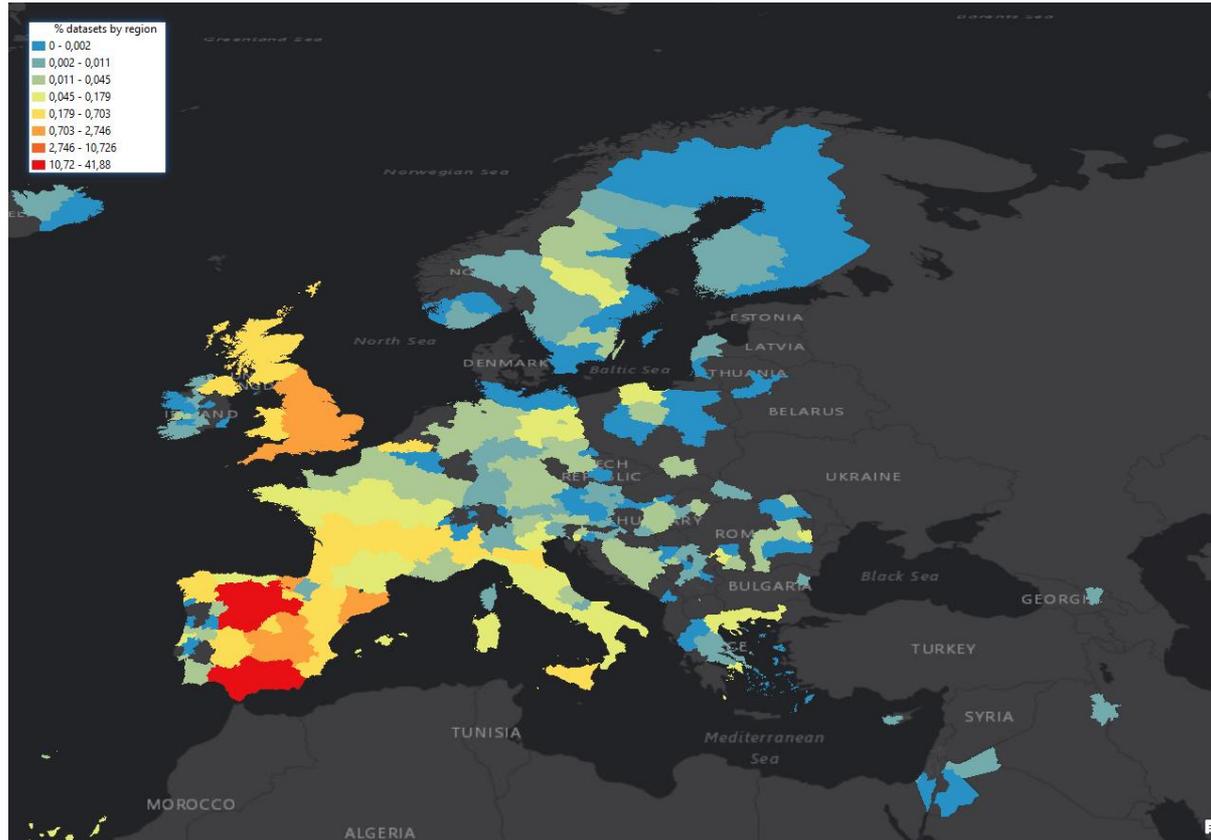
Transformación

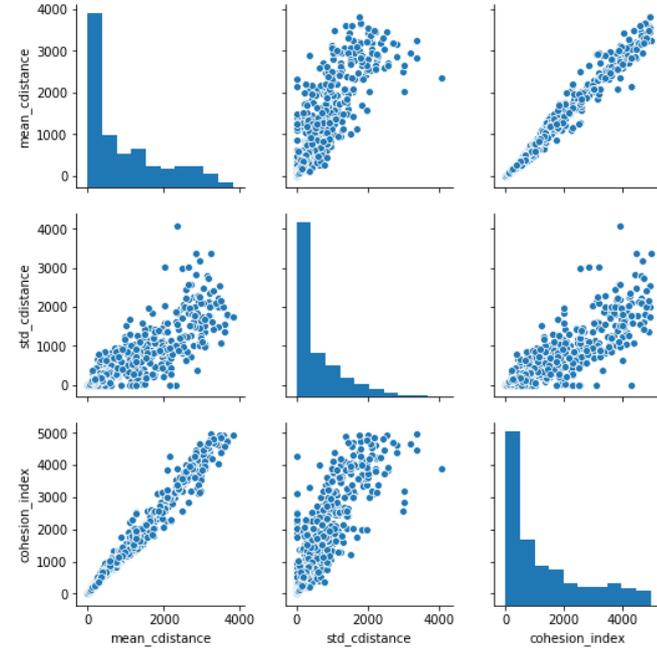
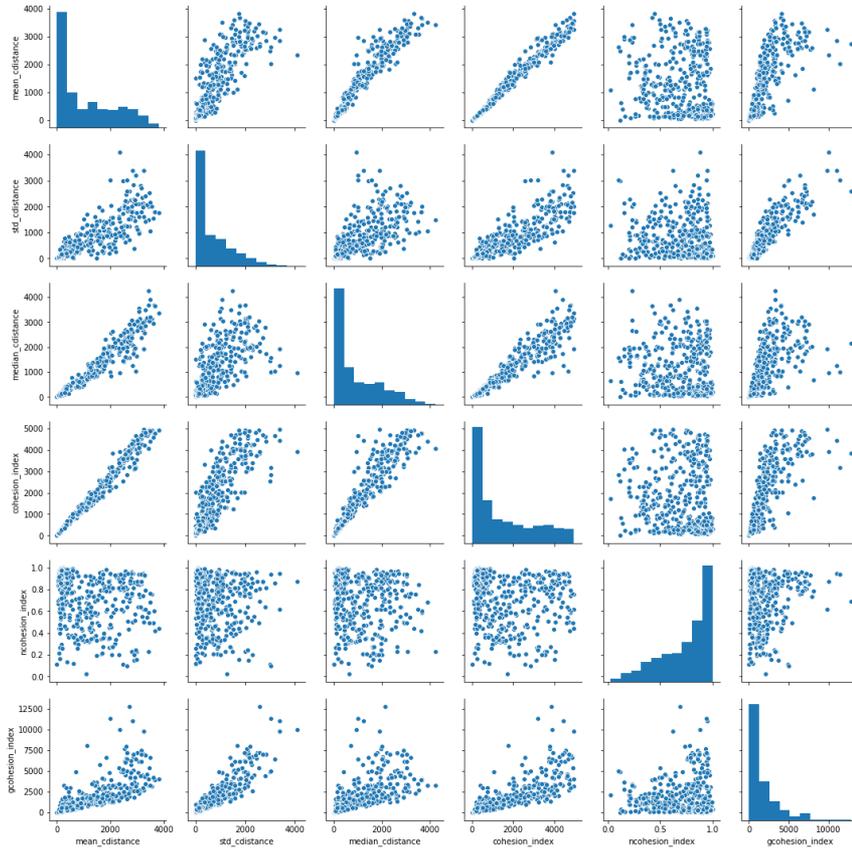


BD
Relacional

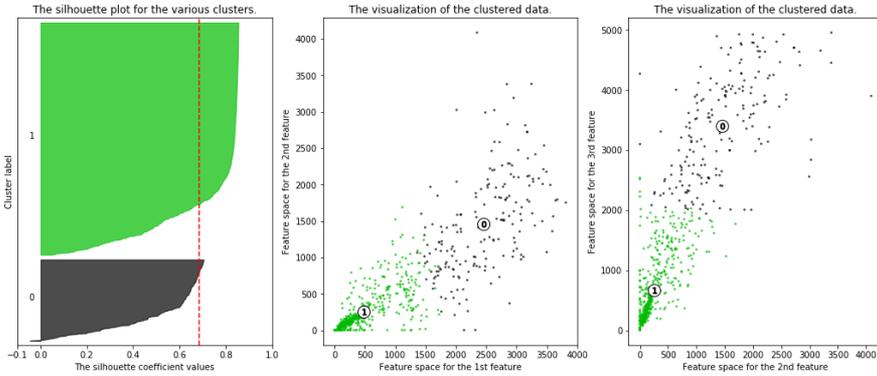
200.000
capas



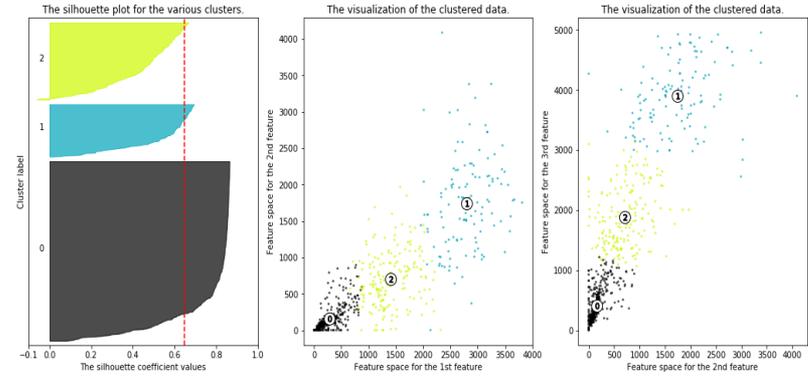




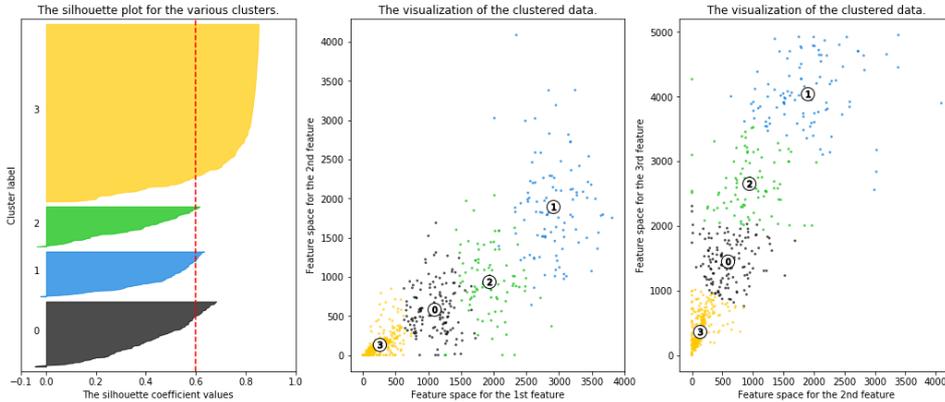
Silhouette analysis for KMeans clustering on sample data with n_clusters = 2



Silhouette analysis for KMeans clustering on sample data with n_clusters = 3



Silhouette analysis for KMeans clustering on sample data with n_clusters = 4



700 capas

300MM elementos

Conclusiones y próximos pasos

- IG ordenada
- Identificación
- Nuevos atributos geométricos
- Clusterización
- Semántica de los atributos

Gracias por su atención

Alejandro.guinea@Geograma.com