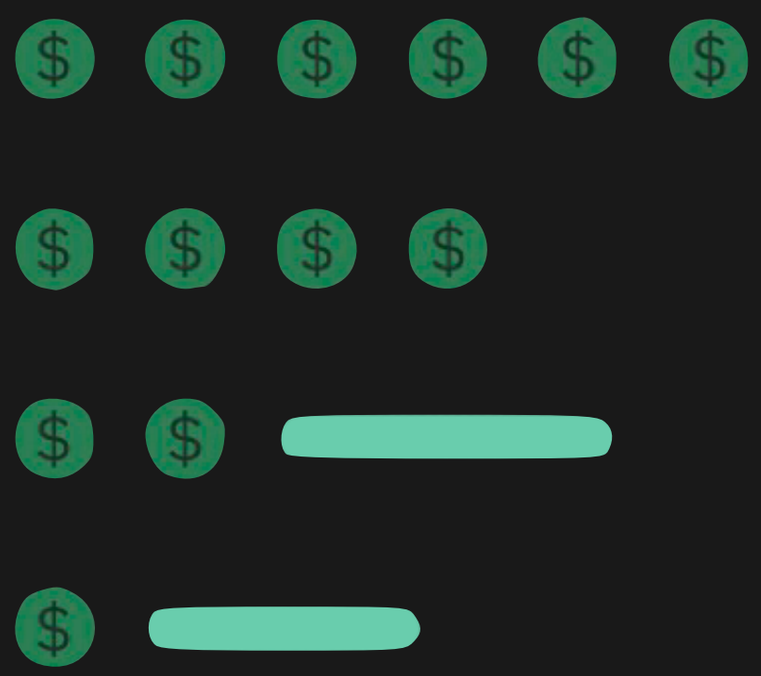
Corrupción en Chile

Analizando datos de corrupción



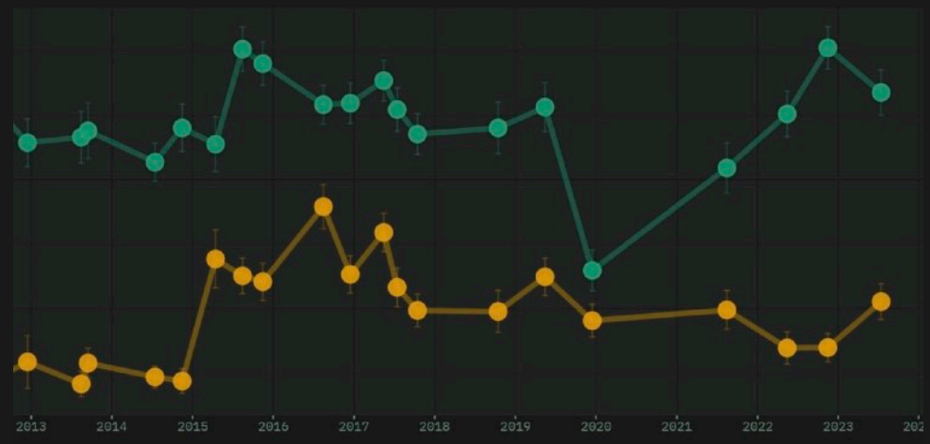
Bastián Olea Herrera

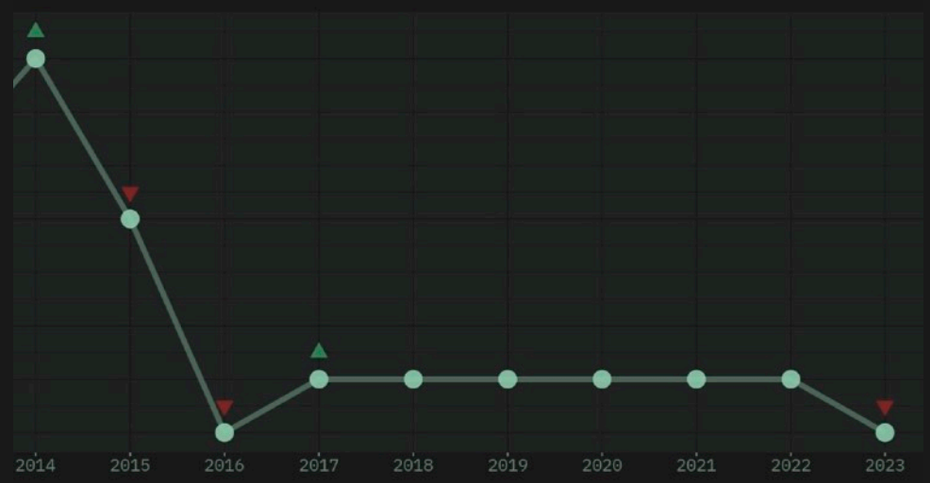
Visualizador de corrupción en Chile

- https://github.com/bastianolea/corrupcion_chile

- ***Metodología***
  - Casos de corrupción donde se involucren recursos públicos

- ***Características***
  - Serie de gráficos actualizados con datos más recientes
  - Énfasis en política



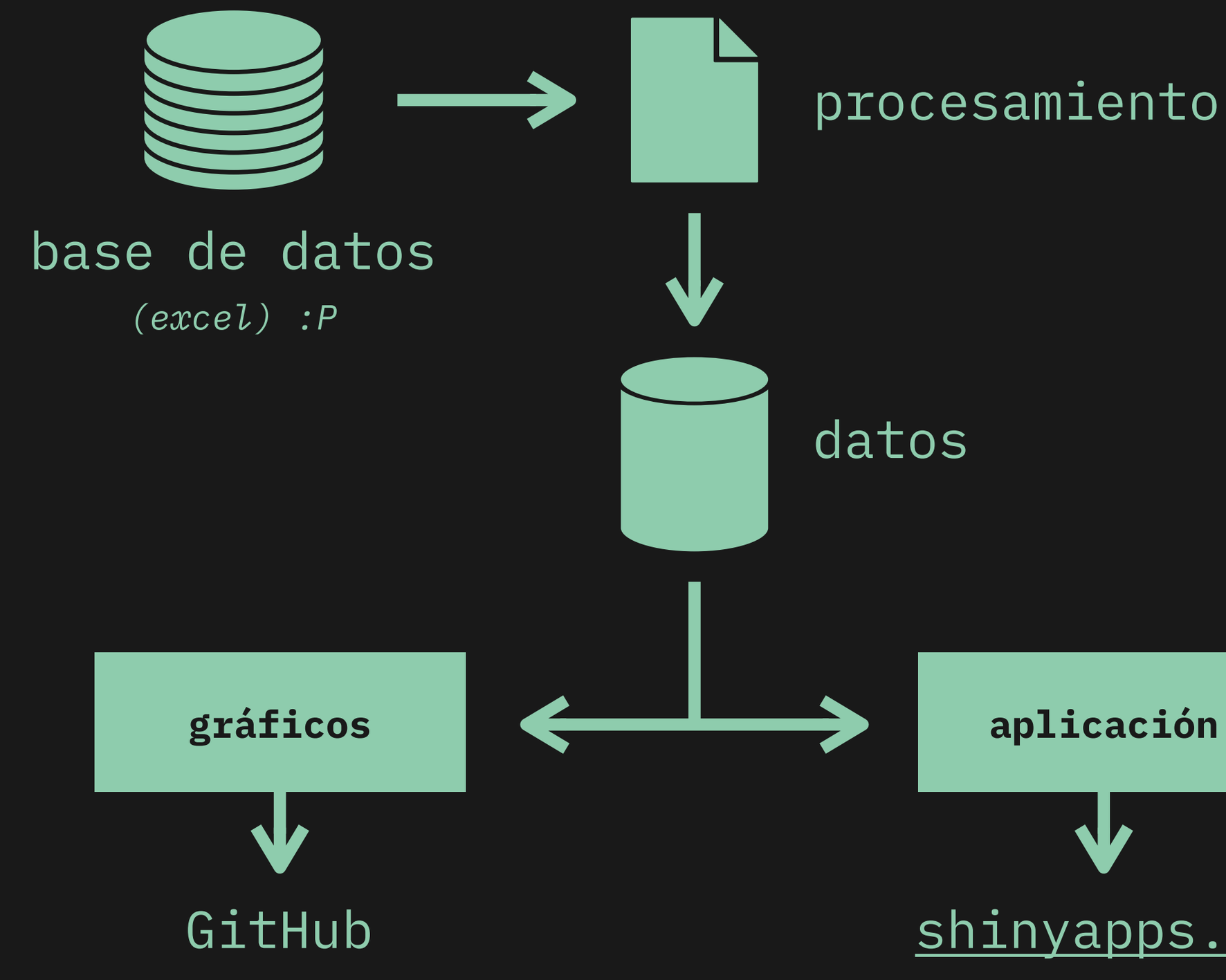


Stack

Tecnologías usadas en la aplicación

- Aplicación programada con el lenguaje **R**
- Framework de aplicaciones web **Shiny**
- Librería de visualización de datos **{ggplot2}**
- Hosting en la nube shinyapps.io

Flujo *Proceso de despliegue*



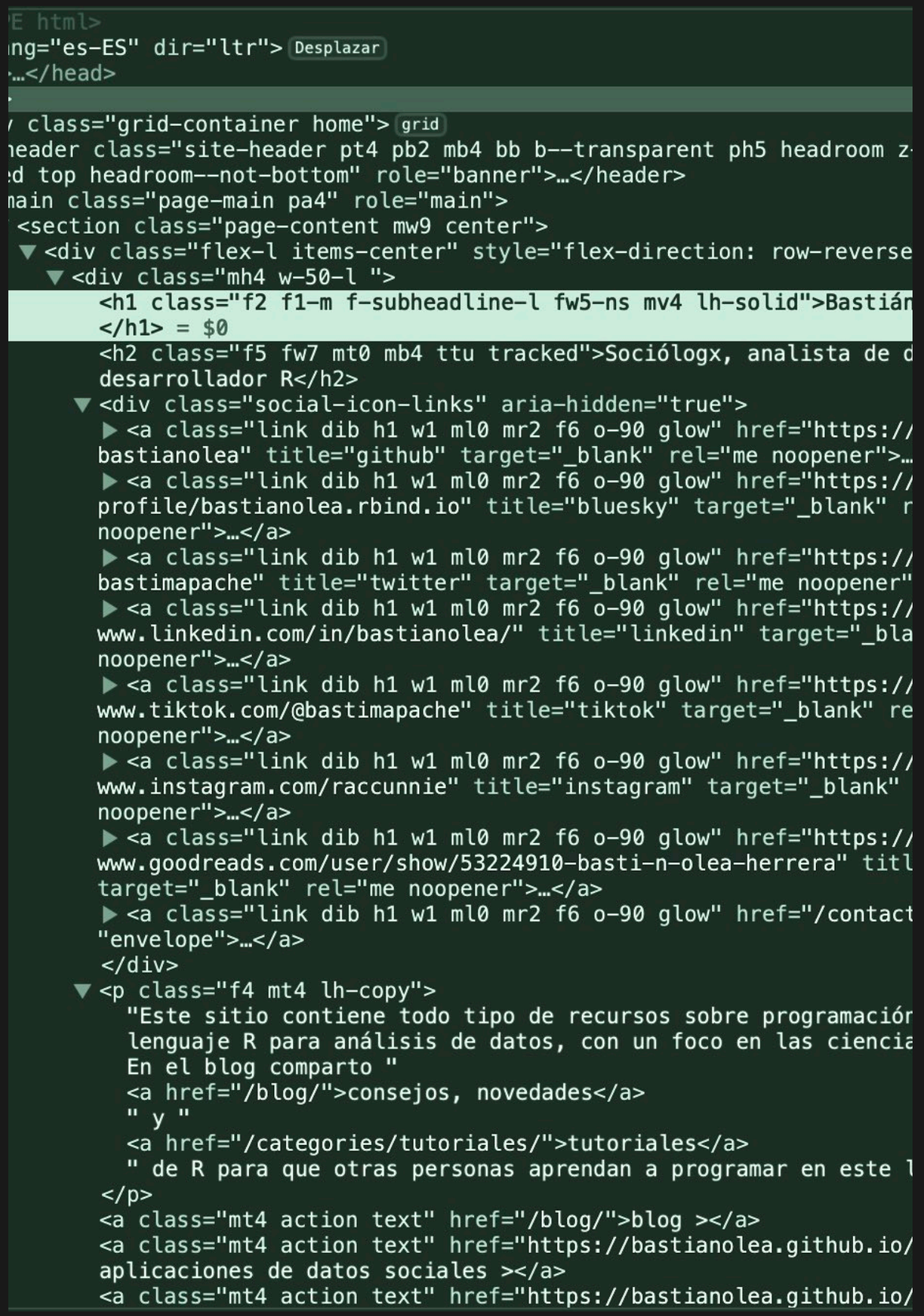
Transparencia

Perspectiva ante una realidad

- Datos abiertos
- Código abierto
- Tecnologías gratuitas
- Participación social
- Exponer información ofuscada

Web scraping

Extracción de datos desde páginas web



¿Qué es el web scraping?

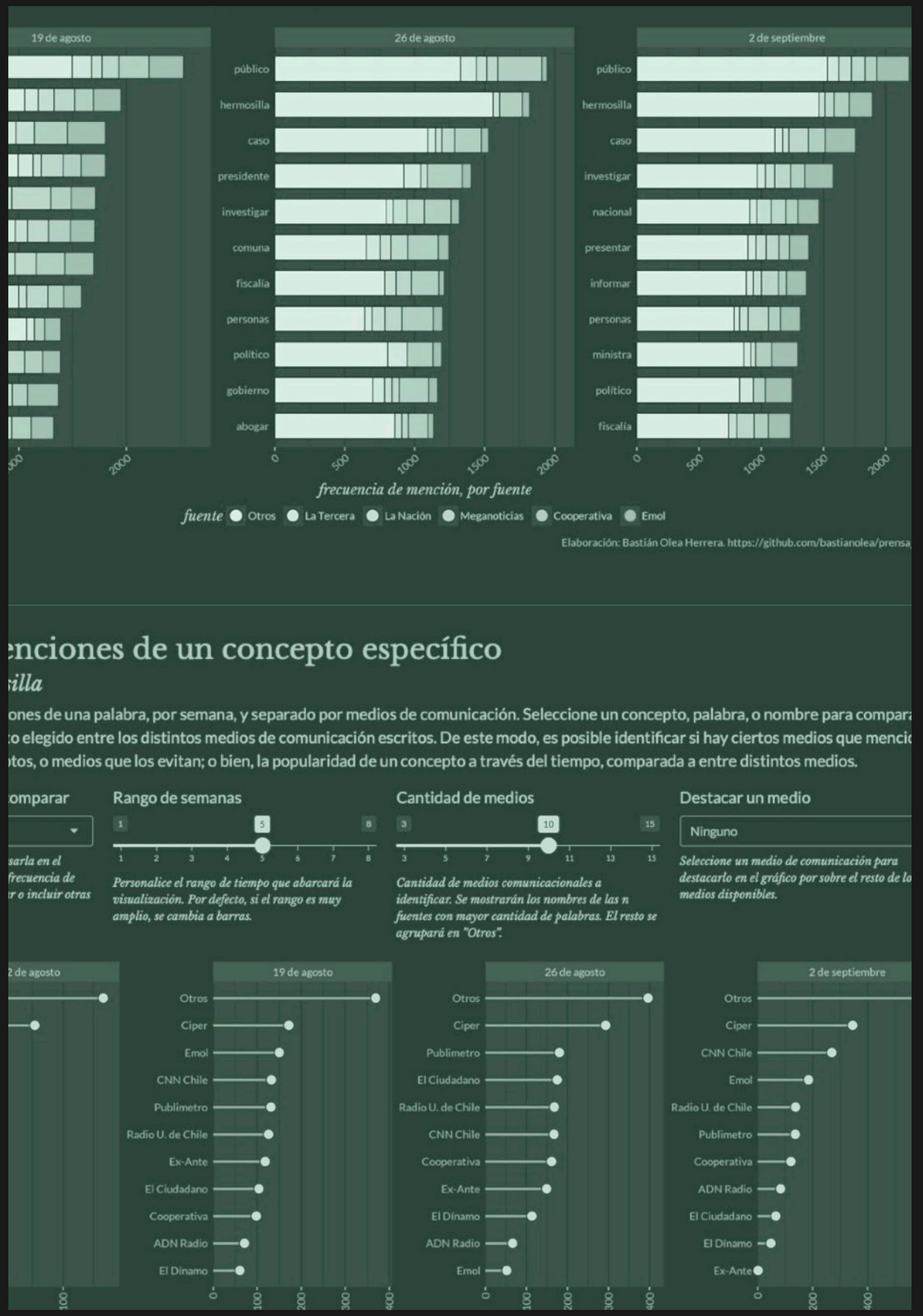
Extracción de datos desde sitios web

- Se utiliza un programa para obtener el código fuente de un sitio web
- Se usan selectores para identificar elementos de un sitio web
- Extracción de datos y organización tabular de los mismos
- Análisis de los datos

> *Ejemplo práctico*

Scraping de prensa chilena

Aplicación de análisis de datos de texto de prensa chilena

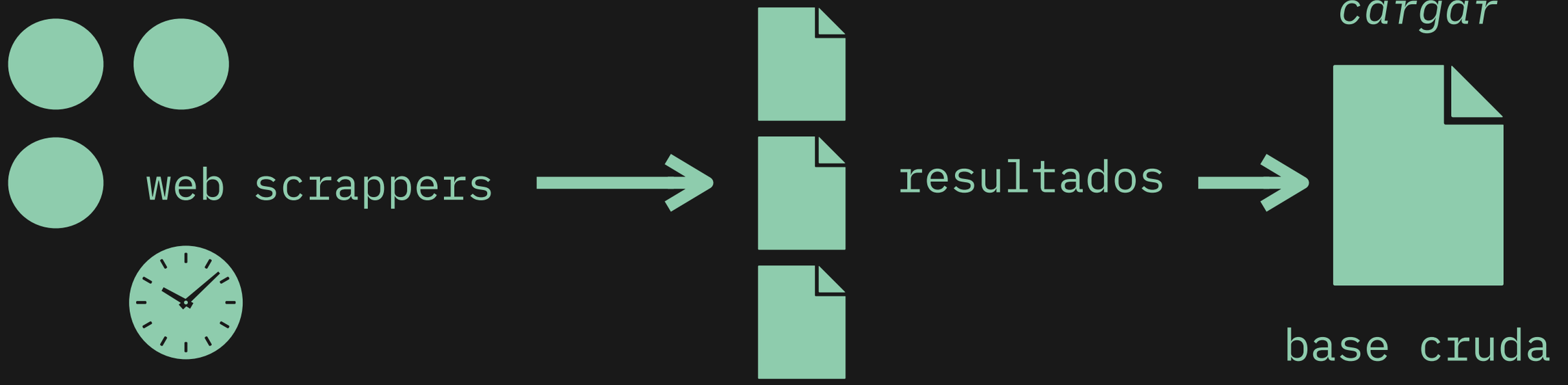


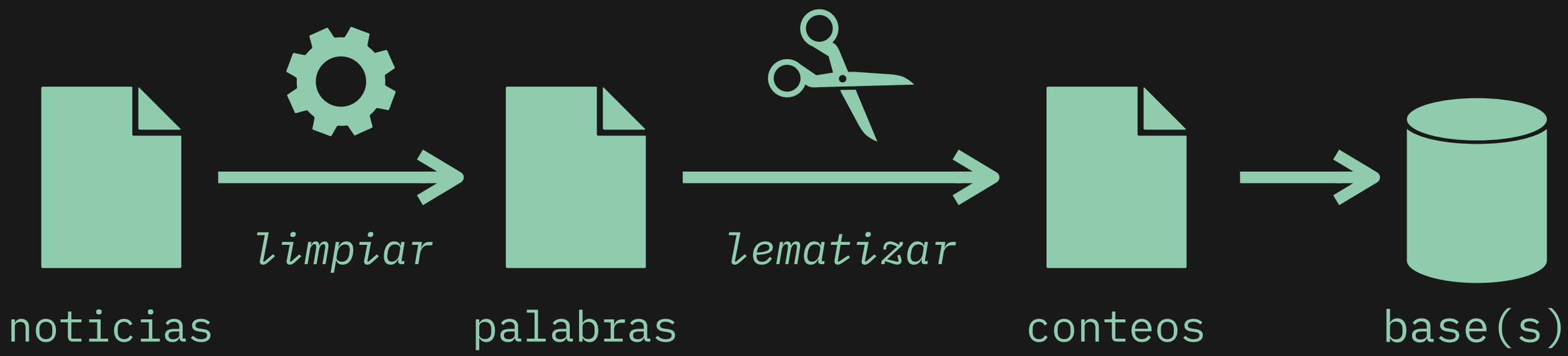
Scraping de prensa chilena

- Web scraping y análisis de texto sobre un corpus de texto de noticias de la prensa chilena, con nuevos datos obtenidos diariamente, y app actualizada semanalmente

- https://github.com/bastianolea/prensa_chile

Flujo *Obtención y procesamiento*





https://bastianoleah.shinyapps.io/prensa_chile/

Procesamiento de texto

- Unión y limpieza de noticias
- Tokenizar texto y conteo de frecuencias
- Legalización y conteo de palabras lematizadas
- Cálculo de correlación entre términos
- Modelamiento de tópicos
- Detección de sentimiento
- Generación de resúmenes de noticias
- Actualización de aplicación

Modelamiento de tópicos

Detección de temas emergentes

- Modelamiento estructural de tópicos (STM)
- Busca modelar tópicos en base a covariantes a nivel de documento
- Puede incluir metadatos
- Permite obtener tópicos latentes desde los datos
- Obtiene el número de tópicos (K) que pueden existir en el texto probando modelos

> *Ejemplo de modelamiento de datos*

> *Ejemplo de análisis de texto*