

# Metodología para la visualización interactiva de datos de la Unión Europea

Roberto Barroso García

Máster universitario en análisis y  
visualización de datos masivos

# Índice

- ▶ Introducción
  - ▶ Descripción
  - ▶ Motivaciones
  - ▶ Objetivos
- ▶ Estado del arte
  - ▶ Trabajos previos
  - ▶ Soluciones propuestas

# Índice

- ▶ Desarrollo de la metodología
  - ▶ Proceso de extracción, tratamiento y carga de los datos
  - ▶ Visualizaciones interactivas
  - ▶ Despliegue de la aplicación
- ▶ Resultados
  - ▶ Demostración
  - ▶ Evaluación
- ▶ Conclusiones
- ▶ Líneas futuras

# Introducción - Descripción

- ▶ Metodología para la visualización interactiva de datos referentes a la Unión Europea.
- ▶ Comparación de datos económicos, sociales y relacionados con la energía, salud y bienestar entre otros.
- ▶ Aplicación web donde se exponen las visualizaciones, divididas en categorías correspondientes a los temas tratados.

# Introducción - Motivaciones

- ▶ Singularidad del tema propuesto.
- ▶ Labor de la UE cuestionada.
- ▶ Reunir información dispersa en otros medios.
- ▶ Datos objetivos al alcance de todos.
- ▶ Facilitar la extracción de conocimiento.
- ▶ Resolver problemas observados en otras metodologías similares.

# Introducción - Objetivos

- ▶ Lograr que el usuario sea capaz de adquirir conocimiento de una forma sencilla e independiente.
- ▶ Establecer los temas que se van a tratar e identificar las posibilidades de captura de los datos.
- ▶ Determinar y describir todos los datos.
- ▶ Desarrollar las visualizaciones interactivas que permitan al usuario adquirir conocimiento.
- ▶ Evaluar la metodología propuesta en un escenario real.

# Estado del arte – Trabajos previos

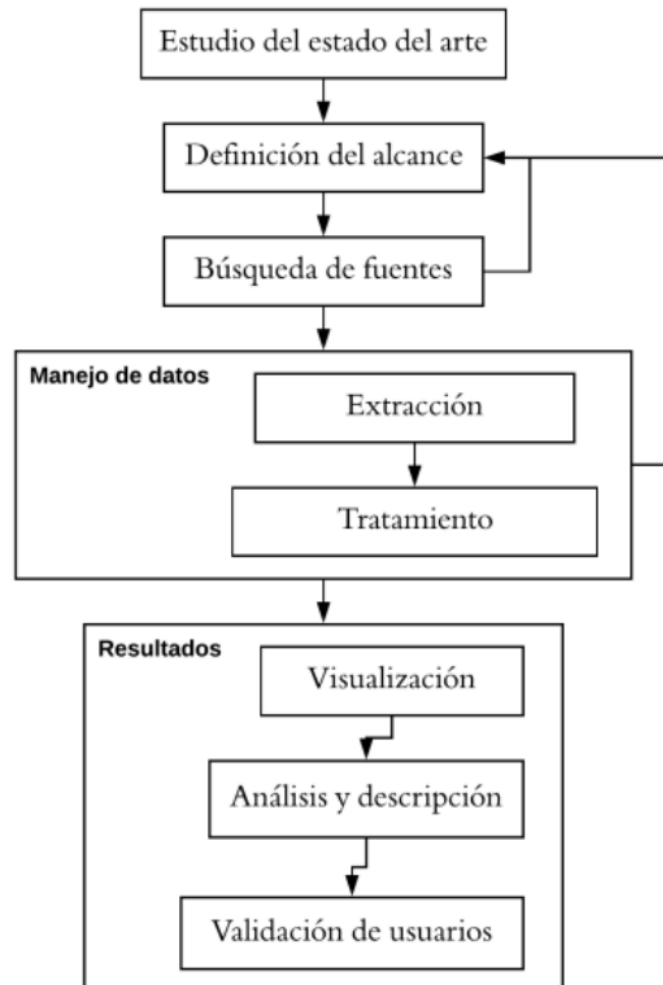
- ▶ Principales observaciones:
  - ▶ Ambigüedad
  - ▶ Falta de contexto
  - ▶ Falta de completitud
  - ▶ Subjetividad
  - ▶ Datos puntuales, sin perspectiva de tiempo
  - ▶ Temática no agrupada por categorías
  - ▶ Visualizaciones estáticas y poco llamativas

# Estado del arte – Soluciones propuestas

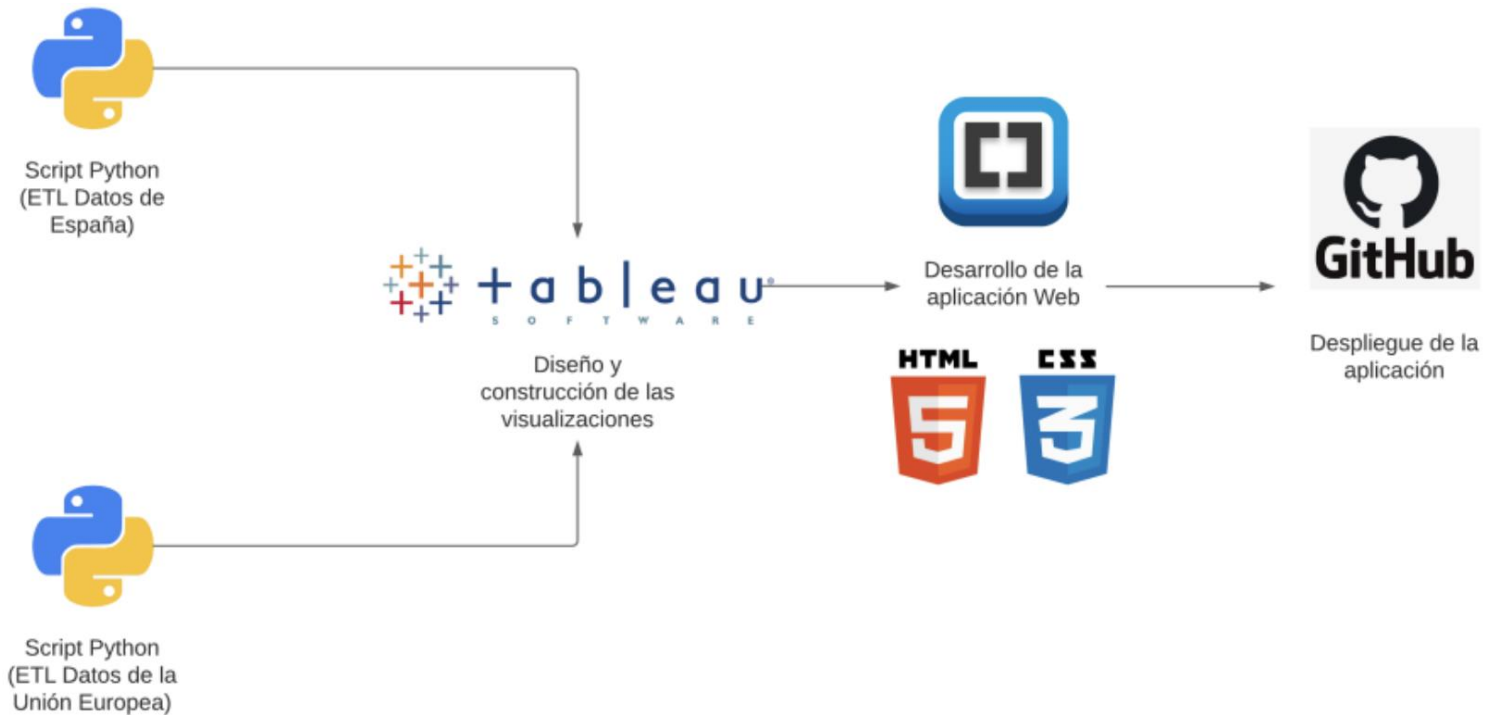
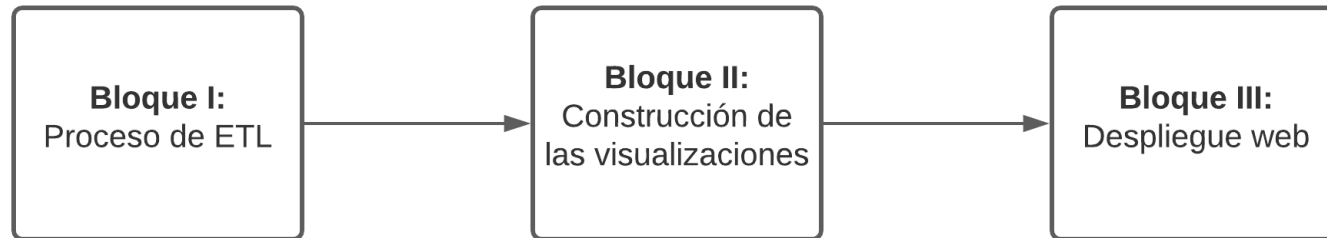
- ▶ Simplicidad y claridad en las visualizaciones, cómodo, intuitivo y funcional.
- ▶ Relacionar las visualizaciones entre si y división en categorías.
- ▶ Datos completos y de fuentes fiables
- ▶ Datos objetivos
- ▶ Añadir evolución en el tiempo
- ▶ Visualizaciones interactivas y estéticas



# Desarrollo de la metodología



# Desarrollo de la metodología



# Desarrollo de la metodología – Proceso ETL

- ▶ Fuentes de los datos:
  - ▶ OurWorldInData
  - ▶ DatosMacro
- ▶ Tecnologías utilizadas:
  - ▶ Python v3.10
    - ▶ Pandas
    - ▶ BeautifulSoup
    - ▶ Request
  - ▶ Jupyter Notebook
- ▶ Carga por Web Scraping y carga desde archivo de datos

# Método Web Scraping

URL donde se encuentran los datos

```
url="https://datosmacro.expansion.com/pib/espana"
```

Petición GET con el "crudo" del contenido HTML

```
html_content = requests.get(url).text
```

Extracción del contenido HTML

```
soup = BeautifulSoup(html_content, "lxml")
```

Extracción de tablas con los datos, del contenido HTML

```
tablas = soup.find_all('table')
```

# Método Web Scraping

## Extracción de los valores de cada tabla

```
tr = tablas[2].tbody.find_all("tr")

df_pib_esp = pd.DataFrame(columns = ['FECHA' , 'PIB', 'PIB_VAR_PORCENTAJE'])

i = 0

for i in range(len(tr)):

    values = []

    for td in tr[i].find_all("td"):

        values.append(td.get("data-value"))

    if values[0] != '0' and values[1] != '0' and values[3] != '0':

        df_pib_esp = df_pib_esp.append({'FECHA' : values[0] , 'PIB' : values[1],
        'PIB_VAR_PORCENTAJE' : values[3]} , ignore_index=True)

    i = i + 1
```

# Método Web Scraping

Modificación de tipos y guardado de datos

```
df_pib_esp[['PIB',  
            'PIB_VAR_PORCENTAJE']].astype(float)  
  
df_pib_esp['FECHA'] = pd.to_datetime(df_pib_esp.FECHA)  
  
df_pib_esp['FECHA'] = df_pib_esp['FECHA'].dt.year
```

Merge de todos los datos en un solo dataset

# Carga desde archivo de datos

## Lectura del dataset descargado manualmente

```
df_population =  
    pd.read_csv("Datos/Generales_UE/Salud_alimentacion_contaminacion/po  
pulation-since-1800.csv")
```

## Filtrado de países de la Unión europea

```
df_population = df_population[df_population.Entity.isin(countries)]
```

## Renombrado de columnas

```
df_population_UE = df_population.rename(columns={'Entity': 'COUNTRY',  
    'Code': 'ISO_CODE', 'Year': 'YEAR', 'Population (historical  
estimates)': 'POPULATION'})
```

# Carga desde archivo de datos

## Filtrado de años necesarios

```
df_population_UE =  
    df_population_UE[df_population_UE['YEAR'] >=  
        min_year]
```

## Guardado de los datos

```
df_population_UE.to_excel('Datos/Generales_UE/UE/Paz_y  
    _seguridad/population_UE.xlsx', index = False,  
    sheet_name='population_UE')
```



# Organización de los datos

## País de referencia

- 1 set de datos
- 56 atributos como PIB, SMI, IPC, etc...
- 60 entradas (1 por cada año)

## Unión Europea

- 30 set de datos
- Formato unificado

# Creación de las visualizaciones interactivas

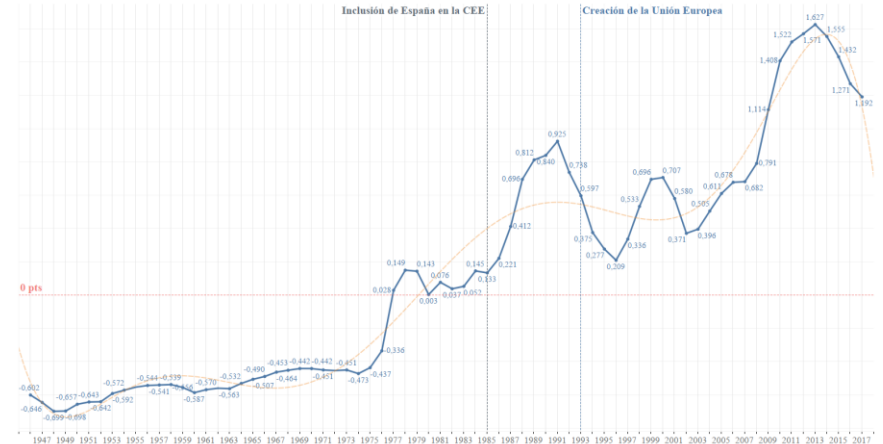
- ▶ Herramienta Tableau
  - ▶ Importación y pretratamiento de los datos
  - ▶ Construcción de las visualizaciones
    - ▶ Tipo de gráfico
    - ▶ Establecer variables, columnas y filas
    - ▶ Añadir filtros
    - ▶ Edición de ejes
    - ▶ Formatear todos los parámetros (tamaños, colores, tipo de fuentes y posicionamiento)
    - ▶ Añadir líneas de referencia y tendencia
    - ▶ Construcción Dashboard
    - ▶ Extracción datos y publicación en línea

# Ejemplos

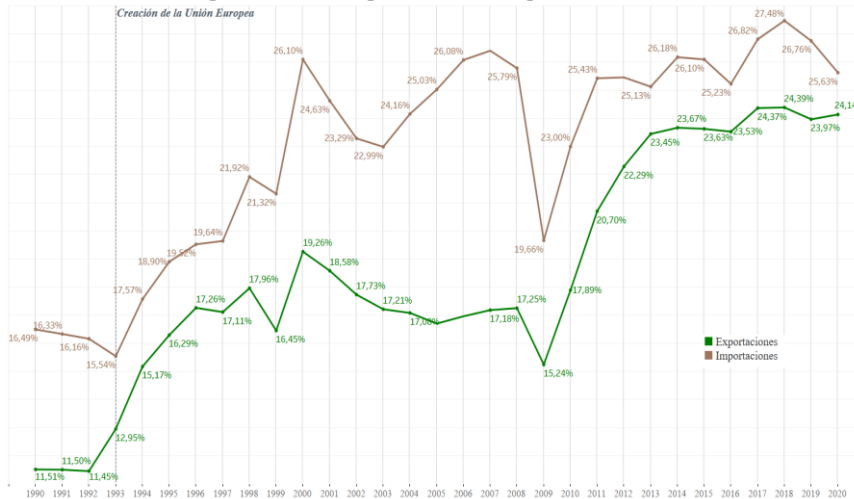
*Tendencias en derechos humanos  
1993 - 2017*



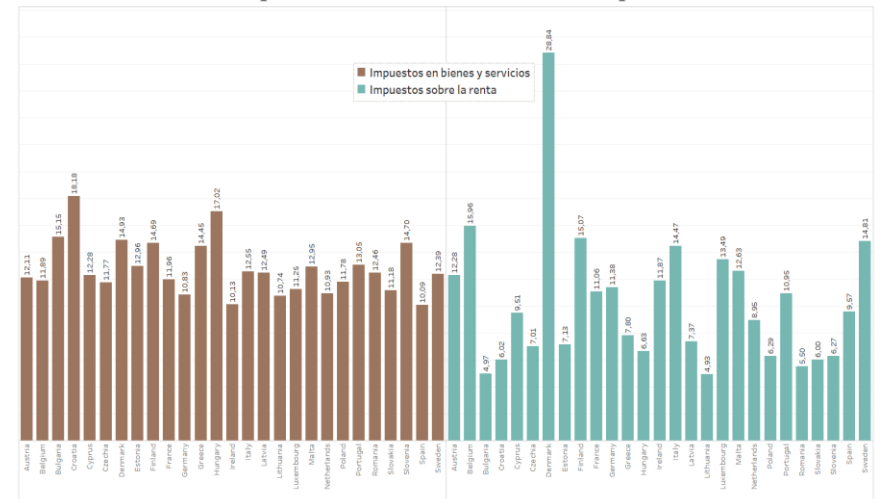
*Derechos humanos en España*



*Exportaciones vs Importaciones en España (% del PIB)*

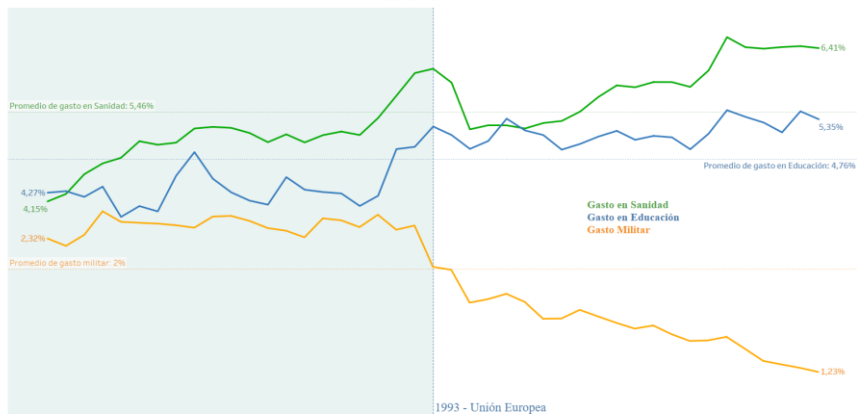


*Impuestos en 2013 en la Unión Europea*



# Ejemplos

**Distribución de gastos en la UE (% del PIB)**  
1972 - 2014



**Índice de desigualdad en la Unión Europea**

Por país - 2019



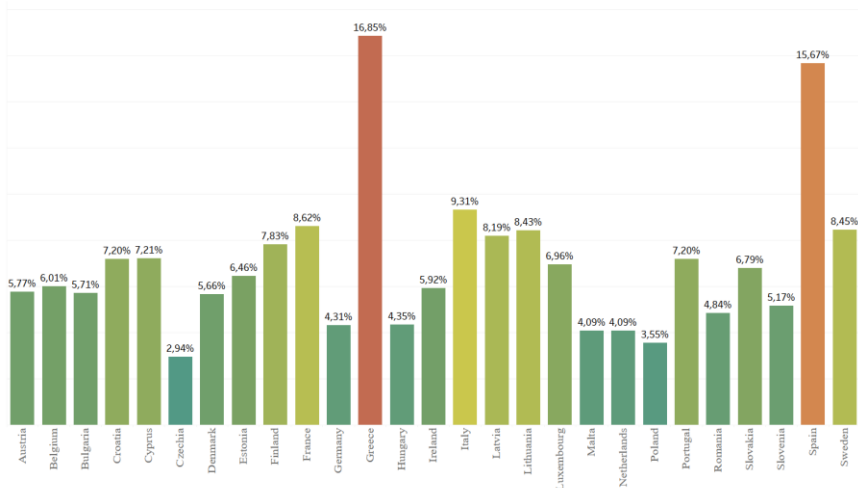
Histórico entre 1980 y 2019



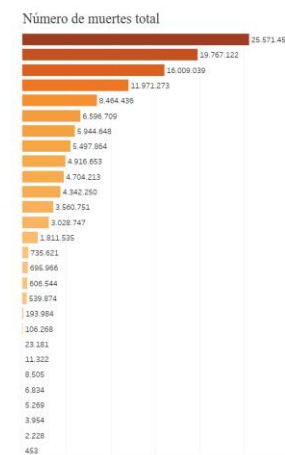
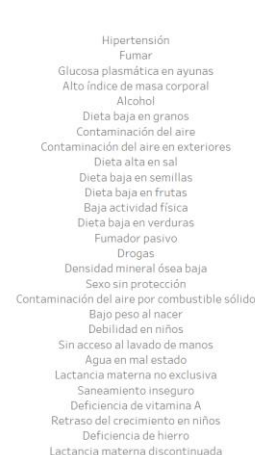
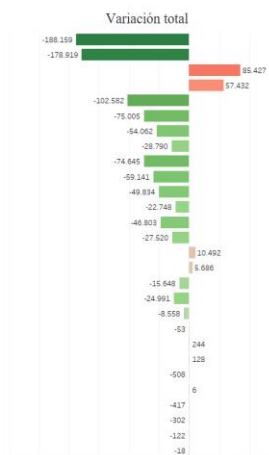
Por país en 2019



**Tasa de desempleo por países en 2020**



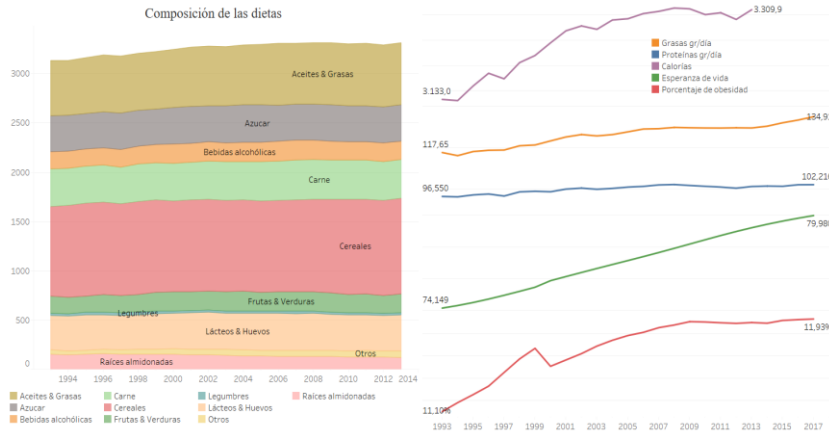
**Muertes asociadas a factores de riesgo en la UE**  
1993 - 2017



# Ejemplos

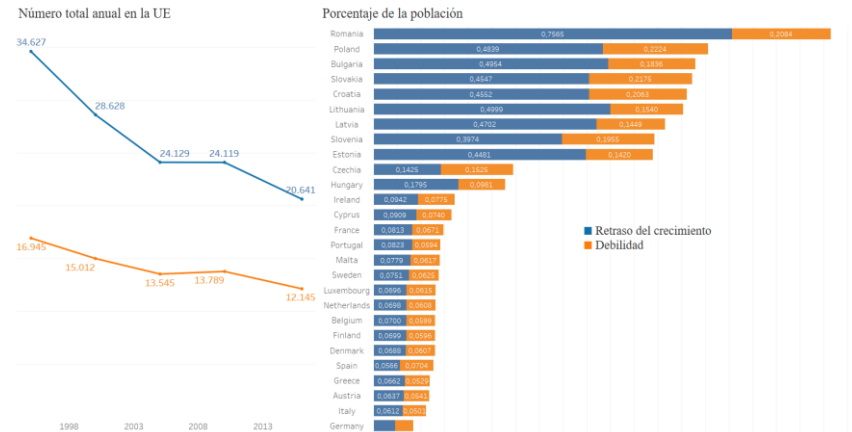
## Alimentación en la Unión Europea

1993 - 2017



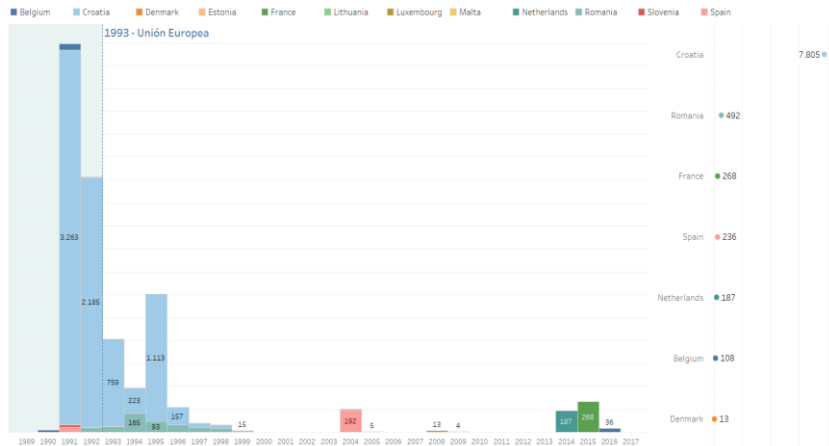
## Retraso del crecimiento vs Debilidad en niños

1993 - 2017



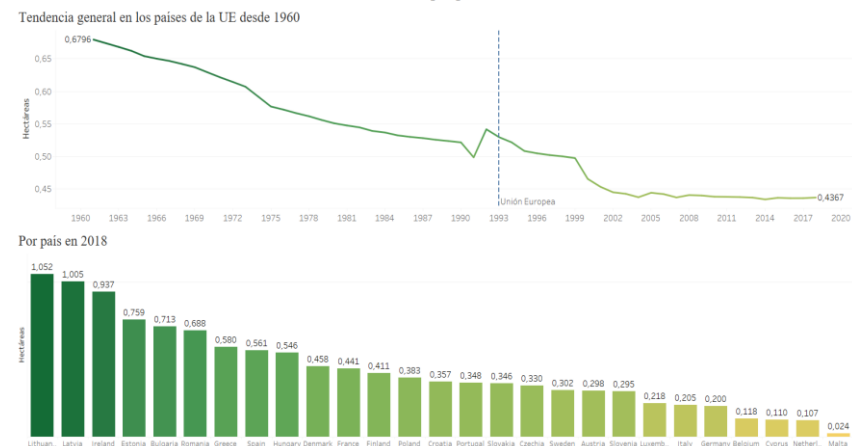
## Muertes por terrorismo en la Unión Europea

1993 - 2017



## Territorios dedicados a fines agrícolas en la UE

Hectáreas por persona



# Despliegue de la aplicación

- ▶ Tecnologías utilizadas:
  - ▶ Brackets – HTML y CSS (Bootstrap)
  - ▶ GitHub Pages
- ▶ Un total de 13 archivos HTML

```
159     <div class="w-100">
160     </div>
161     <div class="col-lg-8">
162         <div class="tableauPlaceholder" id="viz1642458069274" style="position:
            relative"><object class="tableauViz" style="display:none"><param
            name="host_url" value="https%3A%2F%2Fpublic.tableau.com%2F" /> <param
            name="embed_code_version" value="3" /> <param name="site_root" value=""
            /><param name="name" value="Viz_ESP&#47;balance_demografico_esp" /><param
            name="tabs" value="no" /><param name="toolbar" value="yes" /><param
            name="animate_transition" value="yes" /><param
            name="display_static_image" value="yes" /><param name="display_spinner"
            value="yes" /><param name="display_overlay" value="yes" /><param
            name="display_count" value="yes" /><param name="language" value="es-ES"
            /><param name="filter" value="publish=yes" /></object></div>
            <script type="text/javascript">
                var divElement =
                document.getElementById('viz1642458069274');
                vizElement = divElement.getElementsByTagName('object')[0];
                if ( divElement.offsetWidth > 800 ) {
                vizElement.style.width='100%';vizElement.style.height=
                (divElement.offsetWidth*0.75)+'px';} else if ( divElement.offsetWidth >
                500 ) { vizElement.style.width='100%';vizElement.style.height=
                (divElement.offsetWidth*0.75)+'px';} else {
                vizElement.style.width='100%';vizElement.style.height='727px';}
                var scriptElement = document.createElement('script');
                scriptElement.src =
                'https://public.tableau.com/javascripts/api/viz_v1.js';
                vizElement.parentNode.insertBefore(scriptElement, vizElement);
            </script>
163     </div>
164     <div class="col-lg-4 align-self-center">
165         <p class="lead">En la figura se indica una comparativa entre el número de
            nacimientos y el número de muertes en España cada año, donde es notable
            el sor paso que ha habido en los últimos años, esto se debe sobre todo a
            la disminución de nacimientos desde el año 1976 que se ha reducido a la
            mitad hasta los años 90, cuando hubo un rebote hasta la crisis de 2009,
            cuando ya se desplomó hasta los 339,206 nacimientos. El número de muertes
            se ha incrementado de forma constante desde 1960 y la población ha
            aumentado en 13M de habitantes desde el año 1970, esto junto a la elevada
            esperanza de vida ha provocado que la pirámide poblacional se encuentre
            descompensada y haya más gente de avanzada edad y menos jóvenes.</p>
166     </div>
```

# Resultados – Demostración y evaluación

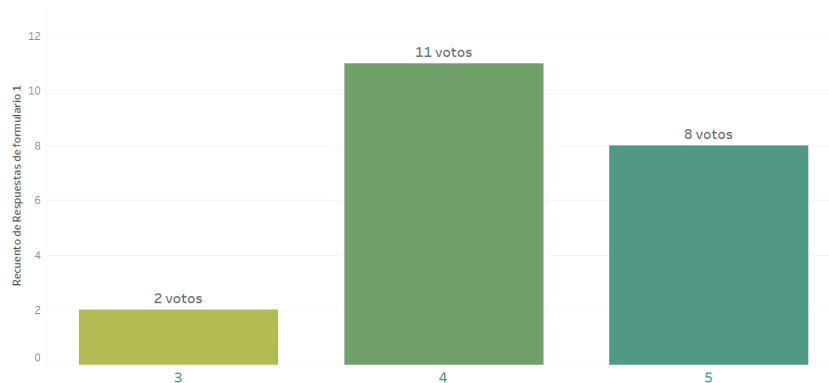
## ► Demostración

### *Cuestionario*

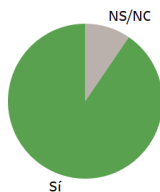
#### Profesiones

Fotógrafo Ciberdefensa Técnico de proyectos. Sociólogo  
Técnico audiovisual Ingeniera de Telecomunicaciones  
Estudiante **Funcionario** Empresaria  
Director empresa Cineasta **Administrativo**  
Médico Periodista

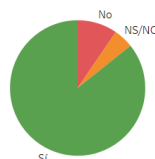
#### Nivel de comprensión



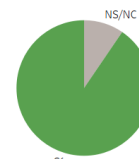
¿Le ha resultado claro y sencillo?



¿Las visualizaciones son fáciles para interactuar?



¿Le ha resultado de utilidad para aumentar sus conocimientos?



# Conclusiones

- ▶ La definición de los temas tratados es clara y acorde a “las prioridades de la UE”.
- ▶ La extracción y transformación de los datos ha sido la esperada.
- ▶ Los usuarios han conseguido interactuar con las visualizaciones sin mayores complicaciones.
- ▶ El objetivo general de aportar conocimiento al usuario de forma sencilla e independiente, se considera cumplido por el feedback aportado y los comentarios propuestos.
- ▶ Los resultados en general han sido muy positivos por la respuesta a la evaluación en un escenario real.



# Líneas futuras

- ▶ Optimizar el tratamiento de los datos
- ▶ Reflejar cambios en tiempo real
- ▶ Cambios en el diseño de la aplicación web
- ▶ Añadir un seguimiento o tracking de la web para analizar el comportamiento de los usuarios
- ▶ Tratar nuevos temas
- ▶ Importante en el ámbito de la divulgación informativa y gran capacidad de mejora.

Muchas gracias por su atención



[www.unir.net](http://www.unir.net)