

CC5208 Visualización de Información

Tarea 3: Visualizaciones con D3

Prof. Benjamin Bustos
Aux. Sebastián Sepúlveda
4 de mayo de 2022

Fecha de entrega: martes 24 de mayo a las 23:59

Esta tarea debe realizarse en forma individual.

Escriba un informe para responder a las actividades descritas en este enunciado. El informe debe tener como máximo seis páginas para responder ambas partes de la tarea, incluyendo imágenes y referencias si las hay. Para la actividad de la **Parte A.1** adjunte a su entrega el archivo `html` que contiene el código de la visualización. Para la actividad de la **Parte A.2** solamente es necesario agregar la visualización realizada a su informe.

1. Parte A (4.5 ptos.)

El objetivo de esta parte es utilizar D3.js para crear visualizaciones Web interactivas y con ellas analizar datos de condición de vulnerabilidad estudiantil en colegios de Chile. Los datos fueron recolectados desde <http://www.junaeb.cl/ive>.

Descripción de los datos

El Sistema Nacional de Asignación con Equidad (SINAE) es un conjunto de criterios que permite identificar distintos grupos dentro de la población de estudiantes de educación básica y media de establecimientos municipales o particulares subvencionados del país, de acuerdo al nivel de vulnerabilidad que presentan. Así, la población identificada como “vulnerable” queda clasificada en tres prioridades: primera, segunda y tercera prioridad, además de dos prioridades que corresponden a datos faltantes o estudiantes que no fueron priorizados.

A continuación se presenta la descripción las prioridades [1] y de los demás datos [2] en los archivos a utilizar:

- ID_RBD: Rol base de datos del establecimiento (nominal categórico).
- DV_RBD: Dígito verificador del RBD (nominal categórico).

- **NOMBRE_ESTABLECIMIENTO:** Nombre del establecimiento (nominal categórico)
- **DEPENDENCIA:** Dependencia del establecimiento, {MUNICIPAL - DAEM, PARTICULAR SUBVENCIONADO, MUNICIPAL - CORPORACIÓN, ADMINISTRACIÓN DELEGADA, SERVICIO LOCAL DE EDUCACIÓN} (nominal categórico).
- **DS_RURALIDAD:** Tipo de área donde se ubica el establecimiento {URBANO, RURAL} (nominal categórico).
- **DS_REGION_ESTABLE:** Código de la región donde se ubica el establecimiento (nominal categórico).
- **DS_PROVINCIA_ESTABLE:** Código de provincia donde se ubica el establecimiento (nominal categórico).
- **DS_COMUNA_ESTABLE:** Nombre de comuna donde se ubica el establecimiento (nominal categórico).
- **PRIMERA PRIORIDAD:** estudiantes con riesgos socioeconómicos (ordinal cero absoluto).
- **SEGUNDA PRIORIDAD:** estudiantes con menor riesgo socioeconómico que en 1a Prioridad, pero que además presenta riesgos socio-educativos como problemas de rendimiento escolar, asistencia o deserción (ordinal cero absoluto).
- **TERCERA PRIORIDAD:** estudiantes con igual riesgo que 2a Prioridad pero sin los problemas mencionados (ordinal cero absoluto).
- **NO_PRIORIZADO_EN_VULNERABILIDAD:** estudiante que no fueron registrados dentro de alguna prioridad de vulnerabilidad (ordinal cero absoluto).
- **SIN_INFORMACION:** estudiante que no se les estudió su nivel de vulnerabilidad (ordinal cero absoluto).

Actividades

1. Construya un diagrama de dispersión (*scatter plot*) de los colegios de educación básica. Para esto, cargue los datos del archivo `ive-basica-2022.csv` y agregue las siguientes características a su visualización:

- Eje x : total de matriculados en el año 2021.
 - Eje y : porcentaje de vulnerabilidad IVE-SINAE.
 - Color: dependencia del colegio.
 - Área: porcentaje de alumnos con primera prioridad de vulnerabilidad.
 - Incluya una leyenda de los colores.
 - Incluya un *tooltip* interactivo que al pasar el mouse sobre cada colegio muestre información acerca de éste.
2. Diseñe y grafique una visualización (que no sea un diagrama de dispersión) que permita observar la distribución de alumnos vulnerables por región con respecto a su dependencia. Tenga en consideración lo siguiente:
- Los datos de alumnos de educación básica y media en los archivos `ive-basica-2022.csv` y `ive-media-2022.csv`, respectivamente.
 - La función `nest()` de D3 le puede ser útil para agrupar datos por región.
 - Visite la página de D3.js d3js.org para explorar opciones de visualizaciones posibles a crear con D3.
 - La visualización no necesariamente debe mostrar todos los datos al mismo tiempo; puede utilizar selectores interactivos que carguen un set de datos por vez.
 - Considere agregar un *tooltip* interactivo que al pasar el mouse entregue un resumen de los datos.

Para la evaluación de este ítem se considerará tanto que funcione correctamente y la visualización creada cumpla con los requisitos anteriores, como que su diseño esté bien justificado. Debe incluir una descripción detallada de por qué eligió la visualización usada y debe justificar su diseño.

3. Realice un análisis de los datos usando las visualizaciones creadas. ¿Qué puede decir de los datos? ¿Qué patrones se observan?

FUENTE:

- [1] <http://junaebabierta.junaeb.cl/catalogo-de-datos/indicadores-de-vulnerabilidad/>
 [2] https://www.junaeb.cl/wp-content/uploads/2013/02/libro_junaeb.pdf

2. Parte B (1.5 ptos.)

Para la siguiente visualización:

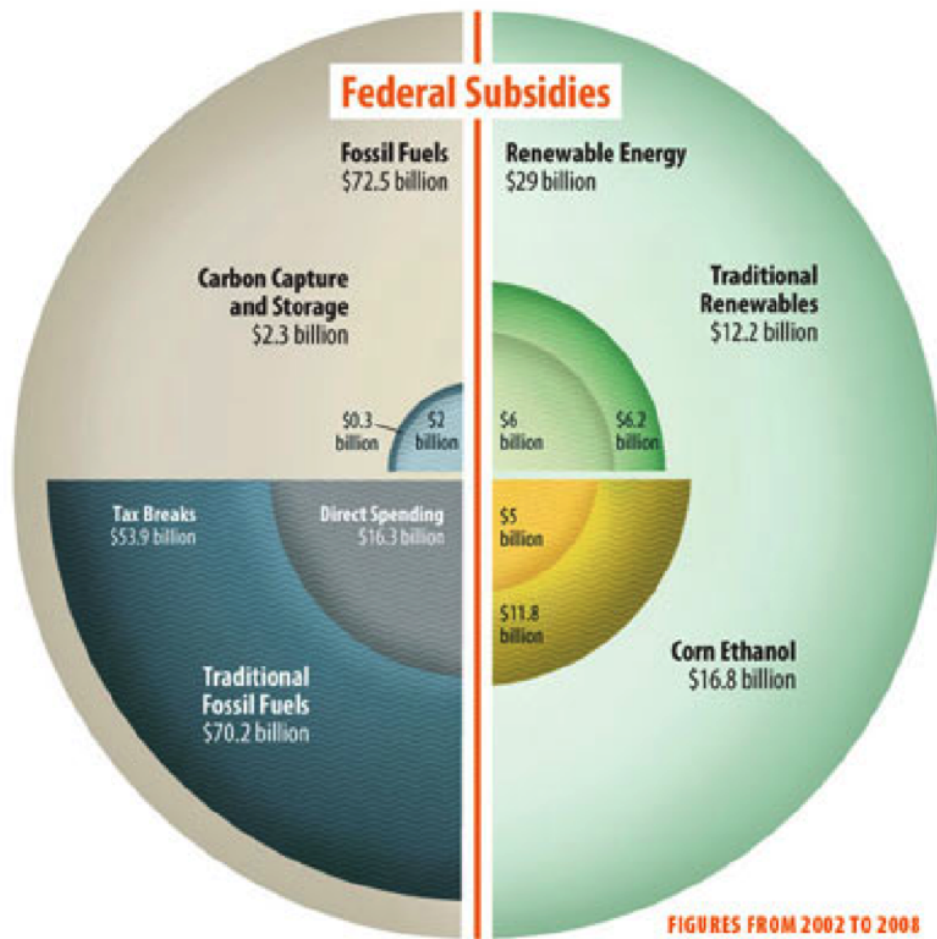


Figura 1: Datos económicos de fuentes de energía

1. Describa el mapeo de atributos a variables visuales (mapeo visual).
2. ¿Qué deficiencias tiene la visualización?
3. Proponga una visualización más expresiva y eficaz, describiendo el mapeo visual de su propuesta y realizando un bosquejo de ésta.

Notas importantes

- La tarea es individual. No olvide leer los documentos “Política de Colaboración” y “Reglas del Curso”, ambos publicados en U-Cursos.
- La tarea debe entregarse a través de la página de U-Cursos, sección Tareas.
- La entrega durante el periodo de atraso definido en U-Cursos no tiene penalización en la evaluación de esta tarea.