



VISUALIZACIÓN DE DATOS

PEC 3

Informe de proyecto de visualización de datos
Acuerdos de paz a nivel mundial

Nadia Nathaly Sánchez Pozo
nsanchezpo@uoc.edu

Índice

Resumen:	1
1. Introducción	1
2. Datos	2
2.1. Descripción de dataset	2
2.2. Proceso de trabajo	3
2.3. ¿Se usarán de las evidencias y conocimientos sobre los datos obtenidos en la PEC2?	3
2.4. Tipos de relaciones que queremos realizar con los datos	4
2.5. Descripción de la gráfica elegida	4
2.6. Descripción de la gobernanza de datos	4
3. Diseño	5
3.1. Colores	5
3.2. Tipografía	5
3.3. Estructura o Plantilla	5
3.4. Bloques	5
3.5. Formato	6
3.6. Mockup	6
Bibliografía	10

Resumen:

El proyecto de visualización de datos “Acuerdos de paz a nivel mundial”, es una interfaz visual para la base de datos de acuerdos de paz (PA-X). La interfaz interactiva permite explorar los acuerdos a nivel país, con una línea de tiempo además de visualizar los datos de mapas, seleccionando filtros para especificar los tipos de acuerdos que deben mostrar las visualizaciones de datos. Esta interfaz permite la exploración de datos temporales, geográficos y temáticos de los acuerdos de paz a nivel mundial, que data desde 1990 hasta 2019 para el programa de investigación de Asentamientos Políticos en la Universidad de Edimburgo.

1. Introducción

La base de datos de acuerdos de paz (PA-X) es un repositorio público de acuerdos de paz a nivel mundial, esta base dispone de acuerdos firmados desde 1990 hasta 2019. Los acuerdos de paz son documentos de libre acceso con el objetivo de contribuir a nuevas investigaciones sobre la paz, podemos hacer uso de dichos documentos accediendo a plataforma PA-X. En este proyecto se ha seleccionado el conjunto de datos completo con el objetivo de realizar una interfaz visual, la cual será capaz de contribuir al análisis exploratorio del conjunto de datos, en este caso se muestra un mapa que permita identificar la distribución geográfica de los acuerdos, para identificar las tendencias de los acuerdos de paz a nivel mundial permitiéndonos tener una idea de los países que han estado involucrados en procesos de paz y tipo de conflicto suscitado. De los países involucrados elegimos Colombia con el objetivo de analizar la evolución temporal de los acuerdos de paz por tipo de conflicto, para intentar identificar los géneros involucrados. Elaboramos diferentes visualizaciones para una mejor interpretación del cambio temporal de los datos.

2. Datos

2.1. Descripción de dataset

Este dataset es interesante para aquellos investigadores que deseen profundizar en analítica de grandes volúmenes de datos ya que el conjunto de datos está formado por 1789 observaciones de 266 variables, creando un esquema de metadatos jerárquico. las cuales tienen parámetros numéricos, categóricos y también tipos texto natural. En este conjunto de datos tenemos algunos metadatos sobre cada columna. Existen variables binarias que toman valores de 0 ó 1, como por ejemplo las variables “Groups”, “State definition” y sus subgrupos. Una descripción detallada del significado de cada columna del dataset, la pueden encontrar en el siguiente: [enlace](#) .

En este proyecto estudiamos las siguientes variables

1. Con: País / Entidad
2. Contp: tipo de conflicto
3. Reg: Región
4. Dat: Fecha firmada
5. Agtp: tipo de acuerdo / conflicto
6. Stage: etapa de acuerdo
7. Loc1GWNO: Indica la ubicación primaria del conflicto dentro del estado y devuelve el código de país de Gleditsch-Ward
8. GeWom: mujeres, niñas y género
9. GaMe: hombres y niños
10. GeFa: familia
11. GeSo: Orientación sexual

En el conjunto de datos podemos apreciar la existencia de valores perdidos, el conjunto de datos está formado por esquema de metadatos jerárquico en su mayoría de tipo binario. Existen otras variables que su único valor es igual a 0 como es el caso de las siguientes variables: GeLgbti, GeLgbtiPos, GOthOth, GeLgbtiNeg, GMigOth, GAgeOth, HrNiOth. La variable Imsrc, toma un único valor igual a “1”.

2.2. Proceso de trabajo.

El proceso de los datos se compone de una fase de “preprocesado” en la cual se realiza la limpieza y reducción de datos para su posterior análisis.

Limpieza del conjunto de datos. Para esto utilizamos la librería pandas de Python, eliminando aquellas variables que tengan más del 5% de valores NaN, eliminamos las variables: 'ThrdPart', 'OthAgr', 'Loc2ISO', 'Loc2GWNO', 'UcdpCon', 'UcdpAgr', 'PamAgr', 'CowWar'.

Reducción, definimos una matriz de correlación para obtener el índice de variables con correlación superior al 80% ya que este tipo de variables altamente correlacionadas aportan información duplicada que está contenida en otras variables. Por lo tanto, eliminamos las siguientes variables 'N_characters', 'GChSubs', 'GDisSubs', 'GAgeSubs', 'GMigSubs', 'GRaSubs', 'GReAntid', 'GReSubs', 'GIndSubs', 'GRefSubs', 'GeMeNu', 'EpsRes', 'HrNiNe', 'CeProv', 'TjAmPro', 'TjRMa'. (Este proceso se especifica en nsanchezpoPEC2.ipynb)

Análisis, cargamos los datos a una herramienta más robusta en este caso tableau, para realizar la representación gráfica de los datos con el objetivo de realizar un análisis visual ya que la visualización facilita la interpretación de los datos. Nuestro objetivo en la presente investigación es analizar la evolución temporal de los acuerdos de paz suscritos en Colombia por tipo de conflicto, para intentar identificar los géneros involucrados en dichos acuerdos.

2.3. ¿Se usarán de las evidencias y conocimientos sobre los datos obtenidos en la PEC2?

A partir del dataset definido en la PEC2, con 1698 observaciones. Realizaremos un mapa para visualizar la distribución de los acuerdos a nivel mundial. En la exploración anterior se identificó que Colombia es uno de los países con mayor número de acuerdos por lo cual centraremos nuestro estudio en dicho país.

2.4. Tipos de relaciones que queremos realizar con los datos.

Vamos a comparar los tipos de acuerdos de paz, y analizar la relación entre género y tipo de acuerdo. Los datos estarán distribuidos por año. Se realizará una comparación a lo largo del tiempo de los diferentes tipos de conflictos suscitados en Colombia.

2.5. Descripción de la gráfica elegida.

Para la exploración general de los datos elegimos un treemap para una visualización jerárquica de los países por número de acuerdos y un Choropleth map porque queremos representar datos geolocalizados.

También elegimos diagramas de barras porque nos permite comparar en el tiempo, además de diferenciar por colores las categorías de las variables.

2.6. Descripción de la gobernanza de datos.

Existirán varias visualizaciones enlazadas en una historia global, se representará una primera visualización interactiva con el objetivo de realizar una exploración general de los acuerdos por país y la distribución geográfica de los países involucrados en los acuerdos de paz, dicha visualización permitirá filtrar los datos por tipo de acuerdo y continente. Para lograr el objetivo de este proyecto se presenta una segunda visualización la que se compone de diferentes graficas que permiten el análisis de la evolución de los acuerdos de paz en Colombia. La gráfica representará la distribución anual por tipo de conflicto. La tercera visualización representará los géneros involucrados por tipo de conflicto, y la distribución anual de géneros involucrados en los acuerdos.

3. Diseño.

En este apartado se especifica las decisiones de diseño,

3.1. Colores.

Los colores evocan emoción expresan significado, colores como el rojo suelen ir asociado con resultado “negativo”, los colores nos permiten resaltar una determinada información que se considere relevante o destacar aquellos puntos donde se ha de centrar el análisis, la gama de colores aconsejada para realizar comparaciones son los colores pastel.

Para el treemap empleamos una gama de colores en donde el color rojo más fuerte es el país con mayor número de acuerdos de paz.

La grafica que representa los géneros se utiliza un color azul para hombres y rosado para mujeres, café familia, naranja orientación sexual.

3.2. Tipografía.

El titulo del proyecto se presenta enmarcado para contextualizar y dirigir la atención hacia la información más importante. En cada gráfica se emplea tipografía legible para las etiquetas. El uso de leyendas y etiquetas es un factor importante en la visualización para la lectura de los datos.

3.3. Estructura o Plantilla.

El proyecto de visualización es un conjunto de visualizaciones enlazadas en una historia global. Donde la primera página es un dashboard de países, tipo de acuerdos y años. La segunda y tercera página es la representación visual del objetivo del proyecto representada es dos dashboard interactivo.

3.4. Bloques.

Página 1 se dividen en dos bloques primario y secundario, el bloque primario se subdivide en dos bloques un bloque por gráfica.

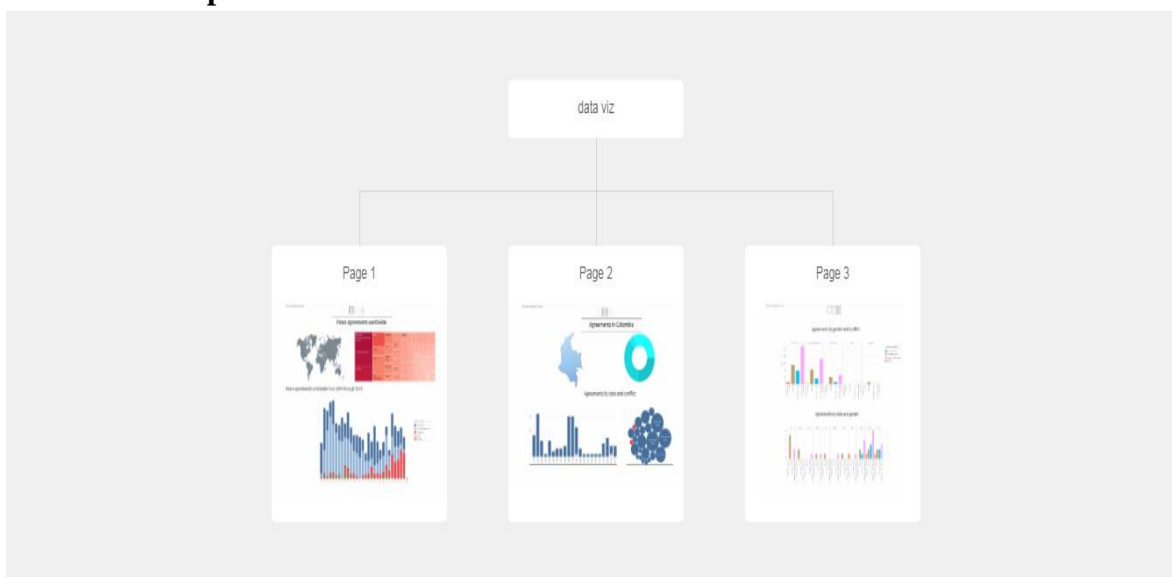
Pagina 2 se divide en cuatro bloques un bloque por gráfica.

Pagina 3 se divide en dos bloques un bloque por gráfica.

3.5. Formato

El formato de este proyecto es digital, todo el contenido será almacenado en un repositorio de github y una visualización compartida en tableau public.

3.6. Mockup

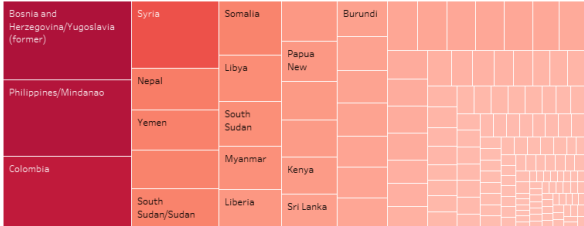


P gina 1.

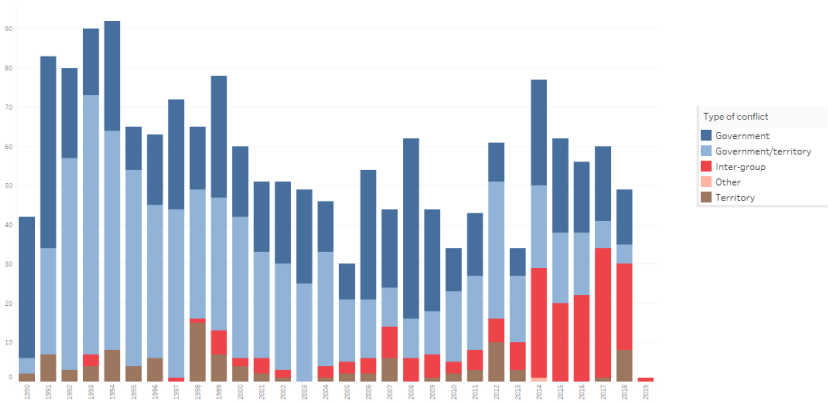
Proyecto de visualizaci n de datos

1	2	3
---	---	---

Peace agreements worldwide



Peace agreements worldwide from 1990 through 2019



Página 2.

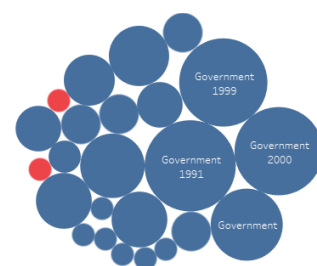
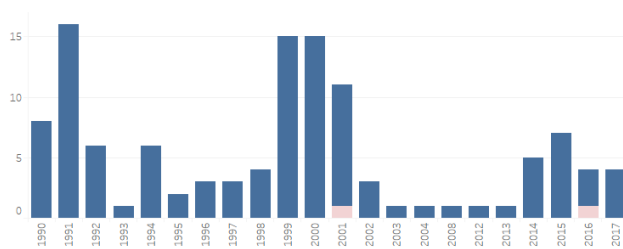
Proyecto de visualización de datos

1	2	3
---	---	---

Agreements in Colombia



Agreements by date and conflict

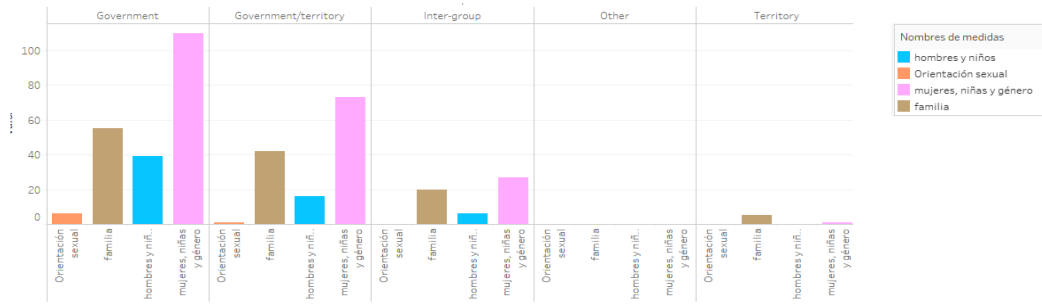


Página 3.

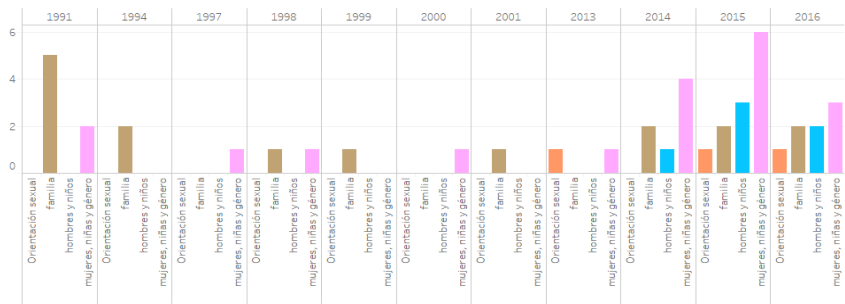
Proyecto de visualización de datos

1	2	3
---	---	---

Agreements by gender and conflict



Agreements by date and gender



Bibliografía

- [1] del Rio Riande, M. G. (2016). Explotación y visualización de datos de investigación en Humanidades. El aporte de las Humanidades Digitales. In InnovaTICS 2016.
- [2] Dürsteler, J. C. (2000). Visualización de información. Gestion.
- [3] Havens, Lucy & Bao, Mengting & Pschetz, Larissa & Bach, Benjamin & Bell, Christine. (2019). PaxVis: Visualizing Peace Agreements.
- [4] Olmeda-Gómez, C. (2014). Visualización de información. El profesional de la información, 23(3).
- [5] Pontis, S. (2007). La historia de la esquemática en la visualización de datos. Visualizar'07.
- [6] Pascual Cid, V., & Minguillón Alfonso, J. Buenas prácticas en visualización de datos.
- [7] Ward, M. O., Grinstein, G., & Keim, D. (2015). Interactive data visualization: foundations, techniques, and applications. AK Peters/CRC Press.