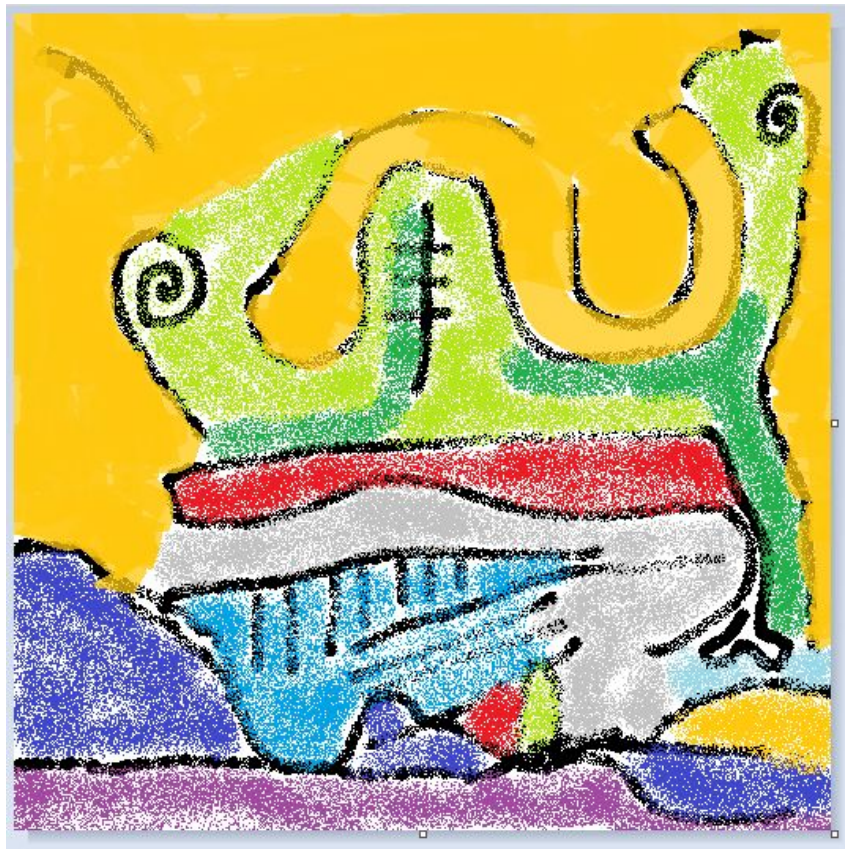


30 años de tratados de paz

Mapa de los compromisos alcanzados



Iván Calvo Revelo

Visualización de datos

Índice

Abstract	3
1. Introducción	4
2. Exploración inicial (PEC 2)	5
2. Datos	7
3. Diseño	10
4. Proyecto final (PEC 4)	12
Agradecimientos	13
Bibliografía	14

Abstract

El presente trabajo pretende ofrecer un dashboard con el que explorar los conflictos bélicos de los últimos treinta años. Partiendo de un conjunto de datos publicado en *Peace Agreements Database*, se realiza una transformación que asigna diversas puntuaciones a los distintos acuerdos de paz, en base al nivel de compromiso alcanzado en una serie de áreas. Estos acuerdos se presentan dentro de su contexto geográfico, político y temporal, ayudando a comprender su influencia en la resolución del conflicto.

1. Introducción

Las metodologías y técnicas pertenecientes al ámbito de la visualización de datos suponen una de las mayores disrupciones ocasionadas por la aparición de la ciencia de datos. Esta disrupción afecta de manera uniforme a la práctica totalidad de ámbitos de conocimiento académico, permitiendo desdibujar la conocida frontera entre las ciencias y las humanidades. De esta manera, podemos ver trabajos de visualización de datos en cualquier área que requiera la presentación formal de una narrativa. Estas narrativas se presentan de muy diversas formas, con distintos propósitos y técnicas, como se expone en [6].

El propósito de este proyecto se inspira en trabajos de presentación de documentos históricos [1] y de visualización de hechos periodísticos [3]. Cada acuerdo de paz es una fuente primaria de información actual, que tendrá una repercusión significativa en la realidad futura. Es, de esta manera, una interfaz entre el periodismo y la historia. La irrupción del dato y su tratamiento para mantenerlo limpio, confiable y relevante son un factor crucial para que la historia contemporánea sea la mejor documentada.

Todo proyecto enfocado sobre las llamadas *Visual Analytics* [5] maximiza su valor cuando se comprende como un esfuerzo iterativo. De esta manera, la primera expectativa que se baraja es la de obtener una mejor idea de en qué manera se podrán idear nuevos planes de análisis que a su vez abran paso a futuras visualizaciones. Se espera que el usuario, al explorar el contexto de un tratado que le sea conocido, proponga nuevas métricas y baremos que reflejen nítidamente aquello de lo expresado en ese tratado que lo hace especialmente relevante.

Además de la incorporación de conocimiento experto al modelo, la visualización ofrecerá un soporte común sobre el que investigadores con diversos perfiles podrán aproximarse desde distintos ángulos a dicho conocimiento. La visualización ubicará en el espacio y en el tiempo la evolución humana de un determinado conflicto, así como el desarrollo de temáticas comunes entre conflictos.

El objetivo más ambicioso de esta visualización es dar herramientas para comparar el estado de conflictos presentes con etapas intermedias de conflictos pasados, estudiando en qué medida es cierto que *la historia se repite*.

2. Exploración inicial (PEC 2)

En este trabajo de visualización se presentan datos asociados a acuerdos de paz firmados entre 1990 y 2020. Estos acuerdos, con sus correspondientes metadatos, han sido recopilados y publicados en www.peaceagreements.org, proyecto dependiente de la *Universidad de Edimburgo*.

Dentro de los datos presentados, encontramos un primer grupo básico de campos relativos al acuerdo, tales como la fecha, participantes, región, etapa dentro del proceso de paz o extensión en número de páginas y de caracteres.

Por otro lado, se presenta una serie de indicadores que miden el grado de compromiso alcanzado en el acuerdo, en base a criterios como la mención explícita de un determinado objetivo o la presentación detallada de acciones destinadas a dichos objetivos. Estos indicadores se agrupan en 12 grandes bloques, cada uno de los cuales se ocupa de alguno de los aspectos fundamentales para la superación de un conflicto armado. En total se presentan más de 200 de estos indicadores.

El dataset cubre más de 1800 acuerdos, enmarcados en aproximadamente 140 procesos de paz.

El dataset cuenta con una estructura jerárquica, con campos cuyo valor detalla en profundidad el valor de un campo previo. Ejemplo de esto son los campos *Stage* y *StageSub*, los cuales describen la etapa en la que se encuentra el proceso de paz en el momento del acuerdo. Estos campos son de tipo categórico.

Los campos relativos a los participantes en el acuerdo, tales como *Parties* y *ThrdPart* son textuales y en general no responden a una estructura predeterminada, con lo que requerirían de un complicado análisis y transformación para incorporarse a la visualización.

Los campos relativos a los compromisos alcanzados son en su mayoría binarios o, en algunos casos, ternarios. Estos campos tienen la propiedad de que, en el caso de los de tipo binario, toman el valor 1 ante la presencia de un compromiso y el valor 0 en su ausencia.

Análogamente, las variables ternarias toman valores tanto mayores cuanto mayor sea el grado de compromiso en su correspondiente aspecto. En esta visualización aprovechamos este hecho para realizar un preprocesado de los datos, calculando una puntuación general para cada uno de los 12 grupos temáticos, que es suma de los valores de sus correspondientes indicadores.

Con respecto a la extensión de los acuerdos,

https://public.tableau.com/profile/ivan7347#!/vizhome/PEC2_16057348283570/Extensiondelostextos muestra que tanto el número de páginas como el número de caracteres son medidas que en la práctica se pueden usar de manera indistinta. En este gráfico además podemos observar que *Final Agreement to End the Armed Conflict and Build a Stable and Lasting Peace* es un

outlier en términos de extensión, con más de 300 páginas, mientras que la extensión típica no supera las 50 páginas.

Mediante un gráfico de barras,

https://public.tableau.com/profile/ivan7347#!/vizhome/PEC2_16057348283570/Totaldecompromisos, podemos comprobar que la prevalencia de los distintos indicadores no es nada homogénea, con grandes diferencias entre áreas temáticas y dentro de las mismas.

Como consecuencia de esto, decidimos filtrar aquellos indicadores que aparecen en menos de 10 ocasiones en el total de acuerdos, obteniendo el gráfico de áreas

https://public.tableau.com/profile/ivan7347#!/vizhome/PEC2_16057348283570/Compromisosexcepcionales, el cual permite, además de comparar sus tamaños relativos, identificarlos y agruparlos de acuerdo a su temática.

A parte de esto, el dashboard

https://public.tableau.com/profile/ivan7347#!/vizhome/PEC2_16057348283570/Distribuciondelosgradosdecompromiso muestra los diagramas de cajas asociados a cada una de las puntuaciones temáticas, confirmando la heterogeneidad de estas puntuaciones, tanto en localización como en dispersión.

Podemos observar una moderada correlación entre las puntuaciones de unas temáticas y otras observando conjuntamente su evolución temporal mediante el dashboard

https://public.tableau.com/profile/ivan7347#!/vizhome/PEC2_16057348283570/Evoluciontemporalde los compromisos. Podemos ver un declive en el volumen de compromisos de paz alcanzados durante los últimos años. Aún sería necesario contrastar esta tendencia con el total de conflictos abiertos para poder distinguir si este descenso supone una menor actividad bélica o una menor actividad diplomática.

Finalmente,

https://public.tableau.com/profile/ivan7347#!/vizhome/PEC2_16057348283570/Compromisosalcanzadosporregion?publish=yes ofrece un dashboard en el que los acuerdos se agrupan geográficamente, ofreciendo una visión agregada del total de compromisos alcanzados en distintas materias en función de la región sobre la que radica el conflicto.

Para obtener estas visualizaciones, en primer lugar he realizado un preprocesado de los datos mediante el siguiente script escrito en python que adjunto en la entrega. Una vez obtenidos los datos en un formato adecuado, he realizado diversas pruebas en Tableau para ayudarme a comprender las características del conjunto de datos en cuestión. De cara a las sucesivas entregas, tengo pensado seguir haciendo uso de python a nivel de preprocesado, para facilitar el tratamiento de los distintos indicadores. Mi objetivo principal es continuar aprendiendo Tableau como herramienta central para la exploración y presentación de conclusiones sobre los datos. Si el tiempo me lo permite, también tengo en mente iniciarme en D3.js como alternativa libre a Tableau, dado que entiendo que ofrece una solución mucho más adaptable.

2. Datos

El conjunto de datos sobre el que construiremos la visualización constará de dos tablas, obtenidas mediante transformaciones sobre el *PA-X dataset*.

El propósito de la primera será ubicar cada tratado dentro de un proceso de paz y de una geografía, así como asociarlo con los países implicados en el mismo.

Nombre	Tipo	Descripción
PPName	Cadena	Nombre del proceso de paz
Agt	Cadena	Título del acuerdo
Con	Geográfico (País)	Lista de países implicados en el acuerdo
Reg	Cadena	Región sobre la que se enmarca el conflicto

De estos campos, el único que supone una transformación es el campo *Con* que originalmente aparece codificado como una lista de países separados por barras '/'. En esta tabla, cada tratado aparece tantas veces como países intervengan en él, indicando cada uno de ellos en su correspondiente fila.

La funcionalidad de esta tabla dentro del dataset será la de componer procesos de paz y países, filtrando el conjunto de tratados que se presentarán cronológica y temáticamente.

Estos campos se visualizarán mediante un mapa geográfico y un listado de procesos de paz, cuyo número de acuerdos vendrá representado mediante barras horizontales.

La segunda tabla contiene la información relativa a los tratados concretos, estructurada de tal forma que permita calcular las puntuaciones temáticas exploradas en la PEC 2.

Nombre	Tipo	Descripción
PPName	Cadena	Nombre del proceso de paz
Agt	Cadena	Título del acuerdo
AgtId	Entero	Identificador del acuerdo
Dat	Fecha	Fecha en la que se firma el acuerdo
N_characters	Entero	Extensión del texto del acuerdo en número de caracteres

Topic	Catagórico	Familia de compromisos, dividida en <i>Subtopic</i> .
Subtopic	Catagórico	Compromiso al que hace referencia el campo <i>Score</i> .
Score	Entero	Nivel de compromiso alcanzado en el acuerdo.

De estos campos, *Topic*, *Subtopic* y *Score* se obtienen transformando las columnas temáticas del dataset original. Cada acuerdo aparecerá en tantas filas como columnas temáticas figuran en el dataset original.

De esta manera, si un determinado acuerdo tiene un valor ‘1’ en el campo *GOthAntid*, una de sus filas tendrá “GOthAntid” como valor del campo *Subtopic* y ‘1’ como valor del campo *Score*. Estas columnas se agrupan jerárquicamente por temática, indicándose en el campo *Topic*, lo que permite el cálculo de las puntuaciones exploradas en la PEC 2.

La siguiente tabla muestra cada temática junto con los posibles compromisos relativos a ella:

Groups	GCh,GChRhet,GChAntid,GChSubs,GChOth,GDis,GDisRhet,GDisAntid,GDisSubs,GDisOth,GAge,GAgeRhet,GAgeAntid,GAgeSubs,GAgeOth,GMig,GMigRhet,GMigAntid,GMigSubs,GMigOth,GRa,GRaRhet,GRaAntid,GRaSubs,GRaOth,GRe,GReRhet,GReAntid,GReSubs,GReOth,GInd,GIndRhet,GIndAntid,GIndSubs,GIndOth,GOth,GOthRhet,GOthAntid,GOthSubs,GOthOth,GRef,GRefRhet,GRefSubs,GRefOth,GSoc,GSocRhet,GSocAntid,GSocSubs,GSocOth
Gender	GeWom,GeMe,GeMeNu,GeMeOth,GeLgbti,GeLgbtiPos,GeLgbtiNeg,GeFa
State definition	StDef,StGen,StCon,StSd,StRef,StSym,StInd,StUni,StBor,StXbor
Governance	Pol,PolGen,PolNewInd,PolNewTemp,ConRen,Cons,Elle,ElecComm,PolPar,PolParTrans,PolParOth,Civso,TraI,Pubad
Power sharing	Polps,PpsSt,PpsSub,PpsEx,PpsOro,PpsOthPr,PpsVet,PpsAut,PpsInt,PpsOth,Terps,TpsSub,TpsLoc,TpsAut,TpsOth,Eps,EpsRes,EpsFis,EpsOth,Mps,MpsMe,MpsJt,MpsPro,MpsOth
Human Rights and Equality	HrGen,EqGen,HrDem,Prot,ProtCiv,ProtGrp,ProtLgl,ProtOth,HrFra,HrfSp,HrfBor,HrfTinc,HrfOth,HrCp,CprLife,CprTort,CprEq,CprSlav,CprLib,CprDet,CprFmov,CprFspe,CprFass,CprTria,CprPriv,CprVote,CprReli,CprOth,HrSec,SerProp,SerWork,SerHeal,SerEdu,SerStdI,SerShel,SerSs,SerCult,SerOth,HrNi,HrNiMe,HrNiNe,HrNiOth,Hrli,HrliMon,HrliBod,HrliOth,HrMob,HrDet,Med,MedGov,MedSubs,MedLog,MedOth,HrCit,CitGen,CitRigh

	ts,CitDef,CitOth
Justice sector reform	JusCr,JusCrSp,JusCrSys,JusCrPow,JusEm,JusJu,JusPri,JusTra
Socioeconomic reconstruction	Dev,DevSoc,DevHum,DevInfra,NEC,NatRes,IntFu,Bus,Tax,TaxPo,TaxRef,TaxOth,Ban,CenBan,BanPers,BanInt,BanXb
Land, property and environment	LaRef,LaRefMan,LaRefRet,LaRefOth,LaNom,LaCH,LaCHTa,LaCHIt,LaCHPro,LaCHOth,LaEn,Wat
Security sector	SsrGua,Ce,CeProv,CeGen,SsrPol,SsrArm,SsrDdr,DdrDemil,DdrProg,SsrInt,SsrPsf,SsrFf,Cor,SsrCrOcr,SsrDrugs,Terr
Transitional justice	TjGen,TjAm,TjAmPro,TjSan,TjPower,TjAmBan,TjCou,TjJaNc,TjJalc,TjMech,TjPrire,TjVet,TjVic,TjMis,TjRep,TjRSym,TjRMa,TjNR
Implementation	ImUN,ImOth,ImRef,ImPK,ImE,ImSrc

Las temáticas se representarán a lo largo del tiempo mediante un gráfico de columnas de iconos, que indicarán la puntuación total de dicha temática en ese año. Al seleccionarse cada una de ellas, se filtrarán los gráficos relativos a las subtemáticas y a los acuerdos individuales.

Las subtemáticas se representarán a lo largo del tiempo mediante un gráfico de áreas apiladas, que mostrará la contribución individual de cada una de ellas a la puntuación temática total.

Finalmente, cada acuerdo se mostrará como una colección de burbujas agrupadas, cuyo tamaño vendrá dado por la longitud en caracteres del acuerdo. Cada burbuja aparecerá tantas veces como compromisos figuren en el acuerdo, dentro de la temática sobre la cual se está filtrando.

3. Diseño

La visualización se estructura en cinco bloques: geografía, proceso, temática, subtemática y acuerdos. El espacio se divide en tres franjas horizontales, de las cuales las dos primeras estarán ocupadas por los cuatro primeros bloques y la restante por el último bloque. Este último bloque es el más importante para la exploración, dado que es el que hace aterrizar la visualización de las tendencias sobre acuerdos concretos.

La distribución se basa en una ramificación del dataset, en la cual la primera franja ofrece una visión general, la segunda muestra tendencias temporales que permiten seguir filtrando para obtener los acuerdos de un determinado año, en relación a una determinada temática y finalmente la tercera franja ofrece datos concretos sobre los acuerdos relevantes. De esta manera, la estructura visual materializa el famoso lema *Overview first, zoom and filter, then details-on-demand* [5].

Pese a la tesis defendida en [2], según la cual la fuente tipográfica tiene un gran impacto en el tono con el cual será interpretada la información presentada, el propósito principal de este proyecto es la exploración de un conjunto de documentos, pudiendo funcionar como motor de recomendación. Dado que se espera que el usuario dedique una cantidad razonable de tiempo en una actividad que supone su participación activa, no parecería razonable dedicar grandes consideraciones al efecto comunicativo que supondría para él recibir el texto con una u otra tipografía.

El color, por el contrario, es un elemento fundamental a nivel semántico, dado que orienta la navegación dentro de la visualización y permite cotejar hechos que aparezcan en distintos gráficos. De esta manera, las áreas apiladas y las burbujas deberán compartir un mismo código de colores, para identificar la contribución de cada hecho dentro de la tendencia.

Si bien el formato será digital, dada la interactividad inherente al gráfico, puede ser interesante explorar la posibilidad de permitir realizar capturas de pantalla que apoyen una determinada narrativa dentro de una investigación. Estas capturas podrán o bien ser impresas o bien compartidas en redes sociales, contribuyendo a la popularización de la herramienta. De esta manera, y mediante los filtros, resaltes y selecciones disponibles en la herramienta, el investigador que se proponga extraer narrativas de los datos podrá guiar la atención del lector hacia los elementos gráficos que sean críticos en su análisis.

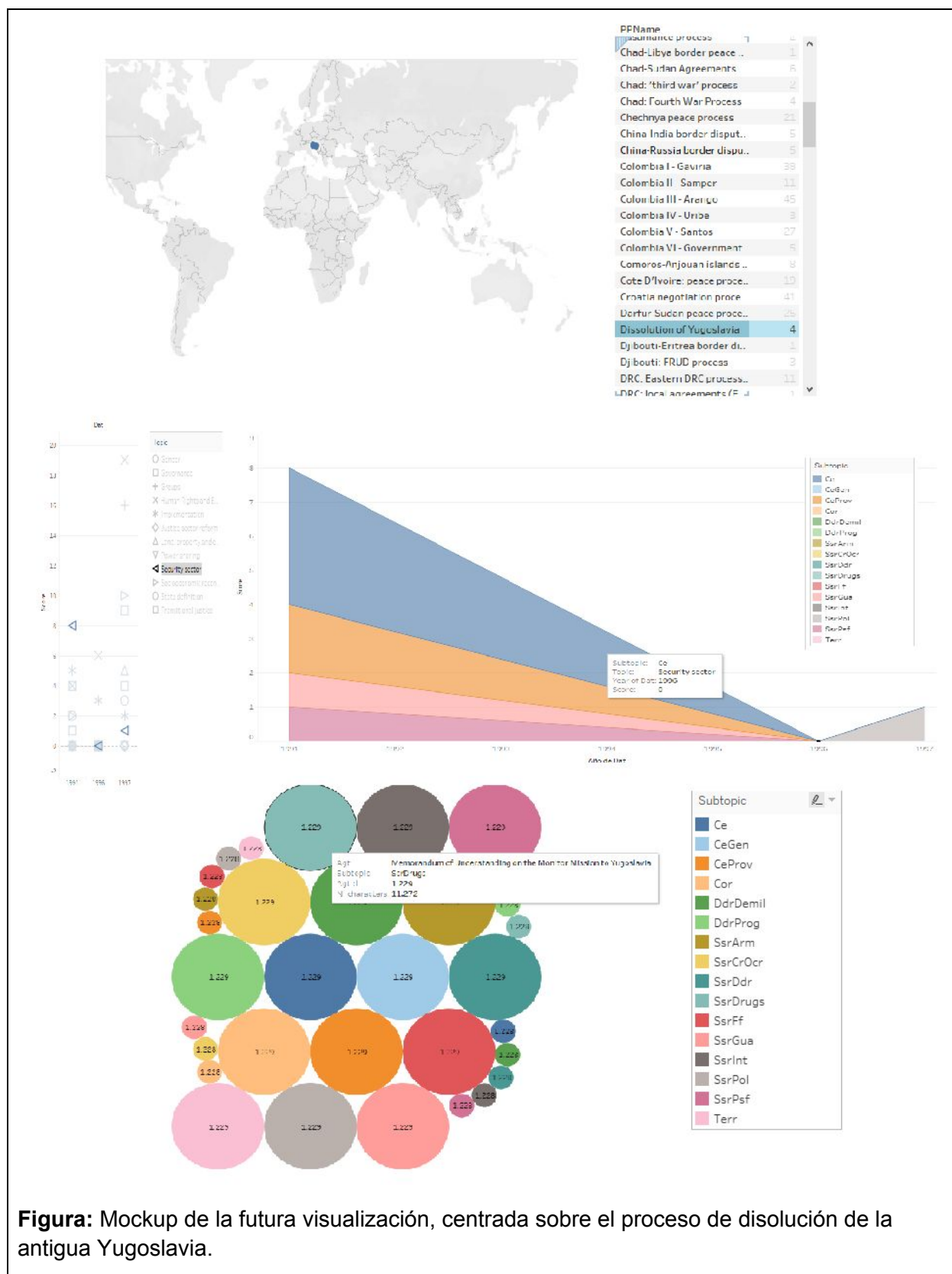


Figura: Mockup de la futura visualización, centrada sobre el proceso de disolución de la antigua Yugoslavia.

4. Proyecto final (PEC 4)

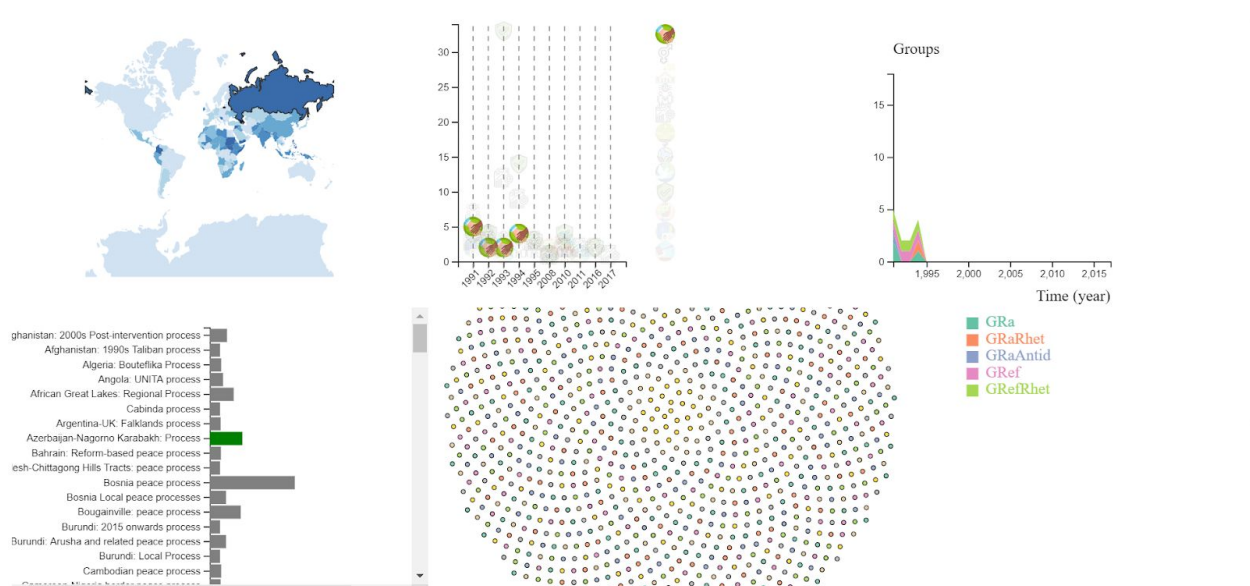
La visualización se encuentra accesible en:

<https://icalvor-uoc.github.io/dataviz/index.html>

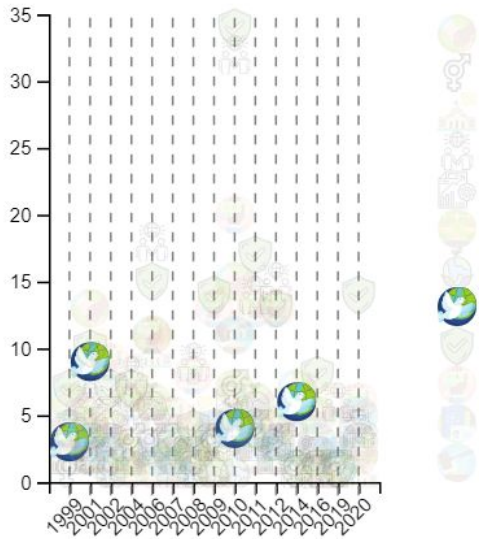
El repositorio que contiene todo el código de la visualización se encuentra en:

<https://github.com/icalvor-uoc/dataviz>

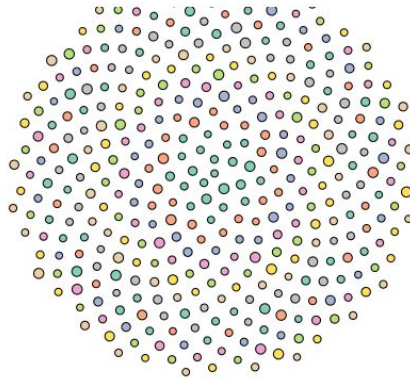
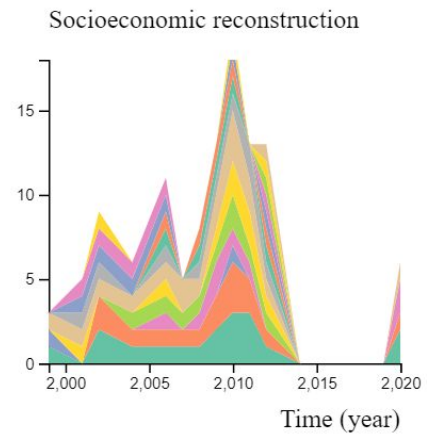
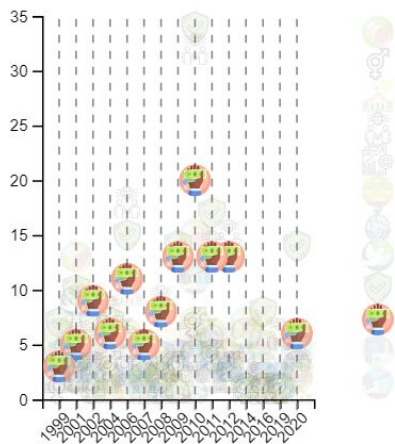
La visualización tiene en su lado derecho los filtros por país y proceso de paz que aplican sobre el resto de la visualización. Pueden añadirse y quitarse países y procesos mediante click izquierdo del ratón, así como fijar el filtro sobre un determinado país o proceso mediante click derecho.



El gráfico superior central indica la evolución temporal de cada temática, pudiendo seleccionarse una de las 12 temáticas para resaltar su evolución dentro del gráfico.

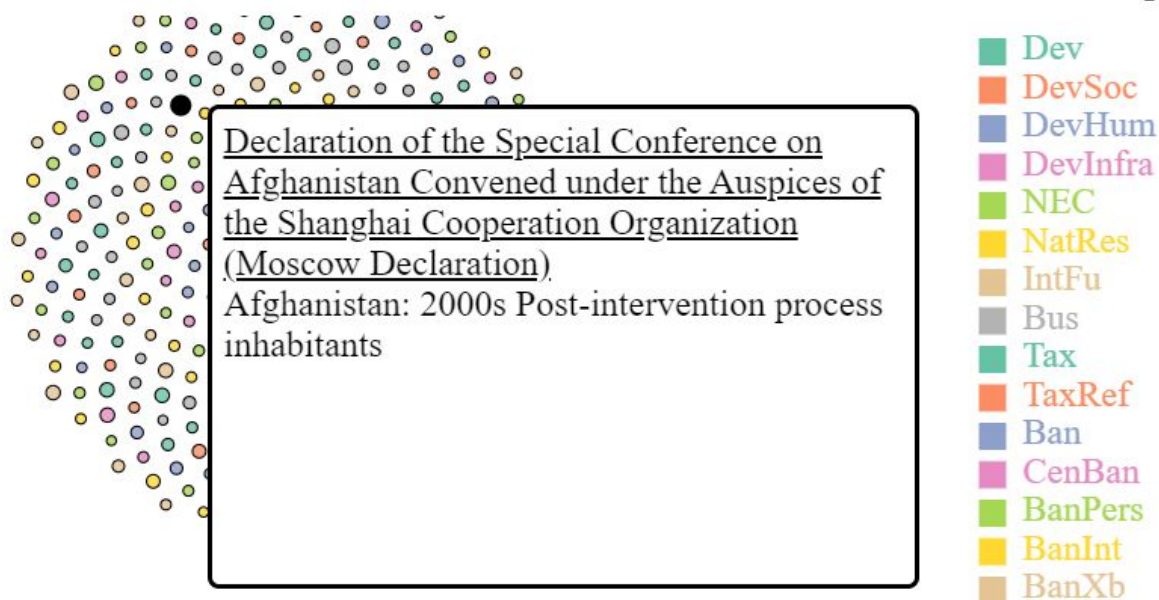


Esta selección actúa a su vez como un filtro que hace zoom sobre los compromisos concretos relativos a esa determinada temática.

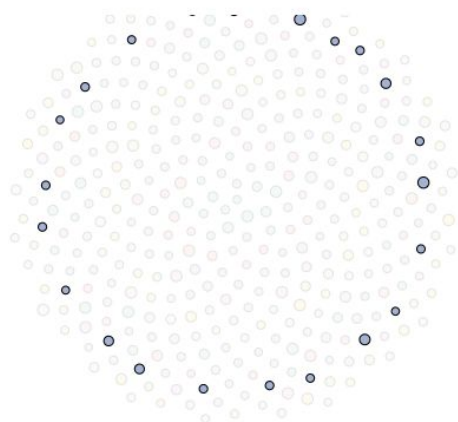
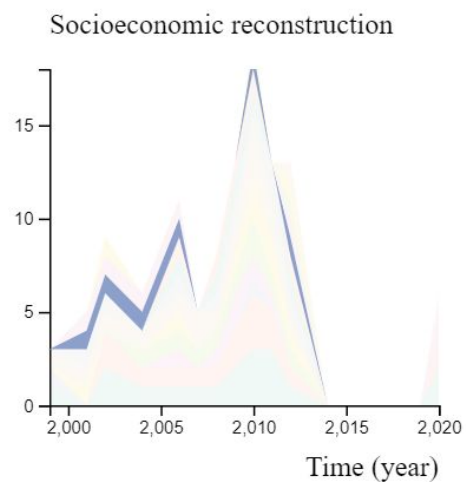
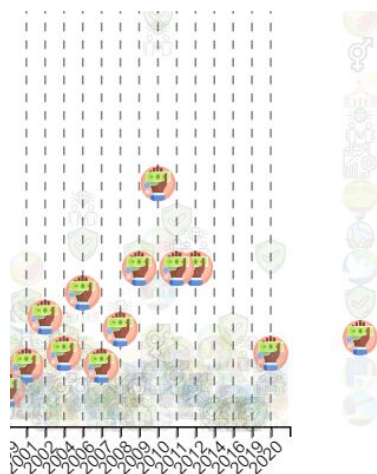


- Dev
- DevSoc
- DevHum
- DevInfra
- NEC
- NatRes
- IntFu
- Bus
- Tax
- TaxRef
- Ban
- CenBan
- BanPers
- BanInt
- BanXb

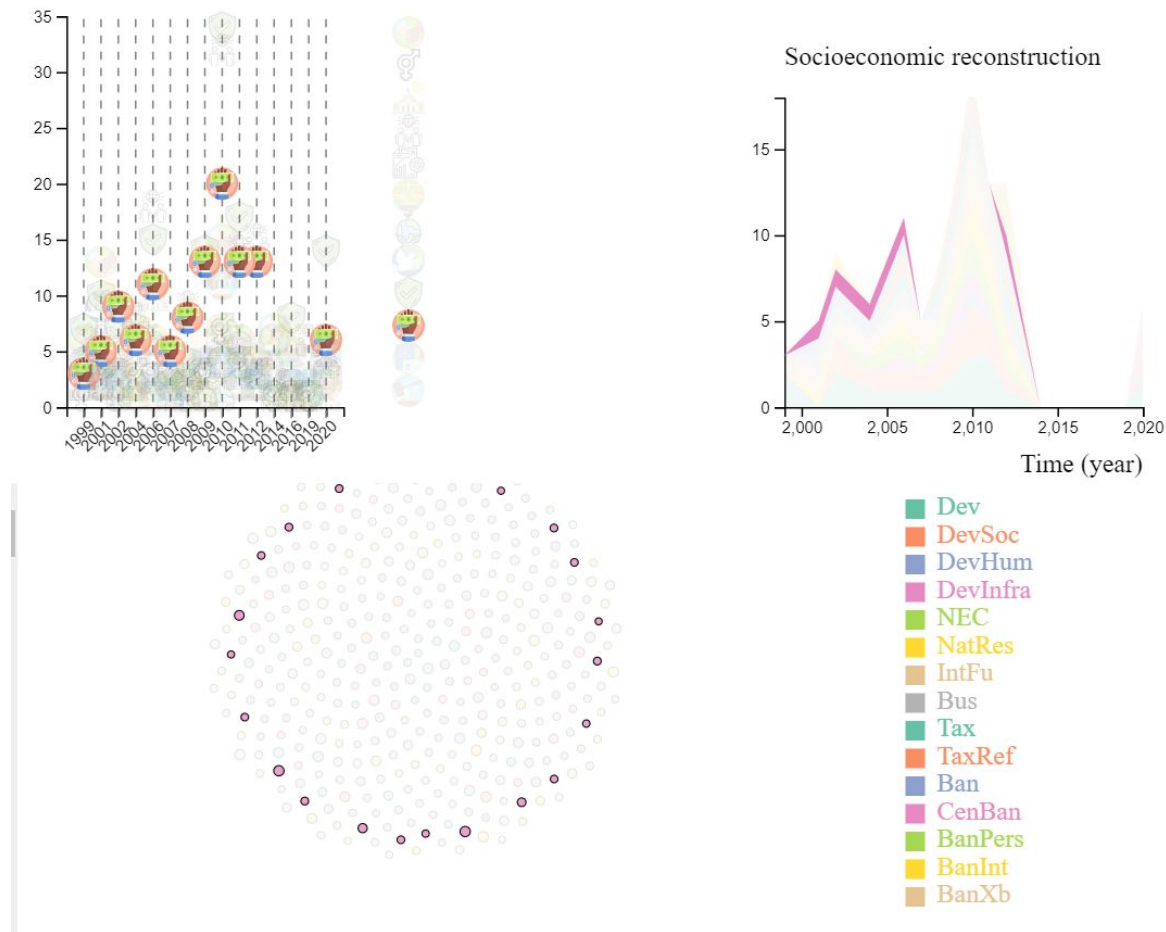
Podemos ver la evolución de los distintos compromisos de la temática seleccionada mediante un gráfico de áreas apiladas. Además, disponemos de un mapa de burbujas que permite explorar los acuerdos de paz en los cuales se ha firmado cada uno de los compromisos.



Para facilitar la navegación, se dispone de una leyenda común entre estos dos últimos gráficos, que permite resaltar los compromisos deseados:



- Dev
- DevSoc
- DevHum
- DevInfra
- NEC
- NatRes
- IntFu
- Bus
- Tax
- TaxRef
- Ban
- CenBan
- BanPers
- BanInt
- BanXb



El principal aprendizaje que considero haber obtenido de este proyecto es una mayor familiaridad con javascript, svg y la librería d3.js en particular. La potencia de esta pila tecnológica, que opera en un bajo nivel de abstracción, hace que las limitaciones técnicas y conceptuales de la visualización vengan dadas por la propia experiencia y no por las características de la herramienta.

La visualización obtenida queda se ajusta en gran medida a lo planificado, con algunas diferencias menores en lo relativo a la distribución del espacio. El esfuerzo realizado ha sido mayor al que habría sido necesario para conseguir una visualización similar mediante tableau. No obstante, la potencia y versatilidad de la librería empleada justifica la inversión de este esfuerzo adicional. Puede considerarse que la curva de aprendizaje (resultados por unidad de esfuerzo) es logarítmica para tableau y exponencial para d3.js.

A nivel técnico, el proyecto se ha llevado a cabo mediante la librería [d3.js](#), escrita en javascript y liberada bajo licencia BSD. La visualización consiste en una página estática alojada en github pages, que visualiza los datos, almacenados en json y csv, mediante código javascript y gráficos svg.

Agradecimientos

- Dataset: <https://www.peaceagreements.org/>
- Componentes básicos de la visualización: <https://www.d3-graph-gallery.com/>
- Iconos: <https://www.flaticon.com/> Autores: freepik, smashicons y eucalyp.

Bibliografía

[1] Codex-atlanticus.it. 2020. *Codex Atlanticus*. [online] Available at: <<https://codex-atlanticus.it/#/>> [Accessed 6 December 2020].

[2] Comicsanscriminal.com. 2020. *Comic Sans Criminal - There's Help Available For People Like You!*. [online] Available at: <<https://www.comicsanscriminal.com/>> [Accessed 6 December 2020].

[3] Procter, R., Vis, F. and Voss, A., 2020. *How Riot Rumours Spread On Twitter*. [online] the Guardian. Available at: <<https://www.theguardian.com/uk/interactive/2011/dec/07/london-riots-twitter>> [Accessed 6 December 2020].

[4] Keim, D., Andrienko, G., Fekete, J., Görg, C., Kohlhammer, J. and Melançon, G., n.d. Visual Analytics: Definition, Process, and Challenges. *Lecture Notes in Computer Science*, pp.154-175.

[5] Heer, J. and Shneiderman, B., 2012. Interactive dynamics for visual analysis. *Communications of the ACM*, 55(4), pp.45-54.

[6] Segel, E. and Heer, J., 2010. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), pp.1139-1148.