

# **Informe del projecte de visualització de dades**

La reconstrucció socioeconòmica als acords de pau

Víctor Boix Bellido

Màster en Data Science

Visualització de dades, PAC3

13/12/2019

## Taula de continguts

<b>1. Introducció.....</b>	<b>3</b>
<b>2. Dades .....</b>	<b>4</b>
2.1. Fonts de dades .....	4
2.2. Neteja i transformació de dades .....	5
2.3. Descripció del dataset .....	6
2.4. Creació de les gràfiques.....	7
2.5. Interaccions amb les dades.....	8
<b>3. Disseny.....</b>	<b>10</b>
3.1. Tipografia i colors .....	10
3.2. Estructura i format .....	10
3.3. Proposta de disseny.....	11
<b>4. Bibliografia .....</b>	<b>13</b>

## Abstract

Aquest treball mostra el procés d'elaboració d'una visualització de dades a partir de la informació dels acords de pau al món des de 1990, proporcionada pel projecte *Peace Agreements* de la Universitat d'Edimburg. L'objectiu de la visualització és representar els acords de pau que inclouen referències a la reconstrucció i el desenvolupament socioeconòmic d'una regió després d'un conflicte, mostrant aspectes com la distribució geogràfica, el nivell de detall i la finalitat dels recursos. En aquest document es detallen les fonts de dades utilitzades, els processos de neteja i transformació de les dades, la creació de les gràfiques i les decisions de disseny que s'utilitzaran per a la creació de la visualització.

## 1. Introducció

Aquest treball parteix de la proposta del professor de l'assignatura de col·laborar amb el projecte *Peace Agreements*, creat per la Universitat d'Edimburg i format per gent que treballa en processos de pau d'arreu del món. La informació que els col·laboradors del projecte han anat recopilant sobre diferents conflictes ha servit per generar una base de dades molt extensa que inclou informació sobre els acords de pau a tot el món des de 1990. L'objectiu del treball consisteix en utilitzar aquest conjunt de dades per a generar una visualització que mostri informació sobre els processos de pau i pugui ser útil als integrants del projecte o a altres persones interessades.

Les motivacions que m'han dut a desenvolupar aquesta proposta són dues. D'una banda, resulta molt interessant treballar amb un conjunt de dades tan extens i totalment desconegut, perquè m'ha obligat a explorar les dades i a cercar un punt de vista interessant per desenvolupar. D'altra banda, resulta motivador realitzar un treball que sobrepassi l'àmbit acadèmic de l'assignatura i pugui resultar útil per a altres persones.

El treball s'ha centrat en estudiar les dades dels acords de pau referents a la reconstrucció i el desenvolupament socioeconòmic d'una regió després d'un conflicte. L'objectiu és donar una visió general sobre les ajudes a la reconstrucció que es destinen a diferents parts del món i, a la vegada, proporcionar informació detallada sobre els recursos destinats a un país, el percentatge d'ajuda humanitària o la distribució de fons internacionals.

La visualització representarà el nombre d'acords de pau a tot el món que inclouen referències al desenvolupament socioeconòmic, mostrant aspectes com la seva distribució geogràfica, el nivell de detall dels acords i la finalitat dels recursos. Haurà de donar resposta a preguntes com:

- Quins són els països que ens els seus acords de pau inclouen més referències al desenvolupament socioeconòmic?
- Quin percentatge d'acords inclouen mencions a l'ajuda humanitària i quins ajudes per a la reconstrucció d'infraestructures?
- Com són els recursos destinats a un determinat país?
- Quin percentatge d'acords rep ajudes de fons internacionals?
- Quins acords detallen terminis i plans de reconstrucció?

## 2. Dades

### 2.1. Fonts de dades

El conjunt de dades utilitzat (*pax\_all\_agreements\_data.xlsx*) es troba disponible a la pàgina web del projecte i conté 1789 registres amb informació sobre els acords de pau signats a tot el món des de 1990; cada registre conté 262 atributs que inclouen característiques generals de l'acord i dades relatives al text de l'acord.

Per a construir la visualització utilitzaré informació sobre la localització dels països que signen l'acord i les mencions a la reconstrucció i al desenvolupament socioeconòmic que hi apareixen. La següent taula mostra els atributs seleccionats de la base de dades per a la realització del treball juntament amb una breu descripció del seu contingut.

Atribut	Tipus	Descripció
<i>Con</i>	text	Nom complet del país o països on s'ha produït el conflicte.
<i>LocISO</i>	categòric	Codi ISO d'un dels països on s'ha produït el conflicte (3 lletres).
<i>Loc2ISO</i>	categòric	Codi ISO del segon país, en cas de conflictes entre països (3 lletres).
<i>Dev</i>	categòric continu	<p>Grau de menció a l'acord sobre la reconstrucció o el desenvolupament socioeconòmic de la regió:</p> <ul style="list-style-type: none"><li>• 0 - Cap menció.</li><li>• 1 - Menció superficial.</li><li>• 2 - Inclou alguns detalls sobre el suport al desenvolupament o la reconstrucció.</li><li>• 3 - Detalla el desenvolupament econòmic, els plans de reconstrucció i el terminis establerts.</li></ul> <p>El valor d'aquest atribut es detalla en 3 subcategories (de tipus binari) que afegeixen informació sobre les referències als recursos per al desenvolupament (només per als acords que tenen un valor <i>Dev</i> &gt; 0):</p> <ul style="list-style-type: none"><li>• <i>DevSoc</i> - Indica si l'acord conté provisions per al desenvolupament socioeconòmic en general.</li><li>• <i>DevHum</i> - Indica si l'acord conté provisions per l'assistència humanitària.</li><li>• <i>DevInfra</i> - Indica si l'acord conté provisions per a la reconstrucció i el desenvolupament de les infraestructures.</li></ul>
<i>IntFu</i>	binari	Indica si l'acord conté referències a fons internacionals.

Taula 1. Atributs seleccionats a la base de dades

## 2.2. Neteja i transformació de dades

Les dades seleccionades per al treball s'hauran de netejar i processar per tal de generar el fitxer necessari per a la visualització. A continuació explico les transformacions que caldrà fer a cada atribut.

El nom del país o països en conflicte es troben a l'atribut *Con*, que no conté cap valor nul i no caldrà modificar-lo.

El codi ISO dels països que signen l'acord es troba distribuït en dues columnes: *Loc1ISO* i *Loc2ISO*. L'atribut *Loc1ISO* conté 44 valors buits, mentre que l'atribut *Loc2ISO* només té 233 valors corresponents a conflictes on hi intervé més d'un país; a més, hi ha conflictes entre més de dos països que no contenen tots els codis corresponents. El primer pas consistirà en inserir tots els codis que falten tenint en compte les dades de l'atribut *Con*; per fer-ho caldrà afegir fins a 7 columnes noves. A continuació, caldrà modificar l'estructura de les dades per tal de tenir el codi ISO en una única columna, és a dir, transformar les dades de format ample a format llarg; això facilitarà la representació geogràfica de les dades i permetrà la consulta i agregació de valors per país. En aquest nou format, les dades d'un mateix conflicte es trobaran duplicades tantes vegades com països hi intervenen; en aquest cas, per evitar modificar els resultats globals he optat per dividir el valor dels atributs quantitatius entre el nombre de països.

L'atribut *Dev* indica el grau o nivell de les referències a la reconstrucció i al desenvolupament socioeconòmic que es troben a l'acord. Es tracta d'un atribut de tipus categòric continu, amb valors entre 0 i 3, i que no conté cap valor nul. No caldrà aplicar-ho cap transformació, però a la llegenda resultarà important afegir-hi una descripció del significat dels seus valors.

Les subcategories de l'atribut *Dev* es troben a les columnes *DevSoc*, *DevHum* i *DevInfra* i només inclouen informació per als països que tenen alguna menció al desenvolupament, és a dir, per als quals el valor de *Dev* és diferent de zero. Es tracta d'atributs binaris que indiquen amb el valor '1' si els tipus de recursos mencionats a l'acord són: socioeconòmics, humanitaris o d'infraestructures. Aquests valors caldrà agrupar-los per països i per grau de menció de l'acord; el procés de transformació consistirà en agregar les dades per codi ISO i grau de menció a la reconstrucció i sumar el valors d'aquestes columnes. D'aquesta manera obtindrè 3 atributs nous de

tipus quantitatiu que indicaran quantes referències als recursos per al desenvolupament es troben a cada país i per cada grau de menció.

L'atribut *IntFu* també és de tipus binari, en aquest cas caldrà fer una transformació semblant a l'anterior, agregant les dades per país i grau de menció i sumant-ne els valors. Aquest nou atribut també serà quantitatiu i indicarà quantes mencions a fons internacionals es troben per a cada país i per a cada grau de menció al desenvolupament.

Finalment, caldrà generar un atribut derivat que compti el nombre d'acords de pau que inclou cada agrupació i que servirà per poder comparar valors i calcular percentatges.

Totes aquestes transformacions les realitzaré utilitzant dues eines. En primer lloc, *Microsoft Excel* em permetrà fer una primera exploració de les dades, seleccionar les columnes necessàries i introduir manualment els valors buits. Una vegada generat un primer fitxer CSV amb les dades necessàries, utilitzaré la llibreria *Pandas* de *Python* per carregar les dades i processar-les de manera més elaborada.

## 2.3. Descripció del dataset

Una vegada efectuat tot el procés de neteja i transformació de dades es disposarà d'un fitxer CSV amb les dades que es mostren a la següent taula.

Nom	Valor	Descripció
País	text	Nom del país o països signants de l'acord
Codi	categòric	Codi ISO del país (3 lletres)
Desenvolupament	categòric continu	Grau de menció a la reconstrucció o desenvolupament (nombre enter entre 0 i 3)
Nombre d'acords	quantitatiu	Nombre d'acords de pau
Socioeconòmic	quantitatiu	Nombre de mencions al desenvolupament socioeconòmic
Ajuda humanitària	quantitatiu	Nombre de mencions a l'ajuda humanitària
Infraestructures	quantitatiu	Nombre de mencions al desenvolupament d'infraestructures
Fons internacionals	quantitatiu	Nombre de mencions a fons internacionals

Taula 2. Atributs del conjunt de dades utilitzat a la visualització

Aquest fitxer l'utilitzaré per a construir les diferents gràfiques que formen la visualització. Per treballar utilitzaré l'eina *Tableau Desktop*, calculant atributs derivats a través de fórmules i programant accions per generar filtres. Els processos de creació de les gràfiques i de les relacions i interaccions amb les dades es detallen a continuació.

## 2.4. Creació de les gràfiques

L'atribut "Nombre d'acords de pau" serà la base de la visualització i servirà per mostrar les dades de cada país en un mapa de color (*figura 1*). Aquest mapa permetrà localitzar visualment aquells països que més acords tenen amb mencions per al desenvolupament i comparar-los amb la resta. La situació dels països al mapa l'obtindrè a partir del valor del codi ISO del país. L'atribut nom del país l'utilitzaré únicament a nivell descriptiu, per a mostrar informació addicional al mapa sobre els països en conflicte.

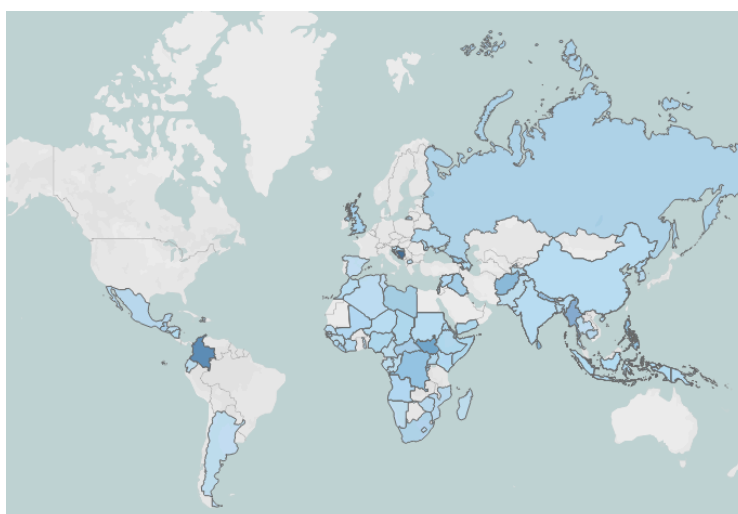
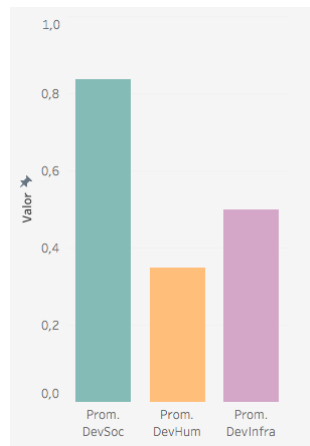


Figura 1. Mapa de color que representa el nombre de mencions a recursos per a al desenvolupament per país.

La visualització també mostrarà la distribució dels recursos per al desenvolupament en funció de la seva finalitat. Els atributs "Socioeconòmic", "Ajuda humanitària" i "Infraestructures" donaran informació desglossada sobre el percentatge de referències a cada tipus de recurs que inclouen els acords seleccionats. Aquest percentatge el calcularé a partir del nombre d'acords i el resultat es mostrarà en un diagrama de barres d'eix fix (*figura 2*). Aquesta representació permetrà comparar-ne

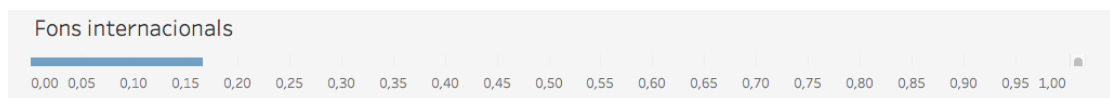


els valors i inspeccionar com és la distribució d'aquests recursos en funció del país o del grau de la referència al desenvolupament.



*Figura 2. Diagrama de barres que mostra el percentatge de referències a cada tipus de recurs*

L'atribut "Fons internacionals" proporcionarà informació addicional que utilitzaré per donar un altre punt de vista a les dades. El seu valor servirà per calcular el percentatge d'acords que inclouen referències a fons internacionals, es representarà en un altre diagrama de barres (*figura 3*) i permetrà comparar les dades entre països i per a diferents paràmetres de la visualització.



*Figura 3. Diagrama que mostra el percentatge de referències a fons internacionals.*

## 2.5. Interaccions amb les dades

Inicialment, tant el mapa com els diagrames de barres mostraran els resultats tenint en compte les dades de tots els països amb acords de pau. Aquests resultats es podran filtrar a partir de diferents paràmetres, com es mostra a continuació.

El valor de l'atribut "Desenvolupament" servirà per filtrar i comparar els resultats en funció del grau de les referències al desenvolupament. El mapa disposarà d'un filtre per mostrar les dades únicament per a determinats valors de l'atribut; per exemple, només per als acords que detallen un pla econòmic i uns terminis per a la reconstrucció (valor 3). Això permetrà veure i comparar la distribució geogràfica i dels recursos per al desenvolupament en funció de la solidesa dels compromisos de reconstrucció adoptats a l'acord.

La visualització també permetrà estudiar les dades per països. En situar el cursor sobre un país s'obtindran detalls sobre els seus acords de pau i, en seleccionar-lo al mapa, s'actualitzaran els diagrames de barres de la visualització amb les dades del país. Això permetrà obtenir informació detallada i visual per a cada país sobre la seva distribució dels recursos per al desenvolupament i el percentatge de referències a fons internacionals. A més, tots els diagrames de barres representaran els seus valors amb un eix fix, de tal manera que l'usuari tingui sempre un punt de referència per a comparar-ne els valors.

Totes aquestes consultes, filtres i interaccions les implementaré utilitzant el programari *Tableau Desktop* i serviran perquè l'usuari pugui explorar les dades i descobrir-hi tendències o diferències importants.

## 3. Disseny

### 3.1. Tipografia i colors

La visualització tindrà poc text, per aquest motiu utilitzaré una única tipografia que doni unitat i consistència al treball; els contrastos entre textos els aconseguiré a través del pes i la mida de la lletra. La font escollida serà de tipus *Sans Serif*, és a dir, de pal o sense ornamentacions; aquest tipus de font facilita la llegibilitat i li dona al text un aspecte modern i diferent al de la premsa escrita. La font triada és *PT Sans*, de la família *Public Type*, una font lliure i d'àmplia distribució que es troba disponible a totes les plataformes i que evitarà problemes de compatibilitat.

A nivell visual utilitzaré colors suaus amb tonalitats fredes. Per al fons he triat un gris clar, uniforme i molt discret, que doni unitat al treball i eviti desviar l'atenció de l'usuari. El mapa mostrarà tots els països del món i marcarà en diferents tonalitats de blau el valor de l'atribut "Nombre d'acords", que és de tipus enter i expressa una quantitat contínua (*figura 1*). El mapa anirà acompanyat d'un diagrama de barres que mostrarà el percentatge d'acords que inclouen referències als recursos socioeconòmics (verd), humanitaris (taronja) i destinats a infraestructures (violeta). En aquest diagrama de barres (*figura 2*) he buscat colors diferents i no graduals per mostrar el caràcter categòric dels seus valors i contrastar amb el blau del mapa, que representa un altre valor.

### 3.2. Estructura i format

La visualització s'estructurarà en 3 blocs. A la part superior hi haurà una capçalera amb el títol i una breu descripció del contingut i objectius del projecte.

A la part central hi haurà el cos de la infografia que contindrà un mapa del món, ocupant la major part de l'espai, i una barra lateral a la dreta. El mapa serà el bloc dominant de la infografia i mostrarà, a través d'un mapa de color, la distribució geogràfica de l'atribut "Nombre d'acords". La barra lateral contindrà un filtre per a les diferents categories de "Desenvolupament" i un diagrama de barres amb la distribució de recursos en funció de la seva finalitat.

A la part inferior hi haurà un diagrama de barres horitzontal a mode de conclusió de la infografia. Aquest diagrama ocuparà tota l'amplada del treball i representarà,

amb una única barra, el percentatge d'acords que mencionen l'aportació de recursos internacionals. A sota d'aquest diagrama hi afegiré informació sobre l'autor, les fonts de dades i la llicència de l'obra.

El projecte es presentarà en format digital a través d'un enllaç obert a *Tableau Public*. Com ja s'ha comentat, la visualització permetrà als usuaris interactuar i explorar les dades a través d'un filtre i de la selecció de determinats valors, com el país o el tipus de desenvolupament.

### 3.3. Proposta de disseny

A continuació adjunto una proposta de disseny per a la visualització. Aquesta proposta és només una maqueta i l'he construïda únicament per tenir una referència en relació a l'estructura i distribució de l'espai. Les dades utilitzades encara no han estat processades i molts detalls de la visualització com el text, les llegendes i els títols no s'han treballat.

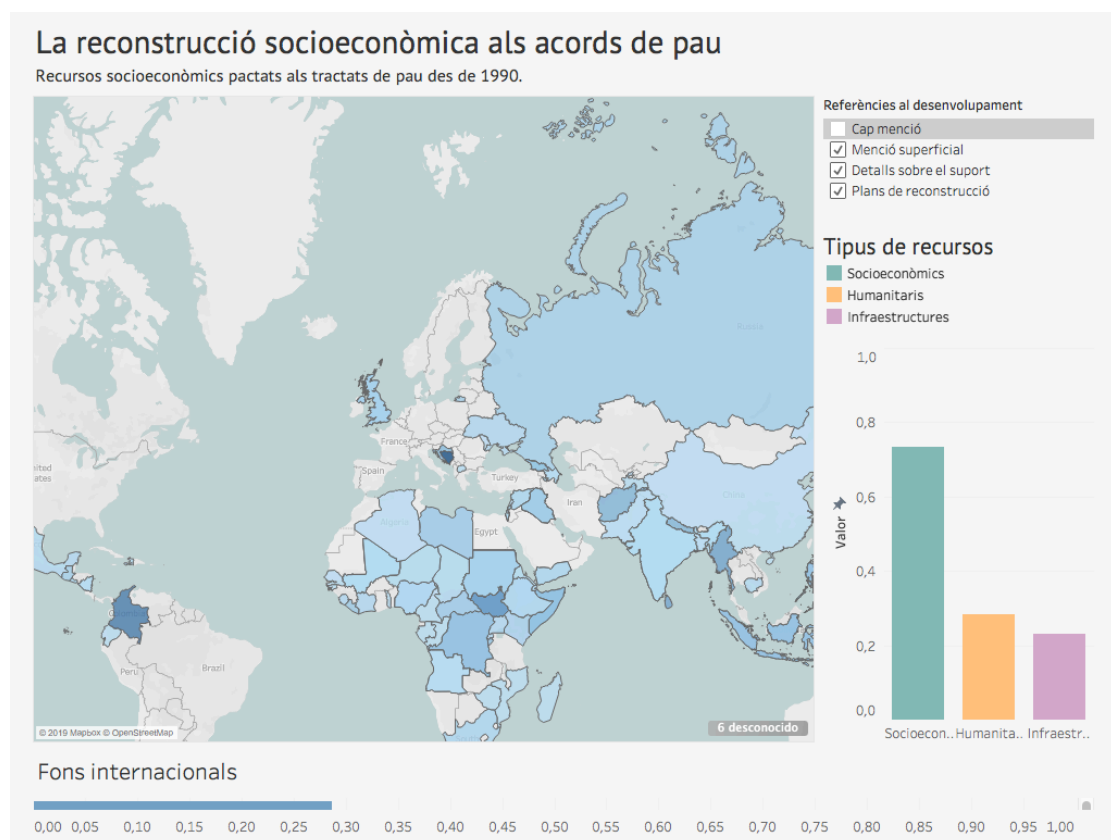


Figura 4. Proposta de disseny per a la visualització.

Com es pot veure, la part més important de la visualització és el mapa, que permetrà consultar i comparar els valors entre països com seleccionar-ne un i visualitzar les dades d'aquest país amb més detall. Al voltant del mapa es mostren diferents elements que ens permeten interactuar-hi a través de filtres i proporcionen informació addicional.

## 4. Bibliografia

- BELL, C.; BADANJAK, S.; FORSTER, R.; JAMAR, A.; MOLLOY, S.; McNICHOLL, K.; NASH, K.; POSPISIL, J.; WISE, L (2019). *PA-X Peace Agreements Database and Dataset, Version 2*. Political Settlements Research Programme, University of Edinburgh, Edinburgh.  
<http://www.peaceagreements.org>
- SKAU, D (2012). *Visually: How to Use Maps in Data Visualization*.  
<https://visual.ly/blog/you-are-here-using-maps-in-data-visualization/>
- CAIRO, A. *El arte funcional: Videos sobre el arte funcional*.  
<https://www.elartefuncional.com/contenidos.html>
- *Google Fonts*. <https://fonts.google.com/>
- *Information is beautiful*. <https://informationisbeautiful.net/>