

Visualització de dades PACX

Josep Alòs Pascual

January 8, 2020

1 Introducció

Aquest document presenta el projecte final de visualització de dades sobre el dataset d'acords de pau. S'explicarà com s'ha implementat la visualització

A continuació es presenta l'estructura del document

Contents

1	Introducció	1
1.1	Resum del projecte	1
1.2	Detalls de la visualització	1
2	Compliment dels objectius inicials	2
3	Que s'ha après en el projecte	3
4	Descripció tècnica del projecte	3
4.1	Preprocessat de les dades	3
4.2	Dades addicionals utilitzades	4
4.3	Eines utilitzades	5
5	Passos futurs i punts a millorar	5

1.1 Resum del projecte

Per determinar com es protegeixen els grups socials d'un territori després d'un conflicte, en aquest projecte s'estudia com diferents acords de pau d'entre el 1990 i 2016 determinen mesures de protecció per aquests grups. En aquest document es presenten les dades i una visualització que serveix per reflectir les conclusions d'aquest estudi de forma senzilla d'entendre però a la vegada útil.

1.2 Detalls de la visualització

- **Títol de la visualització:** Protecció dels diferents grups socials en acords de pau
- **URL de la visualització:** <https://pax-protection-analysis.herokuapp.com/>
- **URL del codi font:** https://github.com/josepalos/pax_protection_analysis
- **Llicència del projecte:** *Creative Commons Attribution-NonCommercial-ShareAlike 4.0 Internacional*¹
- **Dades utilitzades:** Conjunt de dades sobre 1.500 acords de pau des del 1990 fins l'actualitat; recopilats per la universitat d'Edinburg. Pot ser obtingut de la seva pàgina <https://www.peaceagreements.org>, i es troba llicenciat sota la mateixa llicència que aquest projecte.

2 Compliment dels objectius inicials

Inicialment es va proposar que aquest projecte complís amb els següents objectius:

1. Oferir una visualització interactiva sobre les dades dels acords de pau mitjançant l'eina **Tableau**.
2. Preprocessar les dades rebudes, eliminant dades no utilitzades i/o convertint les variables que s'utilitzen al format adequat; utilitzant **Python**.
3. Centrar la visualització a l'estudi de les proteccions que reben els diferents grups d'una societat després d'un conflicte.
4. Avançar des d'una visió general de les dades cap a visualitzacions específiques sobre el tema d'interès.

El primer objectiu s'ha complert, si bé no de la forma esperada. S'ha optat per canviar l'eina que genera la visualització de **Tableau** a **D3**. Això s'ha fet ja que, un cop el desenvolupador s'ha familiaritzat amb els conceptes bàsics d'aquesta nova llibreria, dona més flexibilitat a l'hora de dissenyar i implementar les visualitzacions.

El segon objectiu s'ha complert, però altre cop canviant l'eina utilitzada. S'ha optat per dur a terme el preprocessat de les dades mitjançant **JavaScript** en el moment de carregar el conjunt de dades a l'aplicació. Aquest canvi s'ha fet ja que, un cop no tenim la restricció de **Tableau** que no es poden preprocessar les dades amb la facilitat que dona un *script* escrit específicament per això, a la vegada que es fa servir **JavaScript** per implementar l'aplicació, no cal afegir un pas que s'ha d'executar a priori de desplegar la visualització.

¹<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Per evaluar els dos últims punts, s'ha d'especificar quines visualitzacions s'havien previst inicialment:

1. Visió general de les dades
 - (a) Evolució de la quantitat d'acords durant els anys
 - (b) Mapa mundial
2. Visió focalitzada en mostrar la distribució dels acords segons quin nivell de protecció ofereixen als diferents grups.
 - (a) Quantitat d'acords que estableixen alguna protecció per algun grup cada any.
 - (b) Quantitat de grups protegits segons el nivell i acords
3. Relacions interessants entre variables dels acords.
 - (a) Relació entre els tipus d'acords i la protecció dels grups
 - (b) Relació entre la protecció dels diferents grups
 - (c) Relació entre la protecció dels grups i mencions al *National Human Rights Institution (NHRI)*

S'han desenvolupat totes les visualitzacions previstes amb l'estructura especificada, per tant també s'ha complert els dos últims objectius. L'última visualització, però, s'ha canviat d'un núvol de punts a un gràfic amb cercles de diferents àrees segons el valor que es vol mostrar, per ser consistent amb el gràfic que es té al costat. A més, en aquests dos gràfics s'ha afegit l'opció de canviar quines variables es comparen mitjançant un selector sota del gràfic.

Referent a la utilitat de la visualització per mostrar el que es vol, considero que en general sí que és una eina bona ja que s'ha seguit el disseny proposat a la pràctica 3, i s'ha comprovat que ens permet evaluar de forma ràpida les relacions entre les variables. Tot i això, els dos últims gràfics no aporten tanta informació com s'esperava, degut a que el pes dels acords que no protegeixen els grups (en el gràfic corresponen al 0, 0) és considerablement superior als altres acords, i no s'aconsegueix veure amb claredat les relacions de protecció.

3 Que s'ha après en el projecte

Durant el desenvolupament d'aquest projecte he après el procés a seguir durant el disseny i implementació d'un projecte de visualització (l'anàlisi de les dades, proposta de disseny, implementació. . .). A més a més, he après a fer servir eines per implementar visualitzacions de dades: **Tableau** en les primeres entregues de les pràctiques, i **D3** en la darrera entrega. Si bé no he explorat molt l'eina **Tableau**, de **D3** he après molt sobre aquesta llibreria, com per exemple com gestiona l'actualització dels elements de la pàgina un cop canvien les dades. A més, m'ha permès integrar-ho amb **React**, un *framework* que estava aprenent

pel meu compte paral·lelament, i per tant m'ha servit per practicar aquesta altra eina. Per últim, també comentar que m'ha semblat molt interessant les dades que s'han fet servir, i si bé ara mateix no en tinc cap ús pràctic a part d'ampliar el coneixement, en un futur és possible que torni a utilitzar les dades descobertes.

4 Descripció tècnica del projecte

En aquest apartat s'explicarà com s'ha dut a terme la implementació de l'aplicació.

4.1 Preprocessat de les dades

Primer de tot s'ha d'eliminar les columnes de dades que no necessitem per l'anàlisi. Els atributs que s'han conservat són:

- Con: País (o països) involucrats en l'acord de pau
- Contp: Tipus de conflicte
- AgtId: Identificador del tractat
- Agt: Nom del tractat
- Dat: Data del tractat. Només s'ha conservat l'any
- GCh: Protecció dels infants
- GDis: Protecció de persones discapacitades
- GAge: Protecció de la tercera edat
- GMig: Protecció de treballadors migrants
- GRa: Protecció de grups racials, ètnics, o nacionals
- GRe: Protecció de grups religiosos
- GInd: Protecció de grups indígenes
- GRef: Protecció de refugiats
- GSoc: Protecció de classes socials
- GOth: Protecció d'altres grups
- HrNi: Provisió d'un cos que controli els drets humans.

Per cada variable de protecció s'ha conservat també les variables de cada nivell. Aquestes són "nom de grup"+Rhet,Antid,Subs,Other.

Per últim, s'ha precalculat un agregat de cada tractat de pau, que consisteix en el recompte de grups que estan protegits per cada nivell, obtenint així tres variables addicionals

- rhetoricalCount
- antiDiscriminationCount
- substantiveCount
- otherProtectionsCount

4.2 Dades addicionals utilitzades

Per generar el mapa s'ha utilitzat les dades proporcionades per **TopoJSON** corresponents a l'àtlas mundial. Es poden obtenir des de la pàgina del projecte (<https://github.com/topojson/world-atlas>) o des de la pàgina *UNPKG*². S'ha preprocessat aquestes dades per tal de tenir un mapeig entre els identificadors del país i el seu nom. A més, s'ha especificat un mapeig addicional per a que els noms dels països coincideixin amb els del conjunt de dades dels acords de pau.

4.3 Eines utilitzades

El projecte s'ha implementat utilitzant el *framework* **React** per tal de gestionar l'estructura de l'aplicació. Aquesta eina permet partir l'aplicació en components reutilitzables per tal de garantir les bones pràctiques de disseny.

Dins d'aquest *framework* s'han creat components per cada una de les visualitzacions. Totes les visualitzacions menys dues s'han implementat utilitzant un element *SVG* controlat per la llibreria **D3**. Les altres dues s'han implementat utilitzant una taula, generada de forma dinàmica amb **React**. Per genera el mapa del món s'ha utilitzat **D3** juntament amb la llibreria **TopoJSON**³.

La càrrega de dades es fa de forma asíncrona un cop es carrega l'aplicació, i les dades es guarden al seu estat. Això fa que, un cop s'han carregat, **React** actualitza els components que la fan servir, i aquests components són els encarregats de passar aquestes dades a la llibreria de visualitzacions.

5 Passos futurs i punts a millorar

A continuació s'esmenten possibles canvis i millores a implementar en futures iteracions del projecte:

- Utilitzar **Bootstrap** per fer la pàgina *responsive*.
- Millorar els processos de càrrega, parseig, i mostra de les dades. Actualment no suposen cap problema ja que el volum de dades usades és petit, però personalment considero que el codi pot millorar un cop tingui més domini de **React** i **D3**.

²UNPKG és una pàgina que actua com a proveïdor de continguts per **NPM**

³<https://github.com/topojson/topojson>

- Afegir filtres a les visualitzacions (per exemple, a la quarta visualització afegir la possibilitat de filtrar per anys).
- Afegir més informació als *tooltips* que apareixen en situar el ratolí sobre els diferents gràfics.
- Afegir textos explicatius a cada gràfic, ja sigui amb una capa desplegable o sota d'aquests.