

PAC 4: Projecte de visualització de dades

Assignatura: Visualització de dades

Alumne: Víctor Boix

Data de lliurament: 08/01/2020

1. Presentació de la visualització

Desenvolupament socioeconòmic als acords de pau

<https://public.tableau.com/profile/vboix2#!/vizhome/Acordsdepau/Visualitzacio>

Aquesta visualització de dades representa el nombre d'acords de pau que inclouen recursos per a la reconstrucció i el desenvolupament socioeconòmic d'una regió després d'un conflicte, mostrant aspectes com la seva distribució geogràfica, la finalitat dels recursos o el nivell de detall de l'acord.

Tota la informació tècnica del projecte (documents, dades, codi i llicència) es troba disponible al repositori: <https://github.com/vboix2/Peace-Agreements>

La visualització de dades i la informació tècnica del projecte també es troben disponibles en anglès als enllaços:

https://public.tableau.com/profile/vboix2#!/vizhome/PeaceAgreements_15783398211110/Dashboard

https://github.com/vboix2/Peace-Agreements/blob/master/README_en.md

2. Projecte

Aquest projecte s'ha centrat en estudiar les dades dels acords de pau que fan referència al desenvolupament socioeconòmic d'una regió després d'un conflicte, donant una visió general sobre la seva distribució. La visualització està formada per un mapa de color que representa el nombre d'acords de pau signats per països; aquestes dades poden filtrar-se segons si els acords inclouen recursos per a la reconstrucció socioeconòmica i també en funció del nivell de detall amb què es

menten aquests recursos. A més, s'inclouen dades sobre la finalitat dels recursos en forma de diagrama de barres, indicant el percentatge d'acords que inclouen suport per al desenvolupament socioeconòmic, l'ajuda humanitària i la reconstrucció d'infraestructures. Finalment, s'afegeix informació addicional que indica el percentatge d'acords que inclouen fons internacionals.

Es tracta del meu primer projecte de visualització de dades, això ha fet que em trobi amb certes dificultats i limitacions, però també m'ha permès adquirir molts coneixements nous sobre el disseny i l'elaboració d'una visualització de dades.

Una dificultat important ha sigut les dimensions i la complexitat del conjunt de dades; per aquest motiu vaig decidir centrar-me en un aspecte concret de les dades, el desenvolupament socioeconòmic, i explorar-lo a fons. L'exploració de les dades em va permetre descobrir moltes relacions i tendències interessants, de les quals només una petita part he pogut plasmar a la visualització final. En aquest sentit, he après que durant la fase d'exploració del conjunt de dades resulta molt important focalitzar els esforços en un únic aspecte que consideri interessant.

Un limitació important a l'hora d'iniciar el projecte és que no havia treballat mai amb un programari de visualització de dades. Sabent que *Tableau Desktop* és una eina de referència en aquest camp i disposant d'un llicència per treballar-hi, vaig decidir utilitzar aquest programari tant per a l'exploració de dades com per a la construcció de la visualització. Tota aquesta pràctica m'ha permès aprofundir força en aquesta eina i adquirir-hi certa competència. Malgrat tot, una limitació important ha sigut que durant la fase de disseny no tenia prou pràctica amb *Tableau* i no podia tenir en compte les possibilitats i les limitacions que m'ofereix aquesta eina. Tot això m'ha permès aprendre que en un projecte d'aquestes característiques és important dissenyar la visualització únicament a partir de les dades i, posteriorment, escollir l'eina o tecnologia que millor s'adapti als objectius establerts.

En general, estic molt satisfet amb el resultat final. Tot i que el conjunt de dades és força complex i requereix tenir coneixements sobre el tema o estar-hi familiaritzat, penso que la visualització resumeix molt bé el punt de vista que he volgut mostrar i permet explorar aquest aspecte fàcilment i de manera visual. En tractar-se del meu primer projecte en aquesta disciplina, he hagut d'invertir-hi molts esforços i dedicar-hi moltes hores que no sempre han resultat productives. Malgrat tot, m'ha permès aprendre molt sobre el disseny i la construcció de visualitzacions de dades i m'ha

ajudat a valorar encara més moltes de les grans visualitzacions que he anat consultant durant aquest temps.

3. Descripció tècnica

Les dades utilitzades formen part del projecte *Peace Agreements* de la Universitat d'Edimburg. El conjunt de dades *pax_all_agreements_data.xlsx* es troba disponible a la pàgina web del projecte i conté 1789 registres i 262 atributs amb informació sobre els acords de pau signats a tot el món des de 1990. Per a l'elaboració de la visualització he seleccionat el nom i l'identificador dels països que signen l'acord (*Con*, *Loc1ISO* i *Loc2ISO*) i les metadades referents a la reconstrucció i el desenvolupament socioeconòmic pactats a l'acord, que són: el nivell de detall dels recursos mencionats a l'acord (*Dev*); si l'acord inclou recursos per al desenvolupament socioeconòmic (*DevSoc*), ajuda humanitària (*DevHum*) i per a infraestructures (*DevInfra*); i si l'acord inclou referències a fons internacionals (*IntFu*).

La selecció dels atributs l'he realitzada editant el fitxer XLSX amb *Microsoft Excel*; a més, ha sigut necessari un primer procés de neteja per tal d'inserir valors buits a les dades. Els atributs *Loc1ISO* i *Loc2ISO* contenen un codi de tres lletres (segons l'estàndard *ISO 3166-1 alpha-3*) que permet identificar els països que intervenen al conflicte; aquests atributs contenen valors buits que he hagut de completar manualment a partir del nom del país. A més, he necessitat afegir 7 atributs més (*Loc3ISO*, *Loc4ISO*...) per poder disposar d'un identificador per als conflictes entre més de dos països. El fitxer resultant l'he exportat en format CSV i es troba disponible al repositori com a *data.csv*.

El procés de neteja i transformació de dades l'he realitzat amb *Python*, utilitzant les llibreries *NumPy* i *Pandas*. El codi font es troba a l'enllaç:

<https://github.com/vboix2/Peace-Agreements/blob/master/data/transform.py>

Aquest fitxer importa les dades en un *DataSet* de *Pandas* i converteix els 9 atributs que contenen el codi dels països del seu format ample a un format llarg, on es dupliquen els registres i es guarda el codi del país en un únic atribut (*Loc*). En aquesta nova representació es disposa d'un registre per cada país i conflicte i facilita la representació geogràfica de les dades. A més, he dividit la resta d'atributs quantitatius entre el nombre de països a cada conflicte per evitar modificar els resultats globals. El

fitxer també genera un nou atribut que representa el nombre d'acords (*N_Acords*) i agrega les dades per països i pel nivell de detall de l'acord (*Dev*), sumant el nombre d'acords de cada tipus. Les dades processades s'exporten en un nou fitxer *data_transformed.csv*, que també es troba disponible al repositori.

La visualització s'ha elaborat amb *Tableau Desktop* a partir de les dades anteriors. Aquestes dades s'han unit amb les del fitxer *CountryCodes_alpha3.csv* per obtenir el nom dels països a partir del seu codi. El codi del país s'ha utilitzat per representar el nombre d'acords de pau de cada país en un mapa de color; aquests resultats es poden filtrar segons el nivell de detall dels acords, on els seus valors s'han substituït per un àlies per facilitar la interpretació de la visualització. A més, s'ha utilitzat el valor de l'atribut Nombre d'acords per calcular atributs derivats i poder representar el percentatge d'acords que inclouen recursos socioeconòmics, ajuda humanitària, desenvolupament d'infraestructures i fons internacionals. Finalment, s'han programat accions per mostrar aquests percentatges per països o a partir del filtre.

Per acabar, aquest treball l'he protegit amb una llicència *Creative Commons Reconeixement 4.0 Internacional (CC BY 4.0)*. Aquesta llicència permet la distribució, modificació i l'ús comercial o privat d'aquesta visualització sempre que se'n citi l'autor, la font original i les modificacions efectuades.