

Sequences sunburst visualization for Peace in the American continent

ANDRÉS LAVERDE MARÍN*

Universitat Oberta de Catalunya
alaverde@uoc.edu

8 de gener de 2020

Resum

Aquest document pretén ser un complement a la visualització final, per tant es tracta d'una continuació dels informes presentats anteriorment. En aquest document definirem l'abast i les raons que ens han portat a dur a terme aquesta visualització final. Per tant en aquest document no definirem res de nou, sinó que argumentarem les decisions que hem pres per obtenir la nostra representació final, definint les dificultats amb les quals ens hem trobat al llarg de la representació i la comparació de l'abast que hem pogut assolir en funció de l'abast definit anteriorment.

I. TÍTOL SELECCIONAT, URL

El títol seleccionat finalment ha estat el mateix que hem definit a la PAC3, ja que concorda perfectament amb la visualització permeten tenir diferents graus d'informació depenent del que l'usuari vulgui saber i filtrar per obtenir informació respecte a un determinat tema.

Podem veure la nostra visualització i els seus components a un GitHub Gist [1], però per veure la visualització sencera i interactuar correctament amb ella hem d'accedir a la pàgina web [2]. A la pàgina de GitHub Gist [1] podem trobar una descripció que significa la variable amb l'objectiu de facilitar al lector la visualització a la pàgina web pròpia d'aquesta [2].

II. RELACIÓ RESPECTE A PACS ANTERIORS

La visualització final ha patit diversos canvis respecte a la definició feta a la PAC3, ja que en fer diferents proves a més atributs veiem que com que les dades majoritàriament no presenten referència als diferents grups i categories de la visualització, Figura 1.

*Data Science Master's degree student

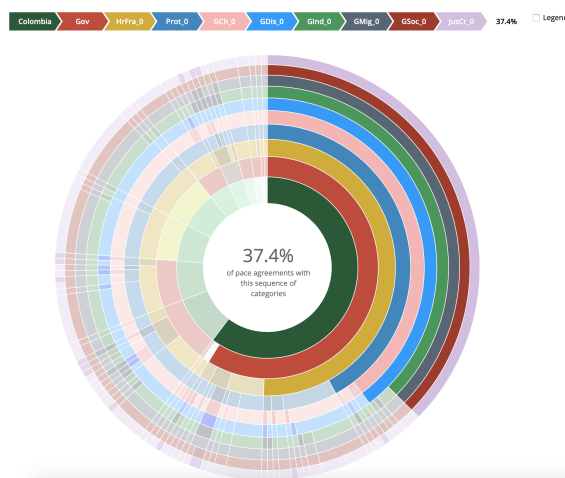


Figura 1: Representació del major grup de dades trobat tras la visualització.

És a dir, les dades no fan referència a cap grup (Gch, GDis, GInd, GMig, GSoc), mesures de protecció (Prot), drets humans (HrFra) o justícia criminal (JusCr, JusCrSp, JusCrSys, JusCrPow, JusNoRef), a causa d'aquest fet hem decidit seleccionar aquelles dades que aportaven més grups diferents de 0 i que ens serveixen per respondre la hipòtesi plantejada a la PAC3, per tal de no crear una visualització

extremadament complexa i que l'usuari pugui obtenir informació ràpidament sense necessitat de conèixer característiques minoritàries.

En altres paraules, hem decidit, reduir la quantitat d'atributs presents a la visualització que havíem definit en la PAC3, ja que aquests atributs eren minoritaris i estratificaven molt la nostra visualització dificultant l'enteniment, el qual és l'objectiu principal d'una visualització. Per tant, aquesta és una bona visualització per les nostres dades, tot i que, si volíem representar tot el conjunt de dades que teníem previst, hauríem d'augmentar tot la mida de la visualització dificultant el seu enteniment, per tant considerem que és una bona visualització, encara que en augmentar les dades augmenta la complexitat d'aquesta, però encara que sigui complexa, et permet obtenir informació nova, ràpidament, sense necessitat d'analitzar totes les dades. Al marge de no haver seleccionat totes les variables definides inicialment, hem mantingut la definició de colors feta a la PAC anterior, però no hem pogut afegir el filtratge per anys definit anteriorment, ja que, per dificultats de coneixement i aprenentatge de D3 no hem estat capaços de dur a terme aquest filtratge, creiem que amb una mica més de temps per llegir i aprendre D3 hauria estat possible aconseguir-ho.

El resultat obtingut després de fer la visualització ens permet respondre la nostra hipòtesi a més de permetre extreure més informació de la que necessitem inicialment per nosaltres gràcies a la interacció que poden tenir els usuaris per filtrar aquesta, per tant, els usuaris, poden extreure la seva pròpia informació en veure la nostra visualització. Per tant després d'haver fet aquesta visualització podem veure que hem obtingut una visualització dinàmica i coherent amb els objectius definits inicialment, encara que no hem implementat tot el que ens hagués agradat, és una primera visualització coherent i robusta per un usuari que no estigui molt familiaritzat amb aquest tipus de temes. Per tant el fet de reduir el nombre d'atributs s'ha fet amb l'objectiu de disminuir la dificultat de comprensió de la visualització. Per tant tot el treball fet prèviament ens ha portat a aquesta

visualització, és a dir, gràcies a tots els conceptes que hem après en aquesta assignatura hem pogut dur a terme una bona visualització, per tant, després de tot l'esforç hem pogut fer una bona visualització, també cal dir que amb una mica més de marge de temps i pot ser uns apunts sobre D3 aquesta visualització hagués pogut arribar a l'excel·lència, ja que en l'àmbit personal la corba d'aprenentatge de D3, tant per dur a terme la visualització com per configurar totes les eines necessàries per veure-la correctament, ha estat molt més alta del que m'hagués agradat per millorar la meua representació.

III. DESCRIPCIÓ TÈCNICA DEL PROJECTE

Per dur a terme aquesta visualització s'ha fet servir principalment python per la neteja i creació del dataset i D3 juntament amb html, css i GitHub Gist per la creació de la visualització. Per facilitar l'enteniment de la descripció tècnica del nostre projecte ho farem d'acord amb les fases per dur a terme la representació.

1. **Netejada de dades:** Per dur a terme la neteja de dades d'aquesta visualització, s'ha fet servir Python, on s'han aplicat els algorismes descrits a la PAC3, podem trobar, tota la neteja de dades al fitxer *Data_clean.ipynb*, a la ruta *visualizatoin/data/* dintre del nostre repositori de GitHub [3].
2. **Preparació de dades:** Per dur a terme la preparació de dades de la visualització, hem hagut de transformar les nostres dades en transaccions per dur a terme la nostra representació, per fer-ho també hem fet servir Python i ho podem trobar en el mateix fitxer definir anteriorment del nostre repositori [3].
3. **Visualització de dades:** Per dur a terme la visualització de les dades hem fet servir la llibreria D3, ho podem trobar al fitxer *sequences.fi.js* que el trobem a la ruta *visualizatoin/* del repositori, en aquest fitxer definim el color de cada atribut i la lectura

del nostre .csv que hem generat als apartats anteriors, un cop fet això, per poder veure la visualització s'ha de definir les característiques del html que es troben al fitxer *sequences_f.css* i per últim unifiquem tots els fitxers en un índex html *index.html* que és el que ens permet veure la visualització, com que el fitxer *sequences_fi.js* és només JavaScript en fer doble click el navegador no l'interpreta com una plana web per tant hem generat una url [2] per poder veure la nostra visualització amb GitHub Gis, d'aquesta manera podem veure la nostra visualització sense necessitat de crear un servidor per veure-la.

i. Comentari

Si volem augmentar les transaccions, és a dir, afegir els atributs que hem decidit no afegir en aquesta visualització, només s'ha de fer una extensió de codi present en *Data_clean.ipynb* [3], d'aquesta manera es poden generar diferents transaccions i contar la freqüència d'aparició de cadascuna.

També val la pena comentar en aquest punt, que tal com havíem pensat la visualització a la PAC3 eliminàvem aquella branca que no fes referència a un determinat atribut, finalment no ho hem fet així, ja que en fer-ho d'aquesta manera perdríem informació a la representació, d'aquí que la nostra visualització tingui els mateixos nivells en totes les categories Figura 1 [2].

També cal dir que per vaure la visualització en una url hem creat un GitHub Gis [1] que ens permet veure tots els fitxers descrits al punt **Visualització de dades** descrit anteriorment.

Tota la documentació tècnica respecte a aquest projecte s'ha fet a la PAC3 i per tant no la repetirem en aquesta PAC.

IV. VISUALITZACIÓ REALITZADA

Tal com hem dit al principi d'aquest document l'objectiu d'una visualització és facilitar l'enteniment i processament de dades i l'objectiu d'aquesta pràctica és justament facilitar l'en-

teniment de com han estat signats els acords de pau al continent Americà des de l'any 1990 fins a l'any 2017, per tant la visualització permet a l'usuari simplement passant el ratolí per cascuna de les categories descrites al *README.md* obtenir el percentatge de la combinació de totes les categories prèvies fins a arribar a aquesta, en altres paraules, cada dònut de la nostra visualització actua com a filtre sobre els dònuts següents, d'aquesta manera aquesta visualització permet que l'usuari pugui filtrar per cadascuna de les categories del *README.md*, d'aquesta manera l'usuari en tot moment té la llibertat de decidir fins a quin punt aplica el límit de la visualització i quina informació extreu d'aquesta dada, d'aquí el títol *Pace in American Continenen*, que no limita en cap moment a una única informació sinó que l'usuari filtra i opte informació en funció del filtre aplicat, és a dir, en aquest cas les dades estan acotades només al continent Americà, cal tenir en compte que en aquesta visualització l'usuari només té uns atributs per realitzar el filtre, els definits al *README.md* i no tots els de la base de dades, però aquests atributs presents han estat seleccionats de tal manera que s'obté el màxim d'atributs amb "_0", és a dir que tenen algun tipus d'informació més interessant per l'usuari, d'aquesta manera no saturem a l'usuari d'informació i té els atributs més rellevants per entendre els tractats de pau i els conflictes als diferents països del continent Americà.

Aquesta pràctica és un complement perfecte de la PAC anterior, ja que el que fa és reduir les dades a aquelles més rellevants per l'usuari i per la nostra hipòtesi intentant contestar aquesta i permeten que els usuaris puguin extreure les seves pròpies conclusions de les dades, és a dir, aquesta pràctica és una continuació de la PAC3 millorant la visualització definida i contestant la hipòtesi inicial d'aquella pràctica.

Finalment s'ha escollit aquesta visualització per la facilitat d'interacció que permet a l'usuari interactuar amb les dades i extreure informació ràpida dels atributs més rellevants de les nostres dades, a més, podem afegir més atributs en un futur en funció de les necessitats que es vulguin representar o els fets que

vulguin destacar de totes les dades. Per tant la selecció d'aquesta visualització ha estat degut a la facilitat d'interacció i l'escalabilitat que presenta en enfront de les dades.

Una millora possible de la nostra visualització seria fer el mateix gràfic únicament per grup minoritari, d'aquesta manera es podria veure en detall cada grup minoritari i quin tipus de referència es fa, nosaltres no ho hem fet, ja que no ens hem centrat únicament en un sinó que hem agafat els més rellevants i els que hem considerat més rellevants per la nostra hipòtesi.

V. CONCLUSIONS

Ala PAC3 vam plantejar la hipòtesi la gran diferència d'acords de pau signats pels diferents països del continent Americà. Per tant després de veure la nostra representació [2], veiem que no tots els països han signat acords de PAU, ja sigui perquè no han tingut processos violents o perquè encara es troben en aquella situació, també podem afirmar que Colòmbia ha estat el país amb més acords de pau signats en aquestes dates, de fet representa un 60.3% dels acords de pau signats en aquest any i un 60.3% d'aquests acords són de tipus governamental, és a dir, són conflictes produïts per ideologies o disputes polítiques, de tots els acords de pau signats durant aquest any un 8.41% han estat acords de pau signats a Colòmbia, són del tipus governamental i han fet referència als drets humans, en diferents nivells.

Per tant, com veiem amb aquest gràfic podem extreure conclusions amb percentatges i comparar els valors pels diferents països així com veure quin tipus de referència s'ha fet en cada país a un determinat col·lectiu. Tornant a les conclusions respecte a la hipòtesi que vam formular a la PAC3, podem afirmar que aquesta hipòtesi és correcta, tal com esperàvem a la PAC3, Colòmbia ha estat el país amb més acords de pau, el que ens indica que és un país amb una història amb molts conflictes violents però a l'hora els diferents governs que ha tingut han fet grans esforços signant acords per tal d'acabar amb aquells, però a l'hora també

podem afirmar que un 37.4% dels acords signats del total no fan referència a Colòmbia i no fan referència a cap col·lectiu, drets humans o ha acabat jutjant al grup violent com a criminals, per tant un 62% dels acords de pau signats a Colòmbia no fan referència a cap dels grups minoritaris ni als drets humans ni s'ha jutjat de cap manera al grup armat involucrat.

Per tant aquest gràfic valida la nostra hipòtesi inicial definida a la PAC3 i a més ens permet extreure la mateixa informació inclús més informació que la que hem tret nosaltres fent l'anàlisi anterior, en poder extreure aquesta informació i molta més demostra que hem triat una bona visualització per l'objectiu definit a la PAC anterior, a més de permetre extreure molta més informació que la que hem extret nosaltres, com per exemple que Colòmbia no ha signat cap acord de pau territorial, mentre que Equador i Perú només han signat acords de pau de territori, el que ens diu que el conflicte hauria estat per la frontera entre aquests dos països i els grups indígenes s'han protegit en algun dels tractats signats, ja que aquests grups es veuen en un percentatge encara que minoritari.

Dit tot això podem concloure que hem triat correctament la nostra visualització (Com a mínim pels nostres objectius) i a més ens permet explorar molta més informació que no s'havia tingut en compte fins ara o no era rellevant per nosaltres. Per últim també val la pena dir que no seria una bona idea tenir aquesta visualització en format paper perquè tenim molts colors i no ens permet veure el percentatge d'aparició per tant l'usuari no podria extreure la informació de la manera com està plantejada la visualització.

REFERÈNCIES

- [1] Repositori de GitHub Gist amb tots els fitxers per la visualització i una previsualització de la mateixa, <https://bl.ocks.org/alaverma/54aaee4bcb86f8809752259e0546bb33>

- [2] Visualització, <https://bl.ocks.org/alaverma/raw/54aaee4bcb86f8809752259e0546bb33/>
- [3] Repositori GitHub, PAC4_Visualization, https://github.com/alaverma/PAC4_Visualization