

A9: Projecte de visualització (PAC4)

Daniel López Ramírez

M2.959 – Visualització de dades

Índex

1.	Presentació de la visualització	3
2.	Projecte realitzat	4
3.	Descripció tècnica	5

1. Presentació de la visualització

Característiques més influents en els acords de pau intraestats

La visualització la podem trobar al següent enllaç:
https://public.tableau.com/profile/daniel.lopez6830#!/vizhome/PAC4_15917263909770/CaracteristiquesAcords

El codi necessari per a la visualització el podem trobar al següent enllaç:
<https://github.com/dlopezra/VisualitzacioDades>

Aquesta visualització representa els atributs més influents en els acords de pau intraestats, mostrant-los de diverses maneres per a poder cercar de la manera més senzilla les característiques dels acords de pau per país, any o característica.

2. Projecte realitzat

Aquest projecte pretén descriure els acords de pau intraestats (1273 acords) avaluant la distribució de les mencions a grups de persones, la governança i la seguretat. Per a poder generar una visualització que fos clara per aquest projecte, he utilitzat tres tipus de visualitzacions que poden interactuar entre elles i que ens permeten cercar les dades de diverses maneres.

Primer de tot he creat un mapa mundi assignant tots els acords de pau a cada país, on es pot observar segons la tonalitat del color escollit quins països tenen més acords de pau. Per a cada país, es poden obtenir les referències utilitzades en els acords de pau.

També he creat un diagrama de barres amb les utilitzades en els acords, que ens permet veure quines són les referències més utilitzades, i ens permet seleccionar-les per veure els països i anys on han estat referenciades.

Finalment, he creat un mapa de calor, amb els països i els anys dels acords, que ens permet cercar per aquests dos paràmetres els acords de pau. També s'aplica una escala de colors per a identificar els països i anys que tenen més o menys acords.

Al ser el meu primer projecte de visualització i amb la poca experiència amb eines de visualització com Tableau Desktop, la realització del projecte ha estat més difícil del que caldria esperar. Primer de tot, he hagut d'entendre les dades que havíem de tractar, tenint en compte que tenia més de 250 atributs diferents i que les dades a tractar no estan dintre del meu àmbit d'estudi. Un cop s'han entès les dades, s'han hagut de tractar i transformar per utilitzar-les per a una finalitat que havíem definit a priori sense un estudi exhaustiu de les mateixes. El procés de tractament de les dades, seguint amb la idea de visualització, ha permès anar refinant les dades, per obtenir un conjunt de dades més petit amb el que poder treballar.

Amb les dades preparades, m'he enfrontat a la creació de la visualització utilitzant l'eina Tableau Desktop. Aquesta eina, sembla una gran eina per a generar visualitzacions, a més és molt reconeguda, i encara que a primera vista, pot semblar una mica complexa, després et permet generar visualitzacions molt interessants. Tot i això, a vegades és una mica rígida, i no permet realitzar algunes configuracions que poden semblar senzilles. També és veritat, que la poca experiència amb l'eina juga un paper important, però per exemple, no he estat capaç de que quan es selecciona al mapa o al mapa de calor un país, el diagrama de barres quedi ordenat per les característiques més utilitzades, igual que poder ensenyar el text dels tractats en una finestra flotant si es seleccionava un país.

El procediment que hem seguit en aquest projecte, m'ha ajudat a entendre millor com tractar els projectes de dades, fent un primer anàlisi profund de les dades, seguit d'una avaluació i tractament de les dades per a treballar en una idea de visualització, fins a arribar a la definició de la visualització amb les dades necessàries per a ella.

El resultat final s'assembla força a la idea inicial que havia dissenyat, tot i que com he comentat anteriorment, per desconexió de l'eina de visualització, no he aconseguit realitzar algunes tasques que crec que podien ser necessàries dintre del projecte de visualització.

3. Descripció tècnica

Per a la realització d'aquest projecte de visualització, he utilitzat dues eines principalment, R i Tableau.

He partit d'un subconjunt de la base de dades Peace Agreements de la Universitat d'Edimburg, seleccionant només els acords de pau intraestats (1273 acords). Amb els acords descarregats en format csv, tant els acords com els seus textos, els he carregat a R per al seu processament.

Utilitzant la funció *merge* de R he fusionat les dades dels dos fitxers, obtenint un dataset de 1273 registres amb 265 variables, incloent també el text dels acords.

Amb les dades fusionades, s'ha revisat si hi havia valors nuls, eliminant del dataset els atributs amb valors nuls (*Loc1GWNO*, *Loc2GWNO*, *UcdpCon*, *UcdpAgr*, *PamAgr* i *CowWar*).

Un cop eliminats els valors nuls, he realitzat una inspecció visual utilitzant la funció *count* de la llibreria *plyr* per observar les variables amb més inferència a la mostra. D'aquesta manera s'ha observat una possible relació entre els atributs *GRa* i *GRe*, i s'ha creat un nou atribut que interrelaciona aquests dos atributs.

Per aprofundir una mica més en la inferència dels atributs, he decidit fer una matriu de correlació entre totes les metadades de la mostra mitjançant dues funcions de correlació. La primera, *cor*, per extreure totes les dades en brut, i la segona *correlate* de la llibreria *corr* per a poder filtrar les correlacions que tinguessin un valor més alt de 0.6.

Revisant les dades de les variables relacionades amb el país (*Con*, *Loc1ISO* i *Loc2ISO*), s'ha observat que, per una banda, *Con* conté també països relacionats amb els conflictes (com per exemple si el país pertanyia a l'antiga Iugoslàvia) i *Loc1ISO* i *Loc2ISO* contenen els codis ISO dels països, però no el nom. Per tant, utilitzant la llibreria *countrycode* de R, afegirem dues variables més amb el nom del país segons les variables abans esmentades.

Amb els atributs preparats, s'han seleccionat els atributs que tractarem a la nostra visualització, tal com s'ha indicat a la pràctica anterior: *Contp*, *PPName*, *Reg*, *Agt*, *Dat*, *Status*, *Stage*, *Part*, *ThrdPart*, *Country1*, *Country2*, *RelGRaGRe*, *Agreement.Text*, *GCh*, *GDis*, *GAge*, *GMig*, *GRa*, *GRe*, *GInd*, *GOth*, *GRef*, *GeWom*, *GeMe*, *GeLgbti*, *GeFa*, *StDef*, *StGen*, *Cons*, *Ele*, *PolPar*, *SsrGua*, *Ce*, *SsrPol*, *SsrArm*, *SsrDdr*, *SsrPsf*.

Amb els atributs seleccionats, he realitzat la transformació en els valors necessaris per a la seva utilització (text, categòric binari, categòric numèric i data), tal com s'indicava a la pràctica anterior:

Atribut	Tipus	Transformació	
Contp	Text	Categòric numèric	Tipus de conflicte: <ul style="list-style-type: none">- Government → 1- Territory → 2- Government/territory → 3- Other → 4
Status	Text	Categòric numèric	Descriu el tipus d'acord: <ul style="list-style-type: none">- Multiparty signed/agreed. → 1- Unilateral agreement. → 2- Status unclear. → 3- Agreement with Subsequent Status. → 4

Stage	Text	Categòric numèric	Estat de l'acord: <ul style="list-style-type: none"> - Pre → 1 - SubPar → 2 - SubComp → 3 - Imp → 4 - Ren → 5 - Cea → 6 - Other → 7
Part	Text	Categòric binari	1 si les parts han signat l'acord, 0 si no l'han signat.
ThrPart	Text	Categòric binari	1 si altres parts han signat l'acord, 0 si no l'han signat
Gch	Numèric	Categòric binari.	1 si l'acord conté referències sobre nens o gent jove i els seus drets, 0 si no les té.
GDis	Numèric	Categòric binari	1 si l'acord conté referències sobre discapacitats i els seus drets, 0 si no les té
GAge	Numèric	Categòric binari	1 si l'acord conté referències sobre la gent gran i els seus drets, 0 si no les té.
GMig	Numèric	Categòric binari	1 si l'acord conté referències sobre immigrants i els seus drets, 0 si no les té.
GRa	Numèric	Categòric binari	1 si l'acord conté referències sobre la raça, ètnia o minories i els seus drets, 0 si no les té.
GRE	Numèric	Categòric binari	1 si l'acord conté referències sobre grups religiosos i els seus drets, 0 si no les té.
GIInd	Numèric	Categòric binari	1 si l'acord conté referències sobre la gent indígena i els seus drets, 0 si no les té.
GOth	Numèric	Categòric binari	1 si l'acord conté referències sobre altres grups i els seus drets, 0 si no les té.
GRef	Numèric	Categòric binari	1 si l'acord conté referències sobre refugiats i els seus drets, 0 si no les té.
Cons	Numèric	Categòric binari	1 si l'acord conté referències sobre la reforma, revisió o creació de la constitució, 0 si no les té.
Ele	Numèric	Categòric binari	1 si l'acord conté alguna referència a eleccions, 0 si no les té.
PolPar	Numèric	Categòric binari	1 si l'acord conté referències sobre la reforma, regulació, o transformació de partits polítics o grups i moviments rebels, 0 si no les té.
Ce	Numèric	Categòric binari	1 si l'acord conté alguna referència al cessament de les hostilitats, 0 si no les té.
SsrPol	Numèric	Categòric binari	1 si l'acord conté referències sobre la vigilància policial o la reforma de la policia, 0 si no les té.
SsrArm	Numèric	Categòric binari	1 si l'acord conté referències sobre la reforma de les forces armades, 0 si no les té.
SsrDdr	Numèric	Categòric binari	1 si l'acord conté referències sobre la desmilitarització, desarmament, desmobilització i la reintegració, 0 si no les té.
SsrPsf	Numèric	Categòric binari	1 si l'acord conté referències sobre com s'han de tractar als grups rebels o de l'oposició, inclosos en noves estructures polítiques, comissions i cossos, 0 si no les té.

Amb les dades transformades, he guardat el nou fitxer ("pax_alldata_1273.csv").

Amb el nou fitxer de dades generat, l'he importat a l'eina de visualització, que en el meu cas ha estat *Tableau*.

Un cop importat el fitxer, he creat tres fulls de treball de Tableau:

- NumeroTractatsPais: mapa que representa les diverses ocurrencies de les metadades amb els diversos països que componen la mostra.
- OcurrenciesMetadades: gràfic de barres per relacionar les ocurrencies de les metadades amb el total d'acords de la mostra.
- TractatsPaisAny: HeatMap per a poder visualitzar la distribució dels acords de pau i les ocurrencies de les metadades de cada país pels diversos anys que componen la mostra.

També s'han creat quatre atributs addicionals per a donar informació sobre el tipus de conflicte, indicant en cada atribut si és d'algun dels tipus de conflicte. D'aquesta manera podem donar més informació sobre els tipus de conflicte per país, indicant si han estat per temes governamentals, territorials o d'altres.

Aquests tres fulls de treball, són els components del Dashboard de Tableau, que serà el component principal de la nostra visualització.

En aquest Dashboard s'han creat les diverses accions necessàries per interrelacionar els fulls de treball:

- Mapa:
 - Si es selecciona un país, els dos diagrames secundaris, el de barres i el heatmap, ensenyaran informació relacionada només d'aquest país.
 - Si es passa per sobre d'un país, apareixerà un pop-up amb la informació sobre els atributs de les metadades d'aquell país.
- Gràfic de barres (Ocurrencia de les metadades):
 - Si es selecciona un atribut, del gràfic, tant el mapa com el heatmap ensenyaran els països que en els seus acords de pau hi ha referències d'aquest atribut, però a més en el heatmap ho podem veure pels anys.
 - Si es passa per sobre d'una metadada, ensenyarà el nombre d'ocurrencies de la metadada en els acords de pau.
- Heatmap: en el cas del mapa de calor, podem realitzar tres tipus de seleccions, per país o per any o país i any. En ambdós casos el comportament en els altres gràfics serà igual.
 - Si es selecciona un país o un any o ambdós, el mapa ensenyarà la informació seleccionada, en el cas del país, només el país i en el cas de l'any, tots els països que tenen acords de pau d'aquell any. En el cas de les ocurrencies de les metadades, depenent del país i/o de l'any s'ensenyaran els valors acumulatius.
 - Si es passa per sobre, ensenyarà informació sobre les dades tant al mapa de calor com al mapa i a les ocurrencies.

La visualització i les dades per a poder reproduir-la, les he deixat al meu repositori de GitHub i l'he protegit amb una llicència del MIT. L'enllaç és el següent:

<https://github.com/dlopezra/VisualitzacioDades>