

## Visualització de dades

### PAC2: Projecte de visualització de dades

**Notes:** Les respostes a les diferents preguntes estan escrites amb **blau** per tal de ser fàcilment identificables.

Igualment, comentar que tant el títol de la visualització com la descripció curta s'han proposat en anglès, per tal de poder facilitar la col·laboració amb les entitats gestores de la PA-X

**Url visualització:** <https://ilergeta.github.io/PAX-Viz/>

Títol de la visualització on es presenten el dataset o datasets escollits.

El títol escollit és "**Beyond Peace Agreements. Analyzing Peace impacts and distribution in a World context.**"

Descripció curta del document i del que s'hi presenta.

This paper and visualization link above are intended to show and explain several graphs aiming to explore multidimensional data from [PA-X Database](#) and trying to obtain valuable information. Additional data as HDI rank ([UNDP source](#)) is also used to enrich original dataset in order to study relation between sensitive countries and this index.

This work is under a [CC BY 4.0 License](#) and Josep M<sup>a</sup> Espasa Verdés is the author.

Què en sabeu de les dades: tipus, estructura, curiositats

Totes les dades s'han obtingut mitjançant descàrrega directa del lloc web de la PA-X, així, el fitxer principal de dades emprades (que es pot obtenir en aquest [enllaç](#)) recull 1789 acords de pau, dels que es relacionen 266 atributs per cadascun, a part d'aquestes dades, també es poden obtenir els documents dels acords de pau (descarregables a través d'aquest [enllaç](#)) on, a part del document de l'acord, es recullen altres 11 atributs.

La descripció de tots els atributs del fitxer de dades principal es troba detallada en el CodeBook editat per les entitats gestores de la PA-X, que es pot descarregar [aquí](#). Com a resum, indicar que es tracten d'atributs de tipologia molt variada: booleans, nombres enters, dates i cadenes de caràcters. Degut al gran nombre d'atributs del dataset, en aquest document únicament s'analitzaran aquells que s'han considerat més interessants, ja sigui pel significat del propi atribut o pel seu comportament.

Com a curiositat, comentar que en la versió del CodeBook descarregada (19 de febrer de 2018), s'ha trobat que l'atribut `Contp` (Conflict type) únicament pot prendre 4 valors, mentre que en la base de dades descarrega, pot prendre un cinquè amb valor "Inter-group", amb el que caldria actualitzar el CodeBook indicant el significat d'aquest valor.

Què hi heu descobert: evidències, tendències, outlayers, etc.

Les evidències, tendències i outlayers que s'han trobat amb les anàlisis que s'han realitzat són les següents:

- El període temporal que abraça la base de dades és suficientment ampli (dades recollides des de l'any 1990 al 2019) com per tenir en compte que un estudi de les dades agregat en els anys, pot no pot recollir la important variabilitat que hi pot haver en aquest període i, per tant, el factor temporal és important tenir-lo en

compte. Per a visualitzar aquest aspecte, únicament cal comparar la figura 1 de la visualització adjunta, en la que es mostra un diagrama de bombolles on la mida indica el nombre d'acords de pau signats, agrupats i acolorits en funció de la regió entre el 1990 i el 2019, amb la figura 2 on es mostra la mateixa visualització però pels anys 1990 i 2018, respectivament. Es pot veure que, més enllà de la variació dels països concrets, les regions del món amb més acords de pau també han variat significativament entre els anys, observant una tendència que passa de la regió d'Amèrica amb alta freqüència al 1990 a la regió de l'Orient mitjà-Àfrica del nord amb alta freqüència al 2018. Per tant, això indica que, en aquest cas, cal introduir una anàlisi temporal, per tal d'aconseguir recollir aquesta variabilitat temporal i, possiblement s'observaria com a l'època dels conflictes dels Balcans sorgiria Europa com una regió amb alta freqüència.

- Tot i la gran riquesa del conjunt de dades PA-X, afegir dades externes aconseguix enriquir el dataset i poder mostrar relacions que, fins llavors no es coneixien. En aquest sentit, en la figura 3 de la visualització adjunta, s'ha incorporat la informació del rànquing del Human Development Index per cada país (es pot trobar més informació a la [wikipedia](https://en.wikipedia.org/wiki/Human_Development_Index), remarcar que millor com més baixa la posició en el rànquing), en l'eix de les X, mostrant-se juntament amb el nombre d'acords signats, en l'eix de les Y, alhora, agrupats per regions i es pot observar clarament la tendència en la que els països amb major freqüència en els acords de pau, amb major mida de bombolla, tenen, habitualment, un rànquing alt d'HDI dins de la regió a la que pertanyen, excepte la regió d'Amèrica que la conflictivitat de Colòmbia, amb un HDI mig dins la regió, fa que surti d'aquesta regla.
- Hi ha atributs com els que recullen les referències en els tractats signats a determinats grups (variables GCh, GDis, GAge, GMig, GRa, GRe, GInd, GOth, GRef i GSoc) que poden assolir quatre valors en funció de si no hi ha referència, si únicament hi ha una menció retòrica, si hi ha una clàusula clarament detallada o si, fins i tot, detalla implementacions en aquests àmbits. S'ha considerat que únicament s'havia de tenir en compte les referències que, com a mínim tinguessin una clàusula detallada. Igualment, per avaluar aquestes referències, s'ha fet comparant-les amb el total d'acords per cada any, obtenint, així, el percentatge d'acords que recullen aquestes referències per cada any, tal com es pot observar en la figura 4 de la visualització. Respecte als resultats obtinguts, comentar que, s'observa una baixa proporció d'acords amb aquestes referències i tot i que tampoc s'observa una tendència en aquests valors, més enllà de mantenir-se en el temps en els valors baixos comentats. Indicar que si s'analitza la distribució del percentatge anual d'aquestes referències apareixen alguns anys que es podrien considerar com outlayers des d'aquest punt de vista. Específicament, caldria remarcar que la referència a treballadors migrants (atribut GMig) és realment baixa, fins i tot, en alguns anys nul·la, fet que si s'analitza la seva distribució, moltes dades anuals surtin com outlayers, tot i que el ràtio anual continuï essent realment baix.
- Unes altres referències en els acords de pau signats que s'han estudiat són les de gènere (variables GeWom, GeMe, GeSo i GeFa) i que es poden veure en la figura 5. De les quatre variables, les que fan referència a les dones sembla que té

una tendència a incrementar-se en el temps, tot i que en els darrers dos anys ha disminuït força. Igualment, s'observa molt poques referències a l'orientació sexual.

- Finalment, la darrera variable que s'analitza és l'evolució de la tipologia de conflicte en els acords, que es pot observar en la figura 6. Comentar que la major part dels conflictes són de tipologia de govern o govern/territori, tot i que amb una clara tendència a disminuir i, per contra, a pujar els conflictes de tipus entre-grups (en l'anterior resposta ja s'ha comentat que aquesta tipologia de conflicte no està especificada en el CodeBook).

Explicar com ho heu descobert: amb quines eines, amb quines operacions.

La metodologia seguida per fer l'anàlisi de l'apartat anterior ha estat la següent:

1. Realitzar amb un jupyter notebook emprant el llenguatge Python3 totes les tasques prèvies de càrrega de dades, neteja i obtenció dels fitxers de dades que posteriorment seran utilitzades per les eines de visualització (fitxers amb format csv).
2. Les eines de visualització emprades han estat, d'una banda, <https://rawgraphs.io/> per realitzar totes les visualitzacions de les figures 1 a 6 i, de l'altra, el programari Tableau Desktop per a obtenir el Dashboard que es mostra en la figura 7.
3. Tota la informació i dades emprades/obtingudes s'ha penjat al repositori de GitHub: <https://github.com/Ilergeta/PAX-Viz>. Amb la següent estructura: dins del directori "data" els diferents fitxers emprats com a font d'informació, en el directori "src" el notebook amb el codi emprat per la realització del punt 1 anteriorment detallat, en el directori "output" els fitxers de sortida després de processar el codi del notebook, en el directori "docs", aquest document en format .pdf i, finalment, en el fitxer "index.html" hi ha la pàgina web que carrega la url <https://ilergeta.github.io/PAX-Viz/> gràcies a GitHub Pages.

El Dashboard o un conjunt de visualitzacions sobre els datasets escollit.

A part de les visualitzacions comentades en els apartats anteriors, en la figura 7 de l'enllaç de visualitzacions <https://ilergeta.github.io/PAX-Viz/> es mostra un primer esborrany de Dashboard, en el que s'ha representat, a la meitat superior, un mapa dels països amb més freqüència d'acords de pau signats en el període (1990-2019) i en la part inferior esquerra es pot veure un gràfic de bombolles amb la distribució dels diferents països segons el seu rànquing en el HDI i en el nombre d'acords de pau signats, finalment, a la part inferior dreta es mostra l'evolució de la tipologia de conflicte, comentar que s'ha definit interacció entre el mapa i el gràfic de bombolles.

Per acabar, crec és important indicar cap a on es vol portar aquest projecte de visualització que acabarà en l'entrega de la PAC4. L'objectiu que em fixo és poder integrar en un Dashboard de Tableau tres representacions: una primera, un mapa com el que s'ha mostrat en la figura 7 però on també es pugui estudiar l'evolució temporal i es pugui seleccionar entre dades per països o agregades en regions, una segona en la que es pugui observar quin impacte tenen els acords de pau signats en diferents àmbits com l'economia, la salut, la qualitat de vida o l'educació de determinats països que es considerin singulars, on s'indicaria amb un gràfic de línies l'evolució de l'indicador segons l'àmbit triat junt amb els valors mitjans dels països de la regió i s'indicaria les dates dels tractats signats i una tercera en la que es mostrin indicadors com la tipologia de conflictes a mode de Race Bar Chart (veure [exemple](#)).