

CREACIÓ DE LA VISUALITZACIÓ

I LLIURAMENT DEL PROJECTE

## “LIFE CYCLE OF PEACE PROCESSES. Visualizing the Peace Agreement Database (PA-X)”



# Índex de continguts

<b>TÍTOL DE LA VISUALITZACIÓ / URL / DESCRIPCIÓ .....</b>	<b>3</b>
Títol de la visualització .....	3
URL de la visualització:.....	3
Descripció del document: .....	3
<b>QUE COMPLEIX I QUÈ NO RESPECTE EL PROJECTE INICIAL (DASHBOARD).</b>	
<b>VALORACIÓ DEL RESULTAT .....</b>	<b>3</b>
Visualització realitzada.....	3
Valoració resultat.....	5
<b>DESCRIPCIÓ TÈCNICA DEL PROJECTE .....</b>	<b>5</b>
Arquitectura de la solució tècnica .....	6
<b>VISUALITZACIÓ REALITZADA .....</b>	<b>8</b>
Objectiu de la visualització.....	8
Nivell de compliment amb la definició del projecte de visualització.....	8
Quines variables s'han seleccionat i perquè?.....	9
Quina visualització s'ha escollit i perquè? .....	11
Resultats obtinguts .....	12
<b>Estructura del ZIP .....</b>	<b>13</b>
<b>BLOC BIBLIOGRAFIA .....</b>	<b>14</b>
Webgrafia .....	14
Bibliografia .....	14
<b>ANNEXOS.....</b>	<b>15</b>
ANNEX 1: Codi Python per la generació de fitxers JSON: fitxer <b>jsonGroups_SingleFile.py</b> .....	15
ANNEX 2: Codi Python per la generació de fitxers JSON: fitxer <b>jsonGroups_MultipleFile.py</b> .....	17

## TÍTOL DE LA VISUALITZACIÓ / URL / DESCRIPCIÓ

Títol de la visualització:

“LIFE CYCLE OF PEACE PROCESSES.

Visualizing the Peace Agreement Database (PA-X)”

URL de la visualització:

<https://public.tableau.com/profile/quim4893#!/vizhome/LifeCycleofPeaceProcesses/Dashboard?publish=yes>

Descripció del document:

Aquest document té per objectiu descriure el procés de creació de la visualització escollida i el lliurament del seu projecte. Pretén recollir en essència quin ha estat el resultat de la visualització realitzada envers l'estudi de la base de dades (PAC2) i la seva planificació i definició (PAC3). Inclou valoracions sobre què s'ha implementat finalment i perquè, així com la descripció de la tecnologia emprada en quan a eines de programari i tècniques de visualització. La part central del document es centra en la pròpia visualització on queda descrit el seu objectiu, com complementa a la seva definició, quines variables selecciona respecte de les existents en la base de dades PA-X, i com han estat transformades les dades per visualitzar-les (sempre tot orientat al resultat final obtingut).

## QUE COMPLEIX I QUÈ NO RESPECTE EL PROJECTE INICIAL (DASHBOARD). VALORACIÓ DEL RESULTAT

Visualització realitzada

Respecte el projecte de definició:

COMPLEIX:

1. Assolir l'objectiu del projecte de visualització, que consisteix en modelitzar els processos de pau de la PA-X. Establir una estructura simplificada de les seves etapes i visualitzar-la de manera entenedora per la major part possible d'usuaris, mitjançant una eina interactiva.
2. L'usuari de la visualització pot identificar d'entre els processos de pau existents, quins tenen major nombre d'acords de pau (Top 10 list)<sup>(9)</sup>.
3. L'usuari disposa d'una llista de **processos de pau** identificant fàcilment quines de les 5 etapes disposa, així com el seu nom. Pot seleccionar-los mitjançant aquest llistat, o bé indicant la seva localització geogràfica en un mapa. En aquest objectiu de modelitzar els processos de pau l'usuari

pot consultar la magnitud o importància de cada fase (visualització estàtica amb rànkings per regió).

4. Respecte la planificació s'ha afegit una visualització dels acords que compren cada procés de pau i la seva classificació per regió i tipus de conflicte així com per etapa de procés de pau.
5. També s'han afegit dos visualitzacions estàtiques que permeten conèixer el ventall de classificacions per les que poden modelitzar els processos de pau, fen més comprensibles els gràfics de barres visualitzats posteriorment.
6. Tal com es va presentar en la definició del projecte, queden definides una àrea d'introducció i descripció, una àrea de selecció i una àrea d'ajuda, tot i que aquesta darrera s'ha reassignat per aprofitar al màxim l'espai disponible.
7. En l'àrea de resultats l'usuari pot visualitzar tan sols un conjunt de les característiques possibles, però la selecció realitzada, és ben significativa. En els gràfics es mostra la importància de cada variable en les diferents etapes d'un procés de pau. S'ofereixen tal com es va planificar gràfiques de barres globals i una gràfica de tendència al llarg de tots els anys.
8. Altres dos objectius del projecte són:
  - a. Incloure un bon nivell d'interactivitat a la visualització. Això ho aconseguim amb les seleccions de processos de pau i amb el gràfic de tendència que permet variar el període en anys analitzat.
  - b. Permetre fer zoom sobre les dades. Amb les seleccions els gràfics són de detall i al fer un clic de ratolí sobre el mar del component de mapa, o bé un clic en zones de no dades tornem a visualitzar un resum global (a vegades es bloquegen les seleccions i aquestes són vies possibles per desbloquejar-les, tal com s'indica en la zona d'ajuda).
9. Finalment també s'ha inclòs un comparador de processos de pau, perquè l'usuari pugui comparar els processos de pau segons la importància de cada grup de població en els acords de pau. Es mostren fluxos de dades per períodes anuals.
10. Aquest comparador implementa diagrames de tipus Sankey mitjançant l'ús de D3.js. Els diagrames de Sankey, són molt útils a l'hora de representar fluxos de dades i poder visualitzar gràficament l'evolució de la importància de cada grup de població.
11. També es compleix amb una idea inicial de no tenir que escollir entre Tableau i D3.js sinó aprofitar les dues tecnologies per obtenir els millors resultats.
12. També s'ha aconseguit seguir les línies de bloc de dades i del bloc de disseny marcades inicialment.

#### NO COMPLEIX

1. No s'ha aconseguit dotar al comparador de la pluralitat d'operacions pensada inicialment.
2. Tampoc millorar l'àrea de resultats. Es presenten dades en gràfics de barres però la intenció era combinar-los amb gràfics de tipus "bullet bars" o de barres però més elaborats, però que al final només s'han considerat els inicialment proposats per representar un valor de mitjana.
3. Tots els aspectes que no s'han complert en la visualització final tenen que veure amb l'ambició de la planificació. Potser s'hauria d'haver afinat més en el cost tecnològic de no haver treballat mai amb les eines de Tableau i D3.js on la corba d'aprenentatge en els inicis, és considerablement més alta.

## Valoració resultat

La valoració del resultat és positiva. El rati temps invertit i resultat té un dèficit temporal considerable motivat per el desconeixement de com s'ha d'enfocar un projecte de visualització, i de no tenir a l'inici una idea clara de com procedir. La valoració però és positiva perquè la realització d'aquest projecte ens fa ser optimistes per el següent, que de ben segur serà molt millor.

L'esforç ha estat major del previst, perquè la voluntat de conjuntar dues tecnologies, comporta moltes dificultats, ja que estan dissenyades per treballar de manera independent i no sempre usen estratègies o filosofies compatibles. En aquest cas com la major part de la feina estava desenvolupada en Tableau s'ha optat per escollir aquesta plataforma com a base, i incloure les personalitzacions en D3.js.

Un fet molt important en el desenvolupament del projecte ha estat treballar molt la primera PAC. Ajuda molt a veure, que no sempre es fàcil aconseguir aplicar la teoria i força fàcil caure en els errors indicats.

## DESCRIPCIÓ TÈCNICA DEL PROJECTE

Per desenvolupar aquest projecte de visualització s'han combinat diferents eines de desenvolupament. Des de el punt de vista de visualització, el programari de Tableau (Tableau Desktop Professional Edition versió 2019.3.3 64bits) i la llibreria gràfica D3 (Data-Driven Documents or D3.js).

### Tableau 2019<sup>(9)</sup>:

- Per definir el Dashboard de la visualització i les interaccions entre els seus components.
- Aporta una plataforma de hosting públic de les visualitzacions realitzades.
- Facilita la comunicació d'events entre els components del Dashboard.
- Ajuda al preprocessament del dataset original.

### D3 (D3.js)<sup>(1,2,4,5,6,10,11)</sup>:

- Consisteix en una biblioteca JavaScript (JS) per visualitzar dades mitjançant estàndards web. D3 ens ajuda a visualitzar dinàmicament les dades mitjançant les tecnologies SVG, Canvas i HTML. D3 combina potents tècniques de visualització i interacció amb un enfocament basat en dades sobre la manipulació de DOM, que ens proporciona les capacitats completes dels navegadors moderns i la llibertat de dissenyar la interfície visual adequada per a les vostres dades.

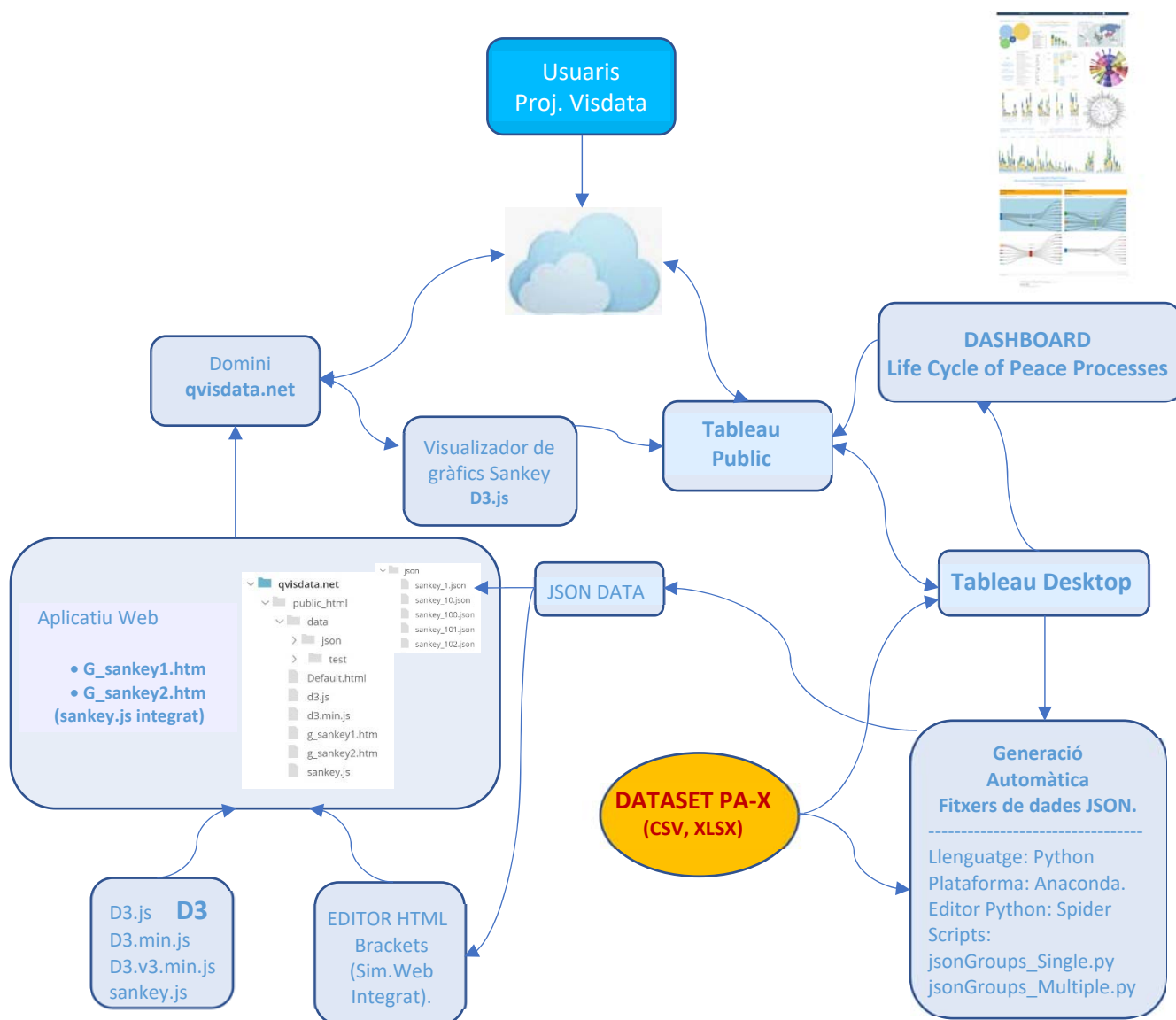
### Brackets:

- Editor HTML amb simulació de servidor Web. Eina orientada al desenvolupament en HTML, CSS, JS. També dona suport al format **JSON**.

### Anaconda Navigator:

- Plataforma integrada de desenvolupament. Conté la plataforma de desenvolupament en Python Spider. Spider s'ha utilitzat per la generació automàtica dels fitxers JSON per poder visualitzar els diagrames de flux de tipus Sankey, usats en el comparador de processos de pau. S'han generat 151 fitxers de dades en format JSON usant dos scripts programats en Python: **jsonGroups\_Single.py** i **jsonGroups\_Multiple.py**.

## Arquitectura de la solució tècnica



**Figura N°1:** Arquitectura del sistema de desenvolupament utilitzat.

En la figura n°1 es pot observar el sistema de desenvolupament global utilitzat i totes les seves connexions. Un cop descarregat de la web de PA-X el dataset de dades, aquest ja es pot treballar amb el full de càlcul Excel i amb el programa de gestió de dades de Tableau (Tableau Prep Builder). Les dades ja havien estat treballades en fases anteriors, per tant per la visualització han estat transformades, però per ser visualitzades en D3. Per això ha fet falta generar els fitxers en JSON específics per la visualització de diagrames de flux de tipus Sankey.

En la següent figura veiem un exemple corresponent al procés de pau nº 1 “Afghanistan 1990s Taliban process”:

```
{
  "nodes": [
    {"node": 0, "name": "1992"},
    {"node": 1, "name": "1993"},
    {"node": 2, "name": "Valor 0"},
    {"node": 3, "name": "Valor 1"},
    {"node": 4, "name": "Valor 2"},
    {"node": 5, "name": "Valor 3"},
    {"node": 6, "name": "Children/Youth"},
    {"node": 7, "name": "Disabled persons"},
    {"node": 8, "name": "Elderly/Age"},
    {"node": 9, "name": "Indigenous people"},
    {"node": 10, "name": "Migrant workers"},
    {"node": 11, "name": "Racial/Ethnic/National groups"},
    {"node": 12, "name": "Refugees/Displaced persons"},
    {"node": 13, "name": "Religious group"},
    {"node": 14, "name": "Social class"},
    {"node": 15, "name": "Other"}
  ],
  "links": [
    {"source": 0, "target": 2, "value": 10},
    {"source": 1, "target": 2, "value": 9},
    {"source": 1, "target": 3, "value": 1},
    {"source": 2, "target": 6, "value": 2},
    {"source": 2, "target": 7, "value": 2},
    {"source": 2, "target": 8, "value": 2},
    {"source": 2, "target": 9, "value": 2},
    {"source": 2, "target": 10, "value": 2},
    {"source": 2, "target": 11, "value": 2},
    {"source": 2, "target": 12, "value": 1},
    {"source": 3, "target": 12, "value": 1},
    {"source": 2, "target": 13, "value": 2},
    {"source": 2, "target": 14, "value": 2},
    {"source": 2, "target": 15, "value": 2}
  ]
}
```

**Figura N°2:** Estructura de dades per el sankey diagram **sankey\_1.json**.

Totes les transformacions a JSON, han estat emmagatzemades a internet en el domini **qvisdata.net**, on queden allotjades les pàgines web que implementen el comparador de processos usant D3.

A la documentació del projecte també s’afegeixen aquest dos scripts en llenguatge Python.

Per visualitzar el comparador implementat en aquest domini s’han utilitzat dos elements de dashboard de tipus “page web objects”. La problemàtica de compatibilitzar dues plataformes ha estat palesa en quan a que aquest objectes accepten en local, visualitzar el comparador però en el moment en que el dashboard es pujat a Tableau Public, no funcionen al imposar-se una restricció de dominis segurs amb protocol https:// de manera obligada.

## VISUALITZACIÓ REALITZADA

### Objectiu de la visualització

L'objectiu de la visualització es pot obtenir del seu títol. "Life Cycle of Peace Processes, Visualizing the Peace Agreement Database PA-X". Volem modelitzar i descriure els processos de pau. Permetre que l'usuari els pugui conèixer com a entitat de la base de dades PA-X. La entitat fonamental de la PA-X són els acords de pau ("agreements"), que són els elements atòmics que conformen els processos de pau. La visualització té per objectiu descriure i caracteritzar els processos de pau, des de que comencen fase ("Starting"), fins que acaven amb un alto el foc ("Ceasefire"). En la PA-X hi ha multitud de possibilitats d'estats o etapes mentre el procés de pau no finalitza. En la visualització es simplifiquen en 4 etapes més l'estat "Unknown" que recull els processos sense informació. Amb la visualització l'usuari pot saber fàcilment en quin estat es troba cada procés de pau, i veure com es comporten 5 característiques escollides per la seva importància: Definició d'estat ("State Definition"), Incidència de cada Grup de població ("Groups"), Drets humans (Human Rights), Gènere (Gender) i Govern ("Governance"). Les gràfiques ofereixen dades del percentatge d'importància de cada subcategoria en cada fase o etapa d'un procés de pau.

Per modelitzar els processos de pau i respondre a la pregunta inherent de com evoluciona un procés de pau i per quines etapes passa la visualització ofereix les gràfiques de barres globals més un gràfic de tendència en el temps. En aquest gràfic, és on podem comparar en el temps les subcategories de la categoria drets humans ("Human rights") per etapes. A més a més, en aquest gràfic, podem variar el període de temps en que volem saber la tendència. En la part final es deixa a l'usuari de la visualització analitzar específicament la categoria **GROUPS** al llarg dels anys. Aquest anàlisi és purament visual observant com es comporten els fluxos de cada nivell d'importància en cadascuna de les subcategories de la categoria GROUPS.

L'usuari disposa d'un punt central d'informació que dona indicacions bàsiques d'ús i dues visualitzacions que li defineix de manera global la ingesta quantitat de categories i subcategories disponibles a la base de dades PA-X, a mode de complement.

Tableau Public tot i no ser la plataforma més ràpida de Tableau, dona les prestacions suficients. Potser les mancances més evidents es demostren en la comunicació d'events entre components, que a vegades no és molt correcte.

### Nivell de compliment amb la definició del projecte de visualització

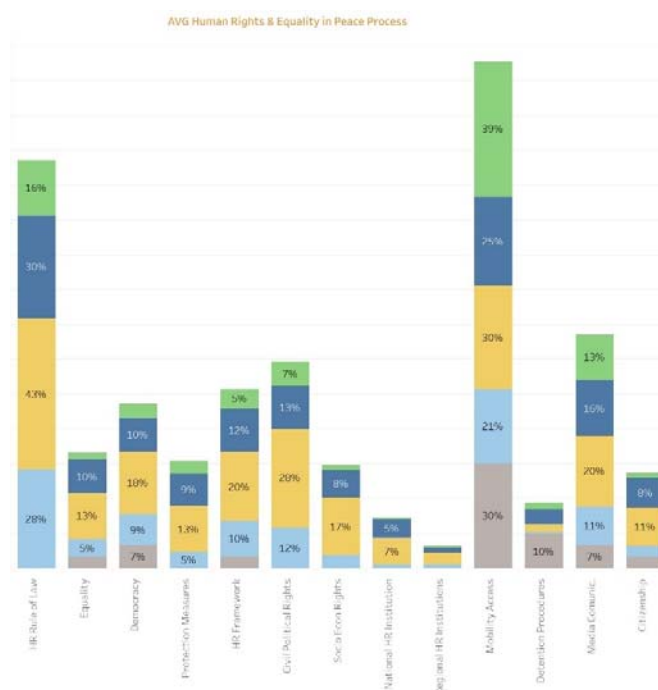
Com ja hem comentat, en el segon apartat d'aquest document ("Què compleix i que no compleix"), el nivell de compliment es molt alt excepte en l'apartat del comparador de processos, on l'expectativa d'implementació del comparador era més alta i finalment s'ha reduït a comparar els processos de pau per una sola característica GROUPS. Tot i això, podem comparar 151 processos de pau, per parelles i conservant un històric dels gràfics que podem netejar en qualsevol moment. En quan a metodologia de procediment queda molt clar com es podrien implementar les següents categories i varietat de gràfics.



## Quines variables s'han seleccionat i perquè?

Les variables escollides per l'anàlisi del percentatge d'importància de cada fase del cicle de vida d'un procés de pau en els gràfics de barres han estat:

- Definició d'estat ("State Definition")
- Incidència de cada Grup de població ("Groups")
- Drets humans (Human Rights)
- Gènere (Gender)
- Govern ("Governance")



**Figura Nº3:** Exemple de gràfics de barres apilades per el cas de drets humans en el procés de pau de Croàcia.

Per el gràfic de tendència, específicament s'ha escollit la categoria de drets humans, per ser una categoria molt important i perquè conté moltes subcategories que poden oferir molta informació en la visualització.

### Agreement Time line for Human Rights content in different categories

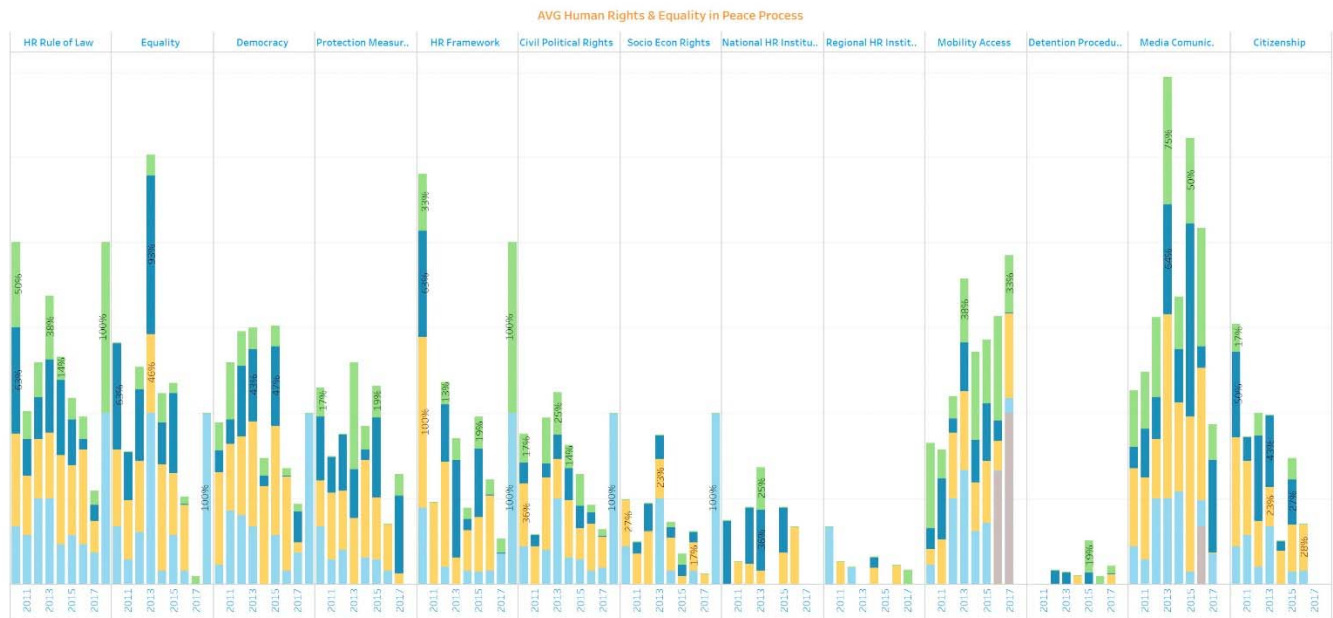
Signed Date (agreement)

1/1/2010



31/1/2018

We can observe the level of significance of each category inside every agreement registered. At difference from the previous chart now we can modify the time line using the "SignedDate" filter control to get a interval to analyze the tendency between the first and second dates.



**Figura N°4:** Exemple de gràfic de barres de tendència per les subcategories de la categoria de drets humans.

En el comparador de processos s'ha escollit específicament la categoria GROUPS i les seves subcategories. No existeix una raó primordial en la seva selecció ja que totes les categories i subcategories són de gran interès, però hem fet una selecció segons el propi criteri d'interès.

### Visual Comparator of Peace Processes

(We can study Groups characteristics using Flowing data with Sankey diagrams)

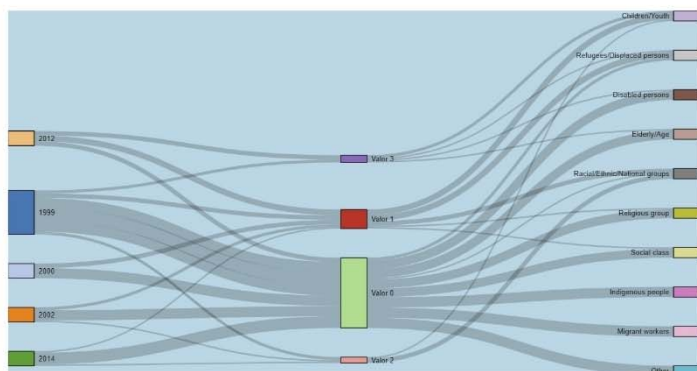
Visualize and compare GROUPS category in pairs and maintaining a history of visualizations.  
To start a new series press the CLEAN button.

#### List of Peace Processes

PROCESS 1:

Sudan: Local Peace Processes

CLEAN

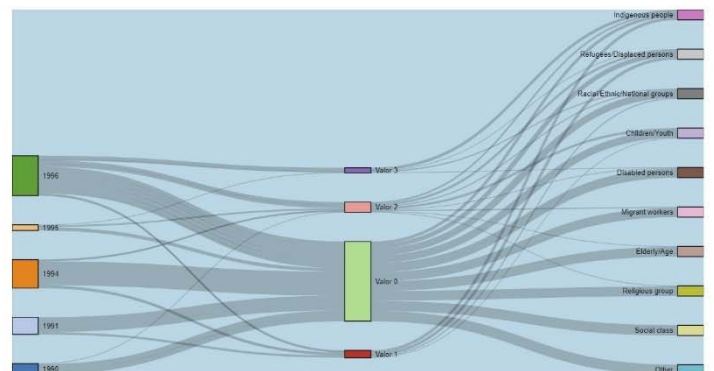


#### List of Peace Processes

PROCESS 2:

Guatemala peace process

CLEAN



**Figura N°5:** Imatge del comparador de processos de pau per fluxos de dades anuals de cada subcategoria de la categoria GROUPS .

## Quina visualització s'ha escollit i perquè?

Hem escollit un projecte de visualització que contempli interactivitat i participació de l'usuari. Amb l'actualització dinàmica dels components de la visualització creiem que es manté l'atenció de l'usuari i se'l convida a participar en la següent secció de la visualització.

S'ha intentat concentrar en la visualització, tot allò que cal que l'usuari tingui coneixement, (tenint en compte que amb tanta informació disponible, no es pot col·locar tot). S'han combinat capacitat de síntesi i capacitat estètica.

Entrant en detall en cadascuna de les visualitzacions incloses en el projecte de visualització, hem implementat cada visualització de menys a més complexitat a mesura que avancem usant un nivell de **comunicació efectiva** que segueix un patró d'esquerra a dreta i de dalt baix.

El primer component de visualització [Figura 6 (1)], és un gràfic de bombolles, que inicialment dona una informació global del nombre d'etapes de cada tipus en el dataset. Ens ha servit perfectament, perquè visualment l'usuari només per magnitud i color identifiqui la fase dominant en el cicle de vida del procés de pau seleccionat (o per defecte en el conjunt global de processos). Els colors utilitzats en el gràfic ja han estat introduïts i identificats en la primera línia de text de la visualització on presenten les fases definides i els seus colors corresponents.

Els següents dos components [Figura 6 (2)], són visualitzacions estàtiques i permeten a l'usuari comparar si el **procés de pau** seleccionat pertany als 10 processos de pau amb més acords de pau o bé la situació en el rànking per regió. En els dos cassos motivem a l'usuari a provar de seleccionar algun d'aquests cassos destacats, i per tant a continuar avançant.

El component de mapa, [Figura 6 (3)], és potser el component de menys interès perquè és molt conegut i utilitzat però ens permet captar l'atenció de l'usuari precisament per la seva popularitat i el domini en el seu ús. L'utilitzem per seleccionar però també per oferir una classificació dels tipus d'acords de pau existents dintre dels processos de pau.

La llista de processos de pau, [Figura 6 (4)], és un element essencial que ens permet oferir la informació clau de la visualització, quins processos de pau tenim i de quantes etapes és composen, es a dir el seu cicle de vida actual. Si volem detall de la composició de cada procés de pau ja podem consultar el següent component, [Figura 6 (5)], situat a la dreta. Aquest component, s'ha dissenyat per representar tots els acords de pau que pertanyen a un procés de pau. A més a més contempla les fases per les quals han passat o es troben donada una selecció en la llista. Per defecte mostra tots els acords de pau). Aquest tipus de visualització, ens permet visualitzar molta informació amb poc espai, utilitzant matrius de símbols amb color.

En el nostre cas també a s'ha tingut en compte el poder posar en pràctica la generació del major nombre de gràfics de visualització per adquirir l'experiència de treballar-los. És el cas de les dues visualitzacions de categories i subcategories incloses a la visualització [Figura 6 (6)]. Per una banda serveix perquè l'usuari pugui adquirir coneixement i per altre ens permet comprovar quins dels dos pot resultar més efectiu, ja sigui per la incorporació dels blocs de color, com per la seva absència i nitidesa.

Els components [Figura 6 (7,8)], ja els hem comentat anteriorment en detall.

## Resultats obtinguts

En la següent figura és mostra un imatge completa de tota la visualització.

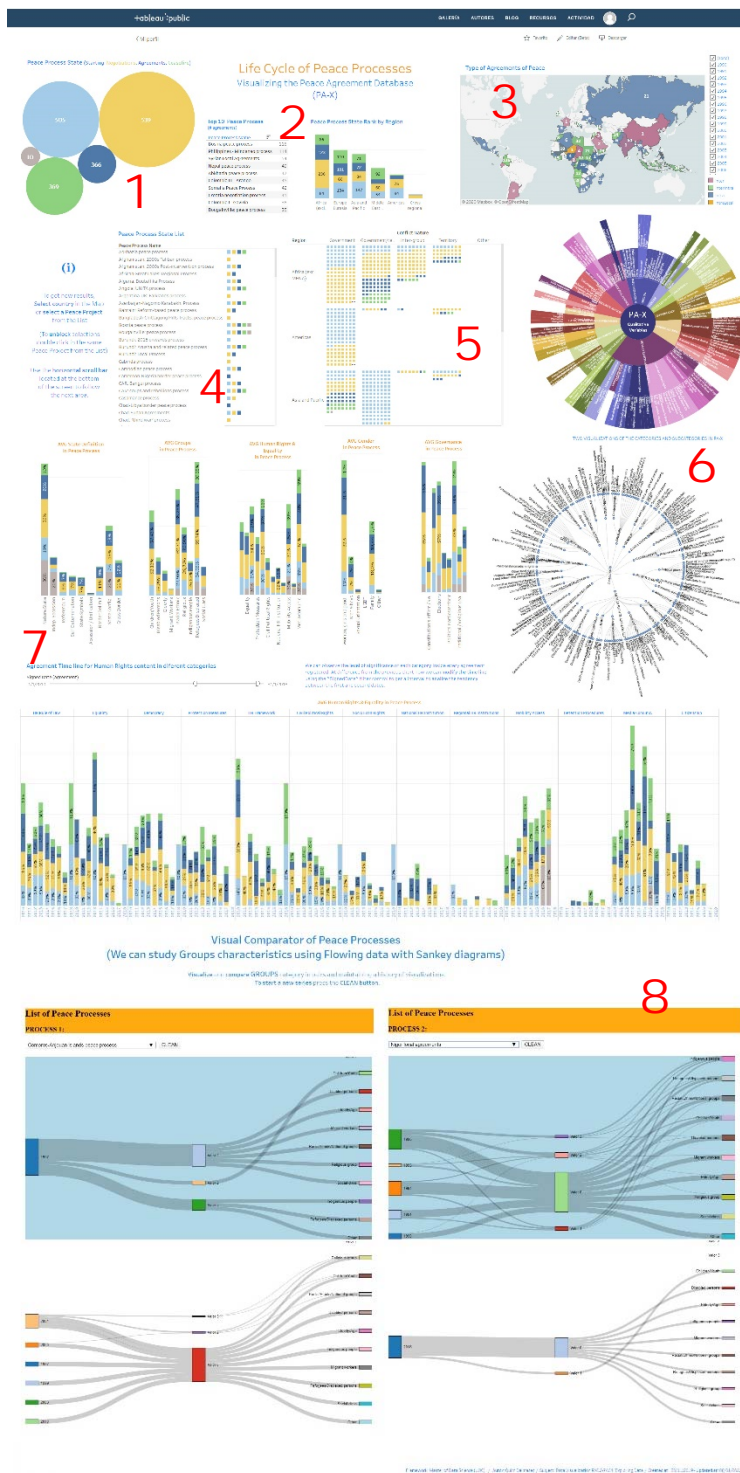
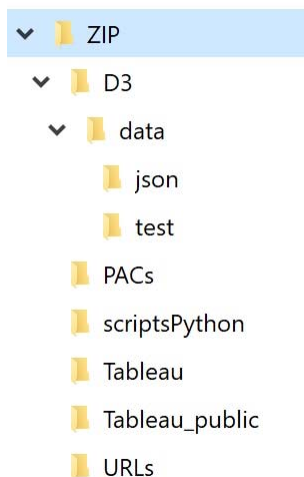


Figura Nº6: Composició i estructura completa de la visualització realitzada.

## Estructura del ZIP

L'estructura de carpetes del ZIP entregat es mostra en la figura següent:



**Figura N°7:** Estructura de directoris del ZIP entregat.

### Carpeta **D3**

Tot el referent al codi, llibreries i html programat per el comparador de processos de pau allotjat en el domini **qvisdata.net**.

Les URL per provar de fer els gràfics sankey de manera directa es troben en la carpeta URLs.

### Carpeta **D3/data/json**

Fitxers de dades en format JSON, creats amb el codi Python per generar els gràfics sankey de cada procés de pau per la categoria GROUPS. En total 151 més algun de test.

### Carpeta **D3/data/test**

Fitxers per generar el gràfic de tipus 'sunburst' (combinat amb el fitxer t7b.htm).

### Carpeta **PACs**

Conté els fitxers de cadascuna de les PACS entregades

### Carpeta **scriptsPython**

Conté els scripts de generació automàtica de fitxers JSON de dades (inclosos en els annexos).

### Carpetes **Tableau** i **Tableau\_public**

Contenen els fitxers de projecte en Tableau. Cal llegir els txt inclosos per utilitzar-los perquè canvien les rutes de les fonts de dades.

### Carpeta **URLs**

Conté en un fitxer text les URL a la visualització final en Tableau públic i al domini **qvisdata.net**.

## BLOC BIBLIOGRAFIA

### Webgrafia

La webgrafia consultada ha estat:

1. **DataViz Gallery: opcions per la visualització de dades en R,Python,D3.js:**  
<https://www.d3-graph-gallery.com/all.html>  
<https://www.data-to-viz.com/>
2. **D3.js GitHub**  
<https://github.com/d3/d3/wiki/Gallery>
3. **Visualizing Data: A Handbook for a Data Drive**  
<http://book.visualisingdata.com/>
4. **DataViz Gallery: opcions per la visualització de dades en R,Python**  
<http://research.nualart.cat>  
<http://research.nualart.cat/peace/cfgal.html#simple>  
<http://research.nualart.cat/peace/>  
<http://research.nualart.cat/area-peaceagreements/>
5. **Web Observable : Jupyter notebooks**  
<https://observablehq.com/>

### Bibliografia

#### Teoria

6. **Andy Kirk** (2012), *Data Visualization: a successful design process* Ed Packt Publishing.
7. **The Encyclopedia of Human-Computer Interaction, 2nd Ed**  
**Stephen Few** (2013), "Data Visualization for Human Perception".  
**Stephen Few** (2007), "DATA VISUALIZATION PAST, PRESENT, AND FUTURE - PERCEPTUAL EDGE".  
**Stephen Few** (2006), "Visual\_Communication- PERCEPTUAL EDGE".
8. **Robert Forster and Christine Bell** (2019), "PA-X Spotlight –Gender - Gender Mainstreaming in Ceasefires: Comparative Data and Examples", Ed UN Women.

#### Tableau

9. **Zen Master** (2018), "Practical Tableau\_ 100 Tips, Tutorials, and Strategies from a Tableau.", Ed O'Really

#### D3.js

10. **Swizec Teller** (2015), "React+d3.js Build data visualizations with React and d3.js". Ed. Lean Publishing.
11. **Michael Heydt** (2015), "D3.js By Example\_ Create attractive web-based data visualizations using the amazing JavaScript library D3.js", Ed. Packt Publishing.

## ANNEXOS

### ANNEX 1: Codi Python per la generació de fitxers JSON: fitxer jsonGroups\_SingleFile.py

```
# -*- coding: utf-8 -*-
"""
Created on Mon Jan 15:12:50 2020

@author: Quim Dalmases
"""

import pandas as pd

idProcesPau=3
maxValue=3
comptador = [0]*(maxValue+1)
groups = {"G Ch":"Children/Youth", "G Dis":"Disabled persons", "G Age":"Elderly/Age", "G Ind":"Indigenous people", \
          "G Mig":"Migrant workers", "G Ra":"Racial/Ethnic/National groups", "G Ref":"Refugees/Displaced persons", \
          "G Re":"Religious group", "G Soc":"Social class", "G Oth":"Other"}

colGroups = "G Ch", "G Dis", "G Age", "G Ind", "G Mig", "G Ra", "G Ref", "G Re", "G Soc", "G Oth"

nodes={}

qjson_head = '{\n"nodes":{\n'
qjson_end = "\n}"

qjson = qjson_head

#for an earlier version of Excel, you may need to use the file extension of 'xls'
df = pd.read_excel(r'.\sankey_Grups1.xlsx')

df['Year'] = df['Signed Date'].map(lambda x: x.year)

idProcesPau = 15
df1=df[df['Peace Process']==idProcesPau]
print(df1.loc[:,["Year", "G Ch", "G Dis", "G Age", "G Ind", "G Mig", "G Ra", "G Ref", "G Re", "G Soc", "G Oth"]])

#df2=df[df['Peace Process']==20].groupby('Year').agg({"G Ch":["sum"], "G Dis":["sum"], "G Age":["sum"], "G Ind":["sum"], "G Mig":["sum"], "G Ra":["sum"], "G Ref":["sum"], "G Re":["sum"], "G Soc":["sum"], "G Oth":["sum"]})
df2=df[df['Peace Process']==idProcesPau].groupby('Year').agg({"G Ch":["mean"], "G Dis":["mean"], "G Age":["mean"], "G Ind":["mean"], "G Mig":["mean"], "G Ra":["mean"], "G Ref":["mean"], "G Re":["mean"], "G Soc":["mean"], "G Oth":["mean"]})
print(df2)

years=[]
for (k,row) in df1.iterrows():
    if str(row['Year']) not in years:
        years.append(str(row['Year']))
years=sorted(years)

#print(years)

m=0
cadena=""
for i in range(len(years)):
    cadena = cadena + '{"node":' + str(i) + ', "name":"' + years[i] + '",\n'
    nodes[str(years[i])]=i
    m=i

#print(cadena)
```



```

m=m+1
for k in range(maxValue+1):
    cadena = cadena + '{"node":' + str(m) + ', "name": "Valor ' + str(k) + '"},\n'
    nodes['Valor ' + str(k)] = m
    m=m+1

for k in groups.values():
    cadena = cadena + '{"node":' + str(m) + ', "name": "' + str(k) + '"},\n'
    nodes[str(k)] = m
    m=m+1

cadena = cadena[0:len(cadena)-2] + '\n],\n"links": [\n'

qjson= qjson_head + cadena

cadena=""
for elem in years:
    df3 = df1.loc[df1['Year']==int(elem),:]
    for (k,row) in df3.iterrows():
        for c in colGroups:
            if row[c]==0:
                comptador[0]=comptador[0]+1
            elif (row[c]==1):
                comptador[1]=comptador[1]+1
            elif (row[c]==2):
                comptador[2]=comptador[2]+1
            elif (row[c]==3):
                comptador[3]=comptador[3]+1

        for i in range(maxValue+1):
            if comptador[i]!=0:
                cadena = cadena + '{"source":' + str(nodes[str(elem)]) + ', "target":' + str(nodes["Valor " + str(i)]) + ', "value":' + str(comptador[i]) + '},\n'
            comptador=[0,0,0,0]

qjson = qjson + cadena

cadena=""
cadena_niv2=""
for c in colGroups:
    for (k,row) in df1.iterrows():
        if row[c]==0:
            comptador[0]=comptador[0]+1
        elif (row[c]==1):
            comptador[1]=comptador[1]+1
        elif (row[c]==2):
            comptador[2]=comptador[2]+1
        elif (row[c]==3):
            comptador[3]=comptador[3]+1
    for x in range(maxValue+1):
        if comptador[x]!=0:
            cadena = '{"source":' + str(nodes["Valor " + str(x)]) + ', "target":' + str(nodes[groups[c]]) + ', "value":' + str(comptador[x]) + '},\n'
        cadena_niv2 = cadena_niv2 + cadena
    comptador=[0,0,0,0]

qjson = qjson + cadena_niv2[0:len(cadena_niv2)-2] + qjson_end
print(qjson)

file2write=open("./json/sankey_" + str(int(idProcesPau)) + ".json",'w')
file2write.write(qjson)
file2write.close()

```



## ANNEX 2: Codi Python per la generació de fitxers JSON: fitxer jsonGroups\_MultipleFile.py

```
# -*- coding: utf-8 -*-
"""
Created on Mon Jan 1 15:12:50 2020

@author: Quim Dalmases
"""

import pandas as pd

idProcesPau=3
maxValue=3
comptador = [0]*(maxValue+1)
groups = {"G Ch": "Children/Youth", "G Dis": "Disabled persons", "G Age": "Elderly/Age", "G Ind": "Indigenous people", \
          "G Mig": "Migrant workers", "G Ra": "Racial/Ethnic/National groups", "G Ref": "Refugees/Displaced persons", \
          "G Re": "Religious group", "G Soc": "Social class", "G Oth": "Other"}

colGroups = "G Ch", "G Dis", "G Age", "G Ind", "G Mig", "G Ra", "G Ref", "G Re", "G Soc", "G Oth"

nodes={}

qjson_head = '{"nodes":['
qjson_end = "']}"

qjson = qjson_head

#for an earlier version of Excel, you may need to use the file extension of 'xls'
df = pd.read_excel(r'.\sankey_Groups1.xlsx')

df['Year'] = df['Signed Date'].map(lambda x: x.year)

for (y,row1) in df.iterrows():
    if row1['Peace Process']!= 'nan':
        idProcesPau = row1['Peace Process']
        df1=df[df['Peace Process']==idProcesPau]
        print(df1.loc[:,["Year", "G Ch", "G Dis", "G Age", "G Ind", "G Mig", "G Ra", "G Ref", "G Re", "G Soc", "G Oth"]])

        #df2=df[df['Peace Process']==20].groupby('Year').agg({"G Ch":['sum'], "G Dis":['sum'], "G Age":['sum'], "G Ind":['sum'], "G Mig":['sum'], "G Ra":['sum'], "G
        Ref":['sum'], "G Re":['sum'], "G Soc":['sum'], "G Oth":['sum']})
        df2=df[df['Peace Process']==idProcesPau].groupby('Year').agg({"G Ch":['mean'], "G Dis":['mean'], "G Age":['mean'], "G Ind":['mean'], "G Mig":['mean'], "G
        Ra":['mean'], "G Ref":['mean'], "G Re":['mean'], "G Soc":['mean'], "G Oth":['mean']})
        print(df2)

        years=[]
        for (k,row) in df1.iterrows():
            if str(row['Year']) not in years:
                years.append(str(row['Year']))
        years=sorted(years)

        #print(years)

        m=0
        cadena=""
        for i in range(len(years)):
            cadena = cadena + '{"node":' + str(i) + ', "name":' + years[i] + '},\n'
            nodes[str(years[i])] = i
            m=i

        m=m+1
        for k in range(maxValue+1):
            cadena = cadena + '{"node":' + str(m) + ', "name": "Valor ' + str(k) + '"},\n'
            nodes['Valor ' + str(k)] = m
            m=m+1
```

```

for k in groups.values():
    cadena = cadena + '{"node":' + str(m) + ', "name":"' + str(k) + '",\n'
    nodes[str(k)] = m
    m=m+1

cadena = cadena[0:len(cadena)-2] + '\n],\n"links": [\n'

qjson= qjson_head + cadena

cadena=""
for elem in years:
    df3 = df1.loc[df1['Year']==int(elem),:]
    for (k,row) in df3.iterrows():
        for c in colGroups:
            if row[c]==0:
                comptador[0]=comptador[0]+1
            elif (row[c]==1):
                comptador[1]=comptador[1]+1
            elif (row[c]==2):
                comptador[2]=comptador[2]+1
            elif (row[c]==3):
                comptador[3]=comptador[3]+1

        for i in range(maxValue+1):
            if comptador[i]!=0:
                cadena = cadena + ' {"source":' + str(nodes[str(elem)]) + ', "target":' + str(nodes["Valor " + str(i)]) + ', "value":' + str(comptador[i]) + '},\n'
            comptador=[0,0,0,0]

qjson = qjson + cadena

cadena=""
cadena_niv2=""
for c in colGroups:
    for (k,row) in df1.iterrows():
        if row[c]==0:
            comptador[0]=comptador[0]+1
        elif (row[c]==1):
            comptador[1]=comptador[1]+1
        elif (row[c]==2):
            comptador[2]=comptador[2]+1
        elif (row[c]==3):
            comptador[3]=comptador[3]+1
    for x in range(maxValue+1):
        if comptador[x]!=0:
            cadena = ' {"source":' + str(nodes["Valor " + str(x)]) + ', "target":' + str(nodes[groups[c]]) + ', "value":' + str(comptador[x]) + '},\n'
            cadena_niv2 = cadena_niv2 + cadena
        comptador=[0,0,0,0]

qjson = qjson + cadena_niv2[0:len(cadena_niv2)-2] + qjson_end
print(qjson)

file2write=open("./json/sankey_" + str(int(idProcesPau)) + ".json",'w')
file2write.write(qjson)
file2write.close()

```