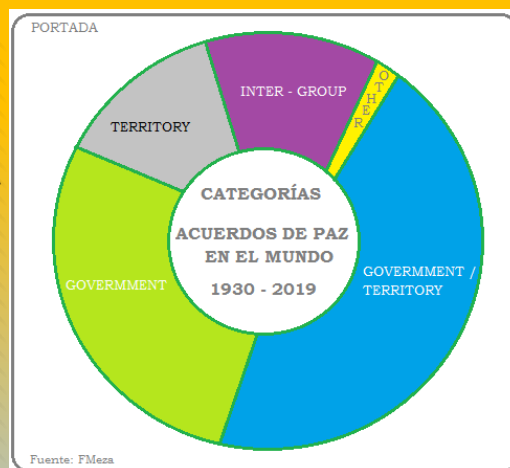


PEC4

Proyecto de Visualización de Datos

**Fernando Meza Ibarra****“ACUERDOS DE PAZ Y SUS
CATEGORÍAS, UNA VISIÓN MUNDIAL”****Mayo de 2020**<https://github.com/fmezaibarra/Visualizaci-n-acuerdos-paz-en-el-mundo>

Máster universitario en Ciencia de Datos

UOC

Contenido del proyecto de visualización de datos

• INTRODUCCIÓN.....	3
• TÍTULO DE LA VISUALIZACIÓN	3
• DESCRIPCIÓN	3
• URL DE LA VISUALIZACIÓN	4
• HIPOTESIS Y EXPLORACION INICIAL DATOS	4
• EL PROCESO DE VISUALIZACIÓN DE DATOS	17
• El diagrama de la Visualización de Datos.....	18
• Comunicar: Tipo de análisis y visualización	19
• Explorando los datos	20
• Datos en contexto	20
• Encontrar patrones y datos extremos	22
• Descripción del dataset	23
• Limpieza de los datos	26
• Relaciones.....	27
• Gráficos seleccionados	27
• Colores a utilizar.....	28
• Tipografías	29
• Plantillas o estructuras	29
• Bloques	29
• Formato	29
• EL PRODUCTO: LA VISUALIZACIÓN FINAL	29
• CONCLUSIONES Y EXPLICACIÓN.....	32
• CÓDIGO Y DOCUMENTACIÓN	33

• INTRODUCCIÓN

A través de la visualización de datos podemos representar de forma gráfica la información y los datos de los acuerdos de paz y acuerdos de entendimiento firmados por diferentes países a nivel mundial. Mediante el uso de elementos visuales, como gráficos y mapas, la visualización de datos ofrece una manera accesible para detectar y comprender las tendencias, los valores atípicos y los patrones en los datos.

En el mundo de los big data, las herramientas y las tecnologías de visualización de datos son esenciales para analizar cantidades masivas de información y tomar decisiones basadas en los datos.

El objetivo de éste proyecto persigue conocer, explorar, transformar los datos propuestos, para poderlos trasladar a visualizaciones utilizando la herramienta especializada TABLEAU, el fin principal es acceder a los datos de una forma rápida, concisa que nos permitan ubicar tendencias y evidencias que nos faciliten la comprensión de una problemática específica.

Con las visualizaciones creadas se persigue que académicos, mediadores, actores en conflicto y ciudadanos civiles que forman parte de la sociedad puedan tener acceso a dicha información de forma visual de una manera sobria, concisa y actualizada; la visualización permitirá navegar y comparar sistemáticamente por las diferentes categorías, muestra cómo se formalizan los procesos de paz y la transición de compromisos negociados en un intento de avanzar hacia la paz mundial.

• TÍTULO DE LA VISUALIZACIÓN

“ACUERDOS DE PAZ Y SUS CATEGORÍAS, UNA VISIÓN MUNDIAL”

• DESCRIPCIÓN

Partiendo de un conjunto de datos (dataset) que reúne de forma integral y consolidada los acuerdos de paz en todas las fases que se firman entre los países en conflicto, además de todos los procesos de paz en los conflictos intra-estatales e inter-estatales, desde 1930 hasta enero de 2020, se realizará una exploración visual que nos permita crear una representación visual interactiva.

Ésta representación involucra a los siguientes actores:

- ✓ **Acuerdo de paz:** Considera el documento formal que está disponible al público, producido después de la discusión con el conflicto protagonistas.
- ✓ **Los protagonistas del conflicto:** Actores estatales y actores no estatales involucrados en conflictos violentos, o sus representantes políticos asociados.
- ✓ **Conflicto:** Violencia armada, que causa más de 25 muertes relacionadas con el conflicto en un año.

- ✓ **Proceso de paz o transición:** Es un intento formal de atraer protagonistas políticos y/o militares de conflicto, a algún tipo de acuerdo mutuo sobre cómo terminar el conflicto.

El conjunto de datos considera un total de 1832 acuerdos en más de 140 procesos de paz, se incluye los textos completos y la codificación cualitativa y cuantitativa en 266 categorías con temas relacionados con política, derecho, seguridad, género, desarrollo e implementación, etc.

La información recogida contiene acuerdos que comprenden; alto el fuego, pre-negociación, sustantivo parcial e integral, y acuerdos de implementación; todos desglosados por país / entidad, región, tipo de conflicto, tipo de acuerdo y etapa del acuerdo.

• URL DE LA VISUALIZACIÓN

A través de la plataforma de Tableau se ha integrado un “Dashboard” que muestra varias visualizaciones de los diferentes tipos de conflicto.

El URL es el siguiente:

<https://public.tableau.com/profile/fernando.meza.ibarra#!/vizhome/Tableau-Paz/Acuerdosdepazysuscategoriasunavisinmundial?publish=yes>


• HIPOTESIS Y EXPLORACION INICIAL DATOS

El motivo principal que induce a la realización de éste proyecto, es conocer de primera mano el alcance de todos los acuerdos de paz entre los países firmantes y poder entender su estado actual y su evolución en el tiempo, la finalidad es saber si éstos han sido extendidos y continuados en fomento de llegar a acuerdos de entendimientos en beneficio de la paz mundial, buscando entenderlos de acuerdo a las categorías en las que han sido contextualizadas.

a) El Dataset – Presentación

A continuación se describen algunas características del conjunto de datos:

Nombre del dataset:	pax_data_1823_agreements_Jan2020release.csv
Formato:	Archivo texto con valores separados por “comas” (csv)
Número de registros:	1832

Número de categorías:	255
Autores:	<p>Bell, Christine, Sanja Badanjak, Juline Beujouan, Robert Forster, Tim Epple, Astrid Jamar, Kevin McNicholl, Sean Molloy, Kathryn Nash, Jan Pospisil, Robert Wilson, Laura Wise (2020). Libro de códigos PA-X, versión 3.</p> <p>Bell, C. y Badanjak, S. (2019) Presentaron una nueva base de datos con acuerdos de paz, en el Journal of Peace Research , 56 (3)</p>
Programa:	Political Settlements Research Programme, University of Edinburgh
Fuente:	<p>Facultad de Derecho, Universidad de Edinburgh</p> <p>Los datos referidos se encuentran disponibles en www.peaceagreements.org</p>
Licencia y términos de uso:	<p>Acceder a https://creativecommons.org/licenses/by-nc-sa/4.0/</p> <div data-bbox="647 945 1347 1529">  <p>Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)</p> <p>Este es un resumen legible por humanos (y no un sustituto) de la licencia. Advertencia.</p> <p>Usted es libre de:</p> <p>Compartir — copiar y redistribuir el material en cualquier medio o formato</p> <p>Adaptar — remezclar, transformar y construir a partir del material</p> <p>La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia</p> </div>

	<p>Bajo los siguientes términos:</p> <p> Atribución — Usted debe dar crédito de manera adecuada, brindar un enlace a la licencia, e indicar si se han realizado cambios. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.</p> <p> NoComercial — Usted no puede hacer uso del material con propósitos comerciales.</p> <p> CompartirIgual — Si remezcla, transforma o crea a partir del material, debe distribuir su contribución bajo la misma licencia del original.</p> <p>No hay restricciones adicionales — No puede aplicar términos legales ni medidas tecnológicas que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.</p>
--	--

Fuente referida: https://www.peaceagreements.org/files/PAX_Leaflet_Version_Last_version.pdf

Conociendo sobre la recolección de información en el Dataset

- ✓ Muchos de los acuerdos se obtuvieron utilizando: colecciones existentes; sitios web específicos del país, sitios web de otros grupos civiles, además de documentación oficial de internacional.
- ✓ La información se recogido de organizaciones y gobiernos que han firmado acuerdos de paz, o mediadores involucrados en conflictos.
- ✓ Los datos recopilados a menudo se generaron cerca del tiempo del conflicto, pero también durante un período de treinta años.
- ✓ En algunas situaciones que involucran acuerdos locales, especialmente el *conflicto sirio*, se han registrado y codificado acuerdos cuyo texto no se han podido verificar. En estos casos ha sido registrado y una advertencia general sobre estos textos se observa en la base de datos.

Conociendo la estructura de información del Dataset

La información dentro del Dataset se estructurado en categorías y subcategorías, así:

- ✓ **Grupos:** Información de niños, jóvenes, personas con discapacidad, personas de la tercera edad, trabajadores migrantes, grupos raciales, grupos religiosos, pueblos indígenas, refugiados, desplazados, grupos sociales, otros.
- ✓ **Género:** Para información de género (mujeres y niñas, hombres y niños), además de información sobre orientación sexual y familia.

- ✓ **Definición de estado:** aplica a la naturaleza del estado; configuración de estado; autodeterminación; referéndum; símbolos de estado; independencia / secesión; adhesión / unificación; límites fronterizos; disposiciones transfronterizas.
- ✓ **Gobernanza:** instituciones políticas (nuevas, reformadas o temporales); Reforma Constitucional; Elecciones; Comisión Electoral; Reforma de partidos políticos; Sociedad Civil; Líderes tradicionales / religiosos; Administración Pública (servicios civiles).
- ✓ **Poder:** Político, Territorial, Económico y Militar.
- ✓ **Derechos humanos e igualdad:** Derechos Humanos / Estado de Derecho; Igualdad; Democracia; Medidas de Protección; Marco de Derechos Humanos; Derechos Civiles y Políticos; Derechos Socioeconómicos; Institución Nacional de Derechos Humanos; Instituciones Regionales o Internacionales de Derechos Humanos; Movilidad / Acceso; Procedimientos de Detención; Media y Comunicación; Ciudadanía.
- ✓ **Reforma del sector de la justicia:** Reforma de la justicia penal y disposiciones de emergencia; Judicial y Tribunales; Prisiones y Detenciones; Leyes Tradicionales / Religiosas.
- ✓ **Reconstrucción Socioeconómica:** Desarrollo o Reconstrucción Socioeconómica; Plan Económico Nacional; Recursos Naturales; Fondos Internacionales; Negocio; Impuestos; Bancos; Reforma Agraria; Derechos Pastorales / Nómadas; Patrimonio Cultural; Ambiente; Derechos de agua.
- ✓ **Sector de Seguridad:** Alto el fuego; Policía; Fuerzas Armadas; Servicios de Inteligencia; Fuerzas Rebeldes, de Oposición y Paramilitares; Retiro de fuerzas extranjeras; Corrupción; Criminalidad; Drogas y Terrorismo.
- ✓ **Justicia transicional:** Amnistía / Perdón; Tribunales y cortes; Liberación de prisioneros; Víctimas; Desaparecidos; Reparación y Reconciliación.
- ✓ **Implementación:** Firmantes de la ONU; Firmantes internacional; Acuerdos de Referéndum; Misiones Internacionales.

A partir de la clasificación de categorías, el Dataset está formado por 265 variables (categorías) y tiene 1832 observaciones o registros.

b) El Dataset – Exploración

Mediante herramientas de exploración de datos y visualizando los datos se recalcan los siguientes hechos:

- ✓ El dataset tiene 16 variables categóricas, correspondiendo la mayoría a Información Básica. Se indican los valores de algunas:

	Con	Agtp	Stage	StageSub
Bosnia and Herzegovina/Yugoslavia (former)	: 133	Inter : 72	Cea :381	MultiIss:250
Colombia	: 128	InterIntra: 236	Imp :331	Ceas :210
Philippines/Mindanao	: 126	Intra :1273	Oth : 1	Iss :191
Syria	: 92	IntraLocal: 251	Pre :510	ImpMod :188
Nepal	: 51		Ren : 42	Proc :168
Yemen	: 50		SubComp:110	Rel :128
(Other)	:1252		SubPar :457	(Other):697
	Reg			Agtp
Africa (excl MENA)	:554	Joint Statement		: 8
Americas	:195	Agreement		: 6
Asia and Pacific	:377	Joint Communique		: 6
Cross-regional	: 34	Joint Press Statement		: 4
Europe and Eurasia	:410	Memorandum of Understanding		: 3
Middle East and North Africa	:262	Protocol of the Gali Meeting between the Georgian and Abkhaz Sides		: 3
(Other)				:1802
	PPName			
Bosnia peace process	: 124			
Philippines - Mindanao process	: 124			
Syrian Local Agreements	: 64			
Abkhazia peace process	: 51			
Colombia III - Arango	: 45			
Somalia Peace Process	: 42			
(Other)	:1382			
				Part
United Nations Security Council				: 16
Israel, signed for by Yitzhak Rabin and Shimon Peres; PLO, signed for by Yasser Arafat				: 10
EZLN, Mexican Government				: 8
National Govt, FARC				: 5
British and Irish Governments				: 5
(Other)				: 4
				:1784
		ThrdPart		
		:885		
William Clinton, U.S. President; Warren Christopher, U.S.; Amr Moussa, Egypt; Bjoern Tore Godal, Norway; Andr V. Kozyrev, Russia; Hussein Ibn Talal, Jordan; Felipe Gonzalez, EU;		: 8		
Royal Government of Norway		: 7		
Agreement unsigned, meeting was chaired by the EU facilitator, Mr. Robert Cooper.		: 5		
/		: 4		
Facilitator: Henry Dunant Centre (signatory)		: 4		
(Other)		:919		

- ✓ La variable “UcdpAgr (Uppsala Conflict Data Program peace agreement ID)”, es de tipo NUMÉRICO, y contienen valores No Disponibles (NA).

UcdpAgr

Min. :2.400e+01
1st Qu.:1.107e+03
Median :1.334e+03
Mean :2.904e+10
3rd Qu.:1.523e+03
Max. :5.172e+12
NA's :1535

- ✓ La variable “**PamAgr** (Peace Accords Matrix agreement ID)”, es de tipo INTEGER y también contiene valores No Disponibles (NA).

```
PamAgr
Min.    : 1.00
1st Qu.: 9.00
Median :17.00
Mean    :17.27
3rd Qu.:26.00
Max.    :34.00
NA's    :1799
```

- ✓ La variable “**CowWar** (Correlates of War)”, es de tipo INTEGER y contiene también valores No Disponibles (NA).

```
CowWar
Min.    :139.0
1st Qu.:817.0
Median :857.0
Mean    :760.3
3rd Qu.:905.0
Max.    :941.0
NA's    :719
```

- ✓ La variable “**GAgeOth**” es una variable binaria, que tiene en todos los registros el valor 0. Ésta variable toma el valor de (1) si el acuerdo establece otras disposiciones para personas mayores o de edad avanzada que no entran en subcategorías anteriores; de lo contrario, (0).

```
GAgeOth
Min.    :0
1st Qu.:0
Median :0
Mean    :0
3rd Qu.:0
Max.    :0
```

- ✓ Las variables “**Loc1GWNO**”, “**Loc2GWNO**”, “**UcdpCon**”, son variables INTEGER y FACTOR respectivamente, las cuales tiene valor NA.

Loc1GWNO	Loc2GWNO	UcdpCon
Min. : 41.0	Min. : 2.0	289 : 129
1st Qu.:365.0	1st Qu.:345.0	308 : 124
Median :531.0	Median :371.0	299 : 93
Mean :528.8	Mean :437.3	389 : 87
3rd Qu.:678.0	3rd Qu.:625.0	309 : 81
Max. :940.0	Max. :770.0	(Other):1206
NA's :14	NA's :1598	NA's : 112

- ✓ Las variables “**Loc1ISO** y **Loc2ISO**”, tienen valores vacíos.

Loc1ISO		Loc2ISO	
PHL	: 159		:1598
BIH	: 135	SDN	: 52
COL	: 129	RUS	: 43
SYR	: 94	IRL	: 34
SSD	: 86	ARM	: 24
GEO	: 62	PER	: 14
(Other)	:1167	(Other)	: 67

- ✓ Existen variables que tienen valores extremos o atípicos (outlayers), se indican algunas.

	Contp
Government	:653
Government/territory	:835
Inter-group	:235
Other	: 1
Territory	:108

- ✓ El resto de variables son de tipo INTEGER.
- ✓ La mayoría de variables INTEGER son variables binarias, es decir que sus valores están comprendidos entre 0 y 1.
- ✓ No existe una variable que propiamente el año de firma de un acuerdo, por ello es necesario crear otra variable que extraiga el año de la variable “Dat”, que contiene la fecha como tal.

c) El Dataset – Procedimientos y Herramientas

Se utilizó al lenguaje de programación R, el cuál es un poderoso entorno en el cual podemos tratar datos y graficar. Algunas características de este lenguaje son:

- ✓ R al estar orientado a las estadísticas, proporciona un amplio abanico de herramientas.
- ✓ Su gran capacidad gráfica, que permite generar gráficos con alta calidad, con sólo utilizar las algunas funciones.
- ✓ R también puede usarse como herramienta de cálculo numérico y a la vez ser útil para la minería de datos.
- ✓ R puede integrarse con distintas bases de datos y existen bibliotecas que facilitan su utilización desde lenguajes de programación interpretados como Perl, Python

y Ruby. Y por supuesto existen proyectos que permiten utilizar R desde Java o .net de Microsoft.

- ✓ Cuenta con un poderoso entorno de desarrollo llamado “R Studio” que es de uso gratuito.

d) Procedimiento

Se ha seguido los siguientes pasos:

- Lectura del Dataset (archivo de datos): Se ha cargado el archivo en formato (csv) en el Dataset denominado “**paz**”.

El código referido en la figura siguiente muestra cómo se ha leído el conjunto de datos y ha sido almacenado en la variable “**paz**”, el cual tiene todos los acuerdos de paz que vamos a tratar:

Explorando Dataset - Acuerdo de Paz en el mundo

Fernando Meza Ibarra - Programa del Master de Ciencia de Datos - UOC

Abril-2020

Análisis de los datos

Ejercicio que realiza un pequeño análisis estadístico descriptivo e inferencial de datos procesados. La data contiene una recopilación de toda la información de acuerdos de paz entre países en todo el mundo.

1. Lectura del Fichero

Carga del Archivo de trabajo: `pax_all_agreements_data.csv`
La data mas de 1832 registros y 265 variables

```
Archivodatos <- "pax_all_agreements_data.csv"
paz<-read.csv(Archivodatos, sep=";", na.strings = "NA")
```

- Descripción visual de las variables: Mediante el comando “**sapply(paz, class)**”, se puede ver el tipo de variables que conforman el conjunto de dato; se aprecian algunas del tipo “Factor e Integer” en su mayoría.

2. Análisis descriptivo visual

Realizamos una breve descripción de los datos

```
sapply(paz, class)
```

```
##      i..Con      Contp      PP      PPName      Reg      AgtId
##      "factor"    "factor"  "integer"  "factor"    "factor"    "integer"
##      Agt        Dat        Status      Lgt N_characters Agtp
##      "factor"    "factor"  "factor"  "integer"  "integer"  "factor"
##      Stage      StageSub    Part      ThrdPart    OthAgr      Loc1ISO
##      "factor"    "factor"  "factor"  "factor"    "factor"    "factor"
##      Loc2ISO     Loc1GWNO    Loc2GWNO    UcdpCon      UcdpAgr      PamAgr
##      "factor"    "integer"  "integer"  "factor"    "numeric"    "integer"
##      CowWar      GCh        GChRhet      GChAntid      GChSubs      GChOth
##      "integer"    "integer"  "integer"  "integer"    "integer"    "integer"
```

Con el comando “**str(paz9)**”, se visualizan las variables, su tipo y algunos valores, la figura siguiente muestra la ejecución de éste comando:

```
str(paz) # Muestra los tipos y algunos valores
```

```
## 'data.frame': 1832 obs. of 265 variables:
## $ i..Con : Factor w/ 167 levels "(Bougainville)/(United Nations)",...: 3 3 3 3 3 3 3 3 3 ...
## $ Contp : Factor w/ 5 levels "Government","Government/territory",...: 2 1 1 1 1 1 1 1 1 ...
## $ PP : int 2 2 2 2 2 2 2 2 2 ...
## $ PPName : Factor w/ 156 levels "Abkhazia peace process",...: 3 3 3 3 3 3 3 3 3 ...
## $ Reg : Factor w/ 6 levels "Africa (excl MENA)",...: 5 5 5 5 5 5 5 5 5 ...
## $ AgtId : int 2232 1739 1923 864 848 849 709 708 707 706 ...
## $ Agt : Factor w/ 1796 levels "'Hudnaa\200\231 (truce) of al-Zabadani, Kefriyya and al-Fu\200\231a",...: 1513 280 304 1739 691 1007 1506 1721 630 1611 ...
## $ Dat : Factor w/ 1503 levels "1990-02-15","1990-03-01",...: 1496 1355 1238 1138 1102 1097 1052 1049 1037 1006 ...
```

Con el comando “**head(paz)**”, podemos visualizar los primeros 6 registros, la figura siguiente muestra una parte de ellos:

```
head(paz) # Muestra las 6 primeras instancias
```

```
##      i..Con      Contp PP
## 1 Afghanistan Government/territory 2
## 2 Afghanistan      Government      2
## 3 Afghanistan      Government      2
## 4 Afghanistan      Government      2
## 5 Afghanistan      Government      2
## 6 Afghanistan      Government      2
##
##      PPName      Reg AgtId
## 1 Afghanistan: 2000s Post-intervention process Europe and Eurasia 2232
## 2 Afghanistan: 2000s Post-intervention process Europe and Eurasia 1739
## 3 Afghanistan: 2000s Post-intervention process Europe and Eurasia 1923
## 4 Afghanistan: 2000s Post-intervention process Europe and Eurasia 864
## 5 Afghanistan: 2000s Post-intervention process Europe and Eurasia 848
## 6 Afghanistan: 2000s Post-intervention process Europe and Eurasia 849
```

Con el comando “**summary(paz)**”, podemos ver un resumen estadístico de las variables.

Una característica importante es que podemos detectar también el número de valores “**Vacíos**” y los valores No Disponibles (**NA**). En lo que respecta a variables numéricas (Integer), también se detecta que existen valores extremos (outlayers).

```
# Muestra NA
summary(paz)
```

La ejecución del comando se muestra en la gráfica siguiente:

##	Loc1ISO	Loc2ISO	Loc1GWNO	Loc2GWNO	UcdpCon
##	PHL : 159	:1598	Min. : 41.0	Min. : 2.0	289 : 129
##	BIH : 135	SDN : 52	1st Qu.:365.0	1st Qu.:345.0	308 : 124
##	COL : 129	RUS : 43	Median :531.0	Median :371.0	299 : 93
##	SYR : 94	IRL : 34	Mean :528.8	Mean :437.3	389 : 87
##	SSD : 86	ARM : 24	3rd Qu.:678.0	3rd Qu.:625.0	309 : 81
##	GEO : 62	PER : 14	Max. :940.0	Max. :770.0	(Other):1206
##	(Other):1167	(Other): 67	NA's :14	NA's :1598	NA's : 112
##	UcdpAgr	PamAgr	CowWar	GCh	
##	Min. :2.400e+01	Min. : 1.00	Min. :139.0	Min. :0.0000	
##	1st Qu.:1.107e+03	1st Qu.: 9.00	1st Qu.:817.0	1st Qu.:0.0000	
##	Median :1.334e+03	Median :17.00	Median :857.0	Median :0.0000	
##	Mean :2.904e+10	Mean :17.27	Mean :760.3	Mean :0.3155	
##	3rd Qu.:1.523e+03	3rd Qu.:26.00	3rd Qu.:905.0	3rd Qu.:0.0000	
##	Max. :5.172e+12	Max. :34.00	Max. :941.0	Max. :3.0000	
##	NA's :1535	NA's :1799	NA's :719		
##	GChRhet	GChAntid	GChSubs	GChOth	
##	Min. :0.00000	Min. :0.000000	Min. :0.0000	Min. :0.00000	
##	1st Qu.:0.00000	1st Qu.:0.000000	1st Qu.:0.0000	1st Qu.:0.00000	
##	Median :0.00000	Median :0.000000	Median :0.0000	Median :0.00000	
##	Mean :0.08188	Mean :0.003275	Mean :0.1048	Mean :0.01092	
##	3rd Qu.:0.00000	3rd Qu.:0.000000	3rd Qu.:0.0000	3rd Qu.:0.00000	
##	Max. :1.00000	Max. :1.000000	Max. :1.0000	Max. :1.00000	

- Para un caso de análisis de minería de datos (DM), es fundamental realizar limpieza y transformación de datos, existe la posibilidad también que en algunos casos se requiera realizar imputaciones, escalamiento o normalización de los datos o crear nuevas variables como se ha mencionado antes (*variable año de firma del acuerdo*).

Las gráficas siguientes muestran algunas transformaciones hechas sobre variables tipo “Factor” y algunas variables “Numéricas”:

3. Algunas Transformaciones - Se acondiciona valores NA a 0

```
# Transformo variables tipo Factor/Numericas en Numericas, eliminando valores NA
```

```
# Uppsala Conflict Data Program Conflict ID.
paz$UcdpCon <- as.character(paz$UcdpCon)
paz$UcdpCon <- ifelse(paz$UcdpCon == "NA", "0", paz$UcdpCon)
paz$UcdpCon <- as.numeric(paz$UcdpCon)
```

```
## Warning: NAs introducidos por coerción
```

```
class(paz$UcdpCon)
```

```
## [1] "numeric"
```

```
# Uppsala Conflict Data Program peace agreement ID
paz$UcdpAgr <- as.character(paz$UcdpAgr)
paz$UcdpAgr <- ifelse(paz$UcdpAgr == "NA", "0", paz$UcdpAgr)
paz$UcdpAgr <- as.numeric(paz$UcdpAgr)
class(paz$UcdpAgr)
```

```
## [1] "numeric"
```

```
# Kroc Institute's Peace Accords Matrix agreement ID
paz$PamAgr <- as.character(paz$PamAgr)
paz$PamAgr <- ifelse(paz$PamAgr == "NA", "0", paz$PamAgr)
paz$PamAgr <- as.numeric(paz$PamAgr)
class(paz$PamAgr)
```

```
## [1] "numeric"
```

```
# Correlates of War
paz$CowWar <- as.character(paz$CowWar)
paz$CowWar <- ifelse(paz$CowWar == "NA", "0", paz$CowWar)
paz$CowWar <- as.numeric(paz$CowWar)
class(paz$CowWar)
```

```
## [1] "numeric"
```

```
# Indicates primary location of intrastate conflict and returns the Gleditsch-Ward country code.
paz$Loc1GWNO <- as.character(paz$Loc1GWNO)
paz$Loc1GWNO <- ifelse(paz$Loc1GWNO == "NA", "0", paz$Loc1GWNO)
paz$Loc1GWNO <- as.numeric(paz$Loc1GWNO)
class(paz$Loc1GWNO)
```

```
## [1] "numeric"
```

```
# Only used in cases of interstate conflict, indicating one of the primary locations of intrastate
# conflict, and returning the Gleditsch-Ward country code
paz$Loc2GWNO <- as.character(paz$Loc2GWNO)
paz$Loc2GWNO <- ifelse(paz$Loc2GWNO == "NA", "0", paz$Loc2GWNO)
paz$Loc2GWNO <- as.numeric(paz$Loc2GWNO)
class(paz$Loc2GWNO)
```

```
## [1] "numeric"
```

Se crea también una nueva variable denominada “Year”, en la cual se registra el año en el cuál se firmó el acuerdo, esta variable es de tipo “Numérica”, para ello se extrae el número del año de la fecha que se encuentra en la variable “Dat”.

Genera nueva columna para el año

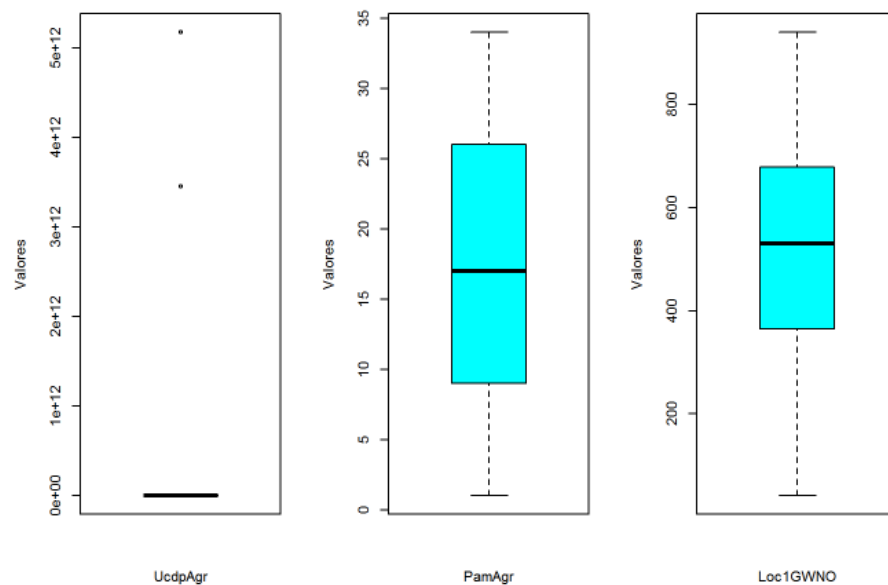
```
# Crea una nueva columna para registrar el año cuando se firmó el acuerdo
paz$Dat <- as.character(paz$Dat)
paz$Year <- substr(paz$Dat, start = 1, stop = 4)
paz$Year <- as.numeric(paz$Year)
class(paz$Year)
```

```
## [1] "numeric"
```

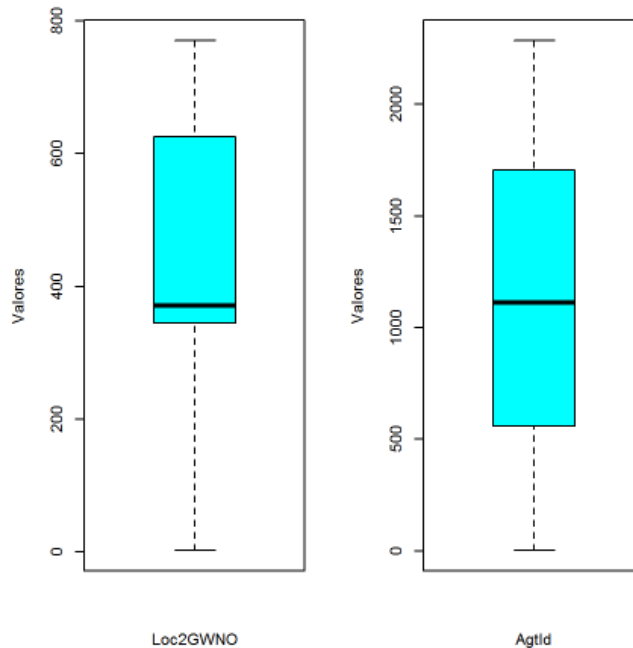
- Visualización de valores atípicos (outliers) mediante diagramas de “bigotes” o cajas, los cuales nos permiten detectar valores inusuales y que están muy alejados. Estos diagramas nos permiten luego corregir los datos extremos.

Diagramas de bigotes para algunas variables cuantitativas - Puedo ver datos extremos

```
par(mfrow=c(1,3))
boxplot(paz$UcdpAgr,xlab = "UcdpAgr", ylab = "Valores",col = "cyan")
boxplot(paz$PamAgr,xlab = "PamAgr", ylab = "Valores",col = "cyan")
boxplot(paz$Loc1GWNO,xlab = "Loc1GWNO", ylab = "Valores",col = "cyan")
```



```
par(mfrow=c(1,3))
boxplot(paz$Loc2GWNO,xlab = "Loc2GWNO", ylab = "Valores",col = "cyan")
boxplot(paz$AgtId,xlab = "AgtId", ylab = "Valores",col = "cyan")
```



- Finalmente, una vez que se ha realizado un pre-procesado del Dataset, generamos un nuevo Dataset que nos permitirá utilizarlo como insumo en la fase de Visualización de datos, que es el paso próximo.

Para ello se utilizará "TABLEAU", el cuál es una herramienta de visualización de datos potente, muy utilizada en el área de la Inteligencia de Negocios (más conocida como Business Intelligence).

TABLEAU simplifica los datos en bruto en un formato muy fácil de entender. La esencia de TABLEAU es simple y a la vez muy relevante: ayudar a las personas y empresas a ver y comprender todos sus datos a través de funciones simples como la de arrastrar y soltar, cualquier persona puede acceder y analizar de forma sencilla datos, e incluso, crear informes y compartir esta información con otros usuarios.

El set de datos ha sido generado a través del procesado que se realizó con ayuda del Lenguaje "R", el cual nos ha dado el archivo resultante "**paz.csv**".

Genero archivo depurado en "paz.csv"

```
write.csv(paz , "paz.csv", row.names = FALSE)
```

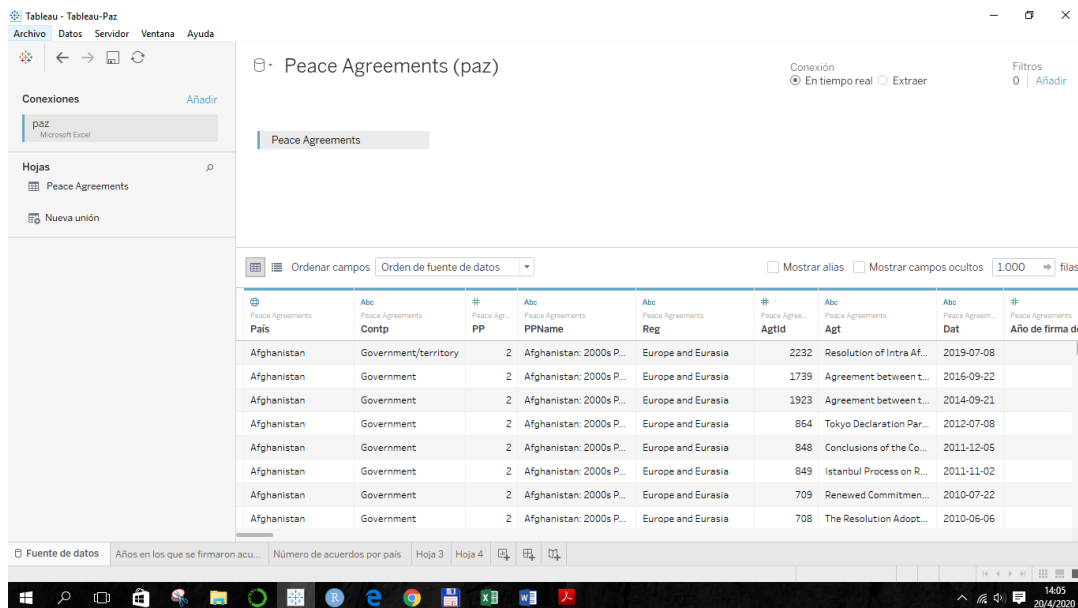
Fin de análisis.

Visualización y Transformaciones de Dataset. UOC - 2020

e) Primer diseño de Visualización

Como primer preámbulo de visualización usando la herramienta TABLEAU, podemos ver algunos gráficos dinámicos que se presentan en un **“Dashboard”**.

Los datos han sido convertidos a formato Excel y cargados a TABLEAU, la gráfica siguiente muestra el conjunto de datos subidos (Dataset **“paz.xlsx”**).



País	Contp	PP	PPNName	Reg	Agtid	Agt	Dat	Año de firma de
Afghanistan	Government/territory	2	Afghanistan: 2000s P...	Europe and Eurasia	2232	Resolution of Intra Af...	2019-07-08	
Afghanistan	Government	2	Afghanistan: 2000s P...	Europe and Eurasia	1739	Agreement between t...	2016-09-22	
Afghanistan	Government	2	Afghanistan: 2000s P...	Europe and Eurasia	1923	Agreement between t...	2014-09-21	
Afghanistan	Government	2	Afghanistan: 2000s P...	Europe and Eurasia	864	Tokyo Declaration Par...	2012-07-08	
Afghanistan	Government	2	Afghanistan: 2000s P...	Europe and Eurasia	848	Conclusions of the Co...	2011-12-05	
Afghanistan	Government	2	Afghanistan: 2000s P...	Europe and Eurasia	849	Istanbul Process on R...	2011-11-02	
Afghanistan	Government	2	Afghanistan: 2000s P...	Europe and Eurasia	709	Renewed Commitmen...	2010-07-22	
Afghanistan	Government	2	Afghanistan: 2000s P...	Europe and Eurasia	708	The Resolution Adopt...	2010-06-06	

Siguiendo la perspectiva de la visualización propuesta, **“Acuerdos de paz y sus categorías, una visión mundial”**, se ha construido un **“Dashboard”** primario al cual se puede acceder a través del link referido al inicio de éste documento.

En éste panel, se puede encontrar o visualizar los siguientes gráficos:

- ✓ Ubicación de países por Tipo de Conflicto
- ✓ Número de acuerdos por país
- ✓ Otros acuerdos por país
- ✓ Años en que los países firmaron acuerdos

• EL PROCESO DE VISUALIZACIÓN DE DATOS

Siendo la Visualización de Datos o Visualización de Información la disciplina que se encarga de generar representaciones visuales de datos abstractos. En general, se puede definir como:

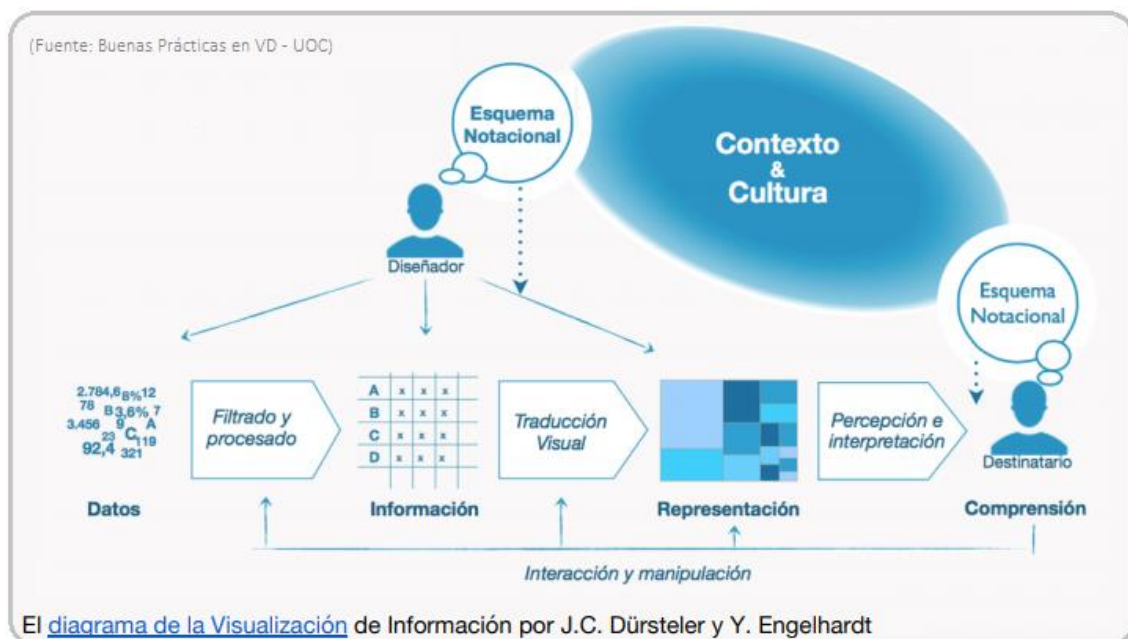
"The use of computer-supported, interactive, visual representations of abstract data to amplify cognition¹"

"El uso de representaciones visuales interactivas y apoyadas por computadora de datos abstractos para amplificar la cognición"

Las diferencias entre "Visualización de Datos" y "Visualización de Información" son mínimas, pero en cierto modo pueden llegar a ser importantes. Como se podrá observar, la diferencia reside en los conceptos de "**dato**" e "**información**". Mientras los "datos" son esencialmente números o incluso texto sin procesar ni manipular; la información es aquello que se obtiene al hacerlo. Además, podemos decir que la información es básicamente datos estructurados y puestos en contexto.

Por lo que se refiere a la definición de visualización, lo más importante es tener en cuenta que, generalmente, nos encontramos con datos abstractos, que no tienen forma. Representarlos visualmente mediante metáforas visuales ayuda a nuestro cerebro a pensar más rápido y mejor, favoreciendo la generación de conocimiento.

- El diagrama de la Visualización de Datos²



El "**Diagrama de la Visualización de Información**" creado por J.C. Dürsteler y Y. Engelhardt en el año 2000 explica el proceso que se debe seguir cuando se visualizan datos. En él podemos ver como el conjunto de datos inicial debe de ser filtrado y procesado para poder llegar a tener información, esto facilita el análisis.

¹ Card et al. 1999

² El artículo "Difference between data and information tiene una buena descripción de estos dos conceptos.
<https://blog.udemy.com/difference-between-data-and-information/>

Esta información deberá ser representada en imágenes, para este proyecto, en imágenes netamente interactivas, que serán consumidas por un usuario influido por su percepción e interpretación.

Estos dos factores se ven afectados además por un esquema “notacional” que usaremos; significa que no todo el mundo interpretará del mismo modo todas las representaciones visuales, sino que la experiencia, el contexto y la cultura adquirida a lo largo del tiempo nos ayudarán a entender mejor algunas de ellas.

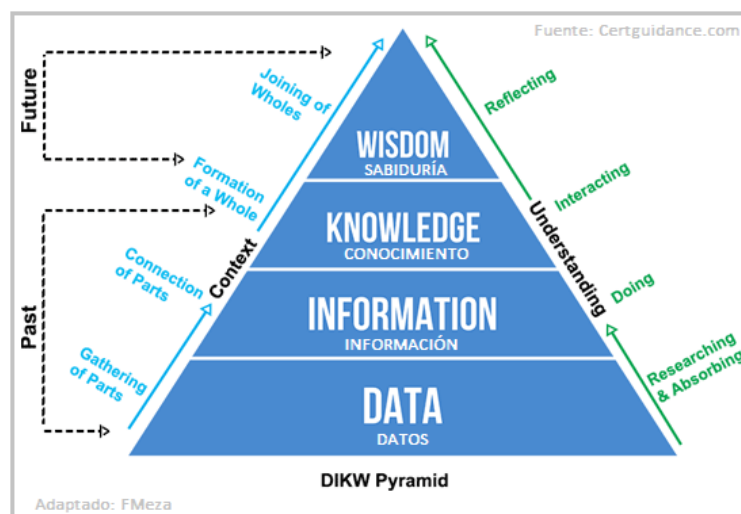
Sería importante poder definir el perfil del usuario consumidor de las visualizaciones para así poderlo estudiar y desarrollar soluciones que se adapten a su contexto y cultura. Finalmente, el diagrama muestra cómo el usuario puede influir en cada uno de estos pasos mediante la interacción. Es decir, hay que tener en cuenta que en algunos sistemas el usuario podrá decidir que representación visual quiere aplicar a los dos, o como limpiar y procesarlos para conseguir información.

Considerando la base de datos mantenida por la Universidad de Edinburg ubicada en Escocia, Reino Unido, la cual se ha configurado para registrar todos los acuerdos de paz entre países con el fin de dar a conocer y facilitar la comprensión de las vías negociadas fuera de conflicto, se implementará un proyecto de Visualización de Información que intenta dar respuesta a los procesos de paz entre países del mundo que han tenido conflictos de gobierno y territoriales principalmente, para entender esta realidad se recurre a los datos creados y mantenidos por esta universidad.

- **Comunicar: Tipo de análisis y visualización**

Desarrollando el concepto de la Jerarquía del conocimiento¹, buscaremos realizar este proyecto para comunicar una idea o el resultado de un análisis hecho sobre estos datos, los cuales se han transformado en información.

El objetivo de la visualización persigue que académicos, mediadores, actores en conflicto y ciudadanos civiles que forman parte de la sociedad puedan tener acceso a dicha información de forma visual de una manera sobria, concisa y actualizada; la visualización permitirá navegar y comparar sistemáticamente por las diferentes categorías, muestra cómo se formalizan los procesos de paz y la transición de compromisos negociados en un intento de avanzar hacia la paz mundial.



Con la información proporcionada, podemos considerar aplicar los siguientes análisis:

Análisis Temporal: En este tipo de análisis se busca entender y descubrir los patrones que pueden tener los datos en el tiempo como por ejemplo ver si una métrica en cuestión crece o decrece, o si existen ciclos estacionales. Veremos el comportamiento de los Acuerdos firmados desde 1930 hasta enero de 2020.

Análisis de Proporciones y Ranking: Dos de los tipos de análisis más utilizados son la comparación de un conjunto de valores que forman parte de un todo y su ranking u ordenación por valor. Por ejemplo, en el conjunto de datos, sería deseable ver qué porcentaje del total ocupa cada categoría (variable "**Contp**"), podremos ver cuál de ellas es la más importante por la incidencia.

Análisis de Relaciones: A menudo lo que se desea de un dataset es descubrir qué relaciones existen entre dos o más variables que lo forman. Para ello, podemos utilizar variables geográficas con el fin de visualizar la presencia de las categorías de acuerdos de país con respecto a la ubicación de los países, para ello una de las mejores representaciones visuales que existen es el gráfico de dispersión o scatter plot, Mapas.

Análisis Multidimensionales: Podemos aplicar éste análisis especialmente sobre variable cuantitativas, que nos den la oportunidad de representarlas en los ejes X e Y. Por ejemplo para determinar cuántos acuerdos se firmaron en un determinado año.

Análisis Espaciales: Los mapas son el mejor sistema visual para hacer análisis de datos espaciales. Parte de la importancia del mapeo de valores en un mapa depende de la proyección que se utilice. Una de las técnicas más utilizada es la de Mercator. Es muy importante tener en cuenta que, el mero hecho de disponer de datos geográficos no significa que la mejor representación a utilizar sea un mapa.

• Explorando los datos

Con los datos existentes tenemos al alcance todo un cúmulo de información que nos permite explorar a través de la herramienta de visualización TABLEAU, con lo que logramos varias perspectivas de comunicación visual, podemos citar algunas que se intentará resolver:

- ✓ Acuerdos de paz en el mundo y su repercusión económica
- ✓ Acuerdos de paz desde una perspectiva de género
- ✓ Acuerdos de paz y los efectos bélicos en el mundo
- ✓ Evolución de los acuerdos de paz en el tiempo
- ✓ Acuerdos de paz por grupos de acuerdos civiles y políticos por región y año
- ✓ Mapas de acuerdos por países firmantes
- ✓ Acuerdos de paz desde una perspectiva de orientación sexual, etc

• Datos en contexto

Para lograr una buena visualización se ha interactuado con los datos que se presentan, así por ejemplo utilizando la aplicación (<https://rawgraphs.io/>) se ha logrado cargar todo el dataset y se ensayado, alternando con el color, el tamaño e incluyendo pequeñas descripciones cortas que resalten las ideas clave de lo que se quiere comunicar.

Las figuras siguientes muestran esta práctica:

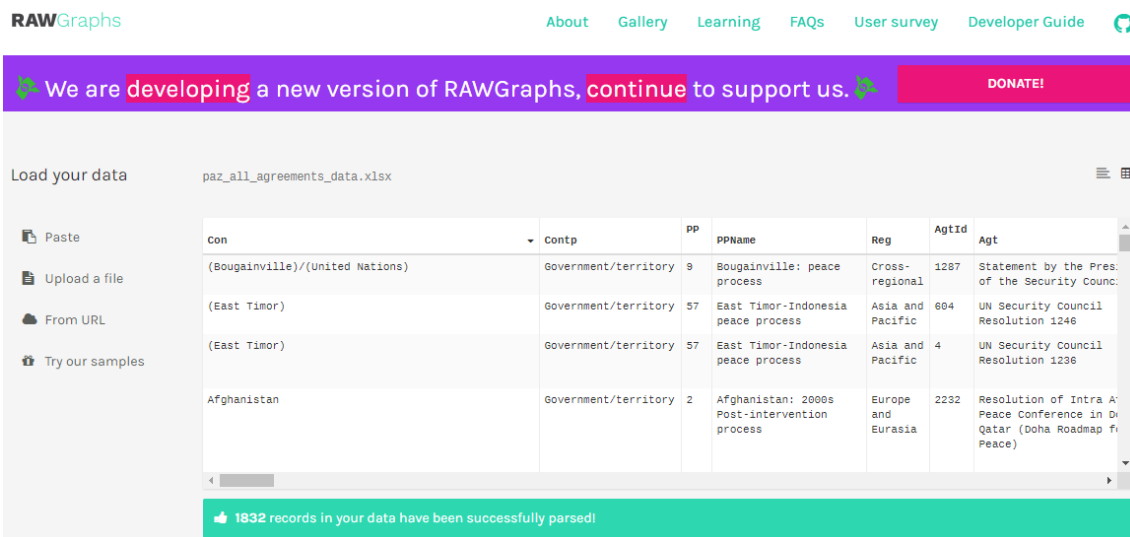


Fig. 1- Uso de RAW para ensayo de visualizaciones.

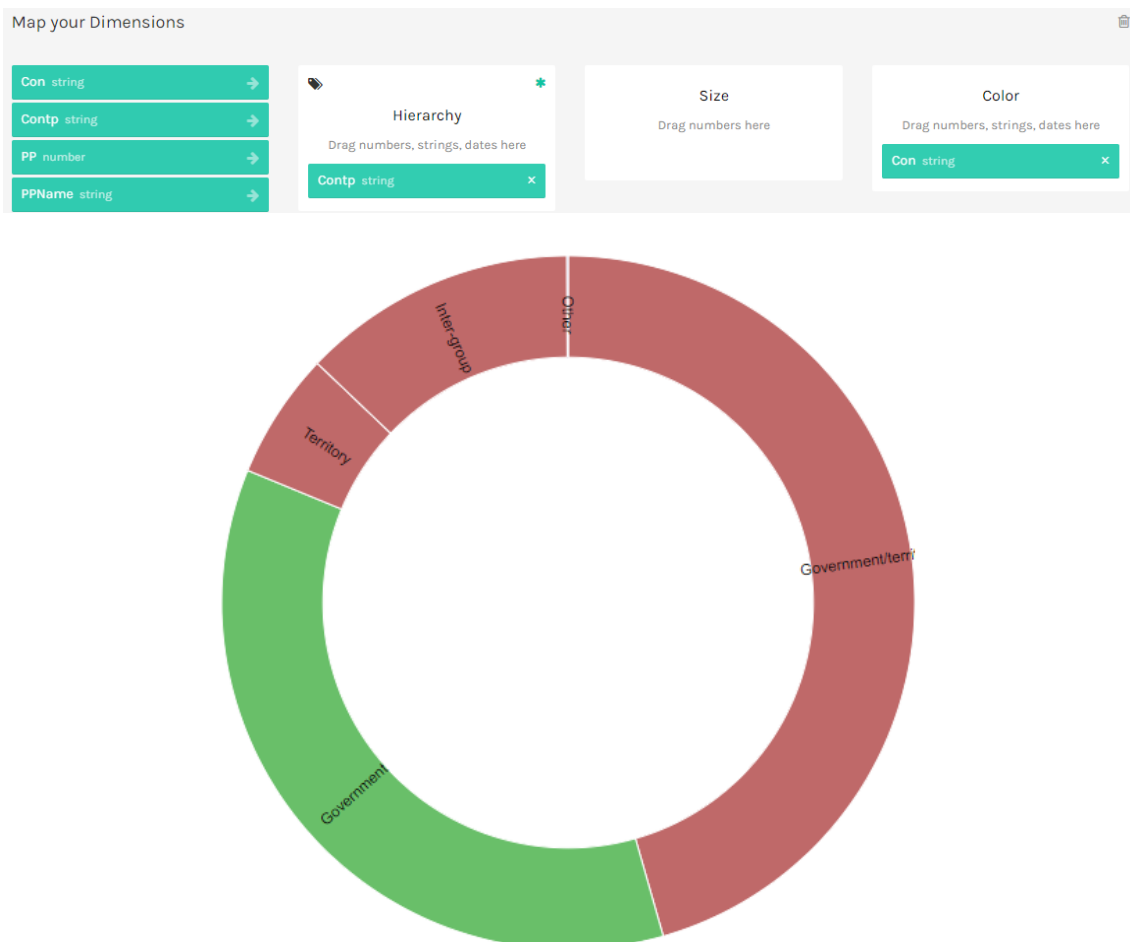


Fig. 2- Visualización con RAW

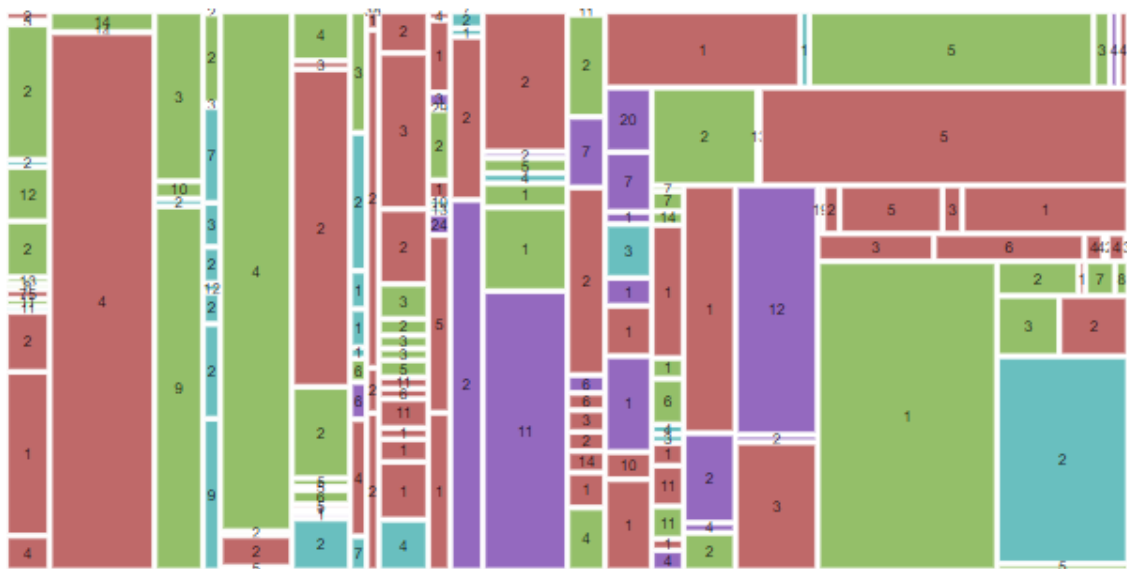


Fig. 3- Visualización con RAW

- **Encontrar patrones y datos extremos**

Se ha ensayado con variables que tienen valores extremos, las figuras siguientes muestran el resultado de mostrar las categorías de acuerdo de paz (Contp), en función de las variables (Lgt y Loc1GWNO)

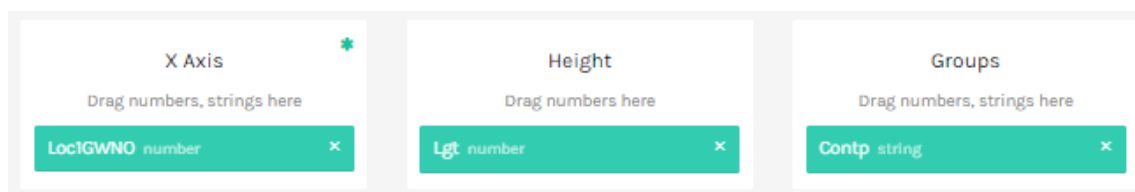


Fig.4- Muestra de patrones y datos extremos de las variables (Lgt y Loc1GWNO)



Fig.5- Muestra visualización de patrones y datos extremos de las variables (Lgt y Loc1GWNO)

• Descripción del dataset

La figura siguiente muestra algunas de las variables y observaciones del conjunto de datos.

Con	Contp	PP	PPName	Reg	AgId	Ag	Dat	Year	Status	Lgt	N	charact	Agtp	Stage	StageSub	Part	ThrdPart
Afghanistan	Government/territory	2	Afghanistan: 2000s Pos Europe and Eurasia	2232	Resolution	2019-07-01	2019	Multipart		2	5235	Intra	Pre	PreMix	Stated to I QatarGerri		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	1739	Agreemer	2016-09-21	2016	Multipart		4	10203	Intra	SubPar	Multiss	President of the IRA		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	1923	Agreemer	2014-09-21	2014	Multipart		4	11110	Intra	Imp	ExtSub	Dr. MohanThe foreg		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	864	Tokyo Dec	2012-07-01	2012	Multipart		14	39746	Interintra	Imp	ExtSub	Not signeiAmbassac		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	848	Conclusio	2011-12-01	2011	Multipart		7	24106	Interintra	Pre	Prin	Not signed, agreemei		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	849	Istanbul P	2011-11-01	2011	Multipart		10	21299	Interintra	Imp	ExtSub	Not signeiNot signei		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	709	Renewed	2010-07-21	2010	Multipart		9	33137	Interintra	Imp	ExtSub	Not signed, agreemei		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	708	The Resol	2010-06-01	2010	Multipart		4	9289	Intra	Pre	PreMix	Not signed, agreemei		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	707	Communi	2010-01-21	2010	Multipart		7	27077	Interintra	Imp	ExtSub	Not signed, agreemei		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	706	Statemen	2009-03-31	2009	Multipart		4	9555	Interintra	SubPar	Multiss	Not signeiNot signei		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	847	Declaratic	2009-03-21	2009	Multipart		3	5677	Interintra	Imp	ExtSub	Not signed, agreemei		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	705	Declaratic	2008-06-11	2008	Multipart		3	8807	Interintra	Imp	ExtSub	Not signeiNot signei		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	1069	Rome Cor	2007-07-01	2007	Multipart		1	3231	Interintra	Imp	ImpMod	Not signed, agreemei		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	1061	Afghanist	2006-02-01	2006	Multipart		17	39498	Interintra	SubPar	Multiss	Afghanist: Listed as c		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	1054	Berlin Dec	2004-04-01	2004	Multipart		3	7634	Interintra	Imp	ExtSub	Not signed, but parti		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	909	Kabul Dec	2002-12-21	2002	Multipart		1	1690	Interintra	SubPar	Multiss	The Transitional Adm		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	908	Communi	2002-01-21	2002	Multipart		5	10729	Interintra	Imp	ImpMod	Not signeiNot signei		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	272	Agreemer	2001-12-01	2001	Multipart		13	23377	Intra	SubPar	Multiss	Ms. AmenAgreemei		
Afghanistan	Government	2	Afghanistan: 2000s Pos Europe and Eurasia	553	Tashkent	1999-07-11	1999	Multipart		5	8371	Interintra	Pre	Proc	Not signeiThe agreee		

Fig: 6- Extracto del dataset de acuerdos de paz

Conociendo sobre la recolección de información en el Dataset

- ✓ Muchos de los acuerdos se obtuvieron utilizando: colecciones existentes; sitios web específicos del país, sitios web de otros grupos civiles, además de documentación oficial de internacional.
- ✓ La información se recogido de organizaciones y gobiernos que han firmado acuerdos de paz, o mediadores involucrados en conflictos.
- ✓ Los datos recopilados a menudo se generaron cerca del tiempo del conflicto, pero también durante un período de treinta años.
- ✓ En algunas situaciones que involucran acuerdos locales, especialmente el *conflicto sirio*, se han registrado y codificado acuerdos cuyo texto no se han podido verificar. En estos casos ha sido registrado y una advertencia general sobre estos textos se observa en la base de datos.

Conociendo la estructura de información del Dataset

La información dentro del Dataset se estructurado en categorías y subcategorías, así:

- ✓ **Grupos:** Información de niños, jóvenes, personas con discapacidad, personas de la tercera edad, trabajadores migrantes, grupos raciales, grupos religiosos, pueblos indígenas, refugiados, desplazados, grupos sociales, otros.
- ✓ **Género:** Para información de género (mujeres y niñas, hombres y niños), además de información sobre orientación sexual y familia.
- ✓ **Definición de estado:** aplica a la naturaleza del estado; configuración de estado; autodeterminación; referéndum; símbolos de estado; independencia / secesión; adhesión / unificación; límites fronterizos; disposiciones transfronterizas.
- ✓ **Gobernanza:** instituciones políticas (nuevas, reformadas o temporales); Reforma Constitucional; Elecciones; Comisión Electoral; Reforma de partidos políticos;

- Sociedad Civil; Líderes tradicionales / religiosos; Administración Pública (servicios civiles).
- ✓ **Poder:** Político, Territorial, Económico y Militar.
 - ✓ **Derechos humanos e igualdad:** Derechos Humanos / Estado de Derecho; Igualdad; Democracia; Medidas de Protección; Marco de Derechos Humanos; Derechos Civiles y Políticos; Derechos Socioeconómicos; Institución Nacional de Derechos Humanos; Instituciones Regionales o Internacionales de Derechos Humanos; Movilidad / Acceso; Procedimientos de Detención; Media y Comunicación; Ciudadanía.
 - ✓ **Reforma del sector de la justicia:** Reforma de la justicia penal y disposiciones de emergencia; Judicial y Tribunales; Prisiones y Detenciones; Leyes Tradicionales / Religiosas.
 - ✓ **Reconstrucción Socioeconómica:** Desarrollo o Reconstrucción Socioeconómica; Plan Económico Nacional; Recursos Naturales; Fondos Internacionales; Negocio; Impuestos; Bancos; Reforma Agraria; Derechos Pastorales / Nómadas; Patrimonio Cultural; Ambiente; Derechos de agua.
 - ✓ **Sector de Seguridad:** Alto el fuego; Policía; Fuerzas Armadas; Servicios de Inteligencia; Fuerzas Rebeldes, de Oposición y Paramilitares; Retiro de fuerzas extranjeras; Corrupción; Criminalidad; Drogas y Terrorismo.
 - ✓ **Justicia transicional:** Amnistía / Perdón; Tribunales y cortes; Liberación de prisioneros; Víctimas; Desaparecidos; Reparación y Reconciliación.
 - ✓ **Implementación:** Firmantes de la ONU; Firmantes internacional; Acuerdos de Referéndum; Misiones Internacionales.

A partir de la clasificación de categorías, el Dataset está formado por 266 variables (categorías) y tiene 1832 observaciones o registros.

Mediante herramientas de exploración de datos y visualizando los datos se recalcan los siguientes hechos:

- ✓ El dataset tiene 16 variables categóricas, correspondiendo la mayoría a Información Básica. Se indican los valores de algunas:

```

UcdpAgr
Min.      :2.400e+01
1st Qu.   :1.107e+03
Median    :1.334e+03
Mean      :2.904e+10
3rd Qu.   :1.523e+03
Max.      :5.172e+12
NA's      :1535

```

- ✓ La variable "**PamAgr** (Peace Accords Matrix agreement ID)", es de tipo INTEGER y también contiene valores No Disponibles (NA).

PamAgr

```
Min.      : 1.00
1st Qu.   : 9.00
Median    :17.00
Mean      :17.27
3rd Qu.   :26.00
Max.      :34.00
NA's      :1799
```

- ✓ La variable “**CowWar** (Correlates of War), es de tipo INTEGER y contiene también valores No Disponibles (NA).

CowWar

```
Min.      :139.0
1st Qu.   :817.0
Median    :857.0
Mean      :760.3
3rd Qu.   :905.0
Max.      :941.0
NA's      :719
```

- ✓ La variable “**GAgeOth**” es una variable binaria, que tiene en todos los registros el valor 0. Ésta variable toma el valor de (1) si el acuerdo establece otras disposiciones para personas mayores o de edad avanzada que no entran en subcategorías anteriores; de lo contrario, (0).

GAgeOth

```
Min.      :0
1st Qu.   :0
Median    :0
Mean      :0
3rd Qu.   :0
Max.      :0
```

- ✓ Las variables “**Loc1GWNO**”, “**Loc2GWNO**”, “**UcdpCon**”, son variables INTEGER y FACTOR respectivamente, las cuales tiene valor NA.

Loc1GWNO	Loc2GWNO	UcdpCon
Min. : 41.0	Min. : 2.0	289 : 129
1st Qu.:365.0	1st Qu.:345.0	308 : 124
Median :531.0	Median :371.0	299 : 93
Mean :528.8	Mean :437.3	389 : 87
3rd Qu.:678.0	3rd Qu.:625.0	309 : 81
Max. :940.0	Max. :770.0	(Other):1206
NA's :14	NA's :1598	NA's : 112

- ✓ Las variables “**Loc1ISO** y **Loc2ISO**”, tienen valores vacíos.

Loc1ISO		Loc2ISO	
PHL	: 159		:1598
BIH	: 135	SDN	: 52
COL	: 129	RUS	: 43
SYR	: 94	IRL	: 34
SSD	: 86	ARM	: 24
GEO	: 62	PER	: 14
(Other)	:1167	(Other)	: 67

- ✓ Existen variables que tienen valores extremos o atípicos (outlayers), se indican algunas.

	Contp
Government	:653
Government/territory	:835
Inter-group	:235
Other	: 1
Territory	:108

- ✓ El resto de variables son de tipo INTEGER.
- ✓ La mayoría de variables INTEGER son variables binarias, es decir que sus valores están comprendidos entre 0 y 1.

No existe una variable que propiamente el año de firma de un acuerdo, por ello se ha creado otra variable que extrae el año de la variable “**Dat**”, que contiene la fecha como tal, esta se denomina “**year**”.

• Limpieza de los datos

Para el proceso de limpieza se utilizó al lenguaje de programación R, el cuál es un poderoso entorno en el cual podemos tratar datos y graficar. Algunas características de este lenguaje son:

- ✓ R al estar orientado a las estadísticas, proporciona un amplio abanico de herramientas.
- ✓ Su gran capacidad gráfica, que permite generar gráficos con alta calidad, con sólo utilizar las algunas funciones.
- ✓ R también puede usarse como herramienta de cálculo numérico y a la vez ser útil para la minería de datos.
- ✓ R también puede integrarse con distintas bases de datos y existen bibliotecas que facilitan su utilización desde lenguajes de programación interpretados como

Perl, Python y Ruby. Y por supuesto existen proyectos que permiten utilizar R desde Java o .net de Microsoft.

- ✓ Cuenta con un poderoso entorno de desarrollo llamado “R Studio” que es de uso gratuito el cuál se ha utilizado en éste proyecto.

- **Relaciones**

Las relaciones que se desea mostrar en la visualización son las siguientes:

- ✓ Categorías de acuerdos de paz y su relación con los países firmantes
- ✓ Categorización de acuerdos de paz por año de firma
- ✓ Categorización de acuerdos de paz y su efecto en aspectos sociales de género

- **Gráficos seleccionados**

Los gráficos del proyecto serán:

Gráficos de Barras

Ya que éste uno de los tipos de representaciones visuales más comunes. Se usa para comparar distintos valores y es muy eficaz puesto que el ojo humano es muy bueno comparando alturas.

En este caso, la disposición de las barras se deberá hacer de izquierda a derecha, que es como convencionalmente se representa la evolución del tiempo. Utilizaremos este tipo de gráfico porque disponemos de datos discretos. Los datos temporales pueden ser categorizados entre discretos o continuos. En el caso de los discretos, los valores que se disponen se refieren a bloques específicos de tiempo, y existe un número finito de valores.

Gráfico Treemap

Otra alternativa para poder visualizar proporciones, o valores que forman parte de un todo, es el Treemap. Este tipo de representación visual es cada vez más común, y tiene la potencia de poder representar jerarquías utilizando el máximo de espacio disponible para el gráfico. La siguiente figura muestra un ejemplo de jerarquía, y como ésta se representa mediante la técnica del treemap (variable “**Contp**”).



Fig.7- Gráficos Treemap que muestra las categorías de acuerdos de paz (variable Contp)

Gráficos de Mapas y Heatmap

Adicional, se intentará realizar visualizaciones para análisis espacial y de relaciones, para ello utilizaremos los heatmaps o mapas de calor que son tablas donde sus celdas están coloreadas en base al valor que albergan, permiten una fácil comprensión. Los mapas son el mejor sistema visual para hacer análisis de datos espaciales, éstos son útiles solo cuando se quieren descubrir patrones geográficos en los datos, en nuestro caso consideramos aspectos como la geo localización de los países.

Los usuarios podrán interactuar con los datos a través de la propia visualización, ya que las mismas son interactivas.

La herramienta TABLEAU que se usará permite al usuario tal interactividad.

A través de un panel (**dashboard**) se permite visualizar gráficos primarios o secundarios, si habrá varias visualizaciones enlazadas en una historia global.

Sobre aspectos de diseño se ha considerado lo siguiente:

- **Colores a utilizar**

La utilización correcta del color considera tener en cuenta dos tipos principales de escalas de color:

- ✓ **La numérica:** que nos permite representar valores numéricos. Ésta está generalmente formada por un único color, la saturación del cual cambia en función del valor a representar. En el caso de disponer de números centrados en un valor, se utilizará un color para aquellos valores por debajo, y otro color para los valores por encima, es decir usaremos una escala divergente
- ✓ **La categórica:** para representar distintos grupos entre sí. Para ello se ha utilizado colores que sean lo suficientemente distintos entre sí para poder distinguir los elementos que forman parte de cada una de las categorías de acuerdos de paz.

- **Tipografías**

Se usará una o dos fuentes en este proyecto, al decidir qué fuentes usar, menos es más. Las tipografías de visualización tienen muchos estilos, como cursiva, negrita, mayúscula y fuente de lujo. Este tipo de letras suelen utilizarse para textos pequeños debido a su naturaleza decorativa. Por ejemplo, títulos, encabezados y diseños gráficos complejos.

- **Plantillas o estructuras**

Se usarán las provistas por TABLEAU.

- **Bloques**

Se aplicará un solo bloque de visualización.

- **Formato**

El proyecto maneja formato digital.

- **EL PRODUCTO: LA VISUALIZACIÓN FINAL**

TABLEAU, la herramienta de visualización

Cómo se refirió al inicio, el Proyecto de Visualización denominado “ACUERDOS DE PAZ Y SUS CATEGORÍAS, UNA VISIÓN MUNDIAL” se encuentra desplegado en el repositorio de TABLEAU³, al cual se accede a través del siguiente enlace de internet:

<https://public.tableau.com/profile/fernando.meza.ibarra#!/vizhome/Tableau-Paz/Acuerdosdepazysuscategorasunavisinmundial?publish=yes>

Dado que ya tenemos un conocimiento cabal y suficiente de los elementos básicos de la teoría de visualización de datos podemos utilizar TABLEAU, la cual es una de las herramientas de software más avanzadas que existen en la actualidad para analizar datos visualmente y generar representaciones interactivas con objeto de narrar historias mediante los datos.

TABLEAU es reconocido como Líder en el Cuadrante Mágico Gartner 2020 para plataformas de análisis e inteligencia empresarial, por encima de “Qlik View” y “Power BI” de Microsoft.

TABLEAU es una herramienta de business intelligence enfocada especialmente en facilitar el análisis visual e interactivo de datos. Dicho de otra manera, TABLEAU se

³ Introducción a Tableau - Víctor Pascual Cid - fuente (FUOC. Fundación por la Universitat Oberta de Catalunya)

puede ver como una especie de Excel u hoja de cálculo avanzada, muy focalizada y preparada para explorar datos de manera visual, al tiempo que permite también generar visualizaciones o dashboards (cuadros de mando) para que otras personas exploren y analicen conjuntos de datos de forma interactiva.

En esencia; TABLEAU nos permite explorar datos y comunicar o narrar historias a través de los mismos, a través de inferencias, las cuáles son las mejores maneras para representarlas, y ayudar al usuario en el proceso de creación de la visualización de datos.

Mediante una licencia de estudiante otorgada para éste proyecto nos permitirá gestionar un espacio donde se alojará el “Dashboard” que muestra las diferentes gráficas elaboradas. Existe una versión pública que aprovecha la mayoría de las funcionalidades de la versión de pago, cuyo ecosistema, que por cierto son bastante altos.

Las credenciales de acceso para ésta licencia son:

Usuario : fmeza@uoc.edu

Contraseña : Tableau\$1xxxxxx5t

TABLEAU, nos permite aplicar numerosas técnicas de análisis y elementos estadísticos a partir del conjunto de datos de “Acuerdos de Paz”, estos elementos los podemos incorporar a nuestras visualizaciones para entender mejor los datos y extraer conocimiento de los mismos a partir del uso de representaciones gráficas.

La preparación de datos

Los datos utilizados en este proyecto se han preparado siguiendo los conceptos de pre procesado, limpieza, integración y validación, permitiendo construir una estructura de fila-columna que será el insumo para el análisis y representación visual, TABLEAU los identifica como *Dimensiones* y *Medidas*.

Las métricas o medidas son las columnas de los datos que contienen valores numéricos y que, por lo tanto, se pueden sumar, es posible calcular su media, etc. Por otro lado, las dimensiones son aquellas columnas que contienen información categórica que nos permitirá segmentar los datos; por ejemplo, una columna que contenga el nombre del país, en la cual veremos el número de acuerdos de paz.

- **Ubicación de países por Tipo de Conflicto**

Muestra un gráfico de mapas dónde se puede localizar a cada país firmante y el tipo de conflicto al que corresponde. Ver Fig.: 8.

Acuerdos de paz y sus categorías, una visión mundial

Ubicación de países por Tipo de Conflicto

Tipo de Conflicto



Fig.: 8 – Ubicación de países por Tipo de conflicto

- **Número de acuerdos por país**

Representa a cada país y el número de acuerdos que ha firmado, ver Fig.: 9.

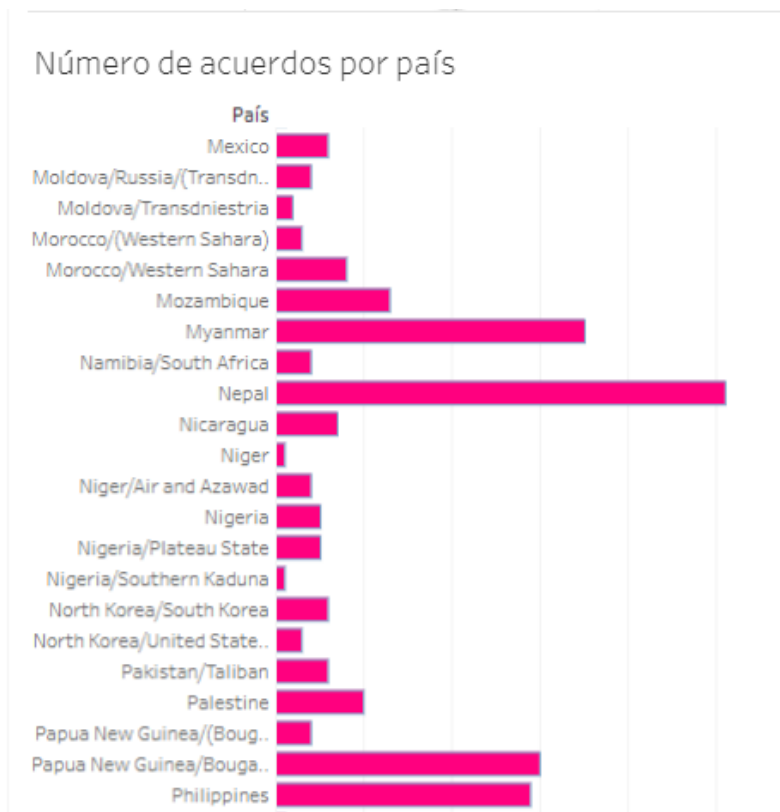


Fig.: 9 – Número de acuerdos

- **Otros acuerdos por país**

Representa el número de acuerdos de las diferentes categorías de acuerdo al tipo de conflictos de un país o países firmantes. Ver Fig.:10.

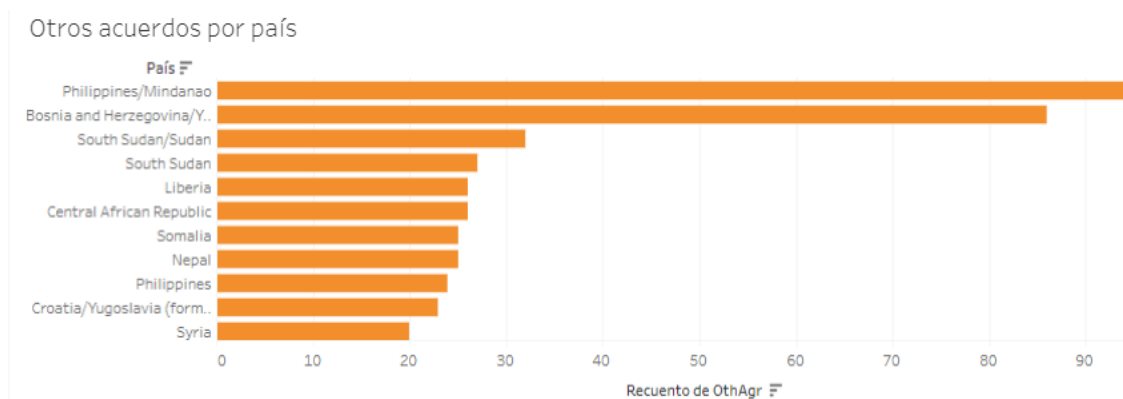


Fig.: 10 – Número de acuerdos por Tipo de conflicto

TABLEAU, incluye el archivo Tableau-Paz.twb, el cual contiene la metadata del “Dashboard” con el nivel de abstracción requerido para representar la “historia” de los acuerdos de paz con perspectiva mundial.

- **CONCLUSIONES Y EXPLICACIÓN**

Con el desarrollo de ésta práctica, se ha logrado entender y explorar datos de gran volumen, nos ha permitido poder realizar o aplicar inferencia estadística fundamental para transformar y visualizar en primera instancia el comportamiento de los datos.

Ha sido enriquecedora la experiencia del tratamiento de datos a través de la herramienta de programación R, la cual tiene su grado de experticia pero es una gran aliada de la visualización de datos.

Con la base concreta sobre la inmensa diversidad de herramientas y técnica de visualización se ha podido conocer a TABLEAU, con la cual se ha construido las diferentes visualizaciones consolidadas en un “Dashboard”.

Como primera práctica se han realizado varias visualizaciones inherentes a los acuerdos de paz en el mundo desde un contexto de categorías según los acuerdos firmados y sus actores involucrados.

Vemos que las representaciones gráficas tienen comportamientos dinámicos lo que facilitan los procesos iterativos con los usuarios lectores, logrando una fácil comprensión de lo que se quiere informar con las mismas.

Con respecto a TABLEAU, podemos decir que a pesar de que existe el acceso a una licencia de “estudiante”, una licencia de funcionalidad completa no está al alcance de

usuarios normales, son las empresas que buscan hacerse de esta herramienta que se mantiene en el “top” de usabilidad en el mundo.

Los profesionales de hoy, tienen en la visualización de datos a un gran aliado en la toma de decisiones. Si éstas son bien tomadas les permitirá mantenerse y mantener a sus organizaciones dentro del concierto empresarial competitivo.

Se intentará en el futuro aprovechar y desarrollar todas las funcionalidades de la herramienta, aunado al conocimiento avanzado sobre técnicas y nuevas tendencias que permitan incursionar más profundamente en éste maravillosos mundo de las representaciones gráficas.

- **CÓDIGO Y DOCUMENTACIÓN**

En el repositorio de GITHUB, podemos encontrar toda la documentación relativa a ésta práctica, el acceso es por el enlace siguiente:

https://github.com/fmezaibarra/Visualizaci-n_acuerdos_paz_en_el_mundo

Fin de Proyecto de Visualización
Creado por Fernando Meza Ibarra
Máster Universitario en Ciencia de Datos – UOC
Mayo 2020