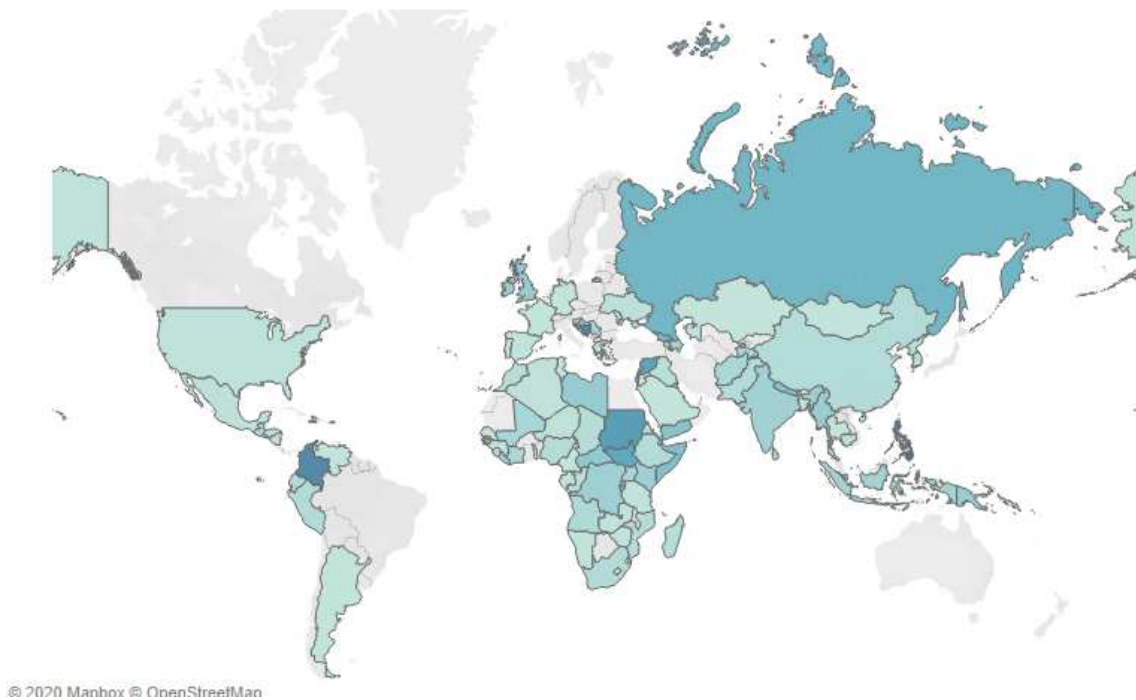


Visualización de datos



PEC4

Proyecto de visualización de datos: Creación de la visualización

Tabla de contenido

1. Título de la visualización y URL.....	2
2. Descripción	2
3. ¿Qué se cumple y que no del proyecto? Valoración del esfuerzo y aprendizaje.....	3
4. Descripción técnica.....	5
5. Visualización realizada.....	6
5.1. Visualización de la PEC2	6
5.2. Visualización de la PEC4	7
6. Anexo I (PEC2)	10
7. Anexo II (PEC3)	23

1. Título de la visualización y URL

Este trabajo ha sido desarrollado en diferentes fases y diferentes entregas. En la primera fase (PEC2) se hizo entrega de una historia titulada: **Peace Agreements (1990 – 2019)** que estaba compuesto por 2 dashboards:

- **Overview of Peace Agreements**
- **Peace Agreements references Women rights**

La URL de esa primera visualización era la siguiente:

https://public.tableau.com/profile/javier.fernandez.martinez#!/vizhome/PEC2_15875421707960/PeaceAgreements1990-2019?publish=yes

En cambio, para la entrega final, he decidido montar mi visualización en un único dashboard, para tener toda la información que quería visualizar de una forma más compacta, y que los filtros funcionasen como filtro único, sin llevar a confusión al usuario, disponiendo de la información necesaria en el mismo cuadro de mando. El título de la visualización se mantiene, y por tanto sigue siendo **Peace Agreements (1990 – 2019)**

La URL de esta visualización final es la siguiente:

https://public.tableau.com/profile/javier.fernandez.martinez#!/vizhome/PEC4_15916113480170/DashboardPAX?publish=yes

Repositorio Github:

Además de este documento, se adjuntan los fuentes de las visualizaciones (Tableau), el código en R de preprocesamiento de los datos y los datos originales y transformados de las PEC2 – PEC4 en el siguiente enlace Github: <https://github.com/javifermar/PAX-Visualizacion>

2. Descripción

Este proyecto consiste en la visualización de datos sobre los acuerdos de paz que se han dado desde el año 1990 hasta el año 2019, basado en los datos obtenidos de la base de datos “Peace Agreements Database” (Bell, Christine, Sanja Badanjak, Juline Beujouan, Tim Epple, Robert Forster, Astrid Jamar, Sean Molloy, Kevin McNicholl, Kathryn Nash, Jan Pospisil, Robert Wilson, and Laura Wise, 2020). En mi proyecto concreto, poniendo el foco especialmente en si en estos acuerdos hay o no menciones respecto a los derechos de los grupos más vulnerables como son los derechos de los niños, las mujeres, los refugiados, las etnias, los grupos religiosos, los LGBTI. También prestaremos especial atención a referencias que tengan que ver con “Derechos humanos”, “Igualdad”, “Democracia” y “Derechos civiles/políticos”. Se han separado 12 tipos de grupos a los que se hacen (o no) referencia y por otra parte se han separado otras 4 variables relativas a los derechos humanos e igualdad.

Gracias a este trabajo podremos comprobar y comparar qué y cuántos acuerdos incluyen referencias a estos aspectos, y podremos ver en qué zonas son más frecuentes, en qué espacio de tiempo y en qué tipo de conflictos. Esta visualización debería ayudarnos a encontrar esas relaciones.

Y por otra parte quiero también ver como inciden los diferentes acuerdos de paz en lo relativo a derechos humanos e igualdad, y concretamente enfocado concretamente en 4 variables que he comentado antes en la parte de los datos, y que son las relativas a referencias en los acuerdos sobre aspectos de: “Derechos humanos”, “Igualdad”, “Democracia” y “Derechos civiles/políticos”.

Para realizar esto, ha sido necesario acceder a una base de datos donde se encuentran estos acuerdos de paz, y sobre él hacer un preprocesamiento de los datos, realizar un análisis descriptivo, limpieza, transformación y enriquecimiento de los datos.

También quiero comentar que en la primera fase del proyecto (PEC2) y cuando aún no tenía demasiado claro sobre qué variables iba a focalizar la visualización, hice un intento de aproximación de cómo se podían visualizar diferentes grupos a los que los acuerdos hacían referencia. De hecho, en el primer dashboard de la PEC2 incluí referencias a la “Democracia”, “Igualdad”, “Niños”, “Refugiados” y “Raciales/Étnicos”. El tema es que solo era capaz de representarlos de forma individual, ya que esos valores los tenía en variables separadas. Además, le conservé los valores originales, de forma que cada una de esas variables tenía 4 posibles valores.

Además, en el dashboard 2 de esa primera entrega, hice unas visualizaciones basadas en los acuerdos que hacían referencia o no a los derechos de las mujeres, combinando diferentes gráficos.

3. ¿Qué se cumple y que no del proyecto? Valoración del esfuerzo y aprendizaje

Entiendo que en ese proyecto se cumple lo que había buscado desde la PEC3, que era el poder agrupar de alguna forma los diferentes grupos sociales. Había intentado sin éxito hacer algo similar en la PEC2 en la que solamente pude poner 5 grupos, y es que, en lugar de tratar los grupos como una única variable, las estaba tomando sueltas, por tanto, no las estaba pudiendo combinar.

Conseguir que los grupos estuviesen juntos en una misma variable, desde luego requería un reprocesamiento de los datos que ya habíamos preparado en la PEC2, a la que ya le habíamos hecho una serie de transformaciones y algunas operaciones para completar datos como por ejemplo los países, y ser capaz de desdoblar aquellos acuerdos en los que estuviese más de 1 país implicado. Ahora tocaba una nueva vuelta de tuerca, y era poder volcar los datos que quería visualizar de forma agrupada en una sola variable. La ventaja que nos proporcionaba esto era que podíamos desde un mismo cuadro hacer una comparativa por ejemplo de la presencia/ausencia de los diferentes grupos en función de las selecciones que tengamos en un momento de terminado. Pero ahora es muy sencillo ver si no aplico ningún filtro, que el grupo social al que más referencias hacen los acuerdos de paz, son sobre los refugiados/desplazados y, en segundo lugar, aunque con una cierta diferencia, a los derechos de las mujeres. Pero si por ejemplo seleccionamos los últimos 10 años (es decir desde 2010 hasta 2019 inclusive) y seleccionamos África como región, entonces vemos que el colectivo que más referencias tienen son los niños (un 35,53% de esos acuerdos tienen referencias a los derechos de los niños). Este reprocesamiento de datos podía hacerlo bien en el origen de los datos, o bien en la propia herramienta de visualización (Tableau) pivotando los datos. En mi caso, como ya había hecho un procesamiento

previo de datos, modifiqué un poco el código de R y añadí lo necesario para poder obtener estos datos también.

De la misma forma también aprendí que en la PEC2 utilicé un mapa, que nos geolocaliza los datos, y nos da una primera aproximación, pero también es cierto que es casi imposible decidir con un mapa qué países tienen más acuerdos cuando los colores entre ellos son similares, y además hay un factor importante, que es la superficie. Los países más grandes son más llamativos. Cuesta por ejemplo ver en el mapa que Filipinas es el país con más acuerdos de paz. Es cierto que para esa PEC2 yo obtuve esos valores con el preprocesamiento de datos que hice en R, pero gráficamente no es tan sencillo de ver que el ranking es Filipinas, Bosnia, Colombia, Sudán, etc. Si miramos el mapa, es posible que nos llame más la atención Colombia por la superficie que ocupa y porque contrasta mejor con los países a su alrededor. Por tanto, el mapa está bien como un vistazo y además suele ser llamativo, y también vale para filtrar (si se conoce la ubicación) o marcar una zona determinada, pero a la hora de comparar no es demasiado efectivo. De ahí que optase por poner un ranking de países, porque ahora es muy sencillo ver qué países tienen más acuerdos de paz de una forma muy intuitiva.

Una cosa que gané y que también perdí. Resulta que cuando hice la PEC2 en la zona de los filtros me decidí por poner un filtro de AÑO. Era un filtro sencillo, un deslizador, que me permitía o tener todos los años, o un año concreto (dentro del rango 1990-2019). Pero me gustaba la idea de poder seleccionar un año desde – año hasta, ya que nos permite jugar por ejemplo con la última década, o si queremos situar algún tipo de conflicto en el espacio tiempo, pues con este filtro lo teníamos arreglado. Entonces me he visto algún tutorial de cómo hacerlo, llegando inicialmente a probar una solución de 2 sliders independientes, pero no era demasiado intuitivo. Al final conseguí poner un único deslizador con deslizamiento desde ambos extremos, con lo cual podemos acotar rangos. ¿Pero qué he perdido? Pues lo que he perdido es ver como de una forma animada la evolución de las gráficas moviendo un único deslizador. En la PEC2 realmente moviendo el deslizador de año, vas viendo en el mapa las diferentes variaciones de conflictos/acuerdos que se van produciendo, y las regiones donde se producen (gracias a los colores). Ahora esa funcionalidad “animada” se pierde porque para mover un año, realmente tengo que mover 2 deslizadores (el desde y el hasta, y aunque están en el mismo objeto, no dejan de ser 2 movimientos, con lo cual se pierde esa “animación”).

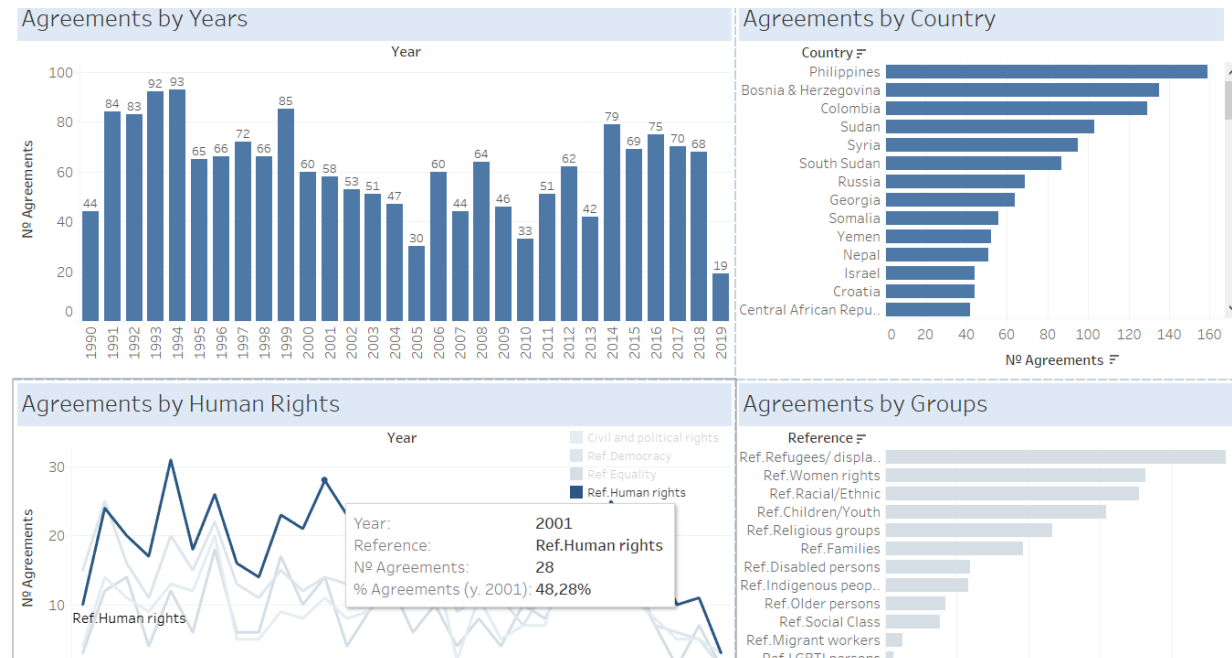
Todo este proyecto ha sido desarrollado con R en la parte de procesamiento inicial del dataset, y aunque la parte más grande de las transformaciones lo hice ya para la entrega de la PEC2, para esta PEC4 tuve que añadir el pivoteo de las columnas que quería destacar en mi visualización y ponerlas en una variable para que entre esos valores se puedan hacer visualizaciones/comparaciones en el mismo gráfico.

En la parte de visualización he usado Tableau. En la valoración del esfuerzo destaco que jamás había utilizado esta herramienta, ha sido mi primer contacto con ella, y la verdad es que la experiencia ha sido positiva. Además, hay muchos videos y documentación que ayuda para poder realizar algunas operaciones que de por sí no parecen triviales.

En esta última entrega también tenía que ver como combinar 2 datasets diferentes, ya que tenía los datos propios del dataset utilizado en la PEC2 y ahora incorporar el dataset de las referencias a los diferentes acuerdos por grupos y por algunas variables de derechos humanos. Y eso también me ha obligado a que en los gráficos tener que jugar con la aparición de valores nulos o de valores que no

procedían para ese gráfico, y hacer que los ocultara, pero sin que los descartara para de esa forma poder controlar bien los conteos y las proporciones.

Una de las cosas que no me fue trivial conseguir fue el obtener en la descripción emergente el gráfico de líneas, cuando quería referirme al porcentaje que supone sobre el total del año (y no sobre el total general). En el ejemplo se puede ver que el número de acuerdos que tienen referencia a los derechos



humanos en el año 2001, son 28 y eso supone el 48,28% de los acuerdos que hubo en el año 2001 (se puede ver en el gráfico de barras superior que el nº de acuerdos de 2001 son 58).

Me costó bastante poder llegar a obtener ese dato, no daba encontrado como relacionarlo con la dimensión para poder obtener la parte correspondiente al total del año.

He intentado utilizar gráficos que sean intuitivos, que sean fáciles de seguir y de entender su contenido, y además que tenga un aspecto amigable y agradable para el usuario. Espero haber conseguido ese objetivo.

4. Descripción técnica

Las herramientas que he utilizado en el desarrollo de este proyecto han sido:

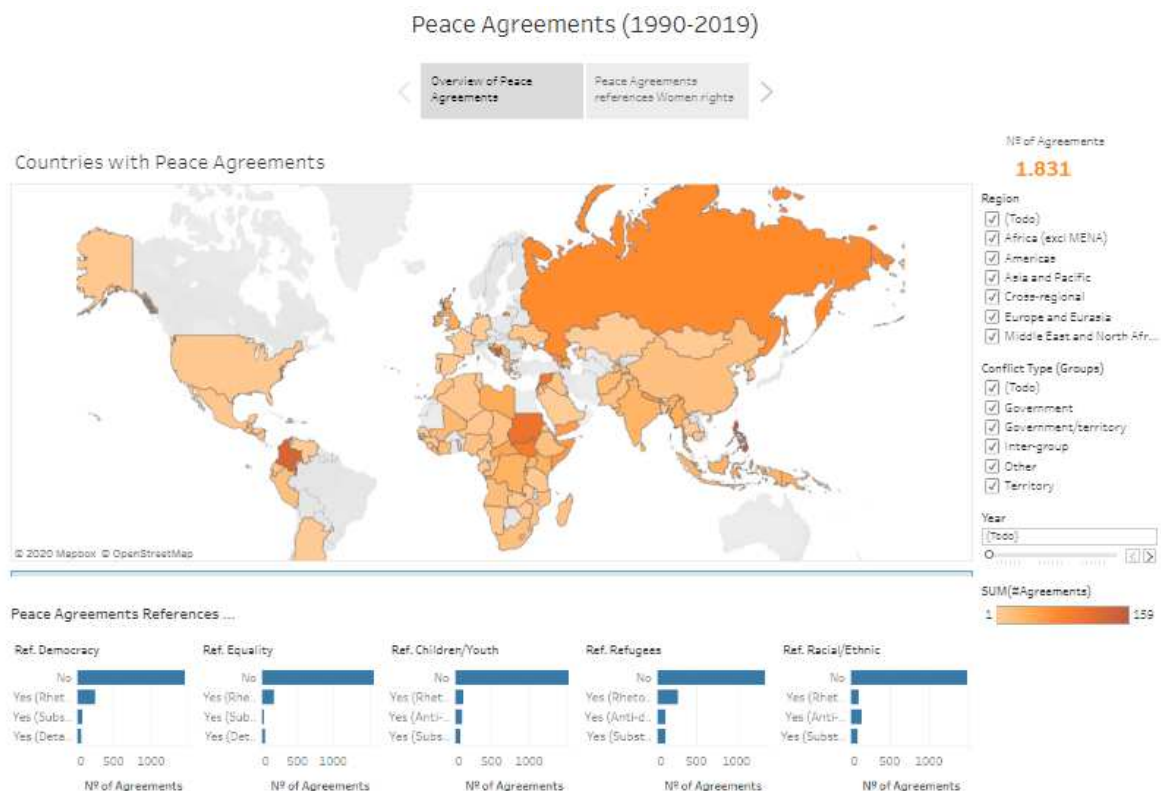
- **RStudio** para el preprocesamiento de los datos (tanto en la PEC2 como en la PEC4). Y respecto a las librerías utilizadas en R, son las siguientes:
 - *readxl*: Usada para poder leer el fichero Excel original con los datos.
 - *writexl*: Usada para poder escribir los ficheros Excel generados con las diferentes transformaciones realizadas. En la PEC2 era solamente 1 fichero, y en la PEC4 con 2 ficheros Excel generados.

- *dplyr*: Usada para manipulación y operaciones con data frames, pero una de las cosas más importantes es que permite la utilización de pipes (%>%).
 - *tidyr*: Para usar *gather()* que nos permite pivotar las columnas.
 - *ggplot2*: Usada para visualizar gráficos.
 - *data.table*: Usada para *data.table()* y para *tstrsplit()* para obtener los múltiples países separados por “/”.
 - *countrycode*: Para obtener los nombres de los países a través de los códigos ISO y los códigos ISO a través de los nombres que hemos separado.
 - *funModeling*: Para usar *df_status()* que nos da estadísticas de los datos.
- **Tableau** para las visualizaciones:
 - *Tableau Desktop*: Para la creación de las visualizaciones. Versión de Tableau Desktop Profesional Edition 2020.1.3 (20201.20.0427.1803) de 64 bits.
 - *Tableau Public*: Para la publicación de las visualizaciones en la web.

5. Visualización realizada

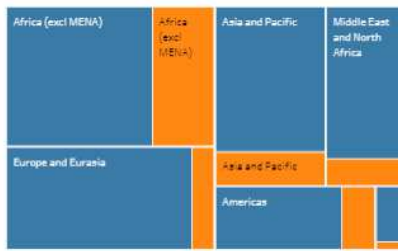
5.1. Visualización de la PEC2

En la PEC2 la visualización presentada fueron 2 dashboards que muestro a continuación

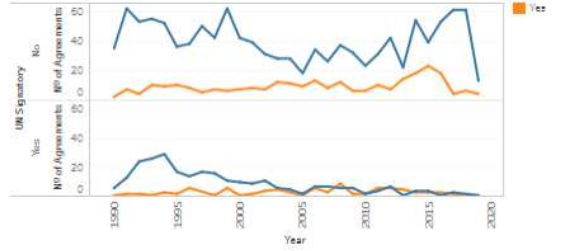


Peace Agreements by References Women Rights

Agreements by Regions



Agreements UN signed/unsigned by Years



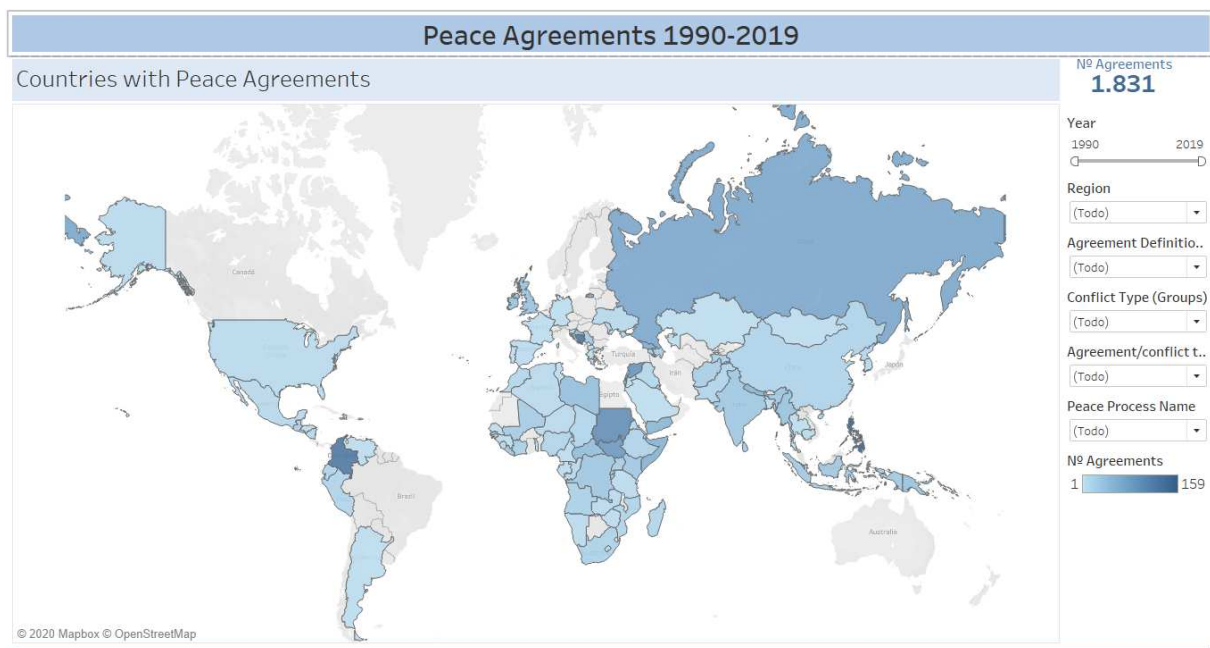
Agreements by Years

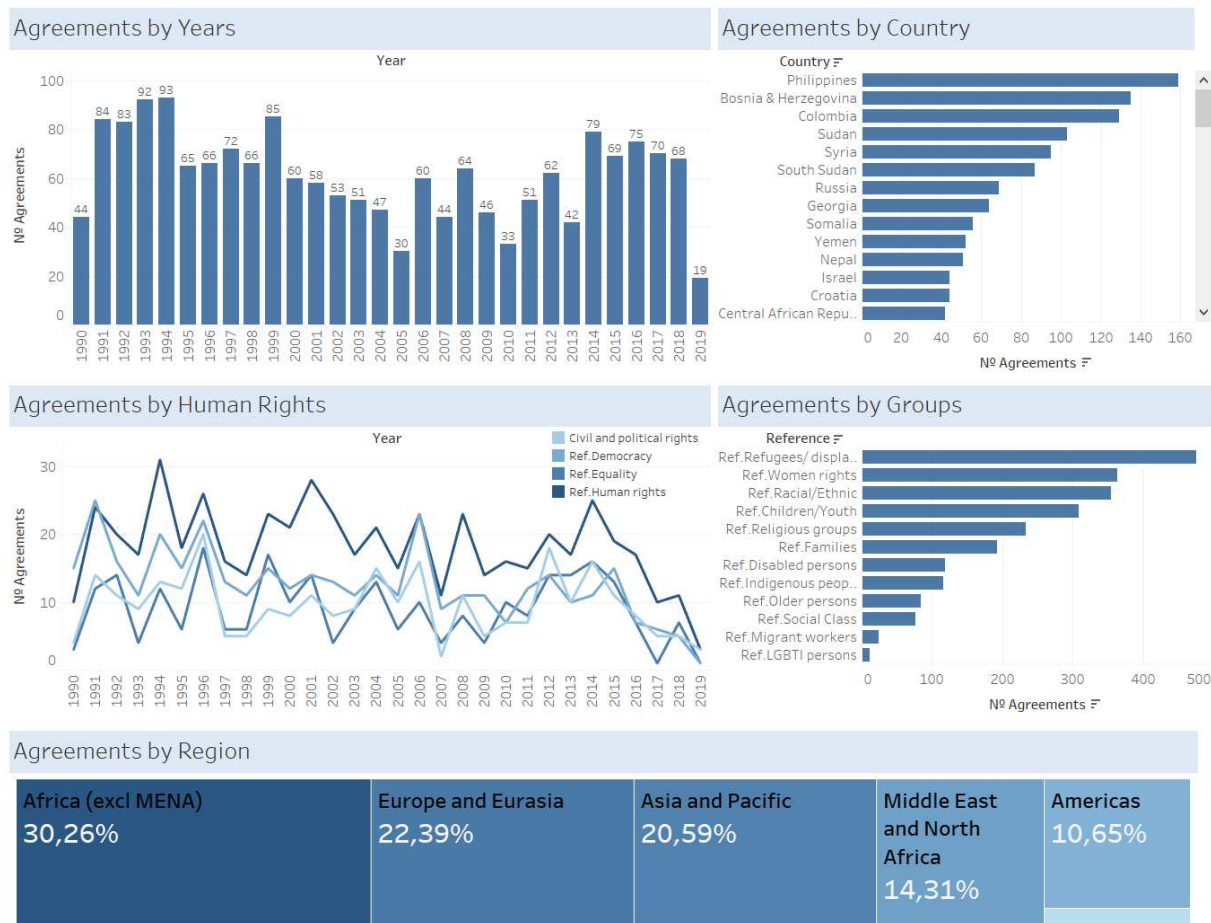


primer dashboard teníamos un mapa, con una serie de filtros (Región, Conflict Type y Year). Y luego teníamos 5 gráficos independientes (porque eran 5 variables independientes) para ver las referencias de los acuerdos a la democracia, a la igualdad, a los niños, a los refugiados y a las razas/étnias. No puse más porque como eran individuales, ocupaban bastante sitio, aunque el objetivo era poder agruparlas de alguna forma.

En el segundo dashboard, basado en la información referente a los acuerdos de paz, y si hacían o no referencia a los derechos de las mujeres. En este caso se implementaron 3 visualizaciones diferentes. Visualizando por regiones con un treemap, visualizando por años con gráficos de barras apiladas y visualizando con un gráfico de líneas teniendo en cuenta los años y si los acuerdos llevaban o no firma de las Naciones Unidas.

5.2. Visualización de la PEC4





Ahora en esta nueva PEC4 pasamos a tener una sola visualización. En ella todos los elementos interactúan con los otros, con lo cual digamos que los elementos están “conectados”. Seguimos viendo un total de los acuerdos en la parte superior derecha (donde pone 1.831 en la foto). Ahí siempre podremos ver los acuerdos que incluye la selección actual.

En esta nueva visualización seguimos manteniendo un mapa, pero hemos cambiado los filtros, permitiendo tener campos nuevos que antes no teníamos (ahora podemos meter más filtros al no estar desplegados por defecto, aunque son filtros de selección también) y en el año he incorporado que el slider sea tanto para el año desde como para el año hasta. Además, el filtro ahora no confunde tanto, ya que antes estaban todos los textos desplegados y marcados.

He incorporado un gráfico de barras para ver los acuerdos por año (en la PEC2 lo teníamos pero apilado teniendo en cuenta los derechos de las mujeres). Es un buen indicador de la evolución del número de acuerdos a lo largo de los años. Además, tenemos el número de acuerdos encima de cada barra. A la derecha he incorporado otro gráfico de barras (aunque en este caso horizontal) para intentar suplir aquello que comentaba en la PEC3 respecto a que con el mapa no es tan fácil hacer un ranking de los países que tienen más acuerdos. Aunque sí podemos ver diferentes tonalidades, el hecho de que los países con mayor superficie sean más llamativos, anula de alguna manera a los países más pequeños. Ahora de una forma muy clara podemos ver en todo momento el ranking de los países (bien con la selección actual o con otra).

Debajo de esto nos encontramos con las funcionalidades que ya quería incorporar desde la PEC2. Y lo hice de 2 formas diferentes. Un gráfico de líneas donde vemos cómo evolucionan los acuerdos en

función de los derechos humanos, y está puesto a propósito debajo del gráfico en el que veíamos el número de acuerdos por año. Ese gráfico además nos indica cuando nos pones encima cual es el porcentaje que supone respecto a los acuerdos totales del año y tiene el efecto “highlight” de destacar la serie cuando pasas por encima de ella.

A la derecha nos encontramos con un gráfico en el que por fin podemos combinar los diferentes grupos sociales de riesgo, sobre los que hacen referencia los diferentes acuerdos. De esta forma ahora podemos ver que colectivos están presentes en más acuerdos, y al tenerlo en un ranking podemos ver como fluctúan en función de esos datos. Por ejemplo, si elegimos el proceso de paz de Afganistán 2000, veremos que los derechos de las mujeres son precisamente los que más referencias tienen en dichos acuerdos, lo cual parece algo bastante evidente que así tuviese que ser en esas zonas de conflicto. Además, al tenerlo como un valor seleccionable, podemos hacer filtros por los diferentes grupos y como se distribuyen a nivel de países y en el tiempo.

Ya por último incluí una idea que comentaba también en la PEC3, que era un treemap de las regiones y poder ver de una forma muy simple qué porcentaje de acuerdos le corresponden a cada región.

6. Anexo I (PEC2)

Tabla de contenido

1. Título de las visualizaciones	11
2. Descripción breve	11
3. Los datos.....	12
4. Exploración de los datos.....	16
5. Procedimientos y herramientas	19
6. Visualización sobre los datos.....	20

1. Título de las visualizaciones

El título de la historia es: **Peace Agreements (1990 – 2019)**. Para esta PEC2 voy a presentar una historia de Tableau. Es una historia que está formada por 2 dashboards:

- **Overview of Peace Agreements**
- **Peace Agreements references Women rights**

El dashboard inicial (Overview of Peace Agreements) consta de las siguientes visualizaciones:

- Countries with Peaces Agreements
- Nº of Agreements
- Peaces References ... (ya que aparecen visualizaciones con diferentes referencias)

El segundo dashboard (Peace Agreements references Women rights) consta de las siguientes visualizaciones:

- Agreements by Regions
- Agreements UN signed/unsigned by Years
- Agreements by Years

La URL de la visualización es la siguiente:

https://public.tableau.com/profile/javier.fernandez.martinez#!/vizhome/PEC2_15875421707960/PeaceAgreements1990-2019?publish=yes

2. Descripción breve

Este documento está basado en los datos obtenidos de la base de datos “Peace Agreements Database” alojada en la web peaceagreements.org, y que corresponde a 1832 documentos sobre acuerdos de paz en todo el mundo. Esta base de datos ha sido creada y está mantenida por la Law School de la University of Edinburgh. Esta base de datos se puede descargar desde el siguiente enlace: https://www.peaceagreements.org/files/pax_all_agreements_data.xlsx

Con esta visualización, vamos a poder ver como se distribuyen los diferentes acuerdos de paz en todo el mundo, permitiéndonos hacer filtros por algunas características, y no solamente por países, como pueden ser las regiones o el tipo de conflicto originado, además en la primera visualización podemos ver algunas características como las referencias de dichos acuerdos de paz sobre niños, personas refugiadas, igualdad, democracia y etnias.

Ya en una segunda visualización está orientado a una referencia concreta, que es si los acuerdos de paz llevan o no referencias a los derechos de la mujer. Estas visualizaciones se pueden ver por regiones y por los diferentes años. Incluso si los acuerdos firmados por Naciones Unidas tienen un impacto más alto en reflejar estos derechos.

Con estas visualizaciones estamos de alguna forma dando importancia a alguna de las variables del dataset que se describen en el siguiente apartado.

3. Los datos

Como ya comenté previamente los datos que hemos utilizado para este trabajo proceden de la base de datos "Peace Agreements Database". Me he bajado el fichero Excel completo, sin filtros previos.

Los datos los he preprocesado con R. La definición de los campos y su significado lo he obtenido también en la propia web de [peaceagreements.org](https://www.peaceagreements.org), es un fichero pdf de 80 páginas, PEACE AGREEMENTS DATABASE AND DATASET V3- codebook - que se puede descargar en: <https://www.peaceagreements.org/files/PA-X%20codebook%20Version3.pdf>. Con este documento podemos entender el significado de cada una de las variables del dataset.

Al cargar los datos en R nos encontramos con que el dataset tiene 1832 registros, y que está compuesto por 265 columnas.

```
> #Número de filas
> nrow(dfpx)
[1] 1832
> #Número de columnas
> ncol(dfpx)
[1] 265
```

Para poder ver de qué tipo son las diferentes columnas podemos utilizar el comando `str()` o en mi caso he utilizado `df_status()` del paquete `funModeling`. Este comando nos va a decir de qué tipo son cada una de las variables, cuantos valores diferentes tiene cada una, cantidad/porcentaje de valores 0, cantidad/pct. de valores nulo y cantidad/pct de valores infinitos.

```
vstatus <- df_status(dfpx)
> table(vstatus$type)

character    numeric
         19         246
```

Nos encontramos inicialmente con que 246 variables son de tipo numérico, y solamente 19 son de tipo carácter. De todas formas de esas variables vamos a ver como son, ya que muchas veces las variables numéricas realmente son variables categóricas. Además vamos a comprobar si hay variables que son constantes y buscaremos variables binarias.

Encontramos 3 variables que son constantes:

```
> vstatus[vstatus$unique == 1,]
  variable q_zeros p_zeros q_na p_na q_inf p_inf  type unique
40  GAgeoth  1832    100    0    0    0    0 numeric      1
45  GMigoth  1832    100    0    0    0    0 numeric      1
172 HrNioth  1832    100    0    0    0    0 numeric      1
```

Encontramos 187 variables que son binarias, de valor 0 y 1

```
> binarias <- vstatus[vstatus$unique == 2 & vstatus$type == 'numeric','variable']
> max(dfpx[,binarias])
[1] 1
> min(dfpx[,binarias])
[1] 0
> length(binarias)
[1] 187
```

Por la documentación que tenemos sobre los campos, algunos de ellos son numéricos pero que corresponden con variables categóricas con valores 0, 1, 2, ó 3 para especificar el nivel de lo que representa la variable en los acuerdos de paz. El 0 indica que no se especifica y el resto de los valores indica en qué grado. Vamos a intentar localizar cuantas variables tenemos así. Y en este caso nos encontramos con 50 variables de ese tipo. Les llamo variables categ4val.

```
> categ4val <- vstatus[vstatus$unique == 4 & vstatus$type == 'numeric','variable']
> max(dfpx[,categ4val])
[1] 3
> min(dfpx[,categ4val])
[1] 0
> length(categ4val)
[1] 50
```

Después de hacer esto realmente nos quedan 25 variables que ni son las constantes, ni son binarias, ni son las de variables categ4val.

Y sobre esas 25 variables además hay algunas con una importante cantidad de valores nulos, y que en algún caso nos va a servir para decidir si es un valor a prescindir. De todas formas este análisis inicial de los datos, junto con el conocimiento de lo que representan, nos ayuda a tomar buenas decisiones. Por ejemplo podemos ver que hay una variable llamada Loc2ISO que tiene 1598 registros nulos (un 87.23% de los registros). Podríamos pensar que esta variable debería ser candidata a ser descartada, pero no es así. Esa variable indica el código de país del segundo país en caso de los acuerdos de paz por un conflicto entre 2 países, por tanto es normal que muchas veces este dato no esté informado (acuerdos sobre 1 solo país o de más de 2 países). Por tanto, si queremos tener los países, este dato cuando está presente, es útil.

```
> restovariables <- vstatus[-which(vstatus$variable %in% c(constantes, binarias, categ4val)),]
> restovariables
```

	variable	q_zeros	p_zeros	q_na	p_na	q_inf	p_inf	type	unique
1	Con	0	0	0	0.00	0	0	character	167
2	Contp	0	0	0	0.00	0	0	character	5
3	PP	0	0	0	0.00	0	0	numeric	156
4	PPName	0	0	0	0.00	0	0	character	156
5	Reg	0	0	0	0.00	0	0	character	6
6	AgtId	0	0	0	0.00	0	0	numeric	1832
7	Agt	0	0	0	0.00	0	0	character	1794
8	Dat	0	0	0	0.00	0	0	character	1503
9	Status	0	0	0	0.00	0	0	character	4
10	Lgt	0	0	0	0.00	0	0	numeric	70
11	N_characters	0	0	0	0.00	0	0	numeric	1693
12	Agtp	0	0	0	0.00	0	0	character	4
13	Stage	0	0	0	0.00	0	0	character	7
14	StageSub	0	0	23	1.26	0	0	character	21
15	Part	0	0	17	0.93	0	0	character	1721
16	ThrdPart	0	0	887	48.42	0	0	character	874
17	OthAgr	0	0	872	47.60	0	0	character	947
18	Loc1ISO	0	0	38	2.07	0	0	character	82
19	Loc2ISO	0	0	1598	87.23	0	0	character	30
20	Loc1GWNO	0	0	14	0.76	0	0	numeric	83
21	Loc2GWNO	0	0	1598	87.23	0	0	numeric	30
22	UcdpCon	0	0	29	1.58	0	0	character	109
23	UcdpAgr	0	0	1230	67.14	0	0	character	293
24	PamAgr	0	0	1479	80.73	0	0	character	34
25	Cowwar	0	0	443	24.18	0	0	character	64

El dataset contiene entre otros los siguientes campos :

"Con": Es el país o países donde se ha originado el conflicto.

"Contp": Tipo de conflicto que aborda el acuerdo. Puede ser Government (disputas ideológicas / políticas), Territory (disputas puramente territoriales), Government/territory, Inter-group Conflicts entre grupos no estatales, acuerdos con gob. Prov- revolucionarios y Other para el resto.

"PP_name": Nombre del proceso de paz.

"PP": Numero del proceso de paz.

"Reg": Región donde se produce el conflicto que aborda el acuerdo.

"AgtId": Identificador único para cada acuerdo. Este nos va a ser muy importante si luego vamos a crear registros para los países implicados en el conflicto.

"Agt": El nombre/descripción del acuerdo.

"Dat": Fecha en la que se firma el acuerdo. En formato AAAA-MM-DD

"Status": Definición del acuerdo y estado. Multiparty signed/agreed. Unilateral agreement. Status unclear. Agreement with Subsequent Status.

"Lgt": Nº de páginas del acuerdo.

"N_characters": Número de caracteres del texto del acuerdo.

"AgtP": Tipo de acuerdo/conflicto. Inter: Conflicto interestatal, InterIntra, Intra e IntraLocal.

"Stage": Fase del acuerdo con 7 valores posibles

"StageSub": Subcategorías de las fases anteriores.

"Part": Partes que firman el acuerdo.

"ThrdPart": Partes que firman el acuerdo como terceros, observadores, garantes o testigos.

"OthAgr": Nombres de cualquier otro acuerdo que mencione el documento en cuestión.

"Loc1ISO": Indica la ubicación primaria del conflicto intraestado y devuelve el código ISO del país.

"Loc2ISO": En conflictos interestatal, código ISO del país2.

"Loc1GWNO": Código Gleditsch-Ward del país indicado en Loc1ISO

"Loc2GWNO": Código Gleditsch-Ward del país indicado en Loc2ISO

"UcdpCon": Es un ID de conflicto (Uppsala Conflict Data Program Conflict ID) (a partir de 2017)

"UcdpAgr": Uppsala Conflict Data Program peace agreement ID.

"PamAgr": ID de acuerdo (Kroc Institute's Peace Accords Matrix agreement ID)

"CowWar": Numero de guerra. Como muchos acuerdos vienen de procesos de paz sobre conflictos

A partir de aquí vienen unas variables identificadas como "Groups" que se refieren a los diferentes grupos. Muchas de estas son las que antes "bauticé" como variables categ4val.

"GCh": Niños/jóvenes. Vale 0: Si el acuerdo de paz no incluye referencias a niños, 1: Si simplemente menciona a los niños, 2: El acuerdo menciona disposiciones referentes a niños, detalle/cláusula clara, 3: Las disposiciones sobre niños figuran, son sustanciales e indican compromiso en la implementación. Se generan subdivisiones binarias en caso de las menciones de los niños en los acuerdos como "GChRhet": Retórica, "GChAntid": antidiscriminación, "GChSubs": Sustantivo y "GChOth": Otro.

De la misma manera que aparece la variable "GCh" para los niños tenemos, "GDis" para personas discapacitadas, "GAge" para las personas mayores/ancianos, "GMig" para trabajadores migrantes, "GRa" para grupos raciales/étnicos, "GRe" para grupos religiosos, "GInd" si los acuerdos hacen referencia a personas indígenas, "GOth" para otros grupos, "GRef": para Refugiados / personas desplazadas y "GSoc" si se hace referencia a clases u origen social.

Del mismo modo se hacen referencias al género, "GeWom" es binaria que valdrá 1 si alguna de las disposiciones del acuerdo de paz abordan específicamente a las mujeres, su inclusión y su derechos. En caso contrario, 0. Lo mismo sobre las referencias al movimiento LGBTI, a las familias, etc.

Al igual que hay una categoría para los grupos, hay diferentes conceptos que se incluyen como la definición de estado, variables relativas al Gobierno, políticas/instituciones, elecciones, sociedad civil, administración pública, repartos de poderes, repartos de territorios, repartos económicos, compartición del poder militar entre los contendientes, derechos humanos e igualdad, democracia, medidas de protección, derechos humanos, civiles y políticos, derechos socio-económicos, instituciones de derechos humanos, menciones a medios de comunicación, ciudadanía, justicia, reconstrucción socio-económica, recursos naturales, negocios, impuestos, bancos, tierras y propiedades, patrimonio cultural, sector de seguridad, cese de fuego, fuerzas armadas, servicios de inteligencia, organizaciones criminales, terrorismo, justicia transicional e implementación.

Este es un resumen del contenido del dataset, aunque para detalles, es conveniente consultar al documento PDF mencionado anteriormente donde vienen explicadas cada una de las variables. Yo no voy a utilizar todas las variables, de hecho incluso me he quedado con unas cuantas, y tampoco las he utilizado todas en la visualización. He renombrado las columnas poniendo un nombre a las columnas que sea mucho más identificativa que las que trae el dataset original.

Además, he hecho algunas transformaciones y enriquecimiento de los datos. Respeto a las transformaciones, alguna simple como por ejemplo en las variables binarias 0-1 reemplazarlas por valores “No”-“Yes” aunque esto se podía hacer fácilmente en Tableau. También lo hice en las variables categ4val asignando las palabras correspondientes a lo que significa.

Una transformación ya algo más compleja que he realizado es que he desdoblado los registros en los que se indicaban Loc1ISO y Loc2ISO generando un registro “duplicado” pero cada uno con el código de país correspondiente. Pero además hay casos como el AgreementID 1807 cuyo “Conflict Originated” es “China/Mongolia/Russia”, ese registro no tiene nada en Loc1ISO, Loc2ISO. Lo que hice para estos casos es desdoblar esos registros separando por la “/” y obteniendo el código ISO de cada uno de los países. Para ese ejemplo concreto, tengo 3 registros (uno para China, otro para Mongolia y otro para Russia). Para esto he utilizado la función “countrycode” del paquete “countrycode” y que incluso en el código fuente que se adjunta con esta PEC2.

Otra de las labores de enriquecimiento, ha sido obtener el nombre de los países. Nosotros solamente teníamos los códigos ISO de los países. Con “countrycode” de R podemos obtener el nombre de los países. Por tanto mi nuevo dataset contiene el código ISO y el nombre del país. El dataset final contiene 2112 registros y 33 columnas. Exportado como “**pax_agreements.xlsx**”

```
> df_status(data_def)
  variable q_zeros p_zeros q_na p_na q_inf p_inf      type unique
1   Conflict originated      0      0      0      0      0      0 character    166
2   Conflict Type          0      0      0      0      0      0 character      5
3   Peace Process Number    0      0      0      0      0      0 numeric    155
4   Peace Process Name      0      0      0      0      0      0 character    155
5   Region                  0      0      0      0      0      0 character      6
6   Agreement ID            0      0      0      0      0      0 numeric   1831
7   Agreement Name          0      0      0      0      0      0 character   1793
8   Date Signed             0      0      0      0      0      0 character   1502
9 Agreement Definition and Status 0      0      0      0      0      0 character      4
10 Agreement/conflict type    0      0      0      0      0      0 character      4
11 Agreement stage           0      0      0      0      0      0 character      7
12 Country ISO               0      0      0      0      0      0 character    99
13 Ref.Children/Youth        0      0      0      0      0      0 character      4
14 Ref.Disabled persons      0      0      0      0      0      0 character      4
15 Ref.Older persons         0      0      0      0      0      0 character      4
16 Ref.Migrant workers       0      0      0      0      0      0 character      4
17 Ref.Racial/Ethnic         0      0      0      0      0      0 character      4
18 Ref.Religious groups      0      0      0      0      0      0 character      4
19 Ref.Indigenous people     0      0      0      0      0      0 character      4
20 Ref.Refugees/ displaced persons 0      0      0      0      0      0 character      4
21 Ref.Social Class          0      0      0      0      0      0 character      4
22 Ref.women rights          0      0      0      0      0      0 character      2
23 Ref.LGBTI persons         0      0      0      0      0      0 character      2
24 Ref.Families              0      0      0      0      0      0 character      2
25 Ref.Human rights          0      0      0      0      0      0 character      2
26 Ref.Equality              0      0      0      0      0      0 character      4
27 Ref.Democracy             0      0      0      0      0      0 character      4
28 civil and political rights 0      0      0      0      0      0 character      2
29 Socio-economic rights     0      0      0      0      0      0 character      2
30 Ceasefire provisions       0      0      0      0      0      0 character      2
31 UN Signatory              0      0      0      0      0      0 character      2
32 International Mission/Force 0      0      0      0      0      0 character      2
33 Country                   0      0      0      0      0      0 character    99

> nrow(data_def)
[1] 2112
> ncol(data_def)
[1] 33
```

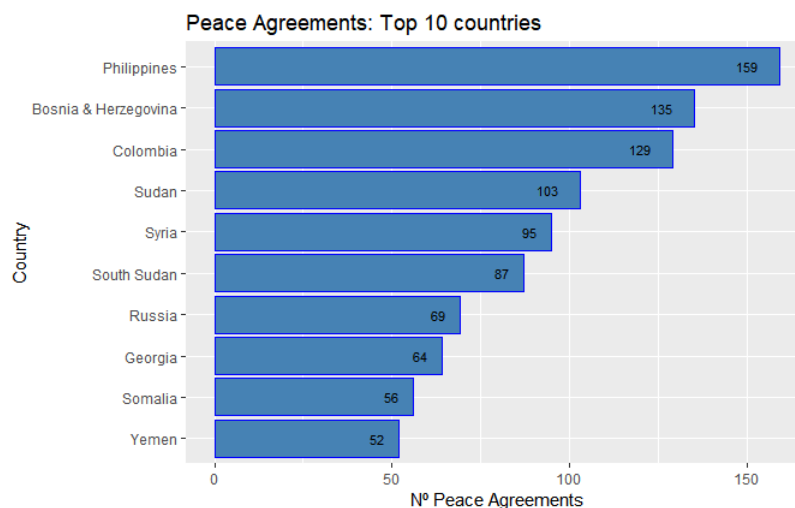

Como se puede ver después de la limpieza y tratamiento de los datos, ya no tengo valores nulos, los nombres de las variables ya son identificativas, es decir claramente podemos ver con qué variables me he quedado del conjunto original. Comentar que la fecha (Date Signed) no la he transformado, ya que esa transformación la he realizado en Tableau, y que las variables booleanas y cate4val que aparecen como character, deberían ser tipo “factor” pero como solamente quería hacer una exportación a Excel, pues han quedado así. Así a simple vista podemos ver que me he quedado prácticamente con todas las variables básicas (omitiendo algunas como el nº de páginas).

4. Exploración de los datos

De una forma simple podemos obtener el TOP 10 de países con mayor nº de acuerdos de paz. Esto sí que fue una completa sorpresa para mí, no me imaginaba ni mucho menos que Filipinas tuviese tantos acuerdos de paz firmados. Apostaba más por la zona del conflicto de Balcanes,

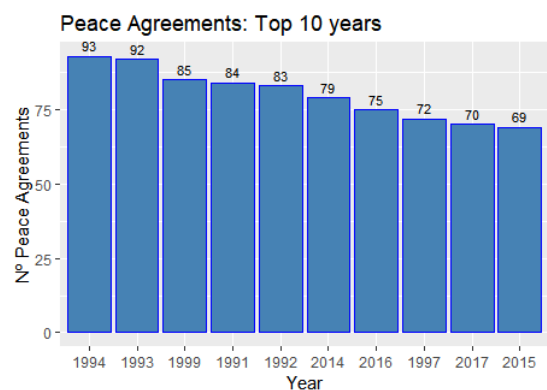
```
> head(sort(table(data_def$Country),decreasing = TRUE),10)
```

Philippines	Bosnia & Herzegovina	Colombia	Sudan	Syria
159	135	129	103	95
South Sudan	Russia	Georgia	Somalia	Yemen
87	69	64	56	52



También pudimos ver gracias a la imagen de antes (df_status previo a la exportación de datos en Excel), que hay 99 países implicados en este dataset.

Se pueden ver los 10 años con más acuerdos de paz

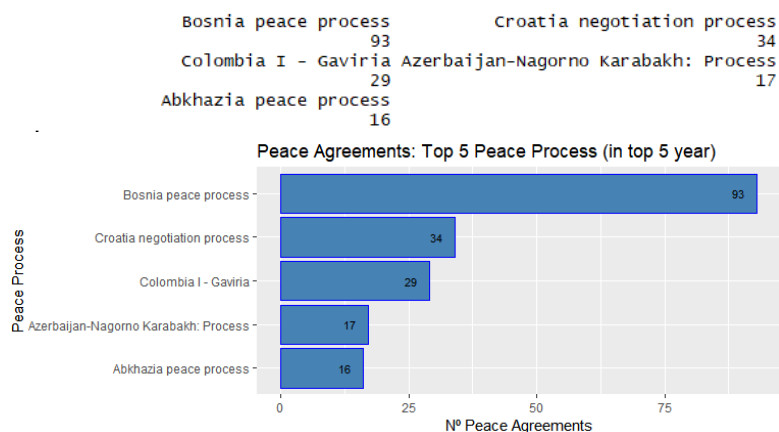


```
> head(sort(table(year(dfpx$Date)),decreasing = TRUE)
```

1994	1993	1999	1991	1992	2014	2016	1997	2017	2015
93	92	85	84	83	79	75	72	70	69

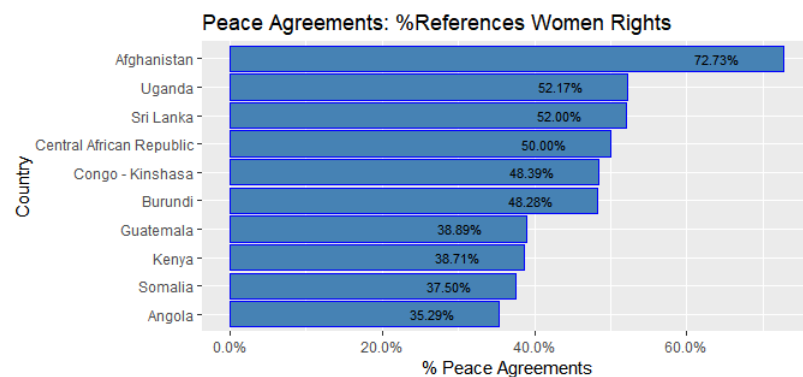
Podemos ver también de los 5 años TOP cuales son los acuerdos más repetidos, es decir cual han sido los procesos de paz más numerosos en esos años.

```
> top_5_year <- as.integer(names(head(sort(table(year(dfpx$Date)),decreasing = TRUE),5)))
> head(sort(table(dfpx[year(dfpx$Date) %in% c(top_5_year),'PName']), decreasing = TRUE),5)
```



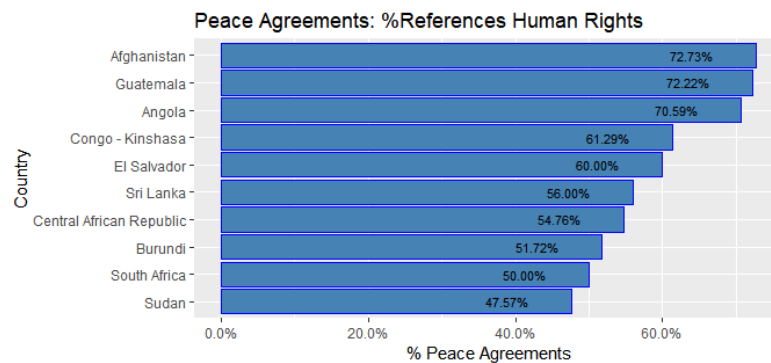
```
> tab_women <- table(data_def$`Ref.women rights`=="Yes", data_def$Country)
> nombres <- colnames(tab_women)
> tab_women <- rbind(tab_women, tab_women[1,]+tab_women[2,])
> tab_women <- rbind(tab_women, tab_women[2,]/tab_women[3,])
> #Países con más de 15 acuerdos, ordenados por porcentaje que incluyen referencias a derechos mujer
> head(sort(tab_women[4,which(tab_women[3,]>=15)],decreasing = TRUE),10)
```

Country	% Peace Agreements
Afghanistan	0.7272727
Uganda	0.5217391
Sri Lanka	0.5200000
Central African Republic	0.5000000
Congo - Kinshasa	0.4838710
Burundi	0.4827586
Guatemala	0.3888889
Kenya	0.3870968
Somalia	0.3750000
Angola	0.3529412



```
> tab_hr <- table(data_def$`Ref.Human rights`=="Yes", data_def$Country)
> tab_hr <- rbind(tab_hr, tab_hr[1,]+tab_hr[2,])
> tab_hr <- rbind(tab_hr, tab_hr[2,]/tab_hr[3,])
> #Países con más de 15 acuerdos, ordenados por porcentaje que incluyen referencias a derechos mujer
> head(sort(tab_hr[4,which(tab_hr[3,]>=15)],decreasing = TRUE),10)
```

Country	% Peace Agreements
Afghanistan	0.7272727
Guatemala	0.7222222
Angola	0.7058824
Congo - Kinshasa	0.6129032
El Salvador	0.6000000
Sri Lanka	0.5600000
Central African Republic	0.5476190
Burundi	0.5172414
South Africa	0.5000000
Sudan	0.4757282



Y aunque esto no es concluyente, ya que estamos utilizando una variable ordinal (que va de 0 a 3 para especificar el nivel de pertenecía de cláusulas y compromisos con refugiados) he hecho una aproximación (ya digo que esto no es exacto) pero nos da una idea de que países son los que tienen acuerdos que incluyen más menciones al tema de los refugiados. Al final es como ir sumando las “puntuaciones” para obtener una primera aproximación.

```
> head(ref_country[order(ref_country$sum_refugees,decreasing = TRUE),],10)
```

	Country	sum_refugees
9	Bosnia & Herzegovina	75
85	Sudan	73
29	Georgia	63
82	South Sudan	60
69	Philippines	43
13	central African Republic	31
19	Congo - Kinshasa	31
71	Russia	28
41	Israel	25
21	Croatia	23

Lo mismo con el tema de las etnias

```
> head(rac_country[order(rac_country$sum_racial,decreasing = TRUE),],10)
```

	Country	sum_racial
9	Bosnia & Herzegovina	82
85	Sudan	61
82	South Sudan	44
60	Nepal	39
10	Burundi	31
13	central African Republic	26
40	Ireland	25
79	Somalia	25
84	Sri Lanka	25
94	United Kingdom	25

Y lo mismo con el tema de referencias a la democracia

```
> head(dem_country[order(dem_country$sum_democracy,decreasing = TRUE),],10)
```

	Country	sum_democracy
16	Colombia	62
60	Nepal	50
85	Sudan	44
82	South Sudan	29
9	Bosnia & Herzegovina	26
32	Guatemala	23
48	Liberia	23
10	Burundi	22
13	central African Republic	22
19	Congo - Kinshasa	19

Resumiendo un poco la exploración de los datos que he realizado se pueden observar cosas como:

- Filipinas es el país con un mayor nº de acuerdos de paz
- Los años donde mayor nº de acuerdos de paz se produjeron fueron los que transcurrieron después del inicio de la guerra de los Balcanes (que fue en 1991). Entre 1992-1994 se produce la guerra Croacia-Bosnia. Y ese en esos años donde aparece muy destacado el proceso de paz “Bosnia peace process”.

- Respecto al punto anterior vemos que los años 1991, 1992, 1993 y 1994 están en el TOP 5 de los años con mayor número de procesos de paz, siendo 1994 el de mayor número (93).
- Respecto a las referencias en los acuerdos de paz sobre los derechos de las mujeres, lo hice respecto al porcentaje de veces que aparecen referencias a esos derechos sobre el total. Y además lo hice sobre aquellos que tuviesen al menos 15 acuerdos, porque salía algún caso de 1 país con 1 único acuerdo, y tenía referencias a derechos de las mujeres por tanto se cumplía en un 100% y no era lo que quería descubrir. Una vez establecido ese criterio, podemos observar que claramente Afganistán es el país más destacado en esa faceta. Y precisamente es uno de los sitios donde las mujeres más tienen que pelear por sus derechos, ya que es uno de los países donde las mujeres están más relegadas precisamente por el hecho de ser mujer. Y le sigue Uganda, otro país donde las mujeres aún mucho tienen que pelear para poder tener una igualdad. En el top 10, hay 8 países africanos, sinceramente nada nuevo, lamentablemente era algo que se podía esperar.
- Del mismo modo que hice con los derechos de las mujeres, lo hice con los derechos humanos, y salió nuevamente Afganistán como nº1, seguido muy de cerca por Guatemala. Debemos tener en cuenta, que Guatemala sufrió una larguísima guerra civil, y entiendo que todo lo que derivó de esa guerra, tuvo que ver en todos esos acuerdos de paz, y en establecer cláusulas para refugiados, derechos humanos, etc.
- Y por último quiero comentar que hice una especie de aproximación, que simplemente es eso, una suma simple de los valores 0, 1, 2 y 3 en algunos aspectos que define el dataset, como por ejemplo las referencias a las razas/etnias. Según las especificaciones del dataset, un 0 significa que no se establece nada sobre el tema. Y un 3 significa que se establecen cláusulas y compromisos. El número indica nivel de compromiso en cuanto al tema en cuestión. Pues vemos claramente que Bosnia lidera esa “clasificación” ya que ahí hubo problemas importantes de raza y por tanto entiendo que los acuerdos de paz necesitaban incluir cláusulas para garantías de las personas. Le siguen países africanos, como Sudán que mantiene un conflicto étnico con Sudán del Sur, donde hay una lucha de las tribus mayoritarias.

5. Procedimientos y herramientas

Las herramientas que he utilizado para esta PEC son: R y Tableau. Como se puede ver en el apartado 4 he estado haciendo algunas comprobaciones y consultas de algunas preguntas, como por ejemplo en que período de tiempo se produjeron más acuerdos de paz, e intentar enlazar esta información con la historia. Por ejemplo, también se pudo comprobar en cuales se incluía de forma más frecuente algunas cláusulas, como derechos humanos, temas vinculados a la democracia precisamente en países donde hay mucho conflicto, temas también de derechos de las mujeres precisamente y que precisamente sale sobre todo en sitios donde apenas se les reconocen, es decir que seguramente hay que hacer mucho trabajo y mucho hincapié en esos asuntos, etc.

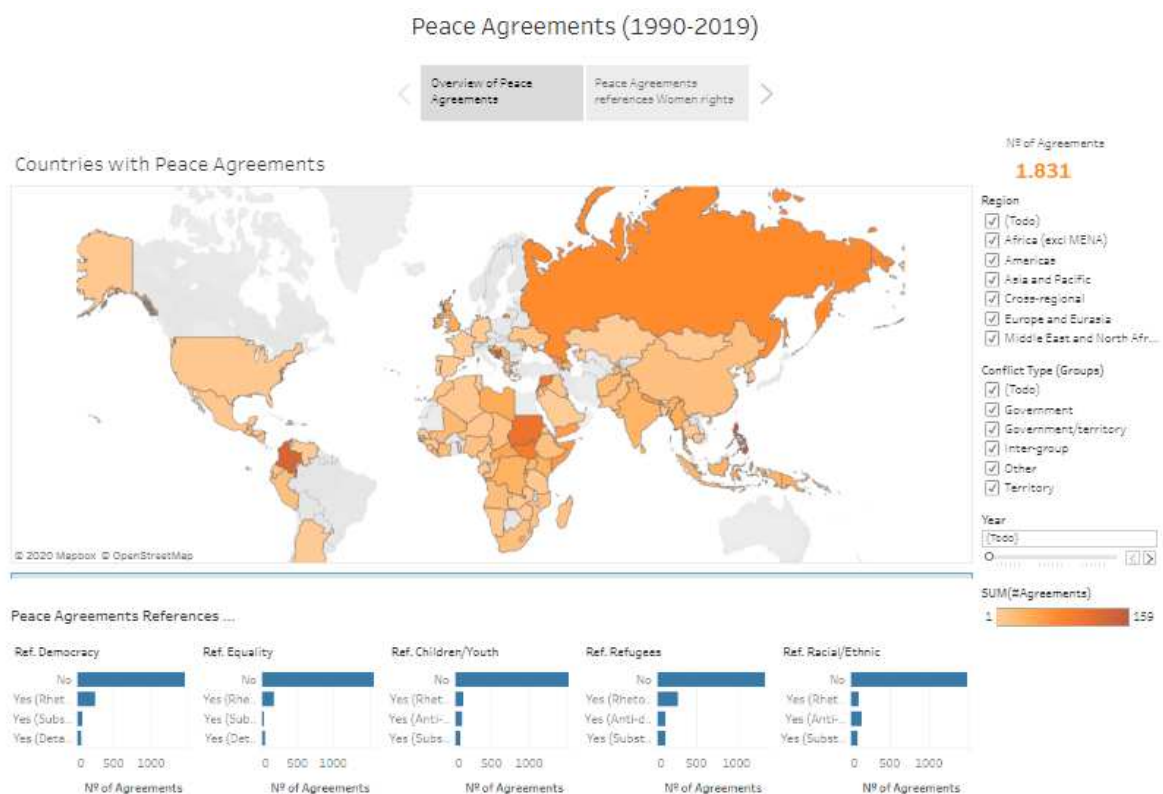
Debido a esto orienté un poco esta primera visualización en poder localizar geográficamente los países de una forma sencilla (a través de un mapa) y algunos filtros que nos ayuden a poder ver de una forma más simple en que zonas se han producido mayor cantidad de acuerdos de paz.

Hace unos cuantos años he trabajado con Qlikview, aunque realmente hice bastante más en la parte de integración (scripting de Qlik) que lo que son cuadros de mando. Sí que hice algunas cosas, pero ya hace bastante tiempo. He tocado muy poquito Qlik Sense, he hecho algunos gráficos sencillos y vi muy de pasada Cognos aunque no he llegado a implementar nada con él. Después como librerías gráficas he utilizado y utilizo básicamente en asignaturas de este máster las librerías ggplot de R, y matplotlib/seaborn de Python.

Tengo previsto conocer Tableau, de hecho es la primera vez que hago algo con él, por supuesto lo conocía de ver un montón de visualizaciones que circulan por ahí en publicaciones y que ves que tienen el logotipo de Tableau. Y por puesto también tengo previsto, y además es un objetivo de esta asignatura, aprender D3.js que tampoco conozco.

6. Visualización sobre los datos

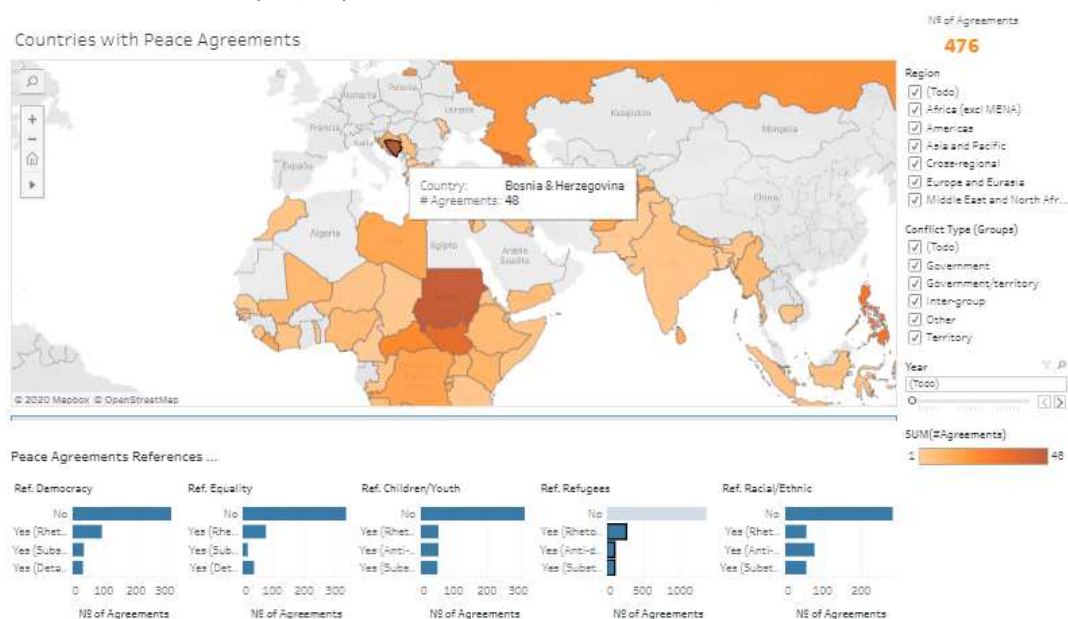
Tal y como comentaba en el primer punto, hay 2 dashboards



En este primer dashboard podemos ver un mapa mundial con y donde vemos reflejado con la intensidad de color de la superficie según el nº de acuerdos de paz que involucre a cada uno de los países. Si pasamos por encima de cada país, nos dice el nombre y el nº de acuerdos. Tenemos disponible unos filtros de región, de tipo de conflicto y un slider de año. De un vistazo rápido podemos ver como Filipinas, Bosnia, Colombia, Siria, etc son los principales países en cuanto a cantidad de acuerdos.

Podemos también ver siempre el indicador nº de acuerdos, que tiene en cuenta la selección actual, y que depende de todos los filtros que tiene este dashboard.

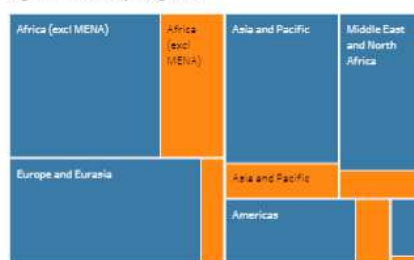
Debajo he puesto 5 indicadores sobre la referencia a la democracia, igualdad, niños, refugiados y de etnia. Respecto a esto he dejado una foto abajo podemos ver como si marcamos en la variable “Ref.Refugees” nos devuelve los países donde existen más acuerdos y que hacen referencia al tema de refugiados/desplazados y coincide con aquella aproximación que había hecho en el punto 4, donde veíamos que respecto a este tema nos salían Bosnia, Sudán, Georgia, y efectivamente eso es lo que también vemos en el mapa (los países con un tono casi marrón).



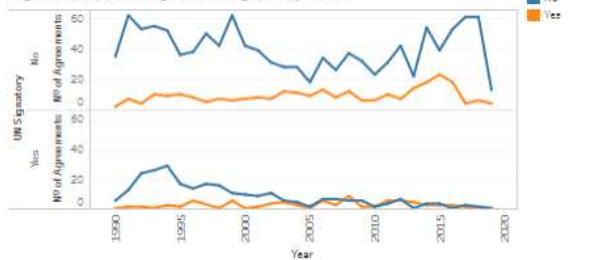
En el segundo dashboard, que se basa en información referente a si los acuerdos incluyen o no cláusulas específicas sobre derechos e igualdad de la mujer.

Peace Agreements by References Women Rights

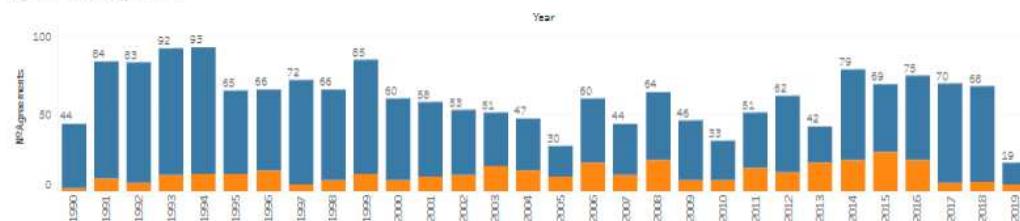
Agreements by Regions



Agreements UN signed/unsigned by Years



Agreements by Years



Aquí podemos ver que aunque en el punto 4 vimos como claramente Afganistán era donde más acuerdos hacían referencia a derechos de la mujer, cuando lo llevamos a regiones, en el cómputo se puede ver como África es quien tiene más. Y por el contrario Europa se destaca precisamente por lo contrario. Quiero pensar que tiene relación en que en Europa la mujer tiene una posición diferente y

por tanto no se hace tan fundamental como en otras zonas como Afganistán o África en general donde están a años luz de los hombres en cuanto a derechos.

También parece haber una tendencia de que aquellos acuerdos en los que la ONU forma parte de las firmas del tratado de paz tienen más referencias porcentuales que cuando no está la ONU como parte firmante.

7. Anexo II (PEC3)

Tabla de contenido

1. Introducción/Comunicación	24
1.1. Abstract	24
1.2. Introducción	24
2. Bloque datos.....	25
2.1. Descripción del dataset	25
2.2. Proceso de trabajo	28
2.3. Relaciones que vamos a realizar	32
2.4. Descripción de la gráfica elegida	33
2.5. Descripción de la gobernanza de datos	35
3. Bloque diseño	36
3.1. Elección de colores.....	36
3.2. Tipografía.....	37
3.3. Plantilla	37
3.4. Formato	38
Bibliografía.....	39

1. Introducción/Comunicación

1.1. Abstract

Este proyecto consiste en la visualización de datos sobre los acuerdos de paz que se han dado desde el año 1990 hasta el año 2019, poniendo el foco especialmente en si en estos acuerdos hay o no menciones respecto a los derechos de los grupos más vulnerables como son los derechos de los niños, las mujeres, los refugiados, las etnias, los grupos religiosos, los LGBTI. También prestaremos especial atención a referencias que tengan que ver con “Derechos humanos”, “Igualdad”, “Democracia” y “Derechos civiles/políticos”.

Gracias a este trabajo podremos comprobar y comparar qué y cuantos acuerdos incluyen referencias a estos aspectos, y podremos ver en qué zonas son más frecuentes, en qué espacio de tiempo y en qué tipo de conflictos. Esta visualización debería ayudarnos a encontrar esas relaciones.

Y por otra parte quiero también ver como inciden los diferentes acuerdos de paz en lo relativo a derechos humanos e igualdad, y concretamente enfocado concretamente en 4 variables que he comentado antes en la parte de los datos, y que son las relativas a referencias en los acuerdos sobre aspectos de: “Derechos humanos”, “Igualdad”, “Democracia” y “Derechos civiles/políticos”.

Para realizar esto, ha sido necesario acceder a una base de datos donde se encuentran estos acuerdos de paz, y sobre él hacer un preprocesamiento de los datos, realizar un análisis descriptivo, limpieza, transformación y enriquecimiento de los datos.

Una vez realizado los procesos anteriores, crear una visualización que sea capaz de dar soporte al objetivo planteado. Se adjunta un mockup como propuesta de visualización.

1.2. Introducción

Este proyecto se ha dividido en 3 fases (que son tres entregas correspondientes a las PEC2, PEC3 y PEC4) y cuyos objetivos/contenidos son:

- PEC2: Familiarización de los datos.
- PEC3: El informe de proyecto.
- PEC4: Creación de la visualización final.

En este documento vamos a abordar lo concerniente a la PEC3, es decir el informe de proyecto donde vamos a describir las partes del proyecto, indicando como se hará cada elemento.

Como ya habíamos anticipado en el documento entregado de la PEC2, este proyecto está basado en los datos obtenidos de la base de datos “Peace Agreements Database” (Bell, Christine, Sanja Badanjak, Juline Beujouan, Tim Epple, Robert Forster, Astrid Jamar, Sean Molloy, Kevin McNicholl, Kathryn Nash, Jan Pospisil, Robert Wilson, and Laura Wise, 2020) alojada en la web peaceagreements.org, y que corresponde a 1832 documentos sobre acuerdos de paz en todo el mundo. Esta base de datos ha sido creada y está mantenida por la Law School de la University of Edinburgh. Estos documentos recogen los acuerdos de paz que van desde el año 1990 hasta el año 2019. Esta base de datos la podemos

descargar completa o con filtros previos. El enlace para descargarla completa es: https://www.peaceagreements.org/files/pax_all_agreements_data.xlsx

En la entrega anterior PEC2, hicimos una exploración global, probando diferentes visualizaciones y ahora nos vamos a poder centrar más en algunas características concretas de los mismos, prestando mayor atención a algunas variables que vamos a destacar.

Con estas visualizaciones estamos de alguna forma dando importancia a alguna de las variables del dataset que se describen en el siguiente apartado. Gracias a ello podemos enfocar por ejemplo las preguntas siguientes:

- De los grupos vulnerables (que describiremos a continuación) ¿Cuáles aparecen más en los acuerdos de paz? ¿Cuáles aparecen menos? ¿Parece haber relación entre ellos?
- ¿En qué zonas son más frecuentes unos grupos que otros? ¿Los grupos más minoritarios, en qué país/región se dan?
- En relación a los derechos humanos e igualdad, ¿tienen relación en función a las regiones y/o tipos de conflicto? ¿Cómo ha sido su evolución a lo largo de los años?

2. Bloque datos

2.1. Descripción del dataset

Como ya comenté previamente los datos que hemos utilizado para este trabajo proceden de la base de datos “Peace Agreements Database”. Me he bajado el fichero Excel completo, sin filtros previos.

Los datos fueron preprocesados con R (cargando el dataset completo en una variable que he llamado dfpax) y referente al dataset hemos obtenido lo siguiente:

- N.º de registros: Se han obtenido 1832, ejecutando `nrow(dfpax)`

```
#Número de filas  
nrow(dfpax)
```

```
## [1] 1832
```

- N.º de columnas: Se han obtenido 265 columnas, con `ncol(dfpax)`

```
#Número de columnas  
ncol(dfpax)
```

```
## [1] 265
```

- Tipos de datos de las columnas: Se puede hacer con `str()` o también con `df_status()` y hemos obtenido que 19 columnas son de tipo carácter (es decir variables categóricas) y 246 columnas son de tipo numérico.

```
#str(df paz)
#summary(df paz)
vstatus <- df_status(df paz, print_results = FALSE)
table(vstatus$type)
```

```
##
## character    numeric
##           19      246
```

Haciendo procesamiento de esos datos, y tal como se ha adjuntado en la documentación de la PEC2, hemos visto que hay **3** variables constantes, **187** variables binarias y **50** ordinales variables que he calificado como categ4val (porque van de 0 a 3). El resto de las variables, son **25**. De esas 25 variables que no son binarias, ni constantes ni ordinales, están las 19 variables categóricas (y 6 variables numéricas (AgtId que es el ID del acuerdo, Lgt el tamaño en páginas del acuerdo, N_characters que es el número de caracteres del texto del acuerdo, PP que el número del proceso de paz, y las variables Loc1GWNO y Loc2GWNO que son los códigos Gleditsch-Ward de los 2 países implicados si es el caso).

```
> restovvariables$variable[which(restovvariables$type=='numeric')]
[1] "PP" "AgtId" "Lgt" "N_characters" "Loc1GWNO" "Loc2GWNO"
> restovvariables$variable[which(restovvariables$type=='character')]
[1] "Con" "Contp" "PPName" "Reg" "Agt" "Dat" "Status" "Agtp" "Stage" "StageSub"
[11] "Part" "ThrdPart" "OthAgr" "Loc1ISO" "Loc2ISO" "UcdpCon" "UcdpAgr" "PamAgr" "Cowwar"
```

La definición de los campos y su significado lo he obtenido también en la propia web de peaceagreements.org, es un fichero pdf de 80 páginas, PEACE AGREEMENTS DATABASE AND DATASET V3- (ref. bibliográfica #3) Con este documento vamos a poder entender el significado de cada una de las variables del dataset.

Para esta parte voy a nombrar aquellas variables que vamos a utilizar:

- **“Con” (Country):** Es el país o países donde se ha originado el conflicto. Si hay varios se separa con el carácter /
- **“Contp” (Conflict Type):** Tipo de conflicto que aborda el acuerdo. Puede ser Government (disputas ideológicas / políticas), Territory (disputas puramente territoriales), Government/territory, Inter-group Conflictos entre grupos no estatales, acuerdos con gob. Prov- revolucionarios y Other para el resto.
- **“PPName” (Peace Process Name):** Nombre del proceso de paz.
- **“Reg” (Region):** Región donde se produce el conflicto que aborda el acuerdo. Los valores posibles son:
 - *Africa (excl. MENA)*
 - *Americas*
 - *Asia and Pacific*
 - *Europe and Eurasia*
 - *Middle East and North Africa*
 - *Cross-regional*
- **“AgtId” (Agreement ID):** Identificador único para cada acuerdo.
- **“Dat” (Date signed):** Fecha en la que se firma el acuerdo.
- **“Status” (Agreement Definition and Status):** Definición del acuerdo y estado de este.
 - *Multiparty signed/agreed:* Acuerdo firmado/aceptado por más de un grupo de los que tienen el conflicto.

- *Unilateral agreement*: Acuerdo de "una parte" en respuesta al acuerdo con la otra parte.
- *Status unclear*: Sin aclarar qué parte del acuerdo se ha firmado o quien lo ha acordado, pero hay indicios de alguna firma
- *Agreement with Subsequent Status*: Se propone un acuerdo y que, aunque no se acepta como un acuerdo entre las partes, es la base para avances posteriores
- **"Agreement Conflict Type"**: Tipo de acuerdo/conflicto. Describe la naturaleza de este.
 - *Inter*: Conflicto interestatal
 - *InterIntra*: Conflicto interestatal/mixto o intraestatal. Son acuerdos interestatales, pero se refieren a conflictos intraestatales porque la mayoría de los componentes del acuerdo/conflicto se producen dentro del estado.
 - *Intra*: Conflictos intraestatales, dentro de las fronteras.
 - *IntraLocal*: Conflictos dentro del estado, pero para resolver problemas locales.
- **"Loc1ISO"**: Indica la ubicación primaria del conflicto intraestado y devuelve el código ISO del país. Para los conflictos interestatales indica una de las 2 ubicaciones principales aleatoriamente (ubic.1 o ubic.2). Si hay muchas ubicaciones el campo se deja vacío.
- **"Loc2ISO"**: Solo se usa en casos de conflictos interestatal, e indica una de las ubicaciones principales, y devuelve el código ISO del país. Al igual que antes se escoge aleatoriamente (ubic.1 o ubic2).

Y a partir de aquí luego ya las variables de grupo, algunas de naturaleza ordinal (que se explicará más adelante que serán transformadas a variables binarias)

- **"GCh"**: Indicador si se incluyen referencias a niños/jóvenes. Es una variable ordinal que va de 0 a 3 y cuyo significado es el siguiente:
 - 0: Si el acuerdo no incluye referencias a niños/jóvenes.
 - 1: Simplemente menciona a los niños/jóvenes.
 - 2: El acuerdo menciona disposiciones referentes a niños/jóvenes, está detallado o incluyen una cláusula clara.
 - 3: Las disposiciones sobre niños/jóvenes figuran y son sustanciales e indican compromiso en la implementación.
- **"GDis"**: Referencias a personas con discapacidades. Ordinal, mismo criterio que el anterior.
- **"GAge"**: Referencias a personas mayores/ancianos. Ordinal, mismo criterio que el anterior.
- **"GMig"**: Trabajadores migrantes. Ordinal, mismo criterio que el anterior.
- **"GRa"**: Grupos raciales/étnicos. Ordinal, mismo criterio que el anterior.
- **"GRe"**: Referencias a grupos religiosos. Ordinal, mismo criterio que el anterior.
- **"GInd"**: Referencias a personas indígenas. Ordinal, mismo criterio que el anterior.
- **"GRef"**: Referencias a refugiados y/o personas desplazadas. Ordinal, mismo criterio anterior.
- **"GSoc"**: Referencias a clases u origen sociales. Ordinal, mismo criterio que el anterior.
- **"EqGen"**: Referencias a compromisos de igualdad, no discriminación, etc. Ordinal.
- **"HrDem"**: Referencias a compromisos con la democracia. Ordinal, mismo criterio anterior.
- **"GeWom"**: Variable binaria que valdrá 1 si alguna de las disposiciones del acuerdo de paz aborda específicamente a las mujeres, su inclusión y sus derechos.
- **"GeLgbti"**: Variable binaria para indicar si hay o no referencias a colectivos/grupos LGBTI.
- **"GeFa"**: Variable binaria para indicar si se incluyen disposiciones relacionadas con las familias.

- **“HrGen”**: Variable binaria para indicar si se incluyen referencias y compromisos con los derechos humanos.
- **“HrCp”**: Variable binaria para indicar si los acuerdos de paz incluyen alguna mención a derechos civiles y políticos.

2.2. Proceso de trabajo

Como he comentado en el apartado anterior, los datos fueron tratados inicialmente con R, y que una vez hemos cargado el dataset nos encontramos con:

- 1832 registros y 265 columnas de las cuales 19 son categóricas y 246 numéricas.
- También hemos visto que 187 son variables binarias, 50 son ordinales con valores de 0 a 3 y 3 columnas son constantes. En la PEC2 se entregó un documento PDF realizado con R Markdown donde se ve en detalle como se han obtenido todos los diferentes tipos y como se fueron realizando las diferentes transformaciones.
- Las variables fueron renombradas de forma que cada columna ya tenga un significado por sí misma, sin necesidad de ir a un catálogo o diccionario de datos para ver los significados. A esta lista le falta la variable “Country” que como no existía en el dataset original, no ha sido necesario renombrarla

```
names(my_dataset)[names(my_dataset)=="Con"] <- "Conflict originated"
names(my_dataset)[names(my_dataset)=="Contp"] <- "Conflict Type"
names(my_dataset)[names(my_dataset)=="PP"] <- "Peace Process Number"
names(my_dataset)[names(my_dataset)=="PPName"] <- "Peace Process Name"
names(my_dataset)[names(my_dataset)=="Reg"] <- "Region"
names(my_dataset)[names(my_dataset)=="AgtId"] <- "Agreement ID"
names(my_dataset)[names(my_dataset)=="Agt"] <- "Agreement Name"
names(my_dataset)[names(my_dataset)=="Dat"] <- "Date Signed"
names(my_dataset)[names(my_dataset)=="Status"] <- "Agreement Definition and Status"
names(my_dataset)[names(my_dataset)=="Agtp"] <- "Agreement/conflict type"
names(my_dataset)[names(my_dataset)=="Stage"] <- "Agreement stage"
names(my_dataset)[names(my_dataset)=="LocISO"] <- "Country ISO"
names(my_dataset)[names(my_dataset)=="GCh"] <- "Ref.Children/Youth"
names(my_dataset)[names(my_dataset)=="GDis"] <- "Ref.Disabled persons"
names(my_dataset)[names(my_dataset)=="GAge"] <- "Ref.Older persons"
names(my_dataset)[names(my_dataset)=="GMig"] <- "Ref.Migrant workers"
names(my_dataset)[names(my_dataset)=="GRa"] <- "Ref.Racial/Ethnic"
names(my_dataset)[names(my_dataset)=="GRe"] <- "Ref.Religious groups"
names(my_dataset)[names(my_dataset)=="GInd"] <- "Ref.Indigenous people"
names(my_dataset)[names(my_dataset)=="GRef"] <- "Ref.Refugees/ displaced persons"
names(my_dataset)[names(my_dataset)=="GSoc"] <- "Ref.Social Class"
names(my_dataset)[names(my_dataset)=="Gewom"] <- "Ref.Women rights"
names(my_dataset)[names(my_dataset)=="GeLgbti"] <- "Ref.LGBTI persons"
names(my_dataset)[names(my_dataset)=="GeFa"] <- "Ref.Families"
names(my_dataset)[names(my_dataset)=="HrGen"] <- "Ref.Human rights"
names(my_dataset)[names(my_dataset)=="EqGen"] <- "Ref.Equality"
```

- Además, haré un cambio respecto a las variables de ordinales que contenían 4 posibles valores en el que se especificaban los siguientes valores: “None”, “Rethorical”, “Anti-discrimination” y “Substantial” que indicaban de alguna forma cuan grado de compromiso existía en cada uno de los acuerdos respecto al grupo que se tratase. Así por ejemplo son las variables “Ref. Children/Youth”, “Ref. Disabled persons”, “Ref. Older persons”, “Ref. Migrant workers”, “Ref. Racial/Ethnic”, “Ref. Religious groups”, “Ref. Indigenous people”, “Ref. Refugees/displaced persons”, “Ref. Social Class”, y de un modo similar con la variable “Ref.Equality”. Estas variables que acabo de mencionar las he transformado a binarias. De forma que, si antes los

valores eran 0, 1, 2, 3; ahora el 0=No y el resto es Yes. El motivo es que además de habría que ver exactamente el compromiso en los 3 niveles, en la PEC2 he dejado todos los valores, y no me terminó de convencer el resultado de tener un NO y tres diferentes SI para cada una de estas variables.

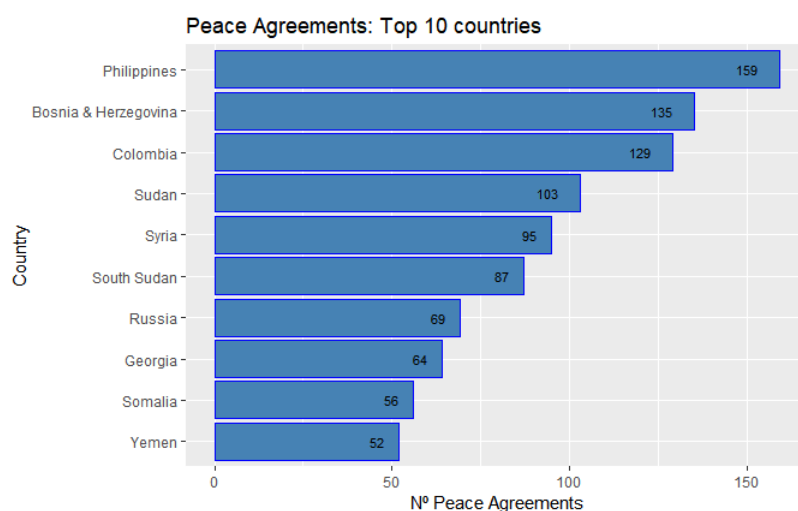
- Hice enriquecimientos de datos, gracias a los campos Loc1ISO y Loc2ISO mayoritariamente, y también gracias a la variable “Con” (conflict originated) en aquellos casos que Loc1ISO y Loc2ISO no tenían valor (por haber más de 2 países involucrados), de forma que he podido desdoblar y obtener los nombres de los países involucrados en los conflictos y acuerdos de paz, pudiendo obtener una nueva variable llamada “Country” que pudiese contener el nombre del país. Debemos tener en cuenta que estamos desdoblado registros, por lo que vamos a tener más registros de los que había originalmente en el dataset. Esto deberemos tenerlo en cuenta a la hora de representar los diferentes acuerdos, utilizando el DISTINCT del “Agreement ID” ya que ahora podemos tener varios registros “duplicados” pero que cambia en cuanto al Código ISO de país, y la variable Country obtenida (el nombre del país).

Analizando los datos pudimos ver como se distribuyen los datos en función de las diferentes variables:

Por ejemplo, podemos ver el TOP 10 de países con mayor número de acuerdos.

```
> head(sort(table(data_def$Country), decreasing = TRUE), 10)
```

Philippines	Bosnia & Herzegovina	Colombia	Sudan	Syria
159	135	129	103	95
South Sudan	Russia	Georgia	Somalia	Yemen
87	69	64	56	52



Ver como se distribuyen los acuerdos entre las diferentes regiones:

```
> sort(table(data_def$Region), decreasing = TRUE)
```

Africa (excl MENA)	Europe and Eurasia	Asia and Pacific
667	506	395
Middle East and North Africa	Americas	Cross-regional
272	211	61

Podemos ver que hay 99 países implicados en estos 1831 acuerdos, cuyos códigos ISO de país son:

```
> unique(data_def$`Country ISO`)
[1] "AFG" "DZA" "AGO" "ARG" "AZE" "BHR" "BGD" "BIH" "PNG" "BDI" "KHM" "CMR" "CAF" "SDN" "TCD" "CHN" "PRK"
[18] "COL" "COM" "NIC" "CIV" "HRV" "COD" "DJI" "IDN" "ECU" "SLV" "ERI" "SOM" "ETH" "KEN" "GAB" "GEO" "GTM"
[35] "GIN" "GNB" "HTI" "HND" "IND" "IRQ" "GBR" "ISR" "LBN" "SYR" "LSO" "LBR" "LBY" "MKD" "MDG" "MLI" "MEX"
[52] "MDA" "MAR" "MOZ" "MMR" "NAM" "NPL" "NER" "NGA" "KOR" "PAK" "PHL" "COG" "RUS" "RWA" "YEM" "SEN" "SRB"
[69] "SLE" "SVN" "SLB" "ZAF" "SSD" "ESP" "LKA" "TJK" "THA" "TGO" "TUN" "UGA" "UKR" "ZWE" "ARM" "VEN" "PER"
[86] "PRT" "KWT" "USA" "IRL" "JOR" "GRC" "SAU" "ZMB" "TZA" "MNG" "KAZ" "DEU" "FRA" "PSE"
> length(unique(data_def$`Country ISO`))
[1] 99
```

Podemos ver la distribución según el tipo de conflicto:

```
> head(sort(table(dfpx$Contp), decreasing = TRUE))
```

Government/territory	Government	Inter-group	Territory	other
835	653	235	108	1

Ver según el proceso de paz, cuando documentos de acuerdos tiene cada uno de ellos (los 6 que más tienen)

```
> head(sort(table(dfpx$PPName), decreasing = TRUE), 6)
```

Bosnia peace process	Philippines - Mindanao process	Syrian Local Agreements
124	124	64
Abkhazia peace process	Colombia III - Arango	Somalia Peace Process
51	45	42

Podemos ver la distribución según el tipo de conflicto:

```
head(sort(table(dfpx$Agtp), decreasing = TRUE))
```

Intra	IntraLocal	InterIntra	Inter
1273	251	236	72

Y según el estado:

```
> head(sort(table(dfpx$Status), decreasing = TRUE))
```

Multiparty signed/agreed	Unilateral document	Status unclear
1691	124	10
Agreement with subsequent status		
7		

Podemos ver los diferentes acuerdos a lo largo de los años (desde 1990 hasta 2019)

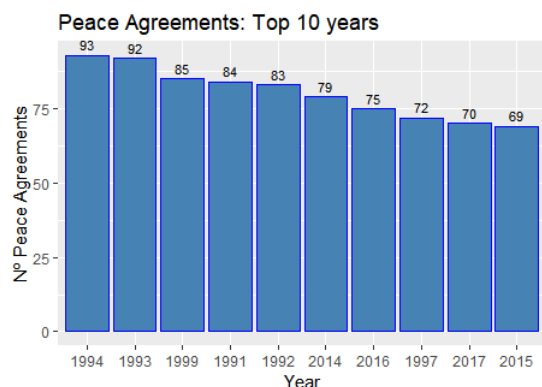
```
> table(year(as.Date(dfpx$Dat)))
```

1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
44	84	83	92	93	65	66	72	66	85	60	58	53	51	47	30	61	44	64	46	33
2011	2012	2013	2014	2015	2016	2017	2018	2019												
51	62	42	79	69	75	70	68	19												

Y por ejemplo ver los 10 años que más acuerdos de paz hubo

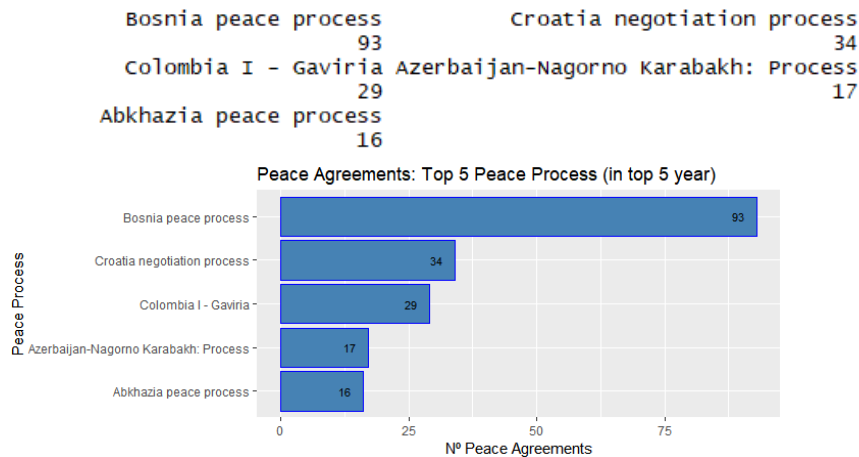
```
> head(sort(table(year(as.Date(dfpx$Dat))), decreasing = TRUE), 10)
```

1994	1993	1999	1991	1992	2014	2016	1997	2017	2015
93	92	85	84	83	79	75	72	70	69



Y podemos ver cuáles son los procesos de paz con mayor cantidad de documentos de acuerdos en los 5 años con mayor número de acuerdos

```
> top_5_year <- as.integer(names(head(sort(table(year(dfpx$Date)), decreasing = TRUE), 5)))
> head(sort(table(dfpx[year(dfpx$Date) %in% c(top_5_year), 'PPName']), decreasing = TRUE), 5)
```



Y en cuanto a los datos de las variables binarias (recordemos que el dataset tenía algunas variables binarias y las que eran ordinales de 0 a 3 las he transformado a binarias también) vemos como se distribuyen sus frecuencias también

	No	Yes
Ref.Children/Youth	1522	310
Ref.Disabled persons	1714	118
Ref.Older persons	1748	84
Ref.Migrant workers	1808	24
Ref.Racial/Ethnic	1477	355
Ref.Religious groups	1598	234
Ref.Indigenous people	1716	116
Ref.Refugees/ displaced persons	1356	476
Ref.Social class	1756	76
Ref.Women rights	1469	363
Ref.LGBTI persons	1821	11
Ref.Families	1640	192
Ref.Human rights	1284	548
Ref.Equality	1561	271
Ref.Democracy	1451	381

En la PEC2 puse alguno de estos grupos en la visualización, pero estaba claro que al ser variables independientes tenía que crear un gráfico para cada una de ellas, cada cual con su eje y sus valores por supuesto. Si quisiera sacar estos datos en un mismo gráfico combinado, debería de juntar todas estas en una sola variable para que uno de los ejes tenga los diferentes valores (los grupos). Para ello si finalmente optamos por esa solución a la hora de construir el gráfico, hay que trasponer las columnas en filas (hacer un reshaping de los datos usando el paquete tidyr, función gather).

Una vez realizadas las transformaciones necesarias, y tal y como hicimos en la PEC2 guardamos el fichero en formato Excel con el comando de R

`write_xlsx(x = list(pax_agreements=data_def), path = './pax_agreements.xlsx', col_names = TRUE)` en el que definimos el nombre de la hoja, el nombre del fichero Excel y que las columnas tengan nombres.

Esos datos serán cargados en Tableau, y en el proceso de carga se establece que la variable “Country ISO” sea una variable de geolocalización, y además que la variable “Date signed” es una fecha, y que se quede con el año correspondiente a esa fecha.

2.3. Relaciones que vamos a realizar

A diferencia de lo que hice en la PEC2 que fue como una visión más general, sin centrarlo en nada demasiado concreto, en esta nueva entrega quiero focalizar la atención concretamente en 2 aspectos diferentes del dataset.

Por un lado, quiero centrar una parte de la atención en los grupos que nos proporciona esta base de datos (las variables Groups que nos proporciona el documento explicativo (Bell, Christine, Sanja Badanjak, Juline Beujouan, Tim Epple, Robert Forster, Astrid Jamar, Sean Molloy, Kevin McNicholl, Kathryn Nash, Jan Pospisil, Robert Wilson, and Laura Wise, 2020) y que se refiere a los grupos vulnerables y con los que hay que tener especial cuidado y atención a la hora de tratar conflictos. Estos grupos que antes fui definiendo en la parte de los datos, son los relativos a “Niños”, “Personas con discapacidad”, “Mujeres”, “Personas mayores”, “Trabajadores migrantes”, “Grupos raciales/etnias”, “Grupos religiosos”, “Indígenas”, “Refugiados/desplazados”, “Clases sociales”, “Personas LGBTI” y “Familias”.

Y por otra parte quiero también ver como inciden los diferentes acuerdos de paz en lo relativo a derechos humanos e igualdad, y concretamente enfocado concretamente en 4 variables que he comentado antes en la parte de los datos, y que son las relativas a referencias en los acuerdos sobre aspectos de: “Derechos humanos”, “Igualdad”, “Democracia” y “Derechos civiles/políticos”.

Respecto a la PEC2, lo había centrado por una parte en algunas de las variables de grupo que aquí comentaba, pero llevado a los 4 niveles (variables ordinales que antes he comentado) y al final acabas obteniendo para cada variable un “No”, y 3 “Yes” pero con diferentes implicaciones en cada uno, con lo cual visualmente no es tan sencillo determinar el impacto de las referencias en cuanto al global de los acuerdos precisamente por la división en diferentes categorías.

Además, también había puesto foco sobre el tema de los acuerdos que tenían o no referencias sobre derechos de las mujeres, pero al estar en un dashboard diferente perdíamos la perspectiva y los filtros del dashboard general y por no repetirlo, lo dejé independiente, pero luego las comparaciones iban por regiones en lugar de por países escogidos libremente. Por tanto, he decidido cambiar un poco la forma de organizar los datos a visualizar de forma que todo tenga relación consigo mismo, y que los filtros funcionen de forma global (que eso ya se podía hacer) pero al menos viendo en la misma pantalla, cuales son los criterios que se han seleccionado, o pudiéndolos cambiar sin tener que cambiar de dashboard.

Organizando todo en un dashboard, donde todo sea accesible desde una sola pantalla mi idea es la siguiente:

- Seguir manteniendo de alguna forma un indicador global que nos muestre el número de acuerdos totales que tenemos en la selección actual.
- Tener un mapa como también tenía en la PEC2, con filtros a su derecha, donde se puedan filtrar años, donde se puedan filtrar regiones, los tipos de conflicto, el estado, etc. Es decir,

utilizar las variables que he dejado en mi dataset como parte útil para filtrar, para poder ir reduciendo o ampliando los focos sobre donde analizamos el impacto en las variables que acabo de poner.

- En la PEC2 tenía el total de acuerdos por año, pero en el dashboard de los derechos de las mujeres. Voy a recuperar ese gráfico, aunque lo más probable es que pase a ser un gráfico de barras simple (era un gráfico de barras apilado). Y sin referencia, ya que el grupo de mujeres podrá ser utilizado de forma independiente.
- Quiero relacionar las diferentes variables de los grupos entre ellas, de forma que dependiendo de las selecciones que tenga en todo el dashboard, de forma que sea visible de forma muy clara qué grupos están representados en los acuerdos y cuáles no, si además son de forma mayoritaria o no. Seguramente nos vamos a encontrar (tal y como hemos visto en la distribución de frecuencias de los datos) con que alguno de los grupos apenas tiene visibilidad, pero al ponerlo de esta forma podremos ver qué países son los que los incluyen.
- También quiero una relación temporal de los grupos referentes a los derechos humanos. Lo planteo para este caso y no para los grupos, porque en los grupos estamos manejando 12 variables, lo cual para un gráfico de líneas se hace cargante y poco claro, y con gráficos apilados tampoco vamos a tener una información muy clara. Con esto podremos tener una comparación temporal y entre ellos mismos, de cuantos acuerdos tienen referencias a esos indicadores referentes a derechos humanos.
- Y aunque el mapa me permite situar geográficamente los países, y además por los colores podemos tener una buena aproximación de los datos (cantidad de acuerdos) es cierto que dependiendo del área de cada país es mucho más fácil (o complicado) el darse cuenta. Por poner un ejemplo, se ve bastante claro que países como Colombia o Sudán tienen muchos acuerdos de paz, pero en cambio es más complicado distinguir que Filipinas y Bosnia que ambos tienen incluso más acuerdos, al ser países con menor superficie, son los que más tienen (porque destacan menos). Mi propuesta es poder tener un ranking de países de forma que no haya que tener demasiado ojo clínico y tener que localizar los países en el mapa, y pasar por encima con el puntero, para ver los datos relativos a ese país. En un ranking es mucho más sencillo de ver y por tanto hacer comparaciones de cuáles son los países que más acuerdos tienen que cumplan con los filtros establecidos.

2.4. Descripción de la gráfica elegida

Para la visualización principal, he dejado un mapa nuevamente, ya que es algo muy intuitivo. Ryan Sleeper en su libro “Practical Tableau” comenta “The power of maps comes from their inherent ability to leverage schemas that your users have been building up for many years.” (Sleeper, Ryan, 2018). Todos tenemos incorporado de forma casi automática el uso de los mapas, y por tanto eso ya de por sí se convierte en una ventaja. Poder situar geográficamente los países, tener la capacidad de evaluar lo que ocurre en los países limítrofes o dentro de la misma zona, y también si una persona quiere buscar los datos de un país o región concreta, debería ser mucho más sencillo encontrarlo en un mapa que tener que buscarlo en una lista ordenada alfabéticamente. Además de cara a la multiselección, es muy sencillo el poder hacerlo gráficamente, sabiendo en todo momento qué es lo que estamos eligiendo. Nuestro dataset además se adapta perfectamente porque tenemos los códigos ISO de los países, y también los propios países (gracias al enriquecimiento de datos) con lo cual podemos situar

todos los registros en un mapa. Respecto a lo que hice en la entrega anterior, al menos cuando vas pasando por encima, y nos pone el nombre del país y el número de acuerdos que le involucran, también voy a añadir el porcentaje sobre el total de acuerdos de paz.

Al igual que hice en la PEC2, en la zona de los filtros (un poquito más arriba) tengo previsto colocar un indicador que nos muestre el número de acuerdos que tenemos en la selección actual que estemos visualizando. De esa forma, siempre tenemos ese dato a la vista, tanto si no hemos realizado ningún filtro (ese número coincidirá con el número de acuerdos, y si hay algún filtro aplicado, ese número varía en función de los filtros).

En el mapa llevaremos incorporados los filtros, un poco al estilo que se hizo en la PEC2, intentando colocar las variables que sean posibles (habrá que ver a la hora de la implementación como quedan los espacios y cuantas cajas de filtros se pueden colocar). La intención es poder tener las variables categóricas, también un selector de año (que me gustaría que fuese con posibilidad de rango, ahora mismo en PEC2 esta con selección año individual) y ver también la posibilidad de incorporar también los grupos vulnerables y un selector para los indicadores de derechos humanos.

He comentado las ventajas de los mapas, pero también tienen sus inconvenientes. Por ejemplo, Gerardo Furtado (Furtado, Gerardo, 2015) nos advierte que con los mapas es difícil hacer comparaciones, y clasificaciones individuales. Tanto sea con valores numéricos en los propios países, como con colores indicativos de cantidades, el problema es hacer una comparación y clasificación ordenada de ellos. Y de ahí que haya surgido la idea también de añadir por tanto un gráfico de barras horizontal (Choudhury, Shilpi, 2013) ya que si vamos a poner nombres de países es mejor esa orientación, además de que para representaciones nominales (en este caso los nombres de los países) se aconseja el uso de bar charts (gráficos de barras horizontales). Gracias a este gráfico, y en conjunción con el mapa, ya tenemos una buena forma de visualizar los países y además clasificarlos por el número de acuerdos (en función de la selección actual).

Otro gráfico que recupero de alguna forma es el que me daba la evolución temporal del número de acuerdos a lo largo de los años. Yo lo tenía en la visualización #2, como un gráfico de barras apiladas en función de la variable binaria que indica si había referencias a los derechos de las mujeres. En este caso en principio la idea es no vincularla a nada, o en su defecto que el vínculo pudiese ser sobre cualquiera de las variables binarias (y no solamente el de las mujeres). La idea inicial es dejarlo como un gráfico de tiempo (por tanto, lo resolveré igualmente con un gráfico de barras) en este caso barras verticales que son adecuadas para representaciones temporales (Yi, Mike, 2019).

Ahora vienen las representaciones de las variables binarias, que hemos dividido en 2 categorías. La categoría de grupos vulnerables (que dijimos que son 12 variables binarias) y la categoría referente a los derechos humanos (que en este caso nos decantamos por usar 4 variables binarias).

Para resolver la visualización correspondiente a las 4 variables relativas a los derechos humanos, he escogido un gráfico de líneas de 4 series, ya que me parece un número adecuado de variables para este gráfico y además nos va a proporcionar una comparativa temporal, la evolución de cada una de las variables en el tiempo y las diferencias entre ellas. (Rost, Lisa Charlotte, 2018).

Y en cambio para la visualización de las variables correspondientes a los grupos sociales, como ya son muchas más variables, un gráfico de series temporal va a quedar muy confuso, se van a superponer

demasiadas líneas y no vamos a ver nada. Entonces lo podremos resolver pues como hicimos con el gráfico de los países, un gráfico de barras horizontal, (Oetting, Jami, 2018) ya que nos va a permitir colocar los nombres de las diferentes variables (las categorías) y así podremos tener una visualización rápida de qué tipo de colectivos está presente en más acuerdos de la selección que haya en el momento. De esa forma podremos comprobar como evolucionan los diferentes grupos vulnerables en función de las zonas donde se produzcan los conflictos.

Un último gráfico, pero este lo considero opcional, porque no tengo claro como va a quedar, es utilizar un treemap, primero para variar un poco más el uso de los gráficos que utilizo, y segundo porque basándome en una recomendación del tutor Mar Canet, de una web de visualización de índices financieros, y la verdad es que los gráficos son muy intuitivos. Voy a intentar hacer una prueba por ejemplo con las regiones, y ver si alguno de los indicadores, y un promedio de ellos puede dar alguna información útil, a la vez que sea muy intuitiva. Digamos que lo pongo en standby, y durante el desarrollo de la PEC4 a ver si soy capaz de montar algo así y que tenga sentido.

2.5. Descripción de la gobernanza de datos

Como vimos cuando describimos el dataset, las únicas 6 variables cuantitativas eran: “AgtId” que es el ID del acuerdo, “Lgt” el tamaño en páginas del acuerdo, “N_characters” que es el número de caracteres del texto del acuerdo, “PP” que es el número del proceso de paz, y las variables Loc1GWNO y Loc2GWNO (correspondientes a códigos de países). Por tanto, si nos fijamos bien, no hay variables “medibles”, y en nuestro caso, y ya lo hicimos así en al PEC2, nuestra variable a medir siempre contar los diferentes AgtId (es decir COUNTD ([AgtId]) en Tableau). Decimos los diferentes porque al haber efectuado transformación y enriquecimiento de datos, ahora nos encontraremos algunos valores de AgtId duplicados.

A través de los diferentes filtros y los diferentes gráficos, todos interactúan con todos. Es decir, los filtros son directos, y luego las selecciones dentro de cada uno de los gráficos hacen de filtro también en todos los demás. Por ese motivo he querido deshacer la historia que había montado en la PEC2, porque comprobé que o bien tenía los filtros desconectados, o si los conectaba (también lo he probado) el problema es que igual seleccionaba unas variables en el primer dashboard, y afectaba al dashboard #2, pero luego vas a ese y no entiendes muy bien qué es lo que está pasando porque yo no repliqué los filtros. Al final lo tomé como historias diferentes, cada una con sus propias selecciones. Para este caso que quiero estudiar ahora, y para centrarme específicamente sobre unas variables concretas, he optado porque todo pertenezca al mismo dashboard y de esa forma los filtros y restos de gráficos que interactúan son visibles y accesibles en todo momento, con lo cual evitamos confusiones.

Al abrir la visualización, sin ningún filtro nos encontramos de lleno con el mapa y sus diferentes colores en función del número de acuerdos por cada uno de los países. A su derecha arriba, siempre el indicador del total de acuerdos “visibles” en ese momento, que es dependiente de las diferentes selecciones. El mapa es el que tiene los filtros, con lo cual aparecen a la derecha de este. Como he comentado en la sección anterior, tengo que esperar a ver cuales son las variables que puedo colocar con filtros. Intentaré colocar las que pueda de forma que el gráfico tenga la mayor usabilidad posible, permitiendo navegar desde muy arriba hasta un nivel de detalle interesante. Este es el gráfico

primario. En la entrega anterior era el correspondiente a “Overview of Peace Agreements” y luego se titulaba “Countries with Peace Agreements”.

Ya por debajo de esto, se coloca el gráfico de barras vertical que nos muestra la distribución de frecuencias de los acuerdos por año. A la derecha de ese, la idea es que se sitúe el ranking de países (el gráfico de barras horizontal), de forma que con este gráfico estamos complementando el mapa tal y como describimos en el apartado anterior. Es decir, hasta aquí tenemos un vistazo general de cómo se distribuyen los acuerdos a través de los países (gracias al mapa), vemos como son sus frecuencias anuales y tenemos un ranking de países para poder hacer comparativas fácilmente entre ellos.

Ya por debajo de estos y haciendo coincidir el ancho del gráfico, voy a colocar el gráfico de líneas temporal para mostrar con 4 series las variables referentes a los derechos humanos. Al quedar por debajo del otro gráfico de barras, y mismo año para ambos, respetamos la dimensión de tiempo y aunque no estén montados en el mismo gráfico podremos ver la evolución de esas variables, y arriba los totales de acuerdos. A ver cómo queda. Y a la derecha de este, y ocupando el mismo espacio que pasaba con el gráfico de países, en este caso colocamos el gráfico de barras horizontales también, pero con los grupos vulnerables, con lo cual tenemos una distribución simétrica.

Como comentaba antes, pretendo que todos los gráficos y filtros interactúen como un conjunto único, de forma que podamos ver las relaciones existentes cuando tocamos algún valor (selección de variables) y como afecta al resto de gráficos y variables.

3. Bloque diseño

3.1. Elección de colores

Para el mapa he cambiado la paleta de colores que había escogido en la PEC2, que estaba basado en el Naranja de Tableau. Me he dejado llevar por un documento de Kennet Field (Field, Kenneth, 2020), en el que nos invita a ser responsables con los datos que presentamos evitando colores sensacionalistas y haciendo un esquema simple de paso de claros a oscuros para que se pueda evaluar de menos a más. La verdad es que el aspecto me gustó más, y por tanto me decanté por cambiarlo. Ya era un degradado secuencial, simplemente he pasado a “Azul” de Tableau.

Además, busqué información relativa al significado de los colores, y cuando se aconseja cada tipo de color, o a que se asocia leyendo algunas documentaciones (Alina Arhipova, s.f.) he visto que el azul proporciona entre otras cosas verdad, seguridad y estabilidad, y me ha parecido apropiado. También revisando más documentación acerca de la elección de colores en una web (Mike Yi, 2019) encontré otro mapa en tonos azules y también hablaba entre otras cosas de la importancia de realizar visualizaciones teniendo en cuenta a las personas daltónicas, y ponía un enlace a una web donde puedes subir imágenes y verlas simulando como las verían y además tiene como muchos parámetros de anomalías de vista, con lo cual hay muchas combinaciones (Anon., s.f.). La verdad es que el resultado que obtuve con el mapa que subí, me pareció adecuado.

Para definir los colores de los gráficos de barras, he leído un artículo que me ha parecido bastante interesante (Stephen Few, 2008) donde nos explica que si el color no agrega significado, entonces es

mejor no usarlos, es decir, no poner gráficos de barras con diferentes colores si cada color no significa nada, ya que nos lleva a intentar interpretar un significado que no existe y por tanto confunde. Con lo cual elijo un tono azul en armonía con el gráfico, para que tampoco resalte demasiado.

Para el gráfico de líneas, fue el que más dudas tuve, porque aquí tengo que comparar en el mismo gráfico diferentes series y por tanto ahora sí que no pueden ser iguales. Pero he visto una web de www.dataquest.io donde (Lisa Charlotte Rost, s.f.) sugiere que en lugar de poner colores a las series, estas queden de un color gris pálido si el usuario no ha seleccionado, o alguna combinación con un gris cálido. Tengo que probar realmente si esta funcionalidad la puedo desarrollar en Tableau, con lo cual no tomo una decisión definitiva de cómo voy a montar estas series. La idea que propone me gusta, al menos en teoría. A ver cuando toque verla de verdad si realmente es posible y si además tiene sentido.

Para los títulos un tono suave, que los textos sean legibles, pero que tampoco sean un foco donde a los usuarios se les vaya la vista. Más intenso sí, el valor indicador del número total de acuerdos.

3.2. Tipografía

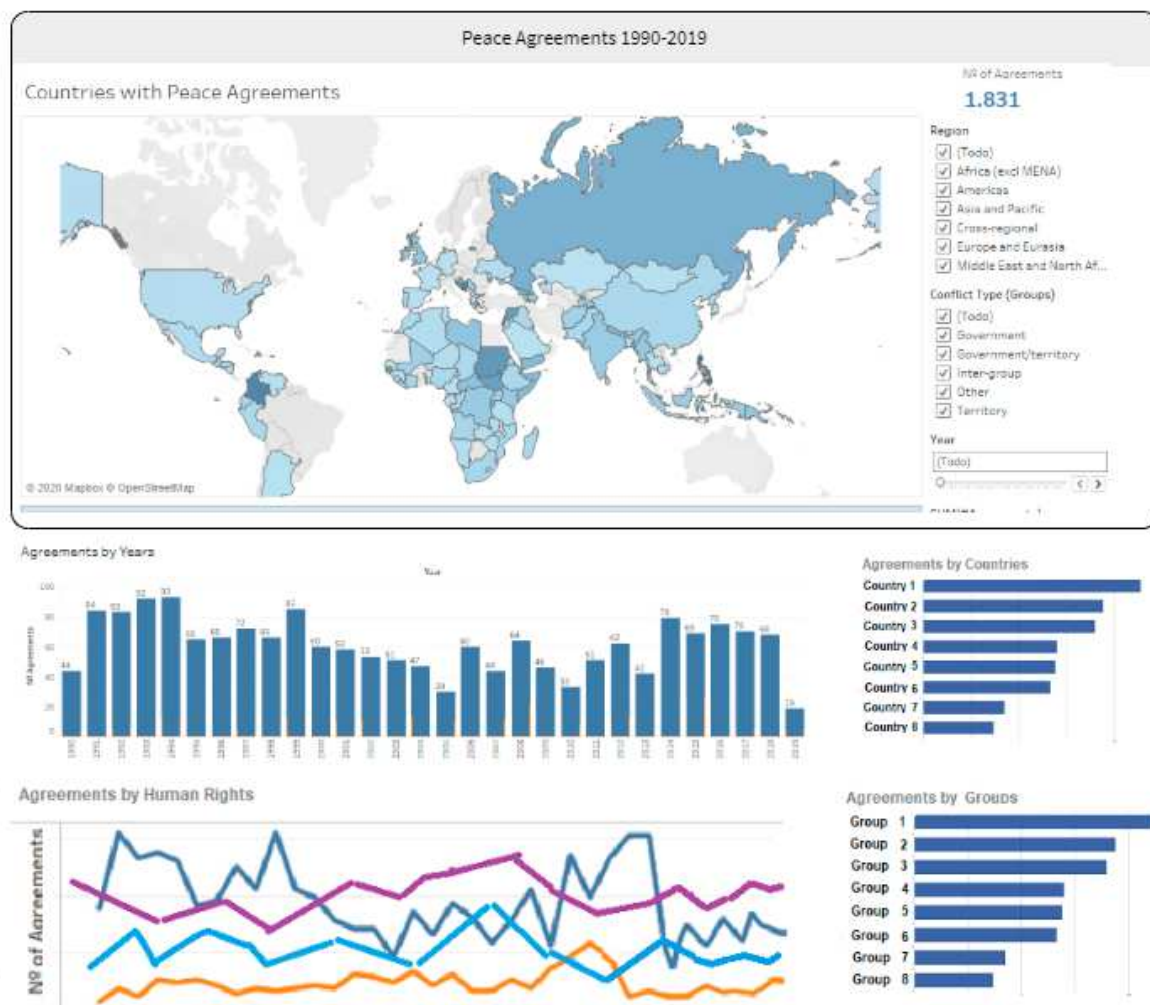
En principio la idea es mantener las fuentes y tamaños utilizados en la visualización entregada con la PEC2 y que es de la siguiente manera:

- Título de la visualización: Tableau Regular 18
- Título de cada uno de los gráficos: Tableau Light 15
- Indicador de número de acuerdos: Tableau Bold 11
- Filtros: Tableau Medium 9
- Descripciones emergentes: Tableau Book 10

3.3. Plantilla

Esto simplemente es un mockup realizado con imágenes que pude obtener de mi propia PEC2 y retocando cosas directamente con una simple aplicación de dibujo.

El filtro no va a quedar así, pero al menos quería que en la maqueta tuviese un aspecto algo presentable. El aspecto a algo similar a ese, haciendo que los filtros podamos poner otras variables (sin desplegar para poder poner más cantidad) y luego tengo que ver como integrar ambos grupos como una única variable que puedan compartir un mismo gráfico (los grupos y los derechos humanos).



3.4. Formato

Formato digital en Tableau Public tal y como se hizo en la entrega de la PEC2. Se adjuntará el link de acceso al documento.

Bibliografía

Alina Arhipova, s.f. *Color in Design: Influence on Users' Actions*. [En línea]

Available at: <https://blog.tubikstudio.com/color-in-design-influence-on-users-actions/>

[Último acceso: May 2020].

Anon., s.f. *Coblis — Color Blindness Simulator*. [En línea]

Available at: <https://www.color-blindness.com/coblis-color-blindness-simulator/>

[Último acceso: May 2020].

Bell, C. and Badanjak, S., 2019. Introducing PA-X: A new peace agreement database and dataset. *Journal of Peace Research*, March. Volumen 56.

Bell, Christine, Sanja Badanjak, Juline Beujouan, Tim Epple, Robert Forster, Astrid Jamar, Sean Molloy, Kevin McNicholl, Kathryn Nash, Jan Pospisil, Robert Wilson, and Laura Wise, 2020. *PA-X Peace Agreements Database and Dataset, Version 3*, s.l.: s.n.

Bell, Christine, Sanja Badanjak, Juline Beujouan, Tim Epple, Robert Forster, Astrid Jamar, Sean Molloy, Kevin McNicholl, Kathryn Nash, Jan Pospisil, Robert Wilson, and Laura Wise, 2020. *PA-X Codebook, Version 3*, Edinburgh: s.n.

Choudhury, Shilpi, 2013. *Choosing the right chart type: Bar charts vs Column charts*. [En línea]

Available at: <https://www.fusioncharts.com/blog/bar-charts-or-column-charts/>

[Último acceso: May 2020].

Field, Kenneth, 2020. *Mapping coronavirus, responsibly*. [En línea]

[Último acceso: May 2020].

Furtado, Gerardo, 2015. *THE PROS AND CONS OF USING MAPS TO VISUALISE DATA*. [En línea]

Available at: <http://theinfologist.com/the-pros-and-cons-of-using-maps-to-visualise-data/>

[Último acceso: May 2020].

Lisa Charlotte Rost, s.f. *What to Consider When Choosing Colors for Data Visualization*. [En línea]

Available at: <https://www.dataquest.io/blog/what-to-consider-when-choosing-colors-for-data-visualization/>

[Último acceso: May 2020].

Mike Yi, 2019. *How to Choose Colors for Data Visualizations*. [En línea]

Available at: <https://chartio.com/learn/charts/how-to-choose-colors-data-visualization/>

[Último acceso: May 2020].

Oetting, Jami, 2018. *Data Visualization 101: How to Choose the Right Chart or Graph for Your Data*. [En línea]

Available at: <https://blog.hubspot.com/marketing/types-of-graphs-for-data-visualization>

[Último acceso: May 2020].

Rost, Lisa Charlotte, 2018. *What to consider when creating line charts*. [En línea]

Available at: <https://blog.datawrapper.de/line-charts/>

[Último acceso: May 2020].

Sleeper, Ryan, 2018. *Practical Tableau: 100 Tips, Tutorials, and Strategies from a Tableau Zen Master*. s.l.:O'Reilly.

Stephen Few, 2008. Practical Rules for Using Color in Charts. *Visual Business Intelligence*, February.

Yi, Mike, 2019. *A Complete Guide to Bar Charts*. [En línea]

Available at: <https://chartio.com/learn/charts/bar-chart-complete-guide/>

[Último acceso: May 2020].