

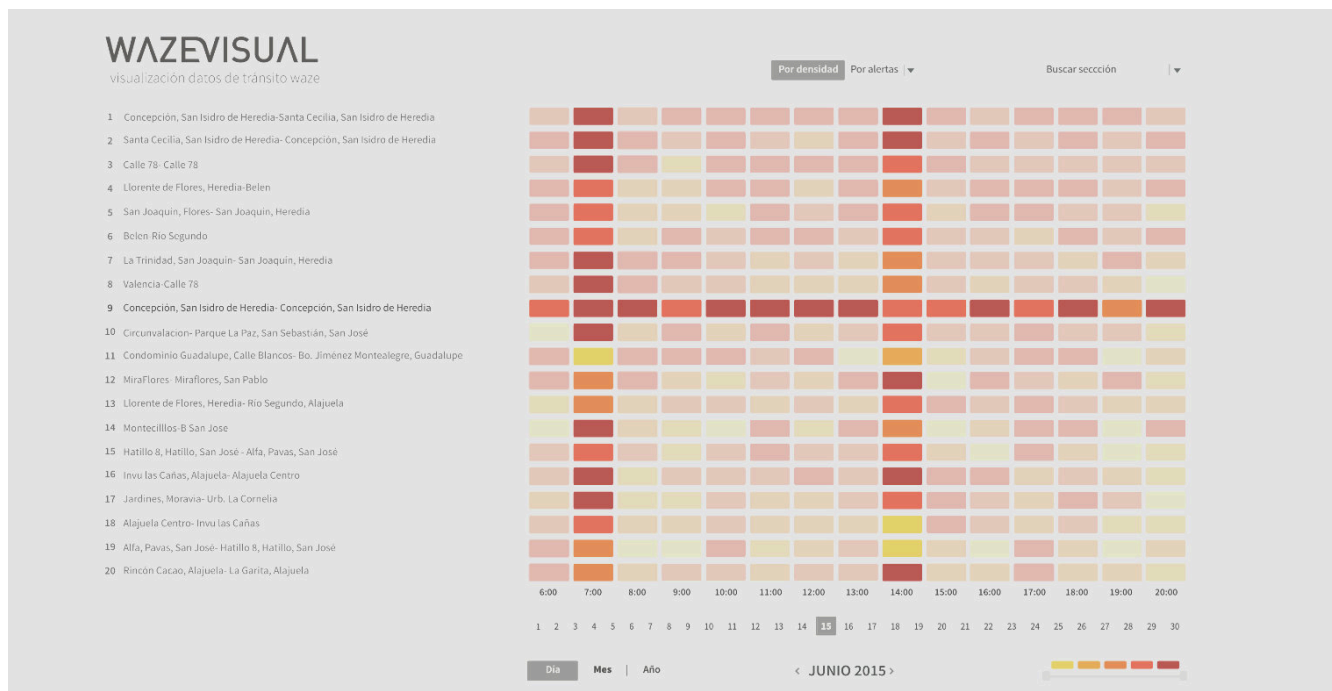
dashboard design cookbook

metodología para el diseño de visualizaciones de datos para la toma de decisiones

ph.h. franklin hernández-castro

instituto tecnológico de costa rica

franhernandez@itcr.ac.cr



Traffic Data Visualization in Costa Rica [1]

1era edición: julio 2021

como citar este trabajo:

Hernández-Castro, F. (2021). dashboard design cookbook, metodología para el diseño de visualizaciones de datos. Escuela de Diseño Industrial, Instituto Tecnológico de Costa Rica. Cartago, Costa Rica. Tomado de www.xxxx.xx en fecha xx/xx/xx

índice

introducción	3
metodología	4
1. búsqueda de los datos pertinentes	5
1.1. formatos	5
1.1.1 archivos de tarjetas	5
1.1.2 tablas	6
2. tipos de datos y sus relaciones	7
2.1. tipos de datos	7
2.2. relaciones entre datos	8
3. definición de las preguntas objetivo POs	11
4. data cleaning (EDA) y data parsing	13
4.1. exploratory data analysis (EDA)	13
4.1.1. completitud	13
4.1.2. coherencia	13
4.2 data parsing	15
5. exploración de paradigmas	17
7. validación	25
8. apéndices	31
8.1. algunas fuentes de datos publicas	31
9. Referencias	32

introducción

el presente documento es un resumen de la metodología que recomiendo para el diseño de "dashboards" o visualizaciones de datos que se usan como fuente de información para tomar decisiones a diario.

para entender mejor de qué tipo de visualización estamos hablando, yo distingo entre **dos grandes tipos de visualizaciones de datos** (ver figura 1 y 2):

- visualizaciones **de datos científicos** (para investigación)
- visualizaciones de datos **para toma de decisiones**

fig 1. visualización científica mostrando más de 140,000 epicentros de terremotos que se originan en las zona de subducción en el centro de costa rica, TEC / OVSICORI-UNA [2]

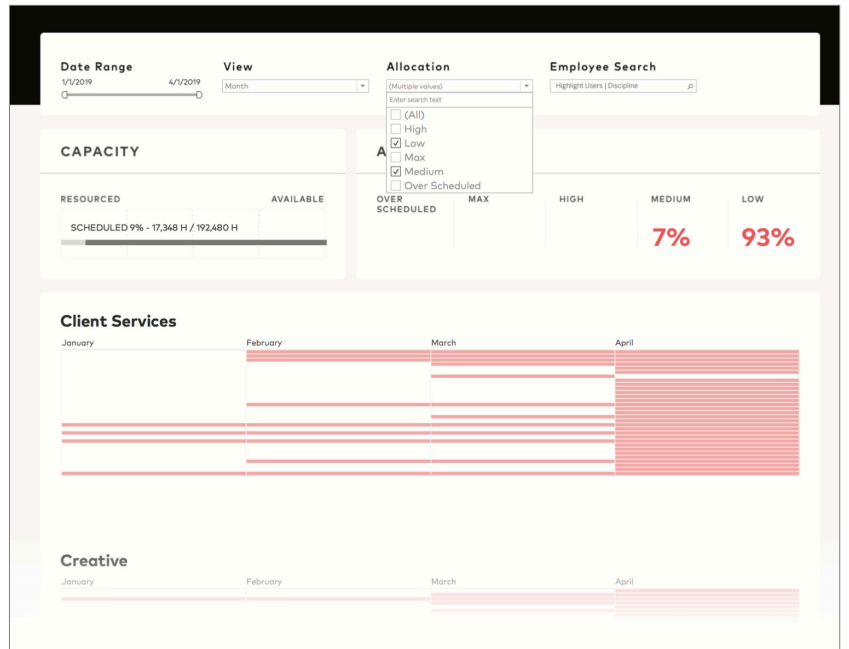
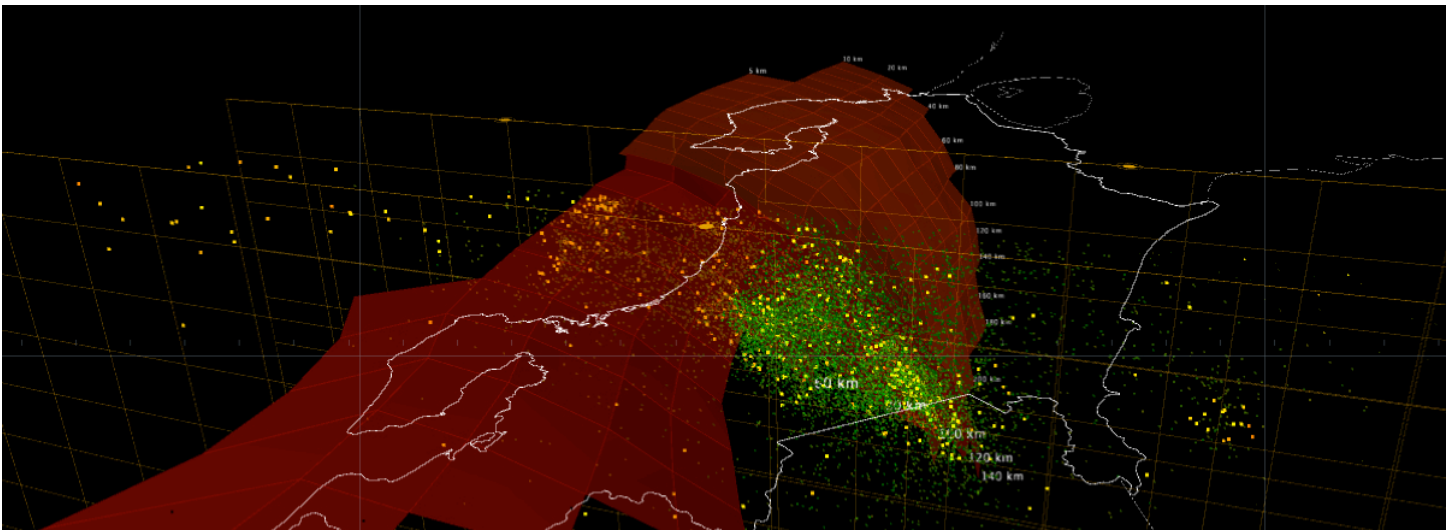


fig 2. visualizaciones de datos para toma de decisiones, dashboard para explorar la ocupación de personal para una empresa, mariana agüero soto, TEC-1S-2019

en este documento trataremos con el segundo tipo de visualización de datos, **diseño de dashboards para toma de decisiones**, que es el más demandado hoy en día en el mercado. sin embargo, muchos conceptos en los procesos de diseño son compartidos por ambos tipos de visualizaciones.

metodología

el proceso metodológico que hemos depurado durante los últimos 10 años **cuenta con siete pasos**. es el resultado de pruebas y errores, ajustando entre la teoría y la práctica con el fin de dar una guía simple y fácil de usar en la mayoría de los casos.

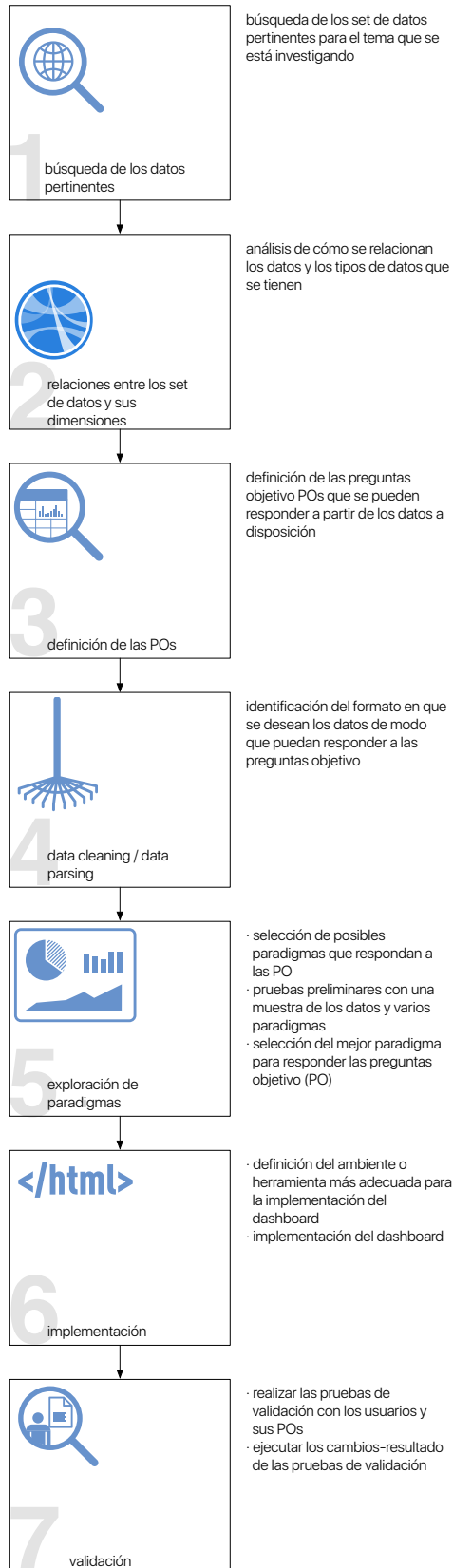


fig 3. metodología básica para diseño de dashboards, ph.d. franklin hernández-castro, TEC 2021

1. búsqueda de los datos pertinentes

cómo actuar en éste paso depende de cuál es el origen del proyecto, habrán proyectos que inician ya con los datos que tiene una empresa y que necesita una herramienta de análisis para los mismos; y habrán proyecto más académicos, de interés personal o de investigación donde se buscan datos para tratar de responder a una pregunta, suposición u hipótesis planteada.

en cualquier caso se debe **hacer una búsqueda, revisión y análisis de los datos** que se tienen, para tener una **idea general de con qué se cuenta y qué tipo de hipótesis se pueden responder** con esos datos.

hay numerosas fuentes de datos públicos de instituciones como la organización mundial de la salud, naciones unidas o gobiernos y universidades, en el apéndice 8.1 hay unas pocas fuentes de los miles que se encuentran en le internet.

1.1. formatos

los datos pueden venir en varios tipos de formatos, al final podemos imaginarnos los datos en dos modalidades con respecto a los formatos, los que vienen en **forma de "archivo de tarjetas"** y los que vienen en **forma de tablas**.

1.1.1 archivos de tarjetas

imaginémonos que los datos están ordenados en forma de un fichero antiguo, como los que aun hay en las bibliotecas. es decir, son tarjetas y cada "muestra" es una tarjeta, misma que podría estar ordenada por el apellido del autor del libro, seguido del nombre; después de eso vienen los datos del libro, la editorial, etc. y una dirección que indica donde encontrarlo en las estanterías.

pues **así son los datos almacenados en bases de datos digitales**, tales como SQL, mySQL, etc., tienen algunas llaves o "**tags**" que son como el apellido del autor en el ejemplo de las tarjetas; es decir, son las "palabras claves" con las cuales se busca esa "muestra" y después tienen "campos" donde está el resto de la información, como la editorial o el año de publicación.

obviamente cuando se trabaja con estas bases de datos se debe respetar un protocolo de acceso que hay que aprender y utilizar al pie de la letra.

por ejemplo, un formato de datos muy usado es el **JSON** que es un **formato nativo de javascript**. este formato es de esta clase "tipo tarjetas" y se define en forma de un archivo de texto y con un protocolo de corchetes "{}" que define la estructura.

en la figura 4 vemos un ejemplo, primero el archivo completo empieza con la apertura de un corchete "{" luego aparece la descripción de los datos, seguido se ve la apertura de un paréntesis cuadrado "[" lo que significa que empieza una "lista" o "array" con las "tarjetas".

la primera tarjeta corresponde a la ciudad de tokyo, se ve como está encerrada en un bloque de corchetes "{}" y dentro de ellos vienen las características, en este ejemplo: "city_ascii", "longitude", "latitude", "country", "iso3", "population" y "id".

y así se conforma la "tarjeta" de esta ciudad, después viene new york y mexico city. así se ve un ejemplo de un archivo en este protocolo que por supuesto puede anidarse tanto como se desee.

```

1  {
2    "description": "most populated cities in te
3    "source": "I forgot. I'm sorry.",
4    "cities":
5      [
6        {
7          "city_ascii": "Tokyo",
8          "longitude": 139.7514,
9          "latitude": 35.685,
10         "country": "Japan",
11         "iso3": "JPN",
12         "population": 35676000,
13         "id": 1392685764
14       },
15       {
16         "city_ascii": "New York",
17         "longitude": -73.9249,
18         "latitude": 40.6943,
19         "country": "United States",
20         "iso3": "USA",
21         "population": 19164071,
22         "id": 1840034016
23       },
24       {
25         "city_ascii": "Mexico City",
26         "longitude": -99.131,
27         "latitude": 19.4424,
28         "country": "Mexico",
29         "iso3": "MEX",
30         "population": 19028000,

```

fig 4. ejemplo de una fracción de un archivo tipo JSON

evento	longitud	latitud	windkph	day	month	year	type	hour	press
1	-79	11.5	0	17	5	1970	-1	18	0
2	-79.2	11.7	3	18	5	1970	-1	0	0
3	-79.7	12.1	11	18	5	1970	-1	6	0
4	-80.1	12.3	7	18	5	1970	-1	12	1007
5	-80.5	12.5	7	18	5	1970	-1	18	0
6	-81	13	11	19	5	1970	-1	0	0
7	-81.5	13.5	11	19	5	1970	-1	6	0
8	-82	14	11	19	5	1970	-1	12	1005
9	-82.5	14.5	11	19	5	1970	-1	18	0
10	-82.5	15.5	18	20	5	1970	0	0	0
11	-82.3	16.2	12	20	5	1970	0	6	0

fig 5. tabla de huracanes con más de 40,000 eventos de huracanes en los últimos 50 años en el mar caribe, 2020, TEC-CIGEFI UCR [3]

1.1.2 tablas

la otra modalidad en la que vienen los datos es en forma de tablas, imaginémonos un excel muy extenso.

si, por ejemplo, tenemos datos de huracanes y en cada fila (o renglón) hay un "evento", es decir, una anotación de dónde estaba un huracán específico en un día específico; entonces las columnas dirán cosas como latitud, longitud, la magnitud del huracán, etc. (ver figura 5).

los formatos en que vienen estos datos son variados, el mismo **formato comercial de excel** es uno de los más comunes de encontrar y la mayoría de las aplicaciones con las que se visualizan los datos son capaces de trabajar con este formato.

el formato **CSV (comma-separated values)** es otro de los más comunes y por demás muy sencillo. es un archivo de texto en el que los datos están separados por comas, como lo dice su nombre, y usa la extensión .csv en el nombre del archivo (hay una versión en el que los datos se separan por "tabs" y se llama tsv).

este tipo de archivos pueden venir con "encabezado" (head) o sin el. en el caso de que vengan con encabezado se trata de la primera línea donde están los "nombres" de cada columna también separados por comas.

un fragmento de un archivo de este tipo se puede ver así:

estacion,day,month,year,precipitacion

7,17,5,1970,13.815146

8,17,5,1970,28.0

10,17,5,1970,19.892172

11,17,5,1970,19.343208

14,17,5,1970,19.38659

22,17,5,1970,35.3

25,17,5,1970,15.0

27,17,5,1970,33.151447

28,17,5,1970,18.24705

29,17,5,1970,27.0

estacion	day	month	year	precipitacion
7	17	5	1970	13.815146
8	17	5	1970	28.0
10	17	5	1970	19.892172
11	17	5	1970	19.343208
14	17	5	1970	19.38659
22	17	5	1970	35.3
25	17	5	1970	15.0
27	17	5	1970	33.151447
28	17	5	1970	18.24705
29	17	5	1970	27.0

fig 6. tabla ejemplo del archivo csv mencionado en el texto

a pesar de que se ve algo confuso por la ausencia de espacios en blanco entre los datos, la realidad es que es muy simple, cada coma representa un salto de columna y por supuesto, cada línea es una línea de la tabla (a la izquierda se ve la tabla que define ese segmento de archivo, fig. 6).

los datos en formato **.csv se pueden "abrir" en un editor de texto simple** o en las herramientas típicas como "excel" o "apple numbers", en esos casos también la herramienta puede "exportar" en este formato.

generalmente estos archivos son muy extensos, en el caso de los huracanes, por ejemplo, el archivo contiene más de 40,000 líneas hacia abajo.

§§§

hay muchos otros formatos de datos, especialmente los desarrollados para tipos de datos específicos como datos geo-referenciados o datos astronómicos, para los más especializados hay herramientas para poder extraer los datos y hacer los análisis. así que es importante saber en cuál formato están los datos que se van usar y cómo podemos acceder a ellos.

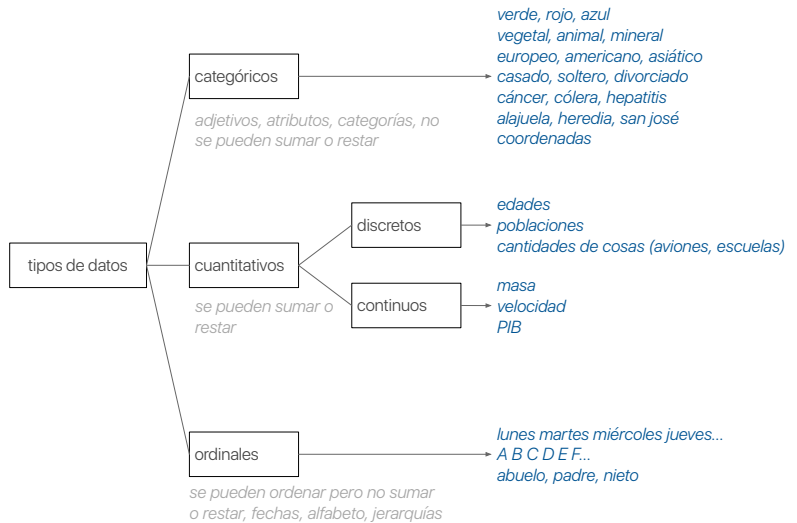
2. tipos de datos y sus relaciones

2.1. tipos de datos

hay muchos modos de clasificar datos dependiendo de para qué se haga la clasificación, posiblemente la más común sería según el tipo de variable que lo albergaría como enteros, strings, floats o booleans.

otra clasificación muy común los clasifica según su condición entre ellos: cuantitativos, categóricos y ordinales (como se resume en la figura 7).

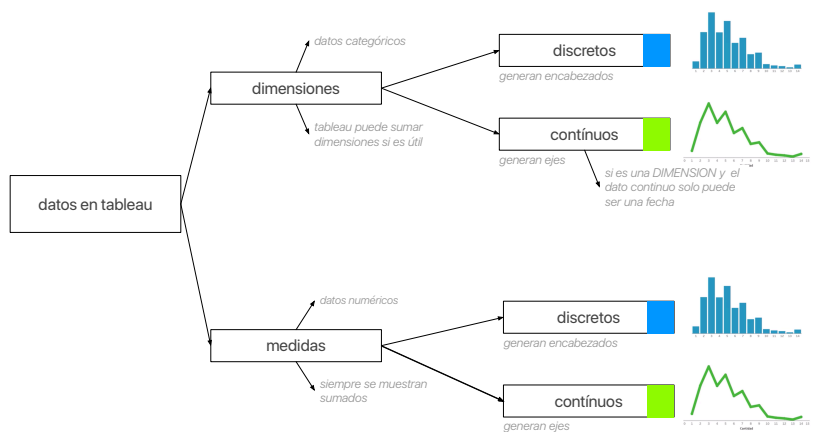
fig 7. clasificación típica de los datos según sus características



sin embargo, para nuestro objetivo vamos a usar una clasificación ajustada para la visualización de datos y que se usa como un estándar *de facto* en la industria; de hecho, es **la clasificación que usa la herramienta tableau** que es una de las herramientas más usadas en la actualidad.

la figura 8 resume esta clasificación. se divide principalmente en los **datos categóricos y los datos numéricos** (aquí dimensiones y medidas); en ambos casos se subdividen en discretos y continuos.

fig 8 clasificación típica de los datos según tableau



cuáles datos son clasificables en una u otra categoría (**discretos o continuos**) en realidad es muy variable y depende de cada caso.

puede que en la visualización que estamos haciendo sea útil convertir un dato continuo en uno discreto y entonces se agruparían, por ejemplo, las medidas diarias en promedios mensuales o semanales.

por supuesto el caso opuesto también sucede, es decir, puede que queramos convertir un dato discreto en un continuo, como por ejemplo,

cuando tomamos los datos de un mes y los distribuimos uniformemente en días o minutos.

estas decisiones se toman en función de cómo queremos representar los datos en la visualización, y eso por supuesto tiene que ver con qué queremos saber de esos datos, comparar, analizar.

sin embargo, el concepto de cuál dato pertenece a cuál tipo es determinante para entender las relaciones entre ellos y así poder tomar decisiones de como visualizarlos.

en resumen podemos pensar en los datos como "dimensiones" que serían esencialmente los datos cualitativos o categóricos y los datos tipo "medidas" que serían los datos numéricos. en ambos casos los datos son "transformables" en datos discretos y continuos, teniendo el cuidado de entender con precisión qué significa esto para cada tipo. la figura 9 nos muestra un resumen.

fig 9 clasificación general de los datos

	contínuos	discretos
dimensiones	dimensiones/contínuas (fechas)	dimensiones/discretas (categorias)
medidas	medidas/contínuas (valores)	medidas/discretas (rangos)

2.2. relaciones entre datos

una vez teniendo claro cuáles datos tenemos y de qué tipo son, pasamos a analizar qué relaciones podrían haber entre ellos. este paso es crucial para el éxito del proyecto.

hablamos de un análisis de cuáles de esos datos podrían funcionar como "entradas", "eventos" u "**observaciones**" (los individuos, las identidades), que son como las "tomas" de cada dato, **las filas o renglones** de la tabla final.

por ejemplo: cada medida que se hizo de un huracán un día y hora específico, o cada caso o paciente en un estudio de salud, cada país en un análisis de producción de CO2, cosas así.

por otro lado están las características o "**propiedades**" de estas "observaciones", éstas son las columnas de la tabla. por ejemplo: velocidad del huracán en ese día y hora, o nivel socioeconómico del paciente específico o la cantidad de toneladas de emisiones de CO2 en un país, en un año en particular.

cómo se ve, las "**observaciones**" pueden contener más de un dato que las defina como tal, es decir para definir un evento, **una "fila" de la tabla** por así decirlo, a menudo no basta con un dato. por ejemplo una observación de un huracán en realidad es una toma de sus características (velocidad, categoría, coordenadas, etc.) en un día en específico; es decir la muestra es el año, el mes, el día y el nombre del huracán; en el caso de nuestro ejemplo de cantidad de toneladas CO2 por país, pues hace falta una precisión de fecha: por mes? por año? y eso representaría en conjunto una "muestra" o "entrada" de la tabla.

así que el primer análisis es tener claro **cuáles datos son observaciones y cuáles propiedades** en cada una de las bases de datos que

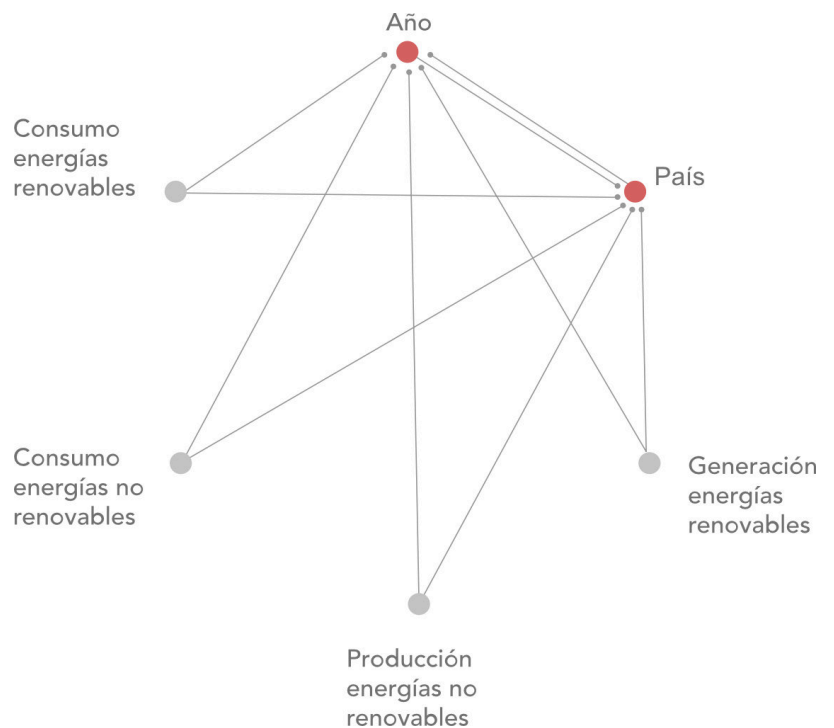
necesitamos para nuestros proyectos. esto en primera instancia parece muy lógico pero en realidad depende de lo que se desee averiguar.

para explicar esto vamos a poner un ejemplo, en un estudio que hicimos acerca del tráfico vehicular en costa rica [1] teníamos los datos de la cantidad de vehículos en cada calle o tramo definido por minuto.

uno podría pensar que las observaciones son cada entrada de datos de un vehículo en una posición (o calle específica) en un día y hora también específicos. ahora imaginemos que nos interesa saber si ¿es cierto que los viernes son los días de mayor tráfico?, es posible que nos convenga ordenar los datos por días de la semana y en ese caso las observaciones sean los días de la semana por horas del día y las condiciones de cantidad de vehículos pasen a ser propiedades de esa observación.

estas relaciones entre los datos terminan en un simple gráfico que la mayoría de las veces se representa como un gráfico de cuerdas. a continuación veremos algunos ejemplos.

fig 10 relación entre datos para un proyecto sobre la generación y consumo de energías renovables y no renovables por país, melany rodríguez montero TEC 1S-2021. [12]

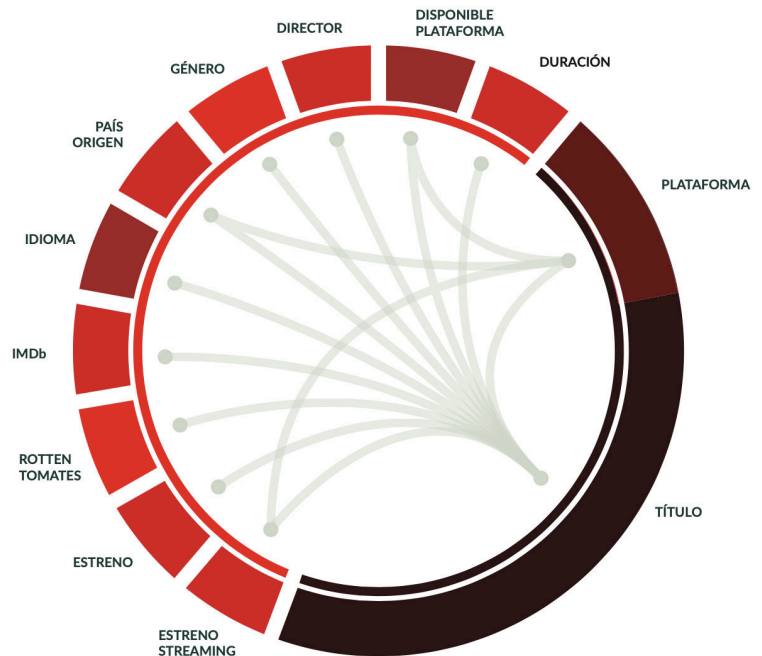


la figura 10, ejemplifica la relación entre los datos para un proyecto que tenía como objetivo saber cuáles países han avanzado más en su conversión a energías renovables en los últimos años.

cómo se ve, en este caso se definió que el país, en conjunto con el año, define la observación. las características serían entonces el consumo y producción (nótese que son cosas distintas) de cada país en cada año. con esta clasificación fue fácil hacer una visualización donde se pudiera ver qué países han contribuido más que otros en este sentido.

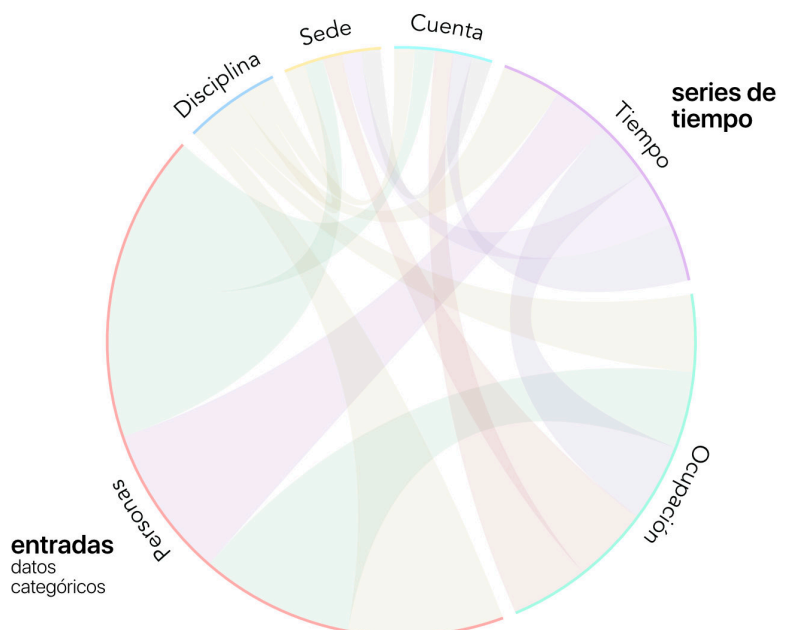
en la figura 11 vemos otro ejemplo, en este caso se trata de un proyecto para analizar la participación de otros países (fuera de estados unidos) en la producción de películas publicadas en la plataforma de streaming netflix. en este caso las películas fueron seleccionadas como las observaciones y las características serían datos como duración, título, país, etc.

fig. 11 relación entre datos para un proyecto sobre la cantidad de películas publicadas por país, mauricio vargas ortiz TEC 1S-2021



en la figura 12, otro ejemplo, en este caso para un análisis del personal de una empresa y datos específicos como: disciplina de procedencia, en qué sede trabajan o qué ocupación desempeñan. como se ve los "empleados" son las "observaciones" y las propiedades son los otros datos mencionados, en este caso el gráfico además incluye las cantidades, el tamaño de los arcos representa la cantidad relativa de cada propiedad en el total.

fig. 12 relación entre datos de un proyecto para un análisis del personal de una empresa y datos específicos como: disciplina de procedencia, en qué sede trabajan o qué ocupación desempeñan, mariana agüero soto TEC 2S-2018



3. definición de las preguntas objetivo POs

el siguiente paso es definir con cuidado las preguntas de investigación que se deben de responder con el proyecto, usamos la metodología y clasificación planteada en el artículo *What for: Classification of Visual Paradigms* [4].

las preguntas objetivo (POs) son **la información que vamos a extraer de los datos**. es el análisis posterior al realizado en el paso anterior (2. tipos de datos y sus relaciones), hecho en forma de preguntas que podrá contestar de manera intuitiva y eficiente la futura visualización.

toda visualización debe tener claro, antes de su diseño, qué preguntas va a responder, ¿para qué se hace el proyecto?; generalmente esta selección se hace en conjunto con los "expertos del dominio".

llamamos "**expertos del dominio**" a los "dueños de los datos", los **científicos o profesionales que trabajan todos los días con los datos** que estamos analizando. se trata de visualizar necesidades que ellos hayan considerado valiosas, en caso de un proyecto científico, serán los expertos del tema y en caso de un proyecto comercial serán los gerentes o estrategas que quieren saber comportamientos de sus productos o tendencias.

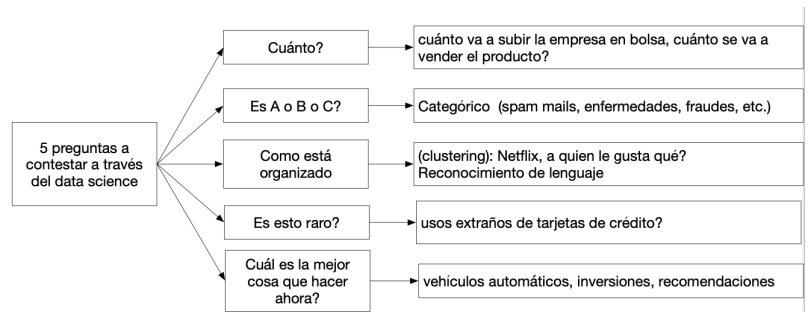
el resultado de esta etapa es muy simple: una pregunta principal y algunas preguntas secundarias; todas ellas que serán respondidas por la visualización y **serán usadas para validar el proyecto**.

sin embargo, eso no significa que es fácil llegar a ellas, la precisión en su redacción debe dar cuenta del análisis que se ha hecho de los datos, **no se deben preguntar cosas que no se pueden responder a partir de los datos** que se tienen, **ni tampoco las preguntas pueden ser tan obvias que se ven en los datos sin análisis**.

como dicta el modelo DIKW [5], de los datos tratamos de adquirir conocimiento, éste es en última instancia, la razón de ser de todo proyecto de visualización. no se trata de solo graficar los datos, se trata de, a través de su análisis y sus relaciones, seamos capaces de generar conocimiento.

mark david en su charla "*data science for every one*" resume que hay cinco tipos de preguntas a contestar en *data science*, (1) cuánto?, (2) es "A" o "B"?, (3) cómo está organizado?, (4) es esto raro?, (5) cuál es la mejor cosa que hacer ahora?, la figura 13 ofrece algunos ejemplos de esta clasificación.:

fig. 13 tipos de preguntas que generalmente se contestan a través de la visualizaciones



por ejemplo para un proyecto de visualización de datos sobre amenazas naturales en costa rica, las preguntas objetivo se veían así:

1. *¿Cuáles fueron las zonas más afectadas?*
 - 1.1. *¿Cuántas afectaciones hubo por provincia?*
 - 1.2. *¿Cuántas afectaciones hubo por cantón?*
2. *¿Cuándo sucedieron las afectaciones?*
3. *¿Cuánta y qué información se obtuvo en los reportes?*

*estudiante Pablo Calderón Bravo, 2S 2019

como se ve, algunas preguntas tienen preguntas subordinadas y otras no, puede que haya una pregunta principal y el resto sean profundizaciones de la misma, en todo caso lo que importa es que las preguntas objetivo resumen las necesidades que va a resolver el diseño.

otro ejemplo para un proyecto sobre el uso y desperdicio de la producción de energía en nuestro país fue:

Pregunta Objetivo Principal

¿Cuál es el grado de sub-utilización de las centrales productoras de energía?

Preguntas Secundarias

¿Qué época o estación del año presenta mayor grado de sub utilización de cada central eléctrica?

¿Cuál es la relación entre la producción solicitada y el grado de sub utilización?

¿Cuál es la diferencia entre la energía real generada y el grado de sub utilización?

*estudiante Josué Porras Fernández, 1S 2019

estas preguntas son importantes pues se convierten en **la guía principal del diseño**, representan los "**casos de uso**" definitivos a los cuáles responderá la visualización.

de hecho, **las POs son parte fundamental de la validación del diseño**, es decir, se usan como base para calificar la eficiencia de las propuestas.

por lo anterior hay que ser muy cuidadoso en la redacción de las POs pues de ellas depende el éxito del proyecto.

4. data cleaning (EDA) y data parsing

4.1. exploratory data analysis (EDA)

para este momento se tienen identificados los sets de datos que se van a usar, las relaciones entre ellos y la información que pretendemos extraer del proyecto.

el siguiente paso es analizar los set de datos y ver si se comportan como uno espera, este proceso se le llama EDA (Exploratory Data Analysis) y contempla dos fases:

4.1.1. completitud

la idea de este paso es definir **qué rangos de datos vamos a usar** en función de **qué tan completos** están los datos, a veces hay que **eliminar un rango o acortar otro** porque en esos trechos los datos no están completos.

es muy común que en lapsos de tiempo, por ejemplo, tengamos que **reducir el rango deseado** porque al final o al inicio del rango deseado no hay suficientes datos o están incompletos. inclusive puede ser que **eliminemos "un año" o "un mes"** porque no hay datos en él o porque solo presenta parte de ellos.

la otra opción es "rellenar" los datos faltantes. esto debe hacerse con mucho cuidado pues no queremos alterar la veracidad de los mismos. en este caso se usan **estrategias de interpolación o extrapolación** para tratar de mantener la tendencia de los datos y al mismo tiempo que estén completos. la decisión queda a discreción del diseñador, sin embargo, **en el reporte debe quedar muy claro cómo y en dónde** se pusieron datos que no vienen de la realidad.

4.1.2. coherencia

la otra actividad típica del EDA es ver si los datos son coherentes como se supone de deben ser. **los datos a menudo vienen "contaminados" por errores** que han sucedido en el tiempo, errores de transcripción por razones ajenas al fenómeno que se está estudiando, etc.

por ejemplo, si tenemos una columna con "edades de los pacientes" y de pronto aparecen valores muy grandes, mayores de 120, pues está claro que los datos están contaminados. entre más cantidad de datos se tengan, más probabilidad hay de que aparezcan estos errores.

fig. 14 típico gráfico de exploración de datos durante la fase EDA, mostrando actividad sísmica en varias estaciones, la mayoría de ellas se comportan en forma coherente y algunas de ellas presentan comportamientos no esperados. autor ph.d. jorge monge fallas [2]



la figura 14 muestra un ejemplo de este tipo de análisis, los colores representan estaciones sísmicas y su actividad en un lapso de tiempo, como se ve, la mayoría de las estaciones tienen comportamientos similares, sin embargo unas pocas de ellas no; saltos entre datos y

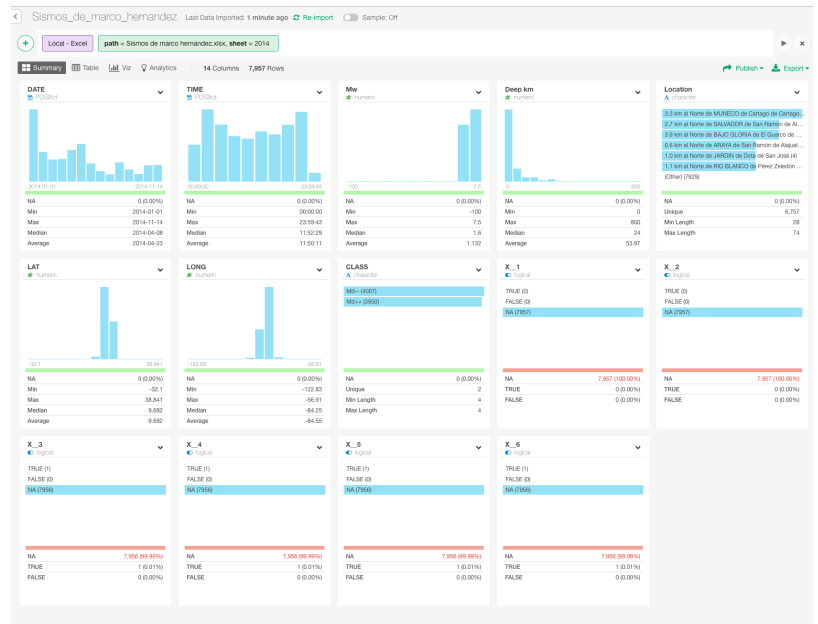
medidas muy altas en comparación a las demás nos insinúan que hay problemas en los datos.

en estos casos, si es posible, se consulta a los expertos del dominio y se define que hacer con esos datos, modificarlos o eliminarlos. a menudo hay explicaciones para el comportamiento anómalo de los datos, por ejemplo, en el caso de las estaciones fue que se re-calibraron y a partir de esa fecha la "base" de los datos cambió. en estos casos es fácil modificar los datos para que sigan siendo útiles, otros eventos clásicos son: se acabó la batería, el día estaba nublado y la estación era solar, un camión chocó con la estación, etc. etc.

§§§

para realizar este análisis se crean muchos tipos de gráficos que son la mejor manera de tener la vista total de todos los datos de una sola vez. herramientas convencionales como "excel" o "apple numbers" se pueden usar en estos casos, siempre que la cantidad de datos lo permita, sino hay un sinnúmero de herramientas especializadas como "exploratory" (<https://exploratory.io>) que, como su nombre lo indica, es una herramienta especializada en esta fase y tiene una versión libre.

fig. 15 resumen de datos en una fase de EDA realizado en la herramienta "exploratory" (<https://exploratory.io>)



en la figura 15 vemos uno de estos análisis realizados en esta herramienta, como vemos, todas las características de cada observación se ven graficadas mostrando su distribución en una pequeña gráfica resumen. abajo de cada una de estas gráficas se encuentra valores, como el valor mínimo de esa propiedad, el valor máximo, la mediana y el promedio, además del porcentaje de "NA" (*non a number*) que es una característica muy valiosa en estos casos, pues representa los datos que no se han reconocido como "medidas".

con un resumen de este tipo es fácil ver dónde debemos profundizar para tomar las decisiones que hemos discutido en los apartados anteriores de eliminación o modificación de los datos anómalos.

4.2 data parsing

una vez con los datos "limpios" pasamos a analizar: cómo los relacionamos?, esto en caso de que vengan de varias fuentes y por lo tanto en "sets" o "tablas" aparte; lo que sucede la mayoría de las veces.

es una cosa simple pero hay que hacerla con cuidado, se trata de encontrar que parte de cada set de datos es compatible con los otros para así poder, por así decirlo, hacer una sola tabla.

poniendo un ejemplo digamos que estamos interesados en una comparación entre el área de la mayoría de los países del mundo y hemos definido que queremos hacer un gráfico como el de la figura 16:

fig. 16 visualización de la mayoría de los países del mundo de acuerdo a sus áreas

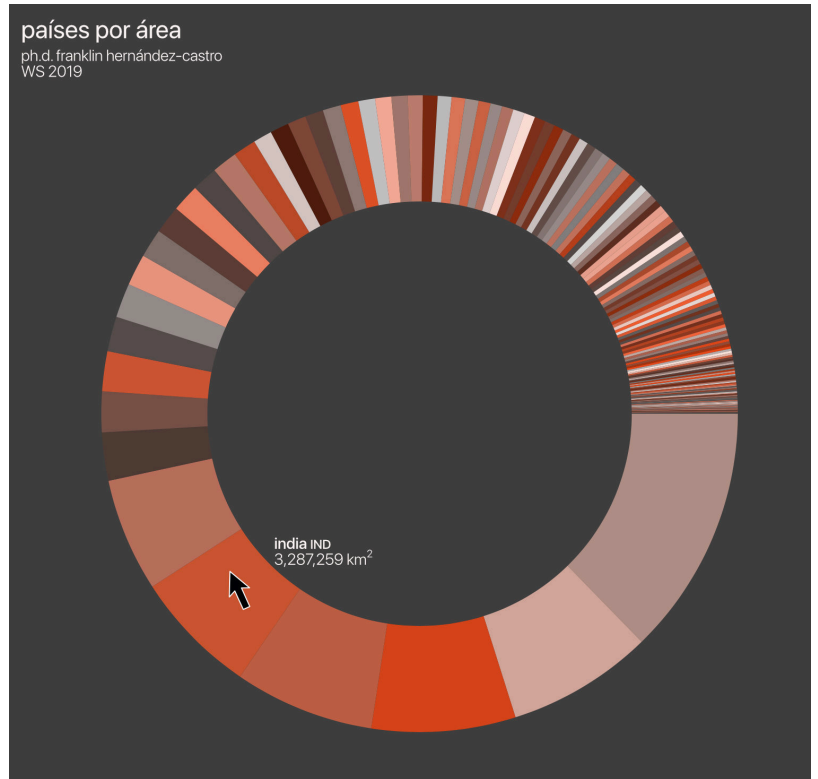


fig. 17 sets de datos involucrados en la visualización de la figura 16

set de datos A		set de datos B		
ISO_A3	sqr_Km	CountryName	ISO3	population
ABW	180	China	CHN	1386395000
AFG	652860	India	IND	1339180127
AGO	1246700	United States	USA	325719178
ALB	28750	Indonesia	IDN	263991379
AND	470	Brazil	BRA	209288278
ARE	83600	Pakistan	PAK	197015955
ARG	2780400	Nigeria	NGA	190886311
ARM	29740	Bangladesh	BGD	164669751
ASM	200	Russian Federation	RUS	144495044
ATG	440	Mexico	MEX	129163276
AUS	7741220	Japan	JPN	126785797
AUT	83879	Ethiopia	ETH	104957438
AZF	86600	Philippines	PHL	104918090
		Egypt	EGY	97553151
		Vietnam	VNM	95540800
		Germany	DEU	82695000

fig. 18 set de datos involucrados en la visualización de la figura 16

CountryName	ISO_A3	sqrKm	population
Russian Federation	RUS	17098250	144495044
Canada	CAN	9984670	36708083
United States	USA	9831510	325719178
China	CHN	9562910	1386395000
Brazil	BRA	8515770	209288278
Australia	AUS	7741220	24598933
India	IND	3287259	1339180127
Argentina	ARG	2780400	44271041
Kazakhstan	KAZ	2724902	18037646
Algeria	DZA	2381740	41318142
Congo	COD	2344860	81339988
Saudi Arabia	SAU	2149690	32938213
Mexico	MEX	1964375	129163276

ahora imaginemos que en nuestra búsqueda de datos hemos encontrado dos tablas; una de ellas con el área que se buscaba y otra con la población. como se ve en la figura 17.

en este caso lo que se busca es un modo de combinar las dos tablas en una sola; propiedades como los nombres de los países no son confiables, pueden variar mucho con los idiomas o los caracteres especiales en algunos idiomas y otros casos (venezuela, república bolivariana de venezuela, r.b. de venezuela?), por otro lado, obviamente el orden es distinto en ambas tablas e igual no se puede partir de los nombres que son los que definirían el orden.

sin embargo, en ambas tablas se encuentran los códigos ISO3 de cada país, este es un estándar internacional para denominar cada país, por lo tanto es una magnífica llave para combinar ambas tablas y queden juntas como en la figura 18.

por cierto, se puede ver en el gráfico que unos 6 países de casi 200 poseen la mitad del territorio habitado.

estas operaciones son posible con varias herramientas; dependiendo de la cantidad de datos involucrados se puede hacer con herramientas clásicas tipo excel. si son muchos datos, aplicaciones como *tableau* o *exploratory* pueden hacer ese trabajo aun con muchísimos datos.

por supuesto esto es solo un ejemplo hecho para fines pedagógicos y entender el concepto de *parsing data*. dependiendo de lo que se esté investigando los procesos de parsing son más complejos, en especial hay que poner atención a qué escalas están los datos y si son comparables para usarlas en el mismo ejemplo. la figura 19 muestra un proceso de parsing data para el proyecto de visualización de huracanes [3].

fig. 19 análisis de parsing entre sets de datos para un proyecto de visualización de la influencia de los huracanes en precipitaciones anómalas en regiones de centroamérica y el caribe

tabla A

estaciones.csv

X1	longitude	latitude
1	-88.3	17.5
2	-89	17.183
3	-88.317	16.983
4	-88.7	17.25
5	-91.503	15.317
6	-89.706	14.234
7	-91.48	14.87
8	-89.584	14.964
9	-88.583	15.733
10	-91.678	14.535
11	-90.408	15.467
12	-90.64	14.801
13	-90.533	14.583
14	-89.867	16.517
15	-89.033	14.533
16	-90.786	14.301
17	-91.05	14.03
18	-91.024	15.044
19	-90.614	15.109
20	-89.876	14.493
21	-91.731	14.633

lista de estaciones y su posición

la tabla A (de estaciones) es simple y no cambia pues solo son las estaciones y sus latitudes y longitudes

la tabla B (precipitacionesMayores) tiene los días en que las estaciones tuvieron altas precipitaciones (más de su percentil 90) cada día puede tener varias estaciones

la tabla C (de fechas) es la tabla de los datos iniciales donde está cada día que se hizo una toma de un huracán con sus respectivas latitud, longitud, fecha, hora, evento (que es el huracán al que pertenece esa fecha), categoría y velocidad

la tabla D ("deDíasHuracanesPorEstación") es un resumen de las estaciones que tuvieron precipitaciones mayores del 90% en esos días, que son días que aparecen en la tabla anterior (tabla C) como que había la presencia de un huracán. esta tabla se genera en processing a partir de la tabla de fechas (tabla C) y la de "precipitacionesMayores" (tabla B) donde están los días que las estaciones tuvieron precipitaciones mayores a su percentil 90

en un cambio de datos las tablas C y D se deben generar de nuevo.

tabla B

presipitacionesMayores.csv

estacion	day	month	year	presipitacion
21	1	1	1970	0.7
62	1	1	1970	3.8
66	1	1	1970	5.4
51	2	1	1970	3.1
65	2	1	1970	10.2
27	3	1	1970	11.774316
28	3	1	1970	8.397335
33	3	1	1970	96.0
42	3	1	1970	16.32813
25	4	1	1970	10.4
33	4	1	1970	26.0
23	5	1	1970	37.1
27	5	1	1970	5.5237885
28	5	1	1970	4.6824675
50	5	1	1970	25.081371

lista de días en los que una o más estaciones tuvieron precipitaciones mayores de su percentil 90

tabla C

1970_2010_28_02_2019clean_filtrados_conEventos.csv

longitud	latitud	windkph	newEvent	day	month	year	type	secuencia	hour	press
-79	11.5	0	0	17	5	1970	-1	1	18	0
-79.2	11.7	3	0	18	5	1970	-1	2	0	0
-79.7	12.1	11	0	18	5	1970	-1	3	6	0
-80.1	12.3	7	0	18	5	1970	-1	4	12	1007
-80.5	12.5	7	0	18	5	1970	-1	5	18	0
-81	13	11	0	19	5	1970	-1	6	0	0
-81.5	13.5	11	0	19	5	1970	-1	7	6	0
-82	14	11	0	19	5	1970	-1	8	12	1005
-82.5	14.5	11	0	19	5	1970	-1	9	18	0
-82.5	15.5	18	0	20	5	1970	0	10	0	0
-82.3	16.2	12	0	20	5	1970	0	11	6	0
-81.9	16.8	12	0	20	5	1970	1	12	12	0
-81.6	17.2	7	0	20	5	1970	1	13	18	993
-81	18	16	0	21	5	1970	0	14	0	0
-80.3	18.2	11	0	21	5	1970	0	15	6	0
-80.2	18.5	5	0	21	5	1970	0	16	12	0
-80	19	9	0	21	5	1970	0	17	18	998
-80.1	19.6	11	0	22	5	1970	0	18	0	1004

lista de días de ocurrencias de huracanes

tabla D

tablaDiasHuracanesPorEstacion_01_03_2019.csv

estacion	day	month	year	presipitacion	fecha
7	17	5	1970	13.815146	0
8	17	5	1970	28.0	0
10	17	5	1970	19.892172	0
11	17	5	1970	19.343208	0
14	17	5	1970	19.38659	0
22	17	5	1970	35.3	0
25	17	5	1970	15.0	0
27	17	5	1970	33.151447	0
28	17	5	1970	18.24705	0
29	17	5	1970	27.0	0
34	17	5	1970	21.5	0
36	17	5	1970	20.5	0
37	17	5	1970	34.12788	0
38	17	5	1970	19.7	0
40	17	5	1970	50.60766	0

esta fecha corresponde al consecutivo de la tabla de días de ocurrencias de huracanes

lista de días de estaciones y su posición

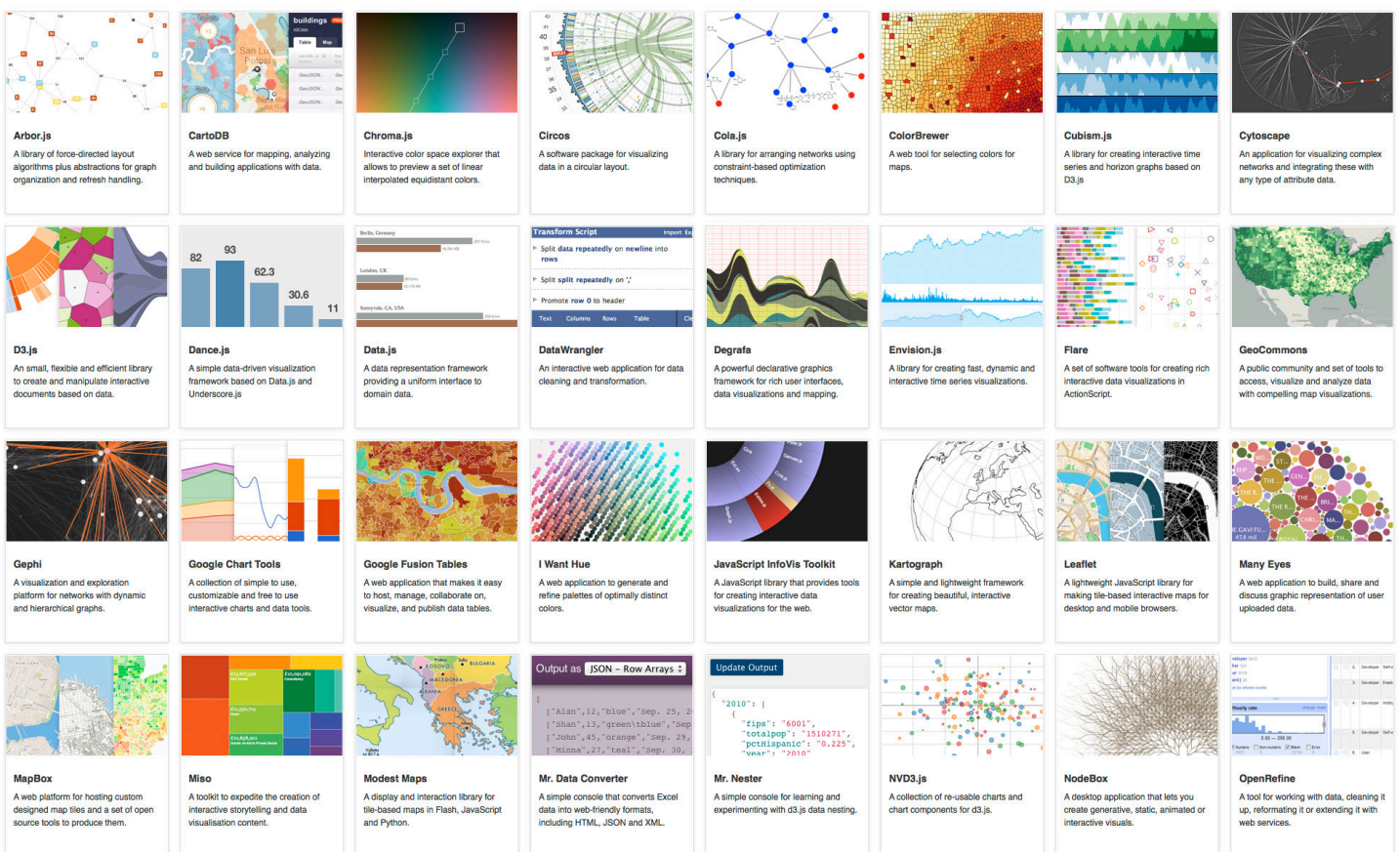
5. exploración de paradigmas

en esta etapa queda la delicada tarea de escoger qué tipo de visualización queremos hacer. esta es una tarea compleja que depende en primera instancia de lo que se desea responder con la visualización, es decir, con las preguntas objetivo. para iniciar el análisis usamos la clasificación descrita en "What for: Classification of Visual Paradigms." [4], donde se clasifican los paradigmas por objetivo.

hay muchos paradigmas posibles y hay también muchas clasificaciones pero pocas se centran en el "para qué" es mejor cada tipo de gráfico.

por ejemplo este catálogo suizo (<http://selection.datavisualization.ch>) es un de los más completos, con información de fuentes y ejemplos (ver figura 20).

fig. 20 catálogo suizo (<http://selection.datavisualization.ch>)



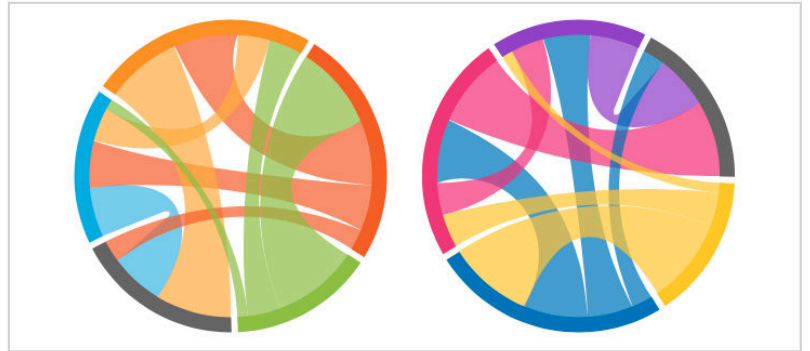
otro catálogo excelente es el de severino ribecca (<https://datavizcatalogue.com/index.html>) ver figura 21, no solo la clasificación sino las fichas de cada paradigma donde se encuentran explicaciones, links y ejemplos de cada uno.

en la figura 22 está el catálogo de la agencia leferdio de dinamarca (<https://datavizproject.com>), que también es excelente con ejemplos en cada paradigma y una propuesta de cómo serían los datos necesarios para cada caso.

fig. 21 imágenes del catálogo excelente es el de severino ribecca (<https://datavizcatalogue.com/index.html>)



Chord Diagram



Description

This type of diagram visualises the inter-relationships between entities. The connections between entities are used to display that they share something in common. This makes Chord Diagrams ideal for comparing the similarities within a dataset or between different groups of data.

Nodes are arranged along a circle, with the relationships between points connected to each other either through the use of arcs or Bézier curves. Values are assigned to each

Anatomy

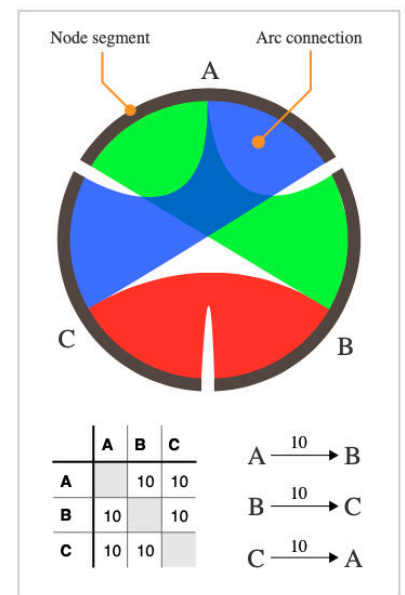
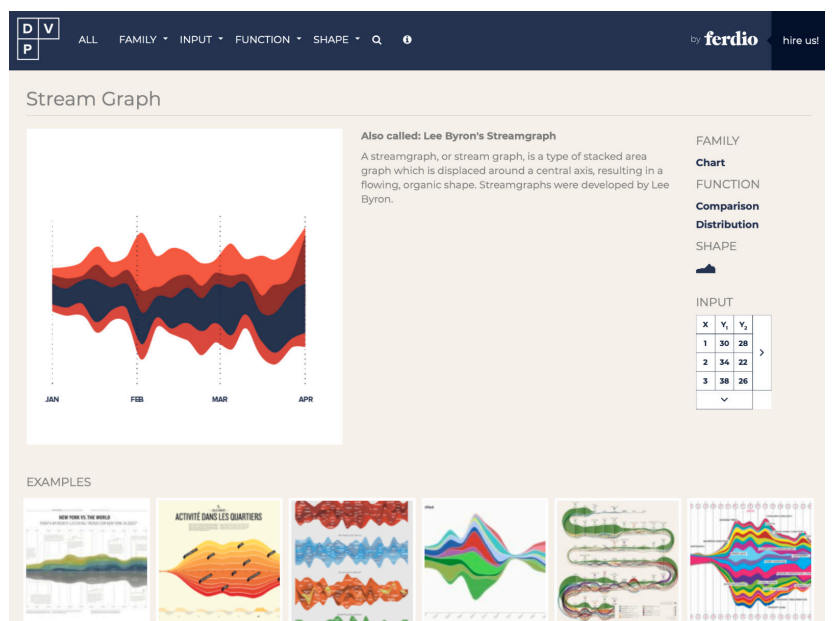


fig. 22 catálogo de la agencia ferdio de dinamarca (<https://datavizproject.com>),



basados en:

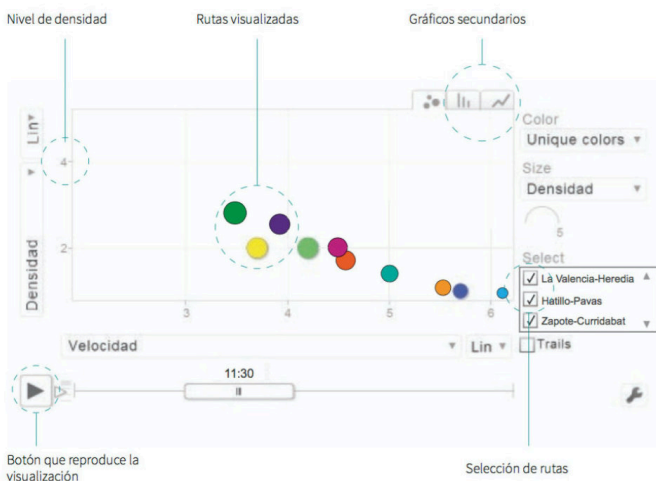
- (1) las preguntas objetivo,
- (2) la clasificación descrita en el artículo mencionado [4], y
- (3) los catálogos y demás fuentes sobre tipos de paradigmas gráficos que se encuentran disponibles

se pasa a pre-seleccionar tres o cuatro posibles paradigmas que podrían resultar útiles para responder lo que queremos. después se toma una pequeña muestra de los datos y se hacen pruebas con estos paradigmas.

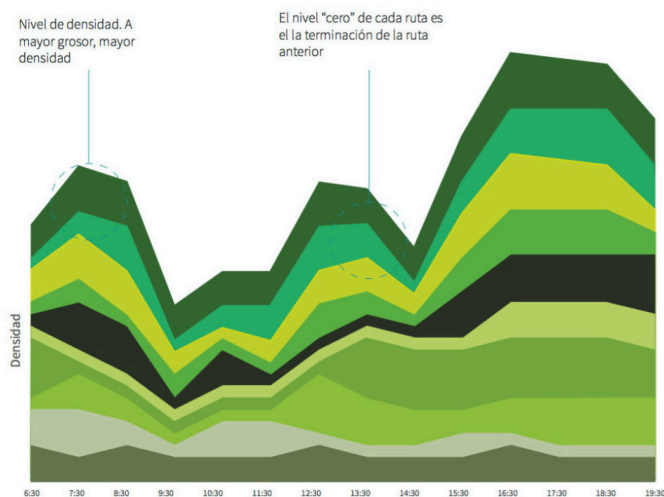
esto se podría hacer en una herramienta especializada o hasta en un programa de diseño como adobe ilustrador®, affinity designer® o figma®, por ejemplo, pues el tamaño de la muestra no es grande, pero claro si se hacen en una herramienta para visualización de datos es más versátil y de paso se va preparando la "implementación" final del proyecto.

fig. 23 las cuatro alternativas que se probaron en el proyecto de visualización del tráfico con los datos de la aplicación WAZE en costa rica

gapminder



streamgraph



veamos un ejemplo:

en el proyecto descrito en el artículo "Traffic Data Visualization in Costa Rica: A Visualization of top 100 routes with the highest traffic density in Costa Rica" [1], las pregunta objetivo se veían así:

1. cuál es el tramo con mayor y menor densidad durante un día específico en costa rica?

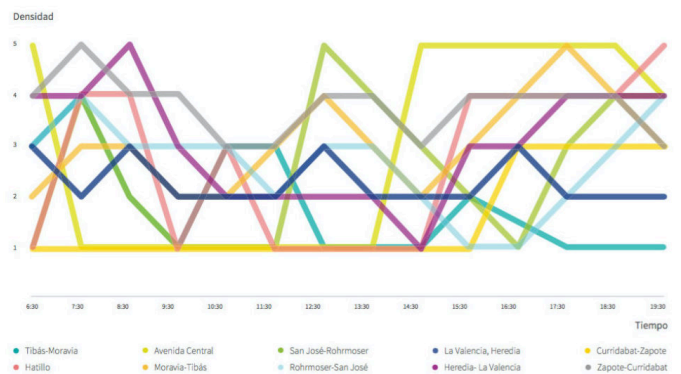
1. comparación de la densidad de dos tramos de ruta durante un día?

2. cuáles son los días con menor densidad en una semana en un tramo específico?

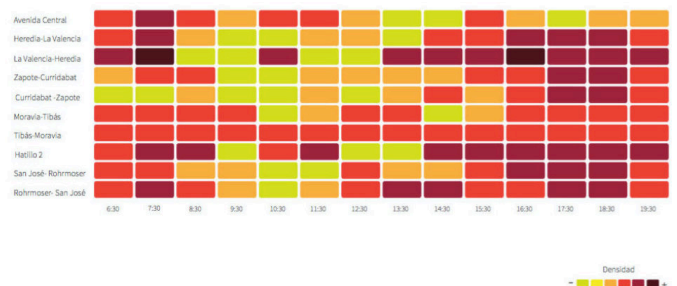
3. cuál es el viernes con mayor densidad de tráfico de un mes en un tramo específico?

así que se buscaron cuatro paradigmas que pudieran responder esas preguntas. (figura 23)

bump chart



heatmap



lo primero que notamos es que las cuatro alternativas, como era de esperarse, tienen un énfasis en mostrar datos que transcurren en el tiempo, como es la naturaleza de los datos de este proyecto.

a partir de estas propuestas se pasa a probar el modo en que cada propuesta responde a las preguntas objetivo y qué tan fáciles son de entender. en otras palabras hacemos una prueba rápida de usabilidad con sus componentes clásicos de eficiencia, eficacia y satisfacción [7][8].

esto se hace haciendo una prueba rápida de pocos usuarios, digamos unos 5, (para detalles sobre pruebas heurísticas de este tipo [9][10][11])

los resultados de las pruebas se resumen generalmente en dos tablas (figura 24), una que resume el modo en que cada paradigma respondió las preguntas objetivo y otra donde se resume la satisfacción (componente de la usabilidad) con la que se realizó la prueba por parte de los testers. en esta etapa no se pide una prueba ISO estándar sino que algo más rápido, más adelante en la validación del proyecto se procede con algo más de profundidad a validar el proyecto.

Table 1. Evaluation of the paradigms in relation to the case studies

Case	paradigm			
	Bump Chart	Gap minder	Stream graph	Heat Map
Case 1: Most and least density in a route section during one day	3	3	2	5
Case 2: Density comparison of two route sections during one day	3	3	2	4
Case 3: The days with the least density in one week	3	3	3	4
Case 4: The Friday with the highest density traffic from one month	3	3	2	2
Total	12	12	9	15

Table 2. Evaluation of the paradigms in relation to usability requirements

Requirements	paradigm			
	Bump Chart	Gap minder	Stream graph	Heat Map
Easy to read	3	4	3	5
Easy to use	4	4	4	4
Facilitates analysis	3	3	3	4
Intuitive	3	4	3	4
Efficient	3	3	4	4
Attractive	2	2	4	5
Allows comparison	3	2	5	4
Learning curve	4	4	3	4
Total	25	26	29	34

como se ve, en ambas tablas el paradigma *heat map* es el claro ganador, así que en este caso se procede a continuar con la implementación de la visualización en este paradigma.

fig. 24 las dos tablas que resumen la evaluación de paradigmas del proyecto de visualización de datos de tráfico WAZE

veamos otro ejemplo, este es el caso de analizar el modo en que los países han estado cambiando a las tecnologías renovables que ya habíamos citado antes. realizado por la estudiante melany rodríguez montero [12], en mi curso de diseño de dashboards.

las preguntas objetivo de este proyecto eran:

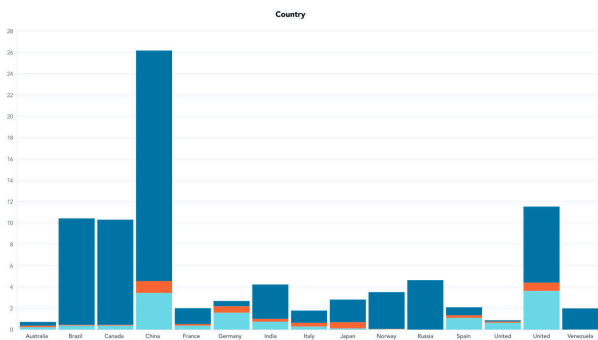
Cuál energía renovable predomina en el mundo y qué país lo maneja mejor?

¿Cuáles países se destacan en las diferentes energías renovables?

¿Cuáles países han transformado hacia energías renovables en forma destacable en los últimos 15 años?

como vemos, de nuevo el tiempo es importante en este proyecto, y como vimos en la figura 10, tenemos los países y su producción y consumo energético dividido en energías renovables y convencionales.

gráfico de barras



icicle

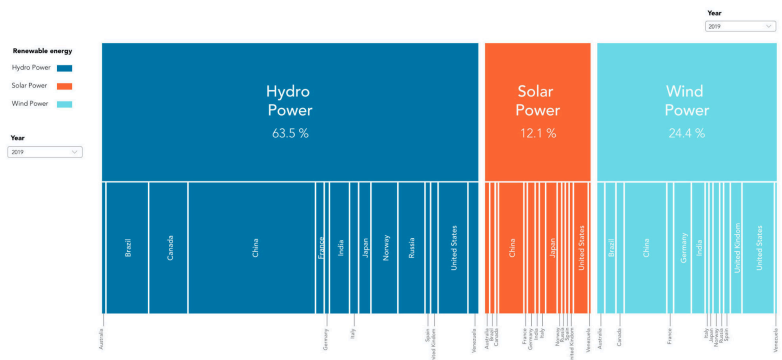


diagrama aluvial

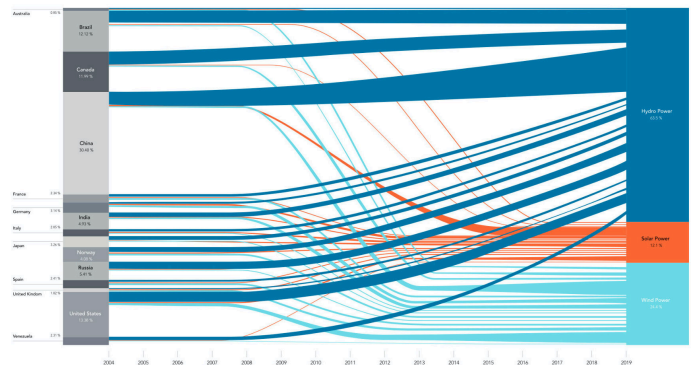


fig 25 alternativas de paradigmas usadas en el proyecto sobre la cantidad de energías renovables y no renovables por país, melany rodríguez montero TEC 1S-2021

en la figura 25 vemos que en este caso se usaron los paradigmas de diagramas de barras, icicle y diagrama aluvial, teniendo en cuenta que el tiempo es importante. sin embargo, en los dos primeros los "años" son un filtro, es decir el usuario debe de cambiar el año para verlo y, como es de esperar, es difícil recordar el comportamiento de los datos en el tiempo mientras se cambian, el diagrama aluvial es muy superior en este sentido.

así que no fue una sorpresa cuando este paradigma ganó en el resumen de las pruebas preliminares. ver figura 26.

fig. 26 tablas de evaluación de paradigmas proyecto sobre la cantidad de energías renovables y no renovables por país, melany rodríguez montero TEC 1S-2021

Caso de uso	Diagrama de barras	Diagrama Icicle	Diagrama aluvial
Caso 01: Visualizar cuál energía renovable predomina y qué país lo maneja mejor.	3	5	5
Caso 02: Observar cuáles países se destacan en las diferentes energías renovables.	4	5	4
Caso 03: Reconocer cuáles países han transicionado destacablemente en los últimos 15 años	1	1	5
Total	8	11	14

Requerimientos	Diagrama de barras	Diagrama Icicle	Diagrama aluvial
Facilidad de lectura	4	3	4
Facilidad de uso	4	4	5
Intutivo	3	4	4
Curva de aprendizaje	5	3	3
Facilita análisis	3	3	4
Permite comparaciones	4	2	4
Visualización específica	3	3	4
Visualización global	4	4	5
Total	30	26	33

§§§

en general este es el modo que se usa para seleccionar el paradigma que mejor se comporta para contestar los objetivos del proyecto.

6. implementación

una vez que se tiene claro cuál paradigma se desea implementar, se procede a esa etapa.

hay muchos modos de implementar un proyecto de visualización, cuando se trata de dashboard, generalmente se usan herramientas diseñadas para estos casos. "exploratory" (<https://exploratory.io>), tableau (<https://www.tableau.com>), powerBI (<https://powerbi.microsoft.com>), etc.

los catálogos de paradigmas como los mencionados en el apartado de cinco: severino ribecca (<https://datavizcatalogue.com/index.html>), catálogo suizo (<http://selection.datavisualization.ch>) o ferdio (<https://datavizproject.com>), a menudo traen también propuestas de

dónde se pueden implementar cada paradigma específico. tómesese en cuenta que hay herramientas muy específicas para casos como datos geográficos entre otros.

claro que algunos diseñadores prefieren empezar desde cero y programar la visualización en algún ambiente que lo permita, por ejemplo processing (<https://processing.org>), D3.js (<https://d3js.org>), openframeworks (<https://openframeworks.cc>), etc.

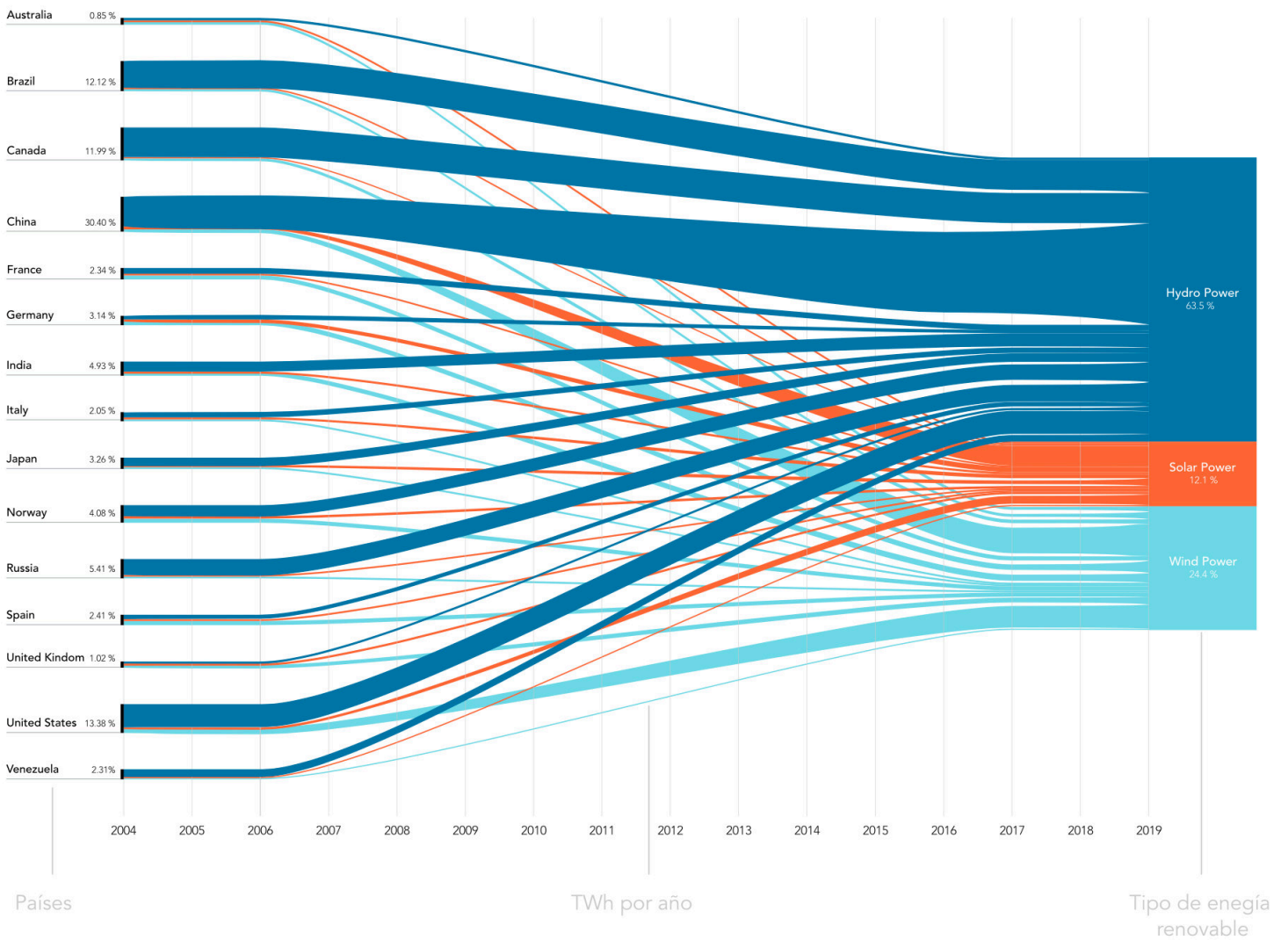
es importante en esta etapa manejar las escalas de color de manera que el usuario pueda entender de manera intuitiva si es una escala de color o el color es un identificador, para profundizar en el tema consultar [13][14][15].

fig. 27 implementación final del proyecto de visualización del tráfico con los datos de la aplicación WAZE en costa rica [1]



fig. 28 implementación final del proyecto sobre la cantidad de energías renovables y no renovables por país, melany rodríguez montero TEC 1S-2021 [12]

Renewable energy consumption



7. validación

una vez implementada la visualización procedemos a validarla con usuarios del dominio, generalmente para esto se usa una prueba ISO 9241-11 [7] para la parte de eficiencia y eficacia y una prueba SUS [11] [14] para completar la parte de satisfacción, terminando con las tablas estándar de estas normas ISO 25062 [15].

para hacer la validación en primera instancia se diseña una prueba heurística. en esta prueba las tareas están definidas por la preguntas objetivo, pues el fin es saber si la visualización sirve para responderlas.

durante la prueba se hacen las anotaciones de tiempo de duración por tarea, errores, solicitudes de asistencia, en fin eficiencia y eficacia (tiempo y completitud de tareas por usuario) y al finalizar la prueba se les pasa a los participantes el cuestionario SUS [11] [14] con el fin de determinar el grado de satisfacción según este estándar.

los resultados de la prueba se resumen en las tablas estandarizadas para este efecto. la primera es la tabla que resumen los participantes, la cantidad de participantes en nuestro caso puede ir de 5 a 8 por "persona" (en el sentido UX del término) [10].

la tabla de participantes oficialmente se ve así:

tabla de participantes

tester #	sexo	edad	educación	ocupación o rol que desempeña	experiencia profesional	experiencia en el manejo de computadoras	experiencia en el producto
T1							
T2							
...							
Tn							

según sea el caso, se puede adaptar, por ejemplo:

Tabla de participantes

	Género	Edad	Educación	Ocupación	Experiencia con el producto
P1	F	54	Secundaria completa	Contadora	—
P2	F	49	Secundaria incompleta	Secretaria	—
P3	F	46	MBA	CEO	—
P4	M	49	Secundaria completa	Economista	—
P5	M	24	Bachiller	Estudiante	—
P6	F	30	Secundaria completa	Entrenadora de tenis de campo	—
P7	M	19	Secundaria completa	Estudiante	—
P8	F	22	Bachiller	Estudiante	—
P9	F	33	Licenciada	Profesora de inglés	—
P10	M	15	Primaria completa	Estudiante	—

fig. 29 tabla de participantes del proyecto sobre la cantidad de energías renovables y no renovables por país, melany rodríguez montero TEC 1S-2021 [12]

el conjunto de tablas que sigue son las tablas por tarea, una tabla típica por tarea se vería así:

tarea n

tester #	tareas realizadas sin asistencia (efectividad % completo)	tareas realizadas con asistencia (efectividad % completo)	tiempo por tarea/tester (minutos)	cantidad de errores cometidos tarea/tester	cantidad de asistencias necesarias
1					
2					
...					
n					
promedio					
desviación estándar					
valor mínimo					
valor máximo					

ejemplo de tablas por tarea:

Tarea 01				Tarea 03				
	Tiempo (s)	Errores	Ayudas		Tiempo (s)	Errores	Ayudas	Observaciones
P1	5.3	0	0	P1	8.3	0	0	---
P2	4.2	0	0	P2	7.3	0	0	---
P3	3.5	0	0	P3	9.5	0	0	---
P4	4.3	0	0	P4	19.5	2	1	Confusión en con la pregunta. Aclaración de significado del cambio de grosor
P5	3.3	0	0	P5	8.6	0	0	---
P6	2.2	0	0	P6	10.3	0	1	Aclaración de significado del cambio de grosor
P7	4.2	0	0	P7	9.4	0	0	---
P8	2.8	0	0	P8	7.8	0	0	---
P9	4.6	0	0	P9	20.6	2	1	Entendió pregunta pero se fijo en el grosor final no cómo cambiaba.
P10	5.9	0	0	P10	15.9	1	1	Repetición de pregunta

fig. 30 ejemplo de tablas por tarea del proyecto sobre la cantidad de energías renovables y no renovables por país, melany rodríguez monterro TEC 1S-2021 [12]

la siguiente es una tabla donde se resumen todas las tareas por usuario, en forma general se ve así:

resumen

tester #	tareas realizadas sin asistencia (efectividad % completo)	tareas realizadas con asistencia (efectividad % completo)	tiempo por tarea/tester (minutos)	cantidad de errores cometidos tarea/tester	cantidad de asistencias necesarias
1					
2					
...					
n					
promedio					
desviación estándar					
valor mínimo					
valor máximo					

satisfacción

tester #	escala SUS
1	
2	
...	
n	
promedio	
desviación estándar	
valor mínimo	
valor máximo	

finalmente se documenta la tabla de resultados de la prueba SUS [11] [14], la tabla en forma general se ve como en la figura 31. por su parte la figura 32 muestra un ejemplo de esta tabla.

§§§

con estas tablas resumen queda completa la validación de la visualización. sin embargo, además de las tablas es aconsejable poner al menos una vistas de la visualización en cada caso de uso correspondiente a cada pregunta objetivo.

fig. 31 modelo de tabla general de satisfacción ISO 25062

Criterio	Totalemente en desacuerdo	En desacuerdo	Neutro	De acuerdo	Totalemente de acuerdo
1. Creo que me gustaría usar esta visualización con frecuencia.	3	5	2	0	0
2. Encontré la visualización innecesariamente compleja.	7	3	0	0	0
3. Encontré la visualización fácil de usar.	0	0	0	2	8
4. Necesito el apoyo profesional para entender la visualización.	9	1	0	0	0
5. Pienso que hay demasiada inconsistencia en esta visualización.	8	1	1	0	0
6. Creo que la mayoría de la gente aprendería a utilizar y entender esta visualización muy rápidamente.	0	0	0	1	9
7. Encontré la visualización muy complicada de entender.	9	1	0	0	0
8. Me sentí muy bien entendiendo el contenido de la visualización.	0	0	0	3	7
9. Tuve momentos de frustración al usar o entender la visualización.	9	1	0	0	0

Calificación SUS: 87.5

fig. 32 ejemplo de tabla de satisfacción en escala SUS del proyecto sobre la cantidad de energías renovables y no renovables por país, melany rodríguez monterero TEC 1S-2021 [12]

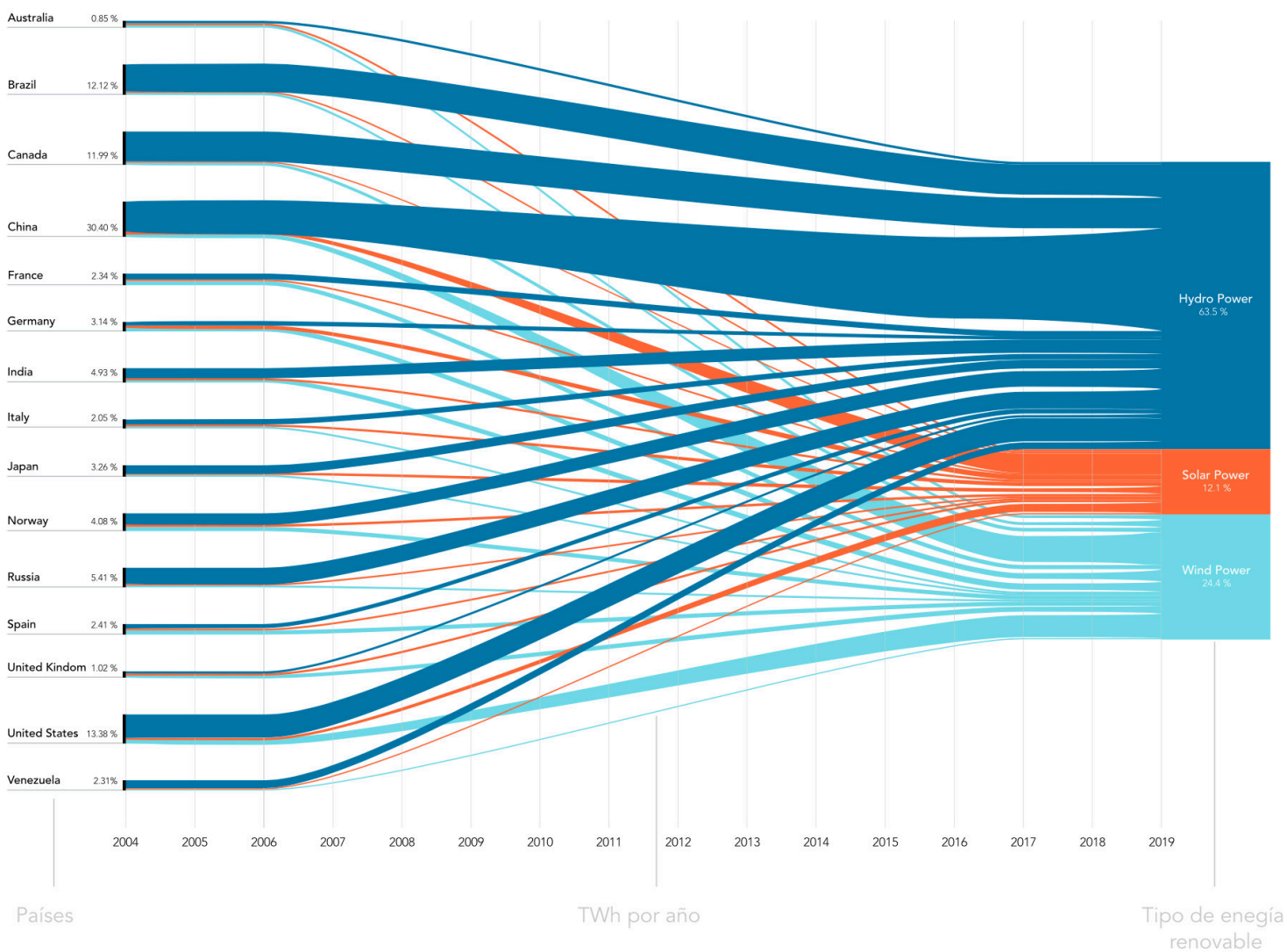
veamos un ejemplo, el proyecto de energías renovables tenía como POs las siguientes:

1. *Cuál energía renovable predomina en el mundo y qué país lo maneja mejor?*
2. *¿Cuáles países se destacan en las diferentes energías renovables?*
3. *¿Cuáles países han transformado hacia energías renovables en forma destacable en los últimos 15 años?*

fig. 33 ejemplo de respuesta a la pregunta objetivo ¿Cuál energía renovable predomina en el mundo y qué país lo maneja mejor? del proyecto sobre la cantidad de energías renovables TEC 1S-2021 [12]

la figura 33 nos muestra la visualización final contestando la primera pregunta objetivo de este proyecto, como se ve, la energía hidroeléctrica es la más usada, seguida de la eólica y en tercer lugar la energía solar.

Renewable energy consumption

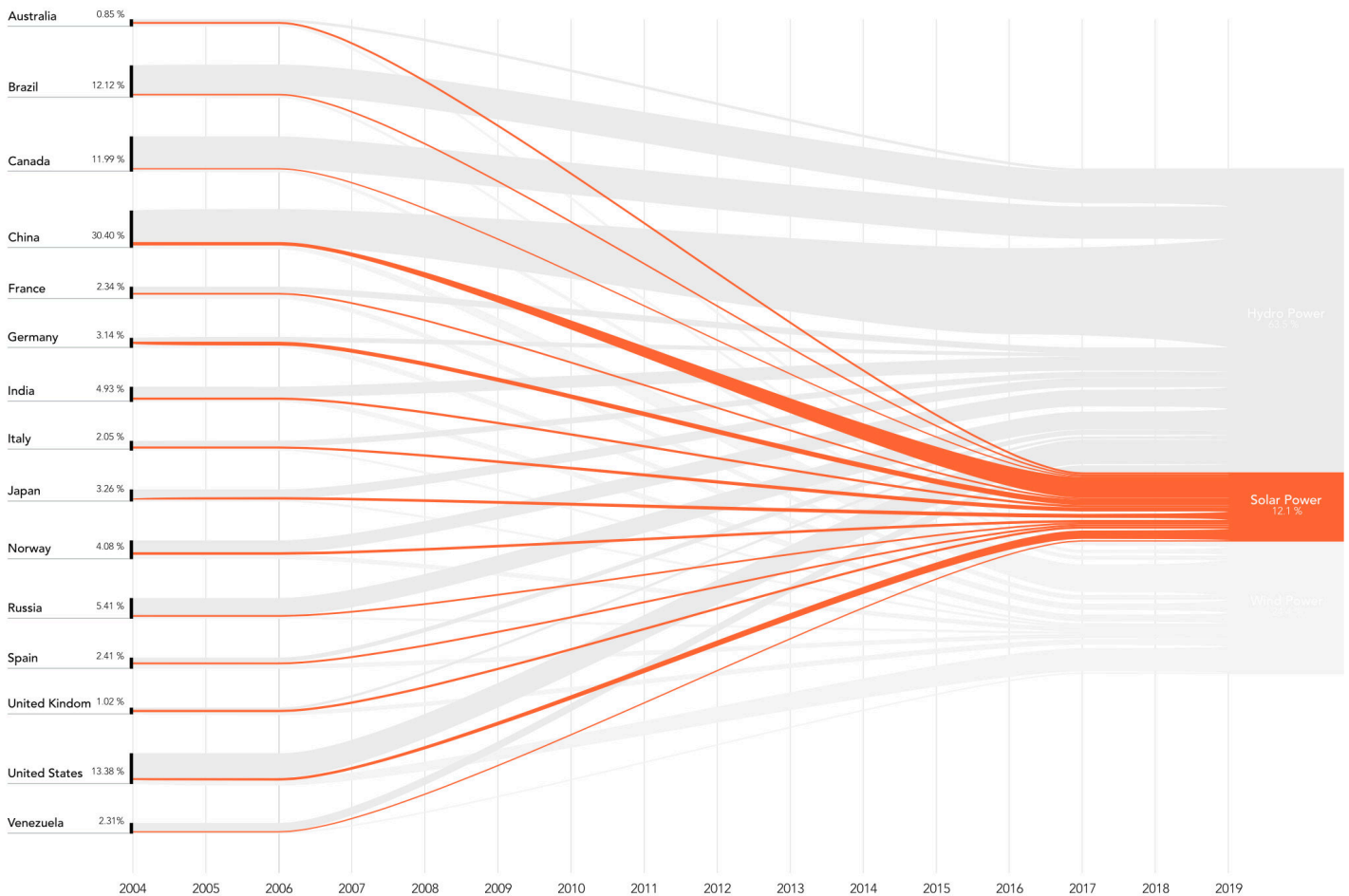


la segunda pregunta de este proyecto es ¿Cuáles países se destacan en las diferentes energías renovables?

en la figura 34, se ve la visualización mostrando el comportamiento de la energía solar y es fácil de ver, a través del grosor de las líneas, que empiezan en el 2004 muy parecidas y que china, estados unidos y de tercero alemania con los países que más han avanzado en energía solar en los últimos 15 años.

fig. 34 ejemplo de respuesta a la pregunta objetivo ¿Cuáles países se destacan en las diferentes energías renovables? del proyecto sobre la cantidad de energías renovables TEC 1S-2021 [12]

Renewable energy consumption

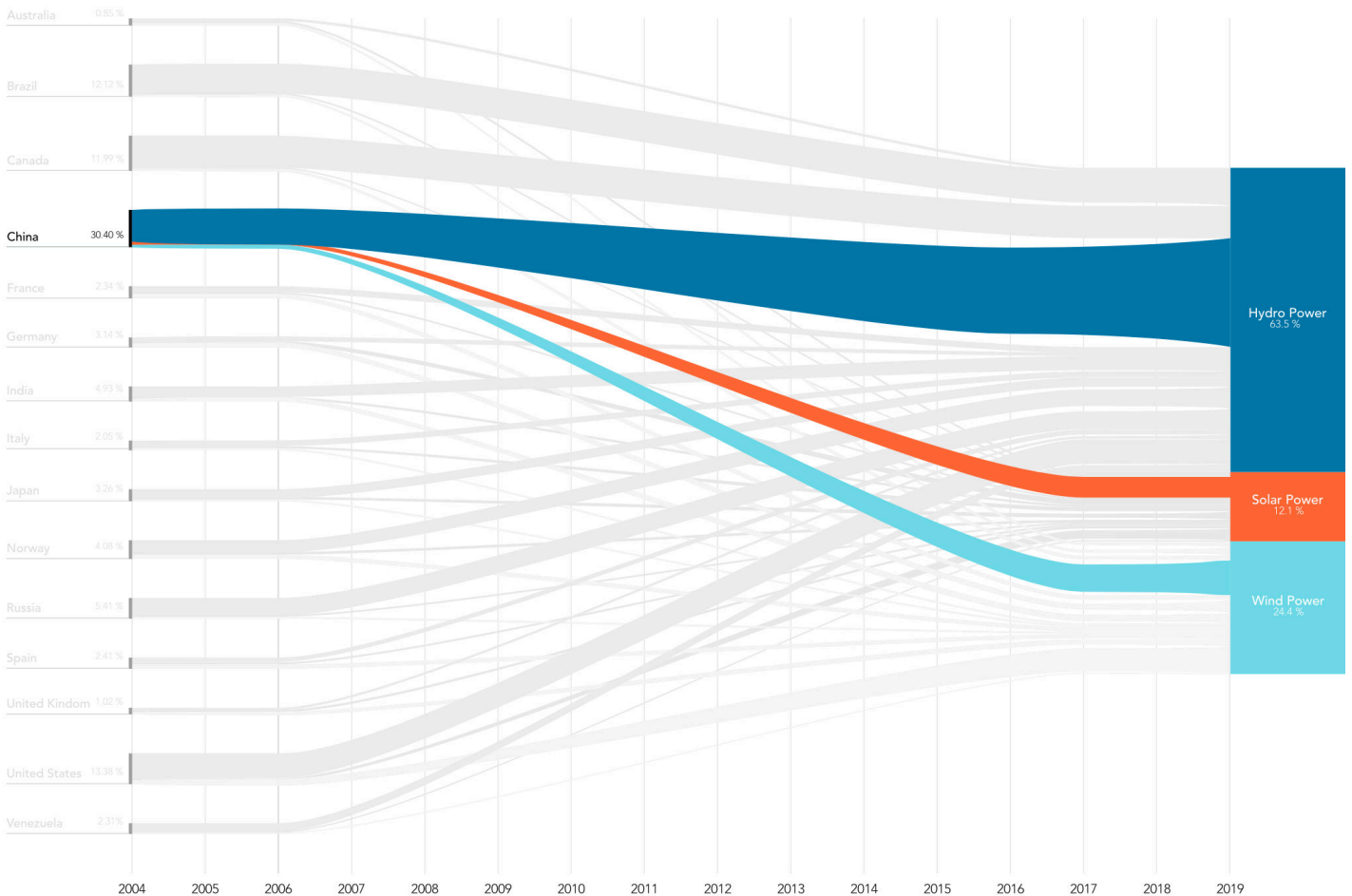


finalmente la tercera pregunta de este proyecto reza: ¿Cuáles países han transformado hacia energías renovables en forma destacable en los últimos 15 años?

en la figura 35 muestra la visualización para este caso de uso, como se ve claramente, china es quién ha aumentado más el grosor de las líneas a través de los años y conforma actualmente la mayor parte de las energías renovables. también es fácil responder, a través de los grosores de línea, cuáles otros países también han aumentado su producción de este tipo de energías.

fig. 35 ejemplo de respuesta a la pregunta objetivo ¿Cuáles países han transformado hacia energías renovables en forma destacable en los últimos 15 años? del proyecto sobre la cantidad de energías renovables TEC 1S-2021 [12]

Renewable energy consumption



estos son buenos ejemplos de una visualización que responde a todas las preguntas objetivo de un modo eficiente, eficaz y satisfactorio.

estas "vistas" de cómo, las diferentes versiones de interacción con la visualización, responden a cada pregunta o caso de uso, son un excelente modo de terminar el informe.

8. apéndices

8.1. algunas fuentes de datos publicas

<http://data.un.org/>

<https://snap.stanford.edu/data/>

<https://unstats.un.org/unsd/demographic/sources/census/wphc/default.htm>

<https://www.springboard.com/blog/data-science/free-public-data-sets-data-science-project/>

<https://www.quandl.com/>

<https://www.census.gov/en.html>

<http://networkrepository.com/>

<http://www.nhc.noaa.gov/data/tcr/index.php?season=2016&basin=atl>

<http://idatosabiertos.org/>

<http://abriendodatos.cr>

<http://www.inec.go.cr/>

<https://www.ministeriodesalud.go.cr/index.php/vigilancia-de-la-salud/estadisticas-y-bases-de-datos/estadisticas>

<http://www.nhc.noaa.gov/data/tcr/index.php?season=2016&basin=atl>

<https://www.nature.com/sdata/publish/for-authors#aims-scope>

<http://www.sciencedirect.com/science/journal/23523409>

<https://data.oecd.org/>

https://www.census.gov/foreign-trade/data/index.html?eml=gd&utm_medium=email&utm_source=govdelivery

<https://trends.google.com/trends/yis/2017/GLOBAL/?playerOpen>

9. Referencias

- [1] F. Hernandez-Castro, J. Jorge Monge-Fallas, S. González Villalobos, E. Barquero Rodríguez, and J. Esquivel Piedra. "Traffic Data Visualization in Costa Rica: A Visualization of top 100 routes with the highest traffic density in Costa Rica." *Ponte: International Scientific Researches Journal*. Vol. 72 | No. 11 | Nov 2016. Florence. Italy.
- [2] Hernández-Castro, F., and J. Monge-Fallas. "Plinius: A Visualization System of Costa Rica's Tectonic Plates." *Scientific Visualization* 11, no. 2 (2019).
- [3] Hidalgo, Hugo G., Eric J. Alfaro, Franklin Hernández-Castro, and Paula M. Pérez-Briceño. "Identification of tropical cyclones' critical positions associated with extreme precipitation events in Central America." *Atmosphere* 11, no. 10 (2020): 1123.
- [4] F. Hernández-Castro and J. Monge-Fallas, (July 2016). What for: Classification of Visual Paradigms. *Ponte: International Scientific Researches Journal*. Vol. 72 | No. 7 | Jul 2016. Florence. Italy.
- [5] Liew, Anthony. "DIKIW: Data, information, knowledge, intelligence, wisdom and their interrelationships." *Business Management Dynamics* 2, no. 10 (2013): 49.
- [6] Ribeca, Severino. "The data visualisation catalogue." URL <https://www.datavizcatalogue.com>. Accessed (2021): 07-17.
- [7] International Standard Organization (1998). Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability (ISO 9241-11) <https://www.sis.se/produkter/miljo-och-halsoskydd-sakerhet/ergonomi/iso9241111998/>
- [8] Hernández-Castro, F. "1 ISO 9241 11" Lecciones. 14 Julio 2021. Video, 12m38s. https://www.youtube.com/watch?v=DozhUxZiz-U&list=PLERsK7uoN4wYgJYnBHxOefpQLtEhLj_AG
- [9] F. Hernández-Castro. Metodología para el análisis y diseño de aplicaciones (usability cookbook). Escuela de Diseño Industrial, Instituto Tecnológico de Costa Rica. Cartago, Costa Rica. (2016). Repositorio TEC: <https://repositoriotec.tec.ac.cr/handle/2238/6776>.
- [10] Hernández-Castro, F. "4 how many testers" Lecciones. 14 Julio 2021. Video, 12m03s. https://www.youtube.com/watch?v=7aSCc2SdP8Y&list=PLERsK7uoN4wYgJYnBHxOefpQLtEhLj_AG&index=4
- [11] Hernández-Castro, F. "3 System Usability Scale" Lecciones. 14 Julio 2021. Video, 11m53s. https://www.youtube.com/watch?v=ZwtWGXZfWPk&list=PLERsK7uoN4wYgJYnBHxOefpQLtEhLj_AG&index=3
- [12] Rodríguez-Montero, M., Hernández-Castro, F. "Energías renovables en el mundo" Escuela de Diseño Industrial, Instituto Tecnológico de Costa Rica. Cartago, Costa Rica. (2021)
- [13] F. Hernández-Castro. Teoría del color (ingredients). Escuela de Diseño Industrial, Instituto Tecnológico de Costa Rica. Cartago, Costa Rica. 2016. <https://repositoriotec.tec.ac.cr/handle/2238/7314>
- [14] F. Hernández-Castro. cooking skills: combinando colores con código. Escuela de Diseño Industrial, Instituto Tecnológico de Costa Rica. Cartago, Costa Rica. 2019. <https://docs.google.com/viewer?sv=1&pid=sites&srcid=ZGVmYXVsdGRvbWFpbXhcmNoaXZvc3JlbGFjaW9uYWRvc3xneDo3ZTgzMzExY2I4YzE5ZjNi>
- [15] F. Hernández-Castro. Cooking skills: diseñando escalas cromáticas por código. Escuela de Diseño Industrial, Instituto Tecnológico de Costa

Rica. Cartago, Costa Rica. 2019. <https://repositoriotec.tec.ac.cr/handle/2238/10362>

[14] J. Brooke, "SUS: A "quick and dirty" usability scale," Usability Evaluation in Industry, pp. 189-194, Jan. 1996.

[15] International Standard Organization (2006). Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Common Industry Format (CIF) for usability test reports (ISO 25062). <https://webstore.iec.ch/publication/11255&preview=1>