Universidad Internacional de La Rioja

Facultad de Ingeniería y Tecnología

Máster Universitario en Análisis y Visualización de Datos Masivos / Visual Analitics & Big Data

Visualización de evolución en el tiempo

Actividad de estudio presentado por:	Juan David Escobar Escobar. Andrés Felipe Leal Mora. Juan Manuel Bautista Correa.
Tipo de trabajo:	Actividad 2
Modalidad:	Individual / Grupal
Profesor/a:	Yusef Hassan Moreno
Fecha:	Febrero 2022

1. Estructuración y análisis de datos

La siguiente imagen muestra una muestra de 5 datos del conjunto de datos original, utilizado como fuente de datos para el desarrollo de la actividad:

	CauseDeathOrInjury	Year	Measure	Value	ValuePercentage	CaseDeathList	Percent
333	Whooping cough	2005 - 2010	Deathsannualchange	-4.473298e-02	-4.473298	2	-4.473298e-02
334	Yellow fever	2017	Percentoftotaldeaths	8.554715e-05	0.008555	1	8.554715e-05
335	Yellow fever	2005 - 2010	Deathsannualchange	-3.874796e-02	-3.874796	2	-3.874796e-02
336	Zika virus	2017	Percentoftotaldeaths	3.467865e-07	0.000035	1	3.467865e-07
337	Zika virus	2005 - 2010	Deathsannualchange	0.000000e+00	0.000000	2	0.000000e+00

Se ignoraron columnas (Location, Age, Sex, Lower value, Upper value) que no generaban valor alguno al momento de representar la información mediante las herramientas de visualización. El siguiente paso fue el análisis de los datos almacenados, mediante la descripción de estos:

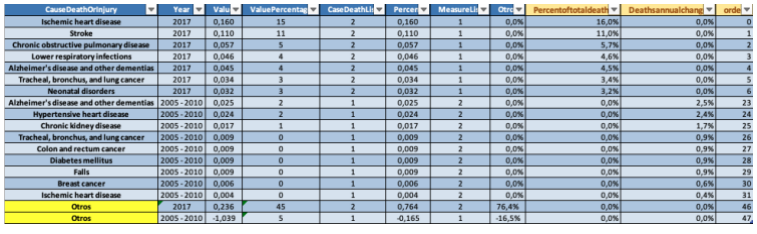
```
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   CauseDeathOrInjury  338 non-null    object
 1   Year                338 non-null    category
 2   Measure             338 non-null    category
 3   Value               338 non-null    float64
 4   ValuePercentage     338 non-null    int64
 5   CaseDeathList       338 non-null    int64
 6   Percent             338 non-null    float64
```

El conjunto de datos se segmento en dos agrupaciones “Cause Death Or Injury” y “Percent total deaths”, el total de registros fue de 338, cada uno de los conjuntos mencionados quedo con 169 luego de la agrupación.

A continuación, se realizo una operación de agregación, etiquetando de una nueva categoría de las medidas, llamada “Otros”, para aquellos registros que poseen un valor por debajo del 1%.

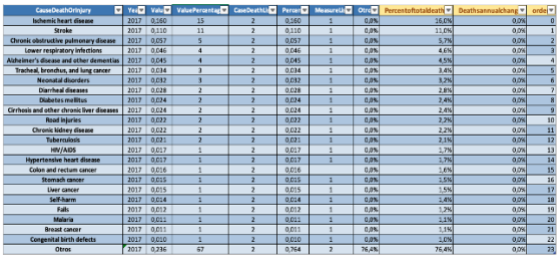
En el siguiente paso de tratamiento de datos se realizo una operación de pivot para los valores de las categorías [“Cause Death Or Injury” y “Percent total deaths” y “Otros”] almacenadas en la variable “Measure”.

La siguiente imagen ilustra muestra de algunos registros la estructura final del Dataset consolidado y agrupado:

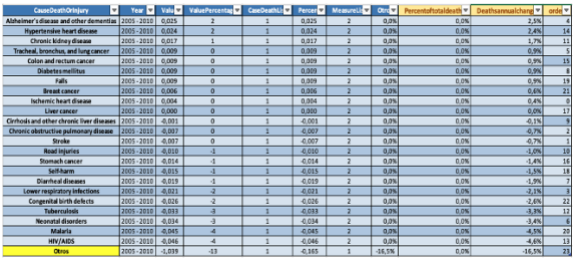


CauseDeathOrInjury	Year	Vale	ValuePercentag	CaseDeathLi	Percer	MeasureLi	Otr	Percentoftotaldeath	Deathsannualchang	orde
Ischemic heart disease	2017	0,160	15	2	0,160	1	0,0%	16,0%	0,0%	0
Stroke	2017	0,110	11	2	0,110	1	0,0%	11,0%	0,0%	1
Chronic obstructive pulmonary disease	2017	0,057	5	2	0,057	1	0,0%	5,7%	0,0%	2
Lower respiratory infections	2017	0,046	4	2	0,046	1	0,0%	4,6%	0,0%	3
Alzheimer's disease and other dementias	2017	0,045	4	2	0,045	1	0,0%	4,5%	0,0%	4
Tracheal, bronchus, and lung cancer	2017	0,034	3	2	0,034	1	0,0%	3,4%	0,0%	5
Neonatal disorders	2017	0,032	3	2	0,032	1	0,0%	3,2%	0,0%	6
Alzheimer's disease and other dementias	2005 - 2010	0,025	2	1	0,025	2	0,0%	0,0%	2,5%	23
Hypertensive heart disease	2005 - 2010	0,024	2	1	0,024	2	0,0%	0,0%	2,4%	24
Chronic kidney disease	2005 - 2010	0,017	1	1	0,017	2	0,0%	0,0%	1,7%	25
Tracheal, bronchus, and lung cancer	2005 - 2010	0,009	0	1	0,009	2	0,0%	0,0%	0,9%	26
Colon and rectum cancer	2005 - 2010	0,009	0	1	0,009	2	0,0%	0,0%	0,9%	27
Diabetes mellitus	2005 - 2010	0,009	0	1	0,009	2	0,0%	0,0%	0,9%	28
Falls	2005 - 2010	0,009	0	1	0,009	2	0,0%	0,0%	0,9%	29
Breast cancer	2005 - 2010	0,006	0	1	0,006	2	0,0%	0,0%	0,6%	30
Ischemic heart disease	2005 - 2010	0,004	0	1	0,004	2	0,0%	0,0%	0,4%	31
Otros	2017	0,236	45	2	0,764	2	76,4%	0,0%	0,0%	46
Otros	2005 - 2010	-1,039	5	1	-0,165	1	-16,5%	0,0%	0,0%	47

Por ultimo se crea una columna calculada llamada “Orden” para generar un orden principal en el Dataset agrupado por el año 2017, organizando los valores de manera ascendente los valores asociados a los % de muertes del 2017, posteriormente se genera la misma columna para el Dataset que almacena la variación porcentual de muertes entre el año 2005 y 2010, conservando el mismo orden asociado a cada enfermedad estipulado en el primer Dataset del 2017, a continuación se presenta una muestra de ambos Datasets.



CauseDeathOrInjury	Year	Valu	ValuePercentag	CaseDeathOr	Percen	MeasureLi	Otro	PercentoftotaldeathS	Deathsannualchang	orde
Ischemic heart disease	2017	0,160	15	2	0,160	1	0,0%	16,0%	0,0%	0
Stroke	2017	0,110	11	2	0,110	1	0,0%	11,0%	0,0%	1
Chronic obstructive pulmonary disease	2017	0,057	5	2	0,057	1	0,0%	5,7%	0,0%	2
Lower respiratory infections	2017	0,046	4	2	0,046	1	0,0%	4,6%	0,0%	3
Alzheimer's disease and other dementias	2017	0,045	4	2	0,045	1	0,0%	4,5%	0,0%	4
Tracheal, bronchus, and lung cancer	2017	0,034	3	2	0,034	1	0,0%	3,4%	0,0%	5
Neonatal disorders	2017	0,032	3	2	0,032	1	0,0%	3,2%	0,0%	6
Diarrheal diseases	2017	0,028	2	2	0,028	1	0,0%	2,8%	0,0%	7
Diabetes mellitus	2017	0,024	2	2	0,024	1	0,0%	2,4%	0,0%	8
Cirrhosis and other chronic liver diseases	2017	0,024	2	2	0,024	1	0,0%	2,4%	0,0%	9
Road injuries	2017	0,023	2	2	0,023	1	0,0%	2,3%	0,0%	10
Chronic kidney disease	2017	0,022	2	2	0,022	1	0,0%	2,2%	0,0%	11
Tuberculosis	2017	0,021	2	2	0,021	1	0,0%	2,1%	0,0%	12
HIV/AIDS	2017	0,017	1	2	0,017	1	0,0%	1,7%	0,0%	13
Hypertensive heart disease	2017	0,017	1	2	0,017	1	0,0%	1,7%	0,0%	14
Colon and rectum cancer	2017	0,016	1	2	0,016	1	0,0%	1,6%	0,0%	15
Stomach cancer	2017	0,015	1	2	0,015	1	0,0%	1,5%	0,0%	16
Liver cancer	2017	0,015	1	2	0,015	1	0,0%	1,5%	0,0%	17
Self-harm	2017	0,014	1	2	0,014	1	0,0%	1,4%	0,0%	18
Falls	2017	0,012	1	2	0,012	1	0,0%	1,2%	0,0%	19
Malaria	2017	0,011	1	2	0,011	1	0,0%	1,1%	0,0%	20
Breast cancer	2017	0,011	1	2	0,011	1	0,0%	1,1%	0,0%	21
Congenital birth defects	2017	0,010	1	2	0,010	1	0,0%	1,0%	0,0%	22
Otros	2017	0,336	67	2	0,764	3	76,4%	76,4%	0,0%	23



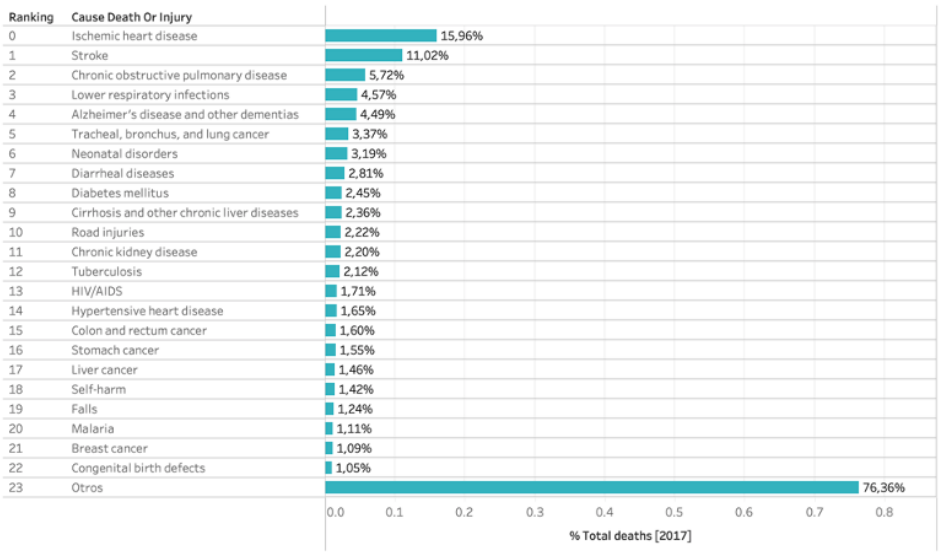
CauseDeathOrInjury	Year	Valu	ValuePercentag	CaseDeathOr	Percen	MeasureLi	Otro	PercentoftotaldeathS	Deathsannualchang	orde
Alzheimer's disease and other dementias	2005 - 2010	0,025	2	1	0,025	2	0,0%	0,0%	2,5%	4
Hypertensive heart disease	2005 - 2010	0,024	2	1	0,024	2	0,0%	0,0%	2,4%	14
Chronic kidney disease	2005 - 2010	0,017	1	1	0,017	2	0,0%	0,0%	1,7%	11
Tracheal, bronchus, and lung cancer	2005 - 2010	0,009	0	1	0,009	2	0,0%	0,0%	0,9%	5
Colon and rectum cancer	2005 - 2010	0,009	0	1	0,009	2	0,0%	0,0%	0,9%	15
Diabetes mellitus	2005 - 2010	0,009	0	1	0,009	2	0,0%	0,0%	0,9%	8
Falls	2005 - 2010	0,009	0	1	0,009	2	0,0%	0,0%	0,9%	19
Breast cancer	2005 - 2010	0,006	0	1	0,006	2	0,0%	0,0%	0,6%	21
Ischemic heart disease	2005 - 2010	0,004	0	1	0,004	2	0,0%	0,0%	0,4%	0
Liver cancer	2005 - 2010	0,000	0	1	0,000	2	0,0%	0,0%	0,0%	17
Cirrhosis and other chronic liver diseases	2005 - 2010	-0,001	0	1	-0,001	2	0,0%	0,0%	-0,1%	9
Chronic obstructive pulmonary disease	2005 - 2010	-0,007	0	1	-0,007	2	0,0%	0,0%	-0,7%	2
Stroke	2005 - 2010	-0,007	0	1	-0,007	2	0,0%	0,0%	-0,7%	1
Road injuries	2005 - 2010	-0,010	-1	1	-0,010	2	0,0%	0,0%	-1,0%	10
Stomach cancer	2005 - 2010	-0,014	-1	1	-0,014	2	0,0%	0,0%	-1,4%	16
Self-harm	2005 - 2010	-0,015	-1	1	-0,015	2	0,0%	0,0%	-1,5%	18
Diarrheal diseases	2005 - 2010	-0,019	-1	1	-0,019	2	0,0%	0,0%	-1,9%	7
Lower respiratory infections	2005 - 2010	-0,021	-2	1	-0,021	2	0,0%	0,0%	-2,1%	3
Congenital birth defects	2005 - 2010	-0,026	-2	1	-0,026	2	0,0%	0,0%	-2,6%	22
Tuberculosis	2005 - 2010	-0,033	-3	1	-0,033	2	0,0%	0,0%	-3,3%	12
Neonatal disorders	2005 - 2010	-0,034	-3	1	-0,034	2	0,0%	0,0%	-3,4%	6
Malaria	2005 - 2010	-0,045	-4	1	-0,045	2	0,0%	0,0%	-4,5%	20
HIV/AIDS	2005 - 2010	-0,046	-4	1	-0,046	2	0,0%	0,0%	-4,6%	13
Otros	2005 - 2010	-0,039	-13	1	-0,165	1	-16,5%	0,0%	-16,5%	23

2. Propuesta grafica

Graphic visualization proposal (laboratory 2)

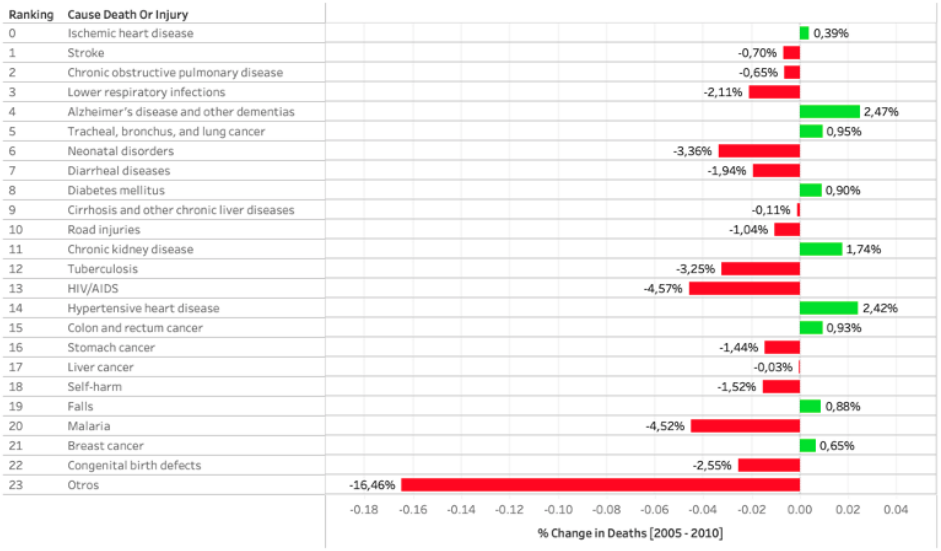
Ranking Causes of Death in 2017

9 de Febrero de 2022



Percentage variation (2005 - 2010)

9 de Febrero de 2022



By: Juan David Escobar Escobar, Andres Felipe leal Mora, Juan Manuel Bautista Correa.

3. Herramientas y recursos

- Tableau Desktop

  https://github.com/juadaves91/UnirVisualizacion/blob/main/Actividad%202/Book2.twb

- Excel 2021 (Dataset estructurado):

  https://github.com/juadaves91/UnirVisualizacion/blob/main/Actividad%202/estructurados_act2.xlsx

- Jupyter Notebook: A continuación, se hace referencia al script generado para el análisis descriptivo de los datos, y la estructuración de estos (https://github.com/juadaves91/UnirVisualizacion/blob/main/Actividad%202/EstructuracionDatosAct2.ipynb).