

GMestrado em Engenharia Informática e Tecnologia Web

22126 – Visualização de informação

Grupo: Claudia Pires (1303334) / Valter Bastos (2302612)

Tópico 4 - Design e Redesenho

RELATÓRIO

1. OBJETIVO

Efetuar um trabalho de redesenho das visualizações de informação concebidas no âmbito do trabalho do tópico 3 sobre exploração de técnicas básicas de conceção de visualizações, com o tema “pobreza” a nível global e as suas origens.

A proposta inicial referia o uso da ferramenta Tableau, mas a opção recaiu sobre a ferramenta Qlik Sense, visto um dos elementos do grupo ter experiência com essa ferramenta, e porque foi possível utiliza-la ao abrigo do “*academic program*” (<https://www.qlik.com/us/company/academic-program>).

Os *datasets* foram extraídos de “*ourworldindata*” e “*Eurostat*”.

2. METODOLOGIA

A definição de pobreza abordada foi descrita no Tópico 3:

(<https://elearning.uab.pt/mod/forum/discuss.php?d=1042577#p3523263>,
<https://elearning.uab.pt/mod/forum/discuss.php?d=1042577#p3526764>)

Em termos gerais, a pobreza é definida como uma condição puramente económica, baseada nos rendimentos líquidos dos agregados familiares.

Para os países em desenvolvimento, conforme declarado pelas Nações Unidas: pessoas que vivem com menos de 3.10\$ por dia (até 2016) ou 3.65\$ por dia a partir de 2017, são consideradas como vivendo na pobreza (1.90\$ e 2.15\$, respetivamente, são os limiares para a pobreza extrema).

Para os Países da UE27, a EUROSTAT define que qualquer pessoa que viva com menos de 60% da mediana de cada país está abaixo da linha de pobreza.

De forma semelhante, para o resto dos países desenvolvidos, o limiar mais comum para a linha de pobreza é 50% da mediana de cada país.

A metodologia abordada neste trabalho assenta nessas definições. Para isso, foram recolhidos vários *datasets* e combinados para gerar esta aplicação.

Iniciou-se o processo descarregando uma lista de todos os países do Mundo com os correspondentes códigos ISO2/ISO3 para facilitar a interseção entre esta lista e outros *datasets*, visto que, foi detetado, que frequentemente o mesmo país surgia com

denominações diferentes consoante a fonte em que era referido. Esta lista também inclui o Nome do País e Continente.

Fonte: <https://github.com/luke/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv>

Em seguida, foram marcados os países que atualmente são identificados como países em desenvolvimento pelas Nações Unidas.

Fonte: Bing Chat – “*Can you please list here all LDC countries*”

Relativamente à distribuição de rendimentos, recolheram-se dados provenientes de "ourworldindata", que permitiram obter informações para 41% dos países em análise. Desse *dataset*, conseguiu-se extrair informação sobre os rendimentos líquidos médios dos agregados familiares para vários anos e rendimento diário limiar para cada decil, de forma a obter-se uma visão mais detalhada da distribuição do rendimento, o que permite apurar a média e também como esse rendimento é distribuído, de forma a identificar-se as disparidades existentes entre os mais pobres e os mais ricos, estando no 1º decil, os 10% mais pobres e no 10º decil, os 10% mais ricos.

Esta abordagem permitiu encontrar a percentagem aproximada da população que vive abaixo dos limiares de pobreza.

Fonte: <https://ourworldindata.org/explorers/incomes-across-distribution-ppp2017?tab=table&time=1990..latest&country=~ALB&pickerSort=asc&pickerMetric=Entity&Indicator=Mean+income+or+consumption%2C+by+decile&Decile=All+deciles&Household+survey+data+type=Income+surveys+only&Period=Day&Show+breaks+between+less+comparable+surveys=false>

Para os países em falta, recorreu-se a um *dataset* contendo a taxa de pobreza (percentagem da população) a 2.15\$/dia e à contagem da população por ano.

Fonte: <https://data.worldbank.org/indicator/SI.POV.DDAY>,
<https://data.worldbank.org/indicator/SP.POP.TOTL>

Por último, recolheu-se dados relativos à divisão da população por género na EU27 e o rendimento médio por género fornecido pela "Eurostat". Com estes dados, foi possível estimar a incidência da pobreza por género e como afeta de forma díspares a população masculina e feminina.

Fonte: https://ec.europa.eu/eurostat/databrowser/view/ilc_di03/default/table?lang=en

3. DESENVOLVIMENTO DAS VISUALIZAÇÕES

A visualização de informação é o uso de representações visuais interativas e com suporte computacional de dados abstratos com o intuito de aumentar o conhecimento (Card, 2008). No decorrer do processo de construção de visualizações, um dos constrangimentos mais frequentes é a dificuldade em encontrar o mapeamento correto entre entidades abstratas e a sua representação espacial (Card, 2008).

A visualização auxilia o conhecimento não por qualquer particular superioridade das imagens sobre qualquer outra forma de pensamento e comunicação, mas apenas porque a visualização ajuda o utilizador a pensar de formas específicas (Card, 2008).

Neste trabalho, sendo o foco principal a exploração de uma primeira ferramenta de conceção de visualizações, foi possível ter a perceção de que as opções a tomar estão direta e indissociavelmente ligadas à natureza dos dados, assim como ao objetivo a atingir com a visualização, pois uma má escolha impede que se cumpra o objetivo de auxiliar o utilizador a adquirir conhecimento.

Uma eficaz transmissão de informações complexas através de visualizações de dados necessita de uma análise cuidadosa da capacidade do utilizador em extrair conhecimento dos dados apresentados (Munzer, 2014), optou-se assim por desenvolver uma aplicação dividida em 3 *sheets* com um total de 12 visualizações.

A aplicação possui uma integração dinâmica, permitindo seleccionar informação, que irá atuar como filtro e condicionar os dados exibidos noutras visualizações relacionadas. O uso de visualizações múltiplas associadas permite que diferentes aspetos dos mesmos dados sejam mostrados em diferentes visualizações, assim cada gráfico oferece uma visão única desses dados, mas juntos enriquecem a informação transmitida. (Munzer, 2014).

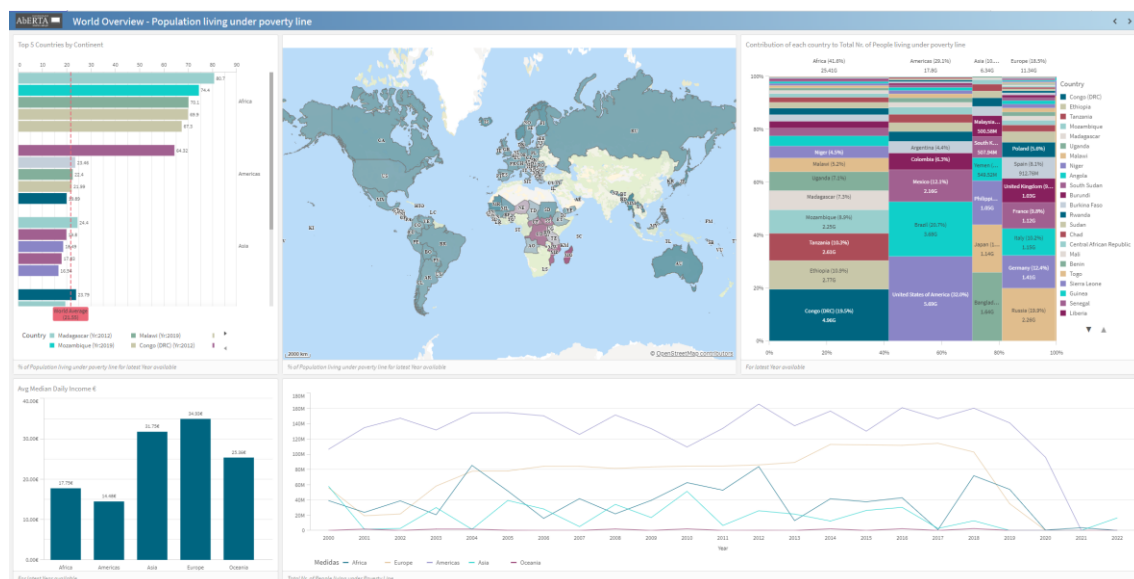


Figura 1 Visão global da Sheet 1- World Overview: Population living under poverty line

A *sheet 1* teve como foco transmitir ao utilizador uma visão global do mundo através de 5 visualizações que proporcionam uma visão global sobre a população mundial que vive abaixo da linha de pobreza ou pobreza extrema, de acordo com a abordagem definida na metodologia anteriormente indicada.

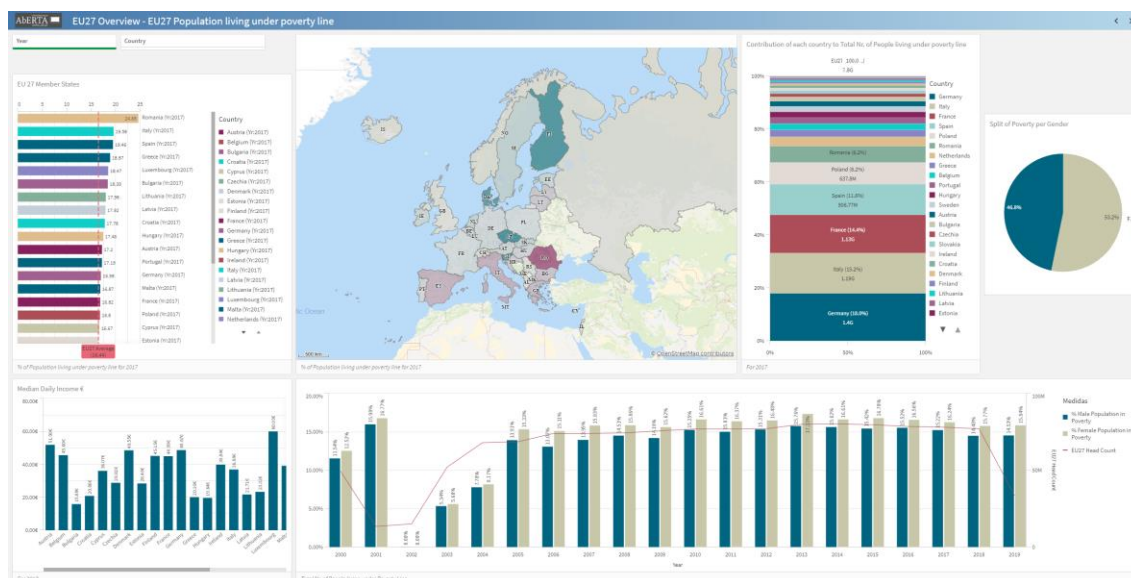


Figura 2 Visão global da *Sheet 2* – EU27 Overview: EU27 Population living under poverty line

A *sheet 2* foi elaborada com o objetivo de transmitir visualizações que proporcionassem uma visão sobre a população dos Estados Membros da UE27 que vivem abaixo da linha de pobreza.

AbERTA Tabular Data

Conti...	Region	Country	ISO2 Code	ISO3 Code	LDC Flag	Year	Latest Year Flag	Rule	Population	Median Daily Income €	% of Population under Poverty Line	Population under Poverty Line
Totals									1,530,987,370			264,802,329
Europe	Rest of the World	Albania	AL	ALB	NO	2017	0	<= 50% Median Daily Income	2,873,457	7.92146395859	17.45	501,365
Europe	Rest of the World	Albania	AL	ALB	NO	2017	0	<= 50% Median Daily Income	2,873,457	7.92146395859	17.45	501,365
Americas	Rest of the World	Argentina	AR	ARG	NO	2017	0	<= 50% Median Daily Income	40,410,674	20.5418095519	18.05	7,294,450
Americas	Rest of the World	Argentina	AR	ARG	NO	2017	0	<= 50% Median Daily Income	40,410,674	20.5418095519	18.05	7,294,450
Europe	EU27	Austria	AT	AUT	NO	2017	0	<= 60% Median Daily Income	8,797,566	51.8955306506	17.20	1,513,440
Europe	EU27	Austria	AT	AUT	NO	2017	0	<= 60% Median Daily Income	8,797,566	51.8955306506	17.20	1,513,440
Europe	EU27	Belgium	BE	BEL	NO	2017	0	<= 60% Median Daily Income	11,375,158	45.5968776081	16.33	1,857,928
Europe	EU27	Belgium	BE	BEL	NO	2017	0	<= 60% Median Daily Income	11,375,158	45.5968776081	16.33	1,857,928
Asia	Rest of the World	Bhutan	BT	BTN	YES	2017	0	<= 1.9565€ Median Daily Income	756,121	-	0.90	6,805
Asia	Rest of the World	Bhutan	BT	BTN	YES	2017	0	<= 1.9565€ Median Daily Income	756,121	-	0.90	6,805
Americas	Rest of the World	Bolivia	BO	BOL	NO	2017	0	<= 50% Median Daily Income	11,192,853	12.7995268401	19.33	2,164,097
Americas	Rest of the World	Bolivia	BO	BOL	NO	2017	0	<= 50% Median Daily Income	11,192,853	12.7995268401	19.33	2,164,097
Americas	Rest of the World	Brazil	BR	BRA	NO	2017	0	<= 50% Median Daily Income	207,833,825	11.1010359096	22.29	46,328,730
Americas	Rest of the World	Brazil	BR	BRA	NO	2017	0	<= 50% Median Daily Income	207,833,825	11.1010359096	22.29	46,328,730
Europe	EU27	Bulgaria	BG	BGR	NO	2017	0	<= 60% Median Daily Income	7,075,947	15.8794392211	18.39	1,301,491
Europe	EU27	Bulgaria	BG	BGR	NO	2017	0	<= 60% Median Daily Income	7,075,947	15.8794392211	18.39	1,301,491

Figura 3 Visão global da *Sheet 3* – Tabular Data

A *sheet 3* contém uma tabela com todos os campos para exploração e extração completamente granular dos dados Fonte.

O foco deste relatório serão a *sheet 1* e *sheet 2*, onde a informação tratada varia, mas, na sua maioria, os tipos de visualizações escolhidos foram os mesmos.

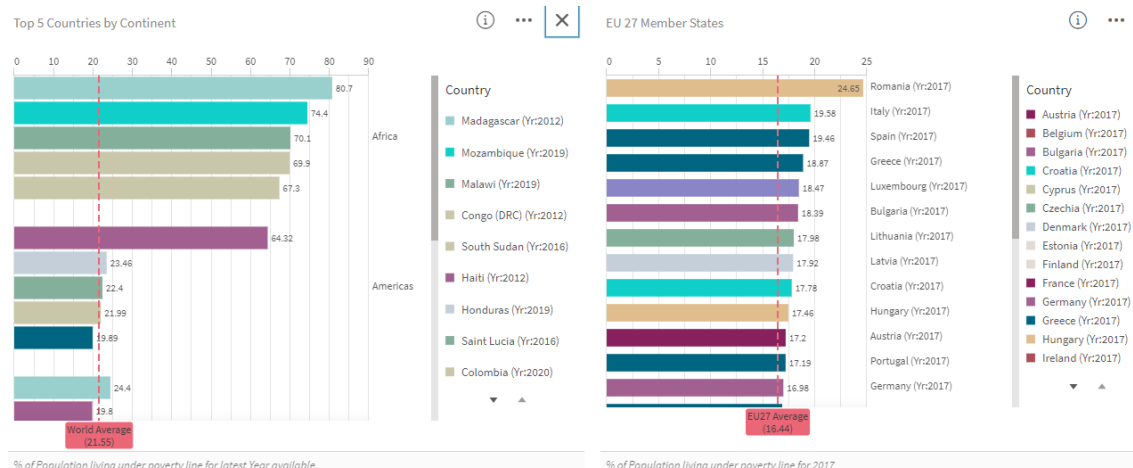


Figura 4 Top 5 Países por Continente / Estados Membros da UE27

No canto superior esquerdo das respetivas *sheets* é apresentado um gráfico de barras horizontais baseado numa estrutura de dados tabular, que permite uma comparação direta entre os itens e onde as marcas (representações visuais dos pontos de dados), são as barras e os canais (meios pelos quais as propriedades dessas marcas transmitem informação) são o comprimento das barras que representam a percentagem da população e a cor que é usada para diferenciar os continentes ou países.

A opção foi por uma visualização simples que permitisse uma análise dimensional, quantitativa e ordinal da correlação de dados extraídos dos *datasets* escolhidos.

A disposição das barras e a clareza na apresentação dos percentuais tiveram como finalidade a redução de carga cognitiva, apresentando informações de forma a minimizar a necessidade de processamento mental pelo utilizador (Munzer, 2014).

Não descurando a importância da precisão temporal na interpretação da visualização, cada barra tem a indicação do ano específico no qual essa informação foi baseada, de forma a promover uma análise fidedigna e a evitar ambiguidades.

Esta visualização permite comparação direta entre os países, e a linha pontilhada vermelha indica a média mundial ou dos Estados Membros da UE27, conforme se trate da *sheet* 1 ou da *sheet* 2, servindo de referência para o utilizador.

No caso da *sheet* 1, optou-se por uma redução de itens, mostrando apenas os 5 países com maior taxa de pobreza por continente, apresentando uma organização implícita dos dados, ordenando os países de cada continente, por percentagem da população que vive abaixo da linha de pobreza.

Esta visualização foi escolhida pela facilidade com que se destacam as disparidades entre os países mais afetados pela pobreza, no seu contexto.

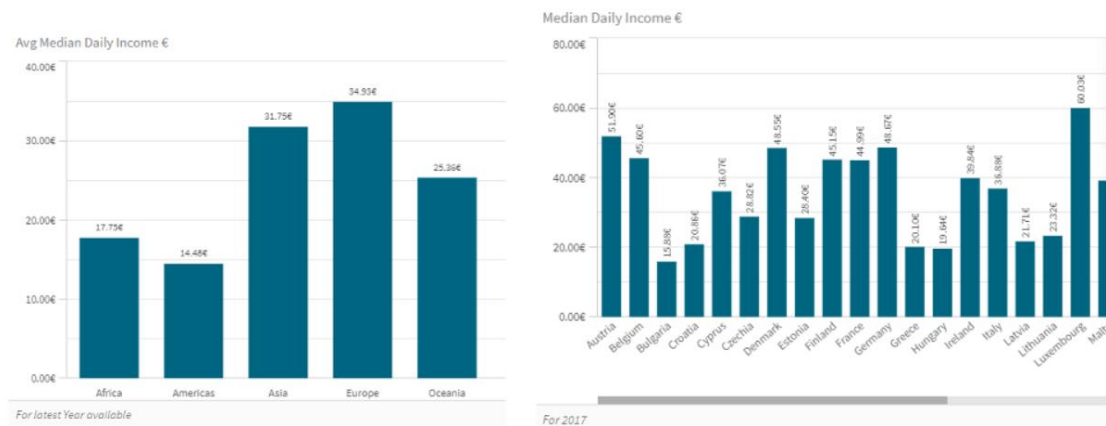


Figura 5 Rendimento Médio Diário dos Agregados

No canto inferior esquerdo, é apresentada uma visualização representando o “Rendimento Líquido Médio Diário dos Agregados Familiares”, na *sheet 1* esse rendimento é apresentado por continente, com possibilidade de selecionar um ou mais continentes e obter o detalhe de países pertencentes a essa seleção, enquanto na *sheet 2* o rendimento é apresentado por Estado Membro.

Trata-se de um gráfico de barras onde as marcas, tal como no anterior, são representadas pelas barras, neste caso vertical. O canal corresponde à altura de cada barra sendo proporcional ao valor que representa, para que facilite a comparação entre os continentes ou países consoante a opção em que o utilizador se encontra.

A cor uniforme foi uma escolha consciente, para transmitir que pertencem a uma mesma categoria de dados, mantendo o foco na altura das barras, como principal meio de interpretação dos dados. Cada barra tem como anotação o valor que representa para facilitar a leitura direta sem necessidade de referência à escala.

As dimensões destes gráficos (por continente e por país) são quantitativas, sendo que no gráfico por continente, resulta numa visualização mais simples e menos granular, e no gráfico por país os dados estão desagregados introduzindo uma maior complexidade e detalhe.

Quanto à categorização, os dois gráficos apresentam diferenças significativas: no gráfico por continentes a categorização é feita a um nível macro, o que facilita comparações em larga escala e tendências gerais, no gráfico dos países, a categorização permite uma análise mais desagregada, o que é útil para identificar padrões e discrepâncias em uma escala mais precisa. A escolha destes dois tipos de categorização permite que o utilizador escolha uma visão geral ou uma investigação detalhada, conforme o propósito da análise.

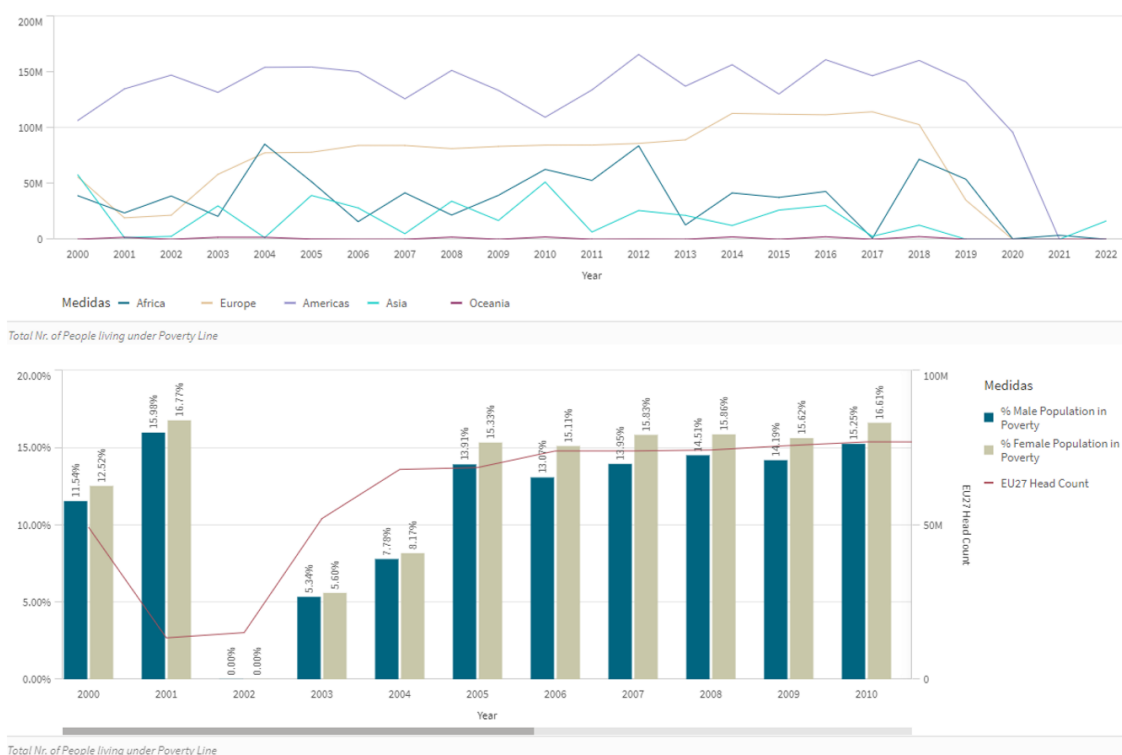


Figura 6 Evolução cronológica da pobreza

Na zona inferior central surge uma visualização que representa cronologicamente o número total de pessoas a viver abaixo da linha de pobreza ou pobreza extrema, por continente.

Nesta visualização foi usado um gráfico de linhas simples no caso da *sheet 1*, onde as marcas são as próprias linhas, codificadas com uma cor diferente, o que serve como um canal de cor para distinguir os diversos continentes (Munzer, 2014), já na *sheet 2*, é apresentado um gráfico de barras combinado com uma linha, sendo neste caso as marcas as barras e os pontos na linha e são utilizados canais de posição vertical para quantificar a percentagem da população masculina e feminina que vive na pobreza, representado pelas barras, e um canal de posição vertical para a linha que representa a população na UE27.

A escolha deste tipo de gráfico recaiu na importância da dimensão temporal da informação quantitativa, utilizando o eixo horizontal (dimensão ordinal) como sequência cronológica e facilitando a análise de padrões, tendências e anomalias. A visualização permite também uma seleção dinâmica de forma a destacar um subconjunto de dados.

A inclusão de anotações numéricas no caso do gráfico da *sheet 2* teve a intenção de melhorar a legibilidade dos dados, permitindo ao utilizador rapidamente ter acesso a informações quantitativas precisas.

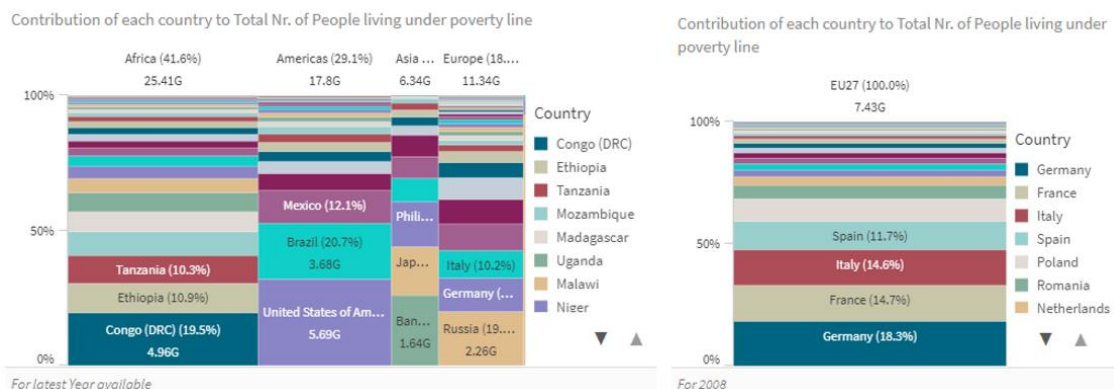


Figura 7 Contribuição absoluta de cada país para o número total de pessoas a viver abaixo da linha de pobreza ou pobreza extrema

Do lado superior direito surge uma visualização que representa a “Contribuição absoluta de cada país para o número total de pessoas a viver abaixo da linha de pobreza ou pobreza extrema”.

O gráfico utilizado nesta visualização foi um gráfico Mekko, onde as marcas são as secções coloridas e o canal de área representa 2 variáveis quantitativas, a percentagem com que cada país contribui para o total da população que vive abaixo da linha de pobreza ou pobreza extrema, e o total cardinal correspondente dessa população. (Munzer, 2014). Assim, esta visualização permite uma análise dimensional, quantitativa e qualitativa. As barras foram ordenadas horizontalmente de acordo com o continente e alinhadas verticalmente para facilitar a comparação entre elas, cada barra é subdividida com recurso do canal de cor para de uma forma intuitiva o utilizador distinguir os diversos países.

A escolha deste tipo de gráfico permitiu mostrar não só a proporção de pessoas que vivem abaixo da linha de pobreza ou pobreza extrema em cada país, mas também o resultado dessa contribuição para os totais de cada continente.

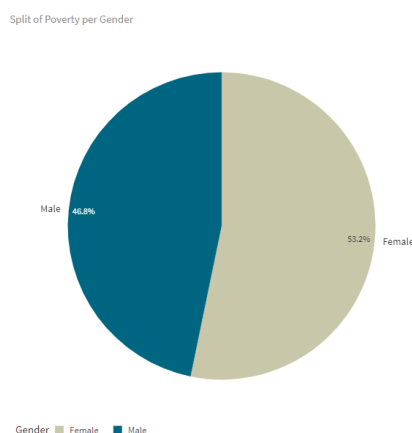


Figura 8 Divisão da pobreza por género

dimensão quantitativa e qualitativa.

No caso da *sheet 2*, é apresentada uma pequena visualização em formato de gráfico circular com a informação da distribuição da pobreza entre os géneros de uma forma direta e perceptível. Neste gráfico as marcas são os setores e o canal de área angular é utilizador para representar a proporção da população masculina e feminina que vive na pobreza.

O gráfico usa um canal de cor para diferenciar as duas categorias, masculino e feminino, e assim ser mais fácil distinguir visualmente a proporção de cada categoria.

Tal como em gráficos anteriores, foi decidido colocar anotação com a percentagem para facilitar a leitura por parte do utilizador. Assim obteve-se uma visualização de

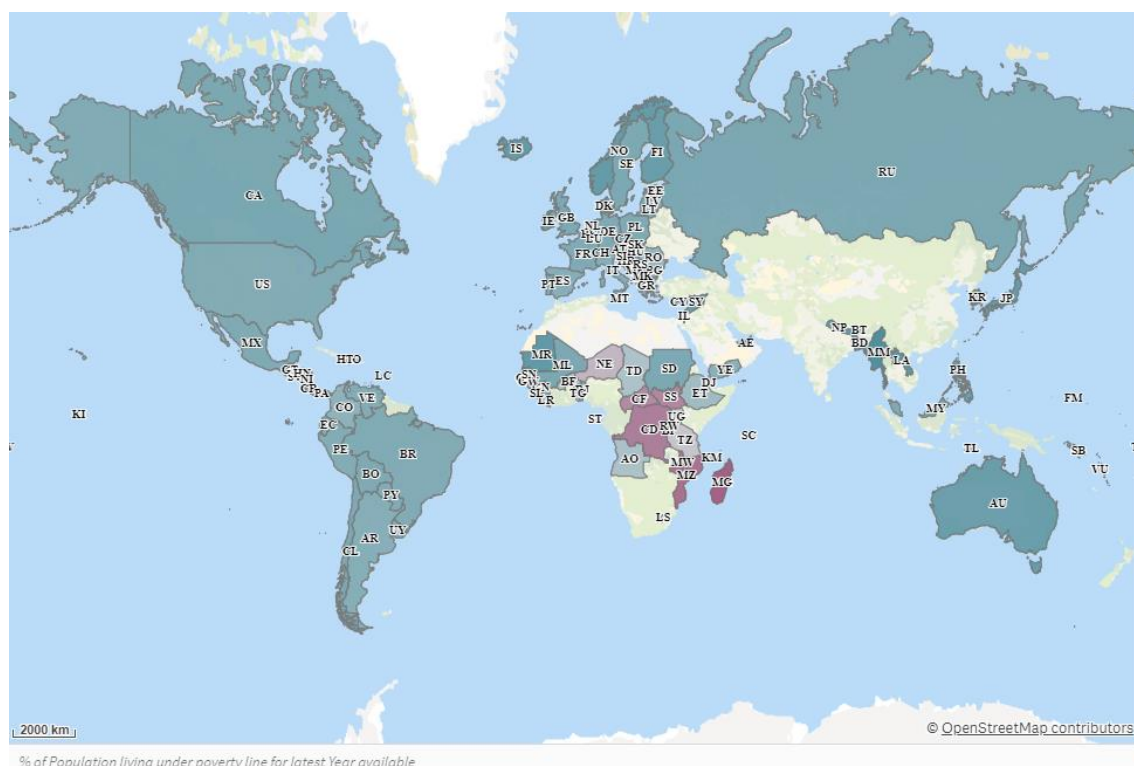


Figura 9 Mapa Mundo

Na zona superior central da aplicação é apresentado um mapa do que tem como objetivo mostrar a incidência da pobreza geograficamente com recurso ao gradiente de cores (entre cores mais vivas a cores mais suaves) para dar ênfase aos países com maior prevalência da pobreza por percentagem da população, o que permite uma rápida identificação de zonas geográficas onde há maior prevalência de pobreza, como é o caso da África Central.

Nesta visualização as marcas são as formas dos próprios países e o canal de cor é usado para codificar a percentagem da população que vive abaixo da linha da pobreza ou pobreza extrema.

Esta visualização tem uma dimensão quantitativa, onde as cores representam dados quantitativos através das suas tonalidades, e uma categorização espacial e ordinal, onde o gradiente de cor sugere uma relação ordinal dos dados, ou seja, países com tons aproximados têm percentagens de pobreza semelhantes. Os países são identificados pelo seu código Alf2 (ISSO 3166) servindo esses códigos como identificadores nominais para cada categoria espacial.

A escolha desta visualização teve como objetivo fornecer uma visão global dentro do âmbito em que o utilizador se encontre, permitir uma fácil identificação dos países para os quais existem dados disponíveis, a forma intuitiva com que o utilizador pode escolher um ou mais países que deseje analisar, seja clicando ou usando a ferramenta laço, permitindo assim uma interação muito positiva, assim como por ser possível ter essa visualização sincronizada com os gráficos de barras e refletir a informação nessas visualizações e vice-versa, o que proporciona uma análise mais detalhada e contextualizada.

4. CONCLUSÕES

O trabalho permitiu a exploração de técnicas fundamentais para a utilização de ferramentas de conceção de visualizações (Qlik Sense neste caso) e aplicação dos conceitos teóricos estudados nos tópicos 3 e 4.

O recurso escolhido revelou uma imensa potencialidade na variedade de visualizações disponibilizadas, na facilidade de integração dos dados para a construção das visualizações, assim como no facto de com um simples botão direito do rato em cima de qualquer gráfico permitir a exportação dos dados subjacentes para excel, bem como gravar essa informação para png ou pdf.

Os *datasets* obrigaram a um moroso trabalho e dada a dificuldade em encontrar dados para todos os anos e para todos os países, optou-se nalguns casos por mostrar os dados do último ano disponível para cada país, o que não seria aconselhável.

A aplicação foi construída para ser visualmente apelativa, mas essencialmente privilegiar a intuitividade e providenciar mecanismos que facilitem a exploração e correta interpretação dos dados, evidenciando os indicadores chave e em determinados casos os padrões nos dados.

Como reflexão final este trabalho permitiu aplicar e compreender os conceitos aplicados nos recursos técnicos indicados pelos docentes, percebendo como se enquadram os conceitos de canal e marca e compreender melhor o significado de dimensões de dados, categorização, dados quantitativos, qualitativos, ordinais, etc.

Bibliografia

Card, S. K. (2008). *The Human Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications 2nd Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Munzer, T. (2014). *Visualization Analysis and Design*. CRC Press.

A aplicação encontra-se disponível no repositório:

https://github.com/MEIW-ORG/VI_World_Poverty

ANEXO 1

Loading Script:

```

SET ThousandSep=', ';
SET DecimalSep='.';
SET MoneyThousandSep=', ';
SET MoneyDecimalSep='.';
SET MoneyFormat='€#,##0.00;-€#,##0.00';
SET TimeFormat='hh:mm:ss';
SET DateFormat='DD/MM/YYYY';
SET TimestampFormat='DD/MM/YYYY hh:mm:ss[.fff]';
SET FirstWeekDay=6;
SET BrokenWeeks=0;
SET ReferenceDay=4;
SET FirstMonthOfYear=1;
SET CollationLocale='en-IE';
SET CreateSearchIndexOnReload=1;
SET MonthNames='Jan;Feb;Mar;Apr;May;Jun;Jul;Aug;Sep;Oct;Nov;Dec';
SET
LongMonthNames='January;February;March;April;May;June;July;August;September;October;November;December';
SET DayNames='Mon;Tue;Wed;Thu;Fri;Sat;Sun';
SET LongDayNames='Monday;Tuesday;Wednesday;Thursday;Friday;Saturday;Sunday';
SET NumericalAbbreviation='3:k;6:M;9:G;12:T;15:P;18:E;21:Z;24:Y;-3:m;-6:u;-9:n;-12:p;-15:f;-18:a;-21:z;-24:y';

set vUSD_to_EUR = 0.91;

//set vPovertyThresholdBefore2017 = 1.90; //Extreme
//set vPovertyThresholdAfter2017 = 2.15; //Extreme

set vPovertyThresholdBefore2017 = 3.10; //Poverty
set vPovertyThresholdAfter2017 = 3.65; //Poverty

//Create Countries definition table
Countries:
LOAD distinct
    [Country],
    [ISO2_Code],
    [ISO3_Code],
    [LDC_Flag],
    [Region],
    [Continent]
FROM [lib:///Pverty DB.xlsx]
(ooxml, embedded labels, table is [Countries])
;

Map_Country_ISO3:
mapping LOAD distinct [Country], [ISO3_Code] Resident Countries;

Map_Country_ISO2:
mapping LOAD distinct [Country], [ISO2_Code] Resident Countries;

Map_ISO3_LDC:
mapping LOAD distinct [ISO3_Code], [LDC_Flag] Resident Countries;

Map_ISO3_Region:
mapping LOAD distinct [ISO3_Code], [Region] Resident Countries;

Map_ISO3_Continent:
mapping LOAD distinct [ISO3_Code], [Continent] Resident Countries;

Map_ISO3_Country:
mapping LOAD distinct [ISO3_Code], [Country] Resident Countries;

Map_ISO3_ISO2:
mapping LOAD distinct [ISO3_Code], [ISO2_Code] Resident Countries;

```

```
Map_ISO2_ISO3:
mapping LOAD distinct [ISO2_Code], [ISO3_Code] Resident Countries;
```

```
Map_ISO2_Country:
mapping LOAD distinct [ISO2 Code], [Country] Resident Countries;
```

```
Map_ISO2_Region:
mapping LOAD distinct [ISO2_Code], [Region] Resident Countries;
```

```
//Incomes Distribution table
//https://ourworldindata.org/explorers/incomes-across-distribution-
ppp2017?tab=table&time=1990..latest&country=~ALB&pickerSort=asc&pickerMetric=Entity&Indicator=Mean+income+or+consumption%2C+by+decile&Decile=All+deciles&Household+survey+data+type=Income+surveys+only&Period=Day&Show+breaks+between+less+comparable+surveys=false
[Incomes Distribution_aux]:
Load
  [ISO3_Code]&'-'&[Year] as [PK_ISO3_YEAR],
  ApplyMap('Map_ISO3_LDC', [ISO3_Code], 'N/A') as [LDC_Flag],
  ApplyMap('Map_ISO3_Region', [ISO3_Code], 'N/A') as [Region],
  *
FROM [lib:///.../Pverty DB.xlsx]
(ooxml, embedded labels, header is 1 lines, table is [Incomes Distribution])
where NUM#([Year]) >= 2000
;

drop field [Entity];
drop field [reporting_level];
drop field [welfare_type];
drop field [Mean income per day $];
drop field [Mean income per year $];
drop field [Median income per year $];
```

```
NoConcatenate
[Incomes Distribution]:
LOAD
    [PK_ISO3_YEAR] ,
    [ISO3 Code] ,
    NUM#([Year]) as [Year],
    If([LDC Flag]='YES',
        if([Year]>=2017, '<= '&$(vPovertyThresholdAfter2017)*$(vUSD_to_EUR)&'€ Median Daily
Income', '<= '&$(vPovertyThresholdBefore2017)*$(vUSD_to_EUR)&'€ Median Daily Income'),
        If([Region]='EU27', '<= 60% Median Daily Income', '<= 50% Median Daily Income')
    ) as [Rule],
    [Median income per day $]*$(vUSD_to_EUR) as [Median
Daily Income €],
    [Poorest decile] as [1 decile]
Mean Daily Income],
    [2nd decile] as [2 decile]
Mean Daily Income],
    [3rd decile] as [3 decile]
Mean Daily Income],
    [4th decile] as [4 decile]
Mean Daily Income],
    [5th decile] as [5 decile]
Mean Daily Income],
    [6th decile] as [6 decile]
Mean Daily Income],
    [7th decile] as [7 decile]
Mean Daily Income],
    [8th decile] as [8 decile]
Mean Daily Income],
    [9th decile] as [9 decile]
Mean Daily Income],
    [Richest decile] as [10 decile]
Mean Daily Income],
    [Poorest decile1] as [1 decile]
Threshold Daily Income],
    [2nd decile1] as [2 decile]
Threshold Daily Income],
    [3rd decile1] as [3 decile]
Threshold Daily Income],
```

```

    [4th decile1]                                as [4 decile]
Threshold Daily Income],
    [5th decile (median)]                        as [5 decile]
Threshold Daily Income],
    [6th decile1]                                as [6 decile]
Threshold Daily Income],
    [7th decile1]                                as [7 decile]
Threshold Daily Income],
    [8th decile1]                                as [8 decile]
Threshold Daily Income],
    [Richest decile1]                            as [9 decile]
Threshold Daily Income],
    [Population]
Resident [Incomes Distribution_aux];

DROP TABLE [Incomes Distribution_aux];

//Poverty headcount ratio at 2.15usd day
//https://data.worldbank.org/indicator/SI.POV.DDAY
NoConcatenate
[PVT_215_aux0]:
Load
*
FROM [lib:///.../Pverty DB.xlsx]
(ooxml, embedded labels, table is [API_SI.POV.DDAY]);

PVT_215_aux1:
//Cross table from YEAR in columns to YEAR in rows
crosstable([Year],[% of Population under 2.15$],1)
Load *
Resident PVT_215_aux0;

drop table PVT_215_aux0;

NoConcatenate
PVT_215:
LOAD
    [Country Code]                                as [ISO3_Code],
    ApplyMap('Map_ISO3_LDC', [Country Code], 'N/A') as [LDC Flag],
    NUM#([Year])                                  as [Year],
    [% of Population under 2.15$]
Resident [PVT_215_aux1];

drop table [PVT_215_aux1];

//Add % of Population under 2.15$ for missing countries in Incomes Distribution
join([Incomes Distribution])
Load
    [ISO3_Code]&'-'&[Year]                        as [PK_ISO3_YEAR]
    ,
    [ISO3_Code]                                    as [ISO3_Code]
    ,
    [Year]                                          as [Year]
    ,
    '<= '&2.15*$(vUSD_to_EUR)&'€ Median Daily Income' as [Rule2],
    [% of Population under 2.15$]
Resident PVT_215
Where [Year]>=2000
and [LDC Flag] = 'YES';

drop table PVT_215;

//-----

//Country Population Table
//https://data.worldbank.org/indicator/SP_POP_TOTL
NoConcatenate
[population_aux0]:
Load
*

```

```
FROM [lib:///Pverty DB.xlsx]
(ooxml, embedded labels, table is [SP.POP.TOTL]);

population:
//Cross table from YEAR in columns to YEAR in rows
crosstable([Year],[Ctr_Population],1)
Load *
Resident population_aux0;

drop table population_aux0;

//Add Countries population for Newly added countries
left join([Incomes Distribution])
Load
    [Country Code]&'-'&[Year]          as [PK_ISO3_YEAR]
    [Country Code]                    as [ISO3_Code]
    NUM#([Year])                      as [Year]
    [Ctr_Population]
Resident population
Where NUM#([Year])>=2000;

drop table population;

//-----

//Population by gender Table
//https://ec.europa.eu/eurostat/databrowser/view/demo_pjan/default/table?lang=en
NoConcatenate
[gender_population_aux0]:
Load
*
FROM [lib:///Pverty DB.xlsx]
(ooxml, embedded labels, table is [gender_pop]);

gender_population_aux:
//Cross table from YEAR in columns to YEAR in rows
crosstable([Year],[Gender_Ctr_Population],5)
Load *
Resident gender_population_aux0;

drop table [gender_population_aux0];

NoConcatenate
gender_population:
Load
    ApplyMap('Map ISO2 ISO3', [geo], 'N/A') as [ISO3_Code],
    NUM#([Year])                             as [Year],
    [Gender_Ctr_Population]                   as [Gender_Ctr_Total_Population]
Resident [gender_population_aux]
Where [sex]='T' and [Gender_Ctr_Population]<>':';

LEFT Join

Load
    ApplyMap('Map ISO2 ISO3', [geo], 'N/A') as [ISO3_Code],
    NUM#([Year])                             as [Year],
    [Gender_Ctr_Population]                   as [Gender_Ctr_Female_Population]
Resident [gender_population_aux]
Where [sex]='F'and [Gender_Ctr_Population]<>':';

LEFT Join

Load
    ApplyMap('Map ISO2 ISO3', [geo], 'N/A') as [ISO3_Code],
    NUM#([Year])                             as [Year],
    [Gender_Ctr_Population]                   as [Gender_Ctr_Male_Population]
Resident [gender_population_aux]
Where [sex]='M'and [Gender_Ctr_Population]<>':';
```

```
drop table [gender_population_aux];

//Add Countries population by gender for all countries
left join([Incomes Distribution])
Load
    [ISO3_Code]& '-'&[Year]           as [PK_ISO3_YEAR]
    [ISO3_Code]                     as [ISO3_Code]
    NUM#([Year])                     as [Year]
    [Gender_Ctr_Total_Population],
    [Gender_Ctr_Female_Population],
    [Gender_Ctr_Male_Population]
Resident gender_population
Where NUM#([Year])>=2000;

drop table gender_population;

//-----

//Income by gender
//https://ec.europa.eu/eurostat/databrowser/view/ilc_di03/default/table?lang=en
NoConcatenate
[gender_income_aux0]:
Load
    *
FROM [lib:///.../Pverty DB.xlsx]
(ooxml, embedded labels, table is [ilc_di03 - MedIncomeGender])
Where [unit] = 'EUR'
and [indic_il] = 'MED_E'
and [age] = 'TOTAL';

gender_income_aux:
//Cross table from YEAR in columns to YEAR in rows
crosstable([Year],[Gender_med_income],6)
Load *
Resident gender_income_aux0;

drop table [gender_income_aux0];

NoConcatenate
gender_income:
Load
    ApplyMap('Map_ISO2_ISO3', [geo], 'N/A')           as [ISO3_Code],
    NUM#([Year])                                       as [Year],
    ApplyMap('Map_ISO2_Region', [geo], 'N/A')         as [Region],
    [Gender_med_income]                               as [Median Daily Total Income €]
Resident [gender_income_aux]
Where [sex]='T' and [Gender med income]<>':.';

LEFT Join

Load
    ApplyMap('Map_ISO2_ISO3', [geo], 'N/A')           as [ISO3_Code],
    NUM#([Year])                                       as [Year],
    ApplyMap('Map_ISO2_Region', [geo], 'N/A')         as [Region],
    [Gender_med_income]                               as [Median Daily Female Income €]
Resident [gender_income_aux]
Where [sex]='F' and [Gender_med_income]<>':.';

LEFT Join

Load
    ApplyMap('Map_ISO2_ISO3', [geo], 'N/A')           as [ISO3_Code],
    NUM#([Year])                                       as [Year],
    ApplyMap('Map_ISO2_Region', [geo], 'N/A')         as [Region],
    [Gender_med_income]                               as [Median Daily Male Income €]
Resident [gender_income_aux]
Where [sex]='M' and [Gender_med_income]<>':.';

drop table [gender_income_aux];
```

```
//Add median income by gender for all countries
left join([Incomes Distribution])
Load
    [ISO3_Code]&'-'&[Year]          as [PK_ISO3_YEAR]
    [ISO3_Code]                    as [ISO3_Code]
    NUM#([Year])                    as [Year]
    [Median Daily Total Income €],
    [Median Daily Female Income €],
    [Median Daily Male Income €]
Resident gender_income
Where NUM#([Year])>=2000
and Region = 'EU27';

drop table gender_income;

//-----

Map_ISO3_MaxYear:
mapping LOAD distinct [ISO3_Code] , max(NUM#([Year])) as MaxYear
Resident [Incomes Distribution]
group by [ISO3_Code] ;

[Population Income]:
Load
    [ISO2_Code]                    as [ISO2 Code],
    [ISO3_Code]                    as [ISO3 Code],
    [Country]                      ,
    [Continent]                    ,
    [Region]                       ,
    [LDC Flag]                     ,
    [Year]                         ,
    [Rule]                         ,
    [Median Daily Income €]        as [Median Daily
Income €],
    [Median Daily Female Income €] as [Median Daily
Female Income €],
    [Median Daily Male Income €]   as [Median Daily Male
Income €],
    [Population]                   ,
    [% Female Population]          as [% Female
Population],
    floor([Population]*[% Female Population]) as [Female Population],
    [% Male Population]            as [% Male Population],
    floor([Population]*[% Male Population])   as [Male Population],
    Num(if(len(keepchar([% of Population under Poverty Line],'0123456789.'))=0
num(keepchar([% of Population under Poverty Line],'0123456789.')=0,
    keepchar([% of Population under 2.15$],'0123456789.'),
    keepchar([% of Population under Poverty Line],'0123456789.')
),'##.00') as [% of Population
under Poverty Line],
    Num([% of Female Population under Poverty Line],'##.00') as [% of Female
Population under Poverty Line],
    Num([% of Male Population under Poverty Line],'##.00') as [% of Male
Population under Poverty Line],
    if([Year]=ApplyMap('Map_ISO3_MaxYear', [ISO3_Code],'N/A'),1,0) as [MaxYear
Flag],
    if([Year]=ApplyMap('Map_ISO3_MaxYear', [ISO3_Code],'N/A'),[Year],'N/A') as [MaxYear]
;
Load
    *,
    IF( [Poverty Line Amt €] <= [1 decile Threshold Daily Income], [Poverty Line Amt
€]/([1 decile Threshold Daily Income]/10),
    IF( [Poverty Line Amt €] <= [2 decile Threshold Daily Income], [Poverty Line Amt
€]/([2 decile Threshold Daily Income]/20),
    IF( [Poverty Line Amt €] <= [3 decile Threshold Daily Income], [Poverty Line Amt
€]/([3 decile Threshold Daily Income]/30),
    IF( [Poverty Line Amt €] <= [4 decile Threshold Daily Income], [Poverty Line Amt
€]/([4 decile Threshold Daily Income]/40),
    IF( [Poverty Line Amt €] <= [5 decile Threshold Daily Income], [Poverty Line Amt
€]/([5 decile Threshold Daily Income]/50),
```

```

    IF( [Poverty Line Amt €] <= [6 decile Threshold Daily Income], [Poverty Line Amt
€]/([6 decile Threshold Daily Income]/60),
    IF( [Poverty Line Amt €] <= [7 decile Threshold Daily Income], [Poverty Line Amt
€]/([7 decile Threshold Daily Income]/70),
    IF( [Poverty Line Amt €] <= [8 decile Threshold Daily Income], [Poverty Line Amt
€]/([8 decile Threshold Daily Income]/80),
    IF( [Poverty Line Amt €] <= [9 decile Threshold Daily Income], [Poverty Line Amt
€]/([9 decile Threshold Daily Income]/90),
    0))))))))) as [% of Population
under Poverty Line],

    IF( [Female Poverty Line Amt €] <= [1 decile Threshold Daily Income], [Female Poverty
Line Amt €]/([1 decile Threshold Daily Income]/10),
    IF( [Female Poverty Line Amt €] <= [2 decile Threshold Daily Income], [Female Poverty
Line Amt €]/([2 decile Threshold Daily Income]/20),
    IF( [Female Poverty Line Amt €] <= [3 decile Threshold Daily Income], [Female Poverty
Line Amt €]/([3 decile Threshold Daily Income]/30),
    IF( [Female Poverty Line Amt €] <= [4 decile Threshold Daily Income], [Female Poverty
Line Amt €]/([4 decile Threshold Daily Income]/40),
    IF( [Female Poverty Line Amt €] <= [5 decile Threshold Daily Income], [Female Poverty
Line Amt €]/([5 decile Threshold Daily Income]/50),
    IF( [Female Poverty Line Amt €] <= [6 decile Threshold Daily Income], [Female Poverty
Line Amt €]/([6 decile Threshold Daily Income]/60),
    IF( [Female Poverty Line Amt €] <= [7 decile Threshold Daily Income], [Female Poverty
Line Amt €]/([7 decile Threshold Daily Income]/70),
    IF( [Female Poverty Line Amt €] <= [8 decile Threshold Daily Income], [Female Poverty
Line Amt €]/([8 decile Threshold Daily Income]/80),
    IF( [Female Poverty Line Amt €] <= [9 decile Threshold Daily Income], [Female Poverty
Line Amt €]/([9 decile Threshold Daily Income]/90),
    0))))))))) as [% of Female
Population under Poverty Line],

    IF( [Male Poverty Line Amt €] <= [1 decile Threshold Daily Income], [Male Poverty Line
Amt €]/([1 decile Threshold Daily Income]/10),
    IF( [Male Poverty Line Amt €] <= [2 decile Threshold Daily Income], [Male Poverty Line
Amt €]/([2 decile Threshold Daily Income]/20),
    IF( [Male Poverty Line Amt €] <= [3 decile Threshold Daily Income], [Male Poverty Line
Amt €]/([3 decile Threshold Daily Income]/30),
    IF( [Male Poverty Line Amt €] <= [4 decile Threshold Daily Income], [Male Poverty Line
Amt €]/([4 decile Threshold Daily Income]/40),
    IF( [Male Poverty Line Amt €] <= [5 decile Threshold Daily Income], [Male Poverty Line
Amt €]/([5 decile Threshold Daily Income]/50),
    IF( [Male Poverty Line Amt €] <= [6 decile Threshold Daily Income], [Male Poverty Line
Amt €]/([6 decile Threshold Daily Income]/60),
    IF( [Male Poverty Line Amt €] <= [7 decile Threshold Daily Income], [Male Poverty Line
Amt €]/([7 decile Threshold Daily Income]/70),
    IF( [Male Poverty Line Amt €] <= [8 decile Threshold Daily Income], [Male Poverty Line
Amt €]/([8 decile Threshold Daily Income]/80),
    IF( [Male Poverty Line Amt €] <= [9 decile Threshold Daily Income], [Male Poverty Line
Amt €]/([9 decile Threshold Daily Income]/90),
    0))))))))) as [% of Male
Population under Poverty Line];

Load
*,
    IF([LDC Flag]='YES',
        IF([Year]>=2017, $(vPovertyThresholdAfter2017)*$(vUSD_to_EUR),
$(vPovertyThresholdBefore2017)*$(vUSD_to_EUR)),
        IF([Region]='EU27', 0.6*([Median Daily Income €])
, 0.5*([Median Daily Income €]))
    ) as [Poverty
Line Amt €],
    IF([LDC Flag]='YES',
        IF([Year]>=2017, $(vPovertyThresholdAfter2017)*$(vUSD_to_EUR),
$(vPovertyThresholdBefore2017)*$(vUSD_to_EUR)),
        IF([Region]='EU27', 0.6*([Median Daily Income €])+([Median Daily Income €]*[% Male
Income]/100))
, 0.5*([Median Daily Income €]))
    ) as [Female
Poverty Line Amt €],
    IF([LDC Flag]='YES',
        IF([Year]>=2017, $(vPovertyThresholdAfter2017)*$(vUSD_to_EUR),
$(vPovertyThresholdBefore2017)*$(vUSD_to_EUR)),

```

```

        If([Region]='EU27', 0.6*([Median Daily Income €]+([Median Daily Income €]*[% Female
Income]/100))
            , 0.5*([Median Daily Income €]))
    )
    Poverty Line Amt €,
    [Median Daily Income €]+([Median Daily Income €]*[% Female Income]/100) as [Median
Daily Female Income €],
    [Median Daily Income €]+([Median Daily Income €]*[% Male Income]/100) as [Median
Daily Male Income €]
;

Load
    ApplyMap('Map_ISO3_ISO2', [ISO3_Code], 'N/A') as [ISO2_Code],
    [ISO3_Code] as [ISO3_Code],
    ApplyMap('Map_ISO3_Country', [ISO3_Code], 'N/A') as [Country],
    ApplyMap('Map_ISO3_Continent', [ISO3_Code], 'N/A') as [Continent],
    ApplyMap('Map_ISO3_Region', [ISO3_Code], 'N/A') as [Region],
    ApplyMap('Map_ISO3_LDC', [ISO3_Code], 'N/A') as [LDC Flag],
    NUM#([Year]) as [Year],
    if(len([Rule])=0, [Rule2], [Rule]) as [Rule],

    if(len([Median Daily Income €])=0,
        keepchar([Median Daily Total Income €], '0123456789.)/365,
        [Median Daily Income €]) as [Median Daily
Income €],
    ((keepchar([Median Daily Female Income €], '0123456789.)*100)
    /
    keepchar([Median Daily Total Income €], '0123456789.))-100 as [% Female
Income],

    ((keepchar([Median Daily Male Income €], '0123456789.)*100)
    /
    keepchar([Median Daily Total Income €], '0123456789.))-100 as [% Male
Income],

    if(len([Population])=0,
        keepchar([Ctr_Population], '0123456789. '),
        keepchar([Population], '0123456789. ')) as [Population],

    (keepchar([Gender_Ctr_Female_Population], '0123456789. '))
    /
    keepchar([Gender_Ctr_Total_Population], '0123456789. ') as [% Female
Population],

    //-----
    (keepchar([Gender_Ctr_Male_Population], '0123456789. '))
    /
    keepchar([Gender_Ctr_Total_Population], '0123456789. ') as [% Male
Population],
    [1 decile Threshold Daily Income] ,
    [2 decile Threshold Daily Income] ,
    [3 decile Threshold Daily Income] ,
    [4 decile Threshold Daily Income] ,
    [5 decile Threshold Daily Income] ,
    [6 decile Threshold Daily Income] ,
    [7 decile Threshold Daily Income] ,
    [8 decile Threshold Daily Income] ,
    [9 decile Threshold Daily Income] ,
    [% of Population under 2.15$]
Resident [Incomes Distribution];

left JOIN ([Population Income])
LOAD DISTINCT
    [ISO3_Code],
    max(NUM#([Year])) as MaxYear
Resident [Population Income]
Group by [ISO3_Code];

Drop Table [Incomes Distribution];
Drop Table [Countries];

```

```

NoConcatenate
[Gender Incomes Distribution]:
Load distinct
  [ISO3 Code],
  [Year],
  'Male' as [Gender],
  [Male Population] as [Gender Population],
  [% Male Population] as [Gender % Population],
  [Median Daily Male Income €] as [Median Daily Gender Income €],
  [% of Male Population under Poverty Line] as [% of Gender Population under Poverty
Line]
Resident [Population Income];

Load distinct
  [ISO3 Code],
  [Year],
  'Female' as [Gender],
  [Female Population] as [Gender Population],
  [% Female Population] as [Gender % Population],
  [Median Daily Female Income €] as [Median Daily Gender Income €],
  [% of Female Population under Poverty Line] as [% of Gender Population under
Poverty Line]
Resident [Population Income];

Drop field [% Male Population] from [Population Income];
Drop field [Male Population] from [Population Income];
Drop field [% Female Population] from [Population Income];
Drop field [Female Population] from [Population Income];
Drop field [Median Daily Male Income €] from [Population Income];
Drop field [Median Daily Female Income €] from [Population Income];
Drop field [% of Male Population under Poverty Line] from [Population Income];
Drop field [% of Female Population under Poverty Line] from [Population Income];

```