

Projet DATA Warehouse

SUICIDE

Réalisé par :

Soumaya EL AMRANI

Encadré par :

Mme Rhanoui Maryem

L'Ecole des Sciences de l'Information

2021-2022

Table des matières

Introduction :	4
Sources de données :	4
Description de la base de données :	5
Outils utilisés :	6
Préparation et Nettoyage de la base de données :	7
Matrice multidimensionnelle :	9
Conception de la DATA Warehouse :	10
Modèle physique des données :	11
Processus ETL :	11
Dimension mental health :	11
Dimension unemployment :	12
Dimension addiction :	14
Dimension population :	14
Dimension Continent :	15
Dimension year :	16
Table de faits : suicide :	17
Résultat final :	18
Analyse des données :	19
Global Suicide:	19
Taux de Suicide au monde et les 10 top pays:	20
Taux de Suicide par tranche d'âge :	21
Taux de Suicide par continent:	22
Taux de mortalité causé par le suicide selon le sexe et les tranches d'âge:	24
Taux de suicide par rapport le taux de chômage :	24
Conclusion :	25

Liste des figures :

Figure 1: Code pour Continent	8
Figure 2: Modification de la colonne âge et création de génération	8
Figure 3 : Conception de la data warehouse	10
Figure 4 : modèle physique des données	11
Figure 5 : Dim_MentalHealth	12
Figure 6: Résultat Dim_MentalHealth	12
Figure 7: Dim_Unemployment	13
Figure 8: Résultat Dim_unemployment	13
Figure 9: Dim_Addiction	14
Figure 10: Résultat Dim_Addiction	14
Figure 11 : Dim_Population	15
Figure 12 : Résultat Dim_Population	15
Figure 13: Dim_Continent	16
Figure 14 : Résultat Dim_Continent	16
Figure 15 : Dim_Year	17
Figure 16 : Résultat Dim_Year	17
Figure 17 : Table de fait Fact_Suicide	18
Figure 18: le résultat de la table de fait	18
Figure 19 : ETL FINAL	19
Figure 20: Taux de suicide par 100K - RSTUDIO	19
Figure 21 : Total de suicide au monde	20
Figure 22: le suicide au monde 2002	20
Figure 23: Les top 10 pays	21
Figure 24: Code pour calculer le suicide par population/continent	21
Figure 25 : le Suicide par âge et continent	22
Figure 26: Code pour le taux de suicide par 100K - Rstudio	23
Figure 27: Le suicide par continent	23
Figure 28: le Suicide par âge et par sexe	24
Figure 29 : Suicide et le chômage	25

Introduction :

D'après les données de l'Organisation mondiale de la santé, environ une personne sur 5 000 à 15 000 meurt par suicide chaque année, avec un taux mondial estimé à 10,5 pour 100 000 habitants contre 11,6 en 2008. Les gens choisissent ainsi de mettre fin à leurs jours pour des raisons complexes, difficiles à déterminer certes mais qu'on peut cerner dans ces fléaux les plus courants :

* **Antécédents de vie:** expérience traumatisante pendant l'enfance et violence sexuelle / physique. * Style de vie: abus d'alcool ou de drogues.

* **Événement de la vie :** perte d'êtres chers ou perte de revenu.

* **Santé mentale:** dépression ou schizophrénie Ici, j'ai exploré quelques facteurs qui, selon moi, auraient un impact sur le taux de suicide.

Sources de données :

Les sources de notre ensemble de données sont les suivantes :

Suicide :

Kaggle | who-suicide-statistics

Population :

Population World – THE WORLD BANK

Chômage :

Unemployment, total (% of total labor force) (modeled ILO estimate)

Santé mentale:

Mental Health and Suicide – Kaggle

Addiction aux drogues :

Drogues Consumption – GitHub

Description de la base de données :

Nous comptons utiliser un dataset «[who-suicide-statistics](#)» de la plateforme publique Kaggle, puis alimenter ce dataset par d'autres bases de données à partir de l'organisation mondiale de la santé et la plateforme publique Kaggle et Github . Et puis ajouter des indicateurs globaux par exemple : la population, le chômage, l'addiction aux drogues et la santé mentale. Les données seront par la suite réparties géographiquement d'une façon équilibrée, et sur plusieurs années entre 1986 et 2015.

Les bases de données avec lesquels j'ai implémenté mon dataset sont les suivantes :

Suicide :

En examinant les données, vous pouvez observer les tendances à long terme et les différences entre les pays, ainsi qu'au sein des pays pour quelques groupes démographiques dont la différence sera considérable dans les deux cas.

Mental Health :

Cette base de données offre le résultat d'une étude qui porte sur les troubles mentaux échelonnés et selon les tranches d'âge ainsi que la situation actuelle de l'individu par rapport à la profession et au traitement des troubles mentaux répartis géographiquement

Chômage :

Cette table réunit la répartition géographique des individus employés et en chômage par rapport à la totalité de la population en se référant à un cadre chronologique

Addiction aux drogues :

Cette table modélise l'utilisation de certains types de drogues(alcool, marijuana,cocaine,heroin,hallucinogen) au cours des années.

Outils utilisés :

SQL Server Integration Services (SSIS)

En utilisant les services d'intégration du serveur SQL, on s'assure que les données sont correctement extraites, et ainsi garantir leur qualité et fournir l'architecture de données appropriée pour des rapports et des analyses efficaces. Les fonctionnalités que SSIS permet sont les suivants :

- ❖ Intégration de données provenant de sources de données multiples et hétérogènes
- ❖ Nettoyage et profilage des données pour améliorer la qualité des données
- ❖ Gestion des données de base
- ❖ Gestion des métadonnées pour garantir la lignée et l'historique des données
- ❖ Gestion de l'information tout au long de son cycle de vie pour améliorer la gouvernance et la conformité

SQL Server

Microsoft SQL Server est un Système de gestion de base de données relationnel et transactionnel développé et commercialisé par Microsoft. Microsoft SQL Server utilise le langage T-SQL (Transact-SQL) pour ses requêtes, c'est une implémentation de SQL qui prend en charge les procédures stockées et les déclencheurs. La dernière version est SQL Server 2017. La première ayant appartenu à Microsoft seul est en 1994.

Microsoft Power BI

Microsoft Power BI est une plateforme de Business intelligence qui fournit aux utilisateurs des outils pour agréger, analyser, visualiser et partager des données. En effet, elle est utilisée pour transformer des données brutes d'une entreprise en des informations utilisées lors de la prise de décision. Elle peut aider à relier des ensembles de données disparates, à transformer et à nettoyer les données en un modèle de données et à créer des tableaux ou des graphiques pour fournir des représentations visuelles des données. Tout cela peut être partagé avec d'autres utilisateurs de Power BI au sein de l'organisation. Power BI peut également fournir des tableaux de bord pour les administrateurs ou les gestionnaires, ce qui permet à la direction d'avoir une meilleure idée de la situation des services.

Visual studio :

Microsoft Visual Studio est une suite de logiciels de développement pour Windows et mac OS conçue par Microsoft. La dernière version s'appelle Visual Studio . Visual Studio est un ensemble complet d'outils de développement permettant de générer des applications web ASP.NET, des services web XML, des applications bureautiques et des applications mobiles. Visual Basic, Visual C++, Visual C# utilisent tous le même environnement de développement intégré (IDE), qui leur permet de partager des outils et facilite la création de solutions faisant appel à plusieurs langages. Par ailleurs,

ces langages permettent de mieux tirer parti des fonctionnalités du framework .NET, qui fournit un accès à des technologies clés simplifiant le développement d'applications web ASP et de services web XML grâce à Visual Web Developer.

RStudio :

RStudio est un environnement de développement gratuit, libre et multiplateforme pour R, un langage de programmation utilisé pour le traitement de données et l'analyse statistique. Il est disponible sous la licence libre AGPLv3, ou bien sous une licence commerciale, soumise à un abonnement annuel.

RStudio est disponible en deux versions : RStudio Desktop, pour une exécution locale du logiciel comme tout autre application, et RStudio Server qui, lancé sur un serveur Linux, permet d'accéder à RStudio par un navigateur web. Des distributions de RStudio Desktop sont disponibles pour Microsoft Windows, OS X et GNU/Linux³.

RStudio a été écrit en langage C++, et son interface graphique utilise l'interface de programmation Qt.

Préparation et Nettoyage de la base de données :

OpenRefine, est un logiciel libre de nettoyage et de mise en forme de données, qui utilise le langage GREL afin de faciliter la préparation des données.

En utilisant le logiciel OpenRefine, j'ai réussi à faire le nettoyage des données, en attribuant aux valeurs manquantes des valeurs significatives :

Valeurs Manquantes → 0: ID, Year,

Valeurs Manquantes → 9999: Tous les champs pour lesquels la valeur 0 est significative.

Valeurs Manquantes → N/A: Tous les champs de type chaîne de caractères. Ensuite, j'ai transformé tout type de données « Female », « Woman ».... Ou « Man », « Male ».... En : Female → F et Male → M Enfin, J'ai mis en majuscule tous type de données chaîne de caractères

Par Rstudio ,

j'ai essayé d'ajouter deux colonnes et de modifier la colonne de l'âge :



Continent : en utilisant la bibliothèque « **rworldmap** » pour lier

Chaque pays à son continent .


```

library(tidyverse) # general
library(ggalt) # dumbbell plots
library(countrycode) # continent
library(rworldmap) # quick country-level heat maps
library(gridExtra) # plots
library(broom) # significant trends within countries
theme_set(theme_light())
# 1) Import & data cleaning
data <- read_excel("c:/Users/pc/Pictures/master.xls")
data <- data %>%
  select(-c(`HDI for year`, `suicides/100k pop`)) %>%
  rename(gdp_for_year = `gdp_for_year ($)`,
         gdp_per_capita = `gdp_per_capita ($)`,
         country_year = `country-year`) %>%
  as.data.frame()
data <- data %>%
  filter(year != 2016) %>% # I therefore exclude 2016 data
  select(-country_year)

minimum_years <- data %>%
  group_by(country) %>%
  summarize(rows = n(),
            years = rows / 12) %>%
  arrange(years)
data <- data %>%
  filter(!(country %in% head(minimum_years$country, 7)))
data$age <- gsub(" years", "", data$age)
data$sex <- ifelse(data$sex == "male", "Male", "Female")

data$continent <- countrycode(sourcevar = data[, "country"],
                             origin = "country.name",
                             destination = "continent")

data_nominal <- c('country', 'sex', 'continent')
data[data_nominal] <- lapply(data[data_nominal], function(x){factor(x)})

```

Figure 1: Code pour Continent

Generation : qui va diviser la population en 4 générations

Age : diviser la colonne selon des tranches

```

data$age <- factor(data$age,
                  ordered = T,
                  levels = c("5-14",
                             "15-24",
                             "25-34",
                             "35-54",
                             "55-74",
                             "75+"))
data$generation <- factor(data$generation,
                         ordered = T,
                         levels = c("G.I. Generation",
                                    "Silent",
                                    "Boomers",
                                    "Generation X",
                                    "Millennials",
                                    "Generation Z"))

data <- as_tibble(data)
global_average <- (sum(as.numeric(data$suicides_no)) / sum(as.numeric(data$population)))
glimpse(data)

```

Figure 2: Modification de la colonne âge et création de génération

Matrice multidimensionnelle :

	Suicide	Mental Health	Chômage	Addiction	Population	Continent	Year	Pays
Quel est la répartition des addictions selon les pays				*		*		*
Les taux des individus employés et ceux en chômage			*				*	*
Les taux de consommation des drogues				*		*	*	*
Quel est le pays avec le taux de suicide le plus élevé et en quelle année	*						*	*
La relation entre self_employed et treatment des mental health		*	*			*	*	*
L'évolution d'utilisation des différentes drogues selon les années				*			*	*
L'évolution de suicide au fil du temps selon les tranches d'âges	*						*	
Le taux de suicide par rapport a la totalité de la population	*				*			
L'utilisation des drogues par rapport au sexe au fil des années	*			*			*	
Le taux de suicide par sexe et année	*						*	

Conception de la DATA Warehouse :

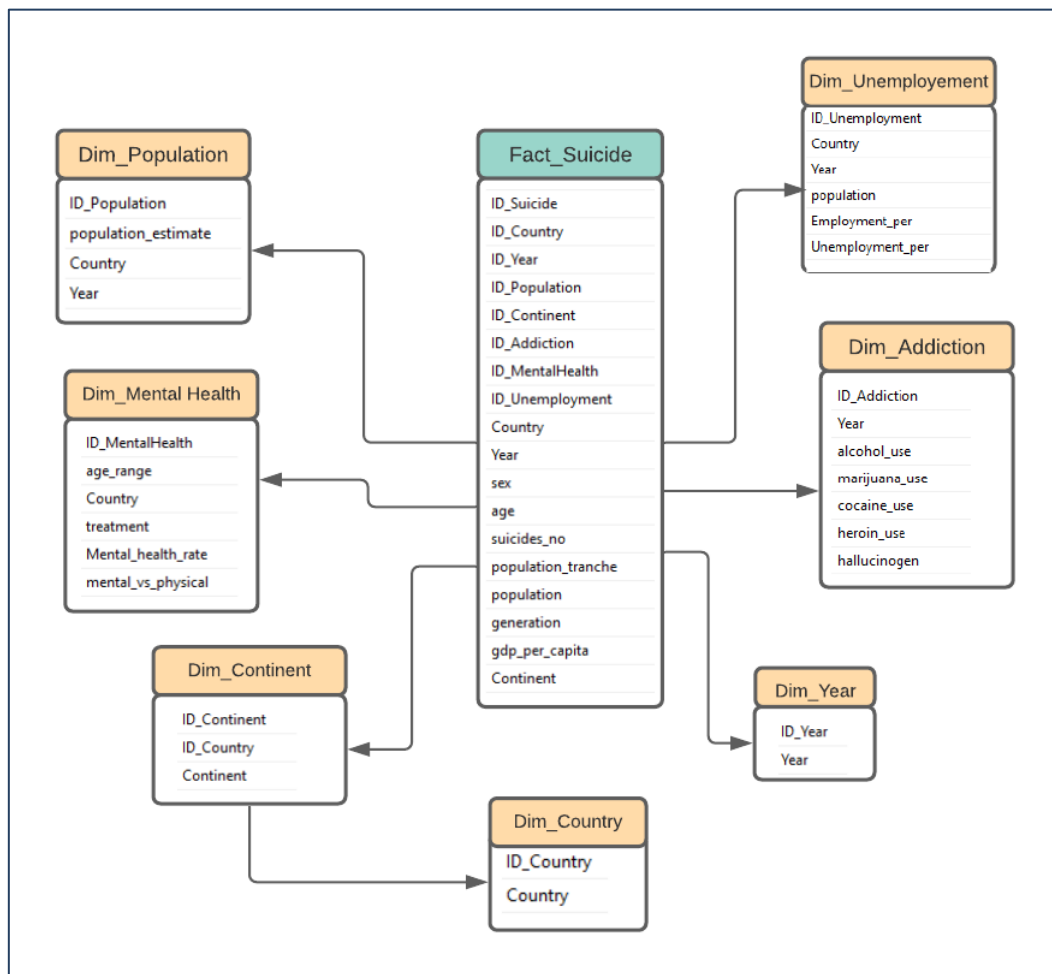


Figure 3 : Conception de la data warehouse

Modèle physique des données

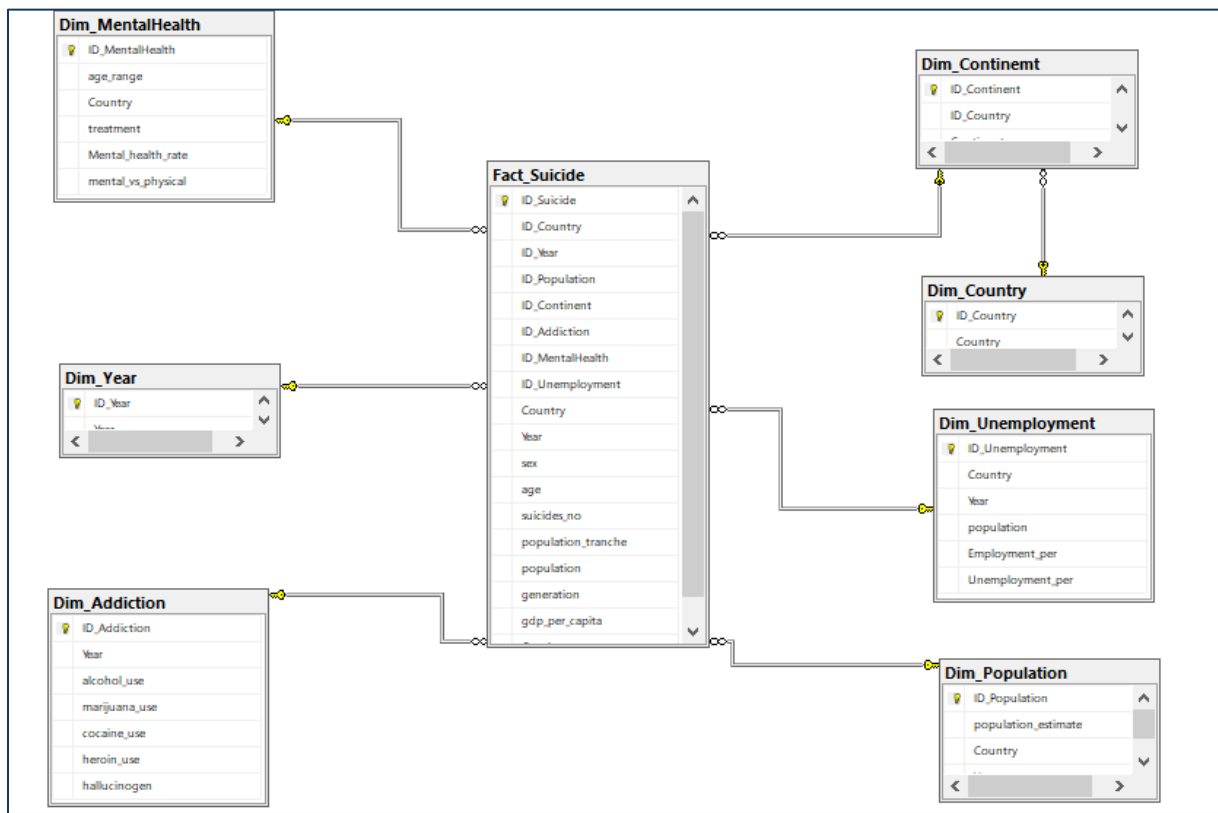


Figure 4 : modèle physique des données

Processus ETL :

Dimension mental health :

Implémentation :

J'ai récupéré les colonnes de la data Warehouse source sur SQL SERVER MANAGEMENT

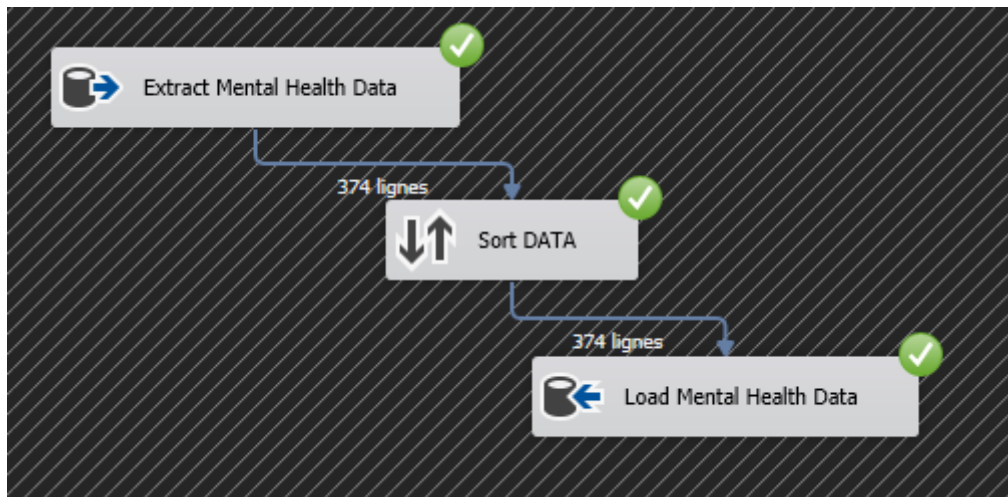


Figure 5 : Dim_MentalHealth

Résultat :

SQLQuery26.sql - D...P-ULTT9B8\HP (59) SQLQuery25.sql - D...P-ULTT9B8\HP (56) SQLQuery24.sql - D...

/***** Script de la commande SelectTopNRows à partir de SSMS *****/

```

SELECT TOP (1000) [ID_MentalHealth]
, [age_range]
, [Country]
, [treatment]
, [Mental_health_rate]
, [mental_vs_physical]
FROM [suicide_DW_project].[dbo].[Dim_MentalHealth]
  
```

100 %

Résultats Messages

	ID_MentalHealth	age_range	Country	treatment	Mental_health_rate	mental_vs_physical
1	1	NULL	NULL	NULL	NULL	NULL
2	2	NULL	NULL	NULL	NULL	NULL
3	3	31-65	Albania	No	22.5	33.0
4	4	21-30	Albania	No	22.5	18.1
5	5	31-65	Albania	No	28.0	10.4
6	6	0-20	Albania	Yes	7.4	2.0
7	7	31-65	Antigua and Barbuda	Yes	28.0	28.4
8	8	0-20	Antigua and Barbuda	No	6.3	3.2
9	9	21-30	Antigua and Barbuda	Yes	22.5	34.0
10	10	31-65	Antigua and Barbuda	No	33.7	10.6
11	11	21-30	Argentina	No	28.0	33.0
12	12	0-20	Argentina	No	5.2	4.1
13	13	31-65	Argentina	No	33.4	8.2
14	14	31-65	Argentina	No	7.4	24.9
15	15	21-30	Amenia	No	33.7	28.4
16	16	31-65	Amenia	Yes	6.3	18.1
17	17	0-20	Amenia	Yes	10.4	4.9

Figure 6: Résultat Dim_MentalHealth

Dimension unemployment :

Implémentation :

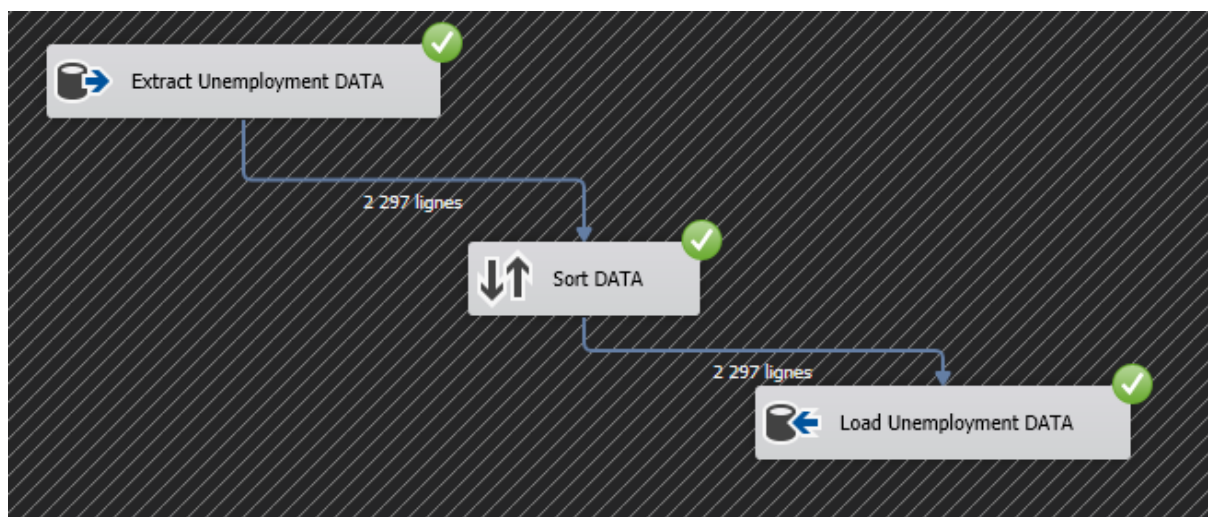


Figure 7: Dim_Unemployment

Résultat :

SQLQuery28.sql - D...P-ULTT9B8\HP (64)) X SQLQuery27.sql - D...P-ULTT9B8\HP (62)) SQL

```

/***** Script de la commande SelectTopNRows à partir de SSMS *****/
SELECT TOP (1000) [ID_Unemployment]
      ,[Country]
      ,[Year]
      ,[population]
      ,[Employment_per]
      ,[Unemployment_per]
FROM [suicide_DW_project].[dbo].[Dim_Unemployment]
  
```

100 %

Résultats Messages

	ID_Unemployment	Country	Year	population	Employment_per	Unemployment_per
27	27	Albania	1999	3,414,760,915	50,4	9,6
28	28	Albania	2000	3,632,043,908	51,6	8,4
29	29	Antig...	1985	240,923,926	49,7	10,3
30	30	Antig...	2006	1,157,005	51,9	8,1
31	31	Antig...	2007	1,311,401	52,4	7,6
32	32	Antig...	2008	1,368,431,037	56,8	3,2
33	33	Antig...	2015	1,364,863,037	54,1	5,9
34	34	Antig...	2013	1,192,925,407	54,2	5,8
35	35	Antig...	2014	1,280,133,333	51,3	8,7
36	36	Antig...	2005	1,022,191	51,9	8,1
37	37	Antig...	2012	1,211,411,704	54,2	5,8
38	38	Antig...	2004	NULL	50,3	9,7
39	39	Antig...	2009	1,224,253,000	54	6
40	40	Antig...	2002	NULL	51,9	8,1
41	41	Antig...	1986	290,440,148	50,1	9,9
42	42	Antig...	1987	337,174,852	54	6
43	43	Antig...	2003	NULL	54,1	5,9

Figure 8: Résultat Dim_unemployment

Dimension addiction :

Implémentation :

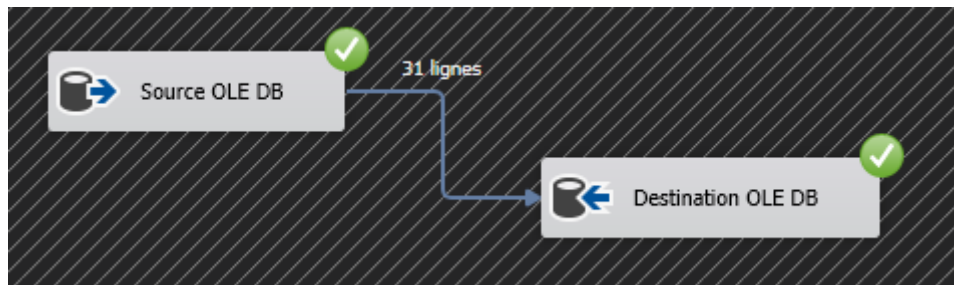


Figure 9: Dim_Addiction

Résultat :

SQLQuery27.sql - D...P-ULTT9B8\HP (62)) SQLQuery26.sql - D...P-ULTT9B8\HP (59)) SQ

/***** Script de la commande SelectTopNRows à partir de SSMS *****/

```
SELECT TOP (1000) [ID_Addiction]
      ,[Year]
      ,[alcohol_use]
      ,[marijuana_use]
      ,[cocaine_use]
      ,[heroin_use]
      ,[hallucinogen]
FROM [suicide_DW_project].[dbo].[Dim_Addiction]
```

100 %

Résultats Messages

	ID_Addiction	Year	alcohol_use	marijuana_use	cocaine_use	heroin_use	hallucinogen
1	1	@YI	3,9	18,1	2,5	10,1	0,2
2	2	@YI	8,5	29,2	3,6	14,5	0,6
3	3	@Y	18,1	40,1	4,7	22,5	1,6
4	4	@YI	29,2	49,3	3,6	28	2,1
5	5	@YI	40,1	22,5	1	33,7	3,4
6	6	@YI	49,3	28	2	33,4	4,8
7	7	@Y	58,7	33,7	3,2	34	7
8	8	@Y	34	33,4	4,1	33	8,6
9	9	@Y\$	33	34	4,9	28,4	7,4
10	10	@Y(28,4	33	4,8	24,9	6,3
11	11	@Y,	24,9	28,4	4,5	20,8	5,2
12	12	@Y0	18,1	24,9	4	16,4	10,4
13	13	@Y4	80,7	18,1	10,4	10,4	10,6
14	14	@Y8	77,5	29,2	33,7	10,6	8,2
15	15	@Y<	75	40,1	33,4	20,8	10,3
16	16	@...	67,2	49,3	14,5	10,3	10,4

Figure 10: Résultat Dim_Addiction

Dimension population :

Implémentation :

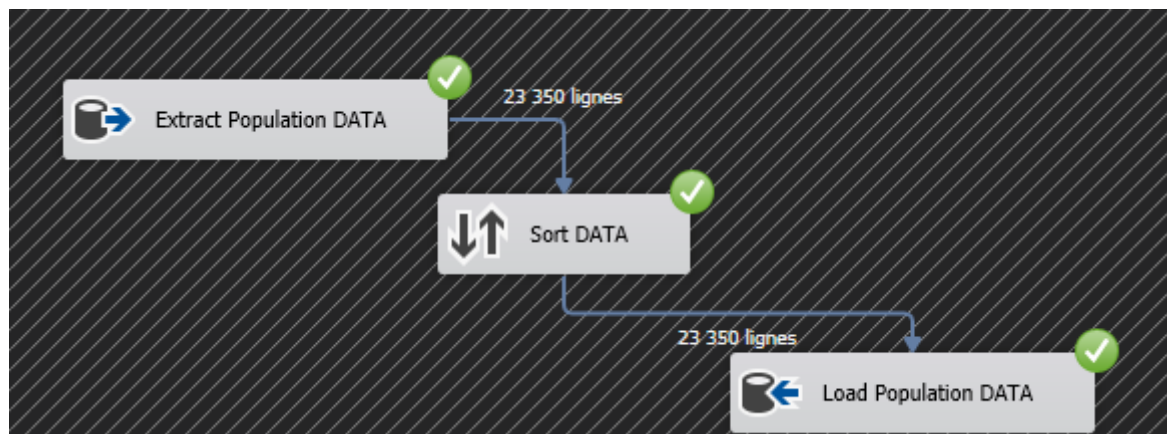


Figure 11 : Dim_Population

Résultat :

SQLQuery22.sql - D...P-ULTT9B8\HP (51)) SQLQuery21.sql - non connecté

```

/***** Script de la commande SelectTopNRows à partir de S
SELECT TOP (1000) [ID_Population]
      ,[population_estimate]
      ,[Country]
      ,[Year]
FROM [suicide_DW_project].[dbo].[Dim_Population]
  
```

100 %

Résultats Messages

	ID_Population	population_estimate	Country	Year
1	1	3,414,760,915	Albania	1999
2	2	2,424,499,009	Albania	1995
3	3	3,314,898,292	Albania	1996
4	4	3,314,898,292	Albania	1996
5	5	3,314,898,292	Albania	1996
6	6	3,314,898,292	Albania	1996
7	7	2,424,499,009	Albania	1995
8	8	3,314,898,292	Albania	1996
9	9	3,314,898,292	Albania	1996
10	10	3,314,898,292	Albania	1996

Figure 12 : Résultat Dim_Population

Dimension Continent :

Implémentation :

Il faut faire un **LOOKUP** pour la récupération de L'ID de la dimension Country pour qu'on puisse lier la dimension de Continent par la dimension des pays.

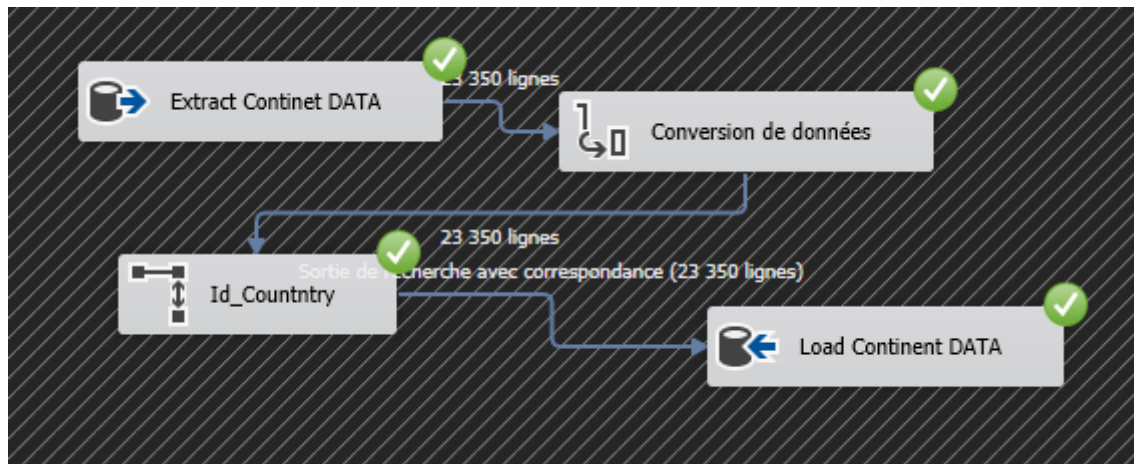


Figure 13: Dim_Continent

Résultat :

SQLQuery25.sql - D...P-ULTT9B8\HP (56) SQLQuery24.sql - D...P-ULTT9B8\HP (56)

```

/***** Script de la commande SelectTopNRows à partir de *****/
SELECT TOP (1000) [ID_Continent]
, [ID_Country]
, [Continent]
FROM [suicide_DW_project].[dbo].[Dim_Continent]
  
```

100 %

Résultats Messages

	ID_Continent	ID_Country	Continent
1	3	7	Europe
2	4	7	Europe
3	5	7	Europe
4	6	7	Europe
5	7	7	Europe
6	8	7	Europe
7	9	7	Europe
8	10	7	Europe
9	11	7	Europe
10	12	7	Europe
11	13	7	Europe
12	14	7	Europe

Figure 14 : Résultat Dim_Continent

Dimension year :

Implémentation :

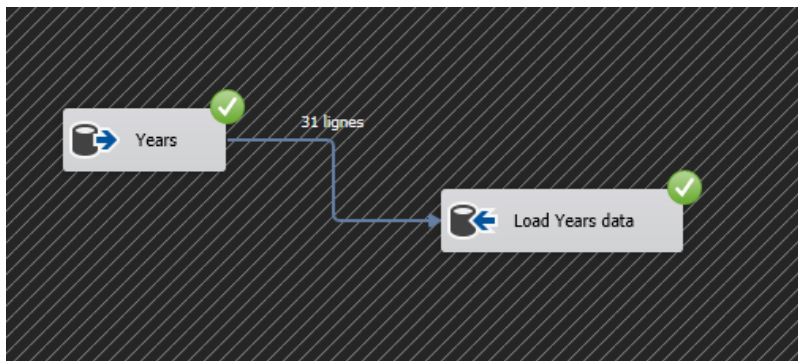


Figure 15 : Dim_Year

Résultat :

SQLQuery24.sql - D...P-ULTT9B8\HP (57) X SQLQuery23.sql - D...

```

/***** Script de la commande SelectTopNRows à pa
SELECT TOP (1000) [ID_Year]
      ,[Year]
FROM [suicide_DW_project].[dbo].[Dim_Year]
  
```

100 %

Résultats Messages

	ID_Year	Year
1	1	1985
2	2	1986
3	3	1987
4	4	1988
5	5	1989
6	6	1990
7	7	1991
8	8	1992
9	9	1993
10	10	1994
11	11	1995
12	12	1996
13	13	1997
14	14	1998

Figure 16 : Résultat Dim_Year

Table de faits : suicide :

Implémentation :

Afin d'implémenter la table de fait , il est nécessaire de faire des **Lookup** afin de récupérer les ID des dimensions pour qu'on puisse lier la table de fait avec l'ensemble de dimensions.

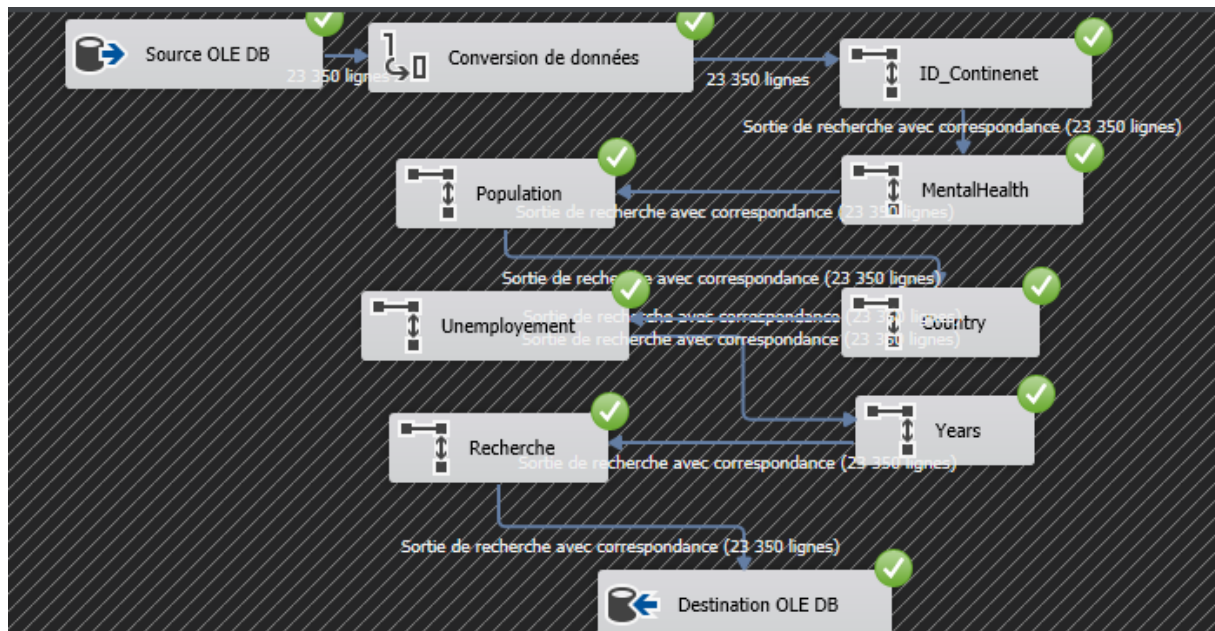


Figure 17 : Table de fait Fact_Suicide

Résultat :

SQLQuery23.sql - D:\P-ULTT988\HP (52) | SQLQuery22.sql - D:\P-ULTT988\HP (51) | SQLQuery21.sql - non connecté | SQLQuery20.sql - non connecté

/***** Script de la commande SelectTopNRows à partir de SSMS *****/

SELECT TOP (1000) [ID_Suicide]

,[ID_Country]

,[ID_Year]

,[ID_Population]

,[ID_Continent]

,[ID_Addiction]

,[ID_MentalHealth]

,[ID_Unemployment]

,[Country]

,[Year]

,[sex]

,[age]

,[suicides_no]

,[population_tranche]

,[population]

,[generation]

	ID_Suicide	ID_Country	ID_Year	ID_Population	ID_Continent	ID_Addiction	ID_MentalHealth	ID_Unemployment	Country	Year	sex	age	suicides_no	population_b
1	3	7	8	1777	3	380	27	155	Austria	1992	Female	25-34	50	660870
2	4	7	8	1777	3	380	27	155	Austria	1992	Female	15-24	33	544520
3	5	7	8	1777	3	380	27	155	Austria	1992	Male	5-14	5	472966
4	6	7	8	1777	3	380	27	155	Austria	1992	Female	5-14	1	448804
5	7	7	9	1777	3	381	27	155	Austria	1993	Male	75+	200	147907
6	8	7	9	1777	3	381	27	155	Austria	1993	Male	55-74	314	627612
7	9	7	9	1777	3	381	27	155	Austria	1993	Male	35-54	377	1059028
8	10	7	9	1777	3	381	27	155	Austria	1993	Male	25-34	228	706145
9	11	7	9	1777	3	381	27	155	Austria	1993	Female	75+	88	339249
10	12	7	9	1777	3	381	27	155	Austria	1993	Male	15-24	119	544899

Figure 18: le résultat de la table de fait

Résultat final :

ETL Final après l'exécution :

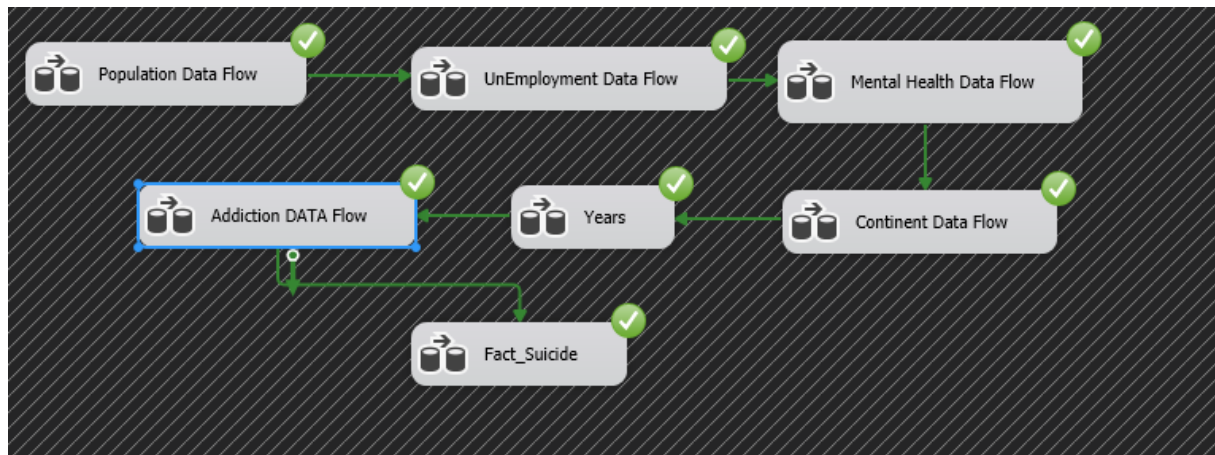


Figure 19 : ETL FINAL

Analyse des données :

Global Suicide:

Le code sur Rstudio pour l'affichage du taux de suicide au monde entre 1985-2015

Alors par ce code en Rstudio , j'ai essayé de calculer le taux de suicide par 100K pour chaque année.

```

data %>%
  group_by(year) %>%
  summarize(population = sum(population),
            suicides = sum(suicides_no),
            suicides_per_100k = (suicides / population) * 100000) %>%
  ggplot(aes(x = year, y = suicides_per_100k)) +
  geom_line(col = "deepskyblue3", size = 1) +
  geom_point(col = "deepskyblue3", size = 2) +
  geom_hline(yintercept = global_average, linetype = 2, color = "grey35", size = 1) +
  labs(title = " Global suicides (per 100k)",
        subtitle = "Trend over time, 1985 - 2015.",
        x = "Year",
        y = "Suicides per 100k") +
  scale_x_continuous(breaks = seq(1985, 2015, 2)) +
  scale_y_continuous(breaks = seq(10, 20))
install.packages("writexl")
  
```

Figure 20: Taux de suicide par 100K - RSTUDIO

On remarque que à 1995 le monde a connu un taux de suicide très élevé :

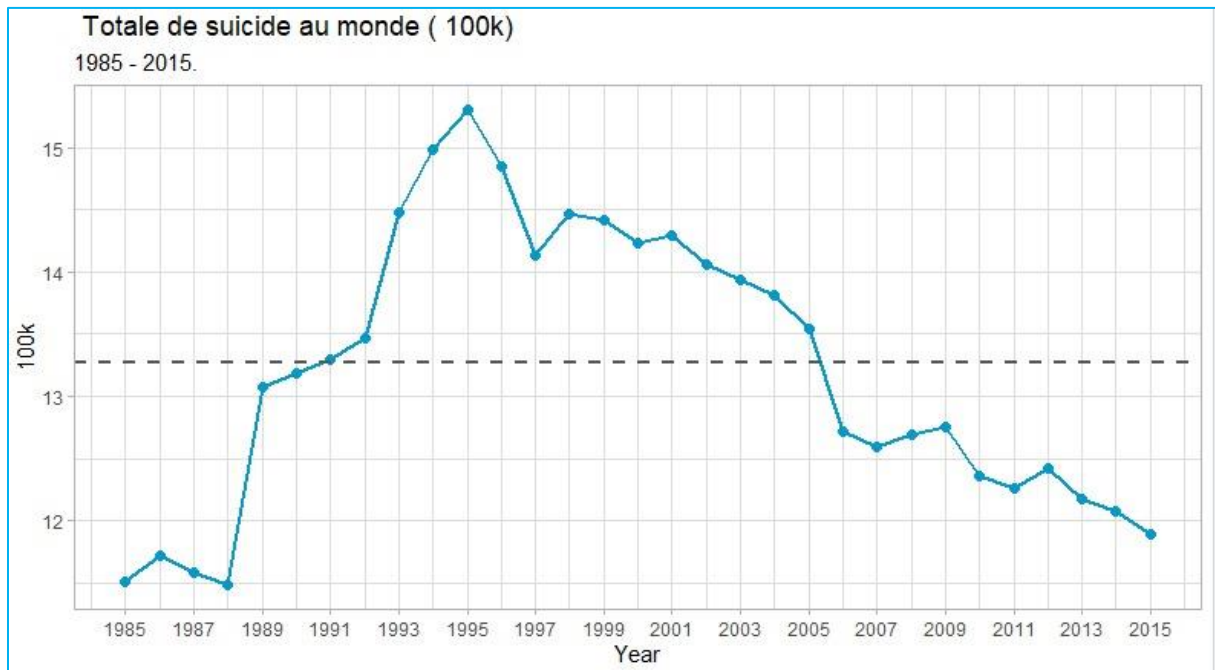


Figure 21 : Total de suicide au monde

Taux de Suicide au monde et les 10 top pays:

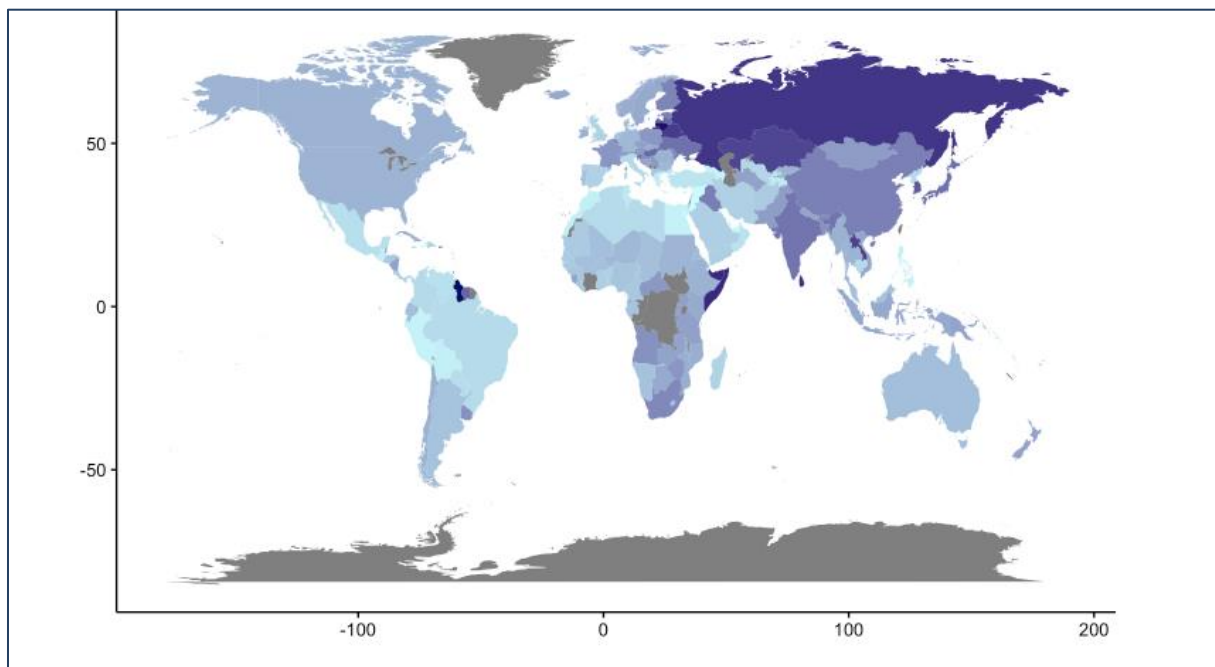


Figure 22: le suicide au monde 2002

On voit que L'Europe et l'Asie connuent des taux de suicide plus élevé en 2005 cependant on trouve que le Maroc est parmi les pays qui connu un taux faible du suicide en 2005

Les 10 top pays dont le taux de suicide le plus élevé en 2002

Country	Suicide Rate
Guyana	35.75287
Lithuania	33.34186
Somalia	29.86416
Sri Lanka	28.10405
Russia	27.87416
Belarus	26.87469
Laos	26.21920
Kazakhstan	25.40460
South Korea	22.40456
Maldives	22.35348

Figure 23: Les top 10 pays

Guyana a pris le dessus par un taux de 35% , et Lithuania dans la 2ème place par un taux de presque 33% ,on voit aussi que Russe a connu aussi un taux élevé de plus que 27% de sa popularité

Taux de Suicide par tranche d'âge :

Le taux de suicide par tranche d'âge dans les 5 continents :

```
data %>%
  group_by(continent, age) %>%
  summarize(n = n(),
            suicides = sum(as.numeric(suicides_no)),
            population = sum(as.numeric(population)),
            suicide_per_100k = (suicides / population) * 100000) %>%
  ggplot(aes(x = continent, y = suicide_per_100k, fill = age)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_hline(yintercept = global_average, linetype = 2, color = "grey35", size = 1) +
  labs(title = "tranche d'age, par continent",
       x = "Continent",
       y = "Suicides per 100k",
       fill = "Age")
```

Figure 24: Code pour calculer le suicide par population/continent

- La tranche d'âge 75+ est la dominante dans les différents continents
- Un taux faible pour la tranche d'âge 5-14
- Et un taux intéressant pour la tranche d'âge 25-34 :
- Plus de 15 personnes/100K pour 5 continents.

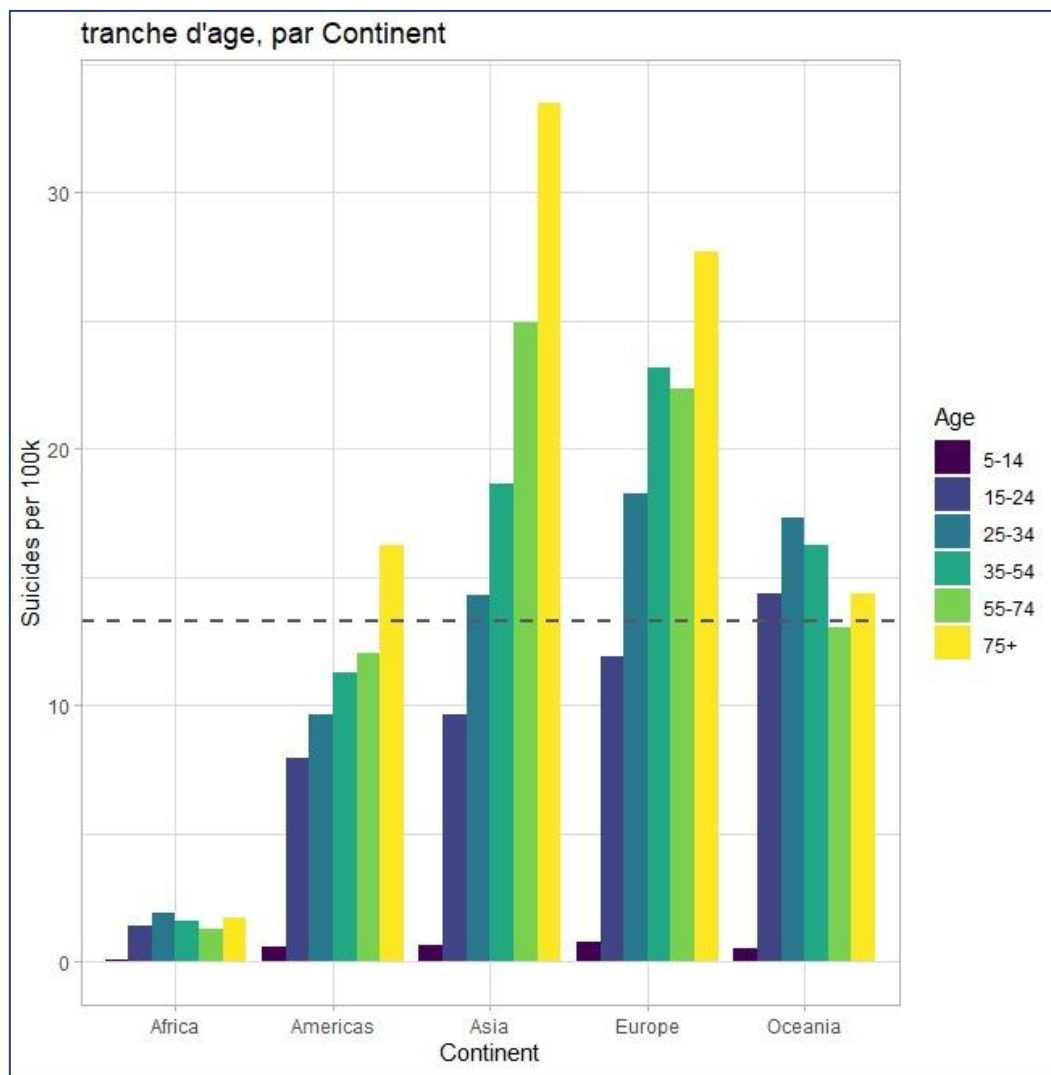


Figure 25 : le Suicide par âge et continent

Taux de Suicide par continent:

Le code Rstudio :


```

continent <- data %>%
  group_by(continent) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) / sum(as.numeric(population))) * 100000) %>%
  arrange(suicide_per_100k)

continent$continent <- factor(continent$continent, ordered = T, levels = continent$continent)

continent_plot <- ggplot(continent, aes(x = continent, y = suicide_per_100k, fill = continent)) +
  geom_bar(stat = "identity") +
  labs(title = "TOTAL Suicide (per 100k), par Continent",
       x = "Continent",
       y = "Suicides per 100k",
       fill = "Continent") +
  theme(legend.position = "none", title = element_text(size = 10)) +
  scale_y_continuous(breaks = seq(0, 20, 1), minor_breaks = F)

continent_time <- data %>%
  group_by(year, continent) %>%
  summarize(suicide_per_100k = (sum(as.numeric(suicides_no)) / sum(as.numeric(population))) * 100000)

continent_time$continent <- factor(continent_time$continent, ordered = T, levels = continent$continent)

continent_time_plot <- ggplot(continent_time, aes(x = year, y = suicide_per_100k, col = factor(continent))) +
  facet_grid(continent ~ ., scales = "free_y") +
  geom_line() +
  geom_point() +
  labs(title = "tendances au fil du temps, by Continent",
       x = "Year",
       y = "Suicides per 100k",
       color = "Continent") +
  theme(legend.position = "none", title = element_text(size = 10)) +
  scale_x_continuous(breaks = seq(1985, 2015, 5), minor_breaks = F)
grid.arrange(continent_plot, continent_time_plot, ncol = 2)

```

Figure 26: Code pour le taux de suicide par 100K - Rstudio

On remarque que le taux de suicide à l'Europe présente le taux le plus élevé mais il déminue au fil de temps .

Le taux de suicide devient de plus en plus intéressant pour l'Amérique au fils de temps

Entre 1985 et 1995 : l'Afrique a connu un taux intéressant pour les gens qui décident de finir leur vie, mais il est devient faible à partir des année 2000 jusqu'à maintenant.

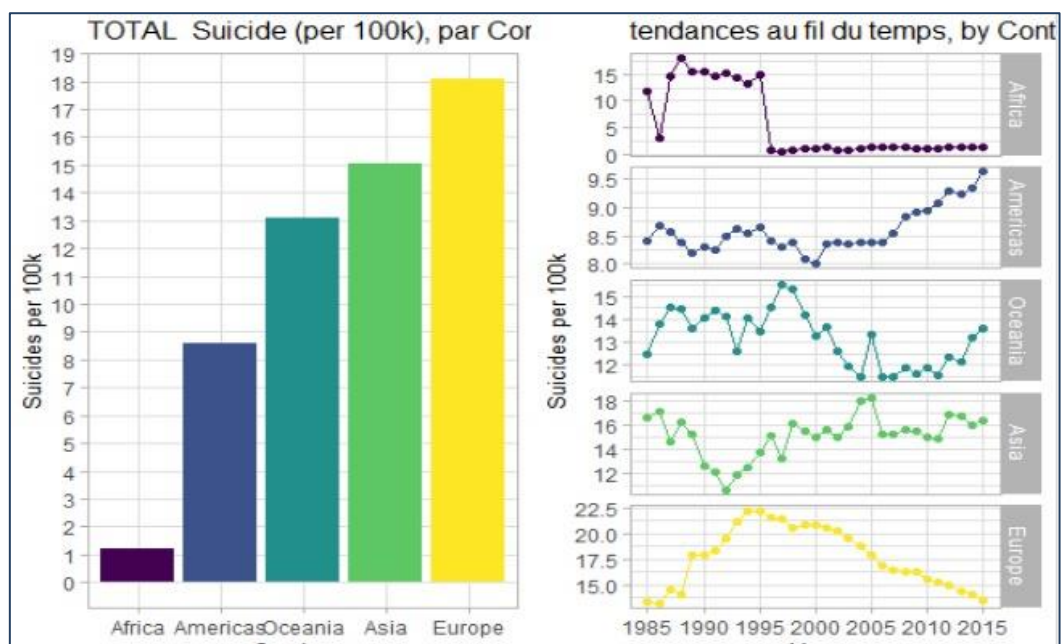


Figure 27: Le suicide par continent

Taux de mortalité causé par le suicide selon le sexe et les tranches d'âge:

- ✓ Les personnes âgées de 15 à 29 ans sont les plus à risque de se suicider.
- ✓ Plus d'hommes se suicident que de femmes de tous âges.

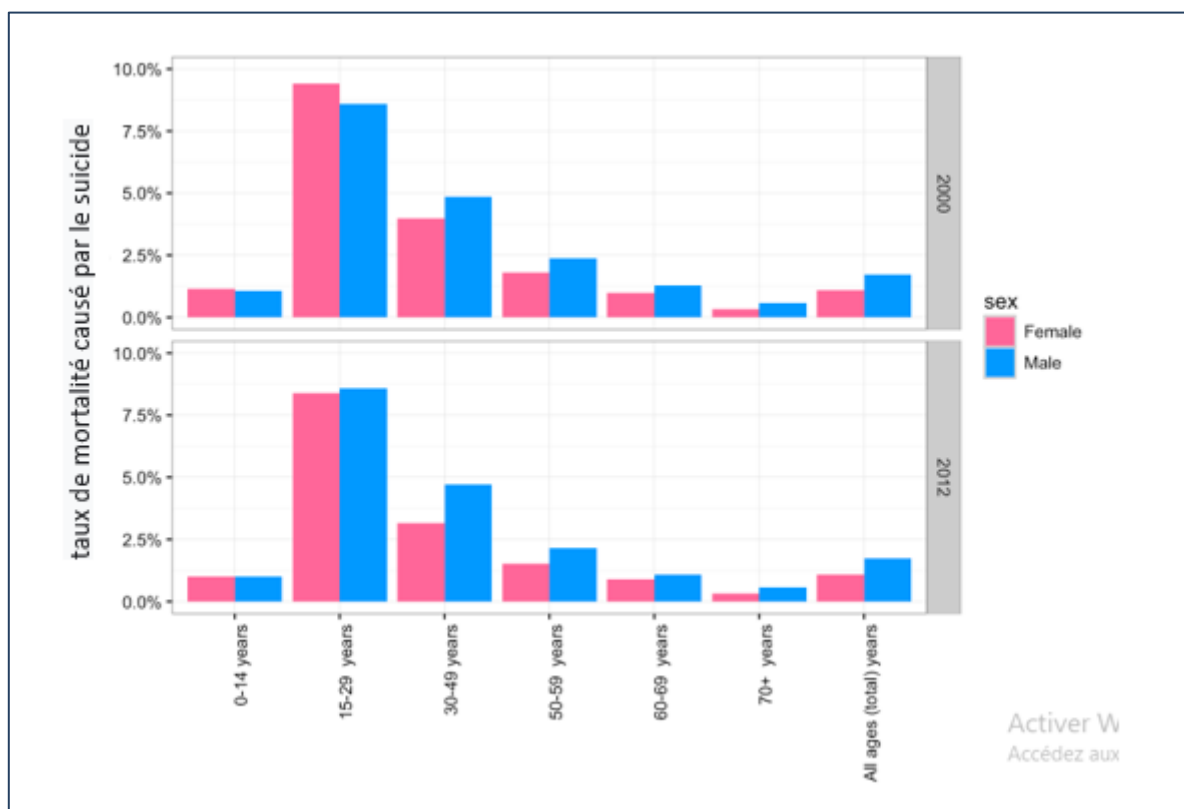


Figure 28: le Suicide par âge et par sexe

Le taux de suicide par rapport le taux de chômage :

Cette analyse se concentre sur l'impact du taux de chômage sur le taux de suicide. La tendance du chômage et du suicide est illustrée dans la figure suivante. Le taux de suicide montré ici est le nombre de personnes qui se suicident pour 100 000 personnes.

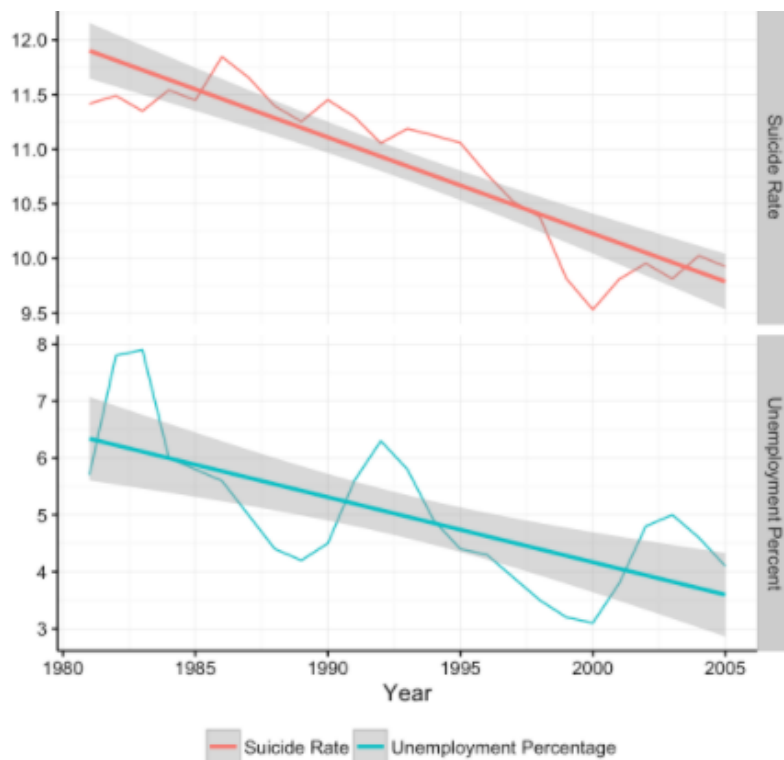


Figure 29 : Suicide et le chômage

Conclusion :

Le suicide reste donc un événement énigmatique : acte radical, destructeur, terminal, il n'en finit pas d'être interrogé et commenté, pour finalement renvoyer dos à dos et congédier toutes les justifications, toutes les explications : acte tantôt conquérant, tantôt désespéré, on le juge tour à tour suprêmement courageux et scandaleusement lâche, marque d'une volonté toute-puissante et anéantissement de cette même volonté, arrachement à l'ordre de la nature, ou reddition définitive ; est-il affirmation ultime et inouïe d'une liberté inconditionnelle, ou destin tragique de l'âme prisonnière de ses conflits ? En somme, il n'y aurait rien à dire du suicide, car on peut tout en dire.