

Causas de Morte ao redor do mundo

Análise Exploratória

Importando Base de dados e imprimindo as primeiras 5 linhas

```
In [2]: import pandas as pd

# Importando base de dados
df = pd.read_csv("./cause_of_deaths.csv")

# Primeiras 5 Linhas do Dataframe
df.head()
```

```
Out[2]:
```

	Country/Territory	Code	Year	Meningitis	Alzheimer's Disease and Other Dementias	Parkinson's Disease	Nutritional Deficiencies	Malaria	Drowning	Interpersonal Violence	...	Diabetes Mellitus	Chronic Kidney Disease	Poisonings	Protein-Energy Malnutrition	Road Injuries	Chronic Respiratory Diseases
0	Afghanistan	AFG	1990	2159	1116	371	2087	93	1370	1538	...	2108	3709	338	2054	4154	5945
1	Afghanistan	AFG	1991	2218	1136	374	2153	189	1391	2001	...	2120	3724	351	2119	4472	6050
2	Afghanistan	AFG	1992	2475	1162	378	2441	239	1514	2299	...	2153	3776	386	2404	5106	6223
3	Afghanistan	AFG	1993	2812	1187	384	2837	108	1687	2589	...	2195	3862	425	2797	5681	6445
4	Afghanistan	AFG	1994	3027	1211	391	3081	211	1809	2849	...	2231	3932	451	3038	6001	6664

5 rows x 34 columns

Quantas entradas temos? de quantos anos? de quantos países? e de quantas doenças?

```
In [3]: df.shape
```

```
Out[3]: (6120, 34)
```

```
In [4]: df['Year'].nunique()
```

```
Out[4]: 30
```

```
In [5]: df['Country/Territory'].nunique()
```

```
Out[5]: 204
```

```
In [6]: dfCauses = df.drop(columns=["Country/Territory", "Year"]).reset_index(drop=True)
dfCauses = dfCauses.melt(id_vars=["Code"], var_name="Cause", value_name="Deaths")
dfCauses["Cause"].nunique()
```

```
Out[6]: 31
```

Nota-se que há 6120 entradas, recorrentes a informações de 204 países ao longo de 30 anos a respeito de 31 causas de mortes distintas.

Abaixo estão listadas as causas de mortes que estamos trabalhando:

```
In [7]: dfCauses["Cause"].unique()
```

```
Out[7]: array(['Meningitis', 'Alzheimer's Disease and Other Dementias',
      'Parkinson's Disease', 'Nutritional Deficiencies', 'Malaria',
      'Drowning', 'Interpersonal Violence', 'Maternal Disorders',
      'HIV/AIDS', 'Drug Use Disorders', 'Tuberculosis',
      'Cardiovascular Diseases', 'Lower Respiratory Infections',
      'Neonatal Disorders', 'Alcohol Use Disorders', 'Self-harm',
      'Exposure to Forces of Nature', 'Diarrheal Diseases',
      'Environmental Heat and Cold Exposure', 'Neoplasms',
      'Conflict and Terrorism', 'Diabetes Mellitus',
      'Chronic Kidney Disease', 'Poisonings',
      'Protein-Energy Malnutrition', 'Road Injuries',
      'Chronic Respiratory Diseases',
      'Cirrhosis and Other Chronic Liver Diseases', 'Digestive Diseases',
      'Fire, Heat, and Hot Substances', 'Acute Hepatitis'], dtype=object)
```

Abaixo está disposta a listagem dos anos. Nota-se que trabalharemos no período de tempo 1990 - 2019.

```
In [8]: df['Year'].unique()
```

```
Out[8]: array([1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000,
      2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011,
      2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019], dtype=int64)
```

Há dados de todos os países para cada um dos anos?

```
In [9]: # Número de entradas (uma por ano) para cada país
df['Country/Territory'].value_counts()
```

```
Out[9]:
```

Afghanistan	30
Papua New Guinea	30
Niue	30
North Korea	30
North Macedonia	30
...	...
Greenland	30
Grenada	30
Guam	30
Guatemala	30
Zimbabwe	30

Name: Country/Territory, Length: 204, dtype: int64

```
In [10]: # Verificar se há dados faltantes
df.isnull().sum()
```

```
Out[10]: Country/Territory      0
Code                          0
Year                          0
Meningitis                    0
Alzheimer's Disease and Other Dementias  0
Parkinson's Disease           0
Nutritional Deficiencies       0
Malaria                        0
Drowning                       0
Interpersonal Violence         0
Maternal Disorders             0
HIV/AIDS                       0
Drug Use Disorders             0
Tuberculosis                   0
Cardiovascular Diseases        0
Lower Respiratory Infections   0
Neonatal Disorders             0
Alcohol Use Disorders          0
Self-harm                      0
Exposure to Forces of Nature   0
Diarrheal Diseases            0
Environmental Heat and Cold Exposure  0
Neoplasms                      0
Conflict and Terrorism         0
Diabetes Mellitus              0
Chronic Kidney Disease         0
Poisonings                     0
Protein-Energy Malnutrition    0
Road Injuries                  0
Chronic Respiratory Diseases   0
Cirrhosis and Other Chronic Liver Diseases  0
Digestive Diseases             0
Fire, Heat, and Hot Substances 0
Acute Hepatitis                0
dtype: int64
```

Observa-se que nossa base de dados está bem completa, sem dados faltantes.

Agora, fica a dúvida, qual a principal causa de morte do planeta?

```
In [11]: # Novo dataset com a frequência de cada causa de morte no mundo
dffrequency = df.groupby("Year").sum(numeric_only=True)
dffrequency
```

Out[11]:

	Meningitis	Alzheimer's Disease and Other Dementias	Parkinson's Disease	Nutritional Deficiencies	Malaria	Drowning	Interpersonal Violence	Maternal Disorders	HIV/AIDS	Drug Use Disorders	...	Diabetes Mellitus	Chronic Kidney Disease	Poisonings	Protein- Energy Malnutrition	Road Injuries	(Resp D
Year																	
1990	432253	560616	147156	756808	840297	460460	372497	302419	336059	56133	...	661085	600925	87951	655975	1112770	3
1991	428621	583166	150875	729145	858984	454375	383689	298271	430725	61890	...	679630	613589	87813	631013	1117024	3
1992	426440	605894	154886	700664	856415	447056	407176	299300	540070	66826	...	702253	630160	88435	606015	1125566	3
1993	420836	629571	160249	674219	862216	445434	432858	293564	664463	71603	...	728077	647255	90036	583919	1137444	3
1994	413799	652176	164381	649801	855671	443350	441971	293148	800169	76717	...	751254	665365	90897	564046	1153642	3
1995	409826	674815	168882	723095	862626	437303	444246	290551	938440	79985	...	773490	683701	90353	641084	1162799	3
1996	417259	696665	173822	671977	872476	423296	432673	287746	1061580	81704	...	799023	704624	88861	593826	1162809	3
1997	400893	717342	179347	647682	892946	413405	427316	289467	1174154	82572	...	827734	731048	87371	572372	1169798	3
1998	393364	738768	185097	620498	901338	407205	431984	289304	1303651	85087	...	854415	758681	86679	549543	1177827	3
1999	390136	761620	191538	593417	893788	397169	440642	288836	1439777	87336	...	879521	785380	87333	527431	1195303	3
2000	385994	786615	198545	568454	882055	387627	450393	287027	1559389	89690	...	905965	814854	88799	505874	1212355	3
2001	379912	814526	206194	542565	912709	375917	451206	282455	1662064	87325	...	934841	842790	89294	483074	1224220	3
2002	372298	845695	214930	517754	929195	361967	462331	276253	1746104	84750	...	970544	874109	89920	462083	1237109	3
2003	369171	877011	222256	417264	959654	347745	446701	269425	1808445	82250	...	1000276	902200	90634	365742	1248216	3
2004	363462	909148	227695	393952	960784	338417	440855	265279	1842954	81843	...	1018138	922921	91455	345527	1258547	3
2005	361809	945619	236498	376740	940540	332705	435946	264407	1832074	84990	...	1050925	956457	92064	329849	1269243	3
2006	358973	982308	242749	362476	941746	319875	427355	259261	1767868	85421	...	1071057	981743	89859	316210	1267492	3
2007	352747	1022057	250475	347407	940193	311889	420025	252551	1669306	85962	...	1092777	1010311	88475	302088	1273318	3
2008	343452	1065297	260382	335623	932910	304870	422800	247899	1559499	87077	...	1122287	1045743	88687	290690	1284333	3
2009	336868	1109405	267738	322879	917796	296871	422750	244175	1453258	87025	...	1142391	1074048	87902	279325	1281669	3
2010	326778	1155944	276188	322187	909814	290831	417449	239749	1364619	88079	...	1165663	1107473	87632	279144	1278839	3
2011	312733	1201138	284488	314327	868564	281106	413880	234669	1283719	89322	...	1197995	1143219	86099	271516	1267737	3
2012	300091	1247515	293123	298117	806183	275585	417694	227177	1198698	90934	...	1236606	1176295	84595	255541	1255450	3
2013	289502	1294701	303488	291451	751776	266827	415620	223473	1126196	94562	...	1277414	1214187	83618	249435	1232221	3
2014	280438	1343756	312928	284578	724324	258489	413371	214881	1072970	99295	...	1320608	1247378	81976	242956	1213989	3
2015	267789	1394942	321782	278793	702890	254237	416362	208450	1029667	105401	...	1366100	1286935	80974	237723	1201418	3
2016	261246	1451840	330736	273191	660364	248811	412500	203737	996098	112875	...	1412896	1325004	79951	232544	1193556	3
2017	250132	1509646	339343	267463	629653	241842	417143	199197	947006	118324	...	1454681	1351349	78484	227038	1188281	3
2018	241666	1568617	351322	258094	631568	240266	419249	198069	892341	123095	...	1501633	1387668	77805	218222	1195919	3
2019	236084	1622426	362702	251411	643201	237069	414157	196306	863056	128048	...	1549593	1426280	77130	212080	1197575	3

30 rows × 31 columns

```
In [12]: # Causa de morte mais frequente no planeta em cada ano
dfMax = dffrequency.idxmax(axis=1)
dfMax
```

```
Out[12]: Year
1990 Cardiovascular Diseases
1991 Cardiovascular Diseases
1992 Cardiovascular Diseases
1993 Cardiovascular Diseases
1994 Cardiovascular Diseases
1995 Cardiovascular Diseases
1996 Cardiovascular Diseases
1997 Cardiovascular Diseases
1998 Cardiovascular Diseases
1999 Cardiovascular Diseases
2000 Cardiovascular Diseases
2001 Cardiovascular Diseases
2002 Cardiovascular Diseases
2003 Cardiovascular Diseases
2004 Cardiovascular Diseases
2005 Cardiovascular Diseases
2006 Cardiovascular Diseases
2007 Cardiovascular Diseases
2008 Cardiovascular Diseases
2009 Cardiovascular Diseases
2010 Cardiovascular Diseases
2011 Cardiovascular Diseases
2012 Cardiovascular Diseases
2013 Cardiovascular Diseases
2014 Cardiovascular Diseases
2015 Cardiovascular Diseases
2016 Cardiovascular Diseases
2017 Cardiovascular Diseases
2018 Cardiovascular Diseases
2019 Cardiovascular Diseases
dtype: object
```

```
In [13]: mode = dfMax.mode()
cases = dfFrequency["Cardiovascular Diseases"].sum(axis=0)
print("A maior causa de morte no planeta de 1990 a 2019:", mode[0], " - tirou a vida de", cases, "indivíduos.")
```

A maior causa de morte no planeta de 1990 a 2019: Cardiovascular Diseases - tirou a vida de 447741982 indivíduos.

```
In [14]: # Outra análise interessante: Qual a causa de morte mais frequente em cada país?
# Para isso, vamos criar um novo dataset com a frequência de cada causa de morte em cada país, em todos os anos

dfCountry = df.drop(columns=["Year"]).reset_index(drop=True)
dfCountry = dfCountry.groupby("Country/Territory").sum(numeric_only=True)
dfCountry
```

Out[14]:

	Meningitis	Alzheimer's Disease and Other Dementias	Parkinson's Disease	Nutritional Deficiencies	Malaria	Drowning	Interpersonal Violence	Maternal Disorders	HIV/AIDS	Drug Use Disorders	...	Diabetes Mellitus	Chronic Kidney Disease	Poisonings	Protein-Energy Malnutrition	In
Country/Territory																
Afghanistan	78666	41998	13397	71453	13924	56536	108228	129621	4282	7094	...	93207	134676	14530	70163	21
Albania	1323	16549	4491	569	0	2397	5242	246	57	634	...	4055	7636	500	526	
Algeria	15685	86914	22943	7138	70	24273	16702	29475	6101	10612	...	89035	154666	12337	6407	31
American Samoa	30	143	69	60	0	120	101	30	15	0	...	970	512	0	60	
Andorra	0	614	137	0	0	0	15	0	85	0	...	198	292	0	0	
...
Venezuela	11615	108735	18573	22554	3726	20273	266071	12739	46090	1664	...	175790	161667	2607	21347	17
Vietnam	38559	369363	83322	48613	17311	214356	47981	13167	148838	19959	...	544222	396874	34681	7366	51
Yemen	21095	31045	7188	68939	143463	27994	17918	53611	6276	3718	...	30812	52119	12561	66731	27
Zambia	98886	13473	4054	95913	205529	12809	30065	28395	1175563	933	...	54098	41751	9056	92915	1
Zimbabwe	41238	20017	5764	66723	118728	18169	32741	29802	1836042	2271	...	71175	49952	9113	65942	1

204 rows × 31 columns

```
In [15]: # Qual será o país que mais teve mortes por cada causa de morte?
dfLeaders = dfCountry.idxmax(axis=0)
dfLeaders
```

```
Out[15]: Meningitis India
Alzheimer's Disease and Other Dementias China
Parkinson's Disease China
Nutritional Deficiencies India
Malaria Nigeria
Drowning China
Interpersonal Violence Brazil
Maternal Disorders India
HIV/AIDS South Africa
Drug Use Disorders United States
Tuberculosis India
Cardiovascular Diseases China
Lower Respiratory Infections India
Neonatal Disorders India
Alcohol Use Disorders Russia
Self-harm India
Exposure to Forces of Nature Haiti
Diarrheal Diseases India
Environmental Heat and Cold Exposure Russia
Neoplasms China
Conflict and Terrorism Rwanda
Diabetes Mellitus India
Chronic Kidney Disease India
Poisonings China
Protein-Energy Malnutrition India
Road Injuries China
Chronic Respiratory Diseases China
Cirrhosis and Other Chronic Liver Diseases India
Digestive Diseases India
Fire, Heat, and Hot Substances India
Acute Hepatitis India
dtype: object
```

```
In [16]: dfCountry.idxmax(axis=0).mode()[0]
```

Out[16]: 'India'

Dado a população numerosa da Índia, faz sentido que ela lidere em número de óbitos para a maior parte das causas. Será que o Brasil lidera em alguma causa? Vejamos:

```
In [17]: # Em quais causas de morte o Brasil aparece no dataset dfLeaders?
dfLeaders[dfLeaders == "Brazil"]
```

```
Out[17]: Interpersonal Violence Brazil
dtype: object
```

O Brasil é o país onde houve mais mortes por Violência Interpessoal nos anos de 1990 a 2019.

Vamos agora construir algumas visualizações.

Visualização

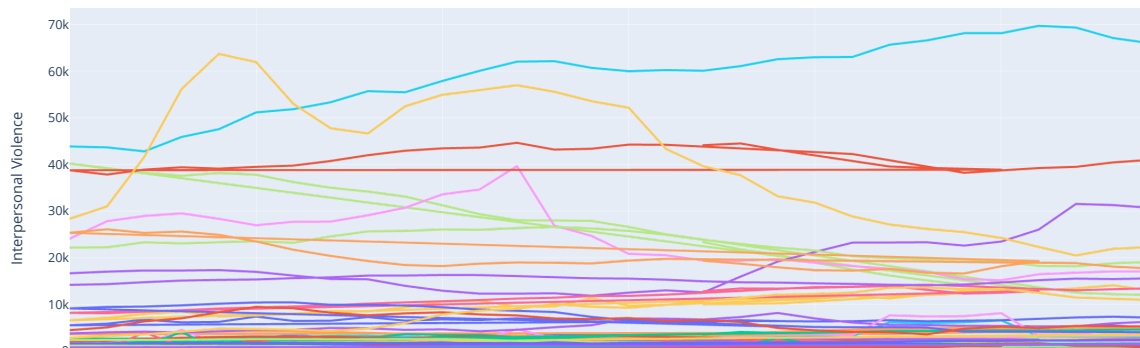
Vamos começar analisando a causa de morte que o Brasil Lidera: a Interpesonal Violence.

```
In [18]: import plotly.express as px

# Gráfico de Linhas para mostrar a evolução do número de mortes por INTERPERSONAL VIOLENCE no Brasil

fig = px.line(df, x="Year", y="Interpersonal Violence", color="Country/Territory", title="Interpersonal Violence")
fig.show()
```

Interpersonal Violence



Faremos, agora, um gráfico de barras que mostra o número de mortes por cada causa no Brasil em 2019 - a data mais recente.

```
In [21]: # Gráfico de barras com o número de mortes em cada país por cada causa em 2019

df2019 = df[df["Year"] == 2019]
df2019 = df2019.loc[df["Country/Territory"] == "Brazil"].drop(columns=["Country/Territory", "Code", "Year"]).reset_index(drop=True)

# Agora usamos melt para transformar as colunas em linhas
df2019 = df2019.melt(var_name="Cause", value_name="Deaths")
bar = px.bar(df2019,
             x="Deaths",
             y="Cause",
             color="Cause",
             color_discrete_sequence=["yellow", "blue", "green"],
)
# Inserir título
bar.update_layout(title="Deaths in Brazil in 2019")

# Não mostrar a Legenda
bar.update_layout(showlegend=False)

bar
```

Deaths in Brazil in 2019



E no mundo? Como se distribuíram as causas em 2019?

```
In [20]: # Gráfico de barras com o número total de mortes no planeta por cada causa de morte em 2019

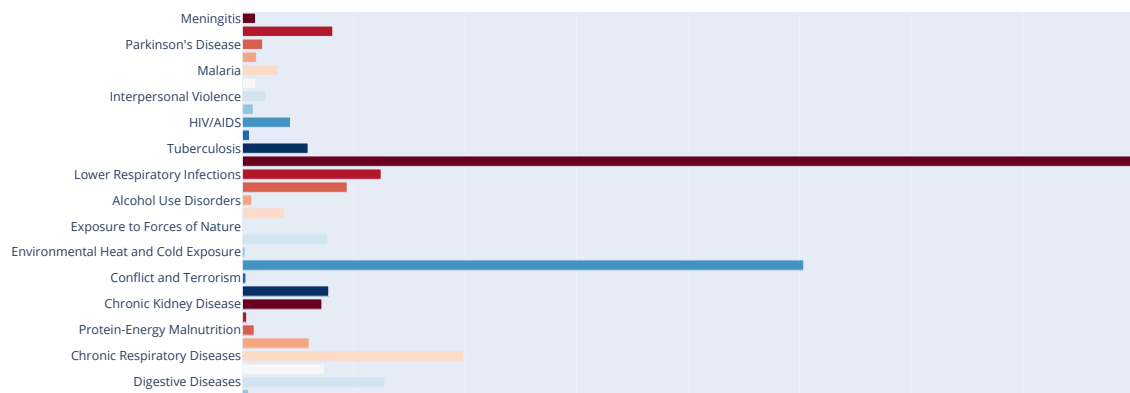
bar2 = px.bar(dfFrequency.loc[2019],
             x=dfFrequency.loc[2019].values,
             y=dfFrequency.loc[2019].index,
             color=dfFrequency.loc[2019].index,
             color_discrete_sequence=px.colors.sequential.RdBu,
)
# Inserir título
bar2.update_layout(title="Deaths in the world in 2019")

# Não mostrar a Legenda
bar2.update_layout(showlegend=False)

# Não mostrar títulos dos eixos
bar2.update_xaxes(title_text="")
bar2.update_yaxes(title_text="")

bar2
```

Deaths in the world in 2019



Essas foram algumas das centenas de visualizações interessantes que essa base de dados nos permite fazer.