

TASK-1 : Create a bar chart or histogram to visualize the distribution of a categorical or continuous variable, such as the distribution of ages or genders in a population.

```
[ ]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[ ]: from google.colab import files
raw = files.upload()
```

<IPython.core.display.HTML object>

Saving WorldBank Data.csv to WorldBank Data.csv

```
[ ]: df = pd.read_csv('WorldBank Data.csv')
```

```
[ ]: df_copy = df.copy()
```

```
[ ]: df.shape
```

```
[ ]: (266, 67)
```

```
[ ]: df.columns
```

```
[ ]: Index(['Country Name', 'Country Code', 'Indicator Name', 'Indicator Code',
          '1960', '1961', '1962', '1963', '1964', '1965', '1966', '1967', '1968',
          '1969', '1970', '1971', '1972', '1973', '1974', '1975', '1976', '1977',
          '1978', '1979', '1980', '1981', '1982', '1983', '1984', '1985', '1986',
          '1987', '1988', '1989', '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', '2020', '2021', '2022'],
          dtype='object')
```

```
[ ]: df.head(10)
```

[ ]:	Country Name	Country Code	Indicator Name	Indicator Code	\
0	Aruba	ABW	Population, total	SP.POP.TOTL	
1	Africa Eastern and Southern	AFE	Population, total	SP.POP.TOTL	
2	Afghanistan	AFG	Population, total	SP.POP.TOTL	
3	Africa Western and Central	AFW	Population, total	SP.POP.TOTL	
4	Angola	AGO	Population, total	SP.POP.TOTL	
5	Albania	ALB	Population, total	SP.POP.TOTL	
6	Andorra	AND	Population, total	SP.POP.TOTL	
7	Arab World	ARB	Population, total	SP.POP.TOTL	
8	United Arab Emirates	ARE	Population, total	SP.POP.TOTL	
9	Argentina	ARG	Population, total	SP.POP.TOTL	

	1960	1961	1962	1963	1964	\
0	54608.0	55811.0	56682.0	57475.0	58178.0	
1	130692579.0	134169237.0	137835590.0	141630546.0	145605995.0	
2	8622466.0	8790140.0	8969047.0	9157465.0	9355514.0	
3	97256290.0	99314028.0	101445032.0	103667517.0	105959979.0	
4	5357195.0	5441333.0	5521400.0	5599827.0	5673199.0	
5	1608800.0	1659800.0	1711319.0	1762621.0	1814135.0	
6	9443.0	10216.0	11014.0	11839.0	12690.0	
7	93359407.0	95760348.0	98268683.0	100892507.0	103618568.0	
8	133426.0	140984.0	148877.0	157006.0	165305.0	
9	20349744.0	20680653.0	21020359.0	21364017.0	21708487.0	

	1965	...	2013	2014	2015	2016	\
0	58782.0	...	102880.0	103594.0	104257.0	104874.0	
1	149742351.0	...	567892149.0	583651101.0	600008424.0	616377605.0	
2	9565147.0	...	31541209.0	32716210.0	33753499.0	34636207.0	
3	108336203.0	...	387204553.0	397855507.0	408690375.0	419778384.0	
4	5736582.0	...	26147002.0	27128337.0	28127721.0	29154746.0	
5	1864791.0	...	2895092.0	2889104.0	2880703.0	2876101.0	
6	13563.0	...	71367.0	71621.0	71746.0	72540.0	
7	106444103.0	...	389131555.0	397922915.0	406501999.0	415077960.0	
8	173797.0	...	8751847.0	8835951.0	8916899.0	8994263.0	
9	22053661.0	...	42202935.0	42669500.0	43131966.0	43590368.0	

	2017	2018	2019	2020	2021	\
0	105439.0	105962.0	106442.0	106585.0	106537.0	
1	632746570.0	649757148.0	667242986.0	685112979.0	702977106.0	
2	35643418.0	36686784.0	37769499.0	38972230.0	40099462.0	
3	431138704.0	442646825.0	454306063.0	466189102.0	478185907.0	
4	30208628.0	31273533.0	32353588.0	33428486.0	34503774.0	
5	2873457.0	2866376.0	2854191.0	2837849.0	2811666.0	
6	73837.0	75013.0	76343.0	77700.0	79034.0	
7	423664839.0	432545676.0	441467739.0	449228296.0	456520777.0	
8	9068296.0	9140169.0	9211657.0	9287289.0	9365145.0	
9	44044811.0	44494502.0	44938712.0	45376763.0	45808747.0	

```

2022
0    106445.0
1   720839314.0
2    41128771.0
3   490330870.0
4    35588987.0
5    2775634.0
6     79824.0
7   464684914.0
8    9441129.0
9   46234830.0

```

[10 rows x 67 columns]

```
[ ]: df.tail()
```

```
[ ]:
Country Name Country Code Indicator Name Indicator Code 1960 \
261 Kosovo XKX Population, total SP. POP. TOTL 947000.0
262 Yemen, Rep. YEM Population, total SP. POP. TOTL 5542459.0
263 South Africa ZAF Population, total SP. POP. TOTL 16520441.0
264 Zambia ZMB Population, total SP. POP. TOTL 3119430.0
265 Zimbabwe ZWE Population, total SP. POP. TOTL 3806310.0

```

```

1961 1962 1963 1964 1965 ... \
261 966000.0 994000.0 1022000.0 1050000.0 1078000.0 ...
262 5646668.0 5753386.0 5860197.0 5973803.0 6097298.0 ...
263 16989464.0 17503133.0 18042215.0 18603097.0 19187194.0 ...
264 3219451.0 3323427.0 3431381.0 3542764.0 3658024.0 ...
265 3925952.0 4049778.0 4177931.0 4310332.0 4447149.0 ...

```

```

2013 2014 2015 2016 2017 2018 \
261 1818117.0 1812771.0 1788196.0 1777557.0 1791003.0 1797085.0
262 26984002.0 27753304.0 28516545.0 29274002.0 30034389.0 30790513.0
263 53873616.0 54729551.0 55876504.0 56422274.0 56641209.0 57339635.0
264 15234976.0 15737793.0 16248230.0 16767761.0 17298054.0 17835893.0
265 13555422.0 13855753.0 14154937.0 14452704.0 14751101.0 15052184.0

```

```

2019 2020 2021 2022
261 1788878.0 1790133.0 1786038.0 1761985.0
262 31546691.0 32284046.0 32981641.0 33696614.0
263 58087055.0 58801927.0 59392255.0 59893885.0
264 18380477.0 18927715.0 19473125.0 20017675.0
265 15354608.0 15669666.0 15993524.0 16320537.0

```

[5 rows x 67 columns]

```
[ ]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 266 entries, 0 to 265
```

```
Data columns (total 67 columns):
```

#	Column	Non-Null Count	Dtype
0	Country Name	266 non-null	object
1	Country Code	266 non-null	object
2	Indicator Name	266 non-null	object
3	Indicator Code	266 non-null	object
4	1960	264 non-null	float64
5	1961	264 non-null	float64
6	1962	264 non-null	float64
7	1963	264 non-null	float64
8	1964	264 non-null	float64
9	1965	264 non-null	float64
10	1966	264 non-null	float64
11	1967	264 non-null	float64
12	1968	264 non-null	float64
13	1969	264 non-null	float64
14	1970	264 non-null	float64
15	1971	264 non-null	float64
16	1972	264 non-null	float64
17	1973	264 non-null	float64
18	1974	264 non-null	float64
19	1975	264 non-null	float64
20	1976	264 non-null	float64
21	1977	264 non-null	float64
22	1978	264 non-null	float64
23	1979	264 non-null	float64
24	1980	264 non-null	float64
25	1981	264 non-null	float64
26	1982	264 non-null	float64
27	1983	264 non-null	float64
28	1984	264 non-null	float64
29	1985	264 non-null	float64
30	1986	264 non-null	float64
31	1987	264 non-null	float64
32	1988	264 non-null	float64
33	1989	264 non-null	float64
34	1990	265 non-null	float64
35	1991	265 non-null	float64
36	1992	265 non-null	float64
37	1993	265 non-null	float64
38	1994	265 non-null	float64
39	1995	265 non-null	float64

```

40 1996          265 non-null    float64
41 1997          265 non-null    float64
42 1998          265 non-null    float64
43 1999          265 non-null    float64
44 2000          265 non-null    float64
45 2001          265 non-null    float64
46 2002          265 non-null    float64
47 2003          265 non-null    float64
48 2004          265 non-null    float64
49 2005          265 non-null    float64
50 2006          265 non-null    float64
51 2007          265 non-null    float64
52 2008          265 non-null    float64
53 2009          265 non-null    float64
54 2010          265 non-null    float64
55 2011          265 non-null    float64
56 2012          265 non-null    float64
57 2013          265 non-null    float64
58 2014          265 non-null    float64
59 2015          265 non-null    float64
60 2016          265 non-null    float64
61 2017          265 non-null    float64
62 2018          265 non-null    float64
63 2019          265 non-null    float64
64 2020          265 non-null    float64
65 2021          265 non-null    float64
66 2022          265 non-null    float64

```

dtypes: float64(63), object(4)

memory usage: 139.4+ KB

```
[ ]: df.describe()
```

```

[ ]:
count    1960    1961    1962    1963    1964  \
mean    2.640000e+02  2.640000e+02  2.640000e+02  2.640000e+02  2.640000e+02
std     3.695439e+08  3.740897e+08  3.808061e+08  3.895039e+08  3.982439e+08
min     2.646000e+03  2.888000e+03  3.171000e+03  3.481000e+03  3.811000e+03
25%     5.132212e+05  5.231345e+05  5.337595e+05  5.449288e+05  5.566630e+05
50%     3.757486e+06  3.887144e+06  4.023896e+06  4.139356e+06  4.224612e+06
75%     2.670606e+07  2.748694e+07  2.830289e+07  2.914708e+07  3.001684e+07
max     3.031474e+09  3.072422e+09  3.126850e+09  3.193429e+09  3.260442e+09

count    1965    1966    1967    1968    1969  \
mean    1.291813e+08  1.320404e+08  1.348980e+08  1.378358e+08  1.408789e+08
std     4.071153e+08  4.164504e+08  4.257424e+08  4.353218e+08  4.452927e+08
min     4.161000e+03  4.531000e+03  4.930000e+03  5.354000e+03  5.646000e+03

```

25%	5.651150e+05	5.691470e+05	5.773872e+05	5.832700e+05	5.875942e+05
50%	4.277636e+06	4.331825e+06	4.385700e+06	4.450934e+06	4.530800e+06
75%	3.084892e+07	3.163010e+07	3.209247e+07	3.249927e+07	3.277149e+07
max	3.328209e+09	3.398480e+09	3.468371e+09	3.540164e+09	3.614573e+09

	...	2013	2014	2015	2016	\
count	...	2.650000e+02	2.650000e+02	2.650000e+02	2.650000e+02	
mean	...	2.927787e+08	2.966774e+08	3.005462e+08	3.044051e+08	
std	...	9.186849e+08	9.301446e+08	9.414558e+08	9.526720e+08	
min	...	1.069400e+04	1.089900e+04	1.087700e+04	1.085200e+04	
25%	...	1.697753e+06	1.743309e+06	1.788196e+06	1.777557e+06	
50%	...	1.014958e+07	1.028212e+07	1.035808e+07	1.032545e+07	
75%	...	6.023395e+07	6.078914e+07	6.073058e+07	6.062750e+07	
max	...	7.229732e+09	7.317970e+09	7.405278e+09	7.492157e+09	

		2017	2018	2019	2020	2021	\
count	2.650000e+02	2.650000e+02	2.650000e+02	2.650000e+02	2.650000e+02	2.650000e+02	
mean	3.082575e+08	3.120276e+08	3.157110e+08	3.192936e+08	3.225180e+08	3.225180e+08	
std	9.638572e+08	9.746880e+08	9.851690e+08	9.952294e+08	1.004211e+09	1.004211e+09	
min	1.082800e+04	1.086500e+04	1.095600e+04	1.106900e+04	1.120400e+04	1.120400e+04	
25%	1.791003e+06	1.797085e+06	1.788878e+06	1.790133e+06	1.786038e+06	1.786038e+06	
50%	1.030030e+07	1.039533e+07	1.044767e+07	1.060623e+07	1.050577e+07	1.050577e+07	
75%	6.053671e+07	6.042176e+07	5.987258e+07	6.170452e+07	6.358833e+07	6.358833e+07	
max	7.578221e+09	7.661777e+09	7.742682e+09	7.820964e+09	7.888161e+09	7.888161e+09	

	2022
count	2.650000e+02
mean	3.254839e+08
std	1.012174e+09
min	1.131200e+04
25%	1.761985e+06
50%	1.052607e+07
75%	6.549775e+07
max	7.951150e+09

[8 rows x 63 columns]

```
[ ]: pd.options.display.max_rows = 300
df.isnull().sum()
```

```
[ ]: Country Name      0
Country Code         0
Indicator Name       0
Indicator Code       0
1960                 2
1961                 2
1962                 2
```

1963	2
1964	2
1965	2
1966	2
1967	2
1968	2
1969	2
1970	2
1971	2
1972	2
1973	2
1974	2
1975	2
1976	2
1977	2
1978	2
1979	2
1980	2
1981	2
1982	2
1983	2
1984	2
1985	2
1986	2
1987	2
1988	2
1989	2
1990	1
1991	1
1992	1
1993	1
1994	1
1995	1
1996	1
1997	1
1998	1
1999	1
2000	1
2001	1
2002	1
2003	1
2004	1
2005	1
2006	1
2007	1
2008	1
2009	1

2010	1
2011	1
2012	1
2013	1
2014	1
2015	1
2016	1
2017	1
2018	1
2019	1
2020	1
2021	1
2022	1

dtype: int64

```
[ ]: df = df.fillna(method = "ffill")
```

```
[ ]: pd.options.display.max_columns=300
df.isnull().sum()
```

```
[ ]: Country Name      0
Country Code         0
Indicator Name       0
Indicator Code       0
1960                 0
1961                 0
1962                 0
1963                 0
1964                 0
1965                 0
1966                 0
1967                 0
1968                 0
1969                 0
1970                 0
1971                 0
1972                 0
1973                 0
1974                 0
1975                 0
1976                 0
1977                 0
1978                 0
1979                 0
1980                 0
1981                 0
1982                 0
```



1983	0
1984	0
1985	0
1986	0
1987	0
1988	0
1989	0
1990	0
1991	0
1992	0
1993	0
1994	0
1995	0
1996	0
1997	0
1998	0
1999	0
2000	0
2001	0
2002	0
2003	0
2004	0
2005	0
2006	0
2007	0
2008	0
2009	0
2010	0
2011	0
2012	0
2013	0
2014	0
2015	0
2016	0
2017	0
2018	0
2019	0
2020	0
2021	0
2022	0
dtype: int64	

```
[ ]: df.dtypes
```

```
[ ]: Country Name      object
      Country Code     object
      Indicator Name    object
```

Indicator Code	object
1960	float64
1961	float64
1962	float64
1963	float64
1964	float64
1965	float64
1966	float64
1967	float64
1968	float64
1969	float64
1970	float64
1971	float64
1972	float64
1973	float64
1974	float64
1975	float64
1976	float64
1977	float64
1978	float64
1979	float64
1980	float64
1981	float64
1982	float64
1983	float64
1984	float64
1985	float64
1986	float64
1987	float64
1988	float64
1989	float64
1990	float64
1991	float64
1992	float64
1993	float64
1994	float64
1995	float64
1996	float64
1997	float64
1998	float64
1999	float64
2000	float64
2001	float64
2002	float64
2003	float64
2004	float64
2005	float64

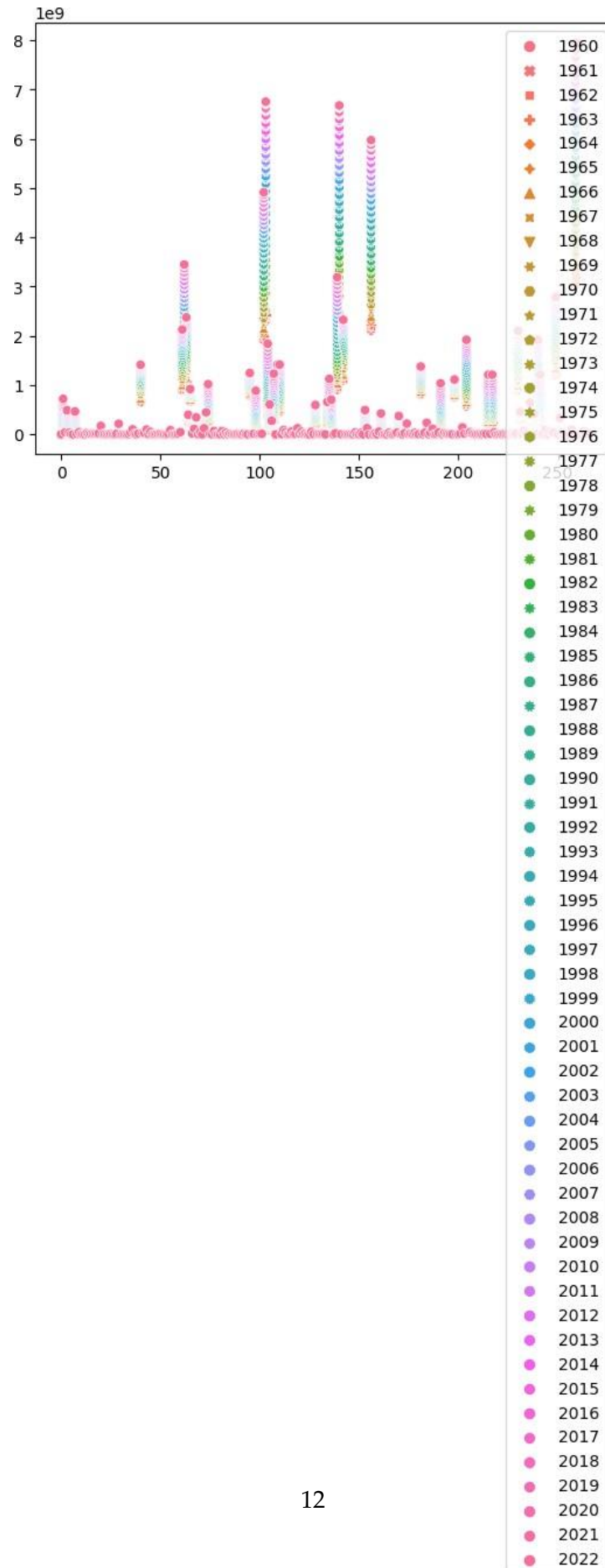
```
2006      float64
2007      float64
2008      float64
2009      float64
2010      float64
2011      float64
2012      float64
2013      float64
2014      float64
2015      float64
2016      float64
2017      float64
2018      float64
2019      float64
2020      float64
2021      float64
2022      float64
dtype: object
```

```
[ ]: df.duplicated().sum()
```

```
[ ]: 0
```

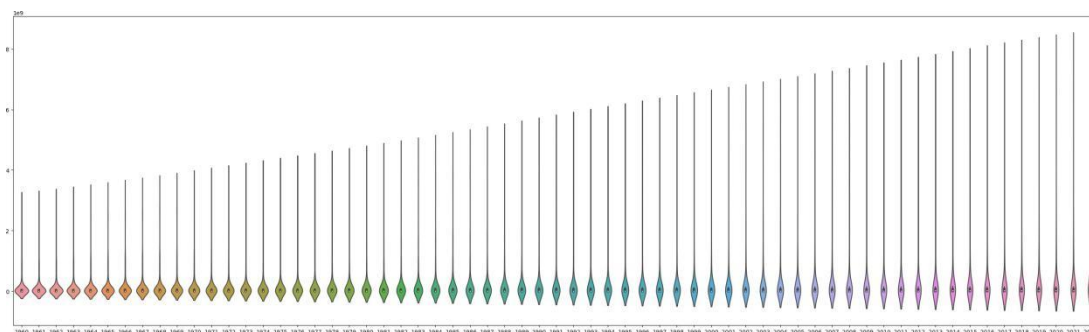
```
[ ]: sns.scatterplot(df)
```

```
[ ]: <Axes: >
```



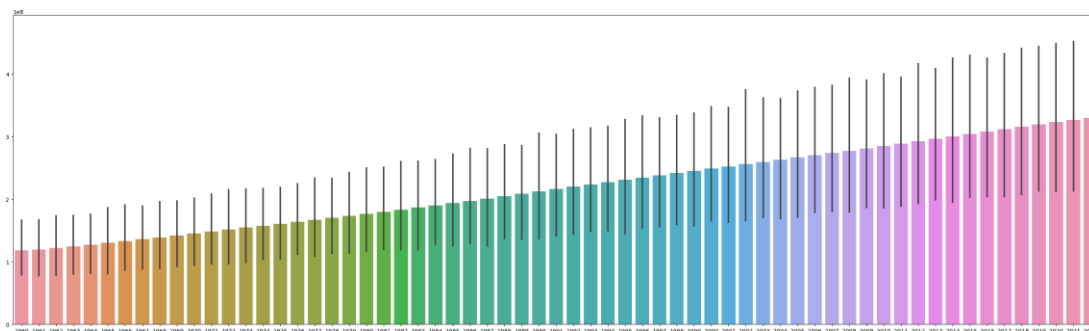
```
[ ]: dim = (35, 10)
fig, ax = plt.subplots(figsize=dim)
sns.violinplot(ax=ax, data=df)
```

```
[ ]: <Axes: >
```



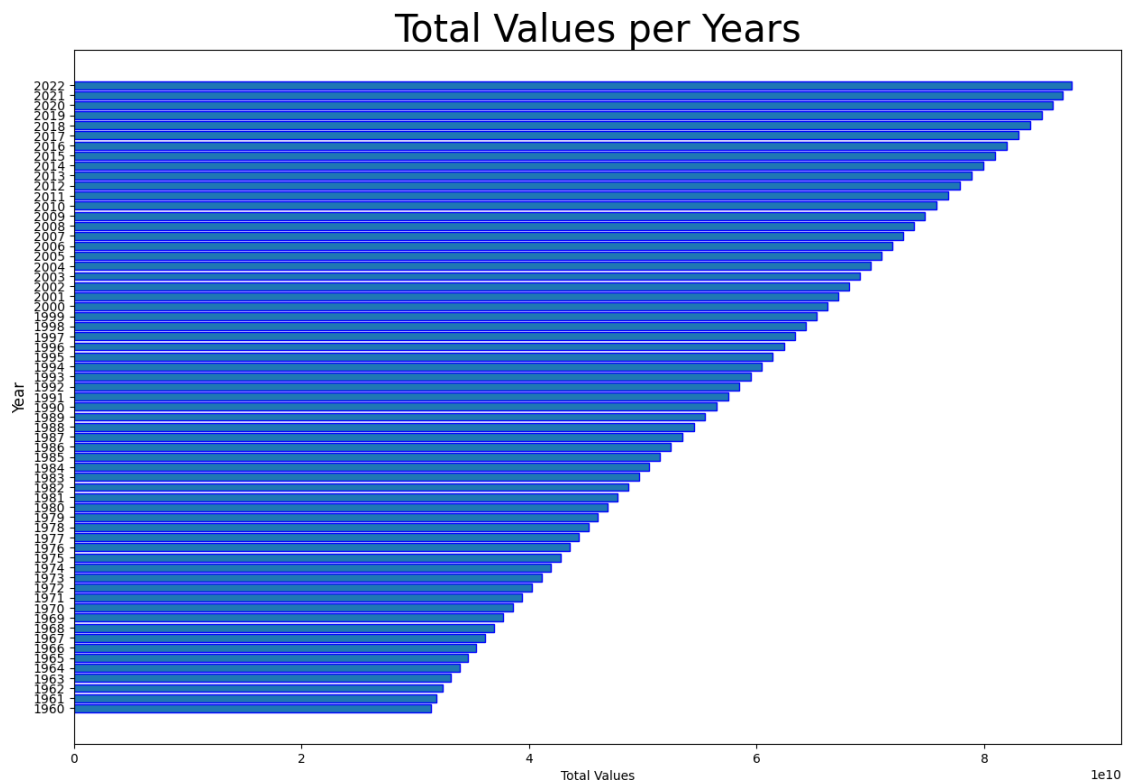
```
[ ]: dim2 = (35, 10)
fig, ax = plt.subplots(figsize=dim2)
sns.barplot(ax=ax, data=df)
```

```
[ ]: <Axes: >
```



```
[ ]: Years = ['1960', '1961', '1962', '1963', '1964', '1965', '1966', '1967', '1968',
              '1969', '1970', '1971', '1972', '1973', '1974', '1975', '1976', '1977',
              '1978', '1979', '1980', '1981', '1982', '1983', '1984', '1985', '1986',
              '1987', '1988', '1989', '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', '2020', '2021', '2022']
```

```
[ ]: Years = df.columns[4:]
values = df[Years].sum()
plt.figure(figsize=(15,10))
plt.barh(Years, values, edgecolor='blue')
plt.xlabel('Total Values')
plt.ylabel('Year', size=12)
plt.title('Total Values per Years', size=30)
plt.show()
```



```
[ ]: for i in Years:
    fig = plt.figure(figsize=(6,8))
    plt.hist(df[i], color='pink', bins=20)
    plt.xlabel(i)
    plt.show()
```

