

# qxdihko2c

December 12, 2023

## #IMPORTING LIBRARIES

pandas (import pandas as pd): For handling structured data with DataFrames.

numpy (import numpy as np): For numerical operations on arrays and matrices.

matplotlib (import matplotlib.pyplot as plt): For creating static visualizations.

seaborn (import seaborn as sb): For creating attractive statistical visualizations.

Setting display options (pd.set\_option): Adjusts options to display all columns and limit rows to 150 for better data exploration.

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

      #display all columns and rows of the dataframe
      pd.set_option('display.max_columns',None)
      pd.set_option('display.max_rows', 150)
```

## IMPORTING DATA SET

Upload a file from your local machine to your Colab environment

```
uploaded = files.upload()
```

Download a file from your Colab environment to your local machine

```
files.download('example.txt')
```

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

```
<IPython.core.display.HTML object>
Saving WorldBank Data.csv to WorldBank Data.csv
```

## REVIEWING THE DATASET

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

## SHALLOW COPYING

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

#EXPLORATORY DATA ANALYSIS

[ ]: df.shape #an attribute of a DataFrame that returns a tuple representing the
      ↪dimensions of the DataFrame.

[ ]: (266, 67)

[ ]: df.head() #show first few rows to learn the structure of the data

[ ]:          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

          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

          1965      1966      1967      1968      1969 \
0   58782.0   59291.0   59522.0   59471.0   59330.0
1  149742351.0  153955516.0  158313235.0  162875171.0  167596160.0
2   9565147.0   9783147.0   10010030.0   10247780.0   10494489.0
3  108336203.0  110798486.0  113319950.0  115921723.0  118615741.0
4   5736582.0   5787044.0   5827503.0   5868203.0   5928386.0

          1970      1971      1972      1973      1974 \
0   59106.0   58816.0   58855.0   59365.0   60028.0
1  172475766.0  177503186.0  182599092.0  187901657.0  193512956.0
2   10752971.0   11015857.0   11286753.0   11575305.0   11869879.0
3  121424797.0  124336039.0  127364044.0  130563107.0  133953892.0
4   6029700.0   6177049.0   6364731.0   6578230.0   6802494.0

          1975      1976      1977      1978      1979 \
0   60715.0   61193.0   61465.0   61738.0   62006.0
1  199284304.0  205202669.0  211120911.0  217481420.0  224315978.0
2   12157386.0   12425267.0   12687301.0   12938862.0   12986369.0
3  137548613.0  141258400.0  145122851.0  149206663.0  153459665.0
4   7032713.0   7266780.0   7511895.0   7771590.0   8043218.0

          1980      1981      1982      1983      1984 \

```

0	62267.0	62614.0	63116.0	63683.0	64174.0
1	230967858.0	237937461.0	245386717.0	252779730.0	260209149.0
2	12486631.0	11155195.0	10088289.0	9951449.0	10243686.0
3	157825609.0	162323313.0	167023385.0	171566640.0	176054495.0
4	8330047.0	8631457.0	8947152.0	9276707.0	9617702.0

	1985	1986	1987	1988	1989	\
0	64478.0	64553.0	64450.0	64332.0	64596.0	
1	267938123.0	276035920.0	284490394.0	292795186.0	301124880.0	
2	10512221.0	10448442.0	10322758.0	10383460.0	10673168.0	
3	180817312.0	185720244.0	190759952.0	195969722.0	201392200.0	
4	9970621.0	10332574.0	10694057.0	11060261.0	11439498.0	

	1990	1991	1992	1993	1994	\
0	65712.0	67864.0	70192.0	72360.0	74710.0	
1	309890664.0	318544083.0	326933522.0	335625136.0	344418362.0	
2	10694796.0	10745167.0	12057433.0	14003760.0	15455555.0	
3	206739024.0	212172888.0	217966101.0	223788766.0	229675775.0	
4	11828638.0	12228691.0	12632507.0	13038270.0	13462031.0	

	1995	1996	1997	1998	1999	\
0	77050.0	79417.0	81858.0	84355.0	86867.0	
1	353466601.0	362985802.0	372352230.0	381715600.0	391486231.0	
2	16418912.0	17106595.0	17788819.0	18493132.0	19262847.0	
3	235861484.0	242200260.0	248713095.0	255482918.0	262397030.0	
4	13912253.0	14383350.0	14871146.0	15366864.0	15870753.0	

	2000	2001	2002	2003	2004	\
0	89101.0	90691.0	91781.0	92701.0	93540.0	
1	401600588.0	412001885.0	422741118.0	433807484.0	445281555.0	
2	19542982.0	19688632.0	21000256.0	22645130.0	23553551.0	
3	269611898.0	277160097.0	284952322.0	292977949.0	301265247.0	
4	16394062.0	16941587.0	17516139.0	18124342.0	18771125.0	

	2005	2006	2007	2008	2009	\
0	94483.0	95606.0	96787.0	97996.0	99212.0	
1	457153837.0	469508516.0	482406426.0	495748900.0	509410477.0	
2	24411191.0	25442944.0	25903301.0	26427199.0	27385307.0	
3	309824829.0	318601484.0	327612838.0	336893835.0	346475221.0	
4	19450959.0	20162340.0	20909684.0	21691522.0	22507674.0	

	2010	2011	2012	2013	2014	\
0	100341.0	101288.0	102112.0	102880.0	103594.0	
1	523459657.0	537792950.0	552530654.0	567892149.0	583651101.0	
2	28189672.0	29249157.0	30466479.0	31541209.0	32716210.0	
3	356337762.0	366489204.0	376797999.0	387204553.0	397855507.0	
4	23364185.0	24259111.0	25188292.0	26147002.0	27128337.0	

	2015	2016	2017	2018	2019	\
0	104257.0	104874.0	105439.0	105962.0	106442.0	
1	600008424.0	616377605.0	632746570.0	649757148.0	667242986.0	
2	33753499.0	34636207.0	35643418.0	36686784.0	37769499.0	
3	408690375.0	419778384.0	431138704.0	442646825.0	454306063.0	
4	28127721.0	29154746.0	30208628.0	31273533.0	32353588.0	
	2020	2021	2022			
0	106585.0	106537.0	106445.0			
1	685112979.0	702977106.0	720839314.0			
2	38972230.0	40099462.0	41128771.0			
3	466189102.0	478185907.0	490330870.0			
4	33428486.0	34503774.0	35588987.0			

[ ]: df.tail()#show last few rows to learn the structure of the data

	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	1966	\
261	966000.0	994000.0	1022000.0	1050000.0	1078000.0	1106000.0	
262	5646668.0	5753386.0	5860197.0	5973803.0	6097298.0	6228430.0	
263	16989464.0	17503133.0	18042215.0	18603097.0	19187194.0	19789771.0	
264	3219451.0	3323427.0	3431381.0	3542764.0	3658024.0	3777680.0	
265	3925952.0	4049778.0	4177931.0	4310332.0	4447149.0	4588529.0	
	1967	1968	1969	1970	1971	1972	\
261	1135000.0	1163000.0	1191000.0	1219000.0	1247000.0	1278000.0	
262	6368014.0	6515904.0	6673981.0	6843607.0	7024196.0	7215835.0	
263	20410677.0	21050540.0	21704214.0	22368306.0	23031441.0	23698507.0	
264	3901288.0	4029173.0	4159007.0	4281671.0	4399919.0	4523581.0	
265	4734694.0	4886347.0	5044163.0	5202918.0	5363423.0	5532842.0	
	1973	1974	1975	1976	1977	1978	\
261	1308000.0	1339000.0	1369000.0	1400000.0	1430000.0	1460000.0	
262	7417736.0	7630190.0	7855657.0	8094985.0	8348182.0	8615301.0	
263	24382513.0	25077016.0	25777964.0	26480300.0	27199838.0	27943445.0	
264	4653289.0	4789038.0	4931249.0	5079672.0	5233292.0	5391355.0	
265	5712712.0	5903530.0	6097083.0	6288387.0	6453044.0	6549349.0	
	1979	1980	1981	1982	1983	1984	\
261	1491000.0	1521000.0	1552000.0	1582000.0	1614000.0	1647000.0	

262	8899922.0	9204938.0	9529105.0	9872292.0	10237391.0	10625687.0	
263	28697014.0	29463549.0	30232561.0	31022417.0	31865176.0	32768207.0	
264	5553462.0	5720438.0	5897481.0	6090818.0	6291070.0	6488072.0	
265	6655833.0	7049926.0	7506526.0	7803855.0	8106356.0	8398567.0	
	1985	1986	1987	1988	1989	1990	\
261	1682000.0	1717000.0	1753000.0	1791000.0	1827000.0	1862000.0	
262	11036918.0	11465444.0	11915563.0	12387238.0	12872362.0	13375121.0	
263	33752964.0	34877834.0	36119333.0	37393853.0	38668684.0	39877570.0	
264	6686449.0	6890967.0	7095185.0	7294325.0	7491275.0	7686401.0	
265	8690515.0	8983044.0	9277488.0	9568739.0	9846346.0	10113893.0	
	1991	1992	1993	1994	1995	1996	\
261	1898000.0	1932000.0	1965000.0	1997000.0	2029000.0	2059000.0	
262	13895851.0	14433771.0	14988047.0	15553171.0	16103339.0	16614326.0	
263	40910959.0	41760755.0	42525440.0	43267982.0	43986084.0	44661603.0	
264	7880466.0	8074337.0	8270917.0	8474216.0	8684135.0	8902019.0	
265	10377815.0	10641501.0	10794918.0	10858594.0	10994041.0	11178171.0	
	1997	1998	1999	2000	2001	2002	\
261	2086000.0	1966000.0	1762000.0	1700000.0	1701154.0	1702310.0	
262	17108681.0	17608133.0	18114552.0	18628700.0	19143457.0	19660653.0	
263	45285048.0	45852166.0	46364681.0	46813266.0	47229714.0	47661514.0	
264	9133156.0	9372430.0	9621238.0	9891136.0	10191964.0	10508294.0	
265	11362401.0	11548364.0	11716454.0	11834676.0	11910978.0	11984644.0	
	2003	2004	2005	2006	2007	2008	\
261	1703466.0	1704622.0	1705780.0	1719536.0	1733404.0	1747383.0	
262	20188799.0	20733406.0	21320671.0	21966298.0	22641538.0	23329004.0	
263	48104048.0	48556071.0	49017147.0	49491756.0	49996094.0	50565812.0	
264	10837973.0	11188040.0	11564870.0	11971567.0	12402073.0	12852966.0	
265	12075828.0	12160881.0	12224753.0	12330490.0	12450568.0	12550347.0	
	2009	2010	2011	2012	2013	2014	\
261	1761474.0	1775680.0	1791000.0	1807106.0	1818117.0	1812771.0	
262	24029589.0	24743946.0	25475610.0	26223391.0	26984002.0	27753304.0	
263	51170779.0	51784921.0	52443325.0	53145033.0	53873616.0	54729551.0	
264	13318087.0	13792086.0	14265814.0	14744658.0	15234976.0	15737793.0	
265	12679810.0	12839771.0	13025785.0	13265331.0	13555422.0	13855753.0	
	2015	2016	2017	2018	2019	2020	\
261	1788196.0	1777557.0	1791003.0	1797085.0	1788878.0	1790133.0	
262	28516545.0	29274002.0	30034389.0	30790513.0	31546691.0	32284046.0	
263	55876504.0	56422274.0	56641209.0	57339635.0	58087055.0	58801927.0	
264	16248230.0	16767761.0	17298054.0	17835893.0	18380477.0	18927715.0	
265	14154937.0	14452704.0	14751101.0	15052184.0	15354608.0	15669666.0	

```
      2021      2022
261  1786038.0  1761985.0
262  32981641.0 33696614.0
263  59392255.0 59893885.0
264  19473125.0 20017675.0
265  15993524.0 16320537.0
```

```
[ ]: 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.info() # Display concise summary of the dataset
```

```
<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
```

```
[ ]: # Display summary for categorical data
df.describe(include='object').T
```

```
[ ]:          count unique           top freq
Country Name    266    266             Aruba     1
Country Code    266    266             ABW      1
Indicator Name   266      1 Population, total  266
Indicator Code   266      1 SP.POP.TOTL  266
```

```
[ ]: data=df.drop(['Country Code','Indicator Name'],axis=1)
```

```
[ ]: # Display summary for numerical data
df.describe().T
```

```
[ ]:          count       mean        std       min      25%      50% \
1960  264.0  1.172712e+08  3.695439e+08  2646.0  513221.25  3757485.5
1961  264.0  1.188807e+08  3.740897e+08  2888.0  523134.50  3887144.0
1962  264.0  1.210511e+08  3.808061e+08  3171.0  533759.50  4023895.5
1963  264.0  1.237333e+08  3.895039e+08  3481.0  544928.75  4139356.5
1964  264.0  1.264378e+08  3.982439e+08  3811.0  556663.00  4224612.5
1965  264.0  1.291813e+08  4.071153e+08  4161.0  565115.00  4277635.5
1966  264.0  1.320404e+08  4.164504e+08  4531.0  569147.00  4331825.0
1967  264.0  1.348980e+08  4.257424e+08  4930.0  577387.25  4385699.5
1968  264.0  1.378358e+08  4.353218e+08  5354.0  583270.00  4450934.5
1969  264.0  1.408789e+08  4.452927e+08  5646.0  587594.25  4530800.0
1970  264.0  1.439693e+08  4.554526e+08  5665.0  595955.00  4615724.0
1971  264.0  1.471163e+08  4.657873e+08  5742.0  608304.50  4715373.5
1972  264.0  1.502115e+08  4.759053e+08  5891.0  625523.00  4823460.5
1973  264.0  1.533501e+08  4.861101e+08  5934.0  646850.75  4921479.0
1974  264.0  1.564755e+08  4.962021e+08  6100.0  648837.75  5022208.5
1975  264.0  1.595368e+08  5.059973e+08  6381.0  662413.75  5115506.5
1976  264.0  1.625684e+08  5.156410e+08  6668.0  695427.00  5228830.0
1977  264.0  1.655894e+08  5.252074e+08  6885.0  736656.25  5343872.0
1978  264.0  1.686657e+08  5.349185e+08  7110.0  787898.25  5432723.5
1979  264.0  1.718262e+08  5.448637e+08  7332.0  809819.00  5558968.0
1980  264.0  1.750146e+08  5.549502e+08  7598.0  818140.50  5728263.0
1981  264.0  1.782819e+08  5.653541e+08  7691.0  828639.25  5874117.0
1982  264.0  1.817011e+08  5.762653e+08  7672.0  841368.75  5999755.0
1983  264.0  1.851494e+08  5.872564e+08  7832.0  857070.00  6144863.0
1984  264.0  1.885888e+08  5.981936e+08  8125.0  873011.75  6311952.5
1985  264.0  1.921012e+08  6.093671e+08  8313.0  889186.50  6444081.5
1986  264.0  1.957174e+08  6.209099e+08  8496.0  907836.50  6538154.0
1987  264.0  1.994330e+08  6.327995e+08  8665.0  927284.50  6714274.5
1988  264.0  2.031770e+08  6.447975e+08  8844.0  956397.75  6894091.0
1989  264.0  2.069199e+08  6.567363e+08  9017.0  989358.75  7007705.5
1990  265.0  2.099328e+08  6.676773e+08  9182.0  1040616.00  7096194.0
1991  265.0  2.136433e+08  6.794607e+08  9354.0  1070266.00  7244891.0
```

1992	265.0	2.172983e+08	6.909190e+08	9466.0	1084441.00	7382050.0
1993	265.0	2.209322e+08	7.022215e+08	9517.0	1097374.00	7494800.0
1994	265.0	2.245370e+08	7.134497e+08	9559.0	1112846.00	7486091.0
1995	265.0	2.281148e+08	7.245981e+08	9585.0	1122457.00	7625357.0
1996	265.0	2.316931e+08	7.357488e+08	9611.0	1150601.00	7683115.0
1997	265.0	2.352589e+08	7.468693e+08	9630.0	1179806.00	7838250.0
1998	265.0	2.388003e+08	7.578820e+08	9634.0	1209832.00	7976789.0
1999	265.0	2.423083e+08	7.687054e+08	9640.0	1240764.00	8009587.0
2000	265.0	2.457971e+08	7.794213e+08	9638.0	1272935.00	8170172.0
2001	265.0	2.492933e+08	7.901104e+08	9621.0	1306590.00	8223941.0
2002	265.0	2.527993e+08	8.006986e+08	9609.0	1341696.00	8372306.0
2003	265.0	2.563022e+08	8.111969e+08	9668.0	1353548.00	8567992.0
2004	265.0	2.598113e+08	8.216948e+08	9791.0	1361172.00	8791853.0
2005	265.0	2.633373e+08	8.321853e+08	9912.0	1369075.00	9026299.0
2006	265.0	2.668790e+08	8.426182e+08	10030.0	1376919.00	9080505.0
2007	265.0	2.704283e+08	8.530054e+08	10149.0	1384861.00	9148092.0
2008	265.0	2.740254e+08	8.635048e+08	10243.0	1444277.00	9219637.0
2009	265.0	2.776655e+08	8.741328e+08	10233.0	1527196.00	9298515.0
2010	265.0	2.813384e+08	8.848783e+08	10241.0	1567220.00	9483836.0
2011	265.0	2.850452e+08	8.958460e+08	10283.0	1609017.00	9726380.0
2012	265.0	2.888896e+08	9.072099e+08	10444.0	1652717.00	9920362.0
2013	265.0	2.927787e+08	9.186849e+08	10694.0	1697753.00	10149577.0
2014	265.0	2.966774e+08	9.301446e+08	10899.0	1743309.00	10282115.0
2015	265.0	3.005462e+08	9.414558e+08	10877.0	1788196.00	10358076.0
2016	265.0	3.044051e+08	9.526720e+08	10852.0	1777557.00	10325452.0
2017	265.0	3.082575e+08	9.638572e+08	10828.0	1791003.00	10300300.0
2018	265.0	3.120276e+08	9.746880e+08	10865.0	1797085.00	10395329.0
2019	265.0	3.157110e+08	9.851690e+08	10956.0	1788878.00	10447666.0
2020	265.0	3.192936e+08	9.952294e+08	11069.0	1790133.00	10606227.0
2021	265.0	3.225180e+08	1.004211e+09	11204.0	1786038.00	10505772.0
2022	265.0	3.254839e+08	1.012174e+09	11312.0	1761985.00	10526073.0

		75%	max
1960	26706062.75	3.031474e+09	
1961	27486939.00	3.072422e+09	
1962	28302886.00	3.126850e+09	
1963	29147077.00	3.193429e+09	
1964	30016841.75	3.260442e+09	
1965	30848919.25	3.328209e+09	
1966	31630099.00	3.398480e+09	
1967	32092469.75	3.468371e+09	
1968	32499274.75	3.540164e+09	
1969	32771488.50	3.614573e+09	
1970	32951857.75	3.690210e+09	
1971	33218150.50	3.767930e+09	
1972	33780171.75	3.843608e+09	
1973	34324598.50	3.920017e+09	

1974	34862533.25	3.995888e+09
1975	35400018.75	4.070022e+09
1976	35920845.25	4.143092e+09
1977	36436755.75	4.215826e+09
1978	36890969.75	4.289796e+09
1979	37287318.75	4.365742e+09
1980	38222997.25	4.442348e+09
1981	39161498.75	4.520917e+09
1982	40119771.50	4.602701e+09
1983	40939798.75	4.684876e+09
1984	41711567.50	4.766641e+09
1985	42420848.00	4.850077e+09
1986	43319791.00	4.936007e+09
1987	44774340.25	5.024289e+09
1988	46263902.75	5.113388e+09
1989	47590215.50	5.202583e+09
1990	47878073.00	5.293395e+09
1991	49937430.00	5.382537e+09
1992	52011554.00	5.470165e+09
1993	52179200.00	5.556623e+09
1994	51921400.00	5.642046e+09
1995	51512800.00	5.726736e+09
1996	51057800.00	5.811580e+09
1997	50594600.00	5.896056e+09
1998	50144500.00	5.979727e+09
1999	49674000.00	6.062289e+09
2000	49176500.00	6.144321e+09
2001	50106657.00	6.226348e+09
2002	51662071.00	6.308141e+09
2003	53205639.00	6.389462e+09
2004	54815607.00	6.470924e+09
2005	56550247.00	6.552700e+09
2006	58143979.00	6.635110e+09
2007	58438310.00	6.717878e+09
2008	58826731.00	6.801710e+09
2009	59095365.00	6.885850e+09
2010	59277417.00	6.970039e+09
2011	59379449.00	7.053995e+09
2012	59539717.00	7.141488e+09
2013	60233948.00	7.229732e+09
2014	60789140.00	7.317970e+09
2015	60730582.00	7.405278e+09
2016	60627498.00	7.492157e+09
2017	60536709.00	7.578221e+09
2018	60421760.00	7.661777e+09
2019	59872579.00	7.742682e+09
2020	61704518.00	7.820964e+09

```
2021 63588334.00 7.888161e+09
2022 65497748.00 7.951150e+09
```

[ ]: data.iloc[105:110,:]

	Country Name	Indicator Code	1960	1961	1962	\		
105	IDA blend	SP.POP.TOTL	120380760.0	123340386.0	126404800.0			
106	Indonesia	SP.POP.TOTL	88382881.0	90816938.0	93345489.0			
107	IDA only	SP.POP.TOTL	273244840.0	280120877.0	287265174.0			
108	Isle of Man	SP.POP.TOTL	48793.0	48422.0	48669.0			
109	India	SP.POP.TOTL	445954579.0	456351876.0	467024193.0			
			1963	1964	1965	1966	1967	\
105	129620425.0	132967846.0	136452685.0	140064295.0	143768745.0			
106	95962527.0	98675061.0	101157868.0	103561105.0	106260749.0			
107	294632894.0	302286207.0	310230648.0	318386686.0	326857777.0			
108	49131.0	49597.0	50057.0	50718.0	51731.0			
109	477933619.0	489059309.0	500114346.0	510992617.0	521987069.0			
			1968	1969	1970	1971	1972	\
105	147587480.0	151541193.0	155624492.0	159733122.0	163971973.0			
106	109138723.0	112149246.0	115228394.0	118347135.0	121504145.0			
107	335669699.0	344654926.0	353539882.0	361723391.0	370219688.0			
108	52910.0	54099.0	55298.0	56431.0	57419.0			
109	533431909.0	545314670.0	557501301.0	569999178.0	582837973.0			
			1973	1974	1975	1976	1977	\
105	168493744.0	173258876.0	178302609.0	183596504.0	189140778.0			
106	124709058.0	127945196.0	131213215.0	134521025.0	137861540.0			
107	379779614.0	389663013.0	399483696.0	409173986.0	419152337.0			
108	58325.0	59179.0	59949.0	60701.0	61506.0			
109	596107483.0	609721951.0	623524219.0	637451448.0	651685628.0			
			1978	1979	1980	1981	1982	\
105	194898741.0	201038256.0	208214579.0	215974148.0	223423864.0			
106	141250964.0	144693087.0	148177096.0	151686337.0	155228658.0			
107	430022867.0	441440230.0	451982661.0	462246693.0	473936564.0			
108	62333.0	63172.0	64022.0	64558.0	64654.0			
109	666267760.0	681248383.0	696828385.0	712869298.0	729169466.0			
			1983	1984	1985	1986	1987	\
105	230352378.0	237102216.0	244392341.0	252086615.0	259986149.0			
106	158790611.0	162331962.0	165791694.0	169135273.0	172421390.0			
107	486705751.0	499790540.0	512976248.0	526144768.0	539531809.0			
108	64598.0	64500.0	64376.0	64663.0	65554.0			
109	745826546.0	762895156.0	780242084.0	797878993.0	815716125.0			

	1988	1989	1990	1991	1992	\
105	268068057.0	276216577.0	284421830.0	292714523.0	300471959.0	
106	175694647.0	178949174.0	182159874.0	185361228.0	188558416.0	
107	552965450.0	566800425.0	580839235.0	594934070.0	610501924.0	
108	66636.0	67740.0	68865.0	69770.0	70307.0	
109	833729681.0	852012673.0	870452165.0	888941756.0	907574049.0	
	1993	1994	1995	1996	1997	\
105	308233041.0	316476824.0	324995506.0	333878799.0	3.428147e+08	
106	191737287.0	194928533.0	198140162.0	201373791.0	2.046280e+08	
107	627217382.0	643572966.0	659768176.0	676066268.0	6.923297e+08	
108	70710.0	71094.0	71454.0	71981.0	7.279300e+04	
109	926351297.0	945261958.0	964279129.0	983281218.0	1.002335e+09	
	1998	1999	2000	2001	2002	\
105	3.518497e+08	3.610078e+08	3.709107e+08	3.811426e+08	3.907044e+08	
106	2.078555e+08	2.109969e+08	2.140724e+08	2.171124e+08	2.201151e+08	
107	7.087355e+08	7.256012e+08	7.426325e+08	7.601358e+08	7.792348e+08	
108	7.372700e+04	7.466800e+04	7.556200e+04	7.639800e+04	7.718300e+04	
109	1.021435e+09	1.040500e+09	1.059634e+09	1.078971e+09	1.098313e+09	
	2003	2004	2005	2006	2007	\
105	3.999955e+08	4.096223e+08	4.193378e+08	4.292508e+08	4.395768e+08	
106	2.230801e+08	2.259386e+08	2.288051e+08	2.317974e+08	2.348583e+08	
107	7.988914e+08	8.181597e+08	8.377179e+08	8.578759e+08	8.781885e+08	
108	7.793900e+04	7.867700e+04	7.941500e+04	8.022800e+04	8.110000e+04	
109	1.117415e+09	1.136265e+09	1.154639e+09	1.172374e+09	1.189692e+09	
	2008	2009	2010	2011	2012	\
105	4.502695e+08	4.614102e+08	4.732045e+08	4.849199e+08	4.958685e+08	
106	2.379365e+08	2.409813e+08	2.440162e+08	2.470997e+08	2.502227e+08	
107	8.981994e+08	9.185279e+08	9.397310e+08	9.616121e+08	9.835899e+08	
108	8.199700e+04	8.291500e+04	8.382800e+04	8.435000e+04	8.433800e+04	
109	1.206735e+09	1.223640e+09	1.240614e+09	1.257621e+09	1.274487e+09	
	2013	2014	2015	2016	2017	\
105	5.064434e+08	5.169001e+08	5.271759e+08	5.373584e+08	5.479850e+08	
106	2.532759e+08	2.562298e+08	2.590920e+08	2.618502e+08	2.644989e+08	
107	1.005081e+09	1.026805e+09	1.049402e+09	1.073443e+09	1.098416e+09	
108	8.414400e+04	8.389600e+04	8.359300e+04	8.345000e+04	8.358000e+04	
109	1.291132e+09	1.307247e+09	1.322867e+09	1.338636e+09	1.354196e+09	
	2018	2019	2020	2021	2022	
105	5.591556e+08	5.706333e+08	5.826371e+08	5.950160e+08	6.077360e+08	
106	2.670668e+08	2.695829e+08	2.718580e+08	2.737532e+08	2.755013e+08	
107	1.123984e+09	1.150416e+09	1.177589e+09	1.205017e+09	1.232235e+09	
108	8.377500e+04	8.393300e+04	8.404600e+04	8.426300e+04	8.451900e+04	

```
109 1.369003e+09 1.383112e+09 1.396387e+09 1.407564e+09 1.417173e+09
```

## 1 HANDLING MISSING VALUES

```
[ ]: data.duplicated()
```

```
[ ]: 0      False
     1      False
     2      False
     3      False
     4      False
     ...
    261     False
    262     False
    263     False
    264     False
    265     False
Length: 266, dtype: bool
```

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

```
[ ]: 0
```

### DETECTION OF MISSING VALUES

```
[ ]: data.isnull().sum()
```

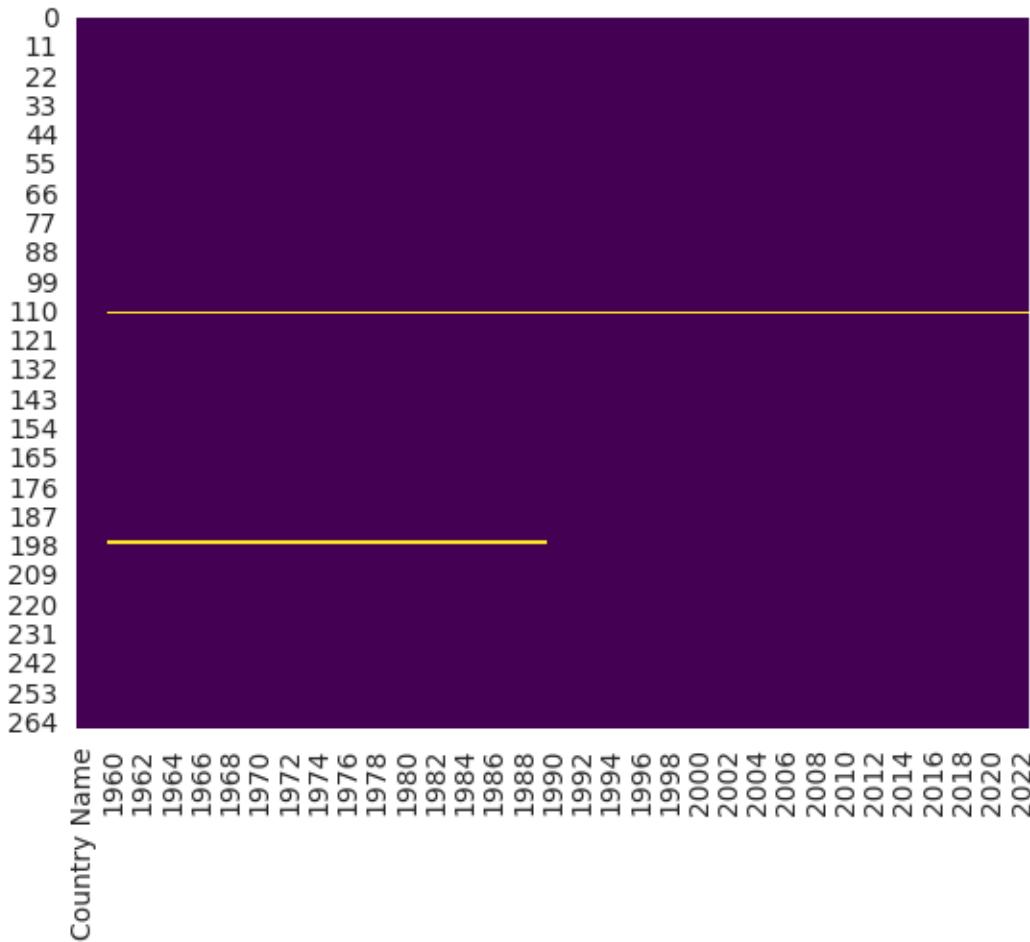
```
[ ]: Country Name      0
     Indicator Code      0
     1960                  2
     1961                  2
     1962                  2
     1963                  2
     1964                  2
     1965                  2
     1966                  2
     1967                  2
     1968                  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
```

```
[ ]: sb.heatmap(data.isnull(), cbar=False, cmap='viridis')
```

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



```
[ ]: (data.isna().sum()*100/df.shape[0]).sort_values(ascending=True)
```

```
[ ]: Country Name      0.00000
    Indicator Code     0.00000
    1993                0.37594
    1994                0.37594
    1995                0.37594
    1996                0.37594
    1997                0.37594
    1998                0.37594
    1999                0.37594
    2000                0.37594
    2001                0.37594
```

2002	0.37594
2003	0.37594
2004	0.37594
2005	0.37594
2006	0.37594
2007	0.37594
2008	0.37594
2009	0.37594
2010	0.37594
2011	0.37594
2012	0.37594
2013	0.37594
2014	0.37594
2015	0.37594
2016	0.37594
2017	0.37594
2018	0.37594
2019	0.37594
2020	0.37594
1992	0.37594
1991	0.37594
1990	0.37594
2022	0.37594
2021	0.37594
1960	0.75188
1961	0.75188
1962	0.75188
1963	0.75188
1964	0.75188
1965	0.75188
1966	0.75188
1967	0.75188
1968	0.75188
1969	0.75188
1970	0.75188
1971	0.75188
1972	0.75188
1973	0.75188
1989	0.75188
1976	0.75188
1977	0.75188
1978	0.75188
1979	0.75188
1980	0.75188
1981	0.75188
1982	0.75188
1983	0.75188

```

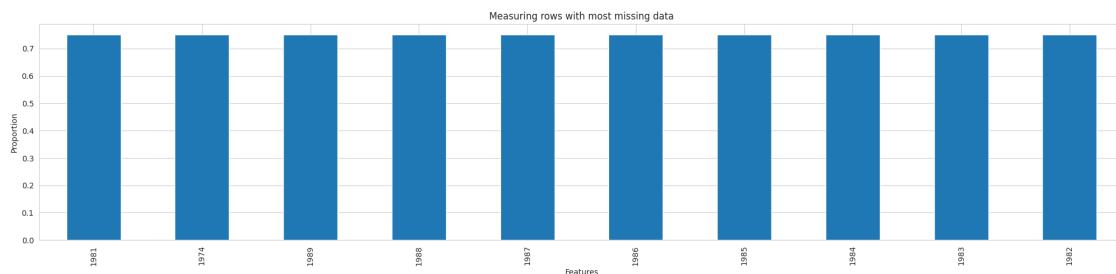
1984          0.75188
1985          0.75188
1986          0.75188
1987          0.75188
1988          0.75188
1975          0.75188
1974          0.75188
dtype: float64

```

```

[ ]: sb.set_style('whitegrid')
plt.figure(figsize=(25,5))
miss_values=data.isna().sum()*100/df.shape[0]
miss_values=miss_values.sort_values(ascending=False)[0:10]
ax=miss_values.plot(kind='bar')
ax.set_title('Measuring rows with most missing data')
ax.set_ylabel('Proportion')
ax.set_xlabel('Features')
ax.tick_params(axis='x',rotation=90)

```



## HANDLING MISSING VALUES

General Recommendations:

Always understand the nature of missing values before deciding on a strategy.

Consider the impact of missing data on your analysis and results.

Document any imputation or handling strategy for transparency.

```
[ ]: data.isna().sum()
```

```

[ ]: Country 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.fillna(method='ffill',inplace=True)
```

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

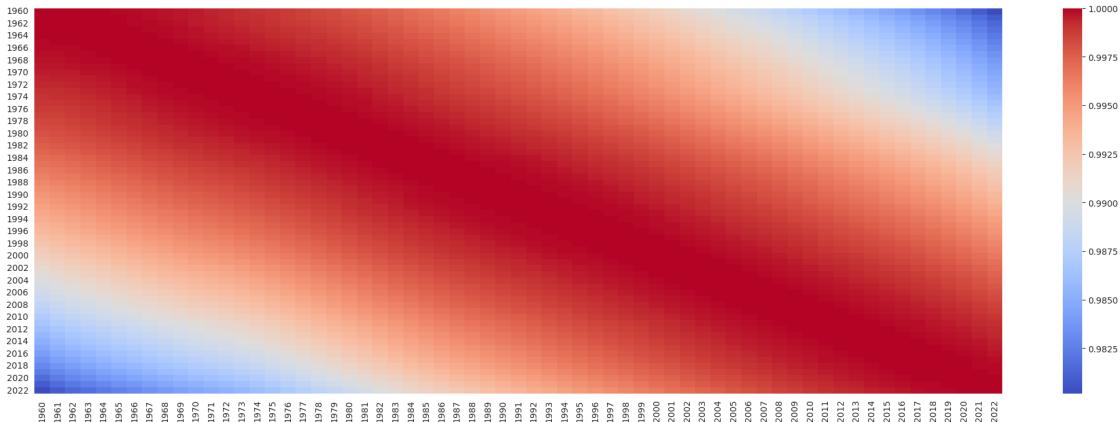
```
[ ]: Country Name      0  
Country Code       0  
Indicator Name     0  
Indicator Code     0  
1960              0  
1961              0  
1962              0  
1963              0  
1964              0  
1965              0  
1966              0  
1967              0  
1968              0  
1969              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
```

```
[ ]: correlation=df.corr()
plt.figure(figsize=(25,8))
sb.heatmap(correlation,cmap='coolwarm')
plt.show()
```

```
<ipython-input-77-643a2a9300a4>:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
```

```
correlation=df.corr()
```



```
[ ]: # sort data based on total population for 2022
total_population = data.sort_values(by='2022', ascending=False)
total_population
```

```
[ ]:      Country Name Indicator Code    1960    1961 \
259          World     SP.POP.TOTL 3.031474e+09 3.072422e+09
103  IDA & IBRD total     SP.POP.TOTL 2.297972e+09 2.329504e+09
140  Low & middle income     SP.POP.TOTL 2.243179e+09 2.274050e+09
156   Middle income     SP.POP.TOTL 2.107416e+09 2.135116e+09
102    IBRD only     SP.POP.TOTL 1.904347e+09 1.926043e+09
..        ...
255  British Virgin Islands     SP.POP.TOTL 7.850000e+03 7.885000e+03
188         Palau     SP.POP.TOTL 9.446000e+03 9.639000e+03
179         Nauru     SP.POP.TOTL 4.582000e+03 4.753000e+03
245         Tuvalu     SP.POP.TOTL 5.404000e+03 5.436000e+03
110 Not classified     SP.POP.TOTL       NaN       NaN
                                         \
1962    1963    1964    1965    1966
259  3.126850e+09 3.193429e+09 3.260442e+09 3.328209e+09 3.398480e+09
103  2.374276e+09 2.431314e+09 2.488809e+09 2.547220e+09 2.608583e+09
140  2.318173e+09 2.374537e+09 2.431355e+09 2.489219e+09 2.550049e+09
156  2.175940e+09 2.228877e+09 2.282094e+09 2.336193e+09 2.393145e+09
102  1.960606e+09 2.007061e+09 2.053555e+09 2.100537e+09 2.150132e+09
..        ...
255  7.902000e+03 7.919000e+03 7.949000e+03 8.018000e+03 8.139000e+03
188  9.851000e+03 1.007600e+04 1.031800e+04 1.056300e+04 1.081300e+04
179  4.950000e+03 5.198000e+03 5.484000e+03 5.804000e+03 6.021000e+03
245  5.471000e+03 5.503000e+03 5.525000e+03 5.548000e+03 5.591000e+03
110       NaN       NaN       NaN       NaN       NaN
```

	1967	1968	1969	1970	1971	\
259	3.468371e+09	3.540164e+09	3.614573e+09	3.690210e+09	3.767930e+09	
103	2.670092e+09	2.733533e+09	2.799372e+09	2.866379e+09	2.933804e+09	
140	2.610859e+09	2.673667e+09	2.738980e+09	2.805573e+09	2.872596e+09	
156	2.449880e+09	2.508411e+09	2.569347e+09	2.631441e+09	2.693867e+09	
102	2.199466e+09	2.250275e+09	2.303176e+09	2.357215e+09	2.412348e+09	
..	...	...	...	...	...	
255	8.337000e+03	8.649000e+03	9.117000e+03	9.581000e+03	9.813000e+03	
188	1.099200e+04	1.107900e+04	1.115900e+04	1.136600e+04	1.174000e+04	
179	6.114000e+03	6.288000e+03	6.508000e+03	6.663000e+03	6.778000e+03	
245	5.657000e+03	5.729000e+03	5.779000e+03	5.814000e+03	5.854000e+03	
110	NaN	NaN	NaN	NaN	NaN	
	1972	1973	1974	1975	1976	\
259	3.843608e+09	3.920017e+09	3.995888e+09	4.070022e+09	4.143092e+09	
103	3.000609e+09	3.068508e+09	3.136140e+09	3.202366e+09	3.268151e+09	
140	2.938845e+09	3.006126e+09	3.073026e+09	3.138386e+09	3.203257e+09	
156	2.755510e+09	2.818049e+09	2.879964e+09	2.940208e+09	2.999857e+09	
102	2.466418e+09	2.520234e+09	2.573218e+09	2.624580e+09	2.675380e+09	
..	...	...	...	...	...	
255	9.907000e+03	1.001800e+04	1.014600e+04	1.028200e+04	1.042800e+04	
188	1.217200e+04	1.254100e+04	1.266200e+04	1.260700e+04	1.253900e+04	
179	6.883000e+03	6.990000e+03	7.096000e+03	7.199000e+03	7.306000e+03	
245	5.891000e+03	5.934000e+03	6.100000e+03	6.381000e+03	6.677000e+03	
110	NaN	NaN	NaN	NaN	NaN	
	1977	1978	1979	1980	1981	\
259	4.215826e+09	4.289796e+09	4.365742e+09	4.442348e+09	4.520917e+09	
103	3.333672e+09	3.400307e+09	3.468586e+09	3.537803e+09	3.609061e+09	
140	3.267883e+09	3.333664e+09	3.401138e+09	3.469529e+09	3.539934e+09	
156	3.059291e+09	3.119391e+09	3.180907e+09	3.244431e+09	3.310520e+09	
102	2.725379e+09	2.775385e+09	2.826108e+09	2.877606e+09	2.930840e+09	
..	...	...	...	...	...	
255	1.057800e+04	1.072700e+04	1.087600e+04	1.110900e+04	1.147200e+04	
188	1.246300e+04	1.238700e+04	1.230000e+04	1.225200e+04	1.238600e+04	
179	7.441000e+03	7.541000e+03	7.586000e+03	7.635000e+03	7.691000e+03	
245	6.984000e+03	7.297000e+03	7.552000e+03	7.731000e+03	7.874000e+03	
110	NaN	NaN	NaN	NaN	NaN	
	1982	1983	1984	1985	1986	\
259	4.602701e+09	4.684876e+09	4.766641e+09	4.850077e+09	4.936007e+09	
103	3.683951e+09	3.759701e+09	3.835415e+09	3.912715e+09	3.992266e+09	
140	3.613997e+09	3.688959e+09	3.763903e+09	3.840450e+09	3.919271e+09	
156	3.379188e+09	3.447962e+09	3.516455e+09	3.586381e+09	3.658613e+09	
102	2.986591e+09	3.042643e+09	3.098523e+09	3.155347e+09	3.214035e+09	
..	...	...	...	...	...	
255	1.188900e+04	1.231500e+04	1.275300e+04	1.320200e+04	1.365600e+04	

188	1.266600e+04	1.297300e+04	1.329900e+04	1.364400e+04	1.398500e+04		
179	7.672000e+03	7.832000e+03	8.125000e+03	8.313000e+03	8.527000e+03		
245	7.993000e+03	8.099000e+03	8.197000e+03	8.329000e+03	8.496000e+03		
110	NaN	NaN	NaN	NaN	NaN	NaN	
	1987	1988	1989	1990	1991	\	
259	5.024289e+09	5.113388e+09	5.202583e+09	5.293395e+09	5.382537e+09		
103	4.073993e+09	4.156275e+09	4.238117e+09	4.319687e+09	4.400104e+09		
140	4.000290e+09	4.081930e+09	4.163156e+09	4.246144e+09	4.326436e+09		
156	3.732821e+09	3.807581e+09	3.881585e+09	3.957385e+09	4.030332e+09		
102	3.274475e+09	3.335242e+09	3.395100e+09	3.454426e+09	3.512456e+09		
..	..	..	..	..	..		
255	1.411700e+04	1.459300e+04	1.509100e+04	1.561700e+04	1.609000e+04		
188	1.430900e+04	1.463200e+04	1.495700e+04	1.529300e+04	1.564000e+04		
179	8.759000e+03	9.022000e+03	9.306000e+03	9.598000e+03	9.902000e+03		
245	8.665000e+03	8.844000e+03	9.017000e+03	9.182000e+03	9.354000e+03		
110	NaN	NaN	NaN	NaN	NaN	NaN	
	1992	1993	1994	1995	1996	\	
259	5.470165e+09	5.556623e+09	5.642046e+09	5.726736e+09	5.811580e+09		
103	4.478741e+09	4.556953e+09	4.634770e+09	4.712145e+09	4.789545e+09		
140	4.404987e+09	4.482824e+09	4.560257e+09	4.637328e+09	4.714458e+09		
156	4.099989e+09	4.168151e+09	4.236409e+09	4.304147e+09	4.371427e+09		
102	3.567767e+09	3.621503e+09	3.674721e+09	3.727381e+09	3.779600e+09		
..	..	..	..	..	..		
255	1.647600e+04	1.685500e+04	1.725500e+04	1.767400e+04	1.812000e+04		
188	1.600400e+04	1.638000e+04	1.677000e+04	1.720900e+04	1.773200e+04		
179	1.012900e+04	1.022600e+04	1.027400e+04	1.031600e+04	1.034500e+04		
245	9.466000e+03	9.517000e+03	9.559000e+03	9.585000e+03	9.611000e+03		
110	NaN	NaN	NaN	NaN	NaN	NaN	
	1997	1998	1999	2000	2001	\	
259	5.896056e+09	5.979727e+09	6.062289e+09	6.144321e+09	6.226348e+09		
103	4.866753e+09	4.943424e+09	5.019067e+09	5.093906e+09	5.168706e+09		
140	4.791350e+09	4.867622e+09	4.942885e+09	5.017754e+09	5.092598e+09		
156	4.438595e+09	4.505163e+09	4.570305e+09	4.635161e+09	4.699674e+09		
102	3.831608e+09	3.882839e+09	3.932458e+09	3.980363e+09	4.027428e+09		
..	..	..	..	..	..		
255	1.859100e+04	1.907900e+04	1.958600e+04	2.010400e+04	2.065700e+04		
188	1.829700e+04	1.885200e+04	1.938000e+04	1.972600e+04	1.982800e+04		
179	1.036300e+04	1.037700e+04	1.038300e+04	1.037700e+04	1.036300e+04		
245	9.630000e+03	9.634000e+03	9.640000e+03	9.638000e+03	9.621000e+03		
110	NaN	NaN	NaN	NaN	NaN	NaN	
	2002	2003	2004	2005	2006	\	
259	6.308141e+09	6.389462e+09	6.470924e+09	6.552700e+09	6.635110e+09		
103	5.243283e+09	5.317576e+09	5.391795e+09	5.465994e+09	5.539727e+09		

140	5.167128e+09	5.241165e+09	5.315126e+09	5.389050e+09	5.462498e+09		
156	4.762356e+09	4.823952e+09	4.885771e+09	4.947164e+09	5.007185e+09		
102	4.073344e+09	4.118690e+09	4.164013e+09	4.208939e+09	4.252600e+09		
..	...	...	...	...	...	...	
255	2.128800e+04	2.198200e+04	2.271500e+04	2.349700e+04	2.432300e+04		
188	1.985100e+04	1.988000e+04	1.990700e+04	1.983100e+04	1.961900e+04		
179	1.035100e+04	1.034400e+04	1.033500e+04	1.031800e+04	1.029400e+04		
245	9.609000e+03	9.668000e+03	9.791000e+03	9.912000e+03	1.003000e+04		
110	NaN	NaN	NaN	NaN	NaN	NaN	
	2007	2008	2009	2010	2011	\	
259	6.717878e+09	6.801710e+09	6.885850e+09	6.970039e+09	7.053995e+09		
103	5.612901e+09	5.686804e+09	5.762172e+09	5.838773e+09	5.917667e+09		
140	5.535561e+09	5.609367e+09	5.684455e+09	5.760865e+09	5.839424e+09		
156	5.066551e+09	5.126726e+09	5.187846e+09	5.249967e+09	5.313918e+09		
102	4.295136e+09	4.338335e+09	4.382234e+09	4.425838e+09	4.471135e+09		
..	...	...	...	...	...	...	
255	2.519100e+04	2.611500e+04	2.704400e+04	2.755600e+04	2.796200e+04		
188	1.936600e+04	1.910200e+04	1.882600e+04	1.854000e+04	1.824000e+04		
179	1.026700e+04	1.024300e+04	1.023300e+04	1.024100e+04	1.028300e+04		
245	1.014900e+04	1.027200e+04	1.040800e+04	1.055000e+04	1.070000e+04		
110	NaN	NaN	NaN	NaN	NaN	NaN	
	2012	2013	2014	2015	2016	\	
259	7.141488e+09	7.229732e+09	7.317970e+09	7.405278e+09	7.492157e+09		
103	5.998534e+09	6.079930e+09	6.161220e+09	6.241659e+09	6.321547e+09		
140	5.919975e+09	6.001075e+09	6.082091e+09	6.162304e+09	6.242060e+09		
156	5.379973e+09	5.447239e+09	5.514389e+09	5.580131e+09	5.644377e+09		
102	4.519075e+09	4.568406e+09	4.617515e+09	4.665081e+09	4.710746e+09		
..	...	...	...	...	...	...	
255	2.842100e+04	2.865700e+04	2.897100e+04	2.936600e+04	2.973900e+04		
188	1.794600e+04	1.780500e+04	1.779600e+04	1.779400e+04	1.781600e+04		
179	1.044400e+04	1.069400e+04	1.094000e+04	1.118500e+04	1.143700e+04		
245	1.085400e+04	1.091800e+04	1.089900e+04	1.087700e+04	1.085200e+04		
110	NaN	NaN	NaN	NaN	NaN	NaN	
	2017	2018	2019	2020	2021	\	
259	7.578221e+09	7.661777e+09	7.742682e+09	7.820964e+09	7.888161e+09		
103	6.401430e+09	6.479098e+09	6.554881e+09	6.628068e+09	6.695340e+09		
140	6.322111e+09	6.400439e+09	6.476995e+09	6.550747e+09	6.618692e+09		
156	5.708165e+09	5.769573e+09	5.828240e+09	5.883693e+09	5.933442e+09		
102	4.755029e+09	4.795958e+09	4.833831e+09	4.867842e+09	4.895307e+09		
..	...	...	...	...	...	...	
255	3.006000e+04	3.033500e+04	3.061000e+04	3.091000e+04	3.112200e+04		
188	1.783700e+04	1.786400e+04	1.791600e+04	1.797200e+04	1.802400e+04		
179	1.168200e+04	1.192400e+04	1.213200e+04	1.231500e+04	1.251100e+04		
245	1.082800e+04	1.086500e+04	1.095600e+04	1.106900e+04	1.120400e+04		

```

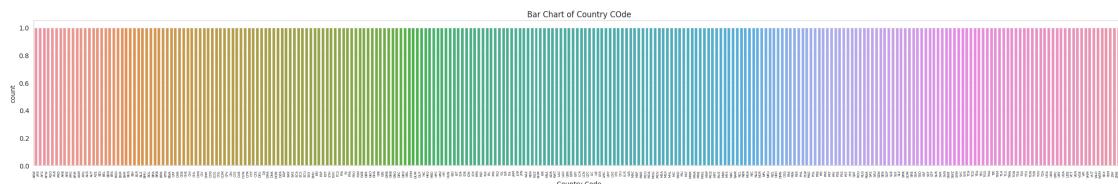
110          NaN          NaN          NaN          NaN          NaN
2022
259 7.951150e+09
103 6.754327e+09
140 6.677949e+09
156 5.974221e+09
102 4.914355e+09
.. ...
255 3.130500e+04
188 1.805500e+04
179 1.266800e+04
245 1.131200e+04
110          NaN

```

[266 rows x 65 columns]

## #UNIVARIATE ANALYSIS FOR CATEGORICAL FEATURES

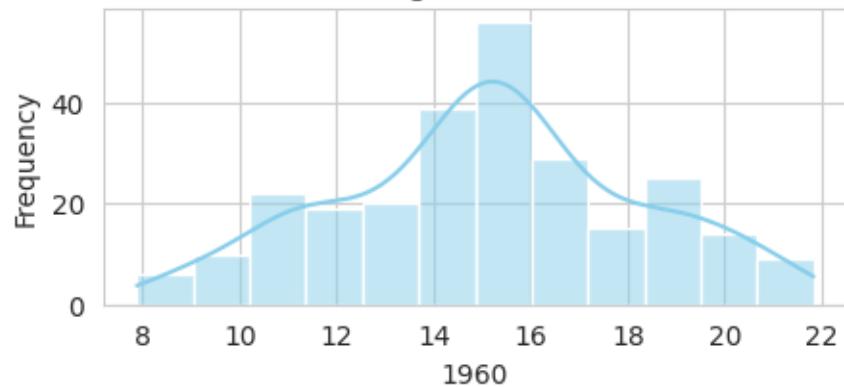
```
[ ]: plt.figure(figsize=(30,4))
sb.countplot(x=df['Country Code'])
plt.title(f'Bar Chart of Country Code')
plt.xticks(rotation=90,ha='right',fontsize=4) #ha='right' means horizontal
    ↪alignment of text in right side
plt.show()
```



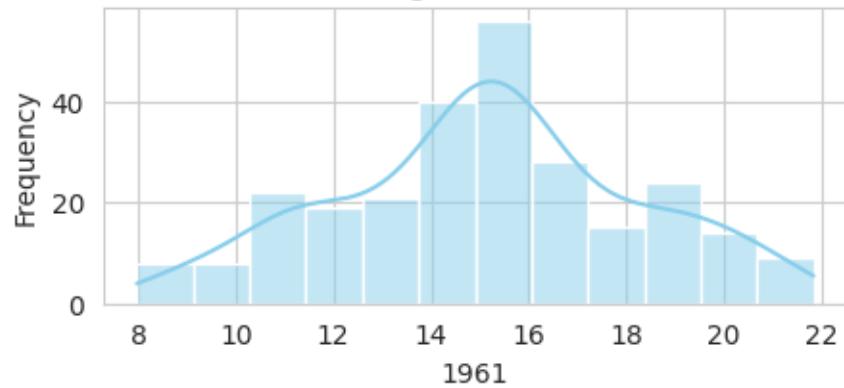
## NUMERICAL FEATURES HISTOGRAM

```
[ ]: for column in data.select_dtypes(include='number').columns:
    # Histogram
    plt.figure(figsize=(5,2))
    sb.histplot(np.log(data[column]), kde=True, color='skyblue') #kde=True adds
        ↪kernel density estimate plot
    plt.title(f'Histogram of {column}')
    plt.xlabel(column)
    plt.ylabel('Frequency')
    plt.show()
```

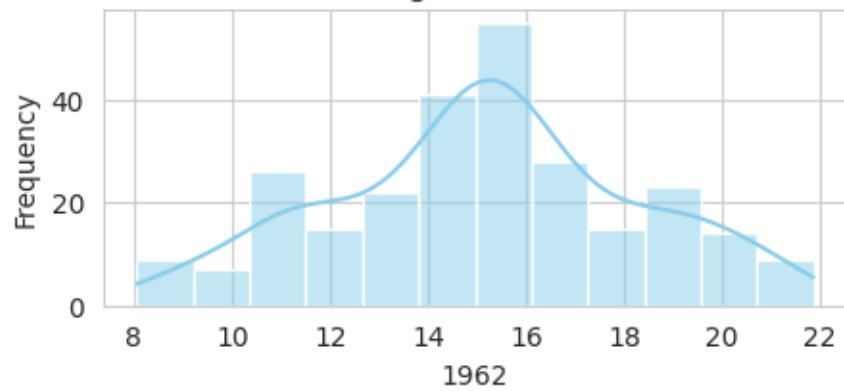
Histogram of 1960



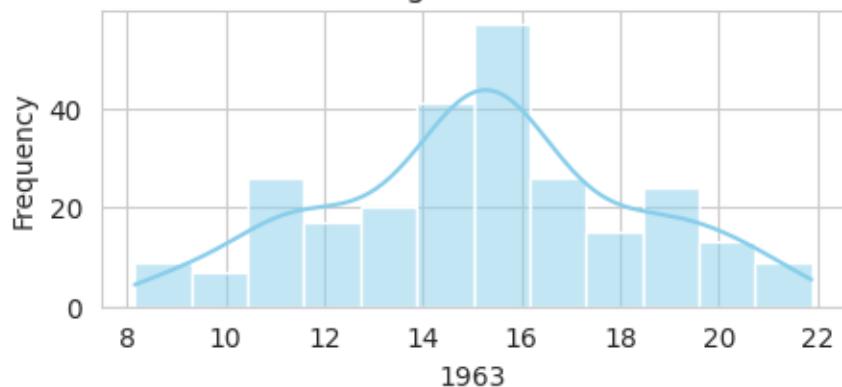
Histogram of 1961



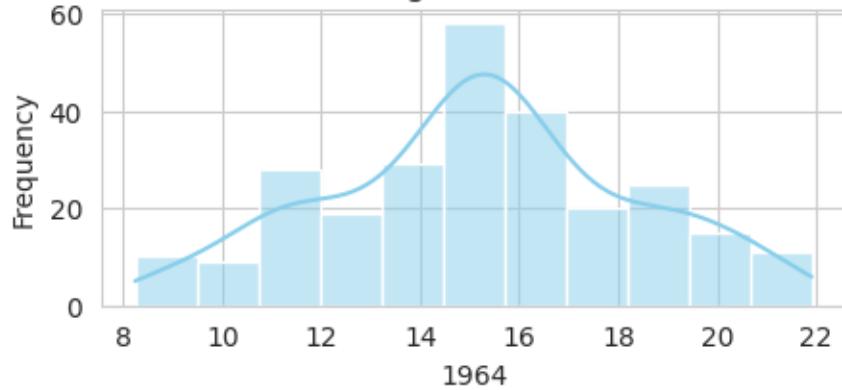
Histogram of 1962



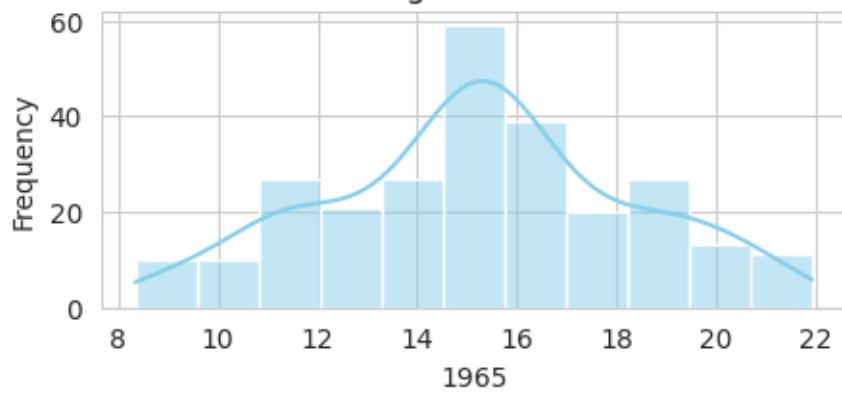
### Histogram of 1963



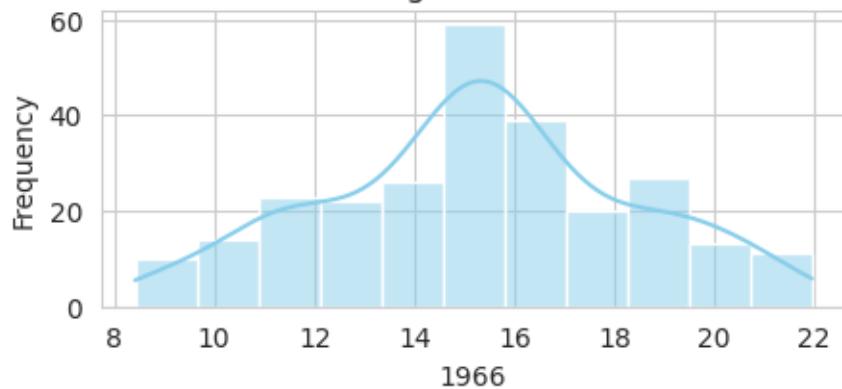
### Histogram of 1964



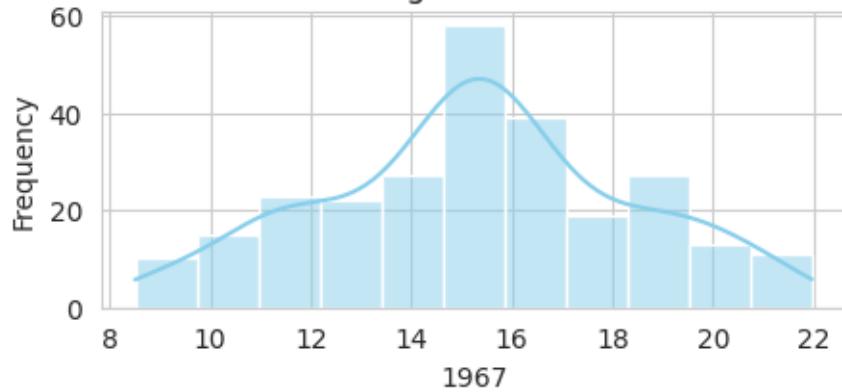
### Histogram of 1965



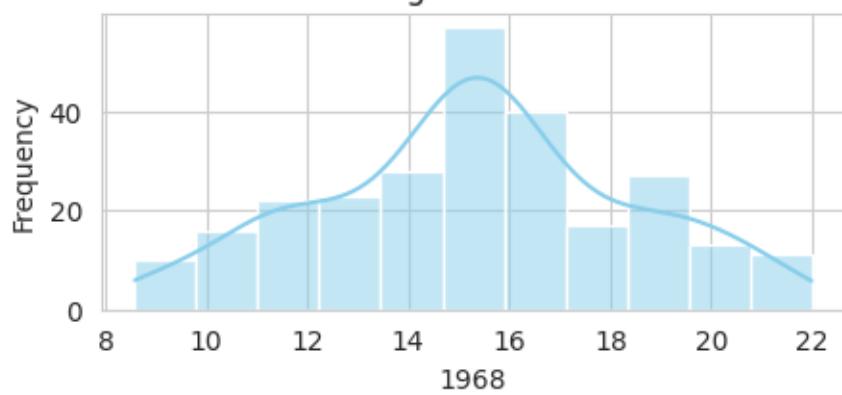
Histogram of 1966



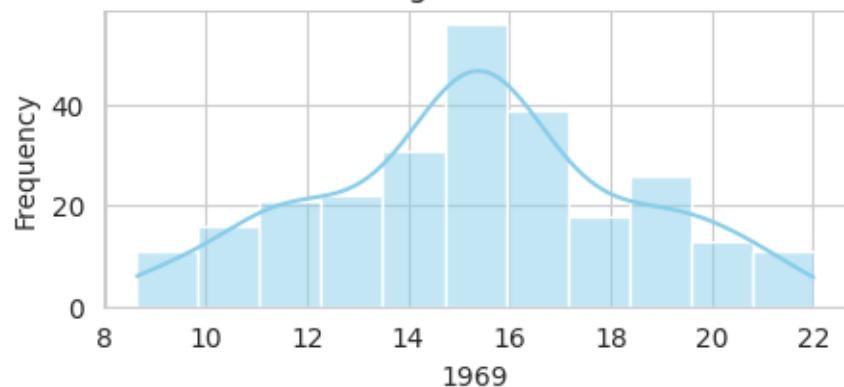
Histogram of 1967



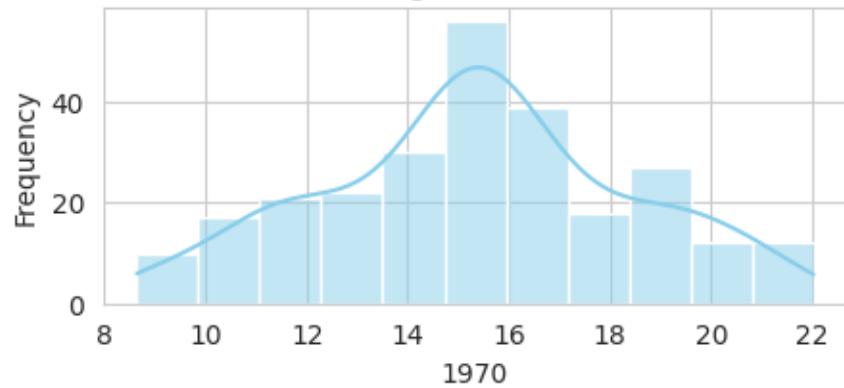
Histogram of 1968



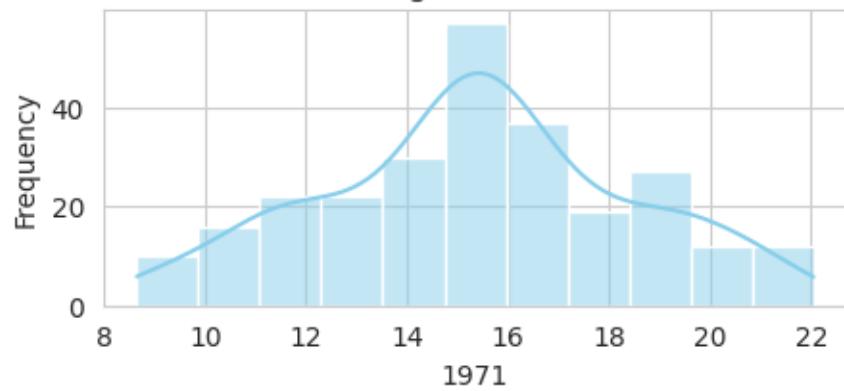
Histogram of 1969



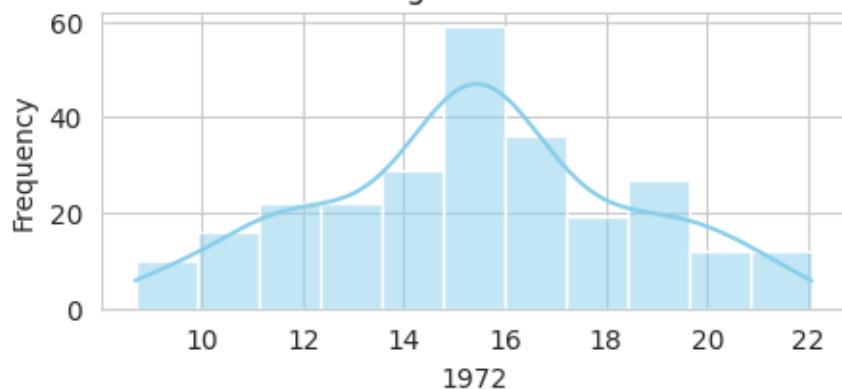
Histogram of 1970



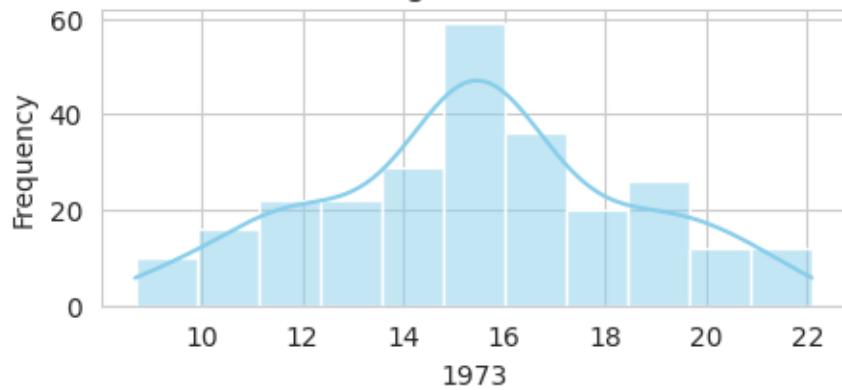
Histogram of 1971



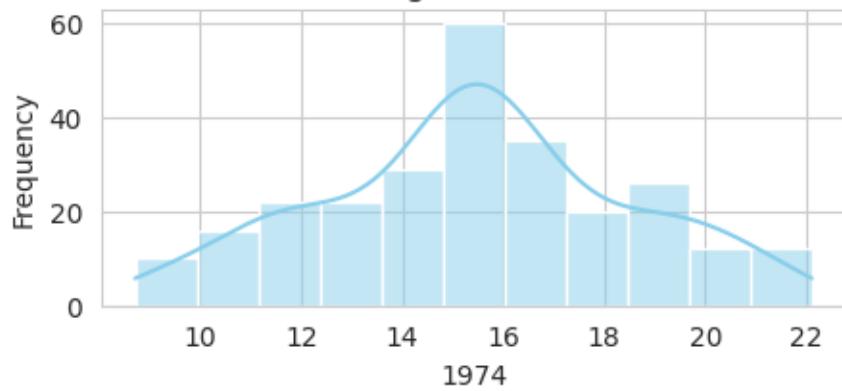
### Histogram of 1972



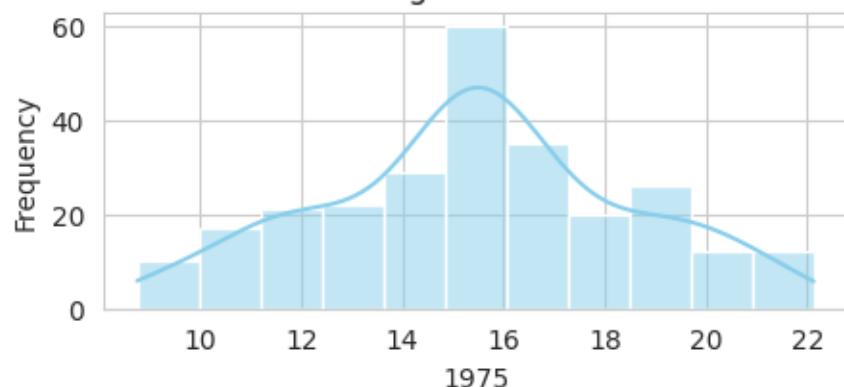
### Histogram of 1973



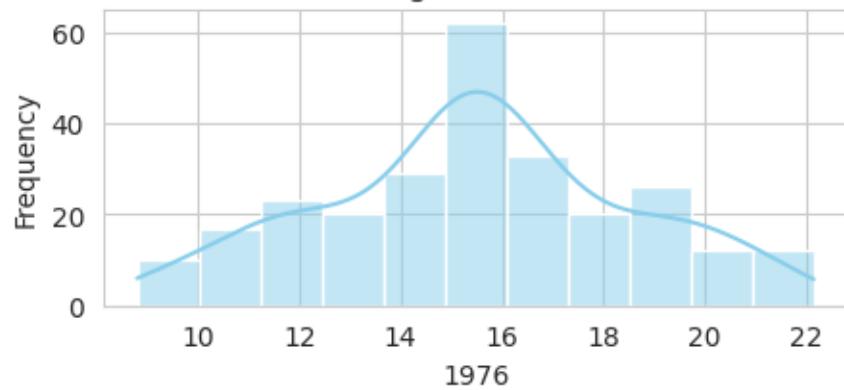
### Histogram of 1974



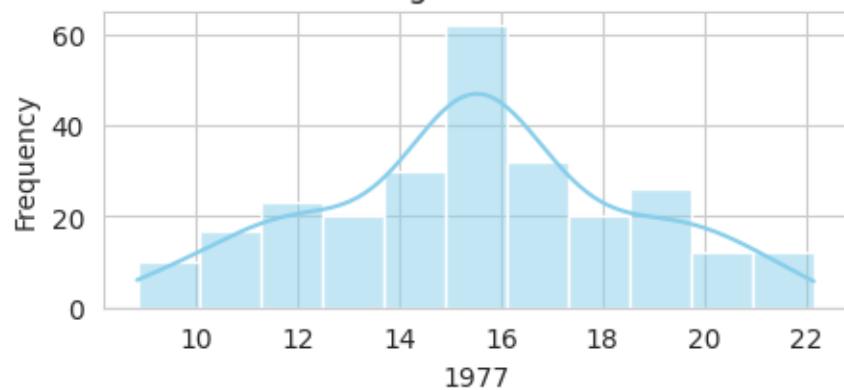
Histogram of 1975



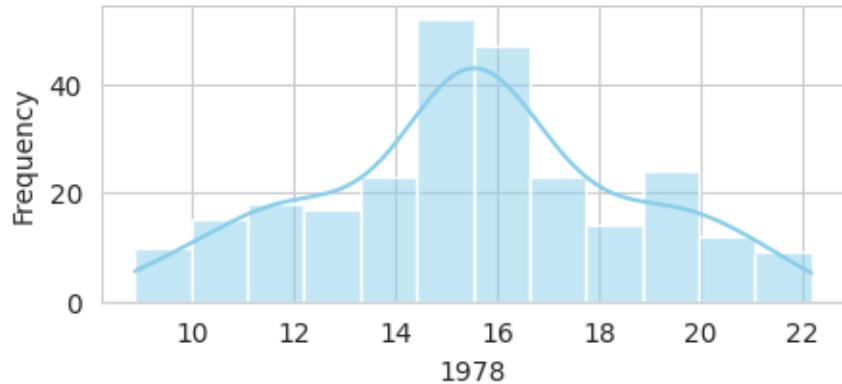
Histogram of 1976



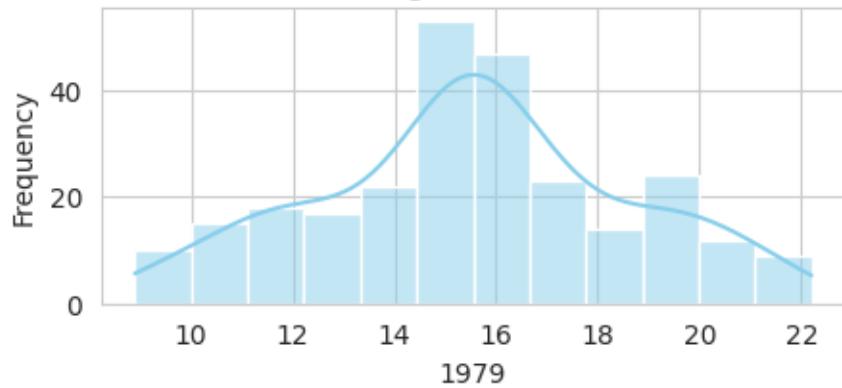
Histogram of 1977



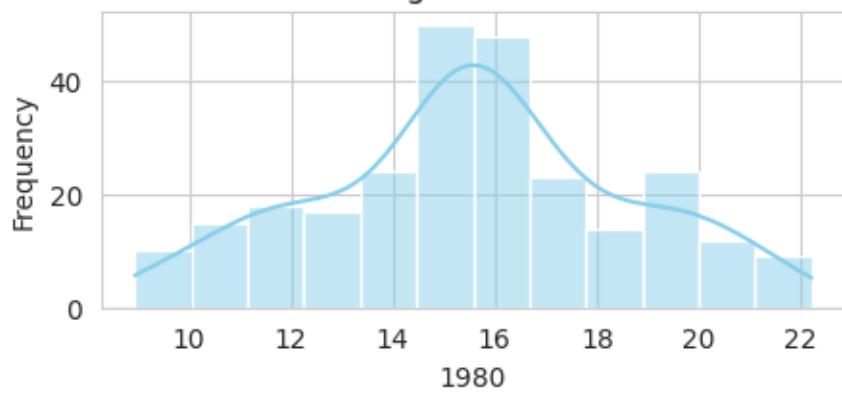
Histogram of 1978



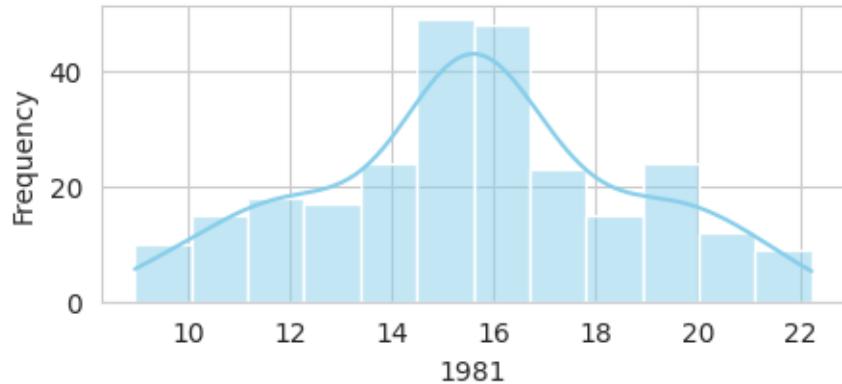
Histogram of 1979



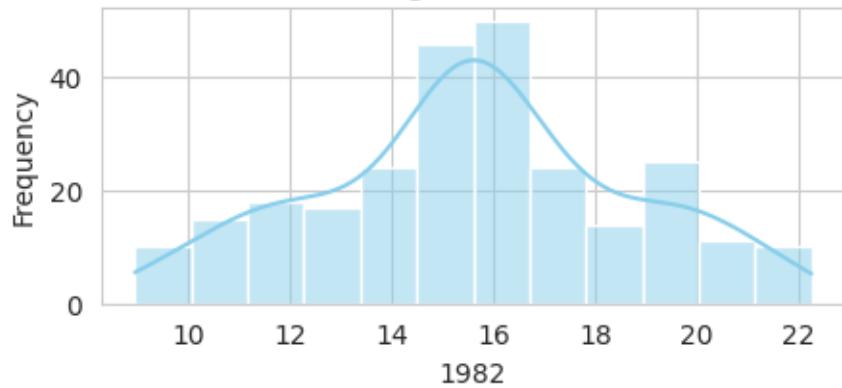
Histogram of 1980



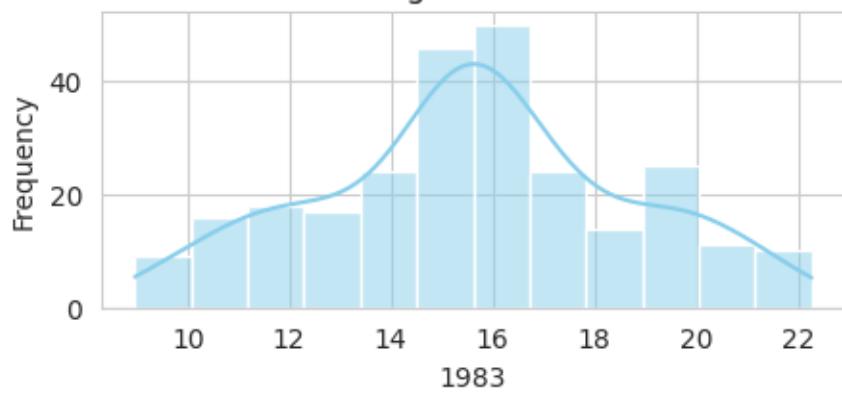
Histogram of 1981



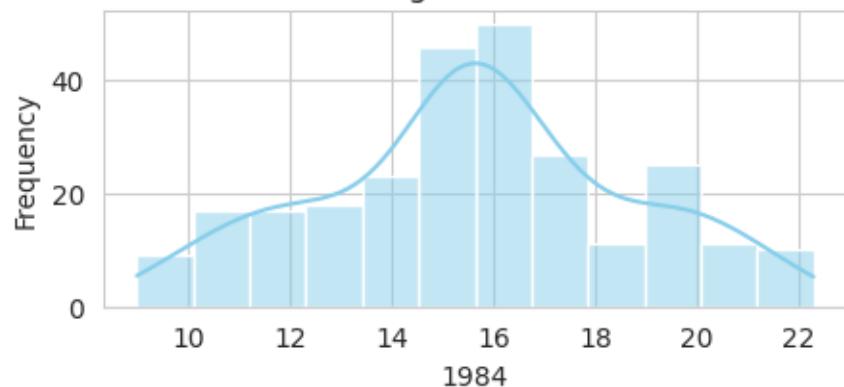
Histogram of 1982



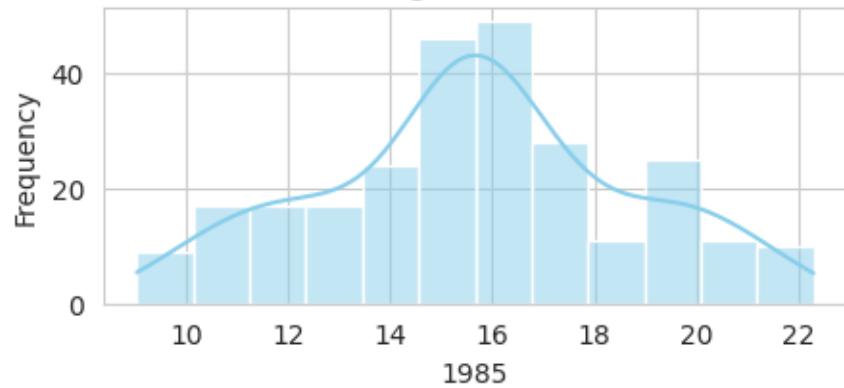
Histogram of 1983



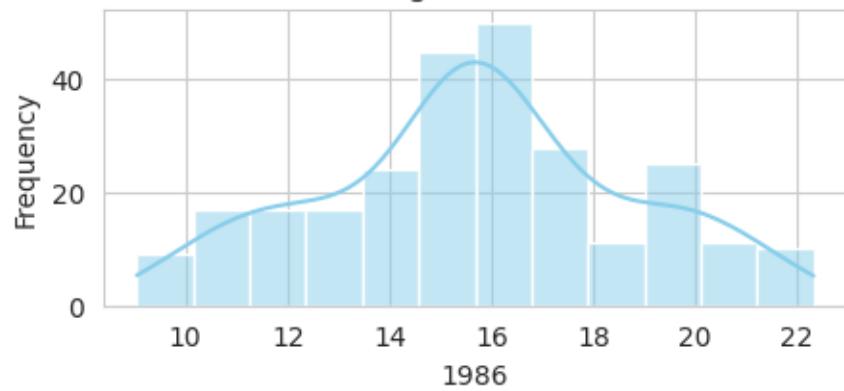
Histogram of 1984



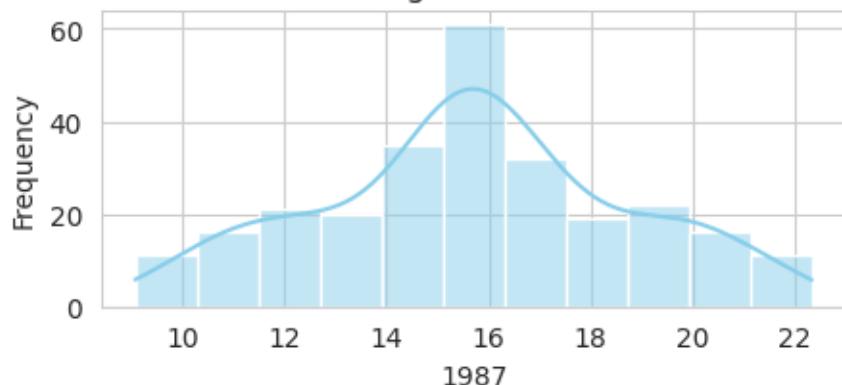
Histogram of 1985



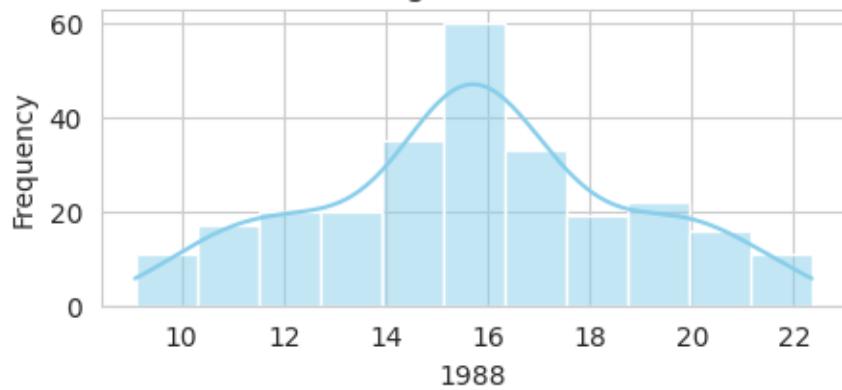
Histogram of 1986



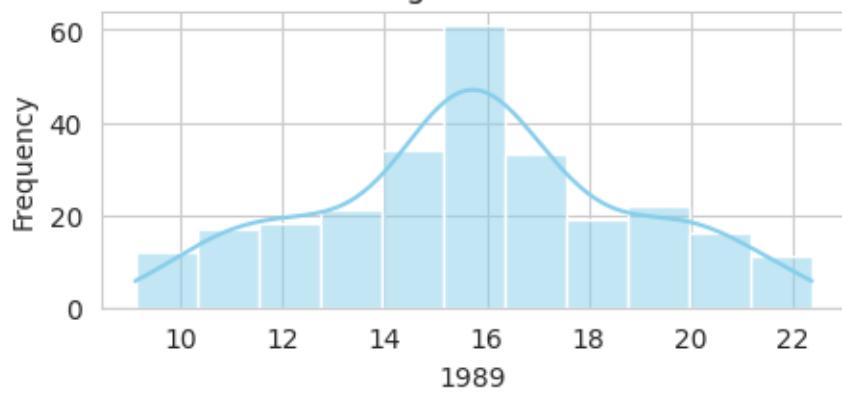
Histogram of 1987

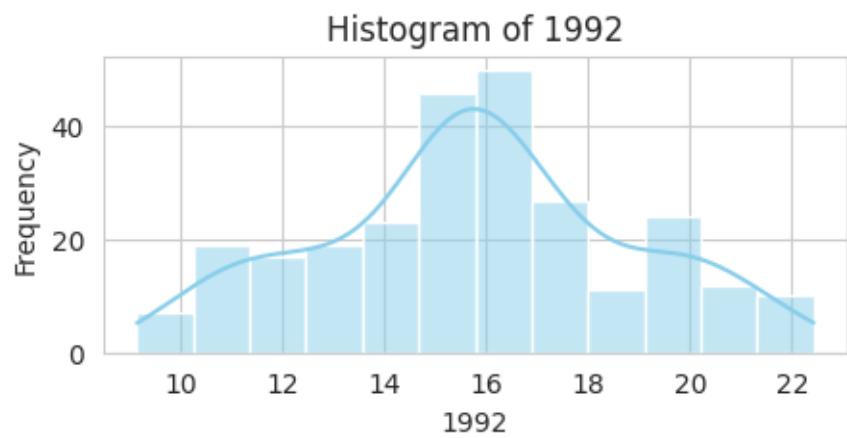
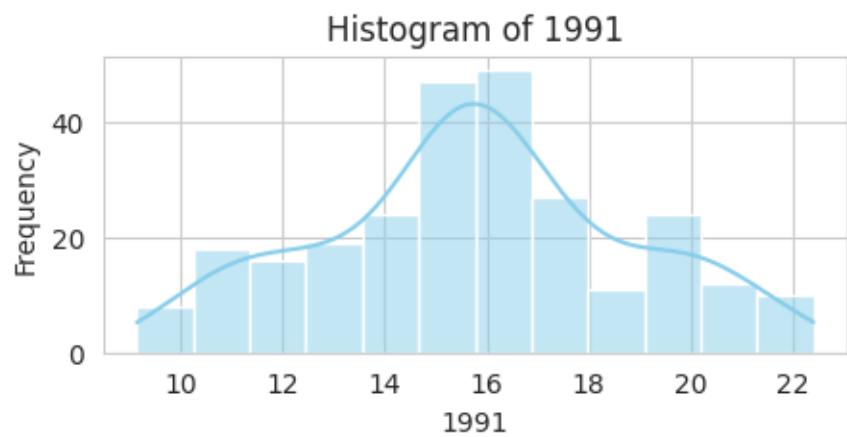
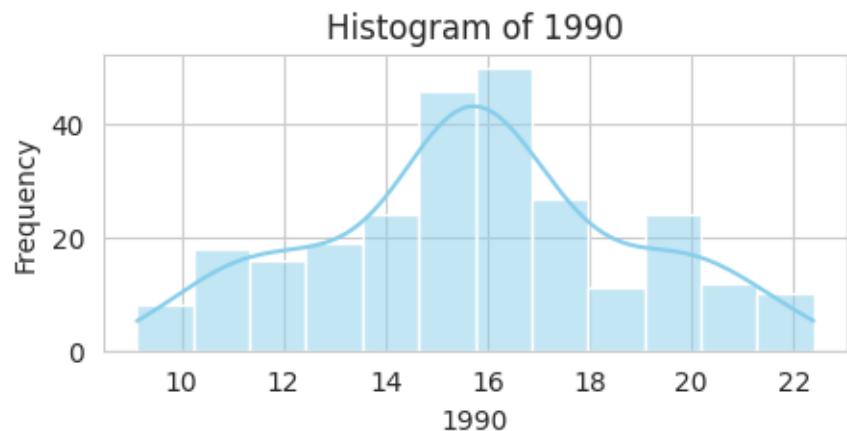


Histogram of 1988

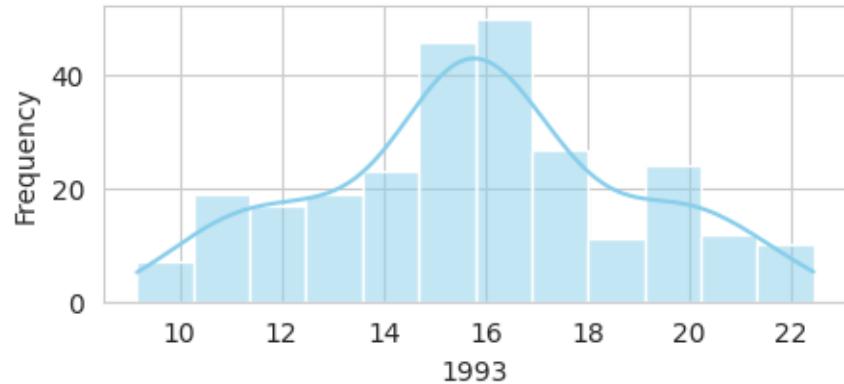


Histogram of 1989

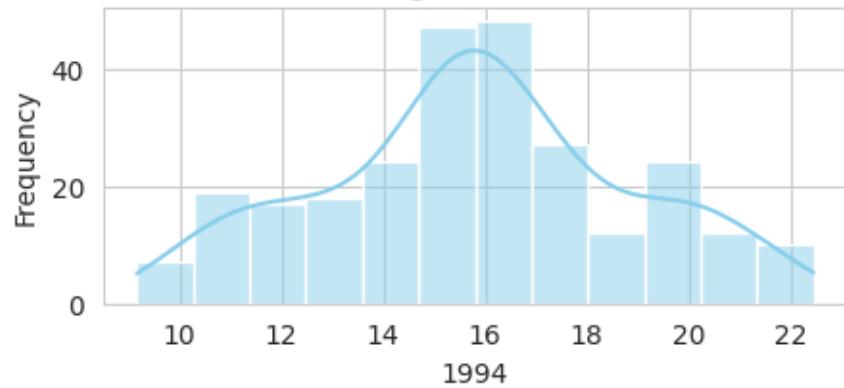




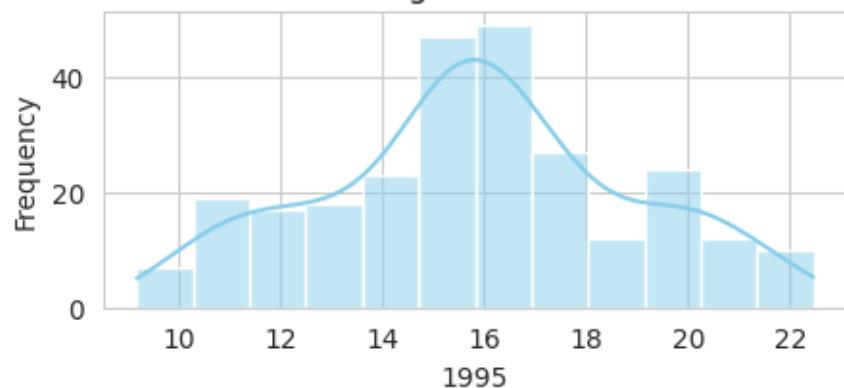
Histogram of 1993

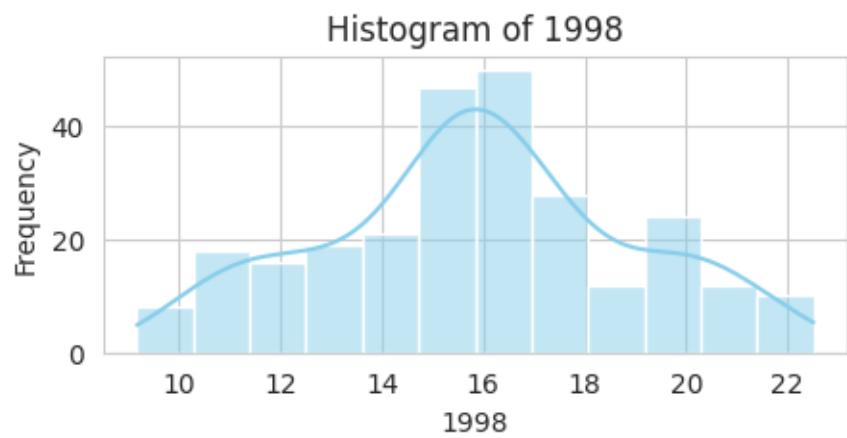
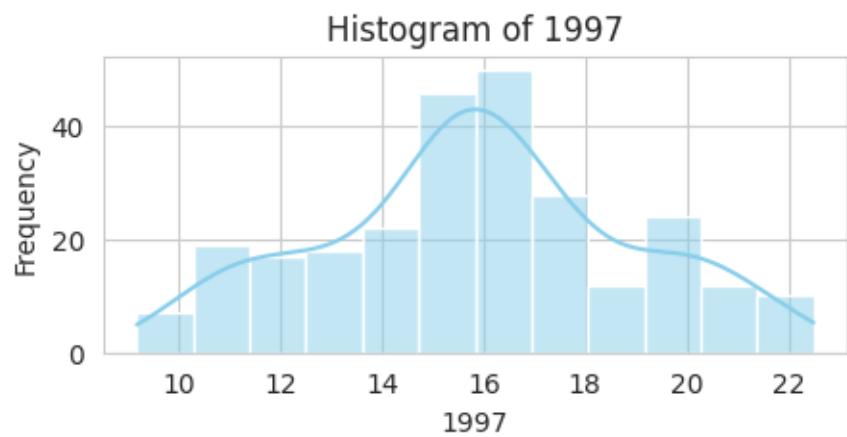
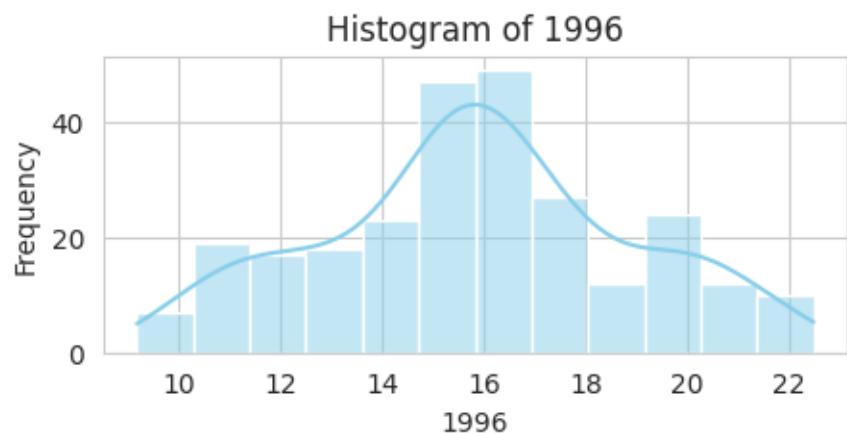


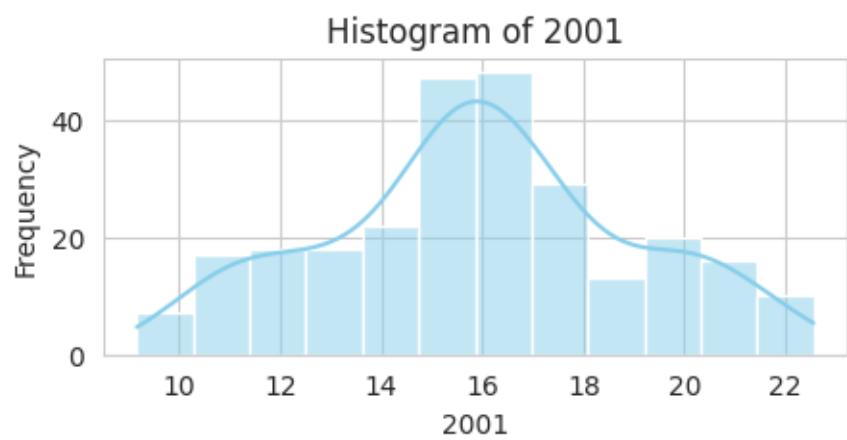
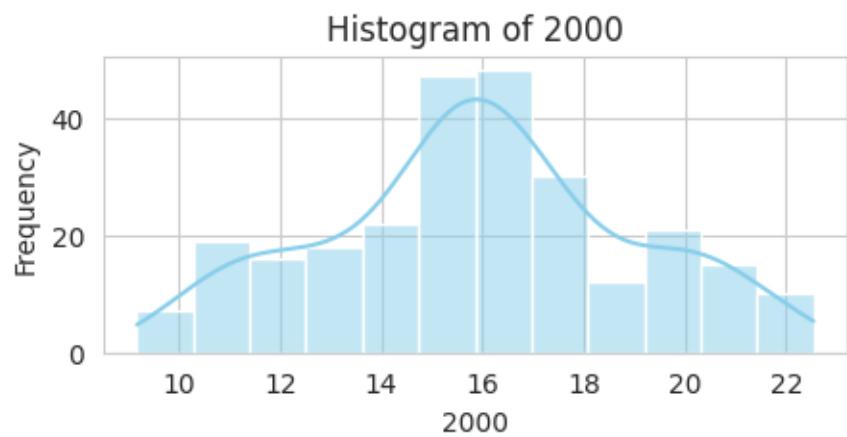
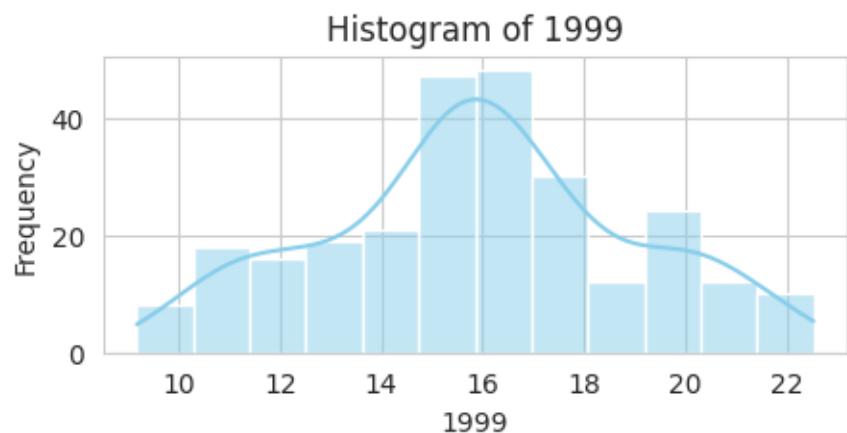
Histogram of 1994



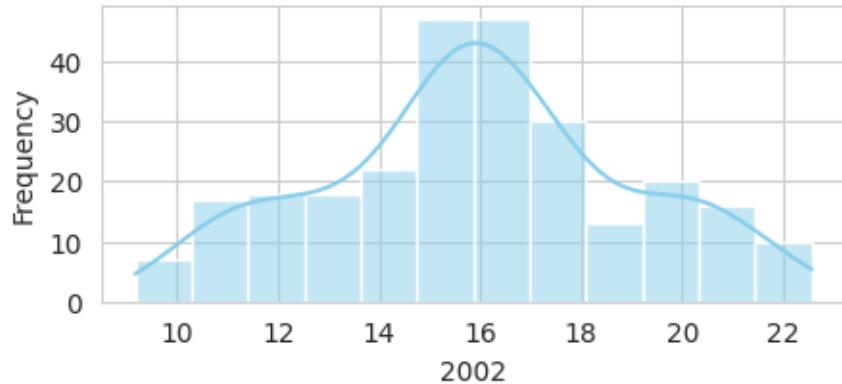
Histogram of 1995



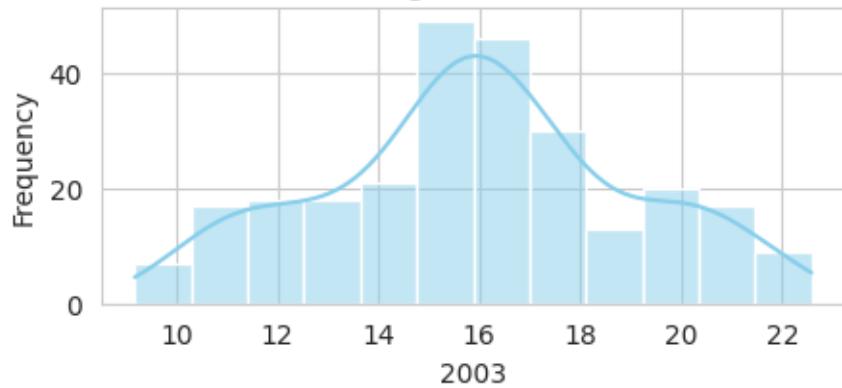




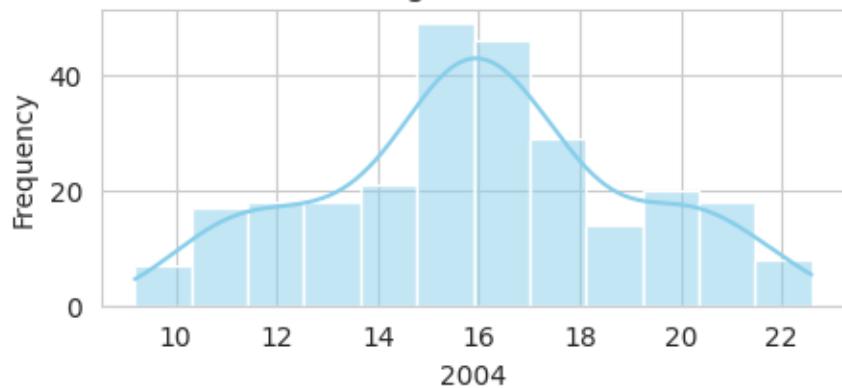
Histogram of 2002

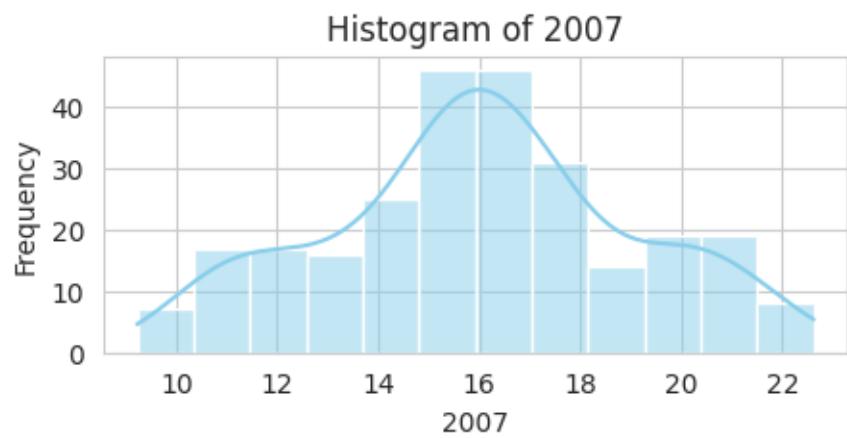
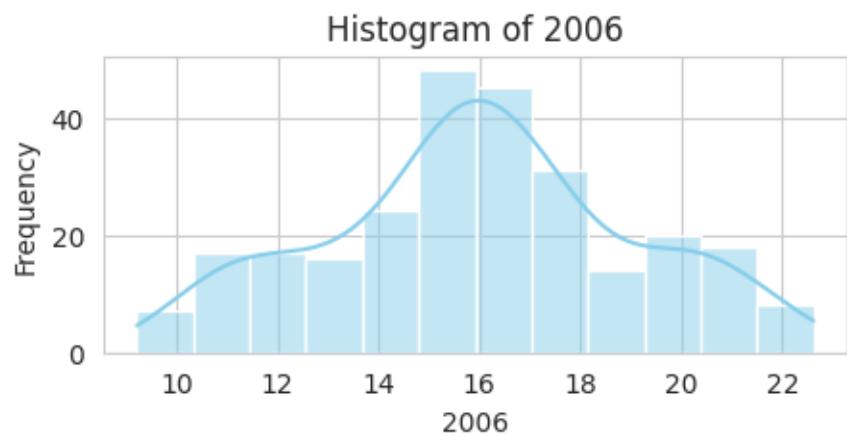
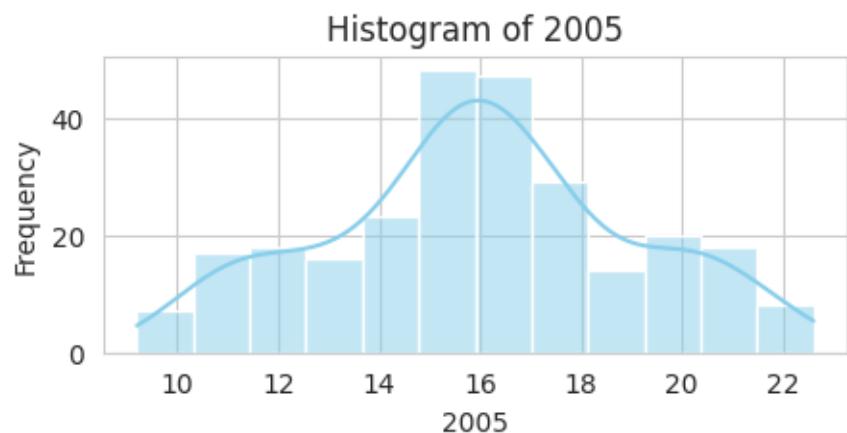


Histogram of 2003

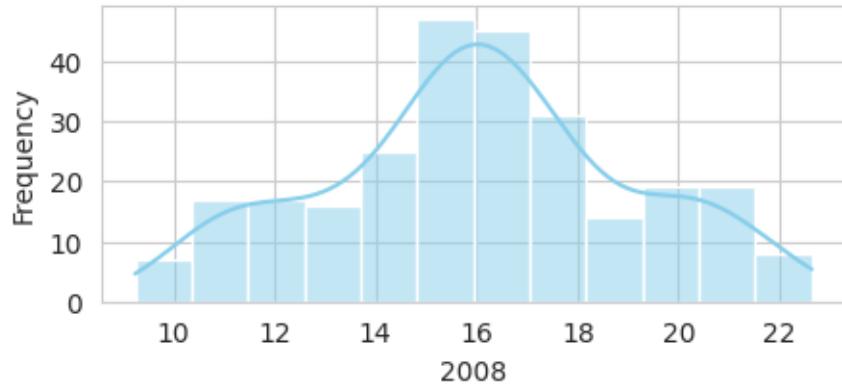


Histogram of 2004

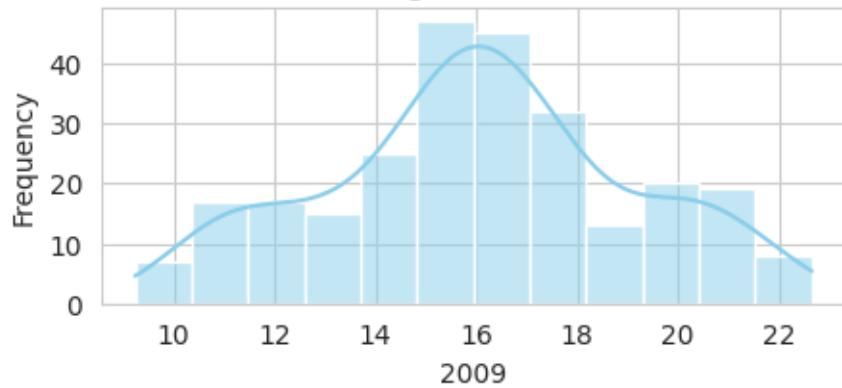




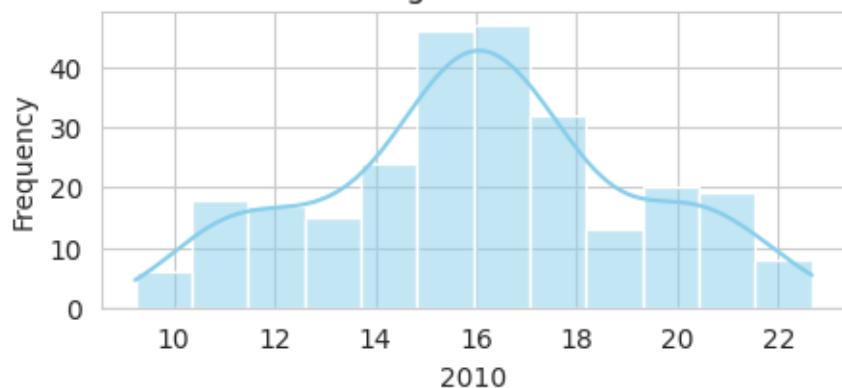
Histogram of 2008



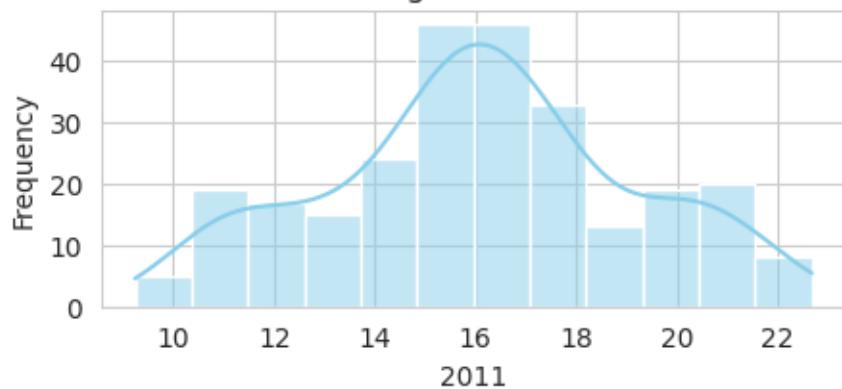
Histogram of 2009



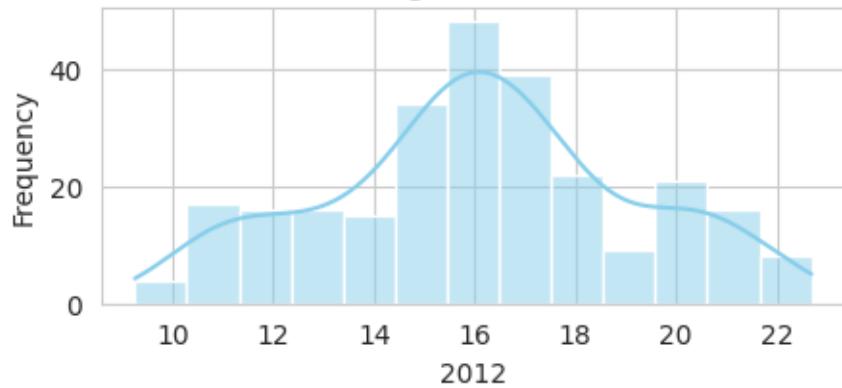
Histogram of 2010



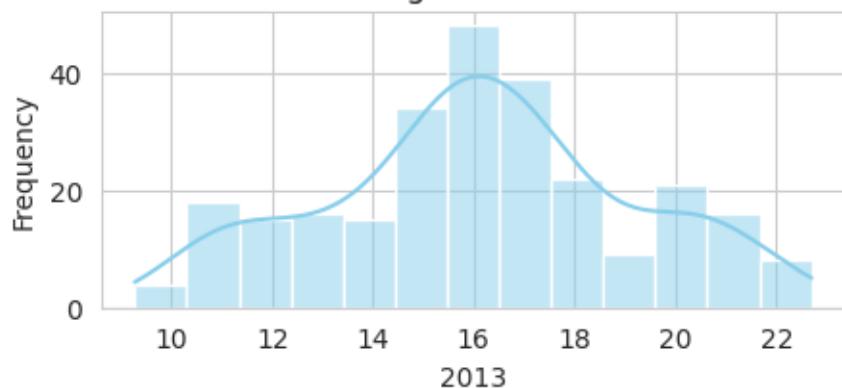
Histogram of 2011

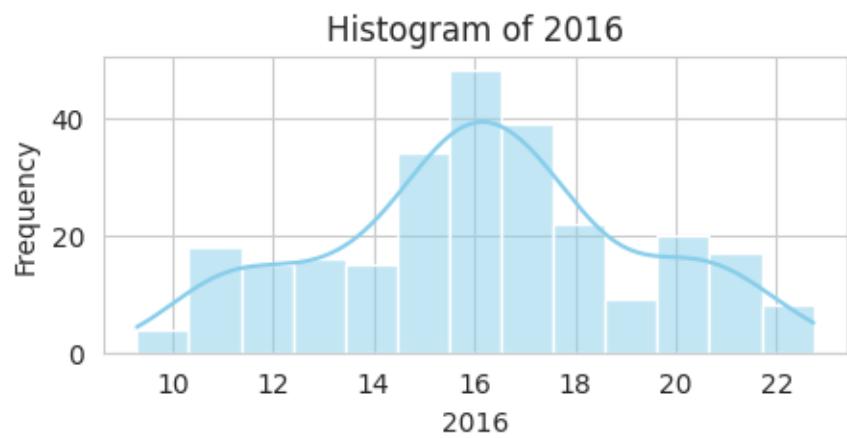
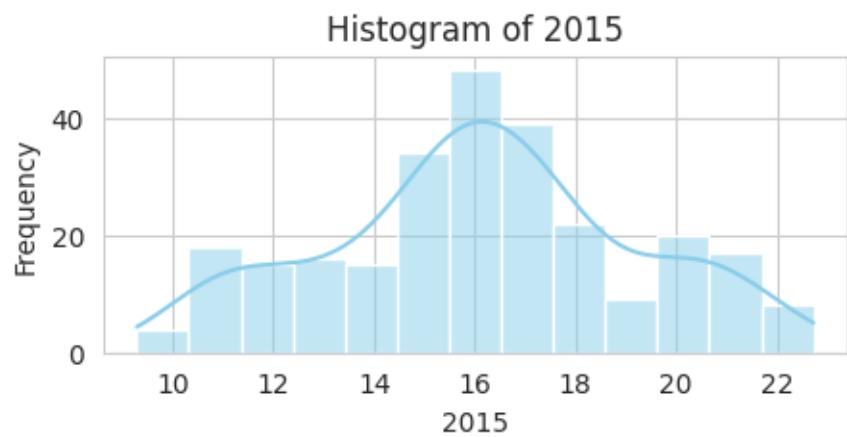
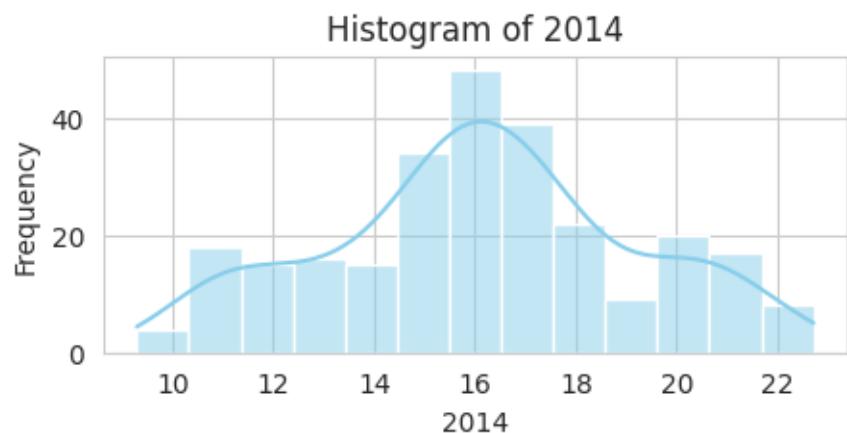


Histogram of 2012

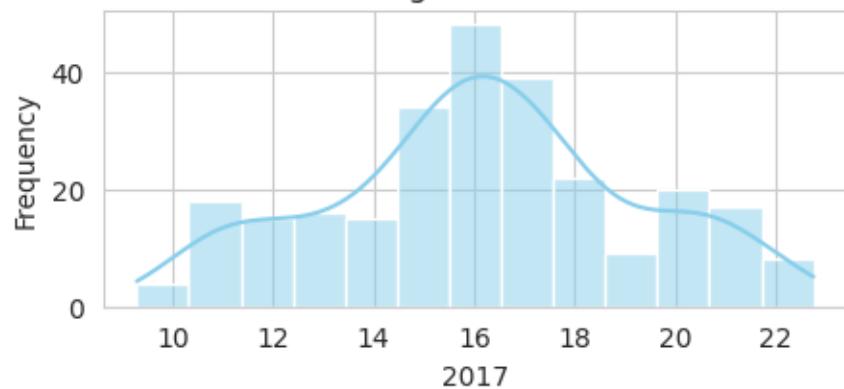


Histogram of 2013

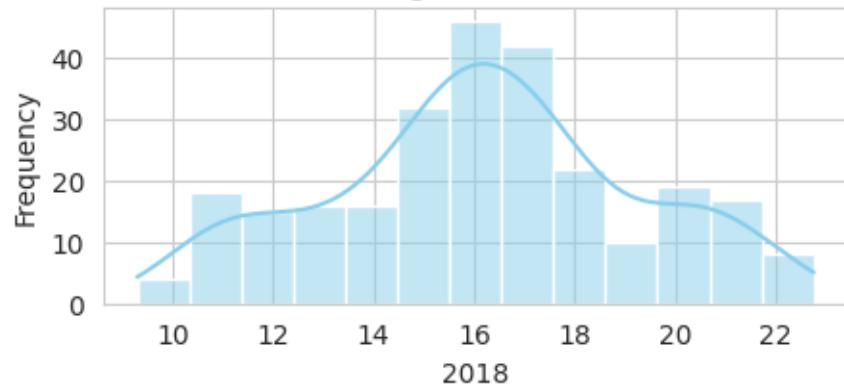




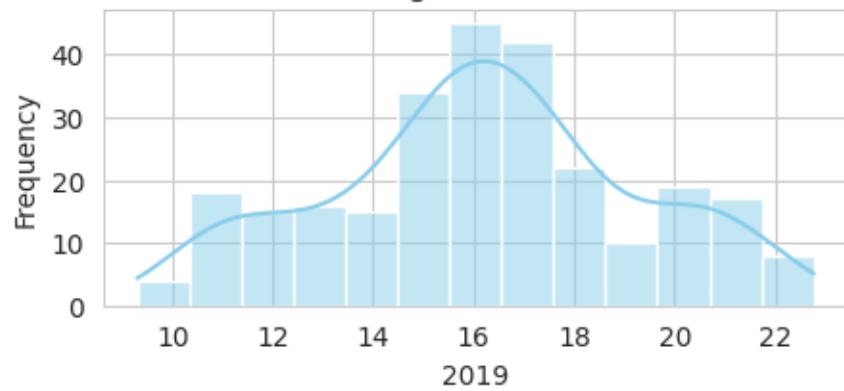
Histogram of 2017



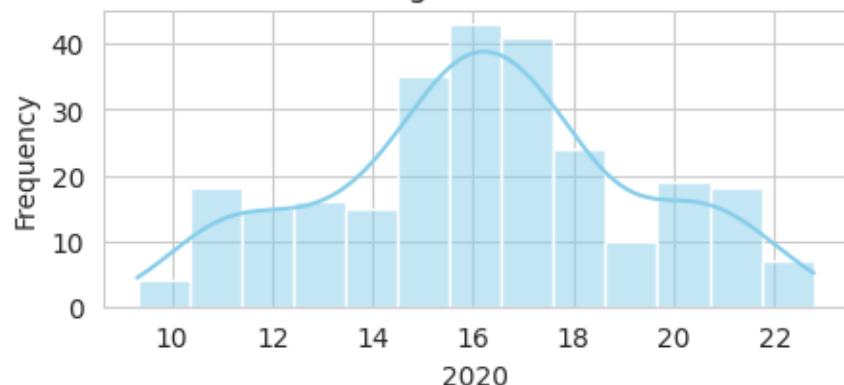
Histogram of 2018



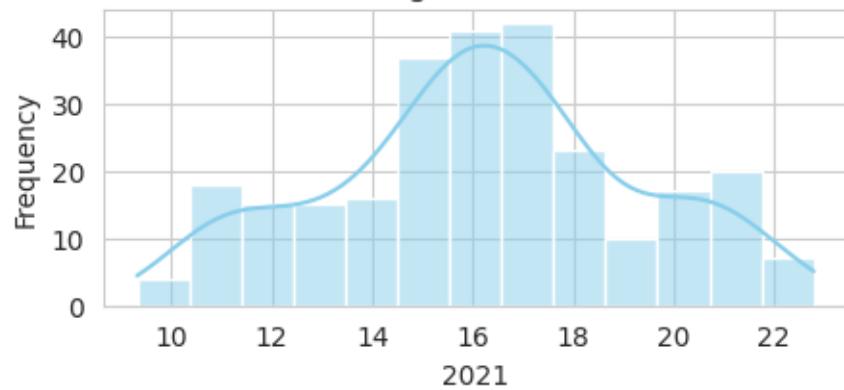
Histogram of 2019



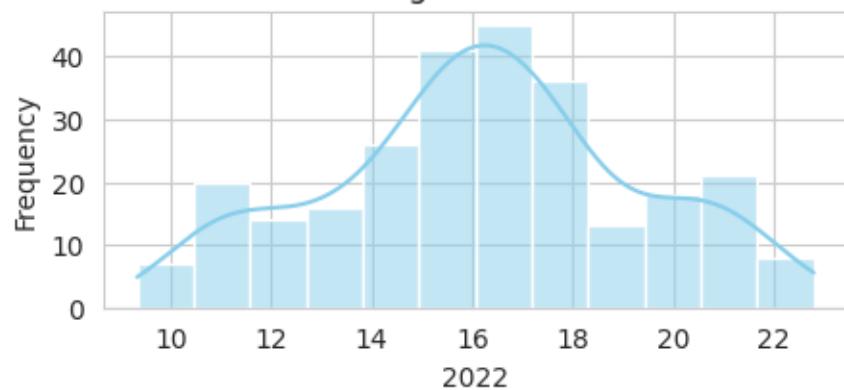
Histogram of 2020



Histogram of 2021

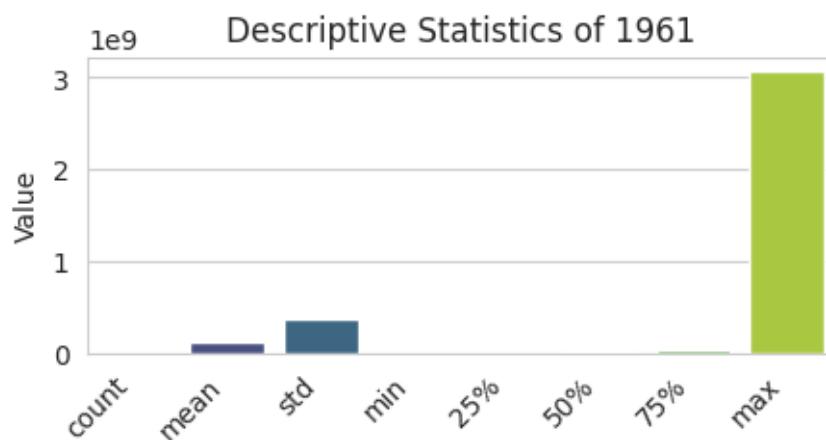
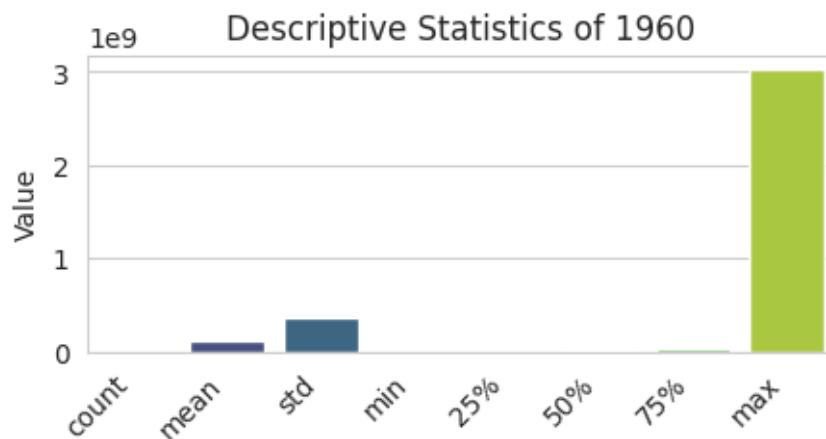


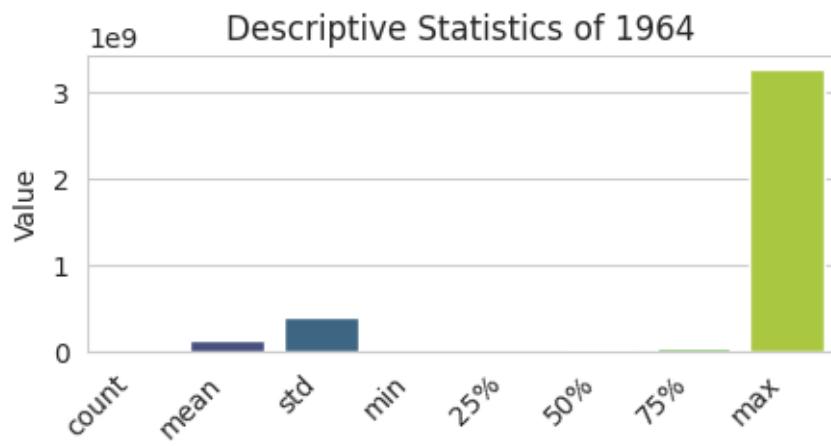
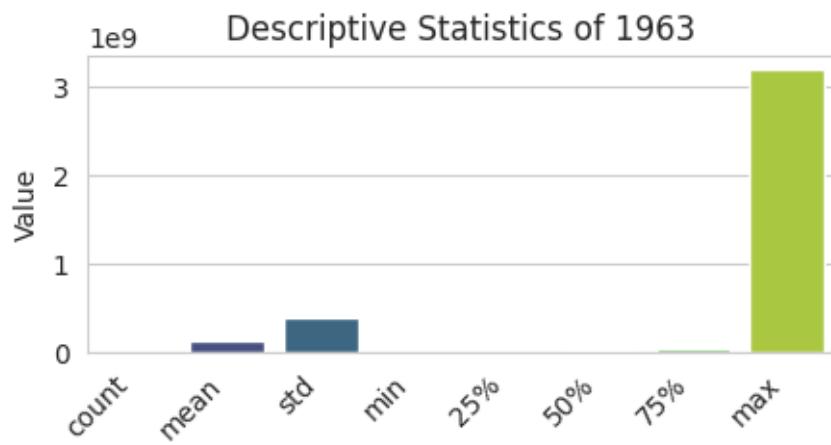
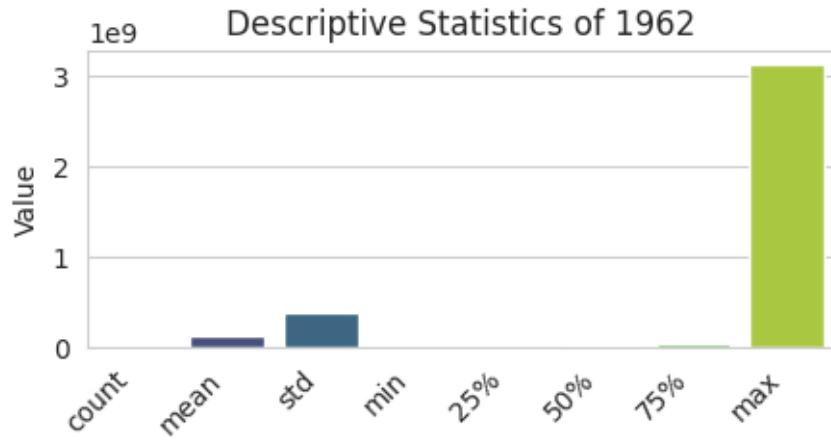
Histogram of 2022

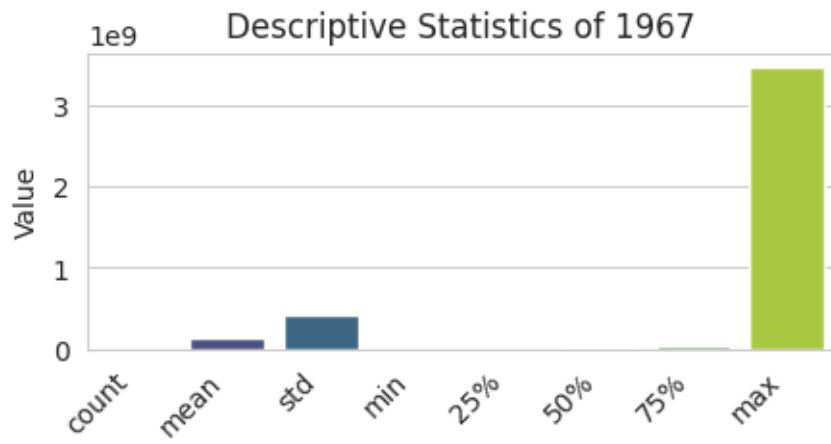
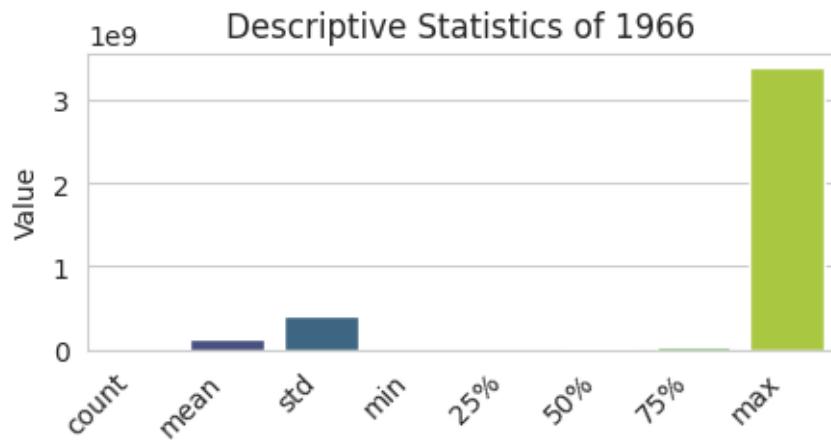
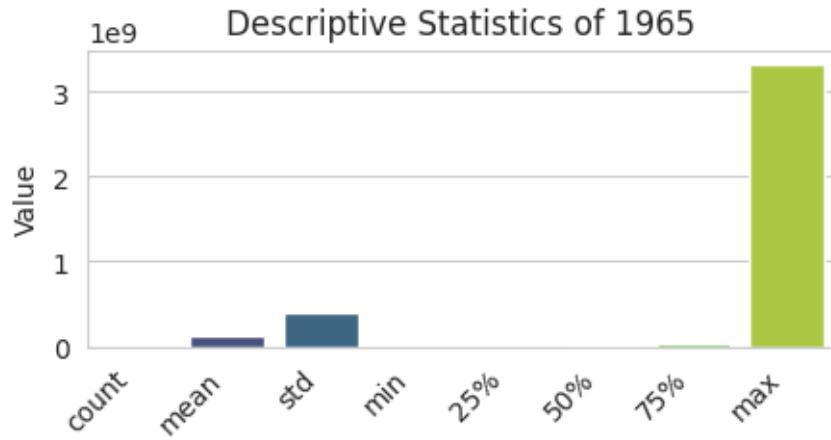


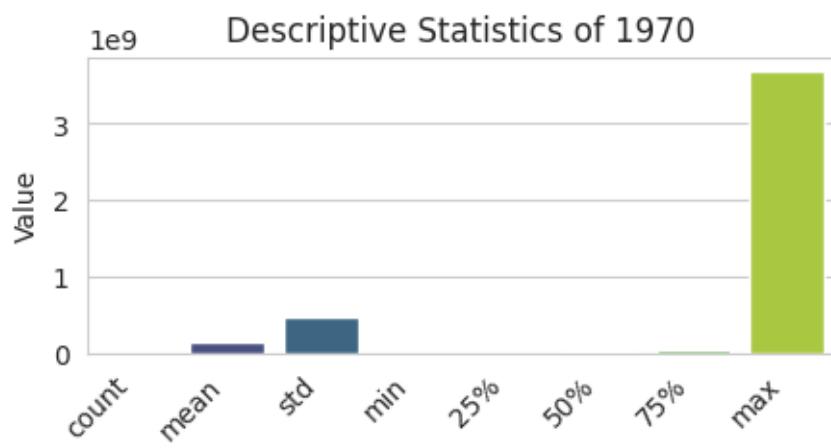
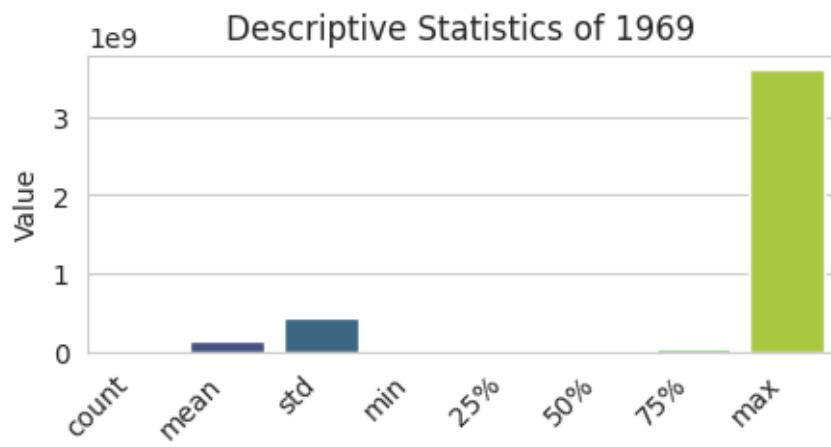
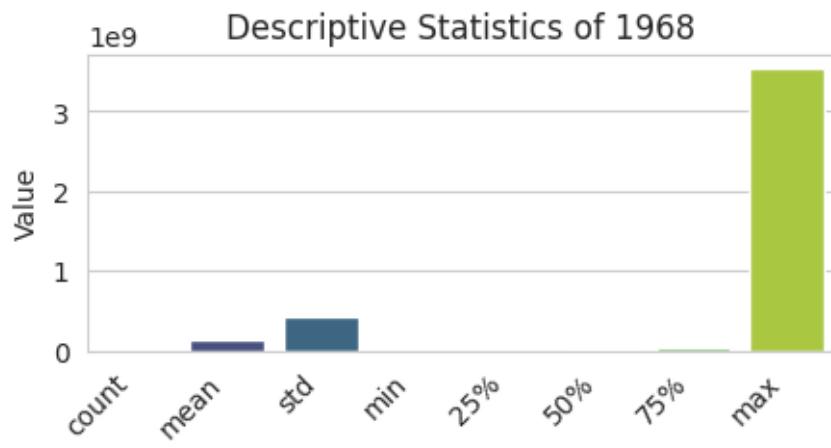
## BAR CHART FOR DESCRIPTIVE STATISTICS

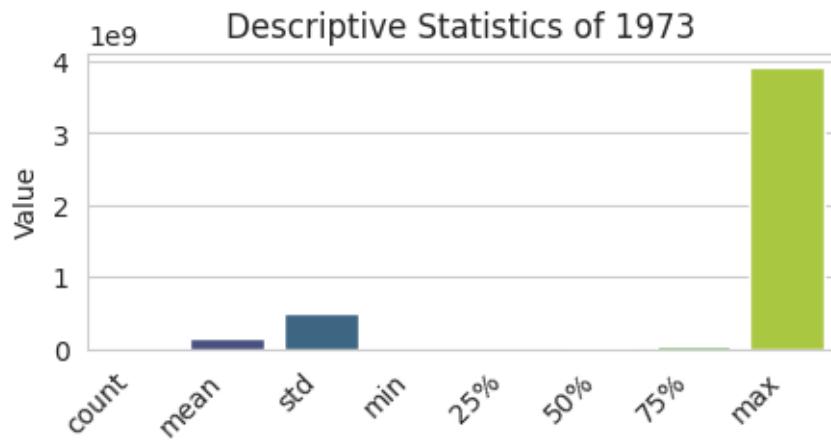
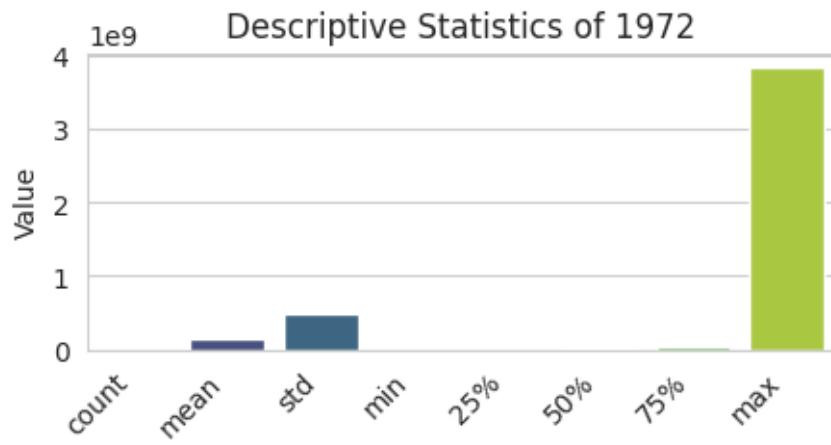
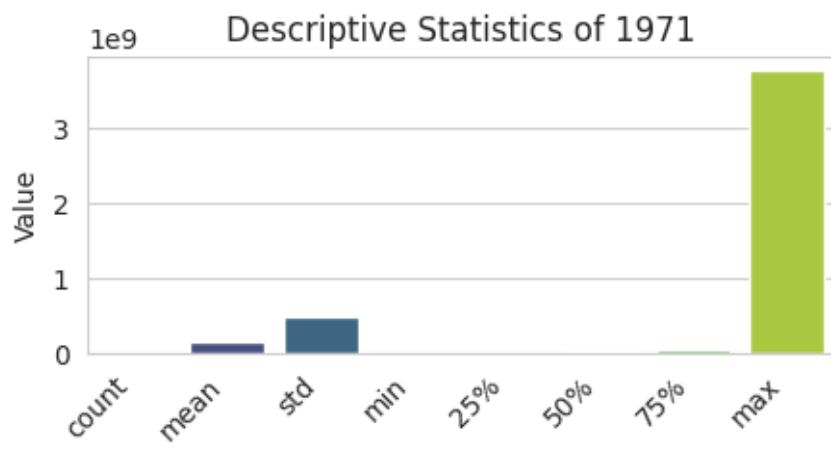
```
[ ]: for column in df.select_dtypes(include='number').columns:  
    # Bar chart for descriptive statistics  
    summary_stats = df[column].describe()  
    plt.figure(figsize=(5,2))  
    sb.barplot(x=summary_stats.index, y=summary_stats.values, palette='viridis')  
    plt.title(f'Descriptive Statistics of {column}')  
    plt.ylabel('Value')  
    plt.xticks(rotation=45, ha='right')  
    plt.show()
```

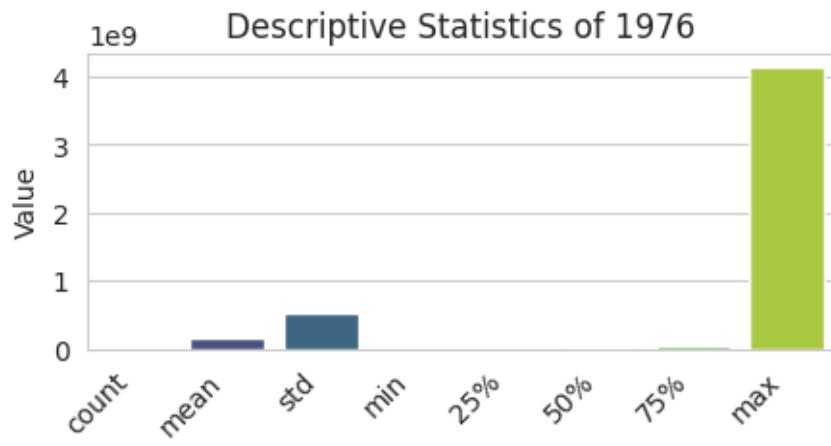
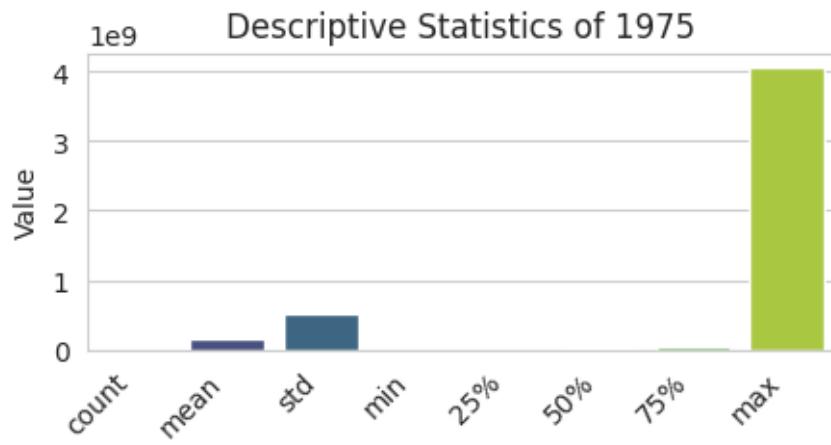
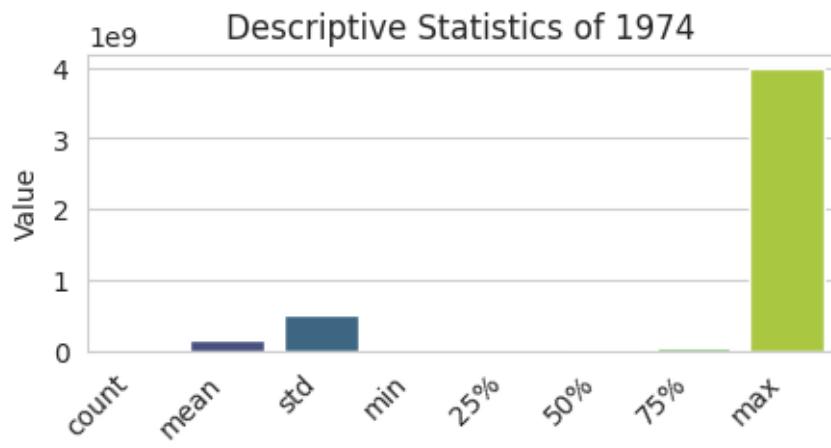


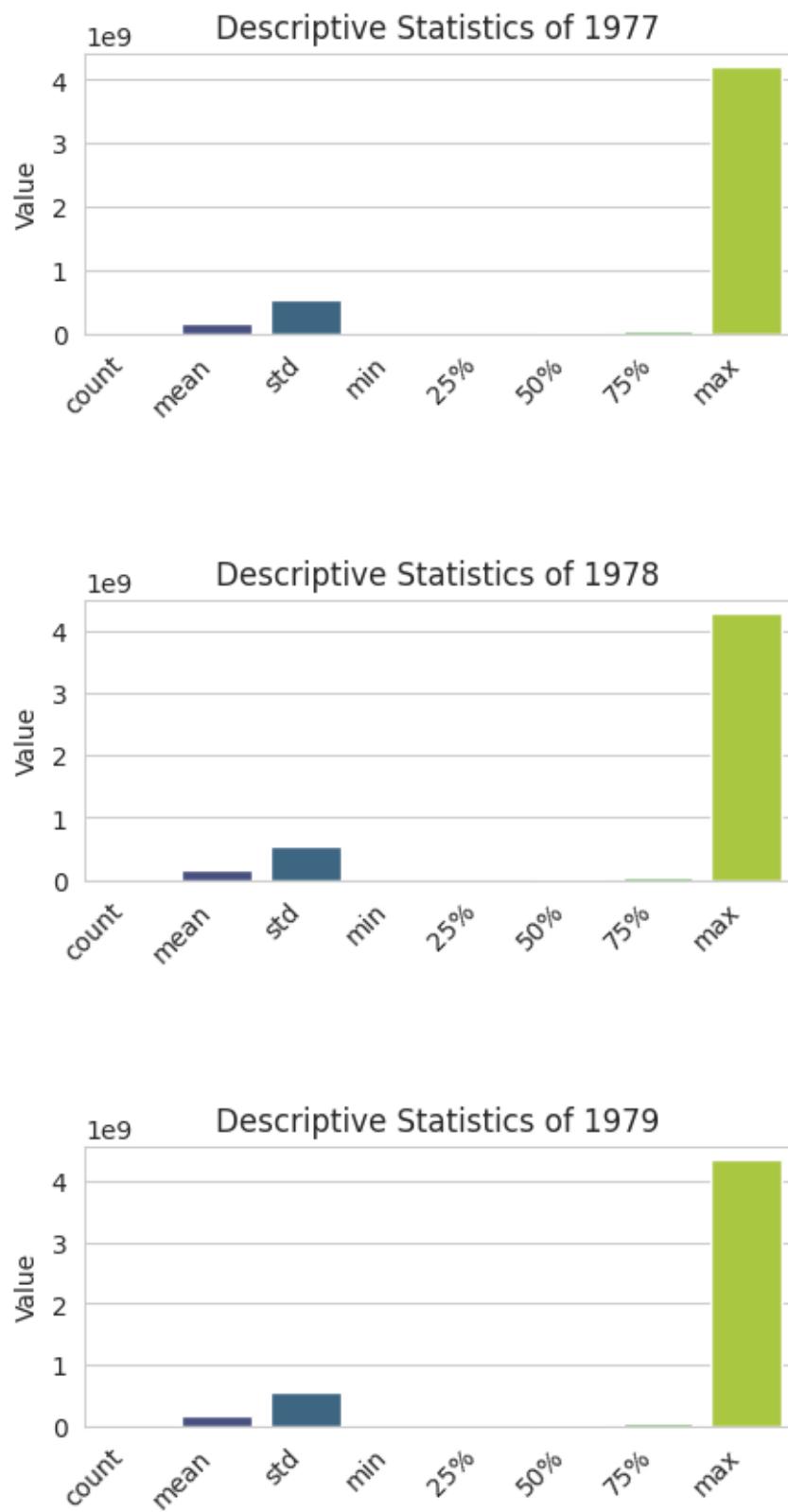


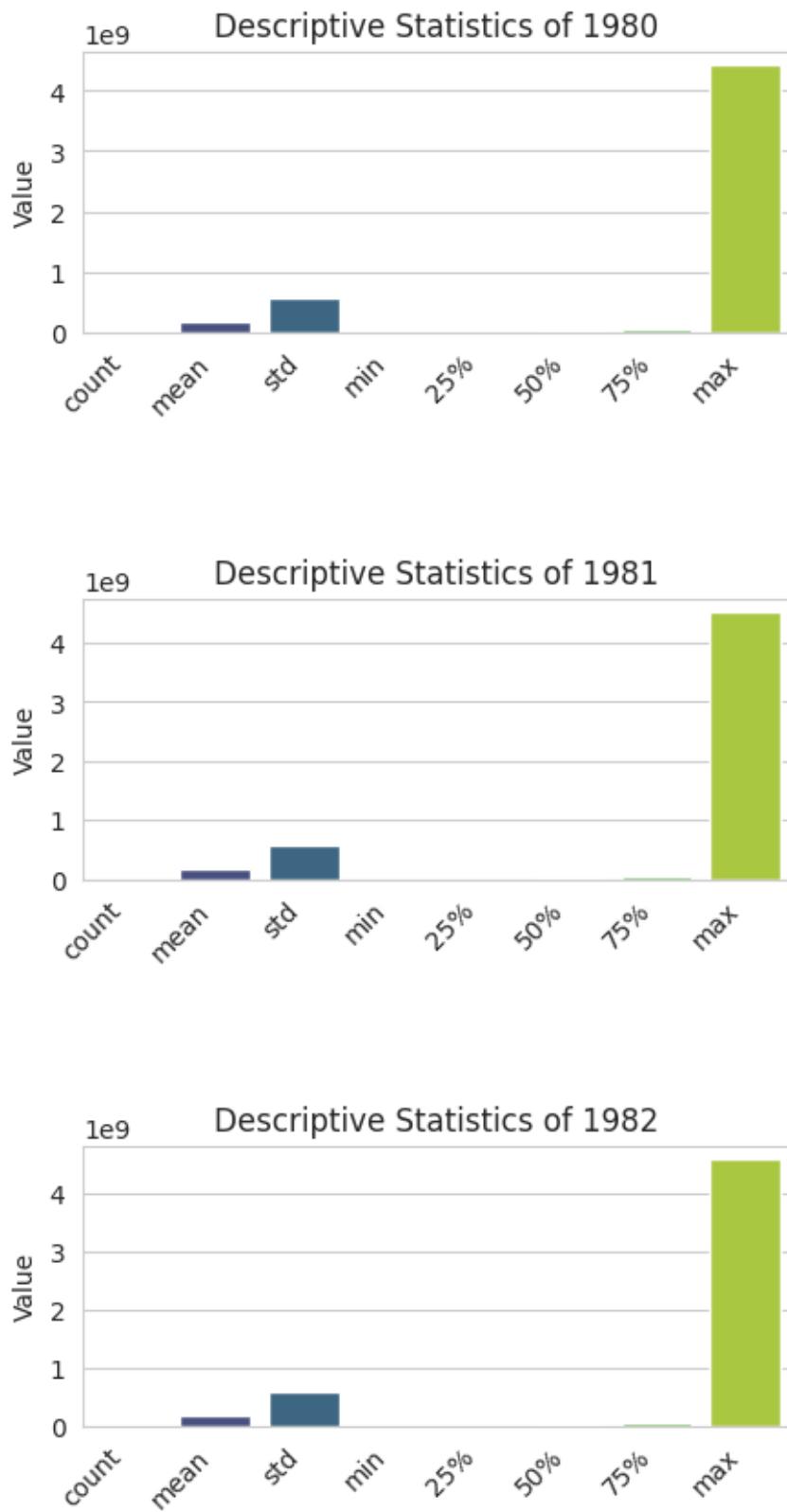


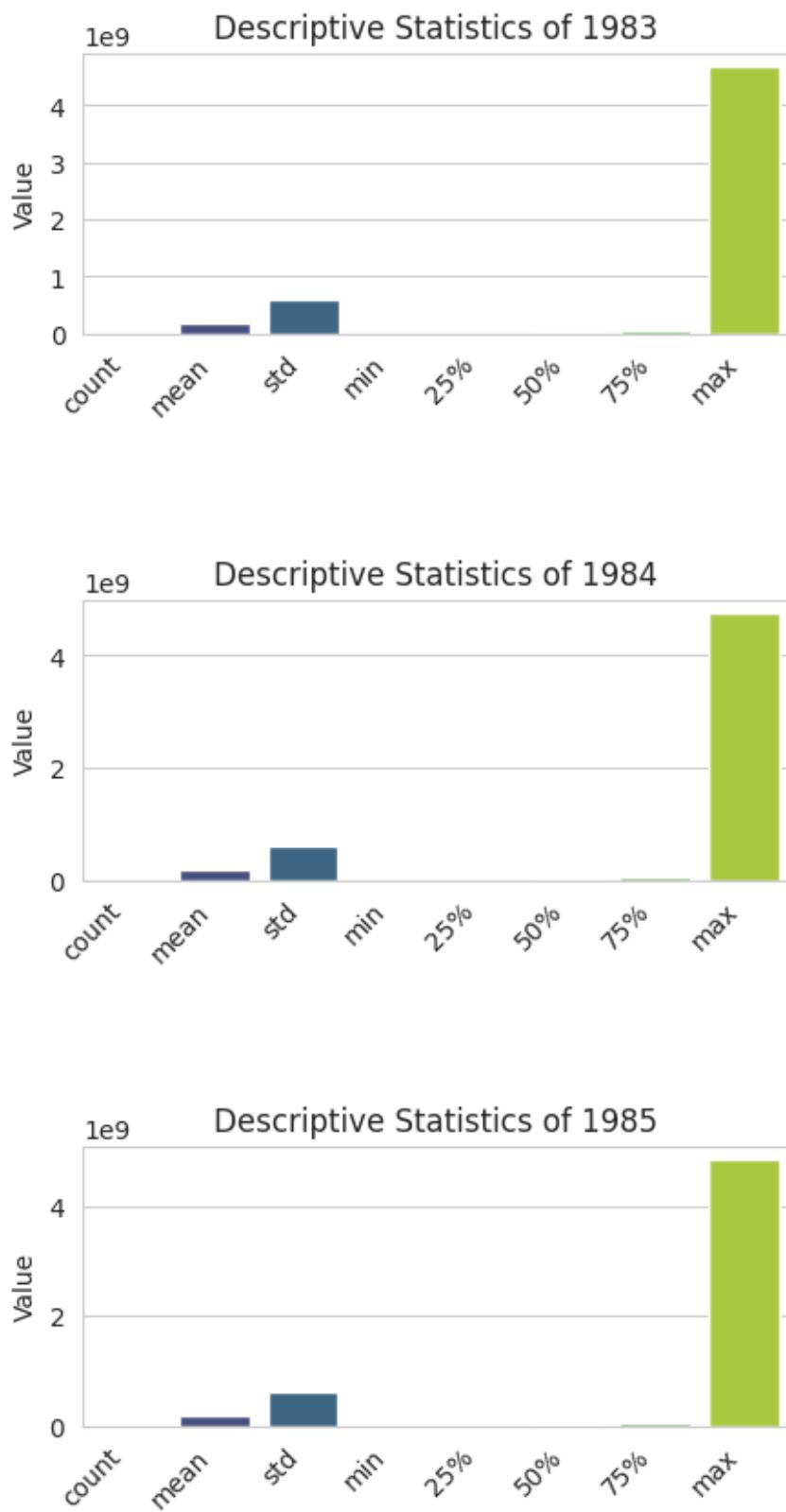


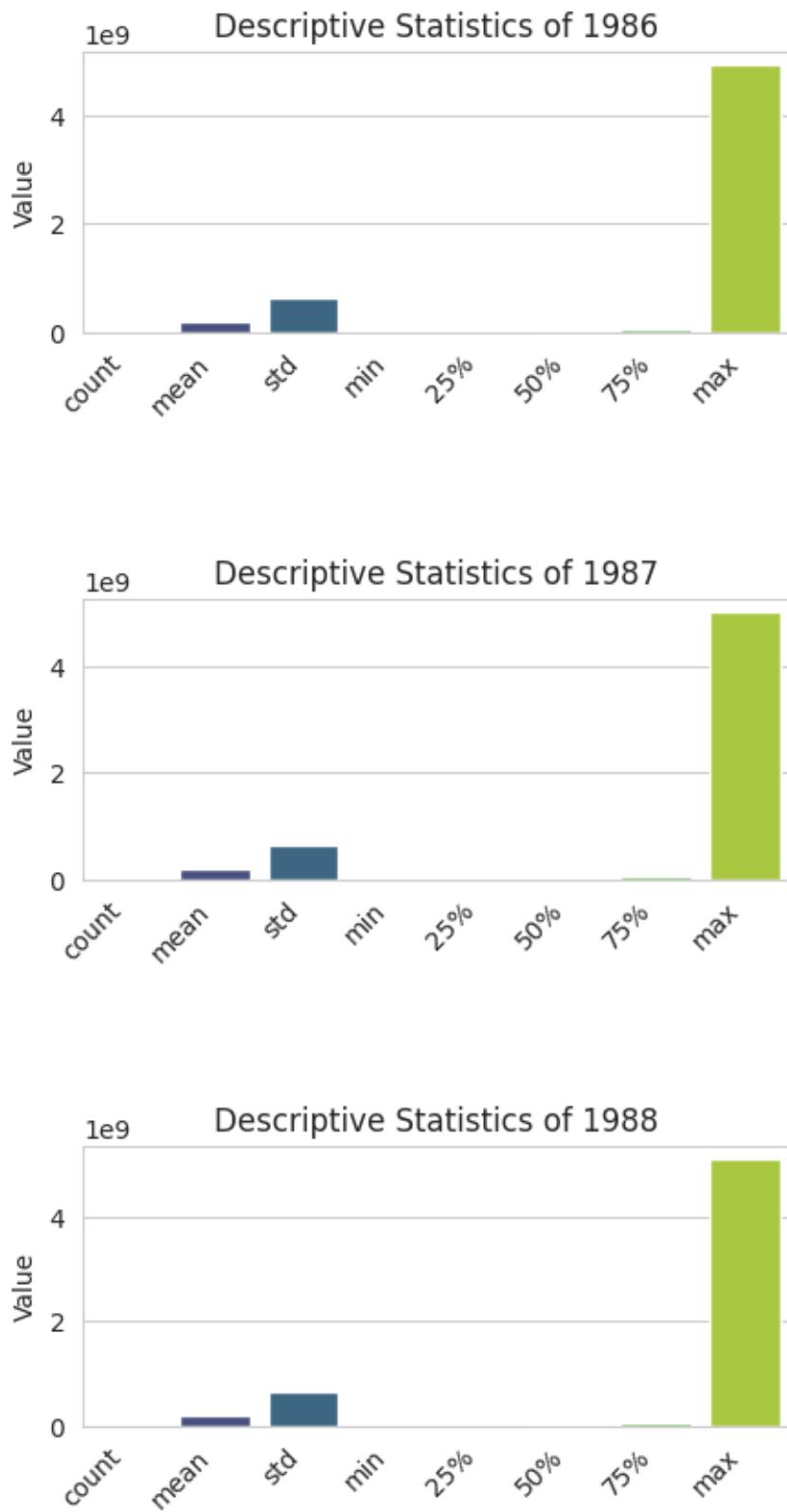


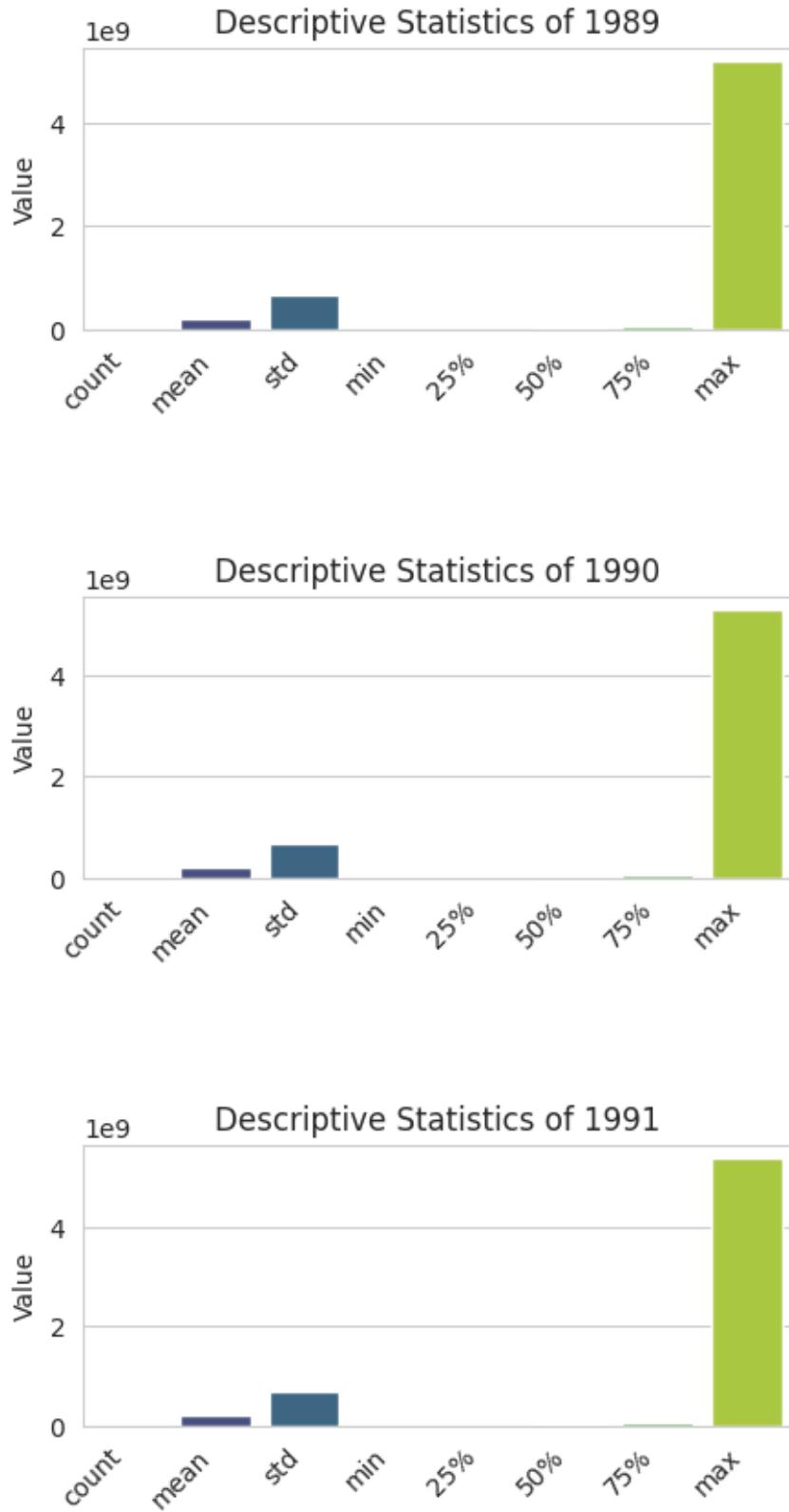


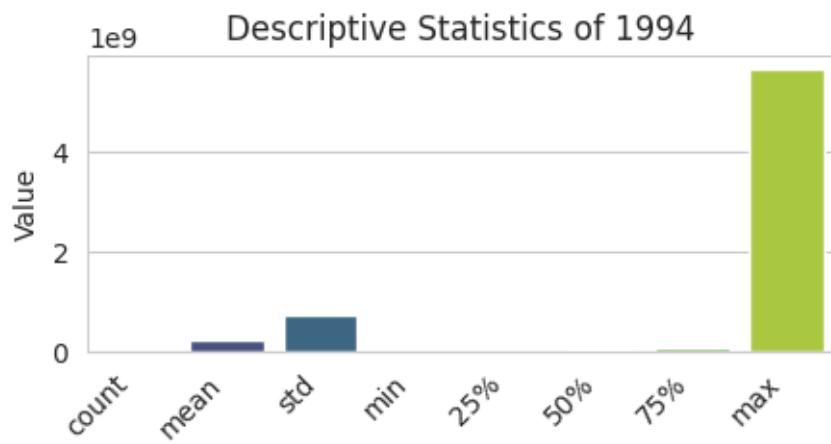
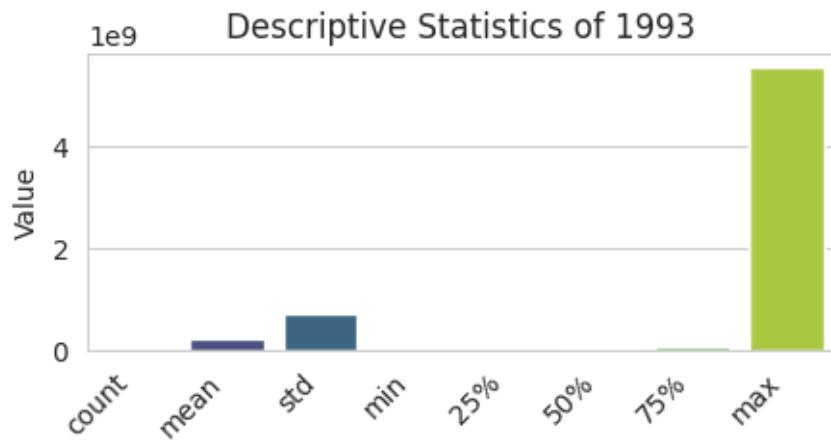
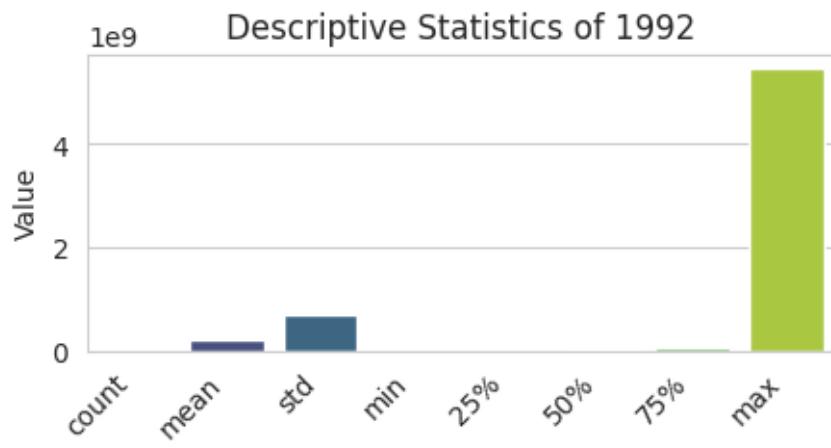


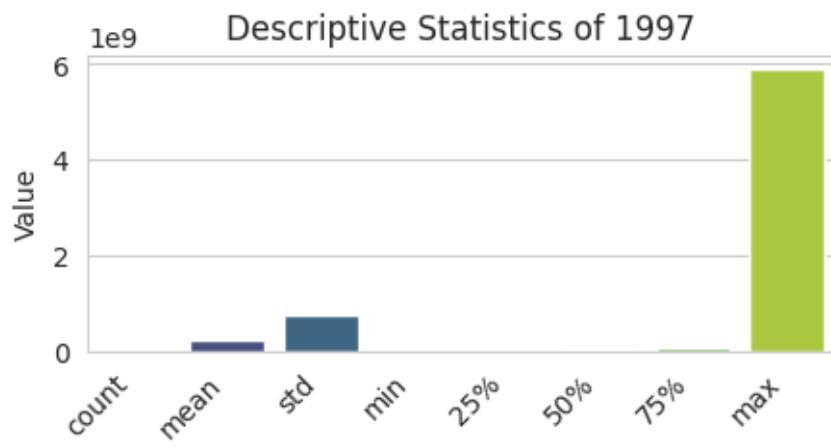
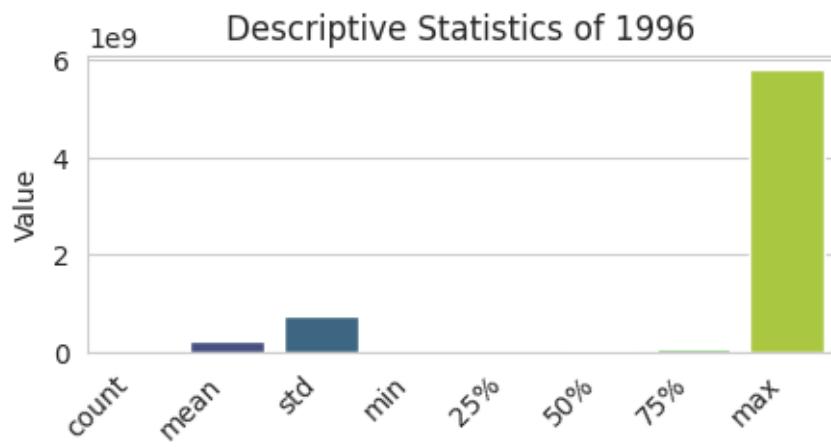
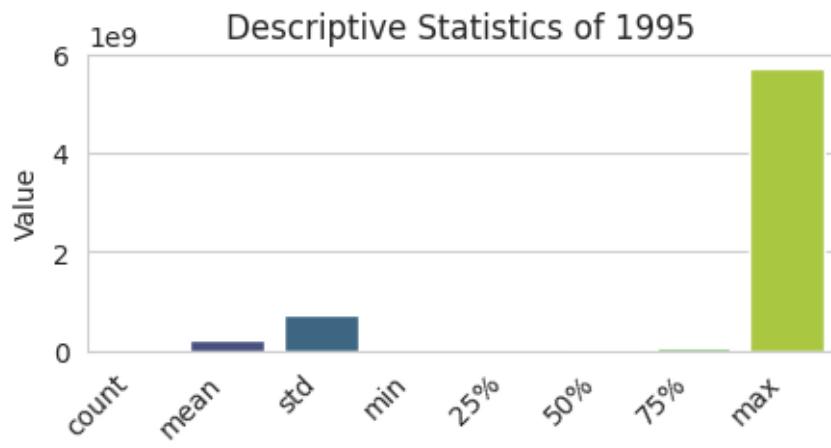


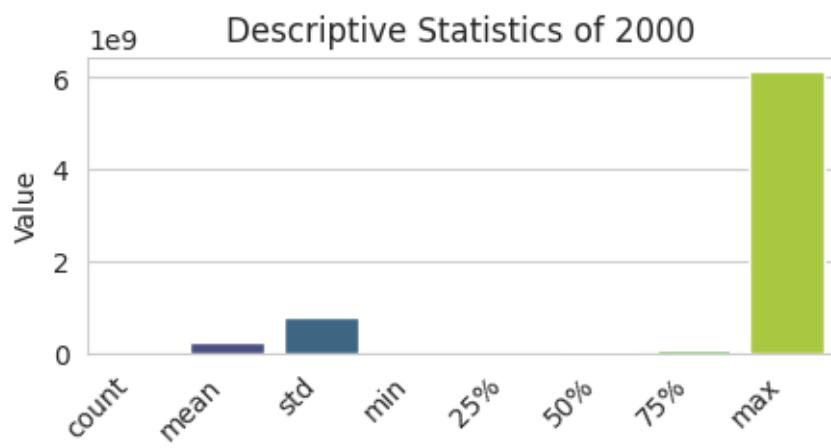
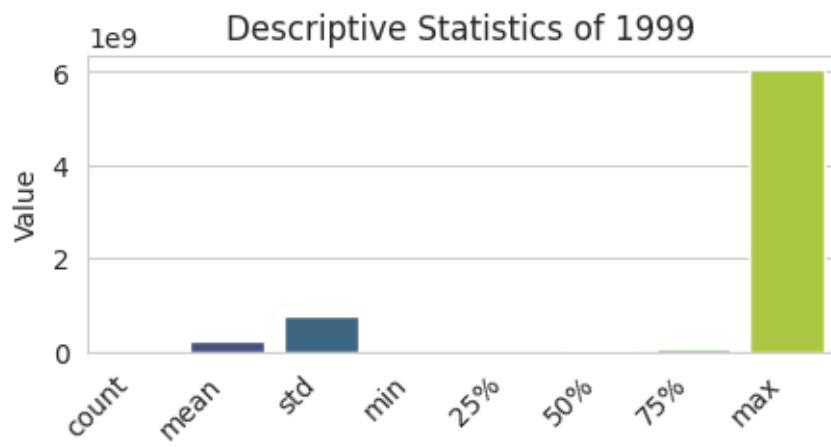
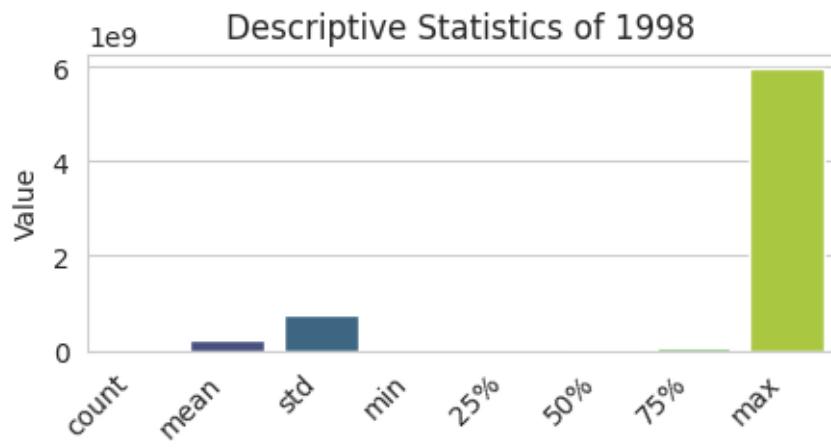


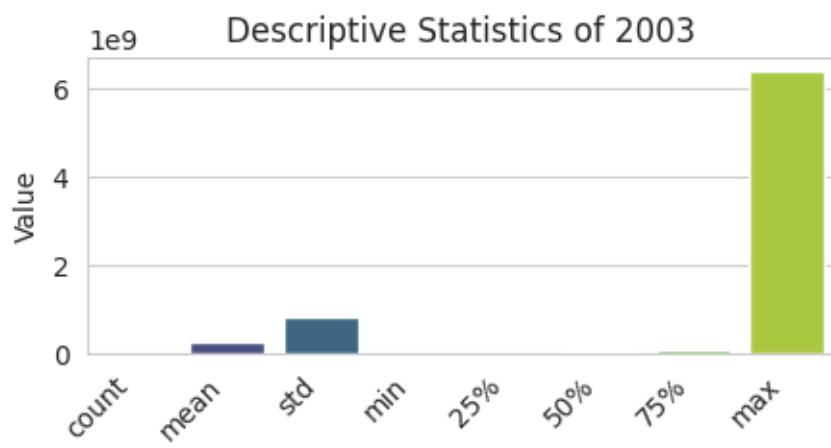
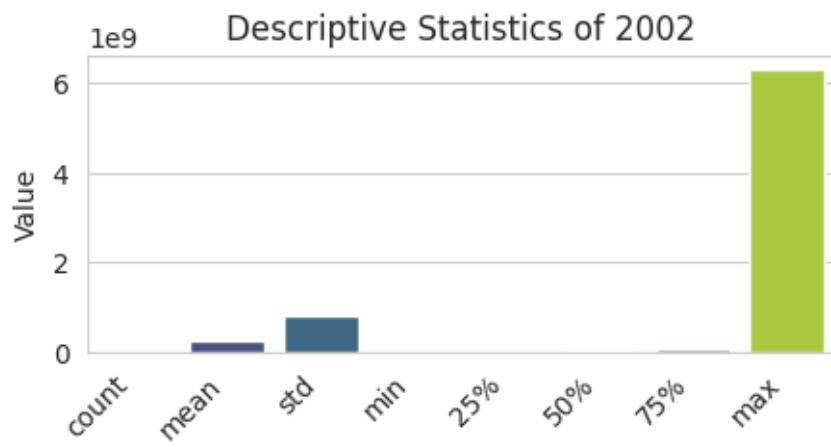
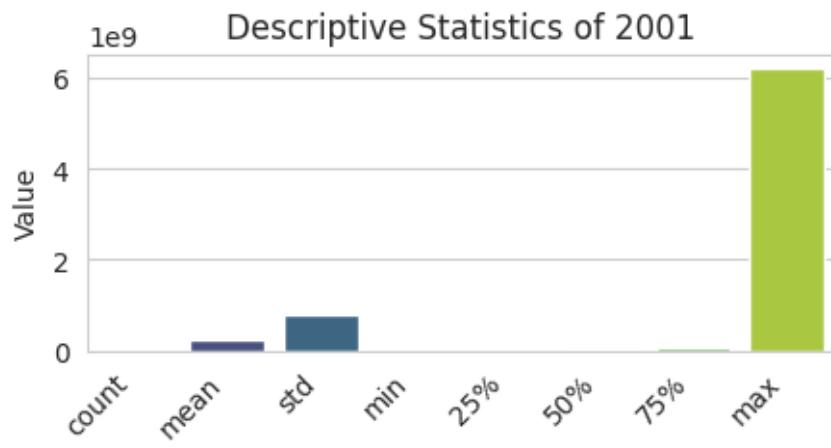


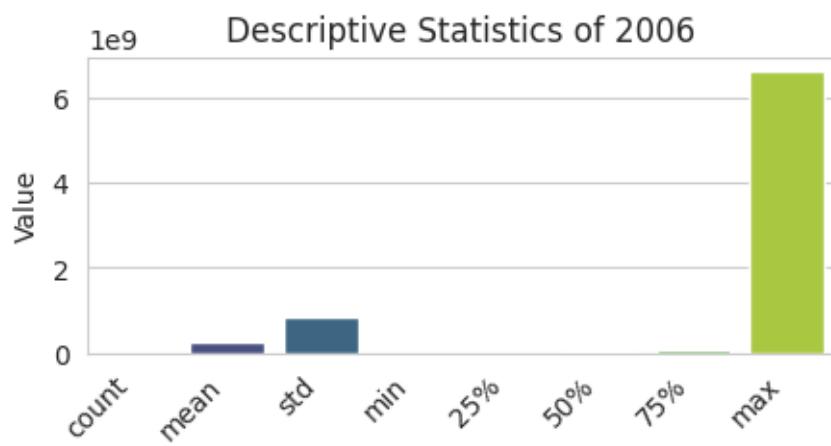
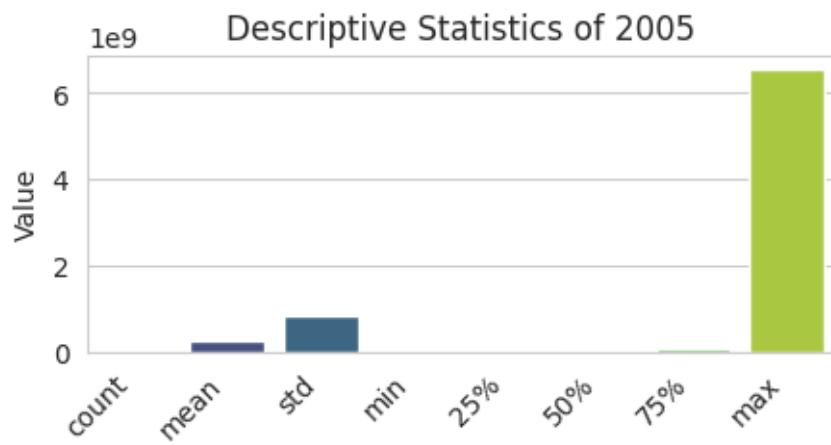
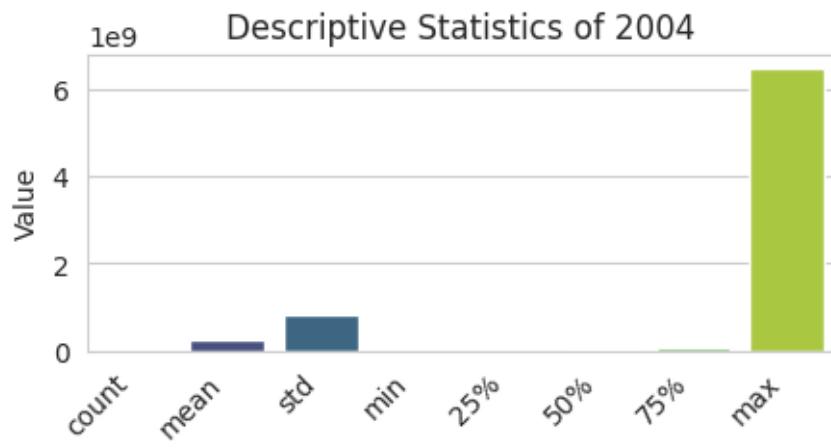


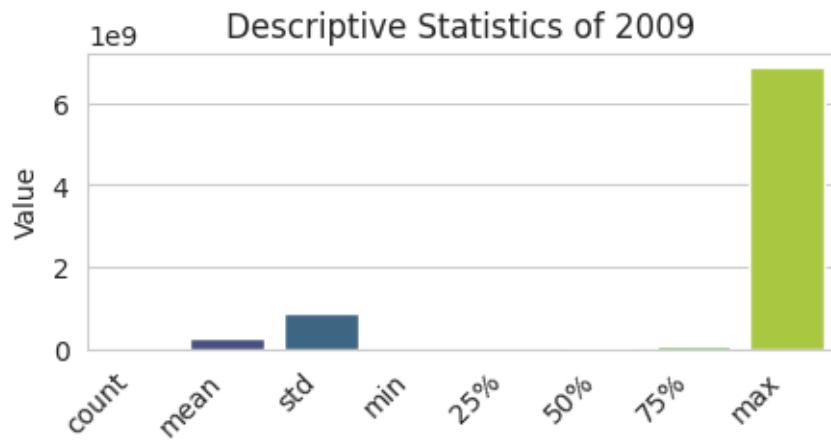
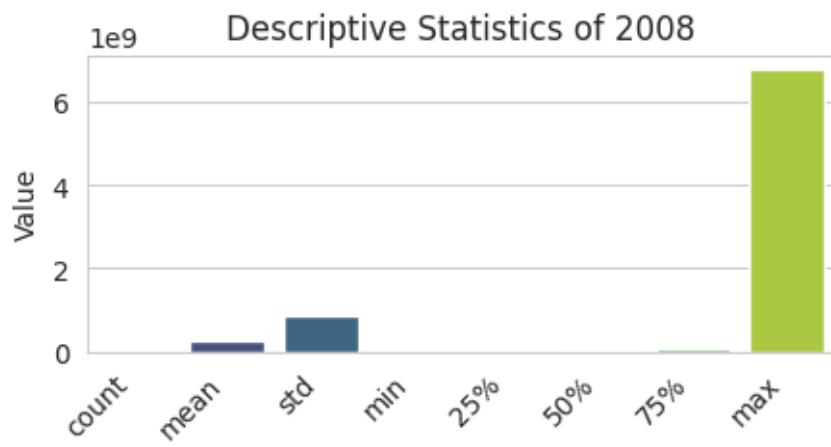
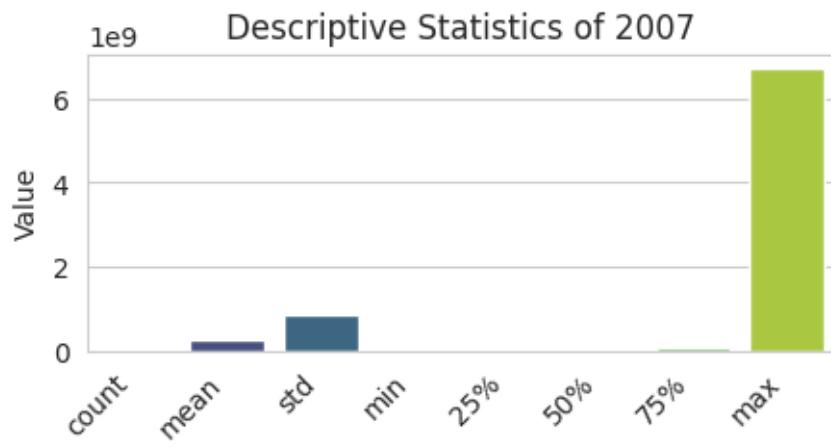


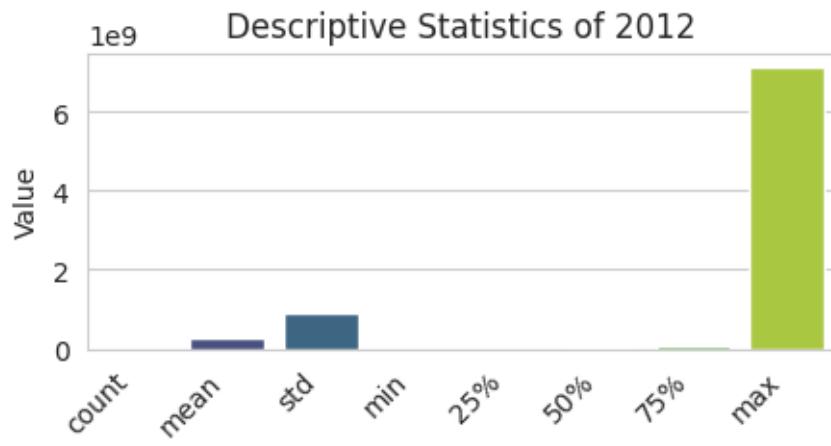
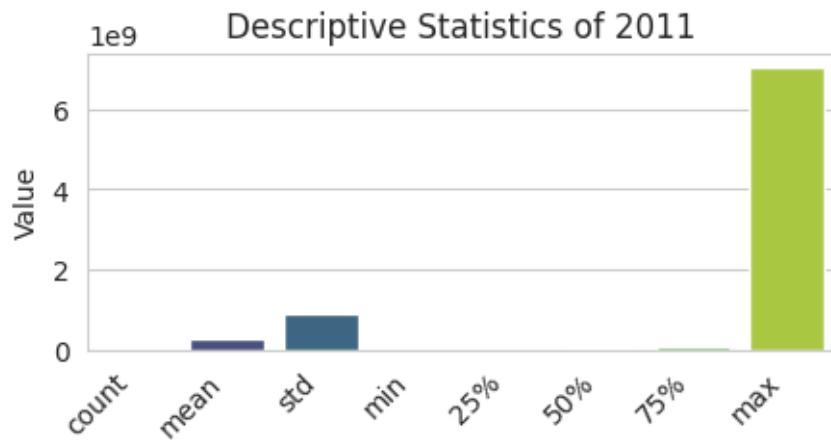
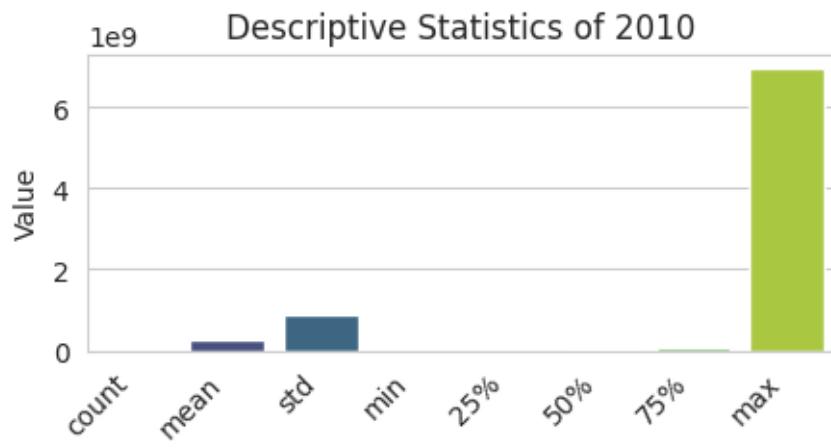


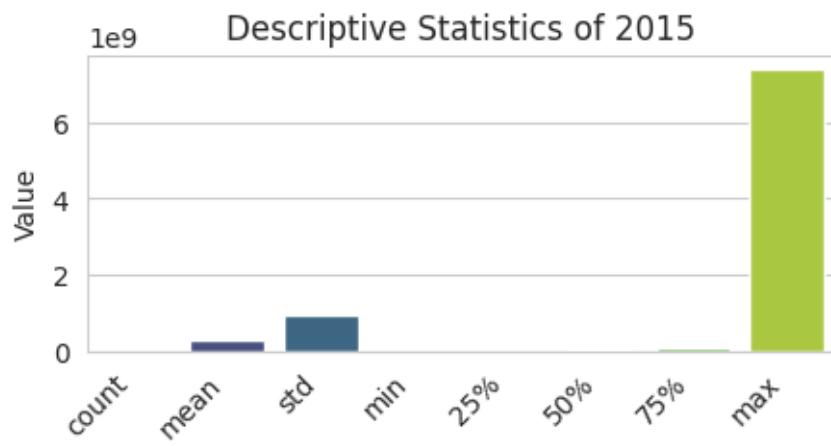
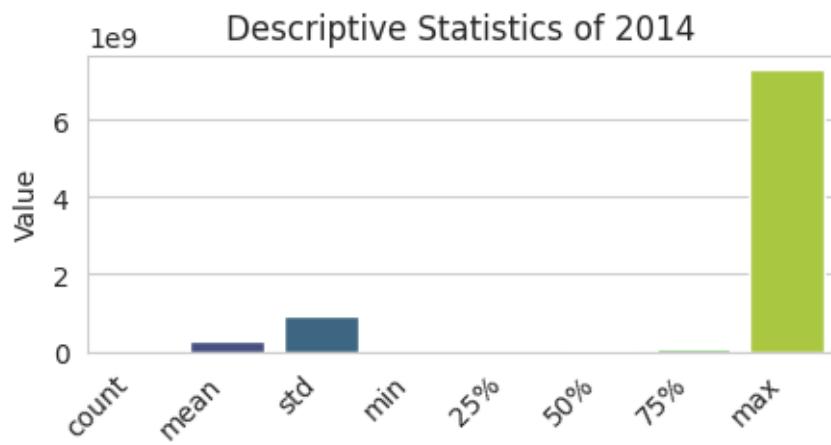
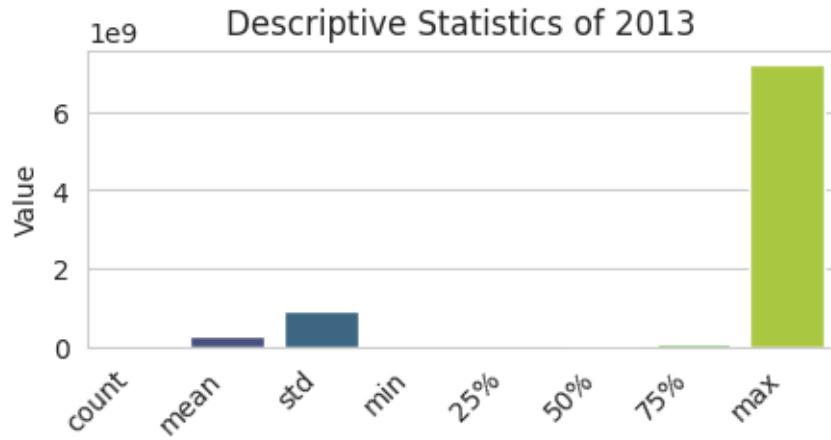


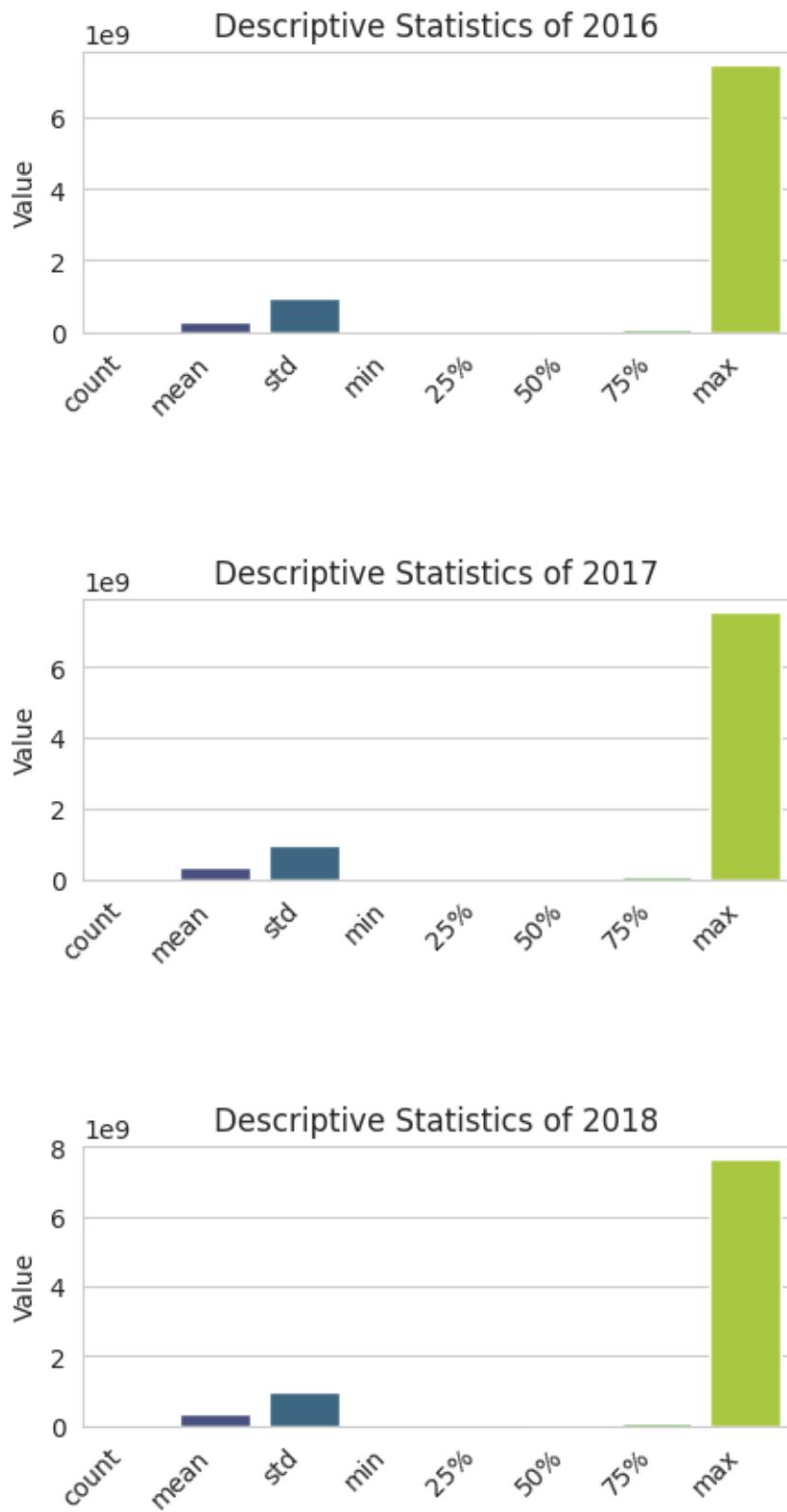


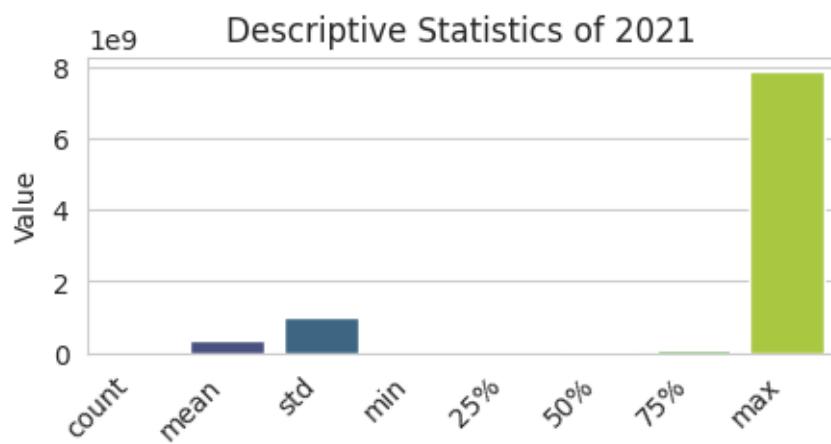
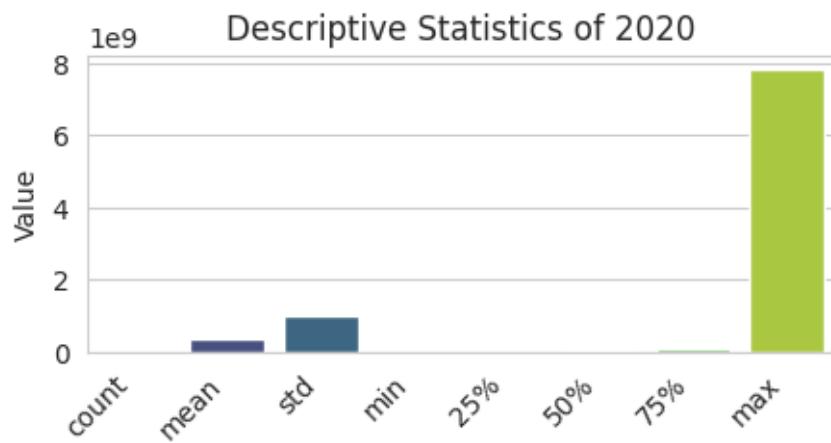
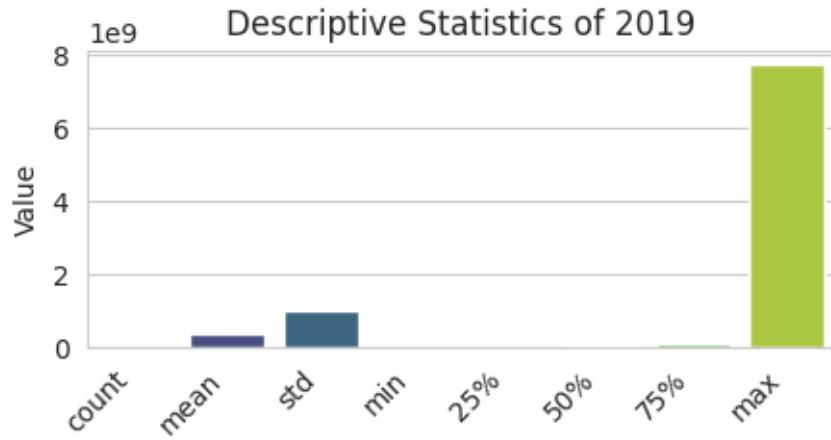


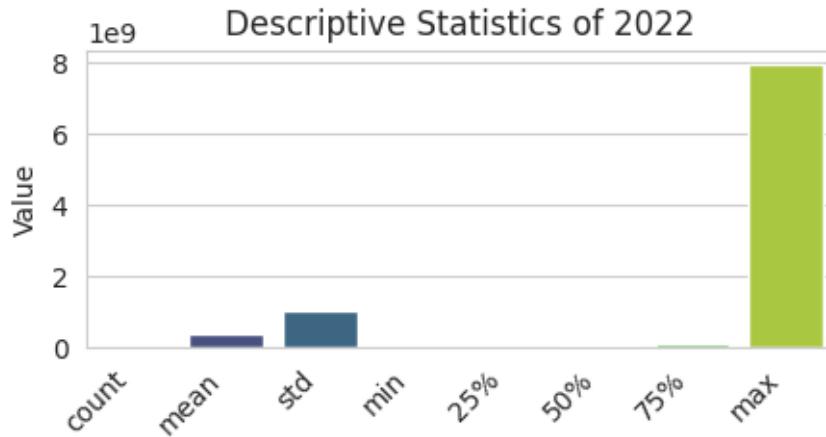












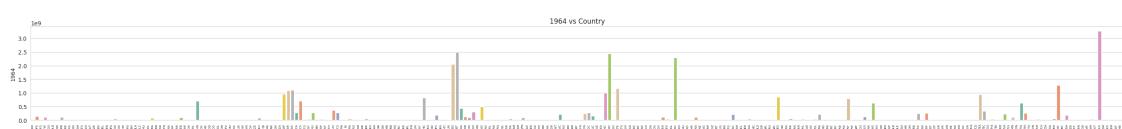
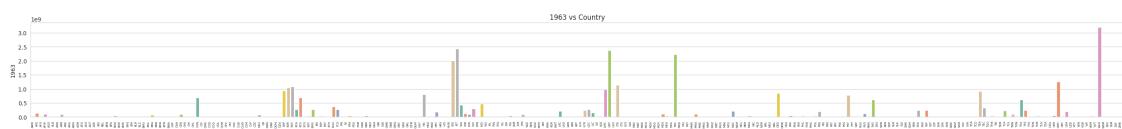
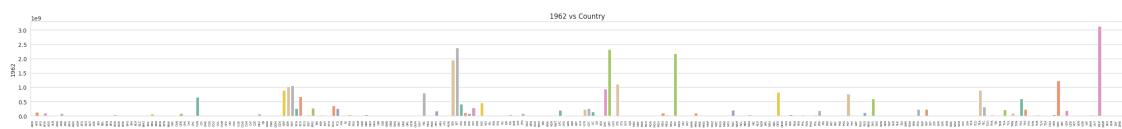
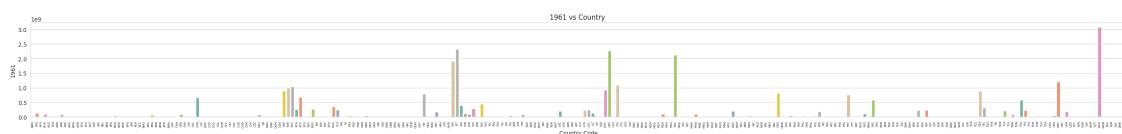
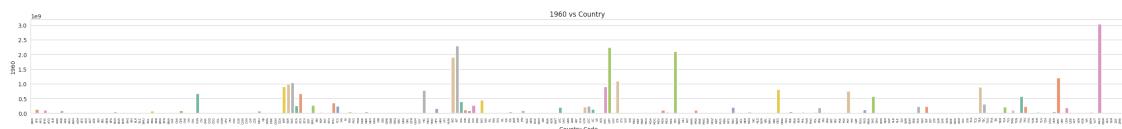
```
[ ]: data['Country Name'].unique()
```

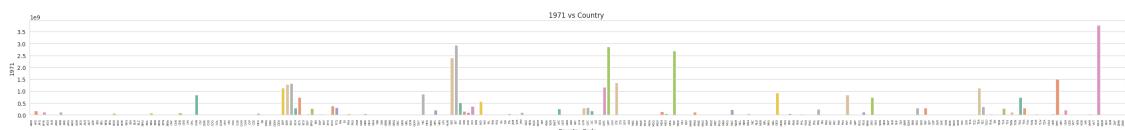
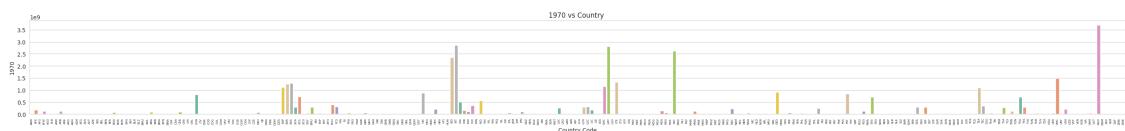
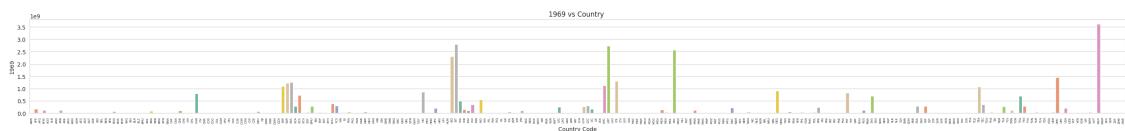
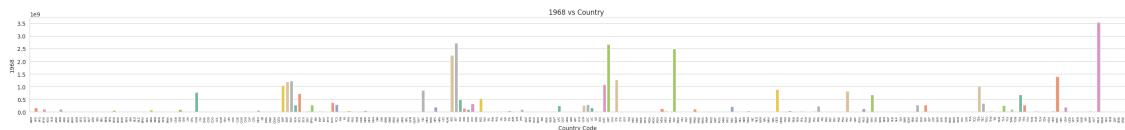
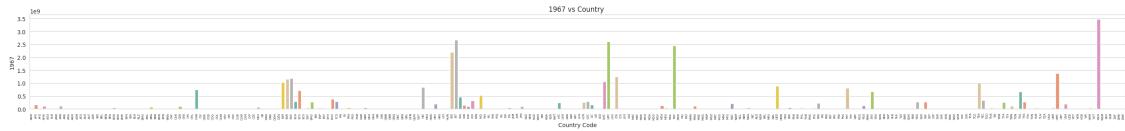
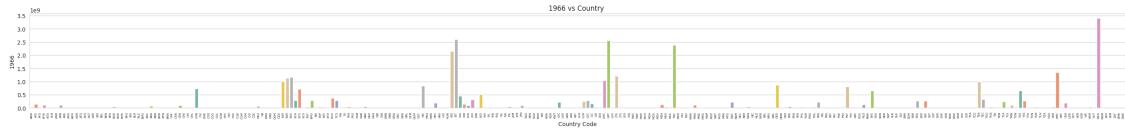
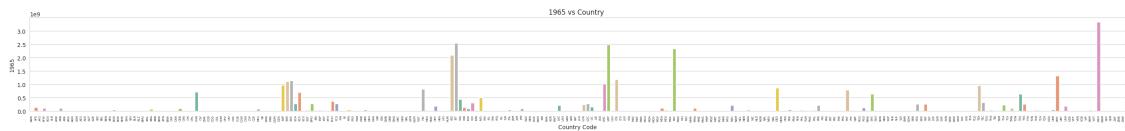
```
[ ]: array(['Aruba', 'Africa Eastern and Southern', 'Afghanistan',
   'Africa Western and Central', 'Angola', 'Albania', 'Andorra',
   'Arab World', 'United Arab Emirates', 'Argentina', 'Armenia',
   'American Samoa', 'Antigua and Barbuda', 'Australia', 'Austria',
   'Azerbaijan', 'Burundi', 'Belgium', 'Benin', 'Burkina Faso',
   'Bangladesh', 'Bulgaria', 'Bahrain', 'Bahamas, The',
   'Bosnia and Herzegovina', 'Belarus', 'Belize', 'Bermuda',
   'Bolivia', 'Brazil', 'Barbados', 'Brunei Darussalam', 'Bhutan',
   'Botswana', 'Central African Republic', 'Canada',
   'Central Europe and the Baltics', 'Switzerland', 'Channel Islands',
   'Chile', 'China', "Cote d'Ivoire", 'Cameroon', 'Congo, Dem. Rep.',
   'Congo, Rep.', 'Colombia', 'Comoros', 'Cabo Verde', 'Costa Rica',
   'Caribbean small states', 'Cuba', 'Curacao', 'Cayman Islands',
   'Cyprus', 'Czechia', 'Germany', 'Djibouti', 'Dominica', 'Denmark',
   'Dominican Republic', 'Algeria',
   'East Asia & Pacific (excluding high income)',
   'Early-demographic dividend', 'East Asia & Pacific',
   'Europe & Central Asia (excluding high income)',
   'Europe & Central Asia', 'Ecuador', 'Egypt, Arab Rep.',
   'Euro area', 'Eritrea', 'Spain', 'Estonia', 'Ethiopia',
   'European Union', 'Fragile and conflict affected situations',
   'Finland', 'Fiji', 'France', 'Faroe Islands',
   'Micronesia, Fed. Sts.', 'Gabon', 'United Kingdom', 'Georgia',
   'Ghana', 'Gibraltar', 'Guinea', 'Gambia, The', 'Guinea-Bissau',
   'Equatorial Guinea', 'Greece', 'Grenada', 'Greenland', 'Guatemala',
   'Guam', 'Guyana', 'High income', 'Hong Kong SAR, China',
   'Honduras', 'Heavily indebted poor countries (HIPC)', 'Croatia',
   'Haiti', 'Hungary', 'IBRD only', 'IDA & IBRD total', 'IDA total',
   'IDA blend', 'Indonesia', 'IDA only', 'Isle of Man', 'India',
```

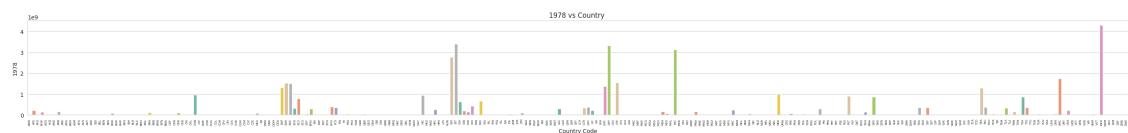
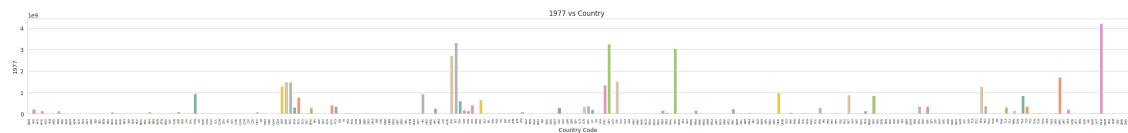
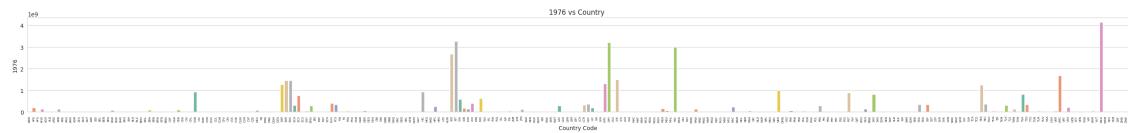
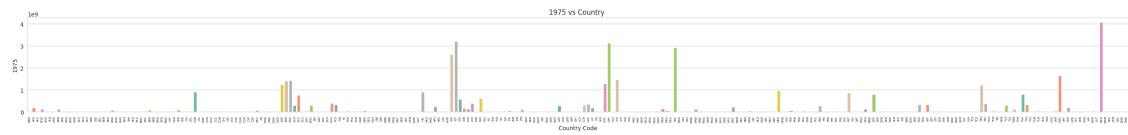
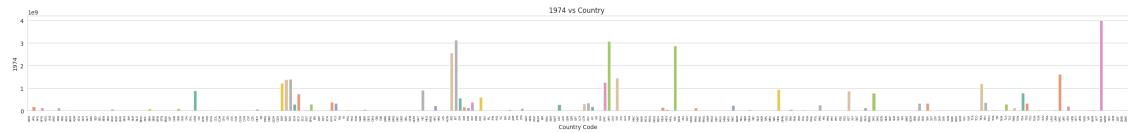
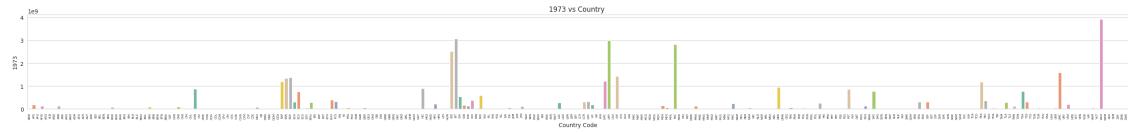
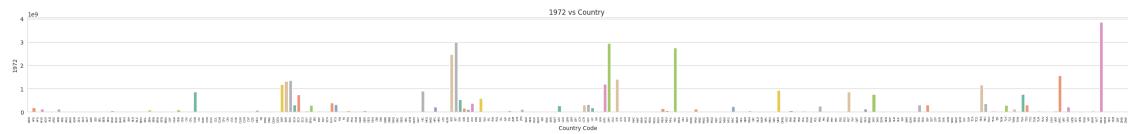
'Not classified', 'Ireland', 'Iran, Islamic Rep.', 'Iraq',  
'Iceland', 'Israel', 'Italy', 'Jamaica', 'Jordan', 'Japan',  
'Kazakhstan', 'Kenya', 'Kyrgyz Republic', 'Cambodia', 'Kiribati',  
'St. Kitts and Nevis', 'Korea, Rep.', 'Kuwait',  
'Latin America & Caribbean (excluding high income)', 'Lao PDR',  
'Lebanon', 'Liberia', 'Libya', 'St. Lucia',  
'Latin America & Caribbean',  
'Least developed countries: UN classification', 'Low income',  
'Liechtenstein', 'Sri Lanka', 'Lower middle income',  
'Low & middle income', 'Lesotho', 'Late-demographic dividend',  
'Lithuania', 'Luxembourg', 'Latvia', 'Macao SAR, China',  
'St. Martin (French part)', 'Morocco', 'Monaco', 'Moldova',  
'Madagascar', 'Maldives', 'Middle East & North Africa', 'Mexico',  
'Marshall Islands', 'Middle income', 'North Macedonia', 'Mali',  
'Malta', 'Myanmar',  
'Middle East & North Africa (excluding high income)', 'Montenegro',  
'Mongolia', 'Northern Mariana Islands', 'Mozambique', 'Mauritania',  
'Mauritius', 'Malawi', 'Malaysia', 'North America', 'Namibia',  
'New Caledonia', 'Niger', 'Nigeria', 'Nicaragua', 'Netherlands',  
'Norway', 'Nepal', 'Nauru', 'New Zealand', 'OECD members', 'Oman',  
'Other small states', 'Pakistan', 'Panama', 'Peru', 'Philippines',  
'Palau', 'Papua New Guinea', 'Poland', 'Pre-demographic dividend',  
'Puerto Rico', "Korea, Dem. People's Rep.", 'Portugal', 'Paraguay',  
'West Bank and Gaza', 'Pacific island small states',  
'Post-demographic dividend', 'French Polynesia', 'Qatar',  
'Romania', 'Russian Federation', 'Rwanda', 'South Asia',  
'Saudi Arabia', 'Sudan', 'Senegal', 'Singapore', 'Solomon Islands',  
'Sierra Leone', 'El Salvador', 'San Marino', 'Somalia', 'Serbia',  
'Sub-Saharan Africa (excluding high income)', 'South Sudan',  
'Sub-Saharan Africa', 'Small states', 'Sao Tome and Principe',  
'Suriname', 'Slovak Republic', 'Slovenia', 'Sweden', 'Eswatini',  
'Sint Maarten (Dutch part)', 'Seychelles', 'Syrian Arab Republic',  
'Turks and Caicos Islands', 'Chad',  
'East Asia & Pacific (IDA & IBRD countries)',  
'Europe & Central Asia (IDA & IBRD countries)', 'Togo', 'Thailand',  
'Tajikistan', 'Turkmenistan',  
'Latin America & the Caribbean (IDA & IBRD countries)',  
'Timor-Leste', 'Middle East & North Africa (IDA & IBRD countries)',  
'Tonga', 'South Asia (IDA & IBRD)',  
'Sub-Saharan Africa (IDA & IBRD countries)', 'Trinidad and Tobago',  
'Tunisia', 'Turkiye', 'Tuvalu', 'Tanzania', 'Uganda', 'Ukraine',  
'Upper middle income', 'Uruguay', 'United States', 'Uzbekistan',  
'St. Vincent and the Grenadines', 'Venezuela, RB',  
'British Virgin Islands', 'Virgin Islands (U.S.)', 'Viet Nam',  
'Vanuatu', 'World', 'Samoa', 'Kosovo', 'Yemen, Rep.',  
'South Africa', 'Zambia', 'Zimbabwe'], dtype=object)

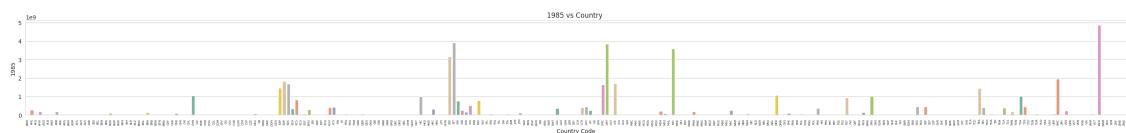
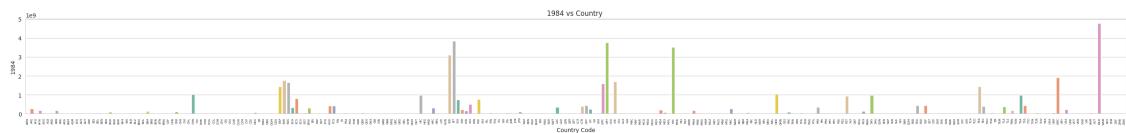
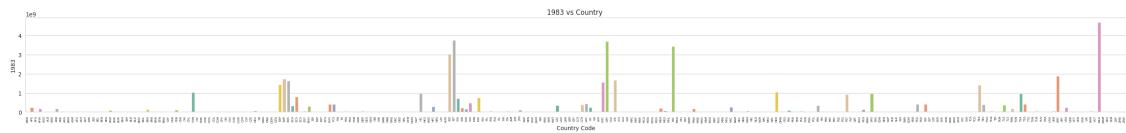
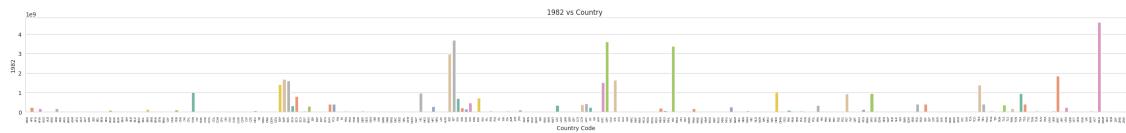
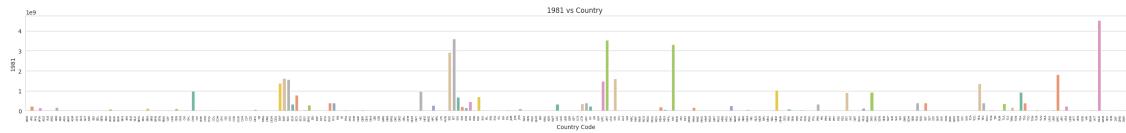
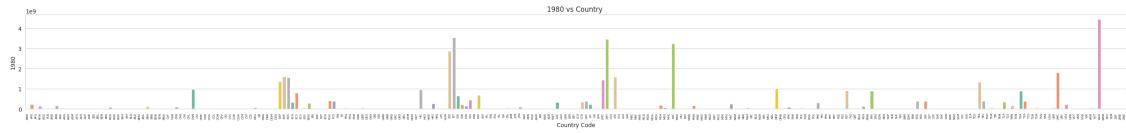
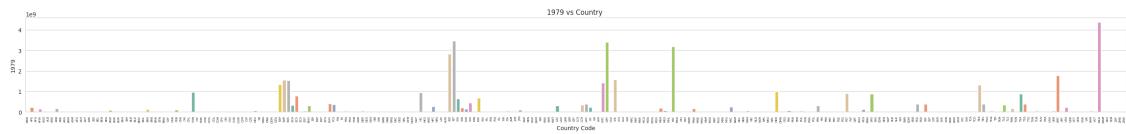
## 2 BIVARIATE ANALYSIS

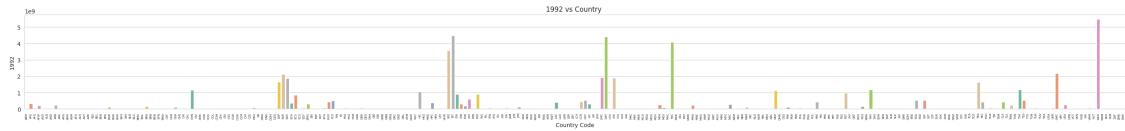
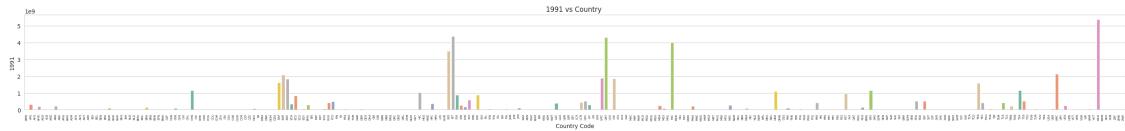
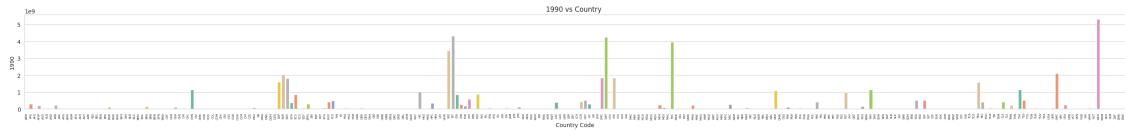
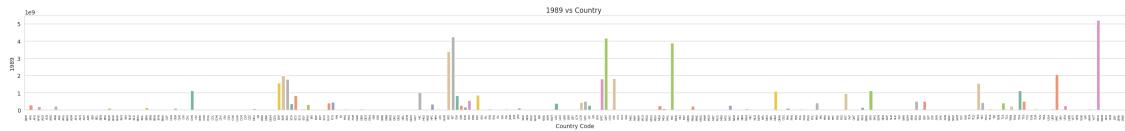
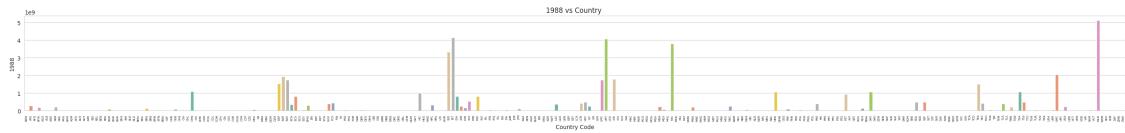
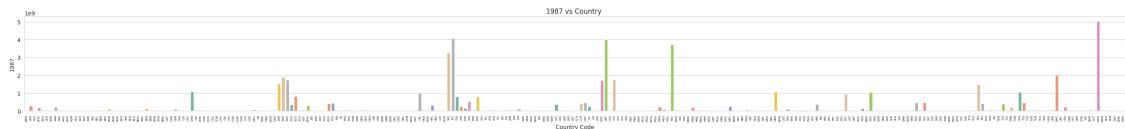
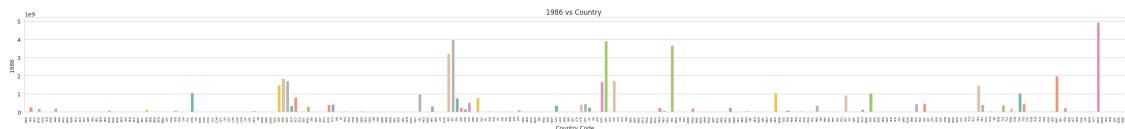
```
[ ]: for column in df.select_dtypes(include='number').columns:  
    # Create a bar plot  
    plt.figure(figsize=(35, 3))  
    plt.xticks(rotation=90, fontsize=5)  
  
    # Use seaborn's barplot to plot the mean 'price' for each category in the  
    # selected numeric column  
    sb.barplot(data=data, x=df['Country Code'], y=column, palette='Set2')  
  
    plt.title(f'{column} vs Country')  
    plt.show()
```

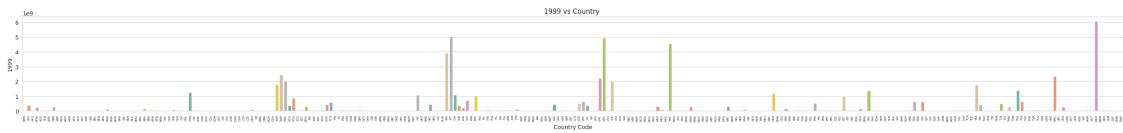
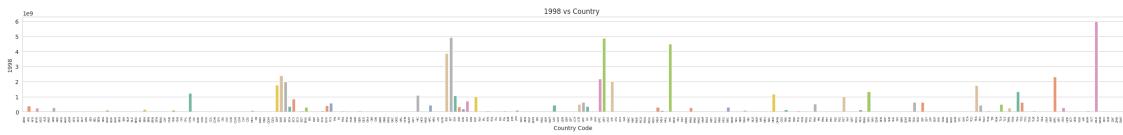
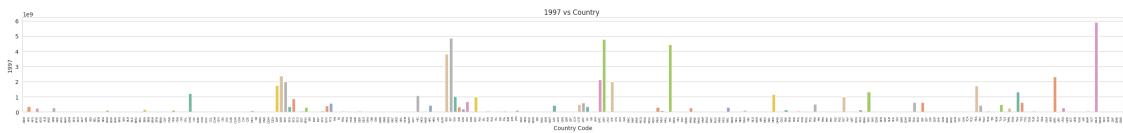
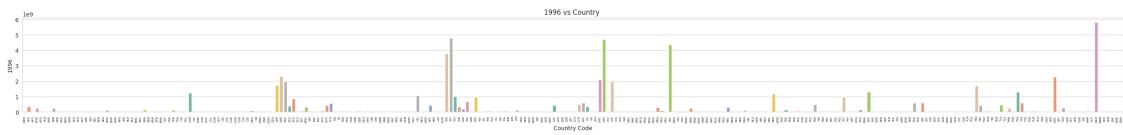
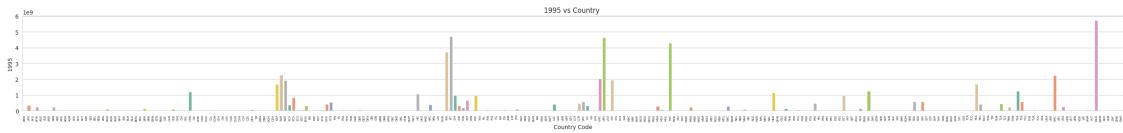
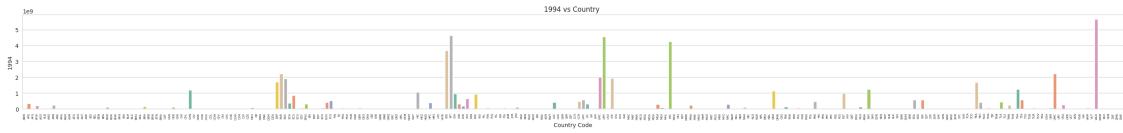
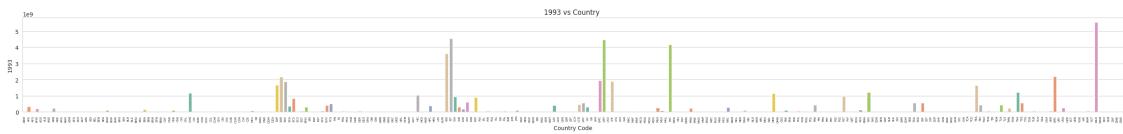


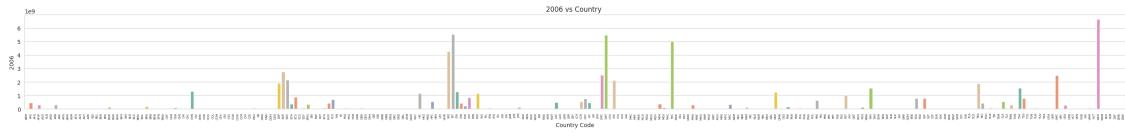
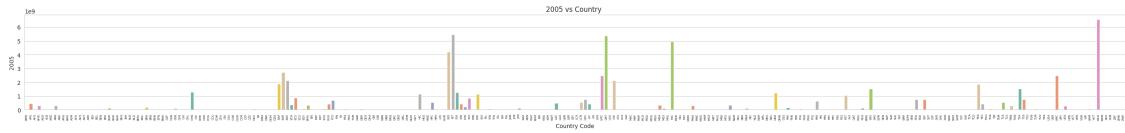
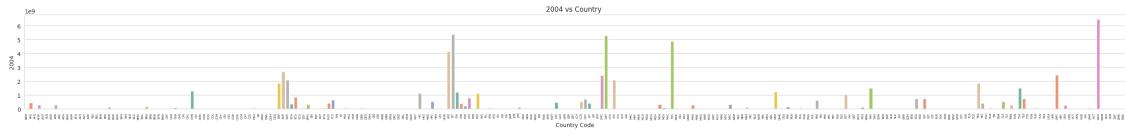
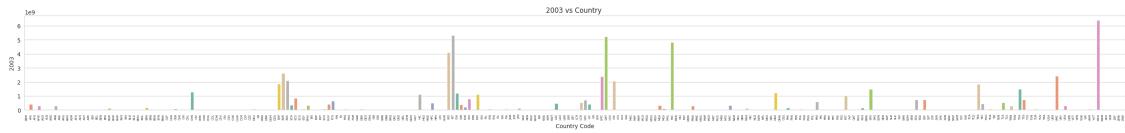
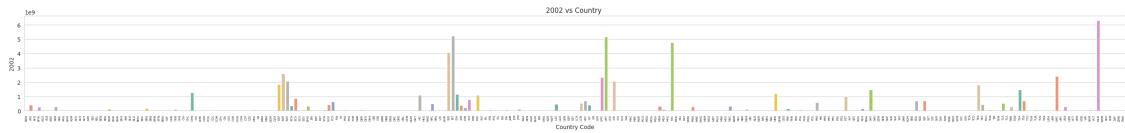
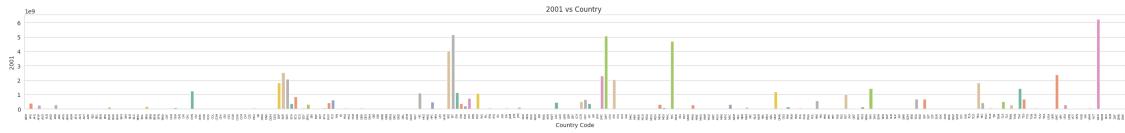
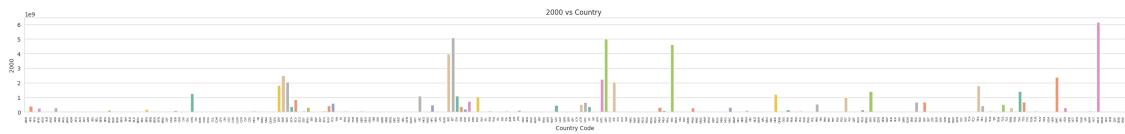


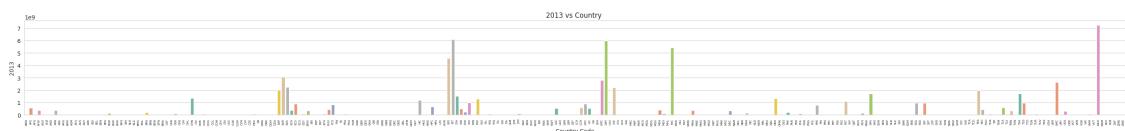
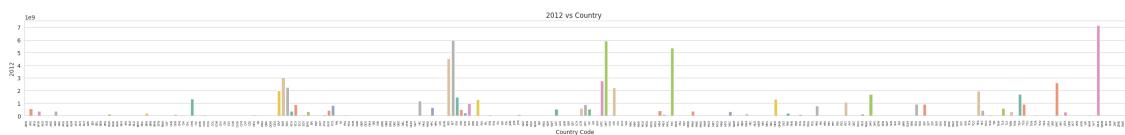
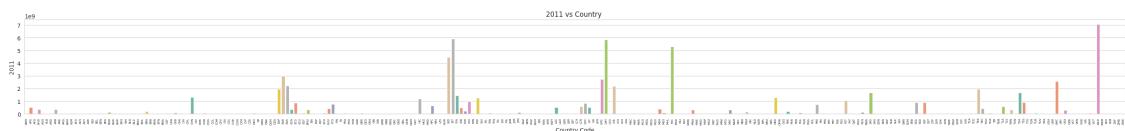
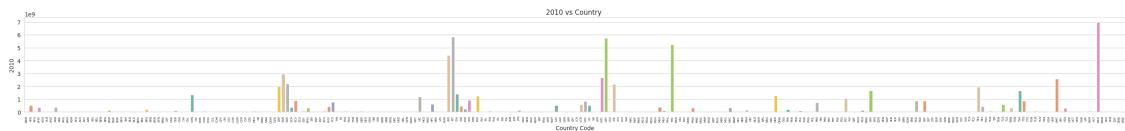
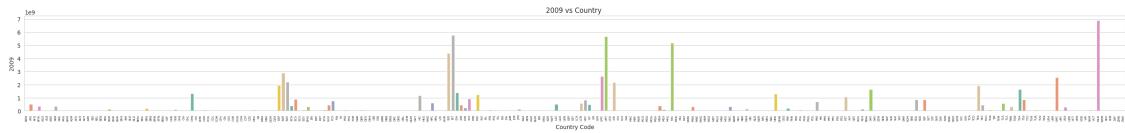
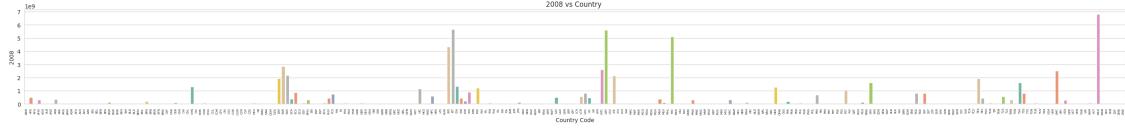
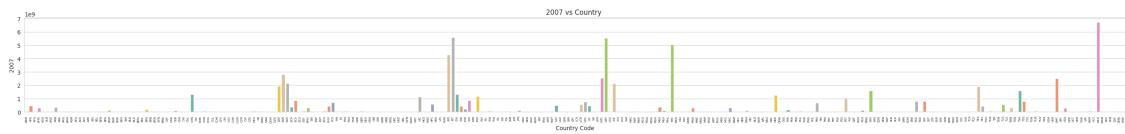
















```
[ ]: # Sort the DataFrame by the '2022' column in descending order
sorted_df = data.sort_values(by='2022', ascending=False)
sorted_dfb = data.sort_values(by='1969', ascending=False)
# Extract the top and bottom ten rows
top_ten_countries = sorted_df.head(10)
top_ten_countriesb = sorted_dfb.head(10)

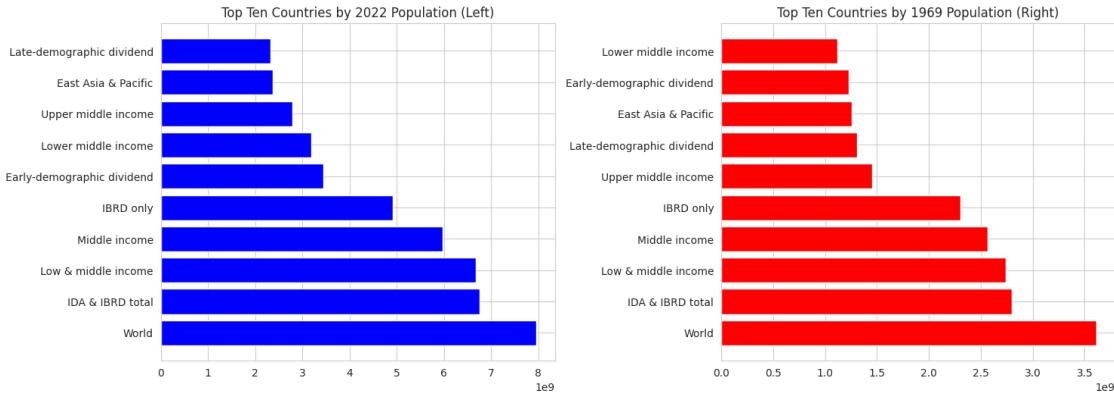
# Create a 2x2 subplot layout
plt.figure(figsize=(15, 10))

# Top left subplot for top ten countries
plt.subplot(2, 2, 1)
plt.barh(top_ten_countries['Country Name'], top_ten_countries['2022'], color='blue')
plt.title('Top Ten Countries by 2022 Population (Left)')

# Top right subplot for bottom ten countries
plt.subplot(2, 2, 2)
plt.barh(top_ten_countriesb['Country Name'], top_ten_countriesb['1969'], color='red')
plt.title('Top Ten Countries by 1969 Population (Right)')

# Adjust layout
plt.tight_layout()

# Show the plot
plt.show()
```



```
[ ]: # Sort the DataFrame by the '2022' column in descending order
sorted_df = data.sort_values(by='2022', ascending=False)
sorted_dfb = data.sort_values(by='1979', ascending=False)
# Extract the top and bottom ten rows
top_ten_countries = sorted_df.head(10)
top_ten_countriesb = sorted_dfb.head(10)

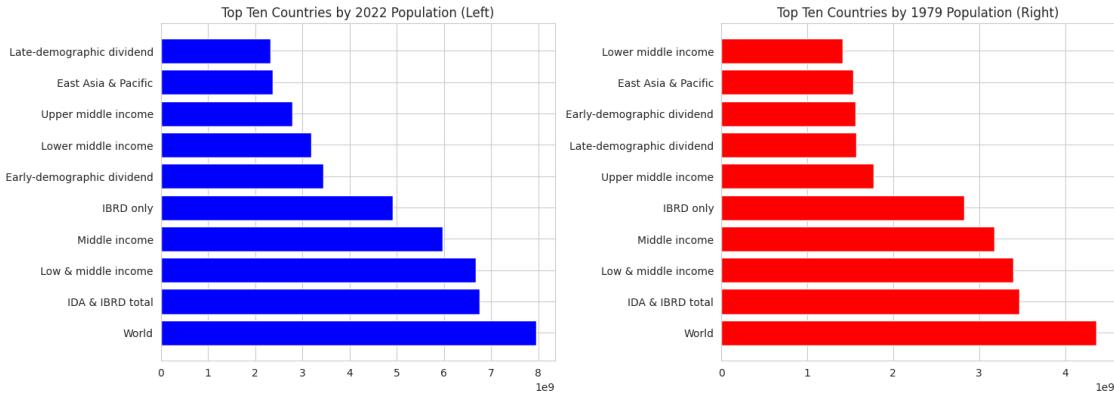
# Create a 2x2 subplot layout
plt.figure(figsize=(15, 10))

# Top left subplot for top ten countries
plt.subplot(2, 2, 1)
plt.barh(top_ten_countries['Country Name'], top_ten_countries['2022'], color='blue')
plt.title('Top Ten Countries by 2022 Population (Left)')

# Top right subplot for bottom ten countries
plt.subplot(2, 2, 2)
plt.barh(top_ten_countriesb['Country Name'], top_ten_countriesb['1979'], color='red')
plt.title('Top Ten Countries by 1979 Population (Right)')

# Adjust layout
plt.tight_layout()

# Show the plot
plt.show()
```



```
[ ]: # Sort the DataFrame by the '2022' column in descending order
sorted_df = data.sort_values(by='2022', ascending=False)
sorted_dfb = data.sort_values(by='1989', ascending=False)
# Extract the top and bottom ten rows
top_ten_countries = sorted_df.head(10)
top_ten_countriesb = sorted_dfb.head(10)

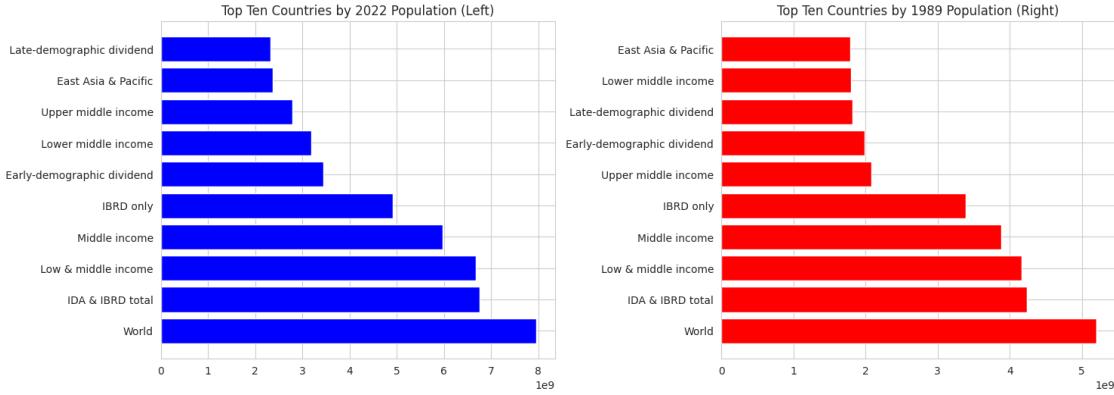
# Create a 2x2 subplot layout
plt.figure(figsize=(15, 10))

# Top left subplot for top ten countries
plt.subplot(2, 2, 1)
plt.barh(top_ten_countries['Country Name'], top_ten_countries['2022'], color='blue')
plt.title('Top Ten Countries by 2022 Population (Left)')

# Top right subplot for bottom ten countries
plt.subplot(2, 2, 2)
plt.barh(top_ten_countriesb['Country Name'], top_ten_countriesb['1989'], color='red')
plt.title('Top Ten Countries by 1989 Population (Right)')

# Adjust layout
plt.tight_layout()

# Show the plot
plt.show()
```



```
[ ]: # Sort the DataFrame by the '2022' column in descending order
sorted_df = data.sort_values(by='2022', ascending=False)
sorted_dfb = data.sort_values(by='1999', ascending=False)
# Extract the top and bottom ten rows
top_ten_countries = sorted_df.head(10)
top_ten_countriesb = sorted_dfb.head(10)

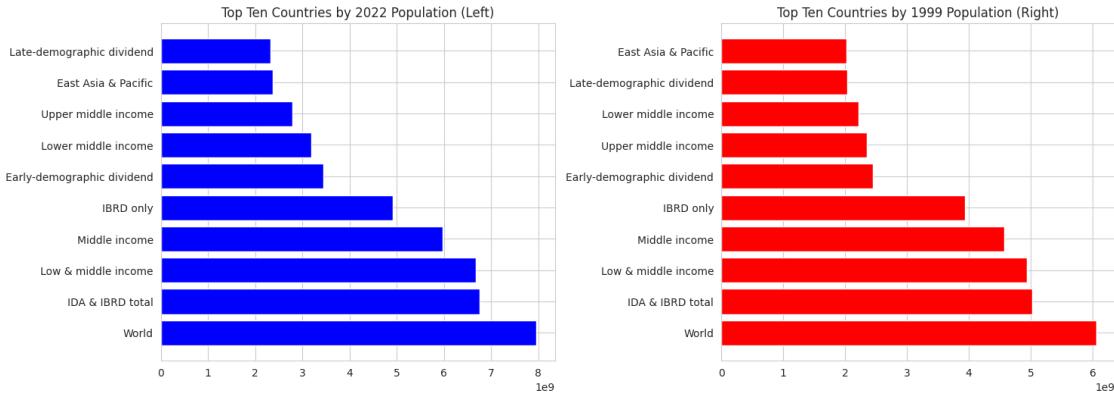
# Create a 2x2 subplot layout
plt.figure(figsize=(15, 10))

# Top left subplot for top ten countries
plt.subplot(2, 2, 1)
plt.barh(top_ten_countries['Country Name'], top_ten_countries['2022'], color='blue')
plt.title('Top Ten Countries by 2022 Population (Left)')

# Top right subplot for bottom ten countries
plt.subplot(2, 2, 2)
plt.barh(top_ten_countriesb['Country Name'], top_ten_countriesb['1999'], color='red')
plt.title('Top Ten Countries by 1999 Population (Right)')

# Adjust layout
plt.tight_layout()

# Show the plot
plt.show()
```



```
[ ]: # Sort the DataFrame by the '2022' column in descending order
sorted_df = data.sort_values(by='2022', ascending=False)
sorted_dfb = data.sort_values(by='2009', ascending=False)
# Extract the top and bottom ten rows
top_ten_countries = sorted_df.head(10)
top_ten_countriesb = sorted_dfb.head(10)

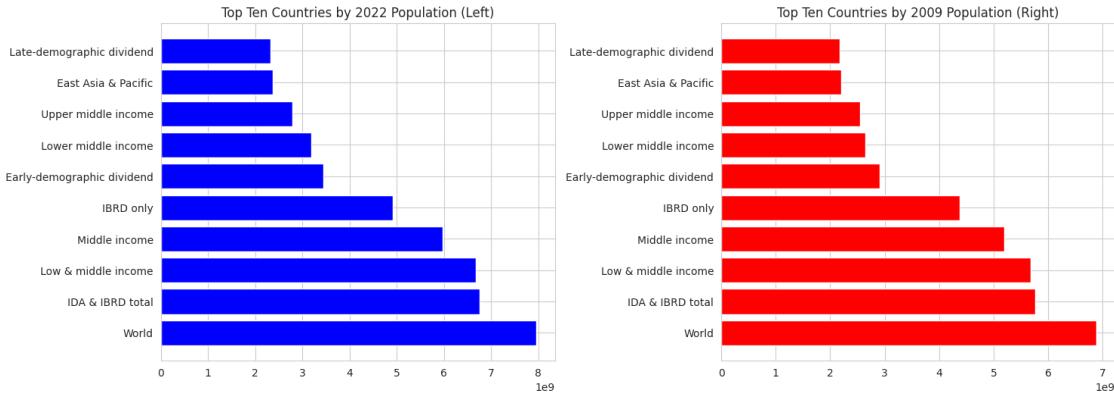
# Create a 2x2 subplot layout
plt.figure(figsize=(15, 10))

# Top left subplot for top ten countries
plt.subplot(2, 2, 1)
plt.barh(top_ten_countries['Country Name'], top_ten_countries['2022'],
         color='blue')
plt.title('Top Ten Countries by 2022 Population (Left)')

# Top right subplot for bottom ten countries
plt.subplot(2, 2, 2)
plt.barh(top_ten_countriesb['Country Name'], top_ten_countriesb['2009'],
         color='red')
plt.title('Top Ten Countries by 2009 Population (Right)')

# Adjust layout
plt.tight_layout()

# Show the plot
plt.show()
```



```
[ ]: # Sort the DataFrame by the '2022' column in descending order
sorted_df = data.sort_values(by='2022', ascending=False)
sorted_dfb = data.sort_values(by='2015', ascending=False)
# Extract the top and bottom ten rows
top_ten_countries = sorted_df.head(10)
top_ten_countriesb = sorted_dfb.head(10)

# Create a 2x2 subplot layout
plt.figure(figsize=(15, 10))

# Top left subplot for top ten countries
plt.subplot(2, 2, 1)
plt.barh(top_ten_countries['Country Name'], top_ten_countries['2022'], color='blue')
plt.title('Top Ten Countries by 2022 Population (Left)')

# Top right subplot for bottom ten countries
plt.subplot(2, 2, 2)
plt.barh(top_ten_countriesb['Country Name'], top_ten_countriesb['2015'], color='red')
plt.title('Top Ten Countries by 2015 Population (Right)')

# Adjust layout
plt.tight_layout()

# Show the plot
plt.show()
```

