

## Trilha 4: Tarefa (Peso 3)

Empregue no mínimo duas bases de dados (tabelas) extraídas de um ou ambos os sites abaixo:

a. <https://wid.world/>;

b. <https://www.gapminder.org/>

E faça uma Exploração dos Dados EDA de seu interesse, empregando o Python e o que você aprendeu até aqui. Os sites apresentam diversos dados sobre as diferenças de desenvolvimento dos países e a desigualdade. Foque em perguntas que são do seu interesse sobre os dados.

Na sua análise você deve responder ao menos quatro perguntas relevantes sobre os dados, incluir na sua análise seleções e visualizações dos dados e um texto de conclusão a cada pergunta.

Peça ao professor modelos de código para sua implementação.

Seu código deve ser minimamente claro e organizado. Não são necessários comentários detalhados, mas deixe claro no seu código cada parte (atividade), cada fase realizada e a resposta final com sua conclusão.

Entrega:

.pdf de print do seu notebook Python.

Compartilhe o endereço do seu notebook Python (Google Drive ou GitHub) 100% executável para abertura no Google Colab.

Não deixe de incluir o seu texto uma introdução justificando a escolha da análise e ao final suas conclusões.

## Introdução

Para realização desta análise foram utilizados os dados do site Gapminder. O dataframe escolhido possui dados sobre a Pobreza no Mundo. Essa atividade permitiu o treino prático em processo EDA bem como o treino em gráficos e análise posterior.

## Perguntas:

### 1. Em que período o índice de Gini melhorou no Brasil?

Com a utilização do gráfico de barras foi verificada a variação do Gini entre 1980 e 2018, onde é possível observar uma piora entre 1980 e 1988, melhora entre 1988 e 1991, piora em 1992 e uma melhora entre 1993 e 2014. A melhora de 20 anos é interrompida e tem início de piora em 2015.

## 2. Como o Brasil se posiciona no índice de Gini em relação aos demais países do mundo?

Comparando os dados de Gini do Brasil com os do Mundo, utilizei o gráfico boxplot para melhor visualização. Considerando quanto maior o índice, pior a desigualdade do país. É possível concluir que o Brasil está com índice muito ruim de Gini, estando muito distante da média mundial.

## 3. Em que período houve mais dados sobre a pobreza no Brasil?

Utilizando um gráfico de linha com a densidade dos dados é possível observar que houve mais dados entre 1985 e 2015, com um crescimento ainda maior entre 2005 e 2010.

## 4. Como a pobreza urbana e rural do Brasil se modificaram com o tempo?

Utiliza-se o gráfico boxplot para analisar a pobreza rural no Brasil. É possível observar que a média da pobreza rural no Brasil está acima da média mundial de pobreza. Realizando essa análise com a pobreza urbana no Brasil verifica-se que a média é menor que a média mundial, está abaixo da média.

Gerando gráfico de barras Pobreza Rural no Brasil entre 1997 e 2002. Apresenta dados claros. Permite inferir situação de queda mas o dataframe não possui dados suficientes dos demais anos. Também utilizando o mesmo tipo de gráfico é possível observar o aumento da pobreza urbana no Brasil no mesmo período. Concluindo, utilizando um gráfico comparativo entre pobreza urbana x pobreza rural no Brasil por período é possível observar aumento da pobreza urbana e redução da pobreza rural no Brasil. Pelo campo valor é possível observar que a pobreza rural é muito maior.

## ▼ Conclusão

Foi possível realizar a análise requisitada utilizando as técnicas aprendidas em aula. Com este treino foi possível identificar dificuldades em alguns processos que precisarão ser aperfeiçoados com muita prática, como uso de dataframes com metadados e dados em diversas disposições.

```
# Importando Bibliotecas e adicionando seus respectivos alias
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```

import matplotlib as mpl
import seaborn as sns
%matplotlib inline

#Importando os df: Gini, Rural, Urban
# Fonte: https://www.gapminder.org/data

dfGini = pd.read_csv('https://raw.githubusercontent.com/rodglins/Python/master/desafios')
dfRural = pd.read_csv('https://raw.githubusercontent.com/rodglins/Python/master/desafios')
dfUrban = pd.read_csv('https://raw.githubusercontent.com/rodglins/Python/master/desafios')

# verificando o formato do df
dfGini.head()

```

	country	1966	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978
0	Angola	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Albania	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	United Arab Emirates	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Argentina	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	Armenia	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```

# tratando os dados, transferindo dados de ano de colunas para linhas, e seus respectivos valores
dfGini = pd.melt(dfGini.reset_index(), id_vars=['country','index'], var_name='ano', value_name='valor')
dfGini.head()

```

	country	index	ano	valor
8683	Zimbabwe	166	2018	50.3
8349	Zimbabwe	166	2016	44.3
7347	Zimbabwe	166	2010	43.2
166	Zimbabwe	166	1966	NaN
333	Zimbabwe	166	1968	NaN

```

# transformando campo ano para o formato data

dfGini.ano = pd.to_datetime(dfGini.ano, format='%Y',errors='coerce')
display(dfGini.dtypes)

```

```
country      object
index        int64
ano          datetime64[ns]
```

```
# comparando dados Gini Brasil x Mundo, com gráfico boxplot
# É possível concluir que o Brasil está com índice muito ruim de Gini, estando muito at
```

```
dfGini_not_null = dfGini[ ~ np.isnan( dfGini.valor )]
```

```
plt.figure(figsize=(12,4))
plt.boxplot(dfGini_not_null[dfGini_not_null.country == 'Brazil'].valor,positions=[1],sh
plt.boxplot(dfGini_not_null.valor,positions=[2],showfliers=False)
plt.title('Gini')
plt.xticks(ticks=[1,2], labels=['Brasil','Mundo'])
```

```
([<matplotlib.axis.XTick at 0x7fe593f6a390>,
  <matplotlib.axis.XTick at 0x7fe594053710>],
 [Text(0, 0, 'Brasil'), Text(0, 0, 'Mundo')])
```



```
# tratando valores vazios
```

```
dfGini2 = dfGini_not_null.sort_values(['country', 'ano'])
```

```
dfGini2.head()
```

	country	index	ano	valor
4677	Albania	1	1995-01-01	27.0
5679	Albania	1	2001-01-01	31.7
6180	Albania	1	2004-01-01	30.6
6681	Albania	1	2007-01-01	30.0
7349	Albania	1	2011-01-01	29.0

```
# resetando o index
```

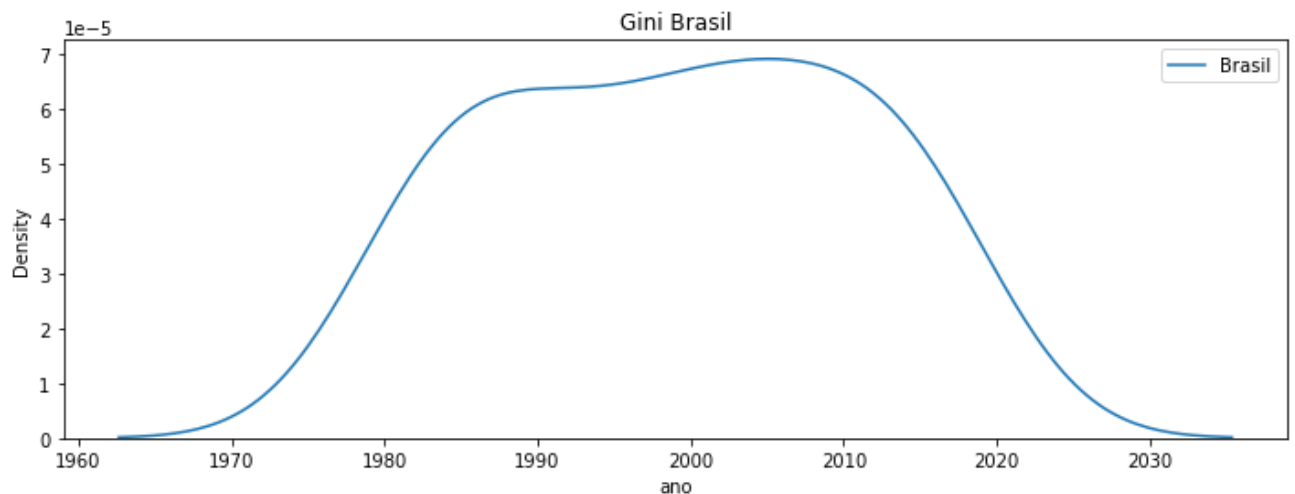
```
dfGini = dfGini.reset_index()
dfGini.head()
```

	level_0	country	index	ano	valor
0	8683	Zimbabwe	166	2018-01-01	50.3
1	8349	Zimbabwe	166	2016-01-01	44.3
2	7347	Zimbabwe	166	2010-01-01	43.2
3	166	Zimbabwe	166	1966-01-01	NaN
4	333	Zimbabwe	166	1968-01-01	NaN

# Mostra a densidade dos dados conforme o período:

```
plt.figure(figsize=(12,4))
sns.kdeplot(dfGini2[dfGini2.country == 'Brazil'].ano,label='Brasil')
# sns.kdeplot(dfGini_not_null.ano,label='Mundo')
plt.title('Gini Brasil')
plt.legend()
plt.show()
```

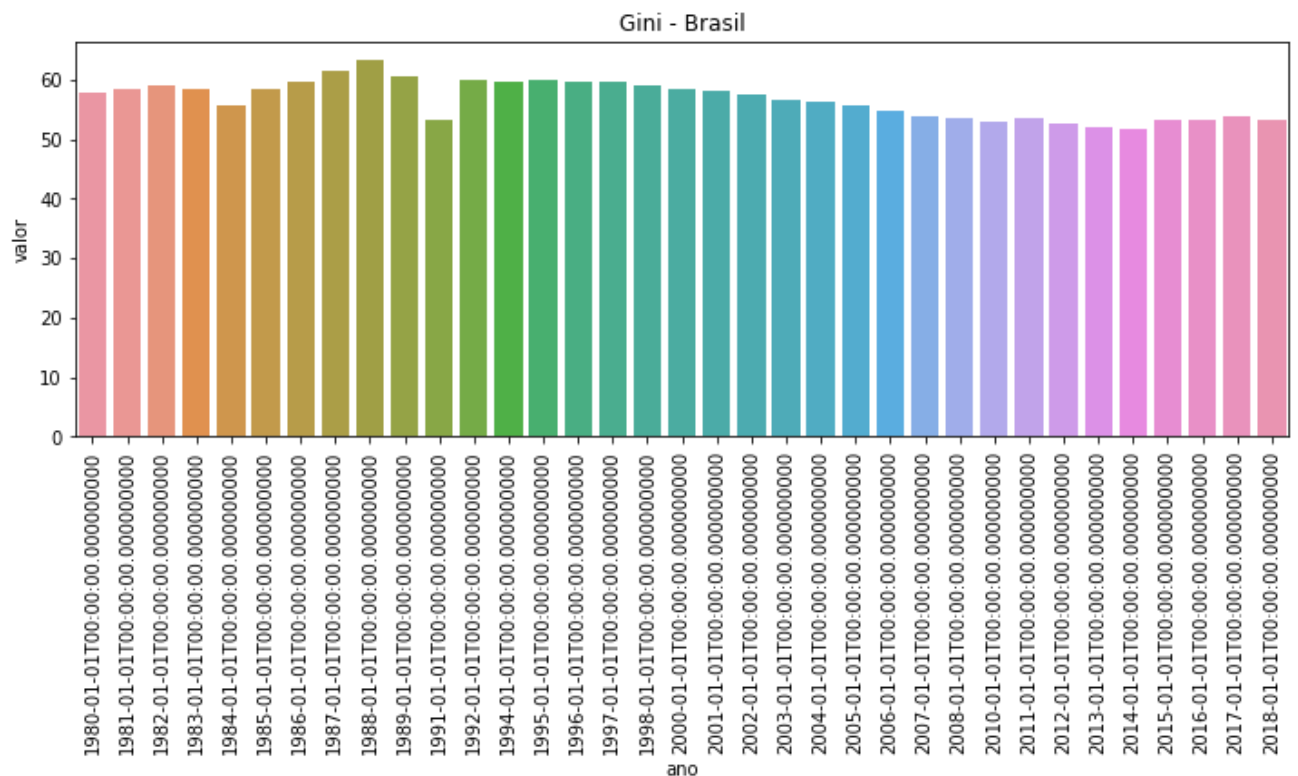
# Houve mais dados entre 1985 e 2015, com um crescimento maior entre 2005 e 2010.



# Verificando a variação do Gini entre 1980 e 2018:

# É possível observar uma piora entre 1980 e 1988, melhora entre 1988 e 1991, piora em  
# A melhora de 20 anos é interrompida e tem início de piora em 2015.

```
plt.figure(figsize=(12,4))
sns.barplot(x=dfGini2[dfGini2.country == 'Brazil'].ano, y=dfGini2[dfGini2.country == 'E
plt.title('Gini - Brasil')
# plt.hlines(dfGini2.valor.mean(),-0.5,7.5,color='r', linestyle=':')
# plt.text(0, dfGini2.valor.mean()+1,'Média de valor')
plt.xticks(rotation=90)
plt.show()
```



## ▼ Tratando dados Dataframe Rural

```
# Convertendo dados de colunas para linhas
```

```
dfRural2 = pd.melt(dfRural.reset_index(), id_vars=['country','index'], var_name='ano',
dfRural2.head())
```

	country	index	ano	valor
<b>759</b>	Zimbabwe	75	1995	48.0
<b>379</b>	Zimbabwe	75	1990	35.8
<b>75</b>	Zimbabwe	75	1984	NaN
<b>151</b>	Zimbabwe	75	1987	NaN
<b>227</b>	Zimbabwe	75	1988	NaN

```
# Convertendo campo ano
```

```
dfRural2.ano = pd.to_datetime(dfRural2.ano, format='%Y',errors='coerce')
display(dfRural2.dtypes)
```

```
country          object
index            int64
ano              datetime64[ns]
valor            float64
dtype: object
```

```
# Gerando gráfico de Pobreza Rural
```

```
# Como o gráfico boxplot é possível observar que a média da pobreza rural no Brasil est
```

```

dfRural3 = dfRural2.sort_values(['country', 'ano'])
dfRural4 = dfRural3[ ~ np.isnan( dfRural3.valor )]

plt.figure(figsize=(12,4))
plt.boxplot(dfRural4[dfRural4.country == 'Brazil'].valor,positions=[1],showfliers=False)
plt.boxplot(dfRural4.valor,positions=[2],showfliers=False)
plt.title('Pobreza Rural')
plt.xticks(ticks=[1,2], labels=['Brasil','Mundo'])

([<matplotlib.axis.XTick at 0x7fe59440bd10>,
  <matplotlib.axis.XTick at 0x7fe59440bf50>],
 [Text(0, 0, 'Brasil'), Text(0, 0, 'Mundo')])

```



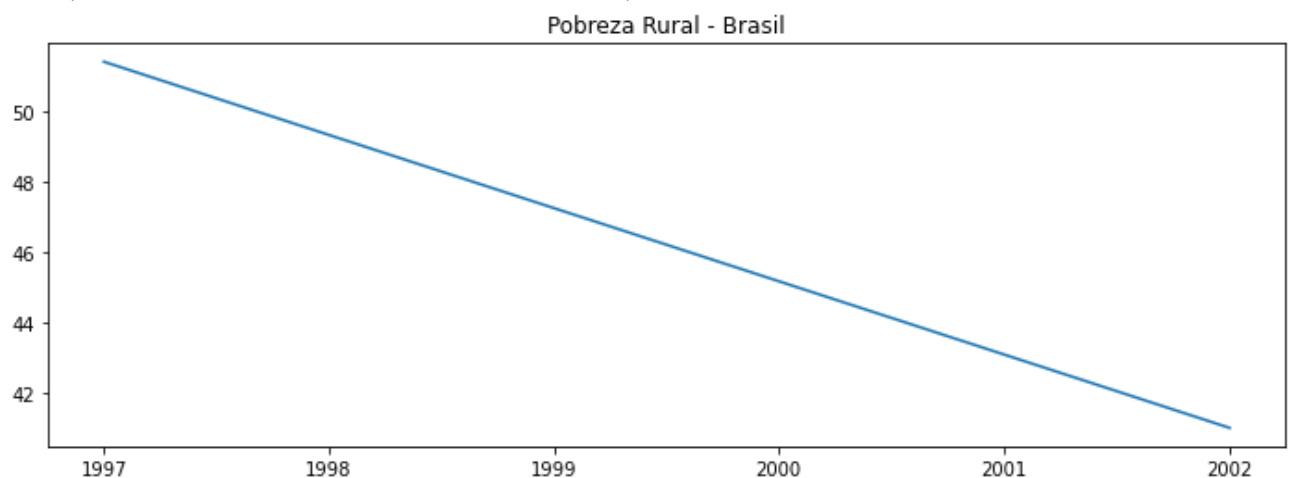
```

plt.figure(figsize=(12,4))
plt.plot(dfRural4[dfRural4.country == 'Brazil'].ano, dfRural4[dfRural4.country == 'Brazil'].valor)
plt.title('Pobreza Rural - Brasil')

```

# Gráfico mostra uma situação de queda

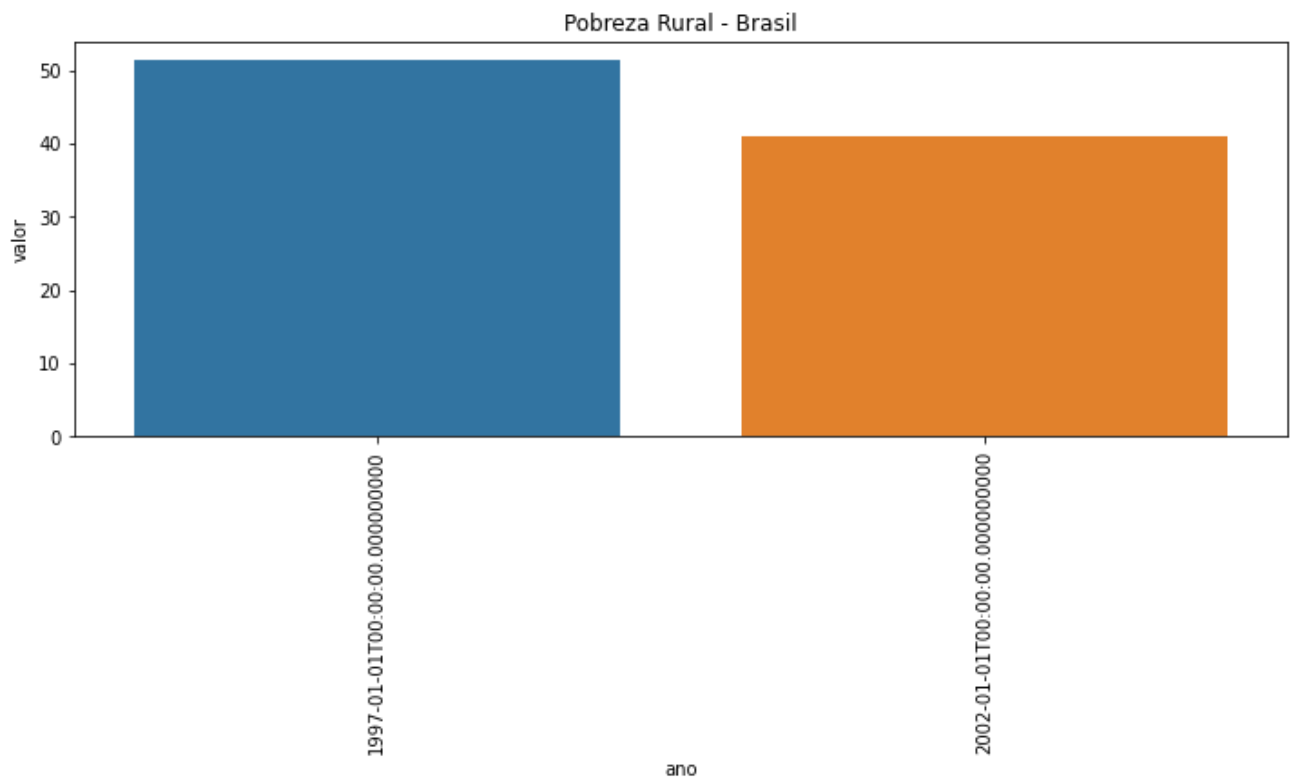
```
Text(0.5, 1.0, 'Pobreza Rural - Brasil')
```



# Gerando gráfico de barras Pobreza Rural no Brasil. Apresenta dados claros.  
 # Permite inferir situação de queda mas o dataframe não possui dados suficientes dos de

```
plt.figure(figsize=(12,4))
```

```
sns.barplot(x=dfRural4[dfRural4.country == 'Brazil'].ano, y=dfRural4[dfRural4.country ==
plt.title('Pobreza Rural - Brasil')
# plt.hlines(dfRural4.valor.mean(),-0.5,7.5,color='r', linestyle=':')
# plt.text(0, dfRural4.valor.mean()+1,'Média de valor')
plt.xticks(rotation=90)
plt.show()
```



```
dfRural4.head()
```

	country	index	ano	valor
1140	Albania	0	2001-01-01	29.6
91	Algeria	15	1987-01-01	16.6
623	Algeria	15	1994-01-01	30.3
913	Armenia	1	1998-01-01	50.8
1065	Armenia	1	2000-01-01	48.7

## ▼ Tratando dados Dataframe Urbano

```
# Convertendo campos de Ano para linhas
```

```
dfUrban2 = pd.melt(dfUrban.reset_index(), id_vars=['country','index'], var_name='ano',
```

```
#Tratando o formato data
```



```

dfUrban2.ano = pd.to_datetime(dfUrban2.ano, format='%Y',errors='coerce')

# Criando um df sem valores vazios

dfUrban3 = dfUrban2[ ~ np.isnan( dfUrban2.valor )]

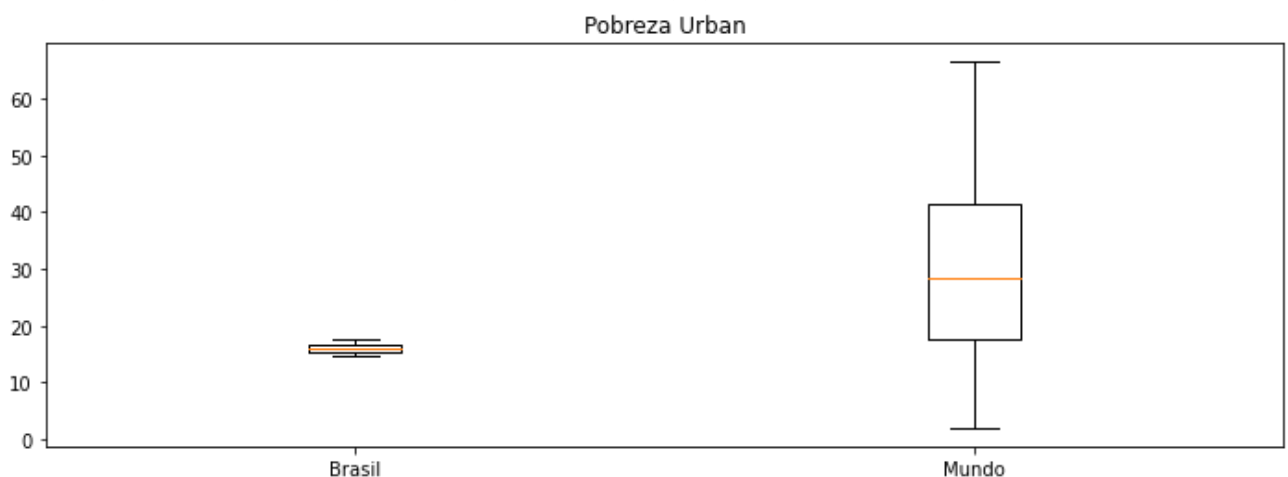
# Ordenando a tabela por País e depois por Ano.
dfUrban4 = dfUrban3.sort_values(['country', 'ano'])

# Gráfico boxplot Pobreza Urbana Brasil e no Mundo
# Podemos observar que a pobreza urbana no Brasil é menor que a média mundial.

plt.figure(figsize=(12,4))
plt.boxplot(dfUrban4[dfUrban4.country == 'Brazil'].valor,positions=[1],showfliers=False)
plt.boxplot(dfUrban4.valor,positions=[2],showfliers=False)
plt.title('Pobreza Urban')
plt.xticks(ticks=[1,2], labels=['Brasil','Mundo'])

([<matplotlib.axis.XTick at 0x7fe5943c3190>,
  <matplotlib.axis.XTick at 0x7fe5943f4790>],
 [Text(0, 0, 'Brasil'), Text(0, 0, 'Mundo')])

```

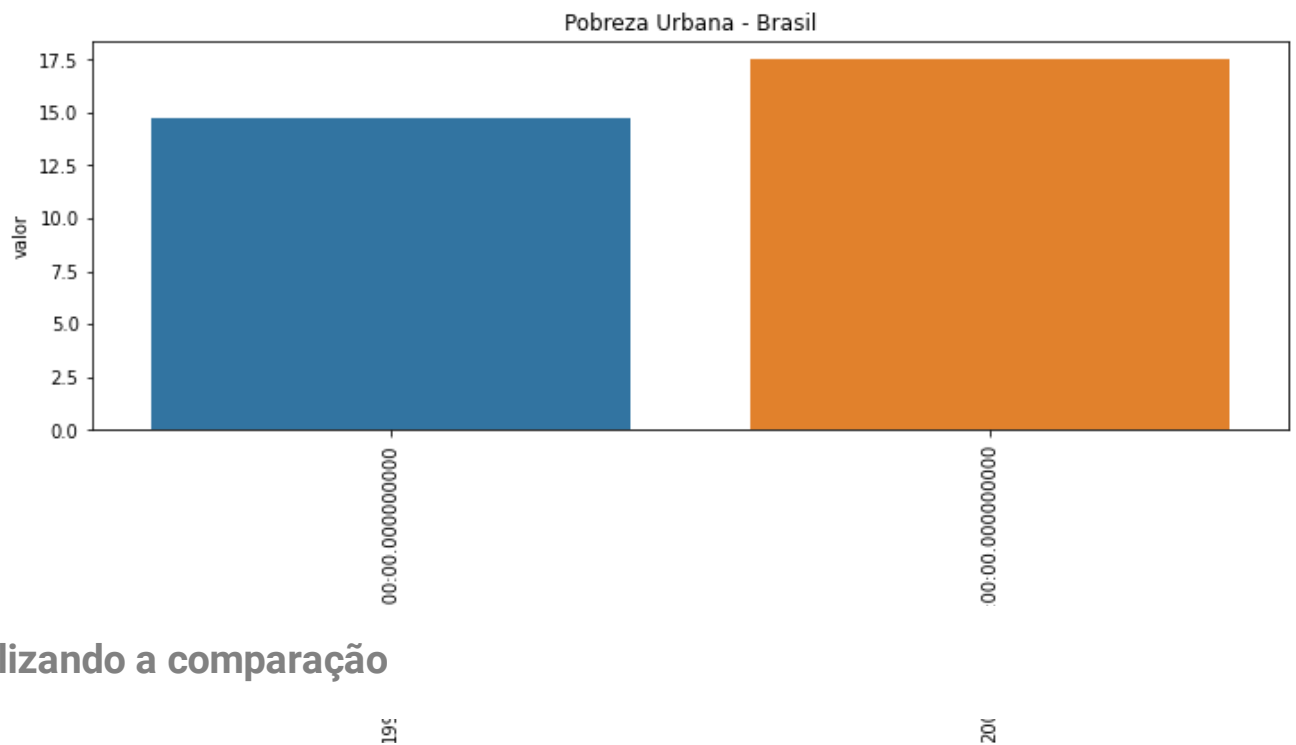


```

# Gráfico de barras Pobreza Urbana no Brasil
# É possível observar o aumento da pobreza urbana no Brasil entre 1997 e 2002.

plt.figure(figsize=(12,4))
sns.barplot(x=dfUrban4[dfUrban4.country == 'Brazil'].ano, y=dfUrban4[dfUrban4.country == 'Brazil'].valor)
plt.title('Pobreza Urbana - Brasil')
plt.xticks(rotation=90)
plt.show()

```



### ▼ Realizando a comparação

```
# Gráfico comparativo entre pobreza urbana x pobreza rural no Brasil
# É possível observar aumento da pobreza urbana e redução da pobreza rural no Brasil.

plt.figure(figsize=(12,5))

plt.subplot(1,2,1)
sns.barplot(x=dfUrban4[dfUrban4.country == 'Brazil'].ano, y=dfUrban4[dfUrban4.country == 'Brazil'].valor)
plt.title('Pobreza Urbana - Brasil')
plt.xticks(rotation=90)

plt.subplot(1,2,2)
sns.barplot(x=dfRural4[dfRural4.country == 'Brazil'].ano, y=dfRural4[dfRural4.country == 'Brazil'].valor)
plt.title('Pobreza Rural - Brasil')
plt.xticks(rotation=90)

plt.show()
```

