```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
import statsmodels.formula.api as smf
import scipy.stats as stats
```

```
#quality of governance dataset import
url = 'https://www.qogdata.pol.gu.se/data/qog_bas_cs_jan24.xlsx'
df = pd.read_excel(url)
df.head()
```

	ccode	cname	ccode_qog	cname_qog	ccodealp	ccodecow	version	ajr_settmort	atop_ally	atop_number	...	wvs_imprel	wv
0	4	Afghanistan	4	Afghanistan	AFG	700.0	QoGBasCSjan24	4.540098	1.0	1.0	...	NaN	
1	8	Albania	8	Albania	ALB	339.0	QoGBasCSjan24	NaN	1.0	8.0	...	2.869328	
2	12	Algeria	12	Algeria	DZA	615.0	QoGBasCSjan24	4.359270	1.0	9.0	...	NaN	
3	20	Andorra	20	Andorra	AND	232.0	QoGBasCSjan24	NaN	1.0	2.0	...	2.034930	2
4	24	Angola	24	Angola	AGO	540.0	QoGBasCSjan24	5.634789	1.0	8.0	...	NaN	

5 rows × 337 columns

```
# I was also interested in looking at GDP like Professor Eirich but with respect to the global peace index.
#Are more peaceful countries generally wealthier than less peaceful ones?

df[['wdi_gdpcapcon2015']].describe()
```

	wdi_gdpcapcon2015
count	188.000000
mean	13726.184306
std	21112.612680
min	263.360992
25%	1939.369629
50%	5395.598633
75%	15333.187744
max	166910.625000

1. Run a simple bivariate regression, and interpret the results. Did the results fit your expectations? Why, why not?

```
#bivariate OLS that accounts for NA values
corruption_gdp = smf.ols(formula = 'wdi_gdpcapcon2015~gpi_gpi', data = df, subset=df['gpi_gpi'].notna()).fit()
print(corruption_gdp.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:      wdi_gdpcapcon2015   R-squared:                       0.267
Model:                            OLS   Adj. R-squared:                  0.262
Method:                 Least Squares   F-statistic:                     55.28
Date:                Sat, 26 Oct 2024   Prob (F-statistic):           7.10e-12
Time:                        17:44:16   Log-Likelihood:                -1700.2
No. Observations:                 154   AIC:                             3404.
Df Residuals:                     152   BIC:                             3411.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   5.347e+04   5619.399      9.515      0.000    4.24e+04    6.46e+04
```

```
gpi_gpi    -2.021e+04   2718.203     -7.435      0.000   -2.56e+04   -1.48e+04
==============================================================================
Omnibus:                       58.537   Durbin-Watson:                   1.791
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              129.142
Skew:                           1.695   Prob(JB):                     9.06e-29
Kurtosis:                       5.938   Cond. No.                         11.6
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

From the results, we see that every unit increase in the global peace index (1 being the most peaceful, and 5 being the least peaceful), GDP drops by approximately 20,210 USD. This is not surprising, as less peaceful countries are often less stable and violence and unrest can lead to the destruction or destabilization of economic and business infrastructure.

```
#quality of governance indicator variable
#scaled from 0 to 1, with 0 being low quality and 1 being the highest quality. Naturally, since this is on the opposite scale
#as the peace index variable, this code will reverse the scales to make sure the multivariate regression is interpretable.
df['new_icrg_qog'] = (1- df['icrg_qog'])*100
df[['new_icrg_qog']].describe()
```

	new_icrg_qog
count	140.000000
mean	45.593585
std	20.090395
min	0.462961
25%	33.333336
50%	50.000000
75%	58.333334
max	94.444444

```
df_filtered = df[['new_icrg_qog', 'gpi_gpi']].dropna()
stats.pearsonr(df_filtered['new_icrg_qog'], df_filtered['gpi_gpi'])
```

```
PearsonRResult(statistic=0.6779091170233515, pvalue=3.180268996950915e-19)
```

The two variables appear to have a moderate to strong positive correlation (0.677), so we will include them in the following multiple regression.

2. Add an additional variable that might mediate or partly "explain" the initial association from that simple regression above -- and explain your results. Did it work out? Yes? No?

```
corruption_actual_gdp = smf.ols(formula = 'wdi_gdpcapcon2015 ~ new_icrg_qog + gpi_gpi', data = df).fit()
print (corruption_actual_gdp.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:      wdi_gdpcapcon2015   R-squared:                       0.690
Model:                            OLS   Adj. R-squared:                  0.685
Method:                 Least Squares   F-statistic:                     141.5
Date:                Sat, 26 Oct 2024   Prob (F-statistic):           4.79e-33
Time:                        17:44:16   Log-Likelihood:                -1386.2
No. Observations:                 130   AIC:                             2778.
Df Residuals:                     127   BIC:                             2787.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     4.994e+04   4248.627     11.755      0.000    4.15e+04    5.83e+04
new_icrg_qog  -794.0024     62.600    -12.684      0.000    -917.876    -670.129
gpi_gpi        472.6273   2725.756      0.173      0.863   -4921.152    5866.407
```

```
==============================================================================
Omnibus:                       42.900   Durbin-Watson:                   1.833
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               98.036
Skew:                           1.349   Prob(JB):                     5.15e-22
Kurtosis:                       6.290   Cond. No.                         265.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

We see that gpi_gpi is no longer significant, and the R^2 value has jumped significantly from 0.267 to 0.690. This indicates that the quality of governance is probably a mediator for global peace index and gdp. Overall, adding the qog variable seems to have worked out well in that it appears to show a clearer representation of the relationship between global peace index and gdp.

3. More on extreme combinations. Find the top 5 entities that are ranked at the top on one variable and ranked at the bottom on another variable. Interpret your results.

```
df['subjective_happiness_rank'] = df['ess_happy'].rank(ascending=False)
df['global_peace_index_rank'] = df['gpi_gpi'].rank(ascending=True)

extreme_comb = df[['cname', 'subjective_happiness_rank', 'global_peace_index_rank']].copy()
extreme_comb['rank_difference'] = extreme_comb['global_peace_index_rank'] - extreme_comb['subjective_happiness_rank']
extreme_sorted = extreme_comb.sort_values(by='rank_difference', ascending=False).head(5)

extreme_sorted
```

	cname	subjective_happiness_rank	global_peace_index_rank	rank_difference
81	Israel	18.0	141.0	123.0
59	France	13.0	70.0	57.0
114	Montenegro	8.0	53.0	45.0
67	Greece	27.0	58.0	31.0
184	United Kingdom of Great Britain and Northern I...	14.0	43.0	29.0

+ Code + Text

We come to an interesting conclusion here, it seems that many of the happiest countries can also be the least peaceful, such as Israel and France given by the difference in ranks. We also notice this for Montenegro. Even though it is ranked 8th in terms of happiness, it seems to be less peaceful, but considering that the United Kingdom is only 15 ranks above it, it's possible that the scaling of the ranks is skewed in a way that is difficult to interpret from this table.