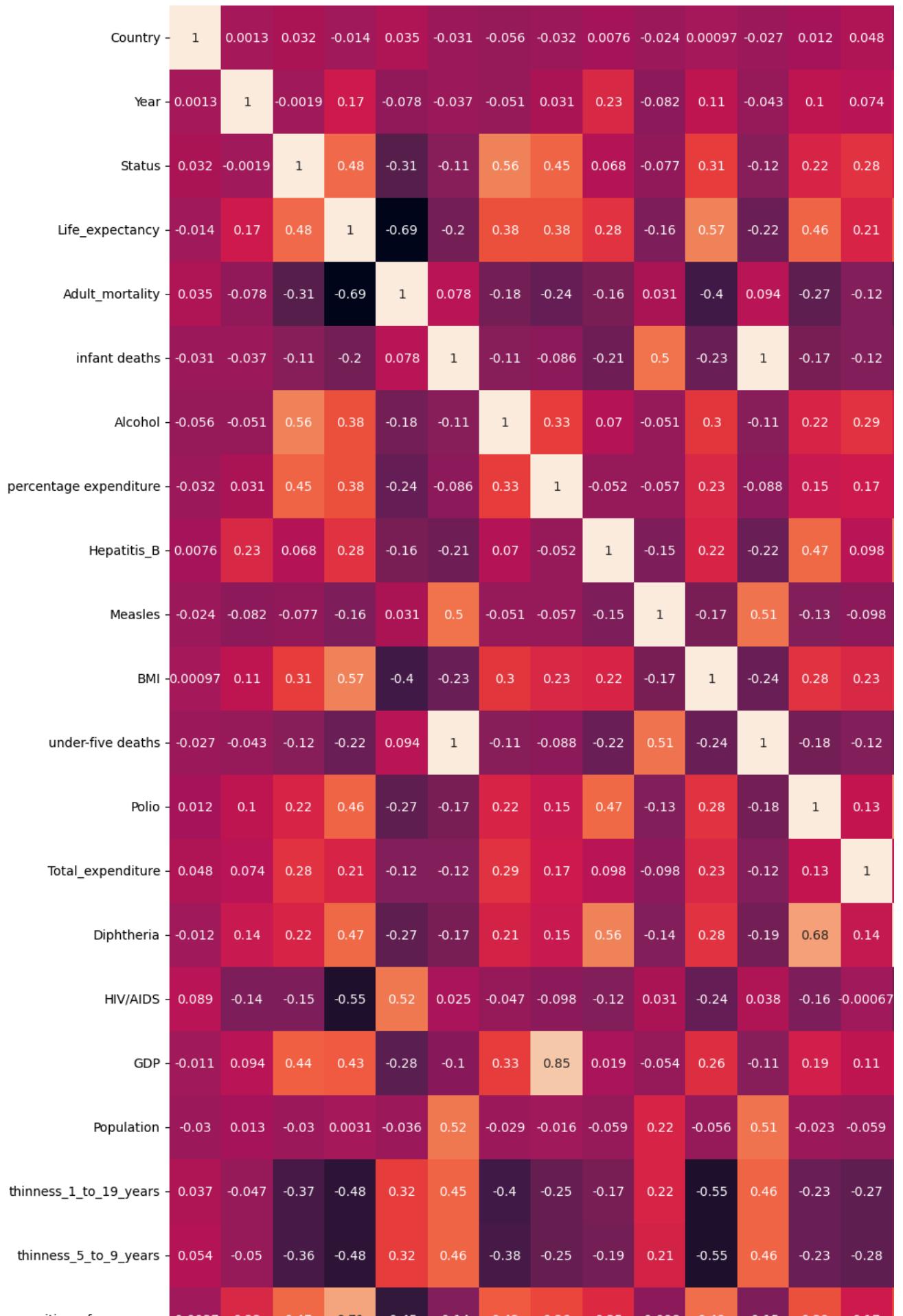
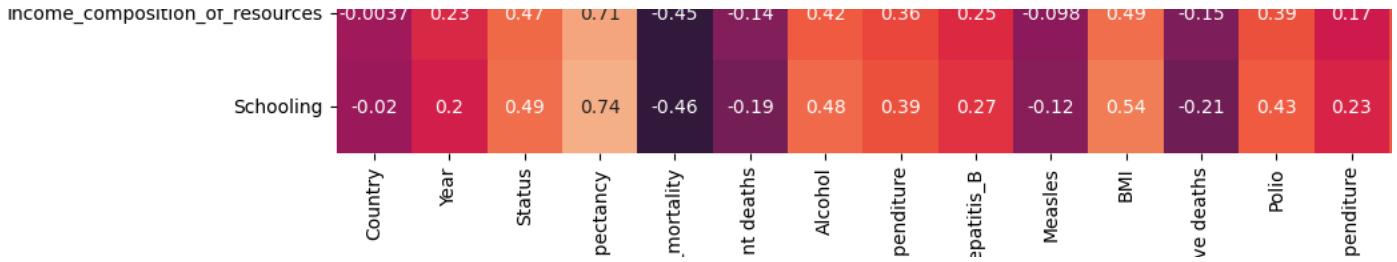


```
%matplotlib inline
import seaborn as sns
import matplotlib.pyplot as plt

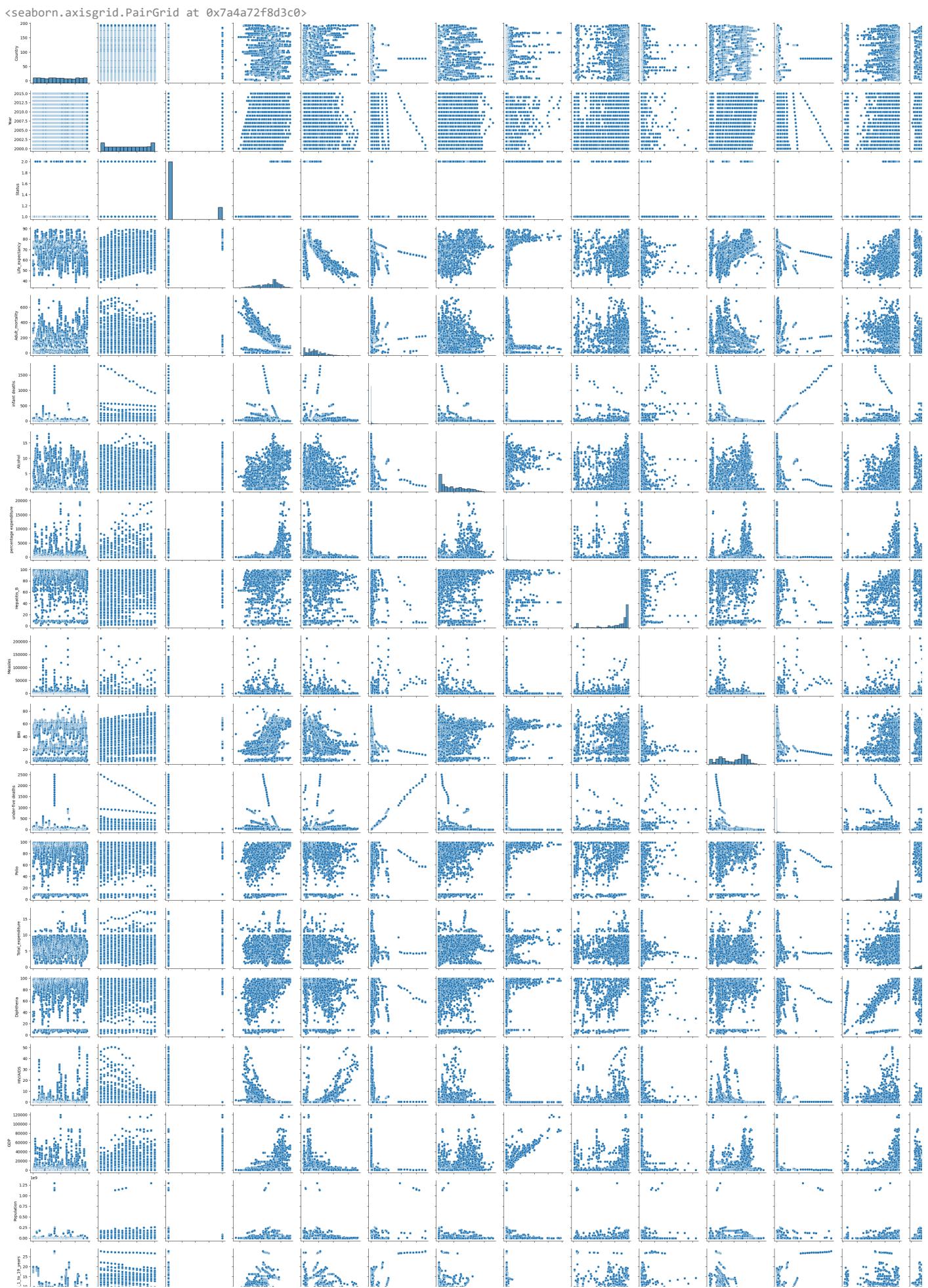
plt.figure(figsize=(20, 20))
sns.heatmap(df.corr(), annot=True)
```

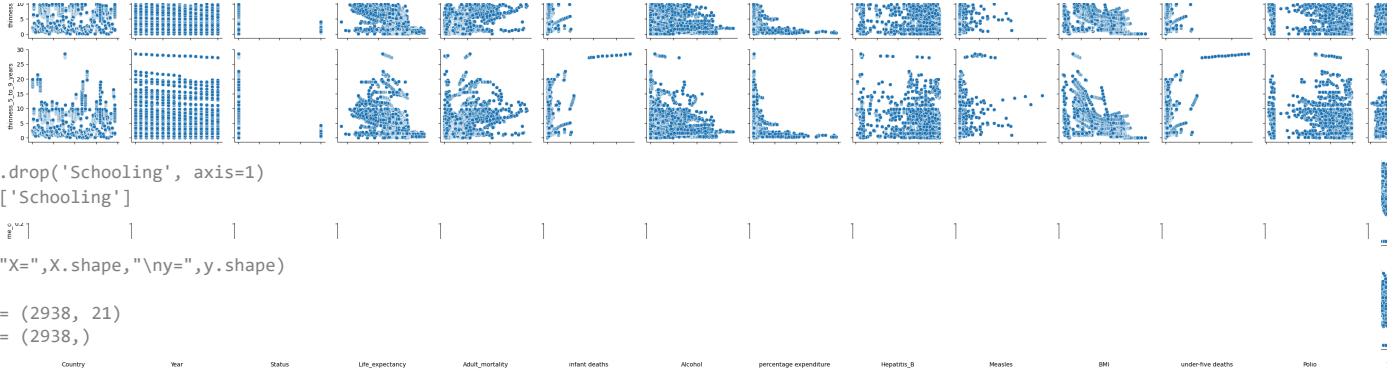
&lt;Axes: &gt;





```
sns.pairplot(df)
```





```
X = df.drop('Schooling', axis=1)
```

```
y = df['Schooling']
```

```
print("X=", X.shape, "\ny=", y.shape)
```

```
X= (2938, 21)
y= (2938,)
```

Country	Year	Status	Life_expectancy	Adult_mortality	infant_deaths	Alcohol	percentage_expenditure	Hepatitis_B	Measles	BMI	underive_deaths	Polio
---------	------	--------	-----------------	-----------------	---------------	---------	------------------------	-------------	---------	-----	-----------------	-------

```
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.linear_model import LinearRegression
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=101)
```

```
X_train.shape
```

```
(2056, 21)
```

```
X_test.shape
```

```
(882, 21)
```

```
model = LinearRegression()
```

```
model.fit(X_train, y_train) #WOW
```

```
▼ LinearRegression
  LinearRegression()
```

```
model.coef_
```

```
array([-1.83391089e-04,  3.65244660e-02,  2.44766274e-01,  1.33924239e-01,
       3.11749369e-04, -2.64554343e-03,  8.83601177e-02,  1.00595956e-04,
      2.95694419e-03,  8.49204866e-06,  1.27277346e-02,  1.50621174e-03,
     4.07145794e-03, -1.98384223e-02, -7.96787574e-04,  8.98069647e-02,
    -1.09839049e-05,  1.03451635e-09, -2.85162958e-02, -1.03738233e-02,
      6.47706664e+00])
```

```
pd.DataFrame(model.coef_, X.columns, columns=['Coefficients'])
```

	Coefficients	
Country	-1.833911e-04	
Year	3.652447e-02	
Status	2.447663e-01	
Life_expectancy	1.339242e-01	
Adult_mortality	3.117494e-04	
infant deaths	-2.645543e-03	
Alcohol	8.836012e-02	
percentage expenditure	1.005960e-04	
Hepatitis_B	2.956944e-03	
Measles	8.492049e-06	
BMI	1.272773e-02	
under-five deaths	1.506212e-03	
Polio	4.071458e-03	
Total_expenditure	-1.983842e-02	
Diphtheria	-7.967876e-04	
HIV/AIDS	8.980696e-02	
GDP	-1.098390e-05	
Population	1.034516e-09	
thinness_1_to_19_years	-2.851630e-02	
thinness_5_to_9_years	-1.037382e-02	
Income_composition_of_resources	6.477067e+00	

```
y_pred = model.predict(X_test)

MAE = metrics.mean_absolute_error(y_test, y_pred)
MSE = metrics.mean_squared_error(y_test, y_pred)
RMSE = np.sqrt(MSE)
```

## ✓ Average Error

MAE

1.273686830093209

## ✓ Mean Square Error

MSE

3.0500293905576346

## ✓ Root Square Mean Error

RMSE

1.7464333341292002

## ✓ average of Schooling

```
df['Schooling'].mean()
```

11.935670524166099

- subtracting y test and y pred yields test residual

```
test_residual = y_test - y_pred
```

```
!pip install hvplot
```

```
Collecting hvplot
  Downloading hvplot-0.9.2-py2.py3-none-any.whl (1.8 MB)
    1.8/1.8 MB 9.7 MB/s eta 0:00:00
Requirement already satisfied: bokeh>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from hvplot) (3.3.4)
Requirement already satisfied: colorcet>=2 in /usr/local/lib/python3.10/dist-packages (from hvplot) (3.1.0)
Requirement already satisfied: holoviews>=1.11.0 in /usr/local/lib/python3.10/dist-packages (from hvplot) (1.17.1)
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from hvplot) (2.0.3)
Requirement already satisfied: numpy>=1.15 in /usr/local/lib/python3.10/dist-packages (from hvplot) (1.25.2)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from hvplot) (24.0)
Requirement already satisfied: panel>=0.11.0 in /usr/local/lib/python3.10/dist-packages (from hvplot) (1.3.8)
Requirement already satisfied: param<3.0,>=1.12.0 in /usr/local/lib/python3.10/dist-packages (from hvplot) (2.1.0)
Requirement already satisfied: Jinja2>=2.9 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (3.1.3)
Requirement already satisfied: contourpy>=1 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (1.2.1)
Requirement already satisfied: pillow>=7.1.0 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (9.4.0)
Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (6.0.1)
Requirement already satisfied: tornado>=5.1 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (6.3.3)
Requirement already satisfied: xyzservices>=2021.09.1 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (2024.4.0)
Requirement already satisfied: pyviz-commits>=0.7.4 in /usr/local/lib/python3.10/dist-packages (from holoviews>=1.11.0->hvplot) (3.0.2)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages (from pandas->hvplot) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->hvplot) (2023.4)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-packages (from pandas->hvplot) (2024.1)
Requirement already satisfied: markdown in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (3.6)
Requirement already satisfied: markdown-it-py in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (3.0.0)
Requirement already satisfied: linkify-it-py in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (2.0.3)
Requirement already satisfied: mdit-py-plugins in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (0.4.0)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (2.31.0)
Requirement already satisfied: tqdm>=4.48.0 in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (4.66.2)
Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (6.1.0)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (4.11.0)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from Jinja2>=2.9->bokeh>=1.0.0->hvplot) (2.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2->pandas->hvplot) (1.16.0)
Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-packages (from bleach->panel>=0.11.0->hvplot) (0.5.1)
Requirement already satisfied: uc-micro-py in /usr/local/lib/python3.10/dist-packages (from linkify-it-py->panel>=0.11.0->hvplot) (1.0.3)
Requirement already satisfied: mdurl~0.1 in /usr/local/lib/python3.10/dist-packages (from markdown-it-py->panel>=0.11.0->hvplot) (0.1.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->panel>=0.11.0->hvplot)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->panel>=0.11.0->hvplot) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests->panel>=0.11.0->hvplot) (2.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->panel>=0.11.0->hvplot) (202)
Installing collected packages: hvplot
Successfully installed hvplot-0.9.2
```

```
import hvplot.pandas
```

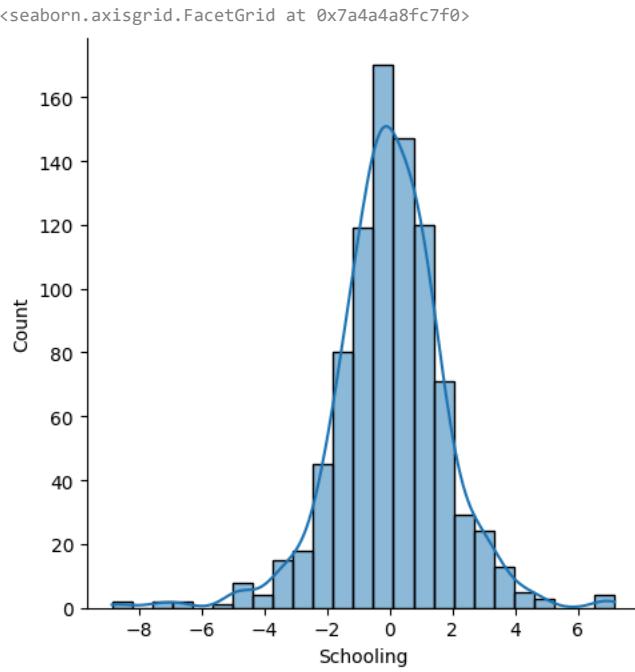
- For some reason, my hvplot does not work, but it has no errors

```
pd.DataFrame({'Error Values': (test_residual)}).hvplot.kde()
```

Notice that this displot has resemblance of a normal distribution, the plot is useful for visualizing the distribution of a dataset and understanding where most of the data

- points lie in terms of the 'Schooling' variable. The bell-shaped curve suggests a normal distribution, indicating that most data points are around the mean value, with fewer occurrences as you move away from the center.

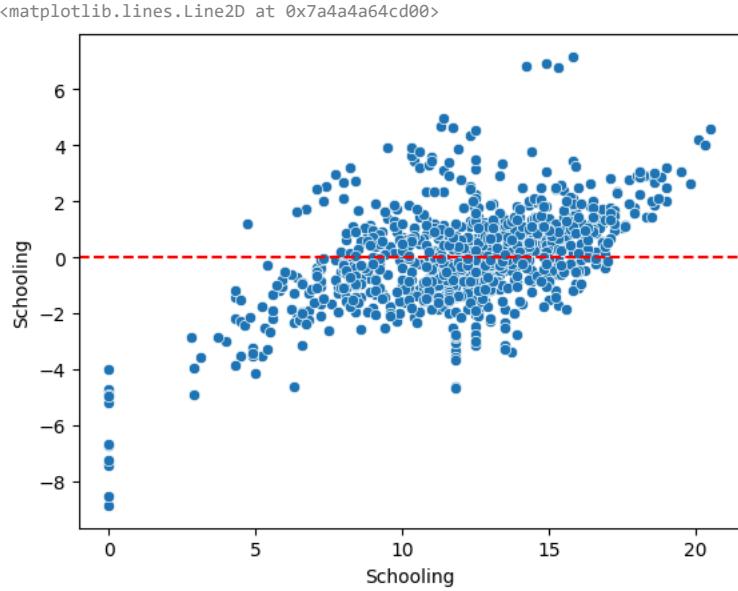
```
sns.displot(test_residual,bins=25, kde=True)
```



This usage of the scatter plot has its similarity with the earlier displot. It could be seen

- ✓ that the values are close to 0 so my explanation for the dependent variable Schooling stands.

```
sns.scatterplot(x=y_test, y=test_residual)
plt.axhline(y=0, color='r', ls='--')
```



In conclusion, the selected dependent variable which is "Schooling" could be predicted

- ✓ successfully using all the other independent variables. Quite useful to implement the Linear Regression process in this environment for front-end and back-end development.