

```
import pandas as pd
import numpy as np

df = pd.read_csv('/content/Dry_Eye_Dataset_with_missing.csv')

num_cols = df.select_dtypes(include=np.number).columns
cat_cols = df.select_dtypes(exclude=np.number).columns

for c in num_cols:
    df[c] = df[c].fillna(df[c].median())

for c in cat_cols:
    df[c] = df[c].fillna(df[c].mode()[0])

df.isnull().sum()
```

	0
Gender	0
Age	0
Sleep duration	0
Sleep quality	0
Stress level	0
Blood pressure	0
Heart rate	0
Daily steps	0
Physical activity	0
Height	0
Weight	0
Sleep disorder	0
Wake up during night	0
Feel sleepy during day	0
Caffeine consumption	0
Alcohol consumption	0
Smoking	0
Medical issue	0
Ongoing medication	0
Smart device before bed	0
Average screen time	0
Blue-light filter	0
Discomfort Eye-strain	0
Redness in eye	0
Itchiness/Irritation in eye	0
Dry Eye Disease	0

dtype: int64

```
import pandas as pd
import numpy as np

df = pd.read_csv('/content/Dry_Eye_Dataset_with_missing.csv')

num_cols = df.select_dtypes(include=np.number).columns

for c in num_cols:
    Q1 = df[c].quantile(0.25)
    Q3 = df[c].quantile(0.75)
    IQR = Q3 - Q1
    lower = Q1 - 1.5 * IQR
    upper = Q3 + 1.5 * IQR
    df[c] = df[c].clip(lower, upper)

df.head()
```

	Gender	Age	Sleep duration	Sleep quality	Stress level	Blood pressure	Heart rate	Daily steps	Physical activity	Height	...	Smoking	Medical issue	Ongoing medication	Sm dev bef
0	F	24.0	9.5	2.0	1.0	137/89	67.0	3000.0	31.0	161.0	...	N	Y	Y	
1	M	39.0	NaN	2.0	3.0	108/64	60.0	12000.0	74.0	164.0	...	N	Y	Y	
2	F	45.0	5.4	1.0	5.0	134/81	95.0	12000.0	93.0	179.0	...	N	N	N	
3	F	NaN	5.4	4.0	5.0	110/90	78.0	19000.0	32.0	160.0	...	N	Y	N	
4	F	42.0	5.7	3.0	2.0	99/67	72.0	4000.0	173.0	179.0	...	N	Y	N	

5 rows × 26 columns

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

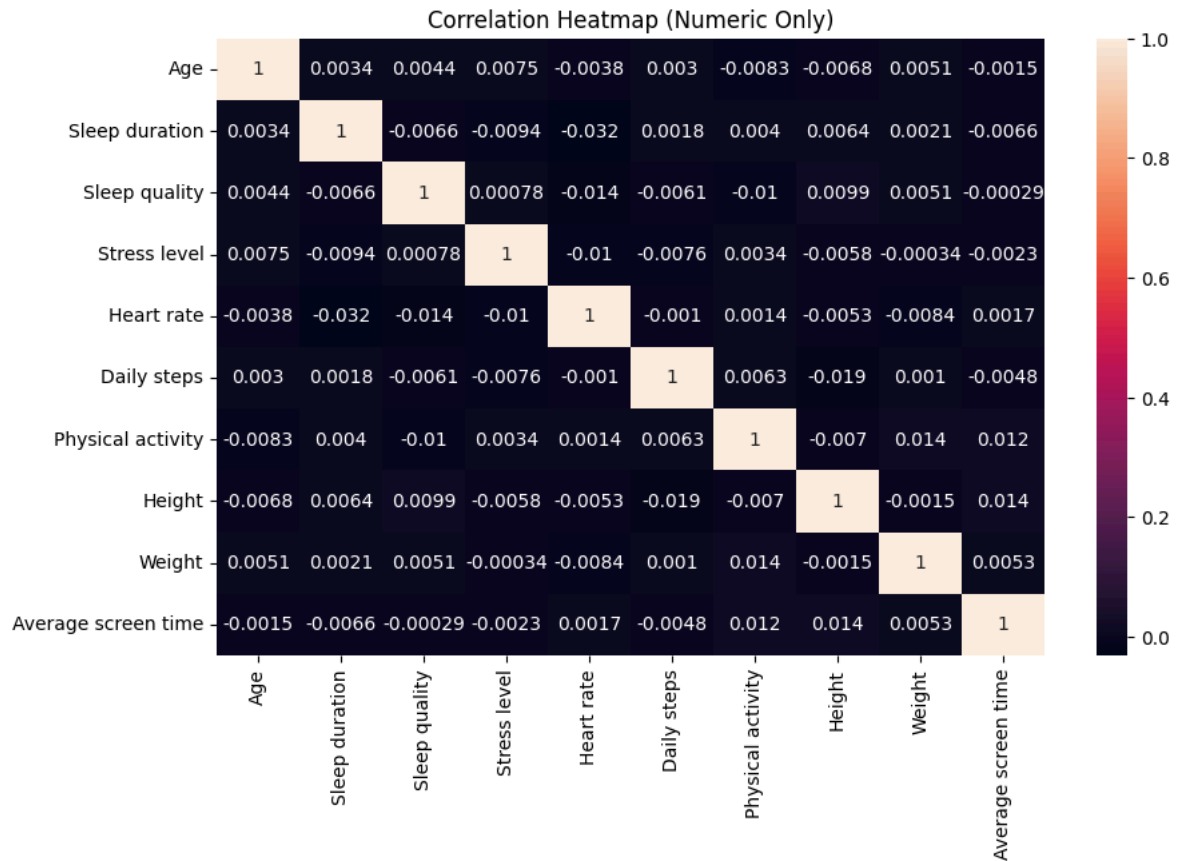
df = pd.read_csv('/content/Dry_Eye_Dataset_with_missing.csv')

num_df = df.select_dtypes(include=[np.number])

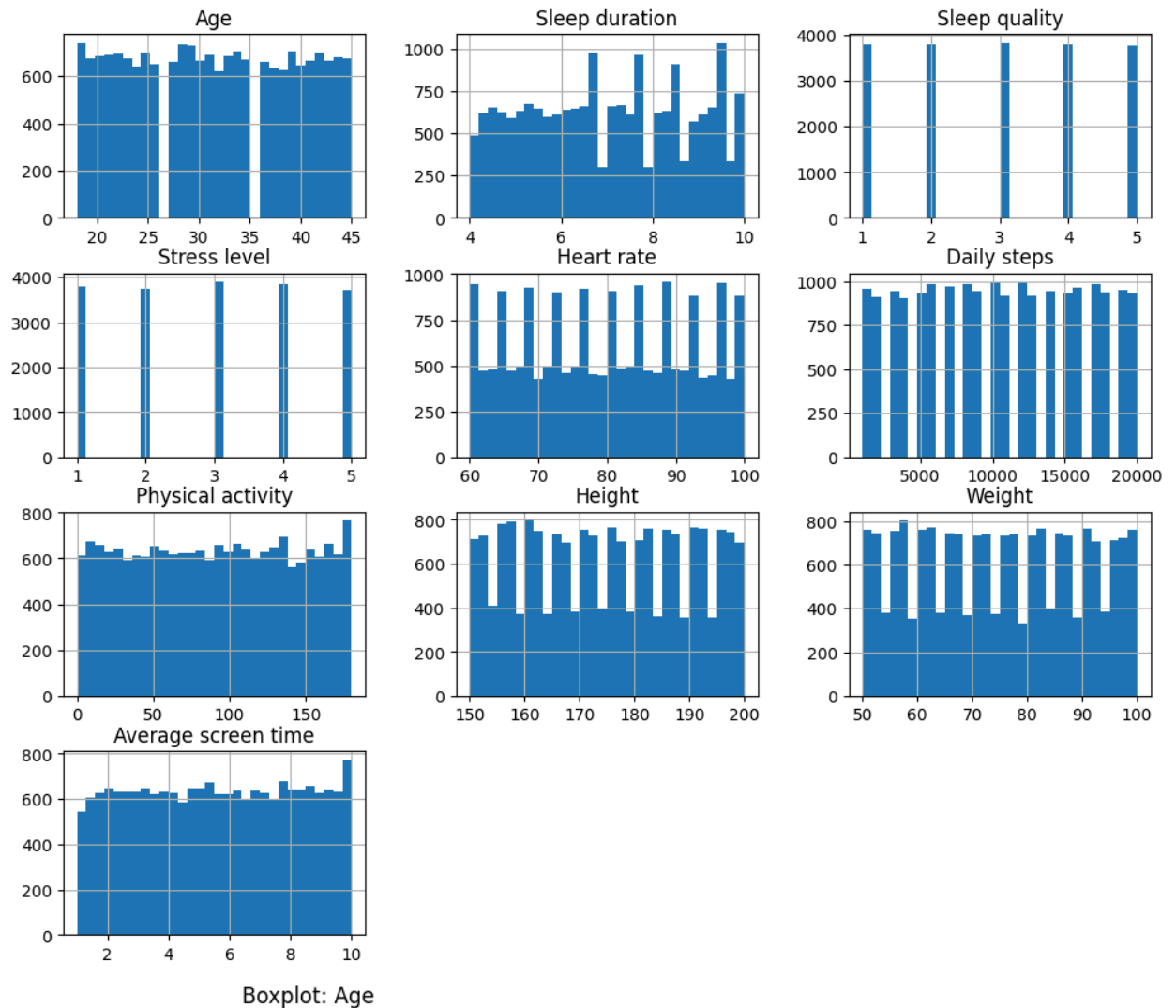
plt.figure(figsize=(10,6))
sns.heatmap(num_df.corr(), annot=True)
plt.title("Correlation Heatmap (Numeric Only)")
plt.show()

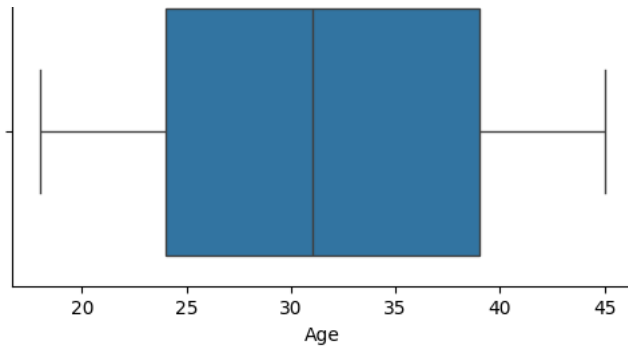
num_df.hist(figsize=(12,10), bins=30)
plt.suptitle("Histograms (Numeric Only)")
plt.show()

for c in num_df.columns[:12]:
    plt.figure(figsize=(6,3))
    sns.boxplot(x=num_df[c])
    plt.title(f"Boxplot: {c}")
    plt.show()
```

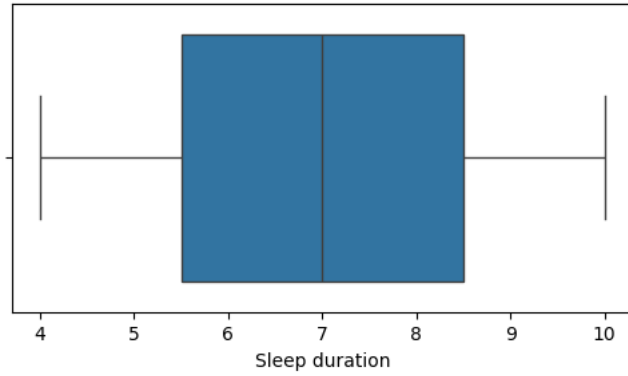



Histograms (Numeric Only)

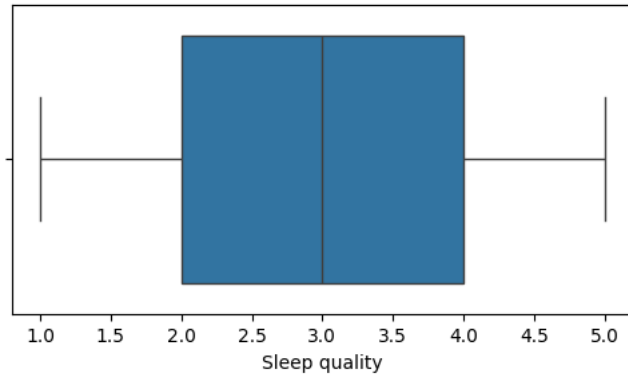




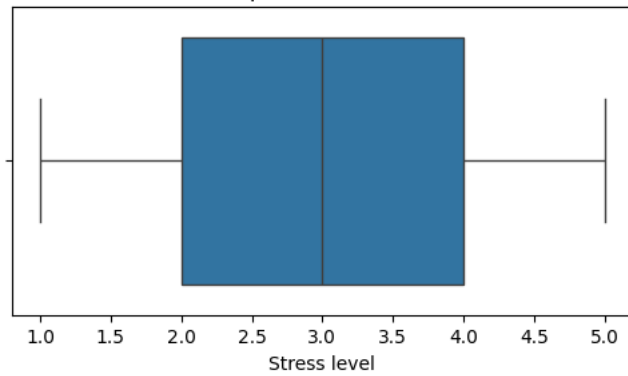
Boxplot: Sleep duration



Boxplot: Sleep quality



Boxplot: Stress level



Boxplot: Heart rate

