```python
import pandas as pd

# Load the dataset from the uploaded file
df = pd.read_csv("Au_nanoparticle_dataset.csv")

# Create a new dataframe by filtering columns
new_df = df[['N_total', 'N_bulk', 'N_surface', 'R_avg']]

# Display the first 20 samples of the new dataframe
print(new_df.head(20))
```

```
    N_total  N_bulk  N_surface     R_avg
0      1599    1014        585   17.3706
1      1642    1034        608   17.6061
2      4637    3365       1272   25.3692
3      7189    5292       1897   29.7011
4     11004    8508       2496   34.2831
5     13375   10768       2607   36.6334
6     13795   11155       2640   37.0108
7     13947   11304       2643   37.1672
8     14020   11357       2663   37.2103
9     14056   11389       2667   37.2467
10     1534     966        568   17.1107
11     1559     974        585   17.2101
12     4356    3099       1257   24.8253
13     6550    4675       1875   29.0376
14    10175    7748       2427   33.4399
15    12393    9842       2551   35.6950
16    13207   10590       2617   36.4657
17    13543   10937       2606   36.7871
18    13713   11079       2634   36.9333
19    13791   11151       2640   37.0104
```

```python
# Calculate mean, standard deviation, and quartiles for each feature
feature_stats = new_df.describe()
print(feature_stats)
```

```
            N_total        N_bulk    N_surface        R_avg
count   4000.000000   4000.000000   4000.00000  4000.000000
mean    3476.786500   2521.550250    955.23625    20.654363
std     3679.286769   2976.232459    721.87022     7.610716
min      236.000000     89.000000    137.00000     8.528600
25%     1061.000000    618.750000    437.00000    15.160725
50%     1867.000000   1199.000000    666.00000    18.629250
75%     4503.000000   3183.000000   1301.75000    25.525125
max    14277.000000  11580.000000   3190.00000    39.239200
```

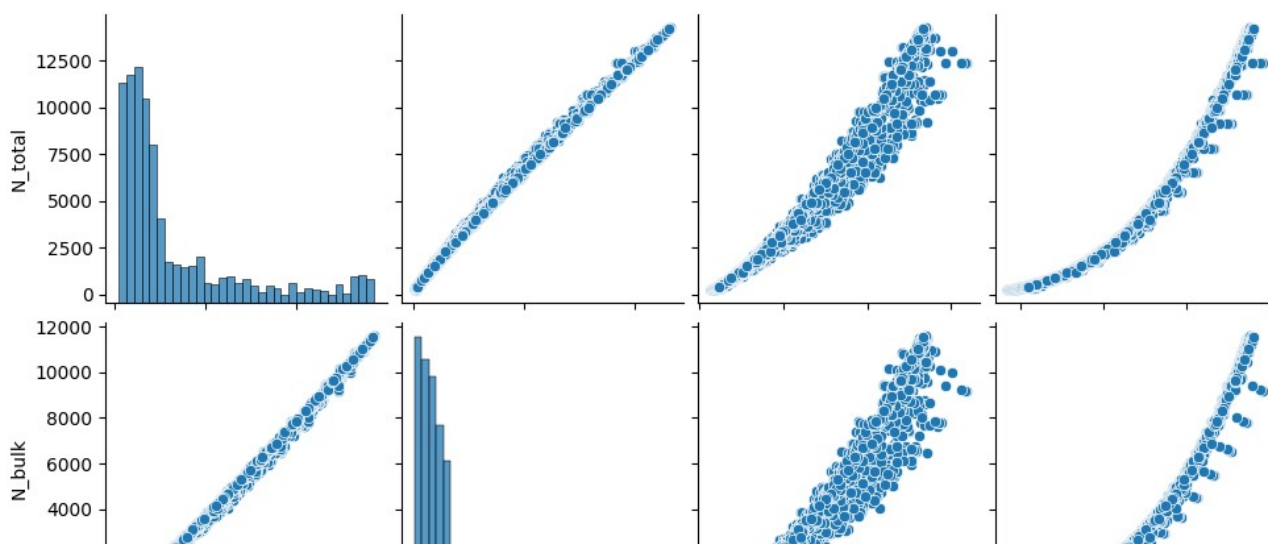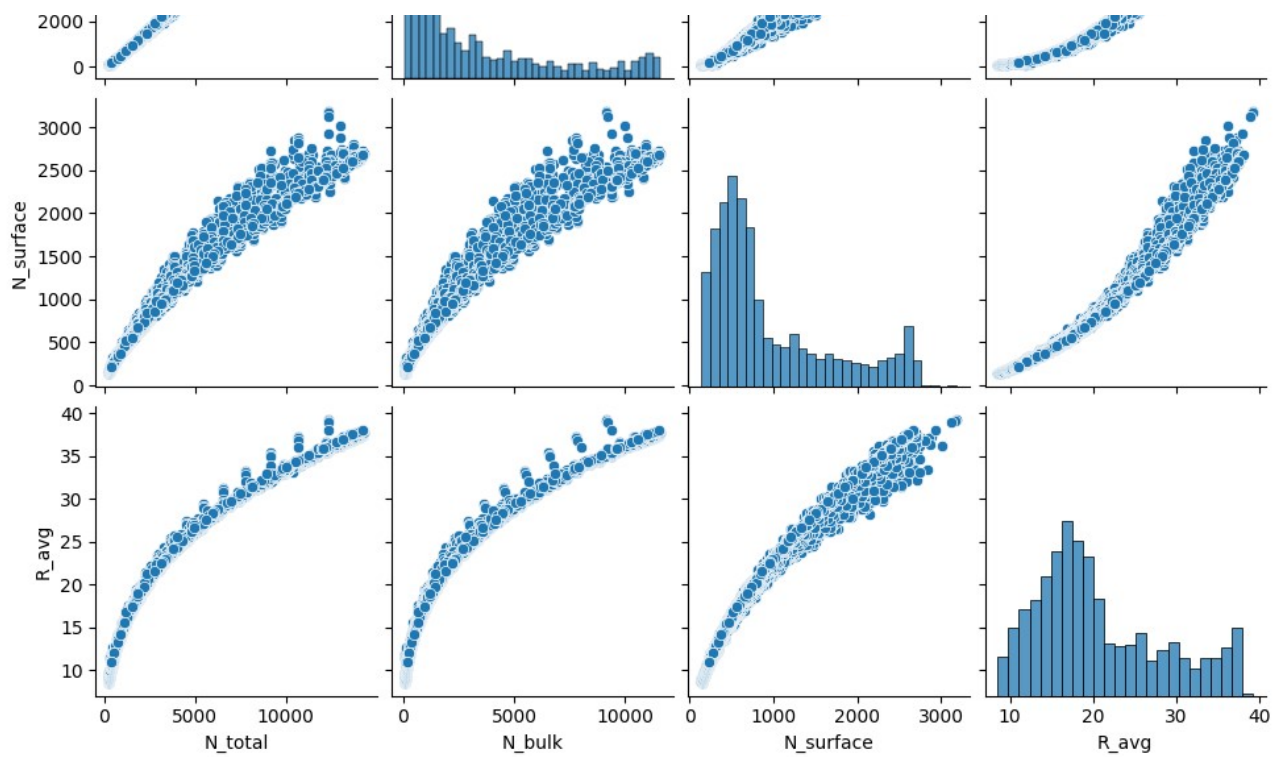✓ 28s   completed at 8:48 PM   ● ✕

```python
# Plot histograms
plt.figure(figsize=(12, 3))
for idx, col in enumerate(new_df.columns):
    plt.subplot(1, 4, idx + 1)
    plt.hist(new_df[col], bins=20, color='blue', alpha=0.7)
    plt.title(col)
plt.tight_layout()
plt.show()
```
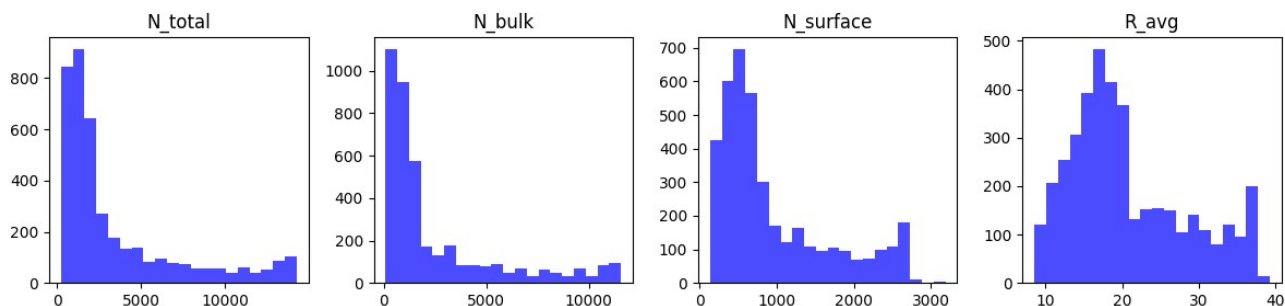


```python
import seaborn as sns

# Visualize scatter plots and histograms using pairplot
sns.pairplot(new_df)
plt.show()
```
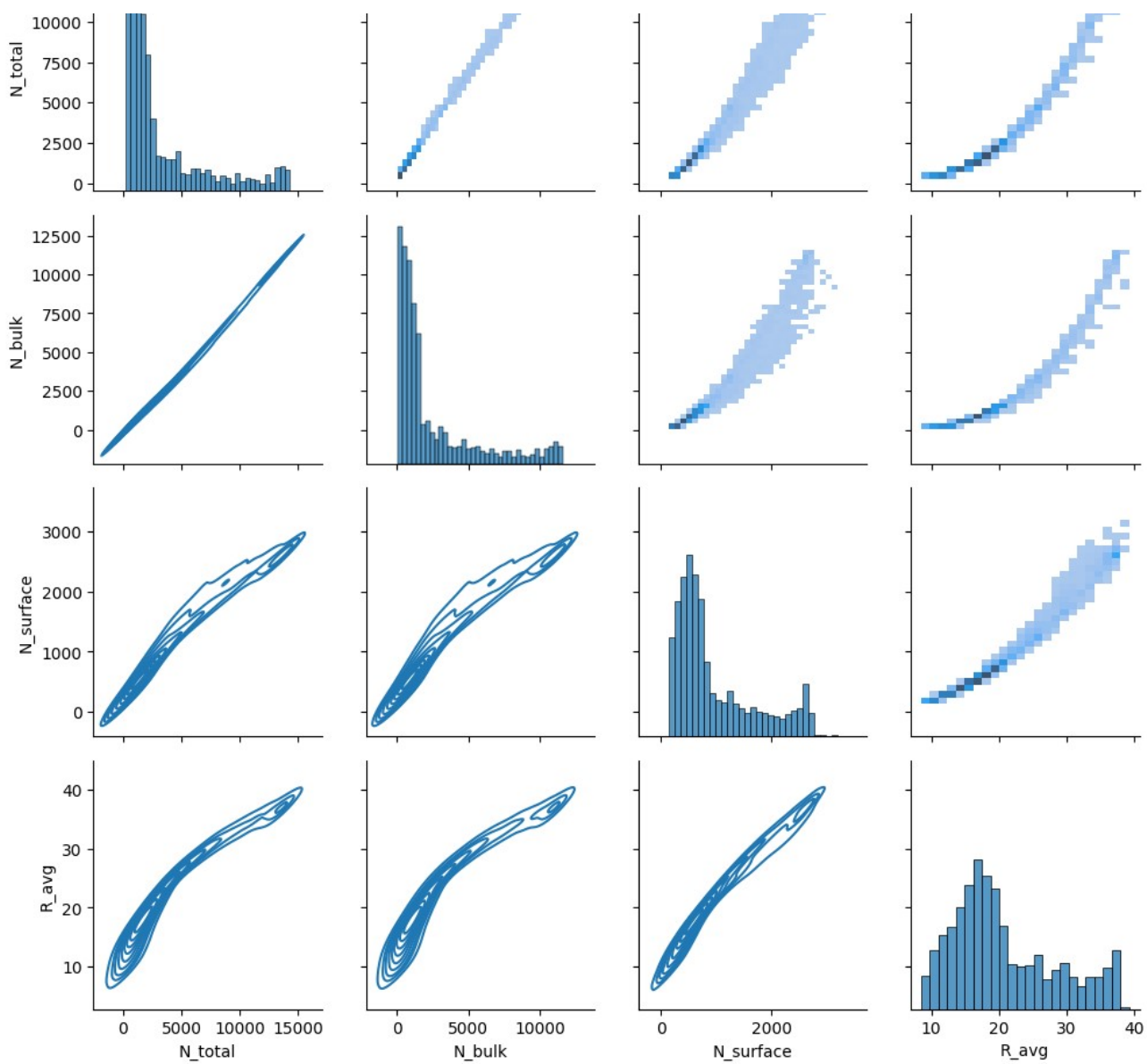
```
# Create a PairGrid with custom plots
g = sns.PairGrid(new_df)
g.map_upper(sns.histplot)   # Bivariate histogram
g.map_diag(sns.histplot)    # Univariate histogram
g.map_lower(sns.kdeplot)    # Bivariate KDE plot

plt.show()
```

Colab paid products - Cancel contracts here