Soil Analysis

Library Import

In [ ]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import math
import os
from scipy import stats
from tabulate import tabulate
from itertools import combinations
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import LabelEncoder
from simple_colors import *
import warnings
warnings.filterwarnings('ignore')
```

Data Loading

In [ ]:

```
soil_nutrient_df = pd.read_csv('Soil Nutrient Composition.csv', index_col='Sample ID.'
soil_particle_df = pd.read_csv('Soil particle analysis.csv', index_col='PARTICLE SIZE
```

In [ ]:

```
soil_nutrient_df.columns = soil_nutrient_df.columns.str.strip()
soil_particle_df.columns = soil_particle_df.columns.str.strip()
```

Data Normalization

In [ ]:

```
numerical_cols = soil_nutrient_df.select_dtypes(include=['float64', 'int64']).columns
scaler = MinMaxScaler()
soil_nutrient_df[numerical_cols] = scaler.fit_transform(soil_nutrient_df[numerical_col
soil_nutrient_df.head()
```

Out[ ]:

	sand	silt	clay	Tex. Class	pH H2O; 1:1	pH KCl	EC	absorbance	ppm(graph)
Sample ID.									
011(27-41)	0.500000	0.857143	0.272727	SANDY LOAM	0.302632	0.345029	0.169683	0.159159	0.159159
2	0.833333	0.428571	0.000000	LOAMY SAND	0.118421	0.140351	0.095023	0.120120	0.120120
17	0.611111	0.857143	0.090909	LOAMY SAND	0.190789	0.140351	0.581448	0.000000	0.000000
013(11-30)	0.555556	0.857143	0.181818	SANDY LOAM	0.302632	0.181287	0.000000	0.048048	0.048048
8	0.833333	0.428571	0.000000	LOAMY SAND	0.151316	0.000000	0.033937	0.102102	0.102102

5 rows × 34 columns

Descriptive Statistics

In [ ]:

```
# Descriptive statistics
soil_nutrient_desc = soil_nutrient_df.describe()
soil_particle_desc = soil_particle_df.describe()

print("\nSoil Nutrient Composition Descriptive Statistics:\n")
soil_nutrient_desc
print("\nSoil Particle Analysis Descriptive Statistics:\n")
soil_particle_desc
```

```
Soil Nutrient Composition Descriptive Statistics:


Soil Particle Analysis Descriptive Statistics:
```

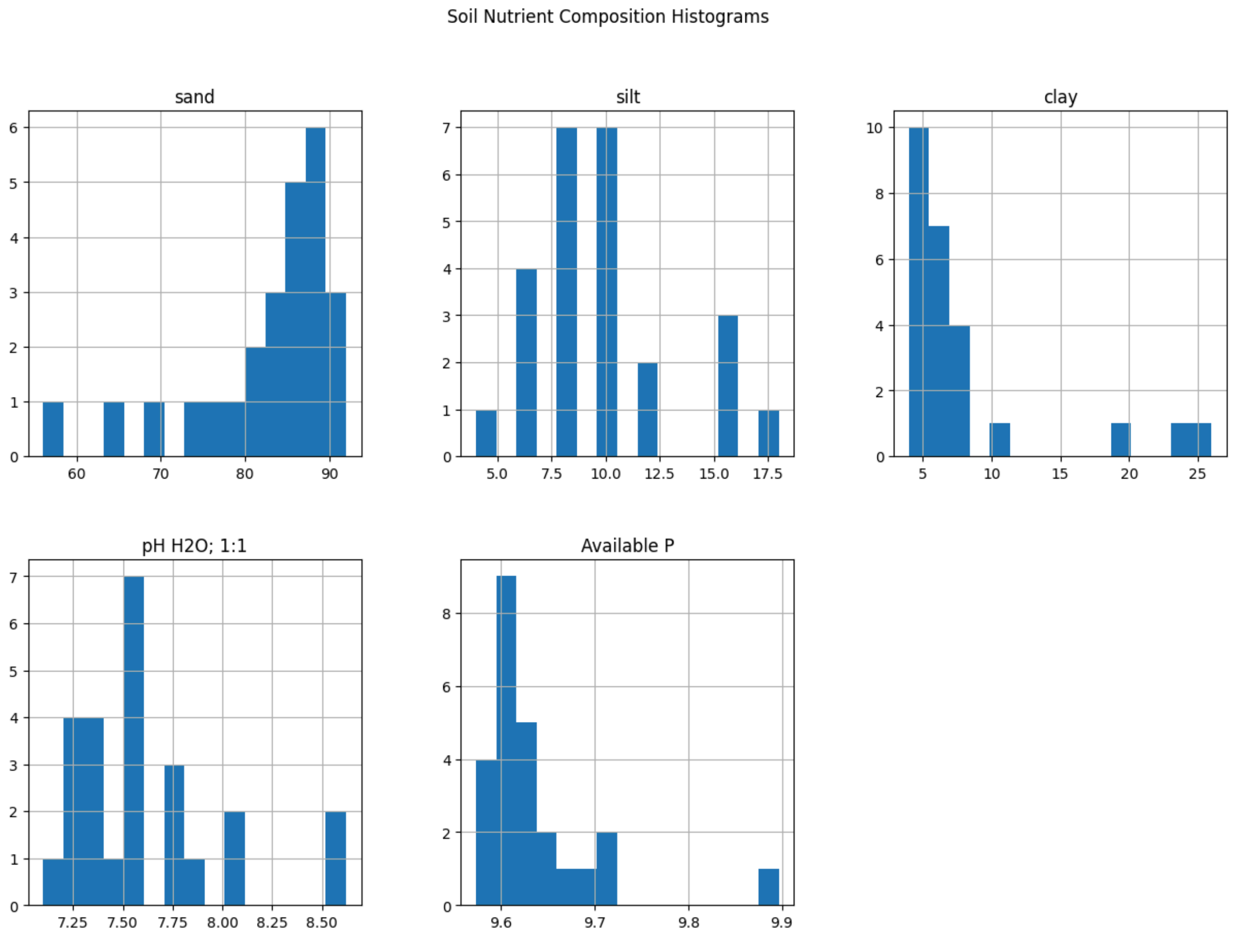
Out[ ]:

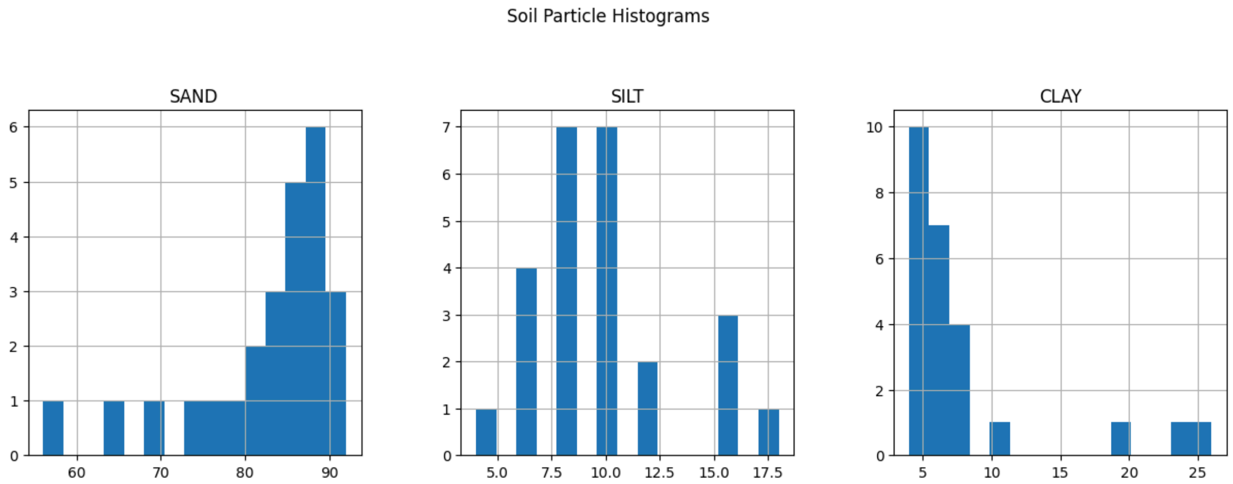
	SAND	SILT	CLAY
count	25.000000	25.00000	25.00000
mean	82.480000	9.76000	7.76000
std	8.761278	3.57398	6.17306
min	56.000000	4.00000	4.00000
25%	82.000000	8.00000	4.00000
50%	86.000000	10.00000	6.00000
75%	88.000000	10.00000	8.00000
max	92.000000	18.00000	26.00000

Virtual Analysis

In [ ]:

```
soil_nutrient_df[['sand', 'silt', 'clay', 'pH H2O; 1:1', 'Available P']].hist(bins=15,
plt.suptitle('Soil Nutrient Composition Histograms')
plt.show()

soil_particle_df.hist(bins=15, figsize=(15, 10), layout=(2, 3))
plt.suptitle('Soil Particle Histograms')
plt.show()
```

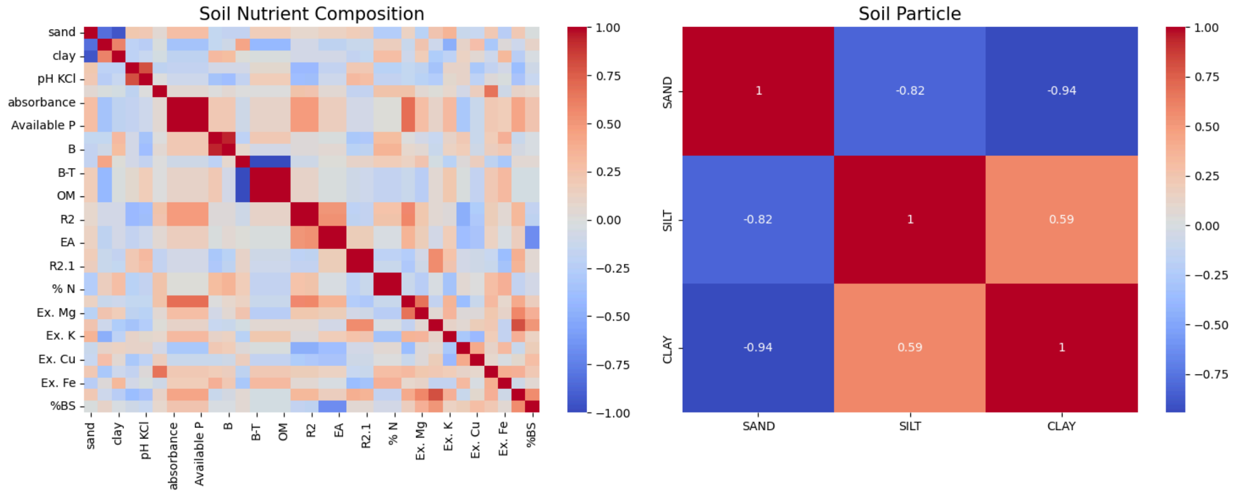




In [ ]:

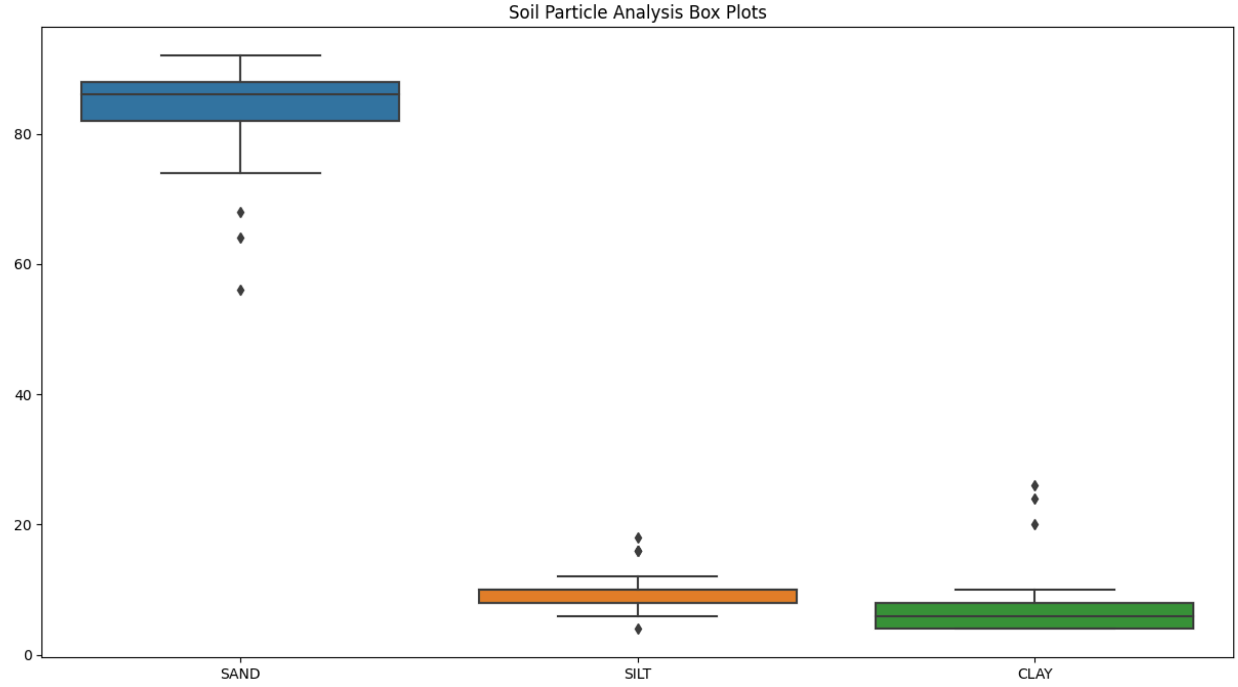
```
soil_nutrient_corr = soil_nutrient_df.drop('Tex. Class', axis = 1).corr()
soil_particle_corr = soil_particle_df.drop('Tex. Class', axis = 1).corr()
```

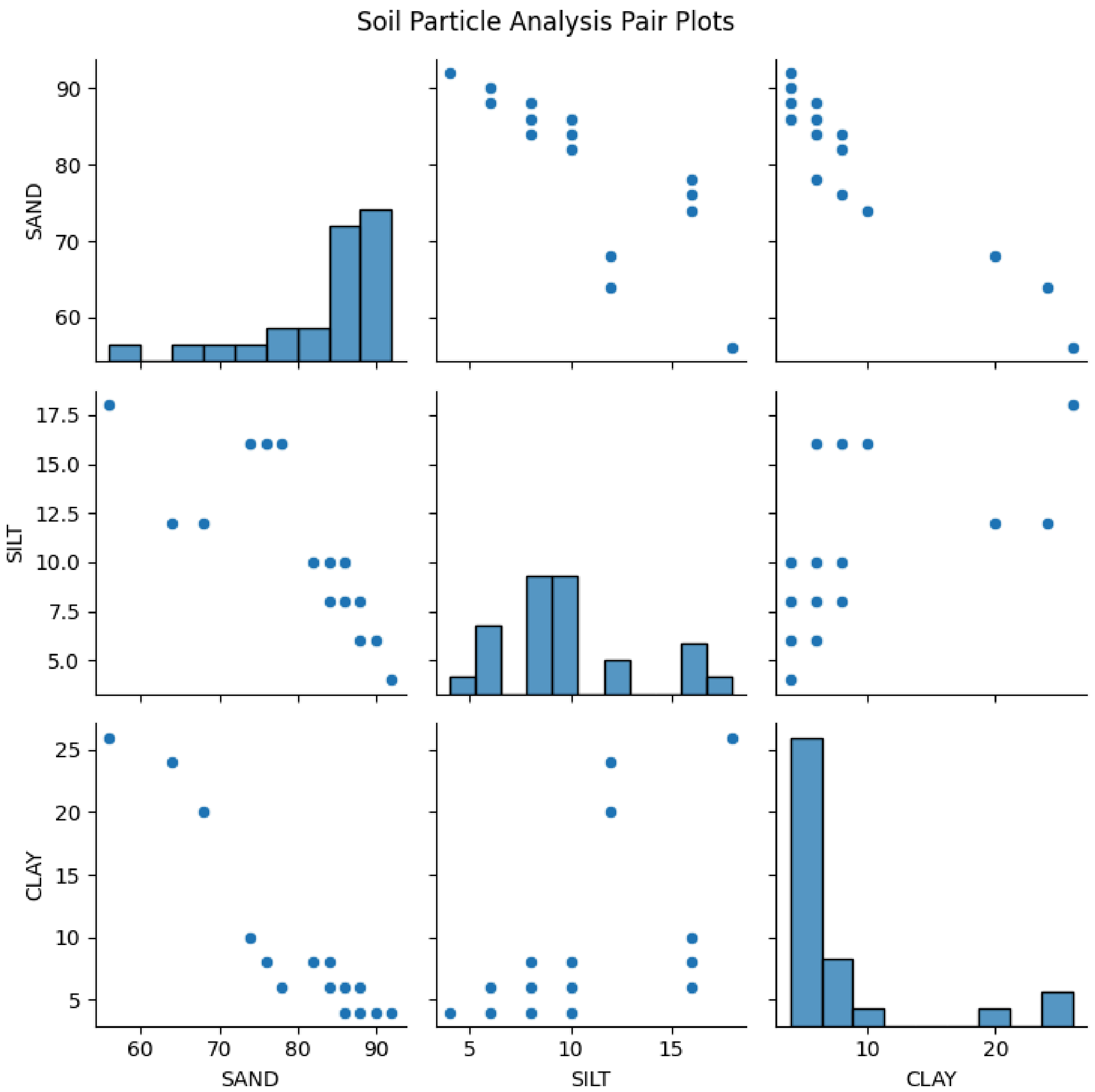
```
plt.figure(figsize=(15, 6))
plt.subplot(1, 2, 1)
sns.heatmap(soil_nutrient_corr, annot=False, cmap='coolwarm')
plt.title('Soil Nutrient Composition', fontsize = 15)

plt.subplot(1, 2, 2)
sns.heatmap(soil_particle_corr, annot=True, cmap='coolwarm')
plt.title('Soil Particle', fontsize = 15)

plt.tight_layout()

plt.show()
```



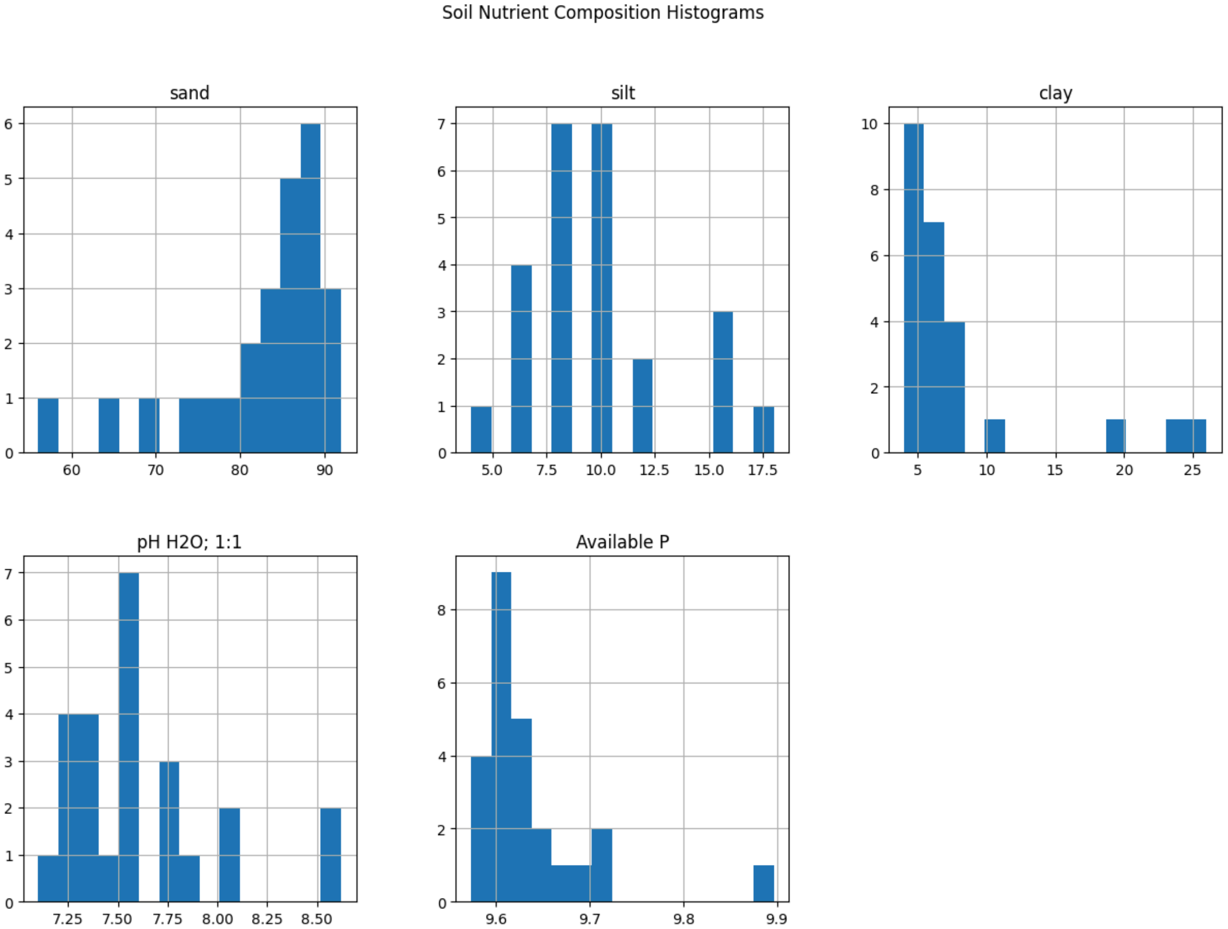
In [ ]:

```
# Box plots for key variables
plt.figure(figsize=(15, 8))
sns.boxplot(data=soil_particle_df)
plt.title('Soil Particle Analysis Box Plots')
plt.show()

sns.pairplot(soil_particle_df)
plt.suptitle('Soil Particle Analysis Pair Plots')
plt.tight_layout()
plt.show()
```


Soil Particle Analysis Box Plots
80
60
40
20
0
SAND
SILT
CLAY



```
In [ ]: # Plot histograms for key variables
        soil_nutrient_df[['sand', 'silt', 'clay', 'pH H2O; 1:1', 'Available P']].hist(bins=15,
        plt.suptitle('Soil Nutrient Composition Histograms')
        plt.show()
```

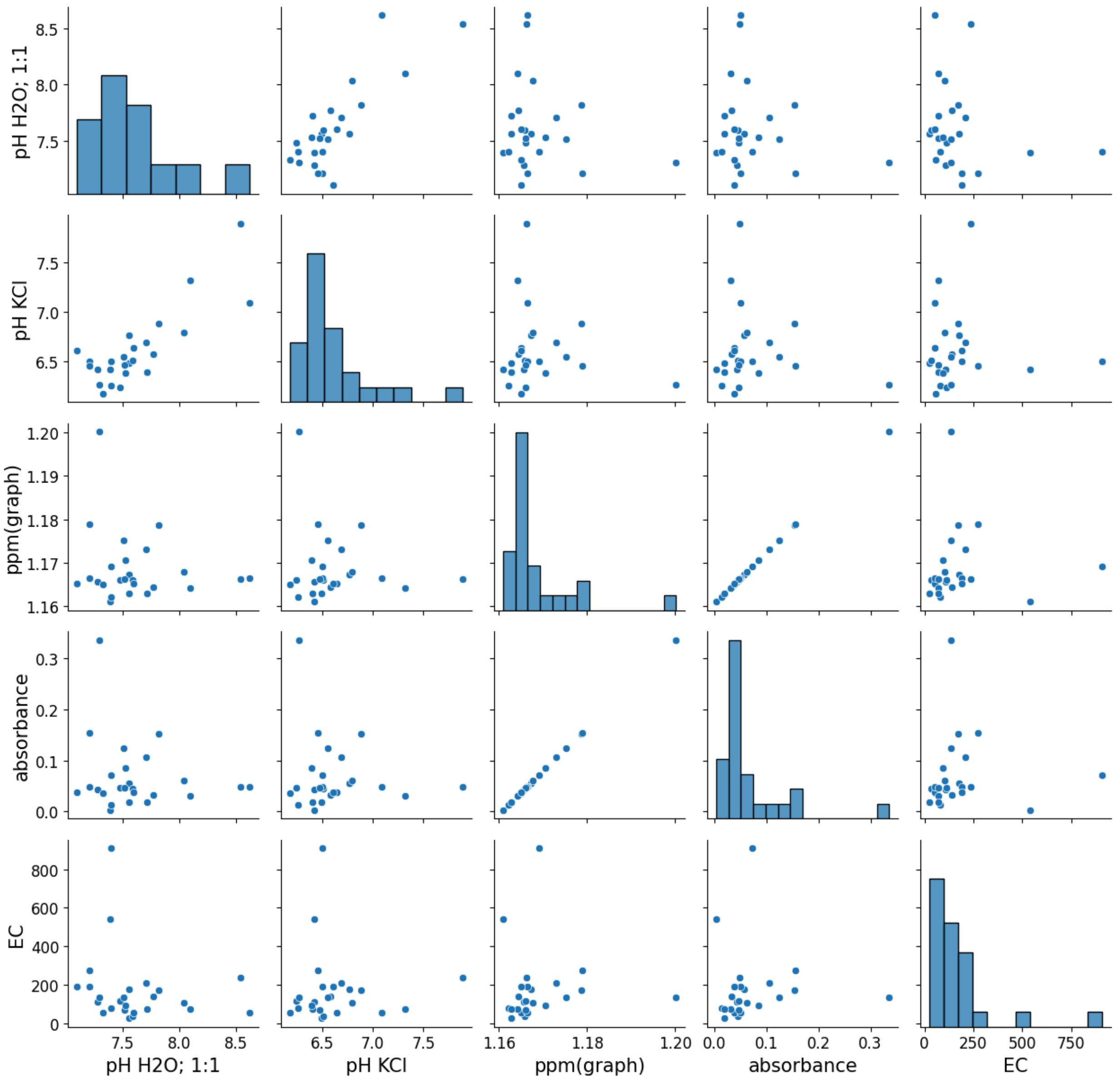


In [ ]:

```python
plt.figure (figsize= (15, 10))
pairplot = sns.pairplot(soil_nutrient_df[['pH H2O; 1:1', 'pH KCl','ppm(graph)', 'absor
plt.suptitle('Soil Nutrient Composition Pair Plots', fontsize = 20, y = 1.02)
plt.tight_layout()
plt.tick_params(axis='x', labelsize=15)
plt.tick_params(axis='y', labelsize=15)
# Adjust the font size of x and y labels
for ax in pairplot.axes.flatten():
    ax.set_xlabel(ax.get_xlabel(), fontsize=15)
    ax.set_ylabel(ax.get_ylabel(), fontsize=15)
    ax.tick_params(axis='x', labelsize=12)
    ax.tick_params(axis='y', labelsize=12)
plt.show()
```

```
<Figure size 1500x1000 with 0 Axes>
```

Soil Nutrient Composition Pair Plots



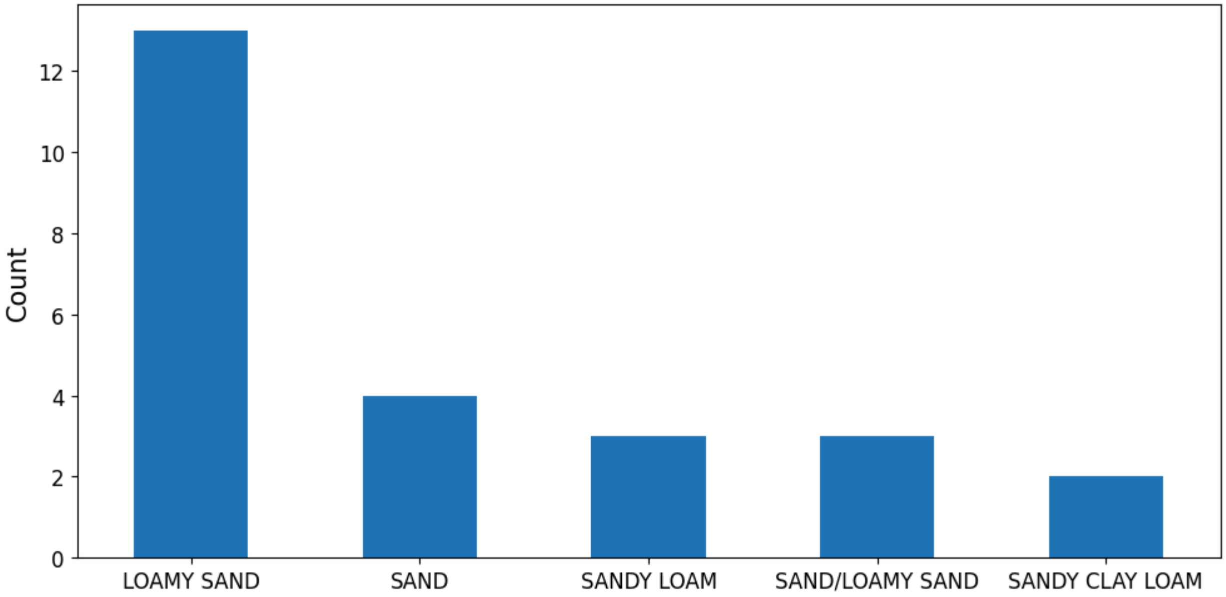
In [ ]:

```
soil_class_counts = soil_nutrient_df['Tex. Class'].value_counts()
file_path = 'Table/soil_class_counts.csv'
soil_class_counts.to_csv(file_path, header=True)

soil_class_counts_df = soil_class_counts.reset_index()
soil_class_counts_df.columns = ['Tex. Class', 'Count']
table = tabulate(soil_class_counts_df, headers='keys', tablefmt='pretty')
print("Soil Class Distribution:")
print(table)
```

```
Soil Class Distribution:
+---+-----------------+-------+
|   |   Tex. Class    | Count |
+---+-----------------+-------+
| 0 |   LOAMY SAND    |  13   |
| 1 |      SAND       |   4   |
| 2 |   SANDY LOAM    |   3   |
| 3 | SAND/LOAMY SAND |   3   |
| 4 | SANDY CLAY LOAM |   2   |
+---+-----------------+-------+
```

In [ ]:

```
plt.figure(figsize=(10, 5))
soil_class_counts.plot(kind='bar', )
#plt.title('Distribution of Soil Classes', fontsize=20)
plt.xlabel('', fontsize=15)
plt.ylabel('Count', fontsize=15)
plt.xticks(rotation=0, fontsize=12)
plt.yticks(fontsize=12)
plt.tight_layout()
plt.show()
```



In [ ]:

```
soil_class_stats = soil_nutrient_df.groupby('Tex. Class').describe()
soil_class_stats
```

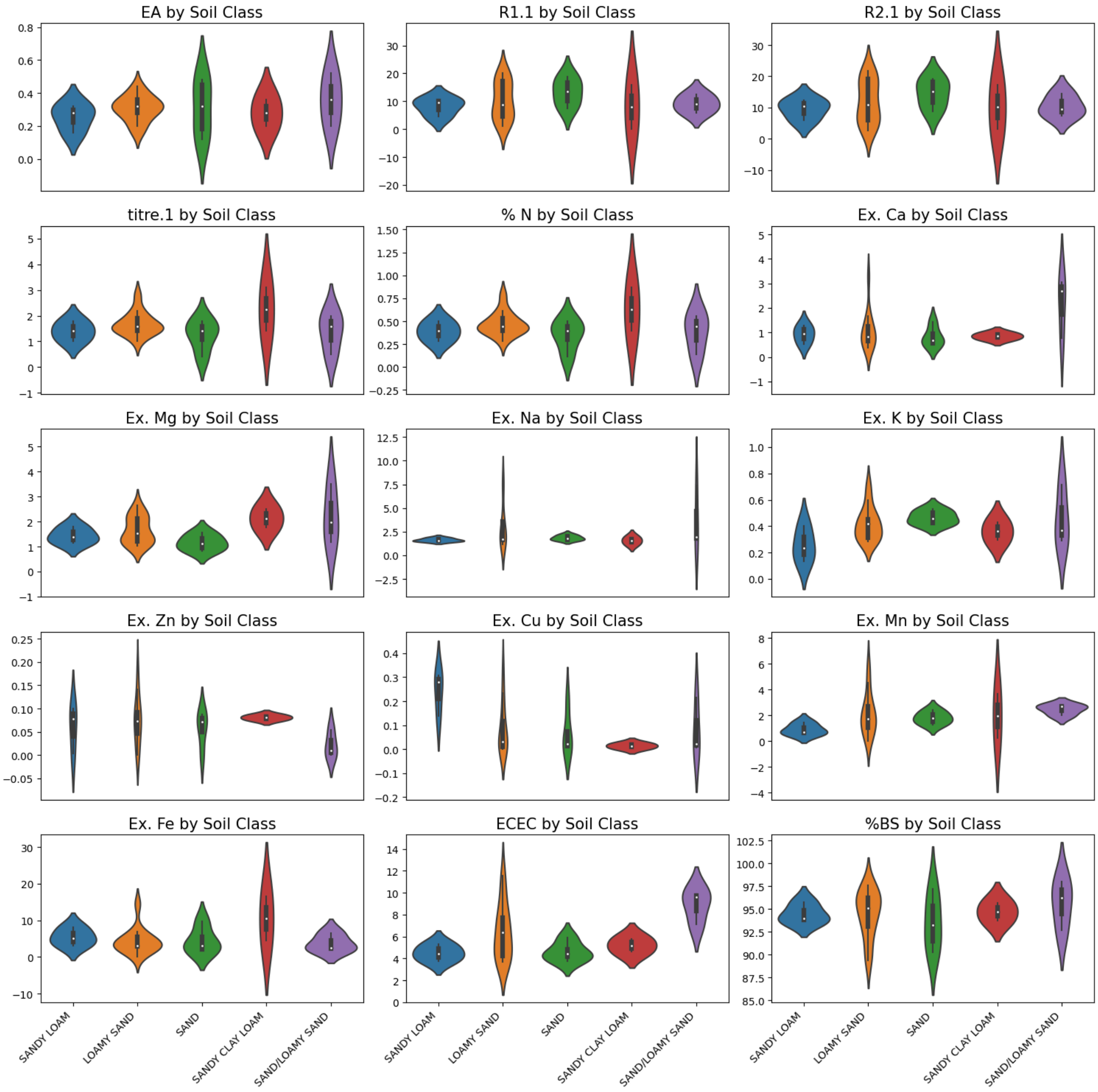
Out[ ]:

	sand								silt		...	
	count	mean	std	min	25%	50%	75%	max	count	mean	...	75
Tex. Class												
LOAMY SAND	13.0	84.615385	2.754949	78.0	84.0	86.0	86.0	88.0	13.0	9.384615	...	7.71231
SAND	4.0	90.000000	1.632993	88.0	89.5	90.0	90.5	92.0	4.0	6.000000	...	4.8610
SAND/LOAMY SAND	3.0	88.000000	0.000000	88.0	88.0	88.0	88.0	88.0	3.0	8.000000	...	9.7519
SANDY CLAY LOAM	2.0	60.000000	5.656854	56.0	58.0	60.0	62.0	64.0	2.0	15.000000	...	5.48841
SANDY LOAM	3.0	72.666667	4.163332	68.0	71.0	74.0	75.0	76.0	3.0	14.666667	...	4.8755

5 rows × 264 columns

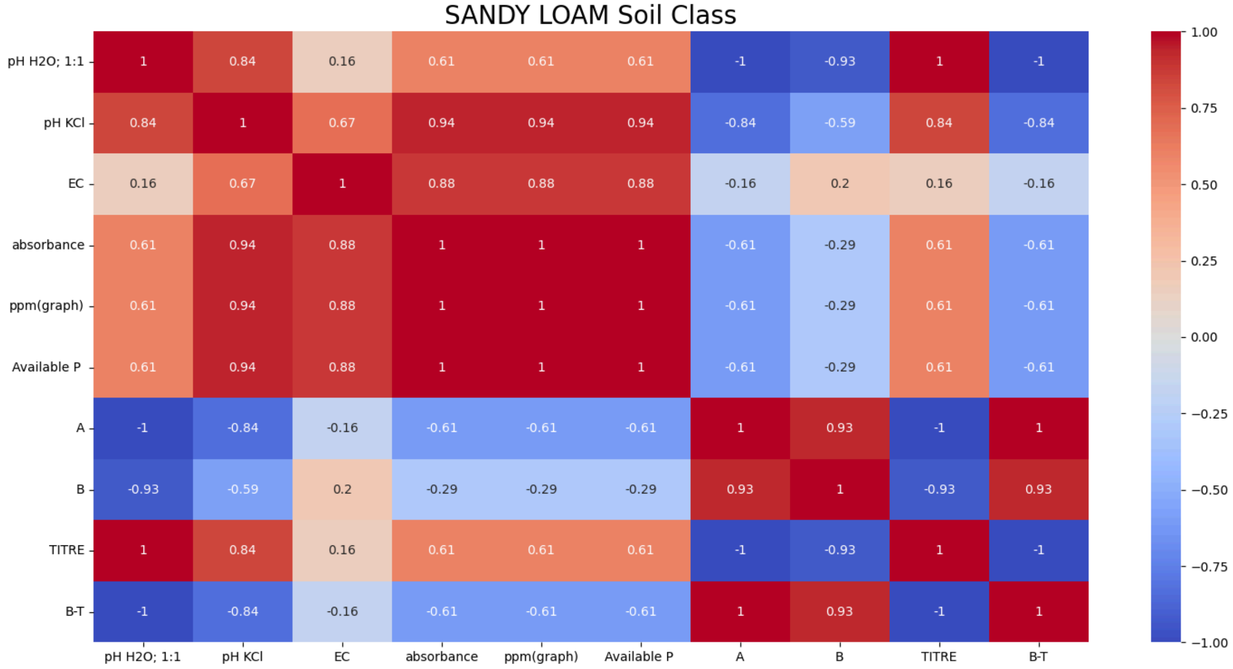
In [ ]:

```
columns_to_plot = soil_nutrient_df.columns[19:]

num_plots = len(columns_to_plot)
num_cols = 3
num_rows = (num_plots + num_cols - 1) // num_cols

plt.figure(figsize=(15, 15))

for idx, col in enumerate(columns_to_plot):
    ax = plt.subplot(num_rows, num_cols, idx + 1)
    plt.subplot(num_rows, num_cols, idx + 1)
    sns.violinplot(x='Tex. Class', y=col, data=soil_nutrient_df, ax=ax)
    plt.title(f'{col} by Soil Class', fontsize=15)
    plt.xticks(rotation = 45)
    plt.xlabel('')
    plt.ylabel('', fontsize=12)
        if idx >= (num_plots - num_cols):
        ax.set_xticklabels(ax.get_xticklabels(), rotation=45, ha='right')
    else:
        ax.set_xlabel('')
        ax.set_xticks([])

plt.tight_layout()
plt.show()
```



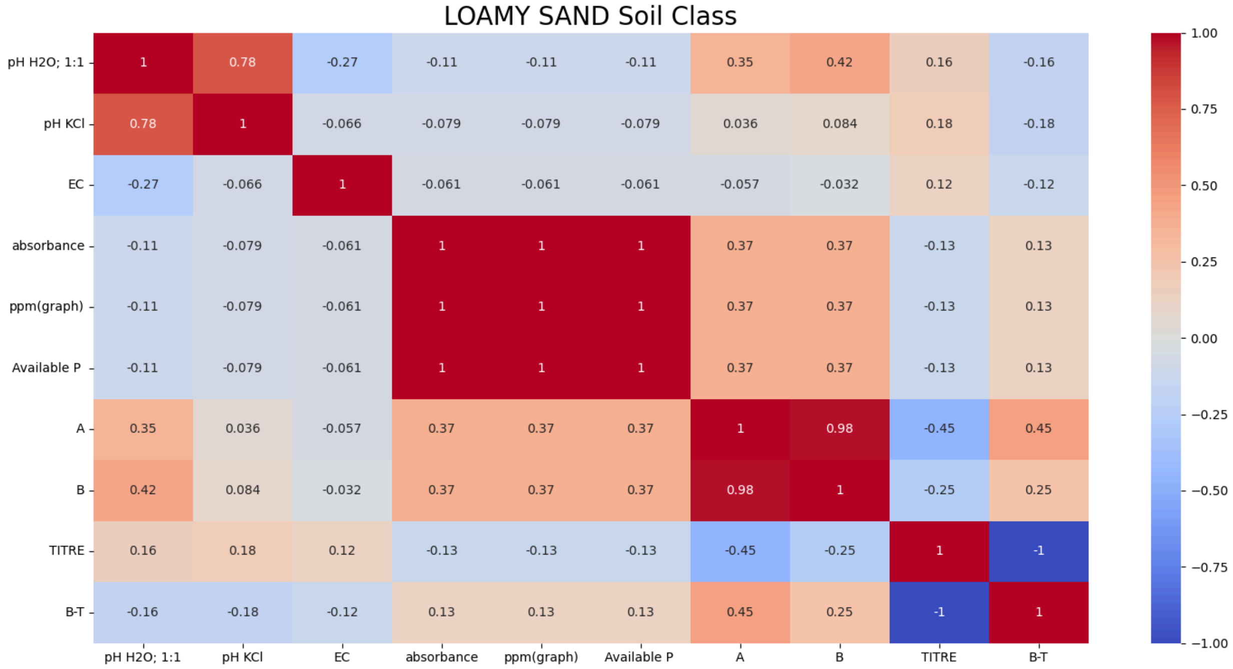
In [ ]:

```
soil_classes = soil_nutrient_df['Tex. Class'].unique()
fig, axes = plt.subplots(len(soil_classes), 1, figsize=(15, len(soil_classes) * 8))

for idx, soil_class in enumerate(soil_classes):
    class_df = soil_nutrient_df[soil_nutrient_df['Tex. Class'] == soil_class]

    class_df_numerical = class_df.select_dtypes(include=['float64', 'int64'])
    class_corr = class_df_numerical.iloc[:, 3:-20].corr()

    sns.heatmap(class_corr, annot=True, cmap='coolwarm', ax=axes[idx])
    axes[idx].set_title(f'{soil_class} Soil Class', fontsize = 20)

plt.tight_layout()
plt.subplots_adjust(wspace=0.2, hspace=0.2)
plt.show()
```



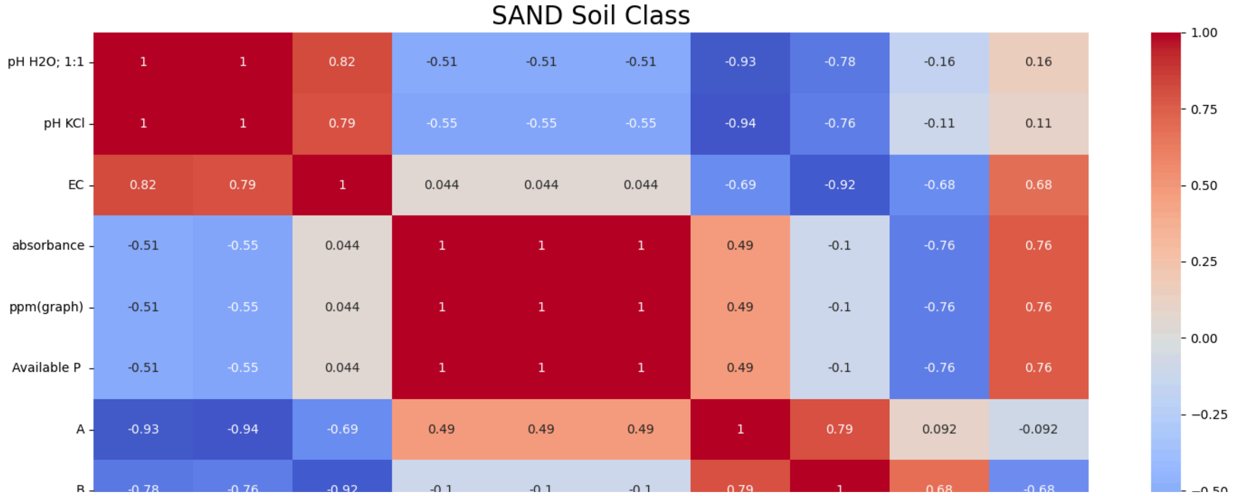
SANDY LOAM Soil Class

	pH H2O; 1:1	pH KCl	EC	absorbance	ppm(graph)	Available P	A	B	TITRE	B-T
pH H2O; 1:1	1	0.84	0.16	0.61	0.61	0.61	-1	-0.93	1	-1
pH KCl	0.84	1	0.67	0.94	0.94	0.94	-0.84	-0.59	0.84	-0.84
EC	0.16	0.67	1	0.88	0.88	0.88	-0.16	0.2	0.16	-0.16
absorbance	0.61	0.94	0.88	1	1	1	-0.61	-0.29	0.61	-0.61
ppm(graph)	0.61	0.94	0.88	1	1	1	-0.61	-0.29	0.61	-0.61
Available P	0.61	0.94	0.88	1	1	1	-0.61	-0.29	0.61	-0.61
A	-1	-0.84	-0.16	-0.61	-0.61	-0.61	1	0.93	-1	1
B	-0.93	-0.59	0.2	-0.29	-0.29	-0.29	0.93	1	-0.93	0.93
TITRE	1	0.84	0.16	0.61	0.61	0.61	-1	-0.93	1	-1
B-T	-1	-0.84	-0.16	-0.61	-0.61	-0.61	1	0.93	-1	1



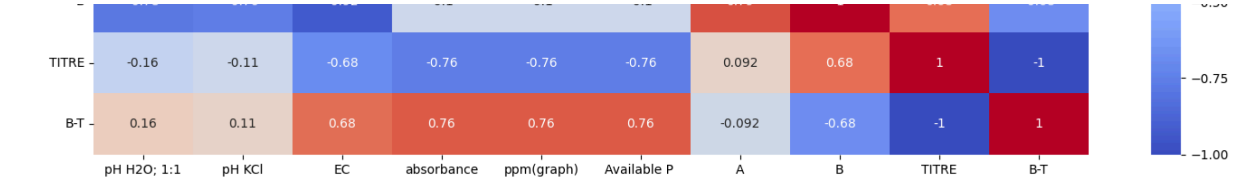
LOAMY SAND Soil Class

	pH H2O; 1:1	pH KCl	EC	absorbance	ppm(graph)	Available P	A	B	TITRE	B-T
pH H2O; 1:1	1	0.78	-0.27	-0.11	-0.11	-0.11	0.35	0.42	0.16	-0.16
pH KCl	0.78	1	-0.066	-0.079	-0.079	-0.079	0.036	0.084	0.18	-0.18
EC	-0.27	-0.066	1	-0.061	-0.061	-0.061	-0.057	-0.032	0.12	-0.12
absorbance	-0.11	-0.079	-0.061	1	1	1	0.37	0.37	-0.13	0.13
ppm(graph)	-0.11	-0.079	-0.061	1	1	1	0.37	0.37	-0.13	0.13
Available P	-0.11	-0.079	-0.061	1	1	1	0.37	0.37	-0.13	0.13
A	0.35	0.036	-0.057	0.37	0.37	0.37	1	0.98	-0.45	0.45
B	0.42	0.084	-0.032	0.37	0.37	0.37	0.98	1	-0.25	0.25
TITRE	0.16	0.18	0.12	-0.13	-0.13	-0.13	-0.45	-0.25	1	-1
B-T	-0.16	-0.18	-0.12	0.13	0.13	0.13	0.45	0.25	-1	1



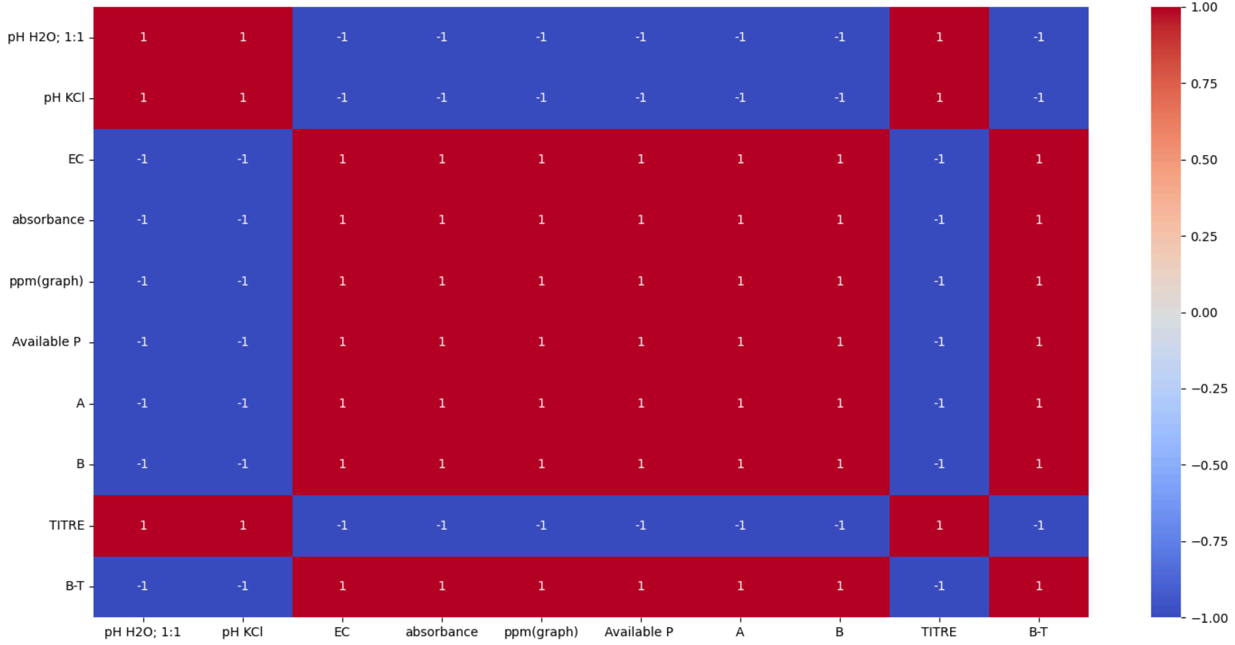
SAND Soil Class

	pH H2O; 1:1	pH KCl	EC	absorbance	ppm(graph)	Available P	A	B	TITRE	B-T
pH H2O; 1:1	1	1	0.82	-0.51	-0.51	-0.51	-0.93	-0.78	-0.16	0.16
pH KCl	1	1	0.79	-0.55	-0.55	-0.55	-0.94	-0.76	-0.11	0.11
EC	0.82	0.79	1	0.044	0.044	0.044	-0.69	-0.92	-0.68	0.68
absorbance	-0.51	-0.55	0.044	1	1	1	0.49	-0.1	-0.76	0.76
ppm(graph)	-0.51	-0.55	0.044	1	1	1	0.49	-0.1	-0.76	0.76
Available P	-0.51	-0.55	0.044	1	1	1	0.49	-0.1	-0.76	0.76
A	-0.93	-0.94	-0.69	0.49	0.49	0.49	1	0.79	0.092	-0.092
B	-0.78	-0.76	-0.92	-0.1	-0.1	-0.1	0.79	1	0.68	-0.68



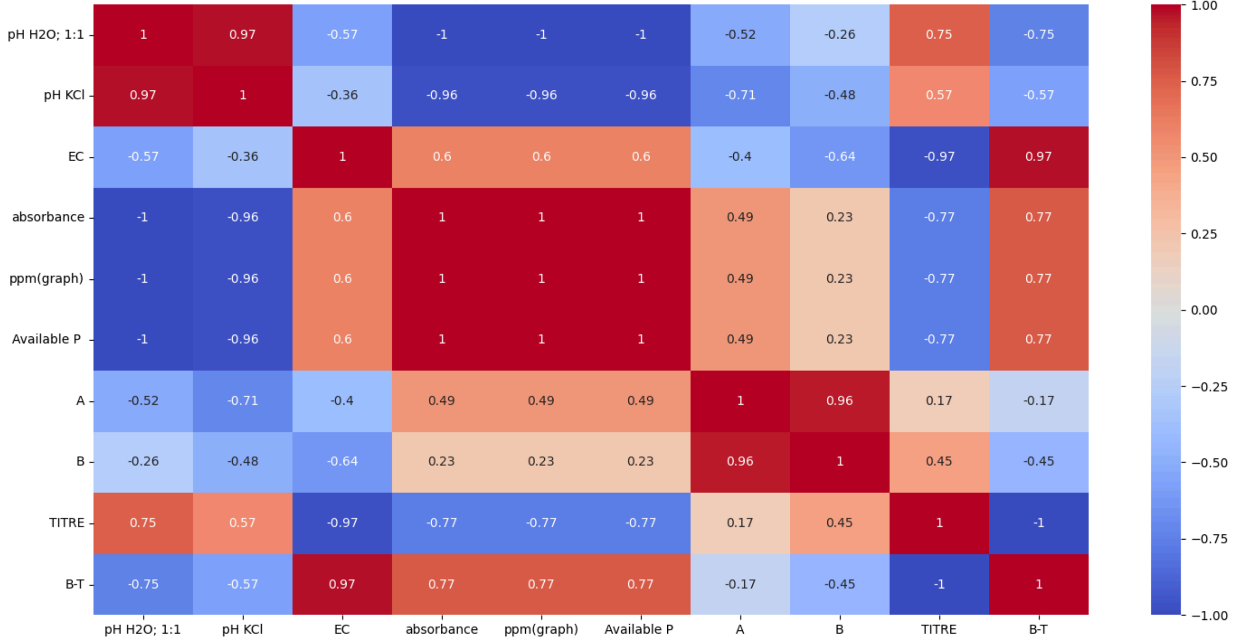
	pH H2O; 1:1	pH KCl	EC	absorbance	ppm(graph)	Available P	A	B	TITRE	B-T
TITRE	-0.16	-0.11	-0.68	-0.76	-0.76	-0.76	0.092	0.68	1	-1
B-T	0.16	0.11	0.68	0.76	0.76	0.76	-0.092	-0.68	-1	1

SANDY CLAY LOAM Soil Class



	pH H2O; 1:1	pH KCl	EC	absorbance	ppm(graph)	Available P	A	B	TITRE	B-T
pH H2O; 1:1	1	1	-1	-1	-1	-1	-1	-1	1	-1
pH KCl	1	1	-1	-1	-1	-1	-1	-1	1	-1
EC	-1	-1	1	1	1	1	1	1	-1	1
absorbance	-1	-1	1	1	1	1	1	1	-1	1
ppm(graph)	-1	-1	1	1	1	1	1	1	-1	1
Available P	-1	-1	1	1	1	1	1	1	-1	1
A	-1	-1	1	1	1	1	1	1	-1	1
B	-1	-1	1	1	1	1	1	1	-1	1
TITRE	1	1	-1	-1	-1	-1	-1	-1	1	-1
B-T	-1	-1	1	1	1	1	1	1	-1	1

SAND/LOAMY SAND Soil Class



	pH H2O; 1:1	pH KCl	EC	absorbance	ppm(graph)	Available P	A	B	TITRE	B-T
pH H2O; 1:1	1	0.97	-0.57	-1	-1	-1	-0.52	-0.26	0.75	-0.75
pH KCl	0.97	1	-0.36	-0.96	-0.96	-0.96	-0.71	-0.48	0.57	-0.57
EC	-0.57	-0.36	1	0.6	0.6	0.6	-0.4	-0.64	-0.97	0.97
absorbance	-1	-0.96	0.6	1	1	1	0.49	0.23	-0.77	0.77
ppm(graph)	-1	-0.96	0.6	1	1	1	0.49	0.23	-0.77	0.77
Available P	-1	-0.96	0.6	1	1	1	0.49	0.23	-0.77	0.77
A	-0.52	-0.71	-0.4	0.49	0.49	0.49	1	0.96	0.17	-0.17
B	-0.26	-0.48	-0.64	0.23	0.23	0.23	0.96	1	0.45	-0.45
TITRE	0.75	0.57	-0.97	-0.77	-0.77	-0.77	0.17	0.45	1	-1
B-T	-0.75	-0.57	0.97	0.77	0.77	0.77	-0.17	-0.45	-1	1