About the Data

This data set shows the different drinking water sources across the country on the year 2020.

- Philippines - Households by Main Source of Drinking Water Census 2020
- Source: https://data.humdata.org/dataset/hh-by-main-source-of-drinking-water-census-2020 (https://data.humdata.org/dataset/hh-by-main-source-of-drinking-water-census-2020)

Load Data

In [60]:

```
import pandas as pd

# loading the data
df = pd.read_excel('data/hh-drinking-water-source-admin3-census2020.xlsx')
df.head()
```

Out[60]:

	Region	Province	Mun	concat	MunCode_New	MunCode_Old	Total Number \nof Households*	Own Use Faucet, Community Water System	Shared Faucet, Community Water System	Own Use Tubed/Pipe Deep Wel
0	NATIONAL CAPITAL REGION (NCR)	NCR, City of Manila, First District (Not a Pro...	CITY OF MANILA	NATIONAL CAPITAL REGION (NCR)NCR, City of Mani...	PH1380600000	PH133900000	483261	118315	27775	280
1	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF MANDALUYONG	NATIONAL CAPITAL REGION (NCR)NCR, Second Distr...	PH1380500000	PH137401000	116505	32880	4687	42
2	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF MARIKINA	NATIONAL CAPITAL REGION (NCR)NCR, Second Distr...	PH1380700000	PH137402000	104404	47226	4447	76
3	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF PASIG	NATIONAL CAPITAL REGION (NCR)NCR, Second Distr...	PH1381200000	PH137403000	212864	89493	6025	142
4	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	QUEZON CITY	NATIONAL CAPITAL REGION (NCR)NCR, Second Distr...	PH1381300000	PH137404000	738330	309249	36489	459

5 rows × 22 columns

Initial Exploration

- I started by looking at the first and last few values (head() and tail()) of the dataset as well as the summary and a descriptive statistic of the available numerical columns using the info() and describe() functions.

```
In [61]: df.head()
```

```
Out[61]:
```

	Region	Province	Mun	concat	MunCode_New	MunCode_Old	Total Number \nof Households*	Own Use Faucet, Community Water System	Shared Faucet, Community Water System	Own Use Tubed/Pipe Deep Wel
0	NATIONAL CAPITAL REGION (NCR)	NCR, City of Manila, First District (Not a Pro...	CITY OF MANILA	NATIONAL CAPITAL REGION (NCR)NCR, City of Mani...	PH1380600000	PH133900000	483261	118315	27775	280
1	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF MANDALUYONG	NATIONAL CAPITAL REGION (NCR)NCR, Second Distr...	PH1380500000	PH137401000	116505	32880	4687	42
2	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF MARIKINA	NATIONAL CAPITAL REGION (NCR)NCR, Second Distr...	PH1380700000	PH137402000	104404	47226	4447	76
3	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF PASIG	NATIONAL CAPITAL REGION (NCR)NCR, Second Distr...	PH1381200000	PH137403000	212864	89493	6025	142
4	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	QUEZON CITY	NATIONAL CAPITAL REGION (NCR)NCR, Second Distr...	PH1381300000	PH137404000	738330	309249	36489	459

5 rows × 22 columns

In [62]:
```
df.tail()
```

Out[62]:

	Region	Province	Mun	concat	MunCode_New	MunCode_Old	Total Number \nof Households*	Own Use Faucet, Community Water System	Shared Faucet, Community Water System
1637	BANGSAMORO AUTONOMOUS REGION IN MUSLIM MINDANA...	INTERIM PROVINCE	MIDSAYAP CLUSTER II	BANGSAMORO AUTONOMOUS REGION IN MUSLIM MINDANA...	PH1999904000	NaN	4274	212	286
1638	BANGSAMORO AUTONOMOUS REGION IN MUSLIM MINDANA...	INTERIM PROVINCE	PIGcAWAYAN CLUSTER	BANGSAMORO AUTONOMOUS REGION IN MUSLIM MINDANA...	PH1999905000	NaN	4356	192	560
1639	BANGSAMORO AUTONOMOUS REGION IN MUSLIM MINDANA...	INTERIM PROVINCE	PIKIT CLUSTER I	BANGSAMORO AUTONOMOUS REGION IN MUSLIM MINDANA...	PH1999906000	NaN	8935	176	143
1640	BANGSAMORO AUTONOMOUS REGION IN MUSLIM MINDANA...	INTERIM PROVINCE	PIKIT CLUSTER II	BANGSAMORO AUTONOMOUS REGION IN MUSLIM MINDANA...	PH1999907000	NaN	7978	34	39
1641	BANGSAMORO AUTONOMOUS REGION IN MUSLIM MINDANA...	INTERIM PROVINCE	PIKIT CLUSTER III	BANGSAMORO AUTONOMOUS REGION IN MUSLIM MINDANA...	PH1999908000	NaN	6566	16	494

5 rows × 22 columns

```
In [63]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1642 entries, 0 to 1641
Data columns (total 22 columns):
 #   Column                                                   Non-Null Count  Dtype 
---  ------                                                   --------------  ----- 
 0   Region                                                   1642 non-null   object
 1   Province                                                 1642 non-null   object
 2   Mun                                                      1642 non-null   object
 3   concat                                                   1642 non-null   object
 4   MunCode_New                                              1642 non-null   object
 5   MunCode_Old                                              1634 non-null   object
 6   Total Number
of Households*                             1642 non-null   int64 
 7   Own Use Faucet, Community Water System                   1642 non-null   int64 
 8   Shared Faucet, Community Water System                    1642 non-null   int64 
 9   Own Use Tubed/Piped Deep Well                            1642 non-null   int64 
 10  Shared Tubed/Piped Deep Well                             1642 non-null   int64 
 11  Tubed/Piped Shallow Well                                 1642 non-null   int64 
 12  Protected Well                                           1642 non-null   int64 
 13  Unprotected Well                                         1642 non-null   int64 
 14  Protected Spring                                         1642 non-null   int64 
 15  Unprotected Spring                                       1642 non-null   int64 
 16  Rainwater                                                1642 non-null   int64 
 17  Surface Watera                                           1642 non-null   int64 
 18  Peddler Includes tanker-truck and cart with small tank   1642 non-null   int64 
 19  Water Refilling Station                                  1642 non-null   int64 
 20  Bottled Water                                            1642 non-null   int64 
 21  Others                                                   1642 non-null   int64 
dtypes: int64(16), object(6)
memory usage: 282.3+ KB
```

In [64]:

```
df.describe()
```

Out[64]:

	Total Number \nof Households*	Own Use Faucet, Community Water System	Shared Faucet, Community Water System	Own Use Tubed/Piped Deep Well	Shared Tubed/Piped Deep Well	Tubed/Piped Shallow Well	Protected Well	Unprotected Well	Protected Spring
count	1642.000000	1642.000000	1642.000000	1642.000000	1642.000000	1642.000000	1642.000000	1642.000000	1642.000000
mean	16062.517052	3541.592570	1537.856273	644.468940	869.188794	222.572473	491.906821	166.775274	539.623021
std	35028.096284	12959.399786	2682.901522	965.179271	1104.238682	339.419088	780.860390	357.021281	826.273025
min	35.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	4968.500000	452.500000	351.250000	71.000000	127.000000	25.000000	56.250000	8.000000	40.000000
50%	8582.000000	1230.500000	910.000000	272.000000	484.000000	102.000000	212.500000	42.000000	232.500000
75%	15390.000000	2986.250000	1882.750000	807.750000	1184.750000	285.000000	602.000000	164.000000	706.500000
max	738330.000000	309249.000000	57556.000000	8529.000000	10834.000000	4530.000000	8471.000000	5962.000000	12553.000000

Handling Missing Values

- Checking if there are null values within the dataset and handling them accordingly such as replacing their values with 0 or removing them completely.

In [65]:
```
df.isnull().sum()
```

Out[65]:
```
Region                                                     0
Province                                                   0
Mun                                                        0
concat                                                     0
MunCode_New                                                0
MunCode_Old                                                8
Total Number \nof Households*                              0
Own Use Faucet, Community Water System                     0
Shared Faucet, Community Water System                      0
Own Use Tubed/Piped Deep Well                              0
Shared Tubed/Piped Deep Well                               0
Tubed/Piped Shallow Well                                   0
Protected Well                                             0
Unprotected Well                                           0
Protected Spring                                           0
Unprotected Spring                                         0
Rainwater                                                  0
Surface Watera                                             0
Peddler Includes tanker-truck and cart with small tank     0
Water Refilling Station                                    0
Bottled Water                                              0
Others                                                     0
dtype: int64
```

- There are null values in the MunCode_Old column which I think should be removed for it is no longer relevant in the code anymore leaving us with a dataset that has no null values. I also dropped the concat column for it's only the region and municipality joined together.

In [66]:
```
# no. of null values in the MunCode_Old column
df['MunCode_Old'].isnull().sum()
```

Out[66]:
```
8
```

In [67]:

```
# dropping both concat and MunCode_Old column
df.drop(['concat', 'MunCode_Old'], axis=1, inplace=True)
df.head()
```

Out[67]:

	Region	Province	Mun	MunCode_New	Total Number \nof Households*	Own Use Faucet, Community Water System	Shared Faucet, Community Water System	Own Use Tubed/Piped Deep Well	Shared Tubed/Piped Deep Well	Tubed/Piped Shallow Well
0	NATIONAL CAPITAL REGION (NCR)	NCR, City of Manila, First District (Not a Pro...	CITY OF MANILA	PH1380600000	483261	118315	27775	2809	1304	310
1	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF MANDALUYONG	PH1380500000	116505	32880	4687	426	63	150
2	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF MARIKINA	PH1380700000	104404	47226	4447	762	70	64
3	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF PASIG	PH1381200000	212864	89493	6025	1428	240	139
4	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	QUEZON CITY	PH1381300000	738330	309249	36489	4594	2719	633

In [68]:

```
# there are no more null values left
df.isnull().sum().sum()
```

Out[68]:

```
0
```

In [69]:

```
df.columns
```

Out[69]:

```
Index(['Region', 'Province', 'Mun', 'MunCode_New',
       'Total Number \nof Households*',
       'Own Use Faucet, Community Water System',
       'Shared Faucet, Community Water System',
       'Own Use Tubed/Piped Deep Well', 'Shared Tubed/Piped Deep Well',
       'Tubed/Piped Shallow Well', 'Protected Well', 'Unprotected Well',
       'Protected Spring', 'Unprotected Spring', 'Rainwater', 'Surface Watera',
       'Peddler Includes tanker-truck and cart with small tank ',
       'Water Refilling Station', 'Bottled Water', 'Others'],
      dtype='object')
```

- Setting the MunCode_New column the index for each element has it's unique Municipality Code.

In [70]:
```python
# setting the MunCode_New column as the new index
df.set_index('MunCode_New', inplace=True)
df.head()
```

Out[70]:

	Region	Province	Mun	Total Number \nof Households*	Own Use Faucet, Community Water System	Shared Faucet, Community Water System	Own Use Tubed/Piped Deep Well	Shared Tubed/Piped Deep Well	Tubed/Piped Shallow Well	F
MunCode_New										
PH1380600000	NATIONAL CAPITAL REGION (NCR)	NCR, City of Manila, First District (Not a Pro...	CITY OF MANILA	483261	118315	27775	2809	1304	310	
PH1380500000	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF MANDALUYONG	116505	32880	4687	426	63	150	
PH1380700000	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF MARIKINA	104404	47226	4447	762	70	64	
PH1381200000	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF PASIG	212864	89493	6025	1428	240	139	
PH1381300000	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	QUEZON CITY	738330	309249	36489	4594	2719	633	

In [84]:

```
# renaming one of the columns
df.rename(
    columns={
        'Total Number \nof Households*':'Total Number of Households'
        },
    inplace=True
)
df.head()
```

Out[84]:

	Region	Province	Mun	Total Number of Households	Own Use Faucet, Community Water System	Shared Faucet, Community Water System	Own Use Tubed/Piped Deep Well	Shared Tubed/Piped Deep Well	Tubed/Piped Shallow Well	P
MunCode_New										
PH1380600000	NATIONAL CAPITAL REGION (NCR)	NCR, City of Manila, First District (Not a Pro...	CITY OF MANILA	483261	118315	27775	2809	1304	310	
PH1380500000	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF MANDALUYONG	116505	32880	4687	426	63	150	
PH1380700000	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF MARIKINA	104404	47226	4447	762	70	64	
PH1381200000	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF PASIG	212864	89493	6025	1428	240	139	
PH1381300000	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	QUEZON CITY	738330	309249	36489	4594	2719	633	

- Deciding whether or not I should replace the columns whose datatypes are objects to be represented as integers.
- There are 3 columns whose datatype is object (Region, Provincem and Mun/Municipality). Since these columns represent no inherent ordering, I decided to leave them as is.

```
In [85]: df.dtypes
```

```
Out[85]: Region                                                     object
Province                                                   object
Mun                                                        object
Total Number of Households                                  int64
Own Use Faucet, Community Water System                      int64
Shared Faucet, Community Water System                       int64
Own Use Tubed/Piped Deep Well                               int64
Shared Tubed/Piped Deep Well                                int64
Tubed/Piped Shallow Well                                    int64
Protected Well                                              int64
Unprotected Well                                            int64
Protected Spring                                            int64
Unprotected Spring                                          int64
Rainwater                                                   int64
Surface Watera                                              int64
Peddler Includes tanker-truck and cart with small tank      int64
Water Refilling Station                                     int64
Bottled Water                                               int64
Others                                                      int64
dtype: object
```

In [86]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 1642 entries, PH1380600000 to PH1999908000
Data columns (total 19 columns):
 #   Column                                               Non-Null Count  Dtype 
---  ------                                               --------------  ----- 
 0   Region                                               1642 non-null   object
 1   Province                                             1642 non-null   object
 2   Mun                                                  1642 non-null   object
 3   Total Number of Households                           1642 non-null   int64 
 4   Own Use Faucet, Community Water System               1642 non-null   int64 
 5   Shared Faucet, Community Water System                1642 non-null   int64 
 6   Own Use Tubed/Piped Deep Well                        1642 non-null   int64 
 7   Shared Tubed/Piped Deep Well                         1642 non-null   int64 
 8   Tubed/Piped Shallow Well                             1642 non-null   int64 
 9   Protected Well                                       1642 non-null   int64 
 10  Unprotected Well                                     1642 non-null   int64 
 11  Protected Spring                                     1642 non-null   int64 
 12  Unprotected Spring                                   1642 non-null   int64 
 13  Rainwater                                            1642 non-null   int64 
 14  Surface Watera                                       1642 non-null   int64 
 15  Peddler Includes tanker-truck and cart with small tank  1642 non-null   int64 
 16  Water Refilling Station                              1642 non-null   int64 
 17  Bottled Water                                        1642 non-null   int64 
 18  Others                                               1642 non-null   int64 
dtypes: int64(16), object(3)
memory usage: 321.1+ KB
```

In [87]:

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

Out[87]:

```
Region                                                     0
Province                                                   0
Mun                                                        0
Total Number of Households                                 0
Own Use Faucet, Community Water System                     0
Shared Faucet, Community Water System                      0
Own Use Tubed/Piped Deep Well                              0
Shared Tubed/Piped Deep Well                               0
Tubed/Piped Shallow Well                                   0
Protected Well                                             0
Unprotected Well                                           0
Protected Spring                                           0
Unprotected Spring                                         0
Rainwater                                                  0
Surface Watera                                             0
Peddler Includes tanker-truck and cart with small tank     0
Water Refilling Station                                    0
Bottled Water                                              0
Others                                                     0
dtype: int64
```

In [88]:
```
df.dtypes
```

Out[88]:
```
Region                                                     object
Province                                                   object
Mun                                                        object
Total Number of Households                                  int64
Own Use Faucet, Community Water System                      int64
Shared Faucet, Community Water System                       int64
Own Use Tubed/Piped Deep Well                               int64
Shared Tubed/Piped Deep Well                                int64
Tubed/Piped Shallow Well                                    int64
Protected Well                                              int64
Unprotected Well                                            int64
Protected Spring                                            int64
Unprotected Spring                                          int64
Rainwater                                                   int64
Surface Watera                                              int64
Peddler Includes tanker-truck and cart with small tank      int64
Water Refilling Station                                     int64
Bottled Water                                               int64
Others                                                      int64
dtype: object
```

- Having a cleaned dataset.

In [89]:
```
cleaned_df = df
cleaned_df.head()
```

Out[89]:

	Region	Province	Mun	Total Number of Households	Own Use Faucet, Community Water System	Shared Faucet, Community Water System	Own Use Tubed/Piped Deep Well	Shared Tubed/Piped Deep Well	Tubed/Piped Shallow Well	P
MunCode_New										
PH1380600000	NATIONAL CAPITAL REGION (NCR)	NCR, City of Manila, First District (Not a Pro...	CITY OF MANILA	483261	118315	27775	2809	1304	310	
PH1380500000	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF MANDALUYONG	116505	32880	4687	426	63	150	
PH1380700000	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF MARIKINA	104404	47226	4447	762	70	64	
PH1381200000	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	CITY OF PASIG	212864	89493	6025	1428	240	139	
PH1381300000	NATIONAL CAPITAL REGION (NCR)	NCR, Second District (Not a Province)	QUEZON CITY	738330	309249	36489	4594	2719	633	

In [90]:

```
# descriptive statistics of the cleaned data (numerical columns only)
cleaned_df.describe()
```

Out[90]:

	Total Number of Households	Own Use Faucet, Community Water System	Shared Faucet, Community Water System	Own Use Tubed/Piped Deep Well	Shared Tubed/Piped Deep Well	Tubed/Piped Shallow Well	Protected Well	Unprotected Well	Protected Spring
count	1642.000000	1642.000000	1642.000000	1642.000000	1642.000000	1642.000000	1642.000000	1642.000000	1642.000000
mean	16062.517052	3541.592570	1537.856273	644.468940	869.188794	222.572473	491.906821	166.775274	539.623021
std	35028.096284	12959.399786	2682.901522	965.179271	1104.238682	339.419088	780.860390	357.021281	826.273025
min	35.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	4968.500000	452.500000	351.250000	71.000000	127.000000	25.000000	56.250000	8.000000	40.000000
50%	8582.000000	1230.500000	910.000000	272.000000	484.000000	102.000000	212.500000	42.000000	232.500000
75%	15390.000000	2986.250000	1882.750000	807.750000	1184.750000	285.000000	602.000000	164.000000	706.500000
max	738330.000000	309249.000000	57556.000000	8529.000000	10834.000000	4530.000000	8471.000000	5962.000000	12553.000000

Analysis

- Understanding the distribution and characteristics of each feature as well as analizing the relationship between multiple variables and plotting them to further visualize their relationship.

In [91]:

```
cleaned_df.columns
```

Out[91]:

```
Index(['Region', 'Province', 'Mun', 'Total Number of Households',
       'Own Use Faucet, Community Water System',
       'Shared Faucet, Community Water System',
       'Own Use Tubed/Piped Deep Well', 'Shared Tubed/Piped Deep Well',
       'Tubed/Piped Shallow Well', 'Protected Well', 'Unprotected Well',
       'Protected Spring', 'Unprotected Spring', 'Rainwater', 'Surface Watera',
       'Peddler Includes tanker-truck and cart with small tank ',
       'Water Refilling Station', 'Bottled Water', 'Others'],
      dtype='object')
```

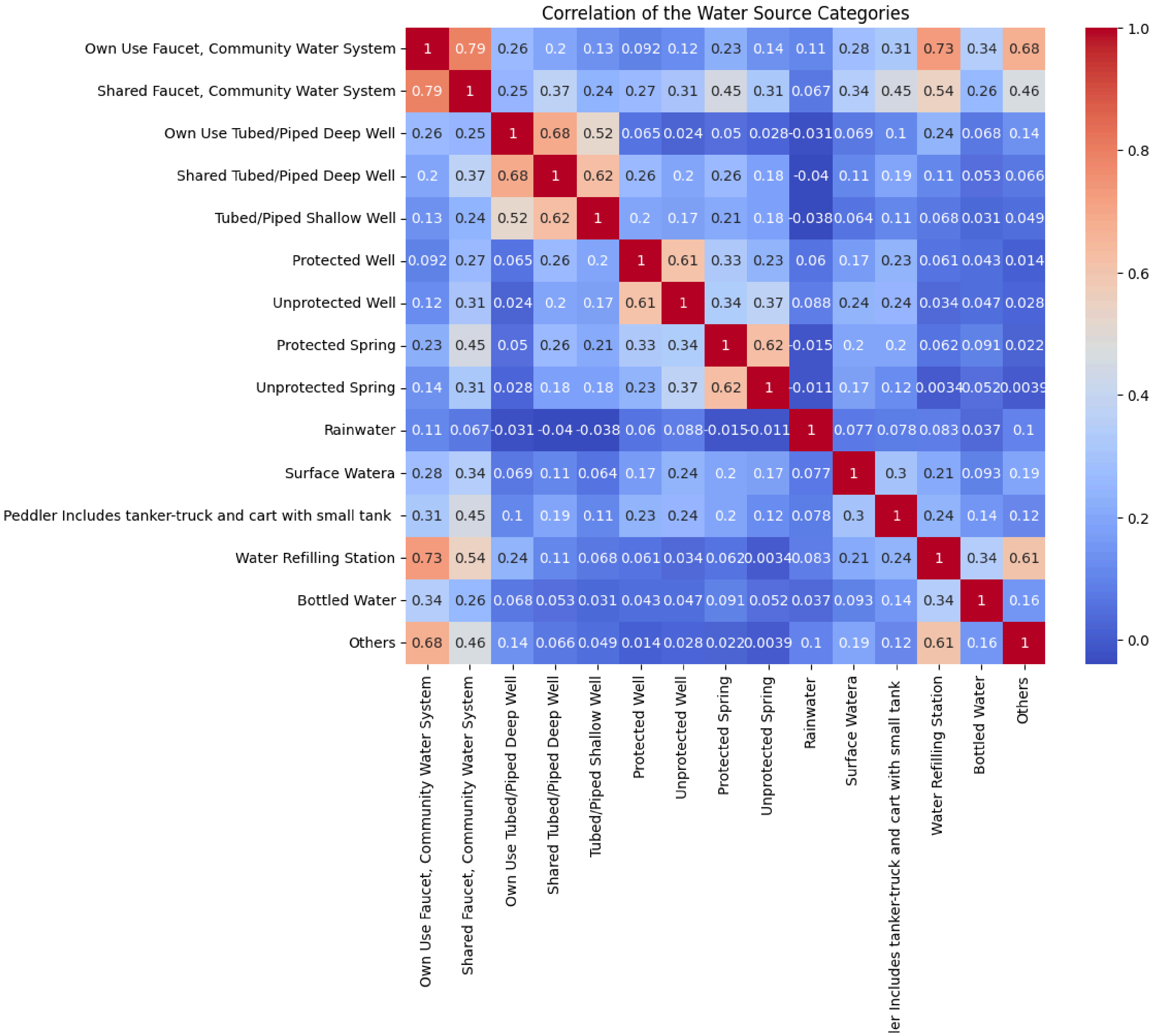
- This heatmap tells us the activity of the data. The highest values is 1 and the rest are below it but not less than 0.

In [105]:

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

# columns I want to use correlation analysis with (water sources)
sources_columns = ['Own Use Faucet, Community Water System', 'Shared Faucet, Community Water System',
                'Own Use Tubed/Piped Deep Well', 'Shared Tubed/Piped Deep Well',
                'Tubed/Piped Shallow Well', 'Protected Well', 'Unprotected Well',
                'Protected Spring', 'Unprotected Spring', 'Rainwater', 'Surface Watera',
                'Peddler Includes tanker-truck and cart with small tank ',
                'Water Refilling Station', 'Bottled Water', 'Others']

# correlation matrix
corr_matrix = cleaned_df[sources_columns].corr()

# using heatmap to visualize it
plt.figure(figsize=(10, 8))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation of the Water Source Categories')
```

Out[105]:

```
Text(0.5, 1.0, 'Correlation of the Water Source Categories')
```


Correlation of the Water Source Categories
Own Use Faucet, Community Water System
Shared Faucet, Community Water System
Own Use Tubed/Piped Deep Well
Shared Tubed/Piped Deep Well
Tubed/Piped Shallow Well
Protected Well
Unprotected Well
Protected Spring
Unprotected Spring
Rainwater
Surface Watera
Peddler Includes tanker-truck and cart with small tank
Water Refilling Station
Bottled Water
Others
1 0.79 0.26 0.2 0.13 0.092 0.12 0.23 0.14 0.11 0.28 0.31 0.73 0.34 0.68
0.79 1 0.25 0.37 0.24 0.27 0.31 0.45 0.31 0.067 0.34 0.45 0.54 0.26 0.46
0.26 0.25 1 0.68 0.52 0.065 0.024 0.05 0.028 -0.031 0.069 0.1 0.24 0.068 0.14
0.2 0.37 0.68 1 0.62 0.26 0.2 0.26 0.18 -0.04 0.11 0.19 0.11 0.053 0.066
0.13 0.24 0.52 0.62 1 0.2 0.17 0.21 0.18 -0.038 0.064 0.11 0.068 0.031 0.049
0.092 0.27 0.065 0.26 0.2 1 0.61 0.33 0.23 0.06 0.17 0.23 0.061 0.043 0.014
0.12 0.31 0.024 0.2 0.17 0.61 1 0.34 0.37 0.088 0.24 0.24 0.034 0.047 0.028
0.23 0.45 0.05 0.26 0.21 0.33 0.34 1 0.62 -0.015 0.2 0.2 0.062 0.091 0.022
0.14 0.31 0.028 0.18 0.18 0.23 0.37 0.62 1 -0.011 0.17 0.12 0.0034 0.052 0.0039
0.11 0.067 -0.031 -0.04 -0.038 0.06 0.088 -0.015 -0.011 1 0.077 0.078 0.083 0.037 0.1
0.28 0.34 0.069 0.11 0.064 0.17 0.24 0.2 0.17 0.077 1 0.3 0.21 0.093 0.19
0.31 0.45 0.1 0.19 0.11 0.23 0.24 0.2 0.12 0.078 0.3 1 0.24 0.14 0.12
0.73 0.54 0.24 0.11 0.068 0.061 0.034 0.062 0.0034 0.083 0.21 0.24 1 0.34 0.61
0.34 0.26 0.068 0.053 0.031 0.043 0.047 0.091 0.052 0.037 0.093 0.14 0.34 1 0.16
0.68 0.46 0.14 0.066 0.049 0.014 0.028 0.022 0.0039 0.1 0.19 0.12 0.61 0.16 1
Own Use Faucet, Community Water System
Shared Faucet, Community Water System
Own Use Tubed/Piped Deep Well
Shared Tubed/Piped Deep Well
Tubed/Piped Shallow Well
Protected Well
Unprotected Well
Protected Spring
Unprotected Spring
Rainwater
Surface Watera
ler Includes tanker-truck and cart with small tank
Water Refilling Station
Bottled Water
Others
1.0
0.8
0.6
0.4
0.2
0.0

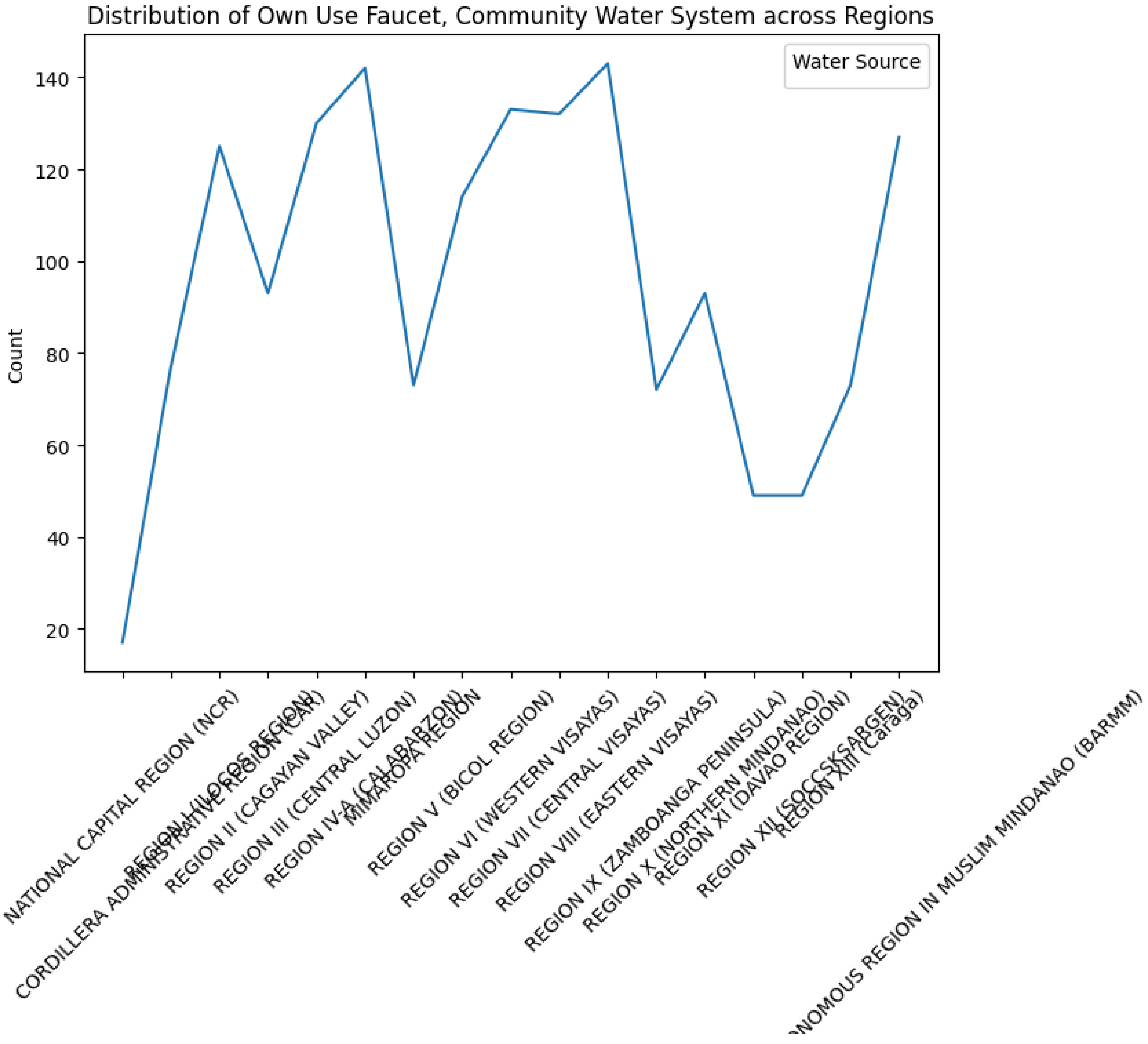
In [100]:

```
# using bar plot to visualize the water source across the regions
plt.figure(figsize=(8, 6))
sns.lineplot(x='Region', y='Own Use Faucet, Community Water System', data=df, estimator='count', ci=None)
plt.title('Distribution of Own Use Faucet, Community Water System across Regions')
plt.xlabel('Region')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.legend(title='Water Source')
```

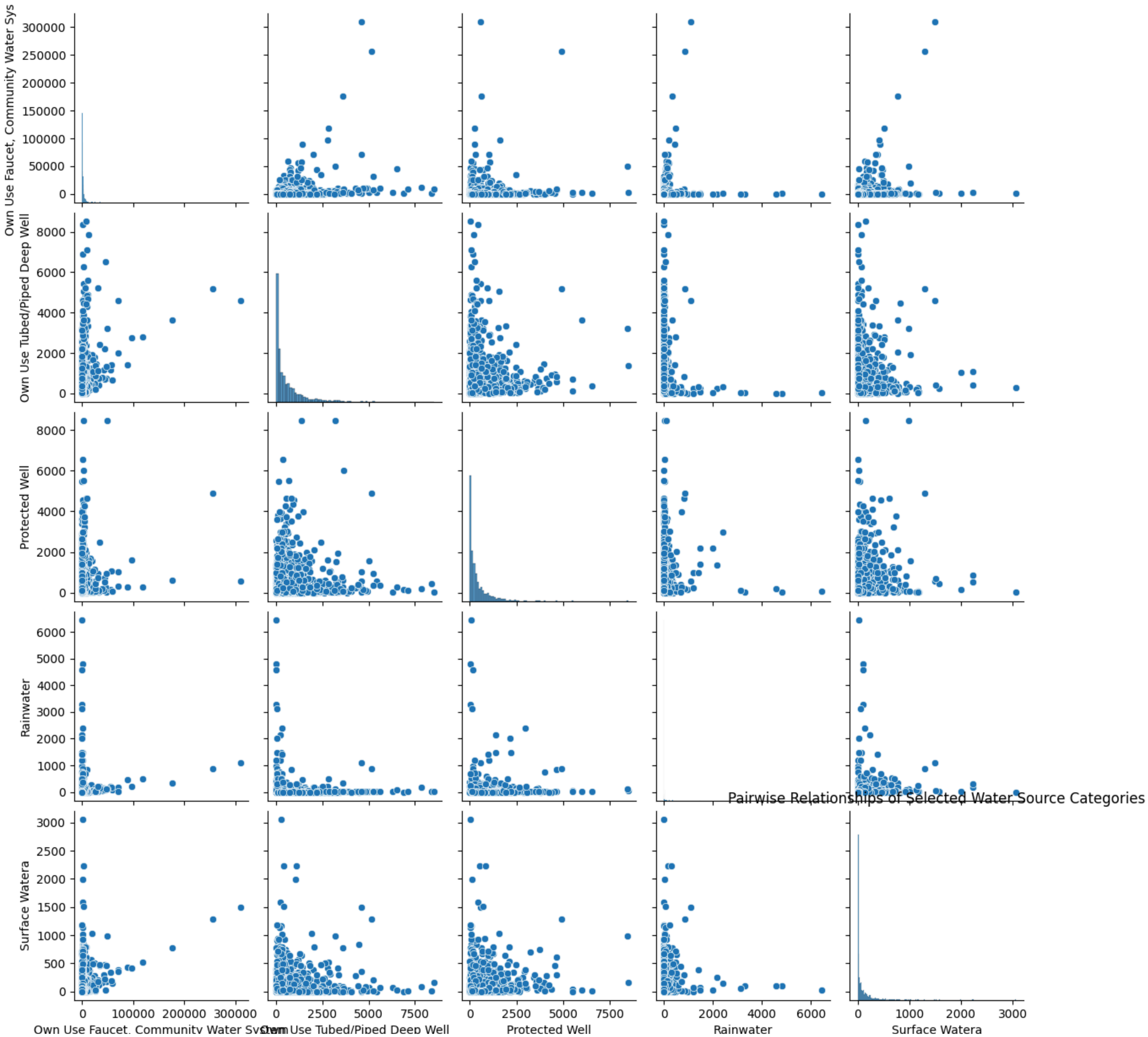
```
<ipython-input-100-f0069355c1a4>:3: FutureWarning:

The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.

  sns.lineplot(x='Region', y='Own Use Faucet, Community Water System', data=df, estimator='count', ci=None)
WARNING:matplotlib.legend:No artists with labels found to put in legend.  Note that artists whose label star
t with an underscore are ignored when legend() is called with no argument.
```

Out[100]:

```
<matplotlib.legend.Legend at 0x780b422703d0>
```


Distribution of Own Use Faucet, Community Water System across Regions
Water Source
Count
140
120
100
80
60
40
20
NATIONAL CAPITAL REGION (NCR)
CORDILLERA ADMINISTRATIVE REGION (CAR)
REGION I (ILOCOS REGION)
REGION II (CAGAYAN VALLEY)
REGION III (CENTRAL LUZON)
REGION IV-A (CALABARZON)
MIMAROPA REGION
REGION V (BICOL REGION)
REGION VI (WESTERN VISAYAS)
REGION VII (CENTRAL VISAYAS)
REGION VIII (EASTERN VISAYAS)
REGION IX (ZAMBOANGA PENINSULA)
REGION X (NORTHERN MINDANAO)
REGION XI (DAVAO REGION)
REGION XII (SOCCSKSARGEN)
REGION XIII (Caraga)
ONOMOUS REGION IN MUSLIM MINDANAO (BARMM)

BANGSAMORO AUTO
Region



Pairwise Relationships of Selected Water Source Categories

- Visualization of the summary statistics.