In [2]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
```

In [3]:
```
G12 = pd.read_csv("g2.csv")
G13 = pd.read_csv("g3.csv")
G14 = pd.read_csv("g4.csv")
G15 = pd.read_csv("g5.csv")
G16 = pd.read_csv("g6.csv")
G17 = pd.read_csv("g7.csv")
G18 = pd.read_csv("g8.csv")
G19 = pd.read_csv("g9.csv")
G21 = pd.read_csv("g21.csv")
```

In [4]:
```
df = pd.concat([G12,G13,G14,G15,G16,G17,G18,G19,G21], ignore_index = True)
```

In [5]:
```
df['STATE'] = df['STATE'].str.upper()
```

In [6]:
```
display(df)
```

	Station Code	Station Name	STATE	Temperature Min	Temperature Max	pH Min	pH Max	Conductivity (µmhos/cm) Min	Conductivity (µmhos/cm) Max	Year
0	15.0	WELL AT KUYYURA, A.P.	ANDHRA PRADESH	23.0	28.0	6.82	7.85	195.0	226.0	2012
1	16.0	WELL AT TADAVAI A.P.	ANDHRA PRADESH	27.0	28.0	7.21	7.72	663.0	1121.0	2012
2	26.0	WELL AT VIJAYWADA, A.P.	ANDHRA PRADESH	26.0	32.0	7.19	7.39	1601.0	1661.0	2012
3	27.0	WELL AT PEDDAVOORA, A.P.	ANDHRA PRADESH	26.0	27.0	7.82	8.80	1420.0	1668.0	2012
4	1513.0	B W. - KRISHNA MURTHY, D.NO. 48-16-43 AUTONAGA...	ANDHRA PRADESH	25.0	30.0	6.80	7.11	1454.0	1575.0	2012
...	...	...	...	...	...	...	...	...	...	...
6311	2533.0	SURI TOWN NEAR BUS STAND	WEST BENGAL	27.0	30.0	7.20	7.90	358.0	651.0	2021
6312	1773.0	TANGRA, CALCUTTA , WEST\nBENGAL	WEST BENGAL	28.0	31.0	7.30	7.40	1435.0	1730.0	2021
6313	1774.0	TOPSIA CALCUTTA, WEST\nBENGAL	WEST BENGAL	29.0	31.0	7.30	7.30	1705.0	1921.0	2021
6314	2546.0	ULUBERIA COLLEGE AT HOWRAH	WEST BENGAL	26.0	30.0	7.20	7.60	909.0	1090.0	2021
6315	2535.0	VISVA BHARATI	WEST BENGAL	30.0	31.0	7.90	8.20	370.0	437.0	2021

6316 rows × 10 columns

In [7]:
```
df.sort_values(by = "Year", inplace = True, ascending = False, ignore_index = 'True')
df
```

Out[7]:

	Station Code	Station Name	STATE	Temperature Min	Temperature Max	pH Min	pH Max	Conductivity (µmhos/cm) Min	Conductivity (µmhos/cm) Max	Year
0	2548.0	HIDCO OFFICE, RAJARHAT	WEST BENGAL	26.0	30.0	7.40	7.50	253.0	1264.0	2021
1	2535.0	VISVA BHARATI	WEST BENGAL	30.0	31.0	7.90	8.20	370.0	437.0	2021
2	2546.0	ULUBERIA COLLEGE AT HOWRAH	WEST BENGAL	26.0	30.0	7.20	7.60	909.0	1090.0	2021
3	1774.0	TOPSIA CALCUTTA, WEST\nBENGAL	WEST BENGAL	29.0	31.0	7.30	7.30	1705.0	1921.0	2021
4	1773.0	TANGRA, CALCUTTA , WEST\nBENGAL	WEST BENGAL	28.0	31.0	7.30	7.40	1435.0	1730.0	2021
...	...	...	...	...	...	...	...	...	...	...
6311	1795.0	B/W. - PANCHAYAT OFFICE ,,BOLLARAM (V) MEDAK D...	ANDHRA PRADESH	26.0	27.0	7.00	7.20	6870.0	7120.0	2012
6312	1794.0	B/W- MANAKONDUR (V) , KARIMNAGAR DIST., A.P	ANDHRA PRADESH	24.0	29.0	7.66	7.75	910.0	944.0	2012
6313	1793.0	O/W- BHOOMAIAH NEAR ASHPONDS OF NTPC , KUNDANP...	ANDHRA PRADESH	25.0	28.0	7.35	7.93	1664.0	1975.0	2012
6314	1792.0	B/W-NEAR CKM COLLEGE , ENUMAMULA (V) ,WARANGAL...	ANDHRA PRADESH	26.0	28.0	6.87	7.38	2377.0	3120.0	2012
6315	1791.0	B/W. - EAST OF SAICHERUVU ,PAIDIPALLY (V),WARA...	ANDHRA PRADESH	28.0	29.0	6.76	7.43	2610.0	4525.0	2012

6316 rows × 10 columns

In [8]:
```
df.dropna(axis = 0,inplace = True)
```

In [9]:
```
result = df.isnull().sum()
result
```

Out[9]:
```
Station Code                  0
Station Name                  0
STATE                         0
Temperature Min               0
Temperature Max               0
pH Min                        0
pH Max                        0
Conductivity (µmhos/cm) Min   0
Conductivity (µmhos/cm) Max   0
Year                          0
dtype: int64
```

In [10]:
```
df.reset_index(inplace=True, drop=True)
df.head(10)
```

Out[10]:

	Station Code	Station Name	STATE	Temperature Min	Temperature Max	pH Min	pH Max	Conductivity (µmhos/cm) Min	Conductivity (µmhos/cm) Max	Year
0	2548.0	HIDCO OFFICE, RAJARHAT	WEST BENGAL	26.0	30.0	7.4	7.5	253.0	1264.0	2021
1	2535.0	VISVA BHARATI	WEST BENGAL	30.0	31.0	7.9	8.2	370.0	437.0	2021
2	2546.0	ULUBERIA COLLEGE AT HOWRAH	WEST BENGAL	26.0	30.0	7.2	7.6	909.0	1090.0	2021
3	1774.0	TOPSIA CALCUTTA, WEST\nBENGAL	WEST BENGAL	29.0	31.0	7.3	7.3	1705.0	1921.0	2021
4	1773.0	TANGRA, CALCUTTA , WEST\nBENGAL	WEST BENGAL	28.0	31.0	7.3	7.4	1435.0	1730.0	2021
5	2533.0	SURI TOWN NEAR BUS STAND	WEST BENGAL	27.0	30.0	7.2	7.9	358.0	651.0	2021
6	3093.0	BORE WELL AT ALLADAPALEM\nVILLAGE, PYDIBHIMAVARAM	ANDHRA PRADESH	29.0	30.0	7.3	7.4	840.0	1481.0	2021
7	3092.0	BORE WELL AT ARINAMA\nAKKIVALASA, SRIKAKULAM	ANDHRA PRADESH	28.0	31.0	7.4	8.1	712.0	1139.0	2021
8	4360.0	BORE WELL AT IDA,\nRAMANAYYAPETA, KKAINADA	ANDHRA PRADESH	27.0	29.0	7.7	7.8	990.0	1280.0	2021
9	3091.0	BORE WELL AT KAPULUPPADA\nDUMPSITE, VISHAKHAPA...	ANDHRA PRADESH	26.0	26.0	7.0	7.1	3340.0	4260.0	2021

In [11]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5703 entries, 0 to 5702
Data columns (total 10 columns):
 #   Column                        Non-Null Count  Dtype  
---  ------                        --------------  -----  
 0   Station Code                  5703 non-null   float64
 1   Station Name                  5703 non-null   object 
 2   STATE                         5703 non-null   object 
 3   Temperature Min               5703 non-null   float64
 4   Temperature Max               5703 non-null   float64
 5   pH Min                        5703 non-null   float64
 6   pH Max                        5703 non-null   float64
 7   Conductivity (µmhos/cm) Min   5703 non-null   float64
 8   Conductivity (µmhos/cm) Max   5703 non-null   float64
 9   Year                          5703 non-null   int64  
dtypes: float64(7), int64(1), object(2)
memory usage: 445.7+ KB
```

In [12]:
```
df['Temperature_Mean']=(df['Temperature Min']+df['Temperature Max'])/2
df['pH_Mean']=(df['pH Min']+df['pH Max'])/2
df['Conductivity_Mean']=(df['Conductivity (µmhos/cm) Min']+df['Conductivity (µmhos/cm) Max'])/2
```

In [13]:
```
df.head()
```

Out[13]:

	Station Code	Station Name	STATE	Temperature Min	Temperature Max	pH Min	pH Max	Conductivity (µmhos/cm) Min	Conductivity (µmhos/cm) Max	Year	Temperature_Mean
0	2548.0	HIDCO OFFICE, RAJARHAT	WEST BENGAL	26.0	30.0	7.4	7.5	253.0	1264.0	2021	28.0
1	2535.0	VISVA BHARATI	WEST BENGAL	30.0	31.0	7.9	8.2	370.0	437.0	2021	30.5
2	2546.0	ULUBERIA COLLEGE AT HOWRAH	WEST BENGAL	26.0	30.0	7.2	7.6	909.0	1090.0	2021	28.0
3	1774.0	TOPSIA CALCUTTA, WEST\nBENGAL	WEST BENGAL	29.0	31.0	7.3	7.3	1705.0	1921.0	2021	30.0
4	1773.0	TANGRA, CALCUTTA , WEST\nBENGAL	WEST BENGAL	28.0	31.0	7.3	7.4	1435.0	1730.0	2021	29.5

In [14]:
```
df=df[['Station Code', 'Station Name','STATE', 'Temperature Min','Temperature Max', 'Temperature_Mean', 'pH Min
```

In [15]:
```
unique_state = df['STATE'].unique()
unique_state
```

Out[15]:
```
array(['WEST BENGAL', 'ANDHRA PRADESH', 'UTTARAKHAND', 'UTTAR PRADESH',
       'TRIPURA', 'TELANGANA', 'TAMIL NADU', 'RAJASTHAN', 'PUNJAB',
       'PUDUCHERRY', 'ODISHA', 'MIZORAM', 'NAGALAND', 'MEGHALAYA',
       'MAHARASHTRA', 'MANIPUR', 'MADHYA PRADESH', 'LAKSHADWEEP',
       'KERALA', 'HIMACHAL PRADESH', 'JAMMU & KASHMIR', 'JHARKHAND',
       'KARNATAKA', 'HARYANA', 'GUJARAT', 'DADRA & NAGAR HAVELI',
       'DAMAN & DIU', 'GOA', 'BIHAR', 'CHANDIGARH', 'CHHATTISGARH',
       'ASSAM', 'PONDICHERRY'], dtype=object)
```

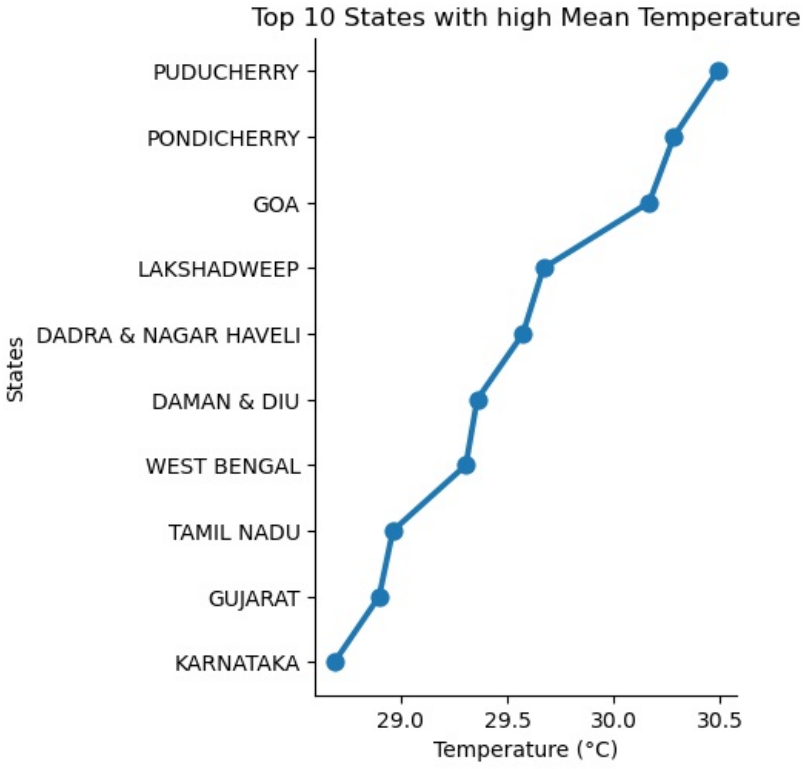
In [16]:
```
sb.color_palette('husl',8)
plt.rcParams['figure.figsize']=50,10
```

In [17]:
```
top_states = df.groupby('STATE')['Temperature_Mean'].mean().sort_values(ascending = False).to_frame().reset_ind
```

In [18]:
```
top_10_states = top_states.head(10)
```

In [19]:
```
sb.catplot(data=top_10_states, x='Temperature_Mean', y='STATE', kind='point')
plt. title( 'Top 10 States with high Mean Temperature')
plt.ylabel( 'States')
plt.xlabel('Temperature (°C)')
```

Out[19]:
```
Text(0.5, 28.999999999999986, 'Temperature (°C)')
```



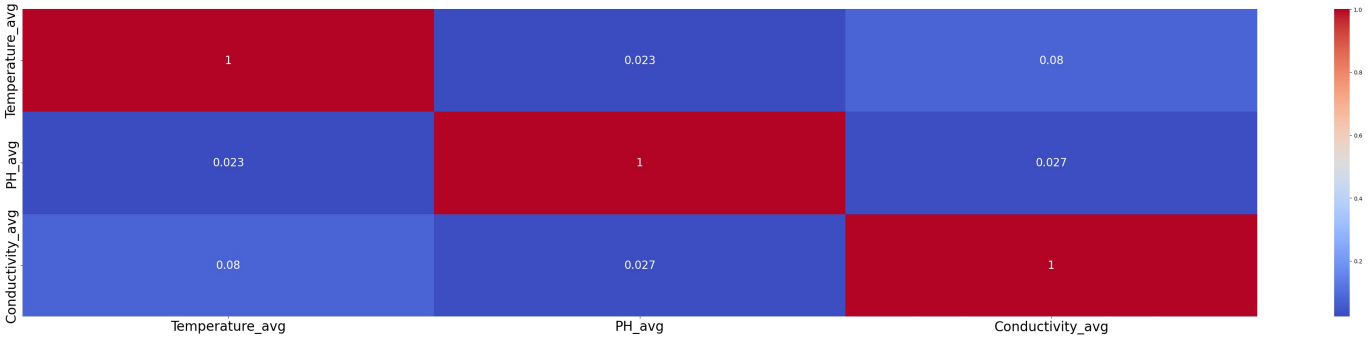
In [25]:
```
df['Temperature_avg'] = (df['Temperature Min'] + df['Temperature Max']) / 2
df['PH_avg'] = (df['pH Min'] + df['pH Max']) / 2
df['Conductivity_avg'] = (df['Conductivity (µmhos/cm) Min'] + df['Conductivity (µmhos/cm) Max']) / 2
```

```
C:\Users\shaik\AppData\Local\Temp\ipykernel_10704\3563509002.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#retu
rning-a-view-versus-a-copy
  df['Temperature_avg'] = (df['Temperature Min'] + df['Temperature Max']) / 2
C:\Users\shaik\AppData\Local\Temp\ipykernel_10704\3563509002.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#retu
rning-a-view-versus-a-copy
  df['PH_avg'] = (df['pH Min'] + df['pH Max']) / 2
C:\Users\shaik\AppData\Local\Temp\ipykernel_10704\3563509002.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#retu
rning-a-view-versus-a-copy
  df['Conductivity_avg'] = (df['Conductivity (µmhos/cm) Min'] + df['Conductivity (µmhos/cm) Max']) / 2
```

In [54]:
```
df['PH_range'] = df['pH Max'] - df['pH Min']
df['Temp_range'] = df['Temperature Max'] - df['Temperature Min']
```

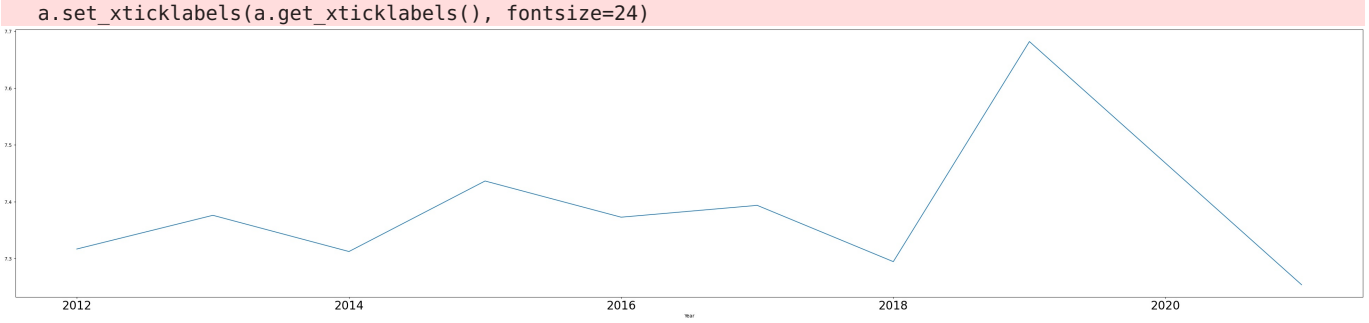
```
C:\Users\shaik\AppData\Local\Temp\ipykernel_10704\1645673603.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#retu
rning-a-view-versus-a-copy
  df['PH_range'] = df['pH Max'] - df['pH Min']
C:\Users\shaik\AppData\Local\Temp\ipykernel_10704\1645673603.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#retu
rning-a-view-versus-a-copy
  df['Temp_range'] = df['Temperature Max'] - df['Temperature Min']
```

In [50]:
```
import seaborn as sns
import matplotlib.pyplot as plt
ax = sns.heatmap(df[['Temperature_avg','PH_avg','Conductivity_avg']].corr(),
            annot=True,
            cmap="coolwarm",
            annot_kws={"size": 20}
           )
ax.set_xticklabels(ax.get_xticklabels(), fontsize=14)
ax.set_yticklabels(ax.get_yticklabels(), fontsize=14)

plt.show()
```

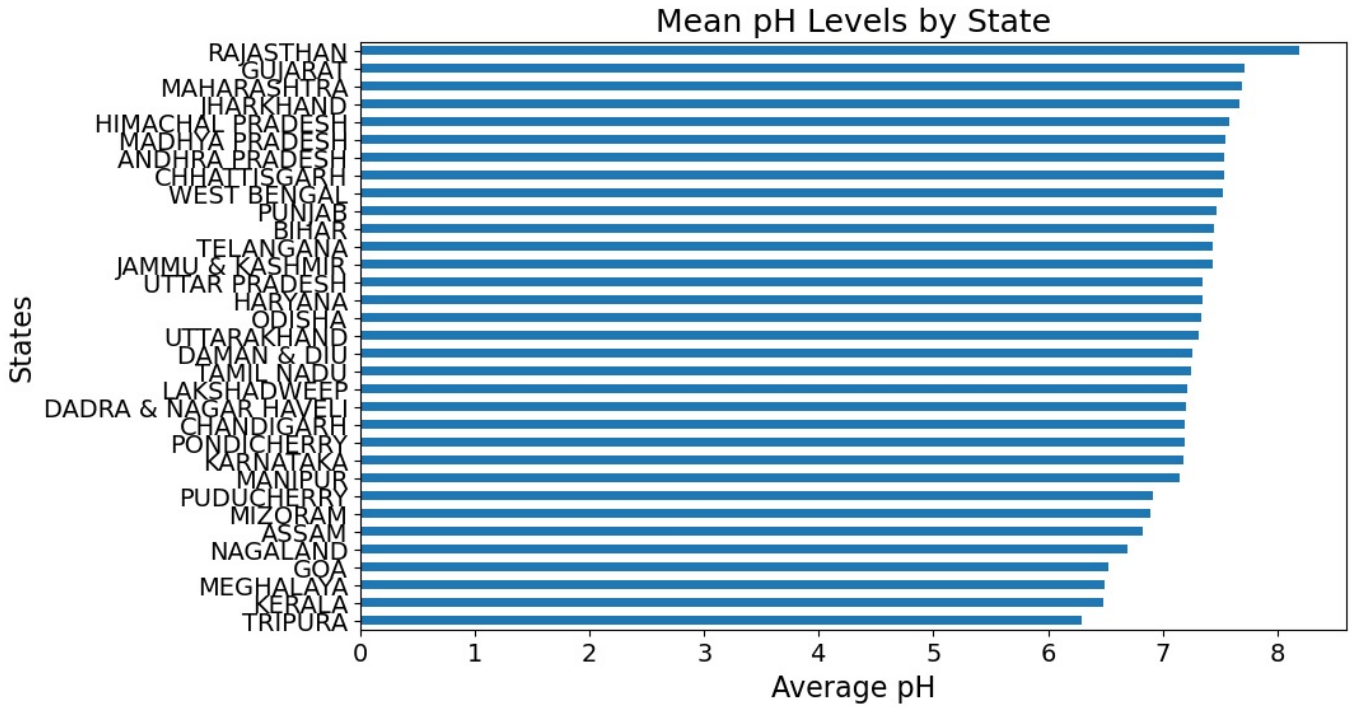


In [45]:
```
a = df.groupby('Year')['PH_avg'].mean().plot(kind='line')
a.set_xticklabels(a.get_xticklabels(), fontsize=24)
plt.show()
```

```
C:\Users\shaik\AppData\Local\Temp\ipykernel_10704\1235596362.py:2: UserWarning: set_ticklabels() should only be
used with a fixed number of ticks, i.e. after set_ticks() or using a FixedLocator.
  a.set_xticklabels(a.get_xticklabels(), fontsize=24)
```

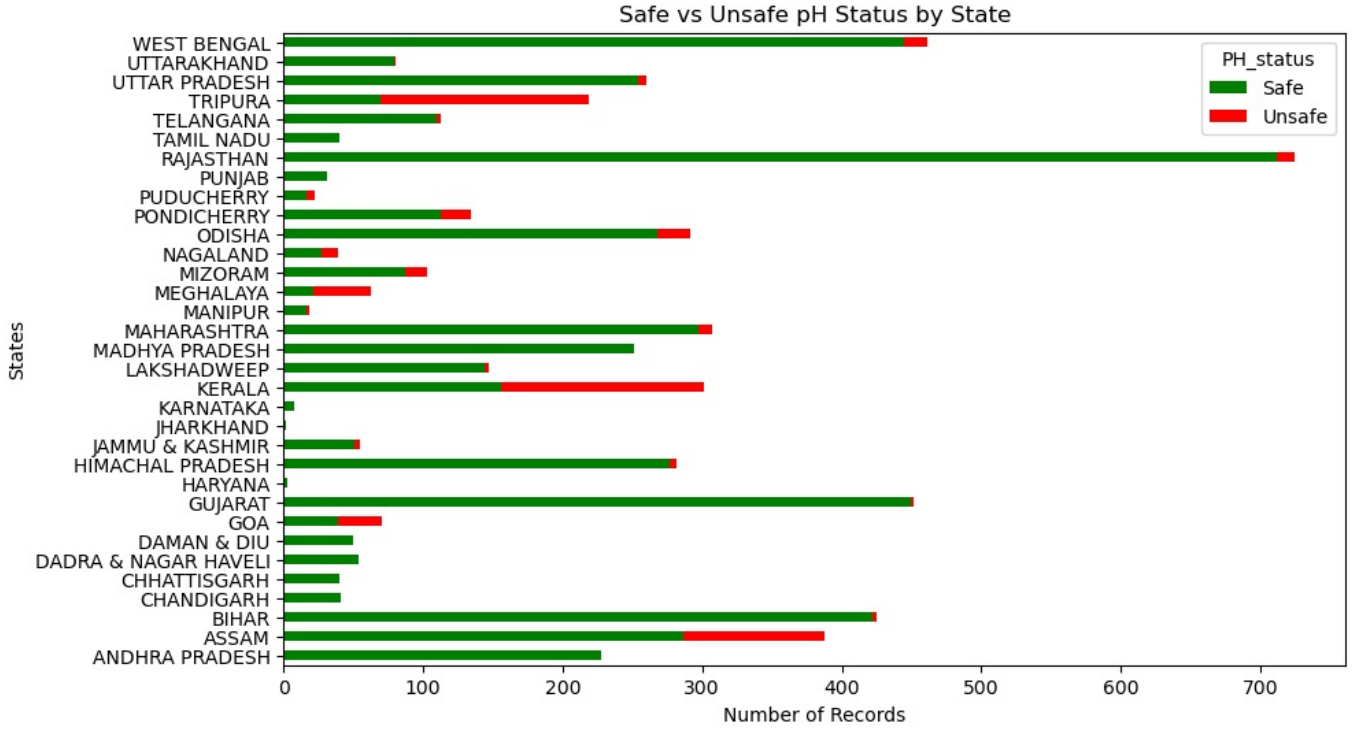


In [46]:
```
ax = df.groupby('STATE')['PH_avg'].mean().sort_values().plot(kind='barh', figsize=(10,6))

# Increase font size of axis values
plt.xticks(fontsize=14)
plt.yticks(fontsize=14)

# Add labels and title (optional)
plt.xlabel("Average pH", fontsize=16)
plt.ylabel("States", fontsize=16)
plt.title("Mean pH Levels by State", fontsize=18)

plt.show()
```



In [48]:

```
df.groupby(['STATE', 'PH_status']).size().unstack().plot(
    kind='barh',
    stacked=True,
    figsize=(10,6),
    color={'Safe':'green','Unsafe':'red'}
)

plt.xlabel("Number of Records")
plt.ylabel("States")
plt.title("Safe vs Unsafe pH Status by State")
plt.show()
```



In [ ]: