Impact of Forest Coverage on Urban Air Quality: A Global Analysis

Author: Avsar Vora

Table of contents

- Introduction
- Dataset Overview
- Data Load and Data Preparation
- Forest Area vs Air Quality Analysis
- Conclusion

Introduction

This research delves into the complex relationship between the forest areas of various countries and environmental factors, specifically focusing on Urban Air Quality metrics such as AQI, CO levels, and Ozone AQI, among others. Through comprehensive dataset analysis, the study seeks to identify patterns and correlations that highlight the broader impact of forest coverage on climate resilience across different nations.

The motivation for this research stems from the urgent need to understand the role of forest coverage in mitigating environmental challenges, particularly in urban areas where air quality issues are most pronounced. Forests play a crucial role in carbon sequestration, pollutant filtration, and overall climate regulation. By investigating the correlation between forest areas and urban air quality metrics, this study aims to provide valuable insights that can inform policy decisions, promote sustainable urban planning, and enhance climate resilience. Understanding these relationships is critical in the context of global environmental changes and the increasing pressure on urban ecosystems to adapt and thrive.

Goals

The main objective is to examine and illustrate the connections between:

- World's Forest Area Ratio Change since 1992
- Forest Area Ratio Change per Country
- Top 10 Countries with Expanding Forest Area Ratio
- Top 20 Countries with Most Forest Area Ratio (2021)
- Comparision of contries Air Quality with Forest cover (2021)

Dataset Overview

Datasource1: World Forest Area

- Metadata URL: https://www.kaggle.com/datasets/webdevbadger/world-forest-area/data
- Data URL: https://www.kaggle.com/datasets/webdevbadger/world-forest-area/download
- Data Source & Data Type: Kaggle - CSV
- License: Creative Commons Attribution 4.0 International License.

The dataset comprises 34 columns, encompassing detailed information such as the country name, country code, and the percentage of forest area as a proportion of the total land area. The dataset spans from 1990 to 2021, with individual columns representing the annual percentage of forest area for each year within this period.

Datasource2: World Air Quality Index

- Metadata URL: https://www.kaggle.com/datasets/adityaramachandran27/world-air-quality-index-by-city-and-coordinates/data
- Data URL: https://www.kaggle.com/datasets/adityaramachandran27/world-air-quality-index-by-city-and-coordinates/download
- Data Source & Data Type: Kaggle - CSV
- License: Creative Commons Attribution 4.0 International License.

This dataset contains detailed information on various countries, including the number of cities, their geographic coordinates (latitude and longitude), and different air quality indices recorded in 2021.

Data Load and Preparation

Import Packages

In [17]:
```
import os
import subprocess
import matplotlib.ticker as mtick
import pandas as pd
import numpy as np
import plotly.express as px
import pycountry
from matplotlib import pyplot as plt
from sqlalchemy import create_engine
import seaborn as sns

plt.style.use('ggplot')
```

```
percent_formatter = mtick.StrMethodFormatter('{x:,.0f}%')
percent_d_formatter = mtick.StrMethodFormatter('{x:,.00f}%')
```

Data Load: Load data from sqlite database for Forest Area and Air Quality Index

In [6]:
```
# If database doesn't exist, run data_collection pipeline.
if not (os.path.exists("../data/processed/world_air_quality_data.sqlite")
        "../data/processed/world_forest_data.sqlite")):
    subprocess.run(["./data_collection.sh"])

# Create sqlite engine to make connection
air_quality_engine = create_engine(f"sqlite:///../data/processed/world_ai
forest_engine = create_engine(f"sqlite:///../data/processed/world_forest_

# Load data from sqlite to pandas data frame
air_quality_df = pd.read_sql_table('world_air_quality_data', air_quality_
forest_df = pd.read_sql_table('world_forest_data', forest_engine)
```

In [7]:
```
air_quality_df.head()
```

Out[7]:

	Country	AQI Value	CO AQI Value	Ozone AQI Value	NO2 AQI Value	PM2.5 AQI Value
0	Afghanistan	86.333333	0.333333	42.000000	0.000000	86.333333
1	Albania	77.111111	1.000000	42.555556	0.555556	76.555556
2	Algeria	106.250000	4.000000	35.000000	25.750000	106.250000
3	Andorra	32.000000	1.000000	32.000000	0.000000	24.000000
4	Angola	85.000000	3.380952	23.190476	2.238095	82.523810

In [8]:
```
forest_df.head()
```

Out[8]:

	Country Name	1990	1991	1992	1993	1994	199
0	Afghanistan	1.852782	1.852782	1.852782	1.852782	1.852782	1.85278
1	Albania	28.788321	28.717153	28.645985	28.574818	28.503650	28.43248
2	Algeria	0.699908	0.696214	0.692519	0.688824	0.685129	0.68143
3	American Samoa	90.350000	90.180000	90.010000	89.840000	89.670000	89.50000
4	Andorra	34.042553	34.042553	34.042553	34.042553	34.042553	34.04255

5 rows × 33 columns

Data Preparation: Prepare data frames for Analysis

In [9]:
```
# Filter world forest data for timeline visualization
world_df = forest_df[forest_df["Country Name"] == "World"].set_index("Cou

# Country vise grouping of data for visualization of Forest area ratio ch
```

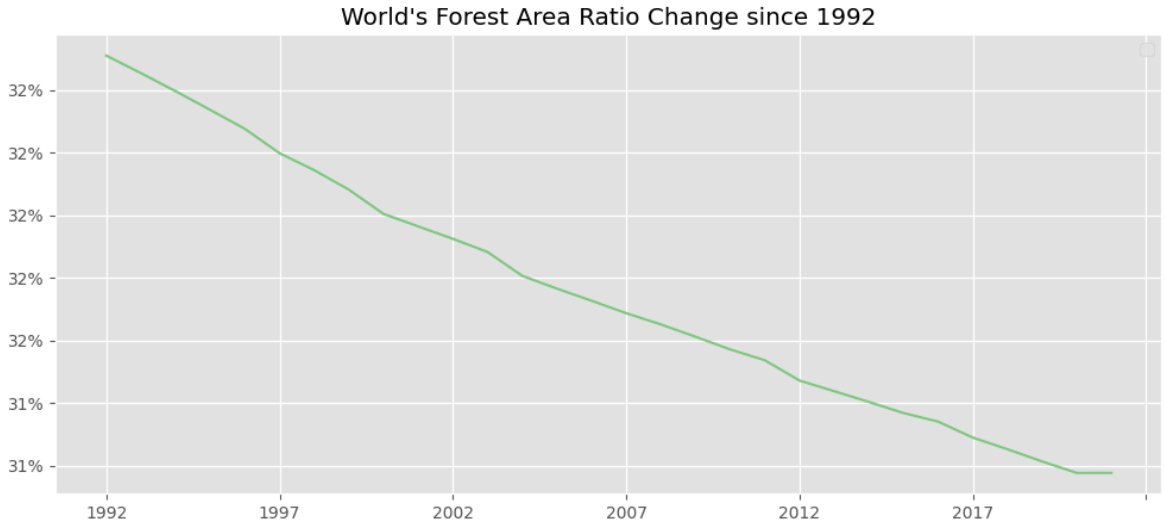
```
countries = []
for country in pycountry.countries:
    countries.append(country.name)
country_df = forest_df[forest_df["Country Name"].isin(countries)]
c_diff_df = country_df.set_index("Country Name").transpose().diff().sum()

# Top 10 countries with expanding forest area data for visualization
top_p_change_df = c_diff_df.sort_values("change", ascending=False).head(1

# Top 20 Countries with Most Forest Area Ratio (2021)
top_df = country_df.sort_values("2021", ascending=False).head(20)

# Combined data for heatmap visulization
df_2021 = country_df[["Country Name", "2021"]].rename(columns={"2021": "F
air_country_df = air_quality_df.groupby("Country")[
    ["AQI Value", "CO AQI Value", "Ozone AQI Value", "NO2 AQI Value", "PM
comb_df = pd.merge(df_2021, air_country_df, left_on="Country Name", right
```

Forest Area vs Air Quality Analysis

World's Forest Area Ratio Change since 1992

The graph illustrates that the world's forest area ratio decreased by 1.2% from 1992 to 2021.

In [10]:
```
fig, ax = plt.subplots(figsize=(12, 5))
world_df.plot(colormap="Accent", ax=ax)
ax.yaxis.set_major_formatter(percent_d_formatter)
plt.title("World's Forest Area Ratio Change since 1992")
plt.legend("")
plt.show()
```

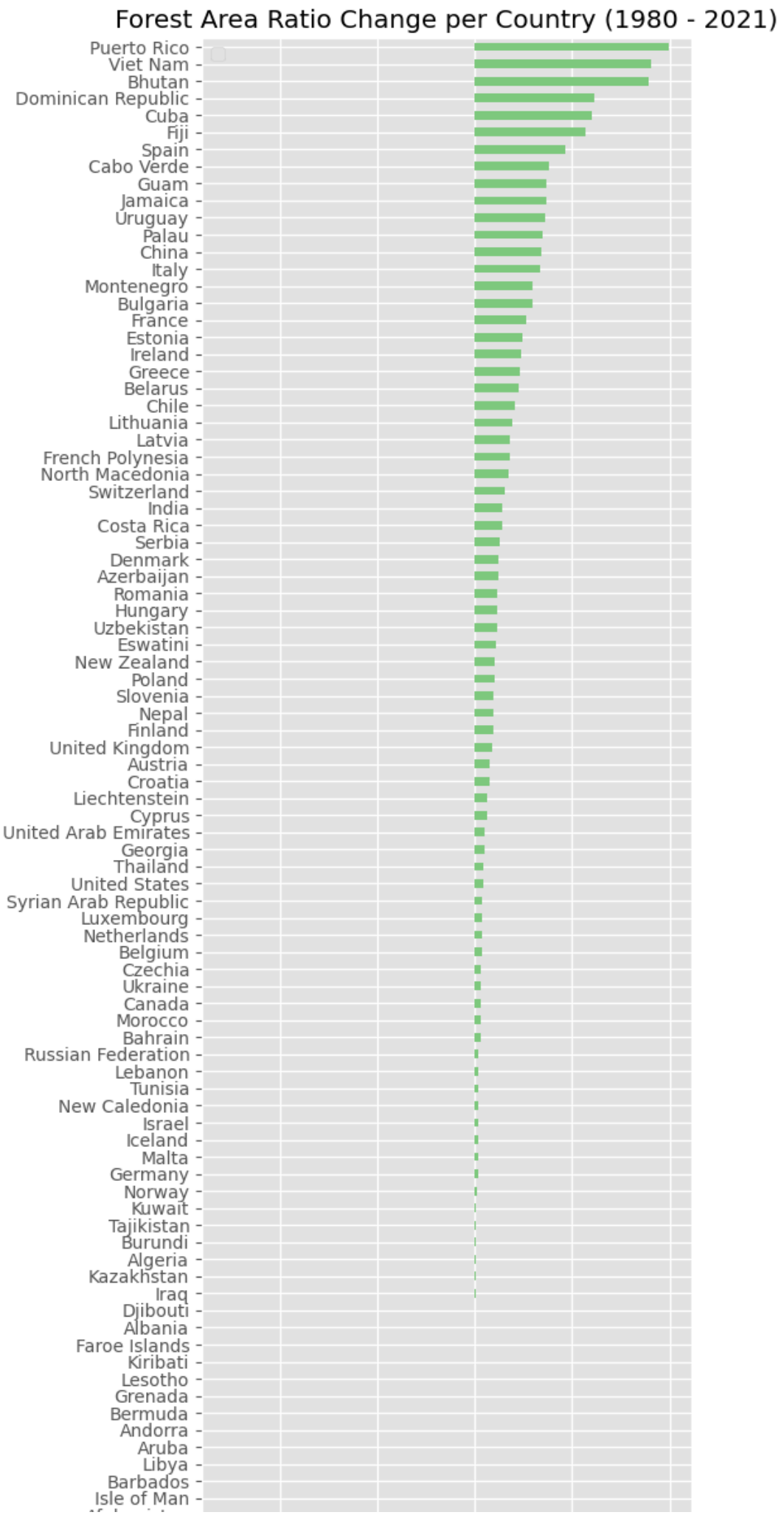


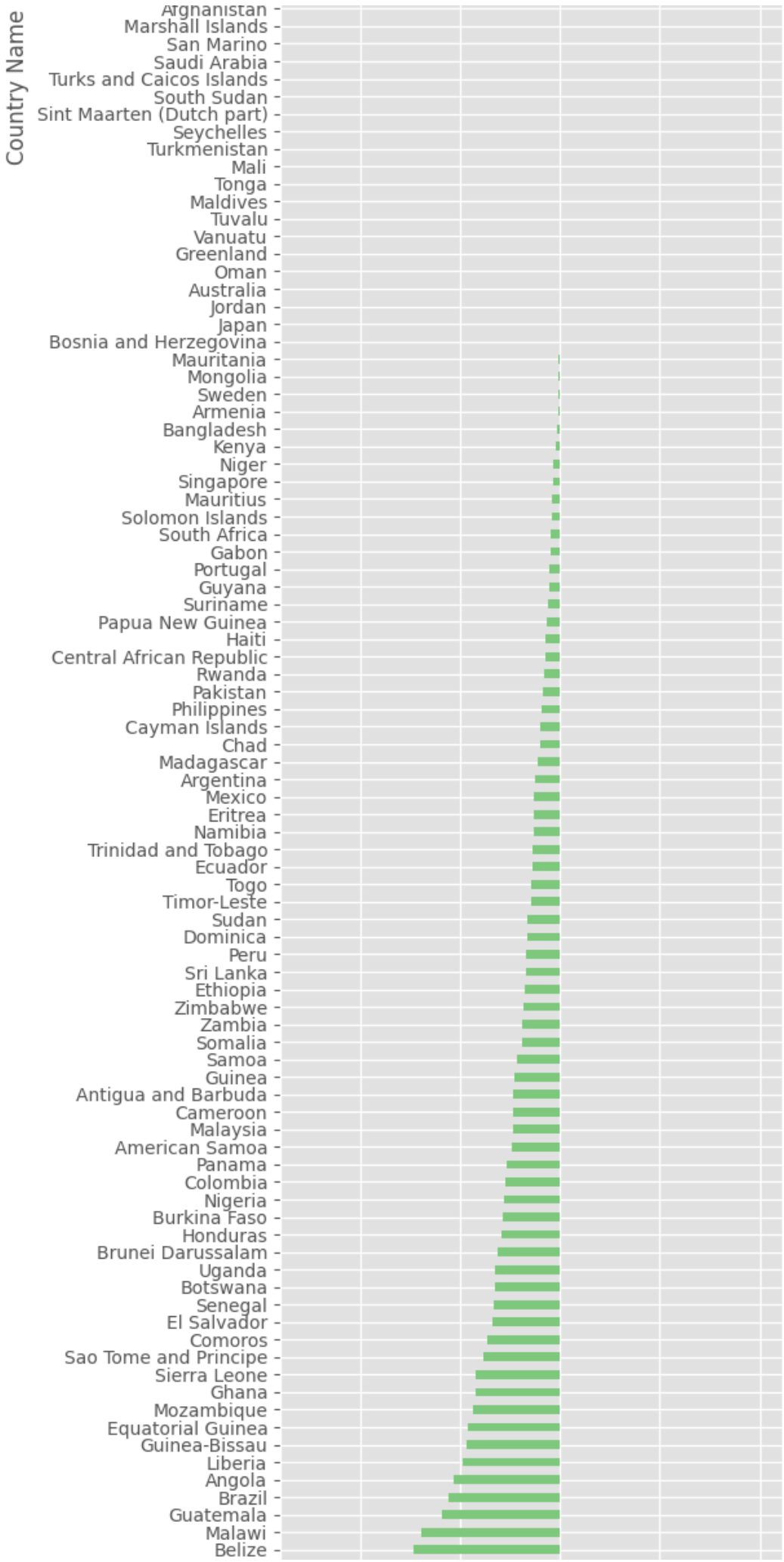
Forest Area Ratio Change per Country

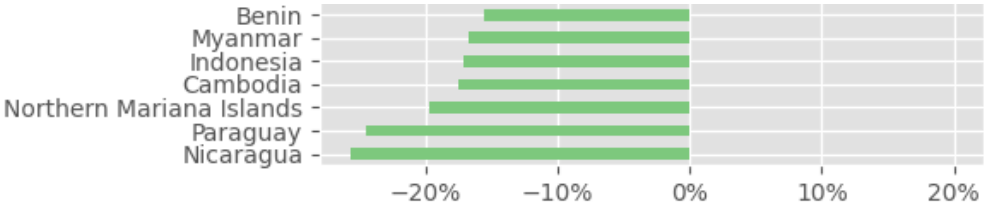
The graph depicts the change in forest area ratio per country, highlighting the varying degrees of deforestation or reforestation across different nations. Notable

trends and significant changes in forest coverage over the observed period are clearly visible.

In [11]:
```
ax = c_diff_df.sort_values("change").plot.barh(x="Country Name", y="chang
plt.title("Forest Area Ratio Change per Country (1980 - 2021)")
plt.legend("")
ax.xaxis.set_major_formatter(percent_formatter)
plt.show()
```


Forest Area Ratio Change per Country (1980 - 2021)
Puerto Rico
Viet Nam
Bhutan
Dominican Republic
Cuba
Fiji
Spain
Cabo Verde
Guam
Jamaica
Uruguay
Palau
China
Italy
Montenegro
Bulgaria
France
Estonia
Ireland
Greece
Belarus
Chile
Lithuania
Latvia
French Polynesia
North Macedonia
Switzerland
India
Costa Rica
Serbia
Denmark
Azerbaijan
Romania
Hungary
Uzbekistan
Eswatini
New Zealand
Poland
Slovenia
Nepal
Finland
United Kingdom
Austria
Croatia
Liechtenstein
Cyprus
United Arab Emirates
Georgia
Thailand
United States
Syrian Arab Republic
Luxembourg
Netherlands
Belgium
Czechia
Ukraine
Canada
Morocco
Bahrain
Russian Federation
Lebanon
Tunisia
New Caledonia
Israel
Iceland
Malta
Germany
Norway
Kuwait
Tajikistan
Burundi
Algeria
Kazakhstan
Iraq
Djibouti
Albania
Faroe Islands
Kiribati
Lesotho
Grenada
Bermuda
Andorra
Aruba
Libya
Barbados
Isle of Man


Country Name
Afghanistan
Marshall Islands
San Marino
Saudi Arabia
Turks and Caicos Islands
South Sudan
Sint Maarten (Dutch part)
Seychelles
Turkmenistan
Mali
Tonga
Maldives
Tuvalu
Vanuatu
Greenland
Oman
Australia
Jordan
Japan
Bosnia and Herzegovina
Mauritania
Mongolia
Sweden
Armenia
Bangladesh
Kenya
Niger
Singapore
Mauritius
Solomon Islands
South Africa
Gabon
Portugal
Guyana
Suriname
Papua New Guinea
Haiti
Central African Republic
Rwanda
Pakistan
Philippines
Cayman Islands
Chad
Madagascar
Argentina
Mexico
Eritrea
Namibia
Trinidad and Tobago
Ecuador
Togo
Timor-Leste
Sudan
Dominica
Peru
Sri Lanka
Ethiopia
Zimbabwe
Zambia
Somalia
Samoa
Guinea
Antigua and Barbuda
Cameroon
Malaysia
American Samoa
Panama
Colombia
Nigeria
Burkina Faso
Honduras
Brunei Darussalam
Uganda
Botswana
Senegal
El Salvador
Comoros
Sao Tome and Principe
Sierra Leone
Ghana
Mozambique
Equatorial Guinea
Guinea-Bissau
Liberia
Angola
Brazil
Guatemala
Malawi
Belize



Top 10 Countries with Expanding Forest Area Ratio

The graph showcases the top 10 countries with the most significant increases in forest area ratio, emphasizing those nations that have successfully expanded their forest coverage. These positive trends highlight effective reforestation and conservation efforts.

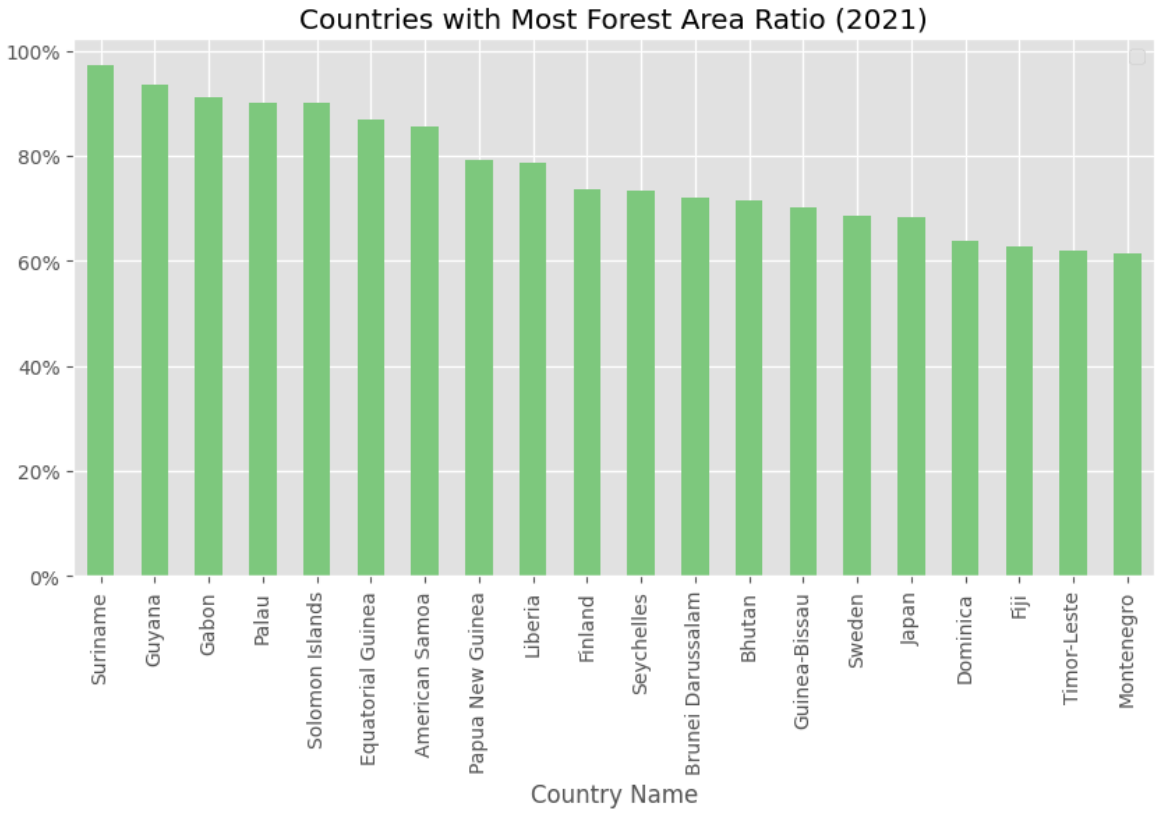
In [12]:
```
fig = px.pie(values=top_p_change_df["change"], names=top_p_change_df["Cou
             title="Top 10 Countries with Expanding Forest Area Ratio")
fig.update_traces(textposition='outside',
                  textinfo='percent+label',
                  marker=dict(line=dict(color='#FFFFFF', width=2)),
                  textfont_size=10)
fig.show()
```

Top 10 Countries with Expanding Forest Area Ratio

Top 20 Countries with Most Forest Area Ratio (2021)

The graph presents the top 20 countries with the highest forest area ratio in 2021, highlighting those with the largest proportion of land covered by forests. These countries play a crucial role in global biodiversity conservation and climate regulation due to their extensive forested areas.

In [13]:
```
fig, ax = plt.subplots(figsize=(10, 5))
top_df[["Country Name", "2021"]].plot.bar(x="Country Name", y="2021", col
plt.title("Countries with Most Forest Area Ratio (2021)")
plt.legend("")
ax.yaxis.set_major_formatter(percent_formatter)
plt.show()
```



Comparision of contries Air Quality with Forest cover (2021)

we are comparing with contries Air Quality using the World Air Quality Index by City and Coordinates dataset. Here we do see somewhat a significant correlation, especially with AQI, Ozone AQI, and PM2.5 AQI values.

In [18]:
```
plt.figure(figsize=(5, 5))
sns.heatmap(comb_df.select_dtypes(include=[np.number]).corr()[["Forest Ar
plt.show()
```

Conclusion

The analysis of 2021 datasets reveals a significant correlation between forest cover and air quality across various countries. Nations with higher forest area ratios tend to exhibit better air quality metrics, underscoring the critical role of forests in enhancing urban air quality and overall environmental health. This finding highlights the importance of forest conservation and reforestation efforts in mitigating air pollution and promoting sustainable urban environments.