Report for ForestQuery into Global Deforestation, 1990 to 2016

ForestQuery is on a mission to combat deforestation around the world and to raise awareness about this topic and its impact on the environment. The data analysis team at ForestQuery has obtained data from the World Bank that includes forest area and total land area by country and year from 1990 to 2016, as well as a table of countries and the regions to which they belong.

The data analysis team has used SQL to bring these tables together and to query them in an effort to find areas of concern as well as areas that present an opportunity to learn from successes.

1. GLOBAL SITUATION

According to the World Bank, the total forest area of the world was 41,282,694.9 sqkm in 1990. As of 2016, the most recent year for which data was available, that number had fallen to 39,958,245.9 sqkm, a loss of 1,324,449 sqkm, or 3.21%.

The forest area lost over this time period is slightly more than the entire land area of Peru listed for the year 2016 (which is 1,279,999.99 sqkm).

2. REGIONAL OUTLOOK

In 2016, the percent of the total land area of the world designated as forest was 31.38%. The region with the highest relative forestation was Latin America & Caribbean, with 46.16%, and the region with the lowest relative forestation was Middle East & North Africa, with 2.07% forestation.

In 1990, the percent of the total land area of the world designated as forest was 32.42%. The region with the highest relative forestation was Latin America & Caribbean, with 51.03%, and the region with the lowest relative forestation was Middle East & North Africa, with 1.78% forestation.

Table 2.1: Percent Forest Area by Region, 1990 & 2016:

Region	1990 Forest Percentage	2016 Forest Percentage
East Asia & Pacific	25.78%	26.36%
Europe & Central Asia	37.28%	38.04%
Latin America & Caribbean	51.03%	46.16%
Middle East & North Africa	1.78%	2.07%
North America	35.65%	36.04%
South Asia	16.51%	17.51%
Sub-Saharan Africa	30.67%	28.79%
World (Total)	32.42%	31.38%

The only regions of the world that decreased in percent forest area from 1990 to 2016 were Latin America & Caribbean (dropped from 51.03% to 46.16%) and Sub-Saharan Africa (30.67% to 28.79%). All other regions actually increased in forest area over this time period. However, the drop in forest area in the two aforementioned regions was so large, the percent forest area of the world decreased over this time period from 32.42% to 31.38%.

3. COUNTRY-LEVEL DETAIL

A. SUCCESS STORIES

There is one particularly bright spot in the data at the country level, China.
This country actually increased in forest area from 1990 to 2016 by 527,229.06 sqkm.
It would be interesting to study what has changed in this country over this time to drive this figure in the data higher. The country with the next largest increase in forest area from 1990 to 2016 was the United States, but it only saw an increase of 79,200 sqkm, much lower than the figure for China.

United States and China are of course very large countries in total land area, so when we look at the largest *percent* change in forest area from 1990 to 2016, we aren't surprised to find a much smaller country listed at the top. Iceland increased in forest area by 213.66% from 1990 to 2016.

B. LARGEST CONCERNS

Which countries are seeing deforestation to the largest degree? We can answer this question in two ways. First, we can look at the absolute square kilometer decrease in forest area from 1990 to 2016. The following 3 countries had the largest decrease in forest area over the time period under consideration:

Table 3.1: Top 5 Amount Decrease in Forest Area by Country, 1990 & 2016:

Country	Region	Absolute Forest Area Change
Brazil	Latin America & Caribbean	541,510 sqkm
Indonesia	East Asia & Pacific	282,193.98 sqkm
Myanmar	East Asia & Pacific	107,234 sqkm
Nigeria	Sub-Saharan Africa	106,506 sqkm
Tanzania	Sub-Saharan Africa	102,320 sqkm

The second way to consider which countries are of concern is to analyze the data by percent decrease.

Table 3.2: Top 5 Percent Decrease in Forest Area by Country, 1990 & 2016:

Country	Region	Pct Forest Area Change
Togo	Sub-Saharan Africa	75.45%
Nigeria	Sub-Saharan Africa	61.8%
Uganda	Sub-Saharan Africa	59.13%
Mauritania	Sub-Saharan Africa	46.75%
Honduras	Latin America & Caribbean	45.03%

When we consider countries that decreased in forest area percentage the most between 1990 and 2016, we find that four of the top 5 countries on the list are in the region of Sub-Saharan Africa. The countries are Togo, Nigeria, Uganda, and Mauritania. The 5th country on the list is Honduras, which is in the Latin America & Caribbean region.

From the above analysis, we see that Nigeria is the only country that ranks in the top 5 both in terms of absolute square kilometer decrease in forest as well as percent decrease in forest area

from 1990 to 2016. Therefore, this country has a significant opportunity ahead to stop the decline and hopefully spearhead remedial efforts.

C. QUARTILES

Table 3.3: Count of Countries Grouped by Forestation Percent Quartiles, 2016:

Quartile	Number of Countries
Q4	9
Q3	38
Q2	72
Q1	85

The largest number of countries in 2016 were found in the Q1 quartile.

There were 9 countries in the top quartile in 2016. These are countries with a very high percentage of their land area designated as forest. The following is a list of countries and their respective forest land, denoted as a percentage.

Table 3.4: Top Quartile Countries, 2016:

Country	Region	Pct Designated as Forest
Suriname	Latin America & Caribbean	98.26%
Micronesia, Fed. Sts.	East Asia & Pacific	91.86%
Gabon	Sub-Saharan Africa	90.04%

Region	Top Quartile	Percent Of Region
East Asia & Pacific	Q4	36.84%
Europe & Central Asia	Q4	63.79%
Latin America & Caribbean	Q3	45.24%
Middle East & North Africa	Q4	38.10%
North America	Q4	100%
South Asia	Q2	62.5%
Sub-Saharan Africa	Q1	56.25%

5. RECOMMENDATIONS

Write out a set of recommendations as an analyst on the ForestQuery team.
  - *What have you learned from the World Bank data?*
  - *Which countries should we focus on over others?*

Recommendations:
  - In recent years, deforestation has increased alarmingly.
  - Although in most regions of the world there was an increase in the area of the forest, it was small compared to the decrease in some areas which caused a decrease of 1.04% in the area of the forest in the world.
  - Most of the countries experiencing a decline in forest area are in Africa.
    A correlation exists between countries with low economic status and those that have lost the most forest area, so they deserve special attention.
    To preserve their forests in the future, these countries will need the help of more resistant countries.

Project Introduction: Setup to complete:
Creating a view:

CREATE OR REPLACE VIEW forestation
(
SELECT r.country_name, r.country_code, region, income_group, la.year,
total_area_sq_mi*2.59 AS total_area,
forest_area_sqkm AS forest_area,(forest_area_sqkm/total_area_sq_mi*2.59)*100 AS
percent_land_forest_area
FROM land_area AS la
JOIN forest_area AS fa
ON fa.country_code = la.country_code
AND fa.year = la.year
JOIN regions AS r
ON r.country_code = la.country_code)

SECTION 1

GLOBAL SITUATION
Instructions:

a. What was the total forest area (in sq km) of the world in 1990?
Please keep in mind that you can use the country record denoted as "World" in the
region table.

Query:
SELECT forest_area FROM forestation
WHERE region = 'World' AND year = 1990

b. What was the total forest area (in sq km) of the world in 2016?
Please keep in mind that you can use the country record in the table is denoted as
"World."

Query:
SELECT forest_area FROM forestation
WHERE region = 'World' AND year = 2016

c. What was the change (in sq km) in the forest area of the world from 1990 to 2016?

d. What was the percent change in forest area of the world between 1990 and 2016?

Query for 1.c, 1.d (including 1.a, 1.b):

```sql
WITH fa_1990 AS (
        SELECT forest_area
    FROM forestation
    WHERE region = 'World' AND year = 1990
), fa_2016 AS (
        SELECT forest_area
    FROM forestation
    WHERE region = 'World' AND year = 2016
)


SELECT f_90.forest_area AS f_90, f_16.forest_area AS f_16, f_90.forest_area-
f_16.forest_area AS fa_change,
      ROUND(((f_90.forest_area-f_16.forest_area)*100/f_90.forest_area)::NUMERIC,2)
AS fa_percent_change
FROM fa_1990 AS f_90, fa_2016 AS f_16
```

e. If you compare the amount of forest area lost between 1990 and 2016,
to which country's total area in 2016 is it closest to?

#for 1.e I created another view called q1 with the answers from question 1a,b,c,d:

```sql
CREATE OR REPLACE VIEW q1 AS(
WITH fa_1990 AS (
        SELECT forest_area
    FROM forestation
    WHERE region = 'World' AND year = 1990
), fa_2016 AS (
        SELECT forest_area
    FROM forestation
    WHERE region = 'World' AND year = 2016
)


SELECT f_90.forest_area AS f_90, f_16.forest_area AS f_16, f_90.forest_area-
f_16.forest_area AS fa_change,
      ROUND(((f_90.forest_area-f_16.forest_area)*100/f_90.forest_area)::NUMERIC,2)
AS fa_percent_change

FROM fa_1990 AS f_90, fa_2016 AS f_16)
```

Query:

```
SELECT country_name, ROUND(total_area::NUMERIC,2) AS total_area, fa_change,
ROUND((fa_change-total_area)::NUMERIC,2) AS diff_result
FROM forestation, q1
WHERE year = 2016 AND (fa_change-total_area) BETWEEN -100000 AND 100000
ORDER BY (fa_change-total_area) ASC
LIMIT 1
```

SECTION 2

REGIONAL OUTLOOK
Instructions:

I created another view called q2 that combined all the relevant data from the
Instructions:

```
CREATE VIEW q2 AS(
SELECT region,year, (SUM(forest_area)/SUM(total_area))*100 AS p_fa
FROM forestation
WHERE year IN (1990,2016)
GROUP BY region, year
ORDER BY region, year ASC)
```

a. What was the percent forest of the entire world in 2016?

Query
```
SELECT ROUND(p_fa::NUMERIC,2) AS p_fa FROM q2
WHERE region = 'World' AND year = 2016
```

Which region had the HIGHEST percent forest in 2016?

Query:
```
SELECT region, ROUND(p_fa::NUMERIC,2) FROM q2
WHERE year = 2016
ORDER BY p_fa DESC
LIMIT 1
```

Which had the LOWEST, to 2 decimal places?

Query:
```
SELECT region, ROUND(p_fa::NUMERIC,2) FROM q2
WHERE year = 2016
ORDER BY p_fa ASC
```

LIMIT 1

b. What was the percent forest of the entire world in 1990?

Query:
SELECT ROUND(p_fa::NUMERIC,2) FROM q2
WHERE region = 'World' AND year = 1990

Which region had the HIGHEST percent forest in 1990?

Query:
SELECT region, ROUND(p_fa::NUMERIC,2) FROM q2
WHERE year = 1990
ORDER BY p_fa DESC
LIMIT 1

Which had the LOWEST, to 2 decimal places?

Query:
SELECT region, ROUND(p_fa::NUMERIC,2) FROM q2
WHERE year = 1990
ORDER BY p_fa ASC
LIMIT 1

c. Based on the table you created, which regions of the world DECREASED in forest area from 1990 to 2016?

Query:
WITH y1 AS(
SELECT region,year, (SUM(forest_area)/SUM(total_area))*100 AS p_fa_16
FROM forestation
WHERE year= 2016
GROUP BY region, year
ORDER BY region, year ASC),
y2 AS (SELECT region,year, (SUM(forest_area)/SUM(total_area))*100 AS p_fa_90
FROM forestation
WHERE year= 1990
GROUP BY region, year
ORDER BY region, year ASC)

SELECT y1.region, ROUND(p_fa_16::NUMERIC,2) AS p_fa_16,
ROUND(p_fa_90::NUMERIC,2) AS p_fa_90,
ROUND(((p_fa_16-p_fa_90))::NUMERIC,2) AS p_diff
FROM y1

```
JOIN y2
ON y1.region = y2.region
WHERE (p_fa_16-p_fa_90) < 0 AND y1.region != 'World'
```

Table 2.1: Percent Forest Area by Region, 1990 & 2016:

```
Query:
WITH y1 AS(
SELECT region,year, (SUM(forest_area)/SUM(total_area))*100 AS p_fa_16
FROM forestation
WHERE year= 2016
GROUP BY region, year
ORDER BY region, year ASC),
y2 AS (SELECT region,year, (SUM(forest_area)/SUM(total_area))*100 AS p_fa_90
FROM forestation
WHERE year= 1990
GROUP BY region, year
ORDER BY region, year ASC)

SELECT y1.region, ROUND(p_fa_16::NUMERIC,2) AS p_fa_16,
ROUND(p_fa_90::NUMERIC,2) AS p_fa_90,
ROUND(((p_fa_16-p_fa_90))::NUMERIC,2) AS p_diff
FROM y1
JOIN y2
ON y1.region = y2.region
```

SECTION 3

COUNTRY-LEVEL DETAIL
Instructions:

```
# Starter window function:
Query:
WITH z1 AS(
SELECT region,year, country_name, (forest_area/total_area)*100 AS p_fa_16,
forest_area AS fa_16
FROM forestation
WHERE year= 2016
ORDER BY region,country_name),
z2 AS (SELECT region,year,country_name, (forest_area/total_area)*100 AS p_fa_90,
forest_area AS fa_90
FROM forestation
```

```
WHERE year= 1990
ORDER BY region,country_name)

SELECT z1.region,z1.country_name, p_fa_16, p_fa_90,(p_fa_16-p_fa_90) AS p_diff,
(fa_16-fa_90)*100/fa_90 AS p_change, z1.fa_16 AS fa_16, z2.fa_90 AS fa_90, (fa_16-
fa_90) AS fa_diff
FROM z1
JOIN z2
ON z1.country_name = z2.country_name


SUCCESS STORIES:

Query1:
WITH z1 AS(
SELECT region,year, country_name, (forest_area/total_area)*100 AS p_fa_16,
forest_area AS fa_16
FROM forestation
WHERE year= 2016
ORDER BY region,country_name),
z2 AS (SELECT region,year,country_name, (forest_area/total_area)*100 AS p_fa_90,
forest_area AS fa_90
FROM forestation
WHERE year= 1990
ORDER BY region,country_name)

SELECT z1.region,z1.country_name, ROUND(p_fa_16::NUMERIC,2) AS p_fa_16,
ROUND(p_fa_90::NUMERIC,2) AS p_fa_90, ROUND(((p_fa_16-
p_fa_90))::NUMERIC,2) AS p_diff,
ROUND(z1.fa_16::NUMERIC,2) AS fa_16, ROUND(z2.fa_90::NUMERIC,2) AS fa_90,
ROUND((fa_16-fa_90)::NUMERIC,2) AS fa_diff
FROM z1
JOIN z2
ON z1.country_name = z2.country_name
WHERE (fa_16-fa_90) > 0 AND z1.region != 'World'
ORDER BY fa_diff DESC
LIMIT 5

Query2:
WITH z1 AS(
SELECT region,year, country_name, (forest_area/total_area)*100 AS p_fa_16,
forest_area AS fa_16
FROM forestation
WHERE year= 2016
```

ORDER BY region,country_name),
z2 AS (SELECT region,year,country_name, (forest_area/total_area)*100 AS p_fa_90,
forest_area AS fa_90
FROM forestation
WHERE year= 1990
ORDER BY region,country_name)

SELECT z1.region,z1.country_name, p_fa_16, p_fa_90,(p_fa_16-p_fa_90) AS p_diff,
ROUND(((fa_16-fa_90)*100/fa_90)::NUMERIC,2) AS p_change, z1.fa_16 AS fa_16,
z2.fa_90 AS fa_90, (fa_16-fa_90) AS fa_diff
FROM z1
JOIN z2
ON z1.country_name = z2.country_name
WHERE (fa_16-fa_90) > 0 AND z1.region != 'World'
ORDER BY p_change DESC
LIMIT 1

a. Which 5 countries saw the largest amount decrease in forest area from 1990 to 2016?
What was the difference in forest area for each?

Query:
WITH z1 AS(
SELECT region,year, country_name, (forest_area/total_area)*100 AS p_fa_16,
forest_area AS fa_16
FROM forestation
WHERE year= 2016
ORDER BY region,country_name),
z2 AS (SELECT region,year,country_name, (forest_area/total_area)*100 AS p_fa_90,
forest_area AS fa_90
FROM forestation
WHERE year= 1990
ORDER BY region,country_name)

SELECT z1.region,z1.country_name, p_fa_16, p_fa_90,(p_fa_16-p_fa_90) AS p_diff,
(fa_16-fa_90)*100/fa_90 AS p_change, z1.fa_16 AS fa_16, z2.fa_90 AS fa_90, (fa_16-
fa_90) AS fa_diff
FROM z1
JOIN z2
ON z1.country_name = z2.country_name
WHERE (fa_16-fa_90) < 0 AND z1.region != 'World'
ORDER BY fa_diff ASC
LIMIT 5

b. Which 5 countries saw the largest percent decrease in forest area from 1990 to 2016? What was the percent change to 2 decimal places for each?

Query:
```
WITH z1 AS(
SELECT region,year, country_name, (forest_area/total_area)*100 AS p_fa_16,
forest_area AS fa_16
FROM forestation
WHERE year= 2016
ORDER BY region,country_name),
z2 AS (SELECT region,year,country_name, (forest_area/total_area)*100 AS p_fa_90,
forest_area AS fa_90
FROM forestation
WHERE year= 1990
ORDER BY region,country_name)

SELECT z1.region,z1.country_name, p_fa_16, p_fa_90,(p_fa_16-p_fa_90) AS p_diff,
ROUND(((fa_16-fa_90)*100/fa_90)::NUMERIC,2) AS p_change, z1.fa_16 AS fa_16,
z2.fa_90 AS fa_90, (fa_16-fa_90) AS fa_diff
FROM z1
JOIN z2
ON z1.country_name = z2.country_name
WHERE (fa_16-fa_90) < 0 AND z1.region != 'World'
ORDER BY p_change ASC
LIMIT 5
```

c. If countries were grouped by percent forestation in quartiles, which group had the most countries in it in 2016?

Query:
```
WITH z1 AS(
SELECT region,year, country_name, (forest_area/total_area)*100 AS p_fa_16,
forest_area AS fa_16
FROM forestation
WHERE year= 2016
ORDER BY region,country_name)

SELECT
CASE WHEN p_fa_16 > 75 THEN 'Q4'
    WHEN p_fa_16 < 75 AND p_fa_16 > 50 THEN 'Q3'
        WHEN p_fa_16 < 50 AND p_fa_16 > 25 THEN 'Q2'
ELSE 'Q1'
END AS quartiles, COUNT(*)
```

```
FROM z1
WHERE region != 'World' AND p_fa_16 IS NOT NULL
GROUP BY quartiles
ORDER BY quartiles DESC
```

d. List all of the countries that were in the 4th quartile (percent forest > 75%) in 2016.

Query:
```
WITH z1 AS(
SELECT region,year, country_name, (forest_area/total_area)*100 AS p_fa_16,
forest_area AS fa_16
FROM forestation
WHERE year= 2016
ORDER BY region,country_name)

SELECT region,country_name, p_fa_16,
CASE WHEN p_fa_16 > 75 THEN 'Q4'
    WHEN p_fa_16 < 75 AND p_fa_16 > 50 THEN 'Q3'
        WHEN p_fa_16 < 50 AND p_fa_16 > 25 THEN 'Q2'
ELSE 'Q1'
END AS quartiles
FROM z1
WHERE region != 'World' AND p_fa_16 IS NOT NULL
ORDER BY p_fa_16 DESC
LIMIT 9
```

Extra Queries:

I wanted to check the correlation between countries with a high percentage of deforestation
and their income levels when writing the recommendations.
Based on the opening data, the view "r_incomes" shows the distribution of income levels by region,
where Q4 represents high income and Q1 represents low income.
In the second view "p_max",
we show the maximum value of each region and what percentage of this value each country represents
As a result of the last query, it will be possible to see what the leading quarter is in each area,
and what percentage it represents.

```sql
CREATE VIEW r_incomes AS(
SELECT region,
CASE income_group
    WHEN 'High income' THEN 'Q4'
    WHEN  'Upper middle income' THEN 'Q3'
        WHEN 'Lower middle income' THEN 'Q2'
ELSE 'Q1'
END AS quartiles, COUNT(*) AS country_count
FROM regions
WHERE region != 'World'
GROUP BY region, quartiles
ORDER BY region,quartiles)

---------

CREATE VIEW p_max AS(
SELECT region, MAX(country_count) AS max_value,
ROUND(MAX(country_count)/SUM(country_count)*100,2) AS p
FROM r_incomes
WHERE region = 'East Asia & Pacific'
GROUP BY region

UNION

SELECT region, MAX(country_count) AS max_value,
ROUND(MAX(country_count)/SUM(country_count)*100,2) AS p
FROM r_incomes
WHERE region = 'Europe & Central Asia'
GROUP BY region

UNION

SELECT region, MAX(country_count) AS max_value,
ROUND(MAX(country_count)/SUM(country_count)*100,2) AS p
FROM r_incomes
WHERE region = 'Latin America & Caribbean'
GROUP BY region

UNION

SELECT region, MAX(country_count) AS max_value,
ROUND(MAX(country_count)/SUM(country_count)*100,2) AS p
```

```sql
FROM r_incomes
WHERE region = 'Middle East & North Africa'
GROUP BY region

UNION

SELECT region, MAX(country_count) AS max_value,
ROUND(MAX(country_count)/SUM(country_count)*100,2) AS p
FROM r_incomes
WHERE region = 'North America'
GROUP BY region

UNION

SELECT region, MAX(country_count) AS max_value,
ROUND(MAX(country_count)/SUM(country_count)*100,2) AS p
FROM r_incomes
WHERE region = 'South Asia'
GROUP BY region

UNION

SELECT region, MAX(country_count) AS max_value,
ROUND(MAX(country_count)/SUM(country_count)*100,2) AS p
FROM r_incomes
WHERE region = 'Sub-Saharan Africa'
GROUP BY region)
--------
SELECT r.region, r.quartiles,CAST(r.country_count AS NUMERIC), CAST(max_value
AS NUMERIC), p FROM p_max AS pm
JOIN r_incomes AS r
ON pm.region = r.region
ORDER BY region
```