

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 41,282,696 in 1990. As of 2016, the most recent year for which data was available, that number had fallen to 39,958,244, a loss of 1,324,452, or 3.2%.

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 1279999.99).

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
Latin America & Caribbean	51.03%	46.16%
Europe & Central Asia	37.28%	38.04%
North America	35.65%	36.04%
Sub-Saharan Africa	30.67%	28.79%
East Africa & Pacific	25.78%	26.36
South Asia	16.51%	17.51%
Middle East & North Africa	1.78%	2.07%

The only regions of the world that decreased in percent forest area from 1990 to 2016 were Latin America & Caribbean (dropped from 51.3% 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 527,229sqkm. 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 79200sqkm, much lower than the figure for China.

China and United States 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 2.14% 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
Togo	Sub-Saharan Africa	5168
Nigeria	Sub-Saharan Africa	106506
Uganda	Sub-Saharan Africa	28091
Mauritania	Sub-Saharan Africa	1940
Honduras	Latin America & Caribbean	36640

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	-75.44%
Nigeria	Sub-Saharan Africa	-61.80%
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
1	85
2	72
3	38
4	9

The largest number of countries in 2016 were found in the first 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
Gabon	Sub-Saharan Africa	90.04
Seychelles	Sub-Saharan Africa	88.41
Guyana	Latin America & Caribbean	83.90
Suriname	Latin America & Caribbean	98.26
Solomon Islands	East Asia & Pacific	77.86
Lao PDR	East Asia & Pacific	82.11
Micronesia, Fed Sts.	East Asia & Pacific	91.86
American Samoa	East Asia & Pacific	87.50
Pala	East Asia & Pacific	87.61

4. 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?*

Following the analysis carried out to explore the World Bank data on deforestation around the world, it was found that between 1990 and 2016, there was a 3.2% decrease in the level of forestation. In other words, several activities which led to an increase in deforestation have been carried out.

It was found that the top 5 countries including Togo, Nigeria, Uganda, Mauritania, and Honduras which had decreased in the forestation area are from the Sub-Saharan Africa region. Activities which contributed to a high level of deforestation have to be put in check in order to encourage forestation.

Generally, we could see from the different quartiles generated that a higher number of countries are distributed in Q1 to Q3, while only 9 countries had their forestation equal to or higher than 75%, most of which are from the East Asia & Pacific region. More awareness has to be created to make known the importance of forestation in order to see significant changes in the coming years.

5. APPENDIX: SQL Queries Used

-- DEFORESTATION EXPLORATION PROJECT

-- Steps to Complete

-- (1). Create a View called forestation by joining all three tables - forest_area, land_area and regions in the workspace.

-- (2). The forest_area and land_area tables join on both country_code AND year.

-- (3). The regions table joins these based on only country_code.

-- (4). In the forestation View, include the following:

-- (4a). All of the columns of the origin tables

-- (4b). A new column that provides the percent of the land area that is designated as forest.

-- (5). Keep in mind that the column forest_area_sqkm in the forest_area table and the land_area_sqmi

--in the land_area table are in different units (square kilometers and square miles, respectively),

--so an adjustment will need to be made in the calculation you write (1 sq mi = 2.59 sq km).

CREATE VIEW forestation AS

```
SELECT f.country_code AS forest_cc,
       f.country_name AS f_name,
       f.forest_area_sqkm AS f_sqkm,
       f.year AS f_year,
       l.total_area_sq_mi AS land_total_area_sqmi,
       r.region AS r_region,
       r.income_group AS r_income_group,
       (f.forest_area_sqkm/(l.total_area_sq_mi*2.59))*100 AS percent_forest_area
```

FROM forest_area f

JOIN land_area l

ON f.country_code = l.country_code AND f.year = l.year

JOIN regions r

ON f.country_code = r.country_code

-- forestation view output

SELECT *

FROM forestation

-- PART 1: GLOBAL SITUATION

-- 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.

SELECT forest_area_sqkm

FROM forest_area

WHERE country_name = '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 denoted as "World" in the region table.

```

SELECT forest_area_sqkm
FROM forest_area
WHERE country_name= 'World'
      AND year=2016;

```

-- c. What was the change (in sq km) in the forest area of the world from 1990 to 2016?
 -- this question can be solved in two different ways;

-- first solution using CASE
 -- The CASE expression goes through conditions and returns a value when the first condition is met
 --(like an if-then-else statement). So, once a condition is true, it will stop reading and return
 --the result. If no conditions are true, it returns the value in the ELSE clause.
 --If there is no ELSE part and no conditions are true, it returns NULL.

```

--CASE Syntax
--CASE
  --WHEN condition1 THEN result1
  --WHEN condition2 THEN result2
  --WHEN conditionN THEN resultN
  --ELSE result
--END;

```

```

SELECT Max(CASE
      WHEN year = 1990 THEN forest_area_sqkm
    END) - Max(CASE
      WHEN year = 2016 THEN forest_area_sqkm
    END) AS diff_forest_area
FROM forest_area
WHERE country_name = 'World';

```

-- second solution

```

SELECT ( sub1.forest_area_sqkm - sub2.forest_area_sqkm ) AS
      diff_forest_area_sqkm
FROM (SELECT f.country_code AS cc,
      f.forest_area_sqkm
    FROM forest_area f
    WHERE f.country_name = 'World'
      AND f.year = 1990) AS sub1
JOIN (SELECT f.country_code AS cc,
      f.forest_area_sqkm
    FROM forest_area f
    WHERE f.country_name = 'World'
      AND f.year = 2016) AS sub2
ON sub1.cc = sub2.cc;

```

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

```

SELECT Abs(( Max(CASE
      WHEN year = 2016 THEN forest_area_sqkm
    END) - Max(CASE
      WHEN year = 1990 THEN forest_area_sqkm

```

```

                END) ) / Max(CASE
            WHEN year = 1990 THEN forest_area_sqkm
                END)) * 100 AS percent_diff_forest_area
FROM forest_area
WHERE country_name = 'World'

```

--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?

--Common Table Expressions (CTE)

--We can define CTEs by adding a WITH clause directly before SELECT, INSERT, UPDATE, DELETE,

--or MERGE statement.

--After you define your WITH clause with the CTEs, you can then reference the CTEs as you would

--refer any other table. However, you can refer a CTE only within the execution scope of the

--statement that immediately follows the WITH clause.

WITH cte AS

```

(
    SELECT l.country_name,
           l.total_area_sq_mi*2.59 AS total_area_sqkm,
           Abs((l.total_area_sq_mi*2.59)-
    (
        SELECT Max(
            CASE
                WHEN year=1990 THEN forest_area_sqkm
            END)- Max(
            CASE
                WHEN year=2016 THEN forest_area_sqkm
            END) AS diff_forest_area
        FROM forest_area
        WHERE country_name= 'World')) AS diff_total_fa_land
    FROM land_area l
    WHERE l.year=2016
    ORDER BY 3
    LIMIT 1)
SELECT *
FROM cte
WHERE diff_total_fa_land IS NOT NULL;

```

-- PART 2: REGIONAL OUTLOOK

-- (2). Create a table that shows the Regions and their percent forest area (sum of forest area divided

--by sum of land area) in 1990 and 2016. (Note that 1 sq mi = 2.59 sq km).

CREATE VIEW regional_table
 AS

```

    (SELECT r.region,
           l.year,
           SUM(f.forest_area_sqkm)                AS
           total_fa_sqkm,
           SUM(l.total_area_sq_mi * 2.59)          AS

```



```

        total_la_sqkm,
        ( SUM(f.forest_area_sqkm) / SUM(l.total_area_sq_mi * 2.59) ) AS
        percent_fa_region
FROM   forest_area f
      JOIN land_area l
        ON f.country_code = l.country_code
        AND f.year = l.year
      JOIN regions r
        ON l.country_code = r.country_code
GROUP BY r.region,
        l.year
--ORDER BY r.region, l.year
)

```

```

-- regional_table view output
SELECT *
FROM regional_table

```

```

-- (2a.1). What was the percent forest of the entire world in 2016?
-- to 2 decimal places?

```

```

-- first solution

```

```

SELECT ROUND(CAST(SUM(total_fa_sqkm)/SUM(total_la_sqkm) AS numeric) *100, 2)
        AS perc_fa_region
FROM   regional_table
WHERE  region = 'World'
        AND year = 2016;

```

```

-- second solution

```

```

-- The CAST() function converts a value (of any type) into a specified datatype.
SELECT ROUND(CAST(percent_fa_region AS numeric) * 100, 2) AS perc_fa_region
FROM   regional_table
WHERE  region = 'World'
        AND year = 2016;

```

```

-- (2a.2) Which region had the HIGHEST percent
--forest in 2016, and which had the LOWEST,
--to 2 decimal places?

```

```

-- regional_table view output

```

```

SELECT *
FROM regional_table

SELECT region, ROUND(CAST(SUM(total_fa_sqkm)/SUM(total_la_sqkm) AS numeric)
                    *100, 2) AS highest_per_fa
FROM regional_table
WHERE year = 2016
GROUP BY region
ORDER BY highest_per_fa DESC;

```

```

-- (2b.1) What was the percent forest of the entire world in 1990?
--to 2 decimal places?

```

```

SELECT ROUND(CAST(SUM(total_fa_sqkm)/SUM(total_la_sqkm) AS numeric) *100, 2)
      AS perc_fa_region
FROM   regional_table
WHERE  region = 'World'
      AND year = 1990;

```

-- (2b.2) Which region had the HIGHEST percent forest in 1990, and which had the LOWEST,
 --to 2 decimal places?

```

-- regional_table view output
SELECT *
FROM regional_table

```

```

-- year 1990 % forest area
SELECT region, ROUND(CAST(SUM(total_fa_sqkm)/SUM(total_la_sqkm) AS numeric)
                    *100, 2) AS highest_per_fa
FROM regional_table
WHERE year = 1990
GROUP BY region
ORDER BY highest_per_fa DESC;

```

```

-- year 2016 % forest area
SELECT region, ROUND(CAST(SUM(total_fa_sqkm)/SUM(total_la_sqkm) AS numeric)
                    *100, 2) AS highest_per_fa
FROM regional_table
WHERE year = 2016
GROUP BY region
ORDER BY highest_per_fa DESC;

```

-- (2c) Based on the table you created, which regions of the world DECREASED
 --in forest area from 1990 to 2016?

```

--solution using a Subquery WITH placement
WITH table_1990 AS (
      SELECT region, ROUND(CAST(SUM(total_fa_sqkm)/SUM(total_la_sqkm) AS numeric)
                        *100, 2) AS highest_per_fa_1990
FROM regional_table
WHERE year = 1990
GROUP BY region),
table_2016 AS (
      SELECT region, ROUND(CAST(SUM(total_fa_sqkm)/SUM(total_la_sqkm) AS
numeric)
                        *100, 2) AS highest_per_fa_2016
FROM regional_table
WHERE year = 2016
GROUP BY region)
--ORDER BY highest_per_fa DESC
SELECT table_1990.region, highest_per_fa_1990, highest_per_fa_2016
FROM table_1990
JOIN table_2016
ON table_1990.region = table_2016.region
WHERE highest_per_fa_1990 > highest_per_fa_2016;

```

-- PART 3. COUNTRY-LEVEL DETAIL

-- (3a). Which 5 countries saw the largest amount decrease in forest area from 1990 to 2016?

What was

--the difference in forest area for each?

--SUCCESS STORIES

WITH top_5_countries

AS (SELECT fa_1990.country_name,
Sum(fa_2016.forest_area_sqkm) - Sum(fa_1990.forest_area_sqkm) AS
diff_fa

FROM forest_area fa_1990

JOIN forest_area fa_2016

ON fa_1990.country_code = fa_2016.country_code

WHERE fa_1990.year = 1990

AND fa_2016.year = 2016

GROUP BY fa_1990.country_name

--ORDER BY 2

)

SELECT t.country_name, r.region,

t.diff_fa

FROM top_5_countries t

JOIN regions r

ON r.country_name = t.country_name

WHERE t.diff_fa IS NOT NULL

AND t.country_name != 'World'

ORDER BY 3 DESC

--LARGEST CONCERNS

WITH top_5_countries

AS (SELECT fa_1990.country_name,
Sum(fa_2016.forest_area_sqkm) - Sum(fa_1990.forest_area_sqkm) AS
diff_fa

FROM forest_area fa_1990

JOIN forest_area fa_2016

ON fa_1990.country_code = fa_2016.country_code

WHERE fa_1990.year = 1990

AND fa_2016.year = 2016

GROUP BY fa_1990.country_name

--ORDER BY 2

)

SELECT t.country_name, r.region,

t.diff_fa

FROM top_5_countries t

JOIN regions r

ON r.country_name = t.country_name

WHERE t.diff_fa IS NOT NULL

AND t.country_name != 'World'

ORDER BY 3

LIMIT 5;

--LARGEST CONCERNS IN %

WITH top_5_countries

AS (SELECT fa_1990.country_name,

```

        Sum(fa_2016.forest_area_sqkm) - Sum(fa_1990.forest_area_sqkm) AS
        diff_fa,
        Sum((fa_2016.forest_area_sqkm) - (fa_1990.forest_area_sqkm))/
Sum(fa_1990.forest_area_sqkm) * 100 AS per_diff_fa
FROM forest_area fa_1990
JOIN forest_area fa_2016
ON fa_1990.country_code = fa_2016.country_code
WHERE fa_1990.year = 1990
AND fa_2016.year = 2016
GROUP BY fa_1990.country_name
--ORDER BY 2
)
SELECT t.country_name, r.region,
t.diff_fa, t.per_diff_fa
FROM top_5_countries t
JOIN regions r
ON r.country_name = t.country_name
WHERE t.diff_fa IS NOT NULL
AND t.country_name != 'World'
ORDER BY 3
LIMIT 5;

```

```

SELECT *
FROM forest_area

```

-- (3b). 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

```

WITH top_5_countries AS
(
    SELECT fa_1990.country_code,
           fa_1990.country_name,
           fa_1990.forest_area_sqkm - fa_2016.forest_area_sqkm AS
fa_diff_askdlnl,
           (fa_2016.forest_area_sqkm -
fa_1990.forest_area_sqkm)/fa_1990.forest_area_sqkm*100 AS percent_diff
FROM forest_area fa_1990
JOIN forest_area fa_2016
ON fa_1990.country_code=fa_2016.country_code
WHERE fa_1990.year=1990
AND fa_2016.year= 2016
--ORDER BY 2
)
SELECT t.country_name,
r.region,
fa_diff_askdlnl,
percent_diff
FROM top_5_countries t
JOIN regions r
ON r.country_code=t.country_code
WHERE fa_diff_askdlnl IS NOT NULL
AND t.country_name != 'World'
ORDER BY percent_diff

```

LIMIT 5;

-- (3c). If countries were grouped by percent forestation in quartiles, which group had the most countries
--in it in 2016?

```
WITH country_fa
  AS (SELECT f.country_code,
            f.country_name,
            (SUM(f.forest_area_sqkm) / SUM(l.total_area_sq_mi * 2.59 )) * 100 AS
country_percent
  FROM forest_area f
  JOIN land_area l
    ON f.country_code = l.country_code
    AND f.year = l.year
  WHERE f.year = 2016 AND l.year = 2016
    AND f.country_name != 'World'
    AND (f.forest_area_sqkm IS NOT NULL AND l.total_area_sq_mi IS NOT
NULL)
```

```
  GROUP BY f.country_code, f.country_name),
quartiles
  AS (SELECT country_code,
            country_name,

            --NTILE(4) OVER(ORDER BY country_percent desc) AS quartile
CASE
  WHEN country_percent >= 75 THEN 4
  WHEN country_percent < 75
    AND country_percent >= 50 THEN 3
  WHEN country_percent < 50
    AND country_percent >= 25 THEN 2
  ELSE 1
  END AS quartile
  FROM country_fa)
```

```
SELECT quartile,
       Count(quartile) AS number_of_countries
FROM quartiles
GROUP BY quartile
ORDER BY quartile;
```

-- (3d). List all of the countries that were in the 4th quartile (percent forest > 75%) in 2016.

```
WITH country_fa AS (
  SELECT
f.country_code,f.country_name,f.year,f.forest_area_sqkm,l.total_area_sq_mi*2.59 AS land_area
  FROM forest_area f
  JOIN land_area l
    ON f.country_code = l.country_code AND f.year=l.year
  WHERE f.year=2016 and f.country_name != 'World'
),
country_percent_fa AS(
  SELECT country_code,country_name,year,(forest_area_sqkm/land_area) *100 AS
country_percent
```

```

        FROM country_fa
    ),
    quartiles AS (
        SELECT country_code, country_percent,
               country_name, year,
               CASE WHEN country_percent >= 75 THEN 4
                    WHEN country_percent < 75 AND country_percent >= 50 THEN 3
                    WHEN country_percent < 50 AND country_percent >= 26 THEN 2
                    ELSE 1
               END AS quartile
        FROM country_percent_fa
    )
    SELECT quartile, q.country_name, r.region, country_percent
    FROM quartiles q
    JOIN regions r
    ON r.country_code = q.country_code
    WHERE quartile = 4
    ORDER BY 3 DESC;

```

-- (3e). How many countries had a percent forestation higher than the United States in 2016?

```

WITH country_fa
    AS (SELECT f.country_code,
              f.country_name,
              f.year,
              f.forest_area_sqkm,
              l.total_area_sq_mi * 2.59 AS land_area
    FROM forest_area f
    JOIN land_area l
    ON f.country_code = l.country_code
    AND f.year = l.year
    WHERE f.year = 2016
    AND f.country_name != 'World'),
    country_percent_fa
    AS (SELECT country_code,
              country_name,
              year,
              ( forest_area_sqkm / land_area ) * 100 AS country_percent
    FROM country_fa),
    quartiles
    AS (SELECT country_code,
              country_name,
              year,
              --NTILE(4) OVER(ORDER BY country_percent desc) AS quartile
              CASE
                WHEN country_percent >= 75 THEN 4
                WHEN country_percent < 75
                  AND country_percent >= 50 THEN 3
                WHEN country_percent < 50
                  AND country_percent >= 26 THEN 2
                ELSE 1
              END AS quartile
    FROM country_percent_fa)
    SELECT COUNT(*) AS number_of_countries

```

```
FROM country_percent_fa
WHERE year = 2016
      AND country_percent > (SELECT country_percent
                             FROM country_percent_fa
                             WHERE country_name = 'United States'
                             AND year = 2016);
```