

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 41282694.9 km² in 1990. As of 2016, the most recent year for which data was available, that number had fallen to 39958245.9 km², a loss of 1324449 km², 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 1279999.9891 km²).

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
Sub-Saharan Africa	30.67	28.79
Europe & Central Asia	37.28	38.04
East Asia & Pacific	25.78	26.36
South Asia	16.51	17.51
Middle East & North Africa	1.78	2.07
North America	35.65	36.04
World	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 527229.062km². 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 79200km², 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 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	541510.00
Indonesia	East Asia & Pacific	282193.98
Myanmar	East Asia & Pacific	107234.00
Nigeria	Sub-Saharan Africa	106506.00
Tanzania	Sub-Saharan Africa	102320.00

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	61.80
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 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
0 - 25%	85
25 - 50%	73
50 - 75%	38
75 - 100%	9

The largest number of countries in 2016 were found in the 0 - 25% 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	98.2576939676578
Micronesia, Fed. Sts.	East Asia & Pacific	91.8572390715248
Gabon	Sub-Saharan Africa	90.0376418700565
Seychelles	Sub-Saharan Africa	88.4111367385789
Palau	East Asia & Pacific	87.6068085491204
American Samoa	East Asia & Pacific	87.5000875000875
Guyana	Latin America & Caribbean	83.9014489110682
Lao PDR	East Asia & Pacific	82.1082317640861
Solomon Islands	East Asia & Pacific	77.8635177945066

94 countries had percent forestation higher than the United States in 2016.

4. RECOMMENDATIONS

We found out, that the forests of the entire world decreased from 41282694.9 km² to 39958245.9 km². That is more than the size of Peru. If things continue like this, in 785 years we won't have any trees left in the world.

There is good news:

Except 2 two regions, Latin America & Caribbean and Sub-Saharan Africa the forestation of the earth increases since 1990. The amount of forest of these two regions is so high, that the total forestation of the world is decreasing.

We must observe why countries like Nigeria keep clearing their forests and find a way to stop it. Finding a solution for Nigeria is necessary, because it ranks on the top 5 lists of the absolute square kilometers decrease in forest and the percentage decrease in forest.

If we can stop deforestation in countries like Brazil, Nigeria and Togo, we are able to increase the amount of other crisis areas and therefore forest globally.

We should investigate how countries like Suriname and Micronesia, Fed. Sts. keep their amount of forest at a high level and we should investigate how countries like United States and China increase their forestation to get a higher number of trees on the earth.

5. Appendix

I create a View called **forestation** by joining all three tables in the workspace:

forest_area, **land_area**, and **regions**. The View contains a column **percent_of_land_is_forest**, it shows how much percent of the total area is forest.

Because there may be unclear results from the data, I must clean it. In the **forestation** View are only countries with complete data.

There are 20 countries with missing data:

```
SELECT COUNT(DISTINCT f.country_name)
FROM forest_area AS f
JOIN land_area AS l
      USING (country_code, year)
WHERE (f.forest_area_sqkm IS NULL
      OR l.total_area_sq_mi IS NULL)
AND (f.year = 1990 OR f.year = 2016)
```

Now I create the View:

```
CREATE VIEW forestation AS
SELECT f.country_code,
       f.country_name,
       f.year,
       f.forest_area_sqkm,
       l.total_area_sq_mi,
       r.region,
       r.income_group,
       f.forest_area_sqkm / (l.total_area_sq_mi * 2.59) * 100 AS
percent_of_land_is_forest
FROM forest_area AS f
JOIN land_area AS l
      ON l.country_code = f.country_code
AND l.year = f.year
JOIN regions AS r
      ON r.country_code = f.country_code
WHERE forest_area_sqkm IS NOT NULL
      AND total_area_sq_mi IS NOT NULL
```

1. Global Situation

To answer the questions a., b., c., and d., I write the follow query:

```
SELECT f90.country_name,
       f90.forest_area_sqkm AS forest_area_1990,
       f16.forest_area_sqkm AS forest_area_2016,
       f90.forest_area_sqkm - f16.forest_area_sqkm AS difference,
       --it's not asked, but I think it looks better with ROUND()
       ROUND(CAST(((f90.forest_area_sqkm - f16.forest_area_sqkm) /
                    f90.forest_area_sqkm)*100 AS NUMERIC), 2) AS
difference_percent
FROM (SELECT * FROM forestation WHERE year = 1990) AS f90
JOIN (SELECT * FROM forestation WHERE year = 2016) AS f16
     USING (country_code)
WHERE f90.country_name = 'World'
     OR f16.country_name = 'World'
LIMIT 1
```

To answer the question e., I write another query:

```
SELECT country_name,
       total_area_sq_mi*2.59 AS total_area_sqkm
FROM forestation
WHERE (total_area_sq_mi*2.59) <
      (SELECT f90.forest_area_sqkm - f16.forest_area_sqkm AS
difference
FROM (SELECT * FROM forestation WHERE year = 1990) AS f90
JOIN (SELECT * FROM forestation WHERE year = 2016) AS f16
     USING (country_code)
WHERE f90.country_name = 'World'
     OR f16.country_name = 'World')
ORDER BY total_area_sqkm DESC
LIMIT 1
```

2. Regional Outlook

a.

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

```
SELECT ROUND(CAST(forest_area_sqkm/(total_area_sq_mi*2.59)*100 AS
numeric), 2) AS forest_percentage_2016
FROM forestation
WHERE country_name = 'World'
AND year = 2016
```

Which region had the **HIGHEST** percent forest in 2016?

```
SELECT region,
ROUND(CAST((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59)*100)
AS numeric), 2) AS forest_percentage_2016
FROM forestation
WHERE year = 2016
GROUP BY region
ORDER BY forest_percentage_2016 DESC
LIMIT 1
```

And which had the **LOWEST**?

```
SELECT region,
ROUND(CAST((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59)*100)
AS numeric), 2) AS forest_percentage_2016
FROM forestation
WHERE year = 2016
GROUP BY region
ORDER BY forest_percentage_2016
LIMIT 1
```


b.

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

```
SELECT ROUND(CAST(forest_area_sqkm/(total_area_sq_mi*2.59)*100 AS
numeric), 2) AS forest_percentage_1990
FROM forestation
WHERE country_name = 'World'
AND year = 1990
```

Which region had the HIGHEST percent forest in 1990?

```
SELECT region,
ROUND(CAST((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59)*100)
AS numeric), 2) AS forest_percentage_1990
FROM forestation
WHERE year = 1990
GROUP BY region
ORDER BY forest_percentage_1990 DESC
LIMIT 1
```

And which had the LOWEST?

```
SELECT region,
ROUND(CAST((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59)*100)
AS numeric), 2) AS forest_percentage_1990
FROM forestation
WHERE year = 1990
GROUP BY region
ORDER BY forest_percentage_1990
LIMIT 1
```

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

```
SELECT pf90.region,
       pf90.fp90 AS i1990_percent_forestation,
       pf16.fp16 AS i2016_percent_forestation
--FROM with subquery
FROM (SELECT region,
ROUND(CAST((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59)*100)
AS numeric), 2) AS fp90
FROM forestation WHERE year = 1990
GROUP BY region) AS pf90
--JOIN with subquery
JOIN (SELECT region,
ROUND(CAST((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59)*100)
AS numeric), 2) AS fp16
FROM forestation WHERE year = 2016
GROUP BY region) AS pf16
--End of subquery, instead of ON, I use USING, it's easier/shorter
USING (region)
```

c.

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

```
SELECT pf90.region,
       pf90.fp90 AS i1990_percent_forestation,
       pf16.fp16 AS i2016_percent_forestation
--FROM with subquery
FROM (SELECT region,
ROUND(CAST((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59)*100)
AS numeric), 2) AS fp90
FROM forestation WHERE year = 1990
GROUP BY region) AS pf90
--JOIN with subquery
JOIN (SELECT region,
ROUND(CAST((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59)*100)
AS numeric), 2) AS fp16
FROM forestation WHERE year = 2016
GROUP BY region) AS pf16
--End of subquery, instead of ON, I use USING, it's easier/shorter
USING (region)
WHERE pf90.fp90 > pf16.fp16
```

3. Country-Level Detail

a.

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

The question in Udacity asks in 3.a. “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?” but the project wants to know the increase, so I calculated the increase.

```
SELECT DISTINCT f.country_name,
               f.region,
               f90.forest_area_sqkm AS forest_area_sqkm_1990,
               f16.forest_area_sqkm AS forest_area_sqkm_2016,
               f16.forest_area_sqkm - f90.forest_area_sqkm AS difference,
               ROUND(CAST(f16.percent_of_land_is_forest -
               f90.percent_of_land_is_forest AS NUMERIC), 2)
               AS difference_percent
FROM forestation AS f
JOIN (SELECT * FROM forestation WHERE year = 1990) AS f90
    ON f.country_name = f90.country_name
JOIN (SELECT * FROM forestation WHERE year = 2016) AS f16
    ON f.country_name = f16.country_name
WHERE NOT f.region = 'World'
GROUP BY f.country_name,
         f.region,
         f90.forest_area_sqkm,
         f16.forest_area_sqkm,
         f16.percent_of_land_is_forest,
         f90.percent_of_land_is_forest
ORDER BY difference DESC
LIMIT 5
```

In this query, I find the country which has the most percentage change in forestation

```
SELECT f90.country_name,
       ROUND(CAST(((f16.forest_area_sqkm -
       f90.forest_area_sqkm) / f90.forest_area_sqkm) * 100 AS NUMERIC), 2)
       AS difference_percent
FROM (SELECT * FROM forestation WHERE year = 1990) AS f90
JOIN (SELECT * FROM forestation WHERE year = 2016) AS f16
    USING (country_name)
GROUP BY f90.country_name,
         f90.forest_area_sqkm,
         f16.forest_area_sqkm
ORDER BY difference_percent DESC
LIMIT 1
```

b.

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

```
SELECT f.country_name,
       f.region,
       f90.forest_area_sqkm AS forest_area_sqkm_1990,
       f16.forest_area_sqkm AS forest_area_sqkm_2016,
       ROUND(CAST(f90.forest_area_sqkm -
f16.forest_area_sqkm AS NUMERIC), 2) AS difference,
       ROUND(CAST(f16.percent_of_land_is_forest -
f90.percent_of_land_is_forest AS NUMERIC), 2)
AS difference_percent
FROM forestation AS f
JOIN (SELECT * FROM forestation WHERE year = 1990) AS f90
ON f.country_name = f90.country_name
JOIN (SELECT * FROM forestation WHERE year = 2016) AS f16
ON f.country_name = f16.country_name
WHERE NOT f.region = 'World'
GROUP BY f.country_name,
         f.region,
         f90.forest_area_sqkm,
         f16.forest_area_sqkm,
         f90.percent_of_land_is_forest,
         f16.percent_of_land_is_forest
ORDER BY difference DESC
LIMIT 5
```

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

```
SELECT f.country_name,
       f.region,
       f90.forest_area_sqkm AS forest_area_sqkm_1990,
       f16.forest_area_sqkm AS forest_area_sqkm_2016,
       ROUND(CAST(f90.forest_area_sqkm -
f16.forest_area_sqkm AS numeric), 2) AS difference,
       ROUND(CAST(((f90.forest_area_sqkm -
f16.forest_area_sqkm) / f90.forest_area_sqkm)*100 AS NUMERIC), 2)
AS difference_percent
FROM forestation AS f
JOIN (SELECT * FROM forestation WHERE year = 1990) AS f90
ON f.country_name = f90.country_name
JOIN (SELECT * FROM forestation WHERE year = 2016) AS f16
ON f.country_name = f16.country_name
WHERE NOT f.region = 'World'
GROUP BY f.country_name,
         f.region,
         f90.forest_area_sqkm,
         f16.forest_area_sqkm,
         f90.percent_of_land_is_forest,
         f16.percent_of_land_is_forest
ORDER BY difference_percent DESC
LIMIT 5
```

c.

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

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

```
SELECT f.quartiles, COUNT(f.country_name)
FROM (
  SELECT country_name,
    CASE WHEN percent_of_land_is_forest >= 0
    AND percent_of_land_is_forest < 25
      THEN '0 - 25%'
    WHEN percent_of_land_is_forest >= 25
    AND percent_of_land_is_forest < 50
      THEN '25 - 50%'
    WHEN percent_of_land_is_forest >= 50
    AND percent_of_land_is_forest < 75
      THEN '50 - 75%'
    ELSE '75 - 100%'
    END AS quartiles
  FROM forestation
  WHERE year = 2016) AS f
GROUP BY quartiles
ORDER BY f.quartiles
```

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

Table 3.4: Top Quartile Countries, 2016:

```
SELECT country_name, region, percent_of_land_is_forest
FROM forestation
WHERE year = 2016
AND percent_of_land_is_forest > 75
ORDER BY percent_of_land_is_forest DESC
```

e.

How many countries had a percent forestation higher than the United States in 2016?

```
SELECT COUNT(country_name)
FROM forestation
WHERE percent_of_land_is_forest > (SELECT percent_of_land_is_forest
  FROM forestation
  WHERE country_name = 'United States'
  AND year = 2016)
AND year = 2016
```

4. # I added this query to get the number of years, how long it takes to erase all forests from earth

```
SELECT CEIL(f16.forest_area_sqkm/((f90.forest_area_sqkm -  
f16.forest_area_sqkm)/(f16.year - f90.year))) AS difference  
FROM (SELECT * FROM forestation WHERE year = 1990) AS f90  
JOIN (SELECT * FROM forestation WHERE year = 2016) AS f16  
USING (country_code)  
WHERE f90.country_name = 'World'  
OR f16.country_name = 'World'  
LIMIT 1
```