Report for ForestQuery into Global Deforestation, 1990 to 2016

Rayner Mendez

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 41,282,694.90 sq km in 1990. As of 2016, the most recent year for which data was available, that number had fallen to39,958,245.90 sq km, a loss of 1,324,449 sq km, 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 1,279,995.05 sq km).

## 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 |
| Sub-Saharan-Africa | 30.67% | 28.79% |
| World | 32.42% | 31.38% |
| Latin America & Caribbean | 51.03% | 46.16% |

The only regions of the world that decreased in percent forest area from 1990 to 2016 were Sub-Saharan-Africa (dropped from 30.67% to 28.79%) and Latin America & Caribbean (51.03% to 46.16%). 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**

### 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 33.55%. 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 2.62%, much lower than the figure for China.

China and the 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. The Dominican Republic increased in forest area by 82.46% from 1990 to 2016.

### 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 sq km |
| Indonesia | East Asia & Pacific | 282,193.98 sq km |
| Myanmar | East Asia & Pacific | 107,234 sq km |
| Nigeria | Sub-Saharan Africa | 106,506 sq km |
| Tanzania | Sub-Sahara Africa | 102,320 sq km |

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 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.

### 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 1 quartile.

There were 85 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 |
| American Samoa | East Asia & Pacific | 87.50% |
| Micronesia, Fed. Sts. | East Asia & Pacific | 91.86% |
| Gabon | Sub-Saharan Africa | 90.04% |
| Guyana | Latin America & Caribbean | 83.9% |
| Lao PDR | East Asia & Pacific | 82.11% |
| Palau | East Asia & Pacific | 87.61% |
| Solomon Islands | East Asia & Pacific | 77.86% |
| Suriname | Latin America & Caribbean | 98.26% |
| Seychelles | Sub-Saharan African | 88.41% |

## 5. RECOMMENDATIONS

*Write out a set of recommendations as an analyst on the ForestQuery team.*

* *What have you learned from the World Bank data?*

*I have learned that my country, The Dominican Republic has been doing pretty good preserving the flora.*

* *Which countries should we focus on over others?*

*Based on my analysis I would focus on the top 10 countries in terms of less forestation. These countries are Togo, Nigeria, Uganda, Mauritania, Honduras, Pakistan, Niger, Kore dem, people’s republic, Zimbabwe, and last but not least Nicaragua.*

*Appendix: SQL Queries*

-- Tables to insert data to local database.  
CREATE TABLE forest\_area (  
 country\_code VARCHAR,  
 country\_name VARCHAR,  
 year SMALLINT,  
 forest\_area\_sqkm DOUBLE PRECISION  
);  
  
CREATE TABLE land\_area (  
 country\_code VARCHAR,  
 country\_name VARCHAR,  
 year SMALLINT,  
 total\_area\_sq\_mi DOUBLE PRECISION  
);  
  
CREATE TABLE regions (  
 country\_name VARCHAR,  
 country\_code VARCHAR,  
 region VARCHAR,  
 income\_group VARCHAR  
);

-- Create a view to join the tables.  
CREATE OR REPLACE VIEW forestation AS  
 SELECT fa.country\_code AS country\_code,  
 fa.country\_name AS country\_name,  
 fa.year AS year,  
 fa.forest\_area\_sqkm AS forest\_area\_sqkm,  
 (la.total\_area\_sq\_mi \* 2.58999 ) AS total\_area\_sqkm,  
 r.region AS region,  
 r.income\_group AS income\_group,  
 (fa.forest\_area\_sqkm / (la.total\_area\_sq\_mi \* 2.58999 )) \* 100 AS forestation\_percentage\_area  
 FROM forest\_area fa  
 JOIN land\_area la  
 ON fa.country\_code = la.country\_code  
 JOIN regions r  
 ON fa.country\_code = r.country\_code  
 WHERE fa.year = la.year;

-- Global Situation  
WITH table\_1990 AS (SELECT country\_code, *SUM*(*COALESCE*(forest\_area\_sqkm, 0)) AS forest\_area\_sqkm  
 FROM forest\_area  
 WHERE year = 1990 AND country\_name = 'World'  
 GROUP BY 1),  
 table\_2016 AS (SELECT country\_code, *SUM*(*COALESCE*(forest\_area\_sqkm, 0)) AS forest\_area\_sqkm  
 FROM forest\_area  
 WHERE year = 2016 AND country\_name = 'World'  
 GROUP BY 1)  
SELECT *ROUND*(*CAST*(a.forest\_area\_sqkm AS DECIMAL), 2) AS year\_1990,  
 *ROUND*(*CAST*(b.forest\_area\_sqkm AS DECIMAL), 2) AS year\_2016,  
 *ROUND*(*CAST*((a.forest\_area\_sqkm - b.forest\_area\_sqkm) AS DECIMAL), 2) AS difference,  
 *ROUND*(*CAST*((((a.forest\_area\_sqkm - b.forest\_area\_sqkm) / a.forest\_area\_sqkm) \* 100) AS DECIMAL), 2) AS percentage  
FROM table\_1990 a  
JOIN table\_2016 b ON a.country\_code = b.country\_code;  
  
SELECT \*,  
 *ROUND*(*CAST*((total\_area\_sq\_mi \* 2.58999) AS DECIMAL), 2) AS total\_area\_sq\_km  
FROM land\_area  
WHERE year = 2016 AND (total\_area\_sq\_mi \* 2.58999) < 1324449  
ORDER BY year, total\_area\_sq\_mi DESC;  
  
------------------------------------------------------------------------------------------------------  
  
-- Regional Outlook  
-- Part 1  
WITH data AS (SELECT  
 r.region,  
 fa.country\_code,  
 fa.country\_name,  
 fa.year,  
 fa.forest\_area\_sqkm AS forest\_area,  
 la.total\_area\_sq\_km AS land\_area,  
 (la.total\_area\_sq\_km - fa.forest\_area\_sqkm ) AS difference,  
 ((fa.forest\_area\_sqkm / la.total\_area\_sq\_km) \* 100) AS forest\_percentage\_in\_country  
FROM forest\_area fa  
INNER JOIN (  
 SELECT country\_code, country\_name, year, (total\_area\_sq\_mi \* 2.58999) AS total\_area\_sq\_km  
 FROM land\_area) la  
 ON fa.country\_code = la.country\_code AND fa.year = la.year  
INNER JOIN regions r  
 ON fa.country\_code = r.country\_code  
WHERE fa.year IN (1990, 2016) AND la.year IN (1990, 2016)  
ORDER BY fa.country\_code)  
  
SELECT region,  
 year ,  
 *ROUND*(*CAST*((*SUM*(forest\_area) / *SUM*(land\_area)) \* 100 AS DECIMAL), 2) AS percentage\_per\_region  
FROM data  
GROUP BY 1, 2  
ORDER BY 3 DESC;  
  
-- Part 2  
WITH data AS (SELECT  
 r.region,  
 fa.country\_code,  
 fa.country\_name,  
 fa.year,  
 fa.forest\_area\_sqkm AS forest\_area,  
 la.total\_area\_sq\_km AS land\_area,  
 (la.total\_area\_sq\_km - fa.forest\_area\_sqkm ) AS difference,  
 ((fa.forest\_area\_sqkm / la.total\_area\_sq\_km) \* 100) AS forest\_percentage\_in\_country  
FROM forest\_area fa  
INNER JOIN (  
 SELECT country\_code, country\_name, year, (total\_area\_sq\_mi \* 2.58999) AS total\_area\_sq\_km  
 FROM land\_area) la  
 ON fa.country\_code = la.country\_code AND fa.year = la.year  
INNER JOIN regions r  
 ON fa.country\_code = r.country\_code  
WHERE fa.year IN (1990, 2016) AND la.year IN (1990, 2016)  
ORDER BY fa.country\_code),  
  
percentage\_1990 AS (SELECT \* FROM data WHERE year = 1990),  
percentage\_2016 AS (SELECT \* FROM data WHERE year = 2016)  
  
SELECT a.region,  
 *ROUND*(*CAST*((*SUM*(a.forest\_area) / *SUM*(a.land\_area)) \* 100 AS DECIMAL), 2) AS percentage\_per\_regionin\_1990,  
 *ROUND*(*CAST*((*SUM*(b.forest\_area) / *SUM*(b.land\_area)) \* 100 AS DECIMAL), 2) AS percentage\_per\_regionin\_2016,  
 CASE  
 WHEN ((*SUM*(b.forest\_area) / *SUM*(b.land\_area)) \* 100) < ((*SUM*(a.forest\_area) / *SUM*(a.land\_area)) \* 100) THEN 'yes'  
 ELSE 'no'  
 END AS decreased  
FROM percentage\_1990 a  
JOIN percentage\_2016 b  
ON a.country\_code = b.country\_code  
GROUP BY 1  
ORDER BY 4 DESC;  
  
-----------------------------------------------------------------------------------------------  
  
-- Country Level Detail  
-- Part 1  
WITH data AS (SELECT  
 r.region,  
 fa.country\_code,  
 fa.country\_name,  
 fa.year,  
 fa.forest\_area\_sqkm AS forest\_area,  
 la.total\_area\_sq\_km AS land\_area,  
 (la.total\_area\_sq\_km - fa.forest\_area\_sqkm ) AS difference,  
 ((fa.forest\_area\_sqkm / la.total\_area\_sq\_km) \* 100) AS forest\_percentage\_in\_country  
FROM forest\_area fa  
INNER JOIN (  
 SELECT country\_code, country\_name, year, (total\_area\_sq\_mi \* 2.58999) AS total\_area\_sq\_km  
 FROM land\_area) la  
 ON fa.country\_code = la.country\_code AND fa.year = la.year  
INNER JOIN regions r  
 ON fa.country\_code = r.country\_code  
WHERE fa.year IN (1990, 2016) AND la.year IN (1990, 2016)  
ORDER BY fa.country\_code),  
  
 data\_1990 AS (SELECT \* FROM data WHERE year = 1990 AND forest\_area IS NOT NULL),  
 data\_2016 AS (SELECT \* FROM data WHERE year = 2016 AND forest\_area IS NOT NULL)  
  
SELECT a.country\_name,  
 a.region,  
 b.year,  
 CASE WHEN b.forest\_area > a.forest\_area THEN 'yes' ELSE 'no' END AS increased,  
 *ROUND*(*CAST*(*COALESCE*(a.forest\_area, 0) AS DECIMAL), 2) AS forest\_area\_in\_1990,  
 *ROUND*(*CAST*(*COALESCE*(b.forest\_area, 0) AS DECIMAL), 2) AS forest\_area\_in\_2016,  
 *ROUND*(*CAST*(*ABS*(a.forest\_area - b.forest\_area) AS DECIMAL), 2) AS difference,  
 *ROUND*(*CAST*(*ABS*((b.forest\_area - a.forest\_area) / a.forest\_area \* 100) AS DECIMAL), 2) AS percentage  
FROM data\_1990 a  
JOIN data\_2016 b  
ON a.country\_code = b.country\_code  
WHERE b.forest\_area > a.forest\_area  
ORDER BY *ABS*(a.forest\_area - b.forest\_area) DESC;  
  
-- Question a  
WITH data AS (SELECT  
 r.region,  
 fa.country\_code,  
 fa.country\_name,  
 fa.year,  
 fa.forest\_area\_sqkm AS forest\_area,  
 la.total\_area\_sq\_km AS land\_area,  
 (la.total\_area\_sq\_km - fa.forest\_area\_sqkm ) AS difference,  
 ((fa.forest\_area\_sqkm / la.total\_area\_sq\_km) \* 100) AS forest\_percentage\_in\_country  
FROM forest\_area fa  
INNER JOIN (  
 SELECT country\_code, country\_name, year, (total\_area\_sq\_mi \* 2.58999) AS total\_area\_sq\_km  
 FROM land\_area) la  
 ON fa.country\_code = la.country\_code AND fa.year = la.year  
INNER JOIN regions r  
 ON fa.country\_code = r.country\_code  
WHERE fa.year IN (1990, 2016) AND la.year IN (1990, 2016)  
ORDER BY fa.country\_code),  
  
 data\_1990 AS (SELECT \* FROM data WHERE year = 1990 AND forest\_area IS NOT NULL),  
 data\_2016 AS (SELECT \* FROM data WHERE year = 2016 AND forest\_area IS NOT NULL)  
  
SELECT a.country\_name,  
 a.region,  
 b.year,  
 CASE WHEN b.forest\_area > a.forest\_area THEN 'yes' ELSE 'no' END AS increased,  
 *ROUND*(*CAST*(*COALESCE*(a.forest\_area, 0) AS DECIMAL), 2) AS forest\_area\_in\_1990,  
 *ROUND*(*CAST*(*COALESCE*(b.forest\_area, 0) AS DECIMAL), 2) AS forest\_area\_in\_2016,  
 *ROUND*(*CAST*((a.forest\_area - b.forest\_area) AS DECIMAL), 2) AS difference,  
 *ROUND*(*CAST*(*ABS*((b.forest\_area - a.forest\_area) / a.forest\_area \* 100) AS DECIMAL), 2) AS percentage  
FROM data\_1990 a  
JOIN data\_2016 b  
ON a.country\_code = b.country\_code  
WHERE a.country\_name != 'World' and b.country\_name != 'World'  
ORDER BY 7 DESC  
LIMIT 5;  
  
-- Question b  
WITH data AS (SELECT  
 r.region,  
 fa.country\_code,  
 fa.country\_name,  
 fa.year,  
 fa.forest\_area\_sqkm AS forest\_area,  
 la.total\_area\_sq\_km AS land\_area,  
 (la.total\_area\_sq\_km - fa.forest\_area\_sqkm ) AS difference,  
 ((fa.forest\_area\_sqkm / la.total\_area\_sq\_km) \* 100) AS forest\_percentage\_in\_country  
FROM forest\_area fa  
INNER JOIN (  
 SELECT country\_code, country\_name, year, (total\_area\_sq\_mi \* 2.58999) AS total\_area\_sq\_km  
 FROM land\_area) la  
 ON fa.country\_code = la.country\_code AND fa.year = la.year  
INNER JOIN regions r  
 ON fa.country\_code = r.country\_code  
WHERE fa.year IN (1990, 2016) AND la.year IN (1990, 2016)  
ORDER BY fa.country\_code),  
  
 data\_1990 AS (SELECT \* FROM data WHERE year = 1990 AND forest\_area IS NOT NULL),  
 data\_2016 AS (SELECT \* FROM data WHERE year = 2016 AND forest\_area IS NOT NULL)  
  
SELECT a.country\_name,  
 a.region,  
 b.year,  
 CASE WHEN b.forest\_area > a.forest\_area THEN 'yes' ELSE 'no' END AS increased,  
 *ROUND*(*CAST*(*COALESCE*(a.forest\_area, 0) AS DECIMAL), 2) AS forest\_area\_in\_1990,  
 *ROUND*(*CAST*(*COALESCE*(b.forest\_area, 0) AS DECIMAL), 2) AS forest\_area\_in\_2016,  
 *ROUND*(*CAST*((a.forest\_area - b.forest\_area) AS DECIMAL), 2) AS difference,  
 *ROUND*(*CAST*(*ABS*((b.forest\_area - a.forest\_area) / a.forest\_area \* 100) AS DECIMAL), 2) AS percentage  
FROM data\_1990 a  
JOIN data\_2016 b  
ON a.country\_code = b.country\_code  
-- I need values where 1990 data is greater than 2016  
WHERE a.forest\_area > b.forest\_area  
ORDER BY 8 DESC  
LIMIT 5;  
  
-- Question c  
-- Part 1  
WITH data AS (SELECT  
 r.region,  
 fa.country\_code,  
 fa.country\_name,  
 fa.year,  
 fa.forest\_area\_sqkm AS forest\_area,  
 la.total\_area\_sq\_km AS land\_area,  
 (la.total\_area\_sq\_km - fa.forest\_area\_sqkm ) AS difference,  
 ((fa.forest\_area\_sqkm / la.total\_area\_sq\_km) \* 100) AS forest\_percentage\_in\_country  
FROM forest\_area fa  
INNER JOIN (  
 SELECT country\_code, country\_name, year, (total\_area\_sq\_mi \* 2.58999) AS total\_area\_sq\_km  
 FROM land\_area) la  
 ON fa.country\_code = la.country\_code AND fa.year = la.year  
INNER JOIN regions r  
 ON fa.country\_code = r.country\_code  
WHERE fa.year IN (2016) AND la.year IN (2016)  
ORDER BY fa.country\_code),  
 data\_2016 AS (SELECT \* FROM data WHERE year = 2016 AND forest\_area IS NOT NULL AND land\_area IS NOT NULL)  
  
SELECT  
 CASE  
 WHEN pt.forest\_percentage\_in\_country <= 25 THEN '0-25%'  
 WHEN pt.forest\_percentage\_in\_country <= 50 THEN '25-50%'  
 WHEN pt.forest\_percentage\_in\_country <= 75 THEN '50-75%'  
 WHEN pt.forest\_percentage\_in\_country <= 100 THEN '75-100%'  
 ELSE 'N/A'  
 END AS quartiles,  
 *COUNT*(country\_name) as countries  
FROM data\_2016 pt  
WHERE country\_name != 'World'  
GROUP BY 1  
ORDER BY 1;

-- Part 2  
WITH data AS (SELECT  
 r.region,  
 fa.country\_code,  
 fa.country\_name,  
 fa.year,  
 fa.forest\_area\_sqkm AS forest\_area,  
 la.total\_area\_sq\_km AS land\_area,  
 (la.total\_area\_sq\_km - fa.forest\_area\_sqkm ) AS difference,  
 ((fa.forest\_area\_sqkm / la.total\_area\_sq\_km) \* 100) AS forest\_percentage\_in\_country  
FROM forest\_area fa  
INNER JOIN (  
 SELECT country\_code, country\_name, year, (total\_area\_sq\_mi \* 2.58999) AS total\_area\_sq\_km  
 FROM land\_area) la  
 ON fa.country\_code = la.country\_code AND fa.year = la.year  
INNER JOIN regions r  
 ON fa.country\_code = r.country\_code  
WHERE fa.year IN (2016) AND la.year IN (2016)  
ORDER BY fa.country\_code),  
 data\_2016 AS (SELECT \* FROM data WHERE year = 2016 AND forest\_area IS NOT NULL AND land\_area IS NOT NULL)  
  
SELECT pt.country\_name,  
 pt.region,  
 *ROUND*(*CAST*(pt.forest\_percentage\_in\_country AS DECIMAL), 2)  
FROM data\_2016 pt  
WHERE *ROUND*(*CAST*(pt.forest\_percentage\_in\_country AS DECIMAL), 2) > 75  
AND pt.country\_name != 'World';