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

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 527229.1 sq km. 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 79200 sq km, 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. 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 5 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 (sq km), 1990 & 2016:

Country	Region	Absolute Forest Area Change
Brazil	Latin America & Caribbean	541510
Indonesia	East Asia & Pacific	282193.98
Myanmar	East Asia & Pacific	107234
Nigeria	Sub-Saharan Africa	106506
Tanzania	Sub-Saharan Africa	102320

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 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
1st	85
2nd	72
3rd	38
4th	9

The largest number of countries in 2016 were found in the first quartile (0-25% of forest area).

There were nine countries in the top quartile in 2016 (>75%). 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
Seychelles	Sub-Saharan Africa	88.41
Palau	East Asia & Pacific	87.61
American Samoa	East Asia & Pacific	87.5
Guyana	Latin America & Caribbean	83.9
Lao PDR	East Asia & Pacific	82.11
Solomon Islands	East Asia & Pacific	77.86

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?*

The World Bank data show that the percent of the total land area of the world designated as forest has decreased from 32.42% to 31.38% from 1990 to 2016. More in particular, the decrease has been identified in the regions of Latin America & Caribbean and Sub-Sahara Africa.

Campaign and attention should be focused therefore on those two regions and in particular on the countries of Nigeria, Togo, Uganda, Mauritania and Honduras, followed by Brazil and Tanziania, if we consider also the loss in area of sq km. Indonesia and Myanamar, in the East Asia & Pacific region, have also seen a significant reduction in forest area.

China and Iceland are instead examples of significant increase in forestation area from 1990 to 2016 (the first in sk qm area, the second in percentage). Both countries should be investigated to understand the basis of such success stories and to apply the learnings to other parts of the World.

5. APPENDIX: SQL QUERIES USED

0.   CREATE A VIEW

To create a view, I decide to use a JOIN (or INNER JOIN) between the three tables. This will cause the loss of the row with country TWN, TAIWAN, CHINA present only in the regions table. For that country, there are no available data regarding forest and land area and therefore it will not be of any use for answering the questions of the project, unless I find relevant information elsewhere (beyond the scope of this project).

```
CREATE VIEW forestation
AS
SELECT f.country_code,
    f.country_name,
    r.region,
    r.income_group,
    f.year,
    f.forest_area_sqkm,
    l.total_area_sq_mi,
    (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 l.country_code=r.country_code;
```

I can than run the code below to see the forestation table:

```
SELECT *
FROM forestation;
```

This table has 8 columns and 5886 rows. From my previous exploration of the different tables, I saw that there are some NULLs for columns forest_area_sqkm and total_area_sq_mi, which cause NULLs also in the resulting percent_forest_area column. If I run for example the code below, I can see that there are 352 rows with NULLs in the derived column. These are missing data (it would not be correct to assume that those Country have 0 skqm of forest). The dealing of such missing data is beyond the scope of this project.

```
SELECT *
FROM forestation
WHERE percent_forest_area IS NULL;
```

1a. total forest area (in sqkm) of the World in 1990:

SELECT forest_area_sqkm
FROM forestation
WHERE country_name='World' AND year=1990;

Output: A single cell with value of 41282694.9

1b. total forest area (in sqkm) of the World in 2016:

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

Output: A single cell with value of 39958245.9

1c. change (in sq km) in the forest area of the World from 1990 to 2016:

I answered this question with two different methods: one using function Lag and one with joining 1990 table with 2016 table.

Method 1:
I build a table with values of forest area for the World for both 1990 and 2016 (in same column).
I can then use LAG to calculate the difference between values originally in different rows.
WITH table1 AS (
      SELECT country_name,
          year,
          forest_area_sqkm,
          forest_area_sqkm-(LAG(forest_area_sqkm) OVER(ORDER BY year)) AS
          forest_sqkm_2016_minus_1990
          FROM forestation
          WHERE country_name='World' AND (year=1990 OR year=2016)
          ORDER by year)
SELECT forest_sqkm_2016_minus_1990
FROM table1
WHERE year=2016;

Output: A single cell gives with value of -1324449 sq km.

Method 2:
Instead of using LAG I can create two different tables, one for year 1990 and one for year 2016 and join them on country_name, in order to calculate the difference between the forest_area_sqkm_2016 and the forest_area_sqkm_1990 columns.

```
WITH table1 AS (
        SELECT country_name,
        year,
          forest_area_sqkm AS forest_area_sqkm_1990
        FROM forestation
        WHERE country_name='World' AND year=1990),
        table2 AS (
        SELECT country_name,
        year,
          forest_area_sqkm AS forest_area_sqkm_2016
        FROM forestation
        WHERE country_name='World' AND year=2016)
SELECT forest_area_sqkm_2016-forest_area_sqkm_1990 AS
forest_sqkm_2016_minus_1990
FROM table1 t1
JOIN table2 t2
ON t1.country_name=t2.country_name;
```

The resulting cell gives again a value of: -1324449 sq km.

1d. percent change in the forest area of the World from 1990 to 2016:

The percentage change is calculated as ((newvalue-oldvalue)/oldvalue)*100

Method 1 (using LAG function):

```
WITH table1 AS (
        SELECT country_name,
        year,
        forest_area_sqkm,
        (LAG(forest_area_sqkm) OVER(ORDER BY year)) AS lag,
        ((forest_area_sqkm-(LAG(forest_area_sqkm) OVER(ORDER BY
        year)))/(LAG(forest_area_sqkm) OVER(ORDER BY year)))*100 AS
        perc_change_forest_1990_2016
        FROM forestation
        WHERE country_name='World' AND (year=1990 OR year=2016)
        ORDER by year)
SELECT perc_change_forest_1990_2016
FROM table1
WHERE year=2016;
```

Output: A single cell with the value of: -3.20824258980244

Method 2:

I create two different tables, one for year 1990 and one for year 2016 and join them on country_name, in order to calculate percent change between the forest_area_sqkm_2016 and the forest_area_sqkm_1990 columns.

WITH table1 AS (
      SELECT country_name,
      year,
      forest_area_sqkm AS forest_area_sqkm_1990
      FROM forestation
      WHERE country_name='World' AND year=1990),
      table2 AS (
      SELECT country_name,
      year,
      forest_area_sqkm AS forest_area_sqkm_2016
      FROM forestation
      WHERE country_name='World' AND year=2016)
SELECT ((forest_area_sqkm_2016-forest_area_sqkm_1990)/forest_area_sqkm_1990)*100
AS perc_change_forest_1990_2016
FROM table1 t1
JOIN table2 t2
ON t1.country_name=t2.country_name;

Output: a single cell with the value of -3.20824258980244

1e. loss of forest area from 1990 and 2016 is closest to total area of country….:
From previous queries, the forest loss from 1990 to 2016 is equal to 1324449 sq km.

SELECT country_name,
      (total_area_sq_mi*2.59) AS total_area_sqkm
FROM forestation
WHERE (total_area_sq_mi*2.59)<=1324449 AND year=2016
ORDER BY total_area_sqkm DESC
LIMIT 1;

Output: country_name Peru, total_area_sqkm 1279999.9891

2 create a table that shows regions and their precent forest area in 1990 and 2016.

To obtain a table with a single column with year either 1990 or 2016, I run the following code:

```
SELECT region,
       year,
       ROUND(CAST(
       ((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59))*100) AS numeric),2) AS
       regional_percent_forest
FROM forestation
GROUP BY region, year
HAVING year=1990 OR year=2016
ORDER BY 1, 2;
```

The resulting output has 16 rows and three columns: region, year (1990 or 2016), regional_percent_forest (rounded to 2 decimal places).

This code is just slightly modified to answer the various questions in section 2 and to fill in table 2.1.

2a. percent forest of the World in 2016. Region with highest and region with the lowest % forest in 2016.

```
WITH table1 AS (
       SELECT region,
       year,
       ROUND(CAST(
       ((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59))*100) AS numeric),2) AS
       regional_percent_forest
       FROM forestation
       GROUP BY region, year
       HAVING year=1990 OR year=2016
       ORDER BY 1, 2)
SELECT *
FROM table1
WHERE year=2016 AND region='World';
```

Output is a row: World, 2016, 31.38

To find the region with highest % in 2016:

```
WITH table1 AS (
        SELECT region,
        year,
        ROUND(CAST(
        ((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59))*100) AS numeric),2) AS
        regional_percent_forest
        FROM forestation
        GROUP BY region, year
        HAVING year=1990 OR year=2016
        ORDER BY 1, 2)
SELECT *
FROM table1
WHERE year=2016 AND region!='World'
ORDER BY regional_percent_forest DESC
LIMIT 1;
```

Output a single row: Latin America & Caribbean, 2016, 46.16

To find the region with lowest % in 2016:

```
WITH table1 AS (
        SELECT region,
        year,
        ROUND(CAST(
        ((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59))*100) AS numeric),2) AS
        regional_percent_forest
        FROM forestation
        GROUP BY region, year
        HAVING year=1990 OR year=2016
        ORDER BY 1, 2)
SELECT *
FROM table1
WHERE year=2016 AND region!='World'
ORDER BY regional_percent_forest
LIMIT 1;
```

Output a single row: Middle East & North Africa, 2016, 2.07

2b. percent forest of the World in 1990. Region with highest and region with the lowest % forest in 1990.

Still just modifying slightly the same code, I can run the following:

```
WITH table1 AS (
        SELECT region,
        year,
        ROUND(CAST(
        ((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59))*100) AS numeric),2) AS
        regional_percent_forest
        FROM forestation
        GROUP BY region, year
        HAVING year=1990 OR year=2016
        ORDER BY 1, 2)
SELECT *
FROM table1
WHERE year=1990 AND region='World';
```

Output is a row: World, 1990, 32.42

To find the region with highest % in 1990:
```
WITH table1 AS (
        SELECT region,
        year,
        ROUND(CAST(
        ((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59))*100) AS numeric),2) AS
        regional_percent_forest
        FROM forestation
        GROUP BY region, year
        HAVING year=1990 OR year=2016
        ORDER BY 1, 2)
SELECT *
FROM table1
WHERE year=1990 AND region!='World'
ORDER BY regional_percent_forest DESC
LIMIT 1;
```

Output is a row: Latin America & Caribbean, 1990, 51.03

To find the region with lowest % in 2016:

```
WITH table1 AS (
        SELECT region,
        year,
        ROUND(CAST(
        ((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59))*100) AS numeric),2) AS
        regional_percent_forest
        FROM forestation
        GROUP BY region, year
        HAVING year=1990 OR year=2016
        ORDER BY 1, 2)
SELECT *
FROM table1
WHERE year=1990 AND region!='World'
ORDER BY regional_percent_forest
LIMIT 1;
```

Output is a row: Middle East & North Africa, 1990, 1.78

Filling in the table 2.1.

To fill in the table 2.1, (Region, 1990 Forest percent, 2016 Forest Percent), I join the regions table with the values of 1990 percentages with the regions table with values of 2016 percentages, instead of modifying the same code as before.

```
WITH table1 AS (
        SELECT region,
        year,
        ROUND(CAST(
        ((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59))*100) AS numeric),2) AS
        regional_percent_forest_1990
        FROM forestation
        GROUP BY region, year
        HAVING year=1990 AND region!='World'),
        table2 AS (
        SELECT region,
        year,
        ROUND(CAST(
        ((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59))*100) AS numeric),2) AS
        regional_percent_forest_2016
        FROM forestation
        GROUP BY region, year
        HAVING year=2016 AND region!='World')
SELECT t1.region,
        t1.regional_percent_forest_1990,
```

t2.regional_percent_forest_2016
FROM table1 t1
JOIN table2 t2
ON t1.region=t2.region
ORDER BY t1.region;

This provides the following output table, which can be used to fill in table 2.1 of the report:

region	regional_percent_forest_1990	regional_percent_forest_2016
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

2c. which regions of the World DECREASED in forest area from 1990 to 2016.

To answer this question, I can add a derived column calculating the difference %forest in 2016 minus % forest in 1990 to the table done to fill in table2.1. Then I can just check when such difference is negative (=decrease in forest %)

WITH table1 AS (
SELECT region,
year,
ROUND(CAST(
((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59))*100) AS numeric),2) AS
regional_percent_forest_1990
FROM forestation
GROUP BY region, year
HAVING year=1990 AND region!='World'),
table2 AS (
SELECT region,
year,
ROUND(CAST(
((SUM(forest_area_sqkm)/(SUM(total_area_sq_mi)*2.59))*100) AS
numeric),2) AS regional_percent_forest_2016
FROM forestation
GROUP BY region, year
HAVING year=2016 AND region!='World')
SELECT t1.region,
t1.regional_percent_forest_1990,
t2.regional_percent_forest_2016,
t2.regional_percent_forest_2016-t1.regional_percent_forest_1990 AS
perc_diff_2016_minus_1990
FROM table1 t1

JOIN table2 t2
ON t1.region=t2.region
WHERE t2.regional_percent_forest_2016-t1.regional_percent_forest_1990 <0
ORDER BY t1.region;

This gives an output with two rows with region, percent in 1990, percent in 2016 and the difference (negative value meaning decrease).

region	regional_percent_forest_1990	regional_percent_forest_2016	perc_diff_2016_minus_1990
Latin America & Caribbean	51.03	46.16	-4.87
Sub-Saharan Africa	30.67	28.79	-1.88

SUCCESS STORIES
The report requires the 2 countries with largest increase in forest area from 1990 to 2016 (increase amount in sqkm).

WITH table1 AS (
    SELECT country_name,
    year,
    forest_area_sqkm AS forest_area_sqkm_1990
    FROM forestation
    WHERE year=1990 AND country_name!='World'),
    table2 AS (
    SELECT country_name,
    year,
    forest_area_sqkm AS forest_area_sqkm_2016
    FROM forestation
    WHERE year=2016 AND country_name!='World')
SELECT t1.country_name,
    forest_area_sqkm_1990,
    forest_area_sqkm_2016,
    forest_area_sqkm_2016-forest_area_sqkm_1990 AS
    diff_forest_sqkm_2016_minus_1990
FROM table1 t1
JOIN table2 t2
ON t1.country_name=t2.country_name
WHERE forest_area_sqkm_2016-forest_area_sqkm_1990>0
ORDER BY diff_forest_sqkm_2016_minus_1990 DESC
LIMIT 2;

This code provides the following outcome:

country_name	forest_area_sqkm_1990	forest_area_sqkm_2016	diff_forest_sqkm_2016_minus_1990
China	1571406	2098635	527229.1
United States	3024500	3103700	79200

The report requests the top country with largest increase in forest area from 1990 to 2016 in % change (not in sqkm).

```
WITH table1 AS (
      SELECT country_name,
      year,
      forest_area_sqkm AS forest_area_sqkm_1990
      FROM forestation
      WHERE year=1990 AND country_name!='World'),
      table2 AS (
      SELECT country_name,
      year,
      forest_area_sqkm AS forest_area_sqkm_2016
      FROM forestation
      WHERE year=2016 AND country_name!='World')
SELECT t1.country_name,
      ROUND(CAST(((forest_area_sqkm_2016-
      forest_area_sqkm_1990)/forest_area_sqkm_1990)*100 AS numeric),2) AS
      perc_change_forest_2016_1990
FROM table1 t1
JOIN table2 t2
ON t1.country_name=t2.country_name
WHERE ((forest_area_sqkm_2016-
forest_area_sqkm_1990)/forest_area_sqkm_1990)*100>0
ORDER BY perc_change_forest_2016_1990 DESC
LIMIT 1;
```

Output: Iceland, 213.66

3A. Five countries with largest amount decrease in forest area from 1990 to 2016; difference in amount for each of those countries.
Again I can answer this with two methods: LAG function (to compare between rows) or joining 1990 and 2016 tables, to compare between columns.

Method 1 (LAG function):
```
WITH table1 AS (
      SELECT country_name,
      region,
      year,
       forest_area_sqkm,
      forest_area_sqkm-LAG(forest_area_sqkm) OVER(PARTITION BY
      country_name ORDER BY year) AS diff_forest_sqkm_2016_minus_1990
```

```
        FROM forestation
        WHERE (year=1990 OR year=2016) AND country_name!='World'
        ORDER BY country_name, year)
SELECT country_name,
        region,
        ROUND(CAST(ABS(diff_forest_sqkm_2016_minus_1990) AS numeric),2) AS
        abs_forest_area_change_sqkm
FROM table1
WHERE year=2016 AND diff_forest_sqkm_2016_minus_1990<0
ORDER BY abs_forest_area_change_sqkm DESC
LIMIT 5;
```

Method 2: joining 1990 and 2016 tables.
```
WITH table1 AS (
        SELECT country_name,
                region,
                year,
                forest_area_sqkm AS forest_area_sqkm_1990
        FROM forestation
        WHERE year=1990 AND country_name!='World'),
        table2 AS (
        SELECT country_name,
                region,
                year,
                forest_area_sqkm AS forest_area_sqkm_2016
        FROM forestation
        WHERE year=2016 AND country_name!='World')
SELECT t1.country_name,
        t1.region,
        ROUND(CAST(ABS(forest_area_sqkm_2016-forest_area_sqkm_1990) AS
        numeric),2) AS abs_forest_area_change_sqkmFROM table1 t1
JOIN table2 t2
ON t1.country_name=t2.country_name
WHERE forest_area_sqkm_2016-forest_area_sqkm_1990<0
ORDER BY abs_forest_area_change_sqkm DESC
LIMIT 5;
```

This code gives as output the following [solution for table 3.1 of the report]:

country_name	region	abs_forest_area_change_sqkm
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

3b. Five countries with largest % decrease in forest area from 1990 to 2016; percent change for each of those countries (2 decimal places).

```
WITH table1 AS (
        SELECT country_name,
                region,
                year,
                forest_area_sqkm AS forest_area_sqkm_1990
        FROM forestation
        WHERE year=1990 AND country_name!='World'),
    table2 AS (
        SELECT country_name,
                region,
                year,
                forest_area_sqkm AS forest_area_sqkm_2016
        FROM forestation
        WHERE year=2016 AND country_name!='World')
SELECT t1.country_name,
    t1.region,
    ROUND(CAST(((forest_area_sqkm_2016-
    forest_area_sqkm_1990)/forest_area_sqkm_1990)*100 AS numeric),2) AS
    perc_forest_area_change_2016_1990
FROM table1 t1
JOIN table2 t2
ON t1.country_name=t2.country_name
WHERE ROUND(CAST(((forest_area_sqkm_2016-
    forest_area_sqkm_1990)/forest_area_sqkm_1990)*100 AS numeric),2)<0
ORDER BY perc_forest_area_change_2016_1990
LIMIT 5;
```

output:

country_name	region	perc_forest_area_change_2016_1990
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

3c. which quartile or percent forestation had the most countries in 2016?

WITH table1 AS(
      SELECT country_name,
           region,
           forest_area_sqkm,
           percent_forest_area,
           CASE WHEN percent_forest_area<25 THEN '1st'
           WHEN percent_forest_area>=25 AND percent_forest_area<50 THEN '2nd'
           WHEN percent_forest_area>=50 AND percent_forest_area<=75 THEN '3rd'
           WHEN percent_forest_area>75 THEN '4th' END AS quartile
      FROM forestation
      WHERE year=2016 AND country_name!='World' AND percent_forest_area IS NOT
      NULL)
SELECT quartile,
      COUNT(*) AS number_of_countries
FROM table1
GROUP BY quartile
ORDER BY number_of_countries DESC;

Outcome (used to fill in table 3.3):

quartile	number_of_countries
1st	85
2nd	72
3rd	38
4th	9

I could add at the end of code **LIMIT 1,** to specifically pull out the group with the highest value (1st quartile).

3d. list countries in 4th quartile in 2016 (percent forest >75%)

WITH table1 AS(
      SELECT country_name,
           region,
           forest_area_sqkm,
           percent_forest_area,
           CASE WHEN percent_forest_area<25 THEN '1st'
                  WHEN percent_forest_area>=25 AND percent_forest_area<50
           THEN '2nd'
                  WHEN percent_forest_area>=50 AND percent_forest_area<=75
           THEN '3rd'
                  WHEN percent_forest_area>75 THEN '4th' END AS quartile
      FROM forestation

```
        WHERE year=2016 AND country_name!='World' AND percent_forest_area
        IS NOT NULL)
SELECT country_name,
        region,
        ROUND(CAST(percent_forest_area AS numeric),2) AS pct_designated_forest
FROM table1
WHERE quartile='4th'
ORDER BY pct_designated_forest DESC;
```

This code provides the information below (used to fill in table 3.4):

country_name	region	pct_designated_forest
Suriname	Latin America & Caribbean	98.26
Micronesia, Fed. Sts.	East Asia & Pacific	91.86
Gabon	Sub-Saharan Africa	90.04
Seychelles	Sub-Saharan Africa	88.41
Palau	East Asia & Pacific	87.61
American Samoa	East Asia & Pacific	87.5
Guyana	Latin America & Caribbean	83.9
Lao PDR	East Asia & Pacific	82.11
Solomon Islands	East Asia & Pacific	77.86

3e. How many countries had % forestation higher than USA in 2016?
The percent of forestation of USA is used to filter the 2016 table.

```
WITH table1 AS(
        SELECT *
        FROM forestation
        WHERE percent_forest_area>
                (SELECT percent_forest_area
                FROM forestation
                WHERE year=2016 AND country_name='United States') AND year=2016
        ORDER BY percent_forest_area)
 SELECT COUNT(*) AS num_countries
 FROM table1;
```

Output equal to 94. In 2016, there were 94 countries with a percent forestation higher than the percent forestation of the USA. This is not requested in the report.