Introduction

You're a data analyst for **ForestQuery**, a non-profit organization. on a mission to reduce deforestation around the world and which raises awareness about this important environmental topic.

Your executive director and her leadership team members are looking to understand which countries and regions around the world seem to have forests that have been shrinking in size, and also which countries and regions have the most significant forest area, both in terms of amount and percent of total area. The hope is that these findings can help inform initiatives, communications, and personnel allocation to achieve the largest impact with the precious few resources that the organization has at its disposal.

You've been able to find tables of data online dealing with forestation as well as total land area and region groupings, and you've brought these tables together into a database that you'd like to query to answer some of the most important questions in preparation for a meeting with the ForestQuery executive team coming up in a few days. Ahead of the meeting, you'd like to prepare and disseminate a report for the leadership team that uses complete sentences to help them understand the global deforestation overview between 1990 and 2016.

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:

   - **All of the columns of the origin tables**
   - 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).

NEXT

1. GLOBAL SITUATION

Instructions:

- Answering these questions will help you add information into the template.
- Use these questions as guides to write SQL queries.
- Use the output from the query to answer these questions.

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.

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."

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?

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?

NEXT

2. REGIONAL OUTLOOK

Instructions:

- Answering these questions will help you add information into the template.
- Use these questions as guides to write SQL queries.
- Use the output from the query to answer these questions.
- 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).
  Based on the table you created, ...

a. What was the percent forest of the entire world in 2016? Which region had the HIGHEST percent forest in 2016, and which had the LOWEST, to 2 decimal places?

b. What was the percent forest of the entire world in 1990? Which region had the HIGHEST percent forest in 1990, and which had the LOWEST, to 2 decimal places?

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

NEXT

3. COUNTRY-LEVEL DETAIL

Instructions:

- Answering these questions will help you add information into the template.
- Use these questions as guides to write SQL queries.
- Use the output from the query to answer these questions.

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?

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?

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

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

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

NEXT

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 1,280,000 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 and 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
World	32.42%	31.38%
Middle East & North Africa	1.78%	2.07%
South Asia	16.51%	17.51%
East Asia & Pacific	25.78%	26.36%
Sub-Saharan Africa	30.67%	28.79%
North America	35.65%	36.04%
Europe & Central Asia	37.28%	38.04%
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**

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 79,200.00 sqkm , 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 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 (sqkm)
Brazil	Latin America & Caribbean	541,510
Indonesia	East Asia & Pacific	282,194
Myanmar	East Asia & Pacific	107,234
Nigeria	Sub-Saharan Africa	106,506
Tanzania	Sub-Saharan Africa	102,320

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
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.

There are 94 countries, in 2016, with a percent forestation higher than the United States.

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	87.50
Guyana	Latin America & Caribbean	83.90
Lao PDR	East Asia & Pacific	82.11
Solomon Islands	East Asia & Pacific	77.86

5. RECOMMENDATIONS

The data from the World Bank suggests the regions with the highest percentage of the world's forest are Sub-Saharan Africa, Latin America & Caribbean, and East Asia & Pacific regions of the country contain most of the world's forests.

It's recommended to target the following countries:
1. **Nigeria**
2. **Togo**
3. **Uganda**
4. **Mauritania**
5. **Tanzania**
6. Honduras
7. Brazil
8. Indonesia
9. Myanmar

These countries had the highest amount of forested area loss or highest percentage loss. Nigeria was listed on both.

The first five countries in the list are from the Sub-Saharan Africa region and are the most concerning. Four out of five in the list had the highest percentage loss of all countries. If you look at the world map some of these countries border each other and some are a distance away. If the issue is not addressed soon the number of countries in Africa added to this list could grow. It's also interesting that these countries are not listed in the top quartile of 2016 percent forestation. A lack of response could trigger a bigger problem and the two Sub-Saharan countries (Gabon and Seychelles) in the top quartile could begin to face declines in forestation levels.

All countries in the list need a response, but it's recommended to first target the Sub-Saharan Africa countries, determine the root cause of the deforestation, and create a plan to address it.

Appendix: SQL Queries

Creation of 'forestation' View'

```sql
CREATE VIEW forestation AS
SELECT r.region,
       fa.country_name AS fa_country_name,
       la.country_name AS la_country_name,
        r.country_name AS r_country_name,
       fa.country_code AS fa_country_code,
       la.country_code AS la_country_code,
        r.country_code AS r_country_code,
       fa.year AS fa_year,
       la.year AS la_year,
       fa.forest_area_sqkm,
       la.total_area_sq_mi,
       r.income_group,
        (fa.forest_area_sqkm / 2.59) / la.total_area_sq_mi*100
            AS percent_forest_area
FROM forest_area fa
     INNER JOIN
    land_area la
    ON fa.country_code = la.country_code
       AND
        fa.year = la.year
      INNER JOIN
    regions r
    ON la.country_code = r.country_code;
```

Question 1: Global Situation

```sql
WITH t1990 AS
(
    SELECT COALESCE (fa_country_name, la_country_name, r_country_name) AS country,
           forest_area_sqkm AS f_area1990_sqkm
    FROM forestation
    WHERE COALESCE(la_year, fa_year) = 1990
          AND
          COALESCE (fa_country_name, la_country_name, r_country_name) = 'World'
),
t2016 AS
(
    SELECT COALESCE (fa_country_name, la_country_name, r_country_name) AS country,
           forest_area_sqkm AS f_area2016_sqkm
    FROM forestation
    WHERE COALESCE(la_year, fa_year) = 2016
          AND
          COALESCE (fa_country_name, la_country_name, r_country_name) = 'World'
),
t1 AS
(
    SELECT t1990.country,
           t1990.f_area1990_sqkm,
           t2016.f_area2016_sqkm,
           (t2016.f_area2016_sqkm-t1990.f_area1990_sqkm) AS change_sqkm,
           ROUND(((t2016.f_area2016_sqkm - t1990.f_area1990_sqkm)/t1990.f_area1990_sqkm*100)::numeric,2)
               AS pct_change
    FROM t1990
        INNER JOIN
        t2016
        ON t1990.country = t2016.country
),
t2 AS
(
    SELECT COALESCE (v.fa_country_name, v.la_country_name, v.r_country_name) AS closest_country_by_diff,
           ABS(ABS(t1.change_sqkm/2.59)-v.total_area_sq_mi) AS diff,
           ROUND((v.total_area_sq_mi*2.59)::numeric,0) AS land_area_sqkm
    FROM forestation AS v,
         t1
    WHERE COALESCE(v.fa_year, v.la_year) = 2016
          AND
          COALESCE(v.fa_country_name, v.la_country_name, v.r_country_name) != 'World'
    ORDER BY 2 ASC
    LIMIT 1
)
SELECT t1.*,
t2.closest_country_by_diff,
t2.land_area_sqkm
FROM t1, t2;
```

Question 2: Regional Outlook

```sql
WITH t1 AS
(
        SELECT  region,
                COALESCE (fa_country_name, la_country_name, r_country_name)
                AS country,
                COALESCE(la_year, fa_year) AS yr,
                (forest_area_sqkm/2.59) AS forest_area_sqmi,
                total_area_sq_mi
        FROM forestation
        WHERE COALESCE(la_year,fa_year) IN (1990,2016)
),
t1990 AS
(
        SELECT region,
                SUM(forest_area_sqmi) AS region_forest_sqmi1990,
                SUM(total_area_sq_mi) AS region_land_sqmi1990
        FROM t1
        WHERE yr = 1990
        GROUP BY region
),
t2016 AS
(
        SELECT region,
                SUM(forest_area_sqmi) AS region_forest_sqmi2016,
                SUM(total_area_sq_mi) AS region_land_sqmi2016
                FROM t1
        WHERE yr = 2016
        GROUP BY region
)
SELECT t1990.region,
        ROUND((region_forest_sqmi1990/region_land_sqmi1990*100)::numeric,2)
            AS   pct_forest_1990,
        ROUND((region_forest_sqmi2016/region_land_sqmi2016*100)::numeric,2)
            AS pct_forest_2016
FROM    t1990
        INNER JOIN
        t2016
        ON t1990.region = t2016.region;
```

Question 3: Country Level Detail
The last part of the query is highlighted in GREEN. This section was changed for results needed to answer questions through Table 3.2. The rest of the questions were answered by their own queries and are listed later.

```sql
WITH t1990 AS
(
    SELECT region,
           COALESCE (fa_country_name, la_country_name, r_country_name) AS country,
           forest_area_sqkm AS f_area1990_sqkm
    FROM forestation
    WHERE COALESCE(la_year, fa_year) = 1990 AND region != 'World'
),
t2016 AS
(
    SELECT region,
           COALESCE (fa_country_name, la_country_name, r_country_name) AS country,
           forest_area_sqkm AS f_area2016_sqkm,
           percent_forest_area AS pct_forest_area_2016
    FROM forestation
    WHERE COALESCE(la_year, fa_year) = 2016 AND region != 'World'
),
t1 AS
(
    SELECT t1990.region,
           t1990.country,
           t1990.f_area1990_sqkm,
           t2016.f_area2016_sqkm,
           ROUND((t2016.f_area2016_sqkm-t1990.f_area1990_sqkm)::numeric,2) AS change_sqkm,
           ROUND(((t2016.f_area2016_sqkm-t1990.f_area1990_sqkm)/t1990.f_area1990_sqkm*100)::numeric,2)
                AS pct_change,
           ROUND(t2016.pct_forest_area_2016::numeric,2) AS pct_forest_area_2016
    FROM   t1990
           INNER JOIN
           t2016
           ON t1990.country = t2016.country
              AND
              t1990.region = t2016.region
    ORDER BY 1 ASC
)
SELECT *
FROM t1
WHERE change_sqkm IS NOT NULL
ORDER BY change_sqkm DESC
LIMIT 2;
```

Replace each of these queries with the green section in the query above.

For 2nd paragraph

```sql
SELECT *
FROM t1
WHERE pct_change IS NOT NULL
ORDER BY pct_change DESC
LIMIT 1;
```

Part B Table 3.1

```sql
SELECT *
FROM t1
WHERE change_sqkm IS NOT NULL
ORDER BY change_sqkm ASC
LIMIT 5;
```

Part B Table 3.2

```sql
SELECT *
FROM t1
WHERE pct_change IS NOT NULL
ORDER BY pct_change ASC
LIMIT 5;
```

*** The remaining queries are to be run independent of queries above ***

Part C Table 3.3

```sql
WITH t1 AS
(
    SELECT region,
           COALESCE (fa_country_name, la_country_name, r_country_name) AS country,
           ROUND(percent_forest_area::numeric,2) AS pct_forest
    FROM forestation
    WHERE COALESCE(la_year, fa_year) = 2016 AND region != 'World'
          AND percent_forest_area IS NOT NULL
),
t2 AS
(
    SELECT country,
           region,
           pct_forest,
           CASE WHEN pct_forest > 75.00 THEN '4'
                WHEN pct_forest > 50.00 AND pct_forest <= 75.00 THEN '3'
                WHEN pct_forest > 25.00 AND pct_forest <= 50.00 THEN '2'
                ELSE '1' END AS quartile
    FROM t1
)
SELECT quartile, COUNT(quartile)
FROM t2
GROUP BY quartile
ORDER BY quartile ASC;
```

Part D Table 3.4

```
WITH t1 AS
(
    SELECT region,
           COALESCE (fa_country_name, la_country_name, r_country_name) AS country,
           ROUND(percent_forest_area::numeric,2) AS pct_forest
    FROM forestation
    WHERE COALESCE(la_year, fa_year) = 2016 AND region != 'World'
          AND percent_forest_area IS NOT NULL
),
t2 AS
(
  SELECT country,
         region,
         pct_forest,
         CASE WHEN pct_forest > 75.00 THEN '4'
              WHEN pct_forest > 50.00 AND pct_forest <= 75.00 THEN '3'
              WHEN pct_forest > 25.00 AND pct_forest <= 50.00 THEN '2'
              ELSE '1' END AS quartile
FROM t1
)
SELECT *
FROM t2
WHERE quartile = '4'
ORDER BY pct_forest DESC;
```

Finding number of countries with higher percent forestation in 2016 than United States

```
WITH t1 AS
(
    SELECT COALESCE (fa_country_name, la_country_name, r_country_name) AS country,
           percent_forest_area AS pct_forest2016
    FROM forestation
    WHERE COALESCE(la_year, fa_year) = 2016 AND region != 'World'
          AND percent_forest_area IS NOT NULL
)
SELECT COUNT(*) countries_higher_than_UnitedStates
FROM t1
WHERE pct_forest2016 > (
                        SELECT pct_forest2016
                        FROM t1
                        WHERE country = 'United States'
                       );
```

Suggestions for students

1 – Look beyond just the numbers for top countries in terms of forestation lost. Look at the other pieces of data captured in the project and see if there is a way to differentiate some of these countries from the rest in order to prioritize action.

2 – Research the topic on the internet to understand why deforestation is happening in places like Africa at an alarming rate. While I did not mention this in my analysis, I learned people in Africa are using inefficient farming methods to supply themselves with food. One tactic to slow the deforestation is to train people on more efficient methods of farming so they don't have to cut down more trees.