11.3 Final Project Step 1

Madeleine Sharp

2021-11-14

Introduction

As the developed world continues to become even more developed, the usage of our planet's natural resources becomes a critical issue to consider. Not only does development impact environmental resource consumption, but one would ascertain that so does population increase – with more inhabitants on planet earth, there are fewer resources per capita. In the United Nations IPCC 2021 report on climate change (IPCC, 2021), a strong call to action was made – primarily citing human influence as one of the primary drivers of environmental decline and thus proposing escalated control and reduction of both human resource usage and human carbon footprints as an imperative solution. While this report focused specifically on climate change and its influencing factors, I would like to explore the issue of the depletion and availability of natural resources and the change in those resources across time. Additionally, I believe it is critical to not only consider historical data in this case, but to predict the landscape (no pun intended) of natural resources in the coming years utilizing predictive modeling of the data. Not only is this a crucial issue to explore and to formulate solutions for as it pertains to the earth, but for every living person now and those to come. By utilizing historical data trends regarding natural resources, we can forecast for tomorrow, and implement plans and solutions today to mitigate further degradation of the environment.

So, What is the Problem?

Given the discussions surrounding climate change, natural resource decline, the degradation of our planet, population overload, and wealth/dvelopment impacting resource usage one would surmise that enough alarm bells have resounded regarding the health and well-being of our planet. Put more plainly, much stress has been placed upon our planet and its resources, and if control is not exercised to minimize any damage, those living on this planet in the generations to come may have a more difficult time living and the landscape of survival may change. The issue of planetary and environmental harm is not just an issue of environmental ethics, but of the ethics surrounding human life, too.

Addressing the Problem

One of the ways in which any problem can be addressed is via conducting some research in the form of data analysis. After all, in an endeavor to obtain any resolve or solution regarding the problem at hand, it is imperative to first truly understand the problem, which includes assessing available and pertinent data.

By reviewing and analyzing any related data, patterns and insights may become clearer as to why the problem is what it is, what its potential roots may be, what relationships exist that may lend credence to the problem, etc. Overall, analyzing the data related to a topic will allow one to glean meaningful insights for proceeding with a solution. After all, one cannot simply be sure of what may be occurring until light is shed upon the matter. In simpler terms, it is important to "trust but verify" - while a hypothesis may be on the table, we cannot assume the hypothesis to be true (even though we may perhaps trust it is an educated guess). We must work towards verifying any hypothesis with data.

In this specific case for my project, I will be attempting to address the problem by analyzing data that may help to answer the below questions:

1. Does a country's population impact its availability or decline of natural resources? More specifically:
   - What is the relationship between a country's total population and its renewable water resources available?
   - What is the relationship between a country's total population and its losses and usage of forest areas?
   - What is the relationship between a country's total population and its material consumption and material footprint?
   - What is the relationship between a country's total population and its total number of critically endangered species?
   - What is the relationship between a country's total population and its total consumption of petroleum and other related liquids?
   - What is the relationship between a country's total population and its percentage of protected terrestrial land?
2. Does a country's wealth impact its availability or decline of natural resources? More specifically:
   - What is the relationship between a country's wealth (GDP) and its renewable water resources available?
   - What is the relationship between a country's wealth (GDP) and its losses and usage of forest areas?
   - What is the relationship between a country's wealth (GDP) and its material consumption and material footprint?
   - What is the relationship between a country's wealth (GDP) and its total number of critically endangered species?
   - What is the relationship between a country's wealth (GDP) and its total consumption of petroleum and other related liquids?
   - What is the relationship between a country's wealth (GDP) and its percentage of protected terrestrial land?
3. Are any of the relationships between a country's wealth (GDP) and its population significant?
4. Which countries have experienced the greatest decline in natural resources including water, forests, natural materials, wildlife, petroleum, and protected land?

5. What do the natural resource changes look like across time?
6. Is there a correlation between those changes and the changes in a country's population or GDP?
7. Based on the historical data that is available related to these topics, can we make accurate predictions regarding the future of our planet's natural resources?

Analysis

Overall, for my analysis portion of the project, I will be focusing on running some specific analyses, including: * Descriptive Statistics - I will be running descriptive summary statistics on each of my original datasets to get an idea and a feel for what the data looks like and to understand what I am looking at from a measures of central tendency and measures of variability standpoint. * Correlational Analyses - I will be running correlations between all variables, specifically doing so in the following manner: * Using Country Population as an independent variable (IV) and assessing its relationship to all of the dependent variables (DVs) in this analysis, which includes renewable water, forest losses and usage, material consumption and footprint, critically endangered species, petroleum and other related liquids consumption, and protected terrestrial land. * Using Country GDP as an independent variable (IV) and assessing its relationship to all of the dependent variables (DVs) in this analysis, which includes renewable water, forest losses and usage, material consumption and footprint, critically endangered species, petroleum and other related l iquids consumption, and protected terrestrial land. * Time Series Plots and Analyses - For those variables that include data for more than a single year (some of my datasets only include data for one year or a most recent year), I will be plotting them across time by country. More specifically: * I will be plotting population as an IV with the countries present in each of the respective DVs (each dataset has different countries, so population will be plotted multiple times for each DV). I will also then plot each respective DV across time. * I will complete the above as well for a country's GDP as an IV.

These measures will allow me to: * Understand my datasets * Assess any relationships present between the variables in question * Assess the movement/direction of countries' population, GDP, and all DVs (for those applicable) across time

Importing Datasets and Cleaning the Data Part I

Water Resources Dataset

1. Organisation for Economic Co-operation and Development. (2021). Freshwater resources (long term annual average). OECD.Stat. https://stats.oecd.org/viewhtml.aspx?datasetcode=WATER_RESOURCES&lang=en

- This dataset includes data regarding freshwater resources by country - with an emphasis on renewable water.
- The data is last updated as of March 2021. The only available data is for the most recent year (2021).
- This data first needs to be exported to an Excel file in order to extract the data.

- On the OECD Stat site, I was able to customize to select and export only those variables for which I am interested in (which was quite helpful)! Once saved as an Excel file (.csv), I am able to import the data within R to be cleaned for use. I am selecting the variables of Total Renewable per Capita (m3/cap) and Total Renewable from this data set. * The Total Renewable per Capita variable is measured in cubic meters per capita. In other words, it measures a block of space that is 1m x 1m x 1m, height x width x depth, which is a highly accurate volume calculation. The figures represent a measurement in the billions (1 = 1 billion). * The Total Renewable variable is also measured in cubic meters. The figures represent a measurement in the billions (1 = 1 billion).
- To clean this data, I will ensure the format of the variable nomenclature schema is the same across the board, thereby being easily read and utilized in my analyses.

Water Dataset: Loaded and Cleaned

```
    Country Total.renewable.per.capita..m3.cap. Total.renewable
1 Australia                             16175.6           407.7
2   Austria                              9603.5            86.0
3   Belgium                              2232.8            25.8
4    Canada                             92967.2          3478.0
5     Chile                             53160.0          1007.5
6  Colombia                                  NA              NA

    country total_renewable_per_capita_m3_cap_2021 total_renewable_2021
1 Australia                                16175.6                407.7
2   Austria                                 9603.5                 86.0
3   Belgium                                 2232.8                 25.8
4    Canada                                92967.2               3478.0
5     Chile                                53160.0               1007.5
6  Colombia                                     NA                   NA
```

Forest Resources Datasets

2. Organisation for Economic Co-operation and Development. (2021). Depletion and growth of forest resources in terms of volume. OECD.Stat. https://stats.oecd.org/viewhtml.aspx?datasetcode=FOREST&lang=en

- This dataset includes data regarding forest resources by country - specifically in terms of depletion and growth.
- This dataset includes data pulled in for the years of 2010-2019 (2019 was the most recent year).
- This data first needs to be exported to an Excel file in order to extract the data.
- On the OECD Stat site, I was able to customize to select and export only those variables for which I am interested in. Once saved as an Excel file (.csv), the data can be reviewed and cleaned in R. I am selecting the variables of Natural Losses (in cubic meters, units are in the thousands in terms of volume) and Intensity of Use of Forest Resources (in ratio units, in terms of volume) across ten years (2010-2019) from this data set. From only looking at 2019 alone, not much data was present.
  - Each of these variables has been pulled into R in its own dataset/dataframe.

- To clean this data, I will ensure the format of the variable nomenclature schema is the same across the board, thereby being easily read and utilized in my analysis.
- This source has the data as two separate data sets - one for natural losses and the other for intensity of usage of forest resources.

Forest Datasets: Loaded and Cleaned

```
      Country X2010.00 X2011.00 X2012.00 X2013.00 X2014.00 X2015.00 X2016.00
1      Latvia  6090.00  6090.00  6090.00  6090.00  6090.00  6070.00  6070.00
2      Norway  3551.00  3585.00  3620.00  3657.00  3694.00  4188.00  4188.00
3     Germany  3697.83  3691.34  3684.46  3685.34  5158.98  5252.59  5064.86
4      Turkey       NA  3245.00  3073.00  4014.00  2567.00  5246.00  3324.00
5   Lithuania  3330.00  3420.00  3310.00  3260.00  3290.00  3220.00  3100.00
6 Switzerland  2648.90       NA       NA       NA       NA  2491.90       NA
  X2017.00 X2018.00 X2019.00
1     6070  5943.00       NA
2     4188  4182.00     4182
3     5023  3747.00       NA
4     2763  3273.00     2810
5     3120  3000.00     3020
6       NA  2395.25       NA

          Country X2010 X2011 X2012 X2013 X2014 X2015 X2016 X2017 X2018 X2019
1  Czech Republic  0.79  0.72  0.70  0.71  0.71  0.74  0.80  0.88  1.15  1.45
2         Estonia  0.60  0.66  0.73  0.71  0.70  0.70  0.75  0.87  0.90  0.80
3         Germany  0.87  0.80  0.81  0.75  0.76  0.78  0.75  0.74  0.88    NA
4 Slovak Republic  0.83  0.79  0.68  0.65  0.78  0.77  0.77  0.78  0.82  0.77
5      Luxembourg    NA  0.72  0.49  0.47  0.53  0.58  0.50  0.57  0.69  0.60
6     Switzerland  0.71    NA    NA    NA    NA  0.70    NA    NA  0.69    NA

      country losses_2010 losses_2011 losses_2012 losses_2013 losses_2014
1      Latvia     6090.00     6090.00     6090.00     6090.00     6090.00
2      Norway     3551.00     3585.00     3620.00     3657.00     3694.00
3     Germany     3697.83     3691.34     3684.46     3685.34     5158.98
4      Turkey          NA     3245.00     3073.00     4014.00     2567.00
5   Lithuania     3330.00     3420.00     3310.00     3260.00     3290.00
6 Switzerland     2648.90          NA          NA          NA          NA
  losses_2015 losses_2016 losses_2017 losses_2018 losses_2019
1     6070.00     6070.00        6070     5943.00          NA
2     4188.00     4188.00        4188     4182.00        4182
3     5252.59     5064.86        5023     3747.00          NA
4     5246.00     3324.00        2763     3273.00        2810
5     3220.00     3100.00        3120     3000.00        3020
6     2491.90          NA          NA     2395.25          NA

          country usage_2010 usage_2011 usage_2012 usage_2013 usage_2014
1  Czech Republic       0.79       0.72       0.70       0.71       0.71
2         Estonia       0.60       0.66       0.73       0.71       0.70
3         Germany       0.87       0.80       0.81       0.75       0.76
4 Slovak Republic       0.83       0.79       0.68       0.65       0.78
5      Luxembourg         NA       0.72       0.49       0.47       0.53
```

```
6     Switzerland       0.71          NA          NA          NA          NA
  usage_2015 usage_2016 usage_2017 usage_2018 usage_2019
1       0.74       0.80       0.88       1.15       1.45
2       0.70       0.75       0.87       0.90       0.80
3       0.78       0.75       0.74       0.88         NA
4       0.77       0.77       0.78       0.82       0.77
5       0.58       0.50       0.57       0.69       0.60
6       0.70         NA         NA       0.69         NA
```

Material Resources Data Sets

3.Organisation for Economic Co-operation and Development. (2021). Material resources. OECD.Stat.
https://stats.oecd.org/viewhtml.aspx?datasetcode=MATERIAL_RESOURCES&lang=en

- This dataset includes data regarding material resources by country - specifically in terms of usage and footprint.
- This dataset includes data pulled in for the years of 2010-2019 and for 2010-2017 (see below).
- This data first needs to be exported to an Excel file in order to extract the data.
- Once saved as an Excel file (.csv), the data can be extracted and imported into R to be cleaned and utilized for analysis. For this data set, I am selecting the variables of Domestic Material Consumption per Capita (kilograms per capita, in the thousands) across ten years (2010-2019) and Material Footprint per Capita (kilograms per capita, in the thousands) across seven years (2010-2017: the data only goes through 2017). From only looking at 2019 alone for material consumption, not much data was present. On the OECD Stat site, I was able to customize to select and export only those variables for which I am interested in.
  - Each of these variables has been pulled into R in its own dataset/dataframe.
- To clean this data, I will ensure the format of the variable nomenclature schema is the same across the board, thereby being easily read and utilized in my analysis.
- This source has the data as two separate data sets - one for material consumption (2010-2019) and the other for material footprint (2010-2017).

Material Datasets: Loaded and Cleaned

```
    Country   X2010   X2011   X2012   X2013   X2014   X2015   X2016   X2017
1 Australia 40.5952 40.7088 39.7431 39.5675 38.6466 38.4333 38.0788 37.7188
2   Austria 18.9726 19.9488 19.3716 18.8691 18.7661 18.2697 19.0816 18.7208
3   Belgium 14.8156 15.5685 14.2317 14.0548 13.7693 13.5473 13.5042 13.8141
4    Canada 28.4746 29.0813 28.3162 30.6897 29.3258 28.9513 28.8247 28.7080
5     Chile 39.1149 39.0211 39.7548 41.0883 40.4474 40.4297 40.4411 40.4049
6  Colombia  5.2268  5.5312  5.6416  5.5965  5.7755  5.8406  5.8721  5.8907
    X2018   X2019
1      NA      NA
2 18.7713 18.6614
3 13.3837 12.3200
4      NA      NA
```

```
5      NA      NA
6      NA      NA

    Country   X2010   X2011   X2012   X2013   X2014   X2015   X2016   X2017
1 Australia 40.7986 40.9456 41.5003 41.1921 42.1842 42.5502 42.8273 43.1128
2   Austria 30.0636 31.0990 30.9061 31.0024 31.5693 31.8904 32.2465 32.5868
3   Belgium 22.8657 22.2866 22.5309 22.8641 23.1293 23.4314 23.7751 24.1195
4    Canada 33.2096 33.5376 34.3866 35.0090 35.0283 34.8016 34.8148 34.8364
5     Chile 15.7977 16.2076 16.1666 17.1028 17.0253 17.0462 17.1084 17.1491
6  Colombia  9.0587  9.8797 10.3391 10.2774 10.4718 10.5186 10.6144 10.6902

    country consumption_2010 consumption_2011 consumption_2012 consumption_20
13
1 Australia          40.5952          40.7088          39.7431          39.56
75
2   Austria          18.9726          19.9488          19.3716          18.86
91
3   Belgium          14.8156          15.5685          14.2317          14.05
48
4    Canada          28.4746          29.0813          28.3162          30.68
97
5     Chile          39.1149          39.0211          39.7548          41.08
83
6  Colombia           5.2268           5.5312           5.6416           5.59
65
  consumption_2014 consumption_2015 consumption_2016 consumption_2017
1          38.6466          38.4333          38.0788          37.7188
2          18.7661          18.2697          19.0816          18.7208
3          13.7693          13.5473          13.5042          13.8141
4          29.3258          28.9513          28.8247          28.7080
5          40.4474          40.4297          40.4411          40.4049
6           5.7755           5.8406           5.8721           5.8907
  consumption_2018 consumption_2019
1               NA               NA
2          18.7713          18.6614
3          13.3837          12.3200
4               NA               NA
5               NA               NA
6               NA               NA

    country footprint_2010 footprint_2011 footprint_2012 footprint_2013
1 Australia        40.7986        40.9456        41.5003        41.1921
2   Austria        30.0636        31.0990        30.9061        31.0024
3   Belgium        22.8657        22.2866        22.5309        22.8641
4    Canada        33.2096        33.5376        34.3866        35.0090
5     Chile        15.7977        16.2076        16.1666        17.1028
6  Colombia         9.0587         9.8797        10.3391        10.2774
  footprint_2014 footprint_2015 footprint_2016 footprint_2017
1        42.1842        42.5502        42.8273        43.1128
2        31.5693        31.8904        32.2465        32.5868
```

```
3        23.1293        23.4314        23.7751        24.1195
4        35.0283        34.8016        34.8148        34.8364
5        17.0253        17.0462        17.1084        17.1491
6        10.4718        10.5186        10.6144        10.6902
```

Endangered Species Data Set

4. Organisation for Economic Co-operation and Development. (2021). Threatened species. OECD.Stat. https://stats.oecd.org/viewhtml.aspx?datasetcode=WILD_LIFE&lang=en

- This dataset includes data regarding threatened and endangered species by country - specifically divided out into animal classes.
- The data is last updated as of March 2021. The only available data is for the most recent year (2021).
- This data first needs to be exported to an Excel file in order to extract the data.
- Once saved as an Excel file (.csv), the data can be extracted into and cleaned in R. For this data set, I am selecting the variable of Number of Critically Endangered Species (total number of species). This data is for "to-date", so the number as of a country's current standing. On the OECD Stat site, I was able to customize to select and export only those variables for which I am interested in.
- To clean this data, I will ensure the format of the variable nomenclature schema is the same across the board, thereby being easily read and utilized in my analysis.

Critically Endangered Species Dataset: Loaded and Cleaned

```
    Country Mammals Birds Reptiles Amphibians Fish Marine.Fish Freshwater.Fis
h
1 Australia      10    17       10         15    7           4
3
2   Austria       4    14        3          1    6          NA
6
3   Belgium       4    20       NA          3    3           2
1
4    Canada      11    26        5          2    9           3
6
5     Chile       3     2        9         10    1          NA
1
6  Colombia       5    16       11         75    7           1
2
  Vascular.plants Mosses Lichens Invertebrates Total
1             198     NA      NA            30   294
2             172     34      57           302   599
3             253     NA      NA           135   421
4             315     75      70           205   727
5              74      1       1            33   135
6             112      6       0             6   241


Attaching package: 'dplyr'
```

```
The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union

    country mammals birds reptiles amphibians fish marine_fish freshwater_fis
h
1 Australia      10    17        10         15    7           4
3
2   Austria       4    14         3          1    6          NA
6
3   Belgium       4    20        NA          3    3           2
1
4    Canada      11    26         5          2    9           3
6
5     Chile       3     2         9         10    1          NA
1
6  Colombia       5    16        11         75    7           1
2
  vascular_plants mosses lichens invertabrates total
1             198     NA      NA            30   294
2             172     34      57           302   599
3             253     NA      NA           135   421
4             315     75      70           205   727
5              74      1       1            33   135
6             112      6       0             6   241
```

Petroleum Data Set

5. U.S. Energy Information Administration. (2021). Consumption: Petroleum and other liquids. EIA. https://www.eia.gov/international/data/world/petroleum-and-other-liquids/annual-refined-petroleum-products-consumption?pd=5&p=0000001001vg000000000000000000000000000000g&u=0&f=A&v=mapbubble&a=-&i=none&vo=value&&t=C&g=00000000000000000000000000000000000000000000000001&l=249-ruvvvvvfvtvnvv1vrvvvvfvvvvvvfvvvou20evvvvvvvvvvnvvvs0008&s=94694400000&e=1577836800000

- This dataset includes data regarding petroleum and other related liquid resources by country - with an emphasis on consumption.
- This dataset includes data pulled in for the years of 2010-2020.
- This data first needs to be exported to an Excel file in order to extract the data.
- On the OECD Stat site, I was able to customize to select and export only those variables for which I am interested in (which was quite helpful)! Once saved as an Excel file (.csv), I am able to import the data within R to be cleaned for use. I am

selecting the variable of Petroleum and Other Liquids Consumption (Mb/d - thousands of barrels per day).

- To clean this data, I will ensure the format of the variable nomenclature schema is the same across the board, thereby being easily read and utilized in my analyses.

Petroleum Dataset: Loaded and Cleaned

```
        Location    X2010    X2011    X2012   X2013    X2014    X2015    X201
6
1   Global/World 88433.47 89005.81 90300.19 91810.3 93354.19 95127.56 96557.9
7
2    Afghanistan    42.76    55.82    49.06   35.08    27.68    34.91    25.5
9
3        Albania    25.95    25.74    22.64   27.17    28.13    26.06    27.1
7
4        Algeria   347.41   349.58   373.79  400.34   415.61   431.27   420.7
2
5 American Samoa     4.14     2.35     2.35    2.35     2.35     2.35     2.3
5
6         Angola   106.95   111.97   119.81  138.74   158.87   156.75   150.6
1
     X2017   X2018 X2019 X2020
1 98890.83 99907.9  <NA>    NA
2    23.63   23.15          NA
3    35.33   33.05          NA
4   407.39  418.17          NA
5     2.35    2.34          NA
6   134.83  133.35          NA

         country pet_use_2010 pet_use_2011 pet_use_2012 pet_use_2013
1   Global/World     88433.47     89005.81     90300.19     91810.30
2    Afghanistan        42.76        55.82        49.06        35.08
3        Albania        25.95        25.74        22.64        27.17
4        Algeria       347.41       349.58       373.79       400.34
5 American Samoa         4.14         2.35         2.35         2.35
6         Angola       106.95       111.97       119.81       138.74
  pet_use_2014 pet_use_2015 pet_use_2016 pet_use_2017 pet_use_2018 pet_use_20
19
1     93354.19     95127.56     96557.97     98890.83     99907.90
NA
2        27.68        34.91        25.59        23.63        23.15
NA
3        28.13        26.06        27.17        35.33        33.05
NA
4       415.61       431.27       420.72       407.39       418.17
NA
5         2.35         2.35         2.35         2.35         2.34
NA
6       158.87       156.75       150.61       134.83       133.35
NA
  pet_use_2020
```

```
1           NA
2           NA
3           NA
4           NA
5           NA
6           NA
```

Protected Land Data Set

6. The World Bank. (2021). Terrestrial Protected Areas (% of Total Land Area). The World Bank. https://data.worldbank.org/indicator/ER.LND.PTLD.ZS

- This dataset includes data regarding protected land by country.
- This dataset includes data pulled in for the years of 2010-2020.
- This data first needs to be exported to an Excel file in order to extract the data.
- The only variable for this data set is protected land (as a percentage of total land area) by country across time. I will be looking at the years of 2010-2020 for population totals.
- To clean this data, I will ensure the format of the variable nomenclature schema is the same across the board, thereby being easily read and utilized in my analysis.

Protected Land Dataset: Loaded and Cleaned

```
                 Country.Name    X2016     X2017     X2018
1                       Aruba  0.53000 18.917576 18.917576
2 Africa Eastern and Southern 16.49366 16.907344 16.907346
3                 Afghanistan  0.10000  0.104707  0.104707
4  Africa Western and Central 15.49113 15.725047 15.725047
5                      Angola  6.97000  6.971427  6.971427
6                     Albania 17.21000 17.736095 17.736095

                      country land_2016 land_2017 land_2018
1                       Aruba   0.53000 18.917576 18.917576
2 Africa Eastern and Southern  16.49366 16.907344 16.907346
3                 Afghanistan   0.10000  0.104707  0.104707
4  Africa Western and Central  15.49113 15.725047 15.725047
5                      Angola   6.97000  6.971427  6.971427
6                     Albania  17.21000 17.736095 17.736095
```

Population Data Set

7. The World Bank. (2021). Population, total. The World Bank. https://data.worldbank.org/indicator/SP.POP.TOTL

- This dataset includes data regarding population totals by country.
- This dataset includes data pulled in for the years of 2010-2020.
- This data first needs to be exported to an Excel file in order to extract the data.
- The only variable for this data set is total population (total number, as is) by country across time. I will be looking at the years of 2010-2020 for population totals.
- To clean this data, I will ensure the format of the variable nomenclature schema is the same across the board, thereby being easily read and utilized in my analysis.

Population Totals Dataset: Loaded and Cleaned

```
                   Country.Name Country.Code    Indicator.Name Indicator.Code
1                         Aruba          ABW Population, total    SP.POP.TOTL
2 Africa Eastern and Southern          AFE Population, total    SP.POP.TOTL
3                   Afghanistan          AFG Population, total    SP.POP.TOTL
4  Africa Western and Central          AFW Population, total    SP.POP.TOTL
5                        Angola          AGO Population, total    SP.POP.TOTL
6                       Albania          ALB Population, total    SP.POP.TOTL
      X2010     X2011     X2012     X2013     X2014     X2015     X2016
1    101665    102050    102565    103165    103776    104339    104865
2 518468229 532760424 547482863 562601578 578075373 593871847 609978946
3  29185511  30117411  31161378  32269592  33370804  34413603  35383028
4 350556886 360285439 370243017 380437896 390882979 401586651 412551299
5  23356247  24220660  25107925  26015786  26941773  27884380  28842482
6   2913021   2905195   2900401   2895092   2889104   2880703   2876101
      X2017     X2018     X2019     X2020
1    105361    105846    106310    106766
2 626392880 643090131 660046272 677243299
3  36296111  37171922  38041757  38928341
4 423769930 435229381 446911598 458803476
5  29816769  30809787  31825299  32866268
6   2873457   2866376   2854191   2837743

                        country country_code         indicator indicator_code
1                         Aruba          ABW Population, total    SP.POP.TOTL
2 Africa Eastern and Southern          AFE Population, total    SP.POP.TOTL
3                   Afghanistan          AFG Population, total    SP.POP.TOTL
4  Africa Western and Central          AFW Population, total    SP.POP.TOTL
5                        Angola          AGO Population, total    SP.POP.TOTL
6                       Albania          ALB Population, total    SP.POP.TOTL
       2010      2011      2012      2013      2014      2015      2016
1    101665    102050    102565    103165    103776    104339    104865
2 518468229 532760424 547482863 562601578 578075373 593871847 609978946
3  29185511  30117411  31161378  32269592  33370804  34413603  35383028
4 350556886 360285439 370243017 380437896 390882979 401586651 412551299
5  23356247  24220660  25107925  26015786  26941773  27884380  28842482
6   2913021   2905195   2900401   2895092   2889104   2880703   2876101
       2017      2018      2019      2020
1    105361    105846    106310    106766
2 626392880 643090131 660046272 677243299
3  36296111  37171922  38041757  38928341
4 423769930 435229381 446911598 458803476
5  29816769  30809787  31825299  32866268
6   2873457   2866376   2854191   2837743
```

GDP (Wealth) Data Set

8. The World Bank. (2021). GDP (current US$). The World Bank. https://data.worldbank.org/indicator/NY.GDP.MKTP.CD

- This dataset includes data regarding GDP totals by country.

- This dataset includes data pulled in for the years of 2010-2020.
- This data first needs to be exported to an Excel file in order to extract the data.
- The only variable for this data set is total GDP (in US dollars) by country across time. I will be looking at the years of 2010-2020 for population totals.
- To clean this data, I will ensure the format of the variable nomenclature schema is the same across the board, thereby being easily read and utilized in my analysis.

GDP Totals Dataset: Loaded and Cleaned

```
                 Country.Name Country.Code    Indicator.Name Indicator.Code
1                       Aruba          ABW GDP (current US$) NY.GDP.MKTP.CD
2 Africa Eastern and Southern          AFE GDP (current US$) NY.GDP.MKTP.CD
3                 Afghanistan          AFG GDP (current US$) NY.GDP.MKTP.CD
4  Africa Western and Central          AFW GDP (current US$) NY.GDP.MKTP.CD
5                      Angola          AGO GDP (current US$) NY.GDP.MKTP.CD
6                     Albania          ALB GDP (current US$) NY.GDP.MKTP.CD
        X2010        X2011        X2012        X2013        X2014        X20
15
1   2390502793   2549720670   2534636872   2727849721   2790849162   29629050
28
2 805795000000 898605000000 915590000000 930086000000 958825000000 8954400000
00
3  15856574731  17804292964  20001598506  20561069558  20484885120  199071114
19
4 580217000000 658428000000 716935000000 807819000000 846943000000 7574920000
00
5  83799496611 111790000000 128053000000 136710000000 145712000000 1161940000
00
6  11926928506  12890765324  12319830252  12776217195  13228144008  113868463
19
        X2016        X2017        X2018        X2019        X2020
1   2983636872   3092430168   3202188607           NA           NA
2 856992000000 964791000000 986611000000 980372000000 900829000000
3  18017749074  18869945678  18353881130  19291104008  19807067268
4 687485000000 680989000000 738131000000 792079000000 786585000000
5 101124000000 122124000000 101353000000  89417190341  62306913444
6  11861200797  13019693451  15147020535  15286612573  14799615097

                      country country_code         indicator indicator_code
1                       Aruba          ABW GDP (current US$) NY.GDP.MKTP.CD
2 Africa Eastern and Southern          AFE GDP (current US$) NY.GDP.MKTP.CD
3                 Afghanistan          AFG GDP (current US$) NY.GDP.MKTP.CD
4  Africa Western and Central          AFW GDP (current US$) NY.GDP.MKTP.CD
5                      Angola          AGO GDP (current US$) NY.GDP.MKTP.CD
6                     Albania          ALB GDP (current US$) NY.GDP.MKTP.CD
         2010         2011         2012         2013         2014         20
15
1   2390502793   2549720670   2534636872   2727849721   2790849162   29629050
28
2 805795000000 898605000000 915590000000 930086000000 958825000000 8954400000
```

```
00
3  15856574731  17804292964  20001598506  20561069558  20484885120  199071114
19
4 580217000000 658428000000 716935000000 807819000000 846943000000 7574920000
00
5  83799496611 111790000000 128053000000 136710000000 145712000000 1161940000
00
6  11926928506  12890765324  12319830252  12776217195  13228144008  113868463
19
          2016         2017         2018         2019         2020
1   2983636872   3092430168   3202188607           NA           NA
2 856992000000 964791000000 986611000000 980372000000 900829000000
3  18017749074  18869945678  18353881130  19291104008  19807067268
4 687485000000 680989000000 738131000000 792079000000 786585000000
5 101124000000 122124000000 101353000000  89417190341  62306913444
6  11861200797  13019693451  15147020535  15286612573  14799615097
```

Datasets After Initial Cleaning

To paint a brief synopsis of what each of the data sets looks like after cleaning, please see below to view the first few lines of the data (I am not including the entire datasets to preserve space within this report).

```
    country total_renewable_per_capita_m3_cap_2021 total_renewable_2021
1 Australia                                16175.6                407.7
2   Austria                                 9603.5                 86.0
3   Belgium                                 2232.8                 25.8
4    Canada                                92967.2               3478.0
5     Chile                                53160.0               1007.5
6  Colombia                                     NA                   NA

      country losses_2010 losses_2011 losses_2012 losses_2013 losses_2014
1      Latvia     6090.00     6090.00     6090.00     6090.00     6090.00
2      Norway     3551.00     3585.00     3620.00     3657.00     3694.00
3     Germany     3697.83     3691.34     3684.46     3685.34     5158.98
4      Turkey          NA     3245.00     3073.00     4014.00     2567.00
5   Lithuania     3330.00     3420.00     3310.00     3260.00     3290.00
6 Switzerland     2648.90          NA          NA          NA          NA
  losses_2015 losses_2016 losses_2017 losses_2018 losses_2019
1     6070.00     6070.00        6070     5943.00          NA
2     4188.00     4188.00        4188     4182.00        4182
3     5252.59     5064.86        5023     3747.00          NA
4     5246.00     3324.00        2763     3273.00        2810
5     3220.00     3100.00        3120     3000.00        3020
6     2491.90          NA          NA     2395.25          NA

          country usage_2010 usage_2011 usage_2012 usage_2013 usage_2014
1  Czech Republic       0.79       0.72       0.70       0.71       0.71
2         Estonia       0.60       0.66       0.73       0.71       0.70
3         Germany       0.87       0.80       0.81       0.75       0.76
4 Slovak Republic       0.83       0.79       0.68       0.65       0.78
```

```
5      Luxembourg         NA       0.72       0.49       0.47       0.53
6     Switzerland       0.71         NA         NA         NA         NA
  usage_2015 usage_2016 usage_2017 usage_2018 usage_2019
1       0.74       0.80       0.88       1.15       1.45
2       0.70       0.75       0.87       0.90       0.80
3       0.78       0.75       0.74       0.88         NA
4       0.77       0.77       0.78       0.82       0.77
5       0.58       0.50       0.57       0.69       0.60
6       0.70         NA         NA       0.69         NA

    country consumption_2010 consumption_2011 consumption_2012 consumption_20
13
1 Australia          40.5952          40.7088          39.7431           39.56
75
2   Austria          18.9726          19.9488          19.3716           18.86
91
3   Belgium          14.8156          15.5685          14.2317           14.05
48
4    Canada          28.4746          29.0813          28.3162           30.68
97
5     Chile          39.1149          39.0211          39.7548           41.08
83
6  Colombia           5.2268           5.5312           5.6416            5.59
65
  consumption_2014 consumption_2015 consumption_2016 consumption_2017
1          38.6466          38.4333          38.0788          37.7188
2          18.7661          18.2697          19.0816          18.7208
3          13.7693          13.5473          13.5042          13.8141
4          29.3258          28.9513          28.8247          28.7080
5          40.4474          40.4297          40.4411          40.4049
6           5.7755           5.8406           5.8721           5.8907
  consumption_2018 consumption_2019
1               NA               NA
2          18.7713          18.6614
3          13.3837          12.3200
4               NA               NA
5               NA               NA
6               NA               NA

    country footprint_2010 footprint_2011 footprint_2012 footprint_2013
1 Australia        40.7986        40.9456        41.5003        41.1921
2   Austria        30.0636        31.0990        30.9061        31.0024
3   Belgium        22.8657        22.2866        22.5309        22.8641
4    Canada        33.2096        33.5376        34.3866        35.0090
5     Chile        15.7977        16.2076        16.1666        17.1028
6  Colombia         9.0587         9.8797        10.3391        10.2774
  footprint_2014 footprint_2015 footprint_2016 footprint_2017
1        42.1842        42.5502        42.8273        43.1128
2        31.5693        31.8904        32.2465        32.5868
3        23.1293        23.4314        23.7751        24.1195
```

```
4        35.0283        34.8016        34.8148        34.8364
5        17.0253        17.0462        17.1084        17.1491
6        10.4718        10.5186        10.6144        10.6902

    country mammals birds reptiles amphibians fish marine_fish freshwater_fis
h
1 Australia      10    17       10         15    7           4
3
2   Austria       4    14        3          1    6          NA
6
3   Belgium       4    20       NA          3    3           2
1
4    Canada      11    26        5          2    9           3
6
5     Chile       3     2        9         10    1          NA
1
6  Colombia       5    16       11         75    7           1
2
  vascular_plants mosses lichens invertabrates total
1             198     NA      NA            30   294
2             172     34      57           302   599
3             253     NA      NA           135   421
4             315     75      70           205   727
5              74      1       1            33   135
6             112      6       0             6   241

          country pet_use_2010 pet_use_2011 pet_use_2012 pet_use_2013
1   Global/World     88433.47     89005.81     90300.19     91810.30
2    Afghanistan        42.76        55.82        49.06        35.08
3        Albania        25.95        25.74        22.64        27.17
4        Algeria       347.41       349.58       373.79       400.34
5 American Samoa         4.14         2.35         2.35         2.35
6         Angola       106.95       111.97       119.81       138.74
  pet_use_2014 pet_use_2015 pet_use_2016 pet_use_2017 pet_use_2018 pet_use_20
19
1     93354.19     95127.56     96557.97     98890.83     99907.90
NA
2        27.68        34.91        25.59        23.63        23.15
NA
3        28.13        26.06        27.17        35.33        33.05
NA
4       415.61       431.27       420.72       407.39       418.17
NA
5         2.35         2.35         2.35         2.35         2.34
NA
6       158.87       156.75       150.61       134.83       133.35
NA
  pet_use_2020
1           NA
2           NA
```

```
3           NA
4           NA
5           NA
6           NA

                     country land_2016 land_2017 land_2018
1                      Aruba   0.53000 18.917576 18.917576
2 Africa Eastern and Southern  16.49366 16.907344 16.907346
3                Afghanistan   0.10000  0.104707  0.104707
4  Africa Western and Central  15.49113 15.725047 15.725047
5                     Angola   6.97000  6.971427  6.971427
6                    Albania  17.21000 17.736095 17.736095

                     country country_code         indicator indicator_code
1                      Aruba          ABW Population, total    SP.POP.TOTL
2 Africa Eastern and Southern          AFE Population, total    SP.POP.TOTL
3                Afghanistan          AFG Population, total    SP.POP.TOTL
4  Africa Western and Central          AFW Population, total    SP.POP.TOTL
5                     Angola          AGO Population, total    SP.POP.TOTL
6                    Albania          ALB Population, total    SP.POP.TOTL
       2010      2011      2012      2013      2014      2015      2016
1    101665    102050    102565    103165    103776    104339    104865
2 518468229 532760424 547482863 562601578 578075373 593871847 609978946
3  29185511  30117411  31161378  32269592  33370804  34413603  35383028
4 350556886 360285439 370243017 380437896 390882979 401586651 412551299
5  23356247  24220660  25107925  26015786  26941773  27884380  28842482
6   2913021   2905195   2900401   2895092   2889104   2880703   2876101
       2017      2018      2019      2020
1    105361    105846    106310    106766
2 626392880 643090131 660046272 677243299
3  36296111  37171922  38041757  38928341
4 423769930 435229381 446911598 458803476
5  29816769  30809787  31825299  32866268
6   2873457   2866376   2854191   2837743

                     country country_code         indicator indicator_code
1                      Aruba          ABW GDP (current US$) NY.GDP.MKTP.CD
2 Africa Eastern and Southern          AFE GDP (current US$) NY.GDP.MKTP.CD
3                Afghanistan          AFG GDP (current US$) NY.GDP.MKTP.CD
4  Africa Western and Central          AFW GDP (current US$) NY.GDP.MKTP.CD
5                     Angola          AGO GDP (current US$) NY.GDP.MKTP.CD
6                    Albania          ALB GDP (current US$) NY.GDP.MKTP.CD
          2010         2011         2012         2013         2014         20
15
1   2390502793   2549720670   2534636872   2727849721   2790849162   29629050
28
2 805795000000 898605000000 915590000000 930086000000 958825000000 8954400000
00
3  15856574731  17804292964  20001598506  20561069558  20484885120  199071114
19
```

```
4 580217000000 658428000000 716935000000 807819000000 846943000000 7574920000
00
5  83799496611 111790000000 128053000000 136710000000 145712000000 1161940000
00
6  11926928506  12890765324  12319830252  12776217195  13228144008  113868463
19
          2016         2017         2018         2019         2020
1   2983636872   3092430168   3202188607           NA           NA
2 856992000000 964791000000 986611000000 980372000000 900829000000
3  18017749074  18869945678  18353881130  19291104008  19807067268
4 687485000000 680989000000 738131000000 792079000000 786585000000
5 101124000000 122124000000 101353000000  89417190341  62306913444
6  11861200797  13019693451  15147020535  15286612573  14799615097
```

Cleaning Part II

Now that each dataset is fully cleaned and ready to be used individually, I will be comparing these datasets for an additional step of cleaning. Because I care to assess the relationships present within the datasets based upon a country's wealth and a country's population, I need to ensure that each country I am assessing has the appropriate data available for all 10 variable measures I am assessing.

Because not every dataset I am utilizing includes the same number of and specific/identical countries, I want to ensure that the available countries for my independent variables (Population Totals and GDP) are compared with the available countries for each of the 10 dependent variables I am assessing (on an individual basis). In other words, before I complete my analysis of, for example, renewable water resources and population totals, I wish to restrict the countries analyzed to only those present in both datasets. This will eliminate any missing data and will allow for less skewed results (especially because I don't care to analyze a specific selection of countries, but rather want to get a glimpse of countries/regions overall, even if that means different sets of countries are present in different datasets). So, for example, my analysis of the relationship between population and renewable water resources may have a different set of countries available than my analysis of the relationship between population and forest losses, and that is okay. As long as I can ascertain overall and assess the relationship of a country's population (or wealth) relative to my other variables individually, that is what I am seeking.

Comparing and Combining the Population Dataset with Each Dependent Variable Dataset

Comparing and Combining the GDP Dataset with Each Dependent Variable Dataset

Descriptive Statistics

To obtain some basic information on each of the datasets, I have run some descriptive summary statistics. These descriptive statistics have been run on the original, cleaned datasets.Primarily, I was interested in viewing the mean (average) for the amounts of the different variables in each of these datasets.

Water Dataset

```
   country           total_renewable_per_capita_m3_cap_2021 total_renewable_20
21
 Length:40          Min.   :   332.2                       Min.   :   1.6  
 Class :character   1st Qu.:  2710.9                       1st Qu.:  44.5  
 Mode  :character   Median :  7035.5                       Median :  91.8  
                    Mean   : 28050.5                       Mean   : 321.2  
                    3rd Qu.: 15623.1                       3rd Qu.: 220.2  
                    Max.   :501420.2                       Max.   :3478.0  
                    NA's   :4                              NA's   :5       
```

From the output of these statistics, we can ascertain the following:

- For all of the countries present within the dataset, the mean total water renewable resources in cubic meters per capita is approximately 28,050.
- The mean for total renewable water overall was approximately 321 unit ratios.

Forest Datasets

```
   country            losses_2010      losses_2011         losses_2012  
 Length:27          Min.   :   11   Min.   :    86.0   Min.   :  85  
 Class :character   1st Qu.: 1149   1st Qu.:   584.5   1st Qu.: 421  
 Mode  :character   Median : 3551   Median :  3502.5   Median :3192  
                    Mean   : 3753   Mean   : 26924.3   Mean   :2930  
                    3rd Qu.: 5235   3rd Qu.:  5074.5   3rd Qu.:4473  
                    Max.   :10300   Max.   :288298.4   Max.   :9126  
                    NA's   :12      NA's   :15         NA's   :13    
  losses_2013       losses_2014       losses_2015        losses_2016    
 Min.   :  62.0   Min.   : 101.0   Min.   :   62.0   Min.   :    62  
 1st Qu.: 435.5   1st Qu.: 526.8   1st Qu.:  605.5   1st Qu.:   473  
 Median :3458.5   Median :2928.5   Median : 4188.0   Median :  3324  
 Mean   :3049.7   Mean   :3135.7   Mean   : 3860.7   Mean   : 28170  
 3rd Qu.:4555.5   3rd Qu.:4841.7   3rd Qu.: 5249.3   3rd Qu.:  5602  
 Max.   :9354.0   Max.   :9768.0   Max.   :10472.9   Max.   :326125  
 NA's   :13       NA's   :15       NA's   :12        NA's   :14      
  losses_2017        losses_2018       losses_2019    
 Min.   :   62.0   Min.   :  62.0   Min.   :  62.0  
 1st Qu.:  292.9   1st Qu.: 337.2   1st Qu.: 259.5  
 Median : 2763.0   Median :2017.6   Median :1687.0  
 Mean   : 3097.0   Mean   :2129.8   Mean   :1799.3  
 3rd Qu.: 5023.0   3rd Qu.:3391.5   3rd Qu.:2967.5  
 Max.   :11994.8   Max.   :5943.0   Max.   :4182.0  
 NA's   :14        NA's   :15       NA's   :21      
```

```
   country             usage_2010        usage_2011        usage_2012    
 Length:32          Min.   :0.3400   Min.   :0.2600   Min.   :0.3400  
 Class :character   1st Qu.:0.5150   1st Qu.:0.5200   1st Qu.:0.4800  
 Mode  :character   Median :0.5600   Median :0.6000   Median :0.5900  
                    Mean   :0.6155   Mean   :0.5887   Mean   :0.5984  
                    3rd Qu.:0.7050   3rd Qu.:0.7050   3rd Qu.:0.6800  
                    Max.   :1.0600   Max.   :0.8000   Max.   :1.1000  
```

```
                      NA's   :10       NA's   :17       NA's   :13
  usage_2013        usage_2014        usage_2015       usage_2016
 Min.   :0.2800    Min.   :0.2300    Min.   :0.250   Min.   :0.2300
 1st Qu.:0.4700    1st Qu.:0.5300    1st Qu.:0.535   1st Qu.:0.5075
 Median :0.5800    Median :0.5600    Median :0.575   Median :0.5650
 Mean   :0.5779    Mean   :0.5806    Mean   :0.609   Mean   :0.5861
 3rd Qu.:0.6850    3rd Qu.:0.7000    3rd Qu.:0.710   3rd Qu.:0.7350
 Max.   :0.8200    Max.   :0.7800    Max.   :0.830   Max.   :0.8000
 NA's   :13        NA's   :15        NA's   :12      NA's   :14
  usage_2017        usage_2018        usage_2019
 Min.   :0.2300    Min.   :0.2000    Min.   :0.1900
 1st Qu.:0.5100    1st Qu.:0.5200    1st Qu.:0.5700
 Median :0.5800    Median :0.6000    Median :0.6100
 Mean   :0.6111    Mean   :0.6365    Mean   :0.6636
 3rd Qu.:0.7475    3rd Qu.:0.6900    3rd Qu.:0.7100
 Max.   :0.8800    Max.   :1.1500    Max.   :1.4500
 NA's   :14        NA's   :15        NA's   :21
```

From the output of these statistics, we can see the average forest losses and usage for each of the years for the years 2010 to 2019 (for all countries).

Material Datasets

```
   country          consumption_2010  consumption_2011  consumption_2012
 Length:194         Min.   : 0.4743   Min.   : 0.5241   Min.   : 0.9092
 Class :character   1st Qu.: 4.0837   1st Qu.: 4.0664   1st Qu.: 4.0649
 Mode  :character   Median : 8.5151   Median : 8.6657   Median : 8.4423
                    Mean   :10.8138   Mean   :10.7458   Mean   :10.5492
                    3rd Qu.:14.6339   3rd Qu.:14.7738   3rd Qu.:13.8664
                    Max.   :46.3321   Max.   :40.7088   Max.   :44.9067
                    NA's   :6         NA's   :6         NA's   :4
 consumption_2013  consumption_2014  consumption_2015  consumption_2016
 Min.   : 0.6498   Min.   : 0.7908   Min.   : 0.7882   Min.   : 0.7777
 1st Qu.: 4.0669   1st Qu.: 4.1804   1st Qu.: 4.1601   1st Qu.: 4.1703
 Median : 8.4396   Median : 8.3607   Median : 8.5396   Median : 8.6610
 Mean   :10.6766   Mean   :10.7393   Mean   :10.8821   Mean   :10.9746
 3rd Qu.:13.6951   3rd Qu.:13.7401   3rd Qu.:13.8193   3rd Qu.:13.6821
 Max.   :44.1823   Max.   :46.8361   Max.   :45.6324   Max.   :46.8435
 NA's   :4         NA's   :4         NA's   :3         NA's   :3
 consumption_2017  consumption_2018 consumption_2019
 Min.   : 0.7676   Min.   : 7.914   Min.   : 8.096
 1st Qu.: 4.1394   1st Qu.:11.878   1st Qu.:11.914
 Median : 8.5203   Median :14.435   Median :14.439
 Mean   :11.1341   Mean   :16.540   Mean   :17.058
 3rd Qu.:14.0049   3rd Qu.:20.320   3rd Qu.:21.777
 Max.   :48.2990   Max.   :34.361   Max.   :31.340
 NA's   :3         NA's   :156      NA's   :159

   country          footprint_2010     footprint_2011     footprint_2012
 Length:179         Min.   :  0.0558   Min.   :  0.0521   Min.   :  0.0723
 Class :character   1st Qu.:  3.2032   1st Qu.:  3.3184   1st Qu.:  3.4726
```

```
 Mode  :character   Median :  8.4407   Median :  8.8245   Median :  8.9873  
                    Mean   : 13.2918   Mean   : 13.7925   Mean   : 13.5416  
                    3rd Qu.: 21.3566   3rd Qu.: 20.6747   3rd Qu.: 20.5718  
                    Max.   :120.8080   Max.   :144.9799   Max.   :115.6759  
                    NA's   :1          NA's   :1          NA's   :1         
 footprint_2013     footprint_2014     footprint_2015     footprint_2016    
 Min.   :  0.3846   Min.   :  0.3923   Min.   :  0.4033   Min.   :  0.4153  
 1st Qu.:  3.5206   1st Qu.:  3.5487   1st Qu.:  3.6027   1st Qu.:  3.6810  
 Median :  8.9823   Median :  9.2585   Median :  9.5491   Median :  9.6862  
 Mean   : 13.8349   Mean   : 13.9098   Mean   : 14.0440   Mean   : 14.2108  
 3rd Qu.: 21.0733   3rd Qu.: 21.2770   3rd Qu.: 21.4568   3rd Qu.: 21.6376  
 Max.   :116.9320   Max.   :116.0565   Max.   :116.6814   Max.   :117.5743  
 NA's   :1          NA's   :1          NA's   :1          NA's   :1         
 footprint_2017    
 Min.   :  0.4273  
 1st Qu.:  3.7452  
 Median :  9.8134  
 Mean   : 14.3795  
 3rd Qu.: 21.8354  
 Max.   :118.4760  
 NA's   :1        
```

From the output of these statistics, we can see the average material consumption and footprints for each of the years for the years 2010 to 2019 (for all countries).

Endangered Species Dataset

```
   country             mammals           birds          reptiles     
 Length:39          Min.   : 0.000   Min.   : 2.00   Min.   : 0.000  
 Class :character   1st Qu.: 2.000   1st Qu.: 9.25   1st Qu.: 1.000  
 Mode  :character   Median : 3.500   Median :16.00   Median : 1.000  
                    Mean   : 4.767   Mean   :16.18   Mean   : 3.348  
                    3rd Qu.: 6.500   3rd Qu.:22.25   3rd Qu.: 4.500  
                    Max.   :13.000   Max.   :42.00   Max.   :11.000  
                    NA's   :9        NA's   :5       NA's   :16      
   amphibians         fish          marine_fish    freshwater_fish
 Min.   : 0.00   Min.   :  1.00   Min.   : 0.00   Min.   :  1    
 1st Qu.: 0.00   1st Qu.:  3.00   1st Qu.: 1.00   1st Qu.:  2    
 Median : 1.00   Median :  6.00   Median : 3.00   Median :  4    
 Mean   : 7.12   Mean   : 13.36   Mean   : 4.65   Mean   : 11    
 3rd Qu.: 3.00   3rd Qu.:  9.75   3rd Qu.: 5.25   3rd Qu.:  9    
 Max.   :75.00   Max.   :135.00   Max.   :34.00   Max.   :101    
 NA's   :14      NA's   :11       NA's   :19      NA's   :14     
 vascular_plants     mosses          lichens       invertabrates  
 Min.   :  4.0   Min.   :  0.00   Min.   :  0.0   Min.   :  1.0  
 1st Qu.: 32.0   1st Qu.: 10.25   1st Qu.: 24.0   1st Qu.: 19.0  
 Median : 74.0   Median : 22.50   Median : 47.0   Median : 45.0  
 Mean   :118.0   Mean   : 36.05   Mean   : 70.7   Mean   : 93.9  
 3rd Qu.:182.5   3rd Qu.: 50.50   3rd Qu.: 73.5   3rd Qu.:130.5  
 Max.   :520.0   Max.   :137.00   Max.   :337.0   Max.   :559.0  
 NA's   :8       NA's   :17       NA's   :19      NA's   :9      
```

```
    total
Min.   :   6.0
1st Qu.:  74.0
Median : 215.0
Mean   : 316.2
3rd Qu.: 429.0
Max.   :1581.0
NA's   :3
```

From the output of these statistics, we can see the average total number of critically endangered species for a variety of wildlife classes for all countries present within the dataset.

Petroleum and Other Liquids Dataset

```
   country            pet_use_2010        pet_use_2011        pet_use_2012
 Length:231         Min.   :    0.04   Min.   :    0.04   Min.   :    0.04
 Class :character   1st Qu.:    7.75   1st Qu.:    8.06   1st Qu.:    8.52
 Mode  :character   Median :   34.74   Median :   35.43   Median :   36.16
                    Mean   :  807.61   Mean   :  812.84   Mean   :  820.91
                    3rd Qu.:  221.44   3rd Qu.:  212.03   3rd Qu.:  217.93
                    Max.   :88433.47   Max.   :89005.81   Max.   :90300.19
                    NA's   :12         NA's   :12         NA's   :11
  pet_use_2013        pet_use_2014        pet_use_2015        pet_use_2016
 Min.   :    0.05   Min.   :    0.05   Min.   :    0.05   Min.   :    0.05
 1st Qu.:    8.93   1st Qu.:    9.00   1st Qu.:    9.03   1st Qu.:    9.30
 Median :   35.84   Median :   37.56   Median :   41.00   Median :   46.83
 Mean   :  834.64   Mean   :  848.67   Mean   :  864.80   Mean   :  873.83
 3rd Qu.:  218.32   3rd Qu.:  208.62   3rd Qu.:  217.98   3rd Qu.:  210.14
 Max.   :91810.30   Max.   :93354.19   Max.   :95127.56   Max.   :96557.97
 NA's   :11         NA's   :11         NA's   :11         NA's   :10
  pet_use_2017        pet_use_2018        pet_use_2019        pet_use_2020
 Min.   :    0.05   Min.   :    0.05   Min.   :   19.68   Min.   :   14.01
 1st Qu.:    8.41   1st Qu.:    9.16   1st Qu.:  176.79   1st Qu.:  148.11
 Median :   46.02   Median :   46.39   Median :  287.25   Median :  244.13
 Mean   :  894.94   Mean   :  904.14   Mean   : 1321.36   Mean   : 1157.35
 3rd Qu.:  214.33   3rd Qu.:  223.20   3rd Qu.: 1273.30   3rd Qu.: 1059.69
 Max.   :98890.83   Max.   :99907.90   Max.   :20542.85   Max.   :18185.91
 NA's   :10         NA's   :10         NA's   :195        NA's   :195
```

From the output of these statistics, we can see the average petroleum consumption for each of the years for the years 2010 to 2020 (for all countries).

Land Dataset

```
   country              land_2016          land_2017            land_2018
 Length:266         Min.   : 0.10   Min.   : 0.00037   Min.   : 0.00037
 Class :character   1st Qu.: 7.33   1st Qu.: 7.63046   1st Qu.: 7.63046
 Mode  :character   Median :14.58   Median :14.80696   Median :14.80696
                    Mean   :16.42   Mean   :16.55093   Mean   :16.55093
                    3rd Qu.:22.42   3rd Qu.:22.53335   3rd Qu.:22.53335
```

```
                         Max.   :62.50    Max.   :54.40416    Max.   :54.40416
                         NA's   :8        NA's   :7           NA's   :7
```

From the output of these statistics, we can see the average percentage of protected land for each year for the years 2016 to 2018 (for all countries).

Population Dataset

```
   country           country_code         indicator          indicator_code    
 Length:266         Length:266         Length:266         Length:266        
 Class :character   Class :character   Class :character   Class :character  
 Mode  :character   Mode  :character   Mode  :character   Mode  :character  
                                                                            
                                                                            
                                                                            
                                                                            
      2010               2011               2012          
 Min.   :1.001e+04   Min.   :1.007e+04   Min.   :1.014e+04  
 1st Qu.:1.523e+06   1st Qu.:1.563e+06   1st Qu.:1.540e+06  
 Median :9.491e+06   Median :9.473e+06   Median :9.825e+06  
 Mean   :2.793e+08   Mean   :2.828e+08   Mean   :2.875e+08  
 3rd Qu.:5.928e+07   3rd Qu.:5.938e+07   3rd Qu.:6.058e+07  
 Max.   :6.922e+09   Max.   :7.003e+09   Max.   :7.086e+09  
 NA's   :1           NA's   :1           NA's   :2          
      2013               2014               2015          
 Min.   :1.021e+04   Min.   :1.029e+04   Min.   :1.037e+04  
 1st Qu.:1.575e+06   1st Qu.:1.610e+06   1st Qu.:1.646e+06  
 Median :9.949e+06   Median :1.002e+07   Median :1.022e+07  
 Mean   :2.913e+08   Mean   :2.950e+08   Mean   :2.988e+08  
 3rd Qu.:6.121e+07   3rd Qu.:6.174e+07   3rd Qu.:6.183e+07  
 Max.   :7.170e+09   Max.   :7.254e+09   Max.   :7.339e+09  
 NA's   :2           NA's   :2           NA's   :2          
      2016               2017               2018          
 Min.   :1.047e+04   Min.   :1.058e+04   Min.   :1.068e+04  
 1st Qu.:1.690e+06   1st Qu.:1.717e+06   1st Qu.:1.740e+06  
 Median :1.036e+07   Median :1.041e+07   Median :1.046e+07  
 Mean   :3.026e+08   Mean   :3.064e+08   Mean   :3.102e+08  
 3rd Qu.:6.187e+07   3rd Qu.:6.192e+07   3rd Qu.:6.193e+07  
 Max.   :7.424e+09   Max.   :7.509e+09   Max.   :7.592e+09  
 NA's   :2           NA's   :2           NA's   :2          
      2019               2020          
 Min.   :1.076e+04   Min.   :1.083e+04  
 1st Qu.:1.752e+06   1st Qu.:1.757e+06  
 Median :1.048e+07   Median :1.053e+07  
 Mean   :3.138e+08   Mean   :3.175e+08  
 3rd Qu.:6.151e+07   3rd Qu.:6.160e+07  
 Max.   :7.673e+09   Max.   :7.753e+09  
 NA's   :2           NA's   :2          
```

From the output of these statistics, we can see the average population for each of the years for the years 2010 to 2020 (for all countries).

GDP Dataset

```
   country           country_code        indicator          indicator_code    
 Length:266         Length:266         Length:266         Length:266        
 Class :character   Class :character   Class :character   Class :character  
 Mode  :character   Mode  :character   Mode  :character   Mode  :character  
                                                                            
                                                                            
                                                                            
                                                                            
      2010                2011                2012          
 Min.   :3.182e+07   Min.   :3.871e+07   Min.   :3.767e+07  
 1st Qu.:7.013e+09   1st Qu.:7.931e+09   1st Qu.:8.448e+09  
 Median :4.125e+10   Median :4.467e+10   Median :4.723e+10  
 Mean   :2.103e+12   Mean   :2.344e+12   Mean   :2.418e+12  
 3rd Qu.:4.936e+11   3rd Qu.:5.551e+11   3rd Qu.:5.651e+11  
 Max.   :6.616e+13   Max.   :7.349e+13   Max.   :7.518e+13  
 NA's   :12          NA's   :10          NA's   :11         
      2013                2014                2015          
 Min.   :3.751e+07   Min.   :3.729e+07   Min.   :3.549e+07  
 1st Qu.:8.486e+09   1st Qu.:9.198e+09   1st Qu.:8.667e+09  
 Median :5.138e+10   Median :5.346e+10   Median :4.994e+10  
 Mean   :2.497e+12   Mean   :2.569e+12   Mean   :2.423e+12  
 3rd Qu.:5.607e+11   3rd Qu.:5.555e+11   3rd Qu.:5.499e+11  
 Max.   :7.732e+13   Max.   :7.945e+13   Max.   :7.503e+13  
 NA's   :10          NA's   :10          NA's   :11         
      2016                2017                2018          
 Min.   :3.655e+07   Min.   :4.062e+07   Min.   :4.259e+07  
 1st Qu.:8.620e+09   1st Qu.:9.752e+09   1st Qu.:1.080e+10  
 Median :5.140e+10   Median :5.402e+10   Median :5.751e+10  
 Mean   :2.464e+12   Mean   :2.639e+12   Mean   :2.804e+12  
 3rd Qu.:5.471e+11   3rd Qu.:6.180e+11   3rd Qu.:5.794e+11  
 Max.   :7.623e+13   Max.   :8.111e+13   Max.   :8.614e+13  
 NA's   :12          NA's   :12          NA's   :12         
      2019                2020          
 Min.   :4.727e+07   Min.   :4.886e+07  
 1st Qu.:1.256e+10   1st Qu.:1.368e+10  
 Median :6.395e+10   Median :7.326e+10  
 Mean   :2.930e+12   Mean   :3.100e+12  
 3rd Qu.:7.466e+11   3rd Qu.:9.139e+11  
 Max.   :8.744e+13   Max.   :8.458e+13  
 NA's   :19          NA's   :41         
```

From the output of these statistics, we can see the average GDP for each of the years for the years 2010 to 2020 (for all countries).

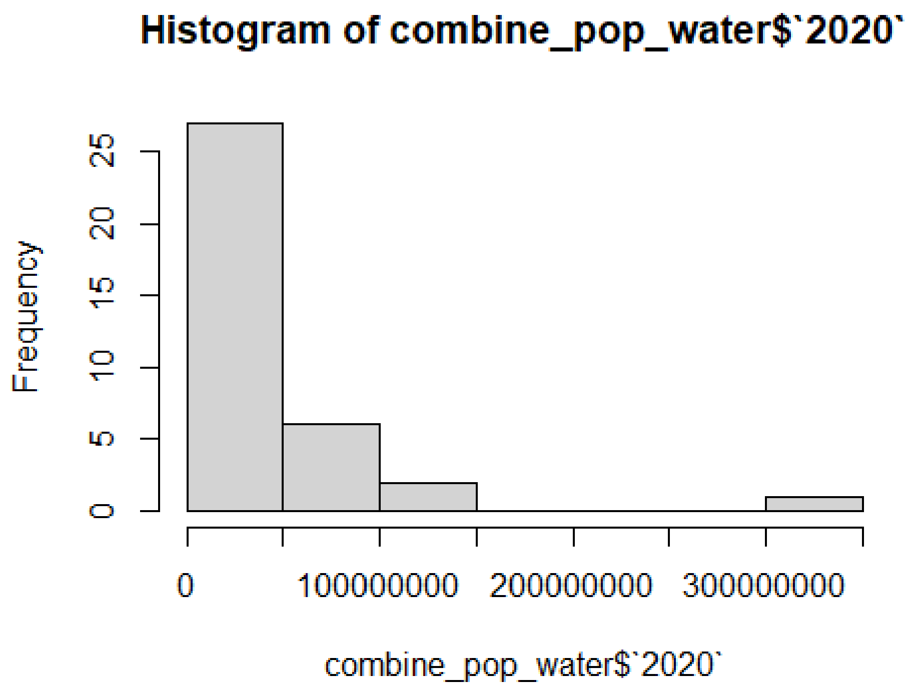
Correlational Analyses

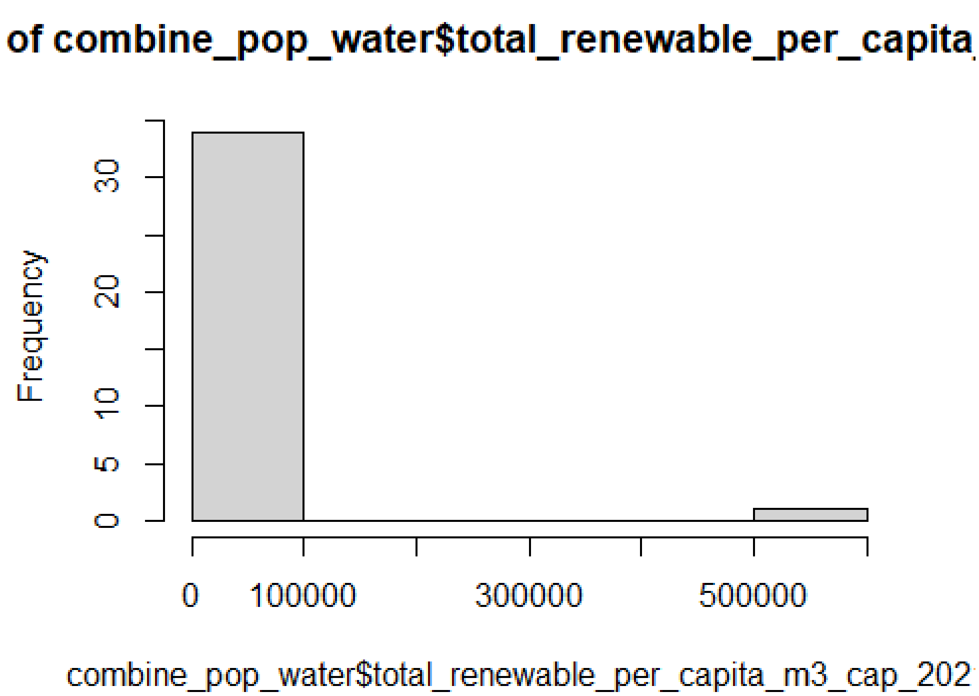
In an endeavor to answer my research questions posed earlier in this final project report, I determined that running some correlational analyses would be beneficial for assessing the relationships below:

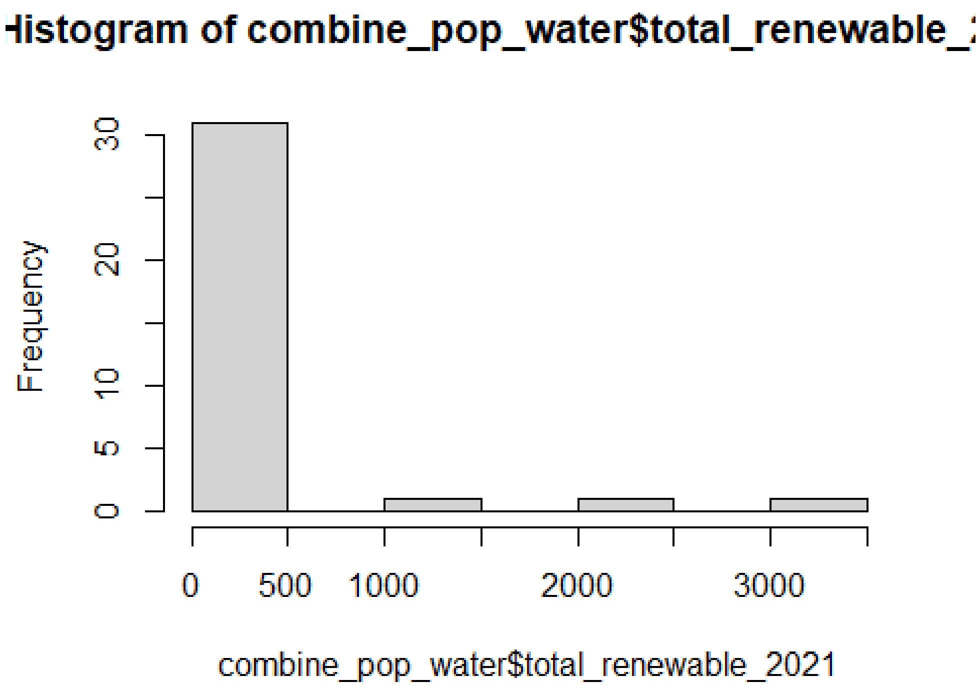
- A country's population and:
  - Renewable water resources
  - Forest losses
  - Forest usage
  - Material consumption
  - Material footprint
  - Critically endangered species
  - Petroleum and other liquids consumption
  - Projected terrestrial land
- A country's GDP and:
  - Renewable water resources
  - Forest losses
  - Forest usage
  - Material consumption
  - Material footprint
  - Critically endangered species
  - Petroleum and other liquids consumption
  - Projected terrestrial land

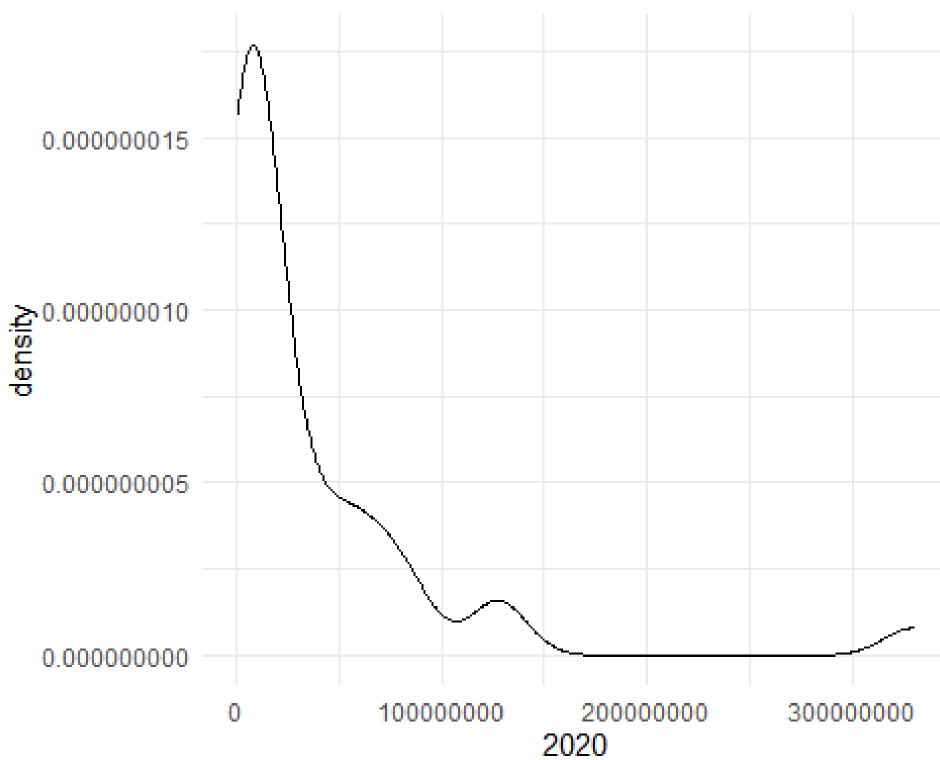
In each of the below sections, I run a correlational analysis for each IV with respect to each DV. It is important to note, that prior to any correlational analysis, I first checked my datasets for their distributions - since none of my data was normally distributed, and since all of the variables I am measuring are numeric (interval/ratio) in nature, each of the correlational tests I run will be a Spearman's Correlation.

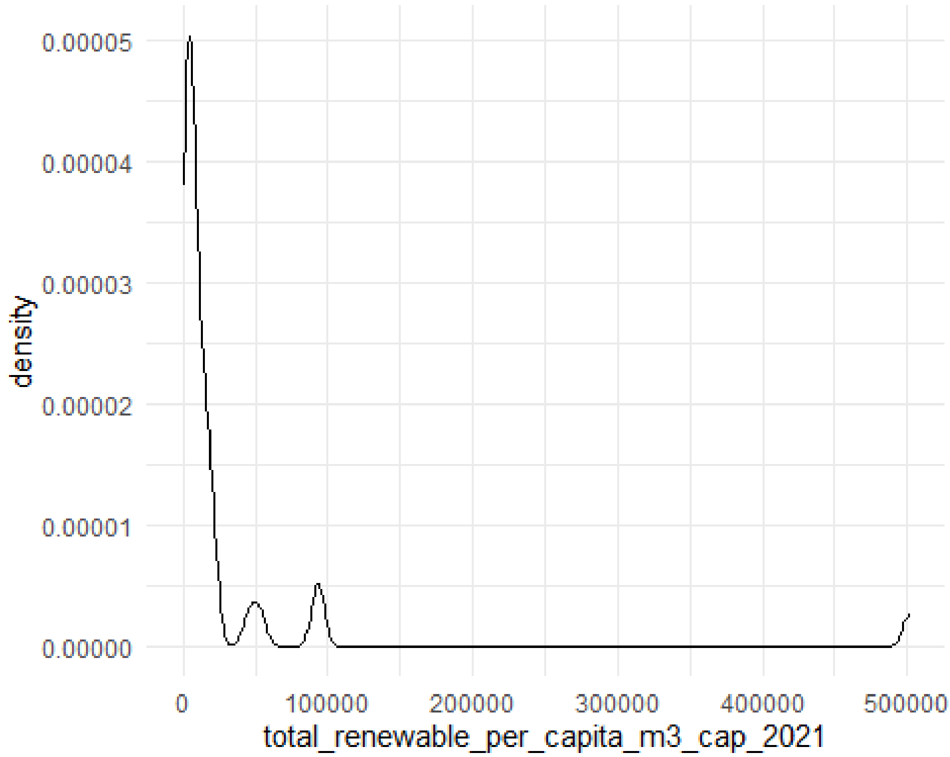
Population and Water Correlation (2020/2021)

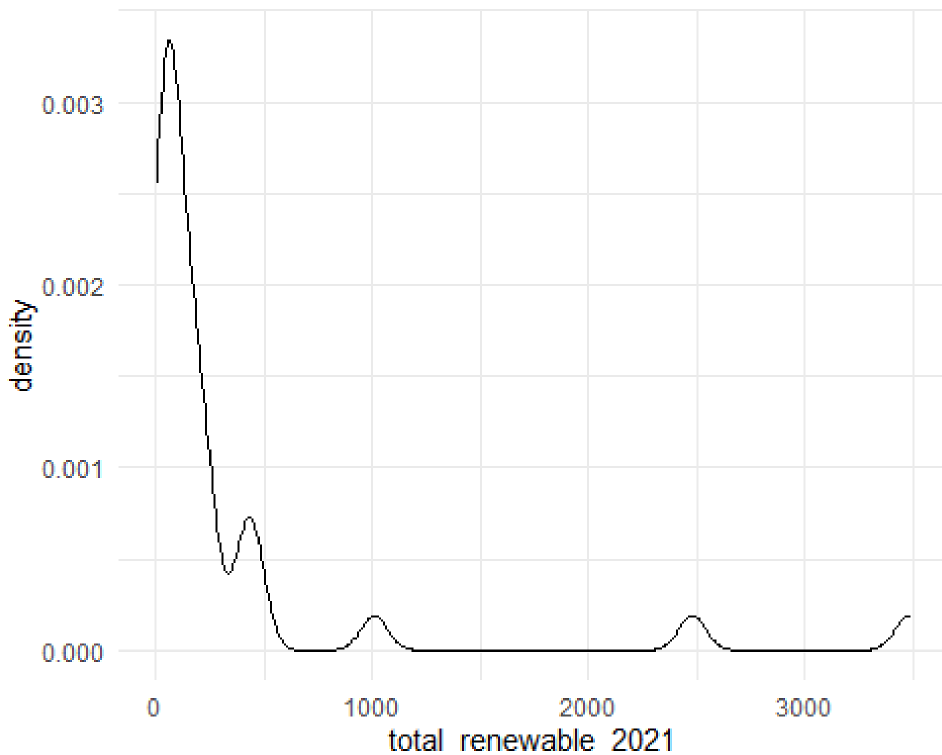





Histogram of combine_pop_water$total_renewable_2
Frequency
0
5
10
20
30
0
500
1000
2000
3000
combine_pop_water$total_renewable_2021


density
0.000000015
0.000000010
0.000000005
0.000000000
0
100000000
200000000
300000000
2020






```
[1] -0.4386555

[1] 0.5824738
```

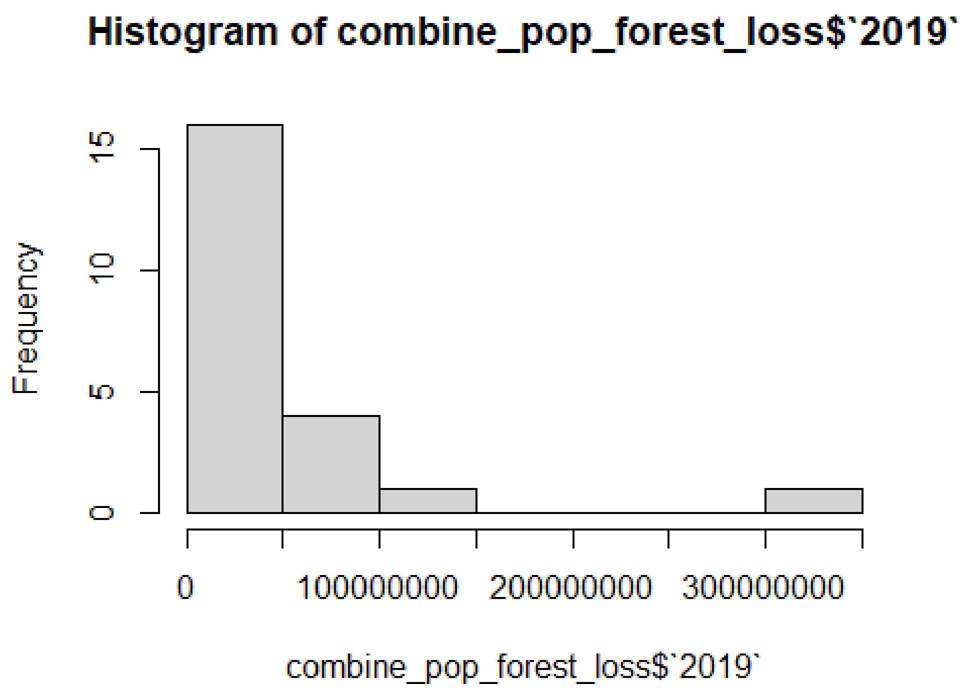
```
    Spearman's rank correlation rho

data:  combine_pop_water$`2020` and combine_pop_water$total_renewable_per_cap
ita_m3_cap_2021
S = 10272, p-value = 0.00839
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.4386555


    Spearman's rank correlation rho

data:  combine_pop_water$`2020` and combine_pop_water$total_renewable_2021
S = 2732.7, p-value = 0.0003012
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.5824738
```

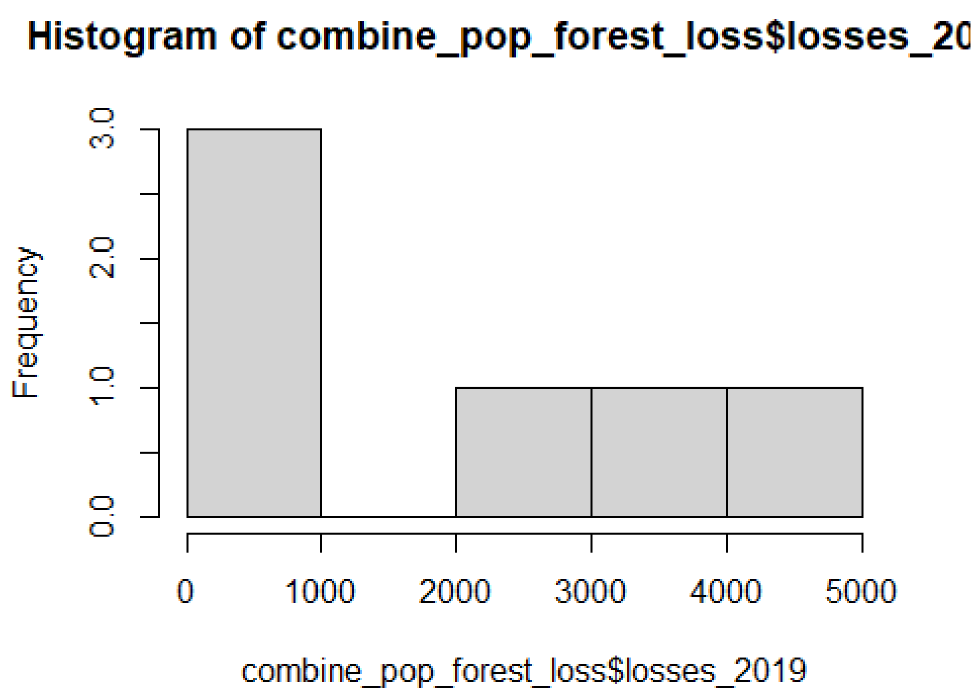
This test is running a correlation between the same countries present in both the population dataset and the water renewables dataset. The test is being run for the year 2020 for population (most recent population data collected) and 2021 for water (most recent data collected). The output of this correlational analysis is:

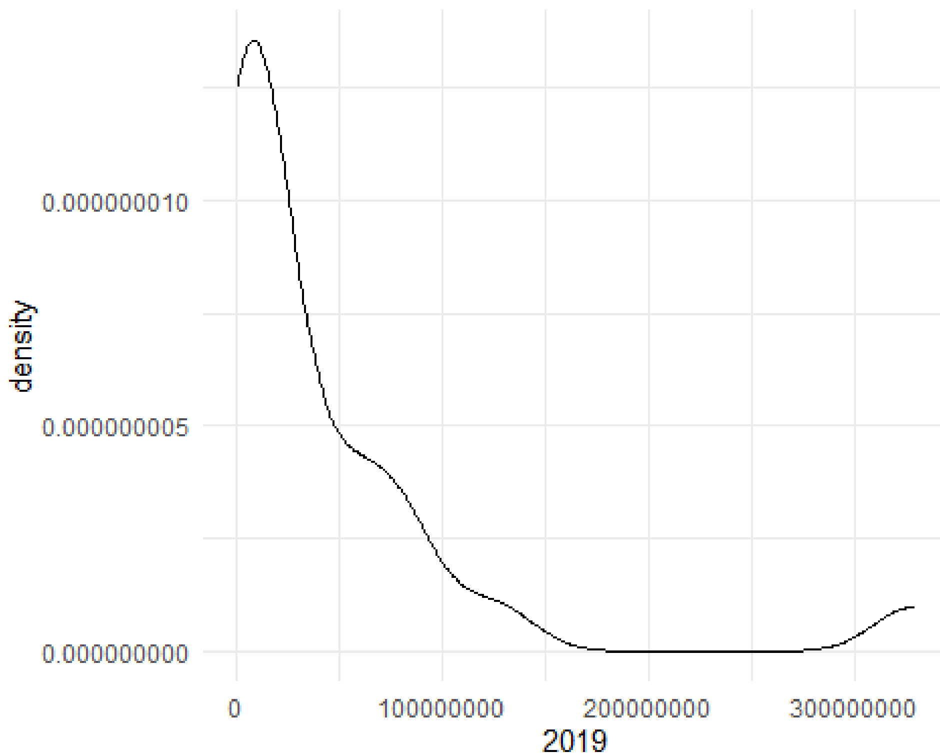
- Population and Renewable Water Per Capita:
  - Spearman's rho value of -0.44, indicating a negative, moderately strong relationship between a country's population and renewable water resources per capita.

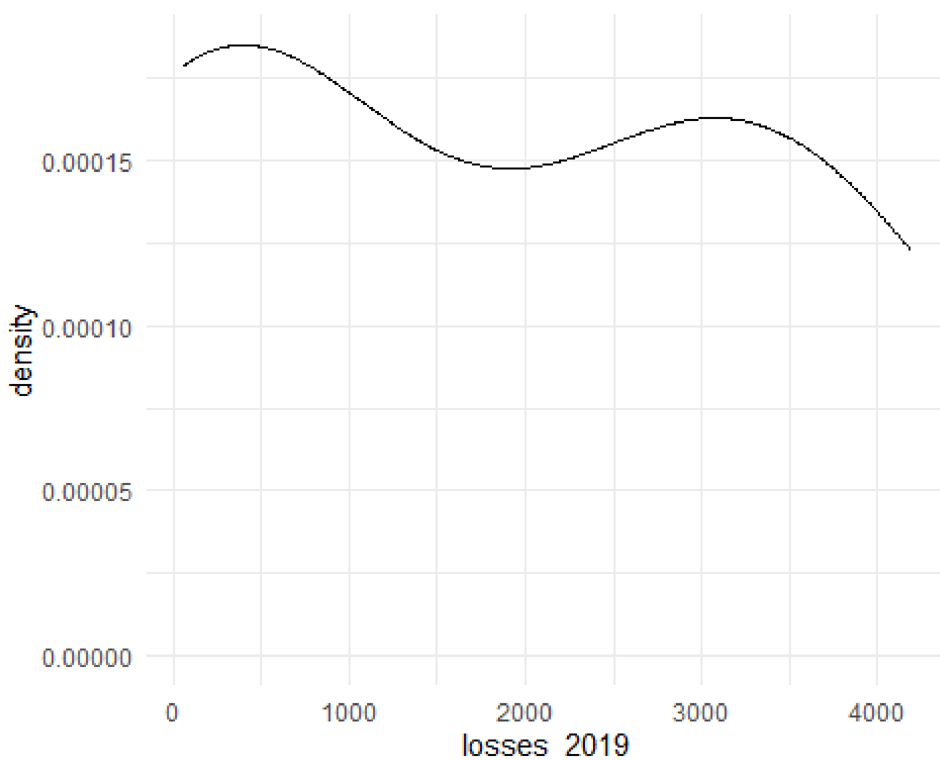
  - This means that as one of the variables goes up (population), the other goes down (renewable water), or vice versa.
  - It appears as though population moderately impacts the amount of renewable water a country has.
  - Because the p-value is less than 0.05, we can determine that the relationship is significant.
- Population and Renewable Water (Total):
  - Spearman's rho value of 0.58, indicating a positive, moderately strong relationship between a country's population and renewable water resources. * This means that as one of the variables goes up (population), the other goes down (renewable water), or vice versa.
  - It appears as though population moderately impacts the amount of renewable water a country has.
  - Because the p-value is less than 0.05, we can determine that the relationship is significant.

Population and Forest Correlation (Forest Losses and Forest Usage) (2019)



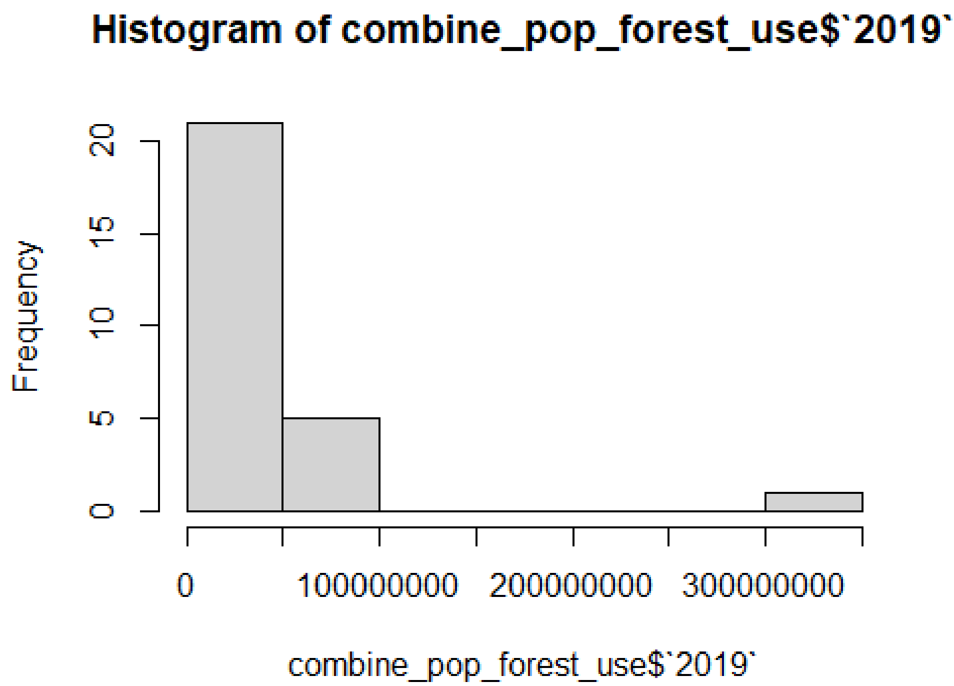


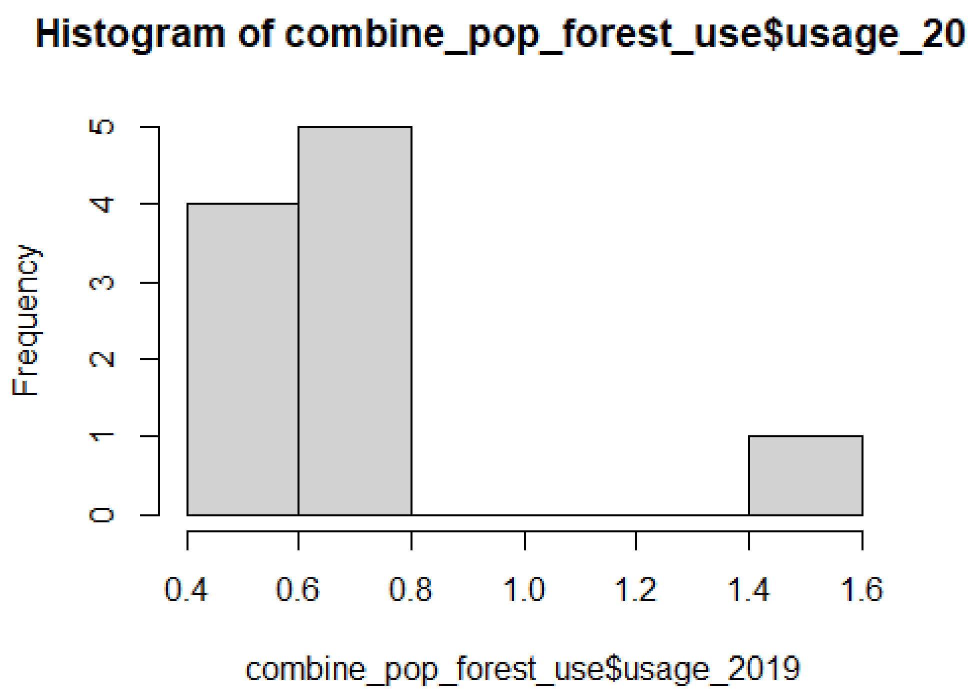


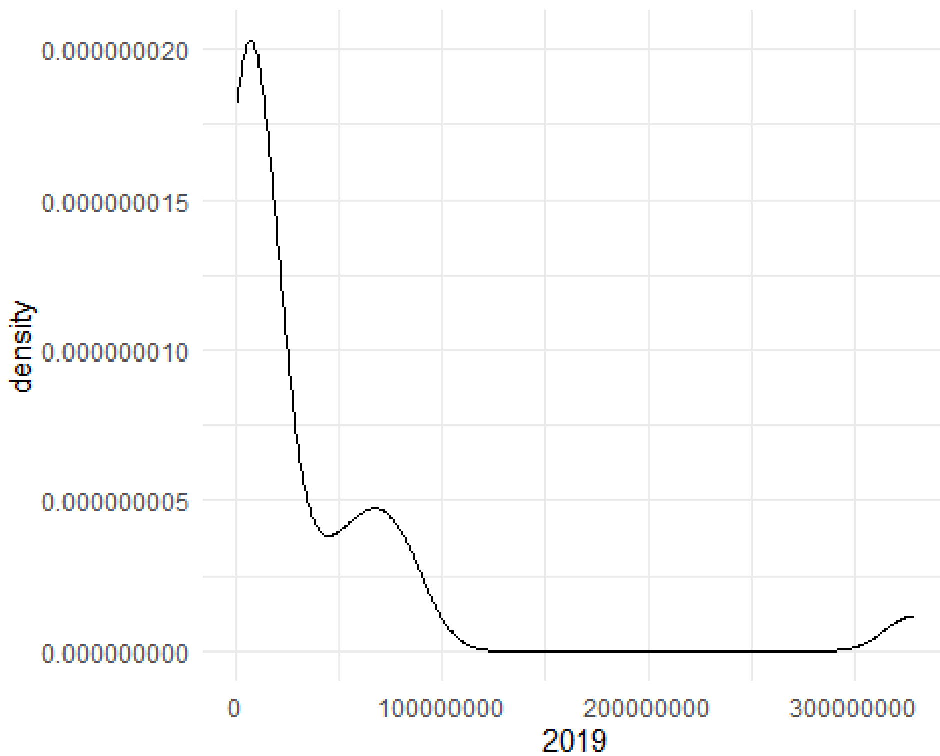


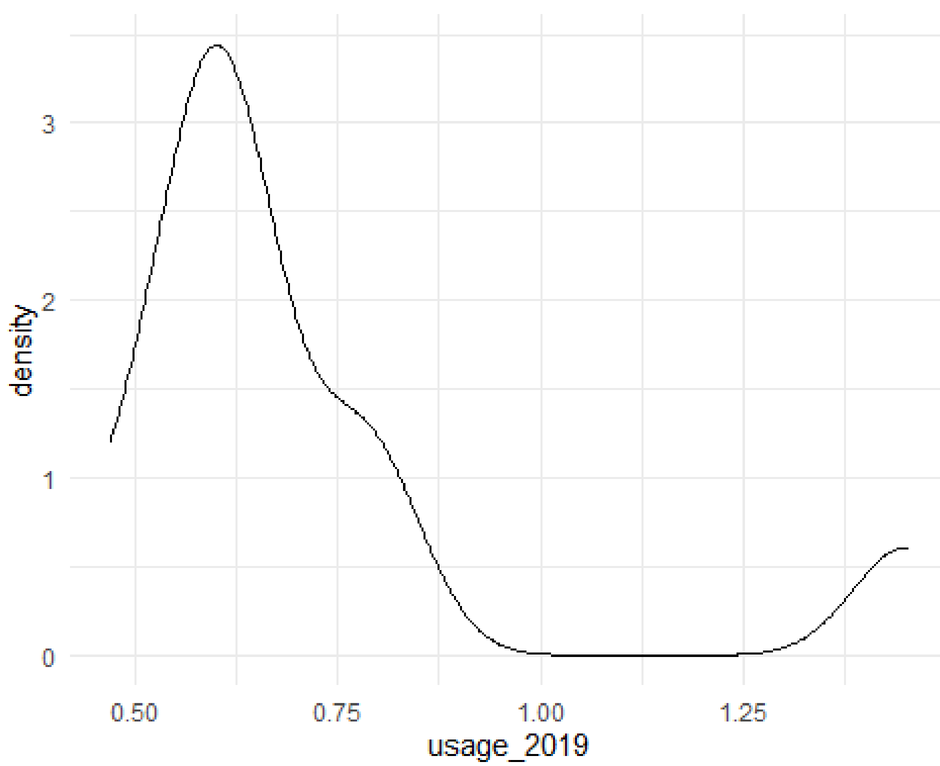

```
[1] 0.08571429


    Spearman's rank correlation rho
```

```
data:  combine_pop_forest_loss$`2019` and combine_pop_forest_loss$losses_2019
S = 32, p-value = 0.8717
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
0.08571429
```


Histogram of combine_pop_forest_use$`2019`
Frequency
0
5
10
15
20
0
100000000
200000000
300000000
combine_pop_forest_use$`2019`


Histogram of combine_pop_forest_use$usage_20
Frequency
0
1
2
3
4
5
0.4
0.6
0.8
1.0
1.2
1.4
1.6
combine_pop_forest_use$usage_2019





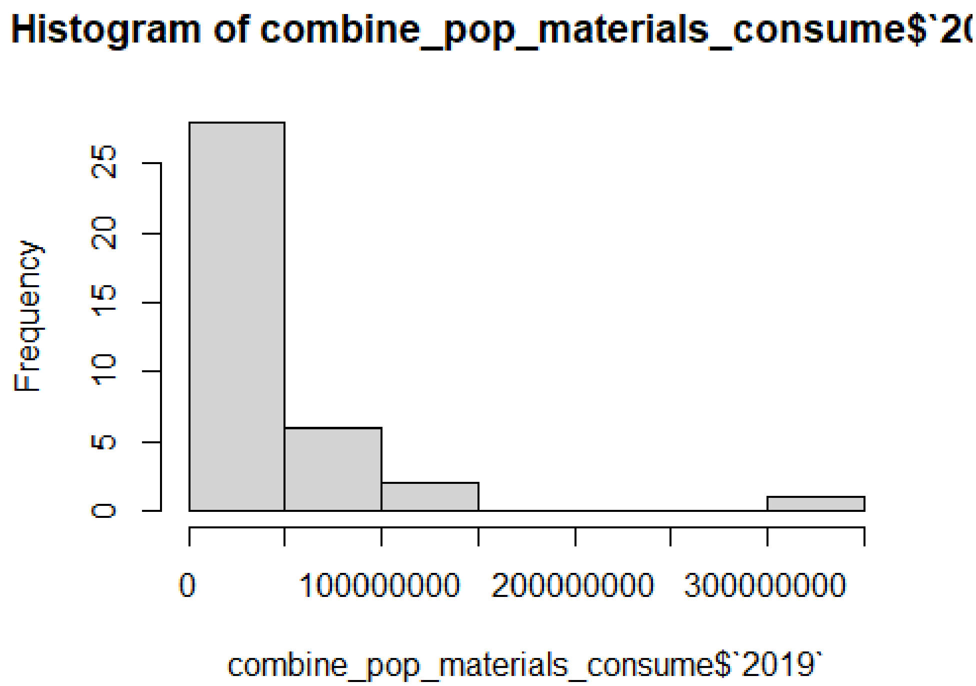

```
[1] -0.1878788


    Spearman's rank correlation rho
```

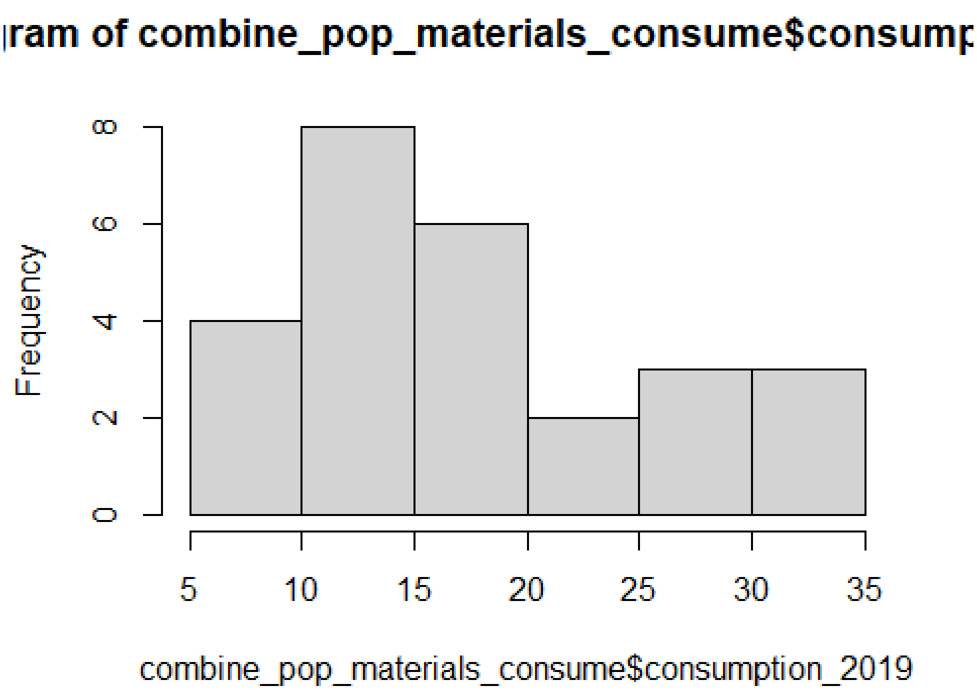
```
data:  combine_pop_forest_use$`2019` and combine_pop_forest_use$usage_2019
S = 196, p-value = 0.6032
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.1878788
```

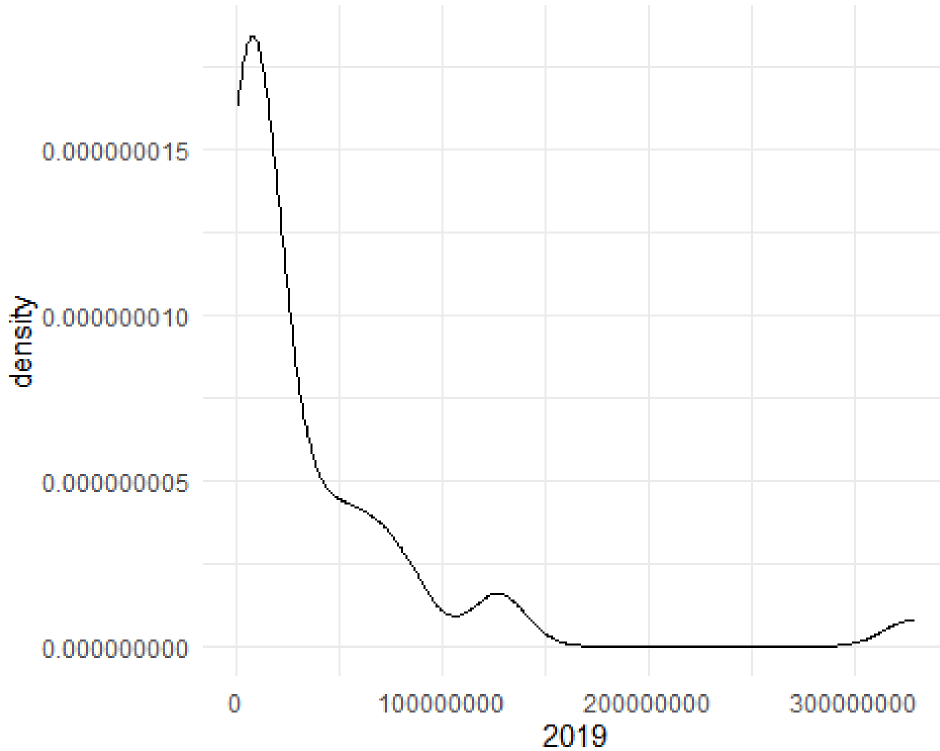
This test is running a correlation between the same countries present in both the population dataset and each of the forest datasets (forest losses and forest usage). The test is being run for the year 2019 for population and forest losses/forest usage, as 2019 is the most recent data collected for the forest datasets. Using 2019 measures for both will allow for a more accurate assessment and outcome. The output of this correlational analysis is:

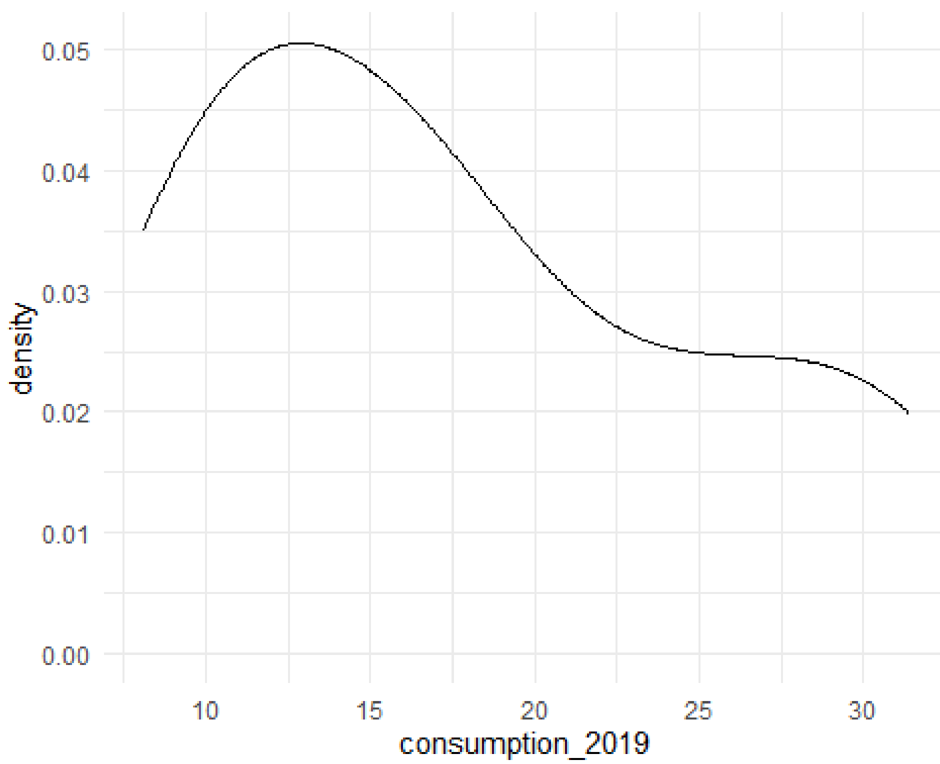
- Population and Forest Losses:
  - Spearman's rho value of 0.09, indicating a positive, weak relationship between a country's population and total forest losses for that country.
  - This means that as one of the variables goes up (population), the other also goes up (forest losses), or vice versa (as one goes down so does the other).
  - It appears as though population does not really impact the forest losses a country has experienced.
  - Because the p-value is NOT less than 0.05, we can determine that the relationship is not significant.
- Population and Forest Usage:
  - Spearman's rho value of -0.19, indicating a negative, weak relationship between a country's population and total forest resource usage for that country.
  - This means that as one of the variables goes up (population), the other goes down (forest usage), or vice versa.
  - It appears as though population does not really impact a country's forest usage.
  - Because the p-value is NOT less than 0.05, we can determine that the relationship is not significant.

Population and Materials Correlation (Material Consumption and Material Footprint) (2019 and 2017, Respectively)



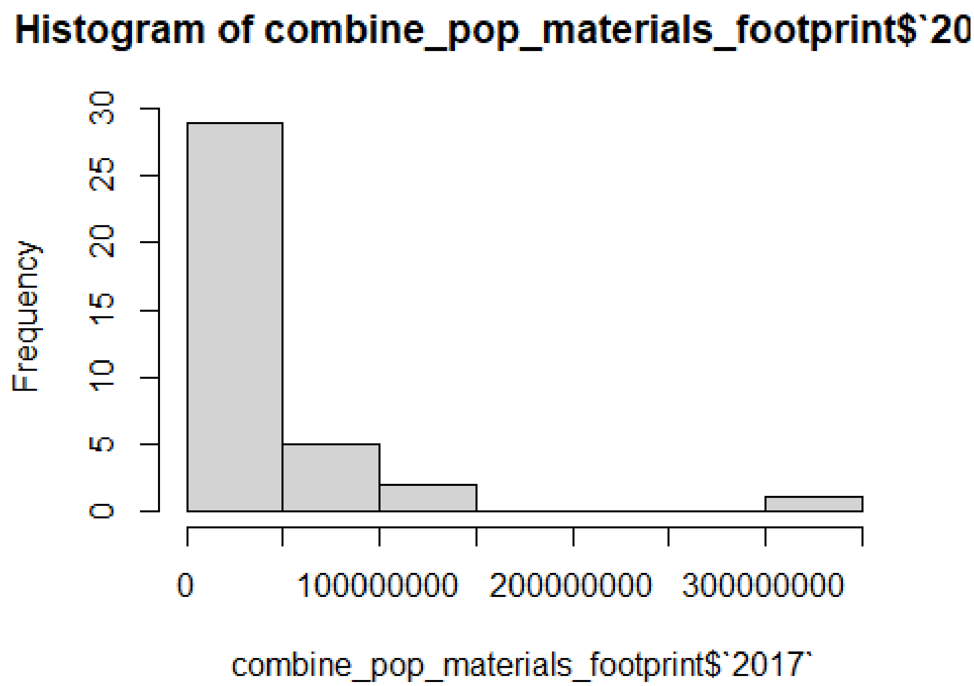


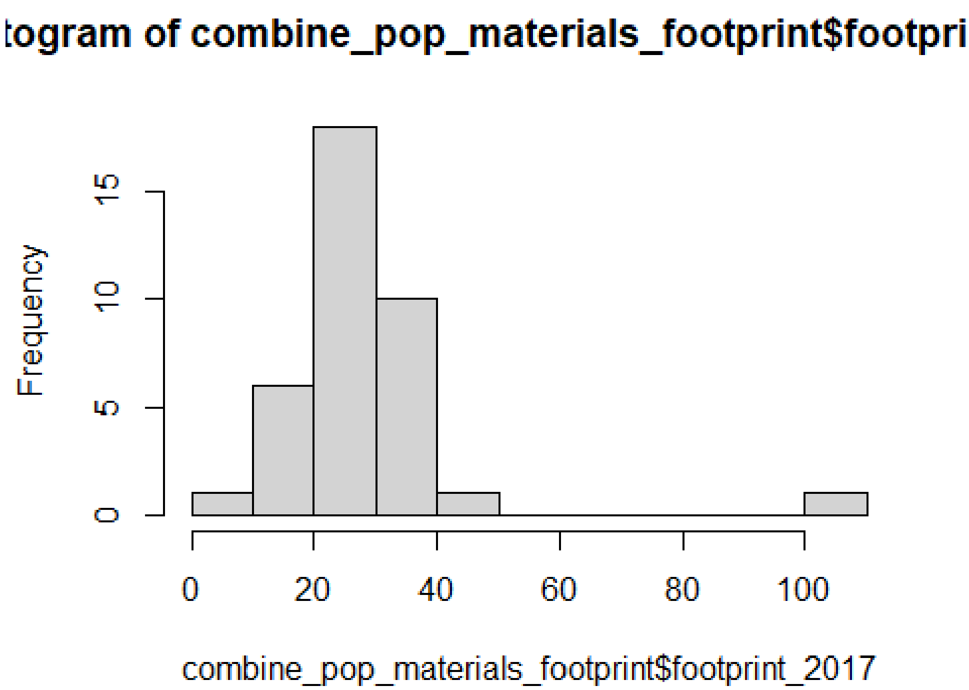


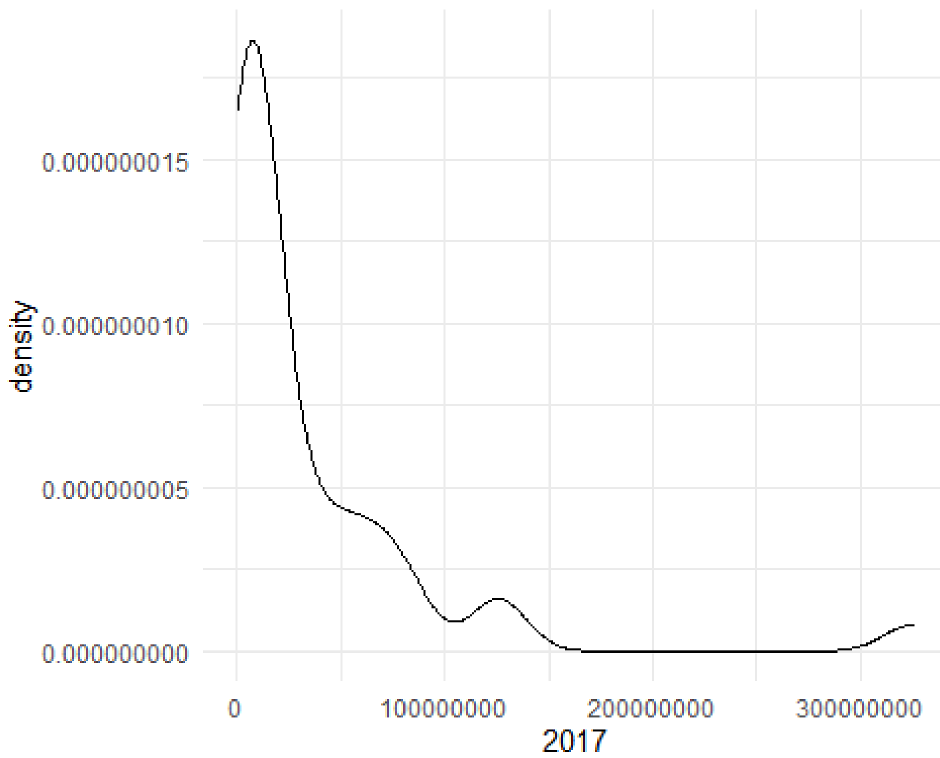


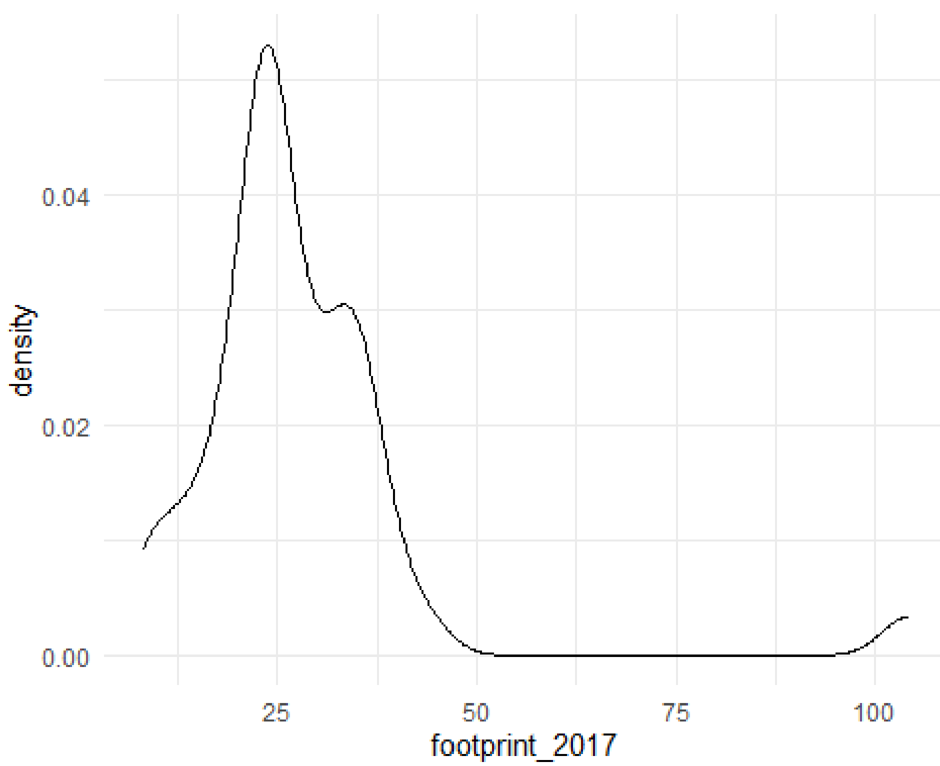

```
[1] -0.6300855


    Spearman's rank correlation rho
```

```
data:  combine_pop_materials_consume$`2019` and combine_pop_materials_consume
$consumption_2019
S = 4768, p-value = 0.0005611
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.6300855
```


Histogram of combine_pop_materials_footprint$`20
Frequency
0
5
10
15
20
25
30
0
100000000
200000000
300000000
combine_pop_materials_footprint$`2017`


togram of combine_pop_materials_footprint$footpri
Frequency
0
5
10
15
0
20
40
60
80
100
combine_pop_materials_footprint$footprint_2017






```
[1] -0.3558559


    Spearman's rank correlation rho
```

```
data:  combine_pop_materials_footprint$`2017` and combine_pop_materials_footp
rint$footprint_2017
S = 11438, p-value = 0.03065
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.3558559
```

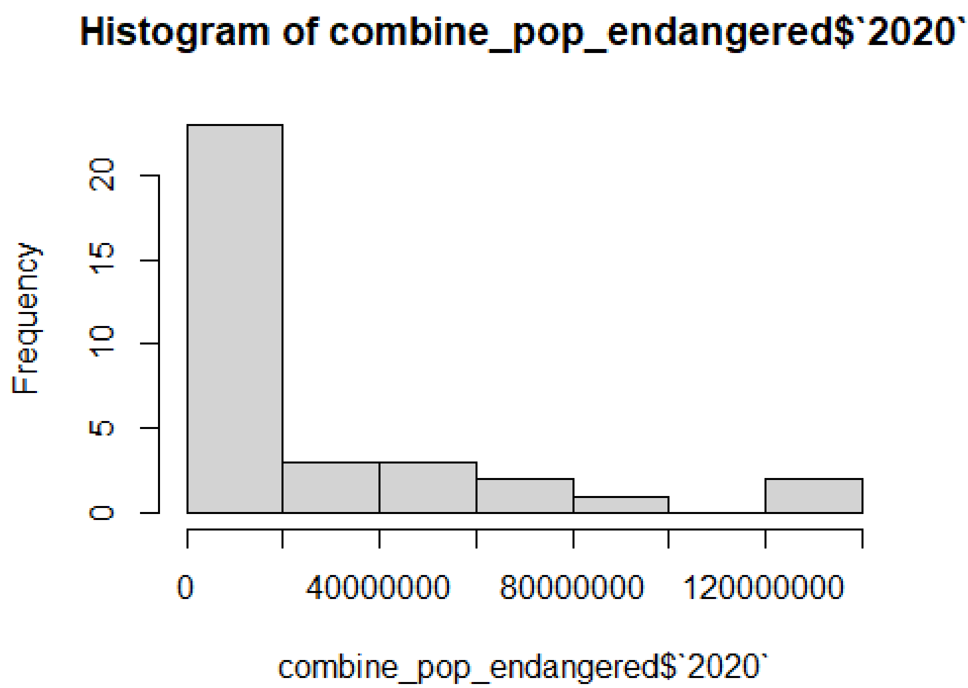
This test is running a correlation between the same countries present in both the population dataset and each of the material datasets (material consumption and material footprint). The test is being run for the year 2019 for population and material consumption (2019 was the most recent data collected for this variable), and 2017 for population and material footprint (2017 is the most recent data collected). Using the same yearly measures for both will allow for a more accurate assessment and outcome. The output of this correlational analysis is:

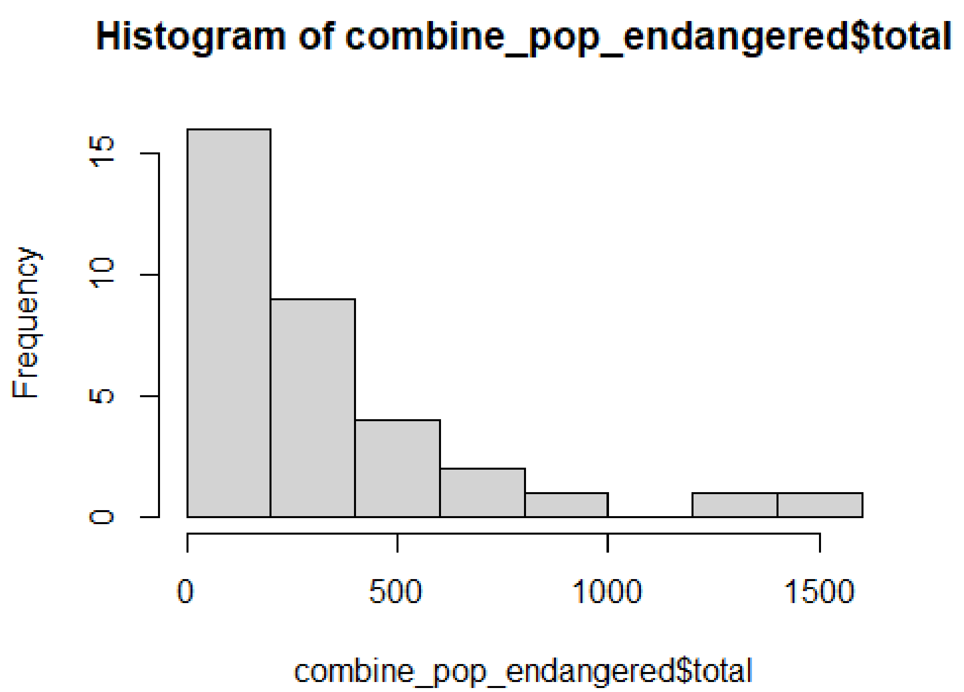
- Population and Material Consumption:
  - Spearman's rho value of -0.63, indicating a negative, stronger relationship between a country's population and material consumption per capita for that country.

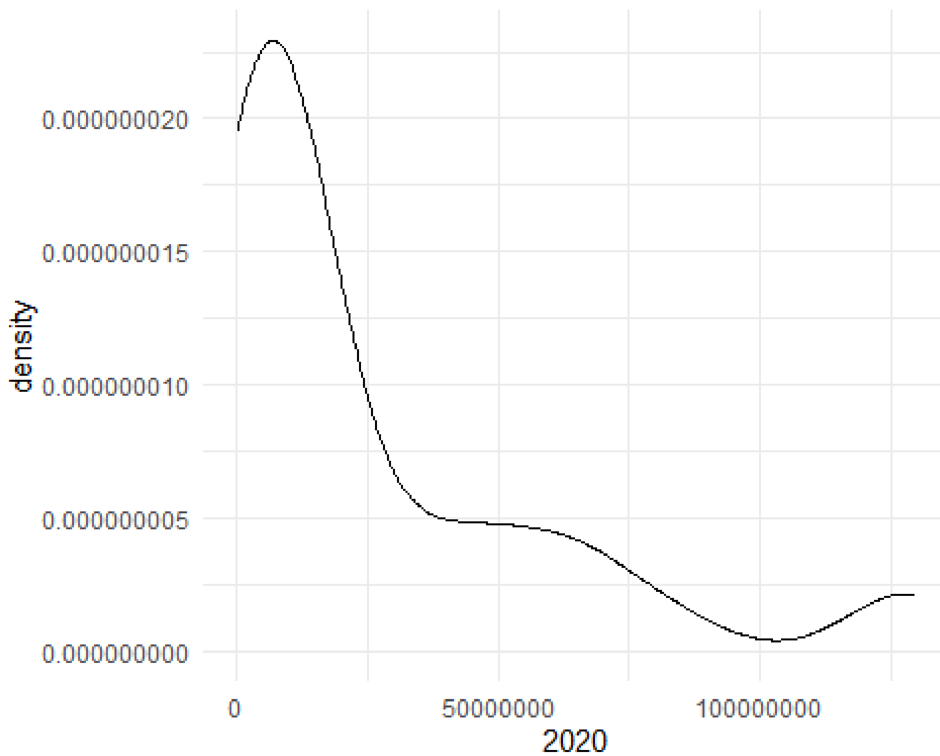
  - This means that as one of the variables goes up (population), the other goes down (material consumption), and vice versa.
  - It appears as though population does impact a country's material consumption.
  - Because the p-value is less than 0.05, we can determine that the relationship is significant.
- Population and Material Footprint:
  - Spearman's rho value of -0.36, indicating a negative, weaker relationship between a country's population and material footprint per capita for that country.

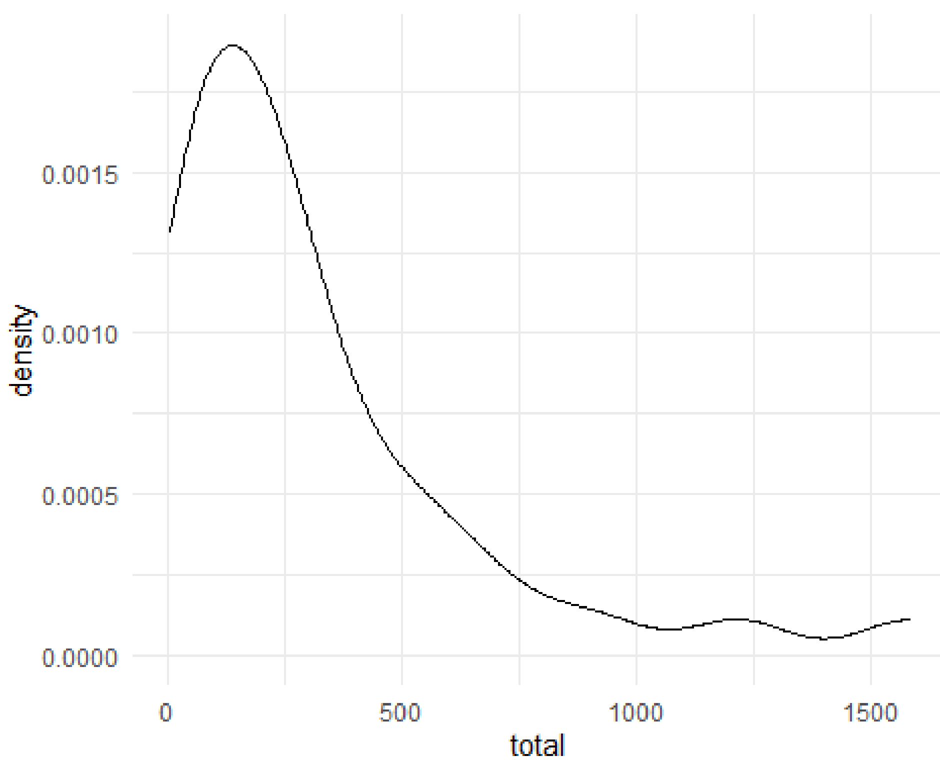
  - This means that as one of the variables goes up (population), the other goes down (material footprint), and vice versa.
  - It appears as though population may slightly (or perhaps, does not really) impact a country's material footprint.
  - Because the p-value is less than 0.05, we can determine that the relationship is significant.

Population and Critically Endangered Species Correlation (2020/2021)









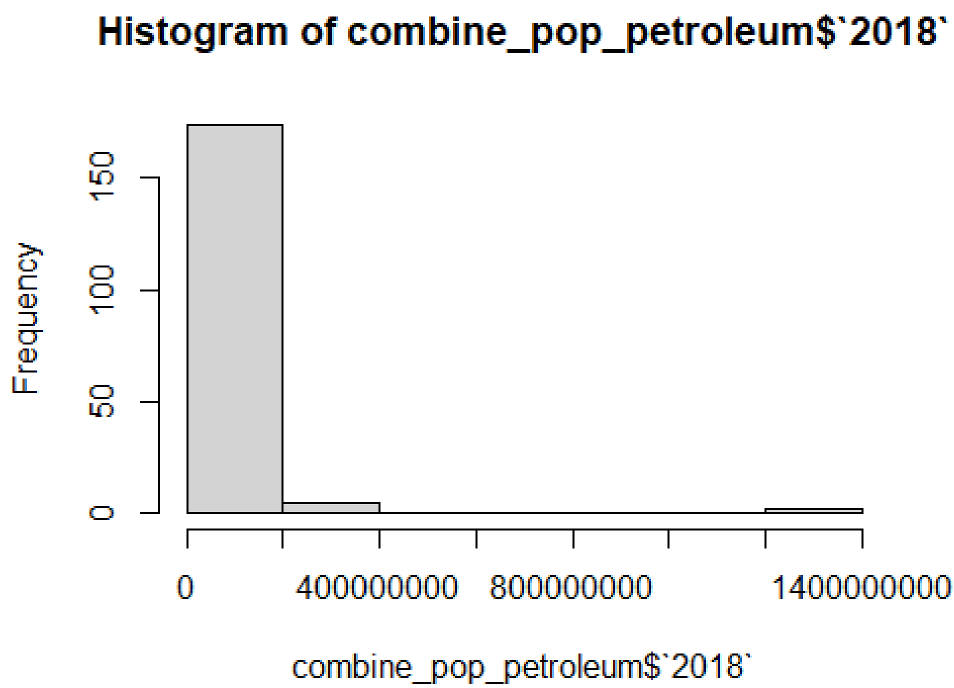

```
[1] 0.4582824


    Spearman's rank correlation rho
```

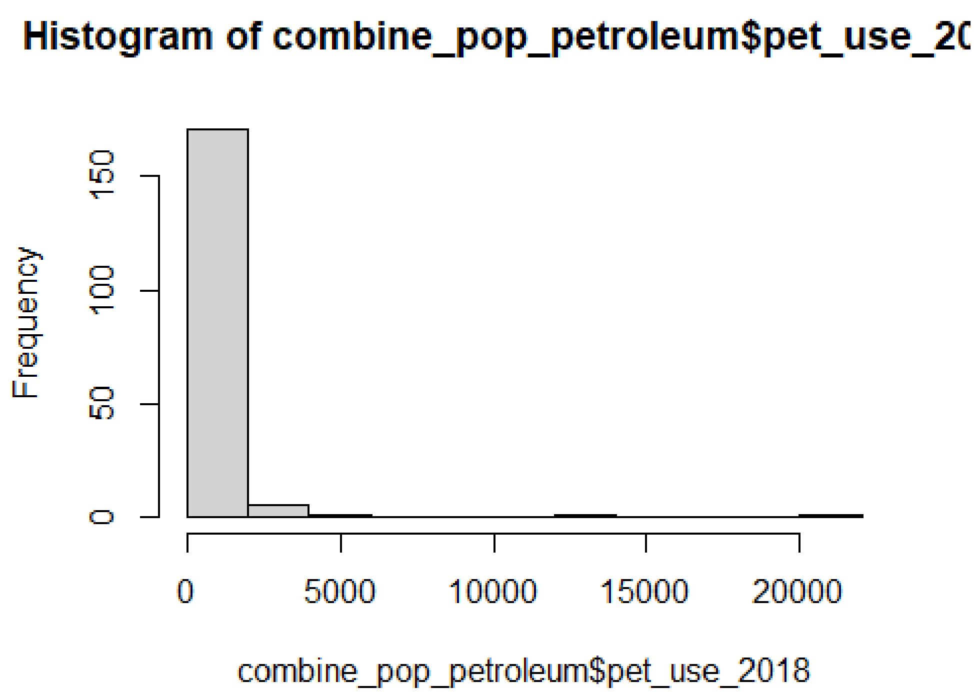
```
data:  combine_pop_endangered$`2020` and combine_pop_endangered$total
S = 3545.5, p-value = 0.006417
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.4582824
```

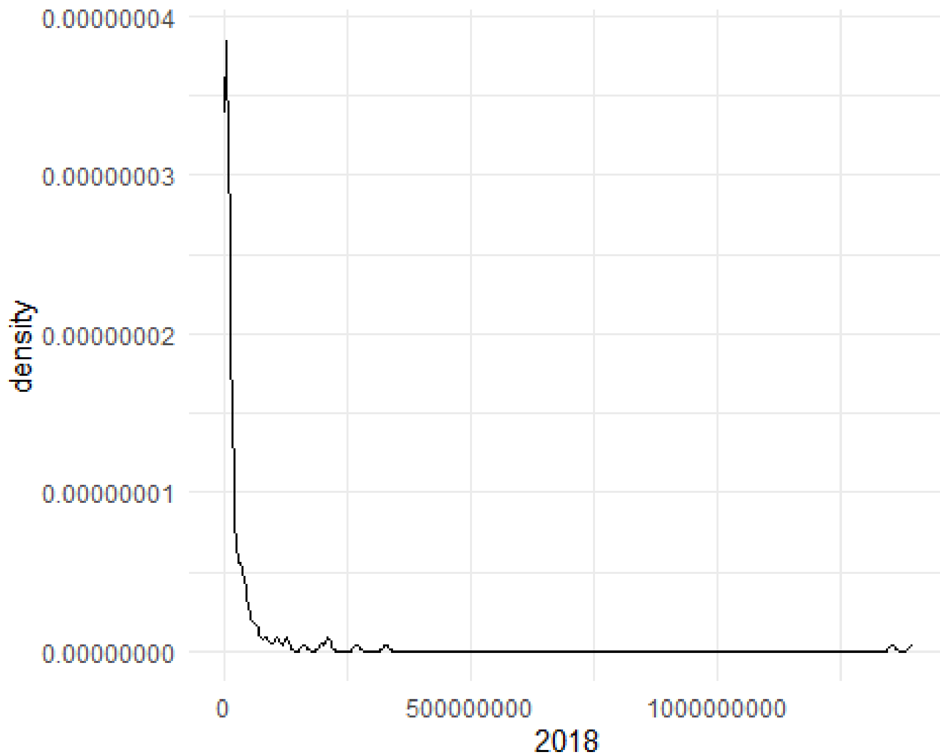
This test is running a correlation between the same countries present in both the population dataset and the critically endangered species dataset. The test is being run for the year 2020 for population and 2021 for critically endangered species (most recent and only data available). The output of this correlational analysis is:

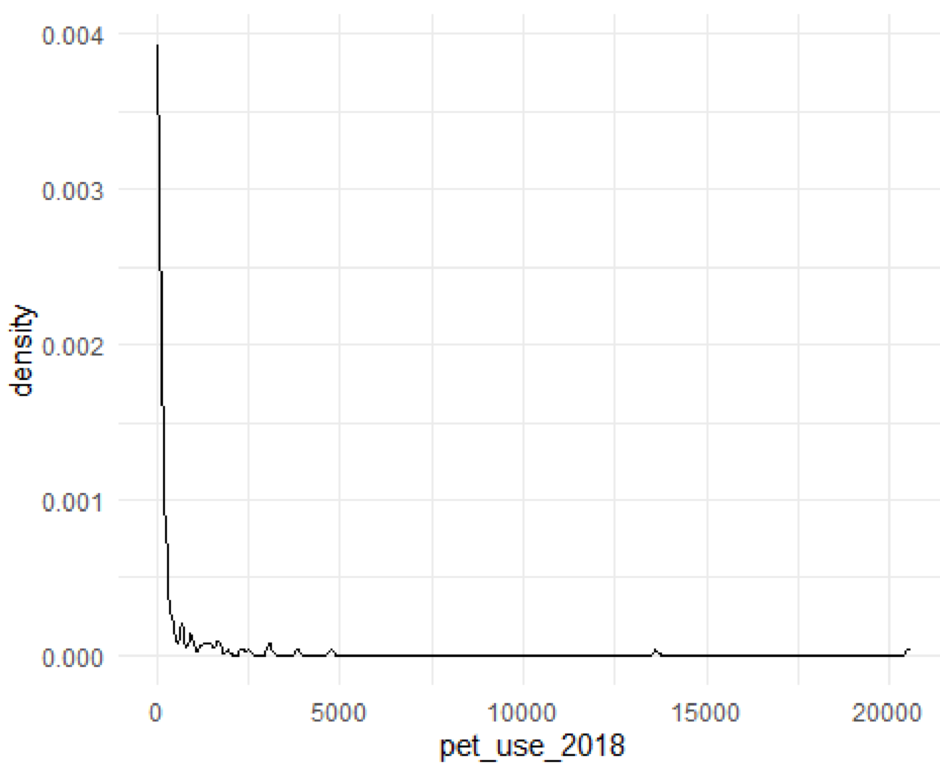
- Population and Critically Endangered Species:
  - Spearman's rho value of 0.46, indicating a positive, moderate relationship between a country's population and its total number of critically endangered species.
  - This means that as one of the variables goes up (population), the other also goes up (endangered species), or vice versa (as one goes down so does the other).
  - It appears as though population may slightly impact a country's total number of critically endangered species.
  - Because the p-value is less than 0.05, we can determine that the relationship is significant.

Population and Petroleum and Other Liquids Consumption (2018)









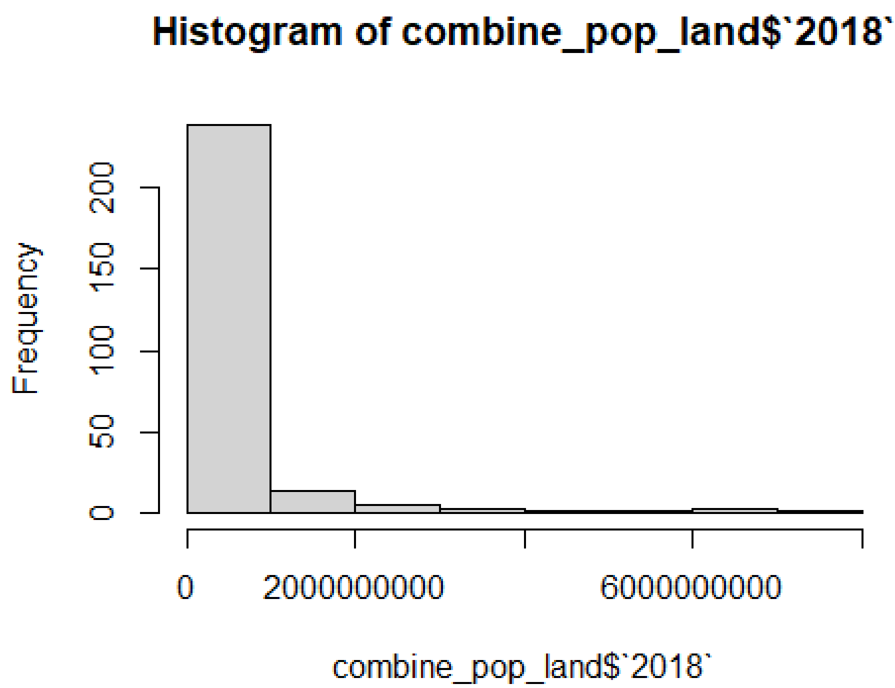

```
[1] 0.7293799


    Spearman's rank correlation rho
```

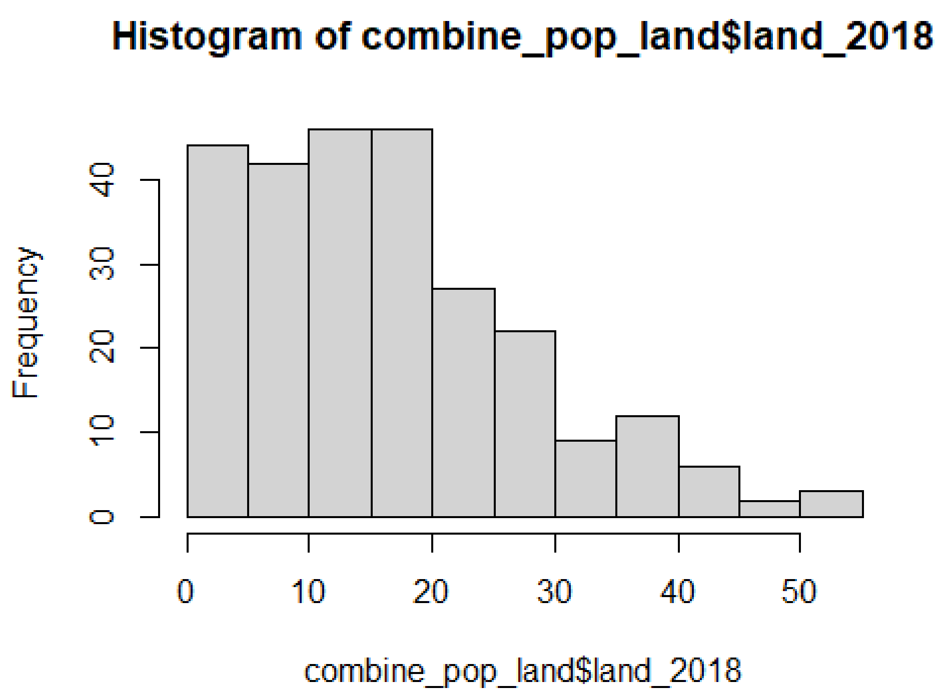
```
data:  combine_pop_petroleum$`2018` and combine_pop_petroleum$pet_use_2018
S = 254364, p-value < 0.00000000000000022
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.7293799
```

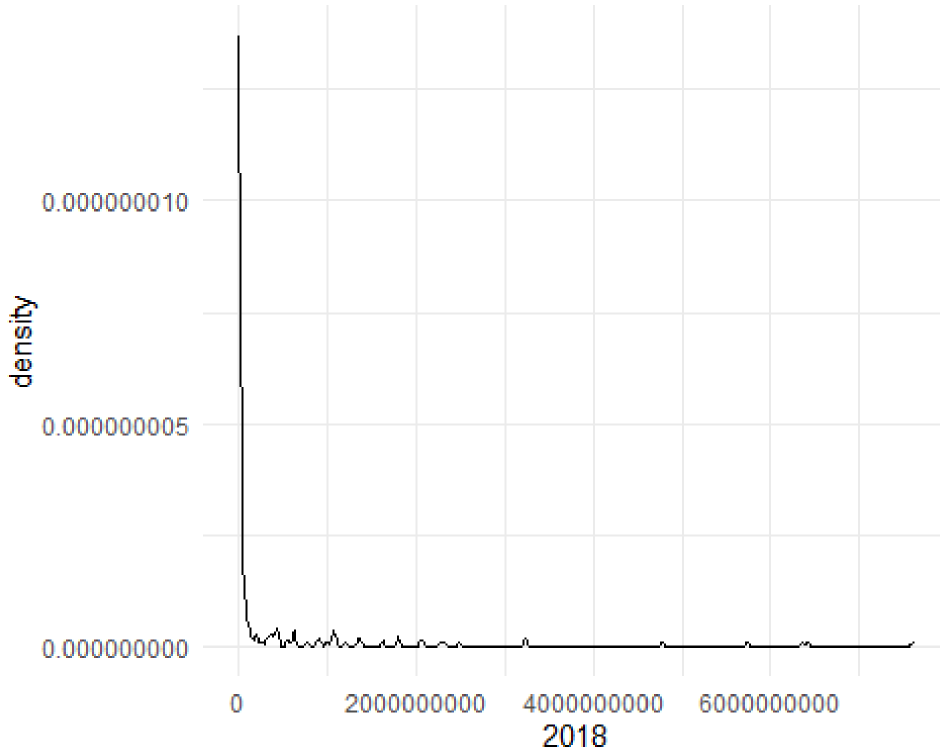
This test is running a correlation between the same countries present in both the population dataset and the petroleum and other liquids consumption dataset. The test is being run for the year 2018 for population and the petroleum dataset (most recent year of data available). Using 2018 measures for both will allow for a more accurate assessment and outcome. The output of this correlational analysis is:

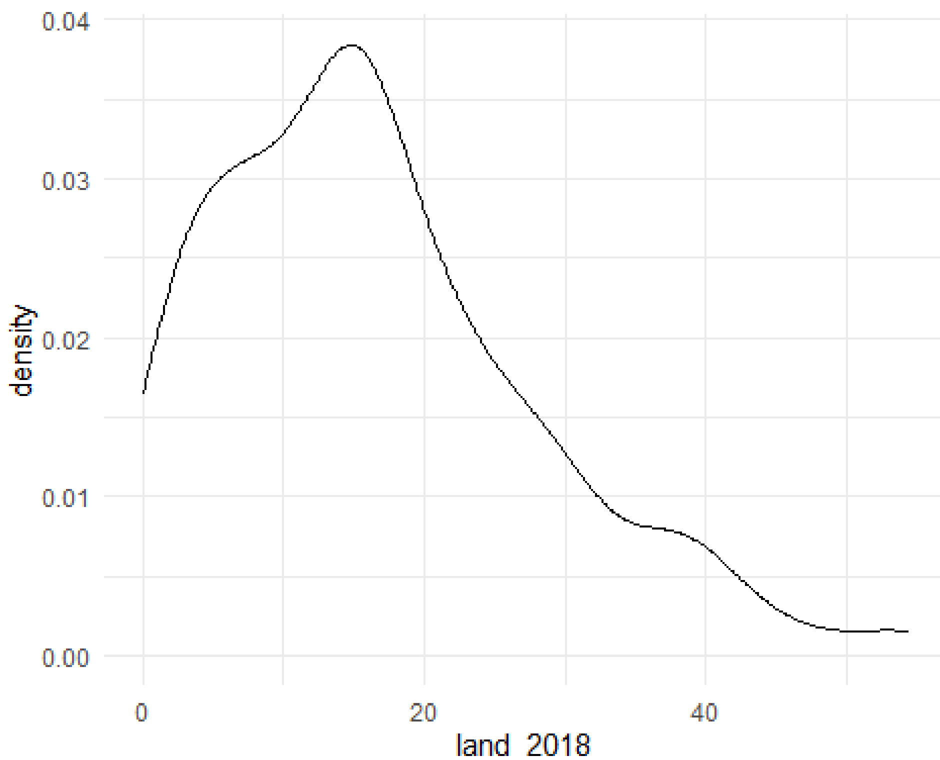
- Population and Petroleum/Other Liquids Consumption:
  - Spearman's rho value of 0.73, indicating a positive, strong relationship between a country's population and its consumption of petroleum and other related liquids.
  - This means that as one of the variables goes up (population), the other also goes up (petroleum consumption), or vice versa (as one goes down so does the other).
  - It appears as though population does have an impact on a country's total amount of petroleum and other liquids consumption.
  - Because the p-value is less than 0.05, we can determine that the relationship is significant (extremely so, in this case).

Population and Protected Terrestrial Land (2018)









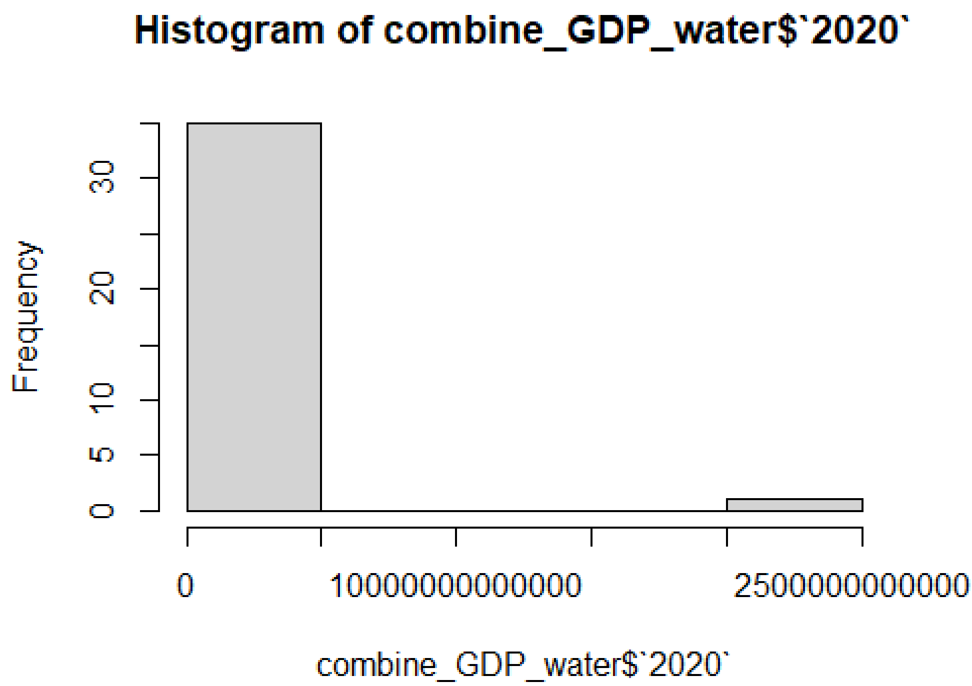

```
[1] -0.01791939


    Spearman's rank correlation rho
```

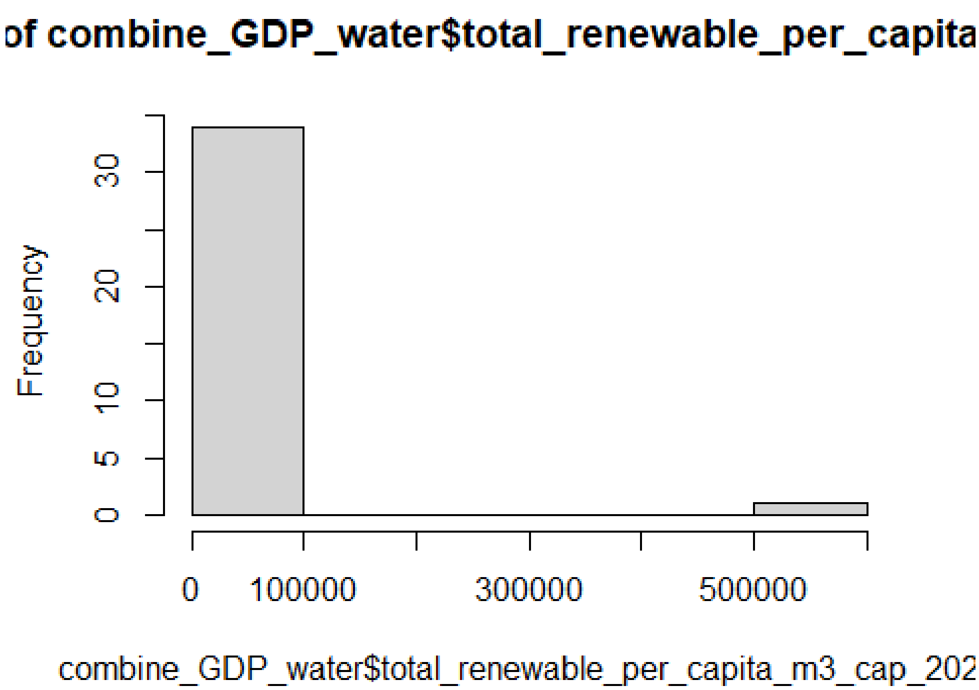
```
data:  combine_pop_land$`2018` and combine_pop_land$land_2018
S = 2913498, p-value = 0.7745
alternative hypothesis: true rho is not equal to 0
sample estimates:
        rho
-0.01791939
```

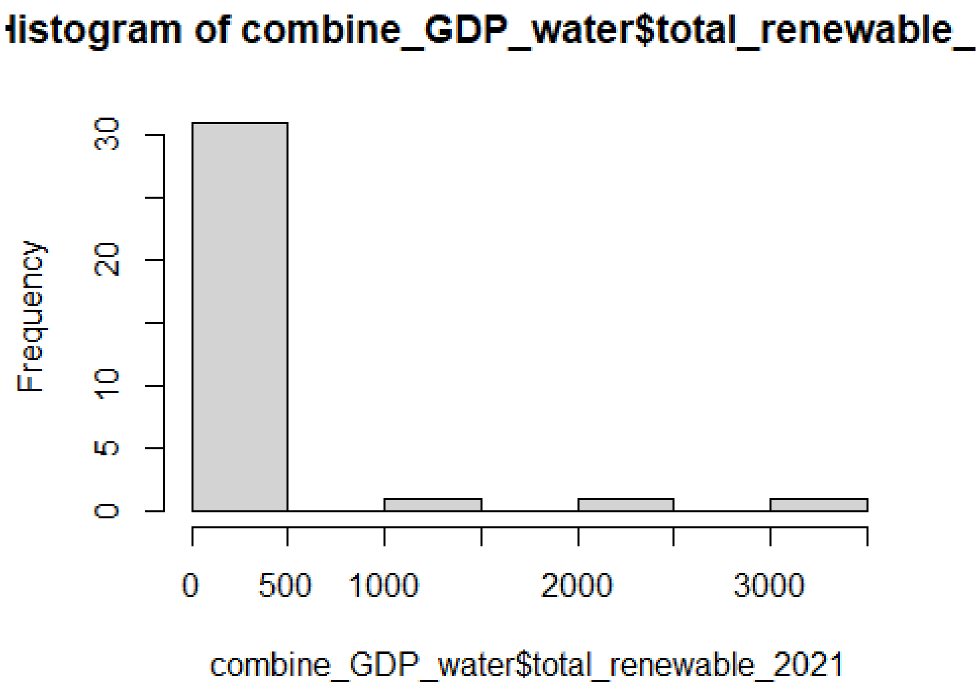
This test is running a correlation between the same countries present in both the population dataset and the protected land dataset. The test is being run for the year 2018 for population and the protected land dataset (most recent year of data available). Using 2018 measures for both will allow for a more accurate assessment and outcome. The output of this correlational analysis is:

- Population and Protected Terrestrial Land:
  - Spearman’s rho value of -0.02, indicating a negative, weak relationship between a country’s population and its percentage of protected land.
  - This means that as one of the variables goes up (population), the other goes down (protected land), and vice versa.
  - It appears as though population does not have an impact on a country’s total percentage of protected terrestrial land.
  - Because the p-value is NOT less than 0.05, we can determine that the relationship is not significant.

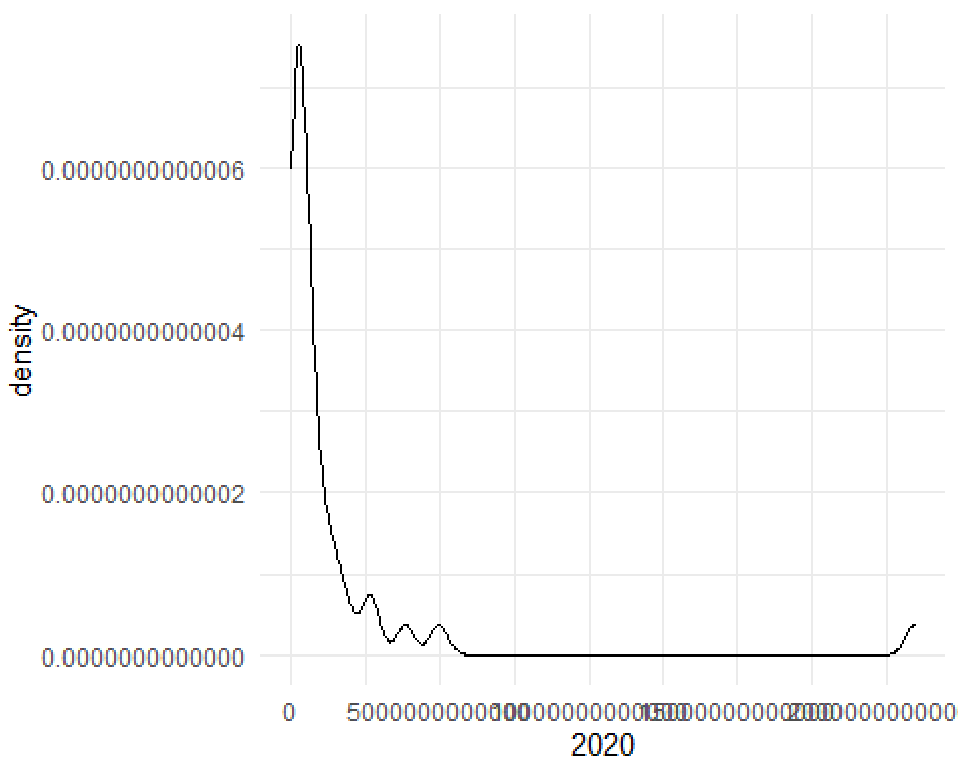
GDP and Water Correlation (2020/2021)



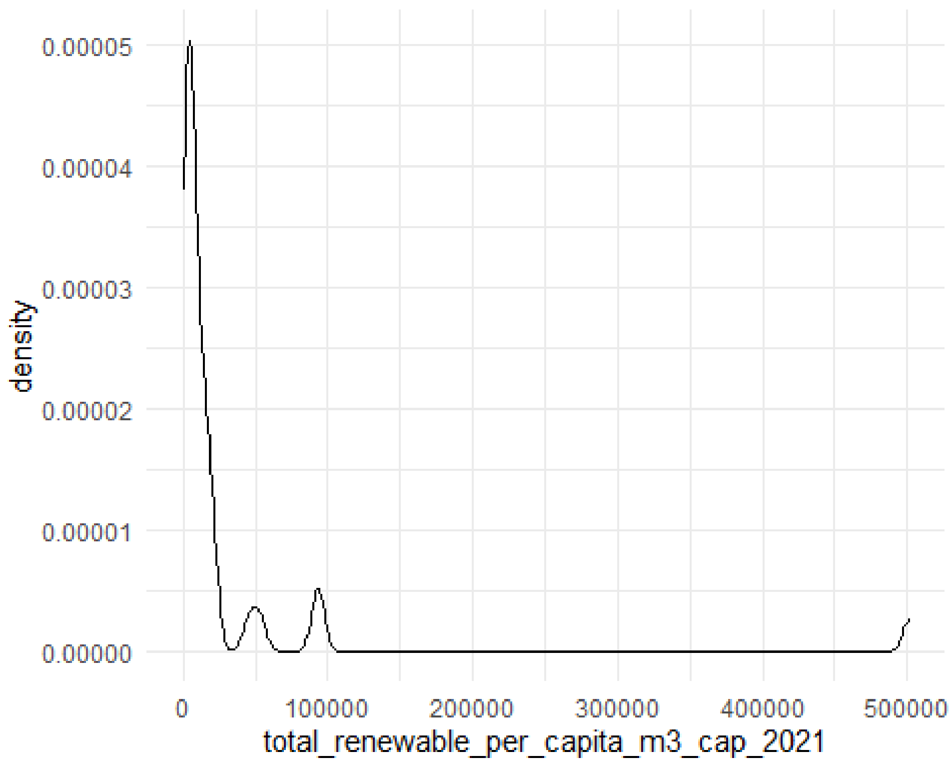


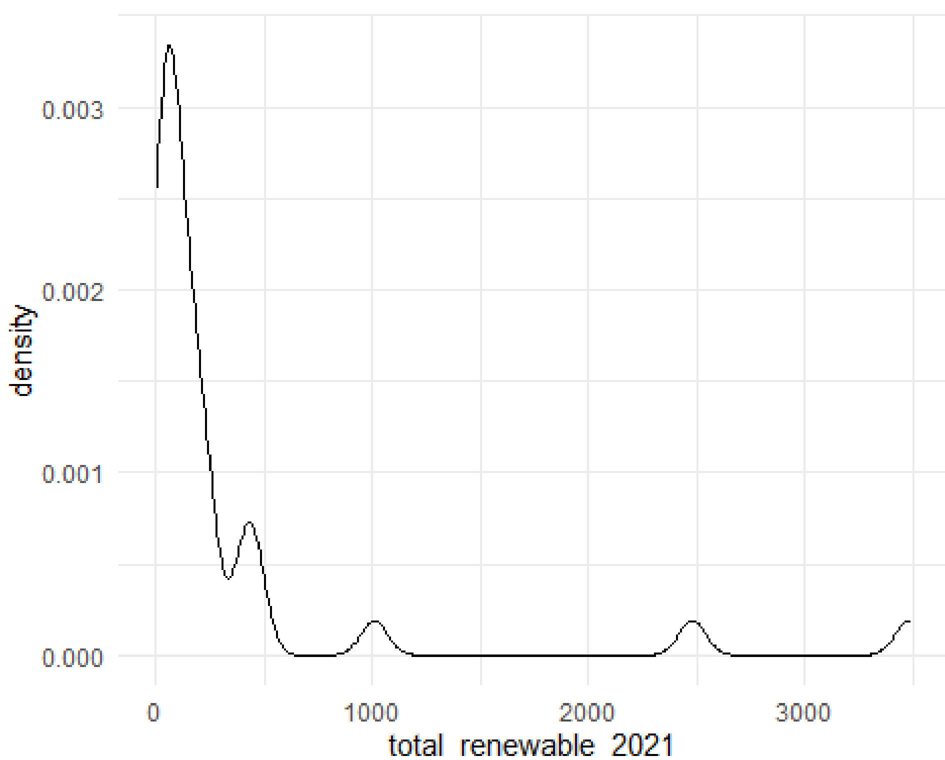


listogram of combine_GDP_water$total_renewable_
Frequency
0
5
10
20
30
0
500
1000
2000
3000
combine_GDP_water$total_renewable_2021



density
0.0000000000006
0.0000000000004
0.0000000000002
0.0000000000000
2020





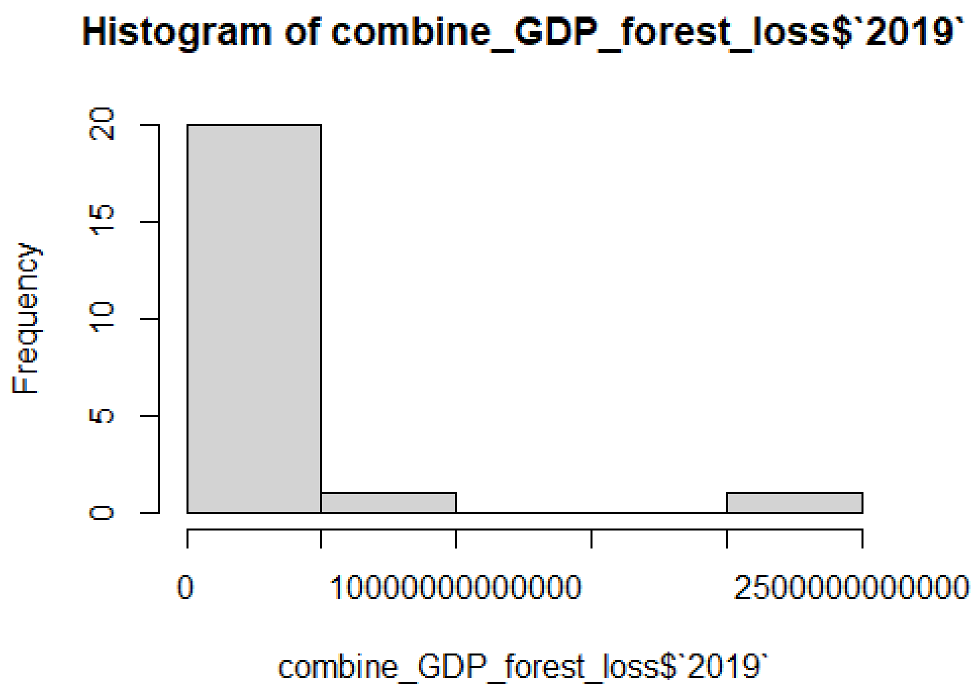

```
[1] -0.3859944

[1] 0.5399954
```

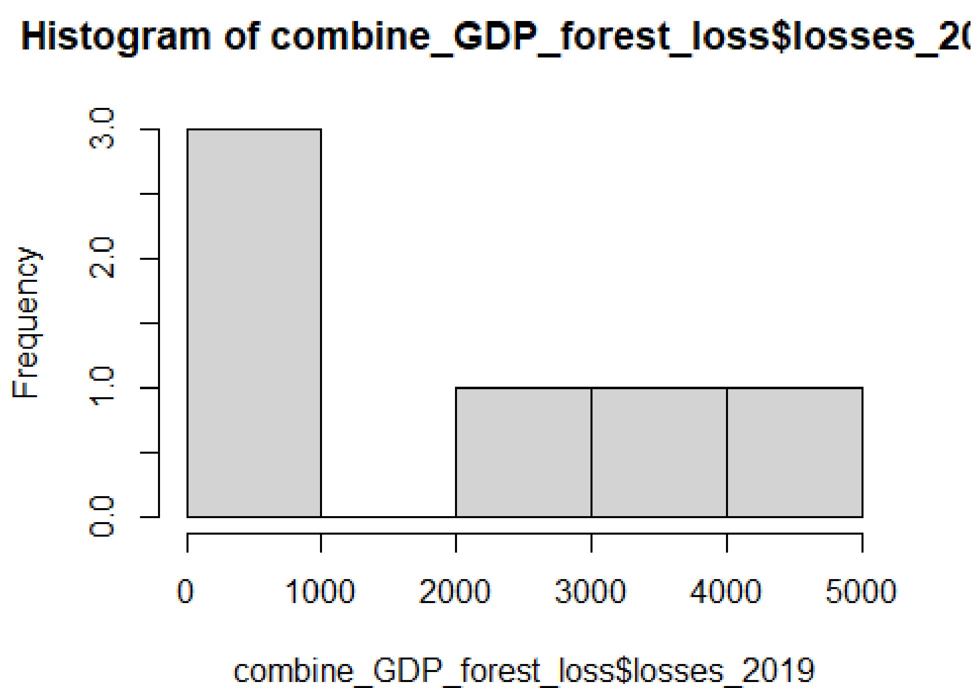
```
    Spearman's rank correlation rho

data:  combine_GDP_water$`2020` and combine_GDP_water$total_renewable_per_cap
ita_m3_cap_2021
S = 9896, p-value = 0.02201
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.3859944


    Spearman's rank correlation rho

data:  combine_GDP_water$`2020` and combine_GDP_water$total_renewable_2021
S = 3010.7, p-value = 0.0009797
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.5399954
```

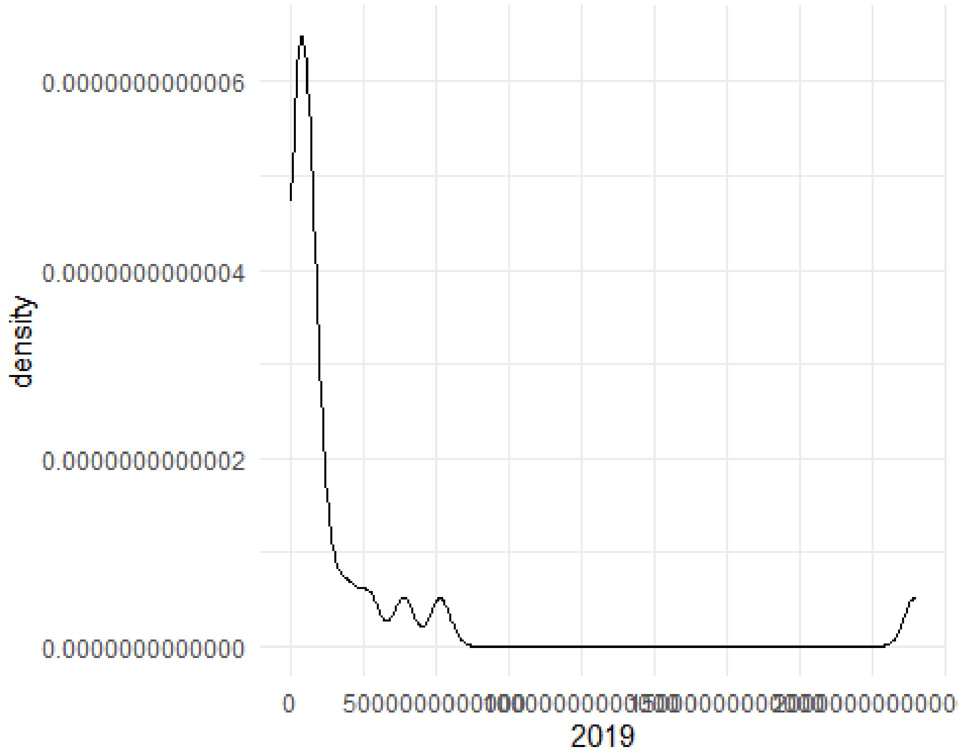
This test is running a correlation between the same countries present in both the GDP dataset and the water renewables dataset. The test is being run for the year 2020 for GDP (most recent GDP data collected) and 2021 for water (most recent data collected). The output of this correlational analysis is:

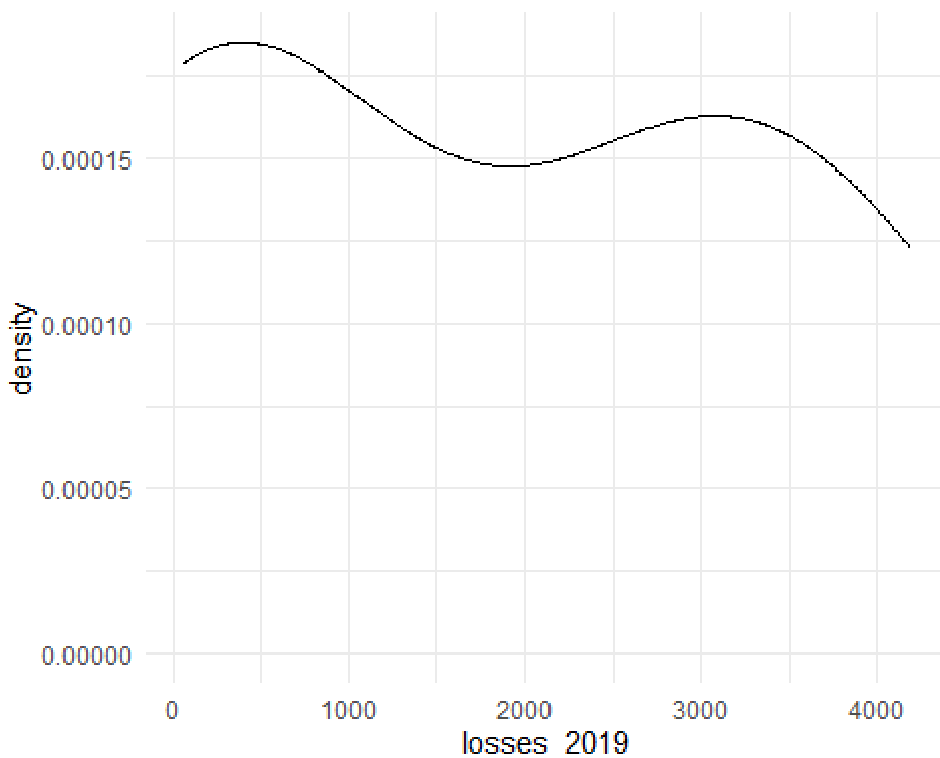
- GDP and Renewable Water Per Capita:
  - Spearman's rho value of -0.39, indicating a negative, moderate relationship between a country's GDP and renewable water resources per capita.
  - This means that as one of the variables goes up (GDP), the other goes down (renewable water), or vice versa.
  - It appears as though GDP may moderately impacts the amount of renewable water a country has.
  - Because the p-value is less than 0.05, we can determine that the relationship is significant.
- GDP and Renewable Water (Total):
  - Spearman's rho value of 0.54, indicating a positive, moderately strong relationship between a country's GDP and renewable water resources.
  - This means that as one of the variables goes up (GDP), the other goes down (renewable water), or vice versa.
  - It appears as though GDP moderately impacts the amount of renewable water a country has.
  - Because the p-value is less than 0.05, we can determine that the relationship is significant.

GDP and Forest Correlation (Forest Losses and Forest Usage) (2019)



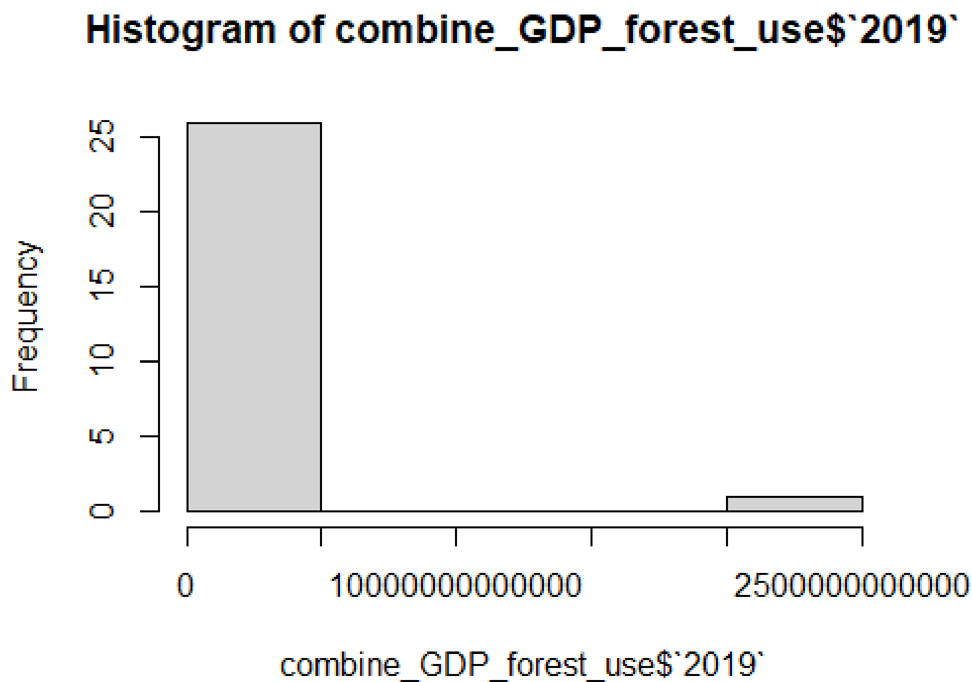


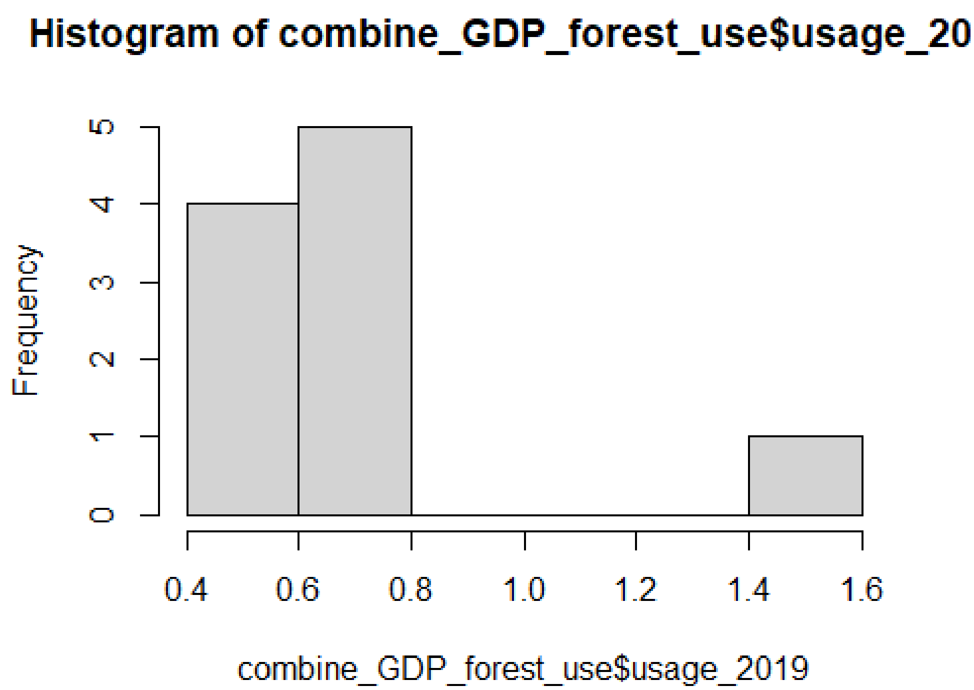


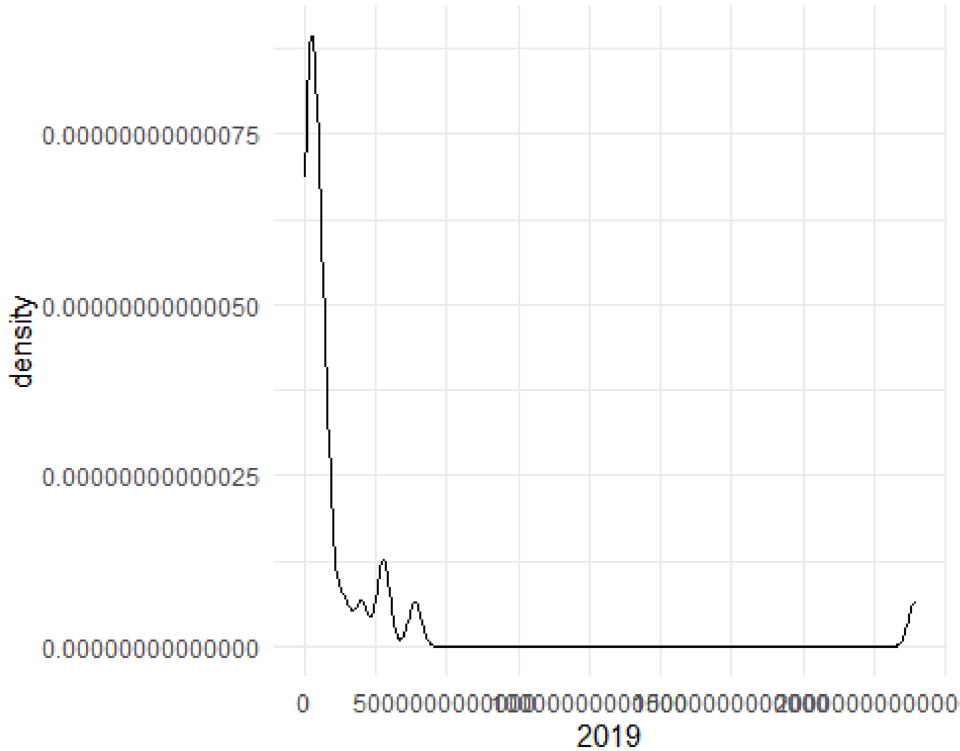


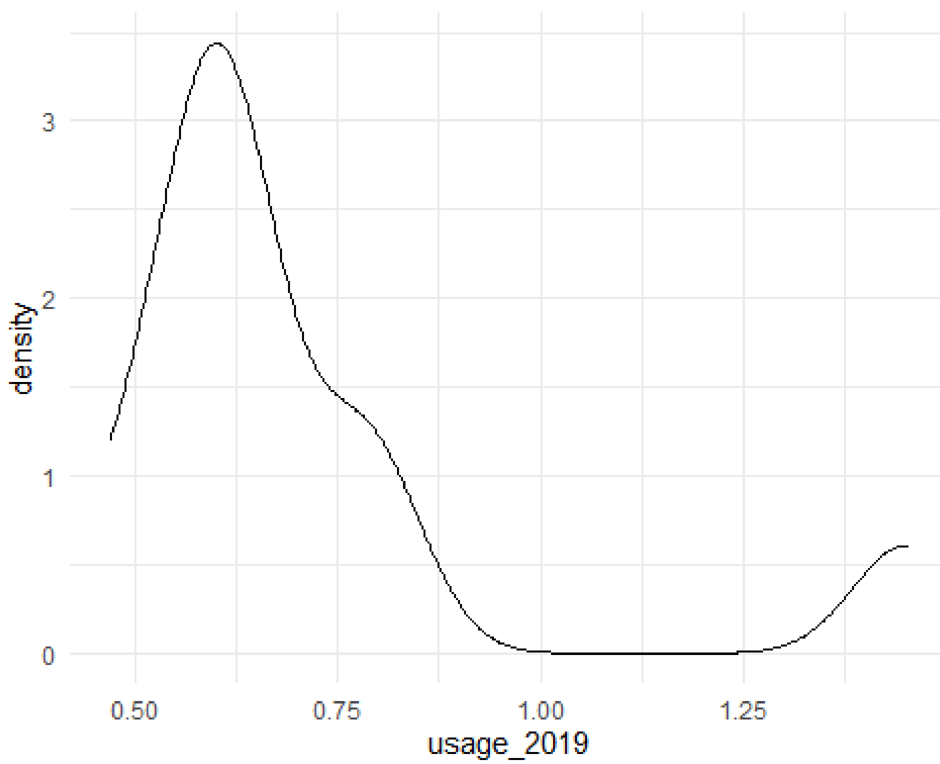

```
[1] -0.08571429


    Spearman's rank correlation rho
```

```
data:  combine_GDP_forest_loss$`2019` and combine_GDP_forest_loss$losses_2019
S = 38, p-value = 0.8717
alternative hypothesis: true rho is not equal to 0
sample estimates:
        rho 
-0.08571429
```


Histogram of combine_GDP_forest_use$`2019`
Frequency
0
5
10
15
20
25
0
10000000000000
25000000000000
combine_GDP_forest_use$`2019`


Histogram of combine_GDP_forest_use$usage_20
Frequency
0
1
2
3
4
5
0.4
0.6
0.8
1.0
1.2
1.4
1.6
combine_GDP_forest_use$usage_2019





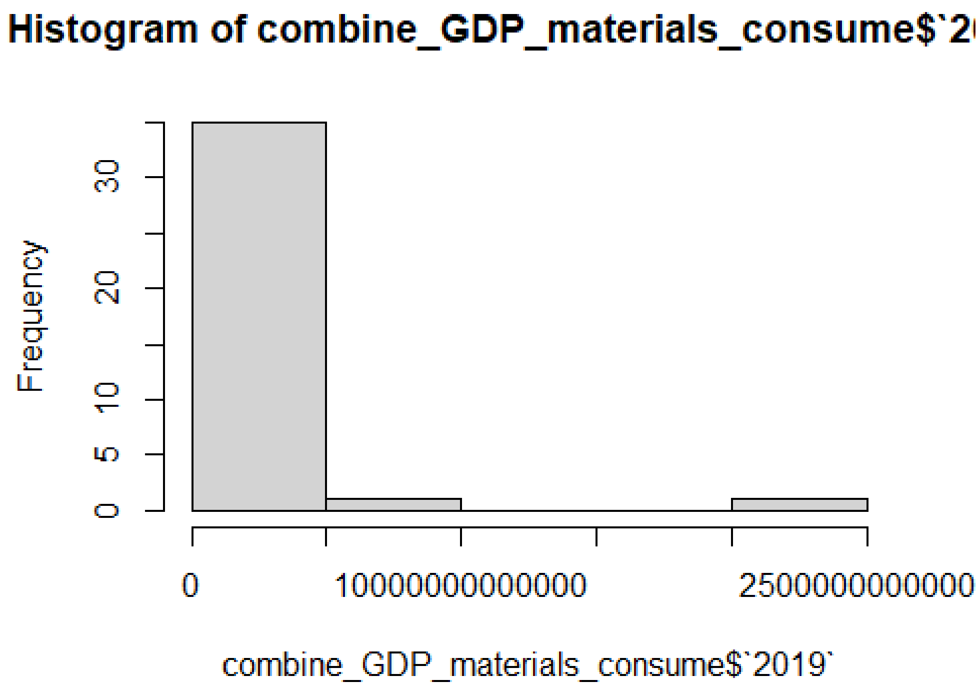

```
[1] -0.4787879


    Spearman's rank correlation rho
```

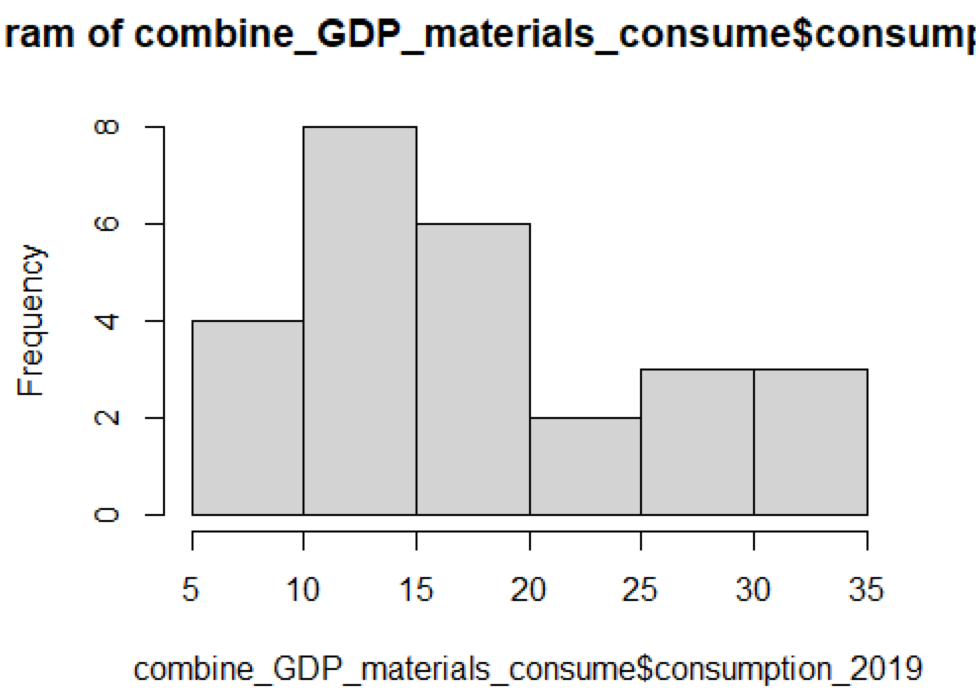
```
data:  combine_GDP_forest_use$`2019` and combine_GDP_forest_use$usage_2019
S = 244, p-value = 0.1615
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.4787879
```

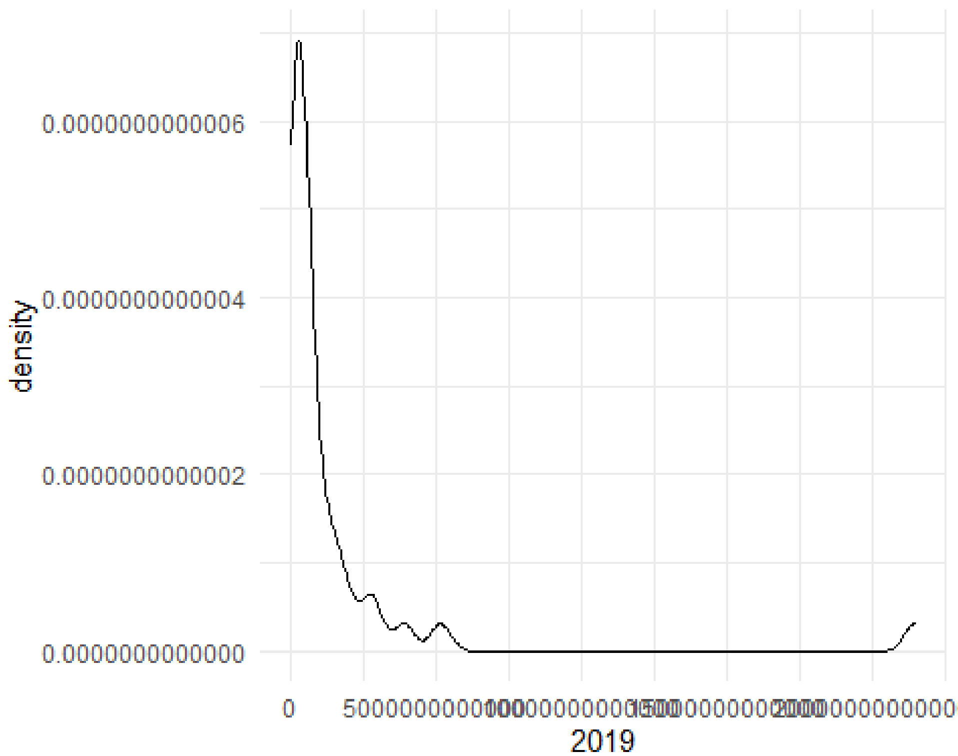
This test is running a correlation between the same countries present in both the GDP dataset and each of the forest datasets (forest losses and forest usage). The test is being run for the year 2019 for GDP and forest losses/forest usage, as 2019 is the most recent data collected for the forest datasets. Using 2019 measures for both will allow for a more accurate assessment and outcome. The output of this correlational analysis is:

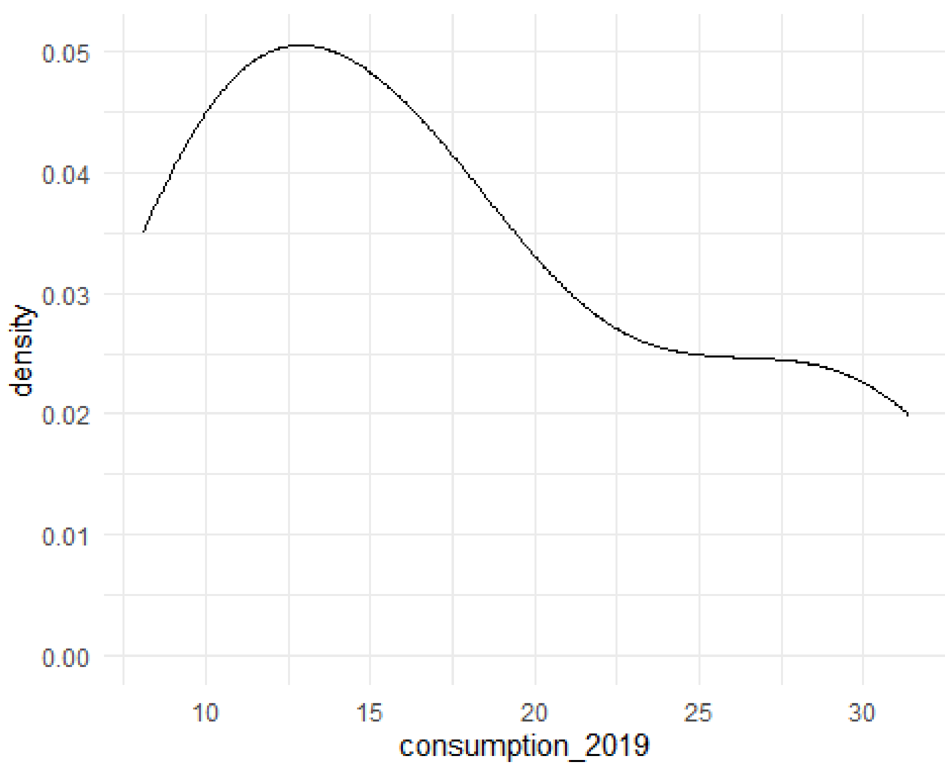
- GDP and Forest Losses:
  - Spearman's rho value of -0.09, indicating a negative, weak relationship between a country's GDP and total forest losses for that country.
  - This means that as one of the variables goes up (GDP), the other goes down (forest losses), or vice versa.
  - It appears as though GDP does not really impact the forest losses a country has experienced.
  - Because the p-value is NOT less than 0.05, we can determine that the relationship is not significant.
- GDP and Forest Usage:
  - Spearman's rho value of -0.48, indicating a negative, moderate relationship between a country's GDP and total forest resource usage for that country.
  - This means that as one of the variables goes up (GDP), the other goes down (forest usage), or vice versa.
  - It appears as though GDP does not really impact a country's forest usage.
  - Because the p-value is NOT less than 0.05, we can determine that the relationship is not significant.

GDP and Materials Correlation (Material Consumption and Material Footprint) (2019 and 2017, Respectively)



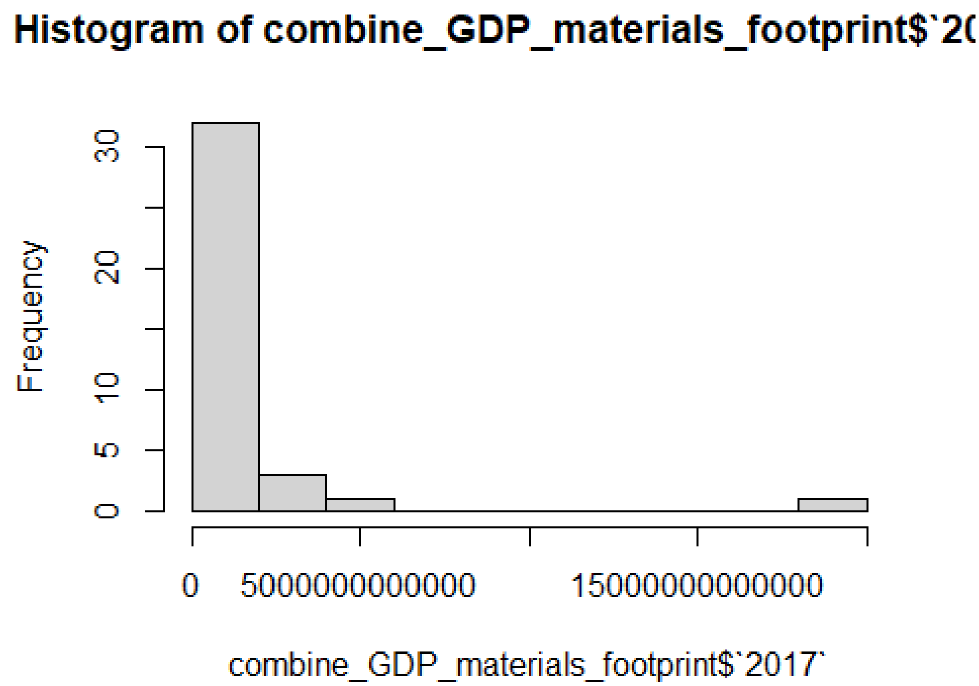


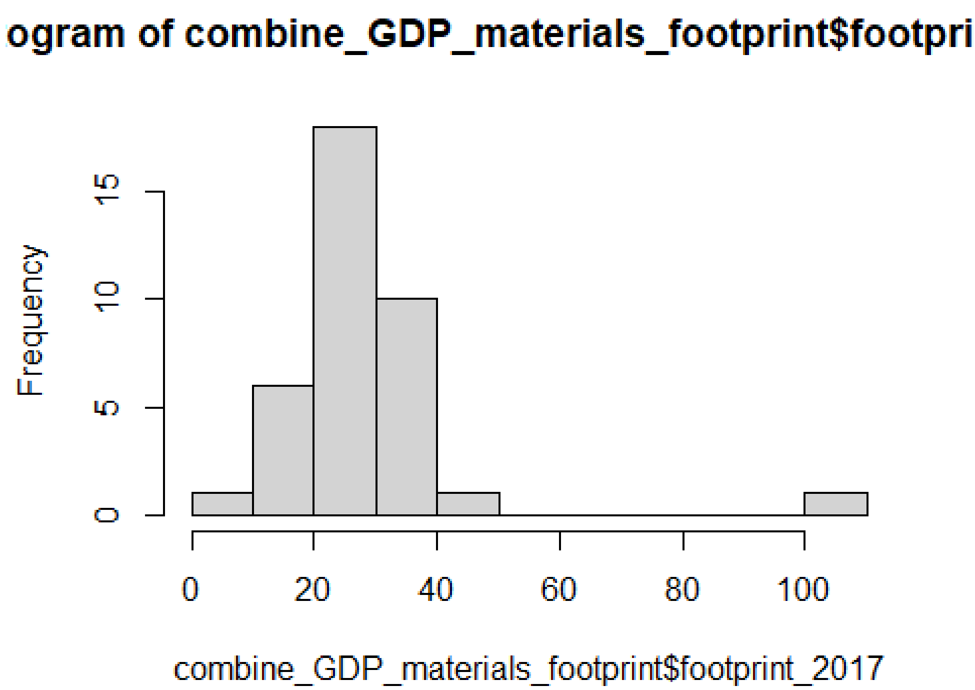


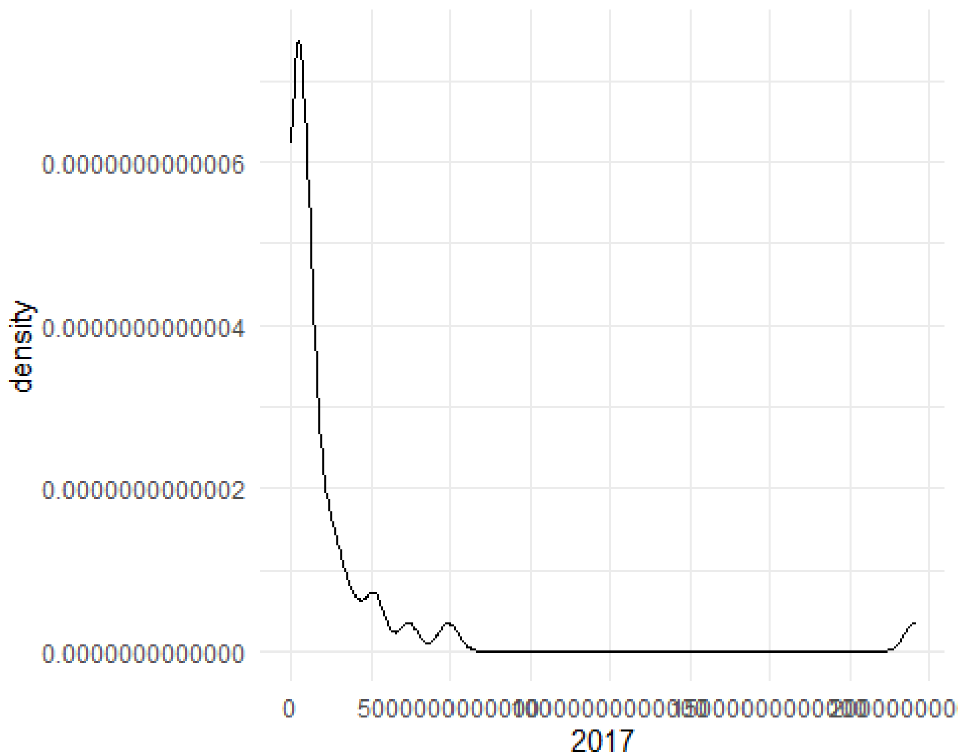


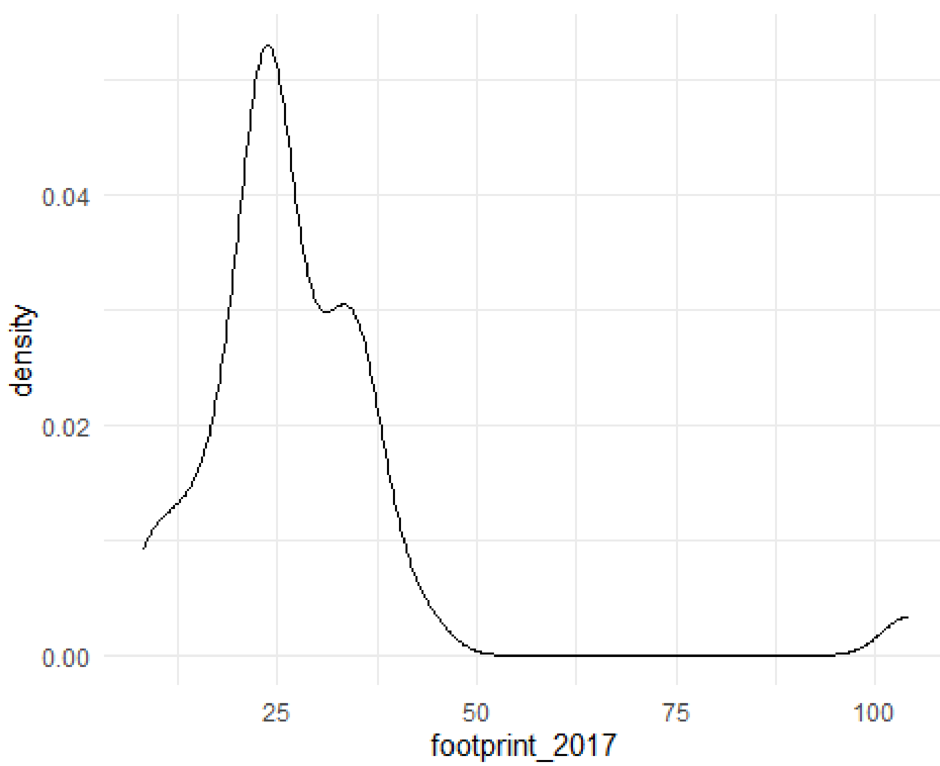

```
[1] -0.4776068


    Spearman's rank correlation rho
```

```
data:  combine_GDP_materials_consume$`2019` and combine_GDP_materials_consume
$consumption_2019
S = 4322, p-value = 0.01361
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.4776068
```


Histogram of combine_GDP_materials_footprint$`20
Frequency
0
5
10
20
30
0
5000000000000
15000000000000
combine_GDP_materials_footprint$`2017`


ogram of combine_GDP_materials_footprint$footpri
Frequency
0
5
10
15
0
20
40
60
80
100
combine_GDP_materials_footprint$footprint_2017





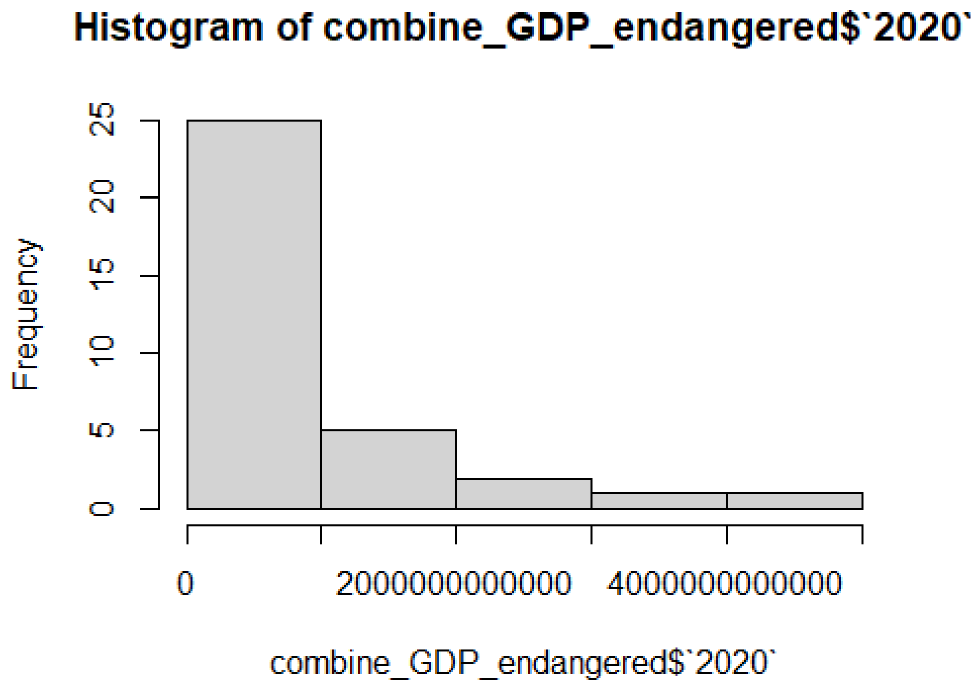

```
[1] -0.08819346


    Spearman's rank correlation rho
```

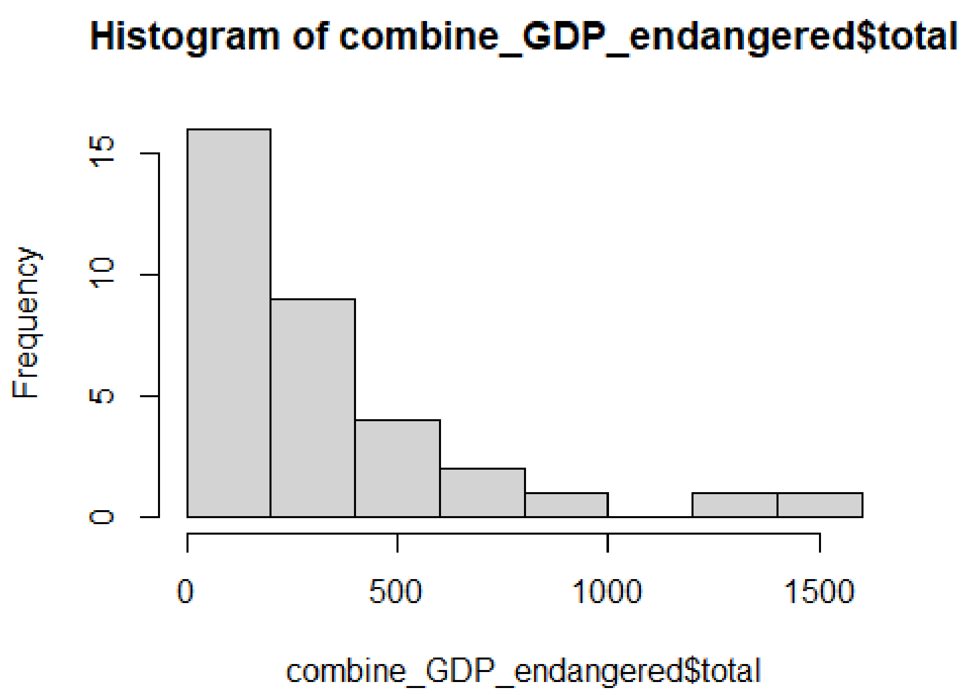
```
data:  combine_GDP_materials_footprint$`2017` and combine_GDP_materials_footp
rint$footprint_2017
S = 9180, p-value = 0.6037
alternative hypothesis: true rho is not equal to 0
sample estimates:
        rho
-0.08819346
```

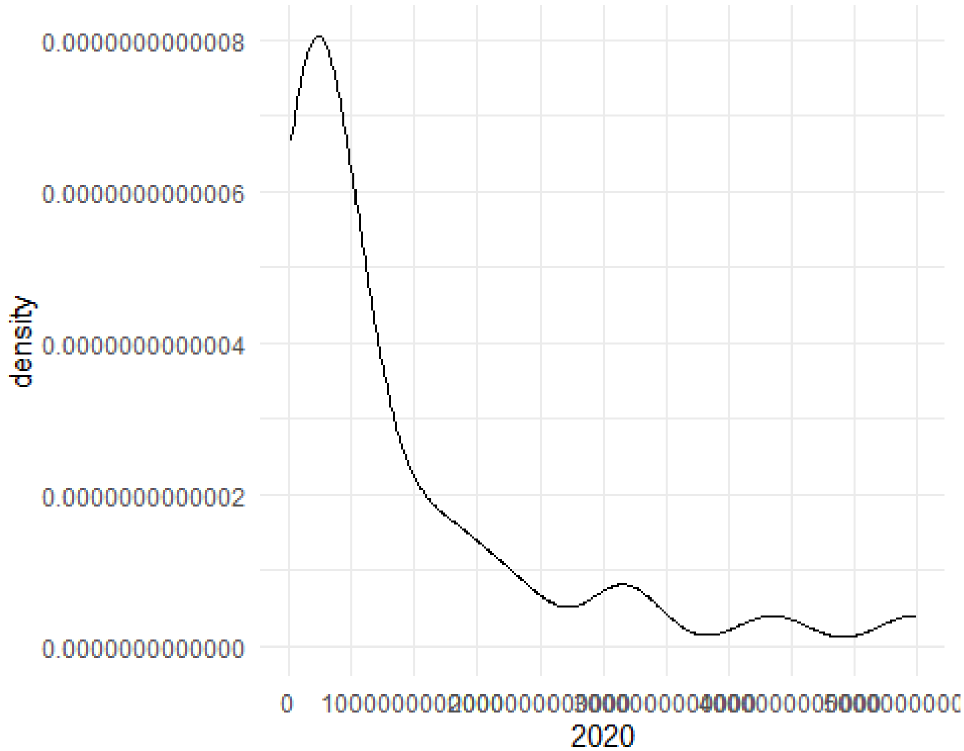
This test is running a correlation between the same countries present in both the GDP dataset and each of the material datasets (material consumption and material footprint). The test is being run for the year 2019 for GDP and material consumption (2019 was the most recent data collected for this variable), and 2017 for GDP and material footprint (2017 is the most recent data collected). Using the same yearly measures for both will allow for a more accurate assessment and outcome. The output of this correlational analysis is:

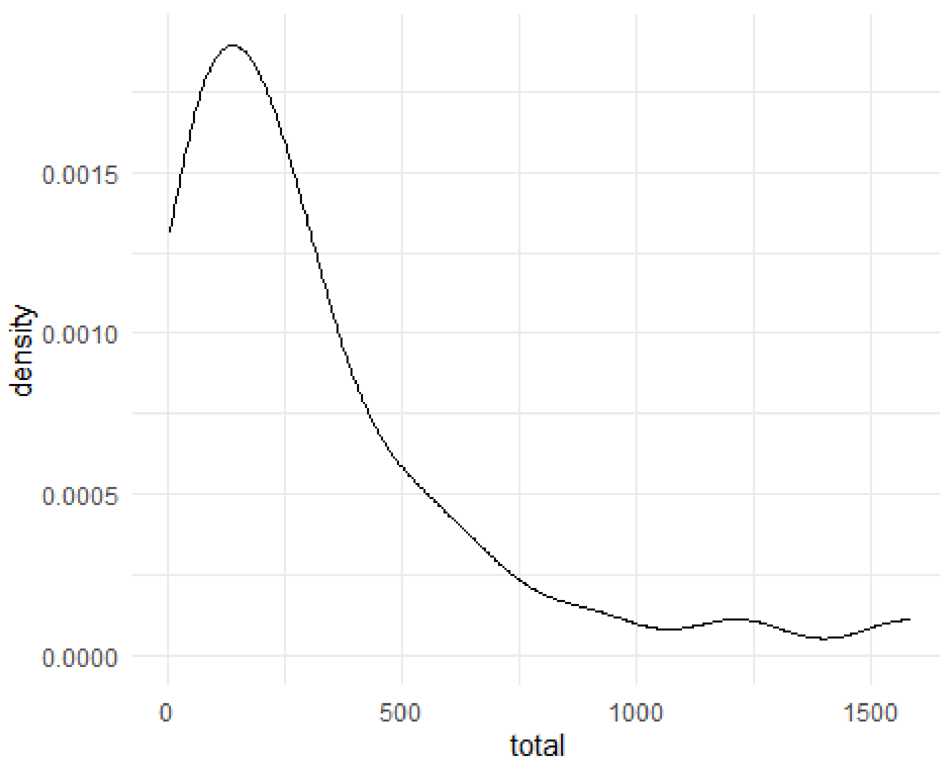
- GDP and Material Consumption:
  - Spearman's rho value of -0.48, indicating a negative, moderately strong relationship between a country's GDP and material consumption per capita for that country.
  - This means that as one of the variables goes up (GDP), the other goes down (material consumption), and vice versa.
  - It appears as though GDP may impact a country's material consumption.
  - Because the p-value is less than 0.05, we can determine that the relationship is significant.
- GDP and Material Footprint:
  - Spearman's rho value of -0.09, indicating a negative, weak relationship between a country's GDP and material footprint per capita for that country.
  - This means that as one of the variables goes up (GDP), the other goes down (material footprint), and vice versa.
  - It appears as though GDP does not impact a country's material footprint.
  - Because the p-value is NOT less than 0.05, we can determine that the relationship is not significant.

GDP and Critically Endangered Species Correlation (2020/2021)









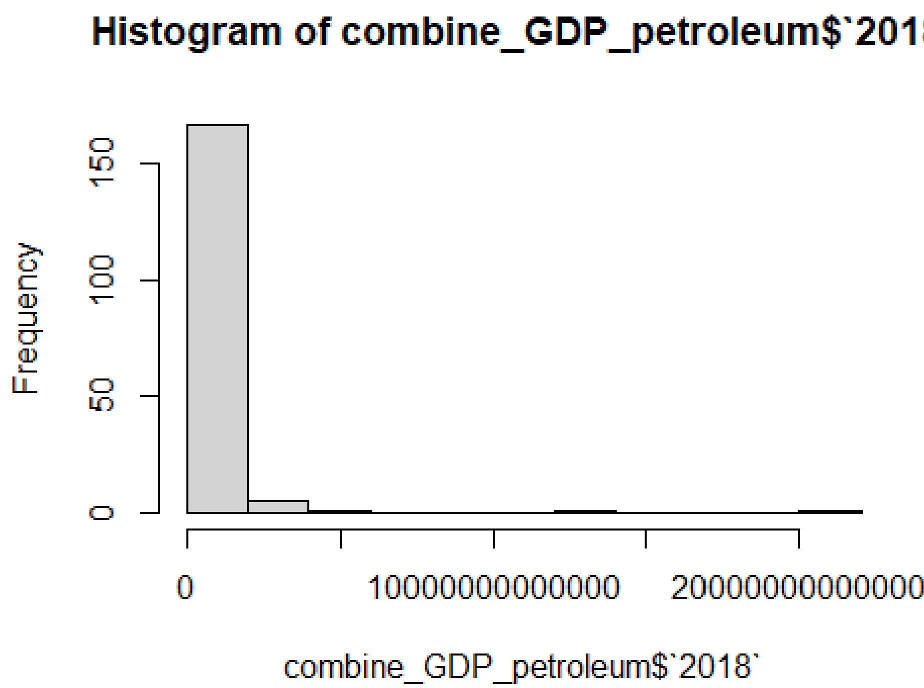

```
[1] 0.5533313


    Spearman's rank correlation rho
```

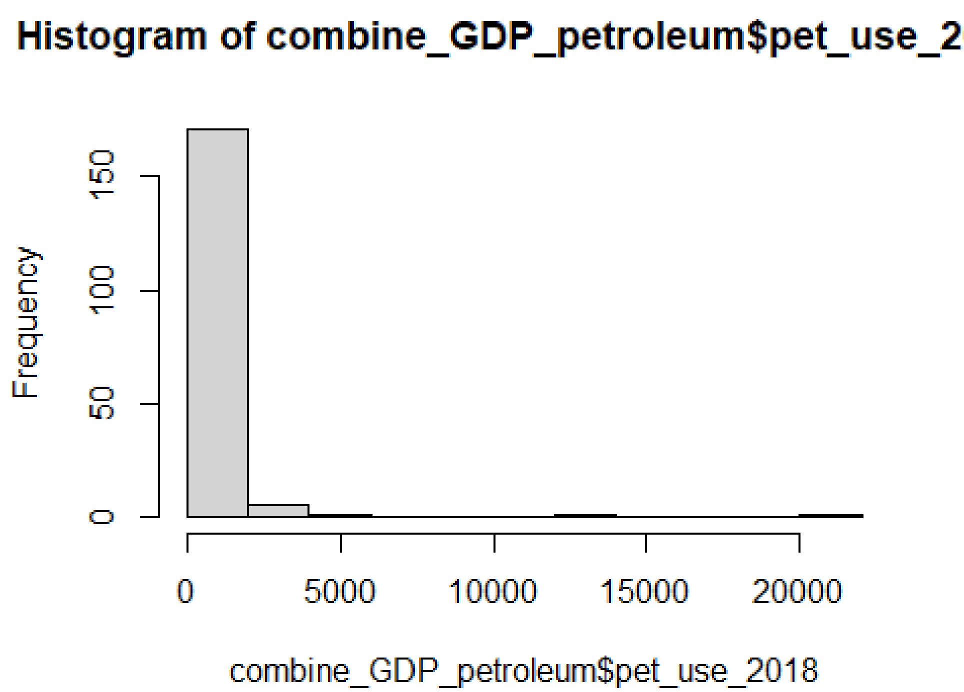
```
data:  combine_GDP_endangered$`2020` and combine_GDP_endangered$total
S = 2923.4, p-value = 0.000688
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.5533313
```

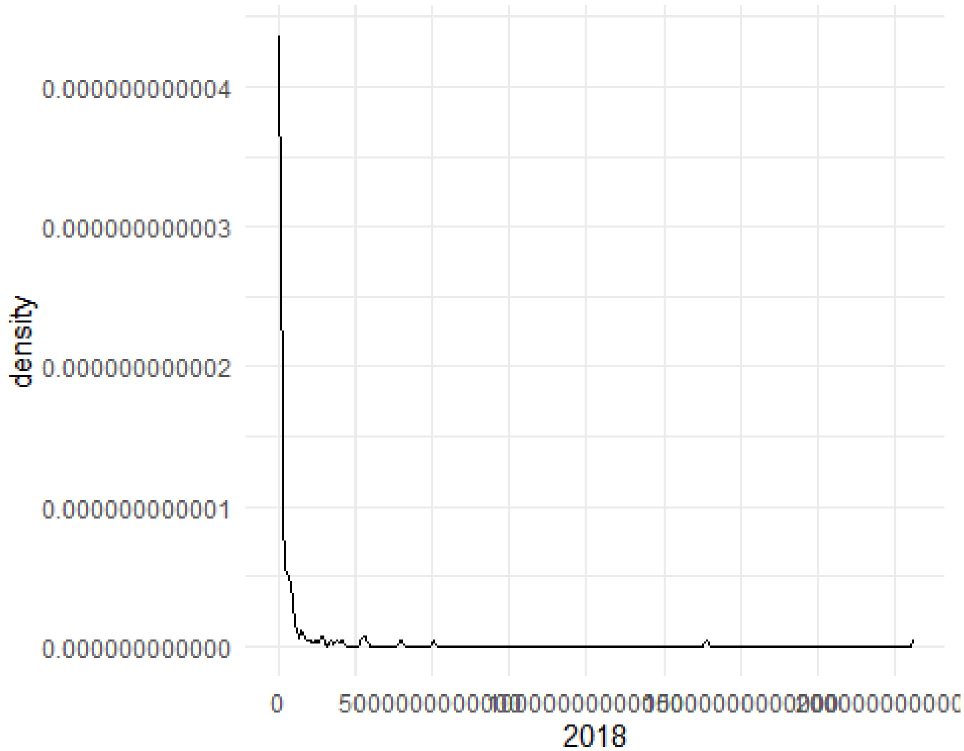
This test is running a correlation between the same countries present in both the GDP dataset and the critically endangered species dataset. The test is being run for the year 2020 for GDP and 2021 for critically endangered species (most recent and only data available). The output of this correlational analysis is:

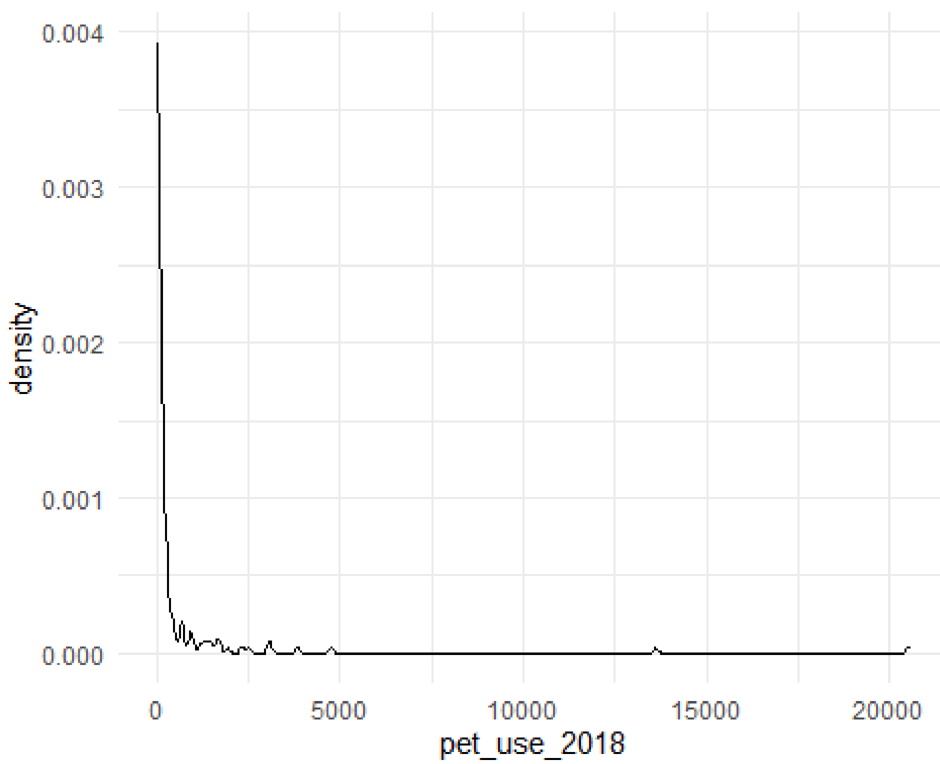
- GDP and Critically Endangered Species:
  - Spearman's rho value of 0.55, indicating a positive, moderately strong relationship between a country's GDP and its total number of critically endangered species.
  - This means that as one of the variables goes up (GDP), the other also goes up (endangered species), or vice versa (as one goes down so does the other).
  - It appears as though GDP may impact a country's total number of critically endangered species.
  - Because the p-value is less than 0.05, we can determine that the relationship is significant.

GDP and Petroleum and Other Liquids Consumption (2018)









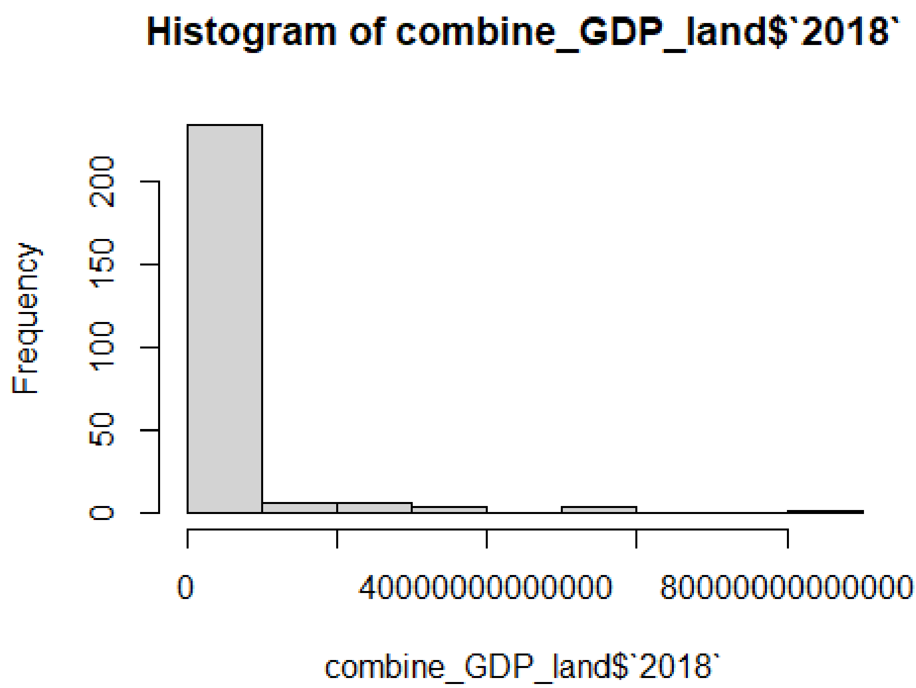

```
[1] 0.9734582


    Spearman's rank correlation rho
```

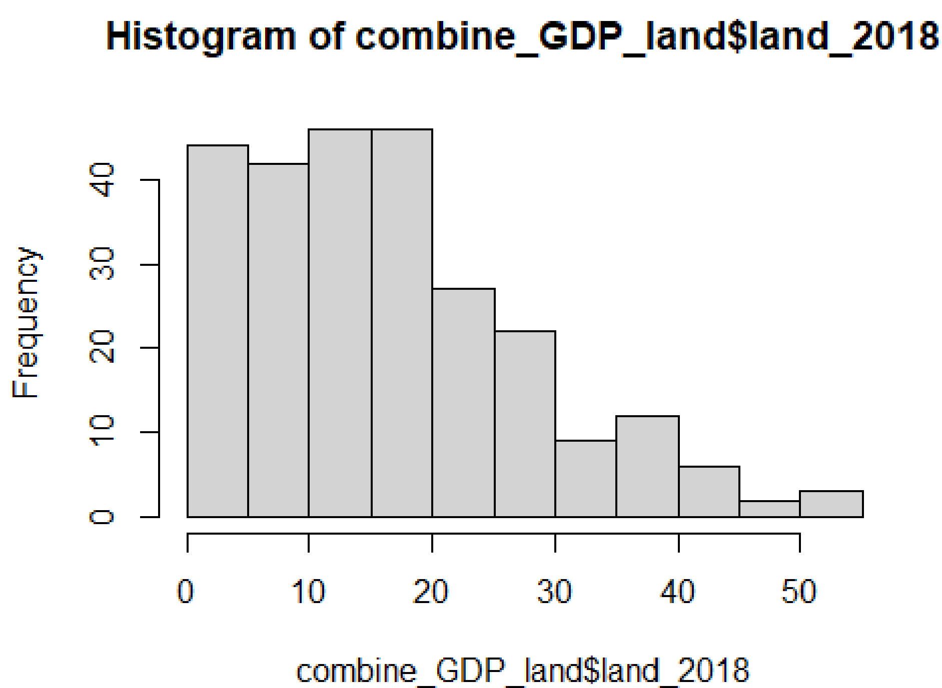
```
data:  combine_GDP_petroleum$`2018` and combine_GDP_petroleum$pet_use_2018
S = 22904, p-value < 0.00000000000000022
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.9734582
```

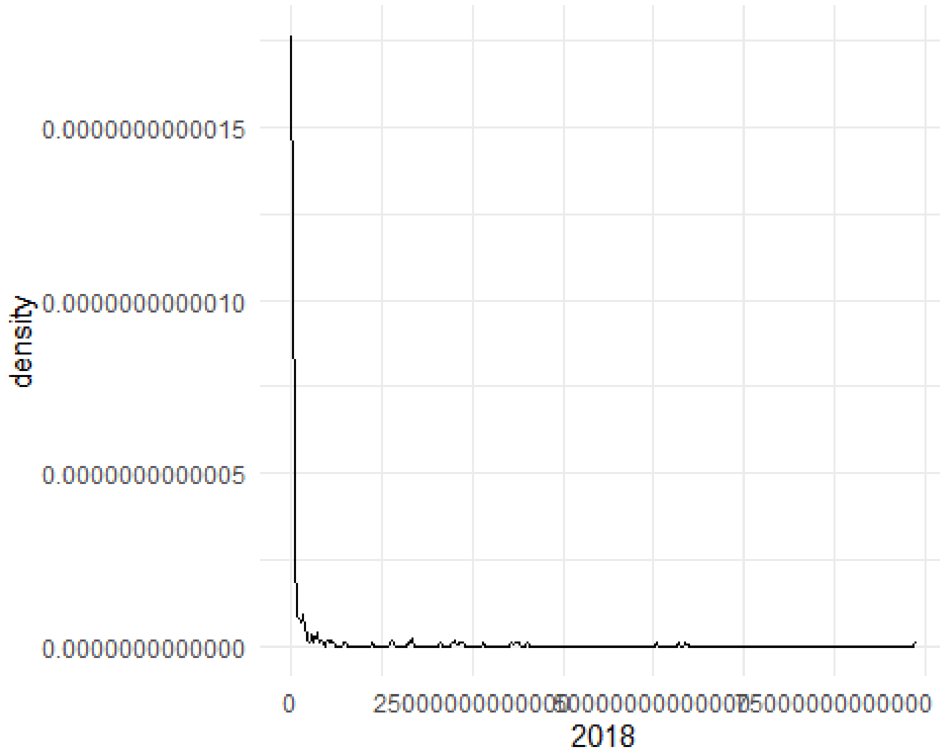
This test is running a correlation between the same countries present in both the GDP dataset and the petroleum and other liquids consumption dataset. The test is being run for the year 2018 for GDP and the petroleum dataset (most recent year of data available). Using 2018 measures for both will allow for a more accurate assessment and outcome. The output of this correlational analysis is:

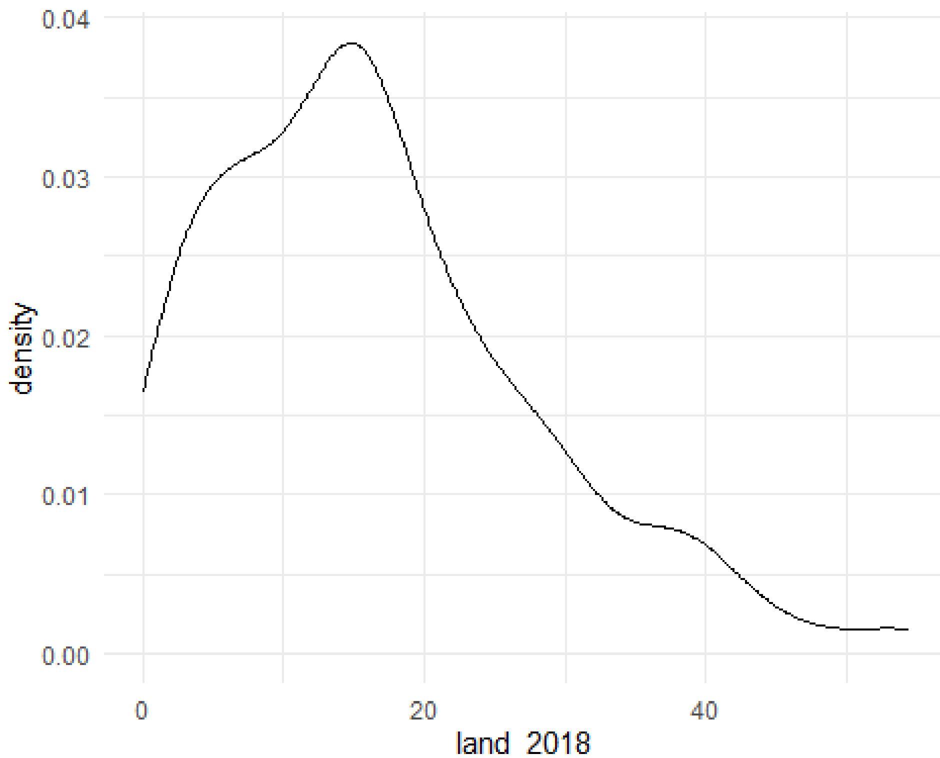
- Population and Petroleum/Other Liquids Consumption:
  - Spearman's rho value of 0.97, indicating a positive, extremely strong relationship between a country's GDP and its consumption of petroleum and other related liquids.
  - This means that as one of the variables goes up (GDP), the other also goes up (petroleum consumption), or vice versa (as one goes down so does the other).
  - It appears as though GDP does have an impact on a country's total amount of petroleum and other liquids consumption.
  - Because the p-value is less than 0.05, we can determine that the relationship is significant (extremely so, in this case).

GDP and Protected Terrestrial Land (2018)










```
[1] 0.07607701


    Spearman's rank correlation rho
```

```
data:  combine_GDP_land$`2018` and combine_GDP_land$land_2018
S = 2377254, p-value = 0.2316
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
0.07607701
```

This test is running a correlation between the same countries present in both the GDP dataset and the protected land dataset. The test is being run for the year 2018 for GDP and the protected land dataset (most recent year of data available). Using 2018 measures for both will allow for a more accurate assessment and outcome. The output of this correlational analysis is:

- GDP and Protected Terrestrial Land:
  - Spearman's rho value of 0.08, indicating a positive, weak relationship between a country's GDP and its percentage of protected land.
  - This means that as one of the variables goes up (GDP), so does the other (protected land), and vice versa (down/down).
  - It appears as though GDP does not have an impact on a country's total percentage of protected terrestrial land.
  - Because the p-value is NOT less than 0.05, we can determine that the relationship is not significant.

Time Series Analyses

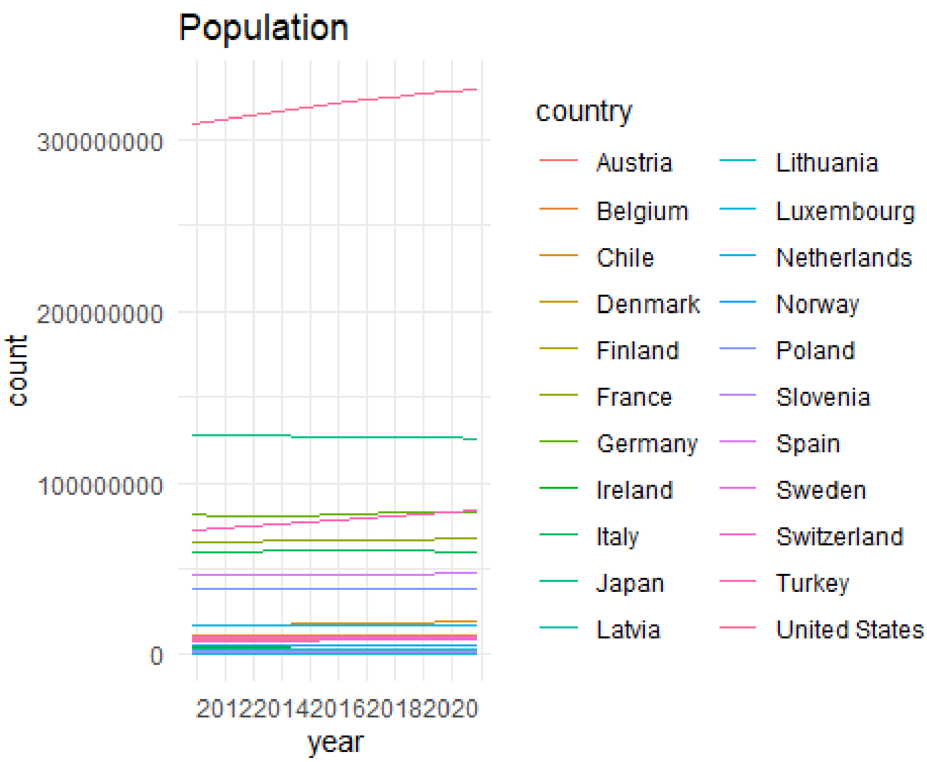
Time series analyses will only be completed on those datasets that have multiple years of data available. Those datasets with only one year available (or only the most recent year of data available) will not be utilized in a time series analysis, as this type of analysis would not apply to those cases. For my analysis, running time series and respective plots will allow me to understand and visualize changes over time for the various variable measures.

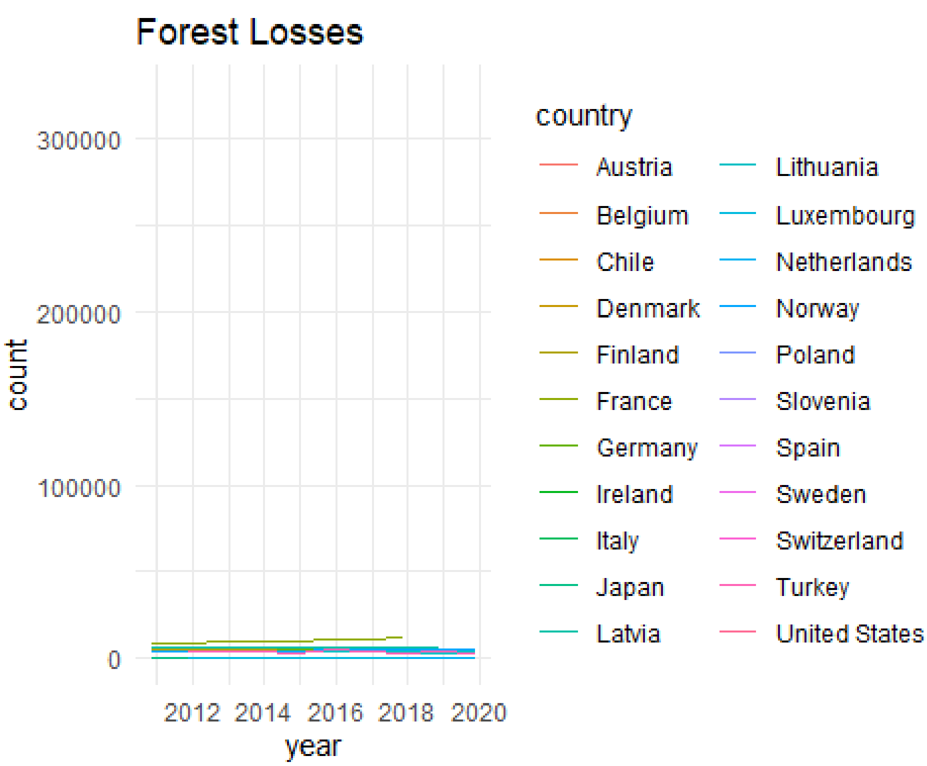
These measures include:

- Population and forest loss
- Population and forest usage
- Population and material consumption
- Population and material footprint
- Population and petroleum consumption
- Population and protected land
- GDP and forest loss
- GDP and forest usage
- GDP and material consumption
- GDP and material footprint
- GDP and petroleum consumption
- GDP and protected land

Population and Forest Loss (2010-2019)

```
Attaching package: 'zoo'

The following objects are masked from 'package:base':

    as.Date, as.Date.numeric


Attaching package: 'reshape2'

The following object is masked from 'package:tidyr':

    smiths


Attaching package: 'lubridate'

The following objects are masked from 'package:base':

    date, intersect, setdiff, union
```






```
Time Series:
Start = 2010
End = 2019
Frequency = 1
```

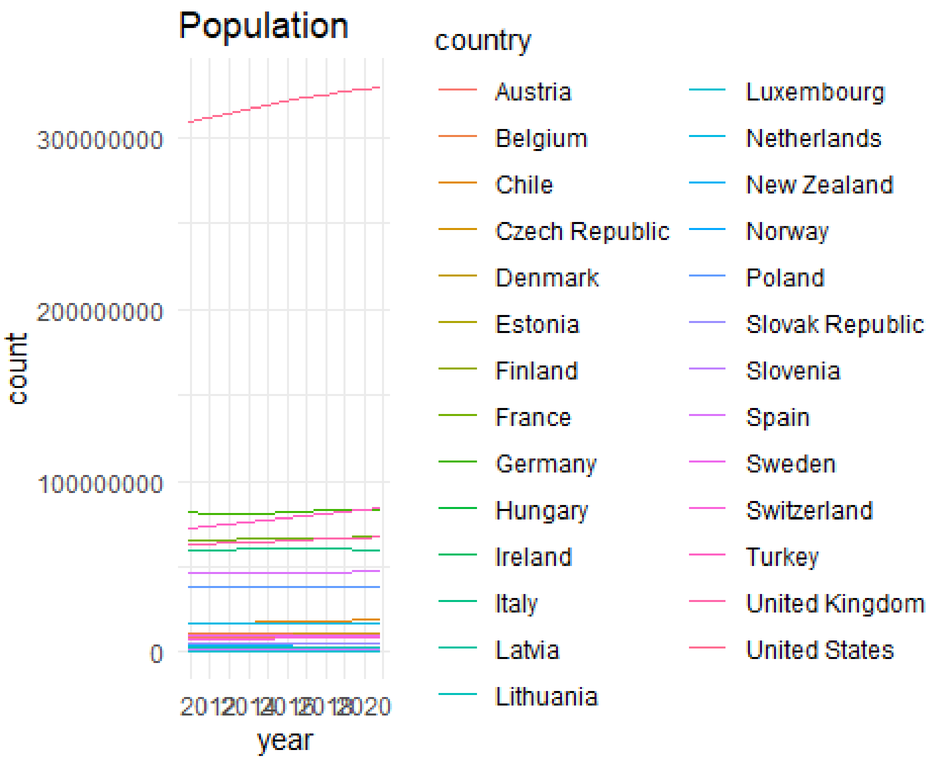
```
     country  year count
2010       1 14933    NA
2011       1 15298    NA
2012       1 15664    NA
2013       1 16029  5710
2014       1 16394    NA
2015       1 16759    NA
2016       1 17125    NA
2017       1 17490    NA
2018       1 17855    NA
2019       1 18220    NA
```

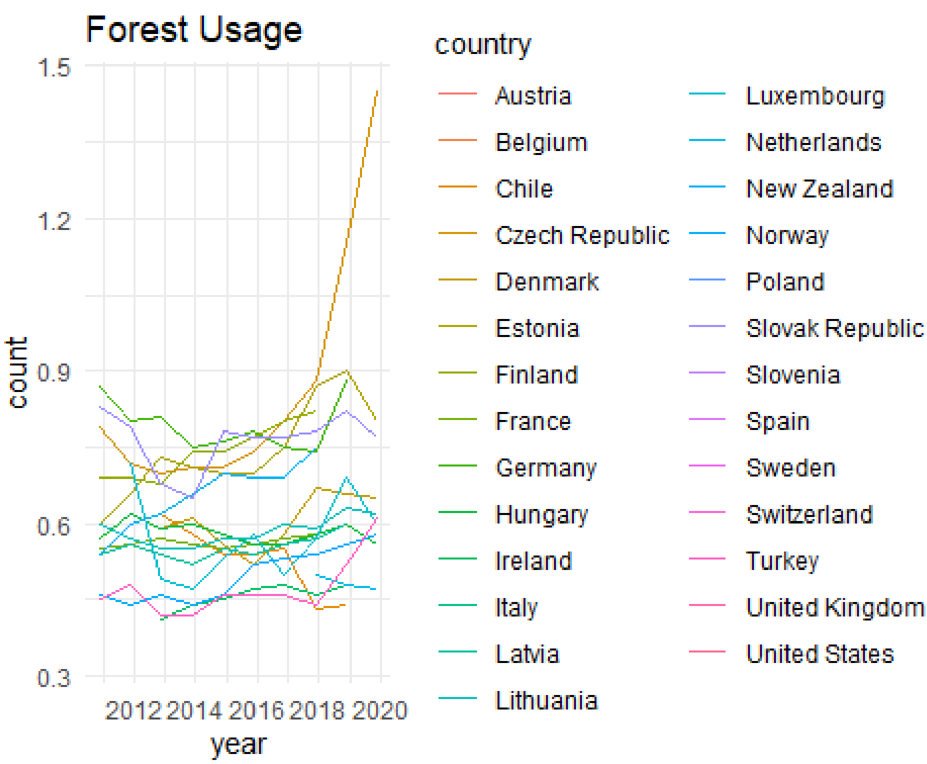
In the above time series plots, we can see:

- The first plot (which is population), shows a slight incline in population for some countries with stagnation or a decline in a few others.
- The second plot, forest losses, indicates relatively similar amounts available between the different countries.

The accompanying output from the time series equation is also present.

Population and Forest Usage (2010-2019)






```
Time Series:
Start = 2010
```

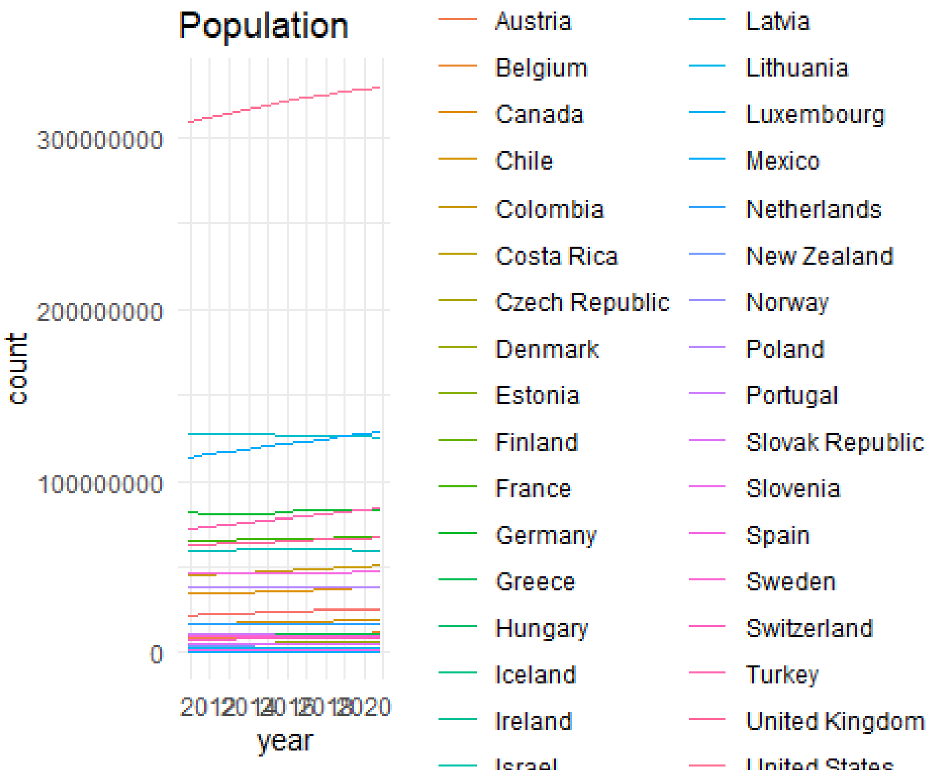
```
End = 2019
Frequency = 1
     country  year count
2010       1 14933    NA
2011       1 15298    NA
2012       1 15664    NA
2013       1 16029  0.82
2014       1 16394    NA
2015       1 16759    NA
2016       1 17125    NA
2017       1 17490    NA
2018       1 17855    NA
2019       1 18220    NA
```

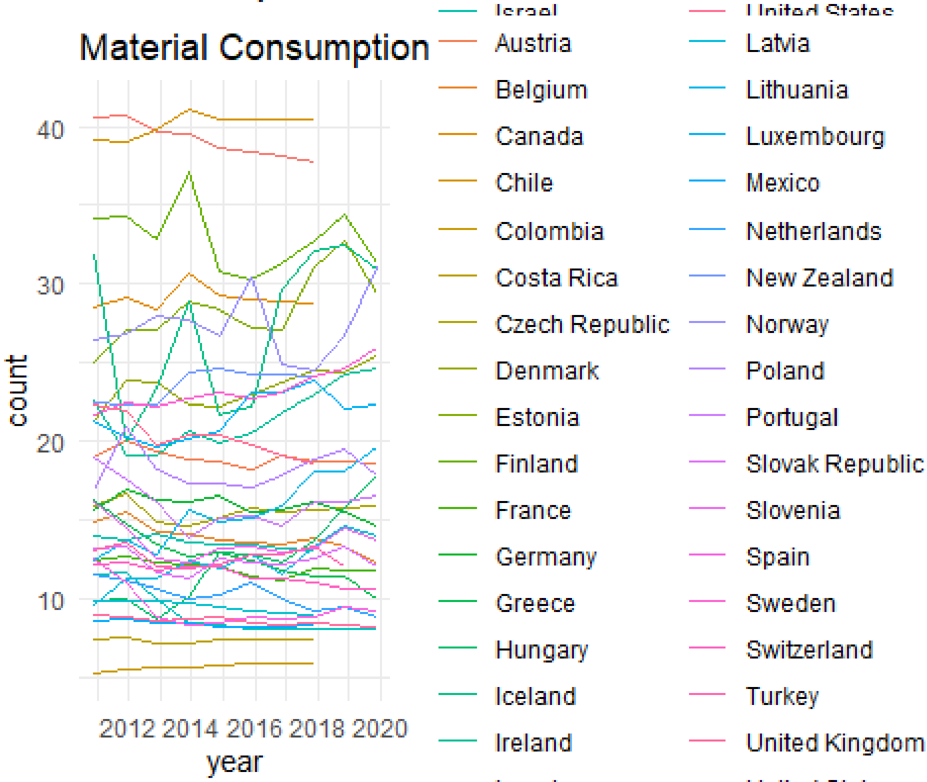
In the above time series plots, we can see:

- The first plot (which is population), is quite similar to the forest losses population plot.
- The second plot, forest usage, indicates some interesting and different trends between countries. The changes appear to be quite dynamic in nature - going up and down with peak (spikes) and valleys.

The accompanying output from the time series equation is also present.

Population and Material Consumption (2010-2019)






```
Time Series:
Start = 2010
```

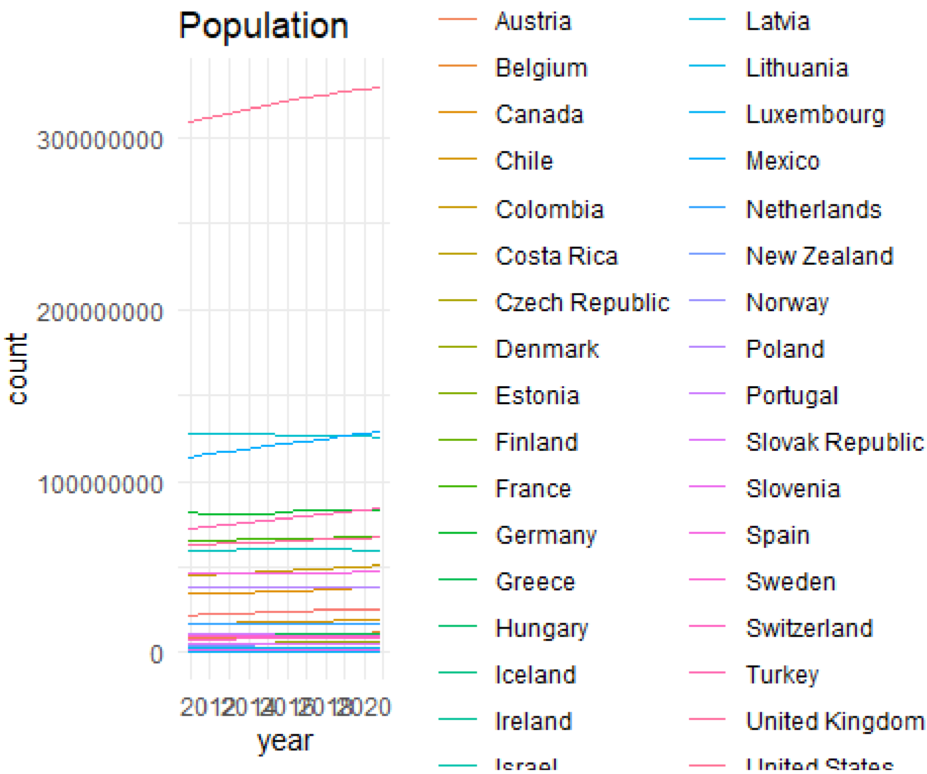
```
End = 2019
Frequency = 1
     country  year   count
2010       1 14933 40.5952
2011       1 15298 40.7088
2012       1 15664 39.7431
2013       1 16029 39.5675
2014       1 16394 38.6466
2015       1 16759 38.4333
2016       1 17125 38.0788
2017       1 17490 37.7188
2018       1 17855      NA
2019       1 18220      NA
```

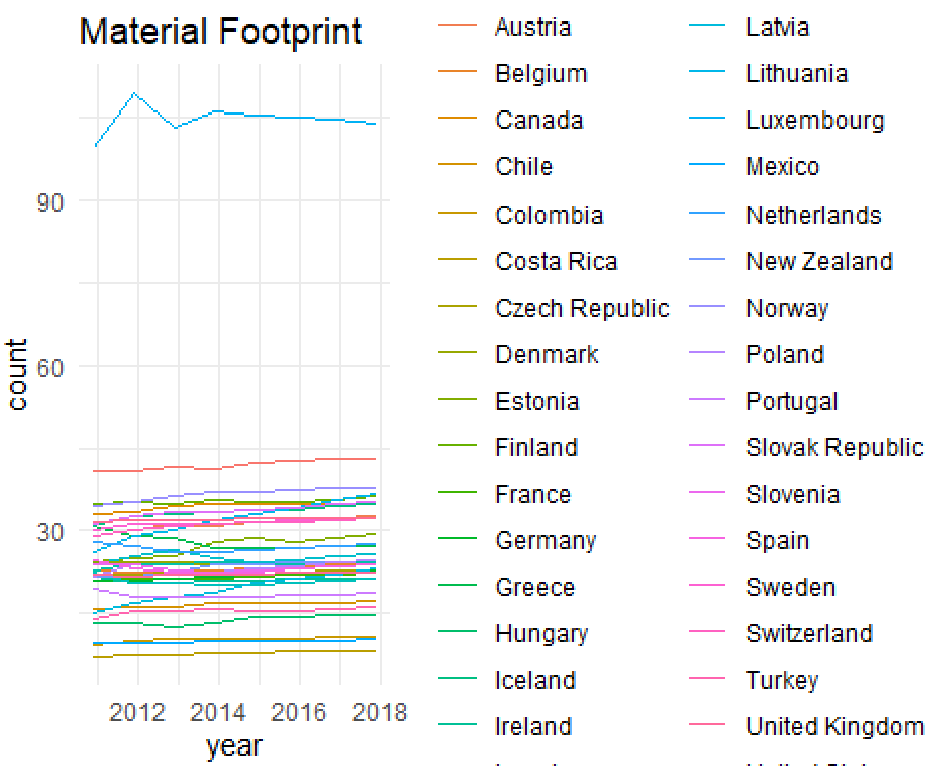
In the above time series plots, we can see:

- The first plot (which is population), is quite similar to the other population plots - pretty steady numbers across time.
- The second plot, material consumption, indicates some interesting and different trends between countries. The changes appear to be quite dynamic in nature - going up and down with peak (spikes) and valleys.

The accompanying output from the time series equation is also present.

Population and Material Footprint (2010-2017)






```
Time Series:
Start = 2010
```

```
End = 2017
Frequency = 1
     country  year   count
2010       1 14933 40.7986
2011       1 15298 40.9456
2012       1 15664 41.5003
2013       1 16029 41.1921
2014       1 16394 42.1842
2015       1 16759 42.5502
2016       1 17125 42.8273
2017       1 17490 43.1128
```

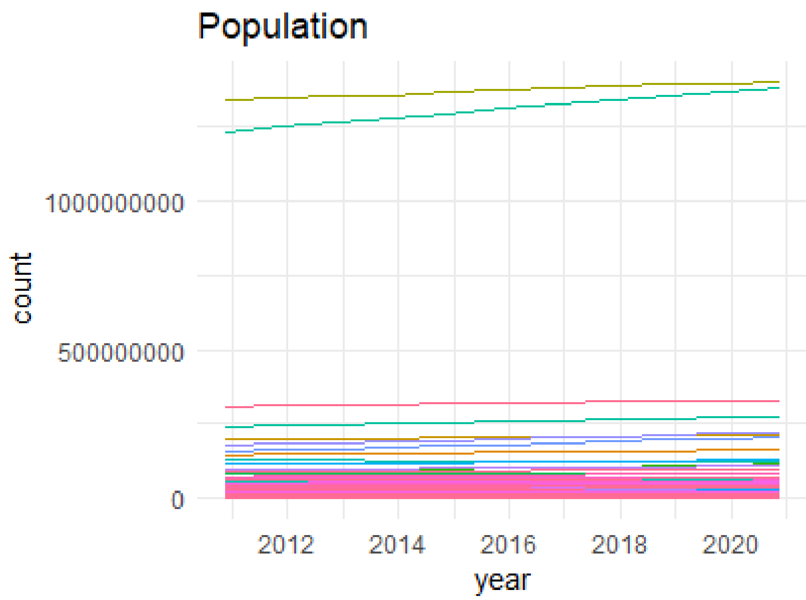
In the above time series plots, we can see:

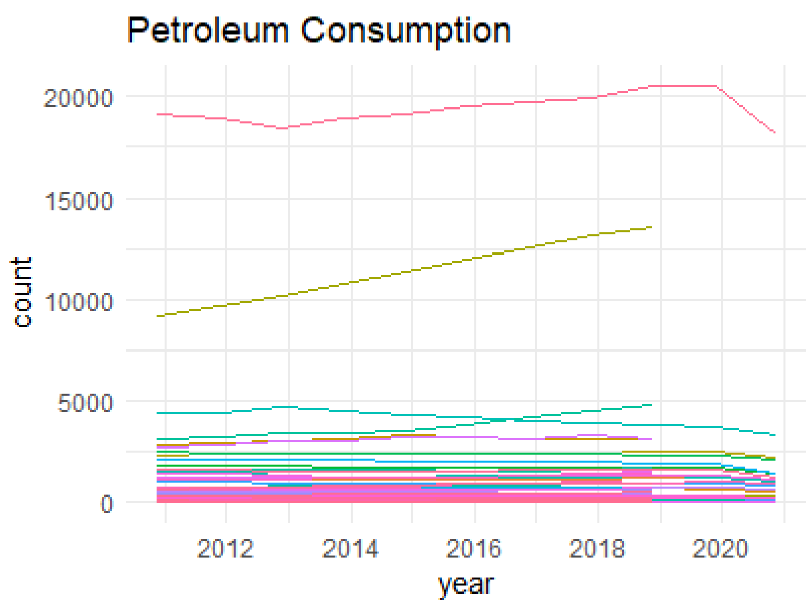
- The first plot (which is population), is quite similar to the other population plots - pretty steady numbers across time.
- The second plot, material footprint, is also pretty steady across time - with a slight increase - and one country with higher numbers than the rest.

The accompanying output from the time series equation is also present.

Population and Petroleum/Other Liquids Consumed (2010-2020)

```
Attaching package: 'cowplot'

The following object is masked from 'package:lubridate':

    stamp
```






```
Time Series:
Start = 2010
End = 2020
Frequency = 1
```

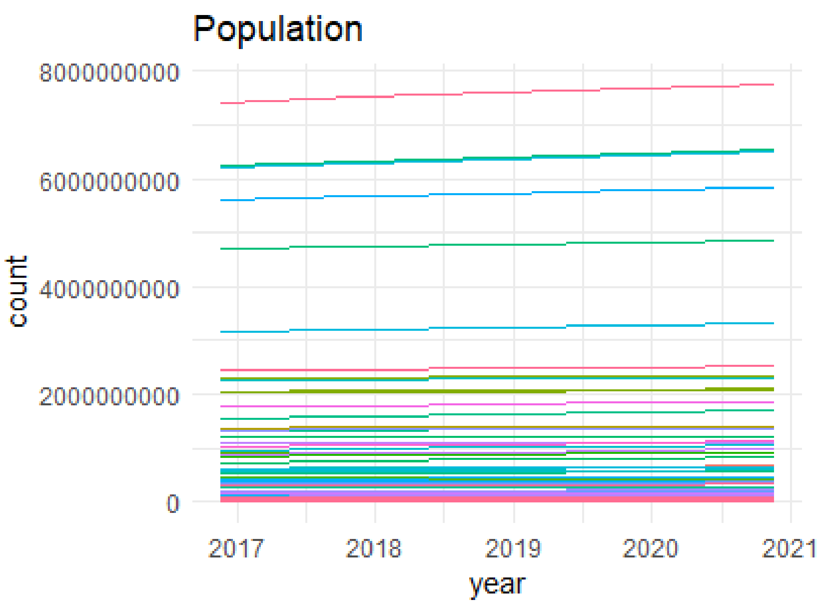
```
     country  year count
2010       1 14933 42.76
2011       1 15298 55.82
2012       1 15664 49.06
2013       1 16029 35.08
2014       1 16394 27.68
2015       1 16759 34.91
2016       1 17125 25.59
2017       1 17490 23.63
2018       1 17855 23.15
2019       1 18220    NA
2020       1 18586    NA
```

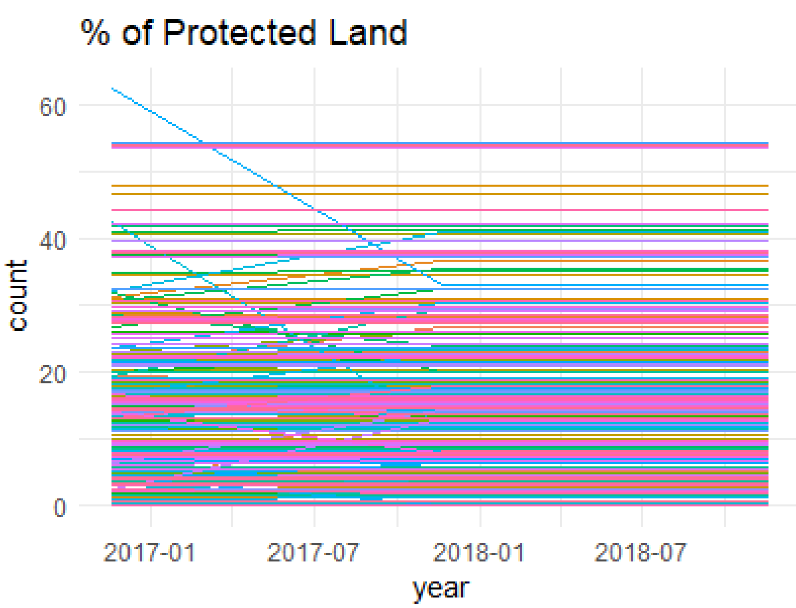
In the above time series plots, we can see:

- The first plot (which is population), is a bit different from the other population plots - there are a couple of country outliers.
- The second plot, petroleum consumption, also includes a couple of country outliers.
- Legends were not included with these plots, as too many countries existed in the population and petroleum datasets (both datasets possessed many countries in common with one another).

The accompanying output from the time series equation is also present.

Population and Protected Land (2016-2018)






```
Time Series:
Start = 2016
```

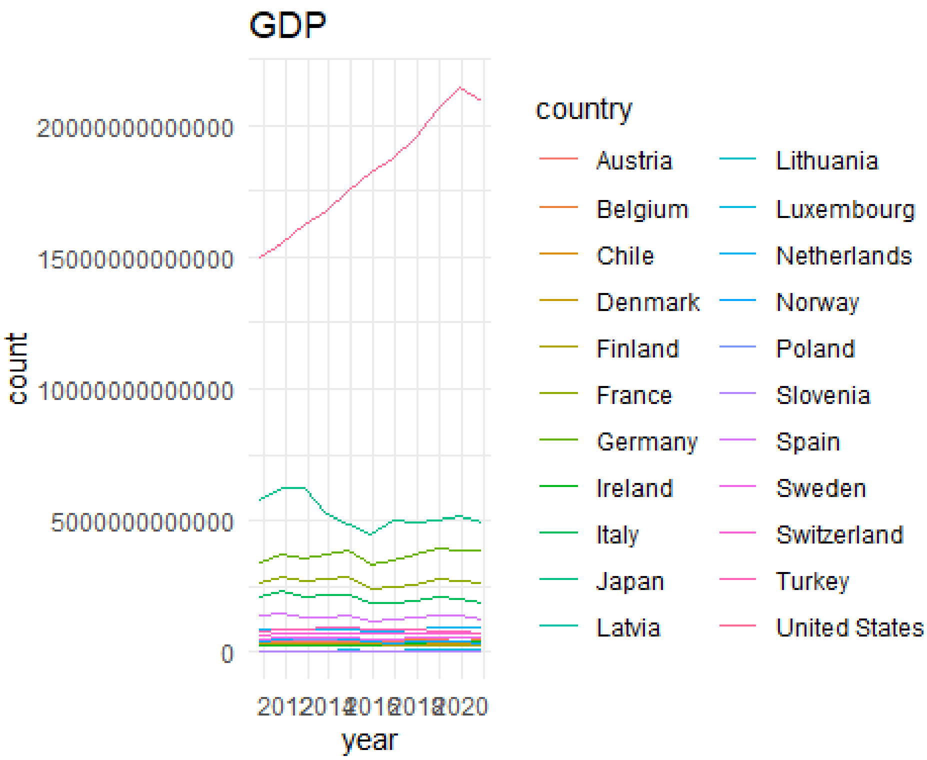
```
End = 2018
Frequency = 1
     country  year    count
2016       1 17125 0.100000
2017       1 17490 0.104707
2018       1 17855 0.104707
```

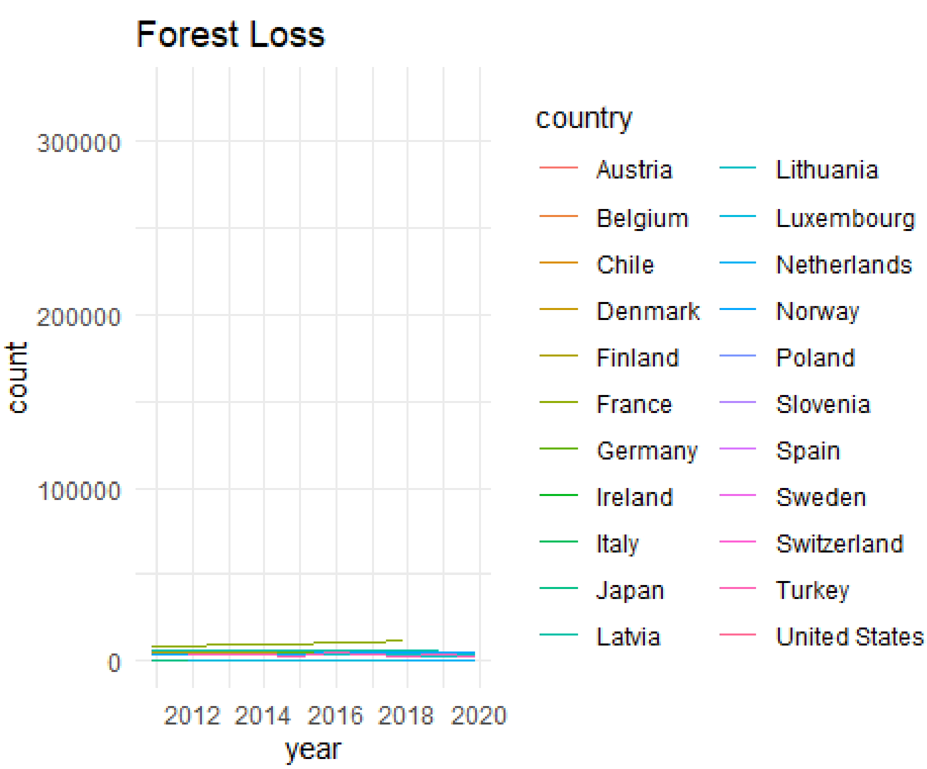
In the above time series plots, we can see:

- The first plot (population) also includes some country outliers, but the population increases are not drastic.
- The second plot, percentage of protected land, indicates that most countries have held steady across the past couple of most recent years of data collection, with a couple of outliers having a decline in protected land and then holding stagnant.
- Legends were not included with these plots, as too many countries existed in the population and land datasets (both datasets possessed many countries in common with one another).

The accompanying output from the time series equation is also present.

GDP and Forest Loss (2010-2019)






```
Time Series:
Start = 2010
```

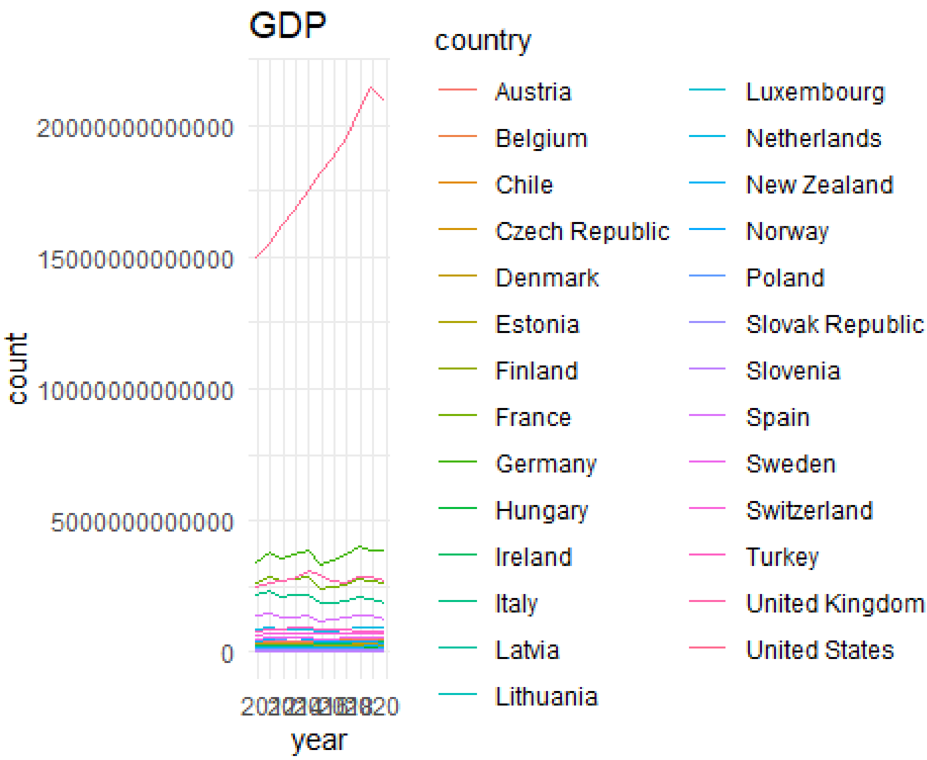
```
End = 2019
Frequency = 1
     country  year count
2010       1 14933    NA
2011       1 15298    NA
2012       1 15664    NA
2013       1 16029  5710
2014       1 16394    NA
2015       1 16759    NA
2016       1 17125    NA
2017       1 17490    NA
2018       1 17855    NA
2019       1 18220    NA
```

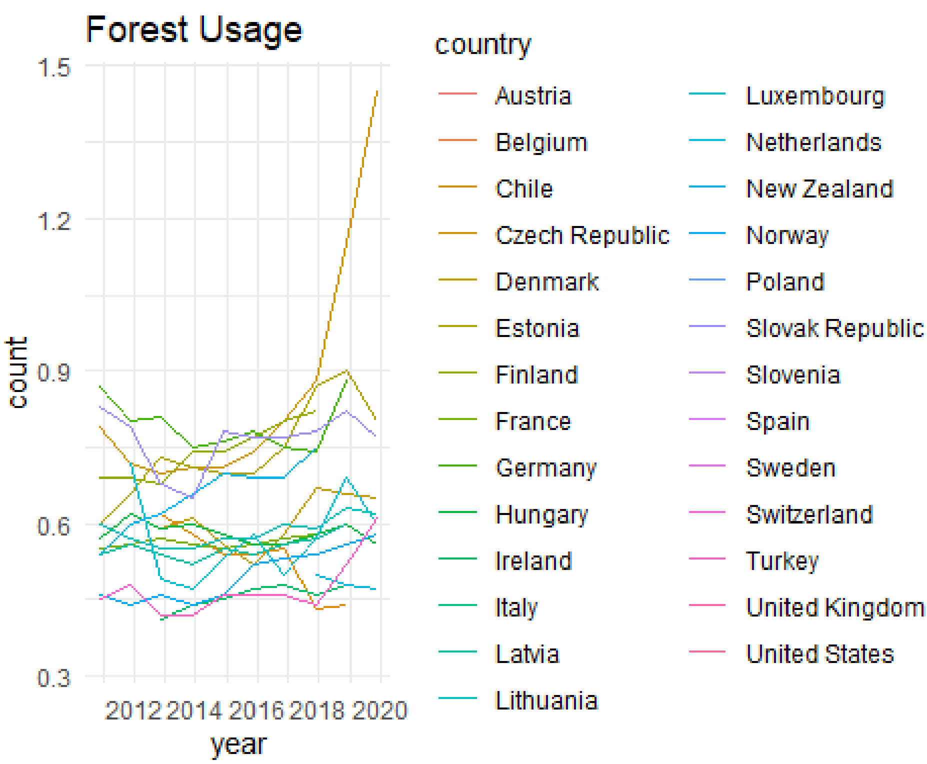
In the above time series plots, we can see:

- The first plot (GDP) shows most countries having a steady wave of GDP, with a couple of outliers (most notably, Austria).
- The second plot, forest loss, indicates that most countries have held steady from 2010-2019.

The accompanying output from the time series equation is also present.

GDP and Forest Usage (2010-2019)






```
Time Series:
Start = 2010
```

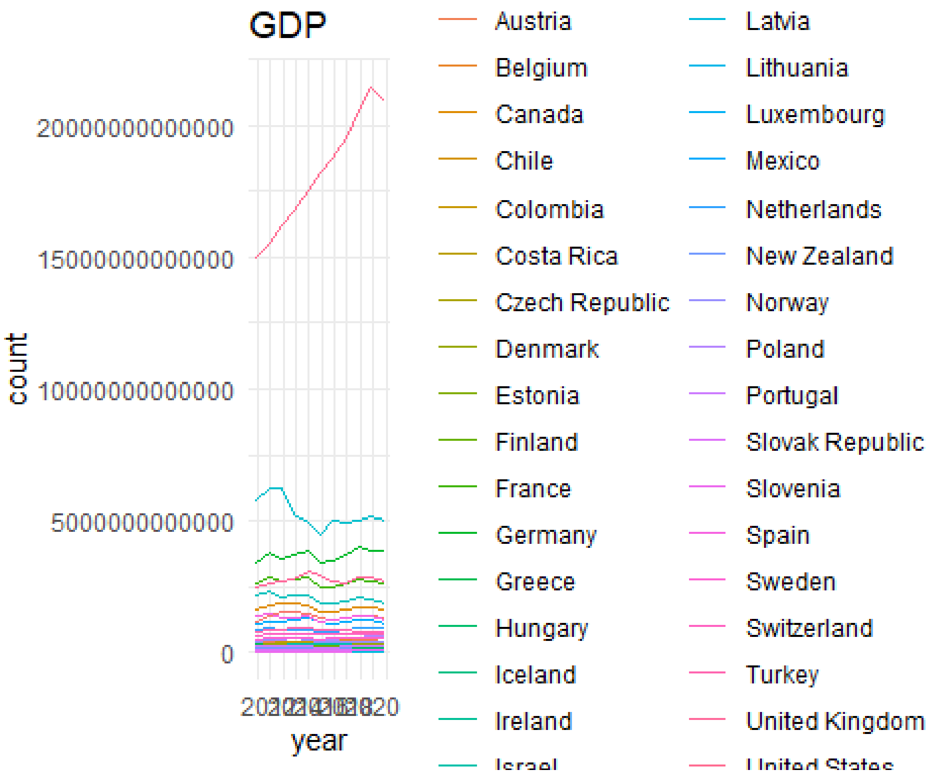
```
End = 2019
Frequency = 1
     country  year count
2010       1 14933    NA
2011       1 15298    NA
2012       1 15664    NA
2013       1 16029  0.82
2014       1 16394    NA
2015       1 16759    NA
2016       1 17125    NA
2017       1 17490    NA
2018       1 17855    NA
2019       1 18220    NA
```

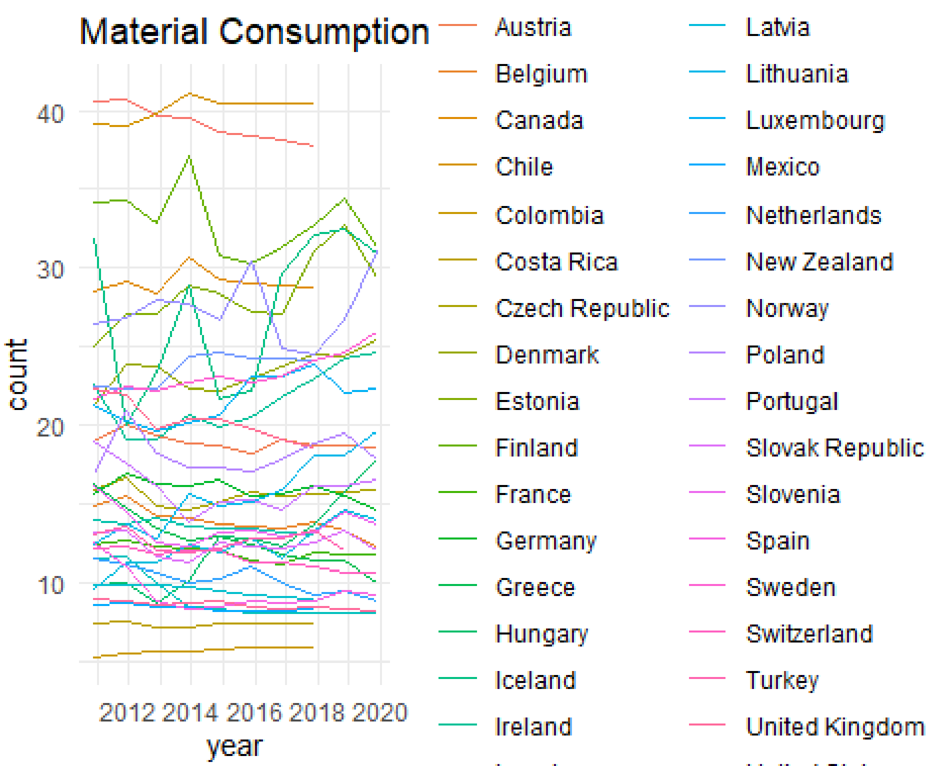
In the above time series plots, we can see:

- The first plot (GDP) shows most countries having a steady wave of GDP, with a couple of outliers (most notably, Austria again).
- The second plot, forest usage, indicates interesting waves of increases and decreases (peaks and valleys) from 2010-2019.

The accompanying output from the time series equation is also present.

GDP and Material Consumption (2010-2019)






```
Time Series:
Start = 2010
```

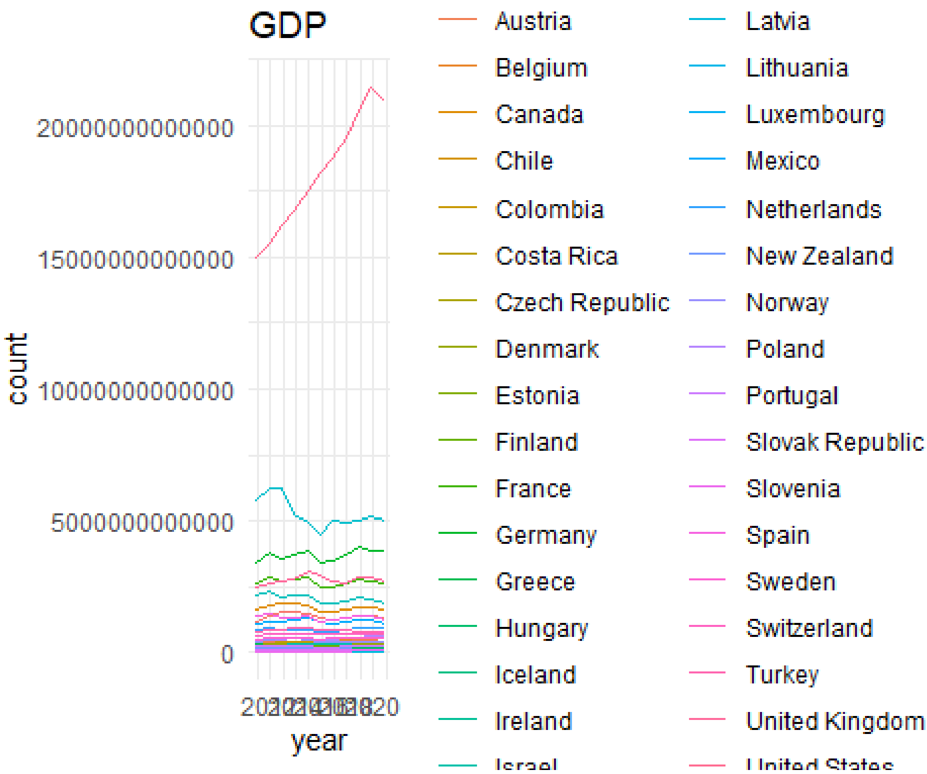
```
End = 2019
Frequency = 1
     country  year   count
2010       1 14933 40.5952
2011       1 15298 40.7088
2012       1 15664 39.7431
2013       1 16029 39.5675
2014       1 16394 38.6466
2015       1 16759 38.4333
2016       1 17125 38.0788
2017       1 17490 37.7188
2018       1 17855      NA
2019       1 18220      NA
```

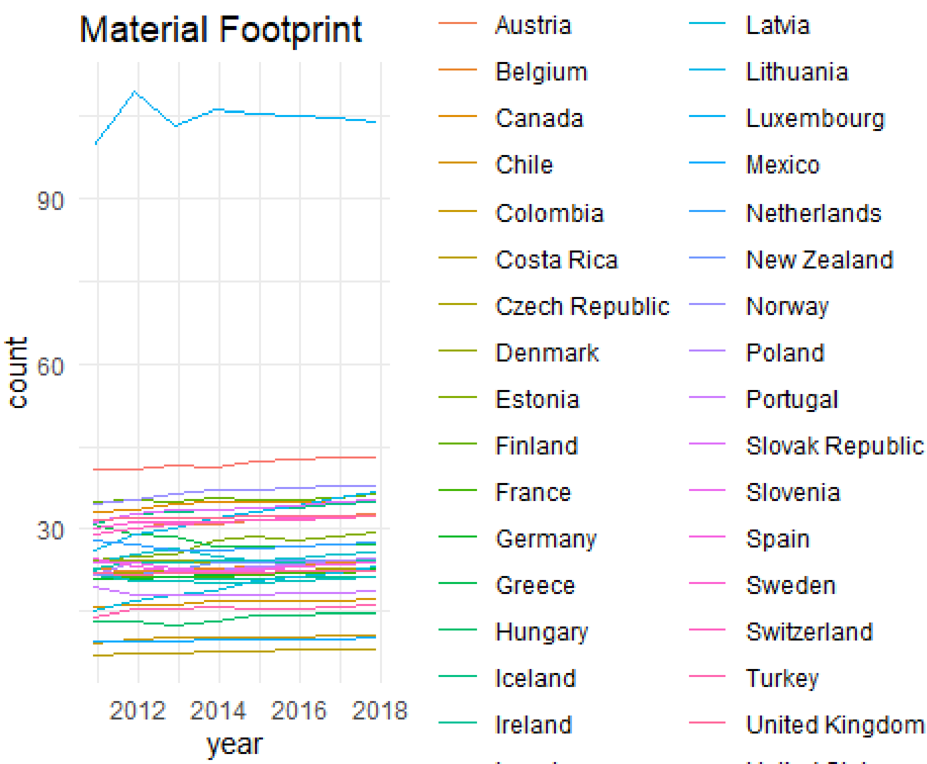
In the above time series plots, we can see:

- The first plot (GDP) shows most countries having a steady wave of GDP, with a couple of outliers (most notably, Austria again).
- The second plot, material consumption, indicates interesting waves of increases and decreases (peaks and valleys) from 2010-2019.

The accompanying output from the time series equation is also present.

GDP and Material Footprint (2010-2017)






```
Time Series:
Start = 2010
```

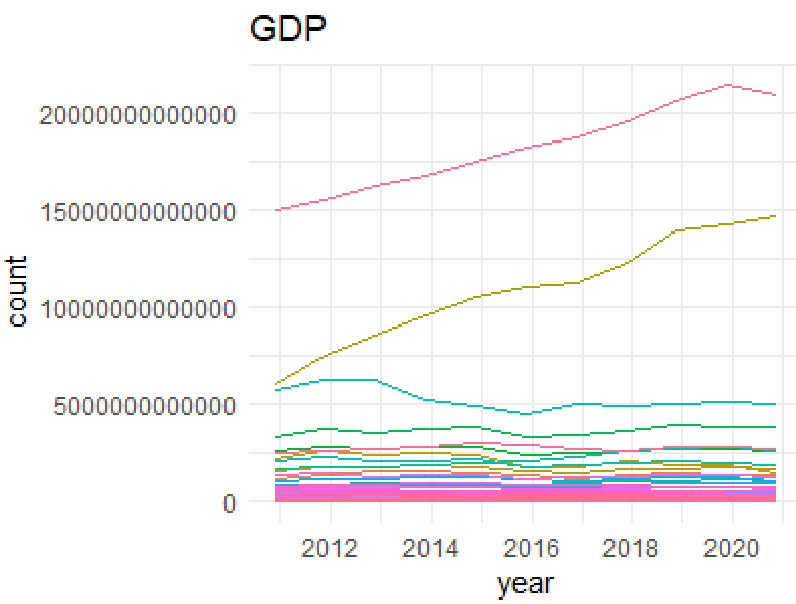
```
End = 2017
Frequency = 1
     country  year   count
2010       1 14933 40.7986
2011       1 15298 40.9456
2012       1 15664 41.5003
2013       1 16029 41.1921
2014       1 16394 42.1842
2015       1 16759 42.5502
2016       1 17125 42.8273
2017       1 17490 43.1128
```

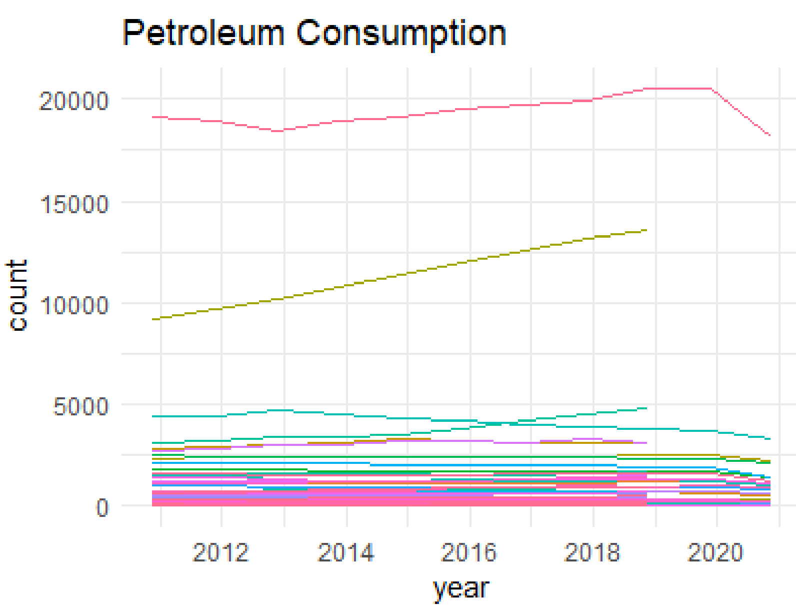
In the above time series plots, we can see:

- The first plot (GDP) shows most countries having a steady wave of GDP, with a couple of outliers (most notably, Austria).
- The second plot, material footprint, indicates that most countries have held steady from 2010-2017.

The accompanying output from the time series equation is also present.

GDP and Petroleum/Other Liquids Consumed (2010-2020)






```
Time Series:
Start = 2010
```

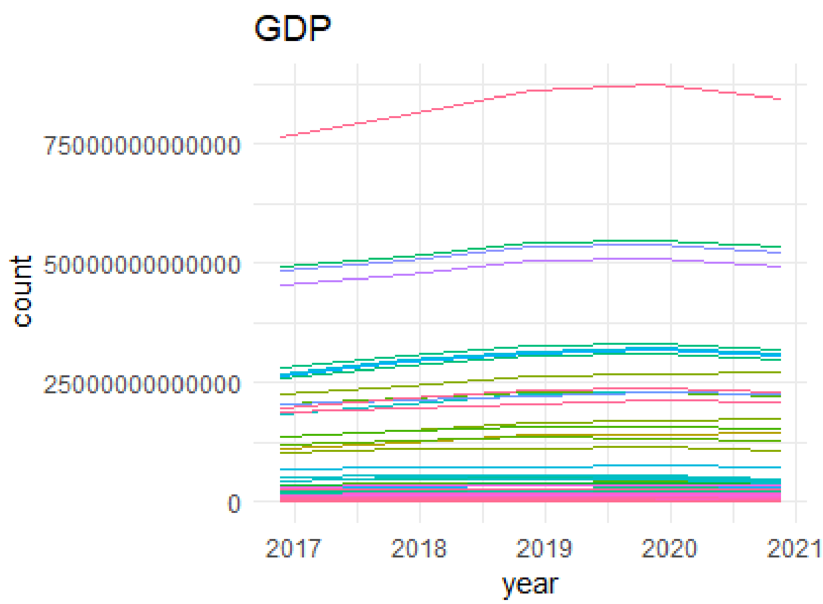
```
End = 2020
Frequency = 1
     country  year count
2010       1 14933 42.76
2011       1 15298 55.82
2012       1 15664 49.06
2013       1 16029 35.08
2014       1 16394 27.68
2015       1 16759 34.91
2016       1 17125 25.59
2017       1 17490 23.63
2018       1 17855 23.15
2019       1 18220    NA
2020       1 18586    NA
```

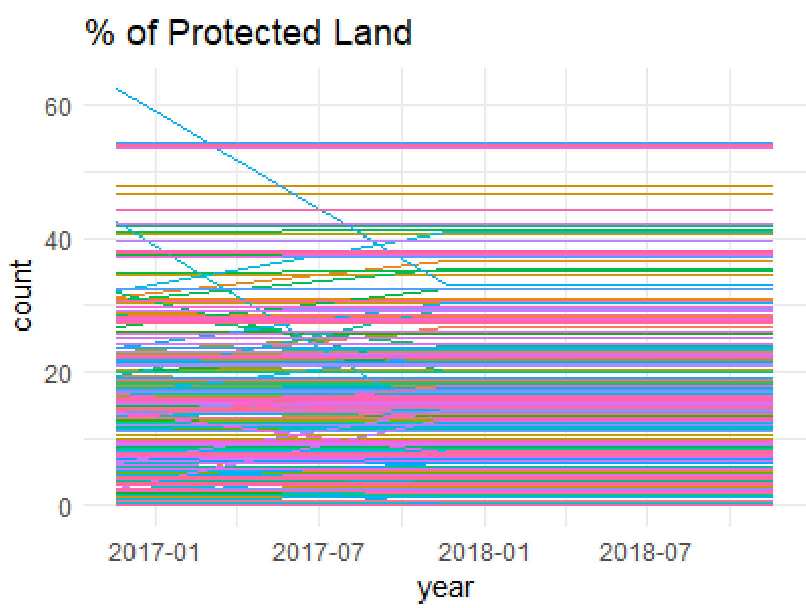
In the above time series plots, we can see:

- The first plot (which is GDP), shows steady waves with a couple of country outliers.
- The second plot, petroleum consumption, also includes a couple of country outliers.
- Legends were not included with these plots, as too many countries existed in the GDP and petroleum datasets (both datasets possessed many countries in common with one another).

The accompanying output from the time series equation is also present.

Population and Protected Land (2016-2018)






```
Time Series:
Start = 2016
```

```
End = 2018
Frequency = 1
     country  year    count
2016       1 17125 0.100000
2017       1 17490 0.104707
2018       1 17855 0.104707
```

In the above time series plots, we can see:

- The first plot (GDP) also includes some country outliers, but increases are not drastic (pretty steady).
- The second plot, percentage of protected land, indicates that most countries have held steady across the past couple of most recent years of data collection, with a couple of outliers having a decline in protected land and then holding stagnant.
- Legends were not included with these plots, as too many countries existed in the population and land datasets (both datasets possessed many countries in common with one another).

The accompanying output from the time series equation is also present.

Implications

The overall implications of my report and analysis include the following:

- It is not totally conclusive the impacts that both population and GDP may have on the dependent variables in question here.
- However, it is important to note that the strongest effects were seen in the relationships between population and petroleum consumption as well as GDP and petroleum consumption.
- The findings herein merely show a small snapshot of data for a small number of variables (perhaps variables that are not necessarily the best indicators), and therefore should be considered as such. In any case, these findings (or lack thereof) may urge others to expand upon this research in an endeavor to validate or refute it.
- No matter the case, it would behoove any of us to proceed with caution when presented with findings related to a problem such as the threat to our planet and the environment.
- Overall, the implication of these analyses indicate it likely is not enough, and it would be paramount for others to conduct additional research and analyses.

Limitations

This analysis was very rudimentary in the grand scheme of research and data science, and therefore was certainly not without its limitations which include but are not limited to:

- Not enough data was available in some of my datasets to conduct a better longitudinal study or better time series analysis. With more data across more years, the findings may have been different from what is present here in this analysis.

- The fact that not all of the yearly data matched or was the same across all variables is also a limitation, and additional sources of data regarding this variables with more similar timeframes may be warranted.
- In addition to simple correlational analyses and time series plots, it would be important to expand upon these analyses and conduct correlational tests between the different time series, as this may offer more robust correlational analyses than the simpler ones I have conducted.
- In order to answer one of my final research questions about the future of our planet, it is suggested and recommended to build a predictive model that can take historical data on these topics (historically past 2010 as well) and forecast what the future of our planet may look like. This is critical for understanding the future and how what we do now may impact that.

Concluding Remarks

Overall, many of us are aware of the need to take care of and look out for our planet and its resources - especially because many of those resources are non-renewable and therefore have a shelf-life. Similarly, this analysis does not begin to touch on the alternative methods for generating or creating new resources, which is another realm to consider entirely - albeit it goes hand in hand with the management and utilization of our current non-renewable resources. While a few of the variables present in this analysis did hold moderately strong relationships with both GDP and population, by and large only a couple (petroleum) were significant strong relationships (and a couple were quite weak). Given this, more research and analyses are likely required to obtain a better picture and landscape of our earth's current state, thus granting more accurate insights for model building and forecasting. No matter the case, our planet's well-being is a hot-button topic that deserves to remain at the forefront of discussions and research - after all, it is our only home.

References

1. IPCC. (2021). AR6 climate change 2021: The physical science basis. IPCC. https://www.ipcc.ch/report/ar6/wg1/#TS
2. Please note: Also see the references listed and included in the "How to Import and Clean My Data" section. Additional resources will be included in the final cumulative project step as needed and as relevant.