Contemporary international relations has been heavily focused on the impending climate crisis throughout the last decade. From signing international climate agreements in Paris to developing more-effective emissions capture technology, it certainly feels like the world stage is converging toward collectively acting to fight climate change. However, the story is never so simple. Just last year, the United States pulled out of the Paris Climate Agreement, citing protection of jobs in Midwestern America.[1] Other countries like the UK are considering leaving as well amidst reports that the world is collectively falling behind the standards set back in 2016.[2] Throughout many lectures, Professor Onder has often stressed the interdependent nature of the world in which we now live, fueled by globalized relationships. The question of interest to me is as follows: what aspect of globalization is having the heaviest impact on the degradation of the planet? If we are more aware of what is impacting our carbon footprint, collectively as nations, we will likely be better equipped to respond to the climate crisis because we will know what we must target. Therefore, I have decided to analyze an ecological footprint database in order to statistically support which globalized activity is most associated with planetary harm.

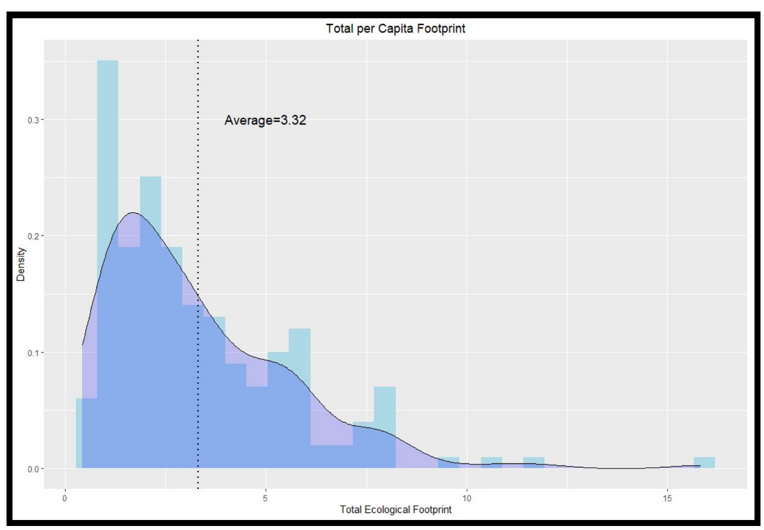
Before I begin my analysis, I will first highlight the parameters of my data, as well as some crucial assumptions that will lay the foundation for my argument. I utilize **R** in order to conduct a statistical analysis on data from the Global Footprint Network, which contains data on population, GDP per capita, total ecological footprint for most countries on the planet, as well as a great number of other variables of lesser interest to the goals of this paper.[3] I will also be using a second dataset from the US Census Bureau (acquired from DataWeb), which contains trade balance information between the United States and most countries listed in the Global Footprint Network data.[4] I have trade balance data all the way back to the year 2000, which should prove useful in our analysis. Since the US is (currently) the largest economy in the world, and in recent history has come out against global climate policy like the

[1] Friedman, 2019.
[2] Plumer & Popovich, 2018.
[3] See **References** section for Data Access and Code for Reproducibility.
[4] DataWeb, 2020.

Paris Agreement, having a trade balance with the US as a metric of interest should prove useful in checking where our priorities lie. Both of these datasets have been merged, and the first few rows (for reference) are available in the supporting information, right after the references, in Table 1.[5] The main variable of choice here is total ecological footprint simply because it is the strongest, most robust metric to view the impact on the planet. Variables such as Earths needed or Countries needed are mostly synonymous, but ecological footprint is at a per capita level, so it is more appealing as a main variable.

To begin our analysis, we first must check for missing data. Figure 1 on the right displays a table listing the proportion of missing data (if any) for each variable in our merged dataset. Here, we are looking for equal proportions to ensure that none of our data has a significantly larger number of missing values over another. All of our predictors in the Global Footprint Network contain a fairly equal proportion of missing data, other than the Human Development Index (HDI). Moreover, despite the proportion of missing data being higher for trade balance with the US, they are all quite equal in missingness, thus it is reasonable for us to continue with our analysis, so long as we are aware that HDI has a slightly higher proportion of missing data.

In this analysis, I am primarily going to focus on the variable "$Total.Ecological.Footprint" because this is the most robust and all-encompassing metric for environmental impact by country. Figure 2



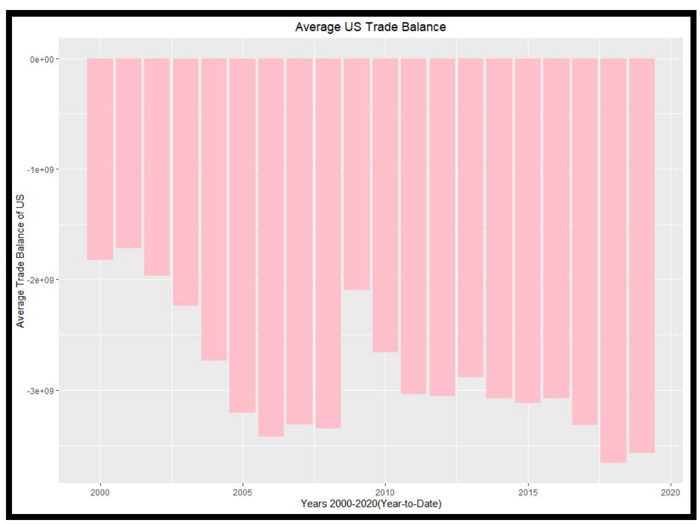
[5] All tables and figures are listed (slightly bigger for viewing) past the references section; PDFs of the figures are available on the Github provided here and in the references section.

on the right shows a histogram with a density plot overlay of total ecological footprint, and we can see the results are heavily skewed towards a smaller footprint. I include an average total footprint for all 188 countries in our data, 3.32, which checks out when considering the density plot over the histogram mostly exists around that average. For comparison, the US has an ecological footprint of 8.22, which is much higher than the international average. Interestingly, there are a healthy number of outliers, some stretching past a footprint value of 10. Does this mean that larger countries simply have a larger ecological impact? Table 2 on the right summarises the footprint per person for the top fifteen polluters in the dataset, and as we can see, this list is far from a simple population issue. For context, I have included the population size (in millions) next to these nations, and most of the top fifteen are staggeringly small. Thus, we still must wonder which is the greatest predictor of total ecological footprint.

	Country <fctr>	Population(Millions) <dbl>	Footprint.Per.Person <dbl>
1	Luxembourg	0.520	15.82
2	Aruba	0.100	11.88
3	Qatar	2.050	10.80
4	Australia	23.050	9.31
5	United States of America	317.500	8.22
6	Canada	34.840	8.17
7	Kuwait	3.250	8.13
8	Singapore	5.300	7.97
9	United Arab Emirates	9.206	7.93
10	Trinidad and Tobago	1.340	7.92
11	Montserrat	0.000	7.78
12	Oman	3.310	7.52
13	Bahrain	1.320	7.49
14	Belgium	11.060	7.44
15	Sweden	9.510	7.25

15 rows

Figure 3 displays trade balance between the US and its international trading partners from 2000 to 2020 (year-to-date). As we can see, we have not had a positive international trade balance for the past 20 years, and this trade has fairly consistently decreased over most of this period. Therefore, if the fifth largest polluter in terms of its ecological footprint has been losing money in trade for nearly two decades, and population does not seem to be a strong metric to judge footprint, what is the driving force behind a high footprint? To answer this question, I begin a regression analysis.



On the right, we can see a simple Pearson's correlation of total ecological footprint on GDP per capita. I chose GDP per capita for a number of reasons. First, our fifteen top polluters were composed of many wealthy nations, so GDP per capita would likely be a strong predictor. Moreover, both total ecological footprint and GDP per capita are measures of individual impact, further making them appropriate to compare. As we can see, their correlation is quite high (~0.8); moreover, the 95 percent confidence interval *contains no value of zero*, so we can sufficiently suggest that GDP per capita is a strong predictor of ecological footprint per person. To further instill this point, I conduct a simple linear regression of Total Ecological Footprint on GDP per capita, and as we can see, their correlation in the model (Multiple R^2=0.6391) is high, and the p-value is far below the one percent level. Therefore, we can further conclude that GDP per capita is a significant linear predictor of Total Ecological Footprint.

```
        Pearson's product-moment correlation

data:  dat$GDP.per.Capita and dat$Total.Ecological.Footprint
t = 17.4, df = 171, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7382666 0.8475257
sample estimates:
     cor
0.799412
```

```
Call:
lm(formula = Total.Ecological.Footprint ~ GDP.per.Capita, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-5.9447 -0.8771 -0.2693  0.6655  4.2801

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.010e+00  1.292e-01   15.56   <2e-16 ***
GDP.per.Capita 8.899e-05  5.114e-06   17.40   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.404 on 171 degrees of freedom
  (15 observations deleted due to missingness)
Multiple R-squared:  0.6391,	Adjusted R-squared:  0.6369
F-statistic: 302.8 on 1 and 171 DF,  p-value: < 2.2e-16
```

For a bit of extra analysis and robustness, I also include a multiple linear regression on the right. Here, I am still using GDP per capita as a predictor, but I also include population (in millions) and a year sampled at random. My sample (using **set.seed()** for reproducibility) selected 2013 at random, so that year of trade balance of the US is included. As we can see, GDP

```
Call:
lm(formula = Total.Ecological.Footprint ~ GDP.per.Capita * Population..millions. *
    `Year 2013`, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-5.7802 -0.8584 -0.2233  0.6159  3.7162

Coefficients:
                                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                                   2.044e+00  1.491e-01  13.711   <2e-16 ***
GDP.per.Capita                                8.768e-05  6.096e-06  14.382   <2e-16 ***
Population..millions.                        -1.659e-03  1.560e-03  -1.063    0.289
`Year 2013`                                  -1.134e-11  2.744e-11  -0.413    0.680
GDP.per.Capita:Population..millions.         -8.036e-08  3.094e-07  -0.260    0.795
GDP.per.Capita:`Year 2013`                    3.180e-16  7.723e-16   0.412    0.681
Population..millions.:`Year 2013`            -5.776e-15  4.528e-14  -0.128    0.899
GDP.per.Capita:Population..millions.:`Year 2013`  7.177e-19  7.761e-18   0.092    0.926
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.407 on 136 degrees of freedom
  (44 observations deleted due to missingness)
Multiple R-squared:  0.6627,	Adjusted R-squared:  0.6453
F-statistic: 38.17 on 7 and 136 DF,  p-value: < 2.2e-16
```

per capita is still the only statistically significant predictor in this model. While our adjusted R^2 did increase from $model_1$ to $model_2$, this is likely because the model is now better at identifying which predictors do well estimating total ecological footprint. Even when our GDP variable is interacted with the other variables in the model, none of them come close enough to being considered strong predictors. Thus, in line with our previous observation of trade balance with the US, trade balance is not a significant predictor of ecological footprint, nor is population size (in millions), as we suspected in our viewing of the top fifteen footprints in our data.

To conclude, GDP is a statistically significant, highly correlated predictor of total ecological footprint per capita, and the reason behind this is intuitive. If the people of a given country have more money, more resources, or a higher standard of living, on average, they will have a higher, more-negative impact on the planet. For example, in the US, we generally have access to constant electricity, plentiful grocery stores, gas-powered vehicles, and many other environmentally-damaging entities. So, it would make sense that we would impact the environment more than someone without these things in another country. In this analysis, we have no support for higher populations leading to higher ecological footprints, so fairly conventional views that large populations, such as China or India, having larger impacts are not supported. Moreover, globalization as a whole may not be responsible for the increased environmental impact we have seen. GDP per capita is likely to increase if we are more connected because we have more access to resources, material goods, trade markets, etc… However, as we saw with the trade balance of the US vs. the world stage, not everything under the umbrella of globalization is having an equally-damaging impact on our environment. To move forward, we must take action to reduce our environmental impact as citizens of high GDP per capita countries. Whether that be using less water, changing dietary habits, or a number of other lifestyle changes, we must act in order to reduce our ecological footprint. Both domestically and internationally, we can use our globalized world to bring about the change necessary to maintain the sanctity of our planet.

References

DataWeb. (2020). "Trade Flow, Trade Balance, and Classification System." *Compiled from the US Census Bureau.* U.S. Department of Commerce. Retrieved from link on 9 April 2020.

Friedman, Lisa. (4 Nov., 2019). "Trump Serves Notice to Quit Paris Climate Agreement." *The New York Times.* Retrieved from link on 9 April, 2020.

Global Footprint Network. (2020). "2016 Global Ecological Footprint: Does your country consume More resources than it produces in a year?" *Global Footprint Network.* Dataset retrieved From Kaggle. Retrieved from link on 9 April 2020.

Plumer, Brad & Nadja Popovich. (7 Dec., 2018). "The World Still Isn't Meeting Its Climate Goals." *The New York Times.* Retrieved from link on 9 April, 2020.

***Note on Data and Code:** Both the data used for this analysis and the code can be found at this Github repository, linked here. This repository contains data retrieved from Kaggle and DataWeb, which are referenced above. RMarkdown was used for better printouts of tables below, but the code found at this repository can be completely reproduced with the proper working directory and libraries installed. **Anything commented out was not used for the paper, but considered during the analysis.**

****Note on DataWeb:** The data acquired from DataWeb was selected using their online toolkit, which requires an account. It is free to create and download the data afterward. I used the following settings in the toolkit, so the downloaded data can be reproduced:

- Step1: Trade Balance, HTS items;
- Step2: Total Exports FAS, Actual and Annual aggregation;
- Step3: Using all Countries, Displaying Separately;
- Step4: HTS-2 Aggregation Level, Aggregate all Commodities;
- Other: No Percentage Column, Displaying All Rows.

Tables and Figures:

Table 1: Head of Merged Data (for Reference):

	Country	Region	Population..millions.	HDI	GDP.per.Capita	Cropland.Footprint	Grazing.Footprint	Forest.Footprint	Carbon.Footprint
	<fctr>	<fctr>	<dbl>	<dbl>	<fctr>	<dbl>	<dbl>	<dbl>	<dbl>
1	Afghanistan	Middle East/Central Asia	29.82	0.46	$614.66	0.30	0.20	0.08	0.18
2	Albania	Northern/Eastern Europe	3.16	0.73	$4,534.37	0.78	0.22	0.25	0.87
3	Algeria	Africa	38.48	0.73	$5,430.57	0.60	0.16	0.17	1.14
4	Angola	Africa	20.82	0.52	$4,665.91	0.33	0.15	0.12	0.20
5	Antigua and Barbuda	Latin America	0.09	0.78	$13,205.10	NA	NA	NA	NA
6	Argentina	Latin America	41.09	0.83	$13,540.00	0.78	0.79	0.29	1.08

6 rows | 1-10 of 42 columns

Fish.Footprint	Total.Ecological.Footprint	Cropland	Grazing.Land	Forest.Land	Fishing.Water	Urban.Land	Total.Biocapacity	Biocapacity.Deficit.or.Reserve	Earths.Required
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
0.00	0.79	0.24	0.20	0.02	0.00	0.04	0.50	-0.30	0.46
0.02	2.21	0.55	0.21	0.29	0.07	0.06	1.18	-1.03	1.27
0.01	2.12	0.24	0.27	0.03	0.01	0.03	0.59	-1.53	1.22
0.09	0.93	0.20	1.42	0.64	0.26	0.04	2.55	1.61	0.54
NA	5.38	NA	NA	NA	NA	NA	0.94	-4.44	3.11
0.10	3.14	2.64	1.86	0.66	1.67	0.10	6.92	3.78	1.82

6 rows | 11-20 of 42 columns

Earths.Required	Countries.Required	Data.Quality	Year 2000	Year 2001	Year 2002	Year 2003	Year 2004	Year 2005	Year 2006	Year 2007	Year 2008
<dbl>	<dbl>	<fctr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
0.46	1.60	6	7388169	5046058	76117226	4592211	132832845	193961136	372336814	414164720	396980773
1.27	1.87	6	13169278	8044924	9195167	5346730	8679422	-18704203	3684980	24441176	28253173
1.22	3.61	5	-1856829642	-1647317983	-1380275406	-4265567506	-6437365253	-9193318409	-14390645242	-16163545046	-18111541546
0.54	0.37	6	-3331017599	-2823595799	-2742045896	-3772413215	-3926976503	-7556360531	-10169048990	-11227393676	-16793704460
3.11	5.70	2	NA	NA	NA	NA	NA	NA	NA	NA	NA
1.82	0.45	6	1597913533	912525300	-1594578954	-733930664	-359436544	-472160809	800742202	1359707410	1716410006

6 rows | 20-31 of 42 columns

Year 2009	Year 2010	Year 2011	Year 2012	Year 2013	Year 2014	Year 2015	Year 2016	Year 2017	Year 2018	Year 2019	Percent Change 2018 - 2019
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>
1395367203	2066147153	2895722389	1484790322	1364602162	738224771	455110273	880158425	927614201	1198295703	714312581	-40.39
32528986	16094068	21148258	35225371	51588239	-17871796	-129183554	-74320256	20490874	-8801831	29157437	NA
-9609040116	-13323630480	-13012310230	-8630171176	-2982167624	-2011979700	-1495815000	-1036658547	-2748698563	-3357285855	-1480240295	55.91
-7916015983	-10646080927	-12094250880	-8333294802	-7299546893	-3680544428	-1640400076	-1605814259	-1794163438	-2173290318	-418370147	80.75
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1669944473	3589381918	5395916417	5907643608	5708789255	6576552830	5410940220	3872968555	4817118349	5094431452	3161775258	-37.94

6 rows | 32-43 of 42 columns

Figure 1: Proportion of Missing Data:

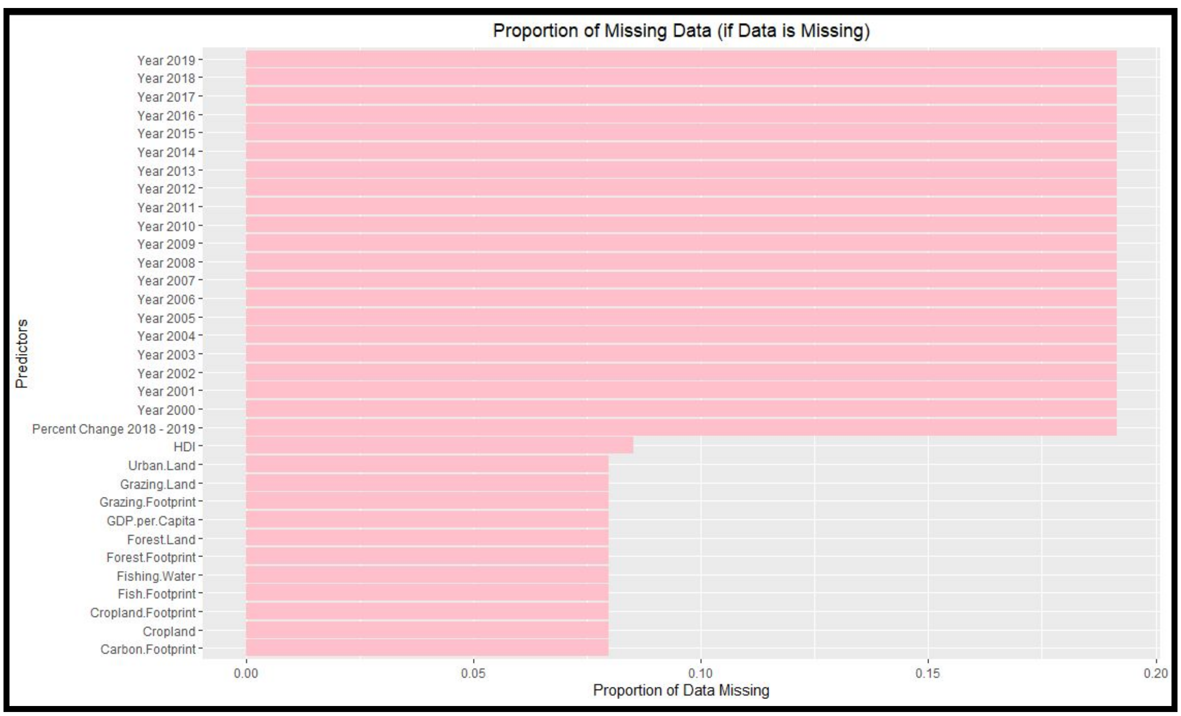


Figure 2: Total Ecological Footprint Histogram/Density Plot:

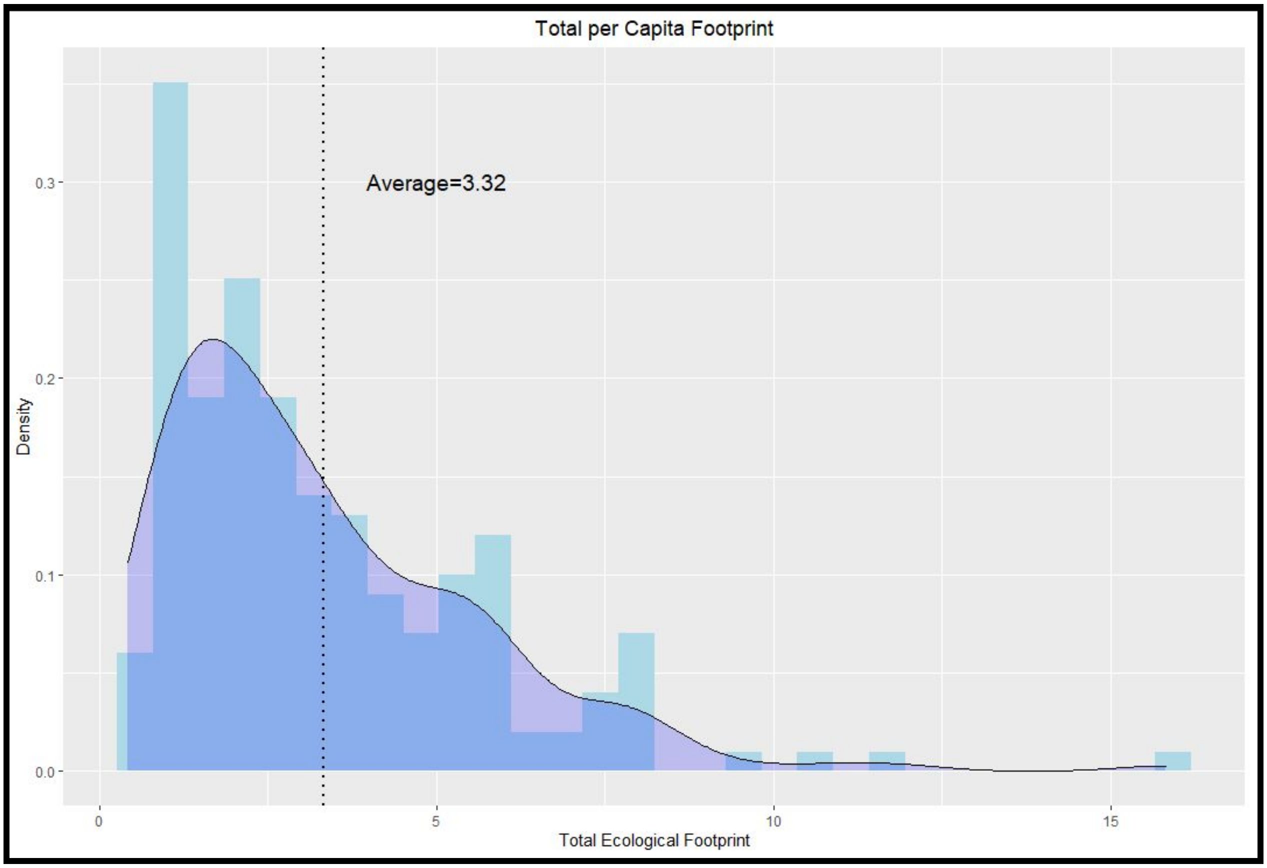


Table 2: Top Fifteen Polluters (in their Footprint) with Population (in Millions) for Comparison:

	Country <fctr>	Population(Millions) <dbl>	Footprint.Per.Person <dbl>
1	Luxembourg	0.520	15.82
2	Aruba	0.100	11.88
3	Qatar	2.050	10.80
4	Australia	23.050	9.31
5	United States of America	317.500	8.22
6	Canada	34.840	8.17
7	Kuwait	3.250	8.13
8	Singapore	5.300	7.97
9	United Arab Emirates	9.206	7.93
10	Trinidad and Tobago	1.340	7.92
11	Montserrat	0.000	7.78
12	Oman	3.310	7.52
13	Bahrain	1.320	7.49
14	Belgium	11.060	7.44
15	Sweden	9.510	7.25

15 rows

Figure 3: US Trade Balance from 2000-2020 (Year-to-Date):

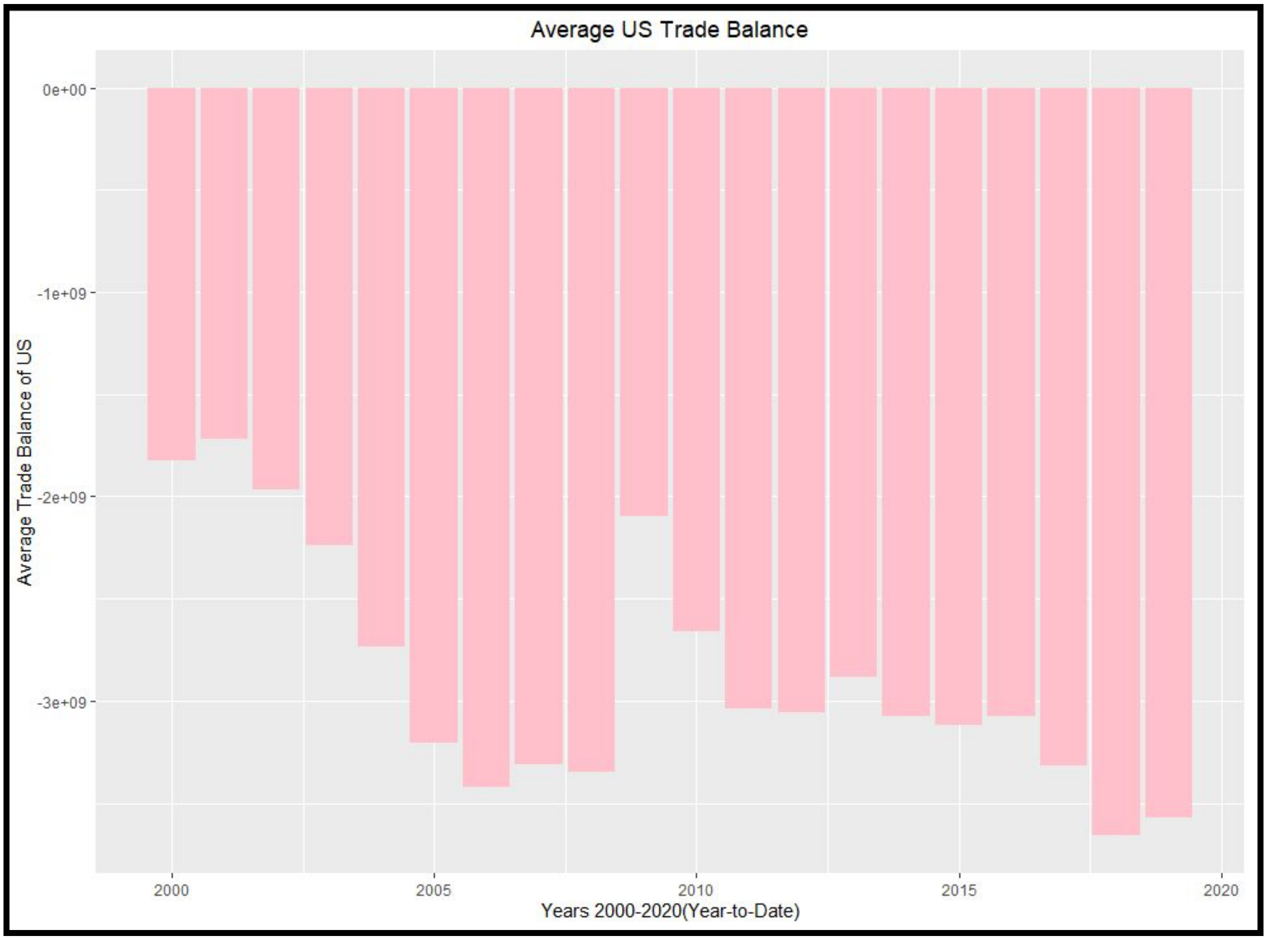


Table 3: Pearson's Correlation Between Total Ecological Footprint and GDP per Capita:

```
        Pearson's product-moment correlation

data:  dat$GDP.per.Capita and dat$Total.Ecological.Footprint
t = 17.4, df = 171, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7382666 0.8475257
sample estimates:
     cor
0.799412
```

Table 4: Simple Linear Regression of Total Footprint and GDP per Capita:

```
Call:
lm(formula = Total.Ecological.Footprint ~ GDP.per.Capita, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.9447 -0.8771 -0.2693  0.6655  4.2801 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)    2.010e+00  1.292e-01   15.56   <2e-16 ***
GDP.per.Capita 8.899e-05  5.114e-06   17.40   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.404 on 171 degrees of freedom
  (15 observations deleted due to missingness)
Multiple R-squared:  0.6391,    Adjusted R-squared:  0.6369 
F-statistic: 302.8 on 1 and 171 DF,  p-value: < 2.2e-16
```

Table 5: Multiple Linear Regression (Footprint, GDP, Population, One Year of Trade):

```
Call:
lm(formula = Total.Ecological.Footprint ~ GDP.per.Capita * Population..millions. * 
    `Year 2013`, data = dat)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.7802 -0.8584 -0.2233  0.6159  3.7162 

Coefficients:
                                                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                      2.044e+00  1.491e-01  13.711   <2e-16 ***
GDP.per.Capita                                   8.768e-05  6.096e-06  14.382   <2e-16 ***
Population..millions.                           -1.659e-03  1.560e-03  -1.063    0.289    
`Year 2013`                                     -1.134e-11  2.744e-11  -0.413    0.680    
GDP.per.Capita:Population..millions.            -8.036e-08  3.094e-07  -0.260    0.795    
GDP.per.Capita:`Year 2013`                       3.180e-16  7.723e-16   0.412    0.681    
Population..millions.:`Year 2013`               -5.776e-15  4.528e-14  -0.128    0.899    
GDP.per.Capita:Population..millions.:`Year 2013`  7.177e-19  7.761e-18   0.092    0.926    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.407 on 136 degrees of freedom
  (44 observations deleted due to missingness)
Multiple R-squared:  0.6627,    Adjusted R-squared:  0.6453 
F-statistic: 38.17 on 7 and 136 DF,  p-value: < 2.2e-16
```