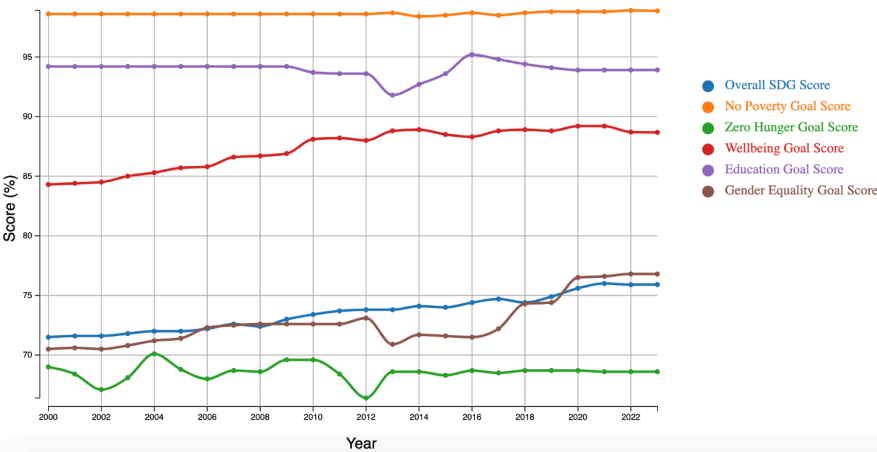


Creators of the Data Visualizations: Sanjum Sahni (ss3873), Grace Wei (gtw25), Arushi Aggarwal (aa2555)

Picture of Data Visualizations:

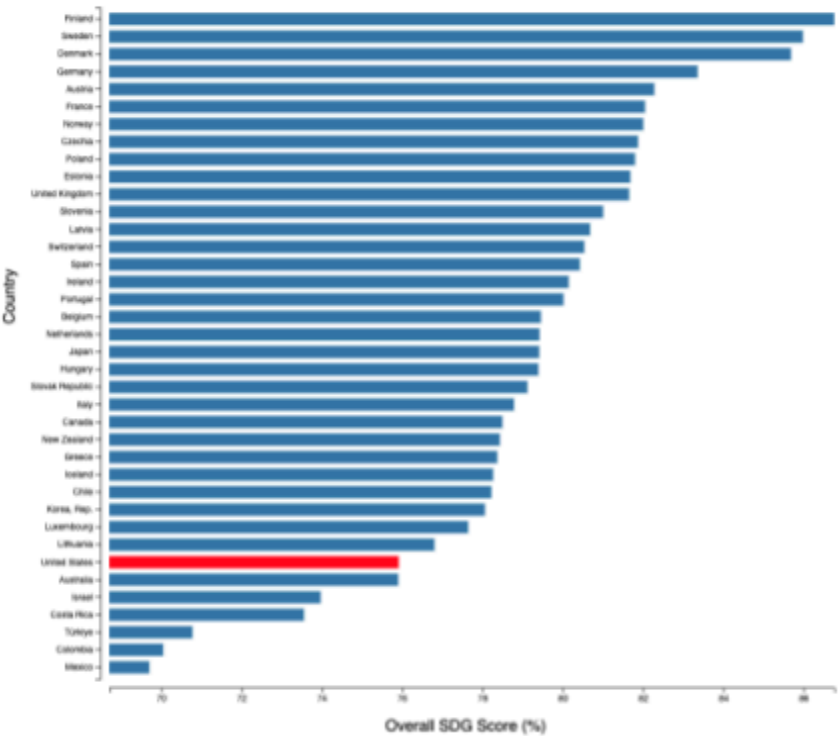
Graph 1: Visualizing Sustainable Development Goals in the United States over Time

This chart shows the United States' performance in five metrics: poverty, hunger, wellbeing, gender equality, and education from the years 2000-2022. We also plot the US's overall SDG score for these years.



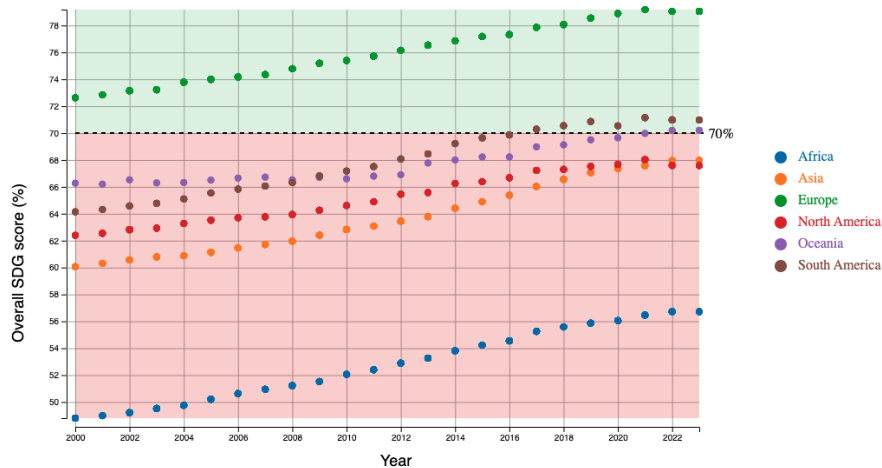
Graph 2: Visualizing Current Performance of Countries in the OECD

This graph shows overall sustainable development Scores for countries in the Organisation for Economic Cooperation and Development (OECD) in the year 2023. The United States is highlighted in red. Out of the 38 countries in the OECD, the United States has the 7th lowest overall SDG score. Top performing countries in the OECD include Finland, Sweden, Denmark, and Germany. Lowest performing countries include Costa Rica, Turkey, Colombia, and Mexico.



Graph 3: Visualizing Sustainable Development Scores for Different Continents over Time

We wanted to see if European countries really stand out in terms of meeting the SDG goals by showcasing a holistic view of all the countries in comparison. By plotting all the continents, Europe, Asia, North America, South America, Oceania, and Africa and their SDG score for each year, we were able to visualize which region of the world is performing well or not. We uncovered a 70% benchmark line through our research about how if a country scores above 70 points they are doing well. From what we can see in this chart, Europe is the only continent that has consistently scored above 70, which was not surprising and was in harmony with our earlier results.



Data Description:

We found our first sustainable development dataset on Kaggle (<https://www.kaggle.com/datasets/sazidthe1/sustainable-development-report>). This dataset's name in our data folder is `sdg_index_2000-2022.csv`. This dataset focuses on global trends in sustainable development goals from 2000 to 2022. The Sustainable Development Goals (SDGs) were designed by the United Nations in hopes of making the world a better and more sustainable place. There are 17 sustainable goals. Each country receives a score on how well they are doing with each goal; this dataset essentially keeps a record of these scores for each country over the past 22 years.

We also found our second sustainable development dataset on Kaggle. This dataset's name in our data folder is `sustainable_development_report_2023.csv`. This data set also keeps a record of each country's score for each sustainability goal but only for the year 2023.

In both of these datasets, these are the variables in common: `country_code`, `country`, `goal_1_score` up until `goal_17_score`. Each of these scores correspond to a specific sustainable development goal, which can be found in the original Kaggle page. The first dataset also has a `year` variable and a `sdg_index_score` (`sdg index score` is an overall sustainability goal score). The second dataset has a `region` variable and an `overall_score` variable (the `overall_score` variable is the same as the `sdg_index_score` variable in the previous dataset).

Going further, for third visualization we wanted to look at the holistic view of all the continents and see how they are performing in terms of the SDG scores. For this reason we performed data pre processing in python to map all the countries to a dictionary of continents that we created and below are screenshots of code for how we created the third column that mapped each continent

to a country. We then also created another column that calculated the averages of the scores for each of the countries in the continent for each year from 2000-2023 so we can plot the values.

Below are screenshots from the data pre-processing.

```
def get_continent(country):  
    return country_to_continent.get(country, 'Unknown')  
  
df = pd.read_csv('sustainable_development_report_2023.csv')  
  
df['continent'] = df['country'].apply(get_continent)  
  
df.to_csv('sustainable_development_report_with_continent_2023.csv', index=False)
```

```
[ ] merged_df = pd.merge(sdg_historical_df, sdg_2023_df[['country_code', 'continent']],  
                        on='country_code', how='inner')  
  
[ ] full_df = pd.concat([merged_df, sdg_2023_df], ignore_index=True)  
  
[ ] full_df = full_df.sort_values(['continent', 'country', 'year'])  
  
[ ] continent_scores = full_df.groupby(['continent', 'year'])['sdg_index_score'].mean().reset_index()  
    continent_scores = continent_scores.rename(columns={'sdg_index_score': 'continent_avg_score'})  
  
    # Merge continent average scores back to the main dataframe  
    result_df = pd.merge(full_df, continent_scores, on=['continent', 'year'], how='left')  
  
    # Save the result to a new CSV file  
    result_df.to_csv('combined_sdg_data_with_continent_scores.csv', index=False)
```

Design Rationale:

For all three of our data visualizations, we wanted to see the trends in how countries have been doing in terms of improving sustainability. Sustainability is crucial in protecting the environment and our homes; for this reason, we wanted to see which countries are performing well in the sustainability sector and what the trends of sustainability have been over the years.

The first data visualization created focused specifically on the United States. Since all three of us consider the United States our home country, we wanted to see how well a nation power like the US has been doing in terms of sustainability and in which categories we could improve more upon. For this reason, in this data visualization, we chose 6 scores to focus on: the `sdg_index` score (the overall score for sustainable development goals), `goal_1_score` (no poverty score), `goal_2_score` (zero hunger score), `goal_3_score` (health and wellbeing goal score), `goal_4_score` (quality education score), and `goal_5_score` (gender equality goal score). The x-axis included the

years from 2000 to 2022 and the y axis was the score scale. We plotted all of the points for all of the goal scores on the graph as circles for the markers because circles are generally good markers to use for points. For each goal, there was a line drawn through their respective points. A line graph with 6 lines on the graph was created where each line represented the United States's trend in the sustainability sector. In order to help the user distinguish between the different lines, we used an ordinal scale with `d3.schemeCategory10`. This made each line a different color. The reason we used `d3.schemeCategory10` was because it outputted colors that were easy to distinguish from for the user. Lastly, for this data visualization, we also created a legend to the side of the graph so the user could easily see what each line represents. The tradeoffs of this graph based on the way we designed it is that if the user is not aware of what the UN sustainability goals are then this graph could possibly take them more time to understand. For example, a lower zero hunger goal score is better than a higher zero hunger goal score which may be not inherently understandable to someone.

After analyzing the trends in the United States' Sustainable Development Goals, we wanted to see how the United States could compare with other countries around the world. We focused on the most recent data, so that we could get a better picture of the US's current performance. We planned to have a side-by-side bar graph of countries' overall SDG scores, as this would allow us to clearly see the differences between countries' scores. However, there were 166 countries represented in the data set, and we couldn't represent all 166 data points using bars, since that would be difficult to see and understand. So, we narrow it down to countries that are in the Organisation for Economic Cooperation and Development (OECD), which is a group of 38 countries that "stimulate economic progress and world trade", including the United States (<https://en.wikipedia.org/wiki/OECD>). These are countries that are the most similar to the United States, as they are democratic and also major economic world powers. So, we want to see how the United States is doing relative to these countries. Originally, we had created a vertical bar chart, but we later switched to a horizontal bar chart because it was a lot easier to understand which bars corresponded with which countries, made the graph more readable, and created a better use of space. We highlighted the bar representing the United States in red in order to create contrast and draw the viewer's attention. We ordered the bars top to bottom from greatest to least SDG score, in order to be able to see the ranking of countries' performances.

While being able to compare the United States, our main focus country, was helpful we wanted to see how we can gain a further down the line holistic view of the world and the global performance in meeting the SDG goals. We realized there is so much depth in understanding how the world is performing in terms of attaining these goals, and through the third visualization, we wanted to see if we can visualize the nuance of each continent's performance. As mentioned earlier, there are 166 countries so we couldn't plot all of them but to still gain the 360 perspective we decided to perform data processing as mentioned above and map each country to one of the 7 continents, North America, Asia, Europe, Africa, Oceania, and South America. Doing this

allowed us to declutter our graph from plotting all the countries yet gain a view on how every region in the world is doing. Further, this chart is supposed to be the end of the report's holistic view on everything so we planned to plot the SDG scores for each country in each of the years. To do this we calculated the main SDG scores for all countries for each year to plot overall scores for each continent. We had to also perform a tradeoff between a scatter plot analysis and a line chart. One of the reasons why we picked the scatter plot over a line chart is because we are representing multiple countries in each continent and creating a line chart led to a bit of a messy graph in interpreting each point. Additionally, we created two rectangles separated by a 70 point benchmark line. Based on research we conducted, we realized there is a 70 point benchmark to see if a country is considered doing well or not in terms of reaching this goal. Adding this line helped us add complexity to the chart and show us what continents are doing well or not in terms of meeting the goals. We separated them with green and red colors based on their positive and negative correlation values and showcased the 70% benchmark lines through the dotted line.

The Story:

The first data visualization shows us the general trends of the first 5 sustainability goals in the United States. We can see that the United States has had a relatively low zero hunger goal score over the years which is a good sign. Over the years, the gender equality score and the overall SDG score has increased but it is still relatively low. The well being goal score has also generally increased and it is relatively high compared to the other scores. The education score in the United States has generally been very high with a dip in 2013. The no poverty goal score is very high and has continuously remained very high over the past 22 years; this is not a good sign as it means there is a significant amount of poverty in the US and it has not improved over the years. It is surprising that the no poverty score is as high as it is; when many people think of the US they think of it as a developed country so it is surprising that there still exists so much poverty here. The trends listed above are the insights we want our viewer to take away from this graph.

The second data visualization shows us the United States' most recent performance overall, compared to the other nations in the OECD. As the graph is ordered, we see that out of the 38 countries in the organization, the United States has the 7th lowest overall SDG score, at a score of about 76%. While this does not give us the details as to which metrics the United States is doing well or poorly in, it gives us the big picture on how the US is performing relative to other similar major countries. Countries with the highest scores include Finland, Sweden, Denmark, Germany, and Austria – all countries located within western Europe (and the top 3 are Scandinavian countries).

Noticing this trend with Europe, we wanted to see if European countries really stand out in terms of meeting the SDG goals by showcasing a holistic view of all the countries in comparison. By plotting all the continents, Europe, Asia, North America, South America, Oceania, and Africa and their SDG score for each year, we were able to visualize which region of the world is

performing well or not. With this, we were able to confirm our suspicions about Europe doing well since they performed marginally better in comparison to the other continents. We uncovered a 70% benchmark line through our research about how if a country scores above 70 points they are doing well. From what we can see in this chart, Europe is the only continent that has consistently scored above 70, which was not surprising and was in harmony with our earlier results. It was surprising however to see how poorly the other continents were doing as they were all below the 70% benchmark line. It makes sense for some continents to not be able to meet the goals as they are not economic powerhouses such as the United States, but overall from all over visualizations we have been to witness that the United States performs poorly in terms of meeting the sustainable development goals.

Through our graph we were able to convey to the viewer which countries were spending their time and resources in aiding the SDG goals and despite the US's vast amount of resources, they weren't able to meet the benchmark set by the United Nations to meet the goals that seek to aid humanity.

Team Contributions:

All of us contributed significantly to this project. We met up each week to discuss plans of what we wanted to accomplish each week. Initially, we all worked together to come up with a project idea we wanted to focus on. Then, we all worked together to find datasets that we could potentially be interested in using for our project. Then, we all planned out how we wanted to create our data visualizations and what exactly we wanted on the data visualizations. The specific contributions of what each team member contributed will be listed below. The parts of the project that took the most time were actually finding a dataset that we wanted to work with, coming up with the data visualizations we wanted to create, and then actually creating the data visualizations. We had some ideas that we wanted to do with our data visualizations. However, we did not necessarily know how to go about carrying through with these ideas. For example, we did not know how to make a legend so that took some time to figure out. Additionally, we did not know how to create a different line for each goal in the first visualization; we discussed some strategies and eventually came up with a solution that worked for us. We also were not sure about how to merge two datasets so we had to do some research to figure that out. These were our most time consuming portions of the project.

Sanjum's Contributions and Time Spent:

- Sanjum helped come up with an idea for the project and helped look for datasets. Eventually, she found a useful dataset on Kaggle that we used for this project. Sanjum also planned out which data visualizations that were to be created. Sanjum was mainly in charge of creating the first visualization. She figured out how to filter the data and create different line graphs for each of the 6 scores we wanted to focus on. Sanjum also figured out how to make labels on the graph, and contributed significantly to the report.

- Time Spent on Project Overall: 9-10 hours

Grace's Contributions and Time Spent:

- Grace helped come up with an idea for the project and helped look for datasets. Grace also planned out which data visualizations that were to be created. Grace was mainly in charge of creating the second data visualization which is a side by side bar graph; she figured out how to create a side by side line graph which we had not done before. Grace also helped figure out how to create a legend for some of the visualizations. She also worked on styling the index page and added to the report.

- Time Spent on Project Overall: 9-10 hours

Arushi's Contributions and Time Spent:

- Arushi helped come up with an idea for the project and helped look for datasets. Arushi also planned out which data visualizations should be created. Arushi was mainly in charge of the third data visualization. She figured out how to merge two datasets together and how to add a new column to the dataset. Additionally, she figured out how to separate the data by continent so she could create her data visualization. She also figured how to showcase the 70% benchmark line with creating two red and green rectangles, and also worked on the report.
- Time Spent on Project Overall: 9-10 hours