SAPIENT is a Command Line Interface Application designed by the **Stockholm Environment Institute (SEI)** to help decision-makers and investors assess the level of alignment that a set of projects have with the **Sustainable Development Goals (SDGs)** or the **European Taxonomy for Sustainable Activities (EU Taxonomy)**. It reads the texts of a folder containing PDF files and creates tabular and graphical outputs on the documents' correlation with the SDGs and the EU Taxonomy.

Citation

If you use SAPIENT, please cite us as:

> Lobos Alva, I.; Cárdenas Vélez, M.; & Betancur Jaramillo, J. C.; Hernández Orozco, E.; Maestre Másmela, D. (2022). SAPIENT - Sustainability mApper for Planning and InvestmENT. https://github.com/Camilo-Betancur/SAPIENT.

Installation

Installing R language and other requirements

First, you must install in your computer and an IDE such as To use SAPIENT, you need to install R language and an Integrated Development Environment (IDE) such as RStudio. Then, you can install the required libraries by running the following command in an R terminal:

```
install.packages(c("caTools", "cli", "DBI", "ggplot2",
                   "ggraph", "glue", "here", "igraph",
```

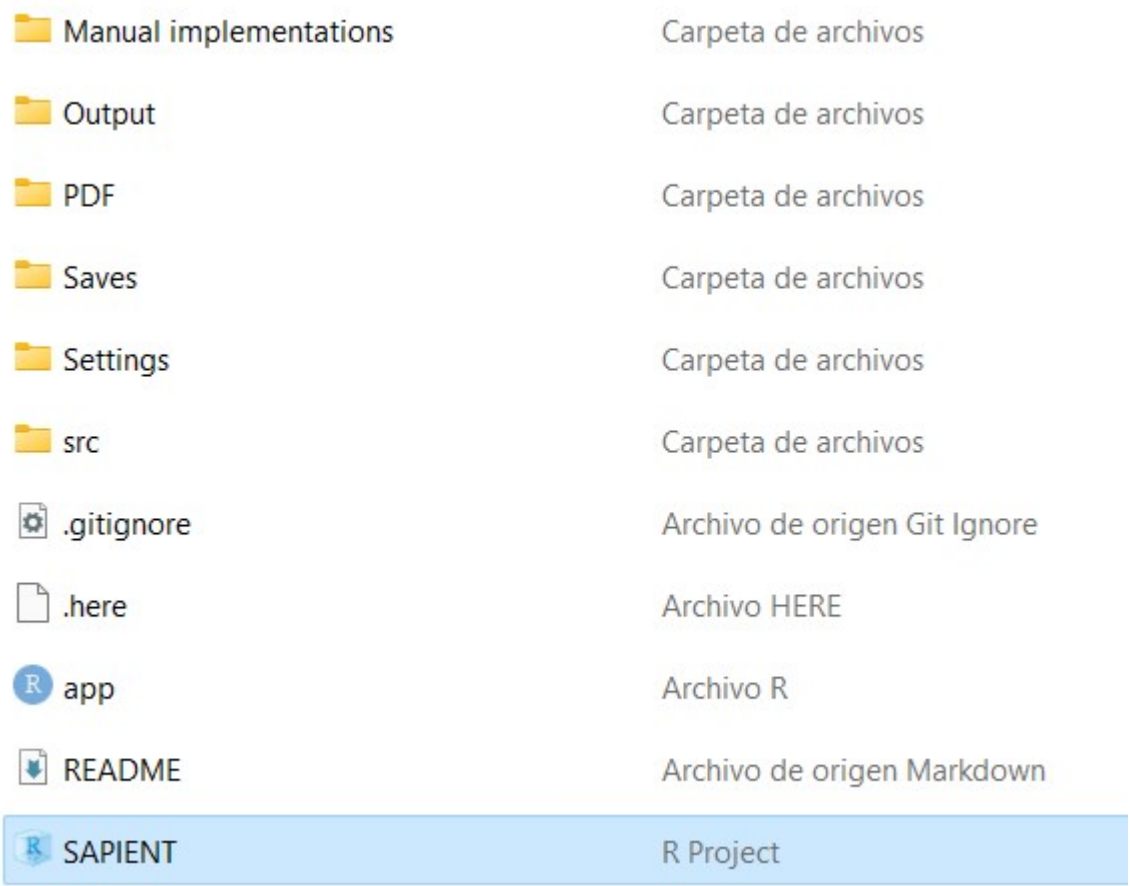
```
                    "jsonlite", "pdftools", "randomForest",
                    "readr", "RSQLite", "showtext", "SnowballC",
                    "sysfonts", "tidygraph", "tidytext",
                    "tidyverse", "tm"
                    ))
```

You can then clone or download the SAPIENT repository from GitHub.

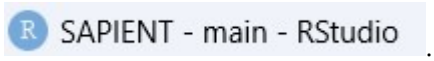
SAPIENT works best with **R 4.2.2** or newer and a machine with 8 GB RAM and 2100 MHz 6-core processor or better if you want to analyze many documents.

Getting Started

When you download/clone the SAPIENT repository, you should get a folder containing the following:

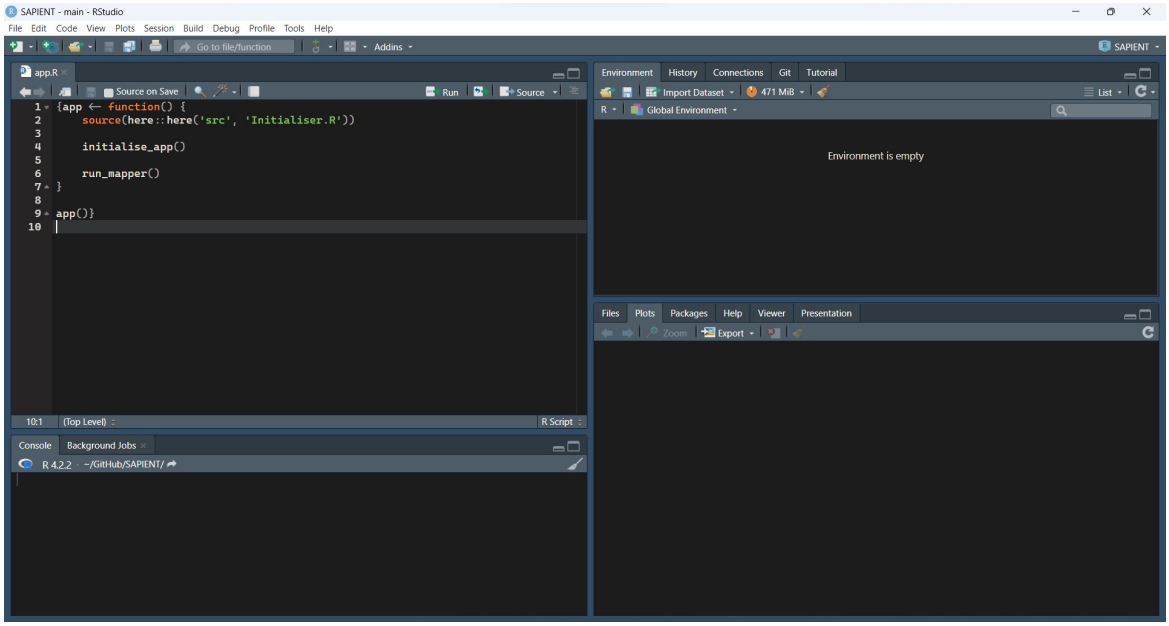


First, you must **open the R Project file named "SAPIENT (*Sapient.RProj*)"**. A new RStudio window will open with the title "SAPIENT - main - RStudio"



Then, you must **go back to the SAPIENT main folder and open the file named "*app.R*"**.

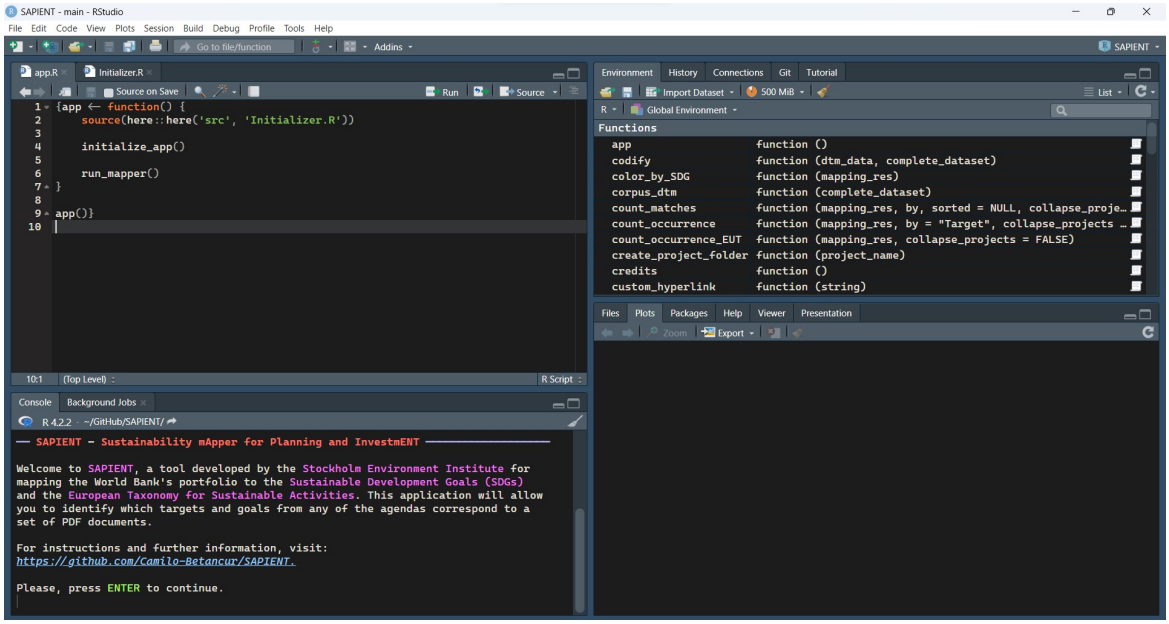
Now, your RStudio window should look something like this:



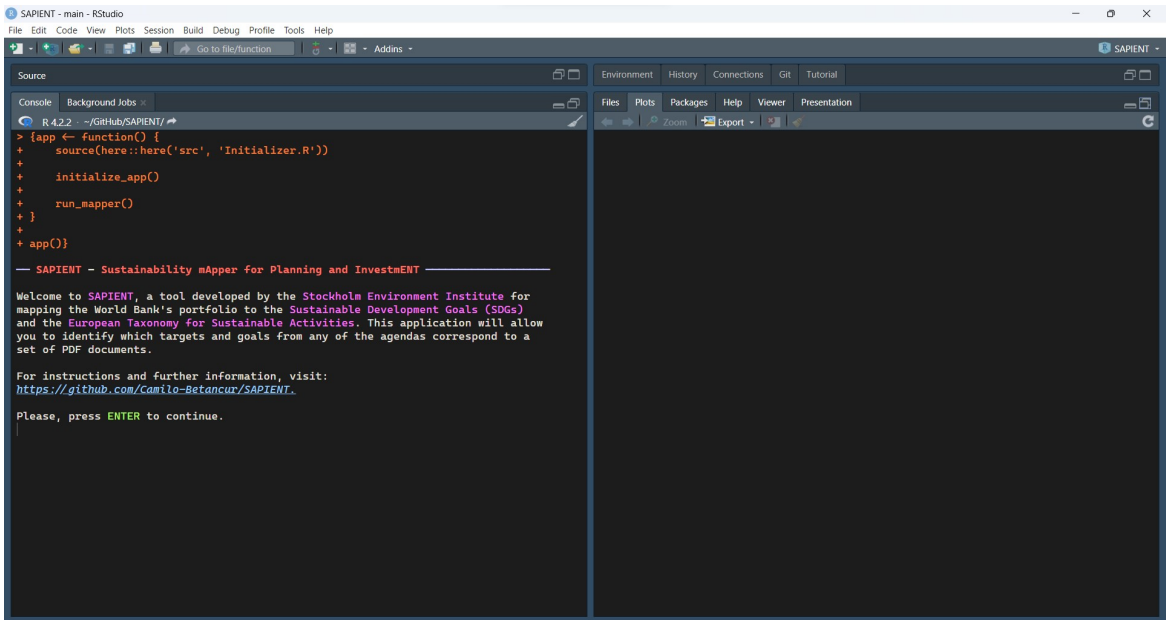
To run the app, you must locate the text cursor at the beggining of the file and press Ctrl + INTRO or click on the "Run" icon in the RStudio interface.

```
{app ← function() {
    source(here::here('src', 'Initialiser.R'))

    initialise_app()

    run_mapper()
}

app()}
```

After running the script, you should get something like this:



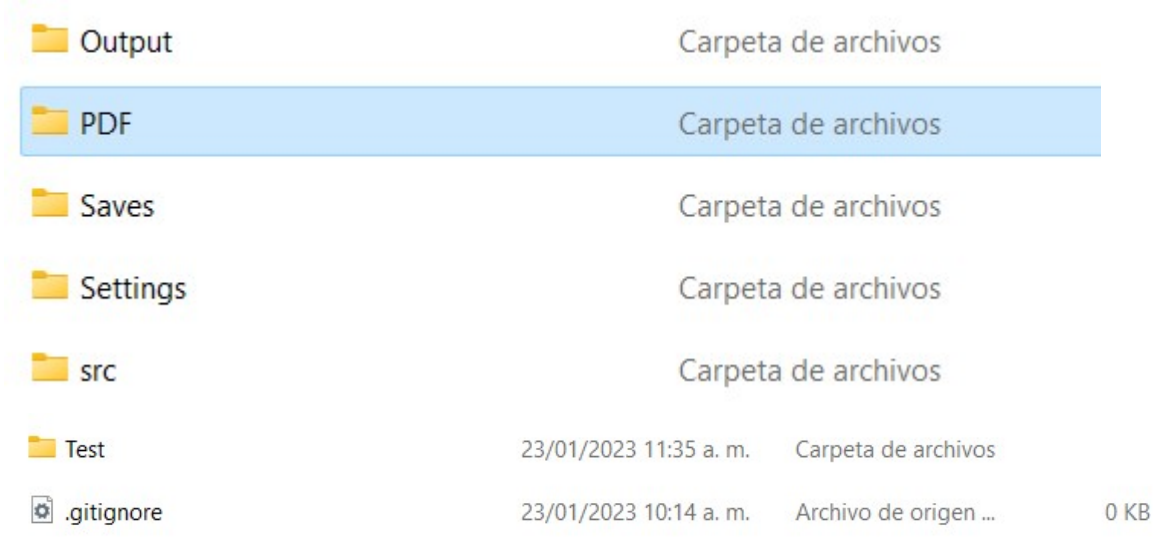
However, as the application runs on the Console and Plots tabs, we suggest you to maximize both.



Using the App

Creating the folder for the PDFs

First, you need to copy the PDF files into the SAPIENT directory. To do this, please go to the main directory and look for the **PDF** folder. Inside it, you must create a folder that will contain the PDF files you want to analyze. You will not be able to select which files of the folder to analyze, so we invite you to think about folders as individual analysis corpus. For our example, we will create a Folder named ***Test***.



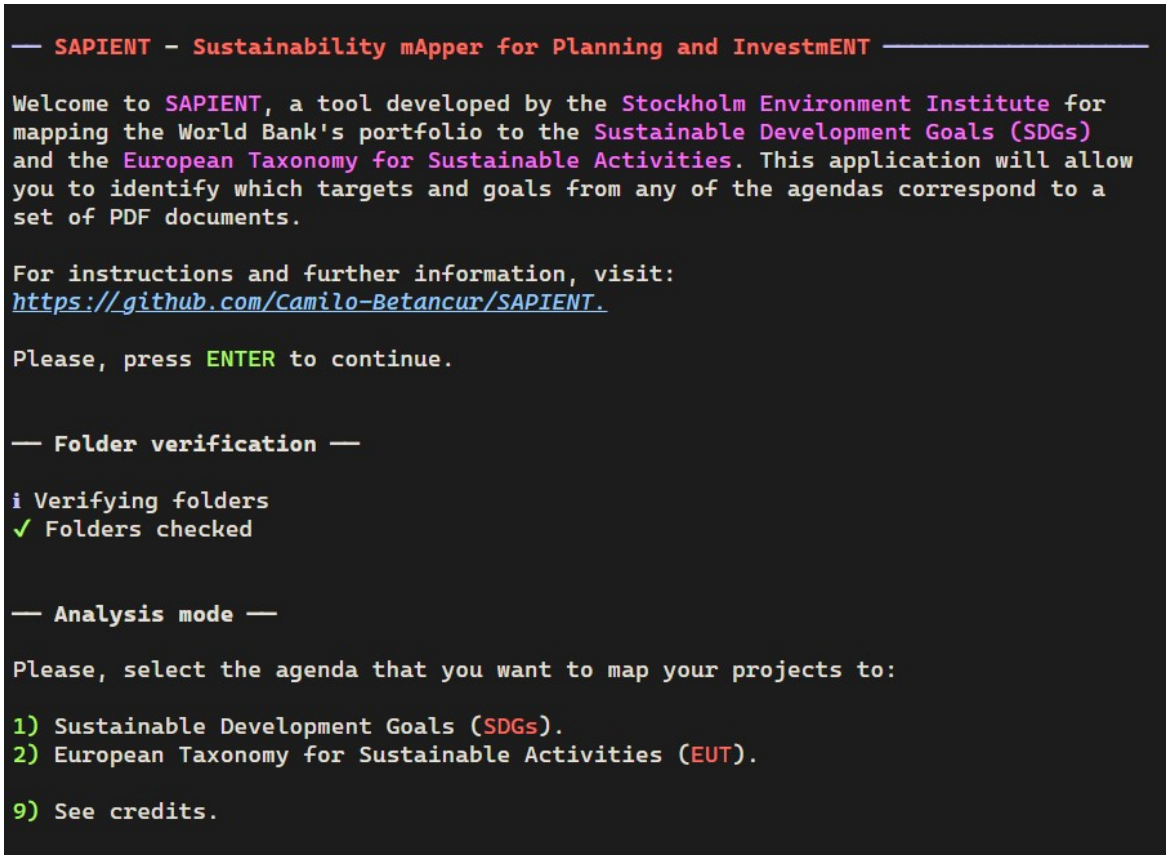
Note: SAPIENT will only be able to analyze the folders that are located in ***SAPIENT/PDF/***.

Running the app

When you run the app, you will see a greeting and contextual information about SAPIENT. Also, you will see a message that indicates the status of the fundamental app directories and a menu for selecting the analysis mode.

The app will give you three options, from which you will have to select writing the option number and pressing INTRO.

- **Option 1:** Sustainable Development Goals (SDGs).
- **Option 2:** European Taxonomy for Sustainable Activities (EUT)
- **Option 9:** Credits.



After choosing an analysis mode, the application will ask you to tell it where will it find the texts to map. They can be located in a folder with PDFs (see Creating the folder for the PDFs) or retrieved from a previously-saved PDF extraction.

For this example, we will choose to **Start analysis from scratch**.

```
── Analysis mode ──

Please, select the agenda that you want to map your projects to:

1) Sustainable Development Goals (SDGs).
2) European Taxonomy for Sustainable Activities (EUT).

9) See credits.
1

── Data source ──

Please, select the source of the projects you want to analyze:

1) Start analysis from scratch.
2) Read saved data.
3) Train the classification model.
1

Write the name of the folder that contains the PDF files and press ENTER to
continue:
Test
```

The app will ask us for the name of the folder containing the projects, which is the one we created before inside ***SAPIENT/PDF/***. After that, the app will extract the texts from the PDFs and will ask you if you want to save the extracted texts into a .json file. If you choose to save it, the resulting file will be saved into ***SAPIENT/Saves/*** and you will be able to use it in future analysis just by choosing the option **2) Read saved data** and writing the project name.

```
Write the name of the folder that contains the PDF files and press ENTER to
continue:
Test

── Extracting texts from the PDF files ──────────────────────────────────────
i Preparing to start extracting the texts
Extracting texts ■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■  100% | ETA:  0s
✓ Done! The process took 0.62 minutes

── Tidying the extracted texts ──────────────────────────────────────────────
i Preparing to start tidying the texts
Tidying texts ■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■  100% | ETA:  0s
✓ Done! The process took 0.43 minutes

✓ Press Y or y and ENTER to save the tidy texts. Press any other key to discard them.
y

✓ Successfully saved the tidy results to the path "Saves/Test.json'"
```

Saving the tabular results

After that, the mapping will start, and soon you will be asked to save or discard the tabular results of the analysis. If you choose to save it, you will find them saved into ***SAPIENT/Output/[Project name]/[Analysis mode]/data***. For our example, we will find the tabular data at:

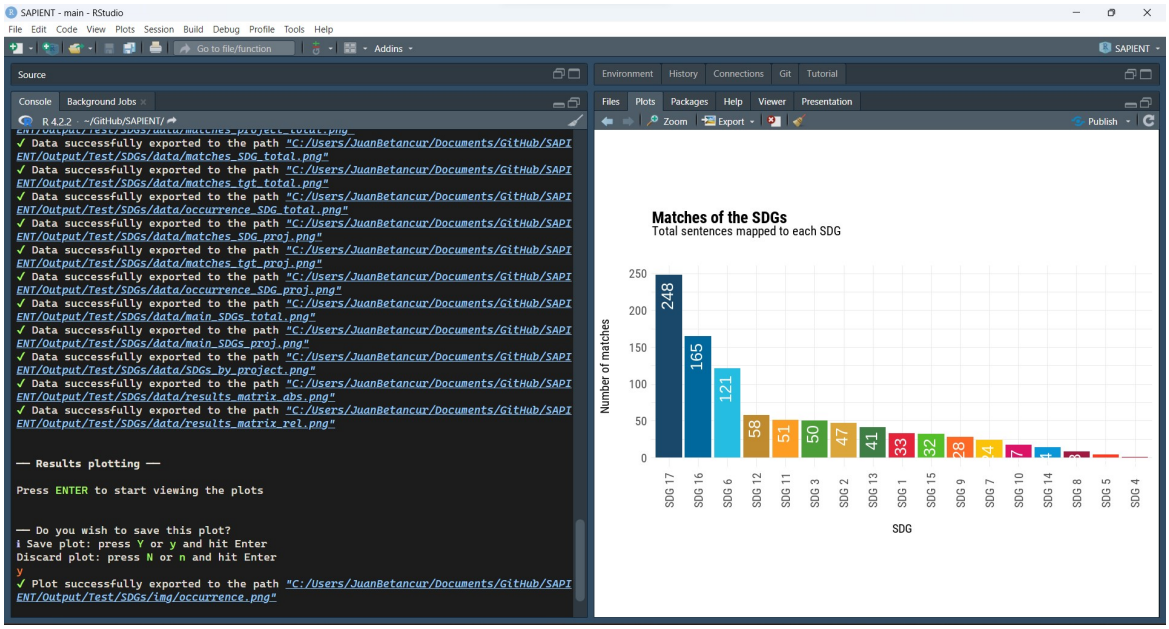
SAPIENT/Output/Test/SDGs/data.

```
── Mapping texts to the Sustainable Development Goals ──────────────────────

── Preparing the test data ──

── Creating corpus
Please wait, this could take several minutes.

Transforming data.
Removing stop words.
Transforming to lowercase.
Removing numbers.
Removing punctuation marks.
Removing stop words.
Stemming document.
Removing extra white spaces.
Creating the Document Term Matrix (DTM).

── Matching the test data with training data ──

── Creating corpus
Please wait, this could take several minutes.

Transforming data.
Removing stop words.
Transforming to lowercase.
Removing numbers.
Removing punctuation marks.
Removing stop words.
Stemming document.
Removing extra white spaces.
Creating the Document Term Matrix (DTM).
✓ Text classification done.


── Do you wish to export the results data? ──

i Save data: press Y or y and hit Enter
Discard data: press N or n and hit Enter
```

```
── Do you wish to export the results data? ──

ℹ Save data: press Y or y and hit Enter
Discard data: press N or n and hit Enter
y
ℹ Results data will be exported
✓ Data successfully exported to the path "C:/Users/JuanBetancur/Documents/GitHub/SAPIENT/Output/Test/SDGs/data/raw_results.png"
✓ Data successfully exported to the path "C:/Users/JuanBetancur/Documents/GitHub/SAPIENT/Output/Test/SDGs/data/matches_project_total.png"
✓ Data successfully exported to the path "C:/Users/JuanBetancur/Documents/GitHub/SAPIENT/Output/Test/SDGs/data/matches_SDG_total.png"
✓ Data successfully exported to the path "C:/Users/JuanBetancur/Documents/GitHub/SAPIENT/Output/Test/SDGs/data/matches_tgt_total.png"
✓ Data successfully exported to the path "C:/Users/JuanBetancur/Documents/GitHub/SAPIENT/Output/Test/SDGs/data/occurrence_SDG_total.png"
✓ Data successfully exported to the path "C:/Users/JuanBetancur/Documents/GitHub/SAPIENT/Output/Test/SDGs/data/matches_SDG_proj.png"
✓ Data successfully exported to the path "C:/Users/JuanBetancur/Documents/GitHub/SAPIENT/Output/Test/SDGs/data/matches_tgt_proj.png"
✓ Data successfully exported to the path "C:/Users/JuanBetancur/Documents/GitHub/SAPIENT/Output/Test/SDGs/data/occurrence_SDG_proj.png"
✓ Data successfully exported to the path "C:/Users/JuanBetancur/Documents/GitHub/SAPIENT/Output/Test/SDGs/data/main_SDGs_total.png"
✓ Data successfully exported to the path "C:/Users/JuanBetancur/Documents/GitHub/SAPIENT/Output/Test/SDGs/data/main_SDGs_proj.png"
✓ Data successfully exported to the path "C:/Users/JuanBetancur/Documents/GitHub/SAPIENT/Output/Test/SDGs/data/SDGs_by_project.png"
✓ Data successfully exported to the path "C:/Users/JuanBetancur/Documents/GitHub/SAPIENT/Output/Test/SDGs/data/results_matrix_abs.png"
✓ Data successfully exported to the path "C:/Users/JuanBetancur/Documents/GitHub/SAPIENT/Output/Test/SDGs/data/results_matrix_rel.png"
```

Viewing and saving the graphical outputs

After saving or discarding the tabular data, SAPIENT will prompt you to view the results plots. After pressing ENTER, the first plot will be displayed into the plots tab at the right side of RStudio. Also, SAPIENT will ask you to save or discard the plot which, if saved, can be found in ***SAPIENT/Output/[Project name]/[Analysis mode]/img***. For our example, the path would be: ***SAPIENT/Output/Test/SDGs/img***.



Contact

For further information, please contact:

- Ivonne Lobos Alva (ivonne.lobos@sei.org),
- Mario Cárdenas (mario.cardenas@sei.org),
- or Juan Camilo Betancur (juan.betancur@sei.org).