For the phase 2 of the project you have to:

* create a git repository (public or private, up to you), with two contributors (one owner and one contributor)
* create a R project from this repository and add to it a data directory with your raw data files (CSV format)
* create a quarto document that load and describe the files. The description must include:
  + a link to the github project;
  + links to the sources (with instructions to reproduce the data selection if direct linking is not possible);
  + a short description of the sources (roughly 500 words);
  + one table per data file that gives basic information about the data file, including its number of rows and number of columns (computed with the R code, not typed down);
  + a short description of the most important variables (one sentence or two per variable should be enough);
  + a description of the data joining, cleaning, etc. you have implemented;
  + a description of you research question (roughly 500 words);
  + if you have a main target variable (e.g. unemployment rate, CO2 emission, etc.) a graphical representation of this target variable.
* make regular commits and push to your repository. Your must demonstrate by the commits and push that the work is shared between the students (even if you do that side by side). This means for instance that A creates the repository and makes the initial commits, then B create the R project and pushes the modification, then A adds the data, etc. You can do that with pull request or with a shared repository (I recommend the second solution if you are new to all of this).
* upload a zip of your work on moodle before midnight on December the 8th. The zip must include at least the html rendering. I strongly recommend not including the html in the git repository;
* tag on github the commit that corresponds to the zip uploaded on moodle;
* invite me as a collaborator if your project is private (I am [fabrice-rossi](https://github.com/fabrice-rossi)).

Notice that by submitting your project, you acknowledge it is the product of you personal work and that AI was used according to the rules presented at the beginning of the course.

This project aims to estimate a gravity model to analyze the impact of internet access on intra-EU trade in digitally deliverable services from 2013 to 2023. The main objective is to explore how internet connectivity influences digital services trade flows within the EU, using a novel combination of trade and internet coverage datasets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Source** | **Data set name** | **Description** | **Number of Obs** | **Number of variables** |
| OECD | [International trade in services dataset](https://data-explorer.oecd.org/vis?fs%5B0%5D=Topic,0%7CTrade%23TRD%23&fs%5B1%5D=Topic,2%7CTrade%23TRD%23%7CTrade%20in%20goods%20and%20services%23TRD_GDS%23%7CTrade%20in%20services%23TRD_GDS_STR%23&pg=0&fc=Topic&snb=4&vw=ov&df%5Bds%5D=dsDisseminateFinalDMZ&df%5Bid%5D=DSD_BOP@DF_TIS&df%5Bag%5D=OECD.SDD.TPS&df%5Bvs%5D=1.0&dq=AUS..S.B..A.USD_EXC.&pd=2009,&to%5BTIME_PERIOD%5D=false&ly%5Bcl%5D=TIME_PERIOD&ly%5Brw%5D=COUNTERPART_AREA) | Annual trade in services statistics in OECD countries. | `N ROW()` | `N COLUMN()` |
| EUROSTAT | [Broadband internet coverage by speed](https://ec.europa.eu/eurostat/databrowser/view/isoc_cbs/default/table?lang=en) | Annual percentage of households with 100 mbs or higher internet connection. |  |  |
| Centre d'études prospectives et d'informations internationales (CEPII) | [The CEPII Gravity database](https://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=8) | For any pair of countries, from 1948 to 2020, Gravity provides all the information required to estimate gravity equations: trade flows, geographical distances, trade facilitation measures, macroeconomic indicators, etc... |  |  |
| IMF | [World Economic Outlook 2024](https://www.imf.org/en/Publications/WEO/weo-database/2024/October/download-entire-database) | The World Economic Outlook (WEO) is a survey of prospects and policies by the IMF staff, usually published twice a year, with updates in between. It presents analyses and projections of the world economy in the near and medium term, which are integral elements of the IMF’s surveillance of economic developments and policies in its member countries and of the global economic system. |  |  |

* digital\_trade\_𝑖𝑗𝑡 = bilateral digitally deliverable trade between country i and j in year t
* distance\_𝑖𝑗 = distance between countries i and j in year t
* gdp\_pc\_𝑖𝑡 = GDP per capita country i in year t
* gdp\_pc \_𝑗𝑡 = GDP per capita country j in year t
* factor(year\_𝑡) = Fixed effects of year t
* same\_continent \_𝑖𝑗 = 1 if countries i and j are in the same continent
* contiguity\_𝑖𝑗 = 1 if countries i and j are neighboring countries
* common\_language\_𝑖𝑗 = 1 if countries i and j share common language
* digital\_serv\_world\_𝑖𝑡 = total digital services trade with partner world of country i in year t
* digital\_serv\_world\_ 𝑗𝑡 = total digital services trade with partner world of country j in year t
* factor(partner) = fixed effects of country j
* internet\_coverage\_ 𝑖𝑡 = internet coverage in country i in year t
* internet\_coverage\_ 𝑗𝑡 = internet coverage in country j in year t
* 𝜀𝑖𝑗 = error term