Data Visualization
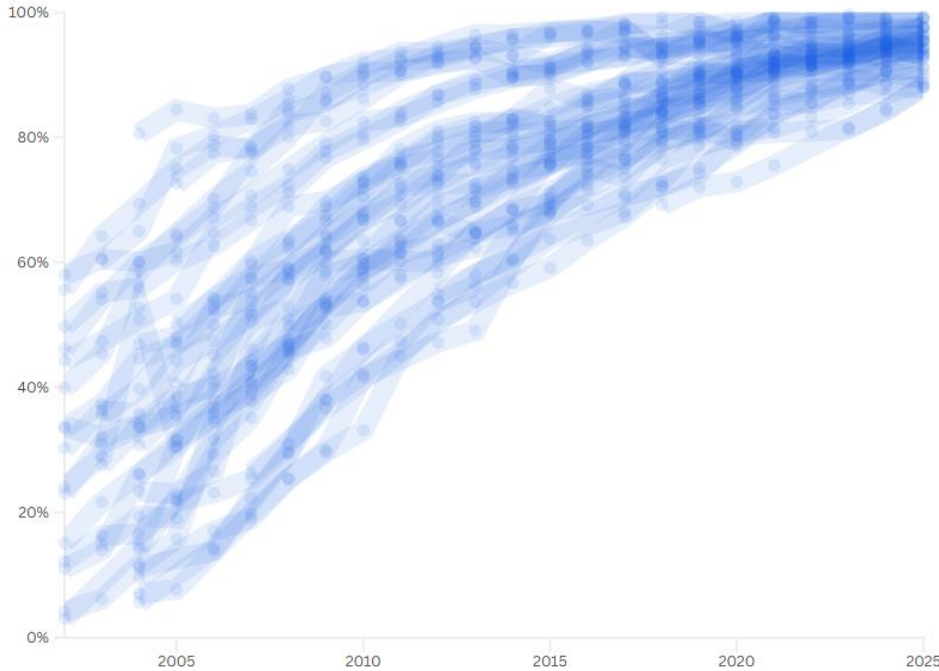
Student Individual Assignment

Source: Eurostat (isoc_ci_in_h)

Digital inclusion in Europe: household internet access over time

Student name: Indra-Theodora Catana
Student e-mail address: indratheodora@gmail.com

Research questions

One of the six EU priorities for 2019-2024 was "A Europe fit for the digital age." Achieving this objective requires, at a fundamental level, that EU citizens have access to the internet in their homes. Therefore, my primary research question is:

What is the level of household internet access across EU and EU partner countries?

- How has the situation evolved over time?
- Are there significant differences between countries?
- Does household composition play a role in the level of internet access?

About Data

Dataset: Households - level of internet access (https://doi.org/10.2908/ISOC_CI_IN_H)
Source: Eurostat
The dataset is shared under the **CC BY 4.0 license**, in accordance with the European Commission's reuse of documents policy (Legal notice – disclaimer and copyright | European Union). That means it allows free use, reproduction, modification, and redistribution of the data, including for academic and commercial purposes, provided that appropriate acknowledgement is given to Eurostat as the source.

Description of the dataset:
It containts annual data from 38 EU and EU partner countries. Each country reports:
• the percentage of households with internet access at home (all forms of internet are included).

In my initial try to understand and analyze the data, I come up with the following question: Are these values exact?
Eurostat is definitely a trustworthy source, but still, it might be important to find whether the values in this dataset represent the exact proportion of households with internet in each country (are they able to find these exact proportions/ the exact numbers?) or they are some estimates of the real proportions. In order to find this, we need to look further into the metadata.

About Data

Looking into the metadata, we find that the values are based on an **annual** *EU survey* *on the use of Information and Communication Technologies (ICT) in households and by individuals*, conducted since 2002. The data is collected by the National Statistical Institutes from each country and transmitted to Eurostat.

So, the values are based on surveys and therefore represent **estimates** of the real proportions.

How good are these estimates?

Eurostat does not publish the exact precision for each estimate, but requires that all countries meet minimum precision requirements defined in the methodological manual. In the metadata, it is stated that "results provide in general good accuracy and reliability". Despite this, in the dataset we can see some values are flagged as 'u', which states for unreliable.

Knowing that we work with estimates instead of exact values might be an important thing to take into consideration when analyzing data and drawing conclusions. Except for this matter, I believe that the data cleaning process will not be too difficult, as the data has already been validated by Eurostat and therefore respects some high quality standards.
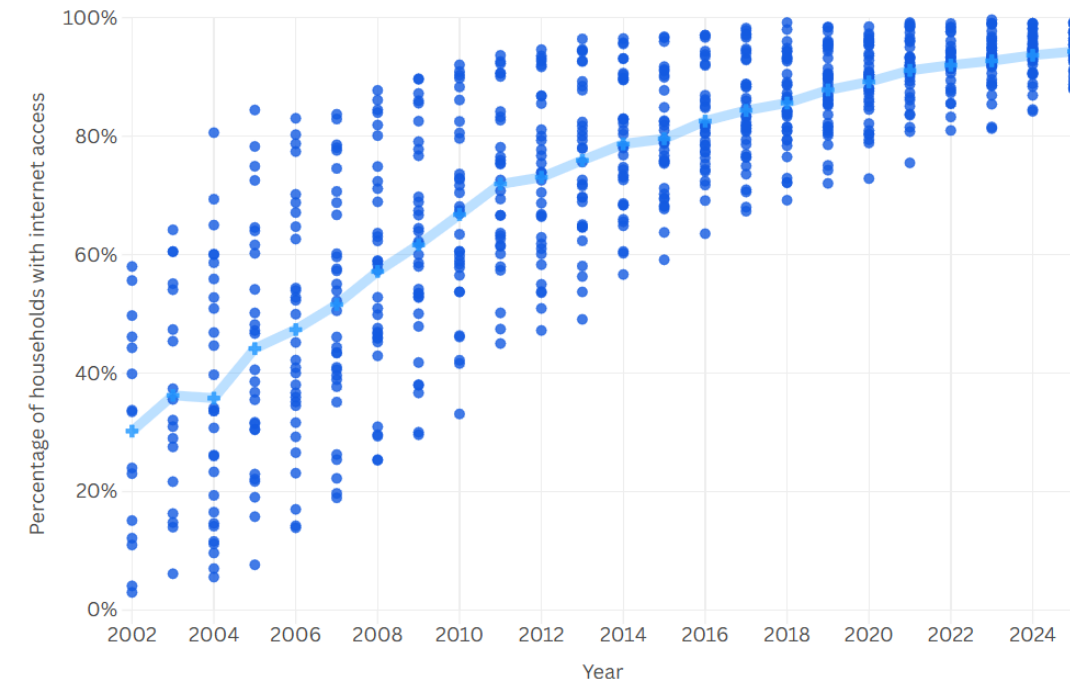
Methodology

- **Data Collection:** Eurostat lets you customize your [dataset](#) by including optional variables. So, before downloading the dataset as a csv file, I enriched it with all the years 2002-2025 and I also added the variables *All types without dependent children* and *All types with dependent children*.

- **Data Processing & Cleaning:** I have used Python and the pandas library to process, clean and transform the data. I have also used matplotlib library for creating graphs which helped me analyze the data.

- **Data Transformation:** I removed redundant columns. I aggregated the data by year and household type to compute yearly averages. I created two cleaned and simplified datasets for the visualizations.

- **Use of AI Tools:** I used ChatGPT to generate Python code for some parts of the data processing and also for creating some initial graphs with matplotlib.

- **Analytical Techniques:** Descriptive statistics (mean, distribution, min/max), Trend analysis (evolution over time), Comparative analysis (household types)

- **Reproducibility:** The data is publicly available on Eurostat's website. I also uploaded [here](#) the two datasets prepared for the data visualization tools and the Python code necessary to produce them.

Insights from the Data

- Nowadays, in 2025, all countries present in our dataset have a high level of household internet access. The minimum value is in Bosnia and Herzegovina – 87.91%, while the maximum value is in Netherlands, where 99.28% of the households have internet access at home.

- However, the values were not always so close. During the years, different countries evolved differently in terms of internet access. The highest difference observable in our dataset happened in 2005: Türkiye – 7.66%   Iceland – 84.41%.
Netherlands, Denmark, Sweden, Germany also had levels of over 60% in 2005.

- The overall minimum value present in our dataset is 0%. It states that, in Luxembourg, in 2004, 0% of the households composed of a single person with dependent children have internet access at home. It does not sound impossible. However, in 2003 the same value was 63.67% and in 2005, 56.47%. So, because Luxembourg is a small country and this household type is relatively rare compared to others, we do not know whether the value was actually true and it just fluctuates a lot, or it was just a bad estiamtion, or a mistake in the database. But, it reminds us that it is important to keep a certain degree of caution when drawing conclusions.

Distribution of household internet access across EU countries by year

Each circle represents a country-year observation; the "+" symbols represent the average of the values in that year
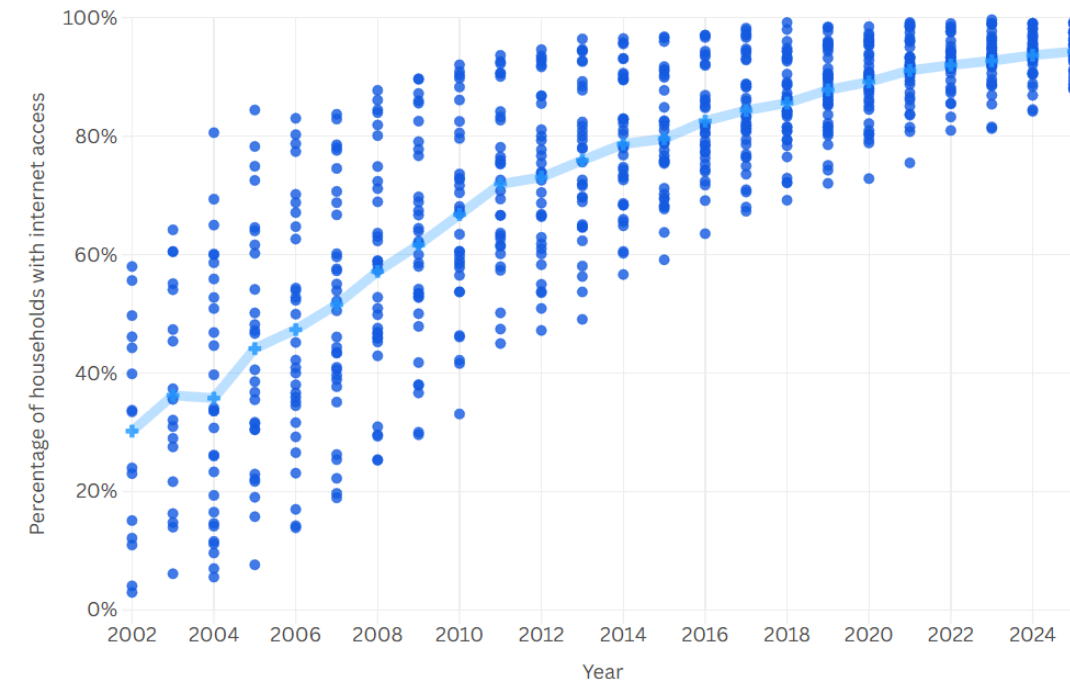
Ideally, in a well-connected digital world, all countries should have a **high** and **similar** level of internet access.

This is the pattern towards which European countries appear to be moving. The graph on the left shows how the distribution of household internet access levels for the countries in our dataset has changed over time. We can see how it started in 2002: there was a lot of variation – some countries already had almost 60% of their households having access to the internet, while other countries just started this process. The problem of big differences between countries in terms of household internet access has continued until the last few years. Starting from 2022, all the countries are over 80% and the distribution is getting tighter and tighter.

Nowadays, the average level of household internet access for the countries in our dataset reaches 94.29%. We can see how this average had increased over the years (blue line), but another remarkable thing is how countries are starting to achieve **similar high** levels of connectivity.

Distribution of household internet access across EU countries by year

Each circle represents a country-year observation; the "+" symbols represent the average of the values in that year

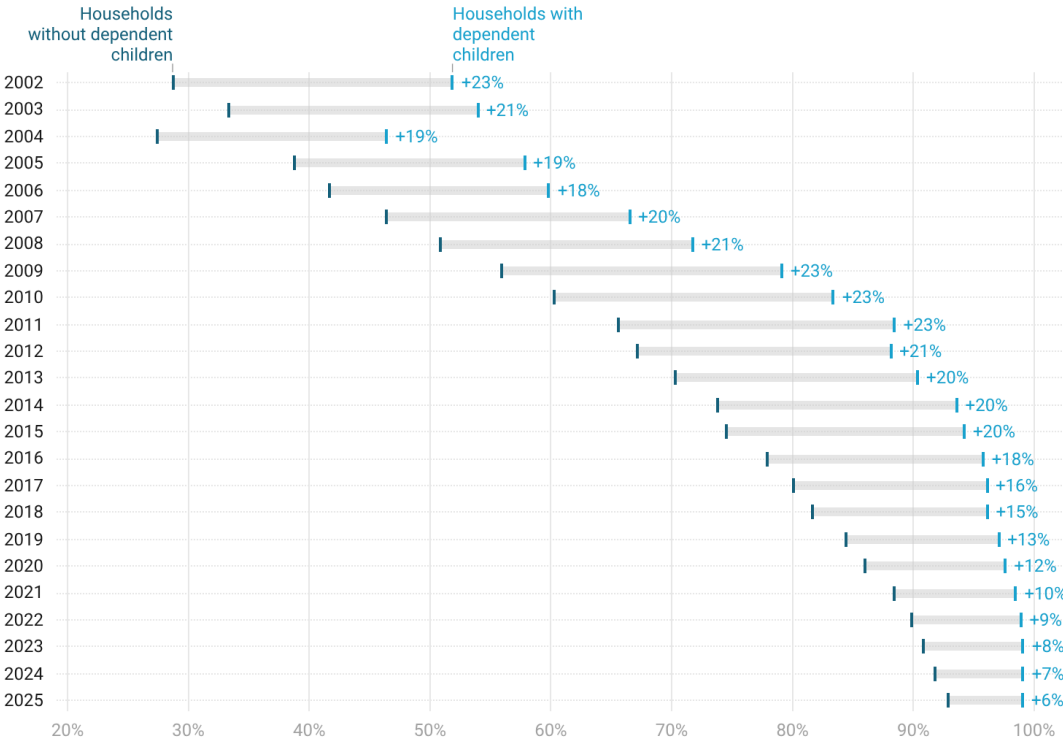Source: Eurostat (isoc_ci_in_h)
Created with Flourish

*It is important to note that the dataset has missing values. We do not have the values for some countries in some years. This is also the reason why in our graph, in 2004, there seems to be a drop in the average. I calculated the average for each year, using the available countries in that year. It just happened that in 2004 many countries with low values have been added for the first time in the dataset.

There are 38 distinct countries in the entire dataset; and every year we have on average data from 32 countries. This reassures us that, despite some missing values, we still have sufficient data every year to have a good representation of how the distribution evolved.

Gap in internet access between households with and without dependent children

Each horizontal segment connects the average internet access levels of households without dependent children (left) and households with dependent children (right) for a given year. The values at the right of the segment represent the difference between the two household types.

Households without dependent children | Households with dependent children

Year	Gap
2002	+23%
2003	+21%
2004	+19%
2005	+19%
2006	+18%
2007	+20%
2008	+21%
2009	+23%
2010	+23%
2011	+23%
2012	+21%
2013	+20%
2014	+20%
2015	+20%
2016	+18%
2017	+16%
2018	+15%
2019	+13%
2020	+12%
2021	+10%
2022	+9%
2023	+8%
2024	+7%
2025	+6%

20%   30%   40%   50%   60%   70%   80%   90%   100%

Source: Eurostat • Created with Datawrapper

We have seen previously that having internet access at home is more and more common across european countries. However, there is still room for improvement and in order to understand how we can improve it, I analyzed the data by household type.
We all know that children have a bigger curiosity towards discovering technology and using it. This may have driven households with dependent children to adopt internet faster than others.
Over the years, there has always been a gap in terms of internet connectivity between households with and without dependent children, but, fortunatelly, as the internet becomes more and more popular, this difference gradually narrows.
In 2025, the average level of internet connectivity for households with dependent children was **98.99%** (for the countries in our dataset).
If we want to improve the digital connectivity in Europe, we should focus on households without dependent children, which are often associated with older adults, that may have difficulties in adopting technology.

LICENSE

Slides are shared with the following license: