



# EMOS Coding lab 2024

## *Statistics Explained through literate programming 2024*

### **1. What you will do – Description of the project and objectives**

Dissemination is an integral part of the whole statistical production workflow. Inspired by the reproducibility movement in Open Science and driven by the opening and sharing of all assets – making not only the data, but also the methods, tools, and software open – the EMOS Coding lab aims to promote an innovative bottom-up approach to dissemination that enables collaboration and increases participative and reproducible forms of statistical services’ design and sharing. Reproducibility along the different stages of the dissemination process supports the objectives of the [European data strategy](#), that expresses a commitment to transparency, collaboration, and accessibility in the digital era, as well as several principles expressed in the [European Statistics Code of Practice](#), most notably principle no. 15 on accessibility and clarity.

This project aims at replicating [Eurostat Statistics Explained](#) articles, statistical products composed of tables, charts and text. The ultimate goal is streamlining the production process for retrieving the data (through calling the [Eurostat Application Programming Interface - API](#)), making calculations, producing tables and visualisations, and more ambitiously integrating text in computational notebooks (e.g. with [R Markdown](#)). In practice, the Statistics Explained pages published on Eurostat websites will be reimplemented in R, following the layout guidelines, which need to be replicated in the programming language. First examples (and templates) of what the project aims for can be explored on [the statistics-coded page](#) of [Eurostat GitHub domain](#). The students will work on case studies consisting of real articles to be updated in the framework of Eurostat official statistics.

### **2. What you will learn – Outcomes and benefits**

The envisioned learning outcomes of the coding lab are the following:

- You will learn about data (and metadata) in official statistics, the way they are formatted, disseminated and shared,
- You will interact with official statistics through the Eurostat API by using dedicated client packages: learn to use the API, query, extract, load and transform data from the Eurostat database,
- You will improve your analytical skills from exploration of datasets to production of visualisations,
- You will learn best practices from Open Science in terms of replicability and reproducibility, including versioning and testing of your code,
- If successful, you will publish your results and reference your work on the statistics-coded page of Eurostat GitHub domain.

### **3. What you will need – Desired/required knowledge and skills**

- The minimum requirement is beginner to intermediate programming skills in R. **This project does welcome students with beginner skills.** Advanced R programmers are more than welcome.
- Prior knowledge of data processing and visualisation tools.



#### 4. How you will work – Organisation of the project

If selected, you will be working on a 3-4 months project under direct tutoring from Eurostat staff. The team will be guided by Matyas Meszaros, PhD in Applied Economics and Andrea Gallelli, PhD in Sociology.

The project work will be conducted remotely. You will work mostly through the collaborative development platform, e.g. GitHub. Technical questions, e.g. regarding code/programming, will be addressed through the ticket issuing facility of the platform. Interactions with the tutors, typically in the form of informal calls and MS Teams chat., will be arranged flexibly depending on the project needs. Note that the timeline of this project is flexible.

#### 5. Organisation

**Duration/workload:** 3-4 months/5 hours a week – 3 weeks equivalent full-time at least. The final time investment will depend on the number of pages the student will want to reproduce.

**Period:** March-June 2024 (flexible)

**Working method:** Remote teams interacting on GitHub and through calls and chat in MS Teams. The working language is English

**Coding expertise:** Minimum beginner to intermediate

**Contact:** EMOS Secretariat at [ESTAT-EMOS@ec.europa.eu](mailto:ESTAT-EMOS@ec.europa.eu)

#### **Deadline for application:**

Selection will be done by Eurostat project leaders based on CVs and motivation letters (up to 400 words) and possibly a short remote interview. All applicants will be informed of the result.

#### 6. Background

[Eurostat](#) is the statistical office of the European Union providing statistics at European level that enable comparisons between countries and regions. We are committed to fostering the education of future official statisticians and to developing new skills in the rapidly changing data environment.

Eurostat and its partners in the European Statistical System and the European System of Central Banks implement the European Master in Official Statistics ([EMOS](#)). EMOS is a shared initiative which aims to include official statistics in master's programme curricula, connect producers of official statistics and academia representatives across Europe and create a pool of highly skilled professional statisticians for recruitment in the ESS and the ESCB.

By organising the Coding lab, Eurostat seeks to form teams of students enrolled in EMOS-labelled master's programmes who will work remotely on the predefined tasks. The project provides a unique learning opportunity on state-of-the-art programming techniques and statistical methodological developments. It is an opportunity to collaborate with Eurostat staff and get further acquainted with the way in which official statistics are produced and disseminated. Students involved in the project will receive a certificate of participation. Depending on the local university rules, the project work may also be recognized officially as part of the study programme, it may grant ECTS credits, or it may serve as a basis for writing a master's thesis (with the support of a university advisor).