PROJECT CONCEPT
ATTITUDES RELATING CLIMATE CHANGE AND AIR QUALITY IN EUROPEAN CITIES

Dataset from EOSC-Future-SP9
Combining environmental data with individuals' attitudes on climate change.

Group:
Github : Rujina
Theory & Background : Mahmudul, Paul
ER Diagram : Abdel
Data Modeling : Paul

ATTITUDES RELATING CLIMATE CHANGE AND AIR QUALITY IN EUROPEAN CITIES

Research Questions:

- Examine, how climate-related attitudes differ between regions.
- Analyze, if exposure to air pollution is associated with thinking about climate change.
- Investigate, if air pollution is related to individuals' perceived responsibility for climate change.

Why Relational Database?

- Data consists of participant-related data and environmental data of participants' region
- Different data should be stored independently
- Increase consistency and reduce redundancy
- Convenient analysis of subsets

Motivation:

Climate change is a topic of high relevance for modern society. We are interested in the link between individuals' attitudes and their perception of environmental problems, which become more important due to climate change. It appears like individuals do not worry about their ecological impact when they don't feel the consequences. Understanding attitudes is relevant to understanding individuals' behavior.

DATASET

- Selecting relevant features from the original dataset (see on next slide)
- Dividing different types of data into different tables
- Creating primary keys when no unique identifier is present
- Through referencing the foreign keys Id, TempID, AirQID and OpinionID, the interviews table can be merged with the variable relevant for analysis
- The regional code can be used to extract participants' countries of origin
- The day, month and year of the interview were used to calculate the date of the interview
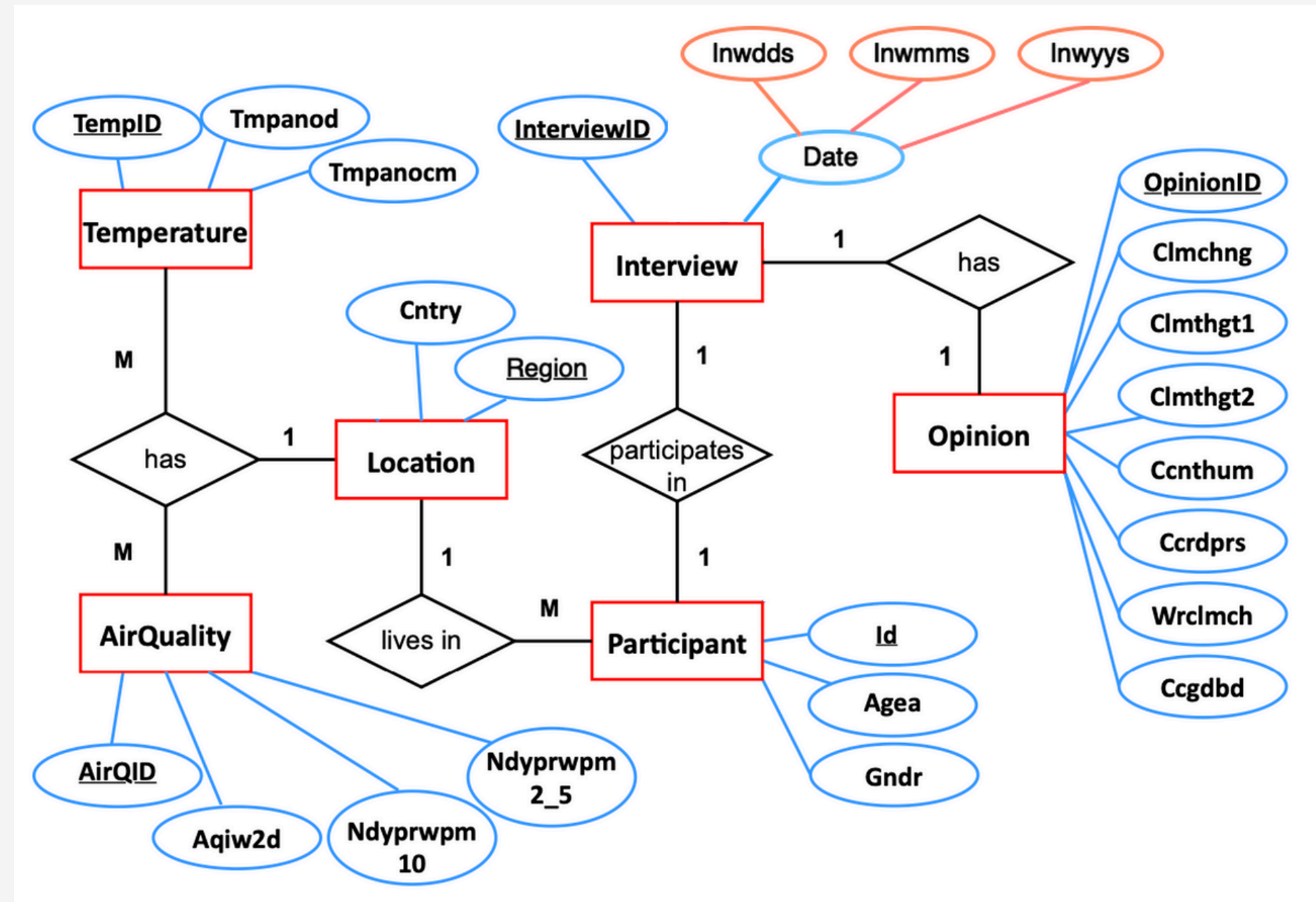
DATA HANDLING

- Pairwise deletion of rows with NAs to ensure referential integrity
- Since the original data file contained participant IDs that are not unique, we created the primary key Id

QUERIES

- Use the regional code to match participants' countries. Group participants by country to obtain how many participants are from what countries (see Figure 2).

- Match participants' opinions on climate change to the weather and air quality (preceding the interview). Examine the correlation between participants' opinions on climate change and the variables on weather / air quality (see Figure 3).

Variables Extracted from Original Dataset

Participant:

- `agea` – Age of respondent, calculated
- `gndr` – Gender

Location

- `region` – Region
- `cntry` – Country

Interview variables

- `inwdds` – Start of interview, day of month
- `inwmms` – Start of interview, month
- `inwyys` – Start of interview, year

Temperature variables:

- `tmpanod` – Temperature anomaly, date
- `tmpanocm` – Temperature anomaly, calendar month

Air quality:

- `aqiw2d` – Worst air quality index level across pollutants, last two days
- `ndyprwpm10` – Number of days with 'poor' air quality level or worse PM10, week before the date
- `ndyprwpm2_5` – Number of days with 'poor' air quality level or worse on PM2.5, week before the date

Opinion on climate change:

- `clmchng` – Do you think world's climate is changing
- `clmthgt1` – How much thought about climate change before today
- `clmthgt2` – How much thought about climate change before today
- `ccnthum` – Climate change caused by natural processes, human activity, or both
- `ccrdprs` – To what extent feel personal responsibility to reduce climate change
- `wrclmch` – How worried about climate change
- `ccgdbd` – Climate change good or bad impact across world
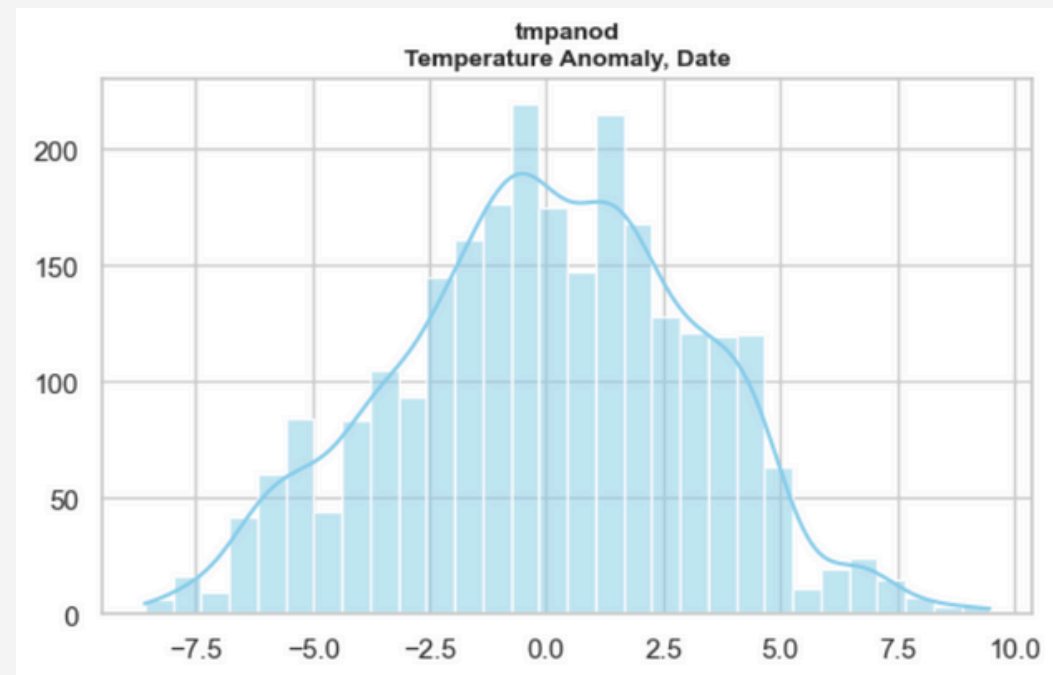
RESULTS

Figure 1: Distribution of temperature anomaly on the day of the interview (example).
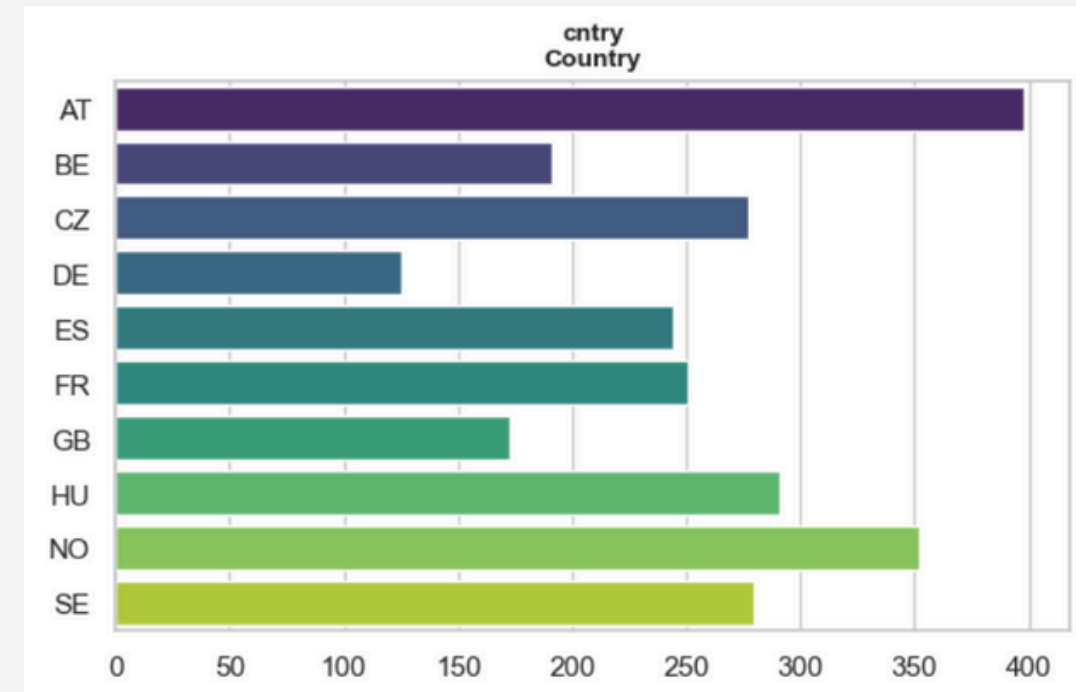

Figure 2: Participants' countries of origin

Figure 3: Correlation Matrix of variables. Variable labels are labeled by color: air quality, weather, opinions on climate change.

Conclusion:

Air quality and extreme weather preceding the interview don't impact interviewee's opinions on climate change.

Limitations:

- Other variables should be examined as well.
- Pearson's correlation only examines linear relationships.
- Measurement of variables is likely not ideal for this research question. E.g. restricted range of opinions on climate change.
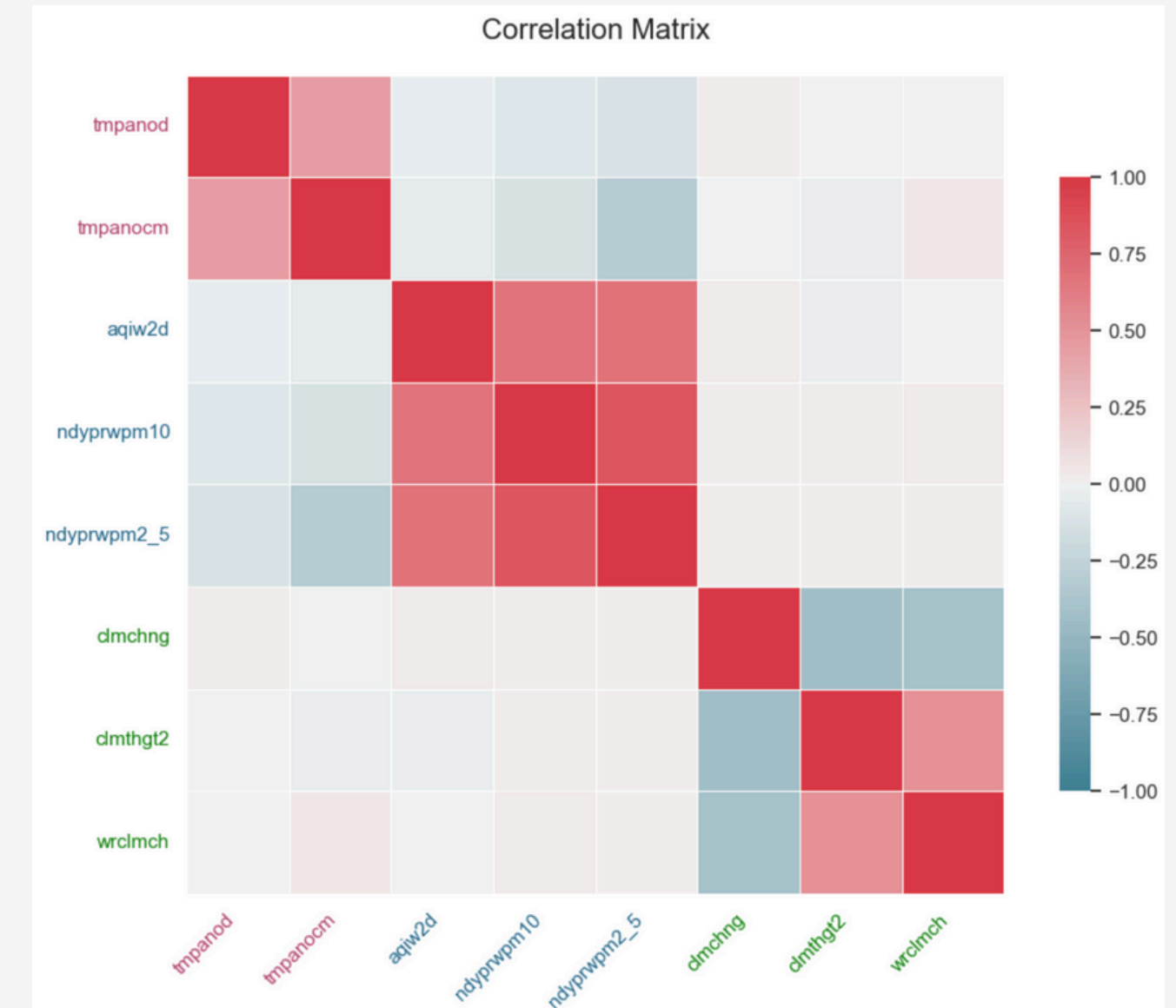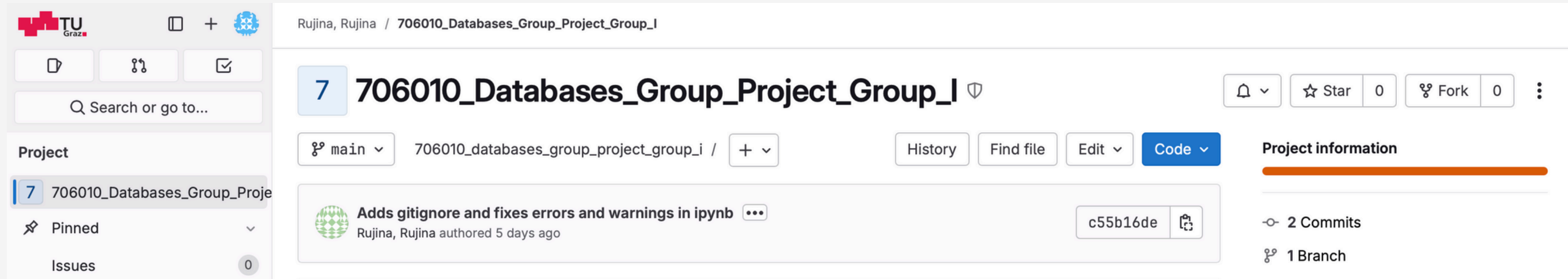
REPRODUCABILITY

Gitlab Page
- Jupyter Notebook: Data ingestion, query execution, data analysis, results and created plots
- Dataset
- Codebook of the original dataset

https://gitlab.tugraz.at/B84B4DB32391348E/706010_databases_group_project_group_i/-/blob/main/Group_I_notebook.ipynb?ref_type=heads