

Text mining in Python for Evaluating the Ethical Foundations in Computer Science



ALLEGHENY COLLEGE

Enpu You, Oliver Bonham-Carter, Janyl Jumadinova

Depts of Computer Science, Allegheny College
Meadville, PA

<https://www.cs.allegheny.edu>

{youe2, obonhamcarter, jjumadinova}@allegheny.edu

PROJECT OBJECTIVES

We present an automated text-mining tool written in Python to measure the technical responsibility of students in computer science courses.

- Our tool automatically collects reflection documents written by students from their GitHub repositories.
- Then, using natural language processing analyzes them for ethical considerations based on pre-determined questions and criteria.
- Using our tool, it is possible to see the progression of a single student's ethical thinking throughout a specific course and throughout the entire computer science curriculum, as well as, to have a grand view of all students' progress in developing an understanding of social responsibility in computer science across all levels of our courses.

TEACHING RESPONSIBLE COMPUTING

Teaching responsible computing is critical in developing software that produces a positive impact on our society, economy, and individuals.

- Each application course in computer science at Allegheny integrates ethical considerations in its pedagogy.
- Broad learning categories include topics of internet health, ethics and responsible computing that are specific to each application course.
- The delivery of these concepts include readings and class discussions, class and lab assignments with heavy software development emphasis.
- Students write reflection reports for these assignments that demonstrate their understanding of relevant issues, ability to analyze and evaluate information, and their capacity for integrating the understanding and analysis of ethical thinking into their own work.

FEATURES

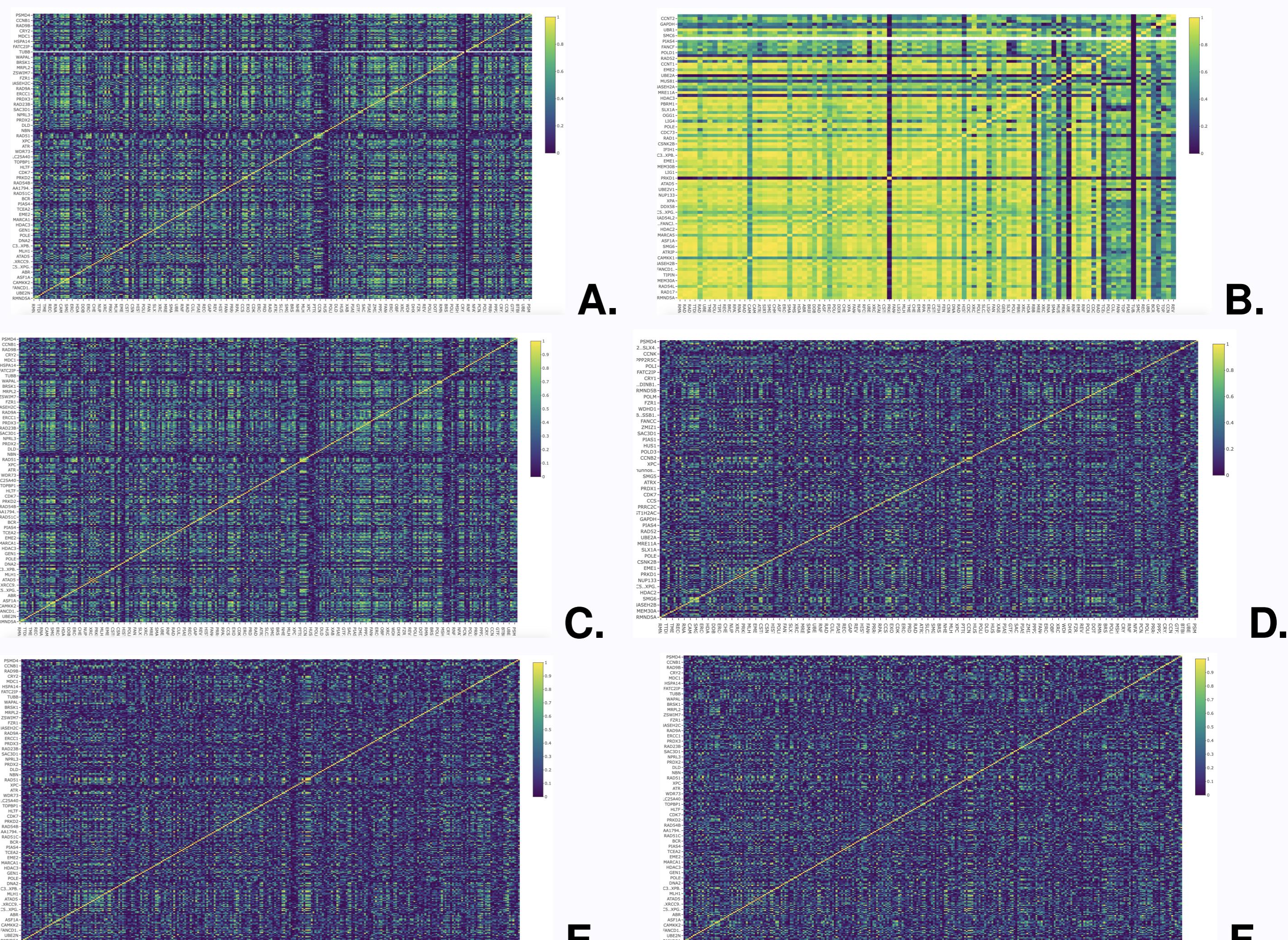


Figure: ???. Heatmaps of R^2 values, derived from normalizing factors

- bla bla bla
- bla bla bla
- bla bla bla

RESULTS

- bla bla bla
- bla bla bla
- bla bla bla

CONCLUSIONS

- bla bla bla
- bla bla bla
- bla bla bla

TEXT MINING TOOL TO DETERMINE ETHICAL PEDAGOGY

Random selection of ten data sets of breast cancer gene expression.

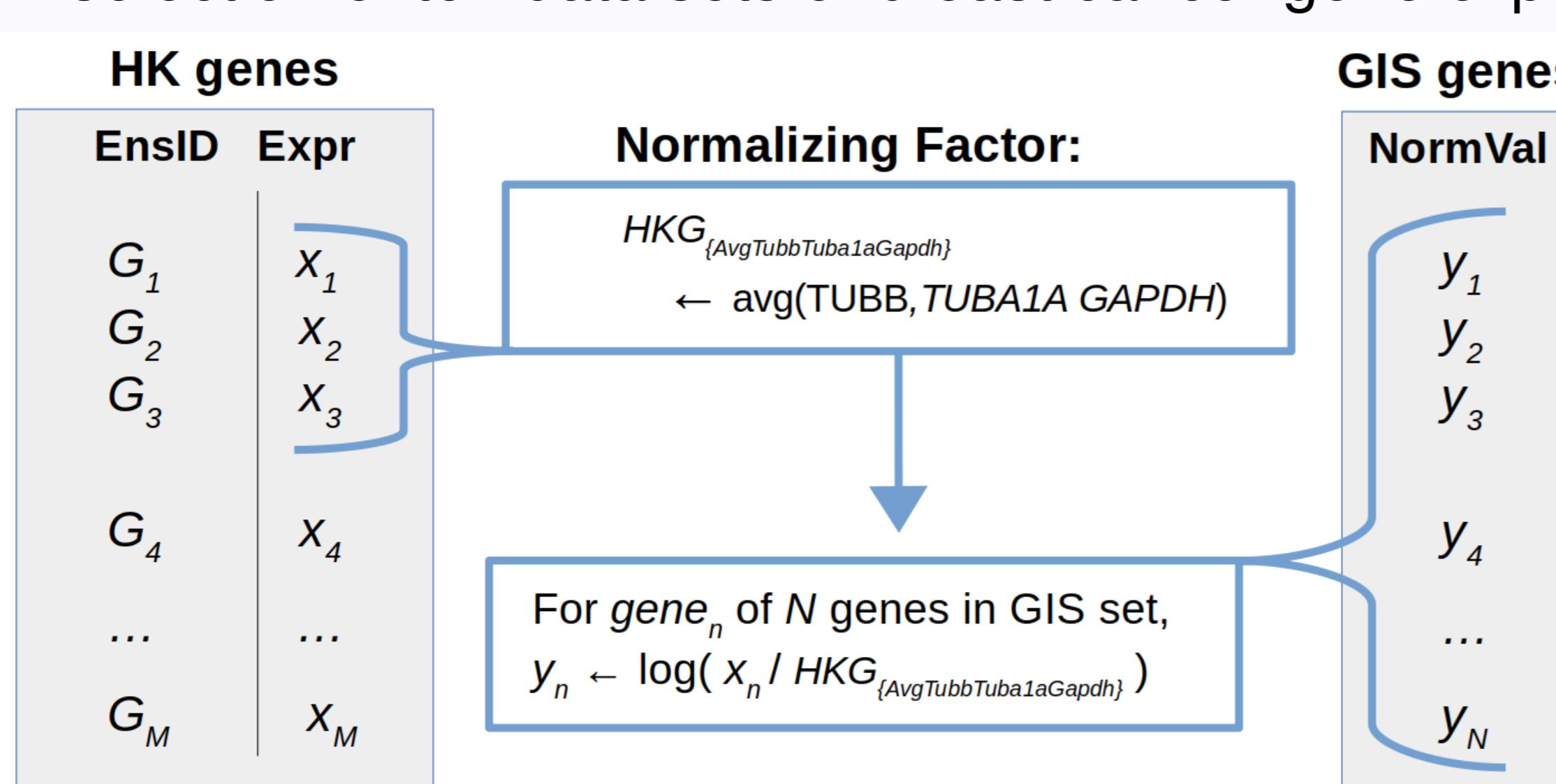


Figure: ???. Determining the normalizing factors

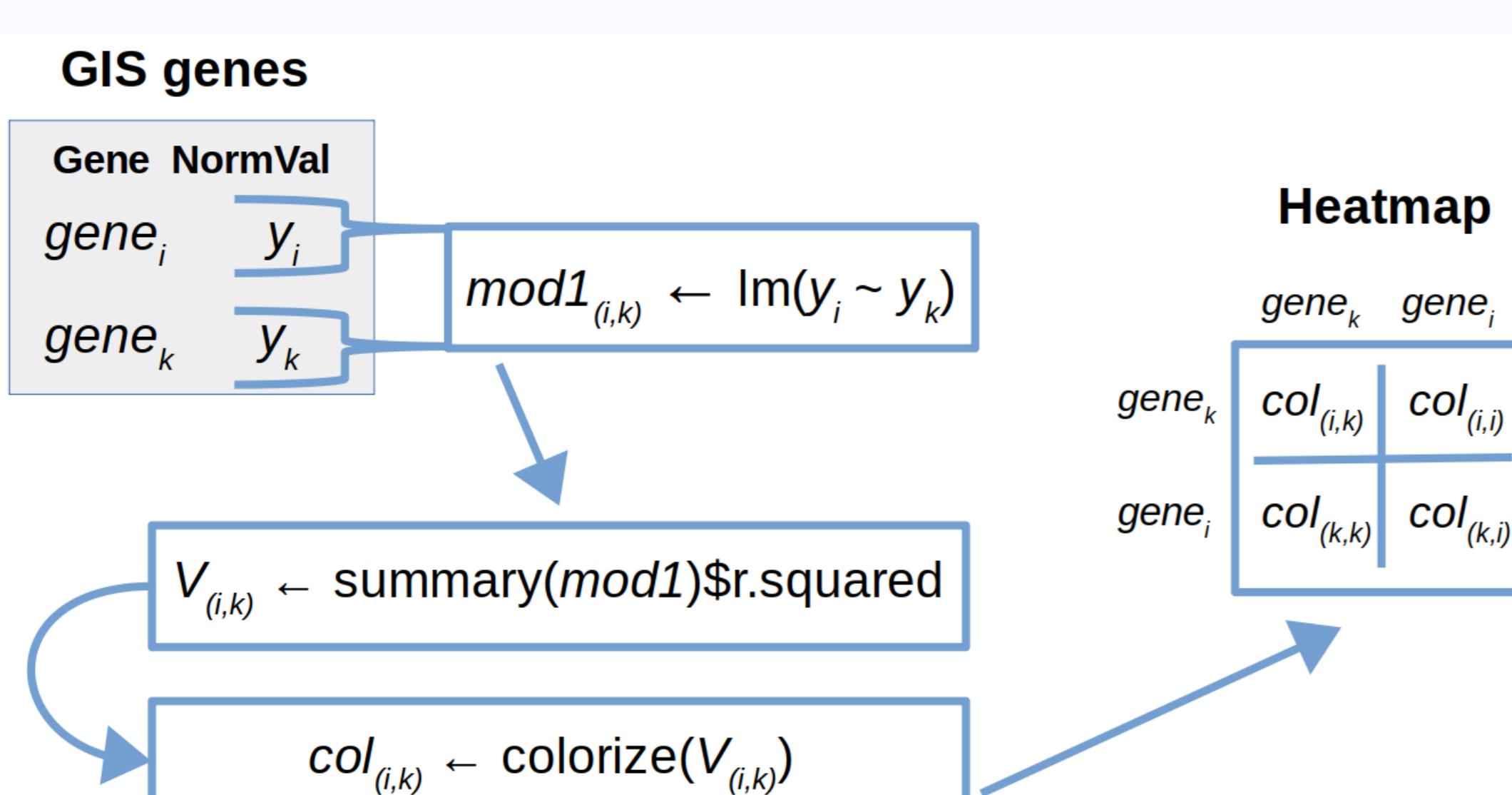


Figure: ???. HouseKeeping genes to normalize data sets

- bla bla bla
- bla bla bla
- bla bla bla