

Webscraping

Peter Ganong and Maggie Shi

November 5, 2024

Table of contents I

Exercise

1. Go to the Tanzania National Examination outcomes ([link](#))
2. Familiarize with the web page and table – **what is the unit of observation?**
3. From looking at the table, **name a detail** about the structure and layout of the table that you anticipate could make scraping difficult
4. Right click over one of the observations and “Inspect” it: **what tag(s) does the data appear to be stored in?**

Exercise: some possible answers for #3...

2. Unit of observation: student-level
3. Possible answers:
 - ▶ Parts of the header are vertical and appear to be in different rows
 - ▶ Missing values for some columns in some rows
 - ▶ “Repeaters” row structure seems to be different – ‘Absent’ value spans several columns
 - ▶ There appear to be 3 other tables: school overall grade summary (top), grade totals and an extra table (bottom)
4. td