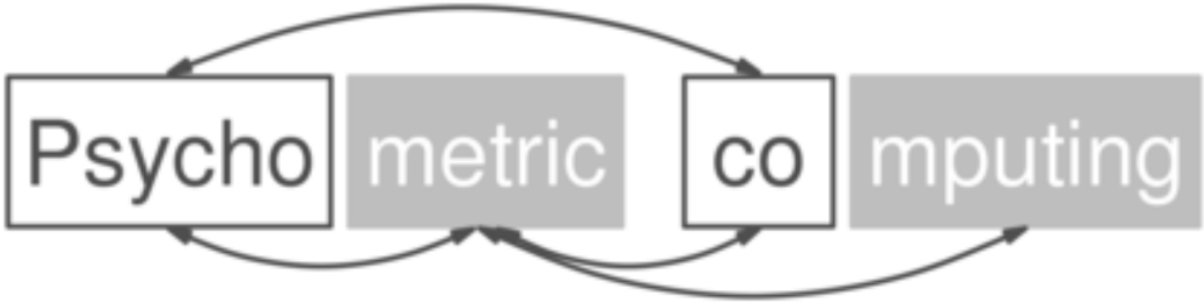
`ShinyItemAnalysis` for Psychometric Training and to Enforce Routine Analysis of Educational Tests

Patrícia Martinková

Dept. of Statistical Modelling, Institute of Computer Science, Czech Academy of Sciences
College of Education, Charles University in Prague

R meetup Warsaw, May 24, 2018

Announcement 1: Save the date for Psychoco 2019!



International Workshop on Psychometric Computing

Psychoco 2019

February 21 - 22, 2019
Charles University & Czech Academy of Sciences, Prague

www.psychoco.org

Since 2008, the international Psychoco workshops aim at bringing together researchers working on modern techniques for the analysis of data from psychology and the social sciences (especially in R).

Announcement 2: Job offers

Job offers at Institute of Computer Science:

- CAS-ICS Postdoctoral position (deadline: August 30)
- ICS Doctoral position (deadline: June 30)
- ICS Fellowship for junior researchers (deadline: June 30)
- ... further possibilities to participate on grants

E-mail at martinkova@cs.cas.cz if interested in position in the area of

- Computational psychometrics
- Interdisciplinary statistics
- Other related disciplines

Outline

1. Introduction
2. `ShinyItemAnalysis`
3. Teaching psychometrics
4. Routine analysis of tests
5. Discussion

Motivation

- To teach psychometric concepts and methods
  - Graduate courses "IRT models", "Selected topics in psychometrics"
  - Workshops for admission test developers
  - Active learning approach w/ hands-on examples
- To enforce routine analyses of educational tests
  - Admission tests to Czech Universities
  - Physiology concept inventories
  - ... tests of various purposes across the world
- Promotion of own psychometrics research
  - Detection of Differential Item Functioning (DIF)

Need for user-friendly and freely available tool

ShinyItemAnalysis Application

ShinyItemAnalysis

Interactive (and step by step) analysis of educational tests and their items

Available as:

- R package
  - Version 1.2.7 now on ▸ CRAN
  - Newest version on ▸ GitHub

```
startShinyItemAnalysis()
```

- Online `shiny` application
  - ICS server in Prague, CZ:

    https://shiny.cs.cas.cz/ShinyItemAnalysis/
  - shinyapps.io:

    https://cemp.shinyapps.io/ShinyItemAnalysis/

Authors and contributors

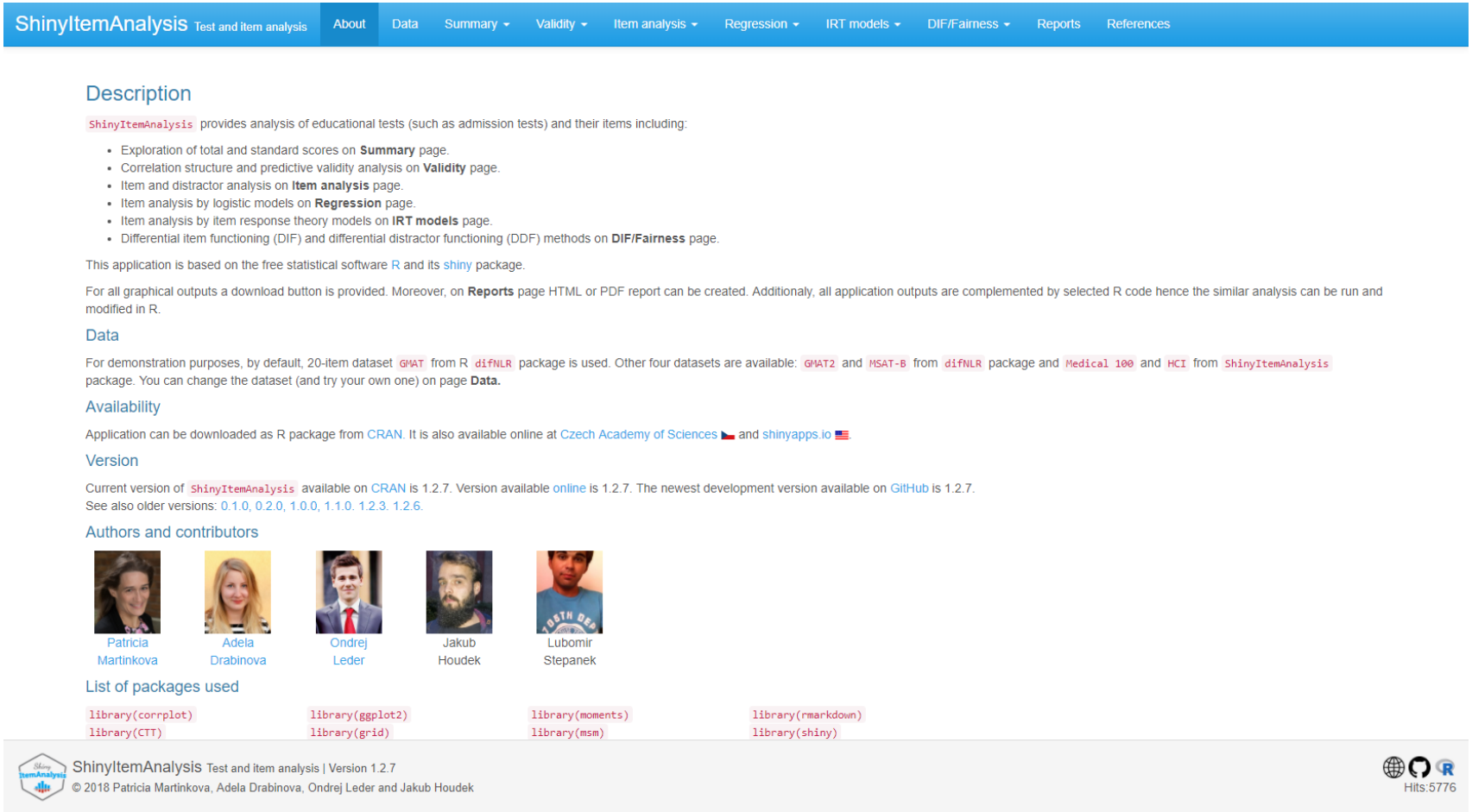




Patrícia Martinková[1,2] Adéla Drabinová[1,3] Jakub Houdek[1,4] Ondřej Leder[3] Lubomír Štěpánek[4,5]

[1] Department of Statistical Modelling, Institute of Computer Science, Czech Academy of Sciences
[2] College of Education, Charles University, Prague
[3] Department of Probability and Mathematical Statistics, Charles University, Prague
[4] Faculty of Informatics and Statistics, University of Economics, Prague
[5] First Faculty of Medicine, Charles University, Prague

ShinyItemAnalysis application



R package ShinyItemAnalysis downloads from CRAN

Package CRAN downloads over time

Enter an R package to see the # of downloads over time from the RStudio CRAN Mirror. You can enter multiple packages to compare them

Packages:

ShinyItemAnalysis

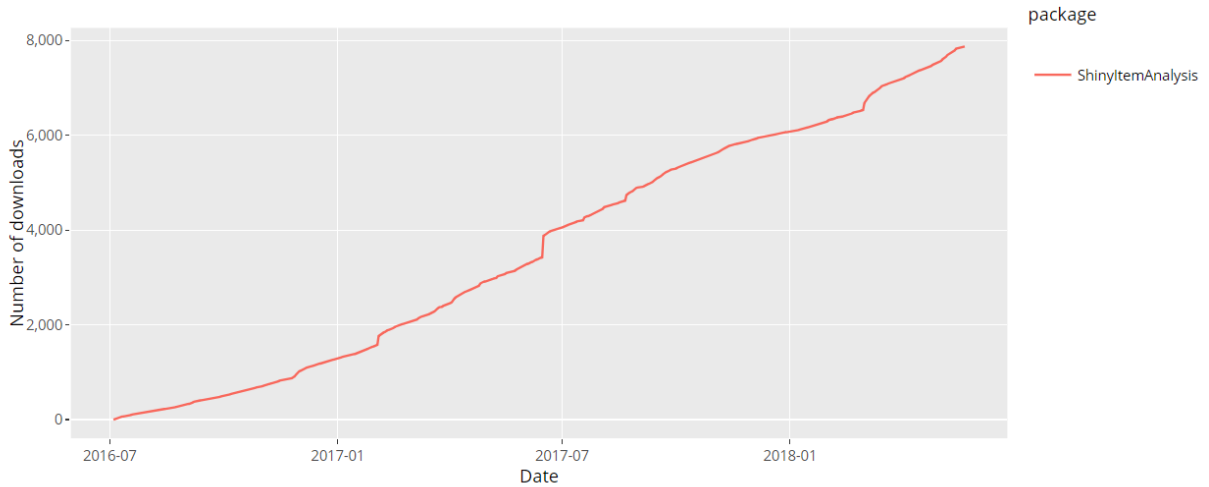
Data Transformation:

- Daily
- Weekly
- Cumulative

Date range: yyyy-mm-dd

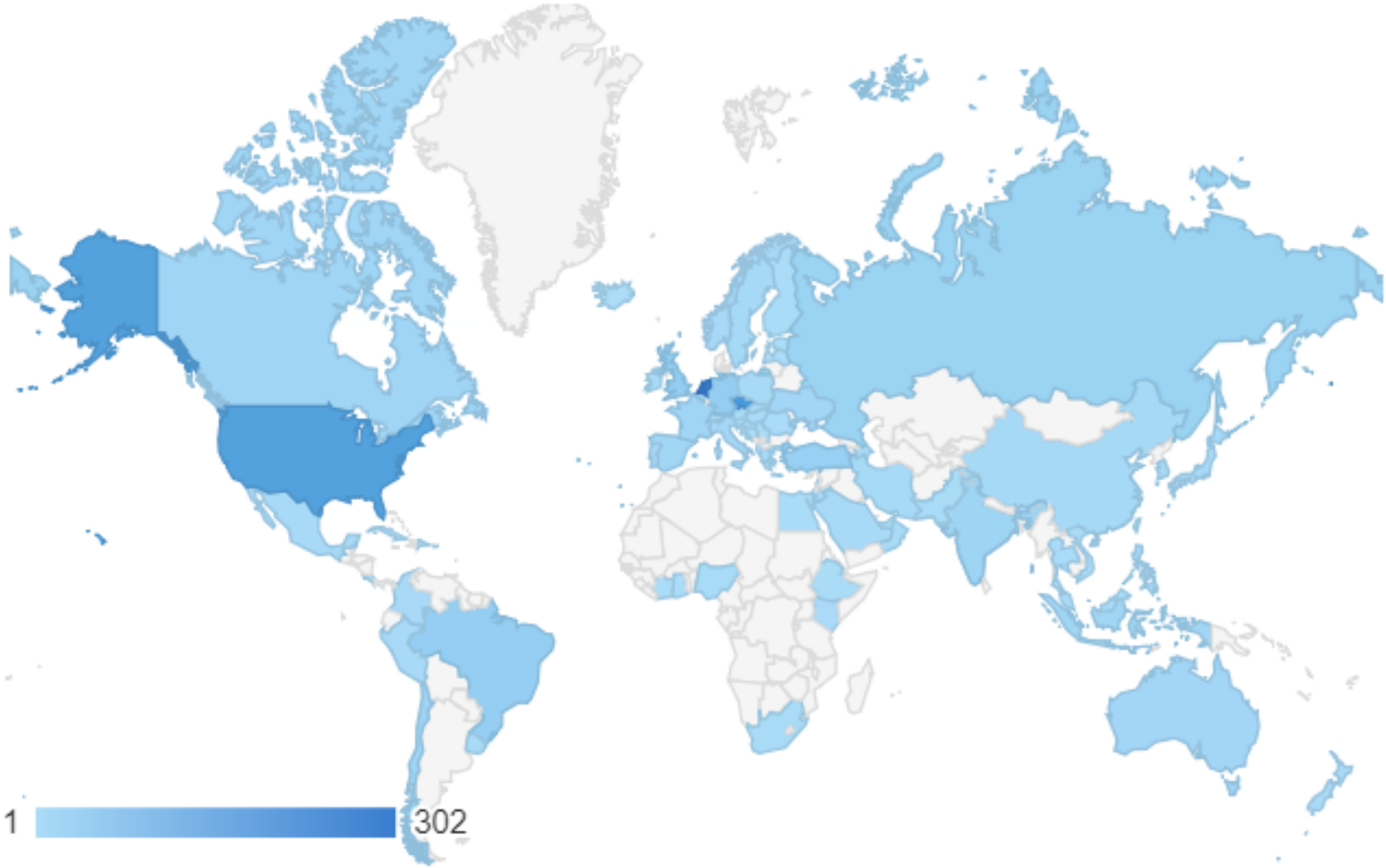
2016-07-04 to 2018-05-22

Created using the cranlogs package. This app is not affiliated with RStudio or CRAN. You can find the code for the app here, or read more about it here.



ShinyItemAnalysis
Title: Test and Item Analysis via Shiny
Latest version: 1.2.7

ShinyItemAnalysis online app is used worldwide!



ShinyItemAnalysis for teaching psychometrics

Who do we teach:

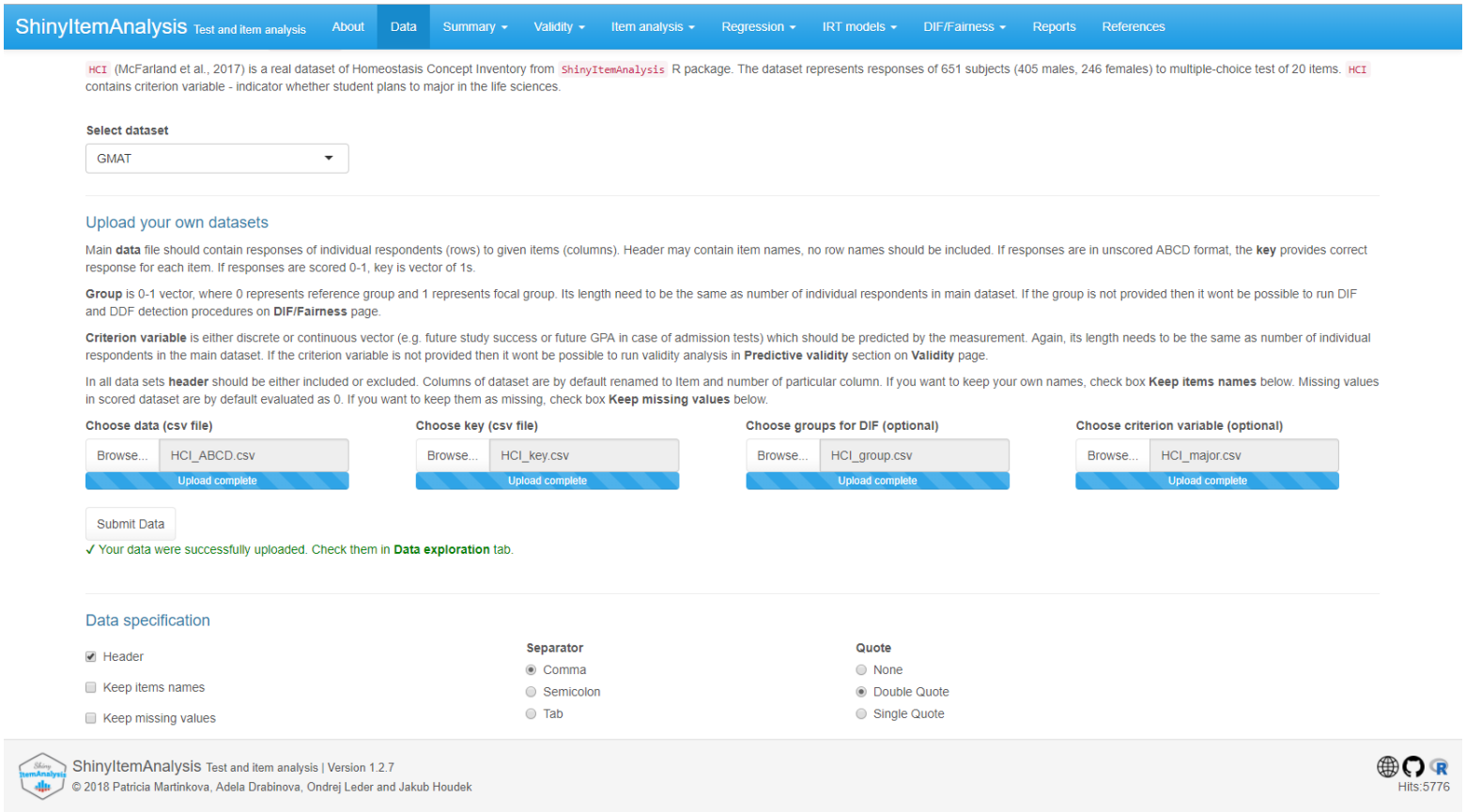
- Graduate students of different fields (Psychometrics ▸ NMST570)
- Faculties, university stakeholders

Some helpful features:

- Example datasets, allows to upload own data
- Building models in a step-by-step way
- Models, estimates, interactive interpretation of results
- Interactive training and exercises
- Provides sample R code

Datasets

- Five toy datasets are available
- Allows to upload and preview one's own dataset



Summary of Total Scores

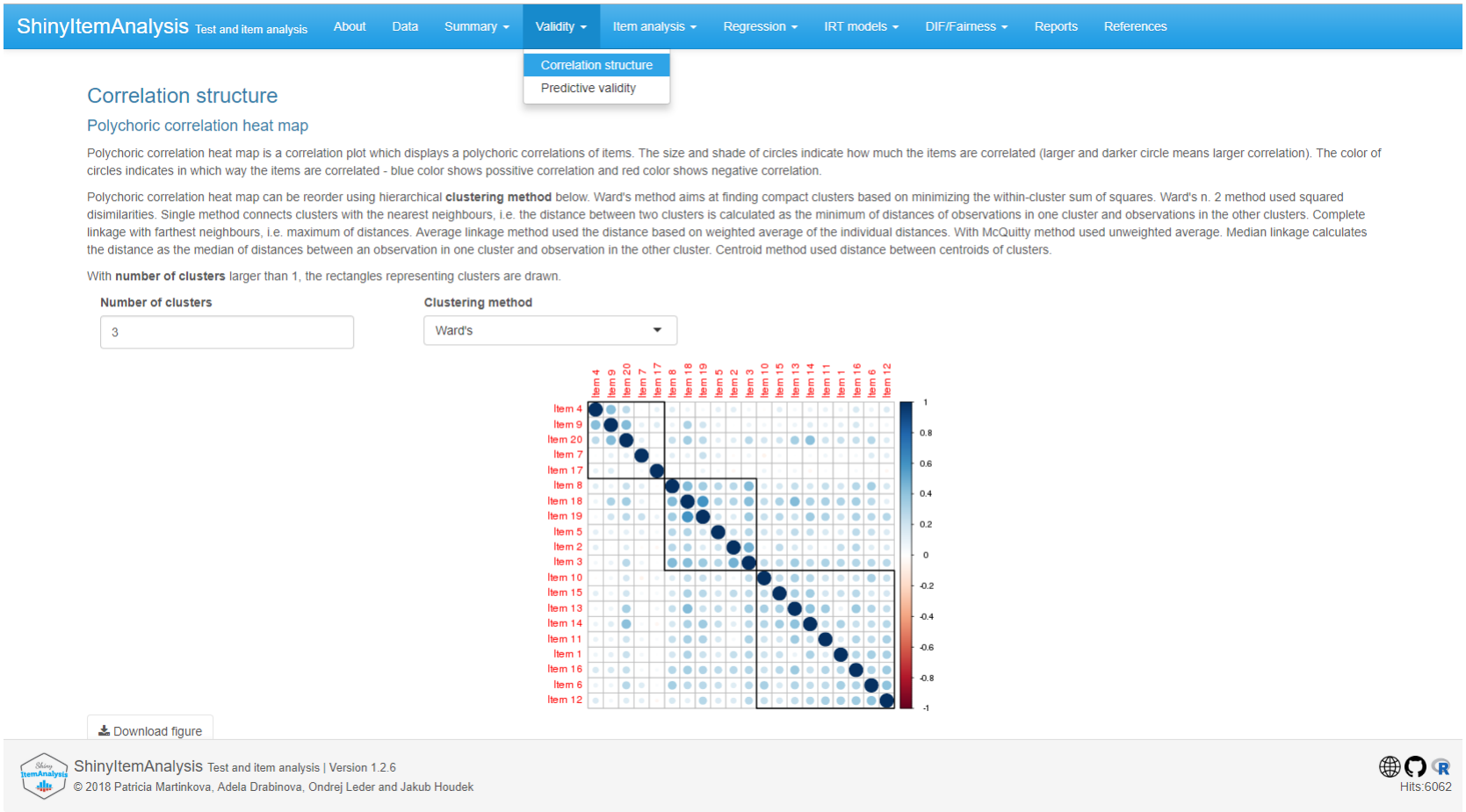
- Summary statistics
- Interactive histogram

Criterion validity

- Ony when criterion variable is available (study success, GPA, etc.)
- Available for total score as well as for items

Correlation structure

- Correlations between items
- Item clusters

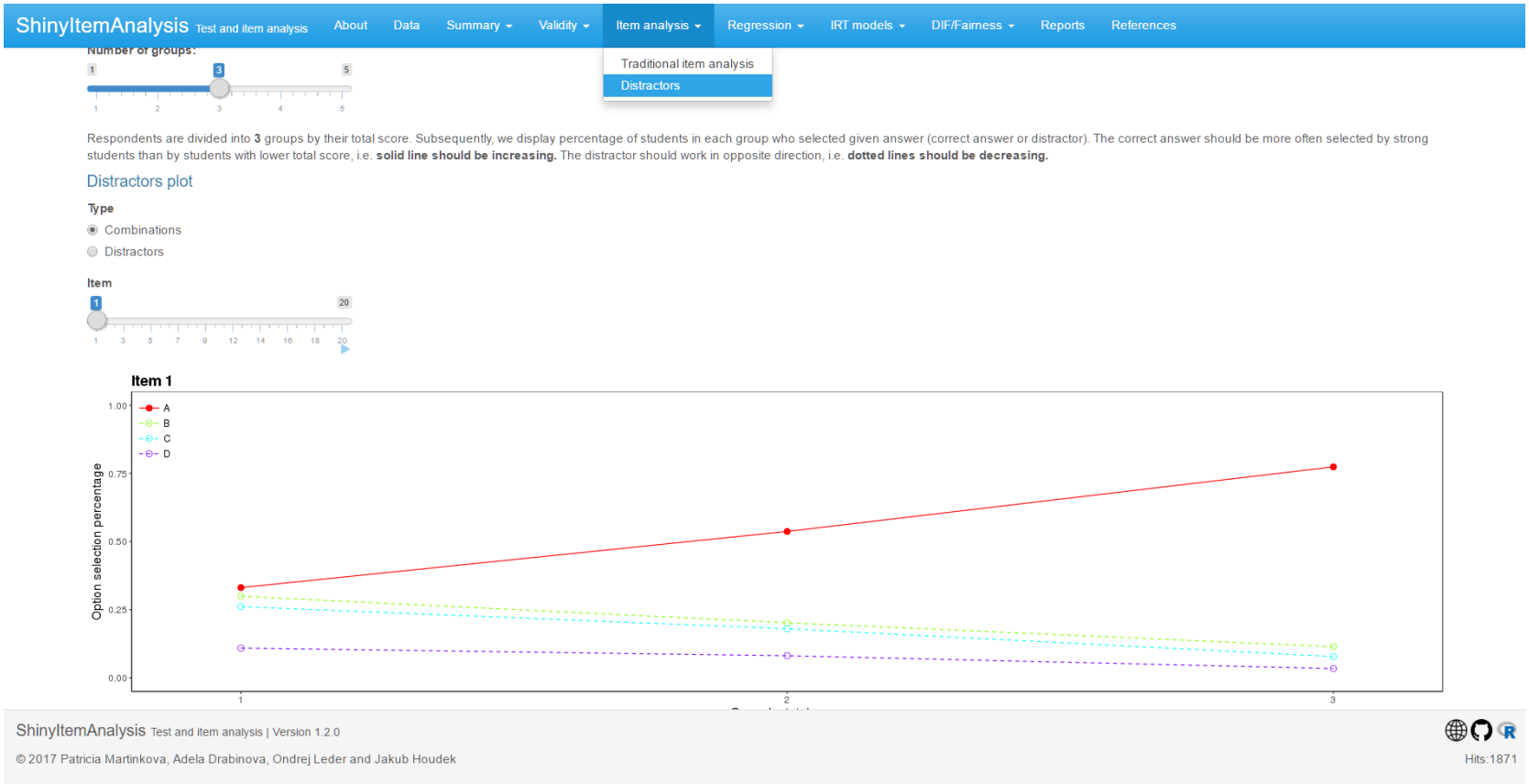


Traditional Item Analysis

- Difficulty, discrimination
- Cronbach's alpha w/o item, index RIT, RIR, etc.

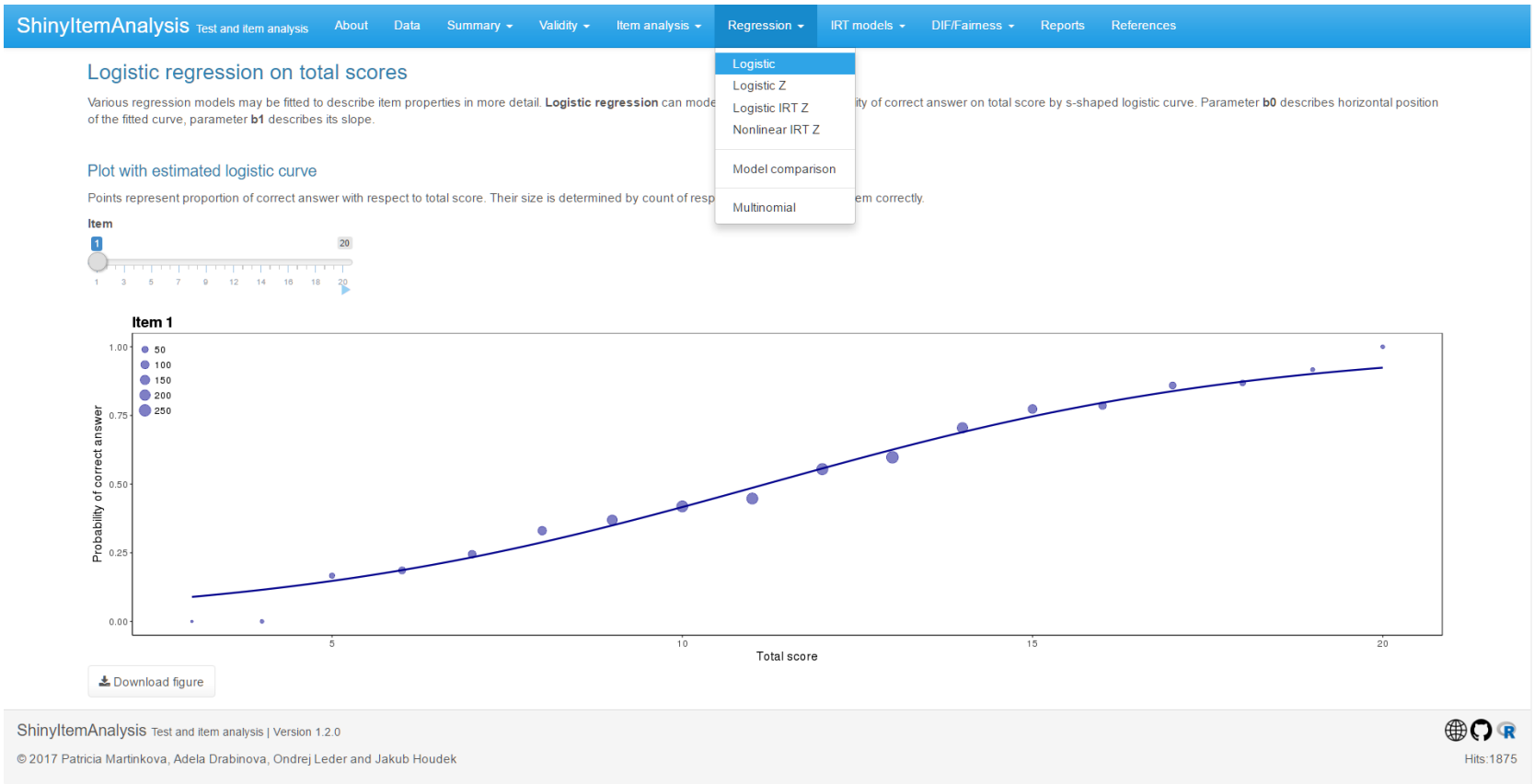
Distractor Analysis

- Displays option selection percentage by total score group
- Number of groups can be changed



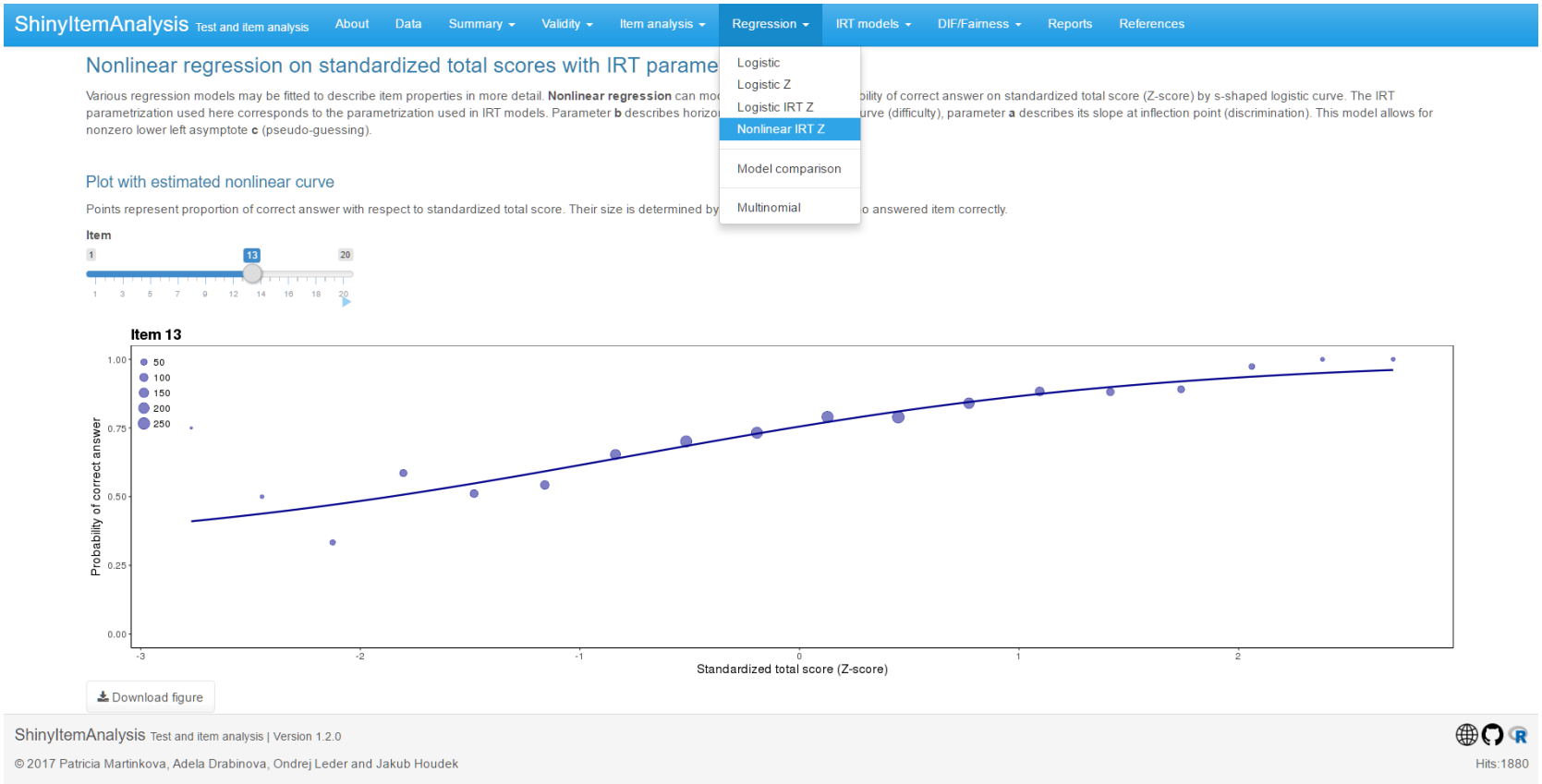
Logistic Regression

- Displays probability of correct answer by total score
- Parameterization can be changed (Z scores, IRT parameterization)



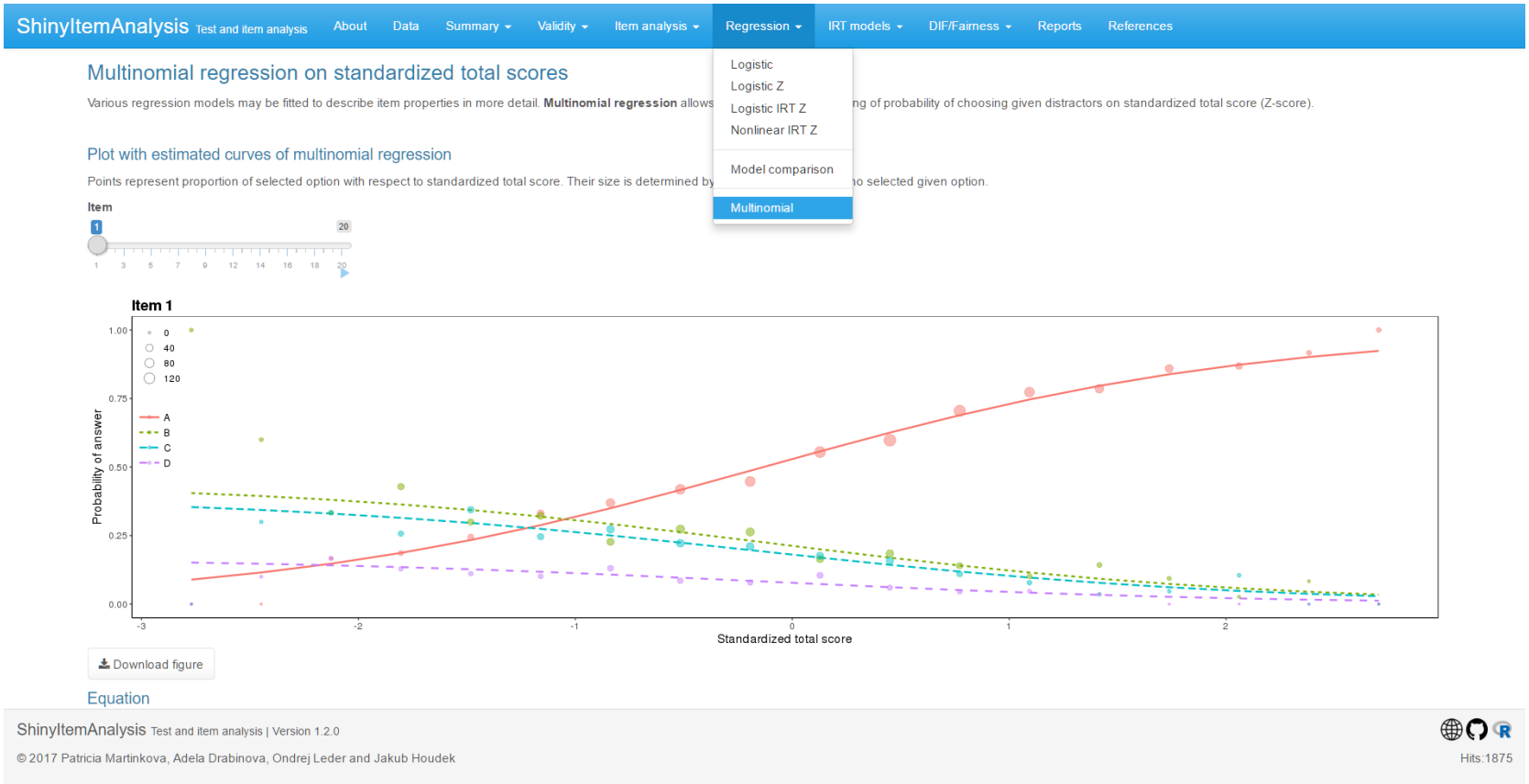
Nonlinear Regression

- Allows for guessing (and inattention)



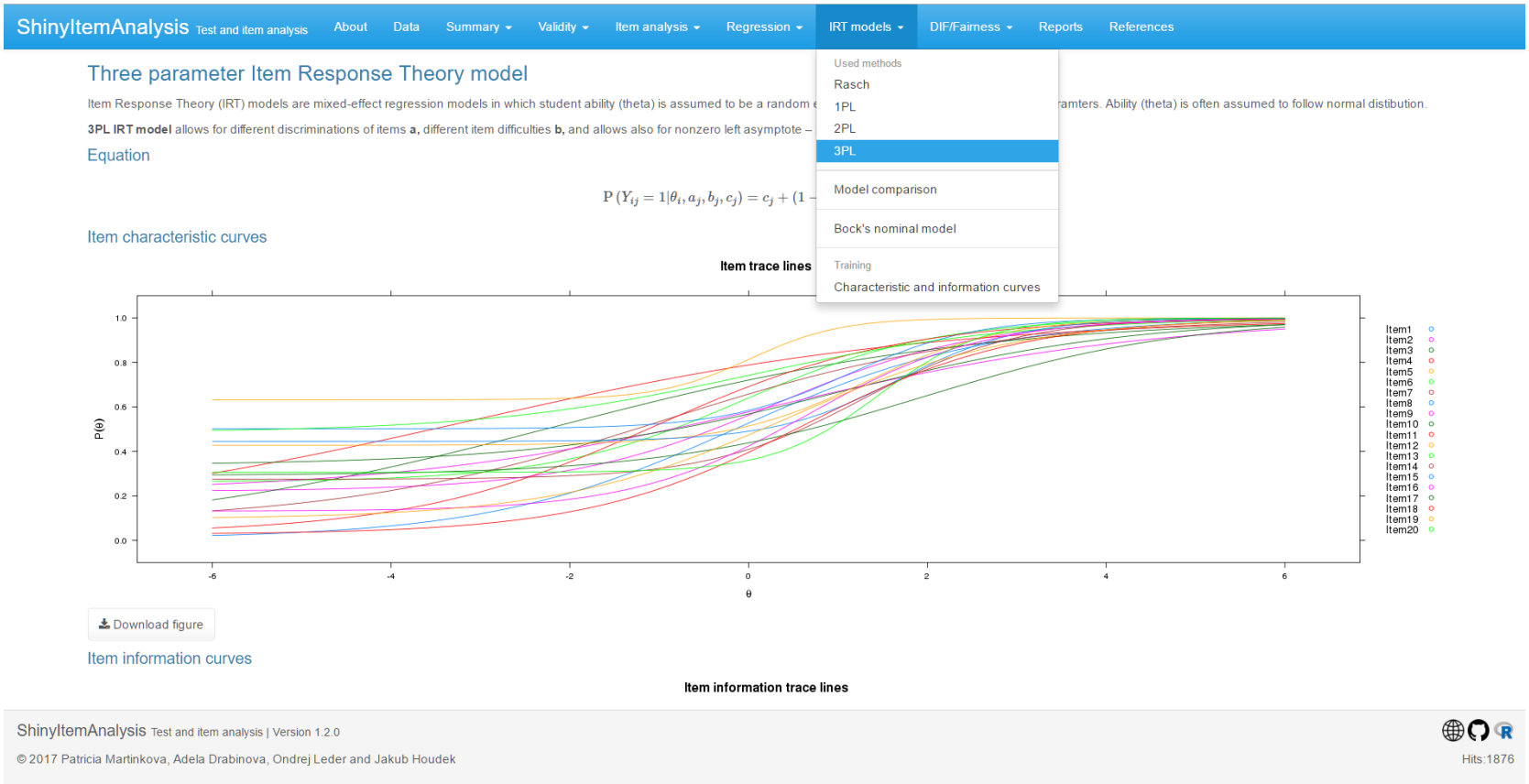
Multinomial Regression

- Allows for joint modeling of distractors



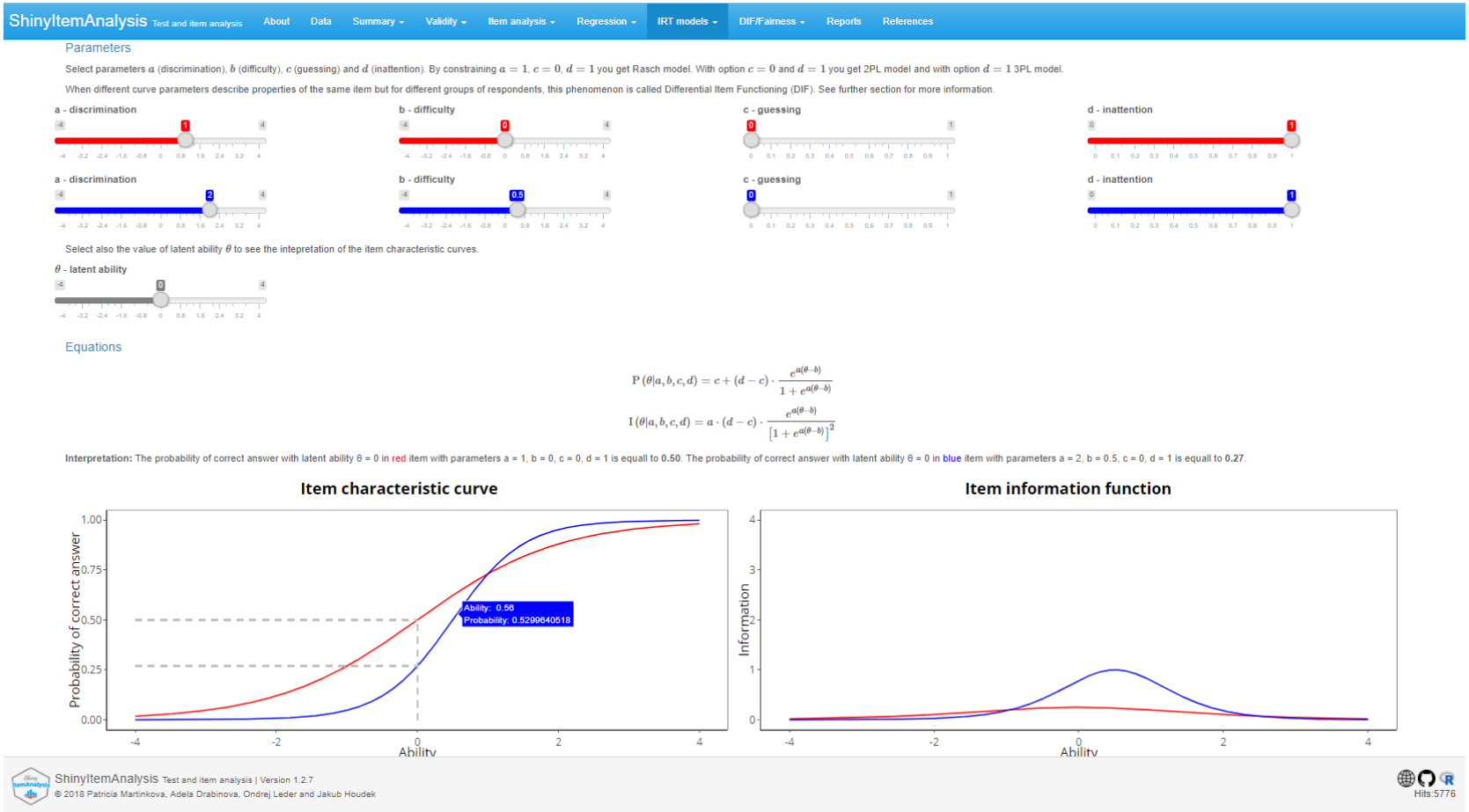
IRT Models

- Conceptualized as nonlinear mixed effect models
- More precise ability estimation



Dichotomous IRT Models - interactive training

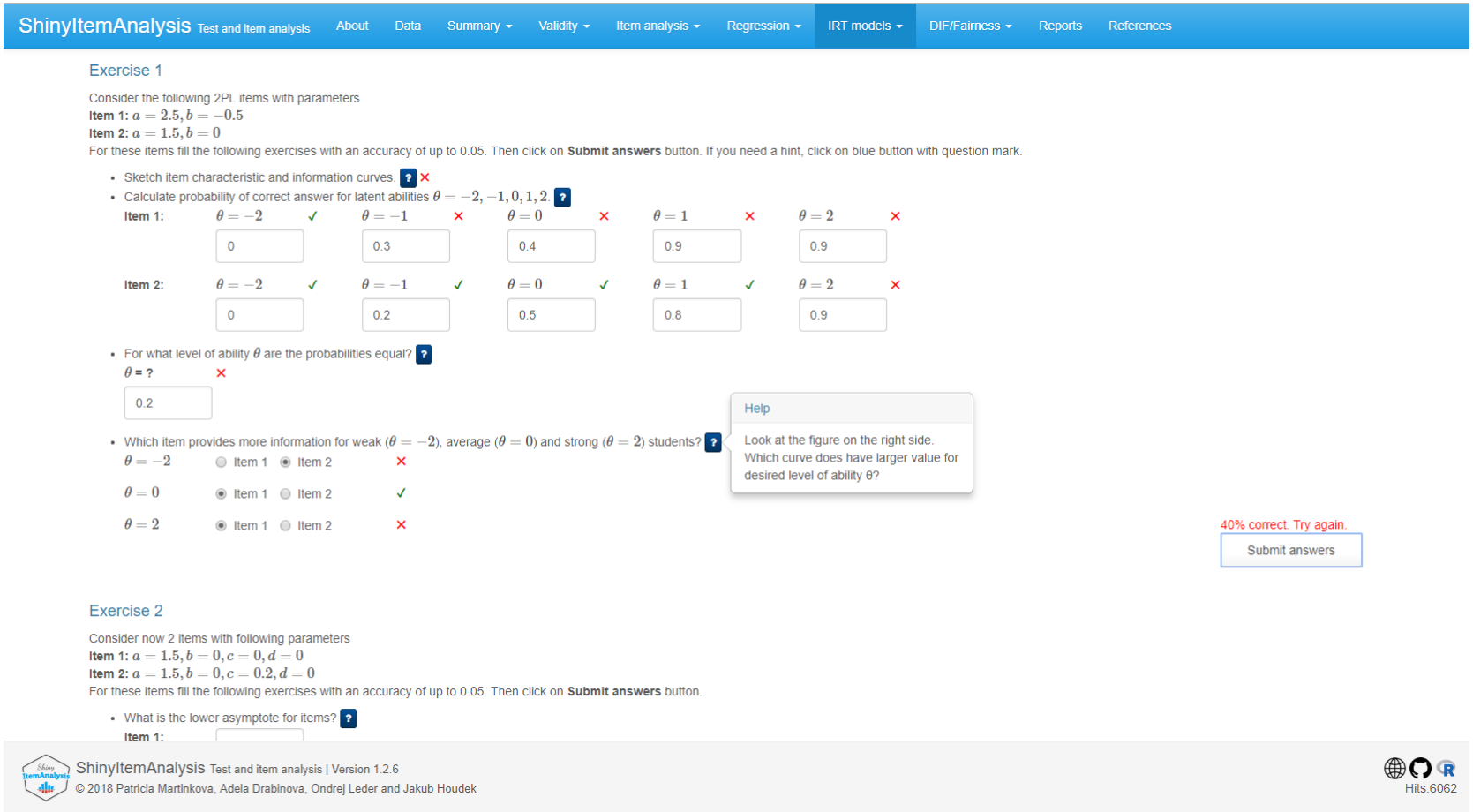
- Plots Item Characteristic and Information Curves (ICC and IIC) based on selected parameters



Polytomous IRT Models - interactve training

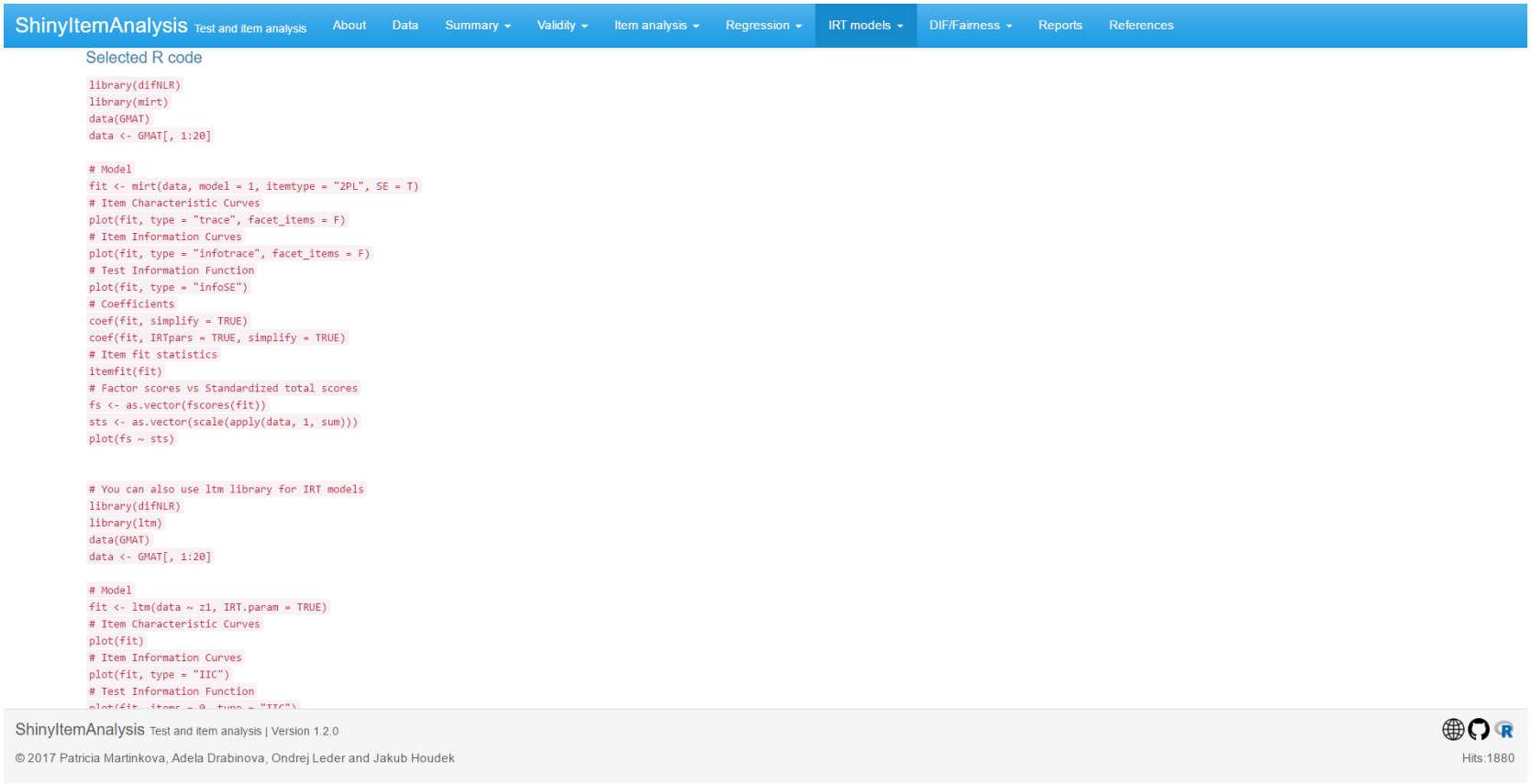
- Plots Category Response Curves and Expected Item Score based on selected parameters

Dichotomous IRT Models - check your understanding



Selected R Code

- Sample R code may be run and modified in separate R session



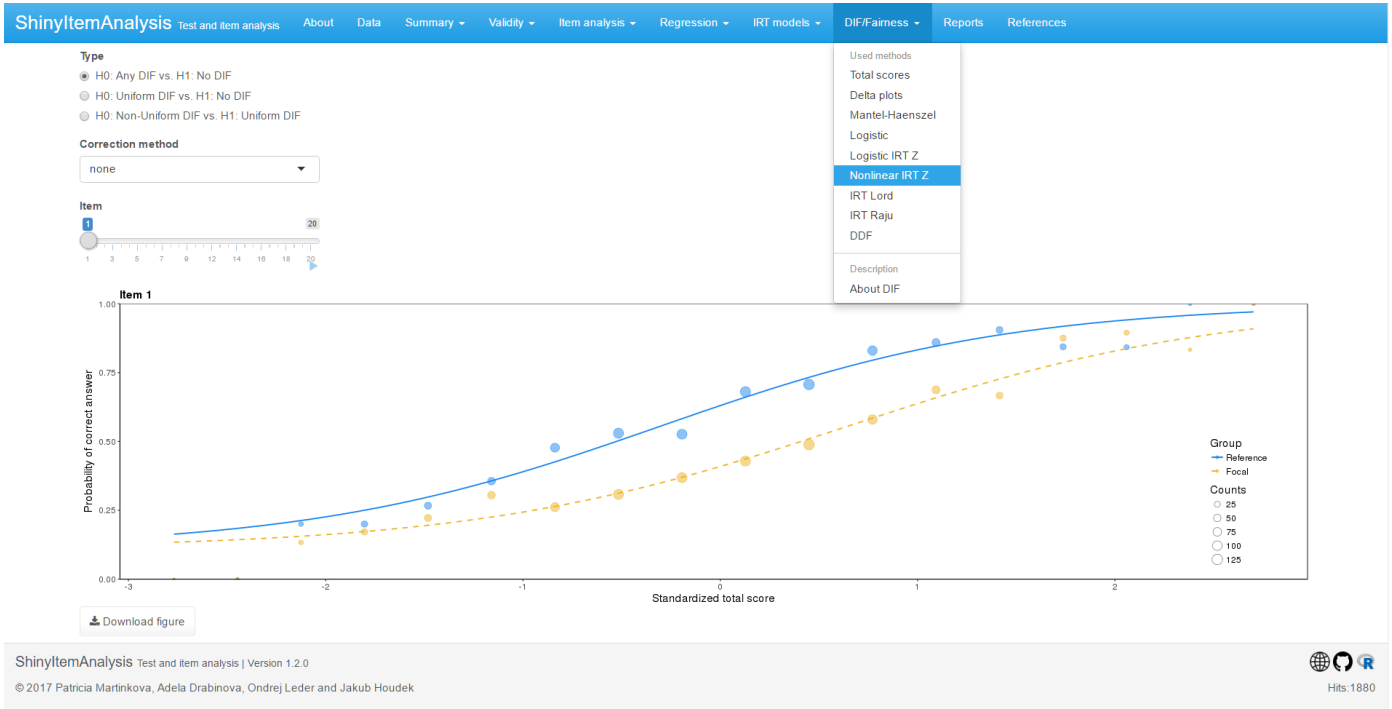
ShinyItemAnalysis to promote our research

App promotes methods and research of our team:

- Detection of Differential Item Functioning (DIF)
- Detection of Differential Distractor Functioning (DDF)
- Why DIF/DDF analysis should be routine part of test development
- etc.

Differential Item Functioning (DIF)

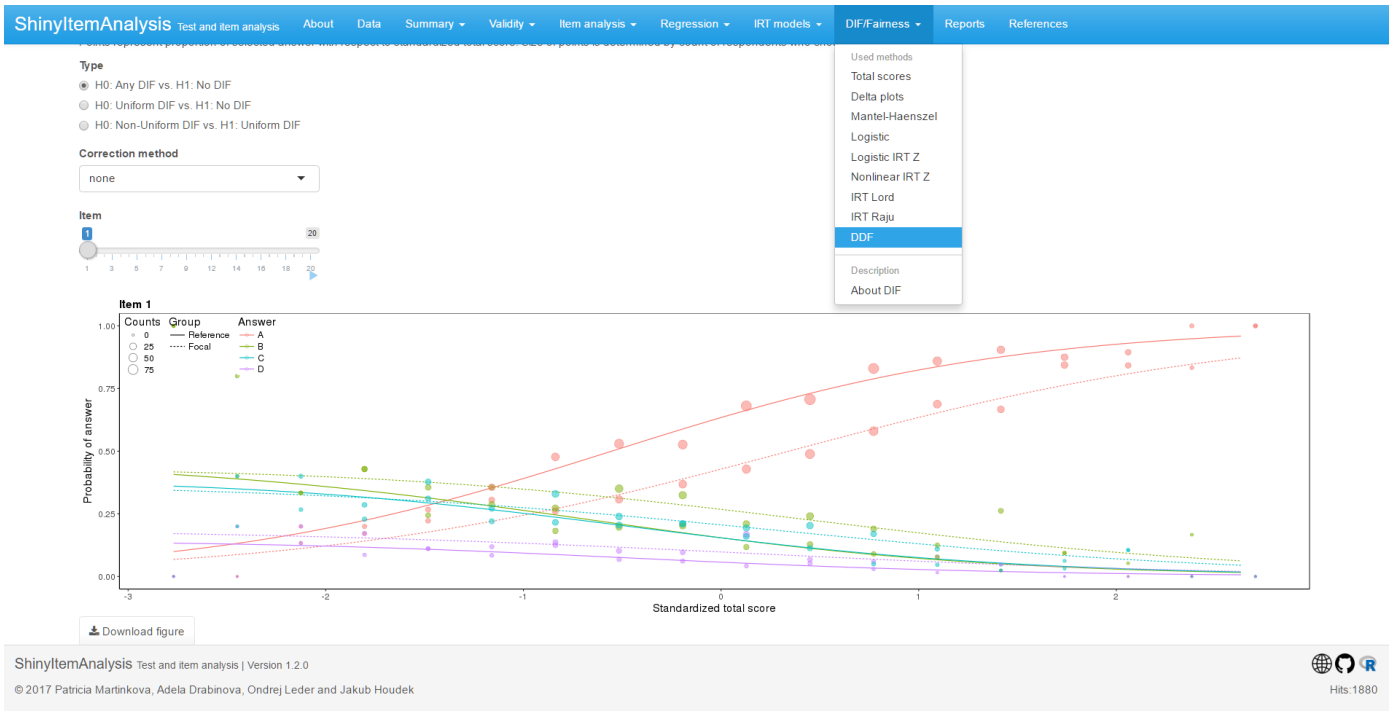
DIF: Students from two groups and *with the same underlying latent ability* have different probability of answering the item correctly.



Drabinová & Martinková (2017): Detection of DIF with Non-Linear Regression: Non-IRT Approach Accounting for Guessing. Journal of Educational Measurement, 54(4), pp. 498-517. doi 10.1111/jedm.12158

Differential Distractor Functioning (DDF)

DDF: Students from two groups and *with the same underlying latent ability* have different probability of selecting given options.



Why DIF Analysis Should Be Analyzed Routinely?

- Dataset HCI: significant difference in total score between males and females, yet no DIF item!
- Simulated GMAT data: total scores may have exactly the same distribution, yet there may be DIF present in some items!

Martinková, Drabinová, Liaw, Sanders, McFarland & Price (2017): Checking Equity: Why DIF Analysis should be a Routine Part of Developing Conceptual Assessments. CBE-Life Sciences Education, 16(2), rm2. doi 10.1187/cbe.16-10-0307

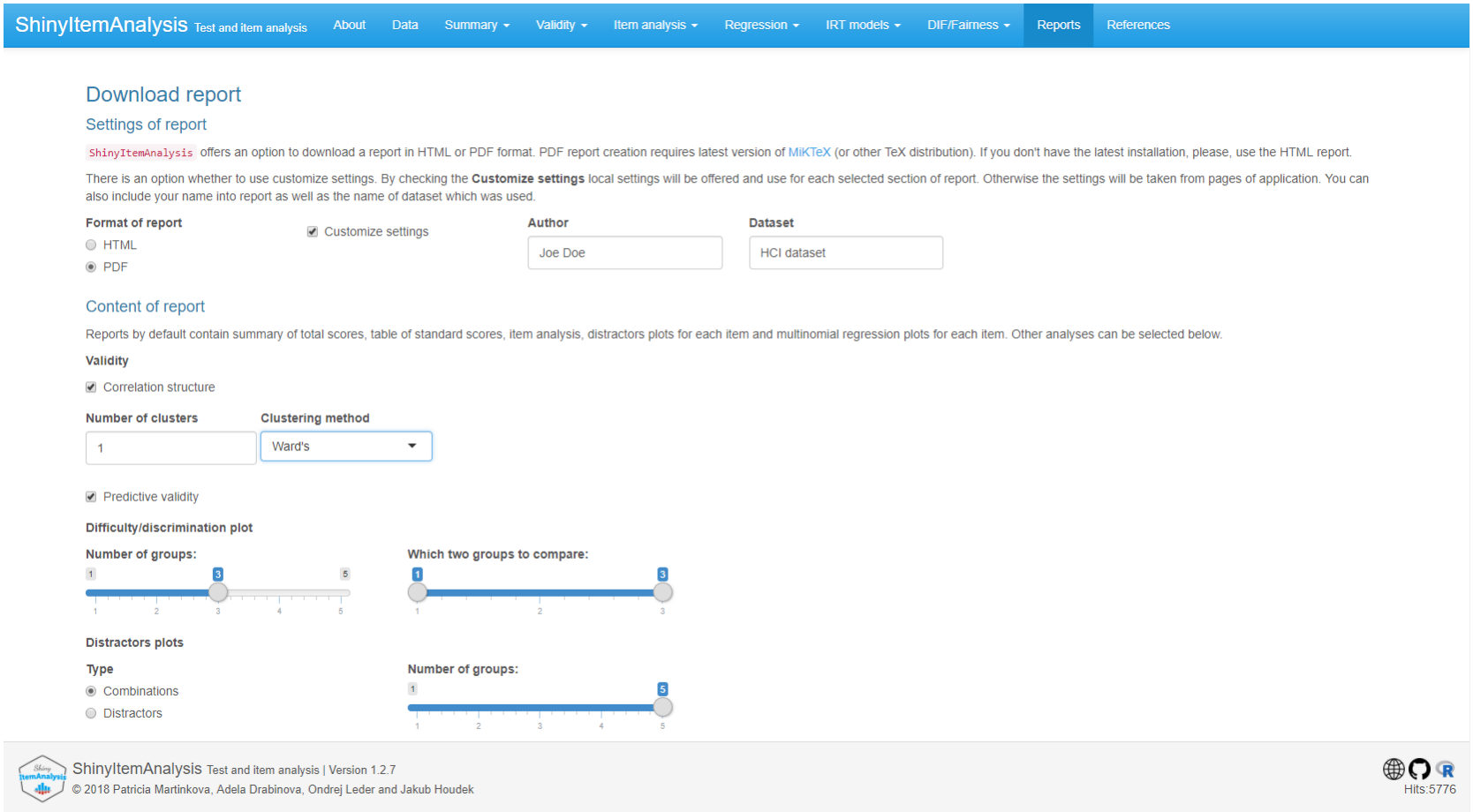
Routine validation of educational tests

Supporting tool for routine validation of educational tests:

- Upload your own data
- Generate PDF/HTML report
- Local or online version

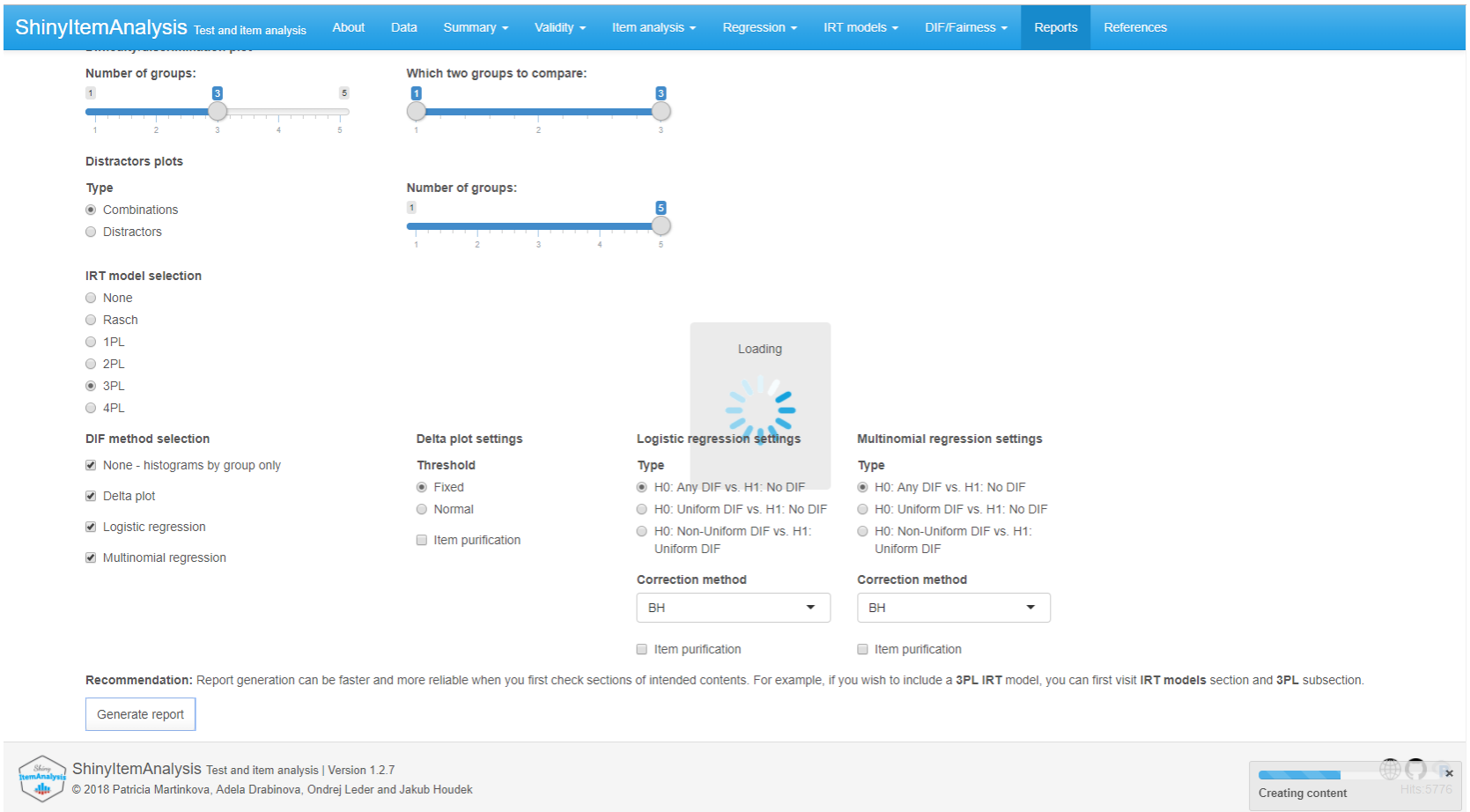
Report generation - settings

- Chose methods, customize settings
- Chose report format (PDF/HTML)

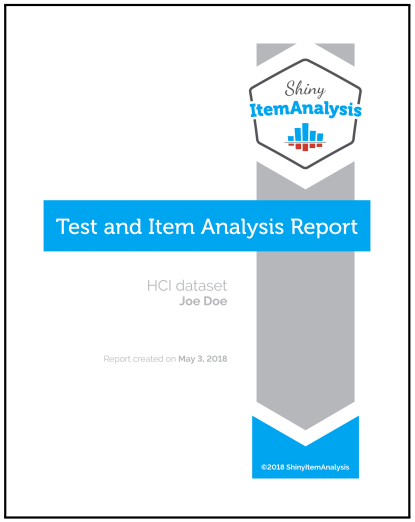


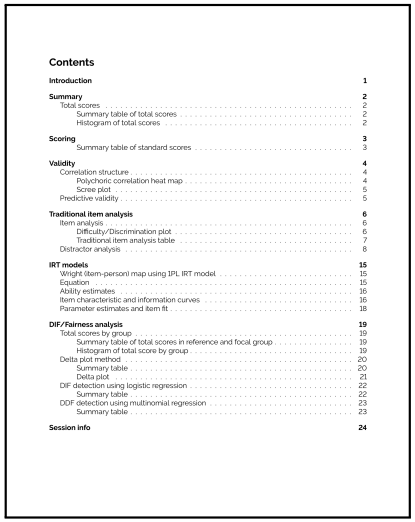
Report generation

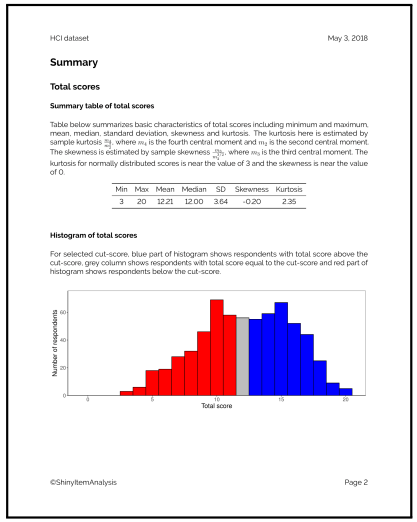
- Generate report (run analyses)
- Download report (compile text into HTML/PDF)

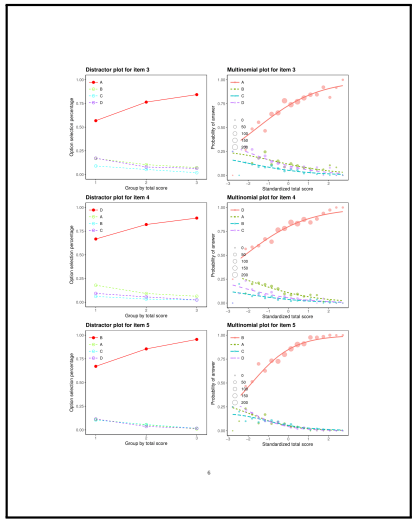


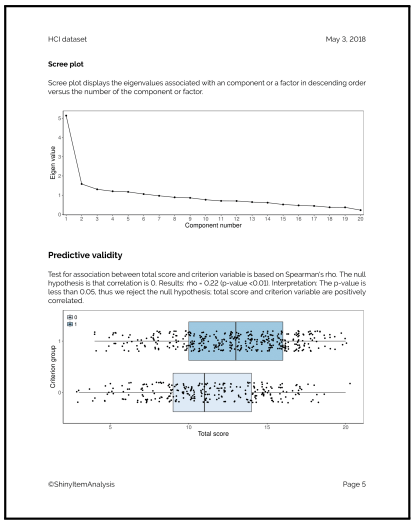
Sample PDF report

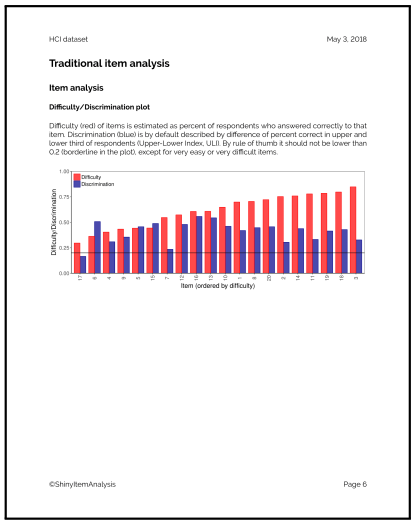


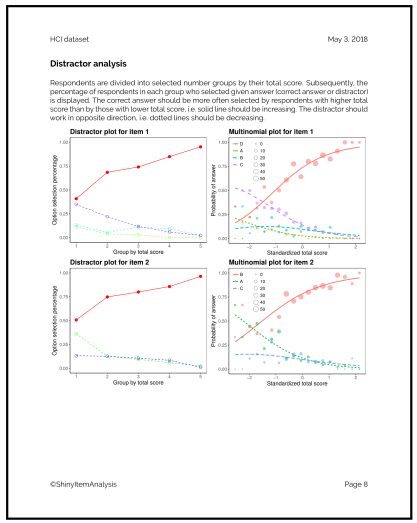


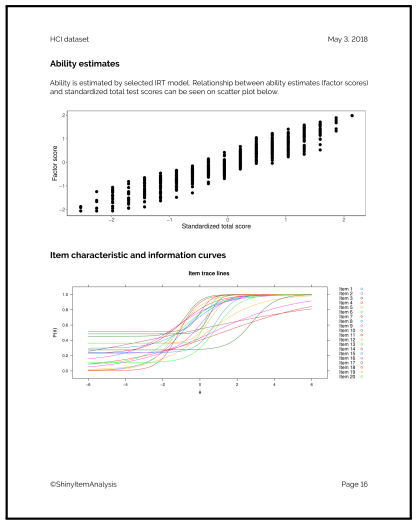




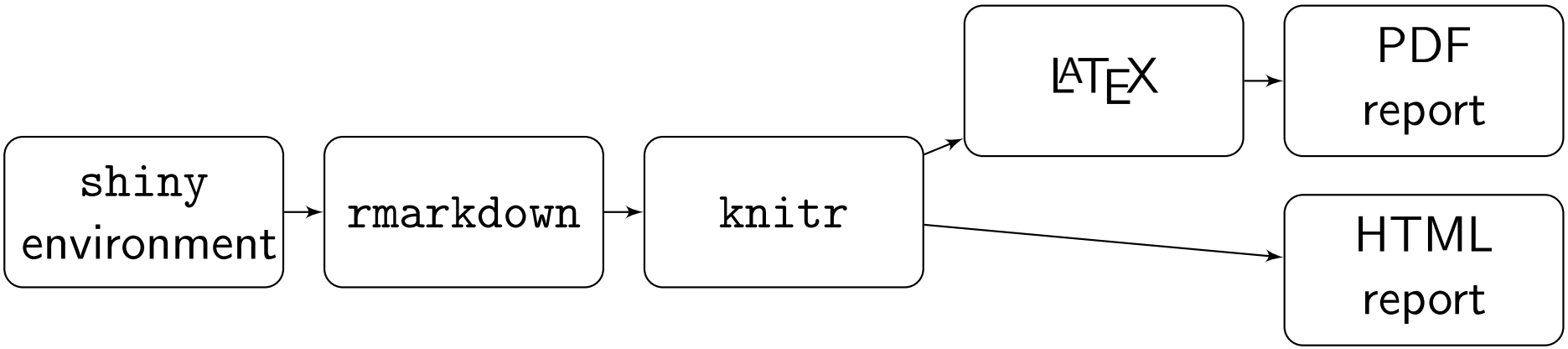








Report generation workflow



- `shiny` provides a user interface
- `rmarkdown` for creating templates for PDF/HTML report generation
- `knitr` for compiling R markdown syntax into HTML/PDF
- TEX for creating PDF reports (latest distribution of LATEX is needed)

Conclusion and Discussion

`ShinyItemAnalysis` is an `R` package and online application for interactive and step-by-step analysis of educational tests. It is useful for:

- TEACHING of psychometrics and educational measurement
  - offers example datasets, upload of new datasets
  - visualization, interpretation of results
  - sample R Code
- ROUTINE VALIDATION OF EDUCATIONAL TESTS
  - generates extensive reports for supplied data

`ShinyItemAnalysis` also promotes our RESEARCH in DIF/DDF detection

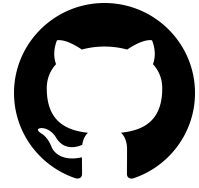
https://shiny.cs.cas.cz/ShinyItemAnalysis/

martinkova@cs.cas.cz

Thank you for your attention!

www.cs.cas.cz/martinkova

martinkova@cs.cas.cz



This research was supported by Czech Science Foundation under grant GJ15-15856Y

References

- Martinková, Drabinová, Leder & Houdek (2017). `ShinyItemAnalysis`: Test and Item Analysis with Shiny. https://shiny.cs.cas.cz/ShinyItemAnalysis/ https://CRAN.R-project.org/package=ShinyItemAnalysis
- Martinková, Drabinová & Houdek (2017). ShinyItemAnalysis: Analýza přijímacích a jiných znalostních či psychologických testů. TESTFÓRUM, č.9, str. 16-35. doi 10.5817/TF2017-9-129
- McFarland, Price, Wenderoth, Martinková, et al. (2017). Development and Validation of the Homeostasis Concept Inventory. CBE Life Sciences Education, 16(2), ar35. doi 10.1187/cbe.16-10-0305
- Martinková, Drabinová, Liaw, Sanders, McFarland & Price (2017). Checking Equity: Why DIF Analysis should be a Routine Part of Developing Conceptual Assessments. CBE-Life Sciences Education, 16(2), rm2. doi 10.1187/cbe.16-10-0307
- Drabinová & Martinková (2017). Detection of DIF with Non-Linear Regression: Non-IRT Approach Accounting for Guessing. Journal of Educational Measurement, 54(4), pp. 498-517. doi 10.1111/jedm.12158