

Now that takes the BISCUIT!

A comparative study of predictive accuracy and interpretability of four statistical learning techniques in personality data, with data missingness conditions

Legend

Sample: 78,828 participants from 200 countries; median age was 33 years; 65% were female.

135 personality items & 27 traits from SPI-27 (Condon, 2018).

Five criteria; bottom left panel.

Source: <https://sapa-project.org>

- Training-test data paradigm
- Regularization penalty
- Variable selection
- Resampling procedure
- Tuning parameter(s)

(a) Split the sample into training (75%) and test (25%) data.

(b) Imposed personality data missingness levels (complete, 25%, 50%, 75%, and 90% missingness).

(c) Multiple imputation using Bayesian principal components analysis.

Elastic net and lasso were the most predictive for complete data and up to 50% data missingness.

Relative to other techniques, BISCUIT performed better as data missingness increased.

Big Five regression performed poorly.

Interpretability was defined as the average number of personality items in an SL technique's models.

Regression using traditional personality traits is usually considered to be the most interpretable, but it requires many personality items.

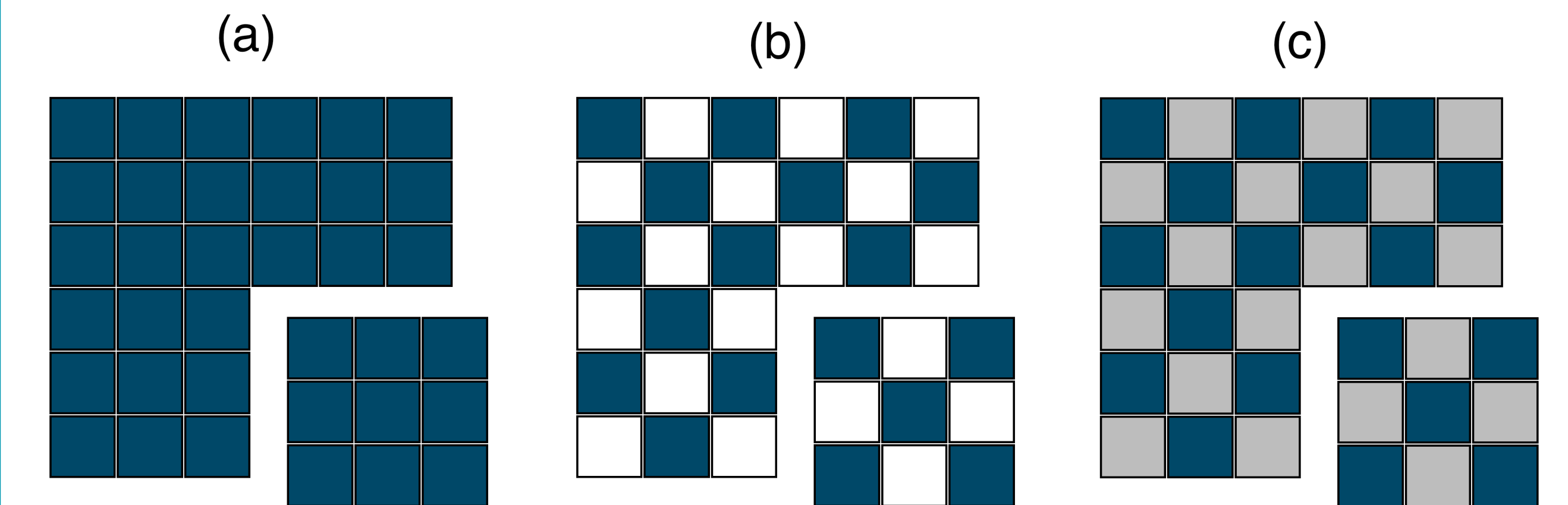
We trained models on levels of data missingness but tested them on complete data.

With few items in its models and little loss in accuracy due to missing data, BISCUIT is cost-effective for building and applying personality-criterion models to real-world predictions.

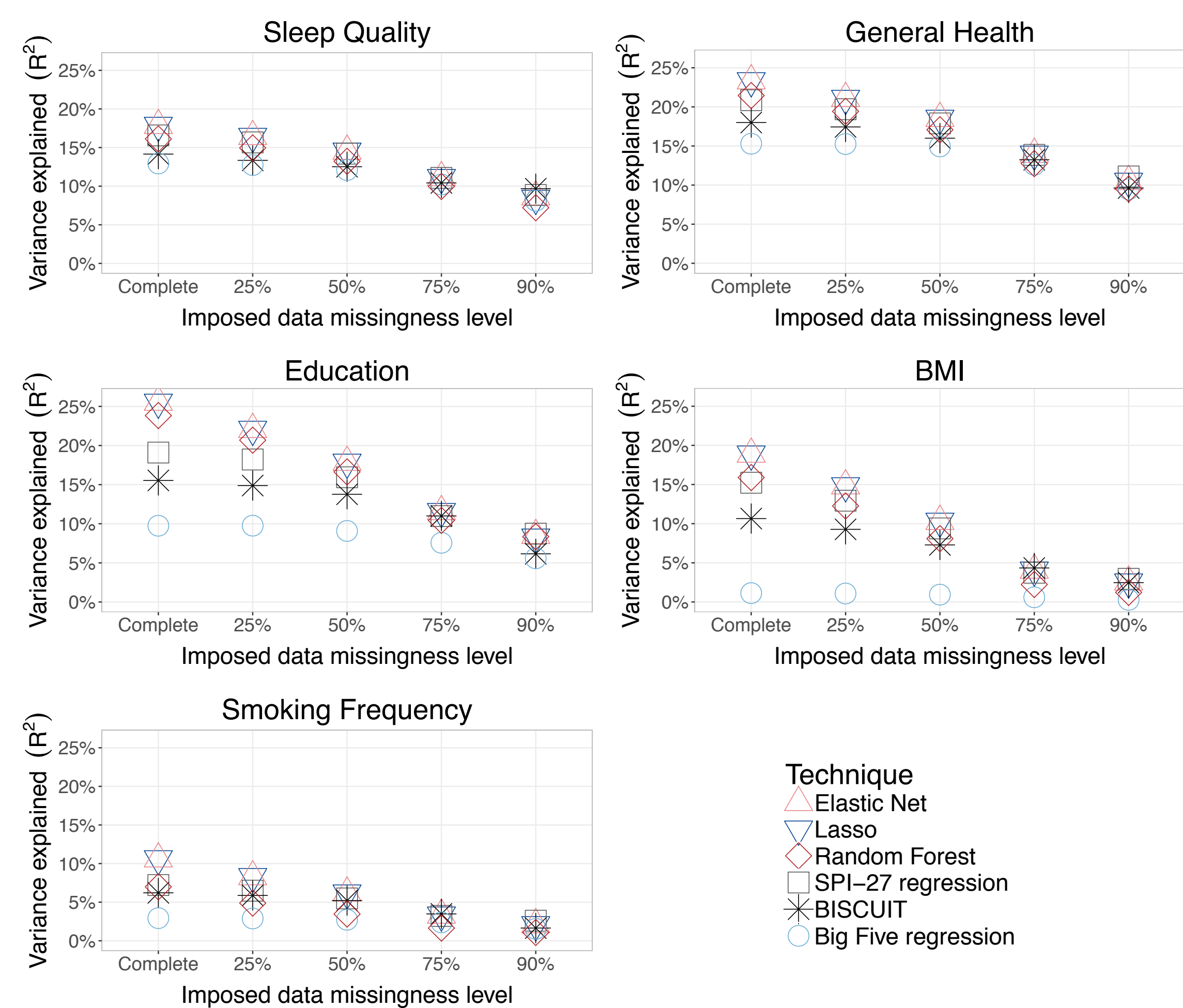
Features of each technique



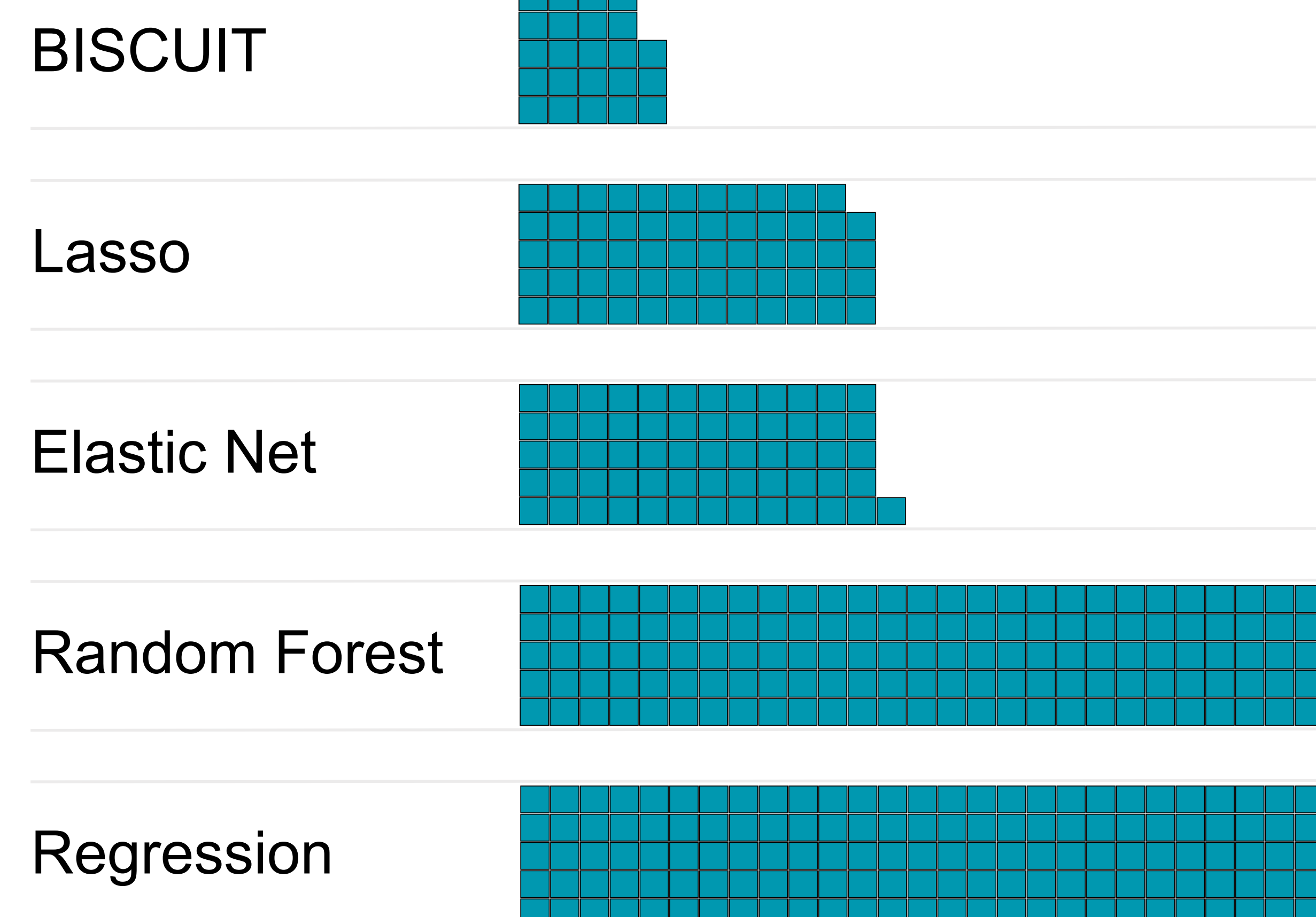
Imposed data missingness conditions: As easy as ABC



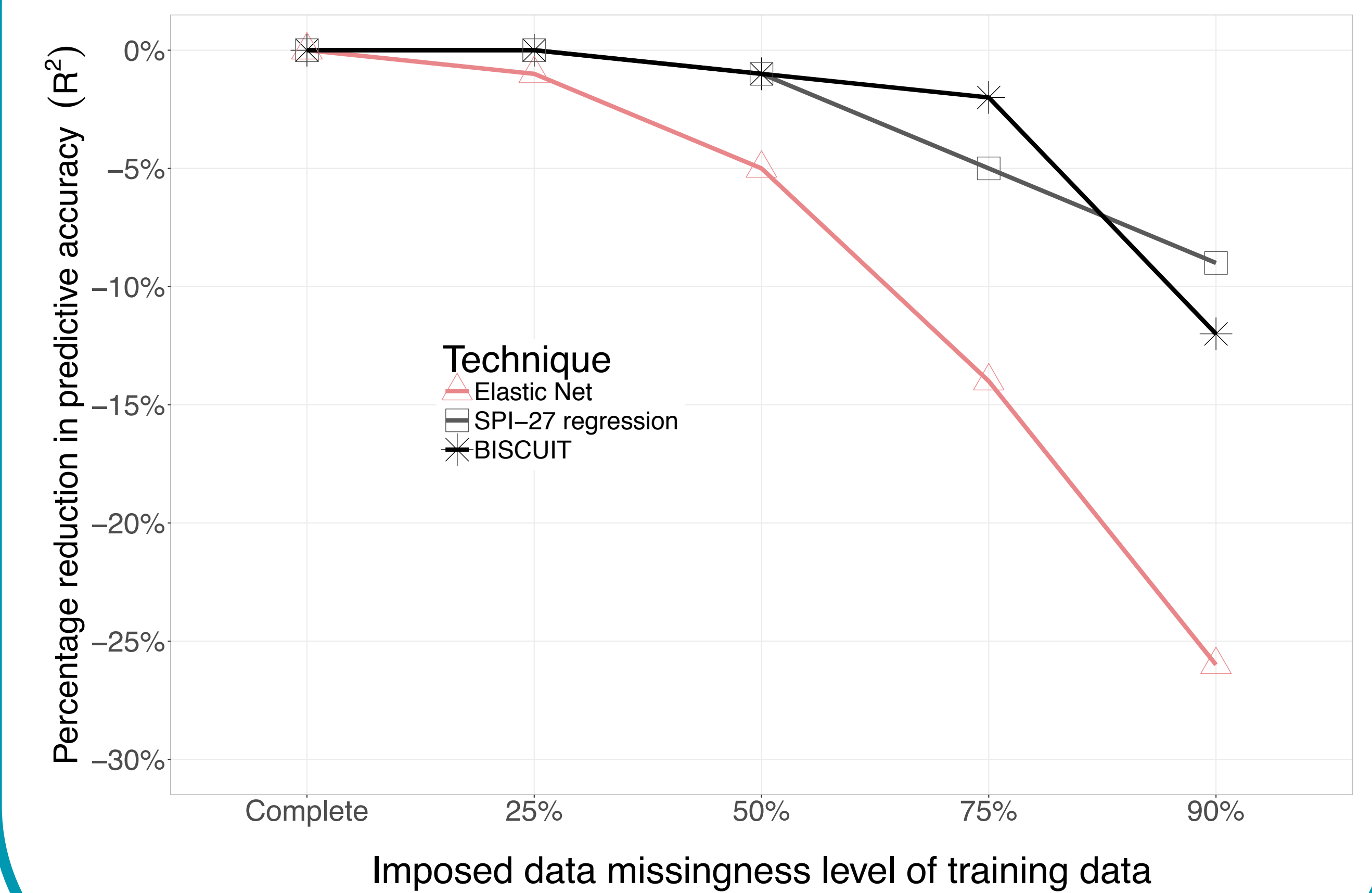
A battle royale of predictive accuracy



Interpretability: Item count for each technique



Almost no loss in predictive accuracy



Lorien G. Elleman¹, Sarah K. McDougald¹, David M. Condon², and William Revelle¹
¹Department of Psychology, Northwestern University; ²Department of Psychology, University of Oregon

Email: lgelleman@u.northwestern.edu

Website: <https://sapa-project.org/lqe>

Twitter: @LorienElleman

Icons from thenounproject.com
 Artists: Eucalyp; Rose Alice Design; Ralf Schmitzer, DE; Royyan Wijaya, ID; Kay Creative

Condon, D. M. (2018). The SAPA Personality Inventory: An empirically-derived, hierarchically- self-report personality assessment model. <https://psyarxiv.com/sc4p9/>

*Best Items Scale (that is) Cross-validated, Unit-weighted, Informative and Transparent

