

Interactive Intelligence

Checklist for Review of Dataset

(Version 1)

We recommend that students or employees wishing to publish on their data and results for a given research project in the form of a dataset asks a fellow student or colleague to review this dataset with regard to the points in this checklist. The purpose of the checklist is to ensure that all data that can be made available is made available, that all analyses were conducted conscientiously by the researchers, that all results are reported accurately, and that all methods are transparent and sufficiently clear to be reproducible.

If you choose to have your code reviewed according to this checklist, we advise you to upload this document together with your dataset to the research data repository of your choice (e.g. 4TU Research Data) upon publication of your work.

I. Basic Data

Paper title:	Everyday Locations as Cues to Smoke
Name(s) of researcher(s):	Alkis Antoniadis
Name of the reviewer:	Ramya Ghantasala
Data repository platform (e.g. 4TU Centre for Research Data):	4TU Centre for Research Data

II. Checklist

Statement	Yes	No
1. The dataset contains a README file that fulfils the requirements of the data repository platform that the researcher wishes to use. If no such requirements can be found, the dataset nonetheless contains a README file that clearly explains the contents of the dataset?	README mostly covers the requirements	4 is missing(not sure if applicable)
2. Either within the README file or within an extra, easily findable file, the researchers have explained their data. This means that, for example, for every column of a tabular dataset, all column names and possible cell values are explained.	README file describes the columns in each dataset.	
3. data is in readily readable file formats. If this should not be the case, the README (or similar) clearly explains the file format and which software can be used to access the contents.	Data is mostly in easily readable file formats.	ipq_data.sav is missing mention of software required to open
4. All data has been anonymized in accordance to promises made in the Data Management Plan.	Unsure of promises made in Data management plan, but data seems anonymized	
5. The analysis file or files contain a header with meta-data (name of author, date of writing, required input files and generated output files).	Name of author, date and required input files are mentioned	Output files don't seem to be mentioned
6. All required input files for the analysis are available in the dataset.	Required input files for analysis are present.	

Statement	Yes	No
7. There is an output file that is generated by the analysis script that neatly combines code and commentary (e.g. markdown output file). This output file is in a readily readable file format (e.g. pdf).	The knitted pdf files are present in Knitted Analysis Files folder	Some portions of Experience_Familiarity_Analysis pdf file, Prior_Sensitivity_Analysis pdf contains code that is outside the display area, and partly unreadable.
8. The analysis script is clean and comprehensible in the sense that: <ul style="list-style-type: none"> • There is sufficient, useful, and clearly written commentary • Irrelevant code (such as old analyses) has been removed • The details of analyses that are not reported in the paper (e.g. assumption checks) are proportional to those that are reported in the paper. This means that unreported analyses should not clutter up the script, making it long and unreadable. 	Analysis script is mostly clean and comprehensible.	In Prior Sensitivity analysis file there are generated plots which do not seem to be used in the thesis report (not sure if required). Experience_Familiarity_analysis contains figure (WAIC comparison) which is not used in the report.
9. The analysis script can be run successfully.	Analysis scripts ran successfully.	
10. All preprocessing steps are clearly described and traceable, especially when preprocessing code cannot be executed because raw data is not available.	Preprocessing steps are described clearly	
11. The analyses and results reported in the manuscript can be found back in the analysis script with labels according to where they appear in the manuscript.	Analysis and results were mostly referenced correctly.	Figure numbers mentioned do not match, but is probably due to only 1 chapter being available.
12. All results reported in the manuscript accurately correspond to the output produced by the analysis script.	Results reported in the manuscript mostly correspond to the results obtained when reproducing.	Experience_Familiarity_Analysis: Some results do not match obtained results from script. (Table 4.3) Sense_of_presence: Rounding is sometimes not consistent between report and obtained results (e.g. 0.9959 is reported as 0.995 instead of 0.996). AUQ, EoMC Confidence intervals in Table 1.4 does not match obtained result.

III. Additional comments by reviewer

- Would recommend naming figures with respect to chapters for easy referencing.
- Rounding is inconsistent in a few places (rounding off vs truncated). Would be good to have a consistent method for rounding.

IV. Review log

Round	Date	Paper Status	Checklist Items	Signature Reviewer	Signature Researcher
1	15/03/2022	Draft MSc thesis sent to thesis committee for further feedback.	12	Ramya Ghantasala	