COG260: Data, computation and the mind

Course project

In this project, you will work with your project partner to formulate a cognitive hypothesis and test it against an extensive public data set. The default data set you will be working with is The World Color Survey or WCS (manual), which includes color naming and behavioural data from 110 non-industrialized languages. You have the option of working individually and/or with an alternative data set of your choice that is similar in complexity to WCS. For instance, you can work with a secondary, large dataset called Database of Cross-Linguistic Colexifications or CLICS (descriptor) that explores colexification (that one word labels multiple meanings), or word-meaning mappings across the world's languages. Note that CLICS is substantially larger than WCS, but for both datasets we provide you with a demo Python Jupyter Notebook to get you started. If you choose to work on a dataset outside WCS or CLICS, you will need to obtain permission from the instructor *prior to* the project proposal due date.

For the analysis of WCS or CLICS, you may work with the demo Jupyter Notebook provided. For analysis with data sets sourced elsewhere, you are responsible to develop your own Jupyter Notebook. All analyses should be performed in Python Jupyter Notebook. It is sufficient to turn in **a single copy** of the proposal and the final report between you and your project partner (make sure to specify your names in the front page). Late submission will receive a deduction of 1 point per delayed hour. Submit your proposal and report in PDF format through Quercus. No other laboratory work will be assigned during the project period.

Detailed description

Stage I. Initial proposal and progress [5pts]
Submit a PDF proposal (about 2 pages) that describes a concrete plan for your project. Your proposal should include 1) a description of a specific and testable cognitive hypothesis with justification, e.g., citable references [1pt] 2) proposed methodologies, models, and analyses [1.5pts] 3) expected results such as any figures and/or tables you will generate from the proposed analyses [1pt] 4) division of labour between you and your collaborator, if you work in pairs [.5pt]. 5) In Appendix (outside the 2-page proposal), provide concrete evidence that you have been making progress toward the project [1pt]; such evidence may include an initial executable Python notebook that processes, visualizes, and/or analyzes the data relevant to the hypothesis you proposed, with corresponding output from the code presented in such forms as tables or figures. Your proposal needs to be feasible given the time frame of the project (about 1 month). Invalid or infeasible proposals will potentially delay your progress; they will be returned and revised. If time permits, you will get a chance to briefly introduce your project proposal to the class after you submit it.

Stage II. Data blitz (final day of class) [10pts]
You and your project partner are required to present in the data blitz (or final project presentation). It is recommended that you split the presentation into two even halves since you will be assessed individually, not as a group. Slides in PDF format should be submitted as PDF through the online submission system at least 1 day prior to the presentation. Failure to conform to the required PDF format will receive 2 points deduction; delayed submission of your slides will incur the same penalty as any late assignment. During the data blitz session, presentation order will be randomized. Failure to attend the data blitz will receive 0 credit. Presentation should be kept under 5-6 minutes and will be timed strictly.

Stage III. Final report (due the Wednesday after data blitz) [15pts]
Submit a PDF document of no more than 6 pages (excluding Appendix for Python code) that includes:

1) *Abstract*: In 1 paragraph, summarize your hypothesis, methods, findings, and conclusions.

2) *Introduction*: Provide background on the project. Justify and articulate your hypothesis. Specify at least 1 alternative hypothesis.
2) *Methods*: Present the methodologies for your analyses in a way that is reproducible. Provide the full set of Python code in Appendix that allows reproduction of your reported findings.
3) *Results*: Summarize the main results from your analyses by making use of figures, tables, and statistics. Explain how these findings support or refute your hypothesis.
4) *Conclusion*: Conclude in 1-2 paragraphs; discuss possible extensions.
5) *References*: List at least 5 full references with in-text citations.

The final report should be written using the following latex template (reports not following this template will receive zero point): `https://github.com/rlevy/cogsci-template`