Using flipbooks to build outcome-oriented tasks

For *Chance Magazine's* 'Taking a chance in the classroom'

Gina Reynolds, March 2020

Chance is a quarterly non-technical statistics magazine published jointly by the American Statistical Association and Taylor & Francis Group. It was established in 1988, and Taylor & Francis has published it since 2012. The magazine sponsors the blog "The Statistics Forum", which allows anyone to post their thoughts on probability and statistics.

https://chance.amstat.org/

https://chance.amstat.org/category/columns/taking-a-chance-in-the-classroom/

CHANCE is copublished quarterly by the American Statistical Association and Taylor & Francis Group The magazine is designed for anyone who has an interest in the analysis of data, informally highlighting sound statistical practice. CHANCE is not a technical magazine, but rather a cultural record of an evolving field, intended to entertain as well as inform.

Since its creation in 1988, CHANCE has covered such topics as the 1990 census adjustment and the redesigned population survey, sports, the environment, DNA evidence in the courts, a variety of medical issues —even how to win on "Jeopardy." CHANCE offers a unique opportunity to reach beyond statistics professionals to a more general audience.

While we invite authors to submit articles in ready-to-publish form, we also encourage prospective authors to submit ideas for articles or articles that are not yet in final form, and we will try to help with their development.

Writing for CHANCE We prefer short manuscripts of 10 to 15 double-spaced typewritten pages. Authors are encouraged to write in a relaxed, expository style; opinion, humor, and historical comments can (and should) be intermingled to make lively reading. Additional suggestions regarding writing for CHANCE are offered under Hints for Authors.

Submitting Files Submit material via email as a PDF document and a Microsoft Word file to the editor at the following address:

Amanda Plunkett CHANCE Editor chancemag.editor@gmail.com CHANCE Style Details for listing and citing references are given in the ASA Magazine Style Guide (PDF).

One goal of data science and statistics classes is to give students tools to make them comfortable preparing their data and analyzing it. This requires not only introducing concepts and tools, but also providing guided practice with these tools. A new way to facilitate this practice with the R statistical programming language is by using the new flipbookr package to build exercises.

The flipbookr package is a tool that was built to showing analysis code pipelines (manipulation, visualization, statistical analysis), step-by-step with output in a slide-show presentation; this yields a movie-like experience linking new code to the change in output that it triggers. Showing how code evolves alongside output can illuminate in-between steps that might not be clear to newcomers if only the initial input and final output of the code pipeline are show.

But flipbookr can also be used in a mode where *only output* is displayed. This mode is useful for the data science and statistics classroom use --- building tasks that are outcome-oriented.

An flipbook designed to build data manipulation practice for example would be organized as follows. There is a reference section of what functions might used in the exercise and a reminder of what each accomplishes. Then a target is introduced --- some final output from a familiar input data set. Subsequently, the in-between states of the data manipulation will be displayed for the first target. Subsequent exercises are structured in the same way; first the target output is presented, and then a walk-though from input through in-between states all the way to the target are presented.

Some students will be able to figure out the steps needed to get to that target, just by looking at the final target. They are encouraged to build the data manipulation pipeline w output, and then compare their strategy with that of the walk-through. Other students may not be able to come up with the entire sequence of steps needed to wrangle to the output. These students may look ahead to the walk through. These students may be able to accomplish the task by taking it step-by-step, or they may be able to at least manage some progress towards the overall target, perhaps getting some assistance from a peer or classmate.

When the manipulation steps are presented in this incremental fashion, students may be able to take on the task or at least some of the task. The subliminal message here is: you may have some big task ahead of you, but try to break up the task into pieces --- it will be more manageable that way. As an instructor the task of assisting students who might have trouble may also be more manageable. Students are less likely to be stumped by the entire data manipulation challenge; and even if they are, the task is already logically broken up for the instructor.

Students will likely approach the instructor about the specific step in manipulation pipeline rather than an amorphous 'I don't know how to get this done!' The exercise is posted online in a slide show that walks through the data manipulation outputs.

One advantage of this exercise is that it de-emphasizes specific tools; where as the focus can sometimes become syntactical, the data transformation itself is more central to the structure of this task (subsetting, keeping columns, summarizing etc). In fact, a large number of statistical software or syntactical paradigms can be used to approach the very general problem presented in the exercises. I teach this using 'the tidyverse' which actually does have some signatures in the output (like there is a separate group_by step when compared to data.table for example). But the manipulation tasks are general, and could also be solved in Stata, SPSS, SAS, R tidyverse, base R, data.table, python's pandas -- something yet to be invented! This shifts the emphasis from *how* (specific implementation and syntax) to

The initial exercise has about 15 wrangling pipelines - some short, some longer -- it is a good amount of practice which is so important for internalizing new tools and modes of thinking. The first half I call 'one-stream' data manipulation and the second half is on summarizing data.

In this essay, I've discussed an exercise for practicing data manipulation, but tasks can also be designed for data visualization. Students are presented with an overall target output (a visualization) given a well known input (a familiar dataset), but are also shown the in-between outputs, breaking up the task into smaller, more manageable tasks.