**ACTION ITEMS**

* Josh emails Hannah Thursday 6/20 to tell her that we are working on our response
* Write response collaboratively in this Google Doc
  + 2-4 sentences for each topic:
    - (1) Tidyverse vs. base,
    - (2) Network data,
    - (3) New walk-through on longitudinal/survey,
    - (4) Audience
* Send to Hannah by email by Wednesday 6/26

***Draft of email to Hannah***

Hi Hannah,

As we mentioned, we really value the feedback we received from reviewers. We have developed a plan to address their comments, and we outline that plan below.

**Draft responses should be in a different color so we know what to copy and paste**

**Tidyverse vs. Base R**

* Community around the tidyverse is more robust
  + Tidyverse and base R are not the same
  + Here are the things that
* Marginal differences in speed irrelevant to our audience (e.g., not trying to optimize function speed, just trying to help people make the move from SPSS to R or Excel to R
* If you learn to code, need to keep knowledge updated anyway (to address the comment from the reviewer about base R) - need to keep tools up to date, etc.
* Maybe talk about intuitive differences
* Feedback overstating the risk a little bit
* It’s not base R vs. tidyverse - the two approaches to R work in complement with one another. However the tidyverse, based on the consistent syntax between associated packages, provides a more beginner-friendly on-ramp to learning R.
* In addition, the tidyverse requires tidy data - data that fits a predetermined structure. This offers beginners a clear and consistent marker of success when wrangling their data by allowing them to self-check responses to three questions:
  + Is each variable in its own column?
  + Is each observation of a variable in its own row?
  + Is there only one value per cell?
* Dave Robinson - arguably one of the best data scientists out there - has great points on tidyverse first, largely that it gets beginners doing powerful things quickly (<http://varianceexplained.org/r/teach-tidyverse/>). In addition, if you *start* with the tidyverse it doesn’t mean we can’t introduce base R concepts later on!
  + What about building off of the cheat sheet Amelia McNamara made, that translates base and tidyverse? We could start incorporating base R in later chapters/working problems

We will improve our proposal by explaining that, while we chose to use Tidyverse functions to meet the goals of the book, there are other approaches in R and other programming languages that can accomplish the same analytic tasks. We will include links to both Tidyverse and base R resources to supplement the reader’s learning. We will also include a short discussion on the benefits of keeping R up-to-date, learning to deal with functions that are no longer supported, and proactively writing code so it is easier to update in response to new versions.

**Network data (keeping, explain why)**

We appreciate the reviewers’ comments about the utility of network data for a range of educational data analysis stakeholders. In response to this comment, we will make a few changes. First, we will emphasize that, while we focus on network data from social media sources (in part due to interest in such data from both researchers and data analysts in schools, districts, and in other organizations), network analysis has been used to analyze, for example, how in-school advice-seeking networks can support the diffusion of innovative teaching practices (e.g., Coburn et al., 2014) or teachers’ use of new technologies (e.g., Frank et al., 2014). Second, we will add another walkthrough on a topic suggested by the reviewer, on longitudinal analysis of survey data (see the next point).

**Potential New Walk-Through on Longitudinal or SurveyMethods**

* Defining the term: does it mean preK - college?
* Survey data?
* What parts of the method are generalizable?
* What questions might an administrator or teacher be interested in? \* could be different than researchers’ goals
* Survey data used to get a perception of baseline in the field -how are they changing over time
  + Also can look at mediation and moderation
* People being concerned about their surveys not being used or if culture is not transparent about how data is being used
  + Can we offer the tools/techniques to help people make sense of survey data?
  + What happens when these techniques touch real people?

In the context of the reviewer’s concern about the utility of the walkthrough on network data (and on the sample size for the walkthrough on online science learners’ motivation), we would like to propose a new walkthrough on longitudinal survey methods. This walkthrough would be of interest to researchers, administrators, and data analysts, as processes of *change over time* are relevant to the data science-related work of individuals from all of these groups. In addition, looking at longitudinal data will allow us to start a conversation around the value of such data (relative to data collected at a single time point), as we can demonstrate how mediation and moderation can be evaluated.

**Teachers work with data as a part of our audience**

* Data analysis is possibly not a great concern
  + more priority on individual action w/ students, might not be aggregating data over time
  + Lots of data in schools, there is an interest in solving problems quantitatively
  + But teachers need to know the next actionable step for the particular student struggling
* In their day to day work, not their main concern
* Different goals
* Broaden the imagination about the questions they could be asking
  + They are thinking within the constraints of available tools
* Teachers are a part of a culture of data use
* Make sure what you do is relevant to students
* Include teachers in your design process

We will change our goals for the K-12 teacher audience from learning how to use R for data science to learning how to participate in the analytic process. This will be a shorter section relative to those aimed at researchers and education consultants, but can serve as an entry point to data science in education for teachers who would like to use a more empirical approach in their practice. This discussion can include best practices for asking productive data science questions, encouraging an expectation of large scale data analysis to answer practical questions, and the importance of including teachers in the analytic design process.

**Minor comments/issues**

* Have a consistent voice throughout (and have a style guide)?

In response to this helpful comment, we will use a common structure (i.e., have the same headings for each of the walkthroughs) as well as a style guide for both the writing (i.e., how we refer to ‘data’ in its singular or plural form) as well as the code (i.e., what conventions we follow for spacing and comments).

* Comment about the value of a dataset with > 1,000 cases -
  + The only science students data set has approximately 700 cases
  + The large-scale assessment walkthrough and the proposed longitudinal walkthrough will each more > 1,000 cases

In response to this comment, we will include a walkthrough on longitudinal survey methods (see above). The online science learners’ motivation data set has approximately 700 cases.

------- ------- ------- ------- ------- ------- ------- ------- ------- ------- ------- ------- -------

HANNAH’S EMAIL

Dear Josh and all

I hope you are all well. I am very pleased to attach three very positive reviews of the book – reviewer 3 also provided some comments in an annotated proposal. I’ve put a quick summary below of main points:

Review 1: very positive. One recommendation only: ‘to flesh out the early chapters about how educators in particular can benefit from data science.’ Notes the possibility of redundancy in the print version is R syntax changes; this is something we can discuss but we would hope sales would be sufficient to support new / updated editions. Also other options like the possibility of uploading updated syntax to an e-resource on the webpage for the book

Review 2: similar note on updating re tidyverse package. Again, let’s discuss – plus suggestion of using base R functions. Some thoughts on Walkthroughs and providing some more detail on those (though probably not necessary to provide full example at this early stage). Combine chapters 14 and 15. And final comment – ensure consistency

of writing and organization throughout given multiple authorship.

Review 3: some comments on market. Not sure I agree with the assertion about social media network data, question 7 – this is an important area. Some specifics, question 15 (and further comments in annotated proposal)

Next steps: we should discuss the software redundancy issues and the market – I think per Reviewer 3’s suggestion, references to K12 teachers may need to be removed as a market sector, but we can discuss this. Josh, we can do this by Skype, or email, or in the review response, whichever feels most comfortable for you. On the latter – the next stage in the process is for you to provide me with a written response to the reviews, outlining what you will change in light of review comments. I will then put everything together for our weekly editorial committee meeting – once I have your response in hand it should take 7-10 days to go to the meeting. All being well I would anticipate offering you a contract shortly after (including provision for hosting the e-version of the book on bookdown as we had discussed).

Thanks all. I am very much looking forward to hearing your thoughts on the next steps, and to working with you further. Great to see such positive feedback on the book!

All best wishes

Hannah

Hannah Shakespeare

Senior Commissioning Editor, Research Methods