# Creation of Learner Personas for a Biomedical Data Science Curriculum

Daniel Y Chen[1,2] and Anne M. Brown[1,2,3]
[1] Research and Informatics, University Libraries, Virginia Tech, Blacksburg, VA
[2] Interdisciplinary Program in Genetics, Bioinformatics, and Computational Biology (GBCB), Virginia Tech, Blacksburg, VA
[3] Department of Biochemistry, Virginia Tech, Blacksburg, VA

## INTRODUCTION

- Learner's prior knowledge and organizational structure in their **mental model** will affect the way they learn
- We created a **self-assessment survey** to create **learner personas** to discern both current knowledge and knowledge gaps/needs
- **Learner personas** are fictional characters that represent key characteristics of a particular learner: **background**, **prior knowledge or experience**, **perception of needs**, and **special considerations**
- The accelerating **changes in medical education** incorporating data science competencies indicate the growing need of data science education in the biomedical sciences
- A **backward design approach** using learner personas for creating lessons kept content creation and teaching focused and maintain alignment with the overall **learning objectives**

## Methods

- 31-question learner **self-assessment survey**
  - Questions on: prior programming, statistics, and data knowledge
- **Hierarchical clustering** using scaled euclidean distance and Ward's clustering criterion to cluster observations
- **Elbow method**, **gap statistic**, **and interpretability** were used to pick the number of clusters (i.e., personas)
- **PCA** (principal component analysis) and **EFA** (exploratory factor analysis) using promax rotation used to validate the survey and simplify the original survey

## Demographics and Clusters
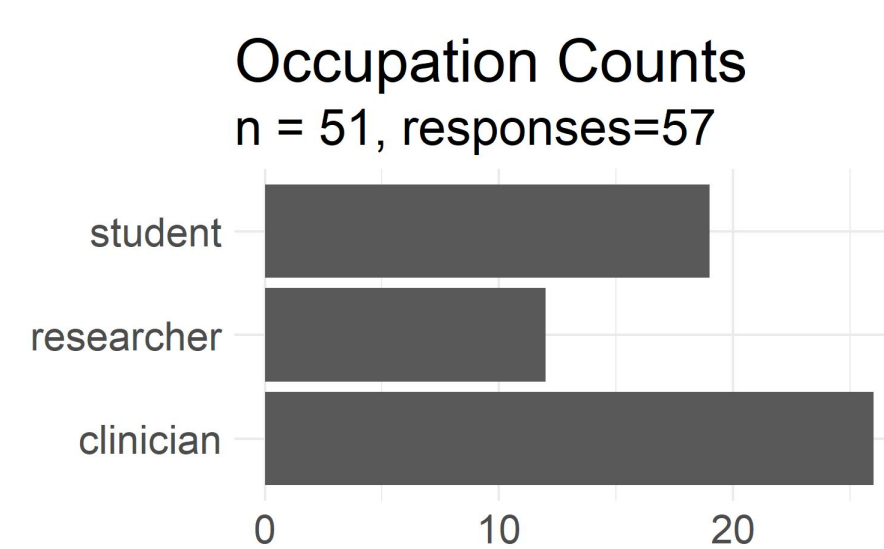
- 57 respondents, 51 consented
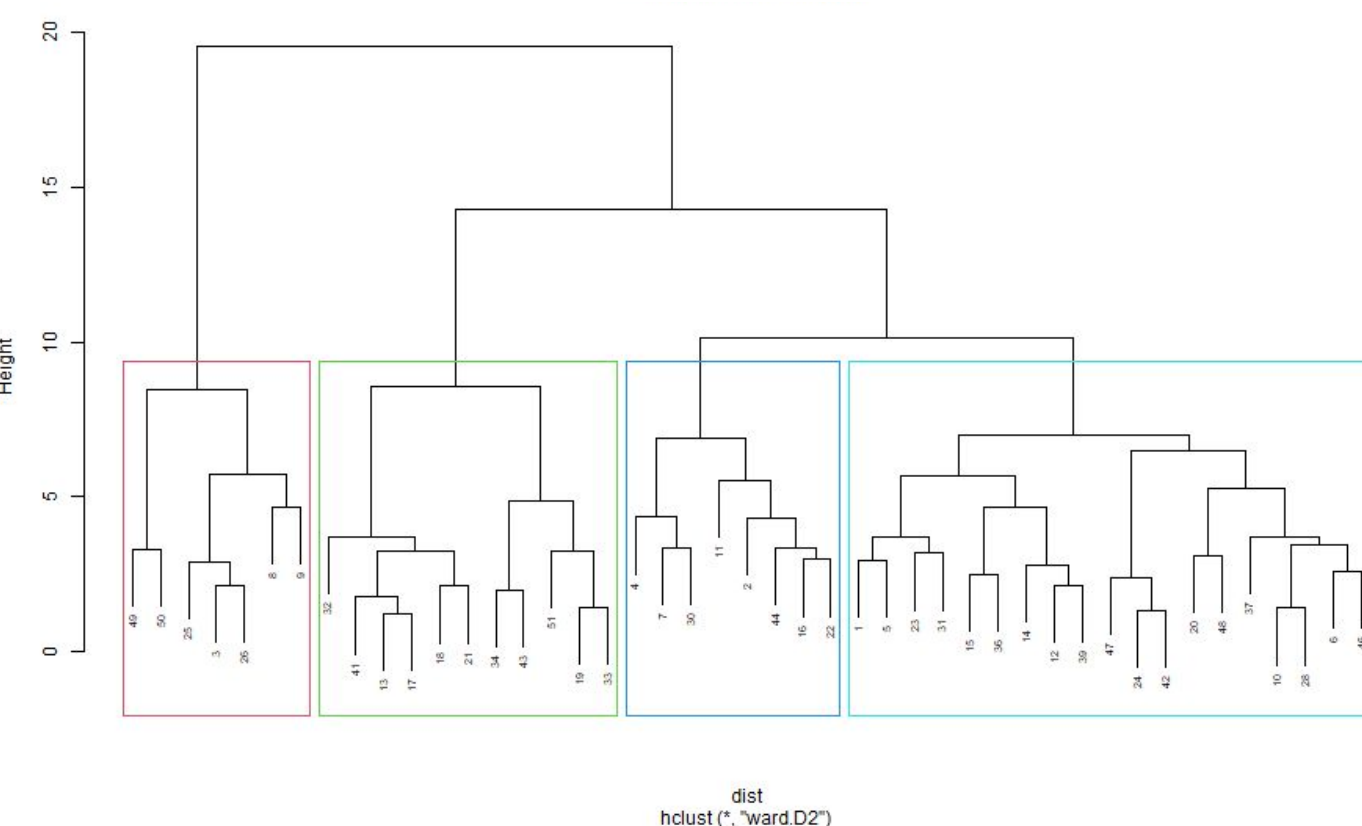- 45 responses were used for the clustering due to missing responses



**Figure 1.** Survey demographic information and clustering. **Top** - Counts of each occupation group. Respondents were able to select more than one occupation. Groups shown were aggregated from original choices **Right** - Dendrogram clusters. Left to right: programmers, clinicians, students, academics
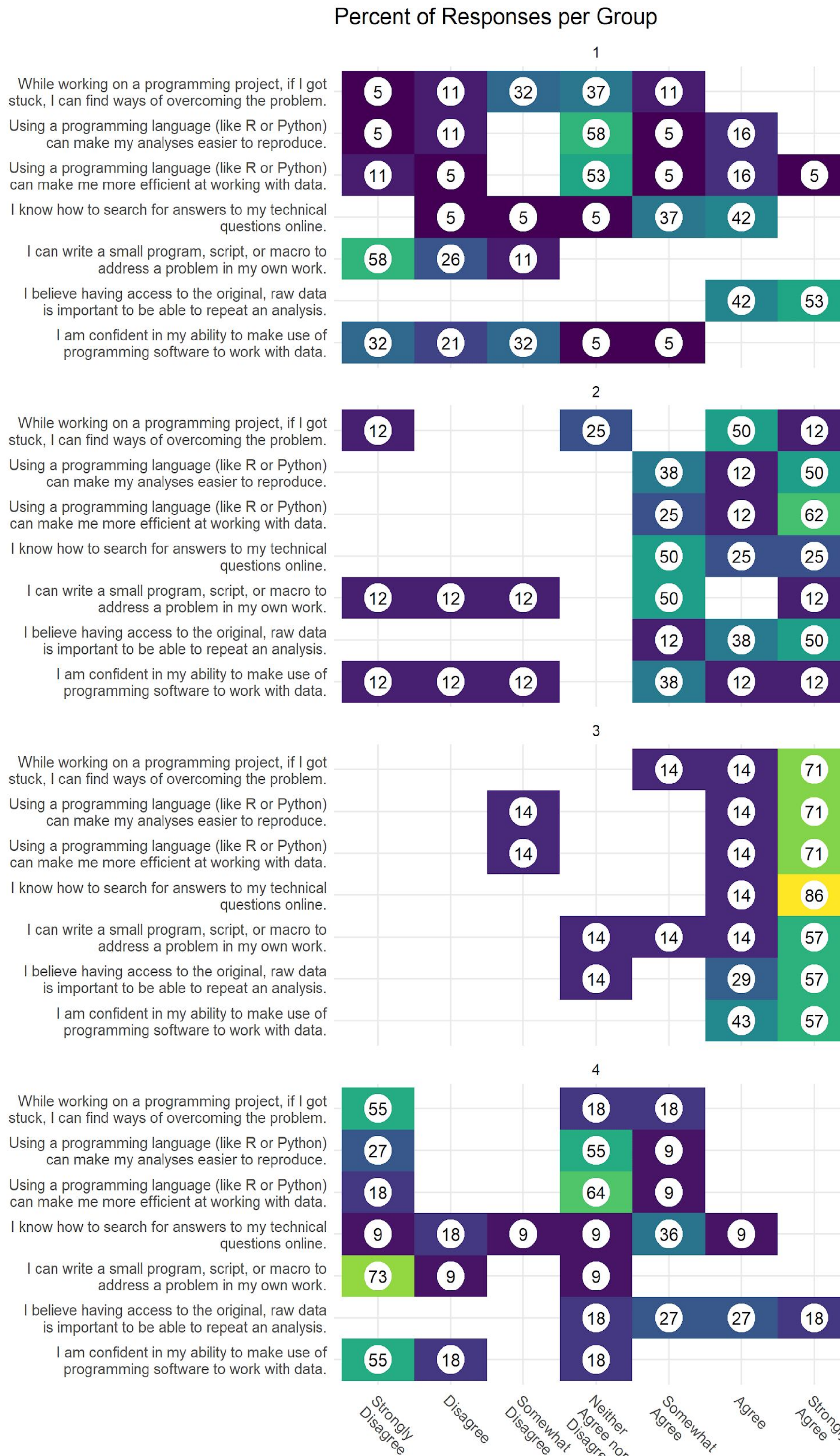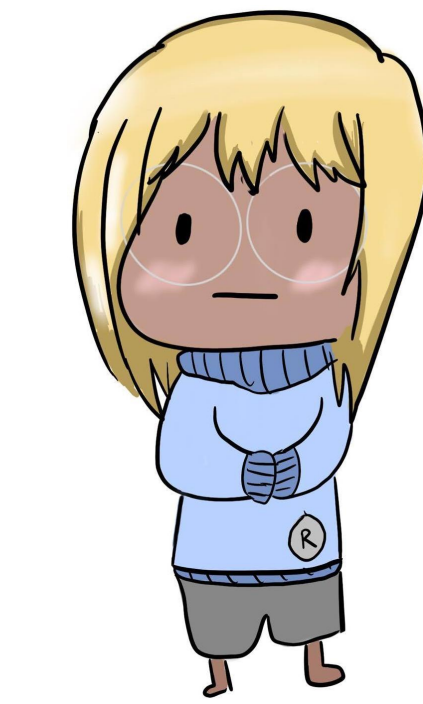
## Personas



**Figure 2.** Responses to likert questions in the learner self-assessment survey by each cluster (i.e., persona). Colors and values show the percent of respondents in each cluster and their agreement with each statement.

Alex Academic

Samir Student

Patricia Programmer

Clare Clinician

## Results

- **Simplified survey** down to 3 questions using the highest EFA factor loading for 3 factors, one for each question domain:
  - **Data**: Do you know what "long" and "wide" data are?
  - **Programming**: How familiar are you with interactive programming languages like Python or R?
  - **Statistics**: If you were given a dataset containing an individual's smoking status (binary variable) and whether or not they have hypertension (binary variable), would you know how to conduct a statistical analysis to see if smoking has an increased relative risk or odds of hypertension? Any type of model will suffice.
- These **personas** were used to create **lesson materials** to teach data science skills to people who work in the medical and biomedical sciences: https://ds4biomed.tech/
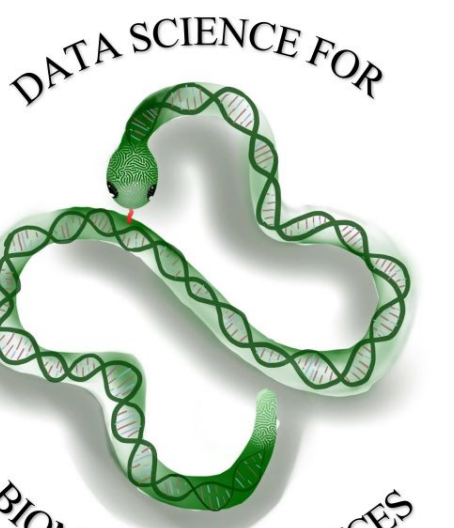
## Conclusion

- Creation of **learner personas** that resonated with workshop attendees
- **Structured learning material** creation based on learner personas will aid in information content and knowledge retention
- The **learner self-assessment survey** and **clustering methodology** can be adapted to other disciplines

## References

Ambrose, Susan A, Michael W Bridges, Michele DiPietro, Marsha C Lovett, and Marie K Norman. *How Learning Works: Seven Research-Based Principles for Smart Teaching.* John Wiley & Sons, 2010.
Jordan, Kari, François Michonneau, and Belinda Weaver. "Analysis of Software and Data Carpentry's Pre- and Post-Workshop Surveys." Zenodo, July 17, 2018. https://doi.org/10.5281/zenodo.1325464.
RStudio, Education Team. "RStudio Learner Personas," 2019. https://rstudio-education.github.io/learner-personas/.
Wilson, Greg. *Teaching Tech Together: How to Make Your Lessons Work and Build a Teaching Community around Them.* CRC Press, 2019.
Zagallo, Patricia, Jill McCourt, Robert Idsardi, Michelle K Smith, Mark Urban-Lurain, Tessa C Andrews, Kevin Haudek, et al. "Through the Eyes of Faculty: Using Personas as a Tool for Learner-Centered Professional Development." *CBE—Life Sciences Education* 18, no. 4 (2019): ar62.

Poster: https://github.com/chendaniely/dissertation-presentations
Twitter: @chendaniely @biochem_anne