

Data Fest 2024

How Best Can We Support Underperforming Students?

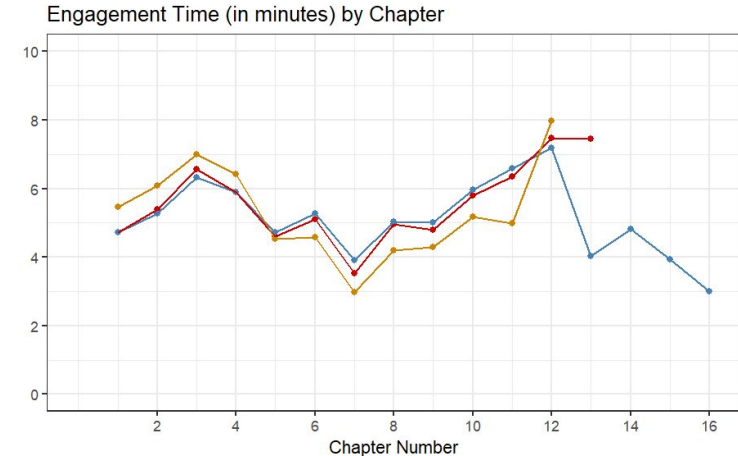
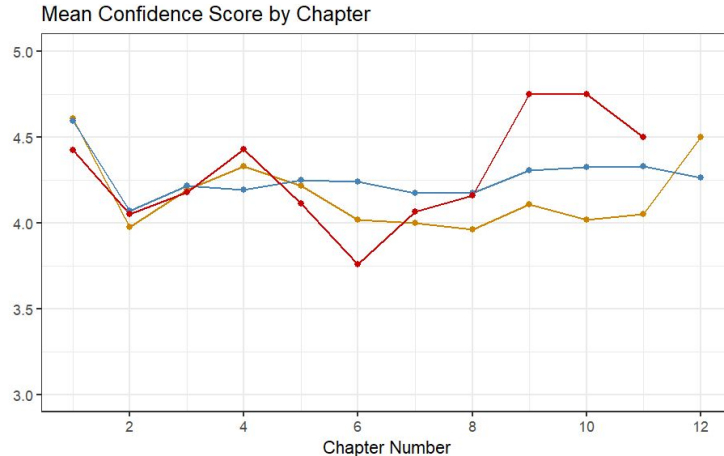
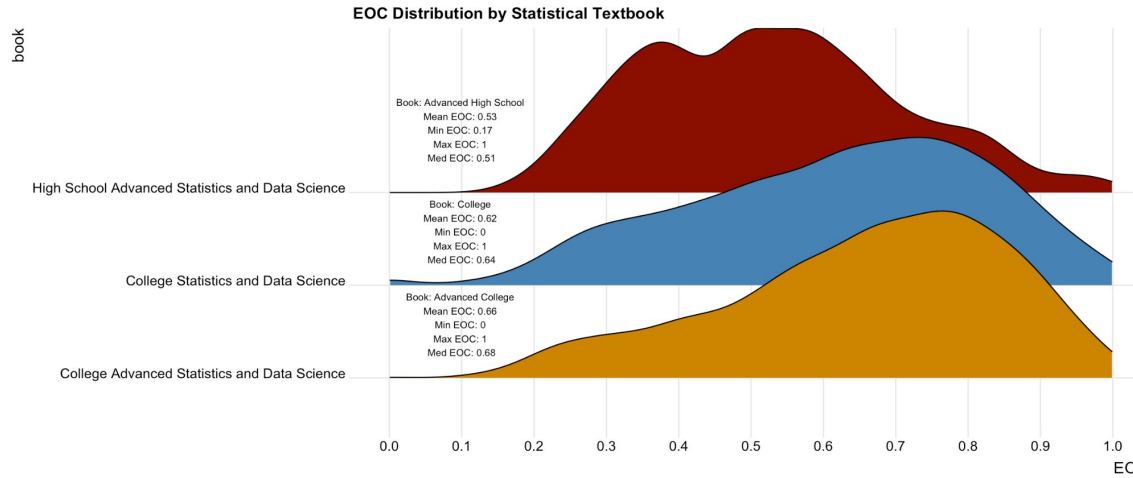


Ella Smith, Colin Hassett, Matthew Huang, Justin Mai

Exploratory Data Analysis

Statistical analysis:
Significant
Tukey: confirmed that
chapters were different
in EOC; when visualizing,
EOC was lower in later
chapters.

P value: Confirmed had
significant evidence of
EOC
Insignificant evidence
that amount of videos
impacts engagement..



Filtering Signal from Noise

Elia's scores
by chapter

[64 50 70 10 5 2 0 ...]

Wants to improve scores for ch 4,5

Student Database

John1
Sama2
⋮

ch1	...	ch4	ch5	...	Performance		Similarity	Relevancy
57		<u>58</u>	<u>62</u>		100		80%	<u>80</u>
51		<u>61</u>	<u>63</u>		112		50%	56
⋮		⋮	⋮		⋮		⋮	

*Note: The specific performance metric we used also takes into account # of attempts:
 $\text{score} * \text{attempts}^{(-1/2)}$

*Note: - The similarity metric considers more than just score data, including book, class, and sentiment
- The specific measure we used is cosine similarity, but others such as correlation or euclidean distance can be used depending on performance considerations

* C1, C2 are weights to scale depending on importance

Performance^{C1} * Similarity^{C2}

Relevancy

Finding Recommendations

Top n Relevant Students:

▲	ch4_attempts	ch5_attempts	ch4_correct	ch5_correct	performance2	cosine_sim	weighted_performance2
1	182	84	91	49	5.272705	0.7670921	3.102619
2	189	86	99	55	4.696956	0.7841144	2.887855
3	185	96	101	52	4.500000	0.8062652	2.925286
4	194	86	86	51	4.769728	0.7118950	2.417272
5	167	84	95	45	5.167494	0.6599228	2.250434

Example Insight:

diff_in_attempts	Named num [1:5]	-50.35	23.61	33.63	84.9	-8.44
------------------	-----------------	--------	-------	-------	------	-------

By finding average **differences** in **attempts** by chapter, we might recommend that Ella study **less** for chapter 1, **more** for 2 and 3, and **much more** for 4.

Other Insights to look for:

- Differences in **sentiment**: Ella's low perception of a chapter's usefulness could prompt an instructor to highlight answering "why should you care?"
- Differences in **engagement time**