The most difficult school subject (according to Jeopardy)

ERHS 535

Jacob Fontenot, Burton Karger, Eric Wendt, Caroline Wendt

12/13/2019

The final report should not exceed 1,500 words. You should aim for no more than three figures and tables.

Part I

Research question

What are the most important and most difficult school subjects according to Jeopardy!?

Introduction

We initially sorted the Jeopardy! data set according to the most commonly asked questions in the history of the show. We found that "What is Australia?" was the most commonly occurring question in the complete Jeopardy! data set. The next 32 most commonly occurring questions were also geography related, indicating knowledge of geography is critical to success in Jeopardy!. This prompted us to consider how a jeopardy-based education would differ from a public school education. Here we investigated if school subjects deemed by us (and students more generally) to be the most important and the most difficult, are also the most important and difficult according to the history of Jeopardy!.

Methods

To evaluate the most difficult and important school subjects according to Jeopardy!, we created a dataframe of school subjects with lists of terms related to those subjects. We obtained these vocab terms from a commonly used online learning platform that provides study tools for various school subjects. We chose to divide the Jeopardy! data set according to science, mathematics, history, english, geography, and art related questions. We matched vocab terms from our vocab dataframe with questions in the Jeopardy! dataframe in order to filter the large Jeopardy! dataframe into a dataframe limited to Jeopardy! questions that matched matched our vocab terms.

To facilitate joining the dataframes, we converted all Jeopardy! questions and school vocab terms into words exclusively limited to lowercase letters with all special characters removed. To account for differences in the number of vocab terms in each subject, before joining the data frames, we randomly sampled 155 vocab terms from each subject. We created a script to match the "question" from the Jeopardy! dataframe to the corresponding key term in the subject data frame. The script expects the standard Jeopardy! dataframe obtained from the Jeopardy! archive as well as a key terms dataframe for a given subject. There should be a subject column with the subject (e.g., `science`) listed for each row next to each vocab term. This column is what is later used to group the Jeopardy! questions by subject. We used `inner_join` from the `dplyr` package to perform matching on the Jeopardy! questions.

We discovered that the structure of the game show changed with regard to round and value after November 26, 2001. That is, values in round one and two doubled their previous values after this date. For instance, the first round value originally ranged from 100 to 500; however, the first round value was doubled to 200 dollars to 1,000 dollars after November 26, 2001. Therefore, we separated the data into two categories to delinate which observations adhered to the particular round-value frameworks before and after this date.

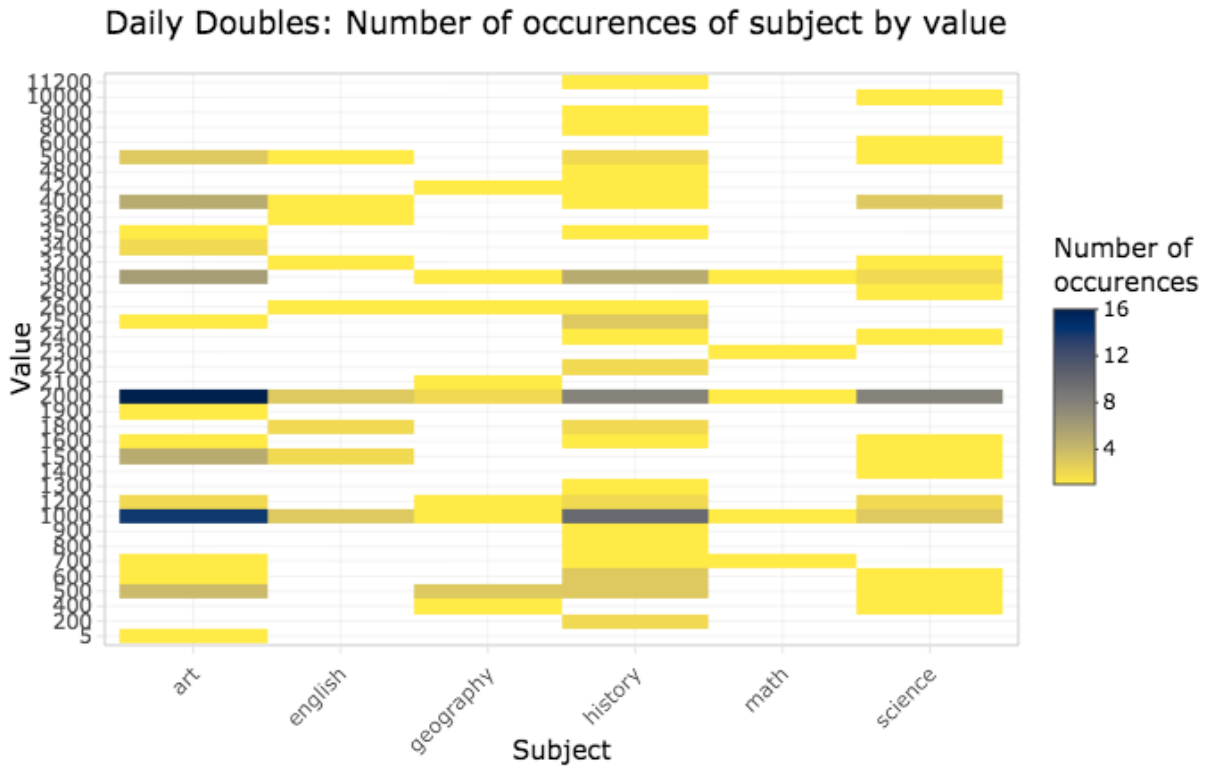
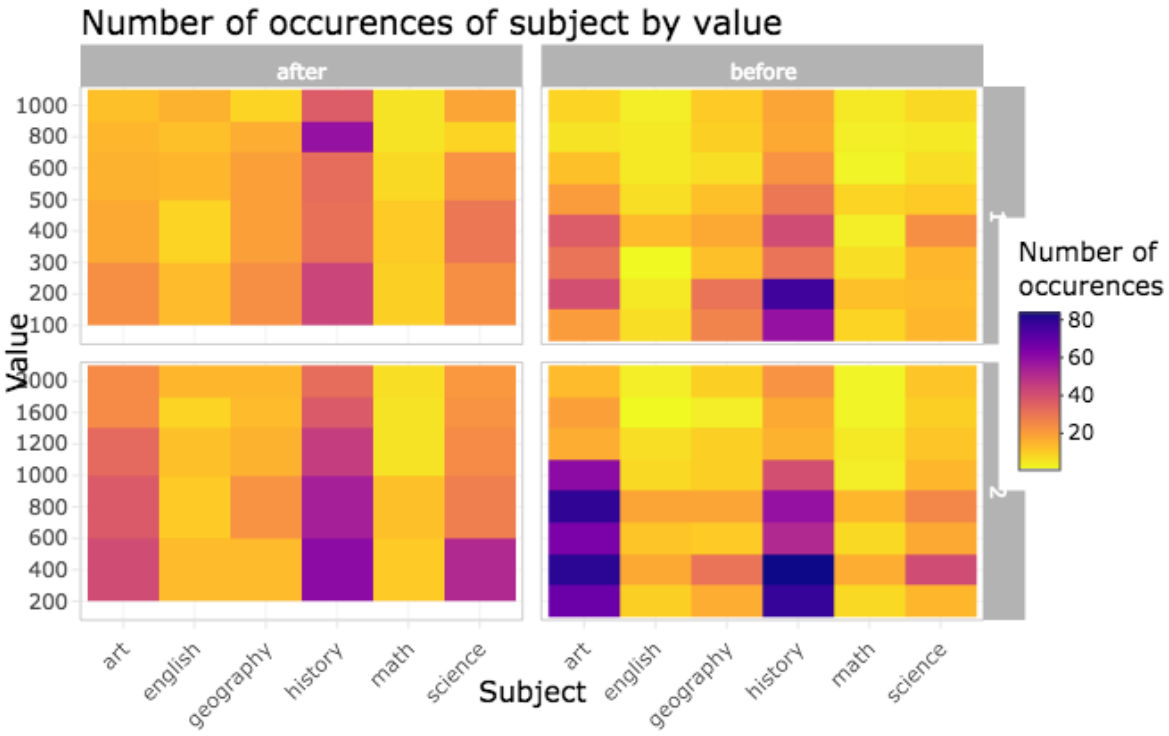
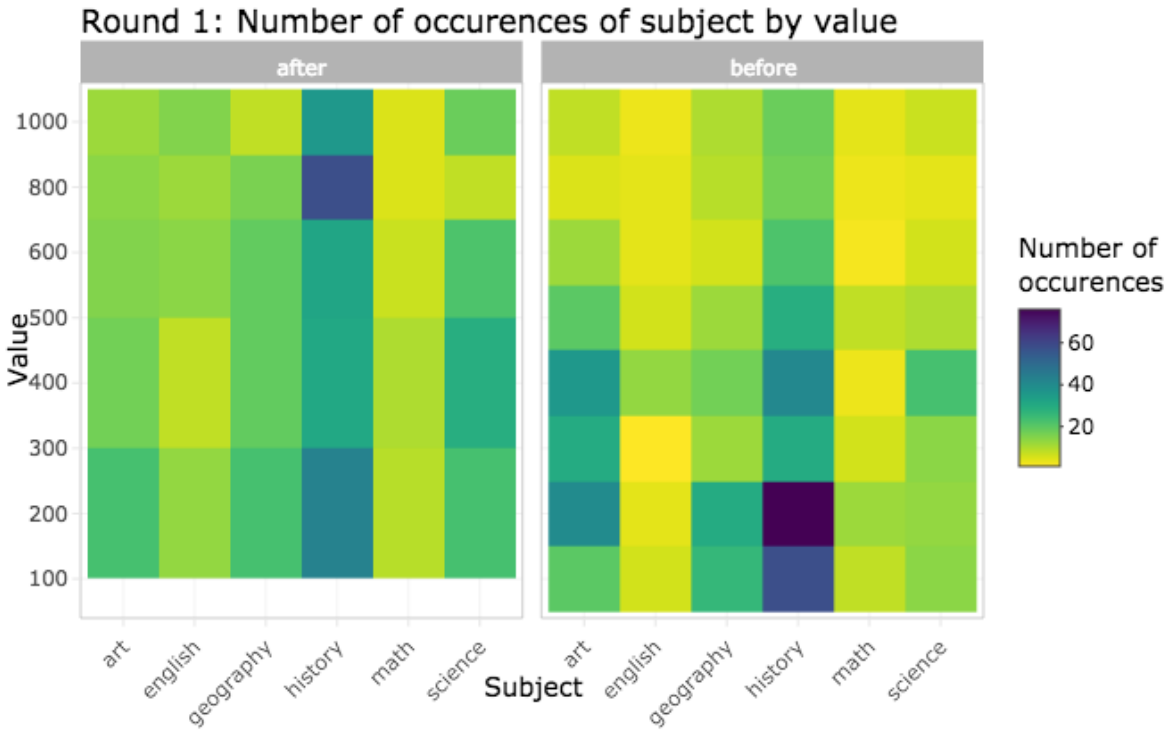


Figure 1: "Heat Map Daily Doubles"

Further, with the joined dataframe in place, we calculated the number of occurrences of each school subject as well as the average value of clues containing each subject for each of the two date ranges. We defined the most important school subject according to Jeopardy! as the topic with the most occurrences, and the most difficult as the subject with the highest average monetary value. We graphed these results using an interactive heat map facetd by both round (e.g., 1 or 2) and date (e.g., before or after November 26, 2001).

Results





What did you find out? Most of these slides should be figures or tables. Discuss your interpretation of your results as you present them. Ideally, you should be able to show your main results in about 3 slides, with one

figure or table per slide.

include plots and tables here see `plots_caroline_final.Rmd`

most important subject most difficult subject

Conclusions

So what? How do your results compare with what other people have found out about your research question? Based on what you found, are there now other things you want to check out?

Part II: Tutorial - For Slides?

Overview of your approach in R: Step us through a condensed version of how you did your project

Interesting packages/techniques

Spend a bit more time on any parts that you found particularly interesting or exciting. Were there packages you used that were helpful that we haven't talked about in class? Did you find out how to do anything that you think other students could use in the future? Did you end up writing a lot of functions to use? Did you have an interesting way of sharing code and data among your group members?

Lessons learned

If you were to do this project again from scratch, what would you do differently? Were there any big wrong turns along the way? Did you find out how to do something late in the project that would have saved you time if you'd started using it earlier?

References

Quizlet Inc. (2019). Quizlet. Retrieved from https://quizlet.com/