The most difficult school subject (according to Jeopardy)

ERHS 535

Jacob Fontenot, Burton Karger, Eric Wendt, Caroline Wendt

12/13/2019

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.

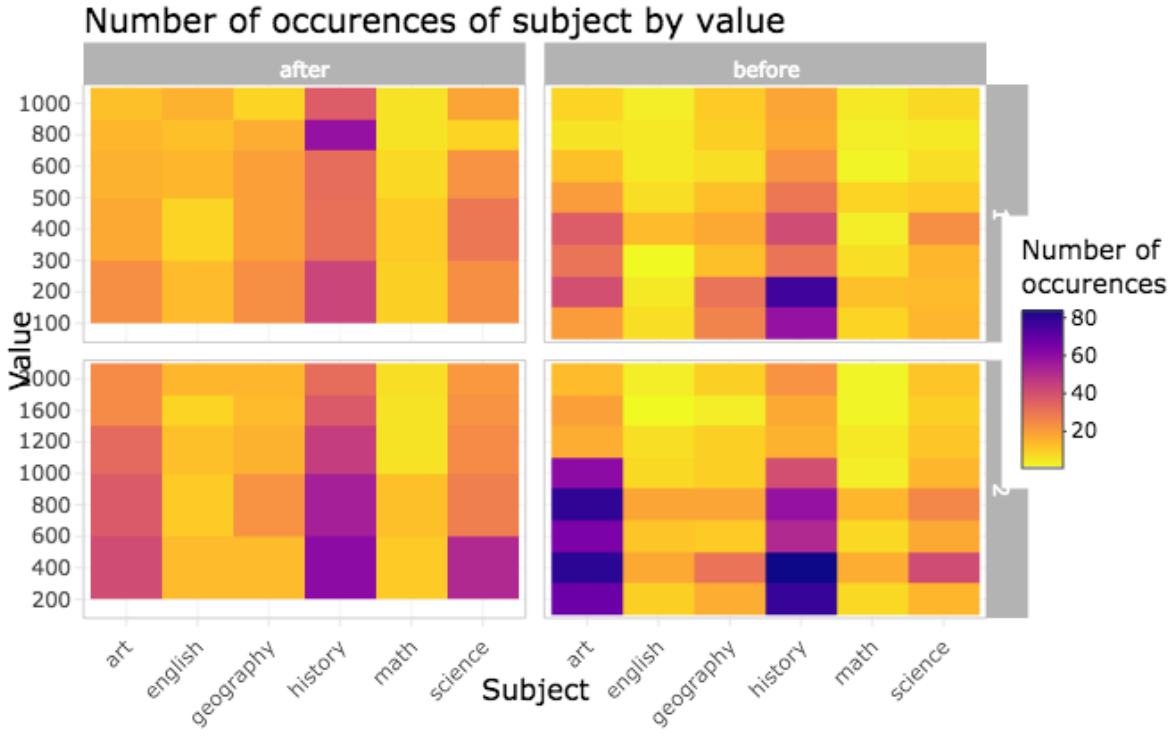
To facilitate joining the dataframes, we converted all Jeopardy! questions and school vocab terms into words exclusively limited to lowercase letters with all special characters removed. To account for differences in the number of vocab terms in each subject, before joining the data frames, we randomly 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. 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

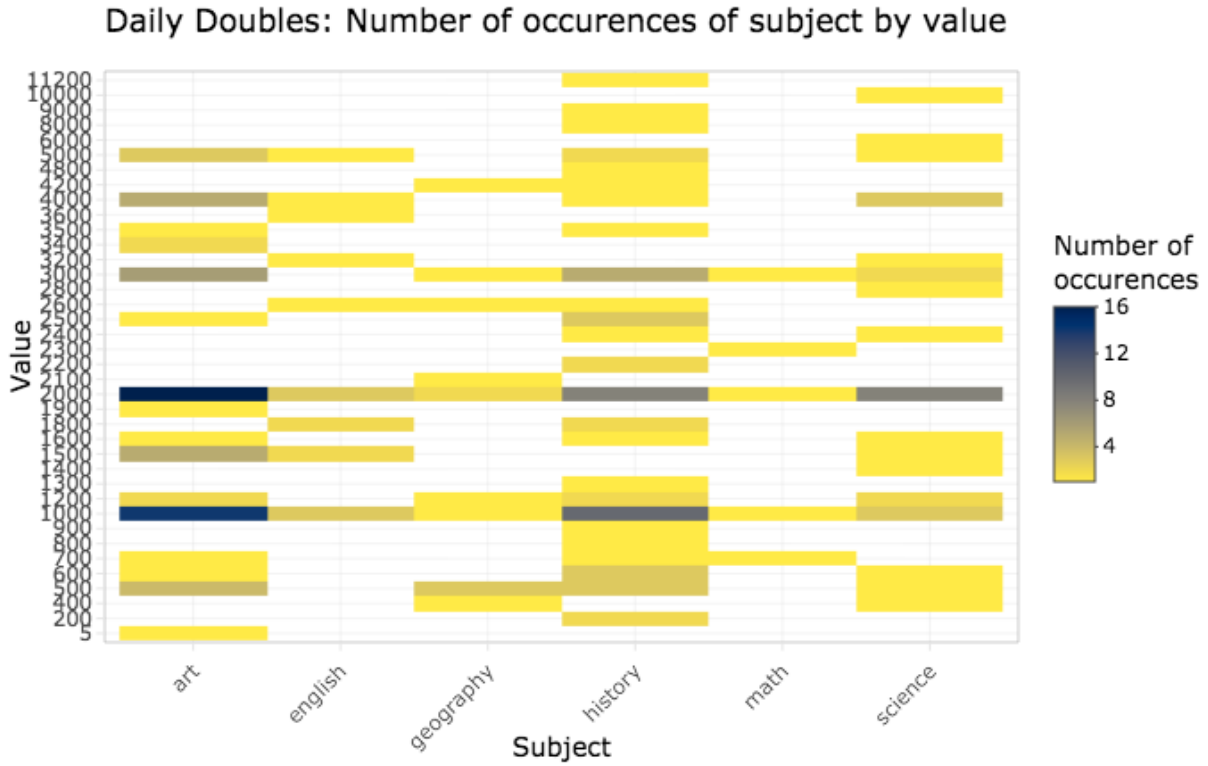
most difficult as the subject with the highest average monetary value. We graphed these results using an interactive heat map facetd by both round (e.g., 1 or 2) and date (e.g., before or after November 26, 2001).

Results

This heat map showing each value for school subjects occurences in regular jeopardy questsions. We see clearly that art and history are the most populous categories and having a fairly high value associated with each question. It is important to note here that the data is also filtered by "after" and "before", meaning before or after November 26, 2001 when the values of the questions doubled.



This heat map here depicts number of occurences of daily doubles per school subject selected against value that was bet per subject. We see that History, science, and art have the most occurences and the highest bets go history, science, art, with the maximum being $11200!!!



	Before November 26, 2001	
subject	mean_value	subject_total
history	679.62	471
art	782.87	397
science	746.59	249
geography	563.98	236
english	645.60	125
math	574.29	105

Kable talbes showing our school subjects of interest ordered by descending values for the total occurences of the subject (subject_total). They are separated by before or after November 26, 2001 when the values were doubled for Jeopardy!
In our before table, highlighted in blue we see that History has the largest amount of occurences, and second is Art followed by Science. However, highlighted in red we see Art has the highest mean_value, meaning when Art was the questions/answer it's was typically valued higher, followed by Science and History.

	After November 26, 2001	
subject	mean_value	subject_total
history	965.05	309
science	925.00	280
art	1034.51	226
geography	842.44	172
english	1007.35	136
math	803.85	78

Figure 1: ""

In the kable table "After November 26, 2001" we see that History and Art still maintain their 1st place rankings for total and mean_value, respectively. However, Science has overtaken Art for second most occurred, and English has overtaken Science for second highest valued!

Conclusions

According to Jeopardy! before and after November 26,2001.

The most important subject is: History!

While the most difficult subject is: Art!

We also compared our results to AP test score stastistics. We looked at "pass rate" and grouped similar subjects together to create a list containing our school subjects ordered by greatest to least difficulty. The list is as follows: Geography, English, History, Science, Math, and Art. From the summary stastics from before 11/26/2001: Art, Science, History, English, Math, Geography. From after 11/26/2001: Art, English, History, Science, Geography, and Math. The list made from the AP test scores doesn't agree much with the statistics from before 11/26/2001 but does with those from after. English, History, and Science represent the 2nd, 3rd, and 4th most difficult subjects in the AP list and the dataset after 11/26/2001, and math is predicted to be easier in both. Since data from after 11/26/2001 is more current, we place more weight on it's agreement with the AP scores, and conclude that a Jeopardy-based education would not differ greatly from a public school education in terms of difficulty.

References

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