

The most difficult school subject
(according to Jeopardy)
ERHS 535

Jacob Fontenot, Burton Karger, Eric Wendt, Caroline Wendt

12/13/2019

Overview

Research question

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

Introduction

- ▶ Initial data investigation showed subject bias
- ▶ Prompted us to consider traditional school subjects
- ▶ Decided to compare “Jeopardy-based” education to public school education

Methods: Overview

- ▶ Created a dataframe of school subjects and related terms (from Quizlet)
- ▶ Separated Jeopardy dataset into 6 subjects (science, math, history, english, geography, and art) by “joining” it with the Quizlet dataframe

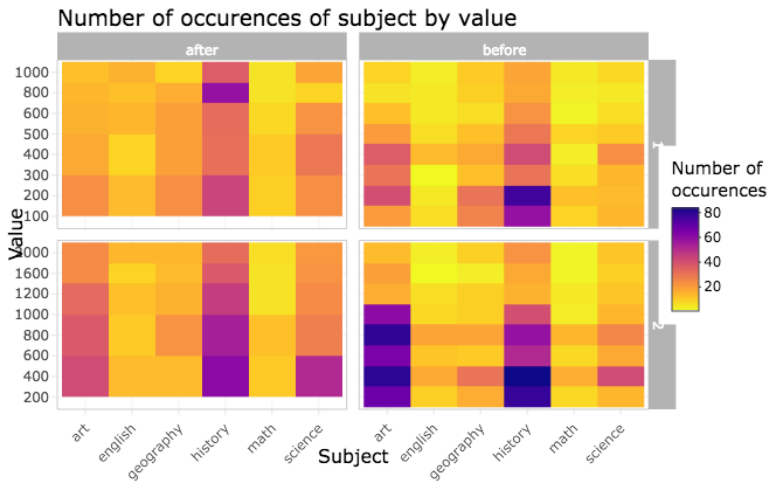
Methods: Joining dataframes

- ▶ Converted all Jeopardy data and school subject data to lowercase and removed all special characters
- ▶ Randomly sampled 155 vocab terms from school subject data
- ▶ Used `inner_join` from the `dplyr` package to match a the “question” to a specific vocab term
- ▶ Grouped the “questions” by subject

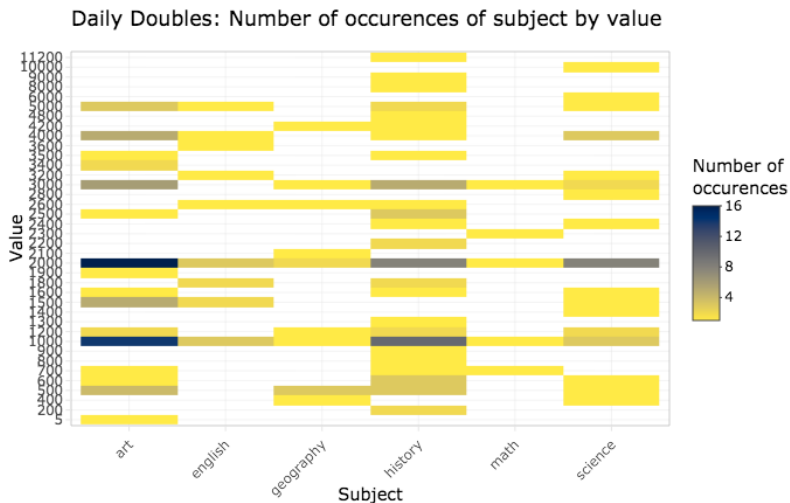
Methods: Determining subject importance and difficulty using joined dataframe

- ▶ Discovered answer value changes after November 26, 2001 and decided to separate data into two categories
- ▶ Calculated number of occurrences of each school subject (most important subject)
- ▶ Calculated average value of all “answers” for each school subject (most difficult subject)
- ▶ Graphed results using interactive heat map faceted by round and date

Results: Heat map for round 1 and 2



Results: Merging round 1 and 2 heat maps



The most difficult school subject (according to Jeopardy)

Results: Summary Statistics before 11/26/2001

subject	Before November 26, 2001	
	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

Results: Summary statistics after 11/26/2001

subject	After November 26, 2001	
	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

Conclusion

- ▶ AP test scores: Geo, English, History, Science, Math, Art
- ▶ Round 1: Art, Science, History, English, Math, Geo
- ▶ Round 2: Art, English, History, Science, Geo, Math

Tutorial

Overview: Approach in R

- ▶ The data for the subjects which, we would use to filter down the Jeopardy!, questions needed to be created and cleaned
- ▶ Initial Jeopardy! and subject data cleaning steps were taken to make the data more uniform (lowercase columns and values) removing special characters and numbers to facilitate our joining.
- ▶ The Jeopardy! and subject terms data were then joined, using `inner_join`, by the “question” column to only keep the rows which are in are in Jeopardy! and subjects terms data frames, and `inner_join` returns all columns.

Overview: Approach in R

- ▶ For the creation of the subject data we had to utilize the `sample_n()` function from `dplyr` since we had more words for some subjects than others, and we chose to randomly select 155 words from each subject.
- ▶ Following our join of the two data frames we had to create one more column to depict the change in value of the question/answers that occurred after November 26, 2001.

Packages

- ▶ packages for data filtering
 - ▶ library(readr)
 - ▶ library(dplyr)
 - ▶ library(stringr)
- ▶ package for tables
 - ▶ library(kableExtra)
- ▶ packages for plots
 - ▶ library(ggplot2)
 - ▶ library(viridis)
 - ▶ library(forcats)
 - ▶ library(plotly)
 - ▶ library(ggthemes)
 - ▶ library(RColorBrewer)

Our Code

```
```{r}
jeopardy data
jeop_data <- read_tsv(
 file = "../data/jeopardy_clue_dataset-master/master_season1-35.tsv") %>%
 dplyr::mutate(question = stringr::str_to_lower(question)) %>%
 dplyr::mutate(question = stringr::str_remove_all(question,
 "[^:alnum:][:blank:]"))
```
```

Our Code

```
```{r}
subjects and terms data
sub_term_data <- read_csv("../data/subjects_and_terms.csv") %>%
 rename(vocab_term = Term,
 subject = Subject) %>%
 dplyr::mutate(vocab_term =
 stringr::str_to_lower(vocab_term)) %>%
 dplyr::mutate(vocab_term =
 stringr::str_remove_all(vocab_term,
 "[^[:alnum:]][:blank:]")) %>%
 dplyr::distinct() %>%
 dplyr::mutate(question = vocab_term) %>%
 dplyr::group_by(subject) %>%
 dplyr::sample_n(size = 155, replace = FALSE) %>%
 dplyr::ungroup()
```
```

Our Code

```
```{r}
jeop_school_joined <- jeop_school_joined %>%
 dplyr::mutate(air_date_type = if_else(air_date < "2006-11-26",
 "before", "after"))
```
```

Lessons learned

- ▶ .Rmd can be a curse and a blessing
 - ▶ Getting exactly what you think your .pdf should look like knitted isn't always the easist thing.
- ▶ GitHub is a wonderful tool for collaborative work
 - ▶ Be very sure to inform others when you have pushed large amounts of changes. . .

References