SMA Capstone

Jiaqi Sun

2023-02-08

Introduction

The client is hoping to analyze the relationship between all these measurement variables, and, if possible, find ways to summarize the scores to compare websites more easily.

Our client wants to validate these hypotheses: Affiliation with parent given information correlates with better content quality; If the webpage is written by a paid consulting company, the readability will be low, and the content is less specific.

This study will be a content and readability analysis of websites published by hospital systems describing spinal muscular atrophy treatment.

Data

Readability will be assessed using the Flesh-Kincaid Grade Level(FLESCH) and the Simple Measure of Gobbledygook (SMOG).
Content analysis will include two parts. The Likert-based 16-item (DISCERN) tool will be applied to each website to assess the quality of the information present; Sites will be analyzed for the presence of the following information (Content) which have been deemed important in informed decision making for SMA treatment options: description of SMA, list of treatment options currently available, dose timing, cost, insurance coverage, side effects, and risk factors. Websites will also be analyzed for the presence of FAQ pages and additional resources available. Themes identified by further analysis may also be included. Scoring will be done by the primary investigator following the guidelines provided by the tools listed and information will be store on a private computer in Excel.

These sections were transferred to Office WordTM software to conduct a calculation of the Flesch-Kincaid Grade Level reading assessment. The calculated value for the Flesch-Kincaid Grade Level corresponds to the grade level required in order to fully understand the text (Yeung et al., 2022). The recommendation from the American Medical Association and National Institutes of Health states that all patient information materials should be between the third and seventh grade level.

Additionally, the Simple Measure of Gobbledygook (SMOG) index was also utilized to analyze readability and was completed manually by the researcher (McLaughlin, 1969). The SMOG index assesses readability by examining polysyllabic words from a selection of sentences in a text. This formula results in a score which is equivalent to the US grade level of education required to understand the written text.

The DISCERN tool was applied to each of the pages to assess the quality of health information. This tool is a 16-item questionnaire with each question scored on a Likert scale of one to five ('very poor,' 'poor,' 'fair,' 'good,' or 'excellent'). A high score indicates the website has more valuable and appropriate content, while a lower score indicates a lack of information (Charnock et al., 1999).

Analysis of content was divided into four categories and twenty-six sub-categories.The four categories were: SMA description, treatment information, pricing, and resources.

If the webpage is written by a paid consulting company(3rd party), the readability will be low, and the content is less specific.

Content Quality by hospital type
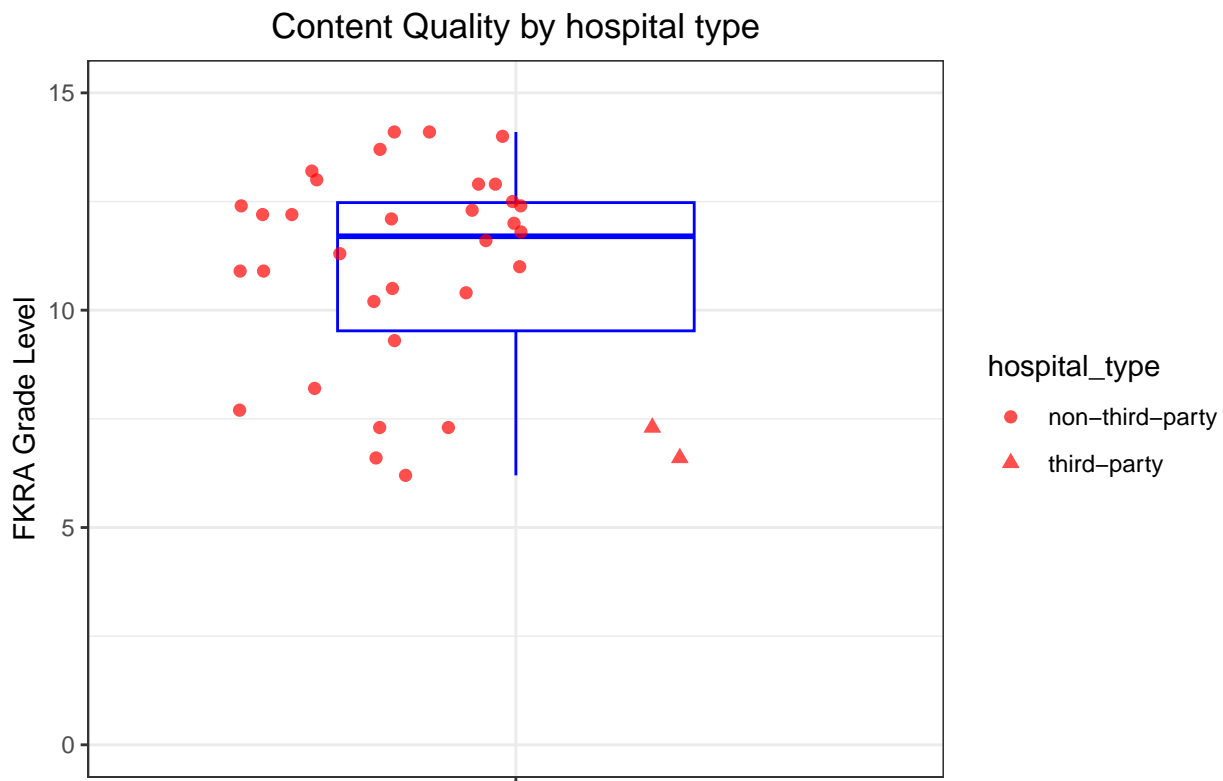
Visualize the distribution of FKRA grade levels for different types of hospitals

Our client wants to validate this hypothesis:
If the webpage is written by a third-party paid company, the readability will be low.

From this plot, we can see that in general, website written by third-party companies tend to require lower level of grade.

Checking the missingness of the data using naniar package in R.

Comparison of the two readability tests (FLESCH and SMOG), see if they're close to each other. –scatterplot
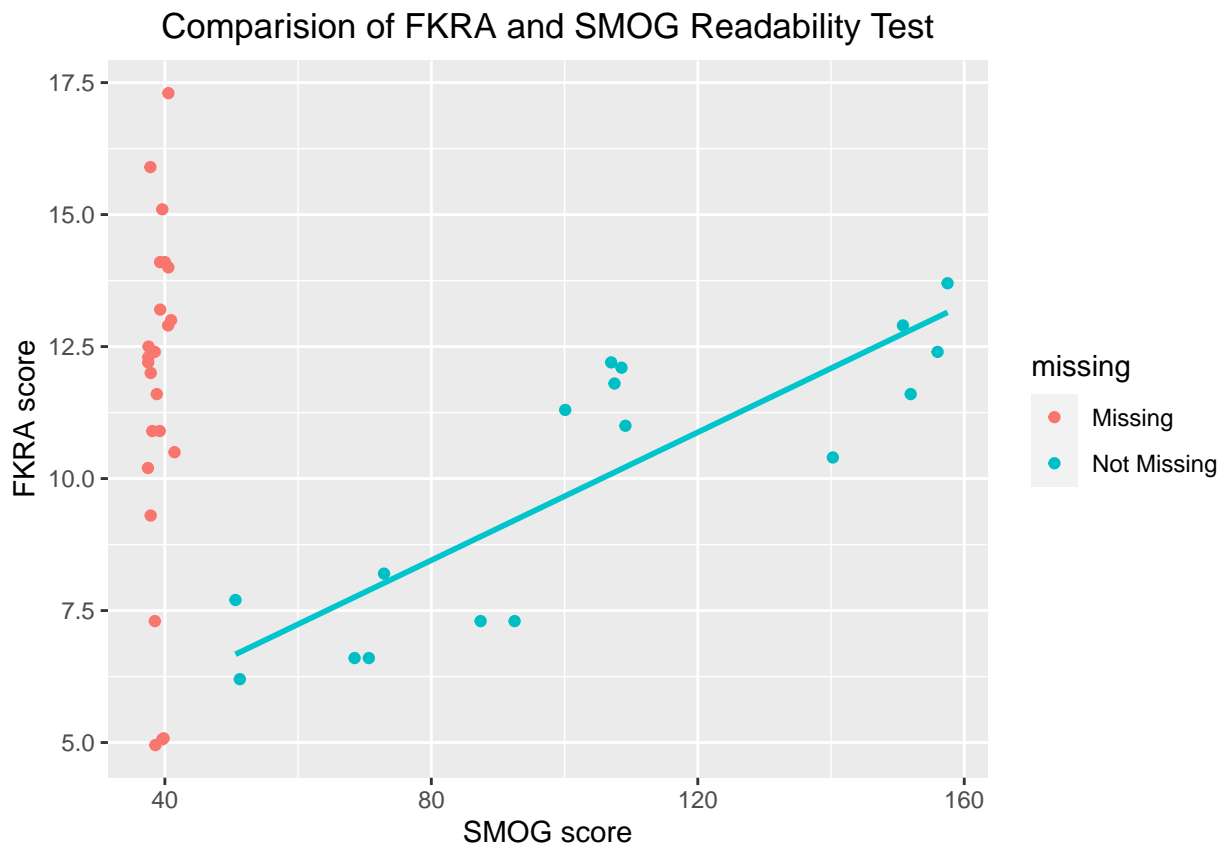
Comparison of the two readability tests (FLESCH and SMOG), see if they're close to each other.

**Comparison of FKRA and SMOG Readability Test**
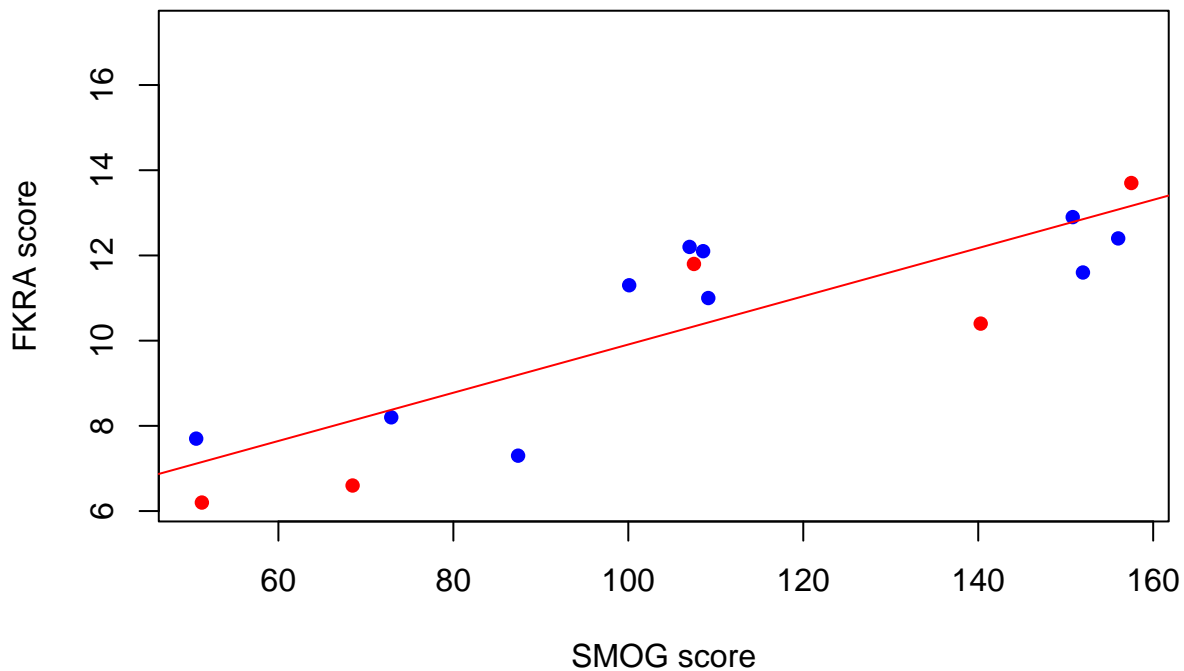
The two reading test are indeed close enough. We suggest use FKRA as it has less missing values.

Comparison of two readability tests (FLESCH VS. SMOG)

Comparision of FKRA and SMOG Readability Test

Positive correlation between SMOG and FKRA scores regression line with missing value shown.

Make boxplots to visualize the main 4 categories of CONTENT, make bivariate 1s and 0s plots. (facet by main category)
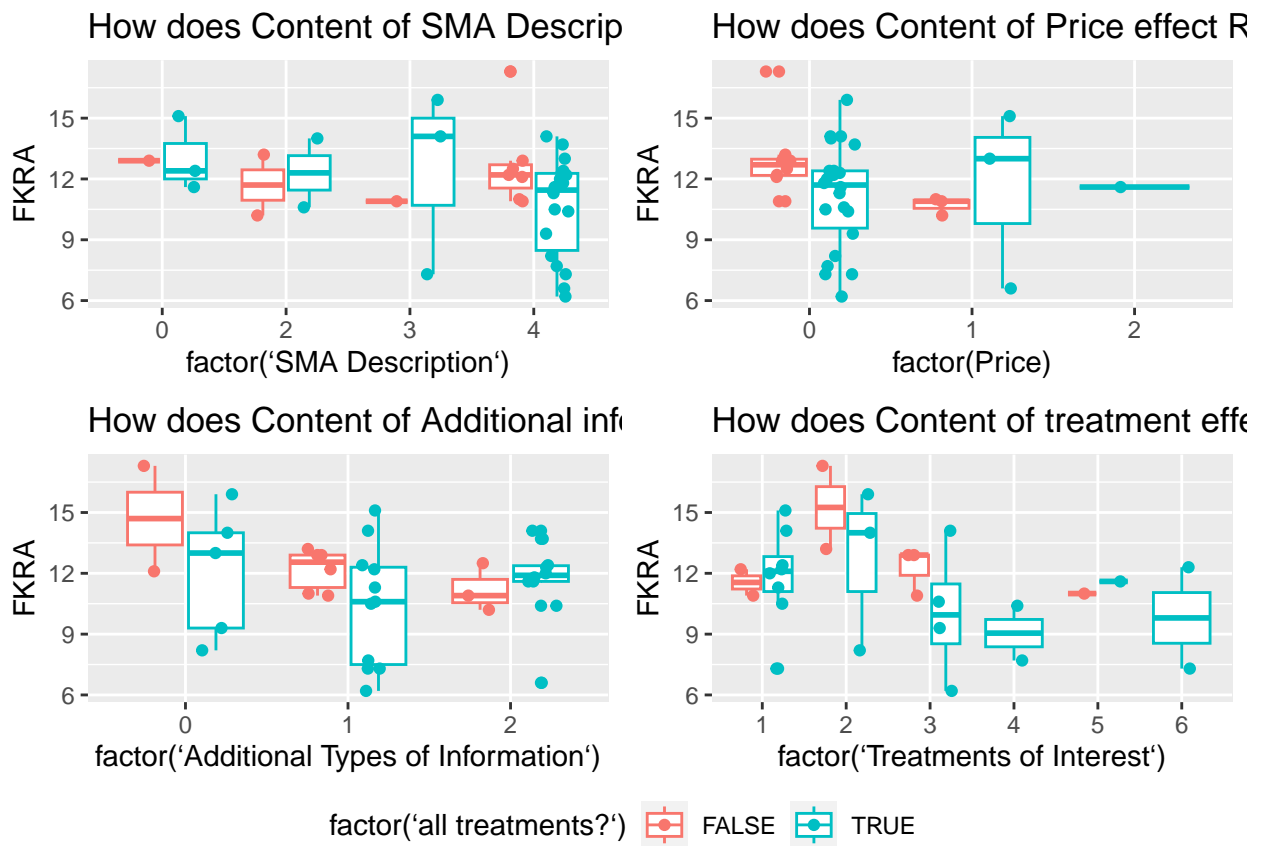
Analysis of content was divided into 4 main categories and 26 sub-categories, jittered side by side

Analysis of DISCERN
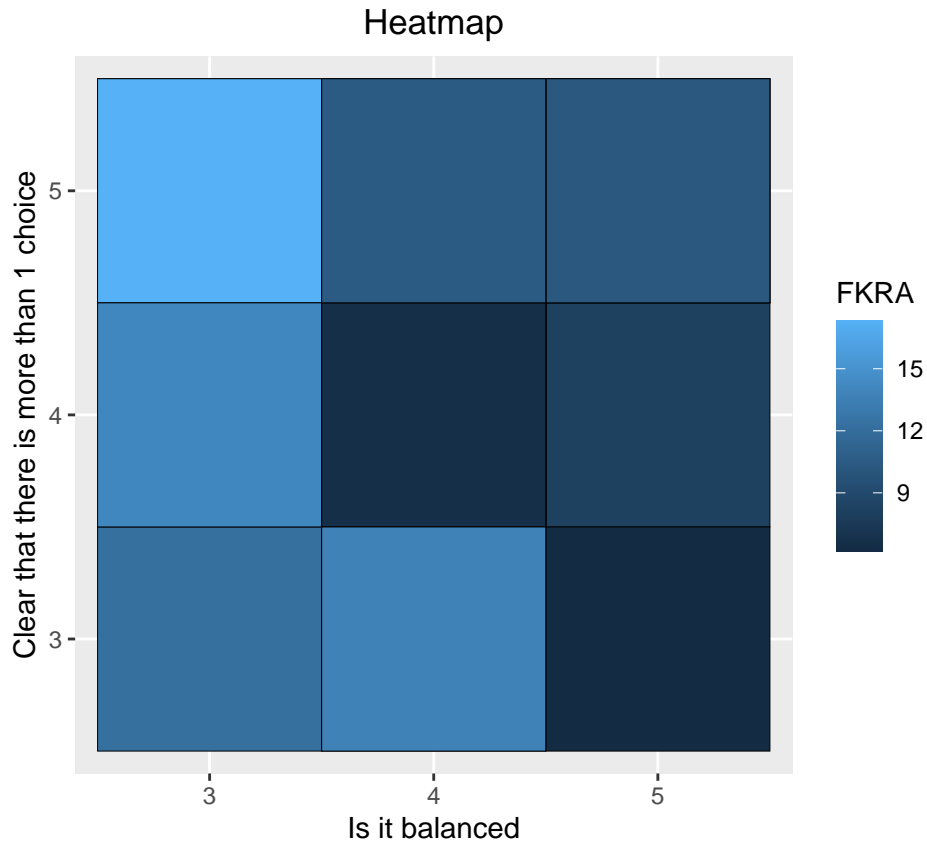
Heatmap

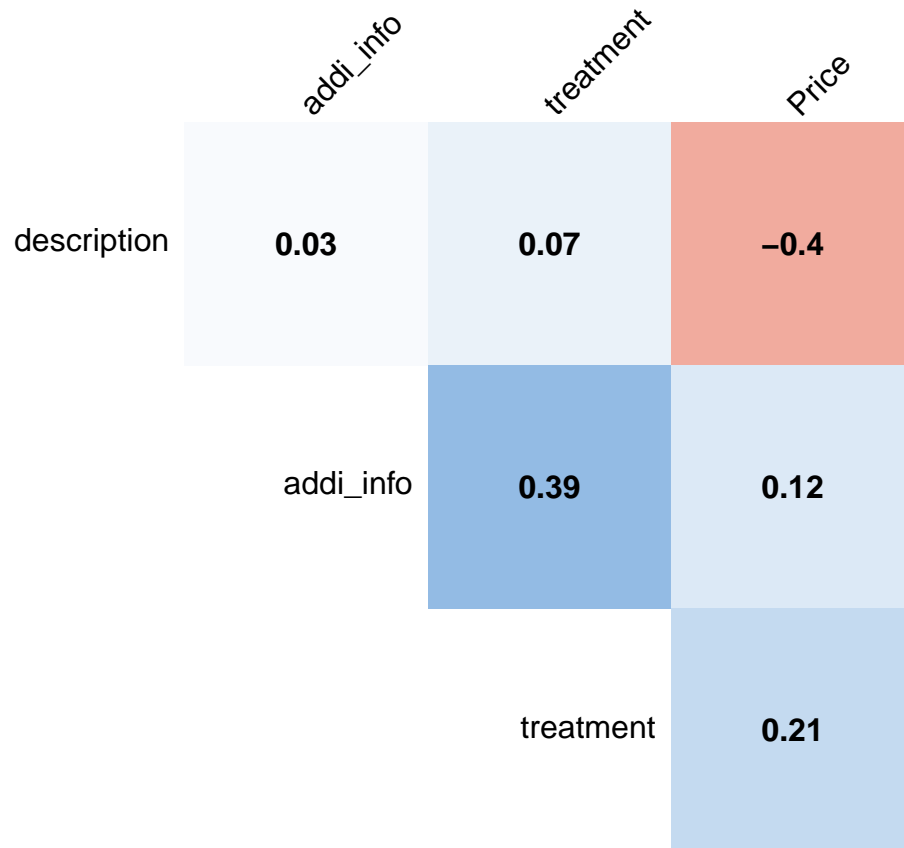A heatmap showing correlations between variables in Discern, with lighter blue require higher level of grade.

Analysis of SMOG

Scatterplot matrix, scatterplots of each pair visualized in left side of the plot and Pearson correlation value and significance displayed on the right side. Use ggpairs to visualize other pairs of variables.
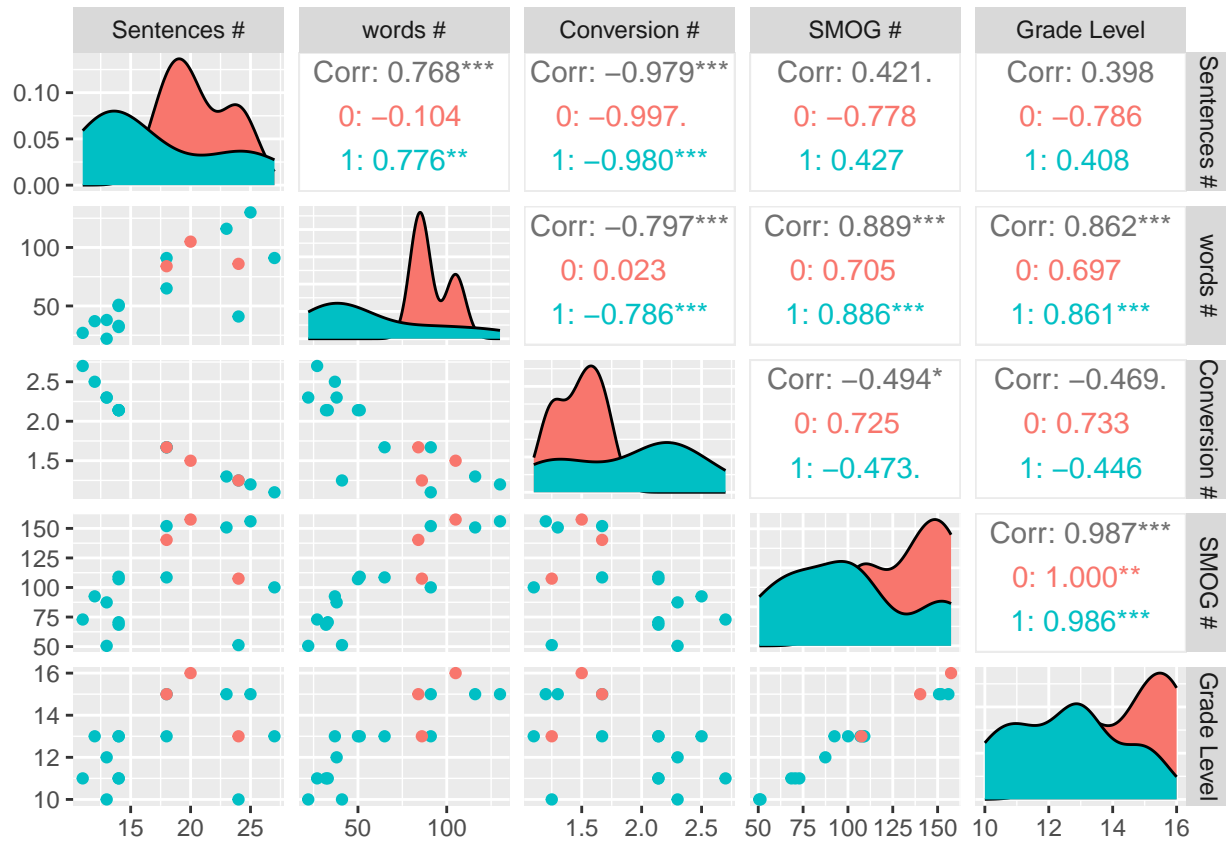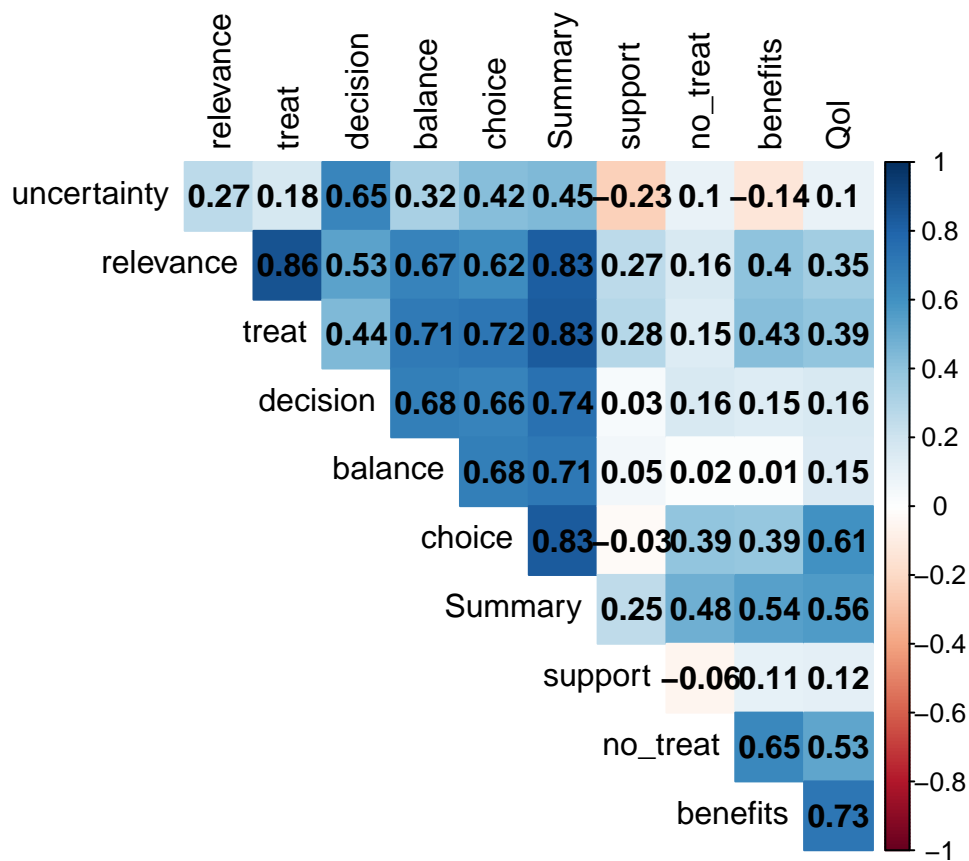
Content Corplot

Pearson-correlation of the variables in Content measurement.

DISCERN Corplot

	relevance	treat	decision	balance	choice	Summary	support	no_treat	benefits	Qol
uncertainty	0.27	0.18	0.65	0.32	0.42	0.45	−0.23	0.1	−0.14	0.1
relevance		0.86	0.53	0.67	0.62	0.83	0.27	0.16	0.4	0.35
treat			0.44	0.71	0.72	0.83	0.28	0.15	0.43	0.39
decision				0.68	0.66	0.74	0.03	0.16	0.15	0.16
balance					0.68	0.71	0.05	0.02	0.01	0.15
choice						0.83	−0.03	0.39	0.39	0.61
Summary							0.25	0.48	0.54	0.56
support								−0.06	0.11	0.12
no_treat									0.65	0.53
benefits										0.73

Pearson-correlation of the variables in Content Discern.

See if the number of treatment described on the website matches the number of treatments in real.
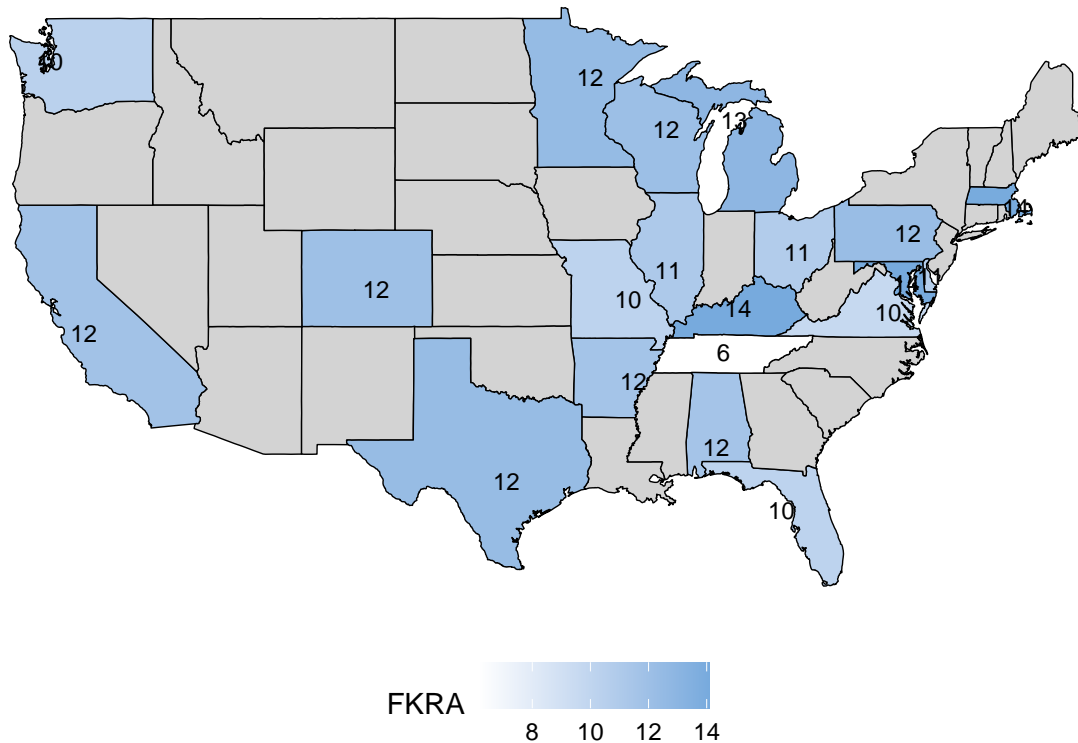
Comparison of number of treatments

# of Treatments offered	0	1	2	3
3	7	6	6	5
2	3	4	3	1
1	0	0	0	0
0	0	1	0	1

# of Treatments mentioned

Count 0 2 4 6
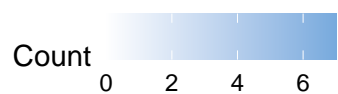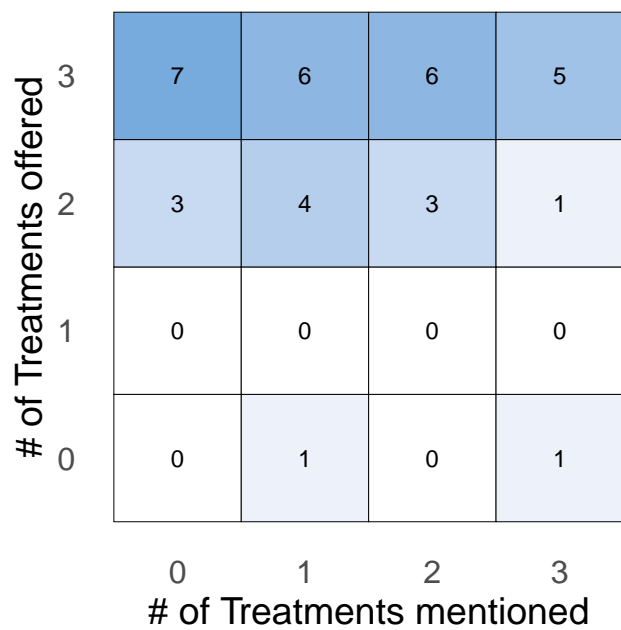
Not exactly the same.

Make a map to see if there State Differences in Reading Score.

FKRA by State

A significant lower score in the state with white color.

Conclusion

In this study, we analyzed the content and readability of websites on spinal muscular atrophy (SMA) treatment from hospital systems. We used the Flesch-Kincaid Grade Level, SMOG indices, and the 16-item DISCERN tool to evaluate readability and content quality. Our findings offer valuable insights into the relationship between these variables and help compare websites more easily. We also examined the client's hypotheses about content quality, affiliation, and the impact of paid consulting companies on readability and specificity.

The study's results can aid organizations in improving their websites' readability and content quality, allowing patients and families to make informed decisions. Our analysis establishes a benchmark for future research and promotes best practices for online health information dissemination. In summary, this project effectively assessed readability and content quality, enhancing our understanding of online health information and facilitating more informed decision-making.