

## Data Visualization Quiz #8: Text and Document Visualization

First Name:

Last Name:

### True or False (Write the entire word in the blank) – 1p Each

\_\_\_\_\_ Tag and Word Cloud visualizations have been shown to underperform relative to other visualizations.

\_\_\_\_\_ Text data is considered ordinal data.

\_\_\_\_\_ Wordles and Word clouds support Shneiderman's Matra.

\_\_\_\_\_ Word frequency is the only useful metric for literary analysis.

\_\_\_\_\_ Topic Modeling aims to extract the major themes from documents.

### Use the following two sentences to answer the next questions about text analysis:

**Sentence 1:** The fox chased the white rabbit.

**Sentence 2:** A brown dog chased a white rabbit.

Convert Sentence 1 into a vector representation, including common words (also called stop words). (1 pt)

Convert Sentence 2 into a vector representation, including common words. (1pt)

Combine the two vector space representations into a matrix suitable for computing distance between the two sentences in vector space. (1 pt)

Change the matrix from the previous question by eliminating columns for stop words (a, an, the, etc.). (1 pt)

Are the two sentences more or less similar after eliminating stop words? Circle only one. (1 pt)

**MORE**

**LESS**