Subject of Focus with Firearms in the 19th and Early 20th Century

Alex Cavallaro, Quantitative Analysis Center, Wesleyan University
Faculty Sponsor: Pavel Oleinikov

Introduction

- Text mining and quantitative textual analysis — the process of representing words and their meanings numerically and apply statistical methods — has become increasingly popular in its use with historians over the past years.
- Few quantitative studies exist examining guns in the late 19th and very early 20th century; most examinations are manual and more qualitative examinations, which are necessarily limited by scale.
- This study seeks to leverage the scalability of quantitative models to explore potential relationships within given texts.

Research Questions

- What sort of gun-related texts appear in newspapers during this time period?
- Is there a correlation between which types of newspaper stories/articles do and don't cover?
- If so, how are they correlated?

Methods

Sample

- Digitized scans of newspapers (n=7,082) from year 1880 to 1922 were downloaded from the Library of Congress' Chronicling America database.
- Newspapers were taken from Oklahoma.

Word Vectors and Cosine Similarity

- Words and texts are capable of being broken down into vectors representing the word.
- The similarity between two of these vectors u and v (and thus their corresponding words) can be measured using cosine similarity with the formula

$$\frac{u \cdot v}{|u||v|} = \cos(\theta)$$

Measures

- Newspaper text was quantified into word vectors.
- Three categories (law-related news stories, gun advertisements, and incident-related news stories) were created and vectorized from within some of the texts
- Cosine similarity was used to classify the topics within a given newspaper with a threshold of 0.625.

Results

Univariate

- 13.3% of newspaper issues contain at least one instance of a gun advertisement, but gun advertisements accounted for 5.7% of actual articles.
- 8.4% of newspaper issues contain at least one article about an incident involving a gun; 3.4% of articles are about gun incidents.
- 1.9% of issue contain at least one article about a gun law; 0.7% of articles are about gun laws

Figure 1.1: Percent of Articles Mentioning Guns by Subject and Year
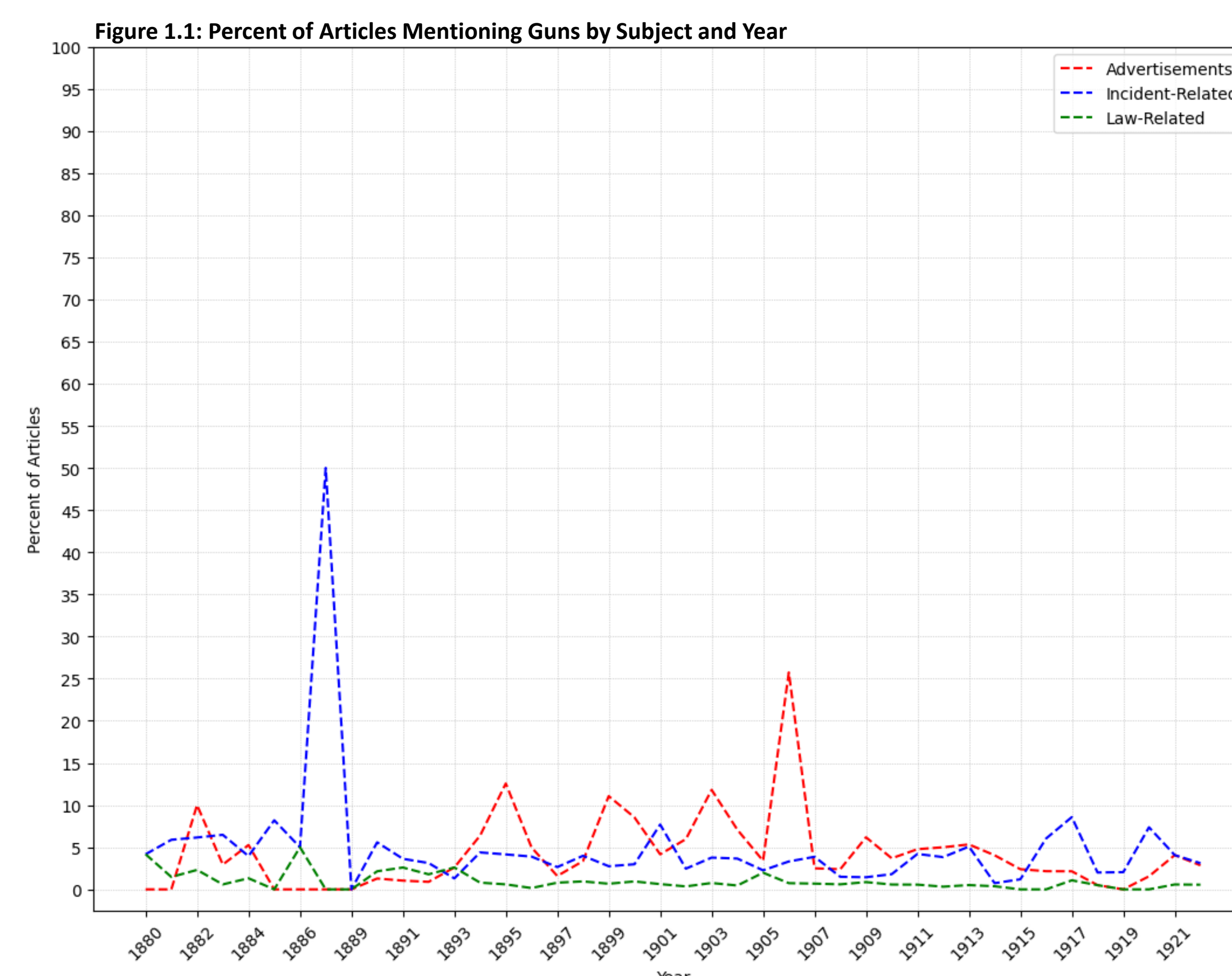
Bivariate

- A Chi-Square test showed that there is a statistically significant relationship between a newspaper featuring **articles about gun incidents** and a newspaper featuring **articles about gun laws** (X2=86.0864, p<0.001).
- A newspaper issue featuring a gun advertisement did not have a statistically significant relationship to a newspaper issue containing an article about a gun law (X2=3.375, p=0.497).
- Similarly, there was no statistically significant relationship between a newspaper issue featuring a gun advertisement and an article about a gun incident (X2=0.028, p=0.998).
- Attempted logistic regressions with a newspaper featuring articles about gun laws produced very inaccurate models (Pseudo R-Squared=0.00826).

Discussion

- A newspaper featuring advertisements for guns is unrelated to whether it features articles about gun incidents or gun laws during this time period.
- The high X2 score between gun incidents and gun laws would indicate a strong correlation, but the inaccuracy of the regression indicates that modeling the data with those variables fails to capture information. Together, these suggest the possibility of a substantial hidden variable that is not in the data.
- Since the data in this study is drawn only from the state of Oklahoma due to circumstantial limitations, further research would be needed with this method across multiple states.