

Jeri E. Wieringa

Ph.D. candidate,
Department of History & Art History,
George Mason University

@jerielizabeth | jeri.elizabeth@gmail.com | jeriwieringa.com

Research Questions

- 1. What is the best strategy for managing time when topic modeling historical documents?
- 2. Can we use topic modeling to see periods of change in a community? To see changes in their conception of time?

Narrative

The question of change over time is central for historians as they trace the changing contours of people, communities, organizations, and states. But computationally tracing changes in discourse over time is not a straightforward process with our current algorithmic tools and methods. Benjamin Schmidt and others have raised concerns about the ways existing tools and methods model change over time. Most text-mining algorithms assume that time is linear and progressive, is experienced as such, and that the data is relatively consistent over time.[1] If we allow that time is relative and that the experience of time shapes the ways a community and its discourses develop, we are faced with the challenge of how to incorporate varied experiences of time into our computational models.

This poster presents my preliminary results applying various periodization schemes to track the development of Seventh-day Adventist discourses around salvation and health over the first 70 years of the denomination's existence.

Methodology

I downloaded 13,000 PDF files from the Seventh-day Adventist *Office of Archives, Statistics, and Research*, split by page, and extracted the text into 197,943 txt files.

I cleaned all of the documents to address standard OCR errors, including line-endings, split words, and burst words. I did not standardize spelling.

I organized the corpus according to five periodization schemes and then preprocessed and modeled all five sets using Gensim.

To evaluate the models, I compared them across one time slice, as demonstrated in this poster. I selected topics related to health and religious beliefs, and examined the ease of labeling, the match between the topic identification and the document, and the distribution of topics over time.

Findings

- 1. The language of the early Seventh-day Adventists poses a challenge for the topic modeling algorithm, as the pervasive use of metaphor and discussion of abstract topics makes it difficult to arrive at “clean” topics.
- 2. Comparing across the periodization schemes, the topics and documents identified are similar, with the greatest variance in the Full Corpus Model.
- 3. The Historical Period Model provides the best correspondence between topics and documents and the greatest ease of labeling.
- 4. All models show an increase in health reform focused pages around 1866 and a relative decline in pages presenting Seventh-day Adventist theology.
- 5. The models provide a useful overview of general trends in the corpus, revealing periods of change in the dominant discourses. However, they are too generalized to reveal shifts in the denomination's conception of time.

Questions for Further Research

- 1. Would better results be achieved using only one title at a time?
- 2. In the cumulative models, would there be a noticeable difference if the documents were weighted by distance from the concluding year?
- 3. What methods can be used in conjunction with topic modeling to surface periods of heightened millennial expectation?

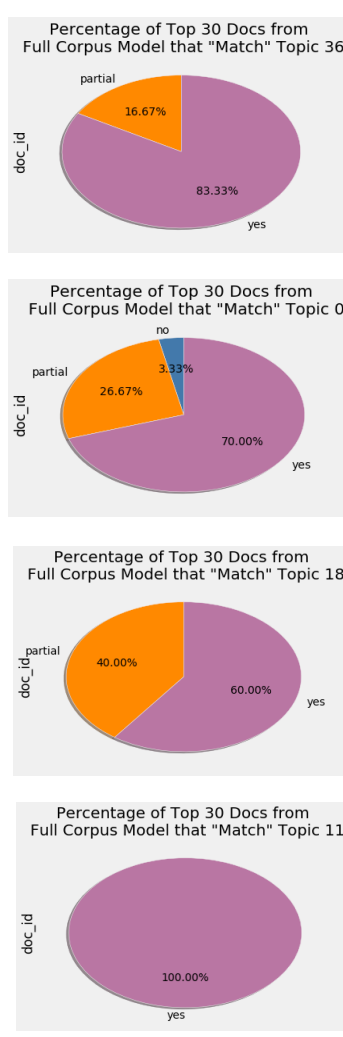
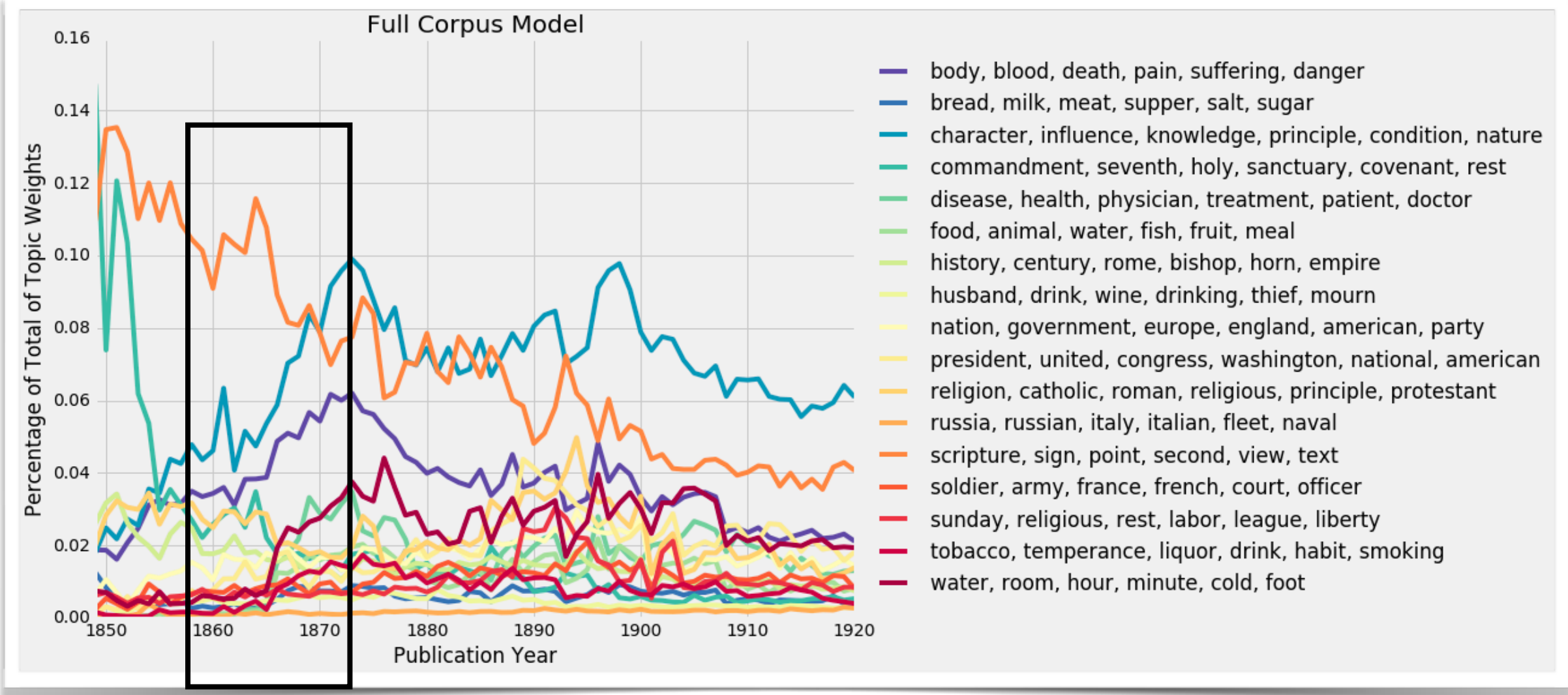
Additional Information

An expanded discussion of methods, references, and the source notebook files are available at <https://github.com/jerielizabeth/DH2017>.

Playing with Time



Full Corpus Model



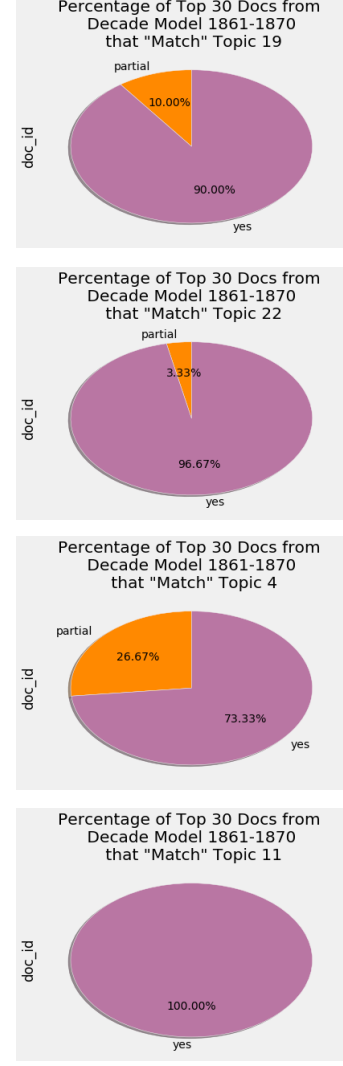
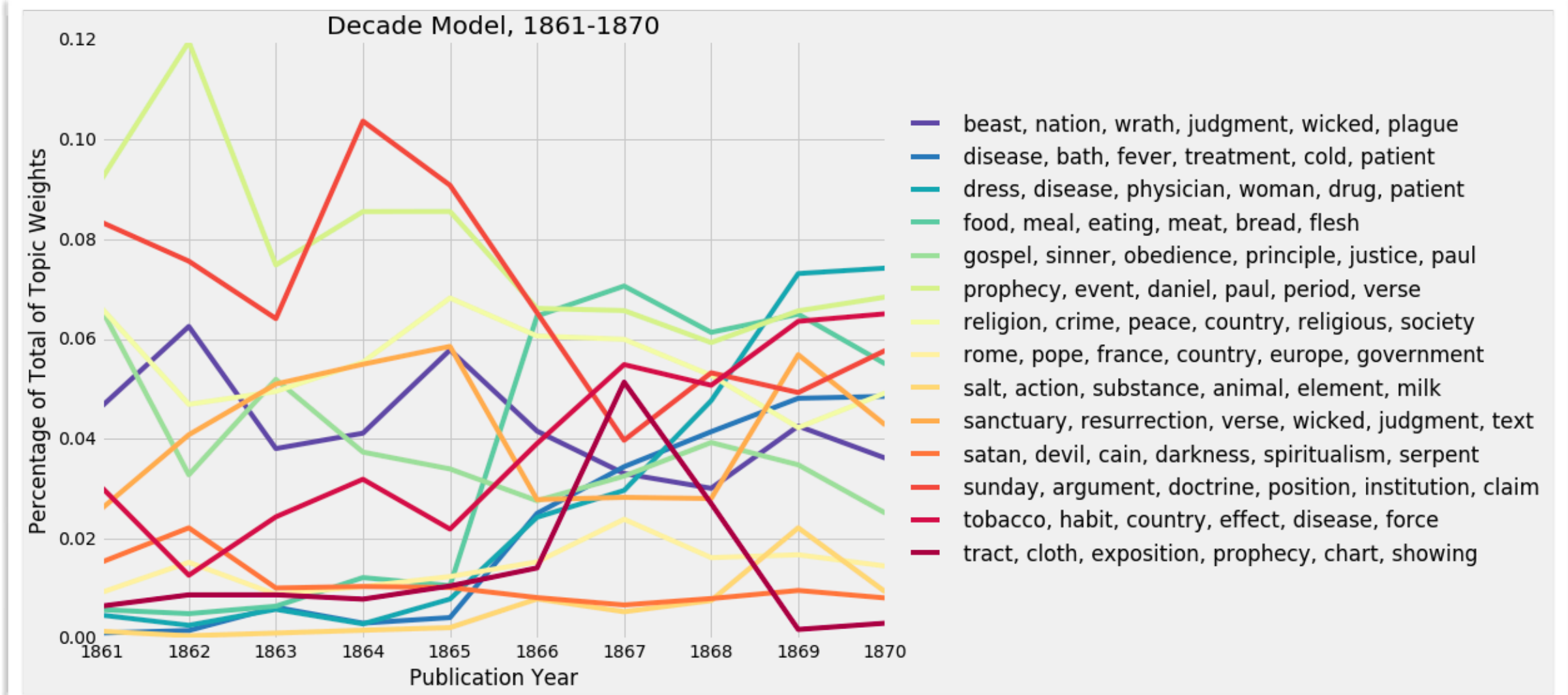
Topic 36
Words: scripture, sign, point, second, view, text, dead, subject, meaning, nature, record, seen, evidence, statement
Hypothesis: The second coming and the Seventh-day Adventist interpretation of scripture.
Final Label: Theological debates, particularly around the nature of the soul.

Topic 0
Words: character, influence, knowledge, principle, condition, nature, study, youth, individual, effort, form, purpose, evil
Hypothesis: Development of character and healthy practices.
Final Label: Health through obedience to natural law; importance of education and parenting in achieving health.

Topic 18
Words: body, blood, death, pain, suffering, danger, exclaimed, trouble, brain, poison, sleep, cell, caused, effect, germ
Hypothesis: Descriptions of consequences of breaking the natural law.
Final Label: Principles of Health Reform; Diseases and their remedies.

Topic 11
Words: commandment, seventh, holy, sanctuary, covenant, rest, keeping, fourth, kept, adventist, change, observance
Hypothesis: Explanations and arguments for Seventh-day Adventist beliefs about the Sabbath and the Sanctuary.
Final Label: Arguments concerning the Sabbath, the Ten Commandments, and God's covenants.

Decade Model, 1861-1870



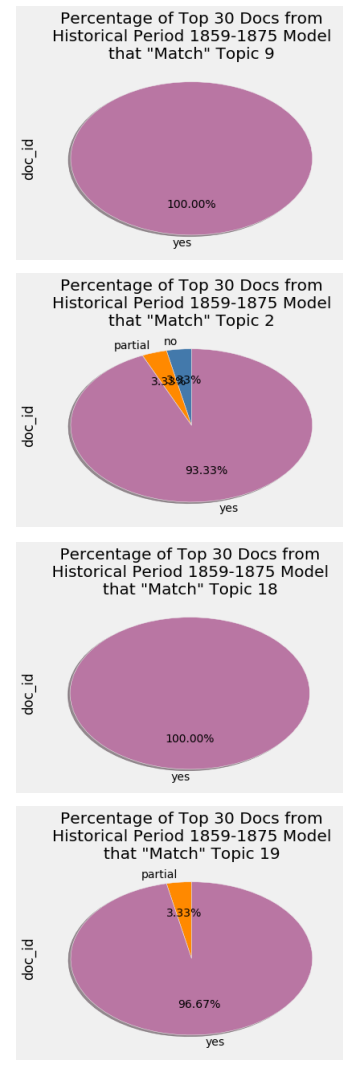
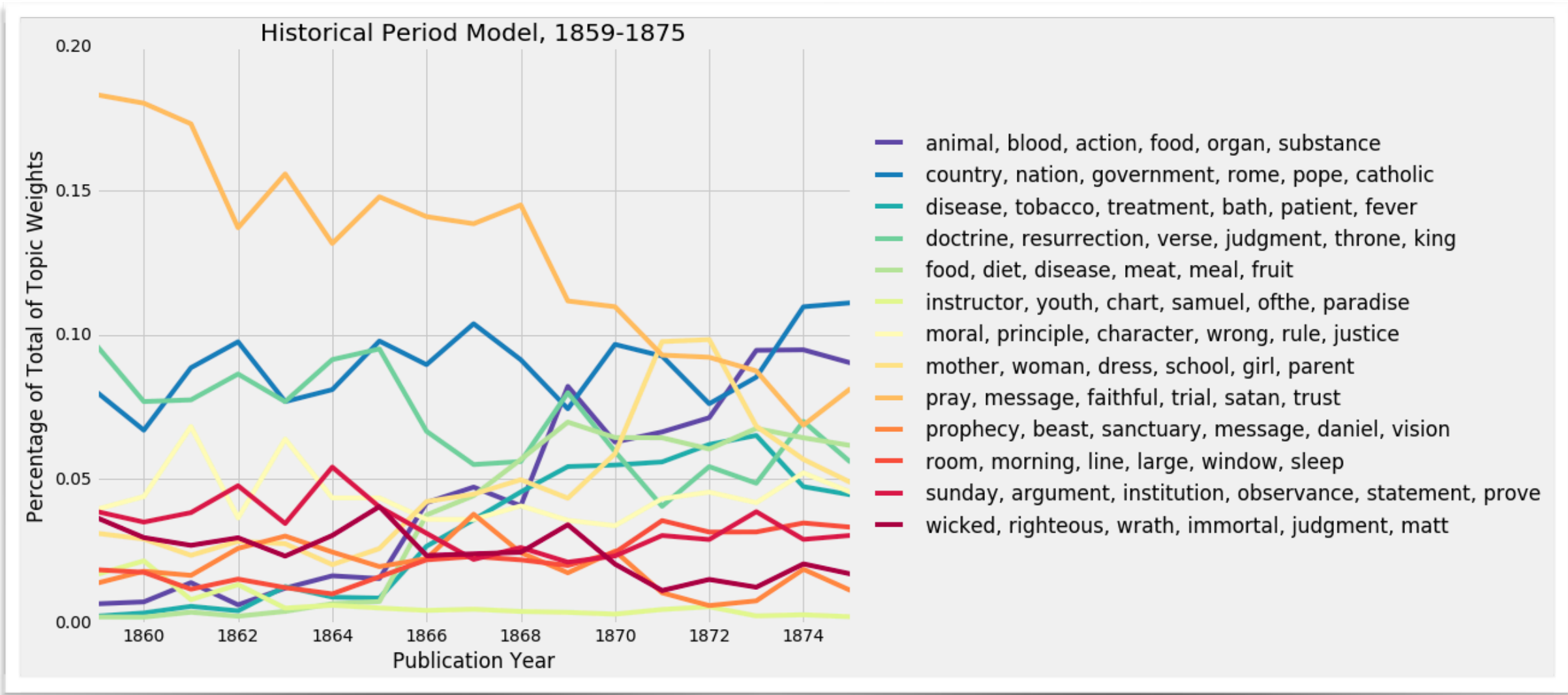
Topic 19
Words: prophecy, event, daniel, paul, period, verse, vision, apostle, prophet, king, history, gospel, chapter, language
Hypothesis: Discussions of prophecies and their interpretation.
Final Label: Defense of Christianity; Arguments concerning Seventh-day Adventist theology.

Topic 22
Words: sunday, argument, doctrine, position, institution, claim, error, change, adventist, observance, prove, statement
Hypothesis: Arguments in favor of Saturday Sabbath.
Final Label: Arguments in defense of Seventh-day Adventism, particularly regarding the Sabbath.

Topic 4
Words: religion, crime, peace, country, religious, society, money, character, conscience, influence, public, moment
Hypothesis: Signs of the second coming, particularly failures of the churches.
Final Label: Signs of the second coming, particularly war; Development of or lack of Christian character.

Topic 11
Words: food, meal, eating, meat, bread, flesh, animal, fruit, diet, stomach, appetite, sugar, drink, vegetable, flour, wheat
Hypothesis: Discussion of the dietary causes of diseases and the virtues of Health Reform.
Final Label: Description of and case for the Health Reform Diet.

Historical Period Model, 1859-1875



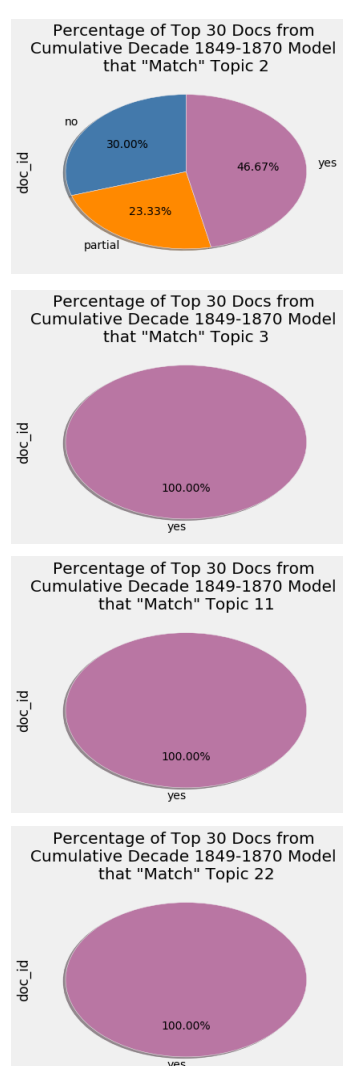
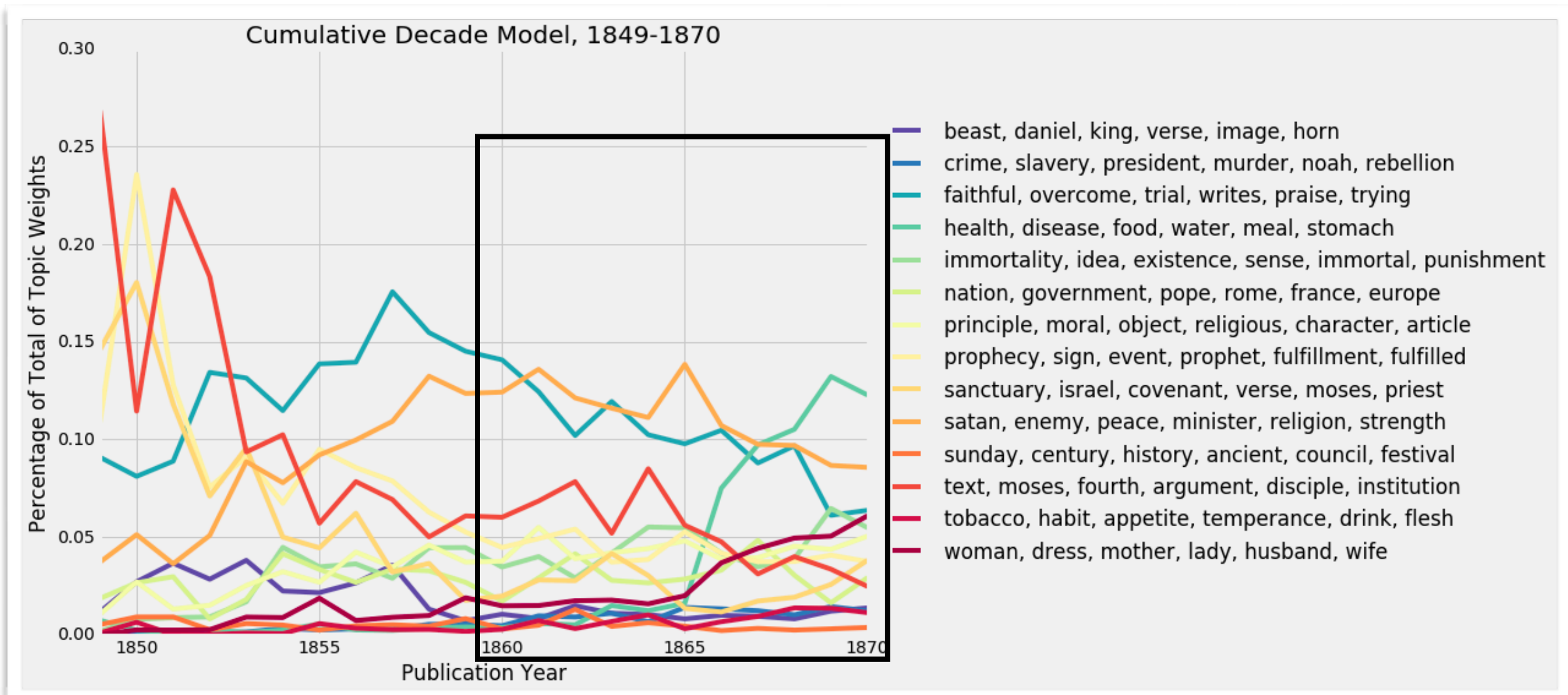
Topic 9
Words: pray, message, faithful, trial, satan, trust, overcome, dear, promise, experience, feeling, seek, mercy, salvation
Hypothesis: Descriptions of Adventist piety, particularly persevering despite delay of the second coming.
Final Label: Letters from readers with testimony and encouragement.

Topic 2
Words: country, nation, government, rome, pope, catholic, europe, religion, history, united, public, england, france
Hypothesis: Interpretations of historical events, framed as a struggle with Catholicism.
Final Label: Interpretation of historical events; Commentary on news and signs of the end of time.

Topic 18
Words: doctrine, resurrection, verse, judgment, throne, king, event, text, peter, idea, future, language, reign, raised
Hypothesis: Discussion of conditional immortality and/or the sanctuary doctrine and Christ's work in heaven.
Final Label: Conditional immortality of the soul and the events of the Final Judgement.

Topic 19
Words: food, diet, disease, meat, meal, fruit, eating, appetite, drink, medicine, drug, physician, salt, wine, flesh
Hypothesis: Discussion of the causes of disease and treatment of disease through embrace health reform.
Final Label: Case for Hygienic Medicine; Description of Health Reform Diet.

Cumulative Decade Model, 1849-1870



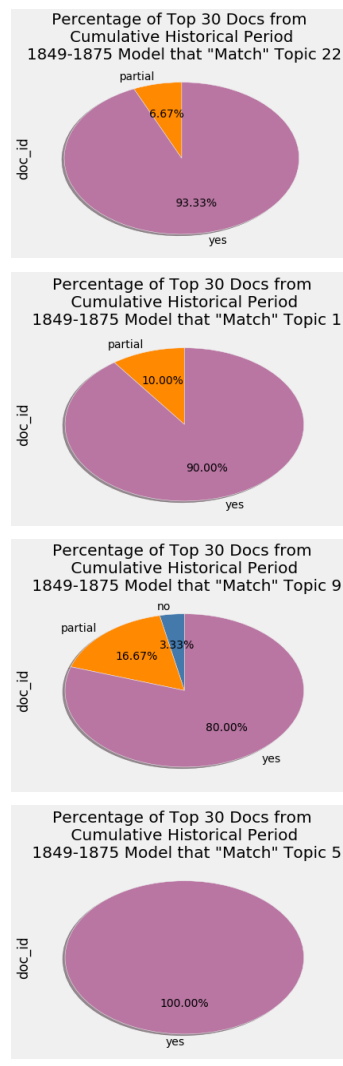
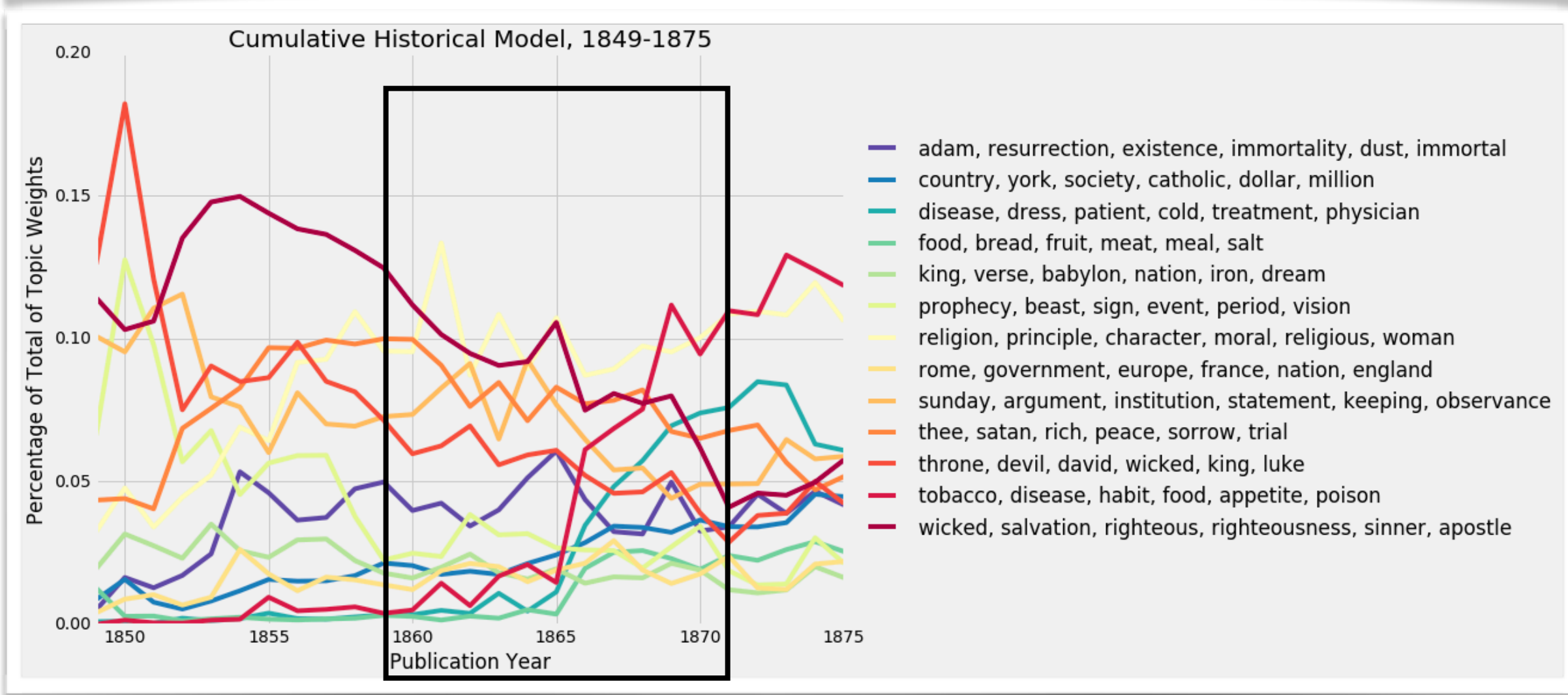
Topic 2
Words: satan, enemy, peace, minister, religion, strength, influence, sinner, devil, mercy, face, cross, deep, danger, turn
Hypothesis: Ambiguous — sense of persevering through some trouble.
Final Label: Unclear. The Christian Life (broadly construed) and the threats of Satan.

Topic 3
Words: faithful, overcome, trial, writes, praise, trying, felt, present, truth, trust, striving, truly, mercy, keeping, rejoice
Hypothesis: Letters from readers offering testimony and expressing belief.
Final Label: Letters from readers with testimony and encouragement.

Topic 11
Words: text, mooses, fourth, argument, disciple, institution, change, proof, creation, prove, new, testament, resurrection
Hypothesis: Arguments for Seventh-day Adventist theology, particularly Seventh-day Sabbath.
Final Label: Debates regarding the Sabbath.

Topic 22
Words: health, disease, food, water, meal, stomach, physician, patient, diet, eating, meat, treatment, bread, cure, fever
Hypothesis: Description of the principles and effectiveness of Health Reform.
Final Label: Applied Hygienic Medicine: descriptions of diseases and their treatments

Cumulative Historical Model, 1849-1875



Topic 22
Words: wicked, salvation, righteous, righteousness, sinner, apostle, jesus, christ, hath, glorious, everlasting, darkness
Hypothesis: Description of the Second Coming.
Final Label: Sermons and lectures on the Judgment and Second Coming.

Topic 1
Words: religion, principle, character, moral, religious, woman, self, wrong, action, object, conscience, rule, crime
Hypothesis: Development of Christian virtues.
Final Label: Warnings against immortal behavior and prescriptions for the development of virtue.

Topic 9
Words: thee, satan, rich, peace, sorrow, trial, seek, self, temptation, pleasure, pray, treasure, suffering, fruit, lead
Hypothesis: Letters from readers offering testimony and stating belief.
Final Label: Exhortations on Christian character and perseverance.

Topic 5
Words: tobacco, disease, habit, food, appetite, poison, psalm, stomach, animal, effect, physical, medicine, action, blood
Hypothesis: Description of the principles and effectiveness of Health Reform.
Final Label: Case for and Principles of Hygienic Medicine.