

# (v. 2) Module 3.b - Analyze a topic cluster

## Survey Flow

ReferenceSurvey: (Ref. Survey) Info about this exercise

Block: WHAT THIS MODULE DOES (1 Question)

Standard: METHOD 1 - SCALED / INTERTOPIC DISTANCE (7 Questions)

Standard: METHOD 2 (HIERARCHICAL CLUSTERING) (1st ITERATION) (6 Questions)

Standard: METHOD 2 (HIERARCHICAL CLUSTERING) (NEXT ITERATIONS) (8 Questions)

Standard: METHOD 2 (HIERARCHICAL CLUSTERING) (SYNTHESIS) (6 Questions)

Standard: METHOD 3: CLUSTERS7D (9 Questions)

Standard: FINAL OBSERVATIONS (6 Questions)

ReferenceSurvey: Ref. Survey Research Takeaways

ReferenceSurvey: (Ref. Survey) Survey end

Page Break

---

IS THIS A TRIAL? Before you begin, please check the choice below if this is just a trial or experimental exercise that should be ignored in collecting outputs from this module.

Trial surveys do not have to be completed. (However, you have to rush through them with empty answers all the way to the finish to clear them from the system. Otherwise, next time you open the same survey, Qualtrics will put you back where you were in the same survey.)

Actual surveys should be completed within one week, after which they will be automatically recorded as finished.

Check this field if this is just a trial: (1)

---

## OPEN RESEARCH POLICY

### **Statement of open research policy**

The modules of the WE1S Interpretation Protocol create records of research that support the project's aim of advancing "open, generalizable, and replicable digital humanities methodology" (see project ["About"](#) statement). This aim is related to current practices of ["open science"](#) and ["open-notebook science."](#) The basic idea is that the materials, methods, steps, and findings of research should be transparent so that other researchers know how conclusions were reached and can test or reproduce the process--an ideal that is especially needed for understanding such machine-learning methods as topic modeling (see [WE1S bibliography on machine-learning interpretability](#)).

In completing this and other modules of the WE1S Interpretation Protocol you are creating an "open notebook" that in principle can be shared with others as a public record of research. *Please be sure that all descriptions, observations, and notes you write in this module are consistent with being part of public record--that is, appropriate in tone, respectful to others, and not in violation of privacy or copyright restrictions.*

*However please also do not let the ideal of open-notebook science inhibit you from speculative or tentative in-progress commentary.* It is understood that research, open or otherwise, is a *process* of finding knowledge with many stages and levels of confidence. (At the conclusion of this and other modules of the Interpretation Protocol, you will be asked to rate your confidence level in the current exercise.)

*In addition, lab notebooks are allowed to be messy as a part of the research process.* It is not

expected that notes written into these modules be copyedited as if for formal publication. (The time for polishing is later when writing up a report synthesizing results from one or more modules of this Interpretation Protocol.)

---

**EXERCISE INFO Please enter the following the following information, which will generate an "Exercise ID".**

Team names and members can be looked up on the [Team Dashboard](#). (Note: at the end of this survey, you can also enter an email address if you wish to receive a copy of the survey with your answers.)

Your name or team name. (If team, then use the format "Team1", "Team2", etc. Please do not use a space): (1) \_\_\_\_\_

Interpretation project number (e.g., "4," indicating that this is the fourth interpretation project for you or your team that the exercise with this module is part of): (2) \_\_\_\_\_

Which stage of your present interpretation project does this module represent? (E.g., if this module is the second you have used in the sequence of modules for the same interpretation project, then write the number "2" here. If multiple team members are working in parallel using various modules, then follow the convention of adding name-abbreviation letters for each member's stage--e.g., "2AL" (for Alan Liu), "2LT" (for Lindsay Thomas), etc. . (3) \_\_\_\_\_

Today's date in the format YYYY-MM-DD (e.g., 2019-07-06) (4) \_\_\_\_\_

The **present** Interpretation Protocol module you are using (e.g., 1, 2, 3a, 4a, etc.) --This is needed to generate an exercise ID. For example, if this is Module 3a, then enter 3a (5) \_\_\_\_\_

---

**G DRIVE INFO Your team folder in the project Google team working space.**

The shared WE1S working space for AM teams in summer 2019 for storing notes, reports, materials, etc. related to interpretation work is in this [Google Drive folder](#). Go there to find the subfolder your team created for your work and enter its URL here.

**Please enter below the URL for your team's subfolder in the common AM Teams workspace:**

\_\_\_\_\_

---

Page Break

---



**EXERCISE ID Exercise ID**

The ID for the present exercise (generated from the information you just provided) is as follows. (Do not change. If you see a mistake, please go "back" in this survey and correct the information that generated the Exercise ID.)

---

**EXERCISE DESCRIP Exercise Description**

To assist in remembering what the exercise using this module is about, please enter below a very brief description of what you are currently doing (e.g., "We are now going to compare the keywords 'humanities' and 'sciences' in the 20190621\_2132\_us-humanities-top-newspapers---topics200 model.")

---

---

---

---

---

## TOPIC MODEL **Topic model you are studying**

For the topic model you are studying, please enter below its "**topic model name**" and "**start page**" URL. A "**topic model name**" looks like this: "20190621\_2132\_us-humanities-top-newspapers---topics200". You can locate this name in the [Registry of WE1S Topic Models](#).

Note that this is the name of the specific granularity of the model you are studying (number of topics). The "**start page**" is the URL of the overall topic model (including all its granularities and their visualizations) listed on this page on the Harbor 10002 server: [Index of / projects/](#). (You can also find this URL in the Registry of WE1S Topic Models.)

A start page URL looks like this: [http://harbor.english.ucsb.edu:10002/projects/20190621\\_2132\\_us-humanities-top-newspapers/](http://harbor.english.ucsb.edu:10002/projects/20190621_2132_us-humanities-top-newspapers/)

"Topic model name" (1) \_\_\_\_\_

"Start page" URL (2) \_\_\_\_\_

---

## RESEARCH QUESTION **Research question you are addressing**

Please enter below information about the research question you are addressing in this exercise. You can find this information in the "Registry of WE1S Research Questions." (If you are taking an overview of a model prior to working on a specific research question, just enter "Taking an overview")

Research question (1) \_\_\_\_\_

Operationalized form of question (2)  
\_\_\_\_\_

Research question ID (use the format "team2-q1", representing in this example team 2's first research question): (4) \_\_\_\_\_

**PREVIOUS MODULES Previous Interpretation Protocol modules you used to address your research question (if any)**

Enter in sequence below any Interpretation Protocol modules you have already completed in previous steps of your current interpretation project. For example, if for steps 1-3 preceding this exercise (where the current exercise is now step 4 in the sequence) you previously used modules 3a, 3b, and 3c in that order, then enter those module numbers in that sequence.

- Step 1 - Module used (1) \_\_\_\_\_
- Step 2 - Module used (2) \_\_\_\_\_
- Step 3 - Module used (3) \_\_\_\_\_
- Step 4 - Module used (4) \_\_\_\_\_
- Step 5 - Module used (5) \_\_\_\_\_
- Step 6 - Module used (6) \_\_\_\_\_
- Step 7 - Module used (7) \_\_\_\_\_
- Step 8 - Module used (8) \_\_\_\_\_
- Step 9 - Module used (9) \_\_\_\_\_
- Step 10 - Module used (10) \_\_\_\_\_

-----  
Page Break \_\_\_\_\_

R1: exercise info

Reminders - Info about this Exercise: **Your name or team name:** `{EXERCISE INFO/ChoiceTextEntryValue/1}` **`{G DRIVE INFO/ChoiceTextEntryValue}`**  
**target="\_blank">Link to your team's folder on the WE1S project Google team drive.**

**Exercise ID:** `{EXERCISE ID/ChoiceTextEntryValue}` **Exercise**  
**Description:** `{EXERCISE DESCRIP/ChoiceTextEntryValue}` **Topic model name:**  
`{TOPIC MODEL/ChoiceTextEntryValue/1} {TOPIC MODEL/ChoiceTextEntryValue/2}`  
**target="\_blank">Start page of model** **Research question (and**  
**ID):** `{RESEARCH QUESTION/ChoiceTextEntryValue/1}` (research question ID #`{RESEARCH QUESTION/ChoiceTextEntryValue/4}`) **Operationalized form of question:** `{RESEARCH QUESTION/ChoiceTextEntryValue/2}` **Previous sequence of interpretation modules used for this question (if any):** `{PREVIOUS MODULES/ChoiceTextEntryValue/1}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/2}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/3}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/4}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/5}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/6}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/7}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/8}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/9}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/10}`

End of Block: INFO ABOUT YOU & EXERCISE (v. 2, created 9 June 2019, rev. 7 July 2019)

---

Start of Block: WHAT THIS MODULE DOES

## MODULE PURPOSE

### What module 3.b does

(Module 3.b, v. 2, created 23 June 2019; last rev. 25 June 2019.)

The purpose of module 3.b of the WE1S topic model Interpretation Protocol is to examine a topic cluster.

### "Meaningful" and "Robust" Topic Clusters Defined.

Intuitively, it should be possible to group together "related" topics based on whether the topics contain the same words in similar proportions. Such groupings may be termed "topic-clusters". The easiest way to conceptualize a topic-cluster is as a group of circles, where each circle represents a topic and is plotted on the X/Y coordinates of a two-dimensional statistical space that is a simplification of the multidimensional statistical "nearness" and "farness" of topics relative to each other. Drawing a larger circle around a set of topic circles that are statistically "near" each other (and "distant" from others) groups that set of circles into a cluster.

However, there are many types of statistical algorithms that can be used to cluster topics (or any data), and the fact that these employ different rules means that they may yield different



apparent topic-clusters. It is thus important to evaluate apparent topic-clusters for their meaningfulness and robustness. “Meaningful” means that topic-clusters contain topics that appear to have genuinely related meanings (rather than being topics that are thrown together as a statistical artifact of the algorithm). “Robust” means that many of the topics in topic-clusters seem to appear “near” each other regardless of which algorithm or topic modeling interface we employ.

**Important caution:** In general, you should consider the analysis of topic-clusters to be only *supplementary* to understanding the main phenomena shown in topic models: the topics themselves (and the relations they show between words and documents in a corpus). Consider that topics are already a reduction of the complexity of relations between words and documents. After all, that is their point: to categorize high-dimension data (e.g., millions of possible relations) into a lower-dimension set (e.g., just 200 topics) making it easier to see patterns. This means that clusters of topics are a *reduction of a reduction* in complexity—a second-order reduction. They might enhance our understanding of a topic model by showing larger-scale patterns at work; but at best they can do so with less confidence. If topic-cluster analysis using one tool seems to show something important, you can increase confidence in the results by seeing if other tools show the same results. Or you can examine a lower-granularity version of your topic model (e.g., 50 topics instead of 200) to see if a cluster you think you have found shows up as one of the grosser topics in such a model.

**A General Note on Methods and Tools in this Module.** You will be using three methods for examining a topic cluster. *Method 1* uses the general-purpose visualization interfaces in the WE1S Topic Model Observatory (Dfr-browser, TopicBubbles, and pyLDAvis). Methods 2 and 3 use the special-purpose tools Clusters7D and DendrogramViewer in the Topic Model Observatory, which may give a surer and clearer understanding of clusters.)

Concise reminders of how to use these various tools to look for clusters follow below. (For more detailed instructions and best practices, consult the [TMO Guide](#) on the relevant tools.

End of Block: WHAT THIS MODULE DOES

---

Start of Block: METHOD 1 - SCALED / INTERTOPIC DISTANCE

R1 exercise info [\\${R1: exercise info/QuestionText}](#)

---

## METHOD 1 INTRO **Method 1 (Scaled / Intertopic distance)**

First you will examine the topic model "space" in which the relative similarity/difference between topics is visualized in the "scaled" view in Dfr-browser and TopicBubbles or the "intertopic

distance" view in pyLDAvis. Here the multidimensional statistical "distance" between topics is shown as projected onto a 2-dimensional space. However, just as constellations in the sky appear the way they do from earth, but may be arranged differently if we were to look at the stars from another planet, so the projection of multidimensional relations onto 2D space can create clusters that are only apparent, while hiding clusters that are not apparent from a particular angle of view. (After using Method 1 to gain an approximate sense of the cluster structure of a topic model, use Methods 2 and 3 for more possibly surer ways of identifying clusters.)

a. b.

c.

**Screenshots: (a) Dfr-browser "scaled view"; (b)TopicBubbles "scaled view"; (c) pyLDAvis "intertopic distance" view**



**METHOD1 - STEP1 Step 1 (of scaled / intertopic distance method) - Get an overview.**

Using a combination of the "scaled" views in Dfr-Browser and Topic Bubbles, and the "intertopic distance" visualization in pyLDAvis, locate topics that appear to cluster together. Then examine some of the topics in the most obvious clusters (those that are largest, most dense, most clearly separated, or otherwise noticeable). Take some notes for yourself about such features as the following: What are the most obvious apparent clusters? Are these clusters characterized by large topics (heavily represented in the corpus)? Do the keywords of topics in the clusters suggest that the topics in fact are similar (share subject matter of discursive features)?

---

---

---

---

---

---

**METHOD1 - STEP2 Step 2 (of scaled / intertopic distance method) - Identify a cluster for study.**

Now single out one topic cluster that you want to study to address the research question you

are working on. Examine some of the topics in the cluster, looking at their keywords and some of the articles associated with them.

---

**LABEL CLUSTER Provide a label for the cluster that describes its meaning.** (A label should be a simple phrase of between one and three words. Avoid making a label into a sentence predicating a specific relationship between terms, which often amounts to a thesis statement.)

Label for cluster you are singling out for study:

---

**TOPICS IN CLUSTER List below by their numbers a few topics that you think can serve as representatives of this cluster** (up to 10 topics).

	Answer 1 (1)	Answer 2 (2)	Answer 3 (3)	Answer 4 (4)	Answer 5 (5)	Answer 6 (6)	Answer 7 (7)	Answer 8 (8)	Answer 9 (9)	Answer 10 (10)
Topic #'s: (1)										

---

Page Break

GO TO NEXT METHOD This completes your use of Method 1 (scaled / intertopic distance in Dfr-Browser, TopicBubbles, and pyLDAvis). Pressing "next" will bring you to Method 2 for cluster analysis.

End of Block: METHOD 1 - SCALED / INTERTOPIC DISTANCE

---

Start of Block: METHOD 2 (HIERARCHICAL CLUSTERING) (1st ITERATION)

R1 exercise info `#{R1: exercise info/QuestionText}`

---

R data you entered 1

**Reminders - Data you entered above:**

The topic cluster you examined using **Method 1** (scaled / intertopic distance) is one you labeled: `#{LABEL CLUSTER/ChoiceTextEntryValue}`

Topics you identified as representative of the above topic cluster are (by topic numbers):

`#{TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/1}`, `#{TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/2}`, `#{TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/3}`, `#{TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/4}`, `#{TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/5}`, `#{TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/6}`, `#{TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/7}`, `#{TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/8}`, `#{TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/9}`, `#{TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/10}`

---

**METHOD 2 INTRO Method 2 (Hierarchical clustering using dendrograms)**

To help confirm or correct the view of possible clusters described above, you will use DendrogramViewer to apply a likely more reliable method called "hierarchical clustering." This method shows visualizations of relations among topics as "dendrograms." In these dendrograms, topics that are statistically close to each other are gathered as "leaves" under the same branch of a branching tree structure ("clade"). (For a fuller explanation, see ["Hierarchical Cluster Analysis"](#), ["How to Read a Dendrogram"](#), or watch the [YouTube video](#)). Ideally, the dendrograms will confirm the existence of clusters you spot with the previous scaled/intertopic distance method or show new clusters. Where the dendrograms show clustered topics that do

not appear in the scaled/intertopic distance method--or where they do *not* show what the scaled/intertopic distance method shows--they are generally to be relied on instead. However, it may be the case that the dendrograms do not give you a confident sense of "real" clusters either after you have examined the topics in an apparent cluster more closely. If that is the case, then you may need to rely on a synthesis of your impressions from all methods combined, or--just as valuable--declare a "null result" (no finding that can be made about a cluster).

1. 2.

3. 4.

**Screenshots: (1) DendrogramViewer main interface; (2) selecting a possible cluster for zoomed-in view; (3) zoomed-in selection; (4) parts of a dendrogram.**

**Multiple Variant Dendrograms in DendrogramViewer:**

DendrogramViewer provides five variant dendrograms for a topic model generated using different statistical algorithms, which are accessed from tabs at the top of the interface (labeled "a" in screenshot 1 above). You will iterate the process of looking at all these dendrograms.



ITERATION 1.i Iteration 1: "Euclidean distance and average linkage"

**Step 1 (of hierarchical clustering using dendrograms method):** To begin with, get an overview using the dendrogram produced for this topic model through "**Euclidean distance and average linkage.**" Make some notes about the "clades" (branches of topics) you see. These groupings are potentially interpretable as "topic clusters." Has the dendrogram identified clades of topics that you did not previously identify as a cluster using the "scaled"/"intertopic distance" views in other tools? Or did the dendrogram make clusters that you previously thought you saw disappear? Other possible questions to consider: which topics seem most closely related in the dendrogram? Do any clusters appear like they might be capturing mega-topics or, more meaningfully, meta-topics?

Take some notes for yourself about such features as the following:      What are the most obvious apparent clusters?    Are these clusters characterized by large topics (heavily represented in the corpus)?    Do these clusters also appear to show up in the "scaled" view of topics in Dfr-Browser and TopicBubbles, and "intertopic distance" view in pyLDAvis?    (Keep in mind that clusters as they appear in Dfr-Browser, TopicBubbles, and pyLDAvis are like constellations of stars--possibly apparent as clusters from one angle of view in the statistical space being visualized, but inapparent from other angles of view. TopicClusters gives the surest indication of clusters.)      Do the keywords of topics in the clusters (as seen in Dfr-Browser, TopicBubbles or pyLDAvis) suggest that the topics in fact are similar (share subject matter of discursive features)?

**Your notes:**

---

---

---

---

---

---

Page Break



ITERATION 1.ii **Step 2 (of dendrogram hierarchical clustering using dendrograms method)**: Now single out one topic cluster that you want to study to address the research question you are working on. (*You can draw a box around a part of the dendrogram to zoom in on it.*) Use the dendrogram to pick out topics of interest in the cluster, and then use Dfr-Browser, TopicBubbles, or pyLDAvis if you want to look at the keywords and also some of the articles associated with the topics.

Take some notes for yourself about such issues as the following:                      Similarities or differences from the cluster you saw in Dfr-Browser, TopicBubbles, and pyLDAvis.

Possible label for the cluster (or revision of the label you previously assigned a cluster you saw in Dfr-Browser, TopicBubbles, and pyLDAvis).      Representative topics by number (or a revision of the representative topics you previously identified).

**Your Notes:**

---

---

---

---

---

---

Page Break

**NEXT ITERATIONS** Next you will iterate by looking at the other available dendrograms provided by DendrogramViewer.

End of Block: METHOD 2 (HIERARCHICAL CLUSTERING) (1st ITERATION)

---

Start of Block: METHOD 2 (HIERARCHICAL CLUSTERING) (NEXT ITERATIONS)

R1 exercise info [\\${R1: exercise info/QuestionText}](#)

---

R data you entered [\\${R data you entered 1/QuestionText}](#)

---

## METHOD 2 ITERATIONS **Method 2 (Hierarchical clustering using dendrograms) (next iterations)**

Above, you used DendrogramViewer to study the dendrogram generated by "Euclidean distance and average linkage." Now repeat the exercise for the following dendrograms (below).

### **Reminder of the two steps for method 2:**

**Step 1: Get an Overview** Examine the dendrograms. Make some notes about the "clades" (branches of topics) you see. These groupings are potentially interpretable as "topic clusters." Has the dendrogram identified clades of topics that you did not previously identify as a cluster using the "scaled"/"intertopic distance" views in other tools? Or did the dendrogram make clusters that you previously thought you saw disappear? Other possible questions to consider: which topics seem most closely related in the dendrogram? Do any clusters appear like they might be capturing mega-topics or, more meaningfully, meta-topics?

Take some notes for yourself about such features as the following: What are the most obvious apparent clusters? Are these clusters characterized by large topics (heavily represented in the corpus)? Do these clusters also appear to show up in the "scaled" view of topics in Dfr-Browser and TopicBubbles, and "intertopic distance" view in pyLDAvis? (Keep in mind that clusters as they appear in Dfr-Browser, TopicBubbles, and pyLDAvis are like constellations of stars--possibly apparent as clusters from one angle of view in the statistical space being visualized, but inapparent from other angles of view. TopicClusters gives the surest indication of clusters.) Do the keywords of topics in the clusters (as seen in Dfr-Browser, TopicBubbles or pyLDAvis) suggest that the topics in fact are similar (share subject matter of discursive features)?



**Step 2: Identify from the dendrogram a particular cluster for study** Now single out one topic cluster that you want to study to address the research question you are working on. (*You can draw a box around a part of the dendrogram to zoom in on it.*) Use the dendrogram to pick out topics of interest in the cluster, and use Dfr-Browser, TopicBubbles, and pyLDAvis, to look at their keywords and also some of the articles associated with them.

Take some notes for yourself about such issues as the following:                      Similarities or differences from the cluster you saw in Dfr-Browser, TopicBubbles, and pyLDAvis.

Possible label for the cluster (or revision of the label you previously assigned a cluster you saw in Dfr-Browser, TopicBubbles, and pyLDAvis).      Representative topics by number (or a revision of the representative topics you previously identified).

---



## ITERATION 2

Iteration 2: "Euclidean distance, single linkage dendrogram"

### Your notes:

---

---

---

---

---

---

Page Break



### ITERATION 3

Iteration 3: "Euclidean distance, complete linkage dendrogram"

**Your notes:**

---

---

---

---

---

---

Page Break

---



## ITERATION 4

Iteration 4: "Euclidean Distance with Ward Linkage"

**Your notes:**

---

---

---

---

---

---

Page Break

---



## ITERATION 5

Iteration 5: "Cosine distance, average linkage dendrogram"

**Your notes:**

---

---

---

---

---

---

Page Break

---

**GO TO SYNTHESIS Pressing "next" will bring you to a step for synthesizing your study of the above variant dendrograms.**

End of Block: METHOD 2 (HIERARCHICAL CLUSTERING) (NEXT ITERATIONS)

---

Start of Block: METHOD 2 (HIERARCHICAL CLUSTERING) (SYNTHESIS)

R1 exercise info [\\${R1: exercise info/QuestionText}](#)

---

R data you entered [\\${R data you entered 1/QuestionText}](#)

---

**LABEL CLUSTER Method 2 (Hierarchical clustering using dendrograms) - Bringing together your views of the different versions of dendrograms**

Now that you have had a chance to study the cluster you are interested and its variants in different dendrograms, provide a label for the cluster that describes its meaning. If the cluster is the same as the one our identified using the "scaled"/"intertopic distance" method earlier and labeled "[\\${LABEL CLUSTER/ChoiceTextEntryValue}](#)", you can here revise the label you previously gave it if needed. Otherwise, enter a new label here. (A label should be a simple phrase of between one and three words. Avoid making a label into a sentence predicating a specific relationship between terms, which often amounts to a thesis statement.)

---

TOPICS IN CLUSTER List below by their numbers a few topics that you think can serve as representatives of this cluster. (You previously listed the following topics in the cluster you singled out using the "scaled"/"intertopic distance" above: [\\${TOPICS IN](#)

[CLUSTER%231/ChoiceTextEntryValue/1/1}](#), [\\${TOPICS IN](#)  
[CLUSTER%231/ChoiceTextEntryValue/1/2}](#), [\\${TOPICS IN](#)  
[CLUSTER%231/ChoiceTextEntryValue/1/3}](#), [\\${TOPICS IN](#)  
[CLUSTER%231/ChoiceTextEntryValue/1/4}](#), [\\${TOPICS IN](#)  
[CLUSTER%231/ChoiceTextEntryValue/1/5}](#), [\\${TOPICS IN](#)  
[CLUSTER%231/ChoiceTextEntryValue/1/6}](#), [\\${TOPICS IN](#)  
[CLUSTER%231/ChoiceTextEntryValue/1/7}](#), [\\${TOPICS IN](#)  
[CLUSTER%231/ChoiceTextEntryValue/1/8}](#), [\\${TOPICS IN](#)  
[CLUSTER%231/ChoiceTextEntryValue/1/9}](#), [\\${TOPICS IN](#)

CLUSTER%231/ChoiceTextEntryValue/1/10}. If the dendrogram method is showing you the same cluster, then reenter those topic numbers here. If the dendrogram method leads you to single out a different cluster or one that only partly overlaps with the previous one you singled out, then enter a new set of numbers here that represents the cluster.

	Answer 1 (1)	Answer 2 (2)	Answer 3 (3)	Answer 4 (4)	Answer 5 (5)	Answer 6 (6)	Answer 7 (7)	Answer 8 (8)	Answer 9 (9)	Answer 10 (10)
Topic #s: (1)										



**SUMMARY NOTES DENDRO Finish your application of Method 2 (hierarchical clustering using dendrograms) by writing an observation synthesizing or evaluating what you learned from the five variant dendrograms.**

(Scroll to bottom of your running notes to add observations.)

---



---



---



---



---

Page Break

GO TO NEXT METHOD This completes your use of Method 2 (hierarchical clustering using dendrograms). Pressing "next" will bring you to Method 3 for cluster analysis.

End of Block: METHOD 2 (HIERARCHICAL CLUSTERING) (SYNTHESIS)

---

Start of Block: METHOD 3: CLUSTERS7D

R1 exercise info [\\${R1: exercise info/QuestionText}](#)

---

R data you entered 1 [\\${R data you entered 1/QuestionText}](#)

---

R data you entered 2

#### Reminders - Data you entered above:

The topic cluster you examined using **Method 2** (hierarchical clustering using dendrograms) is one you labeled: [\\${LABEL CLUSTER/ChoiceTextEntryValue}](#)

Topics you identified as representative of the above topic cluster are (by topic numbers):

[\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/1}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/2}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/3}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/4}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/5}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/6}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/7}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/8}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/9}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/10}](#)

---

#### METHOD 3 INTRO **Method 3 (Cluster analysis using Clusters7D)**

The Clusters7D tool in the Topic Model Observatory performs a second-order topic modeling of your topic model (creating 20 "topics" or clusters out of the original topics in the model). It then displays the results using two versions of the pyLDavis interface (see screenshots). In both views, clusters are represented in the left panel, with the size of the circles representing the relative weight of the cluster's topics in the model. The variation lies in the right panel, where one version of Clusters7D shows the *topics* in a cluster (and ranks them by their weight in the

cluster) and the other version shows the *words* associated with the cluster (and ranks them by their weight). For more detailed instructions and best practices for using Clusters7D, see *TMO Guide*, [chapter 7](#).

1. 2.

**Screenshots: (1) Clusters7D with topic-clusters at left, and *topics* ranked at right; (2) Clusters7D with topic-clusters at left, and *words ranked* at right .**



### METHOD 3 - STEP 1 **Step 1 (of Clusters7D method) - Get an overview**

Freely explore your topic model with Clusters 7D using the the "best practices" for the tool suggested in the *TMO Guide*, [chapter 7](#). (See especially [section 1](#) for exploring the topics in topic-clusters; and [section 2](#) for exploring the words in topic-clusters.)

Take some notes for yourself about such features as the following:      What are the most obvious apparent clusters? Are these clusters characterized by large topics (heavily represented in the corpus)? Do these clusters also appear to show up in the "scaled" view of topics in Dfr-Browser and TopicBubbles and "intertopic distance" view in pyLDAvis, or in the dendrograms you explored using DendrogramViewer?      Using the capabilities of Clusters7D to explore what *other* topic-clusters a topic or word is associated with, are you able to gain a fuller *gestalt* ([WordNet definition of "gestalt"](#)) of the multi-level complexity of your topic model--like getting a sense of the relations between all the parts, great and small, of a biome? **Your notes (scroll to bottom to add notes related to Clusters7D):**

---

---

---

---

---



**METHOD 3 - STEP 2 Step 2 (of Clusters7D method) - Identify a cluster for study.**

Now single out one topic cluster that you want to study to address the research question you are working on. Examine some of the topics in the cluster, looking at their keywords and some of the articles associated with them.

---

**LABEL CLUSTER Provide a label for the cluster that describes its meaning** or revise a label you previously assigned using other methods of cluster analysis. (A label should be a simple phrase of between one and three words. Avoid making a label into a sentence predicating a specific relationship between terms, which often amounts to a thesis statement.)

Label for cluster you are singling out for study:

---

---

**TOPICS IN CLUSTER List below by their numbers a few topics that you think can serve as representatives of this cluster** (up to 10 topics)--or a revision of the representative topics you previously identified using other methods of cluster analysis.

	Answer 1 (1)	Answer 2 (2)	Answer 3 (3)	Answer 4 (4)	Answer 5 (5)	Answer 6 (6)	Answer 7 (7)	Answer 8 (8)	Answer 9 (9)	Answer 10 (10)
Topic #'s: (1)										

---

Page Break

**GO TO LAST STEPS This completes your use of Method 3 (Clusters7D), the last method of cluster analysis in this exercise. Pressing "next" will move you on to bringing everything together in a final observation.**

End of Block: METHOD 3: CLUSTERS7D

---

Start of Block: FINAL OBSERVATIONS

R exercise info [\\${R1: exercise info/QuestionText}](#)

---

R data you entered 1 [\\${R data you entered 1/QuestionText}](#)

---

R2 data you entered [\\${R data you entered 2/QuestionText}](#)

---

R data you entered 3

**Reminders - Data you entered above:**

The topic cluster you examined using **Method 3** (Clusters7D) is one you labeled: [\\${LABEL CLUSTER/ChoiceTextEntryValue}](#)

Topics you identified as representative of the above topic cluster are (by topic numbers):

[\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/1}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/2}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/3}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/4}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/5}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/6}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/7}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/8}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/9}](#), [\\${TOPICS IN CLUSTER%231/ChoiceTextEntryValue/1/10}](#)

---



R data you entered 3 **Reminders from your previous notes.**

(You can tinker with these notes here if you like to include a revised version in the final transcript of the output from this exercise. But it is likely more useful to draw on what you need from these prior notes and reweave them as needed in your final observations below.)

---

---

---

---

---



**FINAL OBSERVATIONS** Please write a final observation about clusters you looked at in this exercise that will help advance your understanding of your research question.

An “observation” should have two parts: a more-or-less objective description in which you have high confidence; and a thesis, hypothesis, interpretation, or suggestion for which you have less confidence before you study the model more closely but that might bear on your research question.

---

---

---

---

---

End of Block: FINAL OBSERVATIONS

---

Start of Block: RESEARCH TAKEAWAYS (v. 2, created 23 June 2019; rev. 27 June 2019)



**TAKEAWAYS Before you finish: Research Takeaways?**

After studying your topic model in this exercise, are there any research takeaways you want to record, store, describe, or log (depending on the nature of the takeaway) to provide key evidence, materials, and data that can be used in writing up a report on your research question

or for future humanities advocacy? For example, a research takeaway can be: Evidence in the form of an important topic(s) or relation between topic(s) and words, documents that you want to describe so that you can find it again; A screenshot of something you found while analyzing your topic model with one of the Topic Model Observatory visualization interfaces;

Citations for or excerpts from key documents associated with a topic (e.g., three newspaper articles you want to use in a report); A text copy of words in a topic ranked by proportional weight; Etc. If you have found such research takeaways, record, store, or describe them as appropriate in your team folder on the WE1S Google team drive. **Reminder:**

The folder for your team ([\\${EXERCISE INFO/ChoiceTextEntryValue/1}](#)) in the WE1S Google team drive is [\\${G DRIVE INFO/ChoiceTextEntryValue}](#)" target="\_blank">here. The main "AM Team Working Space" for all the teams in the WE1S Google team drive is [here](#). You may also want quickly to log or make a note about your takeaways in the field below in order to have a record of them associated with this survey. That way, when you read over your survey results, you can remember what you took away and where you stored or annotated it.

P.S. Some of the Topic Model Observatory interfaces may also allow you to use the [Hypothes.is](#) web annotation tool to highlight or annotated them in your browser for private or shared use.

---

---

---

---

---



### COMPANION Q's? **Companion questions?**

"Companion questions" may be defined as questions that help provide context for the research question you are currently addressing or that are follow-on questions you might want to pursue later. WE1S asks you periodically to consider if companion questions have occurred to you because it's a good way to prevent the kind of "tunnel vision" that research can sometimes create.

This is best explained by way of example. Suppose that your research question concerns the degree and nature of discussion about the "humanities crisis." A tunnel-vision view of the problem would be to consider only discussion of the "humanities crisis." A contextually wider view would consider companion questions such as the following: what is the degree and nature of discussion about *other* kinds of crisis by comparison? For example, in what other contexts

(political, economic, spiritual, etc.) does the word "crisis" tend to appear? Do mentions of "crisis" in all these contexts tend to spike up together, or be concentrated in certain kinds of media or sources? Etc.

Having advanced on your research question using the present module, have any "companion questions" occurred to you? **If so, please set them down here so that they can be remembered for possible future use:**

---



---



---



---



---

End of Block: RESEARCH TAKEAWAYS (v. 2, created 23 June 2019; rev. 27 June 2019)

Start of Block: SURVEY END (v.2, created 20 June 2019, rev. 28 June 2019)

CONFIDENCE?

**Final consideration -- Confidence assessment**

It is possible after conducting an interpretation exercise to conclude that you have either high or low confidence in the results (with confidence descending all the way down to a null result).

Please rank your level of confidence in the results of this interpretation exercise on the following scale:

	Null Result (no confidence) (1)	(4)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	Valid Result (full confidence) (13)
Confidence level: (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**SUGGESTIONS? Suggestions about this module?**

Do you have any suggestions about this module of the WE1S Interpretation Protocol (problems, improvements)?

---

---

---

---

---

---

**EMAIL A COPY? Before you conclude this module of the WE1S Interpretation Protocol, do you wish to have a report of your answers emailed to you?**

Choose "yes" if you have completed the module successfully as part of an interpretation project. The plain-text, HTML, and PDF copies of your survey should be kept as part of the documentation for answering a research question.

- Yes (4)
- No (5)

---

*Display This Question:*

*If Before you conclude this module of the WE1S Interpretation Protocol, do you wish to have a report... = Yes*



**EMAIL ADDRESS Please provide your email address to receive a copy of this survey.**

You will receive by email a text transcript of the questions and answers in this survey, plus links to HTML and PDF versions. You should keep these as documentation of your interpretation project. **Caution:** The online HTML and PDF versions linked in the email will expire after 31 days.)

Good practice is to copy the emailed transcript as text into a Google Doc in your team folder (in a subfolder called "Qualtrics survey transcripts") and also to download and store there the PDF version. Name the transcripts in the following format (for example): "Module3a (2019-07-11) -

history (Tarika)" (In this example, "history" has been added because it was the keyword that Module3a was used to explore; and "(Tarika)" was added to identify the specific person completing the survey if ithe whole research team did not do so.)

**Caution:** The online HTML and PDF versions ilinked in the email you receive will expire after 31 days.

**Your email:**

---

---

**END This is the end of the current module of the WE1S Interpretation Protocol.**

**IMPORTANT: To submit your answers and have Qualtrics mark your survey as completed, please press "Save à Next"**

End of Block: SURVEY END (v.2, created 20 June 2019, rev. 28 June 2019)

---