(v. 2) Module 2 - Examine representative topics (and clusters)

Survey Flow

ReferenceSurvey: (Ref. Survey) Info about this exercise

Block: WHAT THIS MODULE DOES (1 Question)

Standard: IDENTIFY THREE TIERS OF TOPICS (5 Questions)
Standard: EXAMINE HIGH-WEIGHT TOPICS (13 Questions)
Standard: EXAMINE MIDDLE-WEIGHT TOPICS (5 Questions)
Standard: EXAMINE LOW-WEIGHT TOPICS (5 Questions)

Standard: EXAMINE CLUSTERS (5 Questions)

Standard: SUMMARY OBSERVATIONS (4 Questions)

ReferenceSurvey: Ref. Survey Research Takeaways

ReferenceSurvey: (Ref. Survey) Survey end

Page Break

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

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.

	J	Ch	nec	k t	his	s fi	eld	ti b	ftł	nis	is	s jı	us	it a	a t	ria	al:	(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

Page Break				



_\		\sim 1 \sim E			
$\vdash x$	-	(110-	11) –	xercise	
		(, ,)	11 / L		: 11

The ID for the present exercise (generated from the information you just provided) is (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 very brief description of what you are currently doing (e.g., "We are now going to comkeywords 'humanities' and 'sciences' in the 20190621_2132_us-humanities-top-news topics200 model.")	npare the

TOPIC MODEL Topic model you are studying

	•	-	• •	g, please enter below its "topic model name" and "sta	
		-		e" looks like this: "20190621_2132_us-humanities-top-	
	spapers	•		You can locate this name in the Registry of WI	
				is the name of the specific granularity of the model you	
	lying (num		. ,	The "start page" is the URL of the overall top	
	•	•	•	and their visualizations) listed on this page on the Harb	
		•	•	ou can also find this URL in the Registry of WE1S Top	ic
	,		URL looks like		
this:	http://har	bor.engli	sh.ucsb.edu:1	0002/projects/20190621_2132_us-humanities-top-	
new	spapers/				
	O "Topic	model n	ame"l (1)		
	○ "Start	page" Uf	RL (2)		
RES	SEARCH (QUESTIC	ON Research	question you are addressing	
Б.					
				ut the research question you are addressing in this	
				on in the "Registry of WE1S Research Questions." (If y	
	•		of a model pr	rior to working on a specific research question, just ente	er
"Tak	king an ov	erview")			
	○ Resea	rch ques	stion (1)		-
	O Opera	tionalize	d form of ques	stion (2)	
-					
	O Posso	rob gues	tion ID (use th	ne format "team2-q1", representing in this example tear	m 2'c
f		•	•	ie format teamz-q1 , representing in this example tear	11 2 5
•		. 5/1 4400			_

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.

	O Step 1 - Module used (1)
	O Step 2 - Module used (2)
	O Step 3 - Module used (3)
	O Step 4 - Module used (4)
	O Step 5 - Module used (5)
	O Step 6 - Module used (6)
	O Step 7 - Module used (7)
	O Step 8 - Module used (8)
	O Step 9 - Module used (9)
	O Step 10 - Module used (10)
Pa	ge 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 2 does

(Module 2, created 19 June 2019; last rev. 20 June 2019.)

Module 2 of the WE1S topic model Interpretation Protocol guides you through examining representative topics and topic-clusters with more precision than similar exercises in Module 1 ("Take an overview of a model").

"Representative" means that you will be observing a range of topics/clusters from the important to those that are minor or outliers. Together with Module 1, Module2 gives you a good initial sense of a model as context for addressing specific research questions.

End of Block: WHAT THIS MODULE DOES

Start of Block: IDENTIFY THREE TIERS OF TOPICS

R1 exercise info \${R1: exercise info/QuestionText}

STEP 1

Step 1. Identify selected topics at three levels of proportional weight

This part of this module asks you to identify topics at three levels of proportional weight in the model. "Proportional weight" means basically how important the topic is, or how often it turns up in the corpus of documents. (A bit more exactly, proportionial weight refers to the percentage of the total number of words ["tokens"] in the corpus that the modelling process has assigned to a particular topic. It is also sometimes thought of as "conditional probability," referring to how likely it is that a topic will be encountered in a model based on a particular corpus of documents.)

Why look for three levels of topics by weight?

Typically, topics in models fall into several levels of proportional weight, which you might imagine as a set of terraces. For example, there will be a high terrace consisting of several topics at extremely high proportional weight (whose percentages are also often near each other). There are often several kinds of such big topics (explained below). Then there might be a noticeable drop-off to the next terrace below, where there is a series of topics whose percentages are in the middle range (and also often near each other in level). Then, beneath there will be lower terraces. Often, you will recognize the difference between terraces by the fact that, for example, one set of topics will have percentages ranging from 2.1 to 2.4; then there will be a noticeable shift in range to the next set with percentages, for example, of 1.5-1.8.

Methods of finding and sorting topics by weight: Use the general-purpose visualization interfaces in the WE1S Topic Model Observatory as follows (see <u>TMO Guide</u> for more detailed how-to's): In Dfr-Browser, go to the "List" tab under "Overview" and sort the topics in descending order of "proportion of corpus." To sort, click on the label "proportion of corpus" or the small arrows next to that label at the right. Each topic's percentage figure (its proportion of the corpus) shows at the right next to a blue bar graph roughly visualizing that percentage.

In pyLDAvis, look at the relative size of the topic circles in the left-hand "Intertopic Distance Map" panel. Click on any individual topic circle to show its proportion of the corpus (e.g., "5.6% of tokens") at the top of the right-hand panel of pyLDAvis.

The size of the topic circles in Topic Bubbles serves a similar purpose

1. 2.

Screenshot 1: pyLDAvis

Screenshot 2: Topic Bubbles

For the purpose of this module, select for examination topics from three levels by proportional weight (high, middle, low) according to the instructions in the following sequence of questions.

NTER HIGH TOPICS Enter in ranked order the topic number "high" proportional weight level (up to a maximum of 10	
O Topic # of high-weight topic (highest weight) (1)	
O Topic # of high-weight topic (2)	
O Topic # of high-weight topic (3)	
O Topic # of high-weight topic (4)	
O Topic # of high-weight topic (5)	
O Topic # of high-weight topic (6)	
O Topic # of high-weight topic (7)	
O Topic # of high-weight topic (8)	
O Topic # of high-weight topic (9)	
O Topic # of high-weight topic (10)	

"middle"-weight topics that interest you or that seem represental numbers below:	ive and enter their topic
O Topic # of middle-weight topic (1)	
O Topic # of middle-weight topic (2)	
O Topic # of middle-weight topic (3)	
O Topic # of middle-weight topic (4)	
O Topic # of middle-weight topic (5)	
ENTER LOW TOPICS Now in the Dfr-Browser topic "list" view (stopics), jump down to the very bottom of the list to identify the lo Enter the topic numbers for up to 5 low-weight topics at that level Topic # of low-weight topic (1)	w topics by proportional weight.
O Topic # of low-weight topic (1)	
O Topic # of low-weight topic (2)	
O Topic # of low-weight topic (3)	
O Topic # of low-weight topic (4)	
O Topic # of low-weight topic (5)	

Start of Block: EXAMINE HIGH-WEIGHT TOPICS

ENTER MIDDLE TOPICS Identify the level of topics that seems to be one level down, which for

the purpose of this exercise we will call the "middle"-weight level. Choose up to 5 such

R1 exercise info	\${R1:	exercise	info/Q	<i>uestionTe</i>	xť
------------------	--------	----------	--------	------------------	----

R data you entered

Reminders - Data you entered above:

High-weight topics you identified: \$\{\text{ENTER HIGH}\}

TOPICS/ChoiceGroup/AllChoicesTextEntry} Middle-wieght topics you identified:

\${ENTER MIDDLE TOPICS/ChoiceGroup/AllChoicesTextEntry} Low-weight topics

you identified: \${ENTER LOW TOPICS/ChoiceGroup/AllChoicesTextEntry}

STEP 2 Step 2. Examine high-weight topics

Topics with very high proportional weight in topic models are often hard to interpret and require special assessment procedures. If possible, separate the 10 high-weight topics you identified above into three classes, which can be called "super-topics," "meta-topics," and "understandable top topics."

A **super-topic** is typically an extremely high-weight topic characterized by words that are hard to understand as a cohesive group and often seem to represent a general, existential, daily, ordinary, or broadly common stratum of social, spatial, and temporal experience (e.g., "day back night man year told home asked room days morning high make hard made"). A super-topic is hard to understand even after sampling some of the documents associated with the topic.

By contrast, a **meta-topic** is typically characterized by topic words that--while also hard to understand as a cohesive group (e.g., containing the same kinds of existential social, space, and time words)--may represent categories or registers of discourse: i.e., the discourses of different social classes, media types, genres, or even sections of newspapers [like "Entertainment" or "Sports"]). For instance, if words like "lol" or "sucks" are top words in such topics, then there is a good chance that the topics are *discourse types* rather than *themes* per se related to students, popular culture, Reddit, etc. Looking at documents associated with the topic can give a better sense of these broad kinds or classes of language.

An **understandable top topic** is distinguished from both super- and meta-topics by being a high-weight topic whose topic words (and associated documents) seem to cohere around understandable themes or issues that you can fairly easily label in a short phrase (e.g., "higher education programs" or "business and careers").

PARSE HIGH TOPICS

For each of the 10 **high-weight** topics you previously identified please examine the topic as follows:

Look at the 10 most frequent words in the topic. Look quickly at the top 10 articles associated with the topic (and read a bit more closely the top 5 articles). Then try to separate out "super-topics" from "meta-topics" from "understandable top topics," and record their topic numbers below. (Not all fields for each kind of topic need to be filled.)

Reminder: High-weight topics you listed in ranked order by weight: TOPICS/ChoiceGroup/AllChoicesTextEntry}	
SUPER-TOPICS Super-topics : (Not all fields need to be filled.)	
O # (1)	
O # (2)	
O # (3)	
O # (4)	
O # (5)	
O # (6)	
O # (7)	
O # (8)	
O # (9)	
O # (10)	

(Not all fields need to be filled.)	
O # (1)	
O # (2)	
O # (3)	
O # (4)	
O # (5)	
O # (6)	
O # (7)	
O # (8)	
O # (9)	
(4.0)	

META-TOPICS Meta-topics:

(1) # (2) # (3) # (4) # (5) # (6) # (7) # (8) # (9) # (10)

UNDERSTANDABL TOPICS Understandable top topics:

(Not all fields need to be filled.)

Page Break —

reminders.block4 \${QID136239283/QuestionText}	
OBSERVE HIGH TOPICS Now that you have sorted the high-weight topics into the categories of "super-topics," "meta-topics," and "understandeable top topics," you make some observations about the latter two categories. (For this purpose, we will now put aside and ignore the "super-topics" unless you s	
anything interesting there, in which case add a comment in the Running Note below	
\odot	
OBSERVE META-TOPICS Observation on meta-topics : write a brief observation of the " meta-topics ," (either generalizing about all the ones you found or focusing on a single example).	
Reminder: meta-topics that you identified: \${META-TOPICS/ChoiceGroup/AllChoicesTextEntry} An "observation" should have two parts: a more-or-less objective description in which you high confidence; and a thesis, hypothesis, interpretation, or suggestion for which you have confidence before you study the model more closely but that might suggest avenues of research.	
·	

LABEL META-TOPICS **Label meta-topics:** Try to label the meta-topics you have discovered. (A label should be a word or short phrase representing only the objective part of your observation, or what the topic is "about". It should not be formed as a sentence, which tends to posit a thesis.)

		(1)
Meta-topic you identified	\${META-TOPICS/ChoiceTextEntryValue/2}	(2)
Meta-topic you identified	: \${META-TOPICS/ChoiceTextEntryValue/3}	(3)
Meta-topic you identified	: \${META-TOPICS/ChoiceTextEntryValue/4}	(4)
O Meta-topic you identified	\${META-TOPICS/ChoiceTextEntryValue/5}	(5)
Meta-topic you identified	\${META-TOPICS/ChoiceTextEntryValue/6}	(6)
O Meta-topic you identified	\${META-TOPICS/ChoiceTextEntryValue/7}	(7)
O Meta-topic you identified	: \${META-TOPICS/ChoiceTextEntryValue/8}	(8)
O Meta-topic you identified	\${META-TOPICS/ChoiceTextEntryValue/9}	(9)
Meta-topic you identified	: \${META-TOPICS/ChoiceTextEntryValue/10	} (10)

Page Break —



OBSERVE UNDERSTANDAB **Observation on understandable top topics**: write a brief observation of the "**understandable top topics**," (either generalizing about all the ones you found or focusing on a single example).

Reminder: understandable top topics that you identified: \${UNDERSTANDABL TOPICS/ChoiceGroup/AllChoicesTextEntry}}

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 suggest avenues of research.

LABEL UNDERSTANDABLE **Label understandable top topics:** Try to label the understandaable top topics you have discovered. (A label should be a word or short phrase representing only the objective part of your observation, or what the topic is "about". It should not be formed as a sentence, which tends to posit a thesis.)

Understandable top topic you identified:\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/1} (1)
Understandable top topic you identified:\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/2} (2)
Understandable top topic you identified:\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/3} (3)
Understandable top topic you identified:\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/4} (4)
Understandable top topic you identified:\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/5} (5)
Understandable top topic you identified:\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/6} (6)
Understandable top topic you identified:\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/7} (7)
Understandable top topic you identified:\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/8} (8)
Understandable top topic you identified:\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/9} (9)
Understandable top topic you identified:\(\frac{\{\text{UNDERSTANDABL}}{\text{TOPICS/ChoiceTextEntryValue/10}\}\) (10)

End of Block: EXAMINE HIGH-WEIGHT TOPICS

Start of Block: EXAMINE MIDDLE-WEIGHT TOPICS

LABEL UNDERSTANDABLE \${R1: exercise info/QuestionText}

R data you entered

Reminders - Data you entered above:

High-weight topics you identified: "Super-topics": \${SUPER-

TOPICS/ChoiceGroup/AllChoicesTextEntry} "Meta-topics": #\${META-

TOPICS/ChoiceTextEntryValue/1} (\${LABEL META-

TOPICS/ChoiceTextEntryValue/1), #\${META-TOPICS/ChoiceTextEntryValue/2} (\${LABEL

META-TOPICS/ChoiceTextEntryValue/2}), #\${META-TOPICS/ChoiceTextEntryValue/3}

(\${LABEL META-TOPICS/ChoiceTextEntryValue/3}), #\${META-

TOPICS/ChoiceTextEntryValue/4} (\${LABEL META-

TOPICS/ChoiceTextEntryValue/4}), #\${META-TOPICS/ChoiceTextEntryValue/5} (\${LABEL

META-TOPICS/ChoiceTextEntryValue/5}), #\${META-TOPICS/ChoiceTextEntryValue/6}

(\${LABEL META-TOPICS/ChoiceTextEntryValue/6}), #\${META-

TOPICS/ChoiceTextEntryValue/7} (\${LABEL META-

TOPICS/ChoiceTextEntryValue/7}), #\${META-TOPICS/ChoiceTextEntryValue/8} (\${LABEL

META-TOPICS/ChoiceTextEntryValue/8}), #\${META-TOPICS/ChoiceTextEntryValue/9}

(\${LABEL META-TOPICS/ChoiceTextEntryValue/9}), #\${META-

TOPICS/ChoiceTextEntryValue/10} (\${LABEL META-TOPICS/ChoiceTextEntryValue/10})

"Understandable top topics": #\${UNDERSTANDABL

TOPICS/ChoiceTextEntryValue/1) (\${LABEL

UNDERSTANDABLE/ChoiceTextEntryValue/1}), #\${UNDERSTANDABL

TOPICS/ChoiceTextEntryValue/2) (\${LABEL

UNDERSTANDABLE/ChoiceTextEntryValue/2}), #\${UNDERSTANDABL

TOPICS/ChoiceTextEntryValue/3} (\${LABEL

UNDERSTANDABLE/ChoiceTextEntryValue/3}), #\${UNDERSTANDABL

TOPICS/ChoiceTextEntryValue/4} (\${LABEL

UNDERSTANDABLE/ChoiceTextEntryValue/4}), #\${UNDERSTANDABL

TOPICS/ChoiceTextEntryValue/5} (\${LABEL

UNDERSTANDABLE/ChoiceTextEntryValue/5}), #\${UNDERSTANDABL

TOPICS/ChoiceTextEntryValue/6} (\${LABEL

UNDERSTANDABLE/ChoiceTextEntryValue/6}), #\${UNDERSTANDABL

TOPICS/ChoiceTextEntryValue/7} (\${LABEL

UNDERSTANDABLE/ChoiceTextEntryValue/7}), #\${UNDERSTANDABL

TOPICS/ChoiceTextEntryValue/8} (\${LABEL

UNDERSTANDABLE/ChoiceTextEntryValue/8}), #\${UNDERSTANDABL

TOPICS/ChoiceTextEntryValue/9} (\${LABEL

UNDERSTANDABLE/ChoiceTextEntryValue/9}), #\${UNDERSTANDABL

TOPICS/ChoiceTextEntryValue/10} (\${LABEL UNDERSTANDABLE/ChoiceTextEntryValue/10})

Middle-weight topics you identified: \${ENTER MIDDLE TOPICS/ChoiceGroup/AllChoicesTextEntry}
Low-weight topics you identified: \${ENTER LOW TOPICS/ChoiceGroup/AllChoicesTextEntry}
STEP 3 Step 3. Examine middle-weight topics
Next, let's look at the middle-weight topics you identified.
o
OBSERVE MIDDLEWEIGHT Observation on middle-weight topics : write a brief observation of the middle-weight topics you identified (either generalizing about all the ones you found or focusing on a single example).
Reminder. middle-weight topics that you identified: \${ENTER MIDDLE TOPICS/ChoiceGroup/AllChoicesTextEntry}
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 suggest avenues of research.

LABEL MIDDLEWEIGHT Label middle-weight topics: Try to label the middle-weight topics you have discovered. (A label should be a word or short phrase representing only the objective part of your observation, or what the topic is "about". It should not be formed as a sentence, which tends to posit a thesis.) Middle-weight topic you identified: topic #\${ENTER MIDDLE TOPICS/ChoiceTextEntryValue/1} (1) Middle-weight topic you identified: topic #\${ENTER MIDDLE TOPICS/ChoiceTextEntryValue/2) (2) Middle-weight topic you identified: topic #\${ENTER MIDDLE TOPICS/ChoiceTextEntryValue/3} (3) Middle-weight topic you identified: topic #\${ENTER MIDDLE TOPICS/ChoiceTextEntryValue/4} (4) Middle-weight topic you identified: topic #\${ENTER MIDDLE TOPICS/ChoiceTextEntryValue/5} (5) End of Block: EXAMINE MIDDLE-WEIGHT TOPICS Start of Block: EXAMINE LOW-WEIGHT TOPICS R1 exercise info \${R1: exercise info/QuestionText} R data you entered Reminders - Data you entered above: **High-weight topics you identified**: "Supertopics": \${SUPER-TOPICS/ChoiceGroup/AllChoicesTextEntry} "Meta-topics": #\${META-TOPICS/ChoiceTextEntryValue/1} (\${LABEL META-TOPICS/ChoiceTextEntryValue/1), #\${META-TOPICS/ChoiceTextEntryValue/2} (\${LABEL} META-TOPICS/ChoiceTextEntryValue/2}), #\${META-TOPICS/ChoiceTextEntryValue/3} (\${LABEL META-TOPICS/ChoiceTextEntryValue/3}), #\${META-TOPICS/ChoiceTextEntryValue/4} (\${LABEL META-TOPICS/ChoiceTextEntryValue/4}), #\${META-TOPICS/ChoiceTextEntryValue/5} (\${LABEL}

META-TOPICS/ChoiceTextEntryValue/5}), #\${META-TOPICS/ChoiceTextEntryValue/6}

(\${LABEL META-TOPICS/ChoiceTextEntryValue/6}), #\${META-

TOPICS/ChoiceTextEntryValue/7} (\${LABEL META-TOPICS/ChoiceTextEntryValue/7}), #\${META-TOPICS/ChoiceTextEntryValue/8} (\${LABEL META-TOPICS/ChoiceTextEntryValue/8}), #\${META-TOPICS/ChoiceTextEntryValue/9} (\${LABEL META-TOPICS/ChoiceTextEntryValue/9}), #\${META-TOPICS/ChoiceTextEntryValue/10} (\${LABEL META-TOPICS/ChoiceTextEntryValue/10}) "Understandable top topics": #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/1) (\${LABEL UNDERSTANDABLE/ChoiceTextEntryValue/1}), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/2) (\${LABEL UNDERSTANDABLE/ChoiceTextEntryValue/2)), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/3} (\${LABEL} UNDERSTANDABLE/ChoiceTextEntryValue/3}), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/4} (\${LABEL UNDERSTANDABLE/ChoiceTextEntryValue/4}), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/5} (\${LABEL} UNDERSTANDABLE/ChoiceTextEntryValue/5)), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/6} (\${LABEL} UNDERSTANDABLE/ChoiceTextEntryValue/6}), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/7} (\${LABEL UNDERSTANDABLE/ChoiceTextEntryValue/7}), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/8} (\${LABEL UNDERSTANDABLE/ChoiceTextEntryValue/8)), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/9} (\${LABEL UNDERSTANDABLE/ChoiceTextEntryValue/9)), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/10} (\${LABEL UNDERSTANDABLE/ChoiceTextEntryValue/10}) Middle-weight topics you identified: #\${ENTER MIDDLE TOPICS/ChoiceTextEntryValue/1} (\${LABEL MIDDLEWEIGHT/ChoiceTextEntryValue/1}), #\${ENTER MIDDLE TOPICS/ChoiceTextEntryValue/2} (\${LABEL MIDDLEWEIGHT/ChoiceTextEntryValue/2}), #\${ENTER MIDDLE TOPICS/ChoiceTextEntryValue/3} (\${LABEL MIDDLEWEIGHT/ChoiceTextEntryValue/3}), #\${ENTER MIDDLE TOPICS/ChoiceTextEntryValue/4} (\${LABEL MIDDLEWEIGHT/ChoiceTextEntryValue/4}), #\${ENTER MIDDLE TOPICS/ChoiceTextEntryValue/5} (\${LABEL MIDDLEWEIGHT/ChoiceTextEntryValue/5}) Low-weight topics you identified: \${ENTER LOW TOPICS/ChoiceGroup/AllChoicesTextEntry} STEP 4 Step 4. Examine low-weight topics Next, let's look at the low-weight topics you identified.



OBSERVE LOW-WEIGHT **Observation on low-weight topics**: write a brief observation of the low-weight topics you identified (either generalizing about all the ones you found or focusing on a single example).

Reminder: low-weight topics that you identified: \${ENTER LOW}
TOPICS/ChoiceGroup/AllChoicesTextEntry}
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 suggest avenues of research.

LABEL LOW-WEIGHT **Label low-weight topics**: Try to label the low-weight topics you have discovered. (A label should be a word or short phrase representing only the objective part of your observation, or what the topic is "about". It should not be formed as a sentence, which tends to posit a thesis.)

Cow-weight topic you identified: topic #\${ENTER LOW TOPICS/ChoiceTextEntryValue/1} (1)	
Low-weight topic you identified: topic #\${ENTER LOW TOPICS/ChoiceTextEntryValue/2} (2)	
Cow-weight topic you identified: topic #\${ENTER LOW TOPICS/ChoiceTextEntryValue/3} (3)	
Low-weight topic you identified: topic #\${ENTER LOW TOPICS/ChoiceTextEntryValue/4} (4)	
Low-weight topic you identified: topic #\${ENTER LOW TOPICS/ChoiceTextEntryValue/5} (5)	
End of Block: EXAMINE LOW-WEIGHT TOPICS	
Start of Block: EXAMINE CLUSTERS	
R1 exercise info \${R1: exercise info/QuestionText}	
R data you entered Reminders - Data you entered above: High-weight topics you identified: "Super-topics": \${STOPICS/ChoiceGroup/AllChoicesTextEntry} "Meta-topicsTOPICS/ChoiceTextEntryValue/1} (\${LABEL META-TOPICS/ChoiceTextEntryValue/1}), #\${META-TOPICS/ChoiceTextEntryValue/2}), #\${META-TOPICS/ChoiceTextEntryValue/2}), #\${META-TOPICS/ChoiceTextEntryValue/3}), #\${META-TOPICS/ChoiceTextEntryValue/3}), #\${META-TOPICS/ChoiceTextEntryValue/3}), #\${META-TOPICS/ChoiceTextEntryValue/4} (\${LABEL META-TOPICS/ChoiceTextEntryValue/4})	s": #\${META- extEntryValue/2} (\${LABEL noiceTextEntryValue/3}
TOPICS/ChoiceTextEntryValue/4}), #\${META-TOPICS/ChoiceTextEntryValue/5}), #\${META-TOPICS/ChoiceTextEntryValue/5}), #\${META-TOPICS/ChoiceTextEntryValue/5}), #\${META-TOPICS/ChoiceTextEntryValue/5}), #\${META-TOPICS/ChoiceTextEntryValue/5}}	

```
(${LABEL META-TOPICS/ChoiceTextEntryValue/6}), #${META-
TOPICS/ChoiceTextEntryValue/7} (${LABEL META-
TOPICS/ChoiceTextEntryValue/7)), #${META-TOPICS/ChoiceTextEntryValue/8} (${LABEL
META-TOPICS/ChoiceTextEntryValue/8}), #${META-TOPICS/ChoiceTextEntryValue/9}
(${LABEL META-TOPICS/ChoiceTextEntryValue/9}), #${META-
TOPICS/ChoiceTextEntryValue/10} (${LABEL META-TOPICS/ChoiceTextEntryValue/10})
      "Understandable top topics": #${UNDERSTANDABL
TOPICS/ChoiceTextEntryValue/1) (${LABEL
UNDERSTANDABLE/ChoiceTextEntryValue/1}), #${UNDERSTANDABL
TOPICS/ChoiceTextEntryValue/2} (${LABEL
UNDERSTANDABLE/ChoiceTextEntryValue/2}), #${UNDERSTANDABL
TOPICS/ChoiceTextEntryValue/3} (${LABEL
UNDERSTANDABLE/ChoiceTextEntryValue/3}), #${UNDERSTANDABL
TOPICS/ChoiceTextEntryValue/4} (${LABEL}
UNDERSTANDABLE/ChoiceTextEntryValue/4), #${UNDERSTANDABL
TOPICS/ChoiceTextEntryValue/5} (${LABEL}
UNDERSTANDABLE/ChoiceTextEntryValue/5}), #${UNDERSTANDABL
TOPICS/ChoiceTextEntryValue/6} (${LABEL
UNDERSTANDABLE/ChoiceTextEntryValue/6)), #${UNDERSTANDABL
TOPICS/ChoiceTextEntryValue/7} (${LABEL
UNDERSTANDABLE/ChoiceTextEntryValue/7}), #${UNDERSTANDABL
TOPICS/ChoiceTextEntryValue/8} (${LABEL
UNDERSTANDABLE/ChoiceTextEntryValue/8}), #${UNDERSTANDABL
TOPICS/ChoiceTextEntryValue/9} (${LABEL
UNDERSTANDABLE/ChoiceTextEntryValue/9}), #${UNDERSTANDABL
TOPICS/ChoiceTextEntryValue/10} (${LABEL UNDERSTANDABLE/ChoiceTextEntryValue/10})
Middle-weight topics you identified: #${ENTER MIDDLE
TOPICS/ChoiceTextEntryValue/1} (${LABEL MIDDLEWEIGHT/ChoiceTextEntryValue/1}),
#${ENTER MIDDLE TOPICS/ChoiceTextEntryValue/2} (${LABEL
MIDDLEWEIGHT/ChoiceTextEntryValue/2}), #${ENTER MIDDLE
TOPICS/ChoiceTextEntryValue/3} (${LABEL MIDDLEWEIGHT/ChoiceTextEntryValue/3}),
#${ENTER MIDDLE TOPICS/ChoiceTextEntryValue/4} (${LABEL
MIDDLEWEIGHT/ChoiceTextEntryValue/4}), #${ENTER MIDDLE
TOPICS/ChoiceTextEntryValue/5) (${LABEL MIDDLEWEIGHT/ChoiceTextEntryValue/5})
Low-weight topics you identified: #${ENTER LOW TOPICS/ChoiceTextEntryValue/1}
(${LABEL LOW-WEIGHT/ChoiceTextEntryValue/1}), #${ENTER LOW
TOPICS/ChoiceTextEntryValue/2) (${LABEL LOW-WEIGHT/ChoiceTextEntryValue/2}).
#${ENTER LOW TOPICS/ChoiceTextEntryValue/3} (${LABEL LOW-
WEIGHT/ChoiceTextEntryValue/3}), #${ENTER LOW TOPICS/ChoiceTextEntryValue/4}
(${LABEL LOW-WEIGHT/ChoiceTextEntryValue/4}, #${ENTER LOW
TOPICS/ChoiceTextEntryValue/5) (${LABEL LOW-WEIGHT/ChoiceTextEntryValue/5})
```

STEP 5

Step 5. Examine clusters of topics in the model

Next, let's see if we can recognize any meaningul "clusters" of topics (topics that are statistically "near" each other in the words and documents associated with them and "far" or separated from other clusters).

Using the specialized "clustering analysis" tools named Clusters7D and DendrogramViewer in the WE1S Topic Model Observatory (see <u>TMO Guide</u> for how-to's and best practices), see if you can identify clusters of topics in the model. A checklist of things you might want to look for:

Topic-clusters in Clusters7D that bulk large in the model (respresented as larger circles in the left panel)

Topic-clusters in DendrogramViewer with many "leaves" in a "clade"

Topic-clusters that seem clearly separated from others *Important methodological note*: The primary value of topic models lies in the topics they surface. You should always have less confidence in the existence and meaningfulness of topic "clusters" shown in visualization interfaces representing in a 2D-graph space the statistical "nearness" of topics. (Clusters7D and DendrogramViewer are more specialized and trustworthy for this task; the "scaled" views in Dfrbrowser or TopicBubbles and "intertopic distance map" view in pyLDAvis are less so.) The reason is that by definition "topics" in a topic model are already a *reduction* of the complexity of the underlying corpus of documents (they reduce the myriad relations of thought and language in those documents to a much smaller number of "topics"). "Clusters" of topics are then a *second-level reduction* or *reduction-of-a-reduction* (they reduce the topics into a much smaller set). With each step in the reduction series, we gain meaningfulness("these documents are 'about' *this*") at the cost of confidence ("does that cluster really exist?").

If you find a topic-cluster that is really interesting or important for your research question, there are a number of things you can do to increase your confidence in the cluster--e.g., look at a low-granularity version of your topic model (say, 50 topics instead of 250) and see if a topic-cluster in the higher-granularity model resolves into a single topic in the lower-granularity model.

(WE1S Interpretation Protocol modules 3.b and 4.b include enhanced, more precise exercises in topic-cluster analysis that try to optimize the interpretive usefulness and confidence-level of "clusters.")

Once you have looked over the possbile clusters in the model, answer the following sequence of questions.

•

or interesting clusters of topics you see. 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 suggest avenues of research.
LABEL CLUSTERS Label topic clusters : Label up to three of the interesting or major clusters you observed above. ((A label should be a word or short phrase representing only the objective part of your observation, or what the topic is "about". It should not be formed as a sentence, which tends to posit a thesis.)
O Cluster 1 (1)
O Cluster 2 (2)
O Cluster 3 (3)
End of Block: EXAMINE CLUSTERS
Start of Block: SUMMARY OBSERVATIONS
R1 exercise info \${R1: exercise info/QuestionText}
R data you entered Reminders - Data you entered above: High-weight topics you identified: "Super-topics": \${SUPER-TOPICS/ChoiceGroup/AllChoicesTextEntry} "Meta-topics": #\${META-TOPICS/ChoiceTextEntryValue/1} (\${LABEL META-TOPICS/ChoiceTextEntryValue/1}), #\${META-TOPICS/ChoiceTextEntryValue/2} (\${LABEL

OBSERVE CLUSTERS Observation on topic clusters: Write a brief observation about major

META-TOPICS/ChoiceTextEntryValue/2}), #\${META-TOPICS/ChoiceTextEntryValue/3} (\${LABEL META-TOPICS/ChoiceTextEntryValue/3}), #\${META-TOPICS/ChoiceTextEntryValue/4} (\${LABEL META-TOPICS/ChoiceTextEntryValue/4}), #\${META-TOPICS/ChoiceTextEntryValue/5} (\${LABEL} META-TOPICS/ChoiceTextEntryValue/5}), #\${META-TOPICS/ChoiceTextEntryValue/6} (\${LABEL META-TOPICS/ChoiceTextEntryValue/6}), #\${META-TOPICS/ChoiceTextEntryValue/7} (\${LABEL META-TOPICS/ChoiceTextEntryValue/7), #\${META-TOPICS/ChoiceTextEntryValue/8} (\${LABEL} META-TOPICS/ChoiceTextEntryValue/8}), #\${META-TOPICS/ChoiceTextEntryValue/9} (\${LABEL META-TOPICS/ChoiceTextEntryValue/9}), #\${META-TOPICS/ChoiceTextEntryValue/10} (\${LABEL META-TOPICS/ChoiceTextEntryValue/10}) "Understandable top topics": #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/1} (\${LABEL UNDERSTANDABLE/ChoiceTextEntryValue/1}), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/2} (\${LABEL} UNDERSTANDABLE/ChoiceTextEntryValue/2}), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/3} (\${LABEL} UNDERSTANDABLE/ChoiceTextEntryValue/3}), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/4} (\${LABEL UNDERSTANDABLE/ChoiceTextEntryValue/4}), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/5} (\${LABEL} UNDERSTANDABLE/ChoiceTextEntryValue/5)), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/6} (\${LABEL} UNDERSTANDABLE/ChoiceTextEntryValue/6}), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/7} (\${LABEL UNDERSTANDABLE/ChoiceTextEntryValue/7}), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/8} (\${LABEL UNDERSTANDABLE/ChoiceTextEntryValue/8}), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/9} (\${LABEL UNDERSTANDABLE/ChoiceTextEntryValue/9}), #\${UNDERSTANDABL TOPICS/ChoiceTextEntryValue/10} (\${LABEL UNDERSTANDABLE/ChoiceTextEntryValue/10}) Middle-weight topics you identified: #\${ENTER MIDDLE TOPICS/ChoiceTextEntryValue/1} (\${LABEL MIDDLEWEIGHT/ChoiceTextEntryValue/1}), #\${ENTER MIDDLE TOPICS/ChoiceTextEntryValue/2} (\${LABEL MIDDLEWEIGHT/ChoiceTextEntryValue/2}), #\${ENTER MIDDLE TOPICS/ChoiceTextEntryValue/3} (\${LABEL MIDDLEWEIGHT/ChoiceTextEntryValue/3}), #\${ENTER MIDDLE TOPICS/ChoiceTextEntryValue/4} (\${LABEL MIDDLEWEIGHT/ChoiceTextEntryValue/4}), #\${ENTER MIDDLE TOPICS/ChoiceTextEntryValue/5} (\${LABEL MIDDLEWEIGHT/ChoiceTextEntryValue/5}) Low-weight topics you identified: #\${ENTER LOW TOPICS/ChoiceTextEntryValue/1}

Low-weight topics you identified: #\${ENTER LOW TOPICS/ChoiceTextEntryValue/1} (\${LABEL LOW-WEIGHT/ChoiceTextEntryValue/1}), #\${ENTER LOW TOPICS/ChoiceTextEntryValue/2} (\${LABEL LOW-WEIGHT/ChoiceTextEntryValue/2}), #\${ENTER LOW TOPICS/ChoiceTextEntryValue/3} (\${LABEL LOW-

(\${LABEL LOW-WEIGHT/ChoiceTextEntryValue/4}, #\${ENTER LOW TOPICS/ChoiceTextEntryValue/5} (\${LABEL LOW-WEIGHT/ChoiceTextEntryValue/5}) Cluster of topics you identified: Cluster 1 (\${LABEL CLUSTERS/ChoiceTextEntryValue/1}), Cluster 2 (\${LABEL CLUSTERS/ChoiceTextEntryValue/2}), Cluster 3 (\${LABEL CLUSTERS/ChoiceTextEntryValue/3}) RUNNING NOTES END Your Running Notes from this exercise: step 5 Step 5. Summary observations to complete this module Now that you have conducted the exercises in this Module 2 for exploring representative topics and clusters, write an observation that synthesizes your understanding of this topic model so 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: SUMMARY OBSERVATIONS

WEIGHT/ChoiceTextEntryValue/3}), #\${ENTER LOW TOPICS/ChoiceTextEntryValue/4}



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 iin 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.

Hypo	Some of the Topic Model Observatory interfaces may also allow you to use the othes.is web annotation tool to highlight or annotated them in your browser for pred use.	
_		
-		
-		
-		
0		

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.

sources? Et	C.									
questions" o	anced on your recurred to you? d for possible	If so,	please		•	•			•	•
										- -
										_
End of Bloc	k: RESEARCH	TAKE	AWAY	/S (v. 2	2, crea	ted 23	June 2	019; re\	/. 27 Ju	_ ine 2019)
Start of Blo	ck: SURVEY E	ND (v.2	2, crea	ited 20	June	2019 , i	rev. 28	June 20	119)	
It is possible	CE? deration Core after conducting nce in the result	ıg an in	terpret	tation e	exercise			•		•
Please rank scale:	your level of co	nfidend	ce in th	ne resu	lts of th	nis inte	rpretatio	n exerc	ise on t	he following
	Null Result (no confidence) (1)	(4)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	Valid Result (full confidence) (13)
Confidence level: (1)	0	C	(C	C	(0	0	0	\circ

SUGGESTIONS? Suggestions about this module?								
Do yo 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)								
O 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.

Yo	ur 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)