Using TDM Studios to help systematically analyze research in music

Colloquium on Text & Data Mining in Libraries 2023

UNIVERSITY OF TORONTO LIBRARIES | MUSIC LIBRARY

James Mason, Metadata and Digital Initiatives Librarian
May 3rd, 2023

Welcome to TDM Studio

Workbench Dashboard

WHAT TO EXPECT

- Requires basic coding in Python or R and text analysis
- Provides options to select content and create datasets
- Offers Jupyter Notebook coding platform

Predict Labels

By using SBERT model and the array of parsed sentences, we will now output a corresponding array where each element is a tuple of the predicted label and an array of the raw possibilities for each label.

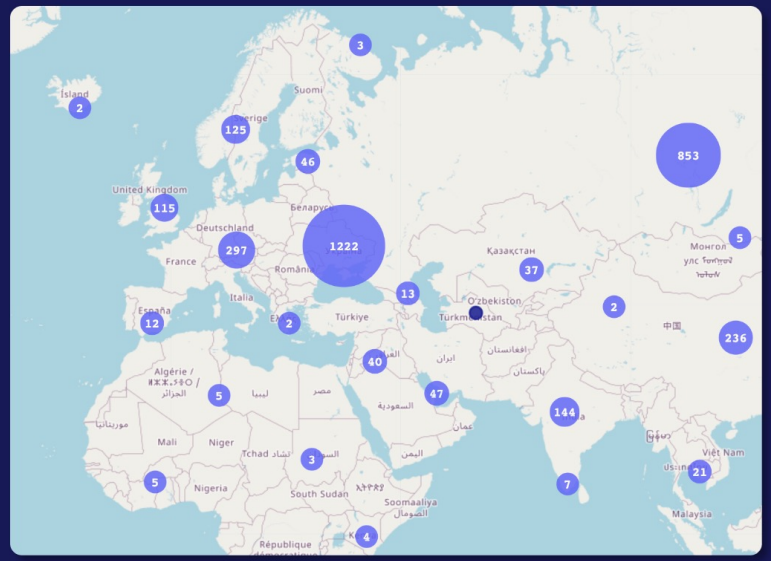
Sample code for text analysis

```
# Set sbert_path to location of SBERT model
sbert_path = 'Overall_emotion_classifier/nli-mpnet-base-v2'
transformer = ST(sbert_path)

# Encode the parsed_sents array
sentence_embeddings = transformer.encode(parsed_sents, show_progress_bar=False)

# Both scaler and sentiment_model should exist before running this cell
if scaler is not None and sentiment_model is not None:
    standardized = scaler.transform(sentence_embeddings)

    y_pred_numeric = sentiment_model.predict(standardized)
    y_pred_string = label_encoder.inverse_transform(y_pred_numeric)

    # Call the predict function on our sentences
    raw_predictions = sentiment_model.predict_proba(standardized)

    results = list(zip(y_pred_string, raw_predictions))

    # Print first element of array
    print(results[0:5])
else:
    print("Please load scaler and sentiment model.")
```

Visualization Dashboard

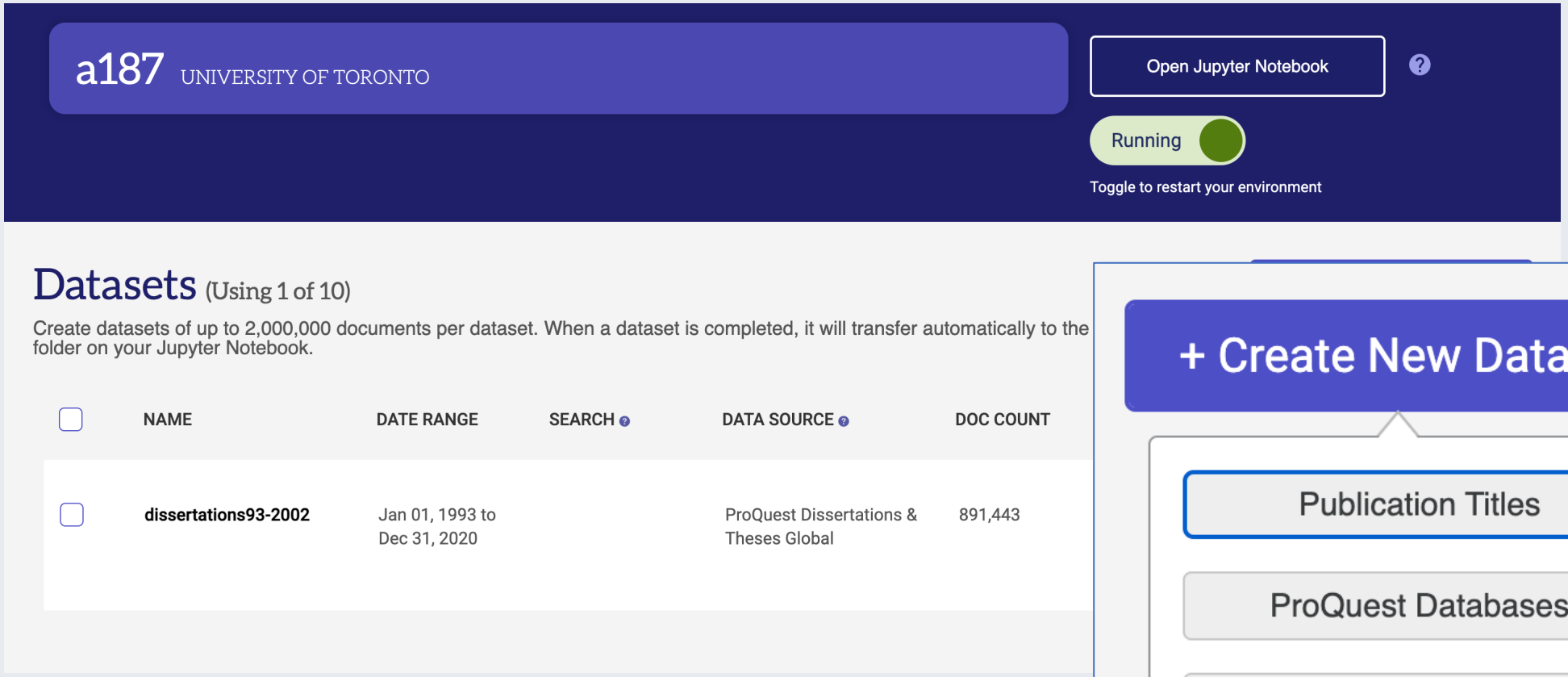
WHAT TO EXPECT

- Requires no coding experience
- Provides the ability to select content specific to your project
- Offers pre-configured data visualization to perform analysis



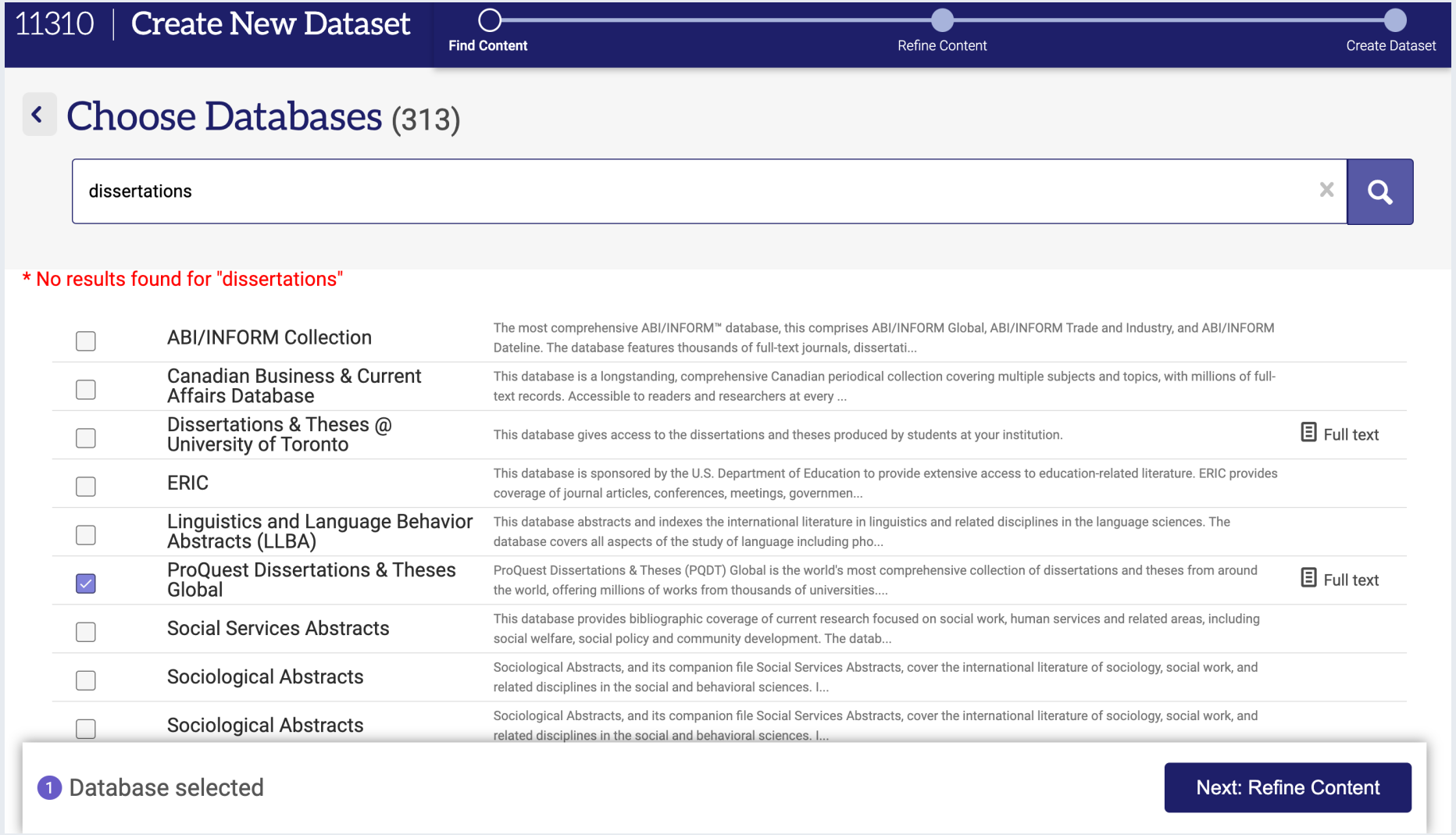
UNIVERSITY OF TORONTO LIBRARIES | MUSIC LIBRARY

James Mason, Metadata and Digital Initiatives Librarian
May 3rd, 2023



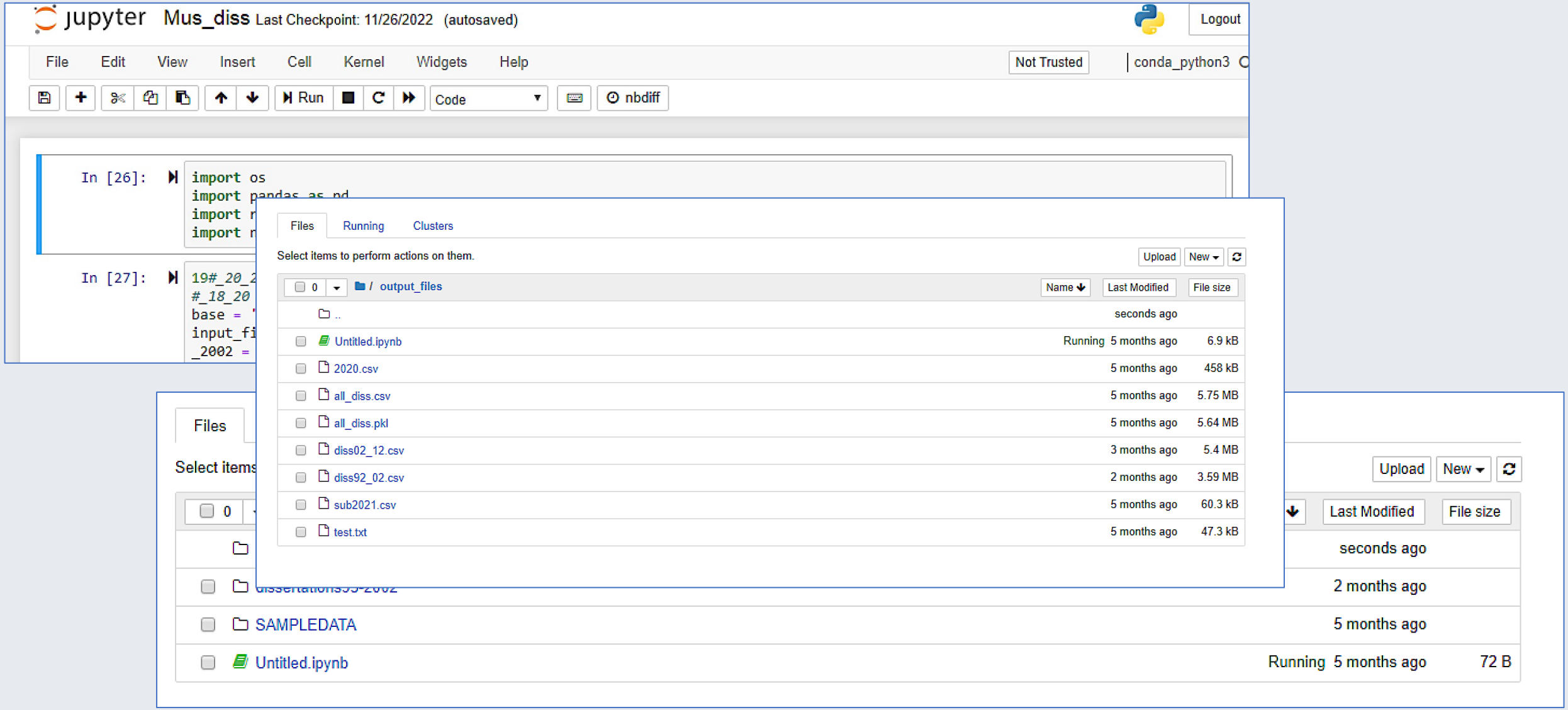
UNIVERSITY OF TORONTO LIBRARIES | MUSIC LIBRARY

James Mason, Metadata and Digital Initiatives Librarian
May 3rd, 2023



UNIVERSITY OF TORONTO LIBRARIES | MUSIC LIBRARY

James Mason, Metadata and Digital Initiatives Librarian
May 3rd, 2023



UNIVERSITY OF TORONTO LIBRARIES | MUSIC LIBRARY

James Mason, Metadata and Digital Initiatives Librarian
May 3rd, 2023

```
for i in input_files:
    elementList = []
    with open(base+"/"+i, "r") as f:
        lines = f.read()
        ClassCode = re.findall(r'<ClassCode>([0-9]{4})',lines)
        #metaElement(ClassCode)
        if ClassCode and "0413" in ClassCode:
            print(i)
            tags = re.findall(r'<([\w\W]*?)>',lines)
            for t in tags:
                print(t)
```

Filtered out dissertations not classified as music

```
DegreeDate = re.search(r'<DegreeDate>([\w\W]*?)<',lines)
metaElement(elementList,DegreeDate.group(1))
RawLang = re.findall(r'<RawLang>([\w\W]*?)<',lines)
metaElement(elementList,RawLang)
Title = re.findall(r'<Title>([\w\W]*?)<',lines)
metaElement(elementList,Title[0])
ClassExpansion = re.findall(r'<ClassExpansion>([\w\W]*?)<',lines)
metaElement(elementList,ClassExpansion)
SchoolLocation = re.findall(r'<SchoolCodeName>([\w\W]*?)<',lines)
metaElement(elementList,SchoolLocation[0])
GenSubjTerm = re.findall(r'GenSubjValue>([\w\W]*?)<',lines)
metaElement(elementList,GenSubjTerm)
DissPaperCategory = re.findall(r'<DissPaperCategory>([\w\W]*?)<',lines)
metaElement(elementList,DissPaperCategory)
DissPaperKwd = re.findall(r'<DissPaperCategory>([\w\W]*?)<',lines)
metaElement(elementList,DissPaperKwd)
GenSubjValue = re.findall(r'<GenSubjValue>([\w\W]*?)<',lines)
metaElement(elementList,GenSubjValue)
```

Metadata elements were gathered by their XML tags

- Title
- Degree date
- Category

James Mason, Metadata and Digital Initiatives Librarian
May 3rd, 2023

- Dataset Export data as csv – to excel

degre	language	title	Classification	institution
1995	['English', 'English']	Vocal lyricism in the melodies of Andre Caplet: Two lecture recitals revealing a singer's perspective	['Music', 'Music education']	University of Maryland, College Park
1995	['English', 'English']	Vocal fatigue in choral singing: Causes and suggestions for prevention voiced by prominent choral directors	['Music', 'Music education', 'Speech therapy']	The Florida State University
1995	['English', 'English']	Violin concerto: For violin and computer-generated tape. (Original composition);	['Music', 'Music education']	Stanford University
1995	['French', 'French']	Une approche poietique du volet composition en education musicale	['Music education', 'Music', 'Cognitive psychology']	Universite de Montreal (Canada)
1995	['English', 'English']	Twentieth century music for unaccompanied trumpet: An annotated bibliography	['Music', 'Music education']	Louisiana State University and Agricultural & Mechani
1995	['English', 'English']	Twentieth century Chinese vocal music with particular reference to its development and nationalistic characterist	['Music education', 'Music']	University of California, Los Angeles
1995	['English', 'English']	Tracking the nature of melodic expectancy development in musical children	['Music', 'Developmental psychology', 'Music education']	University of Washington
1995	['English', 'English']	The value of analysis in the maturation of cognitive musicianship: An experiential chronicle	['Music education', 'Music']	Duquesne University
1995	['English', 'English']	The Twelve Bagatelles and Sonata-Fantasia of George Rochberg: A performer's analysis	['Music', 'Music education', 'American studies']	Southwestern Baptist Theological Seminary
1995	['English', 'English']	The solo pianist: A critical analysis of concepts of musical giftedness	['Music education', 'Music', 'Psychology']	Concordia University (Canada)
1995	['English', 'English']	The relationships among choral performance quality, choral student emotive and aesthetic perception, and audie	['Music education', 'Music']	The University of Utah
1995	['English', 'English']	The relationship between text and music in the choral works of Robert H. Young	['Music', 'Biographies', 'Music education']	The Southern Baptist Theological Seminary

- First, classification terms were considered
- A Counter object was created for all the terms

	A	B	C	D	E
1	Dissertation year	Number of dissertations	number of classification ter	number of unique terms	terms as percent
2	2022	103	431	71	4.184466019
3	2021	165	715	90	4.333333333
4	2020	135	548	78	4.059259259
5	2019	152	612	93	4.026315789
6	2018	151	518	67	3.430463576
7	2017	161	492	70	3.055900621
8	2016	172	519	68	3.01744186
9	2015	164	521	65	3.176829268
10	2014	208	657	79	3.158653846
11	2013	187	567	65	3.032085561

James Mason, Metadata and Digital Initiatives Librarian
May 3rd, 2023

Here the classification terms were :

- Normalized
- Tokenized
- Stemmed
- Lemmatized
- Counted

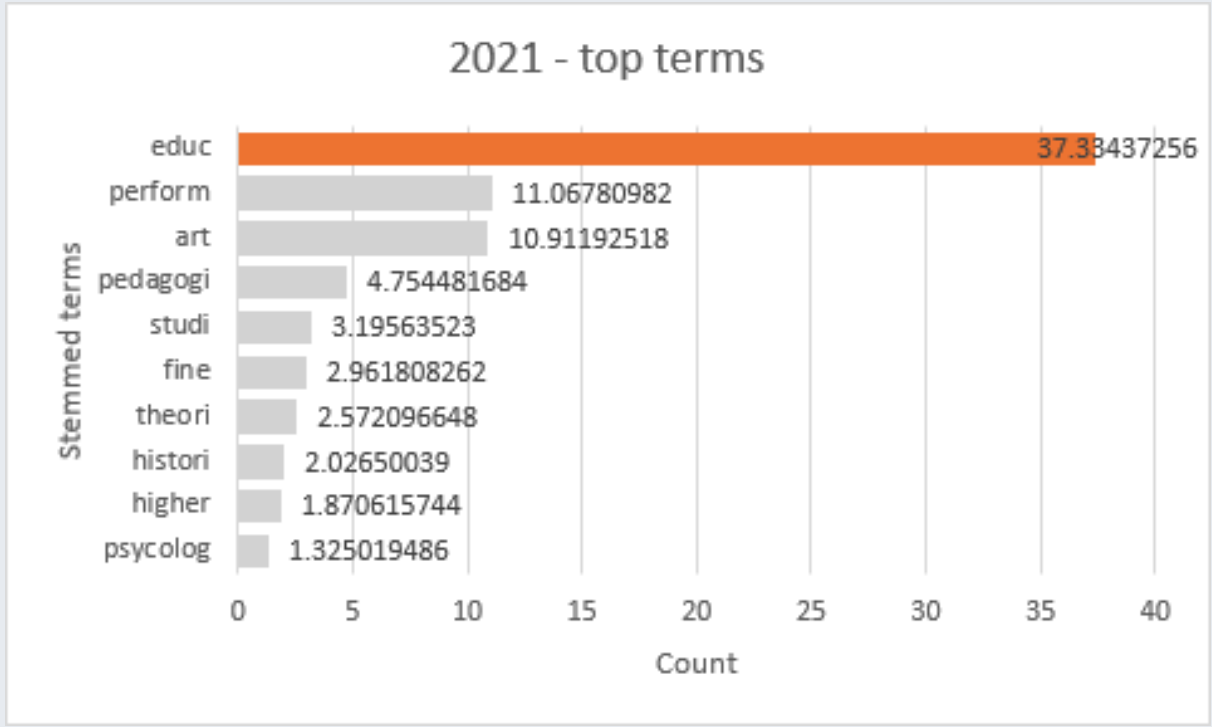
	Tokenized	Stemmed	Lemmatized
music education	education	educ	education
	music	music	music

Classification Term	coun	Tokenized	count	Stemmed	count	Lemmatized	count
teacher education	12	education	451	educ	479	education	451
music education	268	music	323	music	378	music	323
musical performances	43	pedagogy	61	perform	142	pedagogy	61
dance	2	performing	99	art	140	performing	99
performing arts	39	technology	8	pedagogi	61	technology	8
cultural anthropology	3	middle	4	studi	41	measurement	1
fine arts	38	musical	55	fine	38	musical	55
music history	21	development	8	theori	33	curriculum	8
reading instruction	1	gender	5	histori	26	psychology	17
latin american studies	3	curriculum	8	higher	24	anthropology	3
secondary education	9	theory	33	psycholog	17	higher	24
educational technology	8	teacher	12	teacher	12	teacher	12

James Mason, Metadata and Digital Initiatives Librarian
May 3rd, 2023

Terms collocated under “studi” -2021

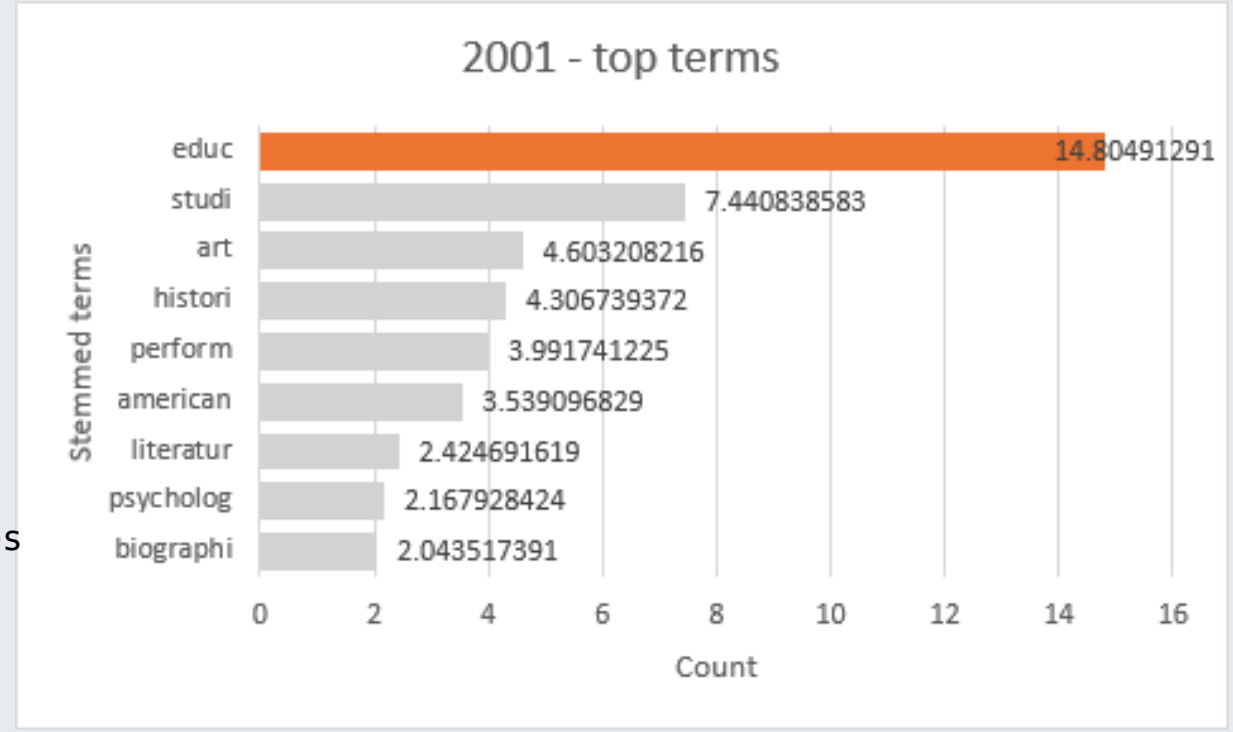
- african american studies
- asian studies
- black studies
- caribbean studies
- east european studies
- ethnic studies
- european studies
- latin american studies
- middle eastern studies
- near eastern studies
- scandinavian studies
- slavic studies
- web studies
- womens studies



James Mason, Metadata and Digital Initiatives Librarian
May 3rd, 2023

Terms collocated under “studi” -2001

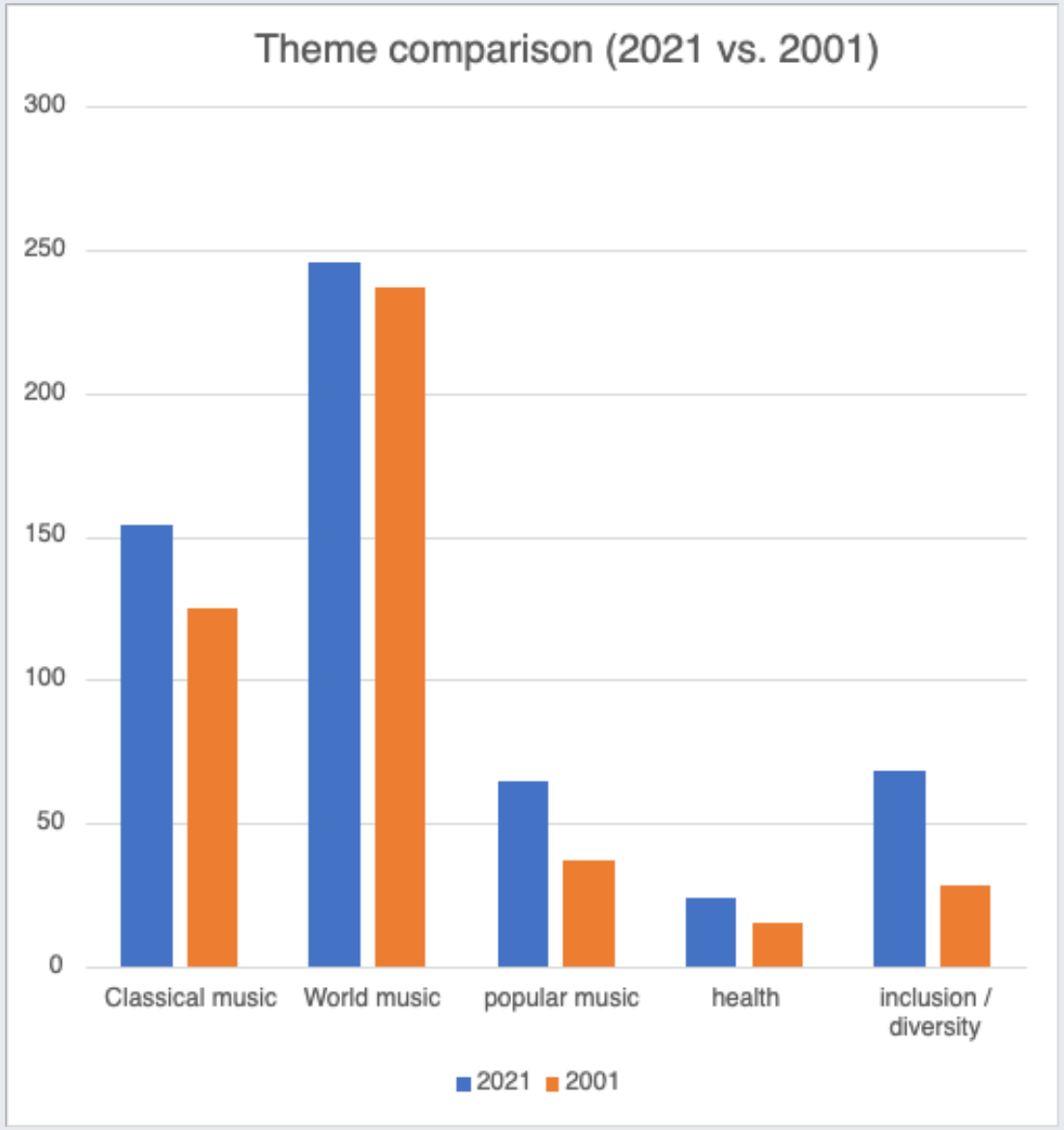
- baltic studies
- southeast asian studies
- environmental studies
- indigenous studies
- north african studies
- pacific rim studies
- scandinavian studies
- peace studies
- sound recording studios
- social studies education
- biblical studies
- disability studies
- asian american studies
- classical studies
- slavic studies
- sub saharan africa studies
- military studies
- translation studies
- east european studies
- holocaust studies
- african studies
- islamic studies
- museum studies
- hispanic american studies
- south african studies
- judaic studies
- caribbean studies
- lgbtq studies
- web studies
- canadian studies
- native american studies
- south asian studies
- regional studies
- european studies
- individual & family studies
- near eastern studies
- asian studies
- latin american studies
- middle eastern studies
- gender studies
- african american studies
- ethnic studies
- film studies
- black studies
- womens studies
- american studies



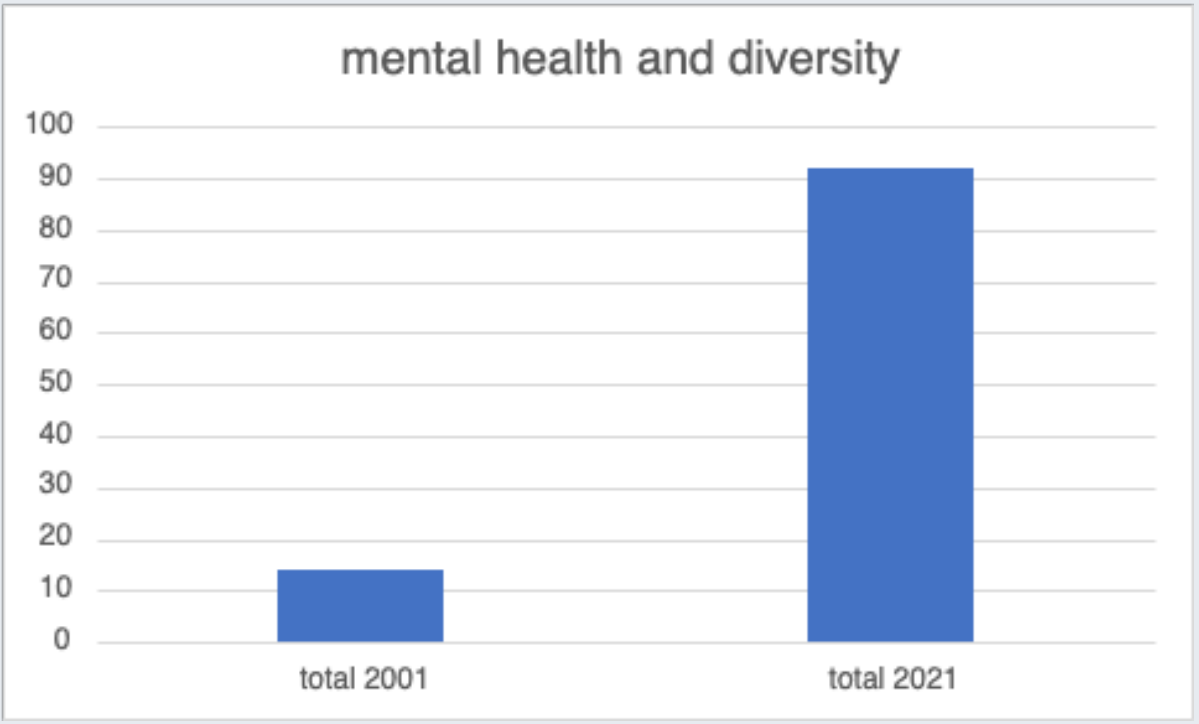
James Mason, Metadata and Digital Initiatives Librarian
May 3rd, 2023

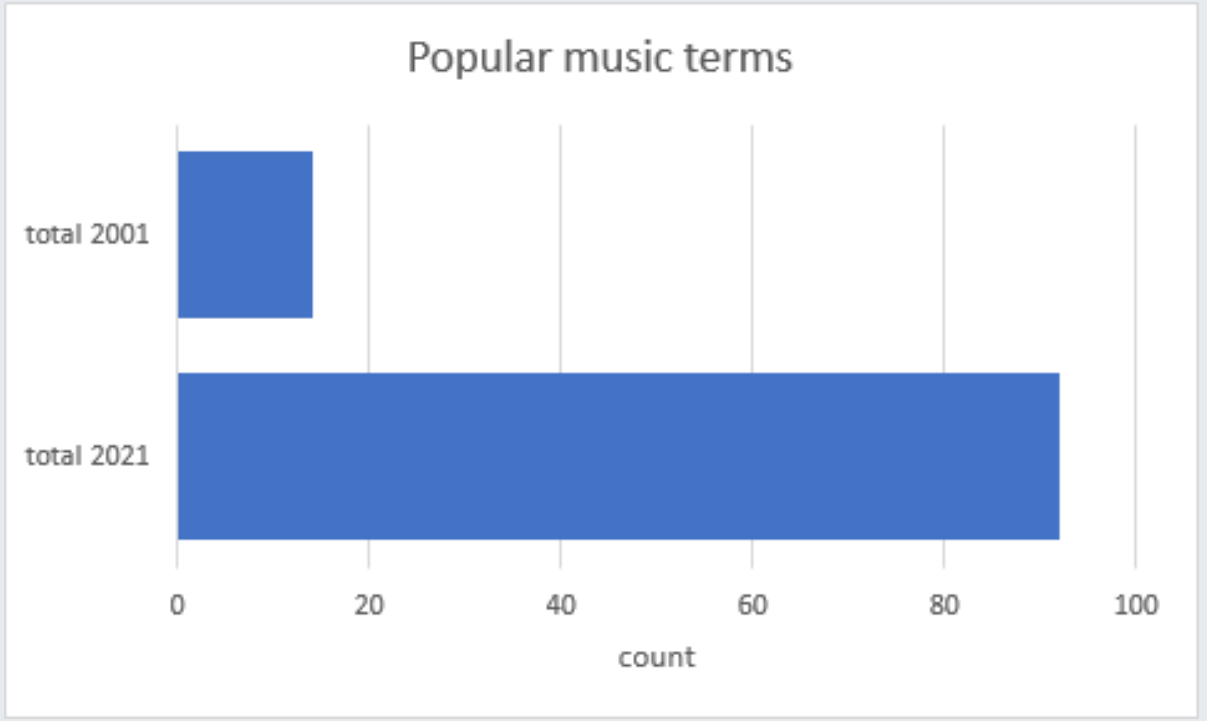
Titles

- Tokenized
- Lemmatized
- Bigrams created
- Topic modelling



James Mason, Metadata and Digital Initiatives Librarian
May 3rd, 2023





Mental health term examples

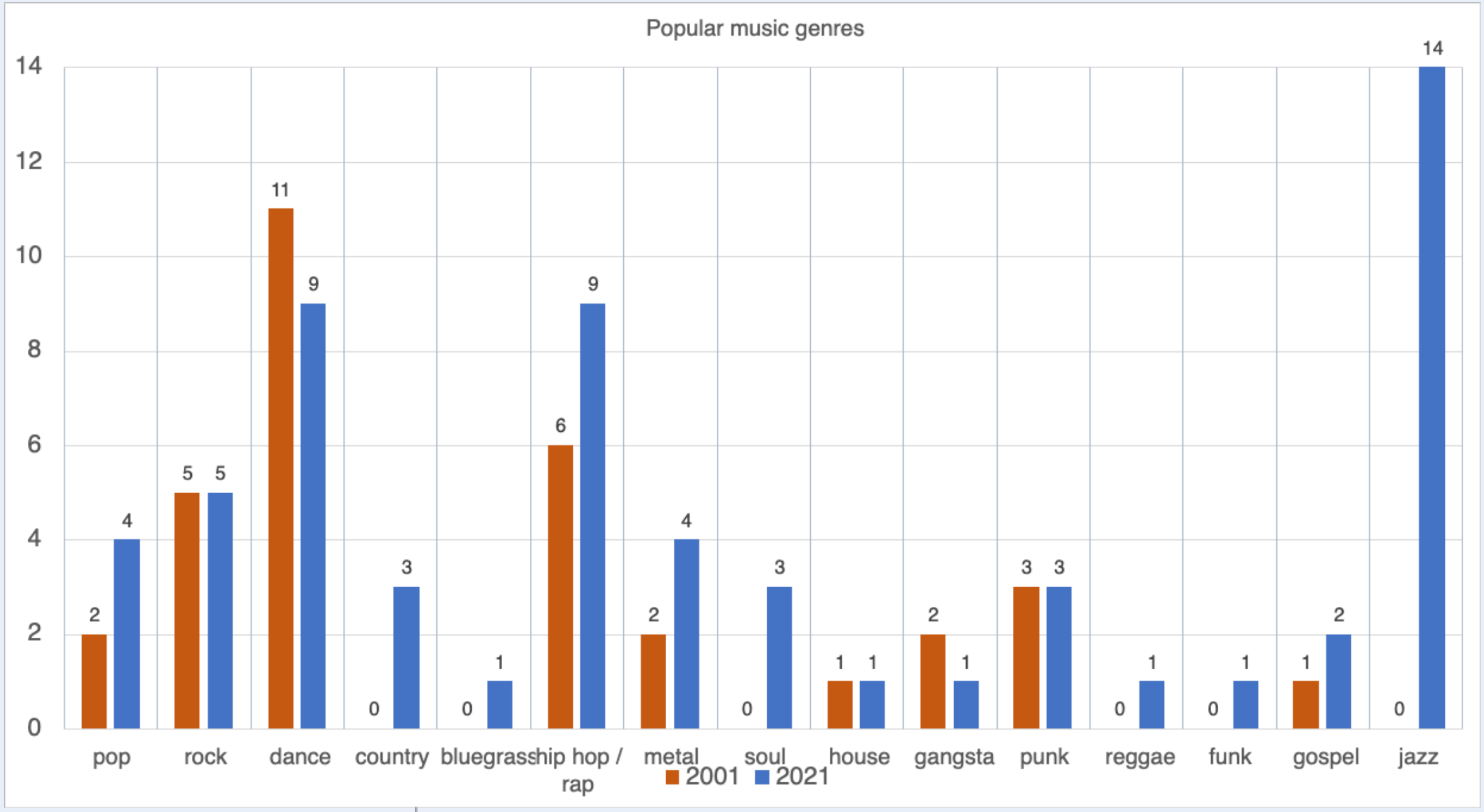
therapy
mental health
disorder
depression
crisis
anxiety

Inclusion term examples

justice
identity
bias
intersectionality
empathy
gender
feminism

UNIVERSITY OF TORONTO LIBRARIES | MUSIC LIBRARY

James Mason, Metadata and Digital Initiatives Librarian
May 3rd, 2023



James Mason, Metadata and Digital Initiatives Librarian
May 3rd, 2023

Nonwestern/art music 2021

James Mason, Metadata and Digital Initiatives Librarian
May 3rd, 2023

Nonwestern/art 2001

James Mason, Metadata and Digital Initiatives Librarian
May 3rd, 2023

Collection assessment
Collection development
Instruction

James Mason, Metadata and Digital Initiatives Librarian
May 3rd, 2023

Thank You

James Mason, Metadata and Digital Initiatives Librarian
j.mason@utoronto.ca