

RELIGIOUS PLACES OF WORSHIP -NC

INLS 490-186

Dr. Ryan Shaw

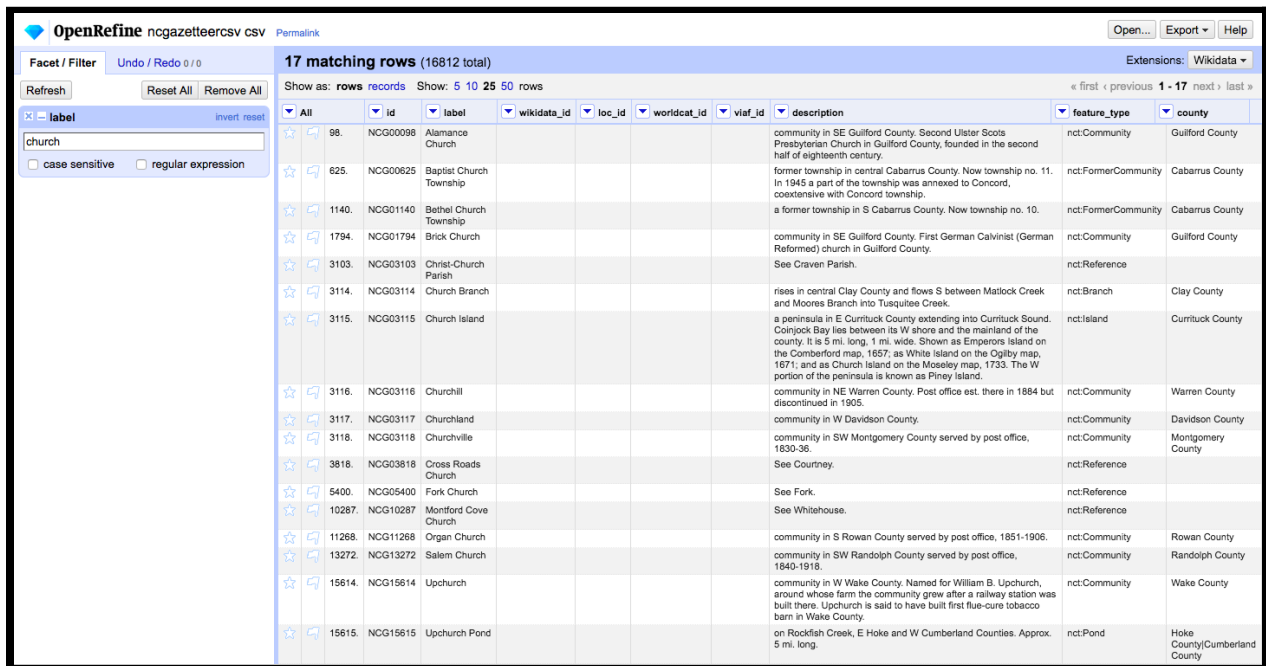
SPRING 2021

BY: Allie Barton, Ally Churchill, & Arsh Sangha

PS. All screenshots are from dataset before column names were edited to match conventions of NC
Gazetteer

INTRODUCTION:

Given the limited number of entries available on the North Carolina Gazetteer regarding religious place names, our team decided to work on adding more place names. When looking at the existing data entries in OpenRefine, there were only 17 entries which contained the word church out of 16812 entries. There were no entries containing the words temple, or mosque. Based on the likely little available data when it comes to religious place names — we decided this would be a great contribution to the existing gazetteer. Furthermore, upon making decisions about our project we decided to aim for quantity over quality because we were hoping for a maximum number of entries as opposed to super detailed, fewer entries. This is something we thought that future groups enrolled in this course could take on to add to the gazetteer.



The screenshot shows the OpenRefine interface with the search term "church" entered in the label facet. The interface displays 17 matching rows from a dataset of 16812 total rows. The columns include id, label, wikidata_id, loc_id, worldcat_id, viaf_id, description, feature_type, and county. The rows list various churches and church-related locations across different counties in North Carolina, such as Guilford County, Cabarrus County, and Warren County.

	All	id	label	wikidata_id	loc_id	worldcat_id	viaf_id	description	feature_type	county
98.	★	NGG00098	Alamance Church					community in SE Guilford County. Second Ulster Scots Presbyterian Church in Guilford County, founded in the second half of eighteenth century.	nct:Community	Guilford County
625.	★	NGG00625	Baptist Church Township					former township in central Cabarrus County. Now township no. 11. In 1945 a part of the township was annexed to Concord, coextensive with Concord township.	nct:FormerCommunity	Cabarrus County
1140.	★	NGG01140	Bethel Church Township					a former township in S Cabarrus County. Now township no. 10.	nct:FormerCommunity	Cabarrus County
1794.	★	NGG01794	Brick Church					community in SE Guilford County. First German Calvinist (German Reformed) church in Guilford County.	nct:Community	Guilford County
3103.	★	NGG03103	Christ-Church Parish					See Craven Parish.	nct:Reference	
3114.	★	NGG03114	Church Branch					rises in central Clay County and flows S between Matlock Creek and Moores Branch into Tusquitee Creek.	nct:Branch	Clay County
3115.	★	NGG03115	Church Island					a peninsula in E Currituck County extending into Currituck Sound. Coinjock Bay lies between its W shore and the mainland of the county. It is 5 mi. long, 1 mi. wide. Shown as Emperors island on the Comerford map, 1657; as White Island on the Ogilby map, 1671; and as Church Island on the Moseley map, 1733. The W portion of the peninsula is known as Piney Island.	nct:Island	Currituck County
3116.	★	NGG03116	Churchill					community in NE Warren County. Post office est. there in 1884 but discontinued in 1905.	nct:Community	Warren County
3117.	★	NGG03117	Churchland					community in W Davidson County.	nct:Community	Davidson County
3118.	★	NGG03118	Churchville					community in SW Montgomery County served by post office, 1830-36.	nct:Community	Montgomery County
3818.	★	NGG03818	Cross Roads Church					See Courtney.	nct:Reference	
5400.	★	NGG05400	Fork Church					See Fork.	nct:Reference	
10287.	★	NGG10287	Montford Cove Church					See Whitehouse.	nct:Reference	
11268.	★	NGG11268	Organ Church					community in S Rowan County served by post office, 1851-1906.	nct:Community	Rowan County
13272.	★	NGG13272	Salem Church					community in SW Randolph County served by post office, 1840-1918.	nct:Community	Randolph County
15614.	★	NGG15614	Upchurch					community in W Wake County. Named for William B. Upchurch, around whose farm the community grew after a railway station was built there. Upchurch is said to have built first flue-cure tobacco barn in Wake County.	nct:Community	Wake County
15615.	★	NGG15615	Upchurch Pond					on Rockfish Creek, E Hoke and W Cumberland Counties. Approx. 5 mi. long.	nct:Pond	Hoke County/Cumberland County

THE ORIGINAL IDEA:

Originally, we were interested in exploring the broad implications of religion across the state of North Carolina. Based on our previous knowledge of the existence of historical religious communities in the state, we explored whether there were any more we could document for the Gazetteer. First, we looked at historical sites that still exist today, which included Old Salem, Bethabara and Bethania. However, while N.C. promotes a number of historic townships, only the aforementioned three were originally religious planned communities. We wanted to expand our scope past religious communities so we could add a breadth of religious information to the Gazetteer.

Next, we explored NCpedia to learn about the religious history of North Carolina and begin to understand the distribution of types of religion. We discovered that N.C. features a number of religions including Judaism, Islam, Hinduism, Catholicism, and many denominations of Protestant Christianity.

After considering the scope of what we could achieve in the semester, we decided to record the geographic information of places of worship in N.C. Our final project is not a comprehensive, all-inclusive account of every place of worship in the state. Religions outside of Christianity are underrepresented in the data we used. Future groups should consider finding other sources to build the index of places of worship in the Gazetteer.

ORIGINAL DATA:

One source that explored the historical settlements in N.C. was a book called Religious Traditions of North Carolina: Histories, Tenets, and Leaders, edited by W. Glenn Jonas, Jr. The work offered a historical perspective on religious communities and distribution in the state, but

provided no information that would have been useful towards our ultimate goal of geocoding for the Gazetteer.

Our main source for gathering information on the places of worship in N.C. came from the Homeland Infrastructure Foundation-Level Data (HIFLD). The Department of Homeland Security (DHS) created a dataset of All Places of Worship in the United States from the IRS master files containing the public national record of all 501(c)(3) non-profit organizations. The dataset was geocoded with the physical locations of each place of worship. We filtered the dataset to only include places in N.C.

The sources we used for religious camps and retreats are the websites <https://www.kidscamps.com/camps/north-carolina-religious-camps.camp> and https://www.retreatfinder.com/Directory/United_States/NC_North%20Carolina.aspx. Both had limited information and we did not consult sources outside of this, but it was a good starting point for our small scale contribution.

DATA MANIPULATION:

The process when examining the data points we collected in our research was filled with several trials and errors. I will take you through a step by step process of these ideas. On the Homeland Security Website, which received its sources from the IRS, we downloaded all of the southern entries which contained religious place names from North Carolina, South Carolina, and Georgia. We proceeded to download this data as a CSV, leaving us with approximately 10,061 entries. Then, we narrowed these entries down to 69006 by using the text filter “NC.”

We then made our first attempt at reconciliation with the Homeland Security Data on OpenRefine. We attempted to reconcile all place names against church buildings in Wikidata.

This process took approximately three hours to complete, and upon retrieving results, there were hardly any. The results we did have when looking at what entries had matched those within Wikidata were not in North Carolina and sometimes were not in the United States.

Upon encountering this stepping stone, we decided the best way to shorten the wait times when reconciling the data was to narrow down our reconciliation process to just the place names that begin with the letter A, followed by place names that begin with the letter B, and so on. We also decided to narrow down counties when reconciling, to hopefully only get matching religious place names that are in North Carolina. In order to do this, we added a text facet on the column subregion. There were 75 counties as well as “NOT AVAILABLE” when applying this text facet. We chose the option “NOT AVAILABLE” and inverted the results, so that any reconciliation process would not attempt to match with counties that were unavailable, and instead only with counties located in North Carolina. When removing the unavailable options, this brought us down to 46 possible counties to match from.

Next, we applied the text filter to our “NAME” column, which included the religious place names. We used the regular expression ^A in order to only match data points of place names that begin with the letter A.

STATE

invert reset

NC

☐ case sensitive
☐ regular expression

SUBREGION

change invert reset

46 choices Sort by: name count Cluster

MADISON -2
MECKLENBURG -38
MITCHELL -1
MOORE -3
NOT AVAILABLE -102 include
ONSLow -5
ORANGE -4
PITT -2
POLK -5
RANDOLPH -11
RICHMOND -2

NAME

invert reset

^A

☐ case sensitive
☒ regular expression

We again reconciled against “church building” in Wikidata, by including the subregion column when it asked for relevant details. We then used the Wikidata code “P131,” which means “located in the administrative territorial entity.”

Reconcile column "NAME"

» Access Service API

Reconcile each cell to an entity of one of these types:

☐ foundation Q157031
☒ church building Q16970
☐ school Q3914

Also use relevant details from other columns:

STADDR

☐

STADDR

☐

CITY_2

☐

SUBREGION

☒

P131

Select an item from the list:

located in the administrative territorial entity P131

the item is located on the territory of the following administrative ent

COUNTRY

☐

LANGCODE

☐

DISTANCE

☐

X

☐

Y

☐

DISPLAYX

☐

Reconcile against type:

Reconcile against no particular type

☒ Auto-match candidates with high confidence

Maximum number of candidates to return


Add Standard Service...

Start Reconciling

Cancel

Though using these methods made the reconciliation process much quicker, we were still not seeing the results that we wanted. There were hardly any matches and the matches the process did find were not located in North Carolina. We repeated this process with the Bs and Cs, and the results were all similar. We then took this knowledge and attempted to reconcile against cities instead of the subregion. Unfortunately, we still were not getting the matching results that we wanted.

The realization dawned that perhaps Wikidata did not have anywhere close to 6900 church building names that our data could match with. We used the following query which searched for how many church buildings are actually located in North Carolina. The number result was 784, which is nowhere close to the 6900 we were attempting to match with. With our failures when trying to reconcile the data within OpenRefine, our best bet for successfully matching entries at this point was to manually enter any matches we could individually find between the main data we were working with and the entries available on Wikidata. At this point, we chose to continue to only focus on Wikidata's subclass of "church building" in order to narrow down our results, since we would now begin entering this data.



The screenshot shows the Wikidata Query Service interface. At the top, there is a header with the Wikidata logo, the text "Wikidata Query Service", and three buttons: "Examples", "Help", and "More tools". Below the header, there is a sidebar with icons for information, expand/collapse, pin, diamond, folder, refresh, and trash. The main area displays a SPARQL query with line numbers 1 through 12. The query is as follows:

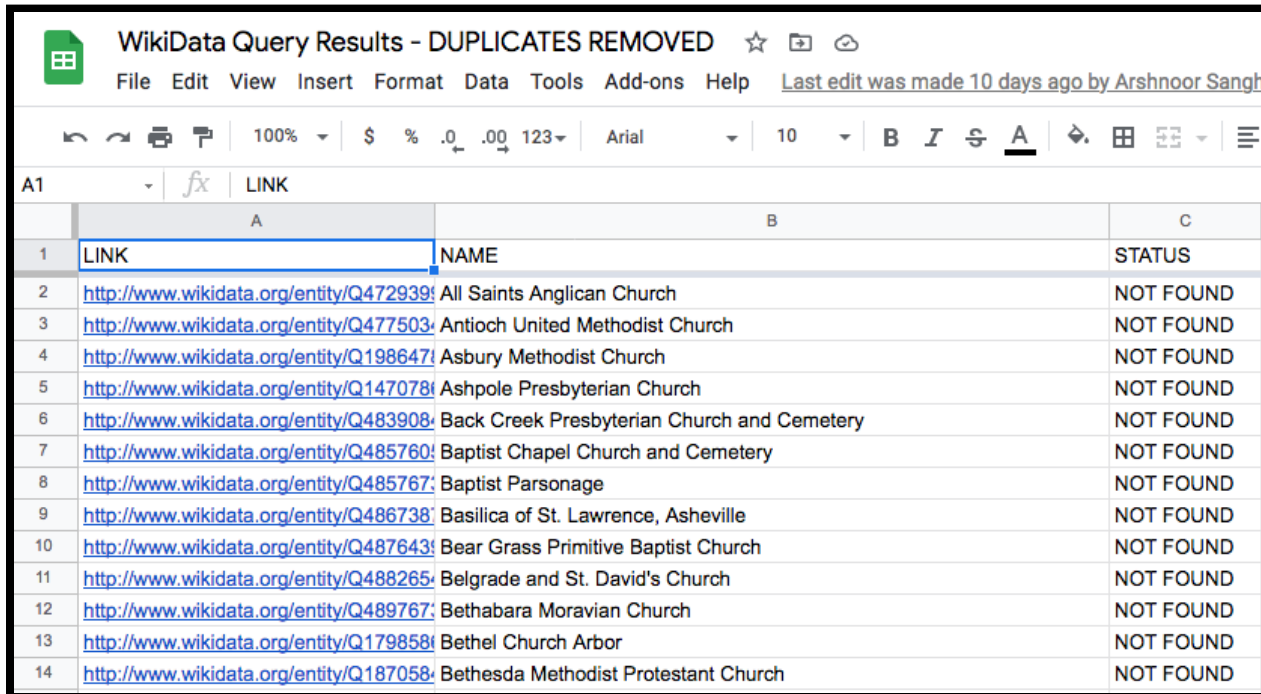
```
1 SELECT ?site ?siteLabel ?locationLabel WHERE {  
2   # this is necessary to keep the query from timing out  
3   hint:Query hint:optimizer "None".  
4  
5   ?site (wdt:P131+) wd:Q1454;  
6     (wdt:P31/(wdt:P279*)) wd:Q16970.  
7  
8   ?site wdt:P131 ?location .  
9  
10  SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }  
11 }  
12
```

After deciding this, we downloaded the 784 results the Wikidata query gave us and downloaded this data as a CSV to begin to work within Google Sheets. We were going to go in and match the results manually. After alphabetizing our place names based on their name, we suspected there might be duplicate Wikidata links. We then removed any link duplicates to further narrow down the data we were working with. A small subset of this downloaded data before and after duplicates were removed is below. With duplicates removed, we were left with 258 entries out of 784.

Before duplicates were removed:

WikiData Query Results - DUPLICATES REMOVED			
File Edit View Insert Format Data Tools Add-ons Help Last edit was made 10 days ago by Arshnoor Sangh			
100% \$ % .0 .00 123 Arial 10 B I S A			
A1	LINK		
	A	B	C
1	LINK	NAME	STATUS
2	http://www.wikidata.org/entity/Q472939	All Saints Anglican Church	NOT FOUND
3	http://www.wikidata.org/entity/Q477503	Antioch United Methodist Church	NOT FOUND
4	http://www.wikidata.org/entity/Q198647	Asbury Methodist Church	NOT FOUND
5	http://www.wikidata.org/entity/Q147078	Ashpole Presbyterian Church	NOT FOUND
6	http://www.wikidata.org/entity/Q483908	Back Creek Presbyterian Church and Cemetery	NOT FOUND
7	http://www.wikidata.org/entity/Q485760	Baptist Chapel Church and Cemetery	NOT FOUND
8	http://www.wikidata.org/entity/Q485767	Baptist Parsonage	NOT FOUND
9	http://www.wikidata.org/entity/Q486738	Basilica of St. Lawrence, Asheville	NOT FOUND
10	http://www.wikidata.org/entity/Q487643	Bear Grass Primitive Baptist Church	NOT FOUND
11	http://www.wikidata.org/entity/Q488265	Belgrade and St. David's Church	NOT FOUND
12	http://www.wikidata.org/entity/Q489767	Bethabara Moravian Church	NOT FOUND
13	http://www.wikidata.org/entity/Q179858	Bethel Church Arbor	NOT FOUND
14	http://www.wikidata.org/entity/Q187058	Bethesda Methodist Protestant Church	NOT FOUND

After duplicates were removed:



	A	B	C
1	LINK	NAME	STATUS
2	http://www.wikidata.org/entity/Q472939	All Saints Anglican Church	NOT FOUND
3	http://www.wikidata.org/entity/Q477503	Antioch United Methodist Church	NOT FOUND
4	http://www.wikidata.org/entity/Q198647	Asbury Methodist Church	NOT FOUND
5	http://www.wikidata.org/entity/Q147078	Ashpole Presbyterian Church	NOT FOUND
6	http://www.wikidata.org/entity/Q483908	Back Creek Presbyterian Church and Cemetery	NOT FOUND
7	http://www.wikidata.org/entity/Q485760	Baptist Chapel Church and Cemetery	NOT FOUND
8	http://www.wikidata.org/entity/Q485767	Baptist Parsonage	NOT FOUND
9	http://www.wikidata.org/entity/Q486738	Basilica of St. Lawrence, Asheville	NOT FOUND
10	http://www.wikidata.org/entity/Q487643	Bear Grass Primitive Baptist Church	NOT FOUND
11	http://www.wikidata.org/entity/Q488265	Belgrade and St. David's Church	NOT FOUND
12	http://www.wikidata.org/entity/Q489767	Bethabara Moravian Church	NOT FOUND
13	http://www.wikidata.org/entity/Q179858	Bethel Church Arbor	NOT FOUND
14	http://www.wikidata.org/entity/Q187058	Bethesda Methodist Protestant Church	NOT FOUND

Next we divided the work and went through each link and began classifying the status of each church building place name as “NOT FOUND” or “ENTERED.” This is explained in more detail in our FINAL PROCESS section. For the sake of simplicity, we usually only added a Wikidata link to our final dataset if we were positive that the church building place names were a match, depending on what information the Wikidata page could give us about each church building.

RELIGIOUS CAMPS

_____ Along with only adding religious place names (which we later simplified down to church building place names), we also thought it would be helpful to the gazetteer to include a few religious camps in North Carolina. We only added data from one source, as mentioned in ORIGINAL DATA. From the data, we added the columns shown below based on what information we could find about these religious summer camps online. Our points included the

name, program address, latitude and longitude, winter address, state, religion of each camp, and a description, website, and phone if they were included.

In retrieving the latitude and longitude of these locations, we utilized Google Sheets geocoding feature. We followed these github directions in order to complete it:

<http://willgeary.github.io/data/2016/11/04/Geocoding-with-Google-Sheets.html>. There weren't

many actual addresses as most locations given were just cities, so the latitude and longitude coordinates are very general.

A	B	C	D	E	F	G	H	I	J
NAME	PROGRAM ADDRESS	LAT	LONG	WINTER ADDRESS	STATE	RELIGION	DESC	WEBSITE	PHONE
A E Finley YMCA	Raleigh	35.7795897	-78.6381787	Raleigh	NC	CHRISTIAN			
Adrian L Shuford Jr YM	Conover	35.7065217	-81.2186933	Conover	NC	CHRISTIAN			
Alamance County Con	Burlington	45.0309042	-93.029523	Burlington	NC	CHRISTIAN			
Asheville JCC	Asheville	35.5950581	-82.5514869	Asheville	NC	JEWISH			
Bertie County YMCA	Windsor	42.3149367	-83.0363633	Windsor	NC	CHRISTIAN			
Boiling Springs YMCA	Boiling Springs	35.046508	-81.9817726	Boiling Springs	NC	CHRISTIAN			
Camp Buc	Sapphire	36.1347413	-115.171593	Sapphire	NC	PRESBYTERIAN			
Camp Celo	Burnsville	44.7677424	-93.2777226	Burnsville	NC	QUAKER			
Camp Creation	Charlotte	35.2270869	-80.8431267	Charlotte	NC	CHRISTIAN			
Camp Crestridge for G	105 Balsam Rd. Ridgecrest, NC 2	35.628346	-82.279417	PO Box 279 Ridgecrest, NC 2	NC	CHRISTIAN	Boys and girls of	https://girls.ridge	828-669-2613
Camp Glen Arden for G	Tuxedo	41.9194716	-87.6621152	Tuxedo	NC	CHRISTIAN	Traditional private girls camp in western North Caro		
Camp Greystone	Tuxedo	41.9194716	-87.6621152	Tuxedo	NC	CHRISTIAN	Owned & directed by the Miller family since 1920, C		
Camp Harrison	Boomer	39.7089926	-104.9338071	Boomer	NC	CHRISTIAN	One & two week, co-ed, residential camp for ages		
Camp Holler	Fletcher	37.06448	-94.4769986	Fletcher	NC	CHRISTIAN			

RELIGIOUS RETREATS

There was not much information on the religious retreats that we could find on the one source we looked at. We did not initiate geocoding because there were no actual addresses in this small amount of data. We did not console other sources in order to see if there were more retreats, or delve into further information about each retreat. An example of our data is shown below.

Name	City	State
Art of Living Retreat Center	Boone	NC
Mountain Light Sanctuary	Asheville	NC
Trinity Center - Retreat at the Beach!	Pine Knoll Shores	NC
Blue Lotus Ayurveda - Panchakarma Clinic & Day Spa	Asheville	NC
Cedar Cross Retreat Center	Louisburg	NC
Christmount Conference Center & Camp	Black Mountain	NC
Heartwood Refuge	Hendersonville	NC
High Pastures Christian Retreat	Burnsville	NC
Hinton Rural Life Center	Hayesville	NC

Finally we added all of our data entries into our manipulated data set, which is the final product we are submitting along with this paper. We had to create new columns for the columns we had included in our religious camps data that were not included in our church building place names, or religious retreat data. Our data was reorganized and realphabitized after adding these extra 73 data entries, so that all columns accurately reflected the information included within them. We ended up with 6979 total new entries to add to the North Carolina Gazetteer.

PROBLEMS ENCOUNTERED:

Though some problems we encountered were outlined in the DATA MANIPULATION section, we encountered a multitude of trials and errors when messing around with our dataset. The main, overarching issue was that Wikidata lacked so much data that could have been beneficial for our reconciliation process. However, there were multiple other smaller details which made our reconciliation difficult.

The first issue I will reiterate were the duplicates in our Wikidata query results. From 784 results, only 258 were unique. The implications of this were that we likely removed some potential accurate matches when removing duplicates that could have been added to our dataset. So many church building place names in North Carolina share the same name, so it is not

surprising that multiple different churches in different counties accidentally shared the same Wikidata link. No online resource when looking for data entries is decidedly foolproof, encompassing all data relevant with no inaccuracies.

Further reasons why we likely did not achieve much success in our automatic reconciliation process is because of name discrepancies. ST. often being substituted for SAINT on our Homeland Security data versus the Wikidata and vice versa is a good example of these discrepancies. More examples are EAST entries being shortened down to E, as well as other numerical directors when it came to names, or AVE versus AVENUE. Because of these minor differences, during our reconciliation process, even if there was a direct match we could later note when completing the process manually, the computer may have perceived two of the same church buildings as different entities because of the slight differences in labeling. There is much to account for when not able to easily match Wikidata entries to our dataset.

Another issue we uncovered when we were doing our manual reconciliation to try to match Wikidata links to our data points was that alternative names existed. In other words, if a church went by a certain name, sometimes on the Wikidata information page, we would see an alternative name the church might also go by. Some of these names were very different. We guess that if sometimes there is an alternative name listed, other times there is likely not one even if a church does have two names. This could result in a further hindrance of our reconciliation process. An example of something like this occurring in our reconciliation process is below. “First Baptist Church” in Eden, North Carolina is also known as Leaksville Baptist Church.

<div> <div>First Baptist Church</div> <div>(Q5452379)</div> </div>			
<div> <div>historic church in Eden, North Carolina, United States</div> <div>Leaksville Baptist Church</div> </div>			<div> <div></div> <div>edit</div> </div>
<div> <div> <div>▼ In more languages</div> <div>Configure</div> </div> </div>			
Language	Label	Description	Also known as
English	First Baptist Church	historic church in Eden, North Carolina, United States	Leaksville Baptist Church
Spanish	Primera Iglesia Bautista (Eden, Carolina del Norte)	No description defined	
Traditional Chinese	No label defined	No description defined	
Chinese	No label defined	No description defined	

FINAL PROCESS AND PRODUCT:

Our final process began with editing the Homeland dataset. The data sheet had a total of 6,908 rows, or unique places of worship in NC. We also had 53 columns in the dataset to work with. While taking a closer look, we found that many of these were redundant, particularly the address columns, which were repeated four times, along with the coordinate columns, which were also repeated four times. In addition to redundancy, we also found that we were dealing with columns with labels that were not self-explanatory, and no keys were found on the Homeland Security or IRS websites (Image below). Columns *J (FOUNDATION)*, *K (ACTIVITY)*, *L (ORGANIZATION)*, *M (TAX_PERIOD)*, *N (ACCT_PD)*, *O (NTEE_CD)*, *P (SORT_NAME)* were all removed due to this reason.

G	H	I	J	K	L	M	N	O	P	Q	R
ZIP	AFFILIATION	RULING	FOUNDATION	ACTIVITY	ORGANIZATION	TAX PERIOD	ACCT PD	NTEE CD	SORT NAME	LOC_NAME	GEOCODED_S
28329		3	201508	15	0	1	201806	6 X20	NOT AVAILABLE POSTAL		M
28303		3	201408	15	0	1	201712	12 X20	NOT AVAILABLE STREETADDRESS		M
27532		3	200712	10	0	1	0	12 X21	NOT AVAILABLE POSTAL		M
28645		3	198210	10	0	5	0	12 NOT AVAILABLE	NOT AVAILABLE POINTADDRESS		M
28540		3	199702	10	1000000	1	0	12 X21Z	NOT AVAILABLE POINTADDRESS		M
28546		3	200706	10	0	1	0	12 X21	NOT AVAILABLE POINTADDRESS		M
27519		3	199411	10	1000000	1	0	12 NOT AVAILABLE	NOT AVAILABLE POINTADDRESS		M
27282		9	196408	10	1002029	5	0	12 NOT AVAILABLE	NOT AVAILABLE POSTAL		M
27406		3	201109	10	0	1	0	12 X20	NOT AVAILABLE POINTADDRESS		M
26206		3	199108	10	1029000	1	0	12 NOT AVAILABLE	NOT AVAILABLE POINTADDRESS		M
27893		3	200007	10	0	1	0	12 X21	NOT AVAILABLE POINTADDRESS		M
28638		9	197210	10	1000000	5	0	12 NOT AVAILABLE	1570 POSTAL		M
27610		3	198608	10	1000000	1	201012	12 NOT AVAILABLE	NOT AVAILABLE POINTADDRESS		M
27532		3	200806	10	0	1	0	12 X20	NOT AVAILABLE POSTAL		M
28540		3	200508	10	0	5	0	12 X21	NOT AVAILABLE POINTADDRESS		M
28070		3	200108	10	0	1	0	12 X20	NOT AVAILABLE POSTAL		M
28163		3	198102	10	1000000	1	0	12 NOT AVAILABLE	NOT AVAILABLE STREETNAME		M
27517		3	200506	10	0	1	0	4 X20	NOT AVAILABLE POINTADDRESS		M
28541		3	198305	10	1000000	1	0	12 NOT AVAILABLE	NOT AVAILABLE POSTAL		M
27591		3	198602	10	1000000	1	0	12 NOT AVAILABLE	NOT AVAILABLE POINTADDRESS		M
28318		3	201111	15	0	1	201712	12 X20	NOT AVAILABLE POINTADDRESS		M
28229		9	194107	10	1029000	5	0	6 NOT AVAILABLE	NOT AVAILABLE POSTAL		M
27868		3	200706	10	0	1	0	12 X20	NOT AVAILABLE POSTAL		M
28052		3	201808	16	0	1	201712	12 X99	NOT AVAILABLE POINTADDRESS		M
27611		3	198511	10	1000000	1	0	12 NOT AVAILABLE	NOT AVAILABLE POSTAL		M
28753		3	201612	16	0	1	201612	12 X20	NOT AVAILABLE STREETADDRESS		M
28345		3	200502	10	0	1	0	12 X21	NOT AVAILABLE POSTAL		M
27703		3	199509	10	1029002	1	0	12 NOT AVAILABLE	NOT AVAILABLE POINTADDRESS		M
27604		3	200903	10	0	1	0	12 X21	NOT AVAILABLE POINTADDRESS		M
27325		3	197012	10	1000000	1	0	12 NOT AVAILABLE	NOT AVAILABLE POINTADDRESS		M
28719		3	200311	10	0	1	0	12 X21	NOT AVAILABLE POSTAL		M
28540		3	201709	15	0	1	201712	12 X20	NOT AVAILABLE POINTADDRESS		M
28603		3	200301	10	0	1	0	1 X21	NOT AVAILABLE POSTAL		M
28070		3	200902	15	0	1	201712	12 S20	NOT AVAILABLE POSTAL		M
27560		3	201406	10	0	1	0	12 X21	NOT AVAILABLE POINTADDRESS		M
27317		9	196408	10	1002029	5	0	12 NOT AVAILABLE	NOT AVAILABLE POINTADDRESS		M
27282		9	196703	10	1000000	5	0	12 NOT AVAILABLE	NOT AVAILABLE STREETADDRESS		M
28205		9	201404	10	0	1	0	12 X21	NOT AVAILABLE POINTADDRESS		M
28734		3	198606	10	1000000	1	0	12 NOT AVAILABLE	NOT AVAILABLE STREETADDRESS		M
28376		3	201706	10	0	1	0	12 X20	NOT AVAILABLE POSTAL		M
27006		9	197410	10	0	5	0	12 NOT AVAILABLE	NOT AVAILABLE POSTAL		M
27524		9	195809	10	0	1	0	12 NOT AVAILABLE	NOT AVAILABLE POSTAL		M
28227		8	195809	10	1000000	1	201712	12 NOT AVAILABLE	NOT AVAILABLE POSTAL		M
28213		9	198804	10	1000000	1	201706	6 NOT AVAILABLE	NOT AVAILABLE POINTADDRESS		M

After deleting all unnecessary columns, we began adding those columns necessary to the overall project - this was limited to just one column - the Wikidata Column. At the end, we were able to bring our columns down from 53 to 29, which allowed us to have a cleaner and more concise view of our data as well as have a slightly easier time actually manually reconciling.

FINAL PRODUCT:

Our edited dataset was now ready. Some of our important retained columns are as follows:

Column B: EIN: Employer Identification Number, registered with state, acts as a sort of Social Security number for organizations

Column C: WIKIDATA: Included the link to wikidata source (when available)

Column K: LOC_NAME: Tells us if the address, in columns E-H is a point address, or in the case of P.O. boxes, a postal address

Column M: MATCH_ADDR: For P.O. boxes addresses in column K, gives further information. If the address is a point address, simply repeats it.

Column O: ADDNUM: Provides a column for the building number for the address (applicable for point addresses). For P.O. Box addresses, values are “NOT AVAILABLE.”

Column P: SIDE: Gives us what side of the road the building is located. For P.O. Box addresses, values are “NOT AVAILABLE.”

Column Q: STPREDIR: Provides the direction of the street direction for point addresses, when available.

Column R: STPRETYPE: Provides the “type” of prefix of the street address. For example, for the street name “New England St,” this column would give the value “New.”

	A	B	C	D	E	F	G	H	I	J	K	L
1	OBJECTID	EIN	WIKIDATA	NAME	STREET	CITY	STATE	ZIP	AFFILIATION	RULING	LOC_NAME	GEOCODED_ST
2	35758	204381243			320	3401 BROUGHT DURHAM	NC		27704	3	200707 POINTADDRESS	M
3	113630	640962994		110 CENTER	303	VINTAGE HI APEX	NC		27539	3	200906 POINTADDRESS	M
4	105328	562118578		121 INC		PO BOX 26 WALLBURG	NC		27373	3	201406 POSTAL	M
5	102139	550851180		1HOPE ASSEMBLY OF C	1922	HENLEY R SANFORD	NC		27330	9	196408 POINTADDRESS	M
6	104610	561873061		1ST NATIONAL ROCK OI	PO BOX 6997	ROCKY MOUNT	NC		27802	3	200307 POSTAL	M
7	116118	810745154		222 INTERNATIONAL	9801	DURANT F RALEIGH	NC		27614	3	201605 POINTADDRESS	M
8	111172	582526882		4 OUTDOORSMEN MINI	800 FAIRVIEW F	ASHEVILLE	NC		28803	3	200012 POINTADDRESS	M
9	87341	460962106		4GIVEN AND FREE MINI	6242 RUFFIN LN	CONOVER	NC		28613	9	199112 POINTADDRESS	M
10	33907	200637923		70X7 MINISTRIES	6804 SIX FORK	RALEIGH	NC		27615	3	200501 POINTADDRESS	M
11	88100	462501395		902 CHURCH	2009 CAREY RC	KINSTON	NC		28501	3	201311 POINTADDRESS	M
12	106076	562562182		A C T S MINISTRIES	6507 RAIN CREI	CHARLOTTE	NC		28262	3	200710 POINTADDRESS	M
13	34669	202093357		A COACHING REVOLUT	2321 FALLS RIV	RALEIGH	NC		27614	3	200503 POINTADDRESS	M
14	55986	320244173		A FAMILY CALLED CHUF	PO BOX 2871	HICKORY	NC		28603	3	200904 POSTAL	M
15	105536	562185132		A GIFT OF LOVE MINIST	PO BOX 30713	CHARLOTTE	NC		28230	3	200302 POSTAL	M
16	97620	522451192		A HAND UP	6212 KNIGHTDA	WENDELL	NC		27591	3	200505 POINTADDRESS	M
17	105658	562220729		A LIGHT IN THE DARKNI	PO BOX 118	UNION GROVE	NC		28689	3	200410 POSTAL	M
18	104816	561946756		A NEW BEGINNING T	700 W ENGLISH HIAH	POINT	NC		27262	3	200201 POINTADDRESS	M
19	105037	562033110		A NEW BEGINNING ASS	898 NEIL MAXW	RAEFORD	NC		28376	9	196408 POINTADDRESS	M
20	56147	320418032		A NEW BEGINNING MIN	PO BOX 112	FOUR OAKS	NC		27524	3	201408 POSTAL	M
21	115060	770677901		A NEW BEGINNING OUT	532 BILTMORE	MOORESVILLE	NC		28115	3	200905 POINTADDRESS	M
22	116173	810842197		A NEW DAY - A NEW HO	6377 OLD STALI	STALEY	NC		27355	3	201601 POINTADDRESS	M
23	92688	475324344		A NEW DAY CHURCH OI	10435 KERNS R	HUNTERSVILLE	NC		28078	3	201708 POINTADDRESS	M
24	105430	562158827		A NEW HAITI	PO BOX 19974	CHARLOTTE	NC		28219	3	200103 POSTAL	M
25	49251	271861057		A PLACE CALLED HOPE	1809 GARNER	RALEIGH	NC		27603	3	201108 STREETADDRE	M
26	103792	561497697		A PLACE IN THE HEART	4221 BECKERDI	SOPHIA	NC		27350	3	198705 POINTADDRESS	M

M	N	O	P	Q	R	S	T	U	V	W	X	Y
MATCH_ADDR	ADDR_TYPE	ADDNUM	SIDE	STPREDIR	STPRETYPE	STNAME	STTYPE	STADDR	CITY_2	COUNTY	POSTAL	X
3401 BROUGHT POINTADDRESS		3401	L	NOT AVAILABLE	NOT AVAILABLE	BROUGHTON	DR	3401 BROUGHT DURHAM	DURHAM		27704	-78.802246
303 VINTAGE HI POINTADDRESS		303	L	NOT AVAILABLE	NOT AVAILABLE	VINTAGE HILL	CIR	303 VINTAGE HI APEX	WAKE		27539	-78.778485
27373, WALLBU POSTAL		NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	WALLBURG	NOT AVAILABLE	27373	-80.137273
1922 HENLEY R POINTADDRESS		1922	L	NOT AVAILABLE	NOT AVAILABLE	HENLEY	RD	1922 HENLEY R SANFORD	LEE		27330	-79.256908
27802, ROCKY I POSTAL		NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	ROCKY MOUNT	NOT AVAILABLE	27802	-77.781242
9801 DURANT F POINTADDRESS		9801	L	NOT AVAILABLE	NOT AVAILABLE	DURANT	RD	9801 DURANT F RALEIGH	WAKE		27614	-78.577952
800 FAIRVIEW F POINTADDRESS		800	R	NOT AVAILABLE	NOT AVAILABLE	FAIRVIEW	RD	800 FAIRVIEW F ASHEVILLE	BUNCOMBE		28803	-82.50635
6242 RUFFIN LN POINTADDRESS		6242	R	NOT AVAILABLE	NOT AVAILABLE	RUFFIN	LN	6242 RUFFIN LN CONOVER	CATAWBA		28613	-81.224138
6604 SIX FORKS POINTADDRESS		6604	R	NOT AVAILABLE	NOT AVAILABLE	SIX FORKS	RD	6604 SIX FORKS RALEIGH	WAKE		27615	-78.639489
2009 CAREY RC POINTADDRESS		2009	R	NOT AVAILABLE	NOT AVAILABLE	CAREY	RD	2009 CAREY RC KINSTON	LENOIR		28501	-77.59327
6507 RAIN CREI POINTADDRESS		6507	L	NOT AVAILABLE	NOT AVAILABLE	RAIN CREEK	WAY	6507 RAIN CREI CHARLOTTE	MECKLENBURG		28262	-80.792755
2321 FALLS RIV POINTADDRESS		2321	R	NOT AVAILABLE	NOT AVAILABLE	FALLS RIVER	AVE	2321 FALLS RIV RALEIGH	WAKE		27614	-78.568877
28603, HICKORY POSTAL		NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	HICKORY	NOT AVAILABLE	28603	-81.39384
28230, CHARLO POSTAL		NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	CHARLOTTE	NOT AVAILABLE	28230	-80.833882
6212 KNIGHTDA POINTADDRESS		6212	R	NOT AVAILABLE	NOT AVAILABLE	KNIGHTDALE-E	RD	6212 KNIGHTDA WENDELL	WAKE		27591	-78.427496
28689, UNION G POSTAL		NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	UNION GROVE	NOT AVAILABLE	28689	-80.948624
700 W ENGLISH POINTADDRESS		700	L	W	NOT AVAILABLE	ENGLISH	RD	700 W ENGLISH HIGH POINT	GUILFORD		27262	-80.013066
898 NEILL MAXI POINTADDRESS		898	R	NOT AVAILABLE	NOT AVAILABLE	NEILL MAXWELI	RD	898 NEILL MAXI RAEFORD	HOKE		28376	-79.225819
27524, FOUR O POSTAL		NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	FOUR OAKS	NOT AVAILABLE	27524	-78.428945
532 BILTMORE I POINTADDRESS		532	L	NOT AVAILABLE	NOT AVAILABLE	BILTMORE	ST	532 BILTMORE I MOORESVILLE	IREDELL		28115	-80.812287
6377 OLD STALI POINTADDRESS		6377	R	NOT AVAILABLE	OLD	STALEY	RD	6377 OLD STALI STALEY	RANDOLPH		27355	-79.593581
10435 KERNS R POINTADDRESS		10435	L	NOT AVAILABLE	NOT AVAILABLE	KERNS	RD	10435 KERNS R HUNTERSVILLE	MECKLENBURG		28078	-80.878381
28219, CHARLO POSTAL		NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	NOT AVAILABLE	CHARLOTTE	NOT AVAILABLE	28219	-80.940407
1809 GARNER S STREETADDRESS		1809	R	NOT AVAILABLE	NOT AVAILABLE	GARNER STATI BLVD		1809 GARNER S RALEIGH	WAKE		27603	-78.65948
4421 BECKERDI POINTADDRESS		4421	L	NOT AVAILABLE	NOT AVAILABLE	BECKERDITE	RD	4421 BECKERDI SOPHIA	RANDOLPH		27350	-79.908193

FINAL PROCESS:

With our final dataset in hand, we now set out to manually reconcile matching values against the Wikidata query results, with the duplicates all removed. This dataset had unique 258 rows, which told us that of our close to 7,000 Homeland values, we would only be able to reconcile a maximum of 3.69% of the data points. With this in mind, we began going through each row in the Wikidata query data sheet and checking if the place matched with a value in our other set. If a match was found, the Wikidata link was added to Column C - WIKIDATA. With the time we had available, we were only able to complete reconciliation upto letter K in the Homeland Dataset - which was a total of 3494 rows. Out of these, only 20 were found to have a match with Wikidata Results, which is a mere success rate of .57%.

ADDITIONAL DATA NOTES:

We found the manual data reconciliation process extremely time-consuming, and with a very low success rate to match. One reason for this was the existence of multiple organizations of

the same name or multiple locations of one organization. For example, there were 25 different entries for “AGLOW INTERNATIONAL,” however these were not duplicates, they were unique locations of one organization (Image below).

	A	B	C	D	E	F	G	H	I	J	K	L
1	OBJECTID	EIN	WIKIDATA	NAME	STREET	CITY	STATE	ZIP	AFFILIATION	RULING	LOC_NAME	GEOCODED_ST
110	57379	341275248		AGAPE WORD CHURCH	8220 COOSA CT	RALEIGH	NC	27616	3	197905	POINTADDRESS	M
111	36142	205292468		AGLOW INTERNATIONAL	1000 PARKVIEW	LUMBERTON	NC	28358	9	197309	STREETADDRESS	M
112	43030	237409850		AGLOW INTERNATIONAL	2109 MILLPINE	RALEIGH	NC	27614	9	197309	POINTADDRESS	M
113	46140	260253609		AGLOW INTERNATIONAL	7 HEATHERFOR	DURHAM	NC	27704	9	197309	POINTADDRESS	M
114	47140	262673024		AGLOW INTERNATIONAL	PO BOX 2427	SANFORD	NC	27331	9	197309	POSTAL	M
115	49856	273218028		AGLOW INTERNATIONAL	PO BOX 29510	CHARLOTTE	NC	28229	9	197309	POSTAL	M
116	50267	274243835		AGLOW INTERNATIONAL	1001 GRAND ST	CONCORD	NC	28027	9	197309	POINTADDRESS	M
117	50492	275047076		AGLOW INTERNATIONAL	124 SCALES ST	ROCKINGHAM	NC	28379	9	197309	POINTADDRESS	M
118	85608	453678795		AGLOW INTERNATIONAL	2905 BENJAMIN	RALEIGH	NC	27610	9	197309	POINTADDRESS	M
119	86065	454723774		AGLOW INTERNATIONAL	105 BROOKME	STATESVILLE	NC	28625	9	197309	POINTADDRESS	M
120	87270	460817229		AGLOW INTERNATIONAL	2107-A KOONCE	GOLDSBORO	NC	27530	9	197309	STREETADDRESS	M
121	88426	463259114		AGLOW INTERNATIONAL	3118 GENLEE D	DURHAM	NC	27704	9	197309	POINTADDRESS	M
122	93584	510215431		AGLOW INTERNATIONAL	105 BROOKME	STATESVILLE	NC	28625	9	197309	POINTADDRESS	M
123	93603	510223542		AGLOW INTERNATIONAL	5309 BALLESTE	HOPE MILLS	NC	28348	9	197309	POINTADDRESS	M
124	93604	510223544		AGLOW INTERNATIONAL	2631 POLO LN	MAIDEN	NC	28650	9	197309	POINTADDRESS	M
125	93606	510223611		AGLOW INTERNATIONAL	222 GLENN BRII	ARDEN	NC	28704	9	197309	POINTADDRESS	M
126	93654	510237927		AGLOW INTERNATIONAL	409 W PHIFER	MONROE	NC	28110	9	197309	POINTADDRESS	M
127	93656	510238719		AGLOW INTERNATIONAL	4104 TRANQUIL	DURHAM	NC	27713	9	197309	POINTADDRESS	M
128	104903	561984193		AGLOW INTERNATIONAL	2329 BERTIE DF	RALEIGH	NC	27610	9	197309	POINTADDRESS	M
129	117105	813819989		AGLOW INTERNATIONAL	2161 EDENDER	STATESVILLE	NC	28625	9	197309	POINTADDRESS	M
130	120573	930764029		AGLOW INTERNATIONAL	8 GENLEE CT	DURHAM	NC	27704	9	197309	POINTADDRESS	M
131	120591	930797100		AGLOW INTERNATIONAL	2803 PARK WES	ALBEMARLE	NC	28001	9	197309	POINTADDRESS	M
132	120604	930815309		AGLOW INTERNATIONAL	8 GENLEE CT	DURHAM	NC	27704	9	197309	POINTADDRESS	M
133	120607	930821961		AGLOW INTERNATIONAL	3500 MALLARD	CHARLOTTE	NC	28269	9	197309	POINTADDRESS	M
134	120633	930906059		AGLOW INTERNATIONAL	3418 OLD MOUNT	MOUNT OLIVE	NC	28365	9	197309	POINTADDRESS	M

Next, we also found Church names that were a perfect match between the Wikidata Query Result of Churches and the Homeland Dataset. However, taking a deeper look, it was found that these locations were in different cities, meaning they were not a map. We actually found a good amount of values in the Wikidata list that were not present in the Homeland list, even though the latter seemed extremely comprehensive.

Another thing to note was more name discrepancies. In addition to different name conventions mentioned earlier in this report, we also found that some names were simplified in one dataset vs. the other. One such example is the Centre Presbyterian Church, Session House and Cemeteries, which is simply listed as “Centre Presbyterian Church” in the Homeland dataset. Another is the Coddle Creek Associate Reformed Presbyterian Church, Session House and Cemetery, which is listed in the homeland dataset as “Coddle Creek A R P Church.”

MOVING FORWARD:

With more time at hand, our group would have continued the manual reconciliation through all of the entries. This would also allow us to find other exceptions in the data that made the automatic reconciliation to fail. Other discrepancies in data, and places of improvement are all areas we would have liked to cover.

For groups who may attempt to continue work on this project, we would suggest creating a list of Wikidata query results more comprehensive than just churches. An overarching list of places of worship from many different religions would allow for further reconciliations to be made. We would also recommend that they look for other data sources beyond the ones we found to help broaden the scope of the project all while increasing accuracy. Finding other sources for a comprehensive list of the places of worship, much like the DHS one we used would also allow them to decide the accuracy of the DHS one. In addition, since our final list of churches also includes religious camps and retreats, though without further information such as addresses, coordinates, etc. as included for the other places of worship, spending time working on these camps and retreats should also be valuable. Furthermore, it could also prove helpful to reach out to the concerned authorities to bring to their attention how inadequate Wikidata is in including the places of worship, particularly in NC. This would also be a great place to present this report to display the inadequacies as found by the conclusion of this project.

CREDIT & ACKNOWLEDGEMENTS:

Each group member was assigned sections of the project to both present, and complete the report on. The sections, as split up by the person who completed them, are below.

ALLIE BARTON: Original Idea, Original Data, normalization of dataset

ALLY CHURCHILL: Introduction, Data Manipulation, Problems Encountered

ARSH SANGHA: Final Process & Product, Additional Data Notes, Moving Forward