

CASE-STUDY

Does Gender Play a Role in the Importance of Religious Identity in 16-Year-Olds in Northern Ireland?

Brigitta
Mafalda Timóteo
Nathan
Yiwen Zhang
Zosya Trimbacher

Table of Contents

Introduction and Why the Research Matters	2
1. Locating and Reviewing the Data File and Documentation	2
2. Files Used and Their Purpose	3 - 4
3. Understanding Variable Types and Levels of Measurement	4 - 5
3.1 Nominal Variables	5
3.2 Ordinal Variables	5
3.3 Scale Variables	5
3.4 Why Variable Type Matters	5
4. Understanding the Dataset Structure	6
4.1 Variables Relevant to the Research Question	6 - 7
5. Coding Scheme and Categories	7
5.1 Gender Variable (rsex)	7
5.2 Religious Identity Importance Variable (relidimp)	7
6. Missing Values	7 - 8
7. Survey Description and Data Collection	8 - 9
Data Cleaning and Preparation	9 - 10
Statistical Analysis and Results	10 - 11
Figure 1. Distribution of Religious Identity Importance	12
Figure 2. Importance of Religious Identity by Gender	12
Figure 3. Gender Difference in Religious Identity Importance	13
Figure 4. Importance of Religious Identity by Place of Residence	13
Interpretation and Discussion of Results	14
Impact of Findings	14 - 15
Limitations	15
Conclusion	15
References	15 – 16

Introduction and why the research matters

This study aims to investigate whether gender plays a role in the importance of religious identity among 16-year-olds in Northern Ireland.

This research question is of interest because religious identity has historically played an important role in shaping social, political, and community relations in Northern Ireland and continues to influence social attitudes today. Understanding whether males and females differ in the importance they place on religious identity may contribute to a better understanding of identity formation and social attitudes among young people.

This research question may be of interest to researchers studying identity and religion, as well as educators, policymakers, and organisations working with young people in Northern Ireland. The analysis uses data from the Young Life and Times (YLT) Survey, a repeated cross-sectional survey of 16-year-olds in Northern Ireland conducted between 2003 and 2012. The survey records the views of young people on a range of issues, including community relations, politics, sectarianism, health, and education.

Review of the Dataset and Documentation

All required data and documentation files were located and reviewed before analysis. The YLT.dta file provides the main dataset, while the Data Dictionary provides the essential information needed to identify variables, categories, measurement levels, and missing value codes. The User Guide explains the survey background, sampling design, data collection method, weighting and missing value conventions. The Readme file provides information on UK Data Archive processing standards and data conversion, while the UKDA Study Information file provides official study metadata and citation information.

For this research question, the key variables are rsex and relidimp. The variable rsex measures gender and is nominal. The variable relidimp measures the importance of religious identity and is mainly ordinal. The dataset inspection confirmed that rsex includes 3,753 male respondents, 5,373 female respondents and 42 missing cases. The variable relidimp includes several valid response categories, as well as missing or non-substantive codes: 9 = Don't know, 99 = Missing and 999 = Not asked. These missing values should be excluded or recoded before the main analysis.

This review ensures that the data are interpreted correctly and that the statistical analysis is based only on valid and meaningful response categories.

1. Locating and Reviewing the Data File and Documentation

Before conducting the analysis, the required data file and supporting documentation for the Young Life and Times Survey, 2003–2012: Teaching Dataset were located and reviewed. This step was necessary to ensure that the dataset was interpreted correctly before any statistical analysis was carried out.

The dataset used in this project is the YLT teaching dataset, which combines selected variables from the 2003–2012 Young Life and Times surveys. The Young Life and Times Survey is a repeated cross-sectional survey of 16-year-olds in Northern Ireland, covering topics such as community relations, health, politics, sectarianism, religion, identity and education. The documentation explains that the 2003–2012 surveys were merged to create one teaching dataset, containing a subset of common variables across the survey years.

The documentation was reviewed for four main purposes:

1. To identify the variables relevant to the research question;
2. To understand the coding schemes and response categories;
3. To identify missing value codes and decide how they should be handled;
4. To understand the survey design, sampling method and data collection process.

2. Files Used and Their Purpose

File	Purpose
YLT.dta	This is the main dataset used for the analysis. It contains the survey responses from the Young Life and Times Survey teaching dataset. The UK Data Archive Data Dictionary (ylt200312_ukda_data_dictionary.rtf) reports that the dataset contains 9,168 cases/respondents and 25 variables, which was later confirmed by loading the file in Python. This file is the basis for all later statistical analysis.
ylt200312_ukda_data_dictionary.rtf	This is the key file for identifying variables, categories, and missing values. It provides variable names, variable labels, measurement levels, value labels and user-missing value definitions. It was used to identify the two main variables for this research question: rsex, which measures gender, and relidimp, which measures the importance of religious identity.
7548_userguide.pdf	This file provides the broader context of the dataset. It explains the aims of the YLT survey, the structure of the teaching dataset, the sampling design, fieldwork process, response rate, weighting, confidentiality and missing value conventions. It was particularly useful for understanding how the data were collected and why some responses are coded as missing, ‘Don’t know’ or ‘Not asked’.

read7548.pdf / read7548.htm	This file provides UK Data Archive processing notes. It explains that the data were processed to the UK Data Archive's A* standard, meaning that checks were carried out on the number of cases and variables, variable labels, value labels, valid coding ranges and confidentiality requirements. This file was used to understand the quality checks and data conversion process applied to the dataset.
UKDA Study Information	This file provides the official metadata for the study, including the study title, study number, principal investigators, data collectors, sponsors, distributor, citation details, copyright information and file update date. It was used to confirm the official source and citation information for the dataset.

3. Understanding Variable Types and Levels of Measurement

Before identifying the variables for the analysis, it is important to understand the different types of variables used in the dataset. In this dataset, each variable represents one piece of information collected from respondents, such as gender, year of survey, religion, or the importance of religious identity. Each row represents one respondent, while each column represents a variable.

The codebook includes a column called Level of Measurement, which explains how each variable should be interpreted. This is important because the level of measurement affects how variables can be summarised, compared and analysed. The YLT codebook classifies variables mainly as nominal, ordinal, or scale variables.

3.1. Nominal Variables

A nominal variable is a categorical variable where the categories have no natural order. The numbers used in the dataset are only codes, not mathematical values.

For example, the variable rsex is nominal:

Code	Meaning
1	Male
2	Female
99	Missing

Here, 1 = Male and 2 = Female are simply category labels. The number 2 does not mean “more” or “higher” than 1. Therefore, gender should be treated as a categorical variable rather than a numerical variable.

Other examples of nominal variables in the dataset include religion, anyrelig, protcath, relschl, and ninatid.

3.2 Ordinal Variables

An ordinal variable is also categorical, but its categories have a meaningful order. However, the distance between categories is not necessarily equal.

For example, the variable `relidimp` asks:

How important is your religious identity to you?

It is coded as:

Code	Meaning
1	Very important
2	Quite important
3	Neither important nor unimportant
4	Not very important
5	Not at all important
6	I don't have a religious identity
9	Don't know
99	Missing
999	Not asked

This variable is mainly ordinal because the valid importance categories move from high importance to low importance. However, code 6 = I don't have a religious identity is slightly different from the 1–5 importance scale, so it should be interpreted carefully. Codes 9, 99, and 999 should not be treated as ordinary categories because they represent non-substantive or missing responses.

Other examples of ordinal variables in the dataset include `placeliv` and `welloff`.

3.3 Scale Variables

A scale variable is numerical and represents a quantity or count. For example, `yearsni` records how many years the respondent has lived in Northern Ireland. This is different from nominal or ordinal variables because the numerical value has a direct quantitative meaning. However, not every numeric-looking variable should be treated as a scale variable. For example, `respondentID` is a numeric identifier, but it should not be analysed as a meaningful number. It is only used to identify respondents.

3.4 Why Variable Type Matters

Understanding variable types is important because it affects the choice of statistical method. For example, since `rsex` is nominal and `relidimp` is ordinal/categorical, the analysis should treat both variables as categorical rather than continuous numerical variables. This also means that frequency tables and cross-tabulations are appropriate for describing the data, and missing or non-substantive codes should be removed before analysis.

In this research question, the key variables are therefore:

Variable	Meaning	Type	Reason
rsex	Gender	Nominal	Categories have no natural order
relidimp	Importance of religious identity	Ordinal	Categories follow an ordered importance scale

This classification ensures that the variables are interpreted correctly and that the analysis is based on their actual measurement level rather than simply treating the numeric codes as ordinary numbers.

4 Understanding the Dataset Structure

The YLT.dta file was loaded and inspected using Python.

The output showed that the dataset contains:

Dataset feature	Result
Number of cases/respondents	9,168
Number of variables	25

The variables included in the dataset are: respondentID, year, rsex, yearsni, placeliv, thisoct, oct2yrs, typeschl, relschl, welloff, sexatt, relarea, anyrelig, religion, relidimp, ninatid, natidimp, protcath, rrelago, rrelfut, relgalwy, mxrlgngh, mxrlgwrk, ownmxsch, and ownreligion.

This confirms that the dataset has the same number of variables and cases as stated in the Data Dictionary.

The dataset is a teaching dataset, meaning that it does not include every original YLT survey question. The user guide states that the teaching dataset contains selected common variables from the 2003–2012 surveys and that some variables have been simplified, recoded or constructed for teaching purposes.

4.1 Variables Relevant to the Research Question

The research question is:

Does gender play a role in the importance of religious identity in 16-year-olds in Northern Ireland?

Based on this research question, the two main variables required are:

Variable name	Variable label	Role in analysis	Measurement level
rsex	Q1. Are you male or female?	Independent variable: gender	Nominal

relidimp	Q45. How important is your religious identity to you?	Dependent variable: importance of religious identity	Ordinal
-----------------	---	--	---------

The variable rsex is nominal because it represents categories of gender without a natural ranking. The variable relidimp is ordinal because most response options follow a ranked order from “Very important” to “Not at all important”.

5. Coding Scheme and Categories

5.1 Gender Variable: rsex

The variable rsex measures the respondent’s gender.

Code	Category	Frequency in dataset
1	Male	3,753
2	Female	5,373
99	Missing	42

The valid categories are 1 = Male and 2 = Female. The value 99 represents a missing response and should not be treated as a valid gender category.

5.2 Religious Identity Importance Variable: relidimp

The variable relidimp measures how important religious identity is to the respondent.

Code	Category	Frequency in dataset
1	Very important	1,849
2	Quite important	2,302
3	Neither important nor unimportant	1,680
4	Not very important	957
5	Not at all important	1,049
6	I don’t have a religious identity	451
9	Don’t know	19
99	Missing	75
999	Not asked	786

For this variable, codes 1–5 represent an ordered scale of importance, ranging from high importance to low importance. Code 6 is a substantive response category, meaning that the respondent states they do not have a religious identity. However, it does not fit neatly into the 1–5 ordinal scale of importance, so it should be interpreted carefully depending on the chosen analysis. Codes 9, 99, and 999 are non-substantive or missing-value codes and should not be treated as ordinary response categories.

6. Missing Values

The user guide explains that missing values can occur for several reasons. Respondents may not answer a question, may not know the answer, or may not have been asked a particular question in a specific survey year. In the YLT teaching dataset, missing values are generally coded as follows:

Code	Meaning
9	Don't know
99	Missing / not answered
999	Not asked

For the variables used in this research question, the missing value treatment is:

Variable	Missing or non-substantive values to exclude
sex	99 = Missing
relidimp	9 = Don't know; 99 = Missing; 999 = Not asked

These values should be excluded or recoded as missing before the main statistical analysis. This is important because including them as ordinary numerical values would distort the results. For example, if 999 = Not asked were treated as a real score in the religious identity variable, it would incorrectly appear much larger than the valid response categories and would seriously bias any analysis.

7. Survey Description and Data Collection

The data used in this case study comes from the Young Life and Times Survey (YLT), a survey focused on young people living in Northern Ireland. The YLT records the views of 16-year-olds on a range of social and political issues, including community relations, health, politics, sectarianism and education. This makes the survey relevant for studying young people's attitudes towards identity, religion and society in Northern Ireland.

The dataset used for this project is the Young Life and Times Survey, 2003–2012: Teaching Dataset. The surveys from 2003 to 2012 were merged into one teaching dataset, containing a subset of common variables from across the time series. The dataset was created to support the teaching and learning of quantitative research skills. It is important to note that this is a teaching dataset, meaning that some variables have been simplified, recoded or constructed for teaching purposes. The user guide also states that, for research purposes, it is advisable to use the original surveys rather than the teaching dataset.

The YLT is a repeated cross-sectional survey. This means that the survey was carried out across different years, but it did not follow the same individuals over time. Instead, different samples of 16-year-olds were surveyed in different years. Therefore, this dataset can be used to compare patterns across groups or survey years, but it should not be interpreted as tracking individual changes over time.

The teaching dataset contains 9,168 cases in total, with the number of respondents varying by year between 2003 and 2012. The documentation also states that the dataset is not weighted. This means that each respondent is treated equally in the analysis, and no weighting adjustment has been applied to make the sample match the wider population distribution. For this reason, the findings should be interpreted as results from this teaching dataset rather than as fully population-weighted estimates for all 16-year-olds in Northern Ireland.

The sampling frame for the YLT survey was the Child Benefit Register, which contained information on young people living in Northern Ireland who had celebrated their 16th birthday during the relevant survey year. This register was used to identify eligible participants for the survey.

The fieldwork was carried out through a mail-out process managed by an independent mailing company. Eligible young people first received an advance letter introducing the survey and giving them the opportunity to opt out. Those who did not opt out were then sent a second letter containing the questionnaire, instructions and a return envelope. A reminder letter with a second questionnaire was sent after 10 days to those who had not responded.

Although postal completion was the main method of data collection, the 2012 survey introduced a mixed-mode approach, allowing respondents to complete the survey by post, online or by telephone. However, post remained the most common mode of response. In 2012, 86% of respondents returned the survey by post, while the remaining respondents completed it online and no respondents completed it by telephone.

This dataset is useful for the present research question because it includes variables related to both gender and religious identity. In particular, the codebook includes RSEX, which records whether the respondent is male or female, and RELIDIMP, which records how important religious identity is to the respondent. These variables allow the analysis to explore whether gender is associated with the importance of religious identity among 16-year-olds in Northern Ireland.

Overall, the YLT teaching dataset provides an appropriate source for this case study because it contains relevant information on gender and religious identity, while also offering a structured dataset suitable for statistical analysis. However, because the dataset is unweighted and designed for teaching purposes, the results should be interpreted carefully and should avoid broad causal or population-level claims.

Data Cleaning and Preparation

The original Stata dataset was imported into Python using pandas and contained 9,168 respondents and 25 variables. A copy of the original dataframe was created before cleaning to preserve the source data.

Responses that were not suitable for the main statistical analysis were then removed. For the gender variable (rsex), cases recorded as “Missing” were excluded. For religious identity importance (relidimp), the non-substantive responses “Don’t know”, “Missing”, and “Not asked” were removed. The valid response “I don’t have a religious identity” was retained because it represents a meaningful answer rather than missing data.

After filtering, unused categorical levels were removed from both variables. No values were imputed, and the original 25 columns were retained. The final analytical dataset contained 8,255 respondents and was exported as a CSV file for use in the statistical analysis and visualisations.

Statistical Analysis and Results

Statistical Analysis

The statistical analysis was conducted in Python using the `pandas` and `scipy.stats` libraries. The cleaned dataset contained 8,255 respondents. All variables included in the analysis were treated as categorical.

Pearson’s chi-square tests of independence were used to examine whether statistically significant associations existed between the selected variables. For each analysis, a contingency table was created using `pd.crosstab()`, followed by the `scipy.stats.chi2_contingency()` function.

For every test, the hypotheses were:

Null hypothesis (H_0): The two variables are independent.

Alternative hypothesis (H_1): The two variables are associated.

A significance level of $\alpha = 0.05$ was used. When the p-value was below 0.05, the null hypothesis was rejected. No effect-size measure was calculated in the notebook; therefore, the results below report only the chi-square statistic, degrees of freedom, and p-value. No additional filtering or recoding was performed within the statistical analysis notebook.

Results

A statistically significant association was found between gender and the importance of religious identity, $\chi^2(5, N = 8,255) = 53.10, p < 0.001$. The null hypothesis of independence was therefore rejected.

Religious identity importance also differed significantly between survey years, $\chi^2(40, N = 8,255) = 152.49, p < 0.001$.

Religion was significantly associated with the social-relations indicator represented by ownreligion, $\chi^2(16, N = 8,255) = 662.72, p < .001$. A statistically significant association was also found between religion and national identity, $\chi^2(24, N = 8,255) = 4,709.34, p < 0.001$.

Perceived financial situation was significantly associated with religious identity importance, $\chi^2(35, N = 8,255) = 97.00, p < 0.001$.

The relationship between survey year and the social-relations indicator was also statistically significant, $\chi^2(32, N = 8,255) = 312.73, p < 0.001$.

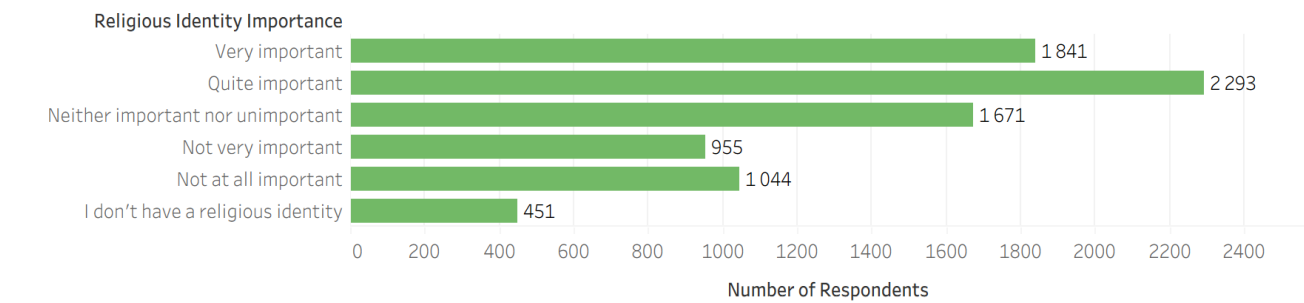
Finally, place of residence was significantly associated with religious identity importance, $\chi^2(30, N = 8,255) = 199.16, p < 0.001$.

Variables examined	χ^2	df	p-value
Gender × religious identity importance	53.10	5	3.2047523704 249576e-10
Survey year × religious identity importance	152.49	40	4.8633058013 049894e-15
Religion × social-relations indicator	662.72	16	1.0975480506 646718e-130
Religion × national identity	4,709.34	24	0.0
Financial situation × religious identity importance	97.00	35	9.8398504572 2986e-08
Survey year × social-relations indicator	312.73	32	8.5054168133 90168e-48
Place of residence × religious identity importance	199.16	30	7.1033250899 8077e-27

Overall, the null hypothesis of independence was rejected for all seven analyses.

Figure 1. Distribution of Religious Identity Importance

Distribution of Religious Identity Importance



Distinct count of Respondent ID for each Religious Identity Importance.

Figure 1 shows the overall distribution of responses for the importance of religious identity. The largest group of respondents considered their religious identity quite important, followed by very important, while the smallest group reported that they did not have a religious identity.

Figure 2. Importance of Religious Identity by Gender ([Tableau](#))

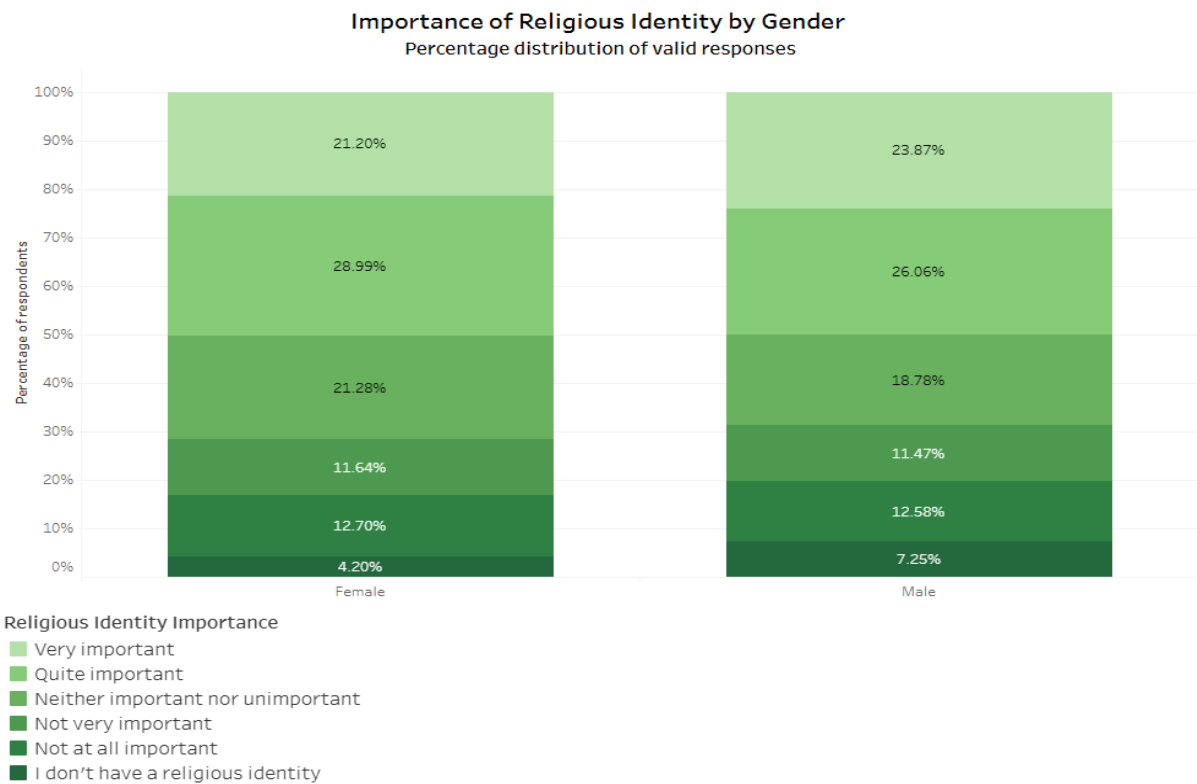
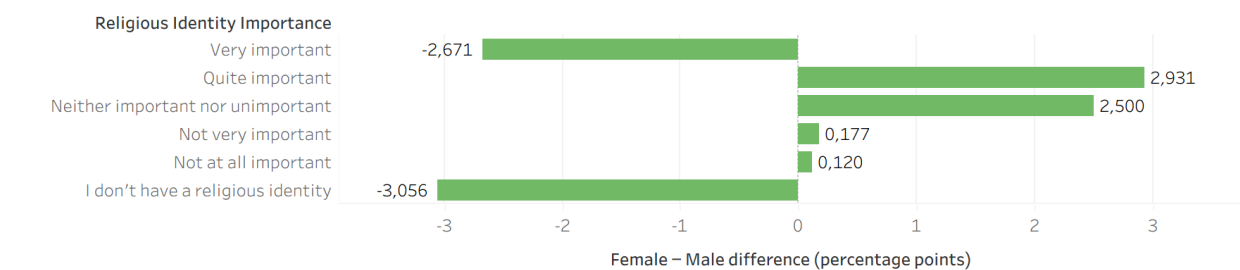


Figure 2 shows the percentage distribution of religious identity importance for male and female respondents. The chart illustrates that the distributions are broadly similar, although some differences can be observed between the two groups.

Figure 3. Gender Difference in Religious Identity Importance

Gender Difference in Religious Identity Importance

Positive values indicate a higher percentage among females and negative values indicate a higher percentage among males.



Sum of Gender Percentage-Point Difference for each Religious Identity Importance. The marks are labeled by sum of Gender Percentage-Point Difference.

Figure 3 highlights the percentage-point differences between female and male respondents for each response category. The largest differences were observed for Quite important, Very important, and I don't have a religious identity.

Figure 4. Importance of Religious Identity by Place of Residence (Valid Responses) ([Tableau](#))

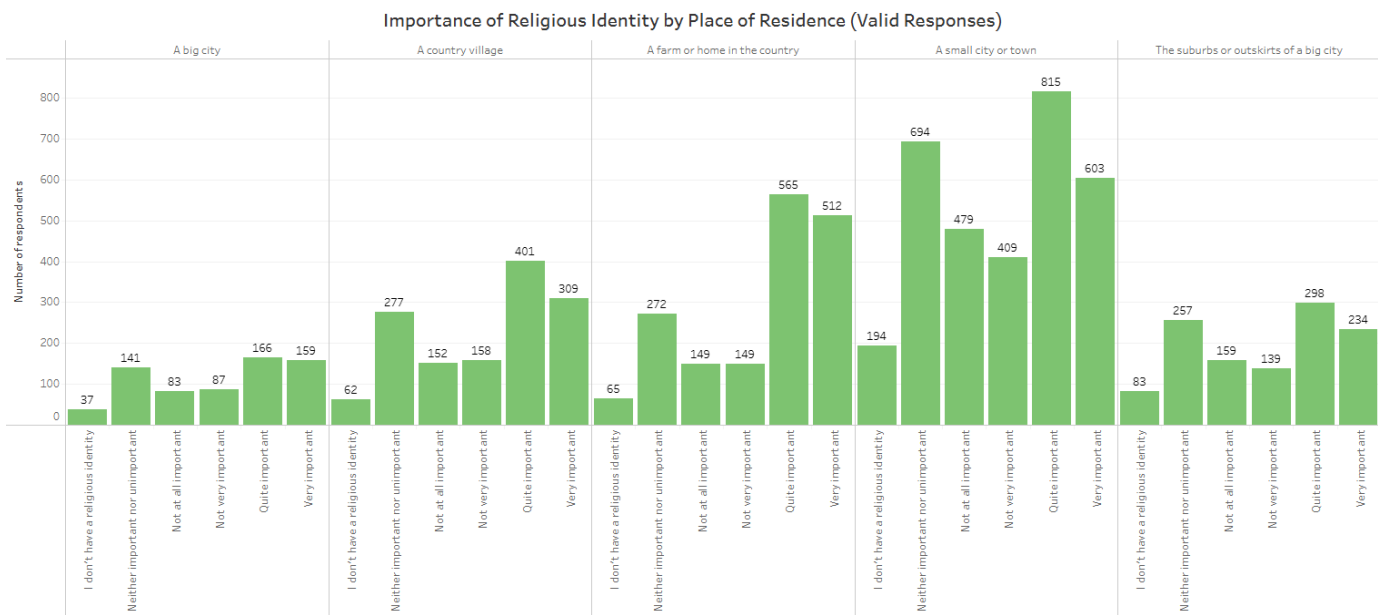


Figure 4 indicates that respondents living in small cities or towns make up the largest group in the sample. Respondents from farms or homes in the country and country villages also make up substantial parts of the sample, while fewer respondents came from big cities and the suburbs or outskirts of a big city.

Interpretation and Discussion of Results

The Chi-square test showed a statistically significant relationship between gender and the importance of religious identity ($\chi^2 = 53.10$, $df = 5$, $p < 0.001$). This means that the null hypothesis was rejected, suggesting that gender and the importance of religious identity are associated among respondents in this dataset.

Figure 2 shows that there are some differences between male and female respondents, although they are fairly small. A slightly higher percentage of males said that their religious identity was very important (23.87%) compared with females (21.20%). Males were also more likely to report that they did not have a religious identity (7.25% compared with 4.20%).

Female respondents were slightly more likely to choose quite important (28.99%) and neither important nor unimportant (21.28%) than male respondents. The percentages for not very important and not at all important were very similar for both groups.

Although the result is statistically significant, the differences between males and females are relatively small. This suggests that gender is associated with the importance of religious identity, but it is probably only one of several factors that influence how respondents answered this question.

Figure 4 shows the number of respondents in each place of residence category. Most respondents lived in small cities or towns, followed by farms or homes in the country. Fewer respondents came from big cities and the suburbs or outskirts of a big city. This chart provides additional information about the sample, but because it shows counts rather than percentages, it should not be used to conclude that one place of residence has higher or lower levels of religious identity than another.

Impact of Findings

The findings suggest that religious identity continues to be important for many 16-year-olds in Northern Ireland. At the same time, the responses show that attitudes towards religious identity are not the same for everyone. While many respondents considered religion to be quite important or very important, others reported that it was not important or that they did not have a religious identity.

These findings may be useful for researchers studying identity, religion, and young people. They may also be useful for educators and policymakers who work with young people from different backgrounds. Overall, the results give a better understanding of how young people in Northern Ireland view religious identity.

However, the results should be interpreted with caution. The differences between males and females were statistically significant, but they were relatively small. Other factors, such as family

background, religion, education, or place of residence, may also influence how important religious identity is to an individual.

Limitations

This study has several limitations.

The dataset combines survey responses collected between 2003 and 2012, so the findings may not represent the views of young people today.

The survey is a repeated cross-sectional survey. This means that different groups of 16-year-olds were surveyed each year rather than following the same individuals over time.

Only gender and religious identity were analysed in this study. Other factors that may influence religious identity were not included in the analysis.

Finally, the Chi-square test only shows whether there is an association between the variables. It does not show that gender causes differences in the importance of religious identity.

Conclusion

This study investigated whether gender plays a role in the importance of religious identity among 16-year-olds in Northern Ireland.

The analysis found a statistically significant association between gender and the importance of religious identity. This means that the null hypothesis was rejected, suggesting that gender and the importance of religious identity are associated among respondents in this dataset.

Although the differences between males and females were statistically significant, they were relatively small. This suggests that gender is only one factor related to religious identity and that other factors may also influence how young people respond.

Overall, the findings show that religious identity remained important for many young people included in the Young Life and Times Survey, while also showing that attitudes towards religion varied between respondents.

References

ARK. (2014). *2003–12 Young Life & Times Survey Teaching Resources*. Queen's University Belfast and Ulster University.

Devine, P. and Calvert, E. (2014). *Young Life & Times Survey, 2003–2012: Teaching Dataset*. UK Data Archive.

Hayward, K. and McManus, C. (2021). *Northern Ireland: Society and Culture*. Queen's University Belfast. Pages 2–7.

Young Life & Times Survey User Guide (UK Data Archive Study 7548).
