

Visualising Identity: A Study of Sexual Orientation and Religion Across England and Wales

Data

The data for this project was downloaded from the Office for National Statistics website. The original file (which can be found in the GitHub repository for this project) includes the code book and metadata on sheets on the corresponding sheets.

The data was a product of the 2021 census – the sheets downloaded contained other demographic data that were collected alongside sexual orientation at the point of the census including: age, sex, employment status, smoking habits and general health.

Research question

The decision to focus on religion and sexual orientation came after a recent discussion that ensued after sharing religious school experiences with friends and colleagues. The question being that, even if one found themselves of a sexual orientation that was deemed anathema by their church/religion, would they still identify with that religion? (which inevitably led to many other questions, outlined below under follow-up). Some churches and religious branches are now very progressive, but a religion, in its very nature, cannot be. Before scrolling through the ONS website I proposed that the data would represent this with a higher percentage of LGBTQIA+ groups identifying as not religious than the heterosexual population.

Project script

Packages and importing the dataset I developed a loop to automatically check the installation of necessary packages, ensuring the script can operate seamlessly on any system without manual adjustments. I first established the *packages* object, which includes the names of all required packages for the project script.

```
packages <- c("here", "readxl", "tidyverse", "shiny", "vcd", "viridis", "scales", "plotly")
```

The *for* loop iterates over each package vector element, using the variable *pkg* to hold the current package name in each iteration.

```
for (pkg in packages) {  
  if (!require(pkg, character.only = TRUE, quietly = TRUE)) {  
    install.packages(pkg)  
    library(pkg, character.only = TRUE)  
  }  
}
```

Inside the loop, the *require()* function checks if the package named by *pkg* is installed and can be loaded. The *character.only = TRUE* argument specifies that *pkg* is a character string (the name of the package). While the *quietly = TRUE* argument suppresses warnings and other messages (to keep the console output

clean as it's checking for multiple packages). If *require()* returns *FALSE* (the package is not available), *install.packages(pkg)* is called to install the package.

After installing necessary packages, *library()* is used to load all of them into the session. The *character.only = TRUE* argument again, indicates that *pkg* is a character string.

The following code then defines the path to the xlsx file using the *here()* function. After which sheet 2a of the file is read into the session skipping the top 4 rows and first column then defining which columns are text/numeric.

```
excel_file <- here("data", "sexualorientationfurtherpersonalcharacteristicsenglandandwalescensus2021.xlsx")

sexualorientationreligion <- read_excel(excel_file,
                                         sheet = "2a",
                                         col_types = c("skip", "text", "text", "text", "text", "text", "text", "text"),
                                         skip = 4)
```

```
## Warning: Expecting numeric in G426 / R426C7: got '[c]'
```

Show entries

Search:

Raw Data

	Sex [note 1]	Age group [note 2]	Sexual orientation	Religion	Population estimate [note 3]	Percentage estimate of group [note 3] [note 4]
1	People	All ages 16 years and over	All usual residents	Buddhist	245508	0.51
2	People	All ages 16 years and over	All usual residents	Christian	23656561	48.71
3	People	All ages 16 years and over	All usual residents	Hindu	823151	1.69
4	People	All ages 16 years and over	All usual residents	Jewish	213619	0.44
5	People	All ages 16 years and over	All usual residents	Muslim	2664722	5.49

Showing 1 to 5 of 540 entries

Previous 2 3 4 5 ... 108 Next

Preparing the data Once I had checked the dataset to ensure that the correct columns/rows had been imported and all observations were there, I then began cleaning and filtering. The console was displaying

the following warning.

```
Warning message:
Expecting numeric in G426 / R426C7: got '[c]'
```

Consulting the ONS metadata I discovered that '[c]' appeared in the cells that were suppressed because they had a value below the threshold. To correct this I created a cleaned dataframe where all cells not including [c] were retained while those including [c] were excluded.

```
sexualorientationreligion_clean <- sexualorientationreligion[!apply(sexualorientationreligion, 1, funct
```

Following that I renamed the variables, see code below.

```
sexualorientationreligion_clean <- rename(sexualorientationreligion_clean,
                                           c( "sexual_orientation" = "Sexual orientation",
                                               "age" = "Age group [note 2]" ,
                                               "sex" = "Sex [note 1]",
                                               "percentage" = "Percentage estimate of group \r\n[note 3]
```

I then used the *dplyr* filtering function to create a new dataframe including only the data that was necessary for the final plot.

```
filtered_data <- sexualorientationreligion_clean %>%
  filter(sexual_orientation != "All usual residents",
         age == "All ages 16 years and over",
         sex == "People")
```

filtered_data is the new dataframe created to store the result of *dplyr*'s filtering operation. This new dataframe contains only the rows from *sexualorientationreligion_clean* that meet the specified conditions. The pipe operator passes the *sexualorientationreligion_clean* dataframe into the next function, *filter()* which specifies the following conditions:

- *sexual_orientation != "All usual residents"* excludes the columns under *sexual_orientation* containing the listed value including only the breakdown of sexual orientations rather than the total of all.
- *age == "All ages 16 years and over"* ensures that only rows where the age column exactly matches "All ages 16 years and over" are retained, excluding the columns breaking the data down by age group.
- *sex == "People"* filters to include only rows where the sex column has the value "People", excluding any rows that might have other breakdowns by sex.

Filtered data

Show entries

Search:

Final Dataframe

	sex	age	sexual_orientation	Religion	Population estimate [note 3]	percentage
1	People	All ages 16 years and over	Straight or Heterosexual	Buddhist	209211	0.48
2	People	All ages 16 years and over	Straight or Heterosexual	Christian	22205800	51.16
3	People	All ages 16 years and over	Straight or Heterosexual	Hindu	747501	1.72
4	People	All ages 16 years and over	Straight or Heterosexual	Jewish	185960	0.43
5	People	All ages 16 years and over	Straight or Heterosexual	Muslim	2410350	5.55

Showing 1 to 5 of 45 entries

Previous

1

2

3

4

5

...

9

Next

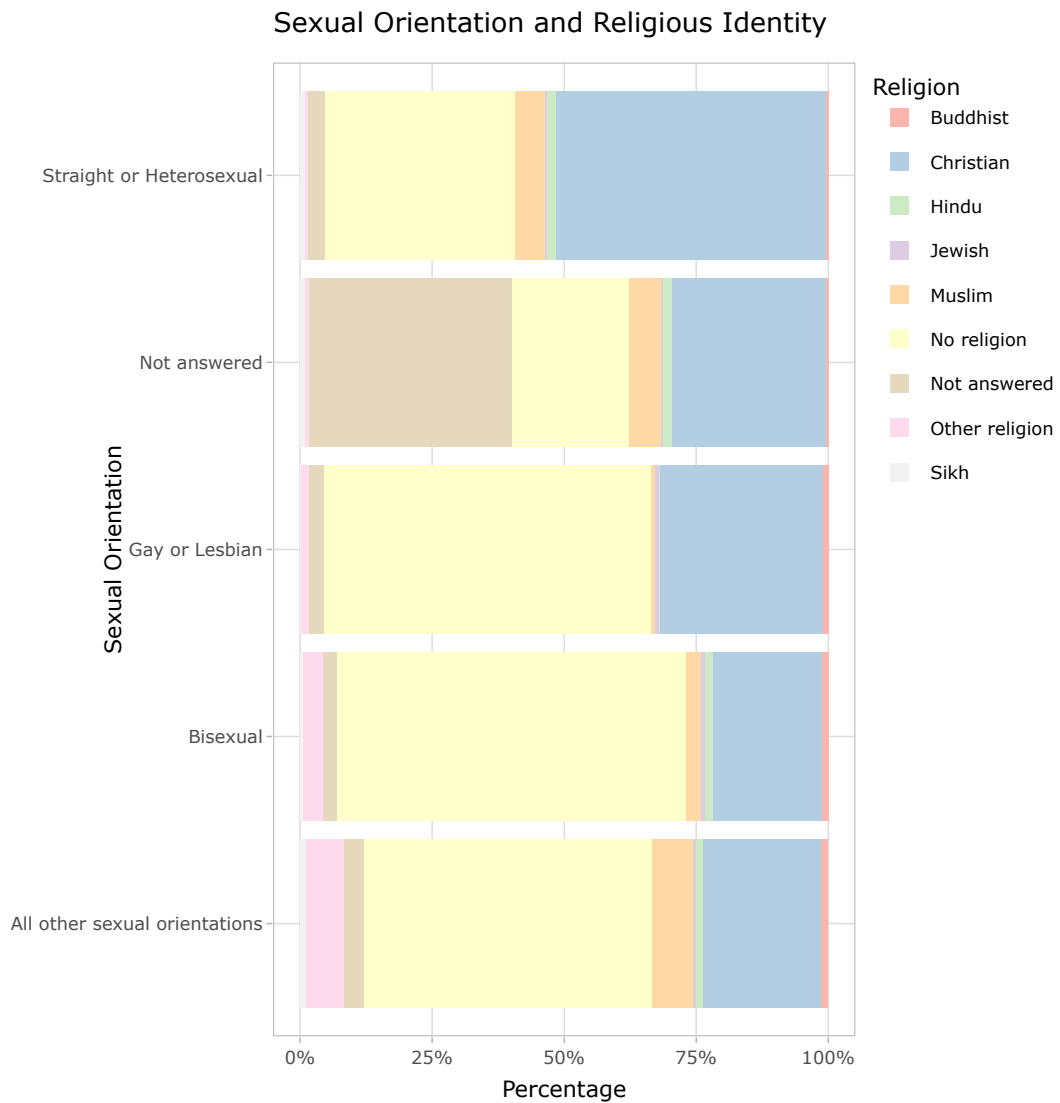
Plot

Creating the plot The bars in the chart are stacked so the different religious categories for each sexual orientation are piled on top of one another to show the composition within each group. The chart's orientation was flipped using `coord_flip()`, making the bars horizontal instead of vertical to make the chart easier to read. To improve aesthetic and readability, I chose a light theme and a colour palette called “*Pastell1*,” which provides subtle yet distinct colours for each religious group. Additionally, the legend, to identify which colours correspond to which religions, is positioned at the bottom of the chart to avoid obscuring any data. I decided after trialling labels on the bars to use the *plotly* library to convert it instead into an interactive visualisation, thus retaining the information without cluttering the plot with too many labels. This transformation allows users to interact with the visualisation directly by hovering over bars to see detailed data labels.

```
p <- ggplot(filtered_data, aes(x = sexual_orientation, y = percentage / 100, fill = Religion, text = pa
  geom_bar(stat = "identity", position = "stack") +
  coord_flip() +
  labs(title = "Sexual Orientation and Religious Identity",
        x = "Sexual Orientation",
        y = "Percentage") +
  scale_y_continuous(labels = percent_format()) +
  scale_fill_brewer(palette = "Pastell1") +
  theme_light() +
  theme(legend.position = "bottom")

interactive_plot <- ggplotly(p, tooltip = "text")

interactive_plot
```



Interpretation The plot clearly demonstrates that among straight or heterosexual individuals, a majority identify as Christian (51.16%), with a substantial group reporting no religious affiliation (35.92%). In

contrast, when we look at individuals identifying as gay or lesbian, there is a noticeable decrease in Christian affiliation (30.77%) compared to heterosexuals, and a significant increase in those reporting no religion (61.92%). This trend is even more pronounced among bisexual individuals, with only 20.76% identifying as Christian and the majority, 65.96%, reporting no religious affiliation. This pattern continues with the other sexual orientation category, where Christian affiliation stands at 22.24%, and a little over half, 54.56%, identify with no religion at all.

These findings align with the initial hypothesis that LGBTQIA+ individuals might be less likely to identify with organized religions, especially Christianity, and more likely to report no religious affiliation, which can be reflective of the historical relationship between religious institutions and the LGBTQIA+ community.

Future work and follow-up questions

The 2021 ONS Census data provides a snapshot of the intersection between sexual orientation and religious identity in England and Wales. It appears that straight individuals tend to align more with Christianity, whereas those who identify as gay, lesbian, or bisexual are more inclined to indicate no religious ties.

Looking ahead, it would be interesting to unpack the reasons behind these patterns. Questions arise about whether cultural acceptance, personal beliefs, or the openness of religious communities play a role in these choices. Using the same dataset it would be useful to investigate the potential generational shifts in attitude towards religion and sexuality by visualising the data with a breakdown of age groups.

Some other follow-up questions worth exploring using the same ONS data:

- Could there be geographic differences within England and Wales that affect the relationship between religion and sexual orientation?
- How might age or the passing of time influence religious identity within the LGBTQIA+ community?