# ENVS450 Social Survey Analysis. Assignment 1: Examining the Association between Age and Religiosity in 21st-Century Britain

**201442927** *University of Liverpool*

For this assignment we examine the association between age and religion, exploring a 2012 Labour Force Survey to see what we can infer about religiosity and secularization in Britain. We offer a univariate description of each variable, before recoding `Religion` as a binary indicator of religiosity. We consider the impact of missing data, and reweight our figures accordingly. The data is then tabulated and visualized, and the confidence levels considered. The analysis is then extended into a third dimension of `Sex`. Pearson's chi-squared test is used to demonstrate the statistical significance of the relationship we find between female-ness and religiosity, which is also tabulated and shown graphically. We conclude with a box-plot visualizing the association between age, religiosity, and sex in Britain.

*Keywords*: Religiosity; Secularisation; Post-Christendom; Post-Christian; Exploratory Data Analysis

### Introduction

The Labour Force Survey is 'the largest household study in the UK' (Office for National Statistics, 2019). As part of the Social Survey Analysis module of our Data Analytics and Society course, we were given a copy of the dataset from QLFS 2012 (Quarter 1), and assigned the task of inferring and describing the association between two variables of our choice. My aim in this exercise is to gain an understanding of the basic principles of statistical social analysis, to demonstrate competence in using the free programming environment R (Ihaka and Gentleman, 1996), and perhaps also to make some comment about British society in the 2010s.

### Examining the Data

An initial examination of the data shows that `qlfs` is a *tidy* (as defined by Wickham, 2014, "Each variable forms a column, Each observation forms a row") table of 84,692 observations, with 28 variables. Four of these are administrative ID values (`Case_ID`, `Person_ID`, `HHold_ID`, `Family_ID`); seven relate directly to work (`WorkStatus`, `NSSEC`, `HoursWorked`, `GrossPay`, `NetPay`, `TravelMode`, `TravelTime`); two to education (`HighestQual` and `DegreeClass`); five to family situation ( `FamilySize`, `FamilyType`, `YoungestChild`, `FamilyToddlers`, `MaritalStatus`) and three to household living situation (`HHoldSize`, `LastMoved`, `Tenure`). We are also told `Sex`, `Age` (which is divided for us into four *ad hoc* `AgeGroup` levels: "16-29", "30-44", "45-64", and "65+"), `CountryOfBirth`, `EthnicGroup`, and `Religion`.

There are zero *NA* responses for `Age`, `Sex`, `GovtRegion`, `MaritalStatus`, `FamilySize`, `FamilyType`, `FamilyToddlers`, and `YoungestChild` (though we must note that respondents with "no child aged <19" have been given the value "-6" for this variable; similarly, non-working adults have been given the value "-7" for the `GrossPay`, `NetPay` and `HoursWorked` variables). There are a small number (61) of *NA* responses for `Tenure`; a significant number (4-10%) for `EthnicGroup` (4473), `CountryOfBirth` (3564), `Religion` (7658), `WorkStatus` (4432), `NSSEC` (4635), `HighestQual` (5090), `DegreeClass` (5085), `HoursWorked` (4432), and `LastMoved` (4457); and a large number (>40%) for `GrossPay` (39064), `NetPay` (39241), `TravelMode` (34737), and `TravelTime` (36244).

`Religion` and `Age` (and by the end, `Sex` too) shall be the focus of our exploratory data analysis.

### Univariate Description

#### *`Religion`*

`Religion` is a categorical nominal factor with 8 levels. In order of frequency, from highest to lowest, these are "Christian" (49,724), " No Religion" (21,723), "Muslim" (2572), "Hindu" (1017), "Any other Religion"

(945), "Sikh" (439), "Jewish" (311), and "Buddhist" (303). Clearly the *central tendency* (ie. for a categorical variable, the *mode*) is "Christian".

The *proportion misclassified* is over two-thirds (67.7%). Note that this is defined here as the *index of dissimilarity* between this dataset when compared with "a theoretical uniform distribution" (Williamson, 2019), not "the true cross-classification" (as for example the phrase means in Cleave et al., 1995).

#### `Age`

Age is a continuous numeric variable, with a minimum value of 16 and a maximum value of 99. Both of these are necessary results of the survey methodology: the Labour Force Survey only focuses on "people aged 16 years and over"(Office for National Statistics, 2019), so it should be impossible to find any respondent aged less than 16; and inspection of `str(qlfs$Age)` shows that the top value of `99` actually means "99 and over". This could be problematic if our analysis were to focus on the very elderly, but as we will be looking at ages across the whole range this need not concern us unduly.

The *central tendency* (in this case, both the *median* and, to two significant figures, the *mean*) is 46. In our survey sample the most common age (the *mode*) happens to be 41. The lower quartile is 32, the upper quartile is 60, and the interquartile range is therefore 28. The skew of the distribution is 0.118, so we can consider it *symmetric*. This is as we might intuitively expect with the mean and median aligned, though Von Hippel (2005) demonstrates "a surprising number of exceptions" to the rule of thumb that "In a skewed distribution, the mean is farther out in the long tail than is the median"(McCabe and Moore, 2003). The standard deviation is 17.5.

### Missing Data

We have already noted that `Age` received a perfect response rate, with zero *NA* values. `Religion`, however, contains 7,658 *NA* responses, which would make it the third most frequent response if counted as a response, with almost three times as many as "Muslim".

Investigation reveals that 3,124 of these responses have "Northern Ireland" as their `GovtRegion` – that is, 100% of the responses from Northern Ireland include no religious data. Since religion in Northern Ireland has a particular significance quite distinct from the rest of the United Kingdom (Mitchell, 2006), it is very unlikely that *the conditional independence assumption* would be valid, so we will not pretend that we can extrapolate from the results of our other data to give representatively reweighted responses for Northern Ireland – instead we will restrict our analysis to Great Britain.

We should also be aware that those included under "Any other Religion" probably include people on opposite ends of the 'religious' spectrum: on the one hand, serious affiliates of the South Asian religion Jainism (Shah, 2014); on the other, perhaps some of those 390,000 people who stated their religion as 'Jedi' in the census (White, 2014). "Situated among different modern interests, such as Christian interreligious polemic, European colonial denials of indigenous religion, and Enlightenment critiques of original religion, this crisis of authenticity has been central to the problem of religion in the modern world" (Chidester, 2003, p.73). However, for the purposes of this analysis we are going to include all responses of "Any other Religion" as 'religious'.

### Recoding

The `qlfs` dataset comes with `Age` already recoded into four *ad hoc* `AgeGroup` sets, but the divisions between these do no reflect equal intervals or equal data. So we have also recoded `Age` into ten deciles: "16-22", "23-29", "30-35", "36-41", "42-46", "47-51", "52-57", "58-63", "64-70" and "71+". Because `Age` is reported imprecisely as an integer number of years, these deciles are not perfectly equal, but this need not concern us very much.

As our analysis concerns the question of secularisation and religiosity, we have recoded the eight `Religion` categories to just two: "Religious" and "Not Religious".

**Reweighting**

Having recoded our variables, we then reweight them by Age Decile to account for the possible impact of non-random response bias. This process is known as *post-stratification*. Since `religion` is the only one of our examined variables that includes *NA* responses, that is the only one that we have to account for. (Note that while we account for *non-response bias*, we do not here make any attempt to correct for possible *response bias* (Lavrakas, 2008)).

In general, the younger the decile, the greater the non-response: the 16-22 group had the highest rate of non-response to the question of religion and was assigned a weighting of 1.135; the 71+ group had the lowest rate and was assigned a weighting of 1.015. But this tendency is not absolutely true: for example, the 36-41 group had a lower rate of non-response than the 47-51 group and so was given a lower weighting.

Having obtained the weightings, the `ENVS450::tally()` function (Williamson, 2019) makes it trivial to use them to correct subsequent exploration of the data. All we have to do is add a simple extra argument to tell the function to use our newly created `Weight` variable:

```
tally(~Religion, data=britain, weights="Weight")
```

**Results**

It is clear from our analysis that in 2012 older people were on the whole more likely to identify themselves as religious. This could be because of a cohort effect (different types of which are examined by Keyes et al., 2010), in which case one might argue from these results that Britain is becoming increasingly non-religious.

Or it could be because of a connection between older age and religious affiliation (as explored by Malone and Dadswell, 2018), in which case it is likely that many of those who responded that they had "No Religion" in 2012, may subsequently gain some sort of religious affiliation. If we were to weigh these two possibilities against each other it would be helpful to compare the 2012 dataset with others allowing us a longitudinal view of what change develops across time.

As responsible statisticians we must also establish the possibility that although there is may be a real correlation between the two variables of age and religiosity, there may not be a causal relationship in either direction. And of course, one should also notice that in the youngest Age Decile, among those aged 16-22, religiosity is actually on the increase – so it is somewhat reductive toconclude that lower religiosity is strictly correlated with younger age.

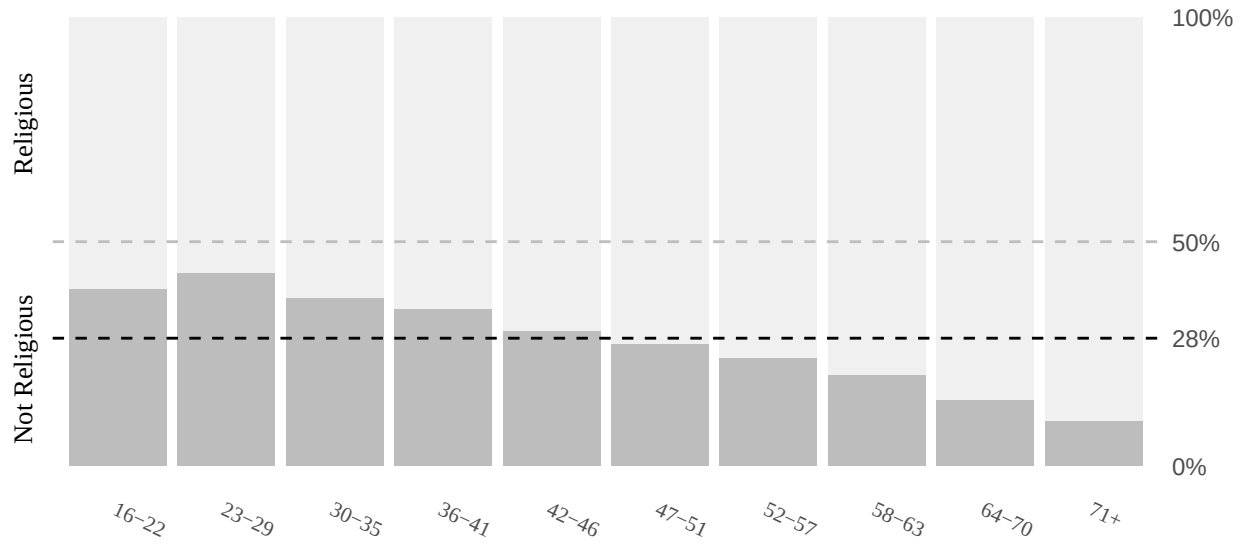


Figure 1: Visualisation of Religiosity by Age Decile.

Table 1: British Religiosity by Age Decile

| Age Deciles | Religious (%) | Not Religious (%) | Total Number |
|---|---|---|---|
| 16-22 | 60.43 | 39.57 | 8836 |
| 23-29 | 56.92 | 43.08 | 8566 |
| 30-35 | 62.46 | 37.54 | 8000 |
| 36-41 | 64.87 | 35.13 | 8670 |
| 42-46 | 69.79 | 30.21 | 7797 |
| 47-51 | 72.7 | 27.3 | 7537 |
| 52-57 | 75.88 | 24.12 | 8136 |
| 58-63 | 79.74 | 20.26 | 7981 |
| 64-70 | 85.36 | 14.64 | 8633 |
| 71+ | 89.89 | 10.11 | 7412 |
| Total | 71.51 | 28.49 | 81568 |

*Source:*
Calculations based on the Quarterly Labour Force Survey 2012.

**Confidence Intervals**

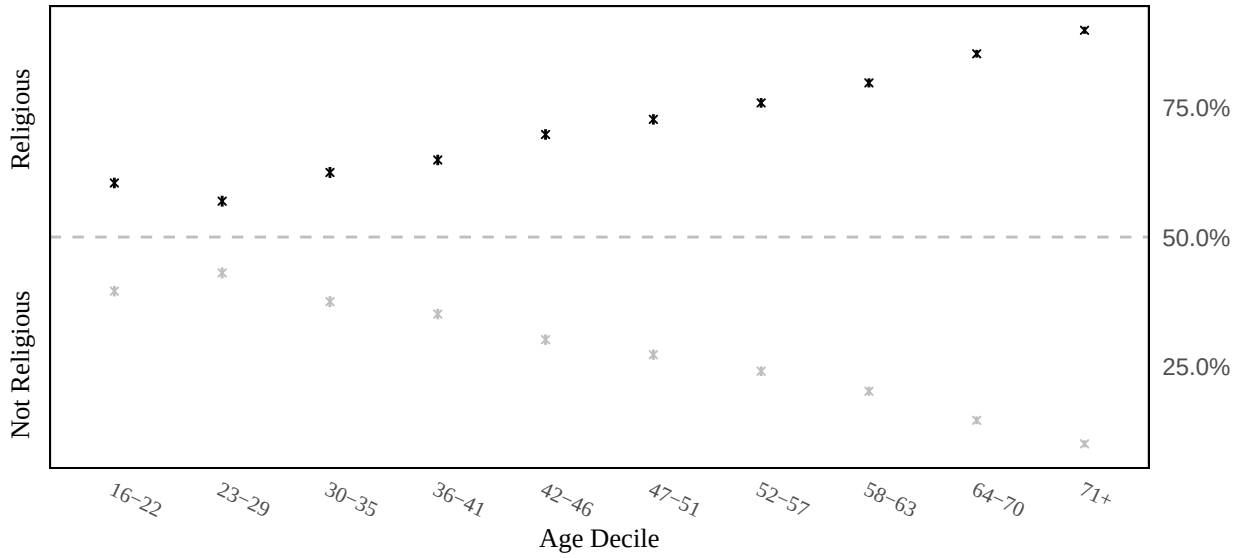


Figure 2: Confidence Intervals for Calculated Religiosity Percentages.

In Figure 2 we have visualized the 95% confidence levels of each point, with a (in the end very small) line showing the range within which the true population value will lie in 95 out of every 100 survey samples, assuming that any difference between the sample estimate and the true population value is attributable to random sampling error alone. The confidence intervals of our points are a long way from overlap, which gives robust support to our analysis on visual inspection (Cumming and Finch, 2005).

Table 2: Religiosity by Sex

| | Male (%) | Female (%) | Total Number |
|---|---|---|---|
| Religious | 45.43 | 54.57 | 58326.2 |
| Not Religious | 53.11 | 46.89 | 23241.8 |
| Total | 47.62 | 52.38 | 81568 |

*Source:*
Quarterly Labour Force Survey 2012.

**Adding `Sex` to the Analysis**

*Univariate Description*

Sex is a binary categorical variable, with a 100% response rate (i.e. there are no *NA* values for this variable in the dataset). Females account for a slightly greater percentage of the sample (51.76% compared to 48.24%), as is the case across the full population (as given by the 2011 UK Census Office for National Statistics (2012)). This disparity can be explained by the greater life expectancy of the female sex.

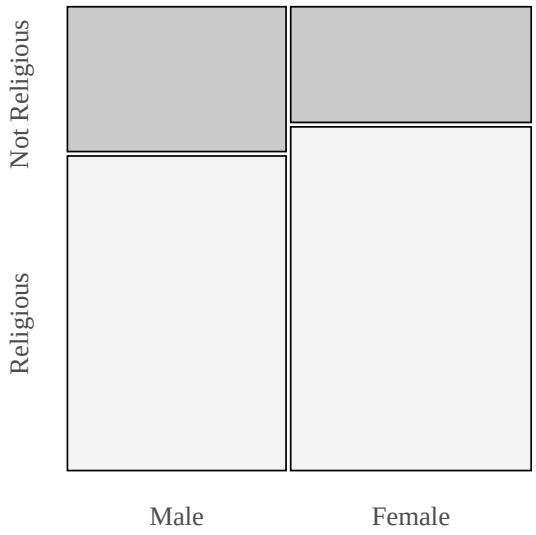


Figure 3: Religiosity by Sex

Figure 3 suggests that there is a correlation between being female and identifying as religious. By applying Pearson's $\chi^2$ test, we can see that this correlation is statistically significant ($\chi^2 = 548.78$, $p = 0.000$).

**Graphical Representation of Three Variables**

Our final visualization is a variable-width box-plot, in the style of McGill et al. (1978). "The width of each box has been made proportional to the square root of the number in the corresponding group"; the middle is fixed at the median; the upper and lower *hinges* are fixed at the upper and lower quartiles; and the upper and lower limits of the *whiskers* are fixed at the upper and lower observations within a distance of 1.5 times the interquartile range from the upper or lower hinge (Wickham et al., 2015).

It is easy to see and interesting to observe that while women are on average older than men in general, when the ages of men and women are compared according to whether or not they are religious, women are on average younger in both categories.

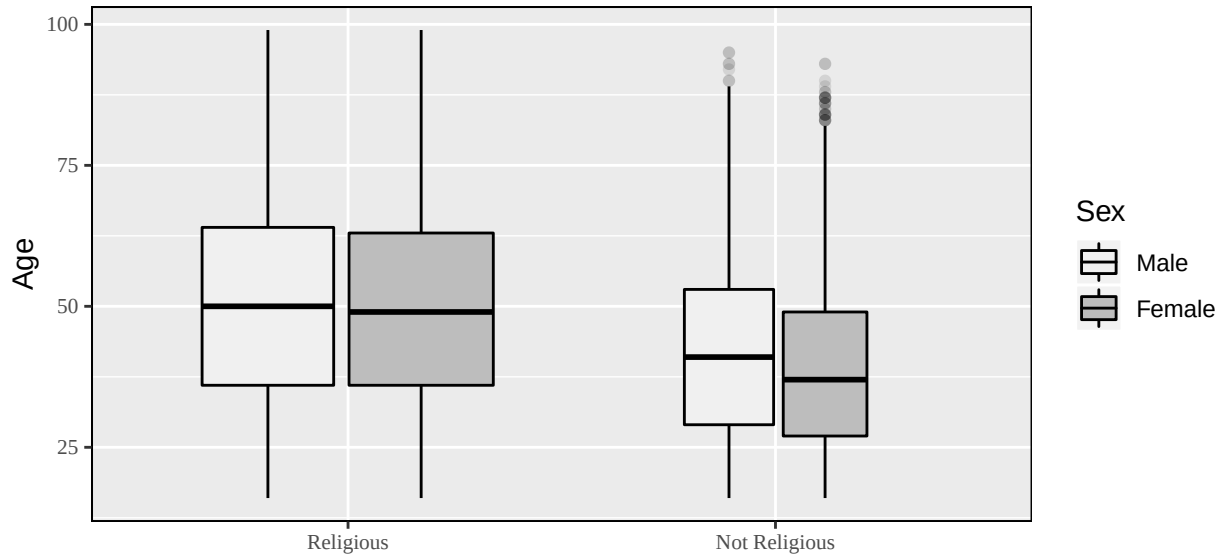


Figure 4: Boxplot showing interrelationship between Religion, Age, and Sex

**Concluding Unscientific Postscript**

Before we conclude our analysis and submit our assignment, we must swiftly set these statistics within the context of 21st Century Britain. While British religiosity may have declined, the institutions of religion continue to sublty shape society: Cranmer et al. (2006) map the various ways that "[c]urrent arrangements [between the state and Church of England] span something more than a merely vestigial residue of the former partnership, especially in the relationship with the sovereign as Supreme Governor of the Church and in episcopal membership of the House of Lord".

"Where God clings to our culture, to our routines of discourse, He is a phantom of grammar, a fossil embedded in the childhood of rational speech". Thus Steiner (1989) paraphrased the view of "Nietzsche (and many after him)" before going on to argue "the reverse", that any coherent account of human speech is "underwritten by the assumption of God's presence".

For some, like the Conservative Scruton (2001), the decline of British religiosity is seen a matter of elegiac sorrow, as "the Anglican Church lost sight of its sacramental character… [and] the Noncomformist churches vanished overnight, and now stand bleak and abandoned in all the towns and villages of England".

For others, like the humanist Copson (2011) the official countiing of religious affiliation is bemoaned as "erroneous numbers" that have "been repeatedly misused" to advocate "that greater public resources should be granted to religious organisations".

On the other hand, Paas (2016) surveys a variety of Christian groups who are untroubled by the decline in religiosity: "those who have been inspired by Anabaptism seem to be particularly attracted to the 'post-Christendom' perspective…", "[while] the 'post-modern' perspective is preferred… by advocates of so-called emerging churches".

Of course, a simple survey response professing some religious affiliation is obviously a very different thing to frequent participation in some particular religious practice. Brierley (2005) p.38 estimated that "over an average month, a total of 10.2% of the population come to church, of which 4.4% is those who come regularly each week and 5.8% is those who happened to come sometime that month".

However, humans being the social creatures that we are, and social proof serving the vital epistemological function that it does (Fallis, 2002), social surveys that ask people about religious subjects will invariably not only provide information about what society believes, but provide impetus to change what society believes. Even when people have just been given eight simple options and asked to tick a box, the subject of religion will always have the potential to disrupt, dismay, and perhaps even occasionally delight.

**References**


Brierley, P.W. (2005). *Painting by Numbers: An Introduction to Church Statistics*. London: Christian Research. ISBN 978-1-85321-158-4. oCLC: 63273648.

Chidester, D. (2003). *Fake Religion: Ordeals of Authenticity in the Study of Religion*. Journal for the Study of Religion. 16(2):pp. 71–97. ISSN 1011-7601.
https://www.jstor.org/stable/24764311

Cleave, N., Brown, P.J., and Payne, C.D. (1995). *Evaluation of Methods for Ecological Inference*. Journal of the Royal Statistical Society. Series A (Statistics in Society). 158(1):pp. 55–72. ISSN 0964-1998. doi:10.2307/2983403.
https://www.jstor.org/stable/2983403

Copson, A. (2011). *If you're not religious, tell the census so*. The Guardian.
https://www.theguardian.com/commentisfree/belief/2011/feb/28/census-religion-question

Cranmer, F., Lucas, J., and Morris, B. (2006). *Church and State: A Mapping Exercise*. London: Constitution Unit. ISBN 978-1-903903-47-6. oCLC: 64571350.

Cumming, G. and Finch, S. (2005). *Inference by eye: Confidence intervals and how to read pictures of data*. American Psychologist. pp. 170–180.

Fallis, D. (2002). *Introduction: Social Epistemology and Information Science*. Social Epistemology. 16(1):pp. 1–4. ISSN 0269-1728, 1464-5297. doi:10.1080/02691720210132752.
http://www.tandfonline.com/doi/abs/10.1080/02691720210132752

Ihaka, R. and Gentleman, R. (1996). *R A language for data analysis and graphics*. Journal of Computational and Graphical Statistics. 5(3):pp. 299–314.

Keyes, K.M., Utz, R.L., Robinson, W., and Li, G. (2010). *What is a cohort effect? Comparison of three statistical methods for modeling cohort effects in obesity prevalence in the United States, 1971–2006*. Social science & medicine (1982). 70(7):pp. 1100–1108. ISSN 0277-9536. doi:10.1016/j.socscimed.2009.12.018.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3469580/

Lavrakas, J., P. (2008). *Response Bias*.
https://methods.sagepub.com/reference/encyclopedia-of-survey-research-methods/n486.xml

Malone, J. and Dadswell, A. (2018). *The Role of Religion, Spirituality and/or Belief in Positive Ageing for Older Adults*. Geriatrics. 3(2). ISSN 2308-3417. doi:10.3390/geriatrics3020028.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6319229/

McCabe, G.P. and Moore, D.S. (2003). *Introduction to the Practice of Statistics*. W. H. Freeman. 4th ed. edn.

McGill, R., Tukey, J.W., and Larsen, W.A. (1978). *Variations of Box Plots*. The American Statistician. 32(1):pp. 12–16. ISSN 0003-1305. doi:10.2307/2683468.
https://www.jstor.org/stable/2683468

Mitchell, C. (2006). *Religion, Identity and Politics in Northern Ireland: Boundaries of Belonging and Belief*. Abingdon, UNITED KINGDOM: Routledge. ISBN 978-0-7546-8226-4.
http://ebookcentral.proquest.com/lib/liverpool/detail.action?docID=429796

Office for National Statistics (2012). *2011 Census*.
https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2
12-17

——— (2019). *Labour Force Survey*.
https://www.ons.gov.uk/surveys/informationforhouseholdsandindividuals/householdandindividualsurveys/labourf

Paas, S. (2016). *Church Planting in the Secular West: Learning from the European Experience*. The Gospel and Our Culture Series. Grand Rapids, Michigan: William B. Eerdmans Publishing Company. ISBN 978-0-8028-7348-4.

Scruton, R. (2001). *England: An Elegy*. London: Pimlico. ISBN 978-0-7126-6805-7. oCLC: 248367258.

Shah, B.V. (2014). *Religion in the Everyday Lives of Second-Generation Jains in Britain and the USA: Resources Offered by a Dharma-Based South Asian Religion for the Construction of Religious Biographies, and Negotiating Risk and Uncertainty in Late Modern Societies*. The Sociological Review. 62(3):pp. 512–529. ISSN 0038-0261. doi:10.1111/1467-954X.12177.
https://doi.org/10.1111/1467-954X.12177

Steiner, G. (1989). *Real Presences: Is There Anything in What We Say?*. London: Faber and Faber. ISBN 978-0-571-14071-8. oCLC: 246831287.

Von Hippel (2005). *Mean, Median, and Skew: Correcting a Textbook Rule*. Journal of Statistics Education. 13(2).
http://jse.amstat.org/v13n2/vonhippel.html

White, P. (2014). *The Census Form*.
https://www2.le.ac.uk/projects/social-worlds/all-articles/sociology/census

Wickham, H. (2014). *Tidy Data*. Journal of Statistical Software. 59(1):pp. 1–23. ISSN 1548-7660. doi:10.18637/jss.v059.i10.
https://www.jstatsoft.org/index.php/jss/article/view/v059i10

Wickham, H., Chang, W., and Henry, L. (2015). *A box and whiskers plot (in the style of Tukey) — geom_boxplot*.
https://ggplot2.tidyverse.org/reference/geom_boxplot.html

Williamson, P. (2019). *Helper functions for ENVS450*.
https://rdrr.io/github/DrPaulWilliamson/ENVS450/

# Appendix: Code Chunks

### Examining the Data

```
# load data
data_filename <- '../data/QLFS 2012 Q1 Adults (Assignment 1).RData'
load(data_filename)
# explore the data
dim(qlfs) # 84692 x 28
names(meta.data)
unique(qlfs$AgeGroup)
head(qlfs)
tail(qlfs)
summary(qlfs)
summary(qlfs$Religion)
summary(qlfs$Age)
# perfect response: Age, Sex, MaritalStatus,
## FamilySize, FamilyType, GovtRegion, YoungestChild, FamilyToddlers
# recoded responses: HoursWorked/GrossPay/NetPay '-7'; YoungestChild '-6'
attributes(qlfs$HoursWorked)
attributes(qlfs$GrossPay)
attributes(qlfs$NetPay)
attributes(qlfs$YoungestChild)
tally(~TravelMode, data=qlfs, format="percent", margin="joint")
```

### Univariate Description

*Religion*

```
# examine Religion
str(qlfs$Religion)
unique(qlfs$Religion)
mode(qlfs$Religion)
religion_df <- tally(~Religion, data=qlfs, format = "count", margin = "joint")
religion_df["Christian",1]
```

```
proportion.misclassified(qlfs$Religion)
```

*Age*

```
str(qlfs$Age)
summary(qlfs$Age)
mode(qlfs$Age)
skew(qlfs$Age)
```

```
var(qlfs$Age)
std.dev(qlfs$Age)
tally(~ AgeGroup, data=qlfs)
```

### Missing Data

```
# no 'religion' responses from N.I.
summary(qlfs$GovtRegion)
religion_df["Muslim",1]
religion.by.region <- tally(Religion~GovtRegion, data=qlfs)
religion.by.region[,"Northern Ireland"]

# restrict analysis to britain
britain <- subset(qlfs, GovtRegion!="Northern Ireland")
britain

# check
region_df <- tally(~GovtRegion, data=qlfs, format = "count", margin = "joint")
region_df
dim(qlfs)[1] - dim(britain)[1] - region_df["Northern Ireland",1]
```

### Recoding

```
# recode age into deciles
groups <- 10
seq(from= 0, to= 1, by= 1/groups)
cumulative_proportions <- seq(from= 0, to= 1, by= 1/groups)
breaks <- quantile( britain$Age, probs=cumulative_proportions , na.rm=FALSE )
britain$AgeDecile <- group.data(britain$Age, breaks= breaks)
levels(britain$AgeDecile)[10] <- "71+" # Reflect the fact that "99" actually means "99+"
tally( ~ AgeDecile, data= britain, format="count", margin="col")
summary(britain$ReligiousCategories)
```

```
# recode Religion
## keep original responses about 'Religion' as 'ReligiousCategories'
britain$ReligiousCategories <- britain$Religion
by_religion <- levels(britain$ReligiousCategories)

## also recode as No Religion/Religious, including NA as 'No Religion'
religious <- c("Religious", "Not Religious", "Invalid Response")
britain$Religion <- recode(britain$ReligiousCategories,
                           "by_religion[1]=religious[2];
                           by_religion[c(2,3,4,5,6,7,8)]=religious[1];
                          NA=religious[3]",
```

```
                        levels = c(religious)
                        )
tally(~Religion, data=britain, format="percent", margin="col")
proportion.misclassified(qlfs$Religion)
```

### Reweighting

```
valid.response <- subset(britain, britain$Religion!="Invalid Response")
valid.count <- tally(~AgeDecile, data=valid.response, format="count", margin="col")
valid.count
total.valid <- valid.count["Total",]
true.count <- tally(~AgeDecile, data=britain, format="count", margin="col")

df <- data.frame(
                  AgeDecile = c( levels(britain$AgeDecile), "Total"),
                  Valid=    as.vector( valid.count ),
                  True=     as.vector( true.count )
                )
df$Weight <- df$True / df$Valid

valid.response <- merge(valid.response, df[ , c("AgeDecile", "Weight") ],
                        by="AgeDecile")
valid.response<- valid.response[order(valid.response$Case_ID), ]
britain <- valid.response
tally(Weight ~ AgeDecile, data=britain)
df
tally(~ AgeDecile, data=britain, margin="col", weights="Weight")
true.count
britain$Religion <- recode(britain$Religion,
                        "by_religion[1]=religious[2];
                        by_religion[c(2,3,4,5,6,7,8)]=religious[1];
                        by_religion[3]=religious[2]",
                        levels = c("Religious", "Not Religious")
                        )
tally(~Religion, data=britain, format="count",margin="joint", weights="Weight")
levels(britain$Religion)
```

### Results

```
# table
religion_by_age <- tally(AgeDecile ~ Religion, data=britain,
                        format="percent", margin="row", weights="Weight")
total <-tally(AgeDecile ~ Religion, data=britain,
              format="count", margin="row", weights="Weight")
total.pct <- tally(AgeDecile~Religion, data=britain,
```

```
                    format="percent", margin="joint", weights="Weight")
total.count <- tally(AgeDecile~Religion, data=britain,
                     format="count", margin="joint", weights="Weight")
decile.labels <- labels(religion_by_age[,1])
religion_by_age[,"Total"] <- total[,"Total"]
frame <- cbind(decile.labels,religion_by_age)
summary_row <- c("Total", total.pct["Total","Religious"]
, total.pct["Total","Not Religious"]
, total.count["Total","Total"])
frame[,1:4]
reframed <- rbind(frame,summary_row)
new_column_labels <- c("Age Deciles", "Religious (%)",
                       "Not Religious (%)", "Total Number")
kable(reframed, col.names=new_column_labels, row.names=FALSE,
      caption = "British Religiosity by Age Decile",
      longtable = FALSE, booktabs = TRUE, format = "latex")
      %>% row_spec(10, hline_after=T)
      %>% footnote(general =
      "Calculations based on the Quarterly Labour Force Survey 2012.",
      general_title="Source:", fixed_small_size = TRUE, threeparttable = TRUE)
```

```
religion.pct.table <- tally(~Religion, data=britain,
                            format="percent", margin="col", weights="Weight")
total.pct.not <- religion.pct.table["Not Religious",]/100

ggplot(data=britain) +
  geom_bar( aes(x=AgeDecile, fill=Religion), position="fill") +
  geom_hline(yintercept = total.pct.not, colour="black", linetype="dashed") +
    geom_hline(yintercept = 0.5, colour="grey", linetype="dashed") +
  scale_y_continuous(labels=percent,
                     position="left",
                     breaks = c(0.0, 0.3219, 0.5, 1.0),
                     sec.axis = dup_axis()) +
  scale_x_discrete(labels=decile.labels) +
  xlab("Age Decile") +
  ylab("Not Religious                  Religious     ") +
  scale_fill_brewer(palette = "Greys") +
  theme(axis.text.x = element_text(angle = 335,
                                   hjust = 0,
                                   family="serif",
                                   size=8),
        axis.title = element_text(family="serif", size=10),
        panel.background = element_blank(),
        axis.ticks = element_blank(),
        axis.text.x.top = element_blank(),
        axis.text.y.left = element_blank(),
        axis.title.x.bottom = element_blank(),
```

```
        axis.title.y.right = element_blank(),
        legend.position = "none")
```

### Confidence Intervals

```
pct_age_religion <- tally(AgeDecile ~ Religion, data=britain,
                          format="percent", margin="row", na.rm=TRUE, weights="Weight")
count_age_religion <- tally(AgeDecile ~ Religion, data=britain,
                          format="count", margin="row", na.rm=TRUE, weights="Weight")
pct.religious <- pct_age_religion[, "Religious"]
pct.not_religious <- pct_age_religion[, "Not Religious"]
denominator <- count_age_religion[, "Total"]
religious.std.error <- ( (pct.religious * (100 - pct.religious) ) / denominator)^0.5
religious.lower.bound <- pct.religious - 1.96*religious.std.error
religious.upper.bound <- pct.religious + 1.96*religious.std.error
non.std.error <- ( (pct.not_religious * (100 - pct.not_religious) ) / denominator)^0.5
non.lower.bound <- pct.not_religious - 1.96 * non.std.error
non.upper.bound <- pct.not_religious + 1.96 * non.std.error
df <- data.frame(pct.religious, religious.std.error, religious.lower.bound, religious.upper.boun
df$AgeDecile <- rownames(df)

ggplot(data=df) +
  geom_point( aes(x=AgeDecile,
                  y=pct.religious/100),
              colour="black",
              fill="white",
              shape=4,
              size=1) +
  geom_point( aes(x=AgeDecile,
                  y=pct.not_religious/100),
              colour="grey",
              fill="grey",
              shape=4,
              size=1) +
  geom_segment( aes(x=AgeDecile, xend=AgeDecile,
                    y=religious.lower.bound/100,
                    yend=religious.upper.bound/100) ) +
  geom_segment( aes(x=AgeDecile,
                    xend=AgeDecile,
                    y=non.lower.bound/100,
                    yend=non.upper.bound/100), colour="grey") +
  geom_hline(yintercept = 0.50,
             colour="grey",
             linetype="dashed") +
  scale_y_continuous(labels=percent,
                     breaks = c(0.25,0.50,0.75),
```

```
                    position="left",
                    sec.axis = dup_axis()) +
  scale_x_discrete(labels=decile.labels) +
  xlab("Age Decile") +
  ylab("Not Religious                  Religious     ") +
  theme(axis.text.x = element_text(angle = 335,
                                   hjust = 0,
                                   family="serif",
                                   size=8),
        axis.title = element_text(family="serif", size=10),
        panel.background = element_blank(),
        axis.ticks = element_blank(),
        axis.text.x.top = element_blank(),
        axis.text.y.left = element_blank(),
        axis.title.y.right = element_blank(),
        legend.position = "none",
        panel.border = element_rect(colour = "black",
                                    fill = NA)
  )
```

### Adding `Sex` to the Analysis

*Univariate Description*

```
tally_sex<- tally(AgeDecile~Sex, data=britain,
                  format="percent", margin="col", weights="Weight")
women <- subset(britain, Sex=="Female")
men <- subset(britain, Sex=="Male")
mean(women$Age)/mean(men$Age)
```

```
# table2
religion_by_sex <- tally(Religion ~ Sex, data=britain,
                         format="percent", margin="row", weights="Weight")
total <-tally(Religion ~ Sex, data=britain,
              format="count", margin="row", weights="Weight")
total.pct <- tally(Religion ~ Sex, data=britain,
                   format="percent", margin="joint", weights="Weight")
total.count <- tally(Religion ~ Sex, data=britain,
                     format="count", margin="joint", weights="Weight")
religion_by_sex[,"Total"] <- total[,"Total"]
rel_labels <- labels(religion_by_sex[,1])
religion_by_sex[,"Total"] <- total[,"Total"]
frame <- cbind(rel_labels, religion_by_sex)
summary_row <- c("Total", total.pct["Total", "Male"],
                 total.pct["Total","Female"], total.count["Total","Total"])
reframed <- rbind(frame,summary_row)
```

```
new_column_labels <- c("", "Male (%)", "Female (%)", "Total Number")
kable(reframed, col.names=new_column_labels,
      row.names=FALSE, caption = "Religiosity by Sex",
      longtable = FALSE, booktabs = TRUE, format = "latex")
      %>% row_spec(2, hline_after=T) %>%
      footnote(general = "Quarterly Labour Force Survey 2012.",
          general_title="Source:", fixed_small_size = TRUE, threeparttable = TRUE)
```

```
# devtools::install_github("haleyjeppson/ggmosaic")
library(ggmosaic)
ggplot(data=britain) +
  geom_mosaic(aes(x = product(Sex), fill = Religion),
              colour = "black", size = .3) +
  scale_fill_brewer(palette = "Greys") +
  theme(axis.text.x = element_text(family="serif",
                                   size=10),
        axis.text.y = element_text(angle=90,
                                   hjust=0.6,
                                   family="serif",
                                   size=10),
        panel.background = element_blank(),
        axis.ticks = element_blank(),
        axis.title.x.bottom = element_blank(),
        axis.title.y.left = element_blank(),
        legend.position = "none"
  )
```

### $\chi^2$ Test

```
table <- tally(Sex ~ Religion, data=britain, weights="Weight")
tally(Sex ~ Religion, data=britain, format="percent",
      margin="col", weights="Weight")
chi_squared_test <- chisq.test(table)
chi_squared_test
```

### Graphical Representation of Three Variables

```
ggplot(data=britain) +
  geom_boxplot(color="black", varwidth=TRUE,
               aes(x=Religion, y=Age, fill=Sex),
               ymin=16, outlier.alpha = 0.1) +
    scale_fill_brewer(palette = "Greys") +
    theme(axis.text.x = element_text(family="serif",
                                     size=8),
```

```
        axis.text.y = element_text(family="serif", size=8),
        axis.title.x = element_blank(),
        panel.border = element_rect(colour = "black",
                                    fill = NA))
```