

Assignment 2: Interpreting Quantitative Findings

University of Glasgow

Student ID: 2819052

Course: Quantitative Methods in the Social Sciences

Number of words: 2747

Contents

1	Introduction	3
2	Data and Method	4
2.1	Sample and Data Collection	4
2.2	Dependent and Independent Variables	6
2.3	Control Variables	7
2.4	Method of Analysis	8
3	Results and discussion	9
4	Conclusion	13
5	References	14
6	Appendix	15

1 Introduction

In 1998, the Good Friday/Belfast Agreement was signed and later validated by the voters, and this agreement marked a new era of hope towards peace, equality, and inclusion in Northern Ireland (Galligan 2013). The main focus on the agreement and referendum was obviously to stop the violent conflict and seek peace, but a part of the agreement also emphasized an equality agenda (Hayward 2021). Thus, the agreement also marked the first Northern Irish formal recognition of women's rights to political inclusion (Galligan 2013). However, a formal recognition does not necessarily imply that gender equality trickles down into societal norms and practices. Therefore, this report examines the contemporary state of women's equality in Northern Ireland.

Perhaps, the most central concept within gender equality is the *gender pay gap*. Disparities in income is an important indicator for gender equality, because it has social, economic, and physiological consequences (Bishu and Alkadry 2017). Research on this area have identified several factors that seems to influence a gender pay gap. One such factor can be inequality in access to workplace authority, where women are denied manager or supervisor position although there were equally qualified (Bishu and Alkadry 2017). Other factors can be discrimination in hiring or promotion processes, but also lack of gender representation can avoid minorities to even apply for a job or promotion (Bishu and Alkadry 2017).

In order to examine the current gender equality in Northern Ireland, we therefore focus on the gender pay gap in Northern Ireland. We ask the following research question: *Does gender affect income in contemporary Northern Ireland?* We employ a deductive approach, where we first formulate a hypothesis and subsequently examine it empirically (Bryman 2016, 33). Therefore, our analysis consider the following null hypothesis (H_0) and alternative hypothesis (H_A):

- H_0 : Being a woman or man does not affect your income.
- H_A : Being a woman rather than a man affects a lower income.

The causal relationship that we examine is thus a negative relationship. Although, we call this relationship causal, it is important to clarify that I do not imply that there is anything biological or deterministic that reduce women's income in general. Rather, the causal link is interpreted as a result of the societal discriminatory practices explained in the previous paragraph. To make a convincing inquiry of this causal relationship, we must also control for a number of other factors related to income and gender. This is important to make sure that our analysis do not simply show a spurious correlation as it could in fact be another factor that influences both gender and income. These control variables are introduced in Data and Method.

2 Data and Method

The research design of this analysis is a cross-sectional design (Bryman 2016, 53). This means that the data for analysis is collected at a single point in time and consists of a sample of respondents from which we seek to infer to a general population of Northern Ireland (Field, Miles, and Field 2012, 36). This section describes this sample and how the data was collected, and then subsequently the operationalized variables to examine our research hypothesis are described.

2.1 Sample and Data Collection

Our sample of analysis consists of data from the Northern Ireland Life and Times Survey (NILT). The data have been collected every year since 1998, but this analysis uses the survey from 2012 (ARK 2012). The respondents for the NILT survey was chosen from a systematic random sample of addresses. From this sample of 2126 addresses, 1204 questionnaires were fully completed using partly face-to-face- and self-completion questionnaires (ARK 2012). Our sample of analysis is further reduced as not all respondents have answered our variables of interest, and we are therefore not able to include those respondents in our analysis. In the tables below, we present descriptive statistics for our variables of interest in both the full NILT sample and our sample of analysis. The first tables shows the categorical variables, where a frequency distribution are shown (Fogarty 2018, 88). The second table shows the numerical variables, where we have calculated the mean and standard deviation (Fogarty 2018, 95). For a full visualization of all variables included in the analysis, see Appendix.

Table 1: Descriptive Statistics for Cleaned and Full Sample (Categorical Variables)

Variable	Cleaned Sample		Full Sample		Test
	N	Percent	N	Percent	
Gender	675		1204		X2=1.022
... Male	284	42.1%	537	44.6%	
... Female	391	57.9%	667	55.4%	
Religion	675		1168		X2=0.849
... Catholic	277	41%	491	42%	
... Protestant	283	41.9%	497	42.6%	
... No religion	115	17%	180	15.4%	
Sexual Orientation	675		1191		X2=3.039
... I am heterosexual or straight	657	97.3%	1173	98.5%	
... I am gay or lesbian (homosexual)	14	2.1%	14	1.2%	
... I am bi-sexual	2	0.3%	2	0.2%	
... Other	2	0.3%	2	0.2%	
Constitutional View	675		1183		X2=0.347
... Unionist	199	29.5%	348	29.4%	
... Nationalist	138	20.4%	255	21.6%	
... Neither	338	50.1%	580	49%	
Trade union membership	675		1179		X2=9.162***
... No	374	55.4%	739	62.7%	
... Yes	301	44.6%	440	37.3%	
Supervisor: No	675		883		X2=0.143
... No	464	68.7%	616	69.8%	
... Yes	211	31.3%	267	30.2%	

Note:

Statistical significance markers: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 2: Descriptive Statistics for Cleaned and Full Sample (Numerical Variables)

Variable	Cleaned Sample			Full Sample			Test
	N	Mean	SD	N	Mean	SD	
Annual Personal Income (GBP)	675	16892.089	13447.704	897	16394.582	13465.9	F=0.526
Age	675	46.763	17.117	1201	49.615	18.53	F=10.81***

Note:

Statistical significance markers: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

As the table shows, our sample is significantly reduced by removing missing values - from 1204 observations to 675 observations. In the table with categorical variables, we conduct a Chi-squared test to check if there is independence between the samples or not, i.e. whether the frequency distribution is significantly different (Fogarty 2018, 176). For the numerical variables, we conduct an Analysis of Variances (ANOVA) to test whether the means in the two samples are significantly different (Field, Miles, and Field 2012, 399ff). For the categorical variables, we see that for the test of trade union membership, the test value is 9.162, which is

significantly different from 0 (with a p-value less than 0.01). This indicates that the frequency distribution in our sample is significantly different from the full NILT sample. There is a larger share (44.6 %) of members of a trade union after removing missing values compared to only 37.3 % in the full NILT sample. Similarly, we see from the ANOVA that the test value (F) for age is significantly different from 0 (with a p-value less than 0.01) meaning that our sample is different from the full NILT sample in the age distribution. We can see that the mean age is only around 46 years compared to 49 years in the full NILT sample. These two findings are important to bear in mind, when we interpret our final conclusion as both age and trade union membership might be related to our dependent variable, income.

The systematic random sampling strategy in the NILT survey is an approach to make the sample representative of our population - in our case all inhabitants of Northern Ireland. It could be relevant to examine the demographic distribution in our sample against the general population of Northern Ireland, but unfortunately that is beyond the scope of this report.

2.2 Dependent and Independent Variables

In order to examine the research question empirically, we operationalize the concepts included in our hypothesis as measurable variables. As the questionnaire was formulated beforehand in the NILT 2012 survey, the data is a secondary source, as we have to find the questions that can represent our theoretical concepts.

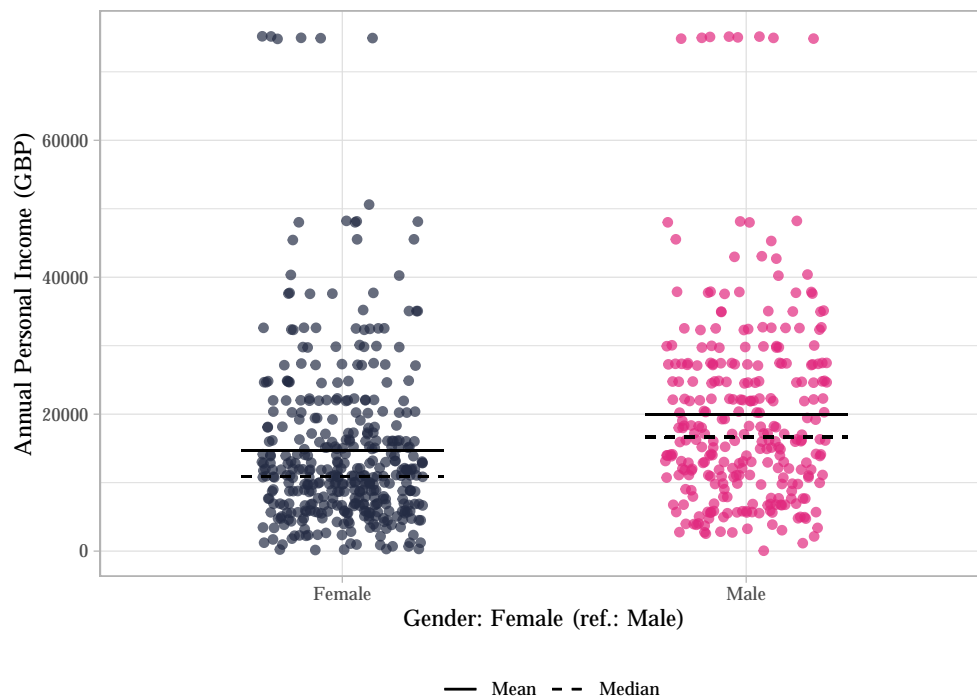
The dependent variable or outcome that we want to explain in this analysis is income. Although income in itself is more concrete concept compared to other theoretical concepts in social science (such as e.g. “solidarity”), it can still be measured in many different ways. In this analysis, we measure income by the question: *What is your personal income before tax and national insurance contributions? (£ per year)*. It is thus a numeric, more specifically a ratio scale with a value of absolute zero - although it is quite uncommon to have a personal income of zero (Fogarty 2018, 56).

The independent variable or explanatory variable of our analysis is gender. Gender is a much more complex concept, and there is a large discussion about the distinction between gender and sex, and the biological and cultural components of these (Unger and Crawford 1993). Quantitative research have mostly been dominated by gender being measured as a binary division between male and female. Likewise, we are also limited to such a measurement of gender in the NILT 2012 survey. In this analysis, we therefore measure gender by a nominal, binary variable of either “male” or “female”.

To consider the distribution of our dependent and independent variable, see the following scatter plot of income and gender. In the scatter plot, we also show the mean and median income for each group of gender

by the solid and dashed lines.

Figure 1: Scatterplot of Income and Gender



Note:
The points are jittered on the xaxis for the purpose of visualization.
There are only two categories for this question.

There are several insights from this visualization. The vertical differences between the descriptive statistics, mean and median signifies that there is a general income difference between males and females, and that females in average have lower incomes than men. Initially, this suggests an acceptance of our alternative research hypothesis, but we will wait with any conclusion till the analysis, where we also control for other related factors. Another interesting insight is that the distribution of income seems to be slightly skewed. For both males and females, there are some respondents with a very high income compared to most other respondents. This can also noticed as there is a difference between the mean and the median for both groups. These respondents have such extreme high incomes that they might be outliers compared to the general trend, and we need to be aware of that in our analysis (Fogarty 2018, 221–22).

2.3 Control Variables

Theoretically, there are other characteristics than gender than might affect your income. Firstly, there are characteristics related to your work that might affect a higher salary that is not rooted in discriminatory

practices. To make sure that an effect of gender is not just a result these factor, we find it relevant to include trade union membership and whether or not someone has a formal responsibility for supervising the work of other employees (ARK 2012). Moreover, there can also be other characteristics that affect a lower income based in discriminatory practices that is not rooted upon one's gender. Therefore, we also include religion, sexual orientation, constitutional view, and age into our analysis as control variables.

2.4 Method of Analysis

To fully examine the relationship between our dependent and independent variable, we employ a multiple linear regression estimated with OLS (Fogarty 2018, 192ff). Thus, we are also able to include the control variables to make the acceptance or decline of our hypothesis more convincing. The formula of the regression of our analysis is thus:

$$\begin{aligned} \hat{income} = & \beta_0 + \beta_1 * gender + \beta_2 * religion + \beta_3 * SexualOrientation \\ & + \beta_4 * ConstitutionalView + \beta_5 * TradeUnionMembership + \beta_6 * Supervisor + \varepsilon \end{aligned}$$

where the main parameter of interest is β_1 for gender.

3 Results and discussion

This report examines the question of whether gender affect income in contemporary Northern Ireland. As described in the Introduction, the analysis therefore seeks to accept whether the null hypothesis or the alternative hypothesis - that being a woman rather than a man affects a lower income. As described in the Data and Method section, our analysis consists of survey data, where our main variables are gender and personal income. To examine our hypothesis, the following table shows the results of two regressions. First, our model is shown as a bivariate regression excluding our control variables, and secondly our full model is shown including our control variables.

Table 3: Regression results

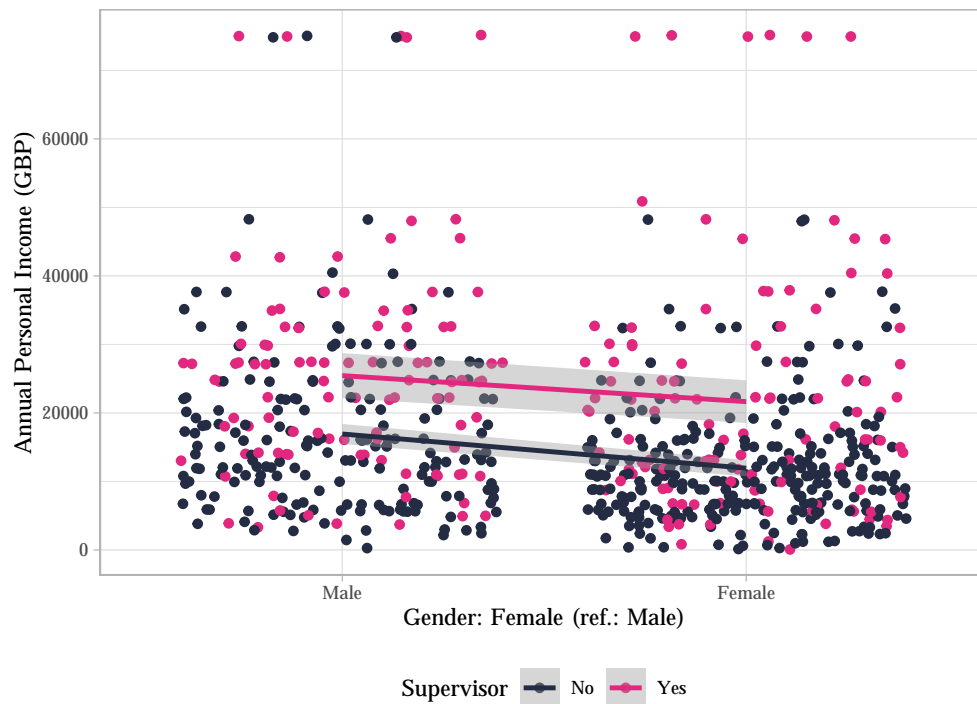
	Dependent Variable	
	Annual Personal Income (GBP)	
	Model excl. controls	Model incl. controls
Gender: Female (ref.: Male)	-5,234.993*** (1,029.652)	-5,068.737*** (994.748)
Religion: Protestant (ref.: Catholic)		465.188 (1,458.367)
Religion: No religion		895.169 (1,533.323)
Sexual Orientation: Homosexual (ref.: Heterosexual)		-6,247.777* (3,437.048)
Sexual Orientation: bi-sexual		-2,826.980 (8,698.806)
Sexual Orientation: Other		1,323.336 (8,737.282)
Constitutional View: Nationalist (ref.: Unionist)		1,788.873 (1,898.294)
Constitutional view: Neither		1,438.036 (1,350.423)
Trade union membership: Yes (ref.: No)		5,277.978*** (977.008)
Supervisor: Yes (ref.: No)		8,648.320*** (1,037.559)
Age		-84.369*** (29.430)
Constant	19,924.510*** (783.659)	17,417.240*** (2,223.337)
Observations	675	675
R ²	0.037	0.183
Adjusted R ²	0.036	0.170
Residual Std. Error	13,206.460 (df = 673)	12,252.840 (df = 663)
F Statistic	25.849*** (df = 1; 673)	13.533*** (df = 11; 663)
Note:	*p<0.1; **p<0.05; ***p<0.01	

The primary finding from the regression results is that the variable gender have an significant effect on income. We see that the p-value is less than 0.01. This indicates that there is indeed a significant difference between the income of men and women. However, a significant p-value does not by itself means that the difference is important or meaningful - we also need to interpret the effect size (Field, Miles, and Field 2012, 57). If we take a look at the effect size (β_1), it has a value of approx -5,068. Thus according to the model, women in average earns 5,068 pounds a year less than men. That seems quite substantial as the average income in our sample is 16,892 pounds a year (See Table 2). Based on our analysis, we therefore accept our alternative hypothesis - that being a woman rather than a man has a significant negative effect in your income.

In the model, where we include our control variables, we see that trade union membership and being a supervisor has a significant positive effect on income, whereas age and to some extent also having a homosexual orientation rather than a heterosexual orientation have a significant negative effect on income. However, the effect of gender is independent of religion, sexual orientation, constitutional view, trade union membership, being a supervisor, and age, as including these control variables into the regression does not render the effect of gender on income insignificant. If we compare the model with and without control variables, we see that both R^2 and adjusted R^2 is increased with indicates that including the control variables does indeed yield a better fit of data.

In this way, our empirical findings support the argument that there is still a gender pay gap in Northern Ireland. Our finding that the gap between men and women is independent of being a supervisor or not is particularly interesting. It suggests that the income gap between men and women exists for both, and the gender pay gap cannot only be explained by men more often being in leadership positions. In the following visualization, this finding can be seen clearly.

Figure 2: Scatterplot of Income, Gender, and Supervisor



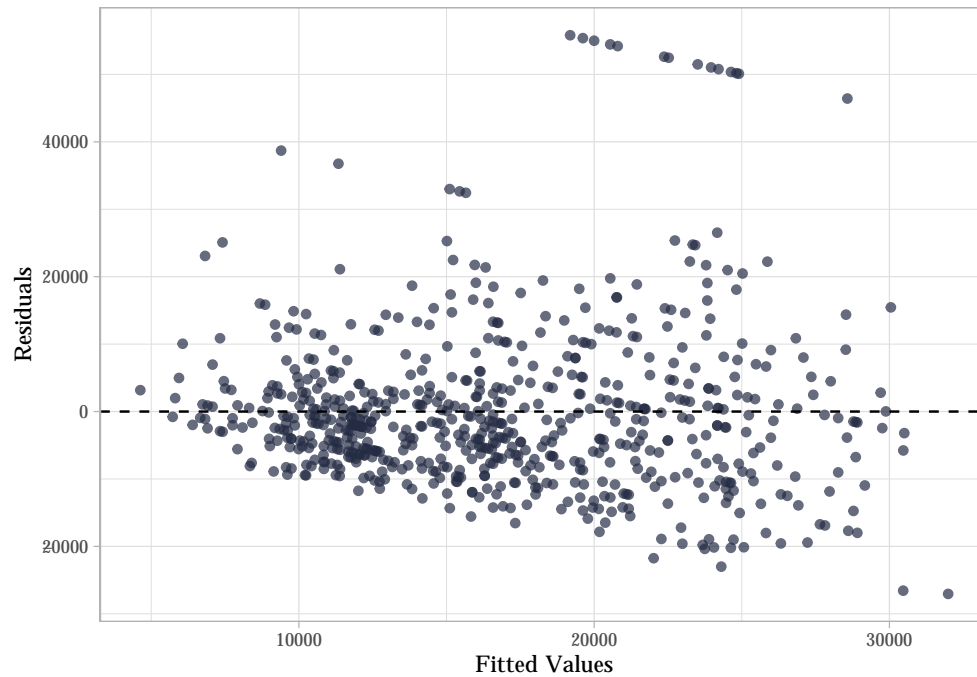
Note:
The points are jittered on the axis for the purpose of visualization.
There are only two categories for this question.

In the visualization, there is also shown two linear regression of gender on income, and we see that both slopes indicate a negative correlation - although the intercepts differ.

Before concluding on our empirical results, it is also important to consider how reliable our regression model is. One relevant concern of our analysis is the influence of outliers, i.e. very influential data points (Fogarty 2018, 221–22; Field, Miles, and Field 2012, 190). As identified in the [Data and Methods] section, the dependent variables seems to have a data points with incomes that are very much higher. Most respondents have an income around 17,000 pounds a year, while a few respondents have a yearly income around more than 60,000 pounds. It is beyond the scope of this report to examine whether these data points are in fact significant outliers, but it is important to notice when interpreting the results.

Another concern to our model is heteroscedasticity, i.e. whether there is not a constant variance of residuals on the levels of the dependent variable, income (Field, Miles, and Field 2012, 272). In the plot below, we visualize the fitted values with the residuals to see if the variances is constant along the x-axis. Below the plot is also shown the results of a Breusch-Pagan test to more formally test for heteroscedasticity (Fogarty 2018, 228).

Figure 3: Scatterplot of Fitted Values and Residuals



Method	Statistic	Pvalue	Parameter
BreuschPagan test	133.494	0	11

Unfortunately, the plot shows that the variance seem to increase for higher values of our dependent variables, income, and the Breusch-Pagan test also suggests that there is heteroscedasticity as the p-value is lower than 0.001. One solution to this problem could be to re-estimate our regression model with robust standard errors, but this has not been possible within the scope of this report (Fogarty 2018, 229).

4 Conclusion

Based on Northern Ireland as a historical case, this report has examined the gender pay gap in the contemporary society of Northern Ireland. Previous research has suggested that inclusion of statements about gender equality does not necessarily trickle down into the everyday practices, and this report therefore sought to examine this causal relation empirically. The analysis tackles the research question: *Does gender affect income in contemporary Northern Ireland?* with the alternative hypothesis that being a woman rather than a man affects a lower income.

Using a survey-based research design, we employed a multivariate linear regression analysis to answer this question. Our analysis showed that there is indeed a gender pay gap in contemporary Northern Ireland. We found that being a woman compared to a man resulted in average of an approx. 5000 pounds lower yearly income - independent of factors such as trade union membership, being a supervisor, age and more. Here, it is particularly interesting that the effect of gender on income is independent of having supervision, because it suggests that the gender inequality exists across having a leadership position or not. Thus, our analysis supports the arguments that gender affects income rooted in discriminatory practices as theory suggests.

However, there are also limitations to our study that it is important to take into concern. Firstly, there might be a few influential data points - so-called outliers - that drives the effect to be larger than the general trend. Secondly, we identified heteroscedasticity in our model, and especially large incomes seems to have larger residuals. These are primarily limitations of the reliability of the study, since it concerns the model and methodological choices (Bryman 2016, 41). If these results can be reproduced with a model with robust standard errors or after a check for outliers, it would increase the reliability of our findings. Another point of discussion is the validity of our findings, and here there is mostly the question of external validity as our sample is reduced significantly during the data-cleaning process (Bryman 2016, 41–42). A further improvement here could have been to directly analyze if the respondents in our data are representative compared to the general population of Northern Ireland.

5 References

- ARK. 2012. *2012 Northern Ireland Life & Times Survey: Technical Notes*. ARK. <https://www.ark.ac.uk/nilt/2012/tech12.pdf>.
- Bishu, Sebawit G., and Mohamad G. Alkadry. 2017. “A Systematic Review of the Gender Pay Gap and Factors That Predict It.” *Administration & Society* 49 (1): 65–104. <https://doi.org/10.1177/0095399716636928>.
- Bryman, A. 2016. *Social Research Methods*. Oxford University Press. <https://books.google.co.uk/books?id=N2zQCgAAQBAJ>.
- Field, A., J. Miles, and Z. Field. 2012. *Discovering Statistics Using r*. SAGE Publications. <https://books.google.co.uk/books?id=wd2K2zC3swIC>.
- Fogarty, B. J. 2018. *Quantitative Social Science Data with r: An Introduction*. Core Textbook. SAGE Publications. <https://books.google.co.uk/books?id=jwJ6DwAAQBAJ>.
- Galligan, Yvonne. 2013. “Gender and Politics in Northern Ireland: The Representation Gap Revisited.” *Irish Political Studies* 28 (3): 413–33.
- Hayward, Katy. 2021. “The 1998 Good Friday (Belfast) Agreement Referendums in Northern Ireland and the Republic of Ireland.” In *The Palgrave Handbook of European Referendums*, edited by Julie Smith, 247–65. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-55803-1_12.
- Unger, Rhoda K, and Mary Crawford. 1993. “Sex and Gender—the Troubled Relationship Between Terms and Concepts.” *Psychological Science* 4 (2): 122–24.

6 Appendix

Figure 4: Bar plot of Gender

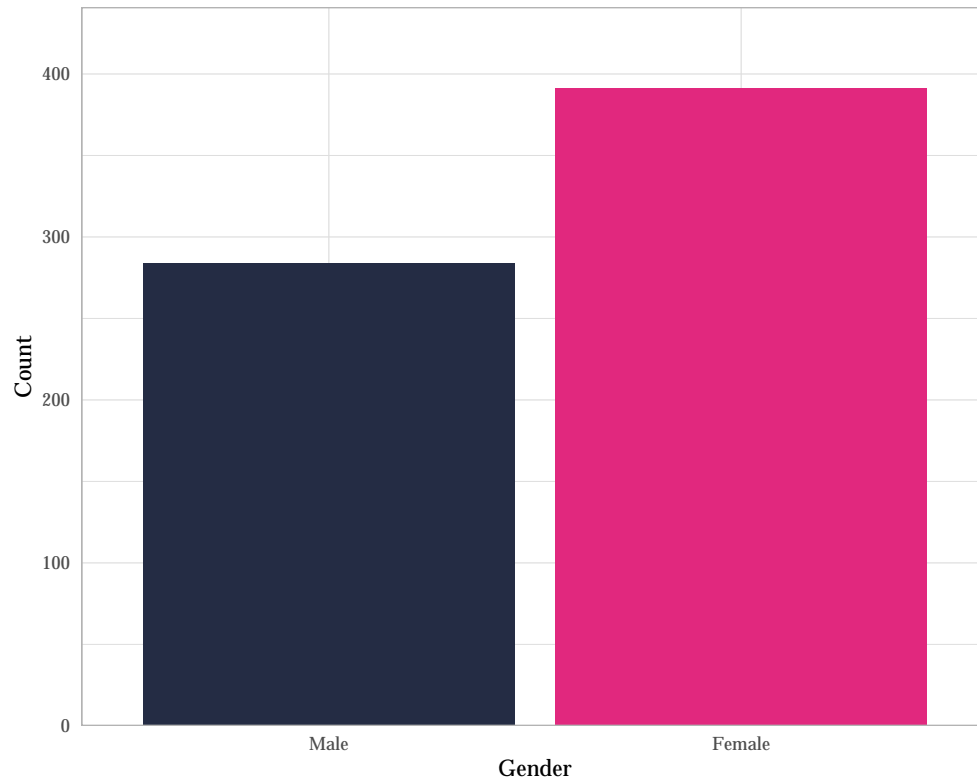


Figure 5: Bar plot of Religion

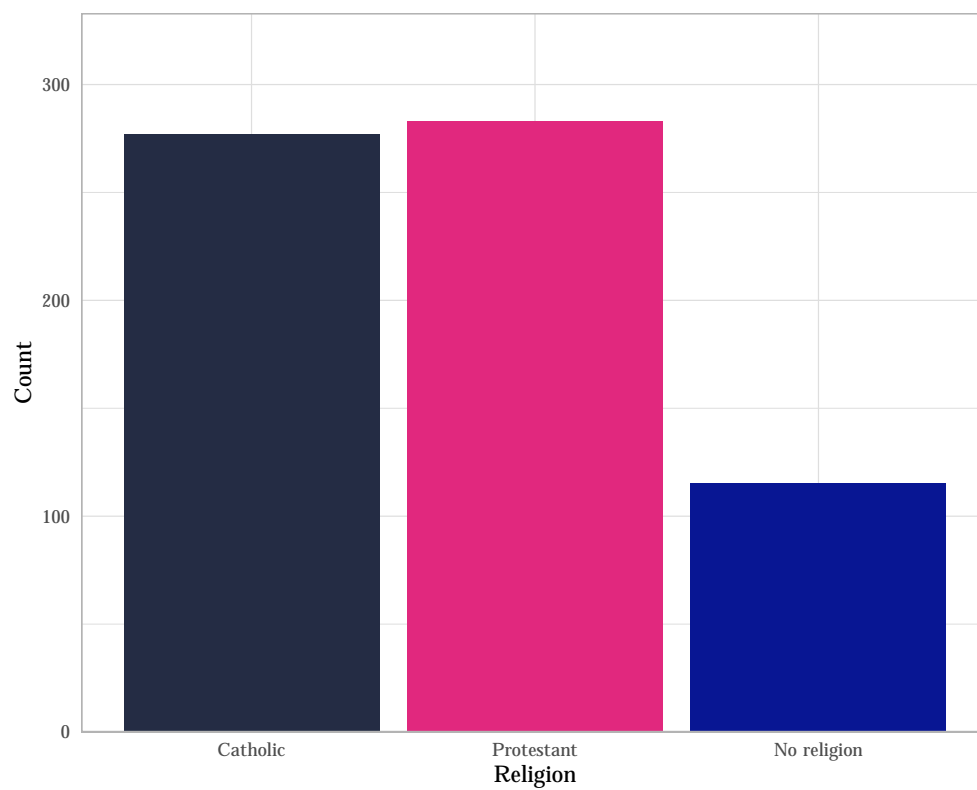


Figure 6: Bar plot of Sexual Orientation

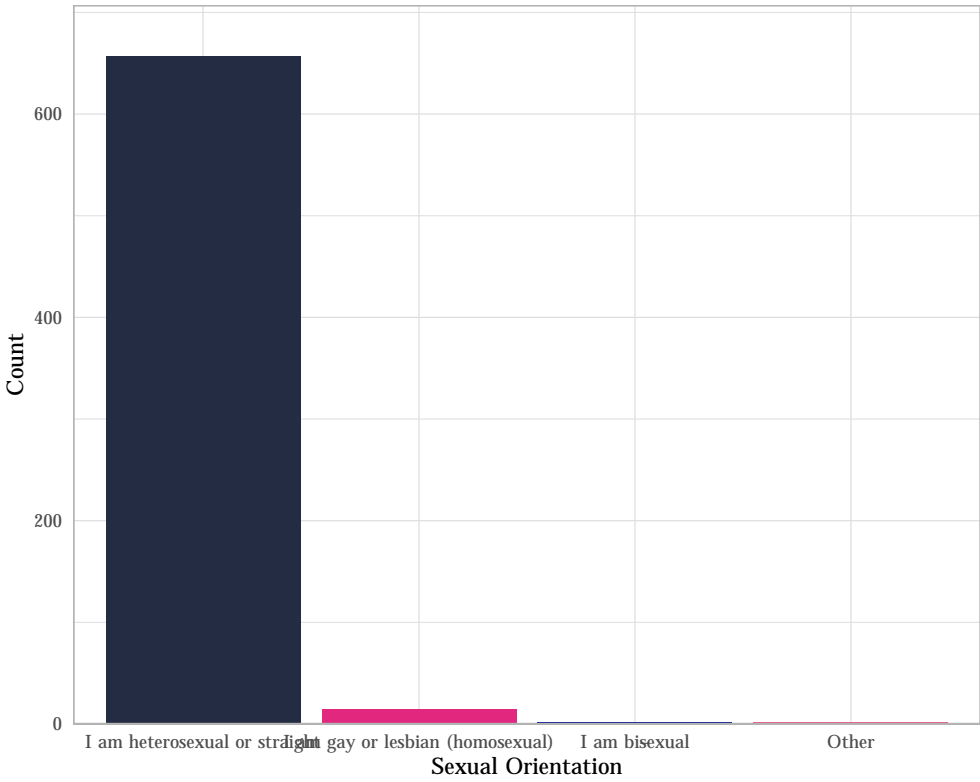


Figure 7: Bar plot of Constitutional View

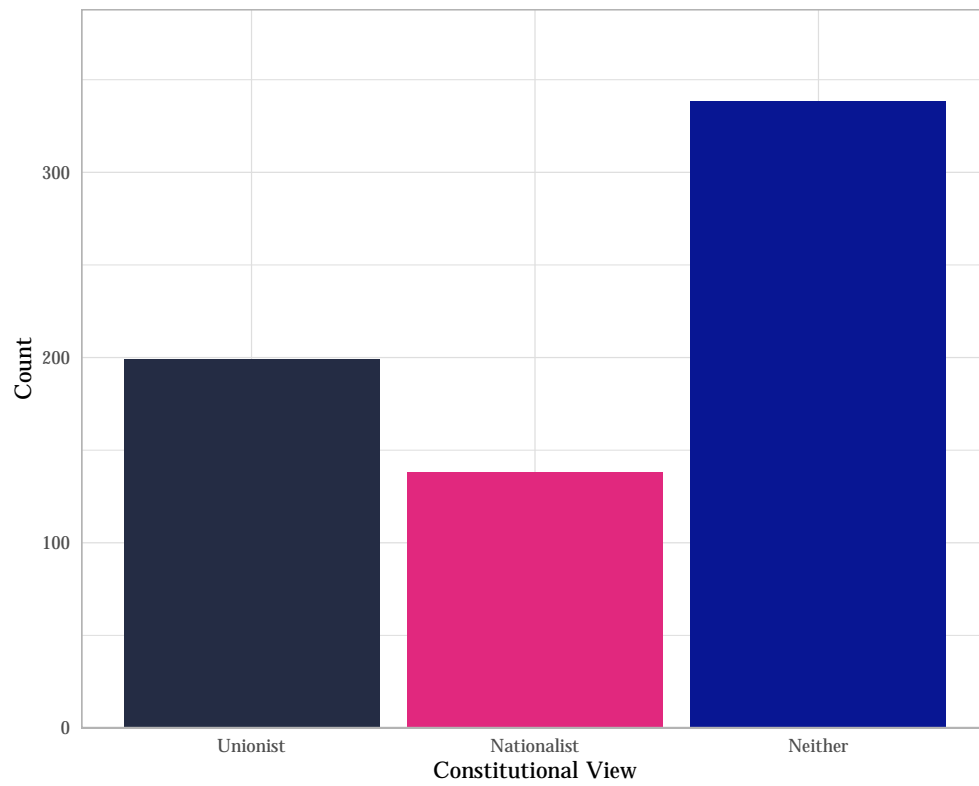


Figure 8: Bar plot of Trade union membership

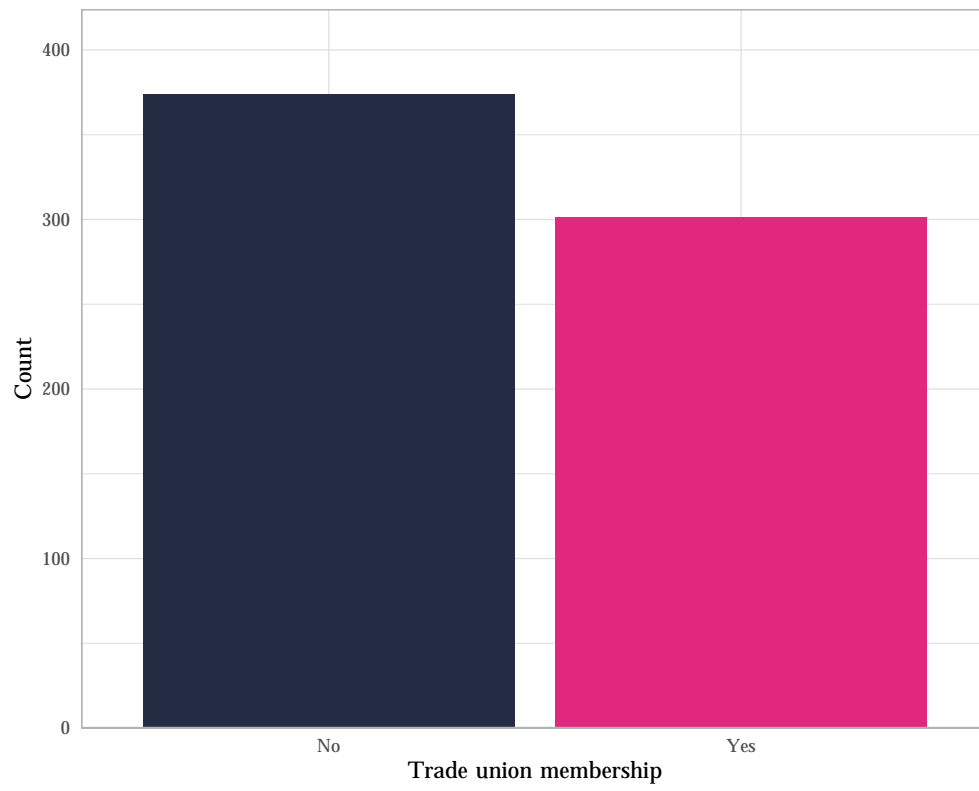


Figure 9: Bar plot of Supervisor

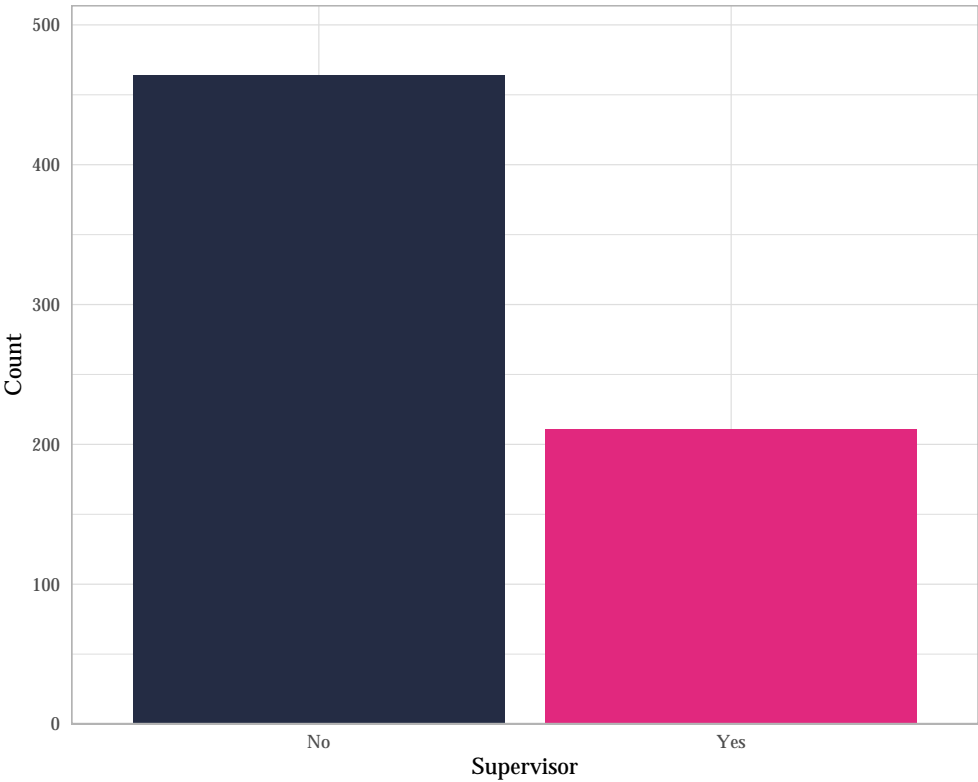


Figure 10: Density plot of Age

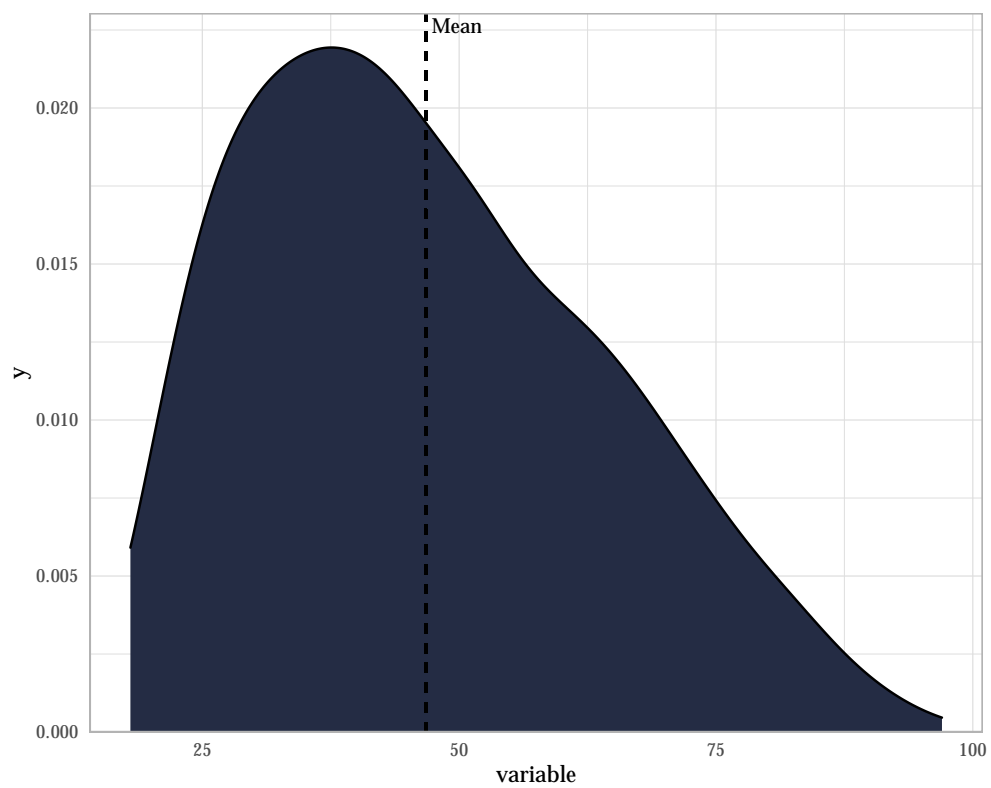


Figure 11: Density plot of Income

