Does Annual Household Income Play a Role in Minority's Political Party Identification?

Introduction

Political system of the United Kingdom

The United Kingdom (UK) is governed through a parliamentary democracy under a constitutional monarchy, where the king (or queen) serves as the ceremonial head of state while political authority is exercised by elected representatives. The UK Parliament consists of two houses: the House of Commons, made up of Members of Parliament (MPs) elected by the public, and the House of Lords, which is primarily appointed. The Prime Minister, as head of government, is drawn from the House of Commons and leads the Cabinet. Executive power is dependent on maintaining the confidence of Parliament, particularly the elected Commons, making it central to the UK's political system.

General elections to the Parliament are held every five years using a first-past-the-post voting system across 650 constituencies, where the candidate with the most votes wins. This system tends to benefit larger political parties, sometimes leading to disproportionate results between national vote share and parliamentary seats.[1] Major political parties include the Labour Party, Conservative Party, and Liberal Democrats, along with regional parties such as the Scottish National Party (SNP) and *Plaid Cymru*, the left-wing nationalist political party in Wales.

In a political system governed by FPTP, the electoral behavior of minority groups can carry considerable weight, particularly in closely contested constituencies. Since FPTP rewards the candidate with the most votes — even without a majority — a relatively small number of voters can tip the balance between the two dominant parties. As a result, understanding how ethnic minority communities align politically is not only important for representation, but also for predicting outcomes in marginal seats that may ultimately determine the composition of government.

Stakeholders in the Survey

The political behavior of people from British backgrounds has been widely researched and is solidly understood. However, the relatively new phenomenon of political participation by immigrants from various backgrounds poses a significant challenge for those seeking to understand or predict the political thinking of the British population, or even to anticipate their reactions to different kinds of events.

The national and local governments, political parties, NGOs, and polling agencies could all be interested in the findings of this survey.

Analyses

In the following, I will perform analyses using five columns of the data:

- Annual Household Income `zqinc`
- Party Identification `bq9_1`
- Strength of Party Identification `bq9_4`
- Whether Political Party Represents R's Views `eq11`

[1] Norton, Philip. *The British Polity.* 5th ed., Longman, 2010.

- Party Represents R's Views Best `eq12`

I will analyze the effect of the income situation of respondents on their political party identification.

I will use `zqinc` as my independent variable.

As my dependent variables of my analyses (described below) i will use the remaining four variables: `bq9_1`, `bq9_4`, `eq11` and `eq12`.

'Annual Household Income' - 'Party Identification'

```
# Load libraries
library(haven)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

Load the data.

```
data <- read_dta("BSE_2010.dta")
```

The dataset is filtered to remove invalid or non-informative responses such as "Refused", "Don't Know", and "Not Stated". This ensures that only valid, analyzable responses are included. The process then groups income (`zqinc`) into broader, interpretable income brackets (e.g., "0-5000", "45001+"), and simplifies party identification (`bq9_1`) into general categories like "Labour", "Conservative", "Other", and "None".

```
data_clean1 <- data %>%
  filter(
    !zqinc %in% c(-2, -1, 17),
    !bq9_1 %in% c(-2, -1, 15)
  ) %>%
  mutate(
    income_group_simple = case_when(
      zqinc == 1 ~ "0-5000",
      zqinc == 2 ~ "5001-10000",
      zqinc == 3 ~ "10001-15000",
      zqinc == 4 ~ "15001-20000",
      zqinc %in% 5:6 ~ "20001-30000",
      zqinc %in% 7:9 ~ "30001-45000",
      zqinc %in% 10:14 ~ "45001+"
    ),
    party_simple = case_when(
      bq9_1 == 1 ~ "None",
      bq9_1 == 2 ~ "Labour",
      bq9_1 == 3 ~ "Conservative",
      bq9_1 %in% 4:12 ~ "Other"
    )
```

```
  )
```

The χ^2 test of independence is appropriate here because both variables — annual household income (grouped categorically) and party identification — are categorical in nature. This test allows us to assess whether the distribution of party preferences differs significantly across income groups. It does not assume a normal distribution and is ideal for frequency data arranged in contingency tables.

```
full_table <- table(data_clean1$income_group_simple, data_clean1$party_simple)
chi_full <- chisq.test(full_table)

print(chi_full)
```

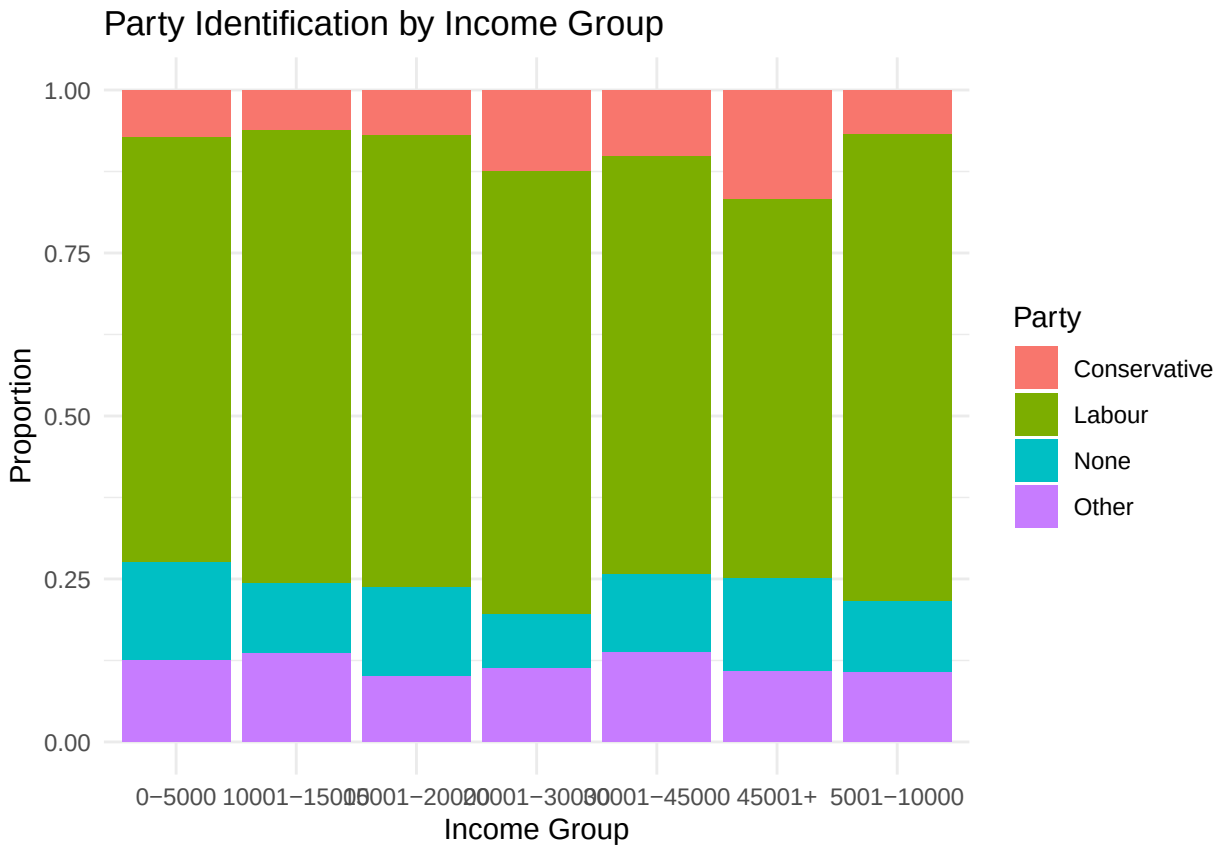
```
##
##  Pearson's Chi-squared test
##
## data:  full_table
## X-squared = 35.967, df = 18, p-value = 0.007125
```

```
if (chi_full$p.value < 0.05) {
  cat("Result: Reject the null hypothesis.\n")
  cat("Interpretation: Income significantly affects minority party identification.\n")
} else {
  cat("Result: Fail to reject the null hypothesis.\n")
  cat("Interpretation: Income does NOT significantly affect minority party identification.\n")
}
```

```
## Result: Reject the null hypothesis.
## Interpretation: Income significantly affects minority party identification.
```

Visualization

```
ggplot(data_clean1, aes(x = income_group_simple, fill = party_simple)) +
  geom_bar(position = "fill") +
  labs(
    title = "Party Identification by Income Group",
    x = "Income Group",
    y = "Proportion",
    fill = "Party"
  ) +
  theme_minimal()
```




```
  coord_flip()
```

```
## <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>
##     aspect: function
##     backtransform_range: function
##     clip: on
##     default: FALSE
##     distance: function
##     expand: TRUE
##     is_free: function
##     is_linear: function
##     labels: function
##     limits: list
##     modify_scales: function
##     range: function
##     render_axis_h: function
##     render_axis_v: function
##     render_bg: function
##     render_fg: function
##     setup_data: function
##     setup_layout: function
##     setup_panel_guides: function
##     setup_panel_params: function
##     setup_params: function
##     train_panel_guides: function
##     transform: function
##     super:  <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>
```

'Annual Household Income' - 'Strength of Party Identification'

In this analysis, the data is first prepared by selecting only the relevant variables: income (`zqinc`) and strength of party identification (`bq9_4`). Invalid responses such as "Refused", "Don't Know", and "Not Stated" are excluded to ensure the remaining data is valid for analysis. Then, income is grouped into broader, interpretable brackets (e.g., "20001-30000", "45001+"), which helps reduce sparsity and improves the interpretability of results. The strength of party identification is recoded into an ordered factor with three clear levels ("Very Strong", "Fairly Strong", "Not Very Strong"), respecting its ordinal nature. This data preprocessing step ensures the variables are clean, well-structured, and statistically appropriate for the Kruskal-Wallis test.

```
data_selected2 <- data %>%
  select(zqinc, bq9_4)

data_clean2 <- data_selected2 %>%
  filter(
    !(zqinc %in% c(-2, -1, 17)),
    !(bq9_4 %in% c(-2, -1, 6))
  )

data_clean2 <- data_clean2 %>%
  mutate(income_group = case_when(
    zqinc %in% 1    ~ "0-5000",
    zqinc %in% 2    ~ "5001-10000",
    zqinc %in% 3    ~ "10001-15000",
    zqinc %in% 4    ~ "15001-20000",
    zqinc %in% 5:6    ~ "20001-30000",
    zqinc %in% 7:9    ~ "30001-45000",
    zqinc %in% 10:14   ~ "45001+"
  ))

data_clean2 <- data_clean2 %>%
  mutate(strength_party_id = factor(bq9_4,
                                    levels = c(1,2,3),
                                    labels = c("Very Strong",
                                               "Fairly Strong",
                                               "Not Very Strong"),
                                    ordered = TRUE))

table(data_clean2$income_group, data_clean2$strength_party_id)
```

```
##
##                Very Strong Fairly Strong Not Very Strong
##   0-5000                36            61              39
##   10001-15000           58           104              78
##   15001-20000           35            82              58
##   20001-30000           50           129              60
##   30001-45000           36           109              63
##   45001+                41           108              64
##   5001-10000            63           123              76
```

The Kruskal-Wallis test is suitable for comparing the strength of party identification, which is an ordinal variable, across multiple independent income groups. Since this variable reflects ordered categories (e.g., "Very strong", "Fairly strong", etc.), and normality cannot be assumed, the Kruskal-Wallis test offers a non-parametric alternative to one-way ANOVA that respects the ordinal nature of the data and does not require equal variances.

```r
kruskal_result2 <- kruskal.test(as.numeric(strength_party_id) ~ income_group,
                                data = data_clean2)

print(kruskal_result2)
```

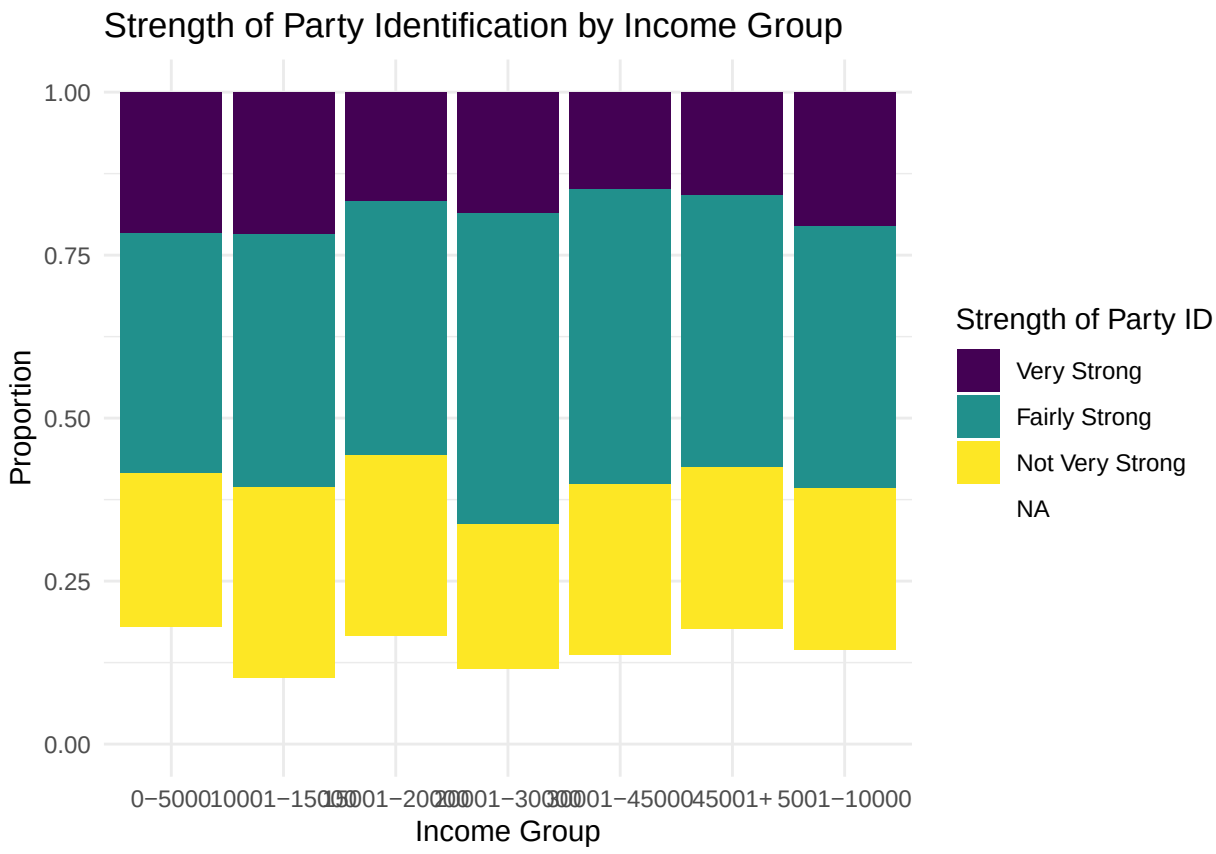
```
##
##  Kruskal-Wallis rank sum test
##
## data:  as.numeric(strength_party_id) by income_group
## Kruskal-Wallis chi-squared = 4.2359, df = 6, p-value = 0.6448
```

```r
if(kruskal_result2$p.value < 0.05){
  cat("Reject H0: Income significantly affects strength of party identification.\n")
} else {
  cat("Fail to reject H0: Income does NOT significantly affect strength of party identification.\n")
}
```

```
## Fail to reject H0: Income does NOT significantly affect strength of party identification.
```

Visualization

```r
ggplot(data_clean2, aes(x=income_group, fill=strength_party_id)) +
 geom_bar(position="fill") +
 labs(x="Income Group", y="Proportion", fill="Strength of Party ID",
      title="Strength of Party Identification by Income Group") +
 theme_minimal()
```




```r
 coord_flip()
```

```
## <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>
##     aspect: function
##     backtransform_range: function
##     clip: on
##     default: FALSE
##     distance: function
##     expand: TRUE
##     is_free: function
##     is_linear: function
##     labels: function
##     limits: list
##     modify_scales: function
##     range: function
##     render_axis_h: function
##     render_axis_v: function
##     render_bg: function
##     render_fg: function
##     setup_data: function
##     setup_layout: function
##     setup_panel_guides: function
##     setup_panel_params: function
##     setup_params: function
##     train_panel_guides: function
##     transform: function
##     super:  <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>
```

'Annual Household Income' - 'Whether Political Party Represents R's Views', 'Party Represents R's Views Best'

In this third analysis, the dataset is preprocessed by selecting only the variables related to income (`zqinc`) and political representation (`eq11`, `eq12`). The filtering step removes invalid or non-responses coded as "Refused", "Don't Know", or "Not Stated", ensuring the data is complete and reliable for analysis. Income is then grouped into broader, meaningful categories to make comparisons more interpretable and statistically stable. Finally, the `eq11` variable — indicating whether respondents feel that any political party represents their views — is recoded into a binary factor ("Yes", "No"), preparing it for categorical analysis such as a χ^2 test. This preprocessing ensures the data structure matches the assumptions of the chosen method.

```
data_selected3 <- data %>%
  select(zqinc, eq11, eq12)

data_clean3 <- data_selected3 %>%
  filter(
    !(zqinc %in% c(-2, -1, 17)),
    !(eq11 %in% c(-2, -1, 5)),
    !(eq12 %in% c(-2, -1, 13))
  )

data_clean3 <- data_clean3 %>%
  mutate(income_group = case_when(
    zqinc %in% 1   ~ "0-5000",
    zqinc %in% 2   ~ "5001-10000",
    zqinc %in% 3   ~ "10001-15000",
    zqinc %in% 4   ~ "15001-20000",
    zqinc %in% 5:6 ~ "20001-30000",
    zqinc %in% 7:9 ~ "30001-45000",
```

```
    zqinc %in% 10:14 ~ "45001+"
  ))

data_clean3 <- data_clean3 %>%
  mutate(represented = factor(eq11, levels = c(1,2), labels = c("Yes", "No")))
```

Analysis 1: Does income affect feeling represented (eq11)? This χ^2 test examines whether the likelihood of feeling politically represented by any party is related to income group. Since both the feeling of being represented (Yes/No) and income group are categorical variables, the test is appropriate for evaluating whether these two factors are statistically associated.

```
table_eq11 <- table(data_clean3$income_group, data_clean3$represented)

chi_eq11 <- chisq.test(table_eq11)

print(chi_eq11)
```

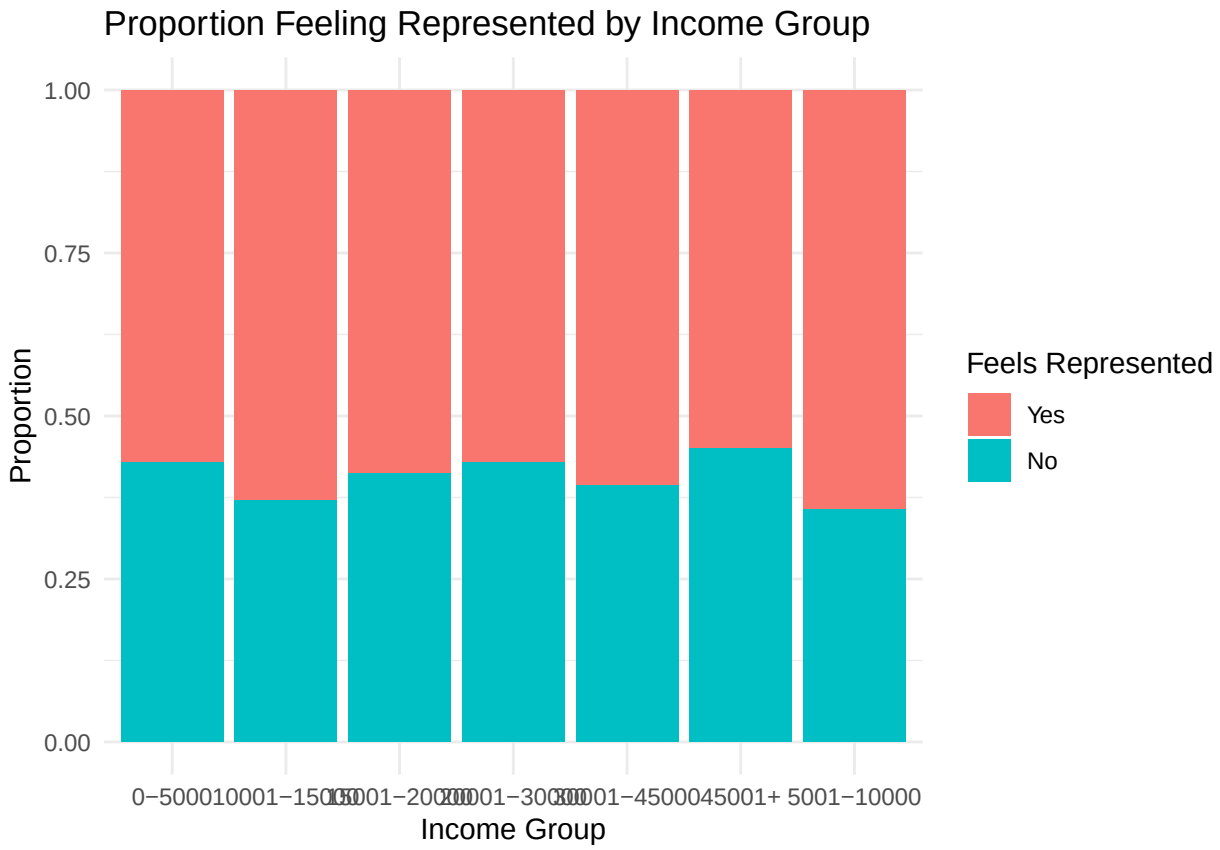
```
##
##  Pearson's Chi-squared test
##
## data:  table_eq11
## X-squared = 6.5674, df = 6, p-value = 0.3627
```

```
if(chi_eq11$p.value < 0.05){
  cat("Reject H0: Income significantly affects whether people feel represented.\n")
} else {
  cat("Fail to reject H0: Income does NOT significantly affect feelings of representation.\n")
}
```

```
## Fail to reject H0: Income does NOT significantly affect feelings of representation.
```

Visualization (Optional)

```
ggplot(data_clean3, aes(x=income_group, fill=represented)) +
  geom_bar(position="fill") +
  labs(title="Proportion Feeling Represented by Income Group",
       x="Income Group", y="Proportion", fill="Feels Represented") +
  theme_minimal()
```




```
coord_flip()
```

```
## <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>
##     aspect: function
##     backtransform_range: function
##     clip: on
##     default: FALSE
##     distance: function
##     expand: TRUE
##     is_free: function
##     is_linear: function
##     labels: function
##     limits: list
##     modify_scales: function
##     range: function
##     render_axis_h: function
##     render_axis_v: function
##     render_bg: function
##     render_fg: function
##     setup_data: function
##     setup_layout: function
##     setup_panel_guides: function
##     setup_panel_params: function
##     setup_params: function
##     train_panel_guides: function
##     transform: function
##     super:  <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>
```

Analysis 2: Does income affect which party best represents views (eq12)? In this final analysis, the data is further refined by focusing only on respondents who indicated that a political party represents their views (`eq11 == 1`) and who selected a party within a defined range (`eq12` values 1 to 9). This ensures the analysis only includes valid, meaningful responses. The selected parties are then grouped into three broader political categories: "Labour", "Conservative block" (including Conservatives, UKIP, and BNP), and "Liberals" (including Liberal Democrats, SNP, Plaid Cymru, Greens, and Respect). This recoding simplifies the analysis while preserving political distinctions. The resulting variable is converted into a factor with a defined order, preparing it for a χ^2 test comparing perceived representation across income groups.

```
data_eq12 <- data_clean3 %>%
  filter(eq11 == 1, eq12 %in% 1:9)

data_eq12 <- data_eq12 %>%
  mutate(representing_party = case_when(
    eq12 == 1 ~ "Labour",
    eq12 %in% c(2, 7, 8) ~ "Conservative block",
    eq12 %in% c(3, 4, 5, 6, 9) ~ "Liberals"
  ))

data_eq12$representing_party <- factor(data_eq12$representing_party,
                                       levels = c("Labour", "Conservative block", "Liberals"))
```

The second χ^2 test explores whether the specific party a respondent feels best represents their views varies across income groups. Both variables are categorical — the respondent's selected party (grouped into blocks like Labour, Conservative block, etc.) and income group. This makes the χ^2 test a suitable choice to detect whether income influences political alignment perceptions.

```
table_eq12 <- table(data_eq12$income_group, data_eq12$representing_party)

chi_eq12 <- chisq.test(table_eq12, simulate.p.value = TRUE, B = 5000)

print(chi_eq12)
```

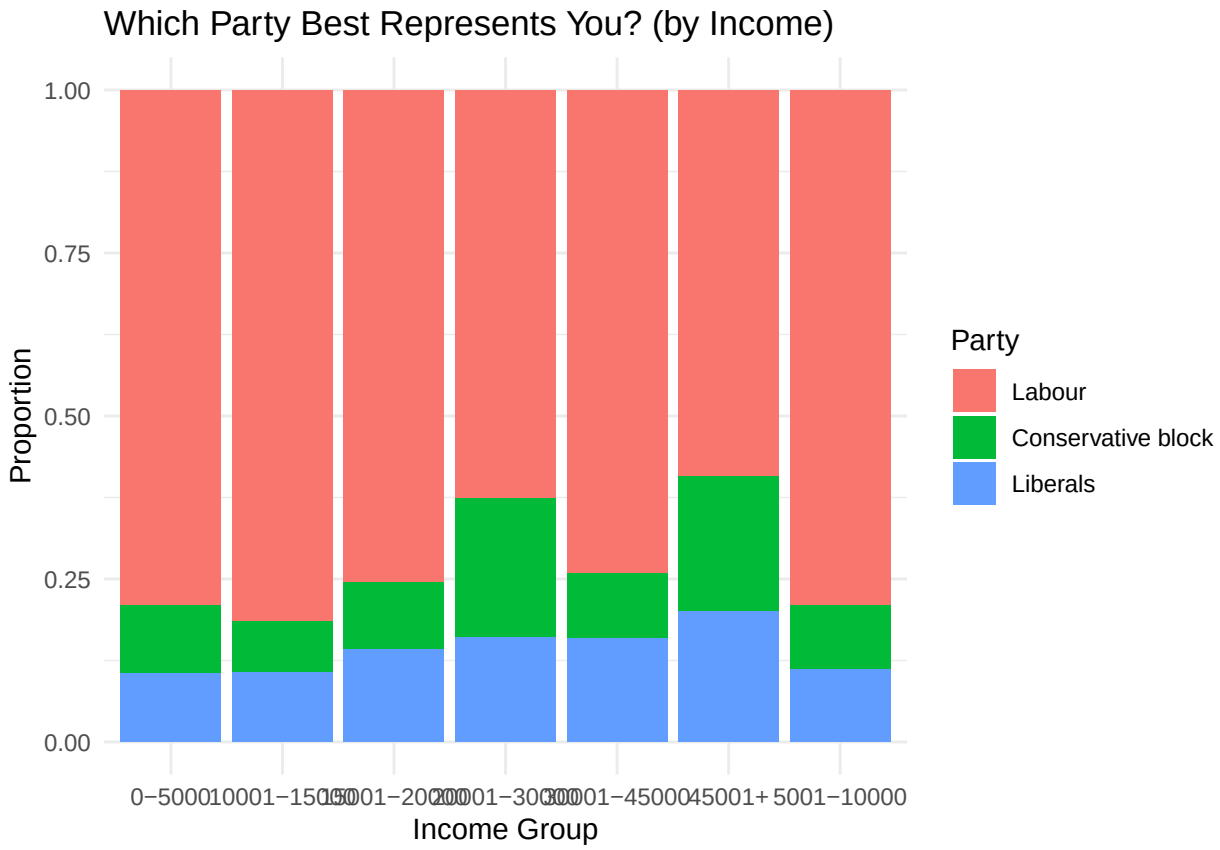
```
##
##  Pearson's Chi-squared test with simulated p-value (based on 5000
##  replicates)
##
## data:  table_eq12
## X-squared = 33.242, df = NA, p-value = 0.0009998
```

```
if(chi_eq12$p.value < 0.05){
  cat("Reject H0: Income significantly affects which party minorities feel best represents them.\n")
} else {
  cat("Fail to reject H0: Income does NOT significantly affect perceived party representation.\n")
}
```

```
## Reject H0: Income significantly affects which party minorities feel best represents them.
```

Visualization

```
ggplot(data_eq12, aes(x=income_group, fill=representing_party)) +
  geom_bar(position="fill") +
  labs(title="Which Party Best Represents You? (by Income)",
       x="Income Group", y="Proportion", fill="Party") +
  theme_minimal()
```




```
  coord_flip()
```

```
## <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>
##     aspect: function
##     backtransform_range: function
##     clip: on
##     default: FALSE
##     distance: function
##     expand: TRUE
##     is_free: function
##     is_linear: function
##     labels: function
##     limits: list
##     modify_scales: function
##     range: function
##     render_axis_h: function
##     render_axis_v: function
##     render_bg: function
##     render_fg: function
##     setup_data: function
##     setup_layout: function
##     setup_panel_guides: function
##     setup_panel_params: function
##     setup_params: function
##     train_panel_guides: function
##     transform: function
##     super:  <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>
```

The effect of nationality

```
# Load and clean the data (adjust path to actual dataset)
data_clean4 <- data %>%
  filter(
    !zqinc %in% c(-2, -1, 17),
    !bq9_1 %in% c(-2, -1, 15),
    zq101 %in% c(3, 4, 7, 8, 9, 11, 12) # Include only selected ethnic groups
  ) %>%
  mutate(
    # Simplified party ID
    party_simple = case_when(
      bq9_1 == 1 ~ "None",
      bq9_1 == 2 ~ "Labour",
      bq9_1 %in% 3:12 ~ "Other"
    ),
    # Recode ethnicity into new groups
    ethnicity = case_when(
      zq101 == 9 ~ "Bangladeshi",
      zq101 %in% c(12, 4) ~ "Black African",
      zq101 %in% c(11, 3) ~ "Black Caribbean",
      zq101 == 7 ~ "Indian",
      zq101 == 8 ~ "Pakistani"
    )
  )

# Convert ethnicity to factor
data_clean4$ethnicity <- factor(data_clean4$ethnicity)

### --- Analysis A: Ethnicity vs. Party Identification ---
ethnicity_party_table <- table(data_clean4$ethnicity, data_clean4$party_simple)
chi_ethnicity_party <- chisq.test(ethnicity_party_table)
print(chi_ethnicity_party)
```

```
##
##  Pearson's Chi-squared test
##
## data:  ethnicity_party_table
## X-squared = 85.597, df = 8, p-value = 3.628e-15
```

```
if(chi_ethnicity_party$p.value < 0.05){
  cat("\nResult: Reject the null hypothesis.\n")
  cat("Interpretation: Ethnicity significantly affects minority party identification.\n")
} else {
  cat("\nResult: Fail to reject the null hypothesis.\n")
  cat("Interpretation: Ethnicity does NOT significantly affect minority party identification.\n")
}
```

```
##
## Result: Reject the null hypothesis.
## Interpretation: Ethnicity significantly affects minority party identification.
```

```
### --- Analysis B: Ethnicity vs. Strength of Party Identification ---
# Load strength of ID and clean
strength_data <- data %>%
```

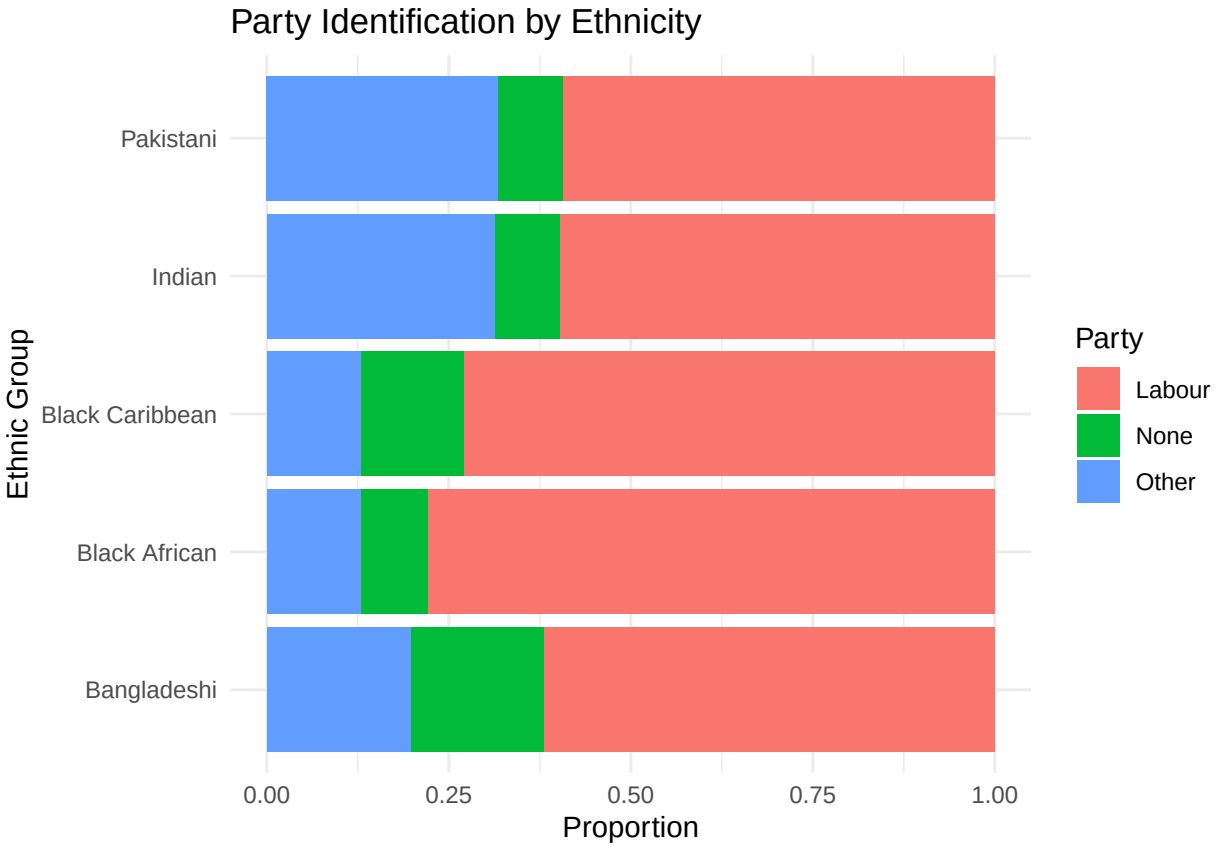
```
  select(zq101, bq9_4) %>%
  filter(zq101 %in% c(3, 4, 7, 8, 9, 11, 12), !bq9_4 %in% c(-2, -1, 6)) %>%
  mutate(
    ethnicity = case_when(
      zq101 == 9 ~ "Bangladeshi",
      zq101 %in% c(12, 4) ~ "Black African",
      zq101 %in% c(11, 3) ~ "Black Caribbean",
      zq101 == 7 ~ "Indian",
      zq101 == 8 ~ "Pakistani"
    ),
    strength_party_id = factor(bq9_4,
                               levels = c(1,2,3),
                               labels = c("Very Strong", "Fairly Strong", "Not Very Strong"),
                               ordered = TRUE)
  )

kruskal_strength <- kruskal.test(as.numeric(strength_party_id) ~ ethnicity, data = strength_data)
print(kruskal_strength)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  as.numeric(strength_party_id) by ethnicity
## Kruskal-Wallis chi-squared = 11.543, df = 4, p-value = 0.02109
```

```
if(kruskal_strength$p.value < 0.05){
  cat("\nResult: Reject the null hypothesis.\n")
  cat("Interpretation: Ethnicity significantly affects strength of party identification.\n")
} else {
  cat("\nResult: Fail to reject the null hypothesis.\n")
  cat("Interpretation: Ethnicity does NOT significantly affect strength of party identification.\n")
}
```

```
##
## Result: Reject the null hypothesis.
## Interpretation: Ethnicity significantly affects strength of party identification.
```

```
### --- Visualization (Optional) ---
ggplot(data_clean4, aes(x = ethnicity, fill = party_simple)) +
  geom_bar(position = "fill") +
  labs(
    title = "Party Identification by Ethnicity",
    x = "Ethnic Group",
    y = "Proportion",
    fill = "Party"
  ) +
  theme_minimal() +
  coord_flip()
```



Discussion

The first χ^2 test showed a significant association between income group and political party identification ($p = 0.007$), suggesting that income plays a role in shaping which party respondents identify with. The visual representation supports this, showing clear variation in party preferences across income levels.

In contrast, the Kruskal-Wallis test found no significant difference in the strength of identification across income groups ($p = 0.645$). This suggests that while income influences which party a respondent identifies with, it does not affect how strongly they feel about that identification.

The relationship between income and whether respondents feel represented by any party was not significant ($p = 0.363$). This implies that minorities across income levels are equally likely to feel either represented or unrepresented in general.

However, income significantly affected which party respondents felt best represents their views ($p = 0.0014$). This aligns with the first test and shows that while the general sense of representation may not change, the specific party seen as representative does vary by income.

Conclusion

Yes, annual household income plays a role in minority political party identification.

While income does not influence how strongly minorities feel about their chosen party or their overall sense of being represented, it significantly affects both which party they identify with and which party they feel best represents their views. These findings highlight the socioeconomic dimension of political alignment within minority communities.