Problem Set 2: Data Visualisation for Social Scientists

Qi Liu
Student No. 25340516

February 4, 2026

1 Introduction

This report details the data manipulation and visualization steps for Problem Set 2, utilizing the National Congregations Study Switzerland (NCSS). The analysis aims to explore income distributions and membership counts across various religious traditions and Swiss regions.

2 Data Manipulation

2.1 Data Loading and Selection

The dataset is imported using a relative path to ensure portability. Only the necessary variables for this problem set—CASEID, YEAR, GDREGION, NUMOFFMBR, TRAD6, TRAD12, and INCOME—are retained.

```
    library(tidyverse)

    # Load the raw dataset
    ncss_raw <- read_csv("data/NCSS_v1.csv", show_col_types = FALSE)

    # Select variables according to the assignment instructions
    ncss <- ncss_raw %>%
    select(CASEID, YEAR, GDREGION, NUMOFFMBR, TRAD6, TRAD12, INCOME)

```

Listing 1: R Code for Initial Data Selection

2.2 Sample Filtering and Descriptive Summaries

The analysis focuses on three specific religious traditions: Christian, Jewish, and Muslim congregations. We calculate the mean and median income for each tradition across the survey waves.

```
    # Filter for Christian (Chretiennes), Jewish (Juives), and Muslim (
    Musulmanes)
    ncss_cjm <- ncss %>%
    filter(TRAD6 %in% c("Chretiennes", "Juives", "Musulmanes"))

    # Generate summary statistics for congregations and income
    summary_stats <- ncss_cjm %>%
    group_by(YEAR, TRAD6) %>%
    summarise(
    n_congregations = n(),
    mean_income = mean(INCOME, na.rm = TRUE),
    median_income = median(INCOME, na.rm = TRUE),
```

```
    .groups = "drop"
    )

```

Listing 2: Filtering Traditions and Summarizing Income

2.3 Binary Income Variable Construction

A categorical variable `AVG_INCOME` is created to identify congregations that meet or exceed the annual average income.

```
    # Calculate the mean income per year
    yearly_means <- ncss_cjm %>%
    group_by(YEAR) %>%
    summarise(avg_yr_income = mean(INCOME, na.rm = TRUE), .groups = "
   drop")

    # Create the AVG_INCOME indicator (1 = Above or Average, 0 = Below)
    ncss_cjm <- ncss_cjm %>%
    left_join(yearly_means, by = "YEAR") %>%
    mutate(AVG_INCOME = ifelse(INCOME >= avg_yr_income, 1L, 0L))

```

Listing 3: Defining the Binary Income Indicator

3 Data Visualisation

3.1 Income Proportions by Religious Classification

Figure 1 shows the proportion of congregations above and below the average income level. This allows for a longitudinal comparison of financial status across traditions.

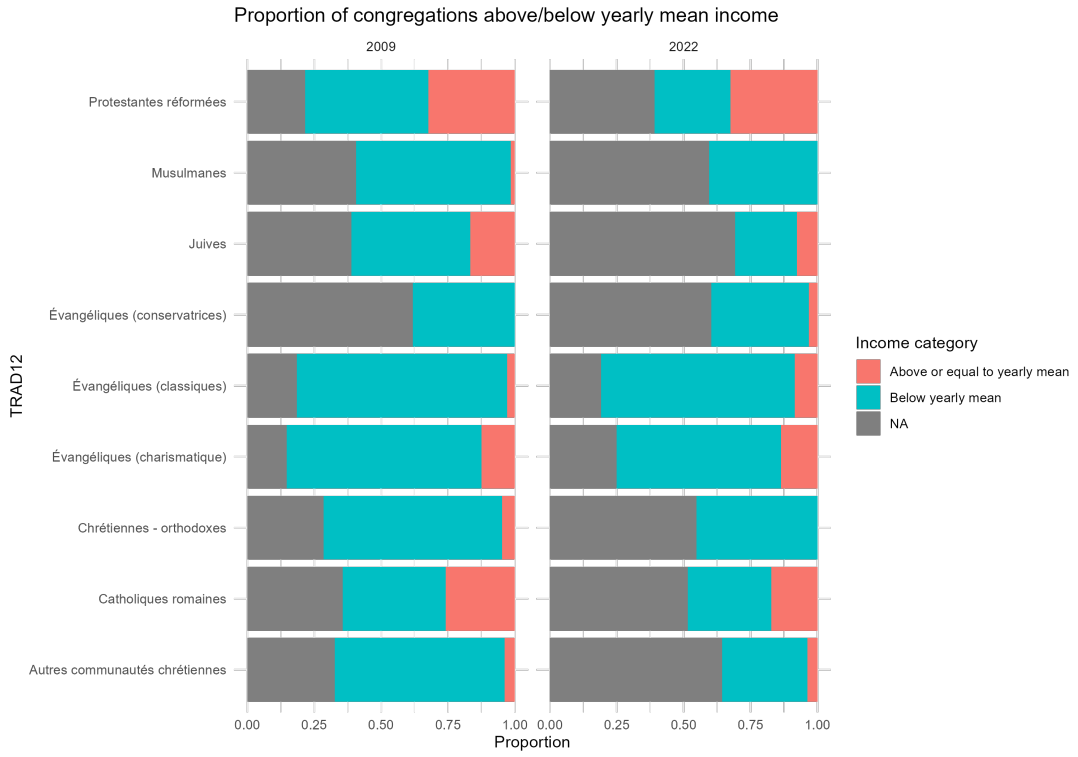

Figure 1: Proportion of congregations above/below average income by TRAD12 and YEAR.

3.2 Total Official Members (2022)

Using the 2022 data wave, Figure 2 displays the total official members aggregated by religious tradition using `geom_col()`.

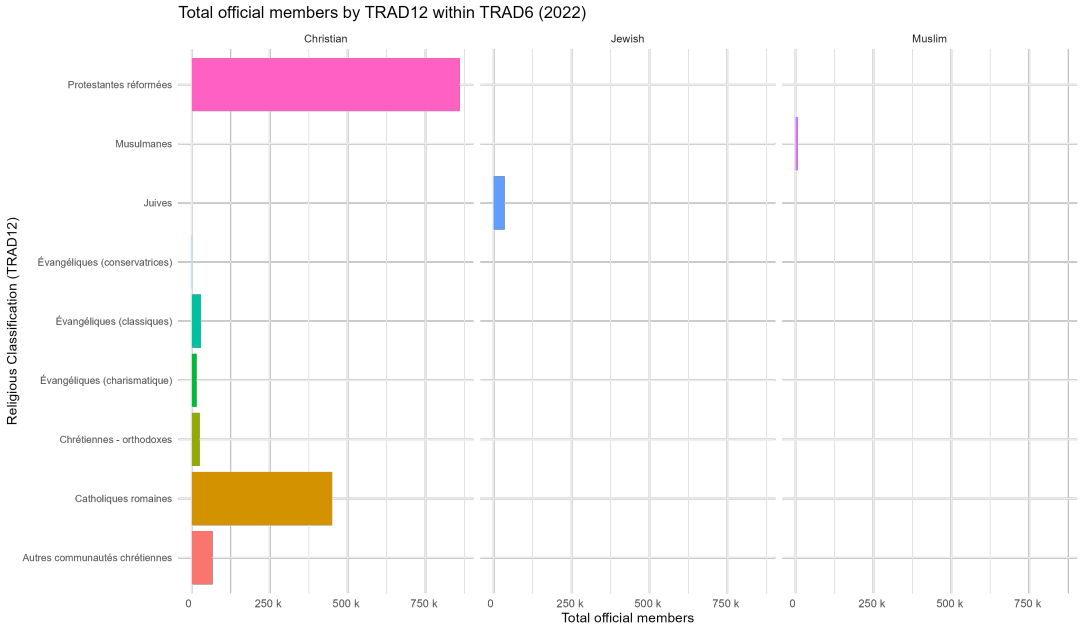

Figure 2: Total official members by TRAD12 within TRAD6 (2022).

3.3 Regional Income Distributions

The ridge plot in Figure 3 illustrates the density and distribution of income across Swiss regions for 2022.

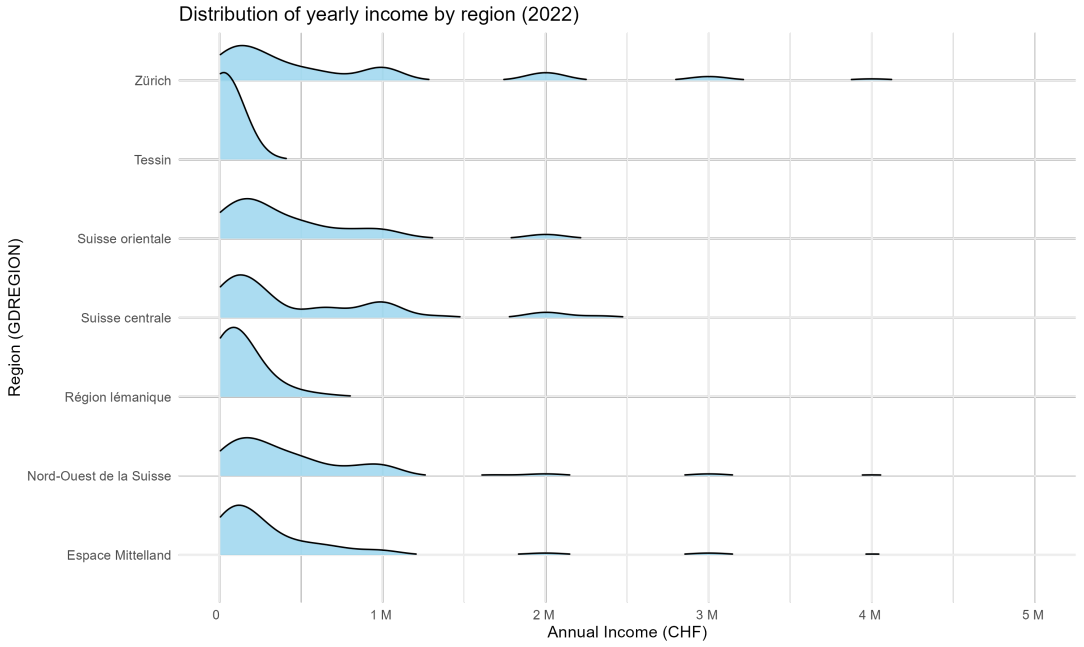

Figure 3: Distribution of yearly income by region (2022).

3.4 Congregational Membership by Region

The boxplot in Figure 4 displays the variance in official member counts per congregation across regions, faceted by geographic location.

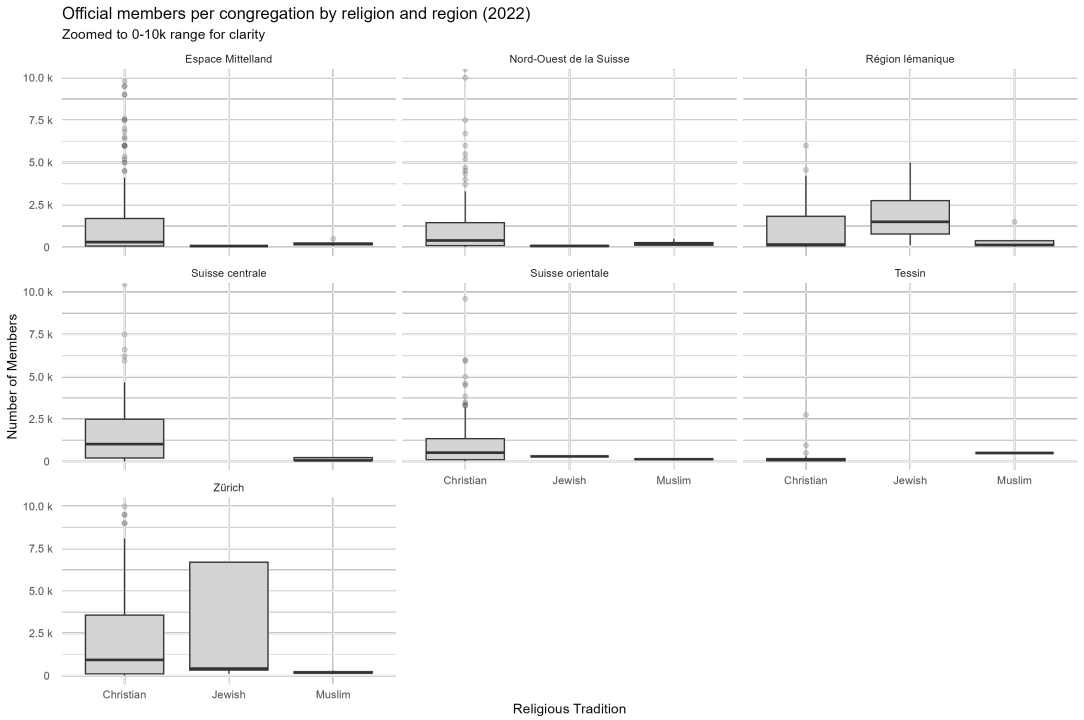

Figure 4: Number of official members per congregation by religion and region (2022).

4 Reproducibility

All figures were generated using the `ggplot2` package and exported to the `outputs/` directory to ensure results are fully reproducible.