Problem Set 2

Data Visualisation for Social Scientists

Due: February 4, 2026

Instructions

- Please show your work! You may lose points by simply writing in the answer. If the problem requires you to execute commands in R, please include the code you used to get your answers. Please also include the .R file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask.

- Your homework should be submitted electronically on GitHub.

- This problem set is due before 23:59 on Wednesday February 4, 2026. No late assignments will be accepted.

Study of Religious Congregations in Switzerland

The data for this problem set come from the National Congregations Study Switzerland (NCSS), which was conducted in 2008–2009 and 2022–2023. The data provide information on organisational structure, staffing, finances, worship practices, youth and educational activities, social composition, external engagement, and inclusion norms. The data were collected using stratified random samples of congregations drawn from comprehensive censuses, with interviews completed by a single knowledgeable key informant in each congregation, most often the spiritual leader.

Data Manipulation

1. Load the NCSS .csv file from GitHub into your global environment. Use the select() function to keep these variables in your dataframe:

   - Congregation ID (CASEID)
   - Year (YEAR)
   - Region (GDREGION)
   - Number of official members (NUMOFFMBR)
   - 6-level religious classification (TRAD6)

- 12-level religious classification (`TRAD12`)
- Total income in last fiscal year (`INCOME`)

```
1  library(readr)
2  library(dplyr)
3  #1
4  setwd("~/Desktop/whiz1/DataViz_2026/problemSets/PS02/my_answers")
5  data <- read_csv("NCSS_v1.csv") %>%
6    select(
7      CASEID,
8      YEAR,
9      GDREGION,
10     NUMOFFMBR,
11     TRAD6,
12     TRAD12,
13     INCOME
14   )
15
16 str(data)
17 head(data)
18 summary(data)
```

2. Filter the dataset so that you only include Christian, Jewish, and Muslim congregations (Chrétiennes, Juives, Musulmanes) using the `TRAD6` variable.

```
1  #2
2  ncss_filtered <- data %>%
3    filter(TRAD6 %in% c("Chrétiennes", "Juives", "Musulmanes"))
4  print(ncss_filtered)
```

3. Compute for the number of congregations by religious classification (`TRAD6`) in each year, as well as the mean and median total income in last fiscal year (`INCOME`) by religious classification and year.

```
1  #3
2  summary_trad6_year <- ncss_filtered %>%
3    group_by(YEAR, TRAD6) %>%
4    summarise(
5      n_congregations = n(),
6      mean_income = mean(INCOME, na.rm = TRUE),
7      median_income = median(INCOME, na.rm = TRUE),
8      .groups = "drop"
9    )
10 print(summary_trad6_year)
```

4. Create a categorical variable for called `AVG_INCOME` that is binary in which $1 = $ "Above average or average income" and $0 = $ "Below average income", which indicates if a congregation is $\geq$ average income or $<$ average income among congregations that year.

```
 1  #4
 2  ncss_filtered <- ncss_filtered %>%
 3    group_by(YEAR) %>%
 4    mutate(
 5      AVG_INCOME = if_else(
 6        INCOME >= mean(INCOME, na.rm = TRUE),
 7          1,
 8          0
 9      )
10    ) %>%
11    ungroup()
12  print(head(ncss_filtered))
```
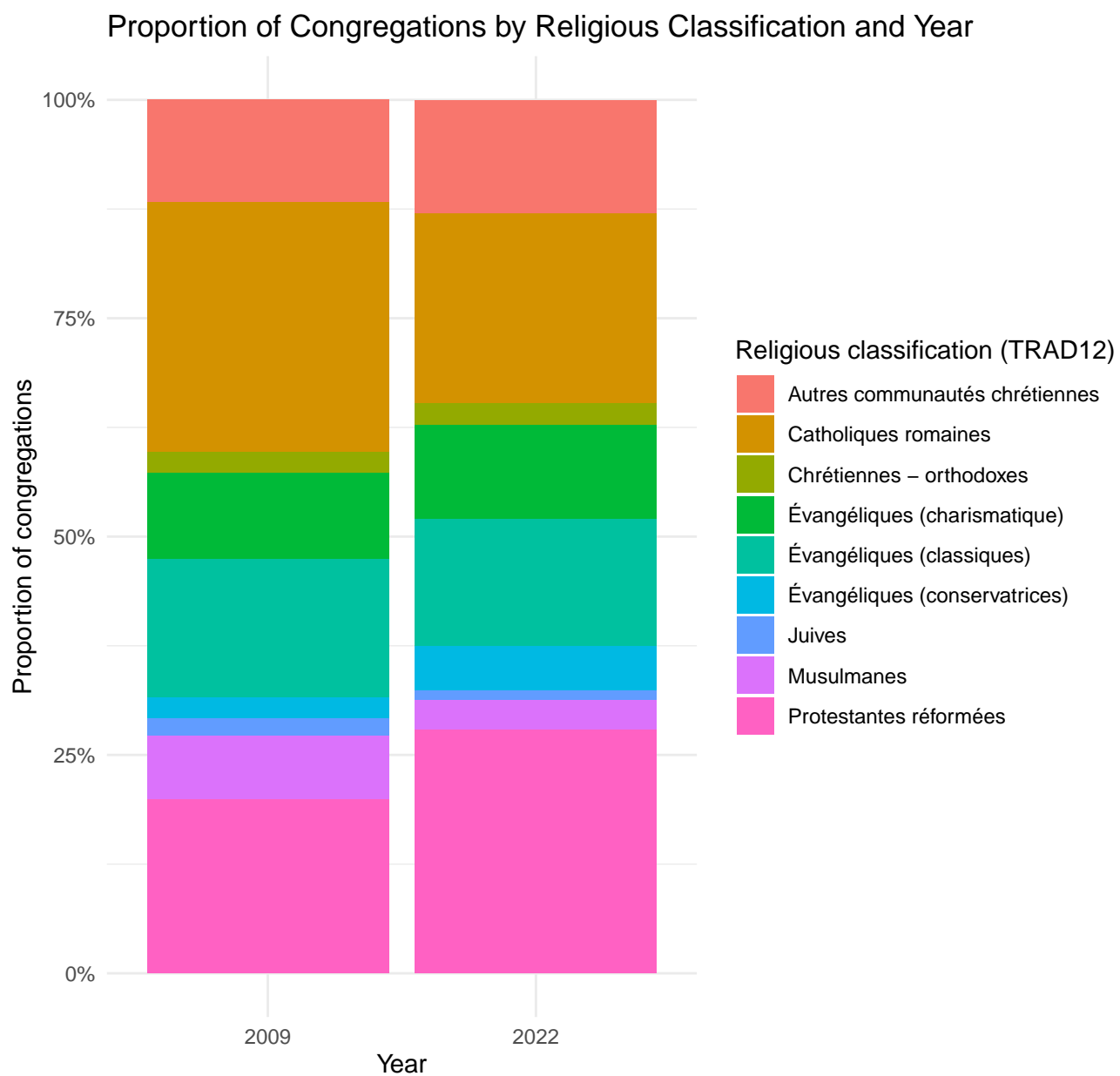
Data Visualization

1. Create a bar plot visualizing the proportion of congregations above and below the average income (AVG_INCOME) in each year by 12-level religious classification (TRAD12). Hint: Use facet() for YEAR.

```
 1  #Data Visualization
 2  #1
 3  library(ggplot2)
 4
 5  ggplot(ncss_filtered, aes(x = factor(YEAR), fill = TRAD12)) +
 6    geom_bar(position = "fill") +
 7    labs(
 8      x = "Year",
 9      y = "Proportion of congregations",
10      fill = "Religious classification (TRAD12)",
11      title = "Proportion of Congregations by Religious Classification and
       Year"
12    ) +
13    scale_y_continuous(labels = scales::percent_format()) +
14    theme_minimal()
```
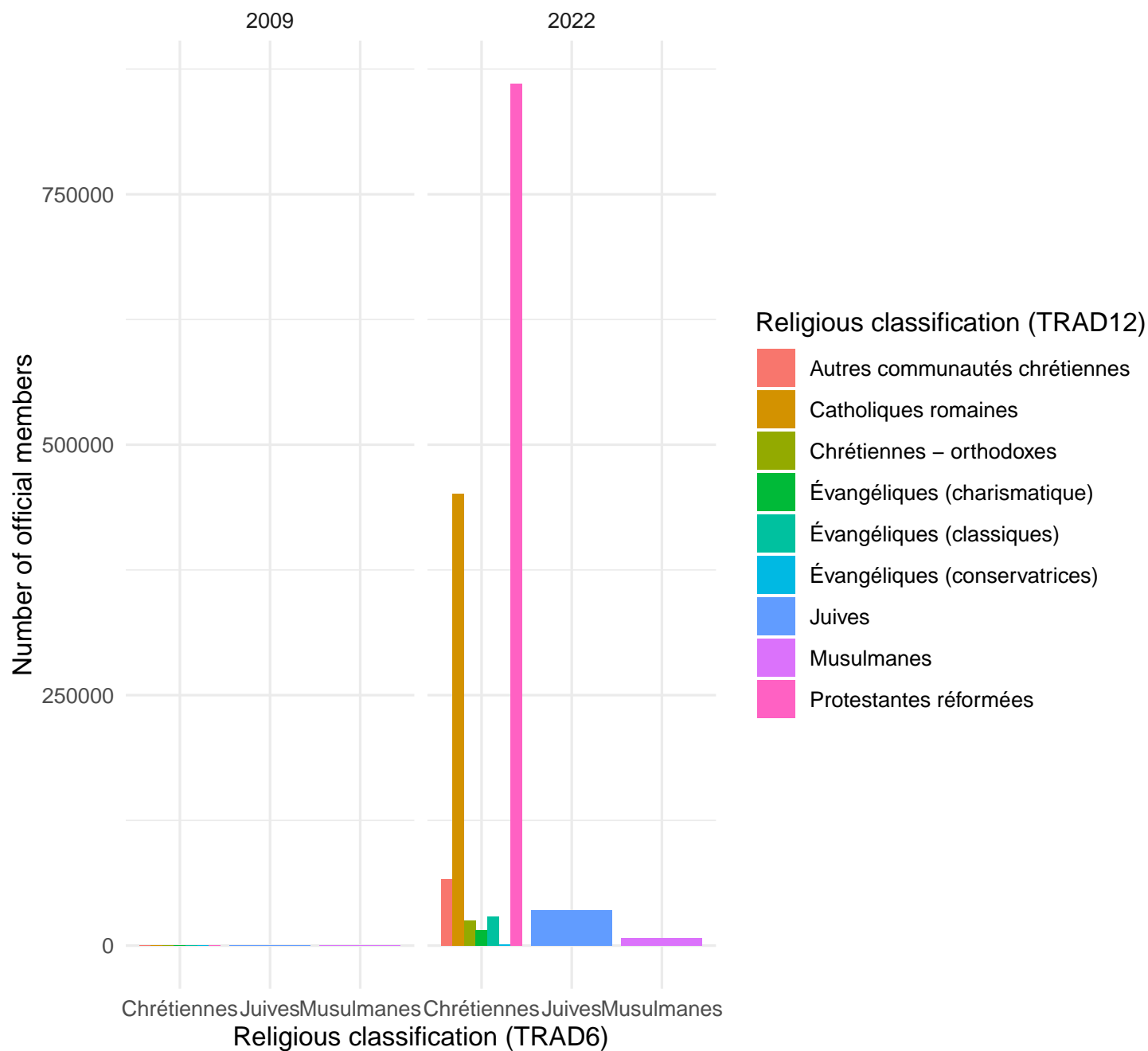
Proportion of Congregations by Religious Classification and Year

2. Make a histogram using `geom_col()` detailing the number of official members using the 12-level religious classification (`TRAD12`) distinguishing between the 6-level religious classification (`TRAD6`) in 2022. Hint: Use `facet()` for `TRAD6`, with `TRAD12` on the x-axis in addition to group/fill with the `position="dodge"`.

```
1  #2
2  ggplot(ncss_filtered, aes(x = TRAD6, fill = TRAD12, weight = NUMOFFMBR))
       +
3    geom_bar(position = "dodge") +
4    facet_wrap(~ YEAR) +
5    labs(
6      x = "Religious classification (TRAD6)",
7      y = "Number of official members",
8      fill = "Religious classification (TRAD12)",
9      title = "Distribution of Official Members by Religious Classification
         and Year"
10   ) +
11   theme_minimal()
```
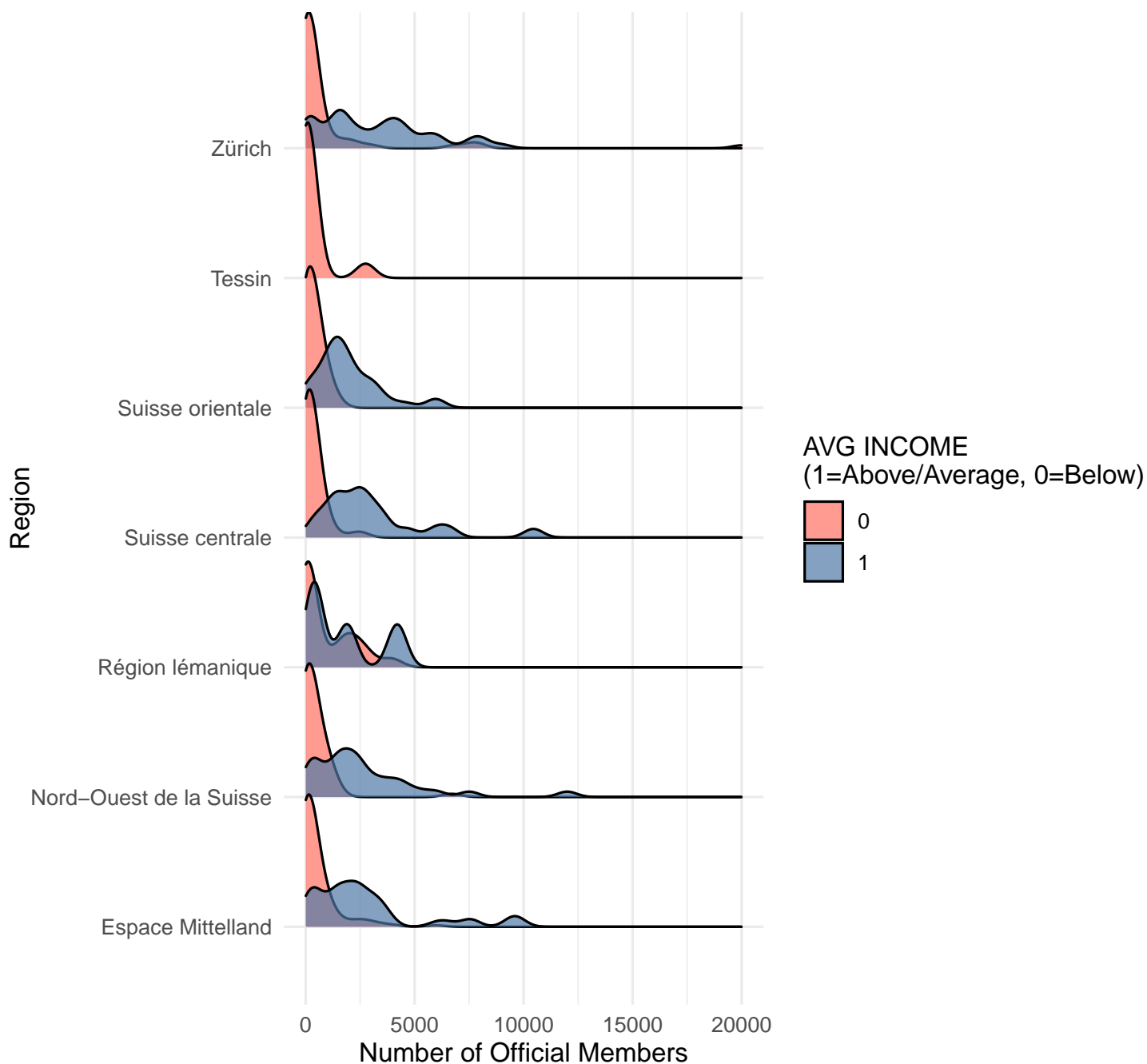
Distribution of Official Members by Religious Classification and Year

Religious classification (TRAD12)

- Autres communautés chrétiennes
- Catholiques romaines
- Chrétiennes – orthodoxes
- Évangéliques (charismatique)
- Évangéliques (classiques)
- Évangéliques (conservatrices)
- Juives
- Musulmanes
- Protestantes réformées

Number of official members

Religious classification (TRAD6)

2009 — Chrétiennes Juives Musulmanes

2022 — Chrétiennes Juives Musulmanes

3. Display the distribution of yearly income (`INCOME`) in 2022 for congregations in each region (`GDREGION`) using ridge plots.

```r
#3
install.packages("ggridges")
library(ggridges)
#Filter for 2022 and remove NAs in NUMOFFMBR or AVG_INCOME
ncss_2022 <- ncss_filtered %>%
  filter(YEAR == 2022, !is.na(NUMOFFMBR), !is.na(AVG_INCOME))

#Ridge plot
ggplot(ncss_2022, aes(x = NUMOFFMBR, y = GDREGION, fill = factor(
    AVG_INCOME))) +
  geom_density_ridges(alpha = 0.7, scale = 1.2) +
  labs(
    x = "Number of Official Members",
    y = "Region",
    fill = "AVG INCOME\n(1=Above/Average, 0=Below)",
    title = "Distribution of Congregations by Income Level and Region
    (2022)"
  ) +
  scale_fill_manual(values = c("0" = "#FF6F61", "1" = "#4E79A7")) +
  scale_x_continuous(limits = c(0, 20000)) +
  theme_minimal()
```

Distribution of Congregations by Income Level and Region (2022)

4. Create a boxplot of the number of official members per congregation in 2022 by religious classification (`TRAD6`) and region (`GDREGION`). Hint: Use `facet()` for `GDREGION`.

```
1  #4
2  ncss_box <- ncss_filtered %>%
3    filter(YEAR %in% c(2009, 2022))
4
5  #Boxplot
6  ggplot(ncss_box, aes(x = GDREGION, y = NUMOFFMBR, fill = factor(YEAR))) +
7    geom_boxplot(outlier.alpha = 0.3, alpha = 0.7, position =
       position_dodge(width = 0.8)) +
8    labs(
9      x = "Region",
10     y = "Number of Official Members",
11     fill = "Year",
12     title = "Number of Official Members by Region (2009 vs 2022)"
13   ) +
14   scale_y_continuous(limits = c(0, 15000)) +
15   scale_fill_manual(values = c("2009" = "#FF6F61", "2022" = "#4E79A7")) +
16   theme_minimal() +
17   theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Number of Official Members by Region (2009 vs 2022)