Problem Set 2

Jeffrey Ziegler

Data Visualisation for Social Scientists

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 NCSS <- read.csv("https://raw.githubusercontent.com/ASDS-TCD/DataViz_2026
    /refs/heads/main/datasets/NCSS_v1.csv")
2 # subset to selected variables
3 NCSS <- NCSS |>
4    select(
5      CASEID,        # Congregation ID
6      YEAR,          # Survey year
7      GDREGION,      # Region
8      NUMOFFMBR,     # Number of official members
9      TRAD6,         # 6-level religion
10     TRAD12,        # 12-level religion
11     INCOME         # Total income last fiscal year
12   )
```

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

```
1 # filter to Christian, Jewish, and Muslim congregations
2 NCSS <- NCSS |>
3    filter(TRAD6 %in% c("Chr tiennes", "Juives", "Musulmanes"))
```

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 NCSS |>
2    group_by(TRAD6, YEAR) |>
3    summarise(
4      n_congregations = n(),
5      mean_income     = mean(INCOME, na.rm = TRUE),
6      median_income   = median(INCOME, na.rm = TRUE),
7      .groups = "drop"
8    )
```

	TRAD6	YEAR	n_congregations	mean_income	median_income
1	Chrétiennes	2009	802	539942.35	200000.00
2	Chrétiennes	2022	1172	474600.50	201000.00
3	Juives	2009	18	330908.73	200000.00
4	Juives	2022	13	2332500.00	115000.00
5	Musulmanes	2009	64	62238.16	25000.00
6	Musulmanes	2022	42	77941.18	42500.00

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  NCSS <- NCSS |>
2    group_by(YEAR) |>
3    mutate(
4      year_mean_income = mean(INCOME, na.rm = TRUE),
5      AVG_INCOME      = if_else(
6        INCOME >= year_mean_income, "Above Average", "Below Average"
7      )
8    )
```

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  NCSS$AVG_INCOME <- as.factor(NCSS$AVG_INCOME)
2  ggplot(NCSS |>
3             group_by(YEAR, TRAD12, AVG_INCOME) |>
4             summarise(n = n()) |>
5             mutate(prop = n / sum(n)),
6           aes(x = TRAD12, y = prop, fill = AVG_INCOME)) +
7    geom_col() +
8    scale_y_continuous(labels = scales::percent_format()) +
9    labs(x = "12-level religious classification (TRAD12)",
10        y = "Proportion of congregations",
11        fill = "Income category") +
12   facet_wrap(~ YEAR) +
13   coord_flip() +
14   theme_bw() +
15   theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

2. *Make a bar plot with `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 `TRAD12` on the x-axis.*

```
1  NCSS_2022 <- NCSS |> filter(YEAR == 2022)
2  ggplot(NCSS_2022 |>
3             group_by(TRAD6, TRAD12) |>
4             summarise(total_members = sum(NUMOFFMBR, na.rm = TRUE)),
5           aes(x = TRAD6, y = total_members, fill = TRAD12)) +
6    geom_col(width = 0.5, position = position_dodge(width=1)) +
7    labs(x = "Religious classification (TRAD6)",
8         y = "Total number of official members\n",
9         fill = "12-level Religious Classification (TRAD12)") +
10   guides(fill=guide_legend(ncol=3))+
11   theme_bw() +
12   theme(axis.text.x = element_text(angle = 45, hjust = 1, size=12),
13         legend.position = "bottom",
14         axis.title = element_text(size=14))
```

Figure 1: Proportion of congregations above and below the average income by year and religious classification.
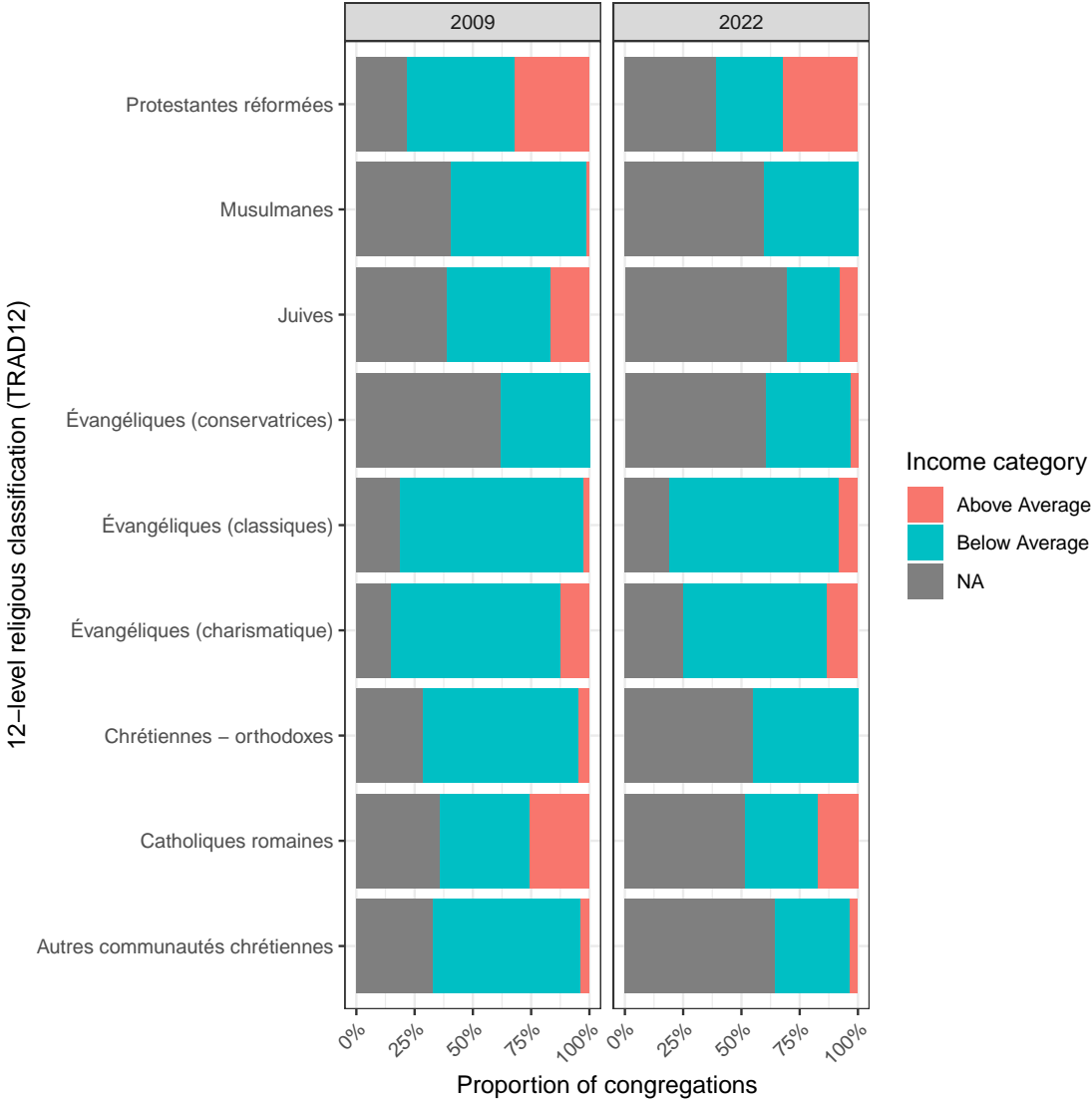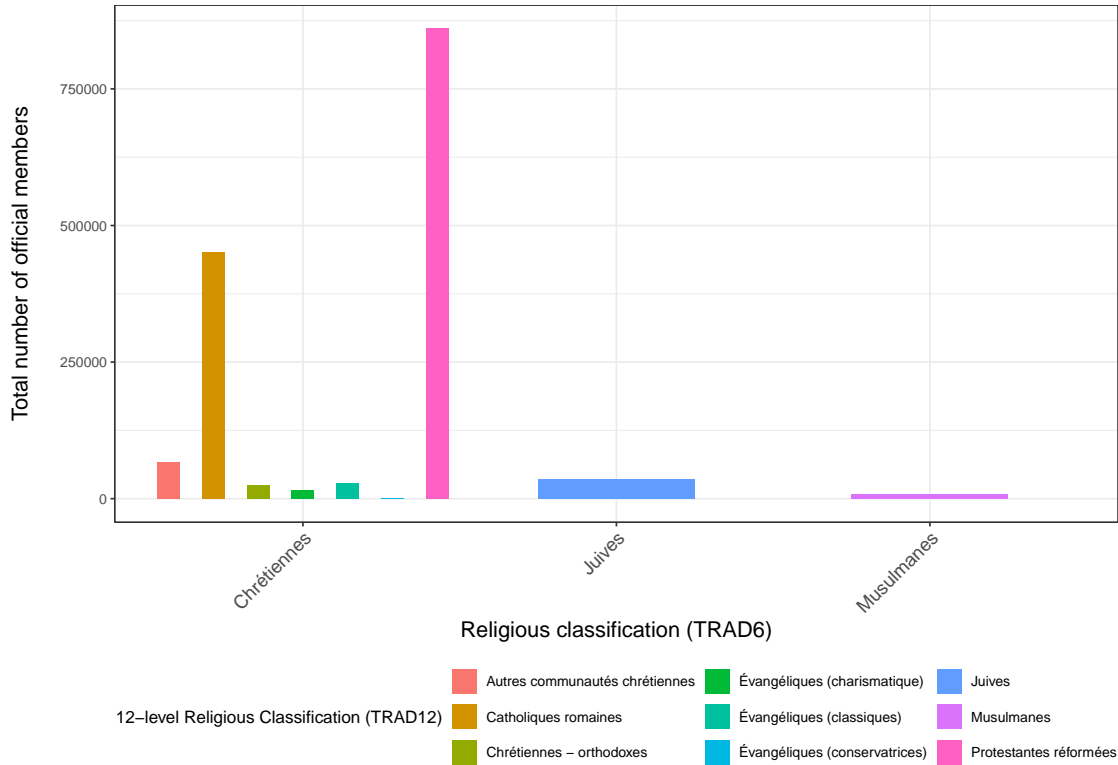
Figure 2: Number of official members by religious classification in 2022.

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

```
ggplot(NCSS_2022, aes(x = INCOME, y = GDREGION, fill = GDREGION)) +
  geom_density_ridges(alpha = 0.7, color = "white", scale = 1) +
  labs(x = "Total income in last fiscal year (INCOME)",
       y = "Region (GDREGION)") +
  theme_bw() +
  lims(x=c(0, 9000000))+
  theme(legend.position = "none")
```

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`.*

```
ggplot(NCSS_2022, aes(x = TRAD6, y = NUMOFFMBR)) +
  geom_boxplot() +
  labs(x = "6-level religious classification (TRAD6)",
       y = "Number of official members per congregation (NUMOFFMBR)") +
  facet_wrap(~ GDREGION) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

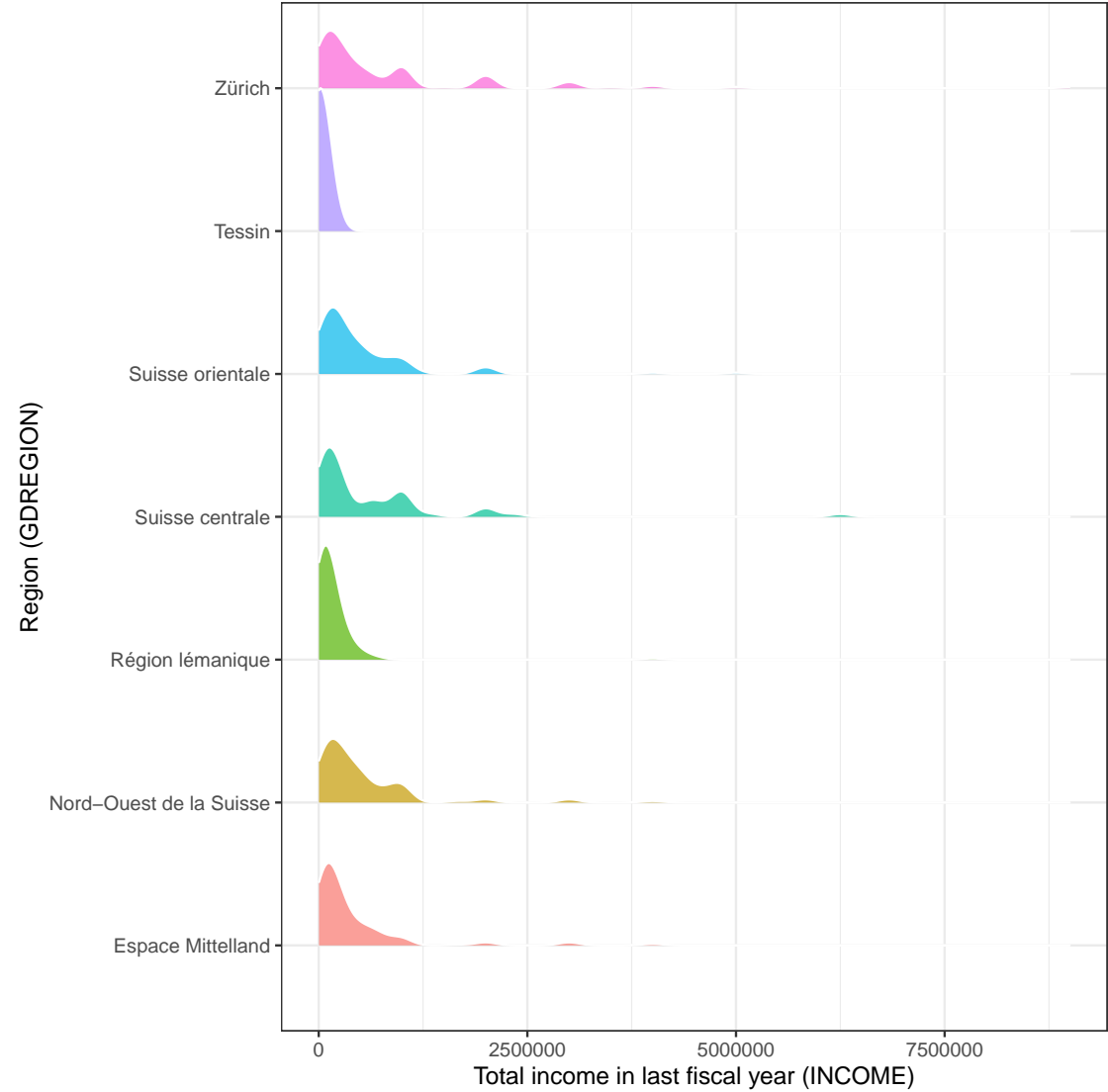Figure 3: Distribution of average yearly income of congregations by region in 2022.

Figure 4: Distribution of the number of official members of congregations by religious classification in 2022.