

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

## 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**.
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 **facet()** for **TRAD6**, with **TRAD12** on the x-axis.
3. Display the distribution of average yearly income (**INCOME**) for congregations in 2022 in each region (**GDREGION**) using ridge plots.
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**.