

# Problem Set 3

## Data Visualisation for Social Scientists

Due: February 18, 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 18, 2026. No late assignments will be accepted.

### Canadian Election Study

The data for this problem set come from the Canadian Election Study (CES) in 2015. The main purpose of the study is to give a comprehensive picture of the Canadian election: why people vote as they do, what changes during campaigns and across elections, and how Canadian voting compares with that in other democracies.

### Data Manipulation

1. Load the CES .csv file from GitHub into your global environment. Filter respondents to only include "high quality" participants:

```
ces2015 <- ces2015 |> filter(discard == 0)
```

2. Recode potential turnout as a binary variable using **voted\_for**. Consider respondents who gave a valid party as having voted, while "none/spoiled ballot" as not having voted. Treat "Don't know" (98) and "Refused" (99) as missing.
3. Create an age variable and group into categories (e.g., <30, 30-44, 45-64, 65+). Year of birth is in age (four-digit year).

## Data Visualization

1. Plot turnout rate by age group.
2. Create a density plot of ideology by party, restricting your sample to respondents with non-missing left–right self-placement (0–10 scale) and those that intended to vote for a main party (e.g., Liberal, Conservative, NDP, Bloc in Quebec, and Green).
3. Produce bar charts of turnout by income (`income_full`), faceted by province.
4. Create your own reusable custom theme. Apply your theme to one of the previous plots and add:
  - (a) An improved title summarizing the main substantive takeaway.
  - (b) A more informative subtitle describing the sample and variables.
  - (c) A caption noting data source, weighting, and key coding decisions.
  - (d) At least one direct annotation using `ggrepel` that calls out a key pattern.