

Problem Set 3

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 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.

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

```
ces2015 <- ces2015 |> filter(discard == "Good quality")
```

```
1 # load data from URL
2 ces2015 <- read.csv("https://raw.githubusercontent.com/ASDS-TCD/DataViz_2026/refs/heads/main/datasets/CES2015.csv")
3 # keep good-quality cases and main weight
4 ces2015 <- ces2015 |>
5   filter(discard == "Good quality") |>
6   # subset to necessary variables
7   select(discard, age, p_voted, p_votechce, province, income_full, p_selfplace)
```

2. Filter the dataset to those participants that answered the question about voting for the past election using `p_voted`. Consider respondents who gave a "Yes" answer as having voted, while "No" as not having voted. Treat "Don't know" and "Refused" as missing.

```
1 ces2015 <- ces2015 |>
2   filter(p_voted == "Yes" | p_voted=="No")
```

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).

```
1 ces2015$age_num <- as.numeric(ces2015$age)
2 ces2015 <- ces2015 |>
3   mutate(
4     age_cat = case_when(
5       2015 - age_num < 30 ~ "<30",
6       2015 - age_num < 45 ~ "30-44",
7       2015 - age_num < 65 ~ "45-64",
8       2015 - age_num >= 65 ~ "65+"
9     ),
10    age_cat = factor(age_cat, levels = c("<30", "30-44", "45-64", "65+"))
11  )
```

Data Visualization

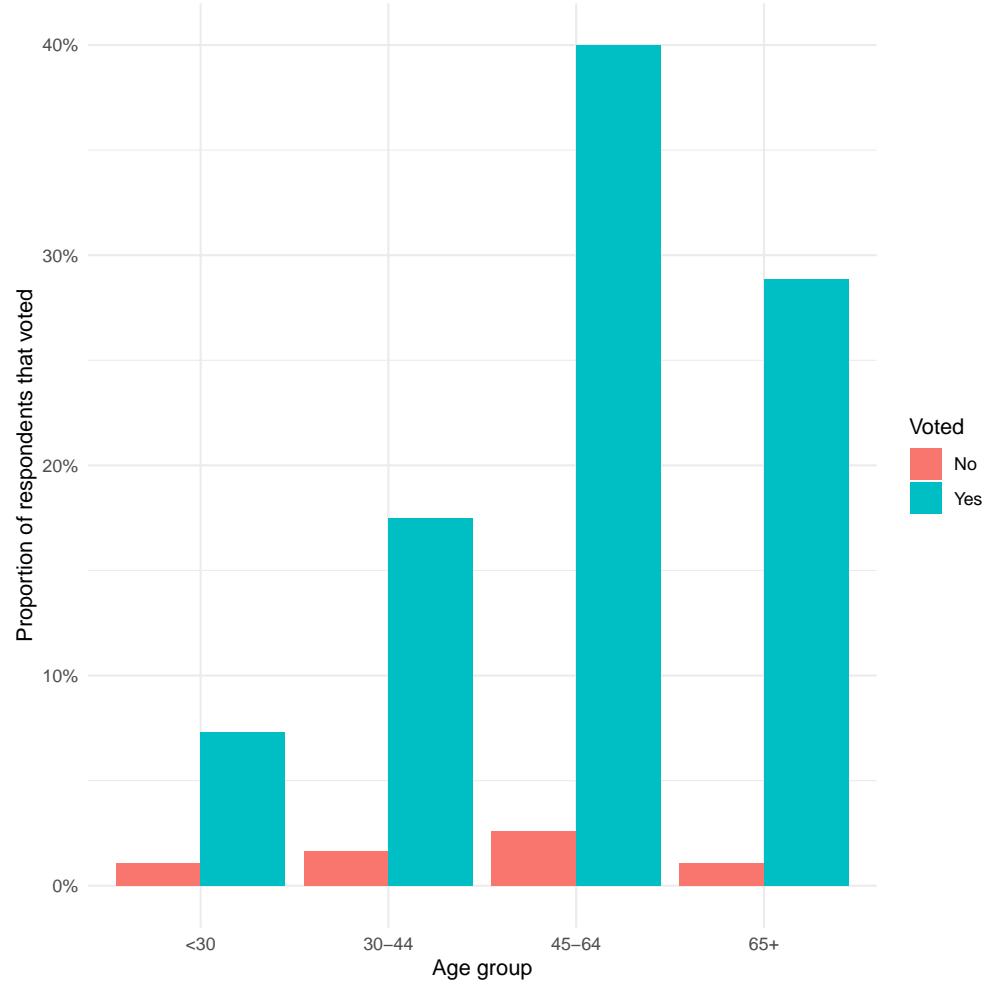
1. Plot turnout rate by age group.

```
1 ggplot(ces2015 |> filter(!is.na(age_cat)),
2   aes(x = age_cat, fill=p_voted)) +
3   geom_bar(aes(y =(..count..)/sum(..count..)), position="dodge") +
4   scale_y_continuous(labels = scales::percent_format()) +
5   labs(
6     x = "Age group",
7     y = "Proportion of respondents that voted",
8     fill = "Voted"
9   ) +
10  theme_minimal()
```

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).

```
1 plot2_data <- ces2015[ces2015$p_votechce %in% c("Conservatives", "Liberal",
2   "ndp", "Green Party", "Bloc Quebecois"),]
2 plot2_data <- plot2_data[plot2_data$p_selfplace!="" & plot2_data$p_
3   selfplace!="1000",]
4 plot2_data$p_selfplace <- as.numeric(plot2_data$p_selfplace)
5 ggplot(plot2_data,
6   aes(x = p_selfplace, y=p_votechce, fill = p_votechce)) +
7   geom_density_ridges(alpha = 0.5, scale = .9) +
8   labs(
      x = "Left-right self-placement (0 = left, 10 = right)",
```

Figure 1: Turnout rate in Canadian federal election as a percentage of all possible voters by age group.

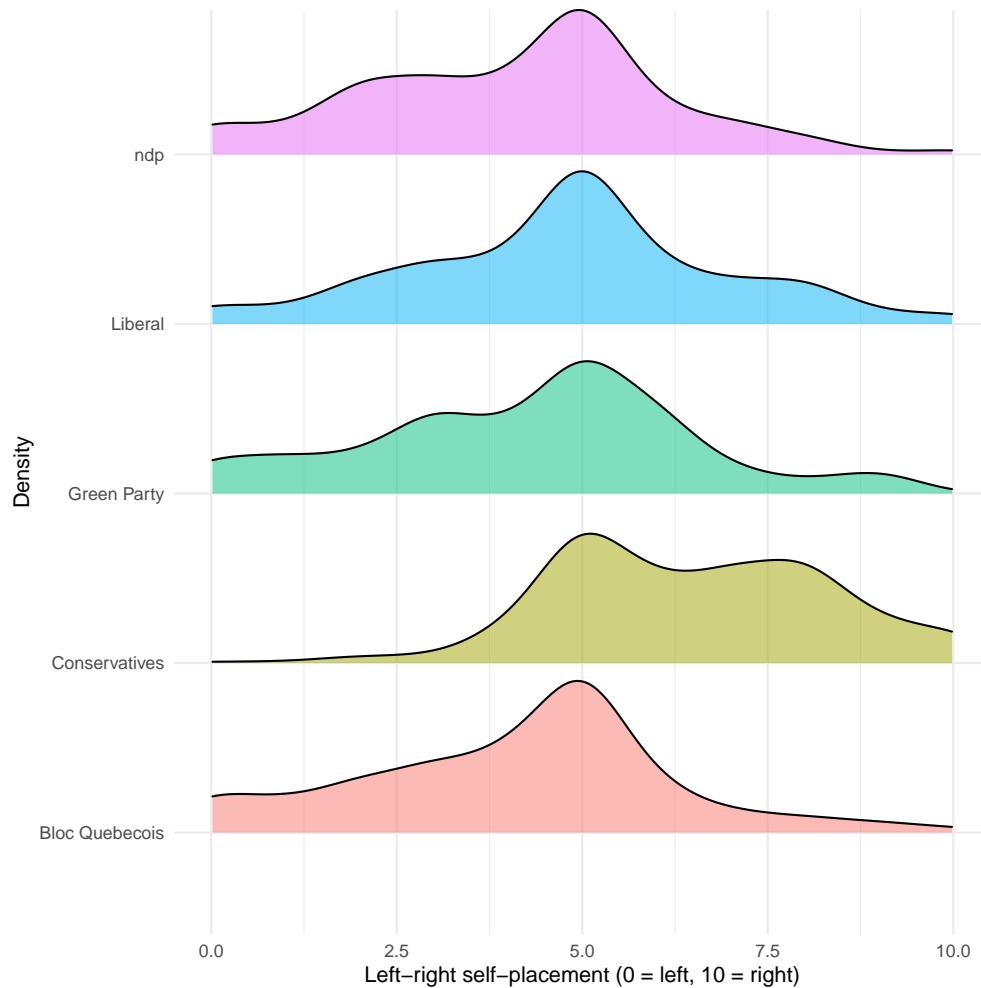


```

9   y = "Density"
10  ) +
11  lims(x=c(0,10))+

```

Figure 2: Ideological placement of party voters in Canadian federal election.



3. Produce histogram counts of turnout by income (*income_full*), faceted by province.

```

1 # get only real income values
2 plot3_data <- ces2015[ces2015$income_full %in% c("less than $29,999", "
  more than $110,000", "between $60,000 and $89,999", "between $30,000
  and $59,999", "between $90,000 and $109,999"),]
3 # make income_full a factor
4 plot3_data$income_full <- factor(plot3_data$income_full, levels = c("less
  than $29,999", "between $30,000 and $59,999", "between $60,000 and $89,999",
  "between $90,000 and $109,999", "more than $110,000"))
5 # change labels for income_full
6 levels(plot3_data$income_full) <- c("<$30k", "$30-60k", "$60-90k", "$
  90-110k", ">$110k")

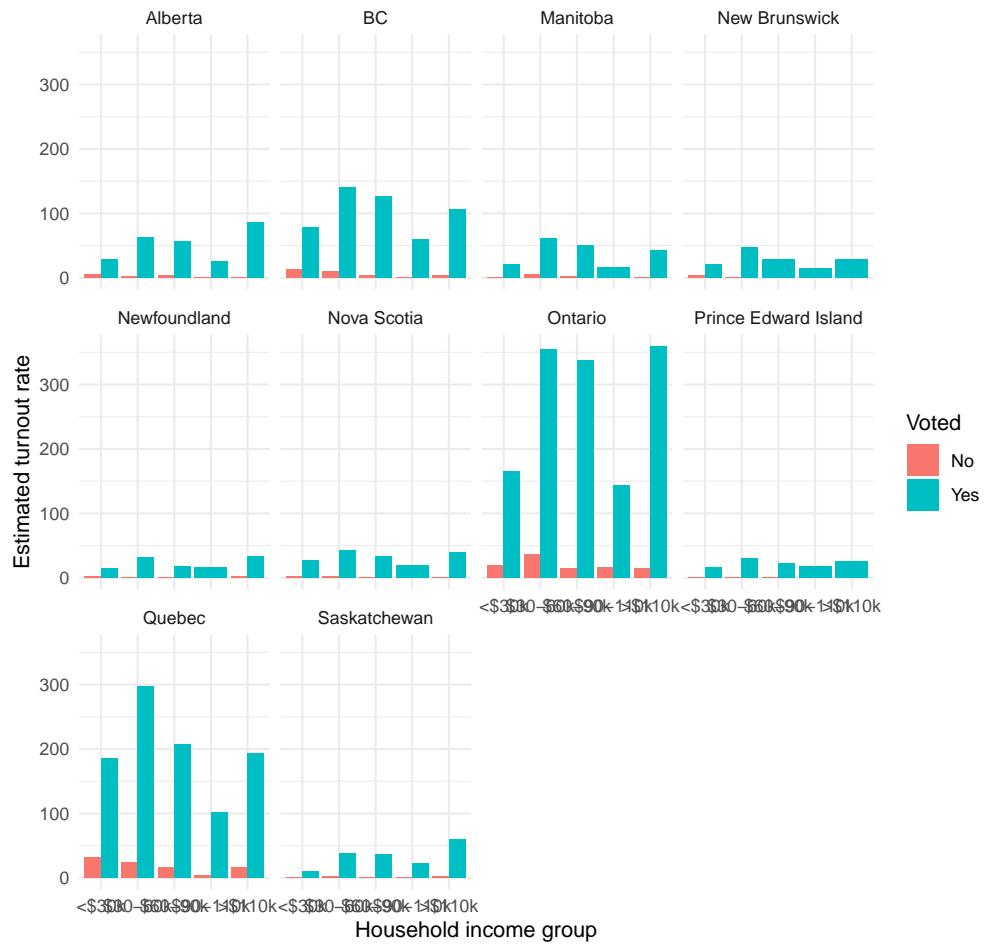
```

```

7 # make province a factor
8 plot3_data$province <- as.factor(plot3_data$province)
9 # change labels for province
10 levels(plot3_data$province) <- c("Alberta", "BC", "Manitoba", "New
    Brunswick", "Newfoundland", "Nova Scotia", "Ontario", "Prince Edward
    Island", "Quebec", "Saskatchewan")
11 plot3 <- ggplot(plot3_data, aes(x = income_full, fill=p_voted)) +
12 geom_histogram(stat="count", position="dodge") +
13 facet_wrap(vars(province)) +
14 labs(
15   x = "Household income group",
16   y = "Estimated turnout rate",
17   fill="Voted",
18   title = "Turnout by income and education, CES 2015"
19 ) +
20 theme_minimal()

```

Figure 3: Turnout in Canadian federal election by region.
Turnout by income and education, CES 2015



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 that calls out a key pattern.

```

1 plot4 <- plot3 +
2   labs(
3     x = "Turnout among potential voters",
4     y = NULL,
5     title = "Turnout does not appear to be linked to income",
6     subtitle = "Turnout by region in Canadian federal election",
7     caption = "Source: CES 2015 combined file",
8     tag = "*Turnout was strongest in Ontario and Quebec"
9   ) +
10  theme_ces() +
11  coord_cartesian(ylim = c(0, 350), clip = "off") +
12  theme(axis.text.x = element_text(angle = 45, hjust=1),
13    plot.tag.position = c(.7, .25))

```

Figure 4: Turnout in Canadian federal election by region.

Turnout does not appear to be linked to income

Turnout by region in Canadian federal election

