

# In-Class Problem Set: Color Encodings with Senate Ideology Data (R + GitHub)

**Goal.** Practice using color intentionally (not decoratively) by visualizing ideology in the U.S. Senate across four time points. You will (i) pull the data from GitHub, (ii) subset to the target Senates, (iii) make one plot for each time point with color encoding ideology, (iv) interpret what you see, and (v) submit via GitHub.

## What to submit (in your GitHub repo).

- A script file: `scripts/lab.R`
- A short write-up: `outputs/writeup.md`
- Four figures (one per time point) saved to `figures/`

## Rules (read carefully).

- Work inside an **R Project**.
- Use a **sequential, hard-coded workflow** (no user-defined functions).
- Your plots must:
  - set a non-default plot background color,
  - use colors that are intuitive for the task,
  - use an accessibility-conscious palette (colorblind-friendly),
  - and **explicitly justify your color scale choice** (sequential vs diverging) in your write-up.
- Save outputs using `ggsave()` (no screenshots).
- Git commands go in the **Terminal tab** (not the R Console).

## Mini codebook (use this; do not guess)

- **Which Senates to use.** For this problem set, use the following Congress numbers to represent the four target time points:
  - **1990** → **101st Congress**
  - **2000** → **106th Congress**
  - **2010** → **111th Congress**
  - **2020** → **116th Congress**
- **What DW-NOMINATE is.** DW-NOMINATE is an ideology scaling procedure based on roll-call voting. In most datasets:
  - `dwnom1` is the primary (left-right) ideological dimension,
  - `dwnom2` is a secondary ideological dimension (often smaller and context-dependent).
- **What you will plot.** Your scatterplots should use the two ideology dimensions (typically `dwnom1` on x and `dwnom2` on y), and **color points by ideology** (typically `dwnom1`) unless your course codebook specifies a different ideology column.
- **Color scale rule (required).**

- If the ideology variable has meaningful sign around 0 (e.g., negative vs positive), use a **diverging** scale centered at 0.
  - If you treat ideology as magnitude only (no meaningful center), use a **sequential** scale.
- In your write-up, state which rule you used and why.

## Questions

- 1. Pull the correct Senate data from GitHub (proof required).**
  - In the **Terminal tab**, run:
 

```
git status
git pull
```
  - Confirm the Senate ideology file exists in your repo (the exact file name/path is in the course repo; for most of you it will be in `data/`).
  - Proof (write-up):** In `outputs/writeup.md`, paste:
    - the output of `getwd()` from inside your R Project, and
    - the output of `list.files("data")` showing the Senate file.
- 2. Load and summarize the dataset.**
  - In `scripts/lab.R`, load the dataset into an object named `df`.
  - Summarize the dataset in a way that supports your next steps. At minimum include:
    - `dim(df)`
    - `names(df)`
    - a focused summary of the key columns you will use (time/Congress, chamber, ideology).
  - Proof (write-up):** Report:
    - number of rows and columns,
    - the column you will use for Congress/time,
    - the column you will use to identify the Senate chamber (if applicable),
    - the ideology columns you will use (e.g., `dwnom1`, `dwnom2`).
- 3. Subset to the four target Senates (required).**
  - Subset the data so it contains only **Senate** observations for the following Congresses:

$$\{101, 106, 111, 116\}.$$

- Save the subset as `df4`.
- Proof (write-up):** Include counts that confirm:
  - only these four Congresses appear in `df4`, and
  - `df4` contains only Senate observations (not House).

- 4. Make four plots (one per Congress), with color encoding ideology (required).**  
For each of the four Congresses, create a scatterplot using the two DW-NOMINATE dimensions:

- x-axis: `dwnom1`
- y-axis: `dwnom2`
- color: ideology (typically `dwnom1`)

**Required design constraints (integrated).** Each plot must:

- set a non-default plot background color,
- use an accessibility-conscious palette,
- use a color scale that is **intuitive for the task**:

- **diverging centered at 0** if ideology sign matters, or
- **sequential** if you treat ideology as magnitude only,
- label axes and the legend clearly (what the variable is).

Save figures to `figures/` with clear names, for example:

- `figures/senate_101.png`
- `figures/senate_106.png`
- `figures/senate_111.png`
- `figures/senate_116.png`

## 5. Interpretation (write-up required).

In `outputs/writeup.md`, write 10–14 sentences answering:

- What does color represent in your plots (which variable, which direction, what range)?
- Compare the earliest vs latest Congress in your set: what changed in ideological separation and dispersion?
- **Color-scale justification (required):** Did you use a sequential or diverging scale? Why does that choice match the meaning of ideology in your plot?
- Accessibility: state one concrete decision you made to improve accessibility (palette choice, contrast with background, labeling).

## 6. Push to GitHub (proof required).

- (a) In the **Terminal tab**, run:

```
git status
git add .
git commit -m "Color encodings lab: Senate ideology plots"
git push
```

- (b) **Proof (write-up):** Paste:

- the output of `git status` after committing (clean working tree), and
- the output of `git log -1`.

## Optional challenge (if you finish early)

Create a second version of one Congress plot that changes exactly **one** design element:

- Switch sequential  $\leftrightarrow$  diverging (and explain why the alternative is worse), *or*
- Keep the same palette but change the background color and explain how contrast changes readability.

In 5–7 sentences, argue which version is better for (i) a general audience and (ii) an expert audience.

## Checklist (before you leave)

- `scripts/lab.R` exists and runs top-to-bottom
- `outputs/writeup.md` exists and includes interpretation + proofs
- Four figures saved in `figures/` (101, 106, 111, 116)
- Work is committed and pushed to GitHub