# DA 332 (2025)

# Assignment 4

---

**Instructions:**

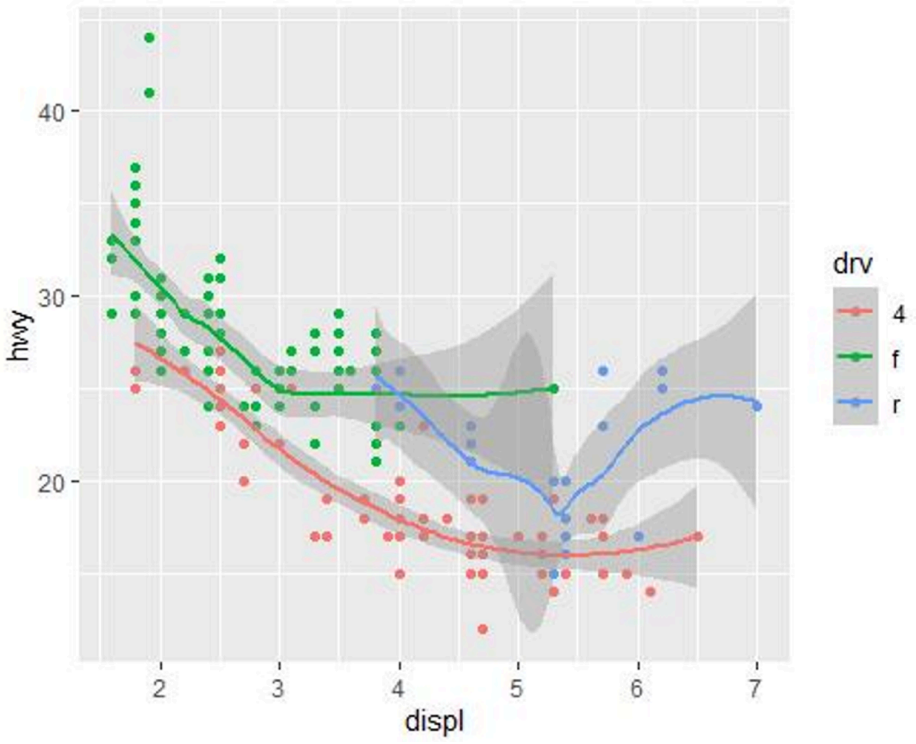
- The instructions in this assignment are specific to R. If you decide to work in another language and you require the data, you can export it from Rstudio/Posit Cloud.
- Submit your results in this form.
- Add a title, axis labels, and a legend to your plots.

---

1. The name of the dataset that you will be working is **mpg**.
    a. You are analyzing the relationship between highway mileage (**hwy**) and the number of cylinders (**cyl**) in a dataset. You want to create a scatter plot where:
        i. Points are colored based on the drivetrain (**drv**).
        ii. A linear regression line (**lm**) is added to show trends.
        iii. The color scheme follows the **Viridis** color scale for better visibility.
        iv. Use appropriate titles and labels for your plot.

    b. Create 3 scatterplots of **displ** vs **hwy** mapping **cyl** to color, shape, and size. How would these aesthetics behave differently for categorical vs. continuous variables?

    c. Recreate the R code necessary to generate this plot.



d. Create a scatter plot of **cty** (city miles per gallon) on the x-axis and **hwy** (highway miles per gallon) on the y-axis. Facet the plot by class (car class) and allow the scales for both axes to vary freely across facets (scales = "free"). Apply the **theme_bw()** theme.

2. Using the **diamonds** dataset,

a. Create a bar plot showing the count of diamonds for each cut category. Use the color variable to fill the bars, and adjust the position to dodge so that bars for each color are side by side. Use **scale_fill_brewer()** to apply a color palette from the **RColorBrewer** package. Apply the **theme_classic()** theme.

b. Convert the bar plot in (A) to a circular (polar coordinate) plot using **coord_polar()**. **Use scale_fill_viridis_d**() to apply a color scale from the **viridis** package. Apply the **theme_minimal()** theme.

c. Create a heatmap that shows the average price of diamonds for each combination of cut and color. Use **geom_tile()** to create the heatmap and map the average price to the fill aesthetic. Use **scale_fill_gradient()** to apply a custom color gradient (e.g., blue to red). Apply the **theme_light()** theme and adjust the legend position to the bottom of the plot.

3. Using the **economics** dataset, create a stacked area plot of **psavert** (personal savings rate) and **uempmed** (median duration of unemployment) over time (**date**). Use **geom_area()** to create the stacked areas and map the variables to the fill aesthetic. Use **scale_x_date()** to format the date axis and **scale_fill_brewer()** to apply a color palette. Apply the **theme_dark()** theme.

4. Try to create a custom theme. See here.

5. Integrate the above with the dashboard you have already worked on. Provide a screenshot.