Tutorial-I In-Class Activity

Dataset: penguins.csv

The Palmer Penguins dataset contains biological observations for three penguin species in Antarctica.

Key Columns include:

- **species**: Adelie, Chinstrap, or Gentoo.
- **body_mass_g**: Total body mass in grams (Numerical).
- **year**: Year of study (2007, 2008, 2009).

Tasks

1. Data Cleaning:
   Download the csv. Filter out any entries that contain missing (NA) values for **your** variables (Hint: use D3's .filter).
2. Bar Graph: Species Distribution:
   Create a bar chart showing the total count of penguins per species. Hint: Use d3.rollups to count the occurrences of each species.
3. Line Graph: Yearly Weight Trends:
   Create a line chart tracking the average body mass (g) of all penguins across the years 2007, 2008, and 2009. Hint: Use d3.rollups with d3.mean to find the average mass per year.

Answer the following questions regarding your visual design in a separate pdf:

- **Q1.** Why did you choose a **Bar Graph** for the counts of species but a **Line Graph** for the average mass over the years? What is the logical difference between the two?
- **Q2.** On your Line Graph, did you start the Y-axis at **0** or a value closer to the data? Explain how your choice affects the visibility of the "trend".
- **Q3.** Explain how the SVG coordinate system (where y=0 is the top) required you to change your height calculations for the bar chart.
- **Q4.** Explain why you used a Band Scale (scaleBand) for species names but a Linear Scale (scaleLinear) for body mass.

Submission format

Zip your

- Html, css, javascript, csv files
- Answer pdf
- Screenshots of both visualisations

into <rollnumber_tutorial1>.zip