Module 1D Questions:

For each of the following scenarios, choose the best visualization below:

- A. Scatterplot
- B. Cumulative frequency plot
- C. Bar plot
- D. Mosaic plot
- E. Boxplot

1. **The table below shows the frequency of a specific single nucleotide polymorphism (SNP) across four different populations:**

Population	Frequency
A	15
B	25
C	45
D	60

Which visualization would best represent the differences in SNP frequency across populations and why?

2. **The following data represents the total number of mutations detected in a genetic sample as a function of increasing sequencing depth:**

Sequencing Depth	Cumulative Mutations Detected
10X	15
20X	40
30X	70
40X	100
50X	120

Which visualization would best represent the accumulation of detected mutations as sequencing depth increases and why?

3. The table below shows the observed genotype frequencies of a SNP in two populations:

Population	Genotype AA	Genotype AB	Genotype BB
A	40%	35%	25%
B	30%	50%	20%

Which of the following visualizations would best represent the proportional relationship of genotypes within and between populations and why?

For the following 2 questions (Q4 and Q5), use the following dataset:

Mouse_ID	Strain	Gene_Expression (GFP)	Mutation_Type	Age (weeks)	Sex
M001	C57BL/6J	12.4	None	10	Male
M002	BALB/c	8.1	SNP	12	Female
M003	C57BL/6J	15.6	InDel	14	Male
M004	DBA/2J	7.3	None	9	Female
M005	BALB/c	9.9	SNP	11	Male
M006	C57BL/6J	14.7	InDel	13	Female
M007	DBA/2J	6.8	None	8	Male

4. You want to compare the **distribution of gene expression levels** across different **mouse strains**. What is the most appropriate plot and why?
5. You want to visualize the **relationship between Mutation_Type and Sex** across your mice. What plot is most appropriate and why?