

NYRaMP-Informatics Workshop, Pre-Test (08/06/2024)

Instructions:

- This is a survey designed to improve your informatics learning
- Keep it for future reference. We will not collect or grade the sheet.

Name: _____

1. Match the genomics terms at left with the description at right:

- | | | |
|---------------------------|-----|---|
| Sanger sequencing | ___ | (A) All proteins in a cell |
| next-gen sequencing (NGS) | ___ | (B) raw DNA sequence outputs from NGS |
| sequencing reads | ___ | (C) all DNA in a cell |
| genome assembly | ___ | (E) RNA-seq |
| genome | ___ | (G) all microbes in the gut |
| microbiome | ___ | (H) whole genome sequences derived from reads |
| proteome | ___ | (I) sequencing by synthesis (e.g., Illumina & PacBio) |
| transcriptome | ___ | (J) chain termination with ddNTPs |

2. To visualize the distribution of a numerical variable, you would use:

- (A) A boxplot
- (B) A histogram
- (C) A bar plot
- (D) A scatterplot

3. To visualize the relation between a numerical variable and a categorical variable, you use:

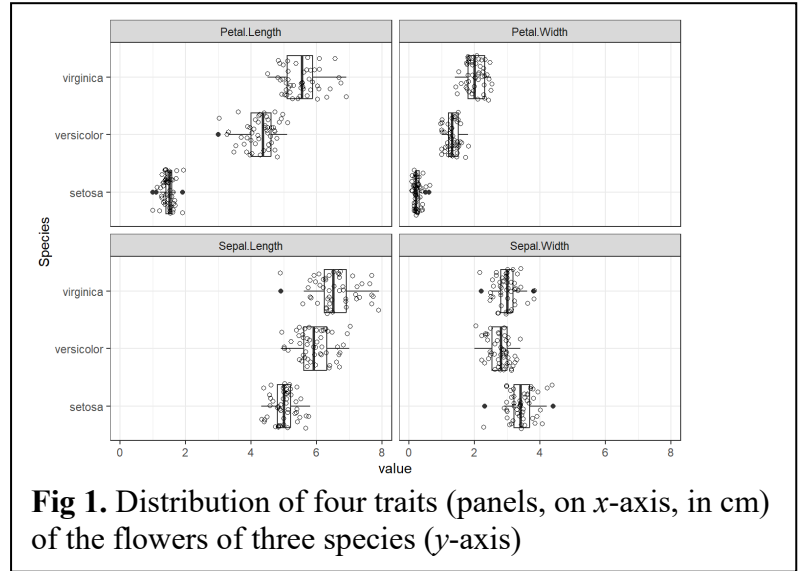
- (A) A boxplot
- (B) A histogram
- (C) A bar plot
- (D) A scatterplot

4. To visualize the relation between two numerical variables, you would use:

- (A) A boxplot
- (B) A histogram
- (C) A bar plot
- (D) A scatterplot

5. Fig 1 is a boxplot. Explain:

- (A) What does the box represent?
- (B) Identify the numerical and categorical variables
- (C) Which trait differs the most and the least between the species, respectively?
- (D) How would you determine the statistical significance of a trait value between two species? Between three species?



6. Fig 2 is a volcano plot. Explain:

- (A) Fold change (FC, x -axis)
- (B) P -value (in \log_{10} scale, y -axis)
- (C) Each point represents a gene. Which genes are significantly over-expressed in this tissue?

