Comp Genomics Workshop-KIZ, Student Survey (10/10/2024)

Instructions:

- This is a survey designed to maximize your learning experience
- It will NOT be graded

Name: _____

1) Please indicate your current level of study

    (A) Undergraduate; (B) Master's; (C) Doctoral; (D) Post-doctoral; (E) Researcher

2) Please describe your experience in the following fields of study

Field of study/research	Years of experience (0 for none)	Description/Explanation (optional)
Molecular Biology		
Genomics		
Transcriptome		
Microbiome		
Population genetics		
Phylogenetics		
Linux commands		
Python or other language		
R programming		
Statistics		

3) 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

4) To visualize the distribution of a numerical variable, you would use:
   a) A boxplot
   b) A histogram
   c) A bar plot
   d) A scatterplot

5) 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

6) To visualize the relation between two numerical variables, you would use:
   a) A boxplot
   b) A histogram
   c) A bar plot
   d) A scatterplot
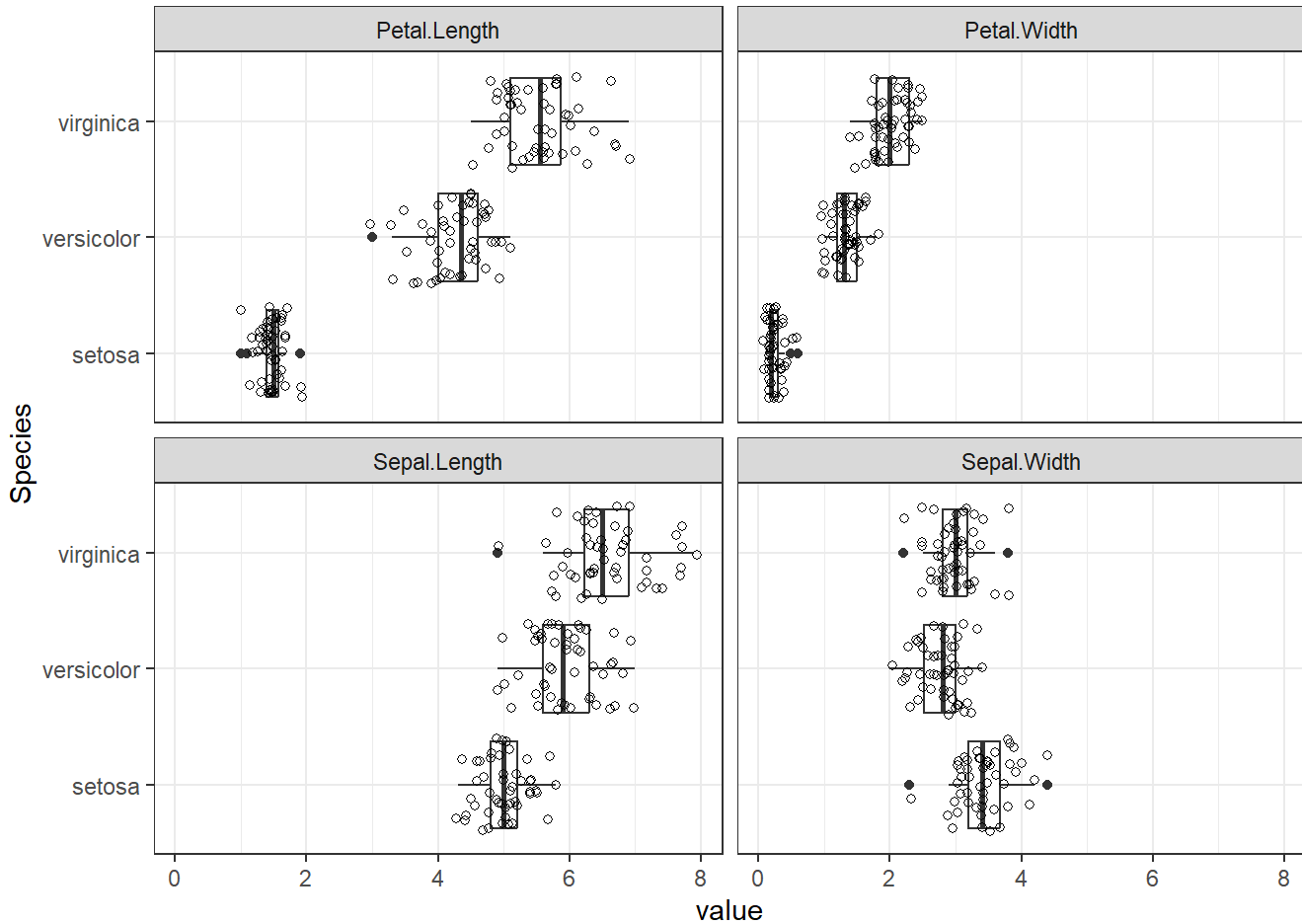
7) With respect to Fig 1, explain:

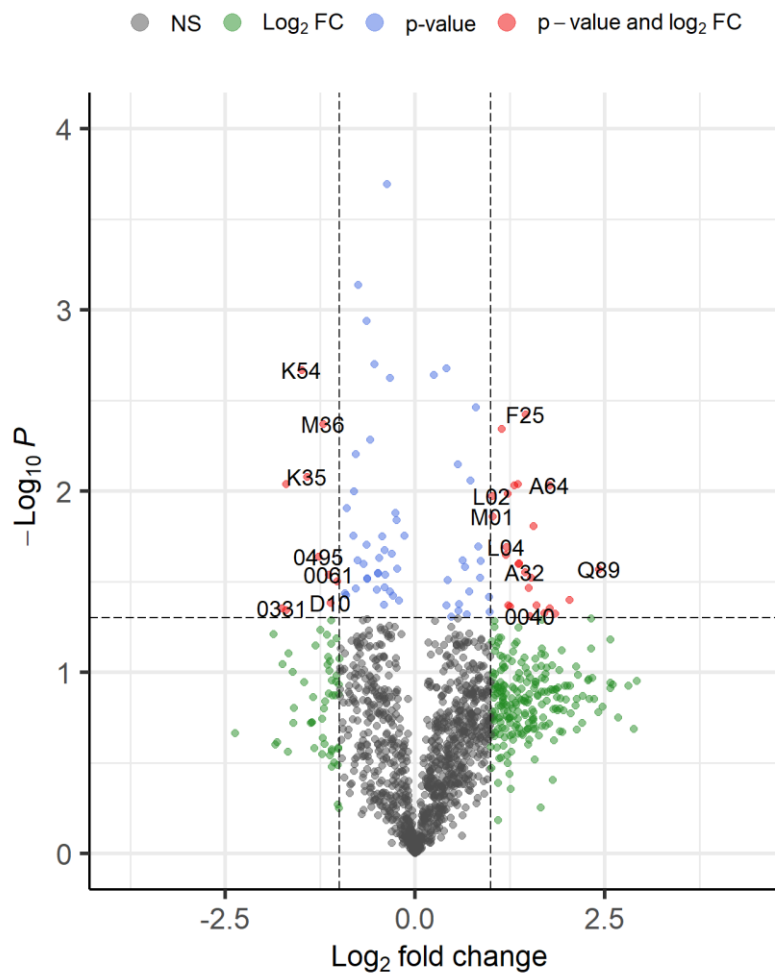   a) What does the box represent?

   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?

Differentially expressed genes during Borrelia burgc

Early vs Late growth

8) Fig 2 is a volcano plot. Explain:

  a) Fold change (FC, *x*-axis)

  b) *P*-value (in log10 scale, *y*-axis)

  c) Each point represents a gene. Which genes are significantly over-expressed?