Name: _____ LA Initials: [ ]

BIOL 275 Lab 6 Describing data

1. List the numerical variable and the categorical grouping variables used in the worked example (`and_vertebrates`).

2. Why is it important to check for and remove missing values before computing summary statistics such as the mean or standard deviation?

3. In your own words, how does a summary table differ from the raw data table?

4. In the `summarize()` call, what does the argument `.by = c(section, species)` do?

5. In this dataset, is the standard error (SE) larger or smaller than the standard deviation (SD)? Why?

6. Name one feature of the data that is visible in the distribution plot but is not visible in the summary table.

7. Why do some `geom_*()` layers use the raw dataset, while others use the summary table?

8. What does it mean to "override" an aesthetic inside a `geom_*()` call?

9. Based on the distribution plot, does the mean or the median better represent a "typical" individual for this dataset? Briefly explain your reasoning.

10. Why is the mean ± confidence interval plot more effective than the distribution plot for comparing group means?