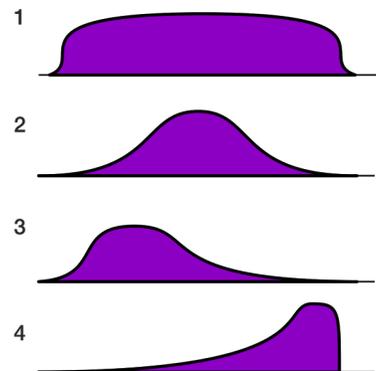
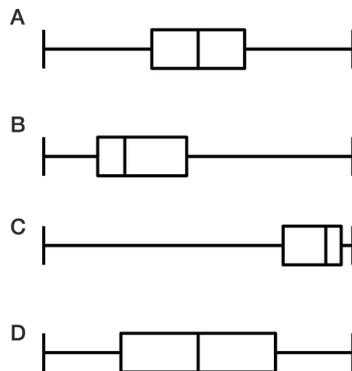


## Homework One

BIO5312

Due 9/5/17 by 5:30 pm

- 1) Identify whether the following variables are numeric or categorical. If numeric, state whether the variable is discrete or continuous. If categorical, state whether the variable is nominal or ordinal (4 pts):
  - a) Time of the fastest runner in the Broad Street Run.
  - b) Milk types, denoted by non-fat, 1%, 2%, or whole milk.
  - c) Circumference of potholes in North Philly.
  - d) Number of red flowers in a garden.
  
- 2) Match the following boxplots to the distribution each represents (4 pts):



- 3) Use R to complete the following questions. *Show **all** of your code in a separate text file with this assignment.* These questions use a dataset called **sparrows.csv** (download here: \_\_\_\_\_). Briefly, in 1898, Hermon Bumpus, an American biologist working at Brown University, collected data on one of the first examples of natural selection directly observed in nature. Immediately following a bad winter storm, he collected 136 English house sparrows, *Passer domesticus*, and brought them indoors. Of these birds, 64 had died during the storm, but 72 recovered and survived. By comparing measurements of physical traits, Bumpus claimed to detect substantial physical differences between the dead and living birds.
  - a) Download the dataset **sparrows.csv**, and read the data frame into R. Explore the data frame using functions such as `head()`, etc. What type/subtype (i.e. numeric/discrete) of variable is each column? (2 pts)

- b) Choose one numeric (not weight or wingspread, please) and one categorical variable each from this dataset. Perform the following tasks. Please limit your text responses to *no more than 2 sentences*. In addition, please ensure that your plots have appropriate axis labels (no dollar signs!).
- i) Make a boxplot and histogram of your chosen numeric variable. Given this visual information, which measure (mean or median) would you choose to represent the location of this distribution and why? (6 pts)
  - ii) Make a barplot of your chosen categorical variable. Briefly describe how the resulting distributions compare. (4 pts)
- c) Use logical indexing and the functions `min()` and `max()` to determine the sex, age, and survival status for the bird with the longest and shortest wingspread. (5 pts)
- d) Compare weight distributions, using both summary statistics and histograms, between birds which survived the storm and birds which died. Based on this descriptive information, do you think weight influenced sparrow survival? Why or why not? (5 pts)