

# Module 2 Lab: univariate analysis

*BIOSTAT 213: Introduction to Computing in the R Software Environment*

*Wednesday, January 23, 2019*

Save your code in a script file. I suggest adding labels and notes to this script file as comments. You may work in groups of 2–3.

1. Download the `unps.csv` data file from the [Collaborative Learning Environment](#). The file, courtesy of Adrienne Epstein (thank you Adrienne!), contains data from children in Uganda. A data dictionary is available on the Collaborative Learning Environment.
2. Import the data into R.
3. Explore the data frame using RStudio's built-in data viewer.
4. Explore the data frame using `names()`, `head()`, `tail()`, `str()`, and `dim()`. Notice, from `head()` and `tail()`, that R is treating the ID variables `hhid` and `pid` as numeric objects, and hiding the full digits using `e+` notation.
5. Use `class()` to verify that `hhid` and `pid` are numeric objects. Convert the variables to character objects. Use `class()` to verify that the variables have been properly converted.
6. How many unique households are there in the data? In other words, how many unique household IDs `hhid` are there in the data?
7. What is the smallest number of characters in `hhid`? What is the largest number of characters in `hhid`?
8. How many waves (`wave`) of data are there?
9. Produce a histogram for age, in months (`age_months`).
10. Add a variable to the data frame for age in years (note: there are 12 months in a year).
11. Compute the percent of children who are male (`male`). Compute the percent of children who are underweight (`underweight`).