

# Homework 1: Intro to R for Epidemiologists

## Instructions

- Due at 11:59 PM on Wednesday, February 4, 2015
- Late assignments will not be accepted
- Covers topics in weeks 1-3
- Submit this assignment as a .R file by e-mail to [jenna.krall@emory.edu](mailto:jenna.krall@emory.edu)
- **Before submitting the assignment, comment out using ## any setwd() or specific file paths to your computer**

You may work with your classmates, but you must type and turn in your own assignment. Do not copy and paste someone else's work. Your code comments should be your own.

## Grading

The assignment is worth 15% of your final grade and will be graded out of 30 total points

- 5 points: Code formatting including proper spacing, indentation, comments, and names
- 8 points: Part 1, Data exploration and summary
- 17 points: Part 2, Plotting

## Data

This assignment uses a dataset containing Forced Expiratory Volume (FEV), a measure of lung function, and covariates.

## References

- FEV: <https://meded.ucsd.edu/isp/1998/asthma/html/spirexp.html>
- Smoking and FEV: <http://www.ncbi.nlm.nih.gov/pubmed/20225780>
- Original dataset from Rosner, Bernard. *Fundamentals of Biostatistics*. Cengage Learning, 2010.
- Data accessed January 5, 2015, <http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets>.

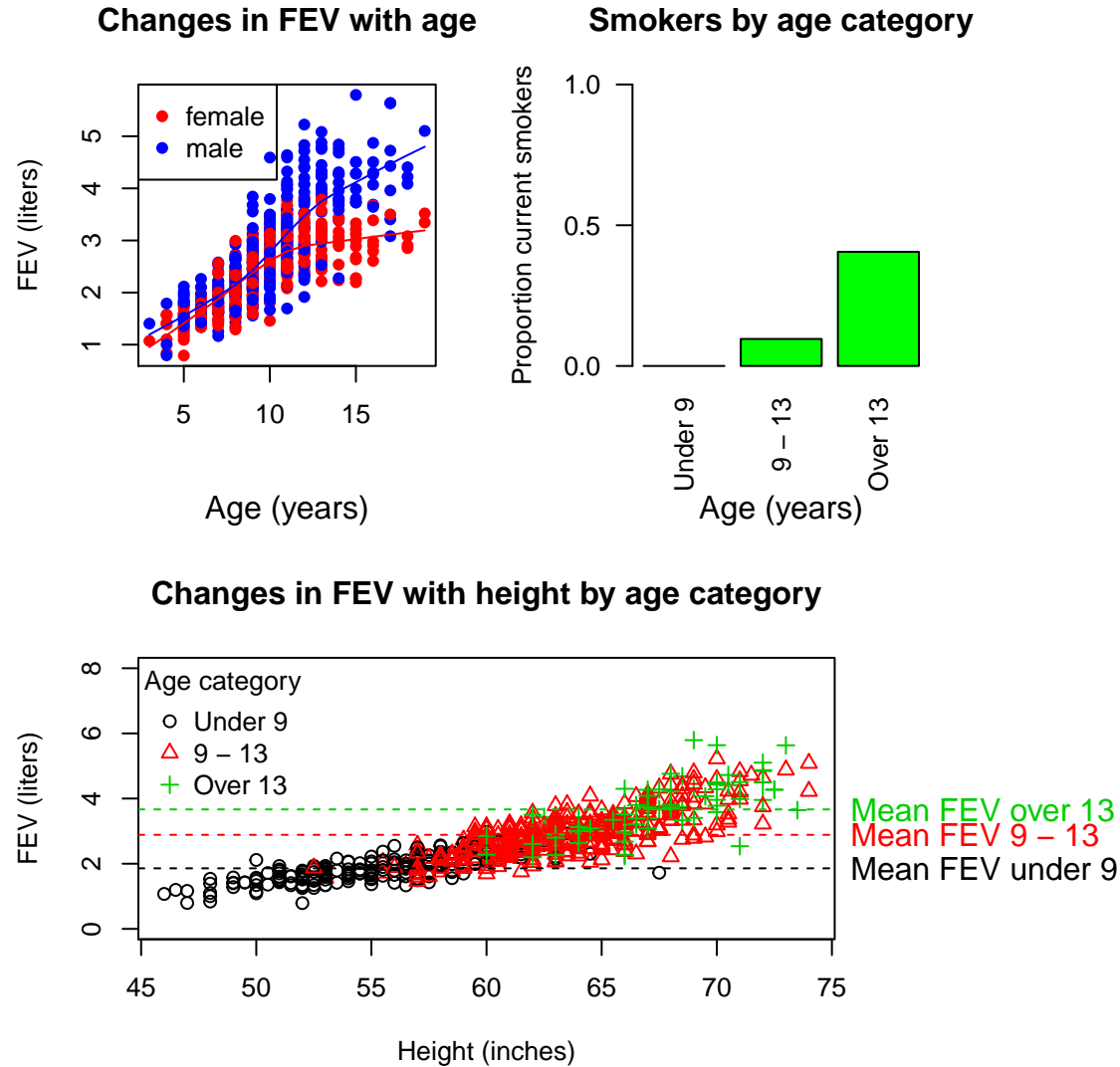
## 1 Data exploration and summary

- a. Read the dataset "FEV.csv" into R.
- b. Create an R vector called `info_fev` with two elements: the number of observations and the number of variables.
- c. Create an R data.frame called `summ_df` with rows corresponding to age, FEV, and height, and columns indicating the number of observations (as an integer), minimum, maximum, mean, standard deviation (all rounded to 2 decimal places), and the number of missing observations (integer) for each variable. The integer variables should be integers in R such that `class(integervariables)` is "integer"
- d. Create a table in R called `tab_sex_smoke` of the percentage of smokers and nonsmokers by sex, rounded to one decimal point. Recall that you can create a table using the R function `table()`
- e. Create a matrix in R called `mean_fev` of the mean FEV by age group (under 9, 9-13, over 13) and by smoking status, rounded to two decimal points.
- f. Create a data frame in R called `fev_smoke` of only female current smokers with FEV less than 2.5, ordered by increasing age then height.

## 2 Plotting

For this question, you need to recreate the following figure. You only need to provide the code to produce the figure.

Hint: You may want to search the R help files for `layout`, `xaxt`, `axis`, `cut`, and `legend`



## 3 Bonus (2 points)

In plot 3, instead of using horizontal lines to represent the mean FEV for each age category, use segments that span the minimum and maximum height for each age category.