Lab 0 - Kenya, Due: (TODO)\_\_\_\_

YOUR NAME HERE

2021-08-19

# Lab 0

## Task 1: Establish a workflow

Open RStudio. Click File > New Project. Hit “Existing Directory” and then “Browse”. By now, you should have created a folder for this course’s lab, and a subsequent folder within for Lab 0. Open the Lab 0 folder and hit “Open”. Then hit “Create Project”.

## Task 2: Familiarize yourself with R Markdown

## Task 3: Install packages and load libraries

1. Note that for this step, you should first install three packages using the following codes; If you have already installed these packages once before, proceed to 3b directly.

install.packages("tidyverse") install.packages("skimr") install.packages("epiR") install.packages("devtools") devtools::install\_github("potato-nathan/epiAssist")

1. Load libraries:

For this lab, we need {skimr} and {tidyverse}

In a fresh code chunk, call in the {tidyverse} and {skimr} packages using the following codes:

library(tidyverse) library(skimr)

# load library {tidyverse}  
  
  
# now load library {skimr}

## Task 4: Load data

Use command readRDS(‘filename.rds’) to load in the dataset. Don’t forget to use the assignment operator to give the dataset a name after importing the data. As an example, you can name it like this:

kenya <- readRDS('data/filename.rds')

# load dataset and give it a name within the r environment

## Task 5: Explore the dataframe

Familiarize yourself with the data by using the commands head(), skim(), ncol(), and nrow().

* Are there any string/character variables?
* Are there any variable or value labels?
* Do any variables have notes?

# find number of columns with ncol(dataName)  
  
# find number of rows with nrow(dataName)

# inspect variable names with names(dataName)  
  
# inspect 3 variable's types with class(dataName$variableName)

# generate summary statistics for your dataset with skim(dataName)

## Task 6: Create variable mage

Create a variable mage for mother’s age (as an integer) at the time of each child’s birth (note – some of these mothers have had multiple children) using the mutate() function. This is calculated from variables b3 (month code of child’s birth) and v011 (month code of mother’s birth). The difference between the values of these variables is months, so divide by 12 to get years. See the data dictionary for a more detailed description of month codes and how to use them. Use the as.integer() function within your mutate() function to truncate the calculated value to an integer.

REMINDER: For your mutate function to work, you have to merge the changes into your dataset. As an example, use this before the mutate function: kenya <- kenya %>%

# create variable mage with a pipe (%>%) and mutate()

## Task 7: Frequency distributions of mage

Suppose you want to break down mage into three categories. First, look at the frequency distribution (one-way frequency table) for mage to see where you might draw lines for your categories using the table() command.

* Are there any missing values for age? If so, how many?
* Which range of ages appear the most frequently in mage?

# generate a table of mage's frequency distribution with table()  
# don't forget to include NA values with the argument useNA = "always"

## Task 8: Create variable magec

Using mage, generate a new variable with three categories: <18, 18-39, and ≥ 40, naming the new variable “magec” (mother’s age categorical). Set the values for magec to be 0,1,2, where 0 corresponds to the youngest age group (<18). Recommended steps:

# create magec with a pipe (%>%) and mutate()

# factor magec (convert it to a categorical variable) with factor()

## Task 9: Cross-tab of mage and magec

Look at a cross-tabulation (two-way table) of mage and magec to ensure that magec was created correctly. Be sure missing values were handled properly (all observations that have a missing value for mage should be assigned the R missing value “NA” for magec.). Type “?table” for help with how to create a 2x2 table. Note: the order of the variables in the command controls which one is in the rows and which is in the columns of the table. Experiment to make your table readable.

# create a cross-tabulation of your new variables with table()

# generate cross-tabulated numbers with a pipe, group\_by(), and count()

## Task 10: Save new dataset

Save the new dataset in the same directory using a different file name with the form shown in the instructions on page 1 using the saveRDS() command.

# save the datset with saveRDS()