Lab 1 Homework: Introduction to Data

Background

The Centers for Disease Control (CDC) collects and monitors data on many health-related issues. The ratio of male to female births can help understand population growth trends and trends in infant mortality. Read https://www.cdc.gov/nchs/data/nvsr/nvsr53/nvsr53_20.pdf for more information about an analysis of sex ratios at birth in the United States.

You recreated some of the displays and preliminary analysis of Arbuthnot's baptism data in the lab manual. Your assignment involves repeating these steps, but for present-day birth records in the United States.

Homework

1. The data are stored in a .csv file called present.csv. It can be found on Canvas. Load the present data into R.
2. What variable (column) names are in this data set?
3. What are the dimensions of the data frame? (How many observations and variables are there?)
4. What years are included in this data set?
5. In the lab1 manual, you observed from the `arbuthnot` data set that boys were born in greater proportion than girls in London from 1629-1710. Now examine the `present` data set - does the relationship hold in the US from 1940-2002? Why or why not?
6. Make a plot that displays the proportion of boys over time. Do you see any trends?
7. Give the plot a title of "Proportion of boys over time ."Which argument did you use to create the title?
8. In how many years did the proportion of boys exceed 0.512?
9. In which year did we see the largest total number of births in the US? (Hint: There are many ways to identify this. To find helpful commands, you can refer to the help files or the R reference card (http://cran.r-project.org/doc/contrib/Short-refcard.pdf).)
10. Save your work as a well-organized and commented R Script. Be prepared to answer quiz questions on this homework.