

Lab 1: Introduction to R

Learning Objectives

Students will be able to:

1. Demonstrate basic R skills
 2. Use basic statistical functions
 3. Visualize single numerical and categorical variables
-

Readings

To prepare for this lab assignment read the following tutorials

1. Fundamentals of R
2. Importing Data in R
3. Summarizing Numerical Data
4. Summarizing Categorical Data

! Deadline: 10/18 @ 11:59pm

Submit a single PDF for your lab report using Quarto

Section 1: Basic R Skills

Vectors

1. Create a vector named `heights` that contains the heights, in inches, of yourself and two students near you. Print the contents of this vector.
2. Create a vector named `names` that contains the names of these people. Print the contents of this vector
3. Try typing `cbind(heights, names)`. What did this command do ? What class is this new object ?

Hint: Try the `class()` function

Downloading Data

4. Download the data set `births.csv` and upload it into R Studio. Store the data frame as `births` variable name
5. Demonstrate that you successfully uploaded the data by printing out the output of `head(births)`
6. How many observations and variables does the data set `births` contain ?

Loading R packages

7. install the `openintro` package (*console only, not in Quarto R chunk*). Verify the installation by printing the output of `find.package("openintro")`
8. Type the command `library(openintro)` to load up the package.
9. install the `tinyploth` package (*console only, not in Quarto R chunk*). Verify the installation by printing the output of `find.package("tinyploth")`
10. Type the command `library(tinyploth)` to load up the package.

Vector Operations

11. Extract the `weight` variable as a vector from the data frame `births`
12. What units do you think the weights are in ?
13. Create a new vector names `weight_in_grams` which are the weights of the babies in grams. You can look up conversion factors online
14. Run the command `weight_in_grams[1:20]` and print out the output of the first 20 weights.

Section 2: Summarize Data (one variable)

1. What is the mean weight of the babies in pounds ?
2. What percentage of the mothers in the sample smoke ?
3. According to the Centers for Disease Control, approximately 9% of adult Americans are smokers. How far off is the percentage you found in (2.) from the CDC's report?

Section 3: Visualizing Data (one variable)

1. Produce a dot plot of the weights in pounds
2. Produce three different histograms of the weights in pounds. Use 3 bins, 20 bins, and 100 bins. Which histogram seems to give the best visualization, and why?

For `barplot()`, (*and other plots()*) we can add several arguments to make the plot more presentable.

- The `col` argument changes the color of the bars.
- `xlab` and `ylab` define the labels for the x- and y-axes.
- `main` gives the plot a title.
- `ylim` can be used to control the range of values displayed on the y-axis.

See `?barplot()` for additional help and customizations.

3. Produce a nicely formatted bar plot that displays the number of mothers in each smoking category
4. We can use the syntax `boxplot(vector1, vector2)` to make a side by side box plot. Create a side by side boxplot of the mother's ages and the father's ages. Which gender tends to be older?

5. Run the command `tinyplo( ~ weight | habit, data = births, type = "hist")`. Describe what this code does. Based on the graph, do you see any major differences between baby weights from smoking moms vs. non-smoking moms?