**Memorandum**

To: NTRHD Intern

From: Dr. Brad Cannell, Director, NTRHD

Re: Data Transfer

**Overview**

The NTRHD is part of a study about the association between chocolate eating habits and academic ability. The study took place at multiple schools, chosen at random, from around the region. Graduate students administered the surveys to students at each school, recorded survey responses electronically, and then emailed them to us. Now, we need your help importing the data into R, doing a little bit of data cleaning, and saving the data again once it’s ready for analysis.

Additionally, we need you to download, clean, and save some NHANES data on alcohol use.

**Task 1.** Click the links below to download the raw data files to your computer:

* [**chocolate study 1.xls**](https://www.dropbox.com/s/9riqfhtf582o2az/Chocolate%20Study%201.xls?dl=1)
* [**chocolate study 2.csv**](https://www.dropbox.com/s/ggln051redw1g98/Chocolate%20Study%202.csv?dl=1)
* [**chocolate study 3.txt**](https://www.dropbox.com/s/90ndmdc463ui88j/Chocolate%20Study%203.txt?dl=1)

**Task 2.** Import **Chocolate Study 1.xls** into your R global environment as **choco\_1**.

|  |
| --- |
| **Question:**  1. Which of the following code chunks will correctly import the **Chocolate Study 1.xls** data into your R global environment as **choco\_1**? |

**Task 3.** View the structure of **choco\_1** using the str() function.

|  |
| --- |
| **Question:**  2. The **choco\_1** data frame contains \_\_\_\_ rows and \_\_\_\_ columns. |

**Task 4.** Import **Chocolate Study 2.csv** into your R global environment as **choco\_2**.

|  |
| --- |
| **Question:**  3. Which of the following code chunks will correctly import the **Chocolate Study 2.csv** data into your R global environment? |

**Task 5.** View the structure of **choco\_2** using the str() function.

|  |
| --- |
| **Question:**  4. The **choco\_2** data frame contains \_\_\_\_ rows and \_\_\_\_ columns. |

**Task 6.** Import **Chocolate Study 3.txt** into your R global environment as **choco\_3**. Make sure to tell R how to correctly identify missing values in the raw data.

|  |
| --- |
| **Question:**  5. Which of the following code chunks will correctly import the **Chocolate Study 3.txt** data into your R global environment? |

**Task 7.** View the structure of **choco\_3** using the str() function.

|  |
| --- |
| **Question:**  6. The **choco\_3** data frame contains \_\_\_\_ rows and \_\_\_\_ columns. |

**Task 8.** Please view the National Health and Nutrition Examination Survey (NHANES) questionnaires, datasets, and related documentation website here: <https://wwwn.cdc.gov/nchs/nhanes/>

**Task 9.** Import the 2015-2016 NHANES Alcohol Use data into R as a data frame named **alq\_i**. Alcohol use was part of the NHANES questionnaire data.

|  |
| --- |
| **Question:**  7. What R package contains the read\_xpt() function we need to use to import the NHANES data in Task 9? |

**Task 10.** Coerce all of the variable names in **alq\_i** to lowercase.

|  |
| --- |
| **Question:**  8. Which of the following code chunks will correctly coerce all of the variable names in **alq\_i** to lowercase? |

**Task 11.** Coerce all of the categorical variables in **alq\_i** to a factor variables. Use the value descriptions from the [2015-2016 NHANES Alcohol Use data codebook](https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/ALQ_I.htm) to make factor levels and labels. Name the factor version of each variable the same name as the original version, but add a **\_f** to the end.

**Task 12.** View the structure of **alq\_i** using the str() function.

|  |
| --- |
| **Questions**  9. The **alq\_i** data frame contains \_\_\_\_ rows and \_\_\_\_ columns.  10. When you view the structure of **alq\_i** using the str() function, what is the vector type shown for **alq101\_f**? |

**Optional**: Please feel free to leave any comments below about the usefulness of this lab. Which parts were helpful? What could I do to improve it? What is still unclear?