# LAB 1F: A Diamond in the Rough

Directions: Record your responses to the lab questions in the spaces provided.
Messy data? Get used to it
Messy data?
The American Time Use Survey
Load and go:  • Just by viewing the data, what parts of our ATU data do you think need cleaning?
Description of ATU Variables
New name, same old data  • Use the example code and the variable information on the previous slide to rename the rest of the variables in atu_dirty.
Next up: Strings
<ul> <li>Numbers are words? (Sometimes)</li> <li>Write down the variables that should be numeric but are improperly coded as strings or characters.</li> </ul>

## Changing strings into numbers

### Mutating in action

• Once you have this code working, use a similar line of code to correctly code the other numeric variables as numbers.

## **Deciphering Categorical Variables**

#### **Factors and Levels**

• Use similar code as we used above to write down the levels for the three factors in our data.

#### A level by any other name...

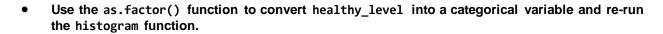
## Allow me to explain

### Finish it off!

- Recode the categorical variable about whether the person surveyed had a physical challenge or not. The coding is currently:
  - "01": Person surveyed *did not* have a physical challenge.
  - "02": Person surveyed did have a physical challenge.
- Write a script that:
  - Loads the atu\_dirty data set
  - Cleans the data as we have in this lab
  - Saves a copy of the cleaned data

## The final lines

# Flex your skills



Recode the healthy\_level categories and re-run the histogram function.