

## 5.3: Tidyverse in Action

To see the tidyverse in action, let's clean up the NHANES dataset. Each individual in the NHANES dataset has a unique identifier stored in the variable `ID`. First let's look at the number of rows in the dataset:

```
nrow(NHANES)
```

```
## [1] 6779
```

Now let's see how many unique IDs there are. The `unique()` function returns a vector containing all of the unique values for a particular variable, and the `length()` function returns the length of the resulting vector.

```
length(unique(NHANES$ID))
```

```
## [1] 6779
```

This shows us that while there are 10,000 observations in the data frame, there are only 6779 unique IDs. This means that if we were to use the entire dataset, we would be reusing data from some individuals, which could give us incorrect results. For this reason, we would like to discard any observations that are duplicated.

Let's create a new variable called `NHANES_unique` that will contain only the distinct observations, with no individuals appearing more than once. The `dplyr` library provides a function called `distinct()` that will do this for us. You may notice that we didn't explicitly load the `dplyr` library above; however, if you look at the messages that appeared when we loaded the `tidyverse` library, you will see that it loaded `dplyr` for us. To create the new data frame with unique observations, we will pipe the NHANES data frame into the `distinct()` function and then save the output to our new variable.

```
NHANES_unique <-  
  NHANES %>%  
  distinct(ID, .keep_all = TRUE)
```

If we number of rows in the new data frame, it should be the same as the number of unique IDs (6779):

```
nrow(NHANES_unique)
```

```
## [1] 6779
```

In the next example you will see the power of pipes come to life, when we start tying together multiple functions into a single operation (or "pipeline").

This page titled [5.3: Tidyverse in Action](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Russell A. Poldrack](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.