

## STAT 102 - Removing Missing Values

### Functions

`na.omit()` - Removes all **rows/observations** with a single missing value for any variable (**most aggressive** way to deal with missing values)

`drop_na()` - Removes **rows/observations** with missing values for specific variable(s) (**moderately aggressive** way to deal with missing values)

`na.rm = TRUE` - Only temporarily ignores N/A as needed before calculating, without removing any **rows/observations** (**least aggressive** way to deal with missing values)

### Example

**colleges has 1285 observations and 27 variables**

```
load("data/colleges.Rdata")
```

**colleges\_aggressive\_removal has 279 observations and 27 variables; I removed all rows with even a single value of N/A (for any variable)**

```
colleges_aggressive_removal <- colleges %>%  
  na.omit()
```

**colleges\_moderate\_removal has 1284 observations and 27 variables; I removed all rows with a value of N/A only for the variable sticker\_price\_2013**

```
colleges_moderate_removal <- colleges %>%  
  drop_na(sticker_price_2013)
```

**colleges\_light\_removal has 1285 observations and 27 variables (the same as the original colleges dataset); no rows were removed, but rather the rows with a value of N/A were temporarily ignored during the calculation**

```
colleges_light_removal <- colleges %>%  
  mutate(mean_sticker_price_2013 = mean(sticker_price_2013, na.rm = TRUE))
```