Lecture 4: Contingency Tables
STAT 310, Spring 2021

A **contingency table** summarizes data for two categorical variables. Each value in the table represents the number of times a particular combination of variable outcomes occurred. For example, here is a contingency table between the variables `PhysActive` and `HealthGen`:

```
nhanes <- readRDS(url("https://ericwfox.github.io/data/nhanes.rds"))
table(nhanes$PhysActive, nhanes$HealthGen)

##
##       Excellent Vgood Good Fair Poor
##   No         48   169  279  150   31
##   Yes       124   301  331   63    4
```

We can use the `addmargins()` function to add the row and column totals:

```
addmargins(table(nhanes$PhysActive, nhanes$HealthGen))

##
##       Excellent Vgood Good Fair Poor  Sum
##   No         48   169  279  150   31  677
##   Yes       124   301  331   63    4  823
##   Sum       172   470  610  213   35 1500
```

In-Class Exercise:

(a) What proportion of participants reported being in excellent health?

(b) What proportion of participants reported being physically active?

(c) What proportion of participants are both physically active and reported being excellent health?

(d) Of the participants who reported being in excellent health, what proportion are physically active?

(e) Of the participants who reported being in poor health, what proportion are physically active?

Contingency tables can be visualized using **stacked** or **side-by-side bar plots**.