Activity 2: Summarizing and Displaying Categorical Data STAT 310, Spring 2023

Below is a contingency table between the variables Smoke100 and HealthGen, from the NHANES data set (lab 2). A contingency table of column proportions is also provided.

```
> nhanes = readRDS(url("https://ericwfox.github.io/data/nhanes.rds"))
```

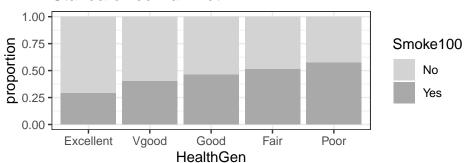
> addmargins(table(nhanes\$Smoke100, nhanes\$HealthGen))

	Excellent	Vgood	Good	Fair	Poor	Sum
No	122	282	328	104	15	851
Yes	50	188	282	109	20	649
Sum	172	470	610	213	35	1500

> prop.table(table(nhanes\$Smoke100, nhanes\$HealthGen), margin = 2)

```
Excellent Vgood Good Fair Poor
No 0.709 0.600 0.538 0.488 0.429
Yes 0.291 0.400 0.462 0.512 0.571
```

Standardized Bar Plot



- (a) What proportion of respondents have smoked at least 100 cigarettes in their lifetime?
- (b) What proportion of respondents have smoked at least 100 cigarettes and are in good health?
- (c) What does 0.291 represent in the contingency table of column proportions?
- (d) What does 0.571 represent in the contingency table of column proportions?
- (e) Do the variables Smoke100 and HeathGen appear to be associated?