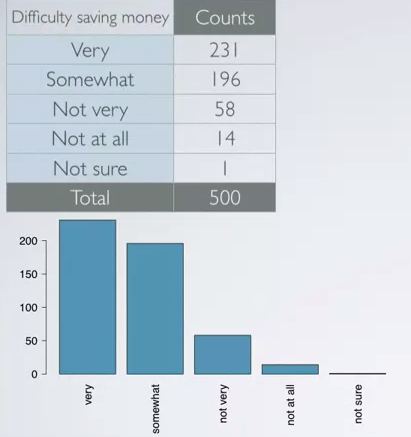
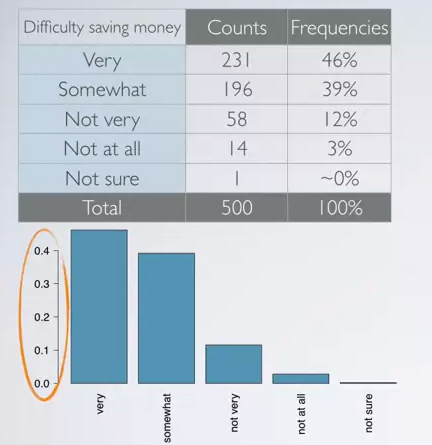
Stats Video Lectures – Exploring Categorical Variables

Week 1, Video 11

1. Describing the distribution of a single categorical variable
   1. Frequency and Bar Plots are useful  
      counts frequencies

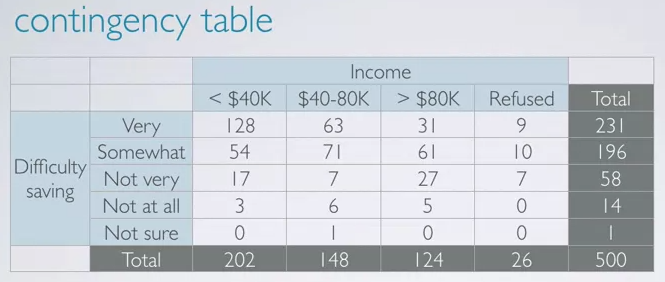
 or 

* 1. Differences between bar plots and histograms
     1. barplots are for categorical data, histograms are for numerical data

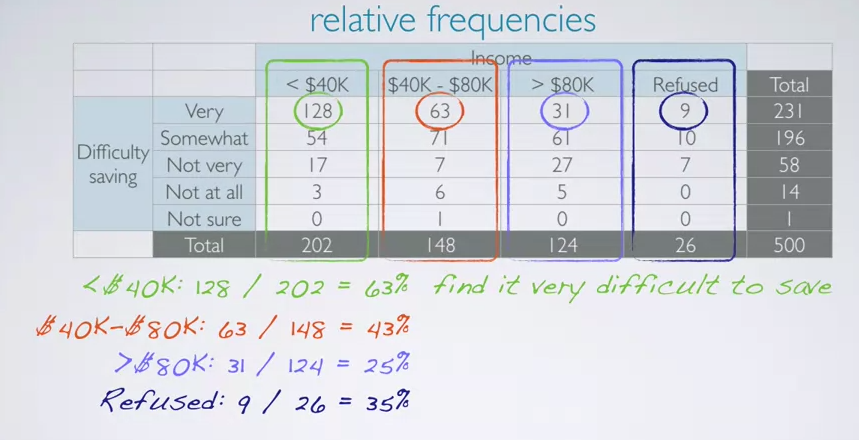


* + 1. x axis on a histogram are numbers (thus not moveable)
    2. x-axis on a bar plot are categories (thus the categories can be rearranged and played with)
  1. Pie Charts
     1. don’t do it

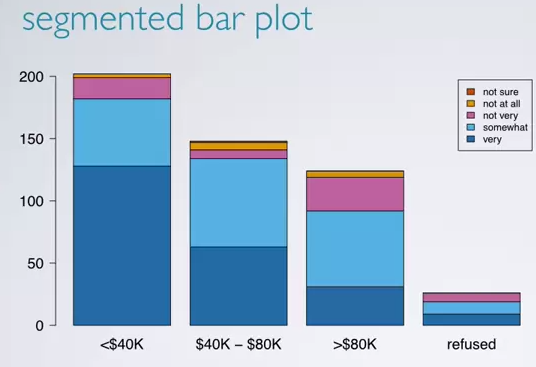
1. evaluating the relationship between two categorical variables
   1. Contingency Table



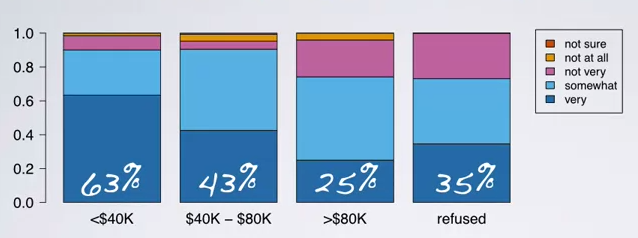
* + 1. relative frequencies are highly useful when the different subcategories are of different sizes



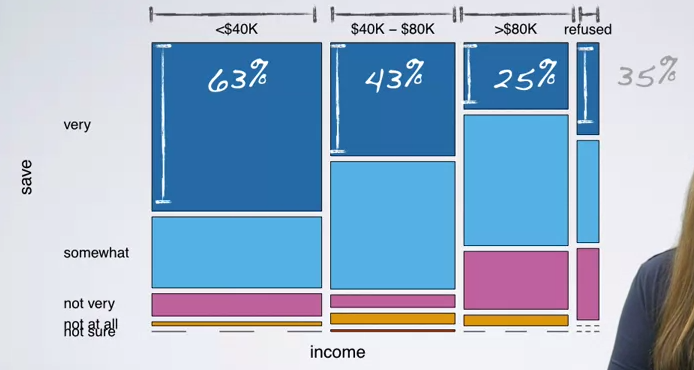
* + 1. this suggests that feelings about difficulty saving money depends on income, so the two categories are dependent on each other.
  1. Segmented bar plot
     1. good for conditional frequency distributions



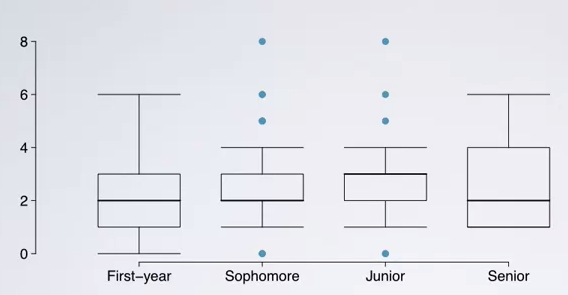
* + 1. not useful for relative frequencies.
  1. Relative Frequency Bar plots
     1. good for showing the frequency difference among the different categories



* 1. Mosiac Plot
     1. shows conditional and marginal distributions
     2. width of the bar shows the marginal distribution



1. evaluate the relationship between a categorical and a numerical variable
   1. Side by Side Box Plots



* + 1. this example shows the number of clubs studies are members of per year in college