

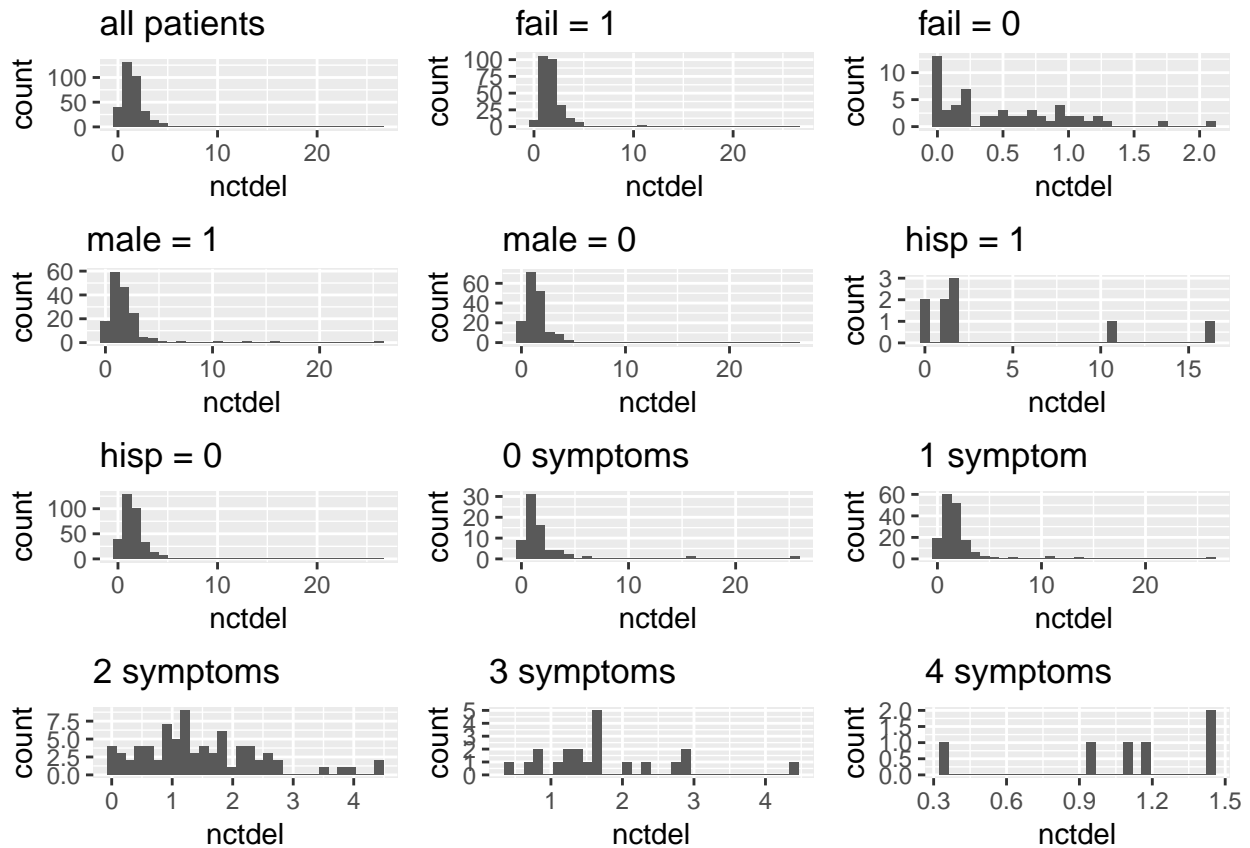
STA 440 Case 2

Annie Lott, Wuming Zhang, William Yang

October 1, 2017

Exploratory Data Analysis

Histograms of each Indicator



To get a better sense of the distributions in the data, we first plot histograms of the nctdel times for each level of all the variables. We notice some general trends here - nctdel time seems to go down with more symptoms, the overall distribution of times for all patients is right-skewed with a mean around 1.5. However, we note that there are only 9 data points for hispanic patients, and only 6 for patients with all 4 symptoms. Inference around these variables will have high variance.