

# Math 70 Homework 8

Alex Craig

## Problem 1

### Instructions

Apply the K-means algorithm to Goldman.imputed.csv data set.

- (a) Use the broken-line algorithm to identify the number of clusters.
- (b) Does clustering reflect gender? To address this question, plot “gender” as 0/1 on the x-axis for individuals 1-1528 and use red/green to color males and females. Save the graphics in the png format with large number of pixels, say, width=6000.
- (c) Upon viewing the graph upon magnification (Windows Photo Viewer works for me), do you think that clustering reflects gender? Hint: Use axes=F and make your own axis(side=1).
- (d) Compute % females in cluster 1 and cluster 2. Does your answer match visual inspection? Display the number of females in the title.

### Solutions