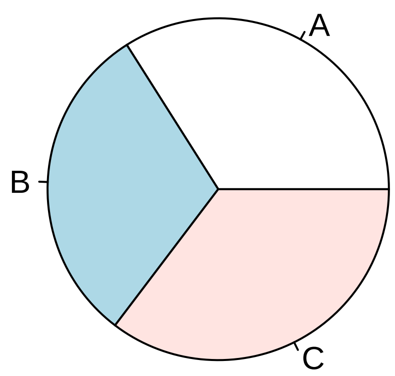
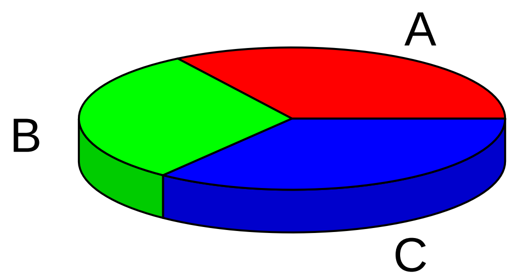
STAT3622 Data Visualization: Quiz 1

27 September 2018 (50min)

Name: ______________________ ID: ______________________ Score: ______________________

1 (70%) The following are 2D and 3D pie charts for a categorical variable with three levels.





a) (10%) From the 2D pie chart, can we draw the conclusion that the three levels have exactly the same frequency? Explain.

b) (10%) Do you think the 3D pie chart is easier to read than the 2D pie chart? Explain.

c) (10%) Does the 3D pie chart bring extra dimensional information compared to the 2D pie chart? Explain.

d) (20%) Suppose we are given a two-way contingency table shown below:

	A	B	C
T	51	46	53
F	15	20	15

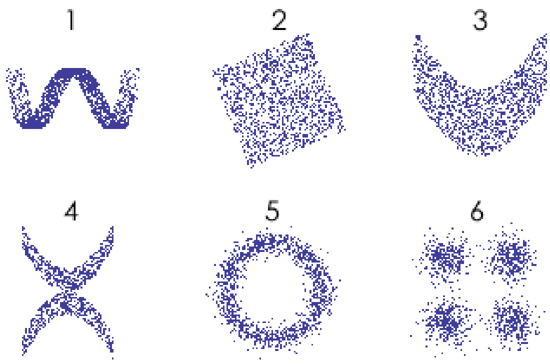
Can you use either 2D or 3D pie chart to visualize such data? If yes, explain how. If no, explain why and suggest a better alternative chart.

e) (20%) Based on the above results, what advices would you give when using 2D and 3D pie charts? List at least three advices.

2 (30%) The Pearson correlation coefficient for two random variables X, Y is defined by

$$\rho = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

where $(\mu_X, \sigma_X), (\mu_Y, \sigma_Y)$ are the means and standard deviations for X and Y, respectively. For the following scatter plots of six bivariate data sets, it is found that they all have the same Pearson correlation coefficient ρ^*.



a) (10%) What is ρ^*? Explain.

b) (10%) Obviously each of the six data sets has some relationship between X and Y. Explain why Pearson correlation cannot discriminate these data sets.

c) (10%) Among the six data sets, which ones (group A) have stronger dependence in (X, Y) than which others (group B)? For each group, find at least one representative.