Danish Iqbal, Kalvin Kao, Matthew Holmes
W200.2 Project 2 Proposal

Titanic Dataset
Description: http://campus.lakeforest.edu/frank/FILES/MLFfiles/Bio150/Titanic/TitanicMETA.pdf
URL: biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/titanic3.xls
Rows: 1309 Passengers, 14 Variables
Github: https://github.com/MIDS-INFO-W18/project2_DI_MH_KK

Features (M, K, D)
Pclass Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)
survival Survival (0 = No; 1 = Yes)
name Name
sex Sex
age Age
sibsp Number of Siblings/Spouses Aboard
parch Number of Parents/Children Aboard
ticket Ticket Number
fare Passenger Fare (British pound)
cabin Cabin
embarked Port of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)
boat Lifeboat
body Body Identification Number
home.dest Home/Destination

Hypothesis (D, M, K)
1. People in lower passenger classes were more likely to die than people in higher passenger classes. Passenger class being a proxy for wealth and status
    a. How does passenger fare correlate to this?
2. What were patterns of mortality between different points of origin?
3. Did more men/women or adults/children survive?
4. What age were most of the survivors?
5. Did families survive more effectively than individuals?
    a. Did parents tend to sacrifice themselves for their children?
6. Where did the iceberg hit and were folks further away more likely to survive?
7. Who was in the lifeboats? It seems more likely they would survive

Plan of Action:
1. EDA
2. Summaries for each variable
3. Correlations between variables, including a multiple linear regression model that seeks to explain survivability
4. Graphs for prioritized variables and questions above