Problem Set 4

You have been given a data set of 154 burn patients with the following variables:

- Obs: The patient number
- Treatment: Whether the patient was treated with routine bathing (0) or body cleansing (1) using chlorhexidine gluconate.
- Male: Male if 1, Female if 0
- White: Self-idenfitied white race if 0, non-white if 0
- PctBurned: Percentage of total body surface burned
- Time: The time on study or until development of a *S. aureus* infection
- Infection: An indicator for the development of an *S. aureus* infection (0 = no infection, 1 = infection)
- type: The type of burn: Scald, Chemical, Flame or Electric

This is a simplified version of the *burn* dataset found in the KMsurv library, and is discussed and analyzed in *Survival Analysis: Techniques for Censored and Truncated Data* by Klein and Moeschberger.

Load this data set by setting your working directory to its location and entering
load("burn.Rda")

Q1. You are interested in the difference in time-until-infection between those treated with the standard bathing protocol vs. those using chlorhexidine. Produce a Kaplan-Meier curve examining this difference.

Q2. Is this difference statistically significant?

Q3. Being a thoughtful, studious epidemiology student, you suspect that there may be confounding between the treatment and the outcome. You set out to use a regression model to estimate a hazard ratio (HR) for treatment type, controlling for these confounding variables.

Describe what steps you take in selecting and evaluating your model, and report your final result.