Questions

1. Below is the summary of built-in lung dataset that ships with survival package:

inst:	Institution code
time:	Survival time in days
status:	censoring status 1=censored, 2=dead
age:	Age in years
sex:	Male=1 Female=2
ph.ecog:	ECOG performance score (0=good 5=dead)
ph.karno:	Karnofsky performance score as rated by physician
pat.karno:	Karnofsky performance score as rated by patient
meal.cal:	Calories consumed at meals
wt.loss:	Weight loss in last six months

(a) Fit the Kaplan Meier estimates of the survival for male and female patients. You should only display the results for day 1, 100 and every 100 days thereafter.

(b) Use the estimates in part (a) to construct the survival curves.

(c) Check whether the survival curves in part (b) are statistically different.

(d) Create a categorical variable on variable age. Label as "young" when the age is less than 62, label as "old" when the age is 62 and above. Fit the Kaplan-Meier estimate of the survival for this new variable.

(e) Fit a Cox Proportional Hazards (PH) Model that uses all the covariates in the data, assuming there are no interactions between the covariates.

(f) Construct a reduced model by removing those insignificant covariate(s). Perform likelihood ratio test to test whether the proposed model is adequate at $\alpha = 0.05$.

(g) Use the model in part (e) to estimate the hazard rate of a female patient who has ECOG score of 3, Karnofsky performance score of 75 (for both physician and patient), calories consumed is 950 and weight loss of 8 in the last 6 months.

(h) Fit an exponential regression model that uses all the covariates in the data, assuming there are no interactions between the covariates. Use AIC and BIC value to compare the performance with the model in part (e).

2. The force of mortality is:

$$\mu_x = \frac{1}{3(120-x)}; \quad 0 \le x \le 120$$

(a) Create a table for μ_x from age 0 to 120. Then, plot the graph.

(b) Derive the ${}_tp_x$. Use R to find the value of ${}_{4|5}q_{30}$.