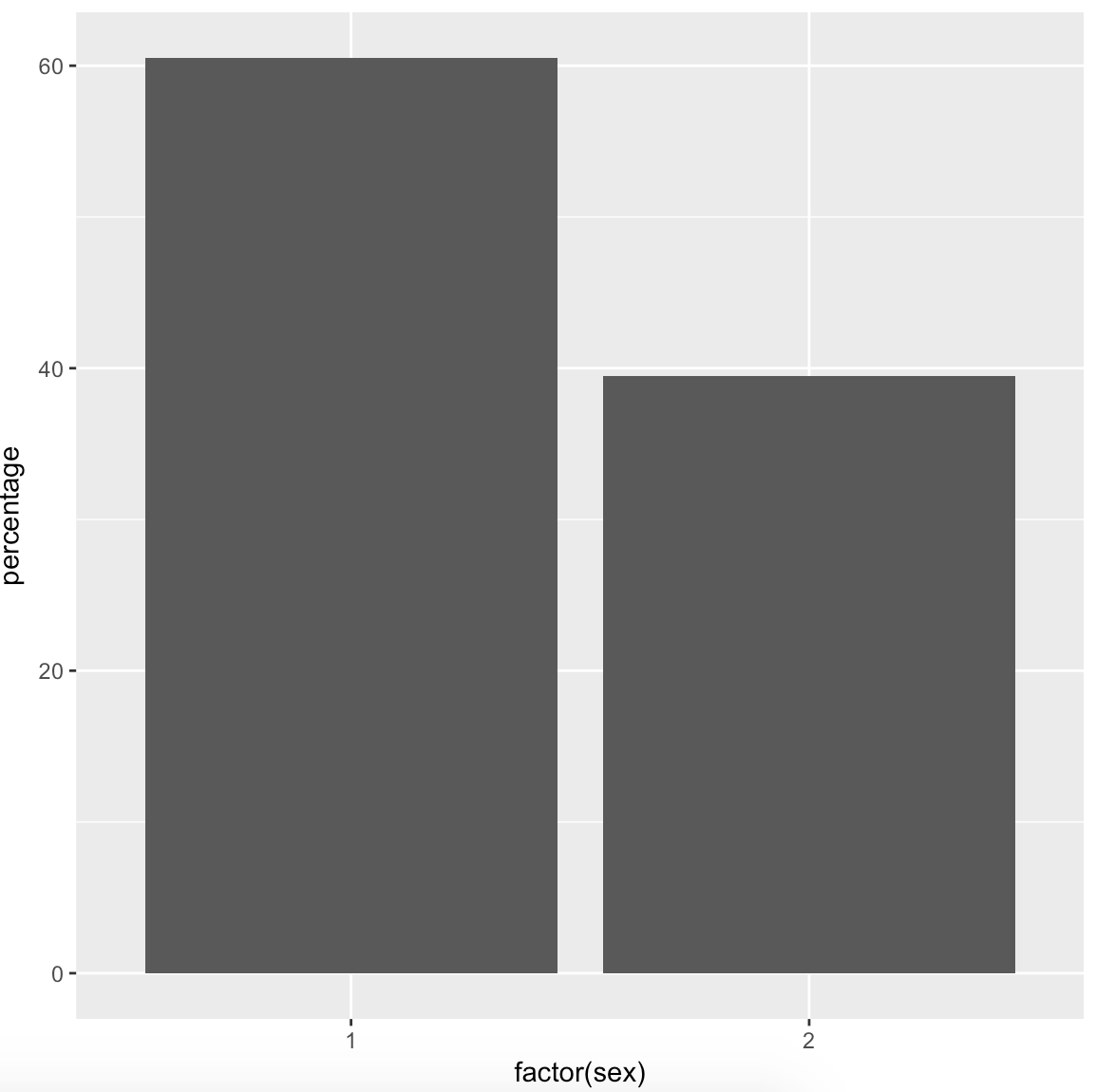
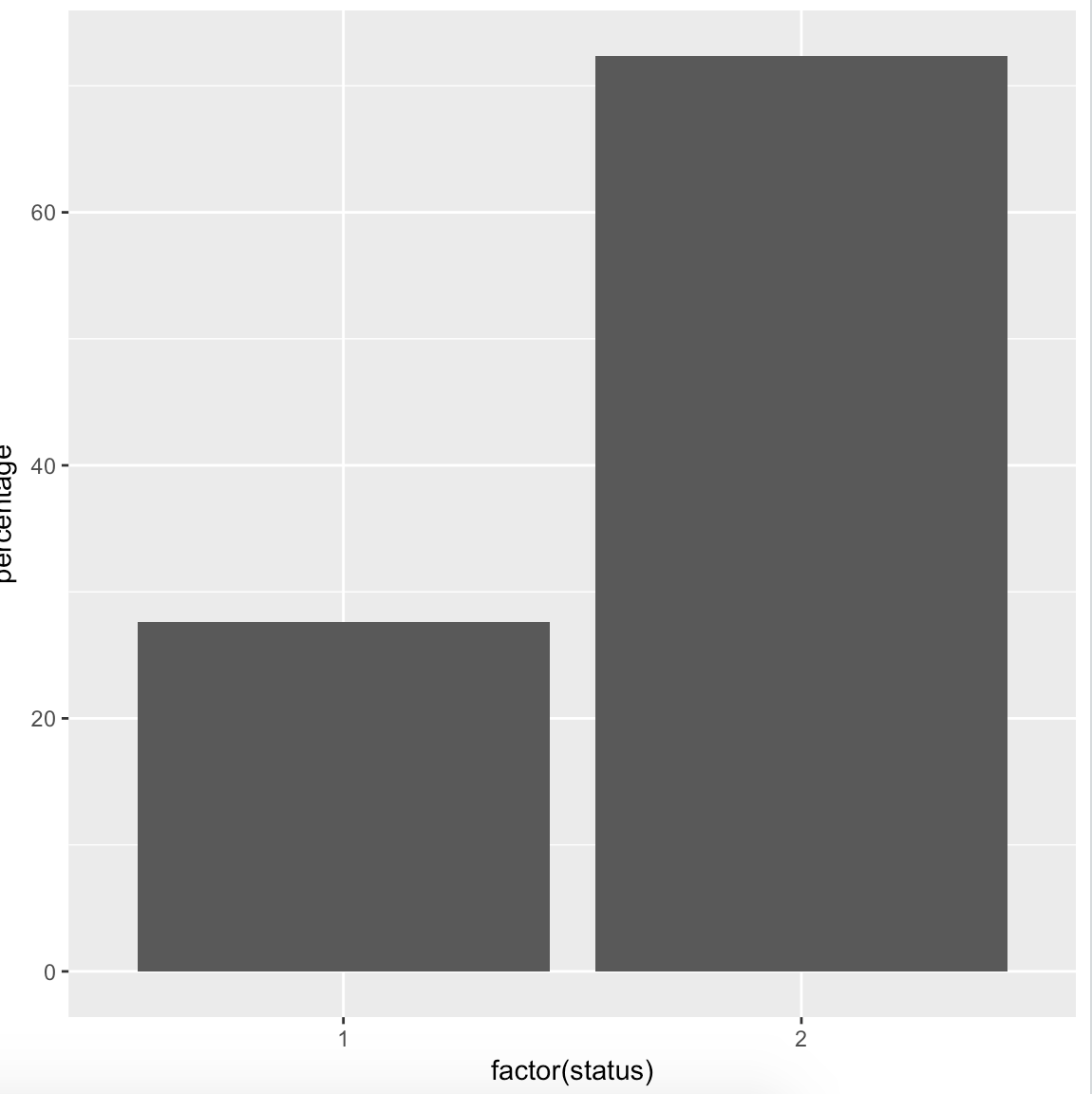
**SURVIVAL ANALYSIS USING**

**COX PROPORTIONAL-HAZARDS MODEL**

**Graphical Representation of Features of Dataset in percentage:**

** Distribution of Sex Distribution of Status.**

****

**1=Male,2=Female 1=Dead ,2=Censored**

**Cox Proportional-Hazards Model:**

The **Cox proportional-hazards model** is essentially a regression model commonly used statistical in medical research for investigating the association between the survival time of patients and one or more predictor variables. It works for both quantitative predictor variables and for categorical variables. Furthermore, the Cox regression model extends survival analysis methods to assess simultaneously the effect of several risk factors on survival time.

**Basics of the Cox proportional hazards model:**

The purpose of the model is to evaluate simultaneously the effect of several factors on survival. In other words, it allows us to examine how specified factors influence the rate of a particular event happening (e.g., infection, death) at a particular point in time. This rate is commonly referred as the hazard rate. Predictor variables (or factors) are usually termed *covariates* in the survival-analysis literature.

The Cox model is expressed by the *hazard function* denoted by h(t). Briefly, the hazard function can be interpreted as the risk of dying at time t. It can be estimated as follow:

h(t)=h0(t)×exp(b1x1+b2x2+...+bpxp)

where,

* *t* represents the survival time
* h(t) is the hazard function determined by a set of p covariates (x1,x2,...,xp)
* Coefficients (b1,b2,...,bp) measure the impact (i.e., the effect size) of covariates.
* h0 is called the baseline hazard. It corresponds to the value of the hazard if all the xi are equal to zero (the quantity exp(0) equals 1). The ‘t’ in h(t) reminds us that the hazard may vary over time.

The quantities exp(bi) are called hazard ratios (HR).

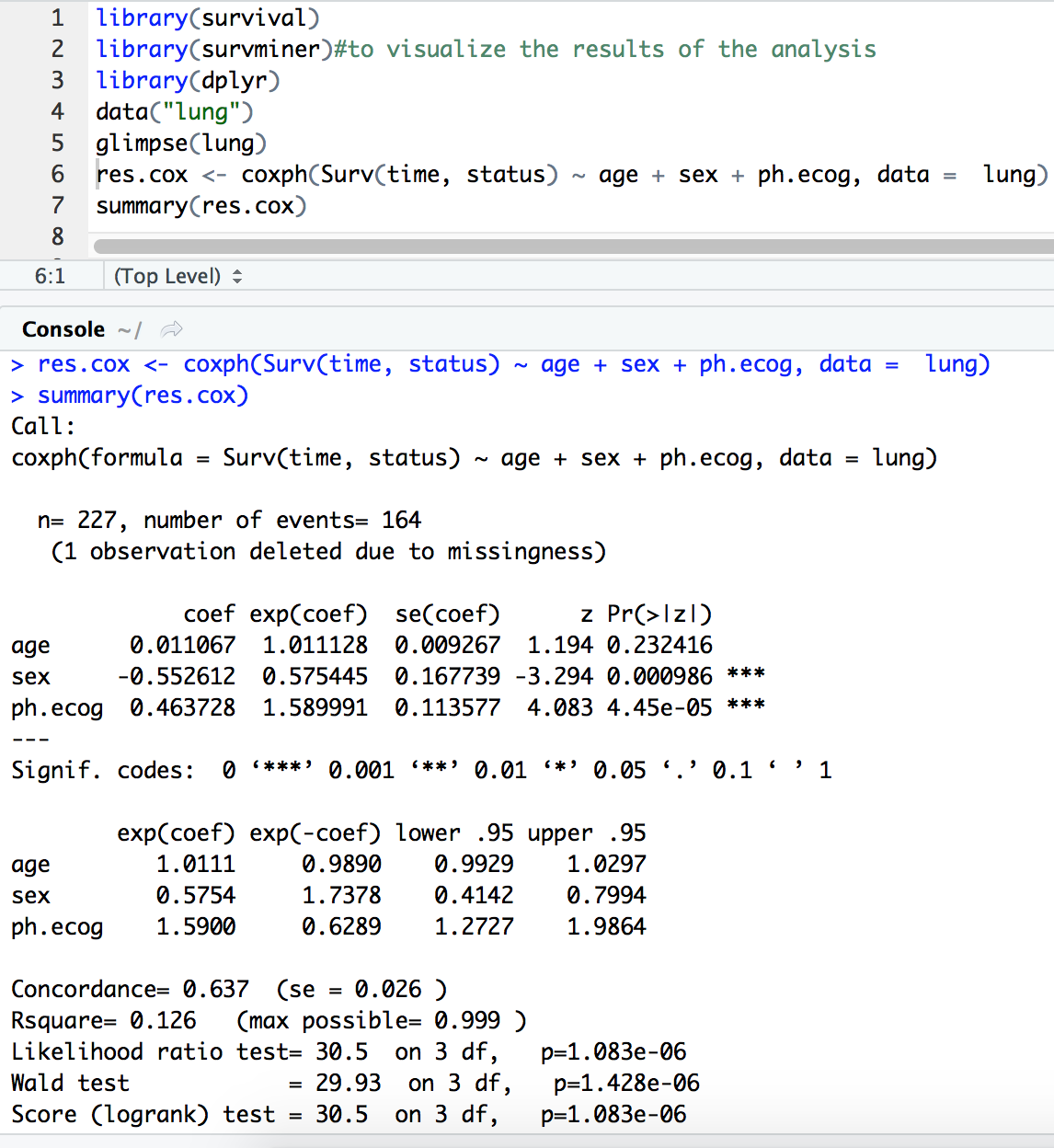
HR above 1 indicates a covariate that is positively associated with the event probability, and thus negatively associated with the length of survival.

In summary,

* HR = 1: No effect
* HR < 1: Reduction in the hazard
* HR > 1: Increase in Hazard

**Multivariate Cox regression analysis:**

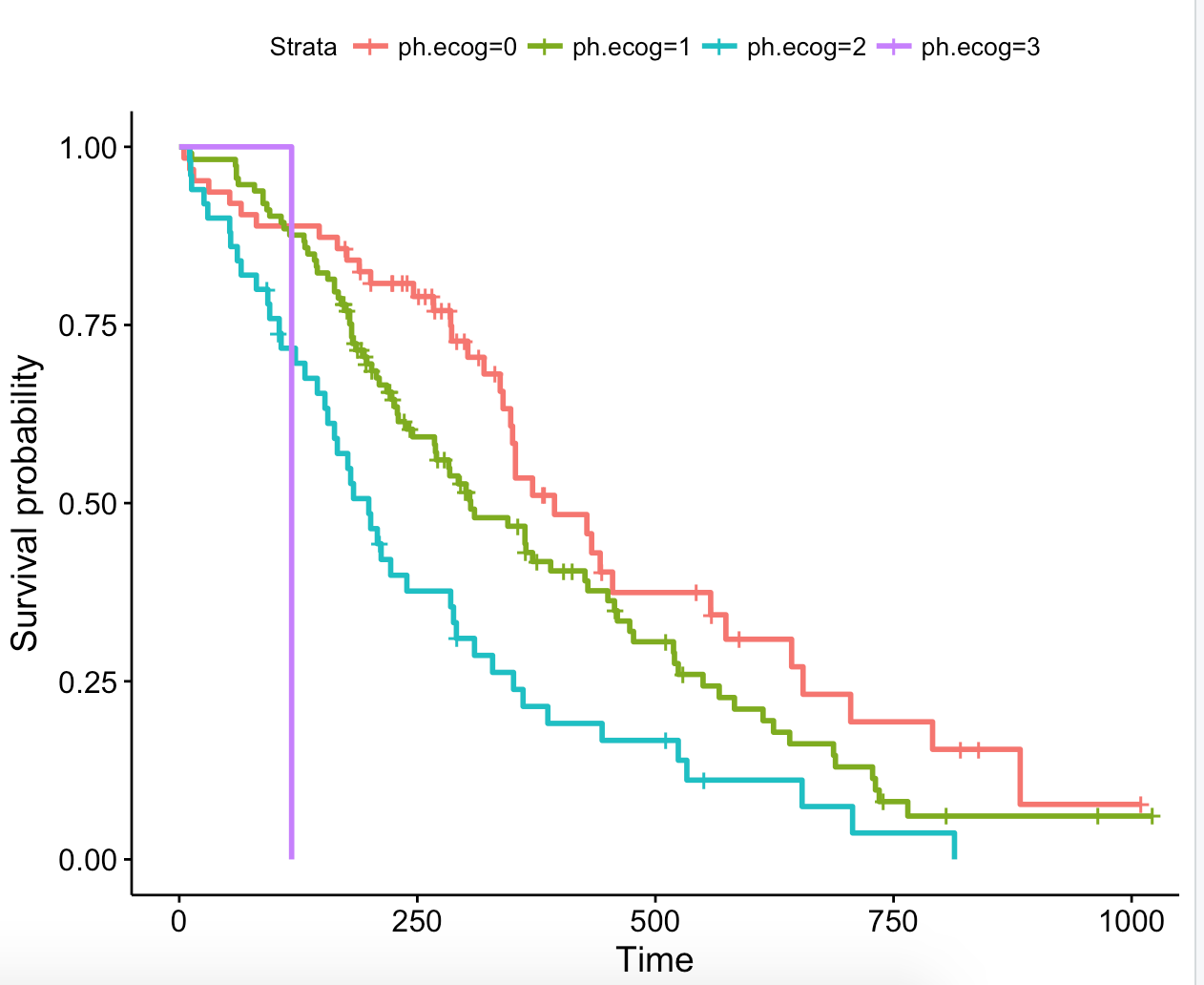
A Cox regression of time to death on the time-constant covariates is specified as follow:



**Interpretations:**

* The p-value for all three overall tests (likelihood, Wald, and score) are significant, indicating that the model is significant. These tests evaluate the omnibus null hypothesis that all of the betas (ββ) are 0. In the above example, the test statistics are in close agreement, and the omnibus null hypothesis is soundly rejected.
* In the multivariate Cox analysis, the covariates sex and ph.ecog remain significant (p < 0.05). However, the covariate age fails to be significant (p = 0.23, which is grater than 0.05).
* The p-value for sex is 0.000986, with a hazard ratio HR = exp(coef) = 0.58, indicating a strong relationship between the patients’ sex and decreased risk of death. The hazard ratios of covariates are interpretable as multiplicative effects on the hazard. For example, holding the other covariates constant, being female (sex=2) reduces the hazard by a factor of 0.58, or 42%. We conclude that, being female is associated with good prognostic.
* Similarly, the p-value for ph.ecog is 4.45e-05, with a hazard ratio HR = 1.59, indicating a strong relationship between the ph.ecog value and increased risk of death. Holding the other covariates constant, a higher value of ph.ecog is associated with a poor survival.
* By contrast, the p-value for age is now p=0.23. The hazard ratio HR = exp(coef) = 1.01, with a 95% confidence interval of 0.99 to 1.03. Because the confidence interval for HR includes 1, these results indicate that age makes a smaller contribution to the difference in the HR after adjusting for the ph.ecog values and patient’s sex, and only trend toward significance. For example, holding the other covariates constant, an additional year of age induce daily hazard of death by a factor of exp(beta) = 1.01, or 1%, which is not a significant contribution.

**Visualizing the Estimated Distribution of Survival Times.**

**Survival probability of ph.ecog Covariate**

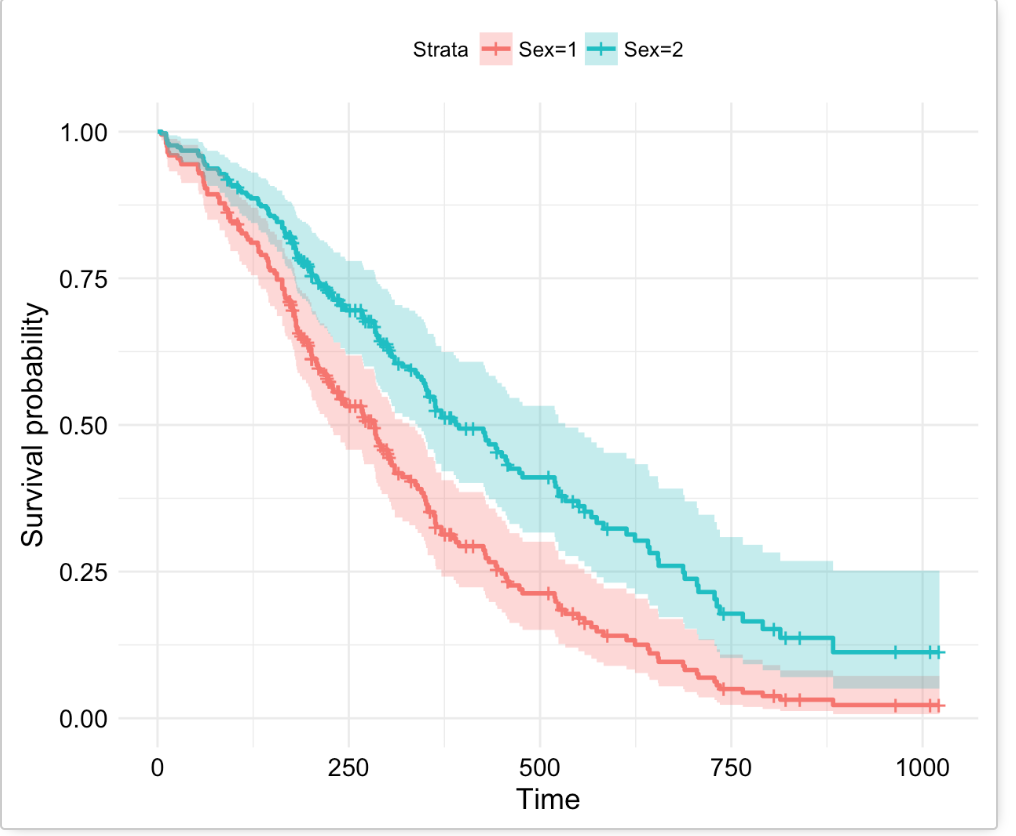
**Rcode:**

fit2 <- survfit(Surv(time,status)~ ph.ecog, data= lung)

ggsurvplot(fit2, data = lung)

* we can see that the ph.ecog variable significantly influence the patient’s  risk of death in this study.
* Strong relationship between the ph.ecog value and increased risk of death. Holding the other covariates constant, a higher value of ph.ecog is associated with a poor survival

**Survival probability of sex covariate**



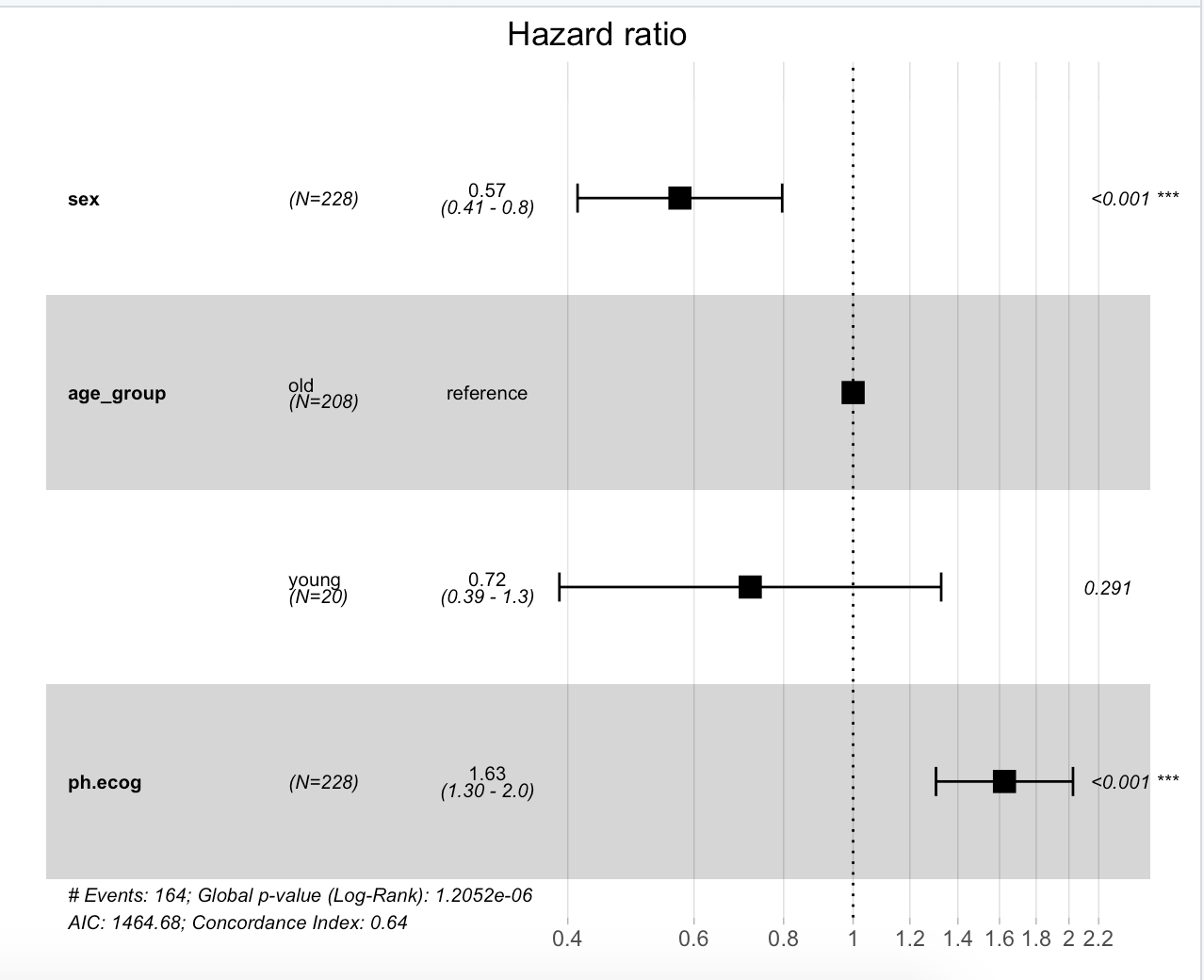
Rcode:

fit <- survfit(Surv(time,status)~ sex, data = lung)

ggsurvplot(fit, data = lung)

* Strong relationship between the patients’ sex and decreased risk of death.
* Holding the other covariates constant, being female (sex=2) reduces the hazard by a factor of 0.58, or 42%. We conclude that, being female is associated with good prognostic.

**Risk of death and respective hazard ratios.**



**Rcode:**

fit.coxph <- coxph(Surv(time, status) ~ sex + age\_group + ph.ecog, data = lung)

ggforest(fit.coxph, data = lung)

* we can see that the sex, ph.ecog variables significantly influence the patient’s  risk of death in this study.
* •Age group has no effect.