Acute myocardial infarction ("heart attack") is caused by the formation of a clot in one of the coronary arteries that supply blood to the heart muscle. Acute MI is a major public health problem. Mortality is substantial in the period immediately after the event (and also during the years after surviving the initial infarction). Some patients die before reaching the hospital. Patients seen in hospitals are reported to have an average mortality within 30 days around 6–15%, with improvement over time. The risk of 30-day mortality strongly depends on various prognostic factors. In younger patients, risks are much lower than in older patients. Other patient demographics are also important (gender, length, weight), as well as the presence of risk factors (hypertension, diabetes, smoking, family history) and the history of previous cardiovascular events (previous MI, angina, stroke, bypass surgery). Relevant presenting characteristics include the location of the infarction and the extent of ECG abnormalities. Very important is the acute state of the patient as reflected by blood pressure, heart rate, and left ventricular function (e.g., presence of shock).

Aim of this study is to find predictors of 30-days mortality.

Data structure:

variable	labels
day30	EVENT 1: death; 0=alive
Sex	0=Male; 1=Female
Age	age in years
a65	age >= 65 =1 ; age < 65=0
killip	killip class (ordinal)
Sho	Shock: Killip class 3/4 versus 1/2 (0/1)
Dia	Diabetes (0/1)
Hyp	Hypotension: systolic BP < 100 (0/1)
Hrt	Heart rate: pulse >80 ("tachycardia", 0/1)
Ant	Anterior infarct location (0/1)
pmi	Previous myocardial infarction (0/1)
Hei	Height in cm
Wei	Weight in kg
Htn	Hypertension history (0/1)
smk	Smoking (1 = never; 2 = ex; 3 = current)
Lip	Lipids: hypercholesterolaemia (0/1)
Pan	Previous angina pectoris (0/1)
Fam	Family history of MI (0/1)
Ste	ST elevation on ECG: number of leads
st4	ST elevation on ECG: >4 leads (0/1)
Ttr	Time to relief of chest pain > 1 h (0/1)

Questions:

1) Build a descriptive table, comparing patients dead *versus* patients alive at 30 days. Insert also a column with the total population descriptive statistics.
2) Perform univariable logistic regression analyses, of all candidate predictors for your model. [**Optional:** For Age is the linearity effect reasonable? How could you model alternatively the age effect?]

3) Build a multivariable logistic regression model starting from the list of significant predictors at univariable analyses. Pay attention to multicollinearity (same variables recoded in different ways).
4) Evaluate model performance in discrimination by means of the AUC under the ROC curve. Evaluate also model performance in terms of calibration.
5) Represent the estimated model by means of a nomogram (hint: R function nomogram {rms})
6) Internally validate the estimated model (hint: R function validate {rms})
7) **Optional:** Setting aside the interpretability of the model, are you able to find a machine learning algorithm that predicts the risk of event with a similar (or better) performance than the logistic model?