绿色表示第一个选项的条件，黄色表示第二个选项的条件。任务是建立一个回归模型，根据数值变量长度 重量 轴距 马力 发票 发动机尺寸 汽缸数以及分类变量产地和类型预测高速公路（MPG_highway）或城市（MPG_city）的汽油消耗量。您可以使用课程中涉及的方法（线性回归、lars/lasso、pls/pcr、对数正态、GLM、splines、loess、带有自定义方程的非线性回归以及这些方法的组合，例如，您可以使用一种方法过滤变量、或用一种模型估计异常值，然后从训练集中移除异常值（不能从测试集中移除观测值），再用另一种程序使用过滤后的数据集建立模型。要评价所建模型的质量，可使用 MAPE 分数 = (1/n) * Σ（|原始值-预测值| / |原始值|）。任何方法都可以用来评估模型的质量：信息标准、基于交叉验证的统计评估、自举法、验证样本等。最终模型将由其中一种方法进行验证，而您不知道该方法的参数（因此不存在拟合），您将被告知 MAPE 评估的结果。您的模型应适用于训练和测试数据集的任何子集（不要假设测试或训练集中有这样或那样的观测数据）。封闭验证程序得出的 MAPE 应小于或等于 0.065。推荐的技术有

- 利用方差分析对分类变量进行转换（对数值进行分组）（切记也要在测试集中进行这些转换）
- 使用逐步法或正则化方法选择重要变量
- 转换输入变量并在回归方程中加入多项式项（切记在测试集和应用模型时都要进行这些转换）。
- 使用非线性依存方程，包括制作自己的 "神经网络"，并使用 nls 将其训练为非线性回归。
- 转换响应或使用具有不同误差分布和链接函数的广义线性（或非线性）模型（切记在 预测后对响应进行适当的重新计算）。

要获得理想的模型质量，通常只需使用上述两种技术即可。

在整个数据集上，用 20×20 个点的统一网格（数据集中的网格应独立生成），绘制最终 模型中一对最重要选定变量的响应依赖性三维图（等高线图），在绘制该图时，应取 该对变量中未包含的其他变量（如有）的平均值。

将整个功能作为一个支持方法的类来实现：

- model<-fit**(train_data)** - 以 train_data（初始汽车**数据**集的行子集）为基础建立模型，包括所有必要的数据预处理和最终模型的构建
- **predict(model, test_data)** - 为 test_data（初始汽车集的行子集）生成一个预测向量，该方法应包括 fit 所做的所有数据预处理（除移除异常值外，可以在不移除观测值本身的情况下修正异常观测值的特征）。
- **plot(model)** - **绘制**三维图或等值线图