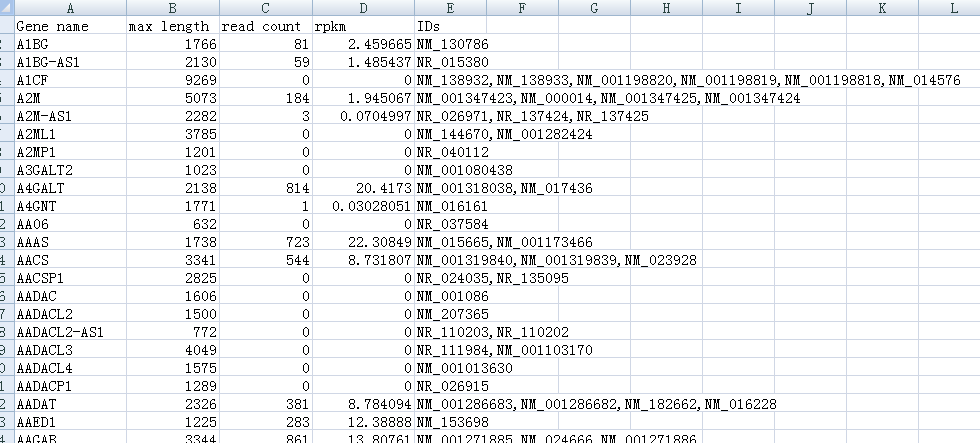
请独立完成以下笔试题，

要求：对于文本操作，不限编程语言；画图部分，请用R语言

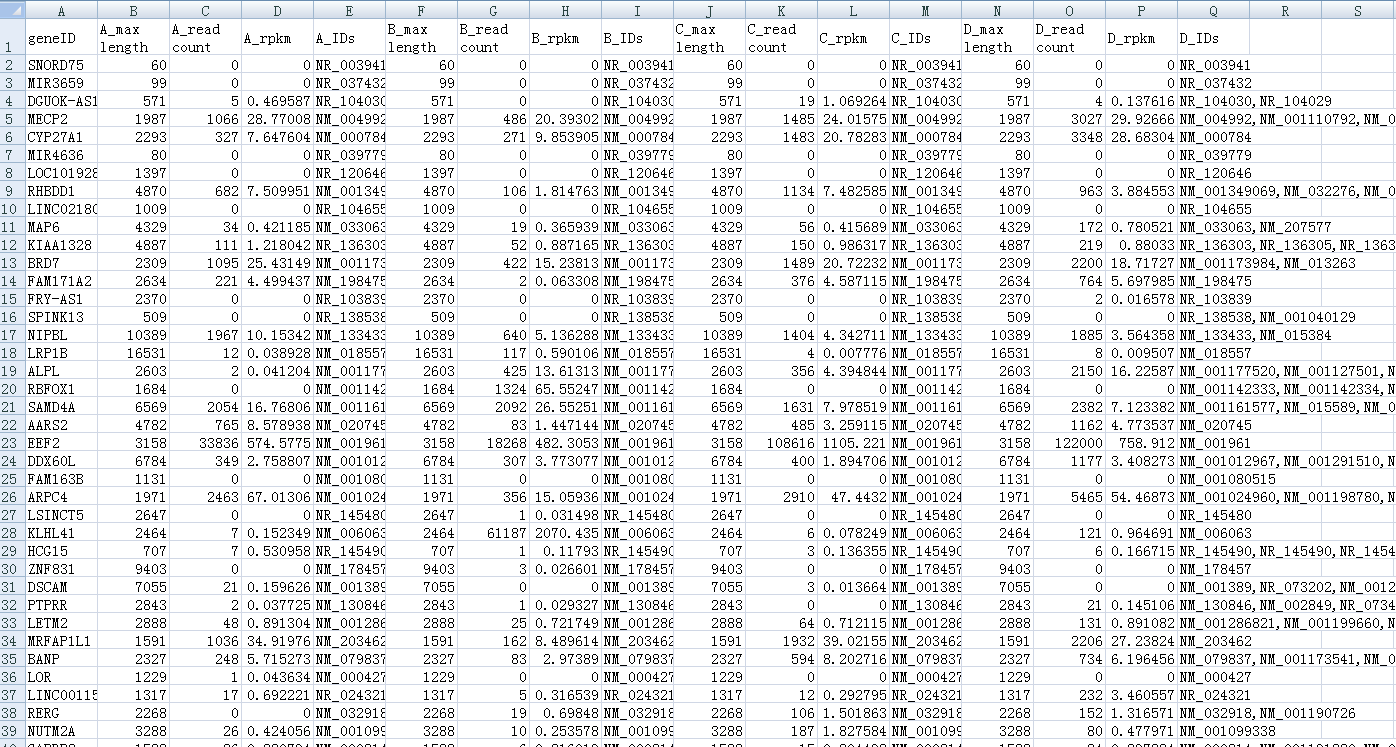
完成后，将程序代码和结果文件一起打包发回

如有必要，可以附带文字说明

1，“各样品定量结果”文件夹下是4个样品的定量文件，每个文件包含了2万多个基因的表达量以及对应信息，这2万多个基因在4个文件中的顺序可能不一样，如下截图：

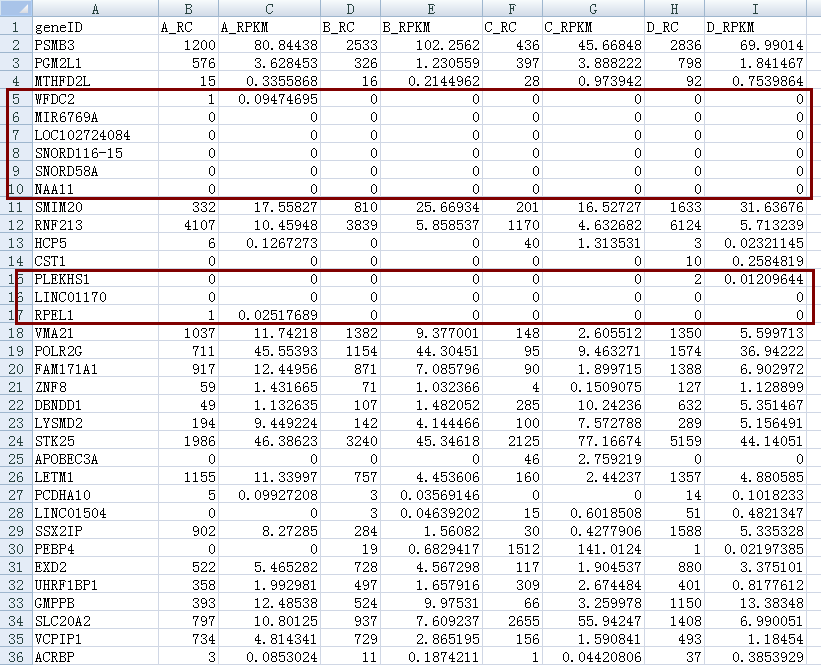


现要求：编写程序，根据基因名，将各样品信息按行合并为一个文件，并命名为merge.txt，合并后结果类似如下截图。为了可重复利用程序，建议所写的程序能够应对任意多个样品。

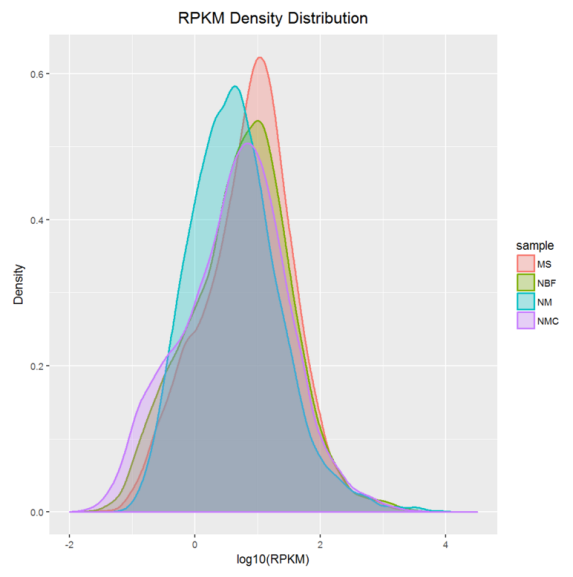
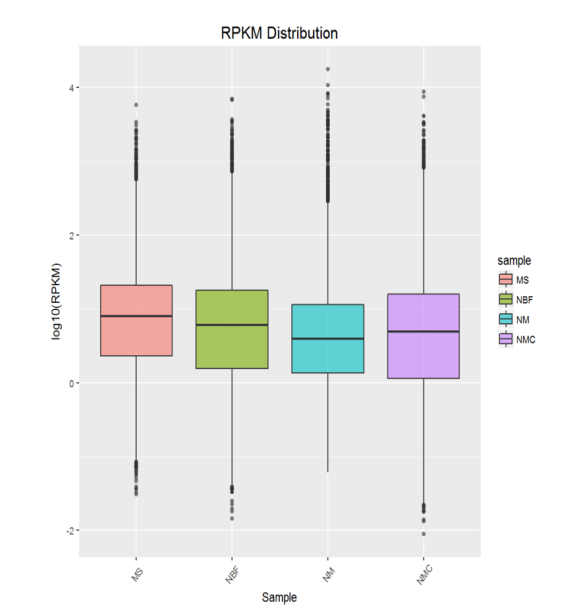


2，RC\_RPKM.tx 文件记录了4个（A、B、C、D）样品2万多个基因的表达情况。截图如下。在分析的时候，因所使用的mapping算法精度高，可认为测序得到 10 reads 以上的基因是有表达的，而 10 reads 以下的基因没有表达。

现要求：编写程序，去掉在四个样品中RC均小于10的基因（如截图中红框部分应去除），结果文件另存为RC\_RPKM\_ge10.txt。为了可重复利用程序，建议所写的程序能够应对任意多个样品。



3，基因表达丰度箱线图和密度图展示。利用“各样品定量结果”文件夹下的4个定量文件，去掉样品中read count小于10的基因，然后对其rpkm取10的对数后（即log10(rpkm)）进行箱线图和密度图展示。示例图如下：



4，“kegg\_pathway\_enrichment.txt”是差异基因进行KEGG通路富集分析的结果。现对前20个通路（即前20行）进行点图展示。

展示效果图如下，要求：横坐标 rich factor = Input number / Background number；纵坐标为通路名称，p value 越小，点图颜色越绿，反之越红；点的大小反映Input number的数值，在图例中标为Gene\_Number，表示有多少个基因富集在该通路中。为了整齐好看，建议按rich factor进行排序。

