# 数据挖掘作业2

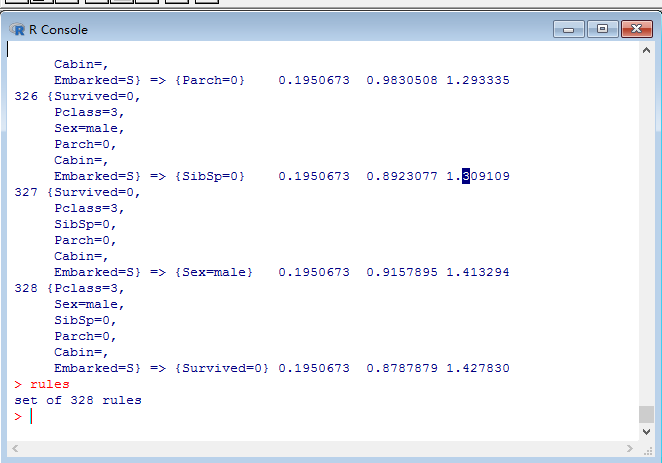
# ——titanic存活数据关联规则挖掘报告

数据集：<https://www.kaggle.com/c/titanic/download/train.csv>

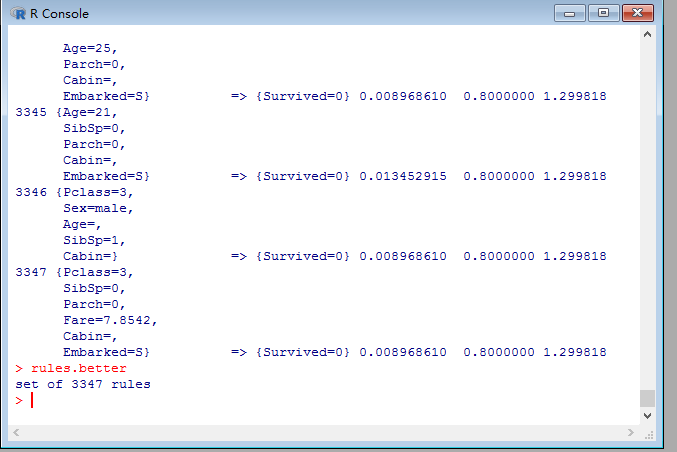
实验环境：win10，R语言3.3.1版本

* 加载库：

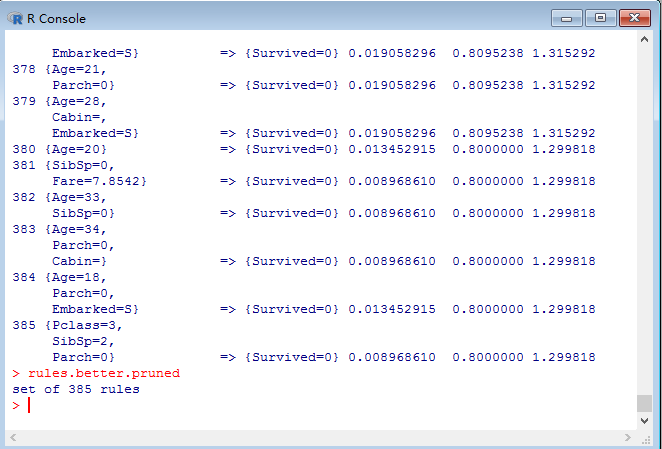
arules，用apriori算法进行关联分析



* 设置条件，按照所需类别进行配置。比如：生存情况，那我的条件就设置为右边的集合rhs为rhs=c("Survived=0","Survived=1")，然后最小关联项、支持度、置信度分别设置为2、0.005、0.8，结果如图：

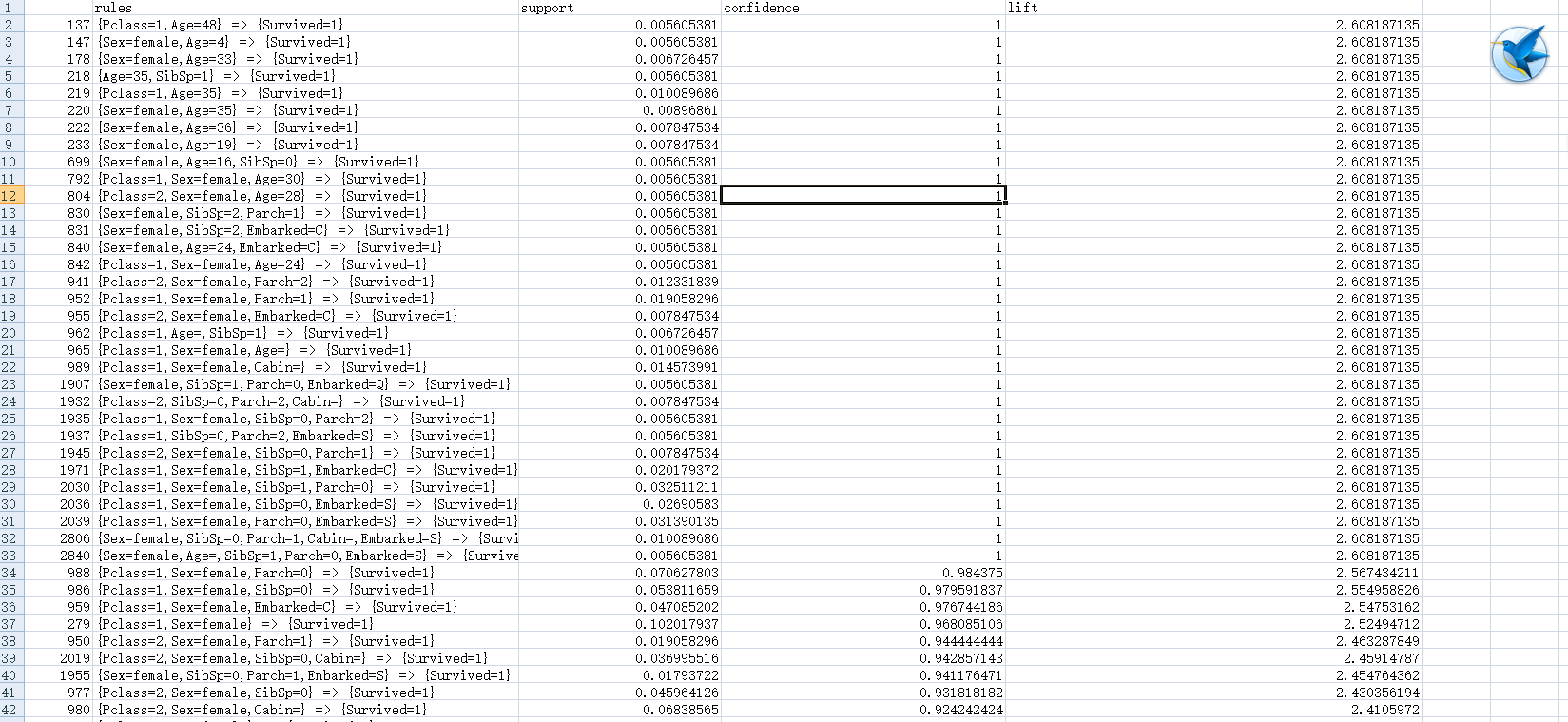


* 去除冗余的规则，结果如图：



* 结果进行评价

由于有385条相关的规则，所以导出结果后进行评价，从挖出的规则来分析，如图：



可以看出最明显的结果是置信度为1的情况下，生存下来的人的条件有这么几个特征：Pclass=1、Sex=female，通俗点讲就是头等舱的女乘客生存几率最大

* 加载arulesViz库，进行画图，结果如图：

