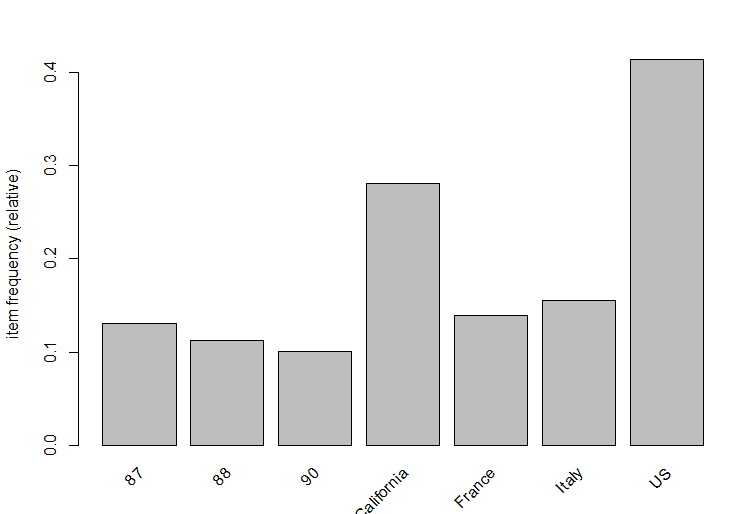
## 数据挖掘大作业二：关联规则挖掘

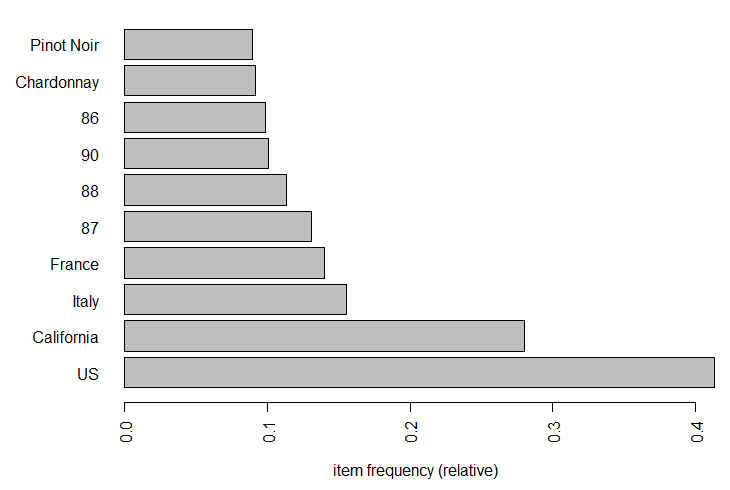
1. 对数据集进行处理，转换成适合关联规则挖掘的形式；
2. 找出频繁项集；
3. 导出关联规则，计算其支持度和置信度;
4. 对规则进行评价，可使用Lift及其它指标, 要求至少2种；
5. 对挖掘结果进行可视化展示。
6. 关联规则需要把源数据的格式转换为稀疏矩阵，1表示访问，0表示未访问
7. 支持度 Support

项集的支持度就是该项集出现的次数除以总的记录数，支持度的意义在于度量项集在整个事务集中出现的频次。我们在发现规则的时候，希望关注频次高的项集。

* 按最小支持度查看

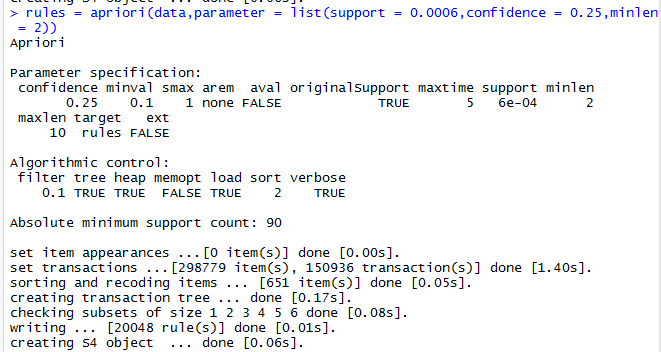


* 按照排序查看

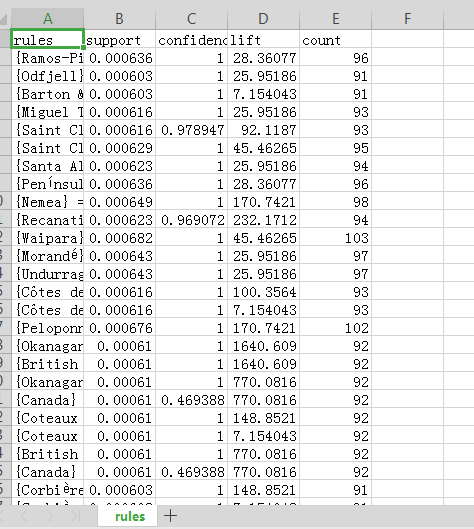


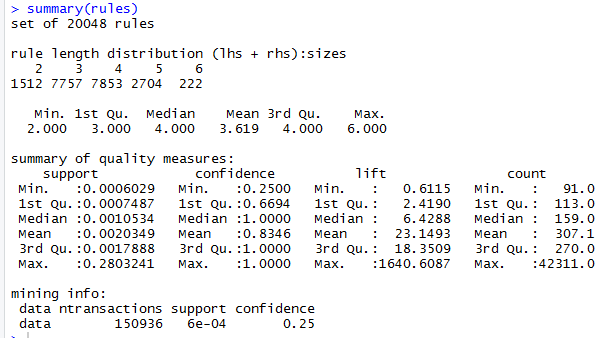
1. 生成规则

从返回的结果看，总共有20048条规则生成。



1. 导出规则，计算其支持度和置信度



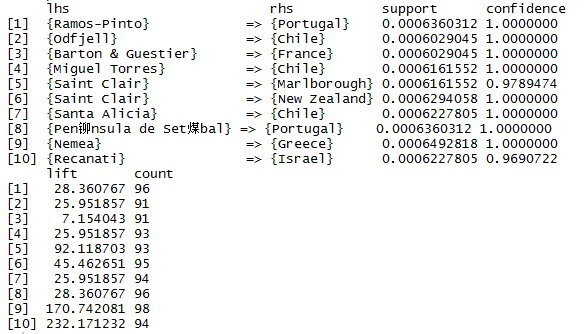


第一部分：规则的长度分布：就是minlen到maxlen之间的分布。如上例，len=2有1512条规则，len=3有7757。同时，rule length的五数分布+均值。

第二部分：quality measure的统计信息。

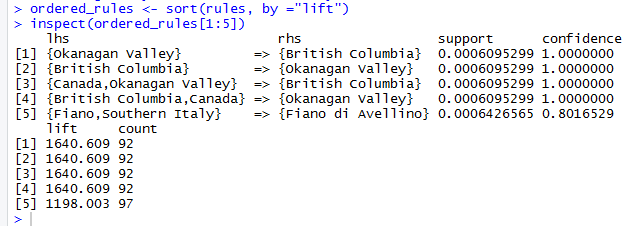
第三部分：挖掘的相关信息

具体规则记录



1. 评估规则

对规则排序



1. 规则展示

