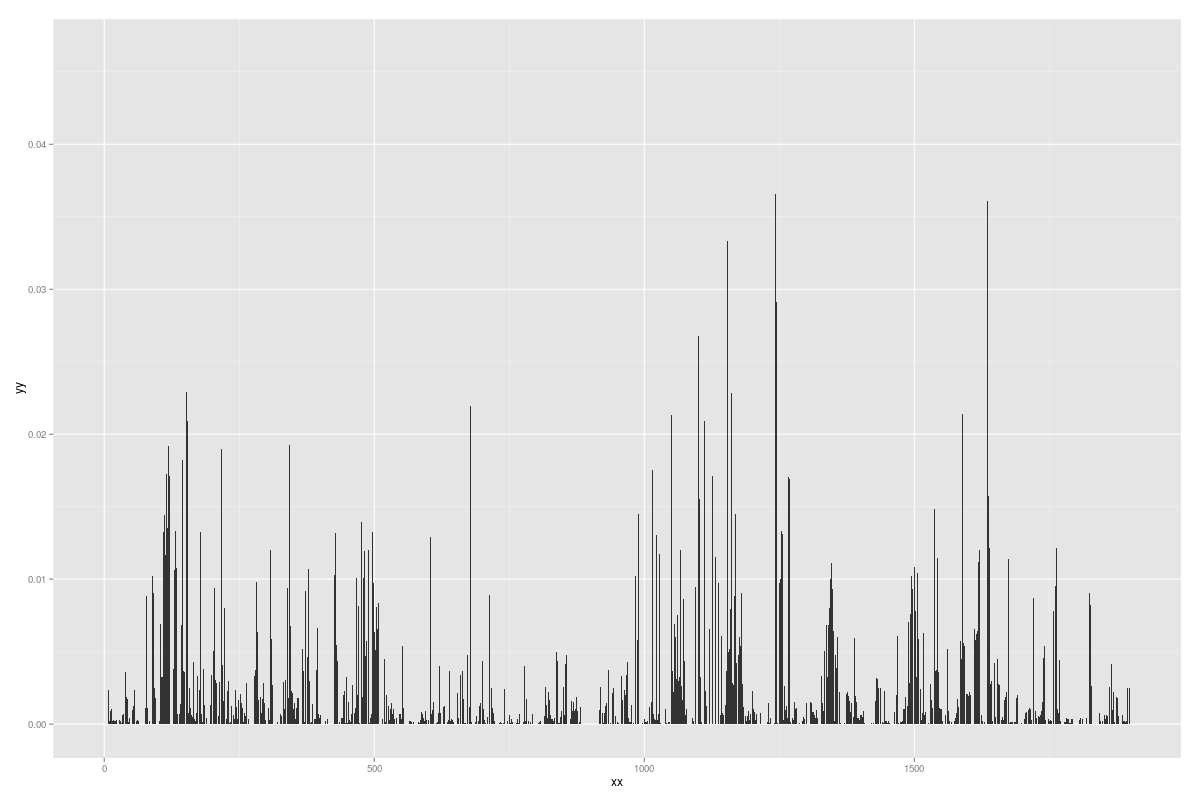
Apriori Algorithm implimentation on Yangshan data.

主要根据 SalesDetails.xlsx 中共 338941 条购物数据, 应用 Association Rules 中的 The Apriori Algorithm, 对购买数据相关关系简单分析.

首先,根据 ProductDetails.xlsx 中的货物信息, 获得共1898 种不同货号的商品, 并将 SalesDetails 中的购买信息按照单号划分为52424 次的购买行为, 每次的购买行为我们用一个 1898 维 0-1向量表示每种商品购买与否,最终形成一个 52424\*1898 的矩阵 XN.( 注意由于ProductDetails 商品货号不全,所以有些购买商品被忽略 ).

根据 XN ,为获得 support > threshold 的m维子集, 首先需要得到m-1维 support > threshold 的子集,这里我们以 m=2 为例, 由于threshold 选择需要人为指定, 我们先展示以下结果:

1. 商品一维子集的 support (基于此可以得到近似的概率分步):



2. support 最高前top.n件商品及相应的support 值: (这里设定 top.n = 20 )

3. 在一维 support >0.02 的子集的二维组合中, support值最高的top.n组商品,及相应的 frequency, lift, confidence 值.

