基于商品名称的自动化分类

数据形式：

序号	商品名称	一级品类	二级品类	三级品类
1	大洋世家 厄瓜多尔白虾（20/30） 1.45kg 30-45只 大虾 海鲜火锅	生鲜	海鲜水产	虾类
2	浓鲜时光 麻辣小龙虾 1600g 4-6钱 净虾850g 盒装	生鲜	海鲜水产	虾类
3	京鲜生 清水熟冻大号麻辣小龙虾尾虾球 500g×3袋 共1500g	生鲜	海鲜水产	虾类
4	Acornfresh 冰岛野生北极甜虾仁熟冻虾仁 北极匮限量捕捞 婴幼儿宝宝营养辅食无污染海鲜 200克*2包	生鲜	海鲜水产	虾类
5	鲜美来 冷冻虾滑150g 袋装 火锅食材 海鲜水产	生鲜	海鲜水产	虾类
6	一米渔 大青虾4斤盒装 海鲜水产鲜冻虾基围虾超大舟山对虾海虾鲜虾青虾	生鲜	海鲜水产	虾类
7	2018新疆吐鲁番葡萄干500g新疆特产绿提子干无核免洗批发干果 葡萄干500g*2袋	食品饮料、保健食品	地方特产	新疆
8	（缓缓节7折）和田大枣特级新疆袋装新货免洗骏枣5斤装红枣干 1000克(500克*2袋)	食品饮料、保健食品	地方特产	新疆
9	中粮初萃物理压榨一级浓香花生油5L 新榨90天之内新油 食用油充氮保鲜包邮2018年新榨	食品饮料、保健食品	粮油调味	食用油
10	英潮鲜椒酱虎邦辣酱山东特产辣椒酱特辣超辣香辣酱 虎皮辣椒酱210g	食品饮料、保健食品	粮油调味	调味品
11	【两件免邮】英潮鲜椒酱虎邦辣酱 辣椒酱组合装 鲁西牛肉酱50g*3罐	食品饮料、保健食品	粮油调味	调味品
12	乌冬面汁，日式乌冬面用汁，带有汤水的乌冬面，味丰出品！	食品饮料、保健食品	粮油调味	调味品
13	开心乐梨汁冰糖小粒黄冰糖350g	食品饮料、保健食品	粮油调味	调味品
14	厨邦葱姜汁料酒500ml瓶装提味增香去腥解腻料酒调味烹饪	食品饮料、保健食品	粮油调味	调味品
15	【买2送2再送杯】红豆薏米芡实茶 祛湿茶200g 大麦苦荞茶养生茶 除湿气茶可去湿气湿热赤小豆薏仁茶	食品饮料、保健食品	茗茶	花草茶
16	聚呈绿茶 龙井 250g 茶叶 雨前西湖龙井【买一送一送同款】龙井茶 散装罐装 2018新茶	食品饮料、保健食品	茗茶	龙井
17	【团购优惠】中粮礼品卡中秋节礼品册团购 福礼398型自选礼品卡册购物卡	食品饮料、保健食品	食品礼券	卡券
18	皇中皇 肇庆特产正宗传统裹蒸粽400克 猪肉绿豆超大粽子 广式早餐粽子 广东老字号 400gx1只	食品饮料、保健食品	食品礼券	粽子
19	口水娃 多味花生 烤肉 五香 香辣 牛肉味30g 可选 不备注随机发 小包包 酱汁牛肉味	食品饮料、保健食品	休闲食品	休闲零食
20	可口可乐 迷你罐组合 200ml*48罐 可乐+雪碧+芬达+零度可乐 碳酸饮料汽水	食品饮料、保健食品	饮料冲调	饮料
21	卓玛景 西藏冰川饮用天然水12L*1桶 弱碱性小分子母婴饮用水 家庭桶装水饮用水 非矿泉水苏打水 老会员专拍	食品饮料、保健食品	饮料冲调	饮用水
22	怡宝 饮用水 饮用纯净水1.555L*12瓶 整箱装	食品饮料、保健食品	饮料冲调	饮用水
23	名仁 苏打水饮料 无糖无汽弱碱性水 375ml*24瓶 整箱装	食品饮料、保健食品	饮料冲调	饮用水

分析任务：

1. 设置好工作路径，找到数据集 catalogs.csv。该数据里面包含每个商品的名称以及对应的一二三级分类。读入该数据，命名为 catalog。然后，使用 head()函数查看数据的基本形式。
2. 数据中包含两个一级品类，即“生鲜”和“食品饮料、保健食品”。请使用 R 中的饼状图对每个一级品类下的各个二级品类的分布情况进行描述分析，并进行文字解读。
3. 利用分词工具对一级品类为“生鲜”的数据进行分词处理，并做出来分词后的词云图
4. 对一级品类为“食品饮料、保健食品”的数据分词结果进行建模前的预处理（参考代码课上的各种预处理操作）
5. 将上一步得到的预处理后的数据按照 8:2 的方式划分训练集和测试集，在训练集上建立朴素贝叶斯模型并在测试集上评估模型的效果并解读模型结果（参考代码课上的建模过程）
6. 尝试改变参数，例如预处理过程中针对高频词和低频词的处理阈值进行调整，以及对构造矩阵时选取的维度（代码课上选择的是 Top 1000）进行调整，再次尝试在相同的数据集上建模并进行对比，谈谈你的发现与理解