

Business Analytics Homework 4

Instructor: Zach Zhizhong ZHOU

1. 在 D 盘建立目录: D:/BA/Homework/HW04, 将附件中的数据放入该目录中, 以该目录做为工作目录。

某企业在中国内地各个地区的产品销售情况如下:

华东地区 (包括山东、江苏、安徽、浙江、福建、上海): 40,000

华南地区 (包括广东、广西、海南): 30,000

华中地区 (包括湖北、湖南、河南、江西): 20,000

华北地区 (包括北京、天津、河北、山西、内蒙古): 25,000

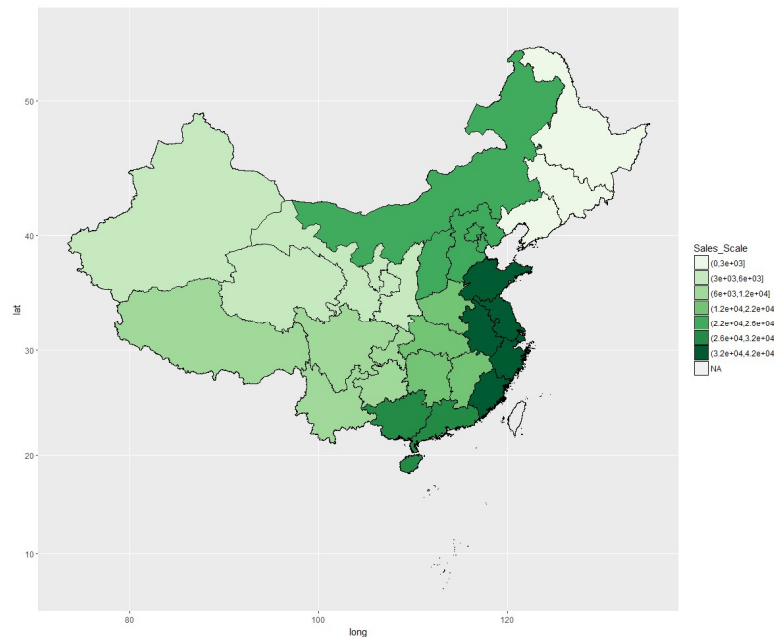
西北地区 (包括宁夏、新疆、青海、陕西、甘肃): 5,000

西南地区 (包括四川、云南、贵州、西藏、重庆): 10,000

东北地区 (包括辽宁、吉林、黑龙江): 2,000

将销量数据画成中国分区统计图。画分区统计图前应将销量取值划分成 7 个区间, 分别是 (0,3000], (3000,6000], (6000,12000], (12000,22000], (22000,26000], (26000,32000], (32000,42000]。分区统计图应使用深浅不同的绿色画出这 7 个取值区间, 取值越高的区间绿色越深 (见下图)。注意中国地图不要漏掉台湾。

你会用到课程文件夹/BA/CHNMap/ 当中的 ProvIndex.csv 和 ProvMapData.csv 数据文件, 但是中国各大区销量的数据文件需要你自已填写。



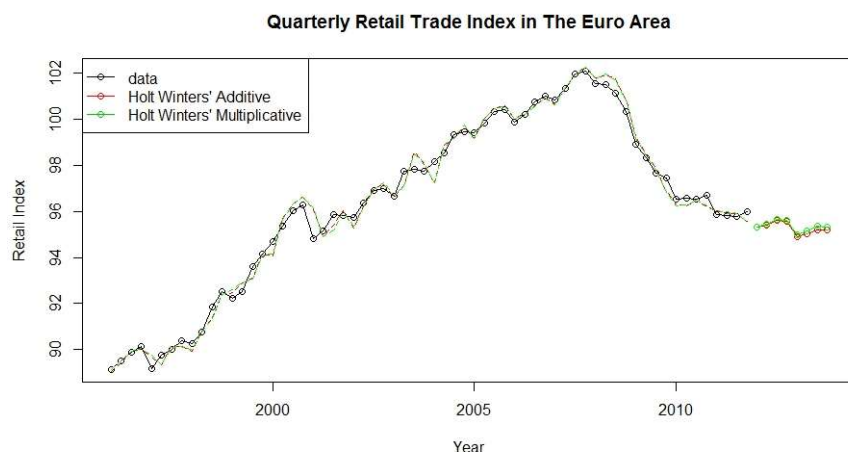
2. PengPai_JieMian.R 文件有部分代码被删掉了, 你需要补全 R 代码, 确保运行这个 R 文件, 可以从澎湃新闻网站提取出以下信息: (1) 一天之内的热门新闻、(2) 三天之内的热门新闻和 (3) 一周之内的热门新闻。并且可以从界面新闻天下栏目提取出以下信息: (4) 界面新闻天下栏目首页中《最新报道》下所有新闻的标题和 (5) 这些新闻的 URL 链接。
3. 在 R Script 中输入 `library(fpp); data(euretail)` 调入名为 `euretail` 的数据集。

a. 分解时间序列:

将时间序列进行分解并画出分解之后各模块的图。

b. Holt-Winters 指数平滑:

对时间序列 euretail 做 Holt-Winters 指数平滑并画图, 应得到以下的图:



注意: 如果画图的时候 legend 函数生成的图例说明太大, 可以在 legend 函数当中加上 `cex=0.8` 或者 `cex=0.5` 调整图例大小。

4. 数据文件 `hmohiv.csv` 记录了对艾滋病患者的观察记录, 其中 `time` 代表观察到的时间长度 (月), `status` 代表状态 (1=死亡, 0=数据删失), `age` 代表年龄, `drug` 代表有无吸毒史 (1=有, 0=无)。使用这个数据文件分析:
- 使用 Kaplan-Meier 方法画出生存函数。
 - 使用 Kaplan-Meier 方法分别画出有吸毒史和无吸毒史两组人群的生存函数, 并比较有吸毒史和无吸毒史两组人群的生存函数在 5% 显著性水平上是否有显著区别。

提交作业的截止时间: **6月24日周六晚上22:00**。

应该通过电子邮件提交到 `zachzhoucourse@hotmail.com`。提交作业的邮件标题是: “BA HW04 你的名字”。

提交文件列表:

- 对题目的解答, 将解答放在一个 Word 或者 PDF 文档里面, 将文档命名为 HW04-你的名字。
- 解答题目的完整代码。

3、第一题画中国地图时用到的销售数据文件,
但是不需要提供 `ProvIndex.csv` 和 `ProvMapData.csv`。

以上代码确保放在以下目录当中可以成功执行: `D:/BA/Homework/HW04`。如果你修改工作目录, 导致代码无法在上面的目录中执行, TA 将不会检查你的代码而是直接给你 0 分!