EPS贸易数据分析

胡华平

2020-12-24

Table of Contents

# 目前进展

数据收集进展情况：

1. EPS数据库。

* HS2002子库（2002-2011年）：正在下载中，一天内可以完成数据下载！
* HS2012子库（2012-2016年）：已经完成下载！可以进行数据分析。

1. 竹云贸易数据库（2017-2020）。受到数据可系统查询设定的约束，下载工作量较大，需要较长时间。

# HS四位编码农产品数量  
n\_prod <- 197  
# 国家数（包括总值）  
n\_country <- 243  
# 年份数（2017-2020年）  
n\_year <- 4  
# 每次操作需要的平均秒数  
n\_operation <- 5  
  
# 估算全部操作需要的时长为 小时数  
(time\_total <- (n\_prod\*n\_country \*n\_year \*n\_operation)/3600/24)

[1] 11

# 面临问题

# 文献讨论

理论模型：

实证模型：

操作步骤和主要过程

主要结论

# 数据分析

## 比较HS国家编码变动

## 合并HS2012数据

下面合并操作只需要运行一次即可。

all\_files <- list.files("../data/eps/import/HS2012")  
n <- length(all\_files)  
  
path\_dir <- here::here("manuscript","data","eps", "import", "HS2012")  
path\_file <- paste0(path\_dir, "/",all\_files)  
  
tbl\_hs2012 <- NULL  
for (i in 1:n) {  
 tbl\_tem <- read\_rds(file = path\_file[i])  
 tbl\_hs2012 <- bind\_rows(tbl\_hs2012, tbl\_tem)  
 print(paste0("成功合并到了第", i,"个文件。共有",n,"个文件！"))  
}  
  
# 写出文件  
write\_rds(tbl\_hs2012, "../data/eps/tbl-hs2012.rds")

## 比较HS产品编码变动

## HS2012数据描述性分析

读取已经合并好的数据，并进行简单清洗：

下面我们进行初步的数据汇总分析。

1.按产品和按月份，来统计加总进口额情况

这是一张表

表 1: 2012年1001产品分月进口总额

|  |  |  |
| --- | --- | --- |
| code\_prod | month | total |
| 1001 | 05 | 16764 |
| 1001 | 03 | 16233 |
| 1001 | 09 | 16232 |
| 1001 | 10 | 11094 |
| 1001 | 02 | 10836 |
| 1001 | 07 | 7978 |
| 1001 | 04 | 7647 |
| 1001 | 08 | 6800 |
| 1001 | 01 | 6673 |
| 1001 | 06 | 5620 |
| 1001 | 11 | 4064 |
| 1001 | 12 | 206 |

下面我们来做一张柱状图：

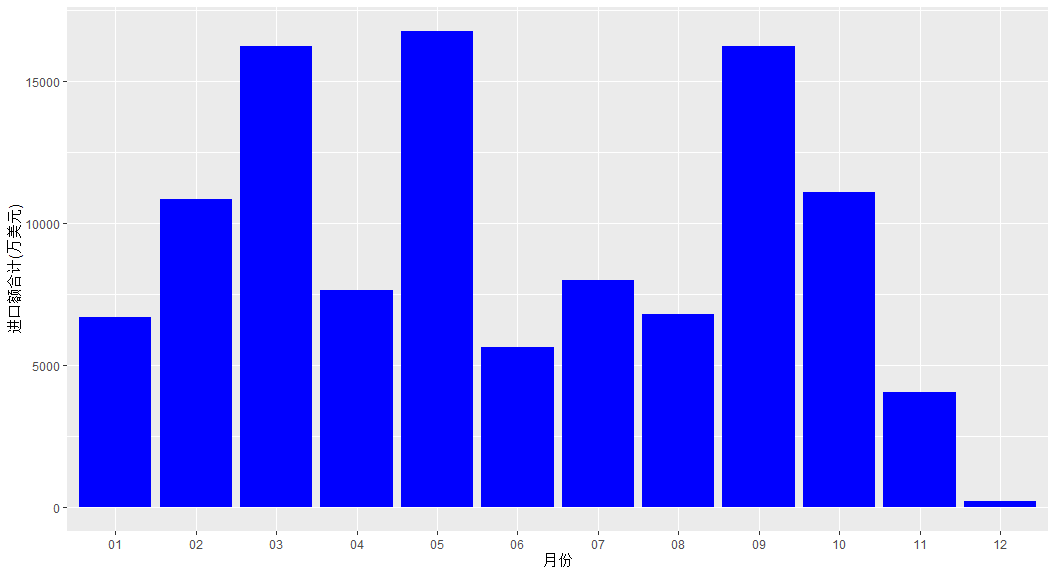


图 1: 一张丑丑的柱状图

上面的表1 和图1 共同表明，分月合计的产品进口贸易额呈现一定的双峰分布形态。

这是另外一张表。它表明什么社么

表 2: 2012年所有产品分月进口总额

|  |  |
| --- | --- |
| month | sum\_pro |
| 09 | 7920468768 |
| 07 | 7678319141 |
| 12 | 7602671762 |
| 05 | 7412775676 |
| 06 | 7258611228 |
| 08 | 7212983278 |
| 11 | 6946904120 |
| 03 | 6911726208 |
| 04 | 6801752643 |
| 10 | 6728003352 |
| 01 | 6010024070 |
| 02 | 5683366755 |

分析国家进出情况（有问题，未完成）。

tbl\_smry <- tbl\_hsclear %>%  
 ungroup() %>%  
 filter(year=="2012") %>%  
 filter(country !="总值") %>%  
 group\_by(code\_prod, month) %>%  
 summarise(n = dplyr::n(country))