

# text\_group

Wanning

2020/12/23

data\_all\_brand是1到11月的汇总数据，data\_search是一些品牌名称，需要再data\_all\_brand中匹配名字，计算广告收入。但是名称的文本很杂乱，不同名字对应的是一个品牌，需要进行文本处理再配对。

```
library(openxlsx)
```

```
## Warning: package 'openxlsx' was built under R version 4.0.3
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.3
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## √ ggplot2 3.3.2      √ purrr   0.3.4
## √ tibble  3.0.4      √ dplyr   1.0.2
## √ tidyr   1.1.2      √ stringr 1.4.0
## √ readr   1.4.0      √ forcats 0.5.0
```

```
## Warning: package 'ggplot2' was built under R version 4.0.3
```

```
## Warning: package 'tibble' was built under R version 4.0.3
```

```
## Warning: package 'tidyr' was built under R version 4.0.3
```

```
## Warning: package 'readr' was built under R version 4.0.3
```

```
## Warning: package 'purrr' was built under R version 4.0.3
```

```
## Warning: package 'dplyr' was built under R version 4.0.3
```

```
## Warning: package 'stringr' was built under R version 4.0.3
```

```
## Warning: package 'forcats' was built under R version 4.0.3
```

```
## -- Conflicts ----- tidyverse_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()
```

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 4.0.3
```

```
library(magrittr)
```

```
## Warning: package 'magrittr' was built under R version 4.0.3
```

```
##  
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:purrr':  
##  
## set_names
```

```
## The following object is masked from 'package:tidyr':  
##  
## extract
```

```
raw_data <- read.xlsx("D:\\liwanning6\\Downloads\\keke\\年度最佳大奖_候选.xlsx", startRow = 1)  
colnames(raw_data) <- c("index", "brand_name")  
  
data_all <- read_csv("D:\\liwanning6\\Downloads\\fanren\\bufanren_combine.csv")
```

```
##
## -- Column specification -----
## cols(
##   日期 = col_character(),
##   广告主pin = col_character(),
##   广告主ID = col_character(),
##   广告主名称 = col_character(),
##   站内外 = col_character(),
##   搜索标识 = col_character(),
##   品牌编码 = col_double(),
##   品牌名称 = col_character(),
##   一级部门 = col_character(),
##   二级部门 = col_character(),
##   三级部门 = col_character(),
##   `产品线(新)` = col_character(),
##   展现 = col_double(),
##   点击 = col_double(),
##   真钱消耗 = col_double(),
##   广告主下单订单行量 = col_double(),
##   广告主收订订单行量 = col_double(),
##   广告主收订订单金额 = col_double(),
##   广告主下单订单金额 = col_double()
## )
```

```
data_search <- raw_data %>%
  select(brand_name) %>%
  filter(!brand_name %in% c("整机业务部", "影像业务部", "智能网络与音频业务部", "电脑配件业务部", "文
仪及配件业务部")) %T>% print()
```

##	brand_name
## 1	联想
## 2	华为
## 3	戴尔
## 4	惠普
## 5	ThinkPad
## 6	华硕
## 7	小米
## 8	Apple
## 9	机械革命
## 10	宏碁
## 11	惠普
## 12	佳能
## 13	索尼
## 14	大疆
## 15	爱普生
## 16	极米
## 17	尼康
## 18	Nintendo Switch
## 19	任天堂
## 20	富士
## 21	华为
## 22	小米
## 23	普联
## 24	漫步者
## 25	索尼
## 26	Apple
## 27	小度
## 28	Bose
## 29	小天才
## 30	beats
## 31	西部数据
## 32	罗技
## 33	三星
## 34	华硕
## 35	希捷
## 36	AOC
## 37	英特尔
## 38	金士顿
## 39	AMD
## 40	闪迪
## 41	得力
## 42	TANGO
## 43	公牛
## 44	晨光
## 45	绿联
## 46	Kindle
## 47	小米
## 48	南孚
## 49	虎牌
## 50	飞利浦

```
data_all_brand <- data_all %>%
  select("日期", "品牌名称", "真钱消耗") %>%
  rename(dt = `日期`, brand_name = `品牌名称`, consumption = `真钱消耗`) %>%
  filter(dt >= "2020-01", dt <= "2020-11") %>%
  group_by(brand_name) %>%
  summarise(across(consumption, ~ sum(., na.rm = TRUE))) %>%
  ungroup() %T>% print()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 17,642 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 ..motomo        0.0571
## 2 ±0              0.669
## 3 0-one           2.99
## 4 0719            6196.
## 5 1=UND           501.
## 6 100% 必霸 (peakpower) 48268.
## 7 100FUN动手乐园      0
## 8 18098           1926.
## 9 1905            0
## 10 1982           0
## # ... with 17,632 more rows
```

grep: general regular expression。 ^: 以xx开头; ?: 有一个或者没有; .\*: 一个或多个字符。该循环目的是将公司的多种名字寻找出来, 每个按照list存储。

```
search_vec3 <- data_search$brand_name

emp_lst2 <- list()

for (brand_name in search_vec3) {
  if (brand_name == "富士") {
    emp_lst2[[brand_name]] <- data_all_brand[grepl(paste0(brand_name, " (", "FUJI", ".*"), data_all_brand$brand_name), ]
  } else {
    emp_lst2[[brand_name]] <- data_all_brand[grepl(paste0("^", brand_name, " (?", ".*"), data_all_brand$brand_name), ]
  }
}
emp_lst2
```

```
## $联想
## # A tibble: 11 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 联想            6504582.
## 2 联想 (IBM)       10.8
## 3 联想 (lenovo)    871401.
## 4 联想 (Lenovo)   212113644.
## 5 联想 (LENOVO)   453330.
## 6 联想 (ThinkCentre) 1027978.
## 7 联想 (ThinkServer) 8877.
## 8 联想 (ThinkStaion) 1886.
## 9 联想 (ThinkStation) 576857.
## 10 联想 (ThinkVision) 147841.
## 11 联想昭阳 (Lenovo) 358.
##
## $华为
## # A tibble: 4 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 华为            6871.
## 2 华为 (HUAWEI)  290378298.
## 3 华为视频        0
## 4 华为智选      1407968.
##
## $戴尔
## # A tibble: 3 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 戴尔            27108.
## 2 戴尔 (DELL)    96531141.
## 3 戴尔亚        0
##
## $惠普
## # A tibble: 4 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 惠普 (H P)      108.
## 2 惠普 (hp)     3790769.
## 3 惠普 (HP)     181817956.
## 4 惠普生 (NEWHOPSON) 39.1
##
## $ThinkPad
## # A tibble: 1 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 ThinkPad      79510234.
##
## $华硕
## # A tibble: 2 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 华硕            0
```

```
## 2 华硕 (ASUS)    109895860.
##
## $小米
## # A tibble: 5 x 2
##   brand_name consumption
##   <chr>          <dbl>
## 1 小米            370263.
## 2 小米 (MI)      51288048.
## 3 小米米家       313273.
## 4 小米涂乐        243.
## 5 小米有品       71767.
##
## $Apple
## # A tibble: 1 x 2
##   brand_name consumption
##   <chr>          <dbl>
## 1 Apple         4141387.
##
## $机械革命
## # A tibble: 1 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 机械革命 (MECHREVO)  20719523.
##
## $宏碁
## # A tibble: 2 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 宏碁 (acer)    28584685.
## 2 宏碁 (ACER)    4921.
##
## $佳能
## # A tibble: 4 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 佳能              0
## 2 佳能 (canon)    90390.
## 3 佳能 (Canon)   34853401.
## 4 佳能 (CANON)   3004527.
##
## $索尼
## # A tibble: 2 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 索尼            4131.
## 2 索尼 (SONY)    70294700.
##
## $大疆
## # A tibble: 2 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 大疆            229578.
## 2 大疆 (DJI)     25900300.
##
```

```
## $爱普生
## # A tibble: 3 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 爱普生          256.
## 2 爱普生 (epson) 327372.
## 3 爱普生 (EPSON) 13388845.
##
## $极米
## # A tibble: 2 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 极米 (XGIMI)    35449407.
## 2 极米熊          53839.
##
## $尼康
## # A tibble: 2 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 尼康            18456.
## 2 尼康 (Nikon)   14107732.
##
## $`Nintendo Switch`
## # A tibble: 1 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 Nintendo Switch 18601807.
##
## $任天堂
## # A tibble: 1 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 任天堂 (Nintendo) 3801483.
##
## $富士
## # A tibble: 2 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 富士 (FUJIFILM)   10449099.
## 2 富士 (FUJIIRYOKI) 0
##
## $普联
## # A tibble: 1 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 普联 (TP-LINK)    10860173.
##
## $漫步者
## # A tibble: 1 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 漫步者 (EDIFIER) 17561417.
##
## $小度
```



```
## # A tibble: 3 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 小度            38126493.
## 2 小度 (Xiao Du)      776.
## 3 小度在家          1375.
##
## $Bose
## # A tibble: 1 x 2
##   brand_name consumption
##   <chr>          <dbl>
## 1 Bose            31732911.
##
## $小天才
## # A tibble: 2 x 2
##   brand_name consumption
##   <chr>          <dbl>
## 1 小天才          31642232.
## 2 小天才酷盖       4447.
##
## $beats
## # A tibble: 1 x 2
##   brand_name consumption
##   <chr>          <dbl>
## 1 beats          13607099.
##
## $西部数据
## # A tibble: 2 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 西部数据 (WD)      13064999.
## 2 西部数据 (Western Digital) 33098.
##
## $罗技
## # A tibble: 4 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 罗技            221.
## 2 罗技 (G)        12125110.
## 3 罗技 (logitech)  130.
## 4 罗技 (Logitech)  7537432.
##
## $三星
## # A tibble: 4 x 2
##   brand_name      consumption
##   <chr>          <dbl>
## 1 三星            212.
## 2 三星 (3-star)      0
## 3 三星 (SAMSUNG)    28655360.
## 4 三星 (SHAN XING)    0
##
## $希捷
## # A tibble: 3 x 2
##   brand_name      consumption
```

```
## <chr> <dbl>
## 1 希捷 0
## 2 希捷 (Seagate) 57267.
## 3 希捷 (SEAGATE) 11716627.
##
## $AOC
## # A tibble: 2 x 2
## brand_name consumption
## <chr> <dbl>
## 1 AOC 23075610.
## 2 AOCG 0
##
## $英特尔
## # A tibble: 1 x 2
## brand_name consumption
## <chr> <dbl>
## 1 英特尔 (Intel) 30836031.
##
## $金士顿
## # A tibble: 2 x 2
## brand_name consumption
## <chr> <dbl>
## 1 金士顿 8191.
## 2 金士顿 (Kingston) 7758220.
##
## $AMD
## # A tibble: 1 x 2
## brand_name consumption
## <chr> <dbl>
## 1 AMD 8430374.
##
## $闪迪
## # A tibble: 2 x 2
## brand_name consumption
## <chr> <dbl>
## 1 闪迪 30882.
## 2 闪迪 (SanDisk) 8659309.
##
## $得力
## # A tibble: 3 x 2
## brand_name consumption
## <chr> <dbl>
## 1 得力 193845.
## 2 得力 (deli) 57026205.
## 3 得力佳 912.
##
## $TANGO
## # A tibble: 1 x 2
## brand_name consumption
## <chr> <dbl>
## 1 TANGO 1229029.
##
## $公牛
## # A tibble: 2 x 2
```

```

##   brand_name    consumption
##   <chr>         <dbl>
## 1 公牛          4.37e-4
## 2 公牛 (BULL)    1.75e+7
##
## $晨光
## # A tibble: 4 x 2
##   brand_name    consumption
##   <chr>         <dbl>
## 1 晨光          232524.
## 2 晨光 (M&G)    20366596.
## 3 晨光 (MG)     236607.
## 4 晨光文具 (M&G) 2345.
##
## $绿联
## # A tibble: 2 x 2
##   brand_name    consumption
##   <chr>         <dbl>
## 1 绿联          0.229
## 2 绿联 (UGREEN) 9125353.
##
## $Kindle
## # A tibble: 1 x 2
##   brand_name consumption
##   <chr>         <dbl>
## 1 Kindle       3512267.
##
## $南孚
## # A tibble: 2 x 2
##   brand_name consumption
##   <chr>         <dbl>
## 1 南孚          7102224.
## 2 南孚酷博        0
##
## $虎牌
## # A tibble: 4 x 2
##   brand_name    consumption
##   <chr>         <dbl>
## 1 虎牌          6620857.
## 2 虎牌 (TIGER)   39949.
## 3 虎牌保险柜      0
## 4 虎牌保险柜 (Tiger safe) 0
##
## $飞利浦
## # A tibble: 6 x 2
##   brand_name    consumption
##   <chr>         <dbl>
## 1 飞利浦          636637.
## 2 飞利浦 (PHILIP) 50.1
## 3 飞利浦 (philips) 293892.
## 4 飞利浦 (Philips) 6105.
## 5 飞利浦 (PHILIPS) 31309115.
## 6 飞利浦 (飞利浦 (PHILIPS) ) 0

```

嵌套的函数：先判断是否有括号，如果有则删去括号内容，如果没有则不做改变。"()"是capture group，意味着是提取的内容。这个函数带入的不是vector，而是单个的变量，于是用map\_chr批量调用。brand\_name\_lgl是判断括号外内容是否匹配，返回TRUE FALSE。

```
brand_name_check <- function(df, brand_name_std) {  
  
  brand_name_modify <- function(brand_name_element) {  
    if (str_detect(brand_name_element, " (")) {  
      brand_name_element <- str_match(brand_name_element, "(.*) (.*) ")[, 2]  
      return(brand_name_element)  
    } else {  
      return(brand_name_element)  
    }  
  }  
  
  brand_name_2 <- map_chr(df$brand_name, brand_name_modify)  
  df["brand_name_lgl"] <- brand_name_2 == brand_name_std  
  df["brand_name_2"] <- brand_name_2  
  
  return(df)  
}
```

map2的前两个变量带入后边的函数中

```
emp_lst2 <- map2(emp_lst2, names(emp_lst2), brand_name_check)  
  
emp_lst2
```

```
## $联想
## # A tibble: 11 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>      <chr>
## 1 联想            6504582. TRUE      联想
## 2 联想 (IBM)       10.8 TRUE      联想
## 3 联想 (lenovo)    871401. TRUE      联想
## 4 联想 (Lenovo)   212113644. TRUE      联想
## 5 联想 (LENOVO)   453330. TRUE      联想
## 6 联想 (ThinkCentre) 1027978. TRUE      联想
## 7 联想 (ThinkServer) 8877. TRUE      联想
## 8 联想 (ThinkStaion) 1886. TRUE      联想
## 9 联想 (ThinkStation) 576857. TRUE      联想
## 10 联想 (ThinkVision) 147841. TRUE      联想
## 11 联想昭阳 (Lenovo) 358. FALSE     联想昭阳
##
## $华为
## # A tibble: 4 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>      <chr>
## 1 华为            6871. TRUE      华为
## 2 华为 (HUAWEI)   290378298. TRUE      华为
## 3 华为视频          0 FALSE      华为视频
## 4 华为智选       1407968. FALSE      华为智选
##
## $戴尔
## # A tibble: 3 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>      <chr>
## 1 戴尔            27108. TRUE      戴尔
## 2 戴尔 (DELL)    96531141. TRUE      戴尔
## 3 戴尔亚          0 FALSE      戴尔亚
##
## $惠普
## # A tibble: 4 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>      <chr>
## 1 惠普 (H P)       108. TRUE      惠普
## 2 惠普 (hp)      3790769. TRUE      惠普
## 3 惠普 (HP)     181817956. TRUE      惠普
## 4 惠普生 (NEWHOPSON) 39.1 FALSE      惠普生
##
## $ThinkPad
## # A tibble: 1 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>      <chr>
## 1 ThinkPad     79510234. TRUE      ThinkPad
##
## $华硕
## # A tibble: 2 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>      <chr>
## 1 华硕          0 TRUE      华硕
```

```
## 2 华硕 (ASUS)    109895860. TRUE    华硕
##
## $小米
## # A tibble: 5 x 4
##   brand_name consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>          <chr>
## 1 小米              370263. TRUE    小米
## 2 小米 (MI)        51288048. TRUE    小米
## 3 小米米家          313273. FALSE   小米米家
## 4 小米涂乐           243. FALSE   小米涂乐
## 5 小米有品          71767. FALSE   小米有品
##
## $Apple
## # A tibble: 1 x 4
##   brand_name consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>          <chr>
## 1 Apple          4141387. TRUE    Apple
##
## $机械革命
## # A tibble: 1 x 4
##   brand_name          consumption brand_name_lgl brand_name_2
##   <chr>              <dbl> <lgl>          <chr>
## 1 机械革命 (MECHREVO)  20719523. TRUE    机械革命
##
## $宏碁
## # A tibble: 2 x 4
##   brand_name          consumption brand_name_lgl brand_name_2
##   <chr>              <dbl> <lgl>          <chr>
## 1 宏碁 (acer)        28584685. TRUE    宏碁
## 2 宏碁 (ACER)         4921. TRUE    宏碁
##
## $佳能
## # A tibble: 4 x 4
##   brand_name          consumption brand_name_lgl brand_name_2
##   <chr>              <dbl> <lgl>          <chr>
## 1 佳能                  0 TRUE    佳能
## 2 佳能 (canon)        90390. TRUE    佳能
## 3 佳能 (Canon)       34853401. TRUE    佳能
## 4 佳能 (CANON)       3004527. TRUE    佳能
##
## $索尼
## # A tibble: 2 x 4
##   brand_name          consumption brand_name_lgl brand_name_2
##   <chr>              <dbl> <lgl>          <chr>
## 1 索尼              4131. TRUE    索尼
## 2 索尼 (SONY)       70294700. TRUE    索尼
##
## $大疆
## # A tibble: 2 x 4
##   brand_name          consumption brand_name_lgl brand_name_2
##   <chr>              <dbl> <lgl>          <chr>
## 1 大疆              229578. TRUE    大疆
## 2 大疆 (DJI)       25900300. TRUE    大疆
##
```

```

## $爱普生
## # A tibble: 3 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>      <chr>
## 1 爱普生          256. TRUE      爱普生
## 2 爱普生 (epson) 327372. TRUE      爱普生
## 3 爱普生 (EPSON) 13388845. TRUE      爱普生
##
## $极米
## # A tibble: 2 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>      <chr>
## 1 极米 (XGIMI)    35449407. TRUE      极米
## 2 极米熊          53839. FALSE      极米熊
##
## $尼康
## # A tibble: 2 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>      <chr>
## 1 尼康          18456. TRUE      尼康
## 2 尼康 (Nikon)   14107732. TRUE      尼康
##
## $`Nintendo Switch`
## # A tibble: 1 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>      <chr>
## 1 Nintendo Switch 18601807. TRUE      Nintendo Switch
##
## $任天堂
## # A tibble: 1 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>      <chr>
## 1 任天堂 (Nintendo) 3801483. TRUE      任天堂
##
## $富士
## # A tibble: 2 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>      <chr>
## 1 富士 (FUJIFILM) 10449099. TRUE      富士
## 2 富士 (FUJIIRYOKI) 0 TRUE      富士
##
## $普联
## # A tibble: 1 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>      <chr>
## 1 普联 (TP-LINK) 10860173. TRUE      普联
##
## $漫步者
## # A tibble: 1 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>      <chr>
## 1 漫步者 (EDIFIER) 17561417. TRUE      漫步者
##
## $小度

```

```
## # A tibble: 3 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>          <chr>
## 1 小度            38126493. TRUE          小度
## 2 小度 (Xiao Du)    776. TRUE          小度
## 3 小度在家        1375. FALSE         小度在家
##
## $Bose
## # A tibble: 1 x 4
##   brand_name consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>          <chr>
## 1 Bose          31732911. TRUE          Bose
##
## $小天才
## # A tibble: 2 x 4
##   brand_name consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>          <chr>
## 1 小天才        31642232. TRUE          小天才
## 2 小天才酷盖    4447. FALSE         小天才酷盖
##
## $beats
## # A tibble: 1 x 4
##   brand_name consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>          <chr>
## 1 beats        13607099. TRUE          beats
##
## $西部数据
## # A tibble: 2 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>          <chr>
## 1 西部数据 (WD)    13064999. TRUE          西部数据
## 2 西部数据 (Western Digital) 33098. TRUE          西部数据
##
## $罗技
## # A tibble: 4 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>          <chr>
## 1 罗技            221. TRUE          罗技
## 2 罗技 (G)        12125110. TRUE          罗技
## 3 罗技 (logitech)  130. TRUE          罗技
## 4 罗技 (Logitech) 7537432. TRUE          罗技
##
## $三星
## # A tibble: 4 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lgl>          <chr>
## 1 三星            212. TRUE          三星
## 2 三星 (3-star)     0 TRUE          三星
## 3 三星 (SAMSUNG) 28655360. TRUE          三星
## 4 三星 (SHAN XING)  0 TRUE          三星
##
## $希捷
## # A tibble: 3 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
```



```

##      <chr>                <dbl> <lg1>          <chr>
## 1 希捷                    0  TRUE          希捷
## 2 希捷 (Seagate)         57267. TRUE          希捷
## 3 希捷 (SEAGATE)        11716627. TRUE        希捷
##
## $AOC
## # A tibble: 2 x 4
##   brand_name consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lg1>          <chr>
## 1 AOC           23075610. TRUE          AOC
## 2 AOCG           0  FALSE          AOCG
##
## $英特尔
## # A tibble: 1 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lg1>          <chr>
## 1 英特尔 (Intel)   30836031. TRUE        英特尔
##
## $金士顿
## # A tibble: 2 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lg1>          <chr>
## 1 金士顿           8191. TRUE        金士顿
## 2 金士顿 (Kingston) 7758220. TRUE        金士顿
##
## $AMD
## # A tibble: 1 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lg1>          <chr>
## 1 AMD           8430374. TRUE        AMD
##
## $闪迪
## # A tibble: 2 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lg1>          <chr>
## 1 闪迪           30882. TRUE        闪迪
## 2 闪迪 (SanDisk)   8659309. TRUE        闪迪
##
## $得力
## # A tibble: 3 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lg1>          <chr>
## 1 得力           193845. TRUE        得力
## 2 得力 (deli)     57026205. TRUE        得力
## 3 得力佳           912. FALSE        得力佳
##
## $TANGO
## # A tibble: 1 x 4
##   brand_name      consumption brand_name_lgl brand_name_2
##   <chr>          <dbl> <lg1>          <chr>
## 1 TANGO          1229029. TRUE        TANGO
##
## $公牛
## # A tibble: 2 x 4

```

```
## brand_name consumption brand_name_lgl brand_name_2
## <chr> <dbl> <lgl> <chr>
## 1 公牛 4.37e-4 TRUE 公牛
## 2 公牛 (BULL) 1.75e+7 TRUE 公牛
##
## $晨光
## # A tibble: 4 x 4
## brand_name consumption brand_name_lgl brand_name_2
## <chr> <dbl> <lgl> <chr>
## 1 晨光 232524. TRUE 晨光
## 2 晨光 (M&G) 20366596. TRUE 晨光
## 3 晨光 (MG) 236607. TRUE 晨光
## 4 晨光文具 (M&G) 2345. FALSE 晨光文具
##
## $绿联
## # A tibble: 2 x 4
## brand_name consumption brand_name_lgl brand_name_2
## <chr> <dbl> <lgl> <chr>
## 1 绿联 0.229 TRUE 绿联
## 2 绿联 (UGREEN) 9125353. TRUE 绿联
##
## $Kindle
## # A tibble: 1 x 4
## brand_name consumption brand_name_lgl brand_name_2
## <chr> <dbl> <lgl> <chr>
## 1 Kindle 3512267. TRUE Kindle
##
## $南孚
## # A tibble: 2 x 4
## brand_name consumption brand_name_lgl brand_name_2
## <chr> <dbl> <lgl> <chr>
## 1 南孚 7102224. TRUE 南孚
## 2 南孚酷博 0 FALSE 南孚酷博
##
## $虎牌
## # A tibble: 4 x 4
## brand_name consumption brand_name_lgl brand_name_2
## <chr> <dbl> <lgl> <chr>
## 1 虎牌 6620857. TRUE 虎牌
## 2 虎牌 (TIGER) 39949. TRUE 虎牌
## 3 虎牌保险柜 0 FALSE 虎牌保险柜
## 4 虎牌保险柜 (Tiger safe) 0 FALSE 虎牌保险柜
##
## $飞利浦
## # A tibble: 6 x 4
## brand_name consumption brand_name_lgl brand_name_2
## <chr> <dbl> <lgl> <chr>
## 1 飞利浦 636637. TRUE 飞利浦
## 2 飞利浦 (PHILIP) 50.1 TRUE 飞利浦
## 3 飞利浦 (philips) 293892. TRUE 飞利浦
## 4 飞利浦 (Philips) 6105. TRUE 飞利浦
## 5 飞利浦 (PHILIPS) 31309115. TRUE 飞利浦
## 6 飞利浦 (飞利浦 (PHILIPS) ) 0 FALSE 飞利浦 (飞利浦
```

do.call行是将emp\_lst2的list调用rbind, 返回df

```
search_df <-  
  do.call(rbind, emp_lst2) %>%  
  filter(brand_name_lgl) %>%  
  group_by(brand_name_2) %>%  
  summarize(consumption = sum(consumption)) %>%  
  ungroup() %>%  
  arrange(desc(consumption)) %T>% print()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 43 x 2  
##   brand_name_2 consumption  
##   <chr>          <dbl>  
## 1 华为          290385169.  
## 2 联想          221706407.  
## 3 惠普          185608833.  
## 4 华硕          109895860.  
## 5 戴尔          96558250.  
## 6 ThinkPad      79510234.  
## 7 索尼          70298831.  
## 8 得力          57220050.  
## 9 小米          51658311.  
## 10 小度          38127269.  
## # ... with 33 more rows
```

还有小的Bug, 比如晨光文具是属于晨光的, 但是被过滤掉了。