

R version 4.0.3 (2020-10-10) -- "Bunny-Wunnies Freak Out"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.73 (7892) x86_64-apple-darwin17.0]

[History restored from /Users/mia/.Rapp.history]

```
> #import libraries
> library(data.table)
data.table 1.13.6 using 1 threads (see ?getDTthreads). Latest news:
r-datatable.com
```

This installation of data.table has not detected OpenMP support. It should still work but in single-threaded mode.
This is a Mac. Please read <https://mac.r-project.org/openmp/>. Please engage with Apple and ask them for support. Check r-datatable.com for updates, and our Mac instructions here: <https://github.com/Rdatatable/data.table/wiki/Installation>. After several years of many reports of installation problems on Mac, it's time to gingerly point out that there have been no similar problems on Windows or Linux.

```
> library(ggplot2)
Error in library(ggplot2) : there is no package called 'ggplot2'
> library(ggmosaic)
Error in library(ggmosaic) : there is no package called 'ggmosaic'
> library(readr)
Error in library(readr) : there is no package called 'readr'
> setwd
function (dir)
.Internal(setwd(dir))
<bytecode: 0x7fed91f18f30>
<environment: namespace:base>
> purchase_behaviour <-
as.data.table(read.csv("QVI_purchase_behaviour.csv"))
> transaction_data <-
as.data.table(readxl::read_xlsx("QVI_transaction_data.xlsx"))
Warning message:
In file(con, "r") :
```

```

cannot open file '/var/db/timezone/zoneinfo/+VERSION': No such
file or directory
> str(purchase_behaviour)
Error in str(purchase_behaviour) : object 'purchase_behaviour' not
found
> str(purchase_behaviour)
Classes 'data.table' and 'data.frame':    72637 obs. of  3
variables:
 $ LYLTY_CARD_NBR : int  1000 1002 1003 1004 1005 1007 1009 1010
1011 1012 ...
 $ LIFESTAGE      : chr  "YOUNG SINGLES/COUPLES" "YOUNG SINGLES/
COUPLES" "YOUNG FAMILIES" "OLDER SINGLES/COUPLES" ...
 $ PREMIUM_CUSTOMER: chr  "Premium" "Mainstream" "Budget"
"Mainstream" ...
- attr(*, ".internal.selfref")=<externalptr>
> str(transaction_data)
Classes 'data.table' and 'data.frame':    264836 obs. of  8
variables:
 $ DATE          : num  43390 43599 43605 43329 43330 ...
 $ STORE_NBR     : num  1 1 1 2 2 4 4 4 5 7 ...
 $ LYLTY_CARD_NBR: num  1000 1307 1343 2373 2426 ...
 $ TXN_ID        : num  1 348 383 974 1038 ...
 $ PROD_NBR      : num  5 66 61 69 108 57 16 24 42 52 ...
 $ PROD_NAME     : chr  "Natural Chip          Compny SeaSalt175g"
"CCs Nacho Cheese    175g" "Smiths Crinkle Cut  Chips Chicken 170g"
"Smiths Chip Thinly  S/Cream&Onion 175g" ...
 $ PROD_QTY      : num  2 3 2 5 3 1 1 1 1 2 ...
 $ TOT_SALES     : num  6 6.3 2.9 15 13.8 5.1 5.7 3.6 3.9 7.2 ...
- attr(*, ".internal.selfref")=<externalptr>
> #change to date format
> transaction_date$DATE <- as.Date(transaction_date$DATE, origin =
"1899-12-30")
Error in as.Date(transaction_date$DATE, origin = "1899-12-30") :
  object 'transaction_date' not found
> transaction_data$DATE <- as.Date(transaction_data$DATE, origin =
"1899-12-30")
Error in as.Date(transaction_data$DATE, origin = "1899-12-30") :
  object 'transaction_data' not found
> transaction_data$DATE <- as.Date(transaction_data$DATE, origin =
"1899-12-30")
> str(transaction_data)
Classes 'data.table' and 'data.frame':    264836 obs. of  8
variables:
 $ DATE          : Date, format: "2018-10-17" "2019-05-14" ...
 $ STORE_NBR     : num  1 1 1 2 2 4 4 4 5 7 ...
 $ LYLTY_CARD_NBR: num  1000 1307 1343 2373 2426 ...
 $ TXN_ID        : num  1 348 383 974 1038 ...
 $ PROD_NBR      : num  5 66 61 69 108 57 16 24 42 52 ...
 $ PROD_NAME     : chr  "Natural Chip          Compny SeaSalt175g"
"CCs Nacho Cheese    175g" "Smiths Crinkle Cut  Chips Chicken 170g"
"Smiths Chip Thinly  S/Cream&Onion 175g" ...
 $ PROD_QTY      : num  2 3 2 5 3 1 1 1 1 2 ...
 $ TOT_SALES     : num  6 6.3 2.9 15 13.8 5.1 5.7 3.6 3.9 7.2 ...
- attr(*, ".internal.selfref")=<externalptr>

```

```

> #examine PROD_NAME
> transaction_data[, .N, PROD_NAME]
      PROD_NAME      N
1:  Natural Chip      1468
2:      CCs Nacho Cheese 1498
3:  Smiths Crinkle Cut Chips Chicken 1484
4:  Smiths Chip Thinly S/Cream&Onion 1473
5: Kettle Tortilla ChpsHny&Jlpno Chili 3296
---
110: Red Rock Deli Chikn&Garlic Aioli 1434
111: RRD SR Slow Rst      Pork Belly 1526
112: RRD Pc Sea Salt      1431
113: Smith Crinkle Cut Bolognese 1451
114: Doritos Salsa Mild  1472
> productWords <-
data.table(unlist(strsplit(unique(transaction_data[,
PROD_NAME]),"")))
>
> setnames(productWords, 'words')
> #remove all words with digits and special characters from product
words through grepl()
> library(stringr)
Error in library(stringr) : there is no package called 'stringr'
> library(stringi)
Error in library(stringi) : there is no package called 'stringi'
> #remove special characters
> productWords$Words <-
str_replace_all(productWords$words,"[[:punct:]]","")
Error in str_replace_all(productWords$words, "[[:punct:]]", "") :
could not find function "str_replace_all"
> install.packages("stringr", repos='http://cran.us.r-project.org')
also installing the dependency 'stringi'

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
stringi_1.5.3.tgz'
Content type 'application/x-gzip' length 13641892 bytes (13.0 MB)
=====
downloaded 13.0 MB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
stringr_1.4.0.tgz'
Content type 'application/x-gzip' length 210650 bytes (205 KB)
=====
downloaded 205 KB

The downloaded binary packages are in
      /var/folders/zp/6cgvvfm974d2t5v6c3hygnv40000gp/T//
Rtmp64WTvU/downloaded_packages

> library("stringr")
>
> productWords$Words <-
str_replace_all(productWords$words,"[[:punct:]]","")

```

```

> ## removing digits
> productWords$words <- str_replace_all(productWords$words,"[0-9]","")
> productWords$words <- str_replace_all(productWords$words,"[gG]","")
> ###count how many times a word appears
> wordsSep <- strsplit(productWords$words,"")
> words.freq <- table(unlist(wordsSep))
> ####sorting them by freq in order of highest to lowest freq
> words.freq <- as.data.frame(words.freq)
> words.freq <-[order(words.freq$Freq, decreasing = TRUE),]
Error: unexpected '[' in "words.freq <-["
> words.freq <- words.freq[order(words.freq$Freq, decreasing = T),]
> words.freq
  Var1 Freq
1      1215
18    i  234
12    e  228
37    s  168
9     C  161
39    t  161
23    l  159
35    r  150
4     a  143
27    n  139
16    h  136
29    o  135
38    S  102
31    p   66
41    u   53
25    m   52
8     c   42
21    k   38
3     &   37
10    d   32
36    R   32
40    T   32
47    y   28
11    D   27
32    P   26
30    O   22
45    W   22
7     B   20
44    w   17
26    M   16
22    K   13
6     b   11
48    z   11
14    f   10
43    V    9
15    F    8
17    H    8
28    N    8
42    v    8

```

```

2      /      7
24     L      7
19     I      5
46     x      4
5      A      3
13     E      3
20     J      3
34     Q      3
33     q      1
> #remove salsa
> transaction_data[, SALSA := grepl("salsa", tolower(PROD_NAME))]
> transaction_data <- transaction_data[SALSA == FALSE,][, SALSA :=
NULL]
> summary(transaction_data)
      DATE          STORE_NBR      LYLTY_CARD_NBR      TXN_ID
Min.   :2018-07-01   Min.     :  1.0   Min.     :  1000   Min.     :
1
 1st Qu.:2018-09-30   1st Qu.: 70.0   1st Qu.:  70015   1st Qu.:
67569
Median :2018-12-30   Median :130.0   Median : 130367   Median :
135183
Mean   :2018-12-30   Mean    :135.1   Mean    : 135531   Mean     :
135131
 3rd Qu.:2019-03-31   3rd Qu.:203.0   3rd Qu.: 203084   3rd Qu.:
202654
Max.   :2019-06-30   Max.     :272.0   Max.     :2373711   Max.     :
2415841
      PROD_NBR      PROD_NAME      PROD_QTY      TOT_SALES
Min.     :  1.00   Length:246742   Min.     :  1.000   Min.     :
1.700
 1st Qu.: 26.00   Class :character   1st Qu.:  2.000   1st Qu.:
5.800
Median : 53.00   Mode  :character   Median :  2.000   Median :
7.400
Mean   : 56.35                      Mean    :  1.908   Mean     :
7.321
 3rd Qu.: 87.00                      3rd Qu.:  2.000   3rd Qu.:
8.800
Max.   :114.00                      Max.     :200.000   Max.     :
650.000
> #outlier in PROD_QTY
> library(tidyverse)
Error in library(tidyverse) : there is no package called 'tidyverse'
> install(tidyverse)
Error in install(tidyverse) : could not find function "install"
> install.packages("tidyverse", repos='http://cran.us.r-
project.org')
also installing the dependencies 'rprojroot', 'pkgbuild', 'diffobj',
'rematch2', 'brio', 'desc', 'pkgload', 'praise', 'ps', 'waldo',
'colorspace', 'sys', 'base64enc', 'testthat', 'farver', 'labeling',
'munsell', 'RColorBrewer', 'viridisLite', 'askpass', 'processx',
'evaluate', 'highr', 'markdown', 'yaml', 'xfun', 'htmltools',
'tinytex', 'backports', 'generics', 'blob', 'DBI', 'tidyselect',
'withr', 'gtable', 'isoband', 'scales', 'curl', 'mime', 'openssl',

```

'clipr', 'BH', 'cpp11', 'callr', 'fs', 'knitr', 'rmarkdown',
'selectr', 'broom', 'dbplyr', 'dplyr', 'forcats', 'ggplot2',
'haven', 'httr', 'jsonlite', 'lubridate', 'modelr', 'purrr',
'readr', 'reprex', 'rstudioapi', 'rvest', 'tidyr', 'xml2'

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
rprojroot_2.0.2.tgz'

Content type 'application/x-gzip' length 96290 bytes (94 KB)

=====

downloaded 94 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
pkgbuild_1.2.0.tgz'

Content type 'application/x-gzip' length 140593 bytes (137 KB)

=====

downloaded 137 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
diffobj_0.3.3.tgz'

Content type 'application/x-gzip' length 992982 bytes (969 KB)

=====

downloaded 969 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
rematch2_2.1.2.tgz'

Content type 'application/x-gzip' length 44807 bytes (43 KB)

=====

downloaded 43 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
brio_1.1.1.tgz'

Content type 'application/x-gzip' length 36742 bytes (35 KB)

=====

downloaded 35 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
desc_1.2.0.tgz'

Content type 'application/x-gzip' length 285234 bytes (278 KB)

=====

downloaded 278 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
pkgload_1.1.0.tgz'

Content type 'application/x-gzip' length 150877 bytes (147 KB)

=====

downloaded 147 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
praise_1.0.0.tgz'

Content type 'application/x-gzip' length 16137 bytes (15 KB)

=====

downloaded 15 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/

```
ps_1.5.0.tgz'
Content type 'application/x-gzip' length 276257 bytes (269 KB)
=====
downloaded 269 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
waldo_0.2.3.tgz'
Content type 'application/x-gzip' length 70175 bytes (68 KB)
=====
downloaded 68 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
colorspace_2.0-0.tgz'
Content type 'application/x-gzip' length 2616761 bytes (2.5 MB)
=====
downloaded 2.5 MB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
sys_3.4.tgz'
Content type 'application/x-gzip' length 47291 bytes (46 KB)
=====
downloaded 46 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
base64enc_0.1-3.tgz'
Content type 'application/x-gzip' length 31695 bytes (30 KB)
=====
downloaded 30 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
testthat_3.0.1.tgz'
Content type 'application/x-gzip' length 2898430 bytes (2.8 MB)
=====
downloaded 2.8 MB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
farver_2.0.3.tgz'
Content type 'application/x-gzip' length 1829409 bytes (1.7 MB)
=====
downloaded 1.7 MB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
labeling_0.4.2.tgz'
Content type 'application/x-gzip' length 60296 bytes (58 KB)
=====
downloaded 58 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
munSELL_0.5.0.tgz'
Content type 'application/x-gzip' length 241504 bytes (235 KB)
=====
downloaded 235 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
```

```
RColorBrewer_1.1-2.tgz'
Content type 'application/x-gzip' length 52944 bytes (51 KB)
=====
downloaded 51 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
viridisLite_0.3.0.tgz'
Content type 'application/x-gzip' length 57005 bytes (55 KB)
=====
downloaded 55 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
askpass_1.1.tgz'
Content type 'application/x-gzip' length 21511 bytes (21 KB)
=====
downloaded 21 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
processx_3.4.5.tgz'
Content type 'application/x-gzip' length 279427 bytes (272 KB)
=====
downloaded 272 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
evaluate_0.14.tgz'
Content type 'application/x-gzip' length 74355 bytes (72 KB)
=====
downloaded 72 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
highr_0.8.tgz'
Content type 'application/x-gzip' length 40856 bytes (39 KB)
=====
downloaded 39 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
markdown_1.1.tgz'
Content type 'application/x-gzip' length 198809 bytes (194 KB)
=====
downloaded 194 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
yaml_2.2.1.tgz'
Content type 'application/x-gzip' length 203104 bytes (198 KB)
=====
downloaded 198 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
xfun_0.20.tgz'
Content type 'application/x-gzip' length 295823 bytes (288 KB)
=====
downloaded 288 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
```



```
htmltools_0.5.1.1.tgz'
Content type 'application/x-gzip' length 231294 bytes (225 KB)
=====
downloaded 225 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
tinytex_0.29.tgz'
Content type 'application/x-gzip' length 114329 bytes (111 KB)
=====
downloaded 111 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
backports_1.2.1.tgz'
Content type 'application/x-gzip' length 86201 bytes (84 KB)
=====
downloaded 84 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
generics_0.1.0.tgz'
Content type 'application/x-gzip' length 69334 bytes (67 KB)
=====
downloaded 67 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
blob_1.2.1.tgz'
Content type 'application/x-gzip' length 45899 bytes (44 KB)
=====
downloaded 44 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
DBI_1.1.1.tgz'
Content type 'application/x-gzip' length 670916 bytes (655 KB)
=====
downloaded 655 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
tidyselect_1.1.0.tgz'
Content type 'application/x-gzip' length 197492 bytes (192 KB)
=====
downloaded 192 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
withr_2.4.1.tgz'
Content type 'application/x-gzip' length 205896 bytes (201 KB)
=====
downloaded 201 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
gtable_0.3.0.tgz'
Content type 'application/x-gzip' length 431500 bytes (421 KB)
=====
downloaded 421 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
```

isoband_0.2.3.tgz'
Content type 'application/x-gzip' length 3683481 bytes (3.5 MB)
=====
downloaded 3.5 MB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
scales_1.1.1.tgz'
Content type 'application/x-gzip' length 552885 bytes (539 KB)
=====
downloaded 539 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
curl_4.3.tgz'
Content type 'application/x-gzip' length 741519 bytes (724 KB)
=====
downloaded 724 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
mime_0.9.tgz'
Content type 'application/x-gzip' length 35303 bytes (34 KB)
=====
downloaded 34 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
openssl_1.4.3.tgz'
Content type 'application/x-gzip' length 2863764 bytes (2.7 MB)
=====
downloaded 2.7 MB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
clipr_0.7.1.tgz'
Content type 'application/x-gzip' length 48725 bytes (47 KB)
=====
downloaded 47 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
BH_1.75.0-0.tgz'
Content type 'application/x-gzip' length 12149712 bytes (11.6 MB)
=====
downloaded 11.6 MB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
cpp11_0.2.6.tgz'
Content type 'application/x-gzip' length 199157 bytes (194 KB)
=====
downloaded 194 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
callr_3.5.1.tgz'
Content type 'application/x-gzip' length 385435 bytes (376 KB)
=====
downloaded 376 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/

```
fs_1.5.0.tgz'
Content type 'application/x-gzip' length 545211 bytes (532 KB)
=====
downloaded 532 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
knitr_1.31.tgz'
Content type 'application/x-gzip' length 1396805 bytes (1.3 MB)
=====
downloaded 1.3 MB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
rmarkdown_2.6.tgz'
Content type 'application/x-gzip' length 3596267 bytes (3.4 MB)
=====
downloaded 3.4 MB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
selectr_0.4-2.tgz'
Content type 'application/x-gzip' length 485312 bytes (473 KB)
=====
downloaded 473 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
broom_0.7.4.tgz'
Content type 'application/x-gzip' length 1776832 bytes (1.7 MB)
=====
downloaded 1.7 MB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
dbplyr_2.1.0.tgz'
Content type 'application/x-gzip' length 790791 bytes (772 KB)
=====
downloaded 772 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
dplyr_1.0.4.tgz'
Content type 'application/x-gzip' length 1250919 bytes (1.2 MB)
=====
downloaded 1.2 MB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
forcats_0.5.1.tgz'
Content type 'application/x-gzip' length 354051 bytes (345 KB)
=====
downloaded 345 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
ggplot2_3.3.3.tgz'
Content type 'application/x-gzip' length 4068756 bytes (3.9 MB)
=====
downloaded 3.9 MB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
```

```
haven_2.3.1.tgz'
Content type 'application/x-gzip' length 1131774 bytes (1.1 MB)
=====
downloaded 1.1 MB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
httr_1.4.2.tgz'
Content type 'application/x-gzip' length 498155 bytes (486 KB)
=====
downloaded 486 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
jsonlite_1.7.2.tgz'
Content type 'application/x-gzip' length 492071 bytes (480 KB)
=====
downloaded 480 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
lubridate_1.7.9.2.tgz'
Content type 'application/x-gzip' length 1608098 bytes (1.5 MB)
=====
downloaded 1.5 MB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
modelr_0.1.8.tgz'
Content type 'application/x-gzip' length 200721 bytes (196 KB)
=====
downloaded 196 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
purrr_0.3.4.tgz'
Content type 'application/x-gzip' length 417900 bytes (408 KB)
=====
downloaded 408 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
readr_1.4.0.tgz'
Content type 'application/x-gzip' length 2450519 bytes (2.3 MB)
=====
downloaded 2.3 MB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
reprex_1.0.0.tgz'
Content type 'application/x-gzip' length 478248 bytes (467 KB)
=====
downloaded 467 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
rstudioapi_0.13.tgz'
Content type 'application/x-gzip' length 271959 bytes (265 KB)
=====
downloaded 265 KB

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
```

```
rvest_0.3.6.tgz'
Content type 'application/x-gzip' length 1510048 bytes (1.4 MB)
=====
downloaded 1.4 MB
```

```
trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
tidyr_1.1.2.tgz'
Content type 'application/x-gzip' length 936440 bytes (914 KB)
=====
downloaded 914 KB
```

```
trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
xml2_1.3.2.tgz'
Content type 'application/x-gzip' length 2335007 bytes (2.2 MB)
=====
downloaded 2.2 MB
```

```
trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
tidyverse_1.3.0.tgz'
Content type 'application/x-gzip' length 433049 bytes (422 KB)
=====
downloaded 422 KB
```

The downloaded binary packages are in

```
/var/folders/zp/6cgvvfm974d2t5v6c3hygnv40000gp/T//
Rtmp64WTvU/downloaded_packages
> prod_qty_200 <- transaction_data %>% filter(PROD_QTY==200)
Error in filter(., PROD_QTY == 200) : object 'PROD_QTY' not found
> prod_qty_200 <- transaction_data %>% filter(PROD_QTY==200)
Error in filter(., PROD_QTY == 200) : object 'PROD_QTY' not found
> is_it_same_customer <- transaction_data %>% filter(LYLTY_CARD_NBR
== 226000)
Error in filter(., LYLTY_CARD_NBR == 226000) :
  object 'LYLTY_CARD_NBR' not found
>
> prod_qty_200 <- transaction_data %>%
filter(transaction_data$PROD_QTY, PROD_QTY==200)
Error in match.arg(method) : object 'PROD_QTY' not found
> #remove customer from list
> transaction_data <- transaction_data[!
(transaction_data$LYLTY_CARD_NBR == 226000)]
> summary(transaction_data)
```

| | DATE | STORE_NBR | LYLTY_CARD_NBR | TXN_ID |
|----------|-------------|---------------|-----------------|----------|
| Min. | :2018-07-01 | Min. : 1.0 | Min. : 1000 | Min. : |
| 1 | | | | |
| 1st Qu.: | 2018-09-30 | 1st Qu.: 70.0 | 1st Qu.: 70015 | 1st Qu.: |
| 67569 | | | | |
| Median : | 2018-12-30 | Median :130.0 | Median : 130367 | Median : |
| 135182 | | | | |
| Mean : | 2018-12-30 | Mean :135.1 | Mean : 135530 | Mean : |
| 135130 | | | | |
| 3rd Qu.: | 2019-03-31 | 3rd Qu.:203.0 | 3rd Qu.: 203083 | 3rd Qu.: |
| 202652 | | | | |

| | | | | | | | |
|----------|-------------|---------|-----------|----------|----------|----------|-----------|
| Max. | :2019-06-30 | Max. | :272.0 | Max. | :2373711 | Max. | : |
| 2415841 | | | | | | | |
| | PROD_NBR | | PROD_NAME | | PROD_QTY | | TOT_SALES |
| Min. | : 1.00 | Length: | 246740 | Min. | :1.000 | Min. | : 1.700 |
| 1st Qu.: | 26.00 | Class : | character | 1st Qu.: | 2.000 | 1st Qu.: | 5.800 |
| Median : | 53.00 | Mode : | character | Median : | 2.000 | Median : | 7.400 |
| Mean : | 56.35 | | | Mean : | 1.906 | Mean : | 7.316 |
| 3rd Qu.: | 87.00 | | | 3rd Qu.: | 2.000 | 3rd Qu.: | 8.800 |
| Max. | :114.00 | | | Max. | :5.000 | Max. | :29.500 |

```
> #count transactions by date
> countByDate <- count(transaction_data, transaction_data$DATE)
Error in count(transaction_data, transaction_data$DATE) :
  could not find function "count"
> install.packages("count", repos='http://cran.us.r-project.org')
Warning messages:
1: package 'count' is not available for this version of R
```

A version of this package for your version of R might be available elsewhere,

see the ideas at

<https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages>

2: Perhaps you meant 'COUNT' ?

```
> T
```

```
[1] TRUE
```

```
> install.packages("COUNT", repos='http://cran.us.r-project.org')
also installing the dependencies 'zoo', 'msme', 'sandwich'
```

```
trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
zoo_1.8-8.tgz'
```

```
Content type 'application/x-gzip' length 1080688 bytes (1.0 MB)
```

```
=====
```

```
downloaded 1.0 MB
```

```
trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
msme_0.5.3.tgz'
```

```
Content type 'application/x-gzip' length 247045 bytes (241 KB)
```

```
=====
```

```
downloaded 241 KB
```

```
trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
sandwich_3.0-0.tgz'
```

```
Content type 'application/x-gzip' length 1452451 bytes (1.4 MB)
```

```
=====
```

```
downloaded 1.4 MB
```

```
trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
COUNT_1.3.4.tgz'
```

```
Content type 'application/x-gzip' length 380569 bytes (371 KB)
```

```
=====
```

```
downloaded 371 KB
```

The downloaded binary packages are in

```

/var/folders/zp/6cgvvfm974d2t5v6c3hygnv40000gp/T//
Rtmp64WTVU/downloaded_packages
> countByDate <- COUNT(transaction_data, transaction_data$DATE)
Error in COUNT(transaction_data, transaction_data$DATE) :
  could not find function "COUNT"
> countByDate <- count(transaction_data, transaction_data$DATE)
Error in count(transaction_data, transaction_data$DATE) :
  could not find function "count"
> library(dplyr)

```

Attaching package: 'dplyr'

The following objects are masked from 'package:data.table':

between, first, last

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```

> countByDate <- count(transaction_data, transaction_data$DATE)
> countByDate
  transaction_data$DATE    n
1:      2018-07-01 663
2:      2018-07-02 650
3:      2018-07-03 674
4:      2018-07-04 669
5:      2018-07-05 660
---
360:    2019-06-26 657
361:    2019-06-27 669
362:    2019-06-28 673
363:    2019-06-29 703
364:    2019-06-30 704
> nrow(countByDate)
[1] 364
> summary(countByDate)
 transaction_data$DATE      n
Min.   :2018-07-01   Min.   :607.0
1st Qu.:2018-09-29   1st Qu.:658.0
Median :2018-12-30   Median :674.0
Mean   :2018-12-30   Mean   :677.9
3rd Qu.:2019-03-31   3rd Qu.:694.2
Max.   :2019-06-30   Max.   :865.0
> #missing date as there are only 364
> ##create sequence of dates and join the count of transactions by
date
> transaction_by_day <- transaction_data[order(DATE),]
> #setting plot themes to format graphs
> theme_set(theme_bw())

```

```
Error in theme_set(theme_bw()) : could not find function "theme_set"
> install.packages("ggplot2", repos='http://cran.us.r-project.org')
trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/
ggplot2_3.3.3.tgz'
Content type 'application/x-gzip' length 4068756 bytes (3.9 MB)
=====
downloaded 3.9 MB
```

```
The downloaded binary packages are in
      /var/folders/zp/6cgvvfm974d2t5v6c3hygnv40000gp/T//
Rtmp64WTvU/downloaded_packages
> theme_set(theme_bw())
Error in theme_set(theme_bw()) : could not find function "theme_set"
> library(theme_set)
Error in library(theme_set) : there is no package called 'theme_set'
> library(ggplot2)
> theme_set(theme_bw())
> theme_update(plot.title = element_text(hjust = 0.5))
> #plot transactions over time
> transOverTime <-ggplot(countByDate, aes(x =
countByDate$`transaction_data$DATE`, y = countByDate$n)) +
+   geom_line() +
+   labs(x = "Day", y = "Number of transactions", title =
"Transactions over time") +
+   scale_x_date(breaks = "1 month") +
+   theme(axis.text.x = element_text(angle = 90, vjust = 0.5))
> transOverTime
> #filter December as there is a spike into individual days
> filterData <- countByDate[countByDate$`transaction_data$DATE` >=
"2018-12-01" & countByDate$`transaction_data$DATE` <= "2018-12-31"]
> ggplot(filterData, aes(x = filterData$`transaction_data$DATE`, y =
filterData$n)) +
+   geom_line() +
+   labs(x = "Day", y = "Number of transactions", title =
"Transactions in December") +
+   scale_x_date(breaks = "1 day") +
+   theme(axis.text.x = element_text(angle = 90, vjust = 0.5))
> #no sales on christmas days
> #####PACK SIZE
> #add column PACK SIZE to data frame transaction_data
> ##### We can work this out by taking the digits that are in
PROD_NAME
> transaction_data[, PACK_SIZE := parse_number(PROD_NAME)]
Error in parse_number(PROD_NAME) : could not find function
"parse_number"
> transaction_data[, PACK_SIZE := parse_number(PROD_NAME)]
Error in parse_number(PROD_NAME) : could not find function
"parse_number"
>
> library(readr)
> library(ggmosaic)
Error in library(ggmosaic) : there is no package called 'ggmosaic'
> hist(transaction_data[, PACK_SIZE])
```



```

Error in `[.data.table`(transaction_data, , PACK_SIZE) :
  j (the 2nd argument inside [...]) is a single symbol but column
name 'PACK_SIZE' is not found. Perhaps you intended
DT[, ..PACK_SIZE]. This difference to data.frame is deliberate and
explained in FAQ 1.1.
>
> transaction_data[, PACK_SIZE]
Error in `[.data.table`(transaction_data, , PACK_SIZE) :
  j (the 2nd argument inside [...]) is a single symbol but column
name 'PACK_SIZE' is not found. Perhaps you intended
DT[, ..PACK_SIZE]. This difference to data.frame is deliberate and
explained in FAQ 1.1.
> ##BRANDS
> #Create a column which contains the brand of the product, by
extracting it from the product name
> transaction_data$BRAND <- gsub("([A-Za-z]+).*", "\\1",
transaction_data$PROD_NAME)
> transaction_data[, .N, by = BRAND][order(-N)]
Error: unexpected input in "transaction_data[, .N, by = BRAND]
[order(-"
> transaction_data[, .N, by = BRAND]
  BRAND      N
1:  Natural 6050
2:    CCs   4551
3:   Smiths 27390
4:   Kettle 41288
5:   Grain  6272
6:  Doritos 22041
7: Twisties  9454
8:    WW   10320
9:   Thins  14075
10:  Burger  1564
11:    NCC   1419
12: Cheezels  4603
13:  Infzns  3144
14:    Red   4427
15: Pringles 25102
16:  Dorito  3183
17: Infuzions 11057
18:   Smith   2963
19: GrnWves  1468
20: Tyrrells  6442
21:    Cobs  9693
22:   French  1418
23:    RRD  11894
24: Tostitos  9471
25:  Cheetos  2927
26: Woolworths 1516
27:    Snbts  1576
28: Sunbites  1432
  BRAND      N
> transaction_data[, .N, by = BRAND, order(-N)]
Error in `[.data.table`(transaction_data, , .N, by = BRAND, order(-
N)) :

```

```

Provide either by= or keyby= but not both
> #Clean brand names – combining names that are the same
> transaction_data[BRAND == "RED", BRAND:= "RRD"]
> transaction_data[BRAND == "SNBTS", BRAND := "SUNBITES"]
> transaction_data[BRAND == "INFZNS", BRAND := "INFUZIONI"]
> transaction_data[BRAND == "WW", BRAND := "WOOLWORTHS"]
> transaction_data[BRAND == "SMITH", BRAND := "SMITHS"]
>
> transaction_data[BRAND == "NCC", BRAND := "NATURAL"]
>
> transaction_data[BRAND == "DORITO", BRAND := "DORITOS"]
>
> transaction_data[BRAND == "GRAIN", BRAND := "GRNWVES"]
>
> transaction_data[, .N, by = BRAND]

```

```

      BRAND      N
1:   Natural 6050
2:      CCs 4551
3:    Smiths 27390
4:    Kettle 41288
5:     Grain 6272
6:   Doritos 22041
7: Twisties 9454
8: WOOLWORTHS 10320
9:     Thins 14075
10:   Burger 1564
11:  NATURAL 1419
12: Cheezels 4603
13:   Infzns 3144
14:     Red 4427
15: Pringles 25102
16:   Dorito 3183
17: Infuzions 11057
18:   Smith 2963
19: GrnWves 1468
20: Tyrrells 6442
21:    Cobs 9693
22:   French 1418
23:     RRD 11894
24: Tostitos 9471
25:  Cheetos 2927
26: Woolworths 1516
27:    Snbts 1576
28: Sunbites 1432

```

```

      BRAND      N

```

```

> #EXAMINE CUSTOMER DATA

```

```

> summary(purchase_behaviour)

```

| LYLTY_CARD_NBR | LIFESTAGE | PREMIUM_CUSTOMER |
|-----------------|------------------|------------------|
| Min. : 1000 | Length:72637 | Length:72637 |
| 1st Qu.: 66202 | Class :character | Class :character |
| Median : 134040 | Mode :character | Mode :character |
| Mean : 136186 | | |
| 3rd Qu.: 203375 | | |
| Max. : 2373711 | | |

```

> #look at lifestage and prem customer
> ##### Examining the values of lifestage and premium_customer
>
> purchase_behaviour[, .N, by = LIFESTAGE][order(-N)]
      LIFESTAGE      N
1:      RETIREES 14805
2:  OLDER SINGLES/COUPLES 14609
3:  YOUNG SINGLES/COUPLES 14441
4:      OLDER FAMILIES  9780
5:      YOUNG FAMILIES  9178
6: MIDAGE SINGLES/COUPLES  7275
7:      NEW FAMILIES  2549
> purchase_behaviour[, .N, by = LIFESTAGE][order(N)]
      LIFESTAGE      N
1:      NEW FAMILIES  2549
2: MIDAGE SINGLES/COUPLES  7275
3:      YOUNG FAMILIES  9178
4:      OLDER FAMILIES  9780
5:  YOUNG SINGLES/COUPLES 14441
6:  OLDER SINGLES/COUPLES 14609
7:      RETIREES 14805
> purchase_behaviour[, .N, by = PREMIUM_CUSTOMER][order(-N)]
      PREMIUM_CUSTOMER      N
1:      Mainstream 29245
2:      Budget 24470
3:      Premium 18922
> #Merge transaction data to customer data
> data <- merge(transaction_data, purchase_behaviour, all.x = TRUE)
> apply(data, function(x) any(is.na(x)))
Error in match.fun(FUN) : argument "FUN" is missing, with no default
> apply(data, 2, function(x) any(is.na(x)))
      LYLTY_CARD_NBR      DATE      STORE_NBR      TXN_ID
PROD_NBR
      FALSE      FALSE      FALSE      FALSE
FALSE
      PROD_NAME      PROD_QTY      TOT_SALES      BRAND
LIFESTAGE
      FALSE      FALSE      FALSE      FALSE
FALSE
PREMIUM_CUSTOMER
      FALSE
>
>
> ### DONE
> write.csv(data, "QVI_data.csv")
>
#####
#####
> #DATA ANALYSIS ON CUSTOMER SEGMENTS
> #Total sales by LIFESTAGE and PREMIUM_CUSTOMER
> total_sale <- data %>% group_by(LIFESTAGE, PREMIUM_CUSTOMER)
> pf.total_sale <- summarise(total_sale, sale_count=sum(TOT_SALES))
`summarise()` has grouped output by 'LIFESTAGE'. You can override

```

using the ``.groups`` argument.

```
> summary(pf.total_sale)
```

| LIFESTAGE | PREMIUM_CUSTOMER | sale_count |
|------------------|------------------|----------------|
| Length:21 | Length:21 | Min. : 10761 |
| Class :character | Class :character | 1st Qu.: 54444 |
| Mode :character | Mode :character | Median : 86338 |
| | | Mean : 85961 |
| | | 3rd Qu.:124648 |
| | | Max. :156864 |

```
> ###creat plot
```

```
> p <- ggplot(pf.total_sale) + geom_mosaic(aes(weight = sale_count, x =  
= product(PREMIUM_CUSTOMER, LIFESTAGE), fill = PREMIUM_CUSTOMER)) +  
labs(x = "Lifestage", y = "Premium customer flag", title =  
"Proportion of sales") + theme(axis.text.x = element_text(angle =  
90, vjust = 0.5))
```

```
Error in geom_mosaic(aes(weight = sale_count, x =  
product(PREMIUM_CUSTOMER, :  
could not find function "geom_mosaic"
```

```
> library(ggplot2)
```

```
> p <- ggplot(pf.total_sale) + geom_mosaic(aes(weight = sale_count, x =  
= product(PREMIUM_CUSTOMER, LIFESTAGE), fill = PREMIUM_CUSTOMER)) +  
labs(x = "Lifestage", y = "Premium customer flag", title =  
"Proportion of sales") + theme(axis.text.x = element_text(angle =  
90, vjust = 0.5))
```

```
Error in geom_mosaic(aes(weight = sale_count, x =  
product(PREMIUM_CUSTOMER, :  
could not find function "geom_mosaic"
```

```
> library(ggmosaic)
```

```
Error in library(ggmosaic) : there is no package called 'ggmosaic'
```

```
> ###sales come from budget - older families, mainstream - young  
singles/couples and mainstream - retirees
```

```
> #Number of customers by LIFESTAGE and PREMIUM_CUSTOMER
```

```
> total_sale <- data %>% group_by(LIFESTAGE, PREMIUM_CUSTOMER)
```

```
> no_of_customers <- summarise(total_sales, customer_count =  
length(unique(LYLT_CARD_NBR)))
```

```
Error in summarise(total_sales, customer_count =  
length(unique(LYLT_CARD_NBR))) :  
object 'total_sales' not found
```

```
> no_of_customers <- summarise(total_sale, customer_count =  
length(unique(LYLT_CARD_NBR)))
```

```
`summarise()` has grouped output by 'LIFESTAGE'. You can override  
using the `.groups` argument.
```

```
\
```

```
> summary(no_of_customers)
```

| LIFESTAGE | PREMIUM_CUSTOMER | customer_count |
|------------------|------------------|----------------|
| Length:21 | Length:21 | Min. : 575 |
| Class :character | Class :character | 1st Qu.:2369 |
| Mode :character | Mode :character | Median :3298 |
| | | Mean :3395 |
| | | 3rd Qu.:4611 |
| | | Max. :7917 |

```
>
```

```
> ##### Create plot
```

```
> p <- ggplot(data = no_of_customers) + geom_mosaic(aes(weight =
```

```

customer_count, x = product(PREMIUM_CUSTOMER, LIFESTAGE), fill =
PREMIUM_CUSTOMER)) + labs(x = "Lifestage", y = "Premium customer
flag", title = "Proportion of customers") + theme(axis.text.x =
element_text(angle = 90, vjust = 0.5))+ geom_text(data =
ggplot_build(p)$data[[1]], aes(x = (xmin + xmax)/2 , y = (ymin +
ymax)/2, label = as.character(paste(round(.wt/sum(.wt),3)*100,
'%'))))
Error in geom_mosaic(aes(weight = customer_count, x =
product(PREMIUM_CUSTOMER, :
could not find function "geom_mosaic"
>
>
> p
Error: object 'p' not found
> #Higher sales driven by more units of chips being bought per
customer
> total_sales_1 <- data %>% group_by(LIFESTAGE,PREMIUM_CUSTOMER)
>
> units <- summarise(total_sales_1, units_count = (sum(PROD_QTY)/
uniqueN(LYLTY_CARD_NBR)))
`summarise()` has grouped output by 'LIFESTAGE'. You can override
using the `.groups` argument.
> summary(units)
  LIFESTAGE          PREMIUM_CUSTOMER      units_count
Length:21          Length:21          Min.      :4.250
Class :character    Class :character    1st Qu.:4.892
Mode  :character    Mode  :character    Median :6.142
                                         Mean   :6.575
                                         3rd Qu.:8.638
                                         Max.   :9.255
> #older and young families buy more chips per customer
> ##average price per unit by LIFESTAGE and PREMIUM_CUSTOMER
> average_price <- data %>% group_by(LIFESTAGE,PREMIUM_CUSTOMER)
> pricePerUnit <- summarise(average_price, price_per_unit =
(sum(TOT_SALES)/sum(PROD_QTY)))
`summarise()` has grouped output by 'LIFESTAGE'. You can override
using the `.groups` argument.
> summary(pricePerUnit)
  LIFESTAGE          PREMIUM_CUSTOMER      price_per_unit
Length:21          Length:21          Min.      :3.685
Class :character    Class :character    1st Qu.:3.748
Mode  :character    Mode  :character    Median :3.823
                                         Mean   :3.833
                                         3rd Qu.:3.924
                                         Max.   :4.074
> ####plot
> ggplot(data=pricePerUnit, aes(weight = price_per_unit,x =
LIFESTAGE, fill = PREMIUM_CUSTOMER)) + geom_bar(position =
position_dodge()) + labs(x = "Lifestage", y = "Avg price per unit",
title = "Price per unit") + theme(axis.text.x = element_text(angle =
90, vjust = 0.5))
>
> #Mainstream midage and young singles and couples are more willing
to pay more per packet of chips compared to their budget and premium

```

```

counterparts.
> #difference in avg price per unit isnt large - check difference is
statistically different
> #Perform an independent t-test between mainstream vs premium and
budget midage and young singles and couples
> # If this p-value is above .05, then there is not a significant
difference in test scores.
> pricePerUnit <- data[, price:=TOT_SALES/PROD_QTY]
> t.test(data[LIFESTAGE %in% c("YOUNG SINGLES/COUPLES", "MIDAGE
SINGLES/COUPLES") & PREMIUM_CUSTOMER == "Mainstream",
price],data[LIFESTAGE %in% c("YOUNG SINGLES/COUPLES", "MIDAGE
SINGLES/COUPLES") & PREMIUM_CUSTOMER != "Mainstream", price],
alternative = "greater")

```

Welch Two Sample t-test

```

data: data[LIFESTAGE %in% c("YOUNG SINGLES/COUPLES", "MIDAGE
SINGLES/COUPLES") & PREMIUM_CUSTOMER == "Mainstream", price] and
data[LIFESTAGE %in% c("YOUNG SINGLES/COUPLES", "MIDAGE SINGLES/
COUPLES") & PREMIUM_CUSTOMER != "Mainstream", price]
t = 37.624, df = 54791, p-value < 2.2e-16
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.3187234      Inf
sample estimates:
mean of x mean of y
 4.039786  3.706491

```

```

> t test - p-value results the unit price for mainstream, young and
mid-age singles and couples ARE significantly higher than that of
budget or premium, young and midage singles and couples.

```

Error: unexpected symbol in "t test"

```

> #CUSTOMER SEGMENTS: MAINSTREAM, YOUNG SINGLES/COUPLES
> segment1 <- data[LIFESTAGE == "YOUNG SINGLES/COUPLES" &
PREMIUM_CUSTOMER == "Mainstream",]
> other <- data[!(LIFESTAGE == "YOUNG SINGLES/COUPLES" &
PREMIUM_CUSTOMER == "Mainstream"),]
> ##### Brand affinity compared to the rest of the population
> quantity_segment1 <- segment1[, sum(PROD_QTY)]
> quantity_other <- other[, sum(PROD_QTY)]
> quantity_segment1_by_brand <- segment1[, .(targetSegment =
sum(PROD_QTY)/quantity_segment1), by = BRAND]
>
> quantity_other_by_brand <- other[, .(other = sum(PROD_QTY)/
quantity_other), by = BRAND]
>
> brand_proportions <- merge(quantity_segment1_by_brand,
quantity_other_by_brand)[, affinityToBrand := targetSegment/other]
>
> brand_proportions[order(-affinityToBrand)]
Error: unexpected input in "brand_proportions[order(-"
>
> ggplot(brand_proportions,
aes(brand_proportions$BRAND,brand_proportions$affinityToBrand)) +

```

```

geom_bar(stat = "identity", fill = "yellow") + labs(x = "Brand", y =
"Customers Affinity to Brand", title = "Favorite brands of
Customers") + theme(axis.text.x = element_text(angle = 90, vjust =
0.5))
>
> ##Preferred pack size compared to the rest of the population
> quantity_segment1_by_pack <- segment1[, .(targetSegment =
sum(PROD_QTY)/quantity_segment1), by = PACK_SIZE]
Error in eval(bysub, parent.frame(), parent.frame()) :
  object 'PACK_SIZE' not found
>
> quantity_other_by_pack <- other[, .(other = sum(PROD_QTY)/
quantity_other), by = PACK_SIZE]
Error in eval(bysub, parent.frame(), parent.frame()) :
  object 'PACK_SIZE' not found
>
> pack_proportions <- merge(quantity_segment1_by_pack,
quantity_other_by_pack)[, affinityToPack := targetSegment/other]
Error in merge(quantity_segment1_by_pack, quantity_other_by_pack) :
  object 'quantity_segment1_by_pack' not found
>
> pack_proportions[order(-affinityToPack)]
Error: unexpected input in "pack_proportions[order(-"
>
>

```