

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]

```
> 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)
```

Warning message:

In file(con, "r") :

cannot open file '/var/db/timezone/zoneinfo/+VERSION': No such file or directory

```
> library(tidyr)
```

```
> install.packages("data.table", repos='http://cran.us.r-project.org')
```

trying URL 'http://cran.us.r-project.org/bin/macosx/contrib/4.0/data.table_1.13.6.tgz'

Content type 'application/x-gzip' length 2335683 bytes (2.2 MB)

=====

downloaded 2.2 MB

The downloaded binary packages are in

/var/folders/zp/6cgvvfm974d2t5v6c3hygnv40000gp/T//
RtmpugsbkQ/downloaded_packages

```
install.packages("gg[]", repos='http://cran.us.r-project.org')
>
> 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//
RtmpugsbKQ/downloaded_packages
> install.packages("tidyr", repos='http://cran.us.r-project.org')
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
```

```
The downloaded binary packages are in
      /var/folders/zp/6cgvvfm974d2t5v6c3hygnv40000gp/T//
RtmpugsbKQ/downloaded_packages
> ##### Point the filePath to where you have downloaded the datasets
to
> ##### assign the data files to data.tables
2021-02-09 21:17:00.673 R[40960:12497545] [QL] Can't get plugin
bundle info at file:///Applications/GarageBand.app/Contents/Library/
QuickLook/LogicXQLGenerator.qlgenerator/
> data <- as.data.table(read.csv("QVI_data.csv", header = TRUE))
> theme_set(theme_bw())
> theme_update(plot.title = element_text(hjust = 0.5))
> data
```

	LYLTY_CARD_NBR	DATE	STORE_NBR	TXN_ID	PROD_NBR
1:	1000	2018-10-17	1	1	5
2:	1002	2018-09-16	1	2	58
3:	1003	2019-03-07	1	3	52
4:	1003	2019-03-08	1	4	106
5:	1004	2018-11-02	1	5	96

264830:	2370701	2018-12-08	88	240378	24
264831:	2370751	2018-10-01	88	240394	60
264832:	2370961	2018-10-24	88	240480	70
264833:	2370961	2018-10-27	88	240481	65
264834:	2373711	2018-12-14	88	241815	16

```

      PROD_NAME  PROD_QTY  TOT_SALES
1:  Natural Chip      Compny SeaSalt175g      2      6.0
2:   Red Rock Deli Chikn&Garlic Aio1i 150g      1      2.7
3:   Grain Waves Sour   Cream&Chives 210G      1      3.6
4:  Natural ChipCo      Hony Soy Chckn175g      1      3.0
5:      WW Original Stacked Chips 160g      1      1.9
----
```

```

264830: Grain Waves Sweet Chilli 210g 2 7.2
264831: Kettle Tortilla ChpsFeta&Garlic 150g 2 9.2
264832: Tyrrells Crisps Lightly Salted 165g 2 8.4
264833: Old El Paso Salsa Dip Chnky Tom Ht300g 2 10.2
264834: Smiths Crinkle Chips Salt & Vinegar 330g 2 11.4
PACK_SIZE BRAND LIFESTAGE PREMIUM_CUSTOMER
1: 175 NATURAL YOUNG SINGLES/COUPLES Premium
2: 150 RRD YOUNG SINGLES/COUPLES Mainstream
3: 210 GRNWVES YOUNG FAMILIES Budget
4: 175 NATURAL YOUNG FAMILIES Budget
5: 160 WOOLWORTHS OLDER SINGLES/COUPLES Mainstream
---
264830: 210 GRNWVES YOUNG FAMILIES Mainstream
264831: 150 KETTLE YOUNG FAMILIES Premium
264832: 165 TYRRELLS OLDER FAMILIES Budget
264833: 300 OLD OLDER FAMILIES Budget
264834: 330 SMITHS YOUNG SINGLES/COUPLES Mainstream
> #Select control stores
> ##### Calculate these measures over time for each store
> Add a new month ID column in the data with the format yyyy-mm
Error: unexpected symbol in "Add a"
> #####Add a new month ID column in the data with the format yyyy-mm
> data[, N. , YEARMONTH:= ]
Error: unexpected '[' in "data[, N. , YEARMONTH:= ]"
> library(lubridate)

```

Attaching package: 'lubridate'

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

```

hour, isoweek, mday, minute, month, quarter, second, wday, week,
yday,
year

```

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

```

date, intersect, setdiff, union

```

```

> monthYear <- format(as.Date(date$DATE), "%Y%m")
Error in date$DATE : object of type 'closure' is not subsettable
> monthYear <- format(as.Date(data$DATE), "%Y%m")
> data[, YEARMONTH := monthYear]
> data$YEARMONTH <- as.numeric(as.character(data$YEARMONTH))
> ### Define measure calculations to use during analysis
> #####For each store and month calculate total sales, number of
customers, transactions per customer, chips per customer and the
average price per unit.
> measureOverTime <- data %>% group_by(STORE_NBR, YEARMONTH) %>%
summarise(totSales = sum(TOT_SALES), nCustomers =
uniqueN(LYLT_CARD_NBR), nTxnPerCust = uniqueN(TXN_ID)/
uniqueN(LYLT_CARD_NBR), nChipsPerTxn = sum(PROD_QTY)/
uniqueN(TXN_ID), avgPricePerUnit = (sum(TOT_SALES)/sum(PROD_QTY)))
Error in summarise(., totSales = sum(TOT_SALES), nCustomers =
uniqueN(LYLT_CARD_NBR), :

```

```

could not find function "summarise"
>
> measureOverTime <- data %>% group_by(STORE_NBR, YEARMONTH) %>%
summarise(totSales = sum(TOT_SALES), nCustomers =
uniqueN(LYLTY_CARD_NBR), nTxnPerCust = uniqueN(TXN_ID)/
uniqueN(LYLTY_CARD_NBR), nChipsPerTxn = sum(PROD_QTY)/
uniqueN(TXN_ID), avgPricePerUnit = (sum(TOT_SALES)/sum(PROD_QTY)))
Error in summarise(., totSales = sum(TOT_SALES), nCustomers =
uniqueN(LYLTY_CARD_NBR), :
could not find function "summarise"
> install.packages("dplyr", repos='http://cran.us.r-project.org')
Warning message:
package 'dplyr' 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>

```

>
> install.packages("dplyr", repos='http://cran.us.r-project.org')
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

```

```

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

```

> measureOverTime <- data %>% group_by(STORE_NBR, YEARMONTH) %>%
summarise(totSales = sum(TOT_SALES), nCustomers =
uniqueN(LYLTY_CARD_NBR), nTxnPerCust = uniqueN(TXN_ID)/
uniqueN(LYLTY_CARD_NBR), nChipsPerTxn = sum(PROD_QTY)/
uniqueN(TXN_ID), avgPricePerUnit = (sum(TOT_SALES)/sum(PROD_QTY)))
`summarise()` has grouped output by 'STORE_NBR'. You can override

```

```

using the `.groups` argument.
> ### Filter to the pretrial period and stores with full observation
periods
> storesWithFullObs <-
as.data.table(table(measureOverTime$STORE_NBR))
> storesWithFullObs <- storesWithFullObs %>% filter(N==12)
Error in filter(., N == 12) : object 'storesWithFullObs' not found
> storesWithFullObs <-
as.data.table(table(measureOverTime$STORE_NBR))
+ storesWithFullObs <- storesWithFullObs %>% filter(N==12)
Error: unexpected symbol in:
"storesWithFullObs <- as.data.table(table(measureOverTime$STORE_NBR)
storesWithFullObs"
> storesWithFullObs <-
as.data.table(table(measureOverTime$STORE_NBR))
+
+ h
Error: unexpected symbol in:
"
h"
> storesWithFullObs <-
as.data.table(table(measureOverTime$STORE_NBR))
> storesWithFullObs <- storesWithFullObs %>% filter(N==12)
> storesWithFullObs<-setNames(storesWithFullObs,c("STORE_NBR","N"))
> preTrialMeasures <- measureOverTime %>% filter(YEARMONTH <
201902,STORE_NBR %in% storesWithFullObs$STORE_NBR)
> #####Now we need to work out a way of ranking how similar
each potential control store is to the trial store. We can calculate
how correlated the performance of each store is to the trial store.
>
> ##### Create a function to calculate correlation for a measure,
looping through each control store.
>
> #FOR SALES
> trialStore_sales <- preTrialMeasures %>% filter(STORE_NBR ==77)
> trialStore_sales <- trialStore_sales %>%
select(STORE_NBR,YEARMONTH,totSales,nCustomers)
> calCorr <- function(preTrialMeasures, trialStore_sales,
trialStoreN){
+
+   calTable = data.table(Store1 = numeric(), Store2 = numeric(),
corr_measure = numeric())
+ stN <- preTrialMeasures %>% select(STORE_NBR)
+ for(i in stN$STORE_NBR){
+
+   contSt <- preTrialMeasures %>% filter(STORE_NBR==i)
+   contSt <- contSt %>% select(totSales)
+ calMeasure = data.table("Store1" = trialStoreN, "Store2" = i,
"corr_measure" = cor(trialStore_sales$totSales,contSt$totSales))
+
+   calTable <- rbind(calTable, calMeasure) }
+   return(calTable)
+ }
> #####FOR CUSTOMERS

```

```

> calculateCorrelation <-
function(preTrialMeasures,trialStore_sales,trialStoreN){
+
+   calTable = data.table(Store1 = numeric(), Store2 = numeric(),
corr_measure = numeric())
+
+   stN <- preTrialMeasures %>% select(STORE_NBR)
+
+   for(i in stN$STORE_NBR){
+
+     contSt <- preTrialMeasures %>% filter(STORE_NBR==i)
+     contSt <- contSt %>% select(nCustomers)
+
+     calMeasure = data.table("Store1" = trialStoreN, "Store2" = i,
"corr_measure" = cor(trialStore_sales$nCustomers,contSt$nCustomers))
+
+     calTable <- rbind(calTable, calMeasure) }
+   return(calTable)
+ }
> ##### Create a function to calculate a standardised magnitude
distance for a measure, looping through each control store
> ##Sales
> calculateMagnitudeDistance1 <-
function(preTrialMeasures,trialStore_sales,trial_storeN){
+   calTable = data.table(Store1 = numeric(), Store2 = numeric(),
YEARMONTH = numeric(),mag_measure = numeric())
+   stN <- preTrialMeasures %>% select(STORE_NBR)
+   for(i in stN$STORE_NBR){
+     contSt <- preTrialMeasures %>% filter(STORE_NBR==i)
+     contSt <- contSt %>% select(totSales)
+     calMeasure = data.table("Store1" = trial_storeN, "Store2" = i,
"YEARMONTH" = preTrialMeasures$YEARMONTH ,"mag_measure" =
abs(trialStore_sales$totSales - contSt$totSales))
+
+     calTable <- rbind(calTable,calMeasure)
+     calTable <- unique(calTable)
+   }
+   return(calTable)
+ }
>
> #####Standardize
> standMag1 <- function(magnitude_nSales) {
+   minMaxDist <- magnitude_nSales[, .(minDist =
min( magnitude_nSales$mag_measure), maxDist =
max(magnitude_nSales$mag_measure)), by = c("Store1", "YEARMONTH")]
+   distTable <- merge(magnitude_nSales, minMaxDist, by =
c("Store1", "YEARMONTH"))
+   distTable[, magnitudeMeasure := 1 - (mag_measure - minDist)/
(maxDist - minDist)]
+   finalDistTable <- distTable[, .(magN_measure =
mean(magnitudeMeasure)), by = .(Store1, Store2)]
+   return(finalDistTable)
+ }
> #####Customers

```

```

> calculateMagnitudeDistance2 <-
function(preTrialMeasures,trialStore_sales,trial_storeN){
+   calTable = data.table(Store1 = numeric(), Store2 = numeric(),
YEARMONTH = numeric(),mag_measure = numeric())
+   stN <- preTrialMeasures %>% select(STORE_NBR)
+   for(i in stN$STORE_NBR){
+     contSt <- preTrialMeasures %>% filter(STORE_NBR==i)
+     contSt <- contSt %>% select(nCustomers)
+     calMeasure = data.table("Store1" = trial_storeN, "Store2" = i,
"YEARMONTH" = preTrialMeasures$YEARMONTH ,"mag_measure" =
abs(trialStore_sales$nCustomers - contSt$nCustomers))
+
+     calTable <- rbind(calTable,calMeasure)
+     calTable <- unique(calTable)
+   }
+   return(calTable)
+ }
> ##Standardize
> standMag2 <- function(magnitude_nSales) {
+ +   minMaxDist <- magnitude_nCustomers[, .(minDist =
min( magnitude_nCustomers$mag_measure), maxDist =
max(magnitude_nCustomers$mag_measure)), by = c("Store1",
"YEARMONTH")]
+ +   distTable <- merge(magnitude_nCustomers, minMaxDist, by =
c("Store1", "YEARMONTH"))
+ +   distTable[, magnitudeMeasure := 1 - (mag_measure - minDist)/
(maxDist - minDist)]
+ +   finalDistTable <- distTable[, .(magN_measure =
mean(magnitudeMeasure)), by = .(Store1, Store2)]
+ +   return(finalDistTable)
+ + }
Error: unexpected '}' in:
"+   return(finalDistTable)
+ }"
> standMag2 <- function(magnitude_nCustomers) {
+   minMaxDist <- magnitude_nCustomers[, .(minDist =
min( magnitude_nCustomers$mag_measure), maxDist =
max(magnitude_nCustomers$mag_measure)), by = c("Store1",
"YEARMONTH")]
+   distTable <- merge(magnitude_nCustomers, minMaxDist, by =
c("Store1", "YEARMONTH"))
+   distTable[, magnitudeMeasure := 1 - (mag_measure - minDist)/
(maxDist - minDist)]
+   finalDistTable <- distTable[, .(magN_measure =
mean(magnitudeMeasure)), by = .(Store1, Store2)]
+   return(finalDistTable)
+ }
> ##### Use the function you created to calculate correlations
against store 77 using total sales and number of customers.
>
> trial_store <- 77
> corr_nSales <- calculateCorrelation(preTrialMeasures,
trialStore_sales, trialStoreN)
Adding missing grouping variables: `STORE_NBR`

```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
Adding missing grouping variables: `STORE_NBR`  
  
>  
> magnitude_nSales <- standMag1(magnitude_nSales)  
>  
> magnitude_nCustomers <-  
calculateMagnitudeDistance2(preTrialMeasures,trialStore_sales,  
trial_store)  
Error: unexpected input in "magnitude_nCustomers <-"  
>  
> magnitude_nCustomers <-  
calculateMagnitudeDistance2(preTrialMeasures,trialStore_sales,  
trial_store)  
Error: unexpected input in "magnitude_nCustomers <-"  
>  
> magnitude_nCustomers <- standMag2(magnitude_nCustomers)  
Error in standMag2(magnitude_nCustomers) :  
object 'magnitude_nCustomers' not found  
>  
> magnitude_nCustomers  
Error: object 'magnitude_nCustomers' not found  
> corr_weight <- 0.5  
>  
> score_nSales <- merge(corr_nSales,magnitude_nSales, by =  
c("Store1", "Store2"))  
> score_nSales <- score_nSales %>% mutate(scoreNSales =  
(score_nSales$corr_measure * corr_weight)+(score_nSales$magnN_measure  
* (1 - corr weight)))
```

```

> score_nCustomers <- merge(corr_nCustomers,magnitude_nCustomers, by
= c("Store1", "Store2"))
Error in is.data.table(y) : object 'magnitude_nCustomers' not found
> score_nCustomers <- score_nCustomers %>% mutate(scoreNCust =
(score_nCustomers$corr_measure * corr_weight)+
(score_nCustomers$magnN_measure * (1 - corr_weight)))
Error in mutate(., scoreNCust = (score_nCustomers$corr_measure *
corr_weight) + :
  object 'score_nCustomers' not found
>
>

```