R Programming Introduction to Workshop Participants

Pramod Verma, Data For Excellence (DFE), GPE, PSA International

05 March 2020

Use of this document

This file is created for the sole use of PSA Data Analytics Technical Workshop participants for demonstration and learning about the R programming language.

Scripting language used

This document is created using R Markdown, a scripting language available as open source from R Foundation.

Loading all the required packages

```
install.packages("readxl",repos = "http://cran.us.r-project.org")
```

```
## package 'readxl' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\pramodkv\AppData\Local\Temp\Rtmpcxp3mh\downloaded_packages
```

```
library(readxl)
```

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

```
## package 'dplyr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\pramodkv\AppData\Local\Temp\Rtmpcxp3mh\downloaded_packages
```

```
library(dplyr)
```

May need to load more libraries/packages depending on local computer/server

Loading the file into R data-frame

```
#Reading the csv file
port = read.csv('D:/Data Analytics/Data Analytics Workshop/Data Analytics
Technical Workshop/Version 2/Data/port.csv')

#Reading the excel file
#port_excel = read_excel('D:/Data Analytics Workshop/Data Analytics Technical
Workshop Singapore/Data/port.xlsx')
```

Analysing the dataset

```
#Looding at the dimension of the dataset (number of rows and columns)
dim(port)
```

```
## [1] 49993    26
```

```
#Looding at the class of the dataset
class(port)
```

```
## [1] "data.frame"
```

```
#Looding at the names of the column of the dataset
names(port)
```

```
##  [1] "ID"                 "CNTR_N"             "CNTR_OPR_C"
##  [4] "CNTR_STATUS_C"      "PURPOSE_C"          "LENGTH_Q"
##  [7] "CNTR_TYPE_C"        "WT_Q"               "LOAD_PORT_C"
## [10] "DISC_PORT1_C"       "DEST_PORT_C"        "DISC_ABBR_VESSEL_M"
## [13] "DISC_SERVICE_C"     "LOAD_ABBR_VESSEL_M" "LOAD_SERVICE_C"
## [16] "LOAD_DT"            "DISC_DT"            "GATE_OUT_DT"
## [19] "GATE_IN_DT"         "BATCH_ID"           "LOGISTICS_PARTNER"
## [22] "DWELL_DAYS"         "PRODUCT"            "COMMODITY"
## [25] "CARGO_OWNER"        "RF_TEMP"
```

```
#Looding at the first few rows
head(port)
```

```
##   ID       CNTR_N CNTR_OPR_C CNTR_STATUS_C PURPOSE_C LENGTH_Q CNTR_TYPE_C
WT_Q
## 1  1 NYKU 9701818         ON             F         T       20          GP
22080
## 2  2 APHU 6602779         EG             F         T       40          GP
22932
## 3  3 PCIU 6007321         PI             F         T       40          RF
15200
## 4  4 HALU 5642133         HE             F         I       40          GP
27900
## 5  5 TEMU 1697343         YM             F         T       20          GP
15227
## 6  6 UACU 3529931         HL             F         T       20          GP
25000
##   LOAD_PORT_C DISC_PORT1_C DEST_PORT_C DISC_ABBR_VESSEL_M DISC_SERVICE_C
## 1       MYPGU        INCCU                    PAC BANDA            PGU
## 2       CNSHA        IDPWG                CMA TANCREDI             ASC
## 3       VNCLI        AUMEL       AUMEL
## 4       CNTXG                                AGAMEMNON            CIS
## 5       THBKK        INMUN       INMUN        MATHU BHUM            TSB
## 6       MYPKG        TRIST                  HONOLULU BR            CIS
##   LOAD_ABBR_VESSEL_M LOAD_SERVICE_C      LOAD_DT      DISC_DT  GATE_OUT_DT
## 1          DONG HAI            CC1 2.018081e+13 2.018080e+13           NA
## 2        MARCO P 118            IDO 2.018072e+13 2.018072e+13 2.018072e+13
```

```
## 3       O KUALALUMPU            AAA 2.018070e+13           NA           NA
## 4                                               NA 2.018092e+13 2.018092e+13
## 5        CONTI CHAMP            CIS 2.018122e+13 2.018121e+13           NA
## 6          HAMBURG E            MD3 2.018093e+13 2.018092e+13           NA
##     GATE_IN_DT BATCH_ID LOGISTICS_PARTNER DWELL_DAYS  PRODUCT  COMMODITY
## 1           NA   201808              <NA>        3.1   Weever   Sea Food
## 2 2.018072e+13   201807              <NA>        7.6    Snail   Sea Food
## 3           NA   201807              <NA>        4.3    Snail   Sea Food
## 4           NA   201809              <NA>        2.7  Sorubim   Sea Food
## 5           NA   201812              <NA>        2.8  pumpkin Vegetables
## 6           NA   201809              <NA>        9.6 delicata Vegetables
##                              CARGO_OWNER RF_TEMP
## 1                                Renault      NA
## 2                              Mitsubishi      NA
## 3                              Mitsubishi       8
## 4                    Magna International      NA
## 5                                 Nissan      NA
## 6 Aviation Industry Corporation of China      NA
```

```
#Looking at all the columns and their types
str(port)
```

```
## 'data.frame':    49993 obs. of  26 variables:
##  $ ID                : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ CNTR_N            : Factor w/ 49839 levels "0526237","AAMU 2604706",..:
28161 235 32621 16162 41583 47444 35444 25528 33586 47145 ...
##  $ CNTR_OPR_C        : Factor w/ 155 levels "AA","AB","AC",..: 99 41 106
61 151 62 119 88 88 62 ...
##  $ CNTR_STATUS_C     : Factor w/ 2 levels "E","F": 2 2 2 2 2 2 2 1 2 1 ...
##  $ PURPOSE_C         : Factor w/ 4 levels "I","P","T","X": 3 3 3 1 3 3 3 3
3 3 ...
##  $ LENGTH_Q          : int  20 40 40 40 20 20 20 40 20 40 ...
##  $ CNTR_TYPE_C       : Factor w/ 7 levels "FR","GP","OT",..: 2 2 5 2 2 2 2
2 2 2 ...
##  $ WT_Q              : int  22080 22932 15200 27900 15227 25000 17880 3880
26433 3660 ...
##  $ LOAD_PORT_C       : Factor w/ 433 levels "","AEAUH","AEDXB",..: 270 86
420 91 367 271 266 188 275 216 ...
##  $ DISC_PORT1_C      : Factor w/ 348 levels "","AEAUH","AEDXB",..: 123 115
16 1 130 273 131 264 191 68 ...
##  $ DEST_PORT_C       : Factor w/ 948 levels "","AEAJM","AEAUH",..: 1 1 27
1 421 1 1 1 1 1 ...
##  $ DISC_ABBR_VESSEL_M: Factor w/ 1587 levels "","A IDEFIX",..: 1260 284 1
11 813 538 1373 578 581 1524 ...
##  $ DISC_SERVICE_C    : Factor w/ 261 levels "","A5C","AA1",..: 170 24 1 52
242 52 210 46 195 185 ...
##  $ LOAD_ABBR_VESSEL_M: Factor w/ 1587 levels "","A IDEFIX",..: 423 779
1211 1 306 528 1175 948 842 1298 ...
##  $ LOAD_SERVICE_C    : Factor w/ 264 levels "","888","A5C",..: 43 99 6 1
53 138 185 170 223 64 ...
```

```
##  $ LOAD_DT           : num  2.02e+13 2.02e+13 2.02e+13 NA 2.02e+13 ...
##  $ DISC_DT           : num  2.02e+13 2.02e+13 NA 2.02e+13 2.02e+13 ...
##  $ GATE_OUT_DT       : num  NA 2.02e+13 NA 2.02e+13 NA ...
##  $ GATE_IN_DT        : num  NA 2.02e+13 NA NA NA ...
##  $ BATCH_ID          : int  201808 201807 201807 201809 201812 201809
201809 201810 201810 201807 ...
##  $ LOGISTICS_PARTNER : Factor w/ 49 levels "Americold Logistics",..: NA NA
NA NA NA NA NA NA NA NA ...
##  $ DWELL_DAYS        : num  3.1 7.6 4.3 2.7 2.8 9.6 3.5 4.3 6.6 3.5 ...
##  $ PRODUCT           : Factor w/ 979 levels "Abalone","Abalone-style
Shellfish",..: 957 801 801 814 679 261 192 801 668 625 ...
##  $ COMMODITY         : Factor w/ 25 levels "Adhesive","Biscuit",..: 22 22
22 22 24 24 22 22 22 24 ...
##  $ CARGO_OWNER       : Factor w/ 101 levels "AB Volvo","Airbus",..: 81 64
64 62 70 7 35 64 26 43 ...
##  $ RF_TEMP           : int  NA NA 8 NA NA NA NA NA NA NA ...

#Looding at the summary of the entire dataset
summary(port)

##        ID                 CNTR_N        CNTR_OPR_C    CNTR_STATUS_C
PURPOSE_C
##  Min.   :    1   TRHU 1605377:    3   MD     : 7393   E: 7449       I:
2837
##  1st Qu.:12499   AAMU 7000400:    2   ON     : 6089   F:42544       P:
40
##  Median :24997   APHU 6410966:    2   MS     : 5398
T:43568
##  Mean   :24997   APZU 3041450:    2   HL     : 4086                 X:
3548
##  3rd Qu.:37495   APZU 4245377:    2   CM     : 3996
##  Max.   :49993   BEAU 2376015:    2   PI     : 3469
##                  (Other)     :49980   (Other):19562
##     LENGTH_Q     CNTR_TYPE_C      WT_Q        LOAD_PORT_C     DISC_PORT1_C
##  Min.   : 0.00   FR:  143    Min.   : 1730   THLCH  : 2580          : 2484
##  1st Qu.:20.00   GP:45829    1st Qu.: 9400          : 2269   THLCH  : 1960
##  Median :40.00   OT:  301    Median :21156   BDCGP  : 2098   CNSHA  : 1664
##  Mean   :30.62   PF:    3    Mean   :18409   IDJKT  : 1483   IDJKT  : 1592
##  3rd Qu.:40.00   RF: 3021    3rd Qu.:27224   CNSHA  : 1405   BDCGP  : 1351
##  Max.   :45.00   TK:  693    Max.   :60000   IDSUB  :  999   ZZZZZ  : 1012
##                  UC:    3                    (Other):39159   (Other):39930
##   DEST_PORT_C       DISC_ABBR_VESSEL_M DISC_SERVICE_C
LOAD_ABBR_VESSEL_M
##          :22497                :10020           :10020                : 9491
##  IDJKT   :  905   NYK JOANNA   :  193    AM1    :  705   MSC RAFAELA :  187
##  BDCGP   :  875   KANWAY GALAX:  157    AM4    :  694   NYK FUSHIMI :  156
##  ZZZZZ   :  766   ATOUT        :  154    IDO    :  613   NYK JOANNA  :  150
##  CNSHA   :  717   SINAR SABANG:  152    UE1    :  581   ATOUT       :  142
##  THLCH   :  672   NYK FUJI     :  148    SW3    :  578   BALTIMORE BR:  141
##  (Other):23561   (Other)      :39169    (Other):36802   (Other)     :39726
```

```
##  LOAD_SERVICE_C     LOAD_DT             DISC_DT           GATE_OUT_DT
##         : 9492   Min.   :2.018e+13   Min.   :2.018e+13   Min.   :2.018e+13
##  AM4    :  851   1st Qu.:2.018e+13   1st Qu.:2.018e+13   1st Qu.:2.018e+13
##  AM1    :  682   Median :2.018e+13   Median :2.018e+13   Median :2.018e+13
##  AM3    :  640   Mean   :2.018e+13   Mean   :2.018e+13   Mean   :2.018e+13
##  FE1    :  591   3rd Qu.:2.018e+13   3rd Qu.:2.018e+13   3rd Qu.:2.018e+13
##  EU1    :  580   Max.   :2.018e+13   Max.   :2.018e+13   Max.   :2.018e+13
##  (Other):37157   NA's   :9474        NA's   :9922        NA's   :40535
##    GATE_IN_DT           BATCH_ID
##  Min.   :2.018e+13   Min.   :201807
##  1st Qu.:2.018e+13   1st Qu.:201808
##  Median :2.018e+13   Median :201809
##  Mean   :2.018e+13   Mean   :201809
##  3rd Qu.:2.018e+13   3rd Qu.:201811
##  Max.   :2.018e+13   Max.   :201812
##  NA's   :40687
##                                     LOGISTICS_PARTNER   DWELL_DAYS
##  UPS Supply Chain Solutions                :  283   Min.   : 2.000
##  Lineage Logistics                         :  257   1st Qu.: 4.000
##  Performance Team                          :  234   Median : 6.000
##  Ingram Micro Commerce & Lifecycle Services:  232   Mean   : 6.014
##  Total Quality Logistics                   :  209   3rd Qu.: 8.000
##  (Other)                                   : 1997   Max.   :10.000
##  NA's                                      :46781
##     PRODUCT           COMMODITY                   CARGO_OWNER
##  Snail  :11397   Sea Food  :35329   Mitsubishi          :11550
##  Crab   :  680   Vegetables: 4710   Dongfeng Motor Group:  792
##  Clam   :  484   Fruits    : 3612   Dell                :  747
##  bean   :  416   Flowers   : 2651   Eni                 :  663
##  Grunt  :  316   Meat      : 1834   Lockheed Martin     :  642
##  (Other):36687   (Other)   : 1844   (Other)             :35586
##  NA's   :   13   NA's      :   13   NA's                :   13
##     RF_TEMP
##  Min.   :-30.00
##  1st Qu.:-20.00
##  Median :-10.00
##  Mean   :-10.23
##  3rd Qu.:  0.00
##  Max.   : 10.00
##  NA's   :46972
```

Creating Variables in R

```
#Simple Variables
x = 5
y = 'Singapore Terminal'

x

## [1] 5
```

```
y
```

```
## [1] "Singapore Terminal"
```

```
# List

a = c(2,3,4,5,8,9)
b = c('Singapore', 'Korea', 'India')
c = rnorm(20)


a
```

```
## [1] 2 3 4 5 8 9
```

```
b
```

```
## [1] "Singapore" "Korea"     "India"
```

```
c
```

```
##  [1]  2.2380139 -0.6758954  1.1034148  0.4127818 -0.7596593  1.7737347
##  [7]  0.9964348  0.4626976  0.9898442  1.7574141  0.5659058  2.0885404
## [13] -1.7541962 -1.3259091 -0.0653206 -1.0131702 -1.0346080  1.3264618
## [19] -0.5713265  0.1899724
```

Basic Data Preparation Activities in R

Selecting Variables

```
# Take the selected columns in new dataset
newdata = select(port, CNTR_N, COMMODITY, CARGO_OWNER)
head(newdata)
```

```
##         CNTR_N  COMMODITY                                CARGO_OWNER
## 1 NYKU 9701818   Sea Food                                    Renault
## 2 APHU 6602779   Sea Food                                  Mitsubishi
## 3 PCIU 6007321   Sea Food                                  Mitsubishi
## 4 HALU 5642133   Sea Food                       Magna International
## 5 TEMU 1697343 Vegetables                                     Nissan
## 6 UACU 3529931 Vegetables Aviation Industry Corporation of China
```

```
# keep the container number and all variables between logistics partner and
cargo owner
newdata = select(port, CNTR_N, LOGISTICS_PARTNER:CARGO_OWNER)
head(newdata)
```

```
##         CNTR_N LOGISTICS_PARTNER DWELL_DAYS  PRODUCT  COMMODITY
## 1 NYKU 9701818              <NA>        3.1   Weever   Sea Food
## 2 APHU 6602779              <NA>        7.6    Snail   Sea Food
## 3 PCIU 6007321              <NA>        4.3    Snail   Sea Food
## 4 HALU 5642133              <NA>        2.7  Sorubim   Sea Food
```

```
## 5 TEMU 1697343              <NA>        2.8  pumpkin Vegetables
## 6 UACU 3529931              <NA>        9.6 delicata Vegetables
##                              CARGO_OWNER
## 1                                Renault
## 2                             Mitsubishi
## 3                             Mitsubishi
## 4                    Magna International
## 5                                 Nissan
## 6 Aviation Industry Corporation of China

# keep the container number and all variables except reefer temperature and
batch id
newdata = select(port, -BATCH_ID, -RF_TEMP)
head(newdata)

##   ID       CNTR_N CNTR_OPR_C CNTR_STATUS_C PURPOSE_C LENGTH_Q CNTR_TYPE_C
WT_Q
## 1  1 NYKU 9701818         ON             F         T       20          GP
22080
## 2  2 APHU 6602779         EG             F         T       40          GP
22932
## 3  3 PCIU 6007321         PI             F         T       40          RF
15200
## 4  4 HALU 5642133         HE             F         I       40          GP
27900
## 5  5 TEMU 1697343         YM             F         T       20          GP
15227
## 6  6 UACU 3529931         HL             F         T       20          GP
25000
##   LOAD_PORT_C DISC_PORT1_C DEST_PORT_C DISC_ABBR_VESSEL_M DISC_SERVICE_C
## 1       MYPGU        INCCU                     PAC BANDA            PGU
## 2       CNSHA        IDPWG                   CMA TANCREDI           ASC
## 3       VNCLI        AUMEL       AUMEL
## 4       CNTXG                                 AGAMEMNON            CIS
## 5       THBKK        INMUN       INMUN        MATHU BHUM            TSB
## 6       MYPKG        TRIST                   HONOLULU BR            CIS
##   LOAD_ABBR_VESSEL_M LOAD_SERVICE_C      LOAD_DT      DISC_DT  GATE_OUT_DT
## 1           DONG HAI            CC1 2.018081e+13 2.018080e+13           NA
## 2        MARCO P 118            IDO 2.018072e+13 2.018072e+13 2.018072e+13
## 3       O KUALALUMPU            AAA 2.018070e+13           NA           NA
## 4                                              NA 2.018092e+13 2.018092e+13
## 5        CONTI CHAMP            CIS 2.018122e+13 2.018121e+13           NA
## 6          HAMBURG E            MD3 2.018093e+13 2.018092e+13           NA
##     GATE_IN_DT LOGISTICS_PARTNER DWELL_DAYS  PRODUCT  COMMODITY
## 1           NA              <NA>        3.1   Weever   Sea Food
## 2 2.018072e+13              <NA>        7.6    Snail   Sea Food
## 3           NA              <NA>        4.3    Snail   Sea Food
## 4           NA              <NA>        2.7  Sorubim   Sea Food
## 5           NA              <NA>        2.8  pumpkin Vegetables
## 6           NA              <NA>        9.6 delicata Vegetables
```

```
##                               CARGO_OWNER
## 1                                 Renault
## 2                              Mitsubishi
## 3                              Mitsubishi
## 4                    Magna International
## 5                                 Nissan
## 6 Aviation Industry Corporation of China
```

Selecting Observations

```
# filter for the selected record in the dataset
newdata = filter(port, CNTR_N=='TEXU 1046448')
head(newdata)
```

```
##    ID       CNTR_N CNTR_OPR_C CNTR_STATUS_C PURPOSE_C LENGTH_Q CNTR_TYPE_C
## 1 985 TEXU 1046448         CM             F         T       20          OT
##    WT_Q LOAD_PORT_C DISC_PORT1_C DEST_PORT_C DISC_ABBR_VESSEL_M
DISC_SERVICE_C
## 1 28300       CNTAO        REPDG       YTLON           APL LION
EU5
##   LOAD_ABBR_VESSEL_M LOAD_SERVICE_C      LOAD_DT      DISC_DT GATE_OUT_DT
## 1       NYK DANIELLA            MXP 2.018093e+13 2.018092e+13          NA
##   GATE_IN_DT BATCH_ID LOGISTICS_PARTNER DWELL_DAYS PRODUCT COMMODITY
## 1         NA   201809              <NA>        6.5 Oilfish  Sea Food
##   CARGO_OWNER RF_TEMP
## 1       Denso      NA

newdata = filter(port, CARGO_OWNER=='Airbus')
head(newdata)
```

```
##    ID       CNTR_N CNTR_OPR_C CNTR_STATUS_C PURPOSE_C LENGTH_Q CNTR_TYPE_C
## 1  14 CXDU 1636132         ON             F         T       20          GP
## 2 192 TGHU 4987564         HL             E         T       40          GP
## 3 225 CRSU 1225190         OU             F         X       20          GP
## 4 350 CAIU 9035123         RC             F         I       40          GP
## 5 401 OOLU 1386683         OR             E         T       20          GP
## 6 408 GESU 6499627         MD             F         X       40          GP
##    WT_Q LOAD_PORT_C DISC_PORT1_C DEST_PORT_C DISC_ABBR_VESSEL_M
DISC_SERVICE_C
## 1 21500       JPUKB        FRFOS                   MOL EMERALD
JSM
## 2  3680       AUSYD        CNSHA                     AL RAWDAH
NEN
## 3 18438                    VNHPH       VNHPH
## 4  7940       THLCH                                 MOL GLOBE
HLX
## 5  2280       INCCU        THLCH
## 6 14267       DKAAR        MYTPP       NLRTM       MANCHESTE MS
LP1
##   LOAD_ABBR_VESSEL_M LOAD_SERVICE_C      LOAD_DT      DISC_DT  GATE_OUT_DT
## 1       MANHATTAN BR            MD2 2.018101e+13 2.018101e+13           NA
```

```
## 2          YM WINNER            MD3 2.018113e+13 2.018113e+13           NA
## 3          KOTA JATI            VCS 2.018113e+13           NA           NA
## 4                                                 NA 2.018123e+13 2.018123e+13
## 5        O KAOHSIUNG            AAA 2.018070e+13           NA 2.018070e+13
## 6        HOPE ISLAND            MAE 2.018121e+13 2.018120e+13 2.018120e+13
##     GATE_IN_DT BATCH_ID LOGISTICS_PARTNER DWELL_DAYS    PRODUCT  COMMODITY
## 1           NA   201810              <NA>        8.0 topinambur Vegetables
## 2           NA   201811              <NA>        3.9  Snakehead   Sea Food
## 3 2.018113e+13   201811     DSC Logistics        6.1   Chimaera   Sea Food
## 4           NA   201812              <NA>        6.7       Swai   Sea Food
## 5 2.018070e+13   201807              <NA>        8.9    Herring   Sea Food
## 6 2.018121e+13   201812  U.S. Xpress Inc.        5.0       Goby   Sea Food
##   CARGO_OWNER RF_TEMP
## 1      Airbus      NA
## 2      Airbus      NA
## 3      Airbus      NA
## 4      Airbus      NA
## 5      Airbus      NA
## 6      Airbus      NA
```

```
# Multiple type of operators can be used to filter records
newdata = filter(port, is.na(RF_TEMP))
head(newdata)
```

```
##   ID       CNTR_N CNTR_OPR_C CNTR_STATUS_C PURPOSE_C LENGTH_Q CNTR_TYPE_C
WT_Q
## 1  1 NYKU 9701818         ON             F         T       20          GP
22080
## 2  2 APHU 6602779         EG             F         T       40          GP
22932
## 3  4 HALU 5642133         HE             F         I       40          GP
27900
## 4  5 TEMU 1697343         YM             F         T       20          GP
15227
## 5  6 UACU 3529931         HL             F         T       20          GP
25000
## 6  7 SIKU 2947904         SA             F         T       20          GP
17880
##   LOAD_PORT_C DISC_PORT1_C DEST_PORT_C DISC_ABBR_VESSEL_M DISC_SERVICE_C
## 1       MYPGU        INCCU                     PAC BANDA            PGU
## 2       CNSHA        IDPWG                 CMA TANCREDI            ASC
## 3       CNTXG                                 AGAMEMNON            CIS
## 4       THBKK        INMUN       INMUN         MATHU BHUM            TSB
## 5       MYPKG        TRIST                   HONOLULU BR            CIS
## 6       MYKUA        INNSA                 SINAR BANDUN            SGX
##   LOAD_ABBR_VESSEL_M LOAD_SERVICE_C      LOAD_DT      DISC_DT  GATE_OUT_DT
## 1           DONG HAI            CC1 2.018081e+13 2.018080e+13           NA
## 2        MARCO P 118            IDO 2.018072e+13 2.018072e+13 2.018072e+13
## 3                                               NA 2.018092e+13 2.018092e+13
## 4        CONTI CHAMP            CIS 2.018122e+13 2.018121e+13           NA
```

```
## 5          HAMBURG E            MD3 2.018093e+13 2.018092e+13           NA
## 6         NYK ATHENA            PS3 2.018093e+13 2.018092e+13           NA
##     GATE_IN_DT BATCH_ID LOGISTICS_PARTNER DWELL_DAYS        PRODUCT
COMMODITY
## 1           NA   201808             <NA>        3.1         Weever   Sea
Food
## 2 2.018072e+13   201807             <NA>        7.6          Snail   Sea
Food
## 3           NA   201809             <NA>        2.7        Sorubim   Sea
Food
## 4           NA   201812             <NA>        2.8        pumpkin
Vegetables
## 5           NA   201809             <NA>        9.6       delicata
Vegetables
## 6           NA   201809             <NA>        3.5 Climbing Perch   Sea
Food
##                              CARGO_OWNER RF_TEMP
## 1                             Renault      NA
## 2                            Mitsubishi      NA
## 3                    Magna International      NA
## 4                              Nissan      NA
## 5 Aviation Industry Corporation of China      NA
## 6                              FAW Group      NA

newdata = filter(port, WT_Q > 20000)
head(newdata)

##   ID       CNTR_N CNTR_OPR_C CNTR_STATUS_C PURPOSE_C LENGTH_Q CNTR_TYPE_C
WT_Q
## 1  1 NYKU 9701818         ON             F         T       20          GP
22080
## 2  2 APHU 6602779         EG             F         T       40          GP
22932
## 3  4 HALU 5642133         HE             F         I       40          GP
27900
## 4  6 UACU 3529931         HL             F         T       20          GP
25000
## 5  9 PONU 0165367         MS             F         T       20          GP
26433
## 6 12 TCLU 3225478         OR             F         T       20          GP
22458
##   LOAD_PORT_C DISC_PORT1_C DEST_PORT_C DISC_ABBR_VESSEL_M DISC_SERVICE_C
## 1       MYPGU        INCCU                     PAC BANDA            PGU
## 2       CNSHA        IDPWG                  CMA TANCREDI            ASC
## 3       CNTXG                                  AGAMEMNON            CIS
## 4       MYPKG        TRIST                   HONOLULU BR            CIS
## 5       MYTPP        MYWSP                  IRENES WARWI            SAE
## 6       IDBUN        TWKHH       CNFOC       MARCOPOLO298            IDO
##   LOAD_ABBR_VESSEL_M LOAD_SERVICE_C      LOAD_DT      DISC_DT  GATE_OUT_DT
## 1           DONG HAI            CC1 2.018081e+13 2.018080e+13           NA
```

```
## 2        MARCO P 118            IDO 2.018072e+13 2.018072e+13 2.018072e+13
## 3                                                NA 2.018092e+13 2.018092e+13
## 4          HAMBURG E            MD3 2.018093e+13 2.018092e+13           NA
## 5       MERATUS MEDA            SSS 2.018102e+13 2.018102e+13           NA
## 6       OOCL JAKARTA            KT3 2.018110e+13 2.018110e+13 2.018110e+13
##     GATE_IN_DT BATCH_ID LOGISTICS_PARTNER DWELL_DAYS  PRODUCT  COMMODITY
## 1           NA   201808              <NA>        3.1   Weever   Sea Food
## 2 2.018072e+13   201807              <NA>        7.6    Snail   Sea Food
## 3           NA   201809              <NA>        2.7  Sorubim   Sea Food
## 4           NA   201809              <NA>        9.6 delicata Vegetables
## 5           NA   201810              <NA>        6.6    Porgy   Sea Food
## 6 2.018110e+13   201811              <NA>        7.0     Cusk   Sea Food
##                               CARGO_OWNER RF_TEMP
## 1                                 Renault      NA
## 2                              Mitsubishi      NA
## 3                    Magna International      NA
## 4 Aviation Industry Corporation of China      NA
## 5                              Continental      NA
## 6                                     Dell      NA
```

Creating/Recoding variables

```
# changes the weight to ton
newdata = mutate(port, WT_Q = WT_Q/1000)
head(newdata)
```

```
##   ID       CNTR_N CNTR_OPR_C CNTR_STATUS_C PURPOSE_C LENGTH_Q CNTR_TYPE_C
## 1  1 NYKU 9701818         ON             F         T       20          GP
## 2  2 APHU 6602779         EG             F         T       40          GP
## 3  3 PCIU 6007321         PI             F         T       40          RF
## 4  4 HALU 5642133         HE             F         I       40          GP
## 5  5 TEMU 1697343         YM             F         T       20          GP
## 6  6 UACU 3529931         HL             F         T       20          GP
##     WT_Q LOAD_PORT_C DISC_PORT1_C DEST_PORT_C DISC_ABBR_VESSEL_M
DISC_SERVICE_C
## 1 22.080       MYPGU        INCCU                     PAC BANDA
PGU
## 2 22.932       CNSHA        IDPWG                   CMA TANCREDI
ASC
## 3 15.200       VNCLI        AUMEL       AUMEL
## 4 27.900       CNTXG                                AGAMEMNON
CIS
## 5 15.227       THBKK        INMUN       INMUN         MATHU BHUM
TSB
## 6 25.000       MYPKG        TRIST                   HONOLULU BR
CIS
##   LOAD_ABBR_VESSEL_M LOAD_SERVICE_C      LOAD_DT      DISC_DT  GATE_OUT_DT
## 1          DONG HAI            CC1 2.018081e+13 2.018080e+13           NA
## 2        MARCO P 118            IDO 2.018072e+13 2.018072e+13 2.018072e+13
## 3       O KUALALUMPU            AAA 2.018070e+13           NA           NA
## 4                                               NA 2.018092e+13 2.018092e+13
```

```
## 5        CONTI CHAMP            CIS 2.018122e+13 2.018121e+13           NA
## 6          HAMBURG E            MD3 2.018093e+13 2.018092e+13           NA
##     GATE_IN_DT BATCH_ID LOGISTICS_PARTNER DWELL_DAYS  PRODUCT  COMMODITY
## 1           NA   201808              <NA>        3.1   Weever   Sea Food
## 2 2.018072e+13   201807              <NA>        7.6    Snail   Sea Food
## 3           NA   201807              <NA>        4.3    Snail   Sea Food
## 4           NA   201809              <NA>        2.7  Sorubim   Sea Food
## 5           NA   201812              <NA>        2.8  pumpkin Vegetables
## 6           NA   201809              <NA>        9.6 delicata Vegetables
##                              CARGO_OWNER RF_TEMP
## 1                                Renault      NA
## 2                             Mitsubishi      NA
## 3                             Mitsubishi       8
## 4                    Magna International      NA
## 5                                 Nissan      NA
## 6 Aviation Industry Corporation of China      NA

# create a new variable to store weight in ton
newdata = mutate(port, WT_Q_TON = WT_Q/1000)
head(newdata)

##   ID       CNTR_N CNTR_OPR_C CNTR_STATUS_C PURPOSE_C LENGTH_Q CNTR_TYPE_C
WT_Q
## 1  1 NYKU 9701818         ON             F         T       20          GP
22080
## 2  2 APHU 6602779         EG             F         T       40          GP
22932
## 3  3 PCIU 6007321         PI             F         T       40          RF
15200
## 4  4 HALU 5642133         HE             F         I       40          GP
27900
## 5  5 TEMU 1697343         YM             F         T       20          GP
15227
## 6  6 UACU 3529931         HL             F         T       20          GP
25000
##   LOAD_PORT_C DISC_PORT1_C DEST_PORT_C DISC_ABBR_VESSEL_M DISC_SERVICE_C
## 1       MYPGU        INCCU                      PAC BANDA            PGU
## 2       CNSHA        IDPWG                  CMA TANCREDI             ASC
## 3       VNCLI        AUMEL       AUMEL
## 4       CNTXG                                  AGAMEMNON            CIS
## 5       THBKK        INMUN       INMUN         MATHU BHUM            TSB
## 6       MYPKG        TRIST                   HONOLULU BR            CIS
##   LOAD_ABBR_VESSEL_M LOAD_SERVICE_C      LOAD_DT      DISC_DT  GATE_OUT_DT
## 1           DONG HAI            CC1 2.018081e+13 2.018080e+13           NA
## 2        MARCO P 118            IDO 2.018072e+13 2.018072e+13 2.018072e+13
## 3       O KUALALUMPU            AAA 2.018070e+13           NA           NA
## 4                                               NA 2.018092e+13 2.018092e+13
## 5        CONTI CHAMP            CIS 2.018122e+13 2.018121e+13           NA
## 6          HAMBURG E            MD3 2.018093e+13 2.018092e+13           NA
##     GATE_IN_DT BATCH_ID LOGISTICS_PARTNER DWELL_DAYS  PRODUCT  COMMODITY
```

```
## 1           NA   201808              <NA>        3.1   Weever   Sea Food
## 2 2.018072e+13   201807              <NA>        7.6    Snail   Sea Food
## 3           NA   201807              <NA>        4.3    Snail   Sea Food
## 4           NA   201809              <NA>        2.7  Sorubim   Sea Food
## 5           NA   201812              <NA>        2.8  pumpkin Vegetables
## 6           NA   201809              <NA>        9.6 delicata Vegetables
##                              CARGO_OWNER RF_TEMP WT_Q_TON
## 1                               Renault      NA   22.080
## 2                             Mitsubishi      NA   22.932
## 3                             Mitsubishi       8   15.200
## 4                    Magna International      NA   27.900
## 5                                 Nissan      NA   15.227
## 6 Aviation Industry Corporation of China      NA   25.000

# create a new flag to indicate heavy containers
newdata = mutate(port, HEAVY_FLAG = ifelse(WT_Q > 20000,
                                    "Heavy",
                                    "Not Heavy"))
head(newdata)

##   ID       CNTR_N CNTR_OPR_C CNTR_STATUS_C PURPOSE_C LENGTH_Q CNTR_TYPE_C
WT_Q
## 1  1 NYKU 9701818         ON             F         T       20          GP
22080
## 2  2 APHU 6602779         EG             F         T       40          GP
22932
## 3  3 PCIU 6007321         PI             F         T       40          RF
15200
## 4  4 HALU 5642133         HE             F         I       40          GP
27900
## 5  5 TEMU 1697343         YM             F         T       20          GP
15227
## 6  6 UACU 3529931         HL             F         T       20          GP
25000
##   LOAD_PORT_C DISC_PORT1_C DEST_PORT_C DISC_ABBR_VESSEL_M DISC_SERVICE_C
## 1       MYPGU        INCCU                     PAC BANDA            PGU
## 2       CNSHA        IDPWG                 CMA TANCREDI             ASC
## 3       VNCLI        AUMEL       AUMEL
## 4       CNTXG                                AGAMEMNON            CIS
## 5       THBKK        INMUN       INMUN       MATHU BHUM            TSB
## 6       MYPKG        TRIST                  HONOLULU BR            CIS
##   LOAD_ABBR_VESSEL_M LOAD_SERVICE_C      LOAD_DT      DISC_DT  GATE_OUT_DT
## 1           DONG HAI            CC1 2.018081e+13 2.018080e+13           NA
## 2        MARCO P 118            IDO 2.018072e+13 2.018072e+13 2.018072e+13
## 3       O KUALALUMPU            AAA 2.018070e+13           NA           NA
## 4                                               NA 2.018092e+13 2.018092e+13
## 5        CONTI CHAMP            CIS 2.018122e+13 2.018121e+13           NA
## 6          HAMBURG E            MD3 2.018093e+13 2.018092e+13           NA
##     GATE_IN_DT BATCH_ID LOGISTICS_PARTNER DWELL_DAYS  PRODUCT  COMMODITY
## 1           NA   201808              <NA>        3.1   Weever   Sea Food
```

```
## 2 2.018072e+13   201807              <NA>        7.6    Snail   Sea Food
## 3           NA   201807              <NA>        4.3    Snail   Sea Food
## 4           NA   201809              <NA>        2.7  Sorubim   Sea Food
## 5           NA   201812              <NA>        2.8  pumpkin Vegetables
## 6           NA   201809              <NA>        9.6 delicata Vegetables
##                               CARGO_OWNER RF_TEMP HEAVY_FLAG
## 1                                 Renault      NA      Heavy
## 2                              Mitsubishi      NA      Heavy
## 3                              Mitsubishi       8  Not Heavy
## 4                      Magna International      NA      Heavy
## 5                                  Nissan      NA  Not Heavy
## 6 Aviation Industry Corporation of China      NA      Heavy
```

Summarizing Data

```
# Calculate the mean length and weight of all the containers
newdata = summarise(port, MEAN_LENGTH = mean(LENGTH_Q, na.rm=TRUE),
                    MEAN_WEIGHT = mean(WT_Q, na.rm=TRUE))
head(newdata)
```

```
##   MEAN_LENGTH MEAN_WEIGHT
## 1    30.62129    18408.55
```

```
# Calculate the mean length and weight of all the containers by commodity
newdata = group_by(port, COMMODITY)
```

```
## Warning: Factor `COMMODITY` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```

```
newdata = summarise(newdata, MEAN_LENGTH = mean(LENGTH_Q, na.rm=TRUE),
                    MEAN_WEIGHT = mean(WT_Q, na.rm=TRUE))
newdata
```

```
## # A tibble: 26 x 3
##    COMMODITY       MEAN_LENGTH MEAN_WEIGHT
##    <fct>                 <dbl>       <dbl>
##  1 Adhesive               29.3      20584.
##  2 Biscuit                31.0      19735.
##  3 Chemicals              29.5      17702.
##  4 Confectionaries        30.2      15466.
##  5 Dairy                  27.4      17438.
##  6 Diary                  30.9      18370.
##  7 Drink                  30.9      19969.
##  8 Eggs                   27.5      16272.
##  9 Films                  33.6      15314.
## 10 Flowers                30.6      18471.
## # ... with 16 more rows
```

Missing Data

```
# what is the proportion of missing data for each variable
Percent_Miss <- colSums(is.na(port))/nrow(port)
as.data.frame(Percent_Miss)
```

```
##                     Percent_Miss
## ID                  0.0000000000
## CNTR_N              0.0000000000
## CNTR_OPR_C          0.0000000000
## CNTR_STATUS_C       0.0000000000
## PURPOSE_C           0.0000000000
## LENGTH_Q            0.0000000000
## CNTR_TYPE_C         0.0000000000
## WT_Q                0.0000000000
## LOAD_PORT_C         0.0000000000
## DISC_PORT1_C        0.0000000000
## DEST_PORT_C         0.0000000000
## DISC_ABBR_VESSEL_M 0.0000000000
## DISC_SERVICE_C      0.0000000000
## LOAD_ABBR_VESSEL_M 0.0000000000
## LOAD_SERVICE_C      0.0000000000
## LOAD_DT             0.1895065309
## DISC_DT             0.1984677855
## GATE_OUT_DT         0.8108135139
## GATE_IN_DT          0.8138539396
## BATCH_ID            0.0000000000
## LOGISTICS_PARTNER  0.9357510051
## DWELL_DAYS          0.0000000000
## PRODUCT             0.0002600364
## COMMODITY           0.0002600364
## CARGO_OWNER         0.0002600364
## RF_TEMP             0.9395715400
```

End of the Script