Week 4 Homework

Mike Kozlowski

2023-02-18

3a.) Write a function that will compute the squares of the values from 1 to 10, and print them out only if the square is an odd number.

```
x = 1:10
xsquared = x^2
xsquared[lapply(x, "%%", 2) == 0]
```

```
## [1]   4  16  36  64 100
```

b.) Create a list of strings of ten items found in your kitchen and a second list of ten items found in your living room (try to include at least two items found in both). Generate from these the list of all items found (no duplicates), the list of items found in both locations and the differences in the two sets.

```
kitchen = c("cleaning towel", "water bottle", "knife", "plate", "bowl", "drinking glass", "blender", "coffe
e grinder", "paper towels", "bread")
livingRoom = c("cleaning towel", "water bottle", "drinking glass", "couch", "tv", "desk", "office chair", "
computer", "textbooks", "bourbon")
union(kitchen, livingRoom)
```

```
##  [1] "cleaning towel" "water bottle"   "knife"          "plate"
##  [5] "bowl"           "drinking glass" "blender"        "coffee grinder"
##  [9] "paper towels"   "bread"          "couch"          "tv"
## [13] "desk"           "office chair"   "computer"       "textbooks"
## [17] "bourbon"
```

```
intersect(kitchen, livingRoom)
```

```
## [1] "cleaning towel" "water bottle"   "drinking glass"
```

```
setdiff(kitchen, livingRoom)
```

```
## [1] "knife"          "plate"          "bowl"           "blender"
## [5] "coffee grinder" "paper towels"   "bread"
```

```
setdiff(livingRoom, kitchen)
```

```
## [1] "couch"        "tv"           "desk"         "office chair" "computer"
## [6] "textbooks"    "bourbon"
```

c.) Use the current assessment data set to find out the list of property owners who own property worth less than $50k and more than $1 million.

```
mydata = read.csv("C:\\Users\\Mike\\Documents\\DAT511\\2-1 class\\2020-2021_Assessment_Roll.csv")
```

```
y = mydata$OWNER1[mydata$TOTAL.VALUE < 50000]
z = mydata$OWNER1[mydata$TOTAL.VALUE > 1000000]
intersect(y,z)
```

```
##   [1] "WILLIAM S HEIN & CO INC"        "CITY BUFFALO PERFECTING TITLE"
##   [3] "CITY OF BUFFALO"                "NATIONAL FUEL GAS DIST"
##   [5] "CITY OF BFLO"                   "CITY OF BUFFALO PERFECTING TIT"
##   [7] "HEALTH SCIENCES CHARTER SCHOOL" "BUFFALO HOUSING ASSOCIATES"
##   [9] "771 WEST FERRY LLC"             "ARCP DG BUFFALO NY LLC"
##  [11] "CITY BUFFALO"                   "UNITED STATES OF AMERICA"
##  [13] "ROSWELL PARK CANCER INSTITUTE"  "1665 MAIN STREET GROUP LLC"
##  [15] "BUFFALO MUNICIPAL HOUSING"      "ELIZABETH PIERCE OLMSTED"
##  [17] "1100 GROUP LLC"                 "945 WEST FERRY, LLC"
##  [19] "CITY OF BUFFALO THE"            "NIAGARA FRONTIER"
##  [21] "WESTERN NEW YORK VETERANS"      "CANISIUS COLLEGE OF"
##  [23] "POLISH COMMUNITY CENTER OF"     "PARKVIEW APARTMENTS HOUSING"
##  [25] "GBUAHN, LLC"                    "1238 GROUP LLC"
##  [27] "CON-RAIL"                       "ARTSPACE AFFORDABLE FAMILY"
##  [29] "BUSCARINO SALVATORE J"          "ELMWOOD SQUARE PRESERVATION,"
##  [31] "NORFOLK SOUTHERN CORP."         "BARNES STEPHEN E"
##  [33] "MICHIGAN-SENECA GROUP INC"      "BUFFALO URBAN RENEWAL AGENCY"
##  [35] "BIG HEART PET BRANDS"           "CHIPPEWA REALTY VENTURES LLC"
##  [37] "HSBC BANK USA, N.A."            "VERIZON NEW YORK INC"
##  [39] "COMMUNITY SERVICES FOR"         "WILLIAM NEWELL STREET LLC"
##  [41] "BUFFALO HISPANIC MANAGEMENT"    "METALICO BUFFALO, INC."
##  [43] "NIAGARA MOHAWK POWER CORP"      "TRUSSO ANTHONY F"
##  [45] "VISCO JEFFREY DR."              "CYCLORAMA BUILDING INVESTOR"
##  [47] "TATANKA DEVELOPMENT CO INC"     "BUFFALO TRANSPORTATION PIERCE"
##  [49] "CITY BUFFALO DIV REAL"          "COUNTY OF ERIE"
##  [51] "1876 BUEHL BLOCK, LLC"          "BUFFALO URBAN RENEWAL"
##  [53] "9274 GROUP INC"                 "HAYES PLACE MANAGEMENT GRP INC"
##  [55] "GENESEE GATEWAY, LLC"           "SENECA HAMBURG LLC"
##  [57] "SSSN HOUSING DEVELOPMENT"       "PEOPLE OF THE STATE OF NY THE"
##  [59] "SHILOH 4 & 7 HOUSING"           "MERCY HOSPITAL OF BUFFALO"
##  [61] "1021 BROADWAY LLC"              "NIA MOHAWK POWER CORP"
##  [63] "BETTER BUFFALO PROPERTIES II"   "MANUFACTURERS & TRADERS"
##  [65] "MILL RACE COMMONS, LLC"         "LAZARUS PROPERTIES LLC"
##  [67] "ST JOHNS BAPTIST CHURCH"        "CHILD & FAMILY SERVICES"
##  [69] "CON RAIL TRANS"                 "STATE OF NEW YORK"
##  [71] "HURON GROUP INC"                "598 MAIN STREET LLC"
##  [73] "SCARP PROPERTY ASSOCIATES LLC"  "STATE OF NEW YORK PEOPLE OF TH"
##  [75] "FOOD BANK OF WESTERN NY INC"    "SAHLEN PACKING COMPANY INC"
##  [77] "BUCKEYE TERMINALS, LLC"         "310 DELAWARE LLC"
##  [79] "1400 WILLIAM ST LIMITED"        "LEWIS STREET APARTMENTS LLC"
##  [81] "GCBC/BENDERSON ASSOCIATES"      "TRIPIFOODS INC"
##  [83] "PAMS PROPERTIES LLC"            "ELMWOOD VILLAGE CHARTER"
##  [85] "PLAZA GROUP 200 LLC"            "PROTECTIVE INDUSTRIES, INC."
##  [87] "ST MARTINS ROMAN CATHOLIC"      "GANSON ENTERTAINMENT LLC"
##  [89] "GOLD WYNN DELSAN COURT LLC"     "100 RIVER ROCK LLC"
##  [91] "NATIONAL GRID"                  "AURUBIS BUFFALO, INC."
##  [93] "LCB CAPITAL LLC"                "SPEEDWAY LLC"
##  [95] "GENERAL MILLS PROPERTIES"       "FASO GROUP LLC"
##  [97] "4628 GROUP INC"                 "BFLO MERCY HOUSING DEV"
##  [99] "N.F.T.A."                       "G&I IX EMPIRE DELAWARE CONSU"
## [101] "UNITED STATES POSTAL"           "PEOPLE INC."
## [103] "CAMPUS WALK ONE LLC"            "TIME WARNER CABLE NORTHEAST"
## [105] "PEOPLE OF THE STATE OF NY"      "YW-WNY HOUSING DEVELOPMEN"
## [107] "S & V ASSOCIATES LLC"           "GRECO PROPERTIES LLC"
## [109] "345 AMHERST ST CO. LLC"         "STATE UNIVERSITY OF NEW YORK"
## [111] "UNILAND PARTNERSHIP OF"         "85 RIVER ROCK ROAD LLC"
```

```
## [113] "ST AMBROSE ROMAN CATHOLIC"      "PLAZA ONE GROUP, INC."
## [115] "CHESED PROPERTIES BUFFALO LLC"  "BELMONT HOUSING RESOURCES FOR"
## [117] "GREAT ARROW ACQUISITION, LLC"   "SISTERS OF CHARITY"
## [119] "BUFFALO ACADEMY OF SCIENCE"     "SISTERS OF CHARITY HOSPITAL OF"
## [121] "1210 GROUP, LLC"                "NHRC REALTY ACQUISITION LLC"
## [123] "HERWOOD PROPERTIES"             "NEWBUFF ASSOCIATES LLC"
## [125] "SAFETEC INTERNATIONAL LLC"      "THE CANISIUS COLLEGE OF"
## [127] "CANISIUS COLLEGE"               "CANISIUS COLLEGE OF BUFFALO"
## [129] "1277 DELAWARE LLC"              "2225 EAST 7 PROPERTIES LLC"
## [131] "CITY OF BUFFALO DIVISION"       "D'YOUVILLE COLLEGE"
## [133] "TOWNSEND STEPHEN C"             "118 GRANT PROPERTY, INC."
## [135] "212 HOLDEN AVENUE, LLC"         "9271 GROUP, LLC"
## [137] "CITY OF BUFFALO BOARD OF"       "PEACE BRIDGE APARTMENTS LP"
## [139] "WEST BUFFALO CHARTER SCHOOL"    "MEDAILLE COLLEGE"
```

d.) Go to the Open Data Buffalo Site, go to the data section and download the Tree Inventory. You can search the site for tree, which should locate this file. Download the Tree Inventory as a CSV file (using the export menu). How many trees are on the inventory? Produce a table of the number of trees by the Council.District.

```
mydata2 = read.csv("C:\\Users\\Mike\\Documents\\DAT511\\Tree_Inventory.csv")
```

```
nrow(mydata2)
```

```
## [1] 133076
```

```
table(mydata2$Council.District)
```

```
##
##   Delaware   Ellicott   Fillmore    Lovejoy     Masten    Niagara      North
##      20650      16624      15361      11971      16008      11157      12754
##      South University
##      16730      11821
```

e.) Load the mtcars data set (using the command data(mtcars)). Convert the cylinders column to a factor. Create a regression model that predicts mpg (mpg is the y or dependent variable) using cyl, disp, hp, wt as predictor variables. Create the model, showing the formula and run an anova using an F test. Which predictors seem to be most important?

```
data(mtcars)
```

```
mtcars$cyl <- as.factor(mtcars$cyl)
regmodel = lm(mpg~cyl + disp + hp + wt, data = mtcars)
anova(regmodel)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## cyl        2 824.78  412.39 66.9607 5.545e-11 ***
## disp       1  57.64   57.64  9.3595  0.005093 **
## hp         1  18.50   18.50  3.0042  0.094894 .
## wt         1  64.99   64.99 10.5526  0.003194 **
## Residuals 26 160.13    6.16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# It appears that horsepower is the biggest predictor in this case.
```

4.) Read Section 3.1 of van der Loo and de Jonge, do problems 3.1.1., 3.1.2, 3.1.3,3.1.4, 3.1.5, 3.1.6, 3.1.7.

3.1.1) Explain which of the following numeric codes can or cannot be stored as integer data in R.

```
a. United States Social security numbers, consisting of nine figures.
These can be stored in integers as the limit for a 32 bit integer is 10 digits long.

b. Universal Product Code-A. This is a barcode convention that labels products with 12 decimal digits.
This cannot be as it is over the digit limit of a 32 bit integer.

c. Phone numbers consisting of 10 digits, all starting with zero.
Since the number start with 0, they will never go over the 32 bit integer limit (2*10^9) so this will also
work.
```

3.1.2) Explain why the following R command yields an error.

read.table(textConnection(“123451234512345”) , colClasses=“integer”) It’s over the limit for what can be stored in a 32 bit integer, it would need to be stored as a double.

3.1.3) Execute the following statement and determine the class of max_int.

max_int <-.Machine$integer.max

Now execute the following statements and explain what happens (recall that the postfix L behind a numeric literal indicates integer).

x1 <- 2L * max_int x2 <- 2 * max_int

```
max_int <-.Machine$integer.max
```

```
#x1 <- 2L * max_int
x2 <- 2 * max_int
```

Adding a postfix L to 2 forces it to be an integer value, which causes an overflow, resuling in an error. Without the L postfix, R seems to recognize the overflow and automatically make x2 be a double in order to have the code execute without error.

3.1.4) Predict the column types of d1 and d2 after executing the following statements.

d1 <- read.table(textConnection(“1324665248”)) d2 <- read.table(textConnection(“4827647632”))

Check your result and explain what happened. If you were asked to read in a file containing 10-digit product codes, what storage format would you choose?

```
#My prediction is that d1 will be an integer, and d2 will be a double, as R seems to be aware of overflowin
g integers and has been assigning var classes as doubles when necessary.
d1 <- read.table(textConnection("1324665248"))
d2 <- read.table(textConnection("4827647632"))
str(d1)
```

```
## 'data.frame':    1 obs. of  1 variable:
##  $ V1: int 1324665248
```

```
str(d2)
```

```
## 'data.frame':    1 obs. of  1 variable:
##  $ V1: num 4.83e+09
```

Since 10 digit product codes have the opportunity to overflow if going above the 32 bit integer limit, I'd use a double.

3.1.5) Determine the machine precision u for numbers in single precision format (hint: the mantissa consists of 23 bits).

```
2^-24
```

```
## [1] 5.960464e-08
```

3.1.6) Compute the smallest representable positive subnormal number in double precision format.

```
2^-52 * 2^-1022
```

```
## [1] 4.940656e-324
```

3.1.7) The fact that roundoff errors are made at each arithmetic operation also implies that computer addition is not associative. That is, in double precision arithmetic it is, in general, not true that (x+y) + z = x + (y + z). Test this using R or any other software you use to do calculations by computing (0.4+0.3) + 0.1 and 0.4 + (0.3+0.1) and comparing the results. How big is the difference?

```
as.double((0.4 + 0.3) + 0.1)
```

```
## [1] 0.8
```

```
as.double(0.4 + (0.3 + 0.1))
```

```
## [1] 0.8
```

Apparently there is no difference, unless I am doing these calculations in the wrong format. I tried both with and without the "as.double" prefix.