

R coding for public policy

Assignment 2

[Name removed]

Assignment Instruction:

Once you download and open Assignment_2.rmd in R studio,

Please complete all the problems in the empty line between “{r}” and “”. You can add more empty lines by press Enter in the empty line. You can also click the green arrow next to each code chunk to check your code. Please include only the relevant codes in the chunks. Once you complete all the problems, click “Knit”.

Submit the word document as an attachment to NYU Classes - Assignment - Assignment 2

#Problem 1. create a numeric vector, a, with 50 elements in the code chunk below

```
a<-c(1:50)
```

#Problem 2. create a character vector, b, with 50 elements in the code chunk below

```
b<-c(1:50, as.character())
```

#Problem 3. create a matrix, c, with the numeric vector a from problem 1 and the character vector b from problem 2, report the dimension of the matrix

```
c <- cbind(a,b)
dim(c)
## [1] 50  2
```

#Problem 4. turn the character vector from problem 2 into a factor

```
b_factor<-factor(b)
```

#Problem 5. import the “2012 SAT Results” dataset from NYC open data; name the dataset SAT; please use the url in the command. The link to the webpage(not the dataset) is here (<https://data.cityofnewyork.us/Education/2012-SAT-Results/f9bf-2cp4>)

```
SAT<-read.csv(url("https://data.cityofnewyork.us/resource/f9bf-2cp4.csv"), header=TRUE)
```

#Problem 6. check the structure and the summary data for SAT; if necessary, change the classes of the columns to be the most appropriate, then re-check the structure and the summary data for SAT

str(SAT)

```
## 'data.frame': 478 obs. of 6 variables:
## $ dbn : Factor w/ 478 levels "01M292","01M448",
...: 1 2 3 4 5 6 7 8 9 10 ...
## $ school_name : Factor w/ 478 levels "47 THE AMERICAN S
IGN LANGUAGE AND ENGLISH SECONDARY SCHOOL",...: 204 438 141 169 309 296 333 11
0 38 1 ...
## $ num_of_sat_test_takers : Factor w/ 175 levels "10","101","102",.
..: 68 166 136 135 98 11 35 42 20 36 ...
## $ sat_critical_reading_avg_score: Factor w/ 164 levels "279","287","300",
...: 34 62 56 93 69 17 145 95 159 74 ...
## $ sat_math_avg_score : Factor w/ 173 levels "312","315","317",
...: 70 87 68 67 94 153 158 82 164 66 ...
## $ sat_writing_avg_score : Factor w/ 163 levels "286","291","297",
...: 45 48 52 41 66 11 147 92 158 69 ...
```

summary(SAT)

```
## dbn
## 01M292 : 1
## 01M448 : 1
## 01M450 : 1
## 01M458 : 1
## 01M509 : 1
## 01M515 : 1
## (Other):472
##
school_name
## 47 THE AMERICAN SIGN LANGUAGE AND ENGLISH SECONDARY SCHOOL
: 1
## A. PHILIP RANDOLPH CAMPUS HIGH SCHOOL
: 1
## ABRAHAM LINCOLN HIGH SCHOOL
: 1
## ABRAHAM LINCOLN YABC/LEARNING TO WORK GED AT ABRAHAM LINCOLN
: 1
## ACADEMY FOR CAREERS IN TELEVISION AND FILM
: 1
## ACADEMY FOR COLLEGE PREPARATION AND CAREER EXPLORATION: A COLLEGE BOARD S
CHOOL: 1
## (Other)
:472
## num_of_sat_test_takers sat_critical_reading_avg_score sat_math_avg_score
## s : 57 s : 57 s : 57
## 54 : 10 367 : 8 385 : 9
## 48 : 8 370 : 8 364 : 8
## 72 : 8 384 : 8 371 : 8
## 9 : 8 398 : 8 381 : 8
## 29 : 7 363 : 7 391 : 8
```

```
## (Other):380          (Other):382          (Other):380
## sat_writing_avg_score
## s      : 57
## 368    : 9
## 370    : 9
## 359    : 8
## 394    : 8
## 351    : 7
## (Other):380

#changing classes
SAT$school_name <-as.character(SAT$school_name)
SAT$num_of_sat_test_takers <-as.numeric(SAT$num_of_sat_test_takers)
SAT$sat_critical_reading_avg_score <-as.numeric(SAT$sat_critical_reading_avg_score)
SAT$sat_math_avg_score <- as.numeric(SAT$sat_math_avg_score)
SAT$sat_writing_avg_score <- as.numeric(SAT$sat_writing_avg_score)
```

#Problem 7. *bonus* Which school has the highest average SAT score? Please create an object named "ANS" for your answer, print ANS.

```
library(sqldf)

## Loading required package: gsubfn
## Loading required package: proto
## Loading required package: RSQLite

#subset of the data for the tests
data_new<-sqldf("select sat_critical_reading_avg_score, sat_math_avg_score, s
at_writing_avg_score from SAT")

#Attempting to convert the "s" to NAs so that na.rm is valid.
data_new[data_new=="s"]<-NA

#convert all to numeric
data_new <- sapply( data_new, as.numeric )

#make school names a vector
school_vector <- as.vector(SAT$school_name)

ANS <- (school_vector[data_new[which.max(rowSums(data_new, na.rm=TRUE))]])
print(ANS)

## [1] "BRONX CENTER FOR SCIENCE AND MATHEMATICS"
```