SM 727

Adam Kaderabek

9/30/2020

# Please write down the R command(s) needed after each question! There is a total of 22 questions.

# Data types

### 1. Which command do you use to determine the type of an object?

The native R command is typeof(), with the object referenced in the parentheses.

### 2. What is the type of vector A?

A <- c("2", "3", "4", "5", "6", "7", "8")  
typeof(A)

## [1] "character"

### 3. Convert A into an integer vector

A1 <- as.integer(A)  
A <- A1  
is.integer(A)

## [1] TRUE

### 4. Create an integer vector B containing the numbers one through ten

B <- c(1L:10L)  
print(B)

## [1] 1 2 3 4 5 6 7 8 9 10

is.integer(B)

## [1] TRUE

### 5. Create a new vector C from B which has the type “double”

C <- as.double(B)  
print(C)

## [1] 1 2 3 4 5 6 7 8 9 10

is.double(C)

## [1] TRUE

### 6. Change the third value of B to “3.5”

print(B)

## [1] 1 2 3 4 5 6 7 8 9 10

B[B==3]<-3.5  
print(B)

## [1] 1.0 2.0 3.5 4.0 5.0 6.0 7.0 8.0 9.0 10.0

### 7. Did this affect the type of B? How?

Yes, the type is now double.

typeof(B)

## [1] "double"

# Reading in data

Download both the Angell.dta (Stata data format) dataset and the Angell.txt dataset from this website <https://stats.idre.ucla.edu/stata/examples/ara/applied-regression-analysis-by-fox-data-files/>

### 8. Read in the .dta version and store in an object called angell\_stata

library(haven)  
AngeLL\_strata <- read\_dta("https://stats.idre.ucla.edu/stat/stata/examples/ara/angell.dta")  
AngeLL\_strata

## # A tibble: 43 x 5  
## city morint ethhet geomob region  
## <chr> <dbl> <dbl> <dbl> <chr>   
## 1 Rochester 19 20.6 15 E   
## 2 Syracuse 17 15.6 20.2 E   
## 3 Worcester 16.4 22.1 13.6 E   
## 4 Erie 16.2 14 14.8 E   
## 5 Milwaukee 15.8 17.4 17.6 MW   
## 6 Bridgeport 15.3 27.9 17.5 E   
## 7 Buffalo 15.2 22.3 14.7 E   
## 8 Dayton 14.3 23.7 23.8 MW   
## 9 Reading 14.2 10.6 19.4 E   
## 10 Des\_Moines 14.1 12.7 31.9 MW   
## # ... with 33 more rows

### 9. Read in the .txt version and store it in an object called angell\_txt

angell\_txt <- read.table("https://stats.idre.ucla.edu/wp-content/uploads/2016/02/angell.txt")  
angell\_txt

## V1 V2 V3 V4 V5  
## 1 Rochester 19.0 20.6 15.0 E  
## 2 Syracuse 17.0 15.6 20.2 E  
## 3 Worcester 16.4 22.1 13.6 E  
## 4 Erie 16.2 14.0 14.8 E  
## 5 Milwaukee 15.8 17.4 17.6 MW  
## 6 Bridgeport 15.3 27.9 17.5 E  
## 7 Buffalo 15.2 22.3 14.7 E  
## 8 Dayton 14.3 23.7 23.8 MW  
## 9 Reading 14.2 10.6 19.4 E  
## 10 Des\_Moines 14.1 12.7 31.9 MW  
## 11 Cleveland 14.0 39.7 18.6 MW  
## 12 Denver 13.9 13.0 34.5 W  
## 13 Peoria 13.8 10.7 35.1 MW  
## 14 Wichita 13.6 11.9 42.7 MW  
## 15 Trenton 13.0 32.5 15.8 E  
## 16 Grand\_Rapids 12.8 15.7 24.2 MW  
## 17 Toledo 12.7 19.2 21.6 MW  
## 18 San\_Diego 12.5 15.9 49.8 W  
## 19 Baltimore 12.0 45.8 12.1 E  
## 20 South\_Bend 11.8 17.9 27.4 MW  
## 21 Akron 11.3 20.4 22.1 MW  
## 22 Detroit 11.1 38.3 19.5 MW  
## 23 Tacoma 10.9 17.8 31.2 W  
## 24 Flint 9.8 19.3 32.2 MW  
## 25 Spokane 9.6 12.3 38.9 W  
## 26 Seattle 9.0 23.9 34.2 W  
## 27 Indianapolis 8.8 29.2 23.1 MW  
## 28 Columbus 8.0 27.4 25.0 MW  
## 29 Portland\_Oregon 7.2 16.4 35.8 W  
## 30 Richmond 10.4 65.3 24.9 S  
## 31 Houston 10.2 49.0 36.1 S  
## 32 Fort\_Worth 10.2 30.5 36.8 S  
## 33 Oklahoma\_City 9.7 20.7 47.2 S  
## 34 Chattanooga 9.3 57.7 27.2 S  
## 35 Nashville 8.6 57.4 25.4 S  
## 36 Birmingham 8.2 83.1 25.9 S  
## 37 Dallas 8.0 36.8 37.8 S  
## 38 Louisville 7.7 31.5 19.4 S  
## 39 Jacksonville 6.0 73.7 27.7 S  
## 40 Memphis 5.4 84.5 26.7 S  
## 41 Tulsa 5.3 23.8 44.9 S  
## 42 Miami 5.1 50.2 41.8 S  
## 43 Atlanta 4.2 70.6 32.6 S

### 10. Drop the first five observations in the angell\_txt object

angell\_txt <- angell\_txt[-c(1:5),]  
angell\_txt

## V1 V2 V3 V4 V5  
## 6 Bridgeport 15.3 27.9 17.5 E  
## 7 Buffalo 15.2 22.3 14.7 E  
## 8 Dayton 14.3 23.7 23.8 MW  
## 9 Reading 14.2 10.6 19.4 E  
## 10 Des\_Moines 14.1 12.7 31.9 MW  
## 11 Cleveland 14.0 39.7 18.6 MW  
## 12 Denver 13.9 13.0 34.5 W  
## 13 Peoria 13.8 10.7 35.1 MW  
## 14 Wichita 13.6 11.9 42.7 MW  
## 15 Trenton 13.0 32.5 15.8 E  
## 16 Grand\_Rapids 12.8 15.7 24.2 MW  
## 17 Toledo 12.7 19.2 21.6 MW  
## 18 San\_Diego 12.5 15.9 49.8 W  
## 19 Baltimore 12.0 45.8 12.1 E  
## 20 South\_Bend 11.8 17.9 27.4 MW  
## 21 Akron 11.3 20.4 22.1 MW  
## 22 Detroit 11.1 38.3 19.5 MW  
## 23 Tacoma 10.9 17.8 31.2 W  
## 24 Flint 9.8 19.3 32.2 MW  
## 25 Spokane 9.6 12.3 38.9 W  
## 26 Seattle 9.0 23.9 34.2 W  
## 27 Indianapolis 8.8 29.2 23.1 MW  
## 28 Columbus 8.0 27.4 25.0 MW  
## 29 Portland\_Oregon 7.2 16.4 35.8 W  
## 30 Richmond 10.4 65.3 24.9 S  
## 31 Houston 10.2 49.0 36.1 S  
## 32 Fort\_Worth 10.2 30.5 36.8 S  
## 33 Oklahoma\_City 9.7 20.7 47.2 S  
## 34 Chattanooga 9.3 57.7 27.2 S  
## 35 Nashville 8.6 57.4 25.4 S  
## 36 Birmingham 8.2 83.1 25.9 S  
## 37 Dallas 8.0 36.8 37.8 S  
## 38 Louisville 7.7 31.5 19.4 S  
## 39 Jacksonville 6.0 73.7 27.7 S  
## 40 Memphis 5.4 84.5 26.7 S  
## 41 Tulsa 5.3 23.8 44.9 S  
## 42 Miami 5.1 50.2 41.8 S  
## 43 Atlanta 4.2 70.6 32.6 S

### 11. Select columns 2 and 3 of the agell\_stata object and store them in a new object called angell\_small

angell\_small <- AngeLL\_strata[c(1:43), c(2,3)]  
angell\_small

## # A tibble: 43 x 2  
## morint ethhet  
## <dbl> <dbl>  
## 1 19 20.6  
## 2 17 15.6  
## 3 16.4 22.1  
## 4 16.2 14   
## 5 15.8 17.4  
## 6 15.3 27.9  
## 7 15.2 22.3  
## 8 14.3 23.7  
## 9 14.2 10.6  
## 10 14.1 12.7  
## # ... with 33 more rows

### 12. Install the “MASS” package, load the package. Then, load the Boston dataset.

library("MASS")  
data(Boston, package = "MASS")

### 13. What is the type of the Boston object?

typeof(Boston)

## [1] "list"

### 14. What is the class of the Boston object?

class(Boston)

## [1] "data.frame"

# Basic data summarizing and description

### 15. How many of the suburbs in the Boston data set bound the Charles river?

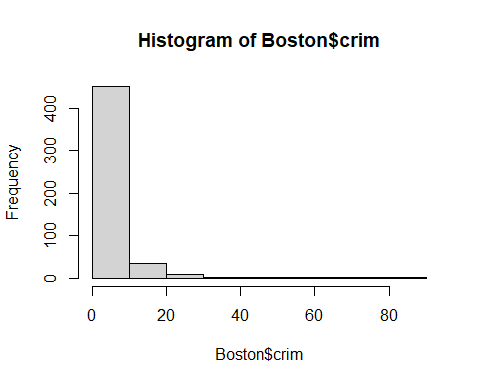
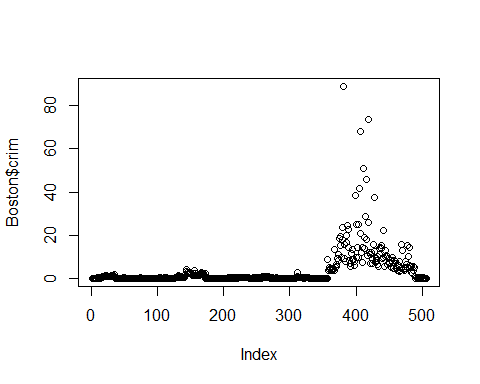
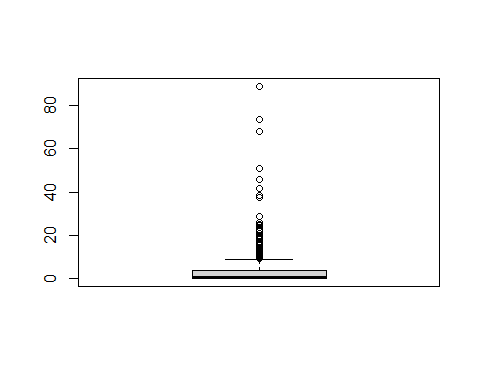
By finding the cases with a value of 1 for the variable “chas” we see the answer is 35.

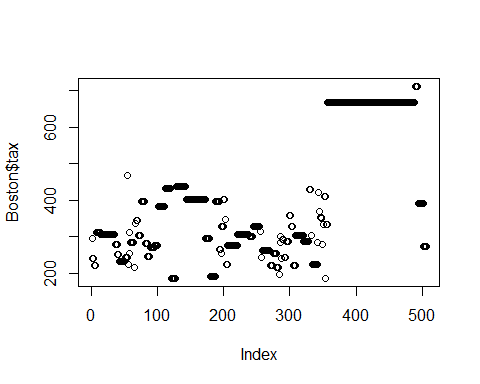
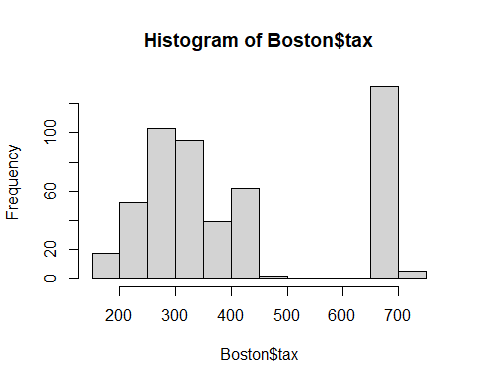
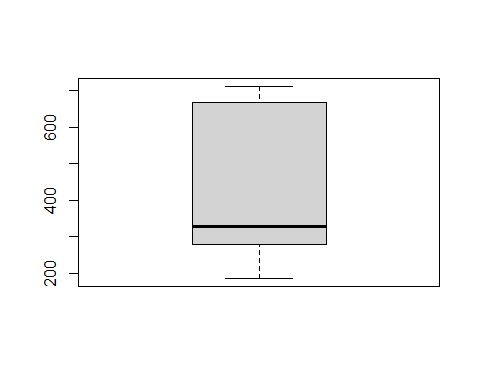
table(Boston$chas, Boston$chas == 1)

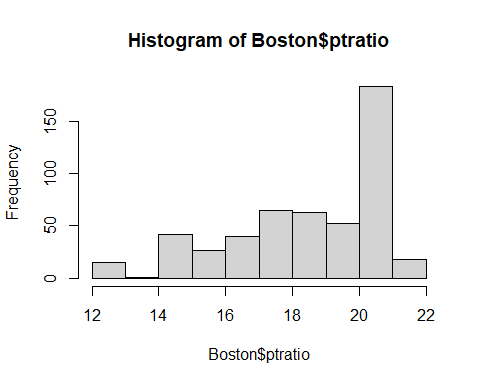
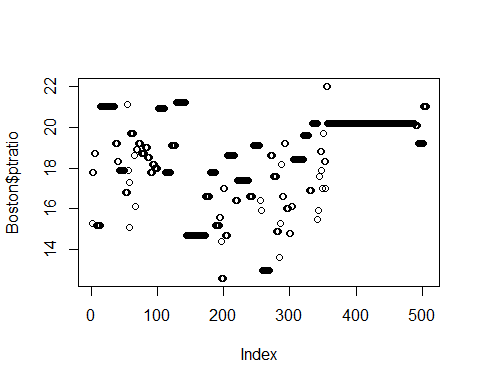
##   
## FALSE TRUE  
## 0 471 0  
## 1 0 35

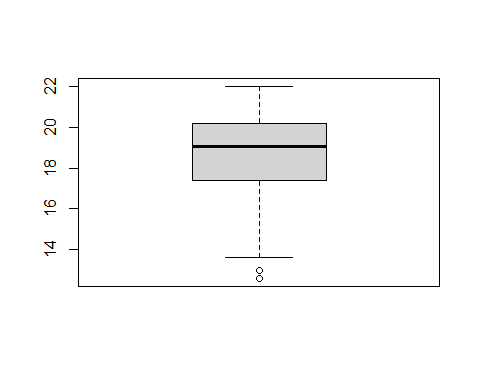
### 16. Do any of the suburbs of Boston appear to have particularly high crime rates? Tax rates? Pupil-teacher ratios? Comment on the range of each variable.

If we review the summary and some simple plots we see that yes there is a lot of variance in all three variables. Within all three we see that some suburbs are much higher than the mean.

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00632 0.08204 0.25651 3.61352 3.67708 88.97620

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 187.0 279.0 330.0 408.2 666.0 711.0

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 12.60 17.40 19.05 18.46 20.20 22.00



### 17. What is the median pupil-teacher ratio among the towns in this data set that have a per capita crime rate larger than 1?

19.29

subpt <- subset(Boston, Boston$crim >= 1)  
summary(subpt$ptratio)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 14.70 20.20 20.20 19.29 20.20 21.20

# Functions

### 18. Write a function that calculates the squareroot of an integer

sri <- function(x) {  
 y <- as.integer(x)  
 sqrt(y)  
 }  
sri(9)

## [1] 3

### 19. Write a function that calculates 95% confidence intervals for a point estimate. The function should be called “my.CI”. When called with “my.CI(2, 0.2)”, the output of the function should read “The 95% CI upper bound of point estimate 2 with standard error 0.2 is 2.392. The lower bound is 1.608.” *Note: the function should take a point estimate and its standard error as arguments. You may use the formula for 95% CI: point estimate +/- 1.96\*standard error.*

### 20. Write a function that converts all negative numbers in the following dataset into NA. Use as little code as possible and try to avoid code repetition.

set.seed(1002)  
df <- data.frame(replicate(10, sample(c(1:10, c(-99,-98,-5)), 6, rep = TRUE)))  
names(df) <- letters[1:6]  
df

## a b c d e f NA NA NA NA  
## 1 -98 6 1 6 7 1 -98 6 5 -98  
## 2 9 -5 10 4 -5 7 -99 3 2 2  
## 3 -5 3 5 3 2 2 5 10 7 -5  
## 4 7 8 -98 9 9 2 10 4 3 -99  
## 5 4 1 5 3 6 6 10 7 -99 6  
## 6 -5 -5 3 9 3 7 10 -98 7 6

NegNA <- function(x) {  
 x[x < 0] <- NA #replace the subset of values in x where x is a negative number (i.e. less than 0) with NA  
 x  
}  
NegNA(df)

## a b c d e f NA NA NA NA  
## 1 NA 6 1 6 7 1 NA 6 5 NA  
## 2 9 NA 10 4 NA 7 NA 3 2 2  
## 3 NA 3 5 3 2 2 5 10 7 NA  
## 4 7 8 NA 9 9 2 10 4 3 NA  
## 5 4 1 5 3 6 6 10 7 NA 6  
## 6 NA NA 3 9 3 7 10 NA 7 6

### 21. Use your function to convert all negative numbers in the dataset into NA without changing the class of the object

NegNA(df)

## a b c d e f NA NA NA NA  
## 1 NA 6 1 6 7 1 NA 6 5 NA  
## 2 9 NA 10 4 NA 7 NA 3 2 2  
## 3 NA 3 5 3 2 2 5 10 7 NA  
## 4 7 8 NA 9 9 2 10 4 3 NA  
## 5 4 1 5 3 6 6 10 7 NA 6  
## 6 NA NA 3 9 3 7 10 NA 7 6

class(df) #object class is still data.frame

## [1] "data.frame"

### 22. Change the function you wrote above such that it turns any negative number into NA!

My function does turn any negative number of an object to NA. It seems it would be more complicated to have done it any other way.