Week 5 R Assignment

Nathan Groom

July 1, 2014

# Appending is when you create a data structure with a certain amount of elements, then add elements as you go along. Preallocating is creating an empty data structure with a predetermined amount of elements that will be filled in later. Conventional R wisdom seems to point to preallocating as the superior method of creating data structures. As the book states, occasionally appending to a list/vector/data.frame is fine but is computationally expensive. It's best to create a list/vector/data.frame as long as its final desired size and then fill it in using appropriate values.

# Appending

# appends another element to our data frame, which in this case is just the number 5

append(theDF, 5) $First [1] 10 9 8 7 6 5 4 3 2 1

$Second [1] -4 -3 -2 -1 0 1 2 3 4 5

$Sport [1] Hockey Football Baseball Rugby Lacrosse Basketball [7] Tennis Curling Cricket Soccer  
10 Levels: Baseball Basketball Cricket Curling Football ... Tennis

[[4]][1] 5 #appends an element, the character 'a', to a list called list3 append(list3, 'a') [[1]][1] 1 2 3

[[2]][1] 99

[[3]][1] "a" #append a factor to a vector vec1. It appears as just a number '1' because there's just 1 level, even though the variable was defined as "data". The vector was created as a vector of integers m <- factor("data") m [1] data Levels: data vec1 <- c(1,5,33,12,9) append(vec1, m) [1] 1 5 33 12 9 1

# Preallocating

# create a data frame with 5 numeric, 5 character, and 5 factor rows

prealdf <- data.frame(colname=numeric(5), colname2=character(5), colname3=factor(5)) > prealdf colname colname2 colname3 1 0 5 2 0 5 3 0 5 4 0 5 5 0 5

# create a vector with 10 empty slots. You can later use [ ] to add elements

vec2 <- rep(NA, 10) vec2 [1] NA NA NA NA NA NA NA NA NA NA

# preallocate a list with 10 null values

preallist <- vector(mode = "list", length = 10) > preallist [[1]] NULL

[[2]] NULL

[[3]] NULL

[[4]] NULL

[[5]] NULL

[[6]] NULL

[[7]] NULL

[[8]] NULL

[[9]] NULL

[[10]] NULL

# Question 2

# Essentially the steps to make a dataframe containing just a few columns from a larger .csv file are as follows: First find out where the desired .csv file is on your computer and import it as whatever you want the table to be called. Mine was Boston <- /Users/nathangroom/Desktop/boston\_house\_prices.csv . Then assign names to the columns you want to use in your data frame. You have to refer to them by number; since I chose nox, crim, and dis, my code was NitrogenOxides <- Boston[,5] , Crime <- Boston[,1] and Distance<-Boston[,8]. Then all you have to do is create the dataset like this: BOSDF1 <- data.frame(NitrogenOxides, Crime, Distance). Originally I had the header in there but I deleted it, which is why the first row is technically labeled row 2.

BOSDF1 NitrogenOxides Crime Distance 2 0.538 0.00632 4.09 3 0.469 0.02731 4.9671 4 0.469 0.02729 4.9671 5 0.458 0.03237 6.0622 6 0.458 0.06905 6.0622 7 0.458 0.02985 6.0622 8 0.524 0.08829 5.5605 9 0.524 0.14455 5.9505 10 0.524 0.21124 6.0821