HW 4

Emil Ramos

February 9, 2018

## 1. FiveThirtyEight Data (30 points):

1. Install the fivethirtyeight package:

install.packages(“fivethirtyeight”)

1. In the listing of Data sets in package ‘fivethirtyeight,’ assign the eighteenth data set to an object ‘df.’

HW4 <- college\_grad\_students

1. Use a more detailed list of the data sets to write out the URL in a comment to the related news story.
2. Using R command(s), give the dimensions and column names of this data frame.

dim(HW4) [1] 173 22

colnames(HW4) [1] “major\_code” “major”  
[3] “major\_category” “grad\_total”  
[5] “grad\_sample\_size” “grad\_employed”  
[7] “grad\_employed\_fulltime\_yearround” “grad\_unemployed”  
[9] “grad\_unemployment\_rate” “grad\_p25th”  
[11] “grad\_median” “grad\_p75th”  
[13] “nongrad\_total” “nongrad\_employed”  
[15] “nongrad\_employed\_fulltime\_yearround” “nongrad\_unemployed”  
[17] “nongrad\_unemployment\_rate” “nongrad\_p25th”  
[19] “nongrad\_median” “nongrad\_p75th”  
[21] “grad\_share” “grad\_premium”

## 2. Data Summary (30 points): Use your newly assigned data frame for Question 2:

1. Write an R command that gives you the column names of the data frame. Right after that, write one that counts the number of columns but not rows. Hint: The number should match one of your numbers in Question 1d for dimensions

colnames(HW4) [1] “major\_code” “major”  
[3] “major\_category” “grad\_total”  
[5] “grad\_sample\_size” “grad\_employed”  
[7] “grad\_employed\_fulltime\_yearround” “grad\_unemployed”  
[9] “grad\_unemployment\_rate” “grad\_p25th”  
[11] “grad\_median” “grad\_p75th”  
[13] “nongrad\_total” “nongrad\_employed”  
[15] “nongrad\_employed\_fulltime\_yearround” “nongrad\_unemployed”  
[17] “nongrad\_unemployment\_rate” “nongrad\_p25th”  
[19] “nongrad\_median” “nongrad\_p75th”  
[21] “grad\_share” “grad\_premium”

ncol(HW4) [1] 22

1. Generate a count of each unique major\_category in the data frame. I recommend using libraries to help. I have demonstrated one briefly in live-session. To be clear, this should look like a matrix or data frame containing the major\_category and the frequency it occurs in the dataset. Assign it to major\_count.

major\_count <- table(HW4$major\_category) major\_count

Agriculture & Natural Resources Arts   
 10 8   
 Biology & Life Science Business   
 14 13   
 Communications & Journalism Computers & Mathematics   
 4 11   
 Education Engineering   
 16 29   
 Health Humanities & Liberal Arts   
 12 15

Industrial Arts & Consumer Services Interdisciplinary 7 1 Law & Public Policy Physical Sciences 5 10 Psychology & Social Work Social Science 9 9  
c. To make things easier to read, put par(las=2) before your plot to make the text perpendicular to the axis. Make a barplot of major\_count. Make sure to label the title with something informative (check the vignette if you need), label the x and y axis, and make it any color other than grey. Assign the major\_category labels to their respective bar. Flip the barplot horizontally so that bars extend to the right, not upward. All of these options can be done in a single pass of barplot(). Note: It’s okay if it’s wider than the preview pane.

barplot(major\_count, horiz = TRUE, col = rainbow(2), xlab = “category”, ylab = “amount”, main = “Major Category”)

1. Write the fivethirtyeight data to a csv file. Make sure that it does not have row labels.

write.csv(bechdel, “fivethirtyeight.csv”, row.names = FALSE)