R Readiness Assignment

Introduction to R, RStudio, dplyr, and ggplot

This faux assignment is intended to give you experience working with the R program. If this were a real assignment, you would submit your responses to each of the questions below in a printed document and label the sections as indicated below within your printed document. All graphics would be resized so that they do not take up more room than necessary and would have an appropriate caption. Any mathematics/equations would also need to be appropriately typeset within the document.

Preparation: Install Packages

Open RStudio and install the following packages, if you have not already installed them:

- dplyr
- ggplot2
- ggthemes
- gridExtra
- sm

Once these have been installed successfully, you should not need to install them again.

Preparation: Script File

Open a new script file. Save the script file as `Assignment-00.R`. Save all of the R syntax you use to answer the questions on this assignment in this script file.

Denote each question in the script file using comments. For example,

```
##################################################
### Question 1
##################################################
```

Add comments throughout your syntax as liberally as you feel is necessary to help you recall what the syntax does in the future.

Part I

In 2013, Andy read 40 books. The number of pages Andy read each month is reported in Table 1.

Table 1 *Number of pages read per month.*

Month	Pages
January	1453
February	422
March	848
April	1679
May	1655
June	1630
July	710
August	557
September	978
October	920
November	647
December	2698

1. Use the `data.frame()` function to enter these data into a data frame called `reading`. There should be two columns in the data frame. The first column you should name `month` and the second should be named `pages`.

2. Use the `sum()` function to find the total number of pages Andy read in 2013.

3. Use the `sm.density()` function from the **sm** package to create a density plot of the marginal distribution of `pages`.

4. Use the `summarize()` function from the **dplyr** package to compute the mean and standard deviation for the total number of `pages` read per month.

Part II

Use RStudio to open the *goodreads-2016.csv* dataset and assign it into an object called `read`. Use the data to answer the following questions.

5. Use **dplyr** to do the following: (1) select only the books on the "read" bookshelf; and (2) compute the total number of pages (use `sum()`) read each month. (Hint: group by month and then summarize. The output should be a data frame with 12 rows and two columns.)

Part III

Use RStudio to open the *beauty.csv* dataset and assign it into an object called `beauty`. Use the data to answer the following questions.

6. Use `ggplot()` to create a scatterplot of the relationship between professors' beauty ratings (`btystdave`) and their average course evaluation rating (`avgeval`). (Put the beauty ratings on the x-axis.) Shange the axis labels so that both the x- and y-axis have labels that suitable describe the variables being plottted. (For help on this, read the *Axes* page of the Cookbook for R website.) Finally, add a figure caption that adequately explains your figure (see http://www.svsu.edu/writingcenter/studentresources/citation/apaformatusingtablesandfigures/). Include this plot in a word-processed document. Resize the plot so it does not take up any more space than neccesary.

Part IV

In this section, you will again, work with the data in the `beauty` object you created in Part III.

7. The variable `female` in the data set is a dummy variable indicating the gender of the professoer; $0 =$ male and $1 =$ female. Use **dplyr** syntax to create a new variable in the dataset called `gender` that has the levels `Male` and `Female` rather than 0 and 1. (There are many ways to do this. For example, see http://www.theanalysisfactor.com/r-tutorial-recoding-values/.)

8. Use `ggplot()` to again creat a scatterplot of the relationship between professors' beauty ratings and their average course evaluation rating. This time, color the observations by gender. Change the point colors to some non-default palette of your choice. Also, facet the plot using gender. Finally, add a regression line to the faceted plot by including the layer `geom_smooth(method = "lm", se = FALSE)`. Be sure the plot has appropriate labels (on both sxes, and on the legend), and has a caption. Include this plot in a word-processed document. Resize the plot so it does not take up any more space than neccesary.