This document needs to be saved to your Root Folder.

**Class Project**

The ***Class Project*** is the only assignment that you will be evaluated on in this class. The purpose of the ***Class Project*** is to apply the skills you learn in class to your own dataset(s) and variables. If you do not have a dataset, you can find one online (just google "R datasets"). There are many freely available datasets online [including hundreds at this website](https://vincentarelbundock.github.io/Rdatasets/datasets.html). You can use multiple script files in the project.

***To pass the class*** you need to:

1. Meet all the requirements in ***Table 1: Project Requirements Table***.
2. Earn 75% of the points in the ***Table 2: Skill List***.

Class Project submissions are sent to the instructor ([belinsky@msu.edu](mailto:belinsky@msu.edu)). The submission needs to include:

1. This document with the ***Table 2: Skill List*** filled out
2. All R script files used in your project
3. All data files used in your project
4. A list of R packages you used in your project (if any) **Packages: tidyr , dplyr, ggplot2,reshape2**

|  |  |
| --- | --- |
| ***Table 1: Project Requirements*** | **Lesson** |
| 1. Script is well commented | 1.1 |
| 1. Use semicolons ( ; ) to end commands throughout your scripts. | 1.1 |
| 1. Consistent alignment of curly bracket ( **{ }** ) | 1.7 |
| 1. Give a brief (1-2 sentences) description in the ***Description*** column of ***Table 2*** for each skill you used. |  |
| 1. Skills completed in ***Table 2*** are commented in your script file where the skill is demonstrated.   The comment should be # SKILL XX where XX is the skill number in ***Table 2***. |  |
| 1. Script lines have no text beyond the 80th character with exception of:   1) long file names (these cannot be broken down)  2) SKILL XX comments | 1.6 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Table 2: Skills List*** (you can use the same line(s) of code to satisfy multiple skills) | **Lesson** | **Max**  **Points** | **Your Points** | **Description** (1-2 sentences and make sure you adjust the points in the ***Your Points*** column) |
| 1. Use of a standard mathematical operations (+, -, \*, /). *Half-point per use/ max of 1 point.* | 1.3 | 1 | 1 | Used addition and division-script 1 |
| 1. Using parentheses to order mathematical operations | 1.3 | 1 | 1 | Used parentheses to order operations-ClassProject\_1.r |
| 1. Using powers or roots | 1.3 | 1 | 1 | Used power for calculation- ClassProject\_1.r |
| 1. Get input from user and save the input to a variable | 1.5 | 1 | 1 | I asked two questions and saved each answer to a variable- ClassProject\_1.r |
| 1. Output mixed message (variables and text) to Console Window | 1.6 | 1 | 1 | Used cat() to output variables and text- ClassProject\_1.r |
| 1. Use of line feed (newline) character | 1.6 | 1 | 1 | Used \n within cat() to create a newline for the text within cat()-ClassProject\_1.r |
| 1. Use of different conditional operators.   *Half point for each unique operator (6 in all).* | 1.7 | 3 | 3 | Used a different conditional operator for each if statement- ClassProject\_1.r |
| 1. Use of if-else structure | 1.8 | 1 | 1 | Created if-else statement which expresses a response based on the input and condition- ClassProject\_1.r |
| 1. Use of if-else-if structure | 1.8 | 1 | 1 | Created if else if statement - ClassProject\_1.r |
| 1. Error condition in an if-else structure | 1.8 | 1 | 1 | Created error statement which expresses an error if input is not included in the statement- ClassProject\_1.r |
| 1. Use of && operator | 1.9 | 1 | 1 | Used && operator within an if statement to select for a condition with both fish length and fish type- ClassProject\_1.r |
| 1. Use of || operator | 1.9 | 1 | 1 | Used || operator within an if statement to express a response if one or the other fish lengths are input- ClassProject\_1.r |
| 1. Using && or || to check conditions on multiple variables | 1.10 | 1 | 1 | Checked conditions fish type and fish size- ClassProject\_1.r |
| 1. Save data from CSV file to a data frame | 2.1 | 1 | 1 | Saved csv to “fishData” at beginning of ClassProject\_2.r |
| 1. Subset one cell in a data frame – save to variable | 2.2 | 1 | 1 | Subset data point for species name of first row ClassProject\_2.r |
| 1. Subset column in a data frame – save to vector | 2.2 | 1 | 1 | Subset fishLength from fishData ClassProject\_2.r |
| 1. Subset individual value in vector | 2.2 | 1 | 1 | Subset individual length from fishLength vector – ClassProject\_2.r |
| 1. Use of sequence | 2.3 | 1 | 1 | Used seq in for loop. seqVal is used as the index variable for the dataset as the loop iterates - ClassProject\_2.r |
| 1. Use of sequence that increases or decreases by number other than 1 | 2.3 | 1 | 1 | Used fishNum in for loop , used sequence 1-10 by 2 - ClassProject\_2.r |
| 1. Sequence to set up iterations in a for() | 2.4 | 1 | 1 | Created sequence from length of vector fish length- ClassProject\_2.r |
| 1. Three uses of state variable. *One point / use* | 2.4 | 3 | 3 | created initial , intermediate and final state variable - ClassProject\_2.r |
| 1. Nested if() inside a for() | 2.4 | 1 | 1 | Nested if within for () to select for condition while for loop iterates – ClassProject\_2.r |
| 1. Use of Boolean values. | 2.4 | 1 | 1 | Created state/Boolean variable to test lower limit of length - ClassProject\_2.r |
| 1. Find max, min, or average value of a vector using for() | 2.5 | 2 | 2 | Used a state variable and a for() to determine min and max of fish length vector – ClassProject\_2.r |
| 1. Ignoring NA values | 2.5 | 1 | 1 | Used na.omit to ignore NA values- ClassProject\_2.r |
| 1. Add vector to data frame | 2.6 | 1 | 1 | Added length in inches to lakeSturgeonData- ClassProject\_2.r |
| 1. Rearrange columns in a data frame | 2.6 | 1 | 1 | Removed and rearranged columns in data frame ClassProject\_2.r |
| 1. Write a data frame to a CSV file | 2.6 | 1 | 1 | Created csv file- ClassProject\_2.r |
| 1. Label x-axis and y-axis and add title to a plot | 2.7 | 1 | 1 | Labeled histogram using ggplot – ClassProject-3.r |
| 1. Adding a legend to a plot | 2.7 | 1 | 1 | Added legend to plot - ClassProject\_5.r, line 63-65 |
| 1. Adding colors to a plot | 2.7 | 1 | 1 | Added color blue to vertical line in histogram – ClassProject-3.r |
| 1. Adding lines or points to a plot | 2.9 | 1 | 1 | Added vertical line to express average length- ClassProject-3.r |
| 1. Histogram | 2.7/2.9 | 1 | 1 | Created histogram using ggplot to visualize distribution of fish length– ClassProject-3.r |
| 1. Barplot | 2.7/2.9 | 1 | 1 | Created barplot using ggpot to visualize total number of fish for survey– ClassProject-3.r |
| 1. Multi-panel scatterplot using pairs() | 2.8/2.10 | 1 | 1 | Created point plot using pairs() – ClassProject\_5.r – line 67 |
| 1. Create and use two of your own functions (2 point/each) | 2.10 | 4 | 4 | Created function to convert length -ClassProject\_4.r  Created function to determine max length for species - ClassProject\_4.r |
| 1. Create a function with at least 2 input parameters | 2.10 | 2 | 2 | Created function with to input parameters-ClassProject\_5.r-line 19-25 |
| 1. Save the return value from your function to a variable | 2.10 | 1 | 1 | Saved return value to “allFishIn” – ClassProject\_4.r |
| 1. Create a function with default parameters | 2.10 | 1 | 1 | Created a default variable within function – ClassProject\_5.r-line 19 |
| 1. Use two functions from another package (1 point/each) | 3.1 | 2 | 2 | Used function “filter” from dplyr to create a vector with only data from a single species – ClassProject\_2.r line 69  Used plot function “ggplot” from ggplot2 – ClassProject\_3.r line 34 |
| 1. Reshape a data frame | 3.2 | 2 | 2 | Used (reshape2) to reshape data frame creating a new column for every country and only including one attribute- ClassProject\_5.r-line 14 |
| 1. Perform operation on multiple columns of a data frame or matrix | 3.2 | 2 | 2 | Used round to eliminate decimal points for all data within the caSchools data frame – ClassProject\_5.r-line 26 |
| 1. Using substring function | 3.2 | 1 | 1 | Used substr to only select the highest grade level provided- ClassProject\_5.r line 28 |
| 1. Rounding values | 3.2 | 1 | 1 | Used function round to round lengths to one decimal ClassProject\_4.r line 52 |
| 1. Renaming columns | 3.2 | 1 | 1 | Renamed row and column ClassProject\_5.r line 16,17 |
| 1. Create a matrix | 3.3 | 1 | 1 | Created matrix with new data set – ClassProject\_5.r line 89 |
| 1. Perform operation on rows, columns, and whole matrix | 3.3 | 1 | 1 | Performed 3 operations on matrix allCrime- ClassProject\_5.r, lines 92-94 |
| 1. Repeat values using the rep() function | 3.4 | 2 | 2 | Used rep to create values for a new vector- ClassProject\_5.r, line 100 |
| 1. Perform up to two ANOVAs and comment on results. *One point each* | 3.4 | 2 | 2 | Performed two different ANOVA tests on crime data – ClassProject\_5.r line 102,114 |
| 1. Perform up to two t-tests and comment on results. *One point each.* | 3.4 | 2 | 2 | Created t-test’s using average reading/math scores between two groups of schools – ClassProject\_5.r line 38,44 |
| 1. Create a boxplots that uses multiple categories | 3.4 | 2 | 2 | Created box plots to illustrate data used in t-test’s – ClassProject\_5.r line 50-53 |
| 1. Randomly sample data | 3.5 | 1 | 1 | Created randomly sampled data from dataframe- ClassProject\_5.r, line 121 |
| 1. Create pseudo-random values using set.seed() | 3.5 | 1 | 1 | Created pseudo-random values using set. Seed- of 25 to create random values from violent rate data ClassProject\_5.r, line 124 |
| 1. Sample from a normal distribution | 3.5 | 1 | 1 | Created sample from a normal distribution - ClassProject\_5.r, line 122 |
| 1. Subset a list | 3.5 | 1 | 1 | Subset list from saved list within data folder ClassProject\_5.r, line 130 |
| 1. Perform two linear regressions and comment on results. *One point each* | 3.6 | 2 | 2 | Created 2 linear regressions using violent crime rate against population and income- ClassProject\_6.r, line 12,17 |
| 1. Add regression line to a plot | 3.6 | 1 | 1 | Added regression lines for both linear models ClassProject\_6.r, line 13 |
| 1. Up to four examples of subsetting a vector using which(). *One-half point each.* | 3.7 | 2 | 2 | Subset 3 different vectors using which()ClassProject\_6.r, lines 21-25 |
| 1. Use grep() on a vector | 3.7 | 2 | 2 | Used grep to return indices of vectors using substring “Kansas” to include Kansas and “kansas” to include Arkansas - ClassProject\_6.r, line 26-27 |
| 1. Up to three examples of using subset vector to index another vector. *One point each.* | 3.7 | 3 | 3 | Used subset vector to index another vector- ClassProject\_6.r, line 28-31 |
| 1. Multiple condition on a subset vector | 3.8 | 2 | 2 | Created multiple condition on a subset vector , in order to subset and index vector– ClassProject\_6.r, line 36,37 |
| 1. Multiple conditions in a grep() | 3.8 | 1 | 1 | Used multiple conditions with grep() using OR ClassProject\_5.r - Line 25 |
| 1. Use of union() or intersect() | 3.8 | 1 | 1 | Used intersect to create a new vector with values in both input vectors– ClassProject\_6.r, line 40 |
| 1. Up to two examples of plotting subset vectors. *One point each.* | 3.7/3.8 | 1 | 1 | Created two plots using subset vectors – ClassProject\_5.r, lines 65-70 |
| 1. Save results as an .rdata file | 3.5 | 2 | 2 | Saved a list of data used to create subset plots- ClassProject\_5.r, line 128 |
| **Total Points**  Add up all the points from the ***Your Points*** column and put the results in this row. |  | **88** | **88** | If you highlight the cell to the left and press F9, the points in the ***Your Points*** column will automatically be added and the results put in the cell. |