R Coding for Data Analysis Final Presentation

For the midterm project you will be asked to identify a dataset of your choosing. You will read this data into R and use the skills taught in class to provide summary information about the data. Your project should be organized akin to a conference presentation of 8-10 minutes with:

● **Background/context** explaining why you chose the data you are working with, what it generally contains and how this data might be useful for analysis.

● **Methods** explaining how you cleaned and transformed your data to allow you to provide meaningful comparisons and relationships within the data. An example of this would be filtering expression data to only include highly expressed rows based on some theoretically meaningful threshold, or creating a dichotomous variable to code a continuous variable as groups for overweight samples of mice vs. non-overweight mice. Your project should explain the decisions you are making and present them in code.

● **Results** should explain some summary and inferential statistics you have uncovered in your data. Examples, building off those in the Methods, could be the proportion of expressed genes relative to the whole or the proportion of overweight mice. Other common summary statistics include looking at frequencies of different groups in your data, mean levels of different variables and how those means vary by different groups. The results should include at least one data visualization (a chart or figure). Examples might include a heatmap of the highly expressed genes, a bar plot comparing activity time of the overweight mice to the non-overweight, a histogram, a scatter plot, etc.

● **Discussion/Future Directions** that should point to ways you might further analyze your data to pull out meaningful insights from your data. This could be a plan for subsequent analysis based on what you discovered by looking at summary statistics of your data.

Your project should be created and submitted as a **.rmd R Markdown** file using one of the two HTML presentation options, Slidy or IoSlides. It should include text explaining your work with code snippets that actually pull in the data, run commands to clean and transform your data and finally generate slides and conduct basic analysis. This should be at least 10 slides in length and explain clearly to individuals without needing additional context what your data contains, why you are working with it, how you are cleaning it and why you are providing general descriptive statistics of different variables.

	4	3	2	1
RMD File	Code presented in your project works, you include clear	Code is present and works, but it is not clear why decisions were	Code works unevenly, formatting does not follow criteria	Code in project does not work, comments do you are running

	explanations of the commands you are running and why	made and a lack of commenting	of an RMD file.	commands
Background	Explanation given about the context of your data, why you have chosen it and what sort of insights you hope to gain from it	Explanation of context given, but does not fully explain why your data is worth examining	Background information is present but unclear or perfunctorily provided.	Context not explained about why you chose your data
Methods	Explains choices for why you are filtering, subsetting and dichotomizing variables, and what thresholds may be of interest	Explanations lacking but decisions on data cleaning and transformation demonstrate an understanding of why you are choosing to organize your data in the way you are.	Lack of explanations and at times data organization and cleaning decisions are unclear.	Lack of explanation for choices in cleaning and organization. Methods present evince a haphazard technique without a clear purpose for analysis
Results	Demonstration of ability to derive descriptive statistics (e.g. percentages, mean, proportions, frequencies) from variables in your data sample. Well formatted chart that includes an explanation of what you are visualizing and why.	Explanations of why you chose the descriptive statistics you did are missing, but you are able to present descriptive statistics and derive them from your data. Chart present but may lack formatting.	Unclear ability to provide descriptive statistics of your data in R. Visualizations lack formatting and is unclear to readers.	Unable to provide descriptive statistics of your data. No chart.
Future Directions	Discussion of what	Discussion of your data is	Restating of descriptive	No explanation demonstrating

	relationships exist in your data that might be worth further explanation	present but lacks evidence of understanding what comparisons might be made in your data.	statistics and methods without explanation of why you did what you did and how this might be useful to further analysis.	understanding of what was found in the summary statistics or potential future avenues for comparisons.