EFR 535 Final Project

A university professor has asked for help understanding the data from their undergraduate chemistry class. The class is quite large (117 students, 2 sections) and rather complex utilizing multiple learning modalities and evaluation types. The instructor has shared a comprehensive, anonymized dataset (over 100 variables). Please use RStudio and the techniques covered in EFR 535 to complete the tasks and answer the questions below. Please write your submission with the chemistry course instructor as your intended reader (i.e., concise writing, no statistical/technical jargon, use visualization best practices). The assignment is worth 75 points towards your final grade (5 points for completing each task below; 25 points for clarity, organization, and visualization accuracy & aesthetics). Due Monday, May 3, by 9 am.

Additional considerations:
- This data is intrinsically messy. There are points that do not add up to a consistent value in the section total, bonus points, and miscellaneous activities for points I do not understand. I have no more information about this data, nor can I get more, so do your best to work with it as it is.
- Your submission must be a Word document created using Markdown with the code for each visualization provided (i.e., don't hide all your code).
- Please attend to the labeling of your visualizations, both the overall plot title as well as axis labels.
- Avoid using the same type of visualization more than once (e.g., if you use a scatterplot for one analysis, use a count or binned scatterplot for another). If you must, be sure to modify the aesthetics.

1. Review the data file. Tidy and manage the data (e.g., variable names) as needed to prepare for analysis. Be sure to consider missing data. (I would have loved to give you the messy data, but I needed to remove the names so I am giving you a fairly clean, merged dataset… but you still should carefully look through it!). Load the data into RStudio. Review the variable types and make any changes you need. You may wish to run descriptive statistics on the variables (if you do, only report the most interesting/unexpected findings… I don't want a summary of every variable). Briefly explain what you did and why. Submit your final imported data file.
2. Who were the students in this class? Visualize student characteristics by gender, program, and level.
3. How were the students evaluated in this class? Visualize the composition of the final course grade: exams, online "Mastering Chemistry" (MC), in class "Team", "Exam Review" sessions, peer evaluations, etc. ~~Note, there was a lab component in this course (see variable "Lab_sec"), but it is not important here.~~
4. How did the students perform in the class overall? Visualize the proportion of the letter grades and the distribution of the total course points.
5. How effectively did students participate in class during "Team" activities (daily group activities in class among three students, includes quizzes and reports)? Visualize the distribution of student Team work. How did Team activity performance relate to exam performance? Correlation
6. Visualize the distribution of engagement among students outside of class in the online "Mastering Chemistry" (MC). Is greater MC performance related to better student performance on exams? Correlation
7. Were any types of students more/less active on Blackboard? Was Blackboard MC activity related to Homework Points, Quiz Points, or Final Course grade? gender/level/program BB total usage
8. What related more strongly to final exam score: online MC performance, in class Team performance, Exam Review sessions? Visualize these relationships and report on their correlations. *Fun bonus*: What are some other variables, if any, that relate to final exam performance other than those listed? correlation
9. How did the Starfish messages translate into student performance in the course overall? *Fun bonus*: What are some variables the instructor could use to identify students early on who struggling and performed poorly in the course in the end? maybe descriptive answer!
10. How did students' performance change over time, if at all? Visualize any two of the following variables over time: (a) Team reports, (b) MC quizzes, (c) MC homework, and (d) exams.

Fun bonus: Did students who completed the SELFI perform better or worse in the course than students who did not complete the SELFI? Correlation