A brief comment about grades: the goal was to get you engaged into a community of practice, to help engage you with the language of data analytics as it relates to multivariate analyses. I recognize this is a core, required course and I understand that passing is defined as a "C" not a "D-" ... I will take that under consideration as you submit your final tasks (one of which will allow you to self-assess and discuss your grade expectations). For me, the journey and lessons learned are way more valuable than any grade you can receive.

The final points consist of:

(1) Your paper-and-pencil notebook (worth 60 points). Prepare a one-page data summary of your efforts (turn in as oldSchoolNotebook.pdf): how many pages of content are written; how many unique dates, quotes, and reflection paragraphs questions have you jotted down; how many quick-notes have you written, and so on. A "master" student would have taken the "training model" I provided, and continued on that path. You will also need to upload a short movie (MP4, AVI, MOV, etc) where you briefly "show-and-tell" your notebook (turn in as oldSchoolNotebook.mp4 or whatever the extension is). It should slow down and zoom in on some initial content (where I was driving the bus) as well as some self-directed content (where you were driving the bus). About 2 minutes, and you should be talking while "showing" ... hence "show-and-tell".

(2) Your workspace audit (worth 50 points). This is an examination of the workspace you have used throughout the semester. Have you installed the version of R, RStudio, Latex, Packages, and so on? Have you used Github regularly to post updates to your "class work history"? Have you installed Dropbox basic (free edition: https://www.dropbox.com/basic)? You will submit a single notebook HTML output (shows a few graphs) and an RDS file that contains the data related to the notebook. I will provide the notebook, you will put it (with all of this other stuff into a "final" folder [living at the same level as "project-measure" but its own folder]). workspaceAudit.html and workspaceAudit.rds

(3) Your formal writeup of Will vs Denzel (worth 100 points). You will submit a raw notebook of your analysis "will-v-denzel.Rmd" and a final-report using a formal writeup "will-v-denzel.pdf"... The final report can utilize the same "report" template as we used for "project-measure" ... you will have to setup that correctly in your final folder. I am giving you an open ended question, and you will build your notebook and report from scratch. You have plenty of other notebooks and report ideas to utilize in your efforts. The PDF report will not have "any code" anywhere, I don't need an appendix. Your audience is your instructor, a data analyst like you. He has asked you a question and wants your opinion. An answer of "they are about the same" or "I can't decide" is likely a "novice" answer. I could see a stellar "mastery" response like that but that would also have to be impressive/stellar analysis. A "novice" answer is about a C, a "mastery" answer is about an A for those of you that are concerned about grades. I would strongly recommend that you provide an overview of the two actors, one strong summary table, and one graphic that aligns with your final answer. The writeup main body should be about 5 pages, 3 is too little and 10 is too much. You can have ENDNOTES, REFERENCES, and TOC that I would not count toward that end. If you have a really important second/third graphic, you could include those in the APPENDIX. I do not want any code in this formal writeup. A good response shows both sides, and does not cherry-pick data on one side to fit a narrative. "Even though X performs better than Y on these factors, I chose Y because ..."

(4) Data audit (50 points). A small portion of this notebook that you will submit as "dataAudit.Rmd" and "dataAudit.html" will be a review and self-assessment of your progress as a data analyst. What skills or tools have you acquired? What feedback would you like to share directly to the instructor? And things like that. Regrettably, this data audit will also be tied to the "primary data collection" for "project measure." There was a lot of replicate data. Maybe it was an error on your part in the XLS --> TXT conversion. So you will need to review the data provided. The other final tasks for this class I consider "public" so they can go into a "final folder" and get posted to Github. This data and its audit should be "SECRET" so adjust accordingly. The biggest "sin" in data analytics would be to fabricate data. And I hope that is not the case. In the real world, you would never work again with a reputable company if you got caught fabricating data or results. If you review the syllabus, there is a section on Academic Integrity, and I have the right to fully implement that policy. You may fail the course if you fabricated data. After I review your audit, and based on the feedback of your peers and the analysis I have already performed, I will formally notify you if I chose to take that path. And it will be a formal submission to the WSU Academic Integrity board. As a course preparing you for the real world, I would prefer not to do that, but I may. It was unfair to your peers. I "jumped in at the last moment" to make certain the data was clean because of several concerns by your peers. I personally felt like I should not be dealing with such ethical/integrity issues at this point. But here we are. The project was intended to be something that could be placed on your resume ... and some of you got really passionate about it and took it very seriously. As a result, it is a portfolio element you can use. As a strong believer in "garbage-in, garbage-out" I could not allow the writeup to occur without first making certain reasonable data was being used. As such, we need to do an audit. Data provenance allows for this. And if you have integrity, you should apologize to me and your peers on Blackboard. You cheated yourself and your peers. And you should discuss in this document what is fair retribution for the violation. Anyway, some of you may be "unnecessarily" hyperventilating. Having a data-entry error on a single element is "systematic bias" it is not blatant or intentional cheating. Having a person that is 5'5" tall with an armspan of someone 6'3" that raises some concerns. And I hope it was some minor error. But you have your original XLS file, your original TXT file. I have both of those. If you have "replicates" you need to explain what happened ... minor infraction ... and what the resulting penalty should be. If you fabricated data for a few "extra credit points" ... major infraction ... and what the resulting penalty should be. An apology is a must in either situation. And a true apology is not: "I am sorry, but ... " A true apology does not try to justify incorrect behavior. Some students in this course cannot walk, yet still were able to measure themselves; some gave the handout to persons diagnosed with COVID; and so on. Anyway, there will be a SECRET notebook you will submit via Blackboard where you will walk-through your data-collection for this project. Sadly, that is a big portion of the "exit interview" opportunity of the final data audit. NOTE: I may remove the current points from Blackboard related to this data-collection for everyone as I review what happened.

Final comment: I enjoyed this opportunity to get in the trenches and code some notebooks for you. I am happy to see the interest in NLP and classifiers, and hope my thought-processes and insights may prepare you for a future that I cannot fully anticipate.

I did create a job-market/resume video (private link) for those that asked: https://youtu.be/AcwtGPs2VVk

As I review the syllabus, there were a few topics I did not cover. I will prepare the "Oh-well" notebook to cover several forms of Regression (linear, nonlinear, probit, logistic) which are obviously used often as classifiers (they can predict). I will also be certain to include mclust2 and the multivariate testing procedures in that notebook. Ultimately, NLP and eigen for features (whether images, words, or social networks) are the future of multivariate analysis, not the past. So I took your feedback and focused the remaining content on those topics.

I do need to apologize to the entire class for allowing the early push to lead to a momentum-killing lull. I lost about 2 weeks of content time during that period. I was getting a feel for your skills and capabilities, but that is not an excuse for allowing the class to slow down. Language immersion with a two-week lull benefits no one. So I do not make excuses, I learned from my experiences, and I apologize to each of you. Most especially to those who were bored-to-tears early on.

I recognize some of these notebooks are very intense with a lot of information. I shared those with you because as an instructor I wanted to give you an information set that would distinguish you from your peers. I did a dissertation on eigenVectorNetwork and shared it with you in a single notebook. That's a lot. Yet, at the same time, when I were your age, I wish I had such a resource. Same with the NLP stuff. I am sharing my life experience with each of you, hoping that it will benefit you in the future. And for those that only understood half of what I shared this semester, it's okay. At least you heard it all. And maybe just hearing it, with some time to process, means you actually learned more than you realize.

I reflect on the delta from Day 1 to now and I am pleased. I hope I influenced you. Maybe not, but each of you certainly influenced me. It is very satisfying to hear new/novel ideas from each of you on your reports and in other discussions; that is how innovation is possible. Personally, I wrote an "n-gram" function that I literally paid workers over $100K to perform; I was able to document my dissertation in a simple notebook with some basic functions; I was able to document some important ideas about "smart-classifying"; I was able to harvest a dataset from IMDB to study the social-network effects of Hollywood. And I was able to be transparent and show you my non-linear way of thinking as I shared by real-time problem-solving strategies.

I am a problem solver, that is my strength. I like hard, multivariate (and multi-dimensional) problems. That may not be your thing. Your thing may be people. That is good, data analytics needs good "liasions". Some of you may like the data wrangling; others may hate it. Some may like structured data formats; others may hate it. It's all good. What matters is that you can recognize who you are as a data analyst and you make efforts to harmonize with that as you try to find a job that aligns with your true self. Some of you have no idea who your true self is ... at this point. Sometimes getting out of our comfort zones helps us find it. Problem-solving and teaching mathematics was not too far afield for me personally, but moving to California and using programming/internet technologies to solve problems and visualize solutions, that was the solidifying self-awareness event for me. I was about 28 when I formally defined myself: "Monte from Montana is a strong, sober mind that likes to solve problems in order to help people." I challenge you to take a moment over the winter holidays and reflect to try and understand and define yourself. And take a moment to thank those that have helped you on the quest (you can review some of this in my dissertation: pg 5-; pg 27).

Best Wishes to Each of You on Your Unique, Individual Life Journey.