I. Final Project Guidelines

Your paper should have the following elements (or something like them).

1. Introduction. Explain what topic you are studying and what question in particular you are asking; this require you to say what are your main independent and dependent variable(s). Then state your hypothesis about how the IV and DV should be related. No literature review is required at all.
2. Description of Data Set and Variables. If you are using the GSS, you can omit this description of data (I know what it is), but if you are using some other data set, I want to know how it was collected, how is it representative (or not), etc. Everyone must tell me what variables you are using and how they have been coded and recoded.
3. Descriptive Statistics. You should have a table with the means, st.devs. etc. of all the variables you are using. With explanation and commentary.
4. Initial Models. Tell me what model you are using and why (logit, probit, LPM, fixed effects, etc.). Start off with a simple model relating you main IV to your main DV. Explain the relationship and why this initial model is insufficient. Maybe you need to make a scale/index of variables. Maybe you need to control for additional factors. Maybe you want to include interaction terms. Maybe you need to check for serial correlation. Etc. Interpret everything correctly (ceteris paribus, on the right scale, etc.)
5. Final Models. Tell me what models are your final best ones. Why are they better than the initial ones and what else do we need to know? Were there some intermediate steps to get to this final model. Maybe have some diagnostics and checks.
6. Conclusion. Tell me what you learned from doing the project. Was your initial hypothesis supported or not. Why or why not? Tell me what else you wish you could have done. Tell me how this work may inform your thesis. Etc.

All in, there is no specific page expectation. I would expect you would need 7-10 tables and approximately 10 pages of text. But you should do as much as you can to satisfy yourself and your curiosity on your topic.

II. Additional Question

In addition, here is a list of additional questions you could ask about your project, as you are putting it together. Do not be overwhelmed by the number of questions here. You don't have to necessarily do them all - but I wanted to give you a sense of the issues you should be dealing with.

1. Have I put a "because" after most of my sentences to show the logic of my statements?
2. Have I motivated the paper sufficiently? Have I explained why I am studying this topic? Have I explained how I expected the *key* variables to perform (direction or sign, magnitude, statistical significance)?
3. Have I been sufficiently explicit about the (re)coding of all of my variables? Have I recoded my variables in the most appropriate way (e.g., changing categorical variables into dummy variables of some kind, logged income or income in 10k of dollars to make easier to interpret)? Have I gotten rid of all meaningless categories, like missing values or don't knows? Am I sure I know what my variables are - did I double-check their meaning from the SDA Archive full question text, or in the real pdf codebook?
4. Have I thought to include appropriate control variables? Typically, researchers include a mix of the following. Demographic variables (race, sex, age, region, size, marital status). Socioeconomic status factors (work status, income, educ, prestg80, class, & sometimes parental SES like paeduc, madeg, incom16, papres80, sibs, etc.). Cultural factors (relig or relig16, attend, pray, polviews, partyid). Personality type traits (coop, happy).
5. Have I explained why I used this given statistical technique (ordinal regression for ordinal dependent variables, etc.)?
6. Does my model make theoretical sense (I use variables that happened prior to the events being predicted? - like I predict your education based on your father's education, not on your current income, for instance.

7. Have I considered possible interactions between my key X variables and other Xs in the model? Have I subset my data along key lines (running separate regressions for men vs. women, or whites vs. blacks, or high SES vs. low SES)? Why would models be different between different subgroups anyway (is R-sq different? are Bs of very different sizes or directions?)?
8. Have I been careful to interpret everything in the most complete, appropriate way (saying, with all else held constant; indicating that certain variables only take on certain ranges; or (for predicted probabiilities from logit, with all else set at their mean)?
9. Have I interpreted my key variables throughout the paper, but have I also discussed some of the other variables to indicate that the model "makes sense" because those other variables are in the expected direction? Have I talked about coefficients, statistical significance and R-sq in most of my tables?
10. Have I explained (though not necessarily shown) the intermediate steps/choices I made on the way to having a number of preferred models?
11. Have I highlighted the limits of my analysis? Have I explained how my variables are sometimes only proxies for the key underlying concepts I am trying to understand? Have I noted other analyses that I wish I had been able to perform but couldn't, either because Greg didn't teach it yet or ran out of time/steam? Have I shown that I am aware that correlation is not causation, but have tried to think how to make the regressions as rigorous as possible? Always think about what the experiment to test your question would be - and why your regression may fall short (or not) of that ideal.
12. What would be the next logical/possible steps with the paper beyond what was done here? Let me know if you are serious about pursuing the topic because I will try to help moving forward.