

Gege Gui

Please be sure that you look over this checklist and make sure you have addressed all of the issues when writing up your final project Qual checklist.

- Make sure that you have a title that highlights the point of your data analysis, for example: “An analysis of data from fivethirtyeight shows Nate Silver is a good statistician”.
- Have a full introduction where you explain why the problem you are working on is of interest, what is the general approach you take to solve it, what methods you plan to employ, and summarizes your results.
- Make sure you include citations in your introduction for any claims you make.
- Be extremely concrete on how you selected the words that you were going to use in your search? What words precisely did you use? How did you choose them? Did you use exact matches or did you allow for fuzzy matching. Be precise in how you collected the data.
- Make sure that all of your figures have figure captions. The figure captions should explain the purpose of the visual and what it describes - for example, “Figure 1. A plot of x versus y. There is an increasing trend in x with y.”
- When describing how you use an R package like rvest, make sure that you explain what you did - did you pull the webpage? Report the exact CSS/Xpath that you used to scrape, the time you used to scrape, etc.
- Make sure you cite R packages using a reference manager (like Bibtex).
- Do not include any R code in your presentation or any R output. If you create a table in R, you should format it as a table in your results.
- Make sure you summarize your final data set you use for analysis. What variables did you collect? How many observations do you have? How much is missing?
- Do not refer to data by file names, make sure you describe each data set, the variables in it, the dimensions, and how you got it in words.
- It appears you trained your model on just one data set - PloS Currents is there any potential bias here? Why not train on a broader subset?
- What can you say about the time trends? Are the differences between journals? Any broad patterns you see? Can you quantify them beyond just visual exploration?
- Make sure you include a conclusion that summarizes what you did and explains the strengths and weaknesses of your analysis.