MOUNTAIN TIGER Reflection

1. How well the team project went overall.

Overall, team Mountain Tiger's final project was a success. At this point in the semester, we knew how to work cohesively as a team as well as the strategies that worked best for us to complete the project. For example, after project 1, we decided to meet as a group right before the project is due and gather all the needed files as a group, so everyone submits the same files. This strategy worked well in terms of everyone remembering to submit everything and avoiding confusion about file names. So, we used it again for the final project by planning ahead of time to meet on the 14th and compile everything we need to submit as a group. Much like our previous projects, we communicated well with each other and divided the workload as evenly as possible.

2. How the research questions were addressed as the project was completed.

In the preliminary project process, we, as a group, decided on an array of possible questions we would like to analyze based initially off of Nolan's presentation. Then, after Chris had merged the data, we had to rule several questions out due to some variables having to be omitted from the combined dataset. We then addressed what statistical methods or visuals would make sense to answer each question (Nolan had some of these already planned out in his preliminary project proposal). For most of our questions, we opted to use scatterplots along with linear regression models and Pearson correlation coefficients to understand what variables did and did not have a relationship with suicide rates.

3. What challenges were faced and how were they overcome.

One of the main challenges that was faced was when merging the three datasets there was a lot of loss of usable data. This was because one of the data sets was organized by country by year by sex by age by generation and another data set was organized just by year and country. For these two data sets to be merged the data set that was organized by five levels had to be aggregated into only two levels, ie. organized by just year and country to then be merged with the other data set. By doing this the observations shrunk from about 3000 to only around 1000. Another issue was that the countries and years between the data sets were not the same so they had to be compared and only the common countries and years were merged, shrinking the data set down to around 900 observations. Finally in the process of merging these two data sets there were found to be some countries that had large amounts of NA values in their columns to the point where they had to be removed. Finally the number of observations had now shrunk down to 800. This then caused the sample points for the analysis to be very small and even caused imputation to be required for some values for any analysis to be viable.

4. What went well on the project and what could have gone better.

Our team process was certainly the best part of our project. All semester we have had a very cohesive, well-maintained team that distributes work evenly and has well structured meeting times and deadlines to hold each other accountable. That was very much true within this project and is something that we believe was a major factor in our success working on and writing about our topic. We also all put good effort into our individual proposals which led to having 4 great topics to choose from for the focus of our final project, which then led to choosing one that not only made good material to analyze, but also lended itself to important, global research questions that mean something to us.

The biggest thing that we believe we could have done better was choosing datasets either more carefully, or looking further into our chosen datasets when choosing a topic. While we believe we were still able to do great work, there was more data loss in merging than we initially anticipated based on the compatibility of the three datasets that we had chosen. This also implies a relatively high amount of missing data, especially in the infant mortality, mortality under 5 years old, and average house-hold income variables explored by Katey. This required the use of KNN imputation, which, while it does happen, would ideally be avoided in the outset by choosing better datasets or possibly choosing a topic with more available, cleaner data to begin with.

5. What are the differences between the preliminary and final Gantt charts and why these differences occurred (see below).

One of the biggest differences between our preliminary and final Gantt charts is the increase in the number of items we listed. The preliminary Gannt chart has 10 items divided between the four of us whereas the final Gantt chart has 18. Part of the reason for this increase is that the final Gantt chart ended up dividing some of the items we had listed as a “team” item in the preliminary chart and assigned them to specific people. This division happened naturally as we worked on the project, especially since we are now well-accustomed to working together and communicating about who will do what part of an assignment. We also received the specific instructions for our final project after completing the preliminary Gantt chart, so some of the tasks in the instructions (such as compiling a readme.txt or an “ideas” directory) were added as needed. This final Gantt chart also has more items than some of our other final Gantt charts, not necessarily because this project is the largest but because of the chaos inherent to finals season. Because we knew that early December is hectic and busy for everyone, we decided to spend some extra time being more specific than usual about who would do what. This strategy helped us plan ahead accordingly and prevented any last-minute scrambling to finish parts of the project. Another difference between our final and preliminary Gantt chart is that in the final chart we split the workload into three “parts” based on the date. We wanted to have our preliminary data analysis done by December 8th and then gave ourselves until the 14th to get the report written up. The third part of our workload is condensed into the last couple of days before the due date, where we planned to meet in-person to create and rehearse our presentation video as well as complete the various project evaluations. The preliminary chart does not have the same divisions based on the date. We intentionally divided up the workload into two “large”

parts and one “small” part because we knew everyone would be busy around finals and wanted to give everyone plenty of time to finish their parts of the project before we met to create the video.

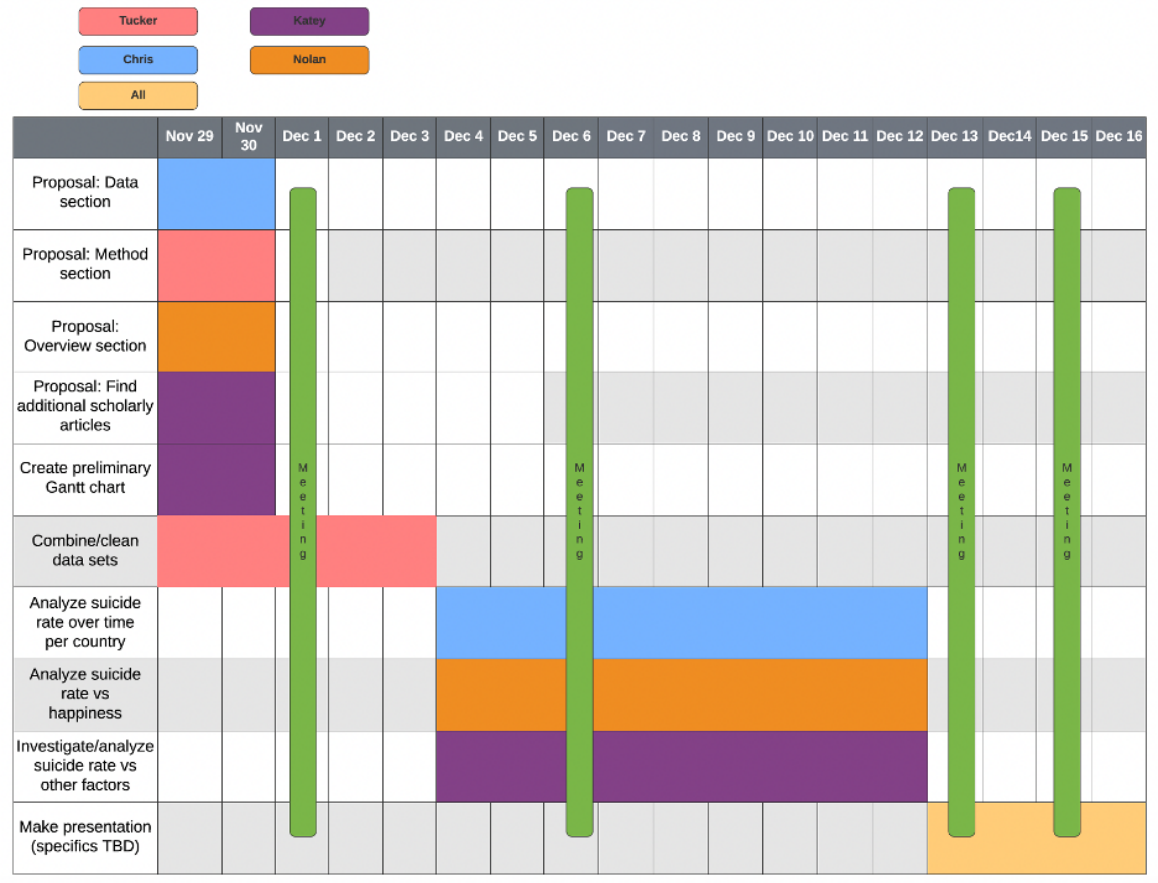

Figure 1 Preliminary Gantt chart, demonstrating the original planned division of labor among Team Mountain Tiger

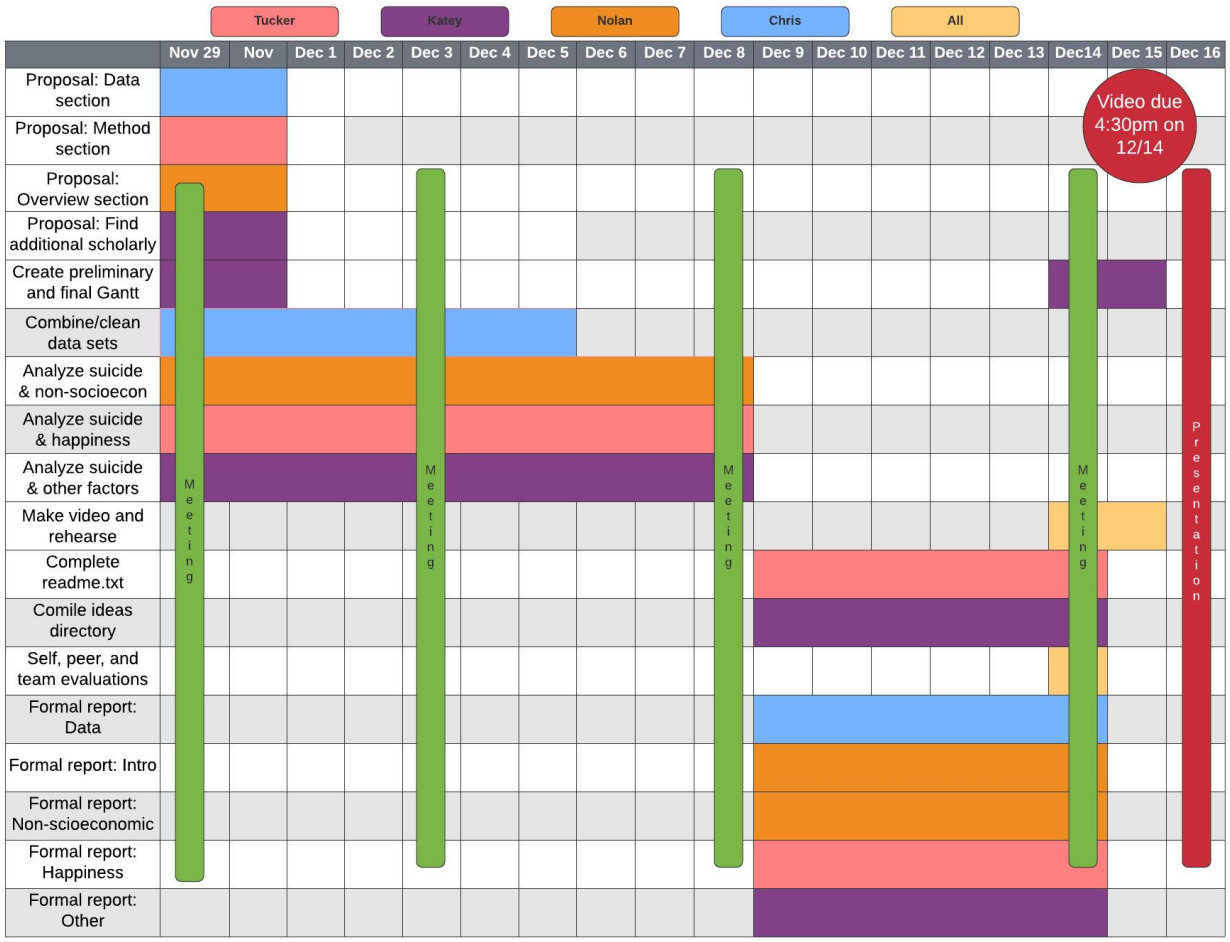

Figure 2 Final Gantt chart, demonstrating the final division of labor among Team Mountain Tiger