Running head: PORTFOLIO PROJECT - OPTION 1: R DATA ANALYSIS

Portfolio Project - Option 1: R Data Analysis

John Imbur

Colorado State University – Global Campus

August 4th, 2019

PORTFOLIO PROJECT - OPTION 1: R DATA ANALYSIS

Portfolio Project - Option 1: R Data Analysis

For my portfolio project I am electing to use the General Social Survey (GSS) which can be found at [https://gssdataexplorer.norc.org](https://gssdataexplorer.norc.org/). This longitudinal study was first begun in 1972 and continues today, measuring and recording a wide set of factors for the specific individuals in the study. Captured in the overall GSS are personal facts about marital and family details, income, race, age, as well as attitudes about an array of different social concerns. For the Portfolio Milestone in Module 4, I chose to look at questions related to the correlation of marital status and happiness, particularly as they may relate to personal details like age, race, religion, politics, work, income, and dependents.

Once I set up an account with NORC, which manages the GSS, I used their GSS Data Explorer to see what data might be available. I requested and received the GSS data from the years 1972 to 2018 in XLS format for the 55 variables I identified as possibly useful. The dataset I received has a total of 64814 records with three tabs comprised of the raw data, an explanation of the specific questions included in the request, and how they were coded for the survey initially.

Once the requested data had been delivered, I looked over how it was laid out, what I would use, and what data transformations would be required. I decided that the religion and political persuasion questions were beyond the scope of what I would need in looking at marriage and happiness. I also weeded out race as all that was captured was black, white, or other and I felt like that was less information than I was comfortable using. Lastly, the details about income, work, and dependents also seemed beyond what was needed for this initial data analysis and were not used.

PORTFOLIO PROJECT - OPTION 1: R DATA ANALYSIS

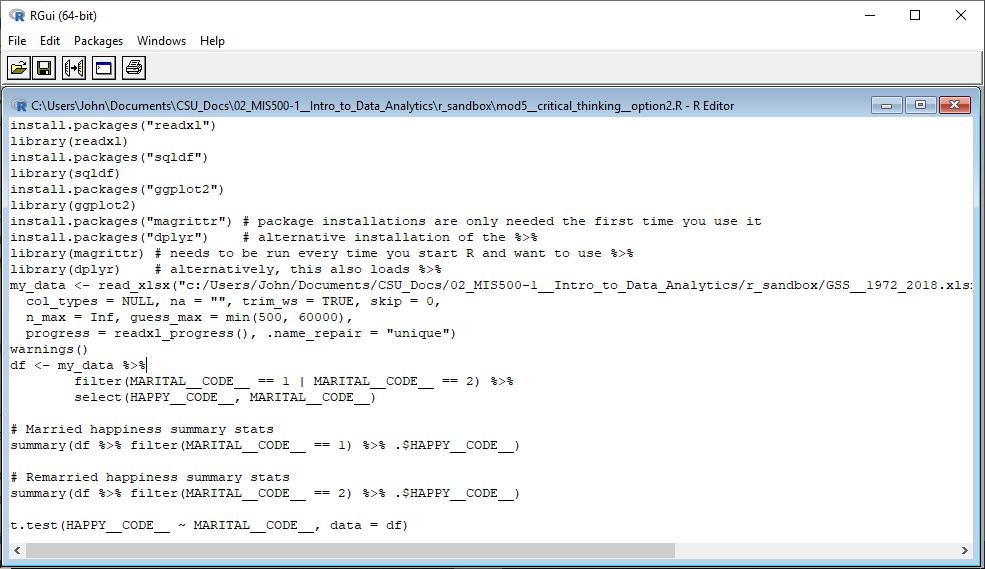
Most of the values to be used then needed to be transformed to a code, specifically individual and marital happiness, work and income satisfaction, social class, and marital status fields. This was done via a simple Excel conditional formula for each field, and I used the naming convention of the field name with the suffix “\_\_CODE\_\_” added. I also added these transformed field names into the second worksheet along with the related conditional formula for later reference as needed.

For storage, the current file includes data from 1972 to 2018. Each year after would require logging in and requesting a new dataset and then ported into the relevant analytic and visualization tool. While the lions share of the data requested and received was not used for this specific project, there are all sorts of further work and study that could be done with the original 55 fields worth of total fields requested. The current file and future data requests would be included in the Portfolio Project section in the GitHub repository along with any resulting reports, taking care to name the data file so that it is apparent the dataset that it comprises.

As stated in the initial proposal in week 4, the question that was to be investigated was whether the degree of participant happiness varied between married participants and divorced participants who have since remarried. This was modelled off an earlier study that was performed with a Denmark population, and I wanted to see if I could replicate the results with an American population. (Hiyoshi, 2015) Happiness in the GSS is measured both for individuals and specific to marriage happiness with the following values: Not Too Happy, Pretty Happy, and Very Happy. I transformed these values to 0, 1, and 2 respectively. Remarried participants were identified as those both marked “Married” and also showing that they have been divorced.

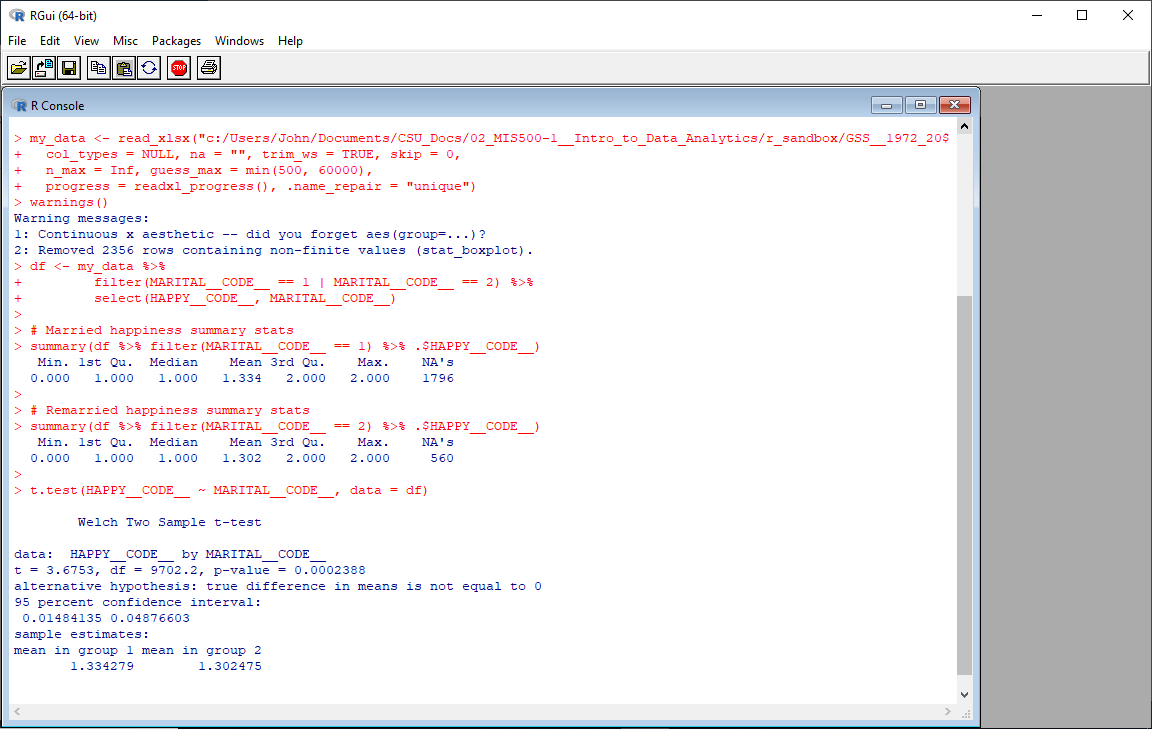
PORTFOLIO PROJECT - OPTION 1: R DATA ANALYSIS

The Null Hypothesis is that there is no statistical difference between the individual happiness value for participants, whether married or remarried. If the Null Hypothesis fails, then the Alternate Hypothesis is that the married participants that have not been divorced are happier. Below are the screen shots of the R code and the output comparing the records with the MARITAL\_\_CODE\_\_ field at 1 and 2 which represent “Married” and “Remarried” respectively. The summaries and the results of the t-test are shown with the interpretation following.



*Figure 1*. R code to generate the both the summaries as well as the t-test result.

PORTFOLIO PROJECT - OPTION 1: R DATA ANALYSIS

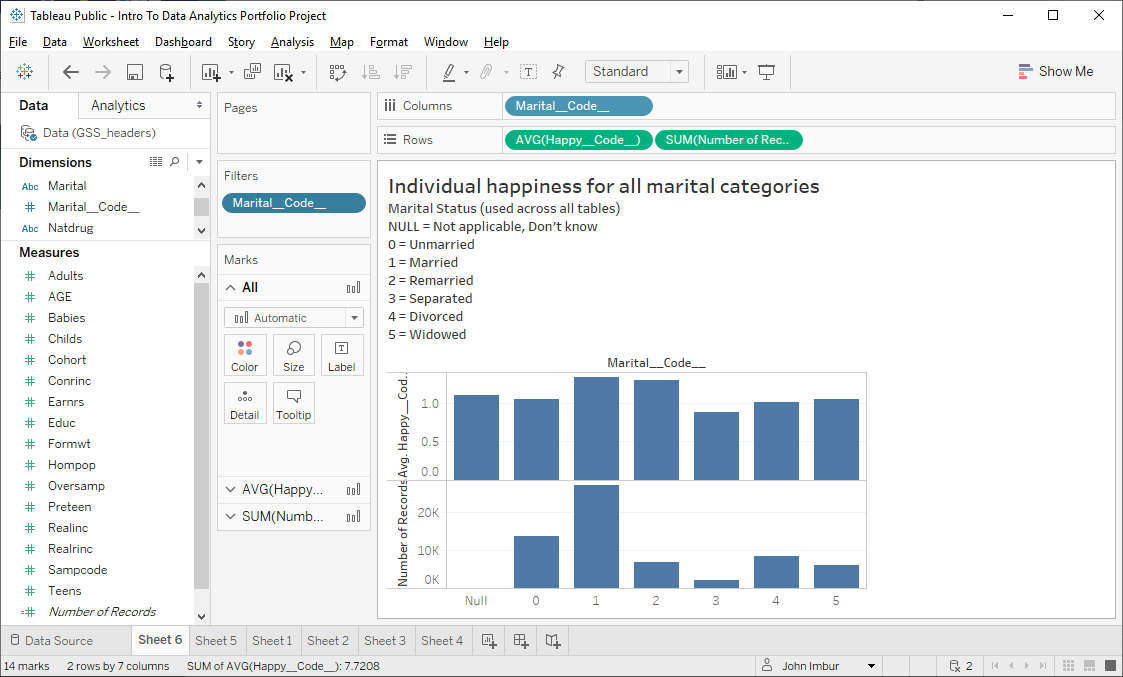


*Figure 2*. R output of both the summaries as well as the t-test result.

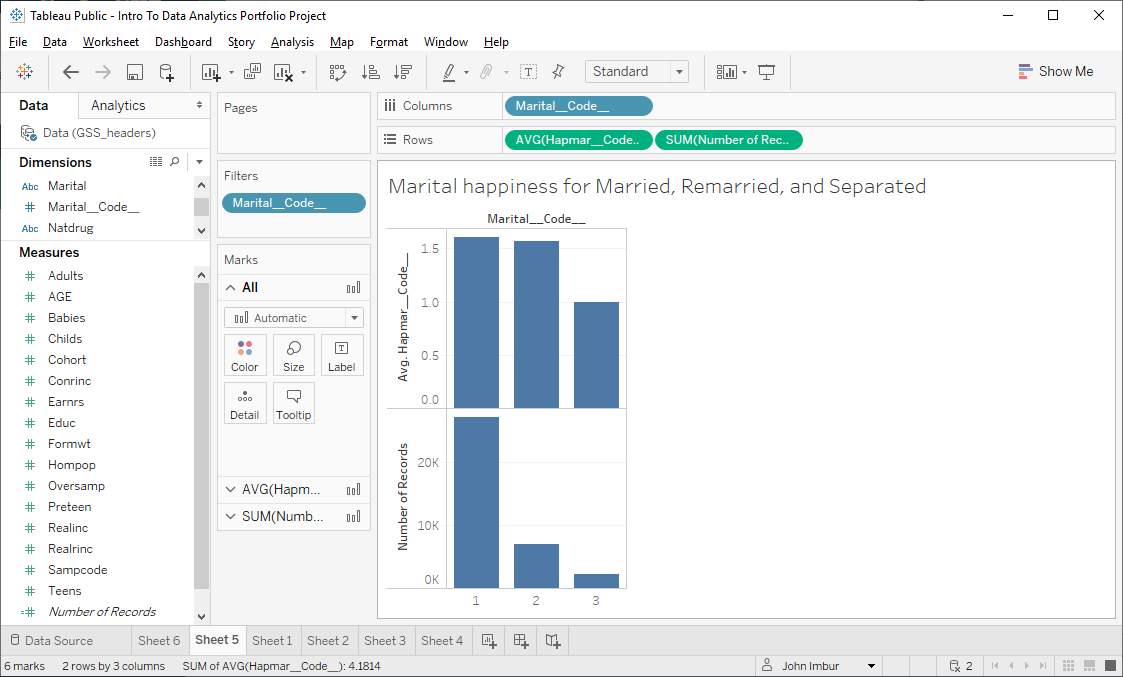
Based on the P-value being well below the standard .05 threshold at 0.0002388, the Null Hypothesis failed. On top of that, the mean of the married cohort is in fact larger than the group of remarried participants, so the Alternate Hypothesis passed.

Below are some graphs generated in Tableau that show both the married and remarried participant values relative to both individual and marital happiness values. I also ran reports, also in Tableau, of the individual happiness grouped by marital status against the following variables: gender, social class, and satisfaction of both job and income.

PORTFOLIO PROJECT - OPTION 1: R DATA ANALYSIS

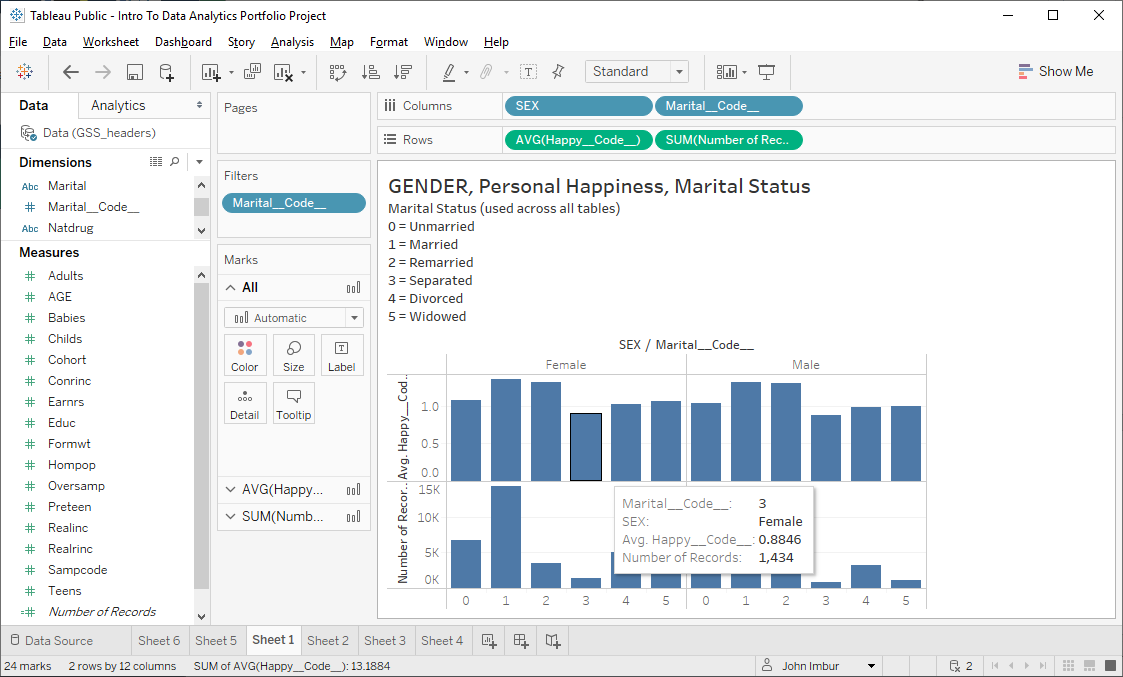


*Figure 3*. Individual happiness for all marital categories, including over total count.

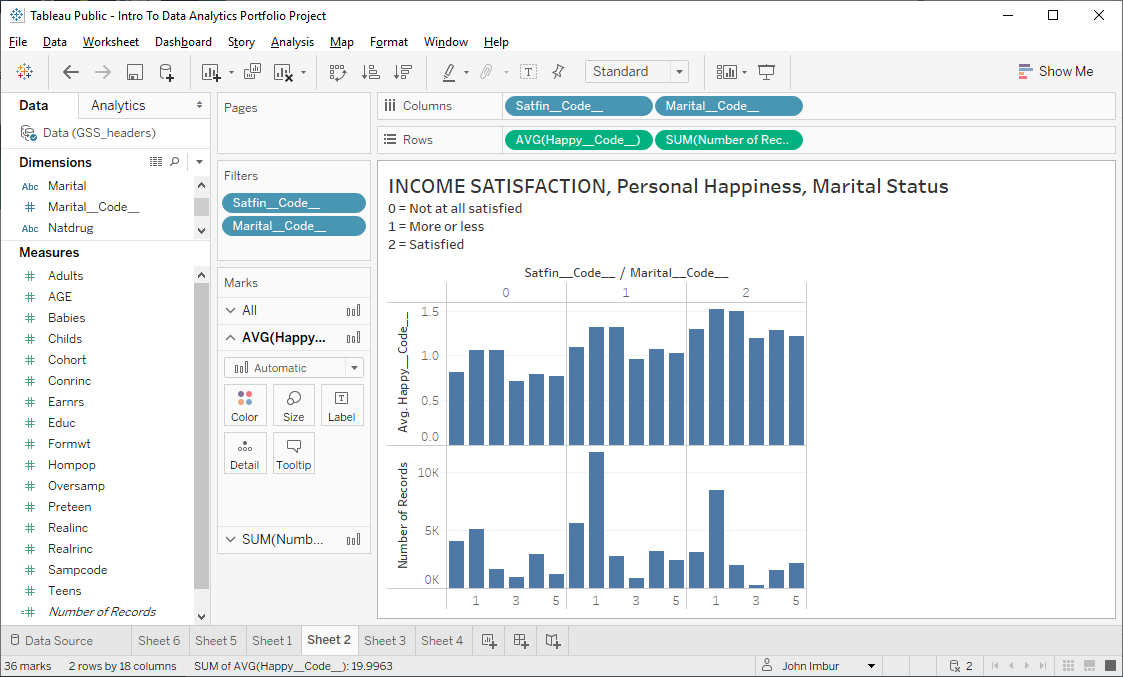


*Figure 4*. Marital happiness across relevant categories.

PORTFOLIO PROJECT - OPTION 1: R DATA ANALYSIS

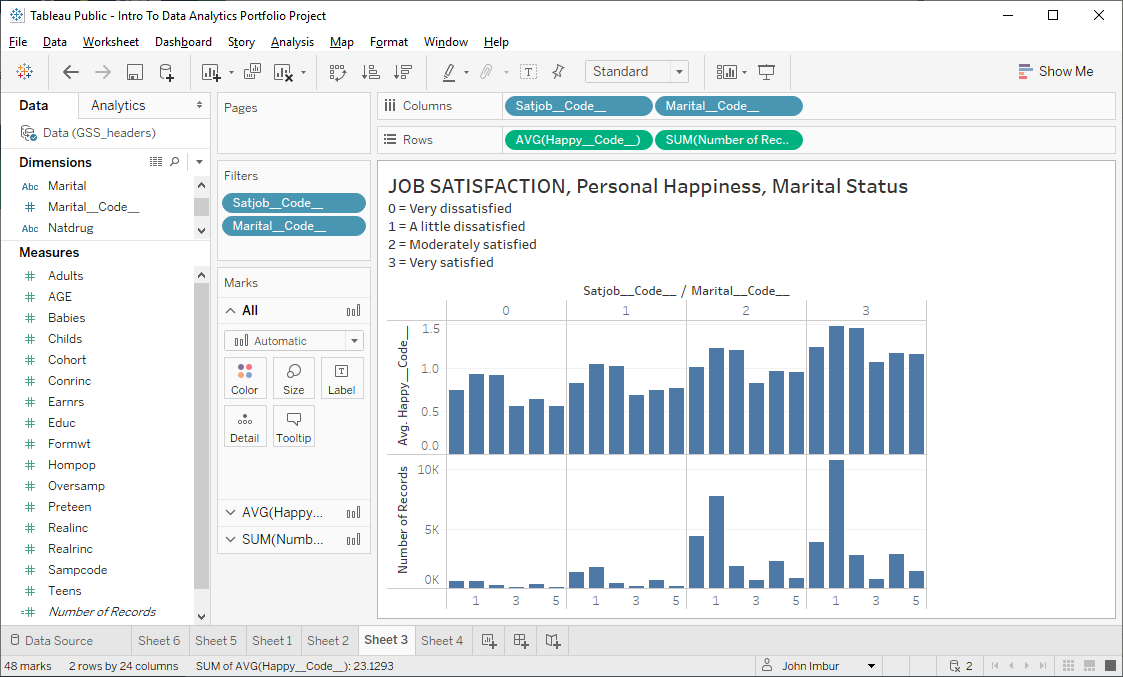


*Figure 5*. Gender, Personal Happiness, Marital Status.

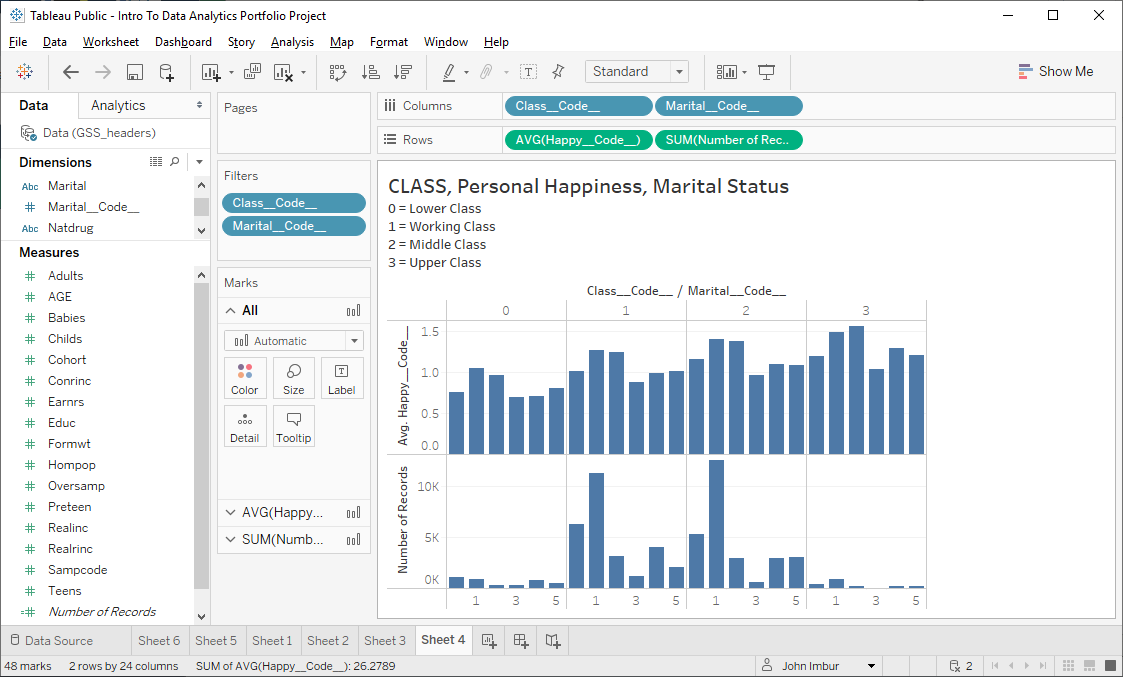


*Figure 6*. Income Satisfaction, Personal Happiness, Marital Status.

PORTFOLIO PROJECT - OPTION 1: R DATA ANALYSIS



*Figure 7*. Job Satisfaction, Personal Happiness, Marital Status.



*Figure 8*. Class, Personal Happiness, Marital Status.

PORTFOLIO PROJECT - OPTION 1: R DATA ANALYSIS

This was a challenging but interesting project and I am glad to have forced myself to stay on target and complete the original topic. I am not very comfortable with the large difference between the overall count for married versus remarried based on the overall divorce rate in America, but these were the values in the dataset. It was interesting how much happier married people are, regardless if remarried, compared to the other categories. It makes sense that people going through a separation would be consistently the lowest.

One thing that caught my eye was the table keying off social class and I could see where looking closer at income might be fruitful further research. Each step up in social class was consistently happier, so much so that someone in the Upper Class going through a separation was equally happy as someone married for the first-time person in Lower Class. I am curious if the number of dependents has a similar correlation.

In addition to GitHub at <https://github.com/johnimbur/CSU-GLOBAL__MIS500>, you can peruse the worksheets created in Tableau at: <https://public.tableau.com/profile/john.imbur#!/vizhome/IntrotoDataAnalyticsPortfolioProject/Sheet1>, <https://public.tableau.com/profile/john.imbur#!/vizhome/IntrotoDataAnalyticsPortfolioProject/Sheet2>, <https://public.tableau.com/profile/john.imbur#!/vizhome/IntrotoDataAnalyticsPortfolioProject/Sheet3>, <https://public.tableau.com/profile/john.imbur#!/vizhome/IntrotoDataAnalyticsPortfolioProject/Sheet4>, <https://public.tableau.com/profile/john.imbur#!/vizhome/IntrotoDataAnalyticsPortfolioProject/Sheet5>, <https://public.tableau.com/profile/john.imbur#!/vizhome/IntrotoDataAnalyticsPortfolioProject/Sheet6>

PORTFOLIO PROJECT - OPTION 1: R DATA ANALYSIS

References:

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research, 12*(3), 229–238. doi: 10.1023/A:1023254226592

Gardner, J., & Oswald, A. J. (2006). Do divorcing couples become happier by breaking up? *Journal of the Royal Statistical Society: Series A (Statistics in Society)., 169*(2), 319-336. doi: 10.1111/j.1467-985X.2006.00403.x

Hiyoshi, A., Fall, K., Netuveli, G., & Montgomery, S. (2015). Remarriage after divorce and depression risk. *Social Science & Medicine, 141*, 109-114. doi: 10.1016/j.socscimed.2015.07.029

Horner, E.M. (2014). Continued Pursuit of Happily Ever After: Low Barriers to Divorce and Happiness. *Journal of Family and Economic Issues, 35*(2), 228-240. doi: 10.1007/s10834-013-9366-z

Lee, S. (2018). Removing the Stigma of Divorce: Happiness before and after Remarriage. *IDEAS Working Paper Series from RePEc*.

NORC at University of Chicago. (2016). About the GSS. Retrieved from: <https://gss.norc.org/about-the-gss>

Whitton, S., Stanley, S., Markman, H., & Johnson, C. (2013). Attitudes Toward Divorce, Commitment, and Divorce Proneness in First Marriages and Remarriages. *Journal of Marriage and Family, 75*(2), 276-287. doi:10.1111/jomf.12008

Yan, X.Y., Huang, S., Huang, C.Q., Wu, W.H., & Qin, Y. (2011). Marital status and risk  
for late life depression: a meta-analysis of the published literature. *Journal International Medicine Research. 39*(4), 1142-1154. doi: 10.1177/147323001103900402