

William Bernard
SOC 321
December 11, 2017
Victoria Sass

Poster Project Report

1 Project Topic and Relevance

I had come across an article that discussed the effects of loss of a loved one on life satisfaction in elderly individuals about mid way through the quarter. Initially I had started looking at the Americans' Changing Lives data set with entirely different intentions. But one I saw that article it lept out to me. I was interested in investigating the effects of loss on life satisfaction. The next question was, how do I bring something new to the table? I tried to answer that question by doing two things. boardng the scope by eliminating age restrictions (looking at individuals from a much wider age range) and by focusing in on a certain type of loss, in this case the loss of a child.

There is a somewhat lacking amount of research that has been done in this field, and it became increasingly apparent as my project evolved that I wasn't just looking at something more than just the effects of child loss on life satisfaction. The disseminating effects of the loss of a child and its impacts on things like marital status, compound to affect and individuals life in different and interesting ways. While my initial question may have seemed somewhat surface level, the investigation has led to some substantial questions that have yet to be answered by the current literature.

This project was not directly associated with the work I will be doing for my honors thesis, even so the topic wass of interest to me and I am glad to have had the opportunity to

investigate it. I feel that the work I did, while not groundbreaking by any means, did its job of creating new and interesting questions that could be explored in the future.

2 Finding Data

The way I found my data was a little backwards. Initially I was interesting in investigating depression and dementia in aging individuals. Towards that end I found the data set, Americans' Changing Lives. I soon realized that the data was not exactly what I needed to investigate the kind of question I was thinking of asking. That being said the data was still intriguing. I took some time to look through it and the measures of life satisfaction jumped out at me. I thought this was a great opportunity to further investigate the ideas I had read about, and decided to switch the focus of my investigation while keeping the same data set.

Ultimately this data set proved to be exactly what I needed. It had concrete measures of life satisfaction, and asked questions pertaining to numerous instances of varying types of loss experienced by the individual (ex. Terminal illness, death of parent, death of spouse, divorce etc.). It was clear to me that this data set answered a lot of questions I wanted answers to. In the end I decided to narrow my focus, honing in on the loss of a child (then controlling for marital status), but there were many other avenues I could have taken if I had more time to investigate further.

Originally I set out to answer a more broad question on the effects of loss on an individual's life satisfaction. Realizing that was too broad I chose to focus on the loss of a child, both because it is not a life event that is not usually associated with age, and because it represented a different dynamic than the loss of a spouse. Over the course of the quarter however after running my regressions, it became clear that marital status was more center

stage. My project evolved naturally, and ultimately how I chose to visualize the data was influenced by the most significant variables I found.

3 Data Structure and Data Munging

Organizing the data was one of the more difficult components of this whole project. Not because I was combining multiple data sets, but because of the sheer size of the data set I was working with. My data set had over three thousand observations and over four thousand variables. Naturally it was a requirement that I reduce that to a manageable number. I spent a considerable amount of time grinding through the code book, finding variables that I thought would be relevant and creating an alternate data set with only the variables I thought I might want to use. Even after doing that I still had something like 60 variables and that was just in the first wave alone. Each subsequent wave asked the same questions. I ultimately decided that I was going to focus on the first wave because subsequent waves lost more and more respondents to death or non-response. That left me with about 60 variables to play with. I used that new data set that I had created for the rest of my time with Americans' Changing Lives. It made it a lot more manageable, and after I changed the names of all the variables it was a lot more accessible too. Luckily for me the data was tidy and did not require much alteration (at least the variables I was using).

The visualizations were fun to create. The most time consuming thing in that process was coercing the variables into numerics for the purposes of plotting (given the categorical nature of many of the questions). Once I had done that it was a matter of displaying the variables I wanted to display in the clearest and cleanest way possible. Changing the way the data was represented was also important. For example, one of the final visualizations facets for

marital status but instead of using counts each bar is representative of the percentage of that facet. This is a much better way to compare the distribution of responses for each facet. In addition to that I decided to forgo the y axis labels and instead label each bar with its corresponding percentage. I thought this was a good way to clearly articulate to the viewer the information exactly.

The statistical analyses were interesting as well. Introducing each new variable one at a time gave me interesting results. At the beginning when though I had an upward trending correlation between child death in the last three years and life satisfaction I was excited. Even though that turned out to be nothing significant It still reminded me of the value of statistical modeling and the power of this kind of research. I was well and truly surprised by how exciting numbers can be.

Some of the coding was a challenge occasionally. I often found myself squinting at the screen and repeating “why?” as my chunk flashed red. I persevered however, and kept at it until the job was done.

4 Reproducible Research

The value of reproducible research is in its transparency. The fact that anyone can look at your findings, take your data, plug it in for themselves and check the legitimacy of your work is invaluable. Without the peer review process it is impossible to call any research academic. Making your research reproducible both helps the academic community grown and provides a check on any kind of misconduct that may be occurring in the academic world.

For the purposes of this project I attempted to make my work reproducible in three significant ways. The first was was making it abundantly clear where I got my data from. Doing

this allows others to access the same information you are. The second was uploading all my progress and code to github. Doing this allows anyone to use or take my code for the purposes of checking my work (encouraging complete transparency). The final way I made my work reproducible was by utilizing packrat. Packrat allows other individuals to run the data using my code no matter the software they use personally (if for example they have updated past the point where my code is operational).

The most useful technique by far was maintaining a github repository. Having all my work in an online repository is convenient as well as transparent, and in that way kills two birds with one stone. Moving forward I plan on continuing to use github whenever possible.