Cate Seymour

Professor Prince Nelson

DS 401

February 25, 2025

In order to obtain a data science minor at Washington and Lee, I must take at least 19 credits, with no more than 9 coming from my Cognitive and Behavioral Science major. For my credits in data science foundations I took CBSC 240 (now 309), Data Science: Mind Analytics; for my credits for statistics I took CBSC 250, Statistics and Research Design; for my computing and programing credits I took CSCI 111, Introduction into Programming; and finally for my elective credits I took SOAN 276, Science and Art of Surveying Research, CHEM 116, Imaging Science in Art, Medicine, and Astronomy, and CBSC 359, Advanced Methods in Cognition and Emotion Research. These classes have allowed me to complement and advance my skills in psychological research and conduct data analysis in an efficient and ethical manner.

My final portfolio includes three projects that showcase the array of tools I have developed while completing my data science minor. The first of these projects was titled, “Does Having Children Affect Religiosity?” and it was for my Science and Art of Surveying Research class. My final project was to interpret some of the data in a GSS data set given to our class. I used SPSS to investigate whether mothers and/or fathers became more religious when they had a child/children. My thesis was that the GSS data will show that with more children the participant has, the stronger their religiosity will appear to be. On SPSS I chose the "child" variable as the DV; "attend", "pray", and "reliten" variables as IVs and performed cross tabs and ANOVA tests

to see if any trends appeared. Based on these arguments and prior research, I believed that I could sustain my thesis that your religiosity increases with each child you have in your care.

The General Social Survey (GSS) data sets have been conducted for around 44 years and the group who organizes it (NORC) is very respected in the social science world. Participants are adults that are a part of households in the United States. They are randomly selected to participate and have the choice to consent or say no. Participants are also told what the survey data is used for, which in this case is to get people 's opinions and demographic information on certain issues that tend to be important in politics and everyday life. Participants can also skip any questions they do not feel comfortable with answering. No names are used in the survey and the government or other organizations are not able to access personal responses. The GSS data sets are never sold to companies for its data and the survey is only shared in summarized form as seen in the all data sets. All this information can also be found on the GSS website very clearly.

My next project is titled, “Do Happier People Take More Risks?” and it was My final project for Statistics and Research design. This project used Qualtrics and R to see if when given a happy video (vs. sad and neutral videos) people were more likely to take risks. This study was based on past studies looking to see if mood influences risk-taking behaviors. This study was a between-subjects design where participants were randomly assigned into a group when taking a Qualtrics survey I created. These results were then put into Excel and R to interpret if the hypothesis was correct using averages, creating graphs, and ANOVA tests which found that it was actually those shown a neutral video who were more likely to take a risk.

Since I was the one to make the Qualtrics survey, I was sure to include an overview of what the survey would entail, along with a question that asks the participant if they consent to being a part of this study. If a participant consented and completed the study, they were informed

at the end of the study what exactly my group was looking to investigate and how their data would be used – in this case, if the video they were shown would affect their risk-taking behavior. Participants were randomly assigned what video they would be shown. It should be noted that if a participant was randomly assigned to the sad video group, the video seen might make the participant feel sad, but all videos were only 60 seconds long and research was done beforehand to make sure the video made participants feel sad momentarily but not depressed or hopeless as that was not the goal of the study. I also made sure to make the survey anonymous so no names were present. Only myself could access the survey data so no other groups or third parties could take participant's data.

My third and final project is, "Speed Dating: Hobbies and Matching" which was my final project for Data Science: Mind Analytics. This entailed doing research off of data from researchers who decided to look at adults to see what attributes they look at when deciding whether to date another adult. The result of this was to have groups of singles in different waves have 4 minute "Hurry Dates" with other single adults where participants would rank their partner based on 6 attributes on a likert scale: attractiveness, sense of fun, sincereness, ambition, intelligence, and shared interests/hobbies. After each date, each person would decide whether or not they would want to "match" with their date in the future. My partner and I used R to see if people having similar hobbies were more likely to match during speed dating, and found that we were able to group hobbies into a 3 factor model that includes "the arts", "sports", and "games" and participants were more likely to match with someone with similar hobbies.

When undertaking this project, my partner and I were able to learn about certain measures the professors who did this study took to contain confidentiality and safety for their participants in the code book. Participants were informed of the study before taking part in the

speed dating exercise. They had knowledge of how the data would be used, as well as knowledge of the confidentiality, which is present in the dataset. No names or identifying data are present. People could use the analysis of the data to try to manipulate people based on what results prove to be effective for getting a match. However, I do not think that this is of significant concern. We have made sure to test our results for differences between genders. We have not tested for the differences between different racial groups because of the added complexity that those models would entail, however in the original data collection they were sure to take measure of both gender and race for all participants.