Cody Myers

Bellevue University

DSC530 – Exploratory Data Analysis

Final Project Summary

This project was centered around the NSFG pregnancy data set that was used in our Think Stats book by Allen B. Downey. I really wanted to do some personal exploration into this data set as it held a wealth of knowledge and I believed that the data could tell some more interesting stories beyond the one’s described in our textbook. When I looked at the data set, I found a few variables that I felt could challenge my own beliefs with data. For this project, I wanted to look at variables that described a person’s socio-economic status and see if there was any correlation between that and the respondent’s pregnancy outcome and the number of times they had become pregnant.

One aspect that I missed, and could be classified as an issue I faced, in this analysis was that I didn’t choose a data set that was well represented in a scatterplot. Most of my data was categories mapped to numerical values, this resulted in very rigid scatterplots that were sometimes difficult to read into. As I look back though, I don’t know if there was much that could be done for this analysis in terms of making the graphs fit the use case of a scatterplot, even after looking for other variables that could help tell a more definite story.

One assumption that I made that was somewhat false showed during the plotting of the poverty percentage level and the respondent’s number of years of formal education. My original assumption was that there would be a clear definition that as the number of years of education increased, so did the respondent’s ability to further themselves from the poverty line. This ended up only being partially true. While there was a definite correlation between being close to the poverty line and having only the minimal amount of years of formal education, we saw something towards the upper end of the poverty percentile spectrum that I found quite odd. It seemed as if there was a very uniform distribution of respondent’s who lived at the highest end of the wealth spectrum while having any level of education. At this time, I have no speculation as to why this is an aspect of our data. I would be curious though to do a future analysis which looked at the wealth distribution as well as the religious beliefs of the respondent’s. My question for that would be, “is there a relationship between having strict and traditional beliefs and a respondent’s wealth?”.

Overall, I thought this was a great project to summarize what we have learned this semester. This analysis unfortunately did not have any profound outcomes that I had hoped for, but nonetheless, I was able to find one interesting outcome in that the wealth distribution in pregnant women is quite even in regards to their education level and if I were to further this study, that is absolutely where I would take this project next.