Elvis Atitsoabui, Sloan Duvall, Jennings Dean, Jaci Crute, Matthew Li

For our project, we surveyed college students to investigate the relationship between field of study, political knowledge, and partisan identification. We also explored the relationship between demographic indicators such as race, and policy opinion, namely support for abortion access. After receiving our survey data, we then proceeded with a data analysis to more so understand the answers from our respondents. We did this by calculating descriptive statistics and conducting hypothesis tests for certain variables in our data set.

Our first hypothesis was aimed at measuring if the association between race and response to one of our questions ,which asked a respondent's stance in abortion, was statistically significant. Race was our independent variable and is nominal. Abortion was our dependent variable and nominal as well. Our Null hypothesis was there was no statistically significant association between race and abortion. Our alternate hypothesis was that there was a statistically significant association between our two variables.To conduct our hypothesis test, we used a chi square test since both of our variables were nominal. After running the chi sq test our results yielded a p value of 1.194e-7. This falls far below our significant level of .05 and thus we can say that the association between the variables race and abortion is most probably due to some significant reason other than chance.Therefore, we reject the null hypothesis.

To further characterize our association we performed a cramer's V test. The result was that our Cramer's V value was .533. This denotes a relatively strong association between abortion and race. Furthermore, to understand the distribution and spread more we calculated the Mode and skewness for abortion. The result was that mode was 4, therefore the 4th choice in our abortion question which is that "By law a woman should always be able to obtain an abortion as a matter of personal choice" was selected the most by respondents. From skewness we found that the value was -1.73 indicating a very powerful right skew for our abortion question, telling us there were more responses toward 4 and 5 then 1 ,2, or 3. Since descriptive statistics for race would not yield much information we decided to conduct descriptive statistics such as Mode and skewness for abortion but by race. This resulted in the modes for American Indian, Asian, African Americans, Other, White, and White/African American being 4,3,4,4,4,4 respectively. Skewness was only calculated for white and African Americans. Their values were -.59 indicating a weak right skew of the data toward the 4 selection, and -1.52 indicating a strong right skew of the data toward the 4 to 5 answer selections. The fact that skewness was not calculated for the other demographics simply means we only received 1 respondent for those demographics.

Our next hypothesis was aimed at testing whether the relation between fields, stem or non stem, and political knowledge was statistically significant. Our variables here were field, which was nominal and Political knowledge, which was interval. This variable was created by taking the sum of the 6 knowledge questions where each question was recoded as 1 if correct and 0 if incorrect. Political knowledge was our independent variable and field was our dependent variable.Our null hypothesis was that the association between political knowledge and fields was not statistically significant. Our alternate hypothesis was that the association between political knowledge and fields was statistically significant. The method we used to test our

hypothesis was ANOVA and TukeyHSD. Our result was that we received a p-value of .0486 which falls slightly below our statistically significant threshold of .05. To further investigate, the Tukey test yielded that the statistical significance came from the association of stem and non stem fields in political knowledge. Its p value was .0513 which is slightly above .05. Based on this we can characterize the association between the fields of stem and non stem as marginally statistically significant for political knowledge despite the p value of .0486. Further since this p value is between .05 and .1 we cautiously fail to reject the null hypothesis. Lastly, the p values for the other combination of distributions, had p values that were above our statistical significance level. For our descriptive statistics, we conducted the measure of mean and skewness on political knowledge as a whole and by fields. As a whole we found that the measures were 3.06 and 0 respectively. The mean here tells us that on average, respondents received a 3 out 6 or 50% for their political knowledge score. The skewness of 0 tells us that for our dispersion, there was no major skew one way or the other making our dispersion somewhat related to the normal curve but just shifted, which was very interesting to find. By field for non stem we have 4.18 and -.78, showing a mean political knowledge score of 69% and a weak right skew of our data toward the 4-6 range of scores. For others, the mean of 2.57 indicates a low political knowledge score of 42% and skew value of -.15 indicating a weak right skew towards the 3-6 range of values. Lastly for stem we had a mean of 2.53, showing a low political knowledge score of 42% and a skewness value of .14 showing a weak left skew toward the 1-3 range of values.

Our last hypothesis was aimed at investigating if the association between Political identification(political party) and Field was statistically significant.Our dependent variable was fields which are nominal. Our independent variable was political identification which was nominal as well. Our Null hypothesis was that the association between the two variables was not statistically significant. Our alternative was that the association was statistically significant. For this hypothesis we conducted a chi sq test and cramer's V. The result was a p value of .7789 for the chi sq test. This indicated that the association between the 2 variables was far above our level of significance .05.Therefore, we fail to reject the null hypothesis. To analyze the association more, our Cramer V value was .215, indicating a moderate close to weak association between our 2 variables. For our descriptive statistics we evaluated the mean,mode, and skewness for the field as a whole and by political Identification. This was done by making stem equal to 1 and non stem equal to 0 and dropping the "other" option in the field since it could be confounding. The results were, as a whole, a mean of .61, mode of 1 and skewness of -.44. The mean of .61 has significance here if we were to think of the field of a respondent as being able to take on only 1(stem) or 0(non stem). This would tell us that a respondent has a probability of .61 of being in the stem field.This could be used to estimate the number of stem students in a sample or population.This becomes more simple by thinking of a binomial random variable for fields. Mode tells us that 1(stem) was the most present value in our fields. Skewness tells us that there was a right skew toward the 1 values in the fields column, however not by much. This can be supported by the mode value. Performing statistics by political Identification yielded that democrats had a mean of .54, mode of 1, and skewness of -.15. Thus we can conclude, a typical college student that is democrat is stem with a probability of .54. Stem was also the most present option in fields for democrat respondents. For democrats we also have a weak right skew toward stem showing most democrats are stem but not by much. For the independent we received values of .73, 1, and -1.02. Therefore, an independent respondent

would be stem with probability .73. Stem is the most present option in fields for independent respondents due to mode 1. Lastly we have a skew to the right towards the stem values for our independent respondent, and this would be a skew by a considerable amount.

In conclusion, these are the results from the various hypothesis tests we conducted on our data to better understand political information from the respondents of our survey. We initially intended to find a connection between partisan identity and major and between partisan identity and election knowledge. However, due to the lack of conservative respondents, we were unable to find these connections.Despite this our data yielded much information.