Politics 101 Homework #2

1. Concept for dependent variable: For this homework, we decided to measure the importance of coastal conservation policies among Californians. We measured this by using three indicators which can be found in the PPIC July 2017 data set: perceived economic importance of ocean and beach condition, perceived personal importance of ocean and beach condition and perceived importance of declining marine life.
    a. Dependent Variable Indicator 1/BCal
        i. Q25: Economic importance of clean coast
            1. "How important is the condition of the ocean and beaches to the economy and quality of life for California's future? Is it very important, somewhat important, not too important, or not important at all?"
        ii. Missing Value/Recodes: First, were coded 8/9 (don't know/refused) into missing values, because the number of total respondents for those categories was less than one percent. We then recoded q25 into BCal for easy viewing purposes. We took the remaining categories and recorded them into two categories 'important' and 'not important'. These 2 categories indicate how respondents view the economic importance of maintaining a clean coast. It should be noted that by condensing the data this way, we had to convert an ordinal data set into nominal data; however, this was necessary as the percentage of respondents in some of the ordinal categories was too small and produced too much skew and kurtosis. We condensed responses 2, 3 and 4 into a single category ('not important') as the percentage of responses were too small on their own. It was challenging to find variables without a large skew and large kurtosis; we assume this is because the population of California is relatively like minded and geographically concentrated. By coding 'important' into 1 and 'not important' into 0, we made the high score equal to perceived economic importance regarding coastal condition.
        iii. Descriptive Statistics: Because the recoded data is nominal, the most important central tendency to look at is the mode. For BCal, the mode is 1; therefore, it can be concluded that the most frequent response for this indicator is that the clean coastal areas are important to the economy.

Mean	.7389

Median	1.0000
Mode	1.00
Skewness	-1.089
Kurtosis	-.816

b. Dependent Variable Indicator 2/ BPer
    i. Q26: Personal Importance of a clean coast
        1. How important is the condition of the ocean and beaches in California to **you personally**? (Is it very important, somewhat important, not too important, or not important at all?)
    ii. Missing Values/Recodes: We re coded 8/9 (don't know/refused) into to missing values, because the number of total respondents for those categories was less than one percent. We then recoded q26 to BPer for easy viewing purposes. We took the four possible responses and condensed them into two categories, 'important' and 'not important'. Again, we realized that we are taking ordinal data and converting it into nominal data; however, the proportions of the uncoded data were too skewed with a very large kurtosis. By coding 'important' into 1 and 'not important' into 0, we made the high score equal to perceived personal importance regarding coastal condition.
    iii. Descriptive Statistics: Again, this indicator is nominal; therefore, we will look again at the mode. For this indicator, the mode is 1, which reflects that the majority of respondents have indicate that clean coastal areas are personally important to the respondents.

Mean	.6483
Median	.5000
Mode	1.00
Skewness	-.552
Kurtosis	-1.066

c. Dependent Variable Indicator 3/ MLife

i. Q22: Perceived importance of marine life health
    1. Thinking about the part of the California coast that is closest to you, please tell me whether you think each of the following is a big problem, somewhat of a problem, or not a problem today. [ROTATE Q22 to Q24] Q22. How about declining marine life? (Is this a big problem, somewhat of a problem, or not a problem today?)

ii. Missing Values/Recodes: Again, we recoded 8/9 (don't know/refused) as it only accounted for six percent of the data. For easy viewing purposes, we recoded q22 to MLife. Looking at the raw data, we decided not to condense any categories, as the percent of respondents in each category were balanced enough so that we did not have to condense the ordinal data (44.8%, 31.4% and 17.9%). However, we did recode the names and values of the categories to keep scoring consistent among all of the dependent variable indicators. By coding 'very important' into 1, 'somewhat important' into .5 and 'not important' into 0, we were able to maintain a consistent direction and range for all of the indicators.

iii. Descriptive Statistics: This question is an ordinal measure; therefore, the meaningful measures of tendency are the mode and median. The mode was 1, which reflects that the majority of respondents though that marine life health was very important, looking at the data, 47.6% of respondents feel into this category. The median was .5, this reflects that the middle of the data falls into the 'somewhat important' category. 31.4% of respondents feel into this category.

Mean	.7115
Median	1.0000
Mode	1.00
Skewness	-.934
Kurtosis	-1.128

iv.

iv.

d. Reliability Analysis: The alpha score suggests that there is an acceptable amount of reliability between the three indicators for the dependent variable, .645. Although the alpha would increase to .653 if we deleted the MLife indicator, it is more useful for us to keep the MLife indicator as we want three indicators, not two. Furthermore, deleting the BCon or BPer in-

dicators would decrease the alpha; therefore, deleting either of these indicators is not a viable option as the alpha would drop to .451 or .509. All in all, the .645 alpha score leads us to conclude the indicators are related enough to make an index.

Cronbach's Alpha	N of Items
.645	3

e.
e.

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
BCal	1.3719	.442	.522	.451
BPer	1.3981	.443	.483	.509
MLife	1.4766	.574	.371	.653

2. Summary Index:
    a. Raw Summary Index Statistics:
    b. Raw Index:

Mean	2.1233
Median	2.5000
Mode	3.00
Std. Deviation	.96400
Variance	.929
Skewness	-.892
Kurtosis	-.447

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	106	6.2	6.7	6.7
	.50	110	6.4	6.9	13.6
	1.00	118	6.9	7.5	21.1
	1.50	124	7.2	7.8	28.9
	2.00	221	13.0	14.0	42.8
	2.50	308	18.0	19.4	62.2
	3.00	599	35.1	37.8	100.0
	Total	1586	92.9	100.0	
Missing	System	122	7.1		
Total		1707	100.0		

3. Recoded Index: After recoding the raw summary index, we saw substantial differences in summary measures. First, we saw the standard deviation decrease when we recoded the data from .964 to .405. This is expected as recoding the data condensed the number of categories into three possibilities (hi, med, low). We decided to cut this data into three categories, to cut the cumulative percent columns into thirds. Although recoding the data in this way does not produce three perfect categories with 33.3% of respondents in each section, it is as close as we could get. Condensing the data in this way is helpful, because it allows us to visualize what the data is telling us with more clarity. Second, the mode

changes from 3 to 1. This indicates that the majority of respondents would support environmental policies. The median goes from 2.5 to .5, this happens because the recoding condenses the categories and balances out the responses; therefore, the middle of the data falls into the 'medium' category in which 37% of respondents fall.

a. Recoded Summary:

N	Valid	1586
	Missing	122
Mean		.5445
Median		.5000
Mode		1.00
Std. Deviation		.40587
Variance		.165
Skewness		-.164
Kurtosis		-1.464

b. Recoded Index:

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	low	458	26.8	28.9	28.9
	med	529	31.0	33.4	62.2
	hi	599	35.1	37.8	100.0
	Total	1586	92.9	100.0	
Missing	System	122	7.1		
Total		1707	100.0		

4. Independent Variables:
    a. Independent Variable 1: Ideological Identification
        i. Q39: Blue
            1. Next, would you consider yourself to be politically?

ii. Concept being measured by IV: The concept being measured is someones ideological identification ranging from liberal to conservative.
iii. Recodes/Missing Variables: We declared missing variables for 8/9 (missing and don't know), because these responses account for only 4% of responses. The raw indicator is ordinal, and when I recoded the data, it remains ordinal even though I condensed the categories from 5 to 3. We did this, because it broke the categories into thirds (or as close as we could get to thirds). Therefore, responses 1,2 became 'liberal' which we gave a score of 1, response 3 become 'moderate' which we indicated with a score of .5, and responses 4,5 became 'conservative' with a score of 0.
iv. Summary Statistics: the relevant summary measure shows that the middle of the data falls at .5, which highlights that the middle of responses indicate a 'moderate' ideology. The mode falls at 1, which indicates that the majority of respondents are liberal.

Mean	.5108
Median	.5000
Mode	1.00
Skewness	-.041
Kurtosis	-1.605

b. Independent Variable 2: Age
   i. D1a: red
      1. Could you please tell me if you are between the ages of…?
   ii. Concept being measured by IV: This independent variable is measuring the age range in which each respondent falls into.
   iii. Recodes/Missing Values: First, we declared missing values for 9 (refuse to answer), as the total number of respondents in that category was less than 1%. In regards to recoding, we left the data in an ordinal format, but we condensed the data from 6 categories to 3 categories. By doing this we condensed 1,2 into '18-34' with the value of 0, 3,4 into '35-54' with the value of .5, and 5,6 into '55 and older' with the value of 1. Again, we made the decision to condense the categories for viewing purposes and to balance the amount of respondents in each category.
   iv. Summary data: The median and mode for this data is .5, which indicates that the most respondents and the middle of respondents fall into the '35-54' category. Cutting the data this way is the most

logical format as it separates respondents into similar age categories and was the closes we could come to three even categories.

Mean	.4935
Median	.5000
Mode	.50
Skewness	.024
Kurtosis	-1.444

c. Independent Variable 3: Level of Education
    i. D7: Green
        1. What was the last grade of school that you completed?
    ii. Concept being measured by IV: This independent variable measures the level of education that the respondent has received.
    iii. Missing values/records: We decided to declare missing values for 6/9 (trade school and refused), because the percentage of respondents in that category were minimal. We left the data in an ordinal format, but we decided to condense the data into three separate categories. 1,2 become 'High school or less', 3 remained 'some college', and 4,5 became 'college grad or more' Originally, we had left the data in 5 categories; however, condensing the data for this IV allowed for easier viewing and did not impact the chi square or ANOVA values in any meaningful way.
    iv. Summary data: The median for this IV is .5 which indicates that the middle of respondents fall into the 'some college' category. The mode is 0 which tells is that the highest number of responses were in the 'HS or less category'.

Mean	.4485
Median	.5000
Mode	.00
Skewness	.191
Kurtosis	-1.478

5. Chi Square Analysis:
    a. Relationship between IV (blue) and Indexed DV

			blue			
			conservative	moderate	liberal	Total
H20	low	Count	176	118	140	434
		% within blue	34.7%	26.7%	24.3%	28.5%
	med	Count	164	163	186	513
		% within blue	32.3%	36.9%	32.3%	33.7%
	hi	Count	167	161	249	577
		% within blue	32.9%	36.4%	43.3%	37.9%
Total		Count	507	442	575	1524
		% within blue	100.0%	100.0%	100.0%	100.0%

i. Cramer's V: .083
ii. Tau-b: .096
iii. Approximate Significance: .000
iv. Analysis: The relationship between political ideology and perceived importance of coastal conservation is a weak but positive relationship demonstrated by the positive tau-b. Tau-b was used because the data is ordinal and a 3x3 square. The .000 significance highlights that the data is statistically significant; however, this is probably because the number of survey respondents is relatively high. The low Tau-b tell us that although statistically significant, the relationship is not of high importance. Looking at the cross tabs, we see some cell percentages differ across the top and bottom rows, this suggests a relationship. The combination of variables was hypothesized to represent a relationship, because we assumed that the more liberal someone was the more value that they would find in conserving the oceans and environment.

b. Relationship between IV (red) and Indexed DV

			red			
			18-34	35-54	55 and older	Total
H20	low	Count	132	173	148	453
		% within red	24.8%	30.6%	31.0%	28.7%
	med	Count	174	181	171	526
		% within red	32.6%	32.0%	35.8%	33.4%
	hi	Count	227	211	159	597
		% within red	42.6%	37.3%	33.3%	37.9%
Total		Count	533	565	478	1576
		% within red	100.0%	100.0%	100.0%	100.0%

i. Cramer's V: .060
ii. Tau-b: -.069
iii. Approximate Significance: .002
iv. Analysis: The relationship between age and the importance of coastal conservation is an incredibly weak, negative relationship. We used tau-b because this is a 3x3 ordinal square. The negative Tau-b tell us that this is an inverse relationship in which the younger respondents are slightly more likely to care about coastal conservation. Scanning the rows shows some relationship between variables, but the slight differences aren't necessarily substantial. Again, the statistical significance is likely due to the high number of respondents in the survey. The small Tau-b tells us that this relationship is even less important than H20s relationship to IV(blue), because it is about a third smaller. Therefore, there does not seem to be a relationship between respondents age and the importance of coastal conservation.

c. Relationship between IV (green) and DV

			green			
			HS or less	some college	college grad or more	Total
H20	low	Count	169	145	130	444
		% within green	28.4%	28.7%	28.2%	28.4%
	med	Count	207	163	156	526
		% within green	34.8%	32.3%	33.8%	33.7%
	hi	Count	219	197	175	591
		% within green	36.8%	39.0%	38.0%	37.9%
Total		Count	595	505	461	1561
		% within green	100.0%	100.0%	100.0%	100.0%

i. Cramer's V: .024
ii. Tau-b: .007
iii. Approx Significance: .750
iv. Analysis: Finally, the relationship between education is very weak and negative. Again, we used tau-b as this data also embodies a 3x3 ordinal square. The approx significance number tells us that this is not a statistically significant measure, because the significance is over .1. The Tau-b is also very small, which shows that this relationship is not important. Finally, there is virtually no relationship that can be seen by scanning the rows of the cross tab, which means that theres not much important information that can be found regarding the relationship between education and costal conservation support. That being said, there is not significance or importance in the relationship between education and perceived importance of coastal conservation.

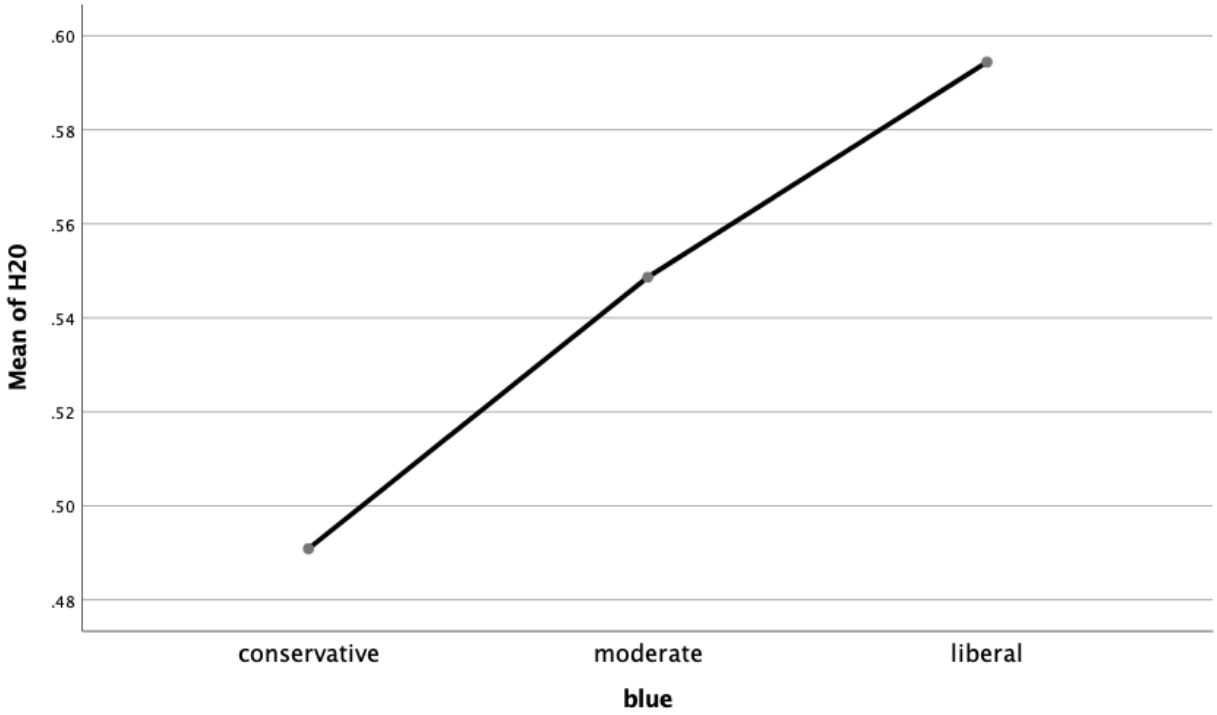
d. Chi Square Analysis for the 3 cross tabs:
   i. Political Ideology (Blue) x Perceived Importance of Coastal Conservation Cross tab:
      1. Pearson's Chi-Square= 20.758 , df= 4, p=.000
      2. Analysis: The results of the chi square test highlight that this is a highly significant relationship, because the p value is .000. This tell us that the relationship is very likely a representation of the population and is very unlikely due to chance.
   ii. Age (red) x Perceived Importance of Coastal Conservation Cross tab:
      1. Pearson's Chi-Square=11.510, df=4, p=.021

2. Analysis: Like the first cross tab, the results of the chi-square test highlight that this is a significant relationship, because the p value falls between .01 and .05. And is likely a representation of the population.

iii. Education (green) x Perceived Importance of Coastal Conservation Cross tab:
   1. Pearson's Chi-Square=.891 , df=4, p=.926
   2. Analysis: The results of the chi-square highlight that this is a non-significant relationship, and that any relationship in the cross tab could be due to chance. We came to this conclusion by looking at the high p value that fell between .1 and 1.

e. ANOVA Analysis:
   i. Political Ideology → Perceived Importance of Coast Conservation

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.889	2	1.445	8.903	.000
Within Groups	246.630	1520	.162		
Total	249.519	1522			

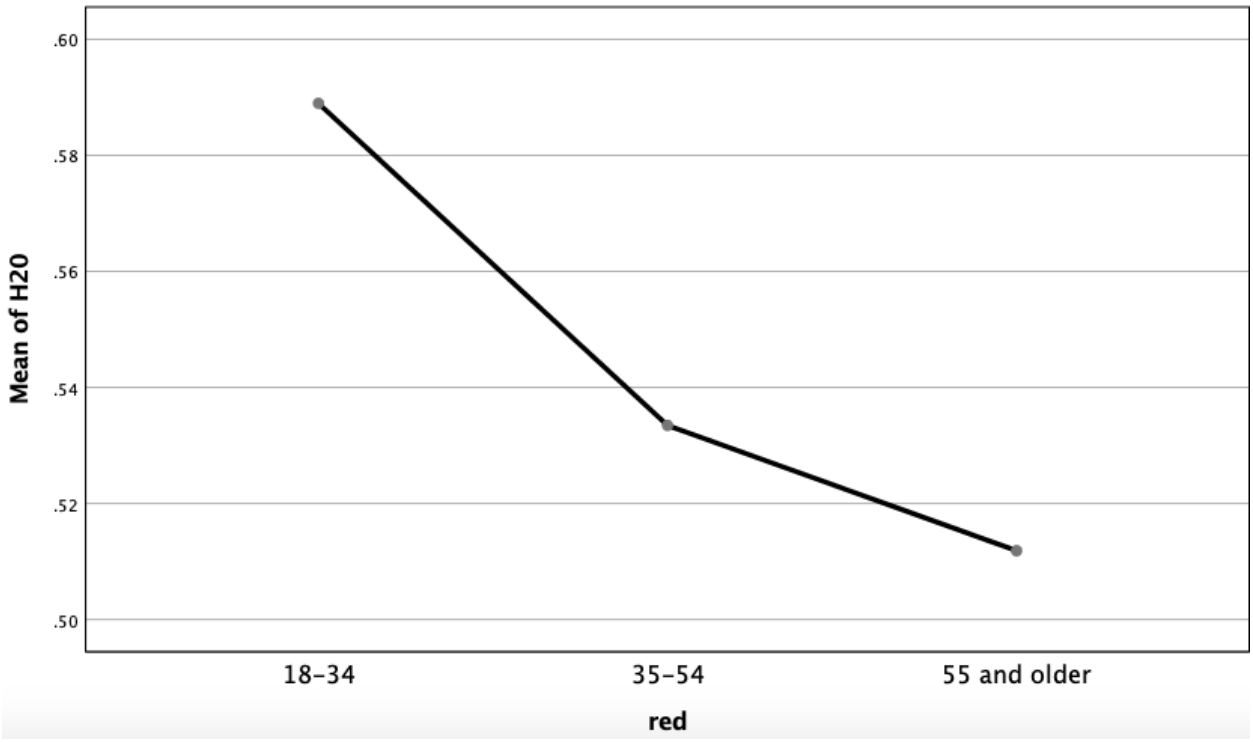


   ii. Analysis: Confirming what we saw in the cross tab. The ANOVA panel shows that the significance is .000 which is a highly significant. This means that the relationship or any differences within the cross tab are not due to chance. In addition, when we look at the

mean differences, we see that one of the sub comparisions is significant, this is demonstrated by an asterisk next to the given values, this means that the means differ enough between the two categories that there is an interesting comparison to be made. We see that the relationship between conservative and liberal has a much greater significance (.000) than conservative and moderate (.089) and/or liberal and moderate (.2). This tells us that the most significant relationship is between the conservative and liberal sub comparision. Therefore, if using this data in the future, we should focus our energy on looking at the differences between those two categories opposed to the data as a whole. However, it should be noted that the mean plot is condensed and only shows numbers between .48 and .6 (on a scale from 0 to 3). This means that even though there is a significant mean difference between the conservative and liberal subcategories, it is relatively small and stays around the middle range of the index.

iii. Age → Perceived Importance of Coast Conservation

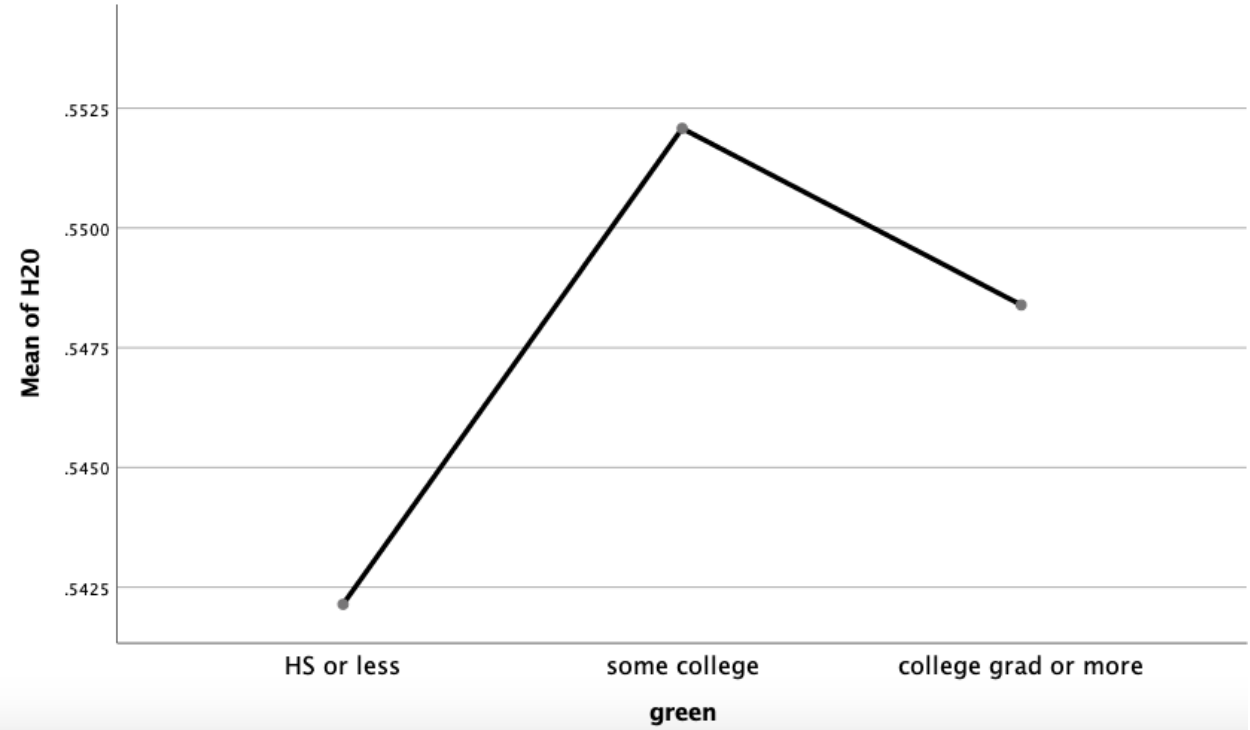
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.628	2	.814	4.969	.007
Within Groups	257.403	1571	.164		
Total	259.031	1573			



iv. Analysis: The ANOVA panel shows that the significance is .007 which indicates a highly significant relationship that is very likely not due to chance. Looking at the mean differences, we see more interesting comparisons between the subgroups. There is an asterisk between the sub comparision which compares the categories '18-34' and '55-older', the significance in for this sub comparision is .011. This tell us that we should direct our energy toward this comparison (young and old) more specifically. The other two sub comparisons significances are .076 and .692 which highlight little to no significance and tell us that the relationship between these subcategories aren't significant. Again, the mean plot is condensed, which also highlights how weak the mean differences are.

v. Education → Perceived Importance of Coast Conservation

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.028	2	.014	.085	.919
Within Groups	255.199	1556	.164		
Total	255.227	1558			



vi. Analysis: Confirming what we saw in our cross tabulation, the ANOVA panel shows that the significance is .919 which further confirms that this is not statistically significant or useful. Looking at the mean differences in the table provided, there are no mean differ-

ences of any significance as their p values are all .1 or higher and do not have any asterisks. Thus, the hypothesis that there is a relationship between level of education and perceived importance of coastal conservation is not supported by the data. Compared to the other two mean plots, this plot is even smaller as all levels of education have means between .53 and .56. This tell us that education makes virtually not difference to someones support regarding costal conservation policies.

6. The independent variable blue, or political ideology, demonstrated the greatest relationship to the indexed dependent variable, which was measuring the perceived importance of coastal conservation policies among Californians. This was demonstrated by having the highest measure of association at .096 for Tau-b (which was slightly higher than the measure of association for our index and age, -.069), and by being the independent variable with the highest measure of significance, at .000 for chi-square and ANOVA, which indicates that the relationship between these two variables, although disappointingly weak (as the tau-b was less than .1) was not due to chance. All in all, the finding in this experiment were weak. No independent variable produced a high level of association with the indexed dependent variable. The most significant relationship that was produced was found between the subcategories of liberal and conservative when looking at ideology and concern about coastal conservation. Those with the most liberal ideologies found coastal conservation to be important; whereas, conservative individuals found this to be slightly less so. The results of this analysis are surprising to us. We expected these IV to be adequate predictors in determining whether or not someone cared about costal conservation. The education IV was especially disappointing, because it is reasonable for one to hypothesize that the more education one has the more they would understand be concerned about coastal conservation. This might signify a failure in the education system to properly educate individuals about environmental policies and climate change. Thinking critically, it could be hypothesized that, because the respondents live in a place that is so close to the coast, the majority of Californians at least sightly support costal conservation as it directly impacts their every day lives. That being said, it would be interesting to run this analysis on a national scale in which respondents are more geographically diverse in the future and compare the differences. Finally, It should be noted that the data was weighted according to the PPIC data codebook specifications. The data was adjusted adjusted to combat oversampling to make data more representative of the entire population of California.

Syntax:

```
*weighting the data*.

Weight by weight.

*univariate statistics for uncoded indicators*.

DATASET ACTIVATE DataSet1.
FREQUENCIES VARIABLES=q22 q25 q26
  /statistics = mode median mean SKEWNESS KURTOSIS.

*recoding the indicators*.

recode q25 (1=1) (2, 3, 4 =0) into BCal.
value labels BCal 1 'important'  0 'not important'.

recode q26 (1=1) (2, 3, 4 =0) into BPer.
value labels BPer 1 'important' 0 ‘not important’.

recode q22 (1=1) (2=.5) (3=0) into MLife.
value labels MLife 1 'very' .5 'somewhat' 0 'not'.

*descriptive statistics for recoded indicators*.

DATASET ACTIVATE DataSet1.
FREQUENCIES VARIABLES=BCal MLife BPer
  /statistics = mode median mean SKEWNESS KURTOSIS.

*Conducting Reliability Analysis*.

reliability /variables= BCal BPer MLife
  /scale('H201') all
  /statistics=descriptive
  /summary=total.

*Constructing the Index*.

compute RawH20 = (BCal + BPer + MLife).
fre var RawH20
 /statistics = mean median mode stddev var skew kurtosis.

*Recoding the Index*.
```

```
recode RawH20 (0 thru 1.51=0) (2.0 thru 2.51= .5) (3 =1) into H20.
value labels H20 0 'low' .5 'med' 1 'hi'.
fre var H20
 /statistics = mean median mode stddev var skew kurtosis.

*creating the predictor of ideological identification*.
recode q39 (1,2 =1) (3 =.5) (4,5 = 0) into blue.
missing value blue (8,9).
value labels blue 1 'liberal' .5 'moderate' 0 'conservative'.
fre var blue.

*running description stats for ideology*.

DATASET ACTIVATE DataSet1.
FREQUENCIES VARIABLES=blue
  /statistics = mode median mean SKEWNESS KURTOSIS.

*creating predictor of age*.
recode D1a (1,2 = 0) (3,4 = .5) (5,6 = 1) into red.
missing value D1a (9).
value labels red 0 '18-34' .5 '35-54' 1 '55 and older'.
fre var red.

*running description stats for age*.

DATASET ACTIVATE DataSet1.
FREQUENCIES VARIABLES=red
  /statistics = mode median mean SKEWNESS KURTOSIS.

*creating predictor of education*.
recode d7  (1,2 =0) (3=.5) (4,5 =1) into green.
missing value d7 (6,9).
value labels green 0 'HS or less' .5 'some college' 1 'college grad or more'.
fre var green.

*running description stats for education*.

DATASET ACTIVATE DataSet1.
FREQUENCIES VARIABLES=green
  /statistics = mode median mean SKEWNESS KURTOSIS.

*Crosstabulation of H20 by ideology, age, education*.
crosstabs tables = H20 by blue, red, green
  / cells = column count
```

```
  /statistics = phi btau chisq.

*One-way ANOVA for ideology*.
oneway H20 by blue
   /statistics=descriptives
   /ranges=scheffe
   /plot means.

*One-way ANOVA by age*.
oneway H20 by red
   /statistics=descriptives
   /ranges=scheffe
   /plot means.

*One-way ANOVA by education*.
oneway H20 by green
   /statistics=descriptives
   /ranges=scheffe
   /plot means.
```