Research Question:

Are there significant differences in attitudes towards immigration between different generations in the United States?

Summary:

In this study, our objective was to determine whether there are significant differences in attitudes towards immigration between different demographics in the United States, mainly focusing on age but also including sex and race. Our team found immigration as a topic to be particularly interesting due to its growing relevance on shaping policy development and economic planning. Understanding how different age groups perceive immigration can be valuable especially as immigration rates change due to population growth, further globalization, environmental factors, and political conflict and instability. To accomplish this, we obtained data from the General Social Survey, conducted data cleaning and processing, and created and used visualizations alongside descriptive statistics to investigate patterns between the chosen variables. The dataset included questions asking the respondents their stances towards immigrants as well as their age, race, and sex. We primarily examined the age of the respondents but we also incorporated the respondent's race and sex into our analysis to validate that age was the most significant factor influencing opinions about immigration. Prior to analyzing the data, we filtered out variables of interest and handled any missing or inaccurate responses. Our main visualizations include kernel density plots to view the distribution of age, our only continuous variable, and count plots to view the frequency of observations for all of our other variables, which are categorical. Our major findings include a definitive presence of a generational divide in attitudes towards immigration. Younger respondents consistently favored more open and liberal immigration policies, believing that immigrants should retain their culture and become a part of American society. They also held more positive views of illegal immigrants. In contrast, older respondents were more likely to support stricter immigration policies, advocating for assimilation and expressing a preference of wanting illegal immigrants to be deported. Interestingly, while race and sex did also play a role in shaping views, age was consistently the most significant factor affecting how individuals responded to questions about immigration.

Data:

Arguably the most important part of this project was the selection of data from the General Social Survey and the cleaning of the chosen variables. The General Social Survey is a publicly available survey data set that provides valuable information on various social attitudes and behaviors of the American population. Since our research question revolved around generational attitudes towards immigration, we selected variables related to this topic, such as questions about support for immigration, what should be done about current illegal immigrants, and the respondent's age. The data we ended up selecting is summarized in Table I. All the selected variables were found in the first chunk of the given parquet files and section 1.000 of the Codebook, which make the parsing and downloading process easy. The selected variables for our

research ended up being ‘adoptus’, ‘immfate’, ‘letin1a’, ‘letinhsp1’, ‘letinasn1’, ‘age’, ‘sex’, and ‘race’.

Variable Name	Label
ADOPTUS	Do you think immigrants living in the U.S. should adopt U.S. culture, keep their own culture, or do both?
IMMFATE	What should be done about immigrants who are currently living in the U.S. illegally?
LETIN1A	Do you think the number of immigrants to America nowadays should be…
LETINHSP1	Do you think the number of immigrants from Latin America who are permitted to come to the United States to live should be decreased, increased or left the same?
LETINASN1	Do you think the number of immigrants from Asia who are permitted to come to the United States to live should be decreased, increased or left the same?
AGE	Respondent’s age.
SEX	Respondent’s sex.
RACE	What race do you consider yourself?

Table I: Variables Selected from the GSS

Next, we had to clean this data. For the most part, the data was already relatively clean since the General Social Survey is a reputable and large-scale project with many checks for clean data already in place- stating that they have “taken great care to keep the survey methodology as comparable over time as possible, which includes everything from keeping the same sampling approach to not changing question wording,” (Smith, et al.). The biggest red flag in the dataset, however, was the large amount of missing values, most likely a symptom of many of our variables having not been added until later versions of the survey. To address this issue, we ultimately decided to exclude any respondents with missing values for our key variables. This choice stemmed from the concern that missing data could bias our analysis or skew the results in unpredictable ways. While dropping responses does sacrifice some information, it's often considered a necessary trade-off when the missing data is substantial. Though alternative techniques do exist, like imputation (when you replace missing values with estimates), some of these can introduce their own new biases and complexities, especially with a large dataset and potentially intricate relationships between variables. Given the resources and scope of this analysis, we opted for the most direct approach to ensure the integrity of our findings.

Additionally, we dropped all responses with "don't know" or "refused" as answers, as these did not provide meaningful information for our analysis. This was a relatively simple

process, as we used the Pandas DataFrame 'dropna()' function to remove rows with empty columns. A command like '`df[~df['adoptus'].isin(['D', 'N', 'I', 'S'])]`' identified and selected all rows that did not contain the "D", "N", "I", or "S", codes, representing "Don't Know", "No Answer", "Not Applicable", and "Skipped on Web", respectively. The cleaning process also involved re-coding certain variables to make them more intuitive and easier to work with.

We first sorted the data by the 'adoptus' column in descending order. This would cause all non-numerical values to go to the top of the data, which was done with an aim to group similar responses together and make it easier to identify and remove any non-numerical responses. In doing so, we found two additional header rows that were inadvertently generated during the execution of the 'get_gss.ipynb' data acquisition script. These extraneous rows were promptly identified and eliminated to ensure the integrity and accuracy of our dataset before proceeding with our data analysis.

We also encountered an issue with the 'age' column being stored as a non-numerical data type. This presented a challenge as the statistical techniques essential for our research are designed to work with numerical data. Understanding this importance of accurately parsing our data by age groups, we took the necessary step of transforming the 'age' column from non-numerical to a numerical data type through the utilization of Scikit-learn's LabelEncoder. This transformation not only ensured compatibility with our data analysis tools but also enabled us to conduct more precise exploratory data analysis. By addressing this data cleaning task early on, we laid a solid foundation for extracting meaningful insights from our dataset.

Next, we dropped any age values less than 18. The General Social Survey only surveys adults that are 18 or older, so any responses from individuals younger than 18 were considered invalid and removed from the dataset. Finally, we replaced values in the 'adoptus' column with more descriptive labels. By default, it contained numerical values from 1.0 to 3.0, with 1.0 representing the response "Adopt US culture", 2.0 representing "Keep their culture", and 3.0 representing "Both". This replacement improved the readability and interpretability of the data. After cleaning the data, we conducted exploratory data analysis to gain insights into the attitudes towards immigration among different generations in the United States.

Results:

The first variable we looked into was ‘adoptus’, which corresponds to the question, “Do you think immigrants living in the U.S. should adopt U.S. culture, keep their own culture, or do both?”. Figure 1. is a visual representation of this question as a kernel density plot compared to the respondents age, where the line hue describes the respondents' answers. The graph shows the ‘keep their culture’ response rises sharply in the beginning, hitting a maximum around age 25, after which it quickly decreases as age increases. The ‘adopt US culture’ response rises later with age, hitting a maximum around age 50 and then

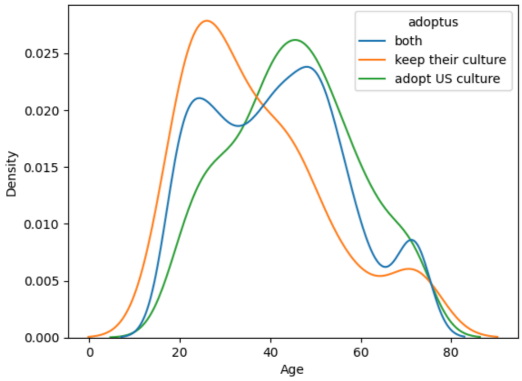

Figure 1. ‘adoptus’ versus ‘age’ represented as a kernel density plot

sharply dropping off. The ‘both’ response varies less with age, with three local maximums that don’t provide us with as much information as the other two responses. We can infer from the kernel density plot that the younger demographic seems to be more supportive of various new cultures brought in by immigrants, allowing them to retain their culture. The older demographic seems to show the opposite, being less supportive of immigrants’ cultures and wanting them to completely assimilate into U.S. culture while leaving their own behind.

Next, we compared ‘adoptus’ and ‘race’ using pandas.crosstab(). The following can be seen in Table II below. All races had a similar percentage of responses that immigrants should both adopt U.S. culture as well as keep their own culture. However, there is a significant difference across race in responses for ‘keep their culture’ and ‘adopt US culture’. Black and Other respondents were very similar, but the White participants held more negative opinions on what immigrants should do once coming to the U.S. 4.5% of White respondents said immigrants should keep their culture, while 14.8% and 11.9% of Black and Other respondents respectively said immigrants should keep their culture. This difference was reflected in ‘adopt US culture’, as 16.5% of White respondents believed immigrants should adopt U.S. culture, while Black and Other response percentages were 7.0% and 7.9% respectively. The data suggests that Black and Other participants have more positive views than White participants on the culture immigrants bring with them, with a lower percentage of White participants saying immigrants should keep their culture and a higher percentage of White participants saying immigrants should adopt U.S. culture.

Then, we compared ‘adoptus’ and ‘sex’ in the same way as we compared ‘adoptus’ and ‘race’. The following can be seen in Table III. We found that response across sex for ‘keep their culture’ varied only 0.1%, while the other two variables had more variation. 5.4% more males than female respondents said immigrants should adopt U.S. culture, and 4.9% less males than females said immigrants should both adopt U.S. culture as well as keep their own culture. This suggests that males have a more negative outlook on the cultures immigrants bring with them into the U.S. than females, although the percent difference is not very significant.

In conclusion, race and age seem to have the largest effect on if U.S. citizens view immigrants and the cultures they bring as positive or negative, with younger respondents of race Black and Other having more positive opinions and immigrant cultures.

The next variable we viewed was ‘immfate’, which corresponds to the question, “What should be done about immigrants who are currently living in the U.S. illegally?”. Figure 2. is a visual representation of this question as a kernel density plot compared to the respondents age, with the hue representing responses. The plot reveals that at approximately age 25, there is a significant increase in responses where illegal immigrants in the United States should either become citizens or become permanent legal residents, with a comparatively low

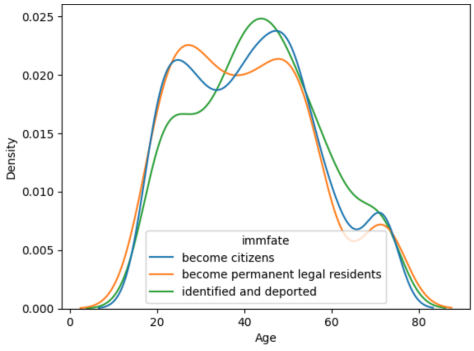

Figure 2. ‘immfate’ versus ‘age’ represented as a kernel density plot

amount saying illegal immigrants should be identified and deported. However, as age increases to approximately 45, citizens who believe illegal immigrants should become permanent legal residents decreased and citizens who believe illegal immigrants should be deported increased dramatically. This peak is hit around age 45-50, after which all responses taper off. We can infer from this pattern that younger U.S. citizens around age 25 hold more positive views on illegal immigrants in the country, responding with the answers that would incorporate illegal immigrants into the U.S. Systems rather than identifying and deporting them. Alternatively, we can also infer that older generations have a more pessimistic view on illegal immigrants, with a peak of responses that illegal immigrants should be identified and deported around age 45.

Continuing with investigating 'immfate', we compared it to the 'race' using pandas.crosstab(). The following can be seen in Table IV below. Black respondents appeared to respond "become permanent legal residents" more than White and Other respondents by almost 8%. The differences within the "become permanent legal residents" responses were smaller, with Other respondents having a higher frequency with about 3.7% more than black respondents and 4.6% more than white respondents. Lastly, White respondents had the highest proportional frequency of "be identified and deported" responses, with 9.4% more than black respondents and 4.9% more than 'other' respondents. Black respondents seem to have a more positive view on illegal immigrants and what should be done with them, in contrast to White respondents who seem to hold a more negative view, wanting in higher percentages to identify and deport illegal immigrants.

Then, we compared 'immfate' and 'sex' in the same way as we compared 'immfate' and 'race'. The results can be seen in Table V below. Similar to race, we find the "become permanent legal residents" response to not vary much by sex. However, females had a 6.7% higher response rate that illegal immigrants in the U.S. should become citizens, while males had a 6.3% higher response rate that illegal immigrants should be identified and deported. This suggests that males have a more negative outlook on illegal immigrants in the U.S. than females, who are more accepting of them.

In conclusion, age and race seem to be the key determining demographics as to how the respondents felt about illegal immigrants, with older, White respondents holding more negative views on illegal immigrants than the younger, Black respondents. Sex seems to play a bigger role in 'immfate' than 'adoptus', but still not overly so.

Another variable we investigated was 'letin1a", which captures responses regarding whether the number of immigrants permitted to come to America nowadays should be "reduced a lot", "reduced a little", "remain the same as it is", "increased a little", or "increased a lot". One of the visualizations created was Fig. 3, a kernel density plot depicting the distribution of these five possible responses across an age range from 0 years old to around 90 years old. For the "increased a lot" response, a prominent peak

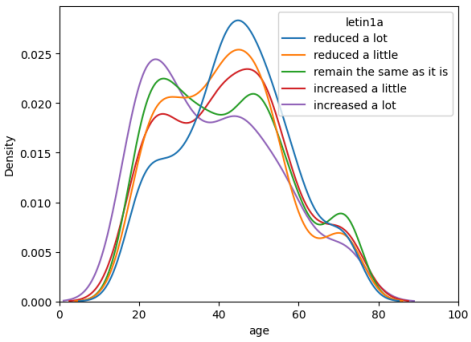

Fig. 3. Age versus opinion on the amount of all immigrants permitted into the U.S.

emerges around the age of 20 and the distribution appears to be skewed to the left, towards the younger ages. In contrast, the plot reveals major peaks around age 50 for the "reduced a lot" and "reduced a little" responses, respectively. The younger demographic appears to be more supportive of an increase in immigrants permitted in compared to those considered middle-aged or older, who were less likely to choose such an option.

A similar trend emerges when looking at 'letinasn1", a variable describing whether the number of Asian immigrants permitted to come to America nowadays should be "left the same", "decreased", or "increased". Looking at Fig. 4, there is again a peak towards the younger ages for "increased" and a peak towards the older ages for "decreased" and "left the same".

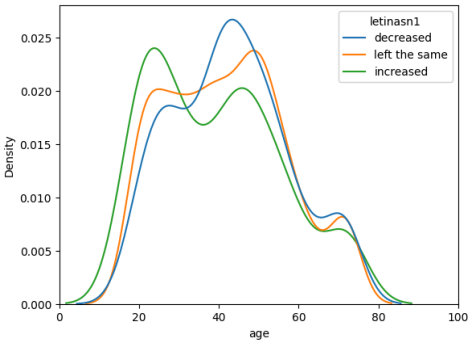

Fig. 4. Sex versus opinion on the amount of Asian immigrants permitted into the U.S.

The final key variable we focused on in our analysis of the GSS was "letinhsp1". Similar to "letinasn1", this variable represented the public opinion of whether the amount of Hispanic immigrants let into should be either "decreased", "increased", or "left the same". To analyze this variable's responses based on age, a kernel density plot was developed: Fig 5. As the variables are very closely related, comparing "letinhsp1" and "letinasn1", can be an effective way to contextualize the distribution and trends of the data. While the two kernel density plots generally have the same trends on a relatively same scale, some slight variations exist, Namely, when looking specifically at the "decreased" line, there is a higher density of responses in the "letinhsp1" plot, mainly in the age range of 40-60, and relative to the "letinasn1" data. From this comparative process, it can be deciphered that while general distributions are similar, a larger percentage of respondents feel that compared to asian immigrants, less Hispanic immigrants should be allowed into the United States.

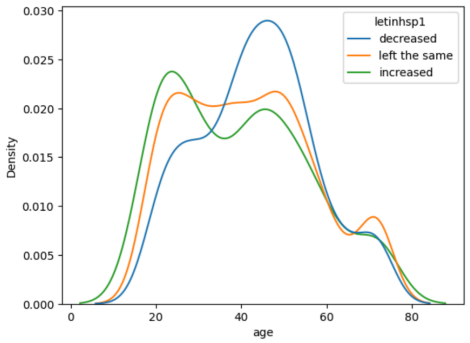

Fig. 5. Sex versus opinion on the amount of Hispanic immigrants permitted into the U.S.

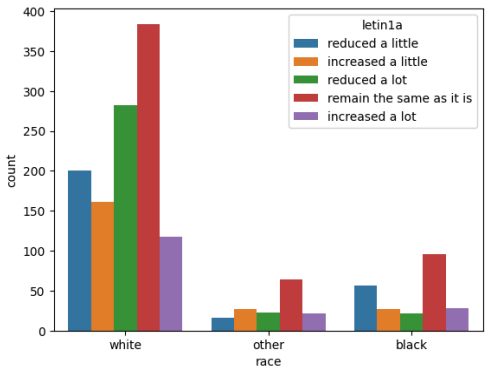

Fig. 6. Race versus opinion on the amount of all immigrants permitted into the U.S.

As for comparing the 'letin1a' variable across different racial groups, as illustrated in Fig. 6, we observe that "remain the same as it is" was the majority response across all groups. There is a similar pattern across all racial groups, though white respondents are more likely to advocate for reducing the number of immigrants "a lot" while black respondents are more likely to advocate for reducing the number of immigrants "a little" as the difference between the next most common responses of the respective racial groups. However, when exploring the effect of sex on 'letin1a', there is almost no discernable difference between male and female opinions on immigration, as seen in Fig. 7. Sex has no significance

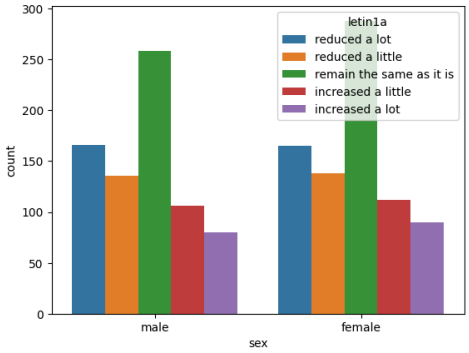

Fig. 7. Sex versus opinion on the amount of all immigrants permitted into the U.S.

on attitudes towards all immigrants in America, while race may have some significance, though not as evidently as age does. Likewise, when investigating the effect of race on the 'letinas1' variable, as illustrated in Fig. 8, we can see that "left the same" was the majority response across all groups. While there is a slight distinction in the next most likely response, for all racial groups, the percentage difference between "decrease" and "increase" was approximately 10% while the percentage of each group wanting the amount of Asian immigrants authorized to come in to remain the same was always slightly above 60% as seen in Table VI. The similarity of the responses across racial demographics reveals that race is not likely a significant factor in attitudes towards Asian immigrants, unlike age.

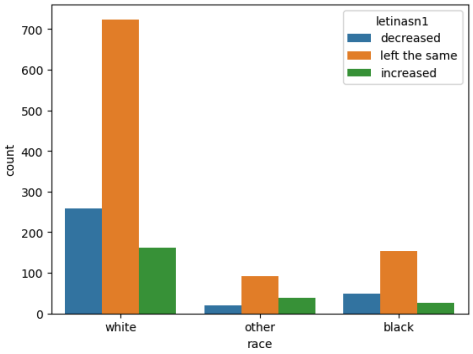

Fig. 8. Race versus opinion on the amount of Asian immigrants permitted into the U.S.

A similar trend emerges when comparing sex and immigration stance, as the counts in Fig. 9. and the percentages in Table VII appear almost identical for male and female, reinforcing how age, not sex or race, is the dominant factor on opinions on the number of Asian immigrants permitted in.

In terms of analyzing "letinhsp1" across racial groups and sex, similar themes can continue to be seen. That is, for each distribution of "letinhsp1" vs characteristics of respondents, we see similar trends to "letinasn1". Importantly however, each with slight spikes for the "decrease" choice, when responding to amounts of Hispanic immigrants allowed in, concentrated within specific demographics. For example, for the data based on race (Table VIII), 31.4% of white respondents hoped to "decrease" the amount of Hispanic immigrants let in, compared to only 22.6% wanting to "decrease" their Asian counterparts. Additionally, when basing "letinhsp1" on sex, there is a general jump in the "decrease" selection for all sexes, when compared to the data of "letinasn1". Notably, the increase in male " decrease" selection was slightly larger.

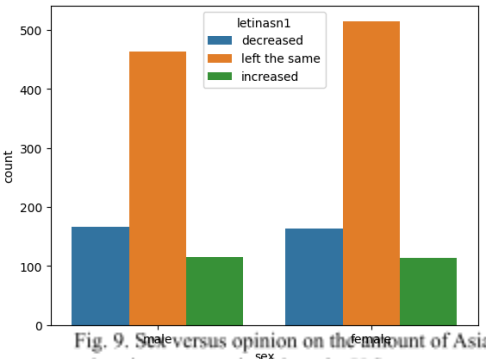

Fig. 9. Sex versus opinion on the amount of Asian immigrants permitted into the U.S.

Overall, based on the data acquired, a general conclusion can be made that a noteworthy percentage of respondents would prefer to "decrease" the amount of Hispanic immigrants allowed in the United States, compared to Asian immigrants.

Conclusion:

Overall, this study concludes that significant differences do exist between generational attitudes towards immigration in the US. We found several concrete pieces of evidence that support this conclusion, including the differing attitudes of generations on immigrant assimilation. Kernel density plots showed a clear shift of opinion with age, as younger respondents very clearly showed a shift in opinion with age, while older respondents displayed a strong preference for assimilation into American culture. By calculating proportions from our analysis, it was revealed that younger generations were more likely to support paths to citizenship or legal residency. In contrast, older generations were significantly more likely to favor deportation. Additionally, percentage comparisons exposed clear negative views on Hispanic immigration compared to Asian immigration. Although many respondents wanted the number of immigrants from both groups reduced, this sentiment was more pronounced for Hispanic immigrants across all age groups.

Causes of these results may be due to several factors, with one prominent reason the overall exposure to diversity that each generation has faced. Younger generations have simply grown up in more diverse environments, leading to a greater acceptance of these different cultures. Older generations may also see immigration as a threat to jobs and resources, something that younger generations are generally less concerned with. Furthermore, the lower ages in this study tended to reflect progressive social values emphasizing multiculturalism and inclusivity, which aligns with typical results in polls for this generation.

For defending our research, our steps and procedures are quite clear, which can be seen directly in our code, and the results in our graphs and tables. As for cleaning data, all our choices along with reasoning are well documented in the “Data” section. On top of this, the questions themselves are procured by the General Social Survey, which is a well known and accredited source. Our research is repeatable and should be free from any criticisms of fraudulent evidence or vague processes.

Additional work still needs to be done so that we can have more information about U.S. views on immigration. For example, why might particular groups have certain dispositions for or against immigration? This question alone should be a whole new study based off of new questions that provide context for respondents’ answers. These questions could be asked in a future GSS, which would maintain the same sampling strategy and surveyed population. Example questions that could be asked could be, “Do you believe immigrants leave a positive impact on the country they come to? (Y/N)” or, “Have you had more or less positive interactions with immigrants? (More/Less)”. Questions such as these start to scratch the surface as to why some groups might have negative or positive outlooks on immigration. Once we draw conclusions from research such as described above and combine it with our own research, we can begin to figure how certain groups feel about immigration as well as why, allowing us to increase education that changes the “why” some might have negative outlooks on immigration, which will change the “how” those feel about immigration.

Going along with this, lawmakers writing immigration laws need to consider the diverse perspectives held by different age groups to create right and fair policies. Recognizing the generational divide can help create conversations between groups with opposing views, potentially reducing social tensions. Also, as younger generations are becoming the more dominant voting party, their more progressive attitudes towards immigration suggest a shift in future immigration policies. It is important to also acknowledge that while our small study does provide valuable insights, it has limitations. Our focus on age as the primary factor might simplify the complex interactions of socioeconomic factors, political affiliations, and personal experiences that shape individual opinions on immigration. Future research should focus on more subtle data collection to hopefully eliminate these biases. Overall, our findings highlight the importance of understanding generational differences when examining public opinion on immigration policy. Addressing these differences is essential for ensuring that immigration policy reflects the values and needs of all generations in American society.

GSS Codebook and Datafile Citation:

Smith, Tom W., Davern, Michael, Freese, Jeremy, and Morgan, Stephen L., General Social Surveys, 1972-2018 [machine-readable data file] /Principal Investigator, Smith, Tom W.; Co-Principal Investigators, Michael Davern, Jeremy Freese and Stephen L. Morgan; Sponsored by National Science Foundation. --NORC ed.-- Chicago: NORC, 2019. 1 data file (64,814 logical records) + 1 codebook (3,758 pp.). -- (National Data Program for the Social Sciences, no. 25).

Grading Criteria:

- Project Concept: What is the research question? Is it well-defined? Are the data appropriate to address the question? References to data documentation or codebooks are essential in this part of the project. What is the research strategy, and is it appropriate to research question?
- Wrangling: How are are missing values handled? For variables with large numbers of missing values, to what extent do the data and documentation provide an explanation for the missing data? If multiple data sources are used, how are the data merged?
- Exploratory Data Analysis and Visualization: For the main variables in the analysis, are the relevant data summarized and visualized through a histogram or kernel density plot where appropriate? Are basic quantitative features of the data addressed and explained? How are outliers characterized and addressed? Are tools for visualizing the relationships between variables used appropriately and interpreted correctly? Are statistics like the mean, variance, median, quantiles, and correlation used appropriately?
- Analysis: What are the main findings of the research? Do the plots and statistics support the conclusions? Is the research strategy carried out correctly? If the research strategy succeeds, are the results interpreted correctly and appropriately? If the research strategy

fails, is a useful discussion of the flaws of the data collection process or the research strategy discussed?

- Replication/Documentation: Is the code appropriately commented? Can the main results be replicated from the code and original data files? Are significant choices noted and explained?
- ^^ All criteria weighted equally (20% each)

Appendix:

Table II
"adoptus" Percentage of Responses by Race

	Response		
Race	Keep Their Culture (%)	Adopt U.S. Culture (%)	Both (%)
White	4.5	16.5	78.9
Black	14.8	7.0	78.2
Other	11.9	7.9	80.1

Table III
"adoptus" Percentage of Responses by Sex

	Response		
Sex	Keep Their Culture (%)	Adopt U.S. Culture (%)	Both (%)
Male	6.8	16.9	76.3
Female	6.9	11.5	81.2

Table IV
"immfate" Percentage of Responses by Race

	Response		
Race	Become Citizens (%)	Become Permanent Legal Residents(%)	Be Identified and Deported(%)
White	64.0	13.9	22.1
Black	72.5	14.8	12.7
Other	64.2	18.5	17.2

Table V
"immfate" Percentage of Responses by Sex

	Response		
Sex	Become Citizens (%)	Become Permanent Legal Residents(%)	Be Identified and Deported(%)
Male	62.0	14.7	23.3

Female	68.7	14.2	17.0

Table VI
"letinasn1" Percentage of Responses by Race

	Response		
Race	Increased (%)	Decreased (%)	Left the Same (%)
White	14.2	22.6	63.3
Black	11.8	21.0	67.2
Other	25.2	13.9	61.0

Table VII
"letinasn1" Percentage of Responses by Sex

	Response		
Sex	Increased (%)	Decreased (%)	Left the Same (%)
Male	15.4	22.4	62.2
Female	14.4	20.6	64.9

Table VIII
"letinhsp1" Percentage of Responses by Race

	Response		
Race	Increased (%)	Decreased (%)	Left the Same (%)
White	16.6	31.4	52.0
Black	15.7	20.1	64.2
Other	22.5	17.2	60.3

Table IX
"letinhsp1" Percentage of Responses by Sex

	Response		
Sex	Increased (%)	Decreased (%)	Left the Same (%)
Male	16.2	30.7	53.1
Female	18.2	25.7	56.1