Data Visualization of Masculinity Survey

Dashboard: https://dabah20.shinyapps.io/shiny_app/

Group member	Email
Daniel Bahrami	dabah20@student.sdu.dk
Hans Pedersen	haped20@student.sdu.dk
Jesper Diederichsen	jedie20@student.sdu.dk
Jonas Beltoft	jobel20@student.sdu.dk

Abstract

In this project, we use data visualization to examine the relationship between masculinity and various demographic factors, such as age, race, and income. The data used in this project comes from a survey conducted by FiveThirtyEight and WNYC, which asked men about their perceptions of masculinity and how it affects their behavior and relationships. By creating charts and plots, we hope to make the data more understandable of how masculinity is perceived and expressed among different groups of men, and how it shapes their behavior and experiences.

Background and Motivation

There are several reasons why we as a group of four males are interested in working on a project that examines the relationship between masculinity and various demographic factors. One reason is that we are personally interested in understanding how masculinity is perceived and expressed in different age and race groups. Additionally, we are interested in exploring the relationship between masculinity, emotional expression and other factors, as this could provide valuable insights into how masculinity shapes men's behavior and interactions with others. Furthermore, working on this project may allow us to contribute to the broader conversation about masculinity and its role in society. Research has shown that men who are more comfortable expressing affection and intimacy are more likely to have strong social support networks and better mental and physical health outcomes. By examining these relationships, we can gain a better understanding of how masculinity impacts men's health and well-being, and identify potential areas for intervention and support.

Therefore, being a group of 4 males, it is certainly interesting for us to look at this data, that can confirm and highlight some issues that men struggle with, across various associations and age groups. We believe there to be certain correlations between your perceived masculinity, and many of the struggles that are asked in the survey, so therefore we want to highlight these the best we can.

Project Objectives

From the data, we have formulated some questions that we want to answer:

- How do questions, such as self-perceived masculinity, come across in different age and race groups?
- What are the factors that a man may worry about in relation to his masculinity, and how do these concerns vary across different races?
- What is the general spread of races over age that has taken this survey?

The motivation behind these questions may be to better understand how masculinity is perceived and expressed among different groups of people. By answering these questions, we may be able to gain insights into how masculinity is constructed and how it varies within and across different social groups. Understanding these dynamics can help us to better understand the role of masculinity in shaping men's behavior and experiences.

Data

The data is from [FiveThirtyEight's study](#) in collaboration with WNYC. FiveThirtyEight is a website that uses data analysis and visualization to cover a wide range of topics, including opinion polls, politics, economics, and sports. It was founded by Nate Silver, a statistician and writer, and takes its name from the number of electors in the United States electoral college. The website began as a blog focused on polling analysis, but has since expanded its coverage to include other areas. It is known for its use of statistical analysis and data visualization to make complex topics more accessible to a general audience.

The poll in which we have obtained our data from, is focused on masculinity and what effects it has on people in different categories. It consists of a series of questions, where the individual can answer typically one of 5 possibilities, usually varying from 'very much' to 'not at all' and 'no answer', with all of them adding up to 100% of people asked. The categories consist of 'Adult men', then three age categories of '18-34', '35-64' and '65 and up'. After these there are categories for 'white' and 'non-white', then 'has children' and 'no children', and lastly sexual

orientation summed up as 'straight' and 'gay/bisexual'. Mainly we'll be using age groups for comparisons, however we also plan to cross compare some of these values, such as masculinity for straight vs gay/bi. In total there are effectively 10 columns and 35 questions, making this data set a decently sized data set to effectively show our points. In order to import and use the dataset, we had to trim the values for the % signs in the data cells. However no bulk cleanup or changes to the data were made.

Data processing

Stacked bar chart - multiple variables

Creating a stacked barchart over the question "Which of the following do you worry about on a daily or near daily basis? (Select all that apply.)" proved rather difficult. The data was not structured in a way we knew how to work with, but we eventually created a stacked barchart with each question represented. However it did not look as we wanted. It wasn't sorted in order, and each question had its own color. It was hard to decode. So we tried to reinterpret the data. We realized that each question was essentially a True or False statement. Either an observation worried about a question, or they did not. This narrowed data down to only have two states, instead of a unique "selected" state for each question. Now we could run over each question and count how many had selected, and how many had not, thereby getting a count. This was then placed in a csv file, with the column headers as: *Question, Selection, Value*. Question being the question that would either be selected, or not. Selection being a categorical variable tired to the value, which represented how many had answered that. So one row could represent: Your height, selected, 800, which means that 800 had selected that they worried about their height. This data could then be grouped by question in R, and stacked in a percentage stacked bar chart. Now we only had two colors, and it was much easier to decode. But it was not enough. We wanted to distinguish answers by race also. So the same method was used, adding another column, *race*, so that the headers would now be: *Question, Race, Selection, Value*. The races could be, All, asian, black, hispanic, other, or white. All being the sum of each of the races, which was the same as before. Now the data could, in r,

be grouped by both Question and Race, and it could then be animated to cycle through the percentage of answers based on race.

Numerical data

Our dataset consisted solely of categorical data. However, when we studied the data closer, we realized that some were categorical ordinal, which meant that it could be said that one answer was greater or lesser than another. This order could be given a numerical value, instead of a categorical one, i.e. the question "In general, how masculine or "manly" do you feel?" with the Answers: *Very maculine, somewhat masculine, not very masculine, not at all masculine, and no answer*, could be mapped to the values 1 - 5. This meant that we could do a scatter plot and a distribution plot.

Visualization

In this section the design and the intended visualization of the masculinity survey will be described.

Design

To visualize the data from a survey on masculinity among men in the age range of 18-65, we would consider using a bar chart or a histogram to show the distribution of responses to each survey question. This would allow the viewer to easily see the most common responses and any significant differences between groups of respondents, such as by age or other demographic factors. We would also consider using color and other visual elements to highlight key findings and make the visualization more engaging and easy to understand. For example, we might use different colors to show the responses of different age groups, or we might use size or shading to indicate the relative importance of different responses. Overall, our goal would be to create a clear and effective visualization that helps the viewer understand the key findings of the survey and any important trends or patterns in the data.

Must-Have Features

When it comes to features, the specific must-have features are the ones that are necessary for the dataset to be able to effectively answer the questions that the group intended to address.

Bar Chart

Bar chart showing the distribution of answers with a filter option to select a question. A bar chart showing the distribution of answers to each question for example the first question "How masculine do you feel?" using a scale of "Not at all", "Not very", "Somewhat" and "Very" as possible answers, there would be 4 bars, each corresponding to a category. Furthermore, the chart will show the distribution for each age group. The height of each bar would represent the number of responses that fall within that category. Each bar would be scaled proportionally to show the relative frequency of each response.

Density Chart

A density chart showing a graphical representation of the distribution of a dataset. In this case, the density chart would show the distribution of responses to the survey question about how masculine the adult men in different race groups feel. The chart would show the number of men in each race group who answered the survey question, as well as the range of responses they provided. For example, the chart might show that a certain percentage of men in each race group feel very masculine, while another percentage feels somewhat masculine, and so on. The chart would allow viewers to quickly see the distribution of responses among the different race groups, providing a visual representation of the data.

Stacked bar chart

A stacked bar chart showing a graphical representation of the correlation between the age and race distribution of all respondents. The chart should give a visualization of the concentration of people in the three different age groups, and also their ethnicity.

The stacked bar chart is particularly useful for comparing the proportions of different categories within the dataset. In this case, the stacked bar chart allows us to easily compare the proportions of different age groups and races among the respondents who have taken the survey. This type of visualization can help identify patterns and trends in the data that may not be immediately apparent from looking at the raw numbers alone. Additionally, a well-designed visual representation of the data can make it easier for others to understand and interpret the results of the survey. It is important to give this kind of visualization to eliminate potential bias. For example, all the white people who answered this survey might have been seniors, so if we made a visualization showing the different races' employment status, it might give the impression that white people don't work at all. Therefore it is important to give a representation of the people who answered the survey. Overall, using a stacked bar chart to visualize the age and race of survey respondents can be a useful way to gain a better understanding of the data and communicate the findings to others.

Results

In this section, the results of creating charts and plots that fulfill the project objectives will be presented, along with any necessary assumptions or data adaptations that were made in order to create the visualizations. It will also be discussed whether the intended goal of the report has been reached, and any other additional findings.
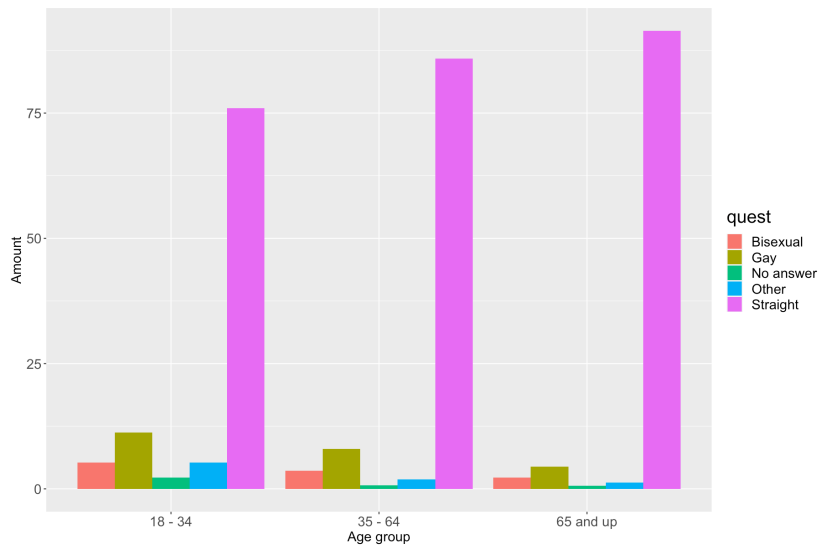
Charts

Understanding our target group

This stacked bar chart shows the concentration of each ethnicity of all of the respondents. This chart does not necessarily answer formulated questions, it still does provide some insight in the data. There are several reasons why it is important and useful to create a chart showing the different ethnicities of respondents in a

survey and the percentage each ethnicity represents. Firstly, such a chart can help to highlight any potential biases or imbalances in the sample population. If certain ethnicities are significantly underrepresented in the survey, it may affect the reliability and generalizability of the results. By creating a chart, it becomes easier to identify any potential issues with the sample and to adjust the analysis accordingly. Finally, a chart like this can be useful for presenting the results of the survey in a clear and concise manner, making it easier for others to understand and interpret the data. By visualizing the data in this way, it becomes easier to see the overall picture and to identify any important trends or patterns. Overall, creating a chart showing the different ethnicities of respondents in a survey and the percentage each ethnicity represents is an important and useful tool for understanding and interpreting the data. We can see on the chart that the white people represent the majority in all age groups, but they have an even stronger representation further up we go in the age groups. This means that our data for races such as black and hispanics become less and less accurate and potentially biased the higher up we look in the age groups of respondents. This is an important factor to consider when analyzing the data.
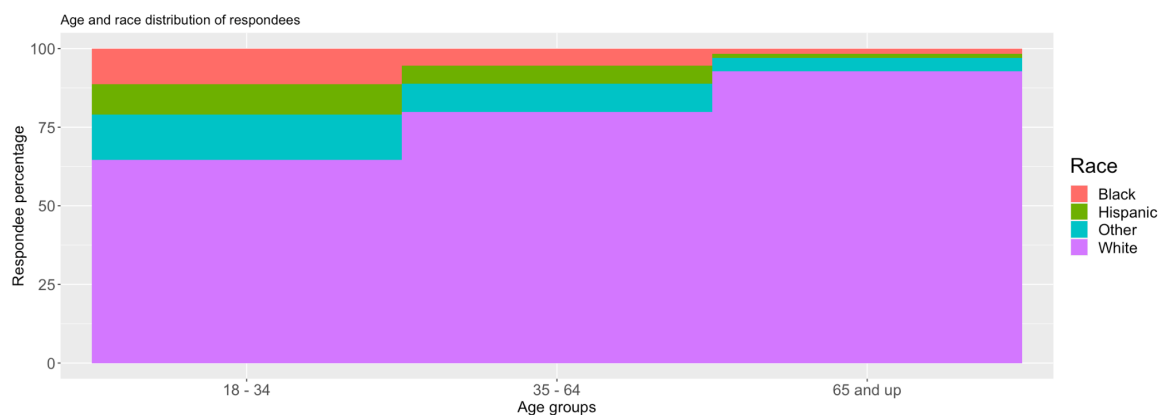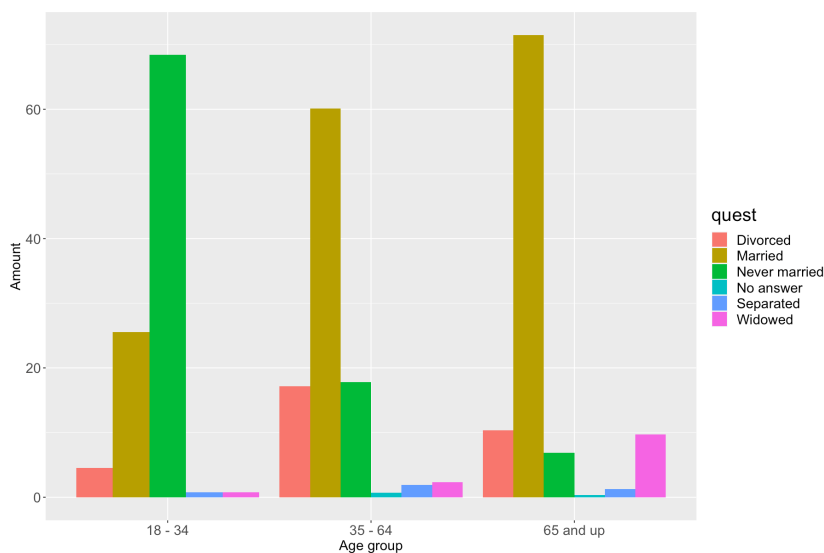

Figure 1: Stacked bar chart

Figure 2: Bar chart - What is your sexual orientation?

This barchart shows us that the majority of respondents of the survey are straight. Using this, we could expand our graphs to see whether there are any difference to how masculine men feel, based on their sexual orientation.

Figure 3: Bar chart - What is your marriage status?

The graph shows the distribution of the marriage status of the respondents. This could be used to investigate whether marriage has any effect on whether men feel masculine or not

Bar chart

The bar chart shows the distribution of answers to each question in the survey on masculinity taken by adult men. The chart includes separate bars for different age groups (18-34, 35-64, and 65+). The height of each bar corresponds to the number of respondents who answered the survey question and fell into the corresponding age and racial group. For example, if the survey asked respondents to rate their level of agreement with statements about traditional masculine traits, the chart would show separate bars for each age and racial group, with the height of each bar representing the number of respondents who agreed with that statement. This bar chart allows the reader to quickly compare the responses of different age and racial groups, providing valuable insights into the ways that these factors may impact beliefs about masculinity.
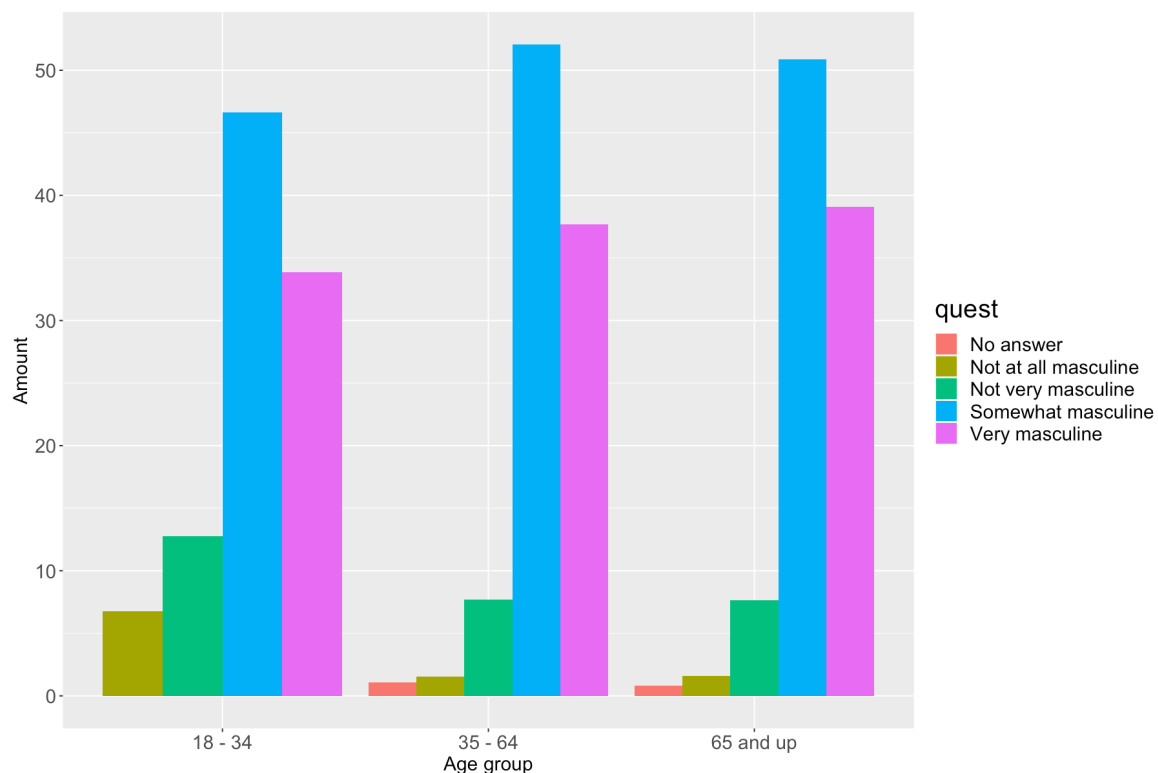

Figure 4: Bar chart - How masculine do you feel?

Horizontal bar chart

The stacked bar chart shows what men worry about, about themselves. Men might feel a lot of pressure from the media, people and so on, as to how they should look

or act, and this graph helps visualize what men actually worry about. All of these factors are grouped on a single chart to easily give a holistic overview of the ranking of these factors. This showed that the vast majority of men do not care about their height, at least not worry about it, but are actually more worried about their weight.
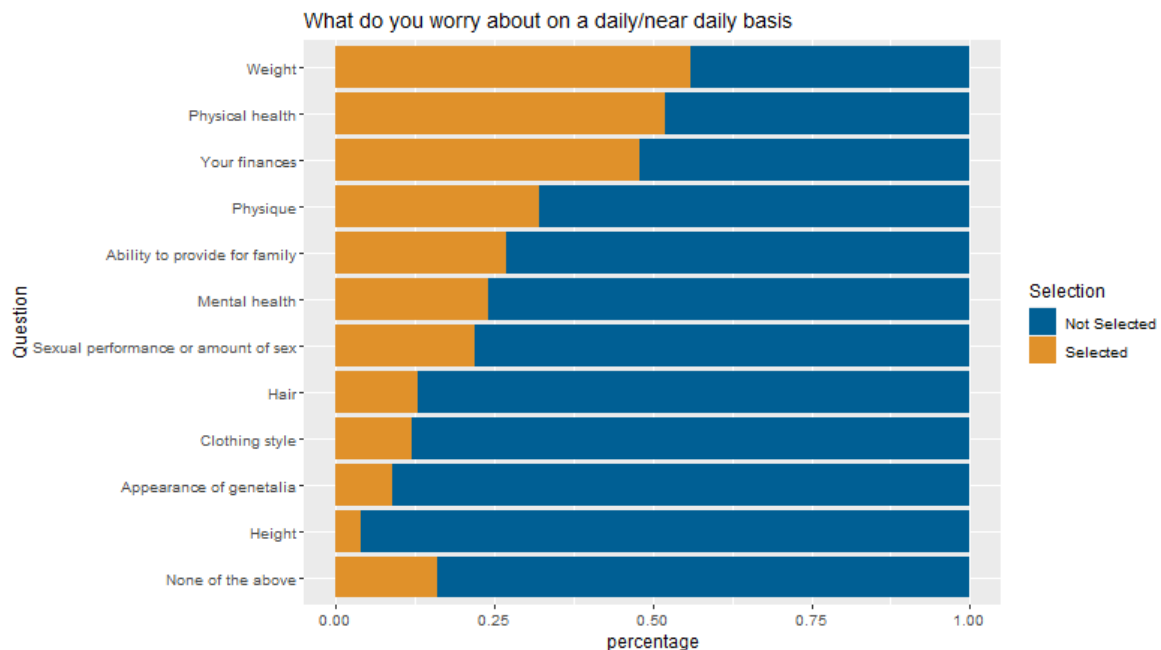

Figure 5: Horizontal bar chart - What do you worry about on a daily/near daily basis?

Animated chart

The stacked bar chart gave a great overview about what men worry about, but we wondered if there was a noticeable difference between what men worry about based on what race they were. A great way to show this was to animate the graph, cycling through respondents based on their race.This shows that generally, the top three factors stay in the top three, meaning that the major factor all men, across race, worry about are Weight, Physical health, and their finances. It was interesting to see that among Asians, their physical health was much more worrying to them than their weight, where weight was higher, or much closer to being higher, among other races. The animated chart is available in our Shiny App.

Count chart

After having mapped the ordered categorical to numeric value, we created a count plot that is a scatter plot, where instead of each observation representing a dot, dots were grouped together, and increased the size of that group. This was done, because the range was small, and the majority of dots would be on top of eachother, and it would not explain anything. The count char visualized questions "In general, how masculine or "manly" do you feel?" and "How important is it to you that others see you as masculine?". This shows a distribution of our observations that the majority of them saw themselves as at least somewhat masculine, but the majority of those who did, also felt that it was important that they were perceived as masculine. Our count char also includes brushing, so that you can mouse over a point, to see details about it.
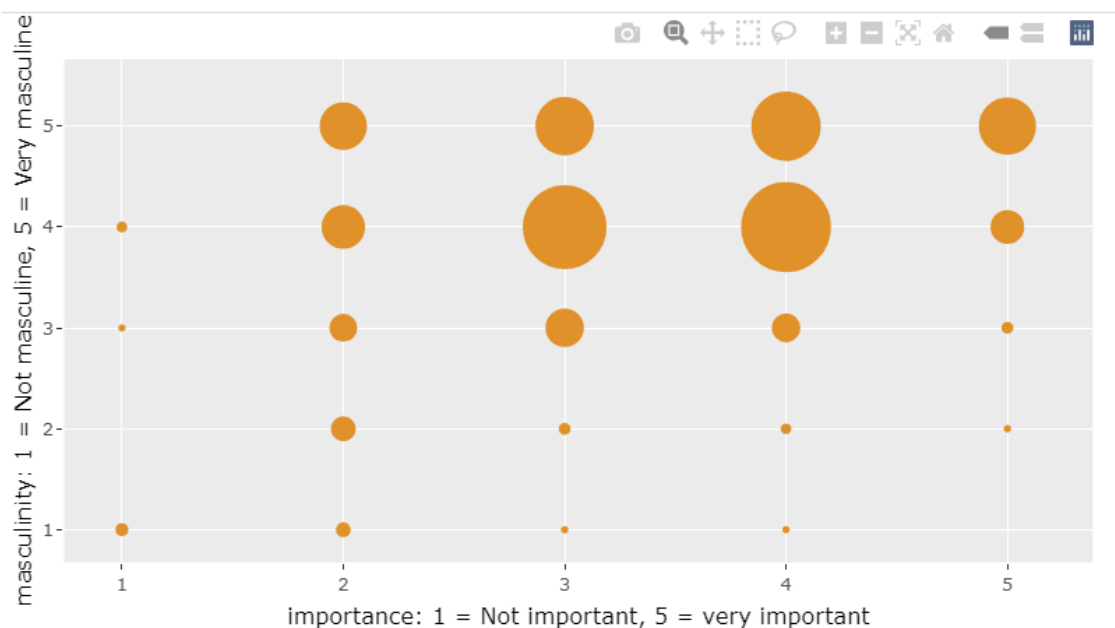

Figure 6: Count chart

Density plot

The density plot shows us how important it is to the participants, that they are perceived as masculine, grouped by race. This shows that whites are, the most in the middle, of caring or not caring, where we see in the other races, that the

distribution curve is going upward, meaning that generally, it is more important to them, that they are perceived as masculine.

Figure 7: Density plot
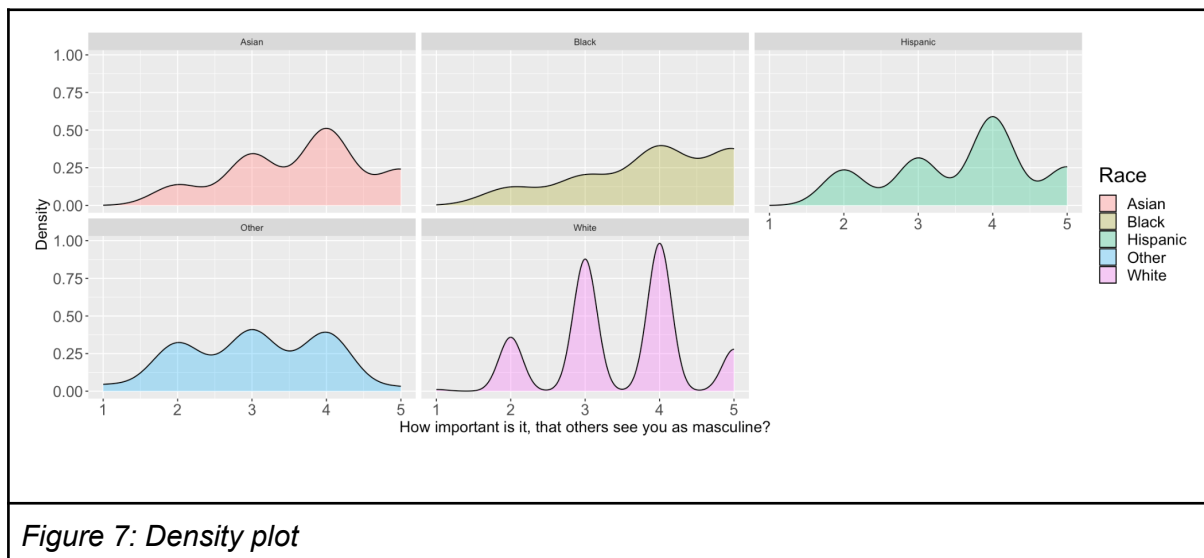
Discussion

In this project, data visualization has been used to examine the relationship between masculinity and various demographic factors, such as age, race, and income. The data used in this project comes from a survey conducted by FiveThirtyEight and WNYC, which asked men about their perceptions of masculinity and how it affects their behavior and relationships. Through the use of charts and plots, the group has been able to visualize and answer the project objectives.
How the answers are distributed in different age groups is clearly seen in the barchart, where a specific question is selected and the distribution of answers are grouped by age.

Conclusion