Data Visualization: Project Report

Heart Disease

Morgan Gant

Rachael Doyle

April 13th, 2022

**Abstract:**

The Personal Key Indicators of Heart Disease data set includes many significant indicators that have been linked to heart disease. Heart disease is one of the leading causes of death for people in the United States. Catching and preventing factors that have the greatest impact on heart disease can reduce cases. Machine learning has allowed specialists to detect patterns in patients and predict the chances of a patient being diagnosed with heart disease. The goal of this project is to determine the following: If race has an influence on heart disease if age and gender are the same (i.e. Are 60–80-year-old white women as likely to develop heart disease as 60-80 year old Asian or black women?), if physical activity can lower your chances of heart disease (even if you smoke), and if there is a relationship between being told you’ve had a stroke and heart disease (If there is, does this relationship stay consistent among races? Age?).

**Table of Contents**

**Introduction 3**

**Data Description 3**

**Analysis**

Initial 7

Heart Disease and Influence on Races 10

Physical Activity and Smoking Effects on Hear Disease 12

Stroke and Heart Disease 13

**Conclusion 15**

**Project Implications 15**

**Further Questions 15**

**References 15**

**Introduction:**

In the United States, Heart Disease is the leading cause of death for men and women. It even expands to the leading cause in most racial and ethnic groups. Every 36 seconds, someone dies from heart disease in the US. Between 2016 and 2017, heart disease cost the US close to $363 billion. These costs include health care services, medicine and declined productivity due to passing away. Heart Disease is a very broad topic that can be broken down into many sub-categories such as Coronary Artery Disease, Heart Attack, Heart Infection, Disease of Heart Muscle, Heart Valve Disease, etc. The most common heart disease is Coronary Artery Disease, in 2019 it killed about 360,900 people. On average about 6.7% of the population has this disease. 2 out of 10 deaths from Coronary Artery Disease occur in adults younger than 65 years old.

There are several risks that can lead to heart disease and a few of them are smoking, high blood pressure and high blood cholesterol. Your lifestyle also has a big impact on your risk of getting heart disease. Things to consider are your activity levels, diet, alcohol consumption and weight.

When coming upon this data set, we viewed it as significant as it targets many high-risk factors. It contains data from a diverse range of subjects which allows for a more thorough investigation of determining who is at the greatest risk to be diagnosed with heart disease based on these specific factors. We will focus on certain key factors we deem might have the strongest relationship with heart disease and how a combination of key factors affects heart disease.

**Data Description:**

The Personal Key Indicators of Heart Disease data set was taken from Kaggle, an interactive site that allows users to post and use data sets, build models, work with others, and solve challenges. This data set was obtained through adult interviews done over the phone by the Behavioral Risk Factor Surveillance System (BRFSS) in collaboration with the CDC in 2020. The original data set contained 401,985 interviews and 300 variables and was condensed down to 319,796 interviews and 18 variables.

**Row and Column Descriptions:**

The data set contains 319,796 conducted interviews (rows) and 18 heart disease related variables (columns). Below is a detailed table of each variable’s name, data type, description and how many missing values it had.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Name** | **Data Type** | **Description** | **Missing values** |
| Heart Disease | Categorical | Does the subject have heart disease? | 0 |
| BMI | Numerical | Subject body mass index | 0 |
| Smoking | Categorical | Have you smoked at least 100 cigarettes in your entire life? [Note: 5 packs = 100 cigarettes] | 0 |
| AlcoholDrinking | Categorical | Heavy drinkers (adult men having more than 14 drinks per week and adult women having more than 7 drinks per week? | 0 |
| Stroke | Categorical | Has the subject ever been told they have had a stroke? | 0 |
| PhysicalHealth | Numerical | Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 was your physical health not good? | 0 |
| MentalHealth | Numerical | Thinking about your mental health, for how many days during the past 30 days was your mental health not good? | 0 |
| DiffWalking | Categorical | Do you have serious difficulty walking or climbing stairs? | 0 |
| Sex | Categorical | Sex of subject- Male or female? | 0 |
| AgeCategory | Numerical | Age category of subject | 0 |
| Race | Categorical | What race is the subject? | 0 |
| Diabetic | Categorical | Is the subject diabetic? | 0 |
| Physical Activity | Categorical | Is the subject physically active? | 0 |
| GenHealth | Categorical | How is the subject’s overall health? | 0 |
| SleepTime | Numerical | How many hours does the patient sleep at night? | 0 |
| Asthma | Categorical | Does the subject have asthma? | 0 |
| KindeyDisease | Categorical | Does the subject have/had kidney disease? | 0 |
| SkinCancer | Categorical | Does the subject have/had skin cancer? | 0 |

**Project Tasks:**

We have selected this data set to analyze with the goal of creating visualizations that highlight the key factors and symptoms that lead to heart disease. Through the creation of charts, graphs, and interactive dashboards, we will address the following questions:

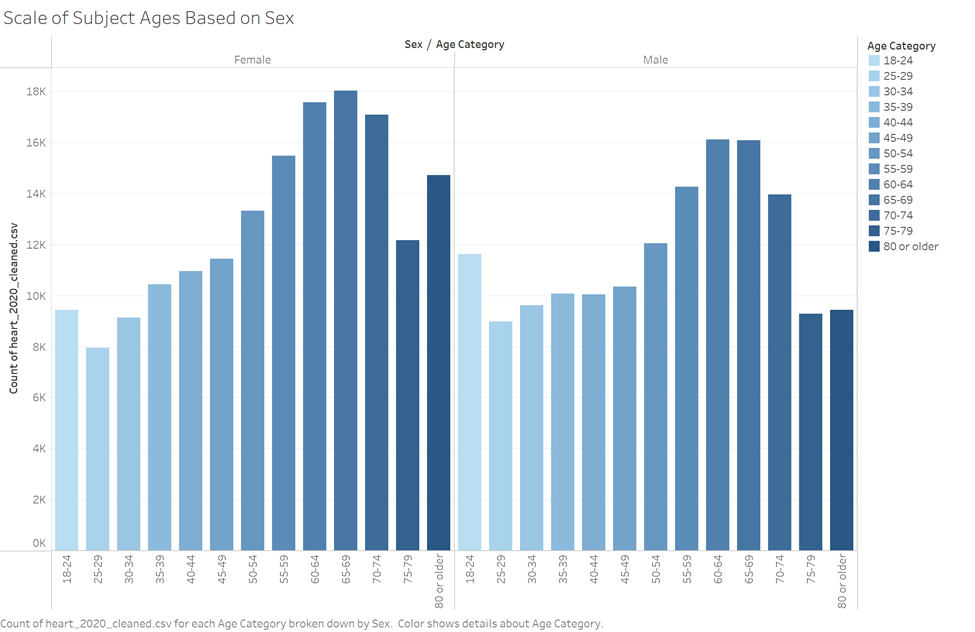
* Does race have an influence on heart disease if age and gender are the same?
  + What races have been recorded during interviews?
  + i.e., Are 60–80-year-old White women as likely to develop heart disease as 60–80-year-old Asian or Black women?

* Can physical activity lower your chances of heart disease even if you smoke?
  + What portion of the interviewed people are smokers?
    - Is there a relationship between smoking and developing heart disease?
  + Is physically active significant to interviewees in this data set?
  + Which age group has the most people diagnosed with heart disease?
* Is there a relationship between being told you’ve had a stroke and heart disease?
  + What portion of the interviewed people have been told they have had a stroke?
  + If there is a relationship, does this relationship stay consistent among races? Age?

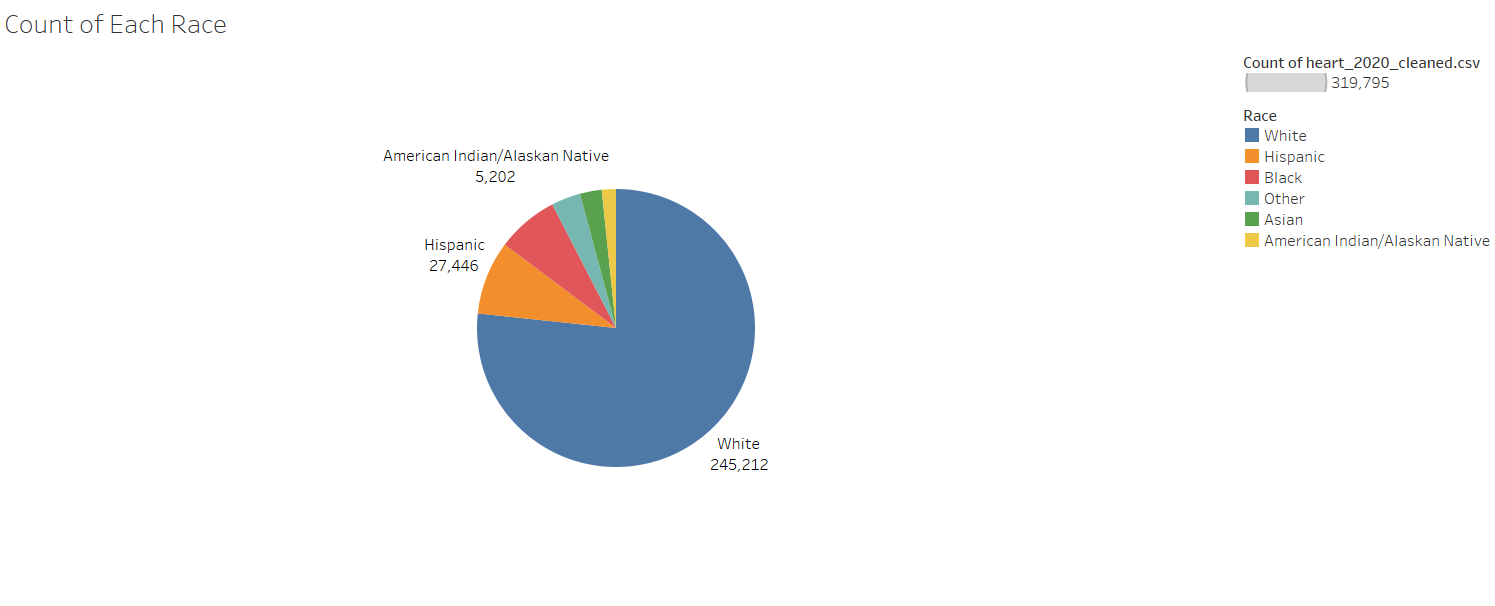
**Analysis:**

When looking at our data we deemed it important to get to know the people interviewed. Before we started analyzing the data for our overall project questions, we analyzed age, race, and selected key factors of heart disease. Below are our findings.

Exhibit 1: *Shows range of ages among sex of those interviewed*



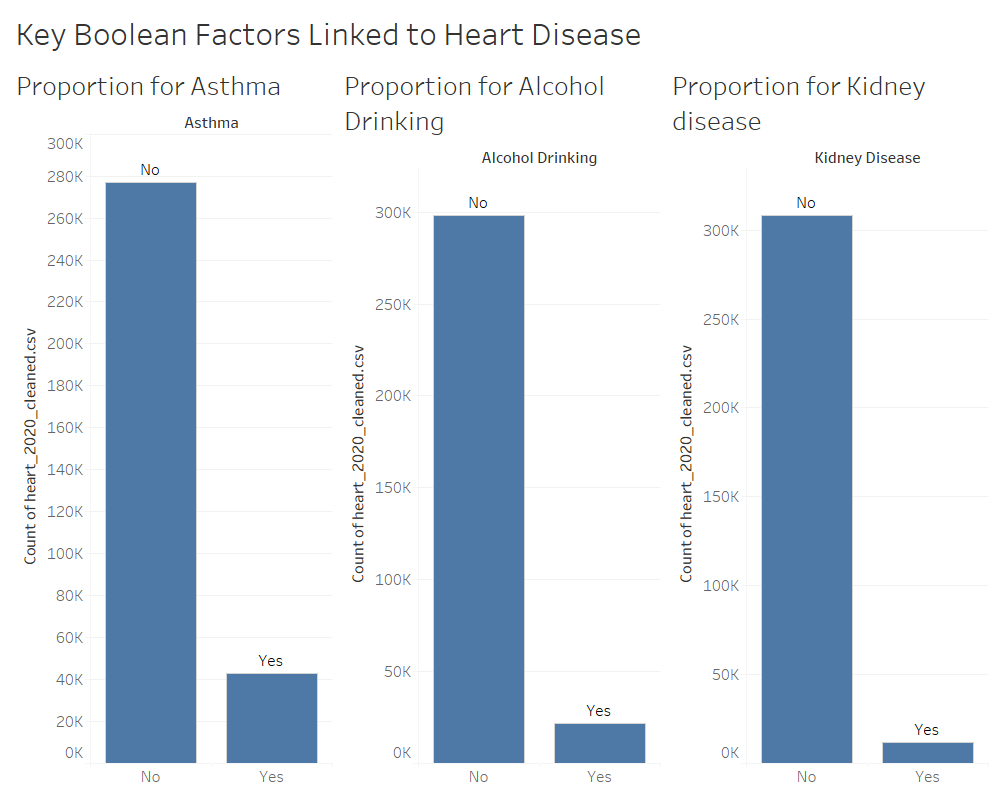
*Exhibit 1 shows most interviewed females and males are older than 50, which is expected because heart disease is most commonly heard of among older people.*

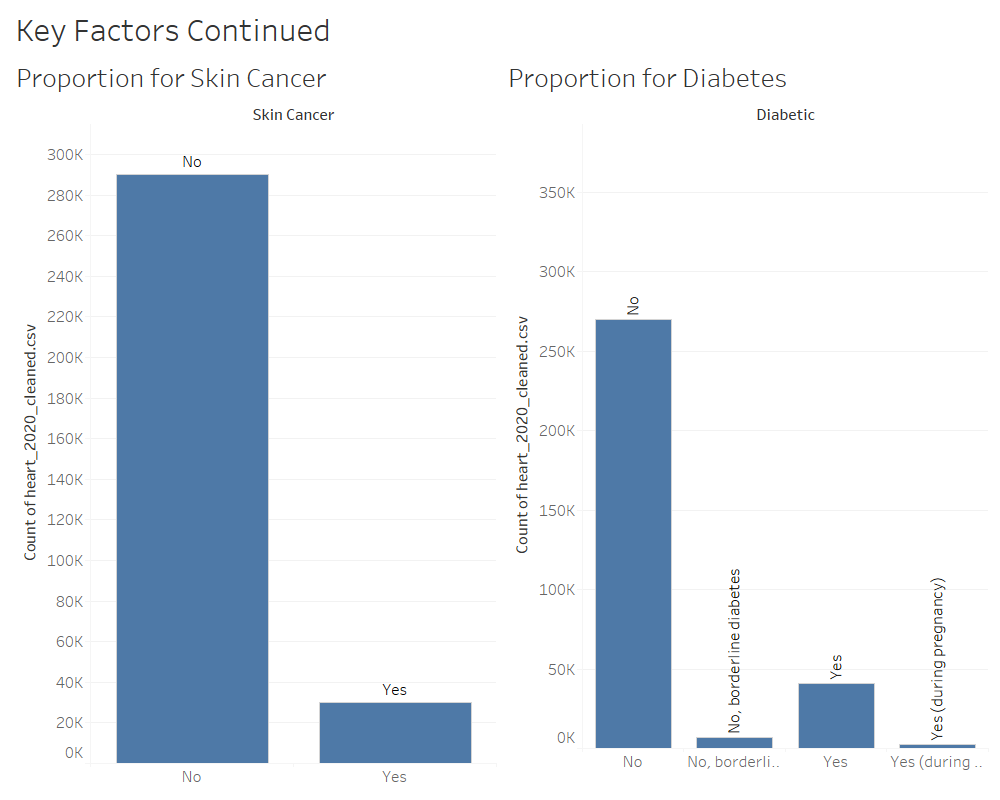
Exhibit 2: *Shows races represented in the data and their count*

|  |  |  |
| --- | --- | --- |
| Race: | Count: | Percentage of Total: |
| White | 245,212 | 76.68% |
| Black | 22,939 | 7.17% |
| Asian | 8,068 | 2.52% |
| Hispanic | 27,446 | 8.58% |
| Other | 10,928 | 3.42% |
| American Indian/Alaskan Native | 5,202 | 1.63% |

*Exhibit 2 shows the proportion of each race within the data set using a pie chart. We can see that the predominant race would be white, and the least common race is American Indian/ Alaskan Native.*

Exhibit 3: *Dashboard of significant Boolean factors*



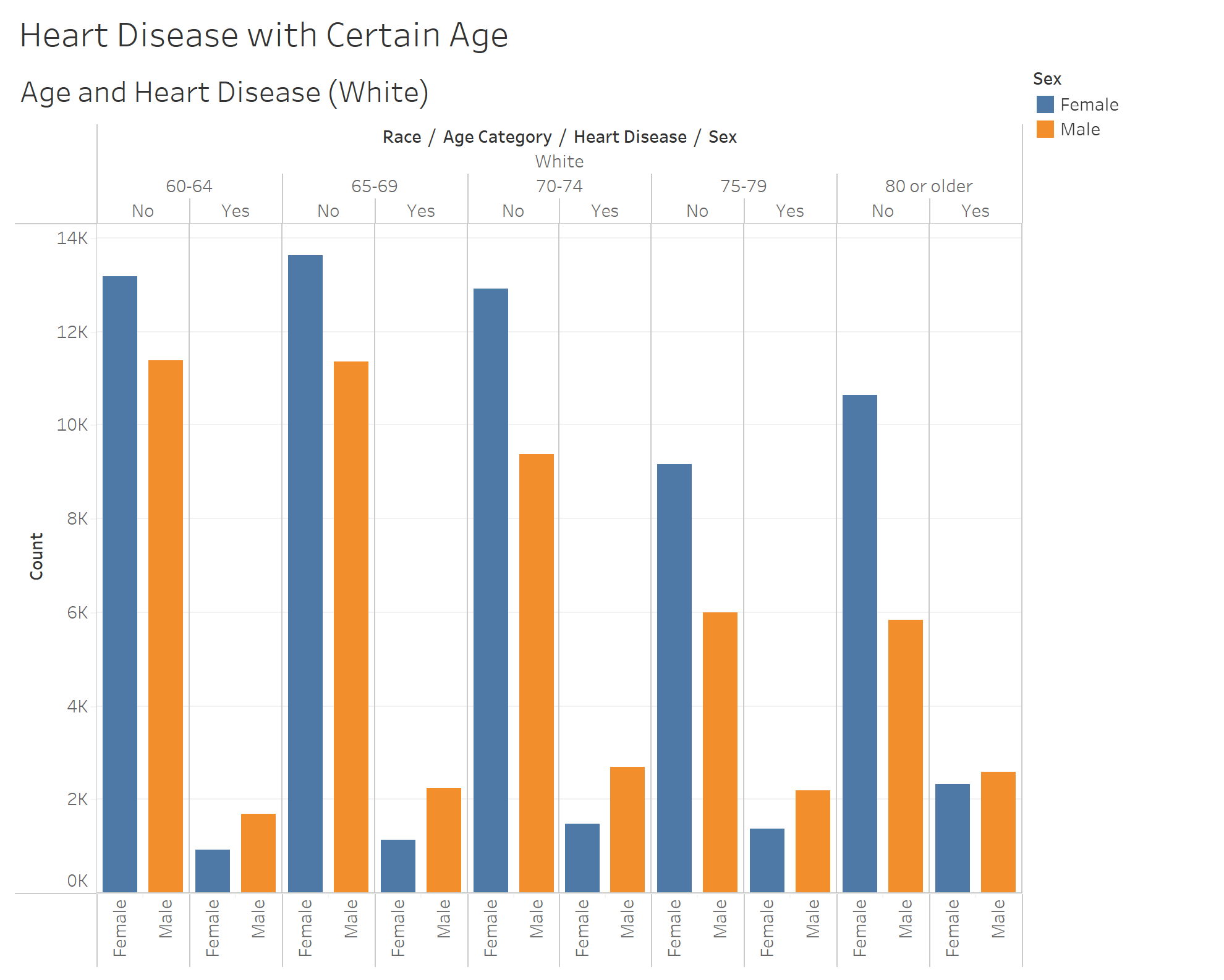


*Exhibit 3, by taking a quick look at these few key factors, we can see that most of the people interviewed do not have these qualities that are presumed to lead to heart disease. Heart disease is not factored into these questions, it was just an over all analysis for certain factors.*

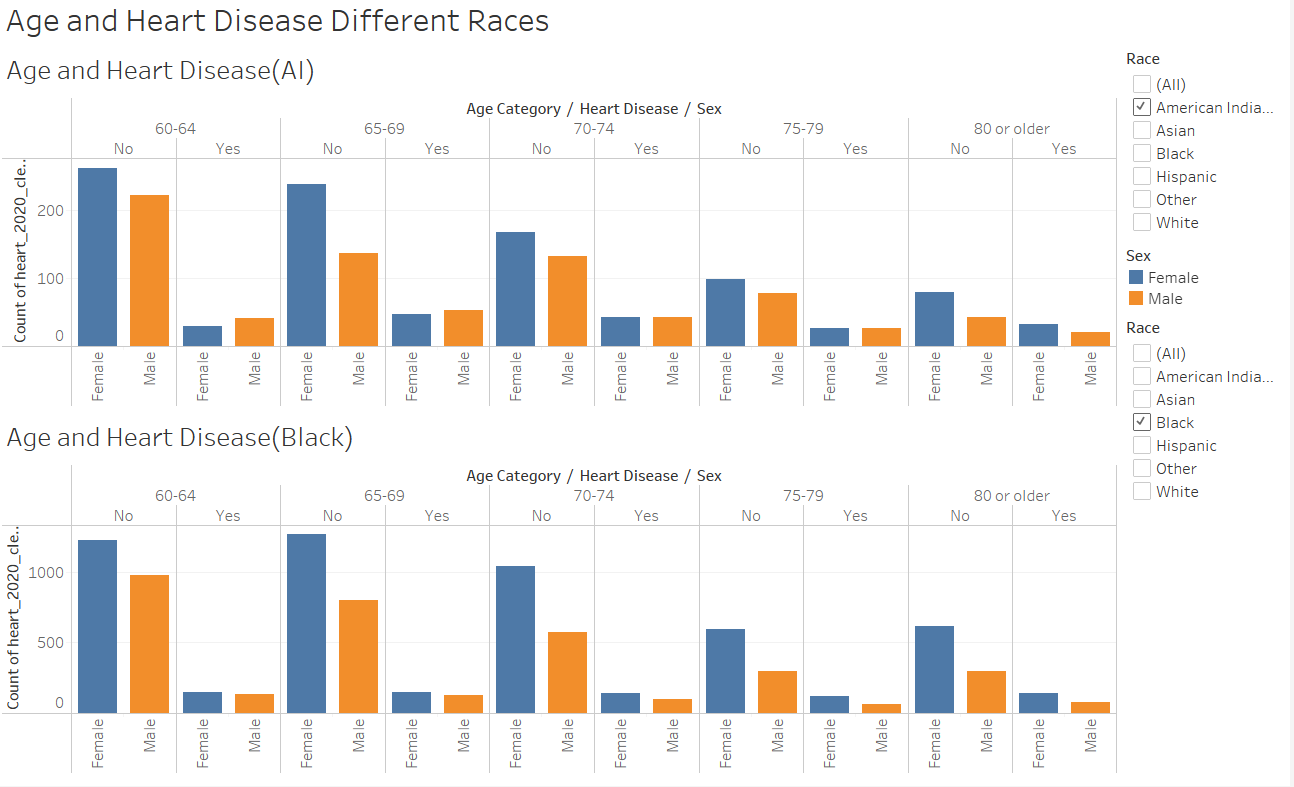
**Heart Disease and Influence on Races**

When looking at the influence on heart disease of different races, we want to determine if age and gender are consistent and how the chances of developing heart disease vary. With our data, we can pick specific races and age categories to focus on and see the proportion of interviewees who have heart disease or not.

***Figure 1.1: Heart Disease in White People Between 60-80***



***Figure 1.2: Heart Disease in Asian and Black People Between 60-80***

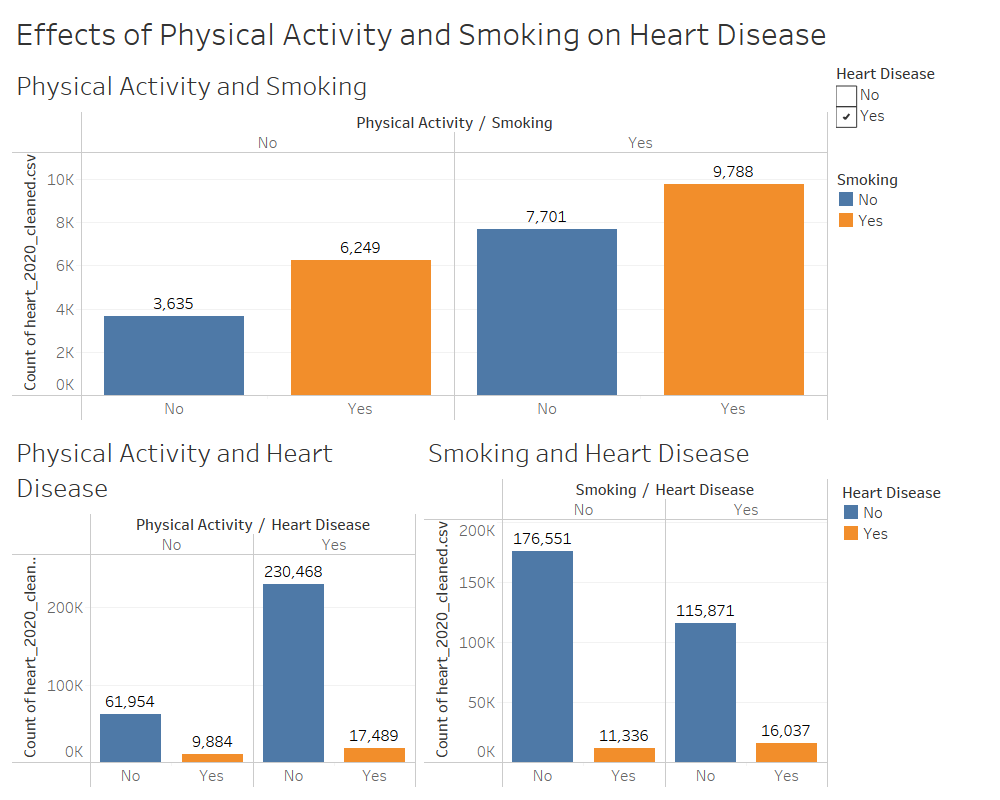


When looking at the data in ***Figure 1.1,*** we saw that most people were of the race white, so this was going to skew the data, and create the trends. Because of this we decided to use the race white as our control and compare all the races separately with this control. Asian, Hispanic, and others all followed the control with their overall trends that we saw between gender, so these charts were left out. However, American Indian/Alaskan Native and Black both had trends that were not seen within the control, as seen in ***Figure 1.2***. In the control we saw that no matter the age group men had the higher number of people with heart disease than women, however this was not the case because when looking at the Black race, there was age groups where more women had heart disease than men, and with the Alaskan Native race there was age groups with little to no difference in the number of men or women with a heart disease. From these charts we can draw the conclusion that most races follow the same patterns of no matter the age group, men had more people with heart disease than women.

**Physical Activity and Smoking Effects on Heart Disease**

When determining factors that lead to heart disease, we thought of smoking and if someone is physically active or not. Smoking is seen as primarily having a damaging effect on your lungs, whether it be popcorn lungs or lung cancer in general, however another large effect of smoking on the body includes the possibility of developing a heart disease. Not being physically active can also be damaging to the body because it can lead to obesity, which also leads to heart disease. Because of these factors on someone's health, it could be assumed that those who are not physically active and are smokers, are going to have a much higher chance of developing a heart disease. This assumption led to the idea of creating a dashboard of comparing the proportion of smoking or not, alongside being physically active or not, among those with a heart disease.

***Figure 2.1: Physical Activity and Smoking Effects on Heart Disease***

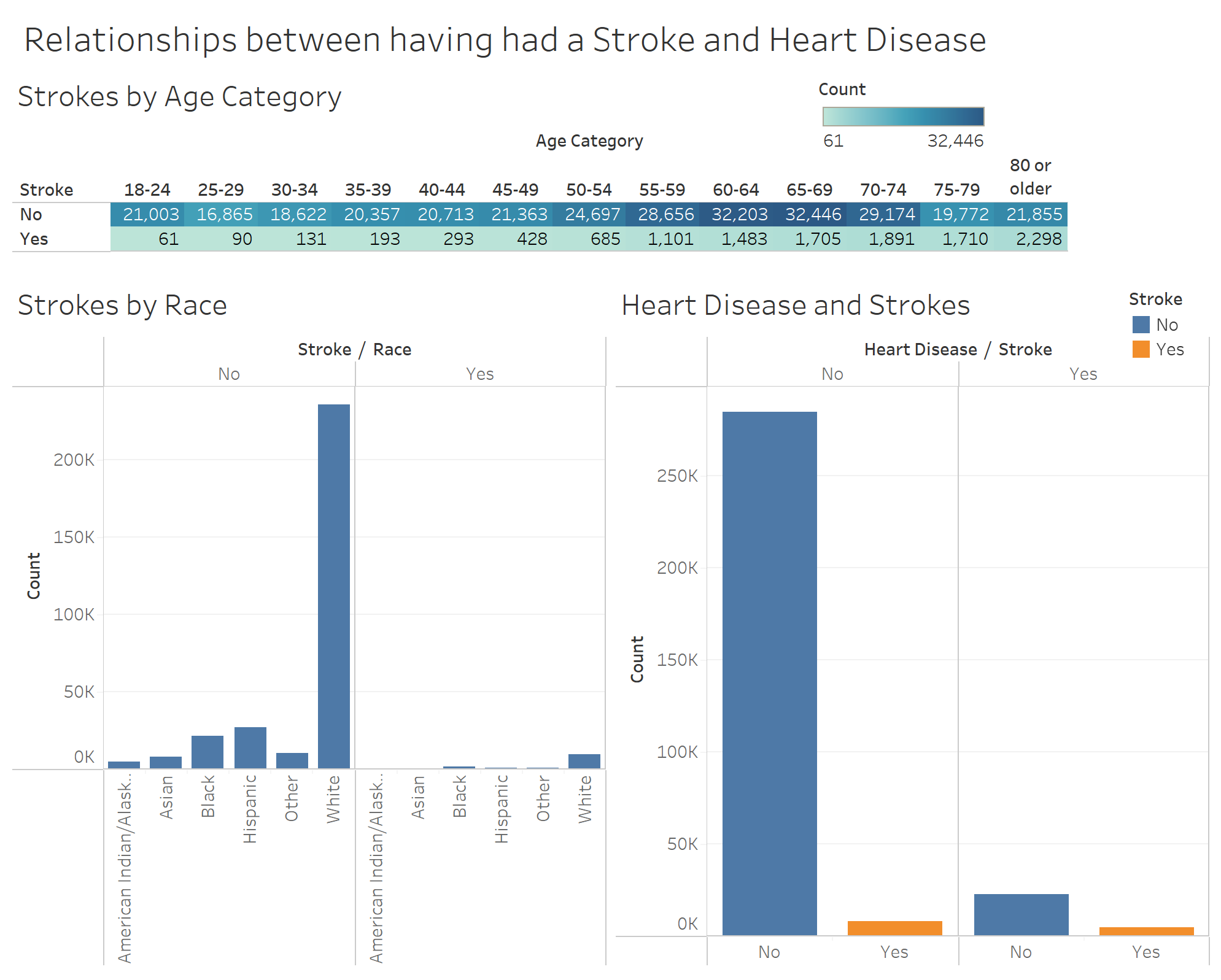


From the above dashboard, ***Figure 2.1,*** we can see that the combination of symptoms that causes heart disease the most is when someone is physically active, and a smoker. This contradicts with the prediction that not being physically active and a smoker is the main cause. There are cases of getting heart disease with any combination of these factors, however those who are physically active still are at a higher risk for developing a heart disease, no matter if they are a smoker or not. Together, these two factors did not create a larger risk factor than when they are looked at individually.

**Stroke and Heart Disease**

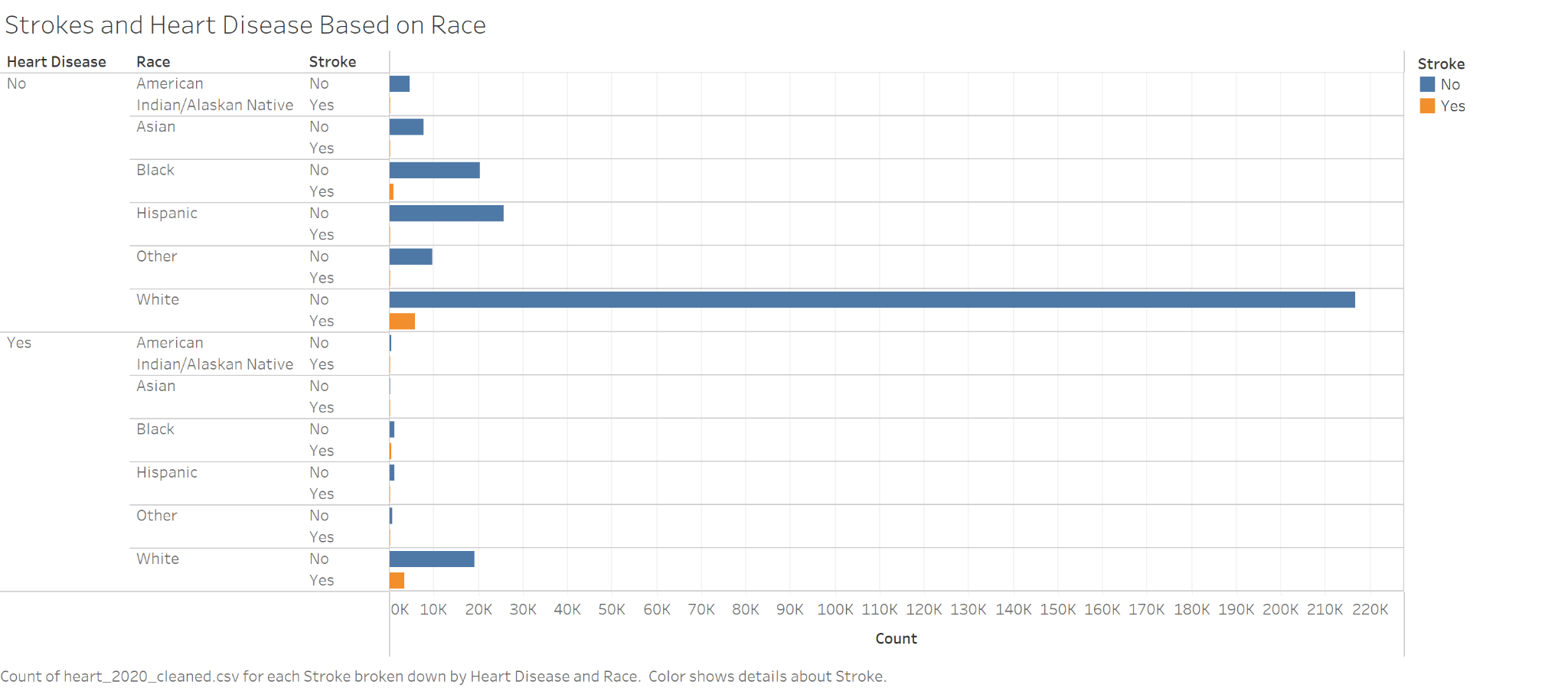
Blood pressure is a common symptom in strokes and heart disease. Strokes occur when your blood pressure gets too high and can cause internal bleeding or blocking of the arteries. High blood pressure can also cause damage to your arteries, causing less blood and oxygen to reach the heart, leading to a heart disease. Because of this significance in both areas, it’s worthwhile to see if there is a relationship between being told you’ve had a stroke and having a heart disease. We’ll look at Strokes by age category, race and by directly comparing it to heart disease.

***Figure 3.1: Heart Disease and Strokes***



Looking at the dashboard above, ***Figure 3.1,*** we have three charts displayed. One showing stroke occurrence differs based on age category, another comparing strokes based on race and the last looking at the proportion of having a heart disease and having a stroke. Focusing on the first chart, stroke occurrences were not common among younger ages including 18 through 50. After this we see a spike in strokes but the majority of the interviewed still reside in not having had a stroke. Next, the chart that shows the races and stroke responses, when looking at proportions, 4% of Whites, 5.4% of Blacks, 2.1% of Hispanics, 4.4% of Other and 5.9% of American Indian/Alaskan Natives reported having had a stroke. There isn't much difference based on racial proportions and no significant racial outliers. Lastly, looking at the chart that simply splits up having had a stroke and heart disease, we can see that the same number of responses that had a stroke was the same when looking at the yes and no responses for also having a heart disease. With all the information provided on this dashboard, we can see that there isn’t significant evidence that links having had a stroke to heart disease. Let's take a deeper look into how each race responds to having had a stroke and heart disease.

***Figure 3.2: Heart Disease and Strokes Based on Race***



The above chart, ***Figure 3.2,*** displays those who have a heart disease, their race and if they have had a stroke or not. We know from previous information that white is the abundant race, so it makes sense that there are a lot more responses from white people. As we look at the proportion of yes for stroke and yes for heart disease for each race, we don’t see a significant difference when comparing it to yes for stroke and no for heart disease. This tells us there is not a significant impact on having heart disease given the races and having been told you’ve had a stroke.

**Conclusion:**

Throughout this project we created visualizations to highlight ideas or assumptions that could be made about the causes of heart disease. We addressed how common heart disease is within each race and age category, the effects that smoking and physical activity has on heart disease, and whether having strokes influences if someone has a heart disease. We concluded that most races follow the pattern of males having a higher proportion of having heart disease when compared to women, despite common beliefs- being a smoker and physically active you’re more acceptable to heart disease than not being physically active and there wasn’t a significant relationship between aching been told you’ve had a stroke and having heart disease.

**Project Implications:**

Surprisingly, we didn’t have many issues while working with this data. One of the hardest things we noticed once we started analyzing the data was that the majority of the interviewed were White. This was difficult when trying to compare races because the data was disproportionate. Instead of comparing each race to each other, we had to compare the race to itself by taking percentages of those that answered yes from the corresponding total racial population. The other implication was that we didn’t have many numerical variables. Most of our data was categorical (Yes and No) so were we limited to the type of graphs we used and how we presented our data. We were still able to explore our data, just had to be repetitive in the chart we used and how we sorted through it.

**Further Questions:**

From reflecting upon the questions we answered, new questions were brought to our eyes. Although we had many variables to base our questions off, we are still missing many possible key factors affecting heart disease. These factors could include the date of which someone is diagnosed with a heart disease, any doctor's notes on their medications, vitals, hormone imbalances, or other pre-existing conditions that could have led to other organ failures. With these additional variables we could ask, what is the average blood pressure of those with a heart disease? What types of medications or combinations of medications lead to heart disease? Do hormone imbalances cause heart disease or symptoms that cause heart disease? Does ‘broken heart syndrome’ lead to heart disease?

**References:**

<https://www.kaggle.com/kamilpytlak/personal-key-indicators-of-heart-disease>

<https://www.cdc.gov/heartdisease/facts.htm>