

Project - Phase III: Dashboard Implementation

Aishwarya Ramaiah Kumar, Anirudh Kumar Kota,

Ganashree Naganahalli Annaiah, Spurty Mutturaj

Department of Information Technology,

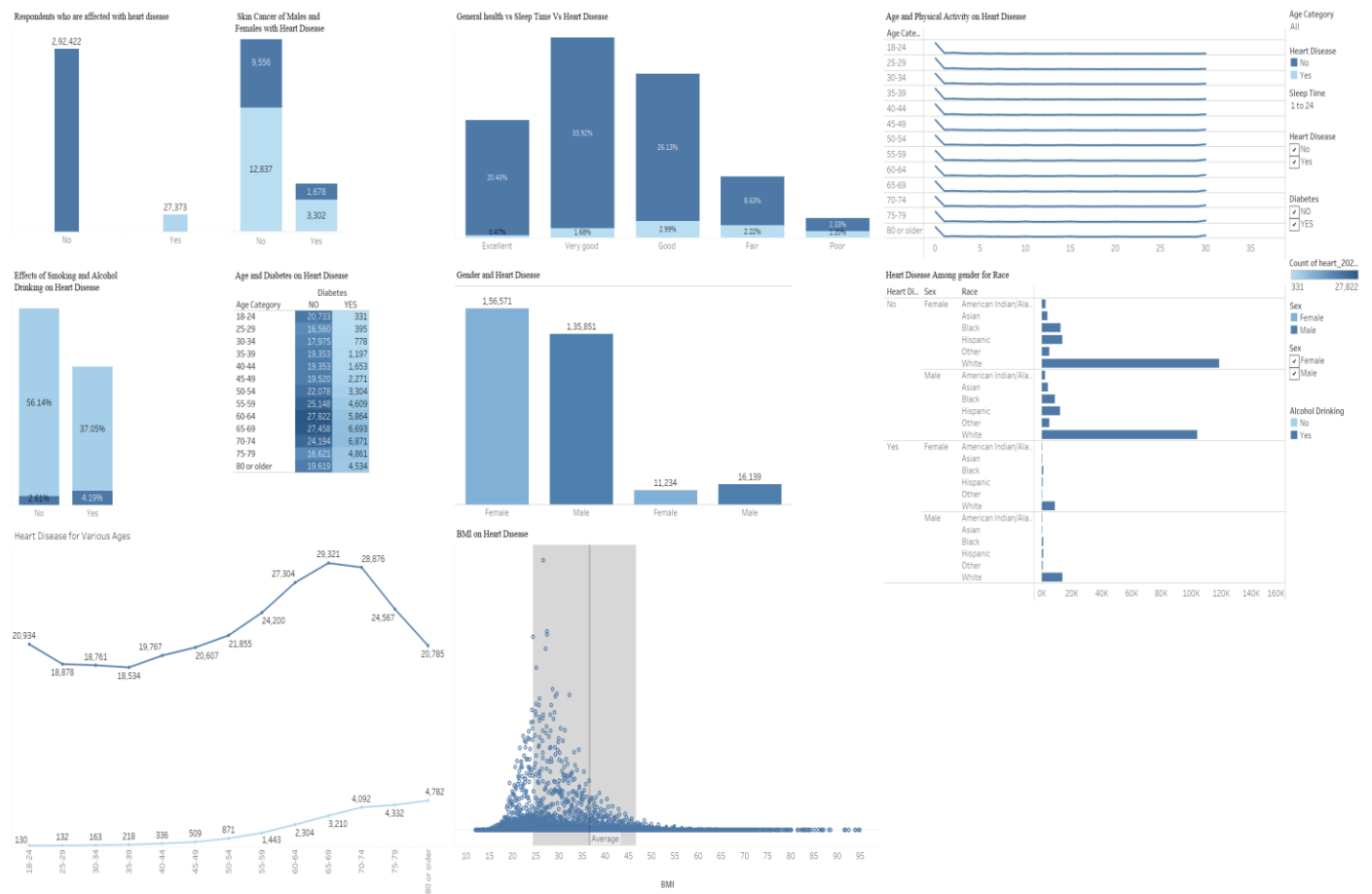
Arizona State University

IFT 598: Data Visualization

Prof. Asmaa Elbadrawy

April 24, 2022

Section 1: The Dashboard



The Dashboard contains all the visualizations that we made. This Dashboard can be used by doctors to plan the treatment based on the analysis done by the visualizations. Scientists can use this for research and development purposes. They can explore and relate many health factors. Also, nutritionists can use this dashboard to create a balanced diet plan for the consultants who approach them based on their health conditions, and age factors.

Section2: The Dataset

The Dataset we selected was personal indicators of **heart disease**. We got this dataset from Kaggle. We selected this dataset because heart disease has become one of the most common diseases and

all age groups of people are suffering from this. Also, according to the Centers for Disease Control and Prevention (CDC), at least one of the three primary risk factors for heart disease is present in half of the Americans that are almost 47 percent. Also, Heart disease is also one of the leading causes of mortality in the United States.

High blood pressure, high cholesterol, smoking, diabetes, obesity (high BMI), insufficient physical exercise, and excessive alcohol use are all risk factors. In healthcare, detecting and avoiding the causes that have the greatest influence on heart disease is essential. As a result of computational advancements, machine learning methods may be used to find "patterns" in data that can be used to forecast a patient's status.

Other data Scientists that used data for:

This dataset may be used to test a variety of machine learning techniques, the most common of which are classifier models like logistic regression, SVM, random forest, etc. However, the classes are not balanced, hence the traditional model application strategy is not recommended. Fixing the weights or under-sampling should improve the findings. Also, there is an app built using a logistic regression model based on this dataset and is present in <https://heart-condition-checker.herokuapp.com/>.

The data originated with the CDC and the Behavioral Risk Factor Surveillance System (BRFSS), which conducts yearly telephone surveys to collect information on Americans' health. Data from 2020 is included in the most recent dataset as of today.

Number of Rows and Columns:

There are 401,958 rows and 18 columns in it. There are 9 Boolean columns, 5 text format columns, and 4 decimal format columns. There were many factors (columns) that directly or

indirectly influence heart disease in this dataset, so we decided to pick the most key factors (Columns) and clean them up so they could be used in machine learning applications.

Dataset attributes and Column data types:

- **Heart Disease:** Defines the value of the respondents having heart disease. Heart Disease is a **Categorical** data type attribute. It has the values like yes or no only.
- **BMI:** Body Mass Index values of respondents in decimal format from 12 to 94.8. BMI is the **ratio** data type attribute.
- **Smoking:** The value is based on the answers yes or no given by respondents to questions like have you smoked at least 100 cigarettes or 5 cigarette packets in your entire life? Smoking is the **categorical** data type attribute.
- **Stroke:** The stroke column is filled with values like yes or no for the question was hit by a stroke ever in their life? The stroke is the **categorical** data type attribute.
- **Alcohol Drinking:** The respondents are asked if they are Heavy drinkers (adult men having more than 14 drinks per week and adult women having more than 7 drinks per week. The responses are recorded as yes or no. Alcohol drinking in **categorical** attribute.
- **Physical health:** About the respondent's physical health including physical illness and injury, for how many days during the past months did they suffered is considered and the input is taken in decimal points up to 30.0. Hence, Physical health is a **ratio** data type attribute.
- **Mental Health:** This has the values given by respondents about their mental health, gauging the decimal value up to 30.0 for how many days during the past month (about 4

and a half weeks) was your mental health not good? Mental health is a **ratio** data type attribute.

- **General Health:** The respondents are asked if their general health is either good or very good. General health is a **categorical** data type attribute.
- **Sex:** The values are either male or female. It is a **categorical** data type attribute.
- **Race:** The values like white, Black, Hispanic, and American Indian/Alaskan Native are recorded from respondents. The race is **categorical** data type attributes.
- **Age Category:** The age category is a **ratio** data type attribute. The age of respondents was recorded if female or male.
- **Diabetic:** The respondents were asked the question if they had diabetics or not. The response was recorded in yes, no, borderline diabetics format. Diabetics is a **categorical** value.
- **Physical Activity:** The answers from respondents were recorded in yes or no format. It is recorded yes if the adults reported they do physical activity or exercise during the past month other than their regular job, otherwise, it is no. Physical attribute is a **categorical** attribute.
- **Sleep Time:** The sleep hours of the respondents are recorded on an average for about 24 hours. The values are recorded in the decimal format of 1 to 24 hours. Sleep time is a **ratio** data type attribute.
- **Asthma:** The values of yes or no were recorded after asking if they ever had asthma or not. Asthma is a **categorical** data type attribute.

- **Kidney Disease:** Based on the yes or no answers of the respondent that if they had any kidney diseases such as kidney stones, bladder infection, or incontinence. Kidney disease is the **categorical** data type attribute.
- **Skin Cancer:** Respondents were asked if they ever had skin cancer. The values were recorded in yes or no format. Skin cancer is a **categorical** data type.
- **Different Walking:** The respondent's answers were recorded to questions like do they have serious difficulty in walking or climbing stairs? The values are taken in yes or no format. The Different walking column is a **categorical** data type attribute.

What can be visualized using this dataset:

We can use this dataset for different visualizations and gain important information. Some examples are:

- We can find out who is more prone for heart disease
- Which age grouped people need to take precautions from getting heart disease.
- Effects of all habits like drinking, smoking, sleepless nights, no exercise and bad mental health can be visualized.

Section 3: Dashboard Users

The Heart Disease dataset is helpful to a variety of professionals related to the medical field. Exploring and visualizing the dataset will help advancement in research and hopefully advance the clinical diagnosis and early detection and treatment of heart diseases.

The prospective users of our dashboard include:

- **Doctors:** The doctors can use the dataset to plan the course of treatment, when the patient has been admitted to the hospital based on his history. Considering factors like diabetes, kidney disease and physical health is important to reach a conclusion.
- **Scientists:** scientists can use it for research and development to predict/analysis the reason/likeliness of a heart disease. The relation between various health factors can also be explored using the dashboard.
- **Nutritionists:** Depending on their health conditions, people will have varying restrictions when it comes to diet. A nutritionist can customize the dietary options available for heart patients based on various parameters.

Section 4: List of Final Questions

1. Visualizing the heart disease for Different Age categories?
2. Respondents who are affected with heart disease.
3. How is heart disease affected by Various sleep timings and Health Condition?
4. How is age category, physical activity and heart disease related?
5. Which gender is most prone to heart disease?
6. How is heart disease affected for Gender and Race?
7. How much will skin cancer be responsible for heart disease among males and females?
8. How are smoking and alcohol drinking contributing to heart disease?
9. How does BMI affect heart disease for a person?
10. How are diabetes, age, and heart disease-related?

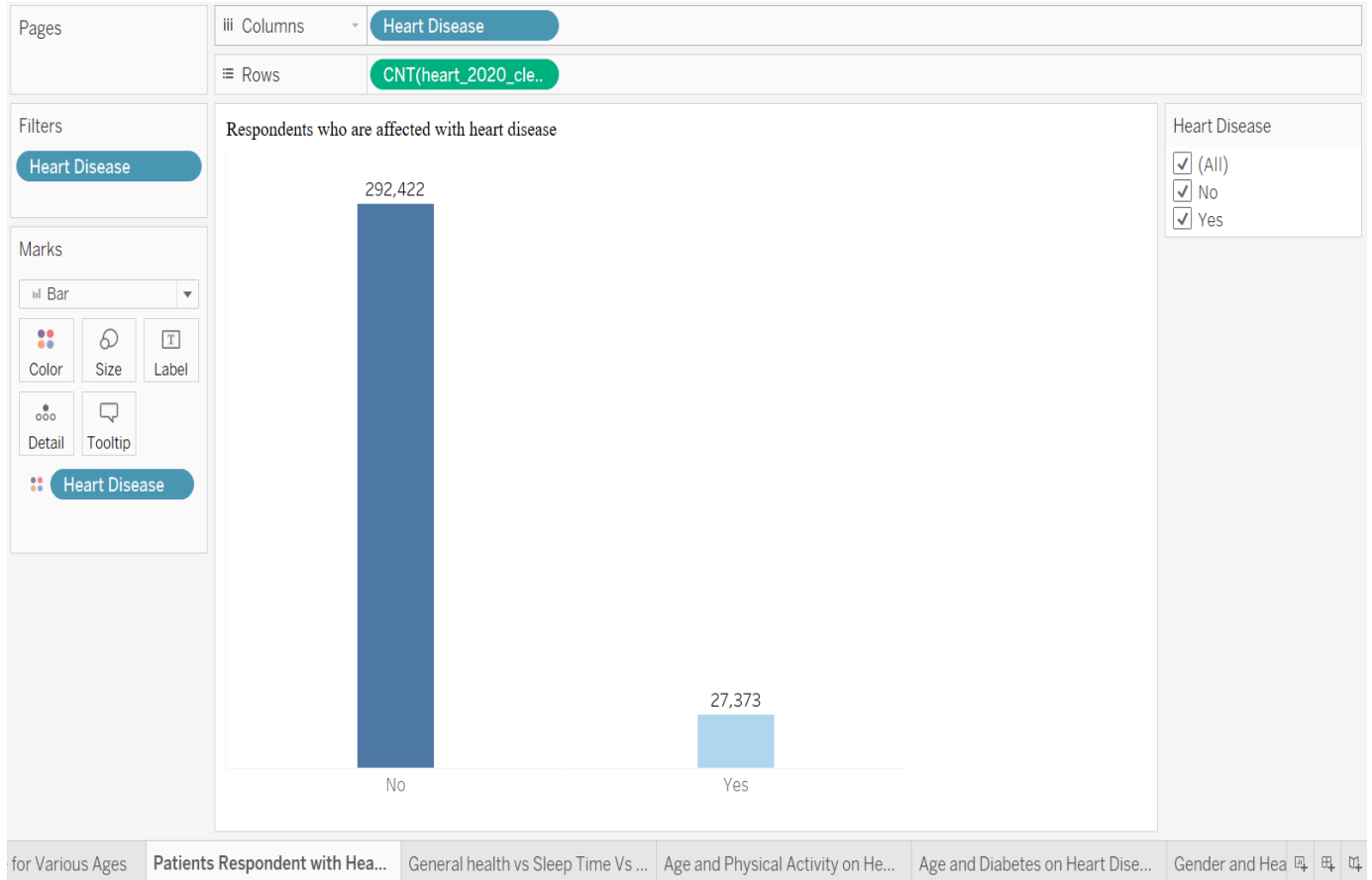
Section 5: Dashboard Plots

1) Visualizing the heart disease for Different Age categories?



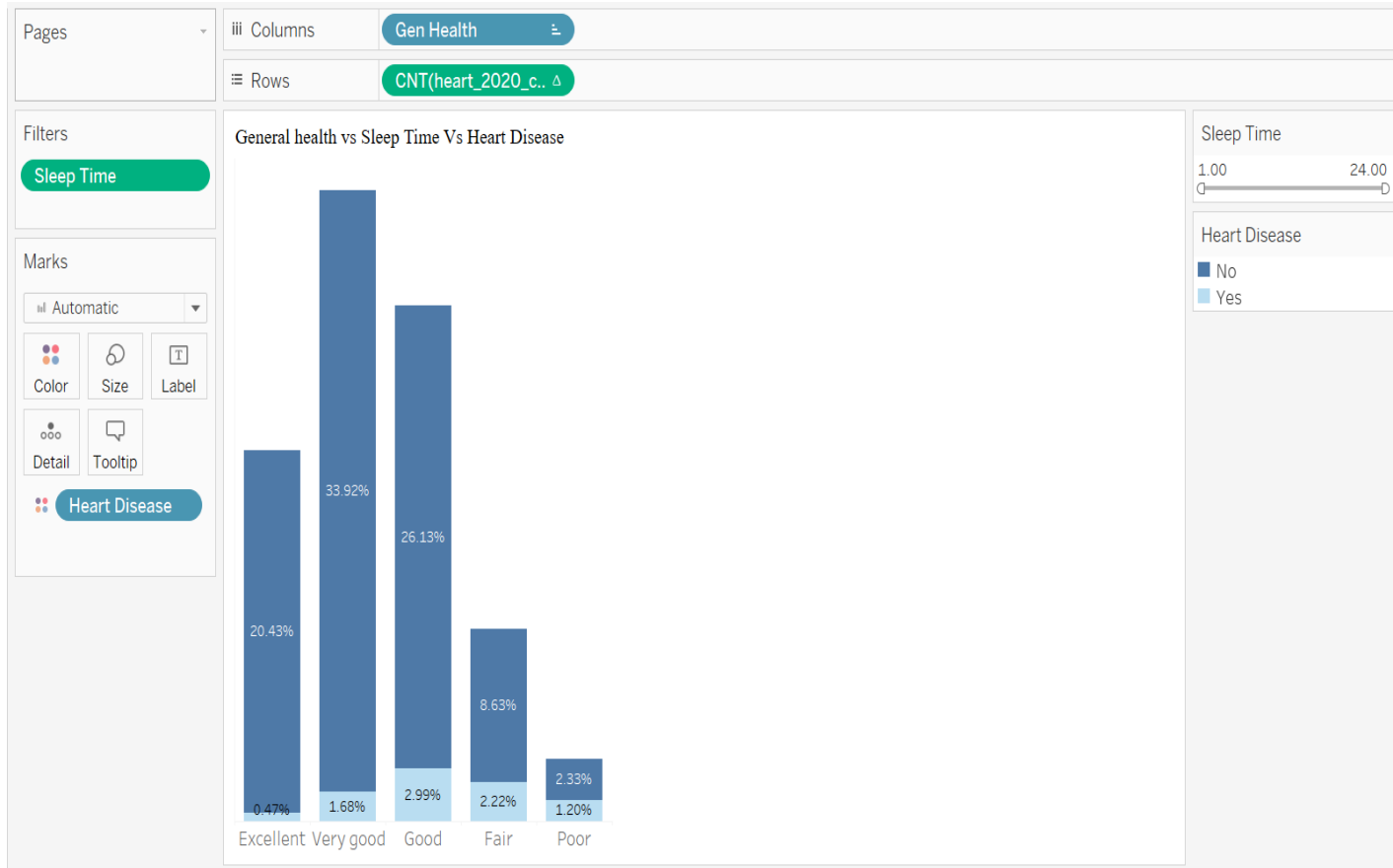
This visualization is created using a Line graph, and an interactivity filter named Age category has been linked to the visualization. It is visible on the right side of the picture and the orange color line indicates the count of people affected with heart disease and the blue color indicates the count of people who are not affected with heart disease. By selecting the different ages in the parameter field, you can get values. The pre-attentive attribute in this chart is the position.

2) Respondents who are affected with heart disease.



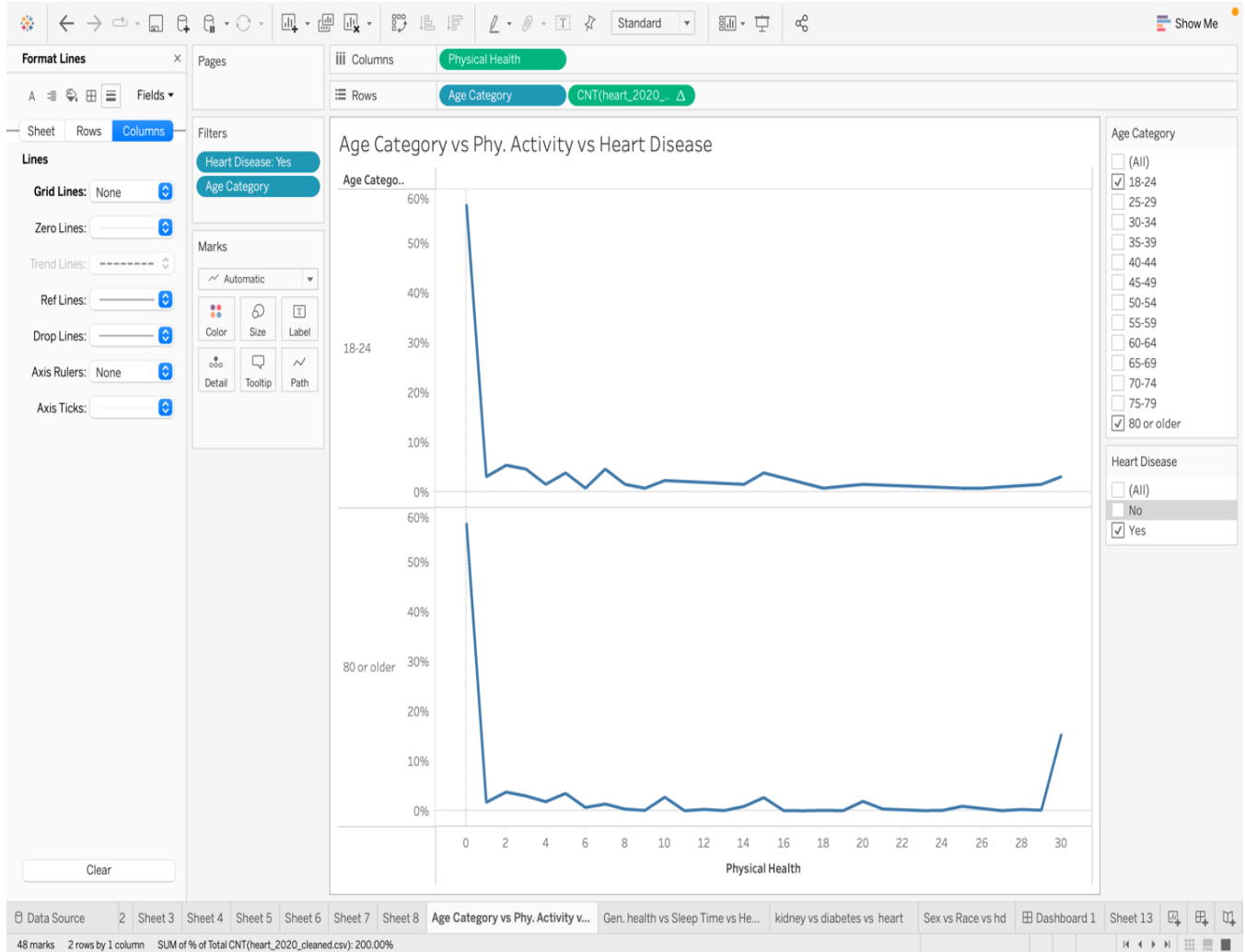
This visualization is created using a Bar chart, an interactivity filter named heart disease has been linked to the visualization. The pre-attentive attribute here is length and color.

3) General health vs Sleep Time Vs Heart Disease



This visualization is created using a bar chart, and an interactivity filter named Sleep time has been linked to the visualization. The pre-attentive attribute in this type of chart is Length and Color. It is visible on the chart that the blue color represents no chances of getting heart disease, and the orange color represents high chances of getting heart disease.

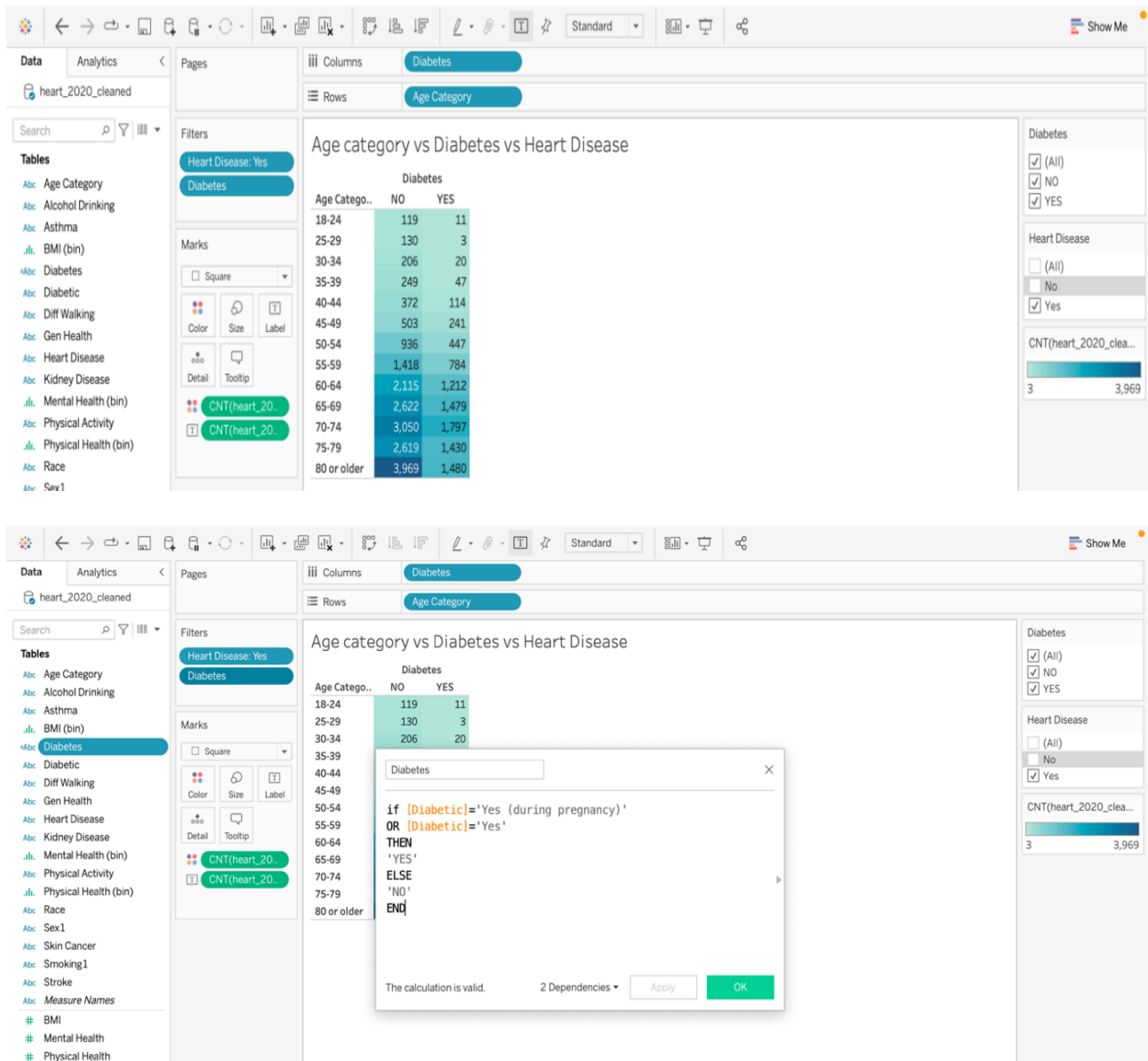
4) Age and Physical Activity on Heart Disease



This visualization is created using the line chart, an interactivity filter named heart disease, and age category has been linked to the visualization. It is visible on the chart that the blue colored line represents the rise and fall of the heart disease rate in individuals of different age groups. The pre-attentive attribute of this graph is the position. On this right-hand side of the chart, we can see the age groups of the people in the checkboxes which when selected show the data of individuals of the respective age group who are most likely to suffer from heart diseases.

5) Age and Diabetes on Heart Disease

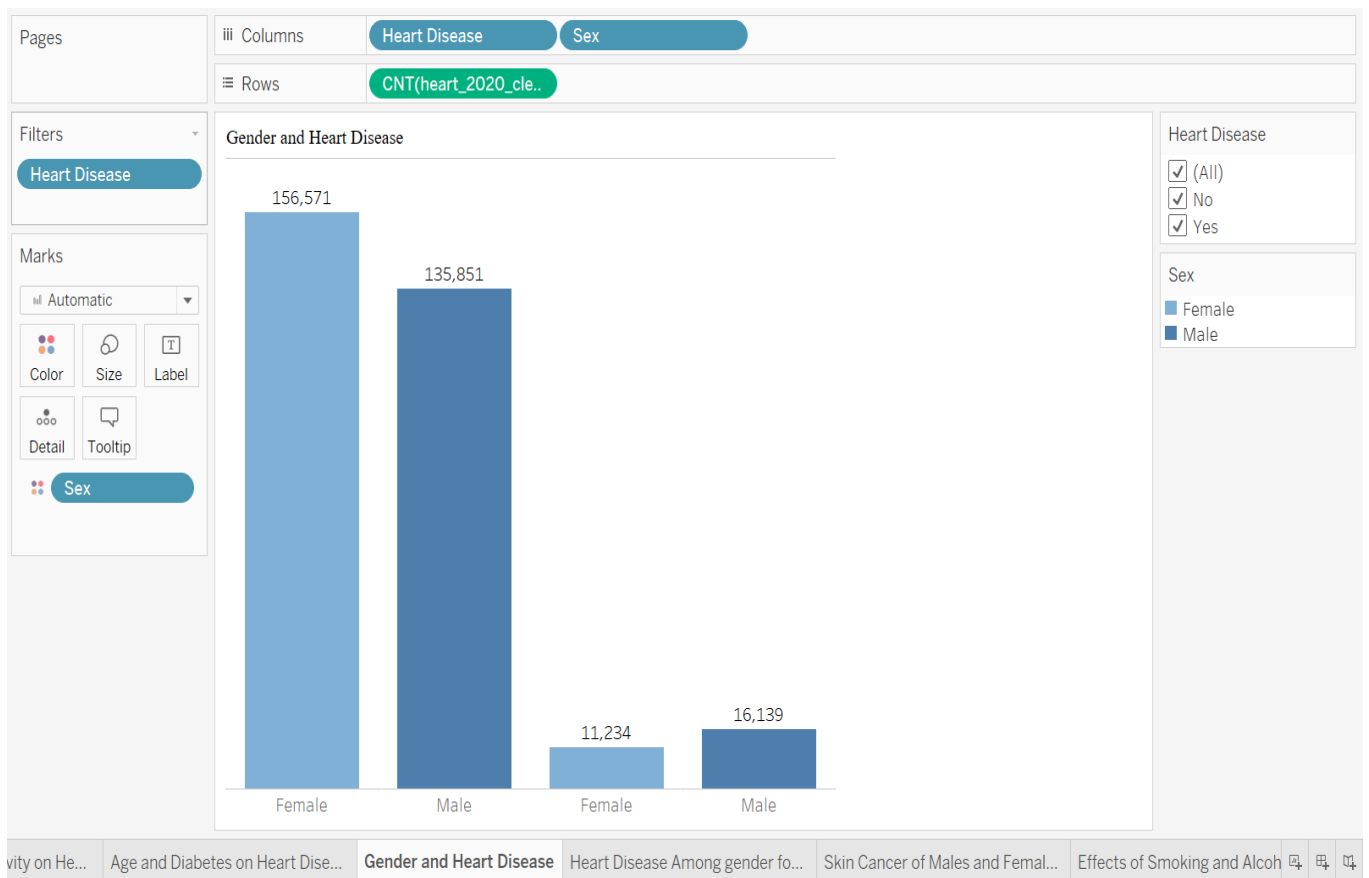
This visualization is created using highlight tables, an interactivity filter named heart disease and diabetes been linked to the visualization. This chart shows the age along with diabetes affects the heart of an individual. The pre-attentive attribute in this chart is highlighted color.



We have added a calculated field for diabetics as it had four different attributes, namely yes. No, no borderline diabetics, yes during pregnancy. The calculated field calculates the value for yes and

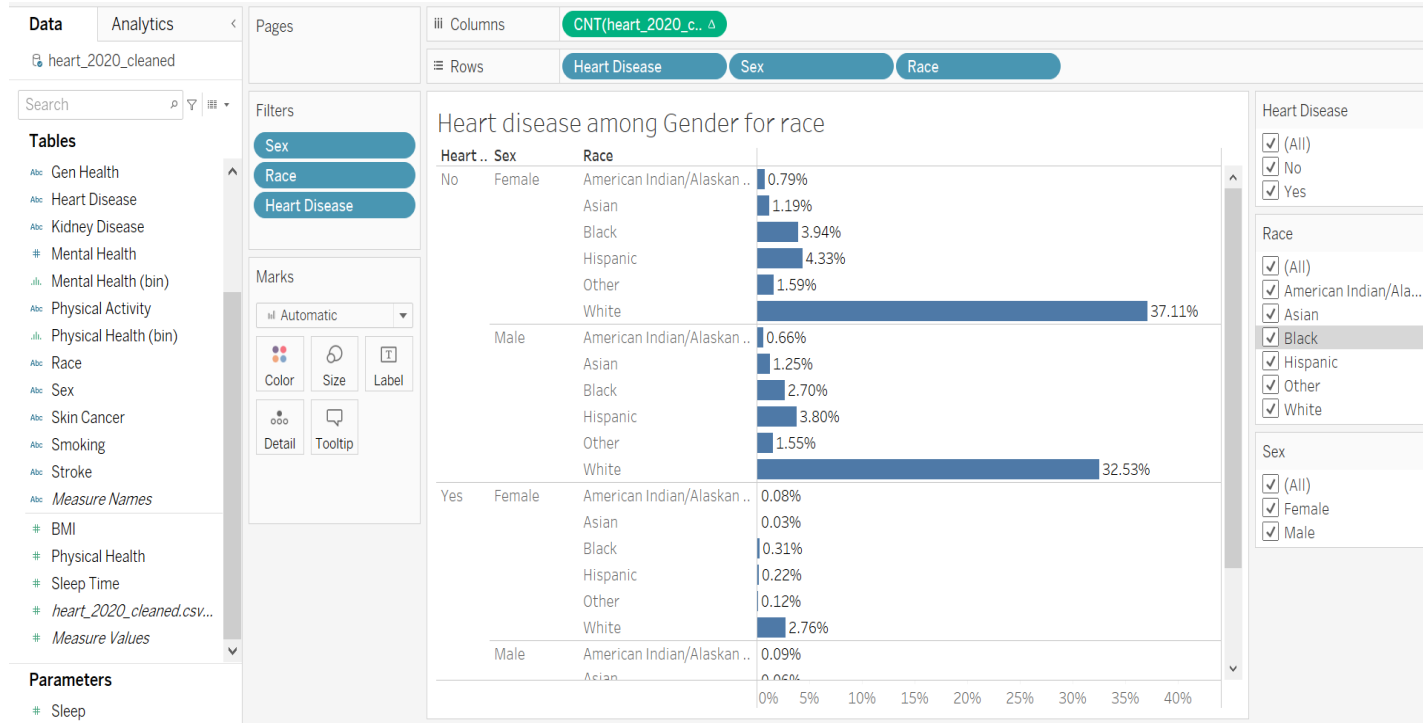
no. It combines yes and yes during pregnancy values and no combines no and no borderline diabetics.

6) Gender and Heart Disease



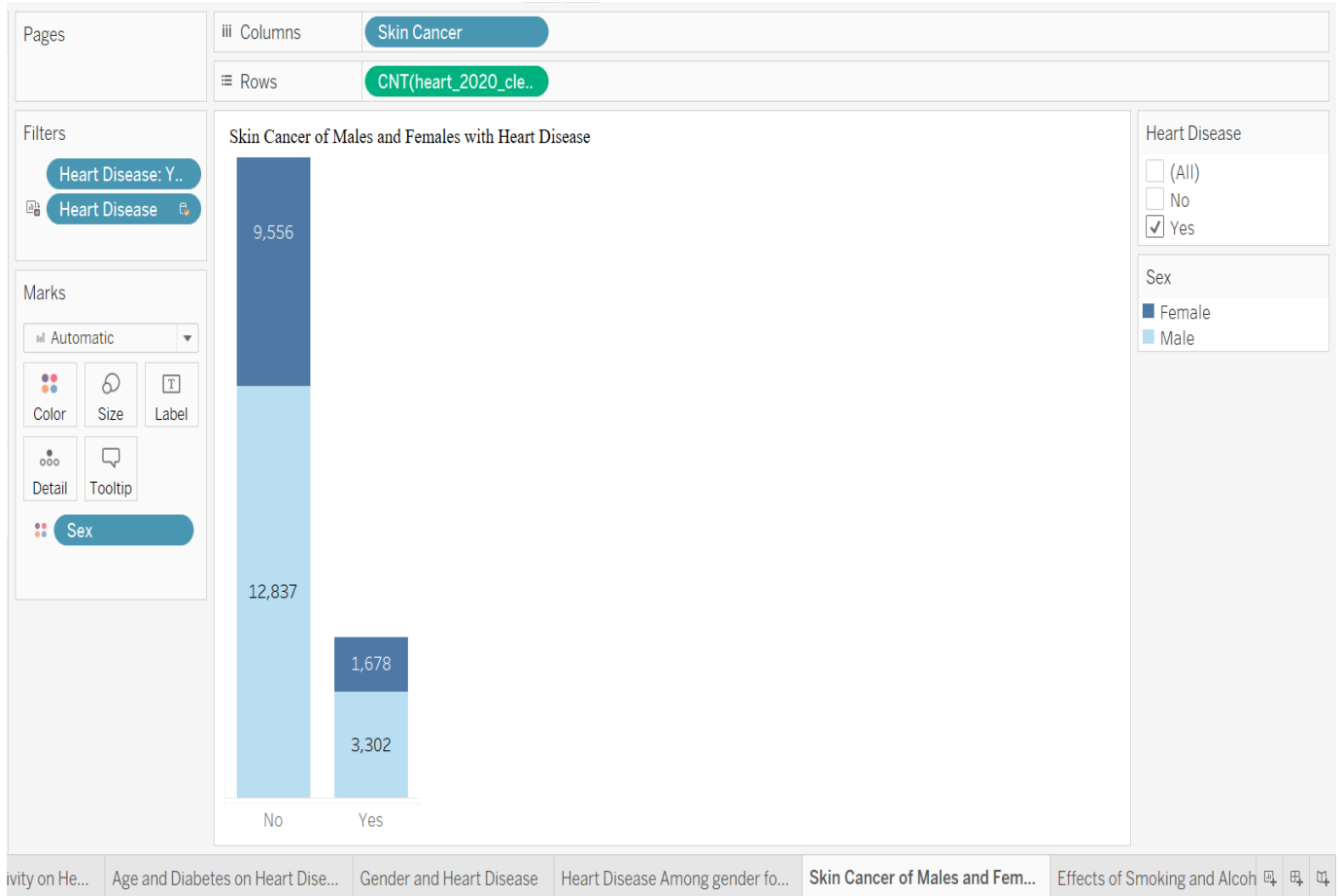
This visualization is created using a bar chart, an interactivity filter named heart disease has been linked to the visualization. It is visible on the chart that the light blue color represents Female, and the blue color represents Male who has a high chance of getting affected by heart disease. The pre-attentive attribute in this chart is length and color.

7) Heart Disease Among gender for Race



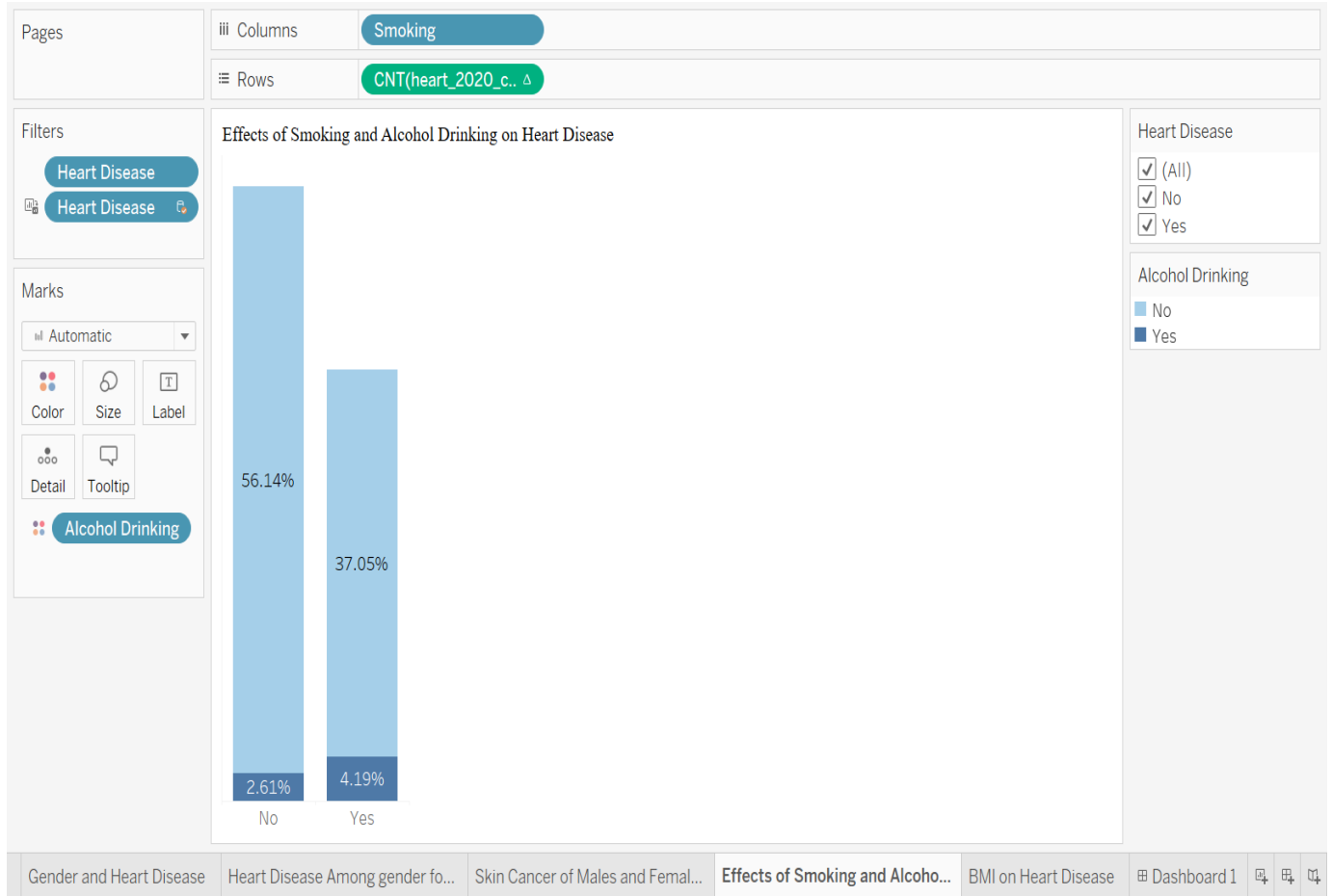
This visualization is created using a bar chart, an interactivity filter named heart disease has been linked to the visualization. It is visible on the right side of the chart that distinct colors represent people from different races which depict their presence in the graph. The pre- attentive attribute here in this graph is the length and the color of the bar graph. This graph acts as an alternative graph for the previous graph which is being shown in a different format.

8) Skin Cancer of Males and Females with Heart Disease



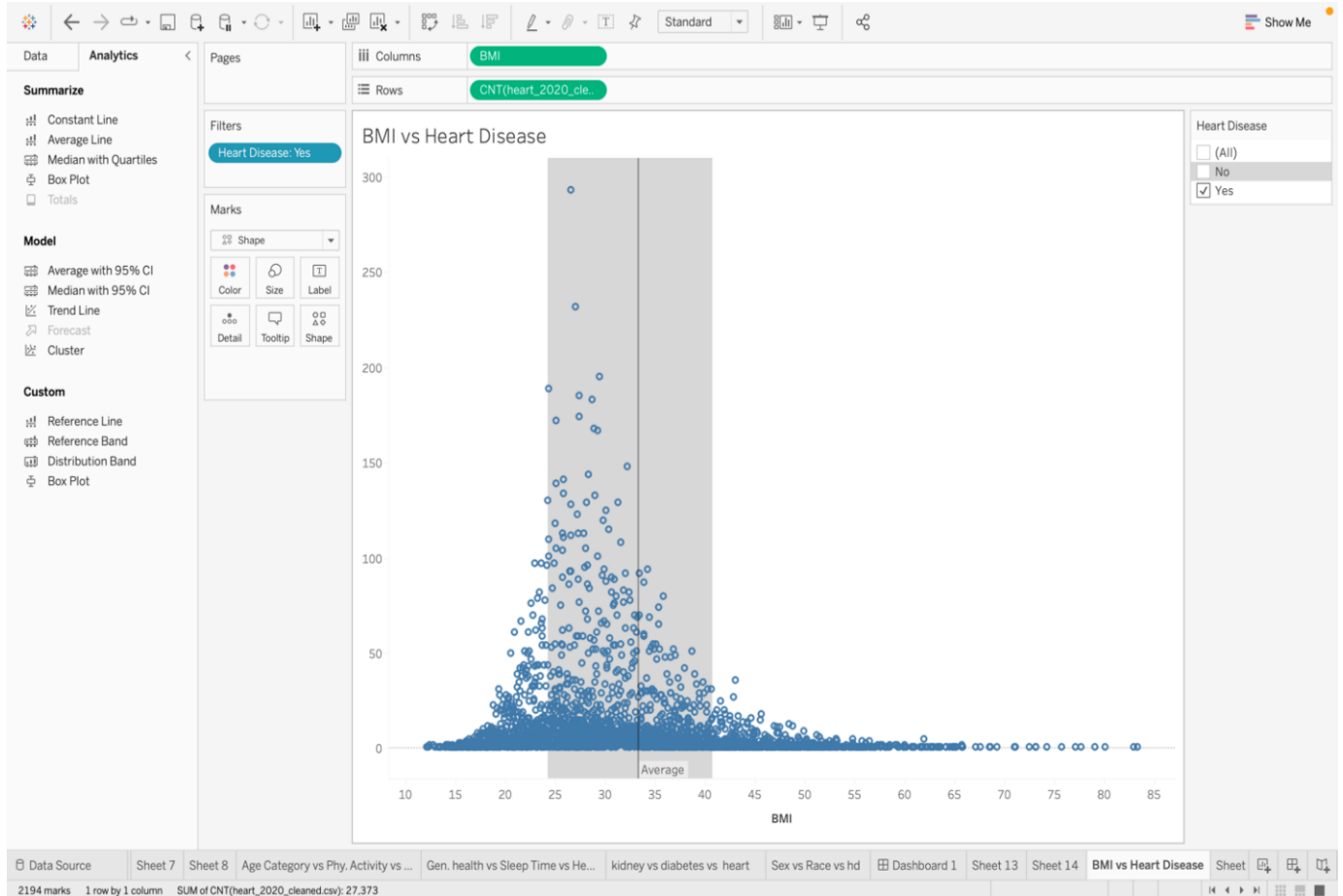
This visualization is created using a bar chart, an interactivity filter named heart disease has been linked to the visualization. It is visible on the chart that the blue color represents Female, and the orange color represents Male who has a high chance of getting affected by heart disease. The pre-attentive attribute in this chart is length and color. This chart represents the individuals who are suffering from skin cancer along with heart disease, with the male and female being represented with the colors.

9) Effects of Smoking and Alcohol Drinking on Heart Disease



This visualization is created using a bar graph, an interactivity filter named heart disease has been linked to the visualization. It is visible on the right side of the chart that distinct colors represent people who smoke and drink alcohol to people who don't with the yes or no being mentioned. The pre-attentive attribute here in this graph is the length and the color of the bars.

10) BMI on Heart Disease



This visualization is created using a scattered chart, an interactivity filter named heart disease has been linked to the visualization. This chart shows the body mass index of an individual and how it affects the heart of an individual, the average of the same is also depicted in this chart. The pre-attentive attribute in this chart is position and color.

Section 6: Dashboard interactivity

We have added interactivities for all the visualizations we have made.

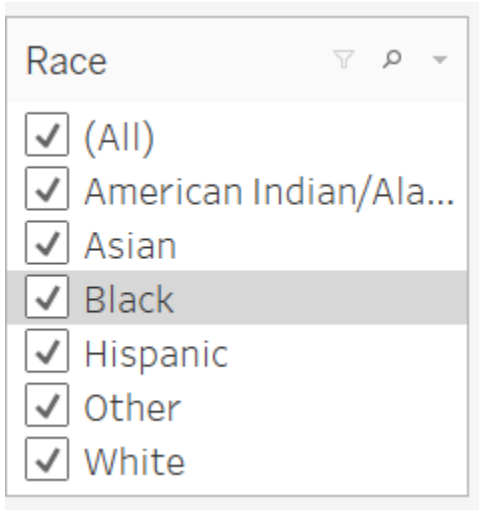
1. We have given an option to choose for displaying the data for having heart disease value as no or yes.



Heart Disease

- ☒ (All)
- ☒ No
- ☒ Yes

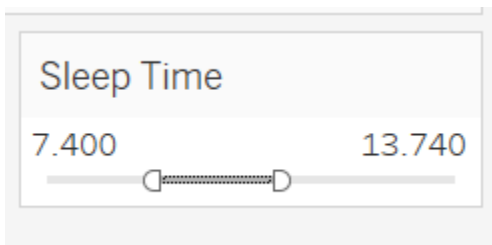
2. We have also given interactivity for choosing among the available age categories values.



Race

- ☒ (All)
- ☒ American Indian/Ala...
- ☒ Asian
- ☒ Black
- ☒ Hispanic
- ☒ Other
- ☒ White

3. We have given the interactivity for choosing the sleep time in hours to see the relation between heart disease and sleep for specific hours.



Sleep Time

7.400 13.740

A horizontal range slider with a dashed line and two circular handles. The left handle is positioned at 7.400 and the right handle is at 13.740.

4. Interactivity for diabetics and heart disease.

Diabetes

☒ (All)
☒ NO
☒ YES

Heart Disease

☐ (All)
☐ No
☒ Yes

5. Interactivity to choose the specific race and gender is also made available.

Race

☒ (All)
☒ American Indian/Ala...
☒ Asian
☒ Black
☒ Hispanic
☒ Other
☒ White

Sex

☒ (All)
☒ Female
☒ Male

We have also added the calculated field as below:

Diabetes

×

```
if [Diabetic]='Yes (during pregnancy) '  
OR [Diabetic]='Yes '  
THEN  
'YES '  
ELSE  
'NO '  
END
```

The calculation is valid.

2 Dependencies ▾

Apply

OK

References:

1. Mural link

<https://app.mural.co/t/ganashreeannaiah1476/m/ganashreeannaiah1476/1649192100047/e984f8d0c8b88e28c0992e6135bec14e6e07c587?sender=u830f511c5b1894991f8d3803>

2. Selected Dataset Link

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

3. Dashboard Link:

https://public.tableau.com/app/profile/ganashree.naganahalli.annaiah/viz/Heart_Disease_Project/Dashboard1?publish=yes