Data Visualization Final Project

Module/Part 1:

My data set is from the U.S. Department of Health and Human Services; it can be found here: https://catalog.data.gov/dataset/conditions-contributing-to-deaths-involving-coronavirus-disease-2019-covid-19-by-age-group. The data set was published by the Centers For Disease Control and is titled "Conditions Contributing to COVID-19 Deaths, by State and Age, Provisional 2020-2023." This dataset consists of all the COVID-19 deaths recorded in the United States and the health conditions (comorbidities) contributing to those deaths. The data is further grouped by age group and each state in the U.S. The data was collected between 2020 and 2023, and no new data points are being added to this data set. I want to pursue healthcare data science, and I was interested in looking at a data set related to public health. During the COVID-19 pandemic there was a lot of discussion around COVID comorbidities, which are health conditions that can increase an individual's risk of having a severe COVID infection or increases an individual's risk of dying from COVID. The data set includes various medical conditions mentioned in conjunction with COVID-19 deaths, these include influenza/pneumonia, chronic lower respiratory diseases, adult respiratory distress syndrome, respiratory failure, respiratory arrest, hypertensive diseases, ischemic heart disease, cardiac arrest, cardiac arrhythmia, heart failure, cerebrovascular diseases, sepsis, diabetes, obesity, Alzheimer's disease, dementia, and renal failure. The attributes included in the data set are the U.S. states, the condition group, condition, ICD-10 code of the condition, age group, and the number of COVID-19 deaths associated with that state and age group; there are also totals for each age group and for the entire U.S. I will look at the total COVID-19 deaths in the United States, broken down by age group for 10 of the deadliest health conditions. The conditions I will look at are hypertensive diseases, diabetes, renal failure, ischemic heart disease, sepsis, chronic lower respiratory diseases, heart failure, cerebrovascular diseases, obesity, and Alzheimer's disease. The age groups are broken down as follows: 0-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, and 85+.

The key questions I would want to answer include: What are the total COVID-19 deaths associated with each health condition? Which health conditions contributed to the highest number of COVID-19 deaths? Which health conditions most or least impacted each age group and are there differences in how each age group was affected? After doing some research on previous visualizations for COVID-19 death data I noticed that most graphs were either a variation of a bar chart or a scatter plot. Multiple bar graphs were often used to depict COVID-19 deaths for different age groups, but these usually only consisted of 2-3 different age groups. Having 8 bars per category (health condition) may be too many bars to efficiently look at the entire visual and may overwhelm the user. Many visuals also used random or rainbow colors or a few mutated colors. Since I have 8 categories for each age group, I would like to have 8 colors that are easy to distinguish from each other without implementing all the colors of the rainbow. Using hue for distinction would be appropriate since I'm using it for categorical encoding. One option would be to use a gradient for the different age groups but trying this the 8 colors were not distinctive enough from each other

to be effective in seeing each age group. I also saw several line graphs, but it was difficult to distinguish the colors of each line since lines can be very thin. Several of the bar graphs I saw had very narrow or short bars which made it difficult to compare each of the bars to each other.

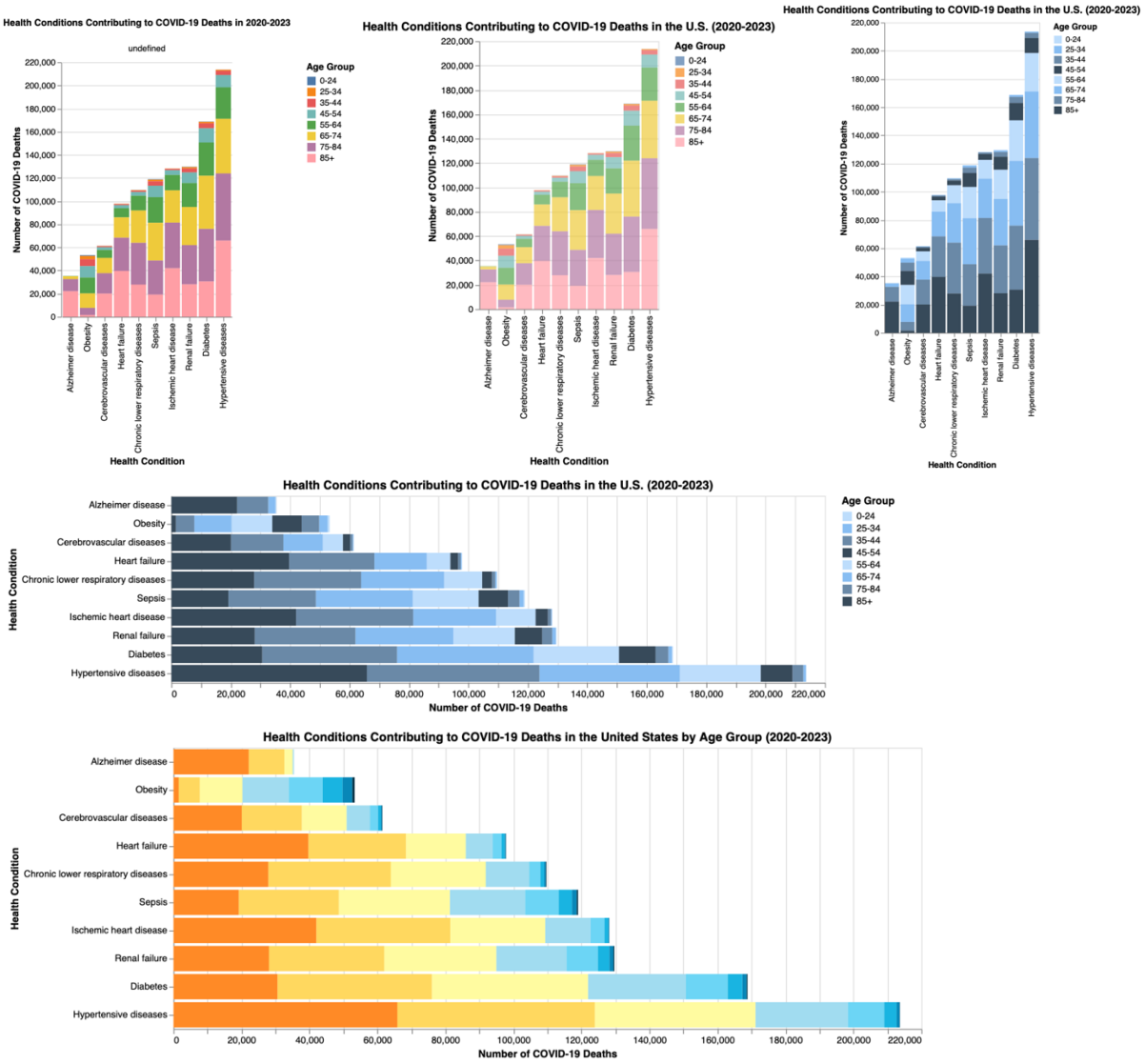
Module/Part 2:

In module 2 I identified two tasks I would like to complete with my visual.

For the first task, the goal is: present/communicate information, the means is: navigate through the data, the characteristics are: identify patterns across each bar (how is the distribution of covid deaths by age group different for each health condition), the target data is: a relative reference frame, the workflow: user will look at the entire graph, distinguish each age group based on color using the legend, then compare the number of deaths there are for each age group within each bar, then compare the bars to each other to see how each condition compares, and the roles: the user has the goal of understanding how different age groups were impacted by different health conditions in contributing to COVID deaths. This gives a better understanding on the most common covid comorbidities for each age group. They also can determine that the majority of covid deaths occurred in people aged 45 and older.

For the second task, the goal: present/communicate information, the means: build connections in the data, characteristics: identify the pattern in the distribution of covid-19 deaths (which conditions lead to the most covid-19 deaths), target data: relative reference frame, workflow: user will look at the entire graph, then use the x-axis to determine the total number of COVID-19 deaths that were associated with each comorbidity, and roles: The user will determine what the different covid comorbidities are, how many deaths were associated with each comorbidity, and determine which comorbidity was the most and least high risk for covid death.

These are some of my low fidelity prototypes that I created with Altair initially (I attached the code for the prototypes and final visual as well at the end):



Module/Part 3:

My preliminary design evaluation, would be an insight-based evaluation, and is as follows:

Target question I want to answer: Can people see the patterns they need to (and correctly) from my visual?

People recruited to answer that question: My family, colleagues/friends, and two acquaintances who are in doctoral programs who frequently review research and their corresponding graphs.

The kind of measures used and what the measures tell me about the core question: Determine if the people I've recruited can complete my two tasks as outlined in my Module 2 reflection. The measure would include insight depth and accuracy. I would want to know what insights my audience finds and if those are accurate with what is depicted in the graph.

The approach to answer that question: I would give each of my participants a survey and ask the following question: What types of insights can you discover or gain in this

visual? Are the colors in the graph easy to distinguish and are they easy to pair with the legend? What are the total deaths counts associated with each health condition (comorbidity)? What types of trends do you notice in the data? Do you have any suggestions for improvement of this visual?

How I instantiate those methods: My participants would complete the survey.

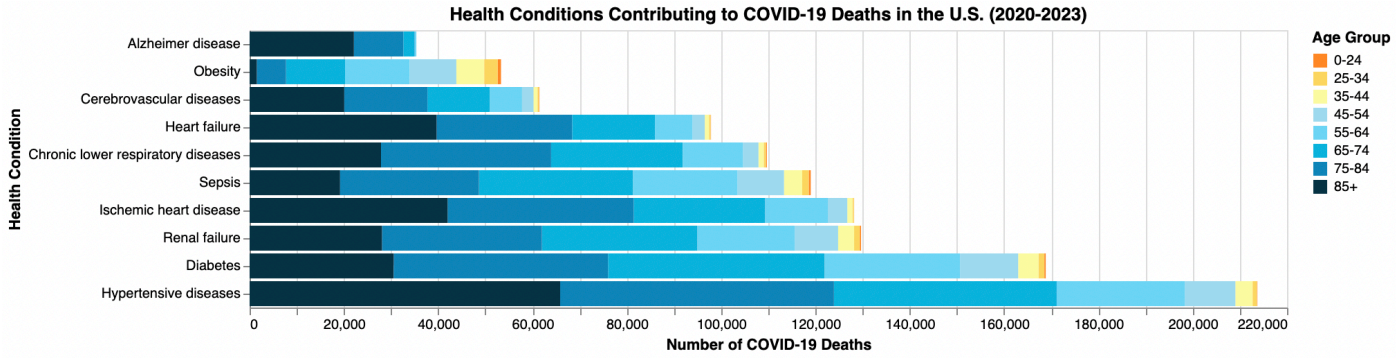
Criteria used to indicate that my visualization is successful: I will review the responses to determine if the goals of my visual were met. This would include an assessment of: Did the insights gained by participants correctly match with the data in the graph? Did participants correctly and easily identify the total number of COVID-19 deaths associated with each health condition? Did participants identify the pattern in COVID-19 deaths across each age group? Were participants able to see the graph clearly with the colors that were chosen? Did participants have any suggestions, and would it make sense to implement those suggestions based on what I've learn in this course? I would know that my visual was successful if the responses reflect that participants were able to easily and accurately identify the patterns and insights I aimed to communicate in my visual.

Module/Part 4

Brief recap of data, goals, and tasks:

The data set used in this visualization was collected by the U.S. Department of Health and Human Services, titled "Conditions Contributing to COVID-19 Deaths, by State and Age, Provisional 2020-2023." My analysis looked at the ten most deadly comorbidities and their corresponding total number of COVID-19 deaths in the United States between 2020 and 2023, these totals were further grouped by age group. The goal of my visual is to communicate information about the total number of COVID-19 deaths associated with each of these health conditions and how the number of deaths vary by age group and health condition. My two tasks included: 1. identify patterns across each bar in the graph (how is the distribution of covid deaths by age group different for each health condition?), and 2. identify the pattern in the distribution of covid-19 deaths (which conditions lead to the most or least number covid-19 deaths?).

Final Visualization Implementation (I also attached a separate, larger file for the image of the visual):



Summary of the key elements of design:

I believe that the horizontal stacked bar graph was the best way to visualize the COVID-19 deaths for each health condition broken down by age group. This kept the graph as simple and easy to read as possible while still displaying all the data. I increased the length and width of each bar to make it easier to identify each age group within each bar. I kept the bars close to each other to make them easier to compare side by side. I made the stacked bar chart horizontal to make the names of each condition easy to read as it would've been difficult to read long names if they were vertical. Instead of a multiple bar chart I chose to use a stacked bar graph because a multiple bar graph would have shown a total of 80 bars, which could've been visually overwhelming. I ordered the bars (health conditions) from top to bottom in ascending order of the number of deaths to make it easier to see how the total number of deaths differed between each condition. I would have liked if all the colors used were a gradient of a hue to show the oldest age group as darkest and the youngest age group as lightest. However, given the number of categories (8 total age groups) it was too difficult to distinguish 8 shades of blue for example, especially since the younger age groups were the smallest. Given that each age group was categorical I made each color distinct and choose the colors for each group based on what would show up best for that category based on its size. Since there were a high number of categories (8) I chose colors that were distinct from each other but made several adjacent colors a gradient so that I could more easily be found on the legend. I used darker colors of blue for the older age groups and shades of orange/yellow for the younger age groups. This highlights that most COVID-19 deaths occurred in people aged 45 and older (as distinguished by the shades of blue). In addition, given the context of the topic, I tried to keep the colors engaging without being too vibrant or a rainbow of colors.

Discussion of my final evaluation approach:

My final evaluation consisted of administering my survey to four different people. I shared my survey via a google doc and my participants wrote in their responses. My four participants included two acquaintances who frequently read/review research, one friend, and one family member.

My survey asked the following questions: What types of insights can you discover or gain in this visual? Are the colors in the graph easy to distinguish and are they easy to pair with the legend? What are the total death counts associated with each health condition (comorbidity)? What types of trends do you notice in the data? Do you have any suggestions for improvement of this visual?

My evaluation results: After administering my surveys and receiving the responses I noticed a few things. One suggestion I had was to create a rainbow of colors to make each age group color distinct. Another suggestion was to only have two different age groups that split the data in half into a 0–54-year age group and a 55+ age group so that only two colors would be necessary. Another suggestion was to make a multiple bar graph with less age groups. Another suggestion I had was to make all 8 age group colors a gradient. Many of my suggestions seemed to be about the colors and the division of age groups. I decided

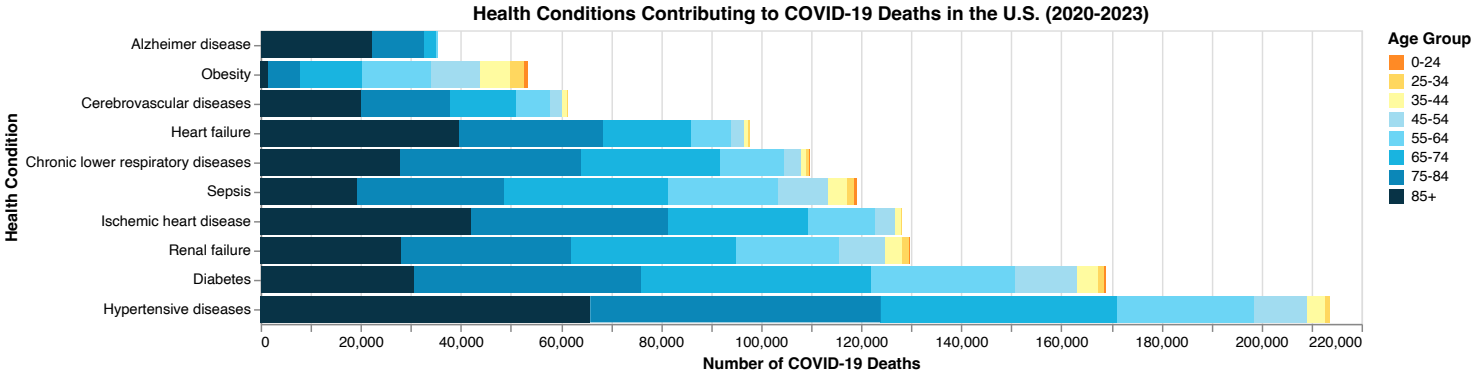
against most of these suggestions because I didn't think rainbow colors were meaningful for the data, I kept all 8 age groups because it highlighted important differences between the age groups, and opted to have a variation of a gradient where possible. All four of the participants were able to determine the total number of COVID-19 deaths per age group, although one participant initially thought that the stacked bars were cumulative. The trends that my participants noticed include: There were a higher portions of COVID-19 deaths for younger age groups associated with obesity and that obesity was the least impacting for the oldest age group (85+). The very youngest age group had the highest number of deaths associated with obesity and the oldest age group had the highest number of deaths associated with hypertensive disease. Overall, Alzheimer disease was associated with the least number of deaths and hypertensive disease was associated with the highest number of deaths. Most deaths occurred in people aged 45 and older.

A synthesis of your findings:

I think that my initial approach was to create low fidelity prototypes in Altair instead of by hand. This allowed me to very quickly try new things, like change the type of graph, color schemes, scales, axes, etc. I went through many changes before arriving at my final visual. I think that I would do my prototyping this way in the future after an initial draft sketch by hand to quickly eliminate the approaches that don't work well.

I think that conducting an insight-based evaluation using a survey was the most appropriate evaluation for this type of visual. I received a lot of feedback this way and was able to understand well what my participants were gaining from my visual. I think my biggest struggle was in choosing color, in the future I would aim to decrease the number of categories to 5-6 or less if it makes sense for the data. This would allow for more selective and intentional color choices. I would have like to have a gradient in my colors, but it would have been difficult for my participants to see every category with the stacked bar chart. I would also consider if using proportion instead of counts is appropriate for the data.

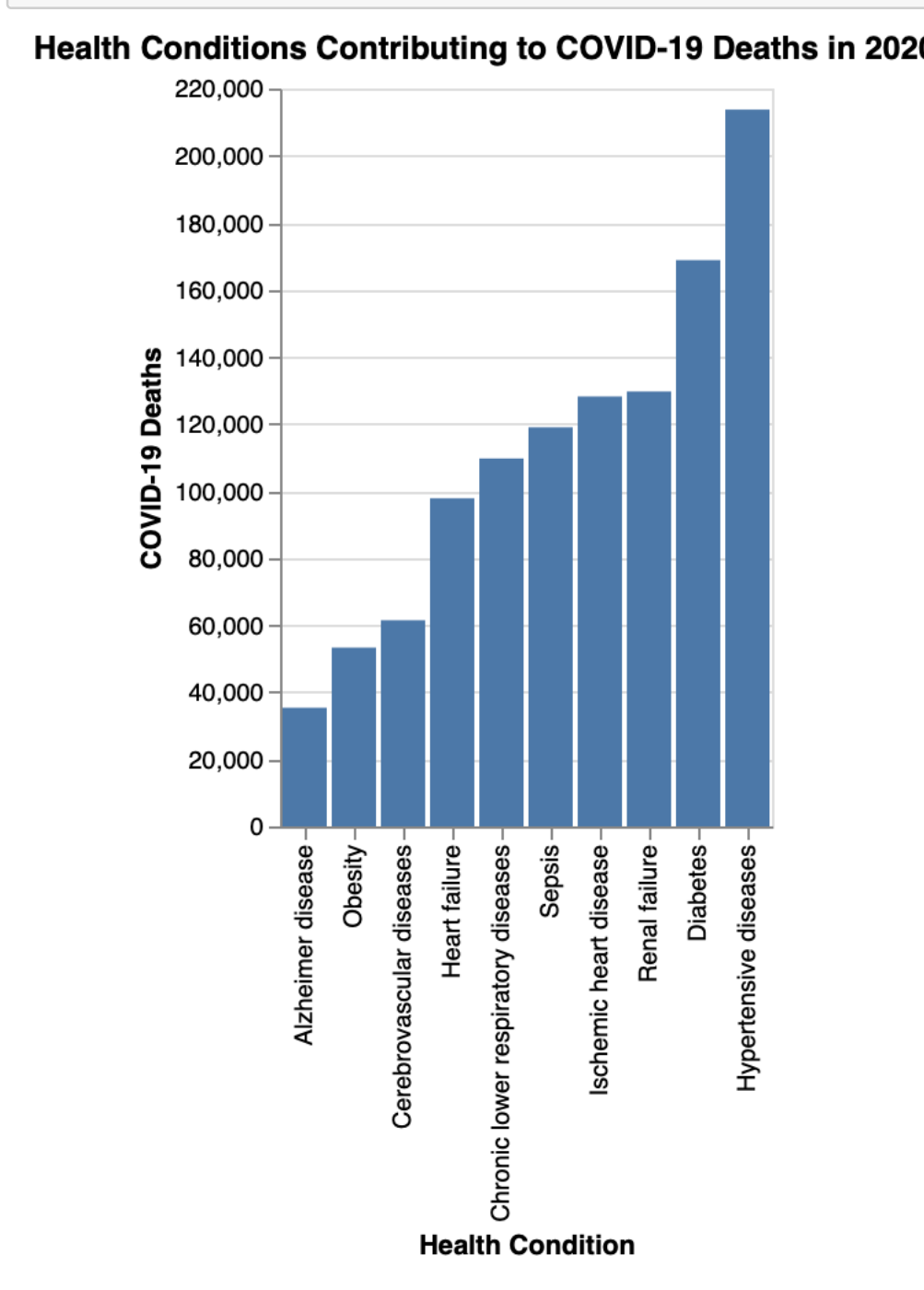
Overall, going through the steps of outlining my goals and tasks was very helpful, this kept my intentions for the visual aligned with its purpose. In the future I would implement the five design sheets process for a more complex visual and would hope to implement more interactive features like reconfiguration to show the age groups in a different order for example for easier comparison, or create an interaction that allowed users to see the proportion of deaths for each condition and age group.


Health Conditions Contributing to COVID-19 Deaths in the U.S. (2020-2023)
Health Condition
Alzheimer disease
Obesity
Cerebrovascular diseases
Heart failure
Chronic lower respiratory diseases
Sepsis
Ischemic heart disease
Renal failure
Diabetes
Hypertensive diseases
0
20,000
40,000
60,000
80,000
100,000
120,000
140,000
160,000
180,000
200,000
220,000
Number of COVID-19 Deaths
Age Group
0-24
25-34
35-44
45-54
55-64
65-74
75-84
85+

In [1]:
```python
import altair as alt
import pandas as pd
```

In [2]:
```python
data = pd.DataFrame({
    'Health Condition': ['Hypertensive diseases', 'Diabetes', 'Renal failure', 'Ischemic heart disease', 'Sepsis', 'Chronic lower respiratory diseases', 'Heart failure', 'Cerebrovascular diseases', 'Obesity', 'Alzheimer disease'],
    'COVID-19 Deaths': [213709, 168830, 129674, 128178, 119000, 109697, 97821, 61455, 53321, 35377]
})

alt.Chart(data).mark_bar().encode(
    x=alt.X('Health Condition', sort=['Alzheimer disease', 'Obesity', 'Cerebrovascular diseases', 'Heart failure', 'Chronic lower respiratory diseases', 'Sepsis', 'Ischemic heart disease', 'Renal failure', 'Diabetes', 'Hypertensive diseases']),
    y='COVID-19 Deaths'
).properties(title="Health Conditions Contributing to COVID-19 Deaths in 2020-2023")
```

Out[2]:



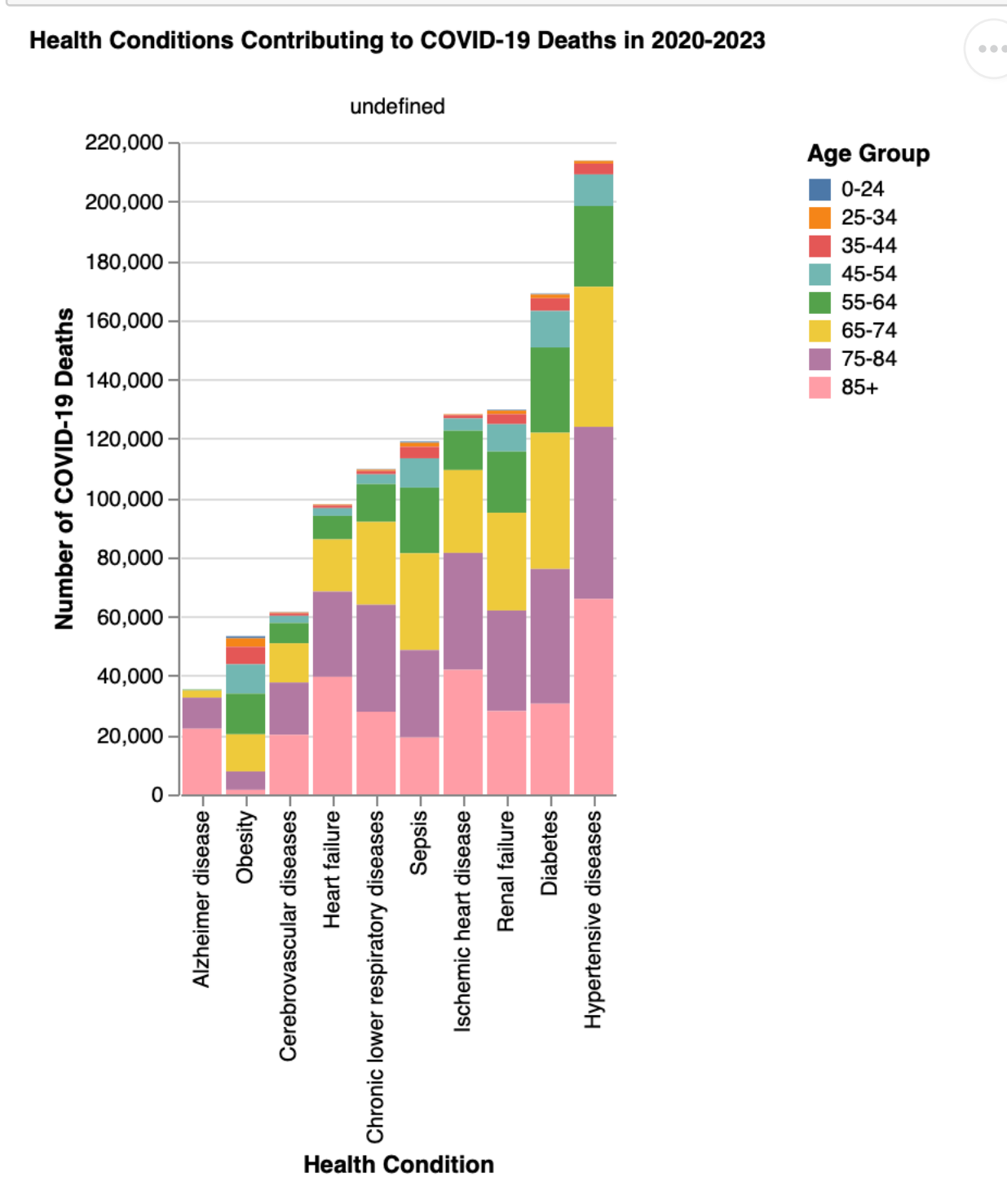
In [ ]:

In [14]:
```python
coviddata = pd.DataFrame({
    'Health Condition': ['Alzheimer disease', 'Alzheimer disease', 'Alzheimer disease', 'Alzheimer disease', 'Alzheimer disease','Alzheimer disease','Alzheimer disease','Alzheimer disease','Obesity', 'Obesity', 'Obesity', 'Obesity','Obesity','Obesity','Obesity','Obesity','Cerebrovascular diseases', 'Cerebrovascular diseases', 'Cerebrovascular diseases', 'Cerebrovascular diseases','Cerebrovascular diseases','Cerebrovascular diseases','Cerebrovascular diseases','Cerebrovascular diseases','Heart failure', 'Heart failure', 'Heart failure', 'Heart failure', 'Heart failure','Heart failure','Heart failure','Heart failure','Chronic lower respiratory diseases', 'Chronic lower respiratory diseases', 'Chronic lower respiratory diseases','Chronic lower respiratory diseases','Chronic lower respiratory diseases','Chronic lower respiratory diseases','Chronic lower respiratory diseases','Chronic lower respiratory diseases','Sepsis', 'Sepsis', 'Sepsis', 'Sepsis','Sepsis','Sepsis','Sepsis','Sepsis','Ischemic heart disease', 'Ischemic heart disease', 'Ischemic heart disease', 'Ischemic heart disease','Ischemic heart disease','Ischemic heart disease','Ischemic heart disease','Ischemic heart disease','Renal failure', 'Renal failure', 'Renal failure','Renal failure','Renal failure','Renal failure','Renal failure','Renal failure','Diabetes', 'Diabetes','Diabetes','Diabetes','Diabetes','Diabetes','Diabetes','Diabetes', 'Hypertensive diseases', 'Hypertensive diseases', 'Hypertensive diseases', 'Hypertensive diseases', 'Hypertensive diseases', 'Hypertensive diseases', 'Hypertensive diseases', 'Hypertensive diseases'],
    'Number of COVID-19 Deaths': [0, 2, 3, 36, 321, 2425, 10476, 22114, 707, 2865, 5910, 9878, 13715, 12516, 6208, 1520, 115, 294, 850, 2425, 6846, 13219, 17636, 20069, 79, 315, 904, 2621,7940, 17580, 28739, 39642,207, 459, 1118, 3286 , 12772, 27920, 36068, 27866, 425,1389, 3867, 9886,22137, 32651, 29455, 19188, 33,289, 1106, 4131,13281, 27901, 39452, 41979, 273,1204,3392,9191,20667, 32952, 33928, 28064, 289,1252,4295,12297,28768, 45930, 45442, 30553, 105,903,3656,10676,27207, 47240, 58062, 65854],
    'Age Group': ['0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+', '0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+','0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+','0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+','0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+','0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+','0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+','0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+','0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+','0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+']
})
```

In [ ]:

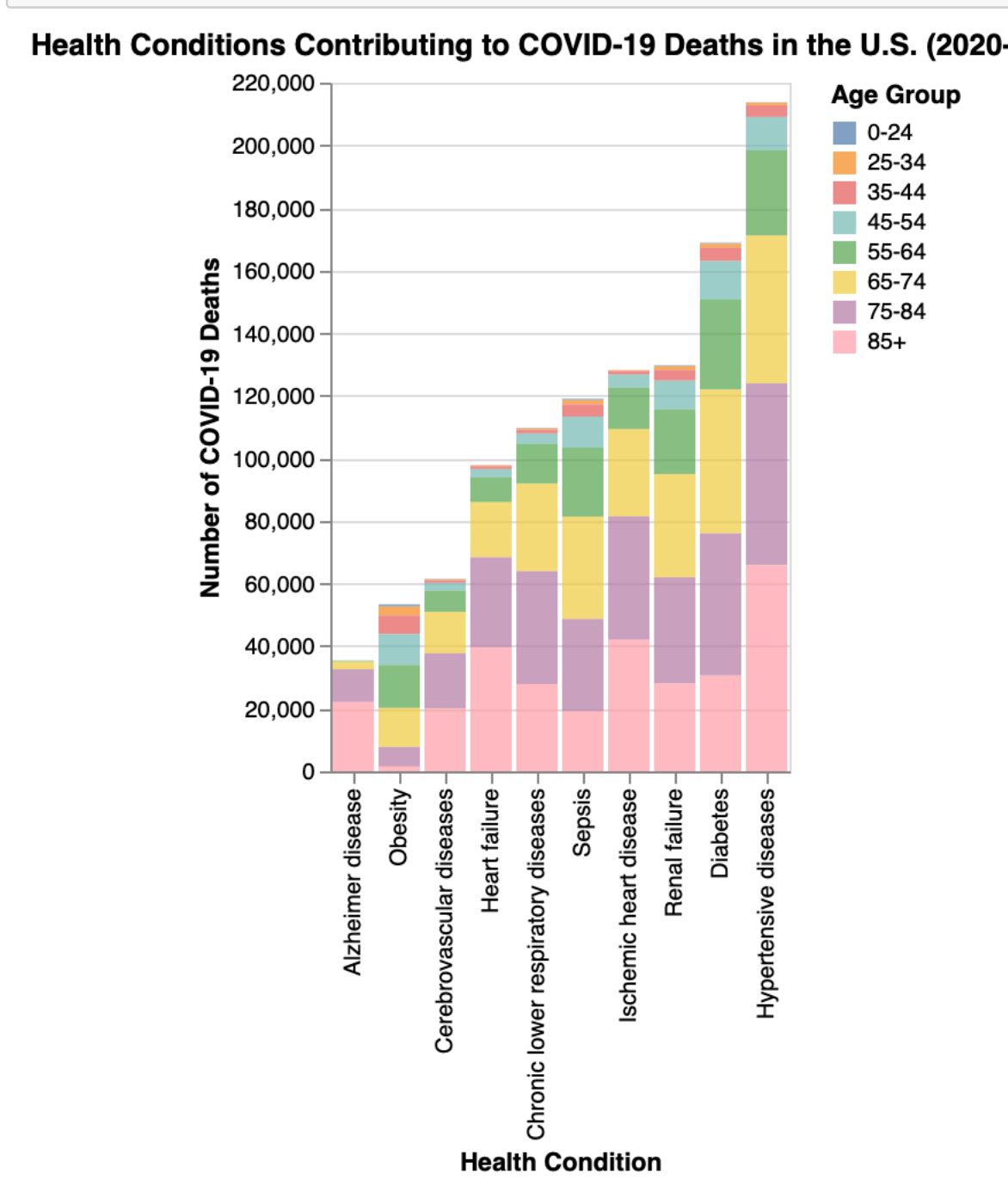
In [15]:
```python
import numpy as np

q13a = pd.DataFrame({
    'Health Condition': ['Alzheimer disease', 'Alzheimer disease', 'Alzheimer disease', 'Alzheimer disease', 'Obesity', 'Obesity', 'Obesity', 'Obesity','Cerebrovascular diseases', 'Cerebrovascular diseases', 'Cerebrovascular diseases', 'Cerebrovascular diseases','Heart failure', 'Heart failure', 'Heart failure', 'Heart failure', 'Chronic lower respiratory diseases', 'Chronic lower respiratory diseases', 'Chronic lower respiratory diseases','Chronic lower respiratory diseases','Sepsis', 'Sepsis', 'Sepsis', 'Sepsis','Ischemic heart disease', 'Ischemic heart disease', 'Ischemic heart disease', 'Ischemic heart disease','Renal failure', 'Renal failure', 'Renal failure','Renal failure','Diabetes', 'Diabetes','Diabetes','Diabetes', 'Hypertensive diseases', 'Hypertensive diseases', 'Hypertensive diseases', 'Hypertensive diseases'],
    'COVID-19 Deaths': [321, 2425, 10476, 22114, 13715, 12516, 6208, 1520, 6846, 13219, 17636, 20069, 7940, 17580, 28739, 39642, 12772, 27920, 36068, 27866, 22137, 32651, 29455, 19188, 13281, 27901, 39452, 41979, 20667, 32952, 33928, 28064, 28768, 45930, 45442, 30553, 27207, 47240, 58062, 65854],
    'Age Group': ['55-64', '65-74', '75-84', '85+', '55-64', '65-74', '75-84', '85+', '55-64', '65-74', '75-84', '85+', '55-64', '65-74', '75-84', '85+', '55-64', '65-74', '75-84', '85+', '55-64', '65-74', '75-84', '85+', '55-64', '65-74', '75-84', '85+', '55-64', '65-74', '75-84', '85+', '55-64', '65-74', '75-84', '85+', '55-64', '65-74', '75-84', '85+']
})


alt.Chart(coviddata).mark_bar().encode(
    x=alt.X('Health Condition', sort=['Alzheimer disease', 'Obesity', 'Cerebrovascular diseases', 'Heart failure', 'Chronic lower respiratory diseases', 'Sepsis', 'Ischemic heart disease', 'Renal failure', 'Diabetes', 'Hypertensive diseases']),
    y='Number of COVID-19 Deaths',
    color='Age Group',
    tooltip=['primary_type:N',
             'year:O',
             alt.Tooltip('sum(Number_of_Incidents):Q',
                         title='Number of incidents')
            ]
).facet(
    column=alt.Column('year:O',
                      header=alt.Header(title="Health Conditions Contributing to COVID-19 Deaths in 2020-2023")),
).resolve_scale(
    x='independent'
).configure_view(
    stroke='transparent'
)
```

Out[15]:



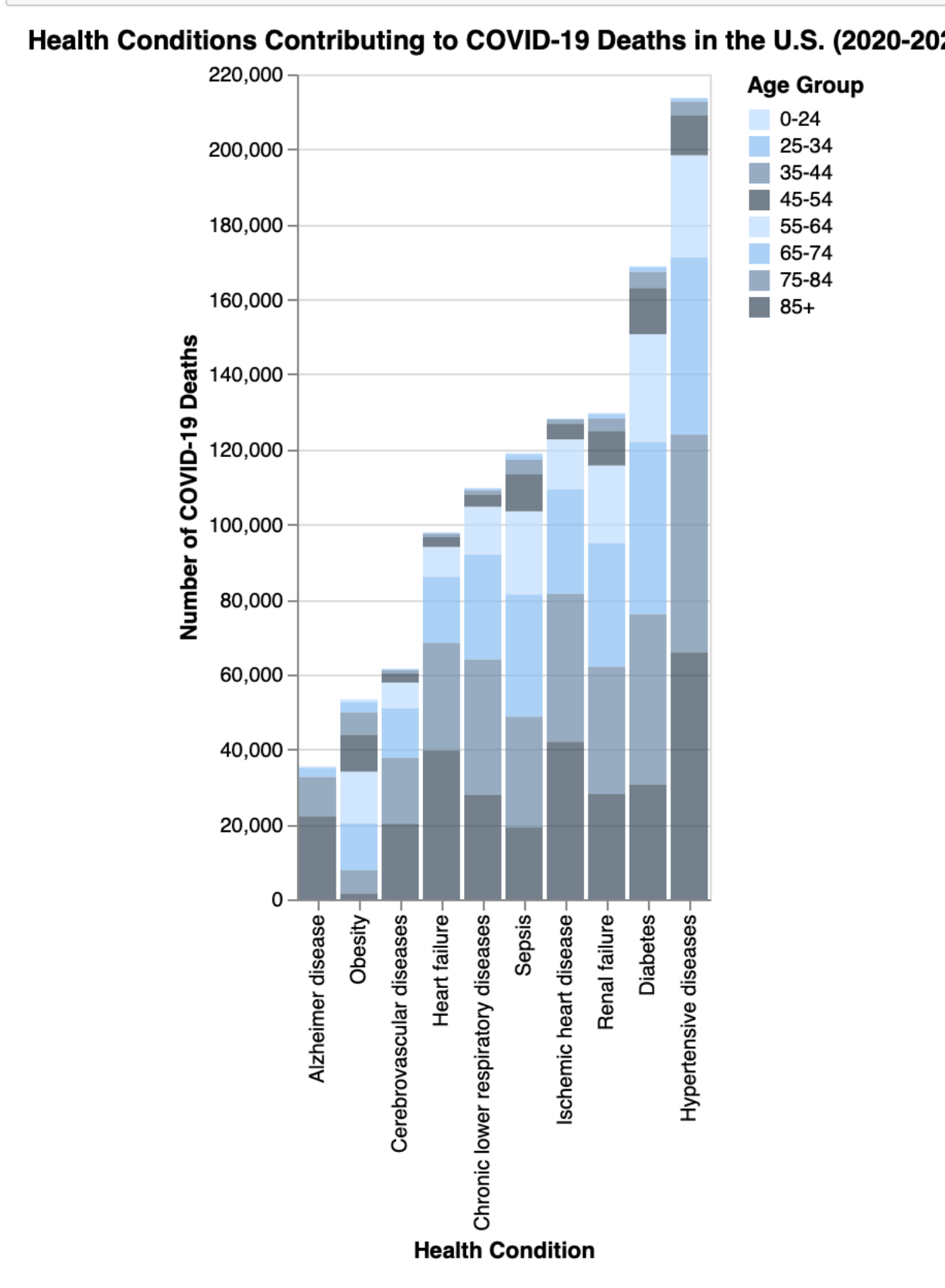
In [7]:
```python
alt.Chart(coviddata).mark_bar(opacity=0.7).encode(
    x=alt.X('Health Condition', sort=['Alzheimer disease', 'Obesity', 'Cerebrovascular diseases', 'Heart failure', 'Chronic lower respiratory diseases', 'Sepsis', 'Ischemic heart disease', 'Renal failure', 'Diabetes', 'Hypertensive diseases']),
    y='Number of COVID-19 Deaths',
    color='Age Group'
).properties(title="Health Conditions Contributing to COVID-19 Deaths in the U.S. (2020-2023)")
```

Out[7]:



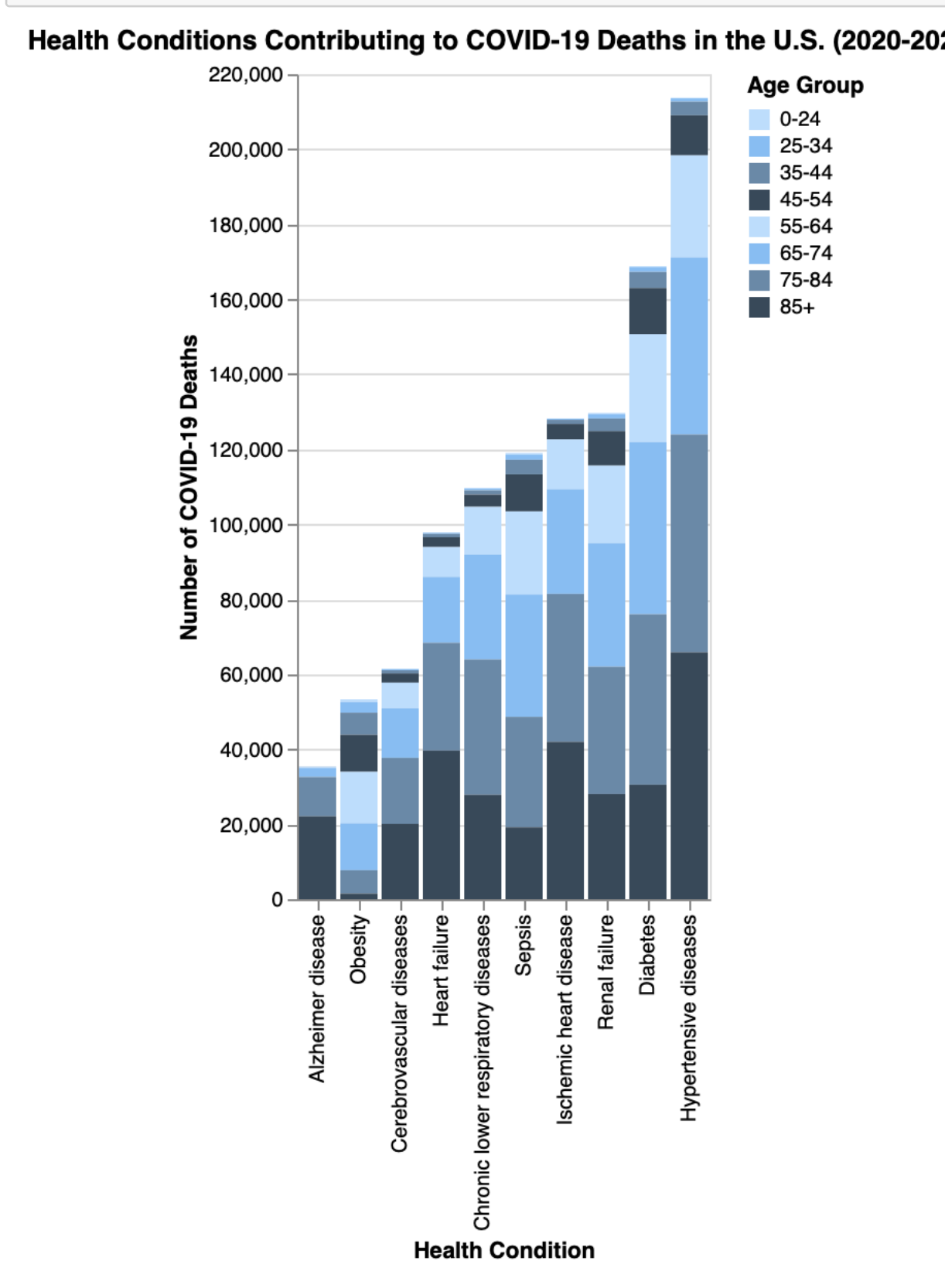
In [16]:
```python
alt.Chart(coviddata).mark_bar(opacity=0.7).encode(
    x=alt.X('Health Condition', sort=['Alzheimer disease', 'Obesity', 'Cerebrovascular diseases', 'Heart failure', 'Chronic lower respiratory diseases', 'Sepsis', 'Ischemic heart disease', 'Renal failure', 'Diabetes', 'Hypertensive diseases']),
    y='Number of COVID-19 Deaths',
    color = alt.Color('Age Group',
                      scale=alt.Scale(domain=['0-24','25-34','35-44','45-54','55-64', '65-74', '75-84', '85+'],
               range=['#bdddfc','#88bdf2','#6a89a7','#384959']))
).properties(height=400,title="Health Conditions Contributing to COVID-19 Deaths in the U.S. (2020-2023)")
```

Out[16]:



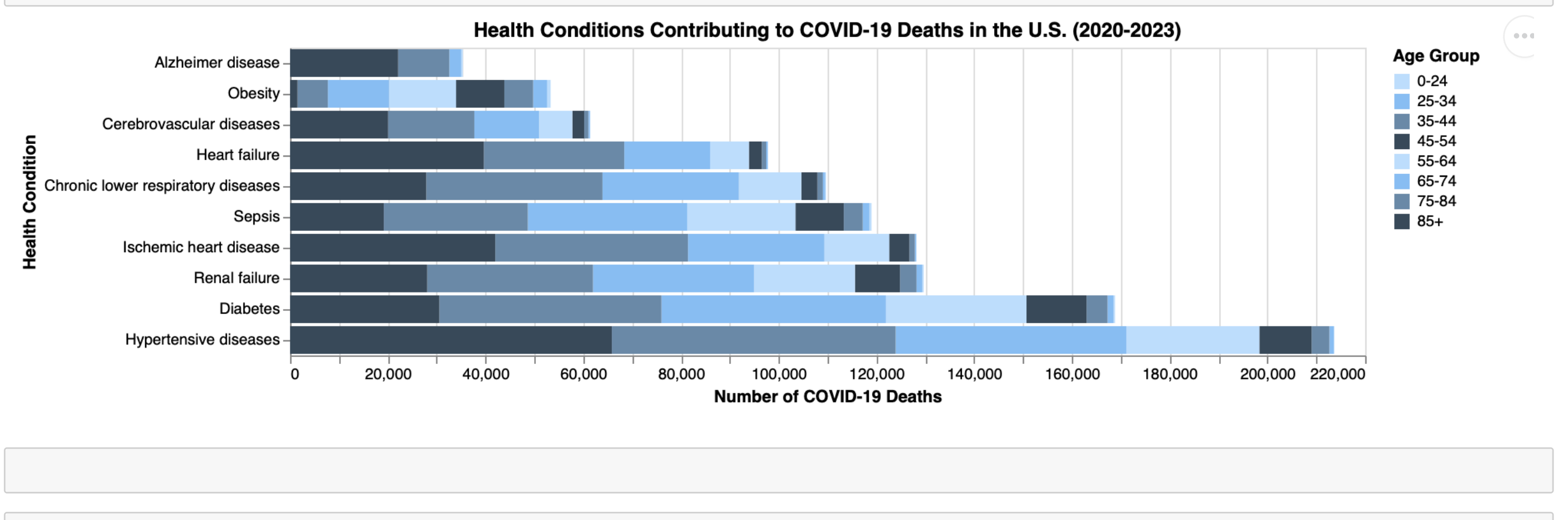
In [17]:
```python
alt.Chart(coviddata).mark_bar().encode(
    x=alt.X('Health Condition', sort=['Alzheimer disease', 'Obesity', 'Cerebrovascular diseases', 'Heart failure', 'Chronic lower respiratory diseases', 'Sepsis', 'Ischemic heart disease', 'Renal failure', 'Diabetes', 'Hypertensive diseases']),
    y='Number of COVID-19 Deaths',
    color= alt.Color('Age Group',
                     scale=alt.Scale(domain=['0-24','25-34','35-44','45-54','55-64', '65-74', '75-84', '85+'],
               range=['#bdddfc','#88bdf2','#6a89a7','#384959']))
).properties(height=400,title="Health Conditions Contributing to COVID-19 Deaths in the U.S. (2020-2023)")
```

Out[17]:



In [18]:
```python
alt.Chart(coviddata).mark_bar().encode(
    y=alt.Y('Health Condition', sort=['Alzheimer disease', 'Obesity', 'Cerebrovascular diseases', 'Heart failure', 'Chronic lower respiratory diseases', 'Sepsis', 'Ischemic heart disease', 'Renal failure', 'Diabetes', 'Hypertensive diseases']),
    x='Number of COVID-19 Deaths',
    color= alt.Color('Age Group',
                     scale=alt.Scale(domain=['0-24','25-34','35-44','45-54','55-64', '65-74', '75-84', '85+'],
               range=['#bdddfc','#88bdf2','#6a89a7','#384959']))
).properties(height=200,width=700,title="Health Conditions Contributing to COVID-19 Deaths in the U.S. (2020-2023)")
```

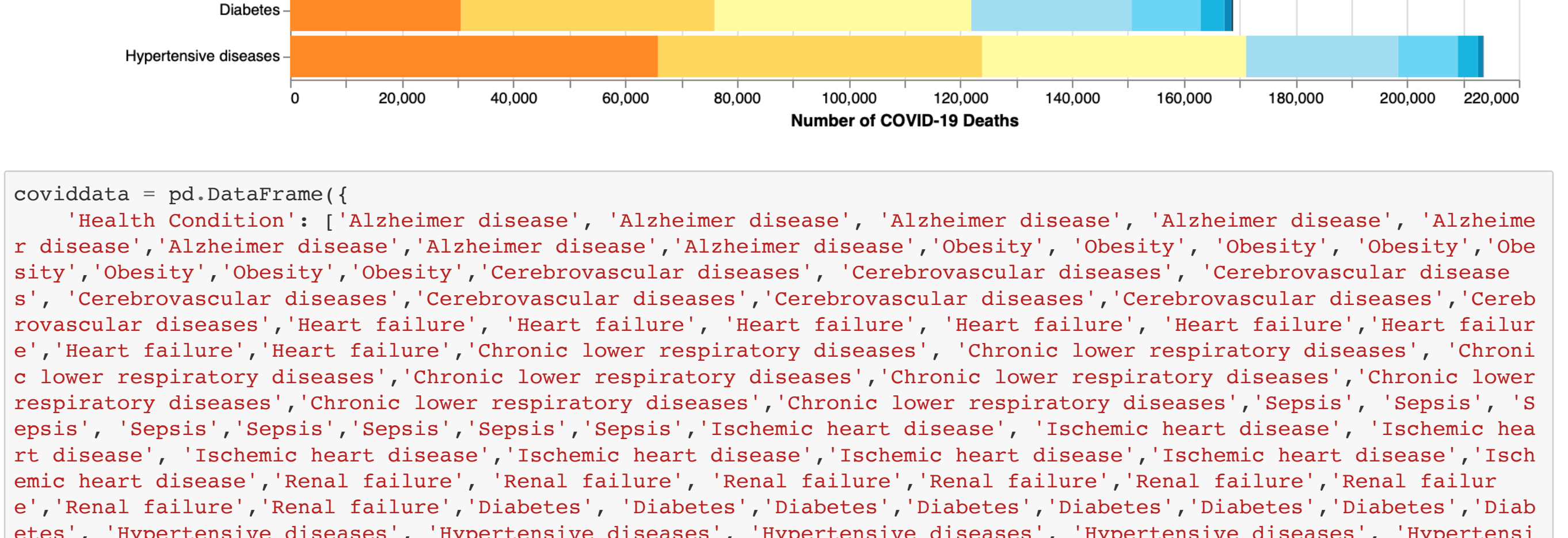
Out[18]:



In [ ]:

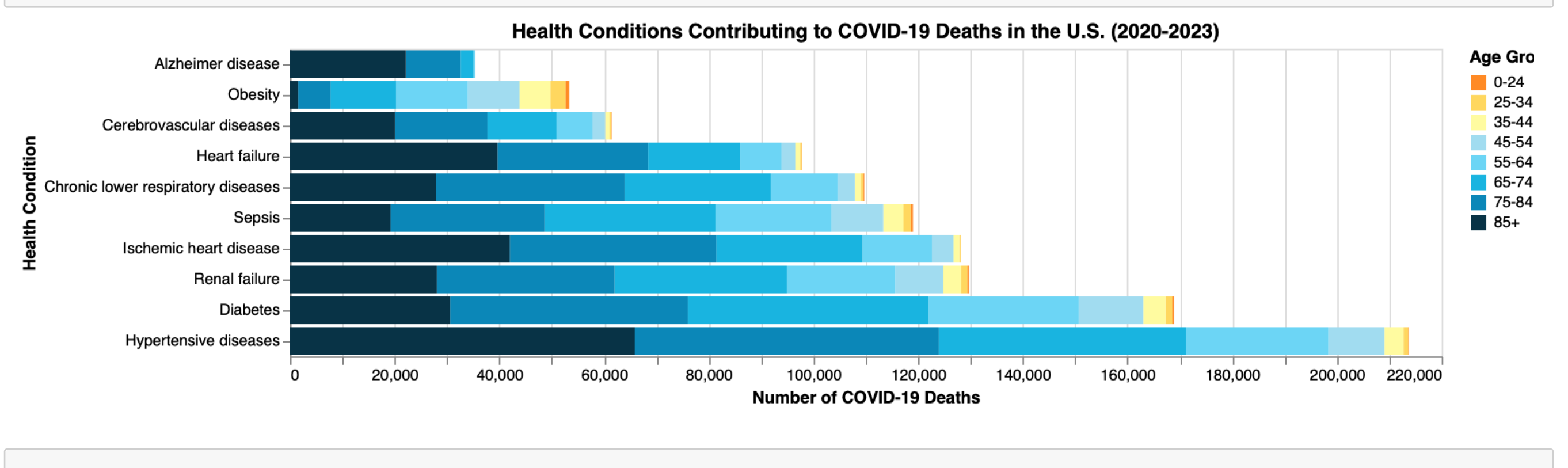
In [13]:
```python
alt.Chart(coviddata).mark_bar().encode(
    y=alt.Y('Health Condition', sort=['Alzheimer disease', 'Obesity', 'Cerebrovascular diseases', 'Heart failure', 'Chronic lower respiratory diseases', 'Sepsis', 'Ischemic heart disease', 'Renal failure', 'Diabetes', 'Hypertensive diseases']),
    x='Number of COVID-19 Deaths',
    color= alt.Color('Age Group',
                     scale=alt.Scale(domain=['0-24','25-34','35-44','45-54','55-64', '65-74', '75-84', '85+'],
               range=['#083346','#0b87b8','#19b4e0', '#6cd5f5','#a1dcf0','#fffba0','#FFD75F','#FF8A25']))
).properties(height=300,width=800,title="Health Conditions Contributing to COVID-19 Deaths in the United States by Age Group (2020-2023)")
```

Out[13]:



In [20]:
```python
coviddata = pd.DataFrame({
    'Health Condition': ['Alzheimer disease', 'Alzheimer disease', 'Alzheimer disease', 'Alzheimer disease', 'Alzheimer disease','Alzheimer disease','Alzheimer disease','Alzheimer disease','Obesity', 'Obesity', 'Obesity', 'Obesity','Obesity','Obesity','Obesity','Obesity','Cerebrovascular diseases', 'Cerebrovascular diseases', 'Cerebrovascular diseases', 'Cerebrovascular diseases','Cerebrovascular diseases','Cerebrovascular diseases','Cerebrovascular diseases','Cerebrovascular diseases','Heart failure', 'Heart failure', 'Heart failure', 'Heart failure', 'Heart failure','Heart failure','Heart failure','Heart failure','Chronic lower respiratory diseases', 'Chronic lower respiratory diseases', 'Chronic lower respiratory diseases','Chronic lower respiratory diseases','Chronic lower respiratory diseases','Chronic lower respiratory diseases','Chronic lower respiratory diseases','Chronic lower respiratory diseases','Sepsis', 'Sepsis', 'Sepsis', 'Sepsis','Sepsis','Sepsis','Sepsis','Sepsis','Ischemic heart disease', 'Ischemic heart disease', 'Ischemic heart disease', 'Ischemic heart disease','Ischemic heart disease','Ischemic heart disease','Ischemic heart disease','Ischemic heart disease','Renal failure', 'Renal failure', 'Renal failure','Renal failure','Renal failure','Renal failure','Renal failure','Renal failure','Diabetes', 'Diabetes','Diabetes','Diabetes','Diabetes','Diabetes','Diabetes','Diabetes', 'Hypertensive diseases', 'Hypertensive diseases', 'Hypertensive diseases', 'Hypertensive diseases', 'Hypertensive diseases', 'Hypertensive diseases', 'Hypertensive diseases', 'Hypertensive diseases'],
    'Number of COVID-19 Deaths': [0, 2, 3, 36, 321, 2425, 10476, 22114, 707, 2865, 5910, 9878, 13715, 12516, 6208, 1520, 115, 294, 850, 2425, 6846, 13219, 17636, 20069, 79, 315, 904, 2621,7940, 17580, 28739, 39642,207, 459, 1118, 3286 , 12772, 27920, 36068, 27866, 425,1389, 3867, 9886,22137, 32651, 29455, 19188, 33,289, 1106, 4131,13281, 27901, 39452, 41979, 273,1204,3392,9191,20667, 32952, 33928, 28064, 289,1252,4295,12297,28768, 45930, 45442, 30553, 105,903,3656,10676,27207, 47240, 58062, 65854],
    'Age Group': ['0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+', '0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+','0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+','0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+','0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+','0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+','0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+','0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+','0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+','0-24','25-34','35-44','45-54', '55-64', '65-74', '75-84', '85+']
})

alt.Chart(coviddata).mark_bar().encode(
    y=alt.Y('Health Condition', sort=['Alzheimer disease', 'Obesity', 'Cerebrovascular diseases', 'Heart failure', 'Chronic lower respiratory diseases', 'Sepsis', 'Ischemic heart disease', 'Renal failure', 'Diabetes', 'Hypertensive diseases']),
    x='Number of COVID-19 Deaths',
    color= alt.Color('Age Group',
                     scale=alt.Scale(domain=['0-24','25-34','35-44','45-54','55-64', '65-74', '75-84', '85+'],
               range=['#FF8A25','#FFD75F','#fffba0','#a1dcf0','#6cd5f5','#19b4e0','#0b87b8','#083346']))
).properties(height=200,width=750,title="Health Conditions Contributing to COVID-19 Deaths in the U.S. (2020-2023)")
```

Out[20]:



In [ ]: