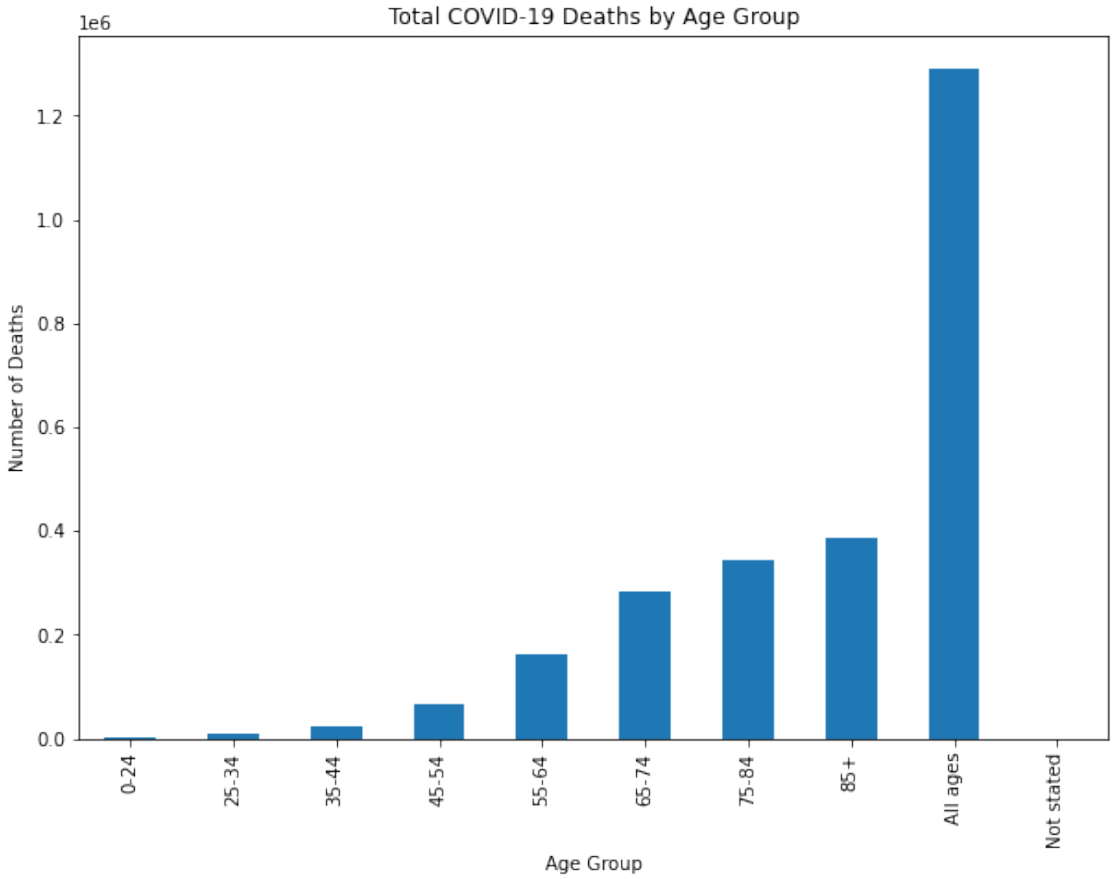
Assignment_2

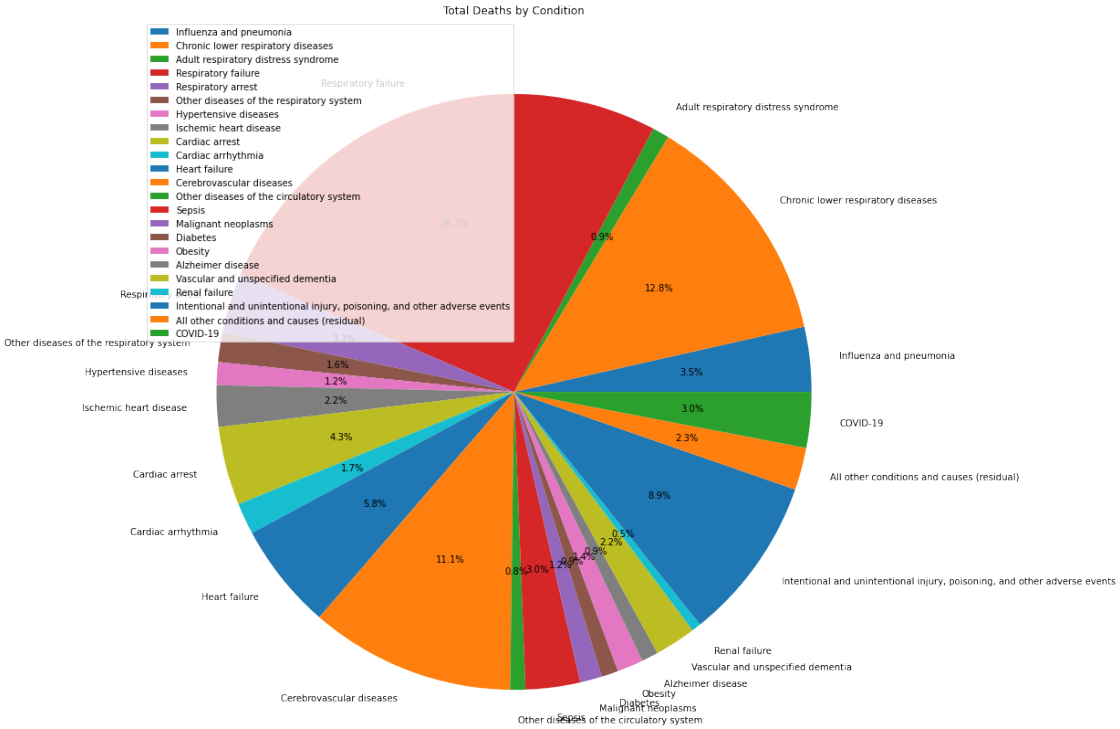
November 19, 2024

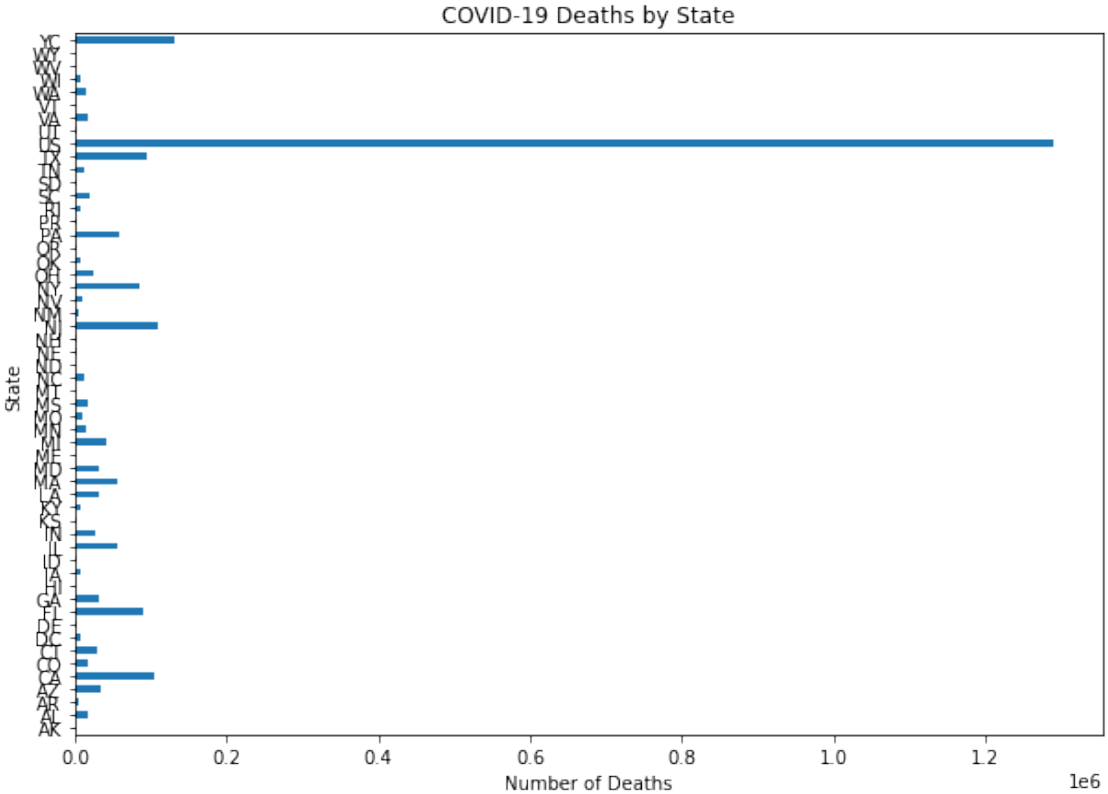
[15]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline
plt.figure(figsize=(10,7))
# Load the CSV file ,into a DataFrame
df = pd.read_csv('coviddeath.csv')
df['Data as of'] = pd.to_datetime(df['Data as of'], format='%m/%d/%Y',␣
 ↪errors='coerce')
df['Start Week'] = pd.to_datetime(df['Start Week'], format='%m/%d/%Y',␣
 ↪errors='coerce')
df['End Week'] = pd.to_datetime(df['End Week'], format='%m/%d/%Y',␣
 ↪errors='coerce')

df['Number of COVID-19 Deaths'] = pd.to_numeric(df['Number of COVID-19␣
 ↪Deaths'], errors='coerce')

# 1. Total COVID-19 Deaths by Age Group
plt.figure(figsize=(10,7))
age_group_totals = df.groupby('Age Group')['Number of COVID-19 Deaths'].sum()
age_group_totals.plot(kind='bar', title='Total COVID-19 Deaths by Age Group')
plt.ylabel('Number of Deaths')
plt.show()

# 2. Total Deaths by Condition
condition_totals = df.groupby('Condition')['Number of COVID-19 Deaths'].sum()
plt.figure(figsize=(20,15))
label = list(df['Condition'].unique())
condition_totals.plot(kind='pie', title='Total Deaths by Condition',␣
 ↪autopct='%1.1f%%',labels = label)
plt.xticks(rotation=45)
plt.legend()
plt.ylabel('')
plt.show()
```

```
# 3. Deaths by State
state_totals = df.groupby('State')['Number of COVID-19 Deaths'].sum()
plt.figure(figsize=(10,7))
state_totals.plot(kind='barh', title='COVID-19 Deaths by State')
plt.xlabel('Number of Deaths')
plt.show()

# 3. Deaths by Condition Group and Age Group
condition_age_group_totals = df.groupby(['Condition Group', 'Age␣
 ↪Group'])['Number of COVID-19 Deaths'].sum().unstack()
plt.figure(figsize=(10,7))
condition_age_group_totals.plot(kind='bar', stacked=True, title='Deaths by␣
 ↪Condition Group and Age Group')
plt.ylabel('Number of Deaths')
plt.show()

# 4. Heatmap of Deaths by Age Group and Condition
heatmap_data = df.pivot_table(values='Number of COVID-19 Deaths', index='Age␣
 ↪Group', columns='Condition', aggfunc='sum')
plt.figure(figsize=(10,7))
sns.heatmap(heatmap_data, annot=True, fmt='g', cmap='viridis')
plt.title('Heatmap of Deaths by Age Group and Condition')
plt.show()
```

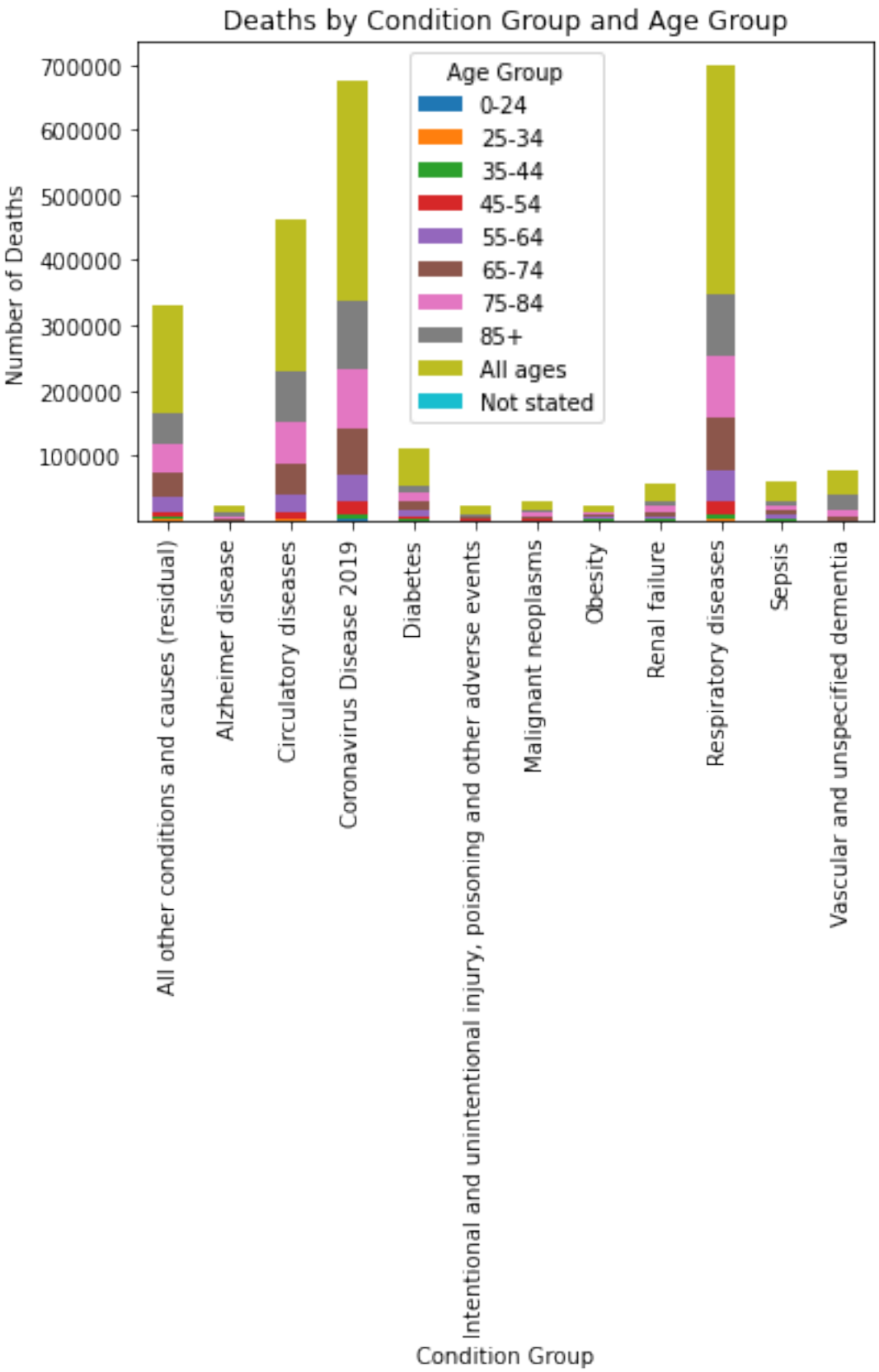
```
<Figure size 720x504 with 0 Axes>
```

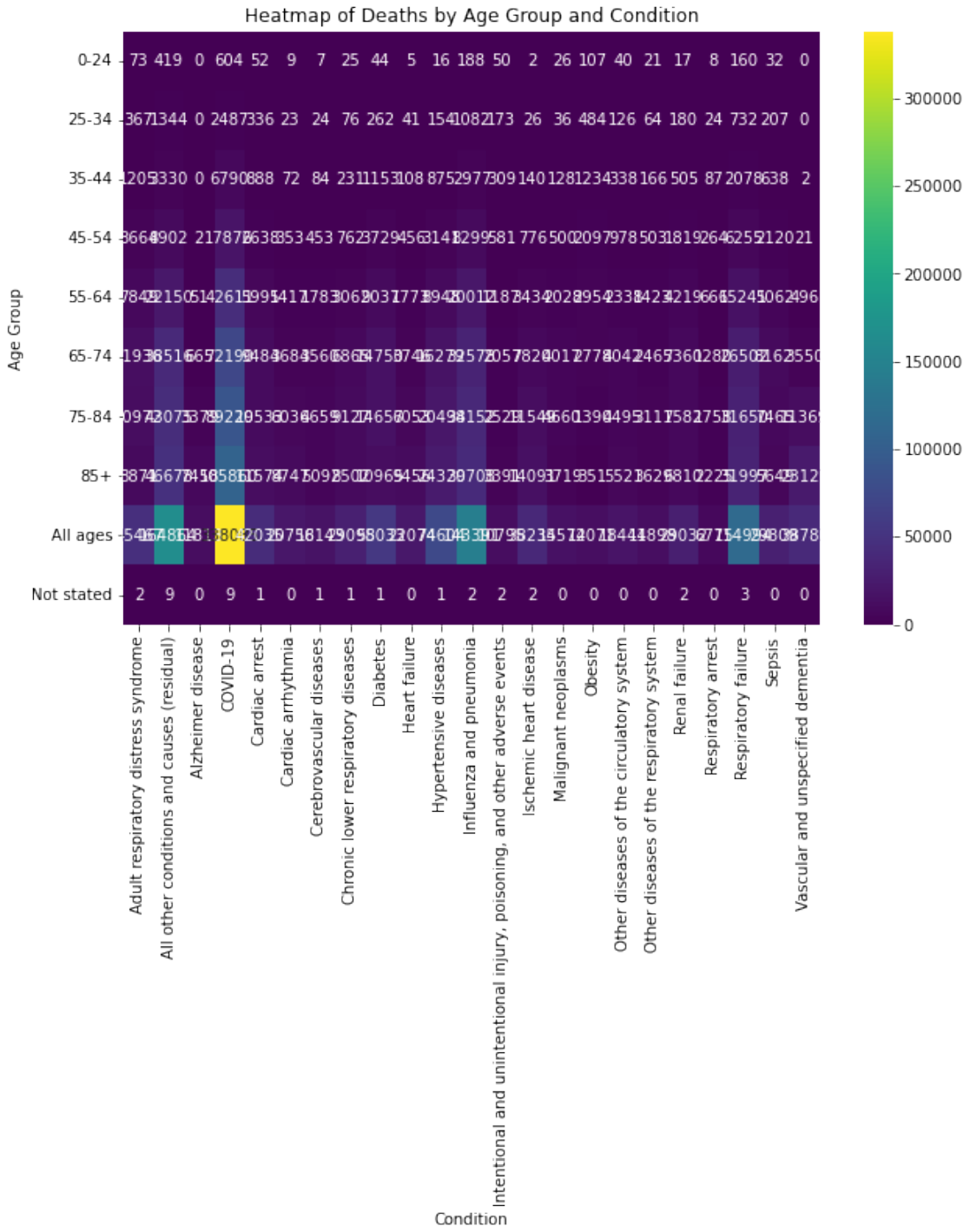

Total COVID-19 Deaths by Age Group
1e6
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Number of Deaths
0-24
25-34
35-44
45-54
55-64
65-74
75-84
85+
All ages
Not stated
Age Group


Total Deaths by Condition
Influenza and pneumonia
Chronic lower respiratory diseases
Adult respiratory distress syndrome
Respiratory failure
Respiratory arrest
Other diseases of the respiratory system
Hypertensive diseases
Ischemic heart disease
Cardiac arrest
Cardiac arrhythmia
Heart failure
Cerebrovascular diseases
Other diseases of the circulatory system
Sepsis
Malignant neoplasms
Diabetes
Obesity
Alzheimer disease
Vascular and unspecified dementia
Renal failure
Intentional and unintentional injury, poisoning, and other adverse events
All other conditions and causes (residual)
COVID-19
26.3%
0.9%
12.8%
3.5%
3.0%
2.3%
8.9%
0.5%
2.2%
0.8%
1.2%
3.0%
0.8%
11.1%
5.8%
1.7%
4.3%
2.2%
1.2%
1.6%
3.3%


COVID-19 Deaths by State
State
Number of Deaths
1e6
0.0
0.2
0.4
0.6
0.8
1.0
1.2

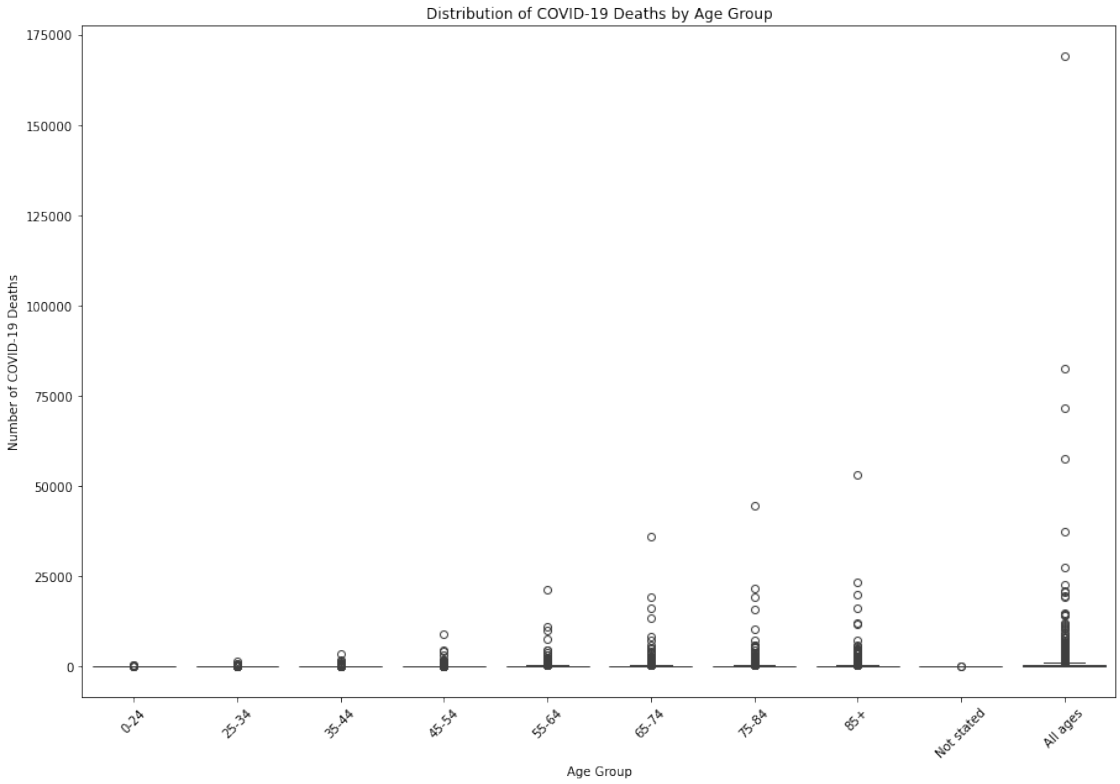
```
<Figure size 720x504 with 0 Axes>
```

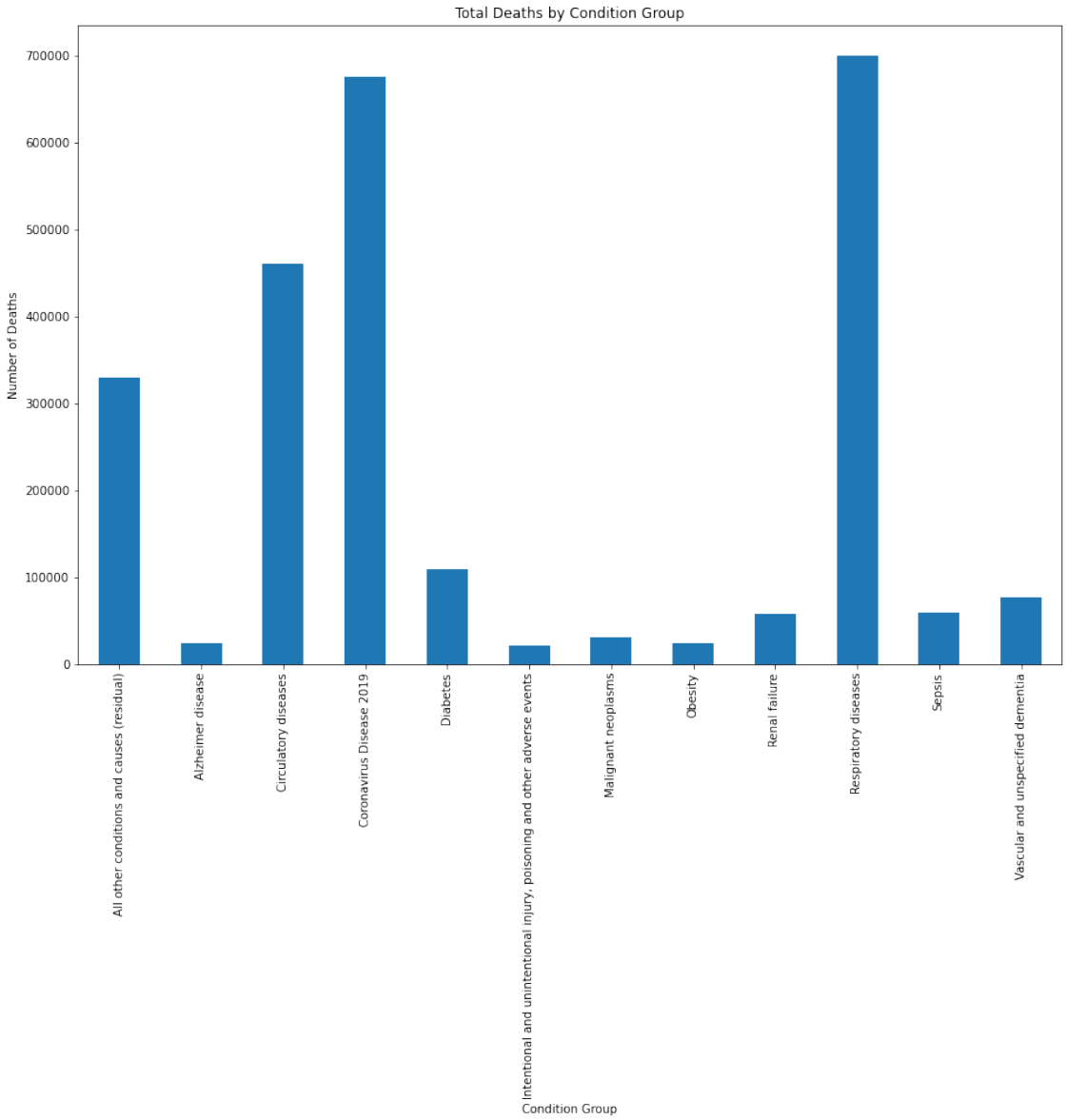

Deaths by Condition Group and Age Group
Number of Deaths
700000
600000
500000
400000
300000
200000
100000
Age Group
0-24
25-34
35-44
45-54
55-64
65-74
75-84
85+
All ages
Not stated
All other conditions and causes (residual)
Alzheimer disease
Circulatory diseases
Coronavirus Disease 2019
Diabetes
Intentional and unintentional injury, poisoning and other adverse events
Malignant neoplasms
Obesity
Renal failure
Respiratory diseases
Sepsis
Vascular and unspecified dementia
Condition Group



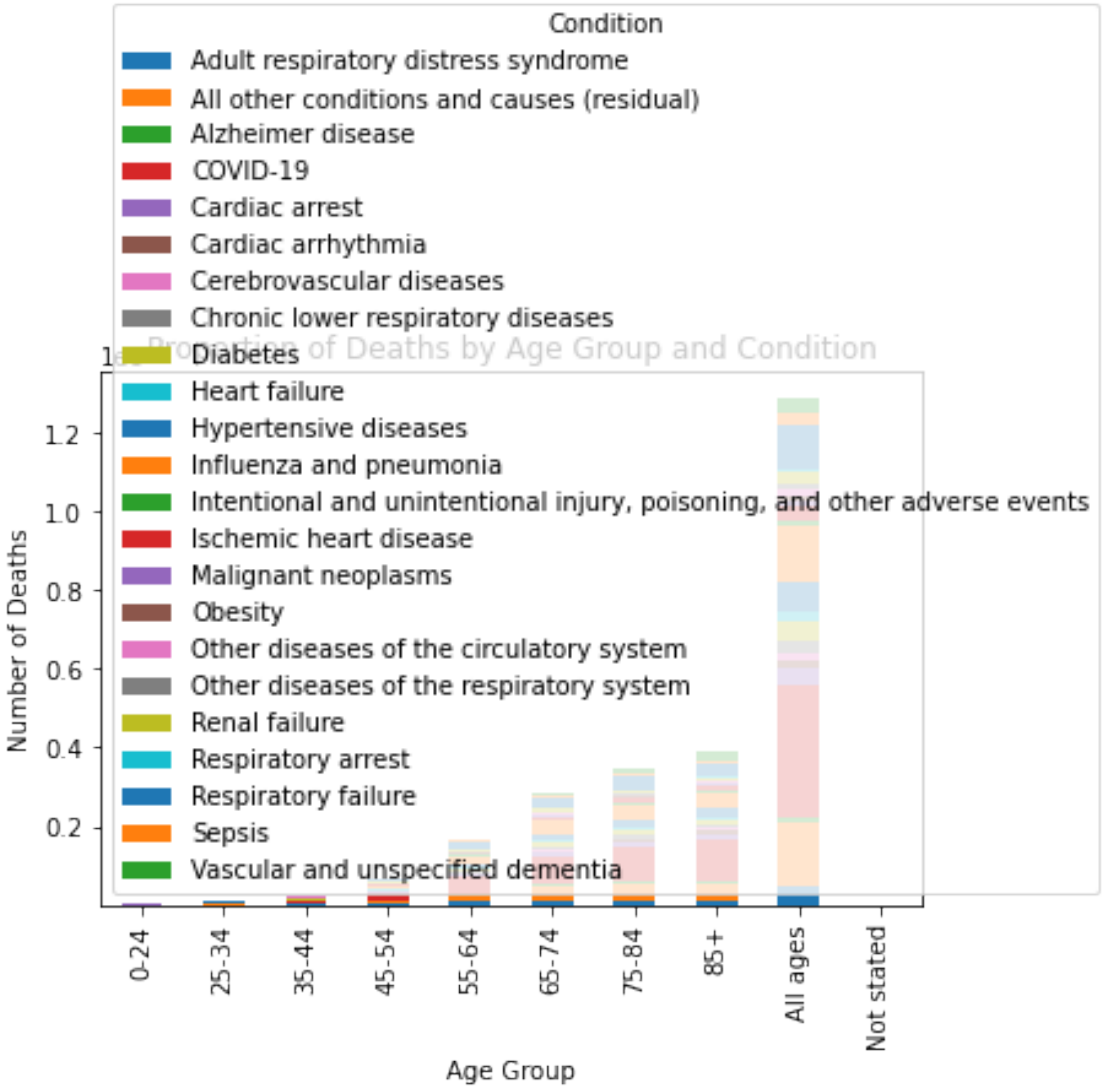
[16]:
```
# 5. Distribution of COVID-19 Deaths by Age Group
plt.figure(figsize=(15, 10))
```

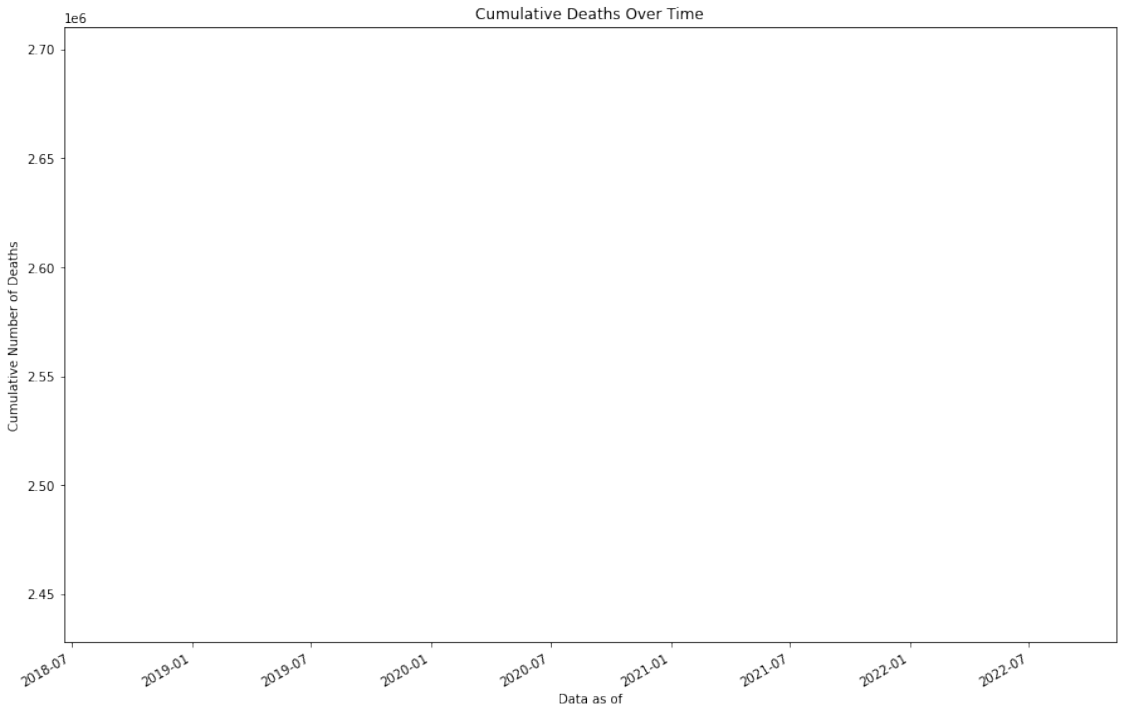
```
sns.boxplot(x='Age Group', y='Number of COVID-19 Deaths', data=df)
plt.title('Distribution of COVID-19 Deaths by Age Group')
plt.xticks(rotation=45)
plt.show()

# 6. Total Deaths by Condition Group
condition_group_totals = df.groupby('Condition Group')['Number of COVID-19␣
 ↪Deaths'].sum()
plt.figure(figsize=(15, 10))
condition_group_totals.plot(kind='bar', title='Total Deaths by Condition Group')
plt.ylabel('Number of Deaths')
plt.show()

# 7. Trend of Deaths by Condition Over Time
# Assuming there's a 'Date' column, you would group by date and condition
# trend_data = df.groupby(['Data as of', 'Condition'])['Number of COVID-19␣
 ↪Deaths'].sum().unstack()
# plt.figure(figsize=(15, 10))
# trend_data.plot(kind='line', title='Trend of Deaths by Condition Over Time')
# plt.ylabel('Number of Deaths')
# plt.show()


# 11. Proportion of Deaths by Age Group and Condition
proportion_data = df.groupby(['Age Group', 'Condition'])['Number of COVID-19␣
 ↪Deaths'].sum().unstack()
plt.figure(figsize=(16, 11))
proportion_data.plot(kind='bar', stacked=True, title='Proportion of Deaths by␣
 ↪Age Group and Condition')
plt.ylabel('Number of Deaths')
plt.show()

# 12. Cumulative Deaths Over Time
# Assuming there's a 'Date' column
cumulative_deaths = df.groupby('Data as of')['Number of COVID-19 Deaths'].sum().
 ↪cumsum()
plt.figure(figsize=(15, 10))
cumulative_deaths.plot(kind='line', title='Cumulative Deaths Over Time')
plt.ylabel('Cumulative Number of Deaths')
plt.show()
```


Distribution of COVID-19 Deaths by Age Group
Number of COVID-19 Deaths
175000
150000
125000
100000
75000
50000
25000
0
0-24
25-34
35-44
45-54
55-64
65-74
75-84
85+
Not stated
All ages
Age Group



<Figure size 1152x792 with 0 Axes>


Condition
Adult respiratory distress syndrome
All other conditions and causes (residual)
Alzheimer disease
COVID-19
Cardiac arrest
Cardiac arrhythmia
Cerebrovascular diseases
Chronic lower respiratory diseases
Diabetes
Heart failure
Hypertensive diseases
Influenza and pneumonia
Intentional and unintentional injury, poisoning, and other adverse events
Ischemic heart disease
Malignant neoplasms
Obesity
Other diseases of the circulatory system
Other diseases of the respiratory system
Renal failure
Respiratory arrest
Respiratory failure
Sepsis
Vascular and unspecified dementia
Number of Deaths
1.2
1.0
0.8
0.6
0.4
0.2
0-24
25-34
35-44
45-54
55-64
65-74
75-84
85+
All ages
Not stated
Age Group



[3]:
```
df.columns
```

[3]:
```
Index(['Data as of', 'Start Week', 'End Week', 'State', 'Condition Group',
       'Condition', 'ICD10_codes', 'Age Group', 'Number of COVID-19 Deaths',
       'Flag'],
      dtype='object')
```

[4]:
```
df.head()
```

[4]:

	Data as of	Start Week	End Week	State	Condition Group	Condition	ICD10_codes	Age Group	Number of COVID-19 Deaths	Flag
0	2020-08-30	2020-02-01	2020-08-29	US	Respiratory diseases	Influenza and pneumonia	J09-J18	0-24	122.0	
1	2020-08-30	2020-02-01	2020-08-29	US	Respiratory diseases	Influenza and pneumonia	J09-J18	25-34	596.0	
2	2020-08-30	2020-02-01	2020-08-29	US	Respiratory diseases	Influenza and pneumonia	J09-J18	35-44	1521.0	
3	2020-08-30	2020-02-01	2020-08-29	US	Respiratory diseases	Influenza and pneumonia	J09-J18	45-54	4186.0	
4	2020-08-30	2020-02-01	2020-08-29	US	Respiratory diseases	Influenza and pneumonia	J09-J18	55-64	10014.0	

```
0  NaN
1  NaN
2  NaN
3  NaN
4  NaN
```

[5]:
```
print(len(df.State.unique()))
print(len(df.Condition.unique()))
print(len(df['Condition Group'].unique()))
```

```
54
23
12
```

[6]:
```
df
```

[6]:

	Data as of	Start Week	End Week	State	Condition Group	Condition	ICD10_codes	Age Group	Number of COVID-19 Deaths	Flag
0	2020-08-30	2020-02-01	2020-08-29	US	Respiratory diseases	Influenza and pneumonia	J09-J18	0-24	122.0	NaN
1	2020-08-30	2020-02-01	2020-08-29	US	Respiratory diseases	Influenza and pneumonia	J09-J18	25-34	596.0	NaN
2	2020-08-30	2020-02-01	2020-08-29	US	Respiratory diseases	Influenza and pneumonia	J09-J18	35-44	1521.0	NaN
3	2020-08-30	2020-02-01	2020-08-29	US	Respiratory diseases	Influenza and pneumonia	J09-J18	45-54	4186.0	NaN
4	2020-08-30	2020-02-01	2020-08-29	US	Respiratory diseases	Influenza and pneumonia	J09-J18	55-64		
…	…	…	…	…	…	…	…	…		
12255	2020-08-30	2020-02-01	2020-08-29	YC	Coronavirus Disease 2019	COVID-19	U071	65-74		
12256	2020-08-30	2020-02-01	2020-08-29	YC	Coronavirus Disease 2019	COVID-19	U071	75-84		
12257	2020-08-30	2020-02-01	2020-08-29	YC	Coronavirus Disease 2019	COVID-19	U071	85+		
12258	2020-08-30	2020-02-01	2020-08-29	YC	Coronavirus Disease 2019	COVID-19	U071	Not stated		
12259	2020-08-30	2020-02-01	2020-08-29	YC	Coronavirus Disease 2019	COVID-19	U071	All ages		

```
4                        10014.0                                NaN
…                          …                                     …
12255                     5024.0                                NaN
12256                     5381.0                                NaN
12257                     4841.0                                NaN
12258                        NaN  Counts less than 10 suppressed.
12259                    20628.0                                NaN

[12260 rows x 10 columns]
```