

▼ MiniProject#2 Exercise 2

```
import pandas as pd
import matplotlib.pyplot as plt
```

▼ NCHS Data Import

```
pdf_data = pd.read_csv("/content/MP2-exercise2_data.csv", header=0)
```

```
pdf_data.head(20)
```



All	Category	All races, both sexes,	Number	Percent of	Rate
-----	----------	------------------------	--------	------------	------

▼ Data Cleansing

```
# Renaming the column depicting Cause of Death
pdf_data.rename(columns = {'All races, both sexes, all ages':'Cause of Death'}, inplace=True)

# Drop column named All as its not adding any value
pdf_data = pdf_data.drop(columns='All')

# Cleaning up Rate column by replacing * with 0
pdf_data.loc[pdf_data['Rate'] == '*'] = 0

# In Rate column, removing ',' to make convert the column to float
pdf_data['Rate'] = pdf_data['Rate'].str.replace(",","").astype(float)

# Check data types of all columns
pdf_data.dtypes

# Remove Cause of Death type 'All Causes' and create a separate dataframe
removed_allcauses_df = pdf_data.loc[~pdf_data['Cause of Death'].str.contains('All causes of death')]

removed_all_causes_df = removed_allcauses_df.loc[~removed_allcauses_df['Cause of Death'].str.contains('All causes of death')]

# Stripping whitespaces
removed_all_causes_df['Cause of Death'] = removed_all_causes_df['Cause of Death'].str.strip()
removed_all_causes_df
```

```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:22: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: <https://pandas.pydata.org/pandas-docs/stab>

▼ Analyzing data for leading cause of death

```

# Filtering by sex and age
removed_all_causes_df = df[df['sex'].isin(['M', 'F']) & df['age'].isin(['109', '111', '113', '120-151'])]

# Removing rate column as its not needed for leading cause of death computation
filtered_df = removed_all_causes_df.drop(columns='Rate')

# Grouping by leading cause of death by age groups
cause_of_death_df = filtered_df.groupby(['Cause of Death']).sum().sort_values('Percent')

cause_of_death_df

#####
# Overall Leading cause of deaths (top 3) are:
# 1.Accidents
# 2.Malignant neoplasms
# 3. Diseases of heart
#####
```

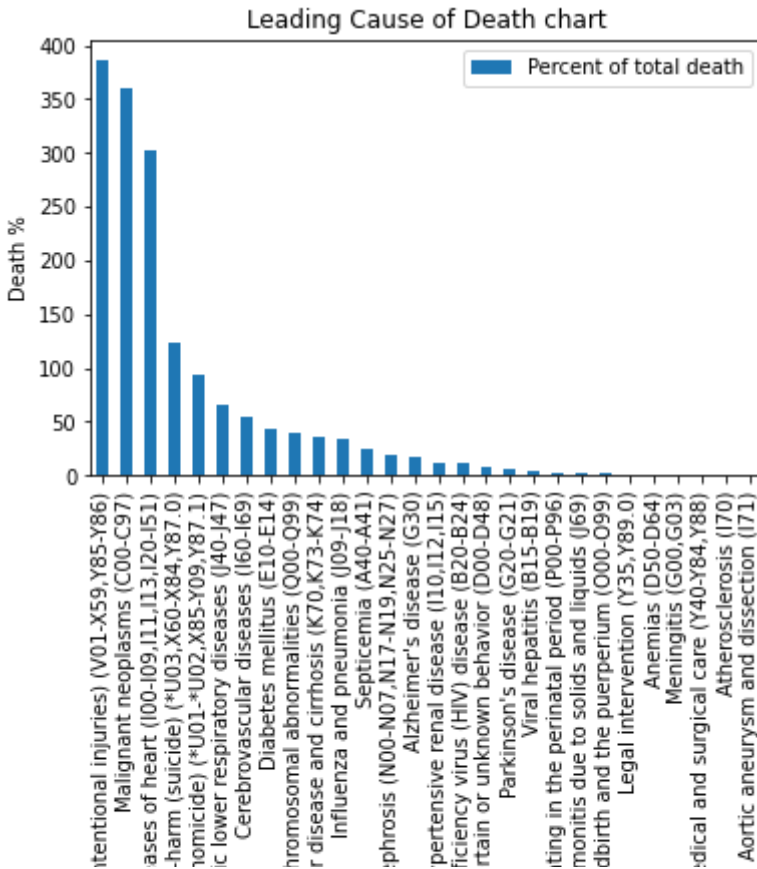
	Percent of total death
Cause of Death	
Accidents (unintentional injuries) (V01-X59,Y85-Y86)	385.9
Malignant neoplasms (C00-C97)	360.1
Diseases of heart (I00-I09,I11,I13,I20-I51)	302.5
Intentional self-harm (suicide) (*U03,X60-X84,Y87.0)	124.1
Assault (homicide) (*U01-*U02,X85-Y09,Y87.1)	92.6
Chronic lower respiratory diseases (J40-J47)	65.5
Cerebrovascular diseases (I60-I69)	54.4
Diabetes mellitus (E10-E14)	43.2
Congenital malformations, deformations and chromosomal abnormalities (Q00-Q99)	39.1
Chronic liver disease and cirrhosis (K70,K73-K74)	35.4
Influenza and pneumonia (J09-J18)	33.0

Plotting leading cause of death

Nephritis, nephrotic syndrome and nephrosis (N00-N07,N17-N19,N25-

10.2

```
# Plot overall leading cause of death for all groups
cause_of_death_df.plot(kind="bar")
plt.xlabel('Cause of Death')
plt.ylabel('Death %')
plt.title('Leading Cause of Death chart')
plt.show()
```



▼ Analyzing for Death trends based on Age

```

AC      Int      m      nd      te      nai      ple      nd      Pre      jik

filtered_df
# Creating data frame for each age group so we can glean clear cause of death
all_ages_df = filtered_df.loc[filtered_df['Category'].str.contains('all ages',na=False)]
all_ages_df = all_ages_df.groupby(['Cause of Death']).sum().sort_values('Percent of total death')
all_ages_df

age_1_4_df = filtered_df.loc[filtered_df['Category'].str.contains('1-4 years',na=False)]
age_1_4_df = age_1_4_df.groupby(['Cause of Death']).sum().sort_values('Percent of total death')
age_1_4_df

age_5_14_df = filtered_df.loc[filtered_df['Category'].str.contains('5-14 years',na=False)]
age_5_14_df = age_5_14_df.groupby(['Cause of Death']).sum().sort_values('Percent of total death')
age_5_14_df

age_15_24_df = filtered_df.loc[filtered_df['Category'].str.contains('15-24 years',na=False)]
age_15_24_df = age_15_24_df.groupby(['Cause of Death']).sum().sort_values('Percent of total death')
age_15_24_df

age_25_34_df = filtered_df.loc[filtered_df['Category'].str.contains('25-34 years',na=False)]
age_25_34_df = age_25_34_df.groupby(['Cause of Death']).sum().sort_values('Percent of total death')
age_25_34_df

age_35_44_df = filtered_df.loc[filtered_df['Category'].str.contains('35-44 years',na=False)]
age_35_44_df = age_35_44_df.groupby(['Cause of Death']).sum().sort_values('Percent of total death')
age_35_44_df

```

```
age_35_44_df
```

```
age_45_54_df = filtered_df.loc[filtered_df['Category'].str.contains('45-54 years',na=I
age_45_54_df = age_45_54_df.groupby(['Cause of Death']).sum().sort_values('Percent of
age_45_54_df
```

```
age_55_64_df = filtered_df.loc[filtered_df['Category'].str.contains('55-64 years',na=I
age_55_64_df = age_55_64_df.groupby(['Cause of Death']).sum().sort_values('Percent of
age_55_64_df
```

```
age_65_74_df = filtered_df.loc[filtered_df['Category'].str.contains('65-74 years',na=I
age_65_74_df = age_65_74_df.groupby(['Cause of Death']).sum().sort_values('Percent of
age_65_74_df
```

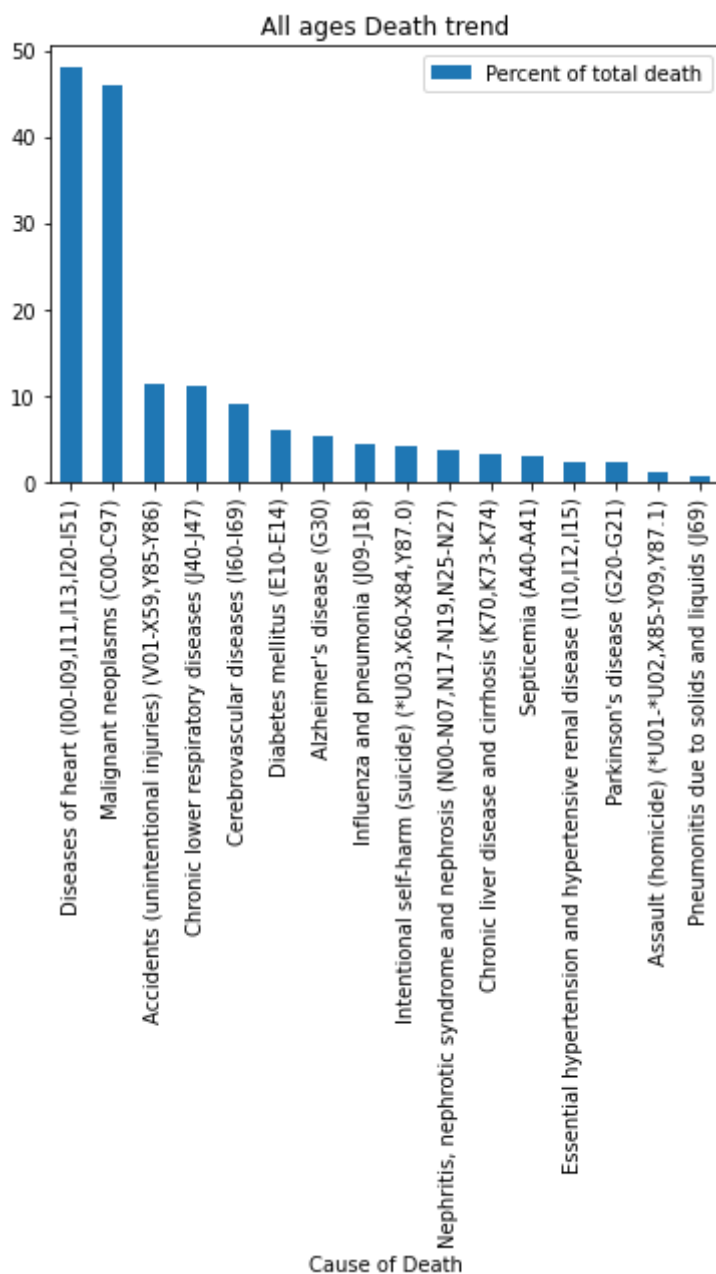
```
age_75_84_df = filtered_df.loc[filtered_df['Category'].str.contains('75-84 years',na=I
age_75_84_df = age_75_84_df.groupby(['Cause of Death']).sum().sort_values('Percent of
age_75_84_df
```

```
age_85over_df = filtered_df.loc[filtered_df['Category'].str.contains('85 years',na=Fa
age_85over_df = age_85over_df.groupby(['Cause of Death']).sum().sort_values('Percent c
age_85over_df
```

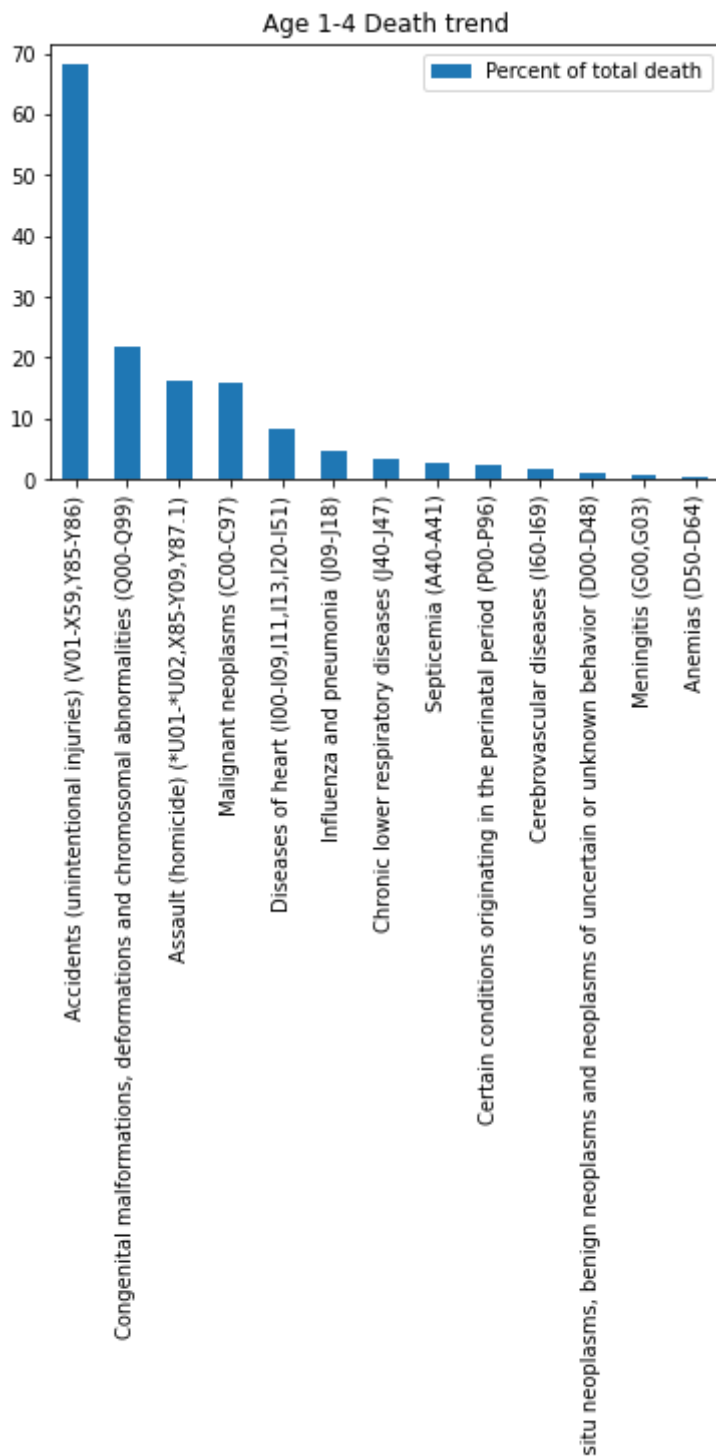
	Percent of total death
Cause of Death	
Diseases of heart (I00-I09,I11,I13,I20-I51)	29.4
Malignant neoplasms (C00-C97)	12.0
Alzheimer's disease (G30)	6.8
Cerebrovascular diseases (I60-I69)	6.6
Chronic lower respiratory diseases (J40-J47)	5.1
Influenza and pneumonia (J09-J18)	3.2
Accidents (unintentional injuries) (V01-X59,Y85-Y86)	2.5
Nephritis, nephrotic syndrome and nephrosis (N00-N07,N17-N19,N25-N27)	2.1
Diabetes mellitus (E10-E14)	2.0
Essential hypertension and hypertensive renal disease (I10,I12,I15)	1.7
Parkinson's disease (G20-G21)	1.3
Septicemia (A40-A41)	1.3
Pneumonitis due to solids and liquids (J69)	1.1
In situ neoplasms, benign neoplasms and neoplasms of uncertain or unknown behavior (D00-D48)	0.6

Plotting Age based trends

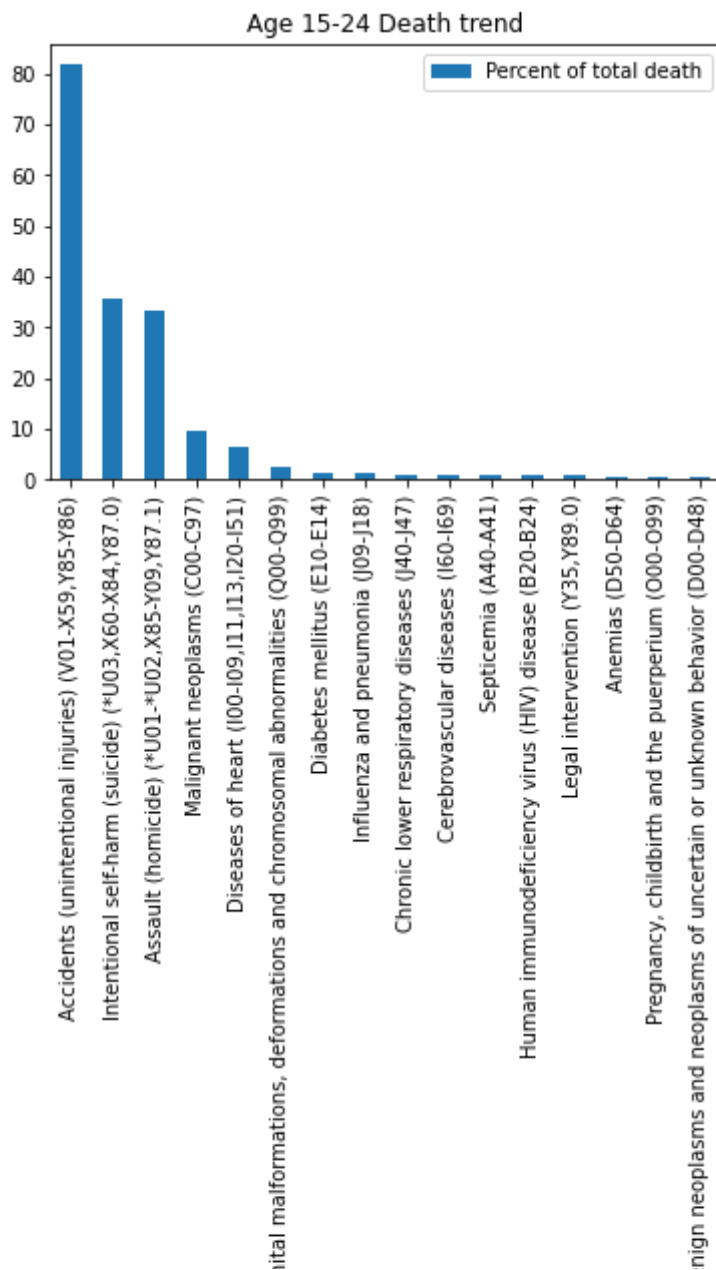
```
# All ages death trend
all_ages_df.plot(kind="bar")
plt.title('All ages Death trend')
plt.show()
```



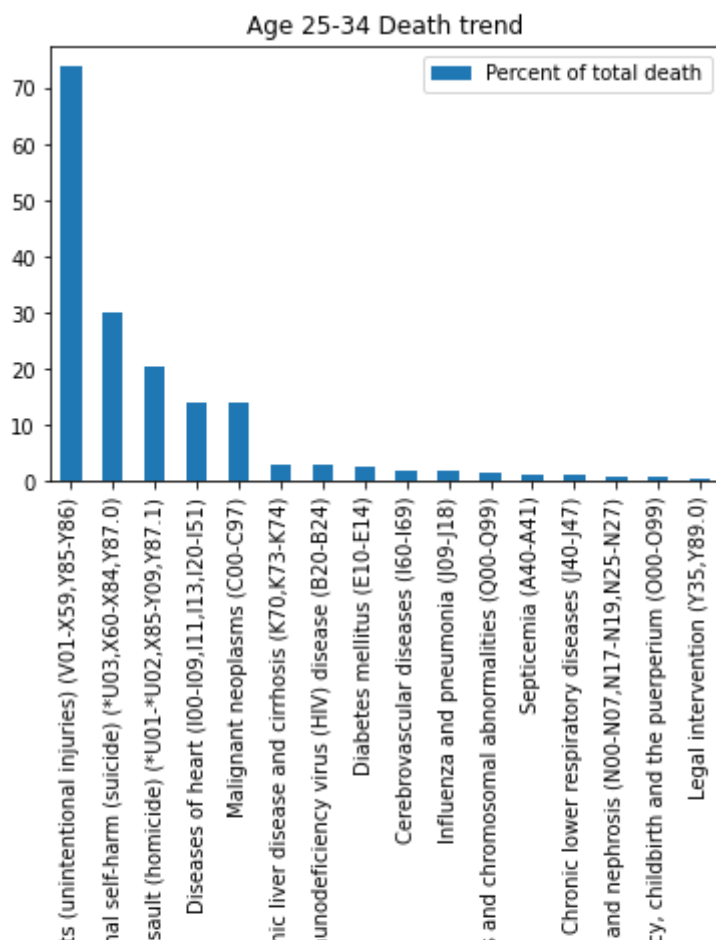
```
# Ages 1-4 death trend
age_1_4_df.plot(kind="bar")
plt.title("Age 1-4 Death trend")
plt.show()
```



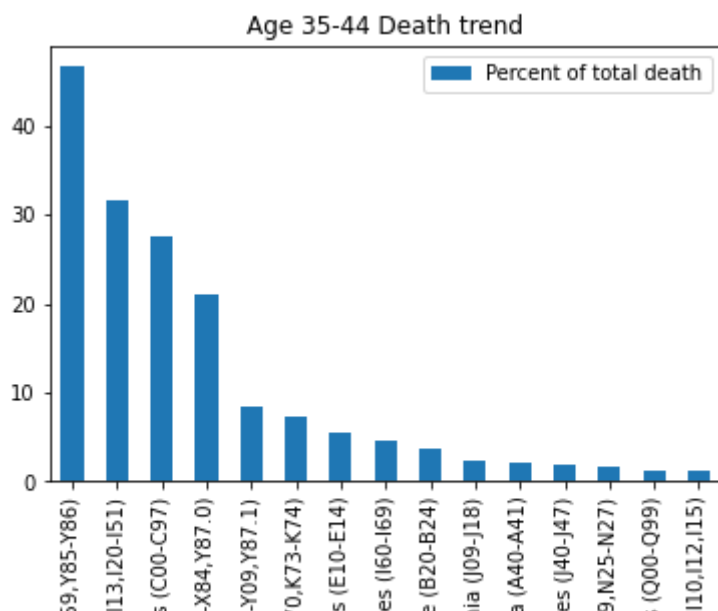
```
# Ages 15-24 death trend
age_15_24_df.plot(kind="bar")
plt.title("Age 15-24 Death trend")
plt.show()
```

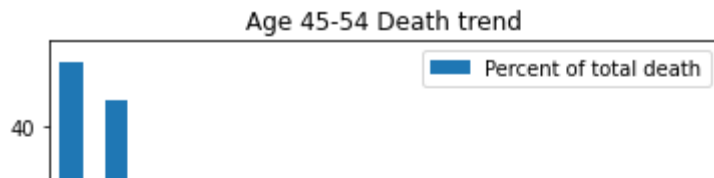
```
# Ages 25-34 death trend
age_25_34_df.plot(kind="bar")
plt.title("Age 25-34 Death trend")
plt.show()
```



```
# Ages 35-44 death trend
age_35_44_df.plot(kind="bar")
plt.title("Age 35-44 Death trend")
plt.show()
```

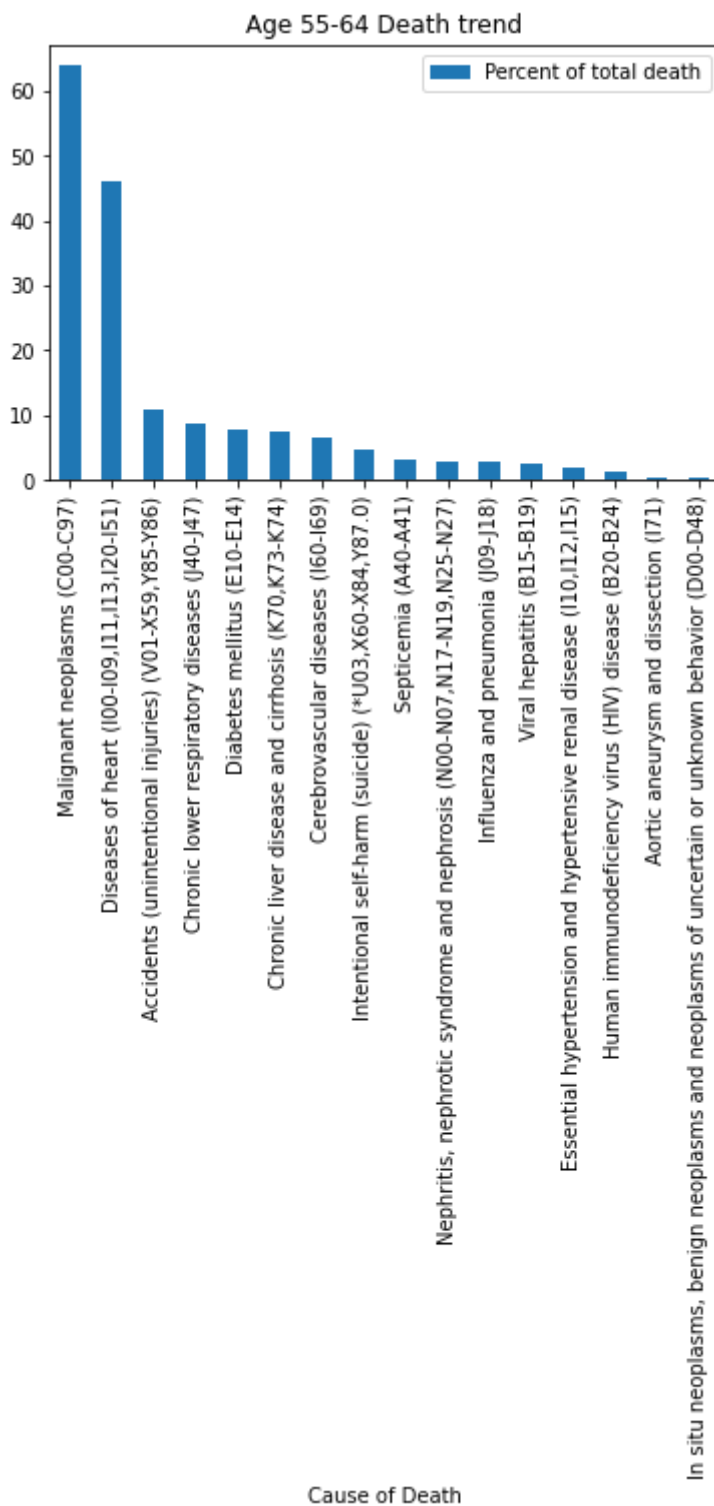


```
# Ages 45-54 death trend
age_45_54_df.plot(kind="bar")
plt.title("Age 45-54 Death trend")
plt.show()
```

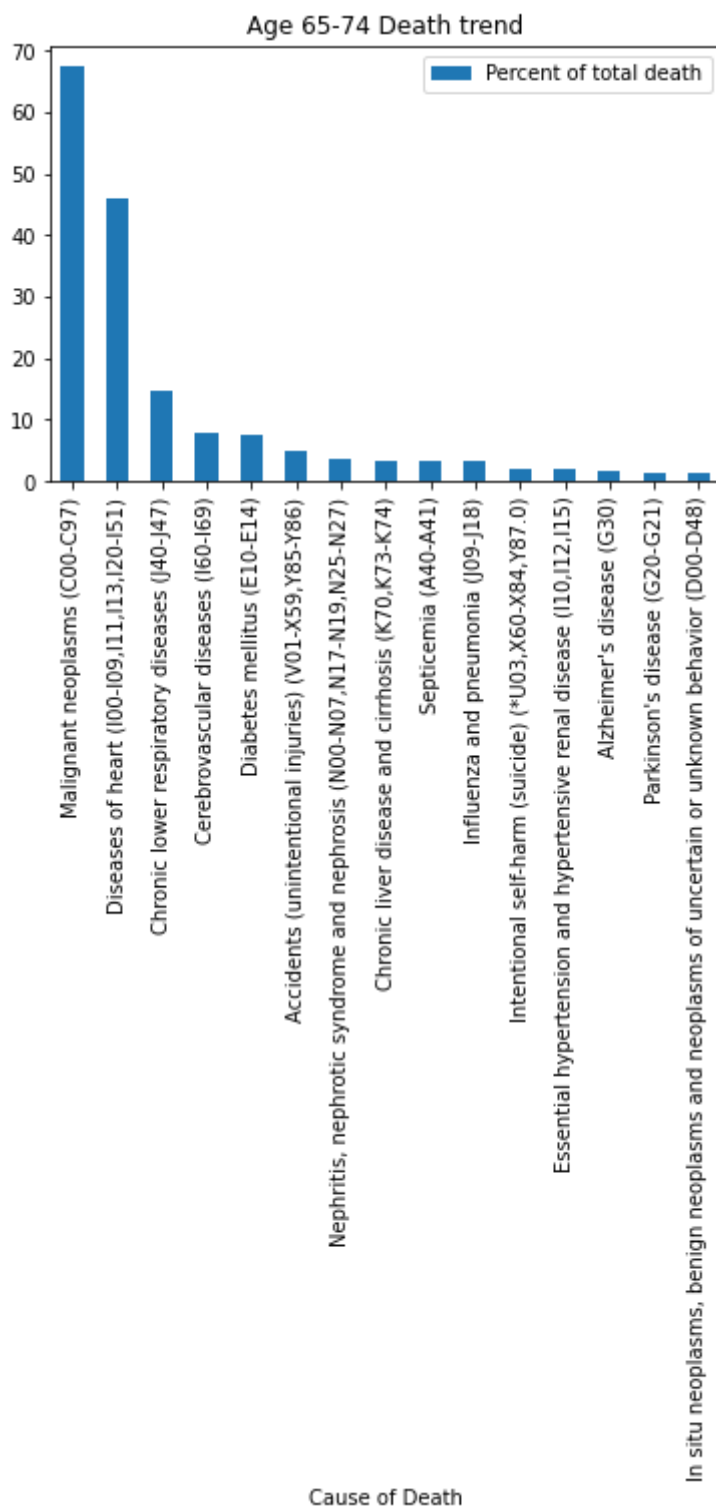


Ages 55-64 death trend

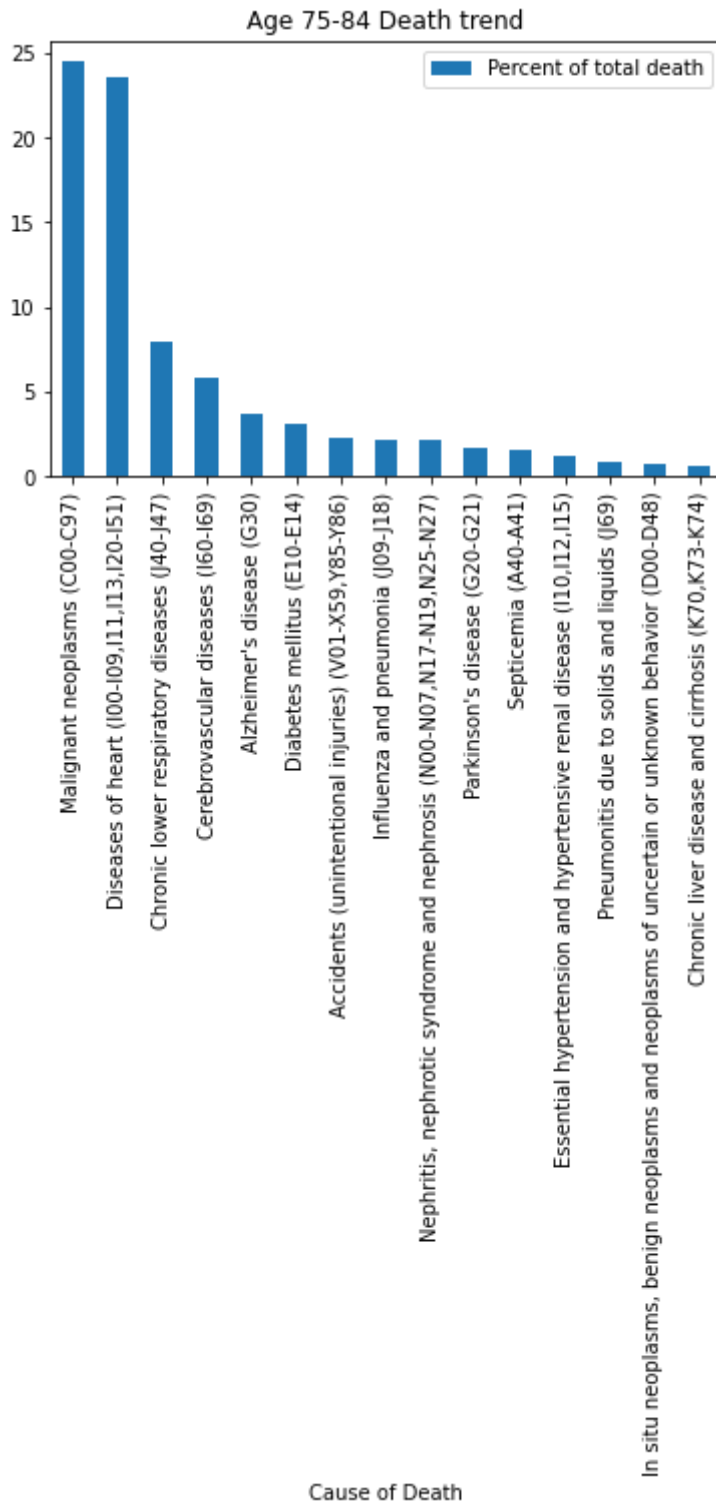
```
age_55_64_df.plot(kind="bar")
plt.title("Age 55-64 Death trend")
plt.show()
```



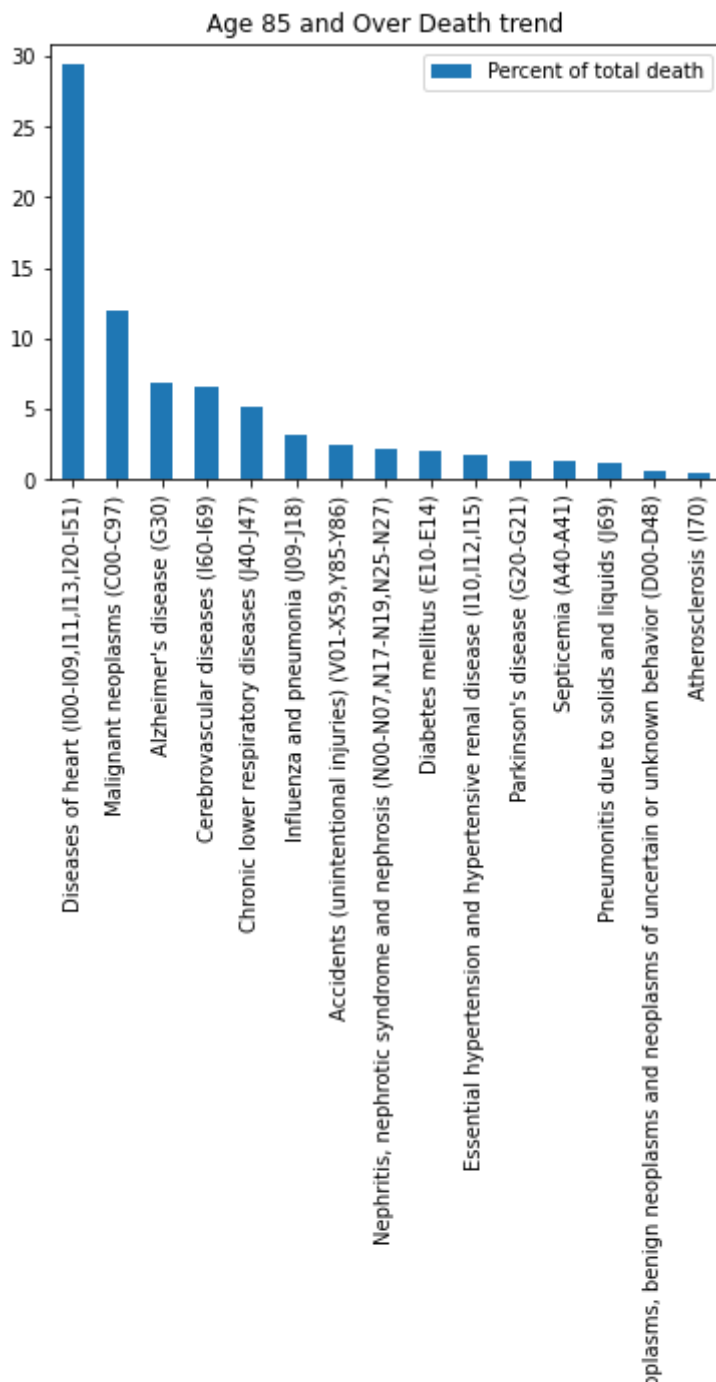
```
# Ages 65-74 death trend
age_65_74_df.plot(kind="bar")
plt.title("Age 65-74 Death trend")
plt.show()
```



```
# Ages 75-84 death trend
age_75_84_df.plot(kind="bar")
plt.title("Age 75-84 Death trend")
plt.show()
```



```
# Ages 85 and Over death trend
age_85over_df.plot(kind="bar")
plt.title("Age 85 and Over Death trend")
plt.show()
```



▼ Analyzing for Sex related trend for cause of death

```
filtered_df
```

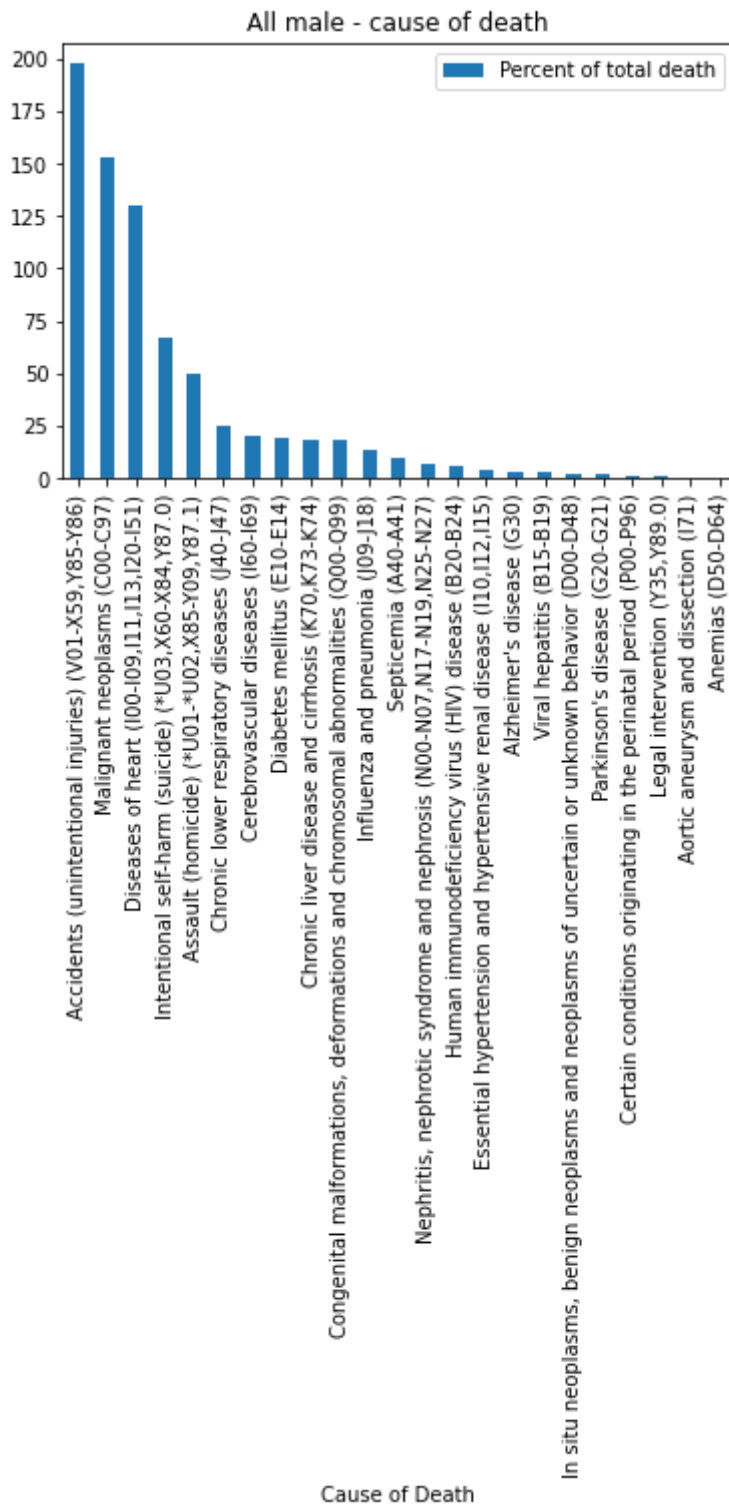
```
all_male_df = filtered_df.loc[filtered_df['Category'].str.contains('male',na=False)]
all_male_df = all_male_df.groupby(['Cause of Death']).sum().sort_values('Percent of total death')
all_male_df
```

```
both_sexes_df = filtered_df.loc[filtered_df['Category'].str.contains('both sexes',na=False)]
both_sexes_df = both_sexes_df.groupby(['Cause of Death']).sum().sort_values('Percent of total death')
both_sexes_df
```

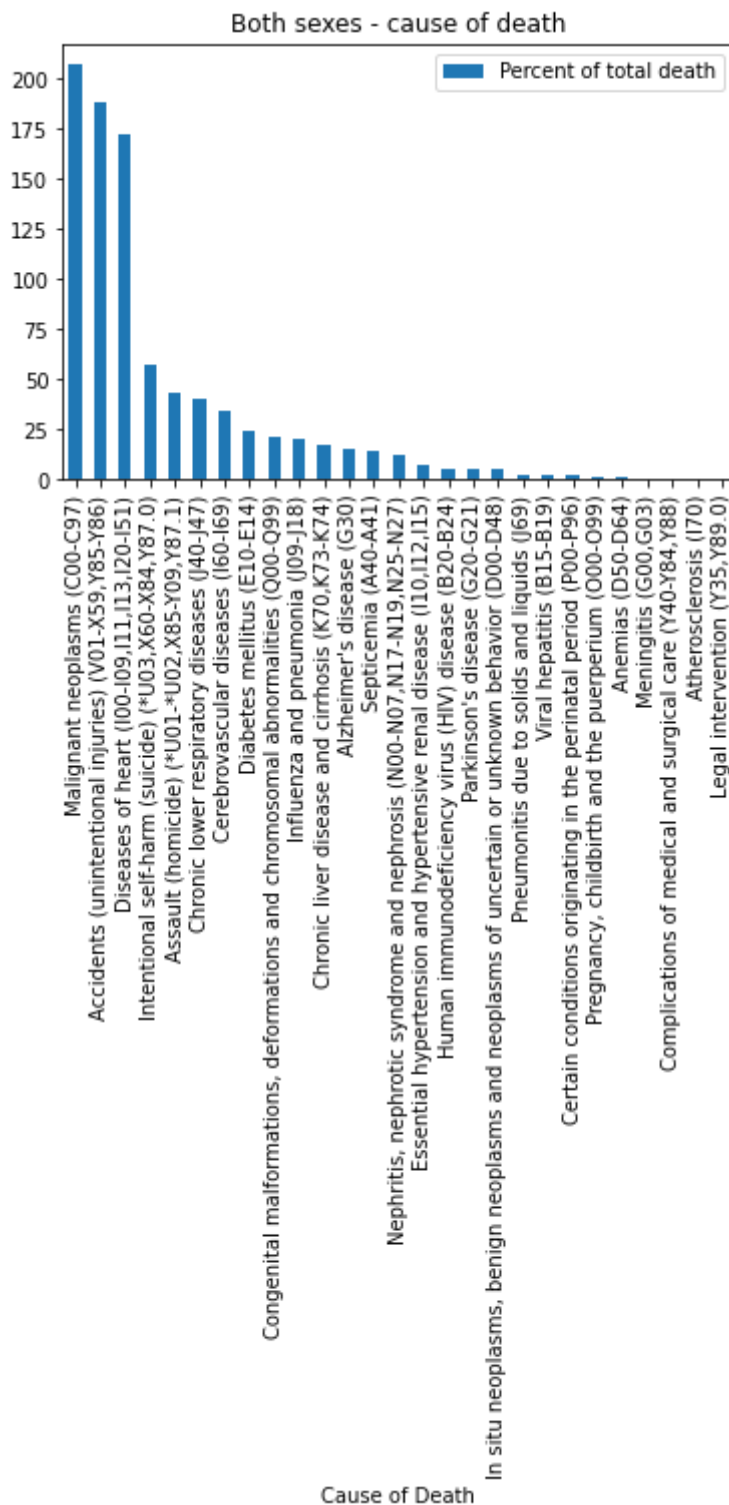
Cause of Death		Percent of total death
Malignant neoplasms (C00-C97)		207.3
Accidents (unintentional injuries) (V01-X59,Y85-Y86)		188.2
Diseases of heart (I00-I09,I11,I13,I20-I51)		172.3
Intentional self-harm (suicide) (*U03,X60-X84,Y87.0)		57.3
Assault (homicide) (*U01-*U02,X85-Y09,Y87.1)		42.8
Chronic lower respiratory diseases (J40-J47)		40.5
Cerebrovascular diseases (I60-I69)		34.6
Diabetes mellitus (E10-E14)		24.4
Congenital malformations, deformations and chromosomal abnormalities (Q00-Q99)		20.9
Influenza and pneumonia (J09-J18)		19.8
Chronic liver disease and cirrhosis (K70,K73-K74)		17.1
Alzheimer's disease (G30)		14.8
Septicemia (A40-A41)		13.8
Nephritis, nephrotic syndrome and nephrosis (N00-N07,N17-N19,N25-N27)		12.0
Essential hypertension and hypertensive renal disease (I10,I12,I15)		7.5
Human immunodeficiency virus (HIV) disease (B20-B24)		5.4
Parkinson's disease (G20-G21)		4.7
In situ neoplasms, benign neoplasms and neoplasms of uncertain or unknown behavior (D00-D48)		4.7
Pneumonitis due to solids and liquids (J69)		2.6
Viral hepatitis (B15-B19)		2.3
Certain conditions originating in the perinatal period (P00-P96)		1.7
Pregnancy, childbirth and the puerperium (O00-O99)		1.6
Anemias (D50-D64)		0.8
Meningitis (G00,G03)		0.6
Complications of medical and surgical care (Y40-Y84,Y88)		0.5

Plotting sex related cause of death


```
all_male_df.plot(kind="bar")
plt.title("All male - cause of death")
plt.show()
```



```
both_sexes_df.plot(kind="bar")
plt.title("Both sexes - cause of death")
plt.show()
```



Analysis Notes:

Based on above trends, we can see that:

1. Top 3 cause of death among most groups in general is: a. Accidents b. Malignant neoplasms c. Disease of heart
2. But during younger age groups, accidents are leading cause of death
3. Older age groups dominate with disease of heart & malignant neoplasms

4. Males leading cause of death is Accidents
5. Females leading cause of death is Malignant neoplasms