Suicides are Preventable, Lets Analyse How

Case Study on Suicide Cases in India

Disclaimer: This is a serious topic viewers are advised
Motive: I believe with the right effort at the right places many lives can be saved

Dataset Provided by Rajanand Ilangovan Kraggle

Link

In [1]:
```
import warnings
warnings.filterwarnings('ignore')

import pandas as pd
import numpy as np
from wordcloud import WordCloud, STOPWORDS
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [2]:
```
df=pd.read_csv('data/Suicides_in_India_2001-2012.csv')
df.head()
```

Out[2]:

	State	Year	Type_code	Type	Gender	Age_group	Total
0	A & N Islands	2001	Causes	Illness (Aids/STD)	Female	0-14	0
1	A & N Islands	2001	Causes	Bankruptcy or Sudden change in Economic	Female	0-14	0
2	A & N Islands	2001	Causes	Cancellation/Non-Settlement of Marriage	Female	0-14	0
3	A & N Islands	2001	Causes	Physical Abuse (Rape/Incest Etc.)	Female	0-14	0
4	A & N Islands	2001	Causes	Dowry Dispute	Female	0-14	0

Cleaning Data

In [3]:
```
modified_df=df[~df["State"].str.contains("Total")]
```

In [4]:
```
states=modified_df['State'].replace(to_replace="A & N Islands", value="Andaman and

states=states.replace(to_replace="D & N Haveli", value="Dadra and Nagar Haveli")

states=states.str.replace(r"\(.*\)","")
```

In [5]:
```
modified_df["State"]=states
```

In [6]:
```
modified_df.drop_duplicates(keep='first',inplace=True)
```

In [7]:
```
modified_df.columns[modified_df.isna().any()].tolist()
```

Out[7]:
```
[]
```

In [8]:
```
modified_df.to_csv('modified_data.csv',index=False)
```

In [9]:
```
modified_df.head()
```

Out[9]:

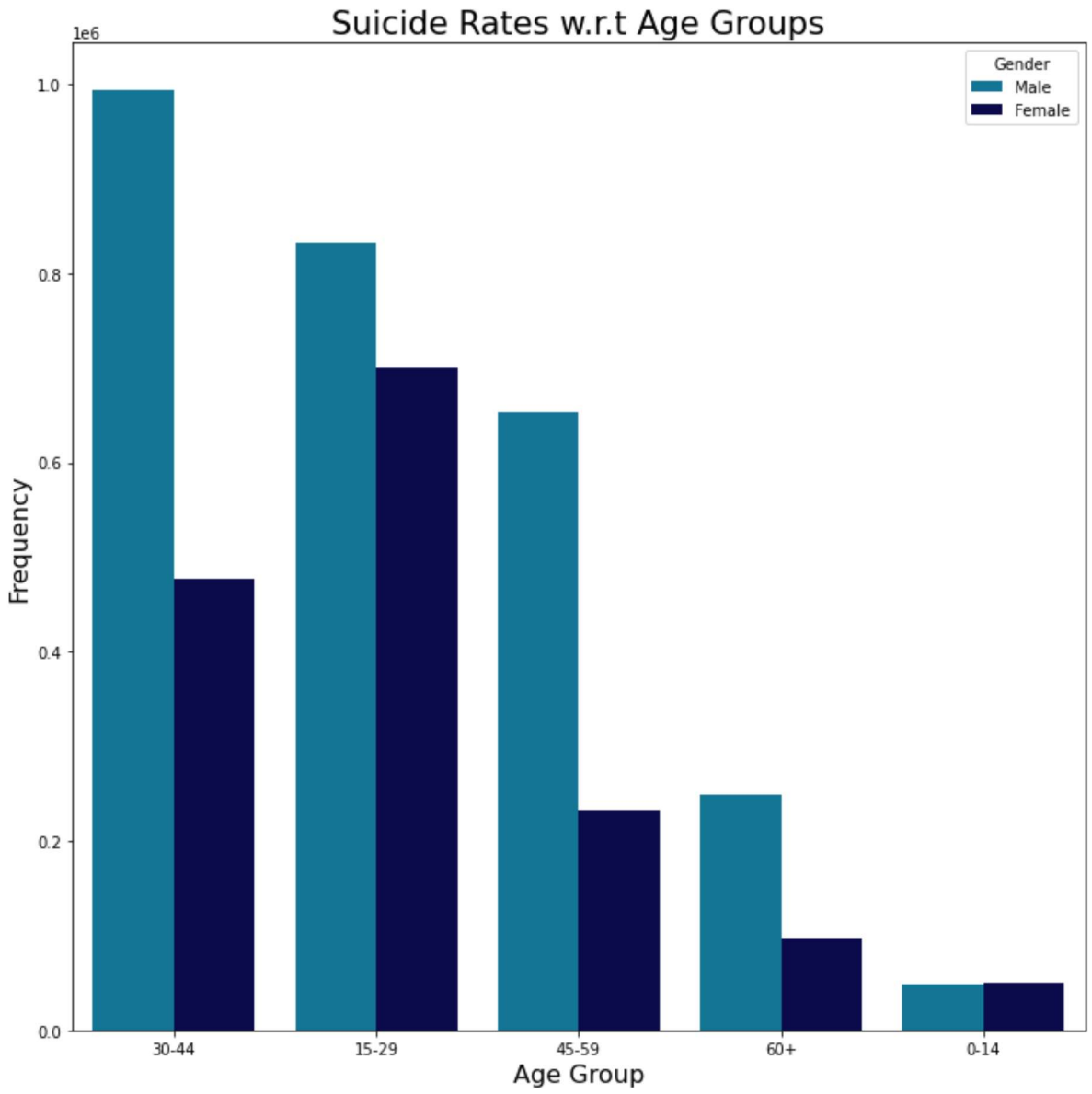
	State	Year	Type_code	Type	Gender	Age_group	Total
0	Andaman and Nicobar Islands	2001	Causes	Illness (Aids/STD)	Female	0-14	0
1	Andaman and Nicobar Islands	2001	Causes	Bankruptcy or Sudden change in Economic	Female	0-14	0
2	Andaman and Nicobar Islands	2001	Causes	Cancellation/Non-Settlement of Marriage	Female	0-14	0
3	Andaman and Nicobar Islands	2001	Causes	Physical Abuse (Rape/Incest Etc.)	Female	0-14	0
4	Andaman and Nicobar Islands	2001	Causes	Dowry Dispute	Female	0-14	0

Which age groups are most effected?

In [10]:
```
age_df=modified_df[['Age_group','Gender','Total']]
```

In [11]:
```
#removing 0-100+ data as it isn't needed for the analysis
age_df=age_df[~age_df['Age_group'].str.contains('0-100+')]
```

In [12]:
```
plt.figure(figsize=(10,10))
age_grp=age_df.groupby(['Age_group','Gender'],as_index=False).sum().sort_values('T
sns.barplot(x='Age_group',y='Total',hue='Gender',data=age_grp,palette='ocean_r')
plt.xlabel("Age Group",fontsize=16)
plt.ylabel("Frequency",fontsize=16)
plt.title("Suicide Rates w.r.t Age Groups",fontsize=21)
plt.tight_layout()
```



Insights:

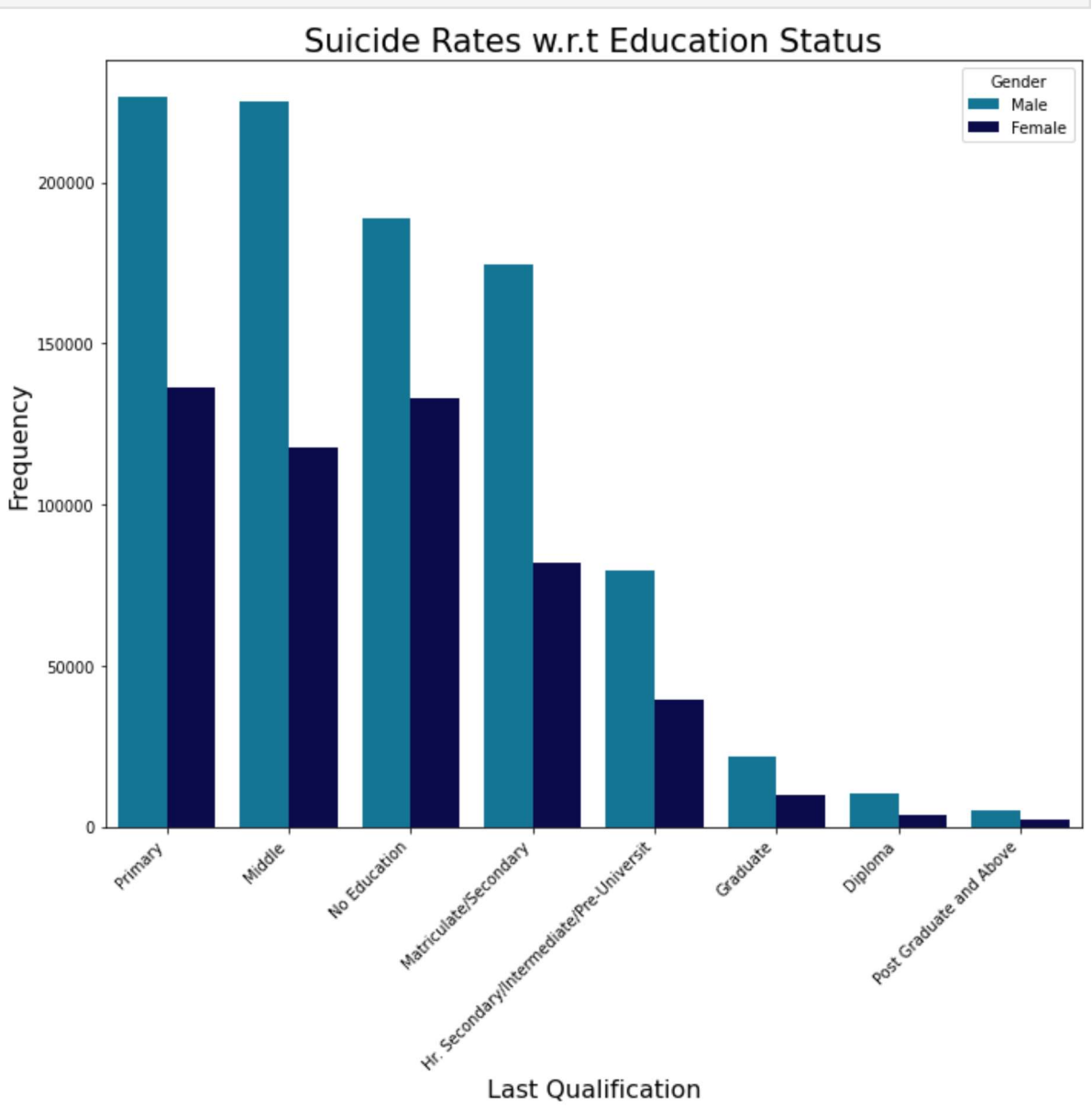
Most suicide rates are seen between the age of 15-44
Mens peaks at age 30-44 whereas Women peaks at age 15-29
ages of 0-14 are the lowest as expected
</p>

How does education play a role in this situation?

In [13]:
```
edu_filt=modified_df['Type_code']=='Education_Status'
```

In [14]:
```
education_df=modified_df[edu_filt][['Type','Gender','Total']]
```

In [15]:
```
plt.figure(figsize=(10,10))
edu_grp=education_df.groupby(['Type','Gender'],as_index=False).sum().sort_values('
sns.barplot(x='Type',y='Total',hue='Gender',data=edu_grp,palette='ocean_r')
plt.xticks(rotation=45,ha='right')
plt.xlabel("Last Qualification",fontsize=16)
plt.ylabel("Frequency",fontsize=16)
plt.title("Suicide Rates w.r.t Education Status",fontsize=21)
plt.tight_layout()
```



Insights:

There is a clear trend as a person gets more educated the lesser are the suicide rates
It should be noted that education may not directly correlate with education as being educated one generally acquires
a stable career and being able to get educated to higher degrees shows they were financially sound from the start. </p>

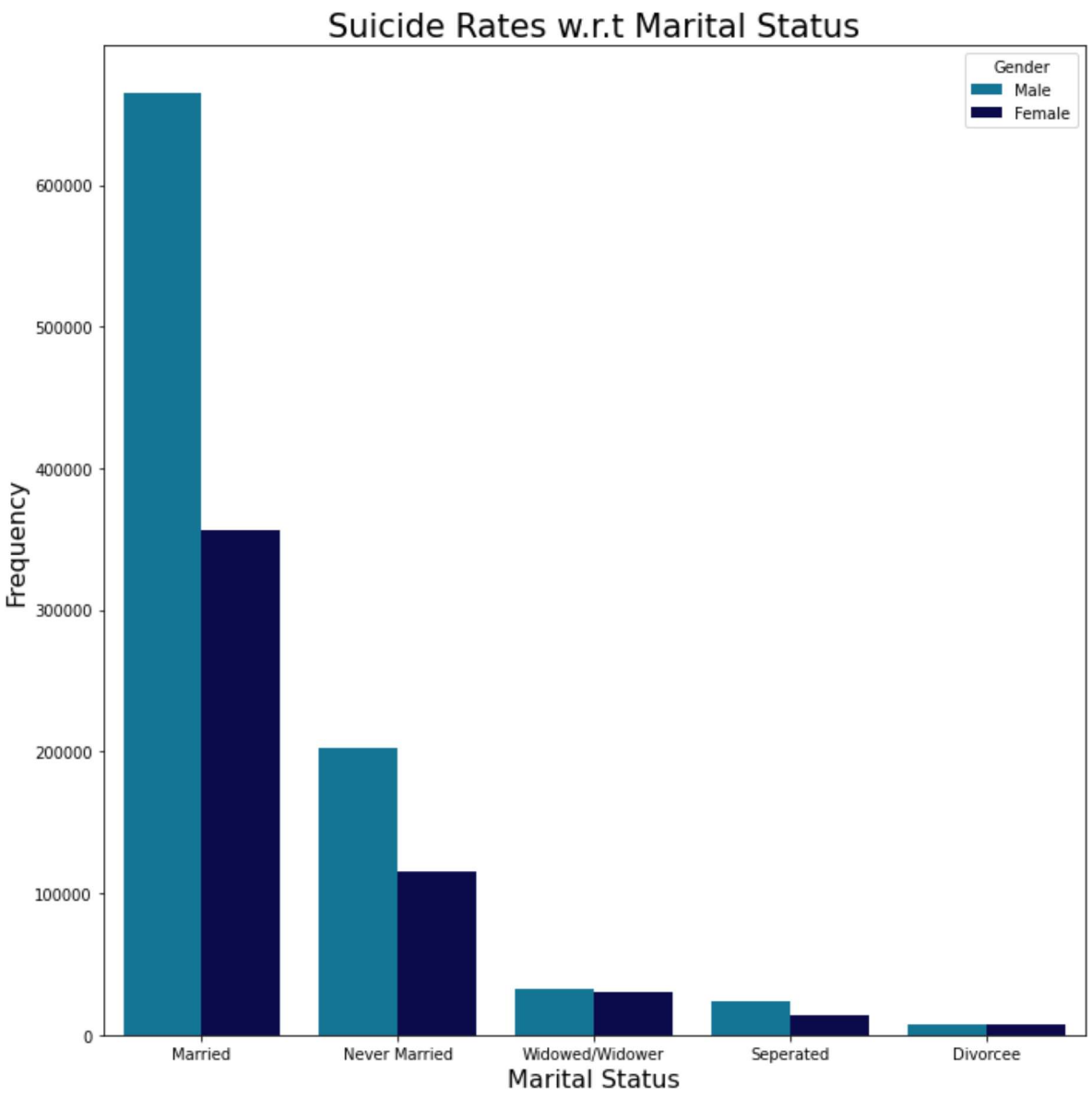
How does Marital Status Effects this situatiion?

In [16]:
```
agefilt=modified_df['Type_code']=='Social_Status'
```

In [17]:
```
social_df=modified_df[agefilt][['Type','Gender','Total']]
```

In [18]:
```
plt.figure(figsize=(10,10))
social_grp=social_df.groupby(['Type','Gender'],as_index=False).sum().sort_values('
sns.barplot(x='Type',y='Total',hue='Gender',data=social_grp,palette='ocean_r')
plt.xlabel("Marital Status",fontsize=16)
plt.ylabel("Frequency",fontsize=16)
```

```
plt.title("Suicide Rates w.r.t Marital Status",fontsize=21)
plt.tight_layout()
```



Insights:

There is a considerable difference between persons who are married and all the rest
It can be concluded that people staying in unhealthy marriages end up being unsatisfied with their lives
This also makes sense as according to Indian culture it is uncommon for divorces to happen forcing this couples
to stay together.

What are the other causes which are observed?

In [19]:
```
causefilt=modified_df['Type_code']=='Causes'
```

In [20]:
```
cause_df=modified_df[causefilt][['Type','Gender','Total']]
```

In [21]:
```
#### Cleaning Cause data
```

In [22]:
```
temp=cause_df['Type'].replace(to_replace="Bankruptcy or Sudden change in Economic"
temp=temp.replace(to_replace="Other Prolonged Illness", value="Prolonged Illness")
temp=temp.str.replace(r"\(.*\)","")
temp=temp.replace(to_replace="Illness ", value="Aids/STD")
temp=temp.replace(to_replace="Physical Abuse ", value="Rape and Abuse")
temp=temp.replace(to_replace="Dowry Dispute", value="Dowry")
temp=temp.replace(to_replace="Fall in Social Reputation", value="deteriorating Soci
temp=temp.replace(to_replace="Not having Children(Barrenness/Impotency", value="Not
temp=temp.replace(to_replace="Not having Children (Barrenness/Impotency", value="No
```

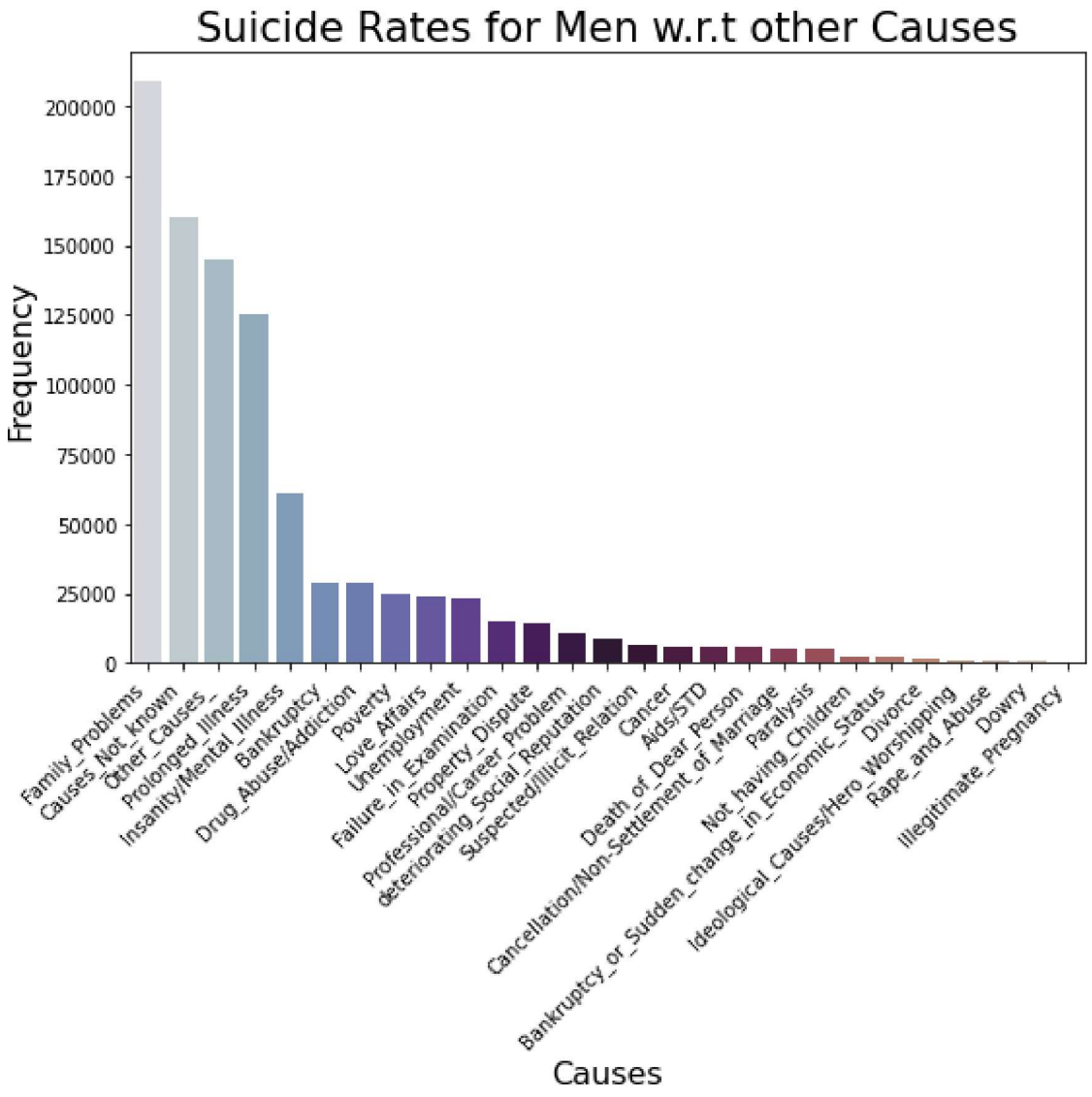
In [23]:
```
cause_df['Type']=temp
```

In [24]:
```
maleFilt=cause_df['Gender']=="Male"
femaleFilt=cause_df['Gender']=="Female"
causeMale=cause_df[maleFilt][['Type','Total']]
causeFemale=cause_df[femaleFilt][['Type','Total']]
```

In [25]:
```
space_series1=causeMale['Type'].str.replace(' ','_')
space_series2=causeFemale['Type'].str.replace(' ','_')
```

In [26]:
```
causeMale['Type']=space_series1
causeFemale['Type']=space_series2
```

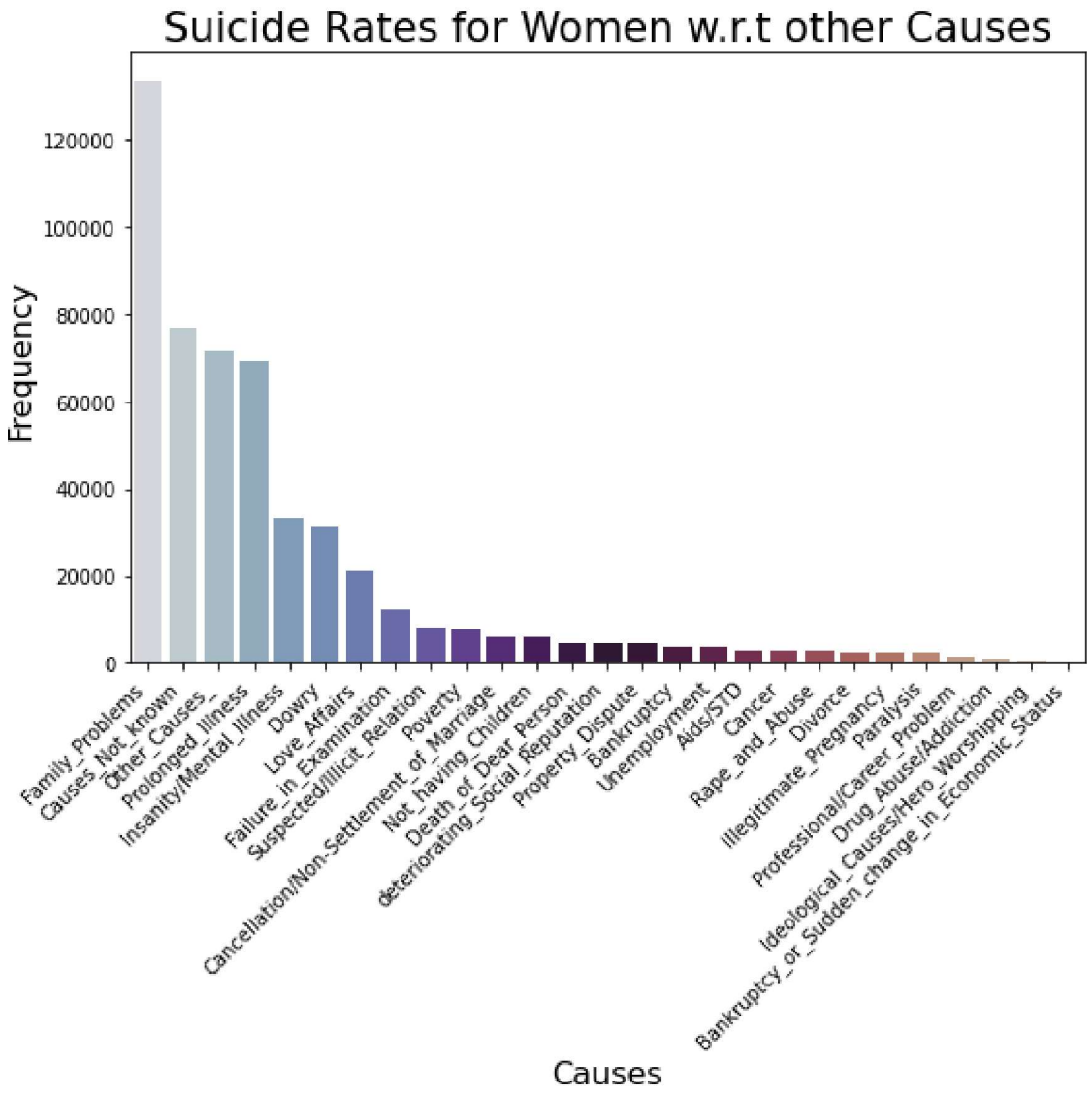
In [27]:
```
plt.figure(figsize=(8,8))
causeMale_grp=causeMale.groupby(['Type'],as_index=False).sum().sort_values('Total'
sns.barplot(x='Type',y='Total',data=causeMale_grp,palette='twilight')
plt.xticks(rotation=45,ha='right')
plt.xlabel("Causes",fontsize=16)
plt.ylabel("Frequency",fontsize=16)
plt.title("Suicide Rates for Men w.r.t other Causes",fontsize=21)
plt.tight_layout()
```



Insights:

Top causes for men to commit suicide are family problems and problems which they aren't comfortable sharing
Bankruptcy is also a big cause for men which isn't the case for women. The reason may be that men are generally the
sole bread winners for their families atleast in rural and orthodox families which increases the stress factor.

In [28]:
```
plt.figure(figsize=(8,8))
causeFemale_grp=causeFemale.groupby(['Type'],as_index=False).sum().sort_values('To
sns.barplot(x='Type',y='Total',data=causeFemale_grp,palette='twilight')
plt.xticks(rotation=45,ha='right')
plt.xlabel("Causes",fontsize=16)
plt.ylabel("Frequency",fontsize=16)
plt.title("Suicide Rates for Women w.r.t other Causes",fontsize=21)
plt.tight_layout()
```



Insights:

Top causes for women to commit suicide are family problems and problems which they aren't comfortable sharing
Dowry is also a big cause for women which isn't the case for men. The reason may be that in rural india malpractices
like dowry still happed causing stress and abuse to the women.

State Wise Analysis

In [29]:
```
state_df=modified_df[['State','Year','Total']]
```

Top 4 States

In [30]:
```
state_top=state_df[['State','Total']]
statetop_grp=state_top.groupby('State')
statetop_grp.sum().sort_values(by='Total',ascending=False).head()
```

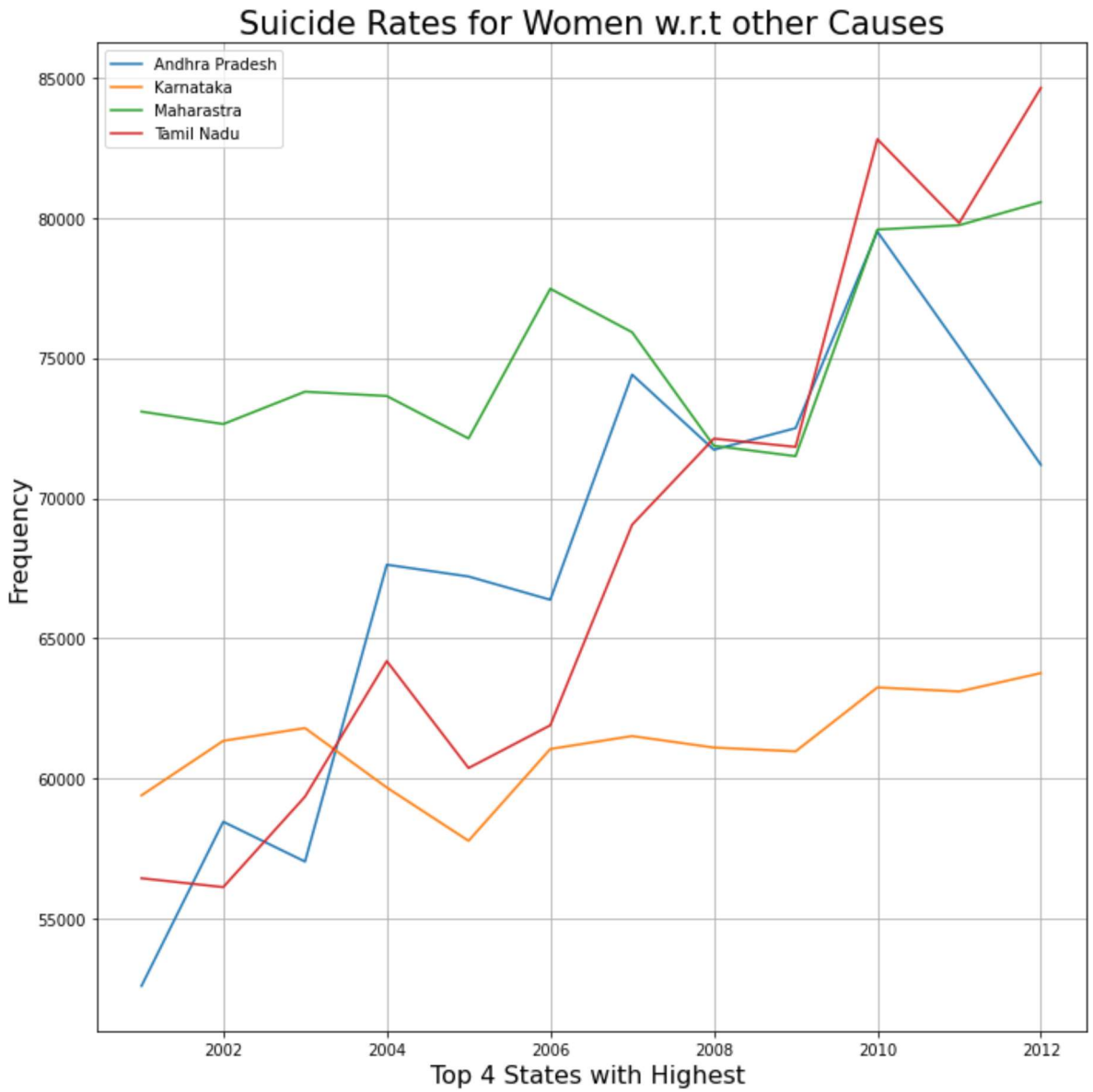
Out[30]:

State	Total
Maharashtra	901945
West Bengal	849936
Tamil Nadu	818691
Andhra Pradesh	814059
Karnataka	734825

In [31]:
```
topstatefilt=state_df['State'].eq("Maharashtra") | state_df['State'].eq("Tamil Nadu
stateTopFour=state_df[topstatefilt]
state_grp=stateTopFour.groupby(['State','Year'])
state_grp.sum().sort_values('Total',ascending=False)
stateGrouped=state_grp.sum()
```

In [32]:
```
x=stateTopFour['Year'].unique()
y_andhra=stateGrouped['Total'].iloc[0:12]
y_karna=stateGrouped['Total'].iloc[12:24]
y_maha=stateGrouped['Total'].iloc[24:36]
y_tamil=stateGrouped['Total'].iloc[36:48]
```

In [33]:
```
plt.figure(figsize=(10,10))
plt.plot(x,y_andhra,label="Andhra Pradesh")
plt.plot(x,y_karna,label="Karnataka")
plt.plot(x,y_maha,label="Maharastra")
plt.plot(x,y_tamil,label="Tamil Nadu")
plt.legend()
plt.grid()
plt.xlabel("Top 4 States with Highest ",fontsize=16)
plt.ylabel("Frequency",fontsize=16)
plt.title("Suicide Rates for Women w.r.t other Causes",fontsize=21)
plt.tight_layout()
```



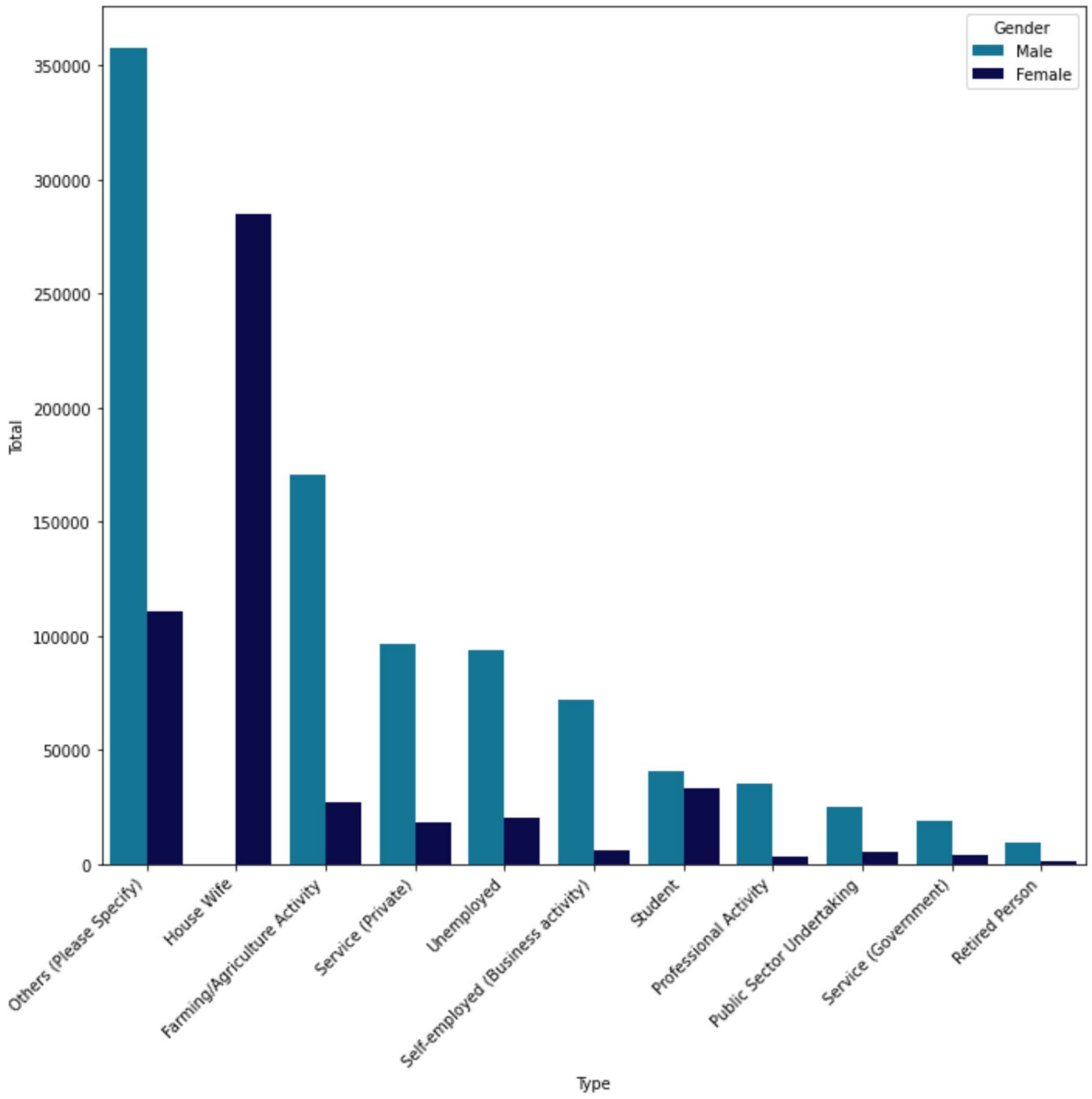
Insights:

The timespan of the data is observed from 2001-2012
Though with govenment sanctions and modernization of society I may had expected for the suicide rates with time but
on the contrary it is seen that the rates has increased. This can be due to less awareness and not normalizing
mental health treatment.

How does Career Status Effects this situatiion?

In [34]:
```
profFilt=modified_df['Type_code']=='Professional_Profile'
profession_df=modified_df[profFilt][['Type','Gender','Total']]
```

In [35]:
```
plt.figure(figsize=(10,10))
prof_grp=profession_df.groupby(['Type','Gender'],as_index=False).sum().sort_values
sns.barplot(x='Type',y='Total',hue='Gender',data=prof_grp,palette='ocean_r')
plt.xticks(rotation=45,ha='right')
plt.tight_layout()
```



Insights:

There are overwhelming number of women who are HouseWifes who end up attempting suicide. Though the data is before
the internet became mainstream, and thus this trend can be expected to even out cause of women empowerment which can only
be validated with more recent data.
Farmers comiting suicide is a well known problem in India and the obsevations validates that.
Men are much more likely to not disclose their professional status. Not being satisfied with their career
leading to commiting suicide seems to be a common trend in men

Conclusion:

- Compiling all the observations it is unfortunate to observe that with time suicide rates are on the rise and globalization and modernization most likely increased stress in the general public. The reason behind this stress can be the increase of competition and the social pressure of not being good enough.
- Pre-teen and teenagers have the lowest suicide rates, where girls have a general high spike in this age gap then usual with prominent reasons being of sexual abuse and academic stress.
- Boys and men are seen having the most cases at the age between 14-44. Most of the major life events like marriage, having kids, graduation, building a career occur in this timespan and men feel the stress of society and family depending on him to keep the family going is a major stress factor. Not being satisfied with salary is seen thus observing causes like bankrupcy, poverty and other money related problems
- Farmer suicide is also a big factor and the steps taken by the government is clearly isn't enough.
- Unhappy Mariage has came up as a big factor for both gender to commit suicide.

Actions:

- Spearding awareness of normalizing divorce is a big step to prevent suicides for both genders. Unhappy marriages/family is the reason for more than 50% people to commit suicide. This will be a big step and deciding factor of suicide prevention.
- Farmers suicide is a major topic which needs to be addressed. Money is a major problem which can be address by making the loan sanction and repayment process easy. It must also be arranged that the famers get a justified marketplace where they can sell their crops at a respectable price.
- Men suicide rates can be drastically reduced by normalizing the treatment of mental health issues and not treating the topic as a taboo. Though difficult for a third-world country like India but having a unemployment benefits where the eligible unemployed people can get a minimum salary will relieve alot of stress from men and in my asumptions will be much effective then the ration program the govenment currently has, though this assumption can only be validated with more data.
- Women suicide rates are also highly preventable if they get a secured upbringing from childhood. Providing education and empower them to make a great career and choosing their own life patners if they choose to marry will be a big factor to save their lives.

Project by:

Deep Saha Link