Analyzing The Road Accidents In India

Organization of this document

1. Importing the required packages

In [128]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import geopandas as gpd
from shapely.geometry import Point, Polygon
import gmplot
```

2. About the dataset

> This dataset contains information of road accidents from 2000-2012.

2.1 Loading the dataset

In [130]:

```
data = pd.read_excel('DATA/datafile.xls')
data.head()
```

Out[130]:

	States	Year	CAUSE	Male	Female	Total	Lat	Long
0	ANDHRA PRADESH	2001	Truck/Lorry (Government)	103	26	129	14.750429	78.570026
1	ANDHRA PRADESH	2001	Truck/Lorry (Private)	2171	414	2585	14.750429	78.570026
2	ANDHRA PRADESH	2001	Bus (Government)	488	119	607	14.750429	78.570026
3	ANDHRA PRADESH	2001	Bus (Private)	289	76	365	14.750429	78.570026
4	ANDHRA PRADESH	2001	Tempo/Vans (Government)	13	3	16	14.750429	78.570026

2.2 Objective

> Analyzing the Road accidents in our country.

3. Understanding Various High Level Information About The Dataset

In [3]:

```
print("Shape of the dataset   :",data.shape)
print("\nColumns in the dataset :",list(data.columns))
print("\n\nMissing Values         : No missing values\n")
print(data.isnull().sum())
print("\nOther information      :\n")
_info=pd.DataFrame(data.describe().T)
_info
```

```
Shape of the dataset   : (6300, 8)

Columns in the dataset : ['States', 'Year', 'CAUSE', 'Male', 'Femal
e', 'Total', 'Lat', 'Long']


Missing Values         : No missing values

States    0
Year      0
CAUSE     0
Male      0
Female    0
Total     0
Lat       0
Long      0
dtype: int64

Other information      :
```

Out[3]:

	count	mean	std	min	25%	50%	7
Year	6300.0	2006.500000	3.452327	2001.000000	2003.750000	2006.500000	2009.250(
Male	6300.0	176.779048	392.862803	0.000000	1.000000	15.000000	149.000(
Female	6300.0	31.743810	67.167001	0.000000	0.000000	3.000000	27.000(
Total	6300.0	208.522857	455.365286	0.000000	1.000000	18.000000	175.000(
Lat	6300.0	22.579335	6.603272	8.900373	19.250232	23.800393	27.333
Long	6300.0	81.618272	7.628217	71.192400	75.980003	78.050006	88.616

4. Segregating the data by states

4.1 Getting list of states

In [4]:

```
states_list=list(data['States'].unique())
print(states_list)
```

```
['ANDHRA PRADESH', 'ARUNACHAL PRADESH', 'ASSAM', 'BIHAR', 'CHHATTIS
GARH', 'GOA', 'GUJARAT', 'HARYANA', 'HIMACHAL PRADESH', 'JAMMU & KA
SHMIR', 'JHARKHAND', 'KARNATAKA', 'KERALA', 'MADHYA PRADESH', 'MAHA
RASHTRA', 'MANIPUR', 'MEGHALAYA', 'MIZORAM', 'NAGALAND', 'ODISHA',
'PUNJAB', 'RAJASTHAN', 'SIKKIM', 'TAMIL NADU', 'TRIPURA', 'UTTAR PR
ADESH', 'UTTARAKHAND', 'WEST BENGAL', 'Andaman And Nicobar', 'CHAND
IGARH', 'Dadra And Nagar Haveli', 'DAMAN & DIU', 'Delhi', 'LAKSHADW
EEP', 'PUDUCHERRY']
```

4.2 Making new dataframes for each state

In [5]:

```
Andhra_pradesh=data[data.States=='ANDHRA PRADESH']
Arunachal_pradesh=data[data.States=='ARUNACHAL PRADESH']
Assam=data[data.States=='ASSAM']
Bihar=data[data.States=='BIHAR']
Chhattisgarh=data[data.States=='CHHATTISGARH']
Goa=data[data.States=='GOA']
Gujarat=data[data.States=='GUJARAT']
haryana=data[data.States=='HARYANA']
Himachal_pradesh=data[data.States=='HIMACHAL PRADESH']
Jammu_kashmir=data[data.States=='JAMMU & KASHMIR']
Jharkhand=data[data.States=='JHARKHAND']
Karnataka=data[data.States=='KARNATAKA']
Kerala=data[data.States=='KERALA']
Madhya_pradesh=data[data.States=='MADHYA PRADESH']
Maharasthra=data[data.States=='MAHARASHTRA']
Manipur=data[data.States=='MANIPUR']
Meghalaya=data[data.States=='MEGHALAYA']
Mizoram=data[data.States=='MIZORAM']
Nagaland=data[data.States=='NAGALAND']
Odisha=data[data.States=='ODISHA']
Punjab=data[data.States=='PUNJAB']
Rajasthan=data[data.States=='Rajasthan']
Sikkim=data[data.States=='SIKKIM']
Tamil_nadu=data[data.States=='TAMIL NADU']
Tripura=data[data.States=='TRIPURA']
Uttar_pradesh=data[data.States=='UTTAR PRADESH']
Uttarakhand=data[data.States=='UTTARAKHAND']
West_bengal=data[data.States=='WEST BENGAL']
Andaman_and_nicobar=data[data.States=='Andaman And Nicobar']
Chandigarh=data[data.States=='CHANDIGARH']
Dadra_and_nagar_haveli=data[data.States=='Dadra And Nagar Haveli']
Daman_and_diu=data[data.States=='DAMAN & DIU']
Delhi=data[data.States=='Delhi']
Lakshadweep=data[data.States=='LAKSHADWEEP']
Pudducherry=data[data.States=='PUDDUCHERRY']
```

5. Analysis on the data

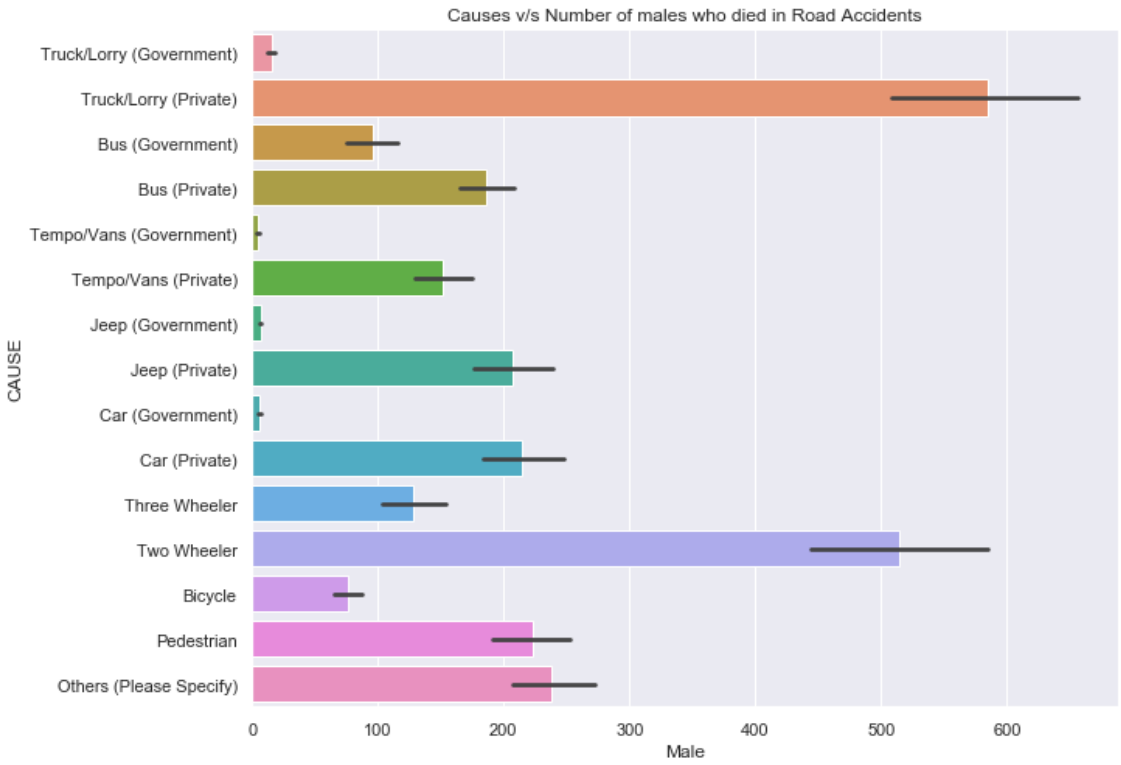
5.1 Number of males who died in the road accidents with their causes

In [60]:

```
sns.set(rc={'figure.figsize':(10,8)})
sns.barplot(x="Male", y="CAUSE", data=data)
plt.title("Causes v/s Number of males who died in Road Accidents")
```

Out[60]:

```
Text(0.5, 1.0, 'Causes v/s Number of males who died in Road Acciden
ts')
```



- Most of the people died by Truck/Lorry and Two wheeler.
- Least number of people died by Tempo/Vans.
- Other than the Bus, No other government vehicles caused a lot of death.

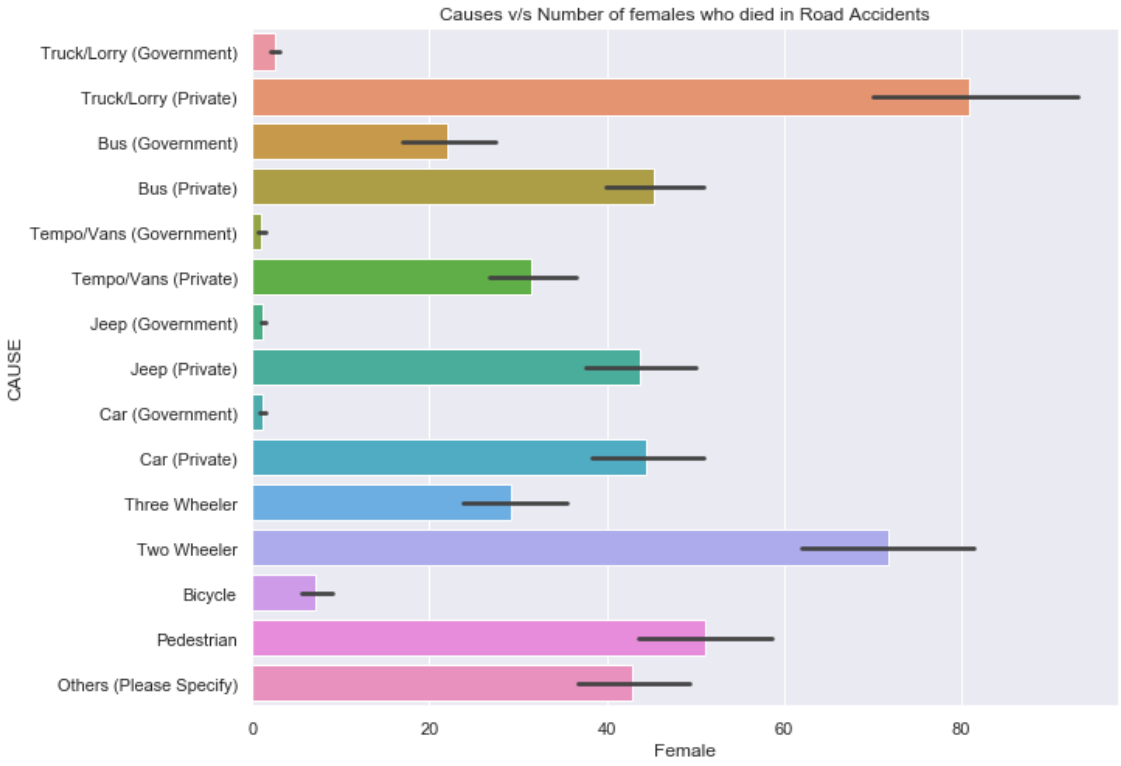
5.2 Number of females who died in the road accidents with their causes

In [66]:

```
sns.set(rc={'figure.figsize':(10,8)})
sns.barplot(x="Female", y="CAUSE", data=data)
plt.title("Causes v/s Number of females who died in Road Accidents")
```

Out[66]:

```
Text(0.5, 1.0, 'Causes v/s Number of females who died in Road Accid
ents')
```



- Relativelty there were more female deaths by private bus.
- Observations made on male deaths satisfy in here too.

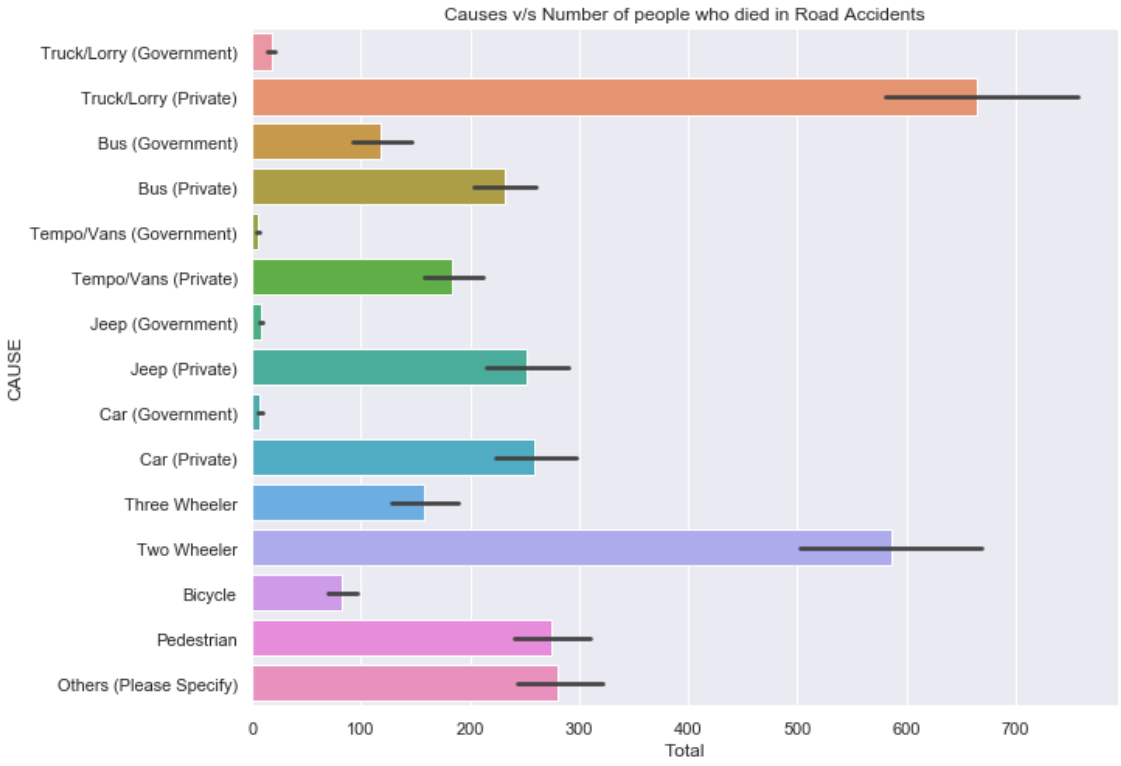
5.3 Both male and female who died in the road accidents with their causes

In [68]:

```
sns.set(rc={'figure.figsize':(10,8)})
sns.barplot(x="Total", y="CAUSE", data=data)
plt.title("Causes v/s Number of people who died in Road Accidents")
```

Out[68]:

```
Text(0.5, 1.0, 'Causes v/s Number of people who died in Road Accide
nts')
```



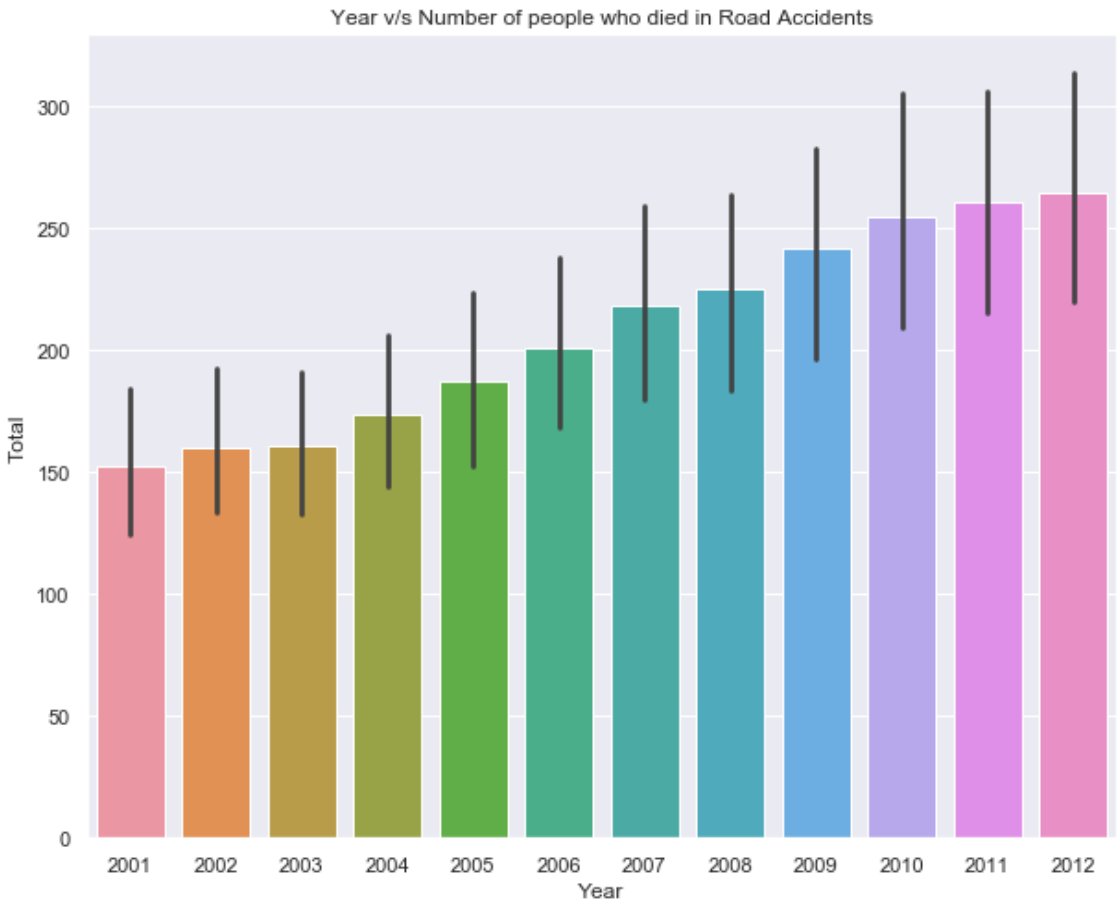
5.4 Number of people died in road accidents in each year

In [72]:

```
sns.barplot(x="Year", y="Total", data=data)
plt.title("Year v/s Number of people who died in Road Accidents")
```

Out[72]:

```
Text(0.5, 1.0, 'Year v/s Number of people who died in Road Accident
s')
```



- Most of the deaths happened in the year 2012.
- Least number of people died in the year 2001.

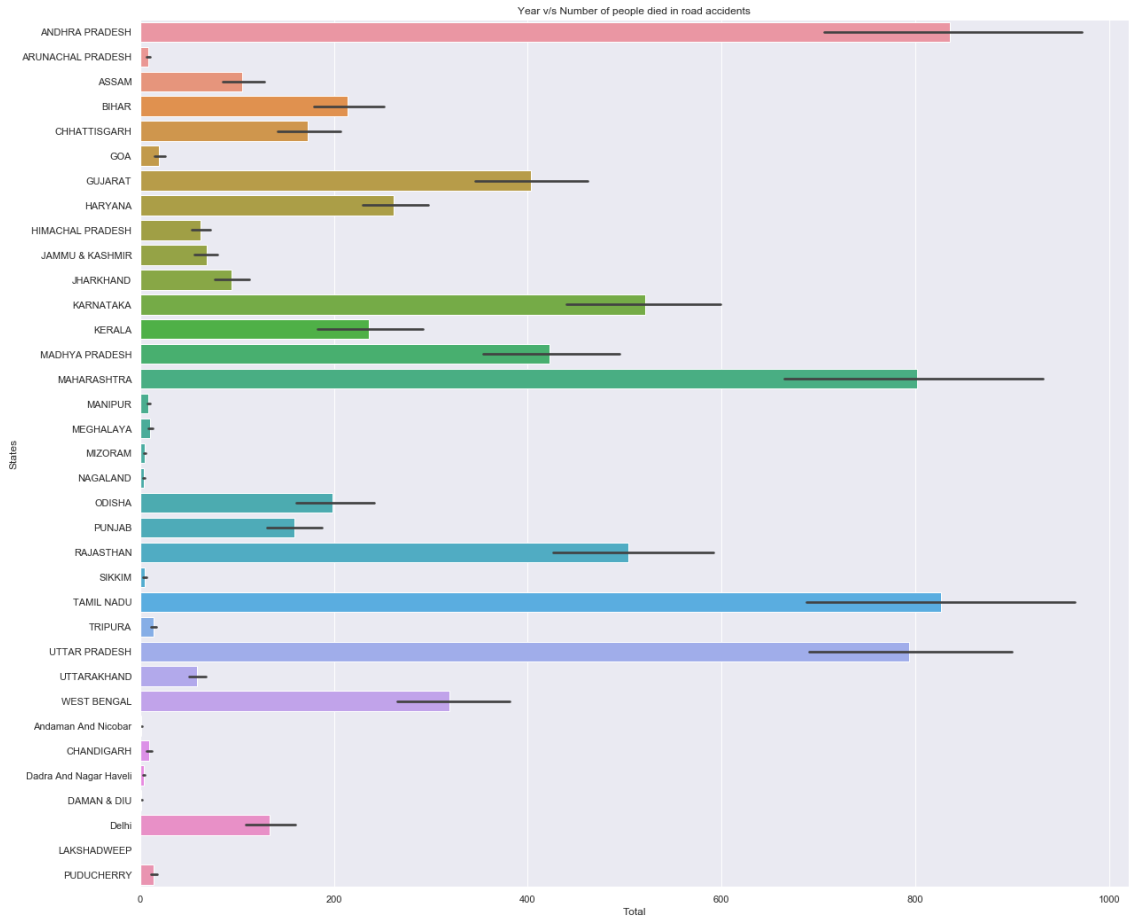
5.5 States v/s Number of people died in road accidents

In [126]:

```
sns.set(rc={'figure.figsize':(20,18)})
sns.barplot(x="Total", y="States", data=data)
plt.title("Year v/s Number of people died in road accidents")
```

Out[126]:

```
Text(0.5, 1.0, 'Year v/s Number of people died in road accidents')
```



- There are huge number of people from Andhra Pradesh who died in road accidents.
- And followed by Maharasthra, Tamilnadu and Uttar pradesh.
- Less number of people died in Nagaland,Mizoram,Lakshwadeep,Daman and diu, Andaman and nicobar,Sikkim,Meghalaya,Manipur,Goa and Arunachal Pradesh

5.6 Heatmap for various features

In [140]:

```
data_cor=data.corr()
sns.set(rc={'figure.figsize':(10,8)})
sns.heatmap(data_cor,annot=True)
```

Out[140]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f03b1103358>
```

- The features **Year** and **Number of people died** are positively correlated.

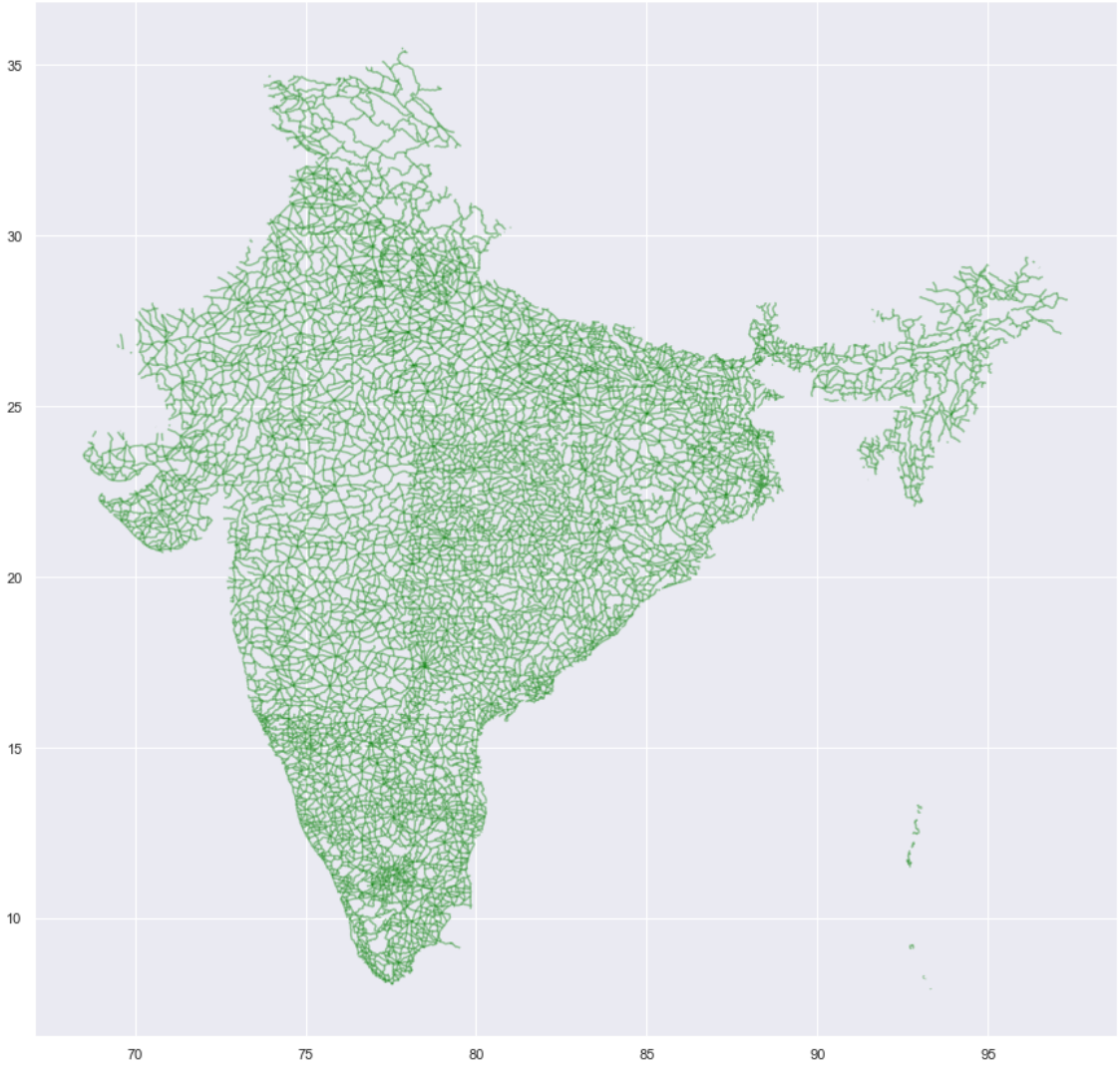
5.7 Creating a map of indian roads and plotting the data

In [121]:

```
# Building an indian roads map from shape file
PATH='/home/chiranjeevi_karthik/Desktop/Accident analysis/Indian Roads Shape FIles/IND_rds/IND_roads.shp'
indian_roads = gpd.read_file(PATH)
crs={'init':'espg:7755'}
cords= [Point(xy) for xy in zip(data['Lat'],data['Long'])]
gdf=gpd.GeoDataFrame(data,crs=crs,geometry=cords)
fig,ax=plt.subplots(figsize=(15,15))
indian_roads.plot(ax=ax,alpha=0.4,color='green')
#gdf[gdf['States']=='ANDHRA PRADESH'].plot(ax=ax,markersize=1,color="red",label="AP")
```

Out[121]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f03b6e8e080>
```



6. Statewise Analysis of Road Accidents

6.1 Andhra pradesh

6.1.1 Number of Males who died every year in Andhra pradesh Road Accidents

In [6]:

```
sns.set(rc={'figure.figsize':(12,8)})
sns.lineplot(x="Year", y="Male",data=Andhra_pradesh)
plt.title("Number of Males who died every year in Andhra pradesh")
```

Out[6]:

```
Text(0.5, 1.0, 'Number of Males who died every year in Andhra prade
sh')
```

- The Number of Males who died in Andhra pradesh linealy increased from 2002 to 2012.

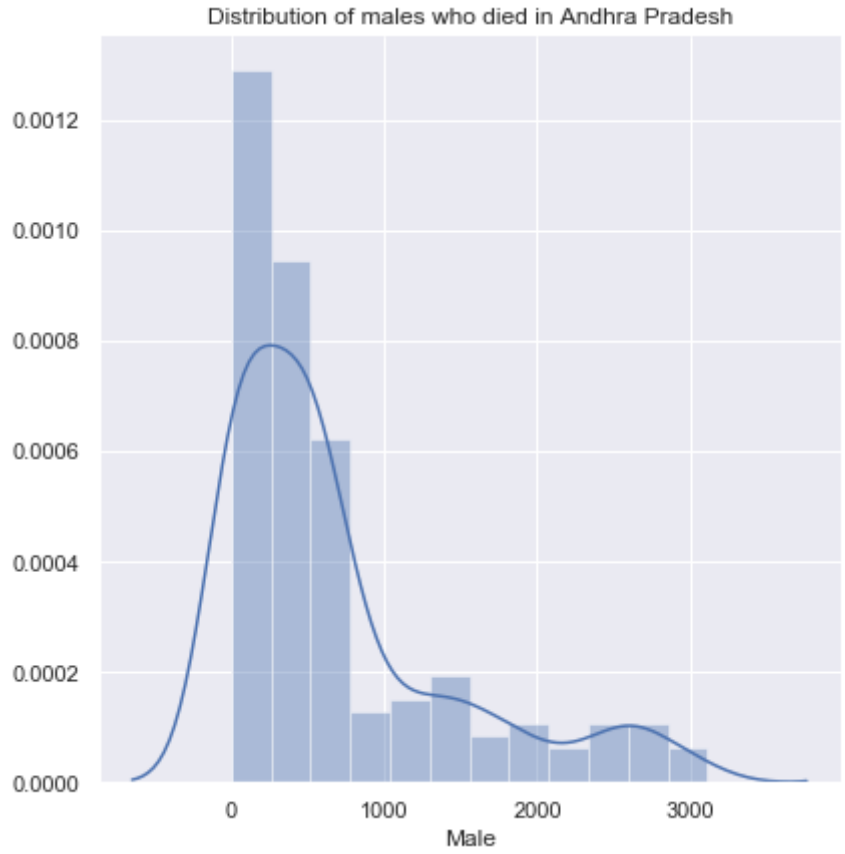
6.1.2 Distribution of males who died in Andhra Pradesh Road Accidents

In [7]:

```
sns.FacetGrid(Andhra_pradesh, height=6).map(sns.distplot, "Male").add_legend()
plt.title("Distribution of males who died in Andhra Pradesh")
```

Out[7]:

```
Text(0.5, 1, 'Distribution of males who died in Andhra Pradesh')
```



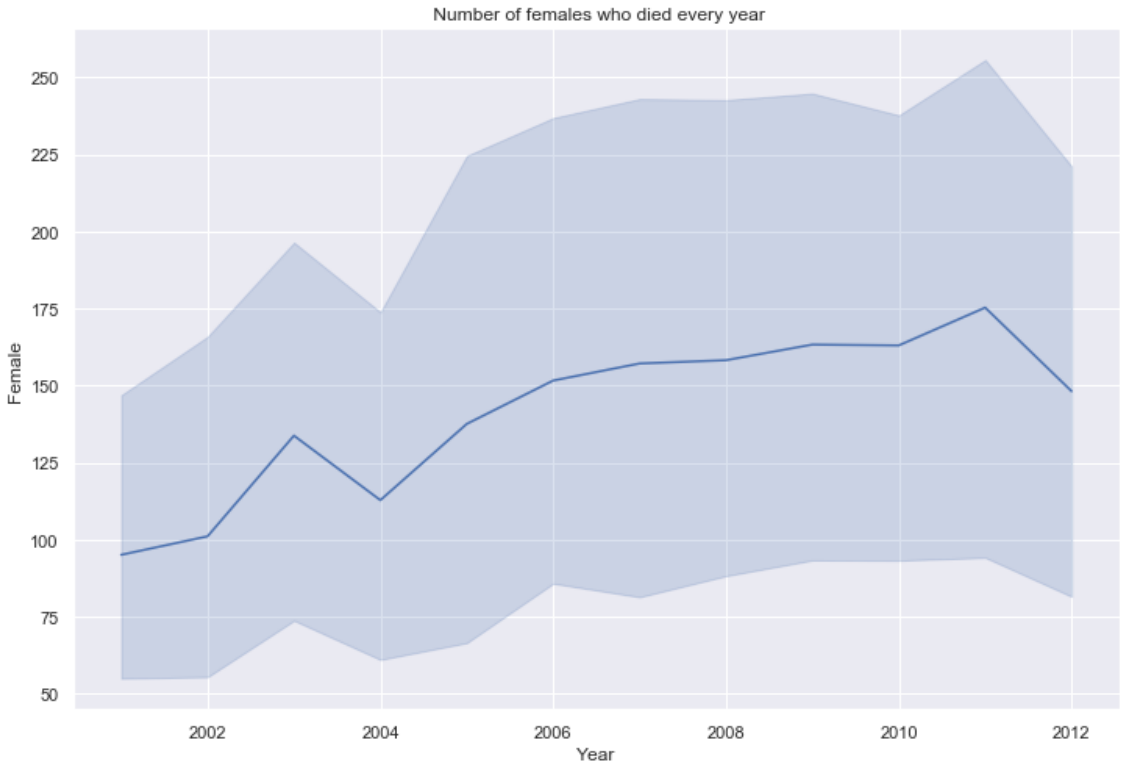
6.1.3 Number of females who died every year in Andhra Pradesh Road Accidents

In [8]:

```
sns.set(rc={'figure.figsize':(12,8)})
sns.lineplot(x="Year", y="Female",data=Andhra_pradesh)
plt.title(" Number of females who died every year")
```

Out[8]:

```
Text(0.5, 1.0, ' Number of females who died every year')
```



- There are relatively less females died in Andhra pradesh than males.

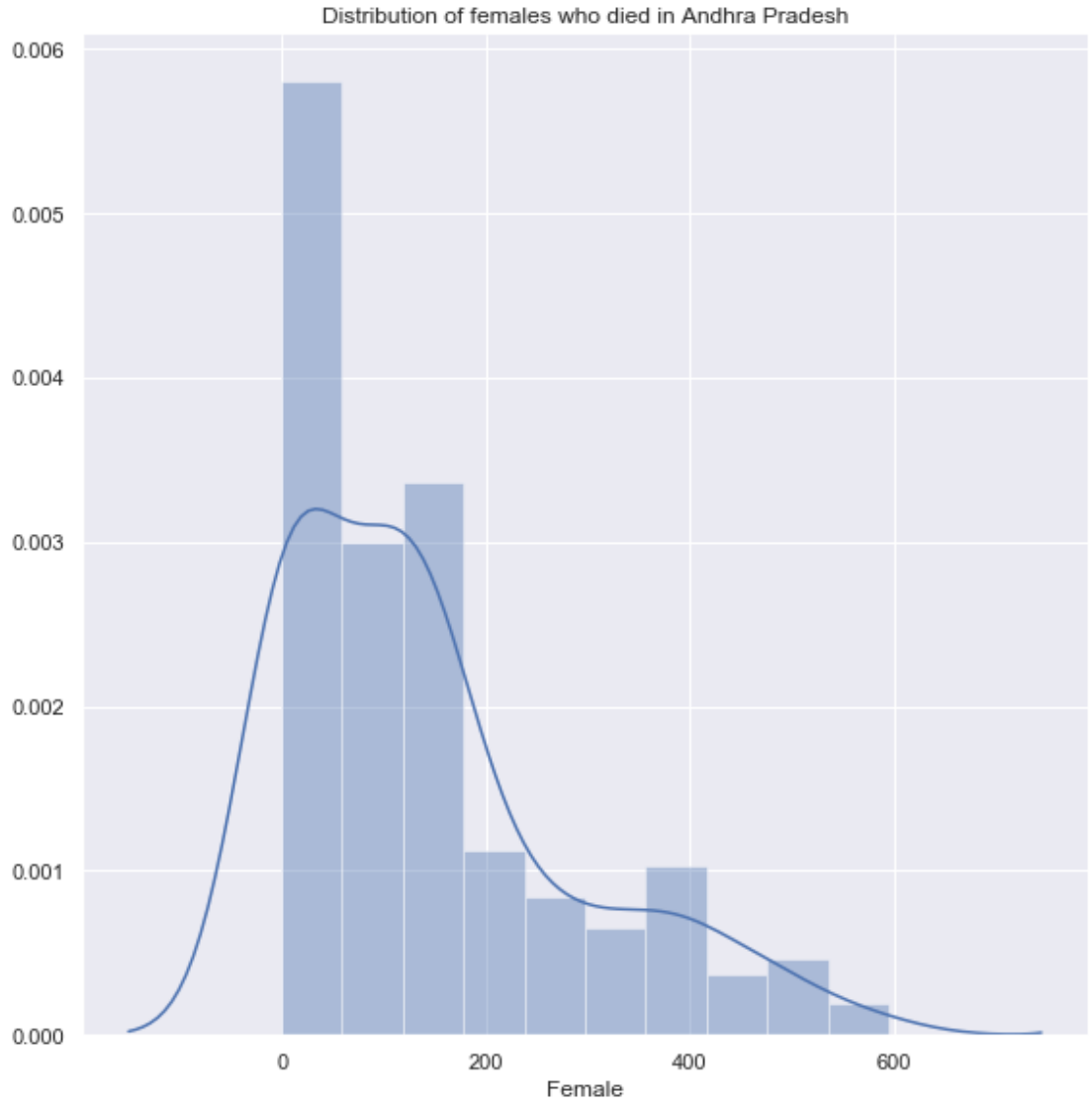
6.1.4 Distribution of females who died in Andhra Pradesh Road Accidents

In [9]:

```
sns.FacetGrid(Andhra_pradesh, height=8).map(sns.distplot, "Female").add_legend()
plt.title("Distribution of females who died in Andhra Pradesh")
```

Out[9]:

```
Text(0.5, 1, 'Distribution of females who died in Andhra Pradesh')
```



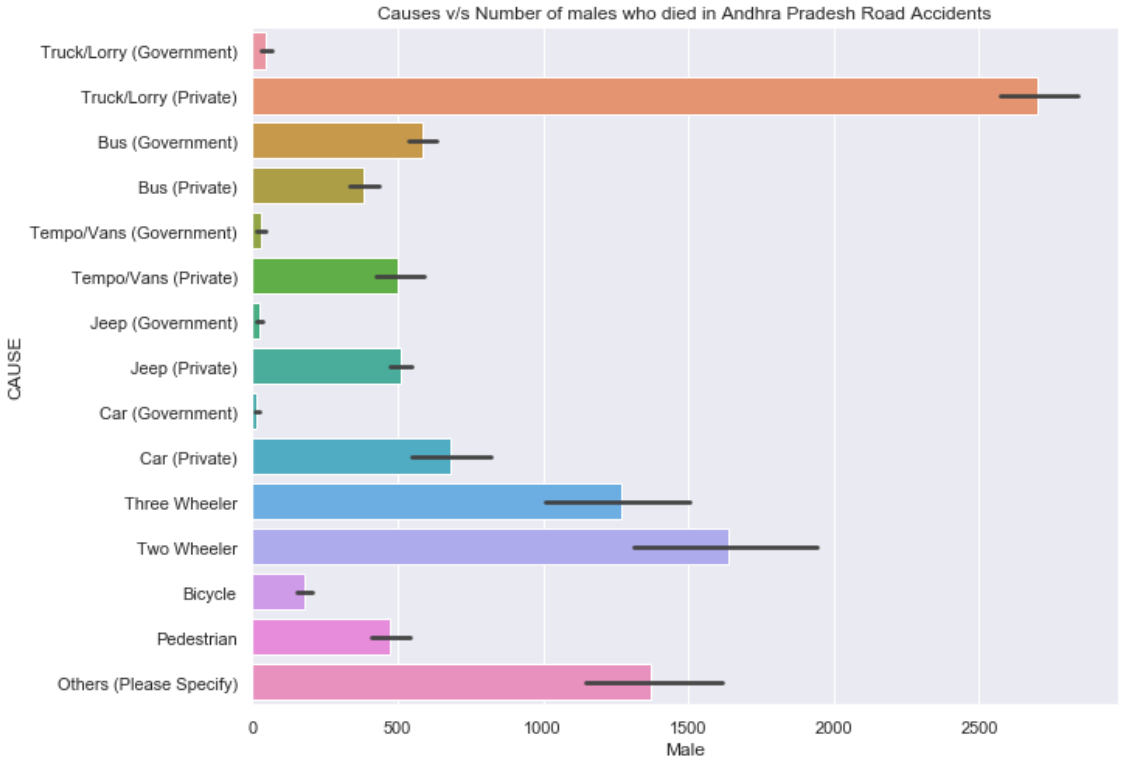
6.1.5 Causes v/s Number of males who died in Andhra Pradesh Road Accidents

In [15]:

```
sns.set(rc={'figure.figsize':(10,8)})
sns.barplot(x="Male", y="CAUSE", data=Andhra_pradesh)
plt.title("Causes v/s Number of males who died in Andhra Pradesh Road Accidents"
)
```

Out[15]:

```
Text(0.5, 1.0, 'Causes v/s Number of males who died in Andhra Prade
sh Road Accidents')
```



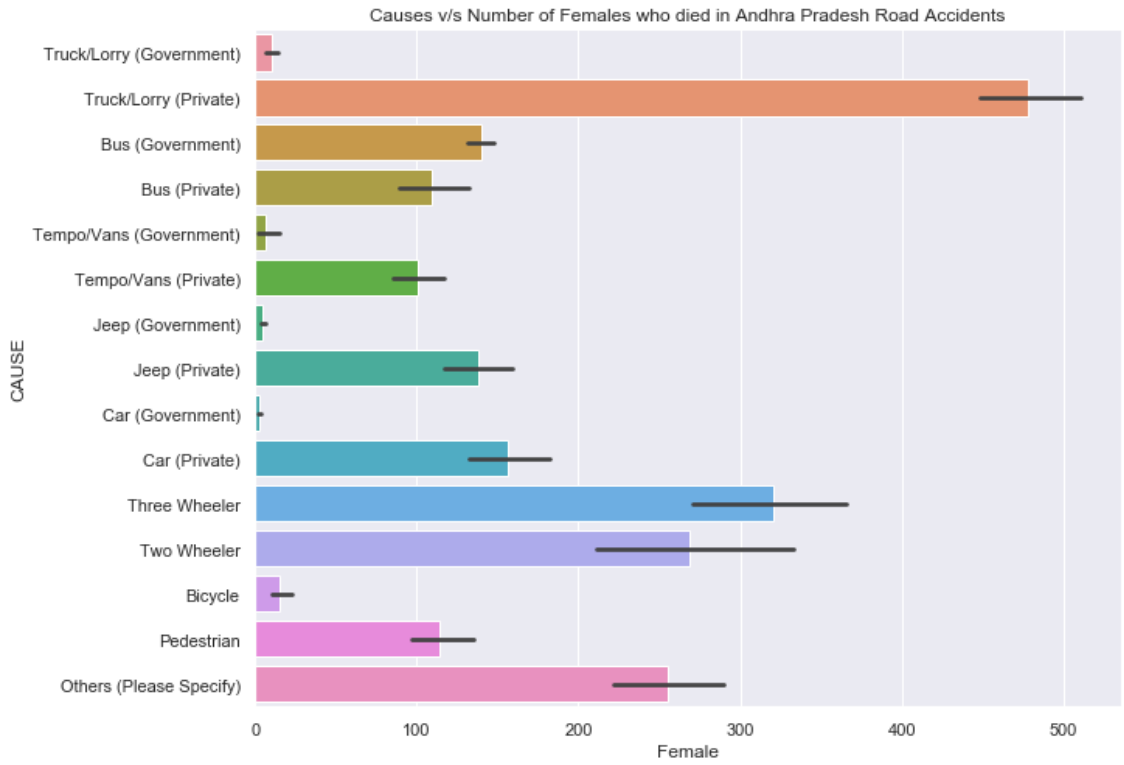
6.1.6 Causes v/s Number of Females who died in Andhra Pradesh Road Accidents

In [19]:

```
sns.set(rc={'figure.figsize':(10,8)})
sns.barplot(x="Female", y="CAUSE", data=Andhra_pradesh)
plt.title("Causes v/s Number of Females who died in Andhra Pradesh Road Accident
s")
```

Out[19]:

```
Text(0.5, 1.0, 'Causes v/s Number of Females who died in Andhra Pra
desh Road Accidents')
```



6.1.7 Map of Accidents For Andhra Pradesh

In [127]:

```
lat=list(Andhra_pradesh['Lat'])
lon=list(Andhra_pradesh['Long'])
gmap = gmplot.GoogleMapPlotter(lat[0], lon[0], 10)
gmap.scatter( lat, lon, '# FF0000', size = 40, marker = True)
gmap.polygon(lat, lon, color = 'cornflowerblue')
gmap.draw( "/home/chiranjeevi_karthik/Desktop/map.html" )
```

7.Conclusions

1. Most of the people died by Truck/Lorry and Two wheeler.
2. Least number of people died by Tempo/Vans.
3. Other than the Bus, No other government vehicles caused a lot of death.
4. Relativelty there were more female deaths by private bus.
5. Most of the deaths happened in the year 2012.
6. Least number of people died in the year 2001.
7. There are huge number of people from Andhra Pradesh who died in road accidents. And followed by Maharasthra, Tamilnadu and Uttar pradesh.
8. Less number of people died in Nagaland,Mizoram,Lakshwadeep,Daman and diu, Andaman and nicobar,Sikkim,Meghalaya,Manipur,Goa and Arunachal Pradesh
9. The features Year and Number of people died are positively correlated.
10. The Number of Males who died in Andhra pradesh linealy increased from 2002 to 2012.
11. There are relatively less females died in Andhra pradesh than males.