

The sea of hopelessness: An exploration of suicide statistics in India.

Name: Sandeep Sethumadhavan

Student ID: 30980283

Tutors: Muhammad (Jameel) Baig and Tam Vo

Introduction:

Whether it be due to pressure, hopelessness or a grave lack of will to survive, for many people suicide is the last straw. It is estimated that suicide is the tenth leading cause of death worldwide and it has rightfully led to a global effort to set up efficient suicide prevention hotlines (Hawton & van Heeringen, 2009). While it is a morbid yet personal decision, it is still possible to explore existing data regarding suicide to help identify possible patterns and trends in order to help prevent, or at least reduce these occurrences in the future.

Being from India and having had my own experiences with depression and having close acquaintances take their own lives, understanding the data behind suicides to help the conversation is a personal mission for me. Through this project, I intend to explore publicly available suicide data for India and visualize my findings.

In my project proposal I identified the following questions that I intended to explore:

1. Has the progress of India as a country over time led to a noticeable decrease in the number of people who have taken their own lives?
2. What have been the major causes of suicide in India and has it changed over time?
3. Is it possible to identify the most vulnerable demography through the lens of socio-economical/educational status?
4. Is there a generational aspect to suicides and what can be understood from the age ranges of the deceased?

In the proposal I also outlined that I would be working with data sources that I have compiled from the years 2013-2018 for the following aspects: professional profile, educational status, and social status of the deceased as well as means adopted and cause for each year.

However, due to issues that I will elaborate in the Data Wrangling section, I modified the time series to be from 2001-2014 and chose to drop the means adopted.

Data Wrangling:

Data compilation:

The data for the project was compiled from various sources and by combining multiple csv files. They include:

Educational status: Each of these data sources carries information about the educational status (level of education) of the deceased for the corresponding years :

2014: <https://data.gov.in/resources/state-wise-educational-status-suicide-victim-during-2014> , 391 rows

2013: <https://data.gov.in/resources/state-wise-educational-status-suicide-victim-during-2013-0> , 343 rows
2001-12: <https://data.gov.in/resources/stateut-wise-educational-status-suicide-victim-during-2001-2012> , 4105 rows

Professional profile: Each of these data sources carries information about the professional profile of the deceased for the corresponding years :

2014: <https://data.gov.in/resources/state-wise-profession-profile-suicide-victim-during-2014> , 976 rows
2013: <https://data.gov.in/resources/state-wise-profession-profile-suicide-victim-during-2013> , 571 rows
2001-12: <https://data.gov.in/resources/stateut-wise-professional-profile-suicide-victim-during-2001-2012> , 6841 rows

Causes: Each of these data sources carries information about the cause of the suicide for the corresponding years :

2014: <https://data.gov.in/resources/state-wise-distribution-suicides-causes-during-2014> , 1210 rows
2013: <https://data.gov.in/resources/state-wise-distribution-suicides-causes-during-2013> , 1065 rows
2001-12: <https://data.gov.in/resources/stateut-wise-distribution-suicides-causes-during-2001-2012> , 12769 rows

Social Status: Each of these data sources carries information about social status(married/unmarried etc) of the deceased for the corresponding years :

2014: <https://data.gov.in/resources/state-wise-social-status-suicide-victim-during-2014> , 313 rows
2013: <https://data.gov.in/resources/state-wise-social-status-suicide-victim-during-2013> , 229 rows
2001-12: <https://data.gov.in/resources/stateut-wise-social-status-suicide-victims-during-2001-2012> , 2737 rows

HDI rankings for India and its states for the period 2001-12 were obtained from:

https://globaldatalab.org/shdi/view/shdi/IND/?levels=1%2B4&interpolation=0&extrapolation=0&nearest_real=0 , 38 rows

The population of India distributed by age group and gender for the period was obtained from:

<https://population.un.org/wpp/DataQuery/>

The population of Indian states distributed by gender for the period was obtained from:

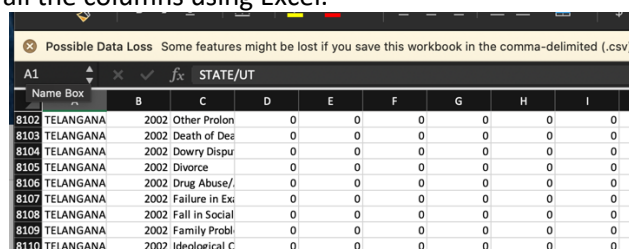
<https://population.un.org/wpp/DataQuery/>

Data Cleaning:

The data cleaning was a laborious process since I had to deal with a large number of data sources. Below I go through the major data cleaning steps that I had to undertake and the tools that I used. (Please note that due to the large number of csv files and datafiles that required cleaning and restructuring, I am highlighting the key and individual examples of an issue rather than documenting multiple examples of the same issues since that would go past the scope of the assignment.)

1. The problem with the state of “Telangana”-

Telangana is a newly formed state in India formed by the splitting of the state of Andhra Pradesh in 2014 (Seshan, 2018) , and hence the census records from the previous years do not include such a state. To combat this issue, for all the CSV files dealing with years 2001-13 I created a new entry called Telangana under the column for states with entries of zero for all the columns using Excel.



	STATE/UT									
8102	TELANGANA	2002 Other Prolon	0	0	0	0	0	0	0	0
8103	TELANGANA	2002 Death of Dea	0	0	0	0	0	0	0	0
8104	TELANGANA	2002 Dowry Dispu	0	0	0	0	0	0	0	0
8105	TELANGANA	2002 Divorce	0	0	0	0	0	0	0	0
8106	TELANGANA	2002 Drug Abuse/	0	0	0	0	0	0	0	0
8107	TELANGANA	2002 Failure in Ex	0	0	0	0	0	0	0	0
8108	TELANGANA	2002 Fall in Social	0	0	0	0	0	0	0	0
8109	TELANGANA	2002 Family Probl	0	0	0	0	0	0	0	0
8110	TELANGANA	2002 Ideological C	0	0	0	0	0	0	0	0

2. Non recognition of transgender people in census:

Indian census data unfortunately didn't recognise transgender people as a part of the

demographic until 2014. So, to help make the data structure uniform to help with the binding, I used R and Excel to encode new columns for the data from 2001-12 and 2013 for transgender people with entries 0.

```

1 #reading csv files for 2001-12,2013,2014
2 Education_2001_12<- read.csv("Educational Status 2001-12.csv", header=T)
3 Education_2013<- read.csv("Educational Status 2013.csv", header=T)
4 Education_2014<- read.csv("Educational Status 2014.csv", header=T)
5
6 #encoding for Transgender people
7 Education_2001_12$Transgender=0
8 Education_2013$Transgender=0
9
10 #reordering the columns
11 Education_2001_12 <- Education_2001_12[, colnames(Education_2001_12)[c(1:5,7,6)]]
12 Education_2013 <- Education_2013[, colnames(Education_2013)[c(1:5,7,6)]]

```

	STATE.UT	Year	CAUSE	Male	Female	Transgender	Total
1	ANDHRA PRADESH	2001	No Education	2117	1657	0	3774
2	ANDHRA PRADESH	2001	Primary	1536	1076	0	2612
3	ANDHRA PRADESH	2001	Middle	1281	722	0	2003
4	ANDHRA PRADESH	2001	Matriculate/Secondary	818	449	0	1267

3. Making the column names uniform for binding purposes:

A major issue I dealt with when binding the CSVs in each group/aspect was the fact that they all had non uniform column names and structures.

An example of this issue is displayed below for the CSV files containing the educational status for 2001-12, 2013 and 2014 respectively.

STATE/UT	Year	CAUSE	Male	Female	Total
ANDHRA PRADESH	2001	No Education	2117	1657	3774
ANDHRA PRADESH	2001	Primary	1536	1076	2612
ANDHRA PRADESH	2001	Middle	1281	722	2003

State/UTs	Year	CAUSE	Male	Female	Total
Total (All India)	2013	No Education	15579	9425	25004
Total (All India)	2013	Primary	19502	10322	29824
Total (All India)	2013	Middle	21679	10095	31774
Total (All India)	2013	Matriculate/Secondary	19416	8180	27596

States/UTs	Year	CrimeHead	Male	Female	Transgender	Total
Andhra Pradesh	2014	No Education	927	470	0	1397
Andhra Pradesh	2014	Primary (upto class-5)	743	385	0	1128
Andhra Pradesh	2014	Middle (upto class-8)	630	343	0	973

This issue was rectified using RStudio, after performing the above-mentioned step of encoding for transgender people.

```

1 #reading csv files for 2001-12,2013,2014
2 Education_2001_12<- read.csv("Educational Status 2001-12.csv", header=T)
3 Education_2013<- read.csv("Educational Status 2013.csv", header=T)
4 Education_2014<- read.csv("Educational Status 2014.csv", header=T)
5
6 #encoding for Transgender people
7 Education_2001_12$Transgender=0
8 Education_2013$Transgender=0
9
10 #reordering the columns
11 Education_2001_12 <- Education_2001_12[, colnames(Education_2001_12)[c(1:5,7,6)]]
12 Education_2013 <- Education_2013[, colnames(Education_2013)[c(1:5,7,6)]]
13
14 #getting column names from Education status 2001-12 dataframe
15 Education_colnames <- c(names(Education_2001_12))
16
17 #setting column names for 2013,2014 dataframes
18 colnames(Education_2013) <- Education_colnames
19 colnames(Education_2014) <- Education_colnames

```

This results in the same column names which can help bind the dataframes easily.

	STATE.UT	Year	CAUSE	Male	Female	Transgender	Total
1	ANDHRA PRADESH	2001	No Education	2117	1657	0	3774

	STATE.UT	Year	CAUSE	Male	Female	Transgender	Total
1	TOTAL (ALL INDIA)	2013	No Education	15579	9425	0	25004

	STATE.UT	Year	CAUSE	Male	Female	Transgender	Total
1	ANDHRA PRADESH	2014	No Education	927	470	0	1397

4. Dealing with erratic spellings of states, causes, Education, social status and professional status:

The different CSV files all had spelling errors and non-uniform encoding which was another issue to deal with. An example of such inconsistencies can be seen below from screengrabs of Educational Status 2001-12 and Educational Status 2014:

JAMMU & KASHMIR	2010	Diploma	2	0	2
JAMMU & KASHMIR	2010	Graduate	4	0	4
JAMMU & KASHMIR	2010	Post Graduate and Above	1	1	2
JAMMU & KASHMIR	2010	Total	122	137	259
JAMMU & KASHMIR	2011	No Education	45	35	80
JAMMU & KASHMIR	2011	Primary	16	27	43
JAMMU & KASHMIR	2011	Middle	28	34	62
JAMMU & KASHMIR	2011	Matriculate/Secondary	32	30	62
JAMMU & KASHMIR	2011	Hr. Secondary/Intermediate/Pre-Universit	11	9	20
JAMMU & KASHMIR	2011	Diploma	5	1	6
JAMMU & KASHMIR	2011	Graduate	4	3	7
JAMMU & KASHMIR	2011	Post Graduate and Above	3	4	7
JAMMU & KASHMIR	2011	Total	144	143	287

Jammu and Kashmir	2014	No Education	34	20	0	54
Jammu and Kashmir	2014	Primary (upto class-5)	17	8	0	25
Jammu and Kashmir	2014	Middle (upto class-8)	32	19	0	51
Jammu and Kashmir	2014	Matriculate/Secondary (upto class-10)	45	19	0	64
Jammu and Kashmir	2014	Higher Secondary/ Intermediate/ Pre-University (upto class-12)	3	17	0	20
Jammu and Kashmir	2014	Diploma/ITI/Certificate	0	0	0	0
Jammu and Kashmir	2014	Graduate and above	1	1	0	2
Jammu and Kashmir	2014	Professionals (MBA; etc.)	0	0	0	0
Jammu and Kashmir	2014	Status not known	18	24	0	42
Jammu and Kashmir	2014	Total	150	100	0	250

As you can see the spelling of Jammu and Kashmir and various educational status are inconsistent. These inconsistencies can be fixed by changing row values using conditions in RStudio as seen below.

```
#changing all the names of states to upper case
Education_2014$STATE.UT <- toupper(Education_2014$STATE.UT)

#changing value of row/name of states based on condition
Education_2014$STATE.UT[Education_2014$STATE.UT== "JAMMU AND KASHMIR"]<- "JAMMU & KASHMIR"
Education_2014$STATE.UT[Education_2014$STATE.UT== "ANDAMAN AND NICOBAR"]<- "A & N ISLANDS"
Education_2014$STATE.UT[Education_2014$STATE.UT== "DADRA AND NAGAR HAVELI"]<- "D & N HAVELI"
Education_2014$STATE.UT[Education_2014$STATE.UT== "DAMAN AND DIU"]<- "DAMAN & DIU"
Education_2014$STATE.UT[Education_2014$STATE.UT== "DELHI"]<- "DELHI (UT)"
Education_2014$STATE.UT[Education_2014$STATE.UT== "TOTAL (UTS)"]<- "TOTAL (UTS)"

#changing value of row/educational status based on condition
Education$CAUSE[Education$CAUSE== "Primary (upto class-5)"]<- "Primary"
Education$CAUSE[Education$CAUSE== "Middle (upto class-8)"]<- "Middle"
Education$CAUSE[Education$CAUSE== "Matriculate/Secondary (upto class-10)"]<- "Matriculate/Secondary"
Education$CAUSE[Education$CAUSE== "Hr. Secondary/Intermediate/Pre-Universit" | Education$CAUSE== "Higher Secondary/ Intermediate/"]
Education$CAUSE[Education$CAUSE== "Diploma/ITI/Certificate"]<- "Diploma"
Education$CAUSE[Education$CAUSE== "Graduate and above"]<- "Graduate"
```

Similar steps are carried for each of the dataframes.

Data Wrangling:

The population of India statistics by age and gender was compiled through separate CSVs generated through queries in the website mentioned under data compilation. Screengrabs of the CSV files are displayed below. The data given is in the 1000s and the demarcation by age are not consistent with the age groups in the suicide data.

ISO 3166-1																
2	numeric code	Location	Time	Sex	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59
3	900	World														
4	1828	Sustainable Development Goal (SDG) regions														
5	921	Central and Southern Asia														
6	5501	Southern Asia														
7	356	India	2001	Male	67 929	64 711	62 080	58 209	51 345	45 469	40 352	36 162	31 895	27 953	20 726	16 119
8	356	India	2002	Male	68 200	65 152	62 551	59 065	52 690	46 347	41 171	36 779	32 561	28 602	22 007	16 442
9	356	India	2003	Male	68 324	65 598	62 979	59 804	54 081	47 256	42 039	37 405	33 264	29 121	23 399	16 855
10	356	India	2004	Male	68 367	66 007	63 383	60 465	55 398	48 257	42 926	38 067	33 962	29 639	24 698	17 456
11	356	India	2005	Male	68 392	66 316	63 783	61 074	56 564	49 380	43 814	38 782	34 630	30 233	25 782	18 296
12	356	India	2006	Male	68 479	66 676	64 198	61 587	57 560	50 598	44 685	39 524	35 242	30 820	26 603	19 279
13	356	India	2007	Male	68 463	66 084	63 631	61 043	57 433	50 434	44 410	39 330	35 042	30 860	26 603	19 279
Population by age and sex (thousands)																
ISO 3166-1																
2	numeric code	Location	Time	Sex	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59
3	900	World														
4	1828	Sustainable Development Goal (SDG) regions														
5	921	Central and Southern Asia														
6	5501	Southern Asia														
7	356	India	2001	Female	60 828	57 790	56 036	52 838	47 035	41 887	37 154	33 204	29 239	25 633	19 905	16 001
8	356	India	2002	Female	61 043	58 121	56 315	53 552	48 145	42 687	37 962	33 842	29 919	26 269	20 945	16 345
9	356	India	2003	Female	61 093	58 487	56 549	54 178	49 282	43 496	38 807	34 487	30 632	26 823	22 073	16 781
10	356	India	2004	Female	61 096	58 830	56 770	54 721	50 353	44 358	39 653	35 159	31 340	27 380	23 146	17 261
11	356	India	2005	Female	61 151	59 067	57 011	55 186	51 305	45 301	40 477	35 872	32 019	27 994	24 079	18 107
12	356	India	2006	Female	61 306	59 385	57 286	55 528	52 143	46 337	41 295	36 623	32 661	28 611	24 822	18 927

Using Excel's inbuilt functions, these groups are summed up and combined to get a new CSV file as seen below:

Indian Population									
Location	Time	Male.upto.14.years	Male.15.29.years	Male.30.44.years	Male.45.59.years	Male.60.years.and.above	Total.Male	Female.upto.14.years	Female.15.29.years
TOTAL (ALL INDIA)	2001	194720000	155023000	108409000	64798000	36090000	559040000	174654000	141760000
TOTAL (ALL INDIA)	2002	195903000	158102000	110511000	67051000	37075000	568642000	175479000	144384000
TOTAL (ALL INDIA)	2003	196901000	161141000	112708000	69375000	38050000	578175000	176129000	146956000
TOTAL (ALL INDIA)	2004	197757000	164120000	114955000	71793000	39025000	587650000	176696000	149432000
TOTAL (ALL INDIA)	2005	198491000	167018000	117226000	74311000	40007000	597053000	177229000	151792000
TOTAL (ALL INDIA)	2006	199353000	169745000	119451000	76702000	41135000	606386000	177977000	154008000

Exploring the data for Causes of suicide, it can be seen that there are a large number of causes which might not be suitable for exploration. Using R, causes that are similar are clustered together and aggregate operations are then performed on the dataframe. Also, one of the causes within the original dataframe was discovered to be an aggregate function itself and hence removed.

```
> unique(Causes[c("CAUSE")])
1          CAUSE
2  BANKRUPTCY OR SUDDEN CHANGE IN ECONOMIC
3  SUSPECTED/ILLCIT RELATION
4  CANCELLATION/NON-SETTLEMENT OF MARRIAGE
5  NOT HAVING CHILDREN(BARRENNESS/IMPOTENCY)
6  ILLNESS (AIDS/STD)
7  CANCER
8  PARALYSIS
9  INSANITY/MENTAL ILLNESS
10 OTHER PROLONGED ILLNESS
11 DEATH OF DEAR PERSON
12 DOWRY DISPUTE
13 DIVORCE
14 DRUG ABUSE/ADDICTION
15 FAILURE IN EXAMINATION
16 FALL IN SOCIAL REPUTATION
17 FAMILY PROBLEMS
18 IDEOLOGICAL CAUSES/HERO WORSHIPPING
19 ILLEGITIMATE PREGNANCY
20 LOVE AFFAIRS
21 PHYSICAL ABUSE (RAPE,INCEST ETC.)
22 POVERTY
23 PROFESSIONAL/CAREER PROBLEM
24 PROPERTY DISPUTE
25 UNEMPLOYMENT
26 CAUSES NOT KNOWN
27 OTHER CAUSES (PLEASE SPECIFY)
28 TOTAL ILLNESS
29 BANKRUPTCY OR SUDDEN CHANGE IN ECONOMIC STATUS
32 NOT HAVING CHILDREN (BARRENNESS/IMPOTENCY)
9748 NOT HAVING CHILDREN (BARRENNESS/IMPOTENCY)
9770 OTHER CAUSES (PLEASE SPECIFY)
14197 BANKRUPTCY OR INDEBTEDNESS
14198 MARRIAGE RELATED ISSUES

#removing redundant aggregate
Causes<-Causes[!(Causes$CAUSE=="TOTAL ILLNESS"),]
#clustering different causes
Causes$CAUSE[Causes$CAUSE=="BANKRUPTCY OR SUDDEN CHANGE IN ECONOMIC" | Causes$CAUSE=="BANKRUPTCY OR SUDDEN CHANGE IN ECONOMIC STATUS" | Causes$CAUSE=="BANKRUPTCY OR INDEBTEDNESS"] <- "BANKRUPTCY OR SUDDEN CHANGE IN ECONOMIC"
Causes$CAUSE[Causes$CAUSE=="CANCELLATION/NON-SETTLEMENT OF MARRIAGE" | Causes$CAUSE=="SUSPECTED/ILLCIT RELATION"] <- "CANCELLATION/NON-SETTLEMENT OF MARRIAGE"
Causes$CAUSE[Causes$CAUSE=="CANCER" | Causes$CAUSE=="PARALYSIS" | Causes$CAUSE=="TOTAL ILLNESS"] <- "CANCER"
Causes$CAUSE[Causes$CAUSE=="DOWRY RELATED ISSUES"] <- "DOWRY DISPUTE"
Causes$CAUSE[Causes$CAUSE=="OTHER FAMILY PROBLEMS"] <- "FAMILY PROBLEMS"
Causes$CAUSE[Causes$CAUSE=="PHYSICAL ABUSE (RAPE; ETC.)"] <- "PHYSICAL ABUSE (RAPE,INCEST ETC.)"
Causes$CAUSE[Causes$CAUSE=="OTHER CAUSES (PLEASE SPECIFY)"] <- "OTHER CAUSES (PLEASE SPECIFY)"

#aggregating the clusters
aggregate_causes <- aggregate(.,~STATE.UT+Year+CAUSE, Causes, sum)
```

Hence, I obtained a reduced set of features which are easier to explore and visualize.

Data Exploration:

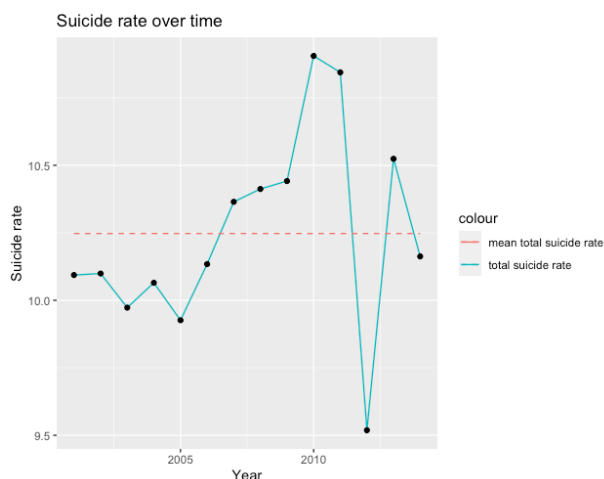
1. National trend of suicide data:

Across India, the rate of suicides seems to show no noticeable decrease. Instead, we can see a steep increase from the period 2005 to 2010, followed by a sharp decrease over the next two years which looks to be an outlier. However, the data from 2013 and 2014 still seem to show the beginning of a new trend of lower suicide rate.

Key Insights:

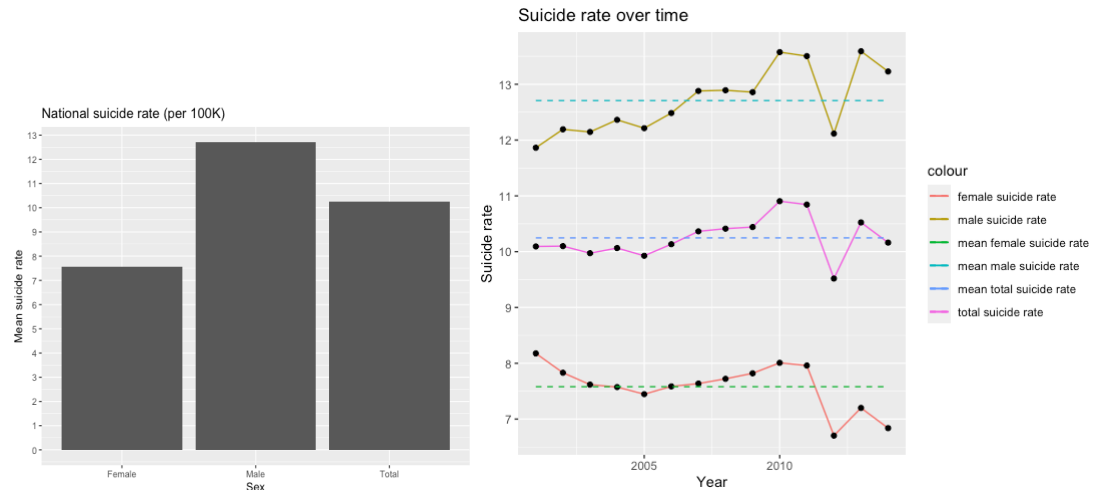
The average suicide rate over the time period 2001 to 2014 is 10.24749.

Using Pearson's correlation coefficient, no relationship was found between the suicide rate and time.



2. National trend by sex:

Sex is a key factor when it comes to the national suicide data. Males are overrepresented with the increase in the gap between male and female suicide rate increasing over the years. The male suicide rate was 1.45 times the female suicide rate in 2001 and it had gone up to 1.93 times in 2014.



3. Suicide trend by States and Union Territories:

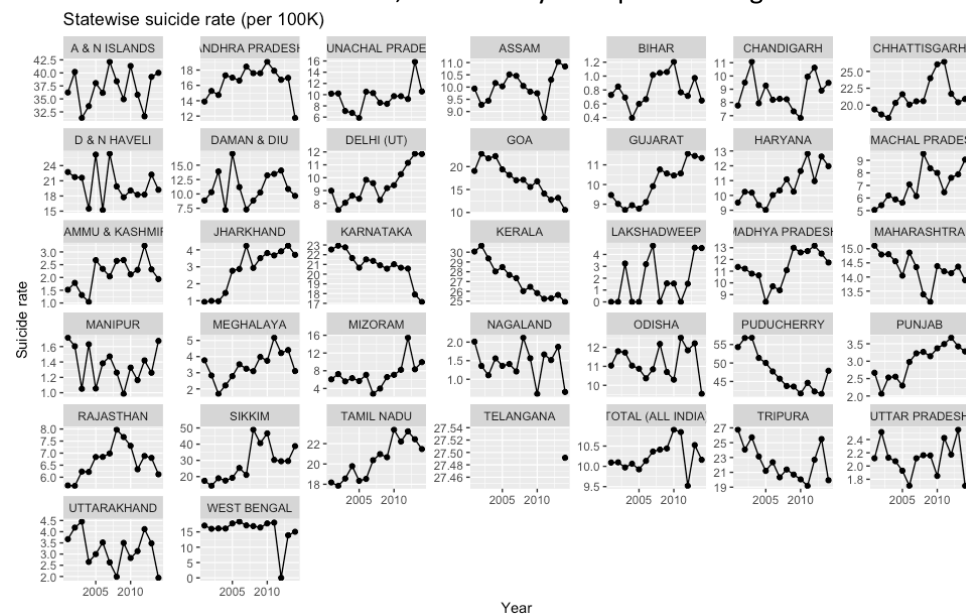
Looking at the suicide trends for the different states and union territories-

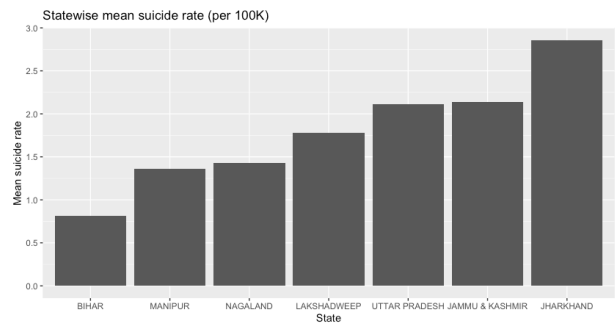
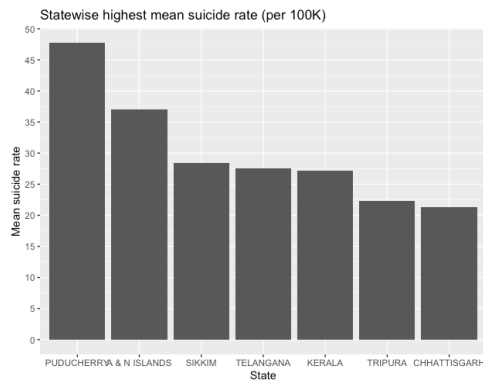
Goa, Karnataka and Kerala are the only states showing visible and consistent decrease in the suicidal rate.

Delhi(UT), Gujarat, Jharkhand and Punjab are on the other end of the spectrum showing worrying constant increases in the suicide rate.

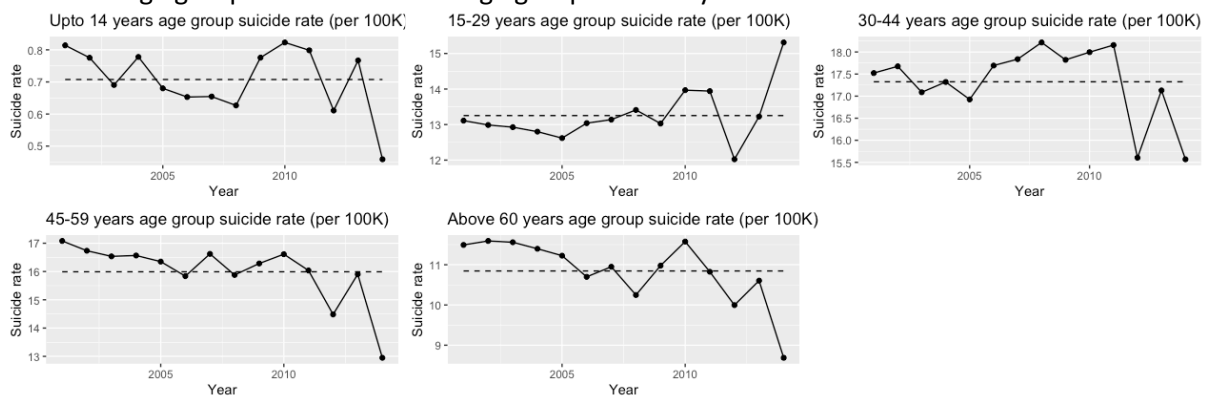
Looking at the suicide rates, Puducherry has by far the highest suicide rate, followed by A&N Islands and Sikkim. While Puducherry has shown a decreasing trend, A&N Islands shows a widely fluctuating rate. Sikkim is the most worrying however, because it has historically shown an increasing trend.

Bihar has the lowest suicide rate, followed by Manipur and Nagaland.

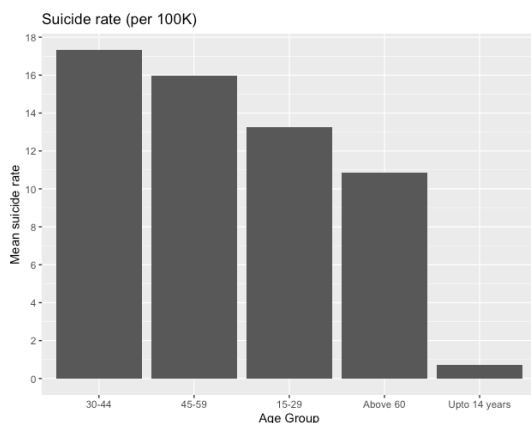




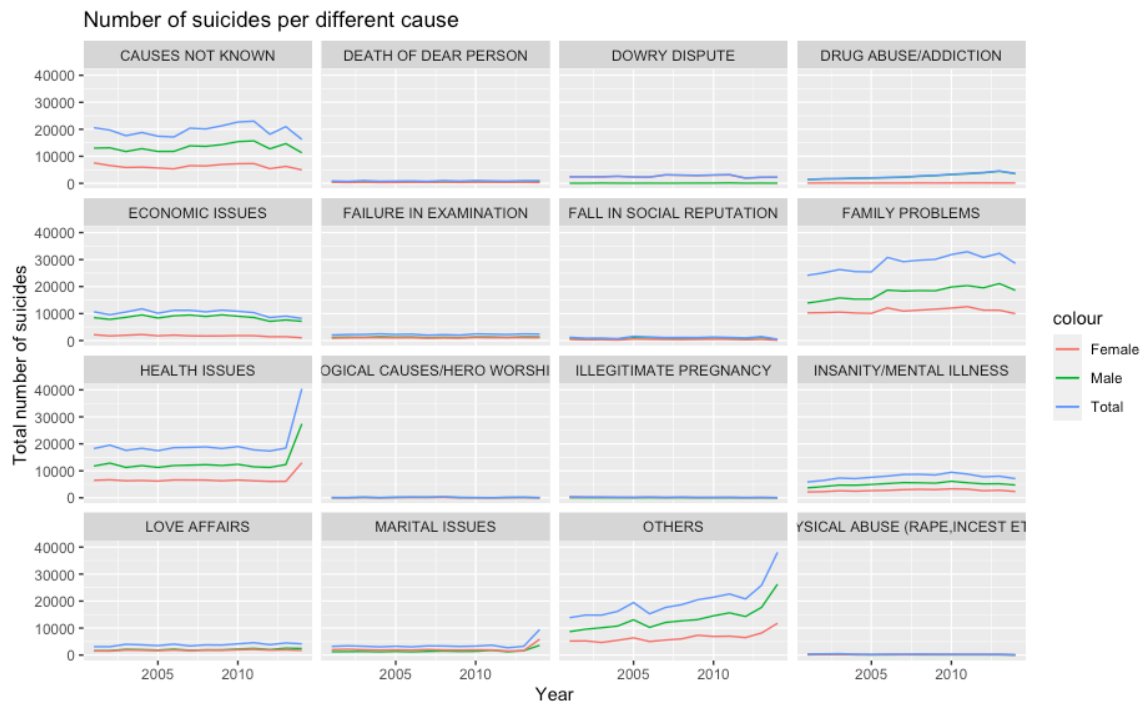
4. When examining the different age groups, a clear increase is seen in the 15-29 age group which is worrying. On the other hand, there has been a steady decline in the suicide rate for the 45-59 age group and the above 60 age group over the years.



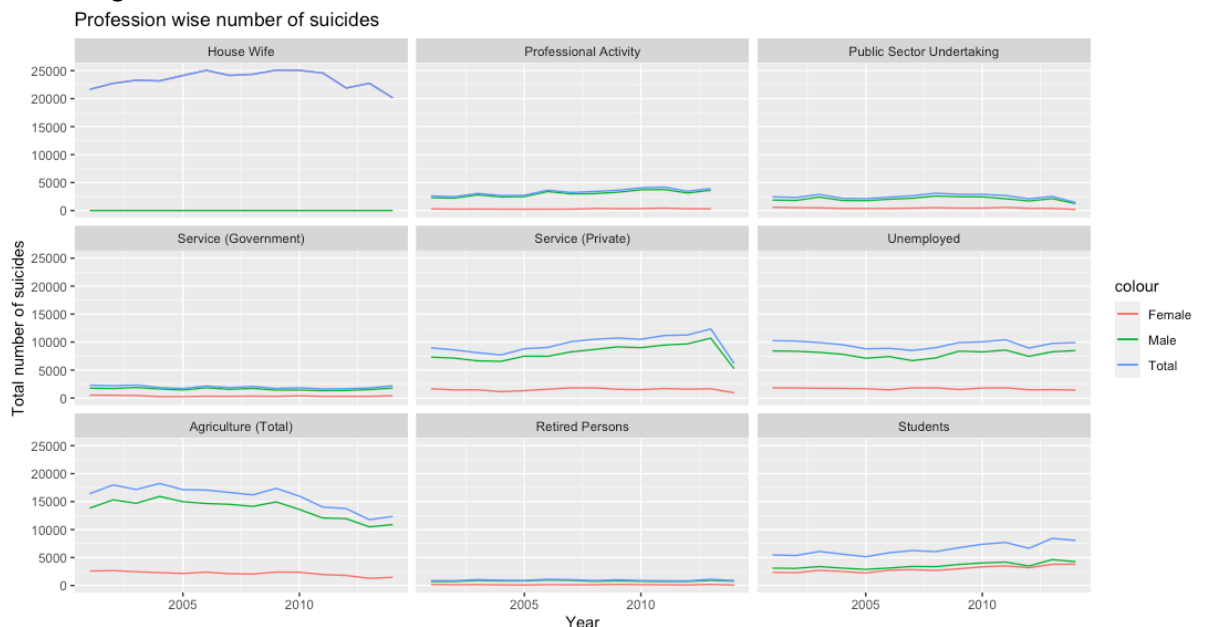
30-44 is the most at risk age group followed by 45-59. The most worrisome is 15-29 which comes in third but it has been showing an increasing trend which initiates concern. People up to 14 years are the least at risk.



5. The data compiling the causes for suicide show some interesting and worrying results. Family problems, Health Issues and Economic Issues are the major causes for suicide in India followed by mental illness (ignoring unknown causes and other causes). This is worrying because health, economic and family problems are issues that everyone is susceptible to. While physical and mental health can be tackled to some extent by improving healthcare facilities, the others are a lot more vague and personal. Health Issues worryingly shows an increasing trend.

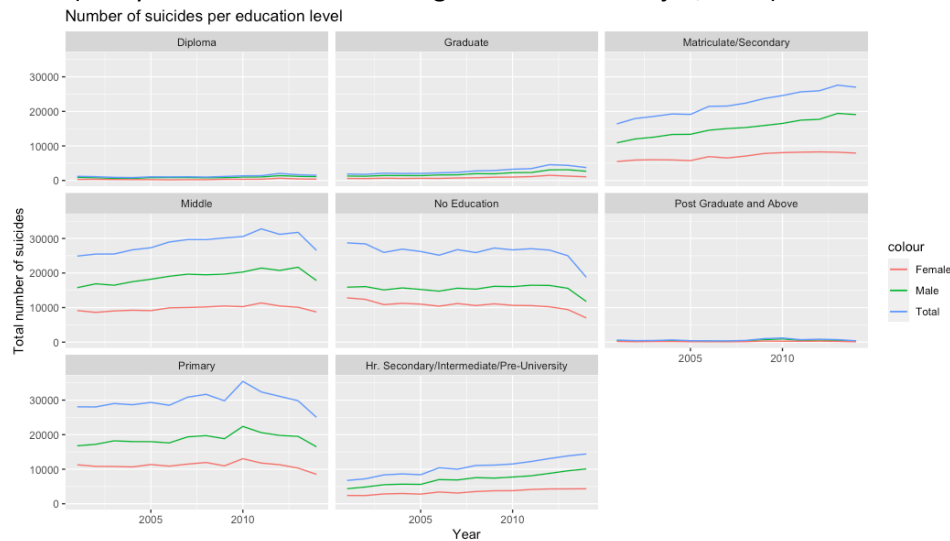


6. Examining professions, it is evident that housewives are the most at risk demographic made up completely of women. However, this might be skewed since less than 20% of women account for labour force in India and the rest remain as housewives due to traditional and cultural norms ("Women in the Workforce - India: Quick Take | Catalyst", 2021). Looking at the rest, Agriculture is the most at risk, followed by people employed in private service, unemployed people and most worryingly students. Agriculture has shown a slowly decreasing trend, however, students have shown an increasing trend to commit suicide.

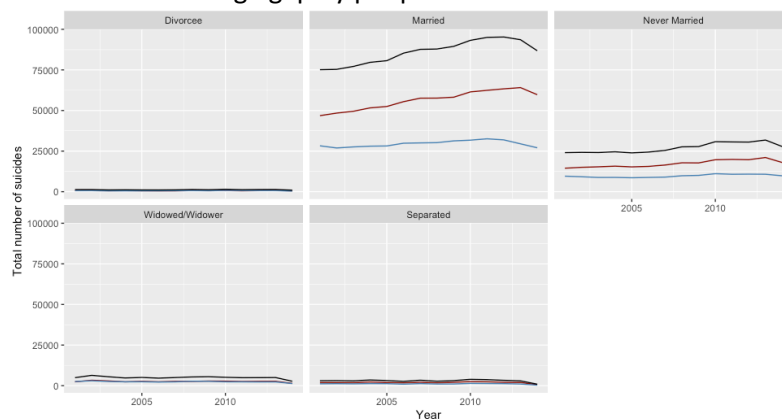


7. Education level shows that people with higher levels of education – Diploma, Graduate, Post Graduate and above are less likely to commit suicide. Whereas the lower educational levels all show high number of suicides. The most worrying trends are that of Matriculate/Secondary and Hr. Secondary both showing highly increasing numbers along with a small bump in the number of graduates committing suicide. This could tie into the

earlier discovered trend of increasing number of students committing suicide which can be tied into high family and societal pressure, failure in examinations and draconian school rules ("Why are students committing suicide? - EastMojo", 2021).



8. Social status doesn't reveal anything too drastic. Most of the deceased were married, followed after a large gap by people who have never married.



Conclusion:

From the data exploration process, I have answered most of my questions.

1. The progress of India as a country over time has had no effect on the suicide rate. In fact, the suicide rate of the country has gone up over the years. This could be due to a number of reasons including the fact that suicide had become normalized through films and popular culture as a sort of last gasp solution to issues.
2. Family problems are the major cause of suicide in India and was only overtaken recently by health issues. Indian culture cultivates close knit families which inevitably leads to a lot of expectations and pressure which could be a key reason for this.
3. Through the lens of socio-economical/educational status, the most vulnerable demographic can be identified as a middle aged married male lacking in educational qualifications working in the agricultural sector. This can be backed up by research which has identified an outbreak of farmer suicides in India (Merriott, 2016).
4. The recent increase in the suicide rate of 15-29 age group and the decrease in the 30-44 and above 60 age group shows a worrying trend of younger people more likely to commit suicide.
5. Men are overrepresented in suicide statistics which highlights the need for better communication and prevention guidelines to deal with this at risk demographic.

Reflection:

From this project, I have gained valuable insights into who might be more vulnerable to taking their own lives and how certain societal privileges give certain demographics a healthy advantage. It also brings into focus an immediate need to improve suicide prevention systems in India.

In hindsight, should have utilized HDI, GDP, literacy rate etc to gain even more of an insight into what are the key factors that could lead to someone being at risk. Also I could have used Excel much earlier to do most of the wrangling and cleaning faster instead of using RStudio.

References:

Hawton, K., & van Heeringen, K. (2009). Suicide. *The Lancet*, 373(9672), 1372-1381. doi: 10.1016/s0140-6736(09)60372-x

Seshan, K. (2018). Telangana: History and the formation of a new state. *Studies In People's History*, 5(1), 72-82. doi: 10.1177/2348448918759870

Women in the Workforce - India: Quick Take | Catalyst. (2021). Retrieved 24 April 2021, from <https://www.catalyst.org/research/women-in-the-workforce-india/>

Why are students committing suicide? - EastMojo. (2021). Retrieved 24 April 2021, from <https://www.eastmojo.com/opinion/2020/01/28/why-are-students-committing-suicide/>

Merriott, D. (2016). Factors associated with the farmer suicide crisis in India. *Journal Of Epidemiology And Global Health*, 6(4), 217. doi: 10.1016/j.jegh.2016.03.003