## **ABOUT THE DATASET**



- Source -Dataset taken from <a href="https://data.gov.in/">https://data.gov.in/</a>
- Name of dataset "State/UT-wise Total Missing and traced Persons form 2017-2021".
- It's the integrated dataset consists of 180 rows and 28 columns which we formed from combining 5 different Datasets of missing, traced and untraced person from 2017,2018,2019,2020,2021 in which the information relates to the gender-wise distribution of person.

Basic Statistics	
Raw Counts	
Name	Value
Rows	180
Columns	28
Discrete columns	2
Continuous columns	26
All missing columns	0
Missing observations	0
Complete Rows	180
Total observations	5,040
Memory allocation	35.5 Kb

<ul> <li>Quality parameters like gender-wise Recovery,         Untraced and missing person's etc by which we         can predict the different cases state wise and year         wise.     </li> </ul>
There are some rows with NA values in the data too.

### PURPOSE

- The purpose of choosing the dataset is to analyze is the recovery rate, unrecovery rate of a person and visualize them using different plots. Visualization is used to highlight the useful information.
- To compare the state-wise and year-wise increase and decrease in number of missing, unrecovery and recovered cases.

## PLOT USED FOR VISUALIZATION

Bar Plot

Line Plot [18]

Choropleth (India) Map



R tables |

# READING THE DATASET



df <- read.csv("Missing Cases Complied (2017-2021).csv")</pre>

#### VIFW

s.	. No	Year	Category	State.UT	Unrecovered.Untraced.Pd	ersons.from.Previous.yearsMale	Unrecovered.Untraced.Persons.from	.Previous.yearsFemale	Unrecovered.Untraced	i.Persons.from.Previous.ye
<i< th=""><th>nt&gt;</th><th><int></int></th><th><chr></chr></th><th><chr></chr></th><th></th><th><int></int></th><th></th><th><int></int></th><th></th><th></th></i<>	nt>	<int></int>	<chr></chr>	<chr></chr>		<int></int>		<int></int>		
1	1	2017	State	Andhra Pradesh		4173		2990		
2	2	2017	State	Arunachal Pradesh		2		93		
3	3	2017	State	Assam		530		1223		
4	4	2017	State	Bihar		1523		2935		
5	5	2017	State	Chhattisgarh		4366		5836		
6	6	2017	State	Goa		475		230		
										<b>)</b>
4										,
Unred	over	ed.Untı	raced.Perso	ons.from.Pre	vious.yearsTotal Per	rsons.Missing.this.yearMale Pe	ersons.Missing.this.yearFemale	··· Total.Recovered.Trace	ed.this.yearTrans	
Unred	over	ed.Unti	raced.Perso	ons.from.Pre	vious.yearsTotal Per <int></int>	rsons.Missing.this.yearMale Pe <int></int>	ersons.Missing.this.yearFemale		ed.this.yearTrans <int></int>	
Unred	cover	ed.Unti	raced.Perso	ons.from.Pre						
Unred	cover	ed.Unt	raced.Perso	ons.from.Pre	<int></int>	<int></int>	<int></int>		<int></int>	
Unred	cover	ed.Unt	raced.Perso	ons.from.Pre	<int> 7168</int>	<int></int>	<int></int>	<del></del> 	<int></int>	
Unrec	cover	ed.Unt	raced.Perso	ons.from.Pre	<int> 7168</int>	(int) 3114 9	<int> 5225</int>	<del></del>  	<int> 7</int>	
Unred	cover	ed.Unt	raced.Perso	ons.from.Pre	<int> 7168 95 1753</int>	sint> 3114 9 997	<pre>5225 90 2302</pre>	 	<int> 7 0 0</int>	

## **DIMENSIONS OF THE DATASET**

### Dimension of Rows and Column

dim(df)

180 - 28

# REPLACING THE "NA" VALUES WITH "0"

df[is.na(df)]=0

The dataset generally comprises of "NA" value in the columns related to transgender.

# DESCRIPTION OF THE DATA

View(summary(df))

•	Var1	Var2	Freq
1		SNo	Min. : 1.00
2		SNo	1st Qu.: 45.75
3		SNo	Median : 90.50
4		SNo	Mean : 90.50
5		SNo	3rd Qu.:135,25
6		SNo	Max. :180.00
7		SNo	NA
8		Year	Min. ;2017
9		Year	1st Qu::2018
10		Year	Median :2019
11		Year	Mean :2019
12		Year	3rd Qu.:2020
13		Year	Max. :2021
14		Year	NA
15		Category	Length:180
16		Category	Class :character
17		Category	Mode :character
18		Category	NA
19		Category	NA
20		Category	NA
21		Category	NA
22		State.UT	Length:180
23		State.UT	Class :character
24		State,UT	Mode :characte
25		State.UT	NA
26		State.UT	NA
27		State.UT	NA
28		State.UT	NA
29		Unrecovered.Untraced.Persons.from.Previous.yearsMale	Min. : 0
30		Unrecovered.Untraced.Persons.from,Previous.yearsMale	1st Qu.: 56
31		Unrecovered.Untraced.Persons.from.Previous.yearsMale	Median: 1258
32		Unrecovered.Untraced.Persons.from.Previous.yearsMale	Mean : 3893
33		Unrecovered.Untraced.Persons.from.Previous.yearsMale	3rd Qu.: 5185

*	Var1	Var2	Freq
33		Unrecovered.Untraced.Persons.from.Previous.yearsMale	3rd Qu.: 5185
34		Unrecovered.Untraced.Persons.from.Previous.yearsMale	Max. :28778
35		${\tt Unrecovered. Untraced. Persons. from. Previous. years Male}$	NA
36		Unrecovered.Untraced.Persons.from.Previous.yearsFemale	Min. : 0.00
37		Unrecovered.Untraced.Persons.from,Previous.yearsFemale	1st Qu.: 66.25
38		Unrecovered.Untraced.Persons.from.Previous.yearsFemale	Median : 1217.00
39		Unrecovered.Untraced.Persons.from,Previous,yearsFemale	Mean : 5047.92
40		Unrecovered.Untraced.Persons.from.Previous.yearsFemale	3rd Qu.: 4889.25
41		Unrecovered.Untraced.Persons.from.Previous.yearsFemale	Max. :39124.00
42		Unrecovered.Untraced.Persons.from.Previous.yearsFemale	NA
43		Unrecovered.Untraced.Persons.from,Previous.yearsTrans	Min. : 0.000
44		Unrecovered.Untraced.Persons.from.Previous.yearsTrans	1st Quz 0.000
45		Unrecovered.Untraced.Persons.from.Previous.yearsTrans	Median: 0.000
46		Unrecovered.Untraced.Persons.from.Previous.yearsTrans	Mean : 1.839
47		Unrecovered.Untraced.Persons.from.Previous.yearsTrans	3rd Qu.; 0.000
48		Unrecovered.Untraced.Persons.from.Previous.yearsTrans	Max. :100.000
49		Unrecovered.Untraced.Persons.from.Previous.yearsTrans	NA
50		Unrecovered.Untraced.Persons.from.Previous.yearsTotal	Min. : 0.0
51		Unrecovered.Untraced.Persons.from.Previous.yearsTotal	1st Qu.: 127.5
52		Unrecovered.Untraced.Persons.from.Previous.yearsTotal	Median : 2779.5
53		Unrecovered.Untraced.Persons.from.Previous.yearsTotal	Mean : 8942.5
54		Unrecovered.Untraced.Persons.from.Previous.yearsTotal	3rd Qu.:10094,8
55		Unrecovered.Untraced.Persons.from.Previous.yearsTotal	Max. :62231.0
56		Unrecovered.Untraced.Persons.from.Previous.yearsTotal	NA
57		Persons Missing this year Male	Min. : 0.00
58		Persons, Missing, this, year Male	1st Qu.: 91.25
59		Persons.Missing.this.yearMale	Median : 1191.50
60		Persons.Missing.this.yearMale	Mean : 3315.67
61		Persons Missing this year Male	3rd Qu.: 4560.75
62		Persons, Missing, this, year Male	Max. :27972.00
63		Persons.Missing.this.yearMale	NA
64		Persons.Missing.this.yearFemale	Min. : 0.0
65		Persons.Missing.this.yearFemale	1st Qu.: 138.8

•	Var1	Var2	Freq
66		Persons.Missing.this.yearFemale	Median: 1727.0
67		Persons.Missing.this.yearFemale	Mean : 6379.3
68		Persons.Missing.this.yearFemale	3rd Qu.: 9444.5
69		Persons.Missing.this.yearFemale	Max. :38506.0
70		Persons.Missing.this.yearFemale	NA :
71		Persons.Missing.this.yearTrans	Min. : 0.000
72		Persons.Missing.this.yearTrans	1st Qu.: 0.000
73		Persons.Missing.this.yearTrans	Median: 0.000
74		Persons.Missing.this.yearTrans	Mean : 6.861
75		Persons.Missing.this.yearTrans	3rd Qu.: 0.000
76		Persons.Missing.this.yearTrans	Max. :488.000
77		Persons.Missing.this.yearTrans	NA
78		Persons:Missing.this.yearTotal	Min. : 0.0
79		Persons.Missing.this.yearTotal	1st Qu.: 215.8
80		Persons.Missing.this.yearTotal	Median : 2854.0
81		Persons.Missing.this.yearTotal	Mean : 9701.8
82		Persons.Missing.this.yearTotal	3rd Qu.:14193.0
83		Persons.Missing.this.yearTotal	Max. :66478.0
84		Persons:Missing.this.yearTotal	NA:
85		Total Persons Missing Male	Min. : 0
86		Total.Persons.MissingMale	1st Qu.: 170
87		Total.Persons.MissingMale	Median : 2930
88		Total.Persons.MissingMale	Mean : 7208
89		Total Persons Missing Male	3rd Qu.: 9064
90		Total.Persons.MissingMale	Max. :56750
91		Total Persons Missing Male	NA
92		Total.Persons.MissingFemale	Min. : 0.0
93		Total Persons Missing Female	1st Qu.: 180.5
94		Total.Persons.MissingFemale	Median : 3573.0
95		Total.Persons.MissingFemale	Mean :11427.2
96		Total.Persons.MissingFemale	3rd Qu.:13638.5
97		Total Persons Missing Female	Max. :68738.0
98		Total.Persons.MissingFemale	NA

•	Var1	Var2	Freq
99		Total.Persons.MissingTrans	Min. : 0.0
100		Total.Persons.MissingTrans	1st Qu.: 0.0
101		Total.Persons.MissingTrans	Median : 0.0
102		Total.Persons.MissingTrans	Mean : 8.7
103		Total.Persons.MissingTrans	3rd Qu.: 0.0
104		Total Persons Missing Trans	Max. :493.0
105		Total.Persons.MissingTrans	NA
106		Total.Persons.MissingTotal	Min. : 0.0
107		Total.Persons.MissingTotal	1st Qu.: 342.5
108		Total.Persons.MissingTotal	Median : 6543.5
109		Total.Persons.MissingTotal	Mean : 18644.3
110		Total.Persons.MissingTotal	3rd Qu.: 23330.5
111		Total.Persons.MissingTotal	Max. :124515.0
112		Total, Persons, Missing Total	NA
113		Total.Recovered.Traced.this.yearMale	Min. : 0.00
114		Total.Recovered.Traced.this.yearMale	1st Qu.: 91.75
115		Total.Recovered.Traced.this.yearMale	Median : 1012.50
116		Total.Recovered.Traced.this.yearMale	Mean : 3335.51
117		Total.Recovered.Traced.this.yearMale	3rd Qu.: 4335.75
118		Total.Recovered.Traced.this.yearMale	Max. :30235.00
119		Total.Recovered.Traced.this.yearMale	NA
120		Total.Recovered.Traced.this.yearFemale	Min. : 0
121		Total.Recovered.Traced.this.yearFemale	1st Qu: 134
122		Total.Recovered.Traced.this.yearFemale	Median: 1464
123		Total.Recovered.Traced.this.yearFemale	Mean : 6214
124		Total.Recovered.Traced.this.yearFemale	3rd Qu.: 8773
125		Total.Recovered.Traced.this.yearFemale	Max. :40095
126		Total.Recovered.Traced.this.yearFemale	NA
127		Total.Recovered.Traced.this.yearTrans	Min. : 0.000
128		Total.Recovered.Traced.this.yearTrans	1st Qu.: 0.000
129		Total.Recovered.Traced.this.yearTrans	Median: 0.000
130		Total.Recovered.Traced.this.yearTrans	Mean : 6.872
131		Total.Recovered.Traced.this.yearTrans	3rd Qu.: 0.000

*	Var1	Var2	Freq
132		Total.Recovered.Traced.this.yearTrans	Max. :434.000
133		Total.Recovered.Traced.this.yearTrans	NA
134		Total.Recovered.Traced.this.yearTotal	Min. : 0.0
135		Total.Recovered.Traced.this.yearTotal	1st Qu.: 223.5
136		Total.Recovered.Traced.this.yearTotal	Median : 2417.0
137		Total.Recovered.Traced.this.yearTotal	Mean : 9556.1
138		Total.Recovered.Traced.this.yearTotal	3rd Qu.:13195.0
139		Total.Recovered.Traced.this.yearTotal	Max. :67018.0
140		Total.Recovered.Traced.this.yearTotal	NA
141		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	Min. : 0.00
142		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	1st Qu.: 34.15
143		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	Median : 47.00
144		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	Mean :48.11
145		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	3rd Qu.: 63.20
146		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	Max. :100.00
147		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	NA's :5
148		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	Min. : 18.70
149		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	1st Qu.: 45.65
150		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	Median : 61.90
151		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	Mean : 59.82
152		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	3rd Qu.: 75.50
153		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	Max. :100.00
154		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	NA's :9
155		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	Min. : 0.00
156		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	1st Qu.: 50.00
157		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	Median :100.00
158		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	Mean : 78.54
159		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	3rd Qu.:100.00
160		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	Max. :100.00
161		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	NA's :135
162		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	Min. : 0.00
163		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	1st Qu.: 39.95
164		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	Median : 53.35

165	Var1	Var2  Percentage.Recoverylotal.Recovered.or, Iracedlotal.Missin	Freq Mean : 54.36
166		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	3rd Qu.: 68.10
167		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	Max. :100.00
168		Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missin	NA's :4
169		Un.Recovered.Untraced.Personsincluding.previous.years.Un	Min. : 0.00
170		Un.Recovered.Untraced.Personsincluding.previous.years.Un	1st Qu.: 52.75
171		Un.Recovered.Untraced.Persons.including.previous.years.Un	Median : 1331.00
172		Un.Recovered.Untraced.Persons.including.previous.years.Un	Mean : 3872.92
173		Un.Recovered.Untraced.Personsincluding.previous.years.Un	3rd Qu.: 5185.00
174		Un.Recovered.Untraced.Persons.including.previous.years.Un	Max. :28778.00
175		Un.Recovered.Untraced.Persons.including.previous.years.Un	NA
176		Un.Recovered.Untraced.Personsincluding.previous.years.Un	Min. : 0
177		Un.Recovered.Untraced.Persons.including.previous.years.Un	1st Qu.: 59
178		Un.Recovered.Untraced.Persons.including.previous.years.Un	Median: 1185
179		Un.Recovered.Untraced.Persons.including.previous.years.Un	Mean : 5213
180		Un.Recovered.Untraced.Personsincluding.previous.years.Un	3rd Qu.: 5452
181		Un.Recovered.Untraced.Personsincluding.previous.years.Un	Max. :35110
182		Un.Recovered.Untraced.Personsincluding.previous.years.Un	NA
183		Un.Recovered.Untraced.Personsincluding.previous.years.Un	Min.: 0.000
184		Un.Recovered.Untraced.Personsincluding.previous.years.Un	1st Qu.: 0.000
185		Un.Recovered.Untraced.Personsincluding.previous.years.Un	Median: 0.000
186		Un.Recovered.Untraced.Personsincluding.previous.years.Un	Mean : 1.828
187		Un.Recovered.Untraced.Personsincluding.previous.years.Un	3rd Qu.: 0.000
188		Un.Recovered.Untraced.Personsincluding.previous.years.Un	Max. :128.000
189		Un.Recovered.Untraced.Personsincluding.previous.years.Un	NA
190		Un.Recovered.Untraced.Persons.including.previous.years.Un	Min. : 0.0
191		Un.Recovered.Untraced.Persons.including.previous.years.Un	1st Qu.: 109.5
192		Un.Recovered.Untraced.Personsincluding.previous.years.Un	Median : 3071.0
193		Un.Recovered.Untraced.Personsincluding.previous.years.Un	Mean : 9088.2
194		Un.Recovered.Untraced.Personsincluding.previous.years.Un	3rd Qu.:11105.2
195		Un.Recovered.Untraced.Personsincluding.previous.years.Un	Max. :58037.0
196		Un.Recovered.Untraced.Personsincluding.previous.years.Un	NA

## STRUCTURE OF THE DATA

str(df)

## REPORT OF THE DATA

library(DataExplorer)

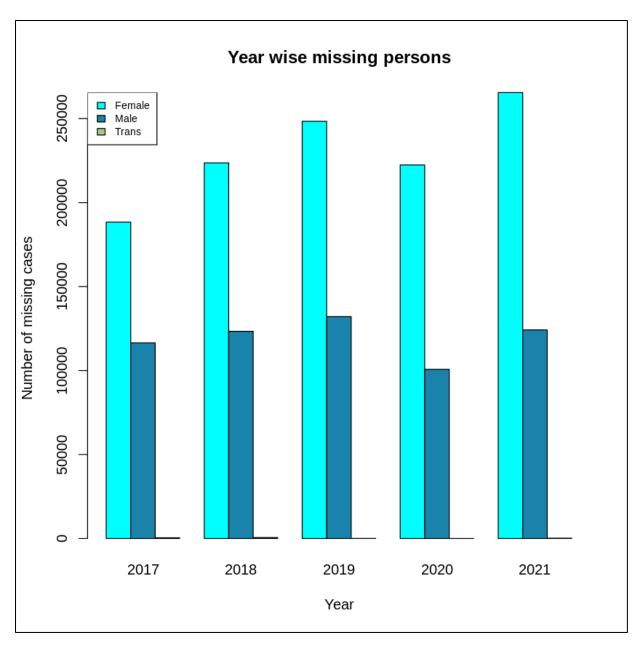
DataExplorer::create\_report(df)

(submitted separately)

# DATA INTERPRETATION

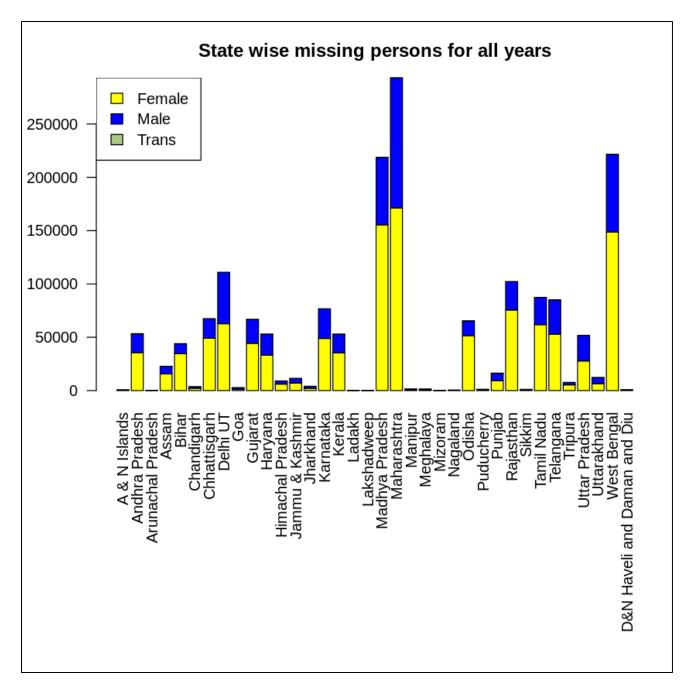
Ques 1: In which year and in which state were total maximum missing cases (male, female and transgender)?

Year	total_missing_female	total_missing_male	total_missing_trans
<int></int>	<int></int>	<int></int>	<int></int>
2017	188382	116467	418
2018	223621	123339	564
2019	248397	132089	40
2020	222395	100748	27
2021	265481	124177	186



(Year-Wise)

Description: The graph shows the number of missing cases from year 2017-2021 using barplot wherein each year consist of three bars (Male,Female,Transgender).



(State-wise)

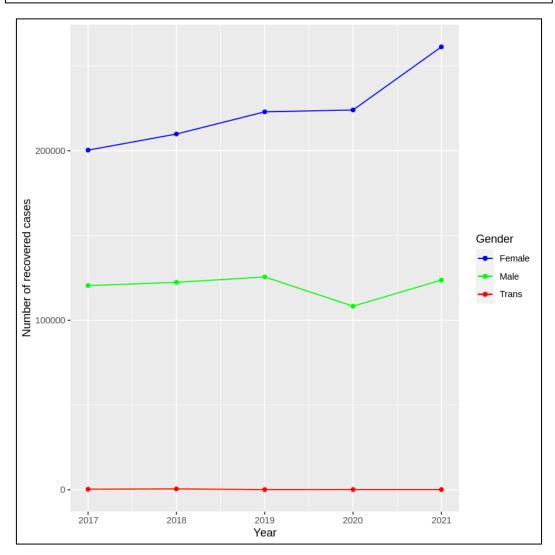
Description: The graph shows the state-wise /UT-wise number of missing cases using barplot wherein each States/UT consist of bars (Male,Female,Transgender).

#### Interpretation:

- Higher number of missing persons can have following possibilities:
  - Young people wanting to become independent.
  - Being a victim of crime-Human trafficking, Kidnapping etc.
  - Mental Health problems.
  - Old age people who couldn't find their way home due to possible loss of memory.
  - Substance Abuse.
- According to the plotted graph the ratio of missing cases of female are much higher as compared to males and the transgender (which has negligible cases).
- There is a steep decline in the no. of missing cases in the year 2020 and a reason for this could be the lockdown due to Covid-19.

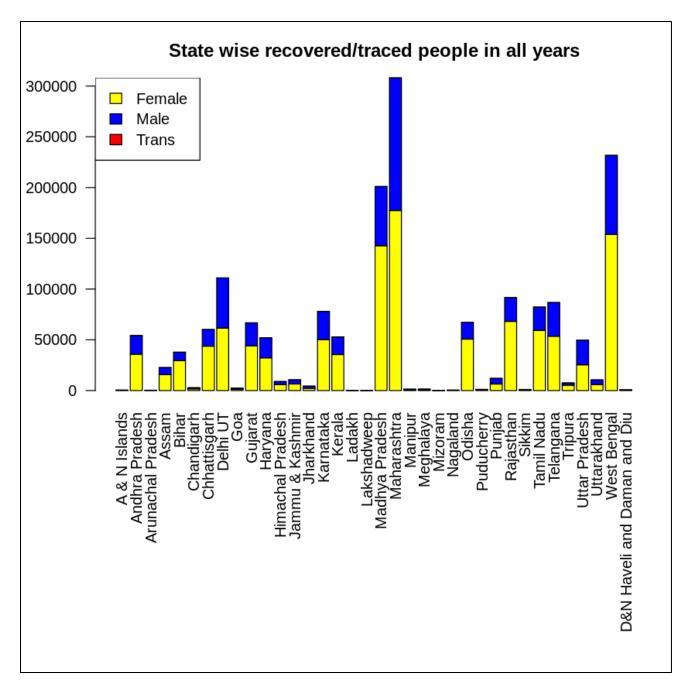
Ques 2: ST/UT wise total recovered cases yearly based?

Year	total_recovered_female	total_recovered_male	total_recovered_trans
<int></int>	<int></int>	<int></int>	<int></int>
2017	200357	120435	346
2018	209852	122384	518
2019	222949	125558	101
2020	224043	108298	142
2021	261278	123716	130



(Year-wise)

Description: The line graph shows the number of recoverd cases from the year 2017-2021 of three bars (Male, Female, Transgender).



(State-wise)

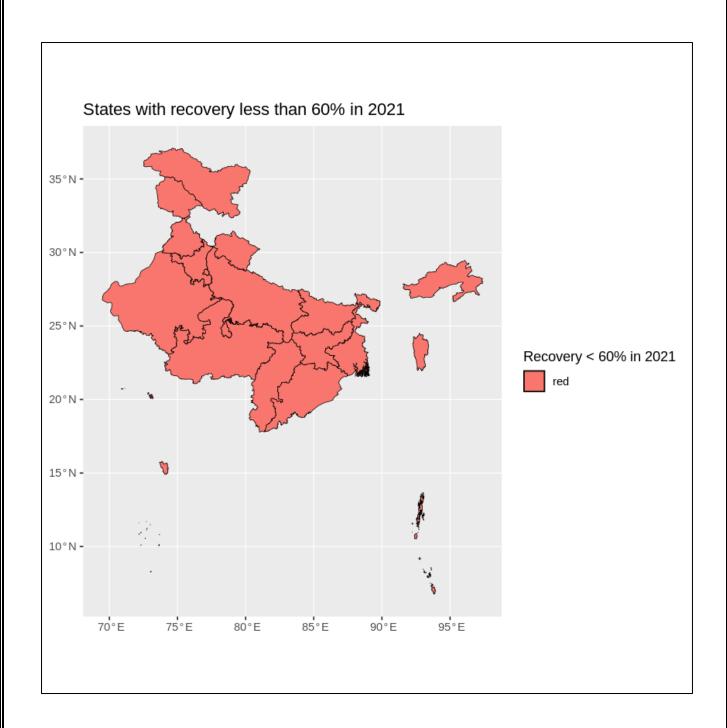
Description: The graph shows the state-wise /UT-wise number of recovered cases using barplot wherein each States/UT consist of bars (Male,Female,Transgender).

Interpretation: Recovery rate of female are much higher than the other two gender over the years but there was certain decline in the recovery rate of the year 2020 due to possible reason of lockdown but amidst these the transgender rate of recovery increase has still not declined in the year 2020.
Another possible reason for a greater number of recovered female cases is due to the larger number of missing cases for female.

# Ques 3: Identify the states with recovery rate less than 60%?

Name	Туре	Year	Percentage.RecoveryTotal.Recovered.or.TracedTotal.Missing100Total
<chr></chr>	<chr></chr>	<int></int>	<db1></db1>
Andaman & Nicobar	Union Territory	2021	28.5
Arunachal Pradesh	State	2021	42.1
Bihar	State	2021	43.9
Chandigarh	Union Territory	2021	29.4
Chhattishgarh	State	2021	42.9
Daman and Diu and Dadra and Nagar Haveli	Union Territory	2021	59.7
Delhi	Union Territory	2021	33.5
Goa	State	2021	35.4
Haryana	State	2021	57.8
Jammu and Kashmir	Union Territory	2021	38.9
Jharkhand	State	2021	46.3
Ladakh	Union Territory	2021	22.8
Lakshadweep	Union Territory	2021	0.0
Madhya Pradesh	State	2021	50.0
Mizoram	State	2021	0.0

Odisha	State	2021	49.2
Punjab	State	2021	16.6
Rajasthan	State	2021	55.3
Uttar Pradesh	State	2021	39.5
Uttarakhand	State	2021	50.7
West Bengal	State	2021	44.5

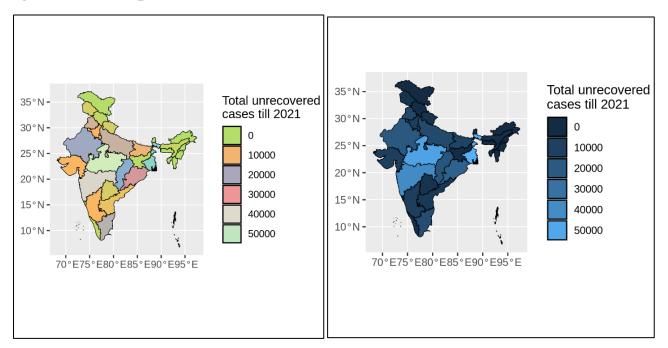


Description: In this Map, the red regions signify the states with recovery rate less than 60% in the year 2021 i.e. [among which the regions with lowest recovery rate are: Punjab, Chandigarh]

Interpretation: On observation the northern region has less recovery rate than the southern region of India. Few reasons could be:

- Faults in Police management system.
- Comparable less literacy rate in few of these northern state.
- More population density, thus more missing cases (as Lakshadweep and Mizoram being exception i.e., they have very few or tends to "1" missing cases).

Ques 4: Interpret the unrecovered cases till 2021?

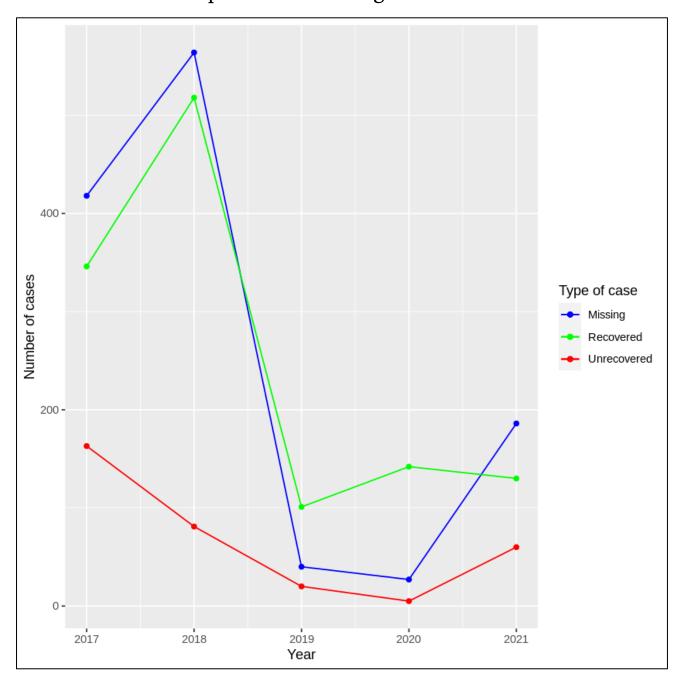


Description: In this Map, we have shown the state-wise total unrecovered cases till 2021.

Interpretation: As we can see from the Map, the states with maximum unrecovered cases are Madhya Pradesh, Maharashtra and West Bengal.

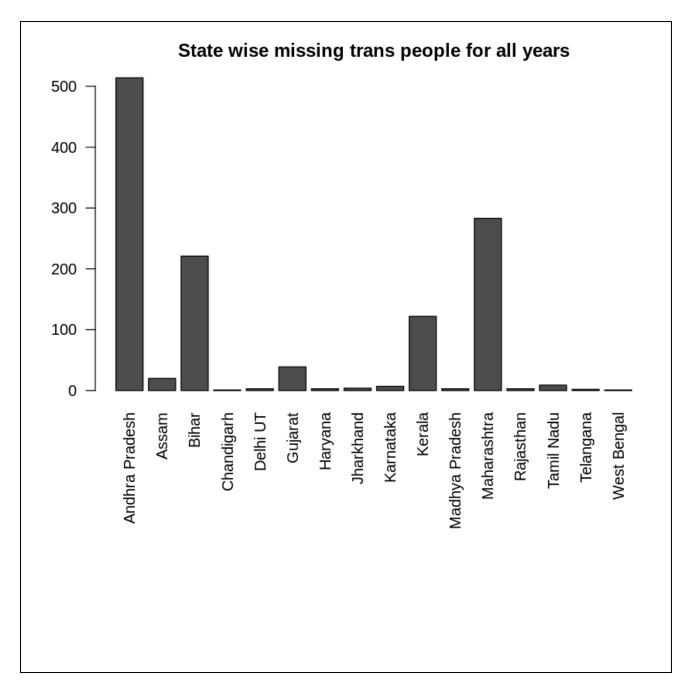
The states with lowest unrecovered cases are Ladakh, Kerela and many of the north-eastern states.

Ques 5: What is the Year-wise and State-wise missing, unrecovered and recovered cases patterns for transgenders?



(Year-wise)

Description: In this Line chart, we have shown the missing, recovered and unrecovered cases for transgender year-wise.



(State-wise)

Description: In this Bar Graph, we have shown the missing cases for transgender state-wise where we have picked the states with no. of cases not equal to zero.

### Interpretation:

- There is rise in the no. of missing and recovered cases in 2018.
- In 2019-2020 the no. of recovered cases is much higher than the no. of missing cases which shows the increased responsibility & positivity of the system and society.
- Andhra Pradesh has the maximum missing cases for transgender followed by Maharashtra & Bihar.
- Reasons for the such low no. of total cases for transgender (as compared to male & female) is either because their population percentage is low in society or because identifying themselves is still considered a 'Taboo in India'.
- Bill- <u>Protection of Rights for Transgender person</u> which was passed in 2019 for the empowerment of the Transgender society can be one of the reason for the decline of missing and recovered cases in 2019. As the particular section must be provided with government support.