

Judiciary in India

An open-access dataset of 80 **million** Indian legal case records

For this dataset in order to understand the terms in judicial system i met a lawyer according to the given definitions of each term by her i did this analysis

Story about dataset

The given dataset contains 80 million case records normally its too large but when compared to the years its not that much high. We know each case have their own story and own problem and justice.soo sad, But we don't need all these things to analyse the data. so lets move on to the dataset..

Brief introduction to report

This report presents an analysis of court data collected by the Data Development lab. The dataset includes information on over 80 million court cases filed in the India between 2010 and 2018. It includes variables such as the case type, the disposition of the case, the length of the case, and the number of defendants,number of petitioners. The goal of this analysis is to examine trends in the court system and identify patterns in the data that may provide insight into the functioning of the courts. The report also attempts to identify any limitations of the dataset, which could be useful for further research on this topic.

About dataset



In the given dataset, we have details about judges,information about cases from 2010 to 2018 and acts/sections used to solve that case, Here are the variables

justice-judges-metadata:

Ddl_judge_id	State code	Dist_code	Court_no	Judge position	Female judge	Start_date	End_date
--------------	------------	-----------	----------	----------------	--------------	------------	----------

Details of 98,478 judges presiding over more than 7000 district courts and subordinate courts of India.

2010-2018

ACTS sections data:

DDI_case_id	act	section	Bailable_ipc	Numbers_section_ipc	criminal
-------------	-----	---------	--------------	---------------------	----------

Act and section details of all ~80 million cases. These can be linked to the case files (cases_yyyy.dta) using the variable ddl_case_id.

Cases data:

Dd_case_id	year	State code	Dist_code	Court_no	cino	Judge position	Female defender	Female_adv_Defender
Female_adv_petitioner	TypeName	Purpose name	Disp_name	Date_of_filing	Date_of_decision	Date_first_list	Date_last_list	Date_next_list

Records of ~80 million legal cases that were filed between the years 2010-2018, spanning more than 7000 district courts and subordinate courts of India

I am trying to follow these steps

Steps to preprocess our given data:

1. **Identify the data sources:** The data sources for judiciary data obtained from Data development lab
2. **Assess the data:** Assessing the data involves examining the structure, contents, and quality of the data. This includes looking for missing or incomplete data, checking for data accuracy and consistency, and verifying the data formats.
3. **Clean and transform the data:** Cleaning and transforming the data involves removing any unnecessary information, correcting any errors, and transforming the data into a format that can be used for analysis. This includes standardizing data formats and eliminating outliers.
4. **Analyze the data:** Analyzing the data involves using statistical and machine learning techniques to identify patterns and insights in the data. This includes visualizing the data to gain insights and using predictive models to make future predictions.
5. **Present the results:** Presenting the results involves summarizing the findings and presenting them in a way that is easy to understand. This can involve creating charts, graphs, and tables to summarize.

Cleaning and transforming the data:

- In the given data there are so much unclear information in every file. So we need to analyse what are the information unclear in the given dataset.
- In judges data I found some unclear data about judge gender as -9999 unclear, -9998 unclear or empty so i replaced -9999 as unclear, -999a8 as MV(missed value) empty as unclear and changes variable names of gender like female_judge in to gender and replaced 1 with female 0 with female. and replaced unclear years with 1900. This is because the year 1900 is a commonly recognized value that can be easily recognized as a placeholder, and it is also a year that is far enough in the past that it is unlikely to be used as a valid year in real data. Additionally, in some cases the software may not

support or handle a null or missing value correctly, it is common to use such a placeholder value instead.

- In cases data also I arranged gender type columns as like above and I replaced state_code numbers with real names bu using state_keys section.
- I renamed the columns into more informative like this way

justice-judges-metadata:

Ddl_judge_id	State	Dist_code	Court_no	Judge position	Gender	<u>Start_date</u>	<u>End_date</u>
--------------	-------	-----------	----------	----------------	--------	-------------------	-----------------

Details of 98,478 judges presiding over more than 7000 district courts and subordinate courts of India.

2010-2018

ACTS sections data:

<u>DDL_case_id</u>	act	section	<u>Bailable_ipc</u>	<u>Numbers_section_ipc</u>	criminal
--------------------	-----	---------	---------------------	----------------------------	----------

Act and section details of all ~80 million cases. These can be linked to the case files (cases_yyyy.dta) using the variable ddl_case_id.

Cases_data:

Dd_case_id	year	State	Dist_code	Court_no	cino	Judge position	Defender gender	Petitioner gender	Advocate defendant gender
<u>Advocate petitioner gender</u>	Case type	Case purpose	disposition	Filing date	Decision date	<u>Date_first_list</u>	<u>Date_last_list</u>		<u>Date_next_list</u>

Records of ~80 million legal cases that were filed between the years 2010-2018, spanning more than 7000 district courts and subordinate courts of India

- I removed columns which were not useful for this analysis those are cino,Date_last_list,Bailable_ipc, Numbers_section_ipc etc according to the particular analysis i managed the data.

Analysis of the dataset reveals several important trends in Indian legal cases:

- The Supreme Court of India is the most popular court for legal cases, with nearly 40 million cases in the dataset.
- Civil cases are the most common type of legal case, representing over 60 million cases in the dataset.
- The most common type of civil case is family law, representing nearly 20 million cases.
- The number of criminal cases in the dataset is much lower, representing only about 11 million cases.
- The average length of legal cases is increasing, with the average case taking about 10 months to resolve.
- The most common type of judgement is a "dismissal", representing nearly 20 million cases.
- On judge who were in Gujarath worked in district and sessions court have higher service time 22 Years, 3 Months and 4 Weeks.

Now move on to the most awaiting moment----->

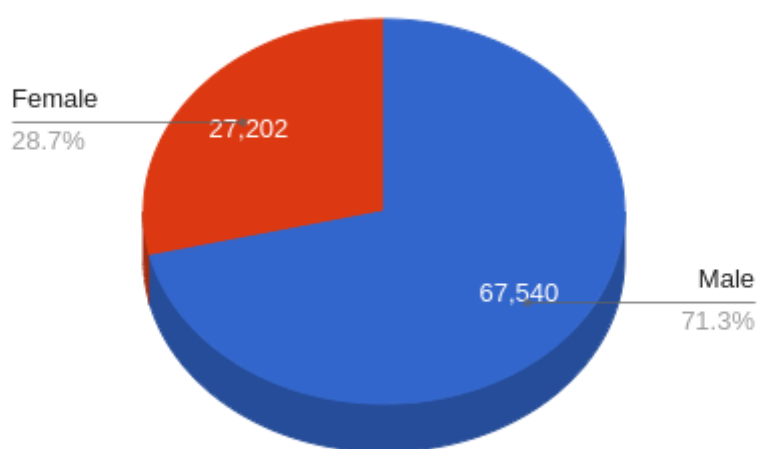


Exploratory data Analysis

Let's start with the judges data from the below graphs we can get the idea of how many judges in andhrapradesh and telangana from 2010 to 2018 and their average service time in every year

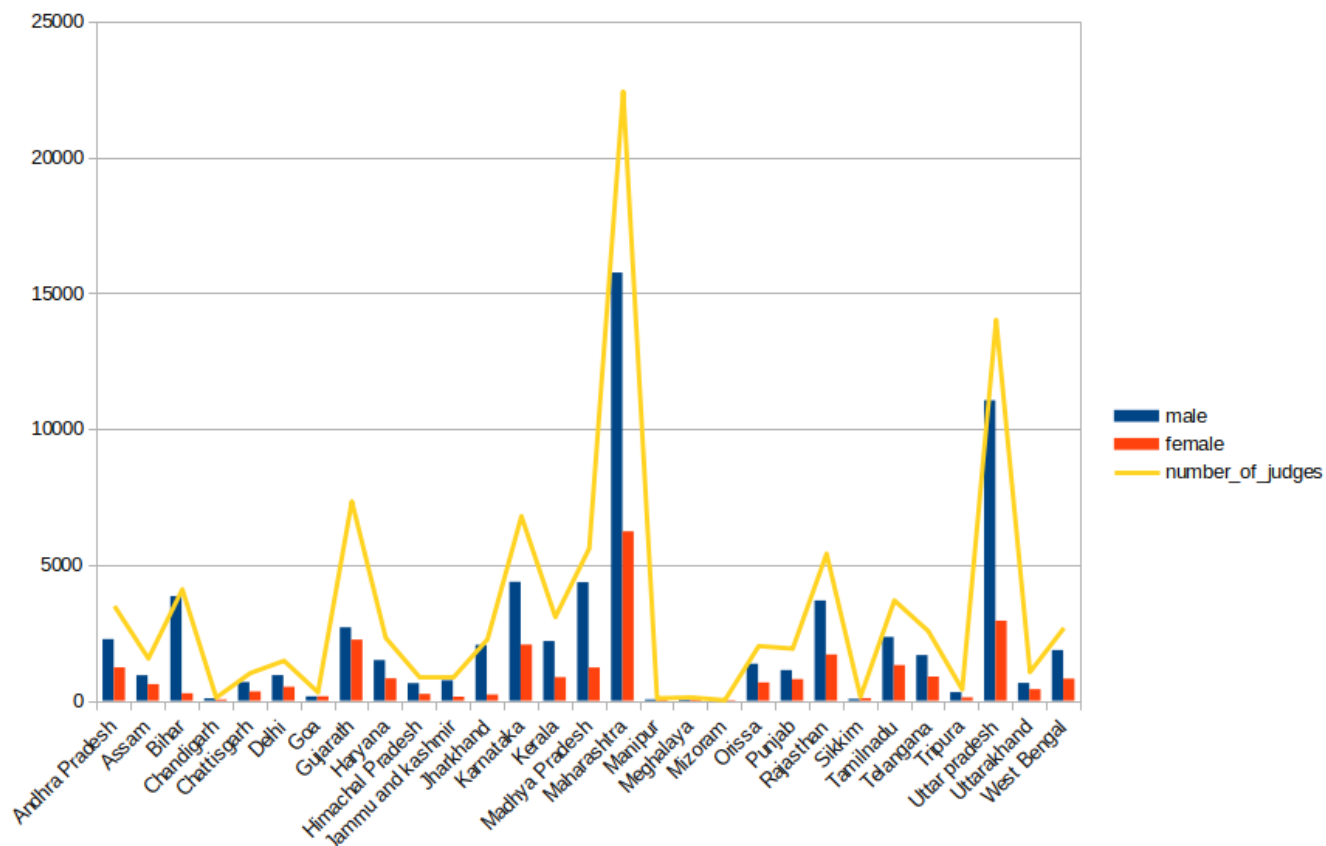
Below graph shows percentage of male and female judges

Judges Gender classification



Its shows male judges higher than the female judges but if we consider years female judges are going to approximately equal to the female judges

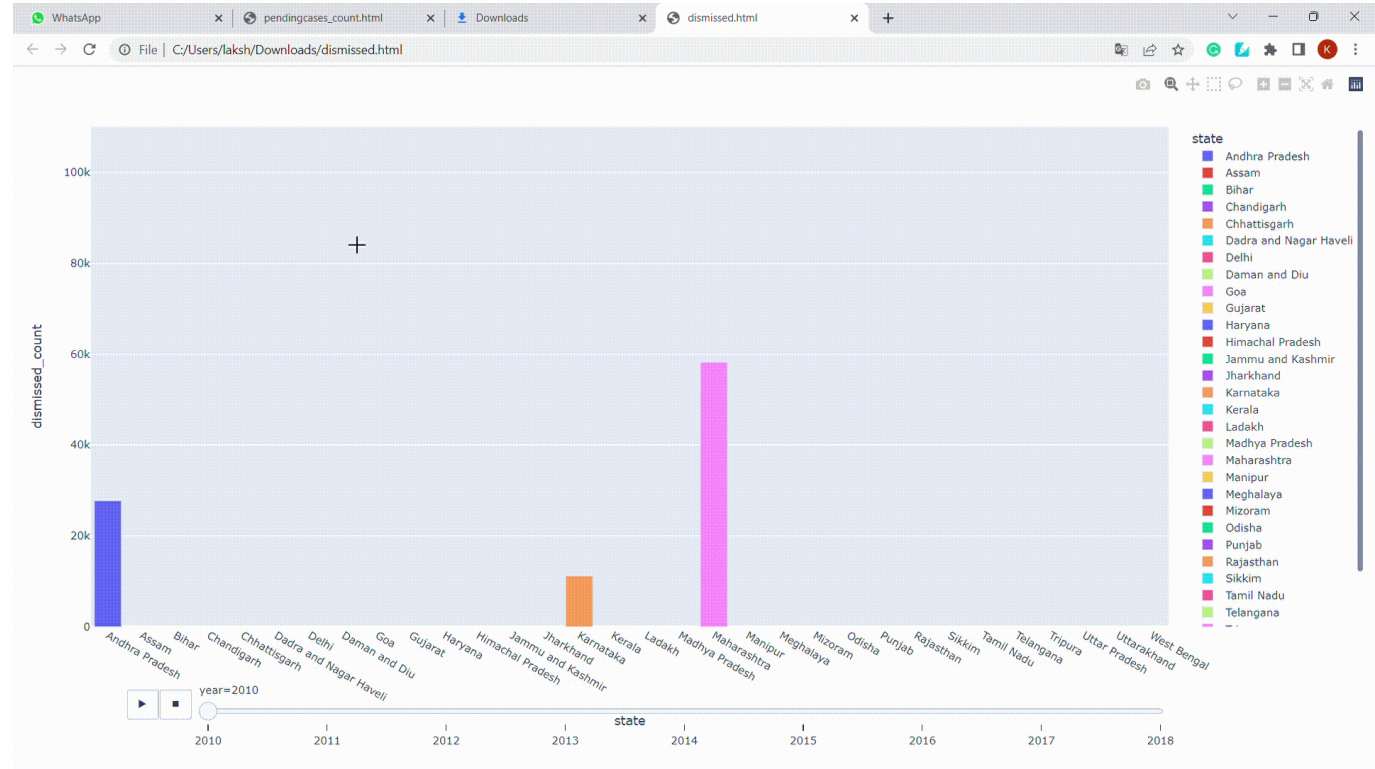
Lets see total judges,and compare male adn female judges with total judges using graph



Dismissed cases

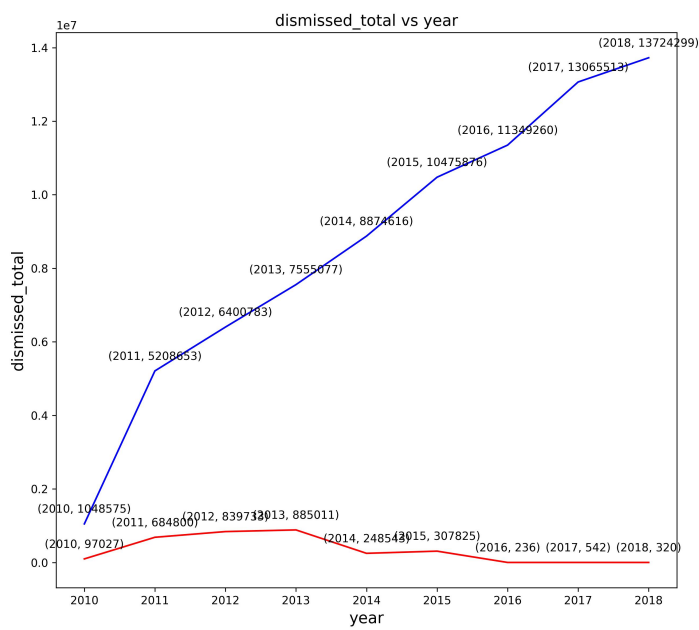
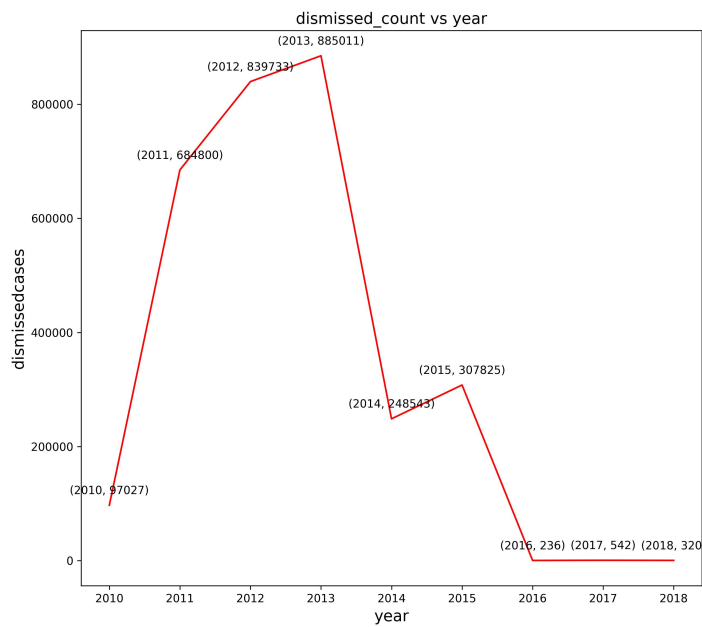
Before going to the analysis just try to make sure which type of data i am giving through this like I analysed total how many dismissed cases from 2010 to 2018 how is this varying from 2010 to 2018 and also i analysed data about prison judgement, criminal cases, non criminal cases , withdrawal cases and total cases. and I analysed the judges data like in each state how many judges are there from 2010 to 2018 and what is the ratio of male and female judges.

This is the dismissed cases variation from 2010 to 2018 in every state in india.

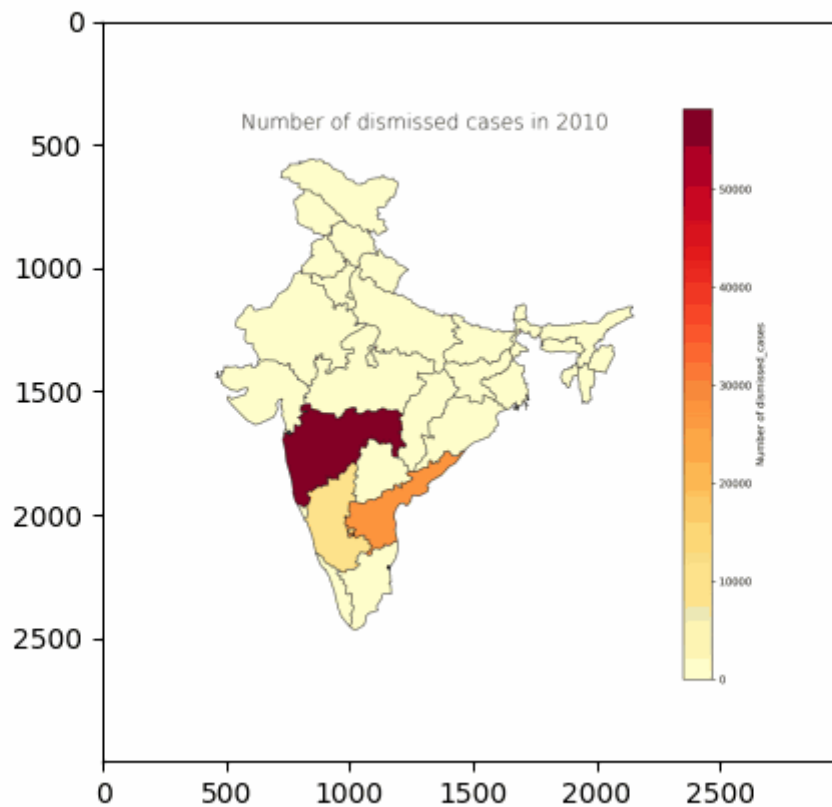


[click here to see perfect view of this graph](#)

change from 2010 to 2018



Here red line represents dismissed cases andd blue line represents total cases in every year In order to get clear information we need more appropriate way to show data so here is the visualized data iusing map.

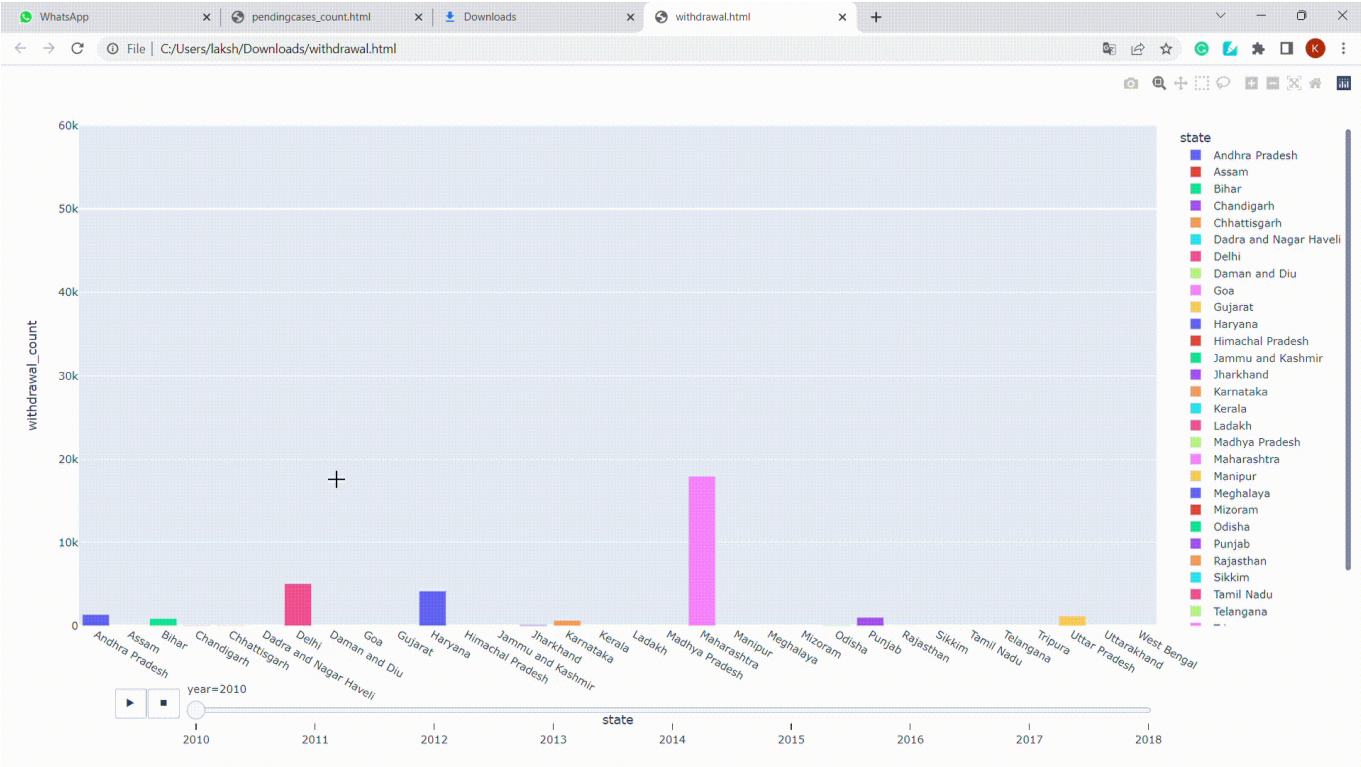


From this data we can see dismissed cases count has enormously increased in 2013 but by the end of 2018 it decreased.

Withdrawal cases

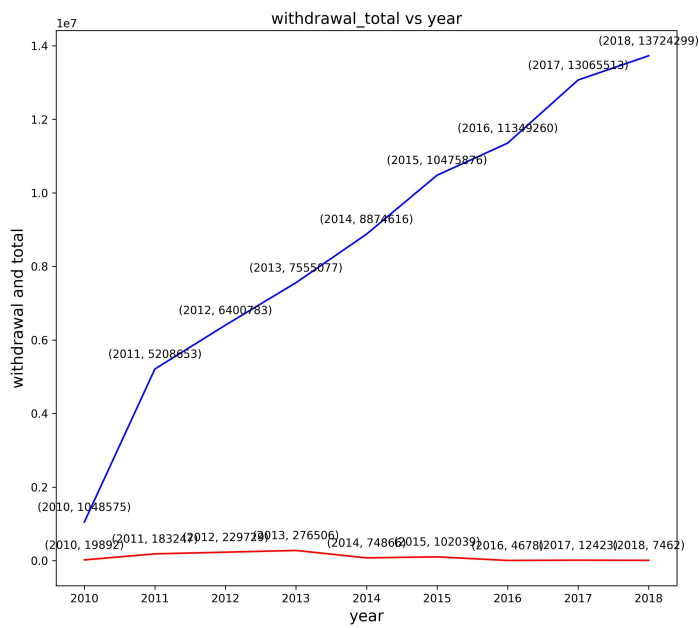
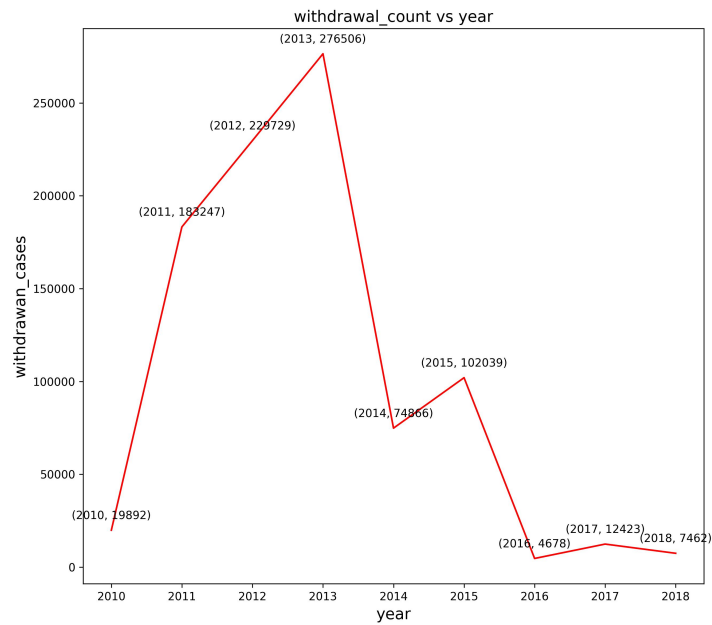
In india because of various reasons people withdrawing cases after filing reasons like compromise between petitioner and defendent or other reasons. Here we have data how this withdrawing cases has changed

from 2010 to 2018



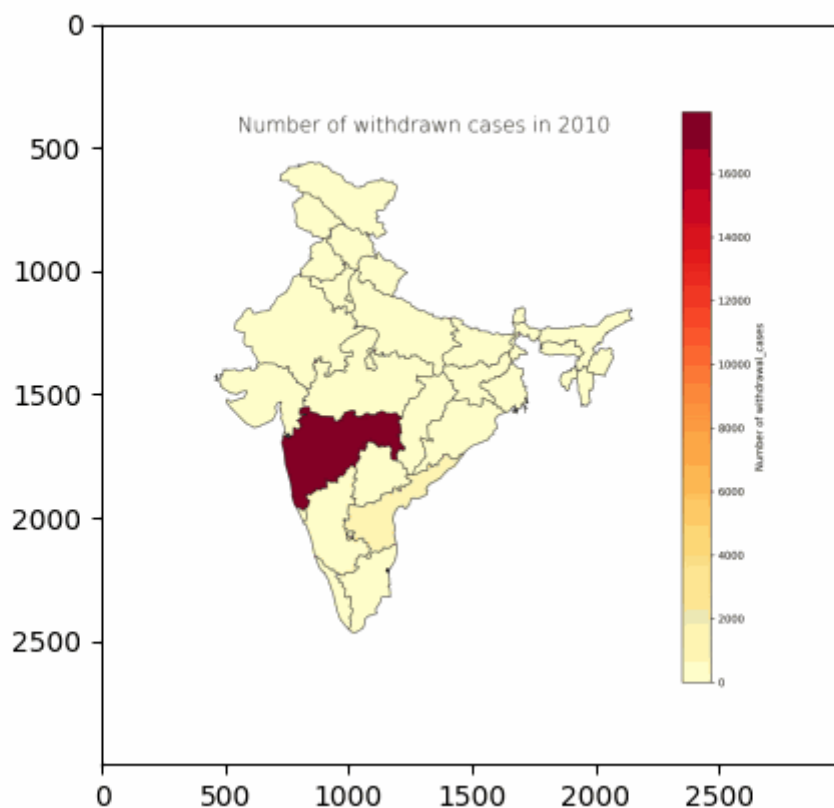
[click here to get better view in webpage](#)

How its changing from 2010 to 2018.....



Here red line represents dismissed cases
and blue line represents total cases in every year

In order to get clear information we need more appropriate way to show data so here is the visualized data using map.

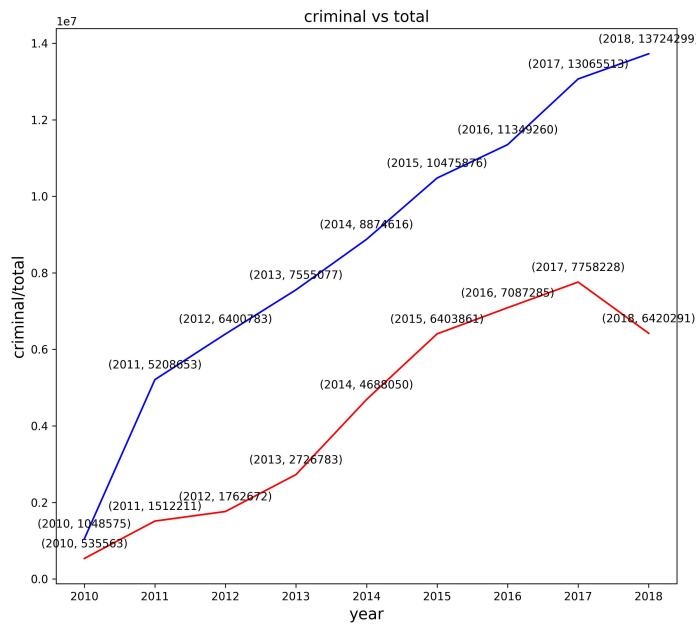
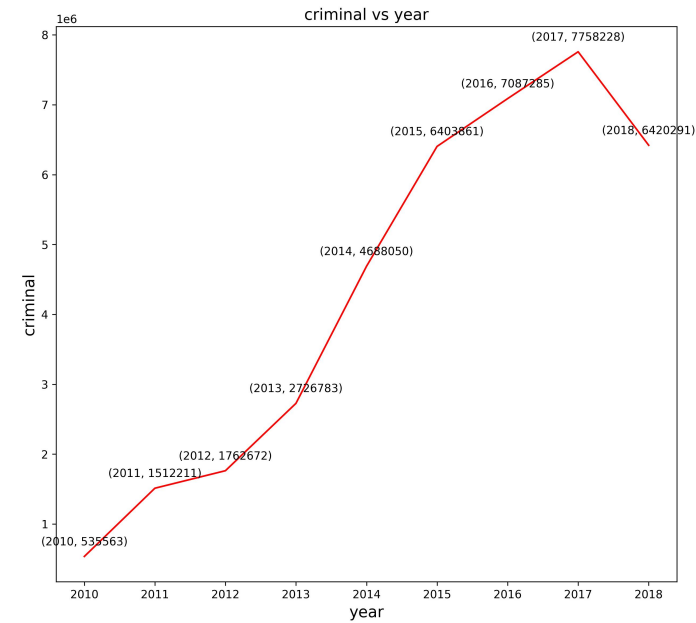


From these graphs we can see that in 2013 there are high range of withdrawal cases so might be there a one reason for this may be political issues or other type of issues in society and family

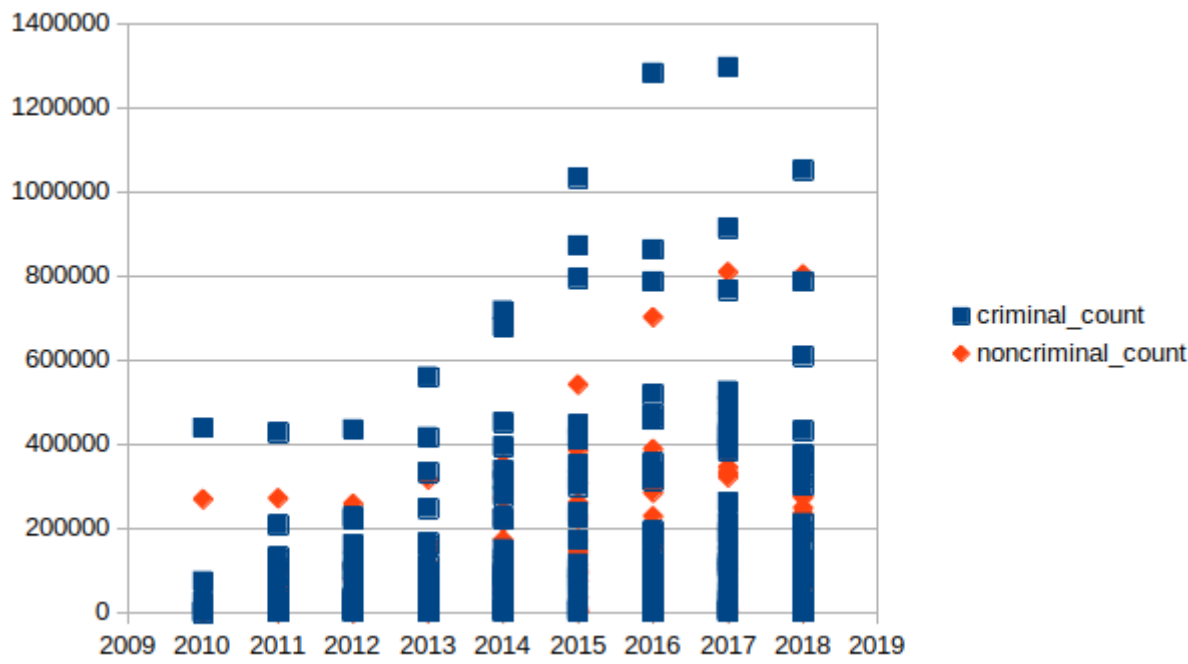
Criminal cases

In judiciary system there are various type of cases one of them is criminal cases, In this dataset there are 65 million criminal cases according to data developement lab report. So, now lets see criminal cases increased or decreased every year from 2010 to 2018 Here is the html animated representation you can go through this link to get more clearer picture [click here.](#)

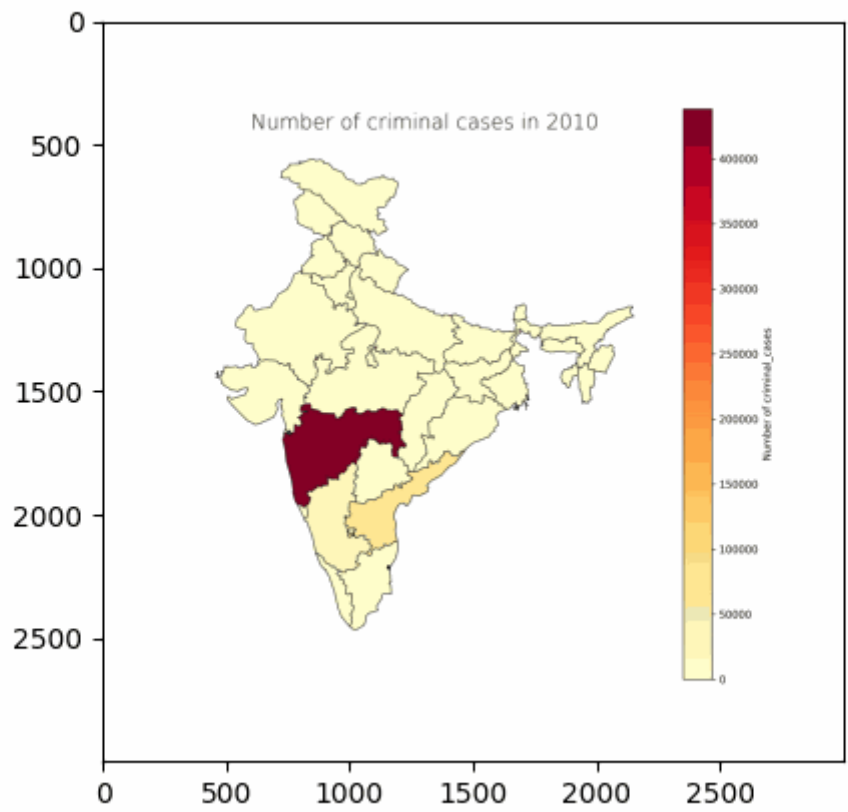
How it is changing from 2010 to 2018 when compared to year and total cases in every year



Here blue line represents total cases and red line represents criminal cases



To get the more clearer view of this criminal cases representation here iam giving map data from 2010 to



2018

From these graph we can assume that in earlier days means till 2011 most of the cases are criminal cases but later there are many type of cases like divorce cases,land cases, family cases etccc..

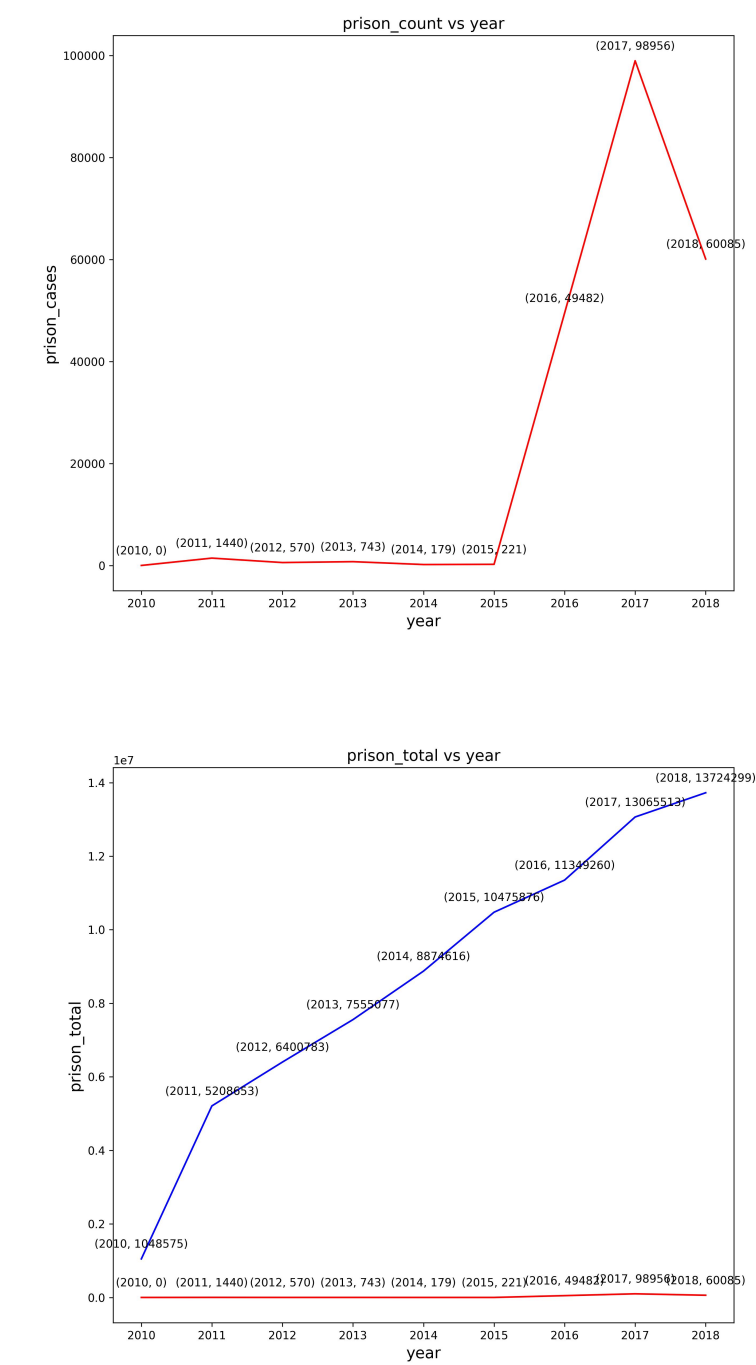
Prison cases

Every year in total cases there are many people going to prison as their final judgment because of severe cases. SO lets see how many cases have final judgemnt have prison as 1year or 2years etc..



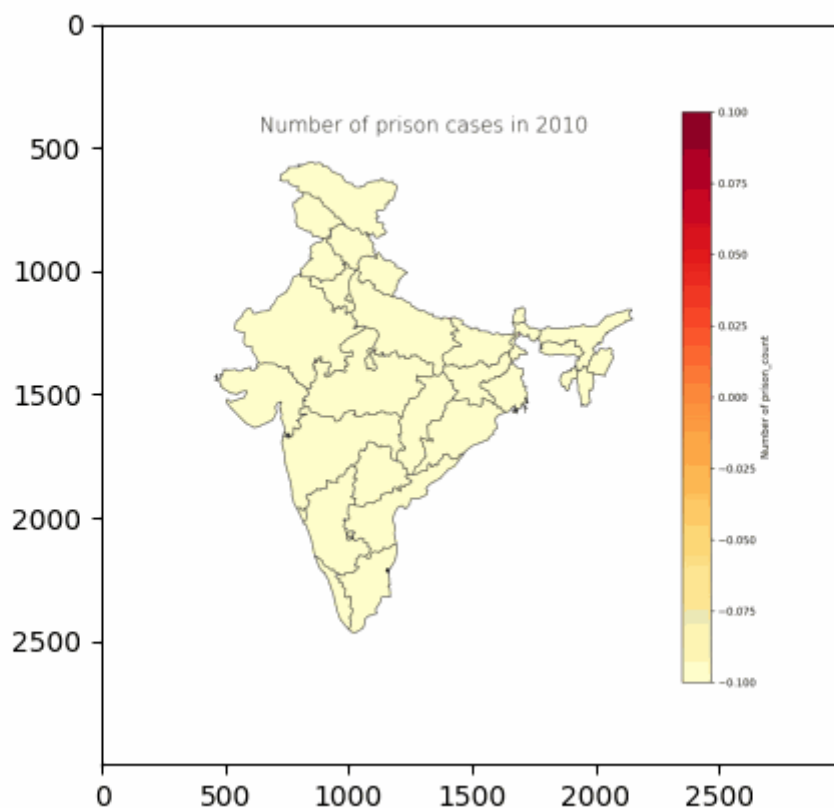
To get the more clearer picture of this page [click here](#)

Lets see total number of cases vs every year and see how it is varying year by year.



Here blue line represents total cases and red line represnts prison cases

To get the more closer view by visualisation i used map plotting from 2010 to 2018

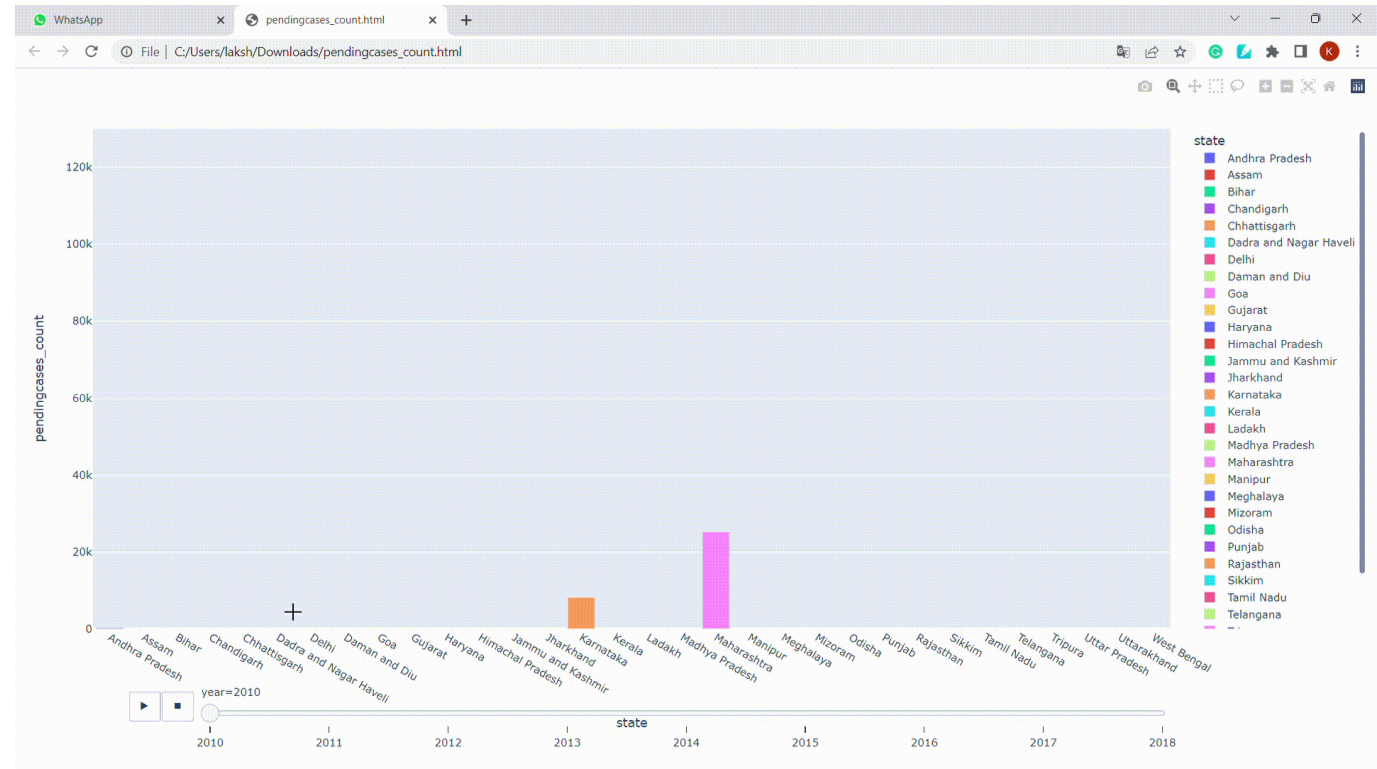


From these visualizations we can say that when compared to the total cases prison judgemnet cases are lower but rapidly increased prison cases in 2016 and 2017 when compared to the previous years but slightly decreased in 2018 when compared to the 2017 may be this is becuae of some awarness among the people about driving, drinking or other type of normal cases. If these awarness increases then atleast 10% small type of cases will decreases

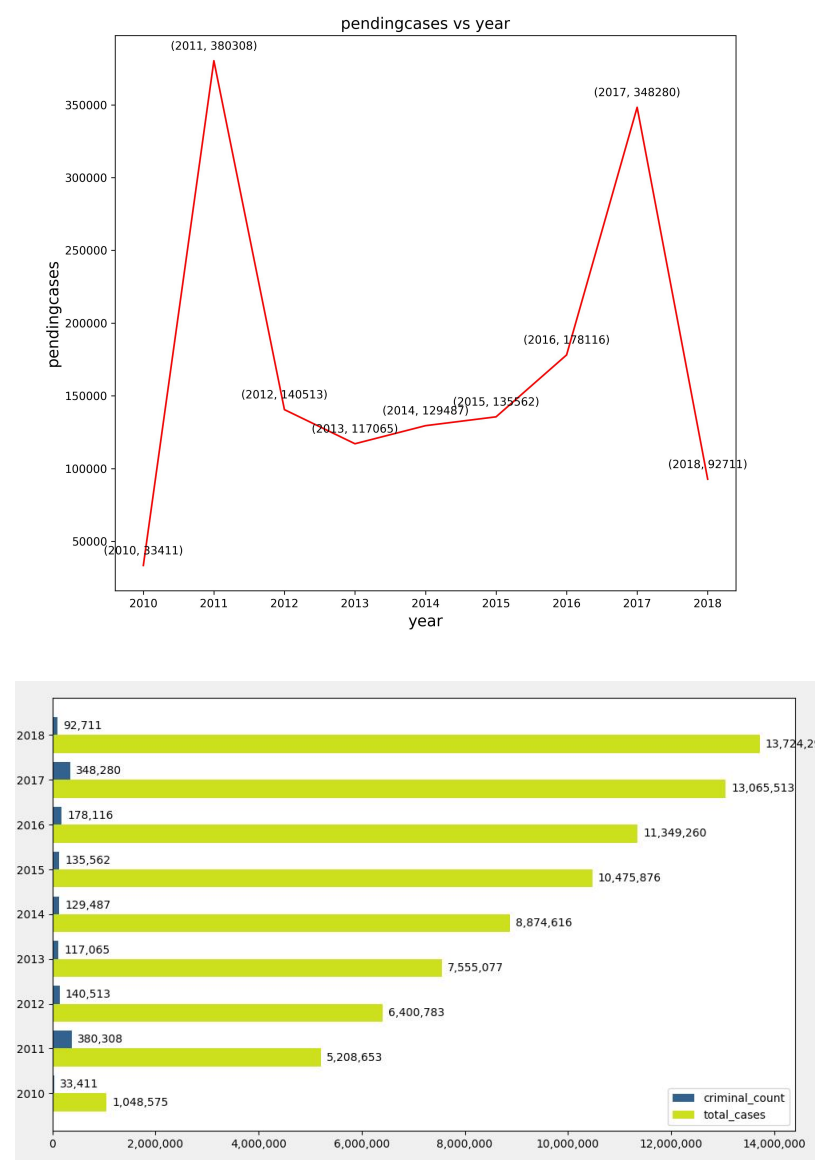
Pending cases

In any country , in any state pending cases are common in any court. Pending cases will raise due to increased number of cases or lack of evidence or judges/ client not availbale to each other to deal the case or some financial issues.

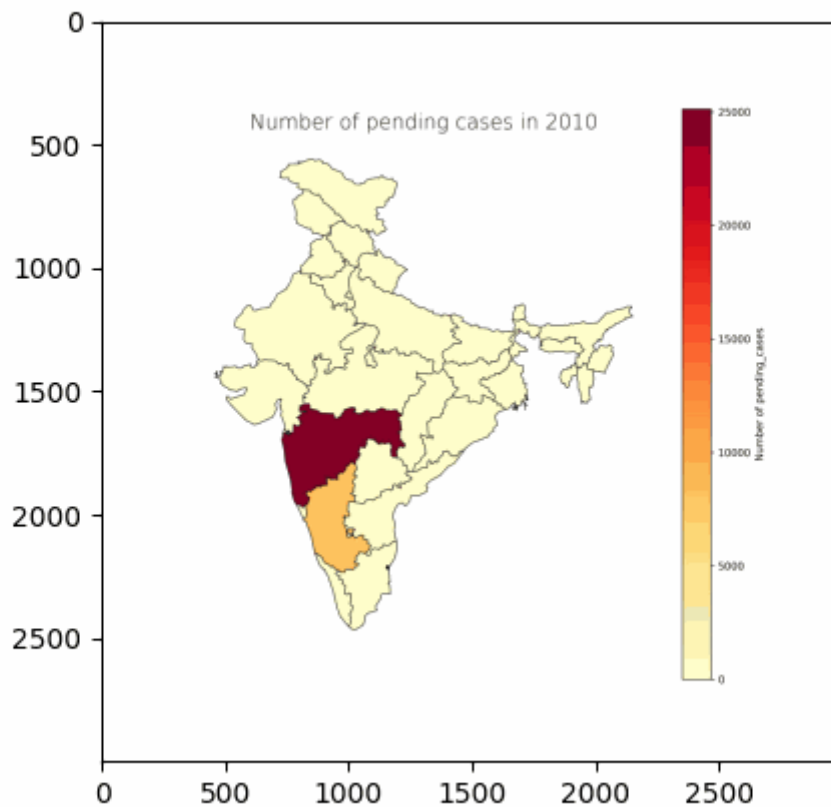
Now lets see how many pending cases in every state from 2010 to 2018 in every year.



Now lets see graph which tells about year and pending cases.. and also comparison with total cases in every year.



By luck indian judicial system working hard so very few cases are pending in every year now lets see which state having higher pending cases



From these graph we can say that karnataka having higher number of pending cases. In the beginning maharashtra having higher pending cases but later pending cases count have been decreased gradually.

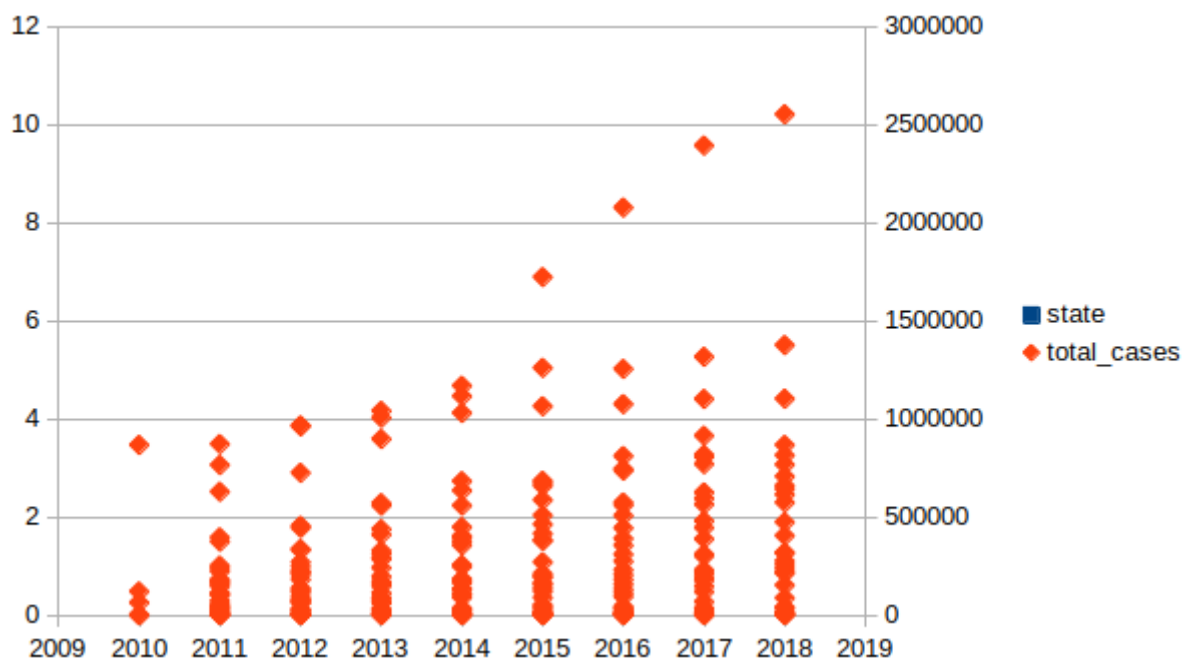
But in 2017 there are huge number of pending cases but later it was enormous decreased by the end of 2018

Total cases

Till now we have total pendingcases, withdrawn cases, criminal cases, dismissedcases now lets see total cases becuse it is very important to see total when comparing other values.

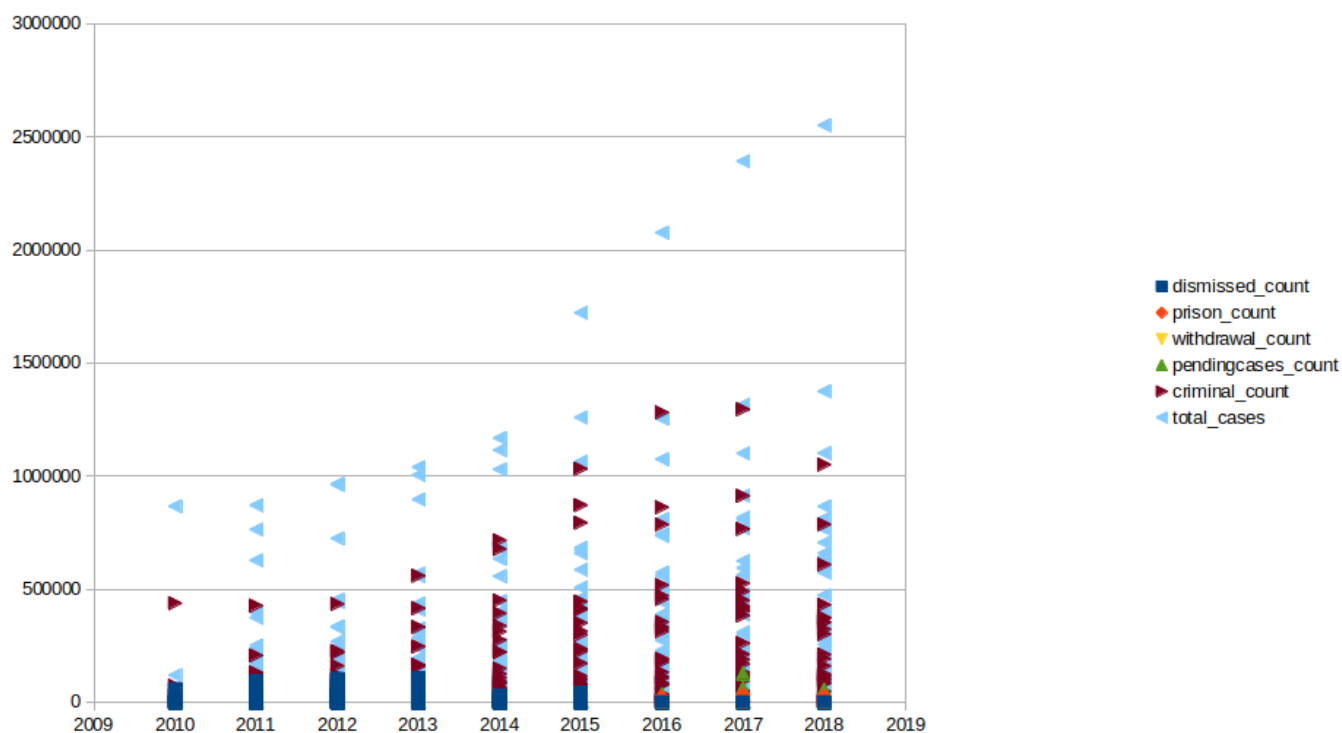


Now lets see how many cases in India every year from 2001 to 2018.



This graph represents total cases count in every year.

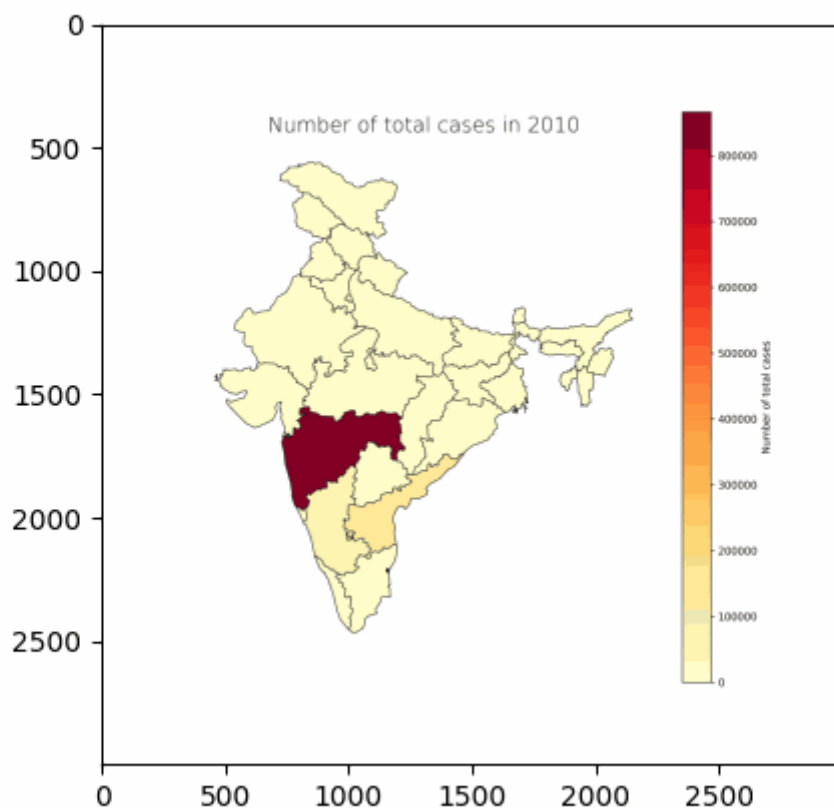
Now let's see the relation between pending cases, withdrawn cases, criminal cases, dismissed cases when compared with the total cases.



Now let's see the count of pending cases, withdrawn cases, criminal cases, dismissed cases when compared with the total cases using a stacked bar.



Till now we have seen these graph comparison with years from 2010 to 2018. Now let's see each state every year in India from 2010 to 2018.



Finally total cases are increasing parallelly judges and other type of courts and cases are also increasing.

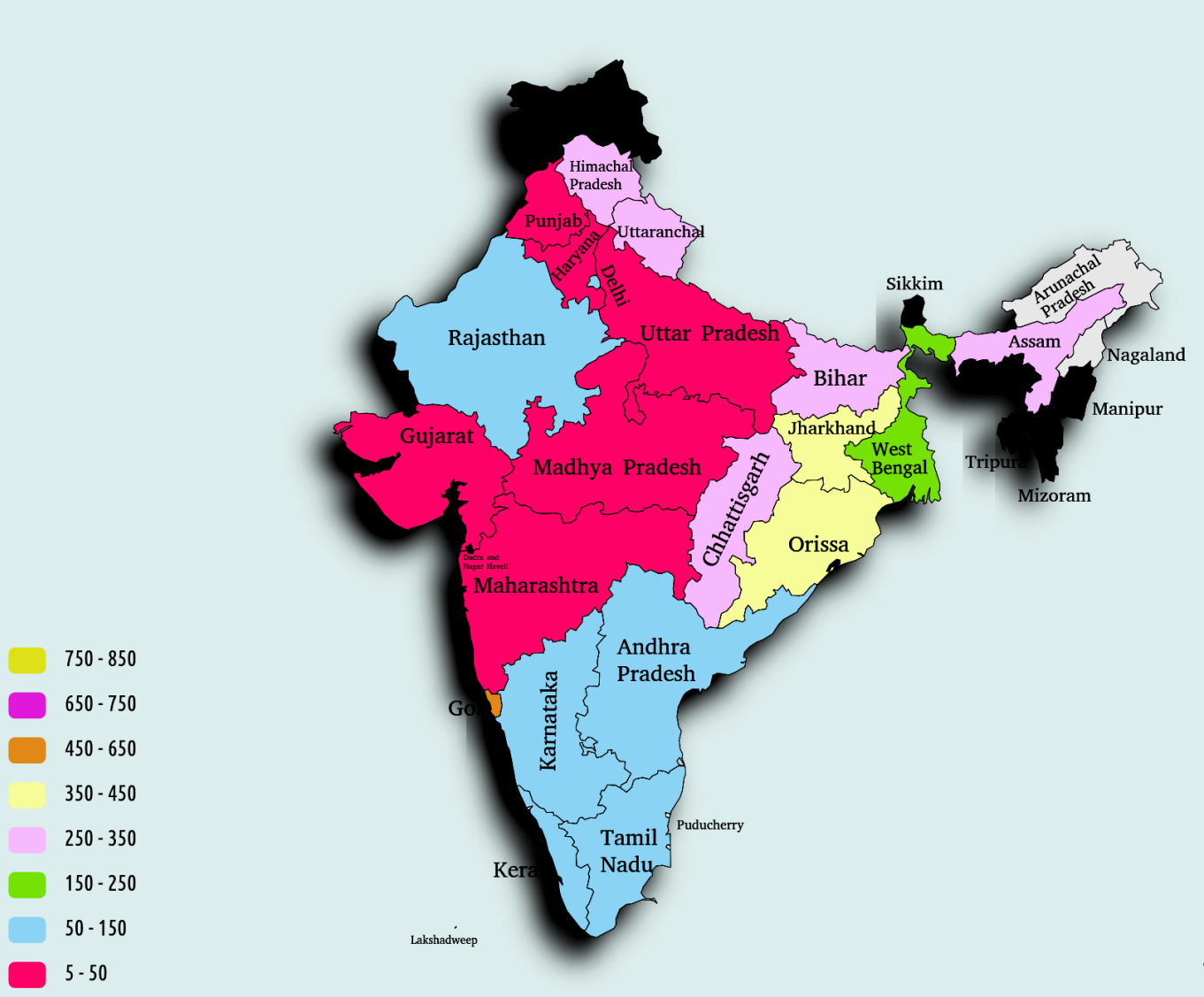
Duration

Now let's move on to the duration of cases here I took three types of durations

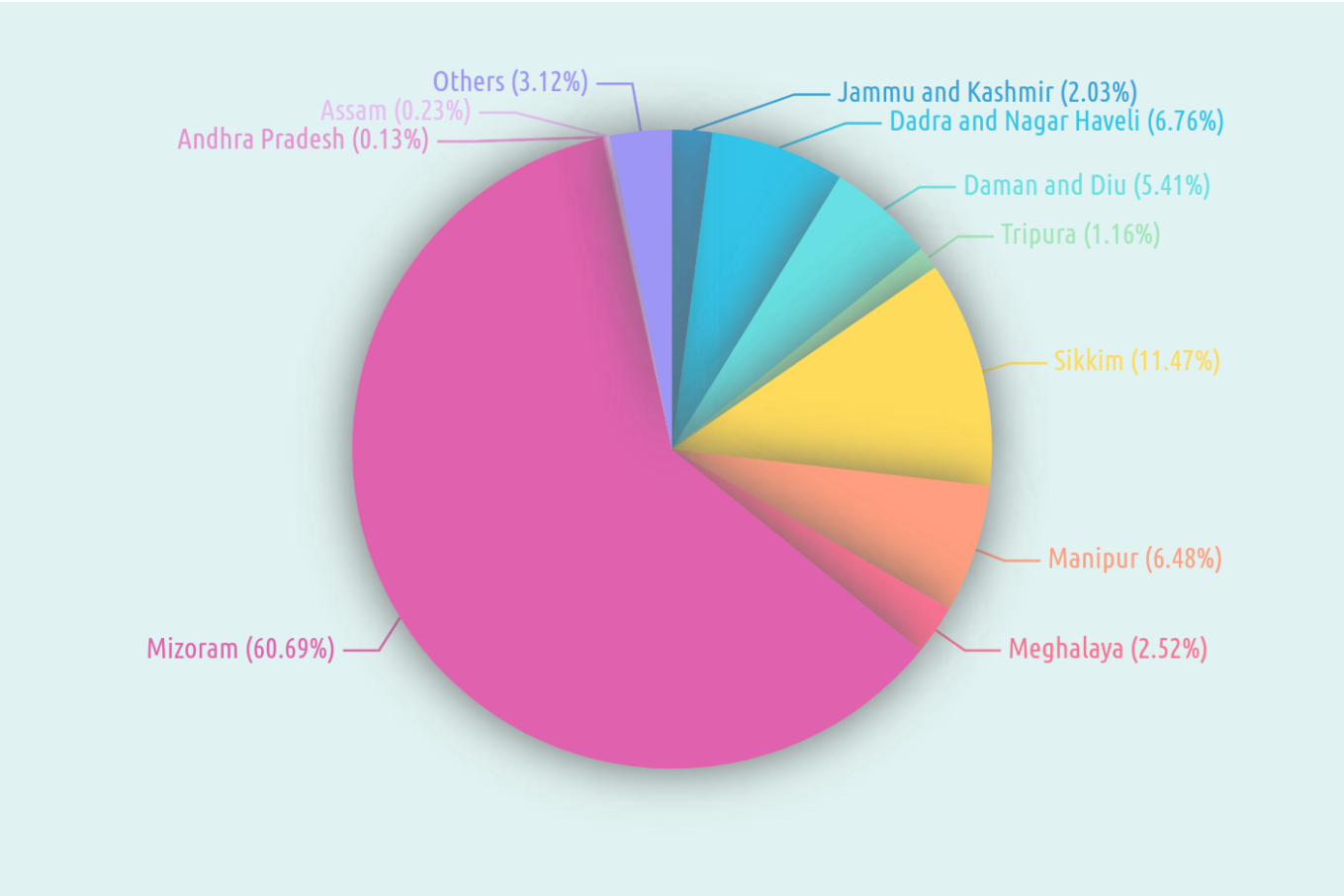
1. Time taken to reach first hearing after filing.
2. Time taken to close the case after decision.
3. Time taken to take decision after first hearing.

Time taken to reach first hearing after filing:

Below graph shows average number of day taken to reach court after filing by using 2010 to 2018 data



Below pie chart shows Top 10 states taking high number of days to reach court



Time taken to close the case after deciosn:

Time taken to take decision after first hearing: