

Rehabilitation of Prisoners in India

FIT5147 : Visualisation Project

Vachan Anand

29880734

Table of Content	
Introduction	3
Design	4
Implementation	6
User Guide	10
Conclusion	13
Reflection	14
Five Design Sheet	15

Introduction

One of the reasons for punishments given to the criminals is that the prisoners do not commit the crime again but sadly this is not the outcome for majority of the cases. The purpose of this assignment is to understand different factors that support or hinder the process of rehabilitation in prisons in India. The data used for the tasks is criminal records for years between 2000 and 2010. The country chosen to understand the relationship among different factors is my country of origin i.e. India.

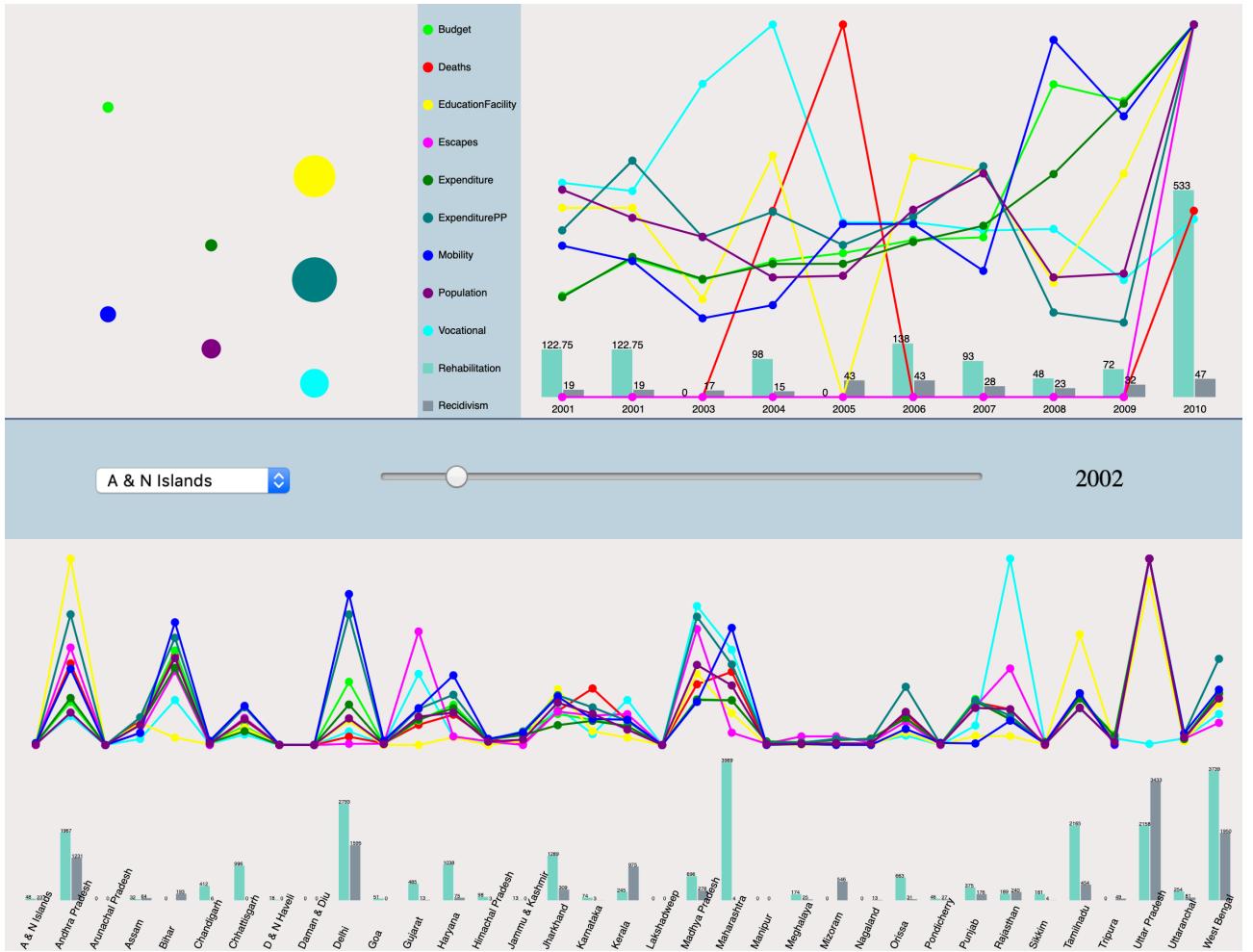
A nation spends millions of dollars on improving the condition of prisons that would in turn support the rehabilitation process of prisoners. But this money if spent without any prior knowledge of the factors that might be effecting the process is waste as the prison system though may be improving still the prisoners are on the same page as before. The visualisation produced in the task not only helps us explore different factors that are affecting the rehabilitation process but also shows the extent to which they may be affecting the process. To add on, the visualisation compares the data between different states within different period of time to draw onto a conclusion.

Personally, I feel it is very important to understand the criminal records and the conditions prisoners are living in to make any improvement with rehabilitation. Only if there was a way to reduce the number of reoffenders and improve number of rehabilitated prisoners, we could be sure that the prison system is efficient enough.

The intended audience for the visualisation is the government authority, specifically in the state government that is responsible to take care of the prisons and prisoners in their region. Using the visualisations they can draw conclusion on the domain they should focus on and invest in to improve the efficiency of the prison system in their particular state.

Design

The visualisation is designed such that the user can easily derive the conclusions on factors he/she may be exploring. The design consists of 3 interrelated graphs. Each graph gives certain level of information about the prison system which when combined can help the user derive a conclusion.



Firstly, a bubble graph is used to understand the factors that may be affecting the rehabilitation process. The graph can be changed across 35 states and 10 time frames ranging from 2001 to 2010. The graphs is updated based on the selection made by the user. The bubbles in the graph indicates different factors the that affect the rehabilitation process. The size of the bubble indicates the extent to with the factors support the process. Each factor is distinguished with different colour that is consistent throughout the graph to avoid any confusion while usage.

Once the user is familiar with the factors that might be affecting the process, the second graph i.e. the line chart gives the user summary of the data within a particular state over a period of 10 years. For this graph the user can choose a state of interest and the data produces a visualisation that

indicates different factors with help of a line and the topic of interest i.e. rehabilitation and recidivism rate in that particular state with help of bars. This graph aids the user to see changing patterning in factors with respect to rehabilitation rate. The trend in the line such as increase in rehabilitation with respect to factors indicates a strong relation and supports the idea of investing more in such factors. From the graph above, education facilities seems like a very strongly related factor for A & N island.

Finally, once the user understands the factors that might be prevalent in their particular state, the third visualisation gives the overview of relation between factors and rehabilitation process throughout the country.

Alternative Method and Justification

The visualisations chosen are such that the user gets in-depth knowledge of factors and their relation on just a click. A variety of other graph very considered such as a pie chart to explore demographic factors of the prisoners or radar graph to draw any conclusion on the relations but were rejected as the quality of exploration and the insight given was compromised in case of a pie chart and the radar graph became densely populated as the number of factors and the states to be compared were high. Hence, after considering the factors above the final design was chosen as above.

Implementation

The data used for this assignment has been taken from dataworld.com. The data was verified from the Indian government website to make sure that the insights obtained from the dataset was reasonable enough to draw any conclusions.

Data Sources :

- <https://data.gov.in/dataset-group-name/prison-statistics>
- <https://data.world/rajanand/prison-in-india>

The tasks starts by reading the relevant data. The data is spread across 11 sheets that is wrangled and cleaned for visualisation purpose.

```
budget_df = read.csv("Budget.csv")
edu_facility_df = read.csv("Education_facilities.csv")
edu_df = read.csv("Education.csv")
inmate_expenditure_df = read.csv("Expenditure_on_inmates.csv")
expenditure_df = read.csv("Expenditure.csv")
escapes_df = read.csv("Inmates_escapee.csv")
population_df = read.csv("jail_wise_population_of_prison_inmates.csv")
movement_df = read.csv("Movements_outside_jail_premises.csv")
recidivism_df = read.csv("Recidivism.csv")
rehabilitation_df = read.csv("Rehabilitation.csv")
vocation_df = read.csv("Vocational_training.csv")
```

As each table consist a number of rows that are not relevant or consistent, those rows are deleted before any visualisations are made. For consistency purpose the data is filtered that lies in the year s between 2001 and 2010 as majority of the data lies in this period.

```
clean_data <- function(df){
  df = df[which(df$year <= 2010),]
  df
}

population_df = clean_data(population_df)
budget_df = clean_data(budget_df)
edu_facility_df = clean_data(edu_facility_df)
edu_df = clean_data(edu_df)
inmate_expenditure_df = clean_data(inmate_expenditure_df)
expenditure_df = clean_data(expenditure_df)
escapes_df = clean_data(escapes_df)
movement_df = clean_data(movement_df)
recidivism_df = clean_data(recidivism_df)
rehabilitation_df = clean_data(rehabilitation_df)
vocation_df = clean_data(vocation_df)
```

Once the data is cleaned it is aggregated for visualisation purposes. The data consist of a lot of columns that we are not interested in. We remove those columns before merging the aggregated data into one data frame.

```
population_ysw = aggregate(population_df[, 18], list(population_df$year,population_df$state_ut_name), sum)
population_sw = aggregate(population_df[, 18], list(population_df$state_ut_name), sum)
budget_ysw = aggregate(budget_df[, 4], list(budget_df$year,budget_df$area_name), sum)
edu_facility_ysw = aggregate(edu_facility_df[, 7], list(edu_facility_df$year,edu_facility_df$state_name), sum)
rehabilitation_ysw = aggregate(rehabilitation_df[, 7], list(rehabilitation_df$year,rehabilitation_df$state_name), sum)
inmate_expenditure_ysw = aggregate(inmate_expenditure_df[, 4], list(inmate_expenditure_df$year,inmate_expenditure_df$area_name), sum)
expenditure_ysw = aggregate(expenditure_df[, 4], list(expenditure_df$year,expenditure_df$area_name), sum)
deaths_ysw = aggregate(deaths_df[, 6], list(deaths_df$year,deaths_df$state_name), sum)
escapes_ysw = aggregate(escapes_df[, 6], list(escapes_df$year,escapes_df$state_name), sum)
movement_ysw = aggregate(movement_df[, 4], list(movement_df$year,movement_df$area_name), sum)
recidivism_ysw = aggregate(recidivism_df[, 4], list(recidivism_df$year,recidivism_df$state_name), sum)
vocation_ysw = aggregate(vocation_df[, 4], list(vocation_df$year,vocation_df$state_name), sum)
```

Normalisation on data plays a very important role in visualisation as the value of the factors may be comparable within the factors but may not make sense while comparing one factor with another. Keeping this in mind the aggregated data is further normalised to make the factors comparable.

```
n=data.frame()
for (i in c(1:length(all_states))){
  t = a[a$population_ysw.Group.2 == all_states[i],]

  for (j in c(3:length(t))){
    mi = min(t[,j])
    ma = max(t[,j])
    t[,j] = (t[,j] - mi)/(ma-mi)
  }

  n = rbind(n,t)
}

n[is.na(n)] = 0

write.csv(n, file = "manipulated_n.csv")
```

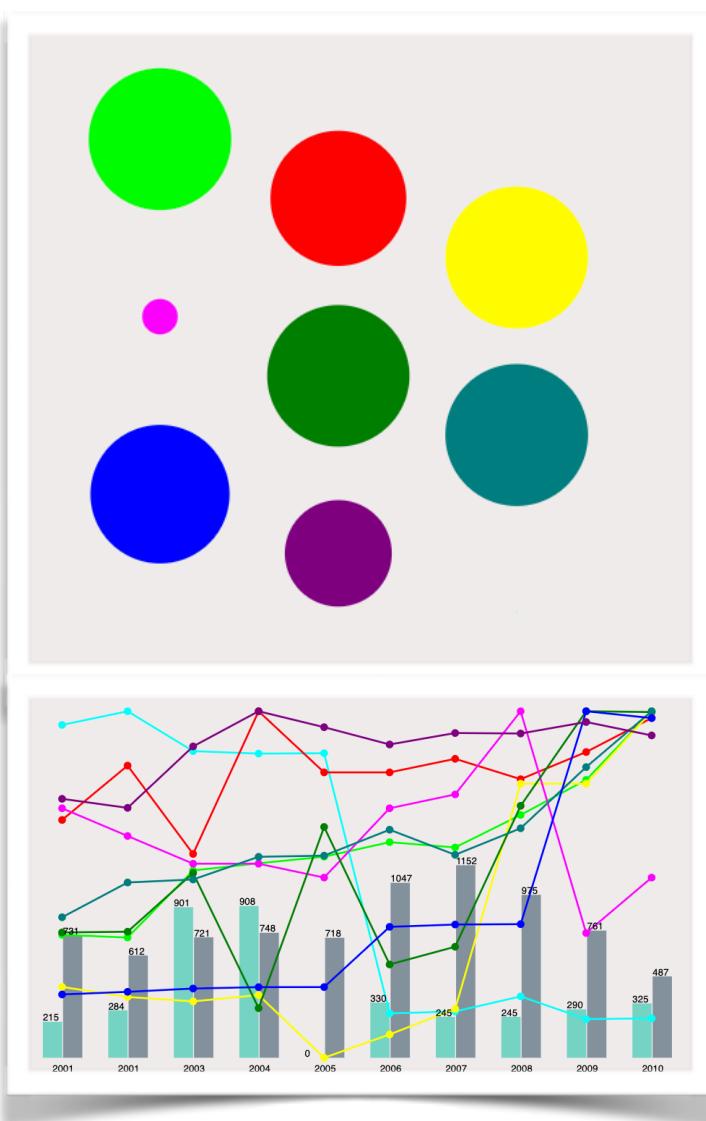
The visualisations are made using

D3 and Javascript : For creating Interactive visualisations using basic SVG element like circles, lines and rectangles.

HTML : For designing the layout of the website.

CSS : For decorating the website for user interface.

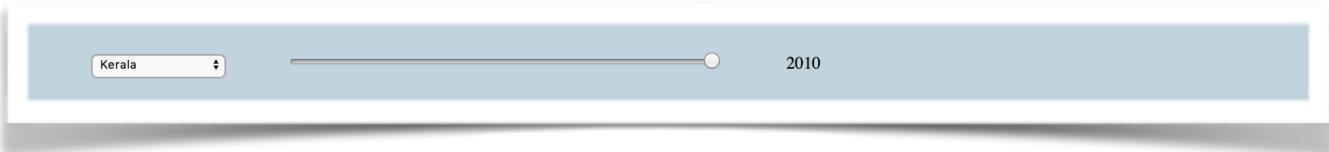
D3 was chosen as it gives a lot of freedom in terms of visualisation effects. Personally, as the industry requires powerful visualisation skills and D3 being one of the most powerful visualisation tool, I wanted to acquire some basic interactive visualisation skill using D3 and Javascript to build on it later.

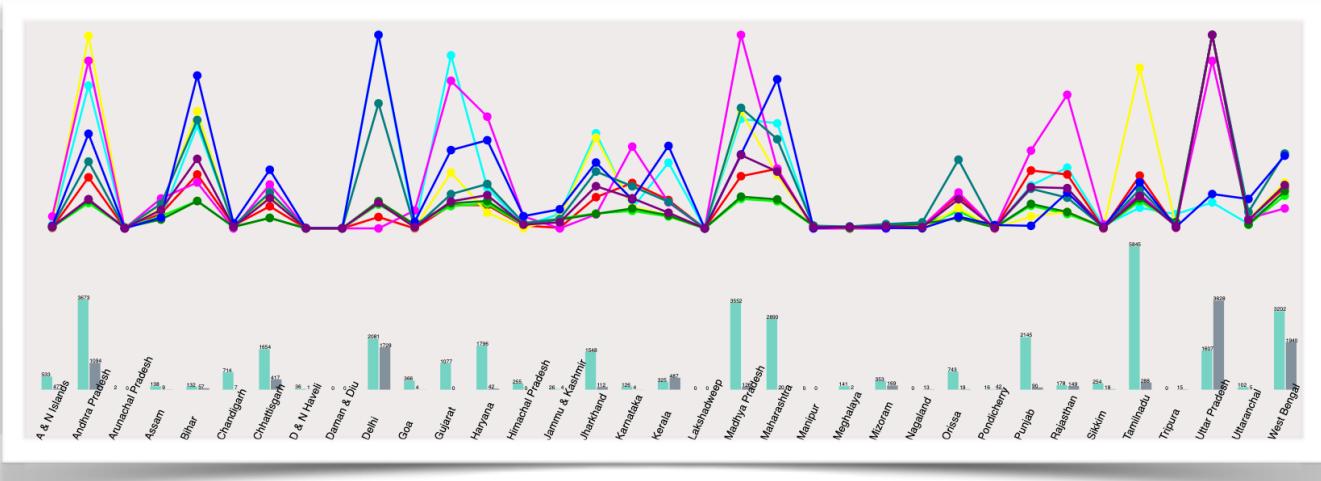


The visualisation is made in 5 segments. First is the bubble graph is made that makes use of circles to explore the factors . Each circle indicates one factor and the radius of each circle shows the extent to which they support or not support the rehabilitation process. Each factor is given different colour that remains constant throughout the visualisation for ease of use. The data used for these circles are normalised to compare one factor with another.

Next using SVG element such as lines, circles and rectangles were made for deriving relation between each factor and rehabilitation and recidivism rate. The data used for this purpose is not normalised as the purpose of this graph os to find patterns.

Next for exploration purpose a slider input and drop down menu is design. This gives user the freedom to explore data at will.

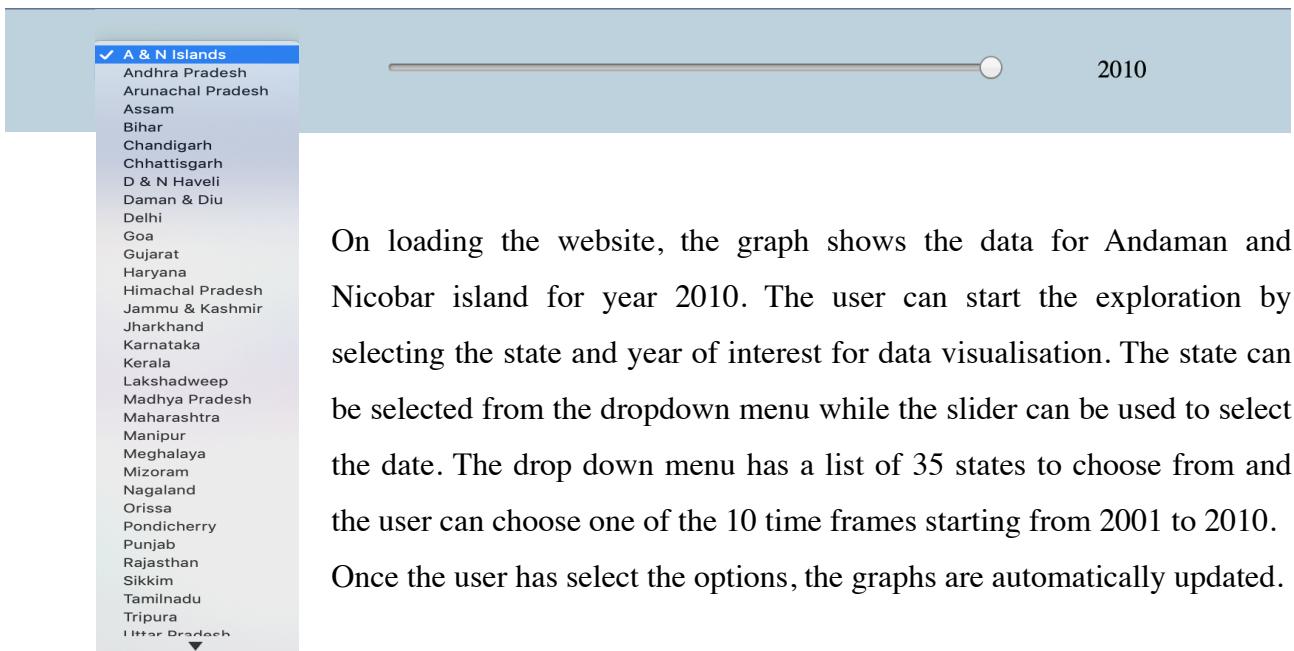




The final graph is designed to give the user an overview of the entire nation. The graph again makes use of SVG element such as lines, circles and rectangles were made for deriving relation between each factor and rehabilitation and recidivism rate in every state for a given time frame.

Lastly the most important part of the visualisation, i.e. a legend is designed. The legend takes care of all three visualisations. The circles in the legend indicates the factors we are making a comparison with. The squares in the legend indicate the area of interest i.e. rehabilitation rate and recidivism rate. The colour combination is purposely kept the same for all the three graphs to avoid any confusion and make it easy for the user to understand the data and learn the results.

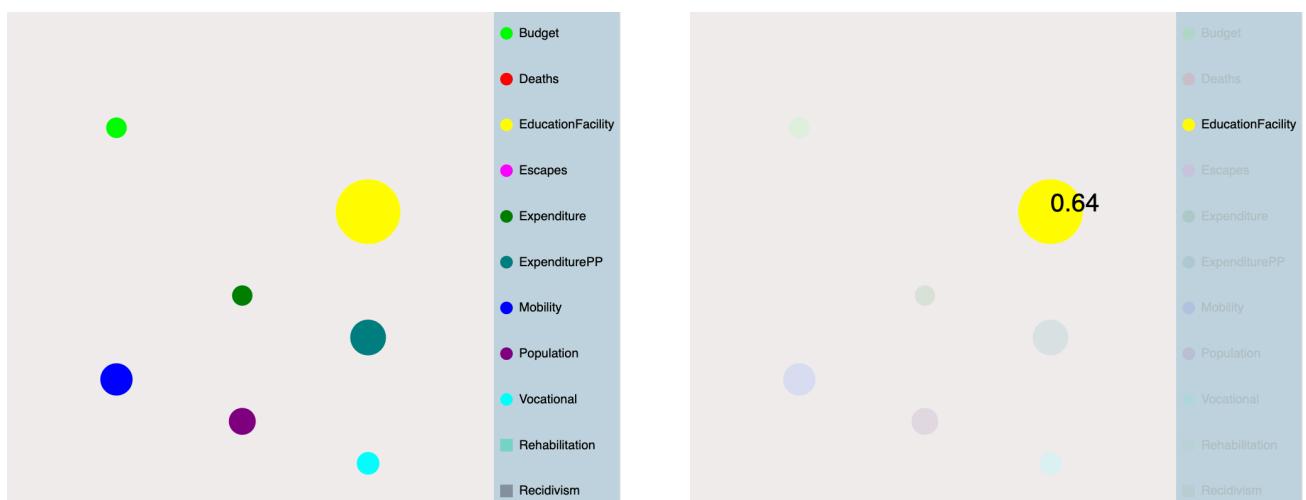
User Guide



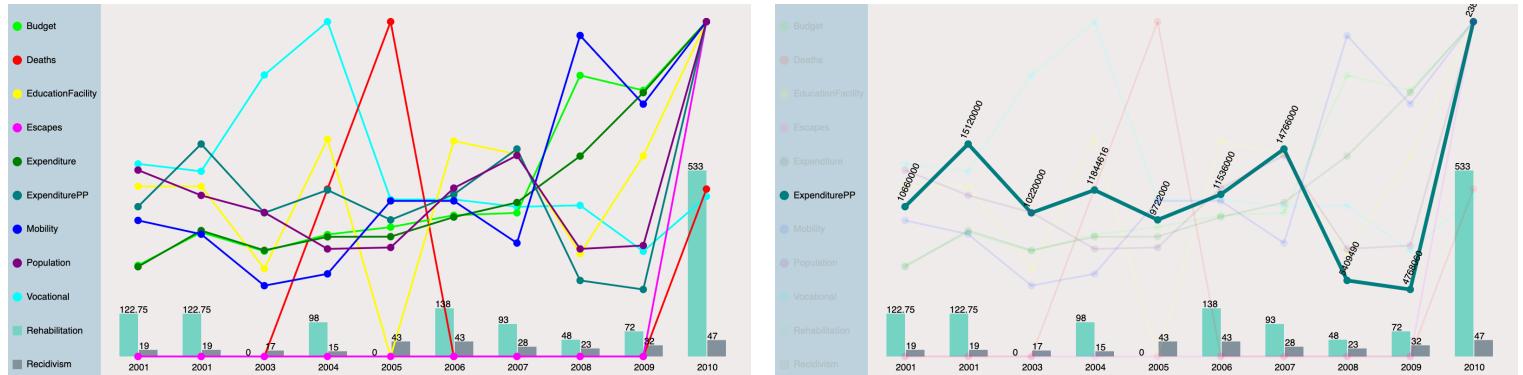
On loading the website, the graph shows the data for Andaman and Nicobar island for year 2010. The user can start the exploration by selecting the state and year of interest for data visualisation. The state can be selected from the dropdown menu while the slider can be used to select the date. The drop down menu has a list of 35 states to choose from and the user can choose one of the 10 time frames starting from 2001 to 2010. Once the user has select the options, the graphs are automatically updated.

Not only are the elements of each graph interactive but also each graph can interact with other graphs. On hovering over any element of the chart, the data value is displayed.

Here we can see that the bubbles on hover highlights the IR factor (Inter-related factor). The IR factor ranges between 1 and 0 where a higher number indicates a more prevalent factor. This graph helps the user understand which factor was more prevalent in a particular state in that particular time. The factor for each state is normalised within their boundaries, i.e. within their state over a time frame of 10 years to make the user compare the factors appropriately.

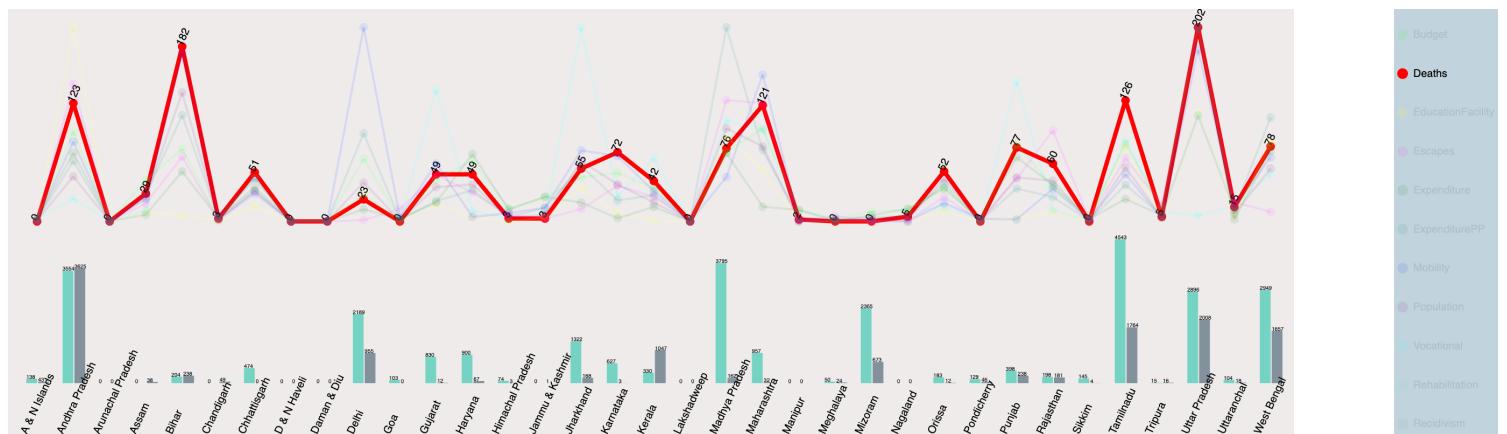
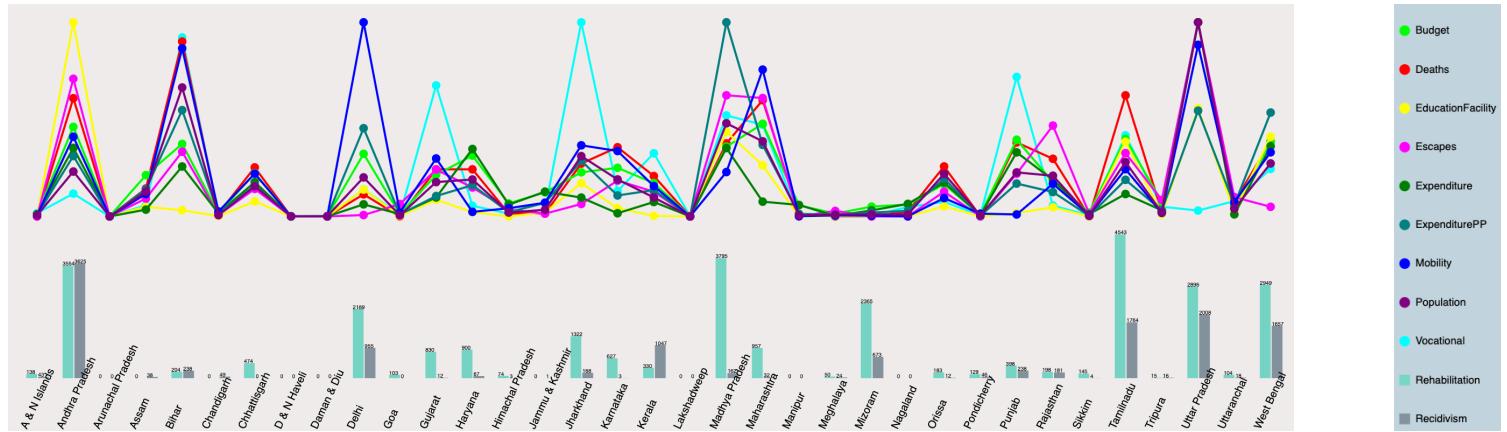


In the next graph, we try to figure out the pattern between different factors and targets i.e. rehabilitation rate and recidivism rate. Each factor is displayed as a line and the target as a bar. The values of the lines are displayed on hovering over the line.

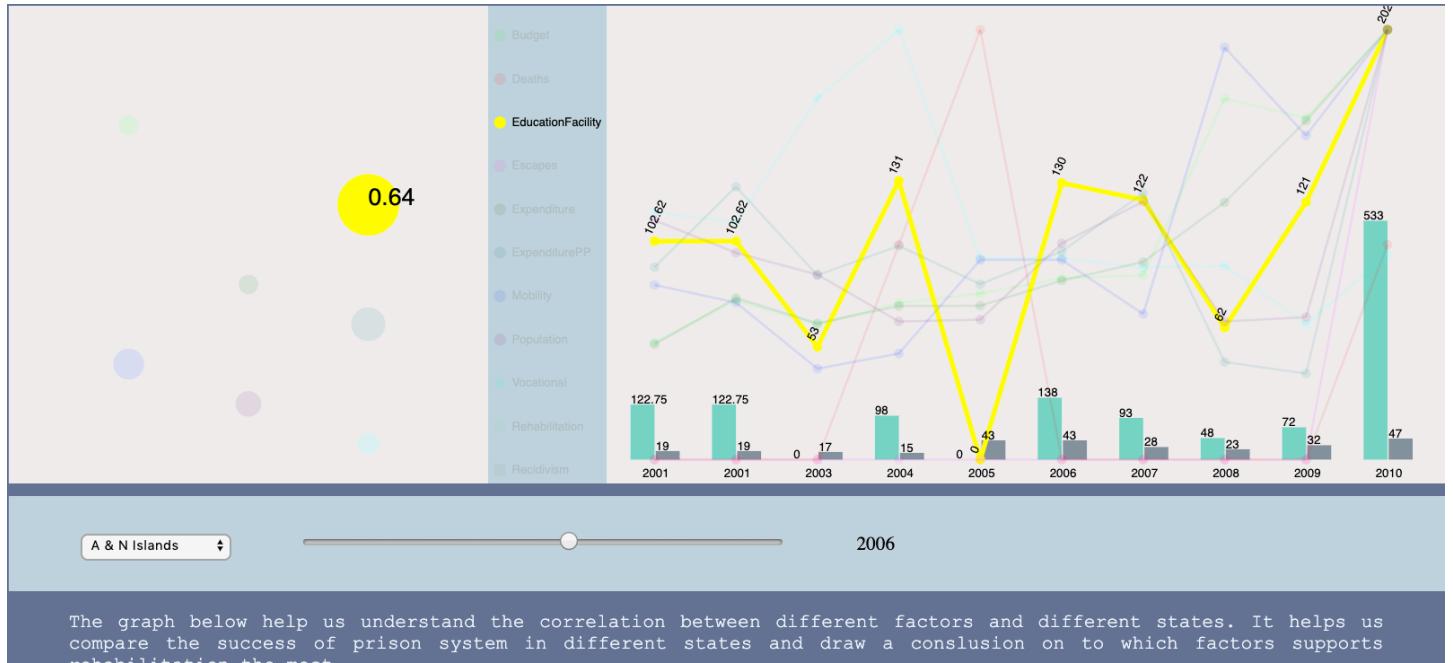


We can see from the graph above that on hover all over factors fade out and the factor we are interested in poppa out. The values of the line is highlighted and the legend highlights the factor hovered over.

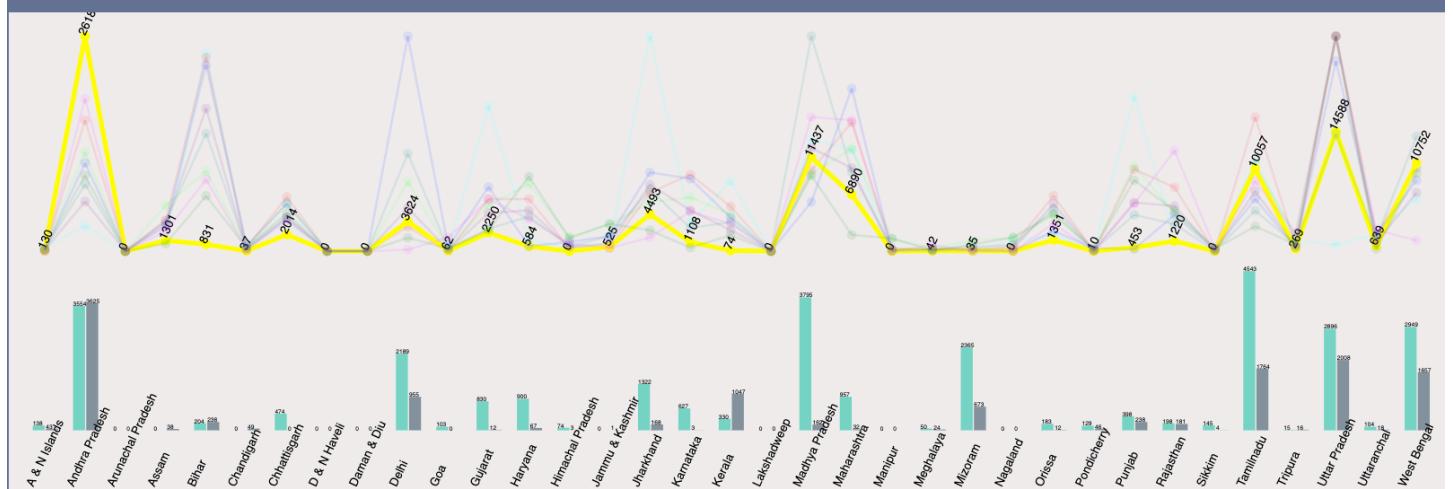
Lastly, the next graph helps the user compare the factors between each state. The interaction in the graph is same as above, that is on hover.



Overall, each graph interacts with other graphs as follows. Each graph is updated in its own way based on the options selected by the user.



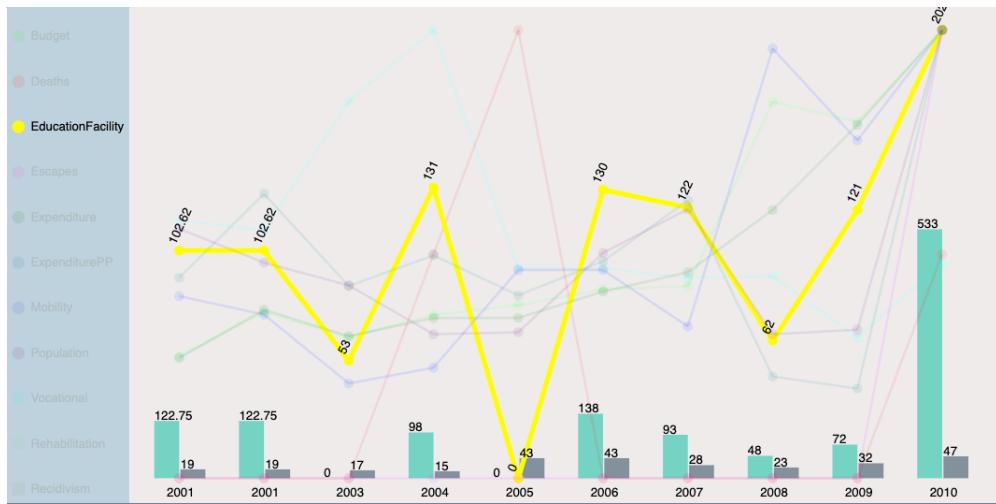
The graph below help us understand the correlation between different factors and different states. It helps us compare the success of prison system in different states and draw a conclusion on to which factors supports rehabilitation the most.



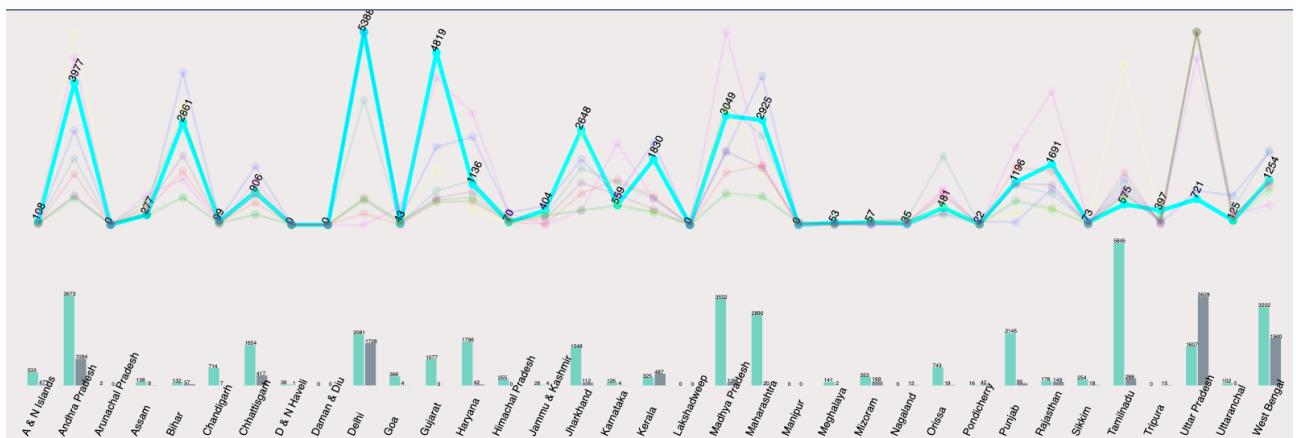
Conclusion

After going through the visualisation we can conclude the following

1. For a few states like A & N Island education facilities inside the jail play a really important role fro rehabilitation of the prisoners.



2. Throughout the country vocational training play some sort of a role in rehabilitation but there was no strong connection found between rehabilitation and training.



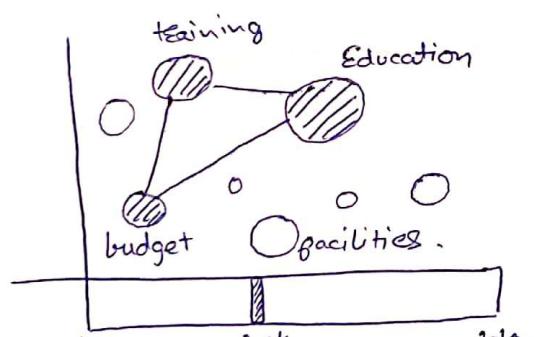
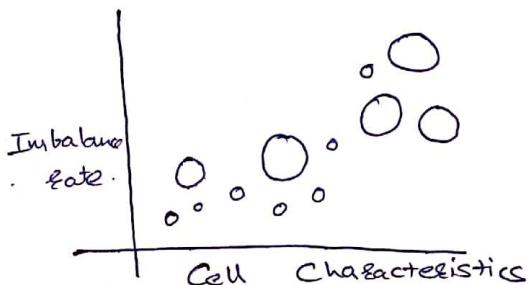
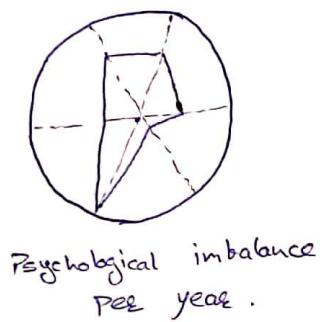
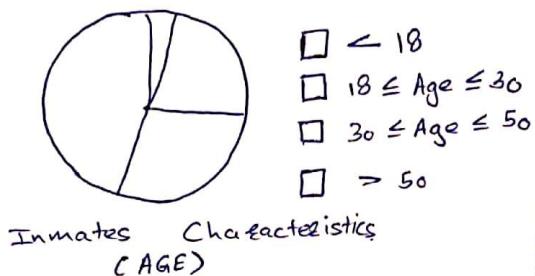
Reflection

After completing the task, the assignment was really helpful in many ways as follows.

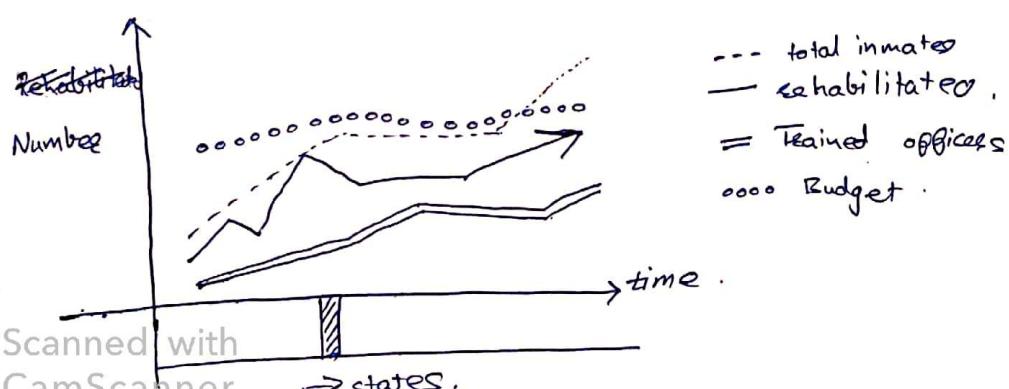
- The task gave me the understanding on making interactive visualisations in D3 and Javascript. It helped my understand the power of D3 and the extent to which it can be used. It provided me with the knowledge of structure of D3 on which I can build in future to improve the quality of visualisations and interactions.
- The task involved cleaning the data and manipulating it to derive some useful results. I realised the concept of garbage-in equals to garbage-out. Therefore, to improve the quality of charts data had to be wrangled in R which gave me an opportunity to work and learn some new concepts in R language.
- As the task involved making interactive charts I realised the importance of user interface and user convenience in creating visualisation. This motivated me to use some HTML and CSS to make the visualisations a bit more appealing for the users.
- One of the most important methodology that the assignment made me familiar with is Five Design Sheet. It made me realise the importance of exploring all possible option before creating a visualisation so as to improve the exploration process.

Five Design Sheet

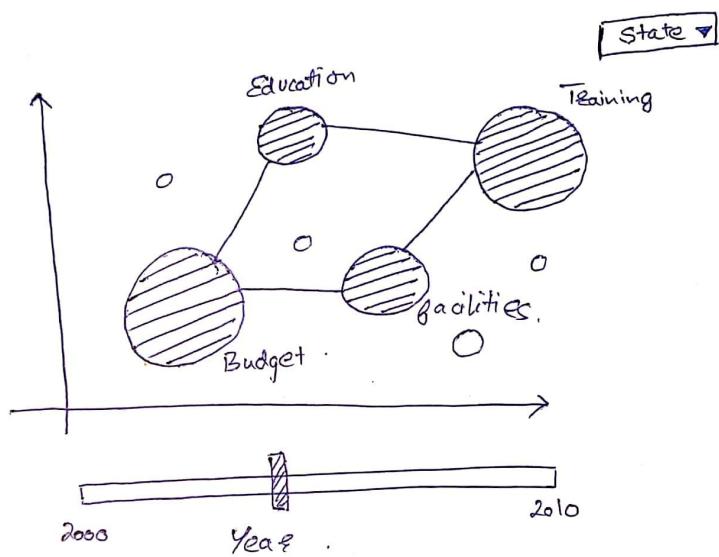
Sheet -1



Major cause of crime.



Sheet - 2



Title : Rehab
of Prisoners
Author : VACHAN ANAND

Date : 24/05/2019

The arc link
the most prevalent
reasons behind
rehabilitations

Operation

- Select time
- Select state.

Focus

On how
details of the
season and
correlation with
rehabilitation.

Sheet - 3



Rehabilitated

Imbalanced

— Budget
= training
--- Total inmates

Budget Training Total inmates

Facilities Literacy

Title : Rehab of
Prisoners
Author : VACHAN ANAND

DATE : 24/05/2019

Operation

- Select state.
- for exploration
- of the details

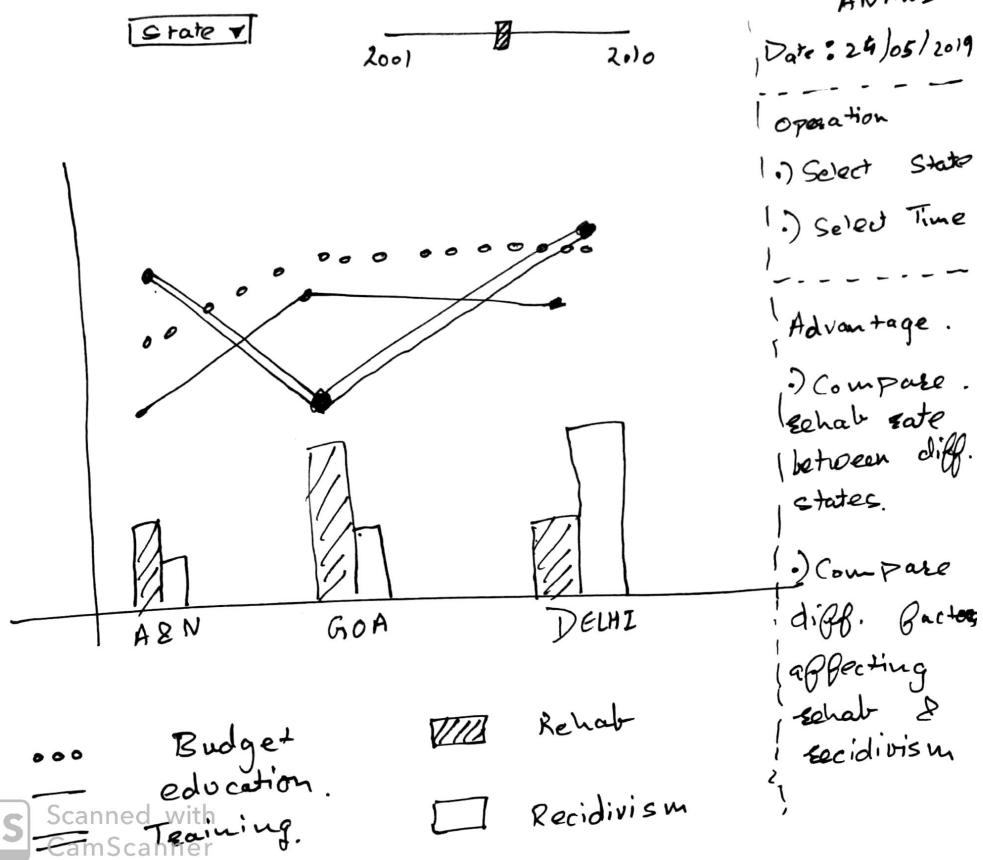
Advantage

• Compare
psychological
imbalance & Rehab
with respect
to different
factors

Disadvantage

- Can not compare
b/w different
states with
respect to time





Author: VACHAN ANAND

Date: 24/05/2019

Operation

1.) Select State

1.) Select Time

1.) Advantage .

2.) Compare .
{ rehab rate
between diff.
states.

3.) Compare
diff. factors
affecting
rehab &
recidivism

