UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO

DATA VISUALIZATION

INDIA'S LITERACY & EMOLOYMENT ANALYTICS
VISUVALIZATION PROJECT

Professor: Prof. Paolo Buono

ALEX PRAVEEN DASS

(MAT NO: 725190)

Table of Contents

Abstract:

Unemployment is the reflection of a static development where strong demographic growth has not reached national production and becomes unproductive. The attenuation of the productive workforce with a high dependency rate due to unemployment also creates immense pressure on a nation's natural resources. This indicator of underdevelopment is one of the major challenges of developing countries like India. The Analysis of unemployment levels is based on primary data, collected by classified survey and then the raw data was calculated and represented by various statistical and mapping techniques. The balance between education and unemployment in the study area is uneven. People who deserve an education do not get decent jobs. As a result, they face issues like stress, unemployment, drug addiction, crime, poverty, etc. in this rural area. Representing this data in an informative and engaging way can be a challenge but data visualization offers many ways to achieve this. This case study focuses on the design of interactive solutions that are measured against existing products by way of a comparative evaluation. In order to study the impact on user performance, efficiency and accuracy are measured for clearly defined tasks carried out on each design. This study will be useful to validate or challenge existing principles. The study will also serve as a case study to support further research into data visualisation evaluation methods. The paper concludes by discussing the findings and possible areas for further research and design.

Introduction:

This is a Case study of Literacy, Labour Participation, Employment, Unemployment and Remedies in India. Economic development results in distribution of wealth. Labour force participation is inherent to any country development. Labour participation depends on social indicators like good health care, access to clean water, high literacy rate etc. Gaps in above indicators results in unbalanced growth and unemployment. Increase in government spending on these indicators results in distribution of income equally.

This analytics contains:

1) Data Dictionary

2) Gender wise Literate comparison

3) State wise Literate comparison

4) Labour Participation

5) Literacy vs Expenditure in Education

6) Labour participation rate.

7) Employment sector wise distribution

8) Unemployment in India

9) Remedies

10) Present and Future

11) Source & References

Existing literature:

1. Data Visualization

Data visualization is the accepted term for the field of graphic communication in which data visualization resides. The foundation of the data analysis and design solutions in this study are well-established principles of human perception and statistical graphics. Clevelend & McGill wrote a fundamental paper in the field, where visual elements such as colour, size, and shape were examined to measure their effectiveness in conveying quantitative information. In addition, Bertin's Semiology of Graphics examines the language of visualization techniques, and Ware's data visualization examines the subject from the perspective of perceptual science. Almost all research on the subject will refer to the work of Edward Tufte, whose books are ground-breaking in their approach to information presentation and visual literacy. The field of interactive data visualization is expanding and continuously evolving due to advancements in technology. Ben Shneiderman has been responsible for many milestones in this area of research. The article Visual Information Search is often referenced as one of the first papers to consider visualizations beyond static representations of data and to highlight the benefits of interactivity. In the paper The Eyes Have It, Shneiderman considers a classification system for interactive visualization techniques classified by data type. The article Visual Information Search is often referenced as one of the first papers to consider visualizations beyond static representations of data and to highlight the benefits of interactivity.

2. Data visualization evaluation

As the demand for new ways to visualize data grows, so does the need for standards and guidelines to support their creation. There is a growing body of work focused entirely on data visualization assessment and the challenges faced by this field that is not common in standard interactive systems assessment. Plaisant addresses these challenges and recommends improving access to data repositories, toolkits and development tools and case studies that can encourage further investment in this area. Carpendale discusses the balance between generalizability, precision and realism and proposes a mixed methods approach to assessments that include both qualitative and quantitative research.

Methodology:

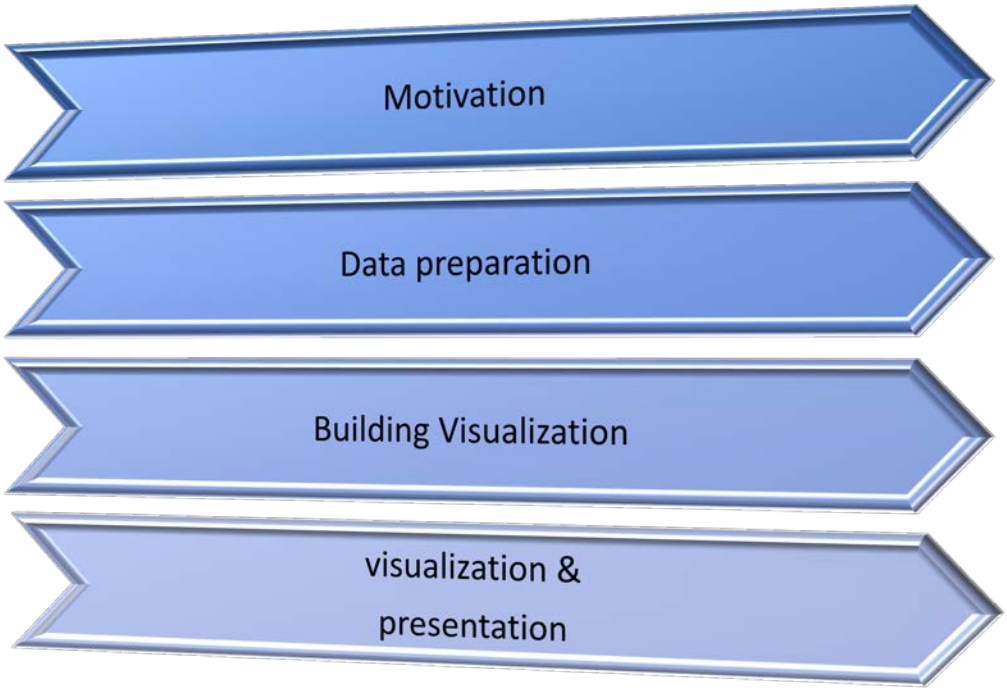
Evaluation Approach:

- Every data visualization project starts with a need, whether that need comes from a problem, decision or clarification, every project has a certain process.
- First, every project needs data to visualize.
- Data is being used and procuring that data is essential as it will form a template for the audience, arguments, and metrics that will all need to be evaluated throughout the steps of the project.
- Next, an argument needs to be made that will use the data to explain, answer, or communicate the point that visualization is made to overcome.

- In general, a data visualization project has these basic steps, but the complexity of each use case is where complexity plays a role. In every data visualization project, there are many things to consider to minimize risk and ensure a successful project.

Research Design

- When designing a project for data analysis, I left wondering where to start?
- Much needs to be done from data collection, cleaning, research, analysis and visualization to derive a perspective that is actionable
- There seems to be no way to approach this problem. However, in order to provide a framework for organizing the work needed for the organization and to provide clear insights from the data, it is useful to imagine it as a cycle with different phases. This section explains the framework and takes us through each step of the project lifecycle to make it easier for us to become familiar with the whole process.



Process of case study

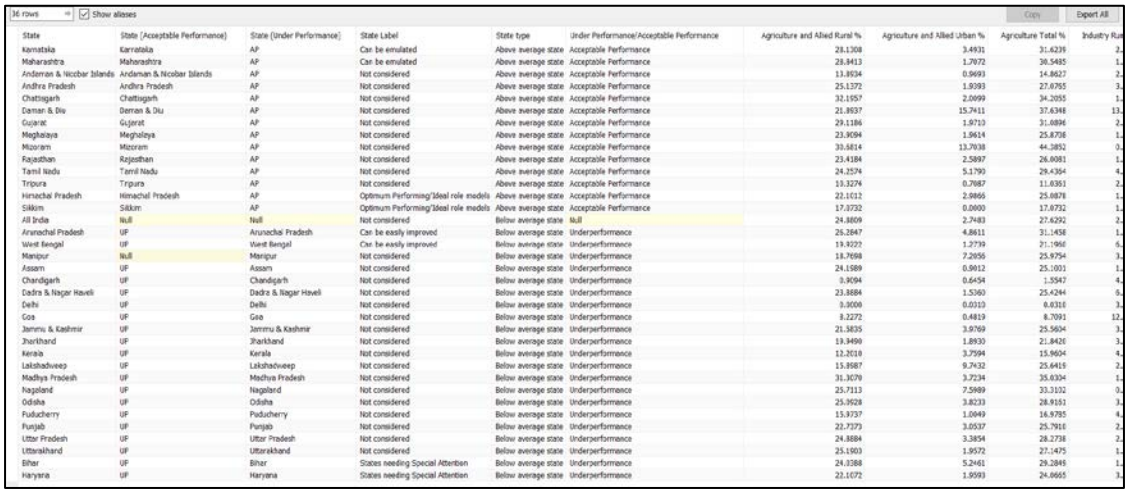
Motivation:

The motivation for this project primarily aims about the India 2025 Development Movement, Here we have constructed a scenario in which a Central Home Minister aims to deal with the current unemployment issues, Illiteracy Rate, Agricultural Protection and State policies for successful Industrialization.

So as soon as the Minister was elected he appointed a group with 10 experts from Human Rights, Reserve Bank of India and Income Tax department research and analysis wing to look into the insights of the ongoing issue to resolve them in his period of Authority.

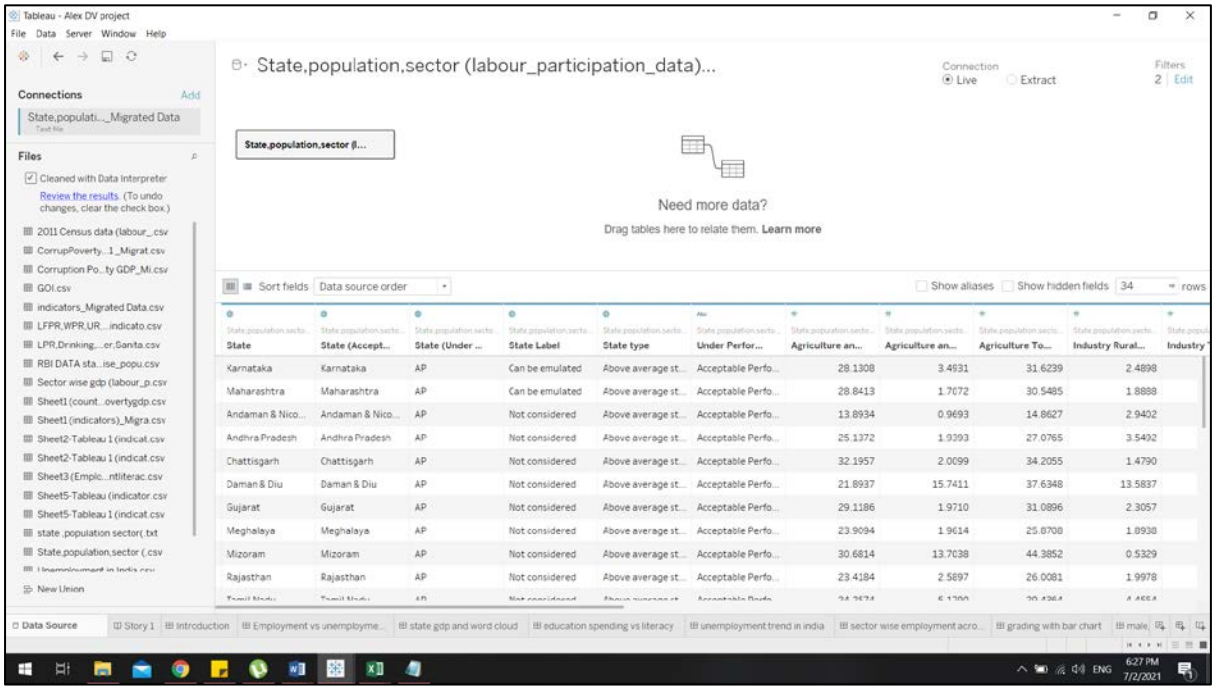
Data preparation

- So, the Home minister's team indicated in the scenario have downloaded the data set from the government's data website which consist of many surveys regarding the GDP, Unemployment, Employment and other related resources for this project.
- They downloaded a multiple data sources in order to do multiple analysis in different charts.
- The data looked already cleaned with desired columns but they were some null values and duplicated or repeated fields.



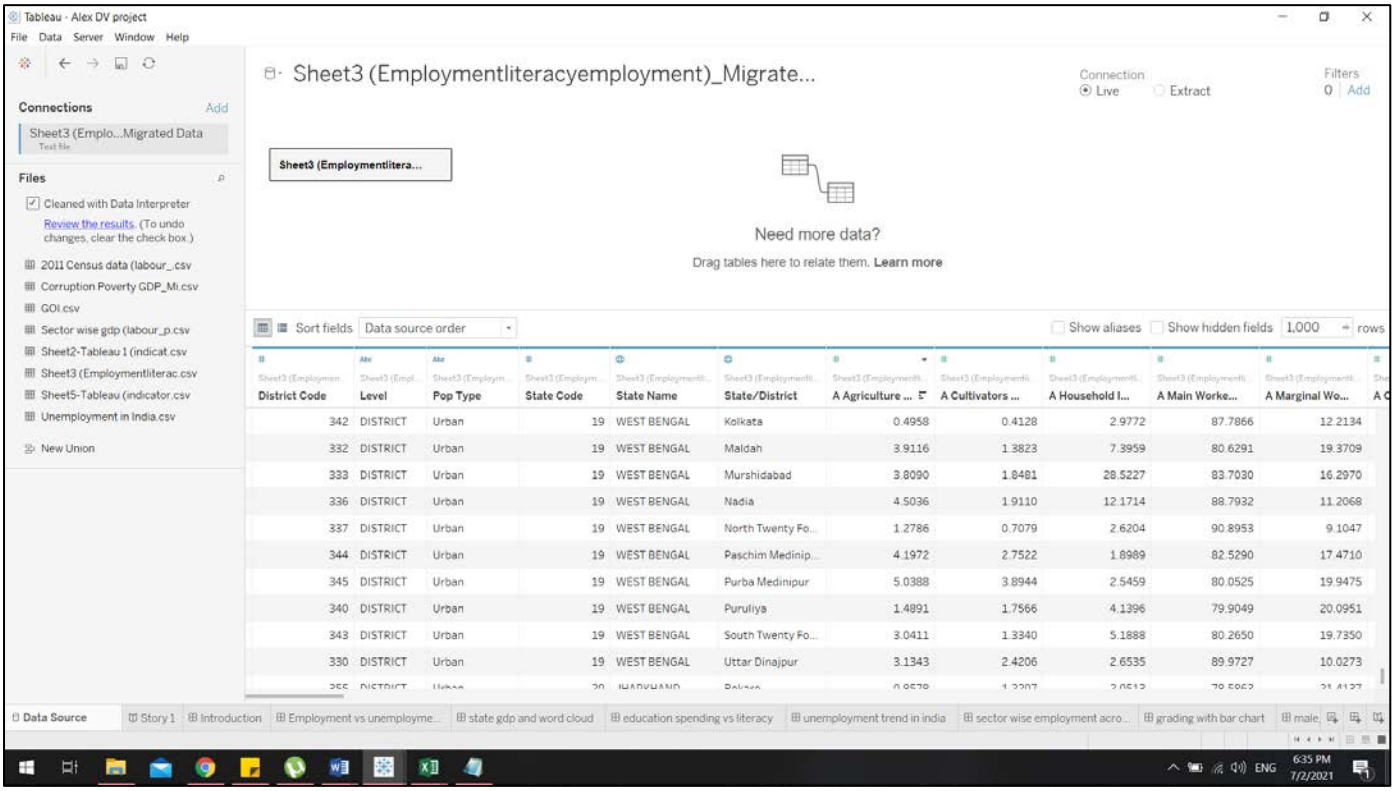
State	State (Acceptable Performance)	State (Under Performance)	State Label	State type	Under Performance/Acceptable Performance	Agriculture and Allied Rural %	Agriculture and Allied Urban %	Agriculture Total %	Industry Rur
Karnataka	Karnataka	AP	Can be emulated	Above average state	Acceptable Performance	28.1308	3.4931	31.6239	2.
Maharashtra	Maharashtra	AP	Can be emulated	Above average state	Acceptable Performance	28.8413	1.7072	30.5485	1.
Andaman & Nicobar Islands	Andaman & Nicobar Islands	AP	Not considered	Above average state	Acceptable Performance	13.8934	0.9693	14.8627	2.
Andhra Pradesh	Andhra Pradesh	AP	Not considered	Above average state	Acceptable Performance	25.1372	1.9393	27.0765	3.
Chattisgarh	Chattisgarh	AP	Not considered	Above average state	Acceptable Performance	32.1957	2.0099	34.2055	1.
Daman & Diu	Daman & Diu	AP	Not considered	Above average state	Acceptable Performance	21.8937	15.7411	37.6348	13.
Gujarat	Gujarat	AP	Not considered	Above average state	Acceptable Performance	29.1186	1.9710	31.0896	2.
Meghalaya	Meghalaya	AP	Not considered	Above average state	Acceptable Performance	23.9094	1.9614	25.8708	1.
Mizoram	Mizoram	AP	Not considered	Above average state	Acceptable Performance	30.5814	13.7038	44.3852	0.
Rajasthan	Rajasthan	AP	Not considered	Above average state	Acceptable Performance	23.4184	2.5897	26.0081	1.
Tamil Nadu	Tamil Nadu	AP	Not considered	Above average state	Acceptable Performance	24.2574	5.1790	29.4364	4.
Tripura	Tripura	AP	Not considered	Above average state	Acceptable Performance	13.3274	0.7687	11.8361	2.
Himachal Pradesh	Himachal Pradesh	AP	Optimum Performing/Ideal role models	Above average state	Acceptable Performance	22.1012	2.9866	25.0878	1.
Sikkim	Sikkim	AP	Optimum Performing/Ideal role models	Above average state	Acceptable Performance	17.3732	0.0000	17.8732	1.
All India	Null	Null	Not considered	Below average state	Null	24.8809	2.7483	27.6292	2.
Arunachal Pradesh	UP	Arunachal Pradesh	Can be easily improved	Below average state	Underperformance	26.2847	4.8611	31.1458	1.
West Bengal	UP	West Bengal	Can be easily improved	Below average state	Underperformance	19.9222	1.2739	21.1960	6.
Manipur	Null	Manipur	Not considered	Below average state	Underperformance	18.7698	7.2056	25.9754	3.
Assam	UP	Assam	Not considered	Below average state	Underperformance	24.1589	0.9012	25.1001	1.
Chandigarh	UP	Chandigarh	Not considered	Below average state	Underperformance	0.9094	0.6454	1.5547	4.
Dadra & Nagar Haveli	UP	Dadra & Nagar Haveli	Not considered	Below average state	Underperformance	23.8884	1.5360	25.4244	6.
Delhi	UP	Delhi	Not considered	Below average state	Underperformance	0.3000	0.0310	0.0310	3.
Goa	UP	Goa	Not considered	Below average state	Underperformance	8.2272	0.4819	8.7091	12.
Jammu & Kashmir	UP	Jammu & Kashmir	Not considered	Below average state	Underperformance	21.5835	3.9769	25.5604	3.
Jharkhand	UP	Jharkhand	Not considered	Below average state	Underperformance	19.9490	1.8930	21.8420	3.
Kerala	UP	Kerala	Not considered	Below average state	Underperformance	12.2019	3.7594	15.9604	4.
Lakshadweep	UP	Lakshadweep	Not considered	Below average state	Underperformance	13.8587	9.7432	25.6419	2.
Madhya Pradesh	UP	Madhya Pradesh	Not considered	Below average state	Underperformance	31.3070	3.7234	35.0304	1.
Nagaland	UP	Nagaland	Not considered	Below average state	Underperformance	25.7113	7.5989	33.3102	0.
Odisha	UP	Odisha	Not considered	Below average state	Underperformance	25.3628	3.8233	28.9161	3.
Puducherry	UP	Puducherry	Not considered	Below average state	Underperformance	15.9737	1.0049	16.9785	4.
Punjab	UP	Punjab	Not considered	Below average state	Underperformance	22.7373	3.0537	25.7910	2.
Uttar Pradesh	UP	Uttar Pradesh	Not considered	Below average state	Underperformance	24.8884	3.3854	28.2738	2.
Uttaralhand	UP	Uttarakhand	Not considered	Below average state	Underperformance	25.1903	1.9572	27.1475	1.
Bihar	UP	Bihar	States needing Special Attention	Below average state	Underperformance	24.0388	5.2461	29.2849	1.
Haryana	UP	Haryana	States needing Special Attention	Below average state	Underperformance	22.1072	1.9593	24.0665	3.

- The data looked already cleaned with desired columns but they were some null values and duplicated or repeated fields.
- Null values are removed manually by the tableau itself and in addition they checked in the data interpreter option which is meant to show the data clean and clear.



- Data cleaning techniques can be performed as batch processing via interacting with data cleaning tools.
- After cleaning, the dataset should be consistent with other data sets involved in the operation.

- Differences identified or eliminated may be substantially due to user input errors, to failures during storage or transmission, or to different data dictionary descriptions of similar items. In different stores. Remaining data from the file.
- The main motto is to remove duplicate or unrelated observations from the data set.
- Duplicate observations will occur most often during data collection. When we combine data sets from multiple places, categorize data, or receive data from data base, there is an opportunity to create duplicate data. Disinfection is one of the biggest areas to be considered in this process.
- Observation mismatch is when you notice observations that do not match the particular problem you are trying to analyse. For example, if you want to analyse data about older member, but your dataset includes older generations, you can delete those irrelevant observations. This can make analytics more efficient and reduce distractions from your main goal as well as create a data set that's more manageable and more productive.



- The main motto is to have a multiple data sources in the same workbook, but not merging or mixing their data.

  The steps are followed as:

  1) Connect to the primary data source as usual and See connecting to the Data for more information.
  2) Create a Worksheet, This Worksheet will use data from your initial data source.
  Create a new Worksheet.
  3) Select Data > New Data Source using the top menu bar.
  4) Connect to a new data source and, if necessary, export the tables.
  5) If we want, we can go back to the last created Worksheet to see the fields in the new data source ready to use.
  Repeat steps as many times we wanted to add as many data sources and worksheets.

Five characteristics of quality data

- **Validity.** The extent to which your data is subject to defined business rules or constraints.
- **Accuracy.** Make sure your data is close to the actual value.
- **Completeness.** The extent to which all required data is known.
- **Consistency**. Make sure your data is consistent within the same dataset and/or across multiple datasets.
- **Consistency.** Data levels are specified using the same units of measurement.
- Having clean data ultimately increases overall productivity and allows the highest quality information to be delivered in your decision making.
- **This data is classified as quantitative data**

Building Visualization

- There is a need behind every project and the need can be as simple as a weekly sales performance dashboard or a sophisticated predictive referral module.
- Addressing these needs with specific measurable goals provides the right framework for providing the right information in the right way. The communication of key performance indicators from the final product is very important for consumers or audience.
- To do this, you must gather requirements, set up design processes, schedule regular discussions with users, and continue these meetings until the project is finally implemented.
- Sample or research questions to help you better understand the project:

Framing a research question

1. **Is male and female are equally illiterate?**
2. **What are the state wise Illiterate and literate contribution?**
3. **How is the Literacy rate and Expenditure related to education?**
4. **What are the distribution of labour force participation rate?**
5. **What is labour force participation vs gross enrolment ratio, mean year in school, literacy rate and poverty line?**
6. **What are the three major sectors people fall under?**
7. **Which states have higher LPR, Literacy, GDP, workers percentage?**
8. **How is the unemployment rate fluctuate throughout the years?**
9. **How corruption is affecting the employment?**
10. **Employment vs Unemployment and how unemployment can be decreased?**
11. **How literacy is affecting the agriculture?**

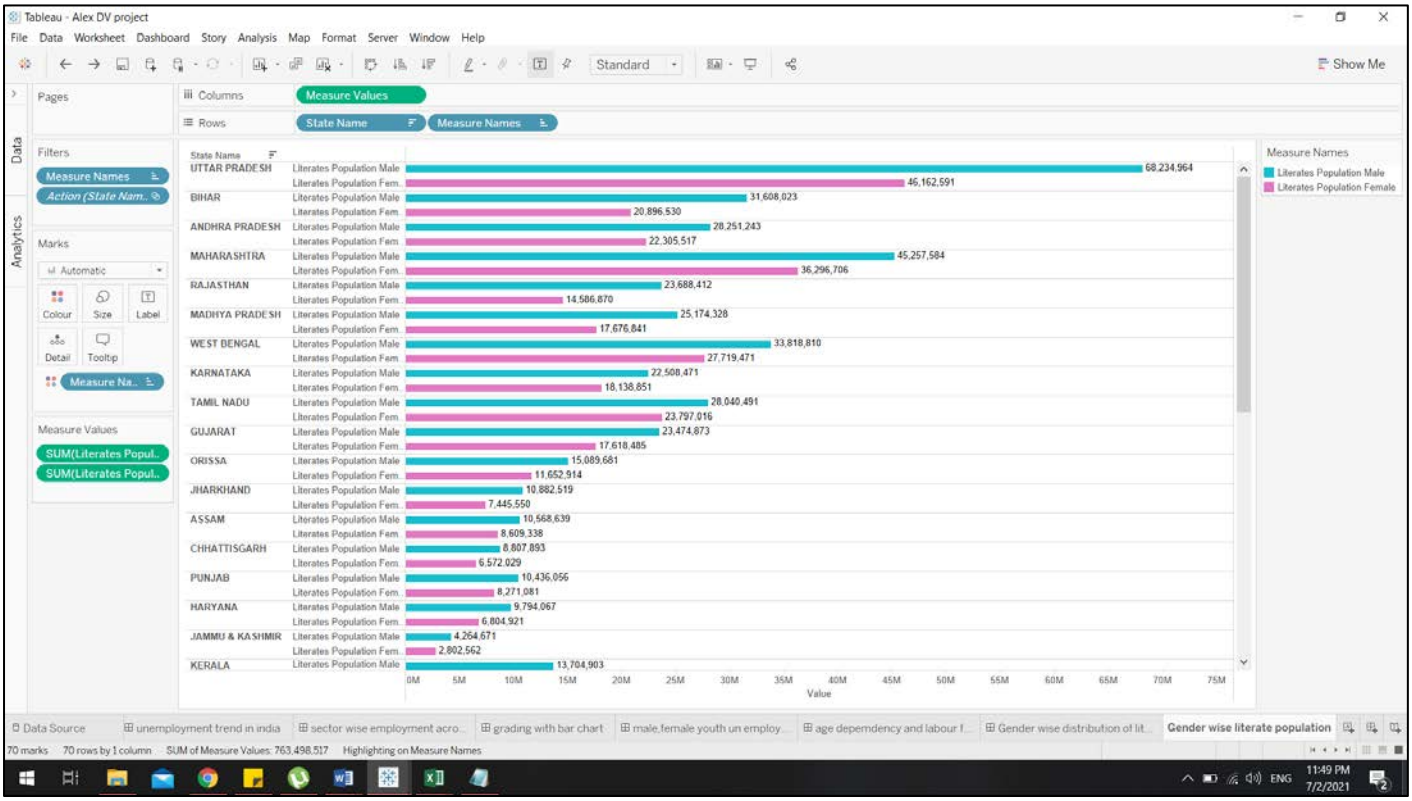
- The Team started to visualize the data with the reference of research question by adapting the **Shneiderman theory to define a data visualization project.**

- They had other theories to follow which is graphical display should show the data induce the viewer to think about the substance rather than about methodology, graphic design and to avoid distorting what the data have to say present many numbers in a small space and make large data sets coherent encourage the eye to compare different pieces of data , reveal the data at several levels of detail serve a reasonably clear purpose: description, exploration, tabulation, decoration be closely integrated with the statistical and verbal Descriptions.

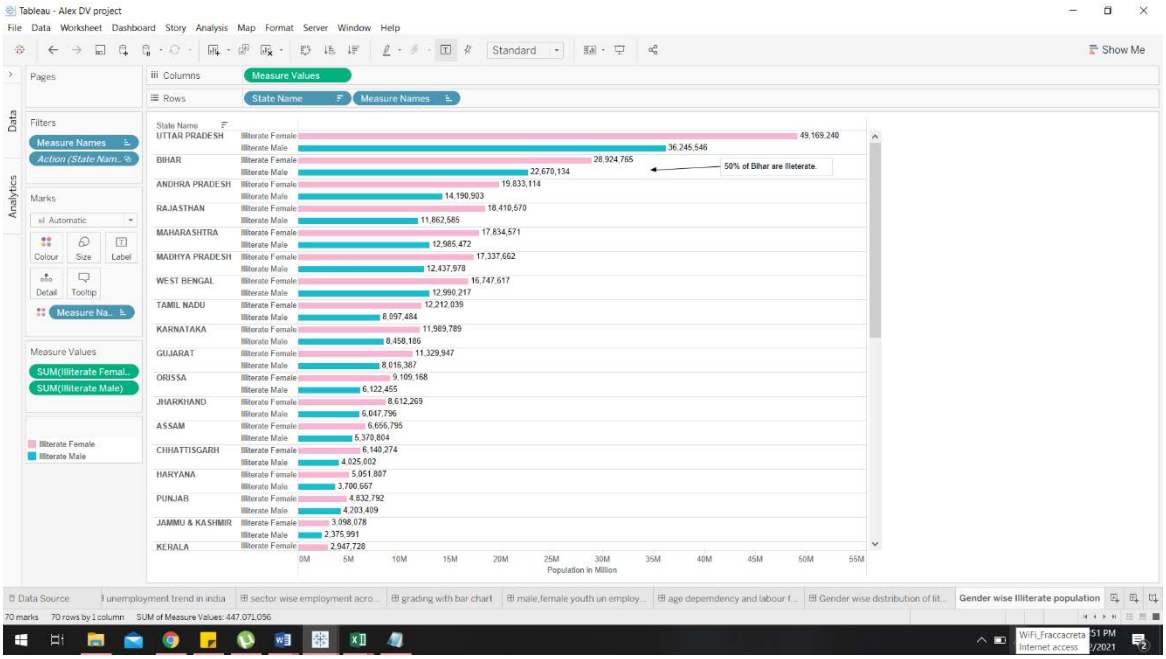
Research questions and its answer by data visualization:

1. *Is male and female are equally illiterate?*

- The main motivation for creating this question and expecting an answer is to visualize and see the distribution of population and visualize if the literacy rate is equal or not.
- To answer this question not only illiterate's data is visualized but also literacy rate proportional to the states should be visualized.



- A horizontal bar chart is used to show the comparison of these data. This chart is the best method for presenting this type of information because the labels are too long to be clearly visible on the x-axis.
- The chart consist of one measure values in column which is nothing but the sum of literate people and the row contain two dimension state name & measure names literate population which is male and female.
- The colour is encoded with blue –male and pink –female.



- The chart consist of one measure values in column which is nothing but the sum of illiterate people and the row contain two dimension state name & measure names illiterate population which is male and female.
- The colour is encoded with blue –male and pink –female.

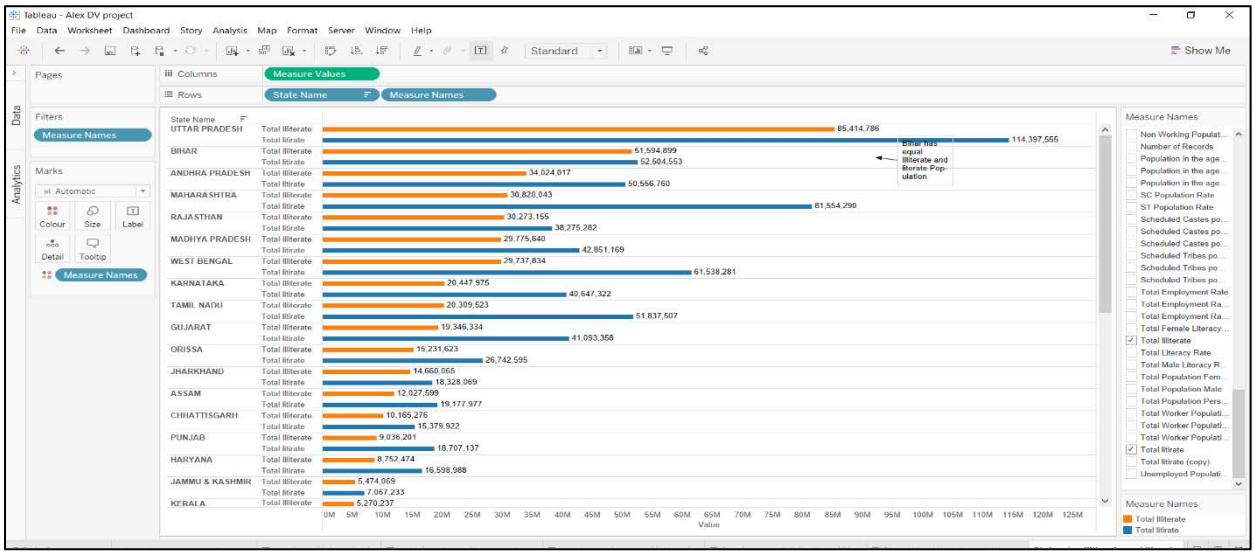
Insights:

- India overall literate and illiterate gap between Male and Female is 8%, which indicates developing economy status. Among Economically well-off states Tamil Nadu and Maharashtra has the high literacy level, hence thin gender literate gap.
- Female literate population in Tamil Nadu is equal to male literate population. Likewise in Gujarat as they have similar population level.
- North Eastern States fares far better than other States.
- Kerala's gender literate gap is the lowest.
- Among southern states Kerala has high literacy level (94% of total population).
- Northern and Eastern states show low literate population against their total population.
- Northern States like Uttar Pradesh and Bihar constitute largest concentration of Illiterate Population.
- Due to population exploitation land availability for economic activity and over dependence on agriculture.
- Gender wise illiterate gap is wider in Northern India.
- In Southern States Female illiteracy is in par with the male illiteracy.

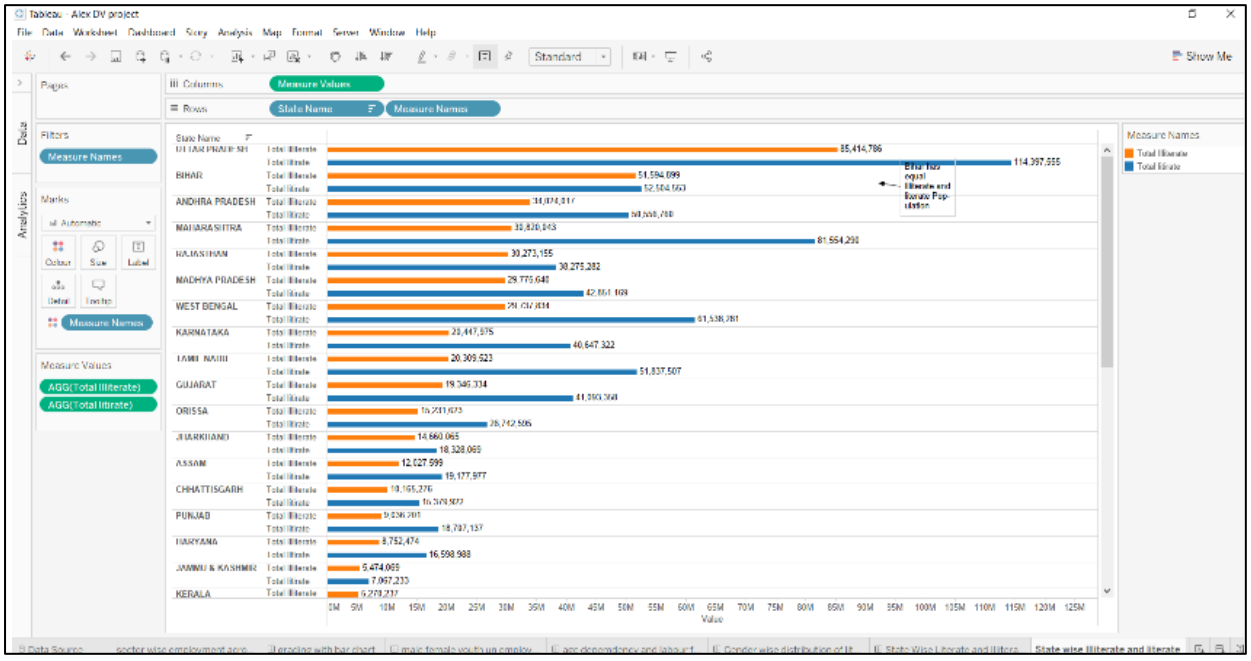
2. What are the state wise Illiterate and literate contribution?

- The last chart we saw gender wise comparison of literate and illiterate distribution, in this research question we are going to see state wise literate and illiterate distribution.
- In order to see this, the crew used horizontal bar chart and multi chart that bar chart with the trend line for two analysis.

State wise LITERATE & ILLTERATE DISTRIBUTION:

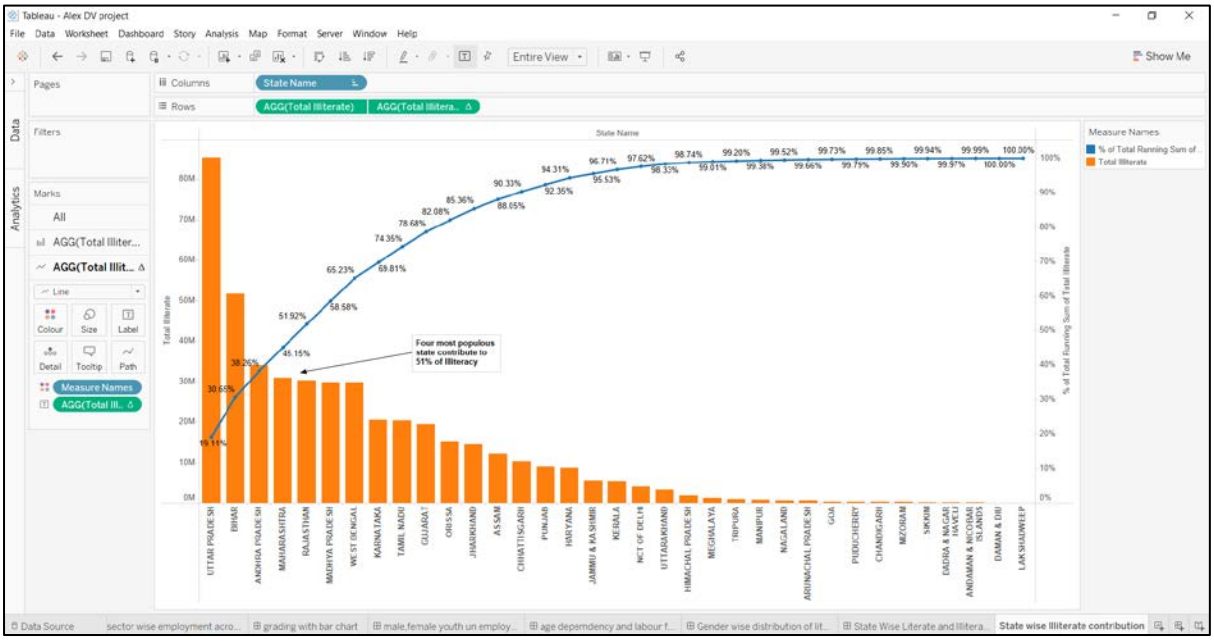


- The data has many fields like caste, agricultural data set, total worker population in which literate and literate has been filtered out dye to details on demand.



- The column contains Measure values nothing but AGG (total illiterate & total literate) (AGG simply indicates an aggregation in a calculated field. For example, if you drag a Sales pill into a view, by default Tableau will wrap it in SUM, and you can edit the aggregation)
- The row contains state name and Measure Names such as Total illiterate &literate.
- The visual colour is encoded with illiterate orange and literate blue.
- **This chart insight is clearly shown about the illiterate and literate counts million and the upcoming chart going to use this count as a percentage of total running sum of total illiterate along table (across)**
- Bihar has equal Illiterate and literate Population.

State wise illiterate contribution:



- Dual-axis combination charts, or Combo Charts, are an effective type of chart for combining views to show relevant information while saving real estate. This chart type is created with one shared axis, such as the X-axis for date, and two separate axes, such as the Y-axis, for two different measures.
- They used the total illiterate dataset in two ways as the bar and lines so that the bar denotes the illiteracy values in numbers and the line chart to depict the percentage values.

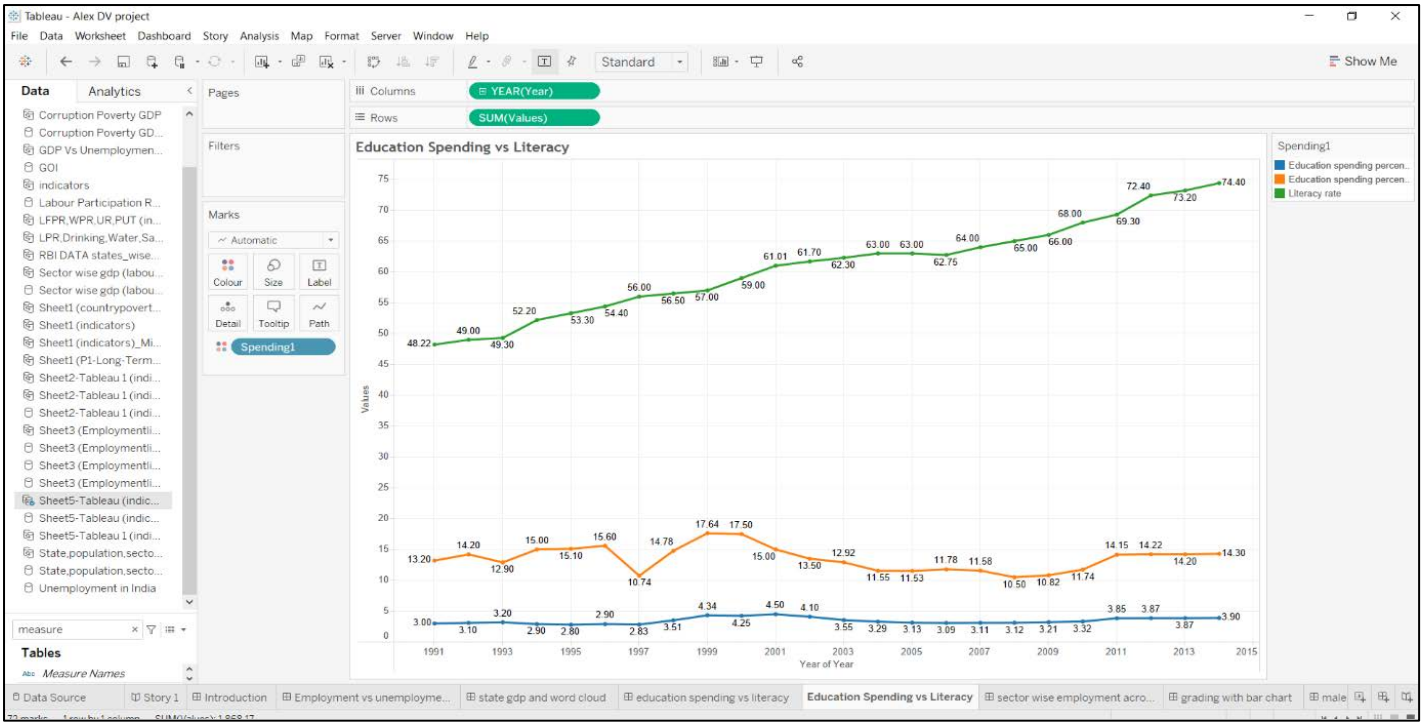
Insights:

- Northern India constitute major part of literate and illiterate population.
- The four most populated States has the high rate of illiteracy.
- The reason being the percentage is less in illiteracy but the states being on the top is because in the total population of a high populated northern state both literacy and the illiteracy rate is at the same ration hence lesser the percentage of illiteracy but higher in number comparative to the other states.
- By analysing this visually we can understand that there is a less gap between literal and illiterate people in Northern States.

3. How is the Literacy rate and Expenditure related to education?

- In order to understand the literacy the expenditure to the education plays a vital role.
- Some countries may provide scholarships, some may not provide scholarship for students but their college fees might be more which affects the literacy rate of a country.
- In order to prove this education spending vs literacy is forecasted in the trend line chart.

- A multiple line graph shows the relationship between independent and dependent values of multiple sets of data. Usually multiple line graphs are used to show trends over time. In the graph, each data value is represented by a point in the graph that are connected by a line.



- Here in this chart a data set which contains **%** up to 75 in Y axis which is an Indian way of calculating a literacy rate.

- The X axis consist of the Year.

- The Line green indicate literacy rate ,Line blue indicates Education spending percent of GDP, The line orange indicates Education spending of the government.

Insights:

- The blue line which starts at 3% and goes up to 3.90% from the span of 1991 -2015. So we can understand that the Education spending percent of GDP in a public university/school hasn't dramatically increased much.

- We also can see the Education spending of the government in 1991 was at 13.2 and by the end of the year in 2015 it was a slight increase to 13.90. But when we see the graph we can understand the spending went up to 17.20 in the year 2000.

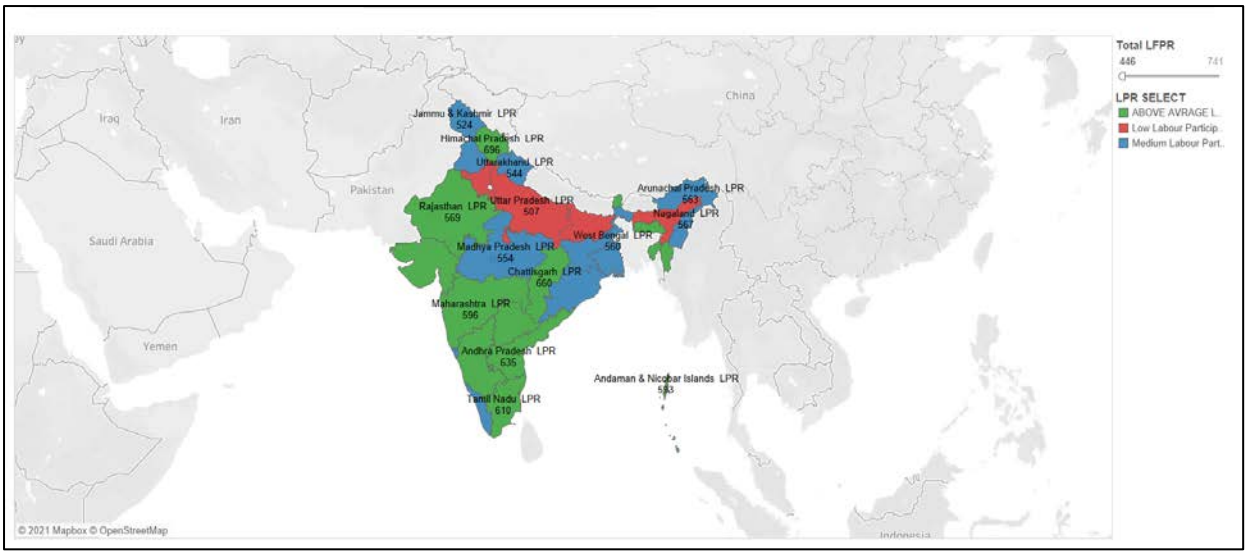
- So when we see the green line for the literacy rate we can see a steady growth at 48.22% in 1991 all the way to 75% by the year 2015.

- As far as our understanding we can conclude that the spending done by the government and the citizens doesn't increase as such but there is a question of how the literacy rate got increased?

- To answer that we understood that almost 64.3% of people get there education in private universities/ schools. Only 22% is being aided by the government the expenditure that is displayed above only indicates the expenditure spent for public institution that's the reason why the literacy rate has been affected much.

4. What are the distribution of labour force participation rate?

- The colour has been visually encoded:
  - Green denotes: Above Average Labour Participation in State
  - Red Denotes: Low Labour Participation Rate in State
  - Blue Denotes: Medium Labour Participation Rate in State



Insights

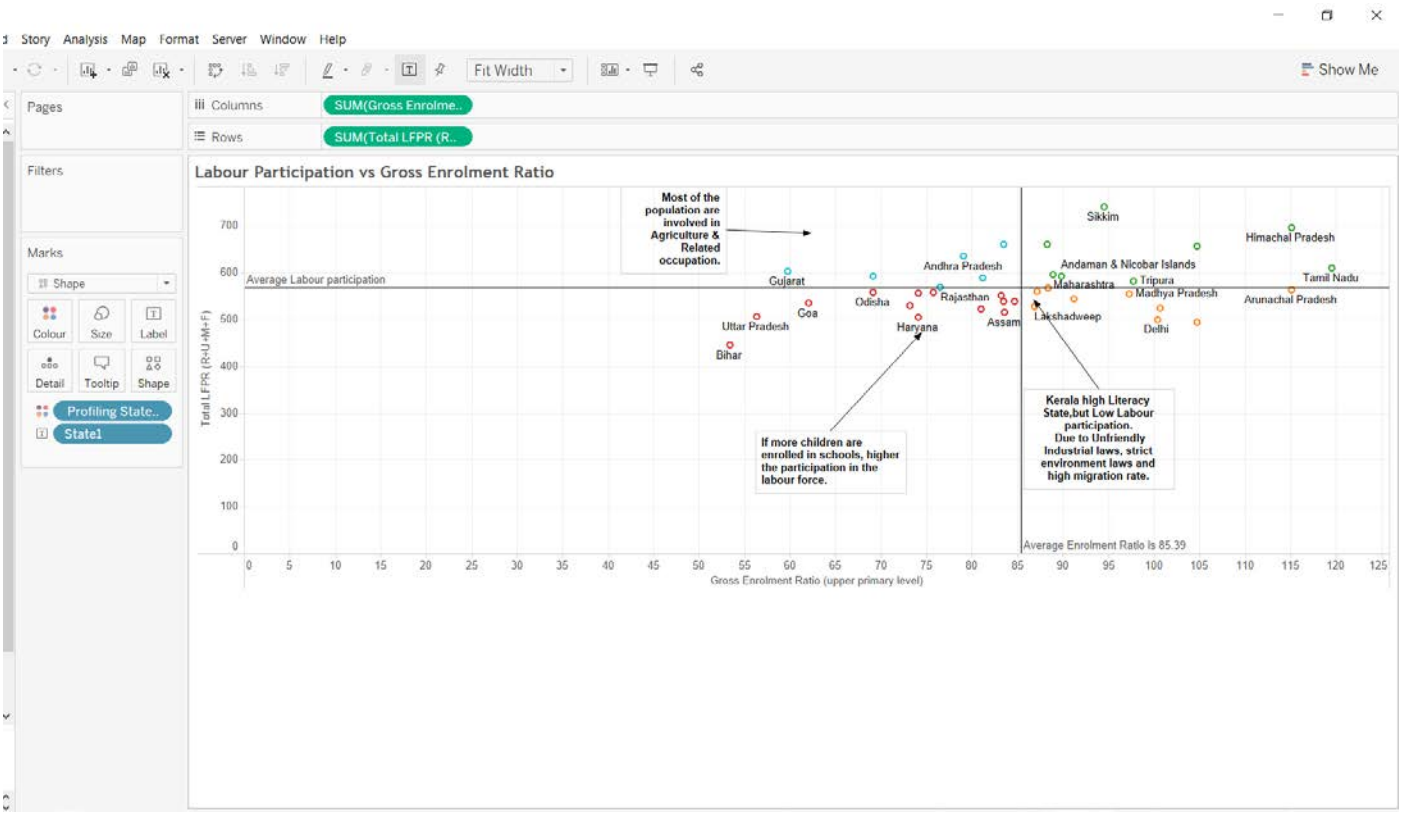
- This chart indicates state Wise Distribution of Labour Force Participation Rate.
- Economically developed States show higher LPR than economically starved states.
- Sikkim, Mizoram, Meghalaya have high LPR even though they are not economically well off.
- This is due to major occupation in Agriculture

5. What is labour force participation vs gross enrolment ratio, mean year in school, literacy rate and poverty line?

Labour force participation vs gross enrolment ratio

- To understand the number of labours in each state and number of students/children getting enrolled in the education system plays a vital role.
- Because for a healthy economy the literacy rate has to be high in order to get enough labour force, so to generate enough literacy in the state we have to be sure about how many institutional enrolments takes place every year.
- To calculate these differences we use Scatter plots in the X axis we have total labour power ratio with the sum of Rural, Urban, Male, Female data. In the Y- axis we have Gross enrolment Ratio, which means the total number of enrolments happen every year.
- Blue Dots Indicate High Labour power ratio with less gross enrolment ratio.
- Red dots indicates low labour power ratio with less gross enrolment ratio.
- Green dots indicates high labour power ratio with high gross enrolment ratio.
- Orange Dots indicates low labour power with high gross enrolment ratio.

Labour force participation vs gross enrolment ratio



- A reference line is being used just to illustrate how narrow the differences are but making a whole new difference.

Insights

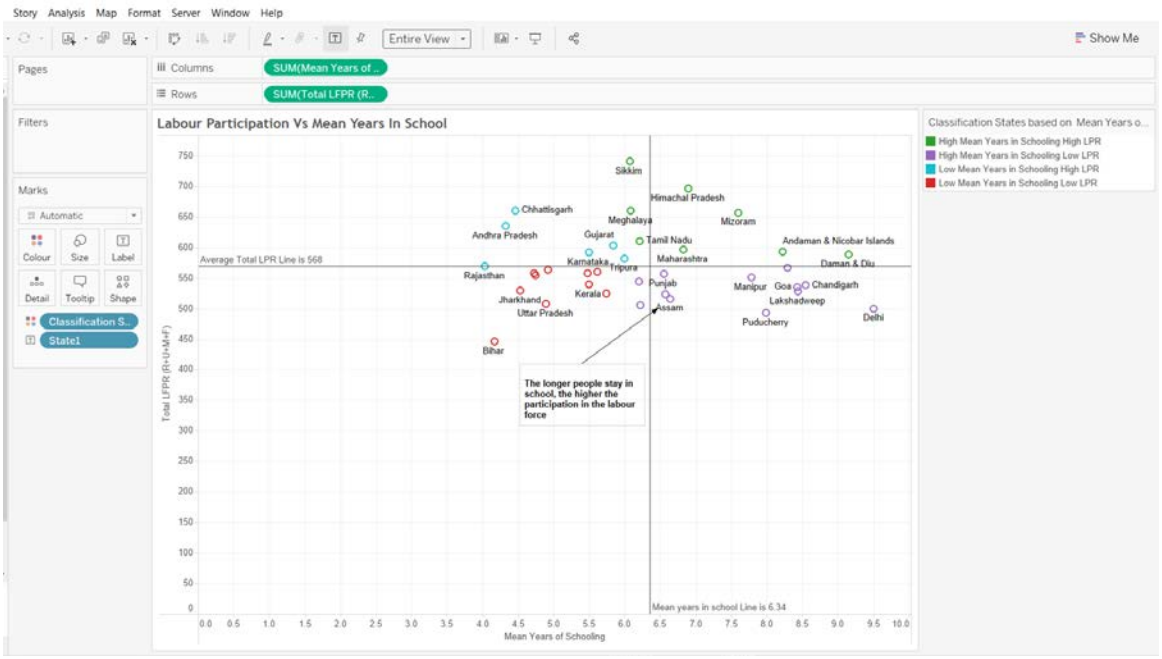
- Some of labour power ratio can be affected due to many other factor in spite of a high literacy rate as an example Kerala State has really high literacy but they have low labour power ratio because of their strict in-state industrial laws, labour laws, high raw material cost for manufacturing, high migration rates etc. causes it to have high literacy but low labour power
- In Tamil Nadu the literacy rate and the labour power ratio is significantly high so at most everyone gets enrolled in education and get an employment afterwards.
- States like Bihar has very poor literacy rate but higher labour ratio because their major occupational sector is Agriculture as both elderly and the youngsters are indulged in agriculture the overall gross enrolment is a huge question.

Labour force participation vs mean year in school

- This graph is to find out the total labour force that is generated along with the same time the Mean year of Schooling
- Mean year in School is nothing but Average number of completed years of education of a country's population aged 25 years and older, excluding years spent repeating individual grades.
- To calculate these differences we use Scatter plots in the X axis we have total labour power ratio with the sum of Rural, Urban, Male, Female data. In the Y- axis we have mean year in school.

- So this is to prove if the median in schooling is higher the literacy rate will be slow so there will be an illiterate set of labour power.
- The green Dots denote that the states which have High Mean years in School with higher labour Power.
- The Red Dots denote that the states which have Low Mean years in School with lesser labour Power.
- The Purple Dots denote that the states which have High Mean years in School with lesser labour Power
- The Blue Dots denote that the states which have Low Mean years in School with higher labour Power

Labour force participation vs mean year in school



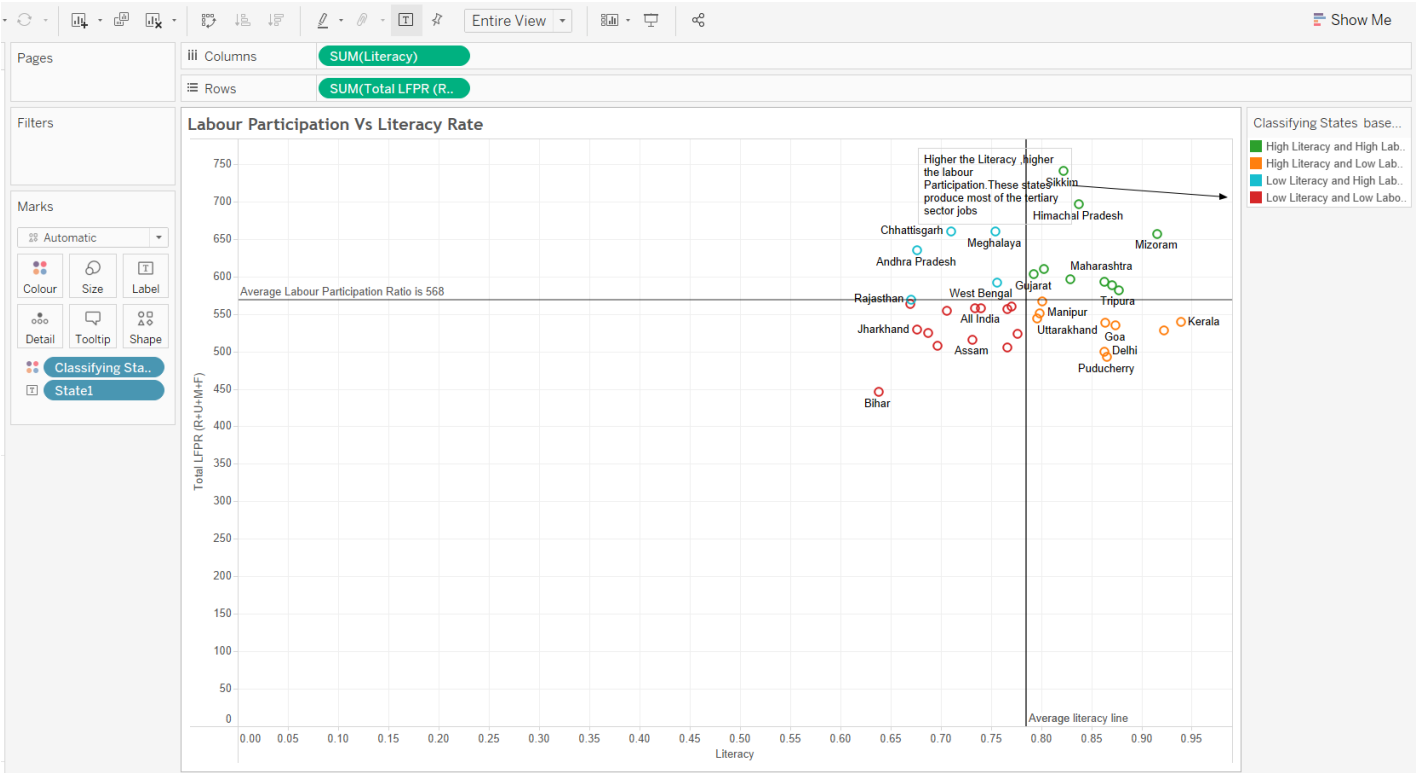
Insights

- The longer people stay in school, the higher the participation in the labour force.
- In State of Kerala the Mean years in Schooling is lesser and have higher literacy rate but due to the other external in-state regulations they have lesser labour force.
- A reference line is being used just to illustrate how narrow the differences are but making a whole new difference.

Labour force participation vs literacy rate

- Literacy rate have direct relationship with LPR and overall economy. Better literacy rate lower the poverty occurrence in the nation.
- To calculate these differences we use Scatter plots in the X axis we have total labour power ratio with the sum of Rural, Urban, Male, Female data. In the Y- axis we have literacy rate.
- The Green Dots Denote States with High Literacy Rates with high labour force ratio.
- The Red Dots Denote States with Low Literacy Rates with Low labour force ratio.
- The Orange Dots Denote States with High Literacy Rates with Low labour force ratio.
- The Blue Dots Denote States with Low Literacy Rates with High labour force ratio.

Labour force participation vs literacy rate



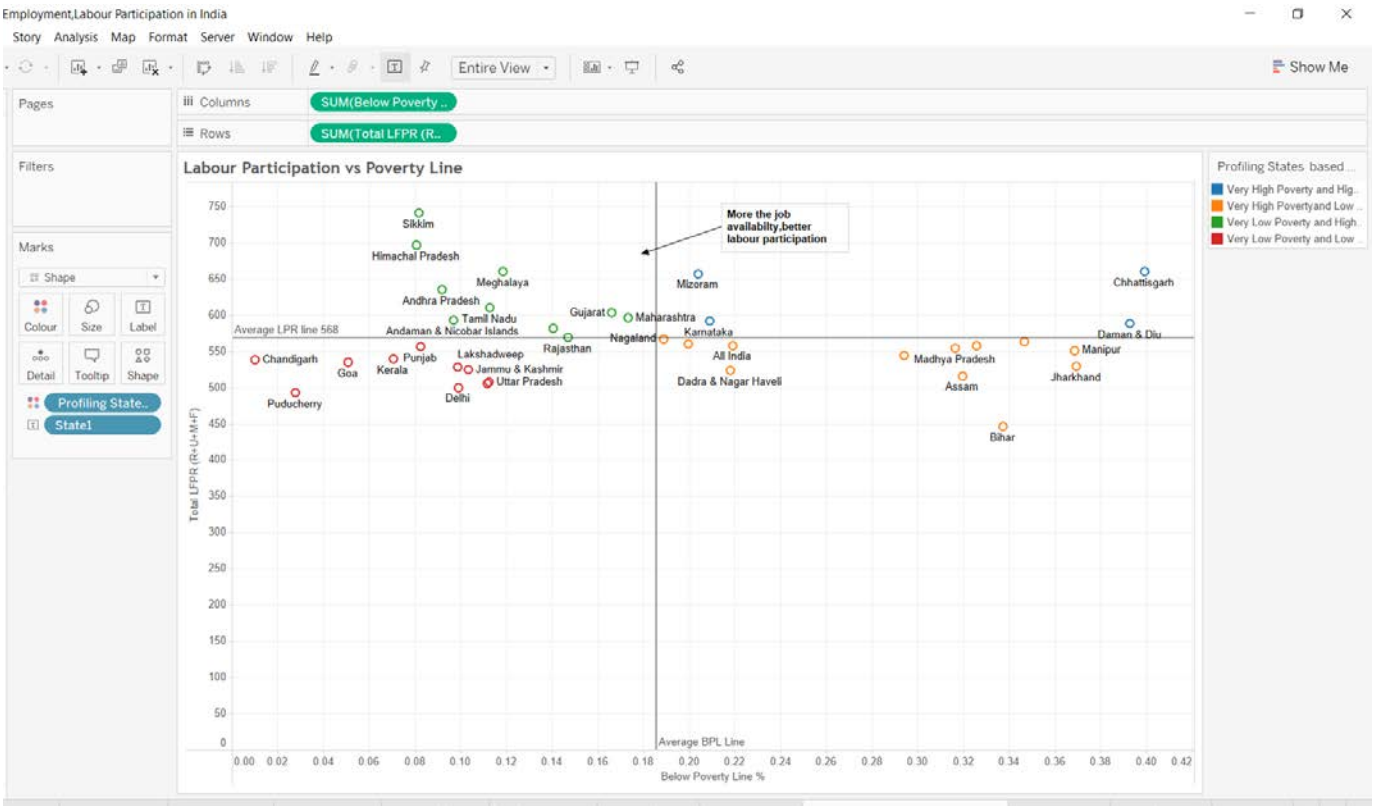
Insights

- Higher the Literacy, higher the labour Participation. These states produce most of the tertiary sector jobs.
- Some States have higher labour force but less rates of literacy. Landing a job in other sectors is always been an issue.

Labour force participation vs Poverty Line

- Poverty line is crucial as the 6% of the total population is in poverty and earn less than 2$ per day
- To calculate these differences we use Scatter plots in the X axis we have total labour power ratio with the sum of Rural, Urban, Male, Female data. In the Y- axis we have Poverty Line.
- The Green Dots Denote States with Very Low Poverty with high labour Participation.
- The Red Dots Denote States with Very Low Poverty with low labour Participation.
- The Orange Dots Denote States with Very high Poverty with low labour Participation.
- The Blue Dots Denote States with Very high Poverty with high labour Participation.

Labour force participation vs Poverty Line

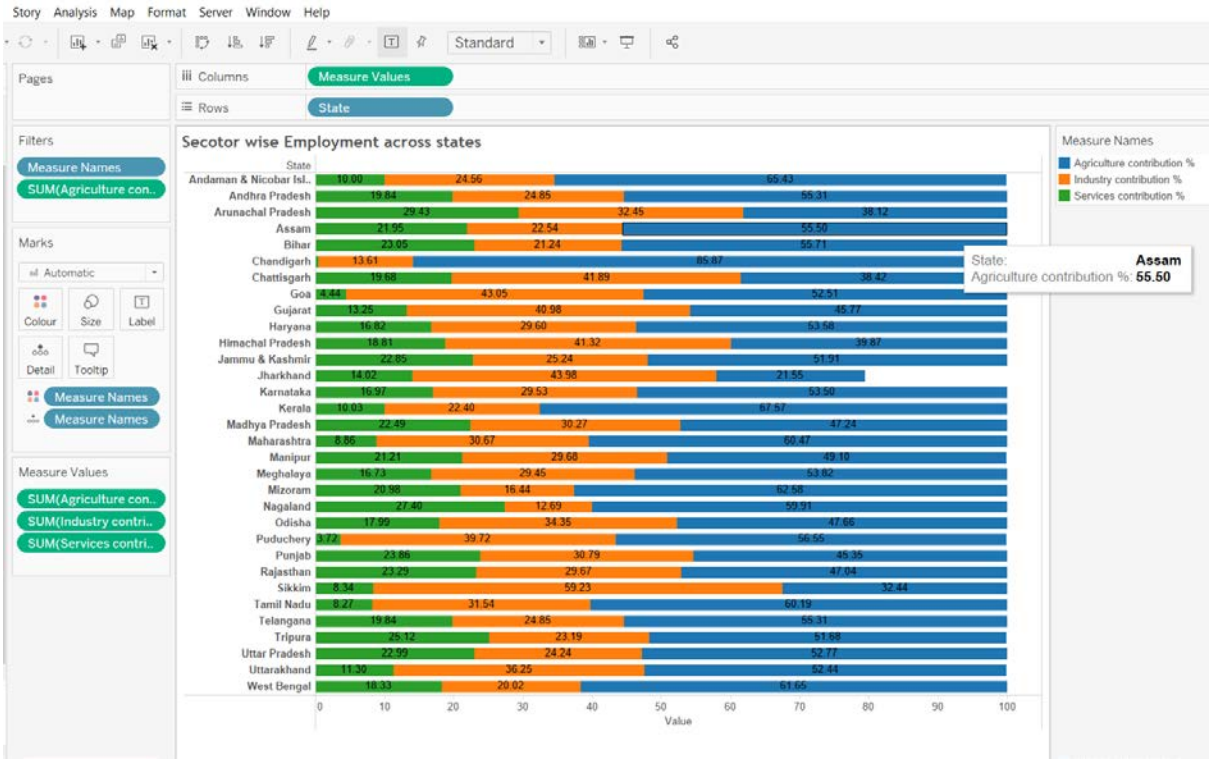


Insights

- More the job availability, better labour participation.
- In Puducherry the population is low and the poverty ratio is also low but since there is less population the labour participation is less.
- When we have better literacy rate we'll have better jobs with better jobs we have less number of poverty occurrence.

6. What are the three major sectors people fall under?

- The main Idea is to showcase the major sectors that the people fall under so the crew used Stacked Bar Chart in order to show the distribution with visual colour encoding.
- Stacked Bar Charts are used to show how a larger category is broken down into smaller categories and what is the relationship of each piece to the total amount? There are two types of Stacked Bar Charts: Simple Stacked Bar Charts place each value for the segment after the previous one.



- The Dataset contains multiple Data fields where service Industrial Contribution and Agricultural Contribution have been filtered out on the details on demand.
- The columns consists of measure values that's just the sum of factors such as service contribution (Government Jobs), Industrial Contribution and Agricultural Contribution.
- The rows indicates the name of the States.
- These columns are visual encoded with muti-colours and details

Insights:

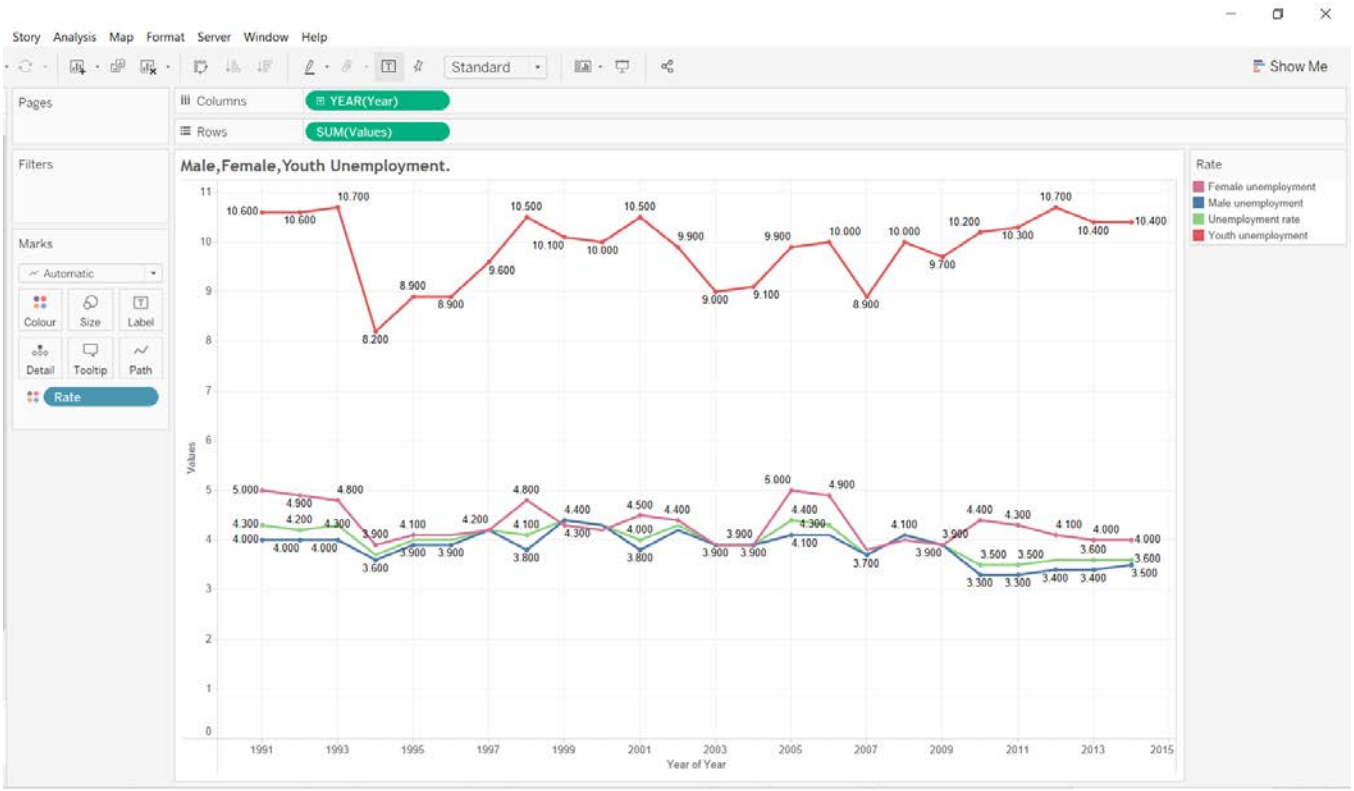
- On average 50% of the population in each state involved in agriculture & related sector. Industrial output from Tamil Nadu is highest than other developed states.
- Industry and service jobs indicates labour force advancement which is a good indicator of development.
- North India depends on Agricultural sector for employment.
- Since north India has higher number in Agricultural sector that has a huge impact in the overall literacy rate.

7. How is the unemployment rate fluctuate throughout the years?

- Unemployment rate of 2 % to 4% is considered healthy according to ILO. Youth Unemployment is higher since India has high population of young people. Except in 1995 youth Unemployment is around 10%.
- Unemployment and underemployment in India are caused by more basic structural factors such as lack of capital, use of capital-intensive technologies, lack of access to land for agricultural household, lack of infrastructure, rapid growth of population resulting in large annual increments in labour force year after year
- In-order to visualize the complete structure we use the line graph for better understanding how the unemployment fluctuates over the course of these years.
- In X axis we have the values in the Y-axis we have years from 1991 -2015.

- The Red Line indicates the country's total youth unemployment.
- The Green Line indicated the country's entire unemployment.
- The Pink Line Denotes the Female Unemployment.
- The Blue Line Denotes the Male Unemployment.

Male, Female, Youth Unemployment



Insights

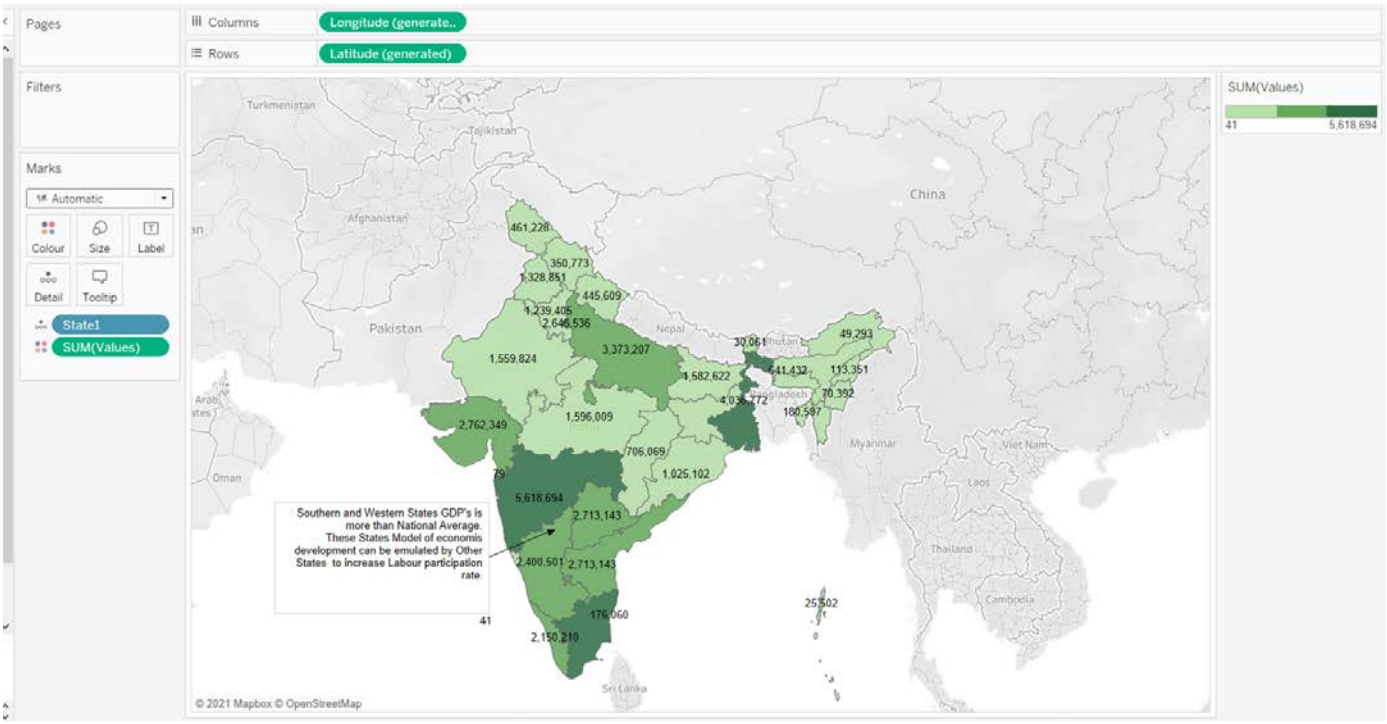
- The Unemployment Ratio between Male and Female are more or less similar and has an even number of fluctuation.
- The overall youth unemployment has been stabilized for the past few years between 2007 and 2015 this is because of the increasing literacy rate because more people are going away from agriculture these years.

8. How corruption is affecting the employment?

Considering Overall GDP

- Corruption is one of causes of Unemployment. Economically starving states show higher corruption rate and unemployment.
- By analysing the overall GDP of each States and comparing it together Southern and Western States GDP's is more than National Average.
- These States Model of economic development can be emulated by Other States to increase Labour participation rate.
- Attracting more people into industrial sectors by giving them a proper literacy rate.

Overall GDP of the States



Insights:

- All the Dark green states contribute more GDP% to the entire country's national average.
- If we consider our previous research we can understand that these states have enough labour participation along with an impressive number of literacy rate hence they contribute more via industrialization.

Unemployment Vs Corruption

- To illustrate and compare the unemployment with the corruption crisis we have used the Word cloud because word cloud is a word visualization that displays the most used words in a text from small to large, according to how often each appears.
- So to pull this off we have used the data from the sum of poverty percentage along with the sum of total political corruption recorded.

Unemployment Vs Corruption Word Cloud

Insights:

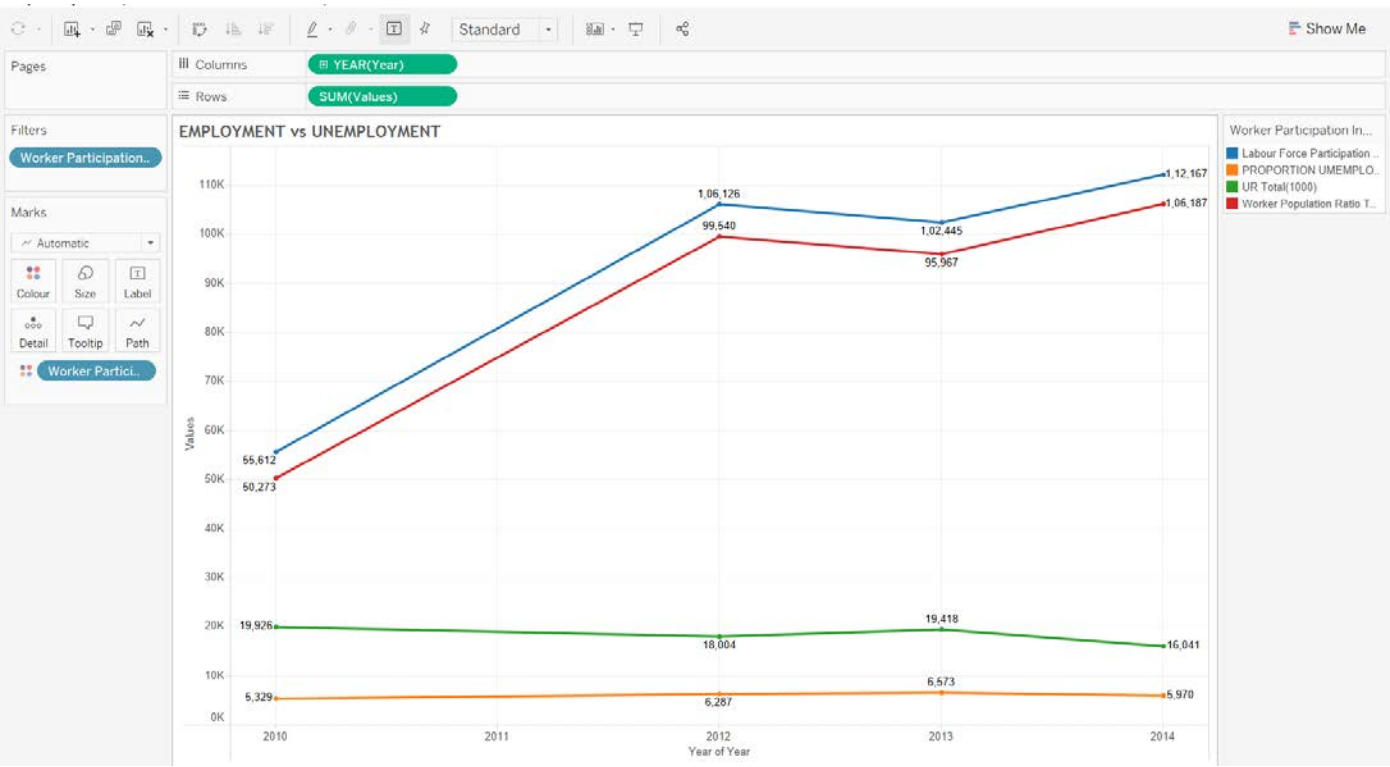
- The state titles with higher number of corruption rate is denoted with a larger size with more bold colour
- These states have higher number of corruption rate and unemployment too.
- Due to the corruption there is a shortage of literacy hence there is unemployment and many of these states depend on the agriculture.

9. Employment vs Unemployment and how unemployment can be decreased to improve the country's GDP

Employment vs Unemployment

- More the Labour Force participation and working population, less Unemployment rate.
- To illustrate how the employment ratio and unemployment ratio is compared we have used the Continuous Multi Line Chart
- The X-axis indicates the values in lakhs and Y- Axis indicates the years from 2010-2014
- The Blue line denotes the total labour Participation from the year 2010 -2014
- The Red Line denotes the total Workers Population Ratio
- The Green line denotes the total Unemployment rate
- The Orange Line denotes the total unemployment proportion of labour force and workforce

Employment vs Unemployment

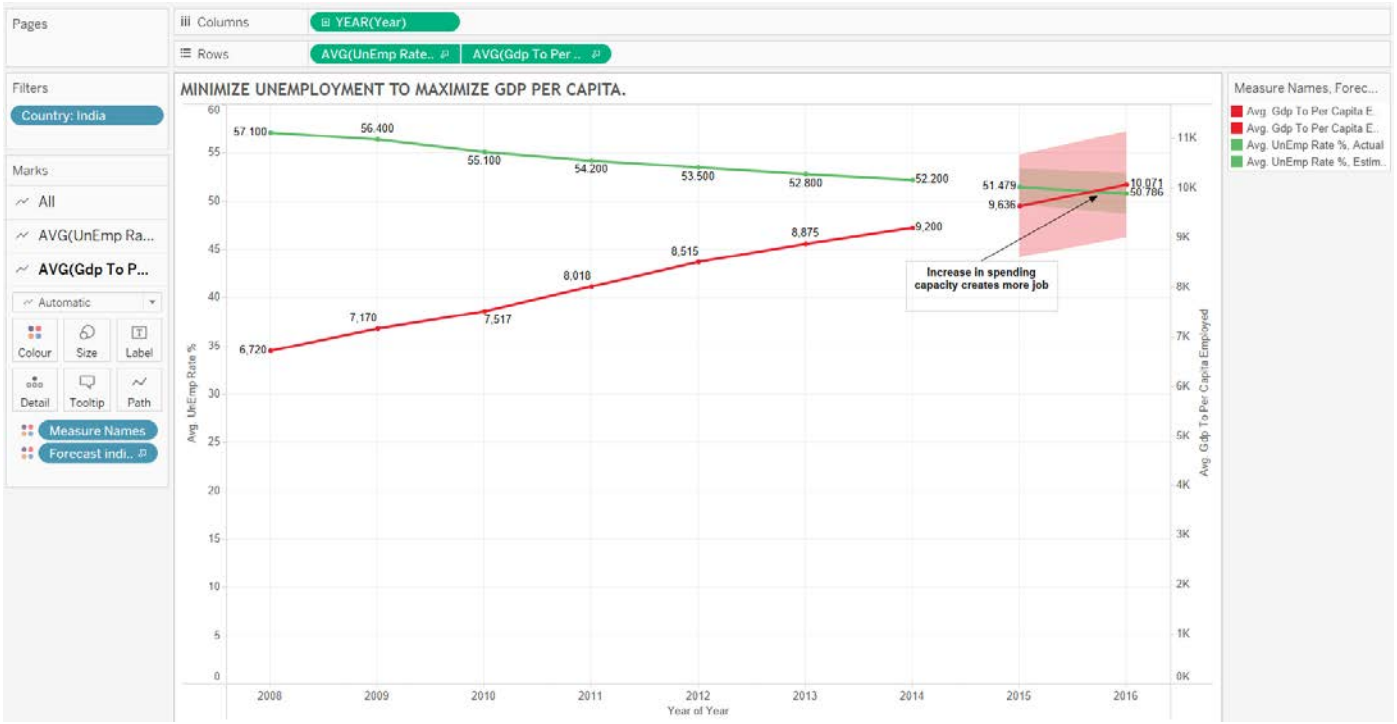


Insights

- As we can see that both the labour force participation and work force population is getting increased dramatically over the past few years.
- This is an occurrence of high increase in population every year. In the year 2030 India is expected to surpass China and become the world's biggest populated county.
- The majority of the labour force participation is in the agriculture sector and the work force is an outcome of the other two sectors, hence there are a high number of illiteracy rate which furthermore causes an impact in the employment exchanges as many of the labour force come from agricultural background that doesn't need any study.
- As we see both the forces gets increased but the unemployment ratio and the employment proportion for unemployment hasn't changed much for the mentioned years that is because with higher work force and higher literacy rated states have these parameters low but our whole intention is to bring this parameters even further in all over Indian States.

10) How unemployment can be decreased to improve the country's GDP?

- More the Labour Force participation and working population, less Unemployment rate.
- Here we have used the Dual Axis Line Chart to understand how dramatically GDP can contribute to lessen the burden of unemployment.
- The left X- axis denotes the average unemployment percentage the right X - axis denotes the average gdp per Captiva and the Y - axis denotes the years from 2008 - 2016
- The green line denotes the average unemployment percentage rate actual values with respect to its estimated value
- The red line denotes the average GDP to per Captiva estimation and Actual GDP per Captiva



Insights

- When the already existing work force are indulged in the work it creates more productivity in the country hence the government get profits and taxes.
- Ditching the corruption and contributing the growing GDP for improving job opportunities for all the people.
- Hence higher the GDP expenditure then the unemployment will gradually decrease just like we have illustrated in the chart.

1) Age Dependency impact on labour force & how literacy is affecting the agriculture?

Age Dependency impact on labour force

- Dependent population are generally economic burden to developing world. But India consider it has future workforce. People have moved away gradually from agriculture related activity. This is due to increase in literacy rate and more job creation.
- The age dependency is the people with 0-15 year old and 60+ years old where they can't work and produce.
- Here to illustrate this we use Dual Axis Line Chart
- The left x axis denotes the age dependency ratio, the right X- axis denotes the labour force and the Y- axis denotes the total number of years from 1991 – 2015

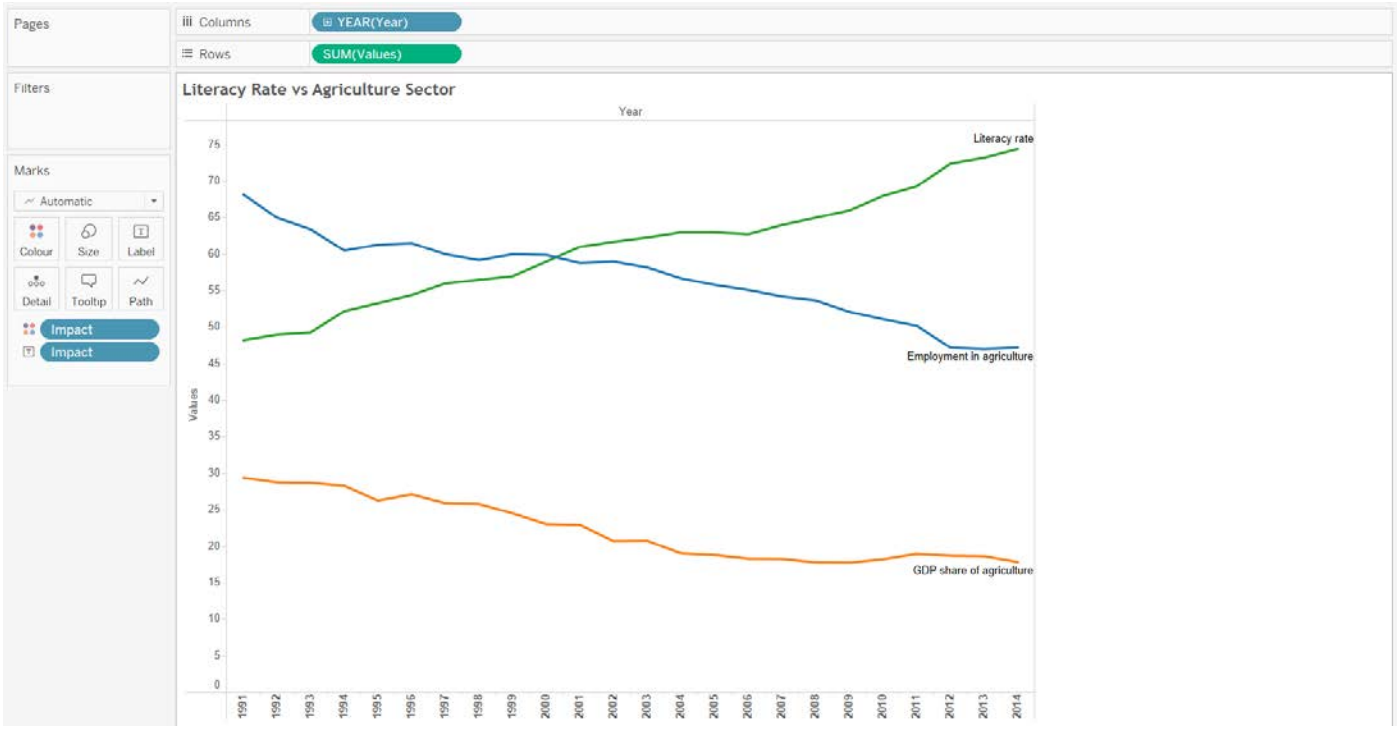
Age Dependency impact on labour force

Insights

- Dependent population are generally economic burden to developing world. But India consider it has future workforce. People have moved away gradually from agriculture related activity. This is due to increase in literacy rate and more job creation.
- There is a flaw in this system, by high literacy rate the job creation using good GDP is fine but still the issue is this all costs a huge impact in the agricultural sector, so demand for food materials is eventually high were the country may be in a situation to import food goods from abroad which will gradually increase the overall money inflation and food costs in the country.

How literacy is affecting the agriculture

- When more people are moving from agriculture to get educated to plot themselves in an industrial sector can have devastating effects.
- From the year 1991 to 2014 the agriculture is dropping a lot due to the industrialization.
- To clearly illustrate this fact we have used Continuous Multi Line Chart
- The X-axis denotes the Values in % and Y –Axis denotes the number of years from 1991 – 2014
- The orange line denotes the GDP generated by the Agriculture sector
- The Green line denotes the Literacy Rate
- The blue line denotes the Employment in Agriculture



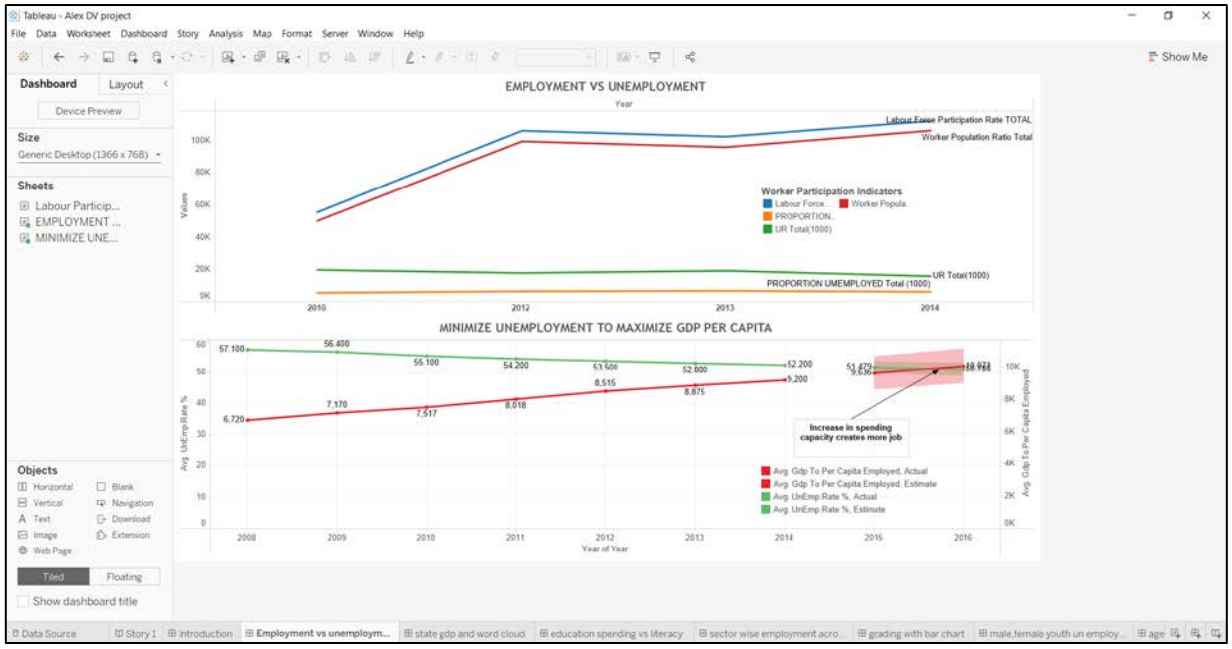
Insights

- As we see when the literacy rate is increased the employment in agriculture is being reduced dramatically
- The overall agricultural GDP contribution has taken a devastating toll in the chart due to this reason]
- This may actually result in a high GDP Average In overall country's job sector but the import of food may cause severe money inflation that can affect people in long term.
- Agricultural sector is also important in the overall GDP production as the 87% of total land in India is meant for agriculture.
- And the national occupation of India is considered to be Agriculture, poor policies and huge corporate involvements have steadily affected the profit percentages in today's agricultural sector.
- Things like land exploitation, fertilizers etc. has also responsible for it major role in this impact on agriculture.

Visualization & presentation:

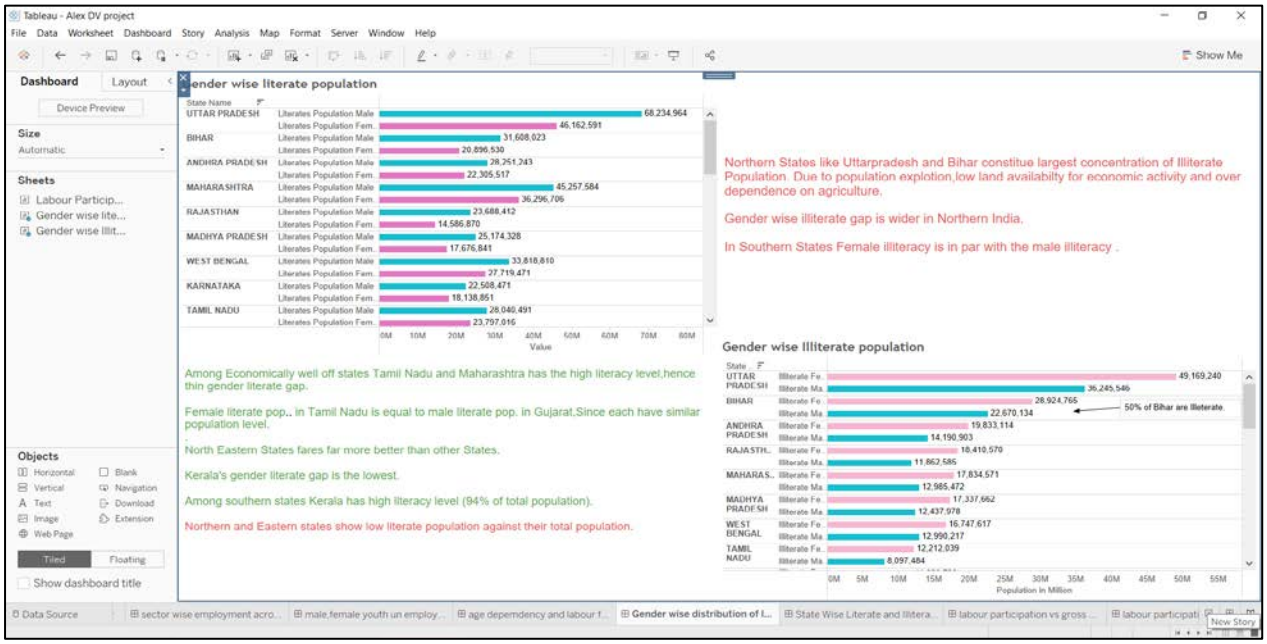
Dashboard:

- The dashboard is a visual display of the most important information needed to achieve one or more goals; consolidated and organized on a single screen so information can be tracked at a glance.
- Dashboards display aggregate information in an intuitive and easy-to-understand manner.
- They provide an overview of the company, business unit, process or project involved in achieving specific goals In addition, they provide easy and immediate access to actionable analytics that can affect an organization's business results
- In this case study the dashboard used not only to present data but also we can easily forecast the trends and analyse the result

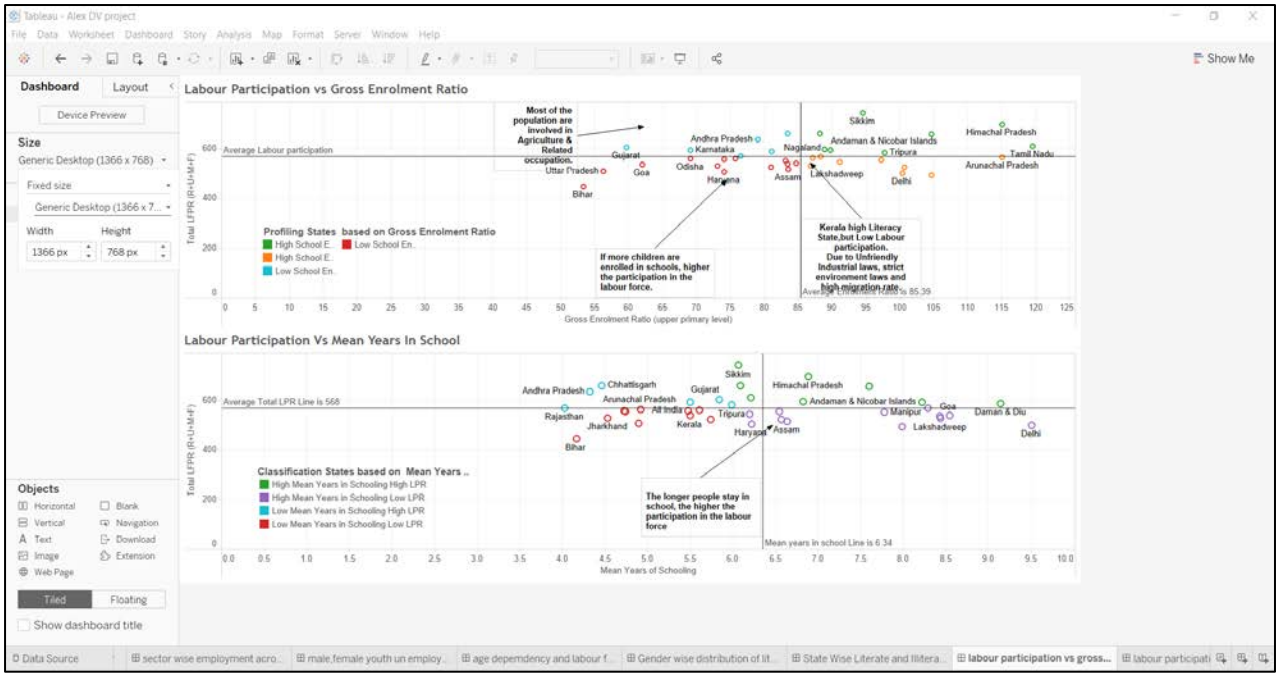


- If we see in the above dashboard, a person can understand the trends of proportions, ratio rate by looking a dash board.

- The dashboard simplifies the complex: we have a lot of information, a lot of data that is constantly changing, and various analytical needs and questions. We want to capture and simplify all this complexity.



- In the above dashboard is not only used to comapre the different charts but also to showcase the insights which audiences can easily understand before even explaing the chart.



- **It tells a clear story**: we can be able to connect data with its context and answer the viewer's questions. The visual layout of the control panel plays a crucial role here.
- **Expresses the meaning of the data**: the selected data visualizations must correctly represent the data and the information you want to extract from them.
- **It reveals details as needed:** we want every viewer to have access to the data they need - not less, but also no more. Some users may need to see a more detailed look at the data.
- Others might be comfortable with an overview.
- While each data dashboard has its own requirements, limitations, and goals, there are certain guidelines that are almost always relevant to creating a dashboard.
- According the data analyst it is a comprehensive snapshot for visualizing the required set of data in a micro level.
- The dashboard software gives insight view of data and help to organize and improve in the relevant areas. Get insight into possible problem areas so stakeholders can handle challenges proactively.

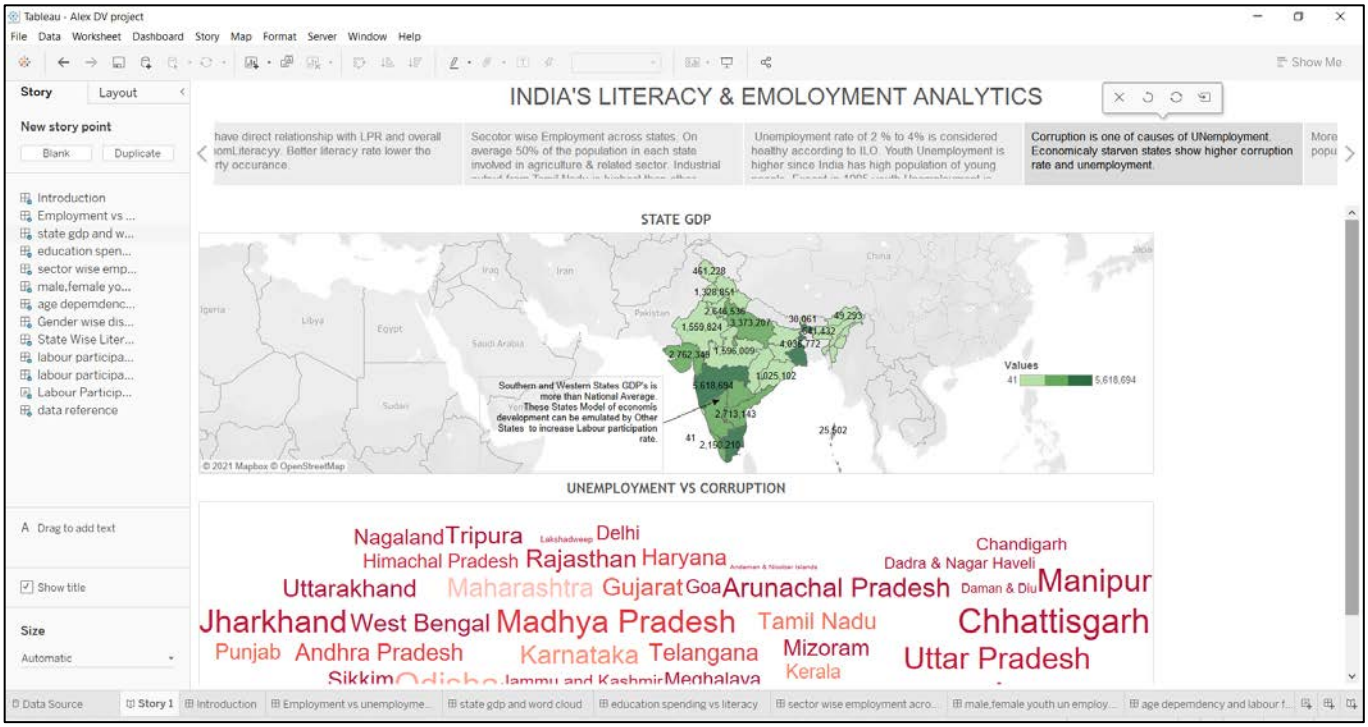
STORY:

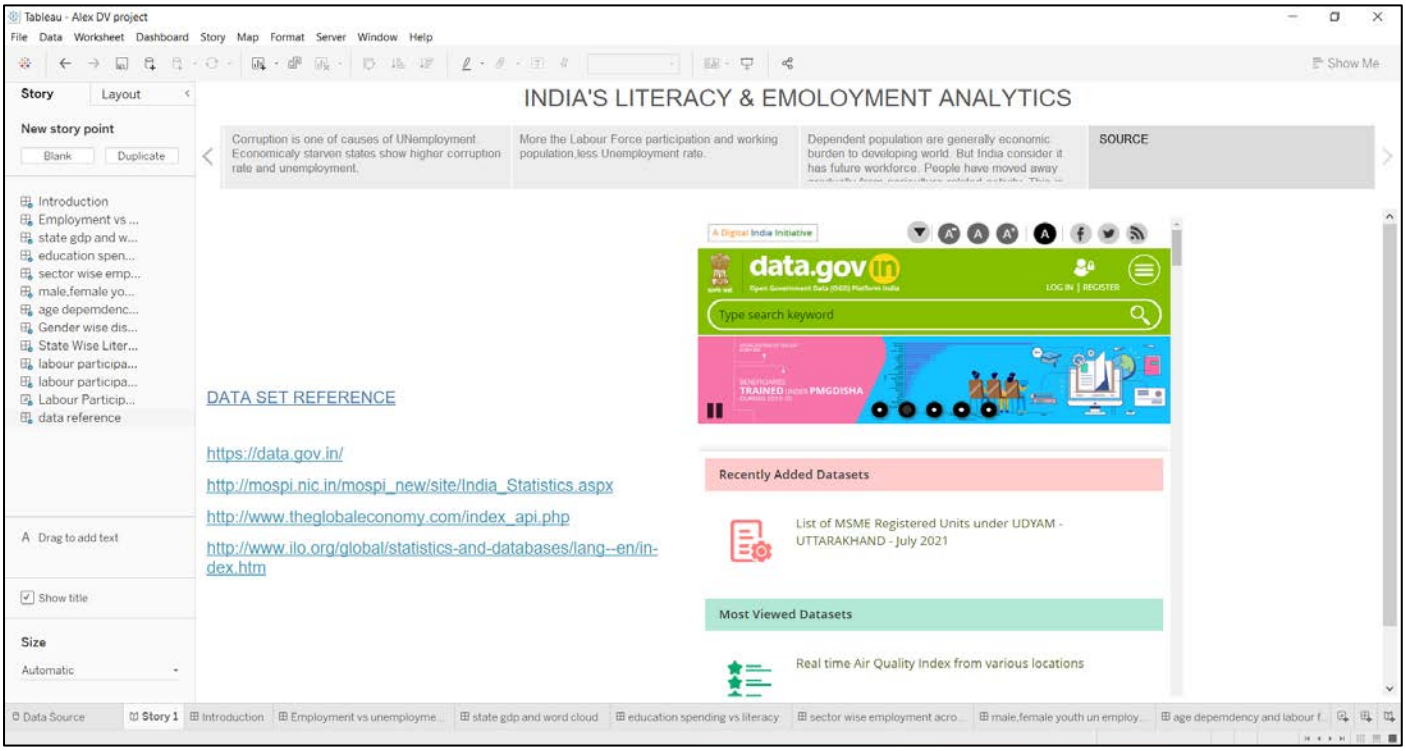
- In Tableau, a story is a sequence of visualizations that work together to convey information. WE can create stories to tell a data narrative, provide context, demonstrate how decisions relate to outcomes, or to simply make a compelling case. Each individual sheet in a story is called a story point. Before creating a story the executive had several checklist for narrating the story.
- **Checklist for Designing Basic Data Narration**
- This checklist is intended to remind you of some important considerations regarding the narration of the data. When developing your own process, try to keep these points in mind.
- Not all of these considerations will apply to every project, so part of your job is to balance and prioritize these factors.
- • **Format** - for example, one page infographic, narrated dashboard, presentation slides, etc.
  • Vehicle presentation - for example, tablet, phone, desktop, slide deck, etc.
- **Data**

- ✓ Remember the potential of creating false stories based on interesting but irrelevant patterns in the data, such as false correlations.
- ✗It is okay to explore story ideas, just validate them in any way you need, including further research and contributions from subject matter experts.
- **Story Structure**
- ✓ Structure the story with something as you work, the story can change and evolve, but the basic structure will help you stay clear, concise, and engaging.
- ✗ don't assume that your audience can look at a few graphs on the dashboard to see what story you want to tell.
- **Presentation**
- ✓Think of an appropriate level of detail to be used for the presentation in relation to the intended audience.
- ✗ do not complicate or simplify the information you submit.
  ✓ Make sure your stories and supporting tables match the data.
- **Graphs**
- ✗ do not play a story with potentially misleading time intervals, nor use axes that distort the scale impression.
- **Design elements**
- ✓Use design elements, such as a selection of colours and languages that match the tone and style of the story and match the needs, interests, and context of the intended audience.
- ✗ do not neglect considerations such as language and colour selection. While the data and graphs are clear, your decisions about the descriptive wording and colour palette of the story can have a big impact on what meaning is ultimately communicated to viewers.
- **Simplicity**
- ✓Put the maximum to simplify, streamline, and sharpen your data story.

- A story should start with an introduction which should be in minimal and clear, it should not exceed the screen size .This introduction part make it clear that what type of analytics has been done.

- Apart from the introduction, it is also be very helpful to put a caption which can be result of an insight or explanation of a certain analysis.





- Telling a story should have a proper continuation and in the end the crew have added reference with the website reference which will be helpful for the audience for their future analytics.
- As overall grouping unrelated data seems to contradict, but it is important to tell a story that has not been observed.

- Data analysis is primarily about finding the stories that the data is trying to tell. Once discovered, they need to be presented effectively. By grouping related data, it is easier to see how they change together.

Findings, Results, Solutions and Conclusion

To conclude the project we have look onto our findings of the Analysis,

1) Literacy

- As of now unlike many other Eastern Countries like UAE etc. Gender based literacy isn't a big issue comparatively.
- The total **Illiteracy rate between men and female in India is just 8%** which is a really good sign of economic development.
- **Northern India constitute major part of literate and illiterate population** meaning that it has equal ratio of literacy and illiteracy.
- Sadly the **4 most populous state contribute higher number of Illiteracy**.
- When it comes to educational expenditure government doesn't spend much, same way **majority of the literal population come from the private schools hence the % of people who couldn't afford to go private** school opt for Public schools but still the chances of **having a school in all the populous states in all the district is a challenge hence the illiteracy rate.**

Solutions

- Building more **schools in rural areas**.
- Overall government expenditure for the Education sector has to be increased.
- Encourage people to **enrol their children in Public Schools and Universities**.
- Creating new **innovative courses** for children to choose from.
- Better literacy rate lower the poverty occurrence.

2) Labour Force Participation

- Most of the labour comes from **three major sectors Agriculture, Industrialization, Public Services**.
- **Agricultural labour power is highest** in all the States in India be it economically Weak or well developed all the **states have 50% of labour power being indulged in Agriculture**.
- **Industrialization** is the major sector that is **second largest contributor to the overall GDP** in spite of **being only 25% to 30%** maximum.
- **Services are the last 10% - 22% percentage of labour power**, departments like Information Technologies, Tourism, Banking, Stock Market, and Medical etc. **Public service in spite being low in labour force they contribute to 60% of overall GDP of the country.**
- **Sothern part of India** contributes to **the higher number of labour power** in all of these sectors.
- Most of the population are involved in Agriculture & Related occupation.
- If more children are enrolled in schools, higher the participation in the labour force.
- Some states **like Kerala has high Literacy**, but **Low Labour participation due to Unfriendly Industrial laws**, strict environment laws and high migration rate.

Solutions

- Fixing the Illiteracy rate in weaker states can be useful for expansion of the services sector as it is 8th place in the top 10 countries with highest service sector with a value of 1.5 trillion $
- **The longer people stay in school, the higher the participation in the labour force.**
- Modification of Industrial laws, Environmental laws and Managing Migration Rate should be implemented.
- **States with higher population tend to have less literacy so if we fix that we get more Labour population and Industrial expansion is possible.**
- Unemployment causes lesser overall GDP that constitute the wastage of the labour power so that has to be fixed.
- **The southern and western States Model of economic development can be emulated by Other States to increase Labour participation rate.**

3) Unemployment Issues

- On average 50% of the population in each state involved in agriculture & related sector. Industrial output from Tamil Nadu is highest than other developed states.
- **Unemployment rate of 2 % to 4% is considered healthy according to ILO**.
- **Youth Unemployment** is **higher** since India **has high population of young people**. **Except in 1995 youth Unemployment is around 10%.**
- More the Labour Force participation and working population, less Unemployment rate.
- Dependent population are generally economic burden to developing world. But India consider it has future workforce. People have moved away gradually from agriculture related activity. This is due to increase in literacy rate and more job creation.

Solutions

- Half of the population fall under the sector that has less GDP contribution and an over exploitation.
- Each sector's total labour force has to be equalized so that every state can improve literacy, Literate people tend to be employed easily and contribute more to GDP.
- Increase in spending capacity creates more job for people struggling with unemployment.
- Shaping up and providing the best for the dependent population because they are the future power resource for future labour power.
- Extending the retirement age limit. Age is just a number if you would ask me so at percent the retirement age in any sector tend to be 60 at present. To note in United States its 67.

4) State GDP and Corruption Issue

- **Southern and Western States GDP's is more than National Average.**
- Corruption is one of causes of Unemployment. Economically starving states show higher corruption rate and unemployment.
- **The Country's overall GDP comes from Agriculture: 16% Industry: 25% Manufacturing: 14% Services: 49.9%.**
- Corruption wastes our taxes or rates that have been earmarked for important community projects – meaning we have to put up with poor quality services or infrastructure, or we miss out altogether.

Solutions

- Increasing literacy rates contributes in fighting corruption
- Dealing with unemployment.
- Creating Job Opportunities in all the sectors
- Strict rules has to be implemented on corruption

5) Conclusion

By performing this research and visualization analysis we can understand these following factors to aim for the India Development Moment 2025,

- Agriculture sector has 50% in overall labour force but its GDP contribution is just 16% because the support for labours indulging in the agriculture tend to be moving away to other sectors which is good till a point. This literally means that almost 50% of the labour population is suffering because of no adequate support from the government. So the first step is to help the farmers to get education and government's participation in preserving and economically developing the farmers will help them to evolve in a much GDP contributing sector. Sadly the most the starving and illiteracy rated population come from the states with only agricultural and related occupation that has to be changed with proper literacy and Good government planned green policies to promote agriculture.

- Providing enough literacy to all the States and encourage people from enrolling into public schools and universities so the government intend to spend more on the education sector to preserve the dependent population as they are the future of the country.

- By satisfying all means of literacy rate there will be likely less unemployment crisis in the country.

- Less unemployment will likely improve the overall GDP in the country.

- To improve the overall GDP in the country instead of completely relying only on the agriculture sector by gaining literacy more people should indulge themselves in the other sectors like industry and services.

- To improve industrial sector environmental policies, industrial laws and Migration issues has to be relaxed and sorted out so that it can contribute more to the GDP.

References:

Dataset References,

https://data.gov.in/

http://mospi.nic.in/mospi_new/site/India_Statistics.aspx

http://www.theglobaleconomy.com/index_api.php

http://www.ilo.org/global/statistics-and-databases/lang--en/index.htm

For the project Structure Ideas,

https://mschermann.github.io/data_viz_reader/how-to-run-a-data-visualization-project.html

- Cleveland, McGill: Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. Journal of the American Statistical Association 79(387) (1984)
- Bertin's: The Semiology of Graphics: Diagrams, Networks, Maps. University of Wisconsin Press, Madison (1983); Translated by Berg, W.J.: Original publisher, Mouton-Gautier. Esri Press, Redlands
- Ware: Information Visualisation: Perception for Design, 2nd edn. Morgan Kaufmann, San Francisco (2004)
- Tufte: The Visual Display of Quantitative Information. Graphics Press, Cheshire (1983)
- Tufte: Envisioning Information. Graphics Press, Cheshire (1990)
- Ahlberg, Shneiderman: Visual information seeking: tight coupling of dynamicquery filters with star field displays. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence, Boston, Massachusetts, United States, pp. 313–317 (1994)
- Shneiderman: The eyes have it: A task by data type taxonomy for information visualisations. In: Proceedings of the IEEE Symposium on Visual Languages (1996)
- Chen: Aesthetics versus functionality. Information Visualisation 1(2), 93–94 (2002)
- Plaisant: The challenge of information visualisation evaluation. In: Proceedings of the Working Conference on Advanced Visual Interfaces, Gallipoli, Italy, May 25-28 (2004)
- Carpendale, S.: Evaluating information visualizations. In: Kerren, A., Stasko, J.T., Fekete, J.-D., North, C. (eds.) Information Visualization. LNCS, vol. 4950, pp. 19–45. Springer, Heidelberg (2008)
- Lam, Bertini, Isenberg, Plaisant, Carpendale: Seven Guiding Scenarios for Information Visualisation Evaluation. Technical Report (2010)
- Bruls, Huizing, Wijk: Squarified TreeMaps. In: Proceedings of Joint Eurographics and IEEE TCVG Symp. on Visualisation. IEEE Press (2000)