Tableau DataViz Challenge

TOPIC : Water & Sanitization
Candidate Name: Beryl Coutinho

Problem Statement: Water plays important role in every single's life. Contamination of water can lead to serious health related issues. Testing water quality has become important. Therefore, to curb this problem it is critical to study the water contamination data and take appropriate steps.

   a. Which state has most contaminated water?

   b. Which district leads in water contamination in states?

   c. What is total amount of quality testing parameter present all over the India?

   d. In western region which state has worst water quality?

   e. How much quality testing parameter are present in north, south, east and west region?

   f. What is the water contamination growth rate of the regions in four years?

   g. What is the district wise growth rate over the years?

Dataset source: https://data.gov.in/catalog/water-quality-affected habitations?filters%5Bfield_catalog_reference%5D=94688&format=json&offset=0&limit=6 &sort%5Bcreated%5D=desc
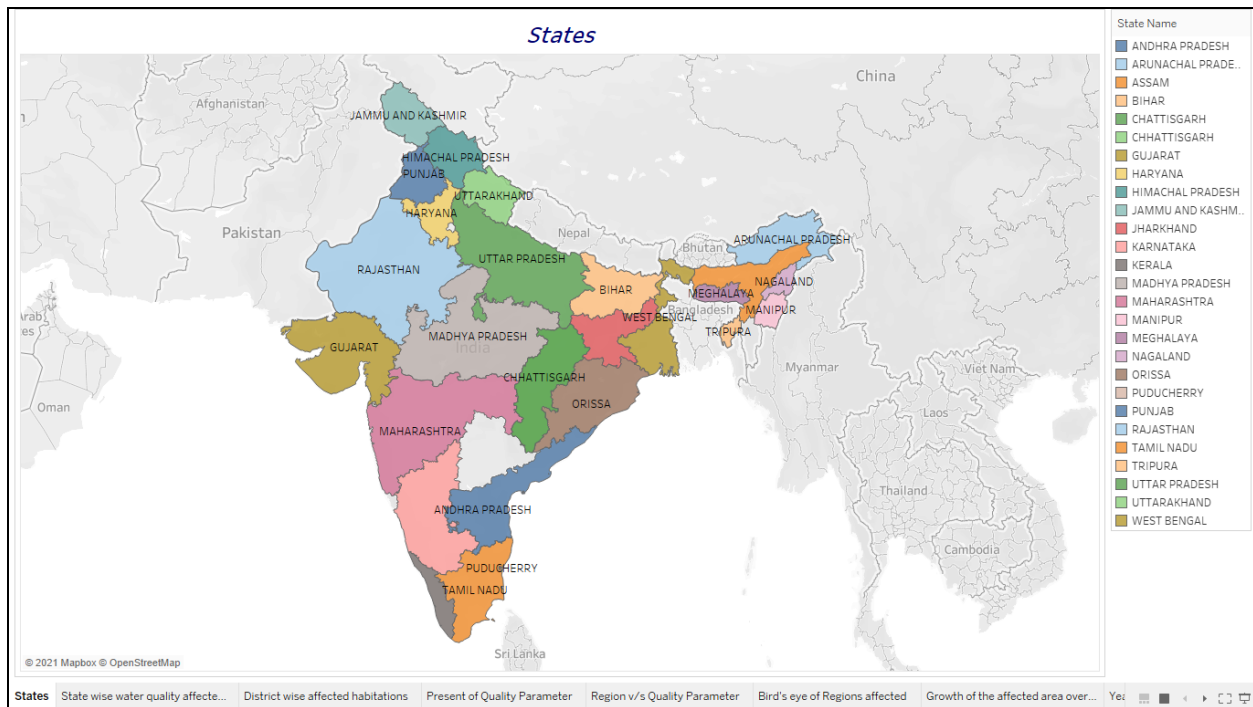
Dataset records: 5,00,000

Fig. 1: States Dataset

Water plays important role in every single's life. Contamination of water can lead to serious health related issues. Therefore, it is critical to study the water contamination data and take appropriate steps to curb this problem. This dataset can be useful for various parties. Government can take appropriate decision by examining the data. This dataset can be also helpful for water filter manufacturing companies to build suitable water filters as per the requirements. This dataset contains data/records of total 26 states of India of four years. Since the dataset contains records/instances of four years it can be used to study the growth pattern of the quality parameter. This dataset contains the records of water quality affected habitations. The dataset contains 8 parameters/attributes like states, districts, panchayat name, habitation, quality parameter, water quality affected habitations, years and regions. Total 5,00,000 records/instances are present in the dataset. Before getting hands on dataset, it's always crucial to know what the dataset is of, what types of attributes are there and instances.
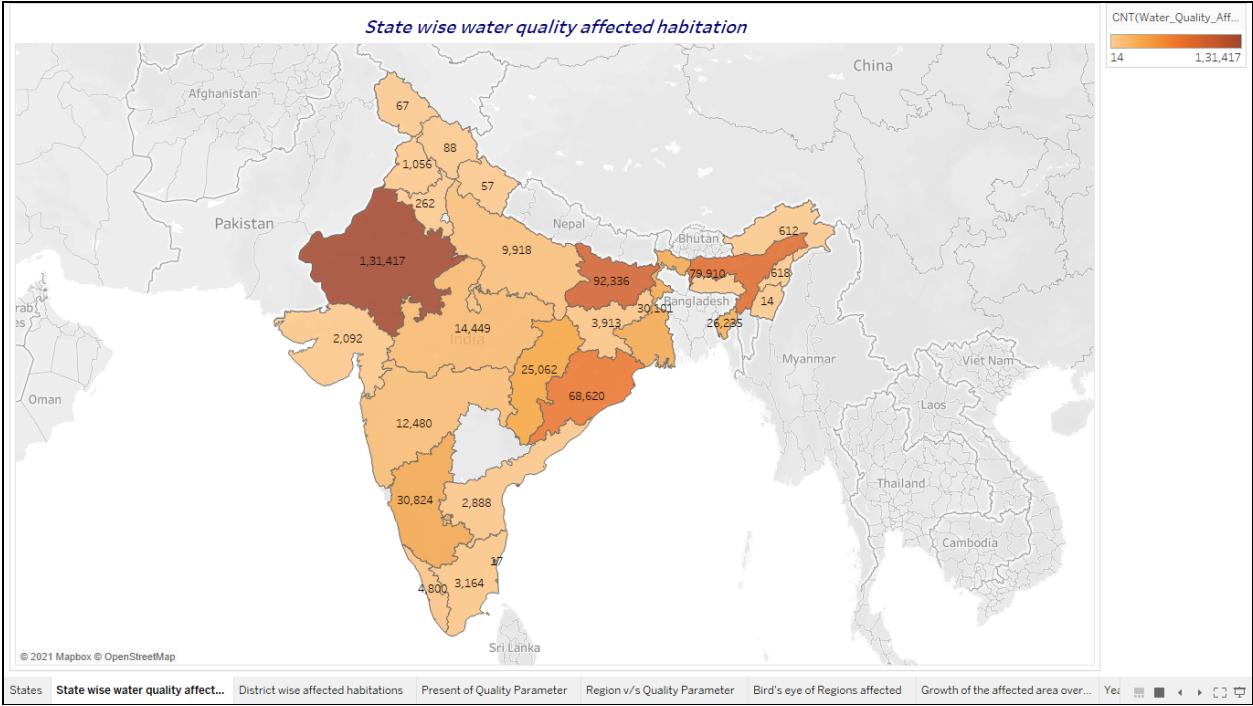
Tableau DataViz Challenge

Fig. 2: State wise quality affected areas.

The above shows the states wise water quality affected habitations. Color shades are used to show the affected areas. Darker the areas/states mean a greater number of affected habitations. This graph will be useful for the government to know badly affected area and take immediate actions on it first.
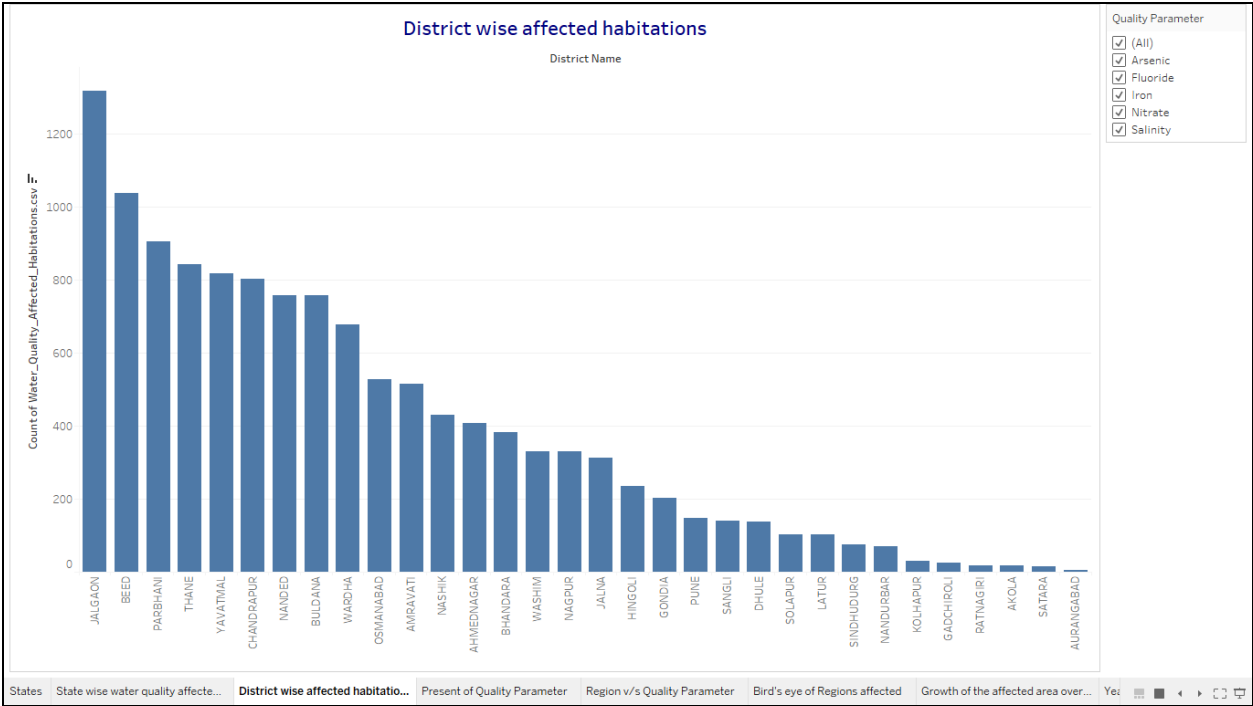
Fig. 4: District wise affected habitations

One of the rules of business intelligence is to unfold actionable insights to make strategic decisions. Equality means each individual or group of people is given the same resources or opportunities while equity recognizes that each person has different circumstances and allocates the exact resources and opportunities needed to reach an equal outcome. Therefore, for a government to take equity-based decision its vital to understand which regions require urgent actions. From the graph it's clear that Jalgaon district exhibit a greater number of affected habitations. Business Intelligence may also demand to know how much quality parameter is present in the particular state. Water quality companies may be interested to look for a particular quality testing parameter like iron and salinity in specific districts to make better business decisions. Quality parameter filter is applied so that we can view graph with different parameters. Therefore, from above graph you can explore graph with different parameters.
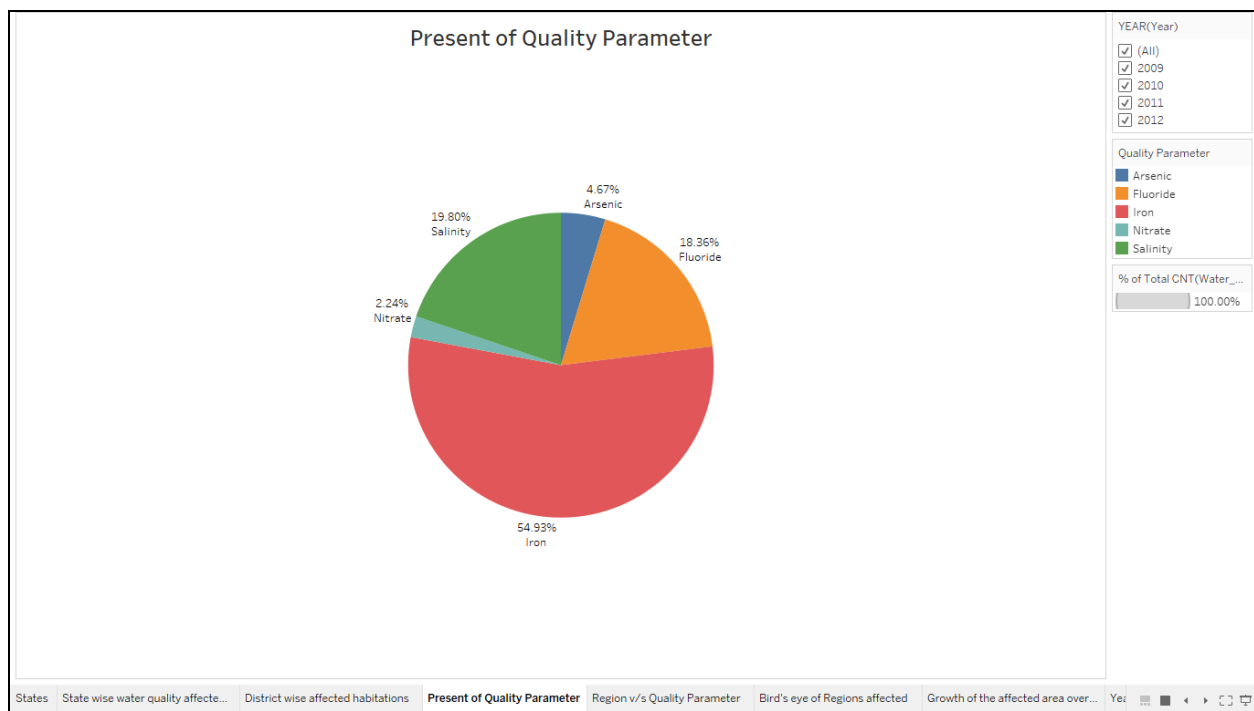

Fig. 5: Total presence of quality parameters.

Total five parameters (iron, nitrate, fluoride, salinity and arsenic) are used to check the quality of the water. Combinations of this parameters can contribute to the contamination of the water which. Businesses may also require to know the presence of parameters all over the states. The above pie chart displays the presence of quality parameters in percentage. Businesses may be also interested to know the total amount of quality parameters detected in different years, to achieve this there is years filter.
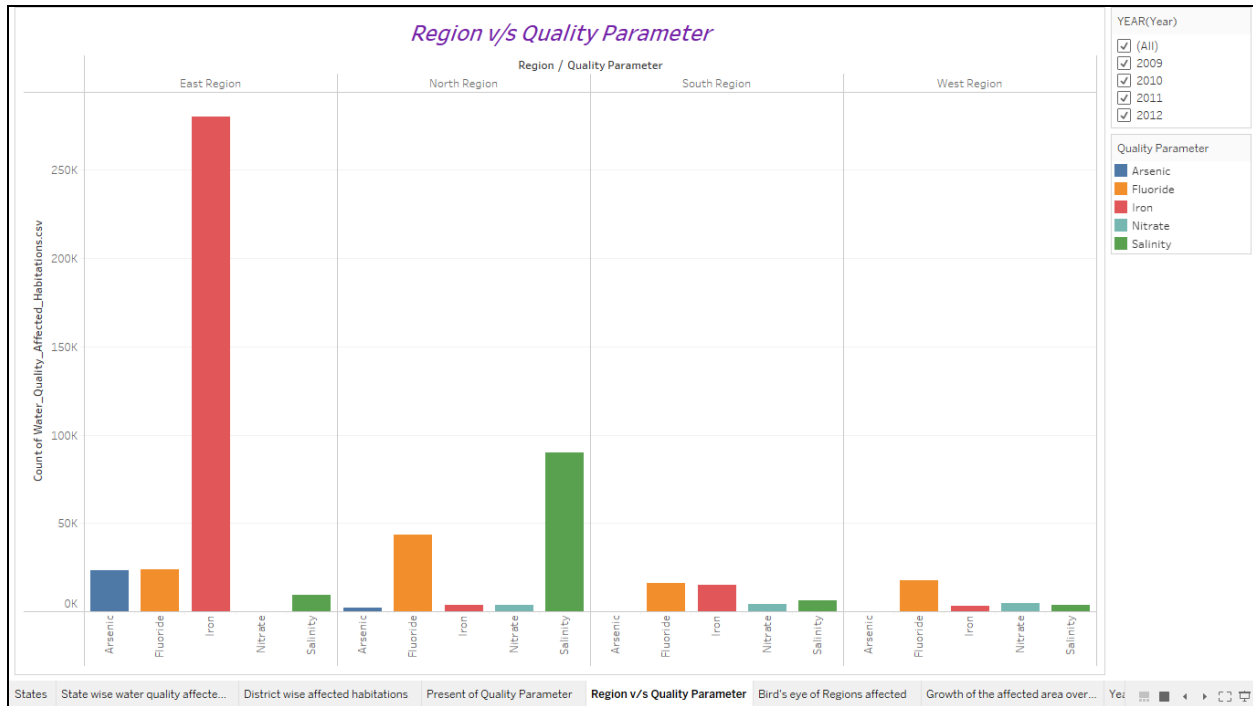
Fig. 6: Region v/s Quality parameter

There are 29 states in India. Looking for individual quality testing parameter of each state will be burdensome and knotty. Thus, we have groups states in four groups like north region, south region, east region and west region. Looking for individual quality testing parameter in above is coherent and uncomplicated and simultaneously we can compare the parameters of the regions. From the above graph its quite recognizable that amount of iron is found maximum in east region and west region states have less amount of iron present. Salinity and fluoride is found in more number in north region.
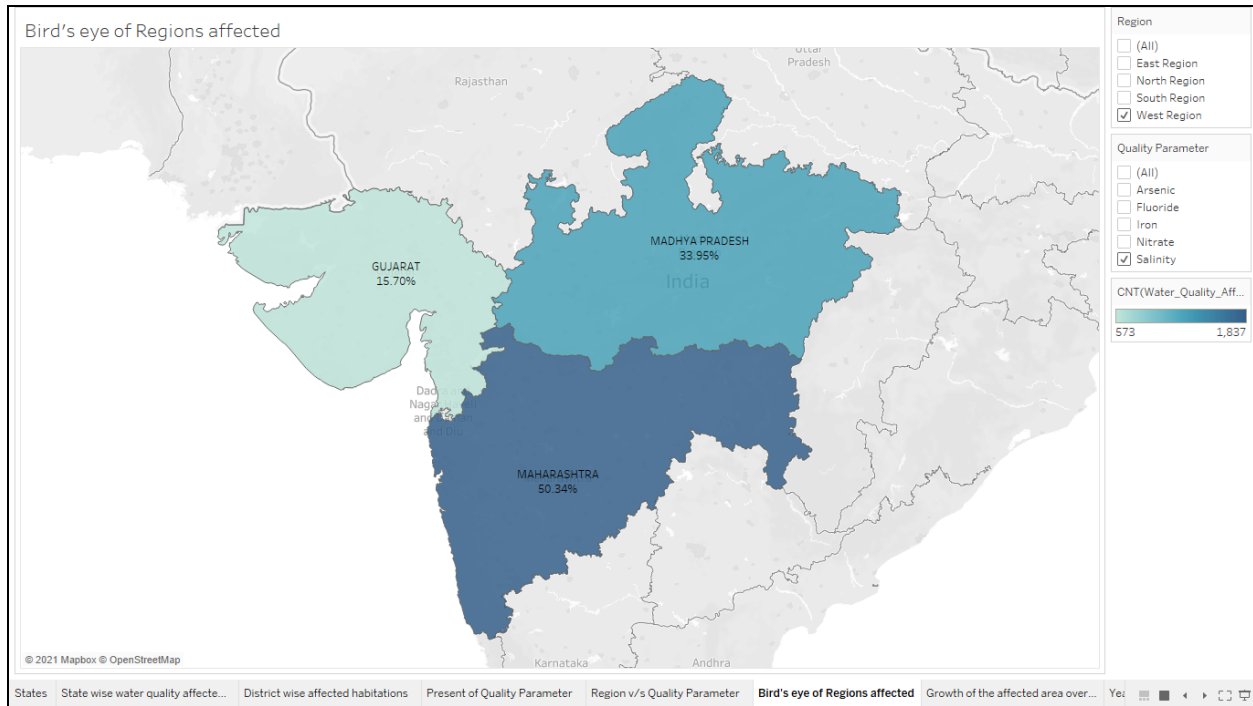
Fig. 7: Bird's eye view of the affected regions

There are 29 states in India. Looking for individual quality testing parameter of each state will be burdensome and knotty. Thus, we have groups states in four groups like north region, south region, east region and west region. Looking for individual quality testing parameter in above is coherent and uncomplicated and simultaneously we can compare the parameters of the regions. Color shades are used to show the affected areas. Darker the areas/states mean a greater number of affected habitations. This graph will be useful for the government to know badly affected area and take immediate actions on it first. We have applied region filters where you can view the state wise affected regions as per the requirements and to make better decision.
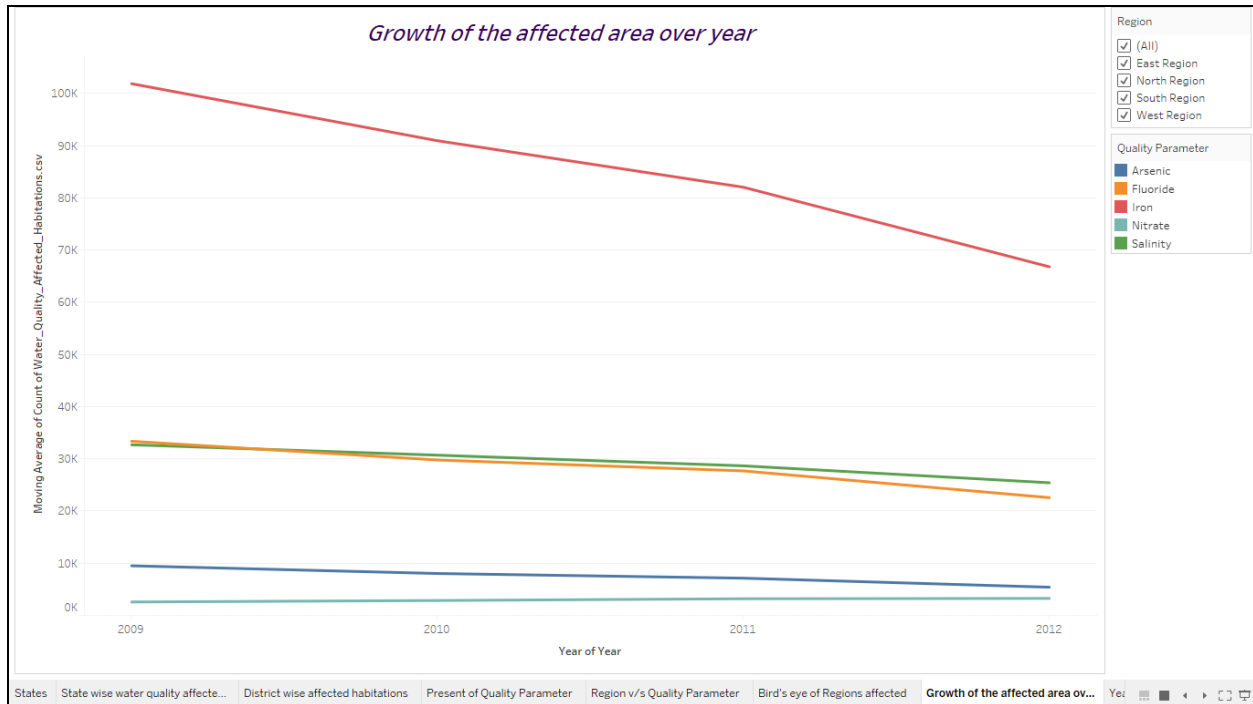
Fig. 8: Growth rate over the year of the quality testing parameter.

A lot of things change over the year. In businesses it's a good practice to look for the pattern change. Similarly, businesses may be interested to know this changing pattern over the year to make decisions accordingly. The lines in above graph represents quality testing parameter change over the year. We have added region filter to examine the growth of the quality testing parameters in particular region. The green color line represents salinity while the orange color stands for fluoride. Iron represented in red. It is observed that the growth of the parameters (iron, arsenic, salinity, fluoride and nitrate) in western region decreases over the year. It is observed that there was a spike increase in iron, salinity and nitrate in western region in 2010 and growth rate decreased in 2012. Similarly, we can apply the filters and examine the growth in various regions like north, south, east and west and take better business decisions.

Fig. 9: Year wise growth of affected habitations.

After studying affected habitations of states and regions businesses may be also interested to examine at the district level for various years and make a comparative study. There are quality testing parameters and districts filters added to the graph to make visualization more understandable and interactive. Year wise you can make comparison by changing the filters of quality testing parameters and districts. In Maharashtra Jalgaon district contributed more in affected habitations. There are five clusters (cluster1-arsenic, cluster2-flouride, cluster3-iron, cluster4-nitrate and cluster5-salinity) formed. This visualization will be helpful for panchayat in districts to take better decisions.
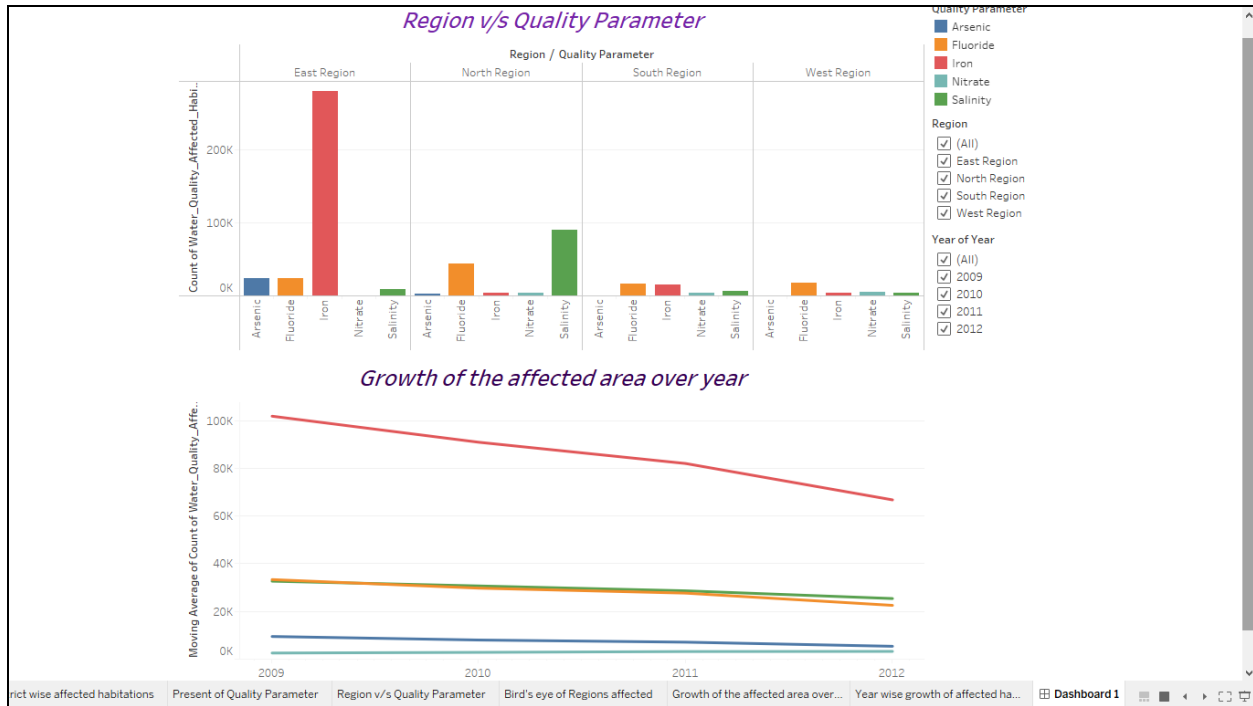
Fig. 10: Region v/s Quality Parameter

What if businesses are interested to know year wise growth rate over the year in a particular region? The above graph answers this question. Region and year filters are applied to the graph so that we can view growth rate in different regions simultaneously and view in different years. In all four years iron was found maximum in east region.
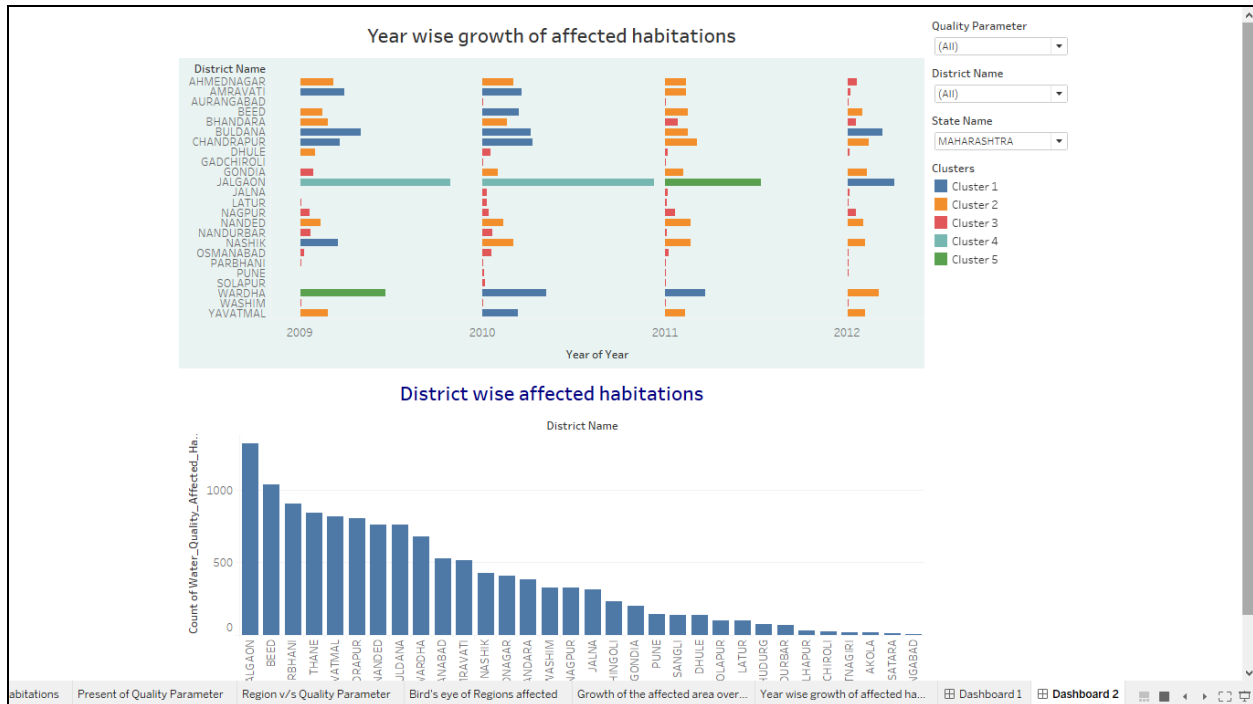
Fig. 11: Year wise growth of affected habitations of districts.

In previous graph we could view the growth of individual parameter region wise but what if businesses want to view the district wise growth affected habitations. The above graph is useful to answer this question. We have applied districts, states and quality testing parameter filters to get insight of all the states, districts and quality parameters. From the above graph its quite evitable that Jalgaon has a greater number of affected habitations.