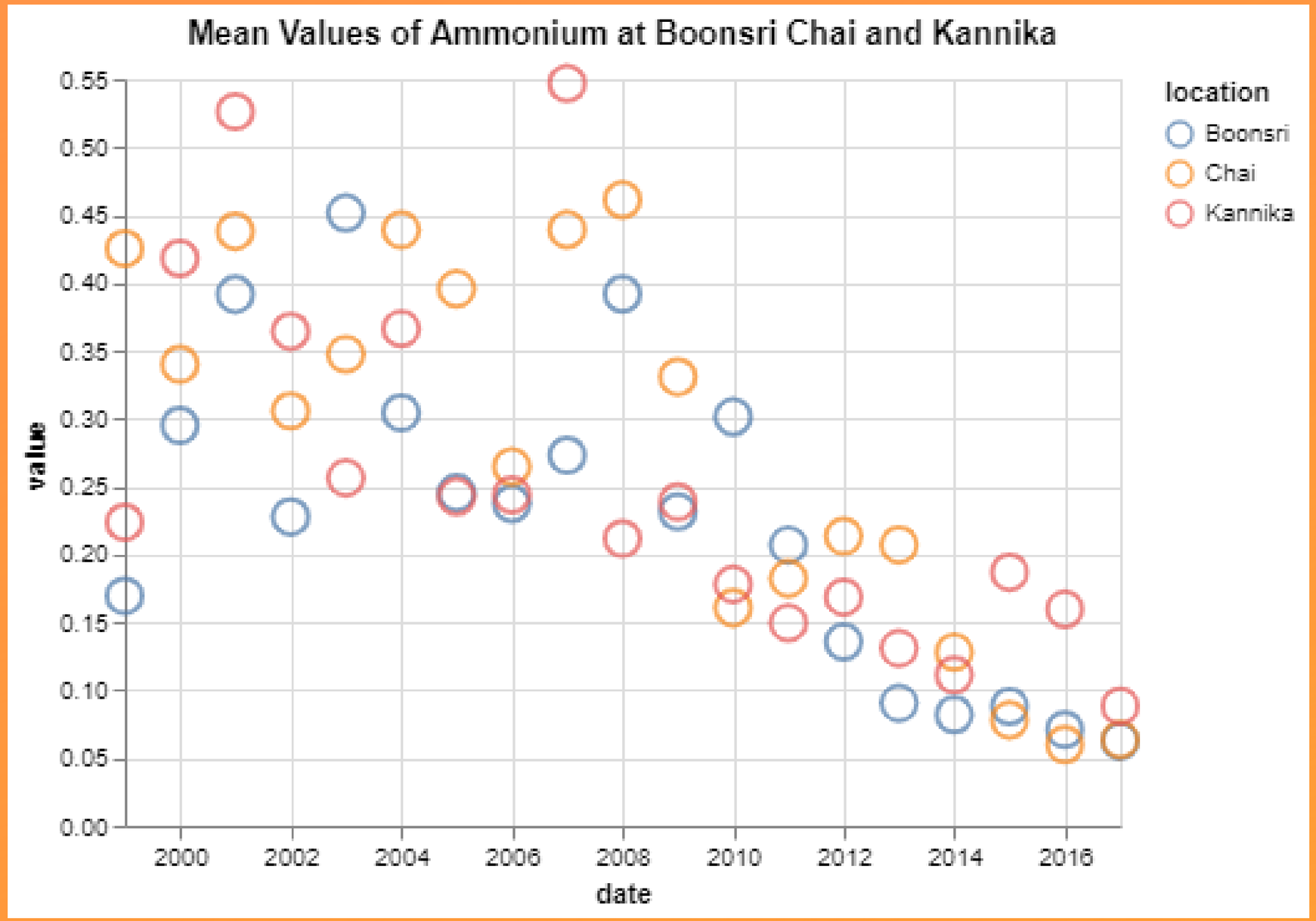
Analysis-1

- **Trends** → Changes over time and/or sensor site
- **Anomalies** → Sudden change over time or one site significantly different from others

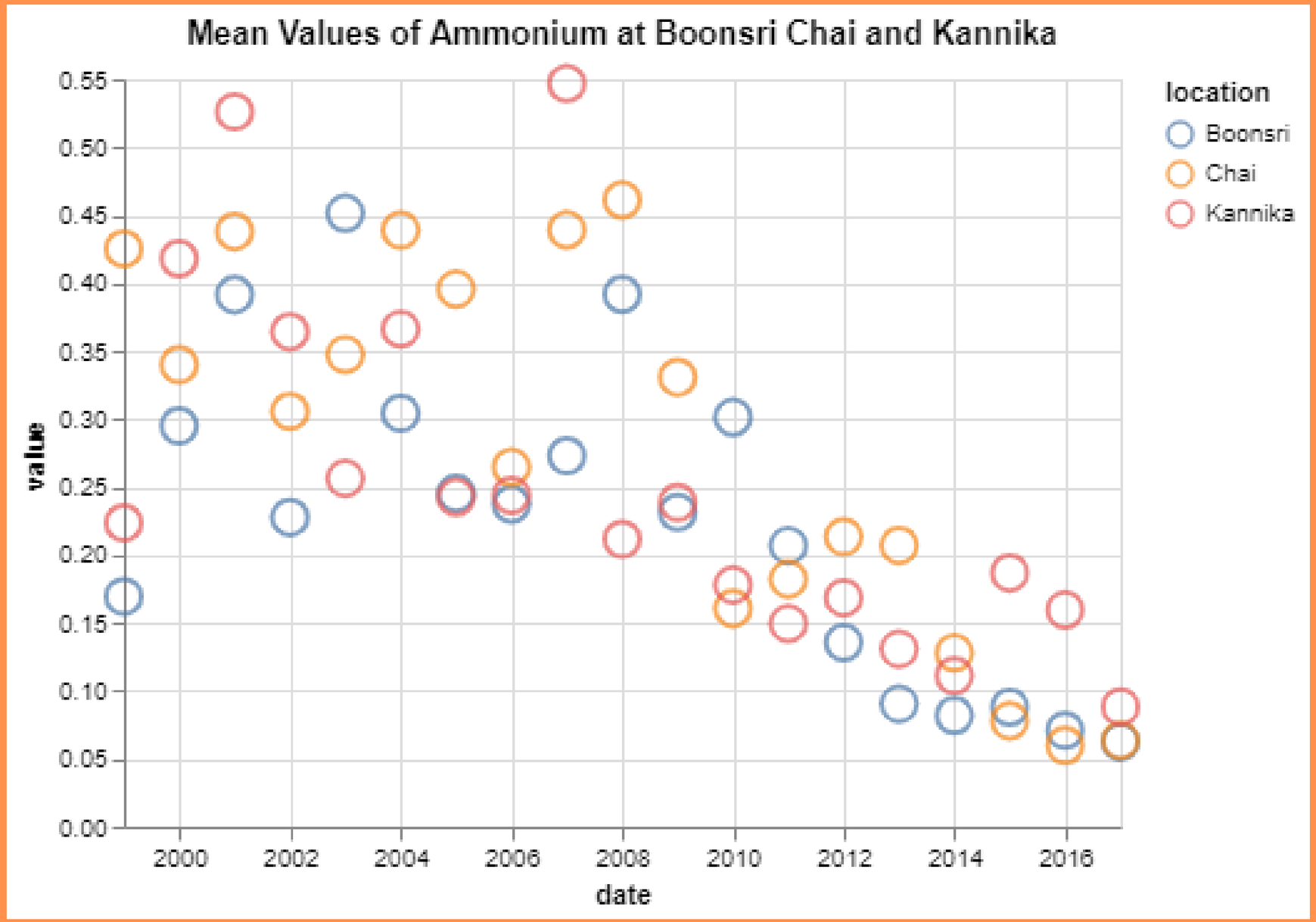
Analysis-2

- **Missing Data** → Missing Data
- **Collectiion Freq.** → Change in collection frequency
- **Unrealistic Values** → Unrealistic values (e.g. water temperature higher than 100 degrees).



The Trend

We can see that the average values of Ammonium at Boonsri, Chai and Kannika declined with the passage of years.



Anomalies

With the decline of values of Ammonium we can see that there are a lot of anomalies. We can see most of them can be found in Chai location such as in year 2007-2008 or in Kannika in year 2007. Despite that the average values have a decreasing graph with time.

For Checking NULL values in the dataset

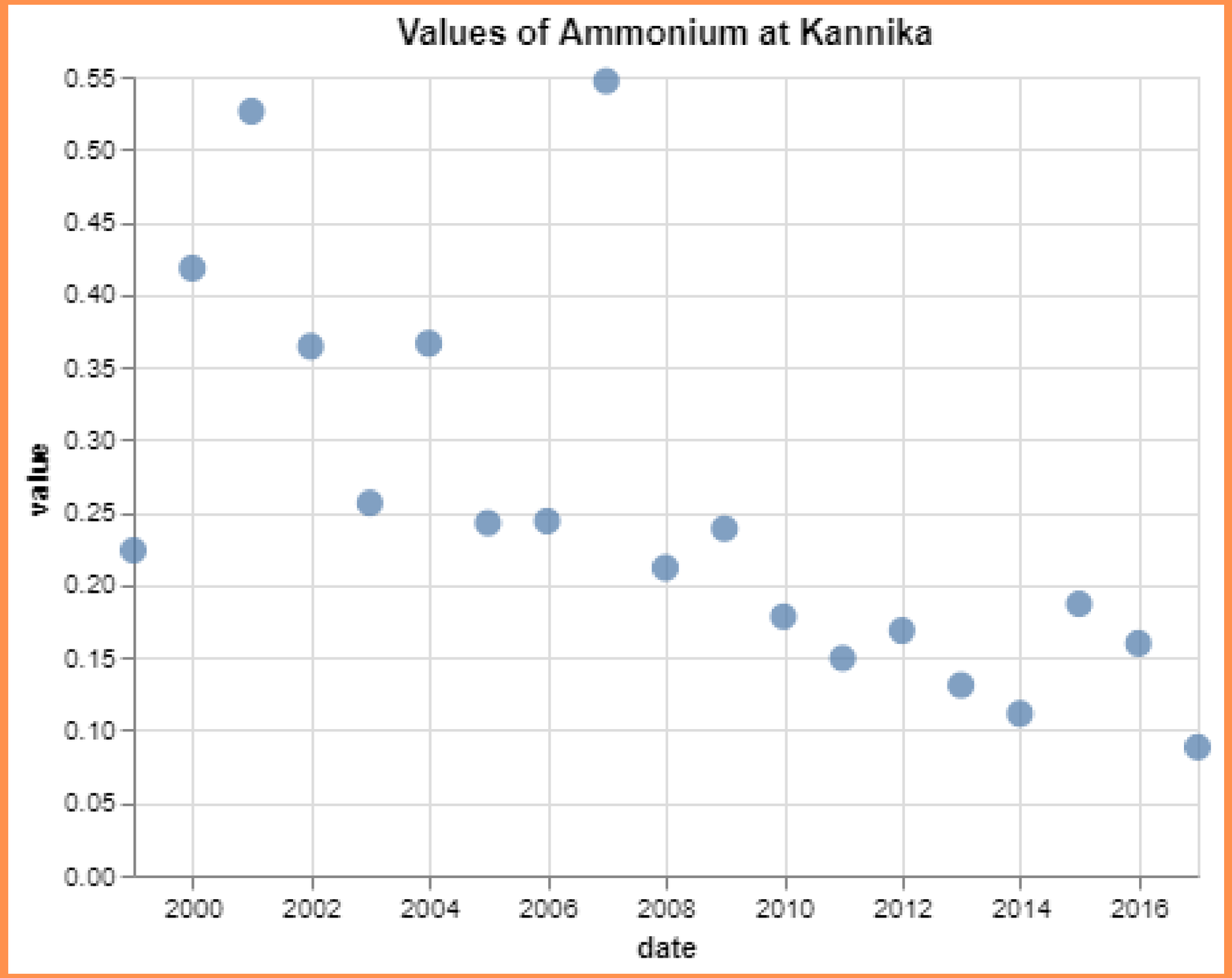
```
df.info() # This gives us all the information about the dataset's insights
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 136824 entries, 0 to 136823
Data columns (total 5 columns):
 #   Column       Non-Null Count   Dtype  
---  ------       --------------   -----  
 0   id           136824 non-null  int64  
 1   value        136824 non-null  float64
 2   location     136824 non-null  object 
 3   sample date  136824 non-null  object 
 4   measure      136824 non-null  object 
dtypes: float64(1), int64(1), object(3)
memory usage: 5.2+ MB
```

Hence we can see that there are no NULL values in the dataset

Missing Data

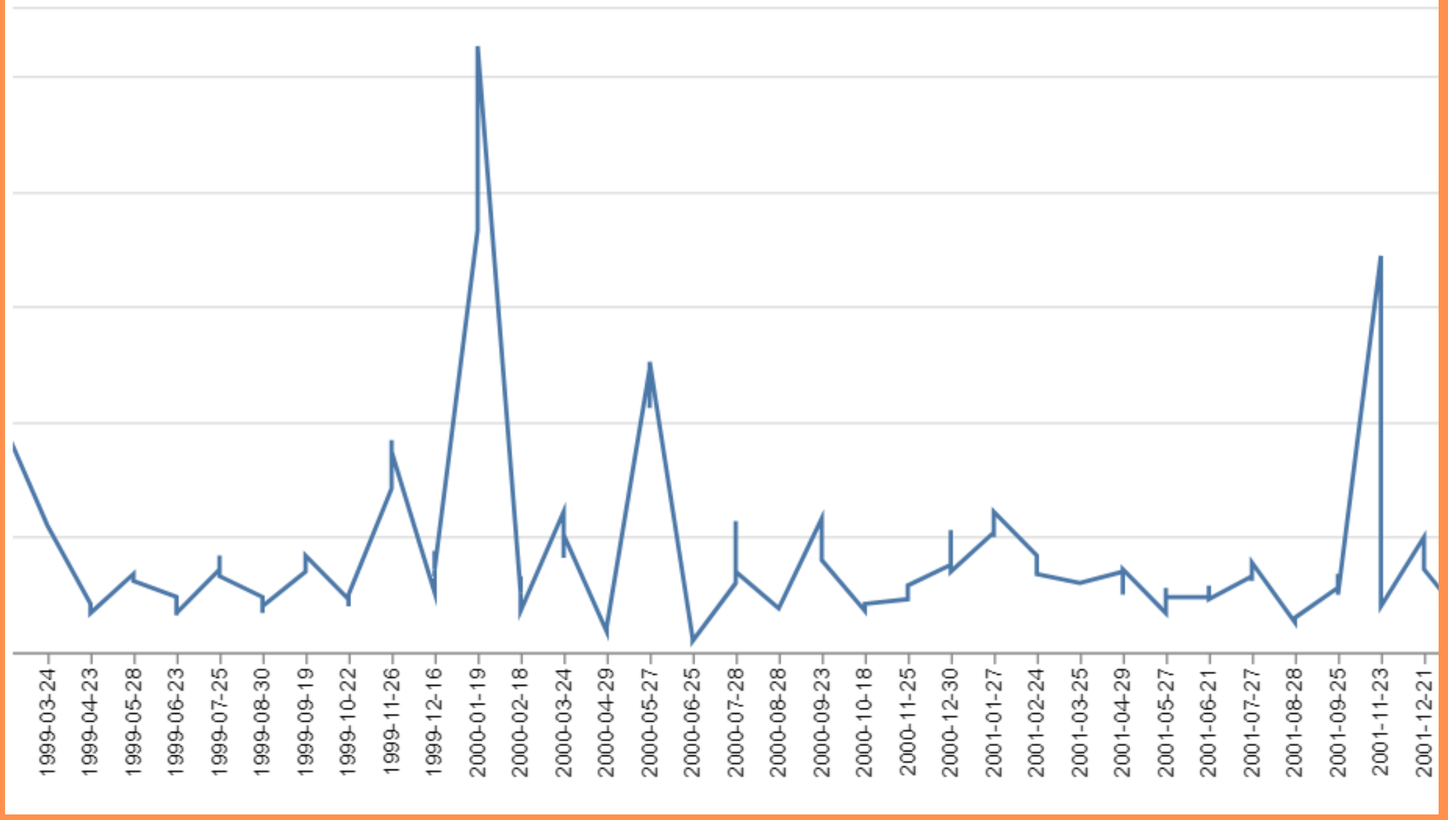
With the help of df.info() we can see that there are no NULL values in the dataset, therefore we do not have to deal with them.



Change in collection frequency

We did filtering, indexing, statistical analysis such as taking mean of chemical values per year to see how the values are changing per year.

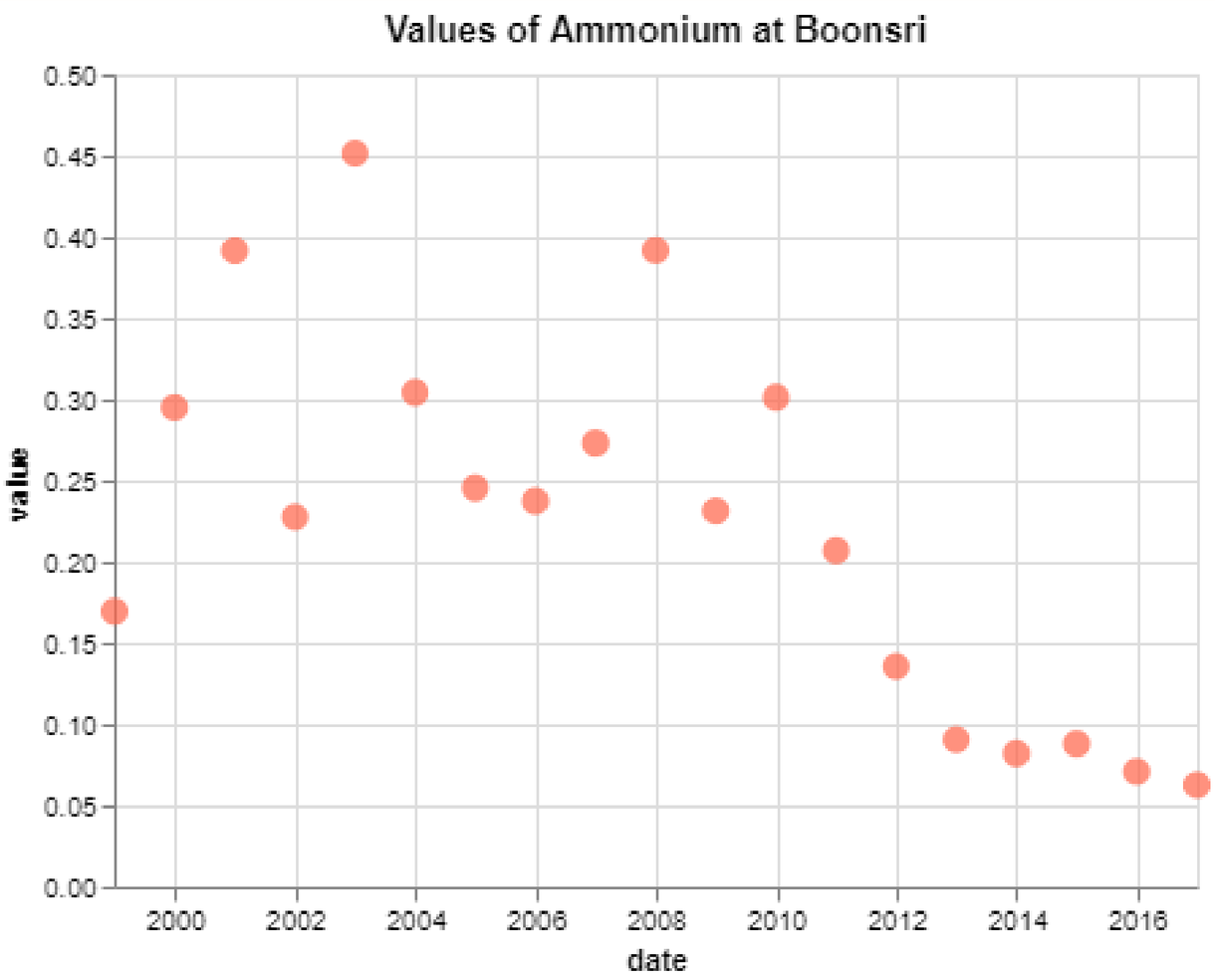
```
alt.Chart(data_sudden_change_Kannika_Ammonium).mark_line().encode(
    x='sample date',
    y='value',
)
```

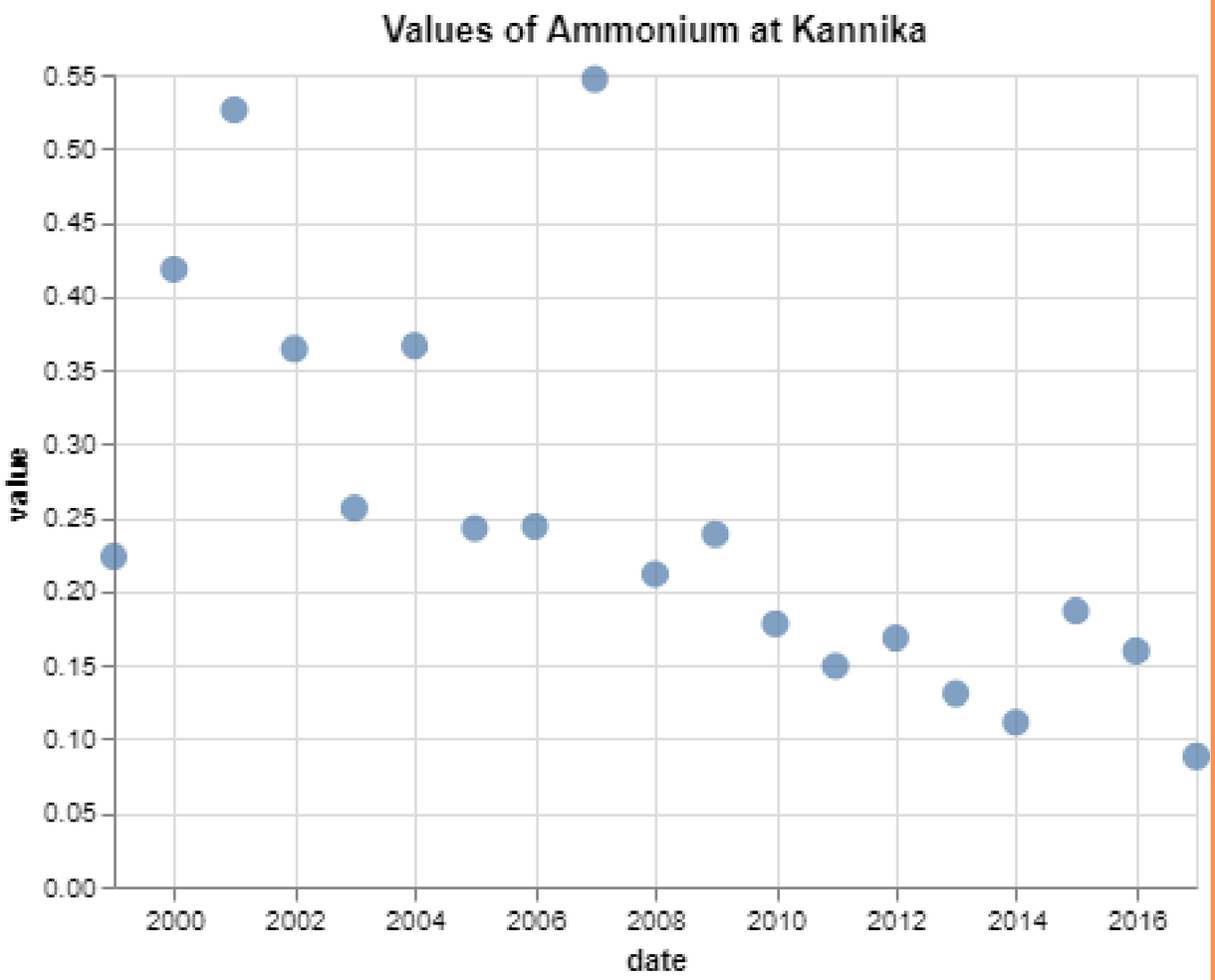


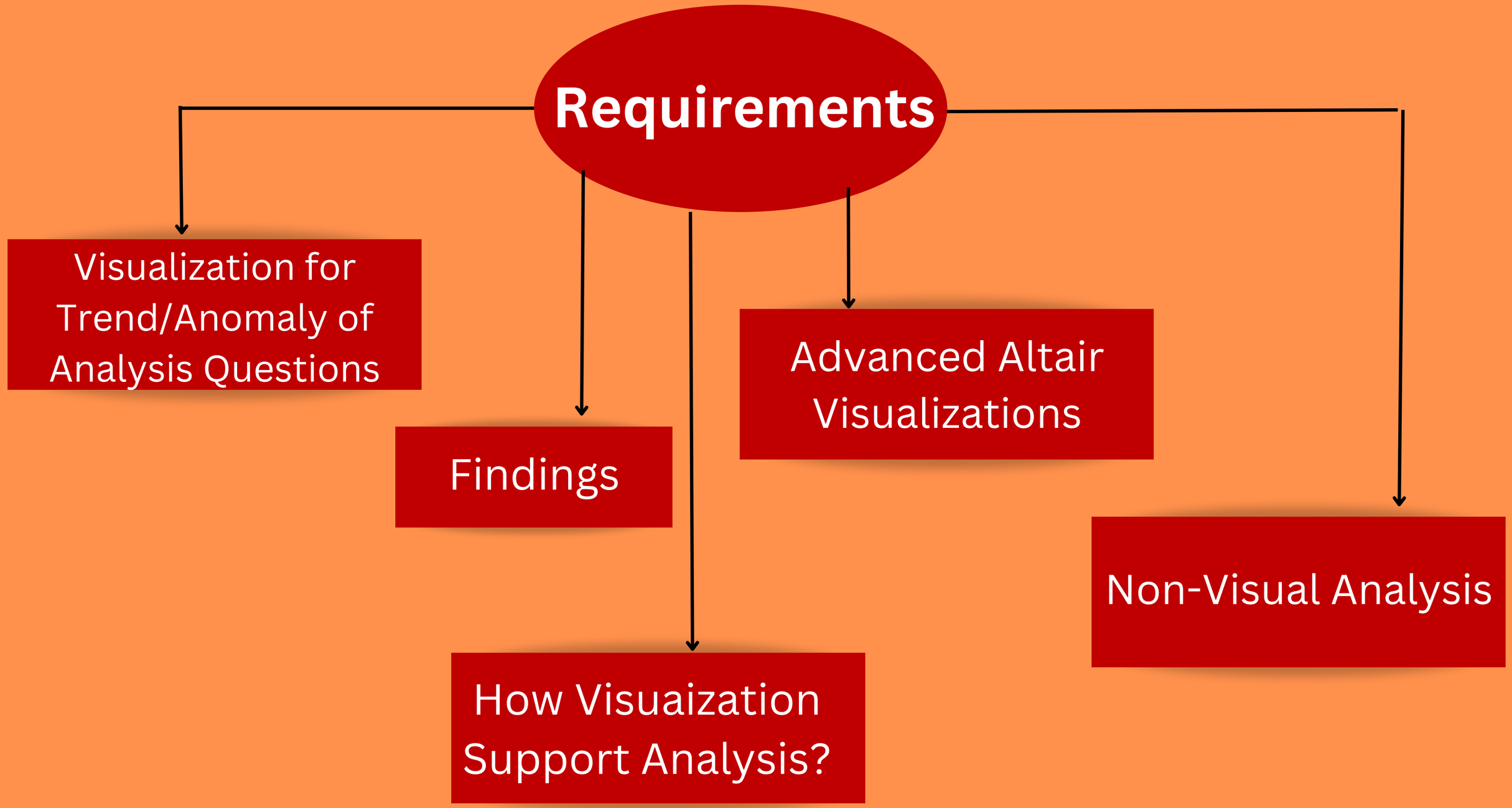
Unrealistic Values

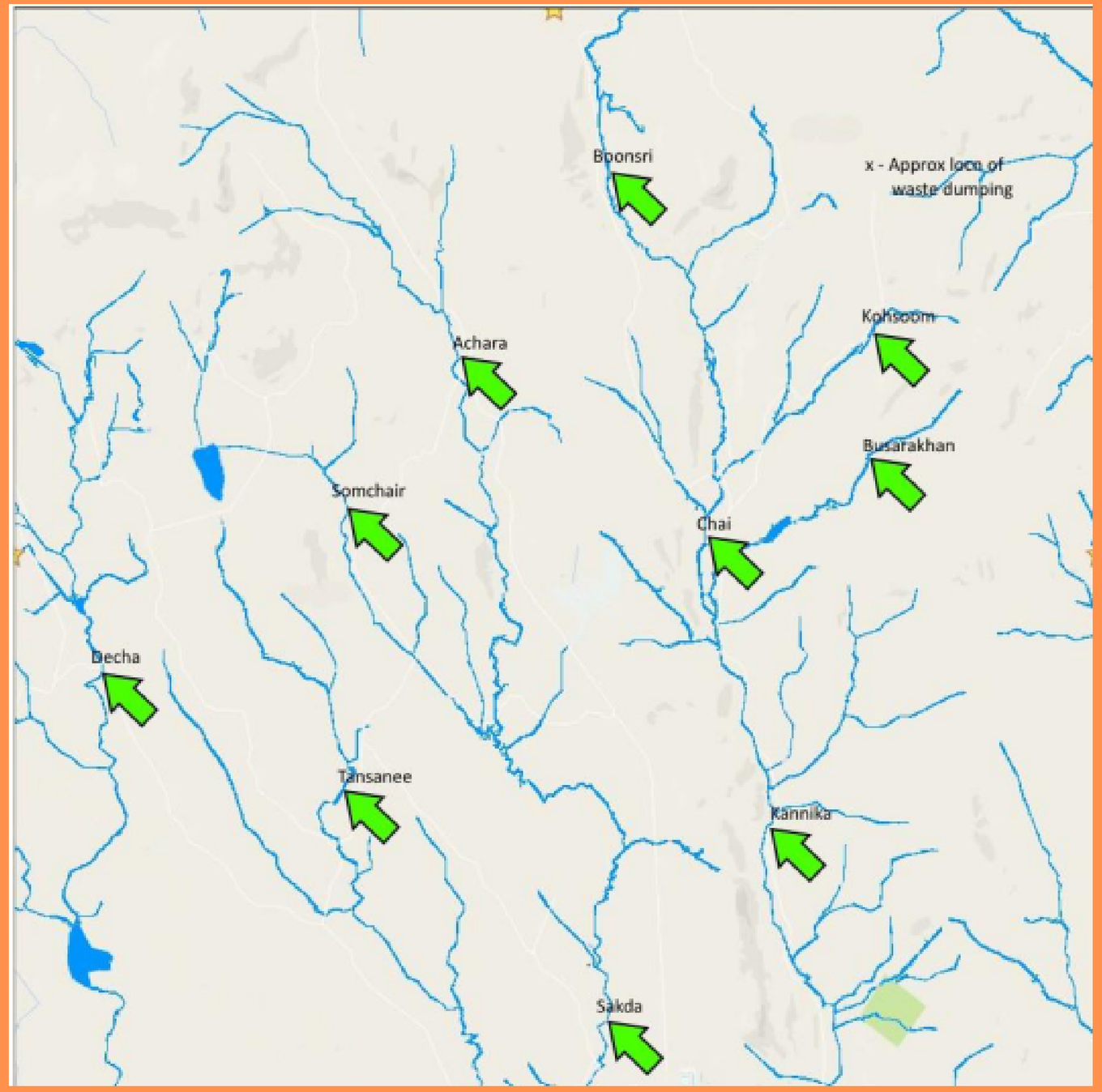
We can see that there are anomalies in the dataset, as we can see in the figure, value of Ammonium at Kannika location rised steepely.
Maximum no of anomalies can be found in Chai.

Altair Visualizations



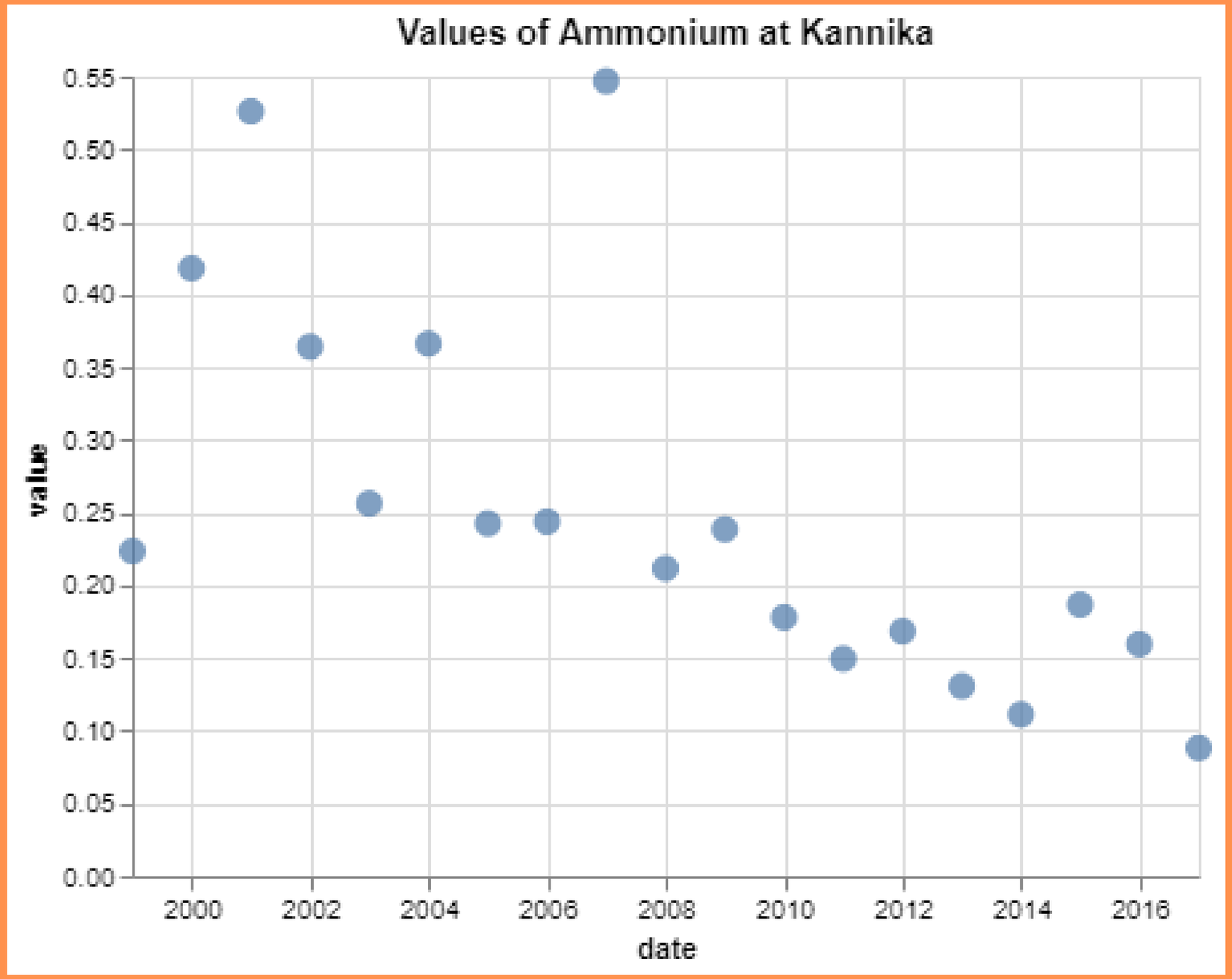



Requirements
Visualization for Trend/Anomaly of Analysis Questions
Findings
How Visuaization Support Analysis?
Advanced Altair Visualizations
Non-Visual Analysis



What are the Findings?

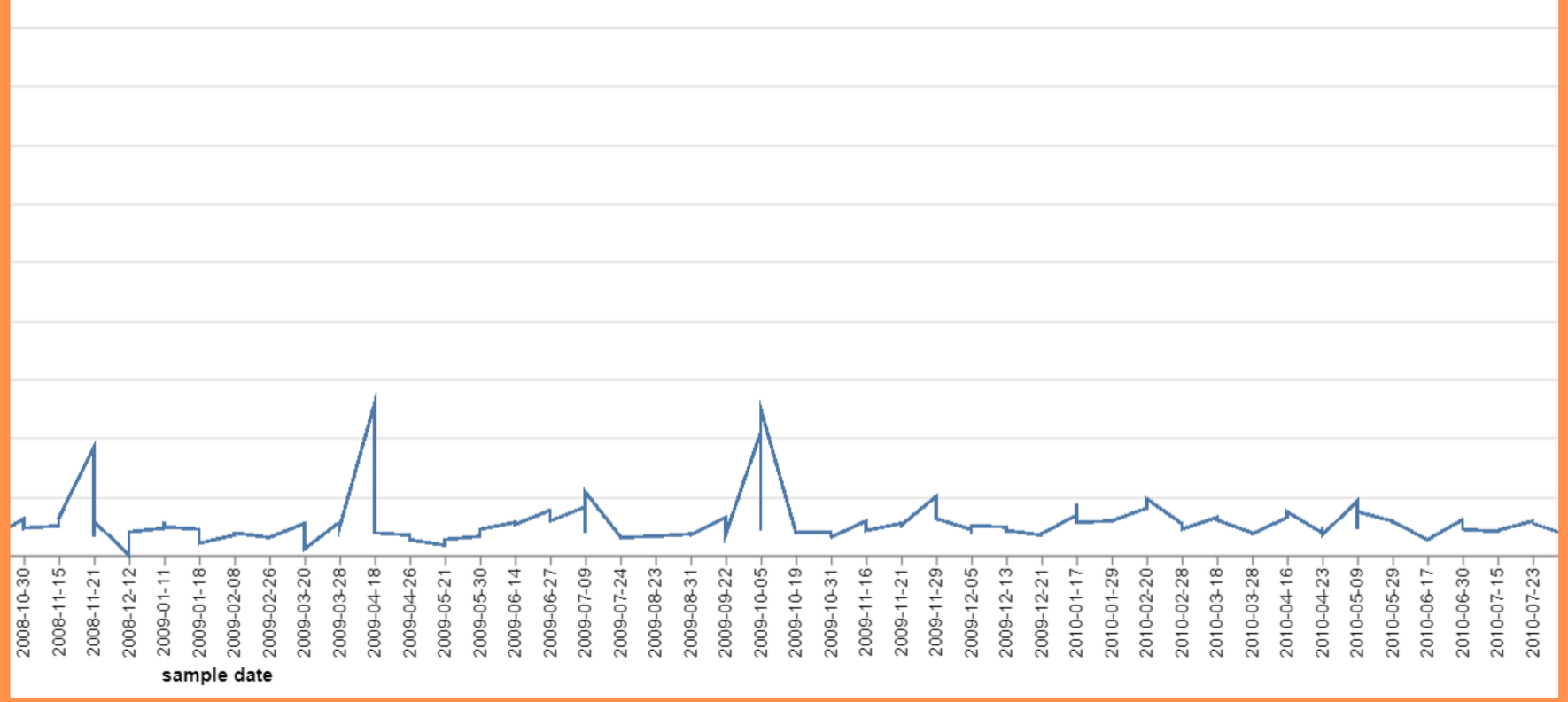
Findings are that the value of "measure" is decresing per year for every "location". Though there are few anomalies but still the values are decreasing.



Visualization for Trend/Anomaly of Analysis Questions

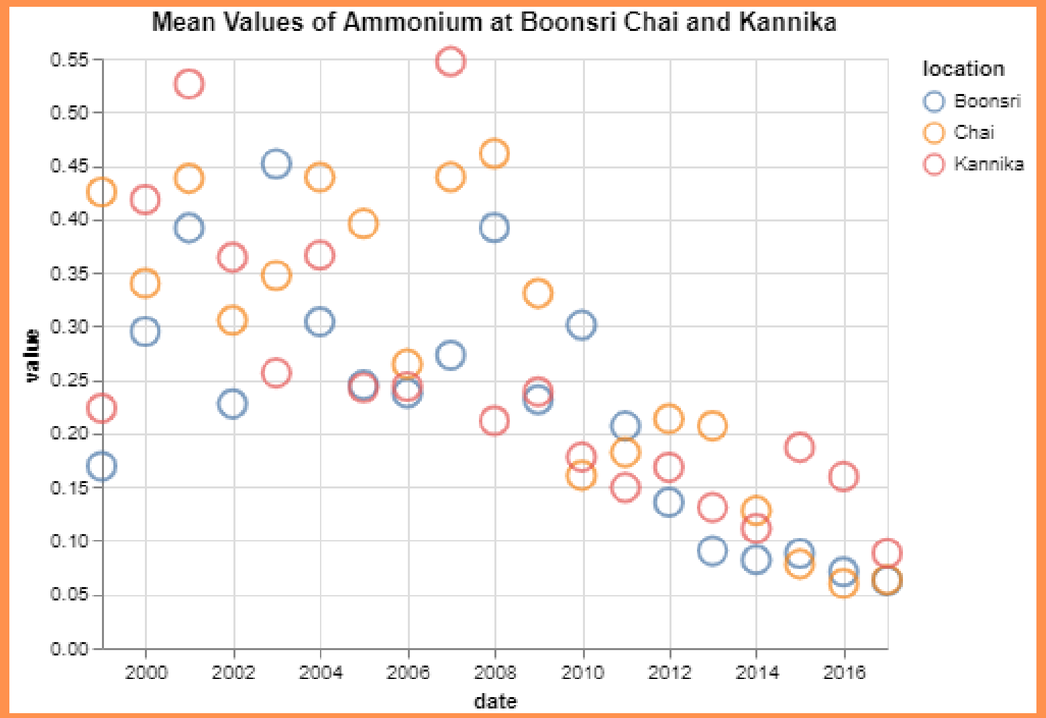
We can easily see the trend is decreasing values of Ammonium with the passage of time. There are few anomalies such as in year 2007 steep rise in value.

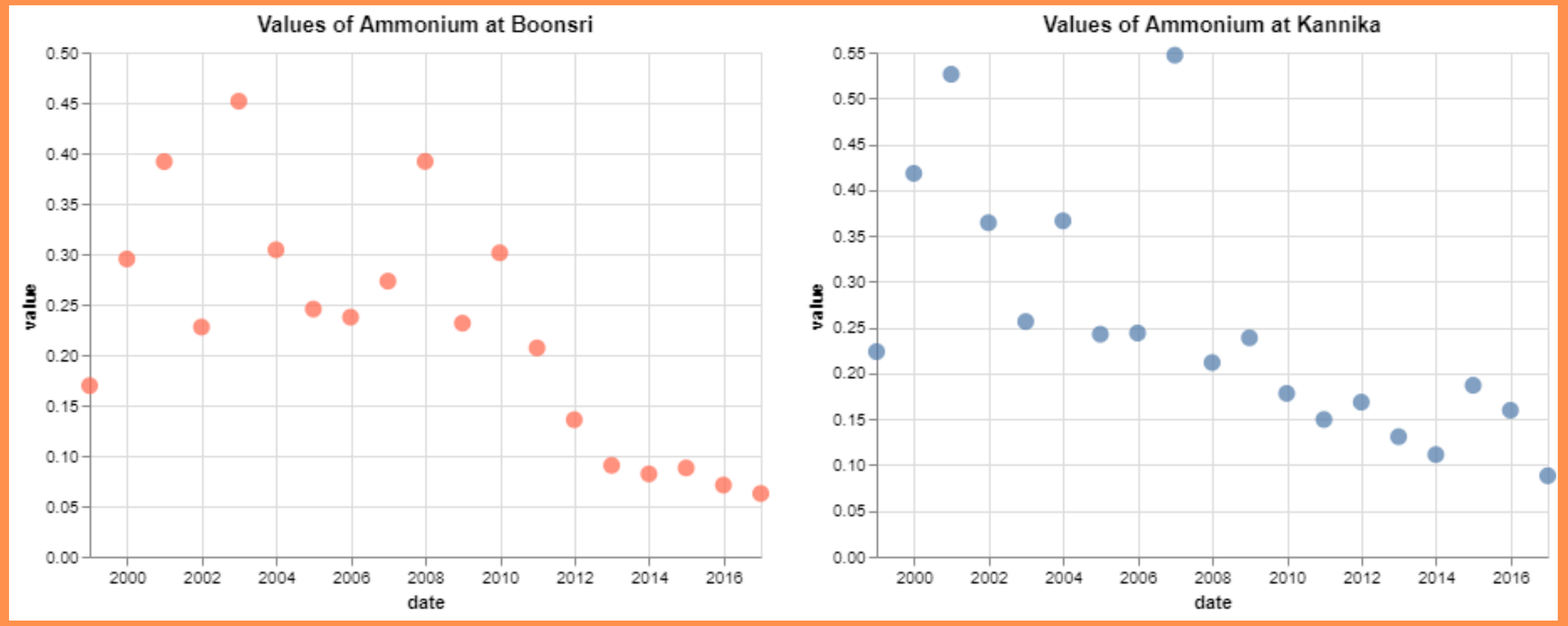
```
alt.Chart(data_sudden_change_Chai_Nitrites).mark_line().encode(
    x='sample date',
    y='value',
)
```



How Visualization Support Analysis?

With the increasing time, the value of measures found at locations such as Chai, Kannika, Boonsri etc. are decreasing and in the later 20's they become stable.





Advanced Altair Visualizations

Advanced techniques of Altair such as Multi Layer Visualization and Chart Concatenation etc. are used.

In [395]:
```
df=pd.read_csv('Boonsong Lekagul waterways readings (2).csv')
df.head(4)
```

Out[395]:

	id	value	location	sample date	measure
0	2221	2.00	Boonsri	11-Jan-98	Water temperature
1	2223	9.10	Boonsri	11-Jan-98	Dissolved oxygen
2	2227	0.33	Boonsri	11-Jan-98	Ammonium
3	2228	0.01	Boonsri	11-Jan-98	Nitrites

In [386]:
```
df.shape  # the Dataset have 136824 rows and 5 columns
```

Out[386]:
```
(136824, 5)
```

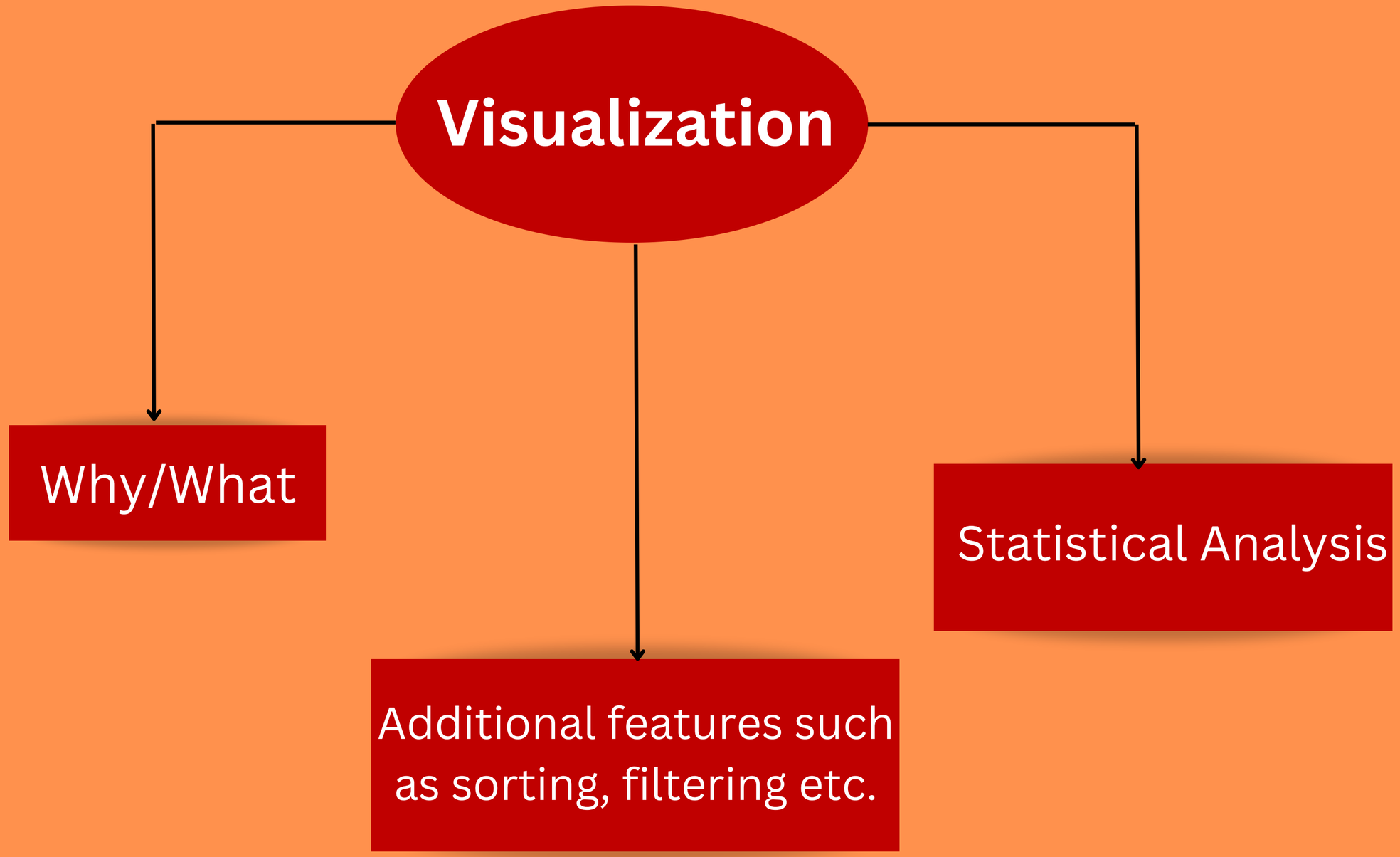
For Checking NULL values in the dataset

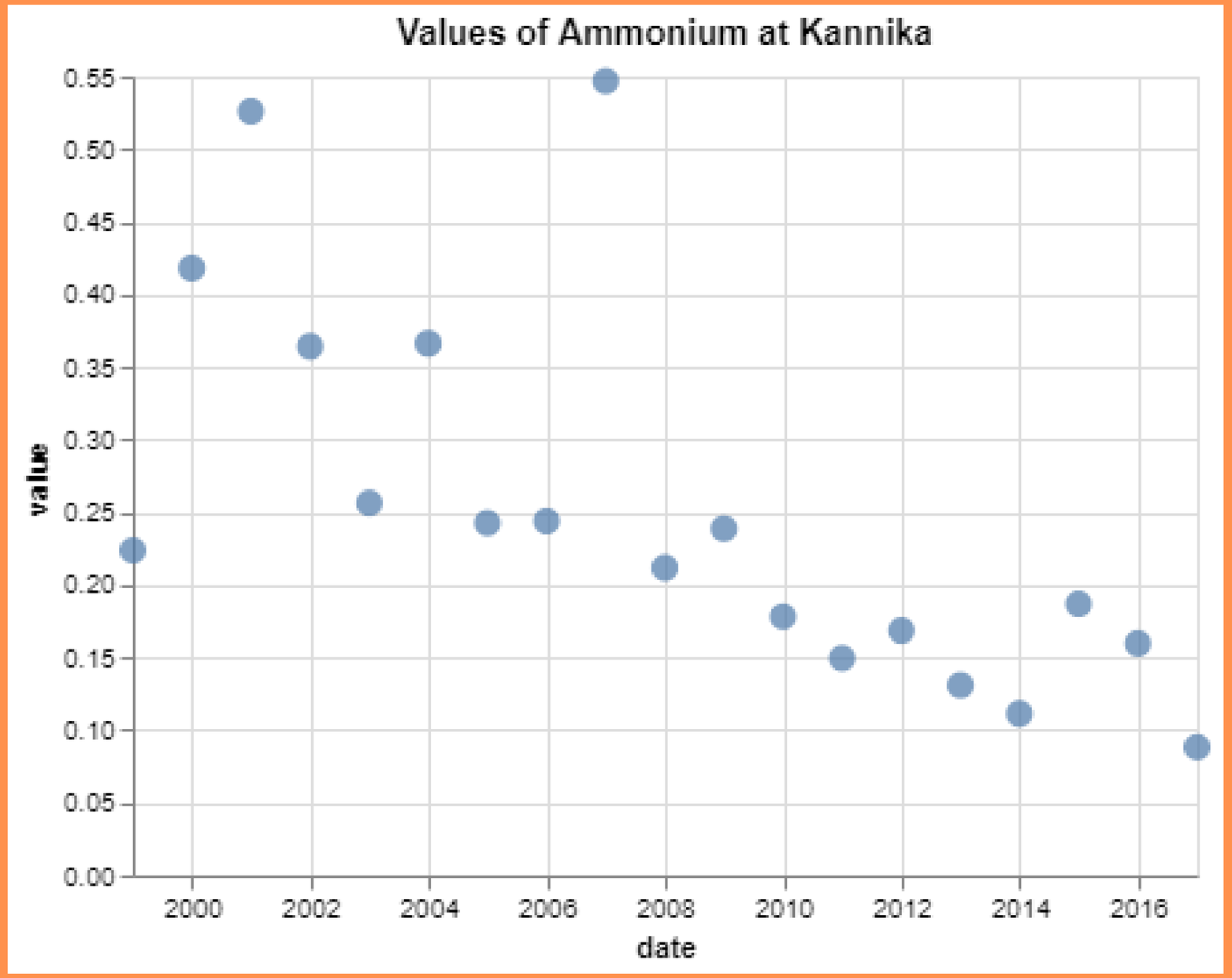
In [391]:
```
df.info() # This gives us all the information about the dataset's insights
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 136824 entries, 0 to 136823
Data columns (total 5 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   id           136824 non-null  int64
 1   value        136824 non-null  float64
 2   location     136824 non-null  object
 3   sample date  136824 non-null  object
 4   measure      136824 non-null  object
dtypes: float64(1), int64(1), object(3)
```

Non-Visual Analysis

Using various functions of pandas such as we can see the that how the value of vaiour measures are changing.


Visualization
Why/What
Additional features such as sorting, filtering etc.
Statistical Analysis



Why/What

These Charts drawn using Altair help us to visualize the data well, shows the trends and anomalies in values, thus helps in visualizing well. Choice of mark is point chart, even though in .ipynb file various other ways if visualization have also been used but preferable is point visualization.

A function returning a dataframe that takes help of dates for statistical analysis of data

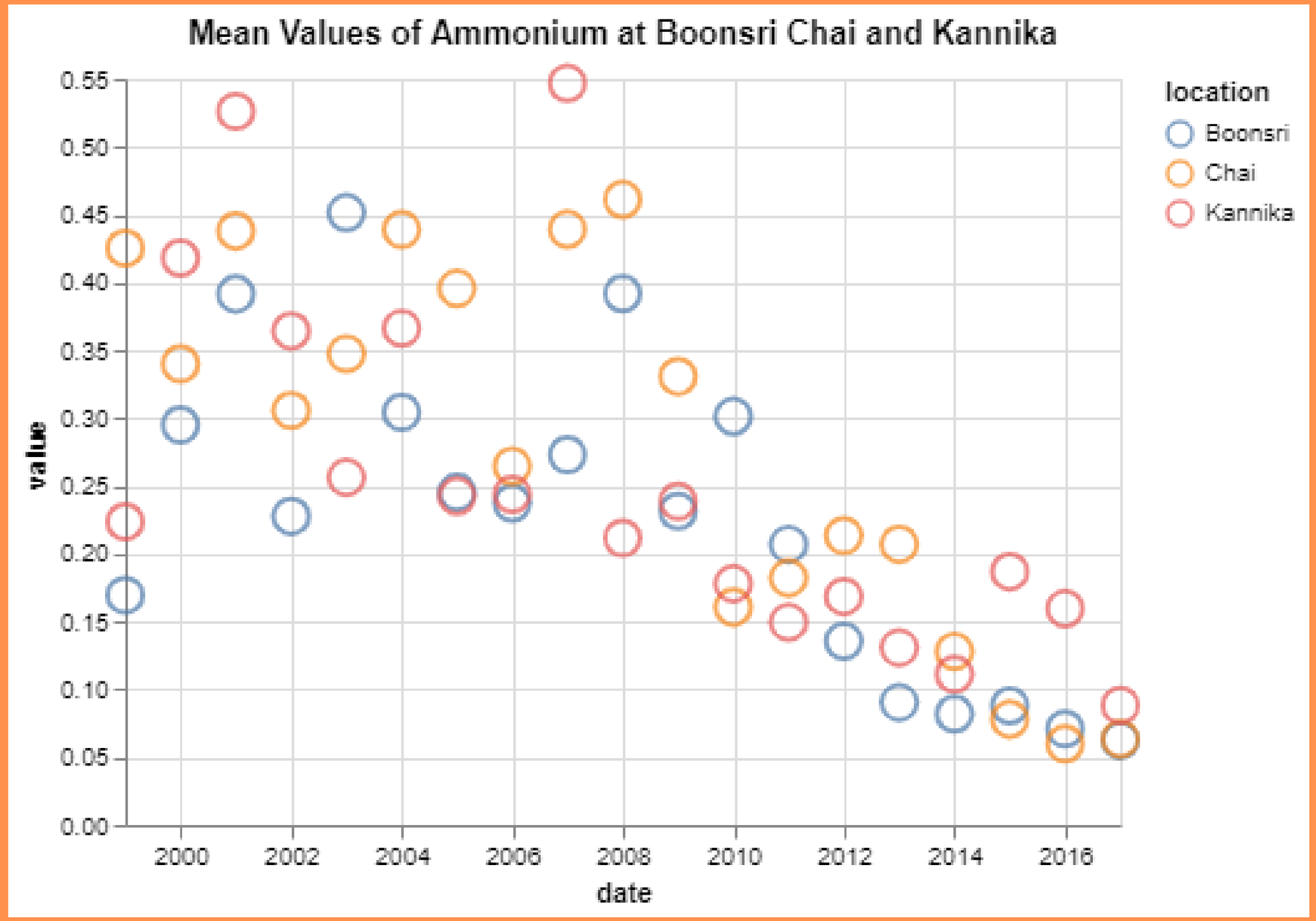
```
def groupbydatesandlocation(df,chemical,site):
    df_mean=df[(df['measure']==chemical)&(df['location']==site)]
    df_mean=df_mean.drop(['measure','id'],axis=1)
    df_mean['date']=pd.to_datetime(df_mean['sample date'])
    df_mean=df_mean.drop(['sample date'],axis=1)
    df_mean = df_mean.set_index('date')
    df_mean=df_mean.resample('A').mean()
    df_mean=df_mean.reset_index()
    return df_mean
```

```
df_Boonsri_ammonium=groupbydatesandlocation(df,'Ammonium','Boonsri')
df_Boonsri_ammonium['location']="Boonsri"
chart_boonsri_mean=alt.Chart(df_Boonsri_ammonium).mark_circle(size=100,
    color='tomato').encode(
    x='date',
    y='value'

).properties(
    title='Values of Ammonium at Boonsri'
)
chart_boonsri_mean
```

Additional features such as sorting, filtering etc.

Techniques such as Indexing of Dataframe according to date and even taking average value of 'measures' per year has been used. This included filtering, sorting and even statistics.



Statistical Analysis

Mean values of "measures" such as Nitrites, Ammonium and Calcium have been evaluated to get a better visual of how the trend has been.

Conclusion

After drawing various types of visualization using Altair such as Circle, Point and line we can evaluate the trend how value of "measures" such as Ammonium, Nitrites and Calcium changes for locations such as Chai, Boonsari, Kannika etc.

The value kept on decreasing and became very less after in 20's. Statistical Analysis, filtering and indexing etc, various techniques were used.