```
import pandas as pd
import numpy as np
import altair as alt
```

```
df=pd.read_csv('Boonsong Lekagul waterways readings (2).csv')
df.head(4)
```

	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

```
df.shape  # the Dataset have 136824 rows and 5 columns
```

```
(136824, 5)
```

```
df.size
```

```
684120
```

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

```
df.columns
```

```
Index(['id', 'value', 'location', 'sample date', 'measure'], dtype='object')
```

```
df.dtypes
```

```
id              int64
value         float64
location       object
sample date    object
measure        object
dtype: object
```

We can see that the column 'sample date' is of type "object" there we have to convert it into "date time" format

```
df['sample date'] = df['sample date'].apply(lambda x: pd.Timestamp(x).strftime('%Y-%m-%d'))   # converting 'sample date' col
```

```
df.head() #df['sample date']=pd.to_datetime(df['sample date'])
```

	id	value	location	sample date	measure
0	2221	2.00	Boonsri	1998-01-11	Water temperature
1	2223	9.10	Boonsri	1998-01-11	Dissolved oxygen
2	2227	0.33	Boonsri	1998-01-11	Ammonium
3	2228	0.01	Boonsri	1998-01-11	Nitrites
4	2229	1.47	Boonsri	1998-01-11	Nitrates

```
df['location'].unique()  # collecting all the 'locations' in the dataset
```

```
array(['Boonsri', 'Kannika', 'Chai', 'Kohsoom', 'Somchair', 'Sakda',
       'Busarakhan', 'Tansanee', 'Achara', 'Decha'], dtype=object)
```

```
sites=['Boonsri', 'Kannika', 'Chai', 'Kohsoom', 'Somchair', 'Sakda','Busarakhan', 'Tansanee', 'Achara', 'Decha']
```

```
df['measure'].unique()  # all the types of "measure" in dataset
```

```
array(['Water temperature', 'Dissolved oxygen', 'Ammonium', 'Nitrites',
       'Nitrates', 'Orthophosphate-phosphorus', 'Total phosphorus',
       'Sodium', 'Potassium', 'Calcium', 'Magnesium', 'Chlorides',
       'Sulphates', 'Iron', 'Manganese', 'Zinc', 'Copper', 'Chromium',
       'Lead', 'Cadmium', 'Mercury', 'Nickel', 'Arsenic',
       'Biochemical Oxygen', 'Chemical Oxygen Demand (Cr)',
       'Chemical Oxygen Demand (Mn)', 'AOX', 'Atrazine', 'Cesium',
       'Macrozoobenthos', 'Total coliforms', 'Fecal coliforms', 'p,p-DDT',
       'gamma-Hexachlorocyclohexane', 'Bicarbonates',
       'Anionic active surfactants', 'Total extractable matter',
       'Fecal streptococci ', 'Petroleum hydrocarbons', 'PAHs',
       'Benzo(a)pyrene', 'Benzo(g,h,i)perylene', 'Benzo(b)fluoranthene',
       'Benzo(k)fluoranthene', 'Fluoranthene', 'Indeno(1,2,3-c,d)pyrene',
       'PCB 28', 'PCB 52', 'PCB 101', 'PCB 138', 'PCB 153', 'PCB 180',
       'Silica (SiO2)', 'Oxygen saturation', 'Total hardness',
       'Total dissolved salts', 'Heptachloroepoxide', 'Heptachlor',
       'Endosulfan (alpha)', 'Endosulfan (beta)', 'p,p-DDD', 'p,p-DDE',
       'alpha-Hexachlorocyclohexane', 'beta-Hexaxchlorocyclohexane',
       'Aldrin', 'Dieldrin', 'Endrin', 'Methoxychlor', 'Simazine',
       'Metolachlor', 'Alachlor', 'Carbonates', 'Total nitrogen',
       'Tetrachloromethane', 'Barium', 'Cyanides', 'Sulfides', 'Selenium',
       'Total organic carbon', '1,2,4-Trichlorobenzene',
       '1,2,3-Trichlorobenzene', 'Pentachlorobenzene', 'Acenaphthene',
       'Acenaphthylene', 'Anthracene', 'Benzo(a)anthracene', 'Chrysene',
       'Naphthalene', 'Phenanthrene', 'Pyrene', 'Hexachlorobenzene',
       'Isodrin', 'Aluminium', 'Dissolved organic carbon',
       'Dissolved silicates', 'Organic nitrogen', 'Fluorene', 'PCB 118',
       'Trifluralin', 'Inorganic nitrogen', 'Berilium', 'Boron',
       'AGOC-3A', 'Methylosmoline', 'Chlorodinine',
       'Total dissolved phosphorus'], dtype=object)
```

We will be dealing with 'sites' in "locations" list and 'measure' in "measures" list

```
locations=['Boonsri','Kannika','Chai']
```

```
measures=['Ammonium','Calcium','Nitrites']
```

Plotting and Visualization of the locations on basis of their 'measure' values

1 For Boonsri and Ammonium

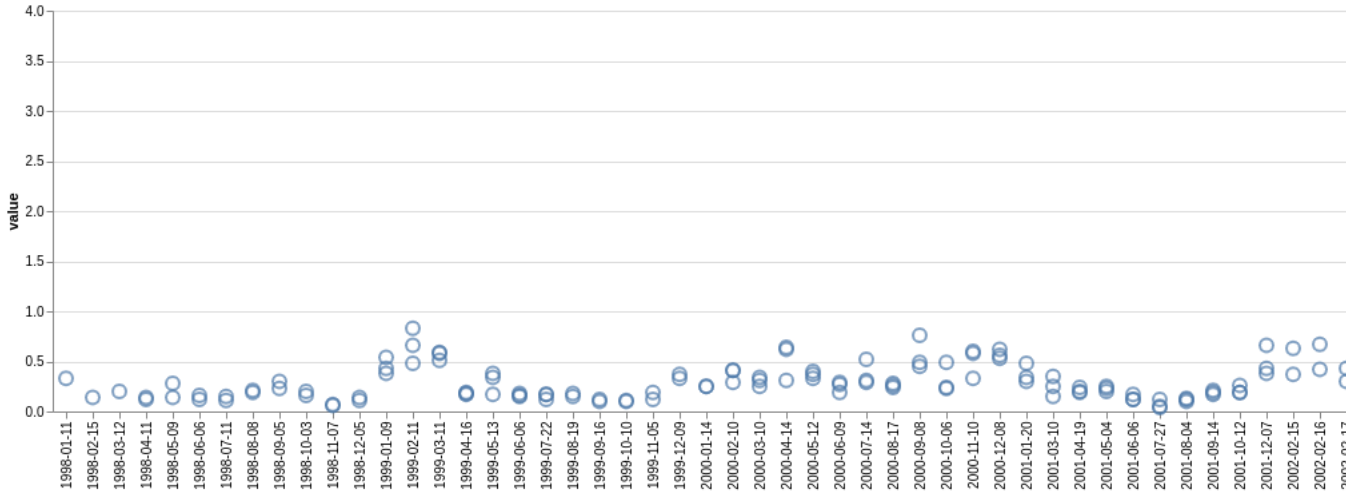
```
df_sudden_change_Boonsri_Ammonium=df[(df['location']=='Boonsri') & (df['measure']=='Ammonium')]
df_sudden_change_Boonsri_Ammonium.head()         # preprocessing of dataset i.e. filtration part
```

	id	value	location	sample date	measure
2	2227	0.33	Boonsri	1998-01-11	Ammonium
226	2267	0.14	Boonsri	1998-02-15	Ammonium
450	2307	0.20	Boonsri	1998-03-12	Ammonium
674	2347	0.12	Boonsri	1998-04-11	Ammonium
708	2710	0.14	Boonsri	1998-04-11	Ammonium

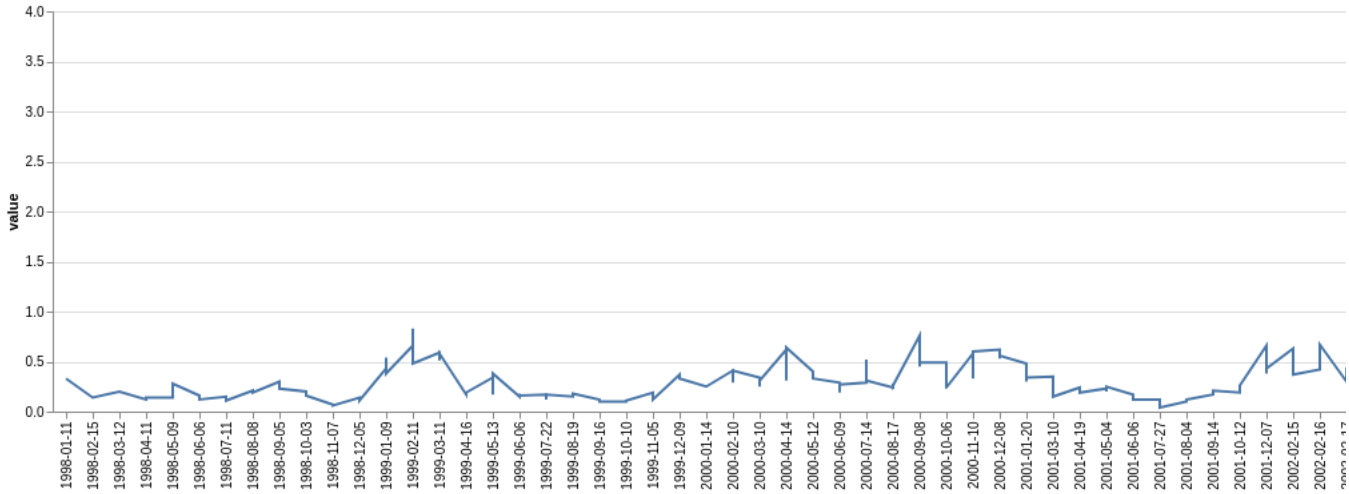
```
data_sudden_change_Boonsri_Ammonium=df_sudden_change_Boonsri_Ammonium.drop(['id','location','measure'],axis='columns')
data_sudden_change_Boonsri_Ammonium.head()       # preprocessing of dataset
```

	value	sample date
2	0.33	1998-01-11
226	0.14	1998-02-15
450	0.20	1998-03-12
674	0.12	1998-04-11
708	0.14	1998-04-11

```
alt.Chart(data_sudden_change_Boonsri_Ammonium).mark_point(size=100).encode(
    x='sample date',
    y='value',
)
```




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



In the beginning years there is no trend in Ammonium at Boonsri but after 2011 it's concentration declined There are very few anomalies such as- In the year 19-04-2007 the value of ammonium is very high or like in the year 31-10-2014 ammonium abruptly raised

2 For Kannika an Ammonium

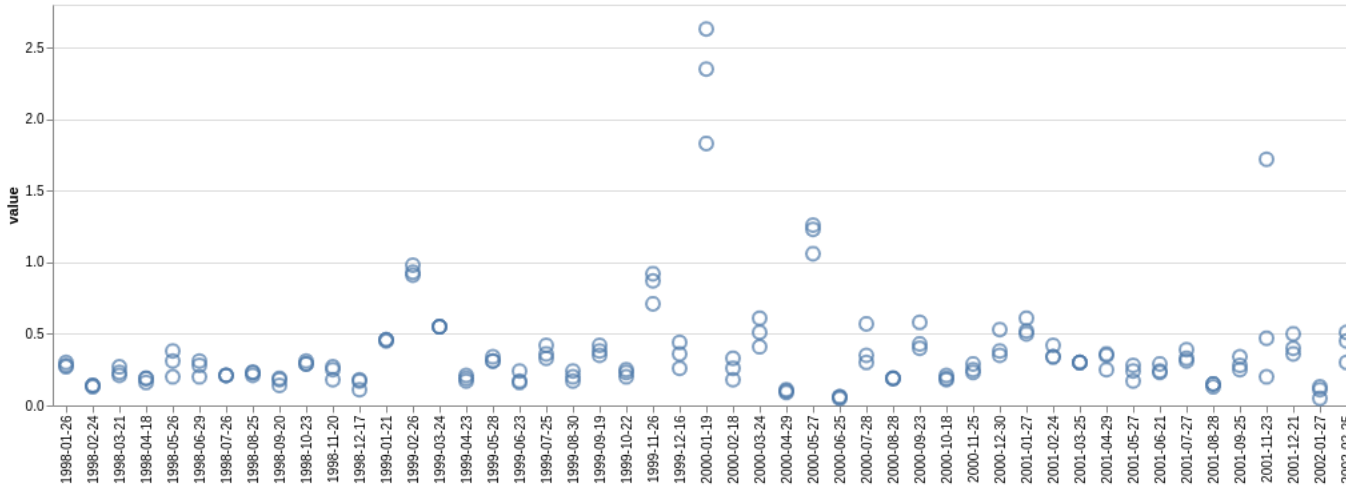
```
df_sudden_change_Kannika_Ammonium=df[(df['location']=='Kannika') & (df['measure']=='Ammonium')]
df_sudden_change_Kannika_Ammonium.head()            # preprocessing of dataset
```

	id	value	location	sample date	measure
34	8297	0.30	Kannika	1998-01-26	Ammonium
66	8778	0.28	Kannika	1998-01-26	Ammonium
98	9261	0.27	Kannika	1998-01-26	Ammonium
354	8337	0.14	Kannika	1998-02-24	Ammonium
386	8818	0.13	Kannika	1998-02-24	Ammonium

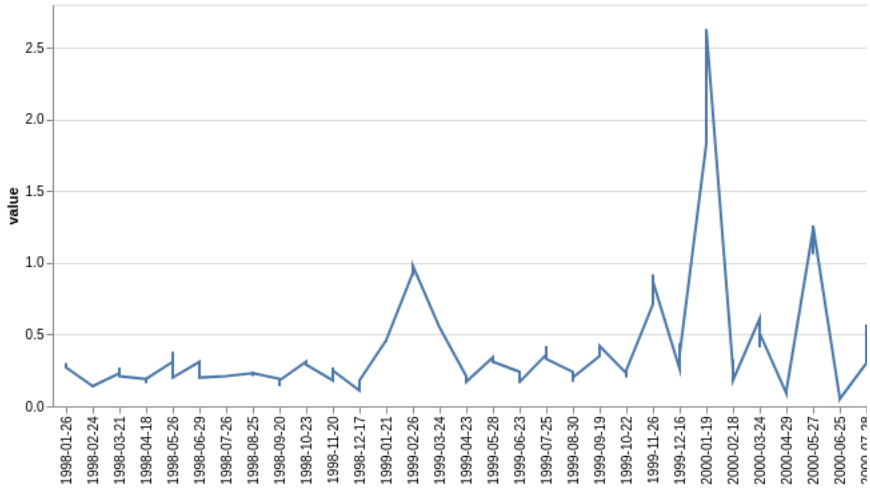
```
data_sudden_change_Kannika_Ammonium=df_sudden_change_Kannika_Ammonium.drop(['id','location','measure'],axis='columns')
data_sudden_change_Kannika_Ammonium.head()         # preprocessing of dataset
```

	value	sample date
34	0.30	1998-01-26
66	0.28	1998-01-26
98	0.27	1998-01-26
354	0.14	1998-02-24
386	0.13	1998-02-24

```
alt.Chart(data_sudden_change_Kannika_Ammonium).mark_point(size=100).encode(
    x='sample date',
    y='value',
)
```




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



We cannot find any pattern in the early years of Ammonium at Kanika and there are many anomalies such as in 2000-01-19 we saw, and since year 2009 it's value kept on decreasing

3 For Chai and Ammonium

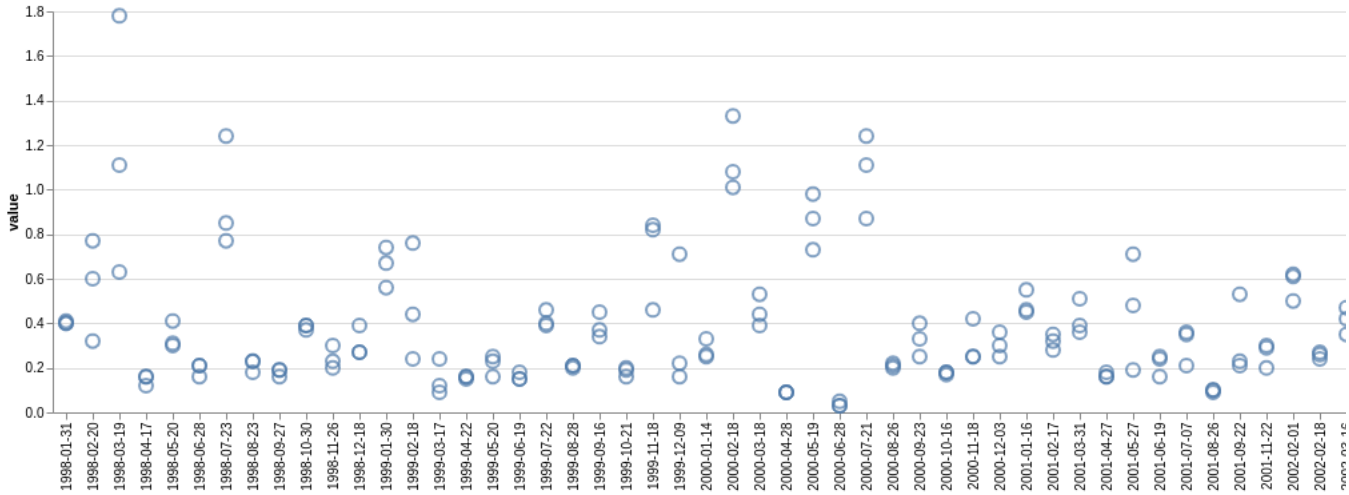
```
df_sudden_change_Chai_Ammonium=df[(df['location']=='Chai') & (df['measure']=='Ammonium')]
df_sudden_change_Chai_Ammonium.head()            # preprocessing of dataset
```

	id	value	location	sample date	measure
130	5054	0.40	Chai	1998-01-31	Ammonium
162	5538	0.40	Chai	1998-01-31	Ammonium
194	6016	0.41	Chai	1998-01-31	Ammonium

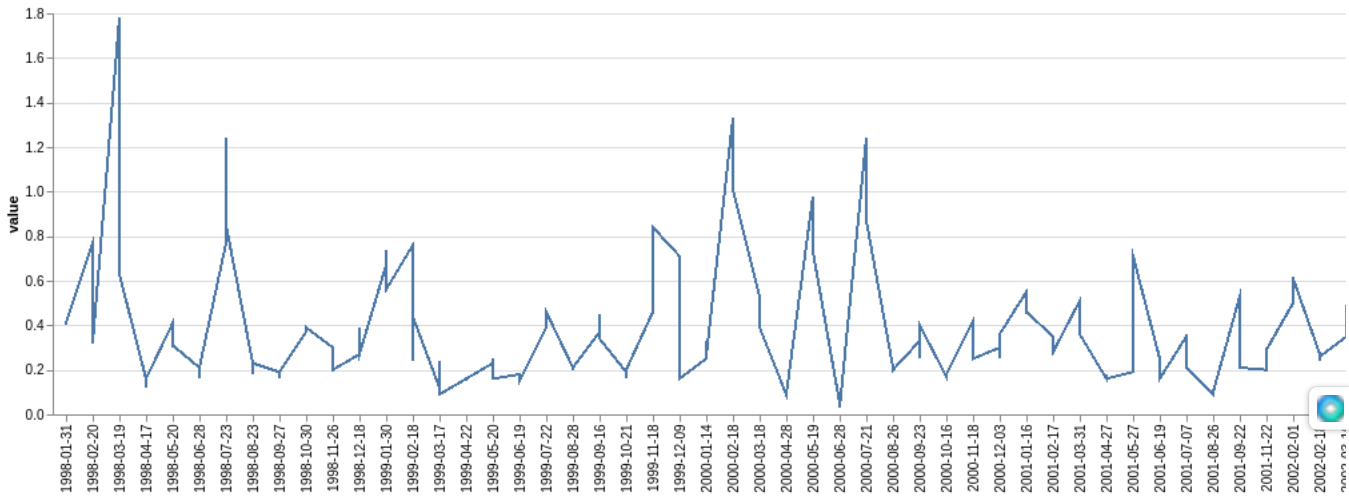
```
data_sudden_change_Chai_Ammonium=df_sudden_change_Chai_Ammonium.drop(['id','location','measure'],axis='columns')
data_sudden_change_Chai_Ammonium.head()           # preprocessing of dataset
```

	value	sample date
130	0.40	1998-01-31
162	0.40	1998-01-31
194	0.41	1998-01-31
258	0.77	1998-02-20
290	0.60	1998-02-20

```
alt.Chart(data_sudden_change_Chai_Ammonium).mark_point(size=100).encode(
    x='sample date',
    y='value',
)
```




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



In the beginning years there is no trend in Ammonium at Chai but after 2011 it's concentration declined There are very few anomalies such as-
In the year 19-04-2007 the value of Ammonium is very high or like in the year 31-10-2014 Ammonium abruptly raised

Chart concatenation of Ammonium at sites 'Chai' and 'Boonsri'

```
chart1=alt.Chart(data_sudden_change_Chai_Ammonium).mark_point(color="tomato",size=100).encode(x="sample date",y="value")
chart2=alt.Chart(data_sudden_change_Boonsri_Ammonium).mark_point(size=100).encode(x="sample date",y="value")
chart1 & chart2
```

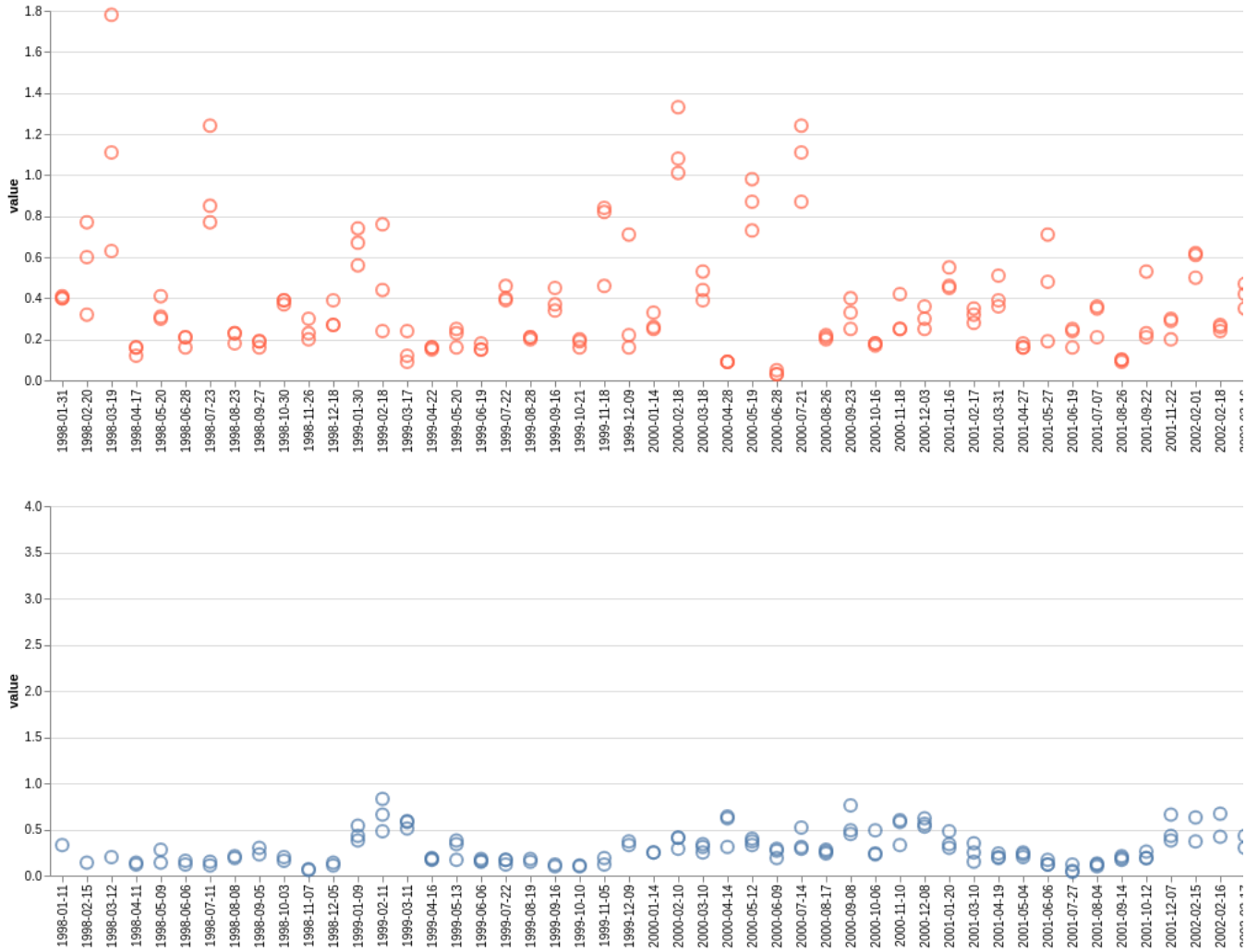


Chart concatenation of Ammonium at sites 'Kanika' and 'Boonsri'

```
chart1=alt.Chart(data_sudden_change_Kannika_Ammonium).mark_point(color="tomato",size=100).encode(x="sample date",y="value")
chart2=alt.Chart(data_sudden_change_Boonsri_Ammonium).mark_point(size=100).encode(x="sample date",y="value")
chart1 & chart2
```

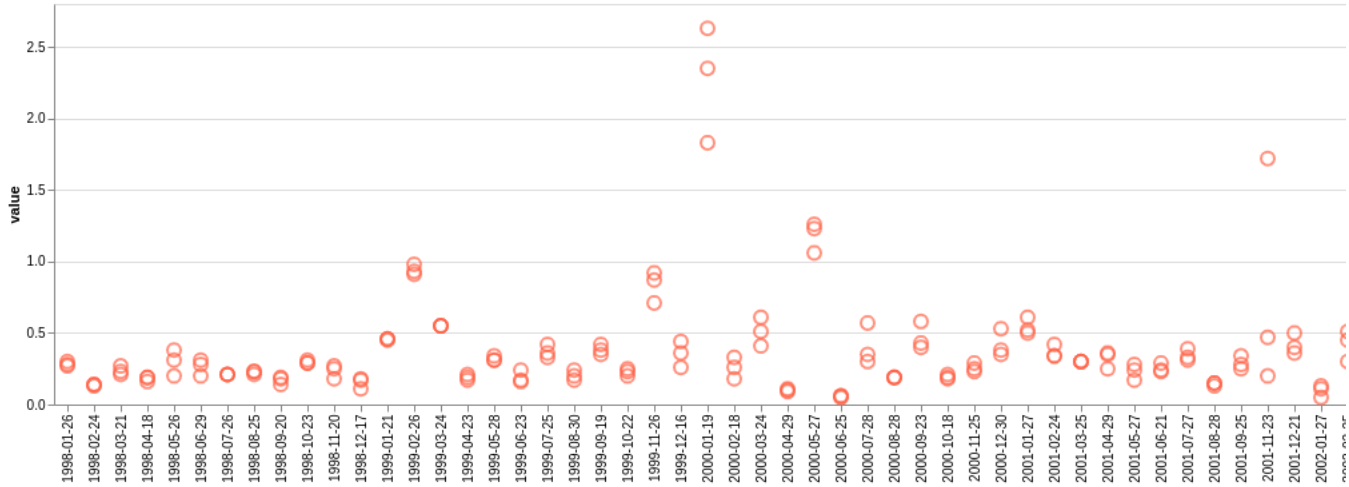
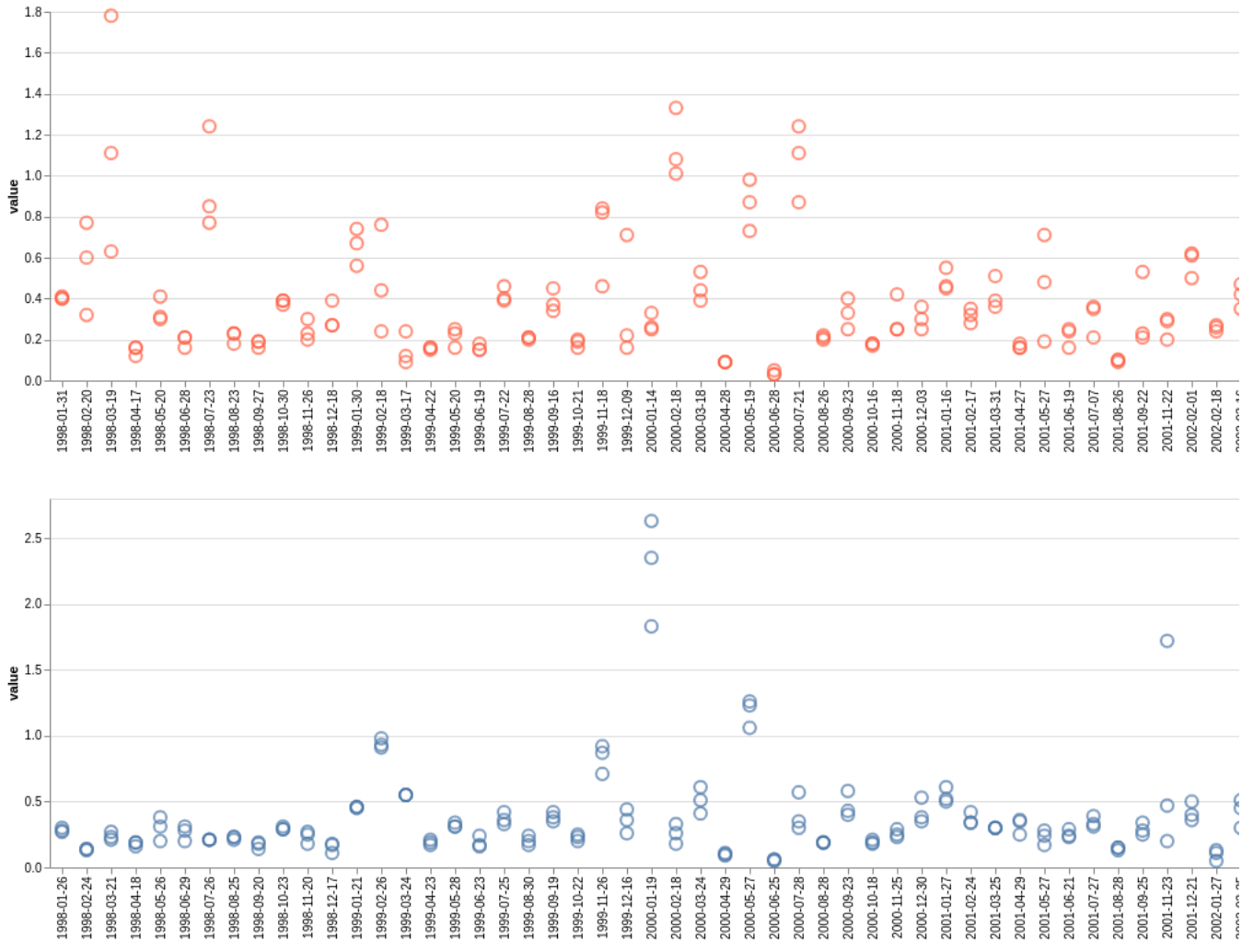


Chart Concatenation of "Ammonium" at sites 'Chai' and 'Kannika'

```
chart1=alt.Chart(data_sudden_change_Chai_Ammonium).mark_point(color="tomato",size=100).encode(x="sample date",y="value")
chart2=alt.Chart(data_sudden_change_Kannika_Ammonium).mark_point(size=100).encode(x="sample date",y="value")
chart1 & chart2
```



1 For Boonsri Calcium

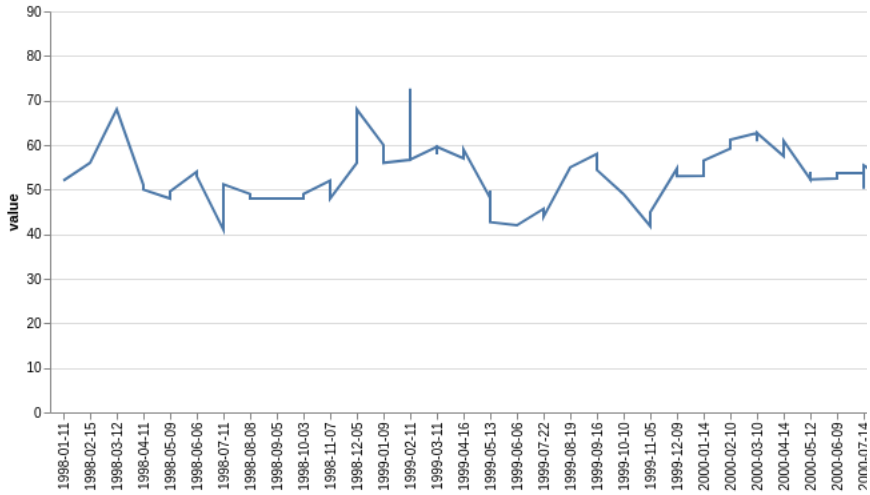
```
df_sudden_change_Boonsri_Calcium=df[(df['location']=='Boonsri') & (df['measure']=='Calcium')]
df_sudden_change_Boonsri_Calcium.head()
```

	id	value	location	sample date	measure
9	2234	52.0	Boonsri	1998-01-11	Calcium
233	2274	56.0	Boonsri	1998-02-15	Calcium
457	2314	68.0	Boonsri	1998-03-12	Calcium

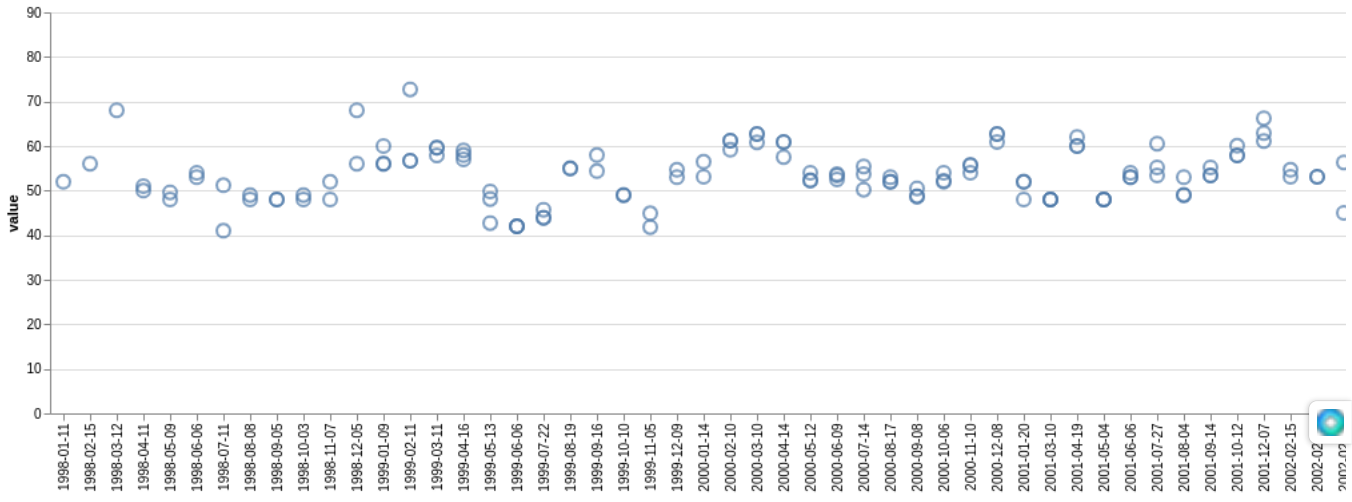
```
data_sudden_change_Boonsri_Calcium=df_sudden_change_Boonsri_Calcium.drop(['id','location','measure'],axis='columns')
data_sudden_change_Boonsri_Calcium.head()
```

	value	sample date
9	52.0	1998-01-11
233	56.0	1998-02-15
457	68.0	1998-03-12
681	51.0	1998-04-11
715	50.0	1998-04-11

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




```
alt.Chart(data_sudden_change_Boonsri_Calcium).mark_point(size=100).encode(
    x='sample date',
    y='value',
)
```



There is no trend or pattern in the values of Calcium at Boonsri and we can also see that value of calcium is varying a lot. There are very few anomalies in the plotting.

2 For Chai and Calcium

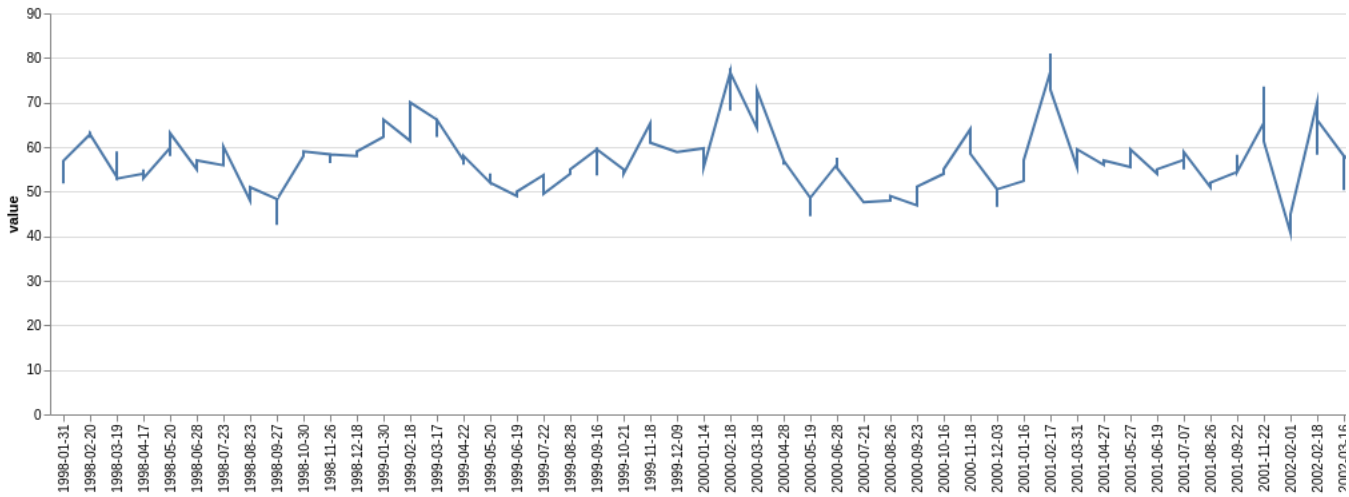
```
df_sudden_change_Chai_Calcium=df[(df['location']=='Chai') & (df['measure']=='Calcium')]
df_sudden_change_Chai_Calcium.head()
```

	id	value	location	sample date	measure
137	5061	52.60	Chai	1998-01-31	Calcium
169	5545	51.80	Chai	1998-01-31	Calcium
201	6023	56.90	Chai	1998-01-31	Calcium
265	5100	62.87	Chai	1998-02-20	Calcium
297	5584	62.02	Chai	1998-02-20	Calcium

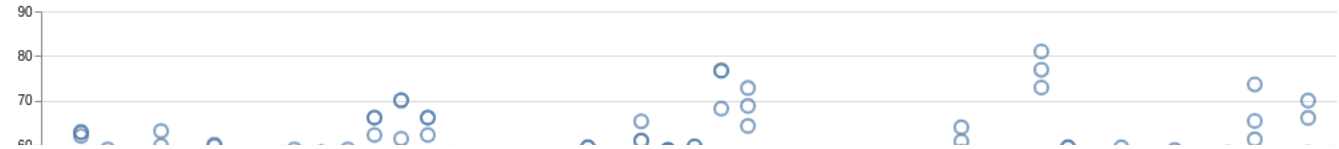
```
data_sudden_change_Chai_Calcium=df_sudden_change_Chai_Calcium.drop(['id','location','measure'],axis='columns')
data_sudden_change_Chai_Calcium.head()
```

	value	sample date
137	52.60	1998-01-31
169	51.80	1998-01-31
201	56.90	1998-01-31
265	62.87	1998-02-20
297	62.02	1998-02-20

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




```
alt.Chart(data_sudden_change_Chai_Calcium).mark_point(size=100).encode(
    x='sample date',
    y='value',
)
```



There is no trend or pattern in the values of Calcium at Chai and we can also see that value of calcium is varying a lot. There are a lot of anomalies in the plotting such as in year 22-05-11 the amount of calcium abruptly fell down.

3 For Kannika and Calcium

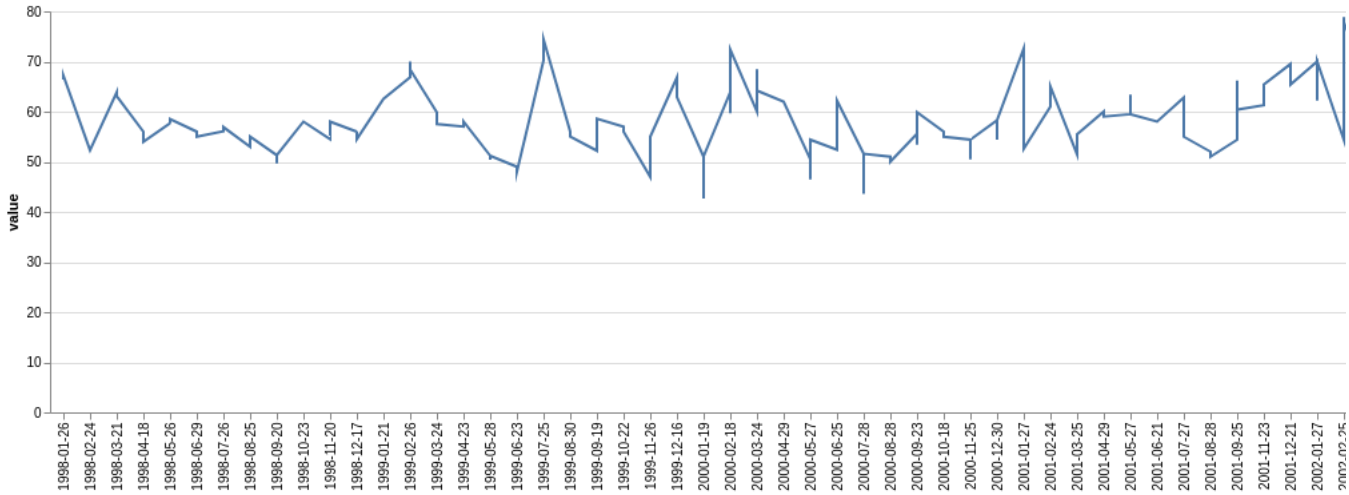

```
df_sudden_change_Kannika_Calcium=df[(df['location']=='Kannika') & (df['measure']=='Calcium')]
df_sudden_change_Kannika_Calcium.head()
```

	id	value	location	sample date	measure
41	8304	67.19	Kannika	1998-01-26	Calcium
73	8785	66.40	Kannika	1998-01-26	Calcium
105	9268	67.20	Kannika	1998-01-26	Calcium
361	8344	52.30	Kannika	1998-02-24	Calcium
393	8825	52.30	Kannika	1998-02-24	Calcium

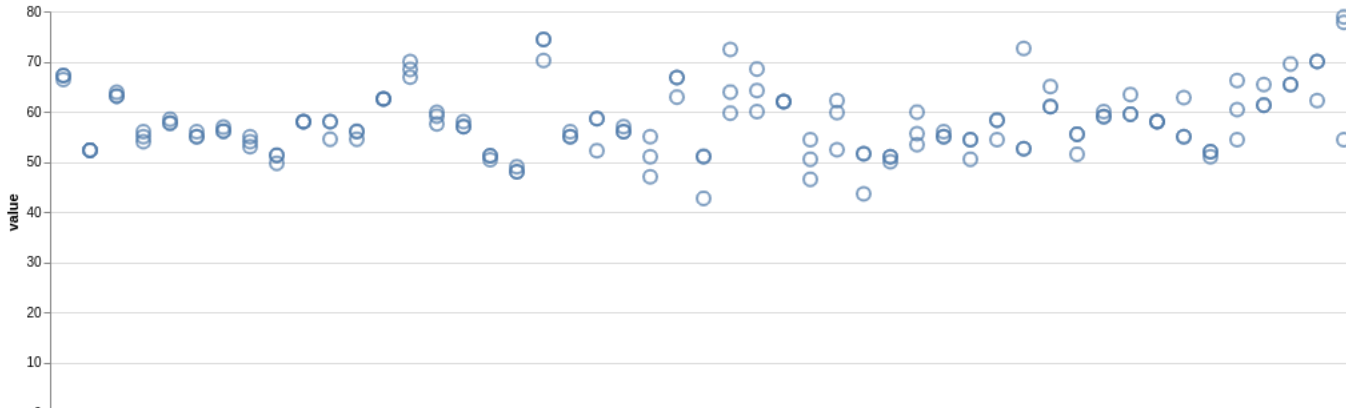
```
data_sudden_change_Kannika_Calcium=df_sudden_change_Kannika_Calcium.drop(['id','location','measure'],axis='columns')
data_sudden_change_Kannika_Calcium.head()
```

	value	sample date
41	67.19	1998-01-26
73	66.40	1998-01-26
105	67.20	1998-01-26
361	52.30	1998-02-24
393	52.30	1998-02-24

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




```
alt.Chart(data_sudden_change_Kannika_Calcium).mark_point(size=100).encode(
    x='sample date',
    y='value',
)
```



There is no trend or pattern in the values of Calcium at Kannika and we can also see that value of calcium is varying a lot. There are a lot of anomalies in the plotting such as in year 19-08-2004 the amount of calcium abruptly fell down.

Chart concatenation of Kannika and Chai for Calcium

```
chart1=alt.Chart(data_sudden_change_Kannika_Calcium).mark_point(color="tomato",size=100).encode(x="sample date",y="value")
chart2=alt.Chart(data_sudden_change_Chai_Calcium).mark_point(size=100).encode(x="sample date",y="value")
chart1 & chart2
```

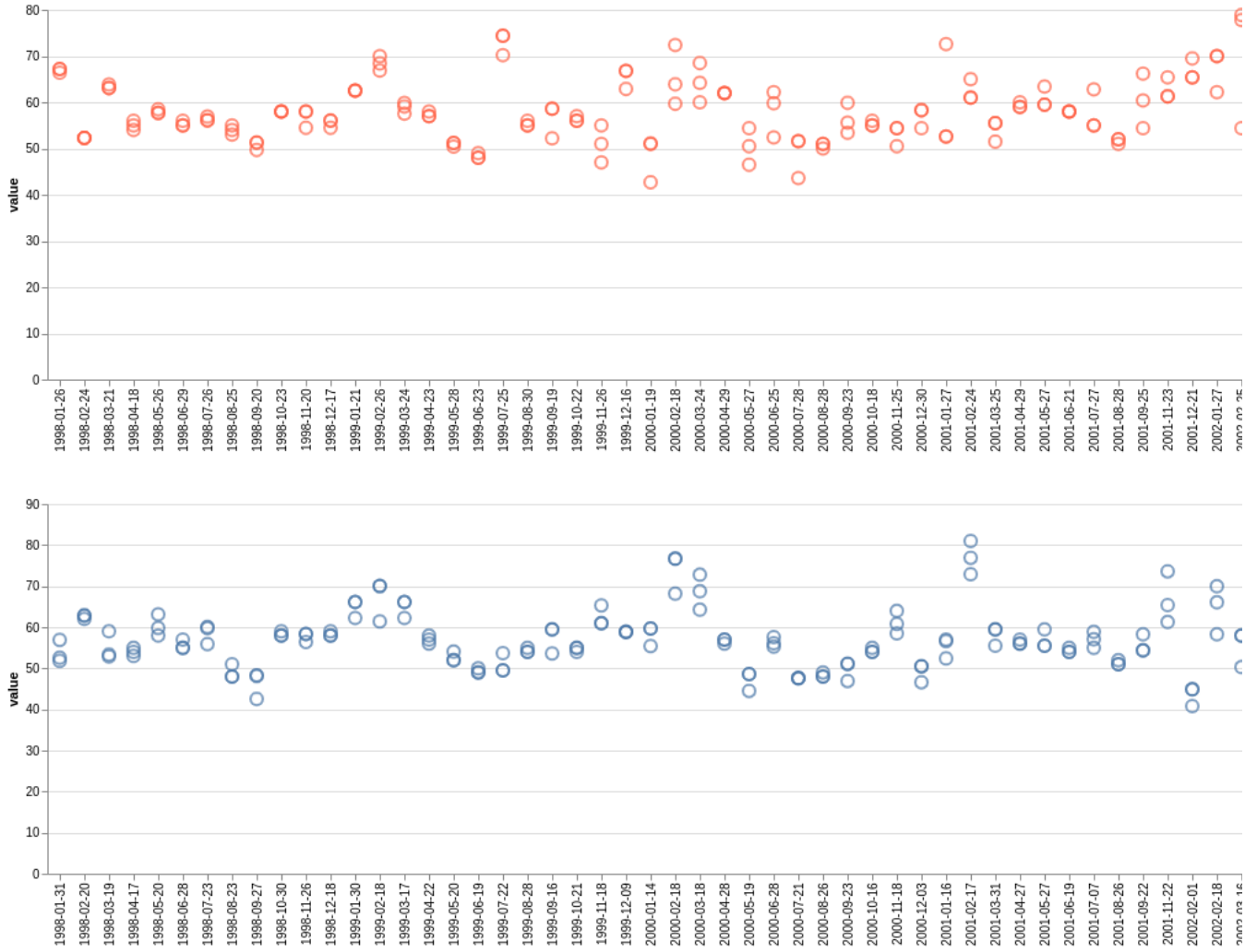


Chart concatenation of Boonsri and Chai for Calcium

```
chart1=alt.Chart(data_sudden_change_Boonsri_Calcium).mark_point(color="tomato",size=100).encode(x="sample date",y="value")
chart2=alt.Chart(data_sudden_change_Chai_Calcium).mark_point(size=100).encode(x="sample date",y="value")
chart1 & chart2
```

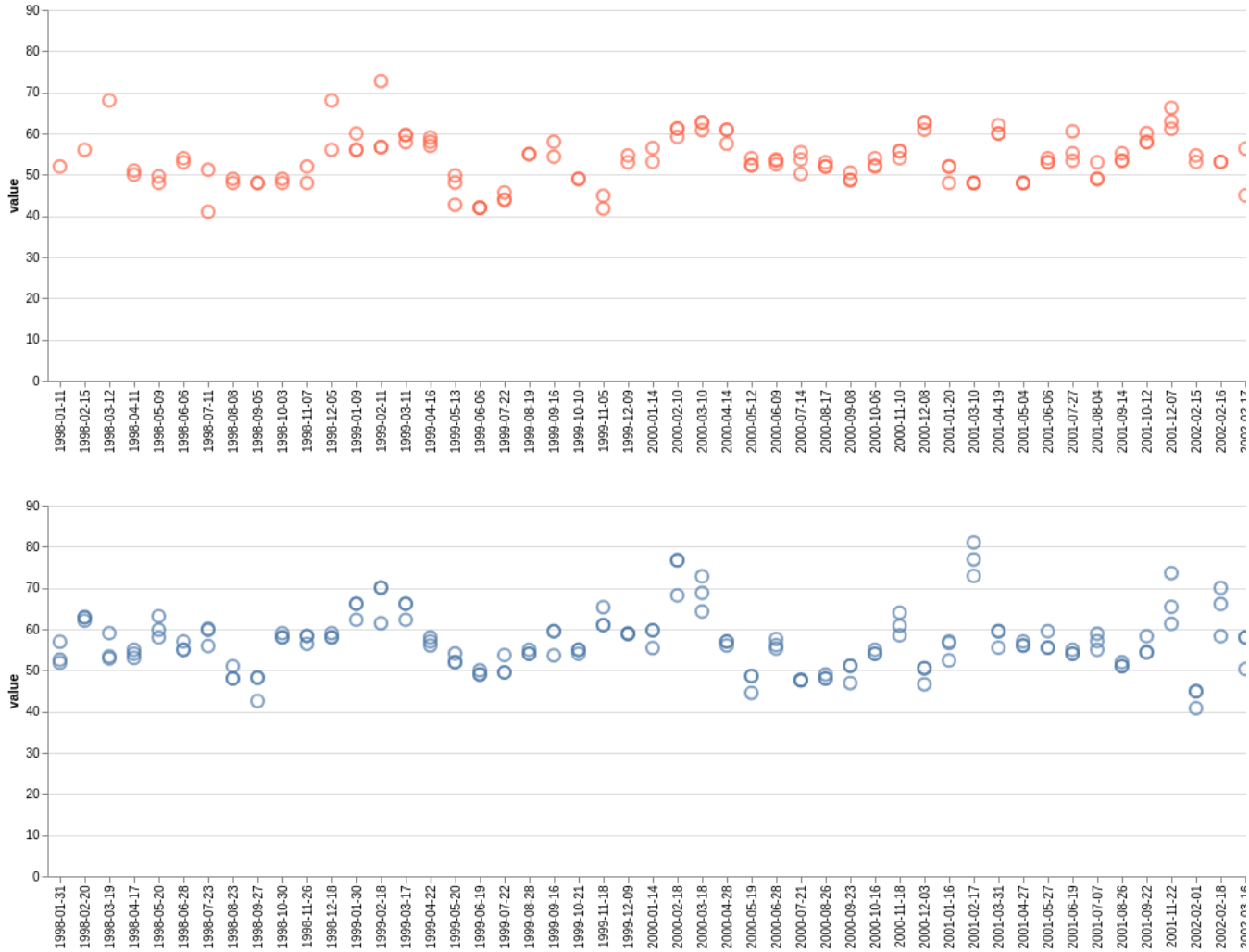
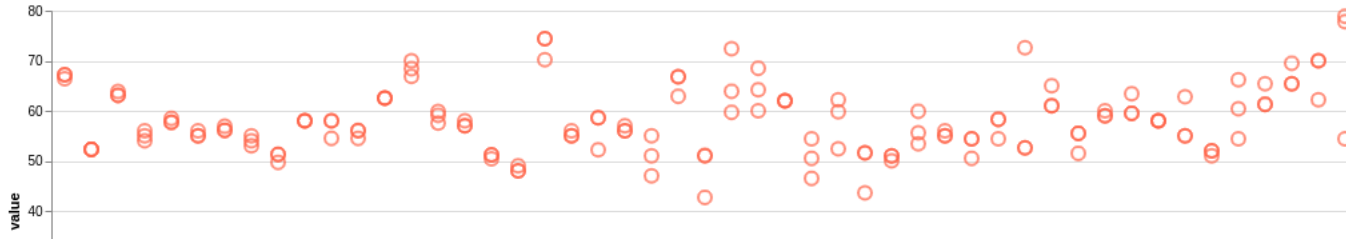


Chart concatenation of Kannika and Boonsri for Calcium

```
chart1=alt.Chart(data_sudden_change_Kannika_Calcium).mark_point(color="tomato",size=100).encode(x="sample date",y="value")
chart2=alt.Chart(data_sudden_change_Boonsri_Calcium).mark_point(size=100).encode(x="sample date",y="value")
chart1 & chart2
```



1 For Boonsri Nitrites

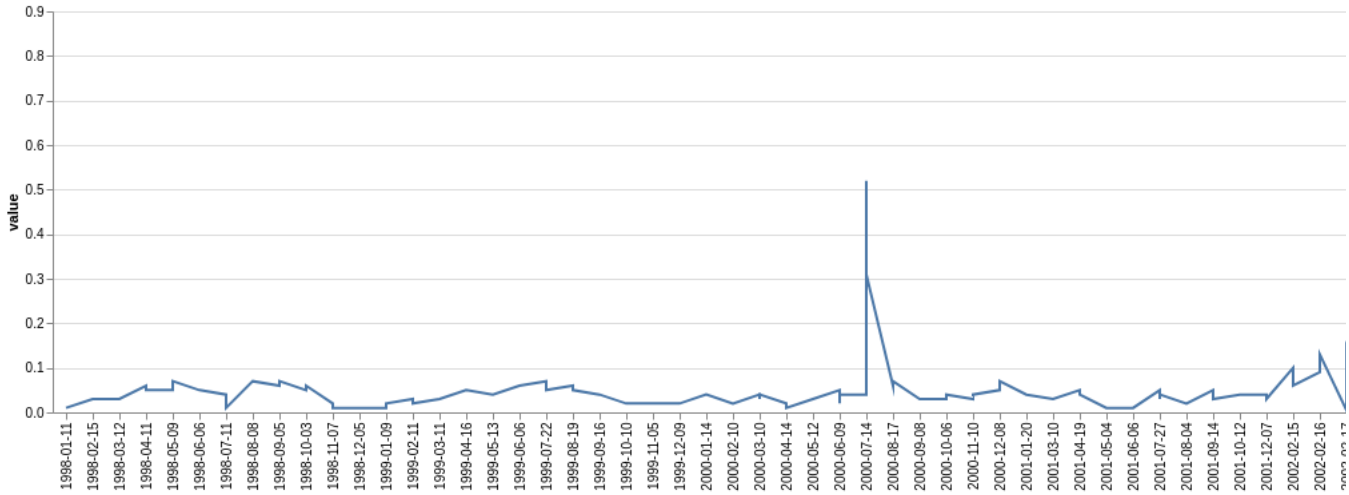

```
df_sudden_change_Boonsri_Nitrites=df[(df['location']=='Boonsri') & (df['measure']=='Nitrites')]
df_sudden_change_Boonsri_Nitrites.head()
```

	id	value	location	sample date	measure
3	2228	0.01	Boonsri	1998-01-11	Nitrites
227	2268	0.03	Boonsri	1998-02-15	Nitrites
451	2308	0.03	Boonsri	1998-03-12	Nitrites
675	2348	0.06	Boonsri	1998-04-11	Nitrites
709	2711	0.05	Boonsri	1998-04-11	Nitrites


```
data_sudden_change_Boonsri_Nitrites=df_sudden_change_Boonsri_Nitrites.drop(['id','location','measure'],axis='columns')
data_sudden_change_Boonsri_Nitrites.head()
```

	value	sample date
3	0.01	1998-01-11
227	0.03	1998-02-15
451	0.03	1998-03-12
675	0.06	1998-04-11
709	0.05	1998-04-11

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




```
alt.Chart(data_sudden_change_Boonsri_Nitrites).mark_point(size=100).encode(
    x='sample date',
    y='value',
)
```

The value of Nitrites at location Boonsri is almost uniform with very few anomalies such as in year 17-08-2000. It stays uniform most of the time

2 For Chai Nitrites

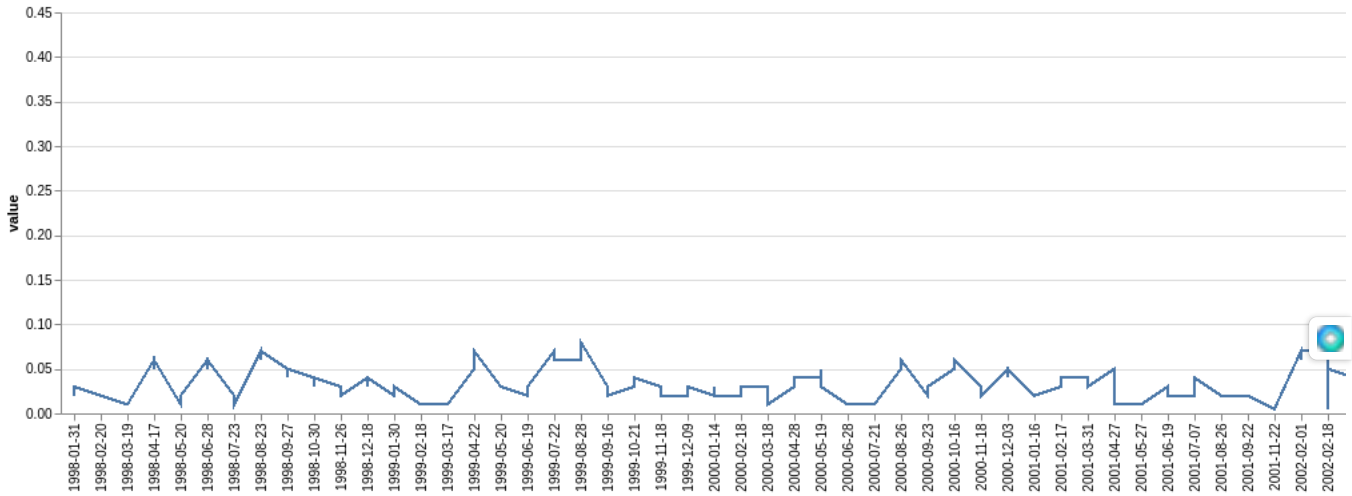
```
df_sudden_change_Chai_Nitrites=df[(df['location']=='Chai') & (df['measure']=='Nitrites')]
df_sudden_change_Chai_Nitrites.head()
```

	id	value	location	sample date	measure
131	5055	0.03	Chai	1998-01-31	Nitrites
163	5539	0.02	Chai	1998-01-31	Nitrites
195	6017	0.03	Chai	1998-01-31	Nitrites
259	5094	0.02	Chai	1998-02-20	Nitrites
291	5578	0.02	Chai	1998-02-20	Nitrites

```
data_sudden_change_Chai_Nitrites=df_sudden_change_Chai_Nitrites.drop(['id','location','measure'],axis='columns')
data_sudden_change_Chai_Nitrites.head()
```

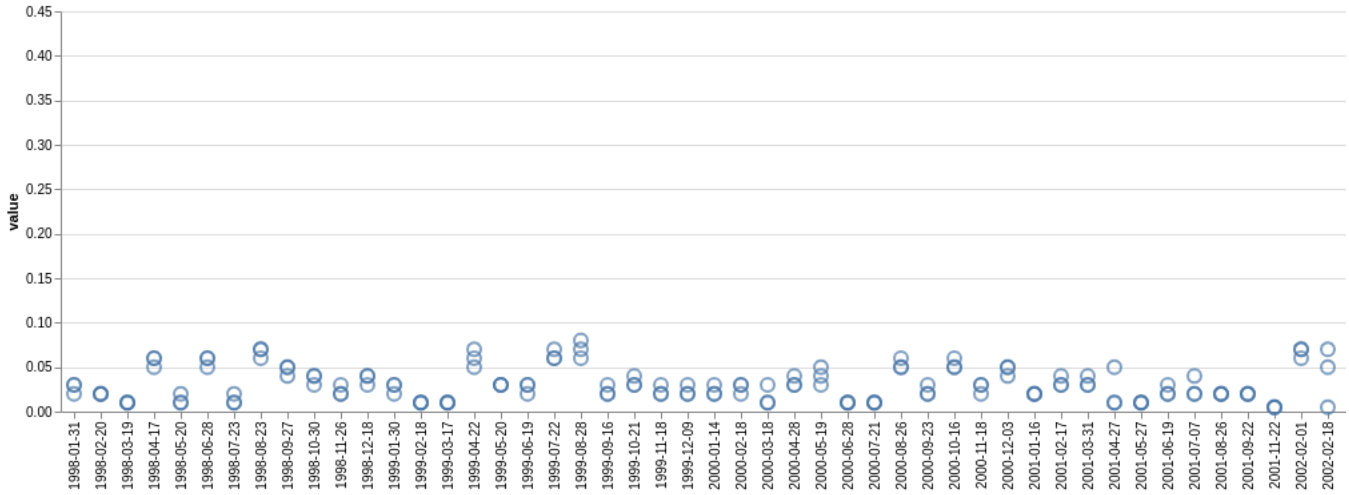
	value	sample date
131	0.03	1998-01-31
163	0.02	1998-01-31
195	0.03	1998-01-31
259	0.02	1998-02-20
291	0.02	1998-02-20

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




```
alt.Chart(data_sudden_change_Chai_Nitrites).mark_point(size=100).encode(
```

```
    x='sample date',
    y='value',
)
```



The value of Nitrites at Chai is varying a lot in non-uniform manner and also there are a lot of anomalies such as in year 13-08-2004 in abruptly increased

3 For Kannika Nitrites

```
df_sudden_change_Kannika_Nitrites=df[(df['location']=='Kannika') & (df['measure']=='Nitrites')]
df_sudden_change_Kannika_Nitrites.head()
```

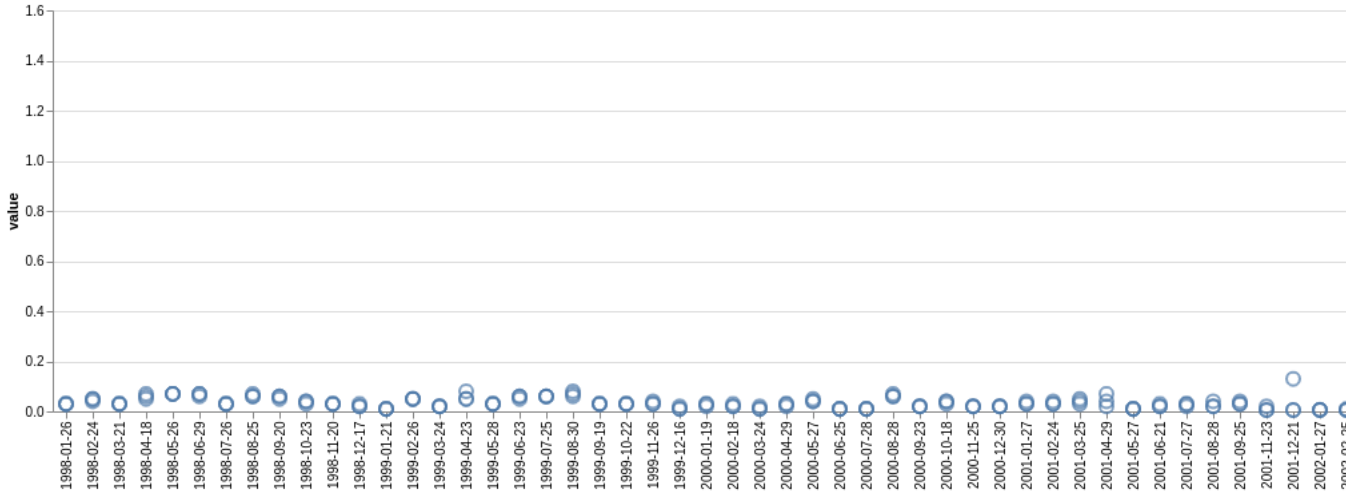
	id	value	location	sample date	measure
35	8298	0.03	Kannika	1998-01-26	Nitrites
67	8779	0.03	Kannika	1998-01-26	Nitrites
99	9262	0.03	Kannika	1998-01-26	Nitrites
355	8338	0.05	Kannika	1998-02-24	Nitrites
387	8819	0.05	Kannika	1998-02-24	Nitrites

```
data_sudden_change_Kannika_Nitrites=df_sudden_change_Kannika_Nitrites.drop(['id','location','measure'],axis='columns')
data_sudden_change_Kannika_Nitrites.head()
```

	value	sample date
35	0.03	1998-01-26
67	0.03	1998-01-26
99	0.03	1998-01-26
355	0.05	1998-02-24
387	0.05	1998-02-24

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


```
alt.Chart(data_sudden_change_Kannika_Nitrites).mark_point(size=100).encode(
    x='sample date',
    y='value',
)
```



The value of Nitrites at Kannika is uniform thorughout and very rare anomalies. It stays uniform with almost linearity.

Chart Concatenation of Kannika and Chai

```
chart1=alt.Chart(data_sudden_change_Kannika_Nitrites).mark_point(color="tomato",size=100).encode(x="sample date",y="value")
chart2=alt.Chart(data_sudden_change_Chai_Nitrites).mark_point(size=100).encode(x="sample date",y="value")
chart1 & chart2
```

Chart Concatenation of Boonsri and Chai

```
chart1=alt.Chart(data_sudden_change_Boonsri_Nitrites).mark_point(color="tomato",size=100).encode(x="sample date",y="value")
chart2=alt.Chart(data_sudden_change_Chai_Nitrites).mark_point(size=100).encode(x="sample date",y="value")
chart1 & chart2
```

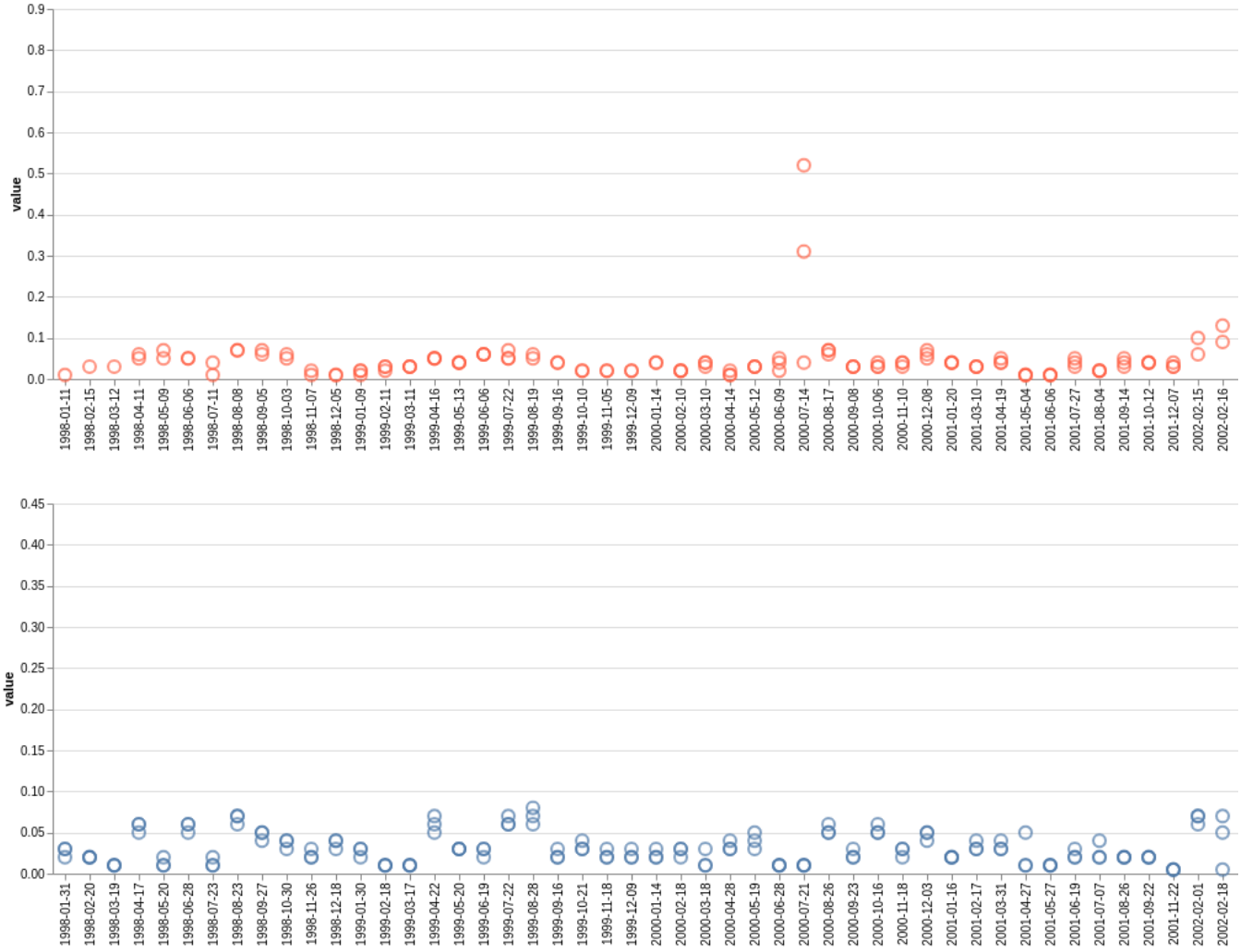
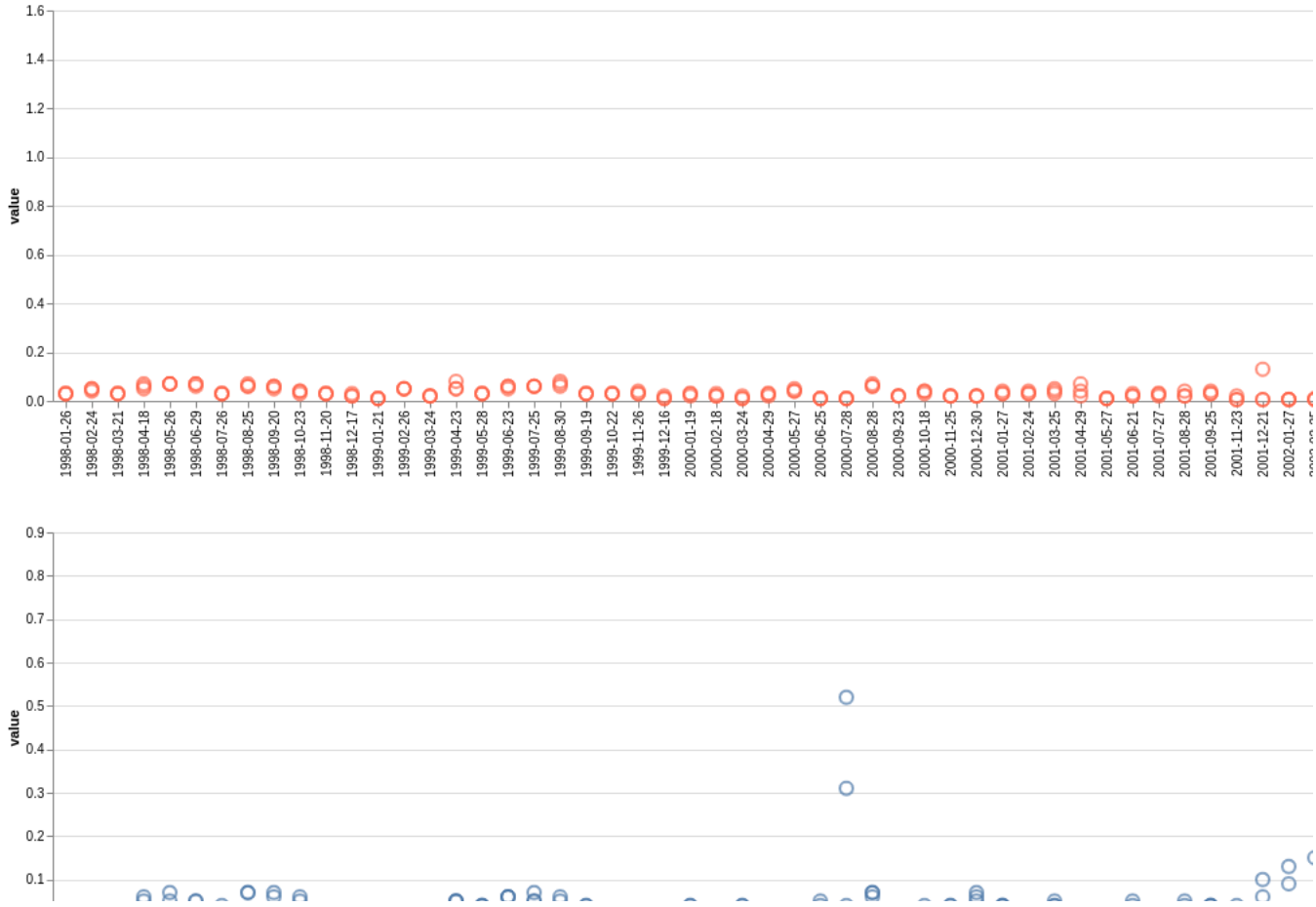


Chart Concatenation of Kannika and Boonsri

```
chart1=alt.Chart(data_sudden_change_Kannika_Nitrites).mark_point(
    size=100,
    color='tomato'
).encode(
    x="sample date",
    y="value")
chart2=alt.Chart(data_sudden_change_Boonsri_Nitrites).mark_point(
    size=100).encode(
    x="sample date",
    y="value")
chart1 & chart2
```




```
df.head()
```

	id	value	location	sample date	measure
0	2221	2.00	Boonsri	1998-01-11	Water temperature
1	2223	9.10	Boonsri	1998-01-11	Dissolved oxygen
2	2227	0.33	Boonsri	1998-01-11	Ammonium
3	2228	0.01	Boonsri	1998-01-11	Nitrites
4	2229	1.47	Boonsri	1998-01-11	Nitrates

Visualization for Ammonium at three Boonsri, Chai, Kannika together

```
df_mixed=df[((df['location']=='Boonsri') | (df['location']=='Chai') | (df['location']=='Kannika')) & (df['measure']=='Ammoni
```

```
df_mixed.head()
```

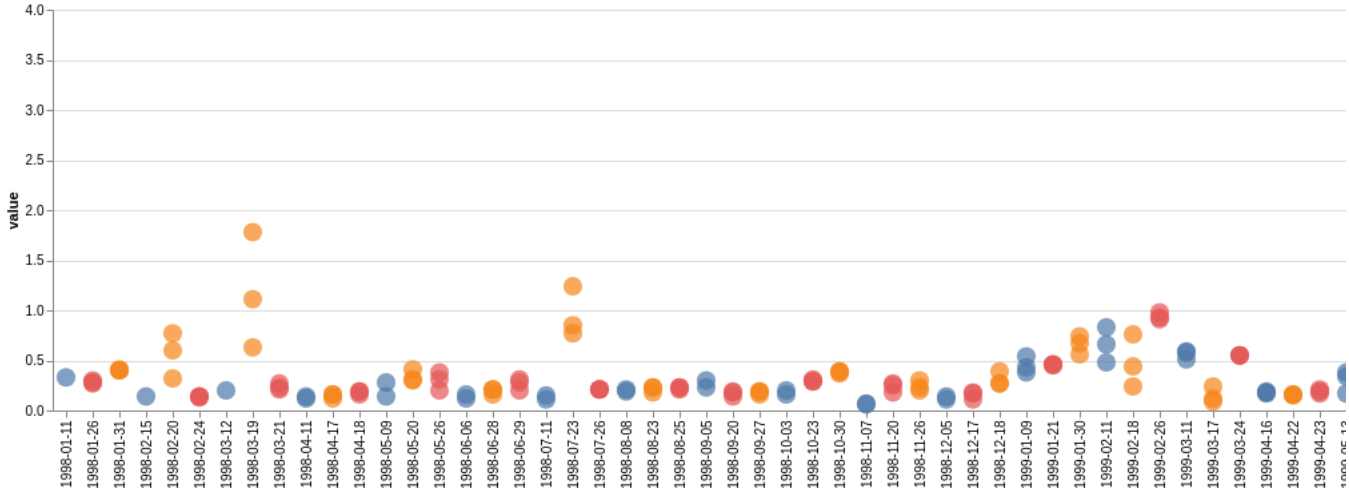
	id	value	location	sample date	measure
2	2227	0.33	Boonsri	1998-01-11	Ammonium
34	8297	0.30	Kannika	1998-01-26	Ammonium
66	8778	0.28	Kannika	1998-01-26	Ammonium
98	9261	0.27	Kannika	1998-01-26	Ammonium
130	5054	0.40	Chai	1998-01-31	Ammonium

```
df_mixed=df_mixed.drop(['id','measure'],axis='columns')
df_mixed.head()
```

	value	location	sample date

```
alt.Chart(df_mixed).mark_circle(size=200).encode(
    x='sample date',
    y='value',
    color='location',

).properties(
    title='Values of Ammonium at "locations"'
)
```



Blue - Boonsri, Yellow - Chai, Red - Kannika

Here we can see that around year 1998 value of Ammonium at 'Chai' is much greater than the others whereas around year 2000 value of Ammonium at Kannika was much greater

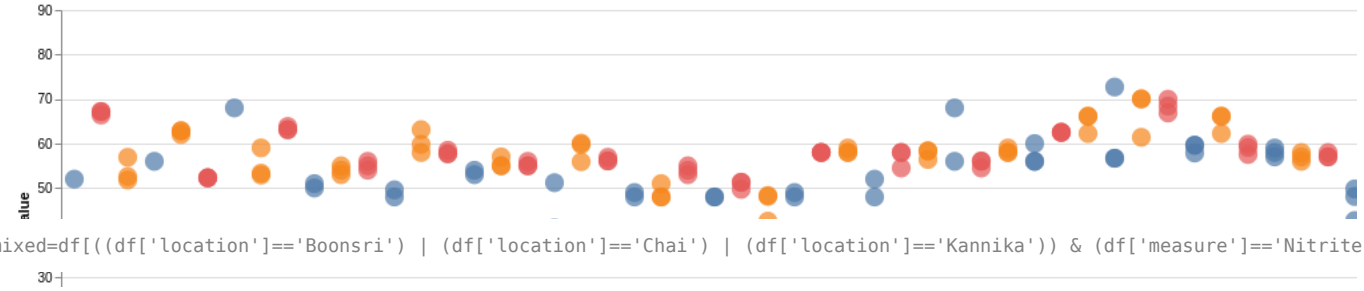
```
df_mixed=df[((df['location']=='Boonsri') | (df['location']=='Chai') | (df['location']=='Kannika')) & (df['measure']=='Calciu
```

```
df_mixed=df_mixed.drop(['id','measure'],axis='columns')
df_mixed.head()
```

	value	location	sample date
9	52.00	Boonsri	1998-01-11
41	67.19	Kannika	1998-01-26
73	66.40	Kannika	1998-01-26
105	67.20	Kannika	1998-01-26
137	52.60	Chai	1998-01-31

```
alt.Chart(df_mixed).mark_circle(size=200).encode(
    x='sample date',
    y='value',
    color='location',

).properties(
    title='Values of Calcium at "locations"'
)
```




```
df_mixed=df[((df['location']=='Boonsri') | (df['location']=='Chai') | (df['location']=='Kannika')) & (df['measure']=='Nitrite
```



```
df_mixed.head()
```

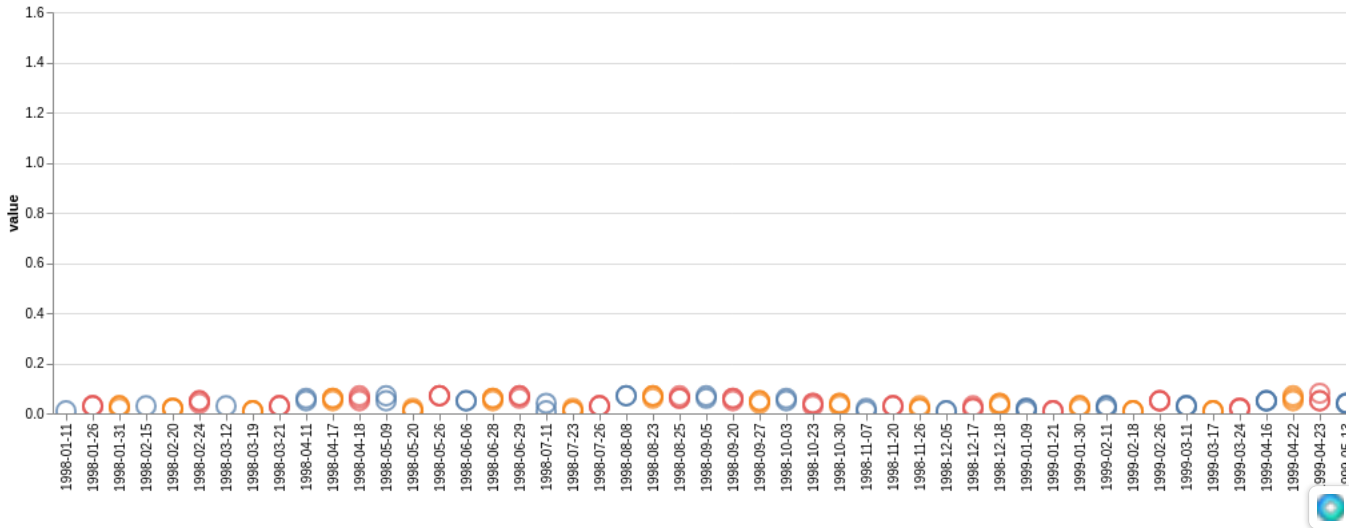
	id	value	location	sample date	measure
3	2228	0.01	Boonsri	1998-01-11	Nitrites
35	8298	0.03	Kannika	1998-01-26	Nitrites
67	8779	0.03	Kannika	1998-01-26	Nitrites
99	9262	0.03	Kannika	1998-01-26	Nitrites
131	5055	0.03	Chai	1998-01-31	Nitrites

```
df_mixed=df_mixed.drop(['id','measure'],axis='columns')
df_mixed.head()
```

	value	location	sample date
3	0.01	Boonsri	1998-01-11
35	0.03	Kannika	1998-01-26
67	0.03	Kannika	1998-01-26
99	0.03	Kannika	1998-01-26
131	0.03	Chai	1998-01-31

```
alt.Chart(df_mixed).mark_point(size=200).encode(
    x='sample date',
    y='value',
    color='location',

).properties(
    title='Values of Nitrites at "locations"'
).interactive()
```



Now we will group them according to dates and then calculate their mean values

```
df=pd.read_csv('Boonsong Lekagul waterways readings (2).csv')
df.head()
```

	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
4	2229	1.47	Boonsri	11-Jan-98	Nitrates

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
```