CIS3200 Term Project Tutorial (Group 2)

Authors:

Instructor: Jongwook Woo

Date 12/18/2020

Lab Tutorial

Christopher Valdepena (cvaldep3@calstatela.edu)

Jiancong Zhang (jzhang100@calstatela.edu)

My Truong (mtruon42@calstatela.edu)

Ryan Applegate(rappleg@calstatela.edu)

Getting started on Elastic Cloud with a Sample Dataset "Superstore"

Objectives

In this hands-on lab, you will learn how to:

- Find clear data sets

- Visit and create an account for Elastic Cloud

- Implement data sets

- Create Simple Graphs Visualizations

- Create GEO_Map_Visualizations

Platform Specifications:
- ElasticSearch & Kibana

- CPU Speed: 3.0GHz

- # of CPU cores: 4

- # of nodes: 1 node

- Total Memory Size: 500MB

**Step 1: Visit and create an account for Elastic Cloud**

1. Go to https://www.elastic.co/cloud/as-a-service

2. Register then Log into your ES (Elastic Cloud) account by using your email account
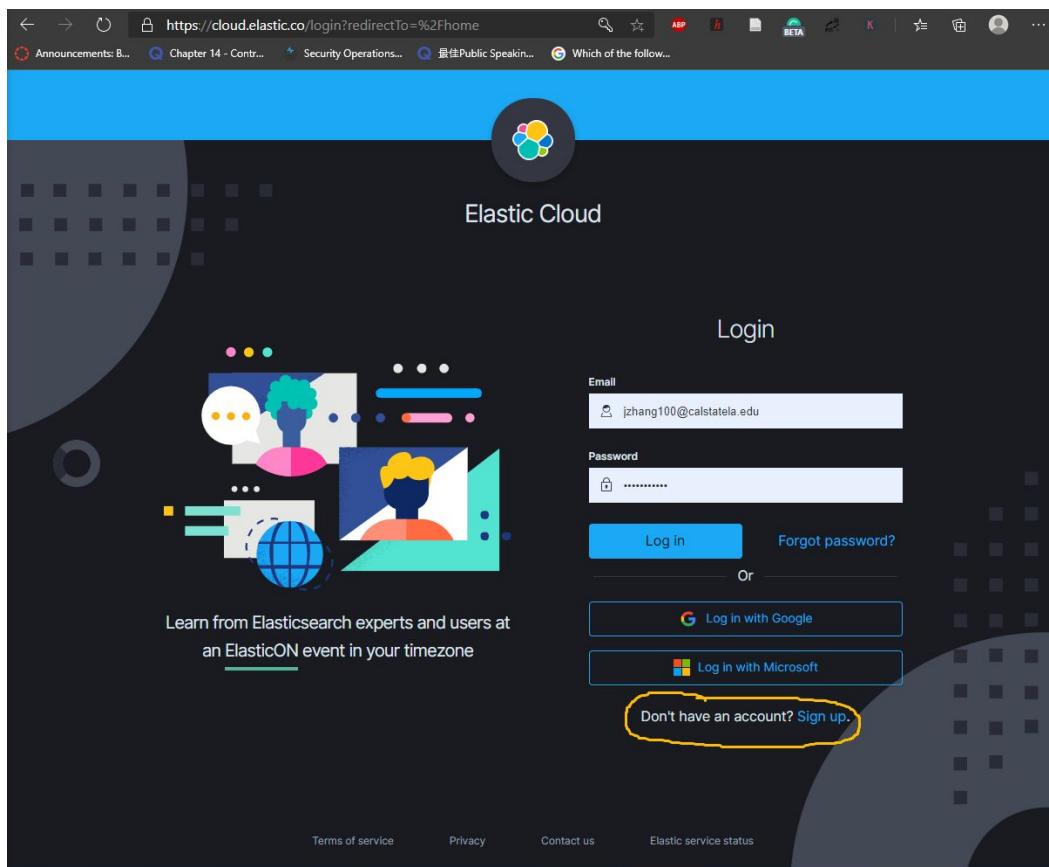
3. Click on the verification link in the email that was sent to your ".edu" address

4. After email verification, you will be prompted to create a password for your Elastic Cloud.

5. Log into your Elastic Cloud account

6. NOTE: If the verification email expires, go to https://cloud.elastic.co/forgot and enter the student email address to trigger a new verification email. Create your first hosted Elasticsearch cluster

Set up the Instances

7. Once you have signed in you will see an overview of your deployments. Click on the "Create Deployment" button

8. Name your deployment, choose the provider and region that you prefer. For trials the deployment size please set up the instances as the next graph showed

9. One thing please pay attention, Since the max memory option may change over time. please just select the maximum option that they provide you.
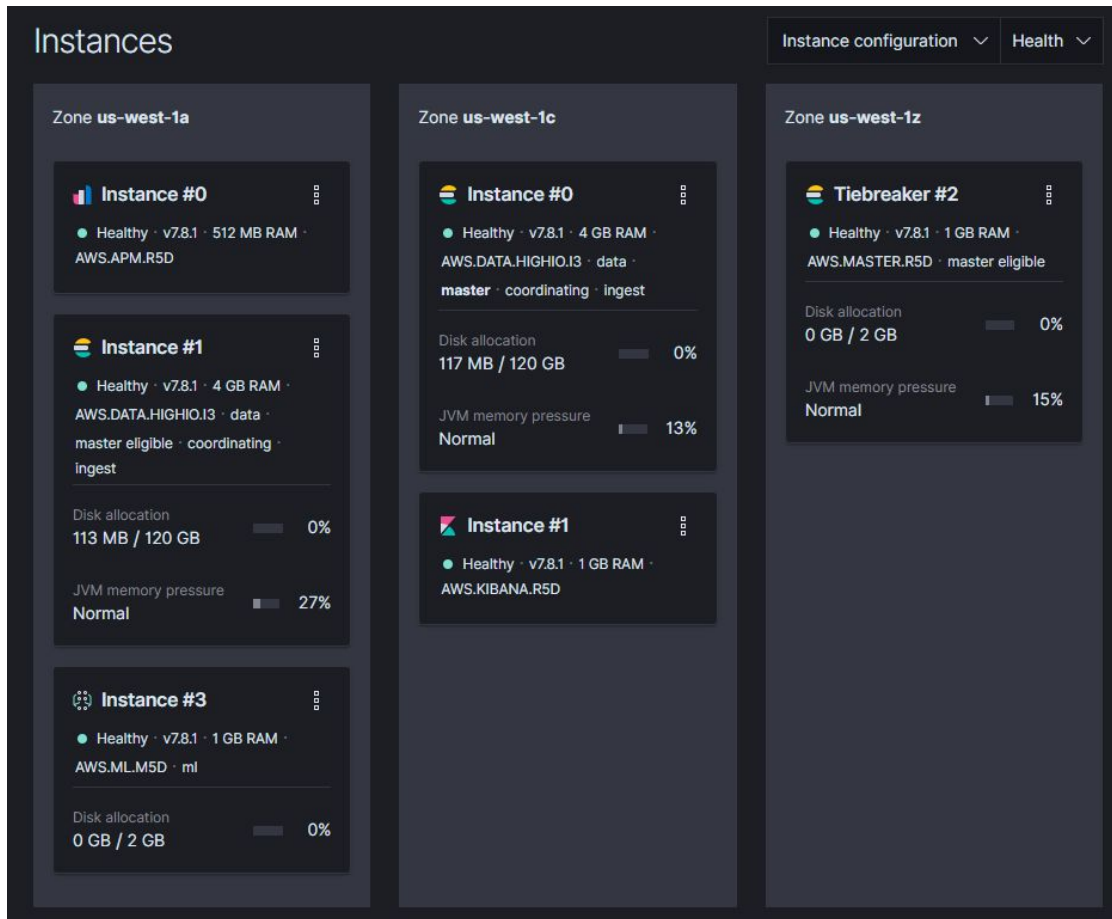
10. After logging into the ES account, you will see the following page as the image below. Then, click on your Kibana:
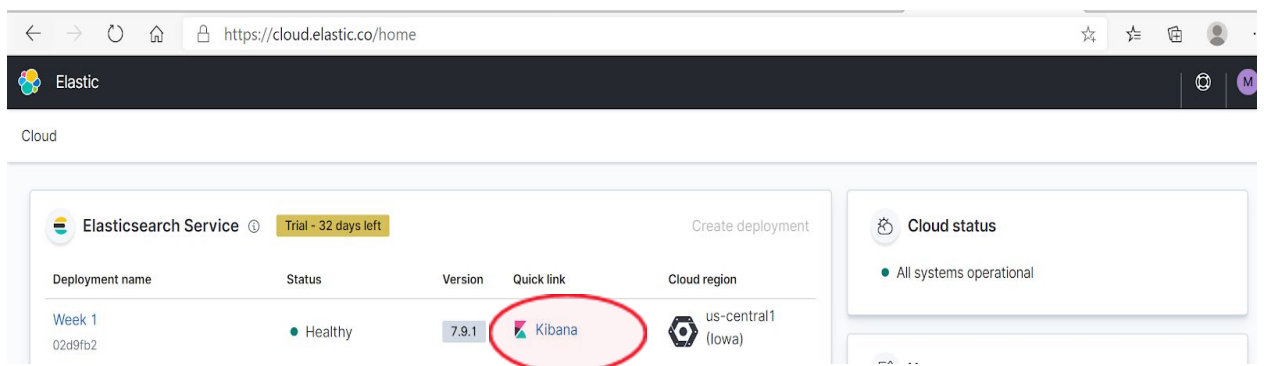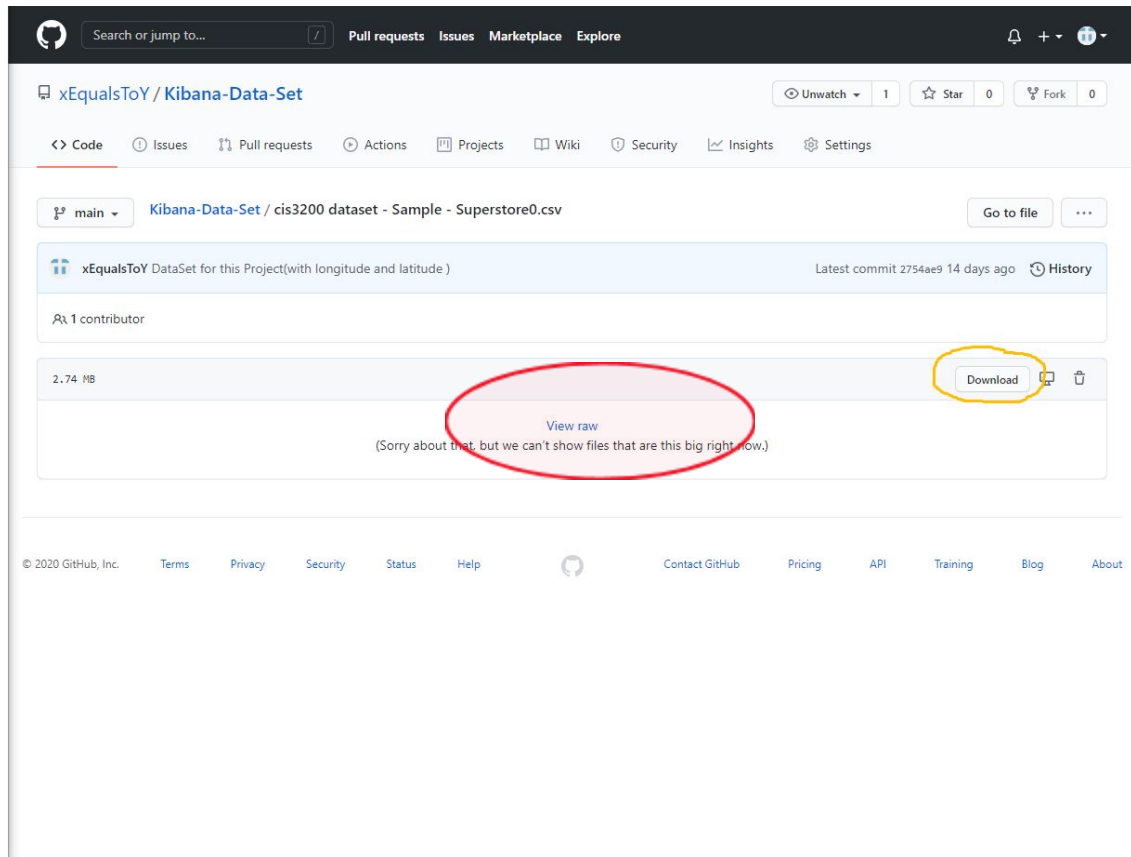
11. Next, you will need to download the data sets from <u>Kibana-Data-Set/cis3200 dataset - Sample - Superstore0.csv at main · xEqualsToY/Kibana-Data-Set (github.com)</u> (see next graph)

12. Come back to the Kibana pages, drag the data set into the red area.

Visualize data from a log file EXPERIMENTAL

The File Data Visualizer helps you understand the fields and metrics in a log file. Upload your file, analyze its data, and then choose whether to import the data into an Elasticsearch index.

The File Data Visualizer supports these file formats:

- Delimited text files, such as CSV and TSV
- Newline-delimited JSON
- Log files with a common format for the timestamp

You can upload files up to 100 MB.

This feature is experimental. Got feedback? Please create an issue in GitHub.

Select or drag and drop a file

13. Then, you will see this page on your screen.

14. Once selecting the *Import* button, we want to navigate to the *Advanced* tab for set the mapping for geo spatial data.

15. Now we add the "coordinates": { "type": "geo_point" } to declare the variable coordinates with the data type geo_point for our geo spatial visuals. We do not have to add any mapping for the *Longitude and Latitude* dimensions since they are already present in the downloaded dataset from our github.

Mappings

```
62        },
63 ▾      "Segment": {
64        |   "type": "keyword"
65        },
66 ▾      "Ship Date": {
67        |   "type": "date",
68        |   "format": "MM/dd/yyyy||M/dd/yyyy||M/d/yyyy||MM/d/yyyy"
69        },
70 ▾      "Ship Mode": {
71        |   "type": "keyword"
72        },
73 ▾      "State": {
74        |   "type": "keyword"
75        },
76 ▾      "Sub-Category": {
77        |   "type": "keyword"
78        },
79 ▾      "coordinates": {
80        |   "type": "geo_point"
81        }|
82    }
```

16. Next, we head over to the *Ingest Pipeline* to declare the values for our *coordinates* dimension.


Ingest pipeline

```
34          }
35       },
36       {
37           "date": {
38               "field": "Order Date",
39               "timezone": "{{ event.timezone }}",
40               "formats": [
41                   "MM/d/yyyy",
42                   "M/dd/yyyy",
43                   "MM/dd/yyyy",
44                   "M/d/yyyy"
45               ]
46           }
47       },
48    } ,
49       "append":{
50       "field":"coordinates",
51       "value":["{{Latitude}}, {{Longitude}}"]
52    },
53       {
54           "convert": {
```

17. Then, create a name for your dataset at the "Index name" tab as you can see in the red circle.
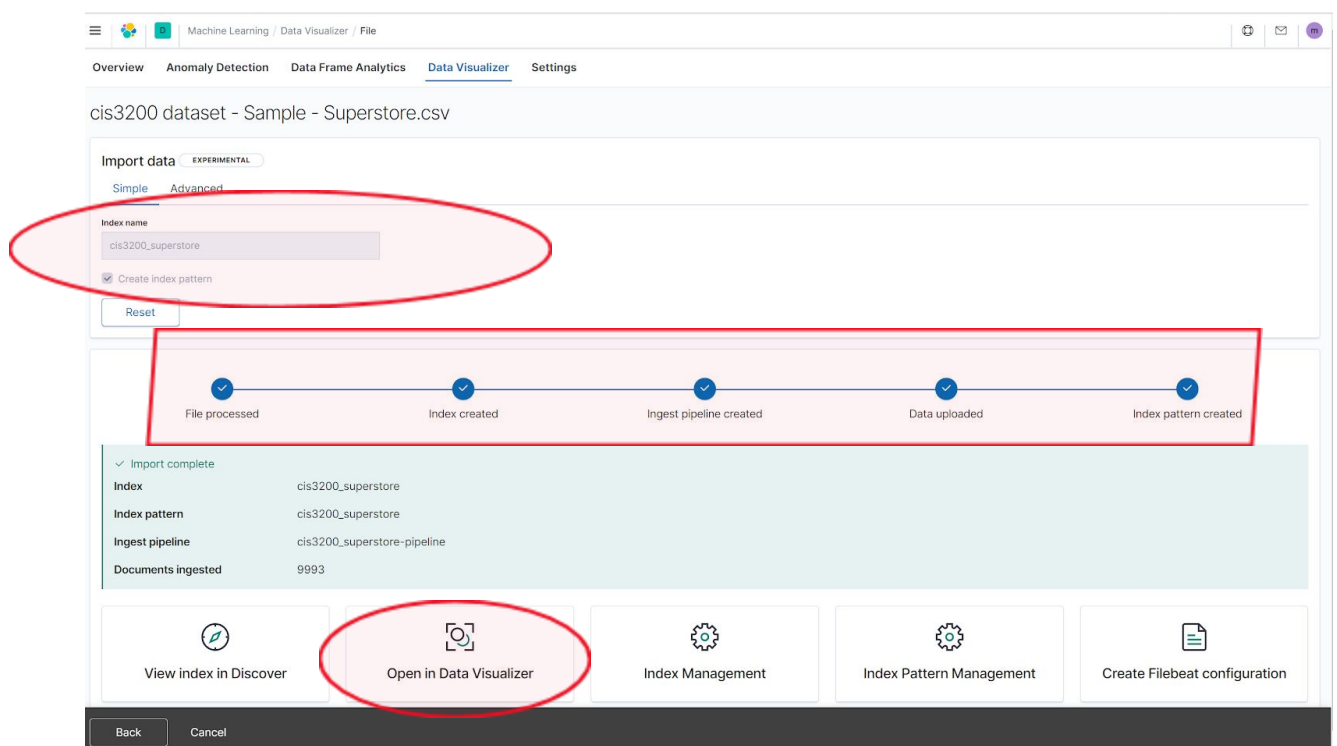
One thing to pay attention to here, make sure to check the "create index pattern" box First!

 After that, please click import in the red circle. Then, you will see the following pipeline as the picture below which shows the imported process with geo_point. If there is no error, the process is completed.
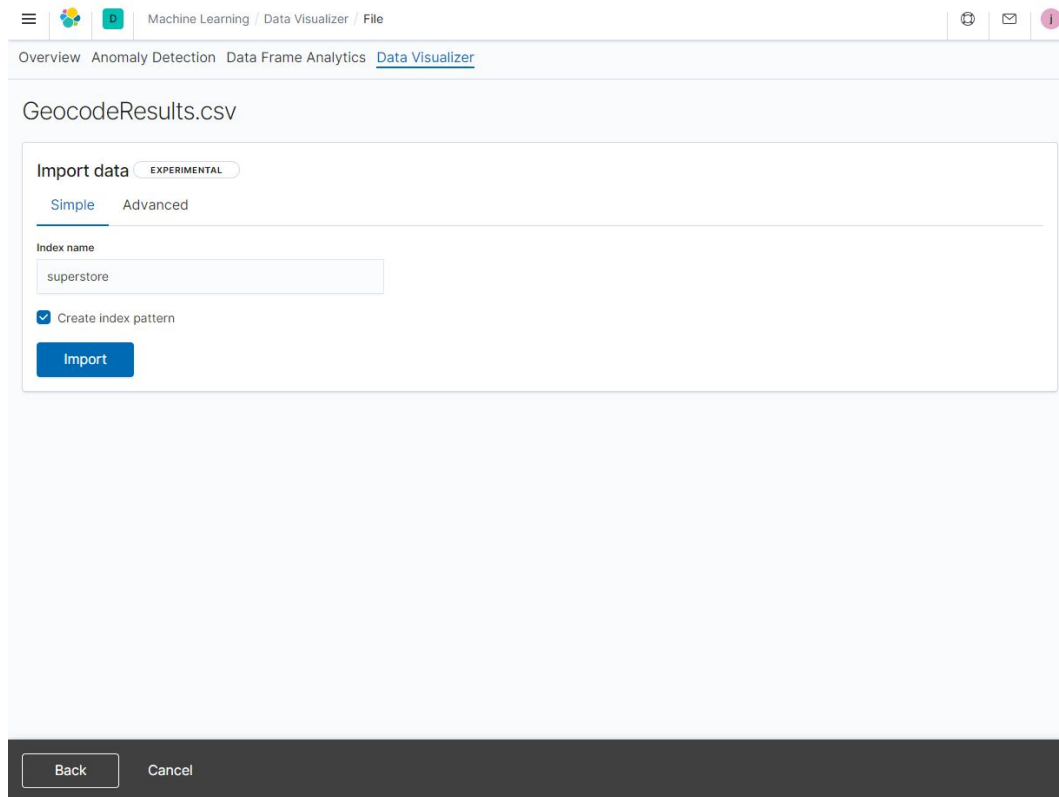
18. Next, click on the "Open in Data Visualizer."

19. Then, it will take you to the window where you can check the time range, and also see metrics and total fields of your dataset.
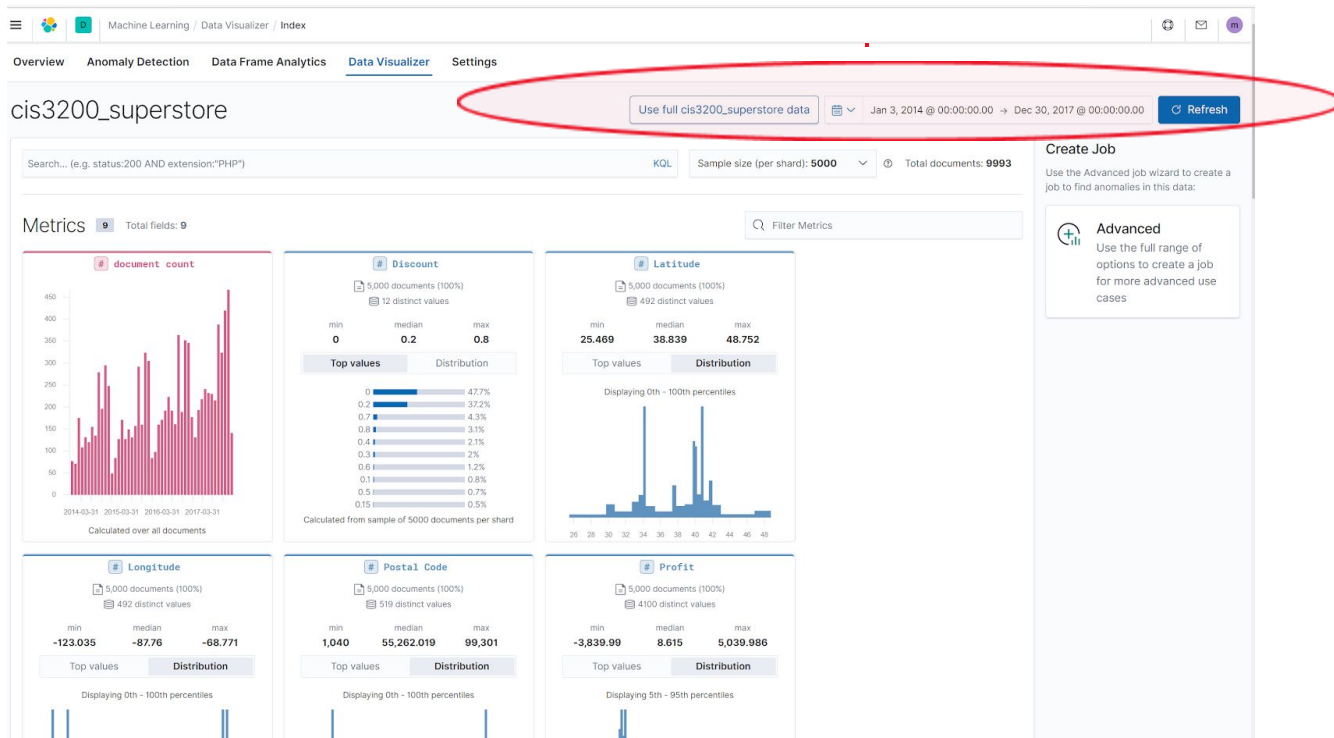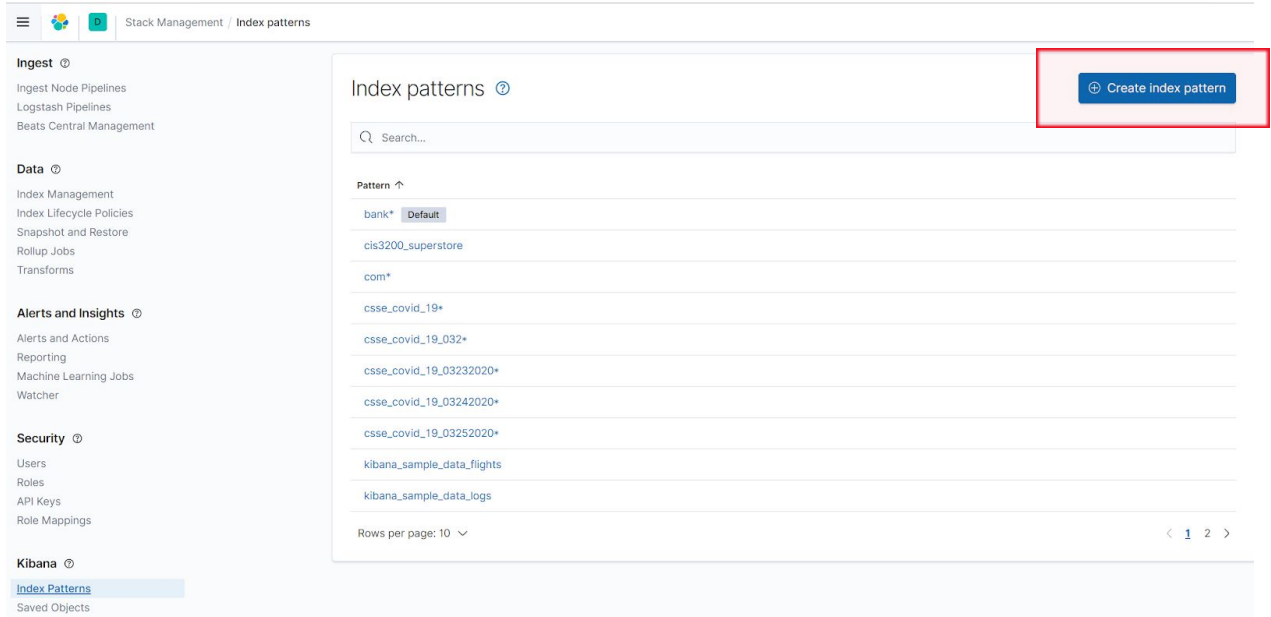

17. Next, click on the "Index Management" under the "Data" tab. It will take you to the step where you can create Index Management.

18. Then, click on "Create index pattern" to create the new one and follow with creating a new name for the index as you can see in the image below with the red circles are:

18. After creating the name, you can continue to click on "Time field" to choose one. This feature is used to filter the data based on time. The drop down fields will display all the time and data related fields from the index.

Now, your new index pattern will look like the image below:

● <mark>Create Simple Graphs Visualizations</mark>

19. Click on "Visualize" tab to start creating the Visualization for the data as follow:

20. Continue to click on the tab "Create visualization" and choose "Vertical Bar" as follow:

Visualizations

<div style="border:1px solid red;padding:4px;display:inline-block;float:right">⊕ Create visualization</div>

🔍 Search...

	Title	Type	Description	Actions
☐	BAR CHART LAB3	▥ Vertical Bar		✎
☐	Bar Example	▥ Vertical Bar		✎
☐	Markdown Example	⊞ Markdown		✎
☐	Pie Example	◔ Pie		✎

Visualizations

New Visualization

Filter

Lens [B]	Area	Controls [E]	Data Table
Gauge	Goal	Heat Map	Horizontal Bar
Line	Maps	Markdown	Metric
Pie	TSVB	Tag Cloud	Timelion
Vega [E]	Vertical Bar		

Select a visualization type

Start creating your visualization by selecting a type for that visualization.

Try Lens, our new, intuitive way to create visualizations.

Go to Lens
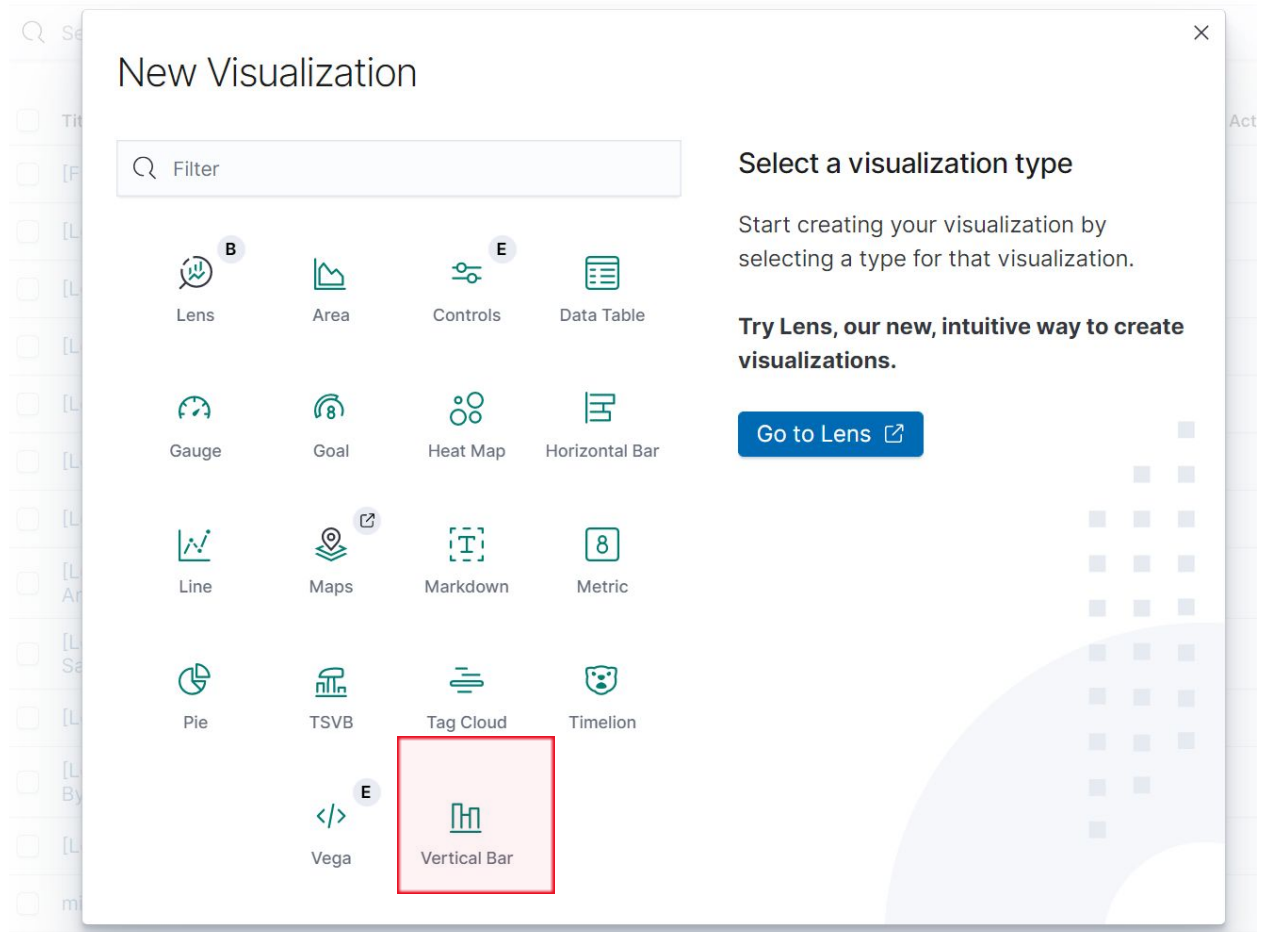
21. At the "New Vertical Bar," click on the index pattern name which you just created in the previous step, "cis3200_superstore*," which is the one we should click to start creating the bar chart:

New Vertical Bar / Choose a source

Search...

Sort ∨   Types 2 ∨

[Flights] Flight Log

bank*

cis3200_superstore

cis3200_superstore*
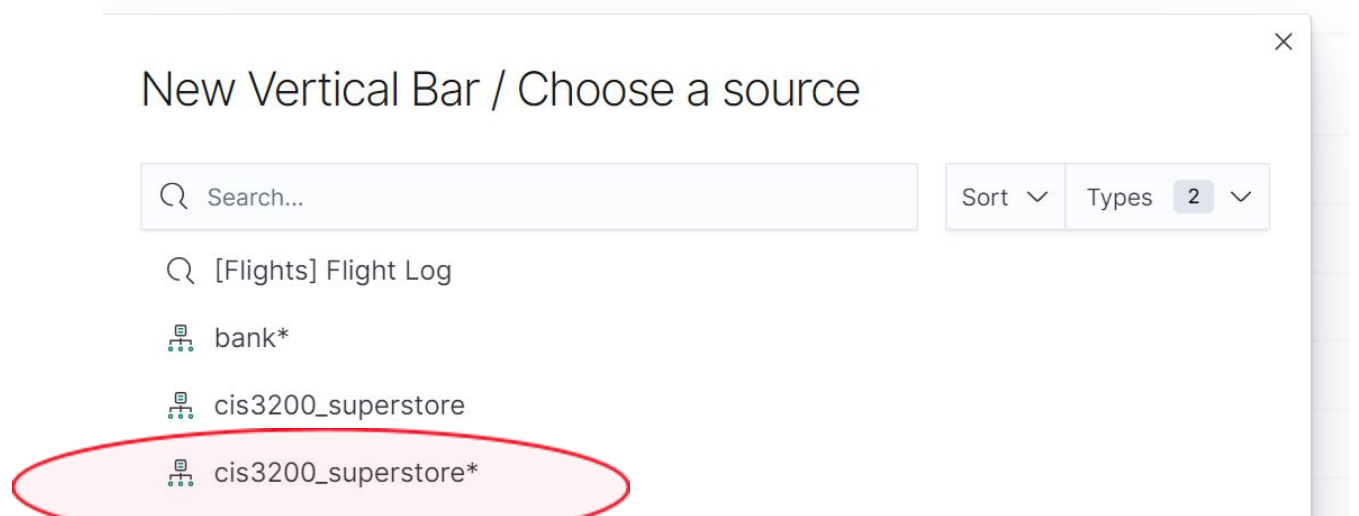
22.Next, you will need to set the time range, please select from Jan.1st 2014, 00:00:00, to Jan 1st 2018, 00:00:00. and then click the blue button named "Refresh".
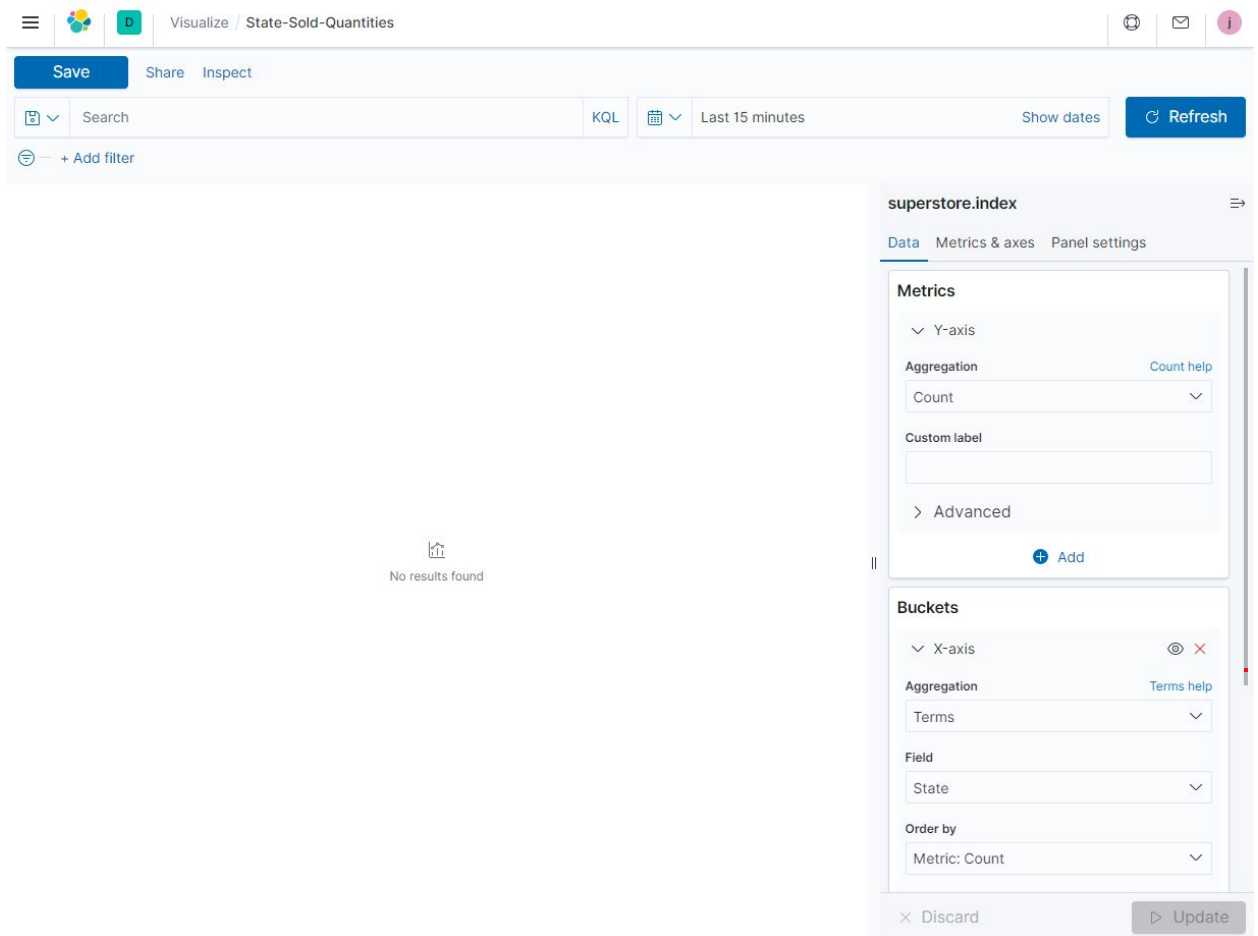
23. After that, follow the graph's right menu setting, make sure the "Group other values in separate bucket" and "Show missing values" both default and they should be off. You should see a graph of each state's sales count.   **Remember to save it!**
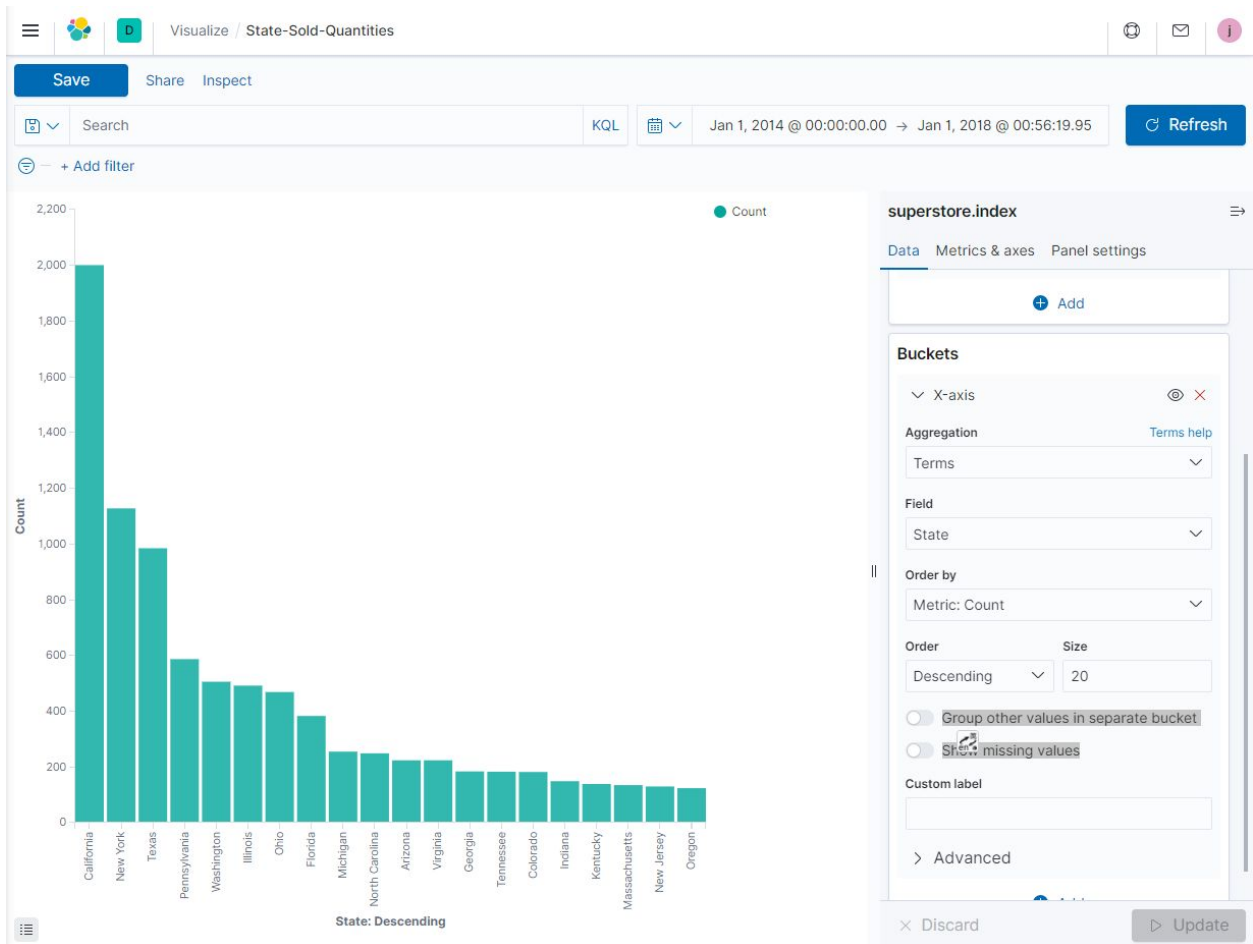
24. Click the left menu and click to visualize. We will do a line chart for this step.(see the yellow circle)

Home

Recently viewed

Kibana

Discover
Dashboard
Canvas
Maps
Machine Learning
Graph
Visualize

Observability

Logs
Metrics
APM
Uptime

Security

SIEM

Management

Dev Tools
Stack Monitoring
Stack Management

Dock navigation

✕ close

KQL   Jan 1, 2014 @ 00:00:00.00 → Jan 1, 2018 @ 00:56:19.95   ↻ Refresh

● Count

superstore.index

Data   Metrics & axes   Panel settings

⊕ Add

Buckets

∨ X-axis   👁 ✕

Aggregation   Terms help

Terms   ∨

Field

State   ∨

Order by

Metric: Count   ∨

Order   Size

Descending ∨   20

⚪ Group other values in separate bucket

⚪ Show missing values

Custom label

＞ Advanced

✕ Discard   ▷ Update

State: Descending

25. Then select Line in the red circle.

26. Select the time range from Jan.1st 2014, 00:00:00, to Jan 1st 2018, 00:00:00. and then click the blue button named "Refresh". Aggregation choose Average, Field "profit".

Add a X-axis , For Aggregation select "Terms", for the Field select "State' and **Save it.**

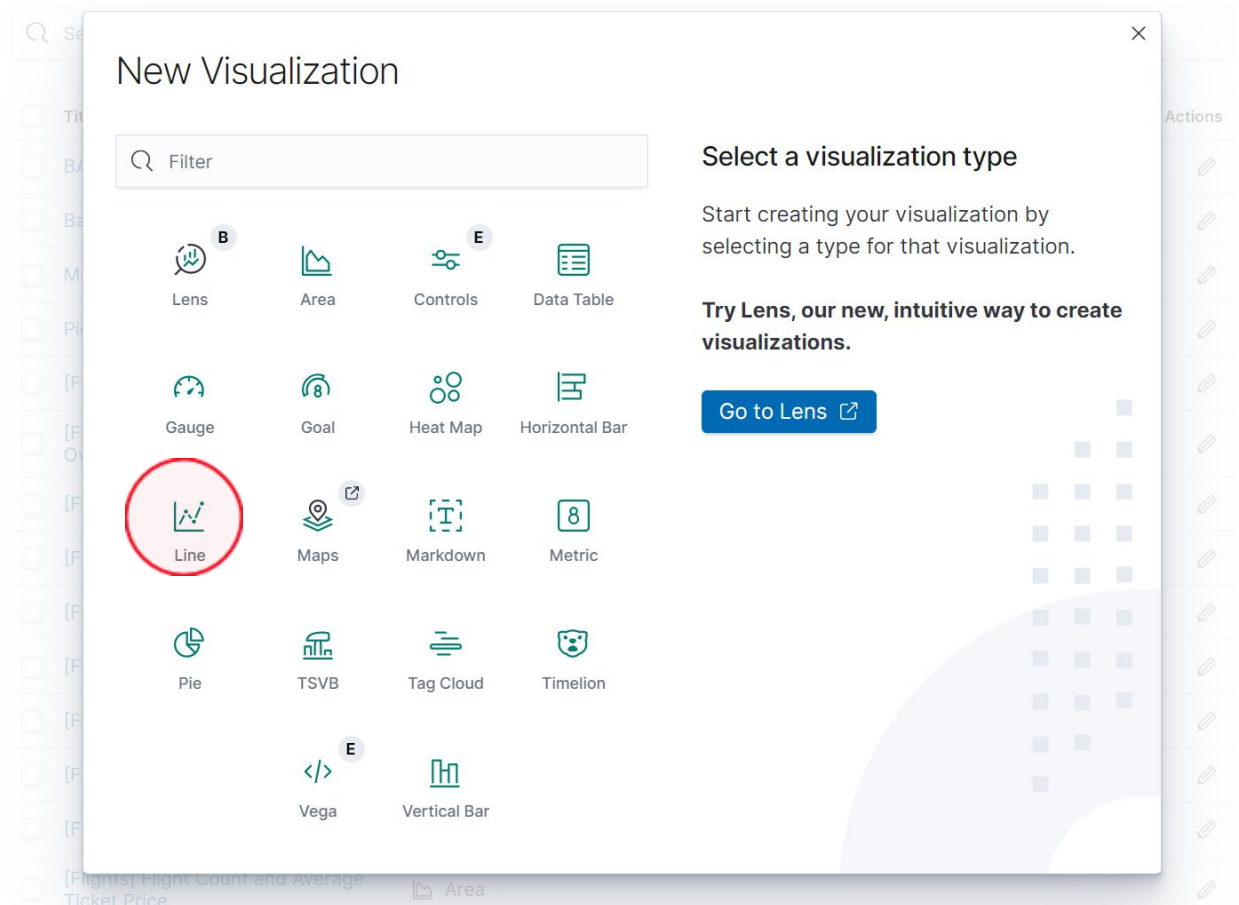**you should have a graph similar to the one at the next.**

26.Click the left menu and click to visualize. We will do a pie chart for this step.(see the yellow circle)
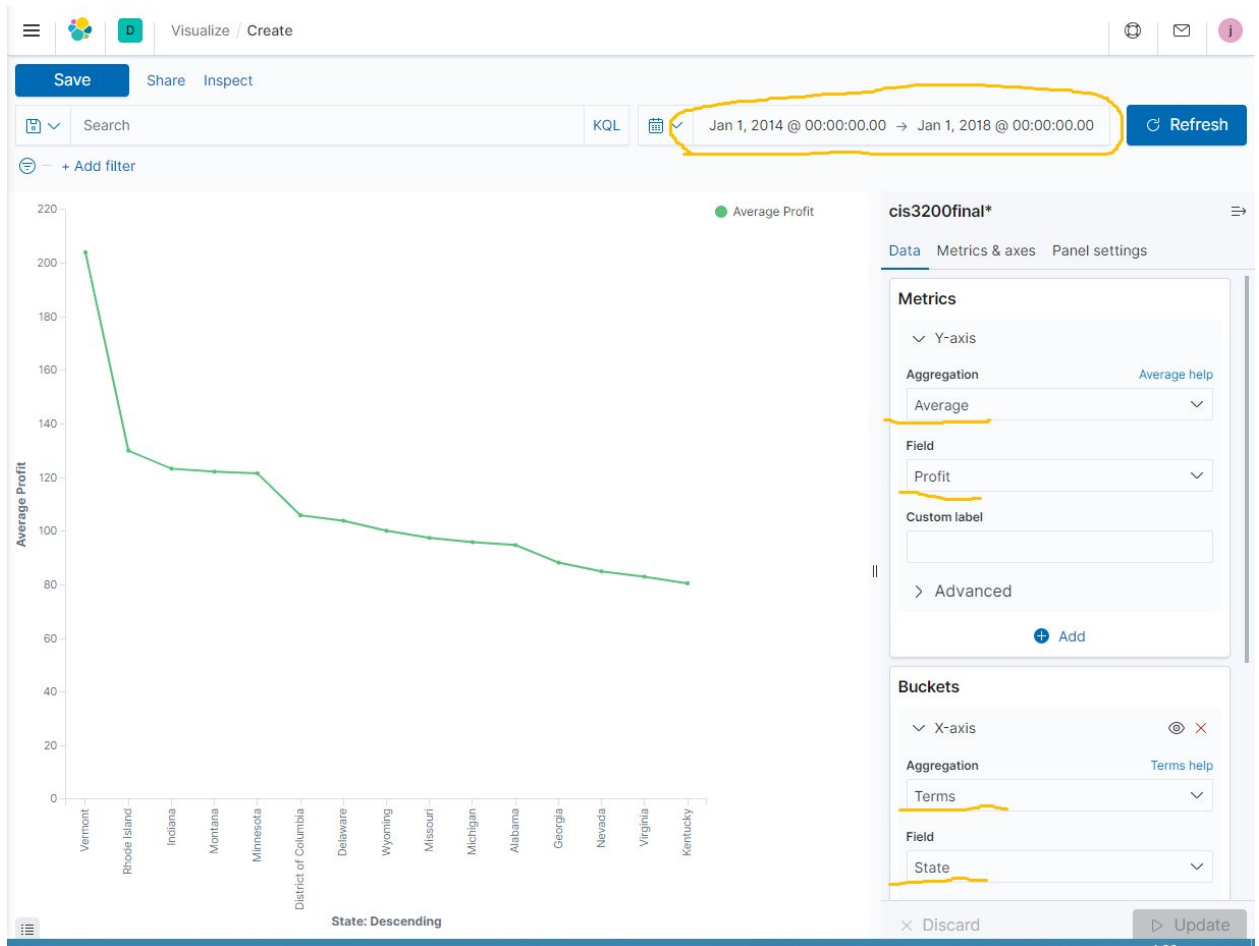
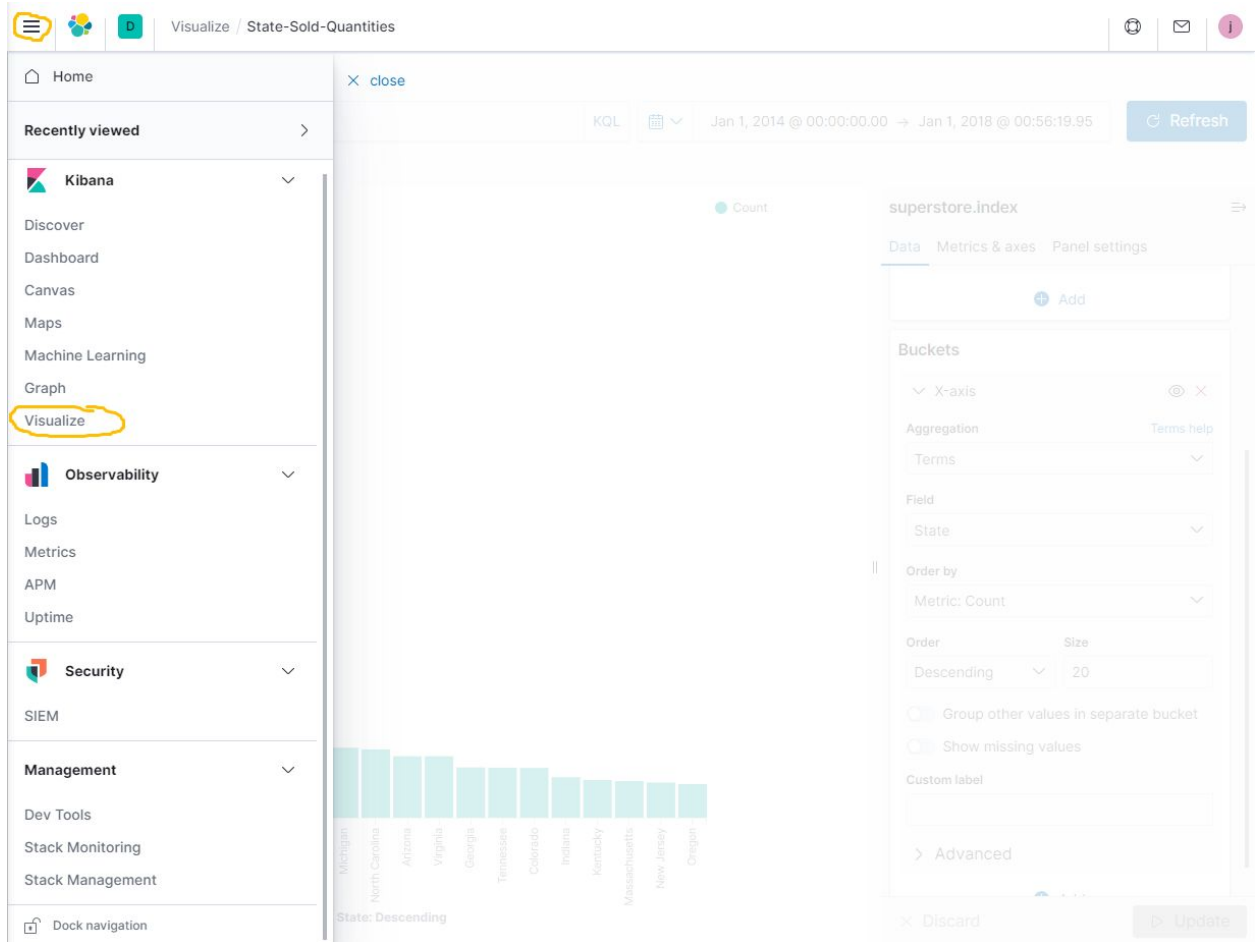27.Then select the Pie In the red circle.
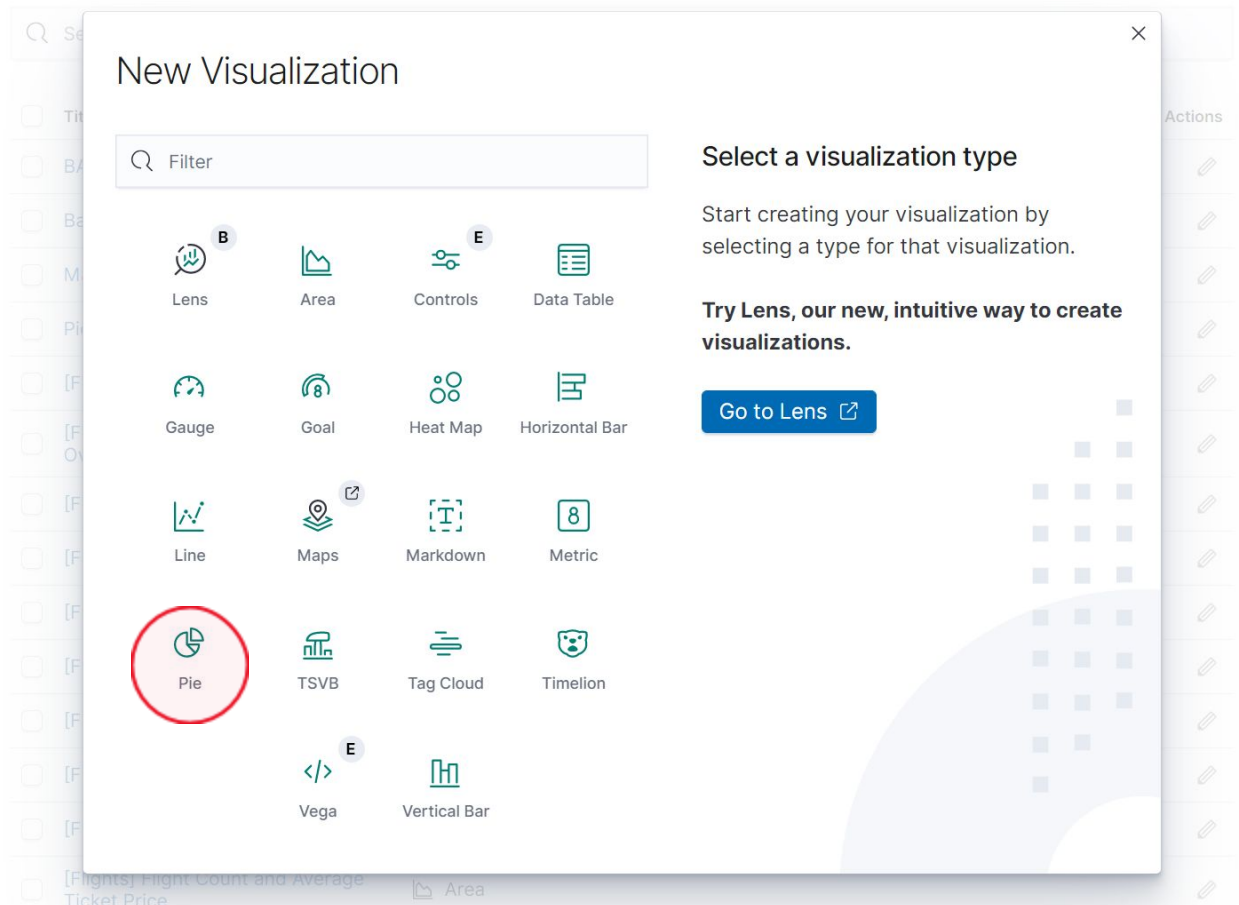
28.Select the time range from Jan.1st 2014, 00:00:00, to Jan 1st 2018, 00:00:00. and then click the blue button named "Refresh".

For Aggregation choose Count, At Buckets add a Split Slices,Aggregation choose Term Filed choose Region , order by Metric:Count . Order Descending size 5. others keep default. You should have the next graph shown on your screen, **Save it.**
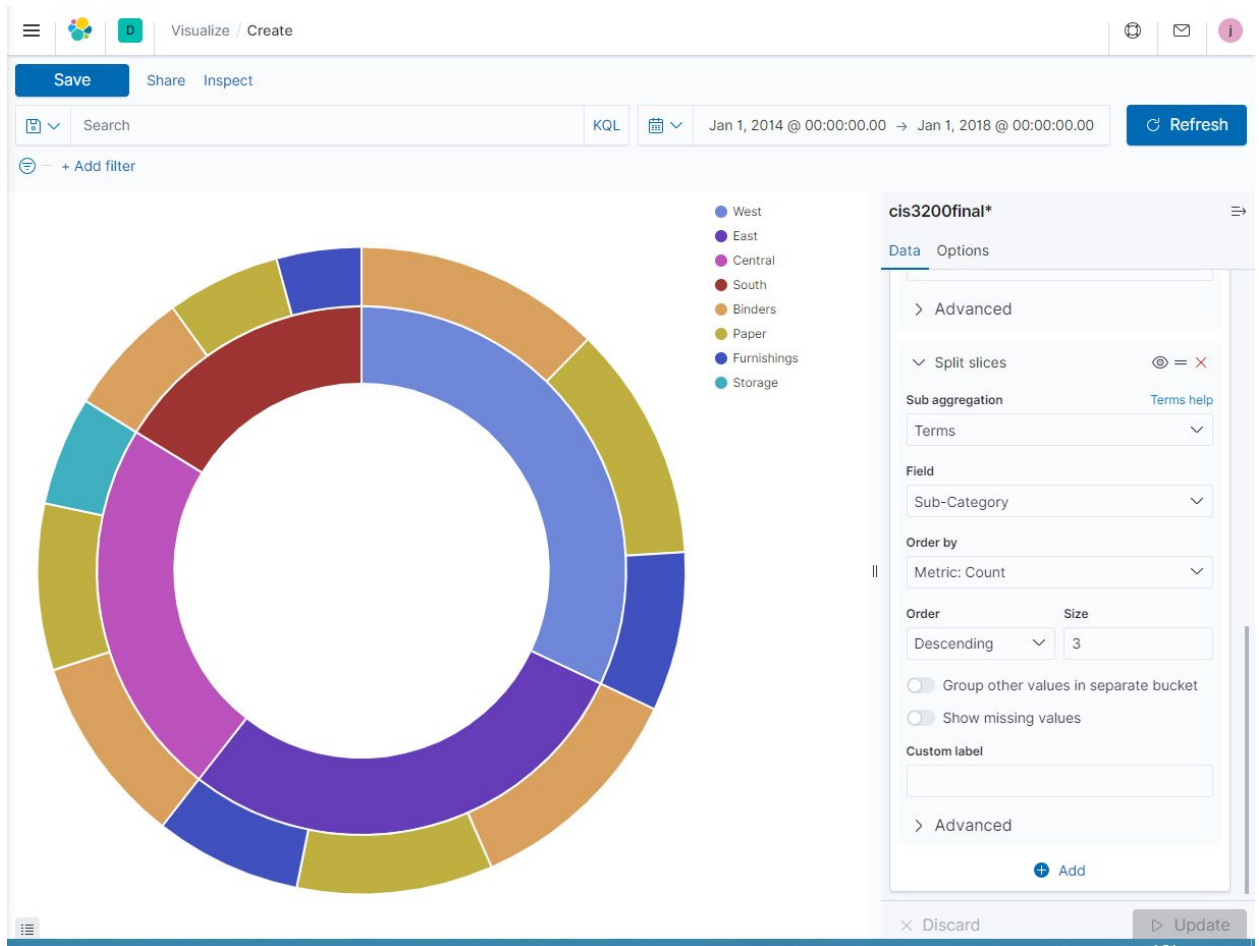
29. Next we will do a quarter sales report at a 4 years's range

Click the left menu and click to visualize.

30.Select Line graph in red circle.

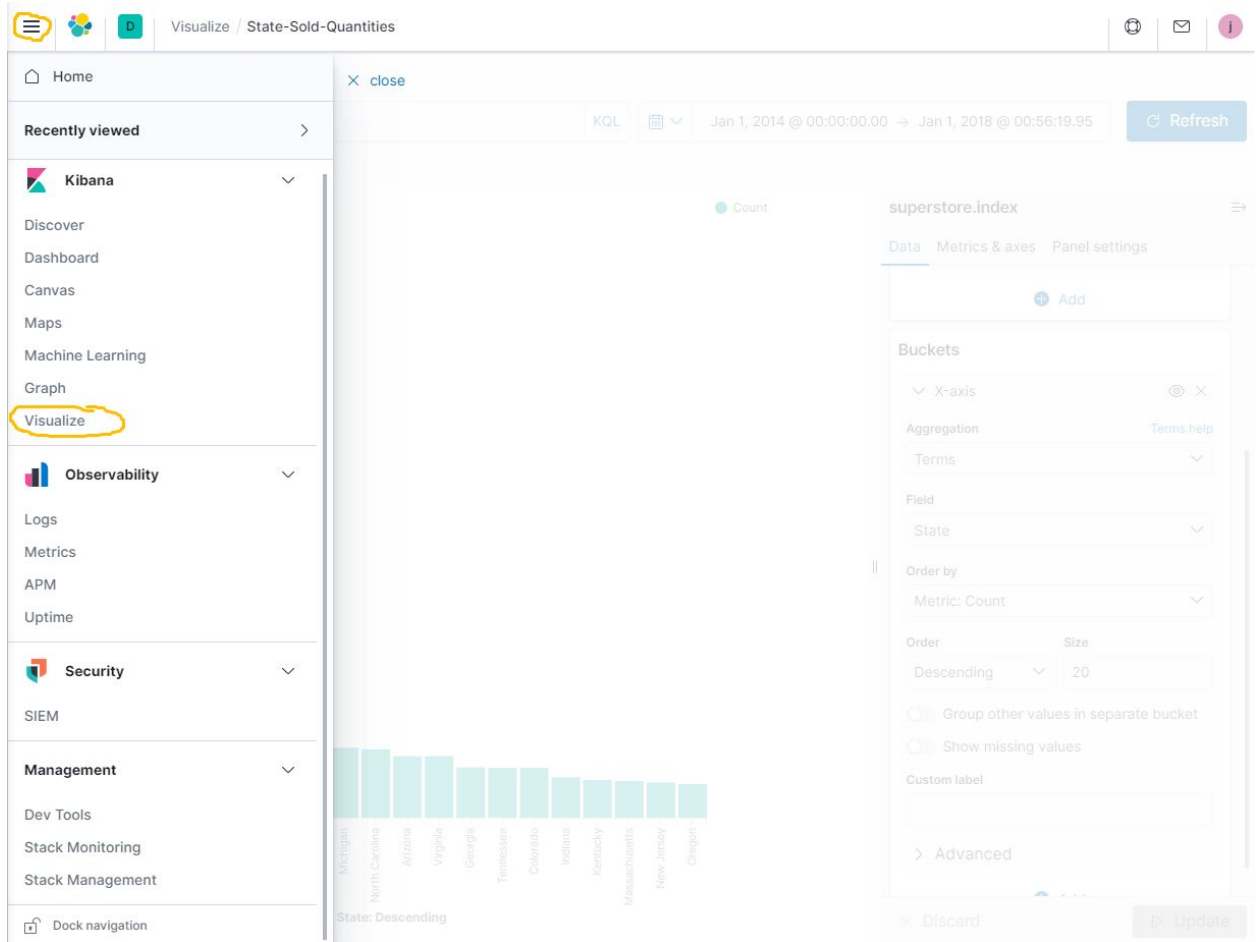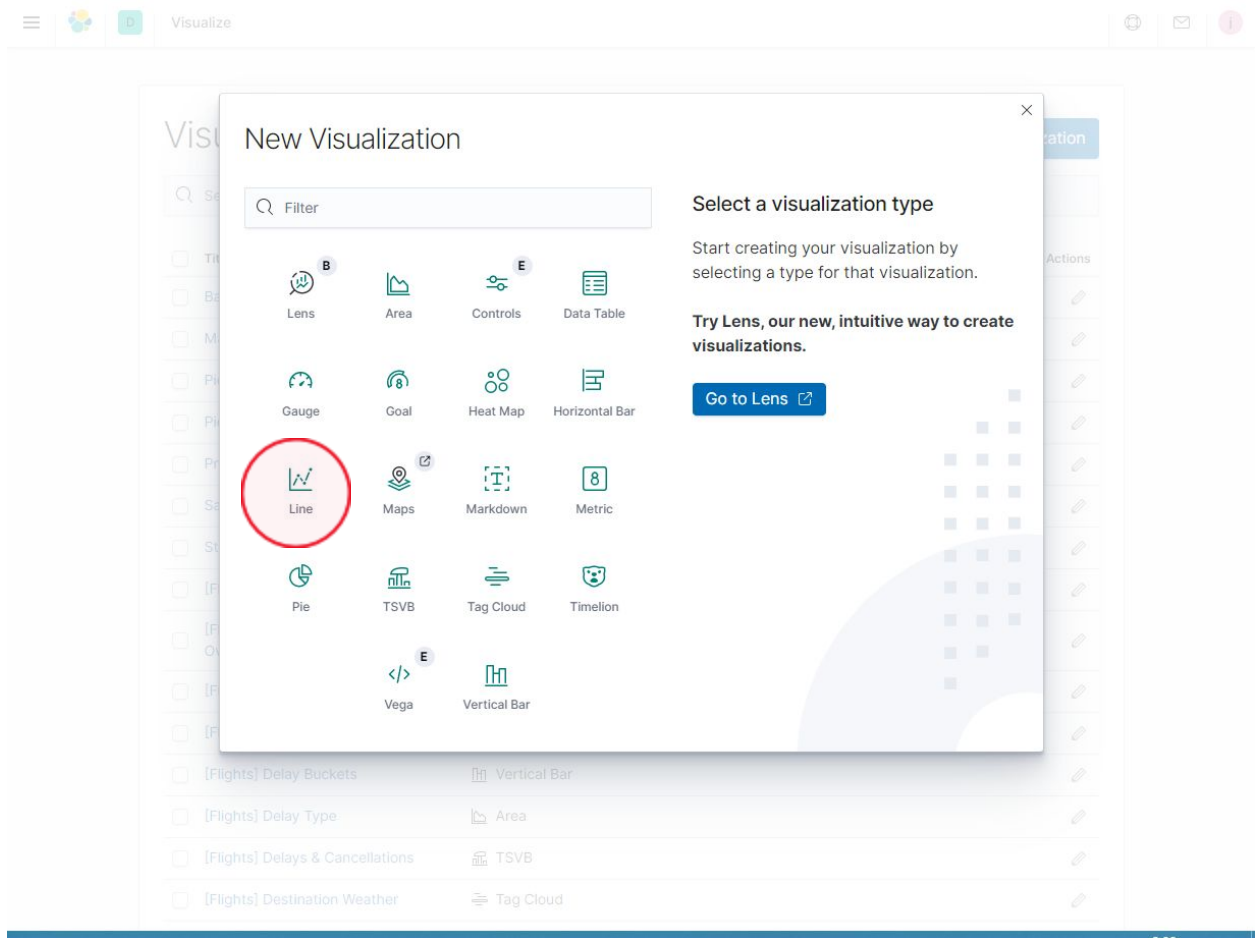31.Select the time range from Jan.1st 2014, 00:00:00, to Jan 1st 2018, 00:00:00. and then click the blue button named "Refresh".   Follow the left menu, you will be able to see this quarter report from the time range Jan 1. 2014 to Jan 1. 2018. Save it.
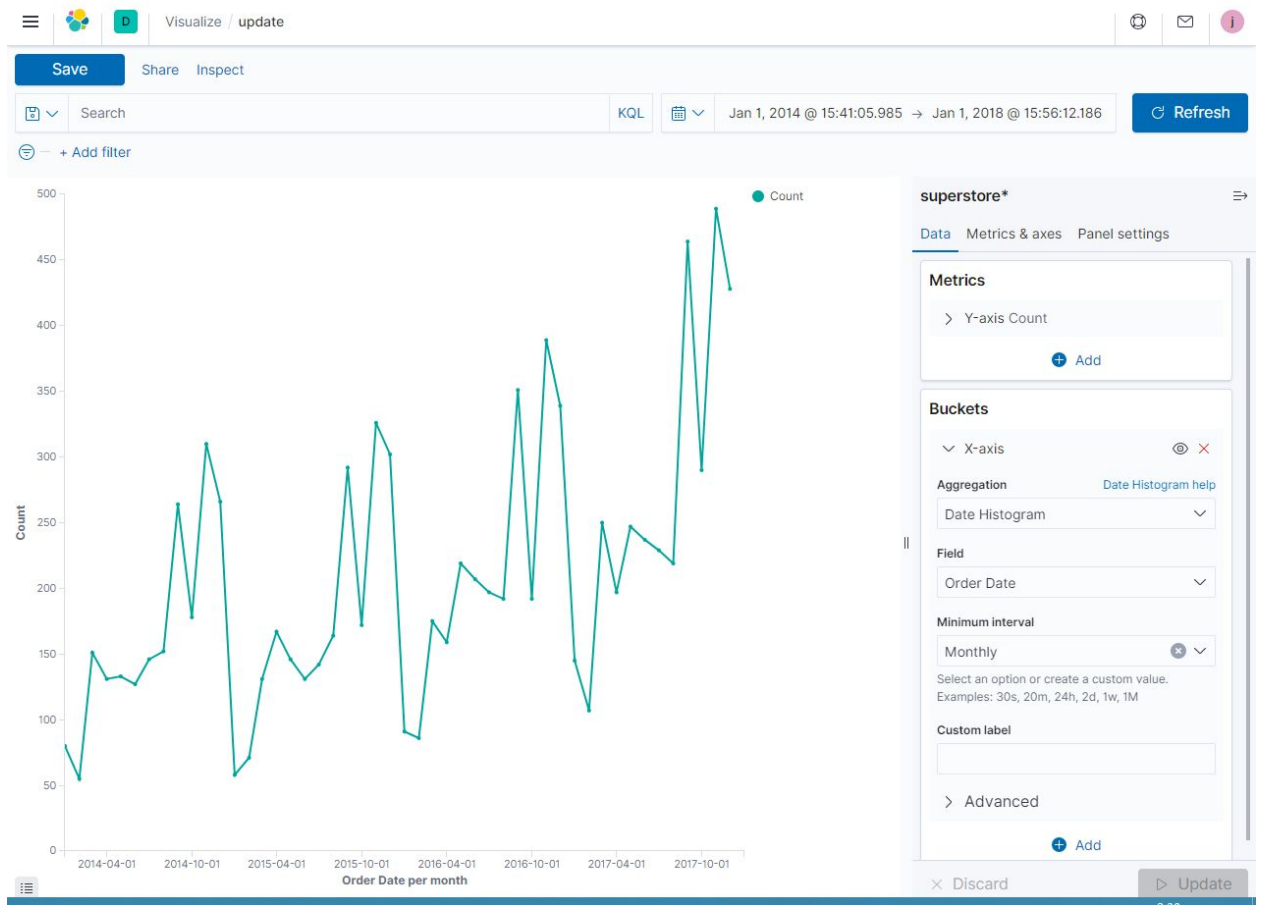
32.Click the left menu and click the dashboard.

33.Click the Create dashboard in the blue button.

Dashboards

<div>Create dashboard</div>

Search...

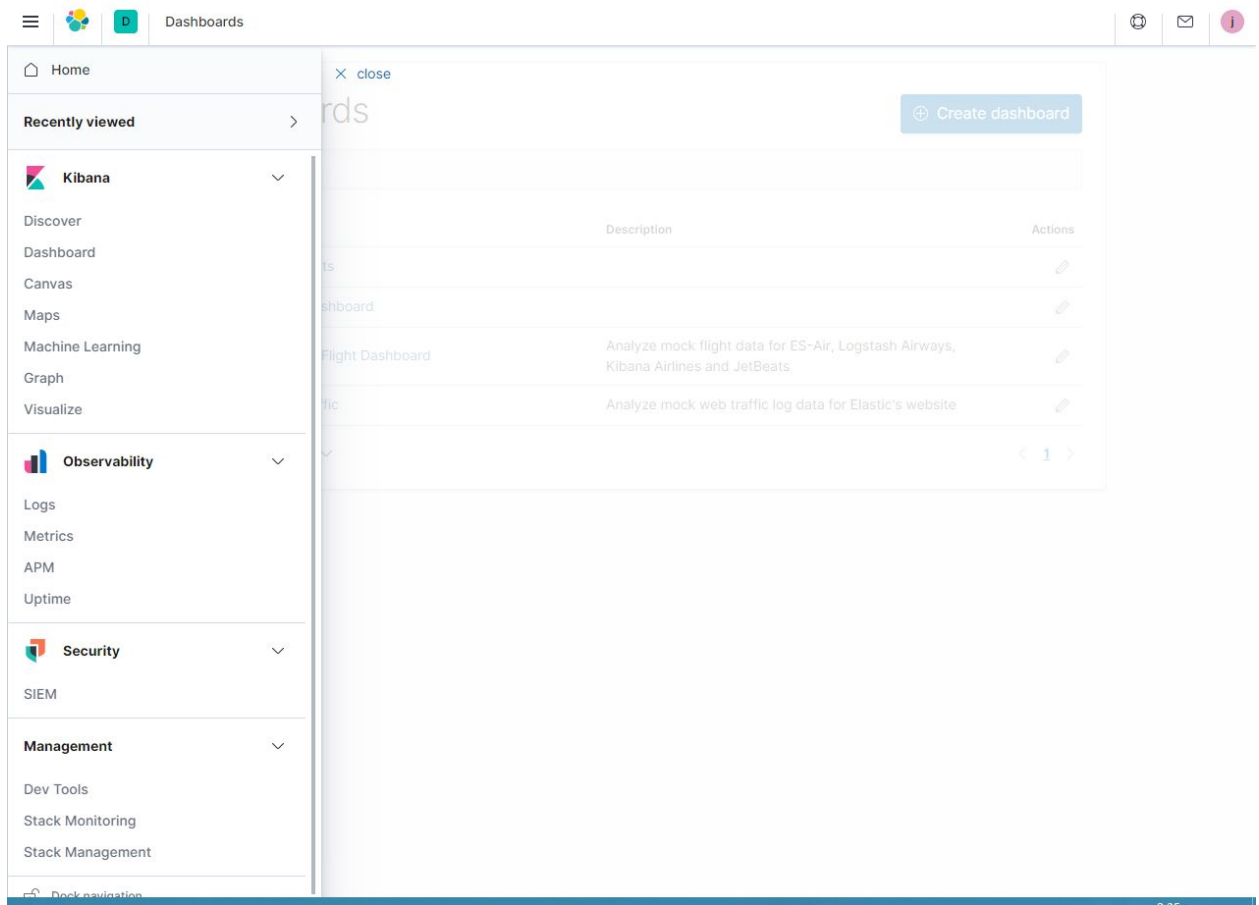|	Title	Description	Actions
☐	3Visualize-Charts		✎
☐	CIS3200FirstDashboard		✎
☐	[Flights] Global Flight Dashboard	Analyze mock flight data for ES-Air, Logstash Airways, Kibana Airlines and JetBeats	✎
☐	[Logs] Web Traffic	Analyze mock web traffic log data for Elastic's website	✎

Rows per page: 20 ⌄   ‹ 1 ›
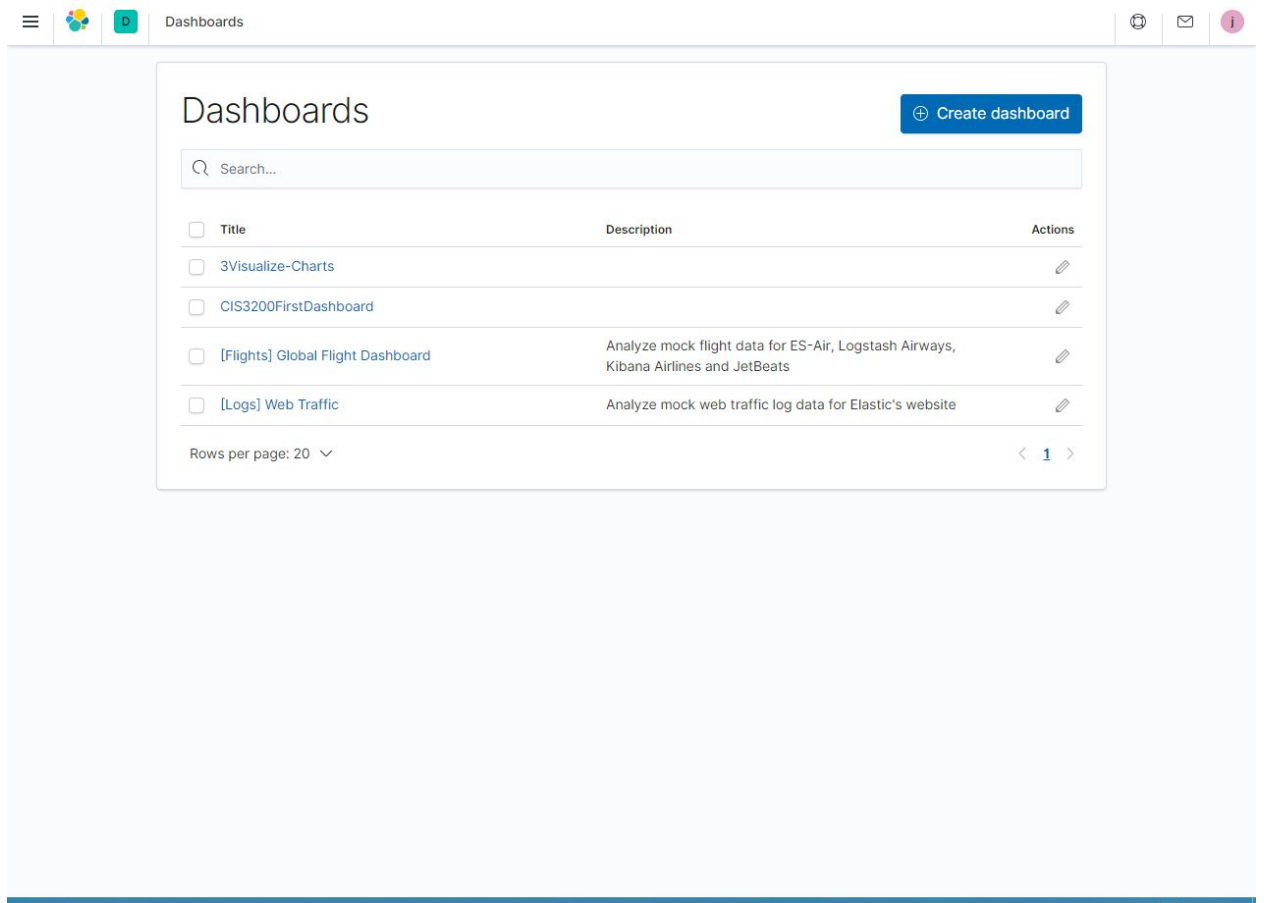
34. Then, Click the Add an existing option. You should be able to add all 4 we have done before. It would look like the one next page.

⊕ **Create new** Save Cancel Add Options Share

🗓 ⌄ Search KQL 🗓 ⌄ Jan 1, 2014 @ 00:00:00.00 → Jan 1, 2018 @ 00:00:00.00 ↻ **Refresh**

— + Add filter

State-Sold-Quantities ⚙

X-axis: State: Descending, Y-axis: Count (2,000 / 1,500 / 1,000 / 500 / 0). Legend: Count. States: California, New York, Texas, Pennsylvania, Washington, Illinois, Ohio, Florida, Michigan, North Carolina, Arizona, Virginia, Georgia, Tennessee, Colorado, Indiana, Kentucky, Massachusetts, New Jersey, Oregon

Pie-Sold ⚙

Legend: West, East, Central, South, Office Supplies, Furniture, Technology

Profit-final ⚙

X-axis: State: Descending (Rhode Island, Montana, District of Columbia, Wyoming), Y-axis: Average Profit (200 / 150 / 100 / 50 / 0). Legend: Average Profit

update ⚙

X-axis: Order Date per month (2015-01-01, 2016-01-01, 2017-01-01), Y-axis: Count (500 / 400 / 300 / 200 / 100 / 0). Legend: Count
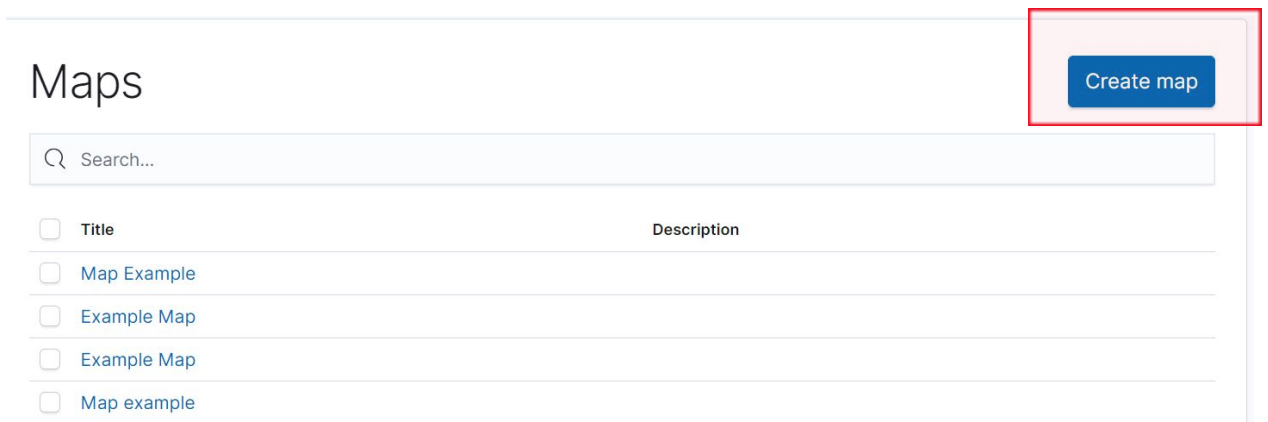
Create GEO Map Visualization

1. Access Kibana and select *Maps* on the left side scroll bar or go to *Visualize* to create a

   map visual.
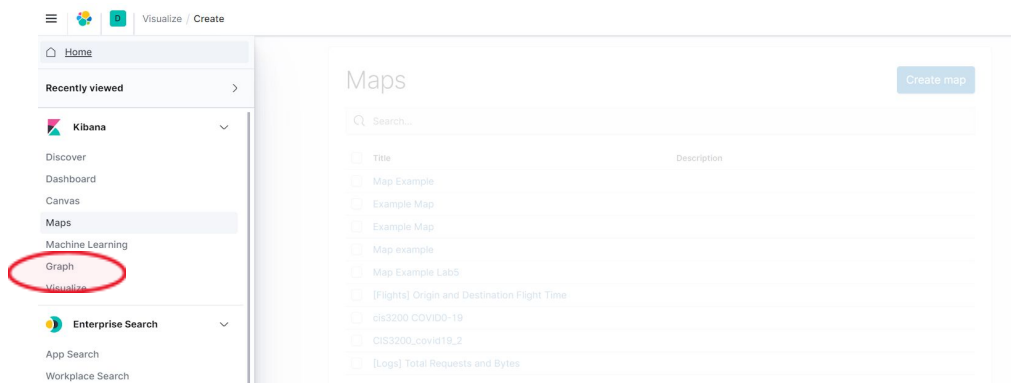
2.If creating a brand new map, select Create Map to start a brand new visualization

3. Set the time frame from 1/1/14 - 1/1/2018 to ensure all of the data within the dataset is accounted for. You want to make sure this is done by selection *Absolute* and manually entering the correct date range.
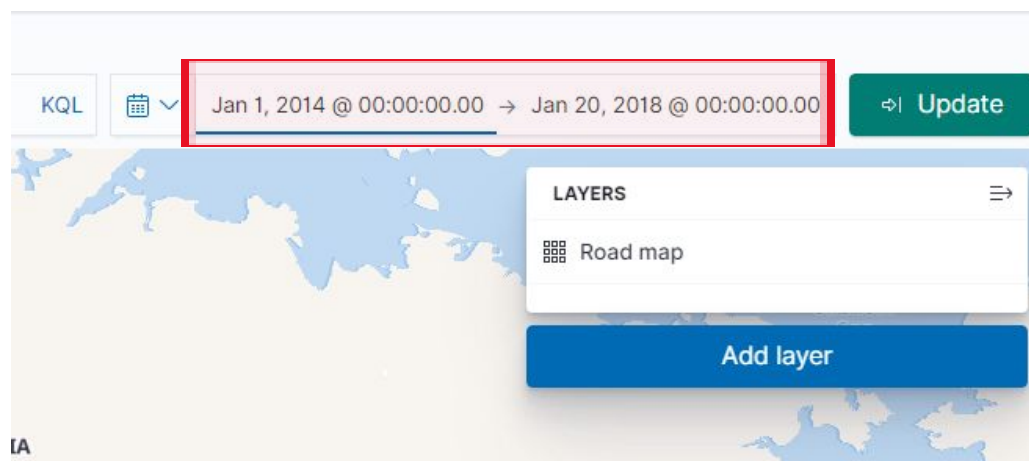
4. Select the Clusters and Grids map option from the menu on the right

Clusters and grids

Geospatial data grouped in grids with metrics for each gridded cell
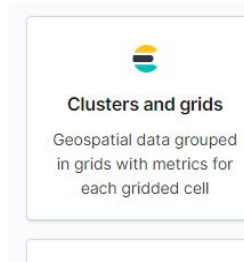
5. Select *Add Layer* and select the index you created in the previous steps. This particular index is named *superstoredata.* Once selected, make sure to click on *Add Layer*
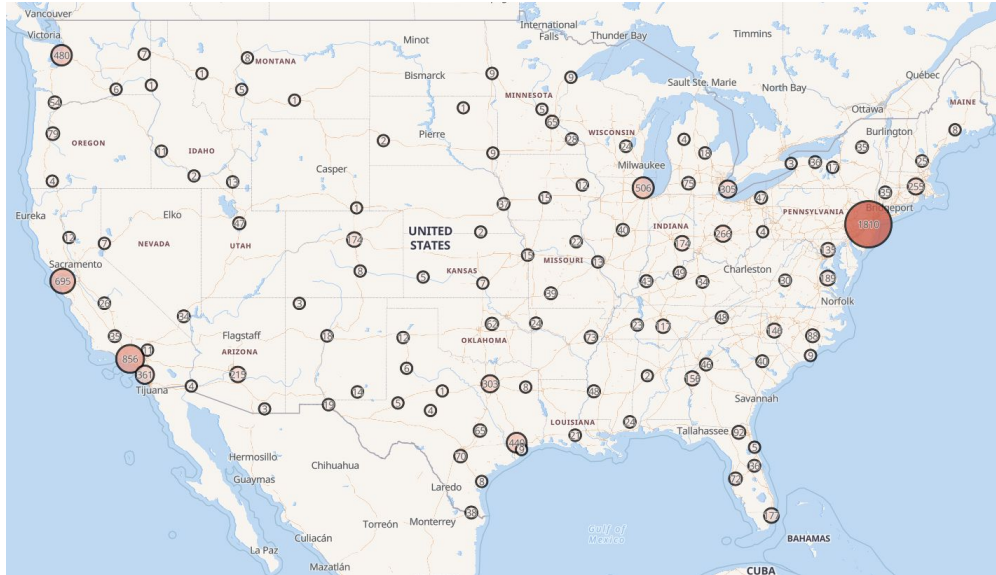
6. In this step we can add some changes to the *Layer Style,* some examples include changing the fill color gradient, border color, and border thickness.

7. Make sure you click *Save and Close* before exiting the layer. Your new geo spatial map is ready for viewing as seen below

References

---

1. URL of Data Source

https://community.tableau.com/s/question/0D54T00000CWeX8SAL/sample-superstore-sales-excelxls

2. URL of your Github

xEqualsToY/Kibana-Data-Set: Data Visuliztion (github.com)

3. URL of References

Elastic Cloud Tutorial: Getting Started with a sample dataset | Elastic Blog