



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:

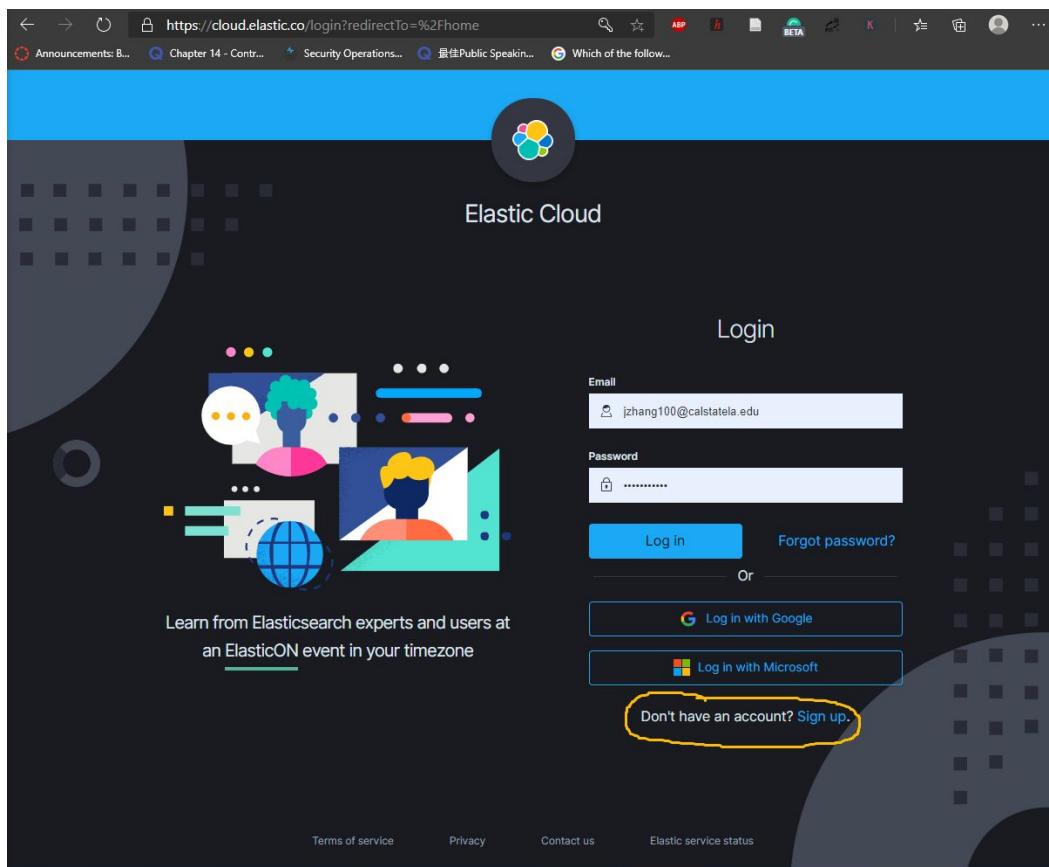
- Find clear data sets
- Visit and create an account for Elastic Cloud
- Implement data sets
- Create Simple Graphs Visualizations
- Create GEO_Map_Visualizations

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

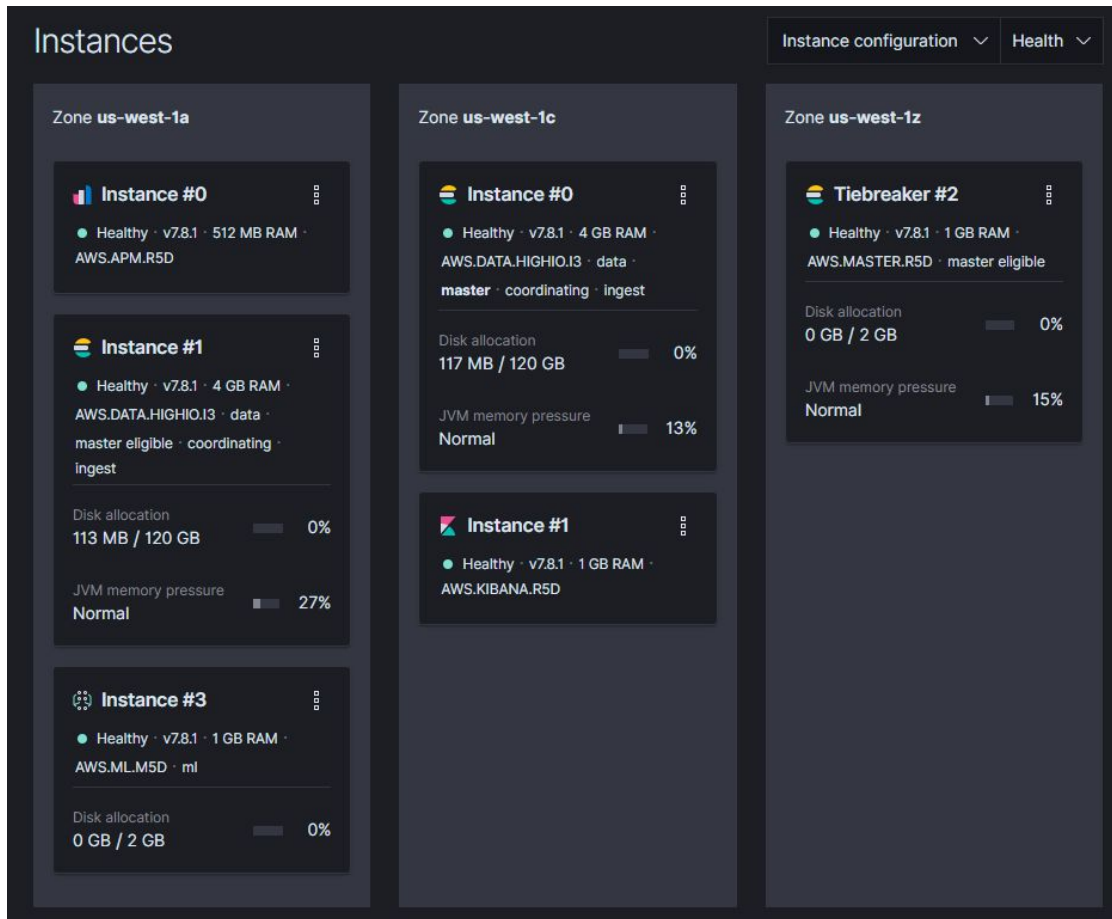
1. Go to <https://www.elastic.co/cloud/as-a-service>
2. Register then Log into your ES (Elastic Cloud) account by using your email account



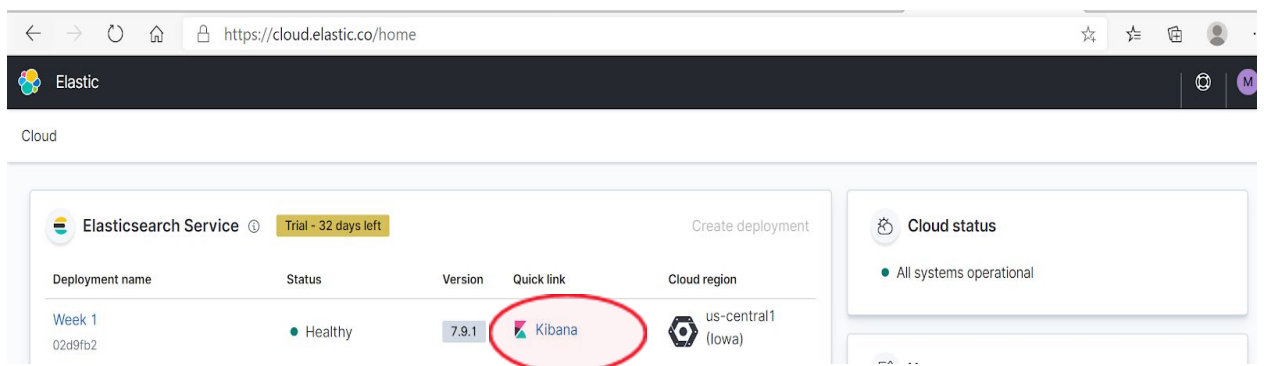
3. Click on the verification link in the email that was sent to your “.edu” address
4. After email verification, you will be prompted to create a password for your Elastic Cloud.
5. Log into your Elastic Cloud account
6. NOTE: If the verification email expires, go to <https://cloud.elastic.co/forgot> and enter the student email address to trigger a new verification email. Create your first hosted Elasticsearch cluster

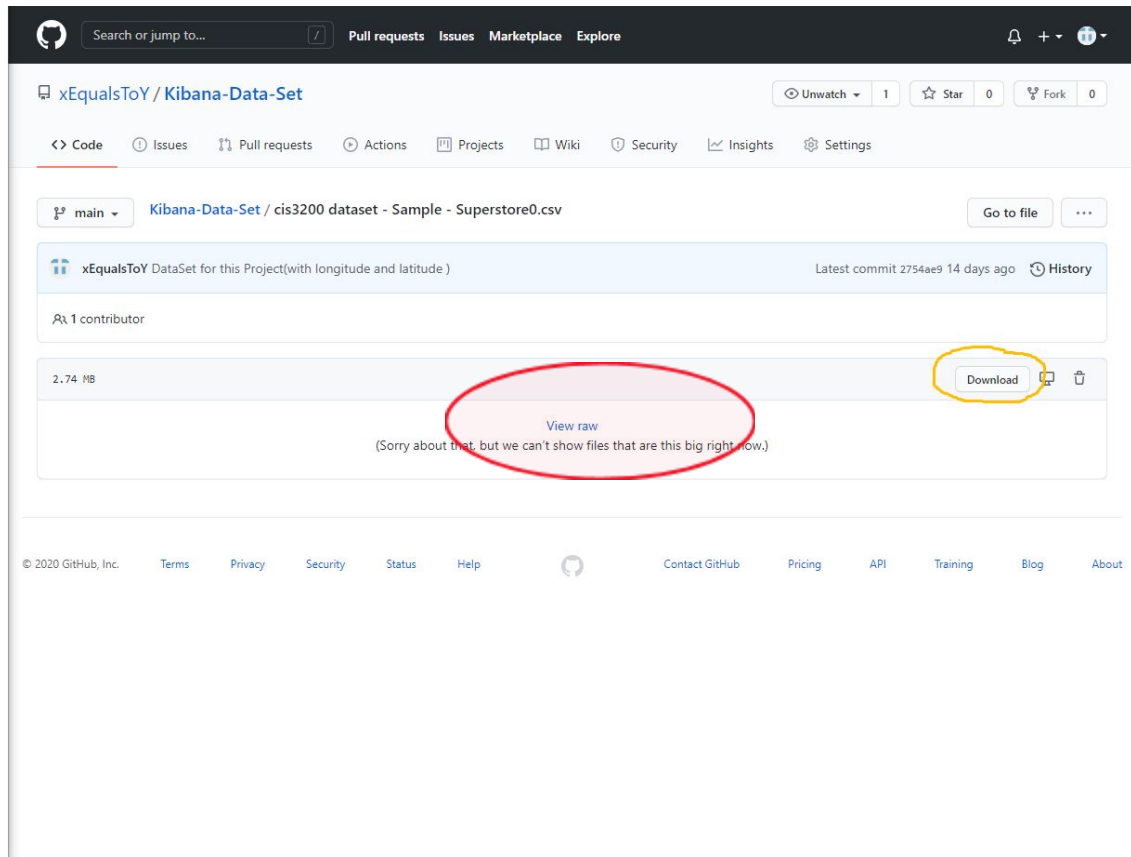
Set up the Instances

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 [Kibana-Data-Set/cis3200 dataset - Sample - Superstore0.csv at main · xEqualsToY/Kibana-Data-Set \(github.com\)](#) (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.

Machine Learning / Data Visualizer / File

Overview Anomaly Detection Data Frame Analytics **Data Visualizer** Settings

cis3200 dataset - Sample - Superstore.csv

File contents
First 1,000 lines

```
1 Row ID,Order ID,Order Date,Ship Date,Ship Mode,Customer ID,Customer Name,Segment,Country,City,State,Full Address,Latitude,Longitude,Postal Code,Region,Product ID,Category,Sub-Category,Product Name,Sales,Quantity,Discount,Profit
2 1,CA-2016-152156,11/8/2016,11/11/2016,Second Class,G-12528,Claire Gate,Consumer,United States,Henderson,Kentucky,United States Henderson Kentucky ,37.8361538,-87.5980134,42420,West,OFF-1A-10001798,Furniture,Bookcases,Bush Somerset Collection Bookcase,261,96,2,0,45,7336
3 2,CA-2016-152156,11/8/2016,11/11/2016,Second Class,G-12528,Claire Gate,Consumer,United States,Henderson,Kentucky,United States Henderson Kentucky ,37.8361538,-87.5980134,42420,West,OFF-1A-10001798,Furniture,Bookcases,Bush Somerset Collection Bookcase,261,96,2,0,45,7336
4 3,CA-2016-138088,6/12/2016,6/16/2016,Second Class,DV-13045,Darrell Van Huff,Corporate,United States,Los Angeles,California,United States Los Angeles California ,34.0522342,-118.2436849,90036,West,OFF-1A-10000240,Office Supplies,Labels,Se
5 4,US-2015-108966,10/11/2015,10/18/2015,Standard Class,SO-20335,Sean O'Donnell,Consumer,United States,Fort Lauderdale,Florida,United States Fort Lauderdale Florida ,26.1224386,-80.1373174,33311,West,OFF-1A-10000577,Furniture,Tables
6 5,US-2015-108966,10/11/2015,10/18/2015,Standard Class,SO-20335,Sean O'Donnell,Consumer,United States,Fort Lauderdale,Florida,United States Fort Lauderdale Florida ,26.1224386,-80.1373174,33311,West,OFF-1A-10000577,Furniture,Tables
7 6,CA-2014-115812,6/9/2014,6/14/2014,Standard Class,BH-11710,Brosina Hoffman,Consumer,United States,Los Angeles,California,United States Los Angeles California ,34.0522342,-118.2436849,90032,West,OFF-1A-10001487,Furniture,Furnishings
8 7,CA-2014-115812,6/9/2014,6/14/2014,Standard Class,BH-11710,Brosina Hoffman,Consumer,United States,Los Angeles,California,United States Los Angeles California ,34.0522342,-118.2436849,90032,West,OFF-1A-10001487,Furniture,Furnishings
```

Summary

Number of lines analyzed 1000

Format delimited

Delimiter ,

Has header row true

Time field Order Date

Time formats MM/dd/yyyy, M/dd/yyyy, MM/dd/yyyy, M/d/yyyy

[Override settings](#) [Analysis explanation](#)

File stats

Row ID	Order ID	Order Date	Ship Date
2,000 documents (100%)	2,000 documents (100%)	2,000 documents (100%)	2,000 documents (100%)

Import Cancel

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

Superstore.csv

Import data **Advanced**

Simple **Advanced**

Index name
Index name

☒ Create index pattern
Index pattern name

Index settings

```
1 {
2   "number_of_shards": 1
3 }
```

Mappings

```
1 {
2   "mappings": {
3     "type": "geo"
4   }
5   "category": {
6     "type": "keyword"
7   }
8   "city": {
9     "type": "keyword"
10  }
11  "country": {
12    "type": "keyword"
13  }
14  "customer_id": {
15    "type": "keyword"
16  }
17  "customer_name": {
18    "type": "keyword"
19  }
20  "discount": {
21    "type": "double"
22  }
23 }
```

Import pipeline

```
1 {
2   "description": "Import pipeline created by file structure folder",
3   "processors": {
4     {
5       "type": "message",
6       "target_field": {
7         "name": "message",
8         "order": 1,
9         "type": "text",
10        "format": "MM/dd/yyyy | M/dd/yyyy | M/d/yyyy | MM/d/yyyy"
11      },
12      "type": "text",
13      "order": 2,
14      "type": "text",
15      "order": 3,
16      "type": "text",
17      "order": 4,
18      "type": "text",
19      "order": 5,
20      "type": "text",
21      "order": 6
22    }
23  }
```

Import

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

```
61 |   "type": "double"
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

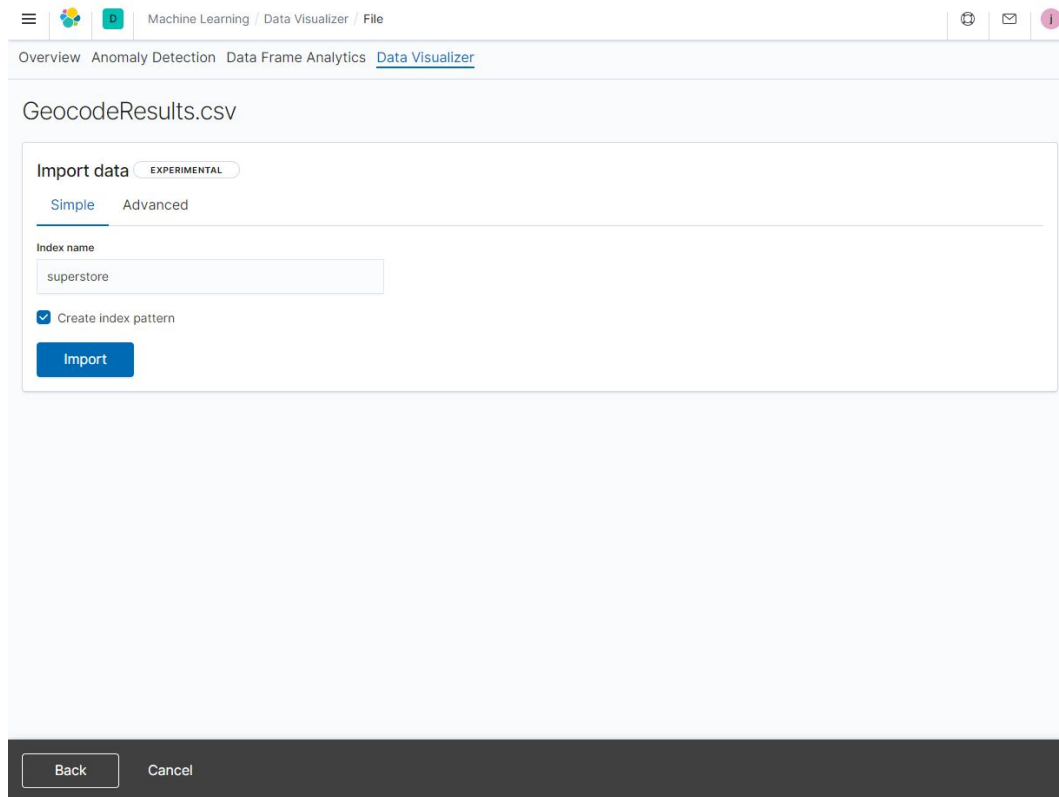
```
34 |   }
35 | },
36 | {
37 |   "date": {
38 |     "field": "Order Date",
39 |     "timezone": "{{ event.timezone }}",
40 |     "formats": [
41 |       "MM/d/yyyy",
42 |       "M/dd/yyyy",
43 |       "MM/dd/yyyy",
44 |       "M/d/yyyy"
45 |     ]
46 |   },
47 | },
48 | },
49 | "append":{
50 |   "field":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

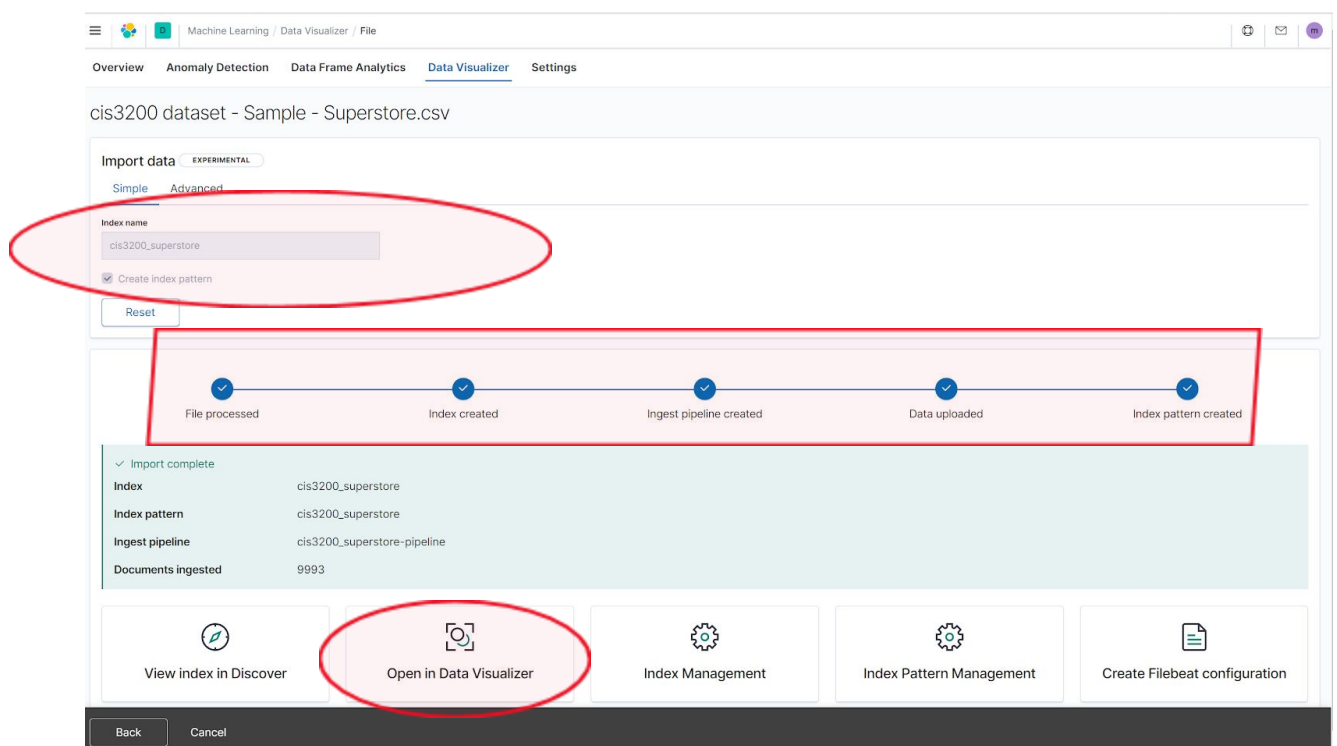
One thing to pay attention to here, make sure to check the “create index pattern” box

First!

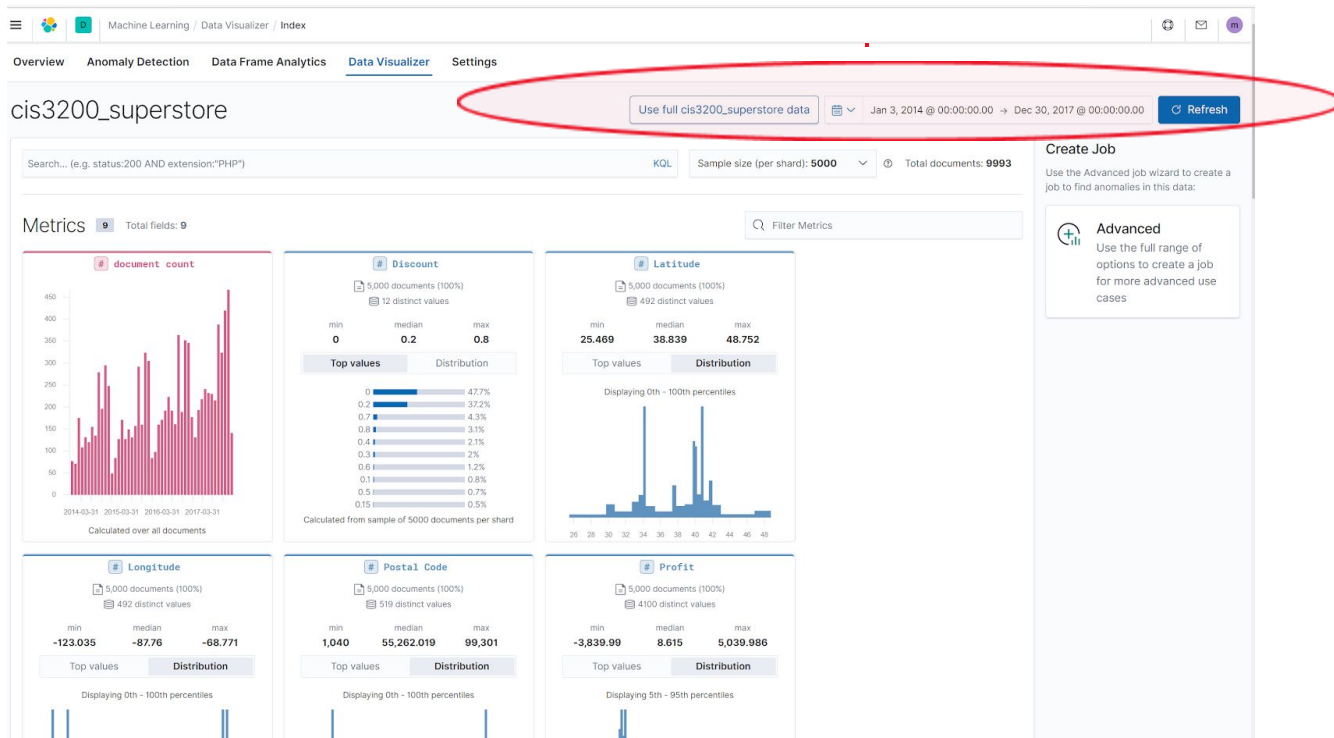
After that, please click import in the red circle. Then, you will see the following pipeline as the picture below which shows the imported process with `geo_point`. If there is no error, the process is completed.



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.

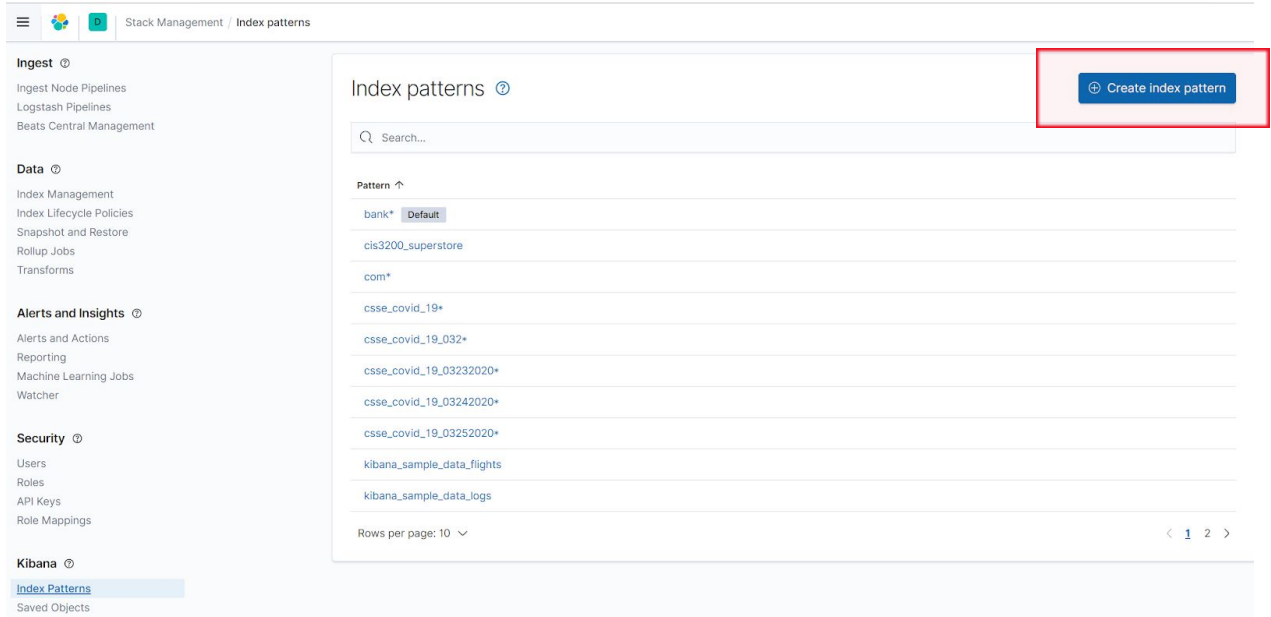


17. Next, click on the “Index Management” under the “Data” tab. It will take you to the step where you can create Index Management.

The screenshot shows the 'Index Management' interface. The left sidebar has a 'Data' tab selected, with 'Index Management' highlighted. The main content area shows the 'Index Management' section with tabs for 'Indices', 'Data Streams', 'Index Templates', and 'Component Templates'. The 'Indices' tab is active. Below the tabs, there is a search bar and a table of indices. A red circle highlights the 'cis3200_superstore' index in the table.

Name	Health	Status	Primaries	Replicas	Docs count	Storage size	Data stream
cis3200_superstore	green	open	1	1	9993	7.4mb	

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:



Step 1 of 2: Define index pattern

Index pattern name

cis3200_superstore*

Use an asterisk (*) to match multiple indices. Spaces and the characters \, /, ?, ", <, >, | are not allowed.

☐ Include system and hidden indices

✓ Your index pattern matches 1 source.

cis3200_superstore	Index
--------------------	-------

Rows per page: 10 ▾

Next step >

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.

Step 2 of 2: Configure settings

cis3200_superstore*

Select a primary time field for use with the global time filter.

Time field Refresh

Order Date ▼

> Show advanced options

< Back

Create index pattern

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

cis3200_superstore*

★ ↺ 🗑️

Time Filter field name: 'Order Date'

This page lists every field in the **cis3200_superstore*** index and the field's associated core type as recorded by Elasticsearch. To change a field type, use the Elasticsearch [Mapping API](#)

Fields (30) Scripted fields (0) Source filters (0)

🔍 Search

All field types ▼

Name	Type	Format	Searchable	Aggregatable	Excluded
@timestamp	date		●	●	✎
Category	string		●	●	✎
City	string		●	●	✎
Country	string		●	●	✎
Customer ID	string		●	●	✎
Customer Name	string		●	●	✎
Discount	number		●	●	✎
Full Address	string		●		✎
Latitude	number		●	●	✎
Longitude	number		●	●	✎

Rows per page: 10 ▼

< 1 2 3 >

● Create Simple Graphs Visualizations

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

The screenshot shows the Kibana interface. In the left sidebar, the 'Visualize' option is highlighted with a red circle. The main content area displays the 'cis3200_superstore*' visualization page. The page includes a search bar, a list of fields (30), scripted fields (0), and source filters (0). The fields listed are: @timestamp, Category, City, Country, Customer ID, Customer Name, Discount, Full Address, Latitude, Longitude, and Rows per page: 10.

Stack Management / Index patterns / cis3200_superstore*

Home

Recently viewed >

Kibana ▾

- Discover
- Dashboard
- Canvas
- Maps
- Machine Learning
- Graph
- Visualize**

Enterprise Search ▾

- App Search
- Workplace Search

Observability ▾

- Overview
- Logs
- Metrics
- APM
- Uptime

Security ▾

- Overview
- Detections

cis3200_superstore*

Time Filter field name: 'Order Date'

This page lists every field in the cis3200_superstore* in use the Elasticsearch Mapping API

Fields (30) Scripted fields (0) Source filters (0)

Search

Name

@timestamp

Category

City

Country

Customer ID

Customer Name

Discount

Full Address

Latitude

Longitude

Rows per page: 10 ▾

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









Visualizations

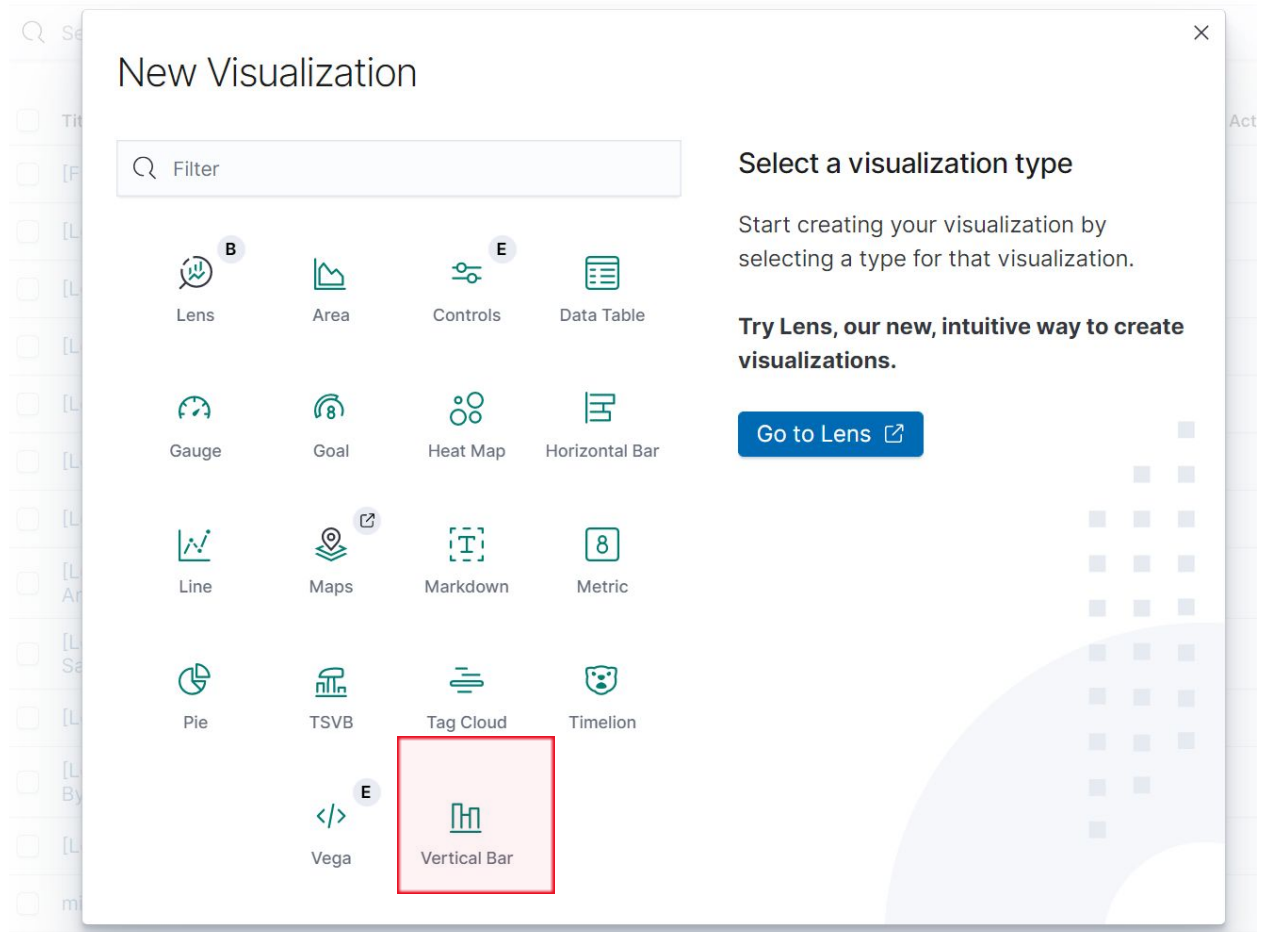
+

Create visualization

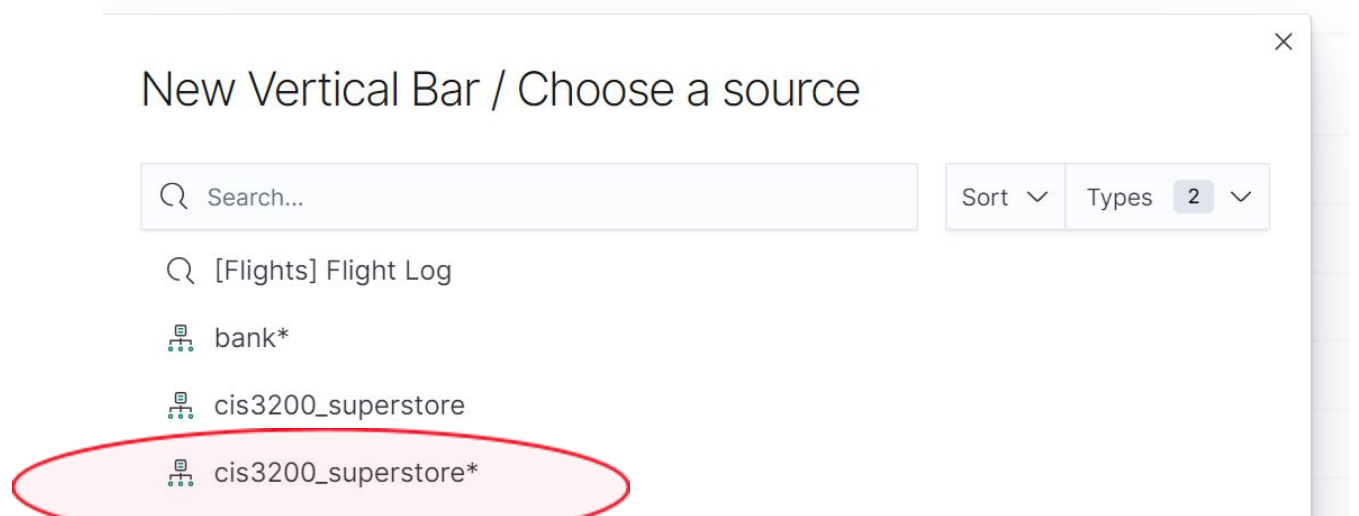
Q

Search...

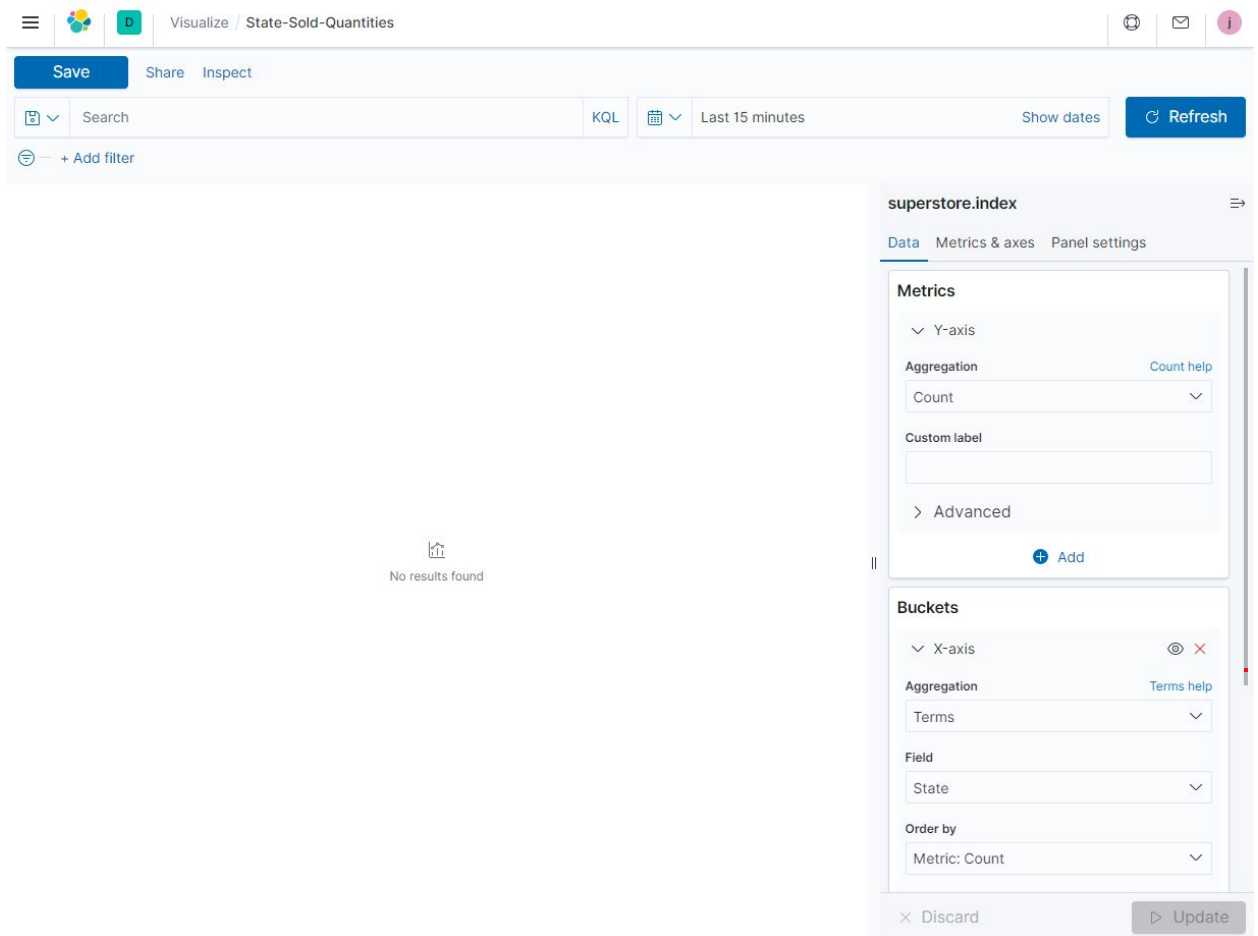
<input type="checkbox"/> Title	Type	Description	Actions
<input type="checkbox"/> BAR CHART LAB3	 Vertical Bar		
<input type="checkbox"/> Bar Example	 Vertical Bar		
<input type="checkbox"/> Markdown Example	 Markdown		
<input type="checkbox"/> Pie Example	 Pie		



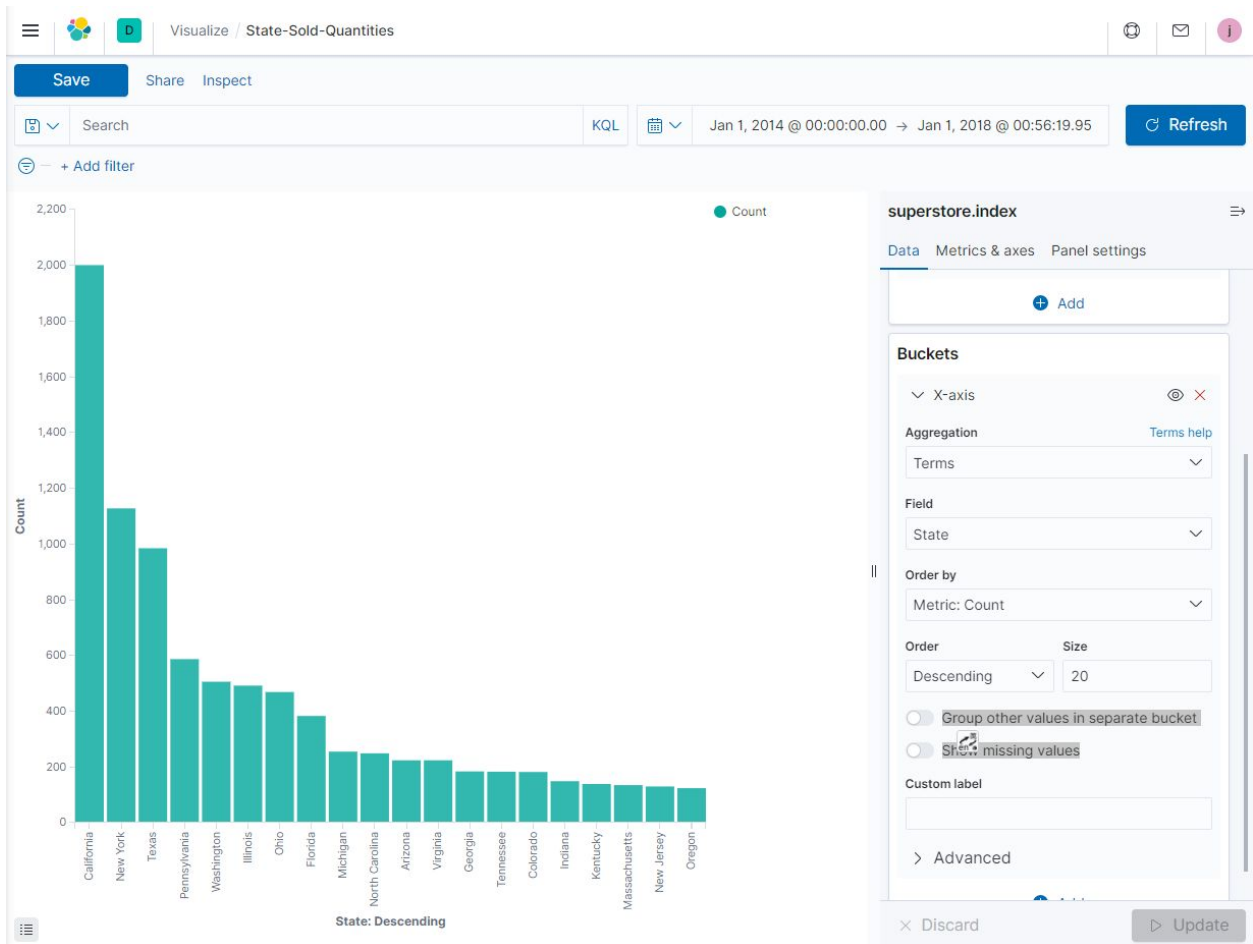
21. At the “New Vertical Bar,” click on the index pattern name which you just created in the previous step, “cis3200_superstore*,” which is the one we should click to start creating the bar chart:



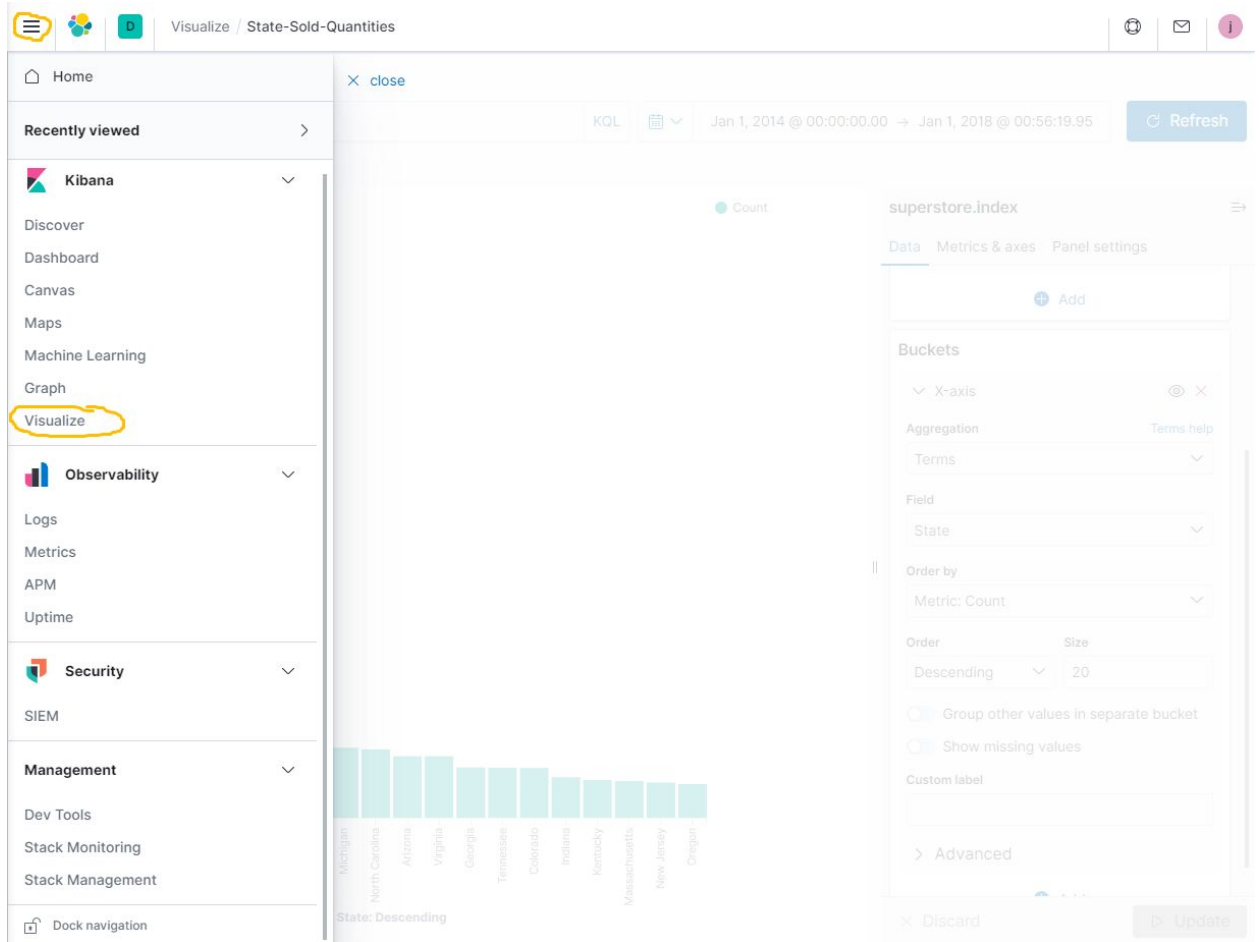
22.Next, you will need to set the time range, please select from Jan.1st 2014, 00:00:00, to Jan 1st 2018, 00:00:00. and then click the blue button named “Refresh”.



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!**



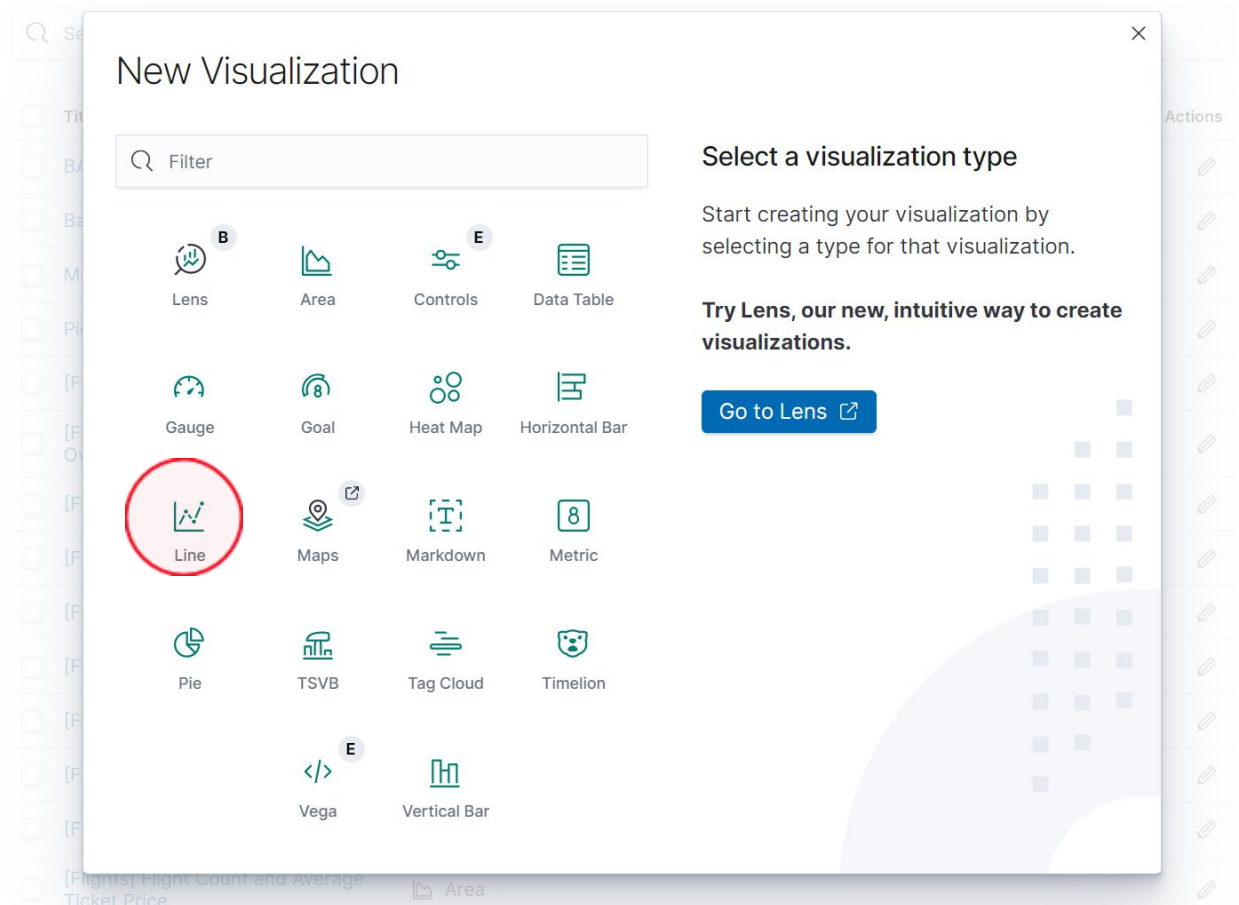
24. Click the left menu and click to visualize. We will do a line chart for this step.(see the yellow circle)



25. Then select Line in the red circle.

Visualizations

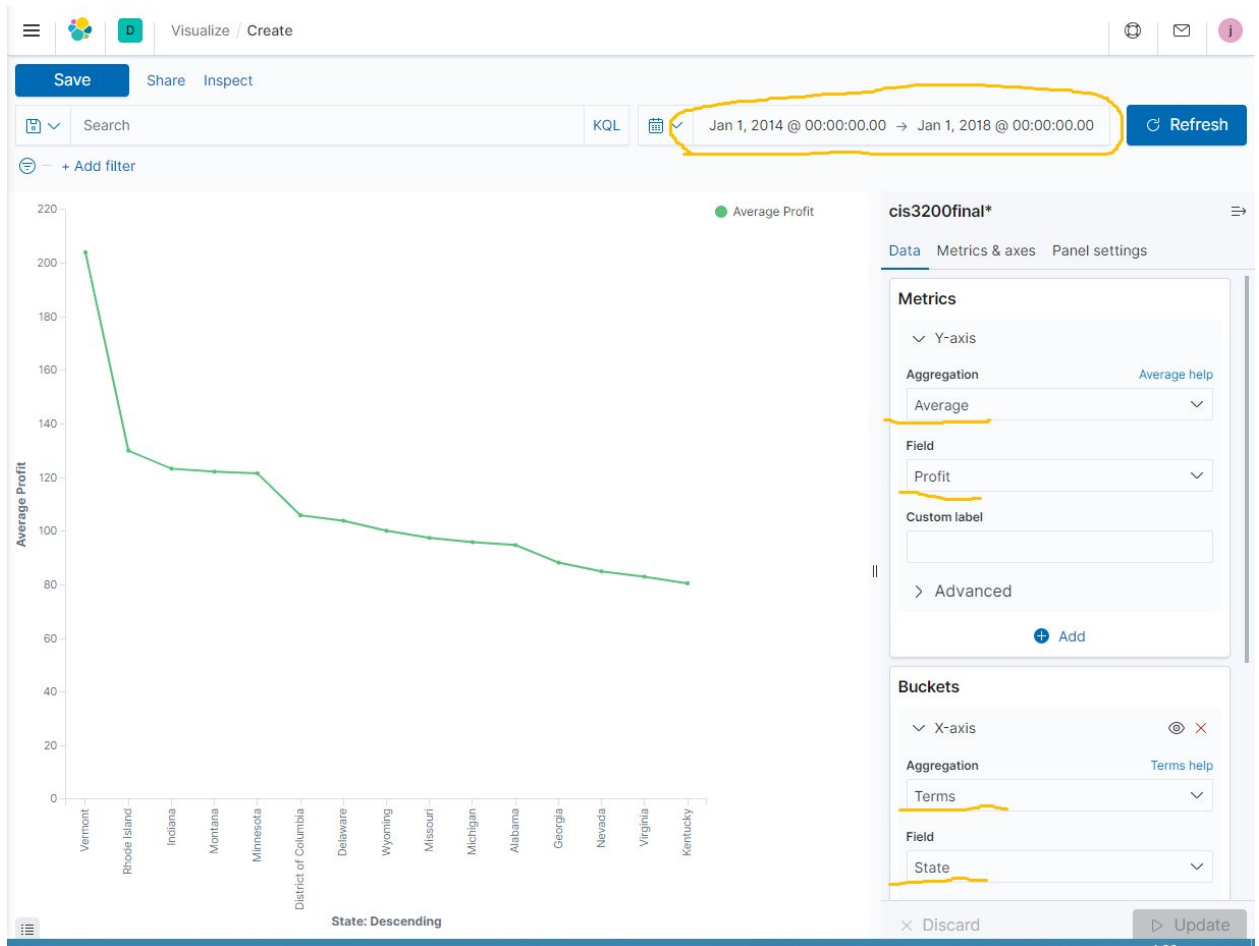
[+ Create visualization](#)



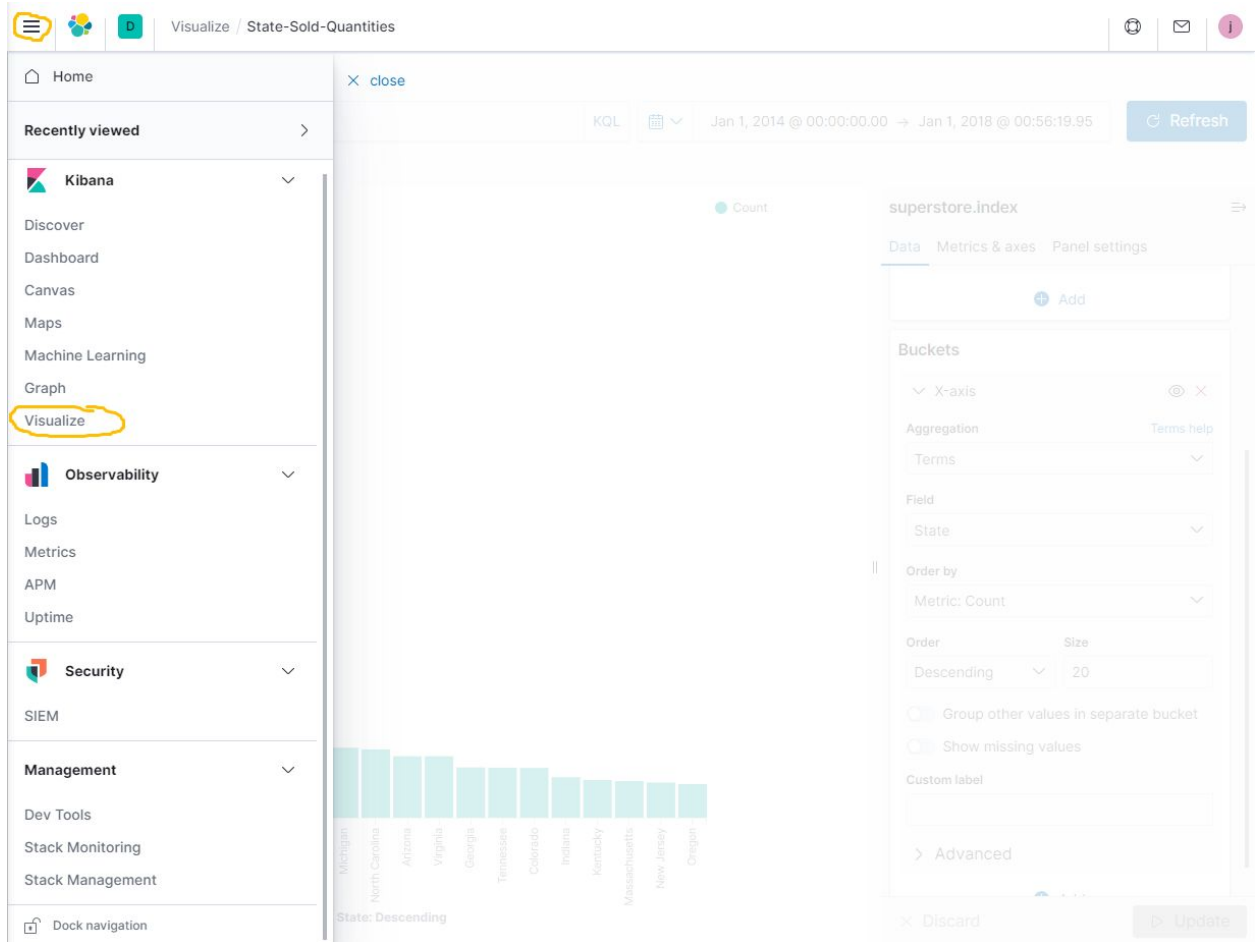
26. Select the time range from Jan. 1st 2014, 00:00:00, to Jan 1st 2018, 00:00:00. and then click the blue button named “Refresh”. Aggregation choose Average, Field “profit”.

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.



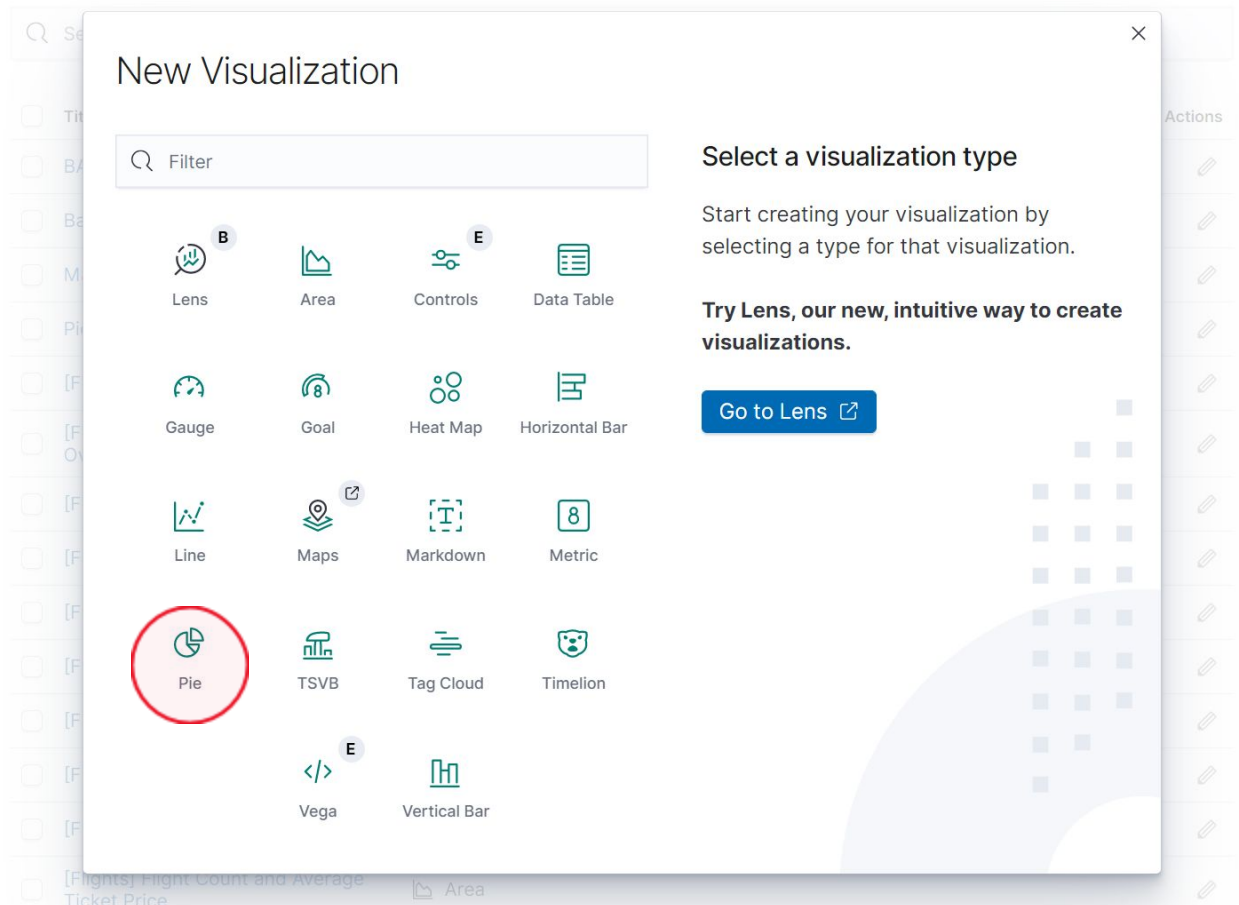
26. Click the left menu and click to visualize. We will do a pie chart for this step. (see the yellow circle)



27. Then select the Pie In the red circle.

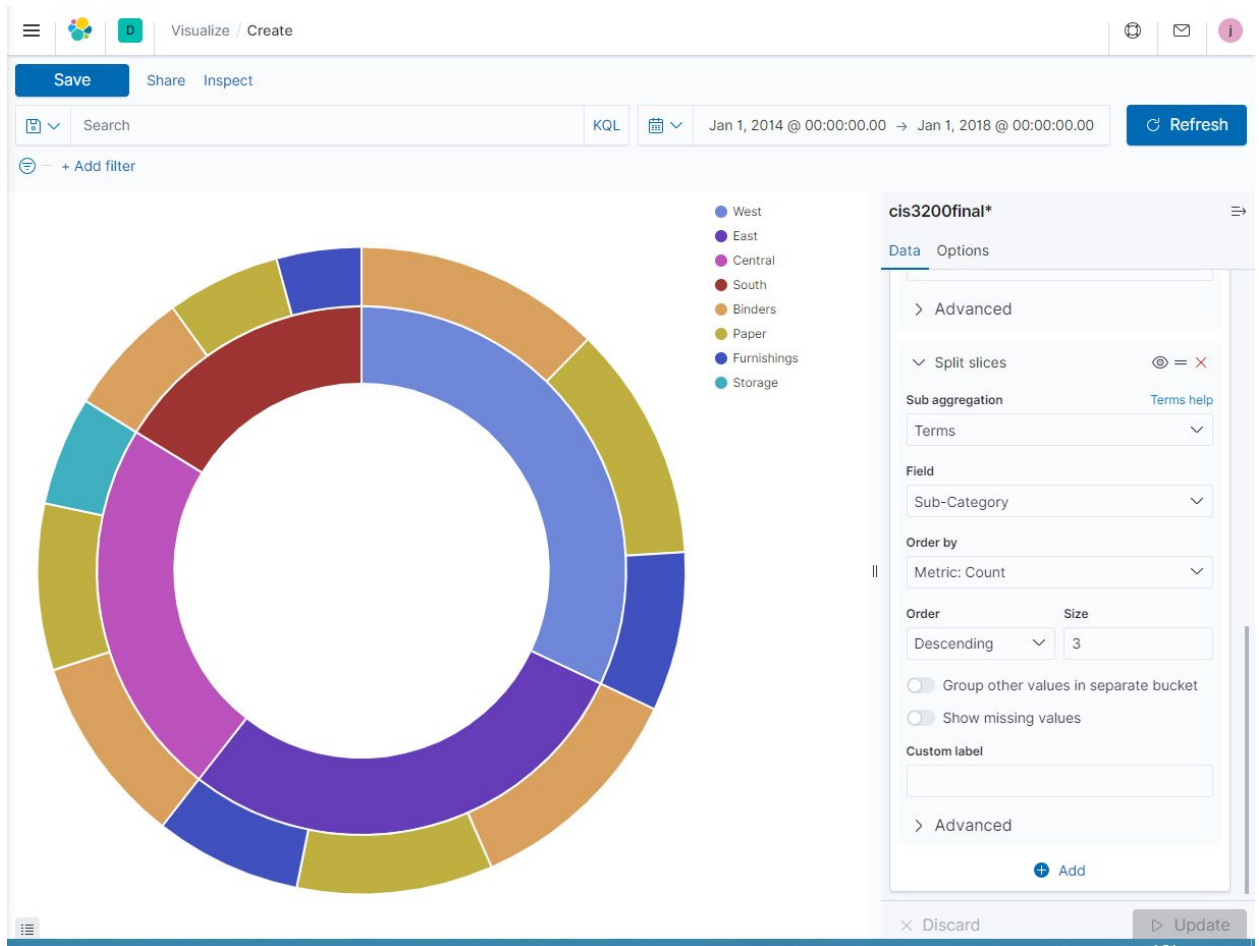
Visualizations

[+ Create visualization](#)



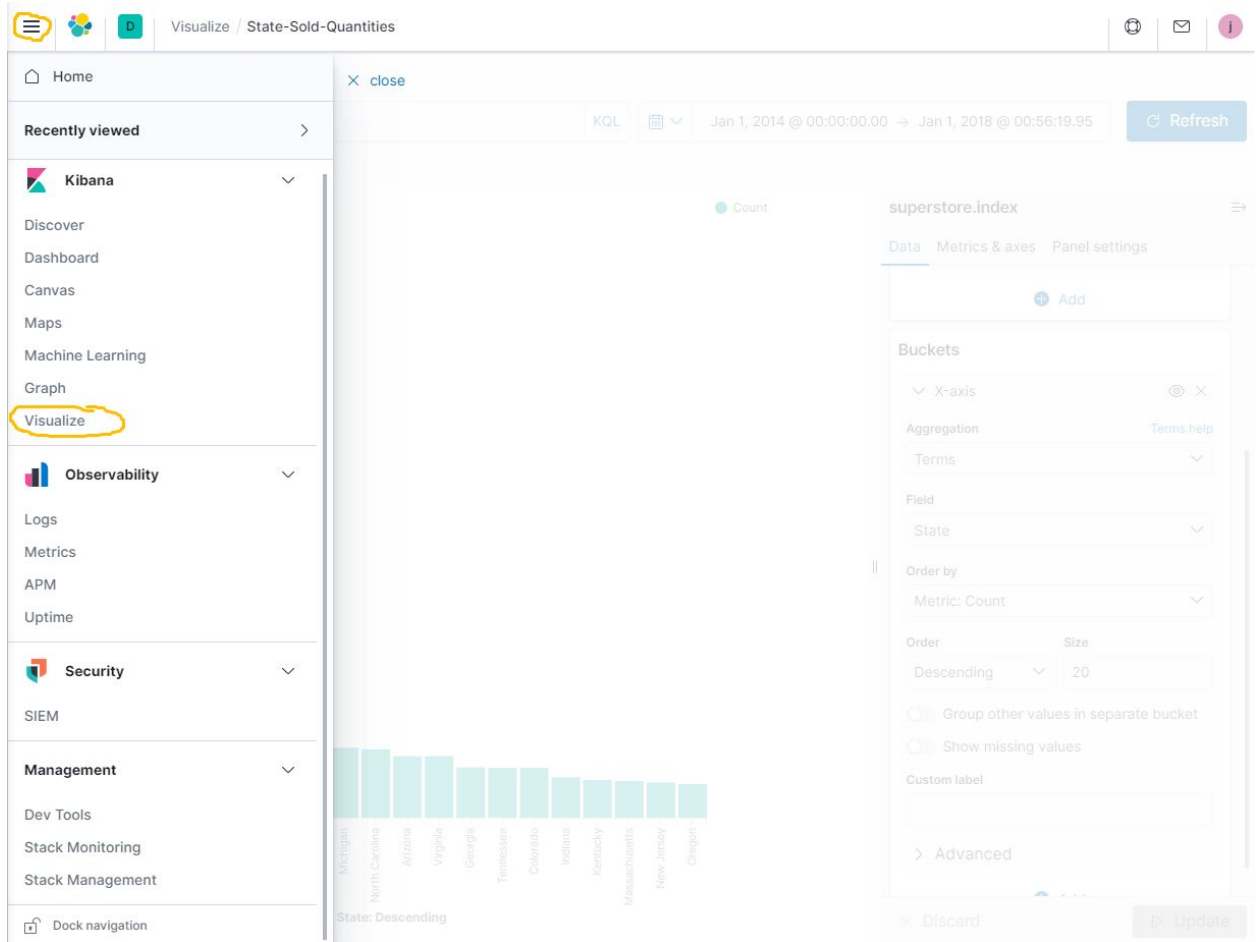
28. Select the time range from Jan. 1st 2014, 00:00:00, to Jan 1st 2018, 00:00:00. and then click the blue button named “Refresh”.

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.**

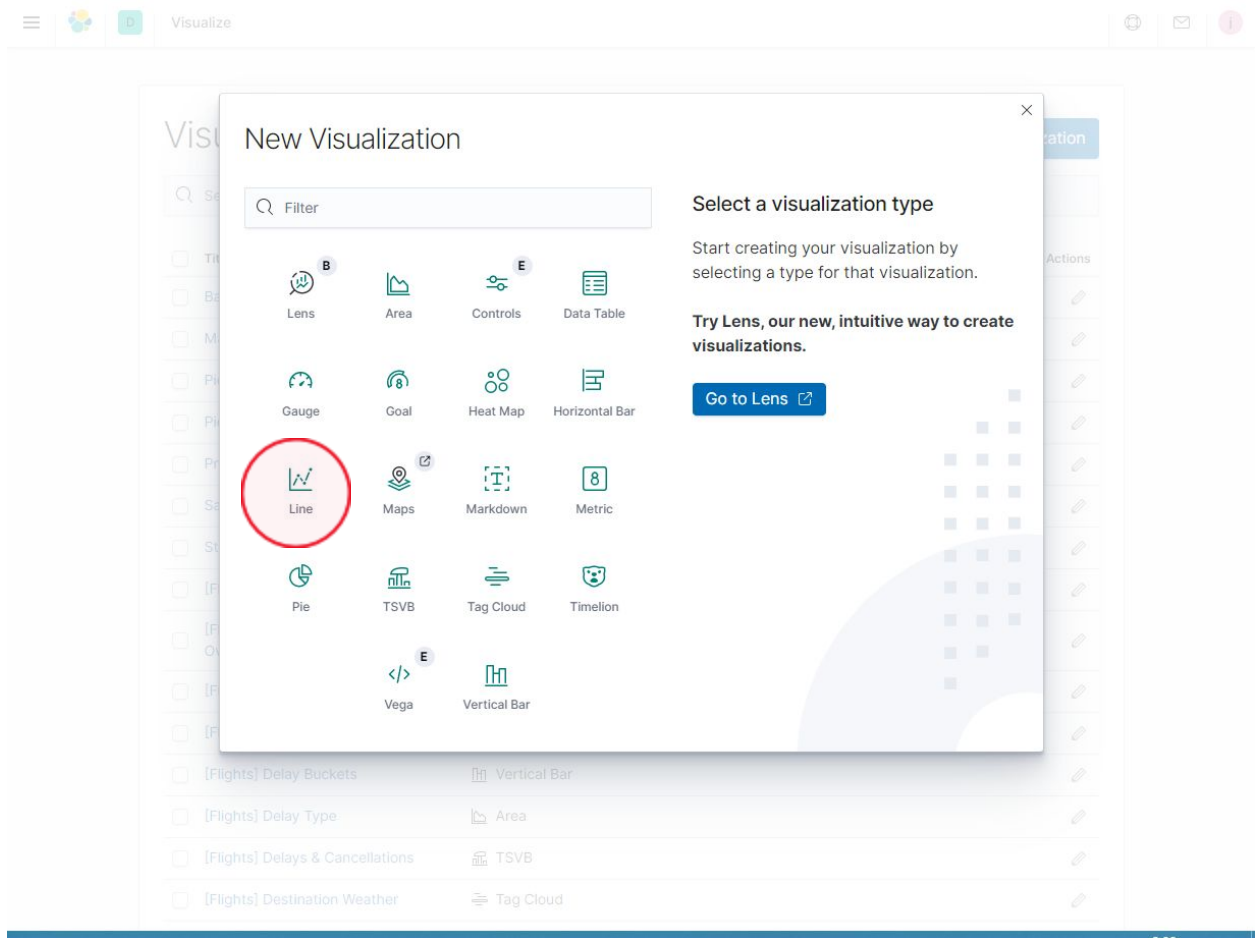


29. Next we will do a quarter sales report at a 4 years' range

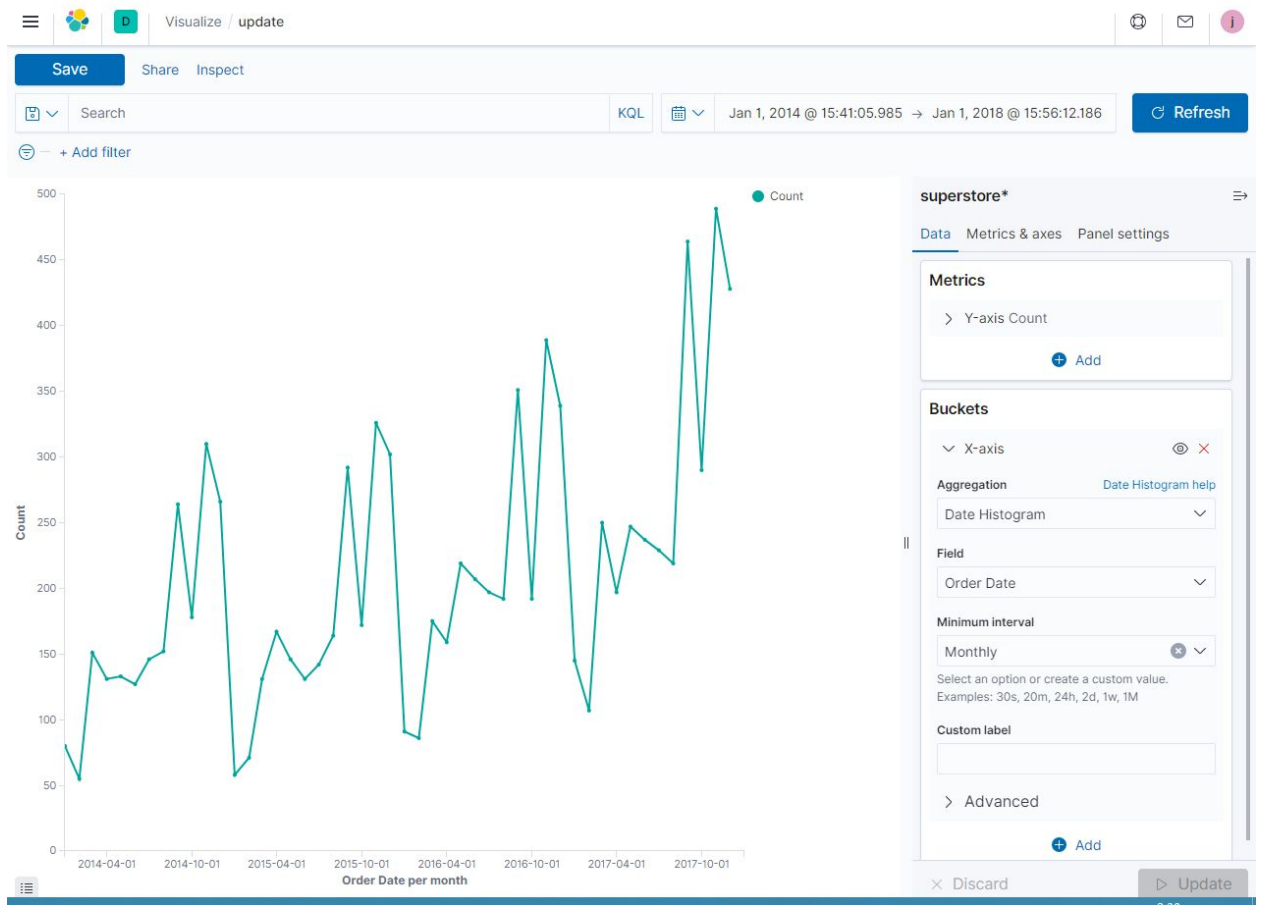
Click the left menu and click to visualize.



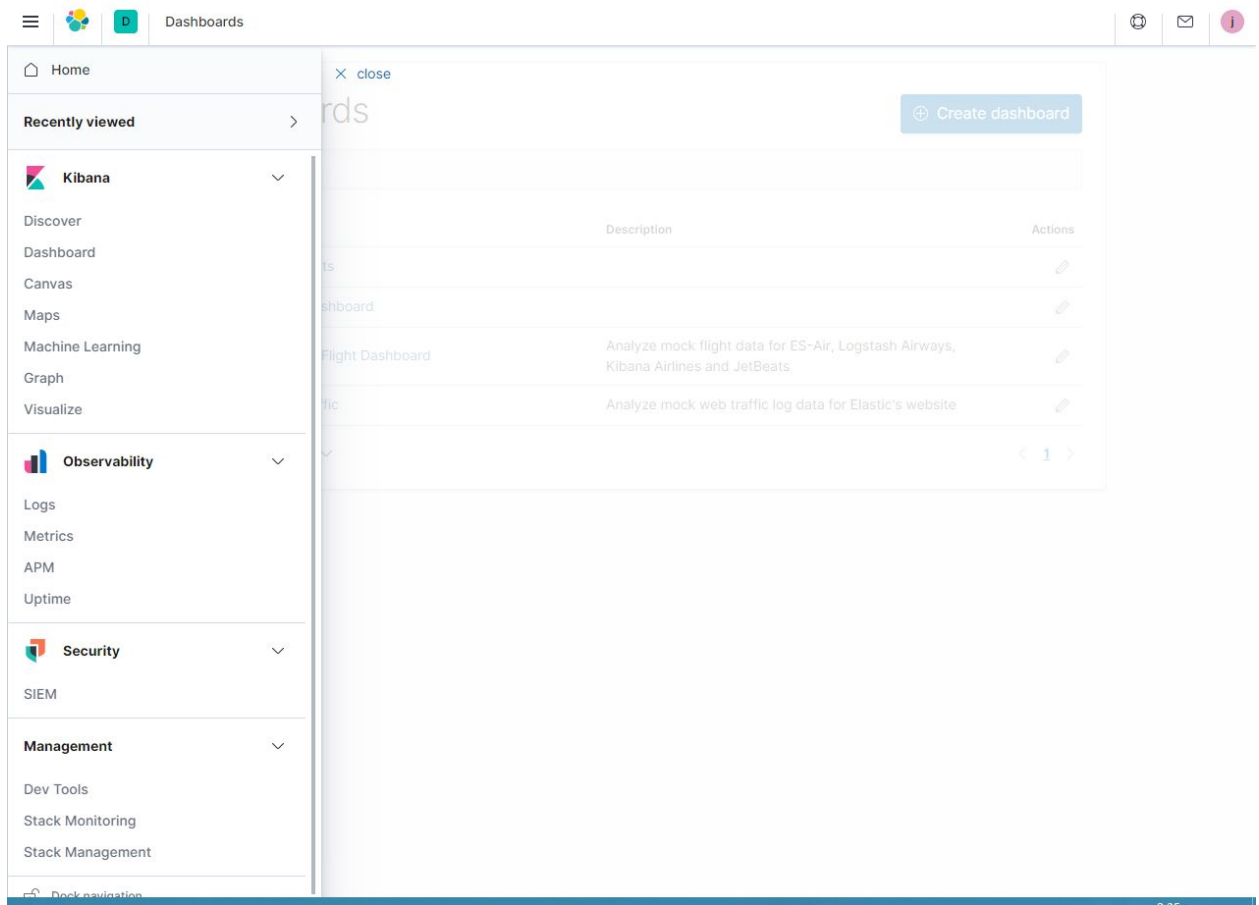
30. Select Line graph in red circle.



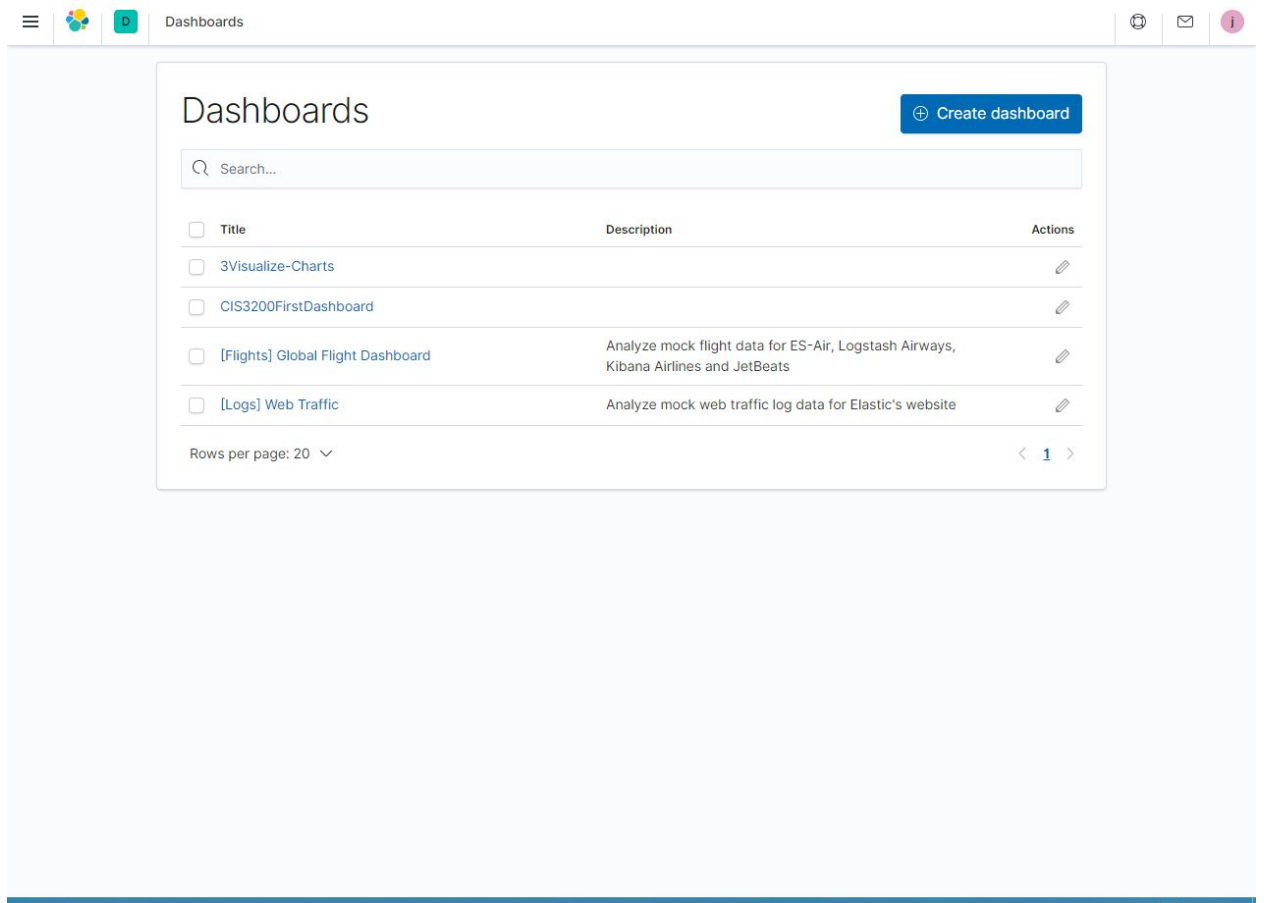
31. Select the time range from Jan. 1st 2014, 00:00:00, to Jan 1st 2018, 00:00:00. and then click the blue button named “Refresh”. Follow the left menu, you will be able to see this quarter report from the time range Jan 1. 2014 to Jan 1. 2018. Save it.



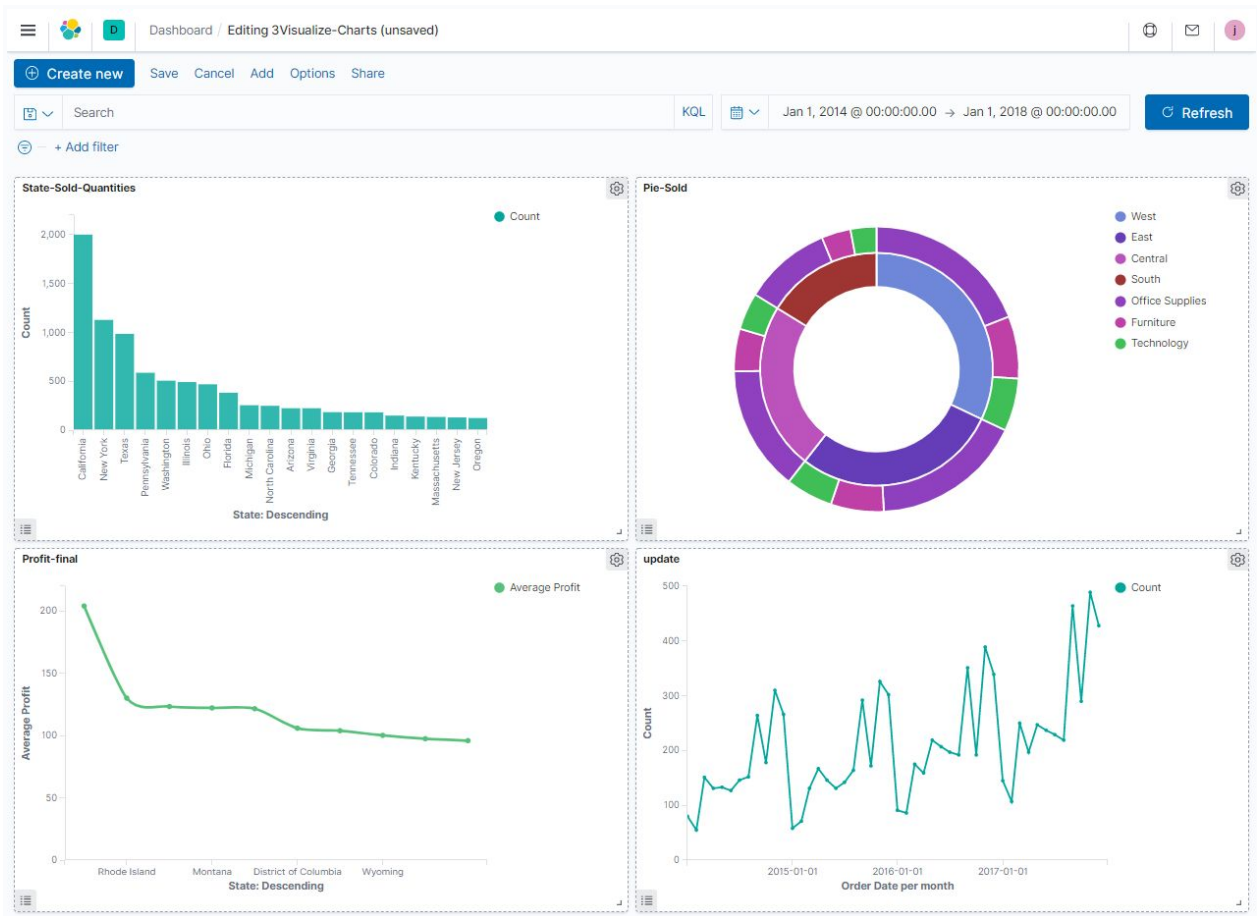
32. Click the left menu and click the dashboard.

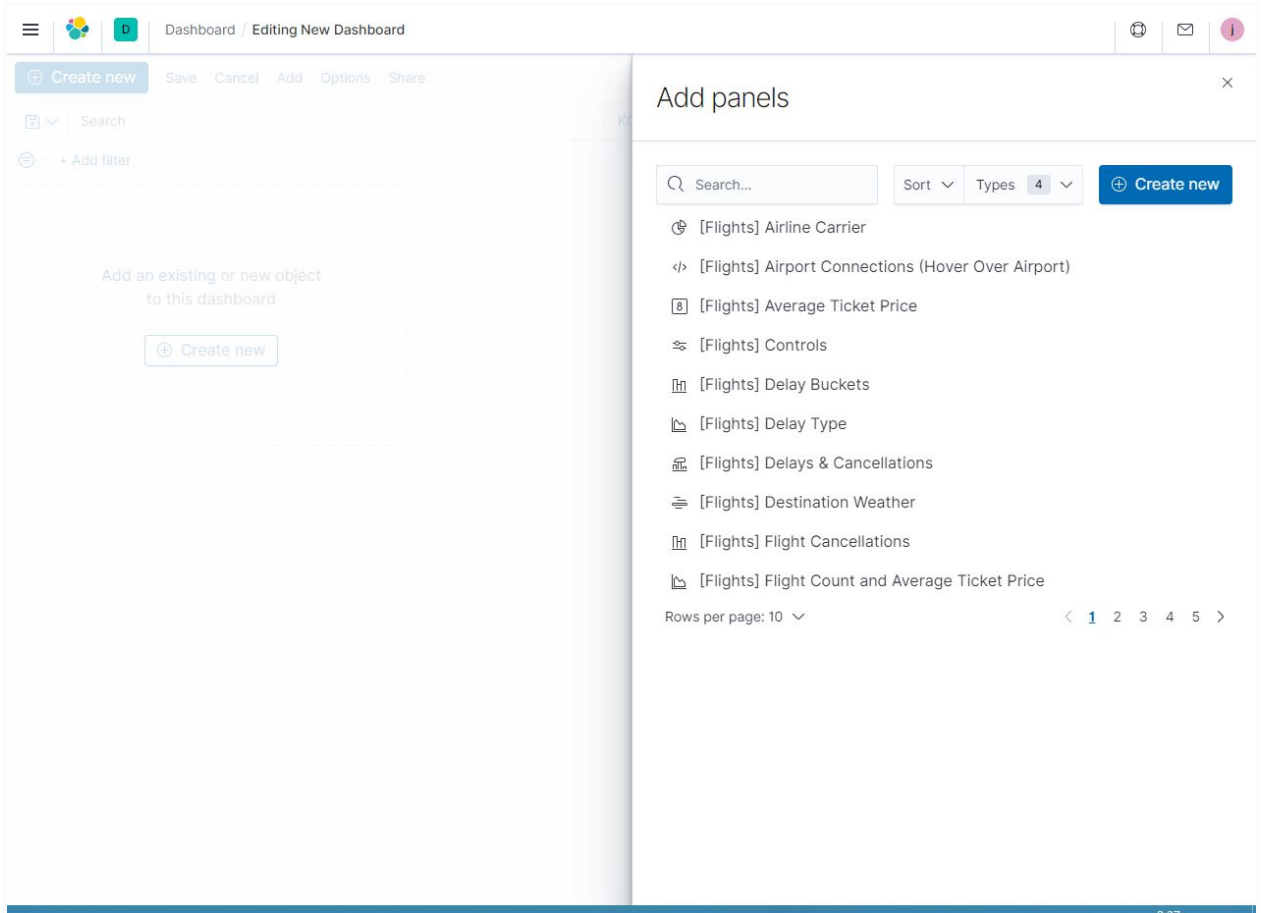


33. Click the Create dashboard in the blue button.



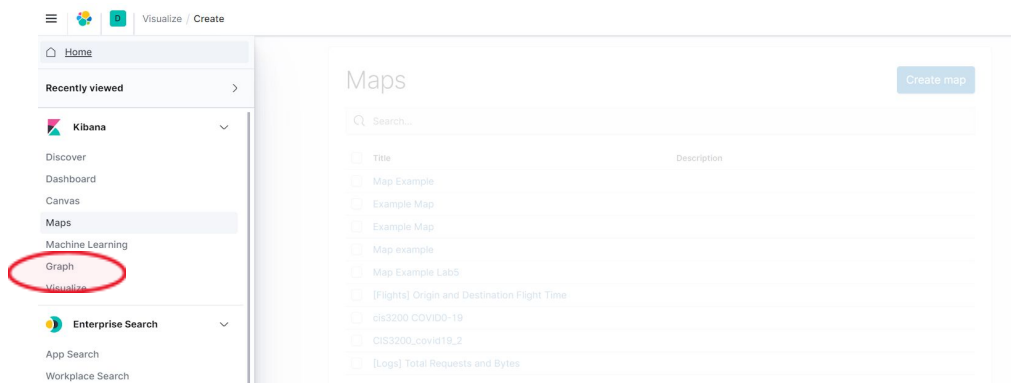
34. Then, Click the [Add an existing](#) option. You should be able to add all 4 we have done before. It would look like the one next page.



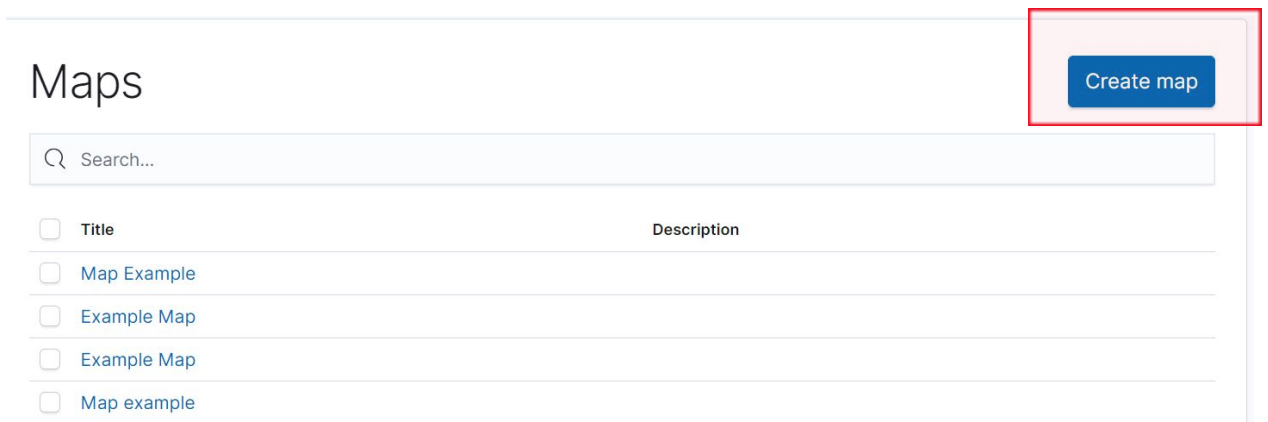


Create GEO Map Visualization

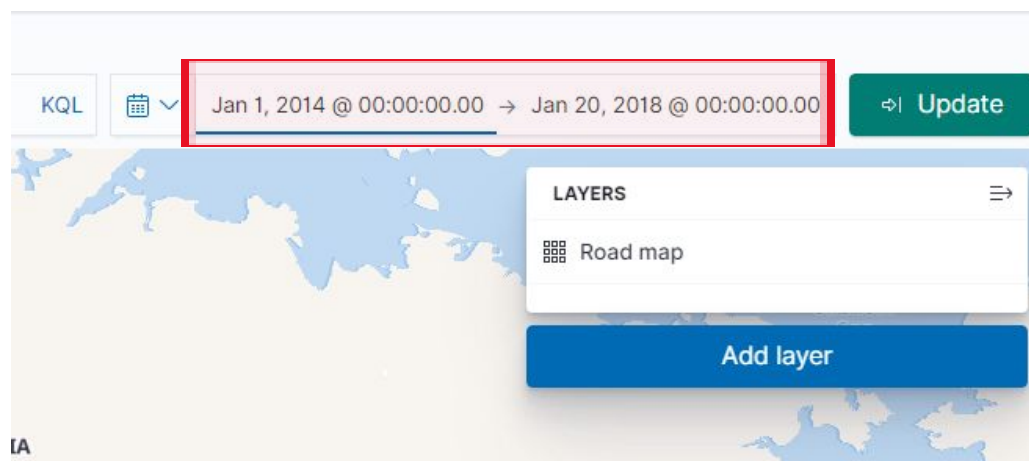
1. Access Kibana and select *Maps* on the left side scroll bar or go to *Visualize* to create a map visual.



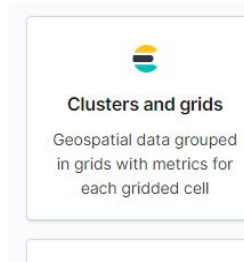
2.If creating a brand new map, select Create Map to start a brand new visualization



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



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*

Add layer

< Change layer

Index pattern

sueprstoredata	✓
logs*	
kibana_sample_data_flights	
csse*	
csse_covid*	
✓ sueprstoredata	
superstoredata	

Cancel

Add layer →

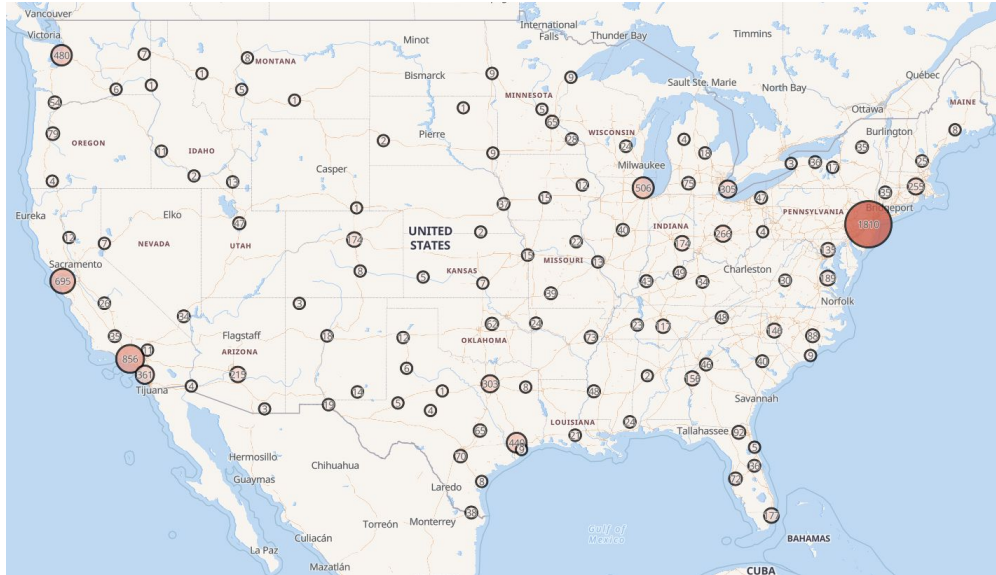
6. In this step we can add some changes to the *Layer Style*, some examples include changing the fill color gradient, border color, and border thickness.

The screenshot shows a 'Layer Style' configuration panel. At the top, there is a toggle switch labeled 'Apply global filter to layer data' which is currently turned on. Below this, the 'Layer Style' section contains several settings:

- Symbol type:** Two buttons, 'marker' (selected) and 'icon'.
- Fill color:** Two rows of dropdown menus. The first row has 'By value' and 'count'. The second row has 'As number' and a color gradient bar.
- Border color:** A dropdown menu with 'Solid' selected, followed by a color swatch and the hex code '#000'.
- Border width:** A dropdown menu with 'Fixed' selected, followed by the value '3' and a 'px' unit label.
- Symbol size:** A dropdown menu with 'By value' and 'count'.

At the bottom of the panel, there are three buttons: 'Cancel', 'Remove layer', and 'Save & close' (which is highlighted in blue).

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/0D54T00000CWx8SAL/sample-superstore-sales-excelxls>

2. URL of your Github

[xEqualsToY/Kibana-Data-Set: Data Visuliztion \(github.com\)](https://github.com/xEqualsToY/Kibana-Data-Set)

3. URL of References

[Elastic Cloud Tutorial: Getting Started with a sample dataset | Elastic Blog](#)