

WELCOME TO THE DATA VISUALIZATION LAB

Part 3 - Data Visualization [25 minutes]

Overview: What is Data Visualization and how do we use it with our data?

Purpose: Create visualization using the flight information answering the question (visually with a density graph):

- What were the most number of flights from destination to origin between (1995-2008) - Route Density

- 1) Open CDP, using the “admin” user within the Test Drive link.

Your link should look something like (remember click the link in your email not the link below)

http://login.trycdp.com/auth/realms/trycdp-trialxx/protocol/saml/clients/samlclient?tn=trialxx_admin@trycdp.com&p=X

*xx represents the trial user #

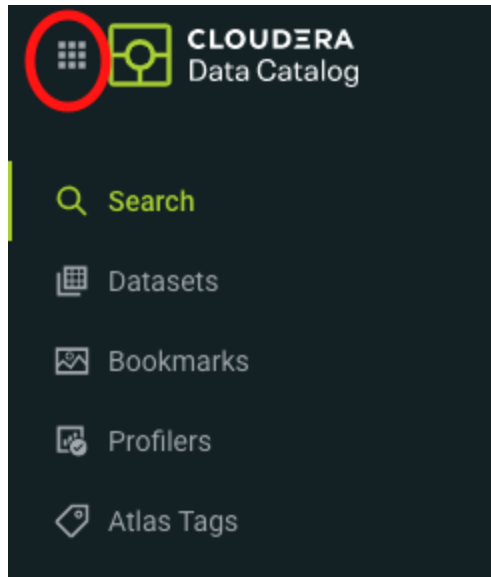
*X represents the password

- 2) Click the “Data Warehouse” within the CDP Home Screen

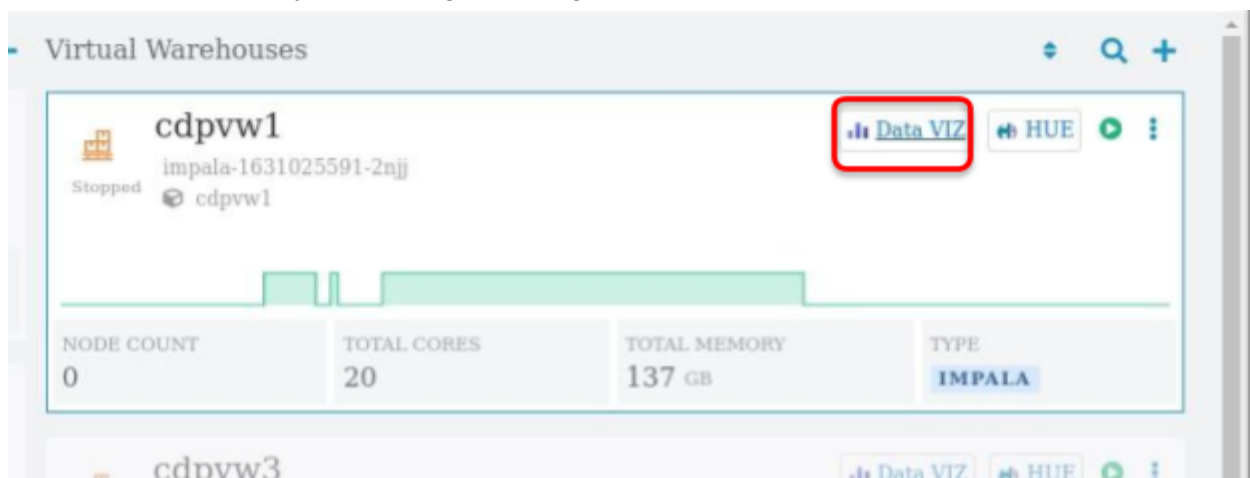


How do you get to the CDP Home Screen?

- From any experience such as “Data Catalog”, click the 9 square at the top left and then click “Home”

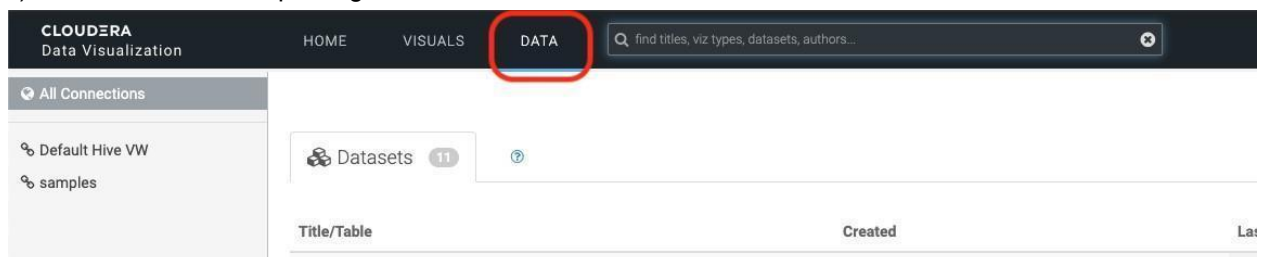


3) Click “Data VIZ” on your existing “Running” Virtual Warehouse.

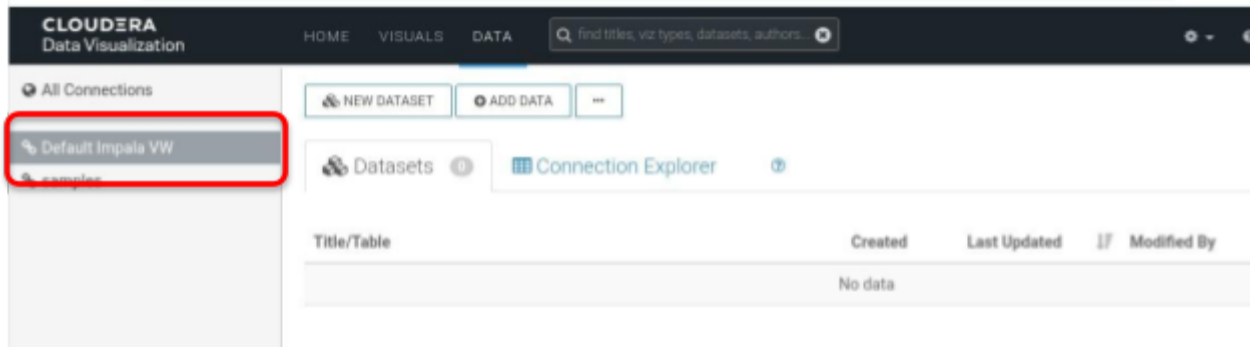


4) You are autologged into Data VIZ

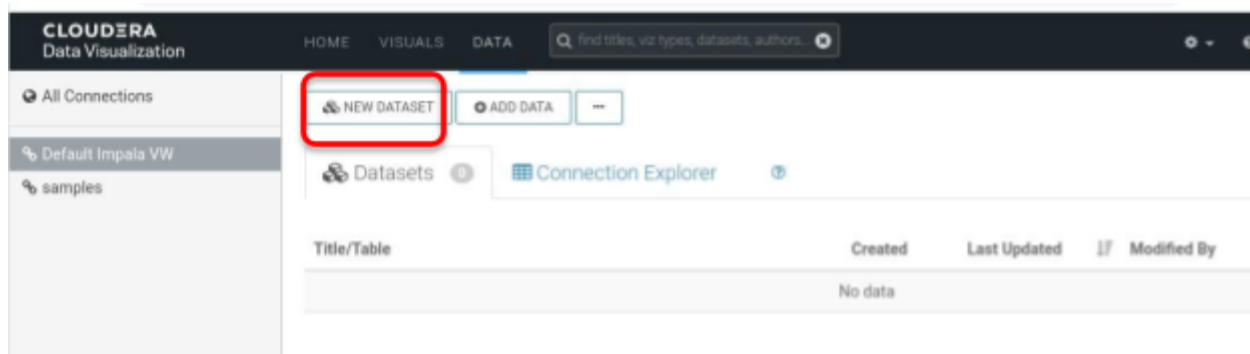
5) Click “DATA” the top navigation bar



- 6) Click “Default Impala VW” to add our dataset.



- 7) Click “NEW DATASET” to add our “flights” data



- 8) Enter a name for the Dataset title naming “mynamehere_airline_new_parquet_flights” Recall Hive prefers ORC and Impala prefers Parquet. Our development team is making both engines like both file formats equally.

New Dataset

Create a dataset from data on this connection. You need to create a dataset before you can create dashboards or apps.

Dataset title *

martyflightsnewparquet

Dataset Source

From Table

Select Database

airlines_new_parquet_cdpvw1

Select Table

flights

CANCEL

CREATE

- 9) Choose the database “airlines_new_parquet_YOURVIRTUALWAREHOUSENUMBER”

New Dataset

Create a dataset from data on this connection. You need to create a dataset before you can create dashboards or apps.

Dataset title *

martyflightsnewparquet

Dataset Source

From Table

Select Database

airlines_new_parquet_cdpw1

Select Table

flights

CANCEL

CREATE

10) Choose the table “flights”

*Need to import multiple databases and tables? You’d use Dataset Source = SQL

New Dataset

Create a dataset from data on this connection. You need to create a dataset before you can create dashboards or apps.

Dataset title *

Dataset Source

From Table

Select Database

airlines_new_parquet_cdpvw1

Select Table

flights

CANCEL

CREATE

11) Click “CREATE”

New Dataset

Create a dataset from data on this connection. You need to create a dataset before you can create dashboards or apps.

Dataset title *

Dataset Source

From Table ▾

Select Database

airlines_new_parquet_cdpvw1 ▾

Select Table

flights ▾

CANCEL

CREATE

12) Click "+" to create a New Dashboard

CLUSTERA
Data Visualization

[HOME](#)
[VISUALS](#)
[DATA](#)

All Connections

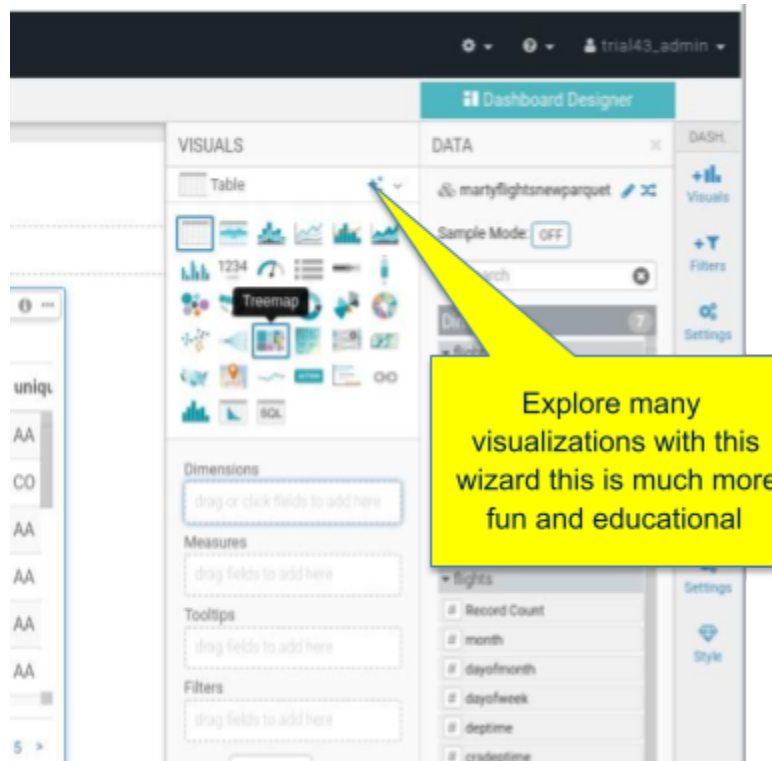
Default Impala VW
 samples

NEW DATASET
 ADD DATA
 ...

Datasets 1
 Connection Explorer

Title/Table		Created	Last Update
martyflightsnewparquet	+	Sep 09, 2021	a few seconds ago
airlines_new_parquet_cdpvw1.flights			

13) Choose “TreeMap” under “VISUALS” or explore with the Wizard



- 14) Drag-and-drop both “dest” and origin” from Dimensions->Flights into Dimensions under Visuals

The screenshot displays the Cloudera Dashboard Designer interface. The main workspace is divided into three primary sections: VISUALS, DATA, and a right-hand sidebar.

- VISUALS:** Contains a grid of visualization types (Table, Bar, Line, etc.) and a 'Dimensions' section. The 'Dimensions' section is highlighted with a red box and contains two items: 'dest' and 'origin', both with a small 'A' icon and a right-pointing arrow.
- DATA:** Shows the data source 'airlines_new_orc.flights' and 'Sample Mode: OFF'. It lists dimensions for the 'flights' table: uniquecarrier, tailnum, origin, dest, cancellationcode, and diverted. It also lists measures: Record Count, month, dayofmonth, dayofweek, deptime, crsdeptime, arrtime, crsarrrtime, and flightnum.
- Right Sidebar:** Includes a 'DASH.' section with 'Visuals', 'Filters', 'Settings', and 'Style' options. Below this is a 'VISUAL' section with a 'Build' button and 'Settings' and 'Style' options. At the bottom is a 'Tags' section with color-coded tags: Red, Orange, Yellow, Green, and Blue.

At the bottom of the interface, there is a 'Limit: 100' input field and a 'REFRESH VISUAL' button.

15) Drag-and-drop “Record Count” from Measures->Flights into Measures under Visuals

The screenshot shows the Cloudera Dashboard Designer interface. The left sidebar contains the 'VISUALS' panel, which is currently set to 'Table' view. Below the visual type selector, there are various visualization icons. The 'Measures' section is highlighted with a red circle, and the 'Record Count' measure is being dragged from the 'Measures' list to the 'Measures' section. A tooltip 'sum(1)' is visible above the 'Record Count' measure. The 'DATA' panel on the right shows the 'airlines_new_orc.flights' dataset. The 'Dimensions' section lists fields like 'uniquecarrier', 'tailnum', 'origin', 'dest', 'cancellationcode', and 'diverted'. The 'Measures' section lists fields like 'Record Count', 'month', 'dayofmonth', 'dayofweek', 'deptime', 'crsdeptime', 'arrtime', 'crsarrrtime', 'flightnum', 'actualelapsedtime', 'crselapsedtime', 'airtime', and 'arrdelay'. The 'Measures' section is currently empty, and the 'Record Count' measure is being added to it. The 'Measures' section is currently empty, and the 'Record Count' measure is being added to it.

16) Click the right arrow next to Record Count and select “Descending” under Order and Top K

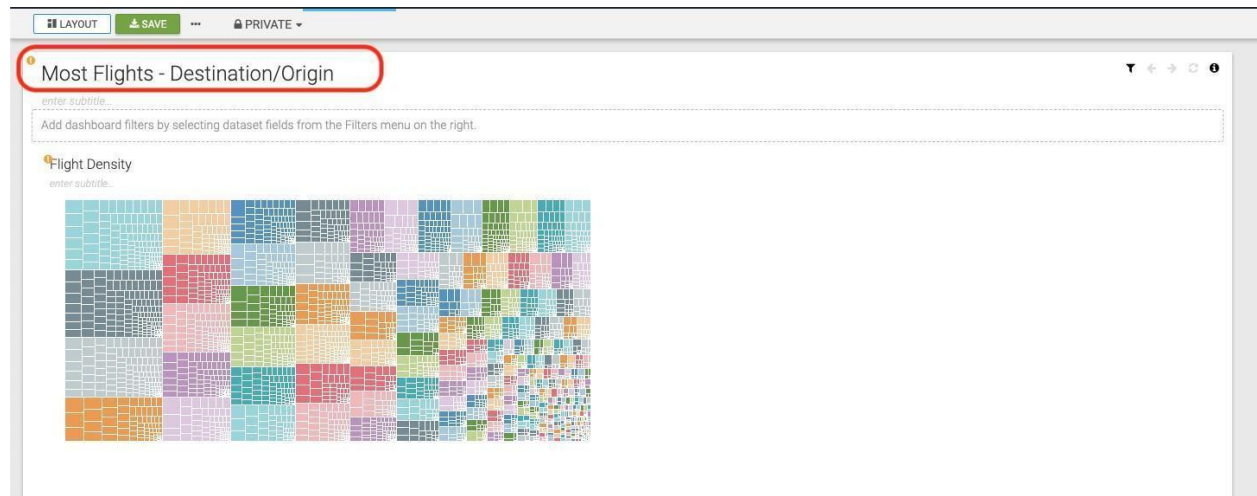
The screenshot shows the Clustera Dashboard Designer interface. On the left, the 'VISUALS' panel displays various chart types, with 'Treemap' selected. Below this, the 'Dimensions' section lists 'dest' and 'origin'. The 'Measure' section shows '# Record Count' with a right arrow icon. Below the measures are sections for 'Tooltips', 'X Trellis', 'Y Trellis', and 'Filters', each with a 'drag fields to add here' prompt. A 'REFRESH VISUAL' button is at the bottom left. The central 'FIELD PROPERTIES' panel is open, showing a list of function categories: 'Date/Time Functions', 'Text Functions', 'Analytic Functions', 'Change Type', and 'Order and Top K'. The 'Order and Top K' category is expanded, showing 'Descending' (selected with a green checkmark) and 'Ascending'. Below these are input fields for 'Top K' and 'Bottom K', both set to 'eg. 100'. A note states 'Top K/Bottom K applies to granular dimensions'. At the bottom of the panel are buttons for 'Enter/Edit Expression', 'Display Format', 'Alias', 'Description', 'Duplicate', 'Save Expression', and 'Remove'. On the right side of the interface, a vertical toolbar contains icons for 'DASH.', 'Visuals', 'Filters', 'Settings', 'Style', 'VISUAL', 'Build', 'Settings', and 'Style'.

17) Click “REFRESH VISUAL”

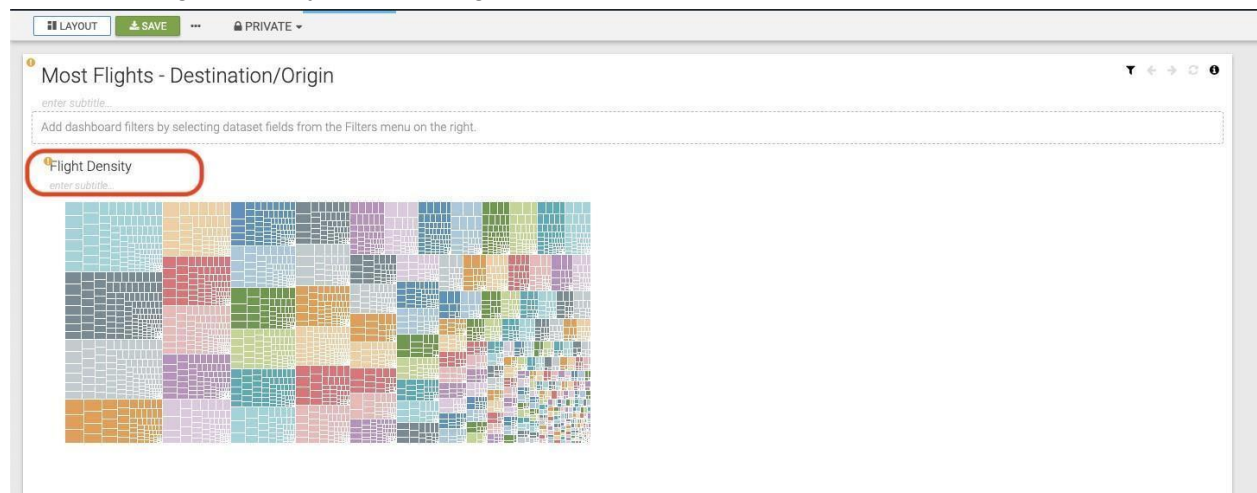
*Notice - you can have other Visuals chosen to be displayed with the Dimensions and Measure(s), then click REFRESH VISUALS

The screenshot shows the 'Dashboard Designer' interface. On the left, the 'VISUALS' panel displays various chart types and a list of fields categorized into Dimensions (dest, origin) and Measures (Record Count). Below these are sections for Tooltips, X Trellis, Y Trellis, and Filters, each with a 'drag fields to add here' prompt. At the bottom of the left panel, a blue button labeled 'REFRESH VISUAL' with a circular arrow icon is highlighted with a red circle. The central 'DATA' panel shows the selected data source 'airlines_new_orc.flights', 'Sample Mode: OFF', a search bar, and lists of Dimensions (6 items) and Measures (24 items). The right sidebar contains navigation links for Visuals, Filters, Settings, Style, and a 'BUILD' button.

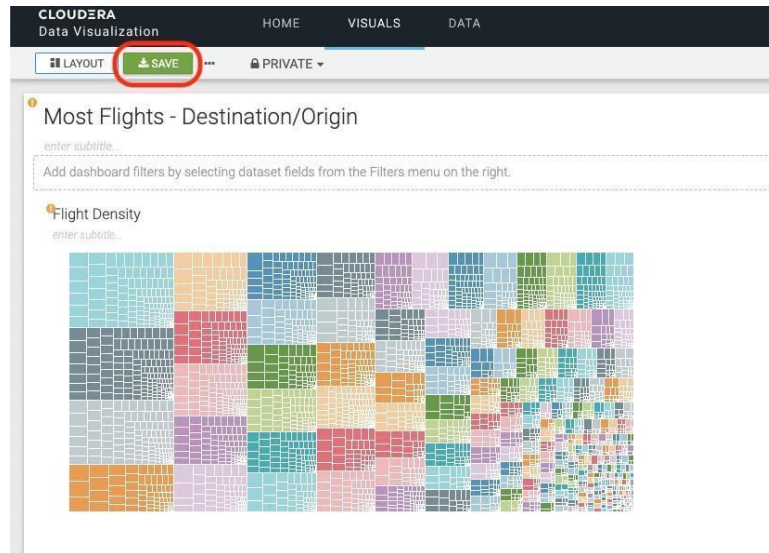
18) Enter a title “Most Flights - Destination/Origin”



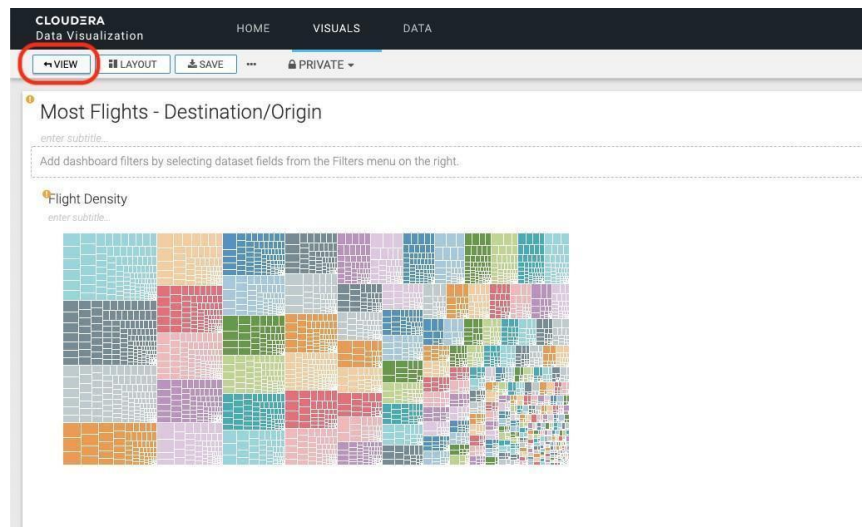
19) Enter “Flight Density” under the graph’s title



20) Click “SAVE”



21) Click “VIEW”



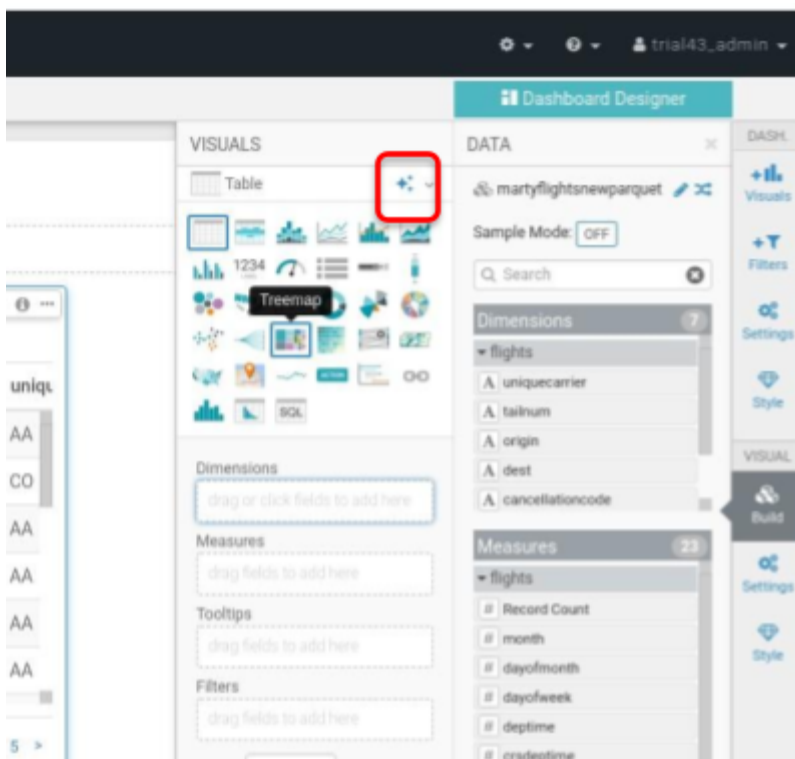
22) Scroll over the graph and click “Expand Visual”



Destinations are displayed



23) Time to explore the other visualization options.



Please scroll through the options and experiment with some of the additional visuals.

