

MIS 0855 – Data Science (Section 006) – Fall 2017 In-Class Exercise (Day 27-28) – Visualizing Network

Objective: Learn how to visualize a network over Tableau

Learning Outcomes:

- Learn how to structure network data in a way that can be visualized as “paths” in Tableau
- Learn how to visualize a network in an interactive manner with a Tableau dashboard.

In this exercise, we will use U.S. domestic flight data from the Airline On-Time Performance Database from the U.S. Department of Transportation (https://www.transtats.bts.gov/Tables.asp?DB_ID=120).

Part 0. Understand the Dataset

1) Download “Flight Paths July 4-8 2016.xlsx” and save it to your computer. Remember where you saved it!

2) Open this Excel file.

origin_dest	carrier	carrier_code	flight_no	date	airport	path_id	latitude	longitude
Origin	Delta	DL	2186	07-04-2016	ATL	DL-ATL-LGA-2186-4	33.6367	-84.4281
Dest	Delta	DL	2186	07-04-2016	LGA	DL-ATL-LGA-2186-4	40.7772	-73.8726
Origin	Delta	DL	1666	07-04-2016	ATL	DL-ATL-RDU-1666-4	33.6367	-84.4281
Dest	Delta	DL	1666	07-04-2016	RDU	DL-ATL-RDU-1666-4	35.8776	-78.7875
Origin	Delta	DL	1641	07-04-2016	ATL	DL-ATL-SAT-1641-4	33.6367	-84.4281
Dest	Delta	DL	1641	07-04-2016	SAT	DL-ATL-SAT-1641-4	29.5337	-98.4698
Origin	Delta	DL	1666	07-04-2016	RDU	DL-RDU-ATL-1666-4	35.8776	-78.7875
Dest	Delta	DL	1666	07-04-2016	ATL	DL-RDU-ATL-1666-4	33.6367	-84.4281
Origin	Delta	DL	1641	07-04-2016	SAT	DL-SAT-ATL-1641-4	29.5337	-98.4698
Dest	Delta	DL	1641	07-04-2016	ATL	DL-SAT-ATL-1641-4	33.6367	-84.4281

This file contains flight path data of four aircrafts from four airlines (Delta, Spirit, United, and Southwest) in July 4-8, 2016.

3) In order to visualize a network, Tableau requires a specific data format as above.

- For each link (path) of a network, there should be a unique path ID (path_id) above.
- The dataset should include the longitude and latitude of the starting and the end points, which are the origin and the destination airports in our case.
- The first column (origin_dest) specifies whether each row represents the starting (“Origin”) or the end point (destination as “Dest”) of the path.
- In other words, for each path, there should two rows in the dataset (one for the starting point and the other for the end point).

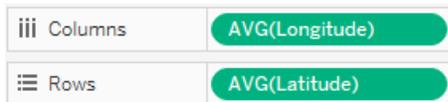
Part 1. Visualizing Flight Paths of Aircrafts

- 1) Start Tableau and connect to "Flight Paths July 4-8 2016.xlsx."
- 2) Select "Flight Paths" as a data source, and go to Sheet 1.
- 3) On Measures, you can see that Latitude and Longitude are recognized as Geographic Measures.

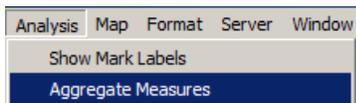
Measures

- # Flight No
- 🌐 Latitude
- 🌐 Longitude
- # *Number of Records*
- # *Measure Values*

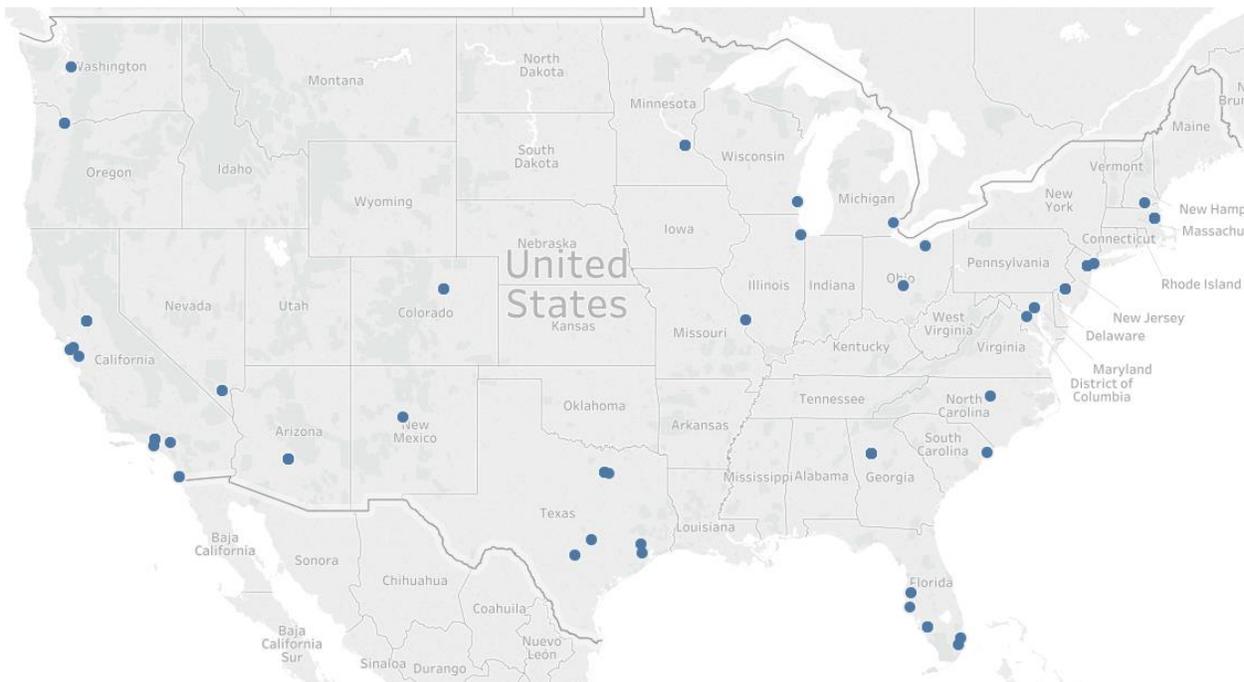
- 4) Drag Latitude to Rows and Longitude to Columns.



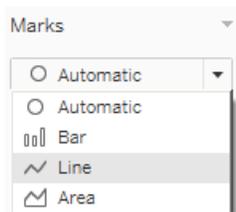
- 5) Click Analysis Menu, and unselect Aggregate Measures.



Doing so will show you the U.S. map with dots that indicate the locations of the airports.



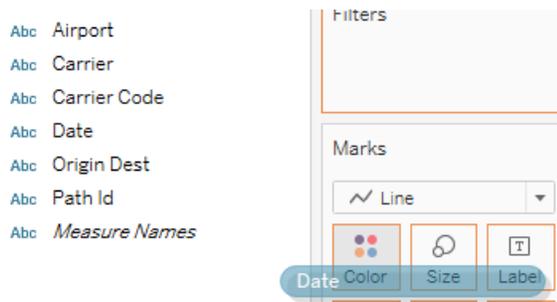
6) Under Marks, change Automatic to Line, which will create weird lines.



7) Drag "Path Id" to Detail under Marks.

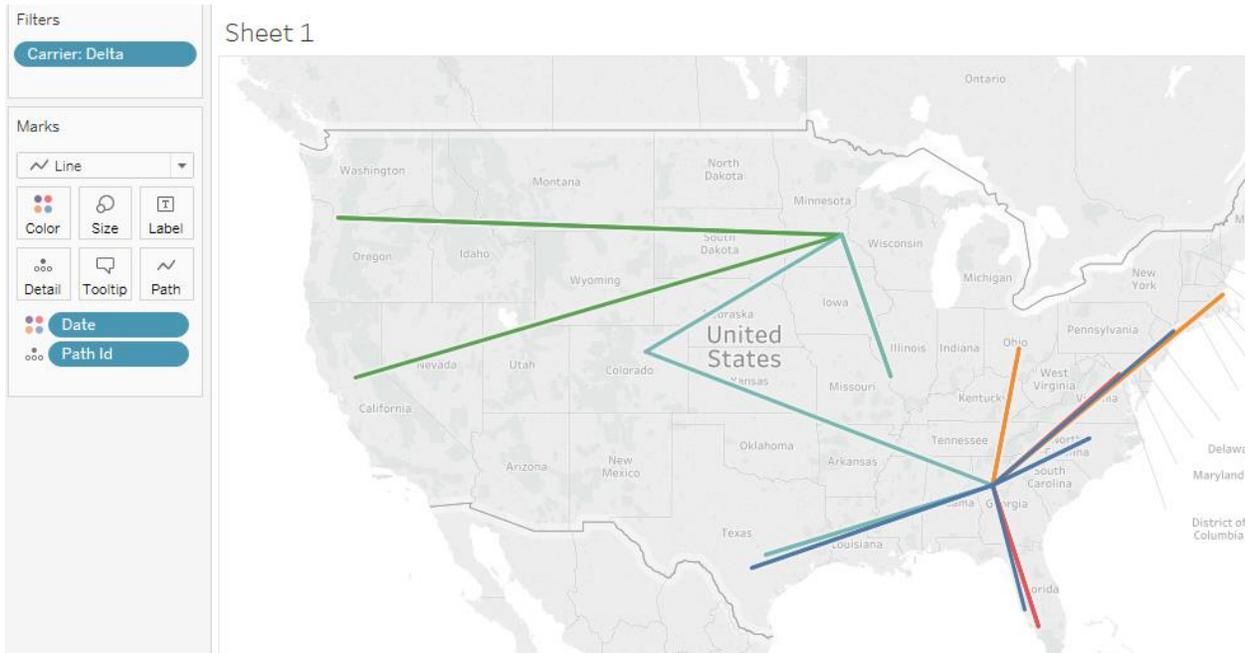


8) Drag Date to Color under Marks.

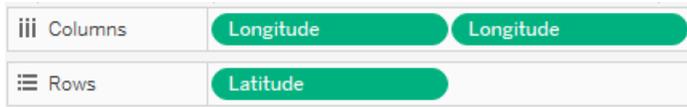


The color in each path represents different flight dates (from July 4 to July 8).

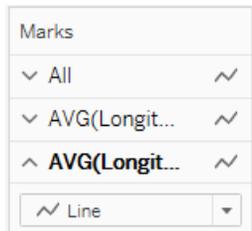
9) Drag Carrier to Filters, and select any airline you want.



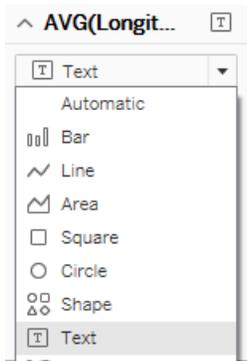
10) Let's try to show the airport code over this map. Drag Longitude again to Columns. You will see the same two maps.



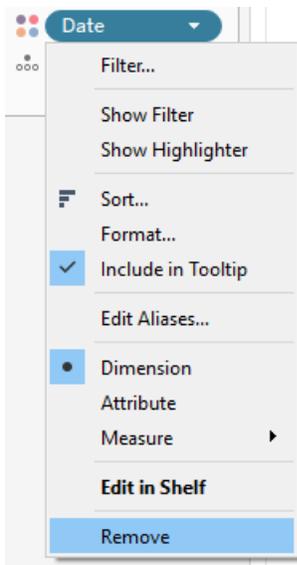
11) Under Marks, click the **second** AVG(Longitude).



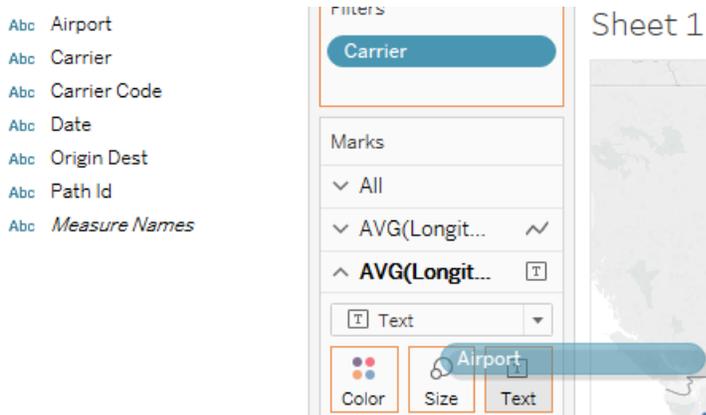
12) Change Line to Text.



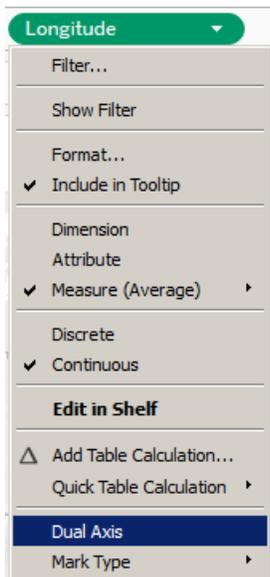
13) Remove Date.



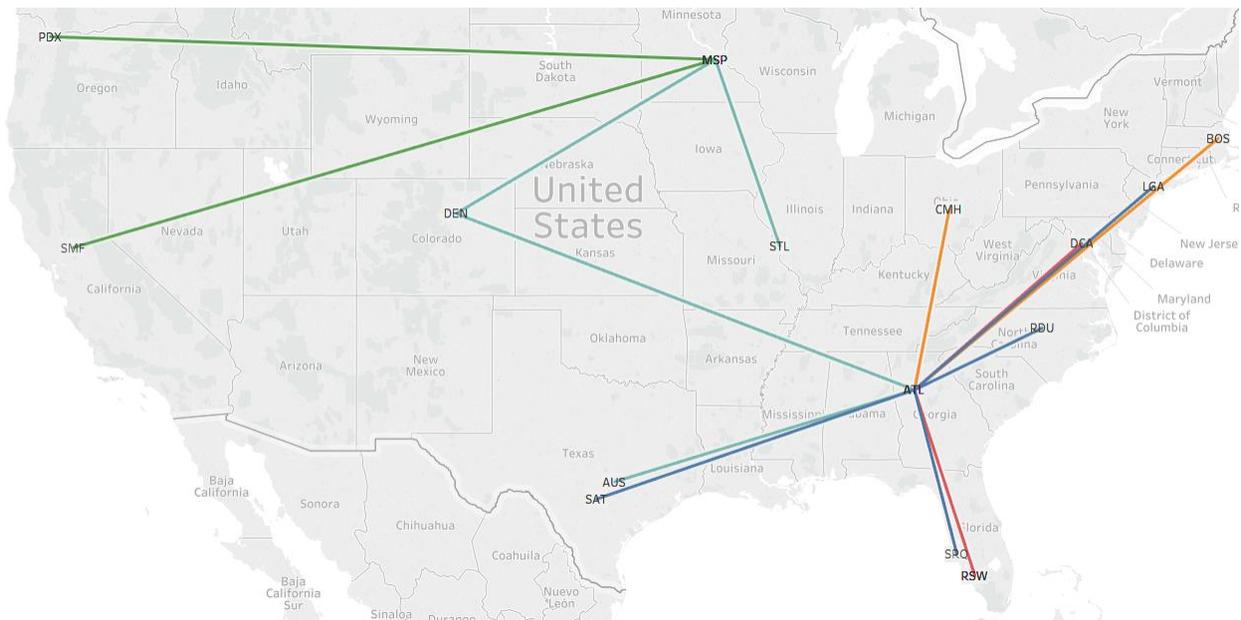
14) Drag Airport over Label under the second AVG(Longitude), which will show the airport codes.



15) Click the second Longitude on Columns and select Dual Axis.



Now, you will see the flight path map accompanied with the airport codes.

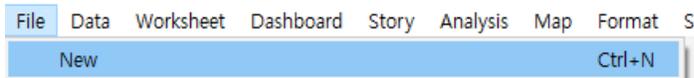


Part 2. Visualizing Flight Networks of Airlines

1) Download "Flight Networks July 4 2016.xlsx."

This file contains the records of all U.S. domestic flights on July 4, 2016, including the number of flights between two cities by each airline as well as the average fare of each route.

2) In Tableau, click File Menu and New.

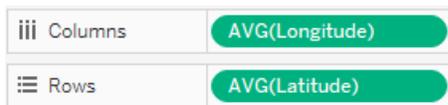


3) Under Data, click "Connect to Data."

4) Select Excel and "Flight Networks July 4 2016.xlsx."

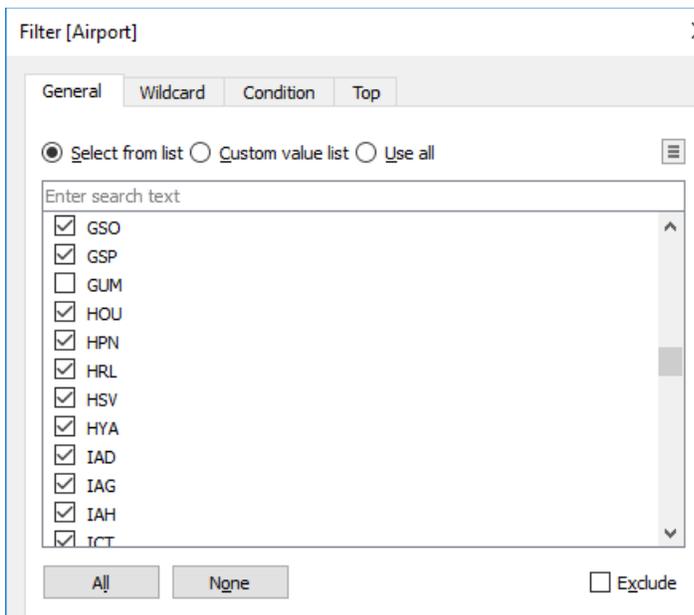
5) Select "Flight Networks" as a data source, and click Sheet 1.

6) Drag Latitude to Rows and Longitude to Columns.

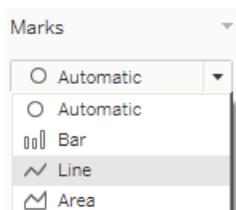


7) Click Analysis Menu, and unselect Aggregate Measures.

8) You will see one dot outside of the Continental U.S., which is Guam, a U.S. territory. Drag Airport to Filter, select All, and unselect GUM (Guam Airport).

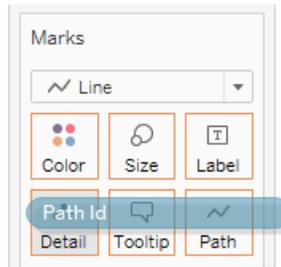


9) Under Marks, change Automatic to Line.

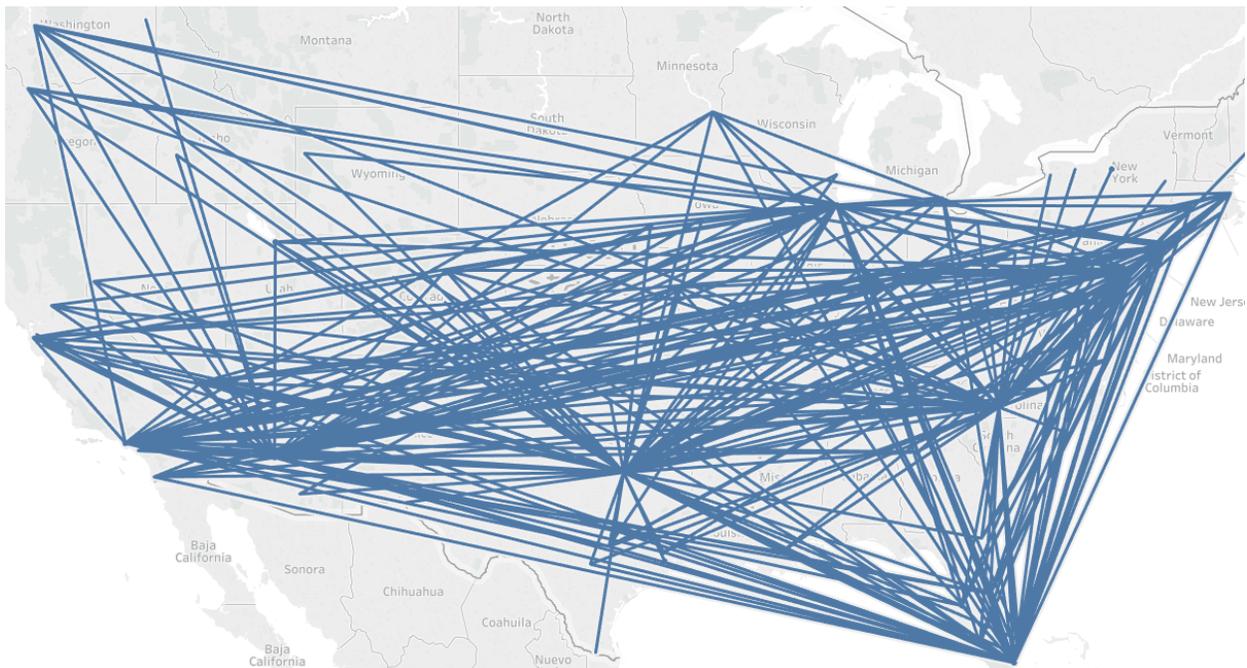
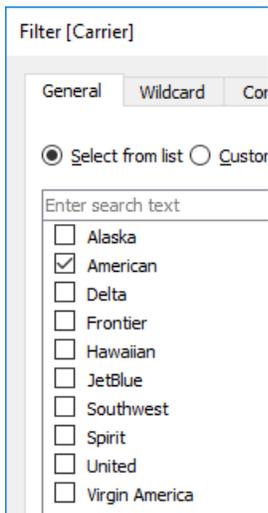


10) Drag Path Id to Detail under Marks.

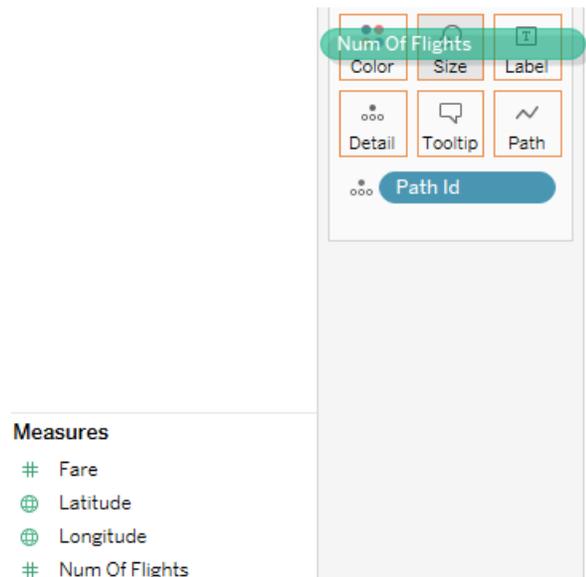
Abc Origin Dest
Abc Path Id
Abc Measure Names



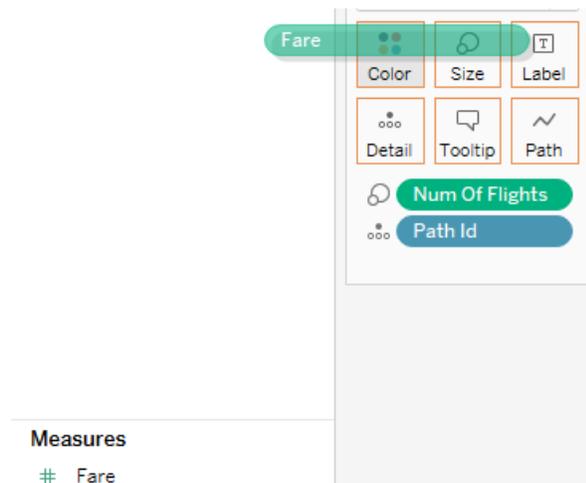
11) Drag Carrier to Filters, and select any one airline you'd like.



12) Drag “Num of Flights” to Size under Marks. This will visualize the frequency of flights by the width of lines.



13) Drag “Fare” to Color under Marks.



14) Under Marks, click Color and Edit Colors.

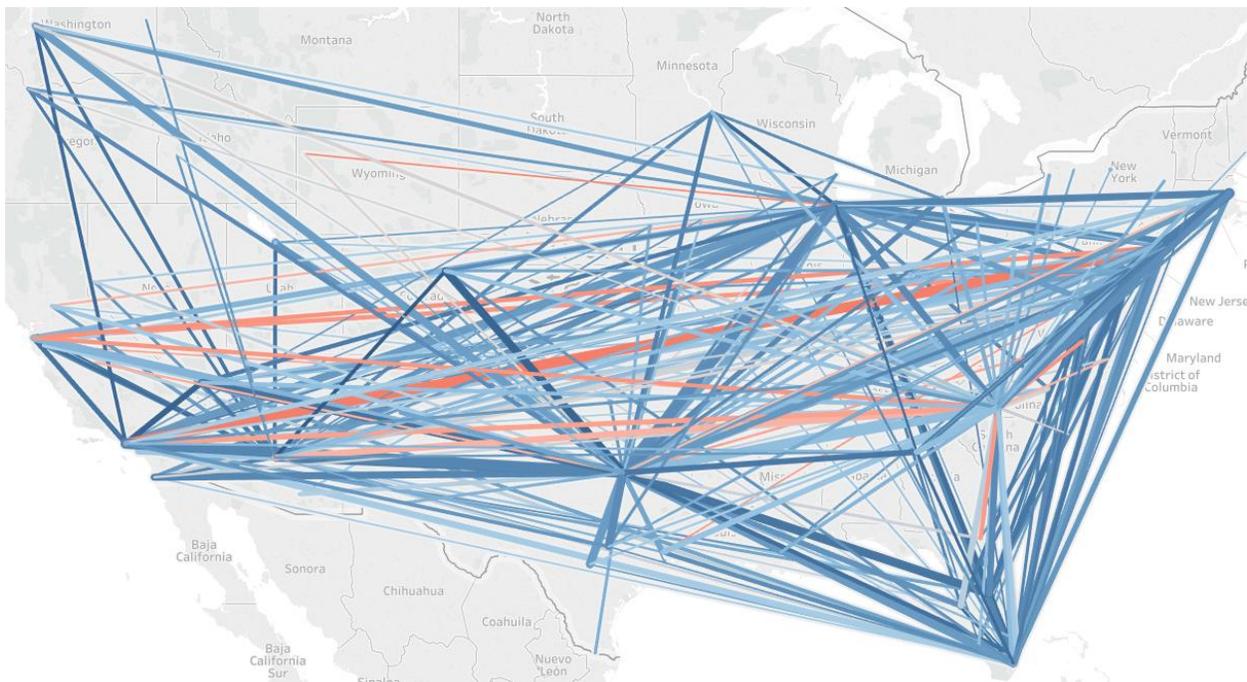
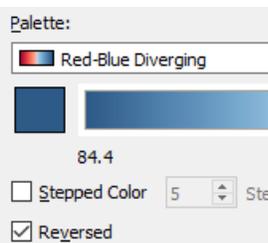
15) Under Palette, select “Red-Blue Diverging.”

Edit Colors [Fare]



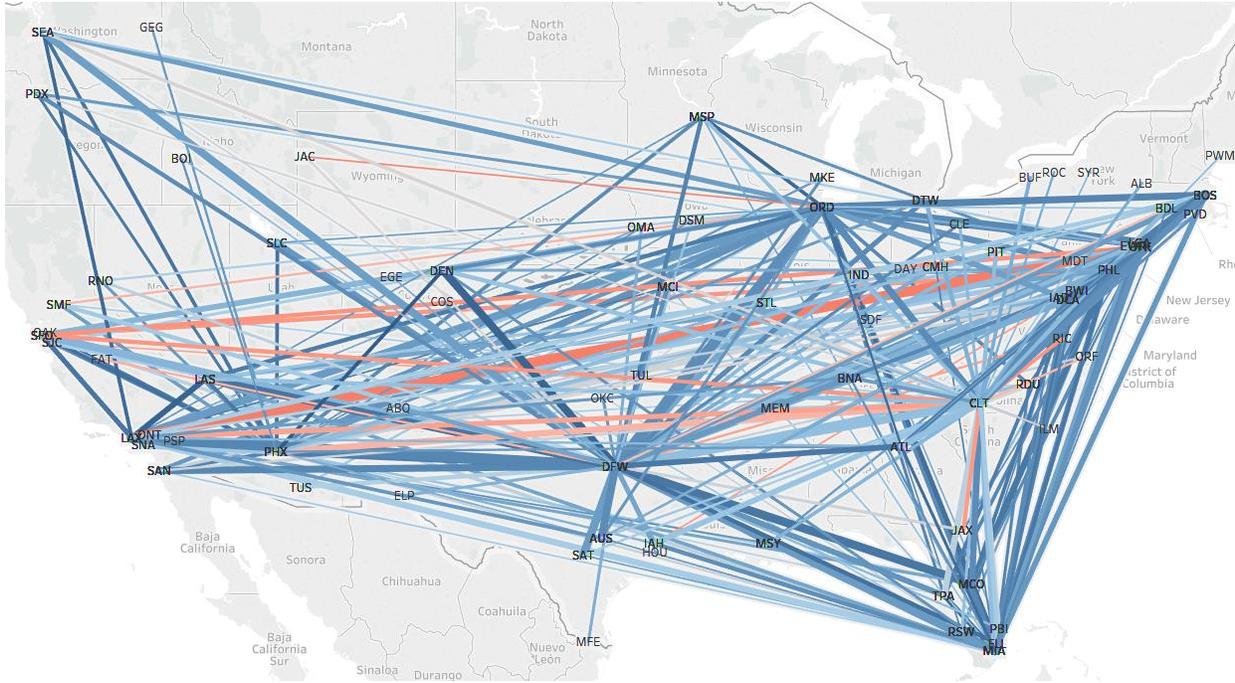
By selecting this palette, the routes with lower fares will appear with red. However, we'd prefer higher fares to be displayed as red, which would be more intuitive.

Select “Reversed” under Palette.



Now we have a map of the flight network, which visualizes both fare (color) and frequency (width).

16) (Do this on your own). As in Part 1, try to display the airport codes over the map.



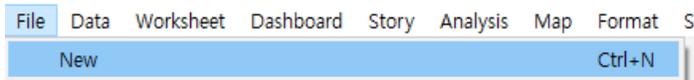
Part 4. Visualizing Capital Bikeshare

1) Download “Capital Bikeshare Trips Apr 1 2016.xlsx.”

In case you haven’t heard of, Capital Bikeshare is a bike-share program in Washington D.C. (similar to Indego in Philadelphia). We are using the bikeshare trip data on April 1st, 2016 obtained from <https://www.capitalbikeshare.com/system-data>.

- “Data” Worksheet is the original data from Capital Bikeshare.
- “Trips” Worksheet is the reorganized data from “Data” in a way that can be structured for network visualization in Tableau.
- There are two account types. Frequent bikeshare riders can register for a membership for one month or more (“Registered”). Occasional riders or tourists can obtain a temporary membership for a single ride (“Casual”).
- We can find where each ride started and ended with “Station IDs.” “Stations” Worksheet provides the location (longitude and latitude) and address of each station. This data is obtained from <http://opendata.dc.gov/datasets/capital-bike-share-locations>.

2) In Tableau, click File Menu and New.



3) Under Data, click "Connect to Data."

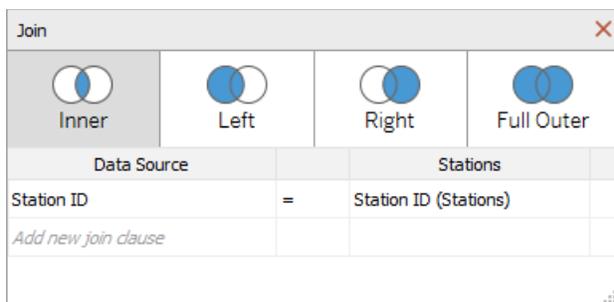
4) Select Excel and "Capital Bikeshare Trips Apr 1 2016.xlsx."

5) Drag "Trips" from left to right. Also, drag Stations next to Trips.

Trips+ (Capital Bikeshare Trips Apr 1 2016)



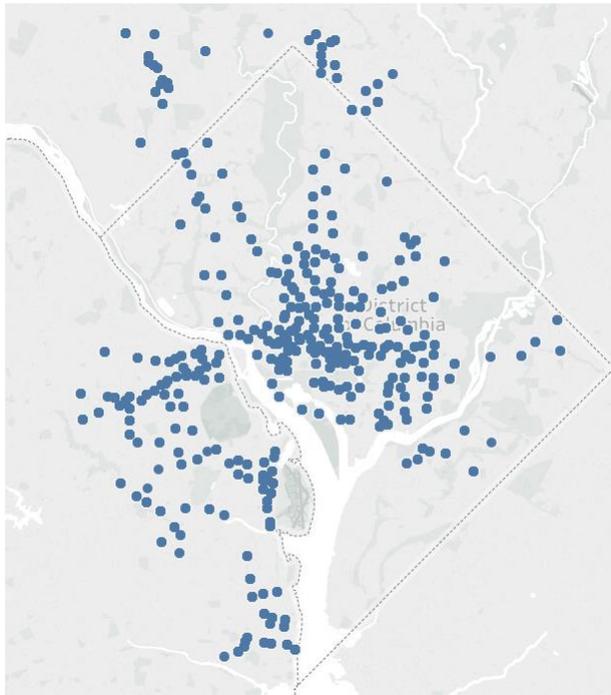
Tableau joins the two worksheets by matching Station IDs.



6) Click Sheet 1.

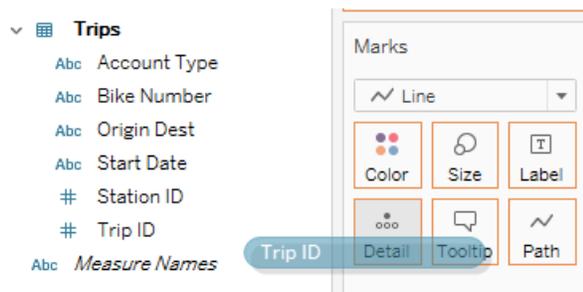
7) Drag Latitude to Rows and Longitude to Columns.

8) Click Analysis Menu, and unselect Aggregate Measures. You will see a map of Washington DC area.

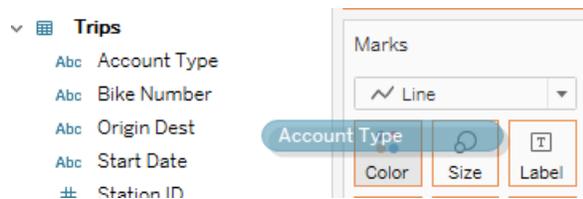


9) Change the chart type from Automatic to Line.

10) Drag Trip ID over Detail under Marks.



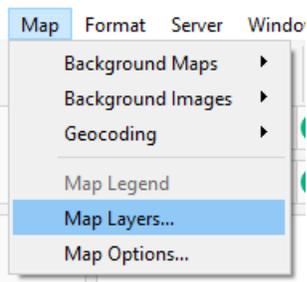
11) Drag Account Type from Trips to Color.



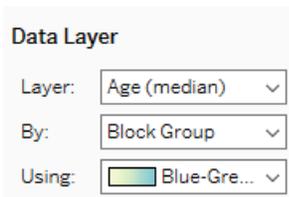
The map will show the two different types of rides in different colors.



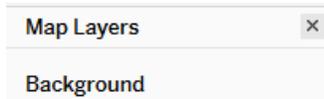
12) Click Map Menu and click Map Layers.



13) On Data Layer panel, select “Age (median)” and “Block Group.” The map will visualize the median age of population in each block in DC Area.

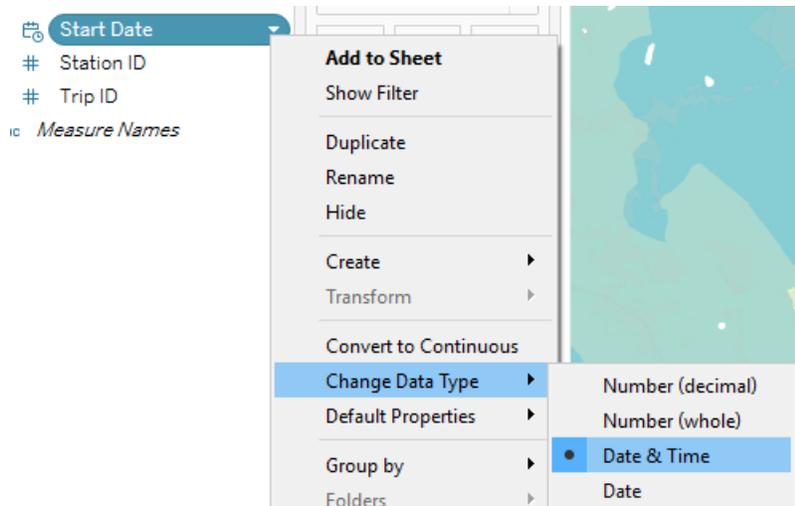


14) Click X icon next to Map Layers.

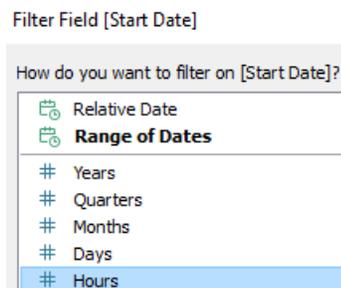


15) Drag Account Type and Bike Number to Filters. For each dimension, click All in the next screen and click Ok.

16) Click Start Date in Trips dimensions, select Change Data Type, and select Date & Time.

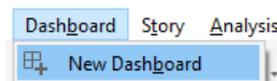


17) Drag Start Date to Filters and select Hours. Click All in the next screen and click Ok.



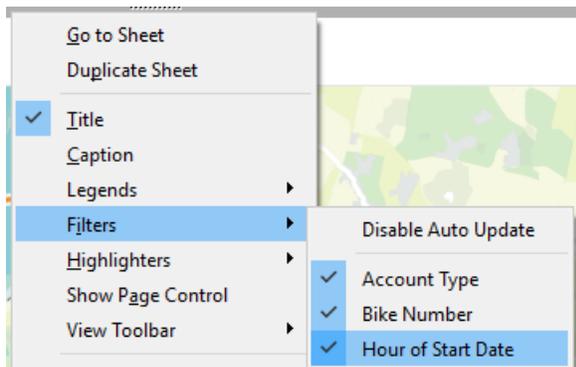
18) Rename Sheet 1 to “Trips in April 2016.”

19) Create a new Dashboard.

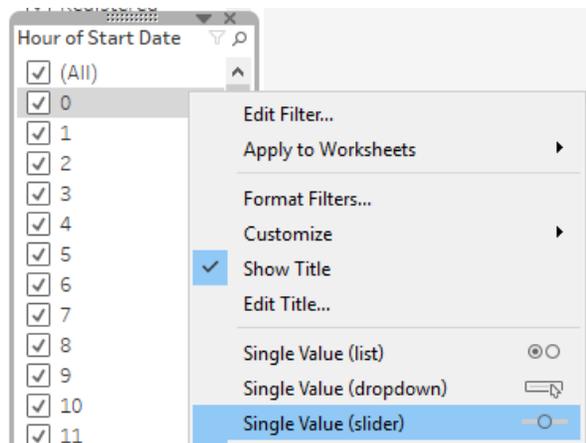


20) Drag “Trips in April 2016” to the right.

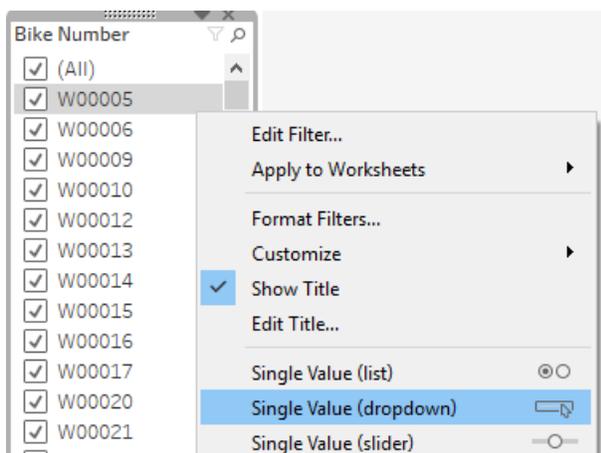
21) Add three filters (Account Type, Bike Number, and Start Date) to the dashboard.



22) Right-click "Hour of Start Date" filter and select "Single Value (slider)."



23) Right-click "Bike Number" filter and select "Single Value (dropdown)."



24) Enjoy the dashboard! Try to find something interesting about Capital Bikeshare.