Flight Data Analysis

Data: Global Flights Network

Source: OpenFlights.org and RITA

Description: Airports and flight route data, aggregated and sliced in various ways.

URL: http://www.visualizing.org/datasets/global-flights-network

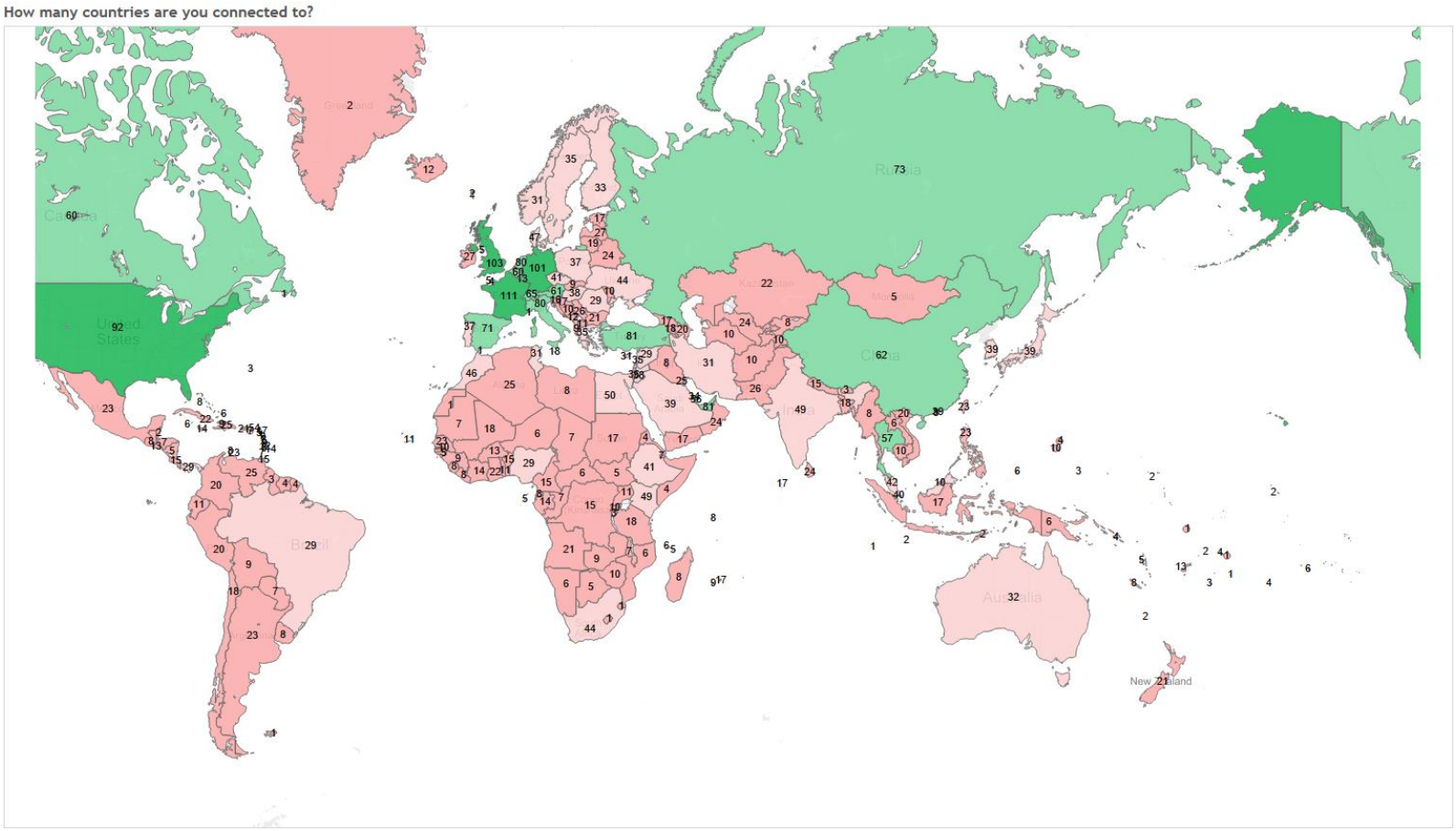
Remarks:

- Data was originally divided onto 6 files
- Data collected primarily for 2010 (except for passenger data which includes additional years)
- I added 1 file using R in order to structure the data for producing 'The Mesh' (R file attached)
- Analysis consists of 7 worksheets, 1 dash board, and 1 story
- Data description available in Appendix A
- I published the story to the following url: https://public.tableausoftware.com/profile/sherif#!/vizhome/SherifMostafa-D3MProject-FlightsAnalysis/TheStory

Table of contents

1. Worksheet: Most Connected Countries

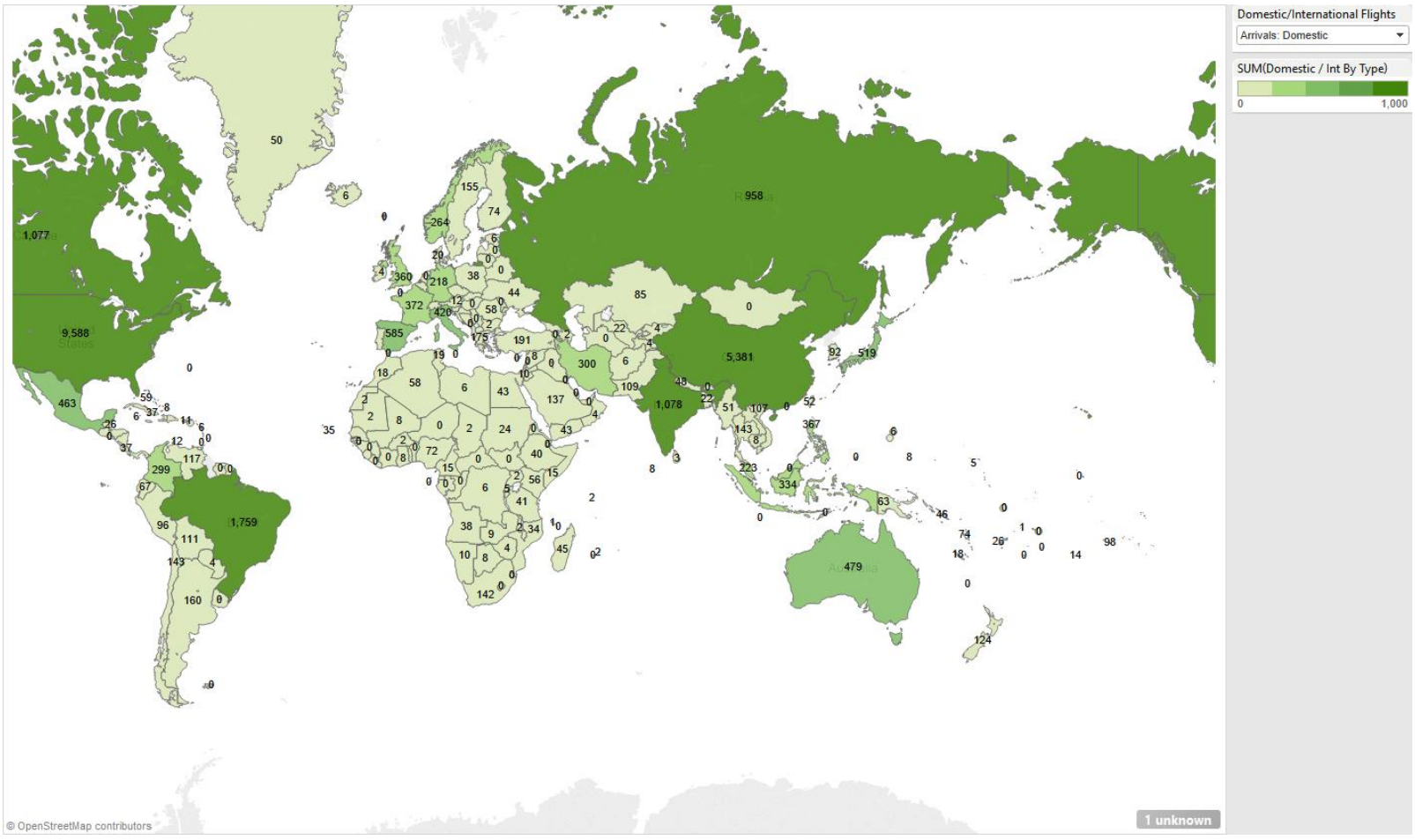
- Presents how many countries are each country connected to
- Used calculated field ‘Country Connections’ in order to filter domestic flights and flag the presence of an international route
- Apparently France has the most connections at 111 countries
- Location, economic welfare play a major role in affecting this measure



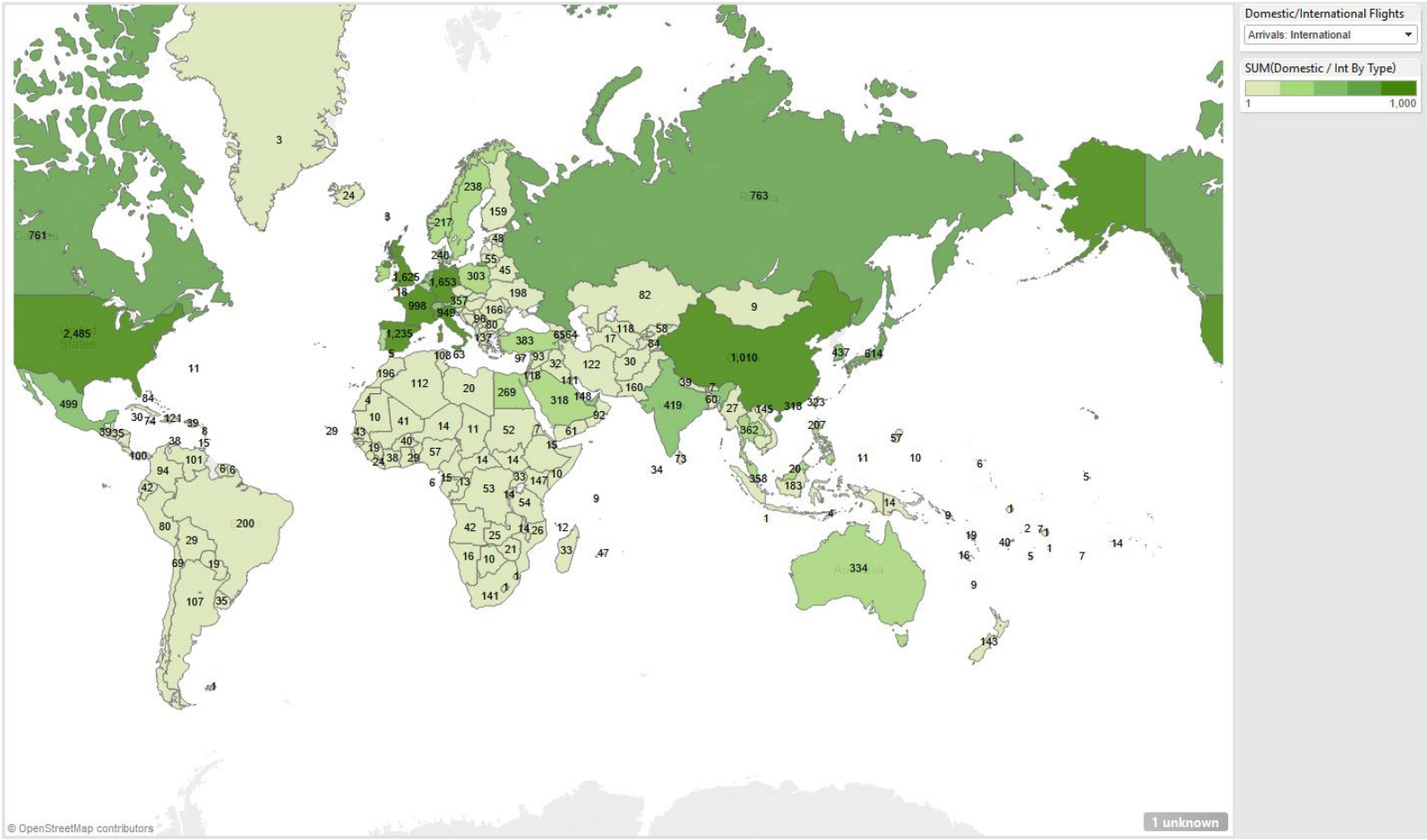
Country Connections

2. Worksheet: Countries by Domestic / Int Arrivals

- Presents how many flights per country by type: Domestic / International & Inbound /Outbound
- Parameters and calculated fields have been used to analyze the proper data upon user selection of the type
- The US has most domestic and international flights
- Not all departure flights have corresponding arrival flight (not symmetric)



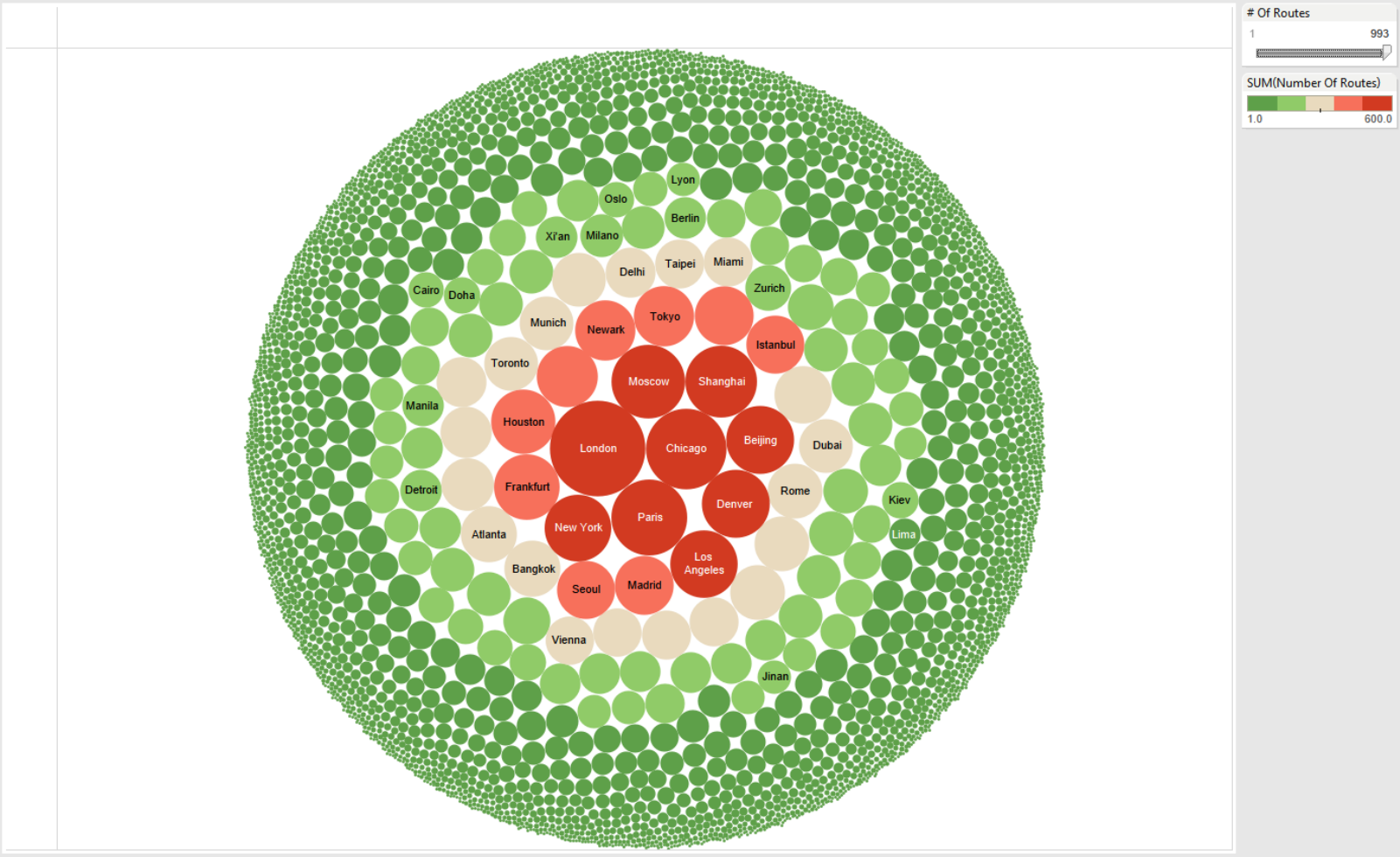
Domestic Arriva



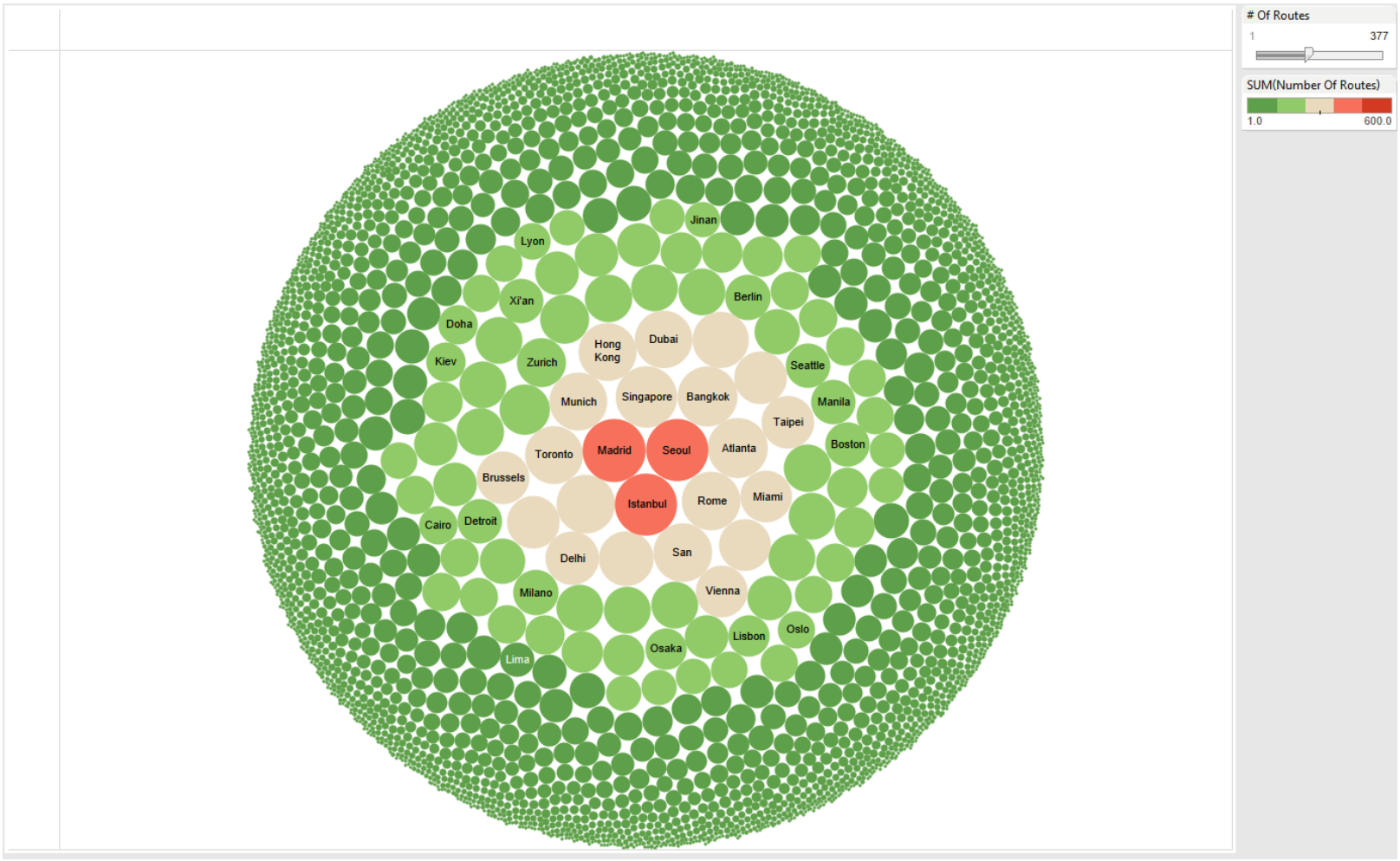
International Arrivals

3. Worksheet: Cities by # Routes (Crystal)

- Summation of the number of routes passing by a city
- It is interesting to lower the number of routes through the quick filter to bring smaller cities in focus



# of routes in cities (no limit on total routes)

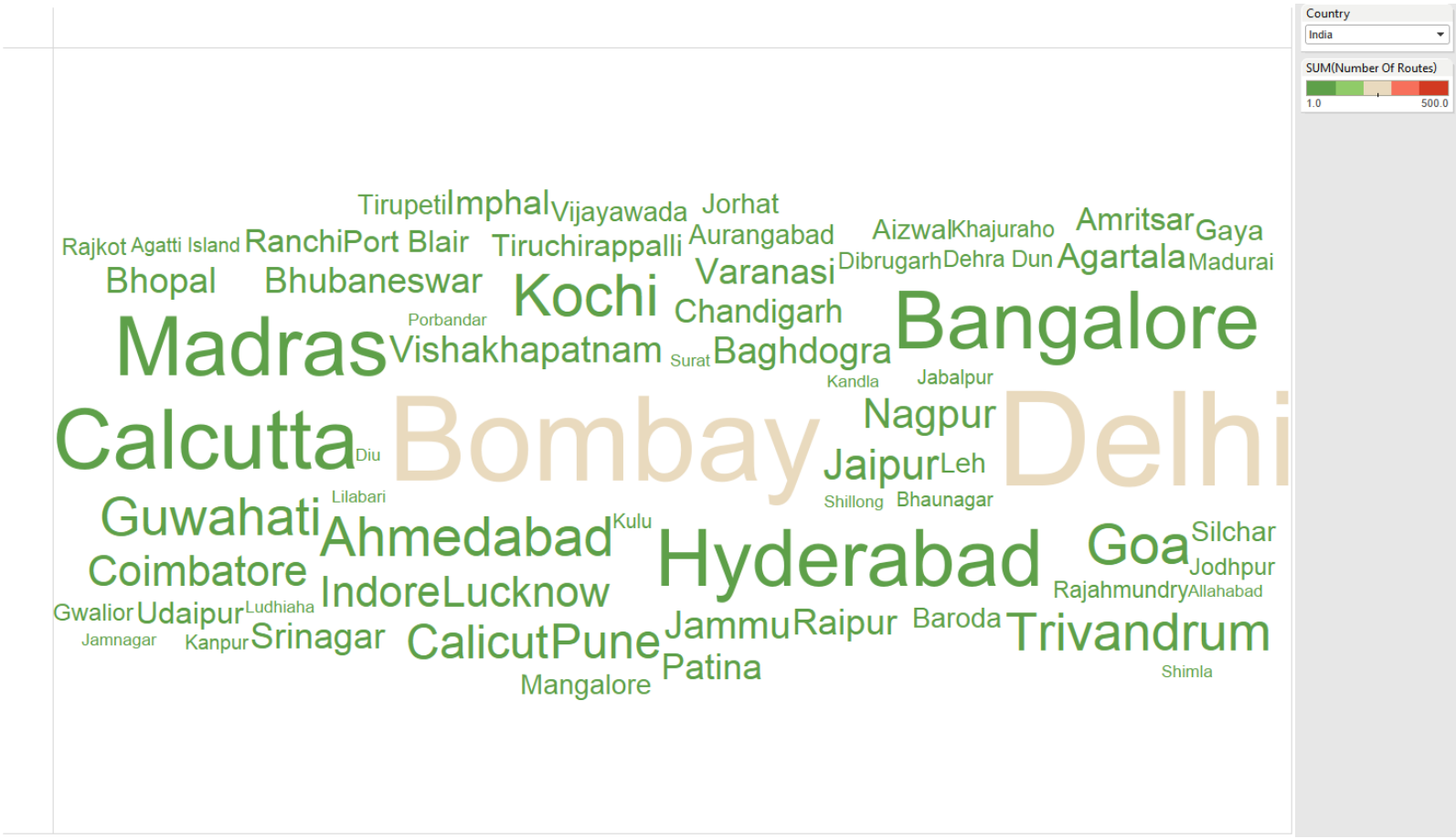


# of routes in cities (limit 377 on total routes)

4. Worksheet: Cities by # Routes (Cloud)

- Summation of the number of routes passing by a city (similar to previous worksheet)
- Trying to fit all the cities in the cloud is in feasible (takes extended amount of time to run), thus I filtered by country

# of routes: cities in United States



# of routes: cities in India

5. Worksheet: City to City 'The Mesh'

- This was the most challenging to produce
- I followed the guide below which required the data in a certain format:

http://kb.tableausoftware.com/articles/knowledgebase/using-path-shelf-pattern-analysis

The original data has the departure and arrival cities in the same record (row) which is incompatible with the required format shown in the diagram.

a. The departure city and arrival cities should be split onto two rows

b. Both rows having the same path ID.

c. Each row would specify a direction using path_order

In order to achieve that structure, I used R to split each record into 2, specifying a path ID and order. Wrote the result in a new file and imported it in tableau. (R code is attached)

Step 1

Create two rows of data for each path you want to trace-the "From" row and the "To" row.

For example, when showing the path of a flight from Berlin to Chrast, you need a row for Berlin and a row for Chrast.

	A	B	C	D	E	F
1	**City**	**Country**	**Path ID**	**Path Order**	**Latitude**	**Longitude**
2	Berlin	UA	Berlin to Chrast	1	52.516667	13.4
3	Chrast	CZ	Berlin to Chrast	2	49.90204971	15.93396
4	Berlin	UA	Berlin to Davle	1	52.516667	13.4
5	Davle	CZ	Berlin to Davle	2	49.89217661	14.40037
6	Berlin	UA	Berlin to Brodek u Prerova	1	52.516667	13.4
7	Brodek u Preova	CZ	Berlin to Brodek u Prerova	2	49.48418986	17.33825
8						

Step 2

For each data pair, create a unique key that identifies them as a pair. In this example, there is a column called Path ID. For the two Berlin to Chrast rows, the Path ID is "Berlin to Chrast."

	A	B	C	D	E	F
1	**City**	**Country**	**Path ID**	**Path Order**	**Latitude**	**Longitude**
2	Berlin	UA	Berlin to Chrast	1	52.516667	13.4
3	Chrast	CZ	Berlin to Chrast	2	49.90204971	15.93396
4	Berlin	UA	Berlin to Davle	1	52.516667	13.4
5	Davle	CZ	Berlin to Davle	2	49.89217661	14.40037
6	Berlin	UA	Berlin to Brodek u Prerova	1	52.516667	13.4
7	Brodek u Preova	CZ	Berlin to Brodek u Prerova	2	49.48418986	17.33825
8						

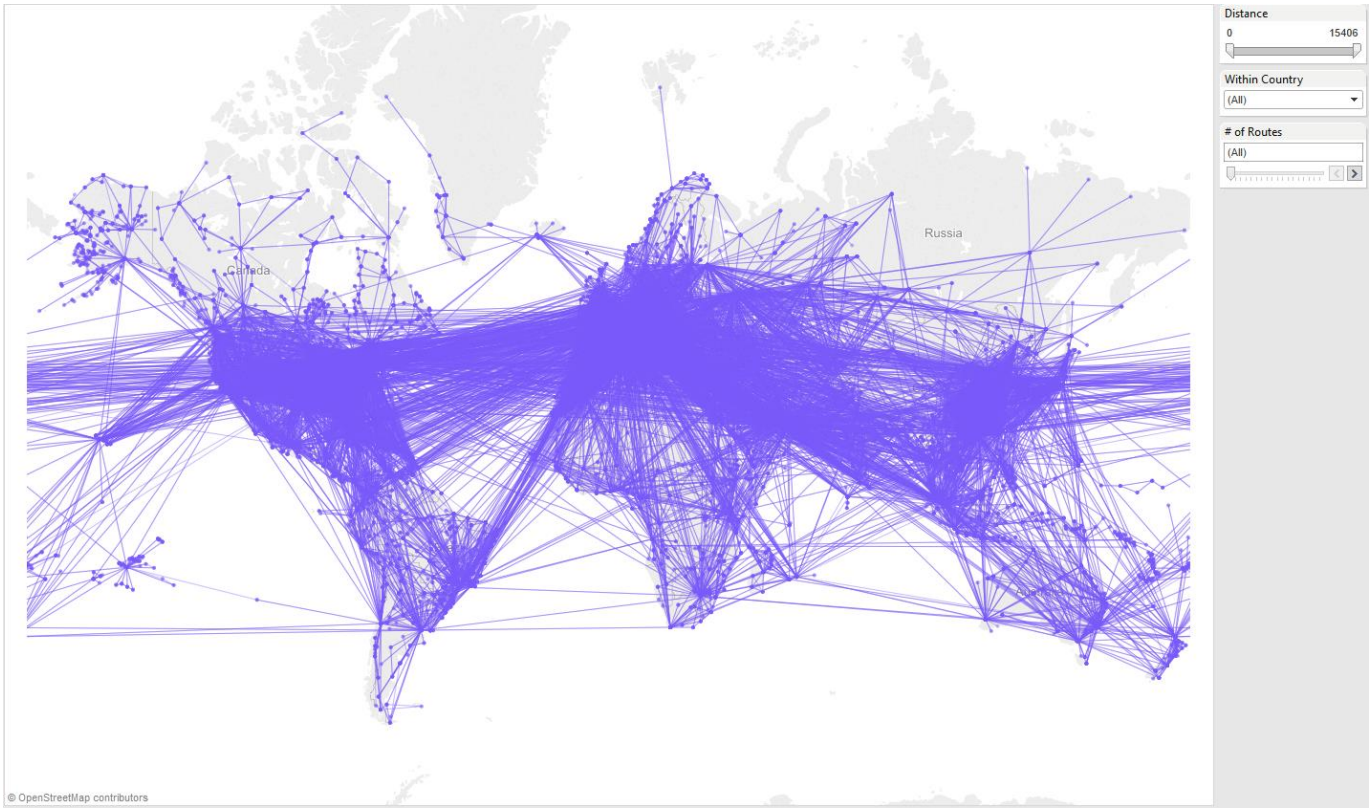
Step 3

Create a column to define the path order: the direction in which the line is drawn.

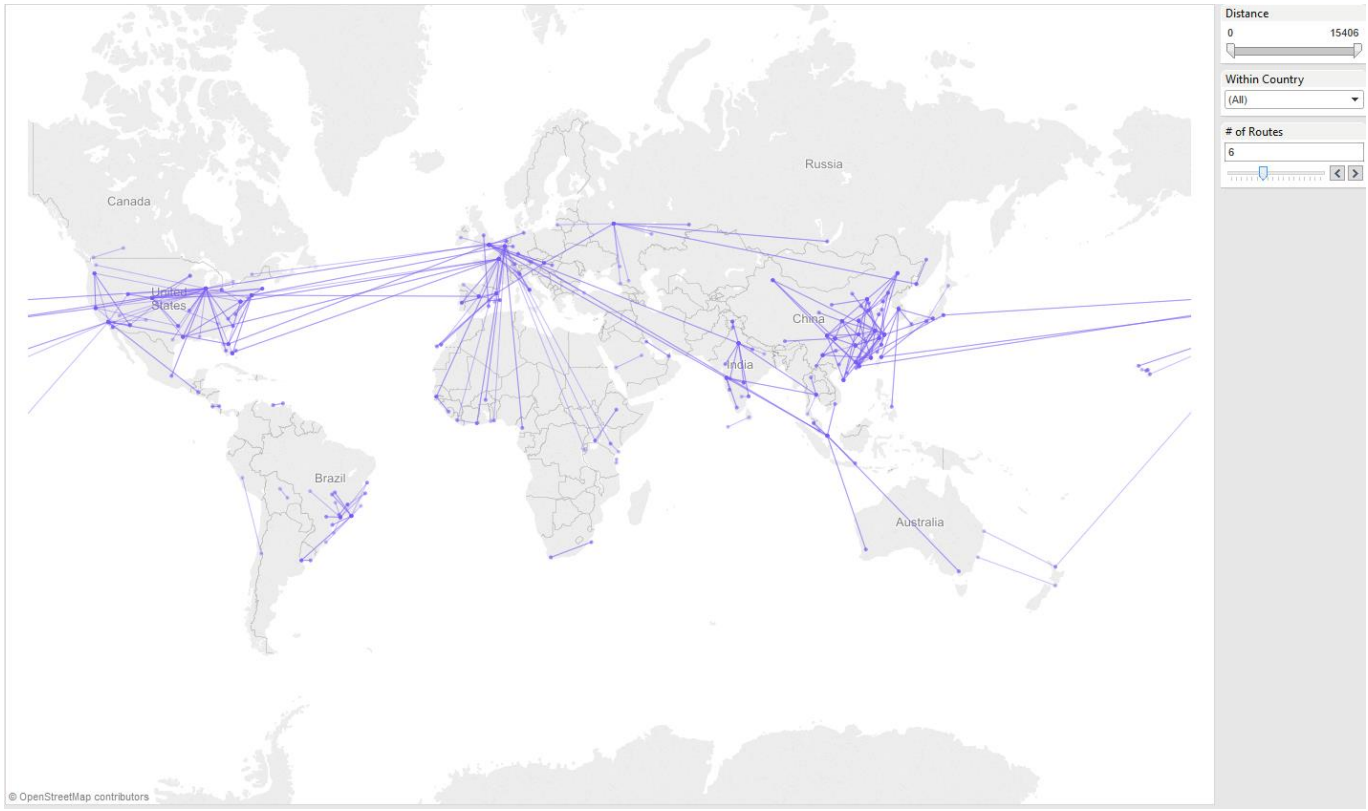
For example, the row for Berlin has a path order of 1 and the row for Chrast has a path order of 2, which indicates that the flight was from Berlin to Chrast.

	A	B	C	D	E	F
1	**City**	**Country**	**Path ID**	**Path Order**	**Latitude**	**Longitude**
2	Berlin	UA	Berlin to Chrast	1	52.516667	13.4
3	Chrast	CZ	Berlin to Chrast	2	49.90204971	15.93396
4	Berlin	UA	Berlin to Davle	1	52.516667	13.4
5	Davle	CZ	Berlin to Davle	2	49.89217661	14.40037
6	Berlin	UA	Berlin to Brodek u Prerova	1	52.516667	13.4
7	Brodek u Preova	CZ	Berlin to Brodek u Prerova	2	49.48418986	17.33825
8						

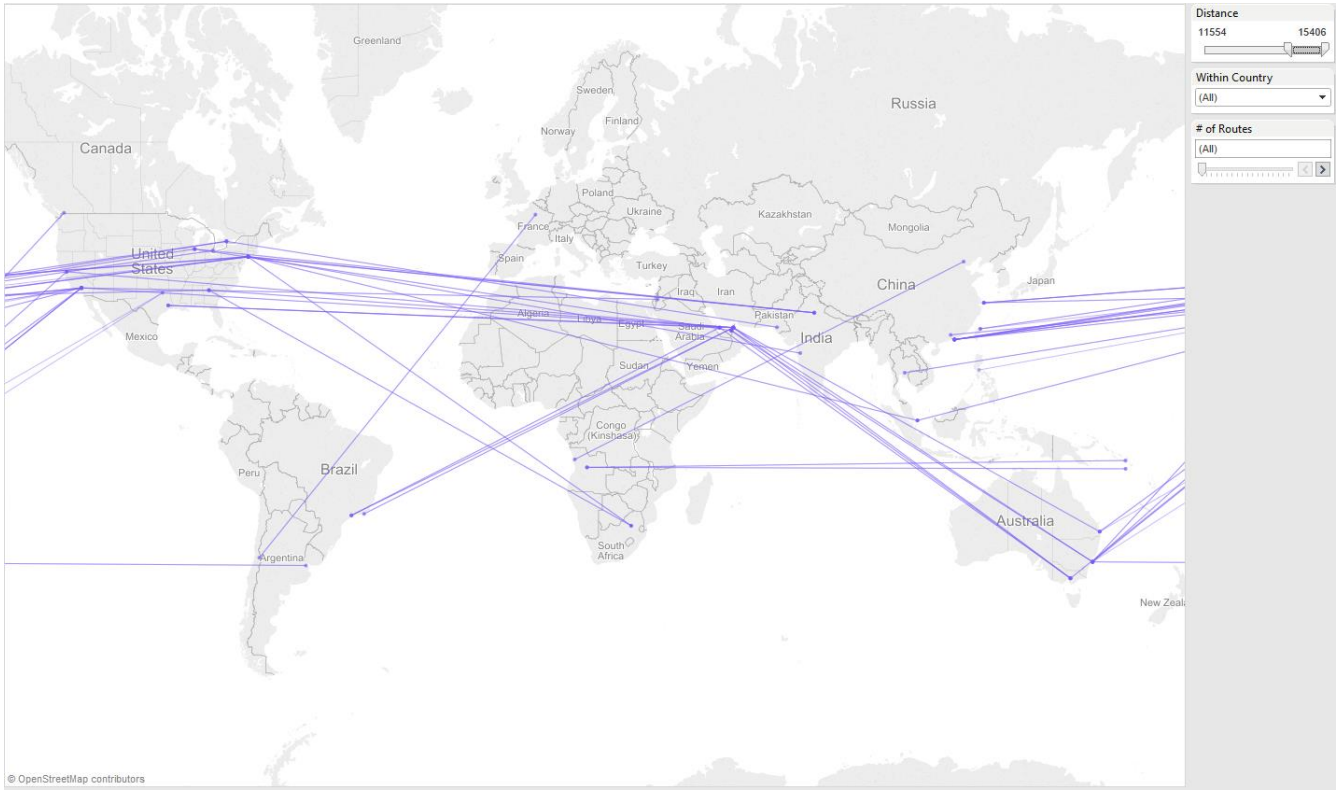
In order to obtain a better visual, limit the route distance using the quick filter, or choose by country. Another filter is the number of routes between the city pair.



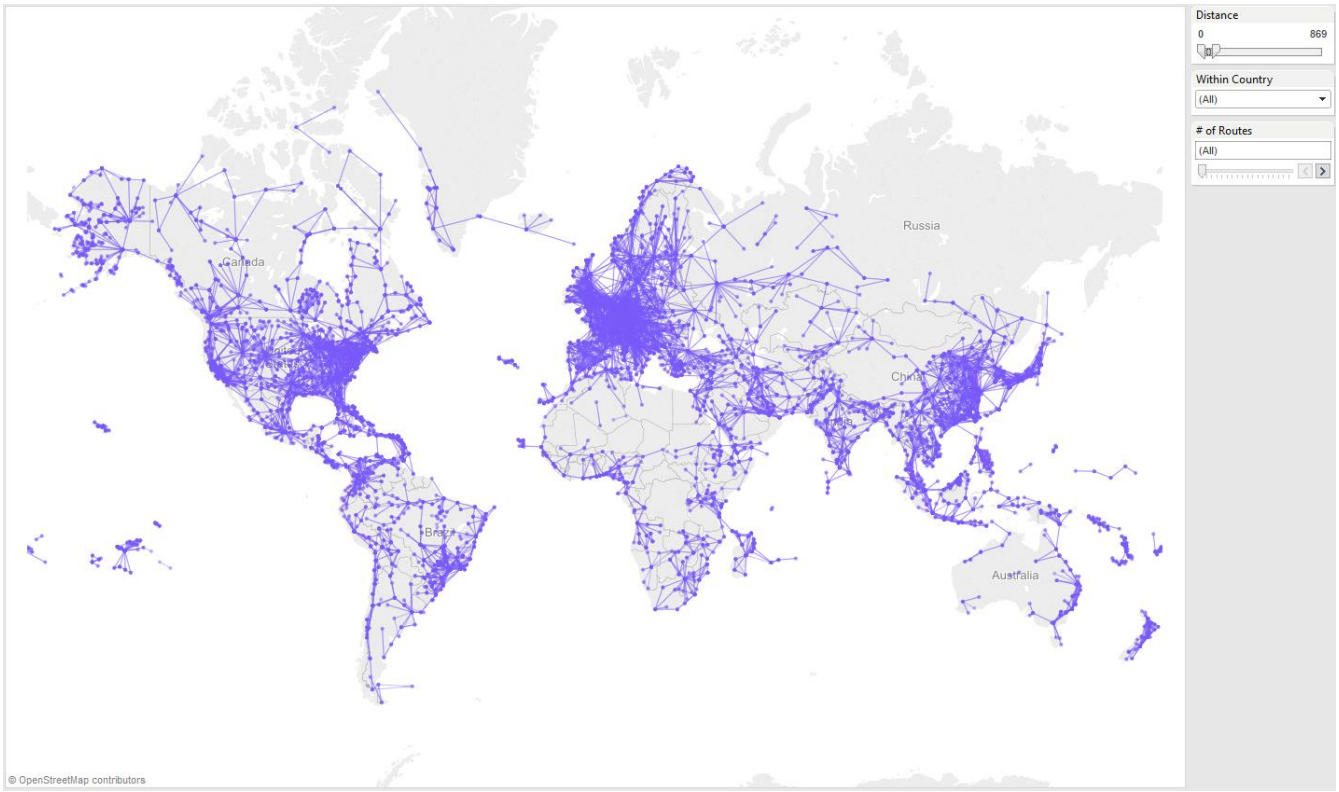
All city pairs, no distance or # of routes limitation



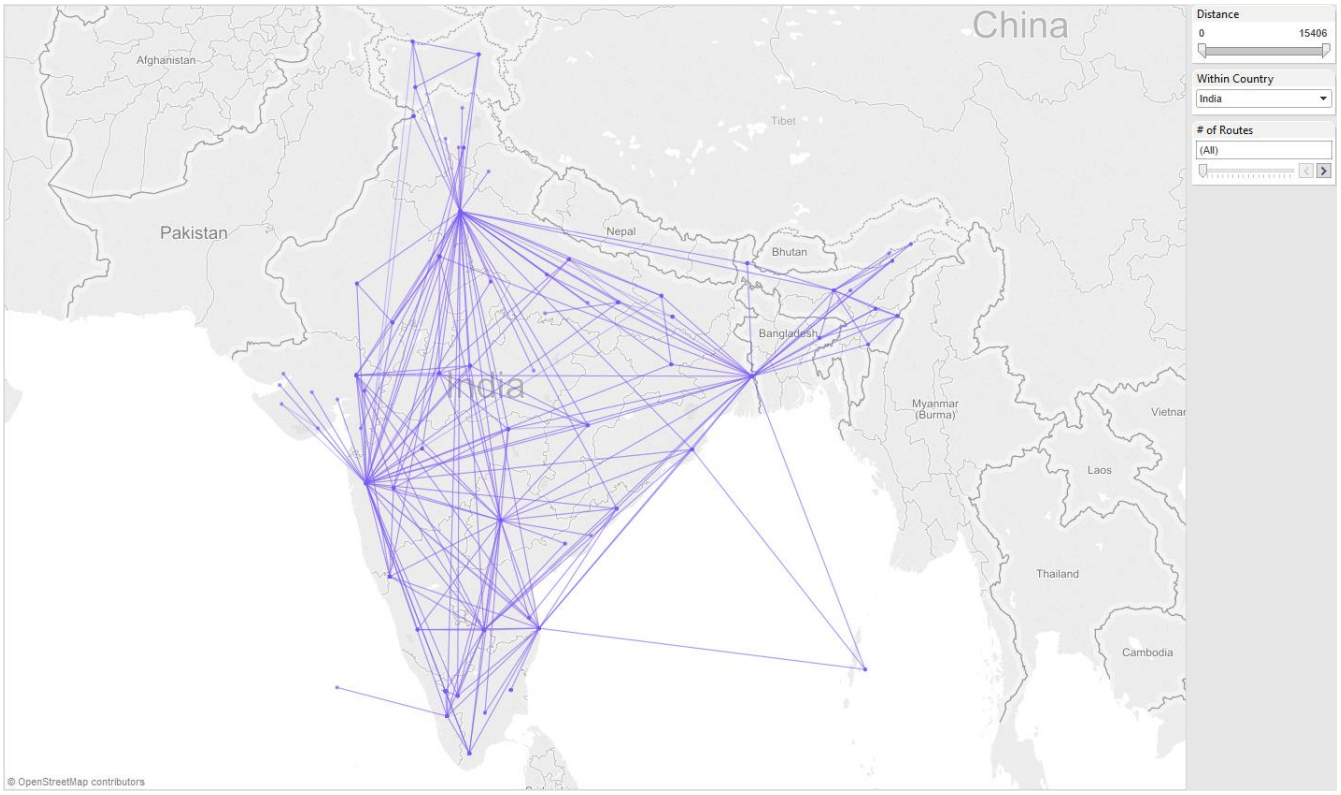
City pairs having 6 routes or more



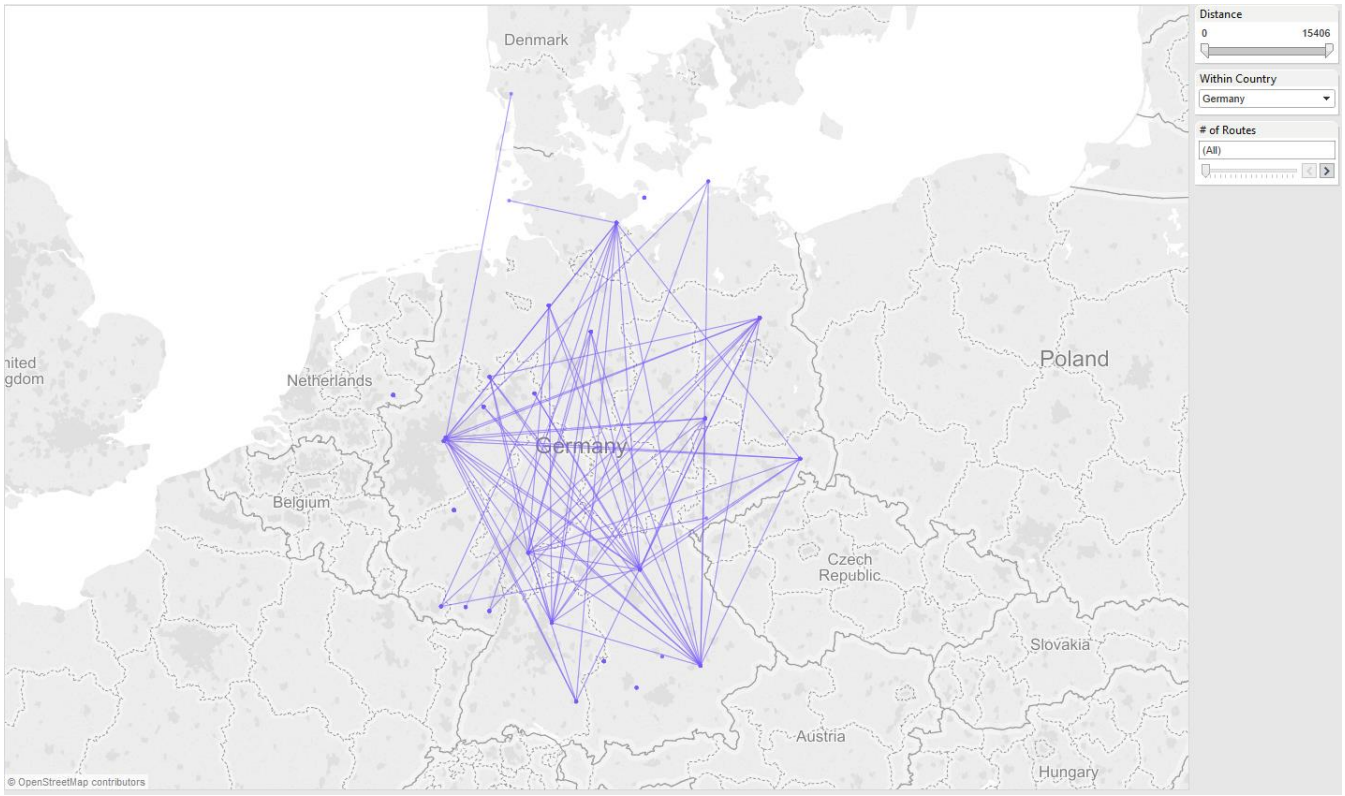
Long distance city pairs



Short distance city pairs



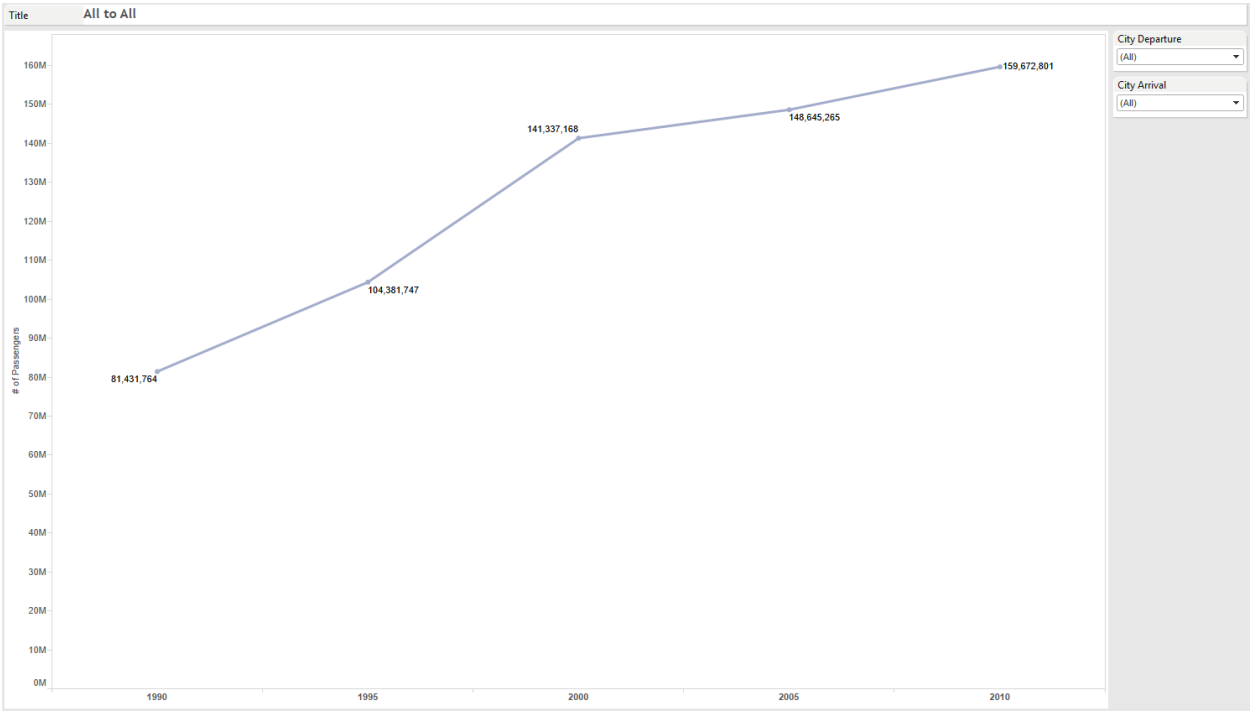
City Pairs in India



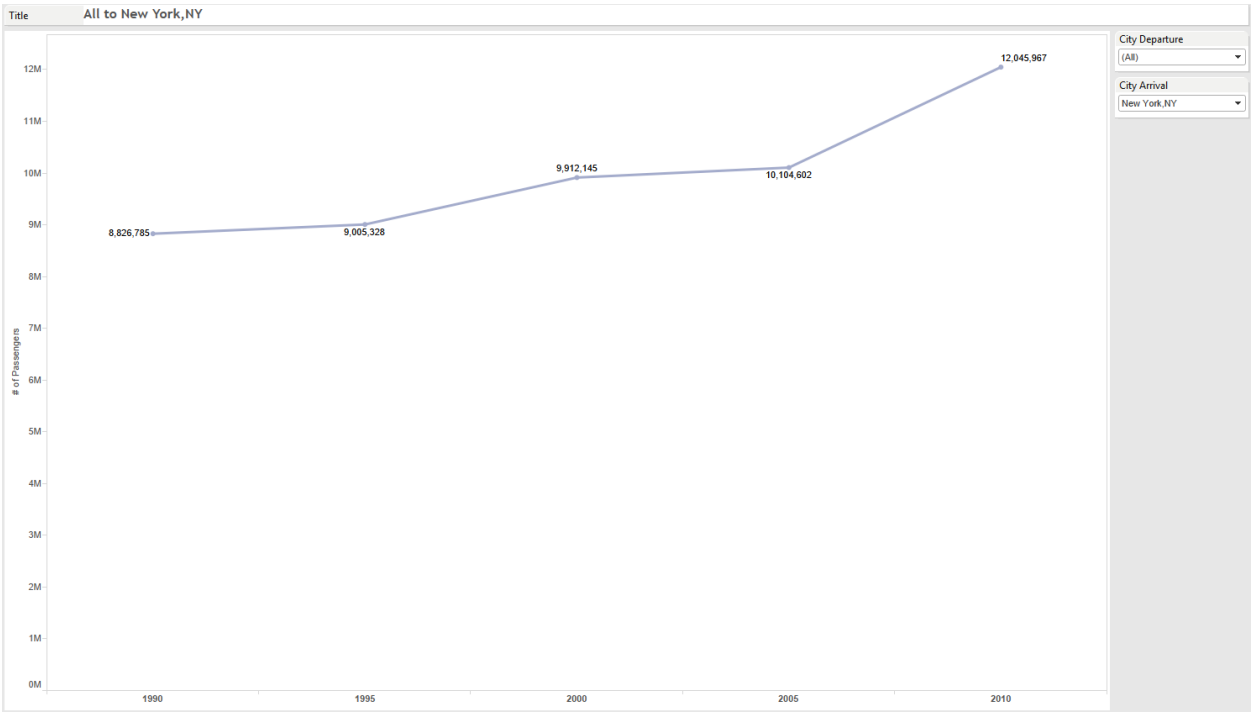
City pairs in Germany

6. Worksheet: Passenger Trends

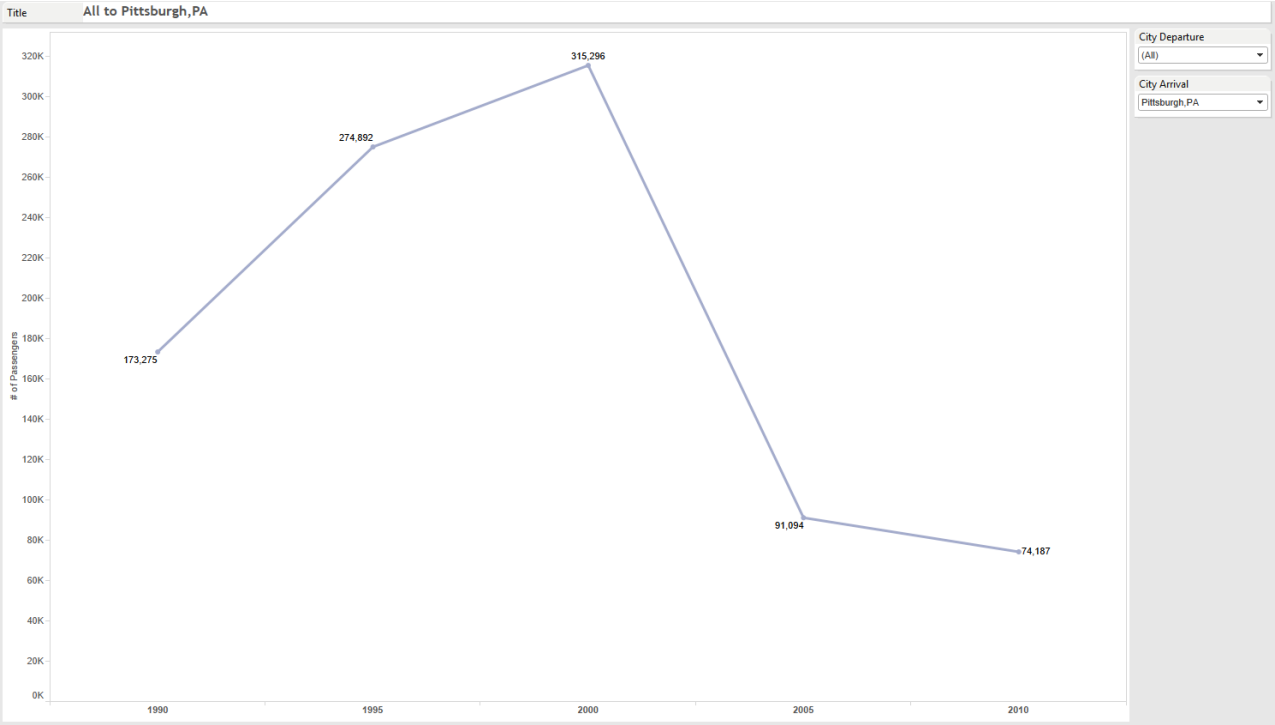
Total number of passengers by year 1990 to 2010 (5 year step), filtered by arrival or departure city



Trend of all passengers



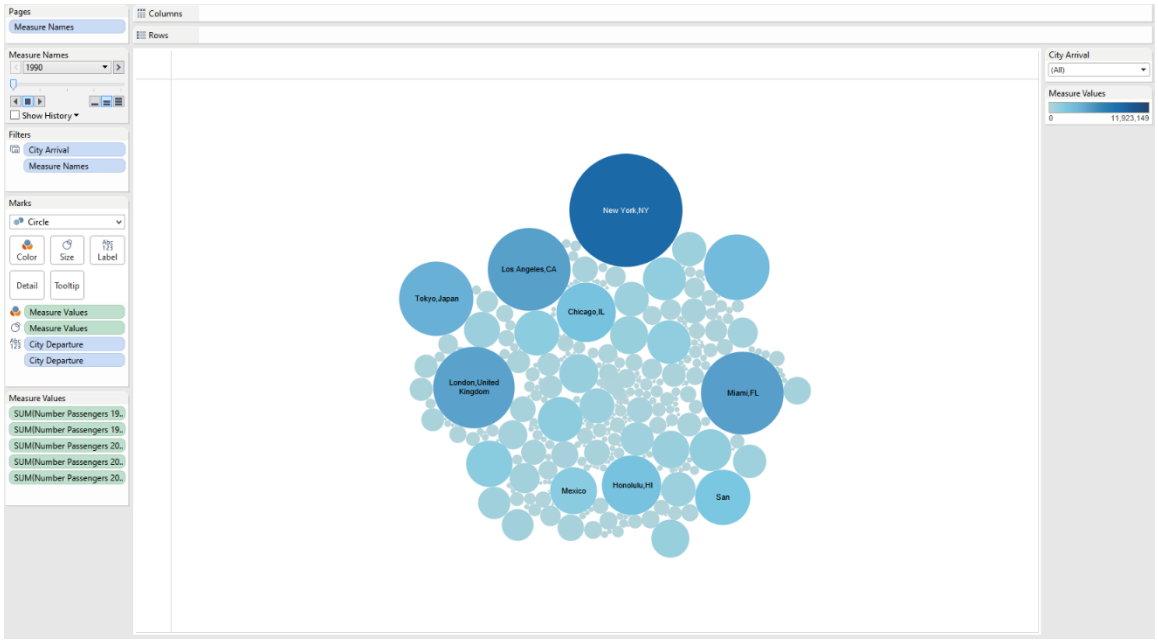
Trend of all passengers going to New York, NY



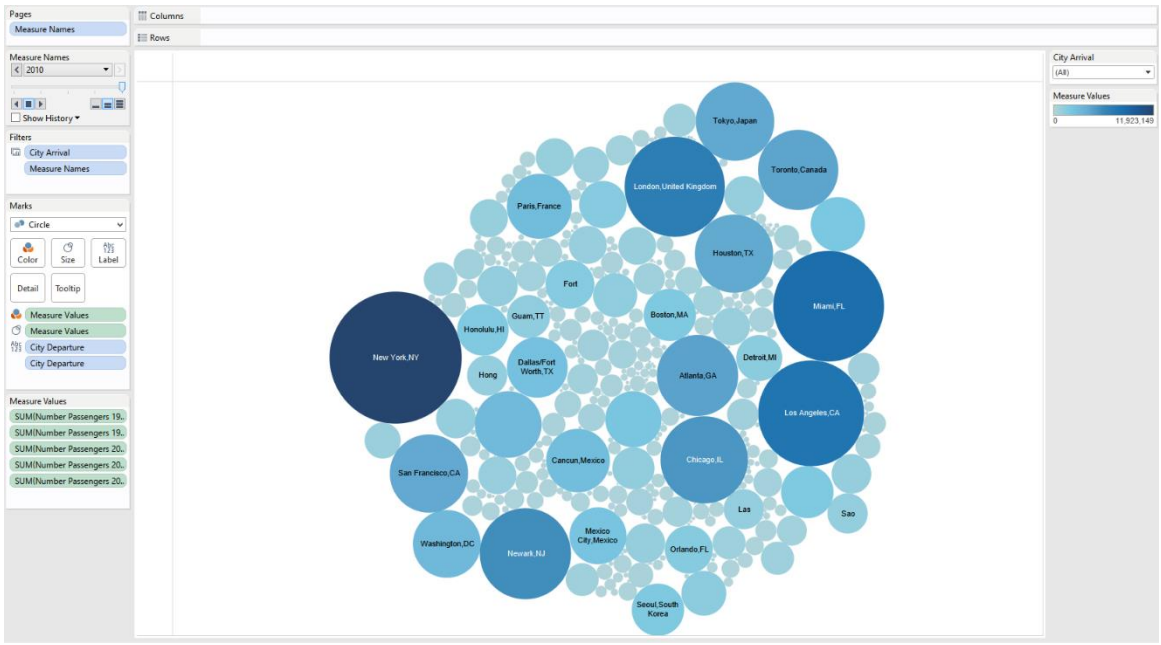
Trend of all passengers going to Pittsburgh, PA

7. Worksheet: Passenger Time Series

Used Pages in order to demonstrate the effect over time for the number of passengers giving from different cities to a given destination city (chosen through the quick filter). Press play on the story or within tableau to view the demo.

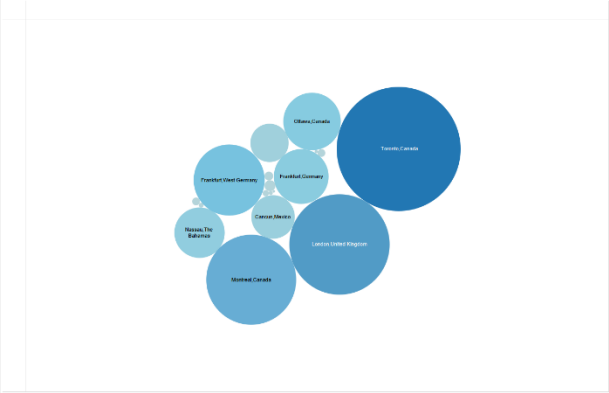


Passengers departing by city to all destinations (1990)

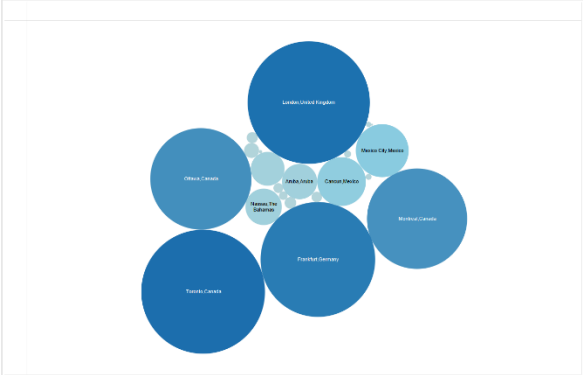


Passengers departing by city to all destinations (2010)

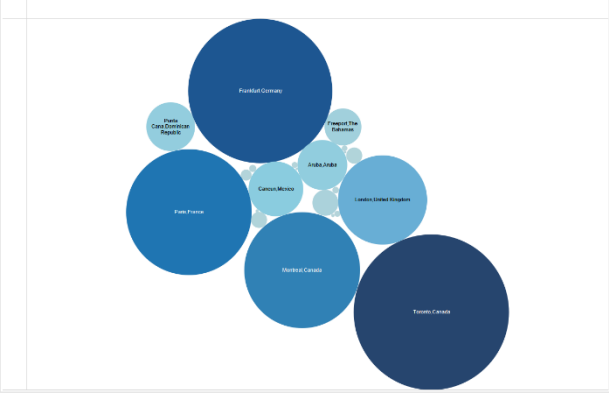
In an attempt to explain the decline in Pittsburgh, PA decline of inbound passengers seen by the previous worksheet, here is the departing passengers to Pittsburgh, PA over time:



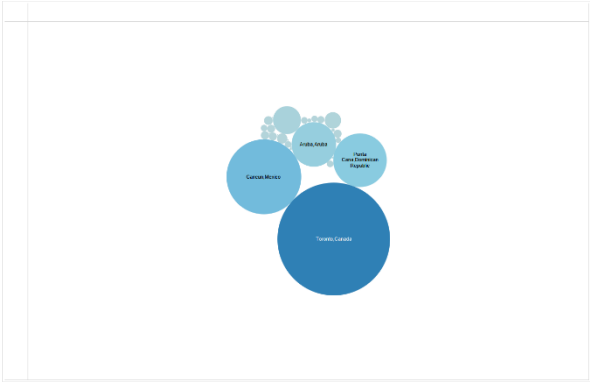
1990



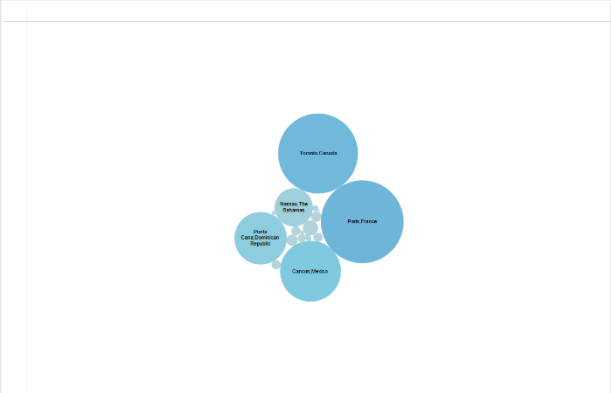
1995



2000



2005



2010

The number of origin cities flying to Pittsburgh, PA as well as total passengers (indicated by the size) have peaked in 2000 then drastically declined until 2010.

Appendix A: Data Description

countries to countries (current active routes)

table (csv)

[country departure name | country arrival name | number of routes]

aprox. dimensions 3x4600

countriesToCountries.csv

countries to countries (current active routes)

network (gml)

~230 nodes, 4600 edges

countriesNetwork.gml

cities to cities (current active routes)

table (csv)

[city departure name | country departure name | long. departure (decimal) | lat. departure (decimal) | city arrival name | country arrival name | long. arrival (decimal) | lat. arrival (decimal) | number of routes | distances (km.)]

aprox. dimensions 5x30000

citiesToCities.csv

all routes (current active routes)

table (csv)Ã¦

[airport departure | city departure | country departure | long. departure (decimal) | lat. departure (decimal) | airport arrival | city arrival | long. arrival (decimal) | lat. arrival (decimal) | airline name | airline country based | route ID | number of stops | distances (km.) | domestic]

aprox. dimensions 12x58000

completeTable.csv

cities (current active routes)

table (csv)

[city name | country name | most active airport | long. airport | lat. airport | number of routes | number of incoming flights | number of outcoming flights | number of incoming domestic | number of outcoming domestic | number of incoming international | number of outcoming international | average distance (km.) | max distance (km.)]

aprox. dimensions 14x3000

citiesTable.csv

cities to cities in time (1990-2010)

table (csv)

[departure city name | arriving city name | number passengers 1990 | number passengers 1995 | number passengers 2000 | number passengers 2005 | number passengers 2010]

aprox. dimensions 7x12000

citiesToCitiesPass.csv