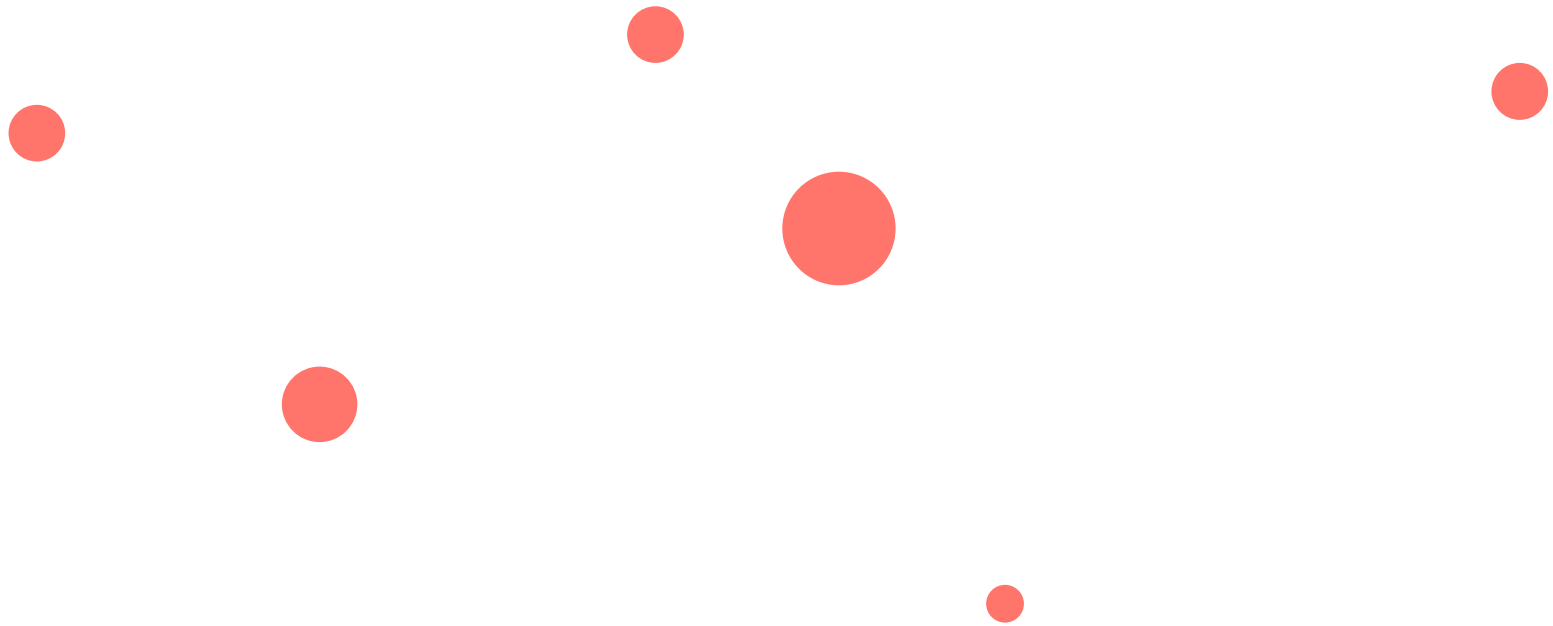
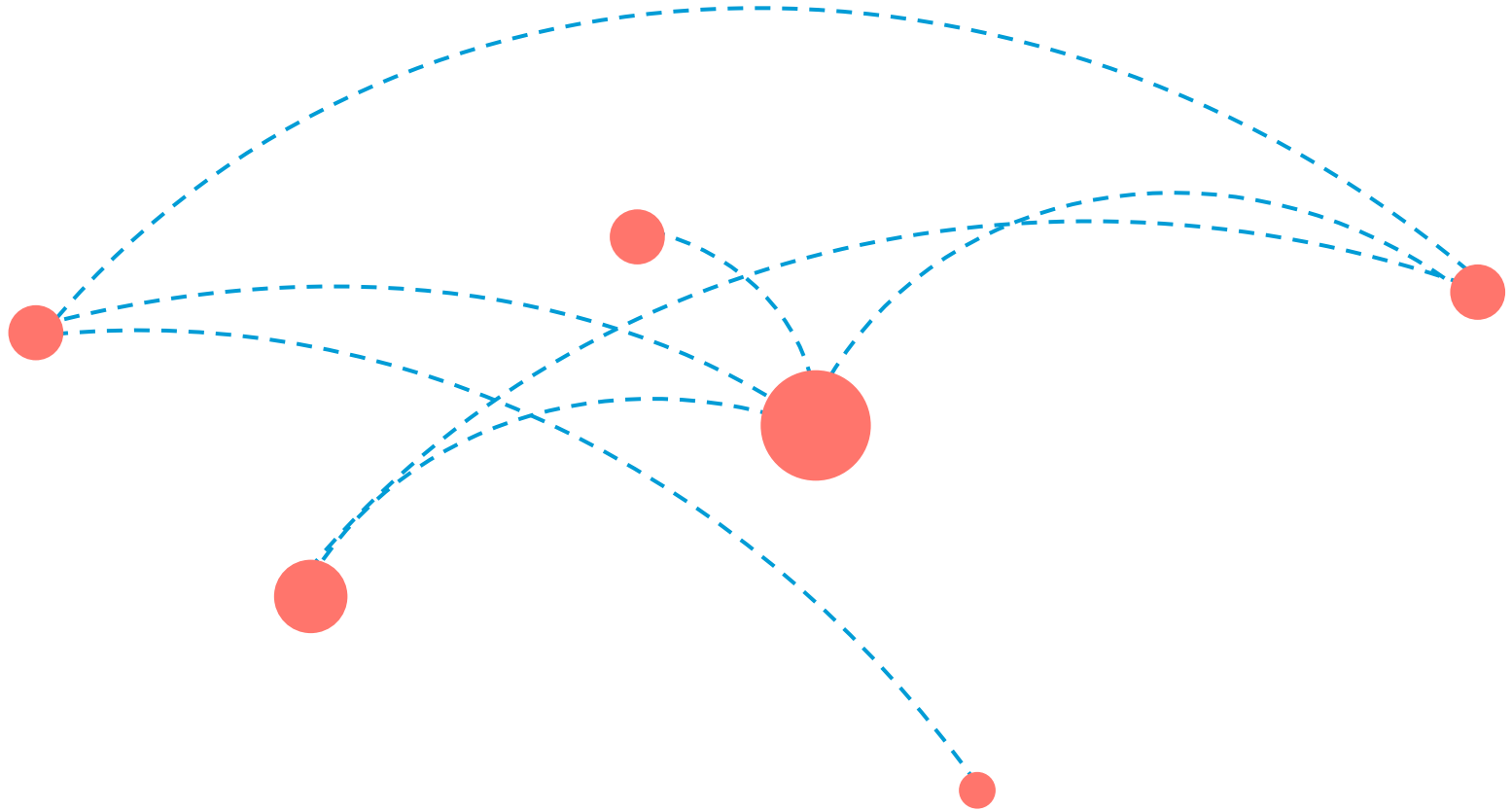
Network Centrality as a Predictor of Airport Passenger Traffic

Charlotte Garcia | Isabelle Grégoire | Daichi Ishikawa | Jedaiah Tan

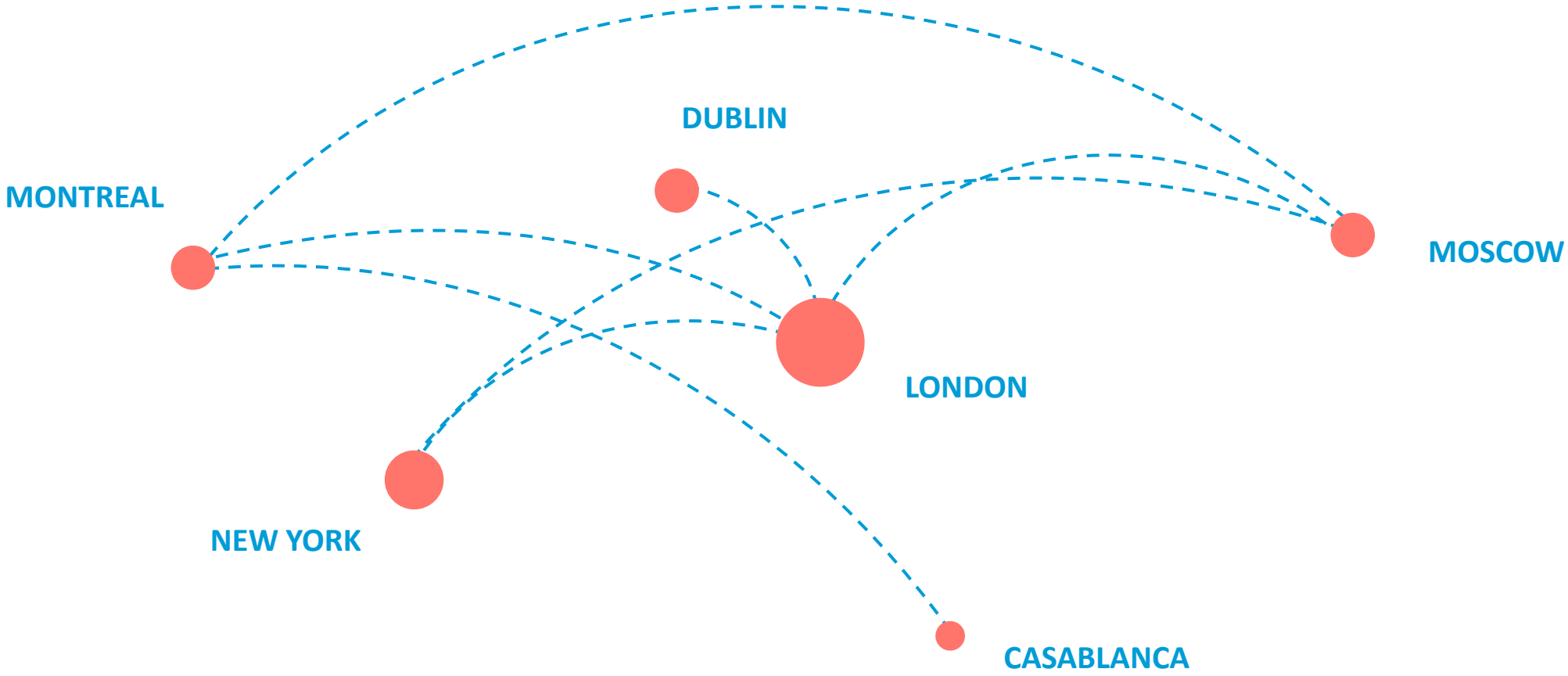
CONTEXT



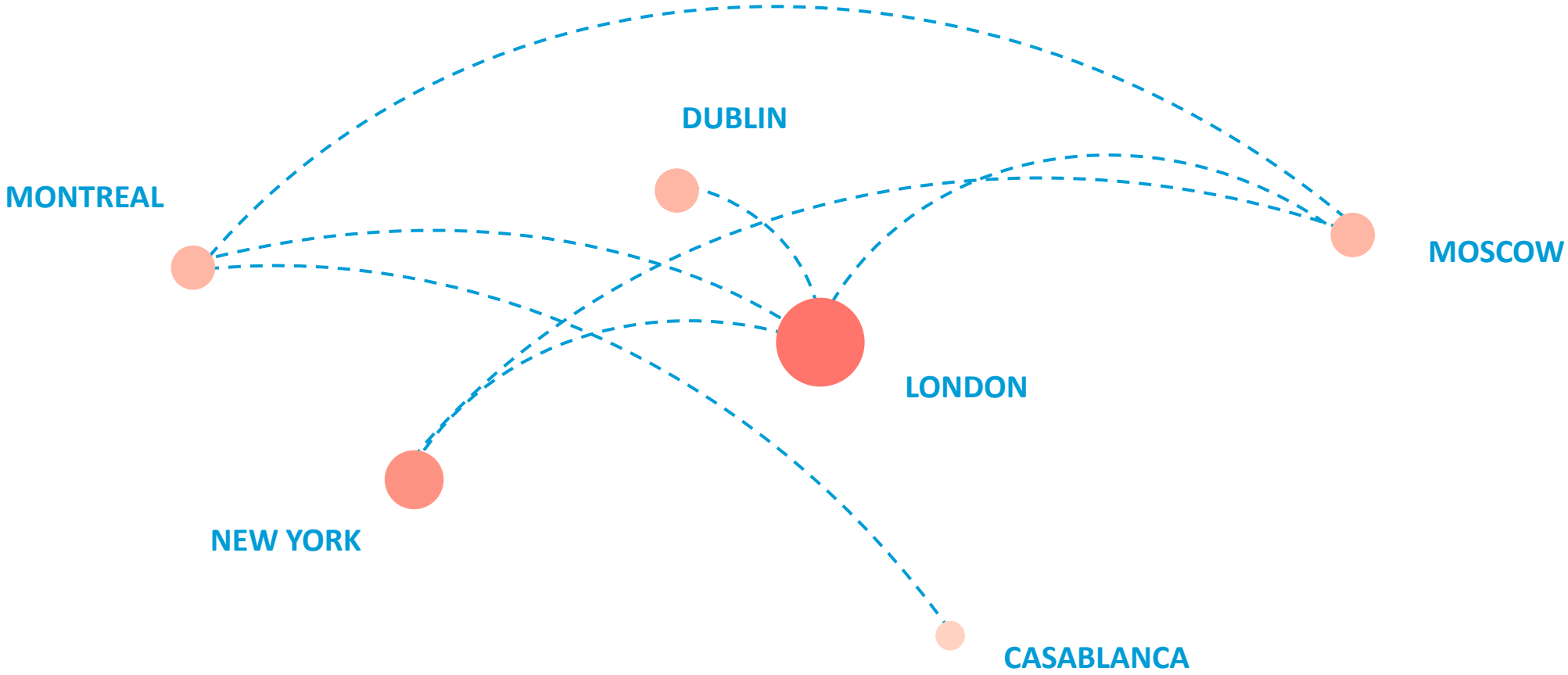
CONTEXT



CONTEXT



CONTEXT



CONTEXT

AIRPORT TRAFFIC = NETWORK CENTRALITY + SOCIO-ECONOMIC FACTORS

HYPOTHESIS

AIRPORT TRAFFIC = NETWORK CENTRALITY + SOCIO-ECONOMIC FACTORS

The more important the airport is in the network, the more flights it will have, and the more passengers use the airport.

- Degrees, Eigenvector, Betweenness, Connectedness, PageRank
- 6977 airports, 59,036 flight routes, 531 airlines
- Openflights.org, US Federal Aviation Administration, US Bureau of Transportation Statistics

HYPOTHESIS

AIRPORT TRAFFIC = NETWORK CENTRALITY + SOCIO-ECONOMIC FACTORS

The higher the disposable income of surrounding population, the more they will travel and use the airport.

- Ideally constrained geographically to 2-3 hour driving radius of airport
- State GDP per capita, population, geographic data
- US Bureau of Economic Analysis, US Census Bureau.

INITIAL MODEL DEVELOPMENT

Initial Regression model explained 85% of variation in passenger traffic, but sacrificed accuracy and efficiency.

MODEL PERFORMANCE METRICS

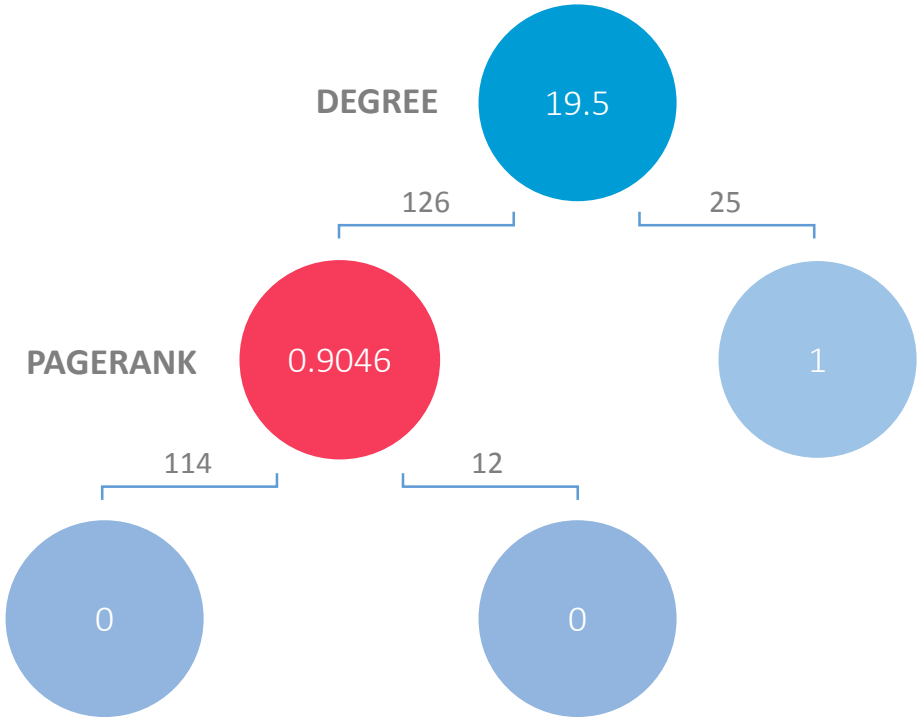
Metric	Value
Residual DF	215
R^2	0.845958678
Adjusted R^2	0.838077495
Std. Error Estimate	2178156.46
RSS	1.02E+15

REGRESSION VARIABLES OUTPUT

Input Variables	Coefficient	Std. Error	P-Value	RSS Reduction (in MM's)
Intercept	-573588.7439	1232598.163	0.642152308	831,610.00
State GDP per Capita Cat.	10.67138133	20.15794071	0.597082173	4,625.25
Latitude	-36160.03963	23485.84356	0.125115514	138,349.00
Longitude	-6295.962901	7443.425333	0.398581192	18,103.00
Degree	286871.599	35787.01387	6.91E-14	5,372,330.00
Betweenness Centrality	-180.7509793	232.7480086	0.438251184	8,757.40
Closeness Centrality	105127.1813	2291704.459	0.963454115	299.34
Eigenvector Centrality	-443074552.4	192114854.3	0.022046504	53,167.30
Clustering Coefficient	-213909.8544	410567.5534	0.60289591	1,645.92
Airport Size_ord	-6627.688737	443505.6719	0.988090823	538.81
Population Cat._ord	101345.1902	110507.3985	0.360123242	3,973.01
Cluster Group_ord	8062.574796	128913.582	0.95018894	18.56

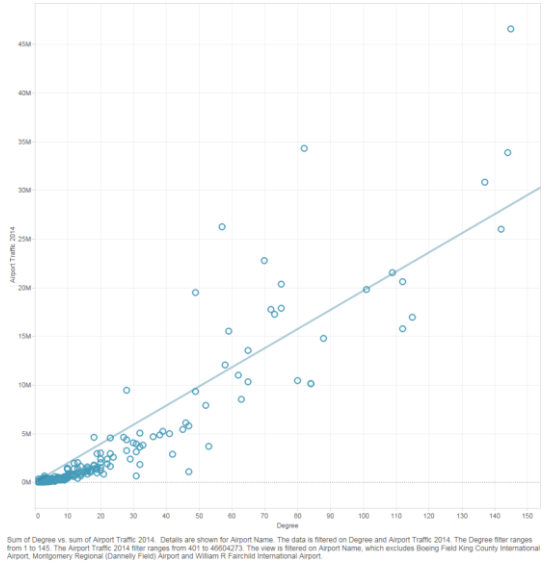
ANALYSIS OF PREDICTOR VARIABLES: DEGREE & PAGERANK

BEST PRUNED & MIN. ERROR TREE



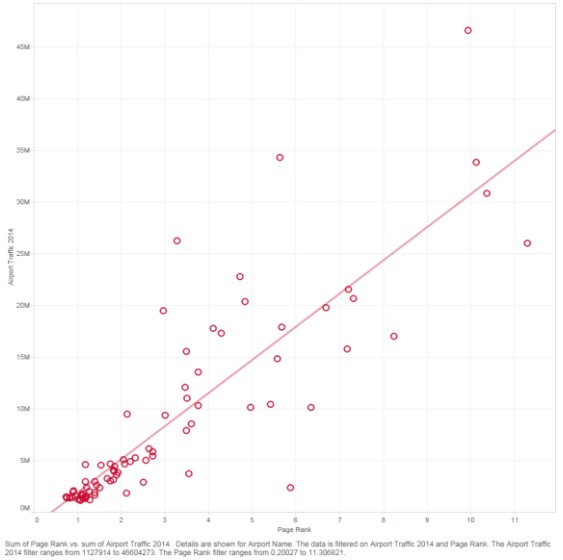
Degree and PageRank coefficients were found to be the most influential predictor variables for airport passenger traffic.

DEGREE COEFFICIENT & TRAFFIC



Sum of Degree vs. sum of Airport Traffic 2014. Details are shown for Airport Name. The data is filtered on Degree and Airport Traffic 2014. The Degree filter ranges from 1 to 145. The Airport Traffic 2014 filter ranges from 401 to 46604273. The view is filtered on Airport Name, which excludes Boeing Field King County International Airport, Montgomery Regional (Dannelly Field) Airport and William R Fairchild International Airport.

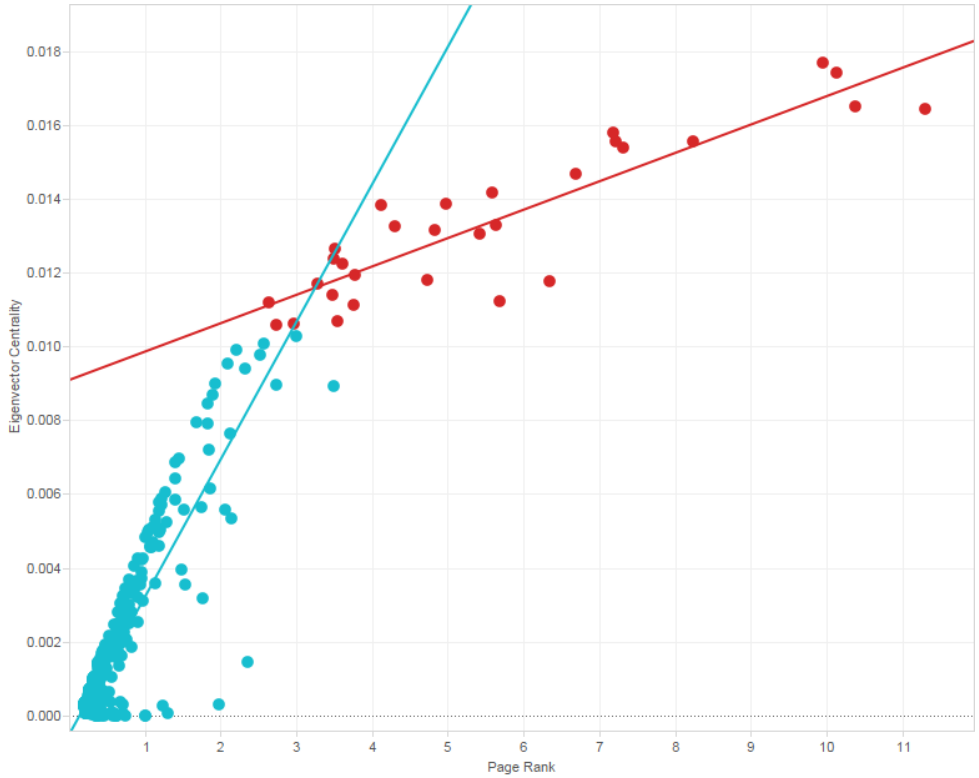
PAGERANK SCORE & TRAFFIC



Sum of Page Rank vs. sum of Airport Traffic 2014. Details are shown for Airport Name. The data is filtered on Airport Traffic 2014 and Page Rank. The Airport Traffic 2014 filter ranges from 1127914 to 46604273. The Page Rank filter ranges from 0.20027 to 11.306821.

ANALYSIS OF PREDICTOR VARIABLES: EIGENVECTOR

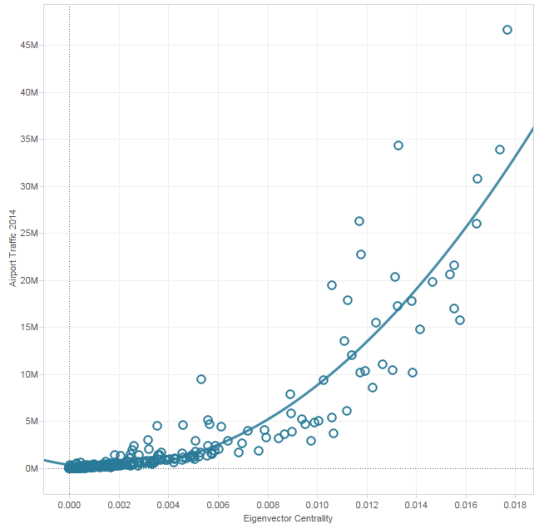
RELATIONSHIP BETWEEN EIGENVECTOR AND PAGERANK



Sum of Page Rank vs. sum of Eigenvector Centrality. Color shows details about Airport Name (group). Details are shown for Airport Name. The view is filtered on Airport Name, which excludes Orlando Sanford International Airport and Ted Stevens Anchorage International Airport.

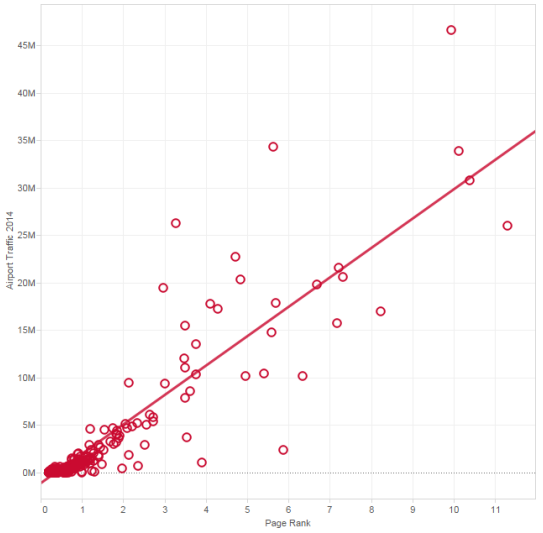
PageRank is a better measure of centrality than Eigenvector for airports due to dampening factor in PR algorithm

EIGENVECTOR & TRAFFIC



Sum of Eigenvector Centrality vs. sum of Airport Traffic 2014. Details are shown for Airport Name.

PAGERANK SCORE & TRAFFIC



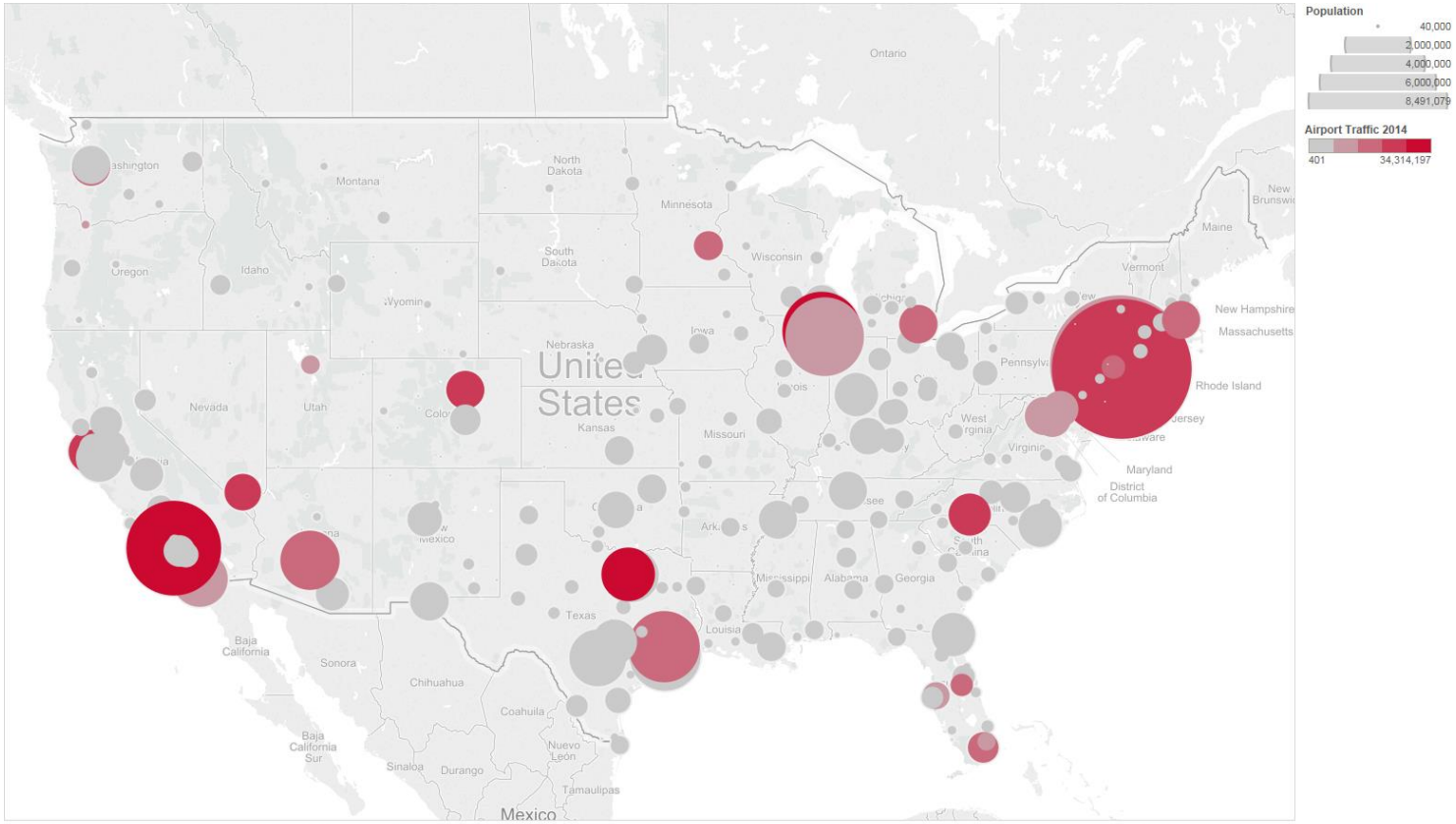
Sum of Page Rank vs. sum of Airport Traffic 2014. Details are shown for Airport Name.

ANALYSIS OF PREDICTOR VARIABLES

There is some correlation between population and airport passenger traffic, but it is not a primary driver.

- On average, a larger population increases airport traffic.
- Potential for stronger correlation if coexistence of multiple airports in a large city was taken into account.

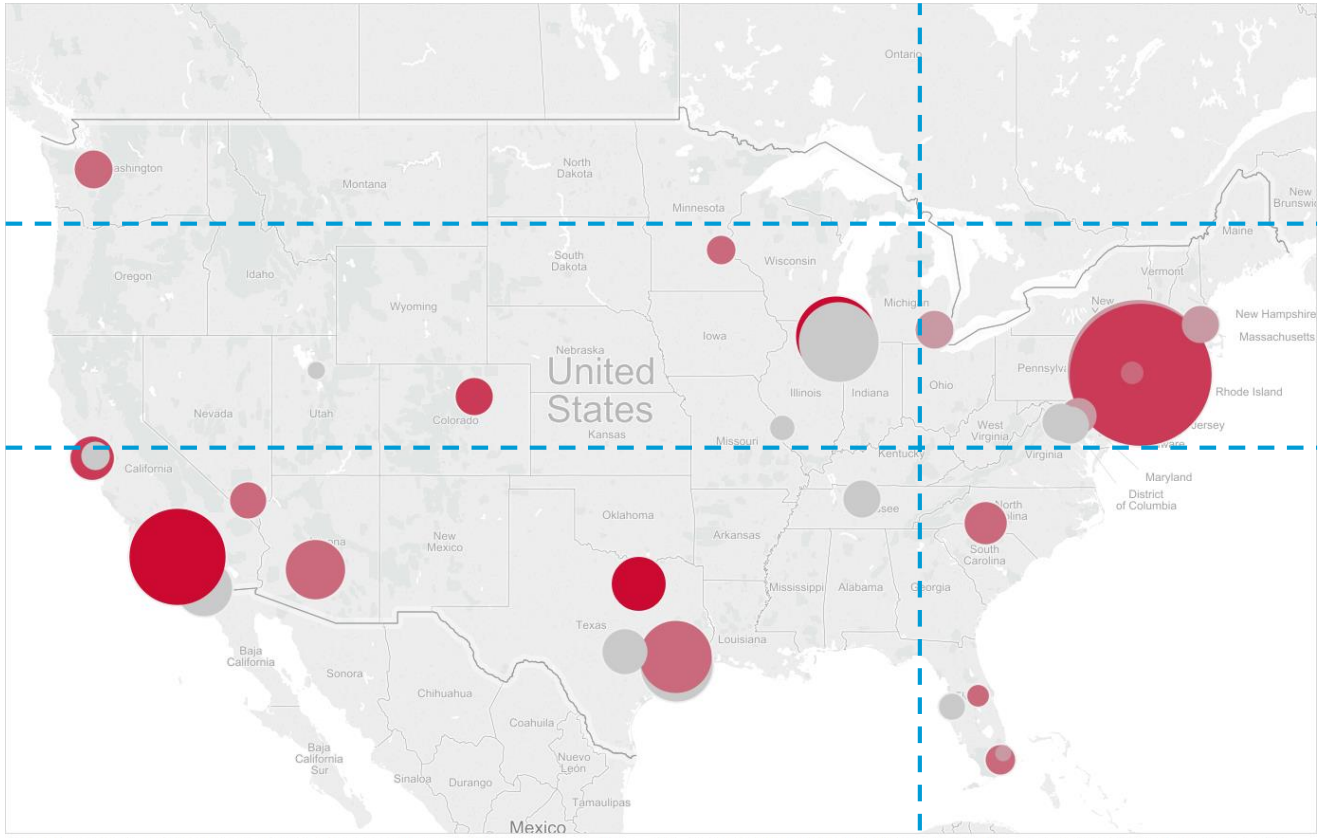
US AIRPORTS BY PASSENGER TRAFFIC AND AREA POPULATION



Map based on average of Longitude and average of Latitude. Color shows sum of Airport Traffic 2014. Size shows sum of Population. Details are shown for State and Airport Name. The data is filtered on Airport Traffic 2014, which ranges from 401 to 45852598. The view is filtered on sum of Population, which ranges from 40,000 to 8,491,079.

ANALYSIS OF PREDICTOR VARIABLES

AIRPORTS WITH OVER 5M IN ANNUAL PASSENGER TRAFFIC



Map based on average of Longitude and average of Latitude. Color shows sum of Airport Traffic 2014. Size shows sum of Population. Details are shown for State and Airport Name. The data is filtered on Airport Traffic 2014, which ranges from 5000000 to 45852598. The view is filtered on sum of Population, which ranges from 40,000 to 8,491,079.

REGRESSION OUTPUT FOR LATITUDE AND LONGITUDE

Input Variable	Std. Error	P-Value	RSS Reduction (in MM's)
Latitude	23485.84356	0.125115514	138,349.00
Longitude	7443.425333	0.398581192	18,103.00

Latitude is more influential than Longitude in predicting airport passenger traffic.

- Likely to do with location of larger population cities along coasts rather than inherent geolocation factors.

ANALYSIS OF PREDICTOR VARIABLES: AIRPORT SIZE

Airport size category exhibited high variance, possibly due to more diversity in the types of flights at large airports.

- Medium sized airports see considerably less traffic than large sized airports
- The high variance in passenger traffic for large-sized airports may be attributed to non-passenger flights, such as cargo and military.

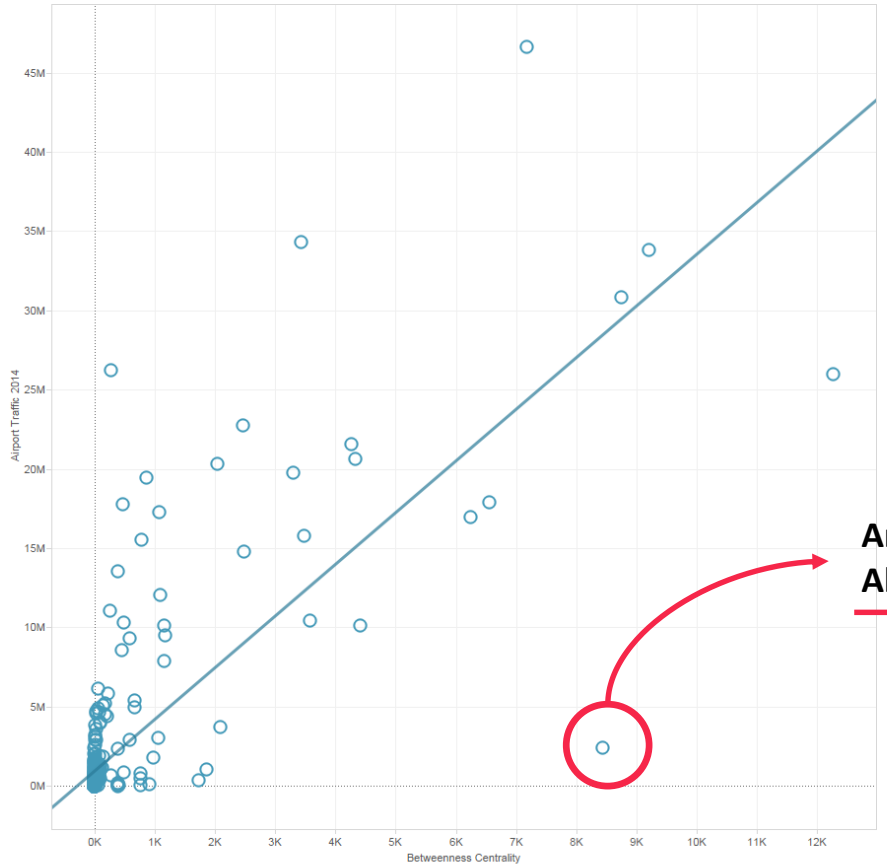
PASSENGER TRAFFIC AT LARGE AND MEDIUM SIZED AIRPORTS

State and Airport Name. Color shows details about Airport Size. Size shows sum of Airport Traffic 2014. The marks are labeled by State and Airport Name. Details are shown for Airport Size. The data is filtered on Airport Traffic 2014, which ranges from 401 to 46604273. The view is filtered on Airport Size, which keeps large and medium.

Airport Size
large
medium

ANALYSIS OF PREDICTOR VARIABLES: BETWEENNESS

PASSENGER TRAFFIC AT LARGE AND MEDIUM SIZED AIRPORTS



Sum of Betweenness Centrality vs. sum of Airport Traffic 2014. Details are shown for Airport Name. The data is filtered on Betweenness Centrality and Airport Traffic 2014. The Betweenness Centrality filter ranges from 0 to 12261.804286. The Airport Traffic 2014 filter ranges from 401 to 46604273.

The significance of betweenness coefficient in the regression model is weak, and not a good predictor variable.

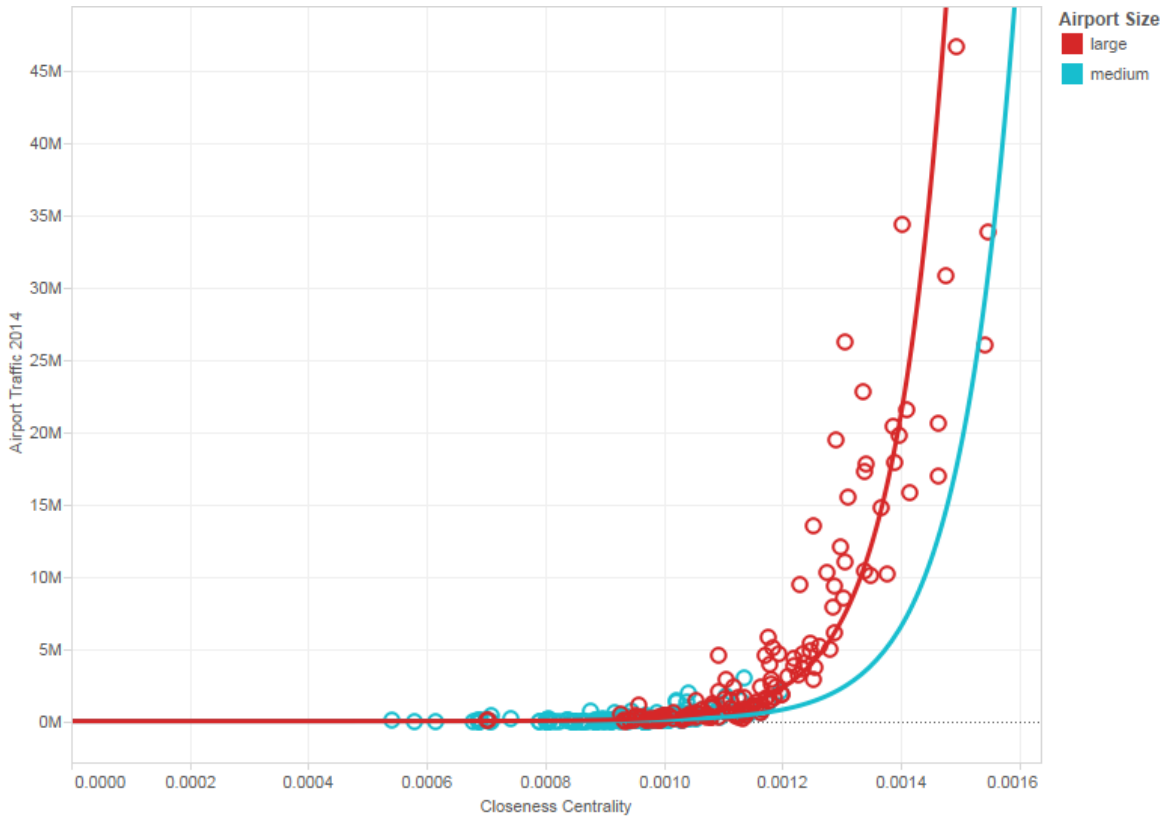
- Measures the role of an airport as a broker in the network
- Apparent randomness may be attributed to the diversity in airport traffic – i.e. cargo, military, etc.
- Missing data on frequency of flights on a given route also contributing to low correlation

ANALYSIS OF PREDICTOR VARIABLES: CLOSENESS

The relationship between closeness and passenger traffic is exponential, not linear.

- Weak relationship in initial MLR explained as exponential variable forced to fit linear model
- Definition: inverse sum of distance between immediate surrounding nodes
- Significant predictor variable in MLR after exponential modification

PASSENGER TRAFFIC AND CLOSENESS CENTRAILITY COEFFICIENT



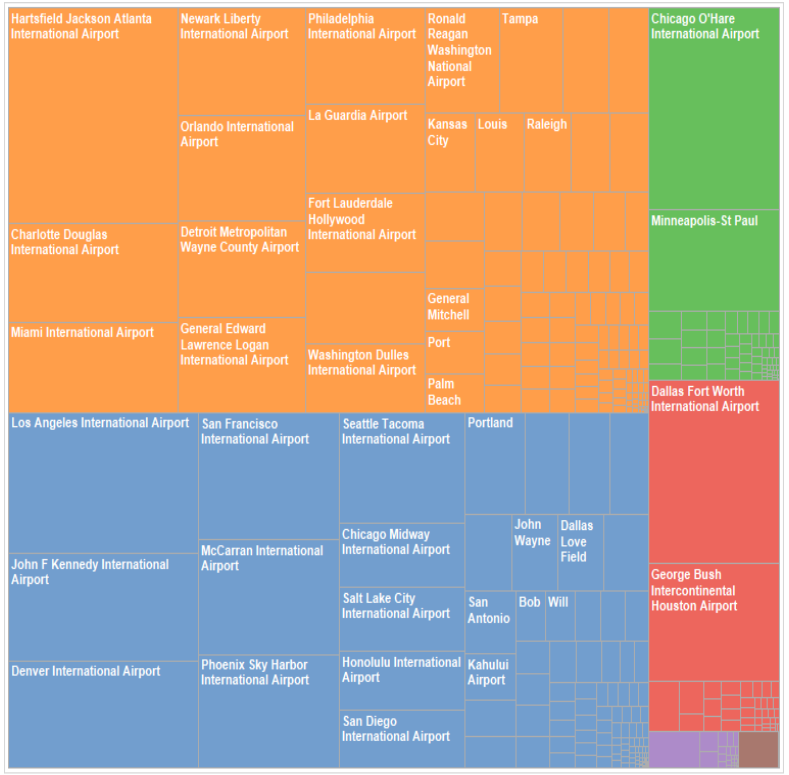
Sum of Closeness Centrality vs. sum of Airport Traffic 2014. Color shows details about Airport Size. Details are shown for Airport Name. The data is filtered on Closeness Centrality and Airport Traffic 2014. The Closeness Centrality filter ranges from 0.000542 to 1. The Airport Traffic 2014 filter ranges from 401 to 46604273. The view is filtered on Airport Name, which excludes Boeing Field King County International Airport, Montgomery Regional (Dannelly Field) Airport and William R Fairchild International Airport.

ANALYSIS OF PREDICTOR VARIABLES: CLUSTER, GDP

Cluster groups and state GDP per capita was not found to be significantly influential in our model.

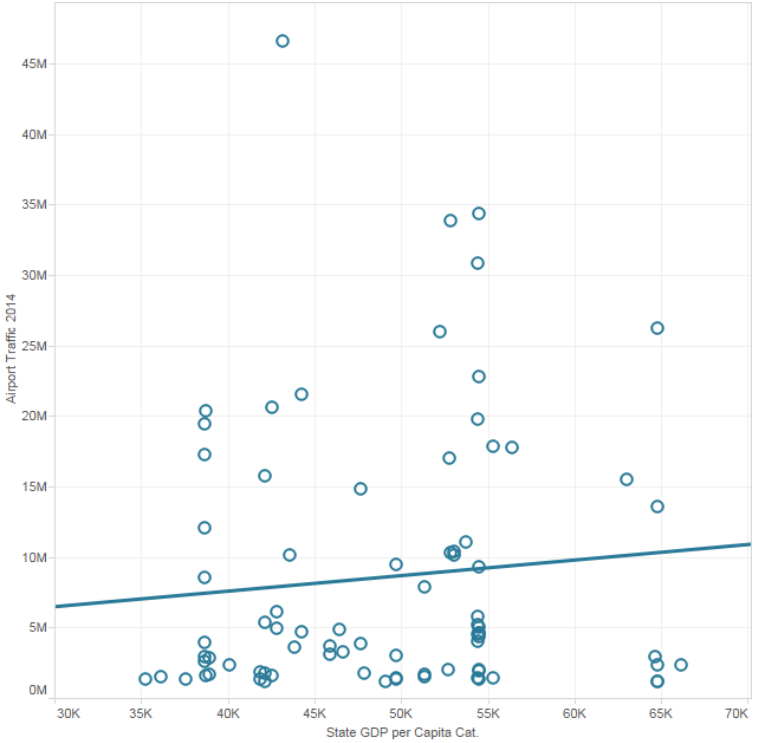
- Cluster group indicate which airports have most flights to one another
- Area-constrained GDP per capita may increase correlation

AIRPORT CLUSTER GROUPS



Airport Name. Color shows details about Clustering Group. Size shows sum of Airport Traffic 2014. The marks are labeled by Airport Name. The data is filtered on Airport Traffic 2014 and Airport Size. The Airport Traffic 2014 filter ranges from 401 to 46604273. The Airport Size filter keeps large and medium. The view is filtered on Clustering Group, which keeps 8 of 8 members.

AIRPORT PASSENGER TRAFFIC AND STATE GDP P.C.



Sum of State GDP per Capita Cat. vs. sum of Airport Traffic 2014. Details are shown for Airport Name. The data is filtered on Airport Traffic 2014 and State GDP per Capita Cat.. The Airport Traffic 2014 filter ranges from 1127914 to 46604273. The State GDP per Capita Cat. filter ranges from 35235 to 66160. The view is filtered on Airport Name, which excludes Boeing Field King County International Airport, Montgomery Regional (Dannelly Field) Airport and William R Fairchild International Airport.

FINAL MODEL

Final Regression model explained 92% of variation in passenger traffic and increased accuracy with only 5 variables.

$$\mathbf{Passengers} = -14558.8262 * \mathbf{Latitude} + 0.585793075 * \mathbf{Population} + 128576 * \mathbf{Degrees} + 0.073626882 * \mathbf{PR'} + 0.147868258 * \mathbf{CC'} + \mu$$

$$PR' = (20743.8 * (PR^4)) - (343625 * (PR^3)) + (1798900 * (PR^2)) - (477376 * PR) + 98401$$

$$\mathbf{CC'} = e^x \quad ; \quad \mathbf{x} = (0.073626882 * CC - 1.29271)$$

MODEL PERFORMANCE METRICS

Metric	Value	% Change
Residual DF	105	-
R^2	0.918918454	8.62%
Adjusted R^2	0.915057428	9.19%
Std. Error Estimate	1874603.67	-13.94%
RSS	3.69E+14	-63.83%

REGRESSION VARIABLES OUTPUT

Input Variables	Coefficient	Std. Error	P-Value	RSS Reduction (in MM's)
Latitude	-14558.8262	6123.15	0.019232	9.12E+14
Population	0.585793075	0.241096	0.016808	6.83E+14
Degree	128576.0803	16997.18	1.55E-11	2.52E+15
PR_Transformed*	0.073626882	0.091514	0.422899	3.50E+13
CC_Transformed**	0.147868258	0.045562	0.001574	3.70E+13

*variable is a four-term polynomial function that returns a predicted value of passengers to be weighted by the regression model
**variable is a logarithmic function that returns a predicted value of passengers to be weighted by the regression model

CONCLUSION

AIRPORT TRAFFIC = LATITUDE + POPULATION + DEGREE + CLOSENESS + PAGERANK

THANK YOU!

Charlotte Garcia | Isabelle Grégoire | Daichi Ishikawa | Jedaiah Tan