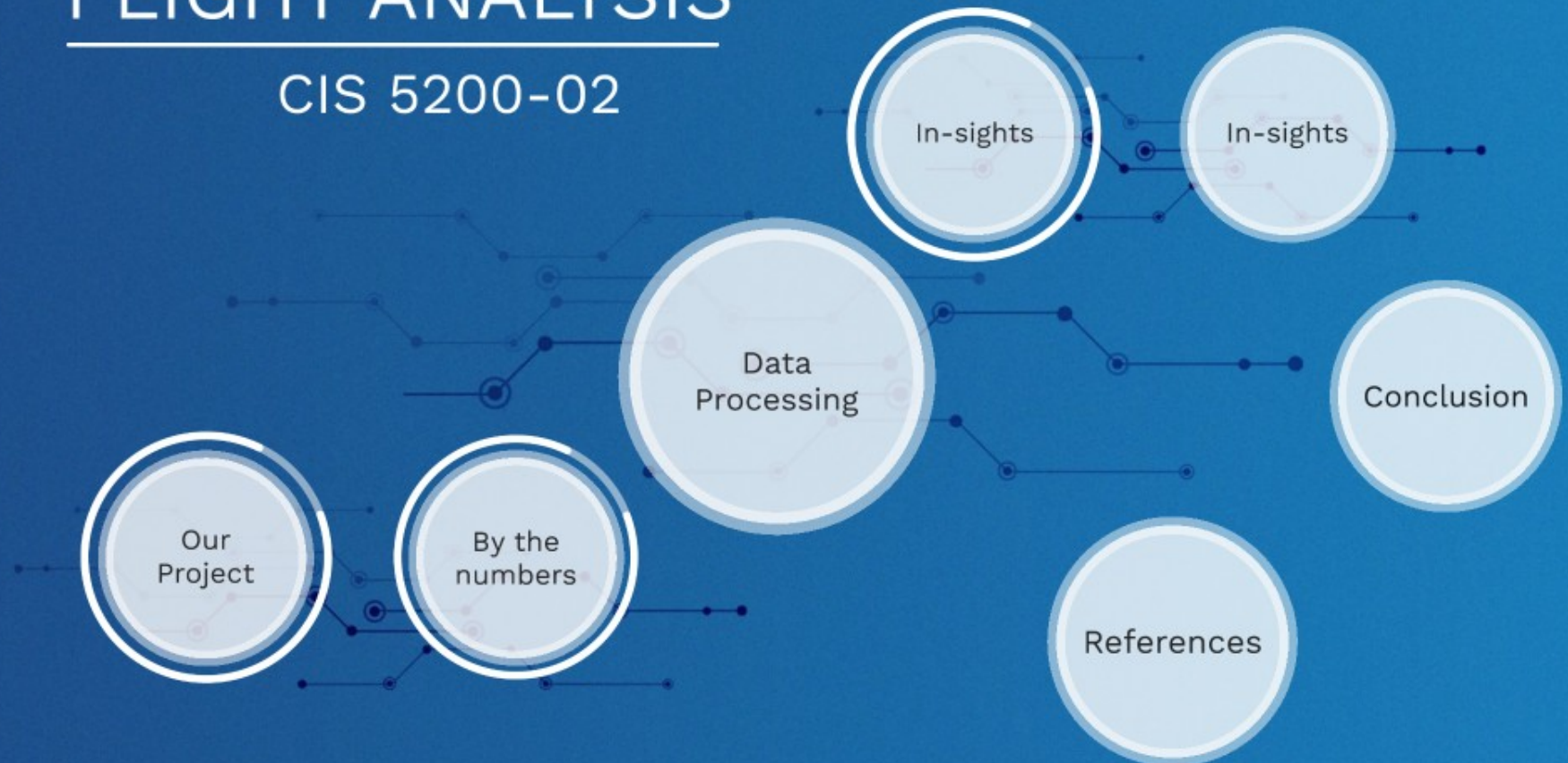


FLIGHT ANALYSIS

CIS 5200-02



GROUP 1

Flight Data Analysis



THE
TEAM

THE TEAM

ANUPAM SAHAY
(CIN : 305903512)

KRITHY NANAIAH
ATRANGADA (CIN :
305903525)

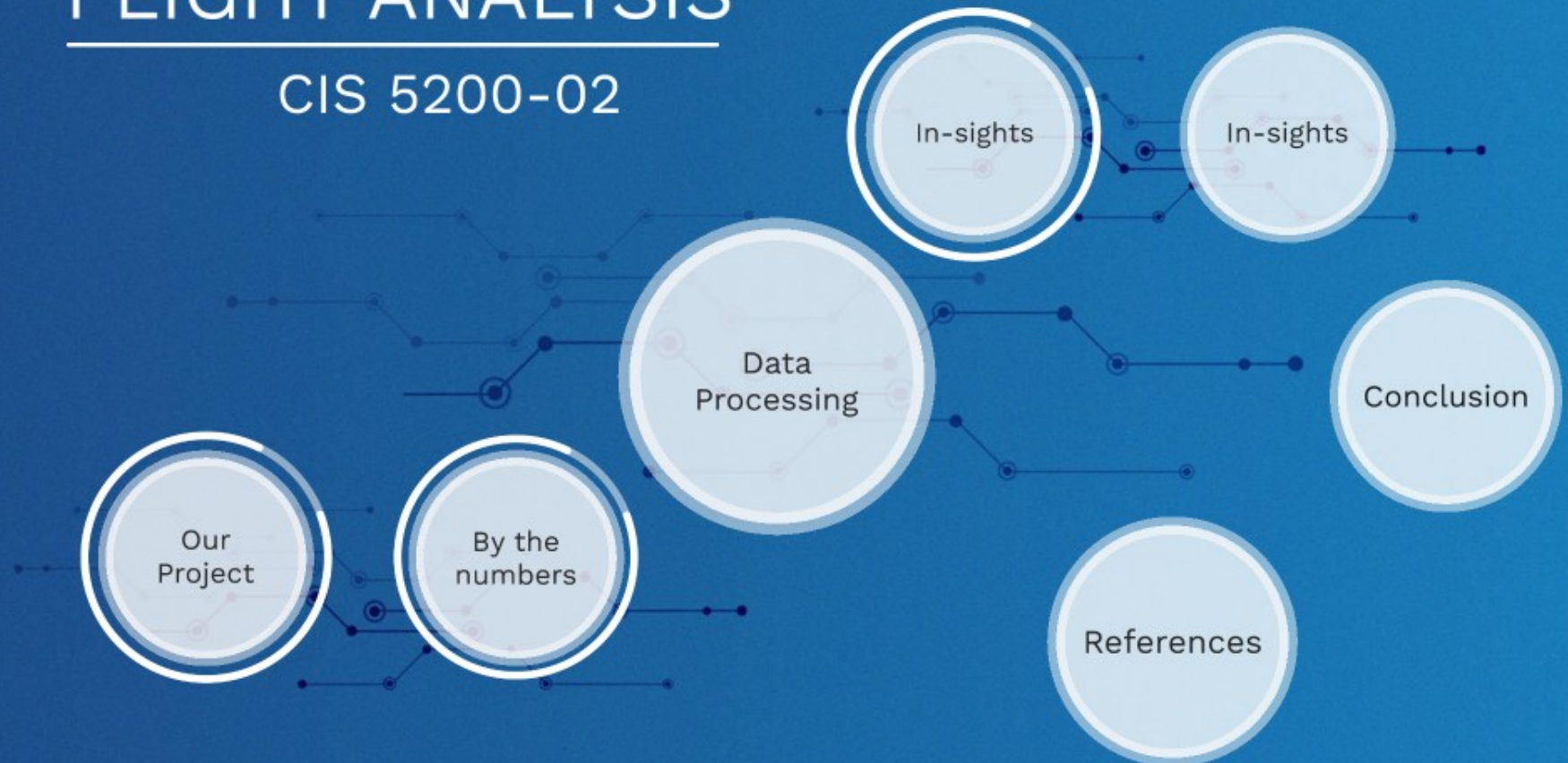
NITESH KAMBOJ
(CIN : 306601664)

SOHONG
CHAKRABORTY
(CIN : 306628834)

ZEESHAN KHAN
(CIN : 306587325)

FLIGHT ANALYSIS

CIS 5200-02



FLIGHT DATA ANALYSIS USING PIG

DATA SIZE : 13 GB (Approx)

No.of years covered :
1987-2008

Data Set Source URL:

[http://stat-computing.org/
dataexpo/2009/the-
data.html](http://stat-computing.org/dataexpo/2009/the-data.html)

Objectives

Specification

Objectives

Learn how to use Pig and Tableau to:

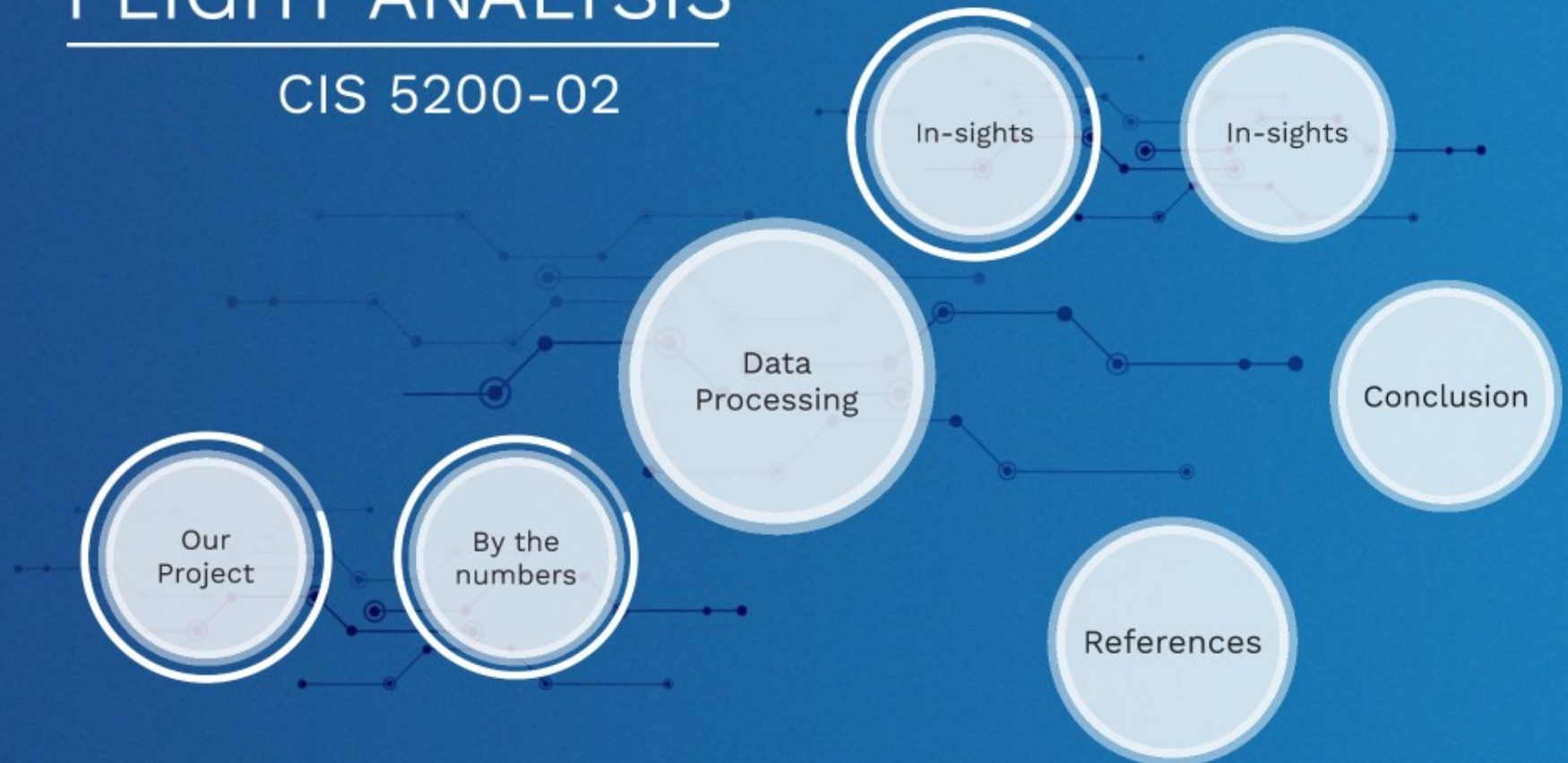
- Filter out the popularity of airline carrier which was most popular in 1987-2008
- Visualize which top monthly outbound flights from all United States Airports.
- Busiest route from LAX (Los Angeles International Airport)
- Longest Flight in 1987 - 2008
- Average Delay ≥ 15 min for arrival & departure

Specification

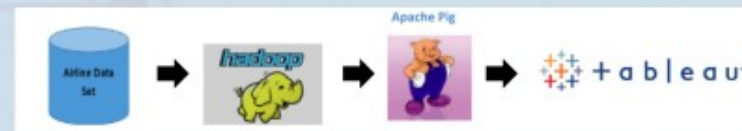
- Cluster Version : IOP 4.2
- Number of Management Nodes : 01
- Number of Data Nodes : 03
- Memory Size : RAM 48GB - Management Nodes, RAM 24GB - Data Nodes
- Disk Size 1TB SATA

FLIGHT ANALYSIS

CIS 5200-02



DATA PROCESSING



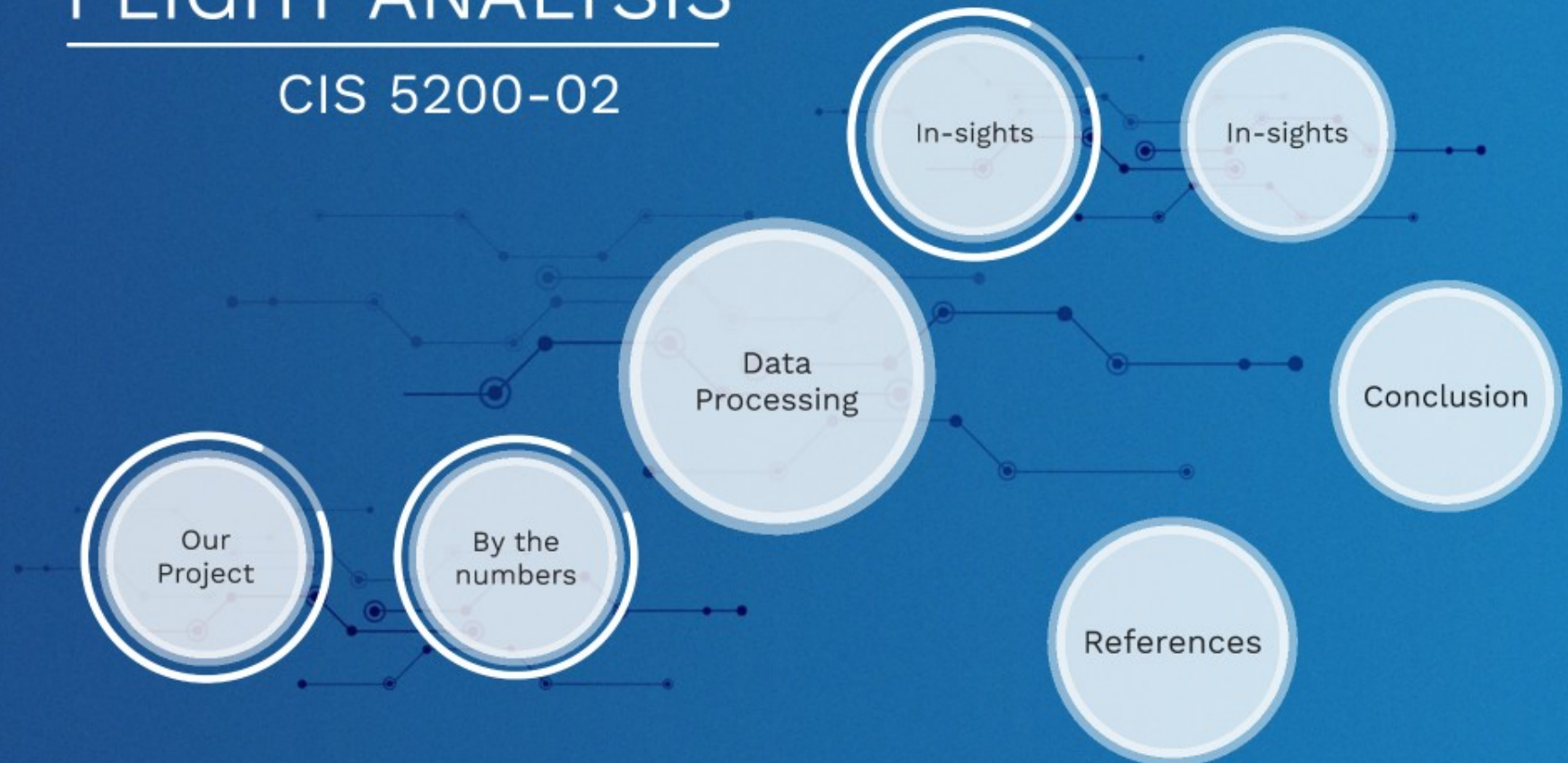


Apache Pig



FLIGHT ANALYSIS

CIS 5200-02



Flight Insights

Insight
1

Insight
2

Insight
3

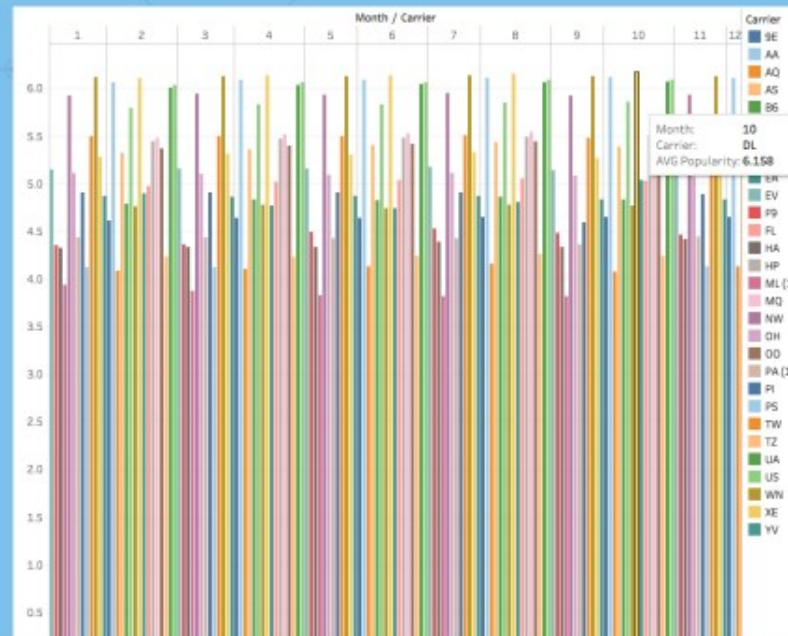
Insight 1

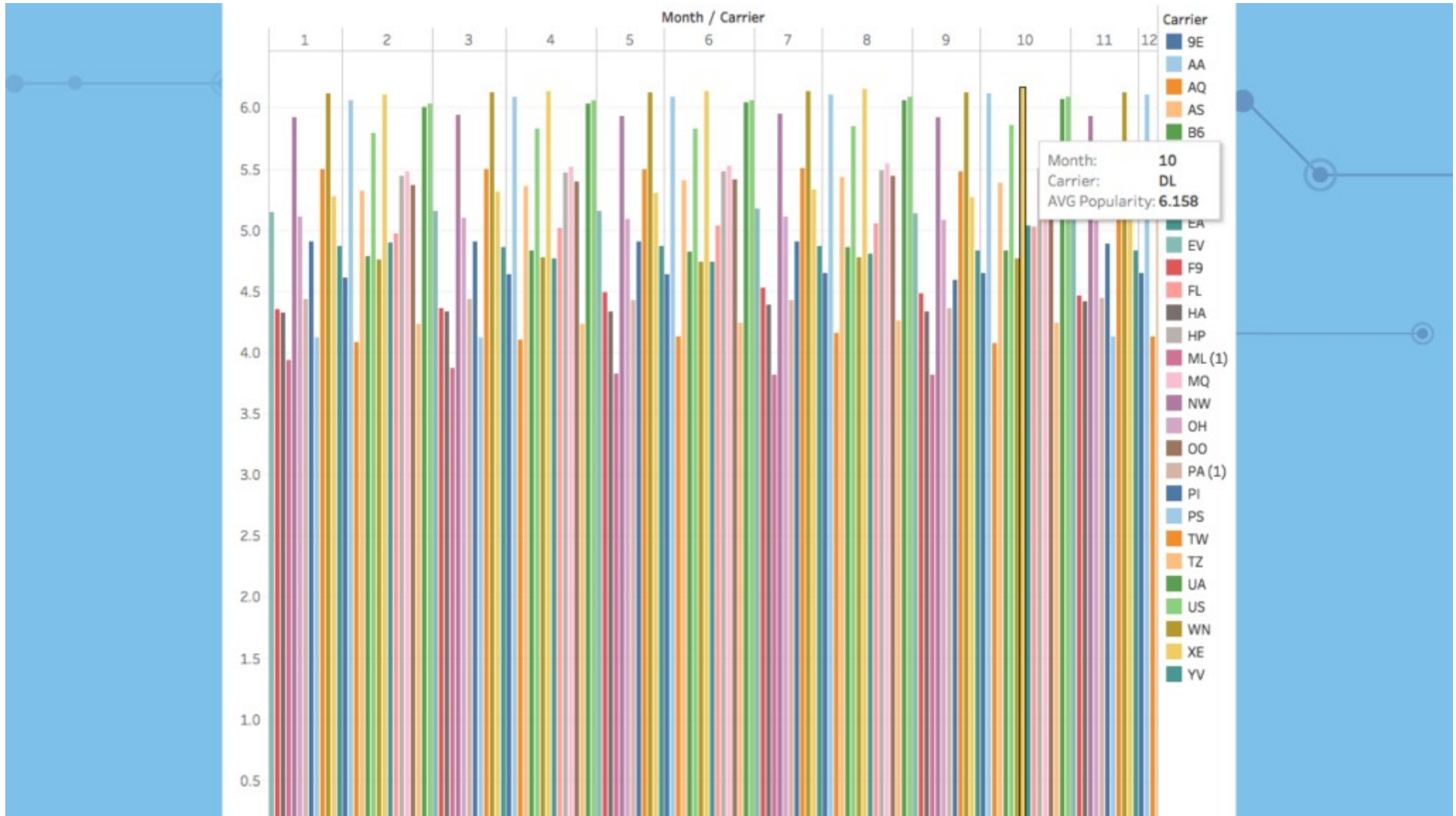
Airline Carrier which was the most popular in a given Year.

> Flight - Delta Airlines

> Month - October (1987-2008)

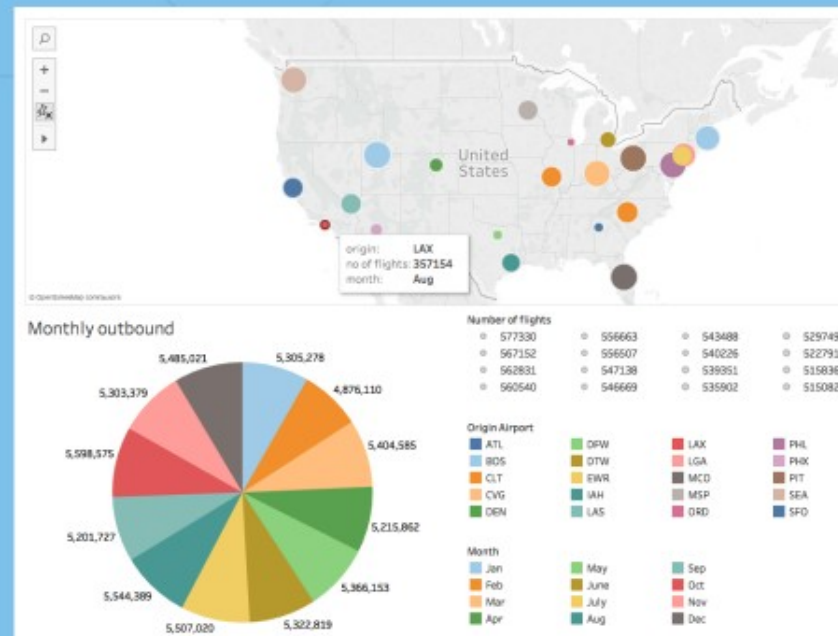
> Avg Popularity - 6.158

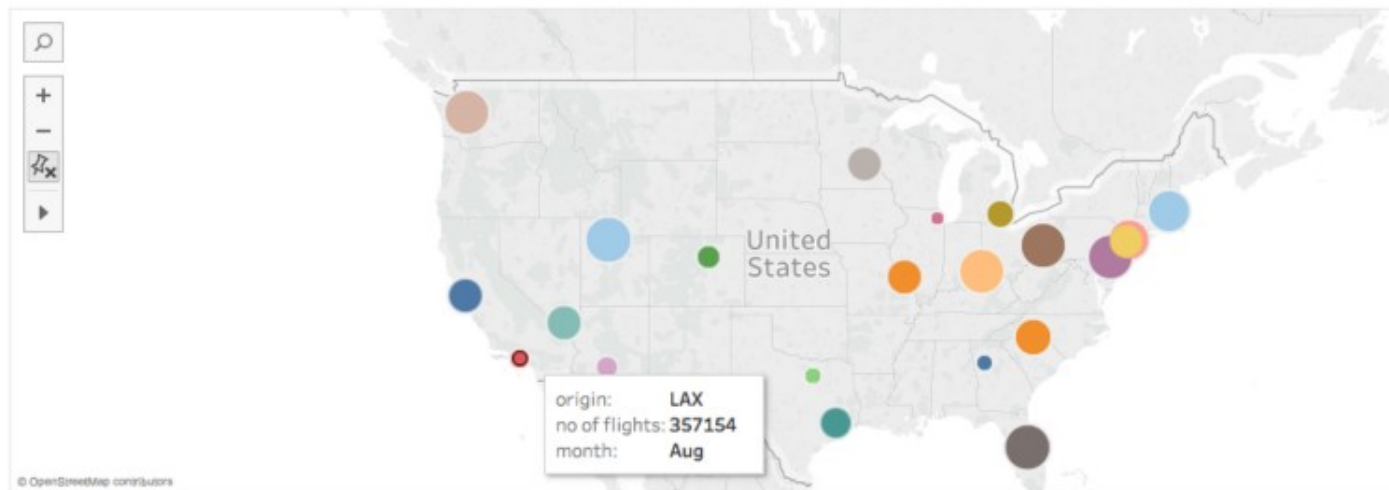




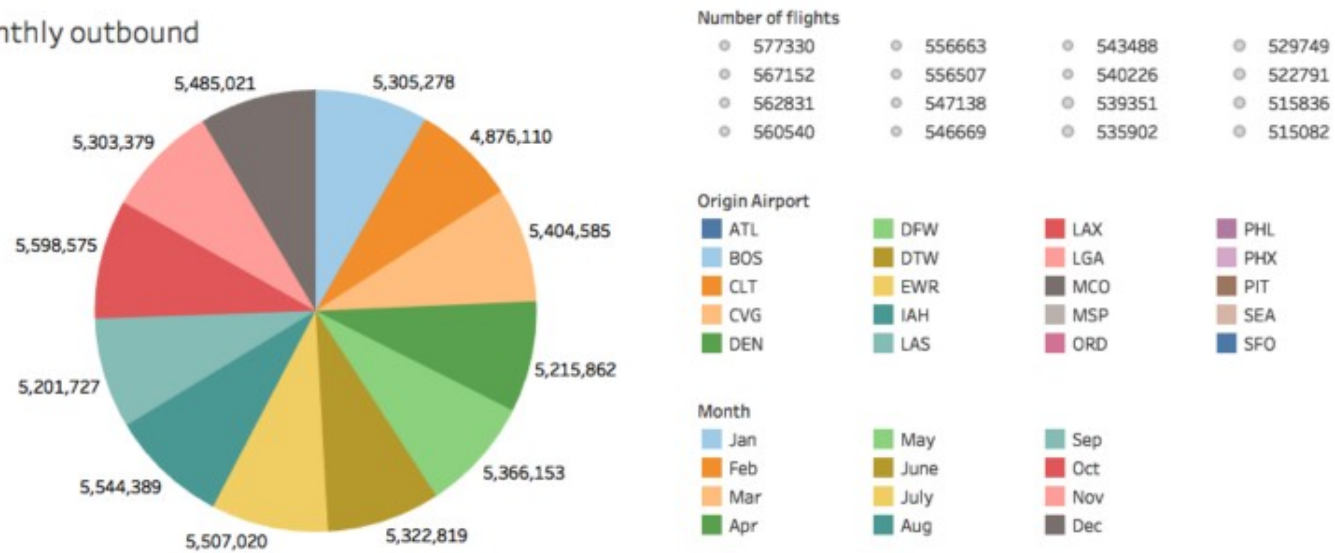
Insight 2

Outbound flights from Top 20 airports on departure basis
> Timespan - 1987 to 2008



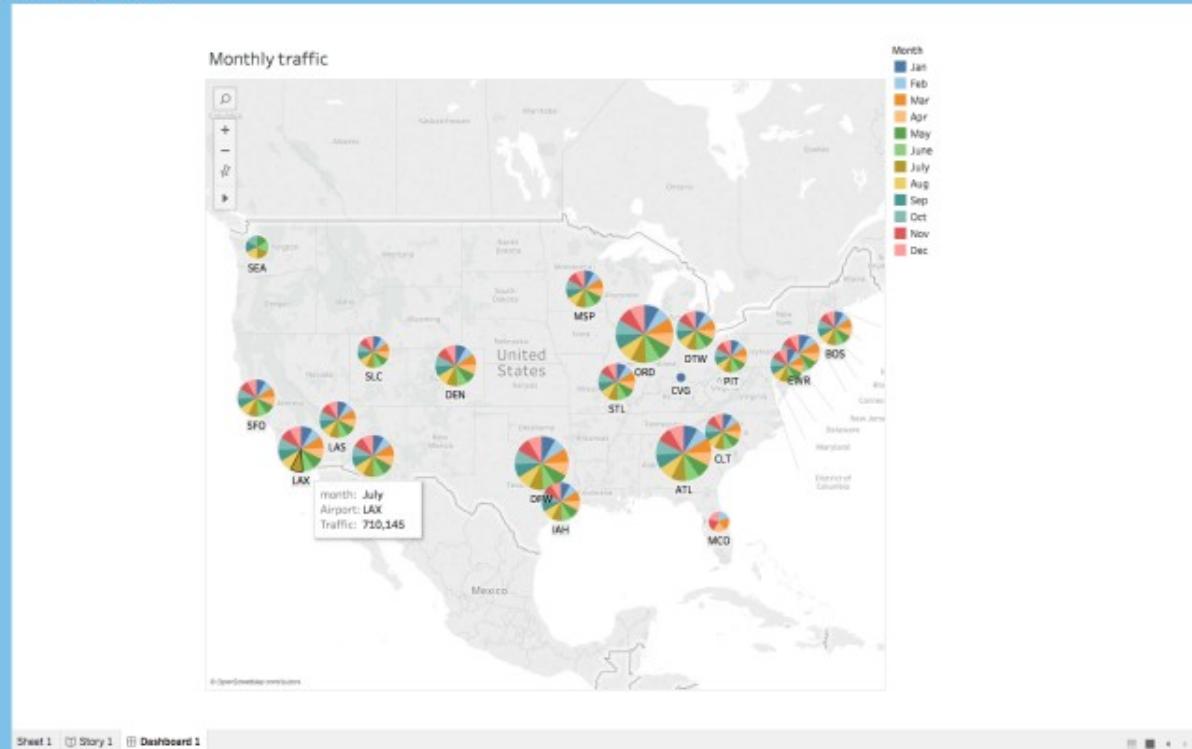


Monthly outbound

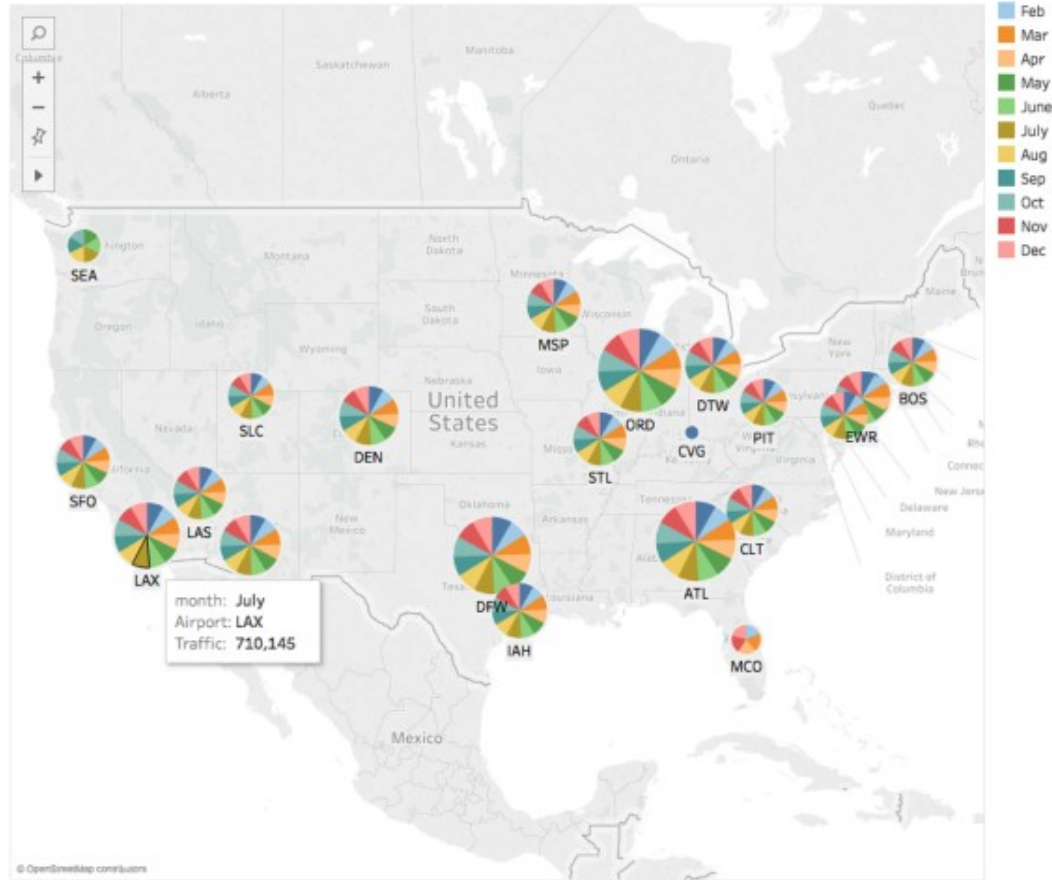


Insight 3

Total flights from Top 20 airports on monthly traffic basis
> Timespan - 1987 to 2008

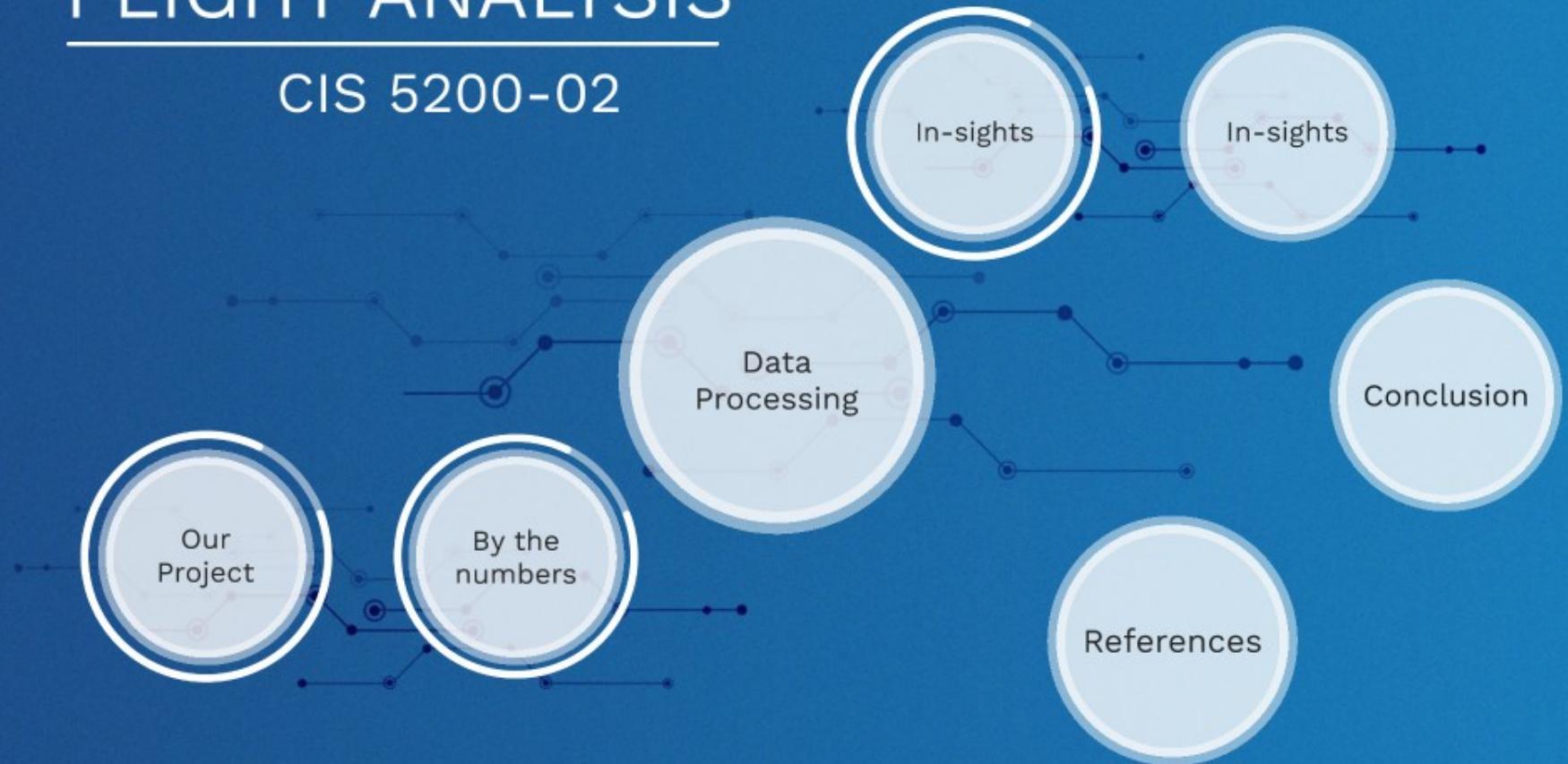


Monthly traffic



FLIGHT ANALYSIS

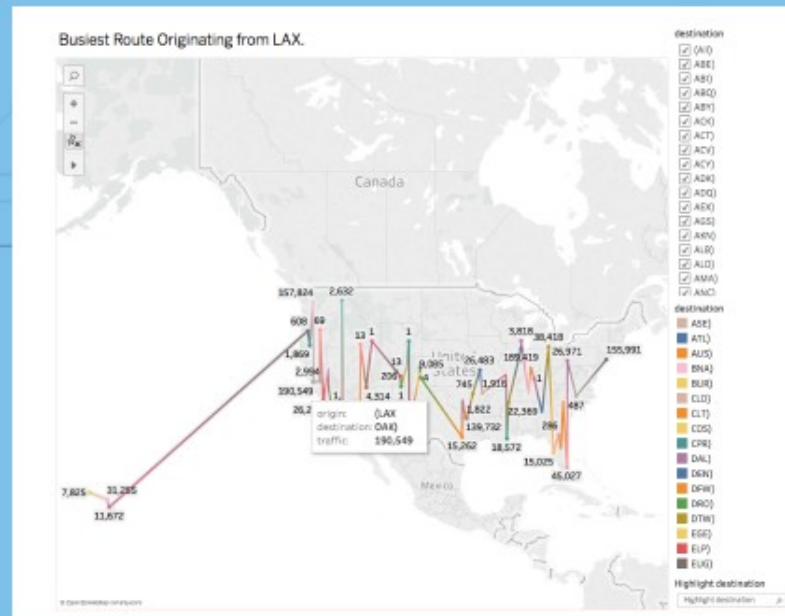
CIS 5200-02



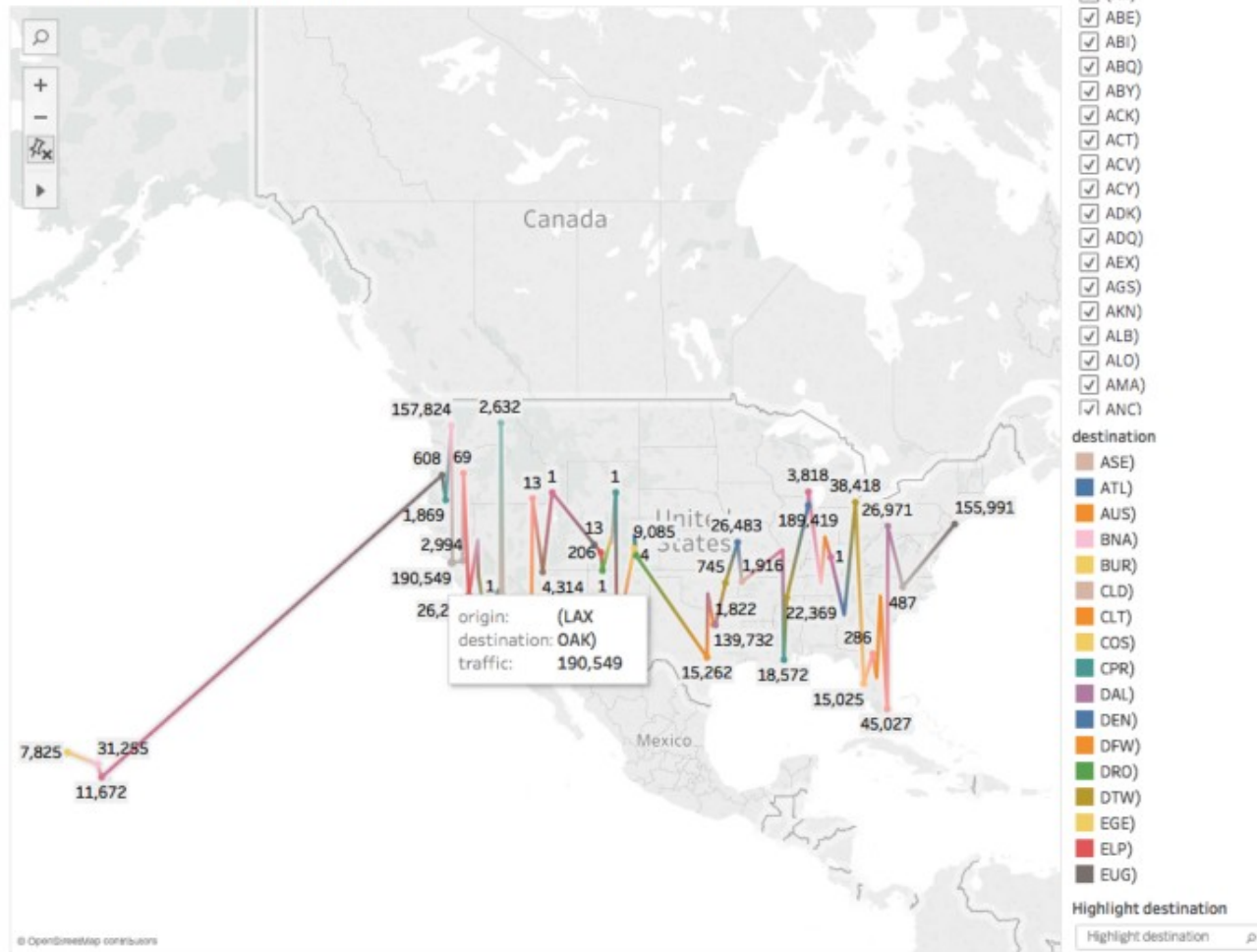


Insight 4

Total flights originating from LAX to other airports
> Timespan - 1987 to 2008

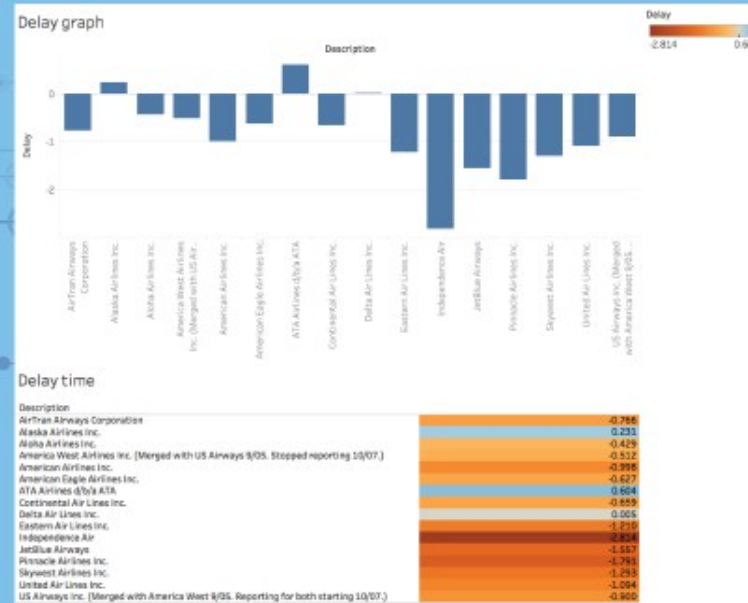


Busiest Route Originating from LAX.

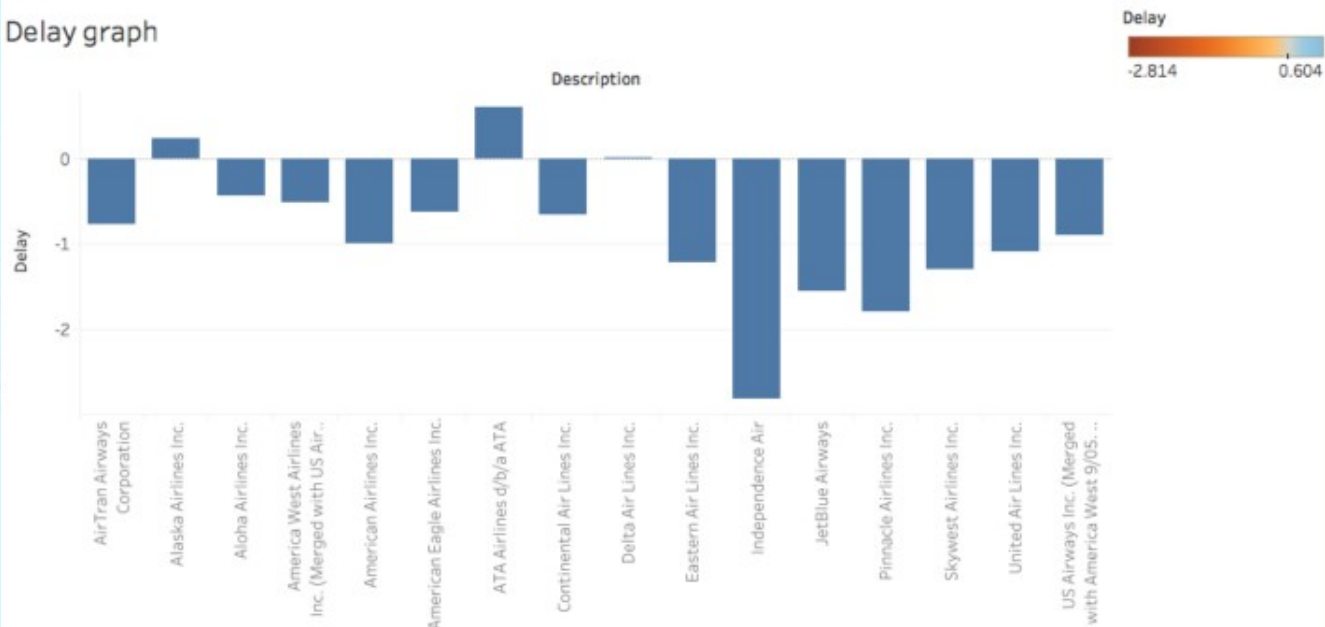


Insight 5

Carrier specific average delay (≥ 15 minutes)



Delay graph

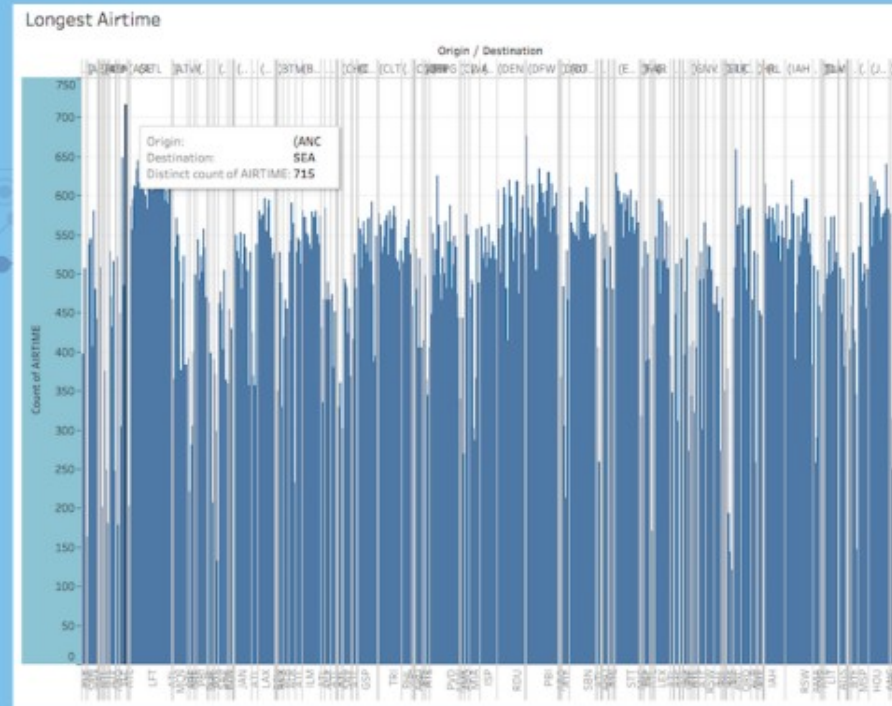


Delay time

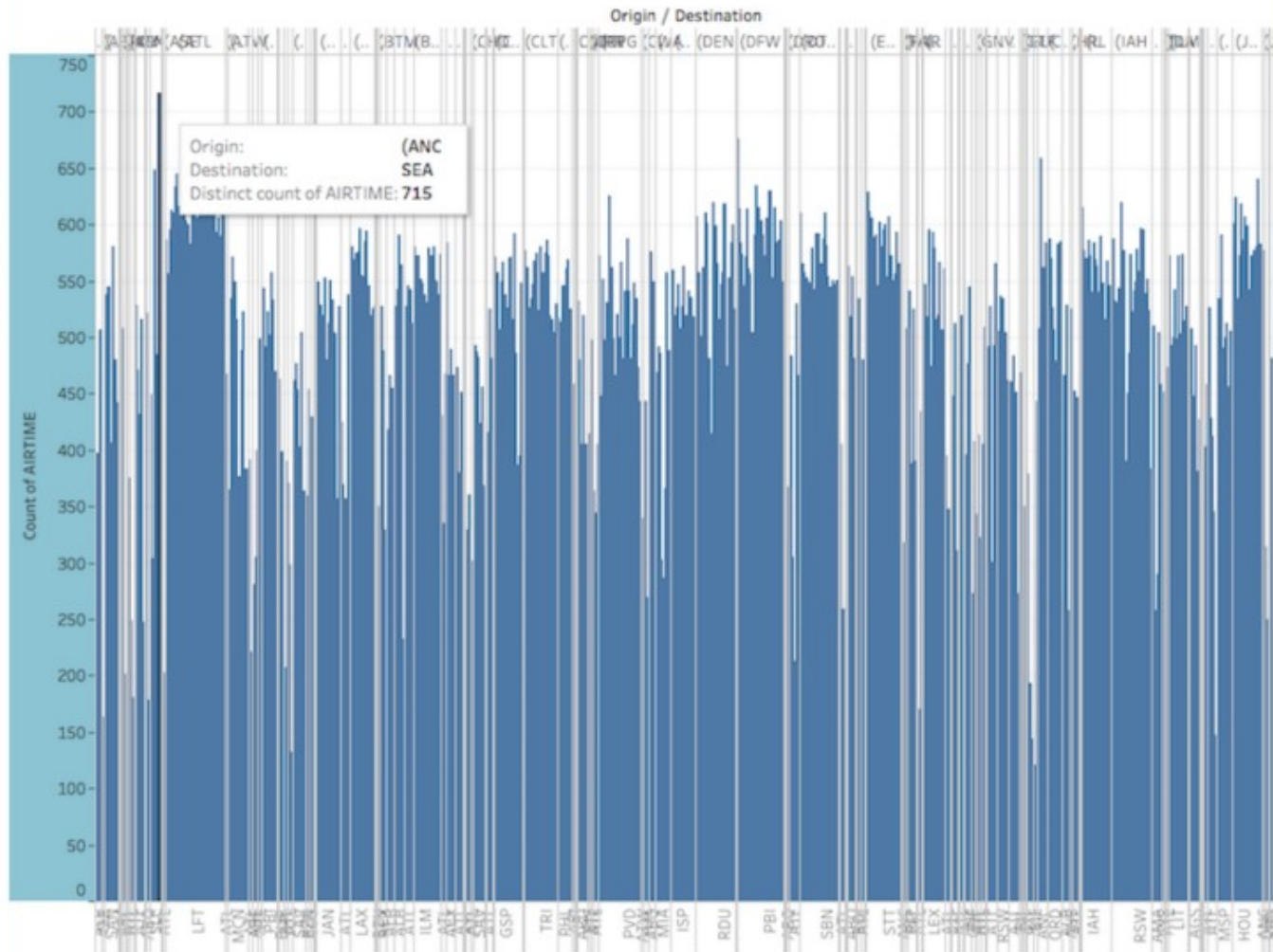
Description	
AirTran Airways Corporation	-0.766
Alaska Airlines Inc.	0.231
Aloha Airlines Inc.	-0.429
America West Airlines Inc. (Merged with US Airways 9/05. Stopped reporting 10/07.)	-0.512
American Airlines Inc.	-0.998
American Eagle Airlines Inc.	-0.627
ATA Airlines d/b/a ATA	0.604
Continental Air Lines Inc.	-0.659
Delta Air Lines Inc.	0.005
Eastern Air Lines Inc.	-1.210
Independence Air	-2.814
JetBlue Airways	-1.557
Pinnacle Airlines Inc.	-1.791
Skywest Airlines Inc.	-1.293
United Air Lines Inc.	-1.094
US Airways Inc. (Merged with America West 9/05. Reporting for both starting 10/07.)	-0.900

Insights 6

Longest flight between two airports

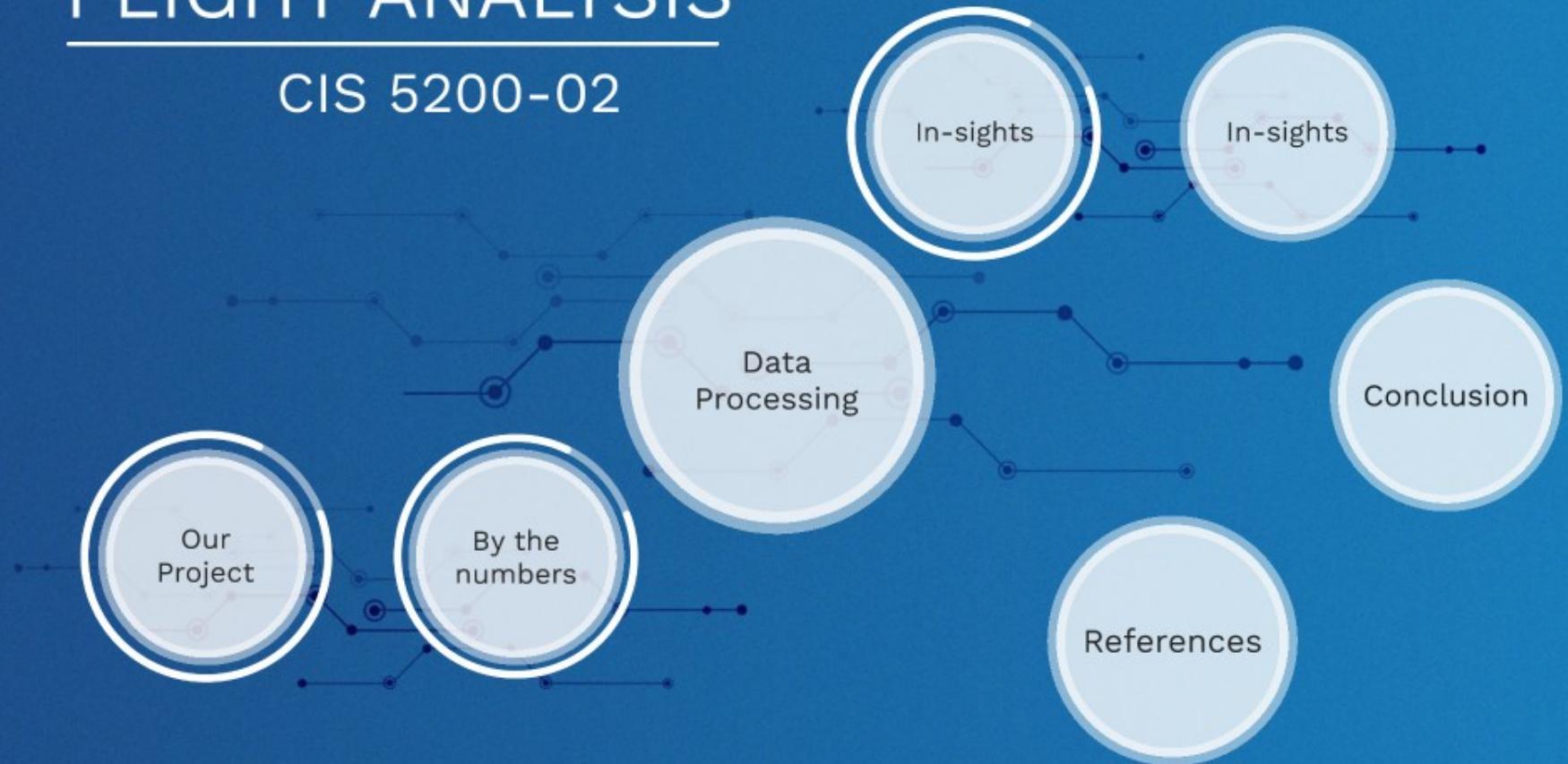


Longest Airtime



FLIGHT ANALYSIS

CIS 5200-02



Conclusion

From the above experimental results, we can see that interesting sets of trends and patterns exists in large data sets which helps us to get a better understanding of the data.

Airline Carrier which was the most popular in a given Year.


Outbound flights from Top 20 airports on departu...

Total flights from Top 20 airports on monthly traff...

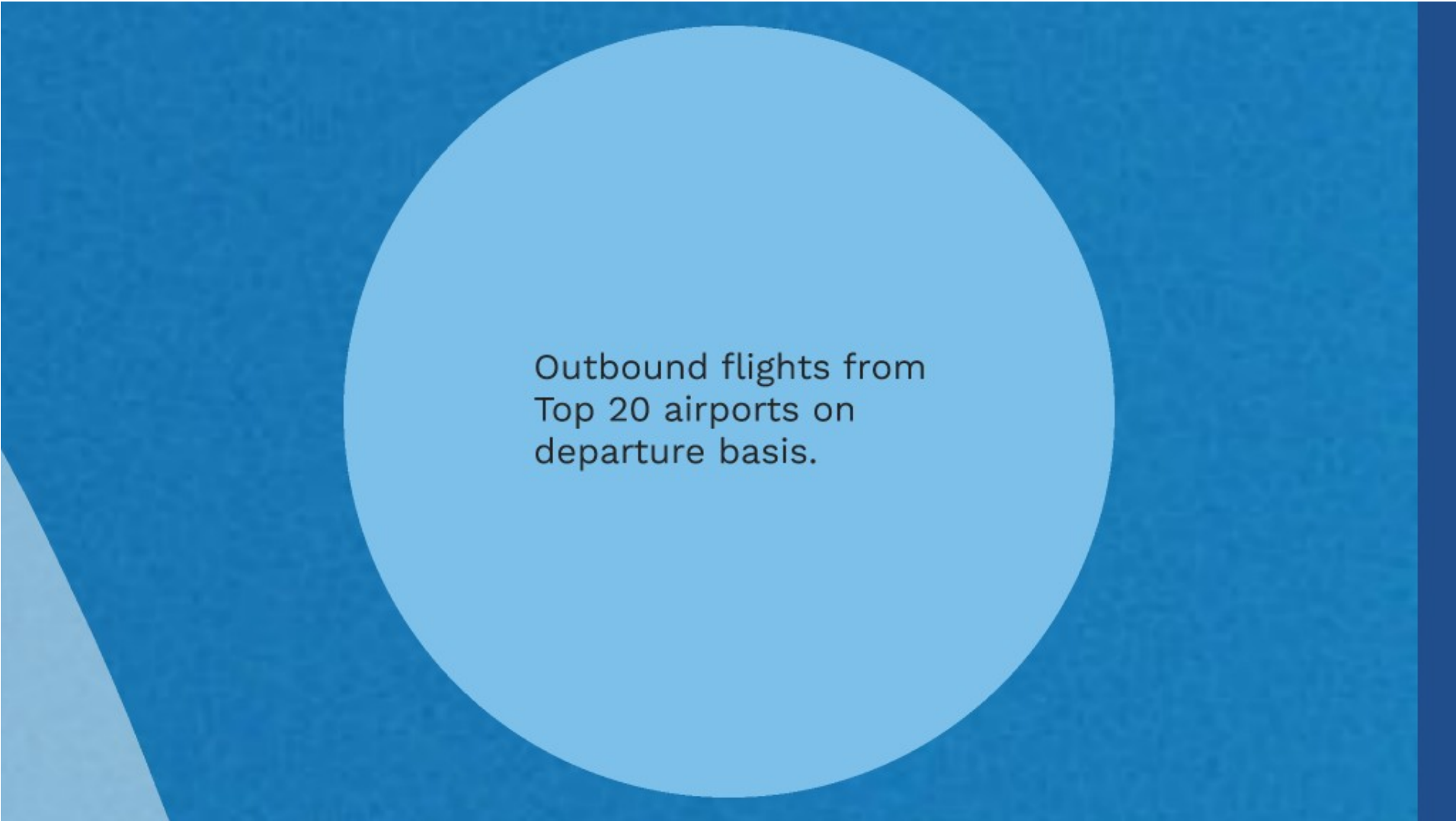
Total flights originating from LAX to other airports

Carrier specific average delay

Longest flight between two airports



Airline Carrier which was
the most popular in a
given Year.



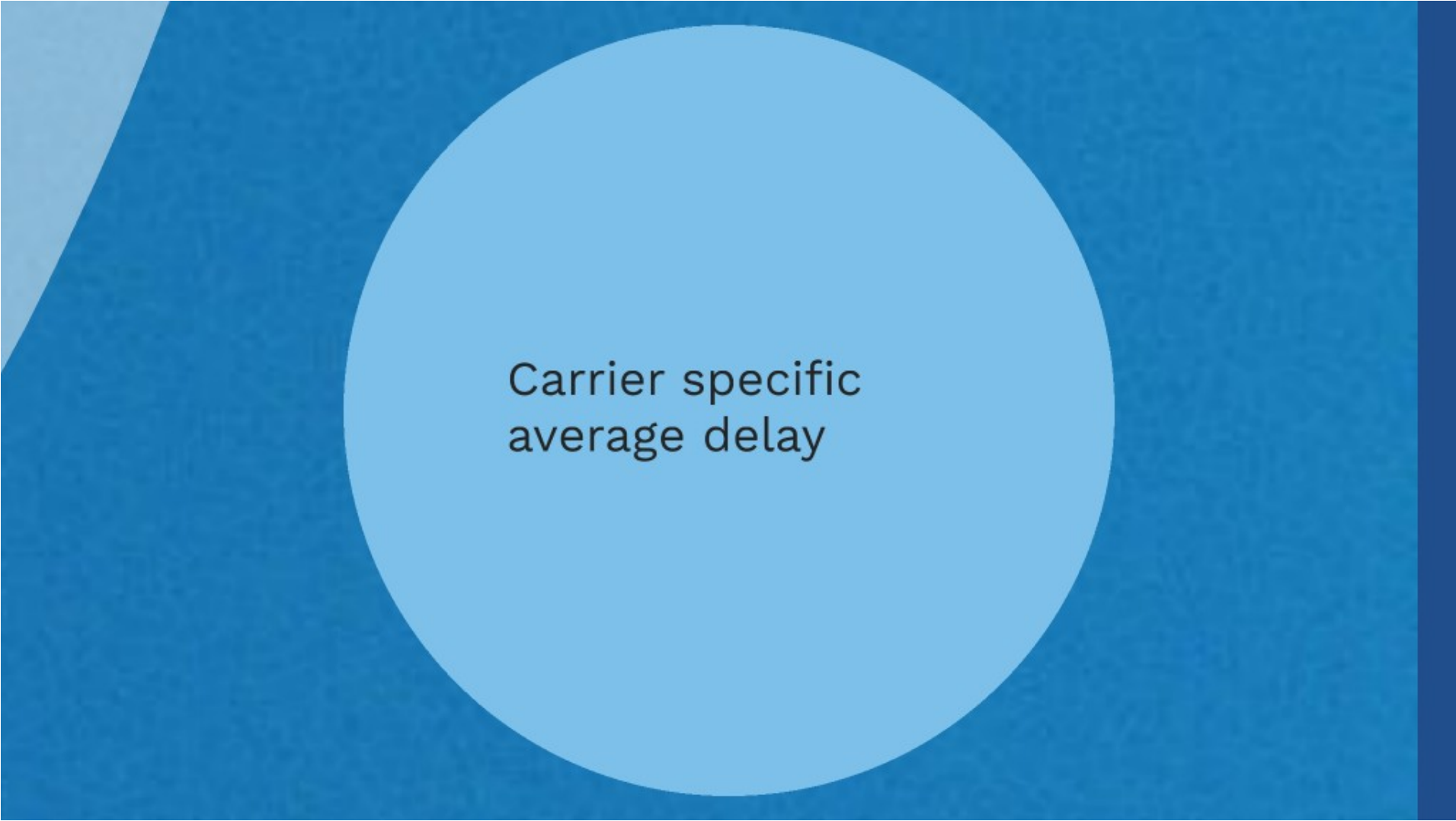
Outbound flights from
Top 20 airports on
departure basis.



Total flights from
Top 20 airports on
monthly traffic
basis



Total flights
originating from
LAX to other
airports



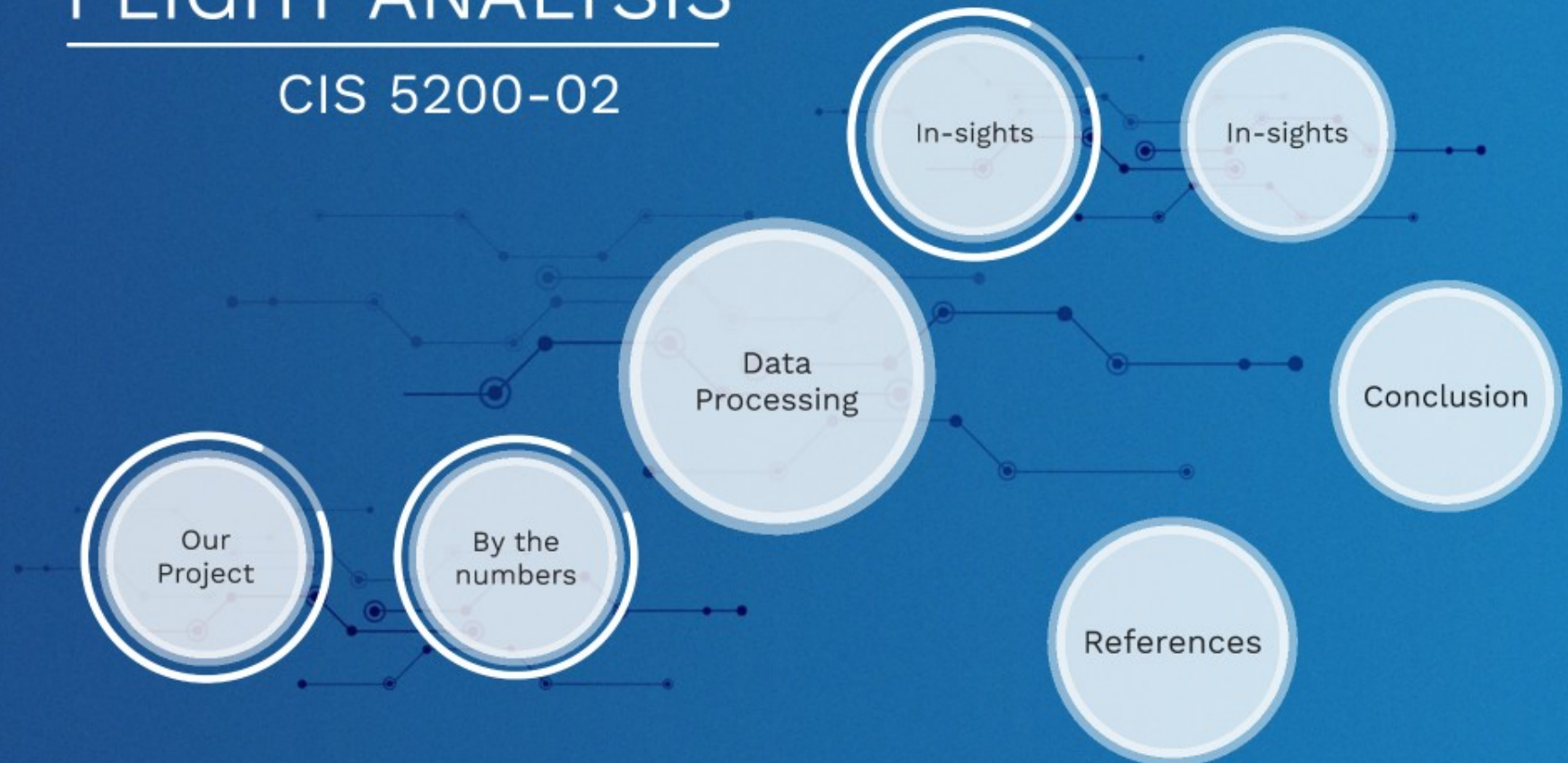
Carrier specific
average delay



Longest flight
between two
airports

FLIGHT ANALYSIS

CIS 5200-02



References

- Airline Data
<http://stat-computing.org/dataexpo/2009/the-data.html>
- Images Courtesy of Google
- Grand Data Source
<http://www.calstatela.edu/centers/hipic/related-site>
- IBM Bluemix
<https://console.bluemix.net/>
- GitHub Link
<https://github.com/Zeebag/5200Project.git>
- Nillohit Bhattacharya and Jongwook Woo, "Airline Data Set Analysis using Big Data in Cloud Computing" in The 2017 Korea Society of Management Information Systems Spring Annual Conference (KMIS 2017), Chonnam University, Korea, June 6 - 9 2017

FLIGHT ANALYSIS

CIS 5200-02

