FlightGlobal

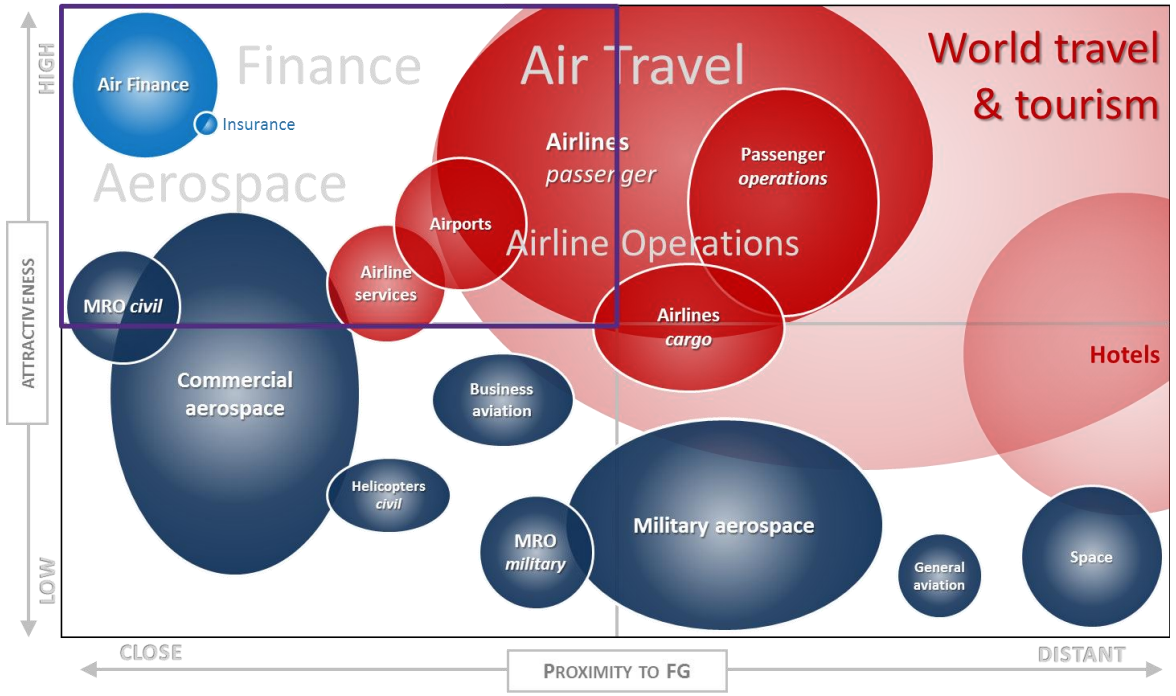
Pioneering Aviation Insight

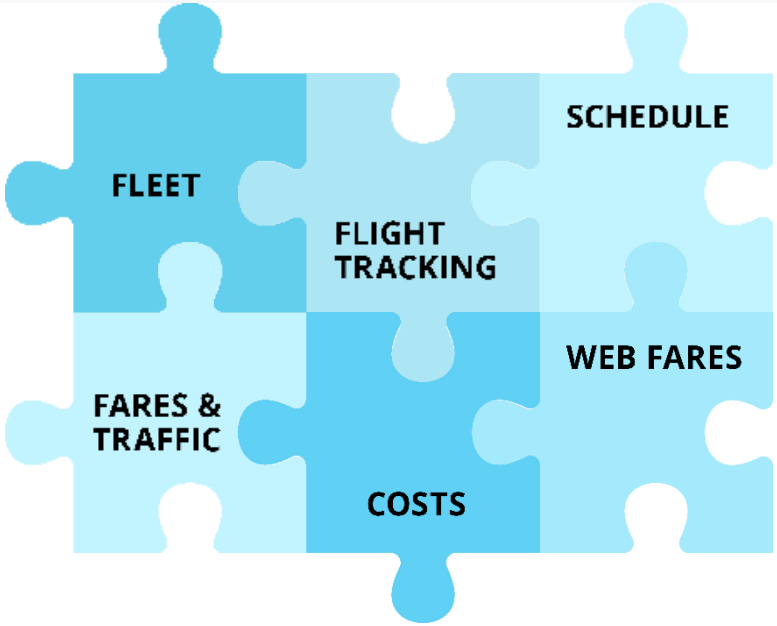
# Enabling Aviation Analytics through HPCC Systems

**HPCC Systems Summit 2016**

Michael Targett, Senior Director, Data

# Aviation information marketplace




**Travel & tourism is a $3 trillion industry**

- FlightGlobal is the industry leading information provider in aerospace and air finance sectors
- Building new capabilities in air travel and airline operations sectors

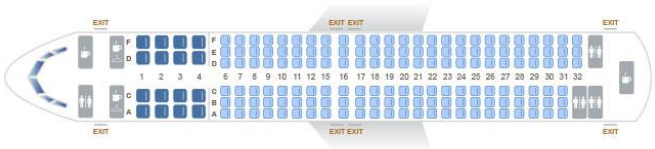
## 6 essential data sets:

Combinations of six macro data sets power market insight and analytics

# Analysing aircraft utilisation by route

## The problem

# Typical analytics use cases




## Capacity

- Who: Network planners, airport business development
- What: What market share / revenue can I expect from flying a new route?
- How: Analyze competitor airline flights already on route by cabin class

## Equipment

- Who: Fleet planners, network planners, schedulers, airport operations
- What: What's the most efficient or competitive aircraft to use on a given route?
- How: Benchmark against other airlines operating those routes

## Product offering

- Who: Revenue managers, OTA
- What: What can passengers expect onboard flights on this route?
- How: Analyse typical aircraft / interior of operator fleet

# Analyzing aircraft utilization by route

**Historical analysis:**

- Based on recorded schedule
- Doesn't reflect changes
- Contains known gaps and inaccuracies

## Data limitations

Analysis based on planned and/or own estimates.

**Forward-looking analysis:**

- Based on submitted schedule
- Extended out on projection
- Assumptions and estimates

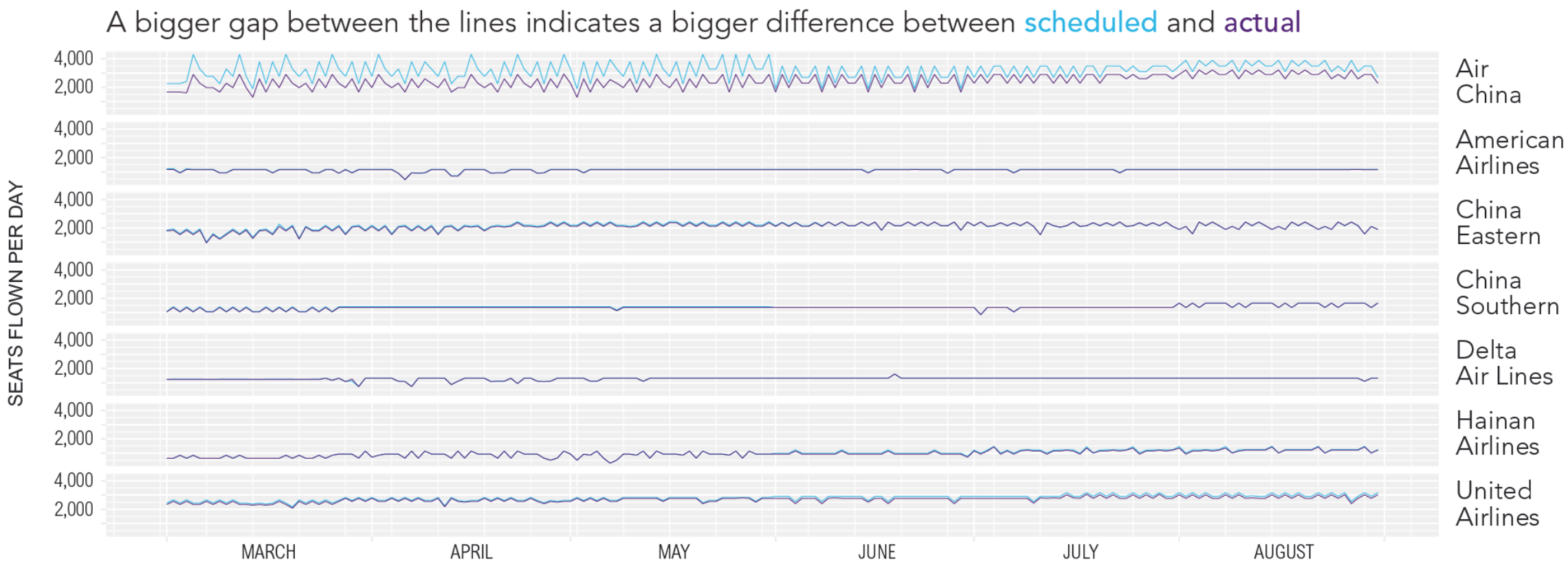
## Hypothesis

*More accurate information will prove a differentiator in the marketplace against other data offerings based on schedule plus estimates*

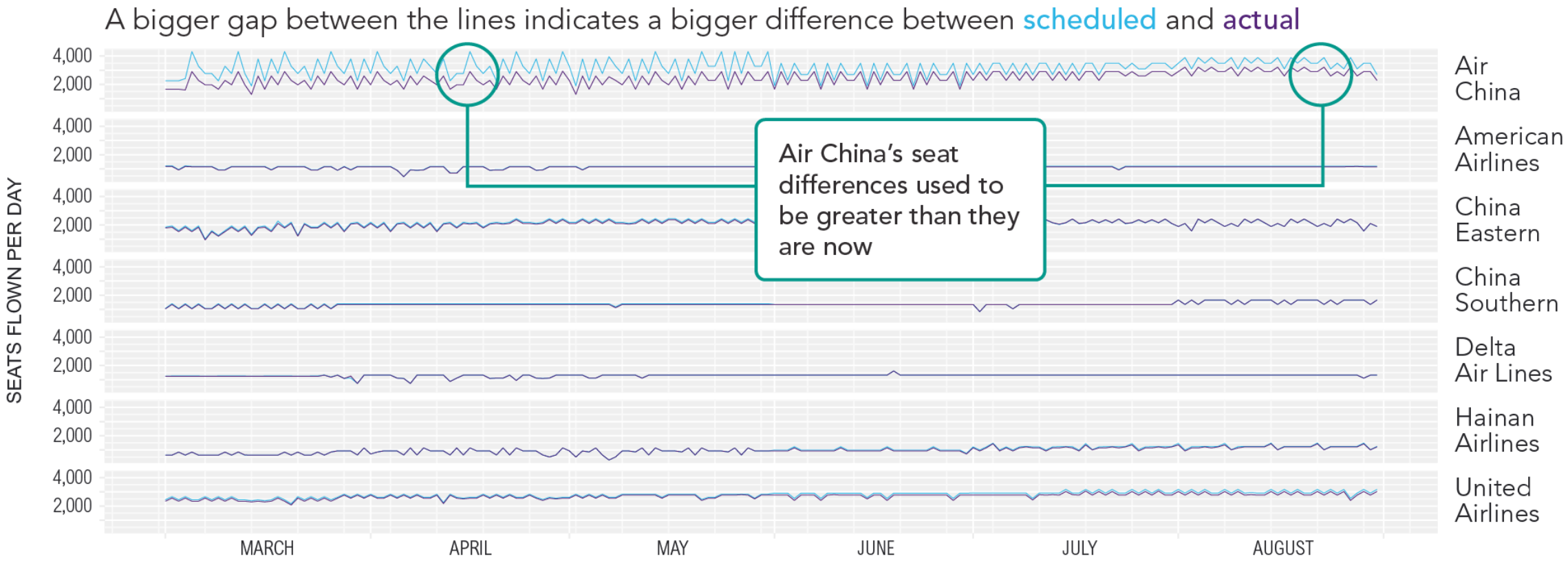
# Forward planning on available capacity

## Case study: USA to China capacity analysis

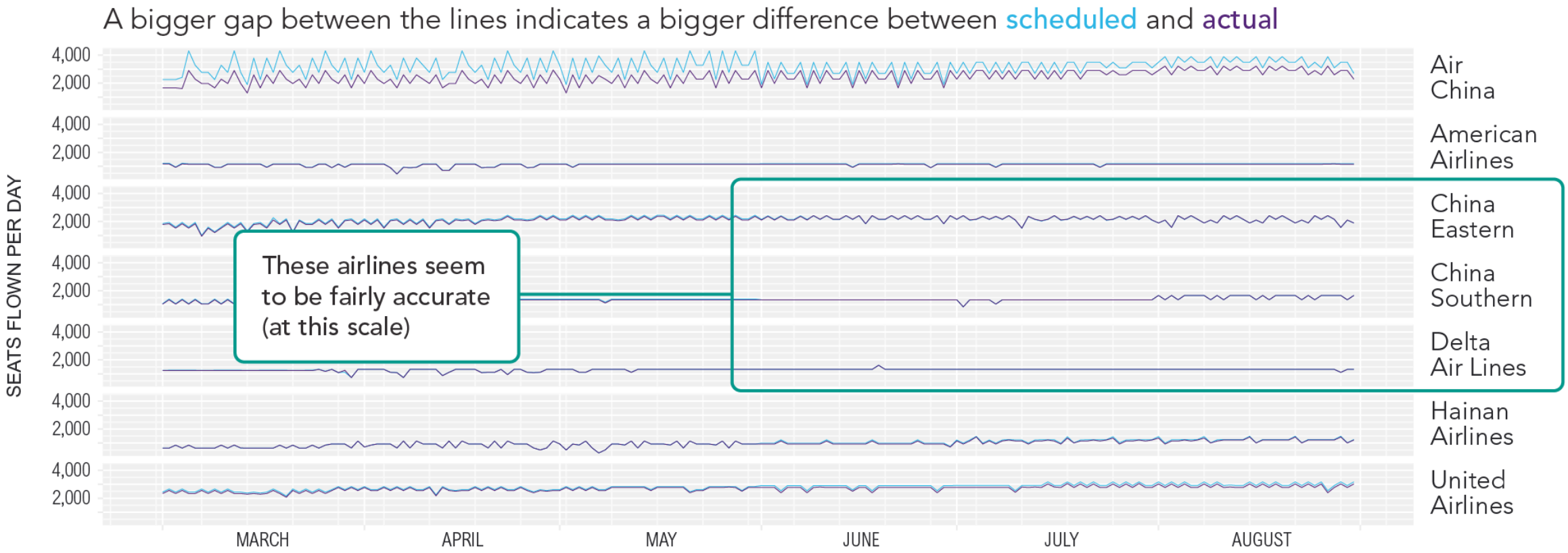
# How do seat counts differ from the schedule?



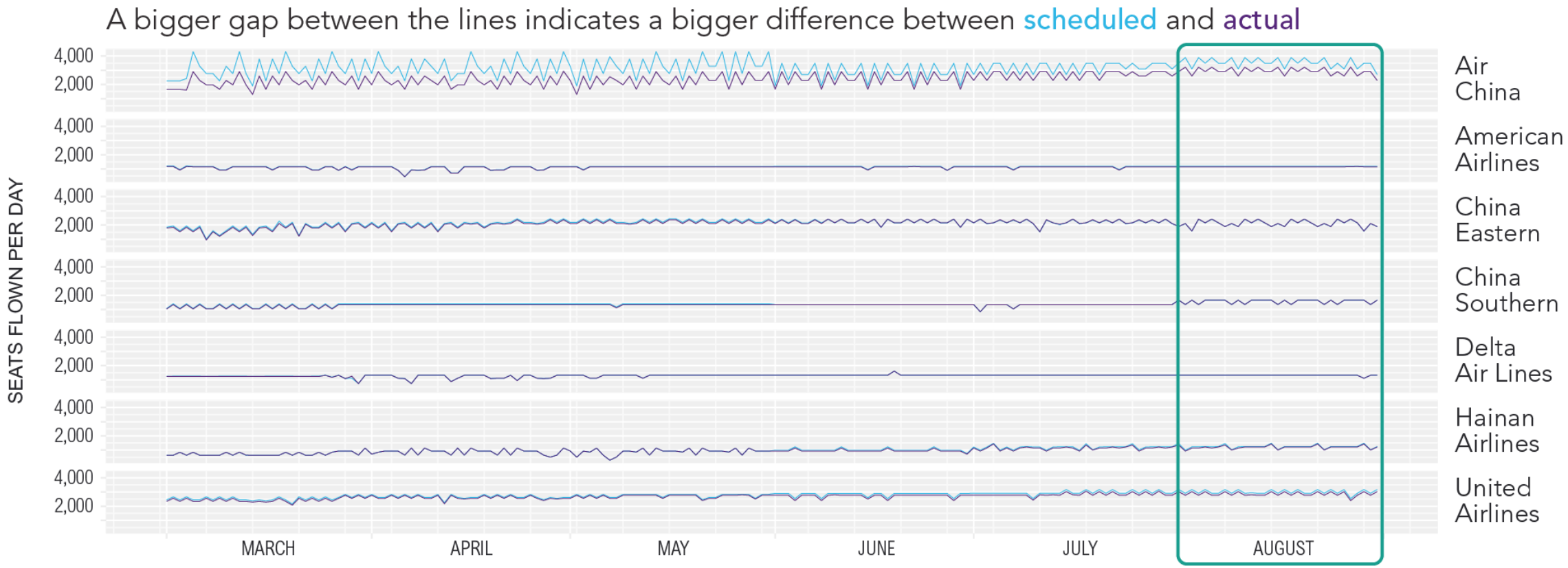
# What can we learn from this?



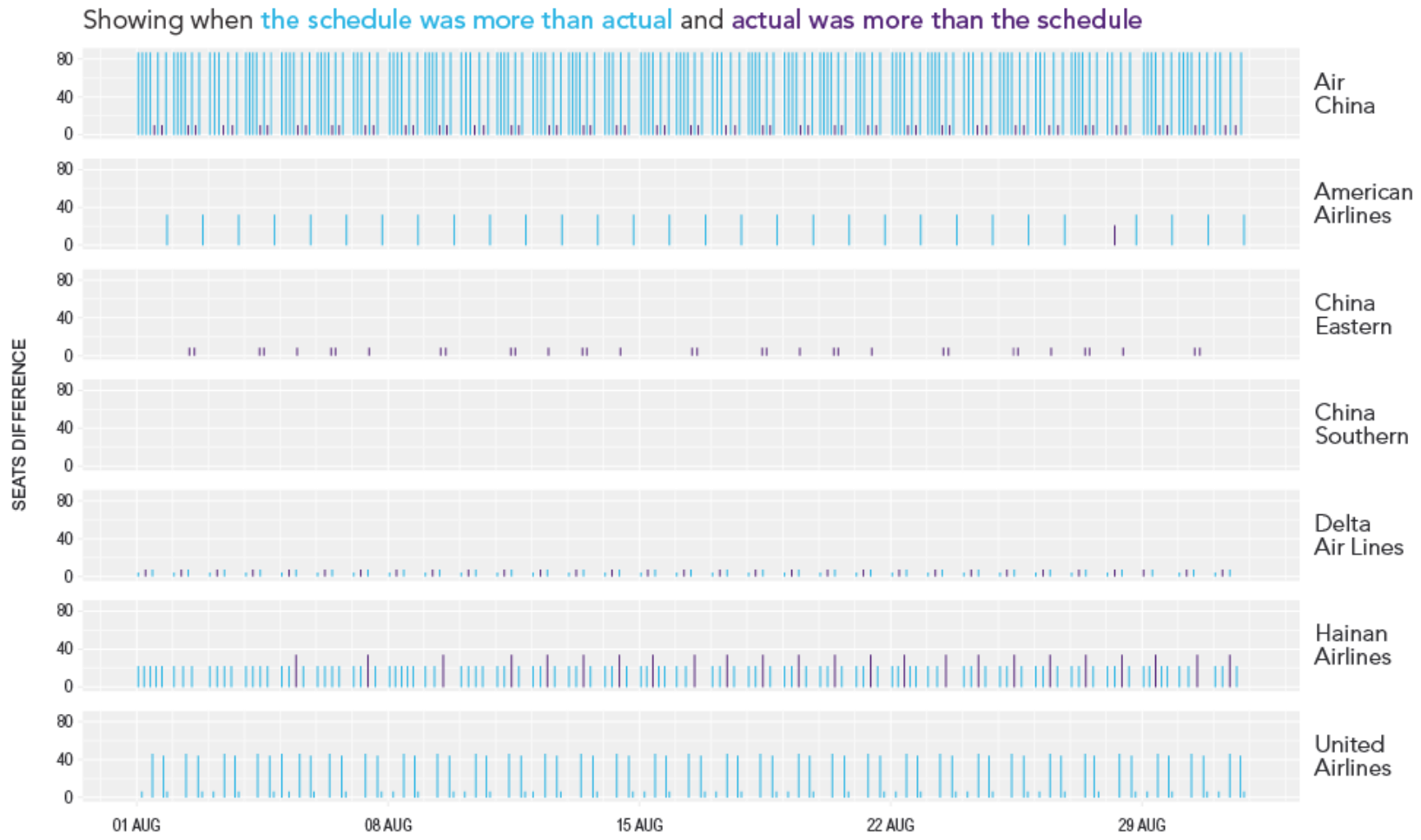
# What can we learn from this?



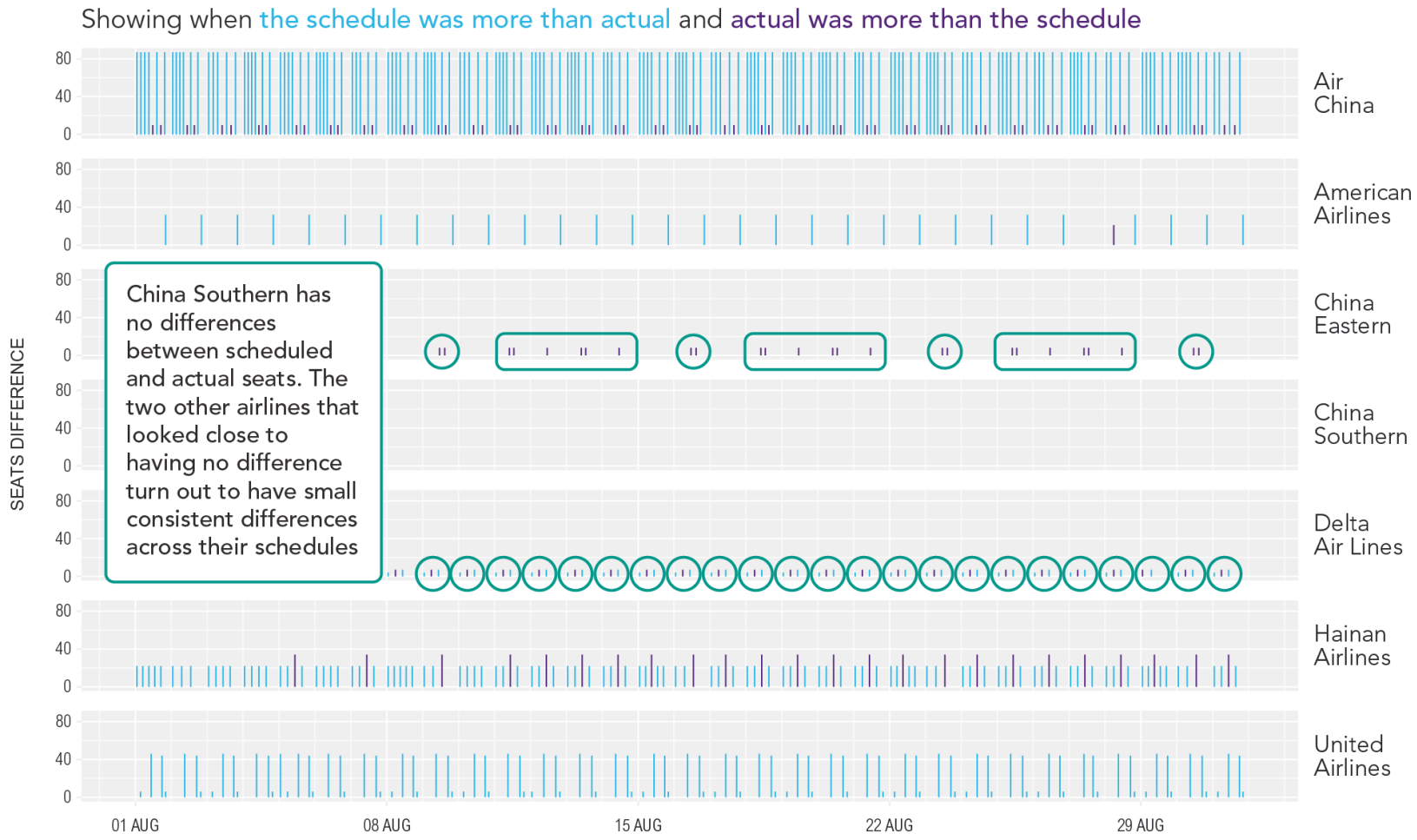
# As an overview this works, but we need to zoom in to see more detail



# Showing seat differences on individual flights allows patterns to be spotted



# Showing seat differences on individual flights allows patterns to be spotted



# Summary table

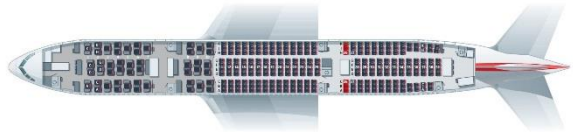
## Airline summary for August

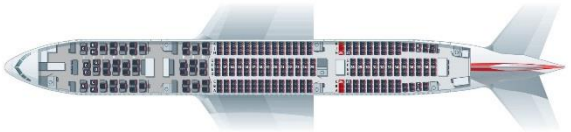
For flights performed by mainline carriers from China to the USA during August, there was a 97.6% match between scheduled and tracked flights. Any flights not matched were discarded in order to not influence the seat capacity comparison.

| | Flights | Scheduled seats | Actual seats | Seat diff. (sch. > act.) | Seat diff. (act. > sch.) | Schedule over-estimation |
|---|---|---|---|---|---|---|
| Air China | 282 | 109,570 | 91,050 | 19,140 | 620 | +18,520 |
| American Airlines | 150 | 35,490 | 34,551 | 960 | 21 | +939 |
| China Eastern | 213 | 65,268 | 65,540 | 0 | 272 | −272 |
| China Southern | 146 | 48,710 | 48,710 | 0 | | |
| Delta Air Lines | 154 | 40,829 | 40,709 | 337 | | |
| Hainan Airlines | 155 | 38,672 | | | | |
| United Airlines | 291 | 93,291 | | | | |

17% over estimation on capacity

More than two additional 777-200 aircraft in typical configuration per day

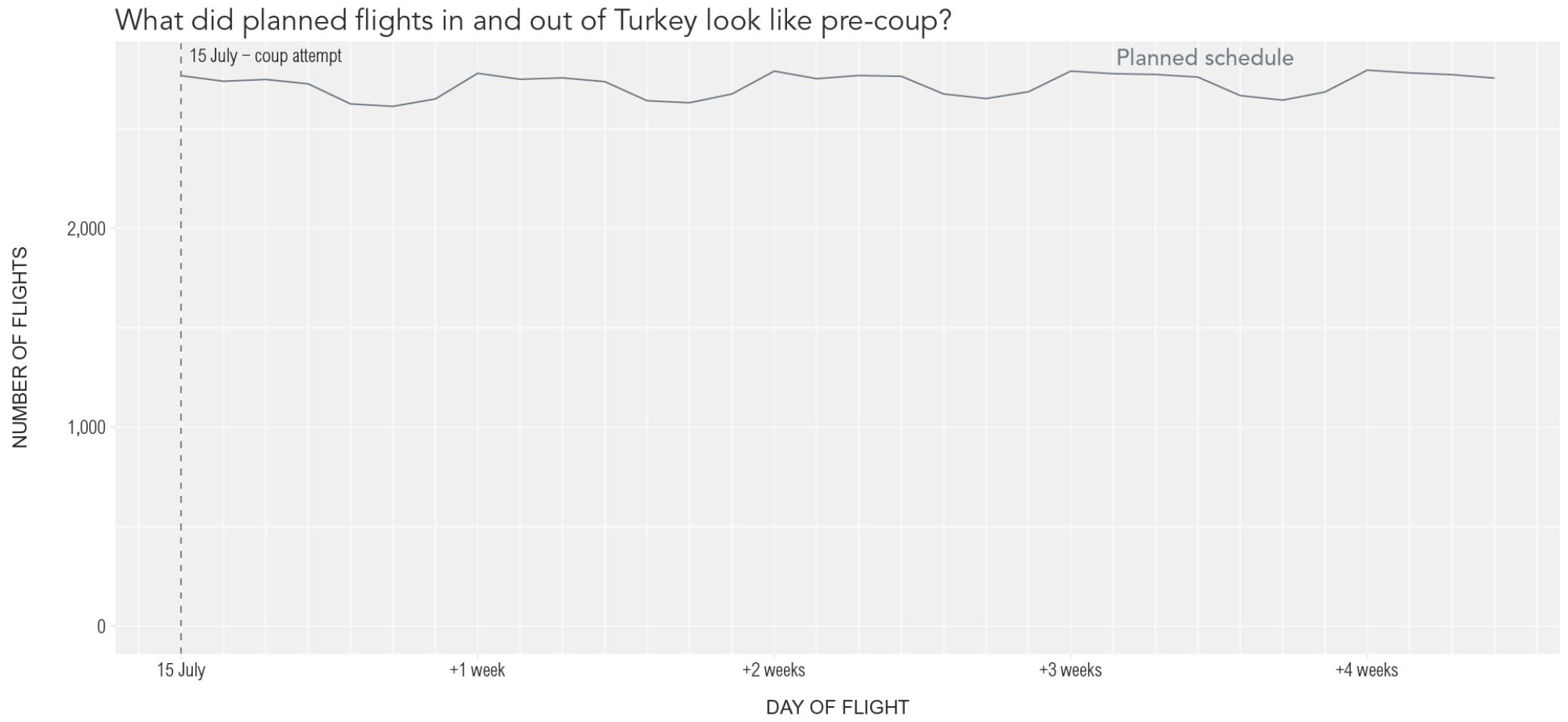




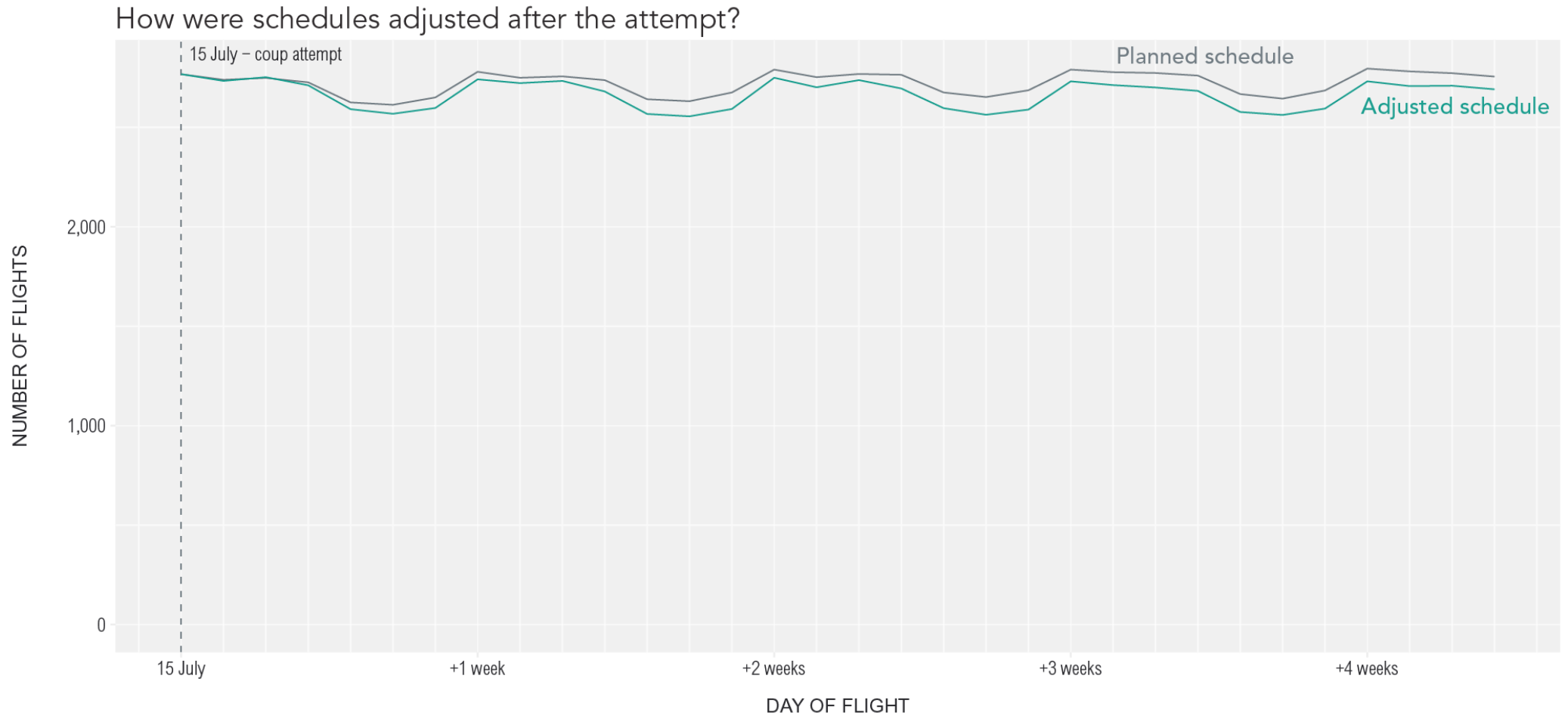
# Changes from schedule to actual

## Case study: Where actual deviates from scheduled

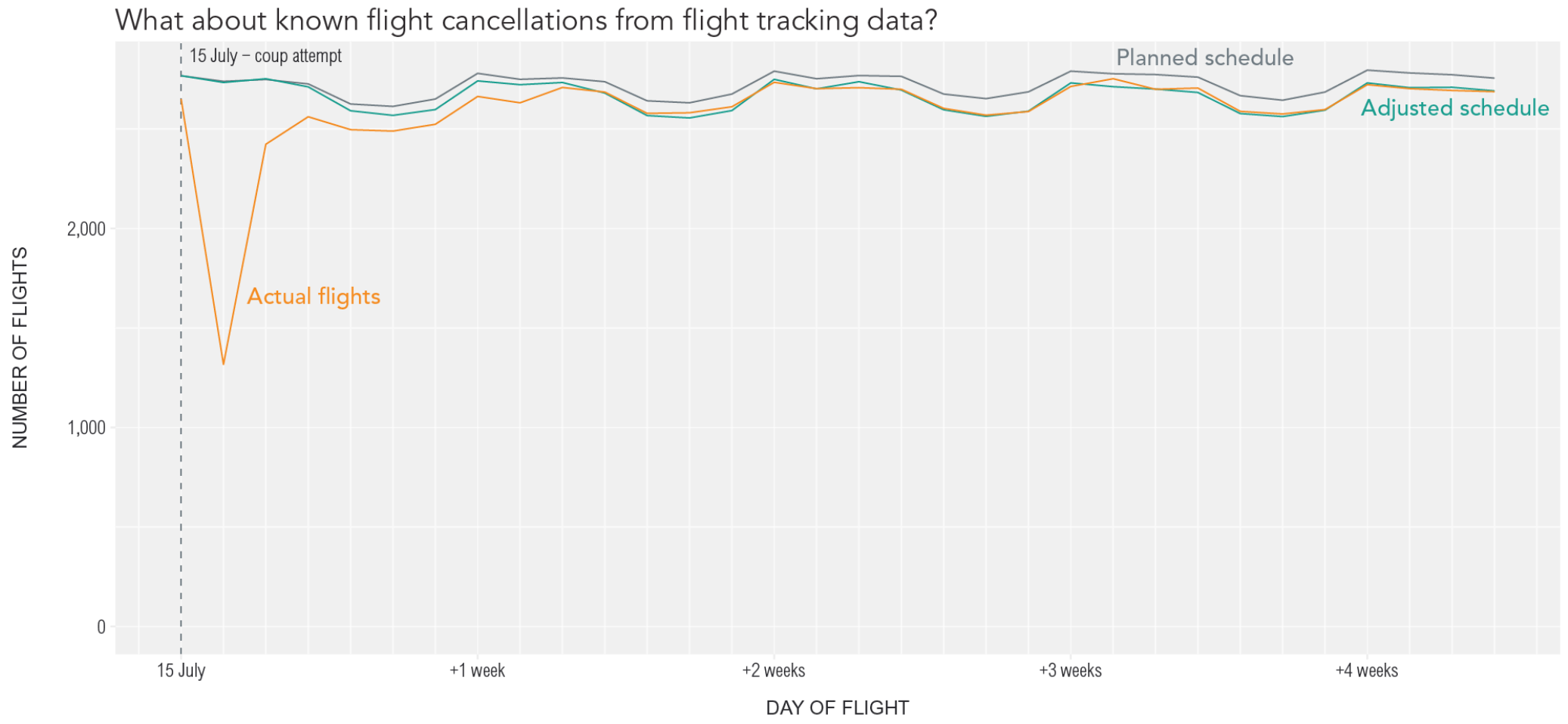
# These are flights according to the planned schedule



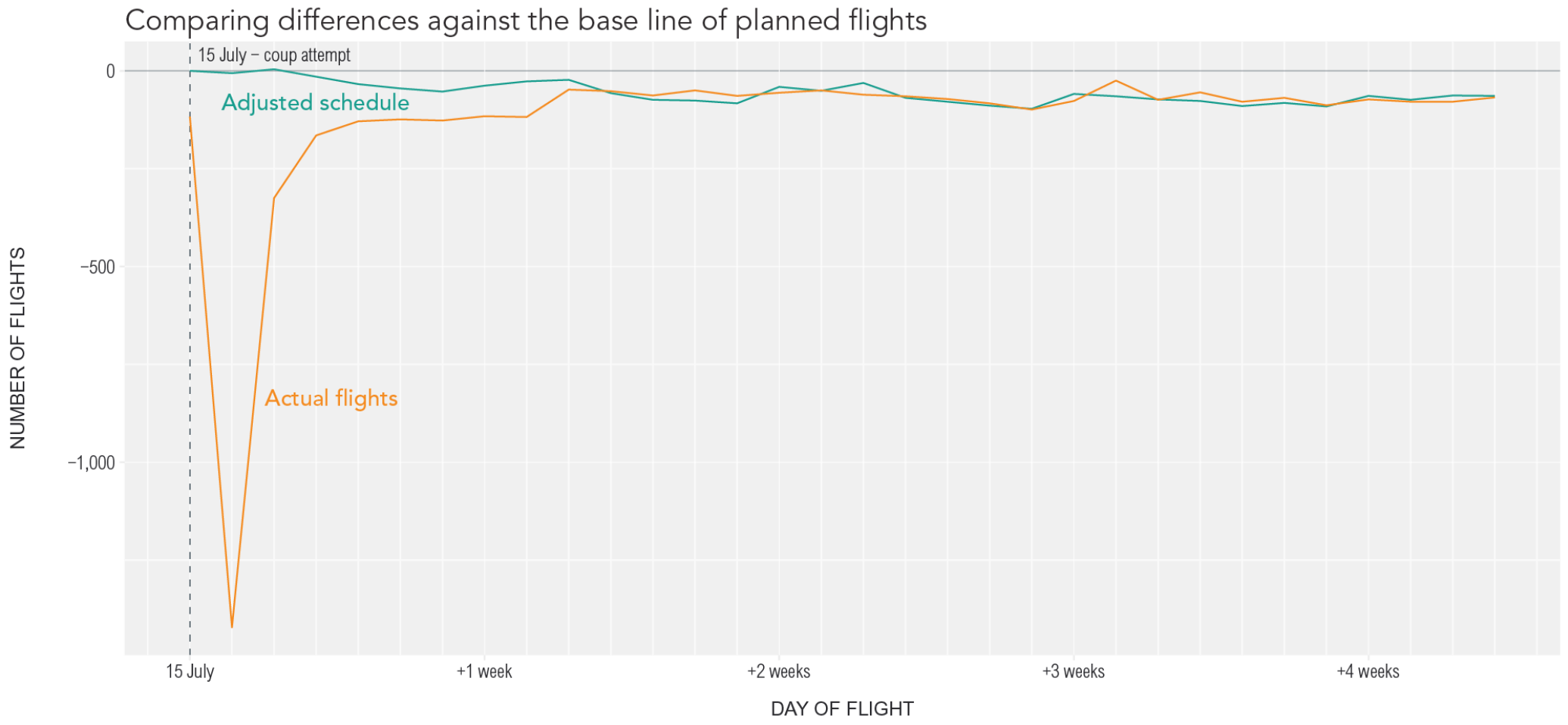
# Plus flights according to an adjusted schedule, taken daily from the time of the coup attempt



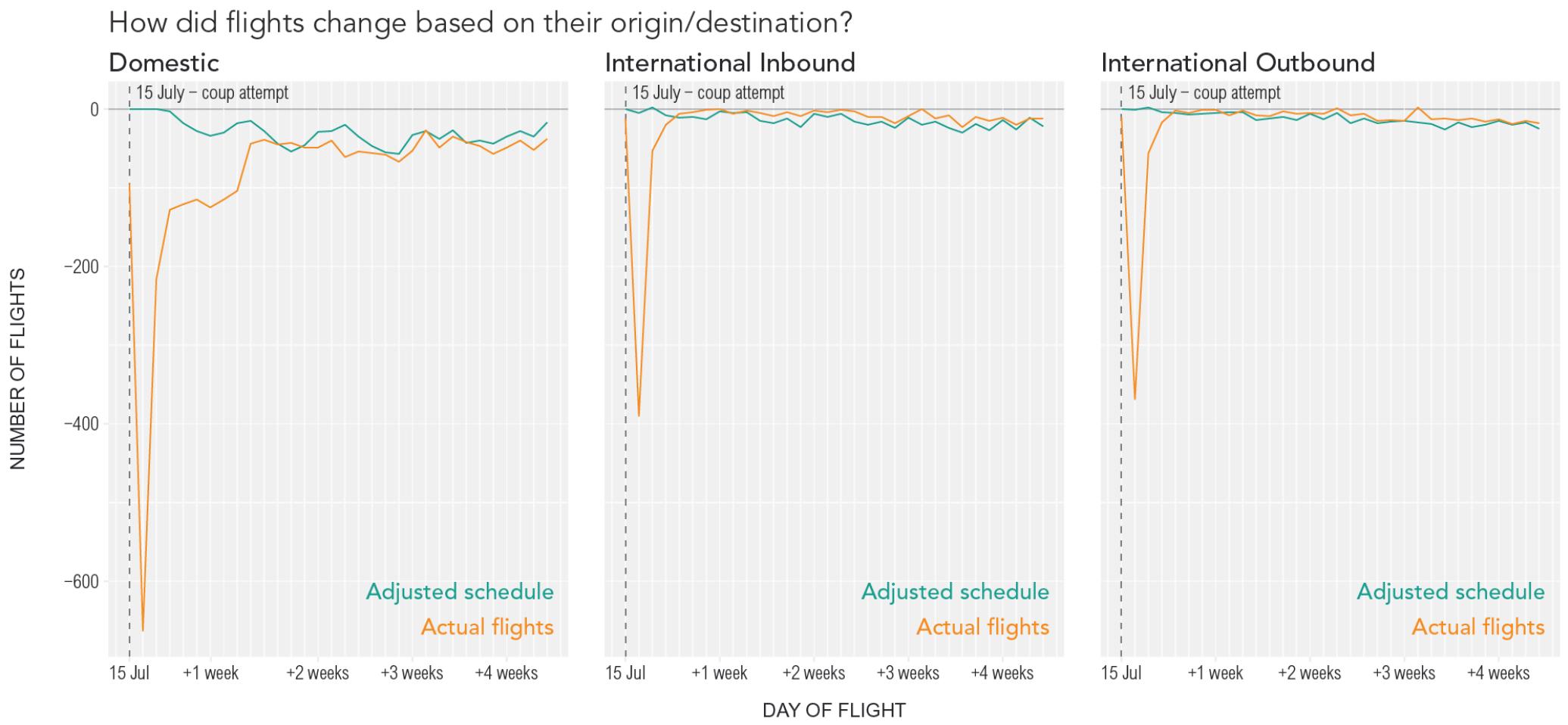
# And bringing back flight tracking data, omitting flights that had a ‘cancelled’ status



# Ways to make this clearer – show differences against a common baseline



# Did it make much difference if the flight was within Turkey?

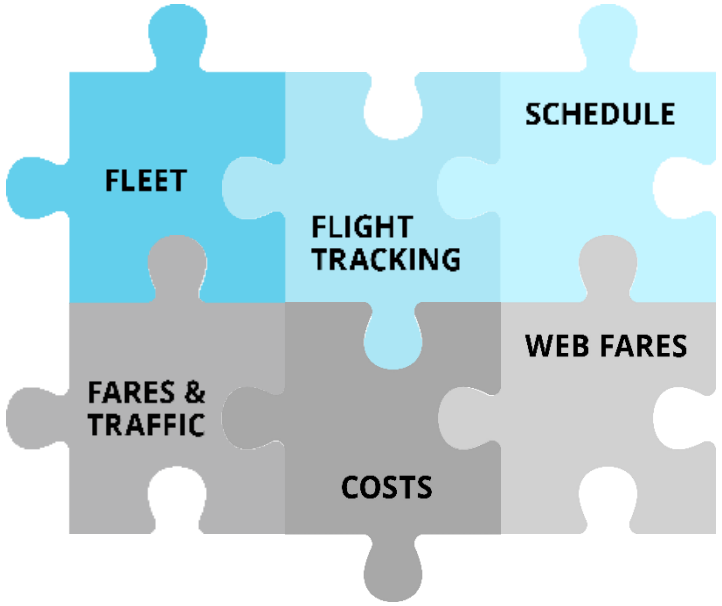


- *How do we help with the problem (and where does HPCC come in)?*

# The Fleet Flight Matching concept

## A new aviation data combination

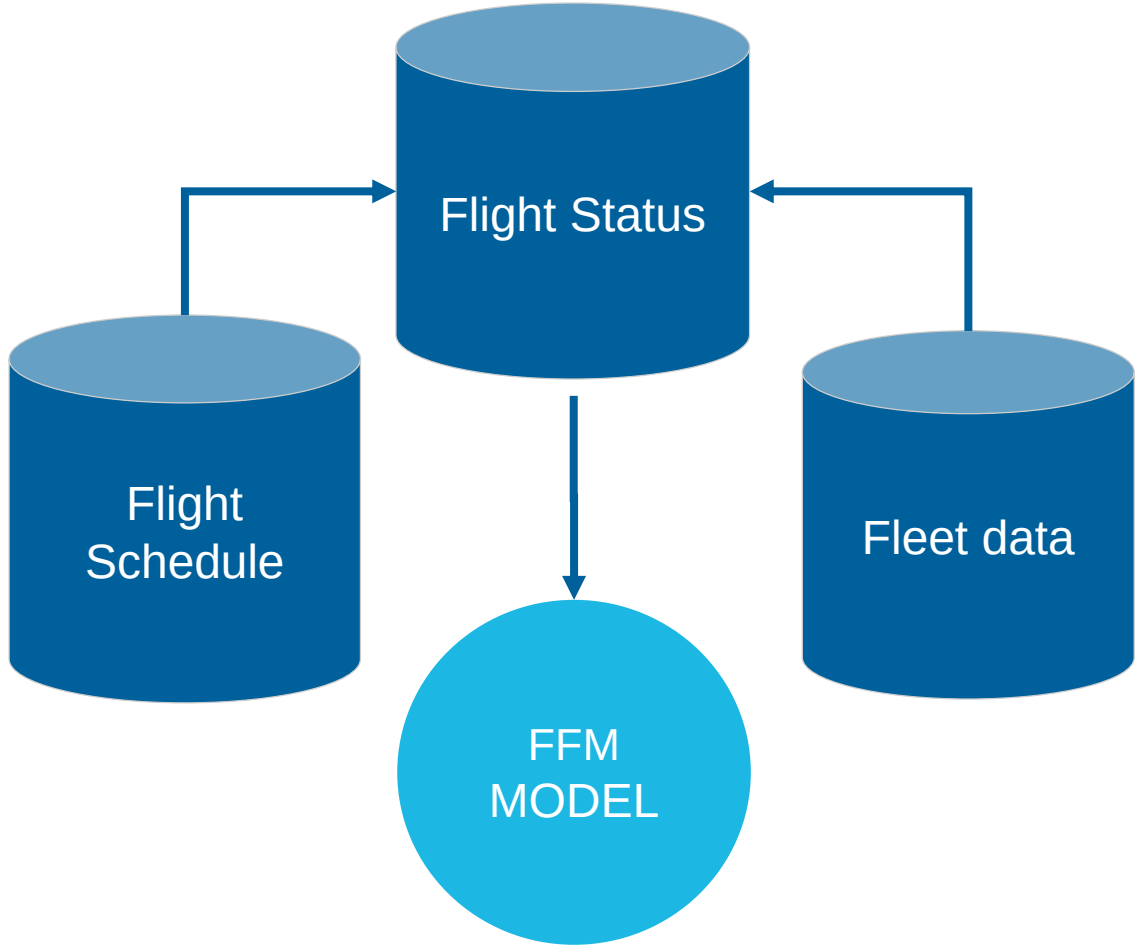
# Matching tail number to flight number using flight status




innovata

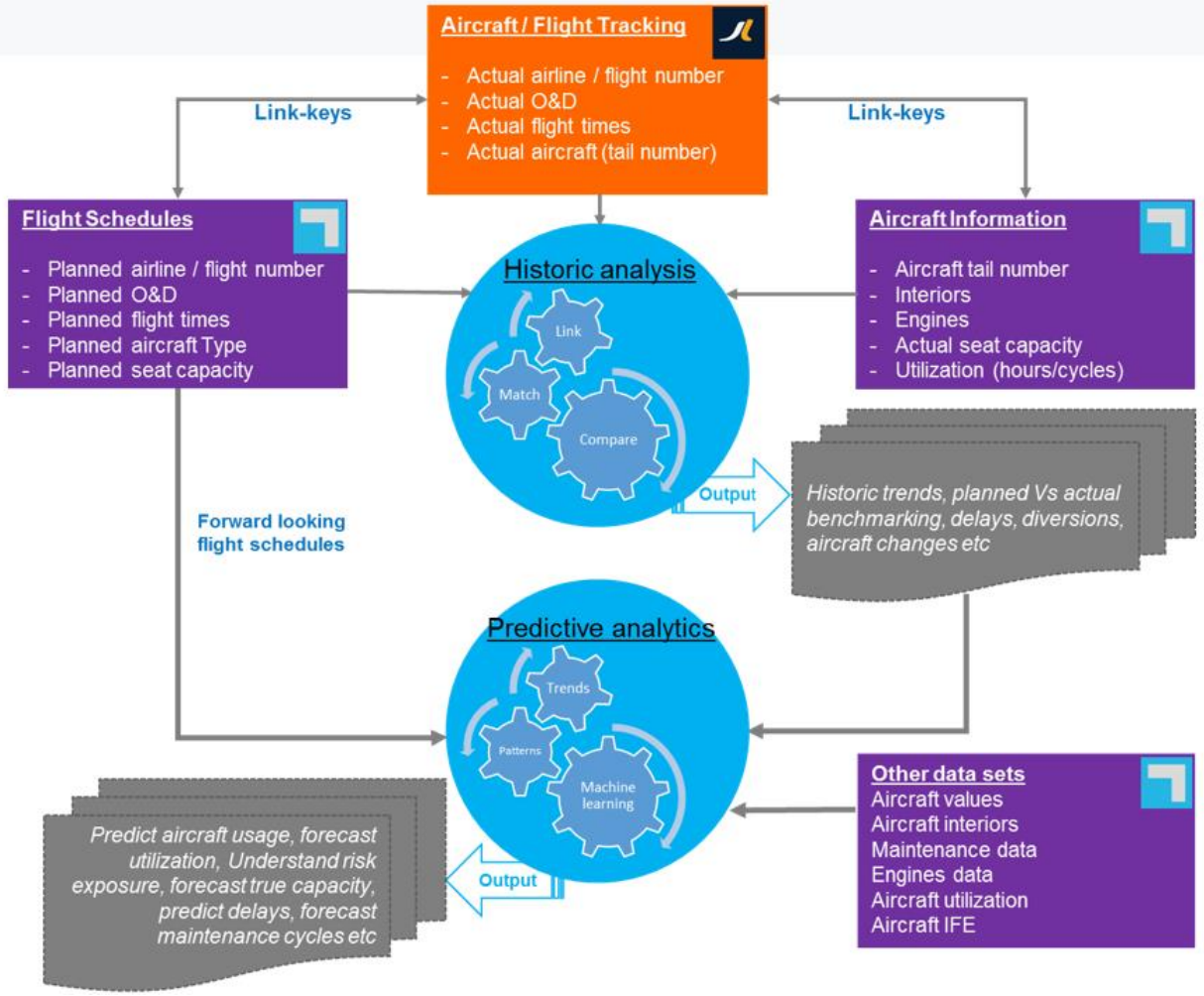
FLIGHTSTATS

FlightGlobal



# Fleet Flight Matching: A new data combination

- Accurate base data is combined to:
  - Improve historical analysis
  - Identify real events and changes to planned events
  - Drive more accurate predictions
- Critical in aiding strategic decision making and analysis. i.e. assumptions based on the most accurate information available



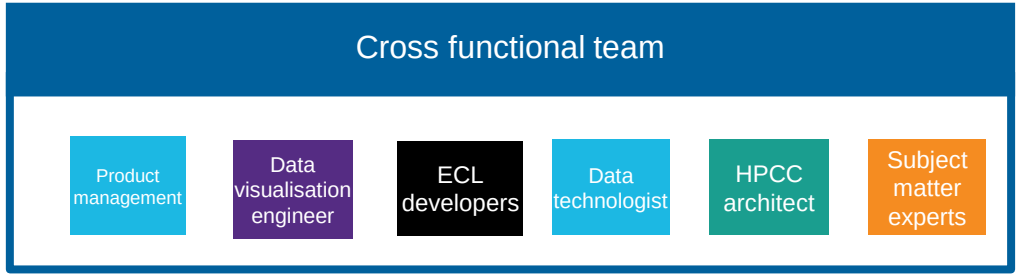
# Fleet Flight Matching capabilities

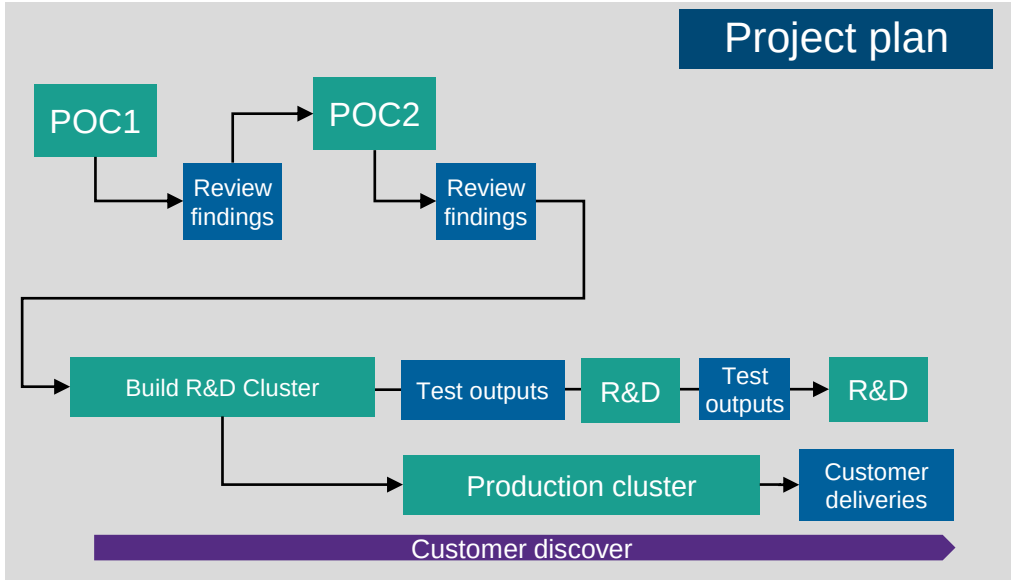
| | HISTORIC | PROJECTION |
|---|---|---|
| **Aircraft attributes**<br>Up to 350 different aircraft attributes:<br>E.g. Seats, engines, service life, IFE | Actual capacity flown | Capacity forecast |
| | Aircraft hours & cycles flown | Utilisation prediction |
| | Actual flights with product offering | Product offering probability mapped to schedule |
| **Aircraft activity**<br>How the aircraft is being used and maintained:<br>E.g. type of route flow, airports visited, time spent in activity | Maintenance flight identification | Maintenance cycle prediction |
| | Turn time analysis | Extra turn opportunity identification |
| | Planned vs real performance | Delay prediction and probability |

# Project approach

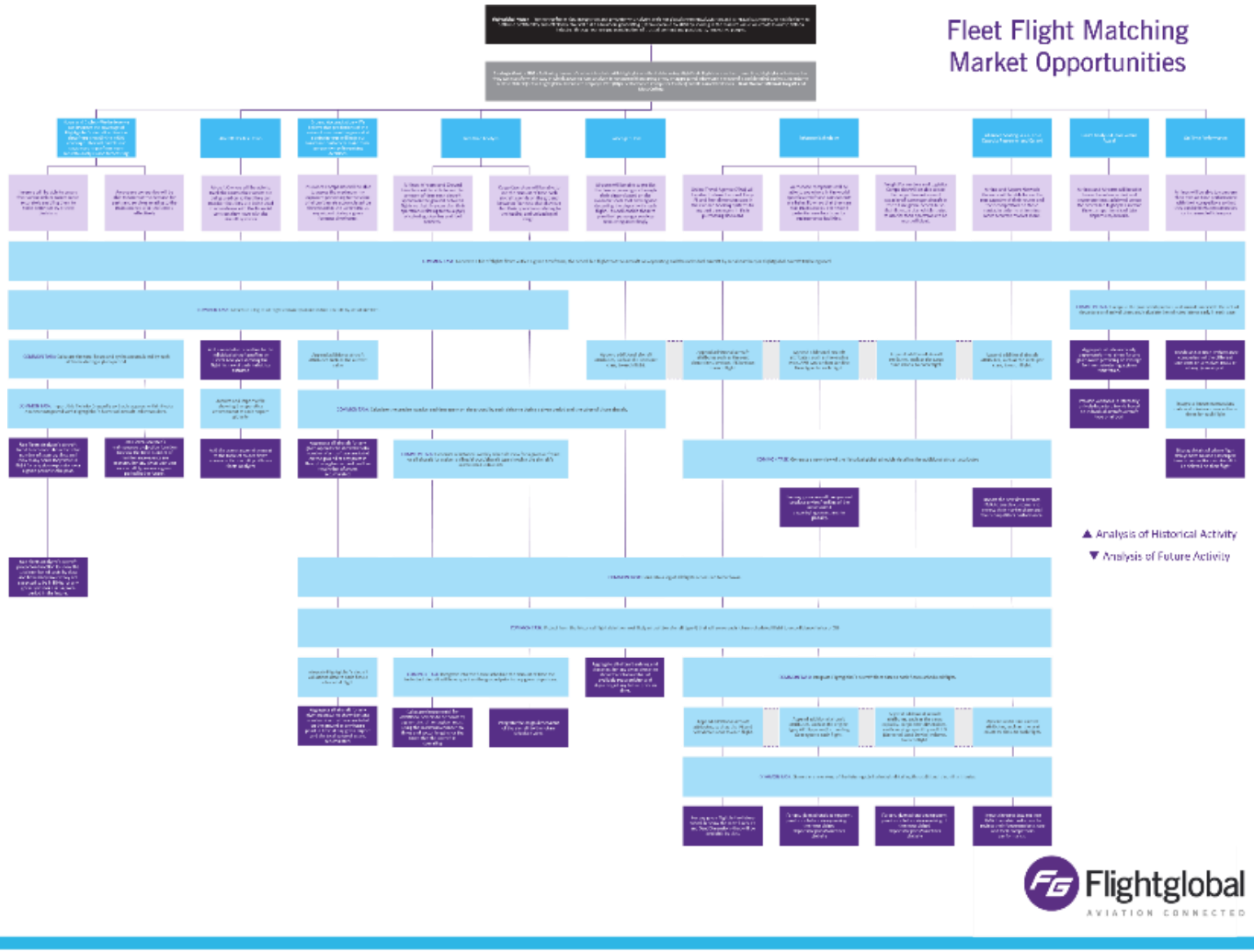
# Technology choice & team

- Technology choice
  - Rapid prototyping and iterative R&D
  - Scale to accommodate a growing dataset
  - Interface to both explore and productionise outputs
- Skills
  - Cross functional team to blend commercial, data, and tech skills
- Objectives
  - Lean value methodology utilised to focus on quick, impactful returns
- Planning
  - Intensive PoCs to make leaps forward followed by consolidation and establishment of dedicated analytics team



# Prioritising customer bets

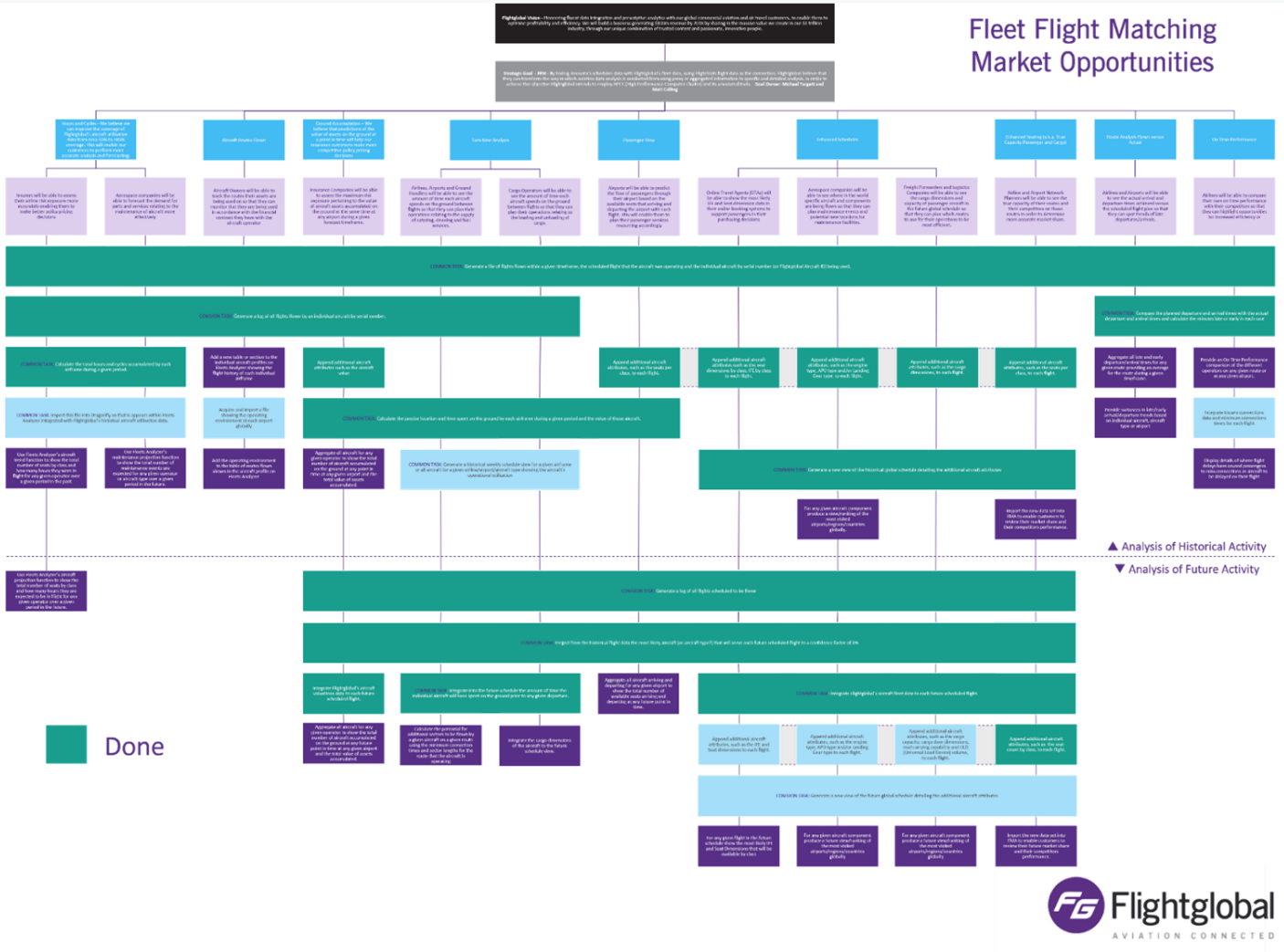


# Common tree outputs

Produce common outputs

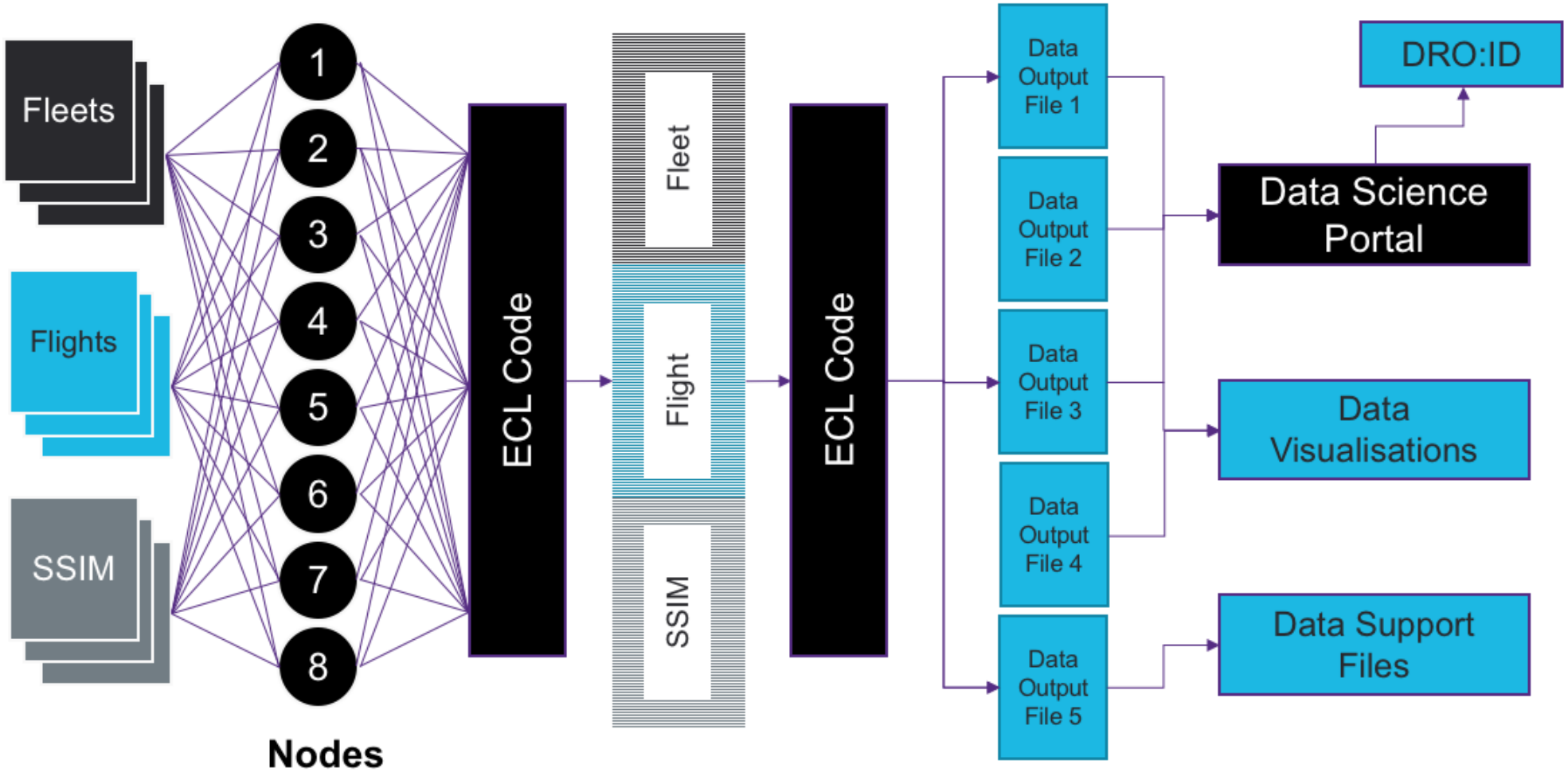
Process into filtered data view

**Analyse, visualise, & communicate**

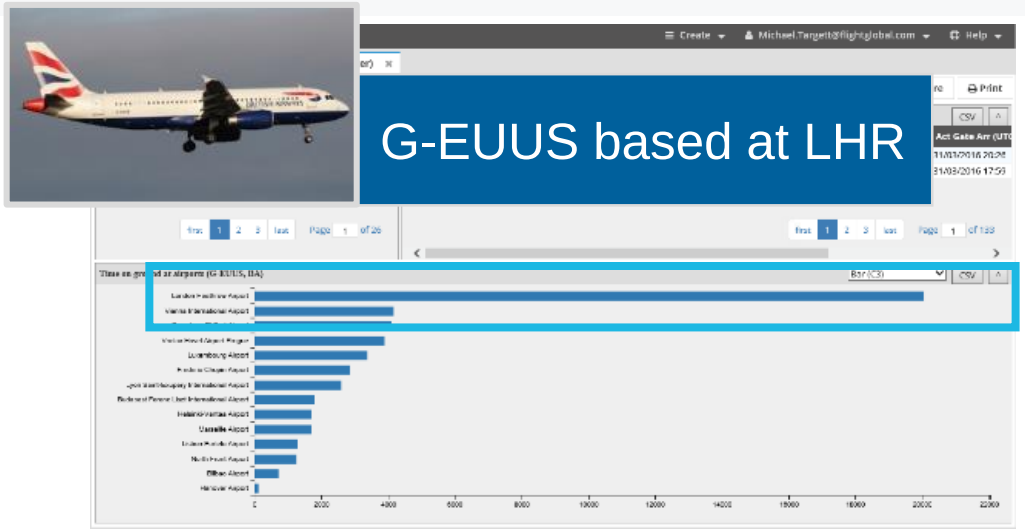


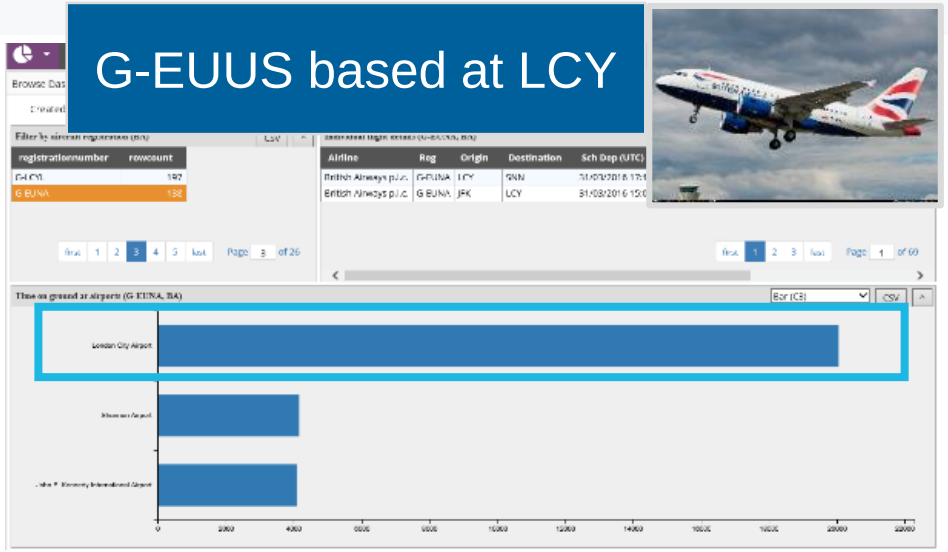
# POC Execution

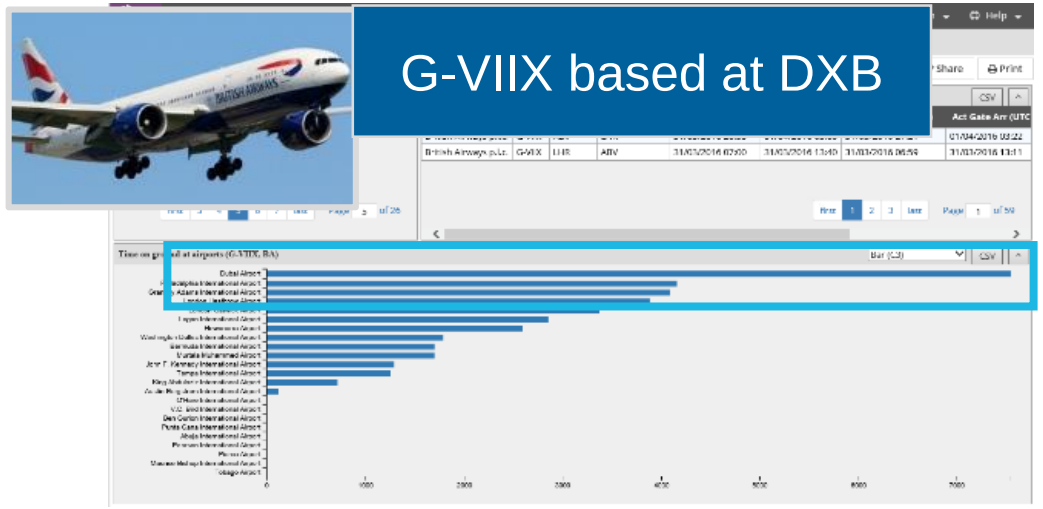
# POC approach

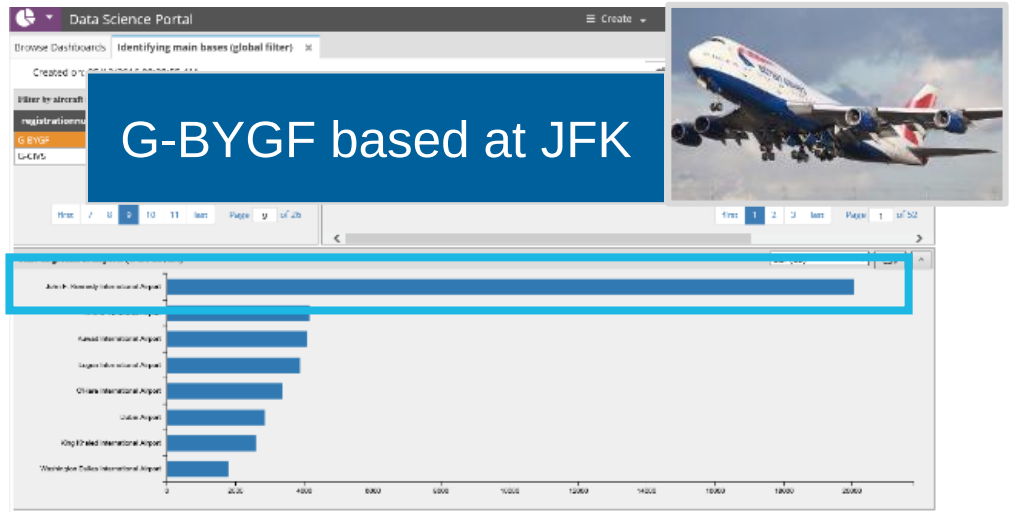


# POC findings explored using Data Science Portal



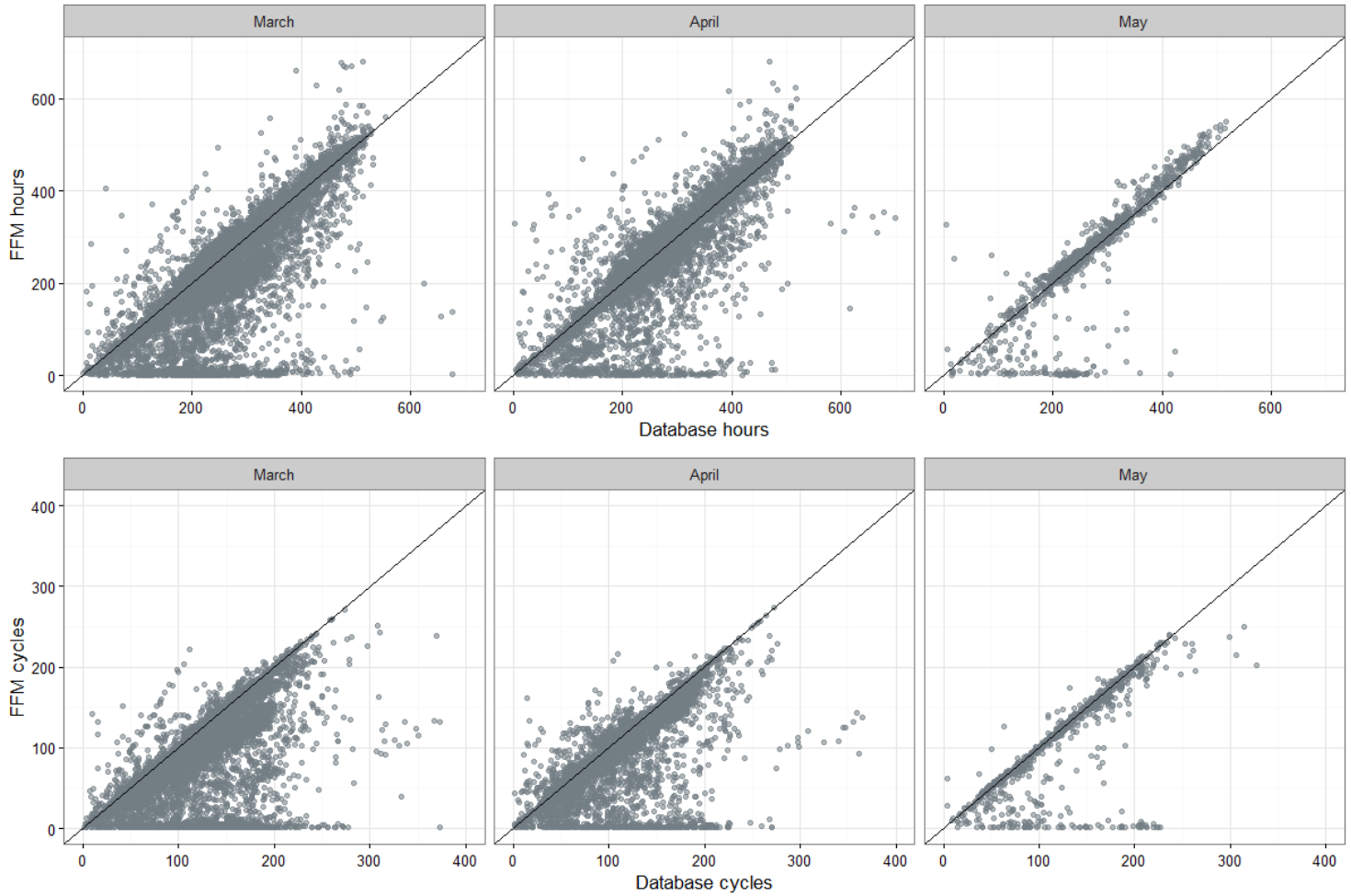



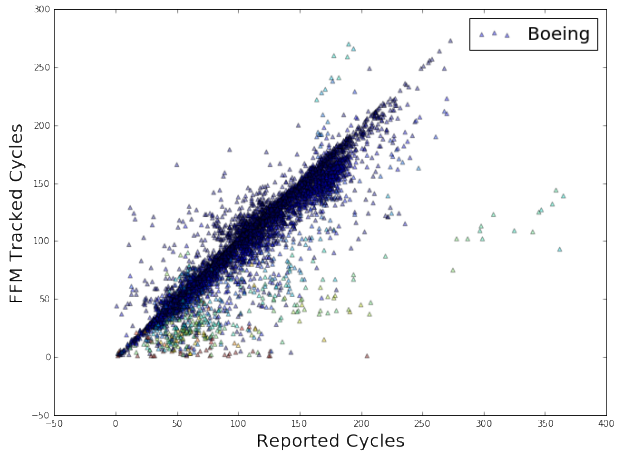



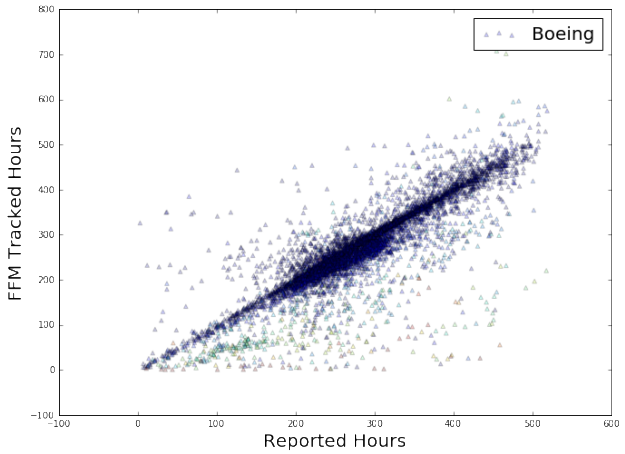
# Initial aircraft utilisation plots

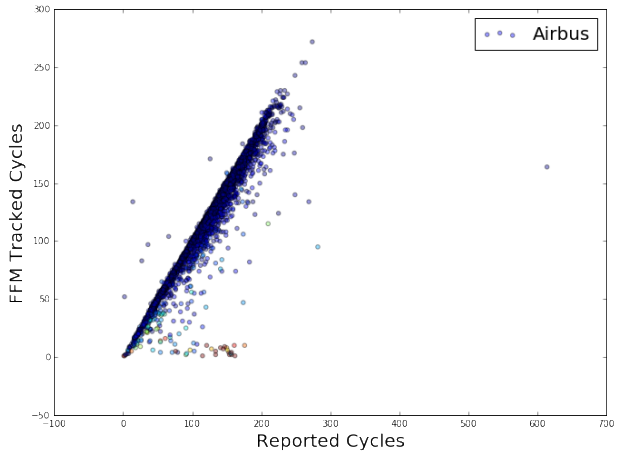
- Aircraft hours and cycles calculated through FFM plotted against known records
- Early results on infilling missing data display significant scatter

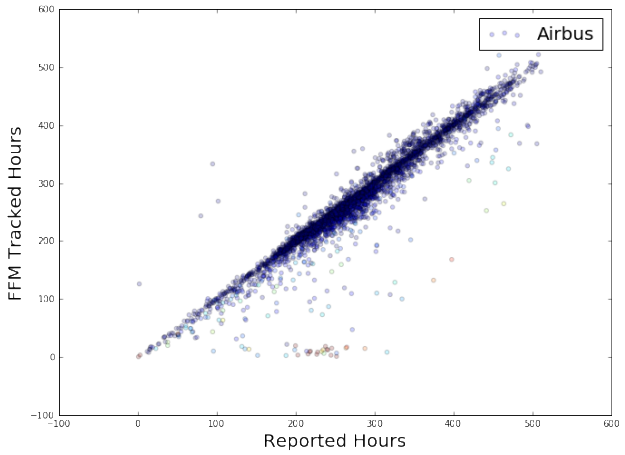


# Improvements yielded through iterations




- Subsequent iterations improved through model adjustments yielded positive results
- Highlighting possible industry benchmarking opportunity

# Customer discovery

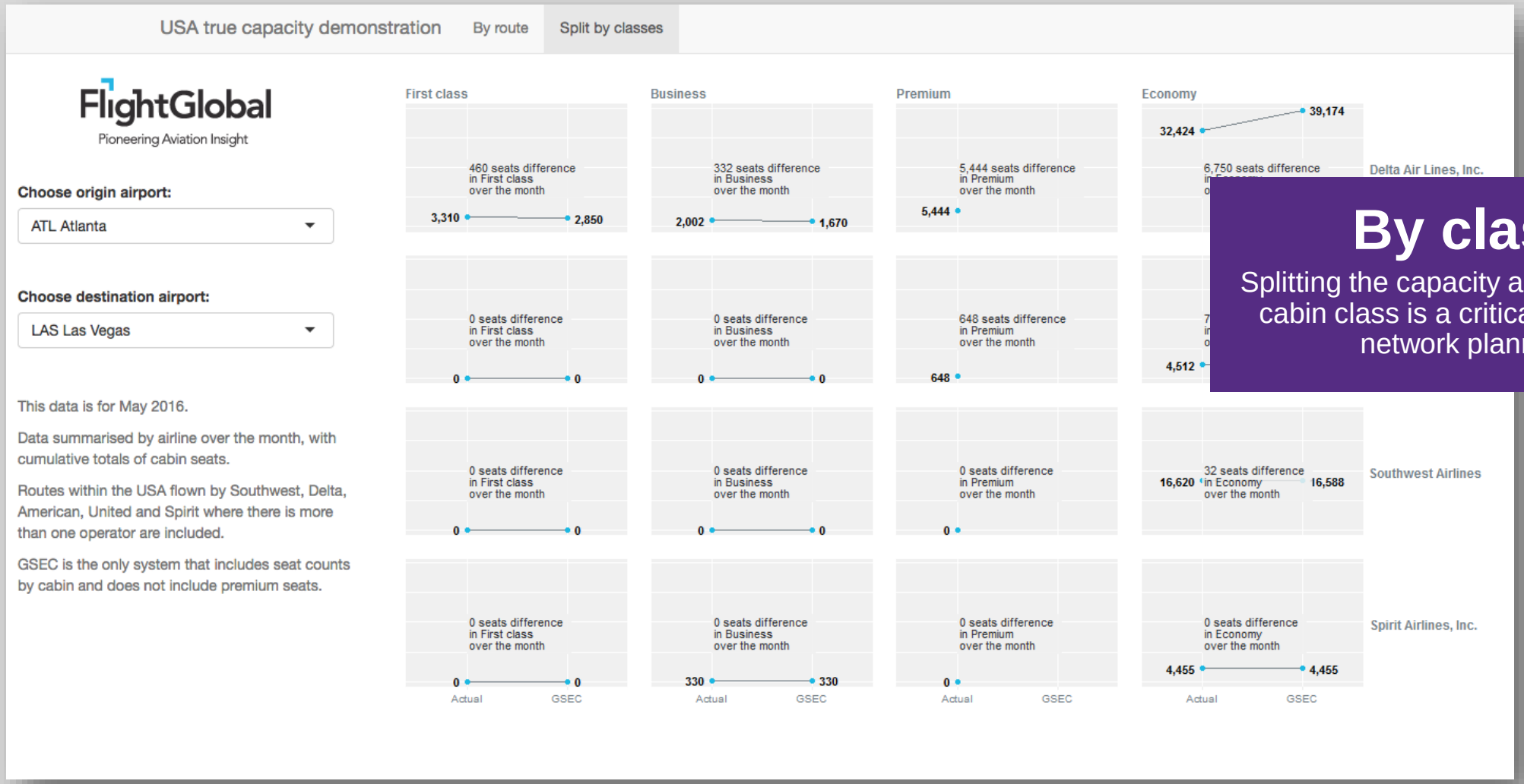
Test, learn, refine

# Examples: True capacity


**Highlighting equipment & seating change**

**Comparing multiple data sources for customer demonstration**

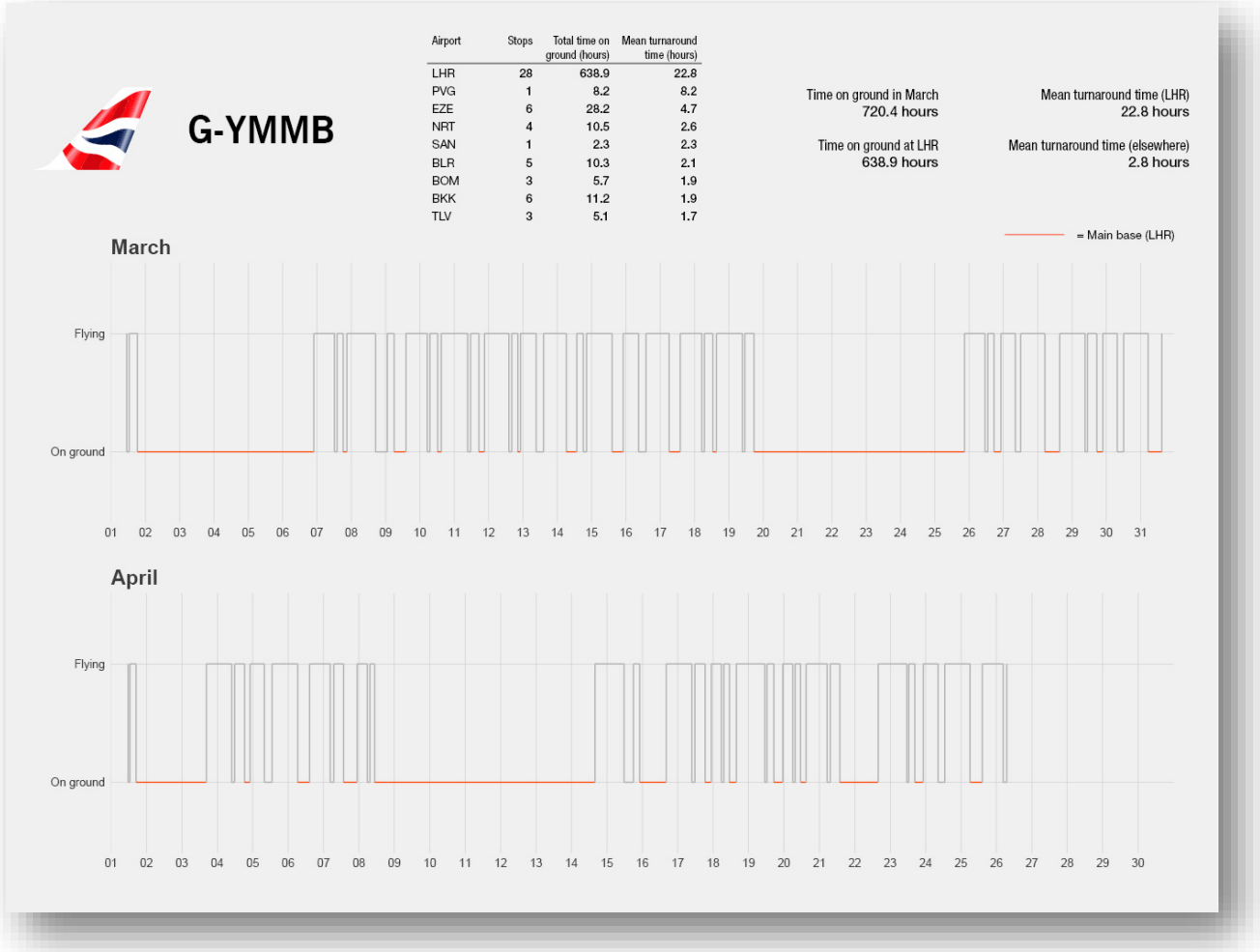
# Examples: True capacity




## By class

Splitting the capacity analysis out by cabin class is a critical feature for network planning

# Examples: Aircraft time on ground



G-YMMB

| Airport | Stops | Total time on ground (hours) | Mean turnaround time (hours) |
|---|---|---|---|
| LHR | 28 | 638.9 | 22.8 |
| PVG | 1 | 8.2 | 8.2 |
| EZE | 6 | 28.2 | 4.7 |
| NRT | 4 | 10.5 | 2.6 |
| SAN | 1 | 2.3 | 2.3 |
| BLR | 5 | 10.3 | 2.1 |
| BOM | 3 | 5.7 | 1.9 |
| BKK | 6 | 11.2 | 1.9 |
| TLV | 3 | 5.1 | 1.7 |

Time on ground in March
720.4 hours

Time on ground at LHR
638.9 hours

Mean turnaround time (LHR)
22.8 hours

Mean turnaround time (elsewhere)
2.8 hours

= Main base (LHR)



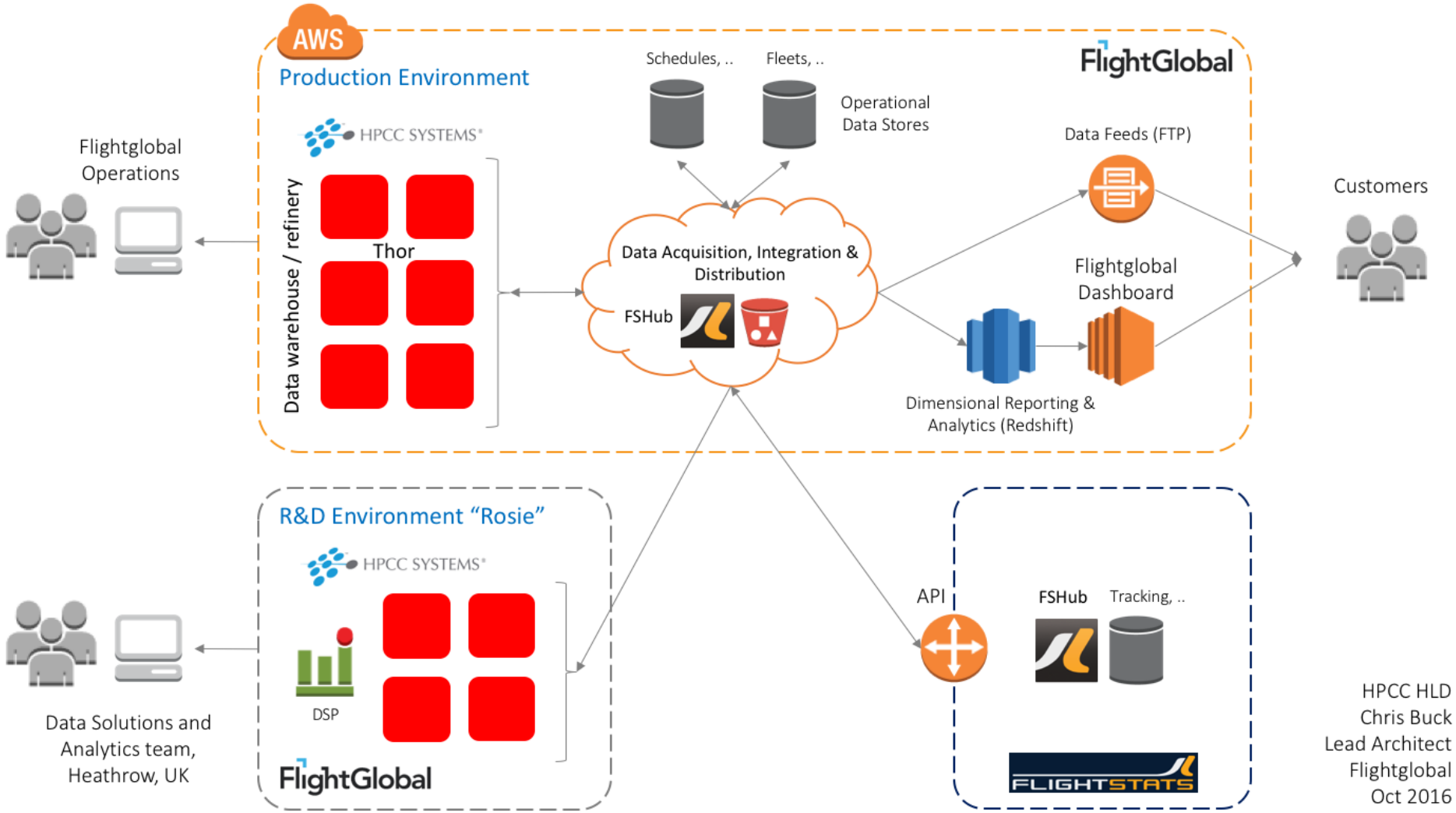
## Aircraft maintenance and movements

FFM facilitates analysis of individual aircraft time/date records

| | |
|---|---|
| Commercial aircraft in service | **28,000 a/c** |
| Daily flight records | **100k rows** |
| Over a 6 month period | **180 days** |
| *For pattern and trend analysis* | **504 billion** |

# Moving FFM into production

# FlightGlobal HPCC HLD




HPCC HLD
Chris Buck
Lead Architect
Flightglobal
Oct 2016

# Future considerations

- Efficient data ingress a priority
- FFM database already bigger than entire Fleet database (over 50 years of aircraft history)
- HPCC x14 faster at running daily matching package than the SQL equivalent.
- Within HPCC, functions like De-Dupe and Rollup quickly and easily deal with common issues such as multiple flight status updates.
- Customer use case include real-time / near real-time data matching

1.5 terabytes from 221 days; 100,000 new flight record per day

Additional datasets mean that scale is essential with the FFM solution

"HPCC technology offers the required **scale**, **processing speed** and **storage capabilities** to turn Fleet Flight Matching into a foundational data set"


Pioneering Aviation Insight