Using Splunk to gain insights into airline safety data

FN1252

Cory Syvenky | Sr. Cloud Analyst, WestJet

September 2018 | Version 1.0

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

splunk> .conf18

CORY SYVENKY

Senior Cloud Analyst, WestJet

splunk> .conf18

Who is the speaker?

@spsavvy

- I’m a Splunk Certified Power User; this is my 3rd .conf.
- I’m an aspiring Data Scientist.
- I’m a Licenced Private Pilot.
- I work in the IT department for a major Canadian airline.
- We use Splunk for Operational Intelligence.

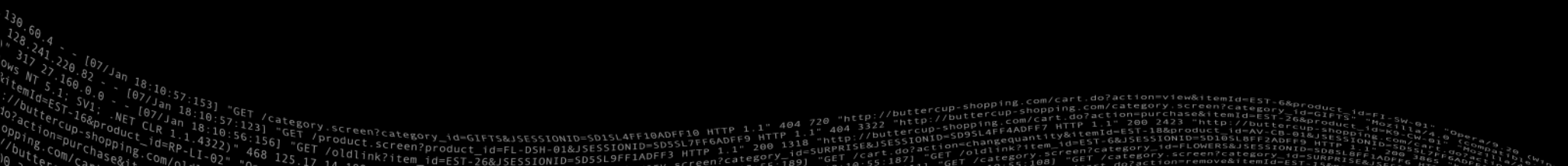
What is going to be covered?

Safety Data from Transport Canada and the NTSB

- Insights from plain text.
- Heatmap by Date.
- Clustered occurrences by Region.
- Review the Canadian and American dashboards.
- Minimal conf needed to process XML.
- Use of OneDrive for data file storage.

Data captured over-the-air via Raspberry Pi based receiver.

- Accessing data from IoT hardware.
- Do commercial aircraft speed?
  - By altitude and proximity to airport.
- Using GPSBabel and Google Maps for 3D visualization.
- Stratux app demo of Orlando Airport.



Why do this?

- Curiosity. Curiosity about flight safety. Curiosity about the challenge.
- Being able to combine two interests (Aviation, Data Visualizations).
- An interest in expanding my ability to gain insights from data.
- Splunk makes it easy.

Incident and accident data from Transport Canada and the NTSB

splunk> .conf18

About these Open Data Sources

- Transport Canada (TC): Freely available data that I have been processing in my local Splunk environment since January ‘17.
- National Transportation Safety Board (NTSB): I have only recently uncovered this data set.
- Data Ingestion, being a good data steward:
  - receive daily summary email from TC.
  - automating the migration of XML data to Splunk index.

Visualizing Aviation Data

Assistance from Splunkbase add-ons makes this work easy

- Wordcloud Visualization for narrative

  https://splunkbase.splunk.com/app/3212

- Calendar Heat Map Visualization for days with highest incidents

  https://splunkbase.splunk.com/app/3162

- Cluster Map for occurrences by region [built-in].

- Dashboards for easy consumption.

Wordcloud app for narrative field

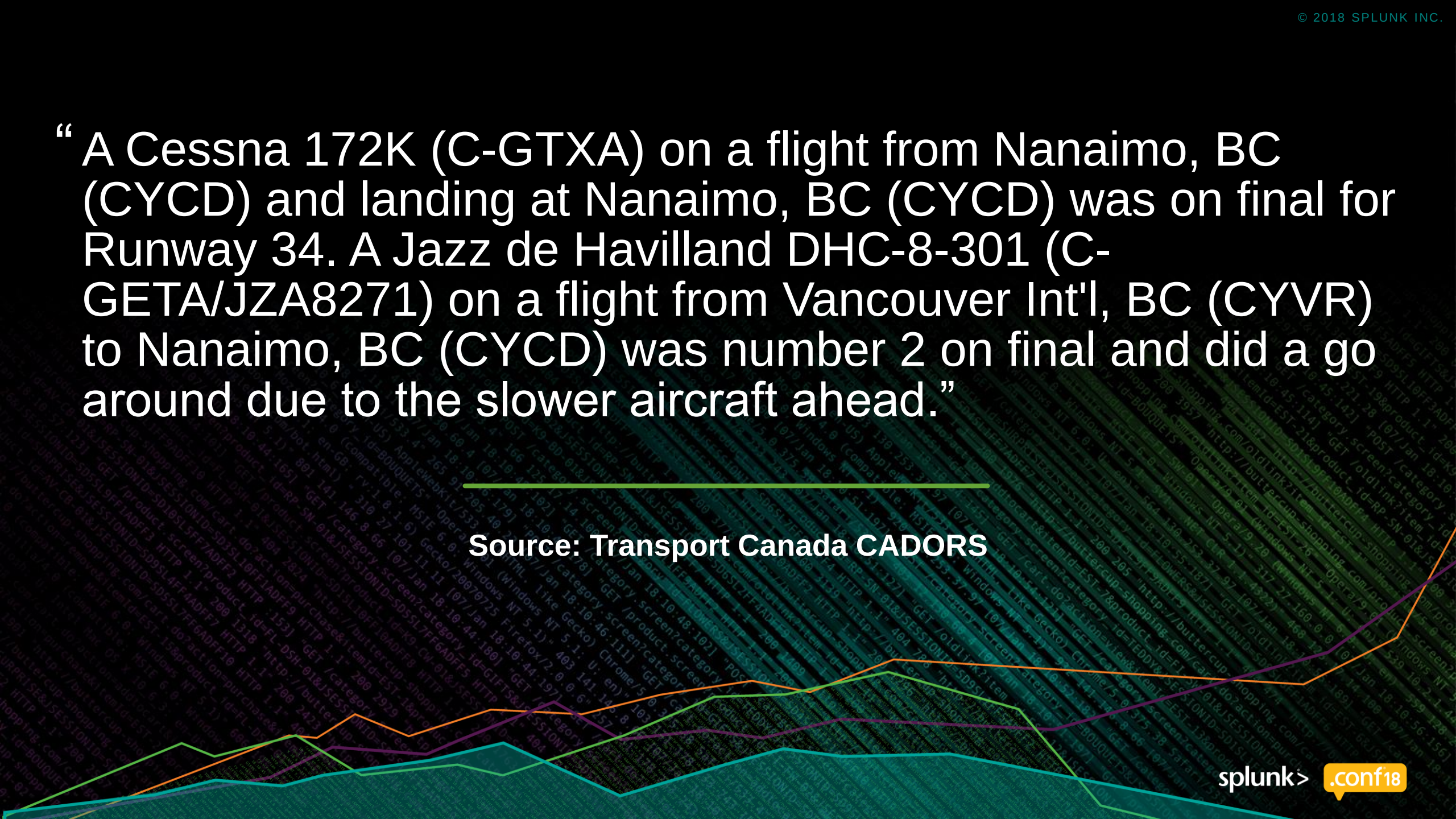
- The approach phase of flight is the most frequent cause of presence in reports (3.5 times more than the next highest cause, which is take-off).
- The narrative is in textual paragraph form from the TC investigators.
- The question:

  What words show up most frequently in these reports?

- The word-cloud represents, in relative size, the frequency of words used as they pertain to the approach phase of flight.

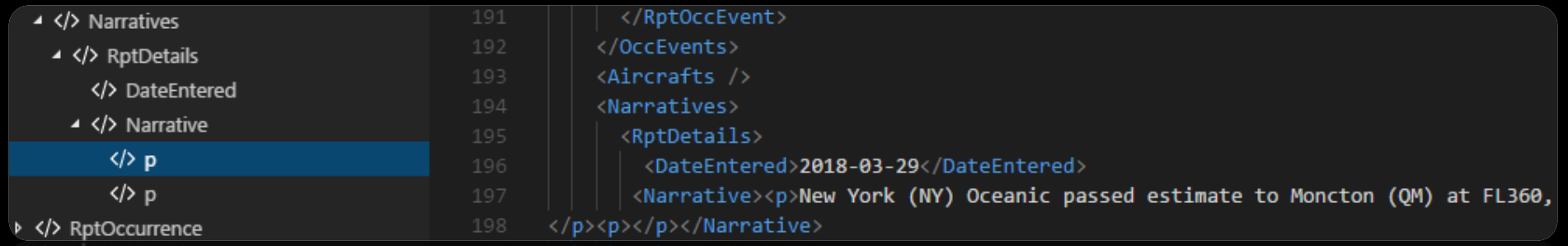
> “A Cessna 172K (C-GTXA) on a flight from Nanaimo, BC (CYCD) and landing at Nanaimo, BC (CYCD) was on final for Runway 34. A Jazz de Havilland DHC-8-301 (C-GETA/JZA8271) on a flight from Vancouver Int'l, BC (CYVR) to Nanaimo, BC (CYCD) was number 2 on final and did a go around due to the slower aircraft ahead.”

Source: Transport Canada CADORS



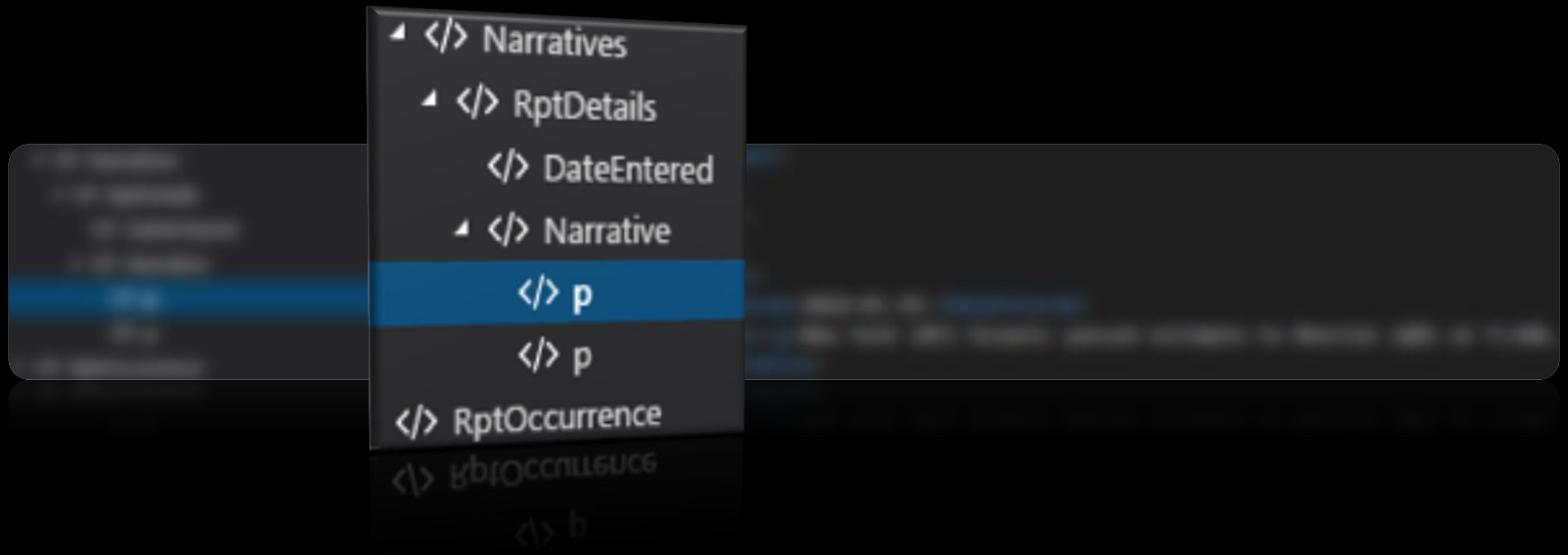
Visual Studio Code

see the structure of the data

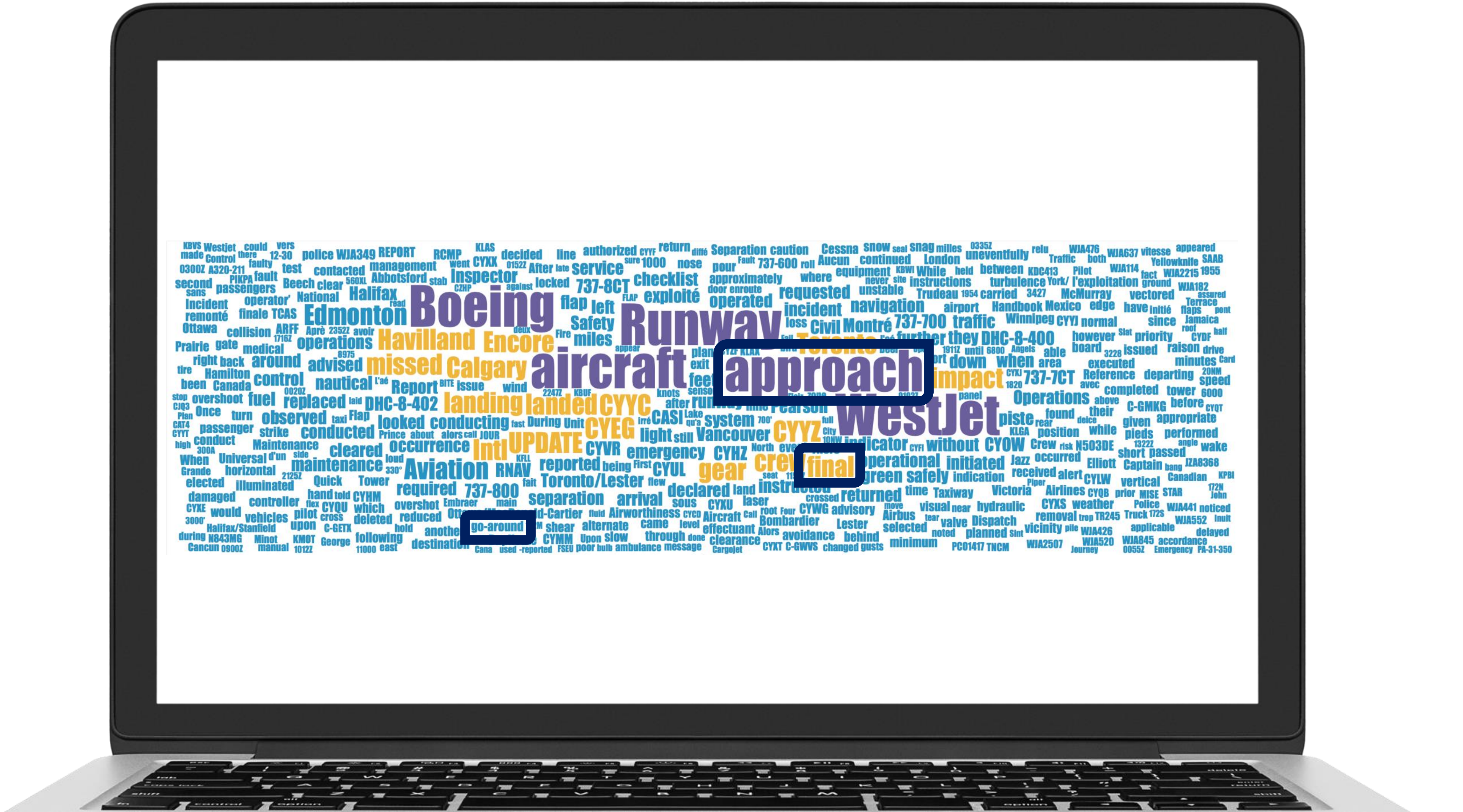


Visual Studio Code

see the structure of the data




police
RCMP
Boeing
Runway
aircraft
approach
WestJet
Edmonton
Havilland
Encore
Toronto
missed
Calgary
impact
landing
landed
CYYC
CYEG
UPDATE
Intl
CYYZ
crew
final
gear
Aviation
operations
Halifax
checklist
exploité
requested
incident
navigation
737-700
traffic
DHC-8-400
737-8CT
737-7CT
DHC-8-402
Operations
Vancouver
emergency
declared
instructed
returned
arrival
separation
Toronto/Lester
737-800
occurrence
maintenance
cleared
conducted
observed
replaced
nautical
control
around
advised
fuel
Pearson
runway
system
indicator
operational
initiated
green
safely
Inspector
service
Safety
miles
flap
left
Civil
Montréal
further
they
down
when
piste
CYOW
CYVR
CYHZ
CYUL
RNAV
Report
REPORT
Abbotsford
Airworthiness
Bombardier
Ottawa/MacDonald-Cartier
Halifax/Stanfield
go-around
Duplicate
clearance
avoidance
minimum
vicinity
removal
Dispatch
weather
Airlines
Winnipeg
Handbook
Mexico
McMurray
Reference
completed
departing
appropriate
performed
Captain
Elliott
Canadian
Cessna
Separation
caution
authorized
decided
uneventfully
turbulence
instructions
unstable
carried
approximately
operated
Trudeau
London
equipment
continued



Calendar Heat Map visualization showing days with highest incidents

- Transport Canada recorded a total of 17,392 CADORS reports in 2017, that's a national average of 47.6 incident/accidents per day. Thus far in 2018, we are at 10,906.

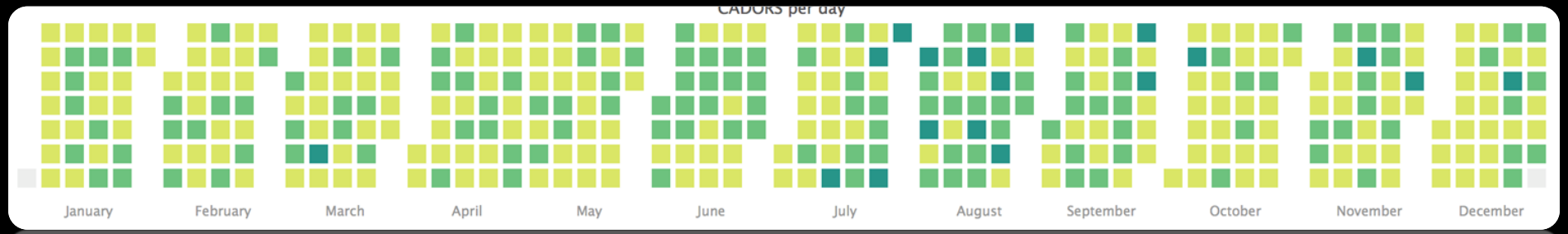
- The question:

  Which days had the highest amount of reports?

- We can easily get access to the days that stand out for volume levels.

Calendar Heat Map

making it easier to cluster by dates



Calendar Heat Map

making it easier to cluster by dates

Calendar Heat Map

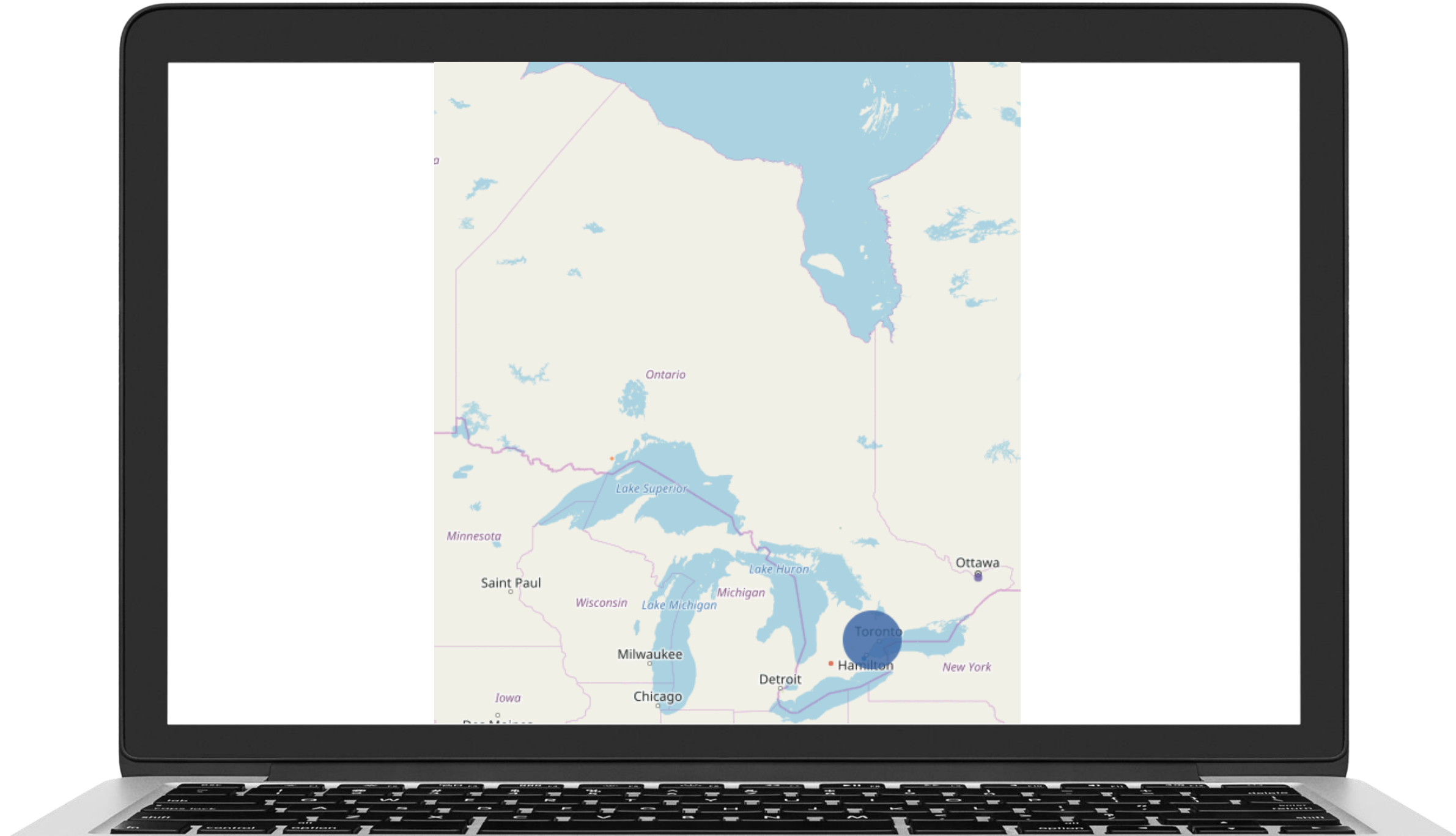
making it easier to cluster by dates

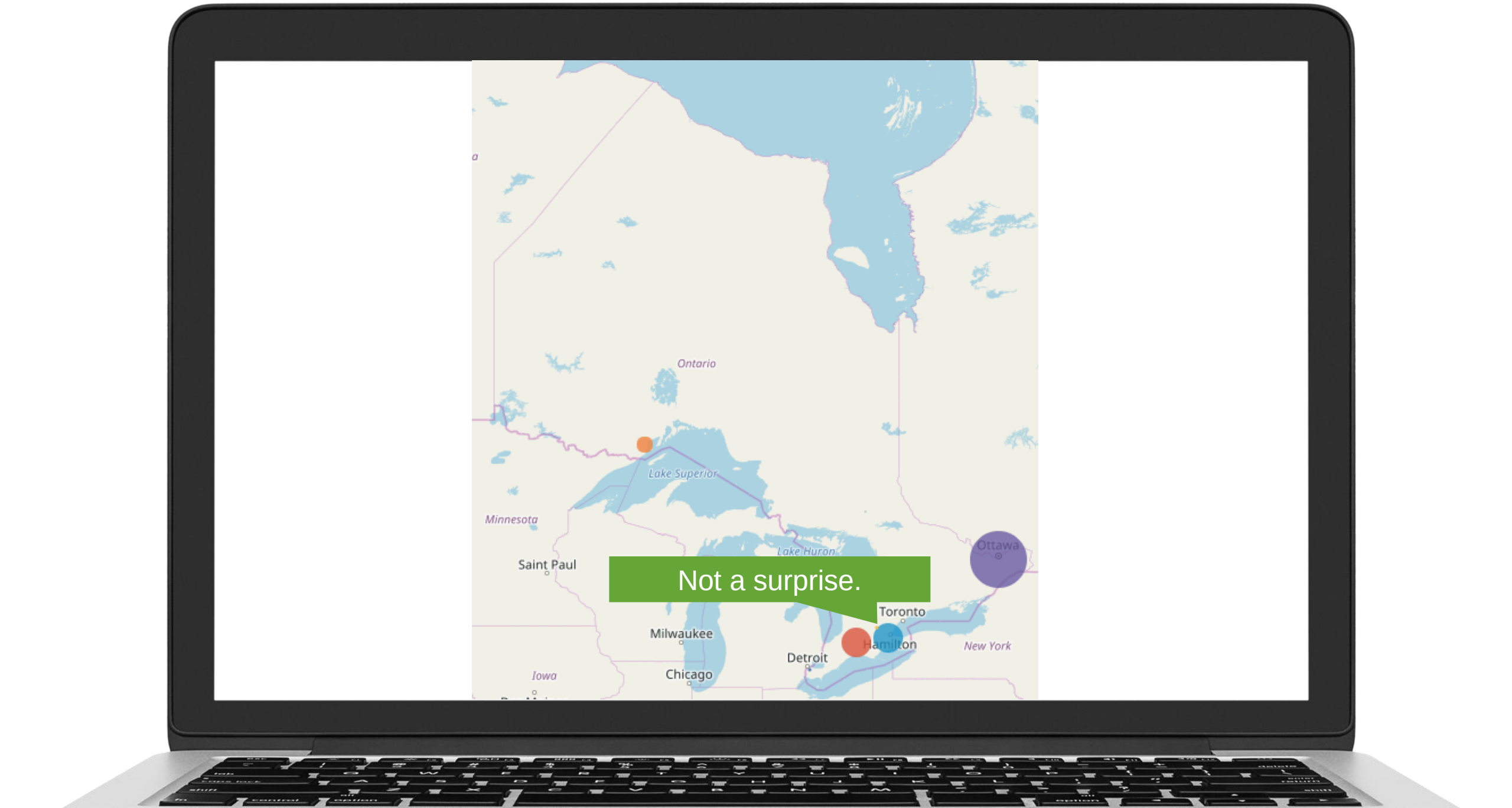
Cluster Map showing occurrences by region

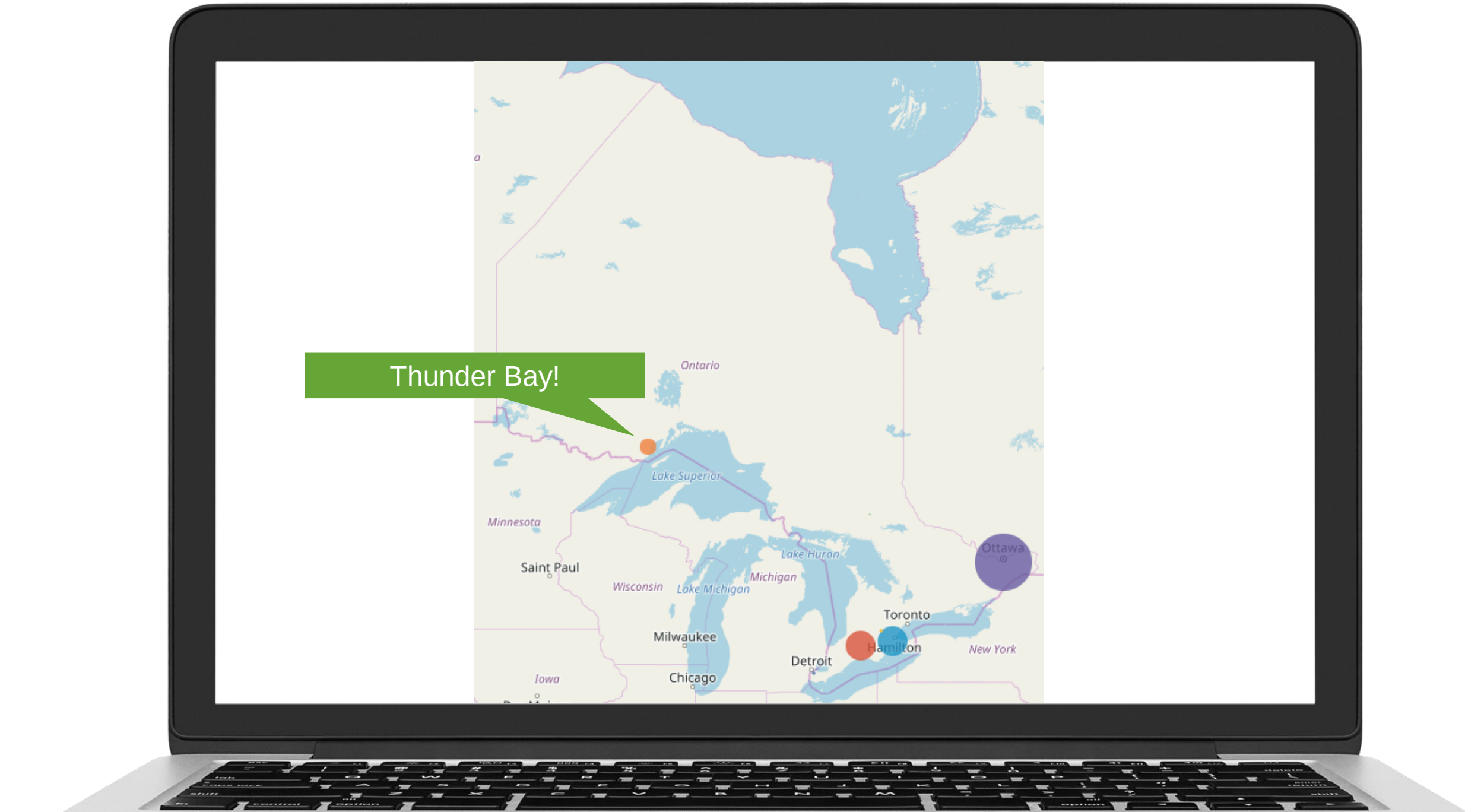
- One of the categorical groupings that Transport Canada has is a data field called "TCRegion".
- The question:

  After the heavy hitters were excluded from the data set, which would be the next highest?

- The frequency of occurrences for some airports is easy to guess, due to the size and the frequency of flights arriving and departing.

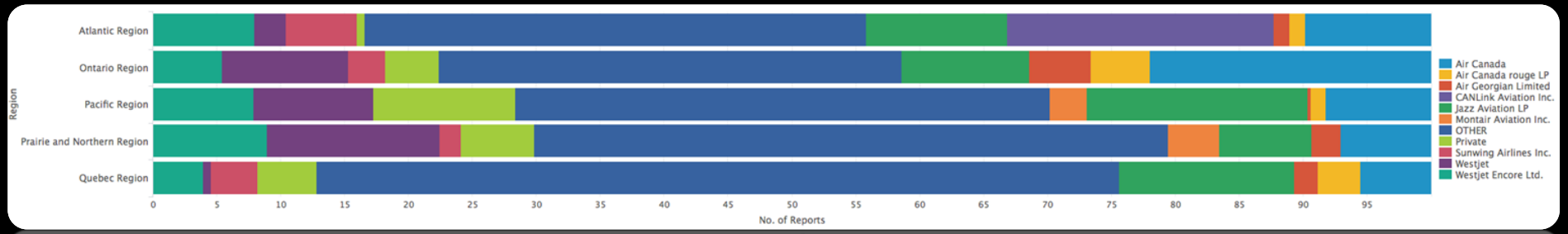






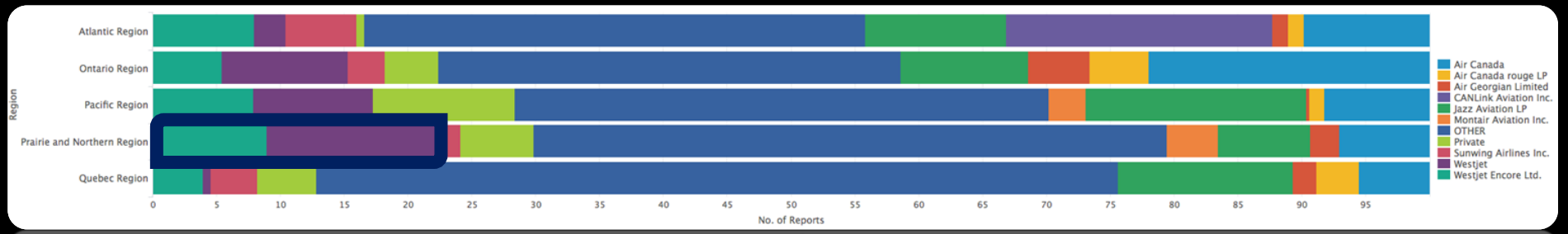
Stacked 100% bar chart

breaking down how each air carrier placed in each national region



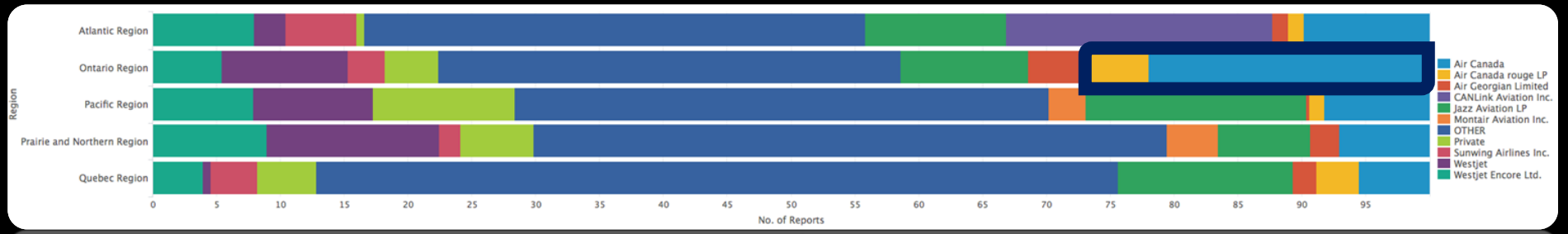
Stacked 100% bar chart

breaking down how each air carrier placed in each national region

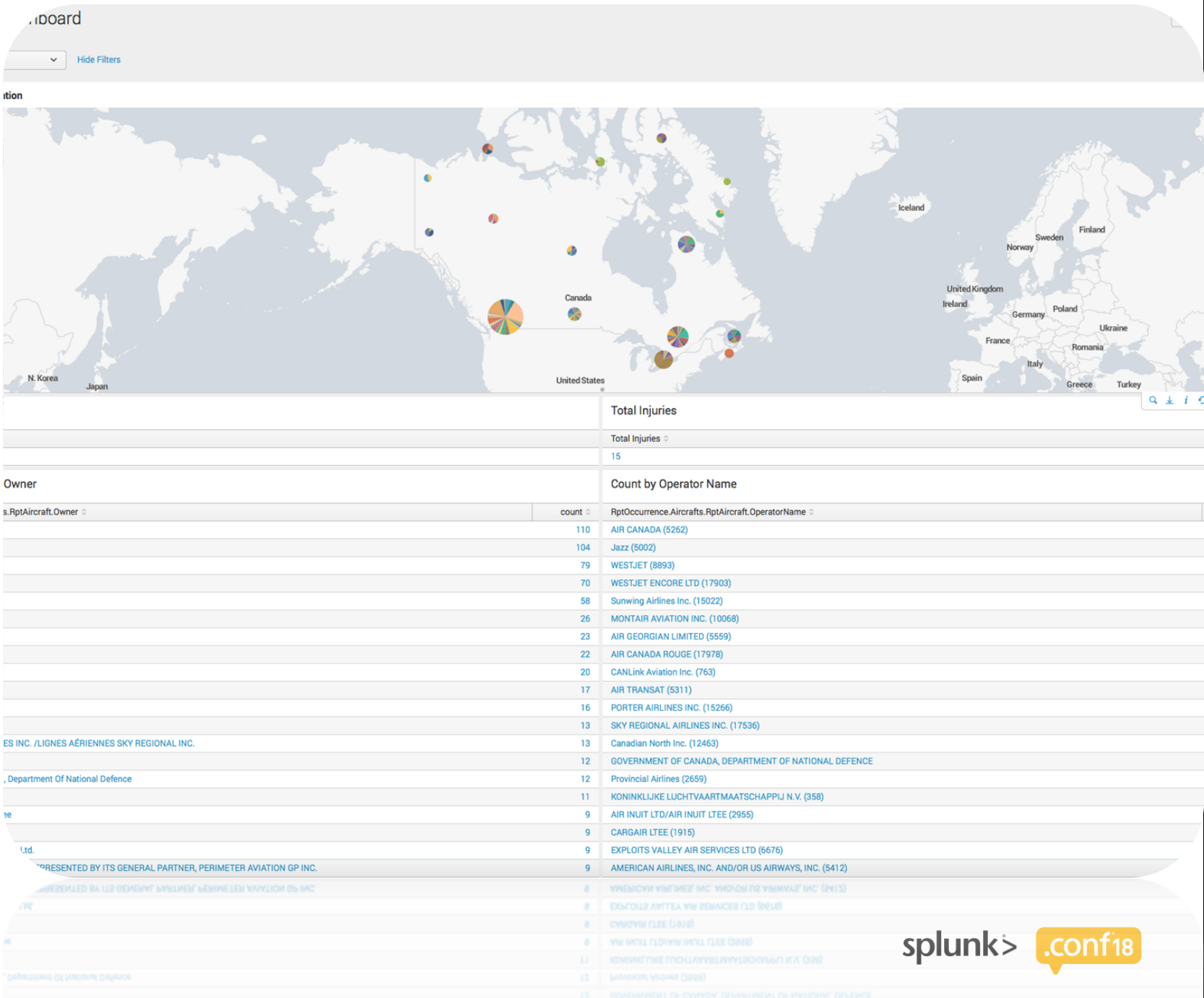


Stacked 100% bar chart

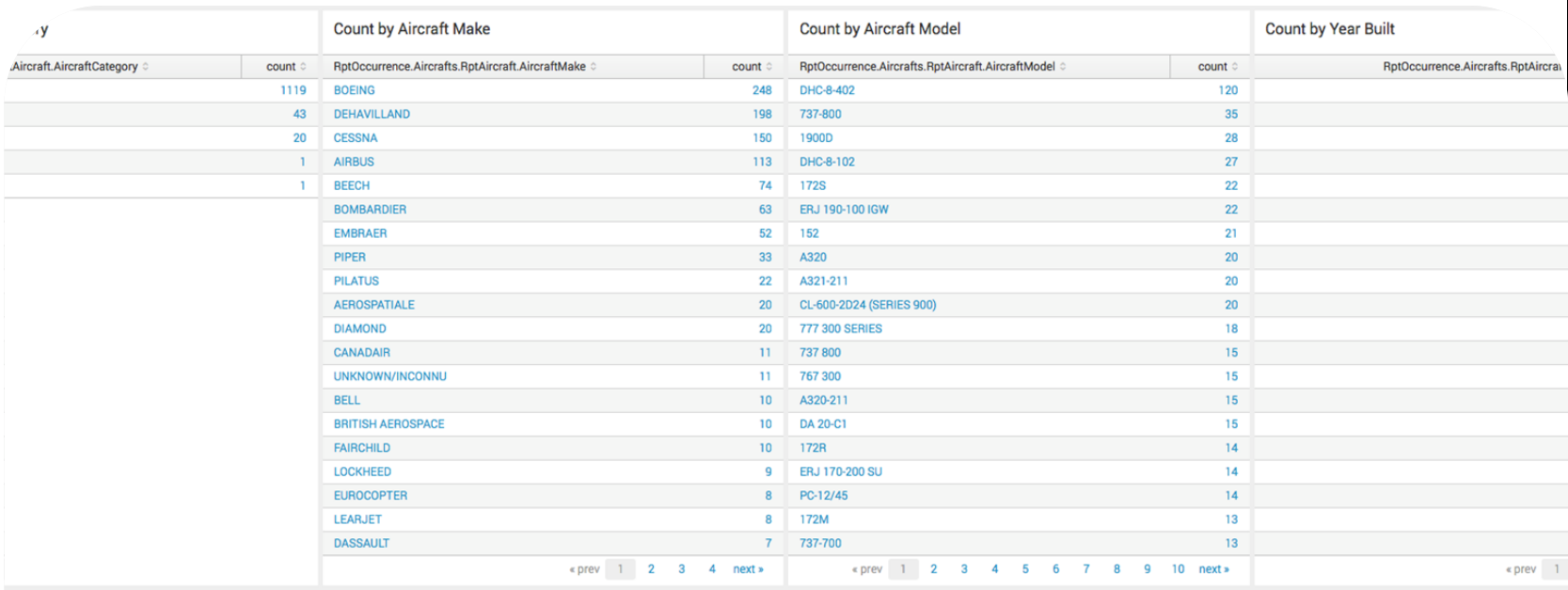
breaking down how each air carrier placed in each national region



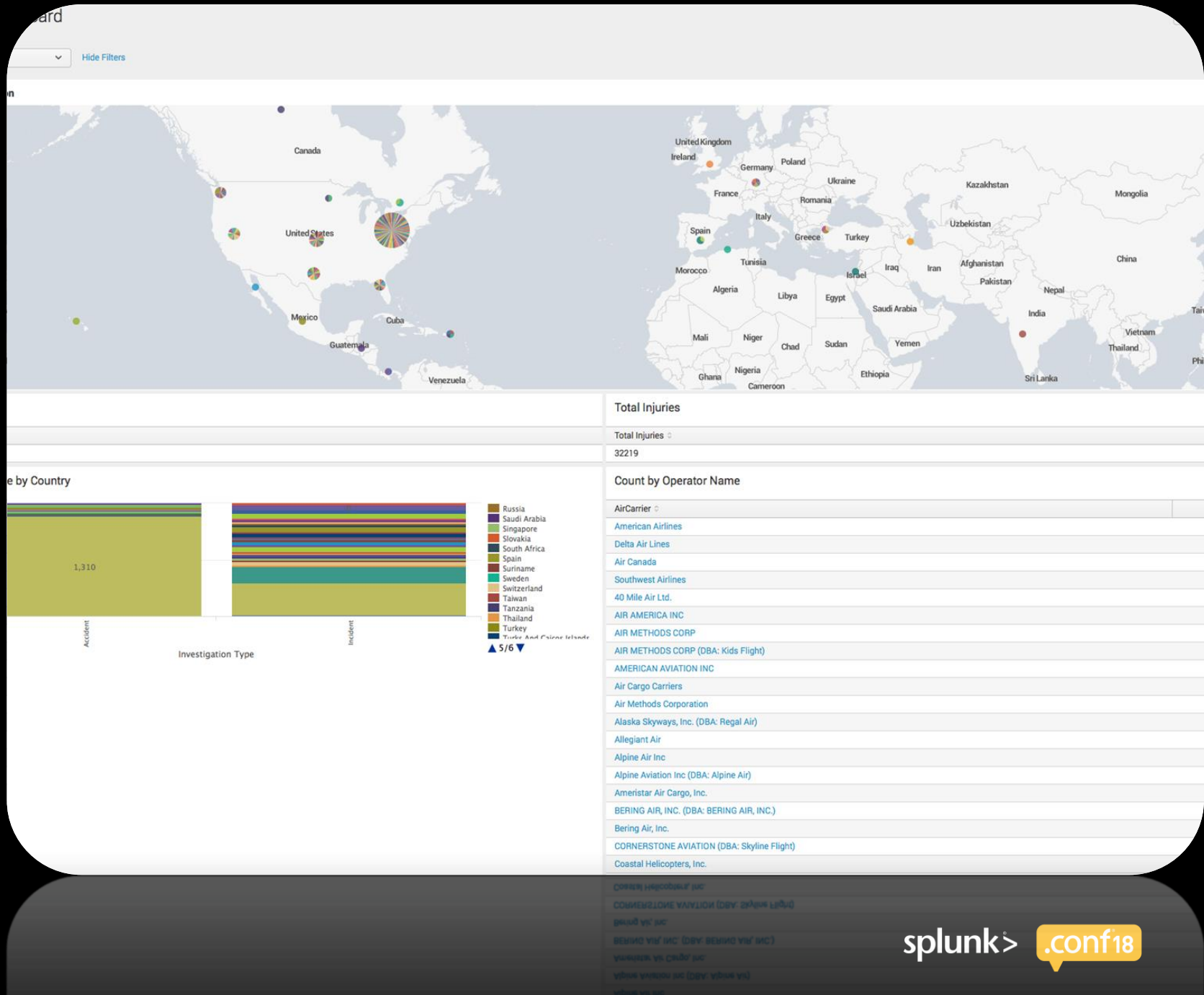
All CADORS Dashboard



All CADORS Dashboard



All NTSB Dashboard



The things that make it go

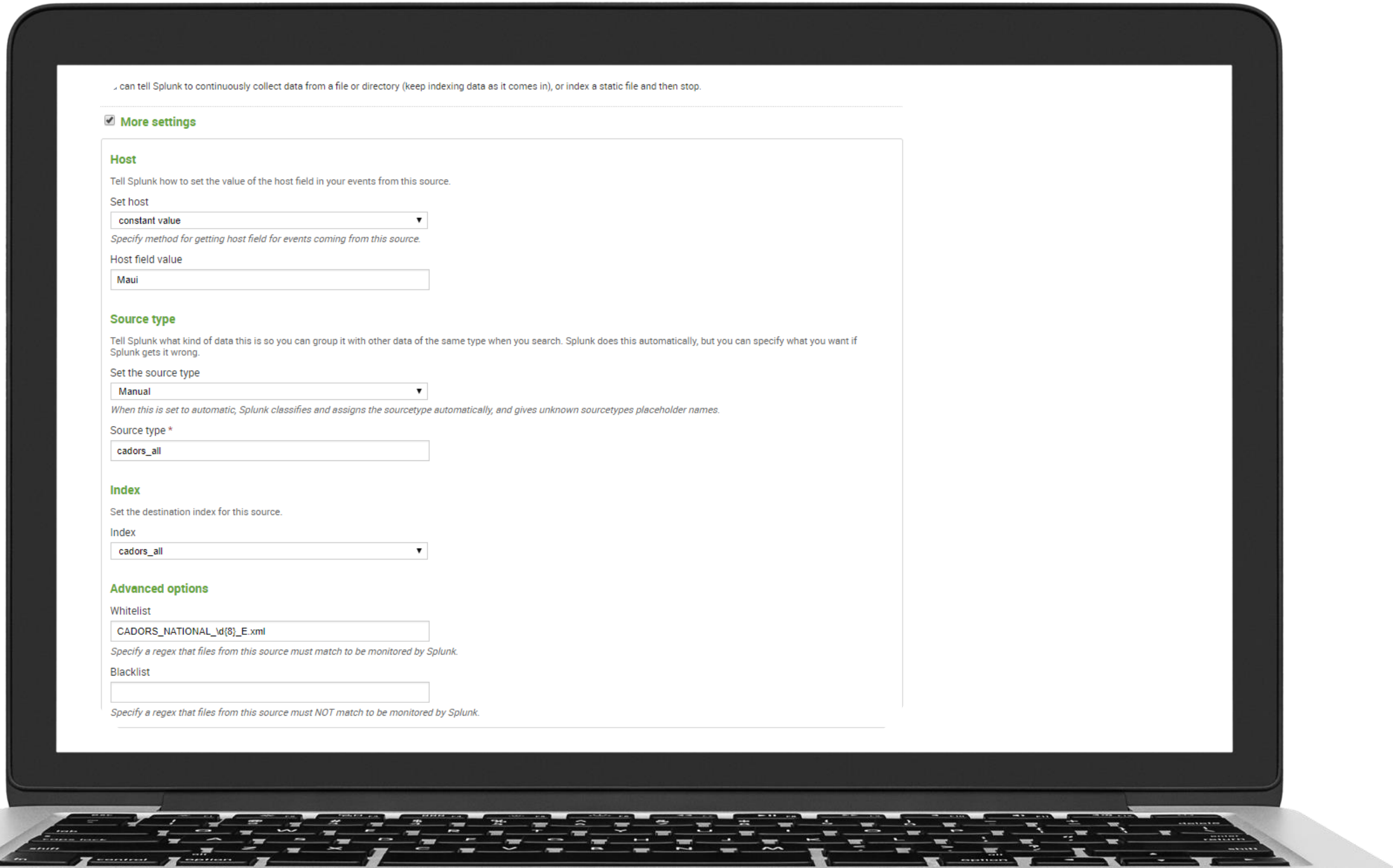
- Splunk app
- SPL: xmlkv, eval, rex, strptime, lookup, geostats, xpath
- inputs.conf to on-board data
- props.conf to parse XML
- Simple XML Dashboard
- File based lookup
  - OurAirports.com
- Event Types
- Workflow Actions

inputs.conf

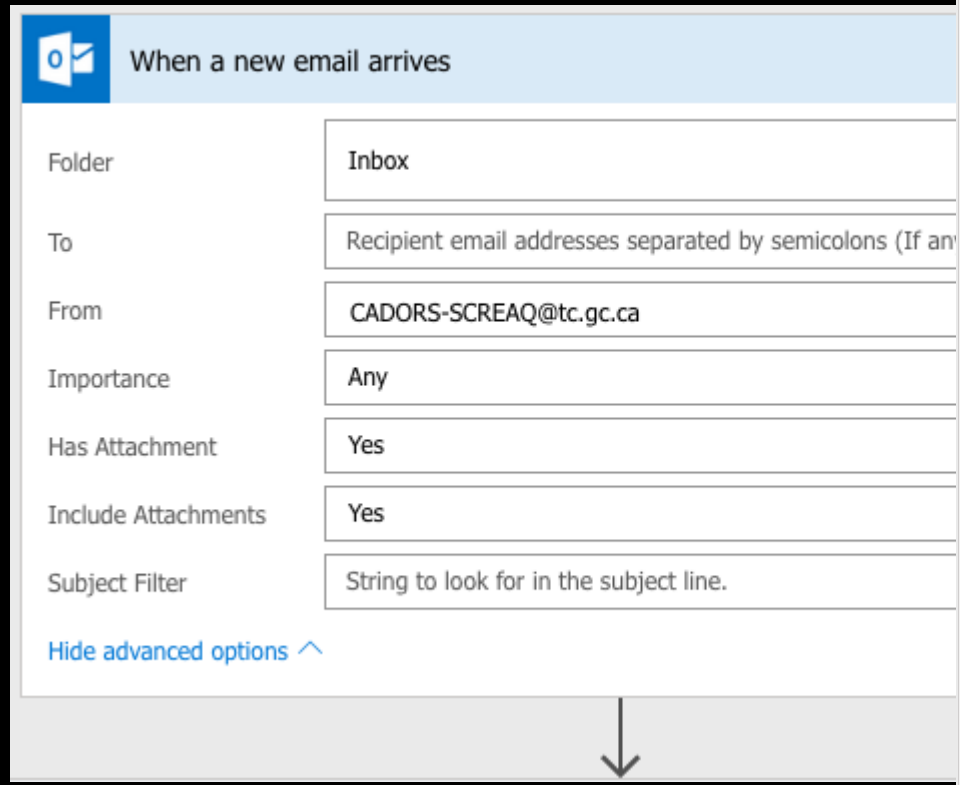
```
[monitor:///Users/cory/Desktop/CADORS]
disabled = 0
index = cadors
sourcetype = cadors
whitelist = CADORS_OPERATOR_REPORT_\d{8}_E.xml

[monitor:///Users/cory/Documents/OneDrive - Force 1024/Attachments/ALL_CADORS/]
disabled = 0
index = cadors_all
sourcetype = cadors_all
whitelist = CADORS_NATIONAL_\d{8}_E.xml
```

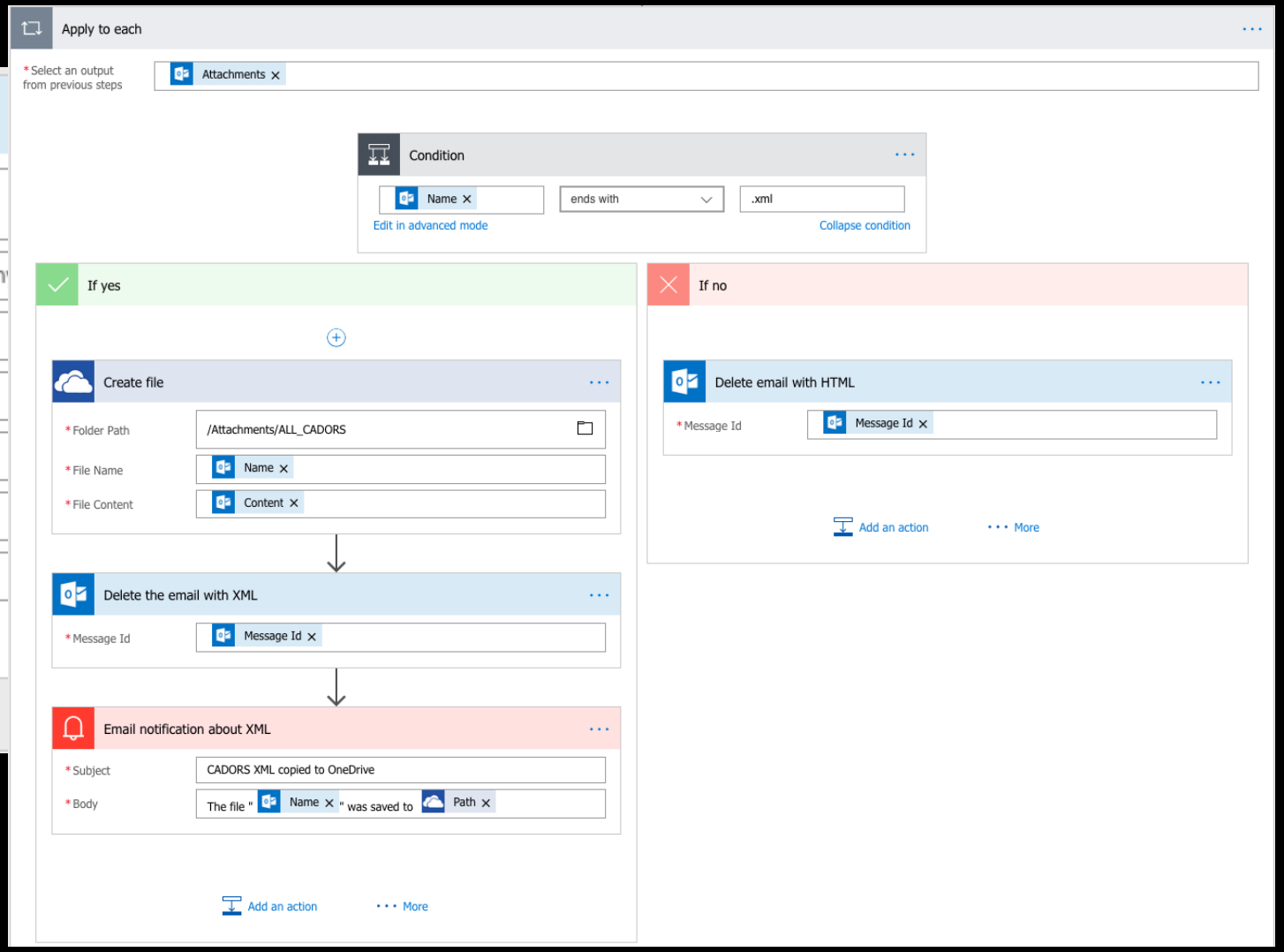
props.conf

```
[cadors]
TIME_PREFIX = <Date>(\d{4}-\d{1,2}-\d{1,2})
MAX_TIMESTAMP_LOOKAHEAD = 130
TIME_FORMAT = %Y-%m-%d
SHOULD_LINEMERGE = true
BREAK_ONLY_BEFORE = <RptOccurrence>
FIELD_HEADER_REGEX = instance\\>
KV_MODE = xml
NO_BINARY_CHECK = true
```



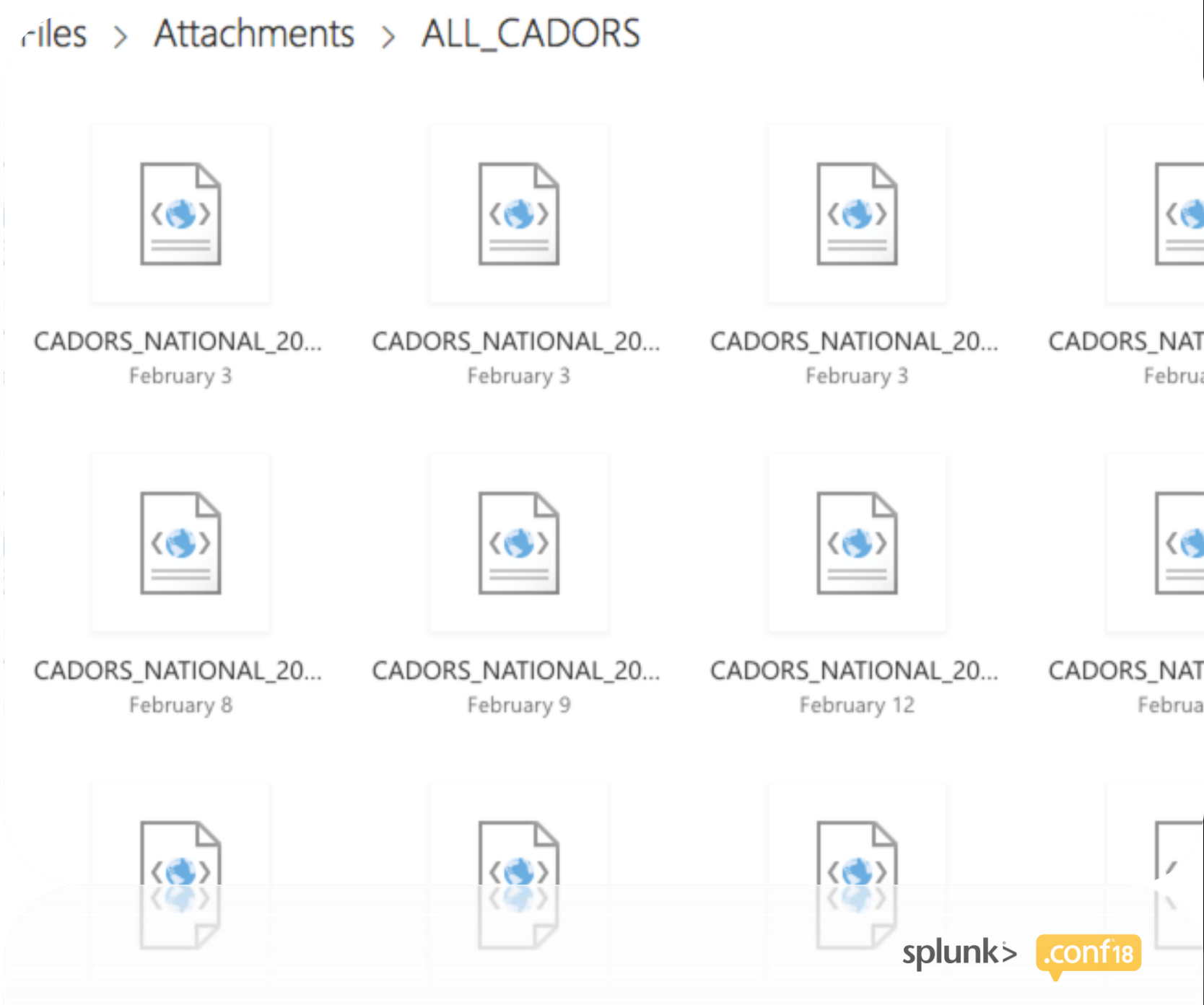
Automated event ingestion with Microsoft Flow





splunk> .conf18

Long term storage of data files on OneDrive



Data captured over-the-air via Raspberry Pi based receiver

About Over-the-Air Data Sources

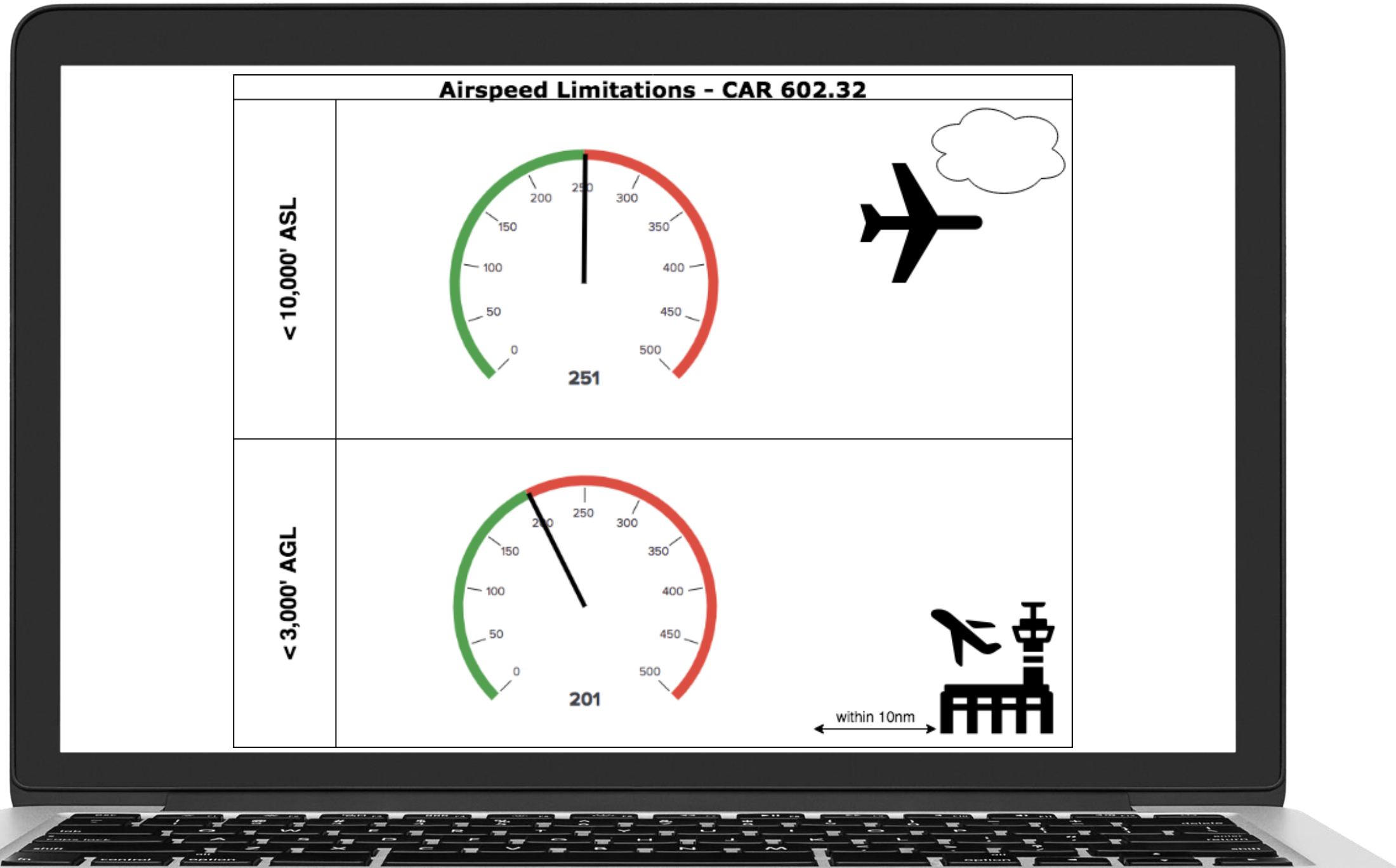
- The Stratux device is an Automatic Dependent Surveillance-Broadcast (ADS-B) receiver made from a Raspberry Pi.

- The Project
  - What? A weather, traffic, GPS receiver.
  - How? Radio antennae capturing over the air signals and logging to text file.
  - Where? http://stratux.me

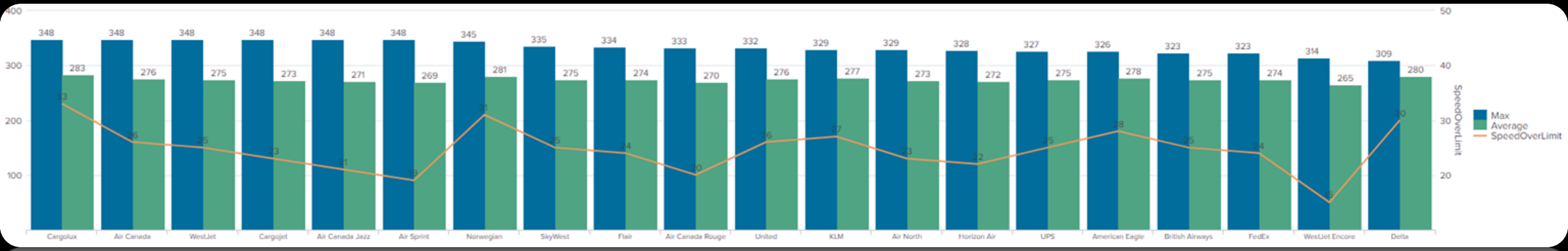
Speed Analysis

- The data set is 5 months worth of logs in the vicinity of Calgary International (CYYC). The data set has > 22 million events.
- I have accurate data recorded from aircraft in the vicinity which ranges from altitude and heading to aircraft identifier and speed.
- The question:

  Do commercial aircraft speed?
- For the first law, the bar chart shows the various carriers that are equipped with ADS-B. For the second law, we'll use a view by Google Earth.
- Future enhancements could easily include mashups with CADORS data or mashup with for-pay service data such as FlightAware.



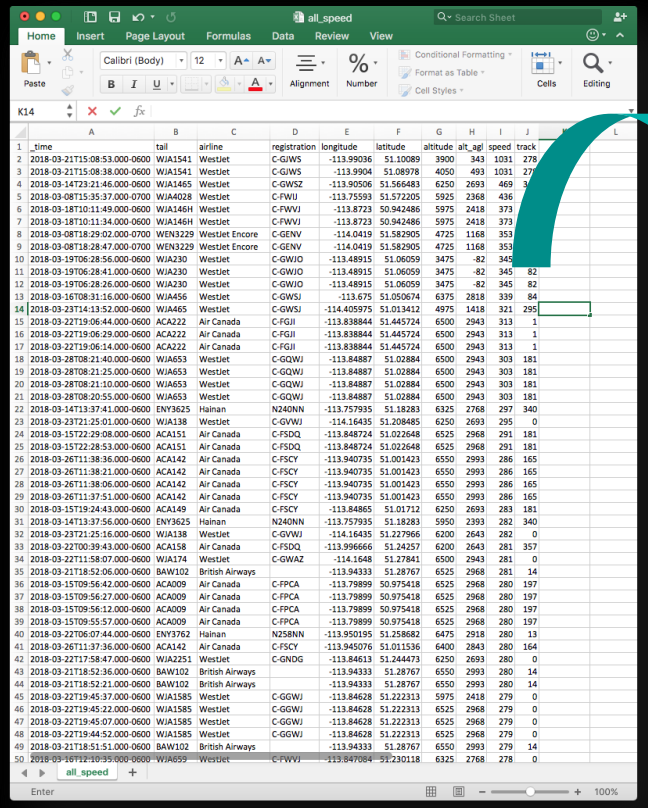
Bar chart with speed overlay

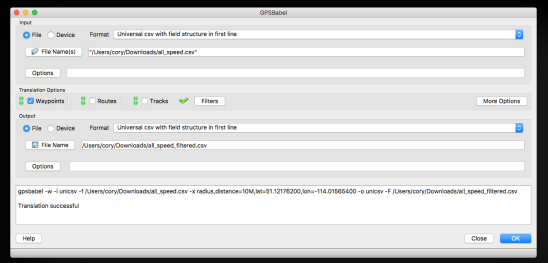


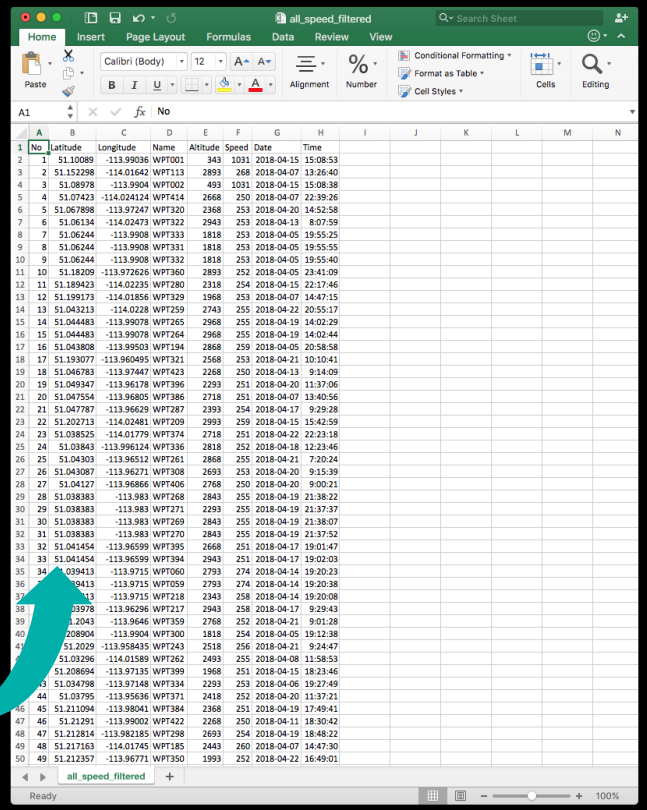
max recorded speed
avg recorded speed
delta between the speed limit and the average speed

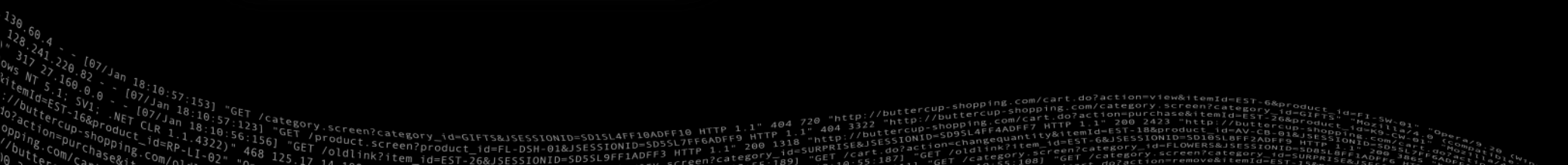
splunk> .conf18

Using GPSBabel to isolate events based on latitude and longitude

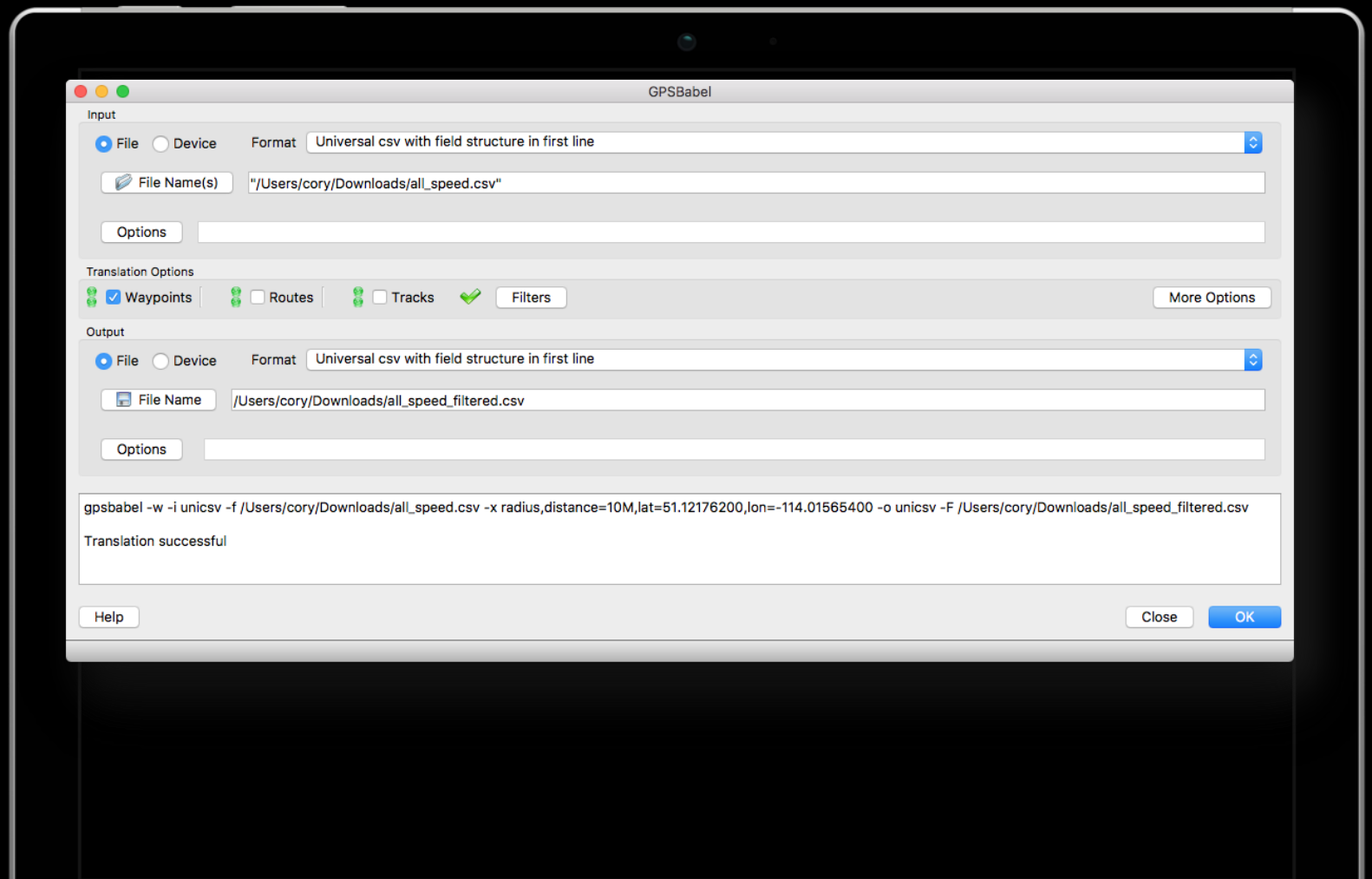








What is GPSBabel?



- “GPSBabel converts waypoints, tracks, and routes between popular GPS receivers such as Garmin or Magellan and mapping programs like Google Earth or Basecamp.”
- “It also has powerful manipulation tools for such data.”
- “It has been downloaded and used tens of millions of times since it was first created in 2001, so it's stable and trusted.”

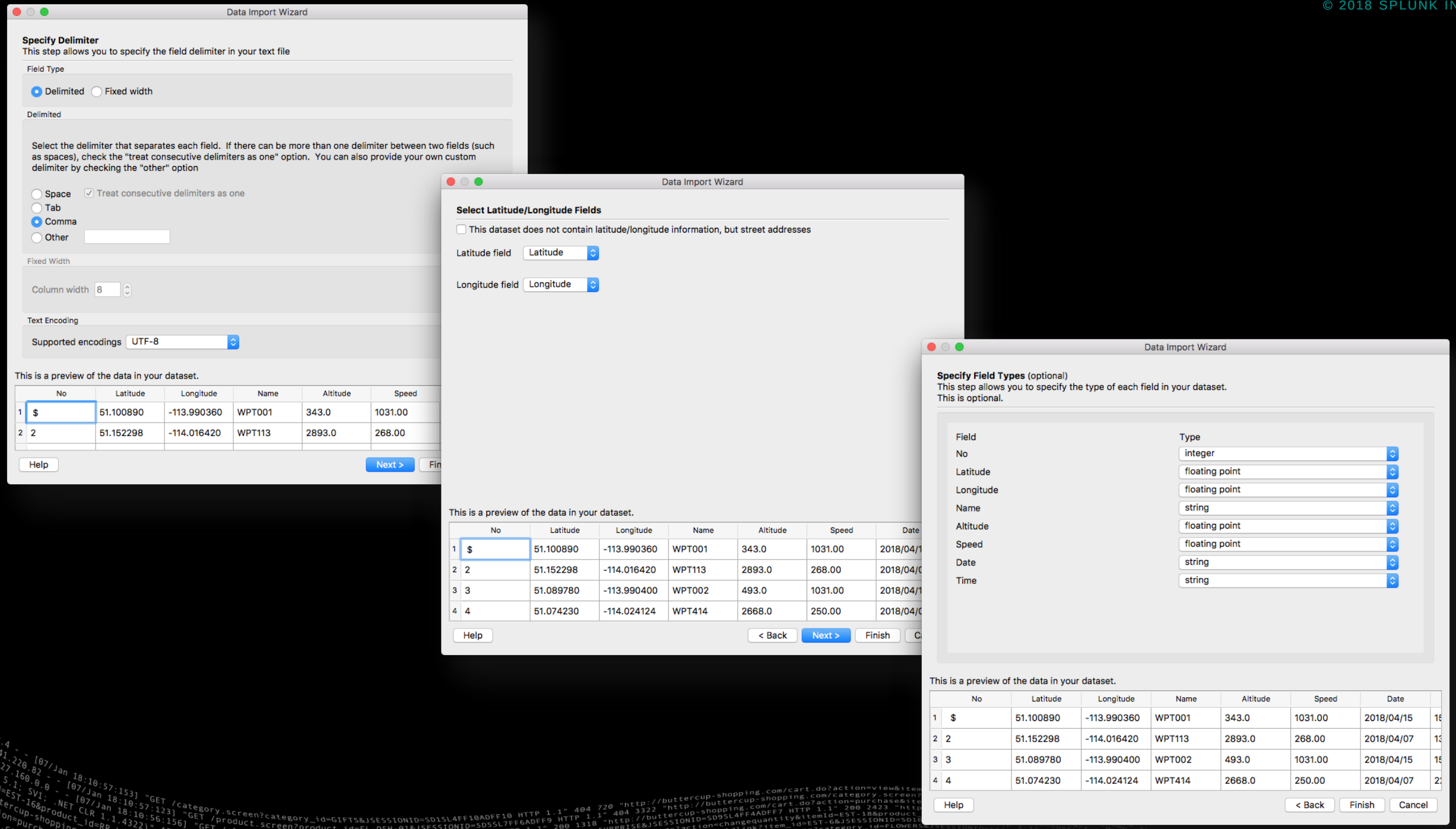
Integrating Google Earth

The 3-step data import process

Import from CSV

Select fields with coordinates

Cast to correct data type



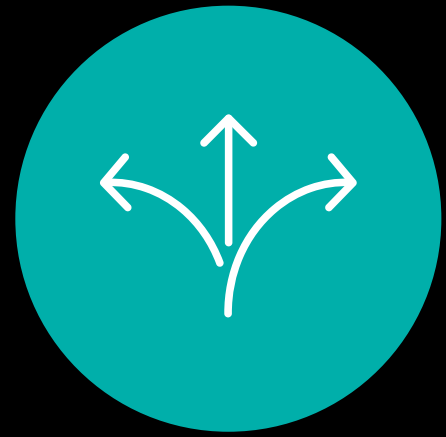
Integrating Google Earth

The 4-step style template creation

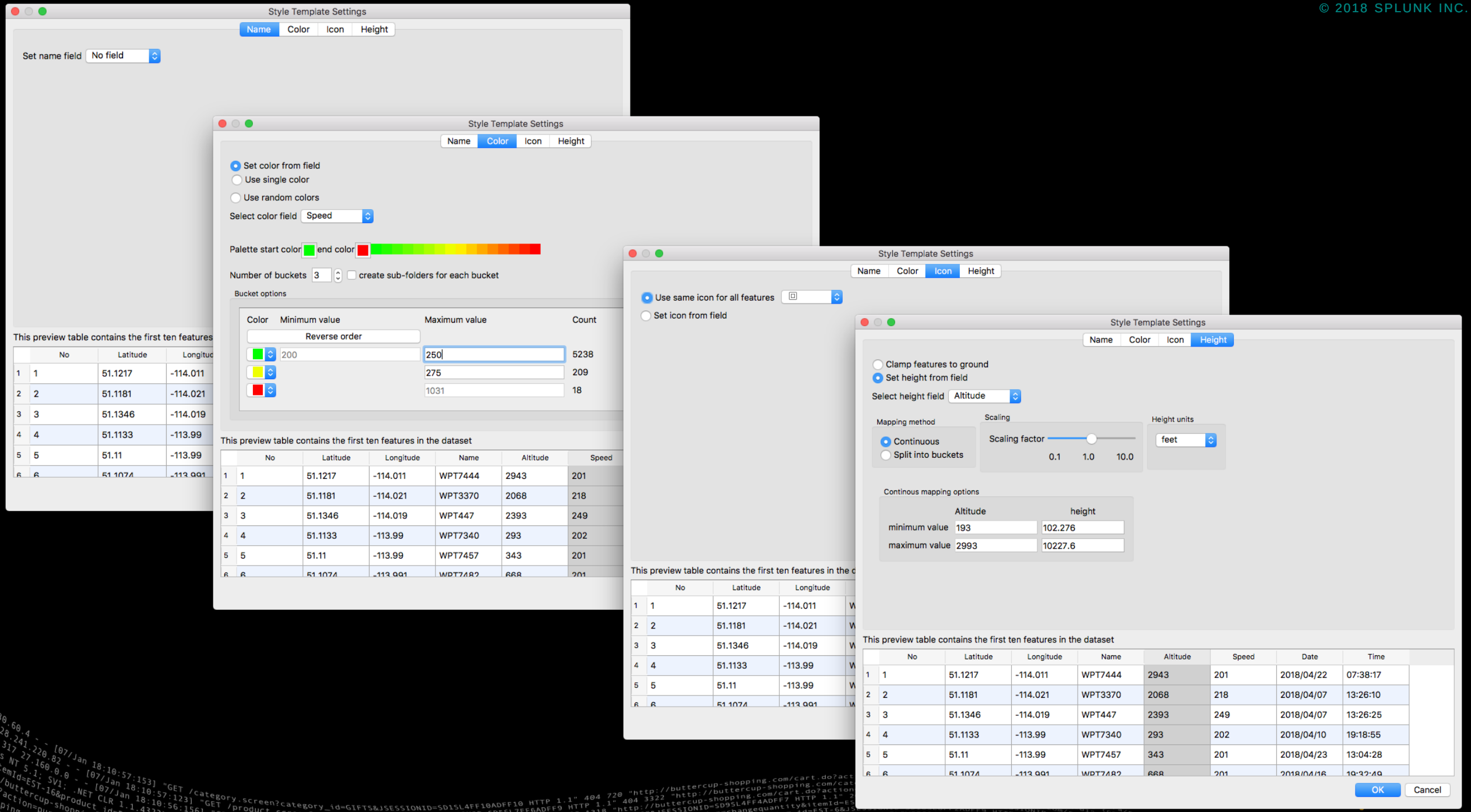
Name

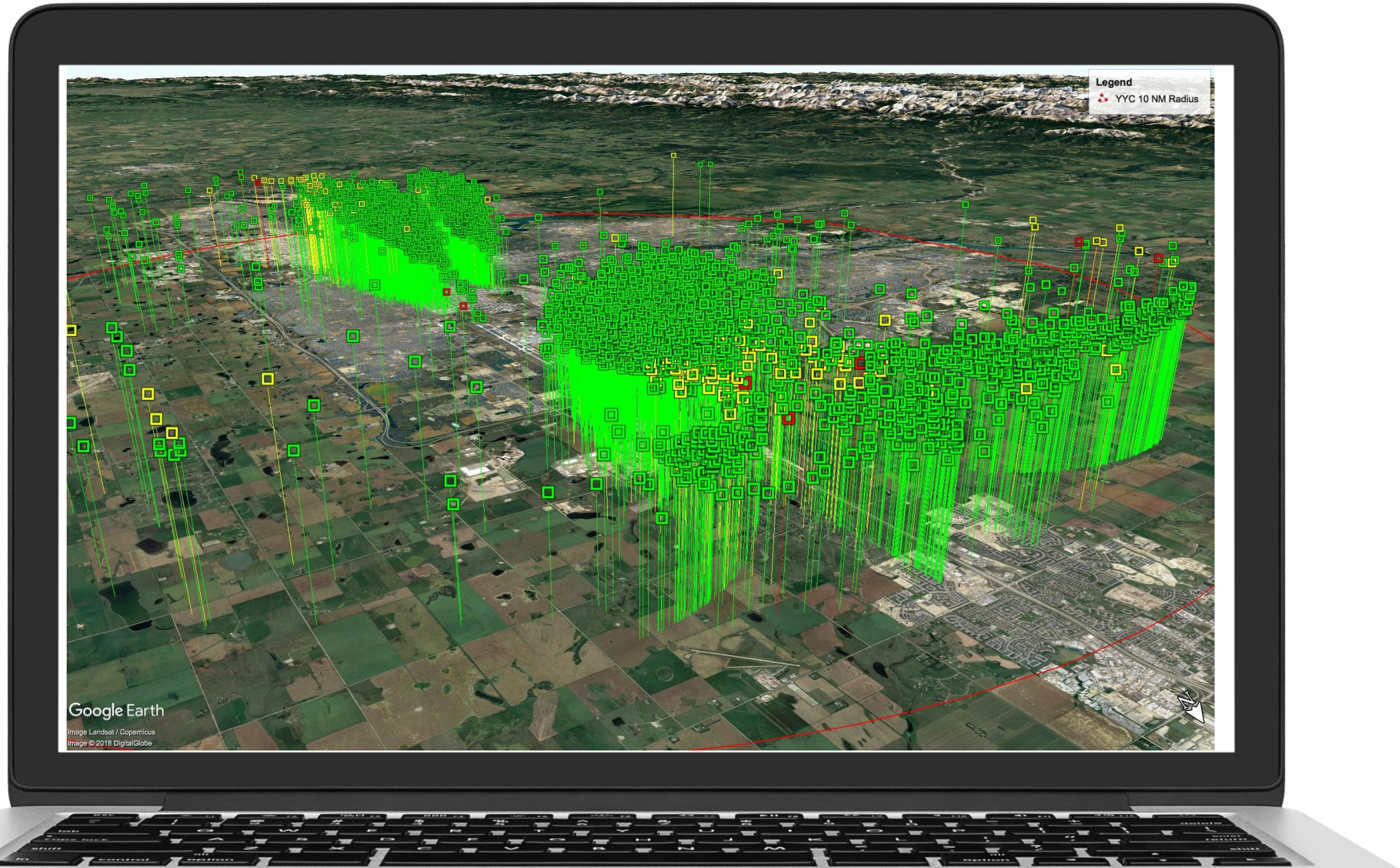
Color

Icon



Height





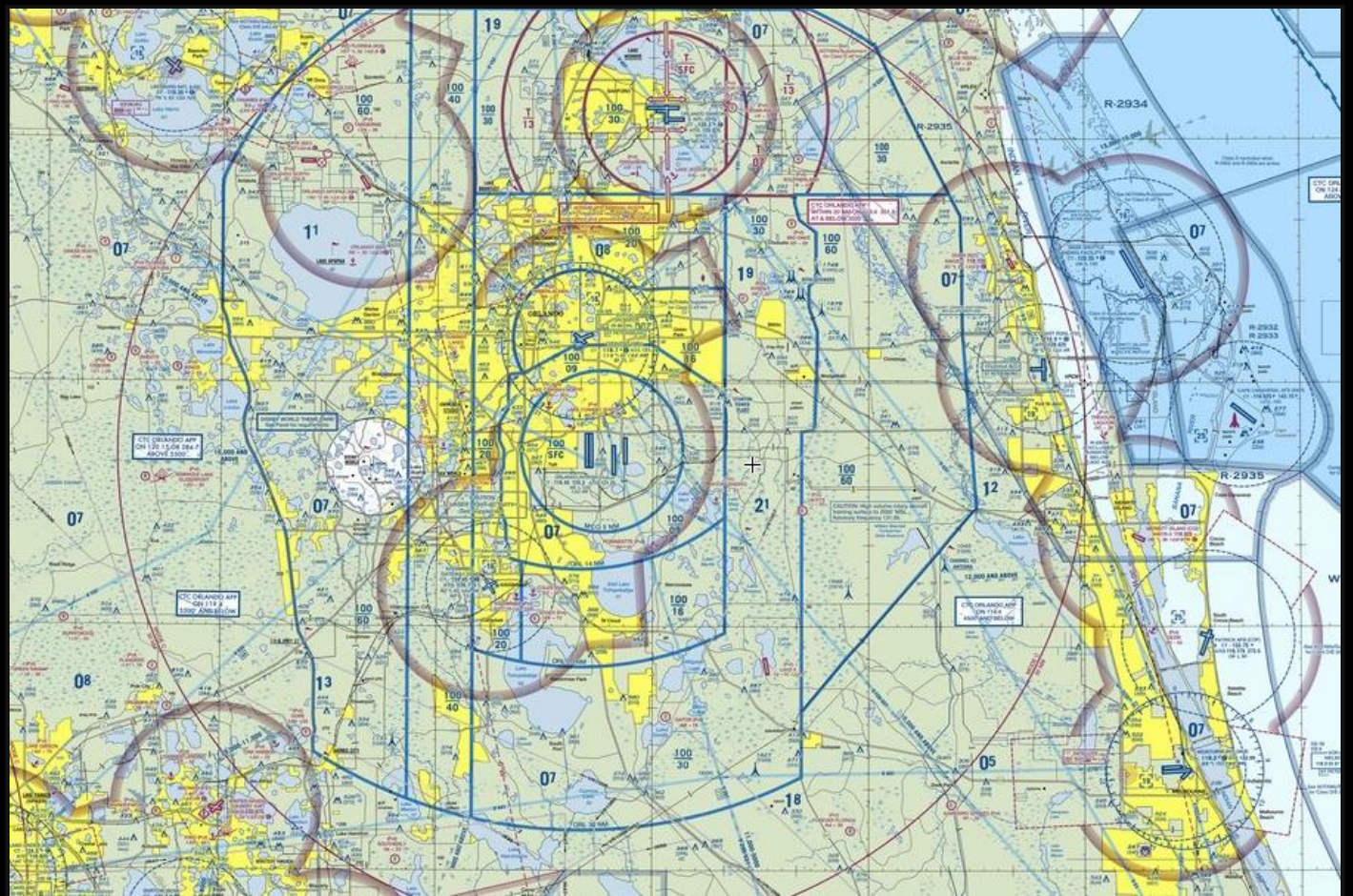
Stratux Demo

Take a peek at the Stratux App

splunk> .conf18

Orlando (MCO) Airspace

Terminal Area Chart



- There are 4 significantly sized airports and 8 smaller within 30 nm of MCO.
- Very busy airspace; Tampa (west), Miami (south), and Kennedy Space Center (east).

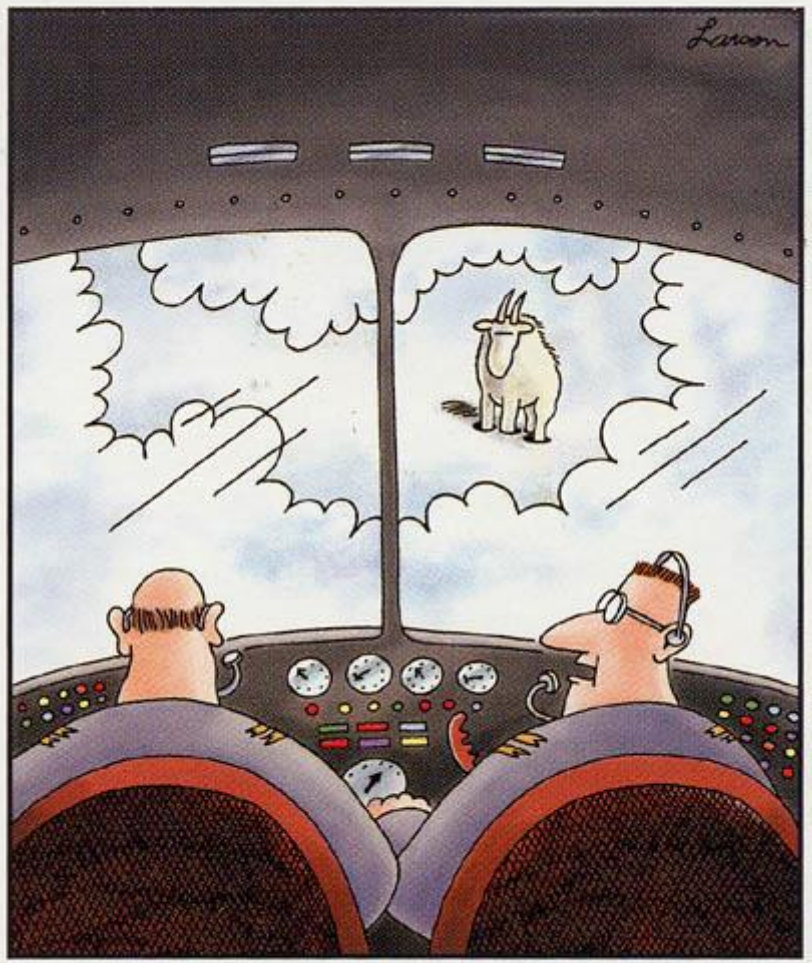
The things that make it go

- Splunk app
- SPL:
  - eventtype
  - eval with case
  - outlier
  - table
- props.conf to parse fields from JSON
- Simple XML Dashboard
- Google Earth
- GPSBabel

props.conf ×

```
[Stratux]
DATETIME_CONFIG =
NO_BINARY_CHECK = true
category = Custom
description = ADS-B Receiver Logs
disabled = false
pulldown_type = true
EXTRACT-icao_addr = Icao_addr\":(?<icao_addr>\d{1,})
EXTRACT-registration = Reg\":\"(?<registration>[\w-]{1,})
EXTRACT-emitter = Emitter_category\":(?<emitter>\d{1,})
EXTRACT-onground = OnGround\":(?<onground>\w*)
EXTRACT-tail = Tail\":\"(?<tail>[\w\s-]{1,})
EXTRACT-signal = SignalLevel\":(?<signal>[-\d.\d]*)
EXTRACT-squawk = Squawk\":(?<squawk>\d*)
EXTRACT-targettype = TargetType\":(?<targettype>\d*)
EXTRACT-positionvalid = Position_valid\":(?<positionvalid>\w*)
EXTRACT-altitude = \"Alt\":(?<altitude>\d*)
EXTRACT-latitude = Lat\":(?<latitude>[-\d.\d]*)
EXTRACT-longitude = Lng\":(?<longitude>[-\d.\d]*)
EXTRACT-gnssdifffrombaroalt = GnssDiffFromBaroAlt\":(?<gnssdifffrombaroalt>\d*)
EXTRACT-altisgnss = AltIsGNSS\":(?<altisgnss>\w*)
EXTRACT-speed = Speed\":(?<speed>\d*)
EXTRACT-track = Track\":(?<track>\d*)
EXTRACT-speedvalid = Speed_valid\":(?<speedvalid>\w*)
EXTRACT-vvel = Vvel\":(?<vvel>[\d-]*)
EXTRACT-prioritystatus = PriorityStatus\":(?<prioritystatus>\d*)
EXTRACT-lastgnssdiffalt = Last_GnssDiffAlt\":(?<lastgnssdiffalt>\d*)
EXTRACT-age = Age\":(?<age>[\d.]*)
EXTRACT-agelastalt = AgeLastAlt\":(?<agelastalt>[\d.]*)
EXTRACT-bearing = Bearing\":(?<bearing>\d*)
EXTRACT-distance = Distance\":(?<distance>\d*)
EXTRACT-extrapolatedposition = ExtrapolatedPosition\":(?<extrapolatedposition>\w*)
```

splunk> .conf18



"Say … what's a mountain goat doing way up here in a cloud bank?"

Takeaways

1. *All CADORS App*

   http://github.com/csyvenky/all_cadors

2. All NTSB App

   https://github.com/csyvenky/all_ntsb

3. Stratux App

   http://github.com/csyvenky/stratux

My Contact Info and Links

- Stratux
  - https://github.com/cyoung/stratux
  - http://stratux.me
- Splunk Addons
  - https://splunkbase.splunk.com
- GPSBabel
  - https://www.gpsbabel.org
- Visual Studio Code
  - https://code.visualstudio.com
- Microsoft Flow
  - https://flow.microsoft.com

@spsavvy

- CADORS
  - http://wwwapps.tc.gc.ca/saf-sec-sur/2/cadors-screaq/m.aspx
- National Transport Safety Board Data
  - https://www.ntsb.gov/_layouts/ntsb.aviation/Index.aspx
- Federal Aviation Administration Data
  - https://www.faa.gov/data_research/aviation_data_statistics

Thank You

Don't forget to rate this session in the .conf18 mobile app