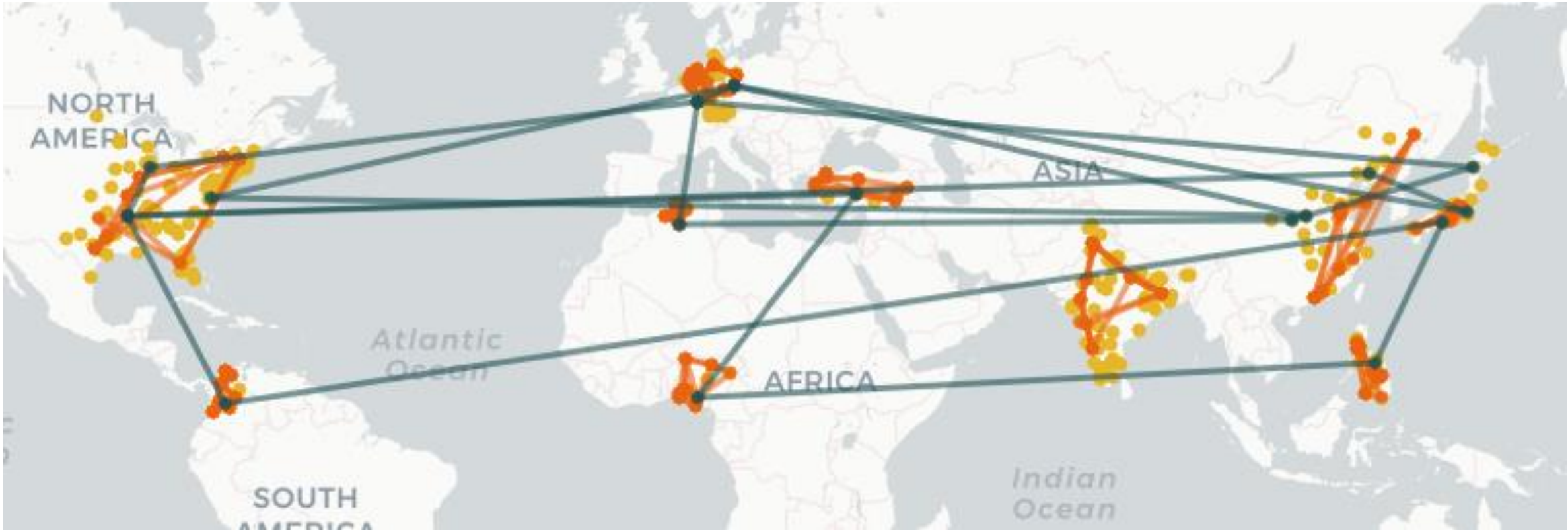
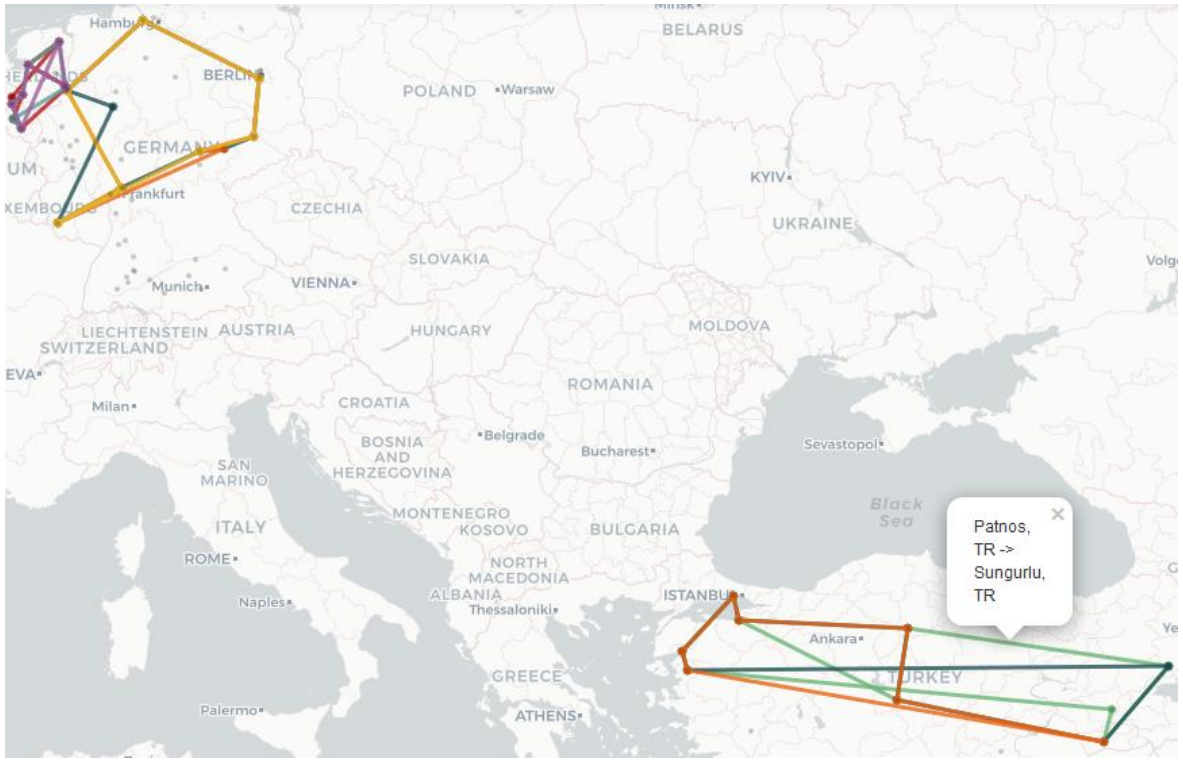
Network Analysis



How many drones are deployed in regional services in Europe?

- Approach
  - check number of drones per service
  - check number of regional services in Europe

- Nine drones are deployed in nine regional services in Europe.



How many station pairs are directly connected in the network?

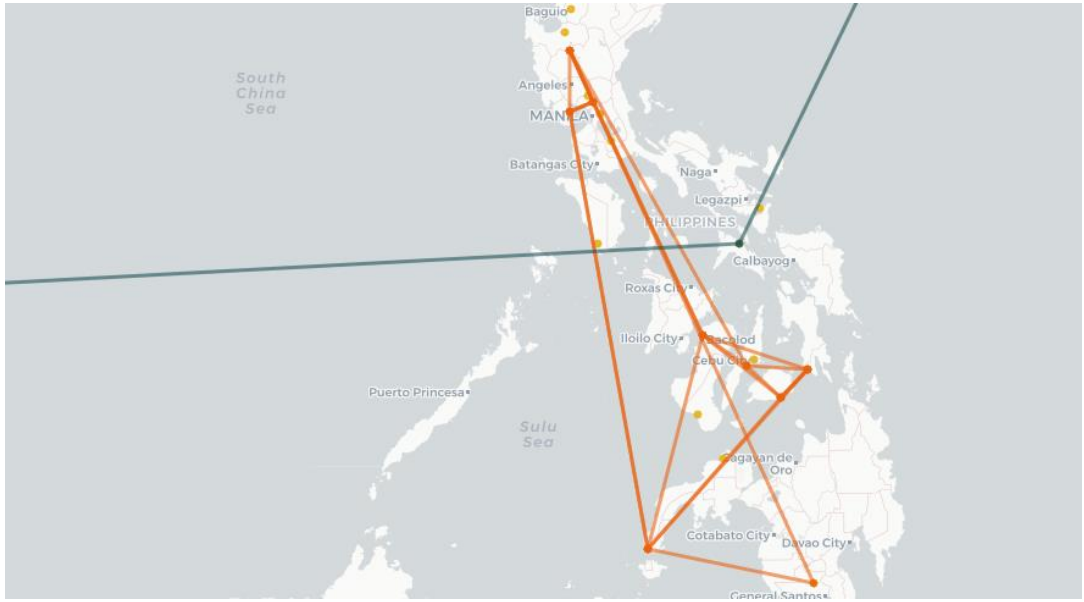
- Approach (by service connected stations)
  - find unique combinations of all stations in service (first) and overall (second)
- Approach (directly connected stations)
  - change directed to undirected network
  - count edges
- Assumption
  - direction of connection is not relevant

- 418 stations are connected without changing services.
- 171 stations are directly connected without changing or waiting at a different station.

Which countries cannot be reached from Germany?

- Approach
  - determine stations and respective countries that have a path from Germany
  - subtract all possible countries

- India and the Netherlands cannot be reached from Germany.
  - The services in the Netherlands and in India are isolated.
  - Other countries with isolated services (China, Philippines) still have connected stations.



What's the fastest route from Utica in the US to Boumerdas in Algeria in the current network?

- Approach
  - weight edges with flight time
  - calculate shortest weighted (=fastest) path
- Assumption
  - 'fast' in respect of flight time only
  - waiting or transfer time is not included
- alternative approaches regarding travel time are attached in the Notebook

Utica → Keene → Lutz → Searcy → Haicheng → Shimotoda → Oschatz → Losheim → Messaad → Thenia → Boumerdas

- 1 day and 46 minutes

What's the average time it takes to get from a random station to another random station?

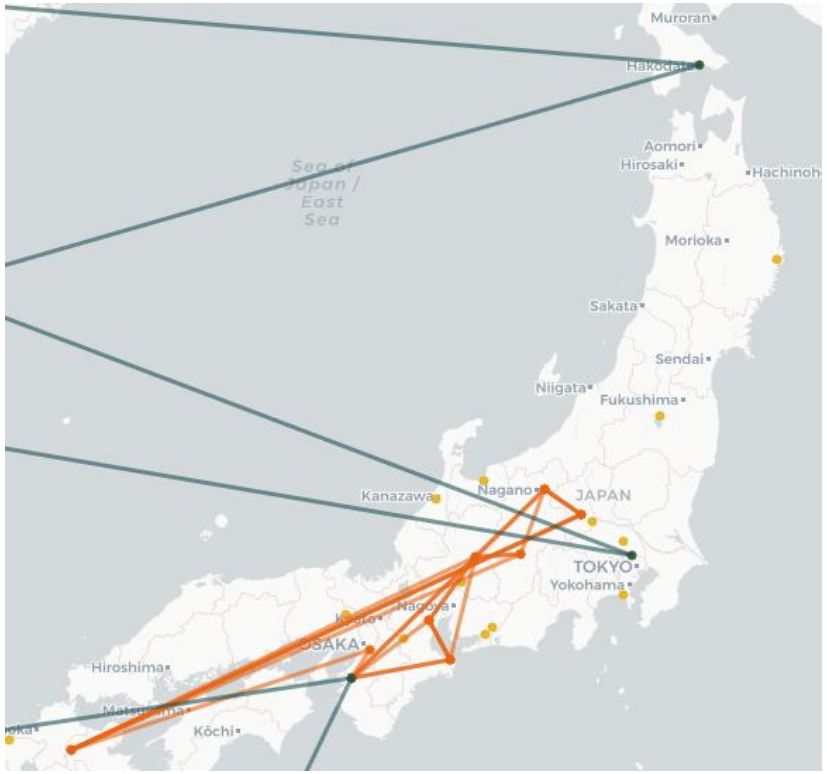
- Approach
  - compute fastest paths for all connections like in 1.4
  - average paths
- Assumption
  - in respect to flight time (not travel time)

- On average flying from a random station to another connected random station takes 19 hours and 40 minutes.

How would you improve this average time if you could set up an additional service?

- Approach
  - compare distance and flight time between two stations
  - connect stations with a low km/hour
- Assumption
  - new connections, not additional scheduled service
  - only already (indirectly) connected stations can be part of the service

- The largest reduction in average flight time would be a service that links Shimotoda in Japan to the Japanese regional network. The same, but to a smaller extent applies for Hakodate in northern Japan.



How would you improve this average time if you could set up an additional service?

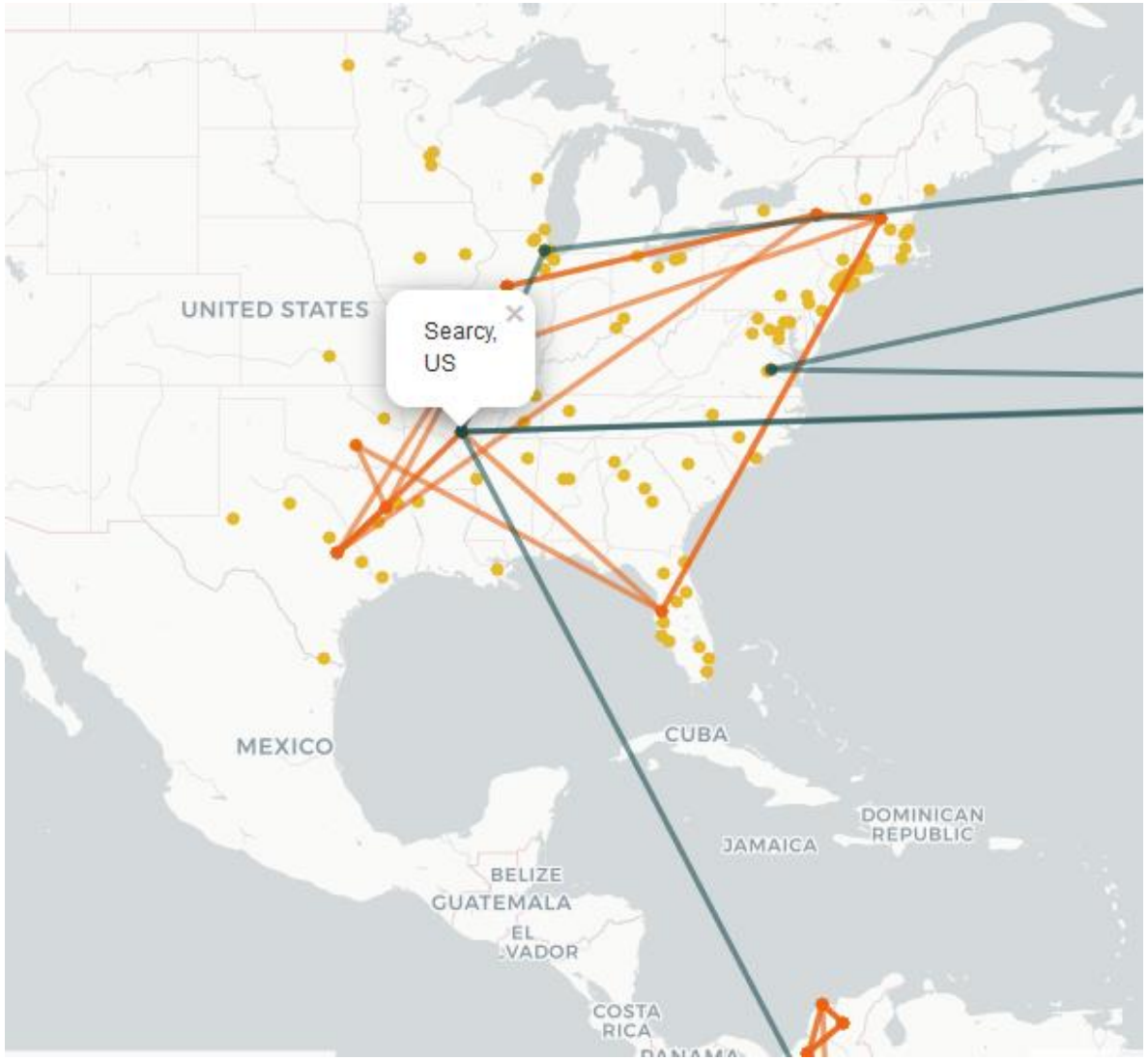
- Fastest connection from Shimotoda to Annaka in current network

Which station in your network would you consider the most important one?

- Approach
  - compute centrality algorithms
- Assumption
  - regarding flight time instead of travel time
  - direct connections don‘t include a stop at another station
  - ‘short‘ refers to the number of changes at stations

- Searcy (US) is the most important station in the network.
  - highest number of directly connected stations
  - highest probability that it is part of the shortest and fastest trips between two random stations
  - shortest and fastest connection to all other stations
- Losheim has outgoing direct connections.
- Messaad has more flights.

Which station in your network would you consider the most important one?



If you could set up the network from scratch with the existing drones, what would you change?

- Trade-off
  - efficiency?
    - fastest regional connections
    - fastest global connections
    - centralized network
  - capacity?
    - decentralized network
  - reliability?
    - decentralized network

- include isolated networks
- connect all stations in a country to regional network
- determine 1 to 2 regional hub
  - connect them to global network
  - assumes unlimited capacity
  - but reduces changes at stations and time

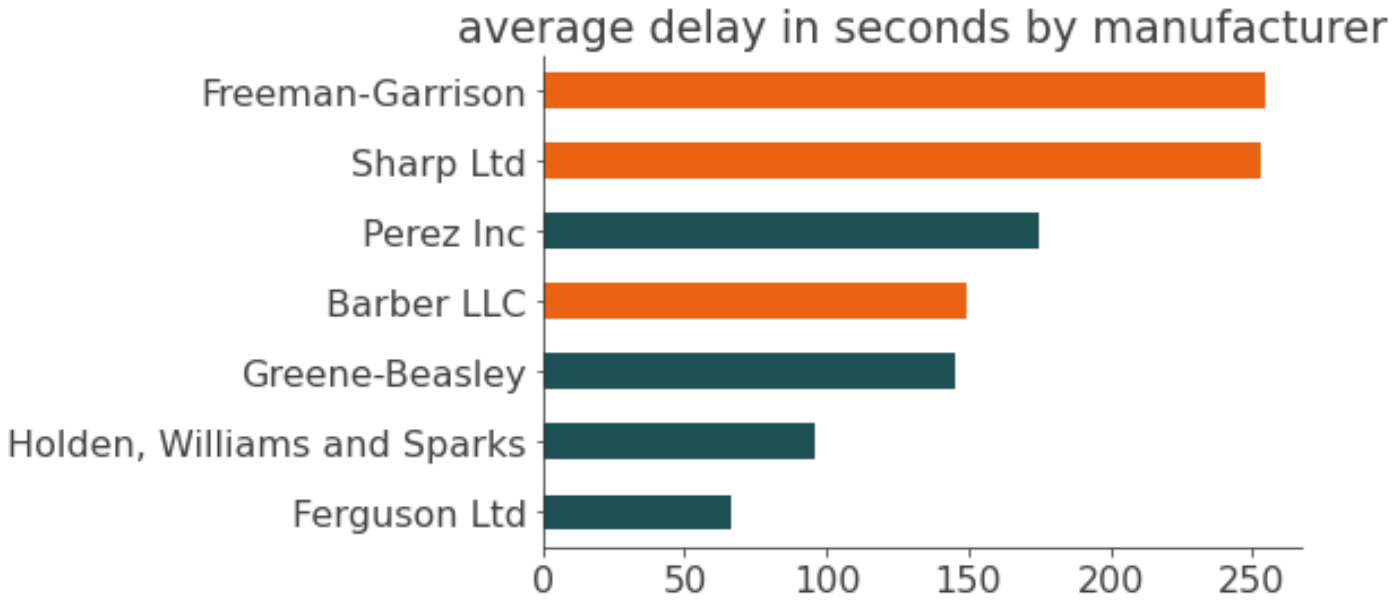
Predicting on-time Arrivals

2.1

Before starting the analysis, you ask your colleagues about their opinion on what influences the on-time arrival of a service. Which statement is correct?

- *"From my former experience as a pilot I can say that drones from 'Barber LLC', 'Sharp Ltd', and 'Freeman-Garrison' are very difficult to get to the destination on time."*

- This is correct for Freeman-Garrison and Sharp Ltd., but not for Barber LLC.
  - controlled for other factors it is also correct for Barber LLC. (see question 2.2)



Before starting the analysis, you ask your colleagues about their opinion on what influences the on-time arrival of a service. Which statement is correct?

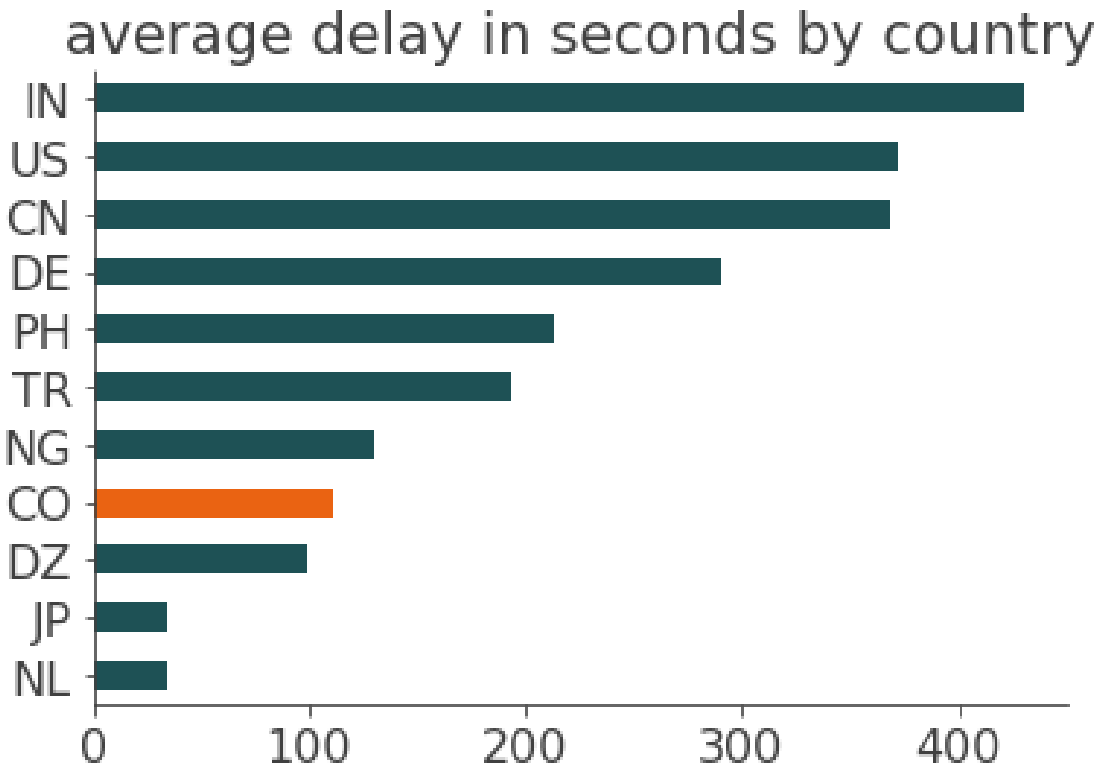
- *"The technology of drones is already so sophisticated that wind has no impact on the arrival time."*

- This is almost correct, there exists a positive but low correlation.
  - controlled for other factors, the statement is not correct (see question 2.2)

Before starting the analysis, you ask your colleagues about their opinion on what influences the on-time arrival of a service. Which statement is correct?

- *"Services in Colombia are almost never on time because of the difficult climate close to the equator."*

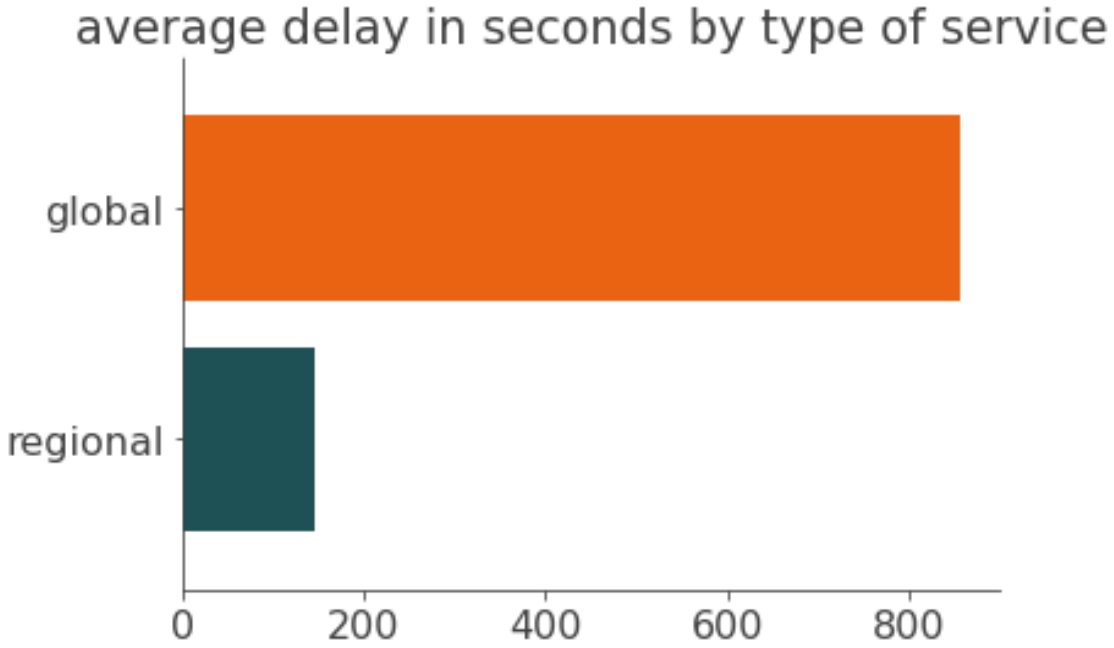
- This is not correct. Services in Colombia are almost always on time.



Before starting the analysis, you ask your colleagues about their opinion on what influences the on-time arrival of a service. Which statement is correct?

- *"Customers pay much more for the global services. So I assume that these services have less delays."*

- This is not correct. On average delays in global services are larger and more frequent.
  - mean and median
  - percentage of delayed flights



Before starting the analysis, you ask your colleagues about their opinion on what influences the on-time arrival of a service. Which statement is correct?

- *"The heavier the drone, the more expensive it is to do a speed up in order to catch up delays. Therefore the bigger drones will have more delays."*

- There are no delayed departure, therefore it is not necessary to catch up a delay from the previous connection.
- But, if delayed during the flight, it is true. The more passengers a drone can carry the more likely that the flight is delayed (positive correlation).

What is the biggest driver of the arrival delay?

- Approach
  - find possible drivers
    - correlated with delays
    - not correlated with each other
  - scale variables
  - execute linear regression
- Assumption
  - observations are independent of each other (no delayed departures)
  - linear relationship

TOP 3

- travel distance in km
  - somehow intuitive, the longer the distance, the longer the travel time and the larger the delay
  - alternatively arrival delay in seconds per km as the dependent variable
- passenger capacity
  - indicates drone size
  - no information about actual number of passengers
- wind speed in knt.

If your only goal is to bring your passengers to the destination on time, which drones would you replace with new ones?

- Approach
  - define delays (> 15 min according to FAA standard)
  - scale variables
  - execute logistic regression
  - construct dataset with drone characteristics and average values for flight attributes (e.g. wind speed)
  - predict delay probability
- Assumption
  - like 2.3

- The drone operating the service World_2, has by far the highest probability of a delay irrespective of wind speed, travel distance and latitude, followed by the drones of World_0 and In_0.

Can you predict the delays of all station arrivals in the last days of July?

- Approach
  - predict delay in sec. → regression
  - start model: Random Forest
  - metric: R-squared
  - calculate model and metric
  - improve model with respect to metric
- alternative approach
  - use time series
  - if observations are independent time series are not necessary

- Yes.
- With Random Forrest the R-squared is 0.98.
  - Actually this is too his, but usual for this data type. The goal is not to predict changes in delays or influencing factors.
- No further improvement.

Which service will have the least delays in the last days of July?

- Approach
  - drop August schedule
  - categorize if predicted delay is larger than 15 min
  - sum by service
- alternative approach
  - predict if service is delayed (0,1)
    - classification
    - Random Forrest Classifier
    - Support Vector Machine
    - …

- TR_1 has only 3 predicted arrivals that are more than 15 minutes delayed.