Stats NZ
Tatauranga Aotearoa

Migration based on the observed behaviour of travellers

New Zealand's approach to administrative data driven migration estimates






Conditions conducive to measuring outcomes

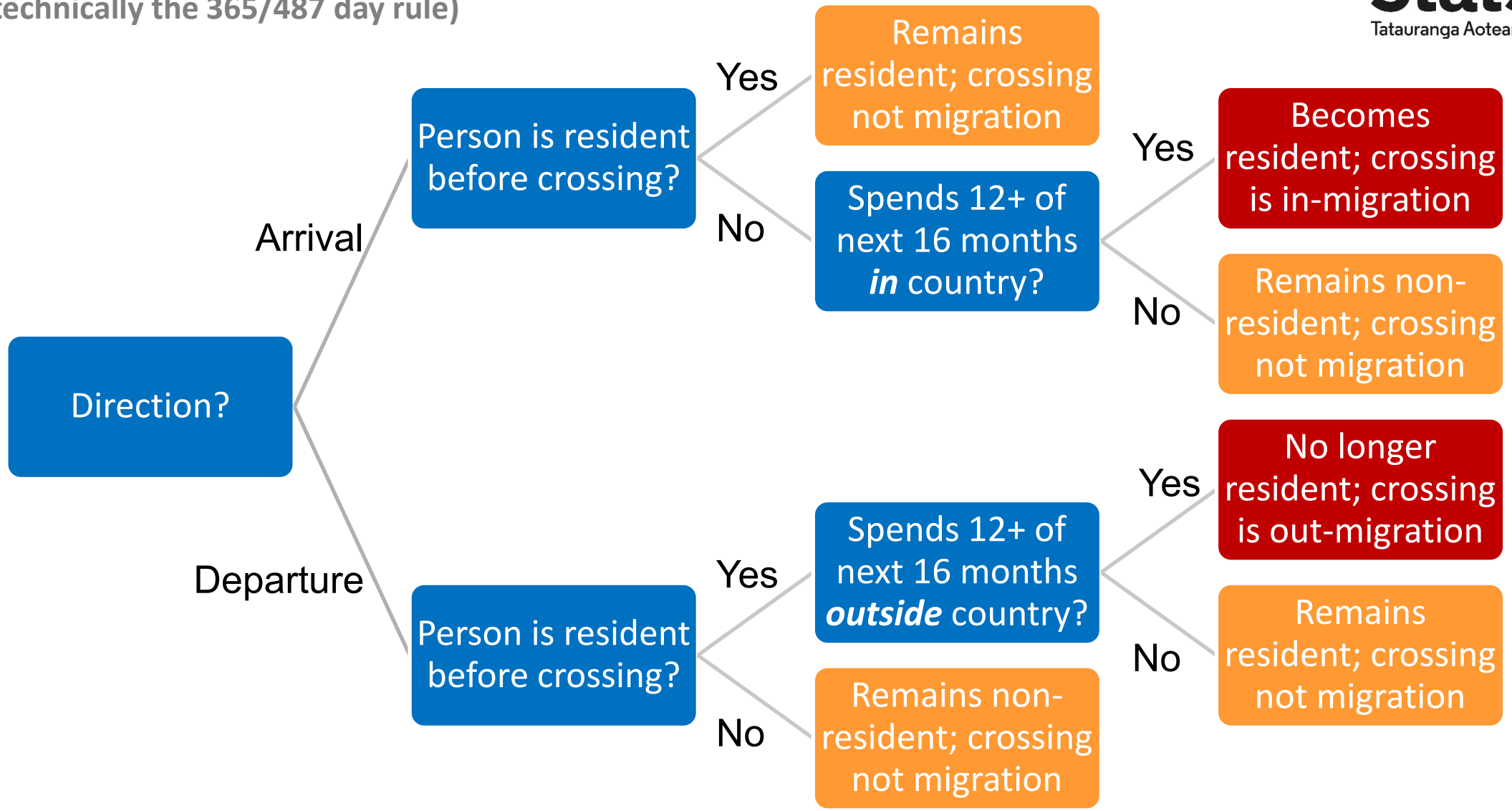
- Awareness of intentions limitations
- Changing nature of migration
  - Temporary visas
  - Migrants making multiple border crossings
- Ability to link arrival and departure records
  - Availability of linking variables
  - Linking methodology/techniques have improved
  - IT grunt
- Growing volume of border crossings
- Proposed removal of the departure card

How outcomes-based approaches work

- Identify unique individuals in the border-crossing data
- Construct a ‘travel history’ for each person
  - Link together all arrivals and departures for the same person
- Apply a classification rule
  - Classify each arrival/departure as a ‘migrant’ vs ‘not migrant’
- 12/16 - month rule is one outcomes-based approach

The 12/16 month rule

(Or technically the 365/487 day rule)



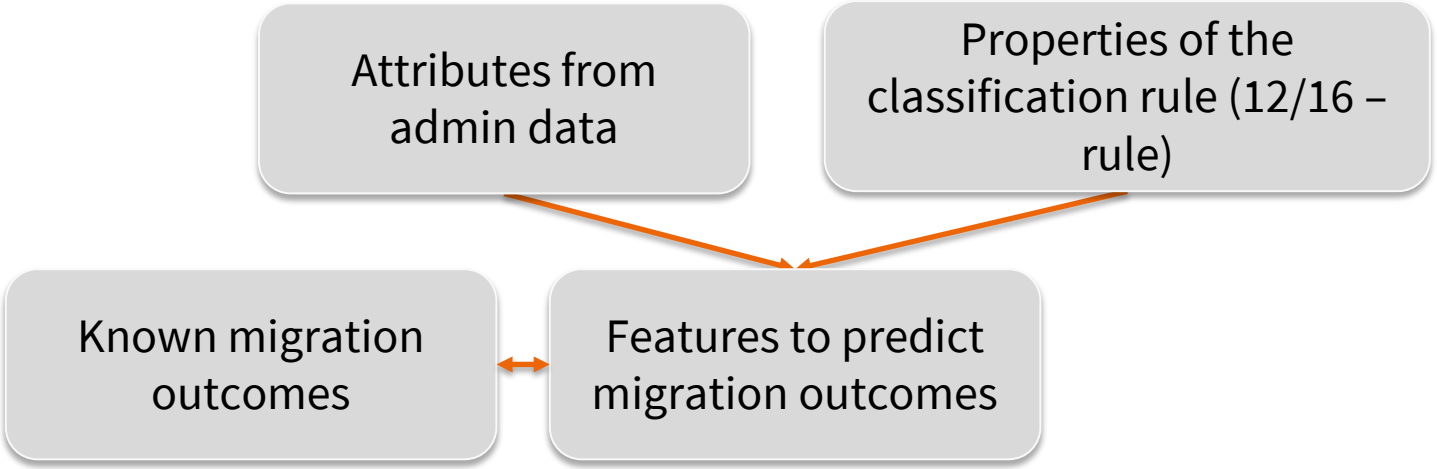
Overcoming the loss of timeliness via prediction

- New Zealand customers cannot wait over 16 months for migration measures
  - They expect results about 1 month after the reference period
- To apply the 12/16 rule to a border crossing, we potentially have to wait 16 months after a border crossing has occurred
- Of all border-crossings (pre-COVID) only around 1 - 2% were migrant crossings
- Can we predict the classifications these crossings?

Unit record modelling with machine learning

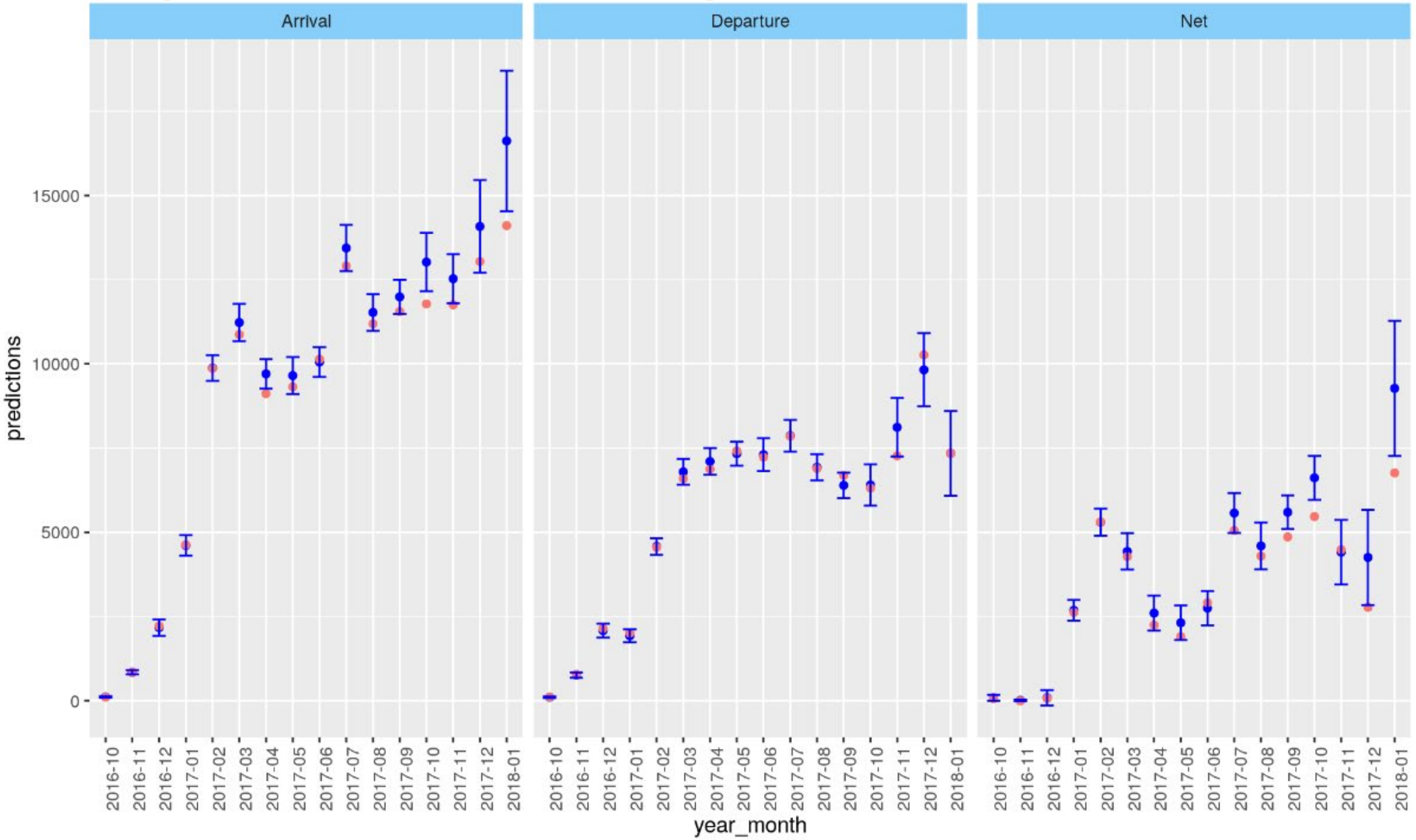
- Machine learning approach to border classification – **at unit record level**
- A gradient-boosted classifier is used, specifically implemented through XGBoost
  - An implementation of boosted trees
  - The algorithm is known to perform well in classification problems
- For each unresolved crossing, classifier estimates probability that it's a migrant crossing
- Integrates some model uncertainty – by running model on several subsamples of training data

Machine learning model features

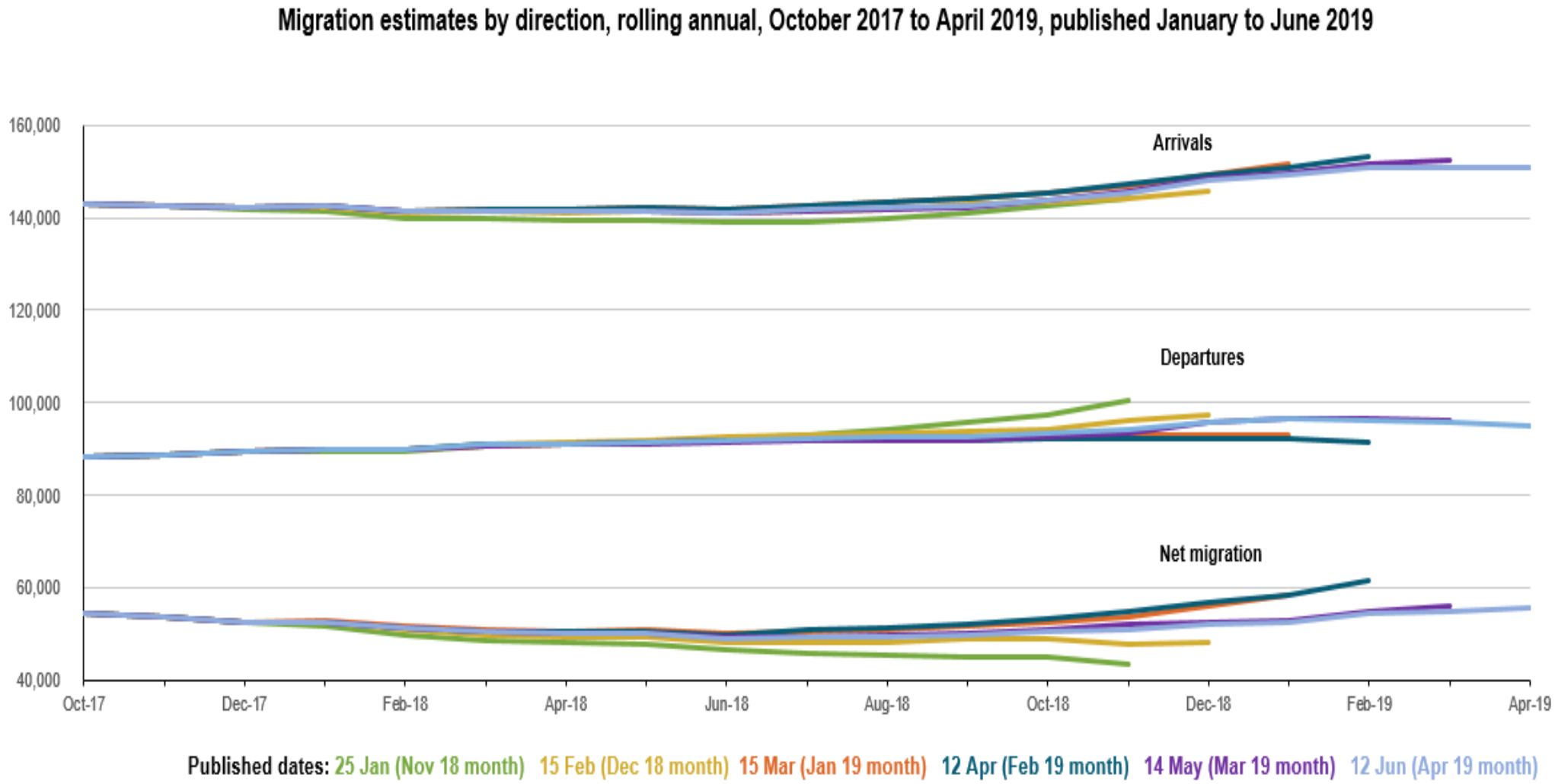


- Days observed: Days passed since the border-crossing occurred
- Days in country: Days in New Zealand since border-crossing
- Count future total: Border-crossings after the one to be classified
- Direction, age, sex, visa type, citizenship region, month of the year

Example of overall estimate performance



Provisional migration estimates



A few concluding remarks & customer/user perspective

- New Zealand uses daily border-crossing data to measure migration
- An outcomes based method, which relies on time spent in/out of the country is used – purely a population concept
  - Independent of legal status, visa type, or other such considerations
  - Independent of the intentions of the traveller – it's simply based on their behaviour
- NZ Customers require timely (around 1 month elapsed time) migration estimates – we use a machine learning model to provide this

Thank you, questions?