

A Comparison of Airline Arrival Delays

Ben Brintz
Katie Eng
Lu Wang
Fangwu Wei

Background and Question

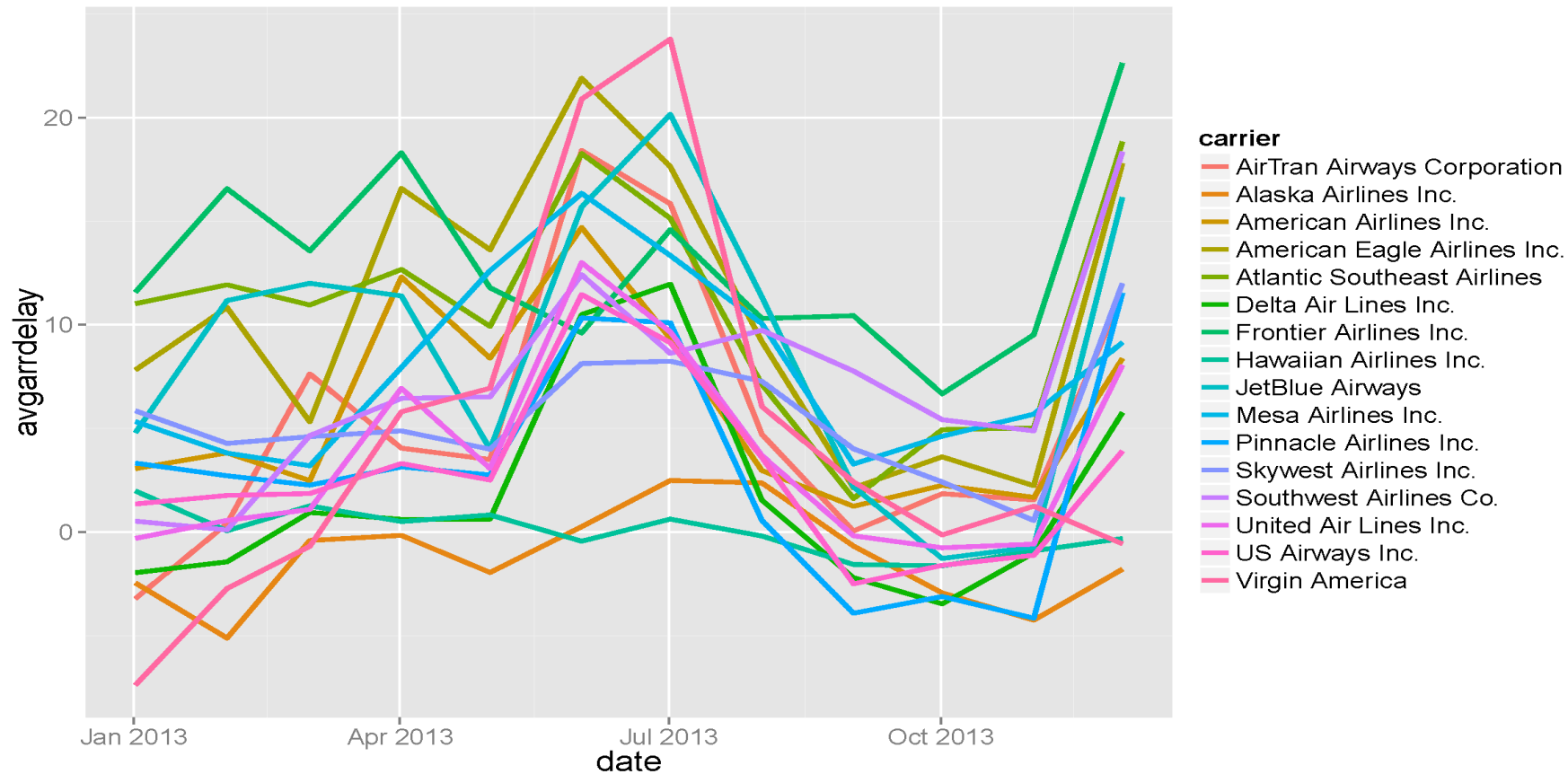
- Bureau of Transportation Statistics - on-time performance database
- Question of interest: Are there any airlines with flight delays that are systematically worse than other airlines?
- Arrival v.s. departure delays
- Population approach v.s. Sampling approach (stratified)

Population Findings



Question answered, right? Wrong, this plot is crap!

Population Findings

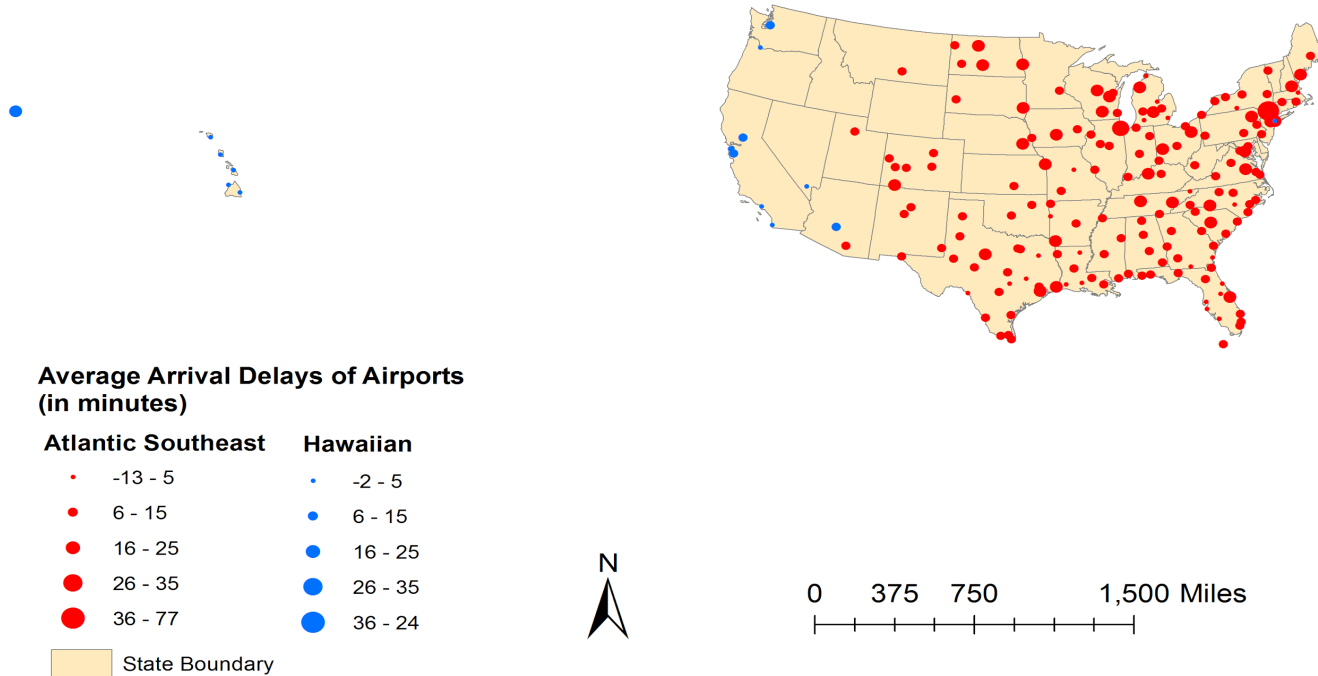


Population Findings

Go to sweet ggvis checkbox example!

Population Findings

On-Time Performance in 2013 (Atlantic Southeast Airlines and Hawaiian Airlines)



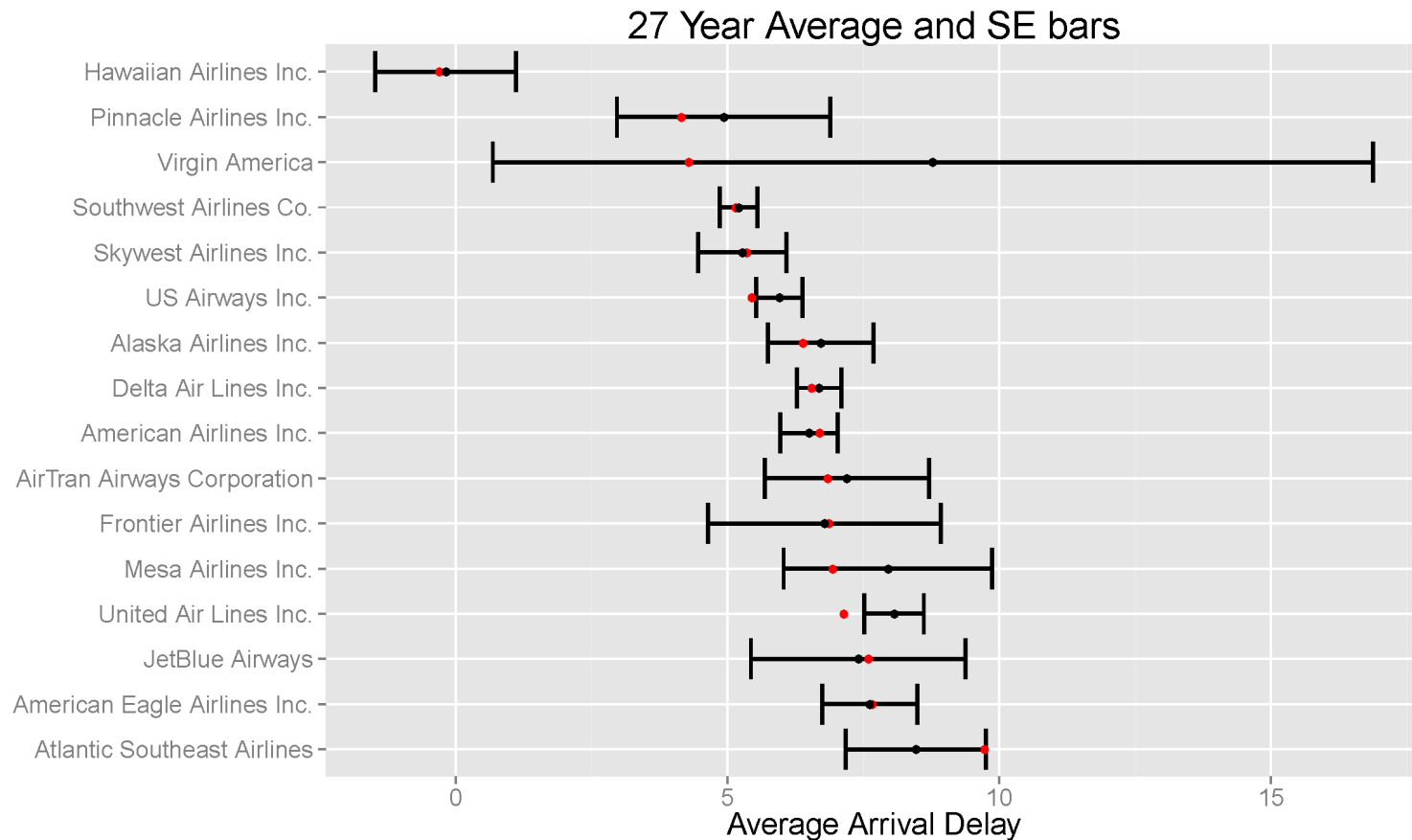
Sampling Findings

Only 2013 reported airlines

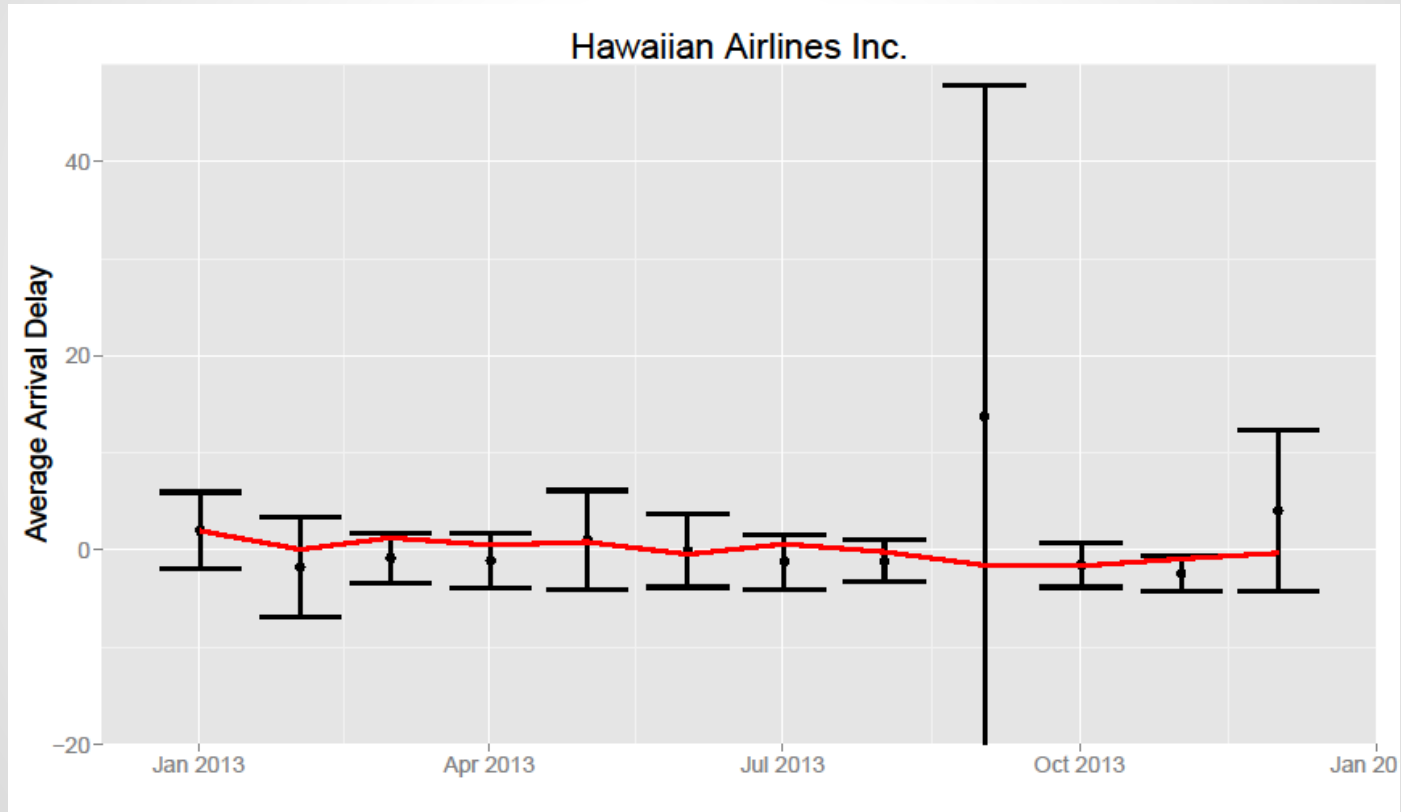
Method 1 - stratified on airlines and randomly sampled from all 27 years

Method 2 - stratified on airlines and month and sampled from only 2013

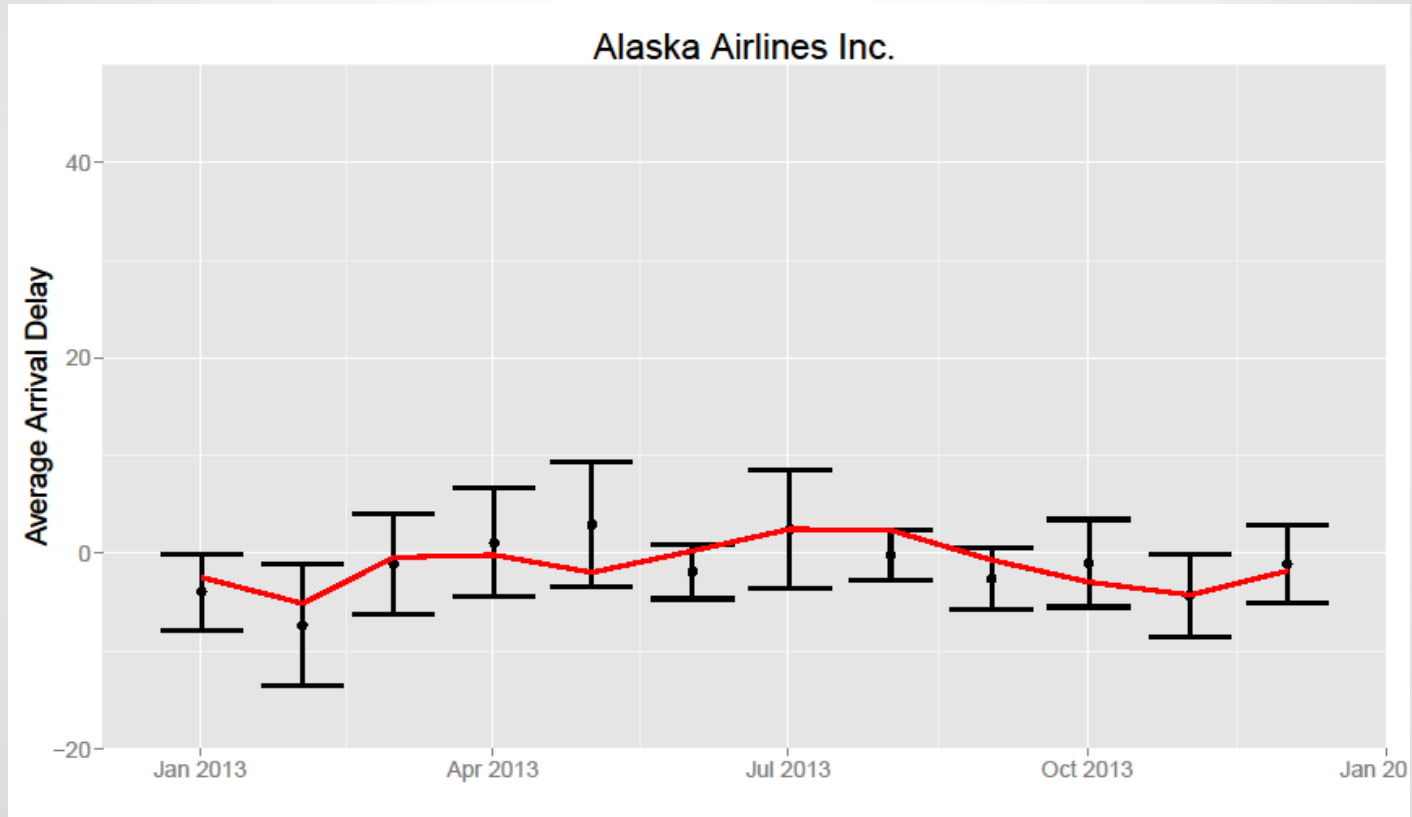
Sampling Findings (Method 1)



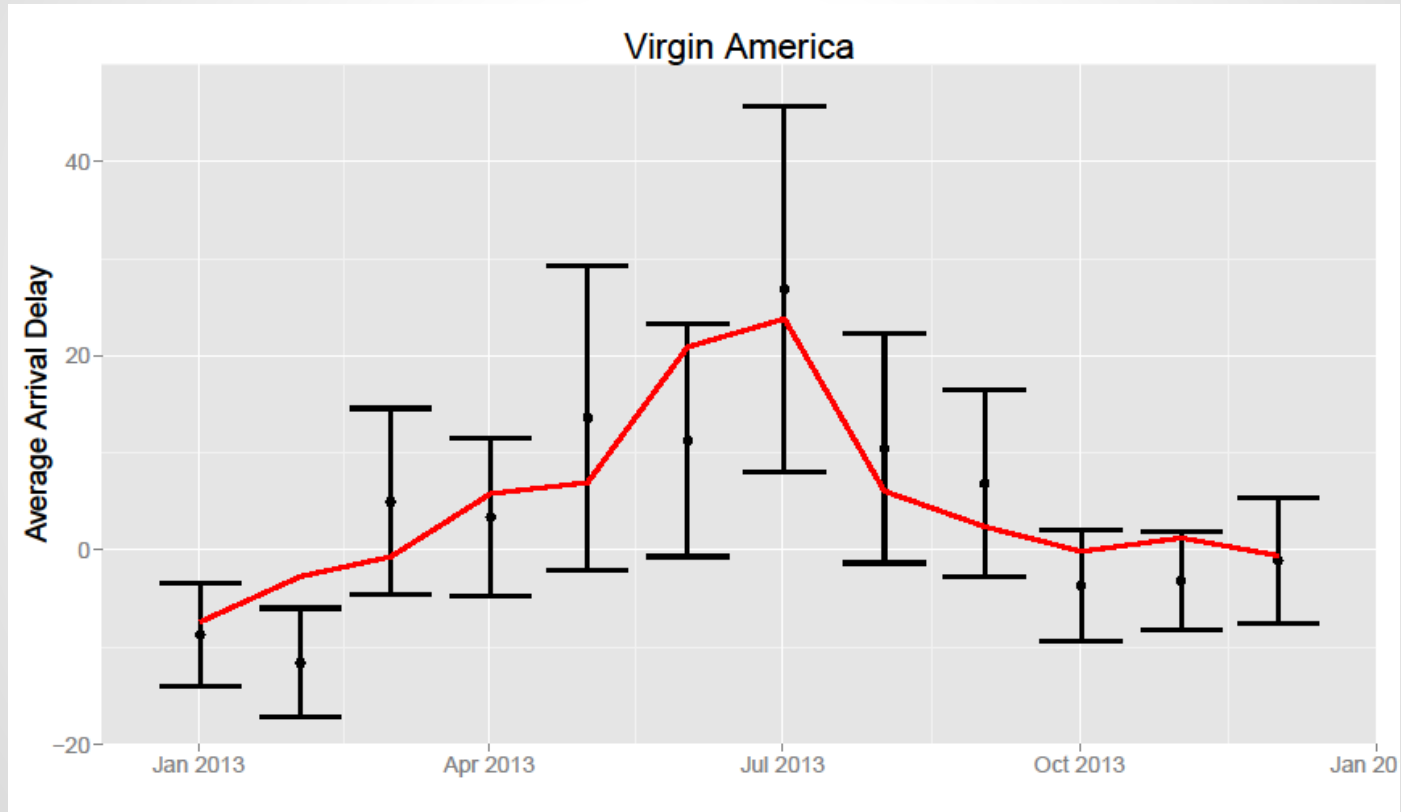
Sampling Findings (Method 2)



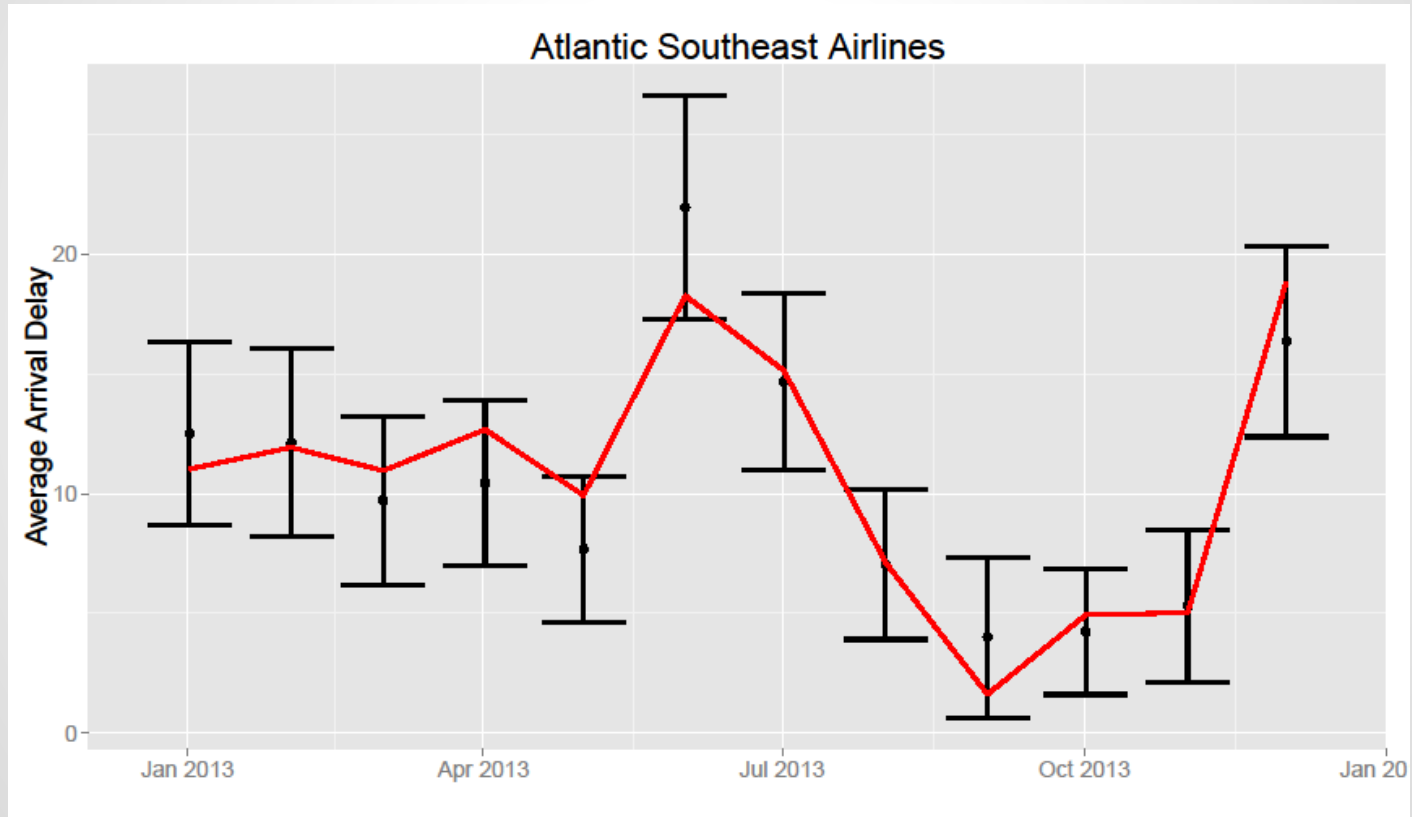
Sampling Findings (Method 2)



Sampling Findings (Method 2)



Sampling Findings (Method 2)



Comparison

Sampling

- only used 2013 airlines
- ignored time (method 1)
- only did 2013 in another sample (method 2)

Population

- looked at all airlines
- averaged over month

Obstacles and Solutions

- Data manipulation couldn't do in dplyr (loops)
- Messy plots with too much data (focus on year and ggvis experimentation)

Spatio-Temporal Visualization

<https://www.youtube.com/watch?v=8UhFhXfBzjE>

<https://www.youtube.com/watch?v=zBHopPrD9rY>

Thank you!
Questions?