Oliver Keyes Research Analyst, Analytics

I got asked a question...

I got asked a question...

How long are mobile readers' sessions?

• Interesting. Enables:

- Interesting. Enables:
- Read time analysis

- Interesting. Enables:
- Read time analysis
- View clustering

- Interesting. Enables:
- Read time analysis
- View clustering
- Defining new terms

How long does a Unique Client spend browsing Wikipedia in a single Session?

• Identifying "unique clients" (UCs)

- Identifying "unique clients" (UCs)
- Identifying "sessions"

Mobile is a particular challenge

- Mobile is a particular challenge
- Readers are a particular challenge

- Mobile is a particular challenge
- Readers are a particular challenge
- Mobile architecture: AOL from the early 2000s

- Mobile is a particular challenge
- Readers are a particular challenge
- Mobile architecture: AOL from the early 2000s
- Reader PII: highly limited

- Mobile architecture:
- On a wireless network? No problem!
- Using a data plan? Problem.

- Mobile architecture:
- On a wireless network? No problem!
- Using a data plan? Problem.
- Very few externally facing IP addresses

- Reader PII:
- (deliberately) limited in our request logs

- Reader PII:
- (deliberately) limited in our request logs
- IP address, User Agent and language variant

- Reader PII:
- (deliberately) limited in our request logs
- IP address, User Agent and language variant
- No cookies, no links to cookies

- Algorithm with the most entropy:
- IP + UA + language variant

- Identifying sessions:
- How do we define a session?

- Identifying sessions:
- How do we define a session?
- Doug Lea: (B M* E)

- Identifying sessions:
- How do we define a session?
- Doug Lea: (B M* E)
- look at when UCs stop making actions

Implementation, the first: Referrer tracking

- Implementation, the first: Referrer tracking
- Some UCs don't have referrers

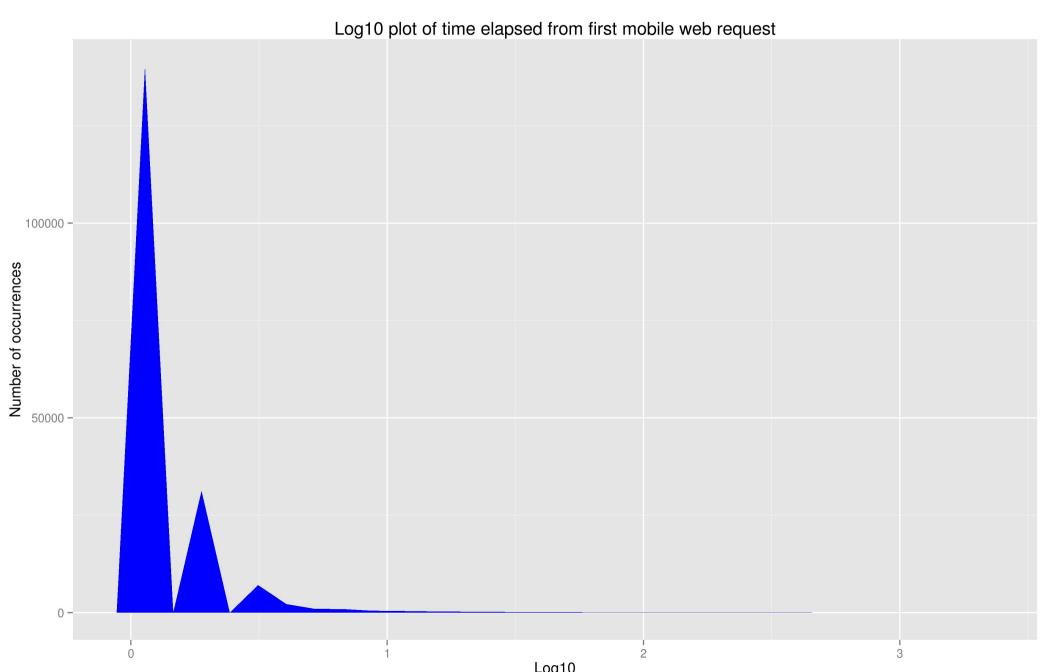
- Implementation, the first: Referrer tracking
- Some UCs don't have referrers
- Some UCs are mad. Well, all users are mad.

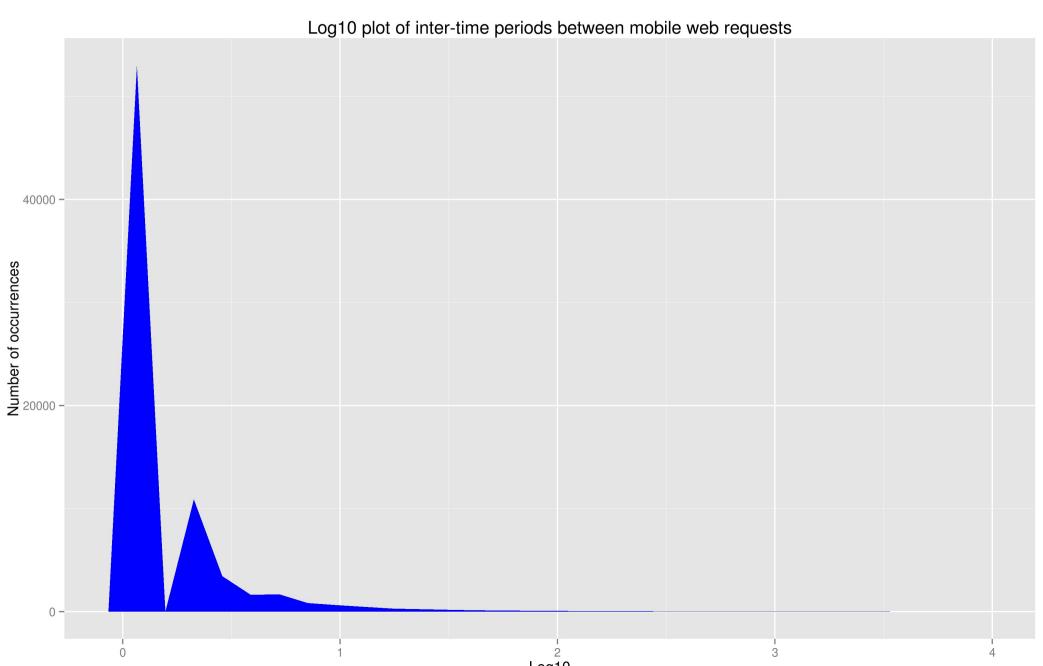
- Implementation, the first: Referrer tracking
- Some UCs don't have referrers
- Some UCs are mad. Well, all users are mad.
- Compensating for this requires heavy tech input.

• Implementation, the second: inter-time analysis

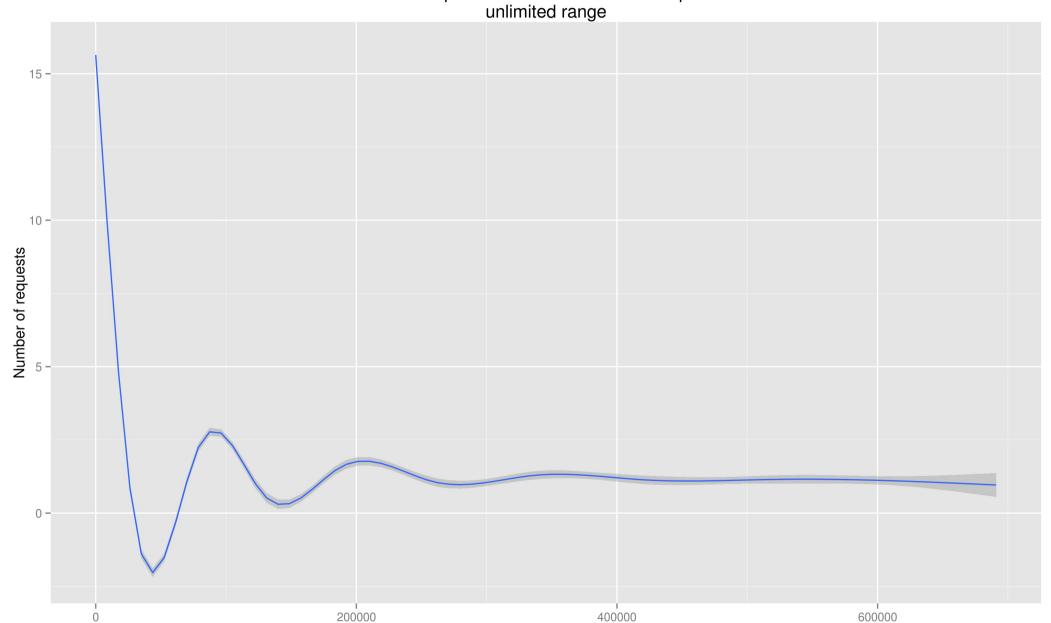
- Implementation, the second: inter-time analysis
- Look at the time between requests, for sampled UCs

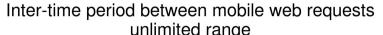
- Implementation, the second: inter-time analysis
- Look at the time between requests, for sampled UCs
- When people stop making actions, that's the end of the session.

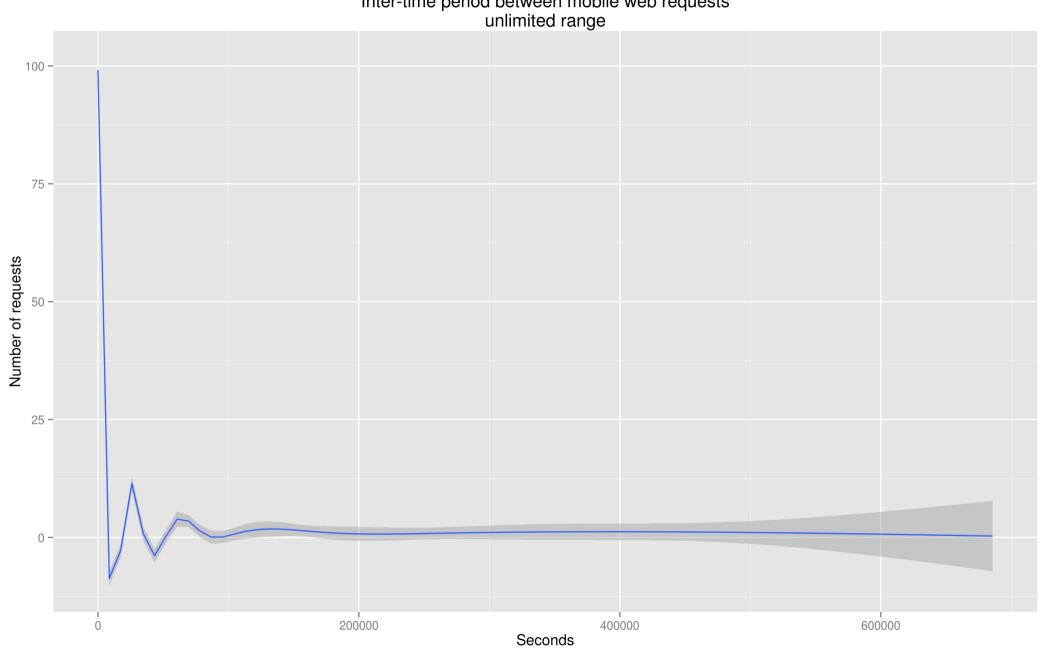


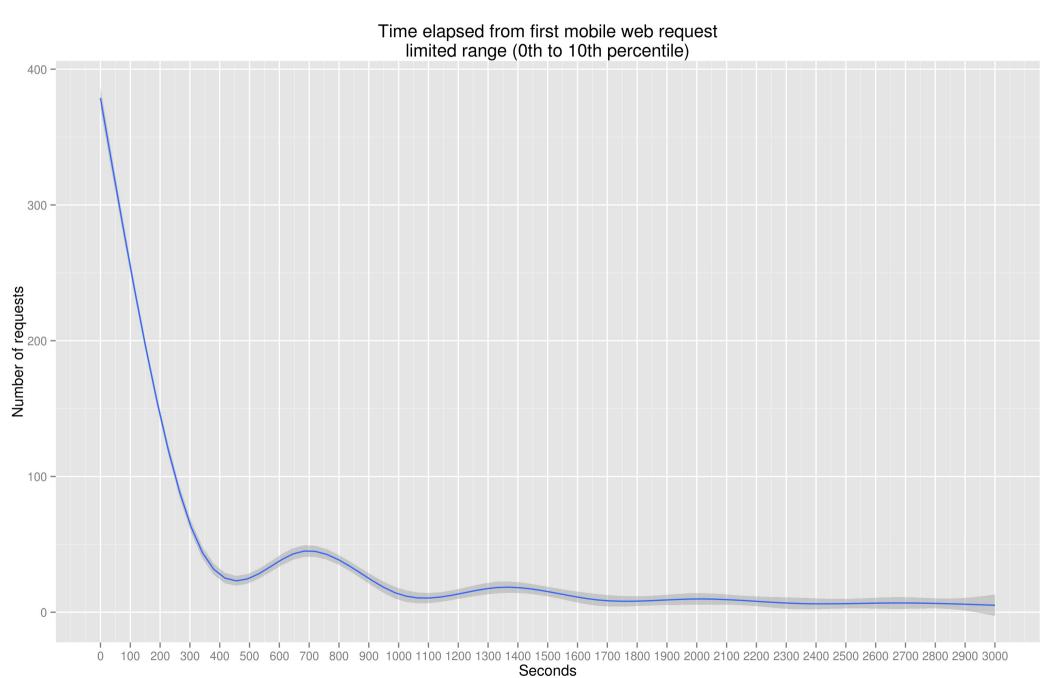


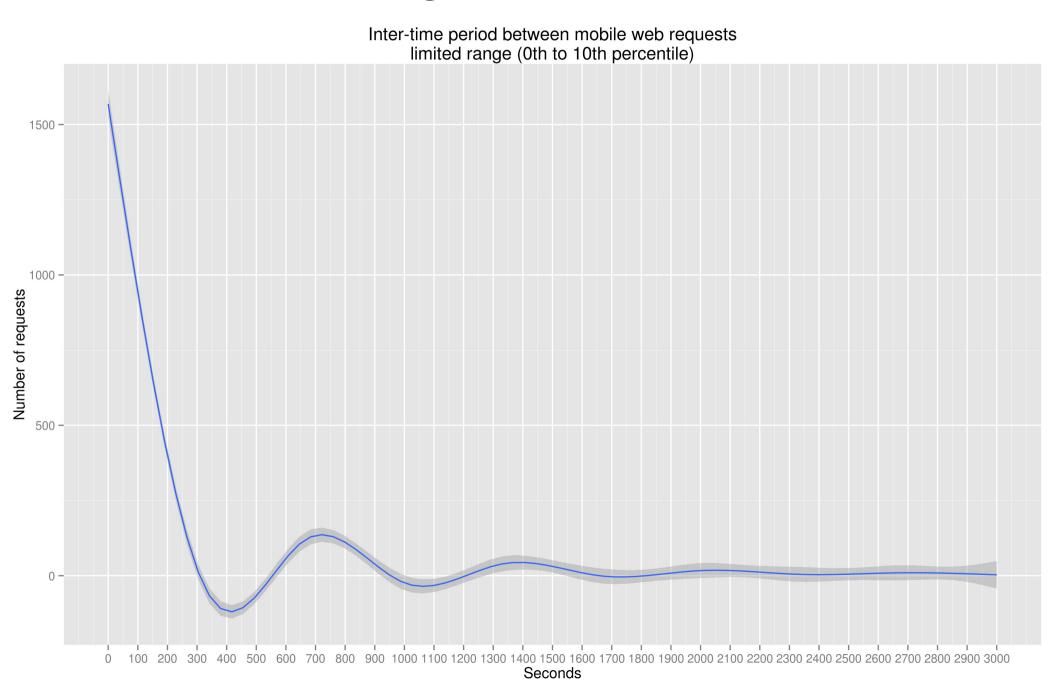
Time elapsed from first mobile web request

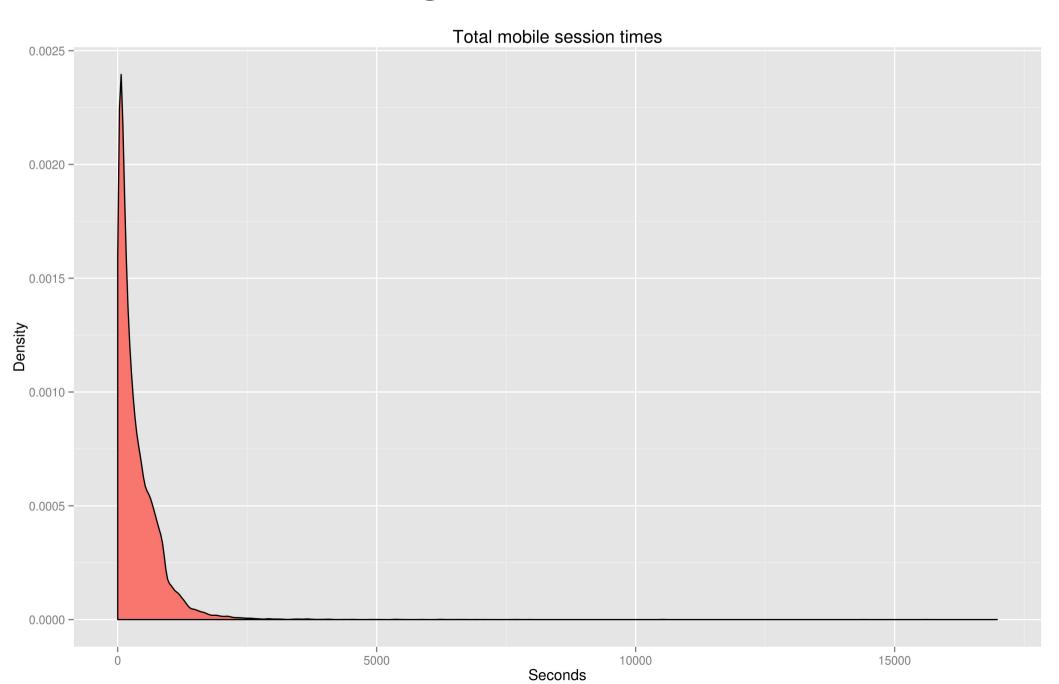


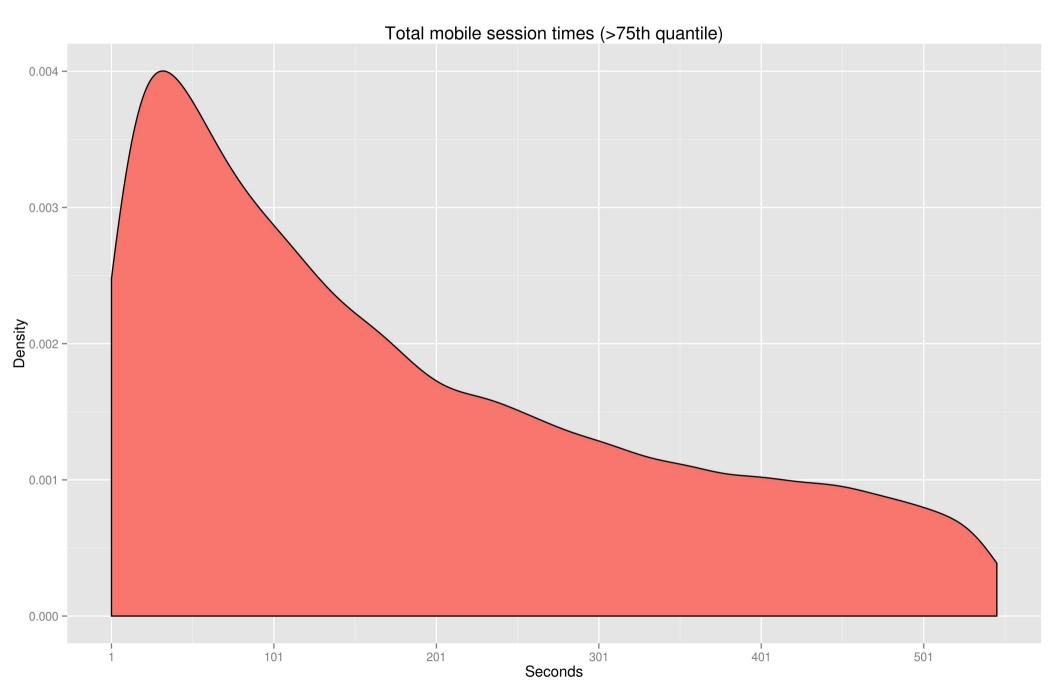




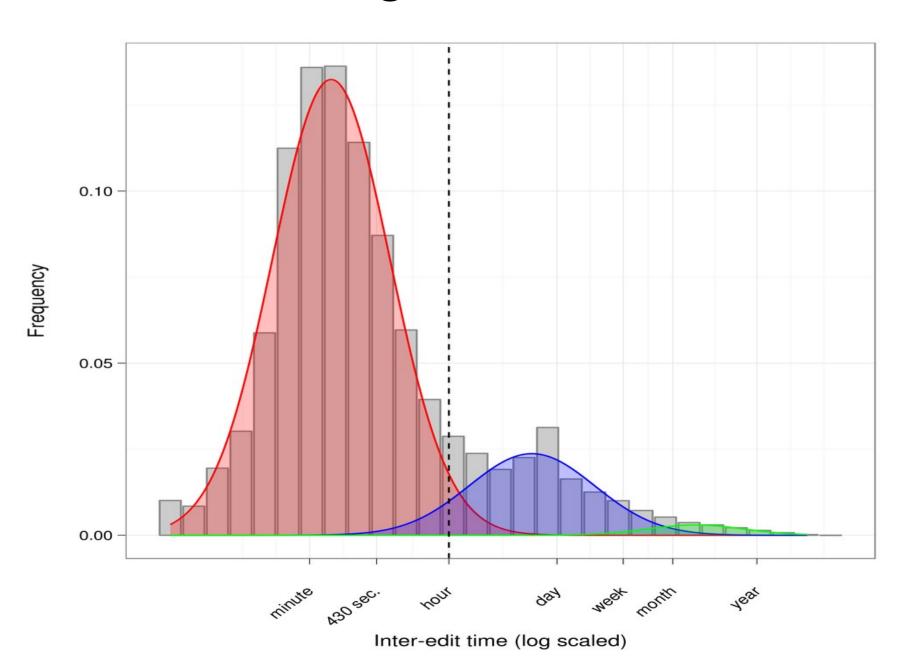


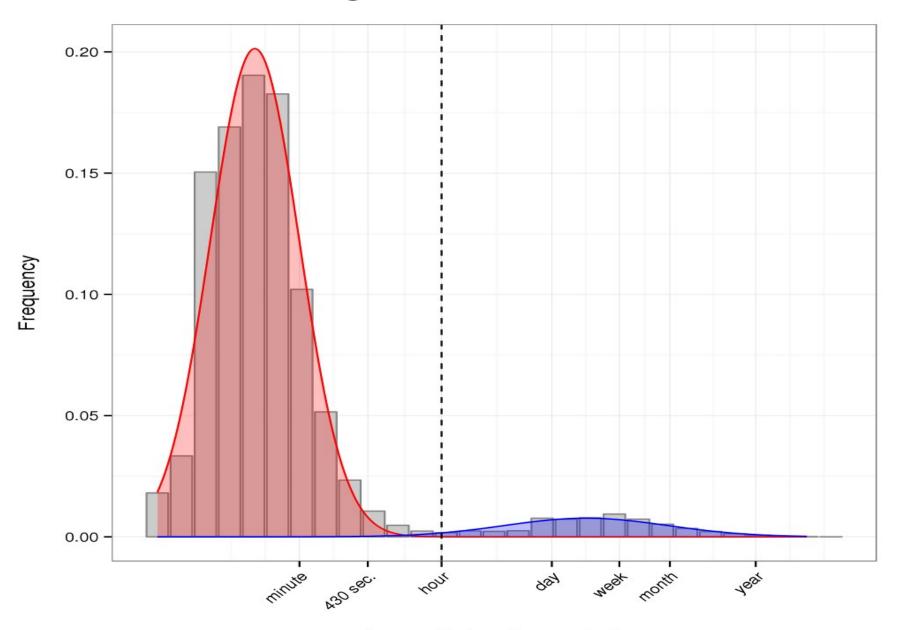




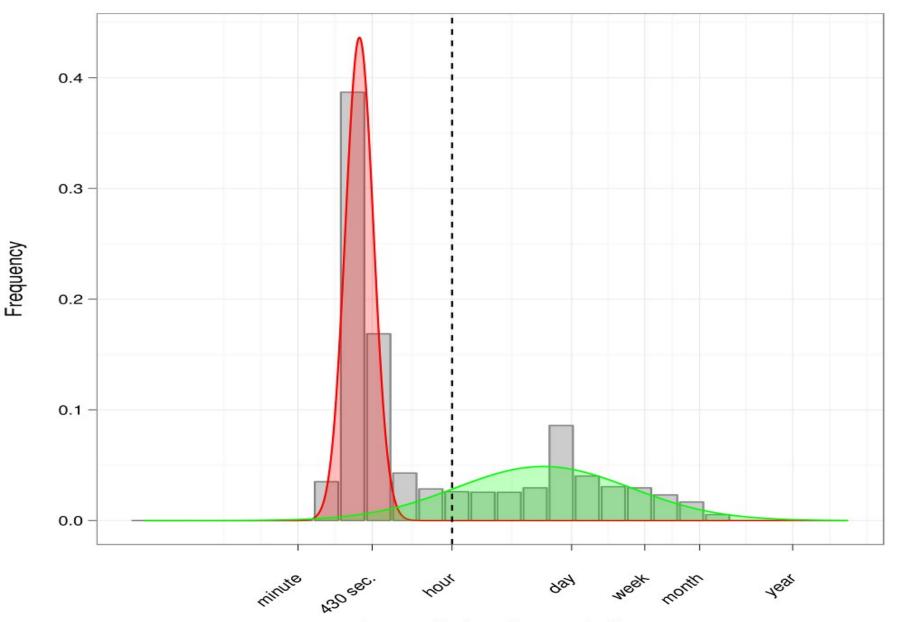


 Still preliminary. But: my gut says that this is accurate. Why?





Inter-edit time (log scaled)



Inter-edit time (log scaled)

Beat on the UC algorithm:

- Beat on the UC algorithm:
- Work out entropy added by each layer

- Beat on the UC algorithm:
- Work out entropy added by each layer
- Compare to mobile edits, where we have UUIDs

- Beat on the UC algorithm:
- Work out entropy added by each layer
- Compare to mobile edits, where we have UUIDs
- Fix session algorithm:
- Make it identify all sessions associated with a UC

- Explore alternative models:
- Breadth-first parser?

- Explore alternative models:
- Breadth-first parser?
- Temporarily include mw.session.id?

Code and graphics: github.com/Ironholds/MobileSessions

Questions?