Lecture 4:
Knowing the Crowd Workers (Cont.)

Requesters and Tasks in Crowdsourcing Markets.

Instructor: Chien-Ju (CJ) Ho

Logistics

- Bid for Presentations
  - Make sure to bid for presentations by the end of today

- If you haven't found a team, talk to people!

- There might be minor changes to reading materials
  - Adding papers to optional readings
  - Update a few lecture titles to better reflect the included papers

Methods for Information Collection (Last Lecture)

- Direct surveys
  - Issues to keep in mind: Sampling bias; Untruthful reports

- Ethnographic analysis (as in the required reading)

- More intricate methods
  - Example 1: How to measure the communication network of the crowd
    - A web app to map networks
    - Impacts of networks to crowdsourcing
  - Example 2: How to measure the size of the crowd

Examples on Networks Induced Collaborations

- https://blog.humancomputation.com/?p=9271
- A half-day hackathon-style experiment

  "Does the fact that workers exist in networks help them solve problems?"

- 100 workers are asked to solve TSP problems
  - Base pay: $0.10
  - Bonus depends on the answer
    - Best answers get $2 bonus
    - Others get bonuses $2 – $0.1 * deduction.
    - For every 100 mile away from the best answer, deduction++.

  - Workers are explicitly told that they can collaborate

Workers started new threads to collaborate on the HIT.
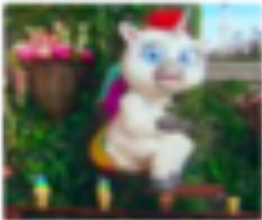
Seattle to
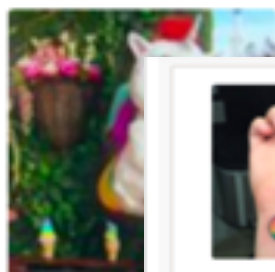
This route

Squatty
Magical H
Unicorn

Contri

HIT P

Joined: M
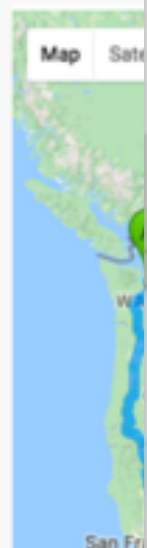
> **Jesicu said:** ↑
>
> *I have Seattle, LA, Denver, Kansas City, Chicago, Atlanta, Houston, Dallas*

> **Jesicu said:** ↑
>
> *By air, the one I found is 4272.3*

I think that's right. It's a straight line into Kansas City then a loop of those last five cities, and you want to go around the loop in the direction that skips the longer of the two routes out of Kansas City, which is KC->Dallas (457.9), and instead go the shorter way to Chicago (407.5).

Kerek, Sep 1, 2016   Last edited: Sep 1, 2016   #3

> **Kerek said:** ↑
>
> *I think that's right. It's a straight line into Kansas City then a loop of those last five cities, and you want to go around the loop in the direction that skips the longer of the two routes out of Kansas City, which is KC->Dallas (457.9), and instead go the shorter way to Chicago (407.5).*

I agree, I'm not sure how 4150 was arrived at with the other one as that comes out to 4488.7 for me.

Starslip, Sep 1, 2016

Methods for Information Collection (Today)

- Direct surveys
  - Issues to keep in mind: Sampling bias; Untruthful reports

- Ethnographic analysis (as in the required reading)

- More intricate methods
  - Example 1: How to measure the communication network of the crowd
    - A web app to map networks
    - Impacts of networks to crowdsourcing
  - Example 2: How to measure the size of the crowd

Delve into the Digital Trace

Martin et al. "Being a Turker", *CSCW'14*

Purposes and Earnings

Why do the work?

- Money
- *"While I do find some of the HITS fun and actually learn an incredible amount by doing HITS, I do it for the cash."*

How much they earn in a year?

- $1,179 to $14,476.93. Rather low-paid!

Why do they "choose this wage"?

- From primary income to make ends meet to supplementary
- A safety-net for some!

Requester-Worker Relationship

- Workers share information on "requesters hall of fame/shame"
  - Good requesters: fair pay, quick approval, willing to communicate
  - Bad requesters: rejections without clear reasons, demeaning comments, unfair treatment

- Similar concepts lead to a plug in for "requester reputation".
  - Turkopticon
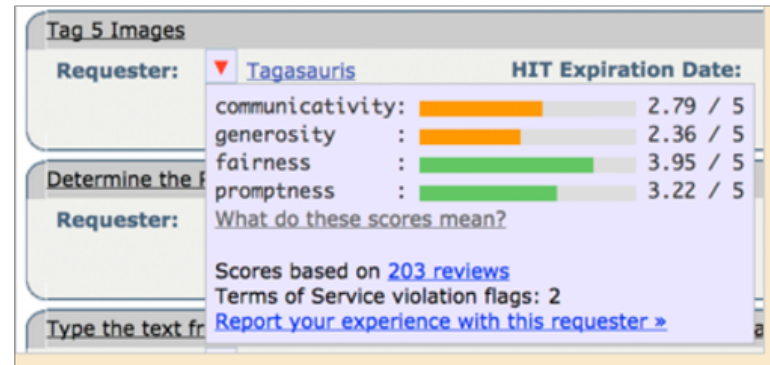
Requester-Worker Relationship

- Workers share information on "requesters hall of fame/shame"
  - Good requesters: fair pay, quick approval, willing to communicate
  - Bad requesters: rejections without clear reasons, demeaning comments, unfair treatment

- Similar concepts lead to a plug in for "requester reputation".

- Asymmetric power and limited communication (provided by the platform):
  - Requesters can reject work but workers can hardly appeal
  - Requesters can block workers but workers can't easily block requesters

Invisible Work

- Search for good HITs / good requesters
- Learn how to best set up computers/ tools
- Learn how to best improve skills in a certain task
- Learn how to manage turking and record keeping

Warm-Up Discussion

- What do you think are the advantages and disadvantages of such methodology (i.e., ethnographic analysis of worker's digital trace)?

- Can you think of ways to address the limitations?

Methods for Information Collection (Today)

- Direct surveys
  - Issues to keep in mind: Sampling bias; Untruthful reports

- Ethnographic analysis (as in the required reading)

- More intricate methods
  - Example 1: How to measure the communication network of the crowd
    - A web app to map networks
    - Impacts of networks to crowdsourcing
  - **Example 2: How to measure the size of the crowd**
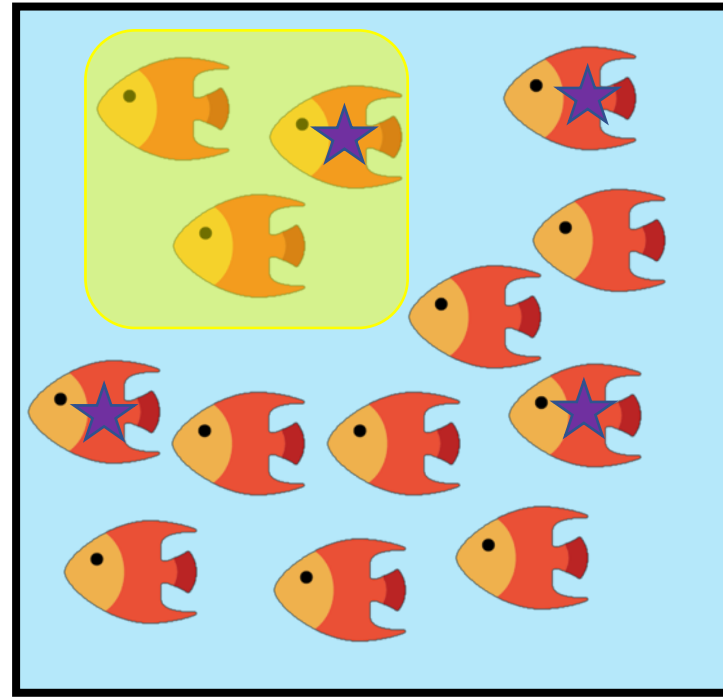
How Many Fishes are in the Lake?
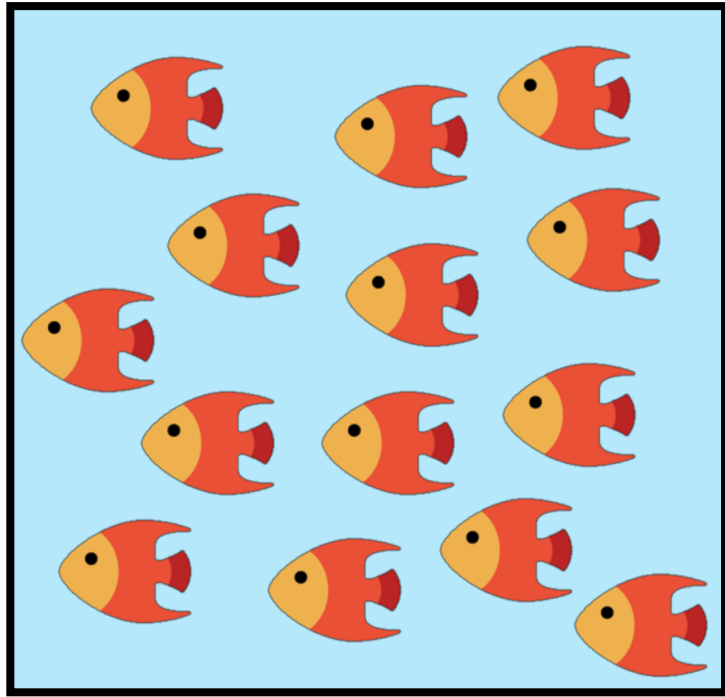
How Many Fishes are in the Lake?

The capture-recapture analysis
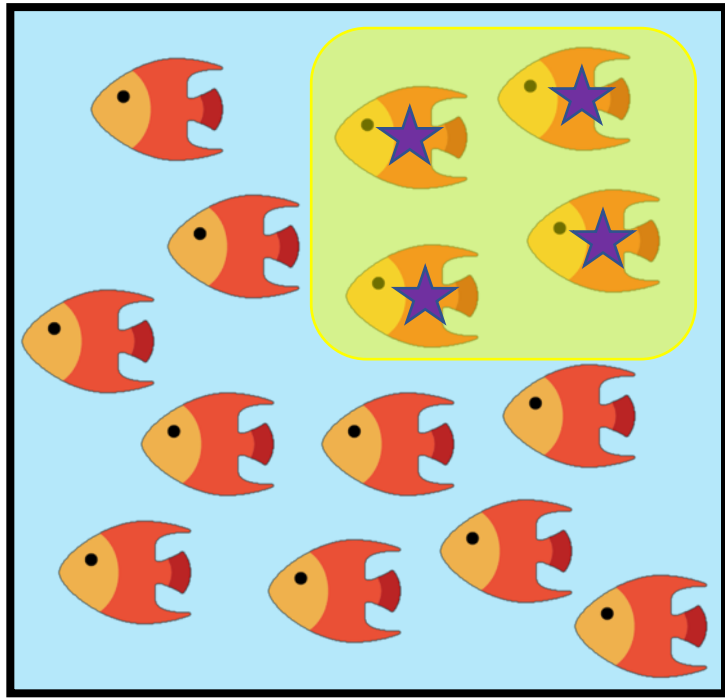
$$n = \frac{N_1}{M} \times N_2$$

$$M = \frac{N_1 N_2}{n}$$

M fish, N_1 captured (& marked) N_2 recaptured, n marked

How Large is the Crowd?

- Applying the capture-recapture analysis to workers who respond to demographic surveys, we concluded that $M = 13,410$.

- What assumptions are made?

  - A1: Closed population model: No workers will leave the worker pool, and no new workers will join the pool

  - A2: Equal propensity of participation: The probability to participate in demographic survey is equal across all workers.

Difallah et al. "Demographics and Dynamics of Mechanical Turk Workers", *WSDM'18*
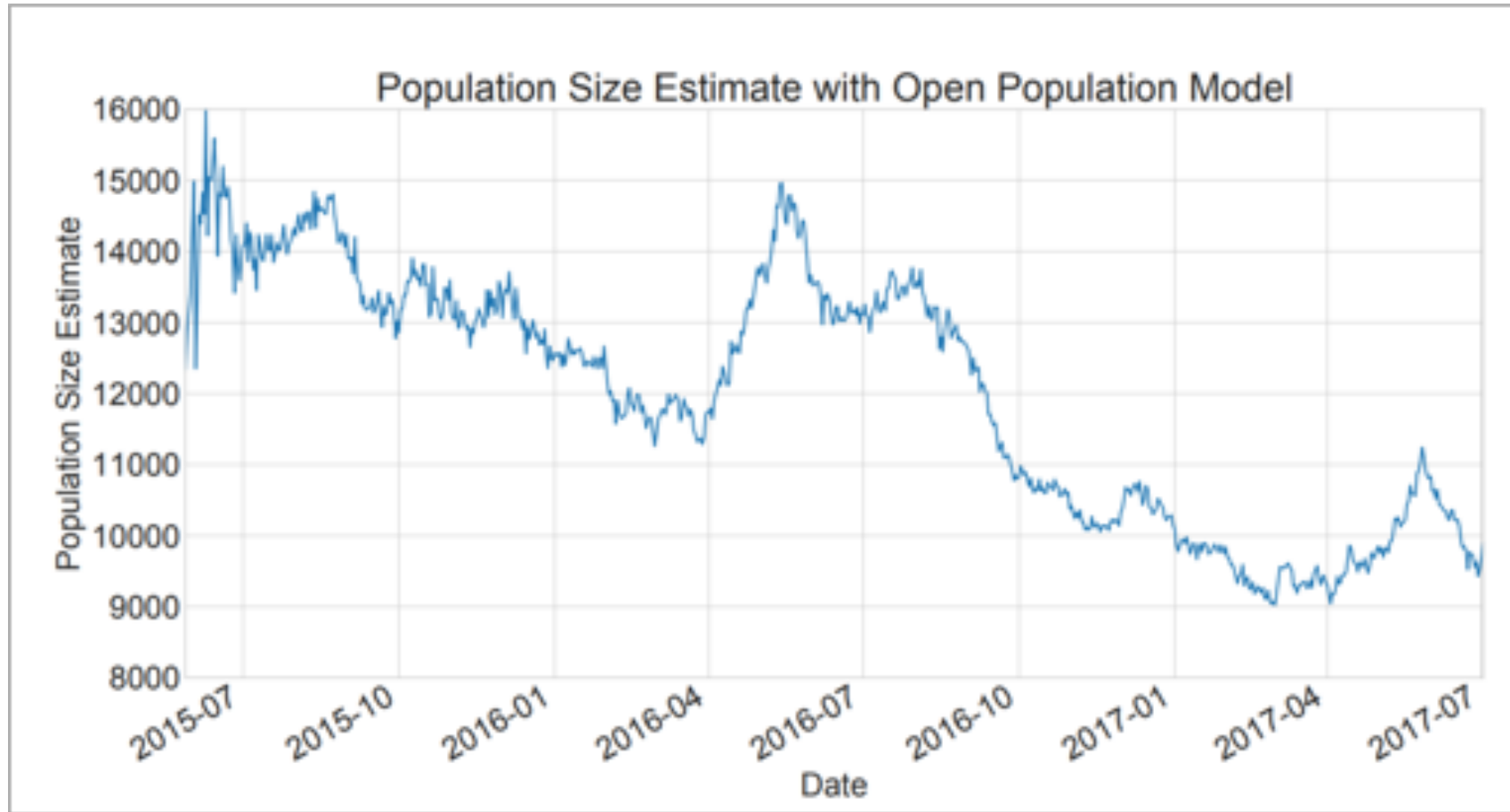
How Large is the Crowd?

Relaxing A1: Open population model

- Survival probability: $S(t) = \exp(-\lambda t)$

- At time $d - t$, N_{d-t} workers participated in the survey

- At time d, assume that there are M_d workers in total. Among N_d workers that participated in the survey, $n_{d,t}$ workers also participated in the survey at time $d - t$.

$$n_{d,t} = \frac{N_{d-t}\exp(-\lambda t)}{M_d} \times N_d$$

$$M_d = \frac{N_{d-t}N_d\exp(-\lambda t)}{n_{d,t}}$$

Difallah et al. "Demographics and Dynamics of Mechanical Turk Workers", *WSDM'18*

How Large is the Crowd?

Half-life of the MTurk worker population is 404 days

Difallah et al. "Demographics and Dynamics of Mechanical Turk Workers", *WSDM'18*

How Large is the Crowd?

- Is "A2: Equal propensity of participation" reasonable?
  - If every worker participate the survey with the same probability, we should see a binomial distribution over # participations with many surveys.

Number of times a worker participated in the 28-month surveys
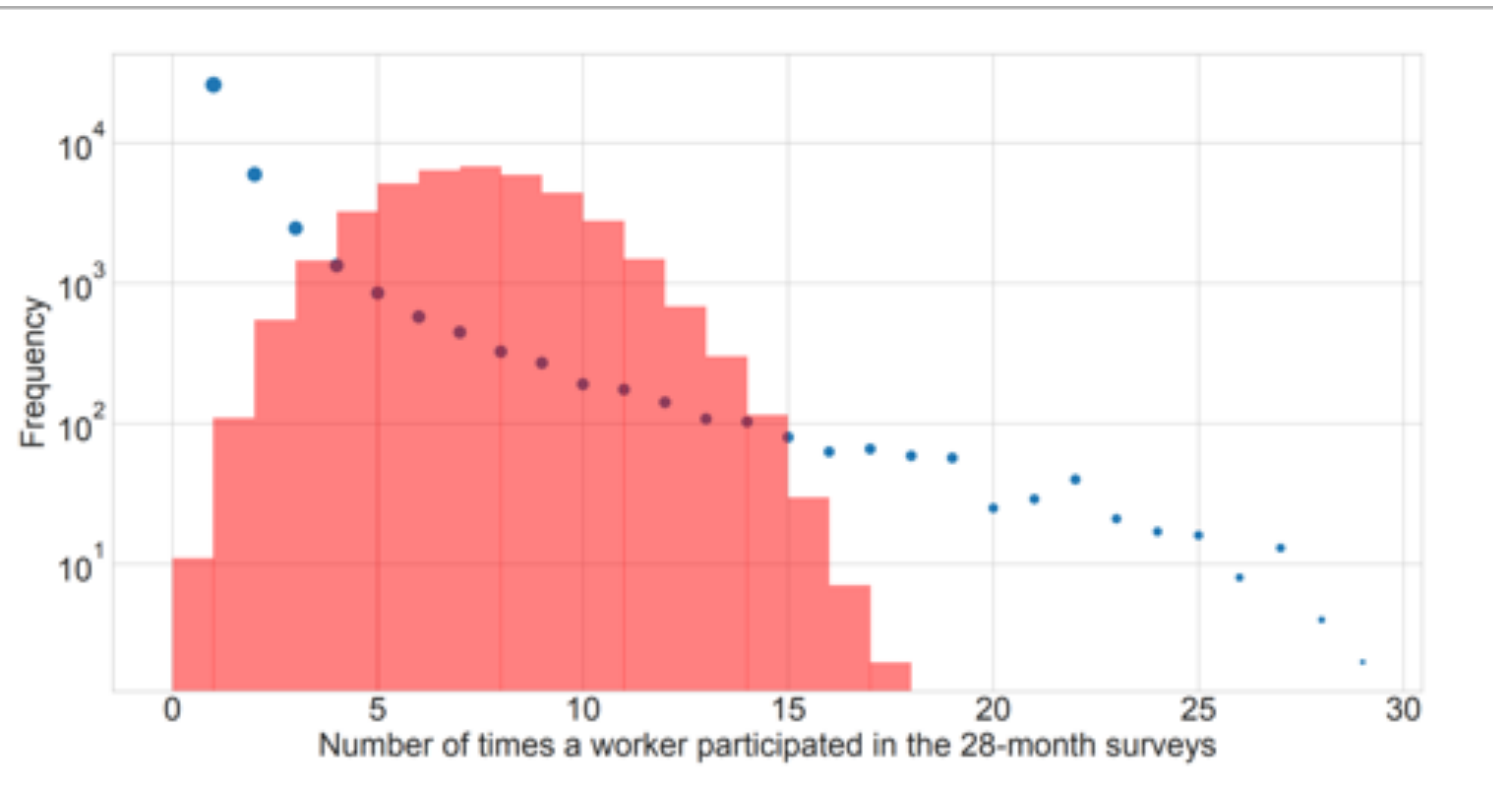
How Large is the Crowd?

Relaxing A2: Accounting for propensity of participating surveys

- Assume there are N workers in total, each worker gets a propensity score $a_i \sim p(a)$

- Capture n_1 workers, $P(capture|a_i) = 1 - (1 - \frac{a_i}{\sum_{j=1}^N a_j})^{n_1} \approx \frac{n_1 a_i}{NE[a]}$

- Recapture n_2 workers, $P(capture, recapture|a_i) = \frac{n_1 n_2 a_i^2}{N^2 E[a]^2}$

- The expected number of workers participating in both surveys:

$$m = N \int \frac{n_1 n_2 a_i^2}{N^2 E[a]^2} p(a) da = \frac{n_1 n_2}{N} (1 + \frac{Var[a]}{E[a]^2})$$

- Following this method, the estimate for the number of Mturk workers is 178,000!

Difallah et al. "Demographics and Dynamics of Mechanical Turk Workers", *WSDM'18*

Data-Driven Approaches

- Want to collect some information. Challenging to obtain enough data points to estimate the information you need.

- Propose models (i.e., making assumptions) of the process. So you only need to estimate a small number of model parameters.

- Using ML approaches to learn those parameters from data.

- Your estimate is only as accurate as the best your model can do.

Discussion

- What additional information do you want to know about Turkers?

- Can you come up with methods to collect that information?

- Does that information help us design better crowdsourcing platforms? How?

Requesters – How to Create Tasks

Creating Tasks on MTurk

- Live Demo of web interfaces
    - Setting task properties (title, description, keywords)
    - Setting payment and time
    - Setting requirements (qualifications)
        - A lot of predefined qualifications:
            - Country, approval rate, etc
        - You can set up your own qualification as well

- Setting up tasks as a webpage
    - You can use any HTML, CSS, or JavaScript to customize your layout.
    - Existing templates for common tasks are available.

- There is a [sandbox](#) website that you can play around.

Creating Tasks

- What if we want to do more complicated things that webpages can't handle?
  - Common approach:
    - Provide a link and ask workers to complete the link outside of Mturk.
    - In the outside server, give a code to Turkers when they finish the work.
    - Ask Turkers to enter the code in Mturk interface.
    - Decide on the payments based on the codes.

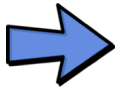- There also exists API and command line tools to automate the process.

Creating Tasks

- Programming frameworks
  - [Turkit](#) (Mar 21 lecture)
  - [Automan](#)
  - …

```
ideas = []
for (var i = 0; i < 5; i++) {
    idea = mturk.prompt(
        "What's fun to see in New York City?
        Ideas so far: " + ideas.join(", "))
    ideas.push(idea)
}

ideas.sort(function (a, b) {
    v = mturk.vote("Which is better?", [a, b])
    return v == a ? -1 : 1
})
```

Task Dynamics in MTurk

Crowd Tasks and Markets

Potential questions of interests

- What are the typical types of tasks on crowdsourcing platforms?
- How does the property of crowd tasks change over time?
- How task features influence the requester's experience?
- How task features influence the worker's experience?
- A market view: supply vs. demand

A Taxonomy of Crowd Tasks

- Surveyed 1,000 workers on CrowdFlower
- Manually annotated the type of tasks and categorized tasks into 6 classes
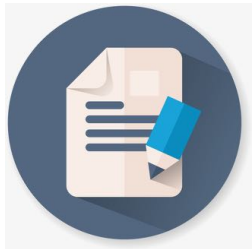
Information Finding

Verification & Validation
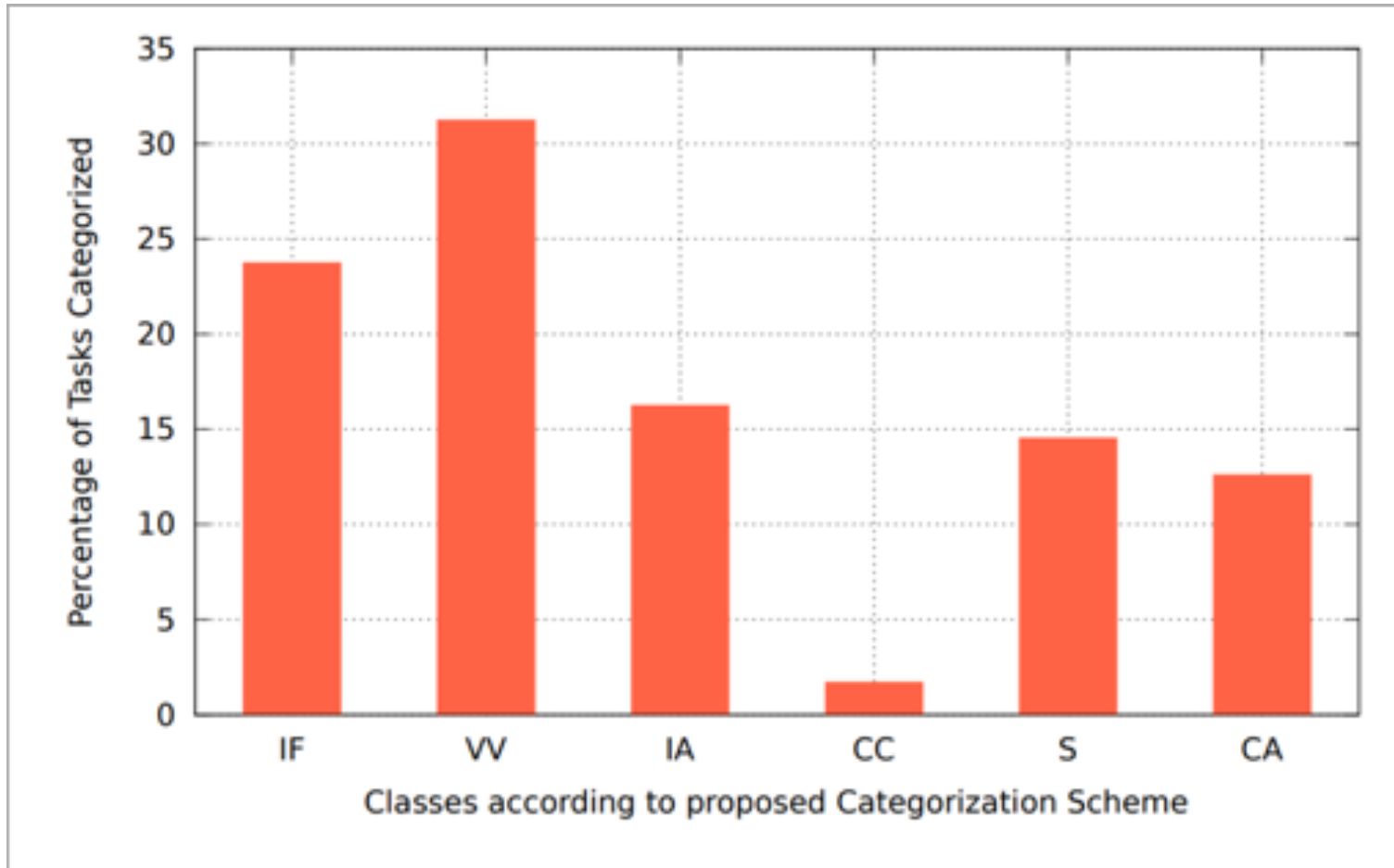
Interpretation & Analysis

Content Creation

Surveys

Content Access

Gadiraju et al. A Taxonomy of Microtasks on the Web. HT'14

# Distribution of Task Types

Problems?

A Case Study of MTurk

- Periodically crawled data from MTurk (data is available at MTurk-Tracker)
- Hourly aggregated data on tasks from 2009 to 2014
  - Task title
  - Task description
  - Task rewards
  - Task qualifications
  - Task keywords
  - Number of tasks available
- 2.5 million batches, 130 million tasks!

Difallah et al. The Dynamics of Micro-Task Crowdsourcing: The Case of Amazon MTurk. WWW'15
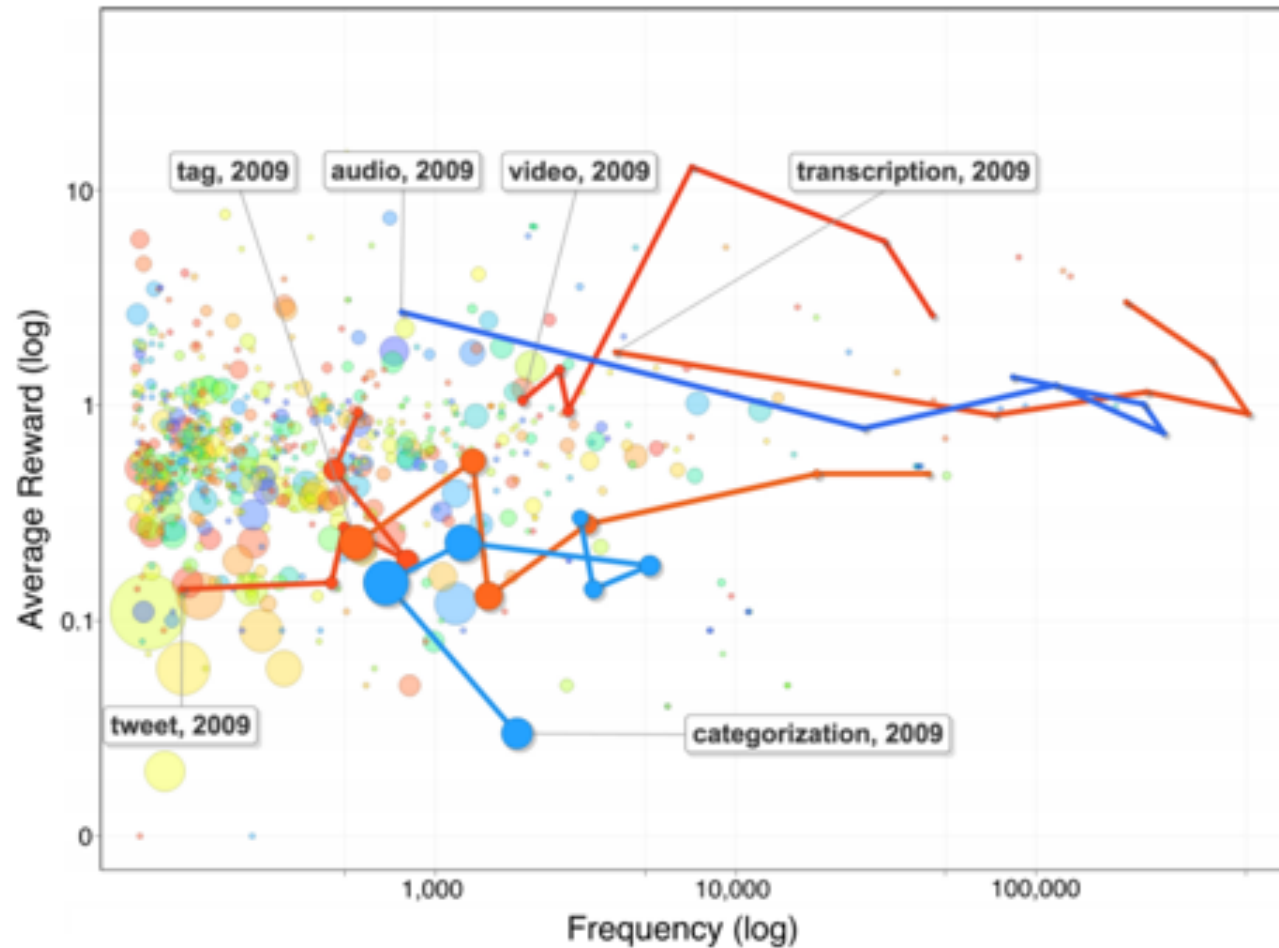
Dynamics of Task Types

- Supervised Learning for Classifying Task Types
  - Get crowd labels for a sample of 5,000 tasks
  - Use tasks with agreement to train a SVM classifier and make predictions on all other tasks
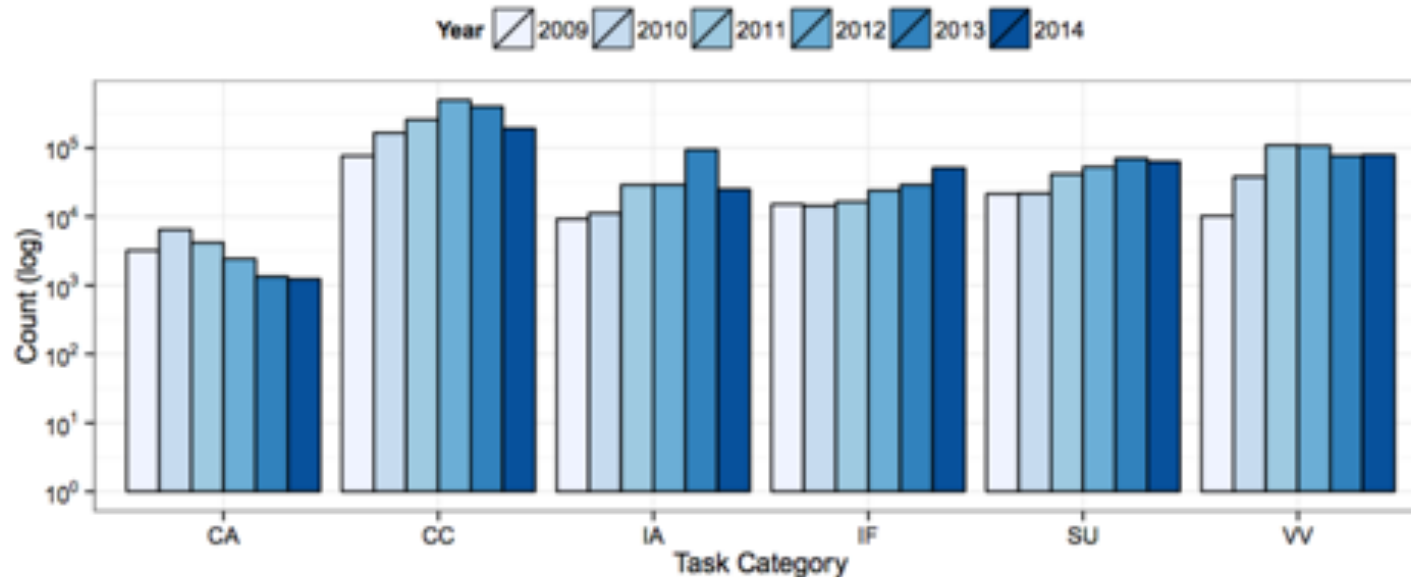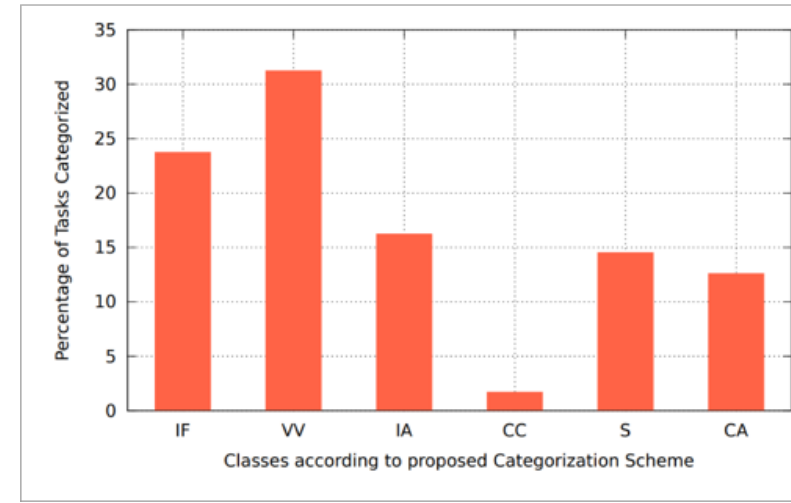
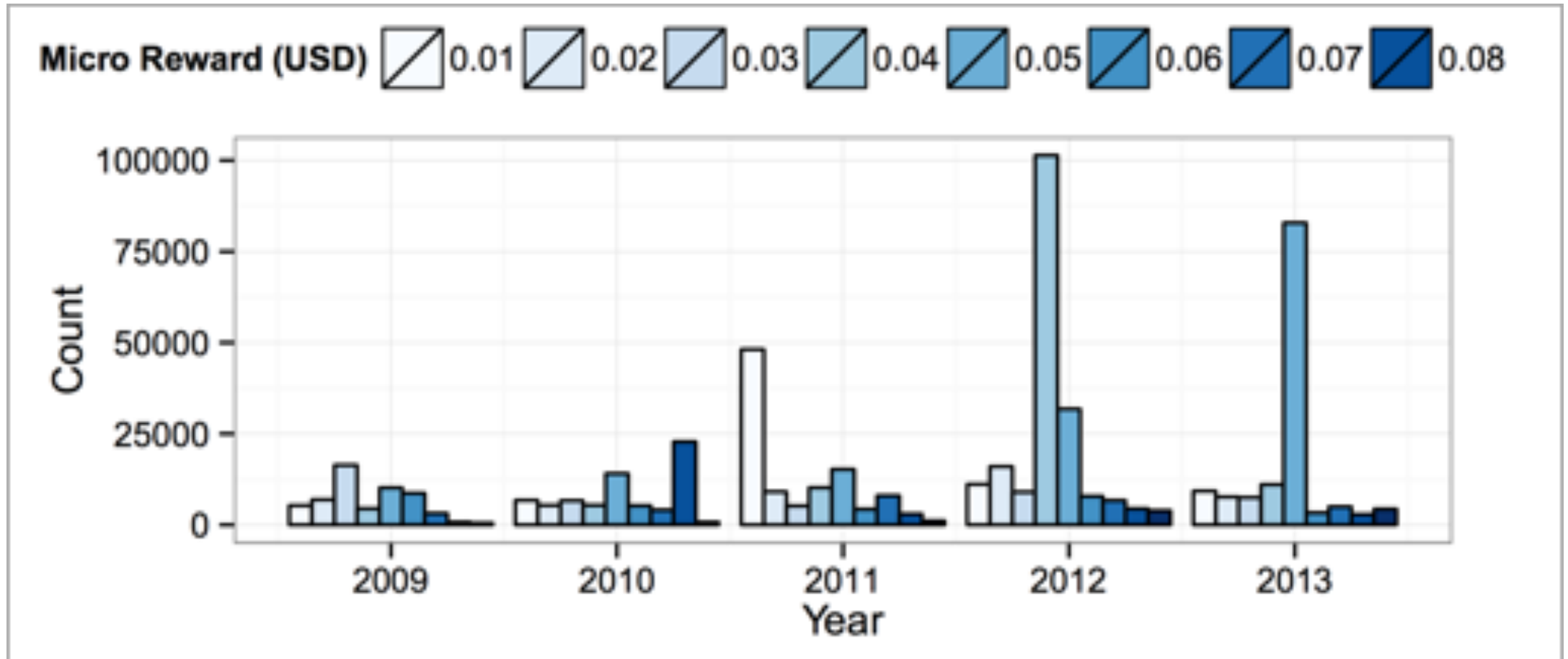Content Access: Decreasing
Other types: Increasing

Differences between
HITs and # Assignments Per HIT
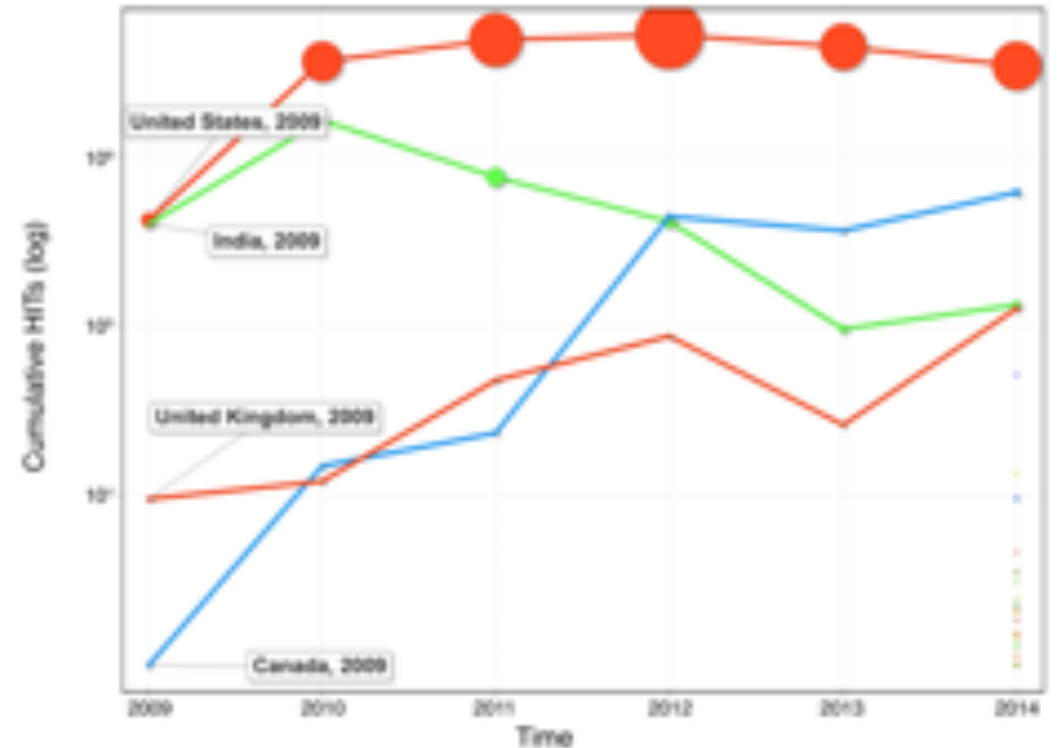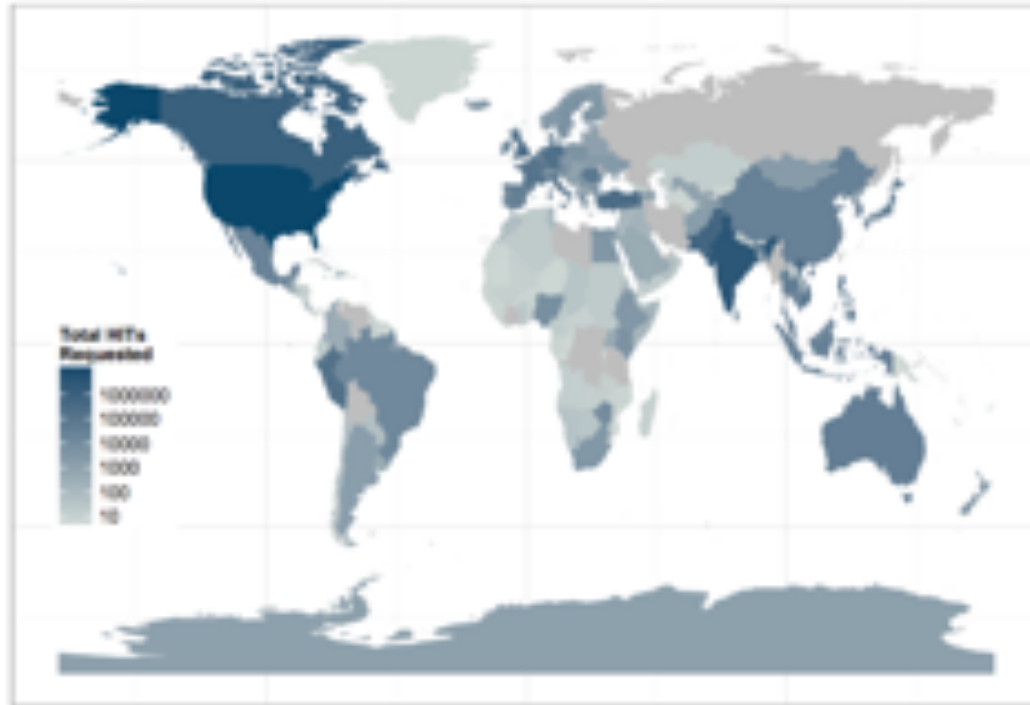
Dynamics of Task Topics

- Tasks of different topics have different "reward levels"

- Tasks related to "audio" and "transcription" increase substantially.

Dynamics of Task Rewards

Task rewards increase over time!

# Dynamics of Task Qualifications (Locations)

A substantial amount of tasks are exclusive to US workers, while tasks exclusive to India workers are decreasing over time!

Task Feature vs. Requester's Experience

How is the *throughout* for a batch of tasks posted by the requester influenced by task features?
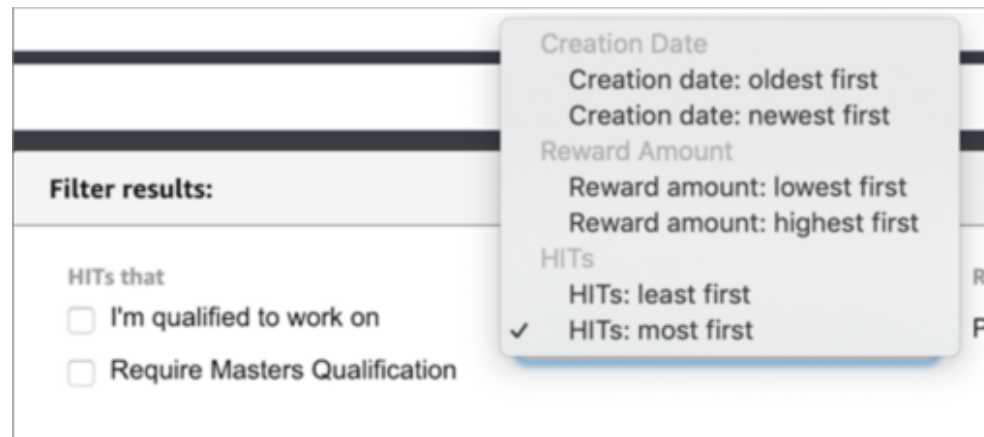
Run a regression with 29 features

Throughput Prediction Results
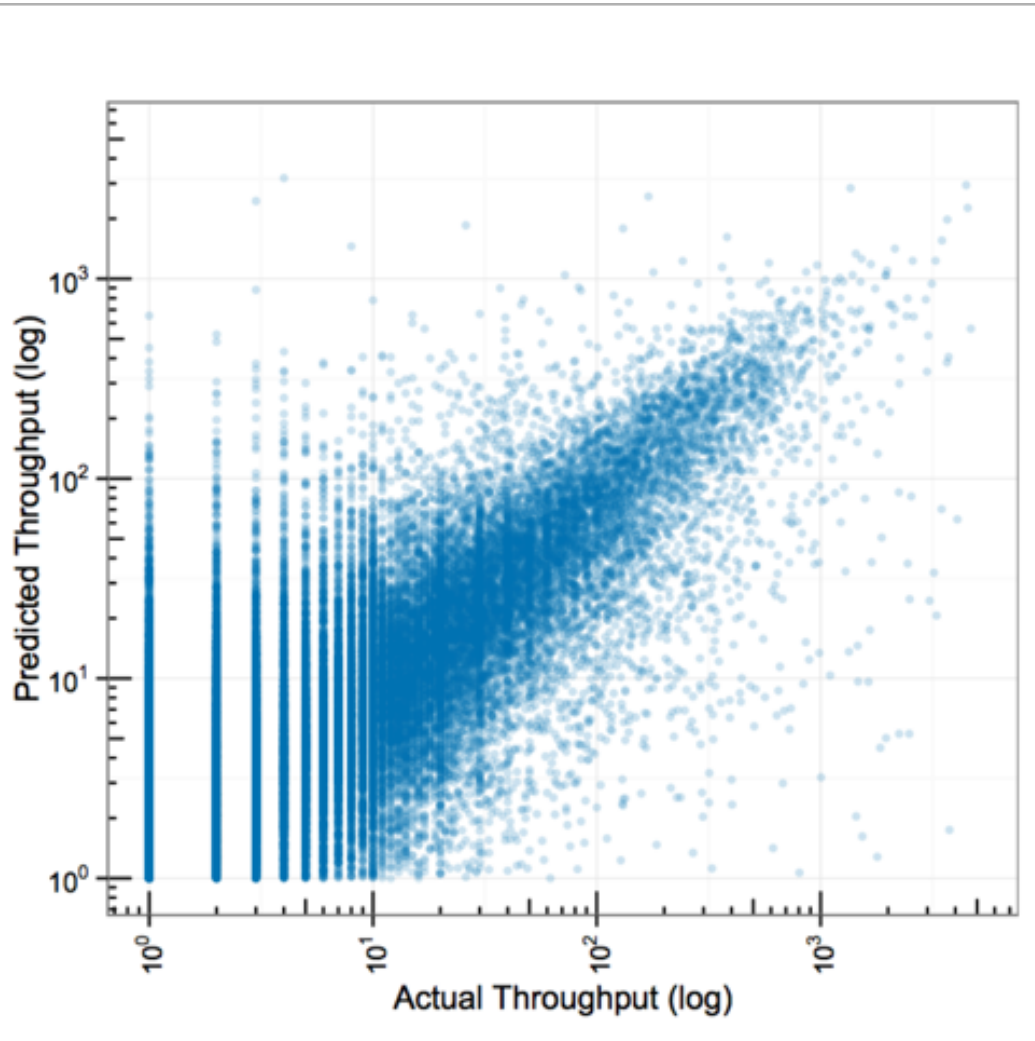
- More accurate when actual throughout is large

- Important features: # of tasks available, age of the task

- Why?

Implications of the Throughput Predictions?

- Important features: # of tasks available, age of the task
    - smaller context switch costs, etc.
   - More importantly, probably due to the MTurk interfaces in showing tasks

Creation Date
   Creation date: oldest first
   Creation date: newest first
Reward Amount
   Reward amount: lowest first
   Reward amount: highest first
HITs
   HITs: least first
✓ HITs: most first

Filter results:

HITs that
☐ I'm qualified to work on
☐ Require Masters Qualification

- Maybe implies the importance of task recommendation / task assignment
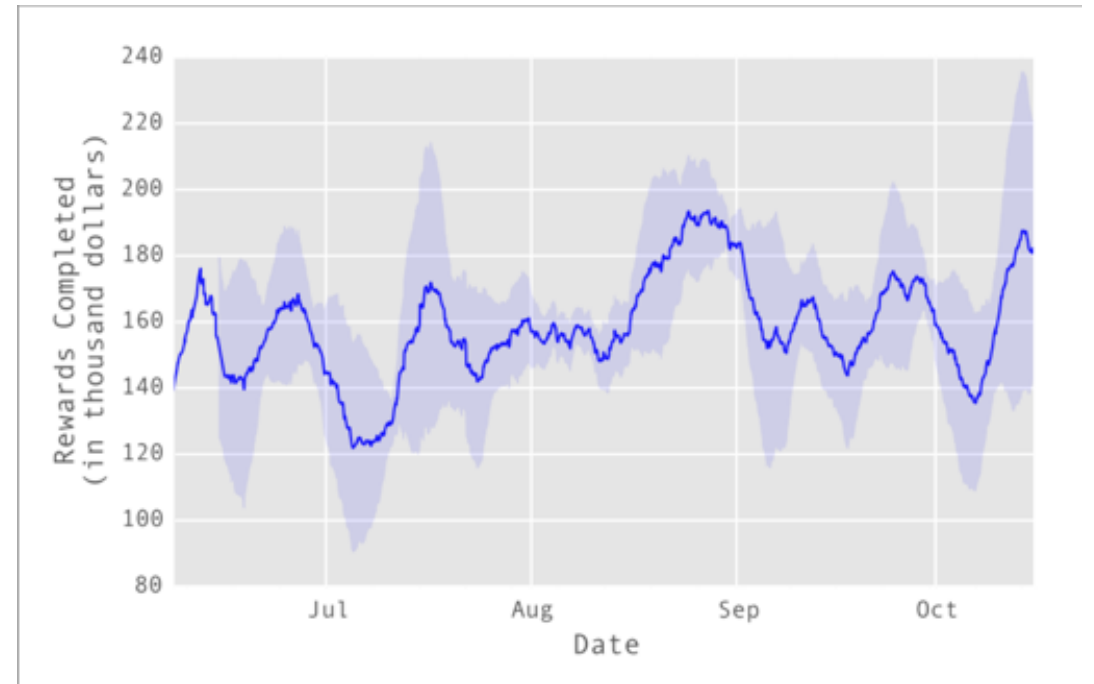
Market View: An Increasing Demand
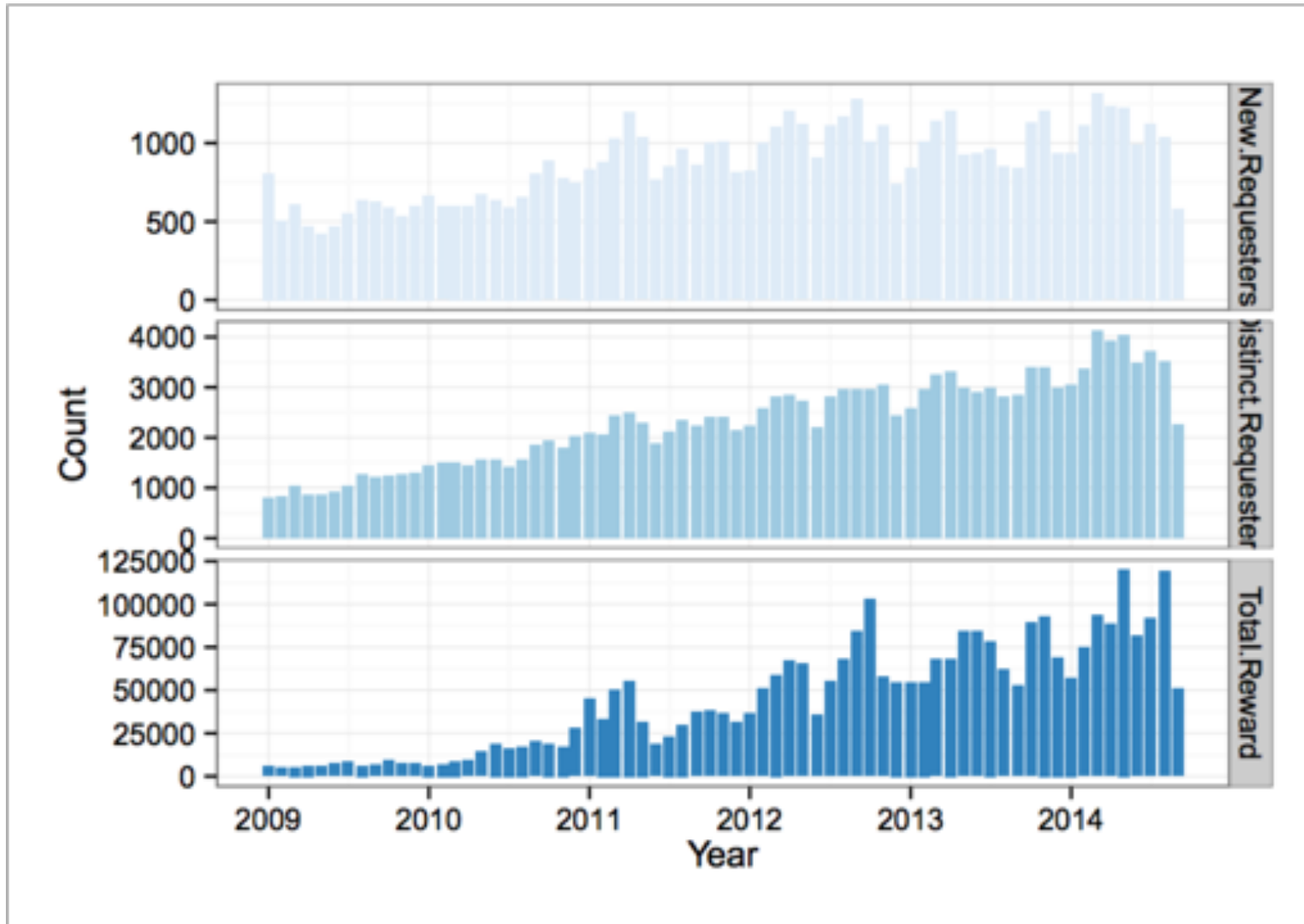
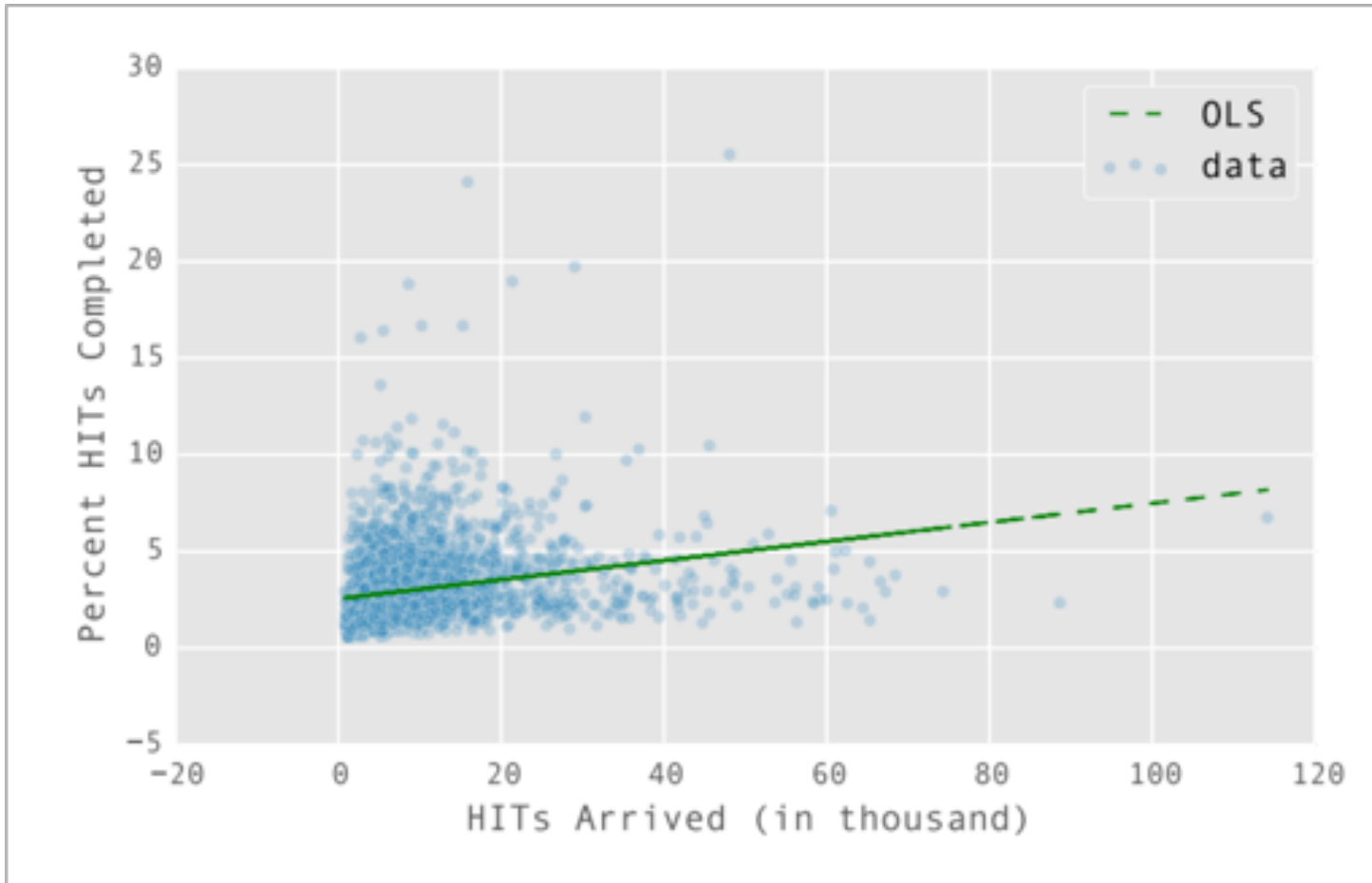# Market View: Periodicity in Supply and Demand
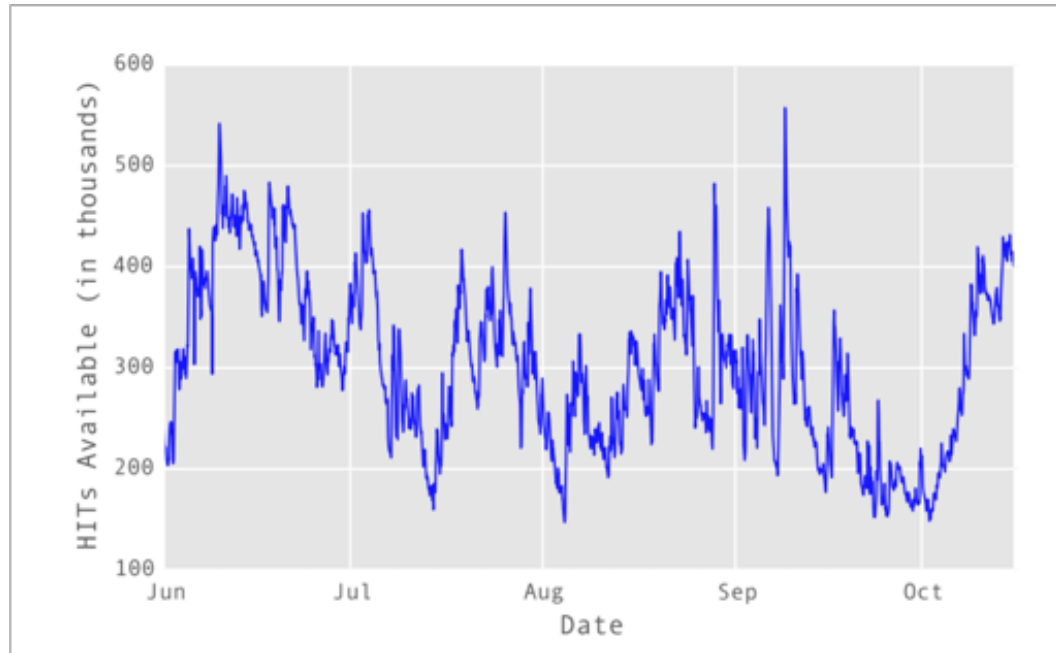
Demand

Supply

Market View: Supply Increases with Demand

- When the demand in an hour increases by 10K, the percentage of completed tasks increase by 0.5%.

Discussion

- If you had this data, what other analyses you want to conduct to better understand crowdsourcing tasks and markets?

- Do you need to collect some additional data for these analyses? If yes, what data do you need?

Beyond Microtask Platforms

- Fiverr: https://www.fiverr.com/
- Upwork: https://www.upwork.com/

- What are the questions you want to ask for understanding the dynamics of these crowdsourcing platforms (especially questions that are different from what you will ask for microtask platforms)?
- What data you would need to collect to answer these questions?