Imagine we have some data...

<table>
<thead>
<tr>
<th>gender</th>
<th>var1</th>
<th>var2</th>
<th>first</th>
<th>last</th>
<th>image</th>
</tr>
</thead>
<tbody>
<tr>
<td>female</td>
<td>0.5</td>
<td>1</td>
<td>sara</td>
<td>annala</td>
<td>img94.jpg</td>
</tr>
<tr>
<td>male</td>
<td>0.6</td>
<td>3</td>
<td>julius</td>
<td>haataja</td>
<td>img69.jpg</td>
</tr>
<tr>
<td>male</td>
<td>1.2</td>
<td>2</td>
<td>ross</td>
<td>meyer</td>
<td>img32.jpg</td>
</tr>
<tr>
<td>female</td>
<td>0.3</td>
<td>1</td>
<td>sarah</td>
<td>lahti</td>
<td>img96.jpg</td>
</tr>
<tr>
<td>female</td>
<td>1.1</td>
<td>5</td>
<td>ada</td>
<td>park</td>
<td>img24.jpg</td>
</tr>
<tr>
<td>female</td>
<td>0.9</td>
<td>2</td>
<td>joan</td>
<td>hernandez</td>
<td>img92.jpg</td>
</tr>
<tr>
<td>female</td>
<td>0.4</td>
<td>1</td>
<td>sofia</td>
<td>korhonen</td>
<td>img87.jpg</td>
</tr>
<tr>
<td>female</td>
<td>0.1</td>
<td>3</td>
<td>helle</td>
<td>kivela</td>
<td>img52.jpg</td>
</tr>
<tr>
<td>male</td>
<td>1.8</td>
<td>4</td>
<td>kasper</td>
<td>johnson</td>
<td>img17.jpg</td>
</tr>
<tr>
<td>male</td>
<td>0.6</td>
<td>2</td>
<td>dirk</td>
<td>luoma</td>
<td>img62.jpg</td>
</tr>
</tbody>
</table>

...but how do we analyze an image variable?
Coding
Manual
translation
Writing
tasks
Building
training sets

Data search/
retrieval/scraping
Content
moderation
Human
subjects
research

UX testing

Categorization
Audio/Video
Transcription
Massively Parallel Human Intelligence Ideal Case for Crowdsourcing
Data Need

Design Data Entry Form

Create HIT(s)

Assignment

Assignment

Assignment

Assignment

Assignment

R

HTML

MTurk

Analyze data

Review
Please look at the following picture:

How **competent** is the person in this photo?

- [ ] 0 extremely incompetent
- [ ] 1
- [ ] 2
- [ ] 3
- [ ] 4
- [ ] 5
- [ ] 6
- [ ] 7
- [ ] 8
- [ ] 9
- [ ] 10 extremely competent

Do you recognize the person in this photo?  [ ] Yes  [ ] No

*If yes, who is it?* __________

[Submit]  [The image did not appear]
# set API keys in environment variables

library("MTurkR")

BulkCreateFromURLs(
  url = paste0("https://example.com/", 1:10, ".html"),
  title = "Image Categorization",
  description = "Describe contents of an image",
  keywords = "categorization, image",
  reward = .01,
  duration = seconds(minutes = 5),
  annotation = "My Project",
  expiration = seconds(days = 4),
  auto.approval.delay = seconds(days = 1)
)
Get back a data.frame:

GetAssignments((annotation = "My Project")

The image coding task with 27,500 images took 225 workers about 75 minutes and cost $412.50

Pay workers with:
ApproveAssignments((annotation = "My Project")
a = GenerateHTMLQuestion(file = "hit.html")

hit = CreateHIT(
    title = "Short Survey",
    description = "5 question survey",
    keywords = "survey, questionnaire",
    duration = seconds(hours = 1)
    reward = .10,
    assignments = 5000,
    expiration = seconds(days = 4),
    question = a$string,
)
GetHIT(hit$HITId)

ExtendHIT(hit$HITId, 
    add.assignments = 500)  
    add.seconds = seconds(days = 1)
)

ExpireHIT(hit$HITId)

ChangeHITType(hit$HITId, 
    title = "New, better title",  
    reward = 5.00
)
**Advanced Features**

- Choose who works for you ⇒ Qualifications and tests
- Monitor HITs ⇒ Notifications
- Sanction and reward workers ⇒ Qualifications, bonuses, and blocks
- Automatic review ⇒ Review Policies
Anatomy of an MTurkR App

CreateHIT() (with Review Policies)

Assignment

Check Known Answer(s)

Reject

Approve

Compare w/ Other Assignments

GetReviewResults()
What’s next?

1. Packages for more crowdsourcing platforms
   - Common interface?

2. HIT templates

3. Performance improvements
# Start Crowdsourcing

# CRAN
install.packages("MTurkR")

# GitHub
install_github("leeper/MTurkR")

# Questions?
# thosjleeper@gmail.com
# https://github.com/leeper/MTurkR/wiki