Getting Started with MTurk Data Collection

Tutorial Slides

Nandana Sengupta, Postdoc @ Knowledge Lab

March 2, 2017

Introduction

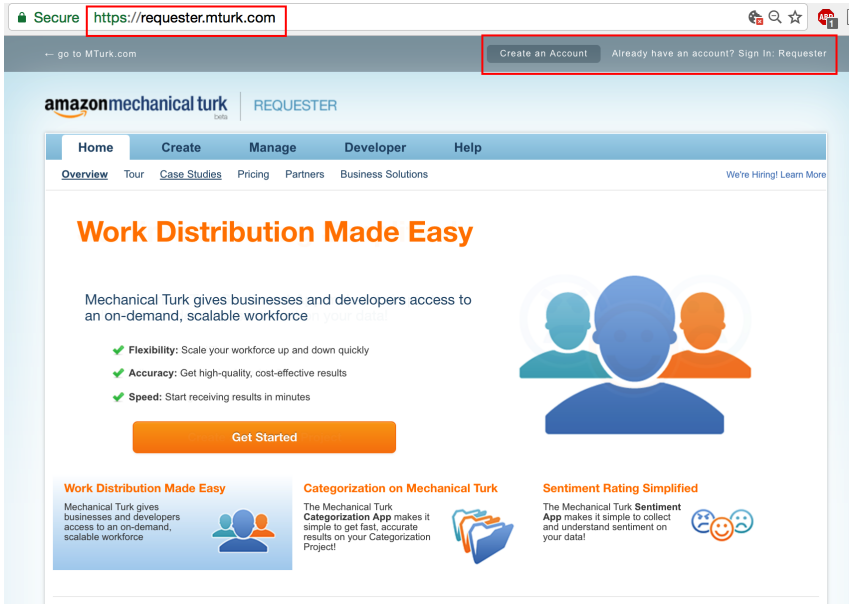
- What is M(echanical) Turk?
  - MTurk is a crowdsourcing internet marketplace. *Requesters* get access to an on-demand workforce (*workers*) to perform surveys and tasks which computers are unable to do.
- Is this tutorial for Workers or Requesters?
  - This tutorial is for MTurk Requesters who want to collect and analyze crowdsourced data.
- What is a HIT?
  - HIT stands for Human Intelligence Task – these are the tasks requesters design and workers complete.
- In this Tutorial:
  - Step by Step MTurk Requester Guide.
  - Some things to keep in mind while using MTurk.
  - Appendix: Simple HTML code for MTurk.

Step by Step MTurk Tutorial

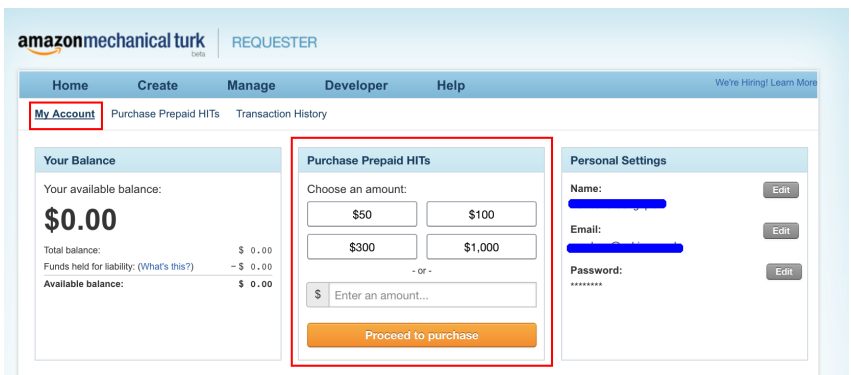
- Step 1: Setup a Requester Account
- Step 2: Purchase Prepaid HITS
- Step 3: Create a new survey
  - Step 3a: HIT properties
  - Step 3b: Design Layout
  - Step 3c: Preview and Finish
- Step 4: Publish HIT
- Step 5: Track Progress and Manage Results
- Step 6: Approve Workers and Download Results

Step by Step MTurk Tutorial

- Step 1: Setup a Requester Account on requester.mturk.com.

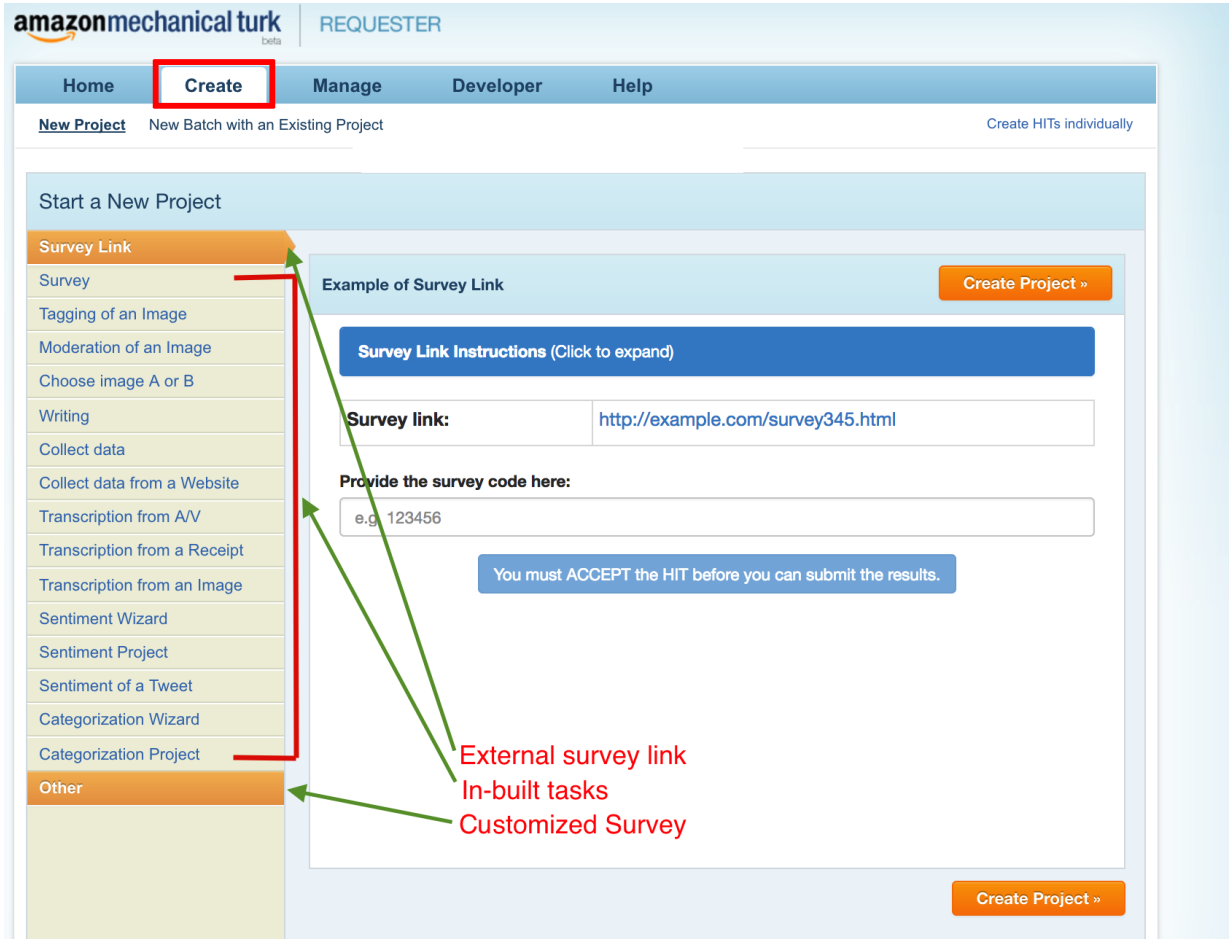


- Step 2: Follow instructions on the 'My Account' tab to Purchase Prepaid HITS.



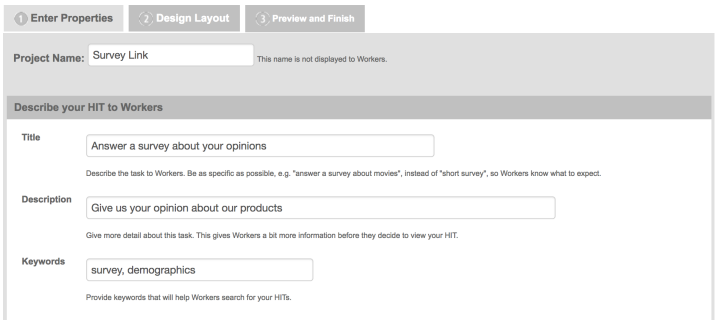
Step by Step MTurk Tutorial

- Step 3: For creating a new HIT click on the 'Create' tab. MTurk provides three types of options:
  - External Survey Link (direct to an app or SurveyMonkey, Qualtrics survey).
  - Customizable in-built HIT options.
  - Customized HITS by changing HTML source code.

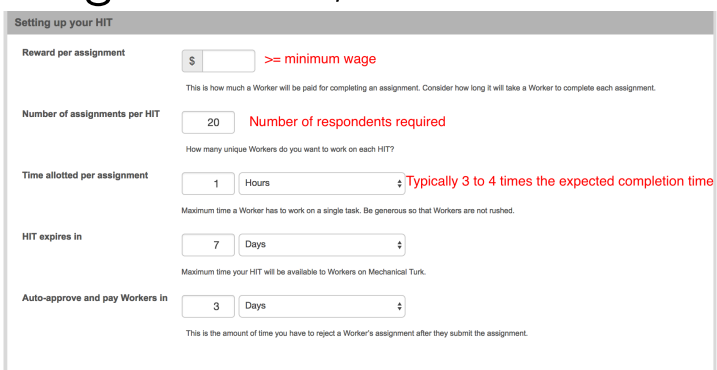


Step by Step MTurk Tutorial

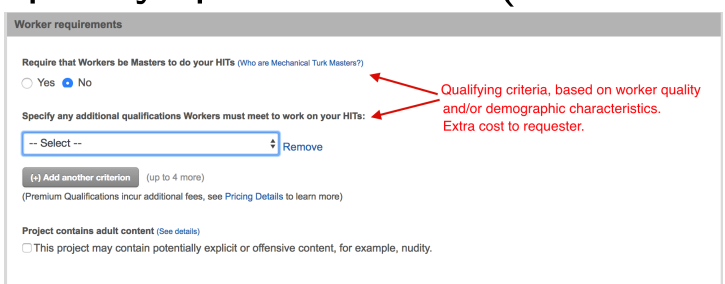
- Step 3a: Once you choose the type of HIT to create, you will be directed to the 'Properties' tab to
  - provide instructions to workers, name the project etc;



  - assign rewards, select number of workers etc;

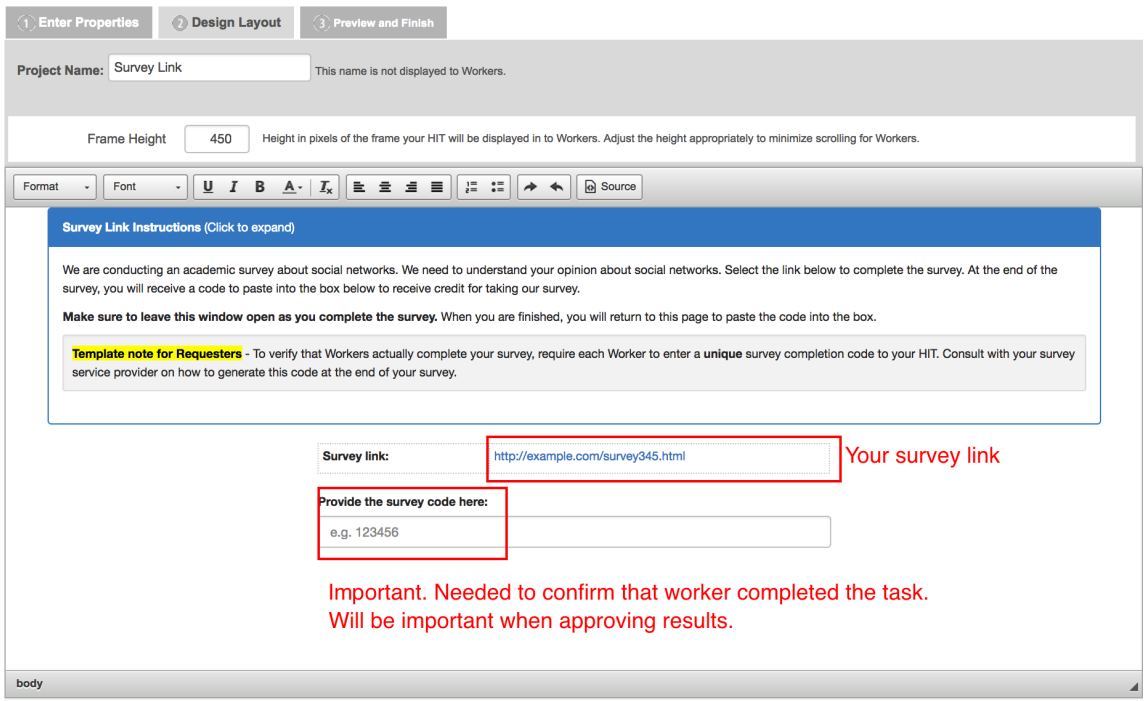


  - specify qualifications (for an added cost).



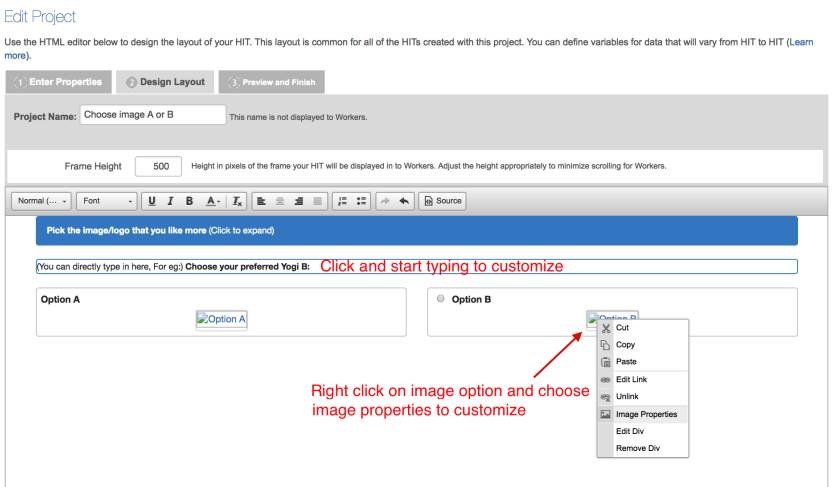
Step by Step MTurk Tutorial

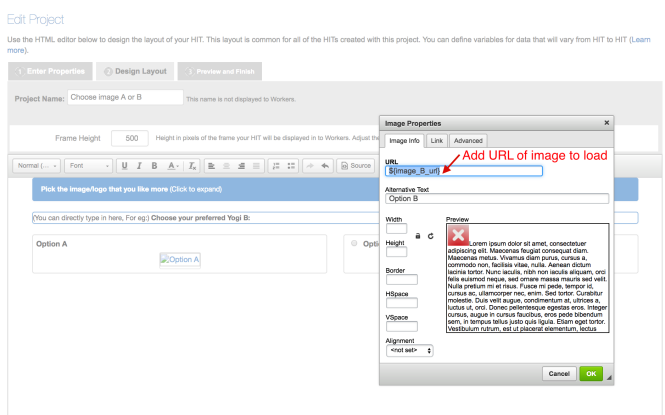
- Step 3b: The next tab is the Design Layout.
- For external survey links the main thing to keep in mind is that the external link provides the worker a survey code (preferably random) at the end of the task.
- With an external link MTurk cannot ensure that the worker completes the task. This is checked by the requester by matching the survey codes once the results are in.



Step by Step MTurk Tutorial

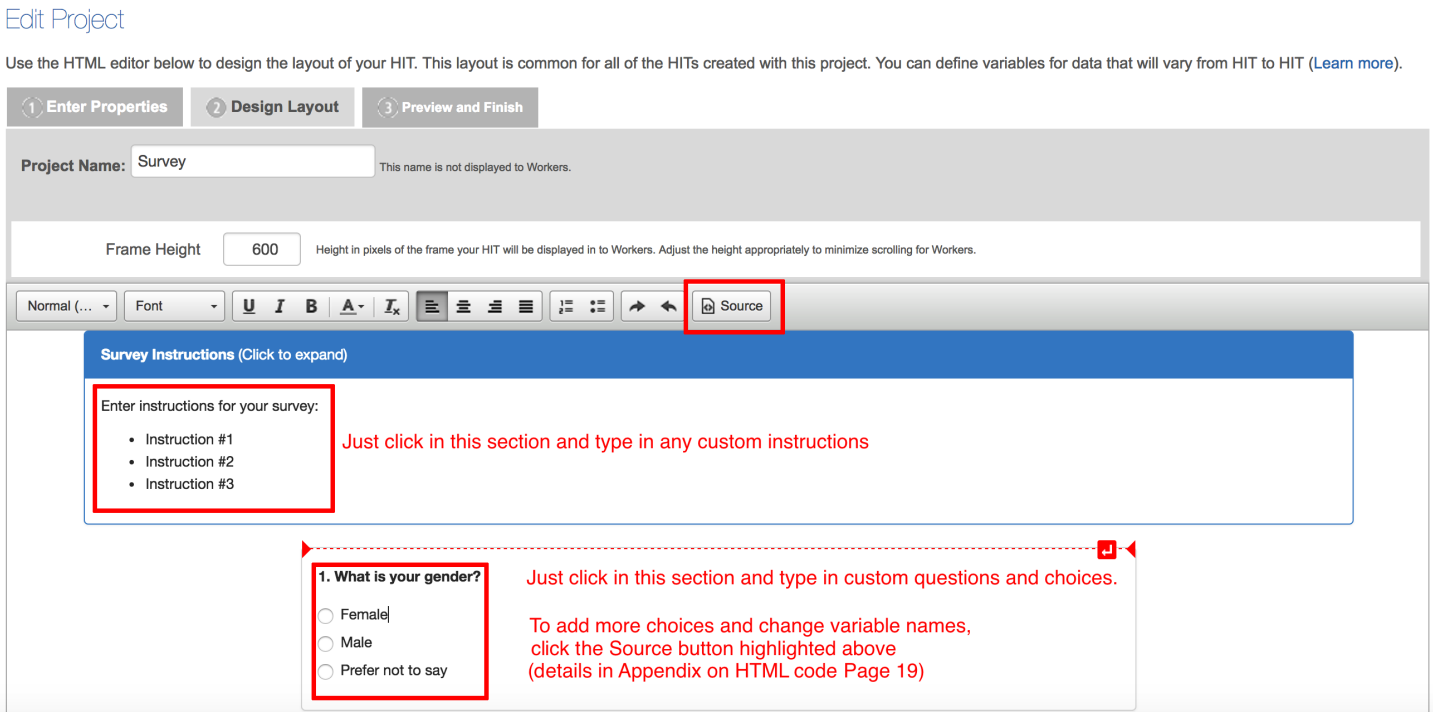
- Step 3b: Inbuilt HITs are customizable in the Design Layout tab without going into source code.
- These customizations are mostly straightforward. The screenshots below provide an example using the inbuilt HIT 'Choose Image A or B'.





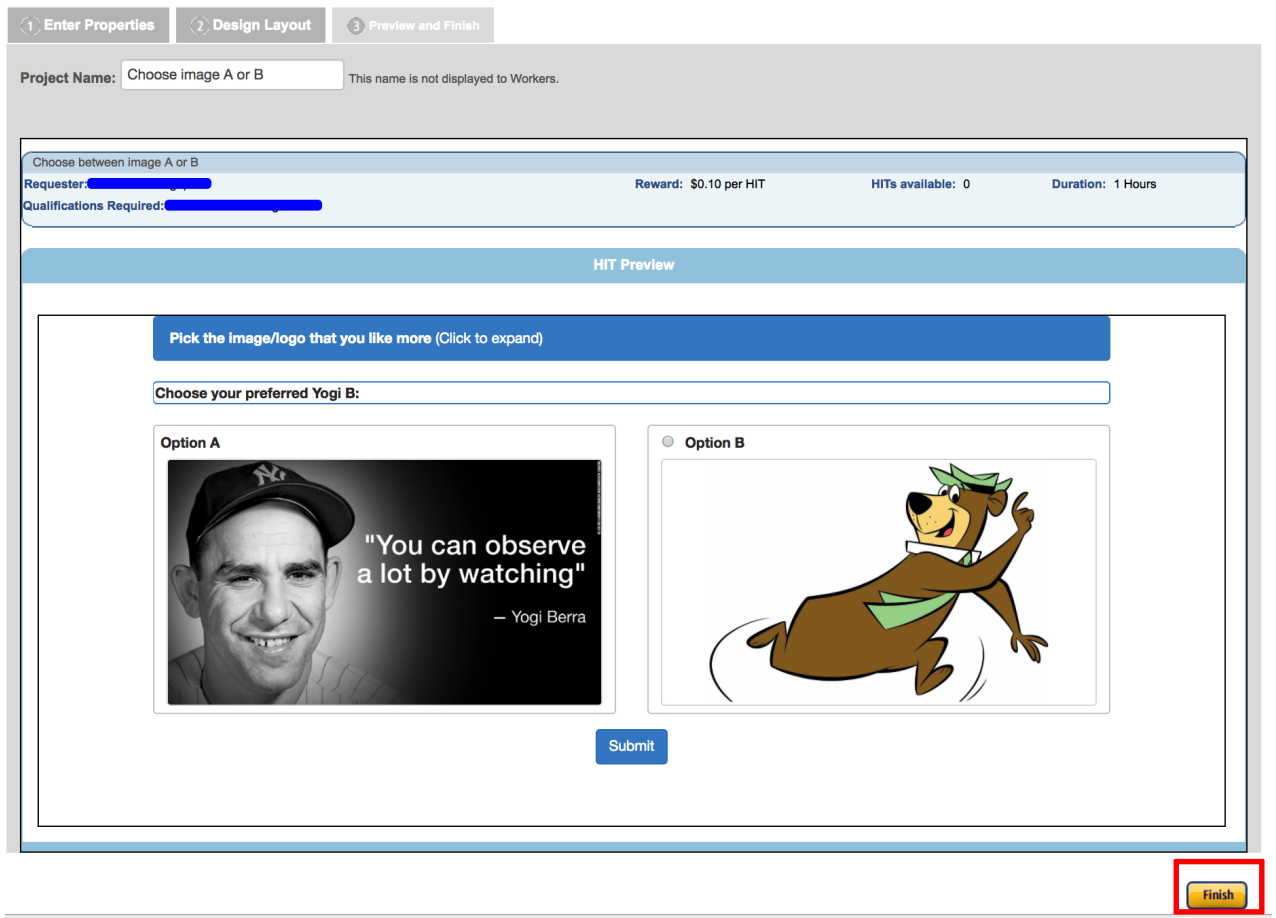
Step by Step MTurk Tutorial

- Step 3b: Inbuilt HITs are customizable in the Design Layout tab without going into source code.
- These customizations are mostly straightforward. Another example is multiple choice questions in the inbuilt HIT 'Survey' very easily customizable within the Design Layout tab, screenshot provided below.



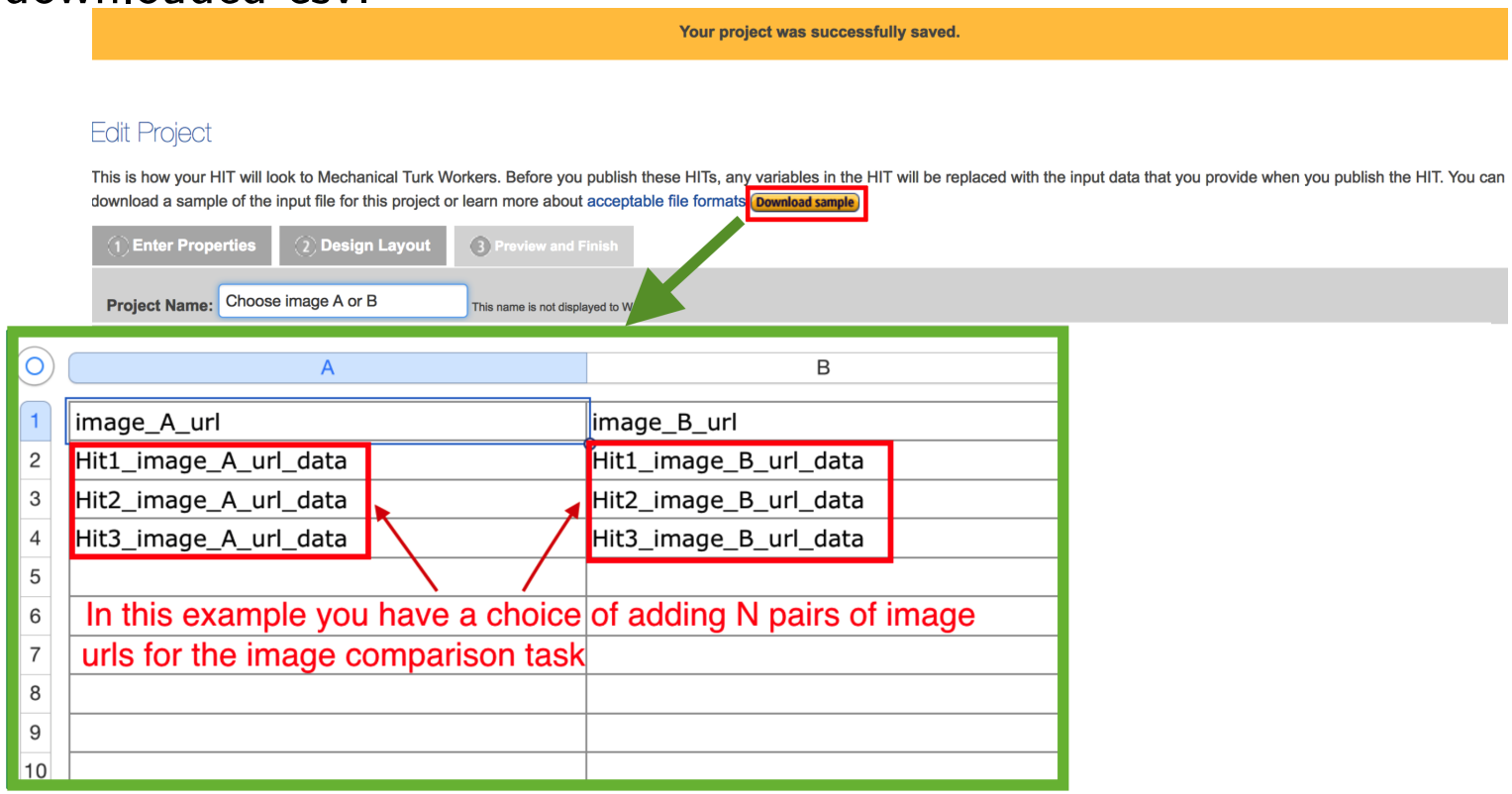
Step by Step MTurk Tutorial

- Step 3c: Once you have finalized the HIT properties and Design Layout, the last tab requires you to 'Preview and Finish' (indicating you are happy with the HIT you designed).



Step by Step MTurk Tutorial

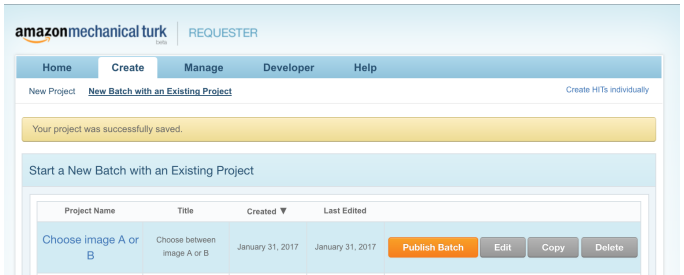
- Step 3c: At this stage, some inbuilt HITs allow you to download a sample csv file. For bulk tasks (like comparing large number of images) you may prepare your input data based on the fields in the downloaded csv.



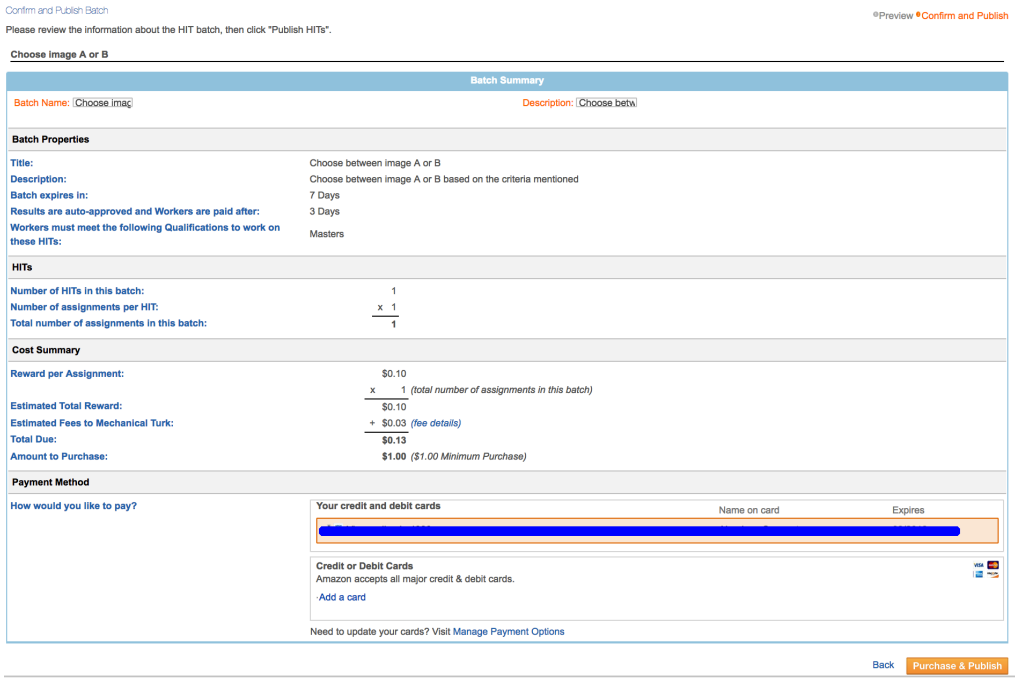
- To change variable names, you will need to go into the source code in the Design Layout tab. This is explained in detail in the Appendix on Page 19.

Step by Step MTurk Tutorial

- Step 4: Once you have created the HIT it shows up under the 'Create' Tab with options to Publish, Edit, Copy or Delete.



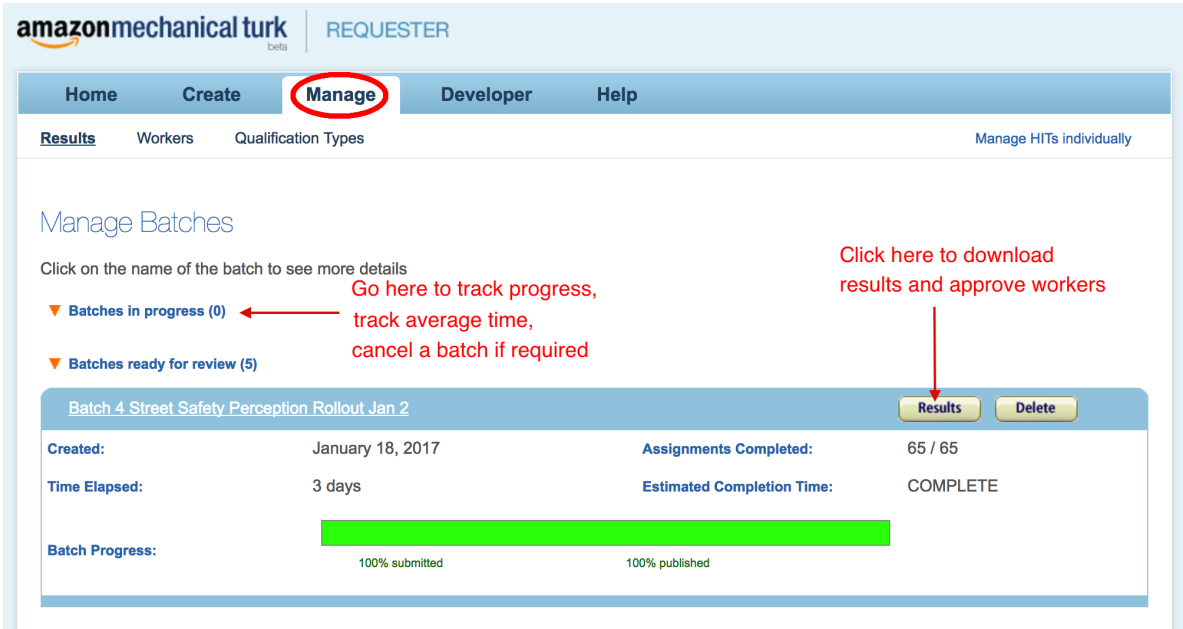
- On selecting 'Publish Batch', you will be directed to a payments page where the fee details and payment details will be confirmed.



- Click 'Purchase and Publish' to make HIT available to workers.

Step by Step MTurk Tutorial

- Step 5: Once the HIT is live you can track its progress under the 'Manage' Tab.
- This is also where completed batches of HITs (i.e. Batches for Review) can be accessed.



- Step 6: Once the HIT batch is completed it is ready for review. Click on 'Results' to Approve or Reject workers, Download CSV files etc. This can also be done before the batch is completed.

Step by Step MTurk Tutorial

Your crowdsourced data is now available for analysis!!!

Manage Batches > **Review Results**

Review Results

Select the check boxes on the left to approve or reject results. You only pay for approved results. To evaluate results offline, select Download CSV.

For additional batch information, view batch details.

[redacted]

Customize View | Filter Results | Upload CSV | Approve All | Download CSV

10 of 10 assignments (FILTER APPLIED: only show assignments that are in 'Approved' status)

Approve | Reject

HIT ID ▲	Worker ID	Lifetime Approval Rate	Tag4	Surveycode
3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	[redacted]	100% (1/1)	yes	xWoshbOy1HAXrqq0kC6PynQ2YR9VEF
3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	[redacted]	100% (2/2)	yes	gitAi0SWP2G1TqejyZr46N4VKgi4ME
3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	[redacted]	100% (1/1)	yes	IG7ZH6t3rrOwo77LP8r5pq0m6jvsai
3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	[redacted]	100% (1/1)	yes	6eZq4ylCB2IadieVaTVoLJXVrQ0KOt
3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	[redacted]	100% (1/1)	yes	7H2fkknL54wnukvTdVSNsI7TchgD1y
3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	[redacted]	100% (3/3)	yes	to6W7NzHoFYWksIDHe4VuOieTqaN6Z
3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	[redacted]	100% (2/2)	yes	XQVDhQMLZN9cGslekNHv2Qt0xTvvwv
3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	[redacted]	100% (2/2)	yes	IT0kl943Pwl5ffura3lREoqwbWO68m
3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	[redacted]	100% (2/2)	yes	7FUD2mSWNZi7qBUNK7rx5TOknxqown
3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	[redacted]	100% (1/1)	yes	YRJjphLymfMPvlQnuX5Jx0Xms3BsuT
HIT ID ▲	Worker ID	Lifetime Approval Rate	Tag4	Surveycode

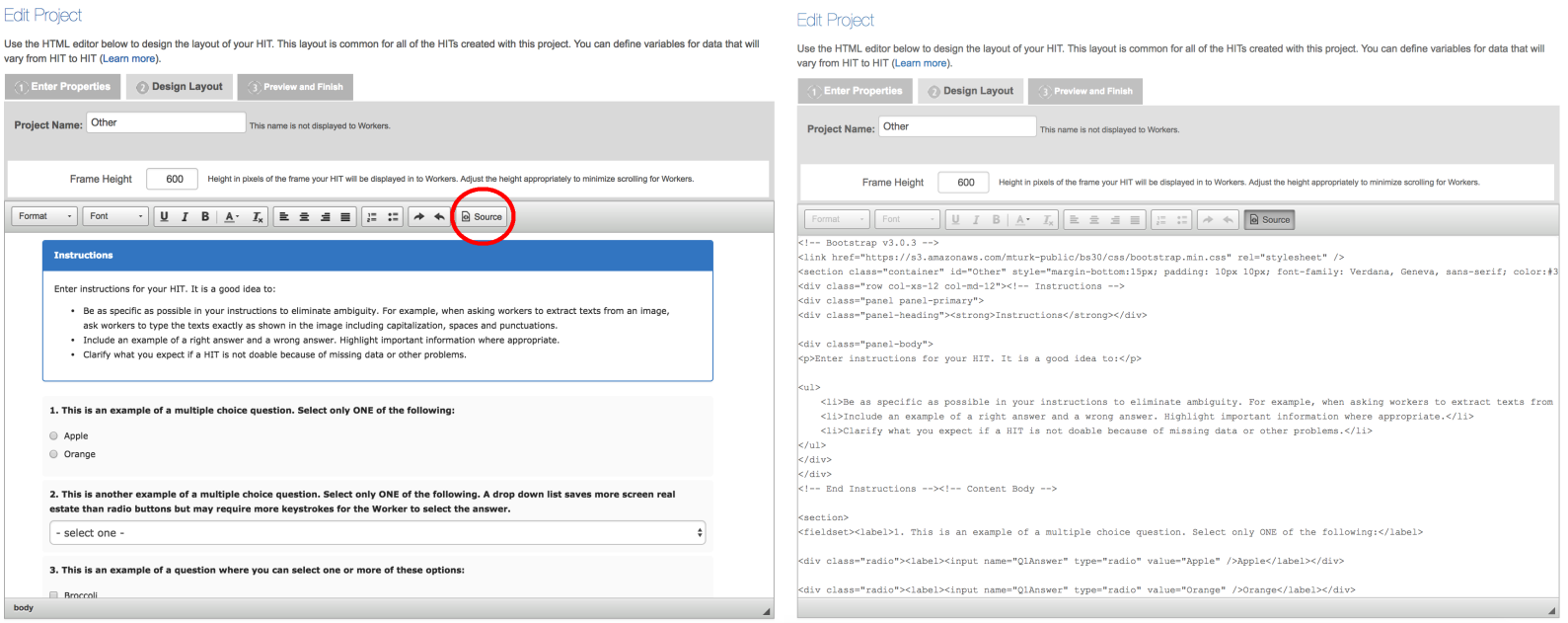
Approve | Reject

Some things to keep in mind when usinh MTurk

- For those interested in trying out different MTurk functionalities without conducting actual experiments, a useful site to checkout is requestersandbox.mturk.com.
- Make sure your HIT is well-designed and not buggy. Also make sure compensation is fair. There are a number of MTurk worker discussion boards where you don't want a bad reputation as a requester!
- While crowdsourcing is an efficient and cost-effective way of collecting data, it may not be suitable for all research questions. For instance, there is a large chance of selection bias in the data (i.e. the workers are not randomly distributed across geographies, ages, income etc.). Make sure crowdsourced data collection is appropriate for your research question.
- Finally, here is an interesting Pew Center report on Crowdsourcing and MTurk if you are interested in knowing more about the profile of MTurk Workers, Requesters and Tasks.

Appendix: Simple HTML code for MTurk

Simple HTML code for MTurk

- For designing custom HITs, its useful to know simple HTML code.
- Source code is the code (written in HTML language) used to generate a web page on a browser – which can be changed by the requester to customize the webpage.
- When creating a HIT, in the Design Layout tab clicking the source code button reveals the source code for the page.
- 

- In the next few slides we'll go over some simple HTML code to help you customize your HITs.

Simple HTML code for MTurk

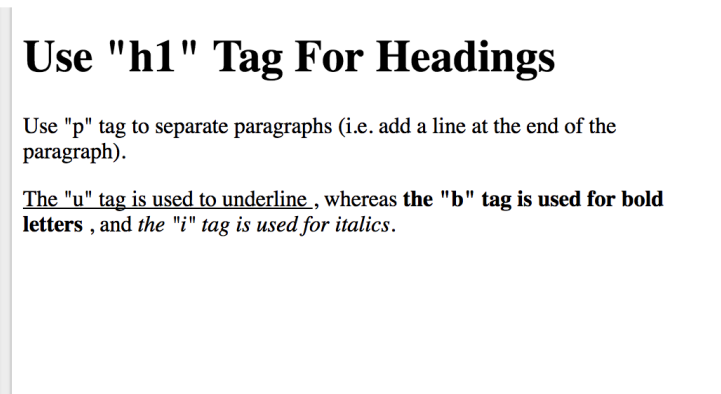
HTML formatting basics

- Basic formatting commands include a prefix and a suffix, for eg:
  - $< p > \cdots < /p >$: paragraph,
  - $< br >$: line break,
  - $< b > \cdots < /b >$: bold font,
  - $< u > \cdots < /u >$: underlined font,
  - $< i > \cdots < /i >$: italic font.
- You can learn more basics here.
- An example:

Source Code

```
<!DOCTYPE html>
<html>
<body>

<h1>Use "h1" Tag For Headings</h1>

<p>Use "p" tag to separate paragraphs (i.e. add a line
at the end of the paragraph).</p>

<p> <u> The "u" tag is used to underline </u>, whereas
<b> the "b" tag is used for bold letters </b>, and <i>
the "i" tag is used for italics. </i> </p>

</body>
</html>
```

Webpage Output



Use "h1" Tag For Headings

Use "p" tag to separate paragraphs (i.e. add a line at the end of the paragraph).

The "u" tag is used to underline , whereas **the "b" tag is used for bold letters** , and *the "i" tag is used for italics*.

Simple HTML code for MTurk

- The next step is setting up forms for user input
  - $<$ form $>$ $\cdots$ $<$ /form $>$: creates a form for user input.
  - The input type defines different types of input elements, eg: text fields, checkboxes, radio buttons etc.
  - The name attribute assigns the variable name, the value attribute assigns pre-determined values for multiple choice questions.
  - Some examples next. You can use these as a base and change the source code in the Design Layout (described in Step 3b on Pages 7-9) to make your own custom HITs.
- Text, number and textbox input

```
<!DOCTYPE html>
<html>
<body>

<p> Here are two input types, the first is text and the
second is a number. </p>

<form>
<b>Describe the weather in one word:</b>
<input type = "text" name = "weather">
<br>
<b>Guess the temperature in Celsius: </b>
<input type = "number" name = "temp">
</form>

<br>
<p> Next is an example of an open-ended question. Use
this when you need Workers to enter multi-line texts. Try
varying cols and rows to see what happens.</p>

<form>
<b>Can you describe the weather in some more detail? </b>
<br>

<textarea  cols="80" name="MultiLineTextInput"
rows="5"></textarea>

</form>
```

Here are two input types, the first is text and the second is a number.

Describe the weather in one word:
Guess the temperature in Celsius:

Next is an example of an open-ended question. Use this when you need Workers to enter multi-line texts. Try varying cols and rows to see what happens.

Can you describe the weather in some more detail?

Simple HTML code for MTurk

- Multiple choice with radio button (one pre-checked)

```
<!DOCTYPE html>
<html>
<body>
<p> <b> What's your favorite pastime? </b> </p>
<form>

  <input type="radio" name="gender" value="Netflix">
Watching Netflix <br>

<input type="radio" name="gender" value="Books">
Reading books <br>

  <input type="radio" name="gender" value="MTurk">
Making MTurk Surveys <br>

  <input type="radio" name="gender" value="Food"> FOOD
<br>

<input type="radio" name="gender" value="Other"
checked> Other <br>
</form>

</body>
</html>
```

What's your favorite pastime?

- ○ Watching Netflix
- ○ Reading books
- ○ Making MTurk Surveys
- ○ FOOD
- ◉ Other

- Multiple choice with checkboxes

```
<!DOCTYPE html>
<html>
<body>

<p> Ready to start running your own customized Mturk
surveys? </p>
<form>
<input type="checkbox" name="ready" value="yes"> Yes
<br>
<input type="checkbox" name="ready" value="no"> No
<br>
<input type="checkbox" name="ready" value="maybe"> Maybe so
<br>
</form>

</body>
</html>
```

Ready to start running your own customized Mturk surveys?

- ☐ Yes
- ☐ No
- ☐ Maybe so

Thank you!!