

# 1 Exploratory Data Analysis

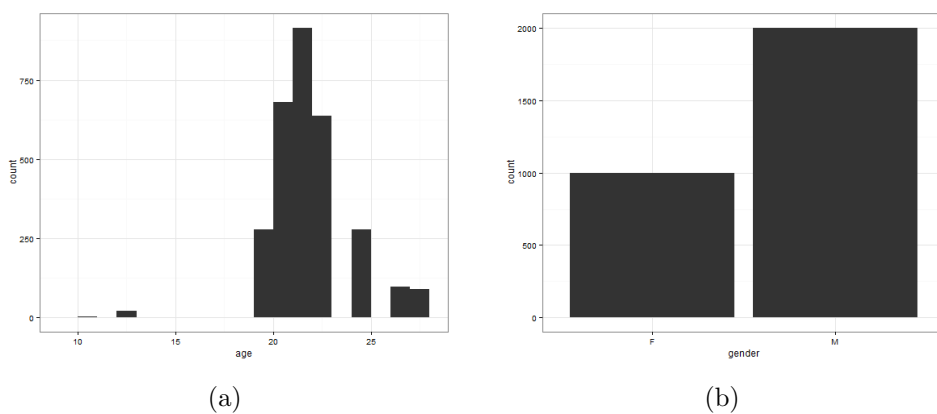


Figure 1: Histogram of age and gender of participants.

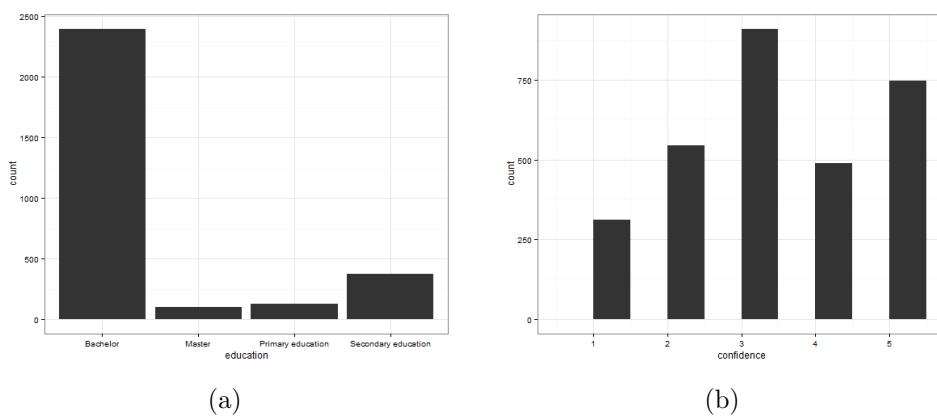


Figure 2: Histogram of education level and confidence level.

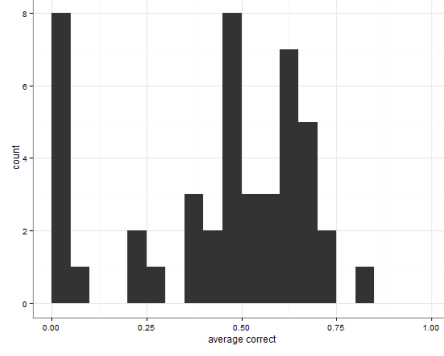


Figure 3: Histogram of user performance on multiple choice tasks.

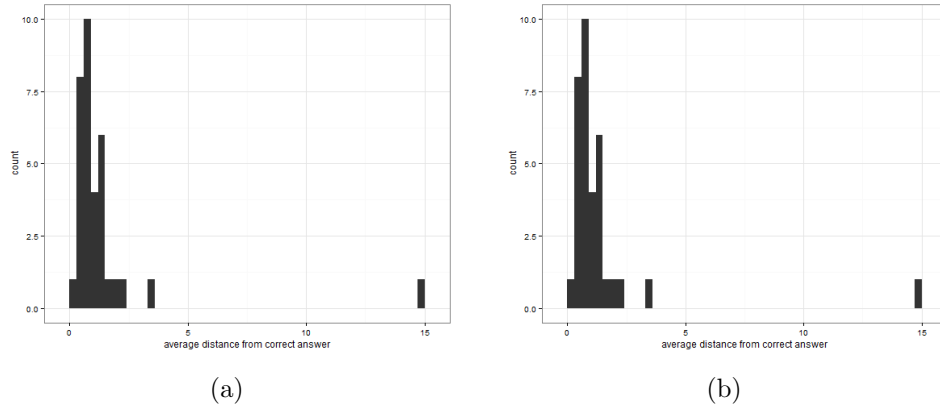


Figure 4: Histogram of aggregate user performance for point estimate tasks.

	type	av_score
1	audio	0.10
2	image	0.71
3	video	0.65

Table 1: Average score by asset type for all domains.

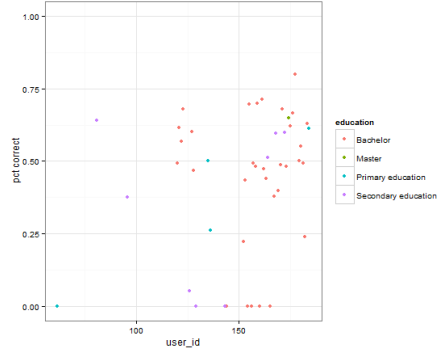


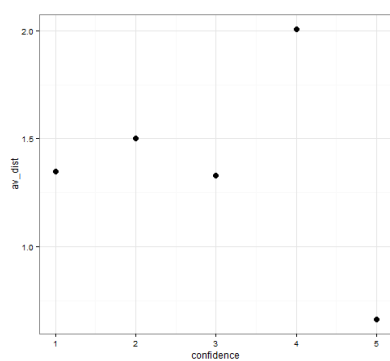
Figure 5: Scatter plot of accuracy by education level on multiple choice tasks.

	education	pct_correct	av_dist
1	Bachelor	0.42	1.46
2	Master	0.52	1.28
3	Primary education	0.45	1.22
4	Secondary education	0.40	1.42

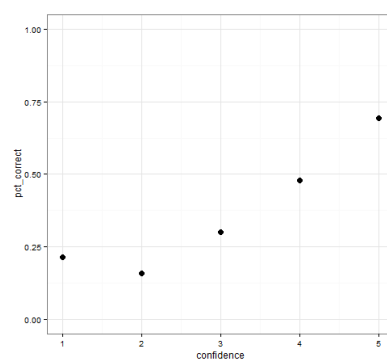
Table 2: Average score by education level all domains.

	confidence	pct_correct	av_dist
1	1	0.23	1.35
2	2	0.17	1.50
3	3	0.31	1.33
4	4	0.52	2.01
5	5	0.75	0.66

Table 3: Average score by confidence level all domains.



(a)



(b)

Figure 6: Scatter plot of percent accuracy by reported confidence level for point estimate and multiple choice questions.

## 2 Crowd Rankings

domain_id	domain_name	crowd_score	crowd_rank
3	MagicTrick	20.00	1.00
7	Landmarks	18.00	0.92
12	Penalties	12.00	0.82
53	MovieSong	13.00	1.00

Table 4: Multiple Choice domains. The Crowd Ranking column contains the percentage of users the crowd performs better than. Score is the number of answers the crowd got right.

domain	mean	median	geom.mean	trunc.mean	trunc.geom.mean
Calories	0.45	0.68	0.63	0.58	0.78

Table 5: Point estimate rankings by domain according to average ranking. Columns represent the ranking by using the corresponding method of aggregation. Average ranking takes the mean of the crowd percentiles for each task.