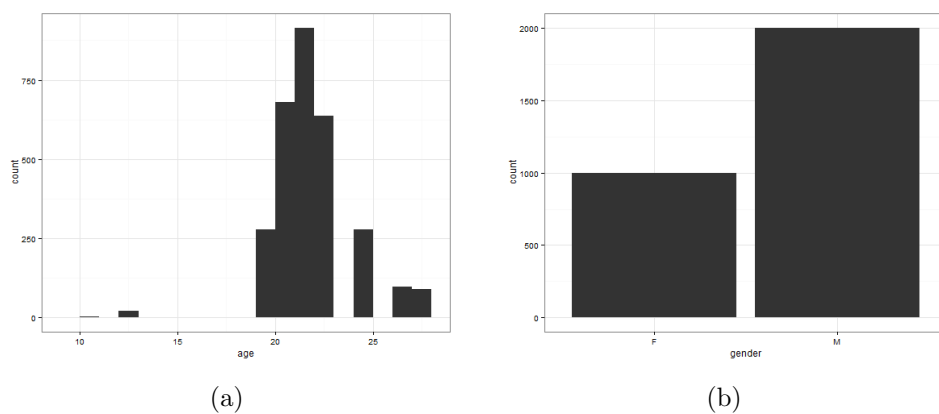# 1 Exploratory Data Analysis

Figure 1: Histogram of age and gender of participants.
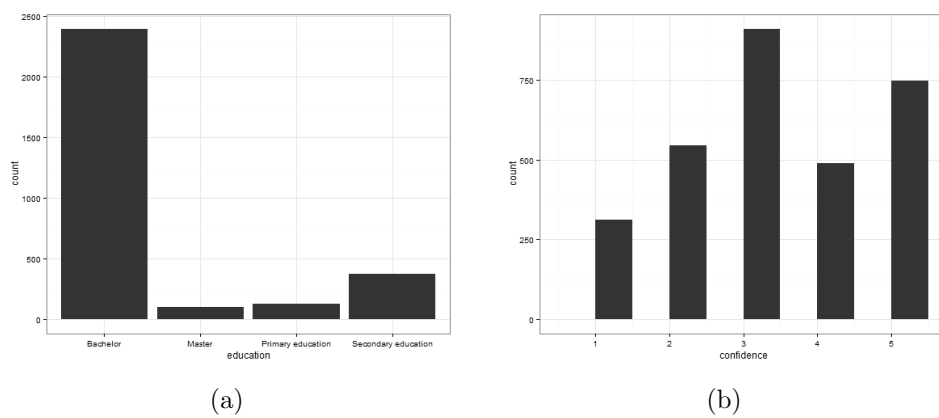

Figure 2: Histogram of education level and confidence level.
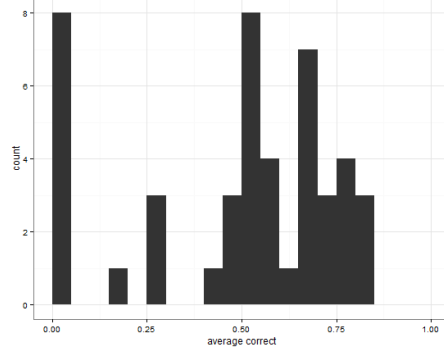
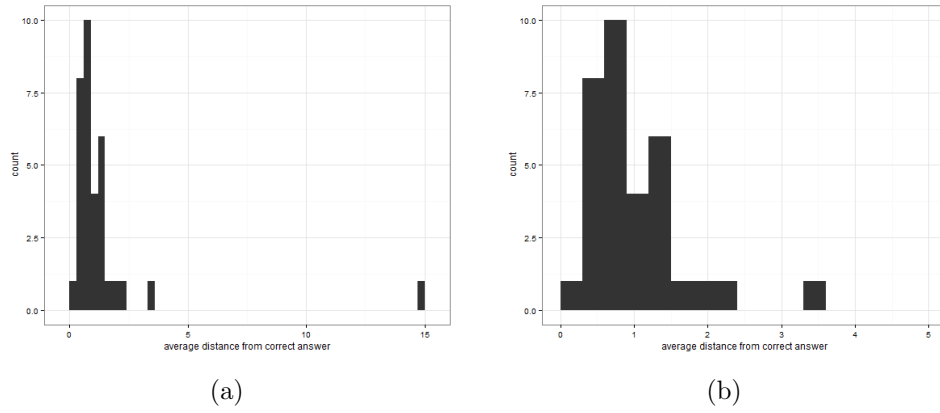Figure 3: Histogram of user performance on multiple choice tasks.

(a)

(b)

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

type	av_score
audio	0.39
image	0.71
video	0.65

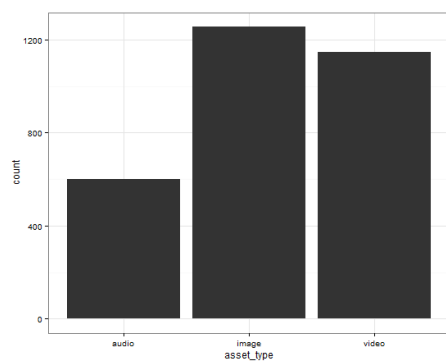Table 1: Average score by asset type for all domains.

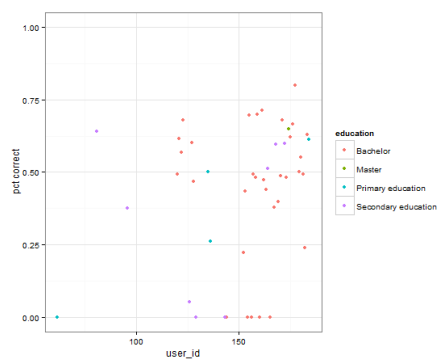Figure 5: Histogram of user performance on multiple choice tasks.

Figure 6: Scatter plot of accuracy by education level on multiple choice tasks.

education	pct_correct	av_dist
Bachelor	0.47	1.46
Master	0.54	1.28
Primary education	0.54	1.22
Secondary education	0.48	1.42

Table 2: Average score by education level all domains.

asset_type	min_time	mean_time	median_time	max_time
audio	3.00	18.04	16.00	45.00
image	2.00	12.26	11.00	44.00
video	2.00	17.73	17.00	45.00

Table 3: Time to answer question by asset type for all domains.

confidence	pct_correct	av_dist
1	0.26	1.35
2	0.23	1.50
3	0.35	1.33
4	0.55	2.01
5	0.86	0.66

Table 4: Average score by confidence level all domains.
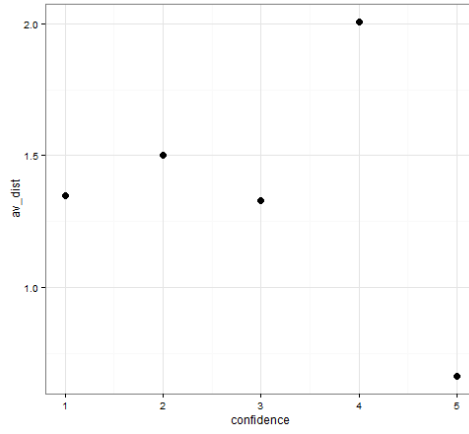
(a)
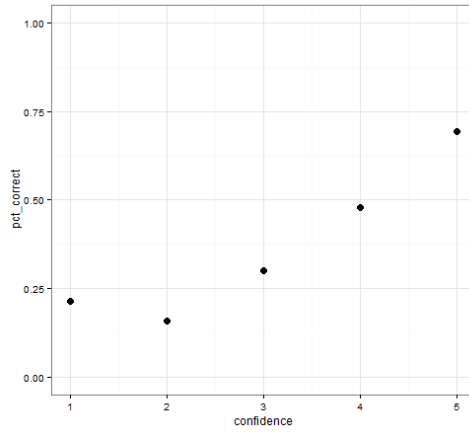
(b)

Figure 7: Scatter plot of percent accuracy by reported confidence level for point estimate and multiple choice quesions.

experimental_condition	av_confidence	time	av_score
control	3.30	16.24	0.59
social	3.54	16.08	0.61

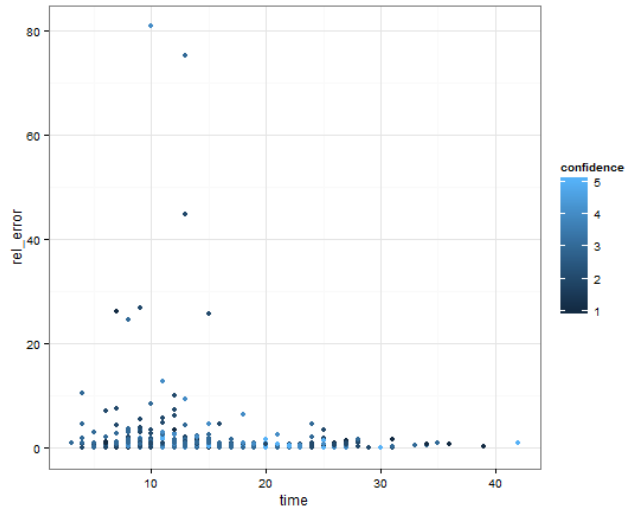Table 5: Statistics for experimental condition.

Figure 8: Relative Error vs time taken to answer question. Colour indicates reported confidence.
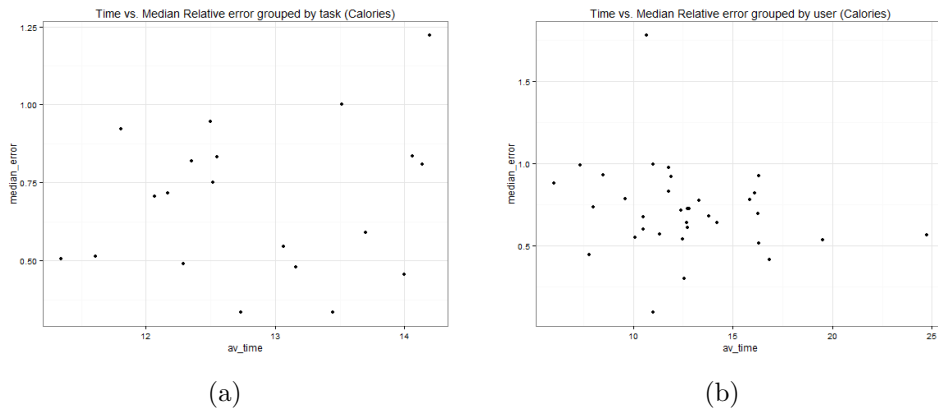

(a)

(b)

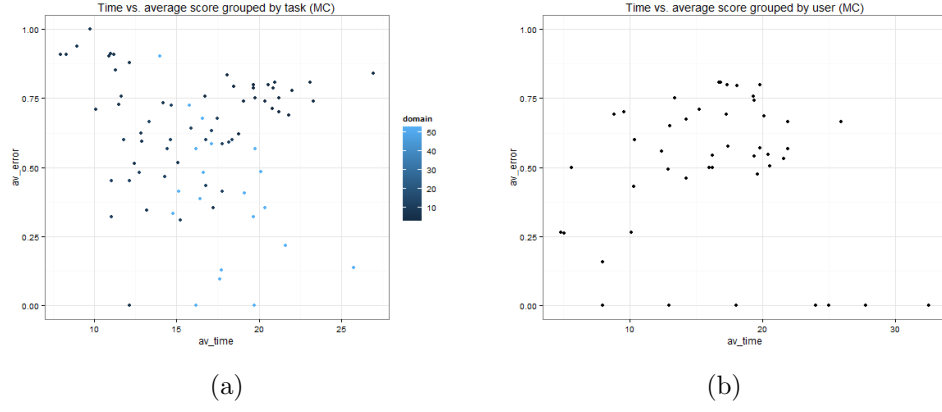Figure 9: Time vs. average score grouped by task (a) and users (b) (point estimate questions)

Figure 10: Time vs. average score grouped by task (a) and users (b) (MC questions)

domain_id	name	confidence	crowd_score_mc	crowd_av_error	crowd_median_err
3	MagicTrick	3.84	20		
7	Landmarks	3.88	18		
12	Penalties	2.75	12		
19	Calories	2.75		1.43	0.69
53	ThemeSongs	2.82	12		

Table 6: Confidence vs domain error.

2 Crowd Rankings

|	x
domain	19.00
abs_error_mean	458.63
abs_error_median	377.15
rel_error_mean	1.43
rel_error_median	0.69
rank_mean	0.45
rank_median	0.68
rank.geom.mean	0.63
rank_trunc_mean	0.58
rank_trunc_geom_mean	0.64

Table 7: Point estimate domains ranked 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.

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	Calories	12.00	0.88

Table 8: 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.
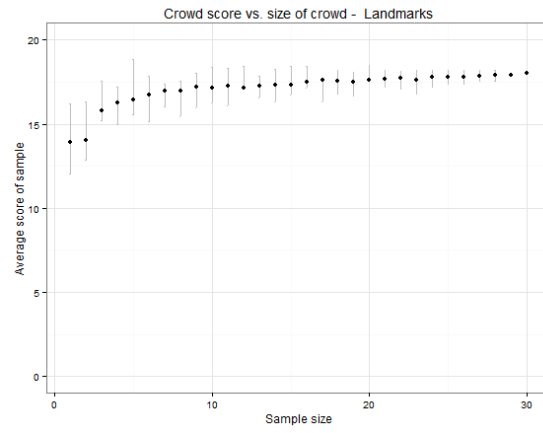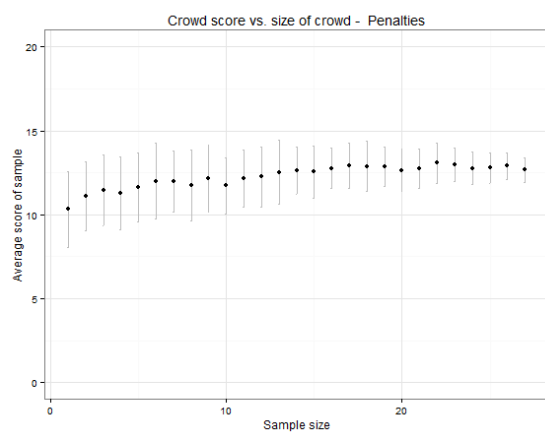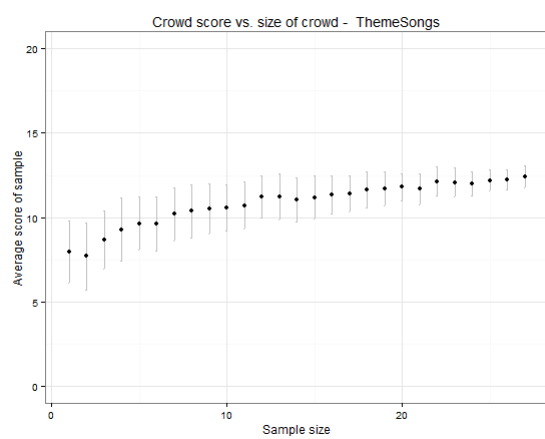
(a)
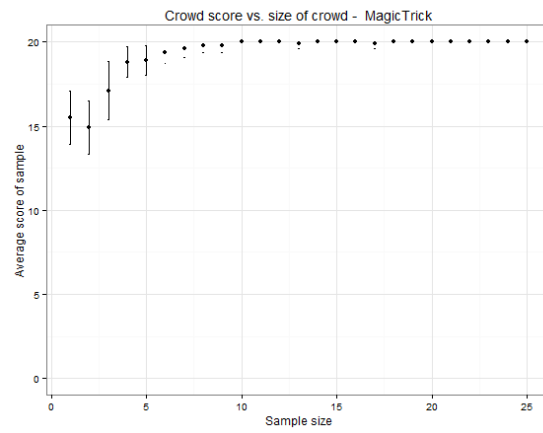
(b)

Figure 11: Sample size vs accuracy for landmarks (a) and magic trick (b).

9

(a)

(b)

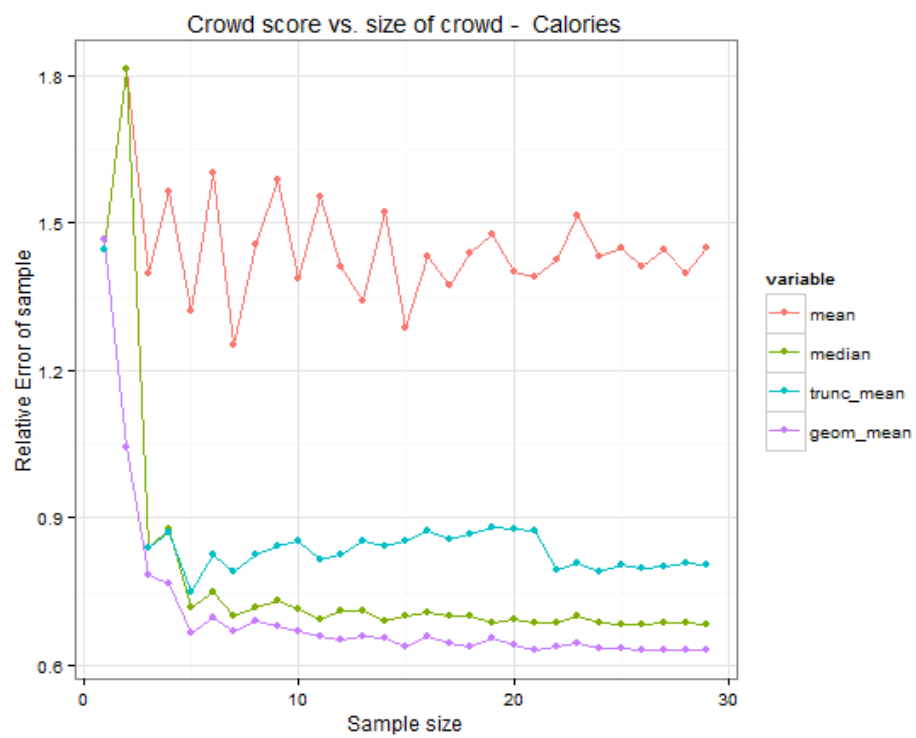Figure 12: Sample size vs accuracy for penalties (a) and theme songs (b).

Figure 13: Sample size vs accuracy for calories.