1 Exploratory Data Analysis
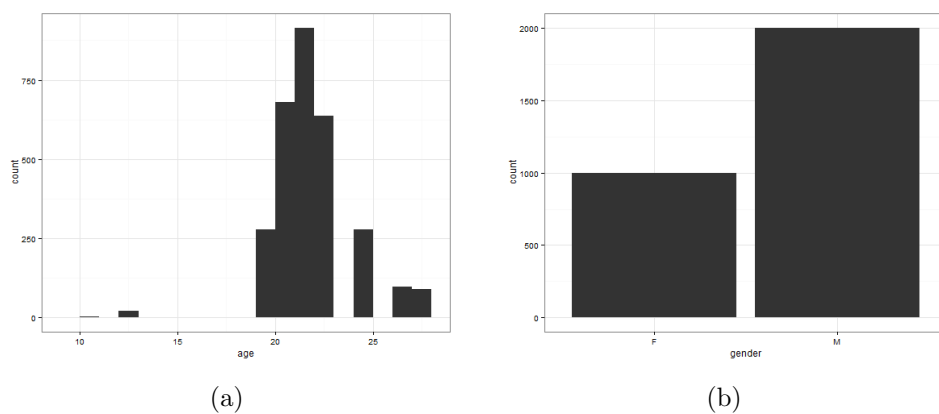

(a)   (b)

Figure 1: Histogram of age and gender of participants.

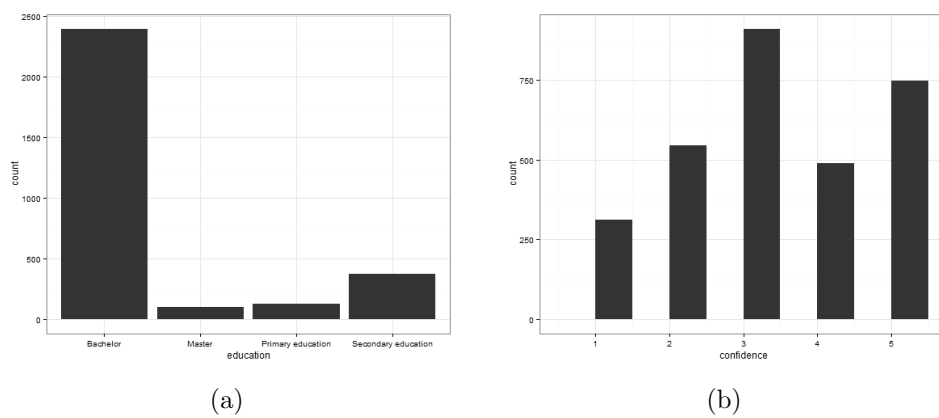(a)   (b)

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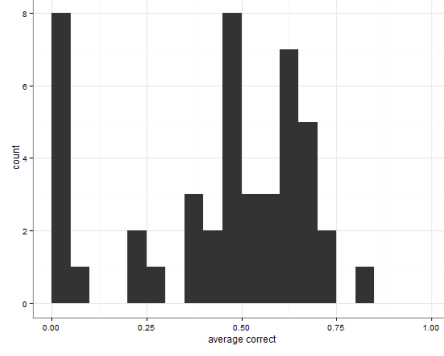
Figure 3: Histogram of user performance on multiple choice tasks.
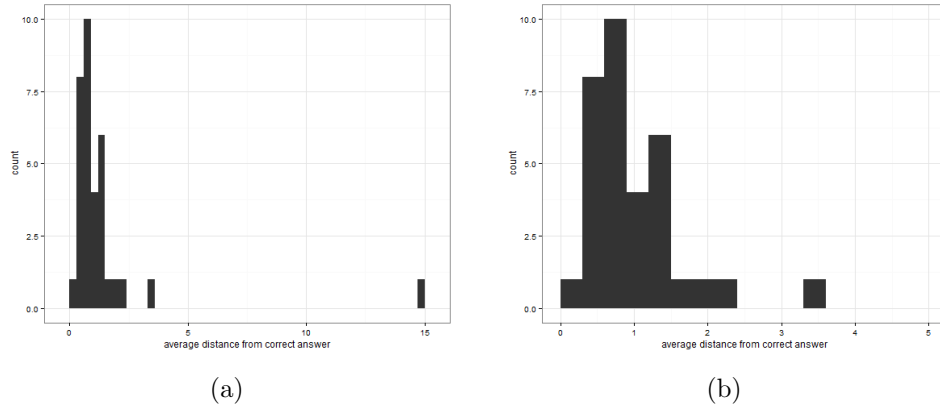

(a)   (b)

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

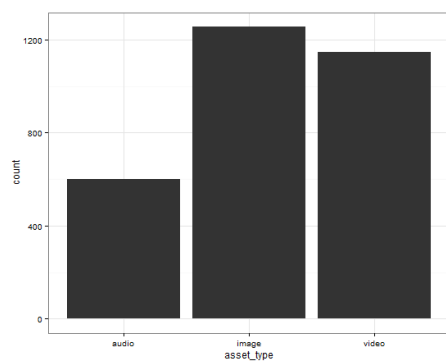|	type	av_score
1	audio	0.10
2	image	0.71
3	video	0.65

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

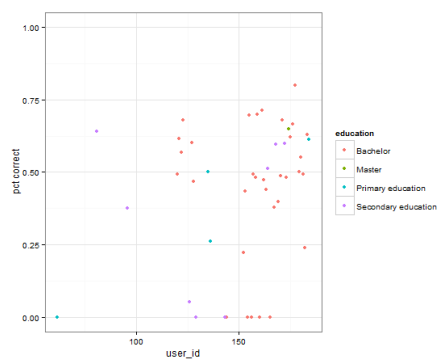Figure 5: Histogram of user performance on multiple choice tasks.

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

	education	pct_correct	av_dist
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.

	asset_type	av_time	median_time
1	audio	18.04	16.00
2	image	12.26	11.00
3	video	17.73	17.00

Table 4: Average and median time to answer question by asset type for all domains.
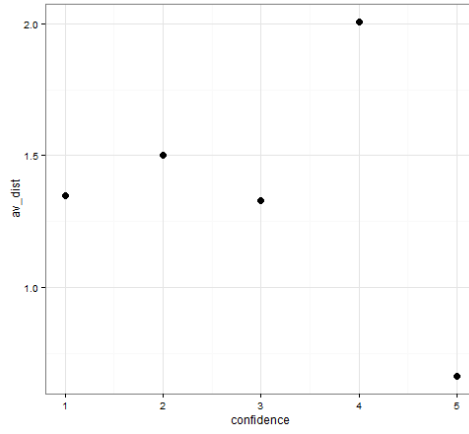
(a)
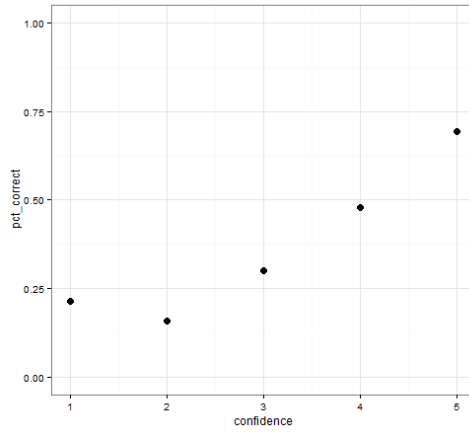
(b)

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

	experimental_condition	time
1	control	15.68
2	social	15.26

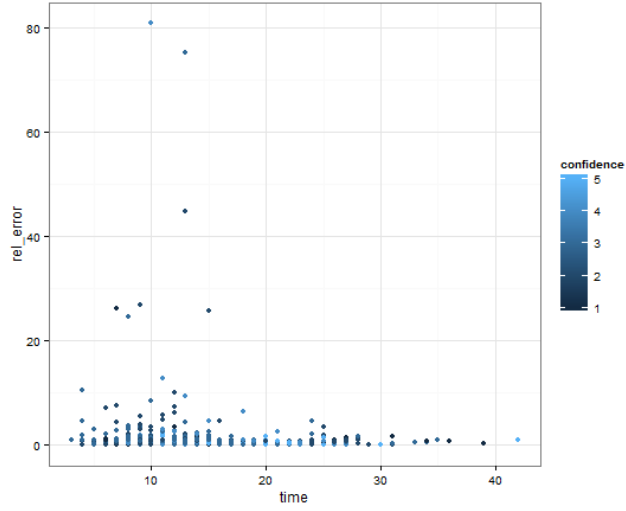Table 5: Average time to answer question by experimental condition.

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

	experimental_condition	av_confidence	time	av_score
1	control	3.28	16.34	0.66
2	social	3.54	16.44	0.68

Table 6: Statistics for experimental condition.

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	4		

Table 7: 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 8: 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
1	3	MagicTrick	20.00	1.00
2	7	Landmarks	18.00	0.92
3	12	Penalties	12.00	0.82
4	53	ThemeSongs	4.00	1.00

Table 9: 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.
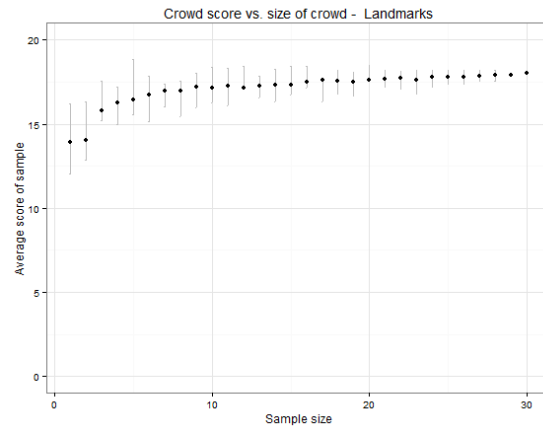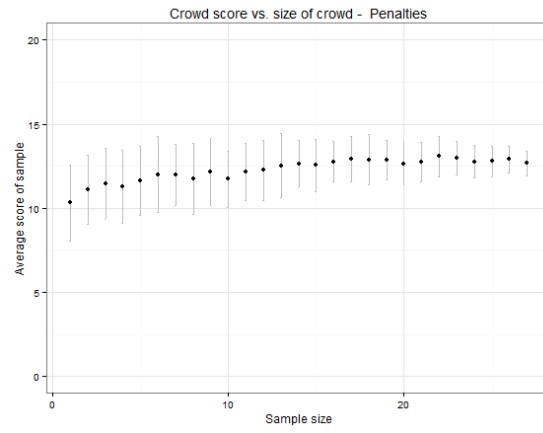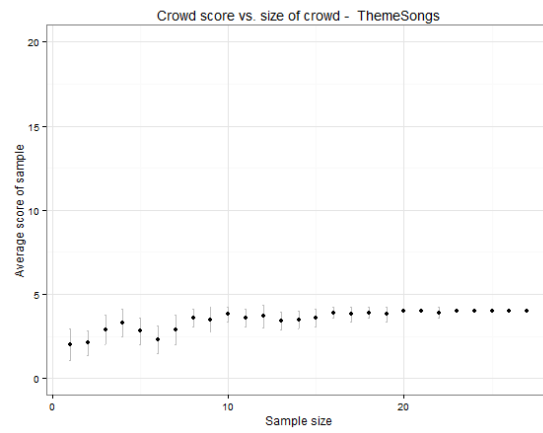
(a)
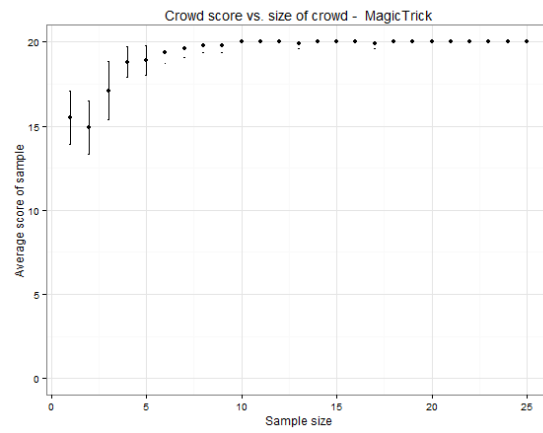
(b)

Figure 9: Sample size vs accuracy for landmarks (a) and magic trick (b).

(a)

(b)

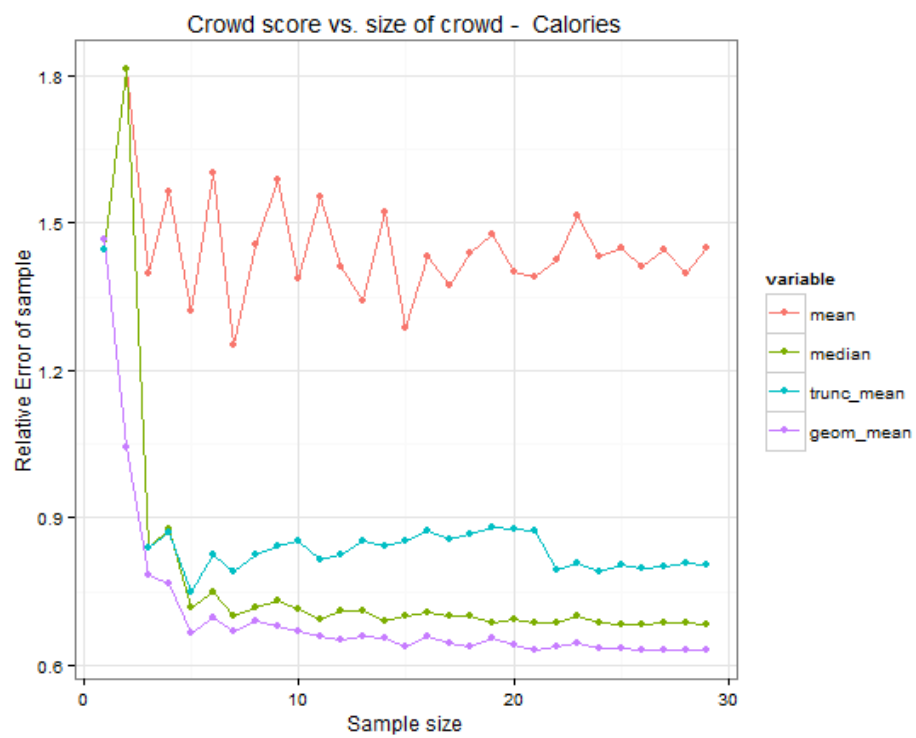Figure 10: Sample size vs accuracy for penalties (a) and theme songs (b).

Figure 11: Sample size vs accuracy for calories.