

Analysis of GroundTruth Data

Terrence Neumann

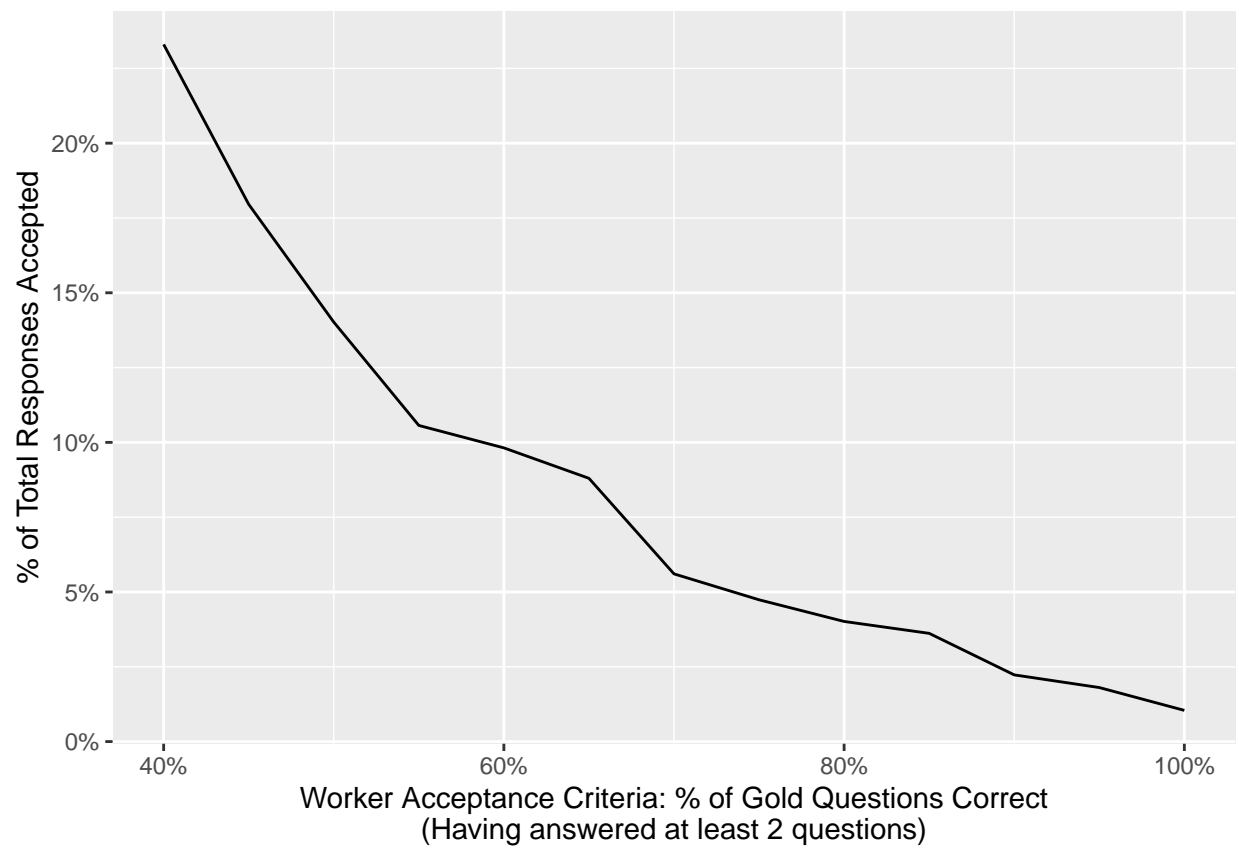
2023-05-03

Contents

Assessing Worker Quality	2
Threshold for Acceptance Analysis	2
Questions Answered per Worker	3
Summary of “Good” Workers	4
Amount Paid to Worker vs % Gold Questions Answered Correctly	6
Variance in Check-Worthiness for “Good” Workers (>80% Gold)	7
Variance in Interest to General Public for “Good” Workers (>80% Gold)	8
Variance of Checkworthiness of GOLD-TRUE Items by Worker (these should be un-checkworthy)	9
Accepted Response Analysis	10
Number of Responses Accepted by Topic	10
Mean Responses by Topic	11
Differences in Response By Gender	12
Largest Gender Differences by Item	14
Differences in Response By Information Literacy	15

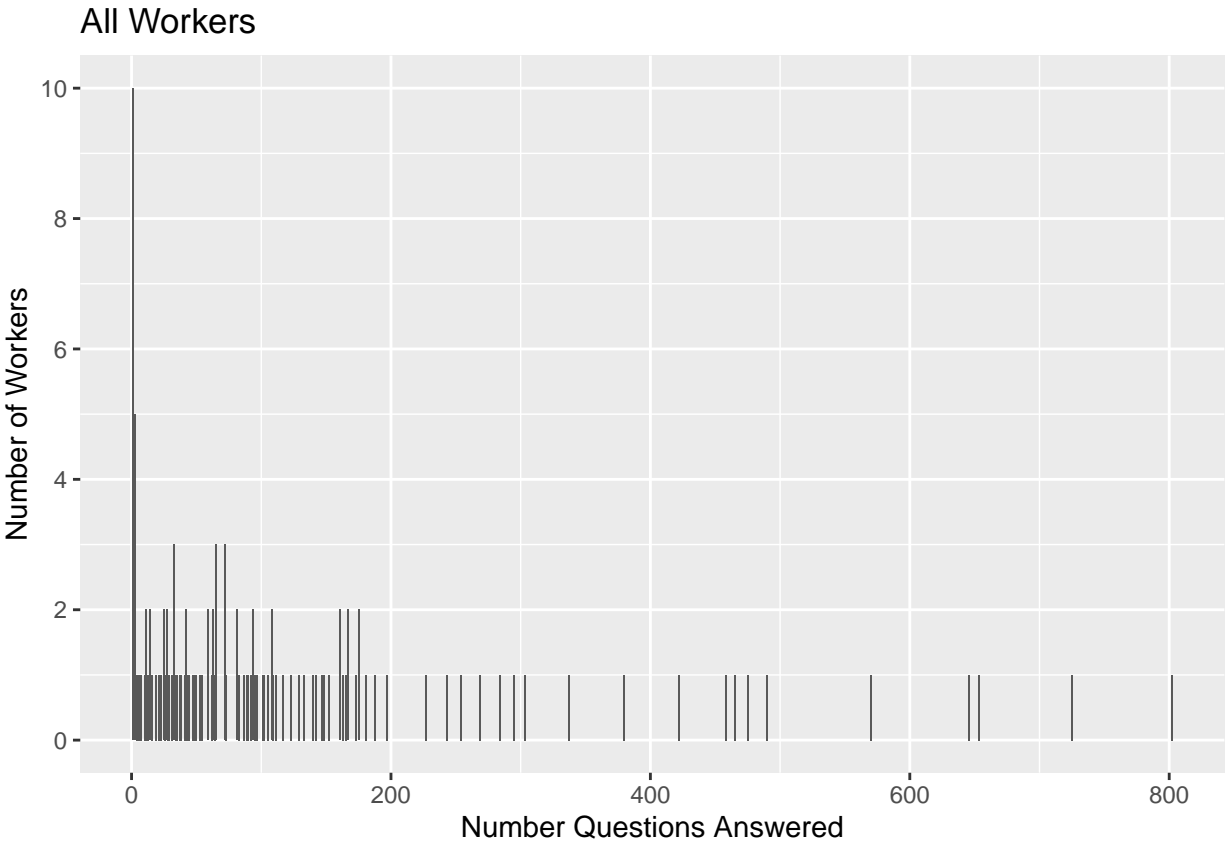
Assessing Worker Quality

Threshold for Acceptance Analysis



A worker acceptance criteria of 80% on simple gold questions drops the percentage of rows accepted to roughly 8%. Acceptance criteria of 85% drops accepted responses to 2%.

Questions Answered per Worker



Summary of “Good” Workers

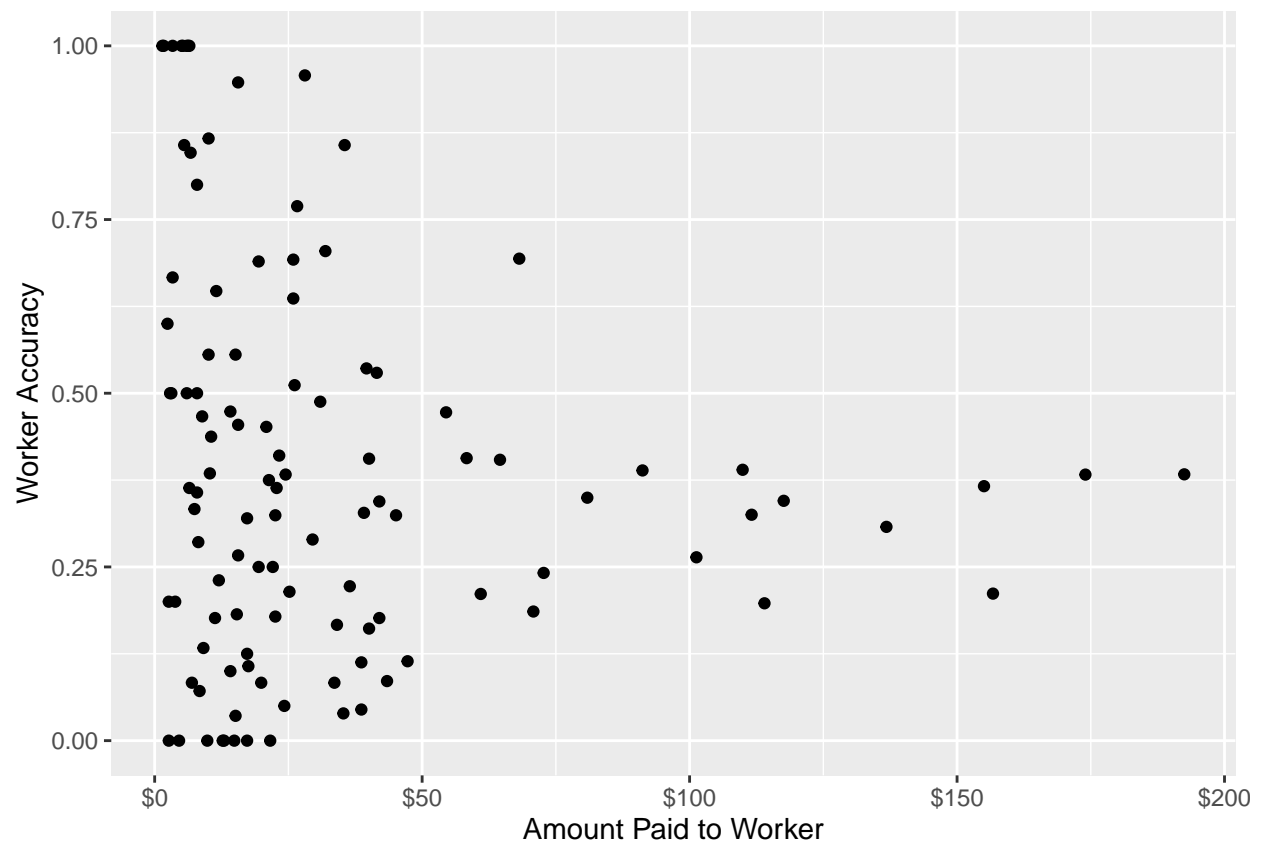
We define good workers as workers with $> 80\%$ of gold questions correct, having answered at least 2 questions.

worker	Age	Education	Gender	Race	SexualOrientation
public.us-east-1.A15QVZHXJYWO31	N/A	NA	N/A	N/A	NA
public.us-east-1.A17PYKQUQH830K	31-40 years old	NA	Female	White	NA
public.us-east-1.A1GQUOWKR6X33C	41-50 years old	Bachelor’s Degree	Female	White	Heterosexual
public.us-east-1.A1MHUF6WJDR4A2	21-30 years old	NA	Female	White	NA
public.us-east-1.A1NA5IPNSNOGYF	51-60 years old	NA	Male	White	NA
public.us-east-1.A1ODA3Q5H6HTDQ	31-40 years old	NA	Female	White	NA
public.us-east-1.A1SCF3L6LTFQHF	21-30 years old	NA	Male	White	NA
public.us-east-1.A1TKVWCJLBIUYF	21-30 years old	NA	Male	Asian	NA
public.us-east-1.A21X60VQIDZU4D	21-30 years old	NA	Female	White	NA
public.us-east-1.A25SH3L1UR6Y1Z	21-30 years old	NA	Male	White	NA
public.us-east-1.A2K5YXJV0IDKH5	N/A	NA	N/A	N/A	NA
public.us-east-1.A2Q99XBBVXFPG7	N/A	NA	Female	White	NA
public.us-east-1.A2WSWG9IQ0ZOC8	31-40 years old	NA	Female	White	NA
public.us-east-1.A314XJY8V1YL12	51-60 years old	Master’s Degree	Female	White	Heterosexual
public.us-east-1.A3HJ1RFM8ZYNGZ	31-40 years old	NA	Male	White	NA
public.us-east-1.ABDWWV9GFHVQ5	31-40 years old	NA	Male	White	NA
public.us-east-1.AI91L9BSP6NBM	31-40 years old	NA	Female	White	NA
public.us-east-1.ANIXYYFREZVH	31-40 years old	Master’s Degree	Male	Asian	Heterosexual
public.us-east-1.APQXZF7GKNTC	31-40 years old	NA	Male	White	NA

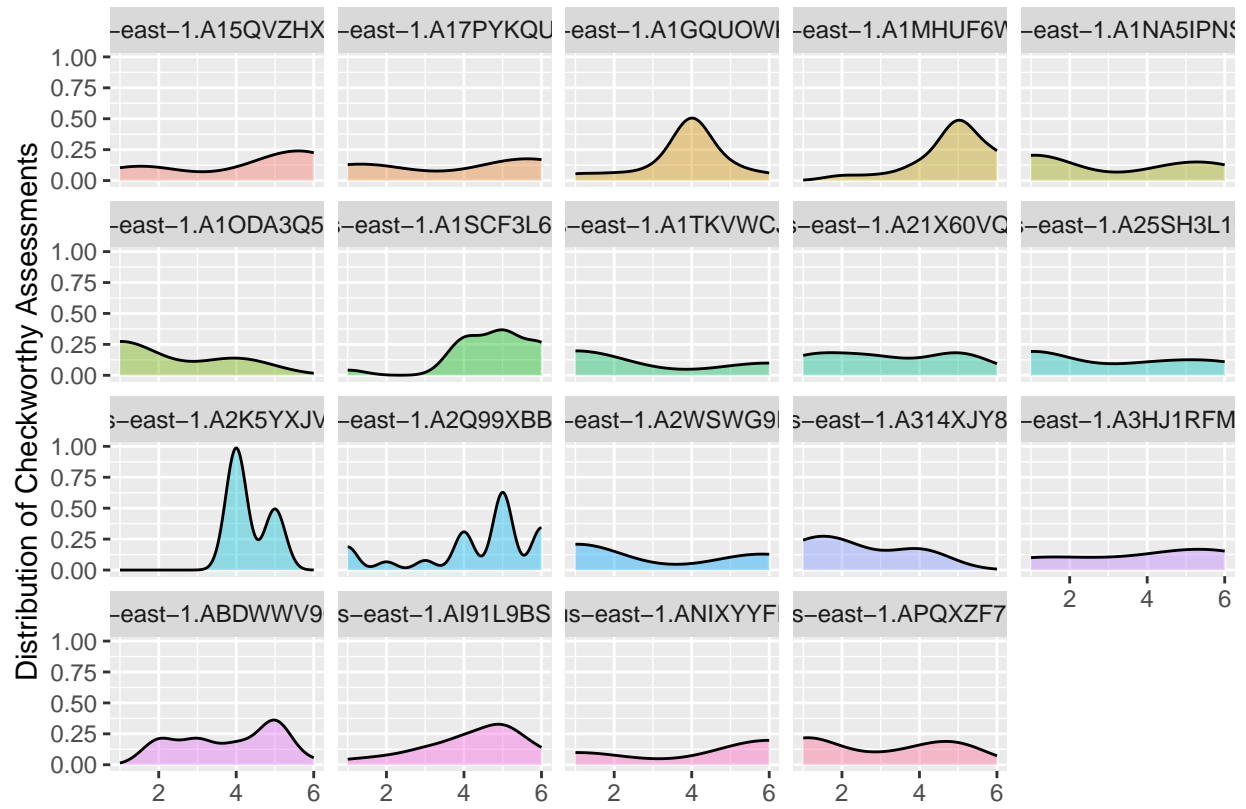
worker	Number Questions Answered	Number Gold Questions	Number Gold Correct	% Gold Correct
public.us-east-1.A15QVZHXJYWO31	26	6	6	-100
public.us-east-1.A17PYKQUQH830K	14	7	7	6
public.us-east-1.A1GQUOWKR6X33C	33	15	12	0

worker	Number Questions Answered	Number Gold Questions	Number Gold Correct	% Gold Correct
public.us-east- 1.A1MHUF6WJDR4A2	22	7	7	2
public.us-east- 1.A1NA5IPNSNOGYF	28	13	11	-100
public.us-east- 1.A1ODA3Q5H6HTDQ	6	3	3	6
public.us-east- 1.A1SCF3L6LTFQHF	65	19	18	6
public.us-east- 1.A1TKVWCJLBIUYF	3	2	2	8
public.us-east- 1.A21X60VQIDZU4D	25	10	10	8
public.us-east- 1.A25SH3L1UR6Y1Z	21	12	12	6
public.us-east- 1.A2K5YXJV0IDKH5	3	1	1	-100
public.us-east- 1.A2Q99XBBVXFPG7	148	49	42	6
public.us-east- 1.A2WSWG9IQ0ZOC8	23	7	6	6
public.us-east- 1.A314XJY8V1YL12	3	1	1	8
public.us-east- 1.A3HJ1RFM8ZYNGZ	7	3	3	6
public.us-east- 1.ABDWWV9GFHVQ5	117	47	45	8
public.us-east- 1.AI91L9BSP6NBM	42	15	13	6
public.us-east- 1.ANIXYYFREZVH	3	1	1	4
public.us-east- 1.APQXZF7GKNTC	27	10	10	6

Amount Paid to Worker vs % Gold Questions Answered Correctly

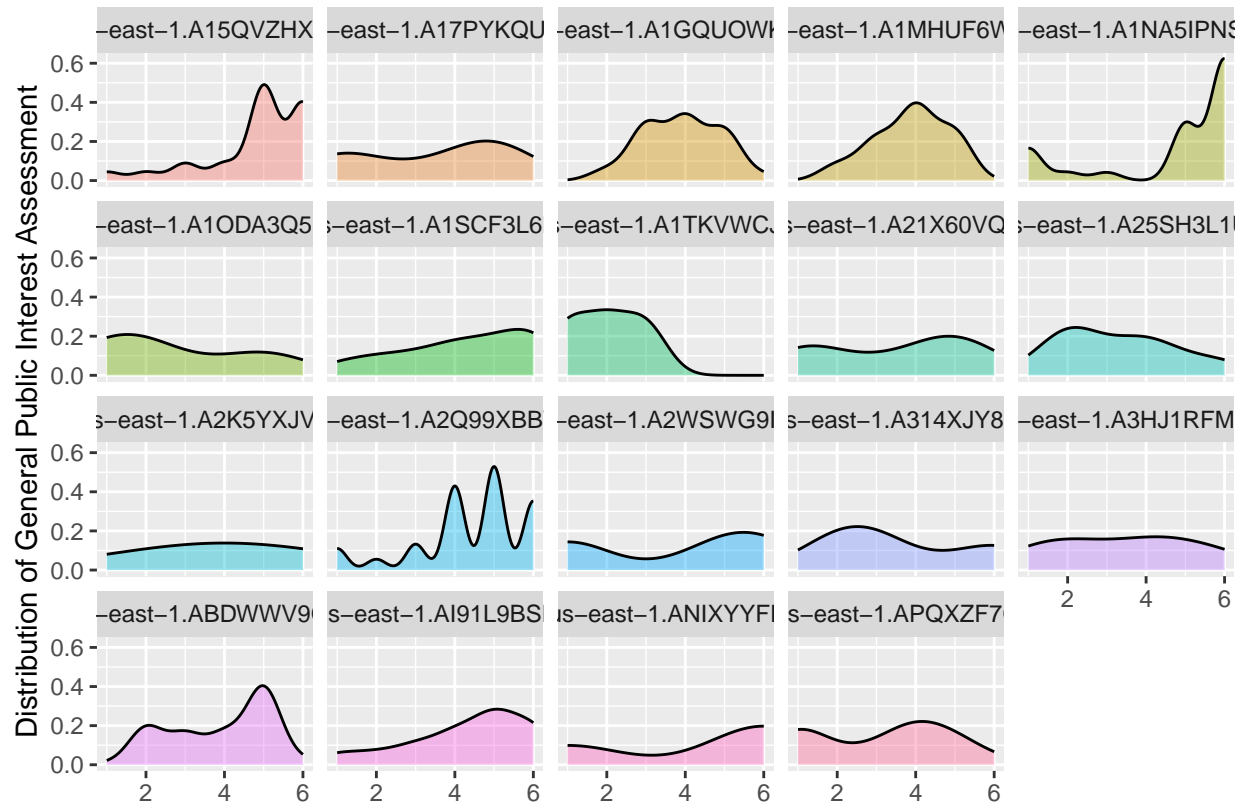


Variance in Check-Worthiness for “Good” Workers (>80% Gold)



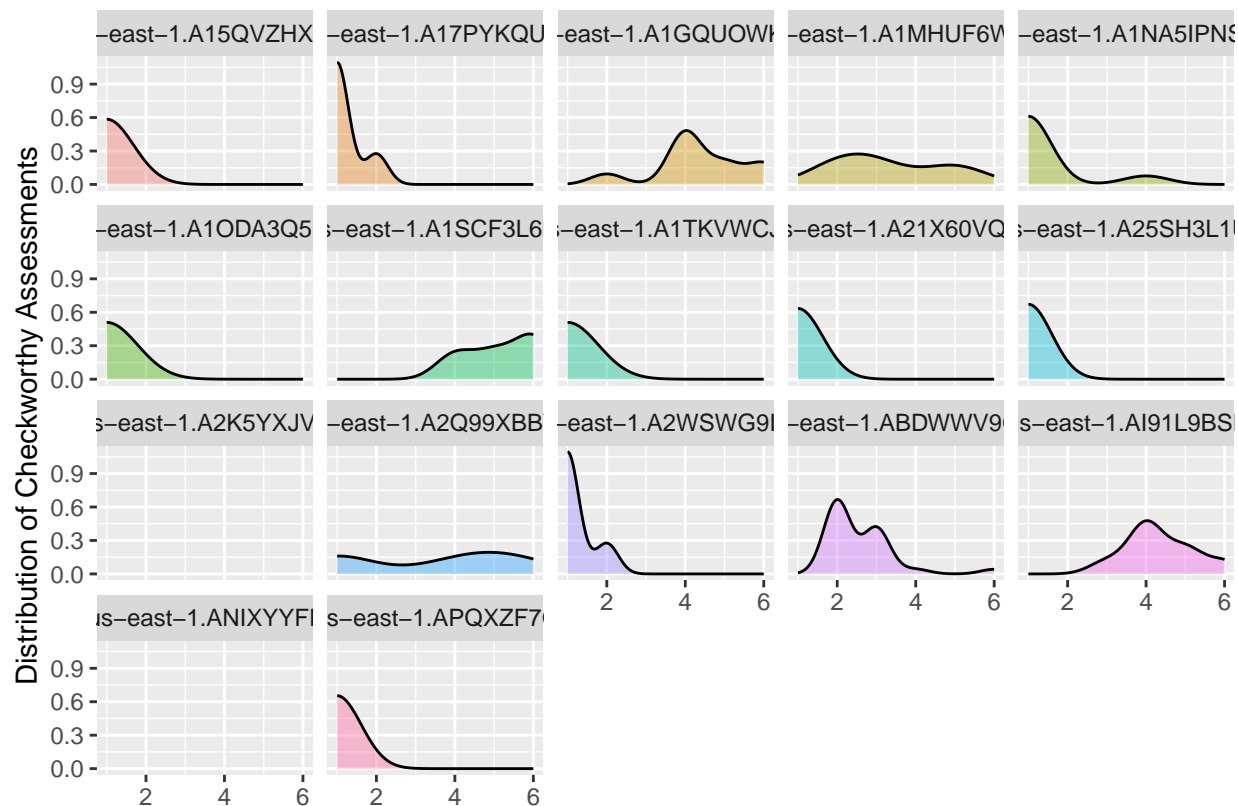
Most workers have high variance, indicating a good-faith effort in differentiating check-worthiness.

Variance in Interest to General Public for “Good” Workers (>80% Gold)



Most workers have high variance, indicating a good-faith effort in differentiating interest to the general public.

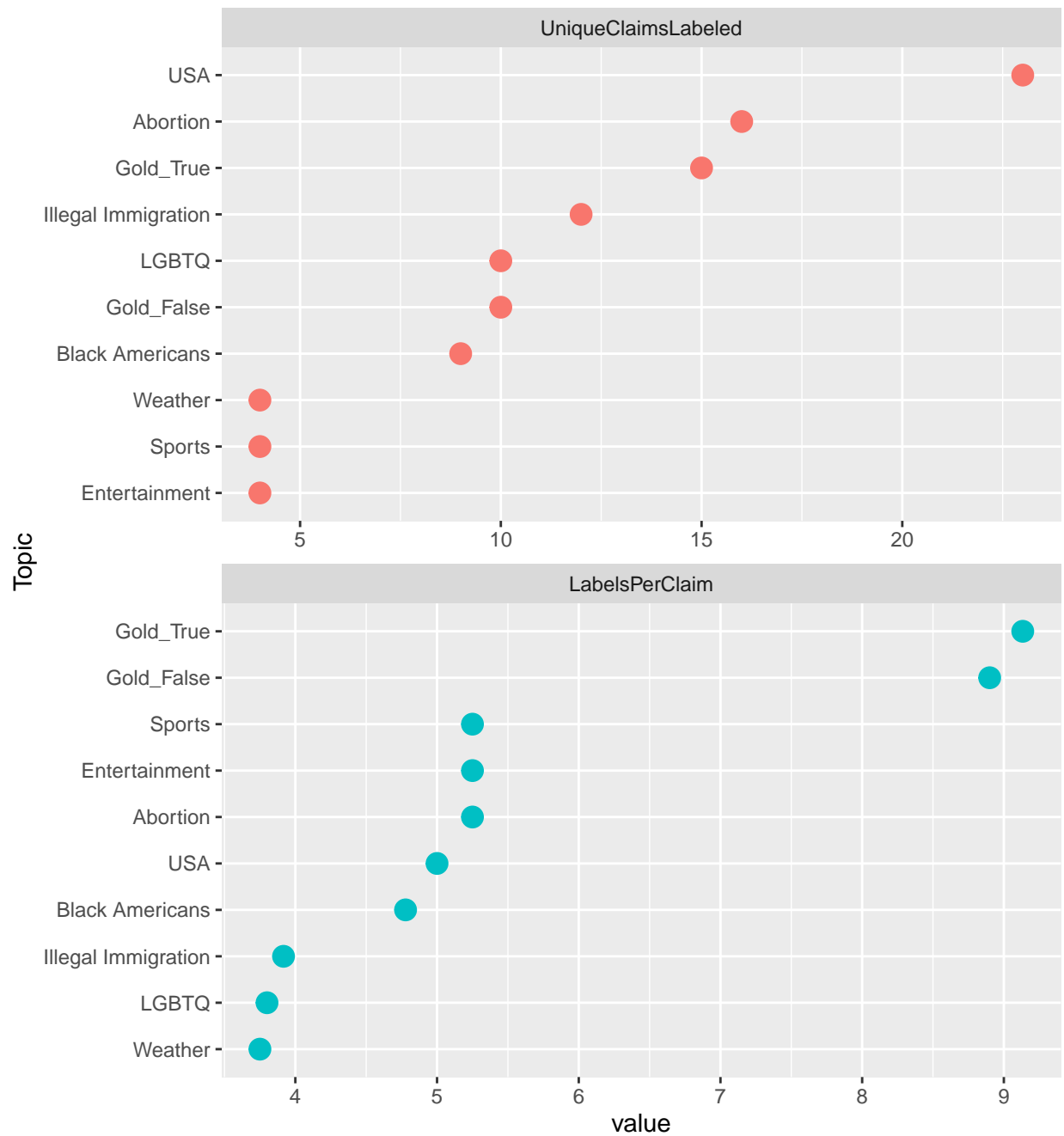
Variance of Checkworthiness of GOLD-TRUE Items by Worker (these should be un-checkworthy)



Most workers rate the gold-TRUE items as having low check-worthiness. This is what we'd expect, and indicative of good-faith effort.

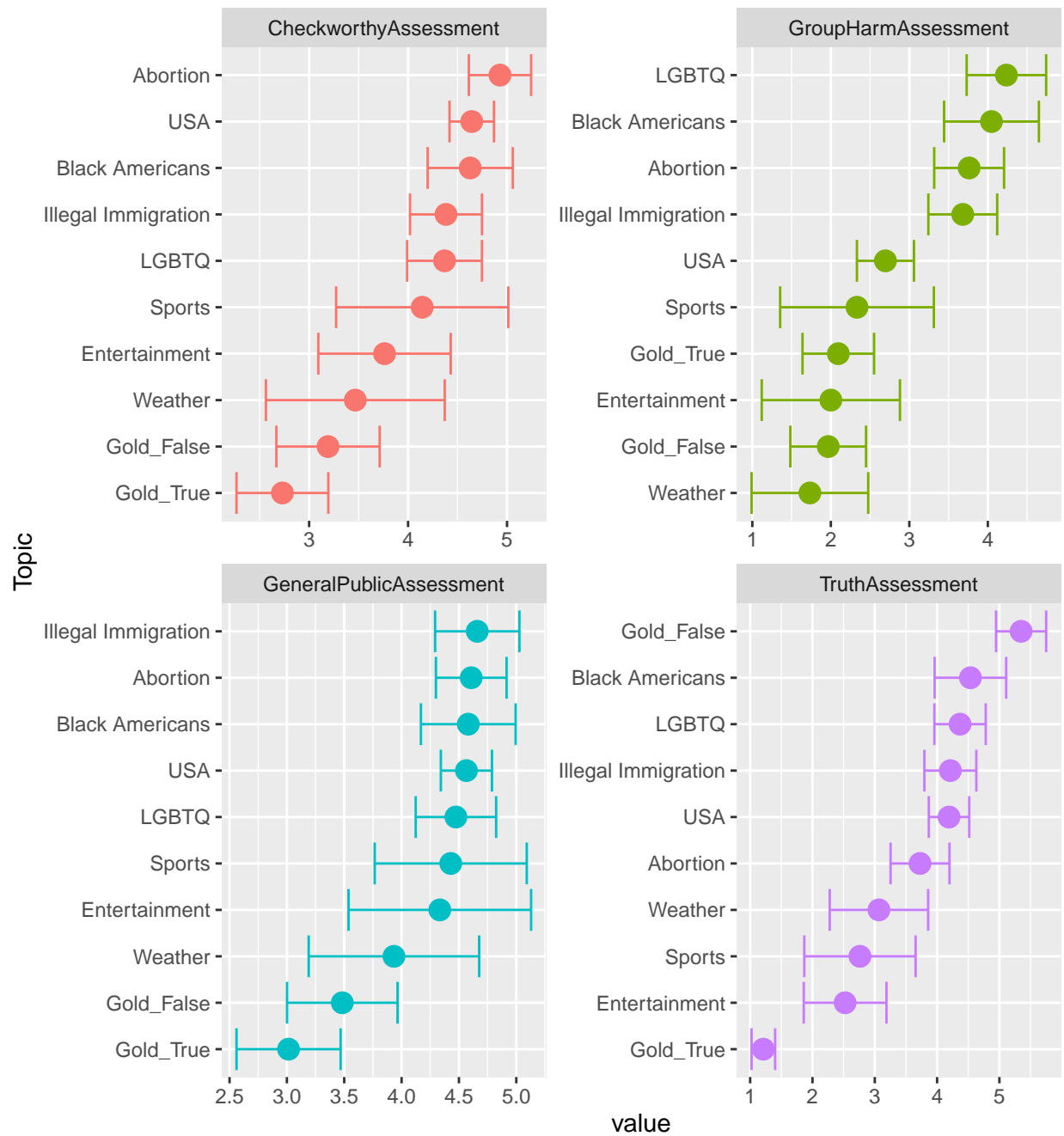
Accepted Response Analysis

Number of Responses Accepted by Topic



Here we see the number of unique claims that have been accepted by topic, as well as the number of annotations per claim within a topic, as not all topics have the same number of claims. The two gold categories have the highest number of annotations per claim, which is expected.

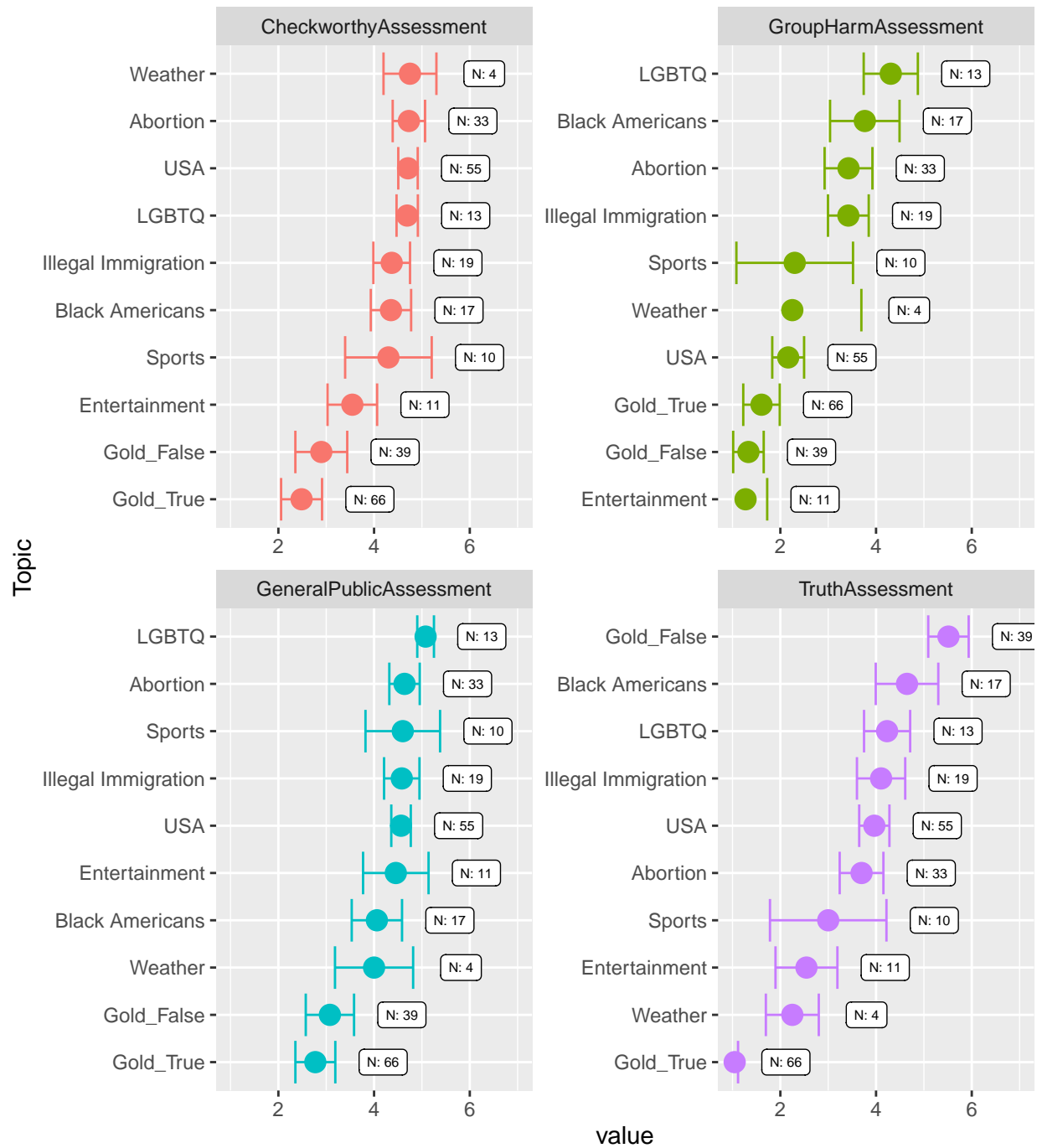
Mean Responses by Topic



This chart shows the average assessment across our accepted sample, as well as ± 1 standard error. The standard error is used to illuminate agreement/disagreement. In the TruthAssessment panel, Gold_False and Gold_True are at their expected extremes with narrow standard errors, indicating these workers are well-calibrated in their assessments.

Differences in Response By Gender

Attitudes of Men Towards Topics



Attitudes of Women Towards Topics



Largest Gender Differences by Item

Table 3: Checkworthiness by Gender

claim	Female	Male	difference
Aborted fetal cells are in artificial flavors added to food products people eat.	6	1.0	5.0
The fastest ever recorded serve in professional tennis was recorded at a speed of 230 miles per hour.	1	5.0	4.0
The state of Hawaii has the highest number of lightning strikes per square mile of any state in the United States.	2	6.0	4.0
A bicycle has two wheels.	5	1.5	3.5
February was chosen as Black History month because it's the shortest month	1	4.0	3.0
Kelly Clarkson, the first American Idol winner, is the ninth best-selling music artist of all time.	5	2.0	3.0

Table 4: Interest to the General Public by Gender

claim	Female	Male	difference
African Americans Make 87% Of Kidney Dialysis Patients In The United States in 2022	5.00	2.00	3.00
A week contains seven days.	1.00	3.75	2.75
Aborted fetal cells are in artificial flavors added to food products people eat.	4.00	1.50	2.50
The fastest ever recorded serve in professional tennis was recorded at a speed of 230 miles per hour.	3.00	5.40	2.40
Illegal immigrants receive welfare benefits.	5.33	3.00	2.33
The iconic red leather jacket worn by Michael Jackson in his “Thriller” music video sold at auction for \$110 million.	2.00	4.33	2.33

An abortion item sparks the largest disagreement in check-worthiness, with women rating it as highly check-worthy and men rating it as not check-worthy.

Differences in Response By Information Literacy

Attitudes of High Media Literacy Towards Topics



Attitudes of Medium Media Literacy Towards Topics



Attitudes of Low Media Literacy Towards Topics

