Kirby Linvill

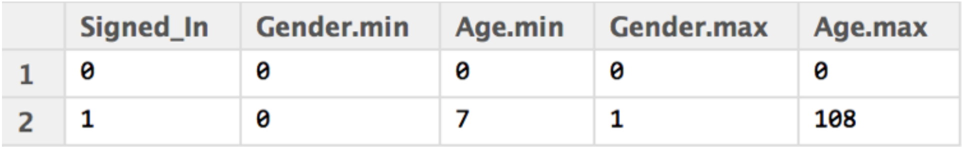
CSCI 183 Data Science HW #1

Main observations:

- If someone is not signed in, their gender is assigned as female and their age is assigned as 0.
- Female and male viewership is about the same across all age ranges (after accounting for not being signed in). The two exceptions are that there are roughly twice as many male viewers as female viewers in the under 18 group, and roughly twice as many female viewers as male viewers in the 65+ group.
- The average (signed in) user age is about 40 and viewership is spread roughly evenly from ages 20 to 50.
- The age distribution is relatively normal with an elongated right tail.
- Click through rate, clicks, and impressions are roughly equivalent across all ages and genders.
- A majority of viewers have 5 impressions and 0 clicks.
- Impressions are independent of whether or not a user is signed in.
- Signed in users are attributed roughly half the clicks of not signed in users.

Observation: If someone is not signed in, their gender is assigned as female and their age is assigned as 0.

Table 1. Gender and Age mins and maxes based on Signed_In status (nyt1.csv)



	Signed_In	Gender.min	Age.min	Gender.max	Age.max
1	0	0	0	0	0
2	1	0	7	1	108

As a result of this observation, any kind of gender or age related statistics should only be based off of the subset of the data for which the user is signed in.

Observation: Female and male viewership is about the same across all age ranges (after accounting for not being signed in). The two exceptions are that there are roughly twice as many male viewers as female viewers in the under 18 group, and roughly twice as many female viewers as male viewers in the 65+ group.

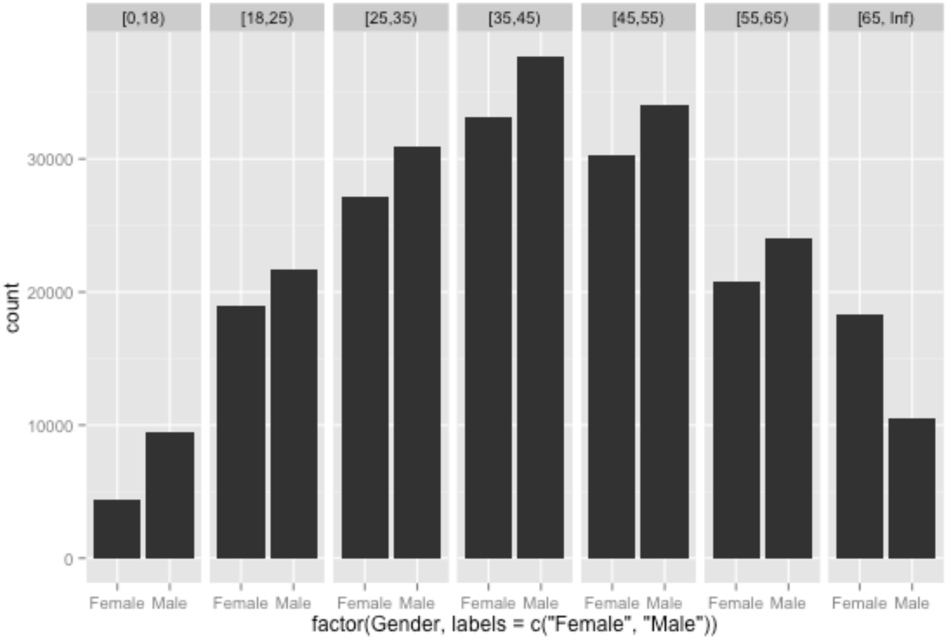
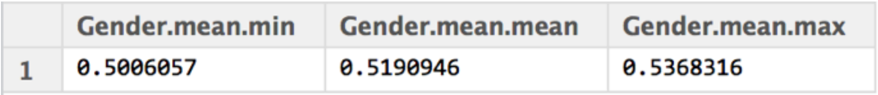


Figure 1. Gender breakdown by Age Group (nyt1.csv)

Table 2. Gender averages across all data



	Gender.mean.min	Gender.mean.mean	Gender.mean.max
1	0.5006057	0.5190946	0.5368316

Observations: The average (signed in) user age is about 40 and viewership is spread roughly evenly from ages 20 to 50. The age distribution looks similar to a chi square distribution.

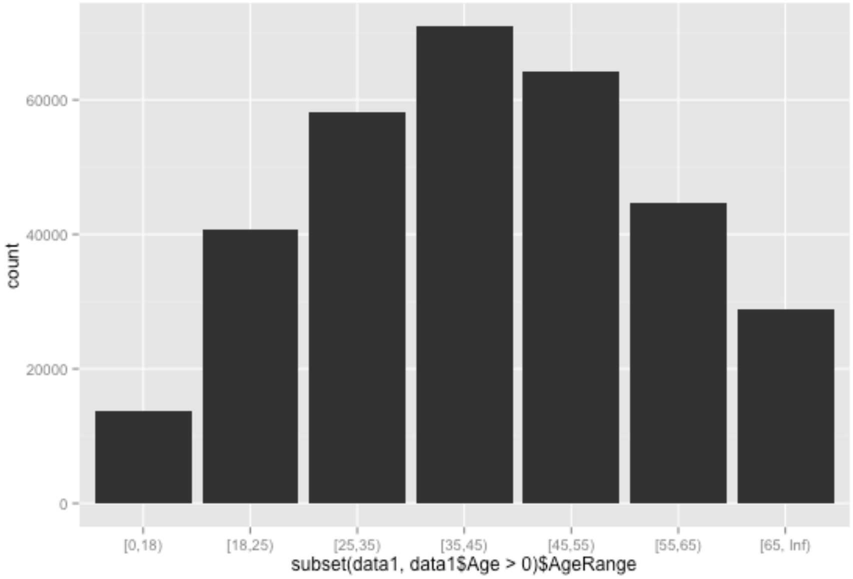


Figure 2. Age Group breakdown (nyt1.csv)

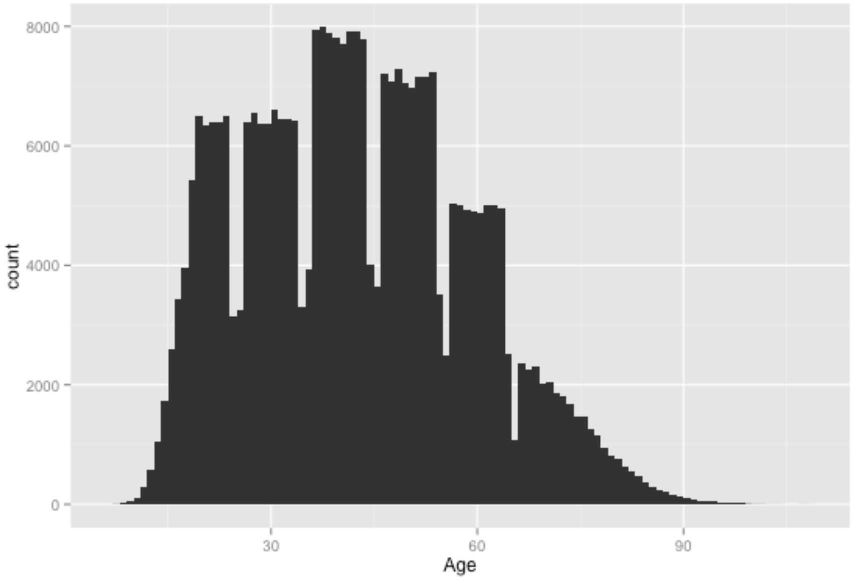


Figure 3. Age breakdown (nyt1.csv)

Objective: Click through rate, clicks, and impressions are roughly equivalent across all ages and genders. By far the largest click through rate is 0.

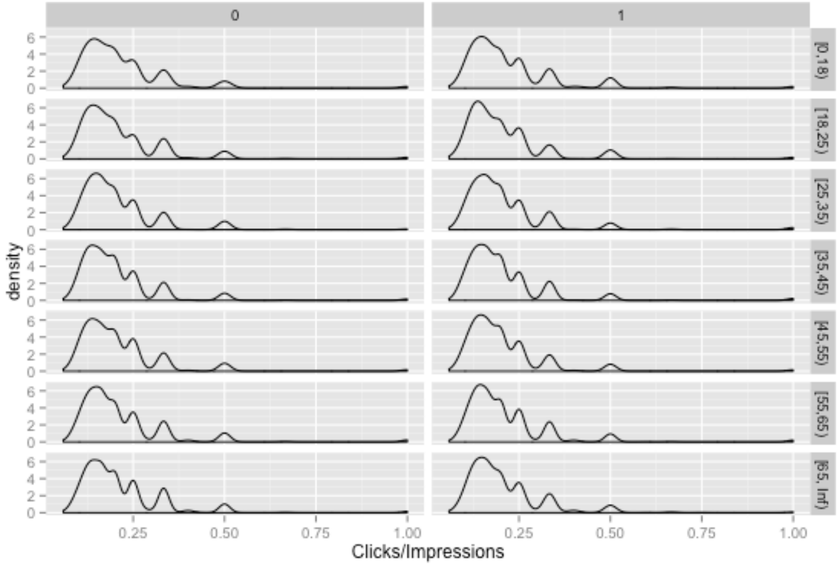


Figure 4. Click-Through Rate by Age and Gender and adjusted to remove NA and 0 values (nyt1.csv). The left column is female behavior and the right column is male behavior.

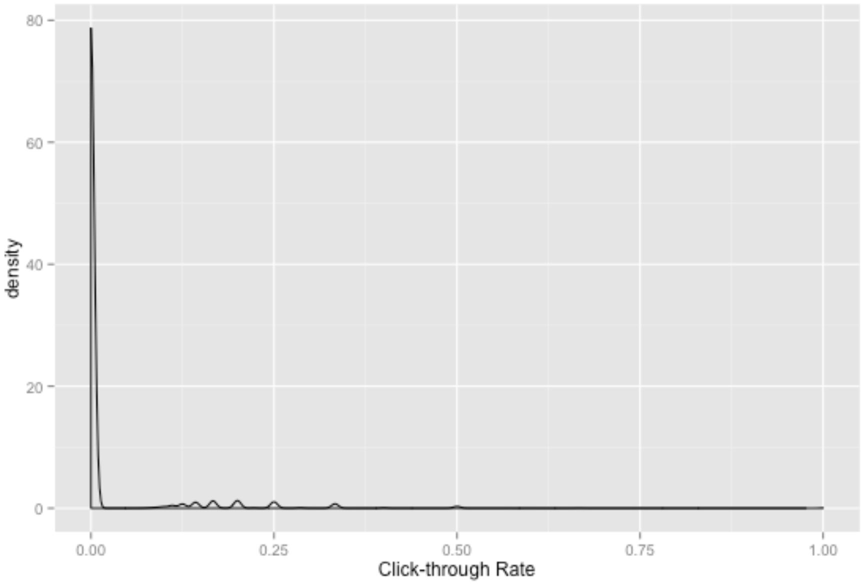


Figure 5. Unadjusted Click-Through Rate (nyt1.csv)

Observation: A majority of viewers have 5 impressions and 0 clicks.

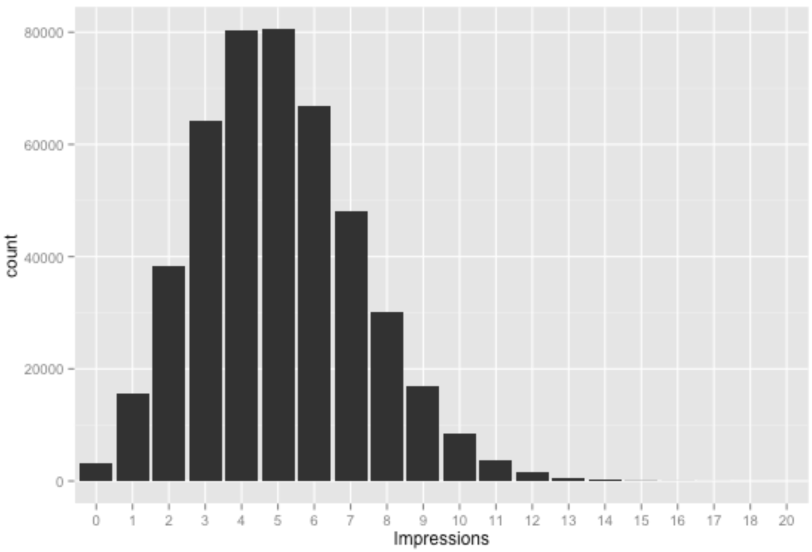


Figure 6. Impressions per viewer (nyt1.csv)

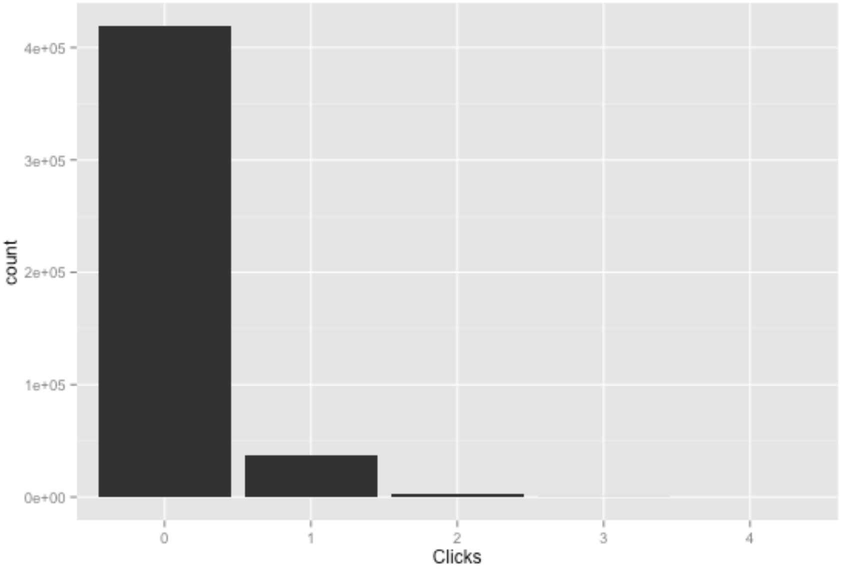


Figure 7. Clicks per viewer (nyt1.csv)

Observation: Impressions are independent of whether or not a user is signed in. Signed in users are attributed roughly half the clicks of not signed in users. This results in signed users having roughly half the click-through rate of non-signed in users.

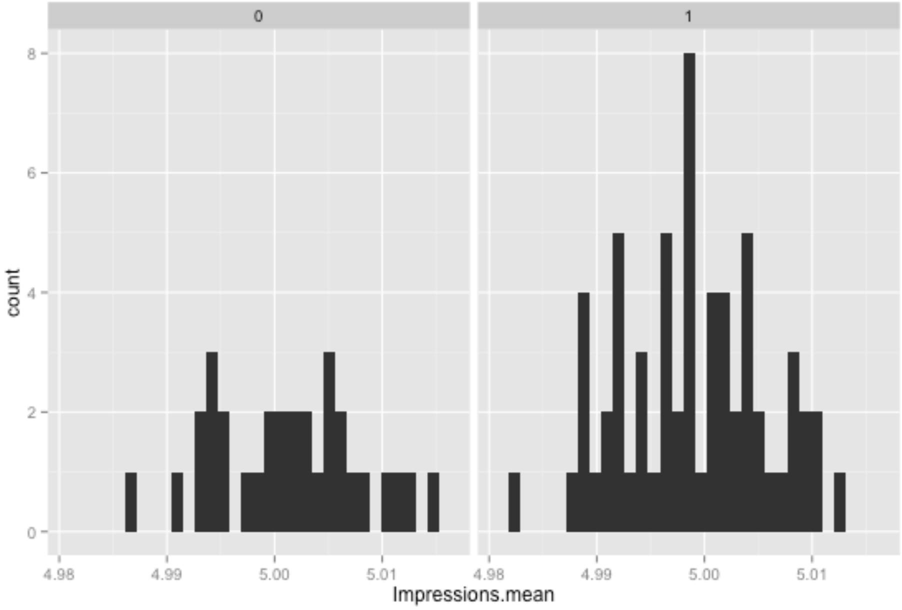


Figure 8. Average Impressions per viewer across all data

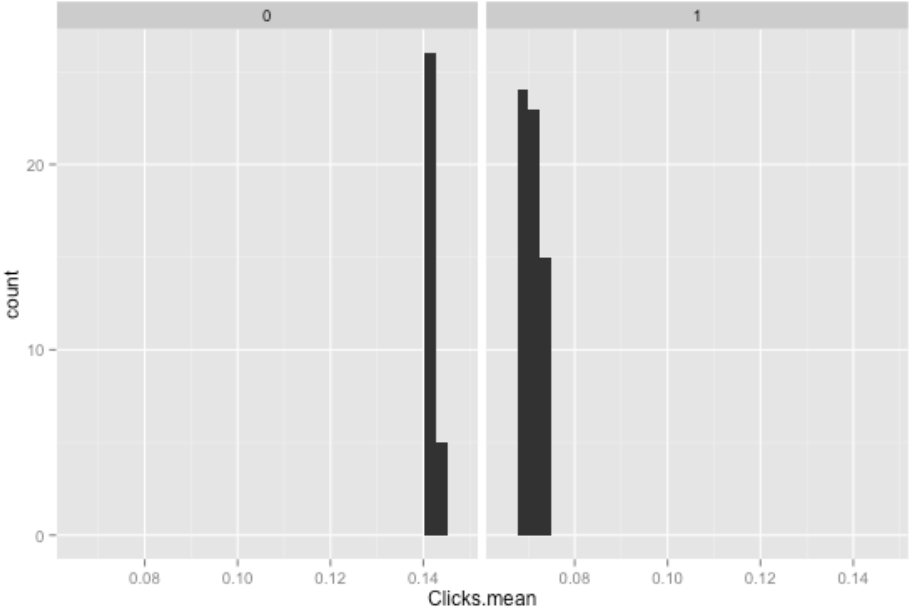


Figure 9. Average Clicks per viewer across all data