Facebook data comes from advertising platform. I calculate men interested in men divided by men interested in men plus men interested in women. The advertising platform allows overall numbers and numbers just for high school students.

The advertising platform does not include data on place of birth. To do this, I used Facebook graph search and the following hack.

Facebook does not allow you to download estimates by place of birth. However, on Facebook graph search, you can search for people by sexuality, place of birth, and place of living.

I estimated gay men by place of birth as follows.

Define g as $P(Gay|LiveState)$
Define p as $P(LiveState|GayBornState)$
Define q as $P(BornState|GayLiveState)$
Define h as $P(LiveState|BornState)$
Define j as $P(BornState|LiveState)$

$$(1) \qquad \frac{q \times h \times g}{p \times j} = \frac{\frac{P(BornState\&Gay\&LiveState) \times P(LiveState\&BornState) \times P(Gay\&LiveState)}{P(Gay\&LiveState) \times P(BornState) \times P(LiveState)}}{\frac{P(LiveState\&Gay\&BornState) \times P(BornState\&LiveState)}{P(Gay\&BornState) \times P(LiveState)}}$$

$$(2) \qquad = \frac{P(Gay\&BornState)}{P(BornState)}$$

The number g comes from the advertising platform. The other numbers come as follows. Search on graph search for men who are interested in men who were born in Mississippi. Then calculate what percent live in Mississippi. And you have p for Mississippi. However, I am now unsure that this gives the proper estimate, as I think this group may be biased.

I have also attempted to do the estimate the following way: I searched for 100 men interested in men who live in a state and were born in the United States. I then searched for 100 men interested in women who live in a state and were born in the United States. I manually recorded the state of birth of all these men.

I then did a weighted average of gay men and straight men born in every state, with weightings based on how many gay men live in every state, as found on the advertising platform.

And this second way seems to give an estimate of more mobility. I am looking into which way is correct.

Craigs List data was downloaded as follows: I found all the Craigs List cities for a given state. (This, for example, is Georgia.) For each one, I went to m4m casual encounters and m4w casual encounters and counted total ads. If there were the maximum number (2,500), I calculated how many days it took to reach the maximum and found a per day number. I then extrapolated for the period included in Craigs List data.

Google gay porn searches can be approximated by dividing gay porn searches by porn searches on Google Trends. The full data uses more searches. Data for searches that include is, husband, and gay can be downloaded from the hack described in my racial animus paper.

Match.com data was found by manually searching the site.