

Research Question: How are females visually represented in the online news nowadays?

Key Variable: The key variable in this research question is the visual representation of females. It is going to be important to figure out exactly what features of visual representation the researcher cares about documenting.

The population: As currently stated, the population is all online news. This is going to be very difficult to work with. Much of this can be fixed through further conceptualization. There are a lot of things that are news and are online. First, is the question focusing on the online presence of traditional news media, or is it open to all kinds of new modes of online news like citizen journalists, bloggers, etc. Even if we do focus in on traditional news outlets online presence those news outlets have significant and diverse presentations of news on the web.

For example, something like CNN will stream their broadcast news directly on the web and make transcripts of their broadcast news available. These kinds of news are online; however, they are really just direct copies of the broadcast news. So if we further refined this to focus on traditional news media's web only content we would need to decide if we wanted to focus on web only video, news articles, Facebook pages, twitter accounts, etc. In short, for this to become an answerable question we need to do some significant conceptualization of exactly what constitutes online news.

Sample: As previously mentioned it is already difficult to describe the population, this makes it a tricky matter to sample. If we can winnow down the idea of the population to be visual representations in major online news outlets then we have at least gotten to the point where we have a manageable theoretical list to work from. I would also suggest focusing on online news articles, as opposed to anything online that could be considered news. I think it would just be too difficult to compare things like news videos with articles and the articles are generally the kinds of things researchers focus on.

There are still a series of difficult decisions to make and justify. Some of the major online news sites are the online presence of major newspapers like the New York Times and the LA Times. Some of the major online news sites are sites for broadcast news, for example the site for CNN. Lastly, there are a range of sites, like Huffington post, that started out strictly as online news and commentary but have at this point become very popular. To select which sites to sample articles from, I would suggest looking at the Alexa traffic ratings for the most popular online news sites and then to use that list to select the news sites that are far and away the most popular.

At this point, the issue now becomes sampling articles from these most popular sites. Here we run into another level of difficulty. Each of the selected sites does not provide a full list of their articles for us to sample from. At this point, it would be ideal if we could have such a list and randomly select articles from it, but as we do not, we are going to make further compromises. In this case, I would suggest building a list of all of the articles that appear on the homepages of each site. Many of these sites have massive amounts of stories linked off their homepages, in many cases as much as 40 articles. So we can use the homepages as the basis to create a list of articles.

This still leaves another sampling issue. We need to decide on what days and times to collect the homepages. Based on our class discussion, I think it would be prudent to take homepages to create two constructed weeks of homepages, sampling two random days from the last year for each of these. With those days in hand, we could visit the Internet Archive to pull up the homepages and build the list of articles. If we take 40 as an estimate of how many articles this is likely to generate per homepage, and the fact that we would be pulling 14 homepages for each of what would likely be about four to six major online news sites we would end up with something on the order of two to three thousand articles. At this point, we could then treat this

constructed list of articles as a frame to sample from to get the desired number of individual articles that we want to code. Depending on the variables we were interested in coding for, and the kinds of statistical tests we were interested in running, we would then randomly select from this list to find the individual articles we would include in our actual coding. I imagine we would want to do something like selecting around 100 articles to work from as this would allow us a large enough N value for a range of statistical comparisons between sub groups.

As evident in my discussion so far, I would be suggesting that the ultimate unit of sampling would be the online news article. These are the individual composed units of content that contain the visual representations of women we are interested in studying. Once we selected the articles we could then code the individual images that feature women included in each article.

Unit of coding: The primary unit of coding would end up being the individual visual representations of women in each article. So we would actually spend most of our time coding the actual images of females included in the articles. As the images are part of the articles, however, it is important to maintain their context as to allow us to ultimately treat the images as independent and as features of the articles.

Variables to code for: There are many potential variables to code the visual representations of women for. The first things that come to mind are evident race (black, white, Hispanic, etc) evident general age group (teen, young, elderly, etc) what part of them is visible (head shot, upper body, full body), facial expression (smiling, frowning etc). These are just a few ideas, as the initial question comes without any kind of context it is difficult to give a particularly useful answer.