

Rebecca Goodwin
rj2339

One scary use of this type of information is the possibility of de-anonymizing data by intersecting different data sets (as we talked about in class.) For instance, one source may release data that has had names scrubbed, but contains both birth date and zip code. Another data set may contain birth date and zip code, but names as well. Combining these two data sets, someone wanting to find out information from a supposedly anonymized data set (the first one) has a fairly high likelihood of being able to know the names of people in the first data set. This is because there are not that many people in the same zip code who also have the same birthday. Another scary possibility that is very real in our current society is the possibility of the government spying on citizens, perhaps by demanding that email, phone, or other technology companies turn over their customer data. This idea is especially frightening since the average citizen probably does not know that her information is being tracked. While we generally trust that the government has its populace's best interests in mind, is a bit disconcerting to imagine what would happen if the government took a more repressive turn. Police could arrest citizens preemptively for, say, having a particularly suspicious combination of visited locations, people contacted, and Google search terms. Even in these actions were all individually perfectly legal and unsuspicious, the power of big data to connect the dots and identify patterns could point to a high probability of a person planning a revolt against the government.

One positive use for this type of data is the customization of experience for a particular user. For instance, Google learns about particular users' search history and customizes their search results based on what it believes a particular user wants. This is good for the user because she now will get search results more relevant to her interests. Another positive use for this type of data is predicting global conflicts. I read an article about a supercomputer that was fed blogs and articles to determine the social climate of a particular area. It found indications of a negative mood in Egypt right before the Gulf War and the Arab Spring in Egypt. So far, it has only worked to "predict" the past, but one can foresee a future in which data analysis becomes good enough to predict conflicts in time to react to them.