CSC2552 Assignment 2

Instructions

• Your work must be your own.
• This assignment is due at 9pm on Monday, April 6 (hard deadline).
• Your assignment must be typed. You may use any word-processing software you like (e.g. LaTeX).

Problem 1 [20 pts]
In order to determine whether small initial successes lock in or fade away, van de Rijt and colleagues (2014) intervened into four different systems bestowing success on randomly selected participants, and then measured the long-term impacts of this arbitrary success. Come up with two other systems in which you could run similar experiments. Evaluate these two systems in terms of issues of scientific value, algorithmic confounding, and ethics.

Problem 2 [20 pts]
There has been substantial debate about experiments using participants recruited from MTurk. In parallel, there has also been substantial debate about experiments using participants recruited from undergraduate student populations. In 1-2 short paragraphs, compare and contrast using Turkers and undergraduates as research participants. Your comparison should include a discussion of both scientific and logistical issues.

Problem 3 [15 pts]
A new experiment aims to estimate the effect of receiving text message reminders on vaccination uptake. A total of 150 clinics, each with 600 eligible patients, are willing to participate. There is a fixed cost of $100 for each clinic you want to work with, and it costs $1 for each text message that you want to send. Further, any clinics that you are working with will measure the outcome (whether someone received a vaccination) for free. Assume that you have a budget of $1,000. Under what conditions might it be better to focus your resources on a small number of clinics and under what conditions might it be better to spread them more widely? Explain in 2 sentences.

Problem 4 [20 pts]
In class, we focused on crowdsourcing (or “human computation”) approaches to mass collaboration. Of the papers we read in class (excluding papers from the Mass Collaboration week), pick one that you think could have benefitted from including a crowdsourced component. Describe the limitation of the paper that you think crowdsourcing could address, describe your crowdsourcing approach, and describe how your approach addresses the limitation.

Problem 5 [25 pts]
Banksy, known for politically-oriented street graffiti, is one of the most famous contemporary artists in the world. But his precise identity is a mystery. In 2008, the Daily Mail newspaper published an article claiming to identify Banksy’s real name. Then, in 2016, Michelle Hauge, Mark Stevenson, D. Kim Rossmo and Steven
C. Le Comber (2016) attempted to verify this claim using a Dirichlet process mixture model of geographic profiling. More specifically, they collected the geographic locations of Banksy’s public graffiti in Bristol and London. Next, by searching through old newspaper articles and public voting records, they found past addresses of the named individual, his wife, and his football (i.e., soccer) team. The author’s summarize the finding of their paper as follows: “With no other serious ‘suspects’ to investigate, it is difficult to make conclusive statements about Banksy’s identity based on the analysis presented here, other than saying the peaks of the geoprofiles in both Bristol and London include addresses known to be associated with [name redacted].”

(a) In 1–2 paragraphs, assess this study using the principles and ethical frameworks in the Ethics chapter of *Bit By Bit*.

(b) The authors included the following ethical note at the end of their paper: “The authors are aware of, and respectful of, the privacy of [name redacted] and his relatives and have thus only used data in the public domain. We have deliberately omitted precise addresses.” Does this change your opinion of the paper? If so, how? Do you think the public/private dichotomy makes sense in this case? Discuss in 2–3 sentences.