Problem Set 3: Study Design

1. **Parameters and statistics.** Identify which value represents the sample mean and which value represents the claimed population mean.

   a. American households spent an average of about $52 in 2007 on Halloween merchandise such as costumes, decorations and candy. To see if this number had changed, researchers conducted a new survey in 2008 before industry numbers were reported. The survey included 1,500 households and found that average Halloween spending was $58 per household.

   b. The average GPA of students in 2001 at a private university was 3.37. A survey on a sample of 203 students from this university yielded an average GPA of 3.59 a decade later.

2. **Air pollution and birth outcomes, scope of inference.** Researchers collected data to examine the relationship between air pollutants and preterm births in Southern California. During the study air pollution levels were measured by air quality monitoring stations. Length of gestation data were collected on 143,196 births between the years 1989 and 1993, and air pollution exposure during gestation was calculated for each birth. (Ritz et al. 2000)

   a. Identify the population of interest and the sample in this study.

   b. Comment on whether or not the results of the study can be generalized to the population, and if the findings of the study can be used to establish causal relationships.

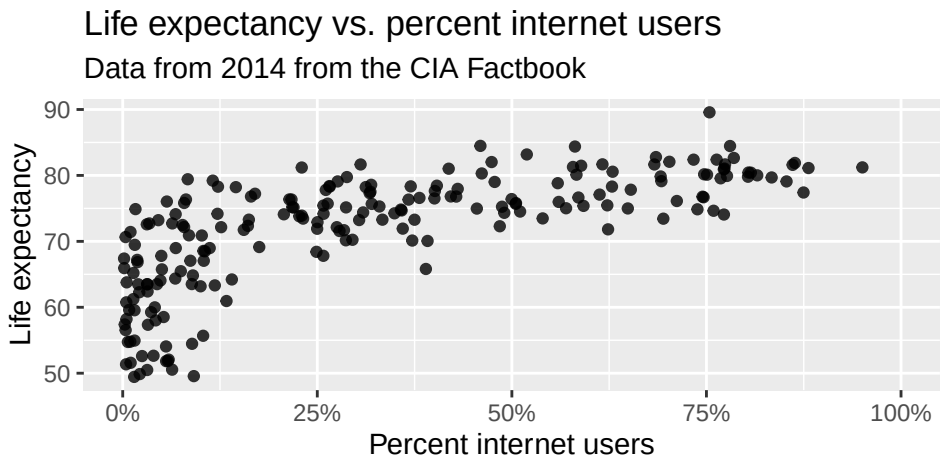
3. **Flawed reasoning.** Identify the flaw(s) in reasoning in the following scenarios. Explain what the individuals in the study should have done differently if they wanted to make such strong conclusions.

   a. Students at an elementary school are given a questionnaire that they are asked to return after their parents have completed it. One of the questions asked is, *"Do you find that your work schedule makes it difficult for you to spend time with your kids after school?"* Of the parents who replied, 85% said *"no"*. Based on these results, the school officials conclude that a great majority of the parents have no difficulty spending time with their kids after school.

   b. A survey is conducted on a simple random sample of 1,000 women who recently gave birth, asking them about whether or not they smoked during pregnancy. A follow-up survey asking if the children have respiratory problems is conducted 3 years later. However, only 567 of these women are reached at the same address. The researcher reports that these 567 women are representative of all mothers.

   c. An orthopedist administers a questionnaire to 30 of his patients who do not have any joint problems and finds that 20 of them regularly go running. He concludes that running decreases the risk of joint problems.

1. **Internet use and life expectancy.** The following scatterplot was created as part of a study evaluating the relationship between estimated life expectancy at birth (as of 2014) and percentage of internet users (as of 2009) in 208 countries for which such data were available.[1]



   a. Describe the relationship between life expectancy and percentage of internet users.

   b. What type of study is this?

   c. State a possible confounding variable that might explain this relationship and describe its potential effect.

2. ***Vitamin supplements.** To assess the effectiveness of taking large doses of vitamin C in reducing the duration of the common cold, researchers recruited 400 healthy volunteers from staff and students at a university. A quarter of the patients were assigned a placebo, and the rest were evenly divided between 1g Vitamin C, 3g Vitamin C, or 3g Vitamin C plus additives to be taken at onset of a cold for the following two days. All tablets had identical appearance and packaging. The nurses who handed the prescribed pills to the patients knew which patient received which treatment, but the researchers assessing the patients when they were sick did not. No significant differences were observed in any measure of cold duration or severity between the four groups, and the placebo group had the shortest duration of symptoms. (Audera et al. 2001)

   a. Was this an experiment or an observational study? Why?

   b. What are the explanatory and response variables in this study?

   c. Were the patients blinded to their treatment?

   d. Was this study double-blind?

   e. Participants are ultimately able to choose whether or not to use the pills prescribed to them. We might expect that not all of them will adhere and take their pills. Does this introduce a confounding variable to the study? Explain your reasoning.

3. ***Stealers, scope of inference.** In a study of the relationship between socio-economic class and unethical behavior, 129 University of California undergraduates at Berkeley were asked to identify themselves as having low or high social-class by comparing themselves to others with the most (least) money, most (least) education, and most (least) respected jobs. They were also presented with a jar of individually wrapped candies and informed that the candies were for children in a nearby laboratory, but that they could take some if they wanted. After completing some unrelated tasks, participants reported the number of candies they had taken. It was found that those who were identified as upper-class took more candy than others. (Piff et al. 2012)

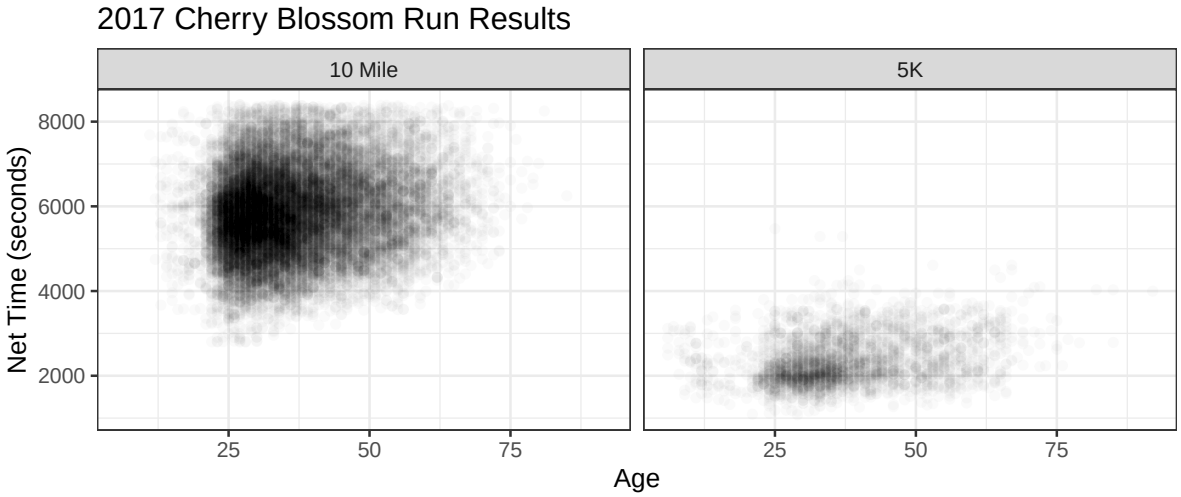
   a. Identify the population of interest and the sample in this study.

[1]The `cia_factbook` data used in this exercise can be found in the **openintro** R package.

b. Comment on whether or not the results of the study can be generalized to the population, and if the findings of the study can be used to establish causal relationships.

4. ***Grammar of Graphics Review.** The data in the graphic below come from the `run17` data set in the `openintro` package. Things to note:

- The points are transparent
- The labs are well formatted
- This graphic has a non-standard theme applied called `theme_bw()`



a. Use your experience with `ggplot2` to recreate the plot.

b. Note three features in the data that are revealed by this graphic and say what they teach us about the 2017 Cherry Blossom Run.

Audera, Carmen, Roger V Patulny, Beate H Sander, and Robert M Douglas. 2001. "Mega-dose vitamin C in treatment of the common cold: a randomised controlled trial." *Medical Journal of Australia* 175 (7): 359–62. http://www.openintro.org/redirect.php?go=textbook-vitamin_C_cold_treatment_2001.

Piff, P. K., D. M. Stancato, S. Côté, R. Mendoza-Denton, and D. Keltner. 2012. "Higher social class predicts increased unethical behavior." *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1118373109.

Ritz, B., F. Yu, G. Chapa, and S. Fruin. 2000. "Effect of Air Pollution on Preterm Birth Among Children Born in Southern California Between 1989 and 1993." *Epidemiology* 11 (5): 502–11. http://www.jstor.org/stable/3703990.