**Statistics Sample**

**A researcher needs your statistical expertise to evaluate an intervention. The six-month intervention was to give subjects a special exercise plan.**

**The researcher randomized 2,500 people to the treatment group (which receives the intervention) and another 2,500 to the control group (does not receive the intervention). The physician wishes to evaluate the subjects’ health outcomes (weight and self-rated health) before the start of the intervention and immediately after the intervention.**

***The physician wants to know whether the new exercise plan will affect one’s overall health (measured by change in weight and self-rated health) differently for those receiving the new exercise plan versus those not receiving the exercise plan.***

**For all questions below, please provide all code (preferably with comments).**

1. Create an analytic dataset by appropriately combining the 4 data files, and provide code for how you arrived at your analytic data set. Describe your workflow.

2. Explore and analyze the data as you see fit. Please show *at least* one plot.

3. Explain what you did in #2 and tell us why you decided on that method.

4. Explain the results to the investigator. Please assume the investigator has limited statistical knowledge. Tell her whether there is or is not a treatment effect and if so, how the two groups differ with respect to their weight profiles and SRH over time.

**What-If Scenarios (brief responses)--**

**5.** What if the investigator wanted to evaluate 16 additional outcomes (in addition to weight and SRH)? She is curious to see if the exercise regime has an impact on body fat %, mental well-being, bone density, etc... She intends to publish the results in a high-impact journal. How would you respond to this? Dimensionality reduction or check the correlation between the data

6. What if the two groups (treatment and control) were not randomized? How might you evaluate the intervention now? What additional information might you require, and which methods could you use? What would be your limitations?