

HW3: Research Reproducibility I

INSTRUCTIONS

This assignment has two parts.

Part 1: Complete the tasks below and turn in by the deadline posted on Blackboard. Include the question 1 text as a response to a survey entitled HW3: Part 1 that is located under assignments in Blackboard.

Part 2: Complete the peer review assessment for two of your classmates RMD files by the deadline posted on Blackboard. The peer review assessment can be found in the peer materials subfolder of the course materials folder.

1. Johnathan Doe (aka JD), is a new Brown School MPH graduate whose first job is as a criminal investigative epidemiologist working for *Epidemiology detectives R us* in Key West, Florida. There is currently an outbreak of an unusual disease among tourists in Key West characterized by lassitude, dehydration, and morning nausea that is going by the name *LDN*. The outbreak was suspected and called into *Epidemiology detectives R us* by a local ER doc. Colonel Bombadini (as he is known) is JD's boss and has been down in Key West working as the chief epidemiologist for 30 years.

Normally things are pretty low key in Key West so the colonel is pretty nervous about the situation and what it will do to tourism (particularly the business at the largest Key West beach bar that he owns and runs on the side). He has put JD on the investigation and told JD that this is an extremely urgent matter that must be solved quickly. JD suspects from the symptoms that have been described to him by the ER doc that they might be dealing with some tainted alcohol being used to make rum runners, the most popular drink in Key West. Using his epidemiology chops, the first thing JD does is design a case-control study where he interviews tourists with LDN at the emergency room and a set of controls without LDN who have shown up for other ailments and injuries (sunburns, jet-skiing injuries, insect bites, etc.). He determines that indeed the cases have consumed rum drinks (including rum runners) at a much higher rate than controls but as a newbie epidemiologist he isn't quite sure that the OR of 3.0 he obtained is correct or if he set up his code right. He has a panic attack that comes and goes because he can't help but worry that maybe he set up the code wrong and that rum drinks are actually protective against LDN. He doesn't want to tell the Colonel about this until he is certain that his results are correct.

Meanwhile, a similar outbreak has been reported over the wire to be occurring on the island of Grand Cayman. He consults with Barbara at the Grand Cayman *Epidemiology detectives R us* office that has also conducted a case-control investigation of their own and found similar results.

Given the similar results, Barbara and JD discuss as a next step looking into the origin of the rum to conduct further investigations into whether there is a certain rum that is being imported into Key West and Grand Cayman that is tainted.

Using this example, describe how the concepts of *reproducible*, *repeatable*, and *replicable* can be applied to determine with additional certainty that rum drinks are making tourists sick in Key West and Grand Cayman.

2. Make an RMD (Notebook) file from your R script code from last week using the following guidelines:

- a. Implement promising practices 1-9 as applicable with the goal that your code and data sources will be easy to follow for peer review
- b. For all questions 1-10 last week include them as level 2 headers prior to each code chunk
- c. Knit your RMD NB code file that includes your result output to an HTML file for submission for peer review

You will be peer and instructor graded based on the readability and reproducibility of your code.

A medal will be awarded for the most and least efficient code.