Reading: “Chapter 12 – Evidence Based Data Analysis” by Roger Peng

1. Reproducibility focuses on the validity of data analysis, and whether certain analyses can be reproduced. On the other hand, replicable is related to address the validity of a scientific claim and if the results of a study or experiment can be replicated with the same conclusion.
2. A). Some benefits of reproducible studies are transparency, data and code for others to analyze, and an increased rate of transfer of knowledge.

B). A limitation to these benefits is that reproducibility does not guarantee validity or correctness of an analysis.

1. “Evidence-based data analysis” is explained in the text to be that for each step in the process of an analyses is backed up by the best method, with statistical research providing evidence of one method over another.
2. The most concerning aspect of the errors in Potti’s research is that it was going to be used to guide therapy in a clinical trial, which put subjects’ lives at risk.
3. Two ways bins can be determined in a “evidence-based histogram” are using Sturges’ formula for the number of bins or David Scott’s method which justified bandwidth/bin size based in integrated mean square criteria. Both options are backed by research and evidence supporting why it should be used.
4. The most important advantage of “evidence based reproducible research is that it guarantees that generally accepted approaches are used to analyze data.
5. “Benchmarking a method” is referring to a using a set of methods on a dataset as a comparison tool for when new methods that improve on the previous methods are made.