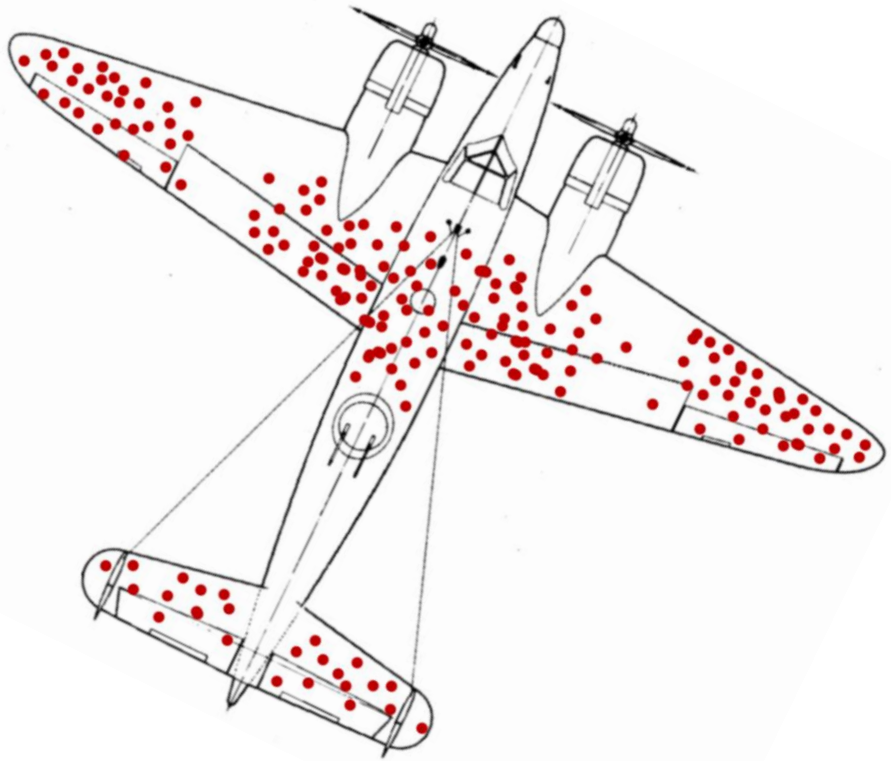
More on Bias: Survivorship Bias in Subsurface Modeling?

Michael Pyrcz, University of Texas at Austin (@GeostatsGuy)

Example shared in my Introduction to Geostatistics class by @uddhav_marwaha (Twitter).

Survivorship Bias: a form of selection bias resulting from selecting samples that "survived" some previous selection process. This often leads to false conclusions. For example, in WWII the Center for Naval Analyses (@CNA_org Twitter) compiled a dataset of bomber damage to assess where reinforcement was needed. Statistician Abraham Wald recognized this was a case of survivorship bias. The plans shot in critical locations did not return to base. Wald suggested reinforcement of locations that were not damaged in planes that safely returned to base!
(https://en.wikipedia.org/wiki/Survivorship_bias#In_the_military)

Is there preselection in our subsurface datasets? For our subsurface projects do we only sample: success cases, producing wells, drill holes with economic ore grades, large fields, clastic depositional settings, marine seismic surveys, high resolution 3D seismic surveys, shallow reservoirs etc. When we pool samples, check for preselection and ensure this is considered in the resulting inferences and decision to export these results. The samples must be representative of the population to which we will apply our model. Of course, this applies to any datasets.



Hypothetical dataset of aircraft damage for planes that returned to based. Source https://en.wikipedia.org/wiki/Survivorship_bias#/media/File:Survivorship-bias.png