Data Science 1

STAT/CS 287

James Bagrow, UVM Dept of Math and Statistics

LECTURE 02

Announcement

Course Teaching Assistant:

Damin Zhu ~ **damin.zhu@uvm.edu**

Help with questions and will hold office hours (Tuesdays 2–3pm)

Today’s Schedule

Intro/Review of probability and statistics

Reading assignment:

Review the rest of today's lecture notes when posted (including properties of expectation, variance)

Motivation

Abraham Wald
1902–1950

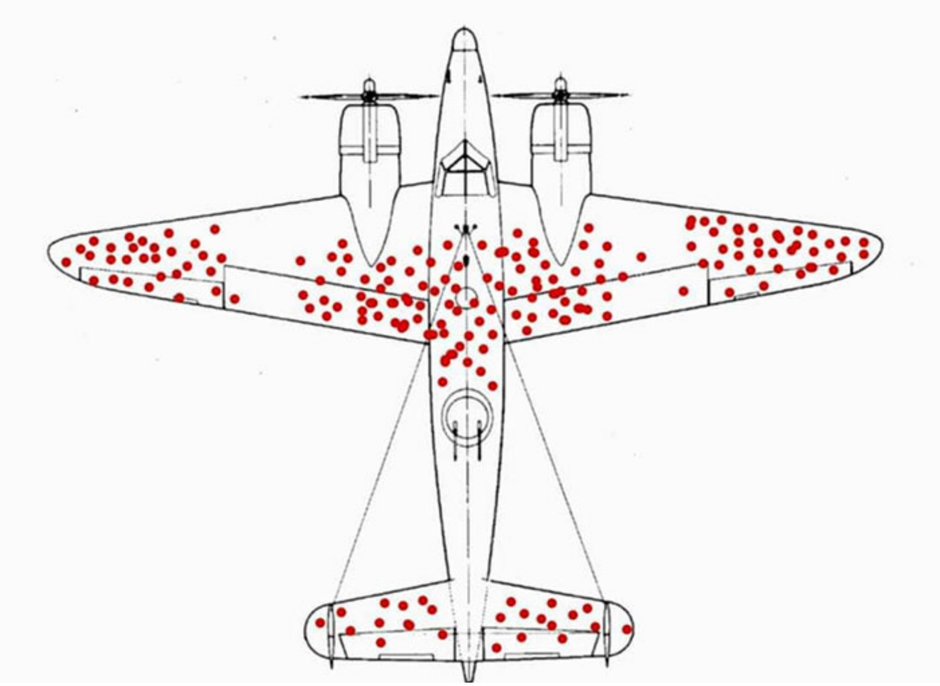
Motivation

Abraham Wald
1902–1950

During WW2: help planes survive combat
→ ***Where* to put bulletproof armor?**

Motivation

Abraham Wald
1902–1950

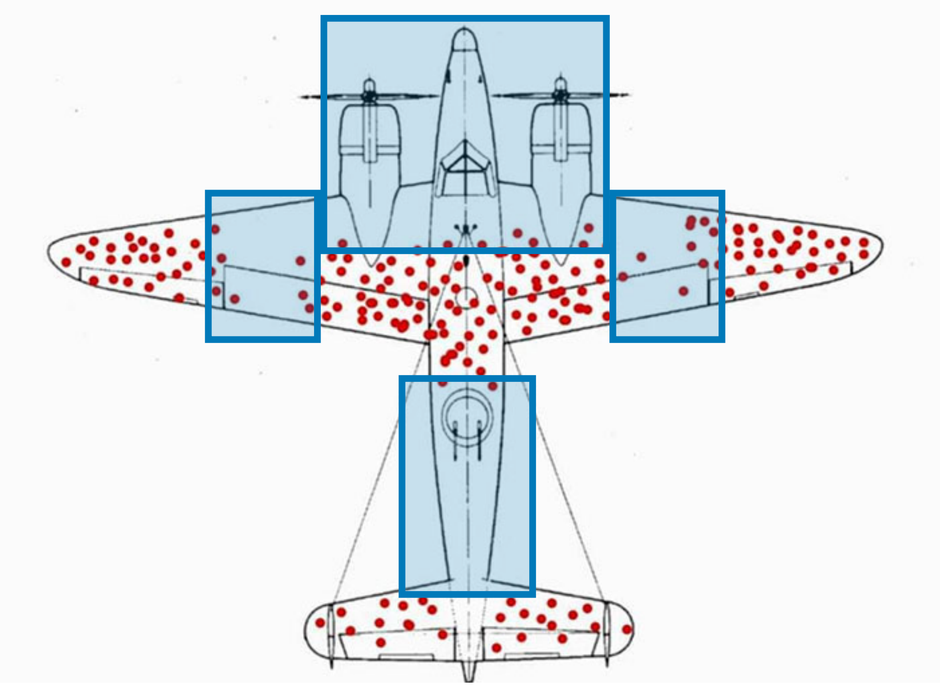


Went to airfields and measured damage on many planes after missions

Drew an interesting conclusion:

Motivation

Abraham Wald
1902–1950



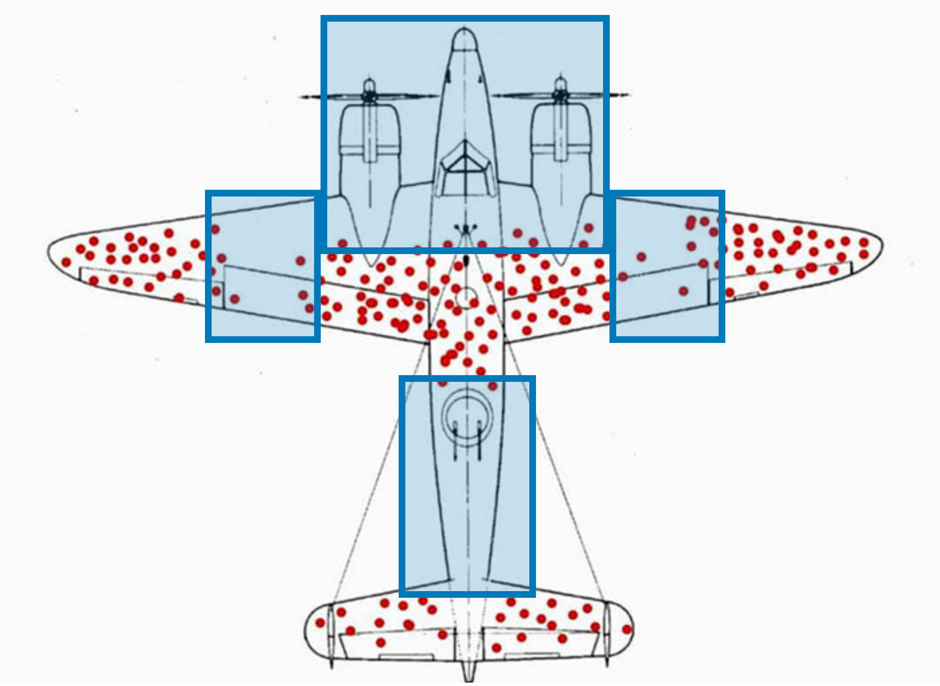
Went to airfields and measured damage on many planes after missions

Drew an interesting conclusion:

Put the armor where the bullet holes **aren't**!

Motivation

Abraham Wald
1902–1950



Went to airfields and measured damage on many planes after missions

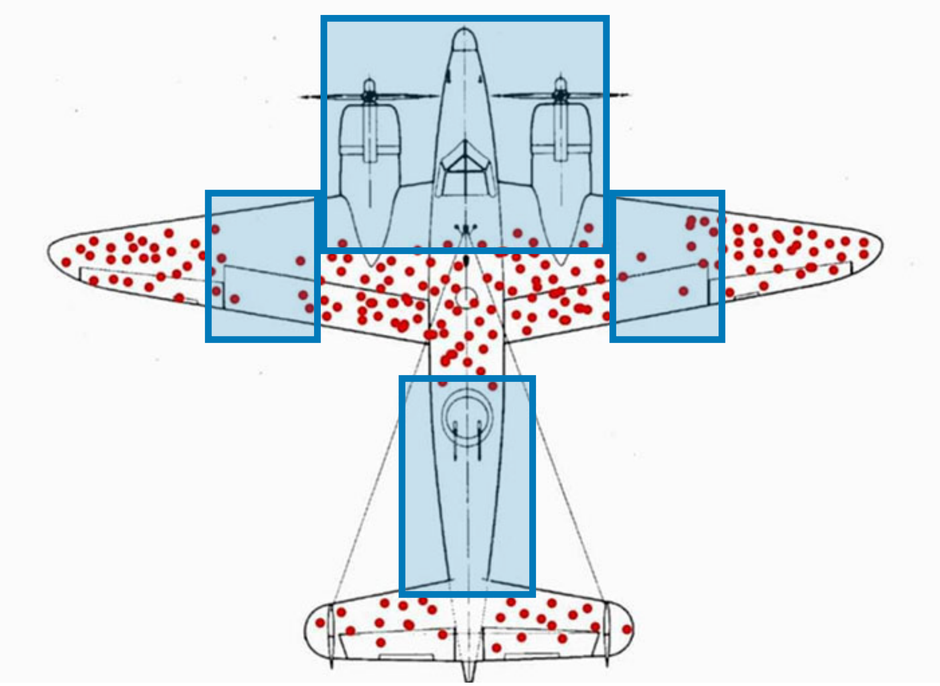
Drew an interesting conclusion:

Put the armor where the bullet holes **aren't**!

Why? Survivorship bias: only *saw the planes that made it back*. His sample of data was conditional!

Motivation

Abraham Wald
1902–1950



Went to airfields and measured damage on many planes after missions

Drew an interesting conclusion:

Put the armor where the bullet holes **aren't**!

Why? Survivorship bias: only *saw the planes that made it back*. His sample of data was conditional!

As data scientists, we should always be reasoning about the *greater context* of the data at hand.
→ Helped by a good grounding of (*conditional*) *probability*

Let's switch to the board