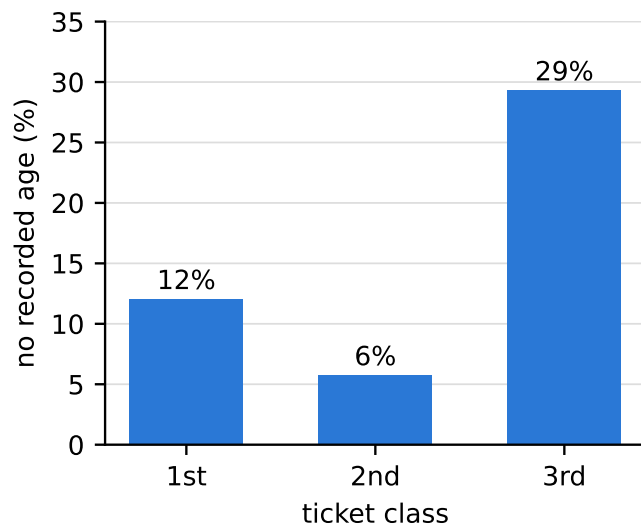


DATA 202 — Missing Data · Week 9 Wednesday

Same passengers as Monday, but all 1,309 of them: 263 have no recorded age. Your professor will tell you when to do each part.

1 · Who lost their age?



Share who survived:

class	age recorded	no age
1st	64%	49%
2nd	44%	25%
3rd	26%	24%

- Which class lost the most ages? _____
- Inside each class, who survived more often: passengers with an age, or without? _____
- With a neighbor: one story that could explain **both** patterns: _____

2 · Name the mechanism

MCAR: pure chance. **MAR**: explained by columns we *can* see. **MNAR**: depends on the missing value itself, or on something never recorded.

For each blank, write MCAR, MAR or MNAR, and why.

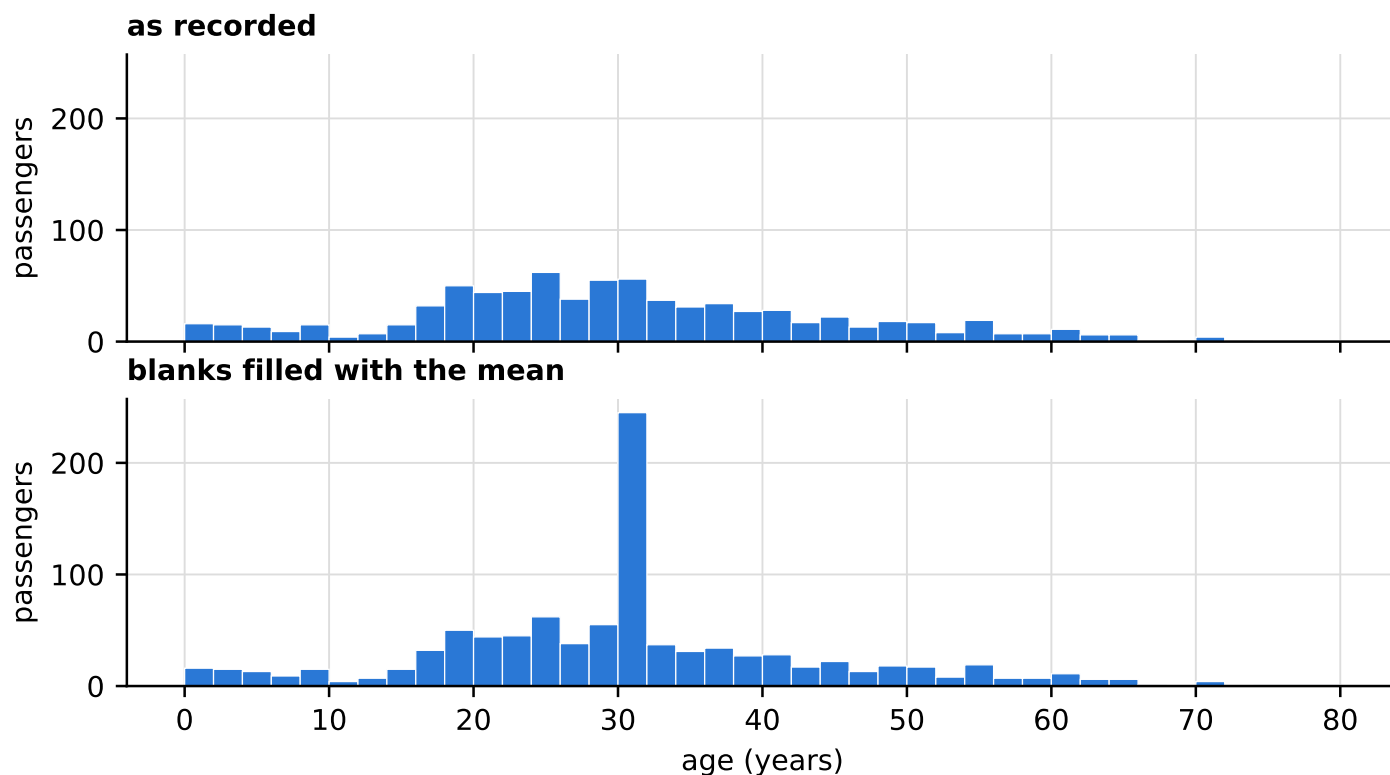
- Two passengers have no port: a clerk smudged their line in the register.

- Age is blank for 29% of 3rd class but 6% of 2nd class; within a class, which ages are blank is luck.

- (*Suppose*) older passengers left their age blank on purpose.

- Cabin is blank for 98% of 3rd class: their tickets rarely listed one.

3 · The spike



The 981 training passengers' ages: as recorded (top), and after filling every blank with the mean age (bottom).

1. **Circle** what the imputer added.
2. How many passengers are now exactly 30.0 years old? _____
3. Standard deviation of age: before _____ years, after _____ years.
4. Median age is 39 in 1st class and 24 in 3rd. Who gets a worse guess from the mean: a missing 1st-class age or a missing 3rd-class age? Why? _____