



Common Data Challenges

Challenge

Example

Missing information across fields	: 90% of cases missing judge info
Use of invalid values	: 30% of charges have invalid class
Inconsistent case outcomes all charges are	: 16% of cases are “open” but disposed
Bad date chronologies	: 11% of cases arrest date precedes disposition
Lack of field parsimony	: 130 unique values within disposition
Poor dataset integrity	: Sealed, expunged, and declined cases permanently removed from dataset
Inconsistent time units recorded in	: Sentence lengths inconsistently days, months, and years within same field



Common Data Challenges

- Lack of category exclusiveness. : Race field categories: Asian, Asian or Pacific Islander, Native Hawaiian or Pacific Islander
- Misdemeanor data are incomplete: Only charging and dispo info is updated. Misdemeanor case closeout lags
- Permanent loss of information. : Key decisions are overwritten with subsequent decisions
- Lots of key info not collected. : Key data not collected: offer and outcomes, def. custody status, case screening decision, drug type, prosecution recommendations, def. indigent and housing status, vic. demographics and relationship to def., flags for special crimes and priors