Since 2014, the Chinese government has taken unprecedented steps in legal transparency, releasing documents related to more than 25 million court cases online. Although the resulting database is far from complete, this new source offers an unusual opportunity to transform our understanding of the Chinese legal system by developing a granular portrait of what happens in courts every day. We offer a first look at the promise and pitfalls of using this new data source, through a hands-on illustration based on a dataset of 697,172 documents from Henan province. Our goals are descriptive (what is available?), prescriptive (what is the best way to use large databases like this one?), and theoretical (what can this data teach us about Chinese institutions?). In particular, we devote special attention to understanding bias in the dataset. This is a critical consideration, not only for the field of Chinese law, but also for anyone poised to dive into the world of text as data.