@Retrosheet @Adi Wyner @Ryan Gross

At the start of this summer, my colleagues and I needed baseball data on the plate-appearance level. Thankfully, Retrosheet is a rich dataset, available for free online, containing every play (moreover, every pitch) for the last century. Retrosheet data, however, exists in an extremely condensed string format, and is difficult to parse. The R package Retrosheet does not help with parsing Retrosheet data, and to my dismay, the R package in fact has errors (you cannot always recover the pitcher of a given plate appearance, and some rows were missing due to errors with `badj` (batting adjustments)).

So, I decided to take care of the dirty work. Rather than having to parse Retrosheet yourself, or use the faulty Retrosheet R package,

I have compiled a cleaned csv file containing every plate appearance from 1990-2000.