**XDF vs CSV**

* **XDF file** = much smaller than CSV file b/c **compressed**
* Main advantage over CSV = can be read + processed much faster (Computation using an XDF file is faster than using data.frame)
* Disadvantage = a format only MRS understands + can work with.
* In order to decide whether we chose XDF or CSV we need to understand the I/O trade-offs involved:
* Converting from CSV to XDF is itself a cost in terms of runtime.
* Once original CSVs are converted to XDFs, runtime of processing (reading from + sometimes writing to) XDFs is lower than if we had directly processed CSVs instead.
* Since an EDA workflow usually consists of cleaning + munging data + feeding it to various modeling + data-mining algorithms, the *initial* runtime of converting from CSV to XDF is quickly offset by the reduced runtime of *subsequently* working w/ XDF files.
* However, one-off kinds of analyses on datasets that’re ready to be fed to the modeling algorithm might run faster if we skip XDF conversion.
* One-off operations are also common in PROD code, such as when a dataset is scored w/ an *already existing model every time new data comes in*.
* In such cases, need to run some benchmarks in order to find the optimal solution.