Sanity check dataset