**Text of some emails from Steve RE these data files**

**First email**

Take a quick look at this and see if it's alright. Not every single column is documented in the codebook, nor are the merge patterns, but it's \*much\* better documented than it was when \*I\* got the data. And I think the experience of reverse-engineering a database is a big part of the job, any how. First step for these students will be for them to devise an E-R diagram and think about what kinds of merges they want to perform to get there.

I've also attached a dataset for each of the potency tests, with each row being a test (so it cannot be used to calculate market averages, which would also require knowing how many products/sales resulted from each tested sample). This is something I made for me but figured I would share.

Take a look at the Introduction and Codebook and see if there is anything that doesn't make sense.

**Second email**

Re timeline: LCB issues data with a full 2 months delay. That runs through the November release, which is what went into the RAND paper. But I can re-run the code to add October; it's not too difficult, just requires running a second computer in the background for a few hours.

Re text analysis: I understand. There is also a "strain name" field which, though we dropped it from our final dataset, because we find it useless, your students can pull in from the inventory table.

Re edible analysis: There's still lots to analyze with regards to edibles but it might involve the more qualitative aspect of googling brand names and product types and then thinking of the appropriate ways to code edibles based on distinctive product categories and how well they're expressed in those text fields.

Remember that this store is on the OR border so someone could attempt to look at changes w/r/t OR policy changes. I can also quite easily give you another store's worth of data (only two new tables would be needed), if you want a control or if you want another way for students to divide and conquer. I didn't do that because I was mindful of file size.

Re testing: I'm not sure analysis of what you describe would work, since it would require for labs to send the same sample to multiple labs. Testing is done before shipping to retailers, right before the construction of the final product.  I don't know if producers shop around for labs in that way, or perhaps a more general way which would not lend to that data analysis. But if you're interested in that, I could provide you a record of all transfers of samples. I could also provide tables relating to plants, "plant derivatives", and "inventory combinations" (i.e. lotting and sublotting), so you could trace samples back to source material, but that's where this stuff gets tough, mostly due to difficulty of doing really big table merges.

On the other hand, there is interesting data analysis work on the lab data to be done. Some of it already has been done, and made quite a splash with regulators, leading to rule changes.

Jim MacRae (not sure I can vouch for him) has been looking at the data and has done high-impact work on labs <http://www.straightlineanalytics.biz/2015/12/labbiness-in-washington-labs-where-is-the-truthiness-when-you-need-it/>

as well as analysis of strains: (3) <http://www.straightlineanalytics.biz/2015/11/cannabis-strain-naming-in-washington/>. I haven't read his work widely but there might be some ideas for studies to reproduce/modify.

**Third email**

In general, keep in mind that I'm withholding the producer-side tables relating to plants, harvests, and lotting, and that I have the files ready deliver if you want more (and harder) space for the students to explore.