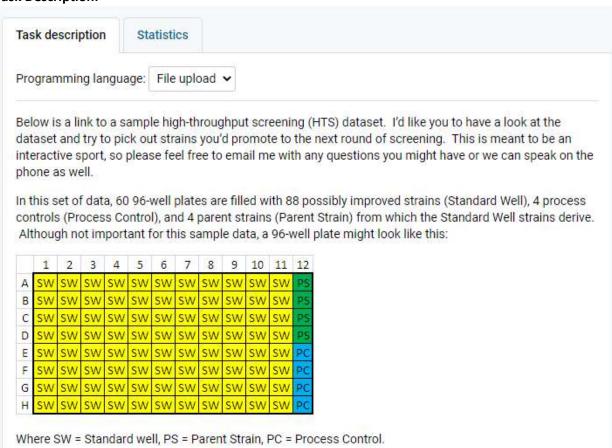**Instructions:**

Please feel free to take all the time you need to complete this and send all your results directly to me at grapov@amyris.com. Please send the results as a reproducible python or R notebook and its rendered .html version. Please do not hesitate to reach out if you have any questions about the challenge data or goals (or anything else).

Task Description:

Statistics

Programming language: File upload ▾

Below is a link to a sample high-throughput screening (HTS) dataset. I'd like you to have a look at the dataset and try to pick out strains you'd promote to the next round of screening. This is meant to be an interactive sport, so please feel free to email me with any questions you might have or we can speak on the phone as well.

In this set of data, 60 96-well plates are filled with 88 possibly improved strains (Standard Well), 4 process controls (Process Control), and 4 parent strains (Parent Strain) from which the Standard Well strains derive. Although not important for this sample data, a 96-well plate might look like this:

	1	2	3	4	5	6	7	8	9	10	11	12
A	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	PS
B	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	PS
C	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	PS
D	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	PS
E	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	PC
F	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	PC
G	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	PC
H	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	SW	PC

Where SW = Standard well, PS = Parent Strain, PC = Process Control.

The goal of our HTS group is to take data like this and pick out standard well strains that have been improved over their parent.

We'd like to get a sense of the way you might go about finding improved strains. We'd like you to try to:

- Pick out improved strains you would promote to the next round of testing.
    - Most Standard Well strains are not improved.
    - There is noise in the assay.
    - What would you promote if you could screen 20, 60, or 180 strains in another round of screening?
- Comment about features you see in the data and how you found them and what you would do with them.
- Send us a few graphical visualizations of the raw or transformed data that you found helpful.

A few things about the data:

- The 60 plates are processed and measured on 3 different robots: Bender, Terminator, and C3P0.
- On each robot, the 20 plates are loaded onto a stacker and are fed into the analyzer one plate at a time. It takes about 3 minutes to measure all wells of a plate.

We've been looking at this sort of data for years now, so we have probably left out some details that are essential for looking at the data in a useful way. As I mentioned above, this is meant to be an interactive process, so please email with any questions or thoughts. If you get stuck, we can talk and try to get you unstuck. We are less interested in your list of hits as we are in the process by which you arrived at that list and the ways you might have visualized the data. Feel free to use any tools that you think you might want to use. If you want the data in a different format, let us know.

Data: Sample data