**Introduction**

Hello, this script will evaluate the percentage success rate for each barcode in your data.

It filters by **gene**, and can also be used to filter by **polymerase**. It will plot these values in a **graph**, and then give you the option to save a **csv file with the % success rates** of every barcode in your filtered set for further analysis/planning.

**You will need:**

* A **csv file** of all amps you may be interested in, with a column headed **'barcode',** a column headed **'sample\_ID'** and a column headed **'pass\_fail'.** If you want to sort by polymerase, you will also need a column headed **'polymerase'**.
  1. Extra columns are not a problem, don't bother trimming them off
  2. Extra amp info for genes you aren't interested in also aren't a problem. Just do a bulk export and use that data.
  3. Data in the barcode column can be formatted however you want (eg 004, 4, 004F-004R) as long as it's the same throughout your csv file. This also applies to Polymerase names (eg GoTaq will be considered different to GOTAQ)
  4. Data in the pass\_fail column needs to be either PASSED or FAILED, as per **your own criteria**.

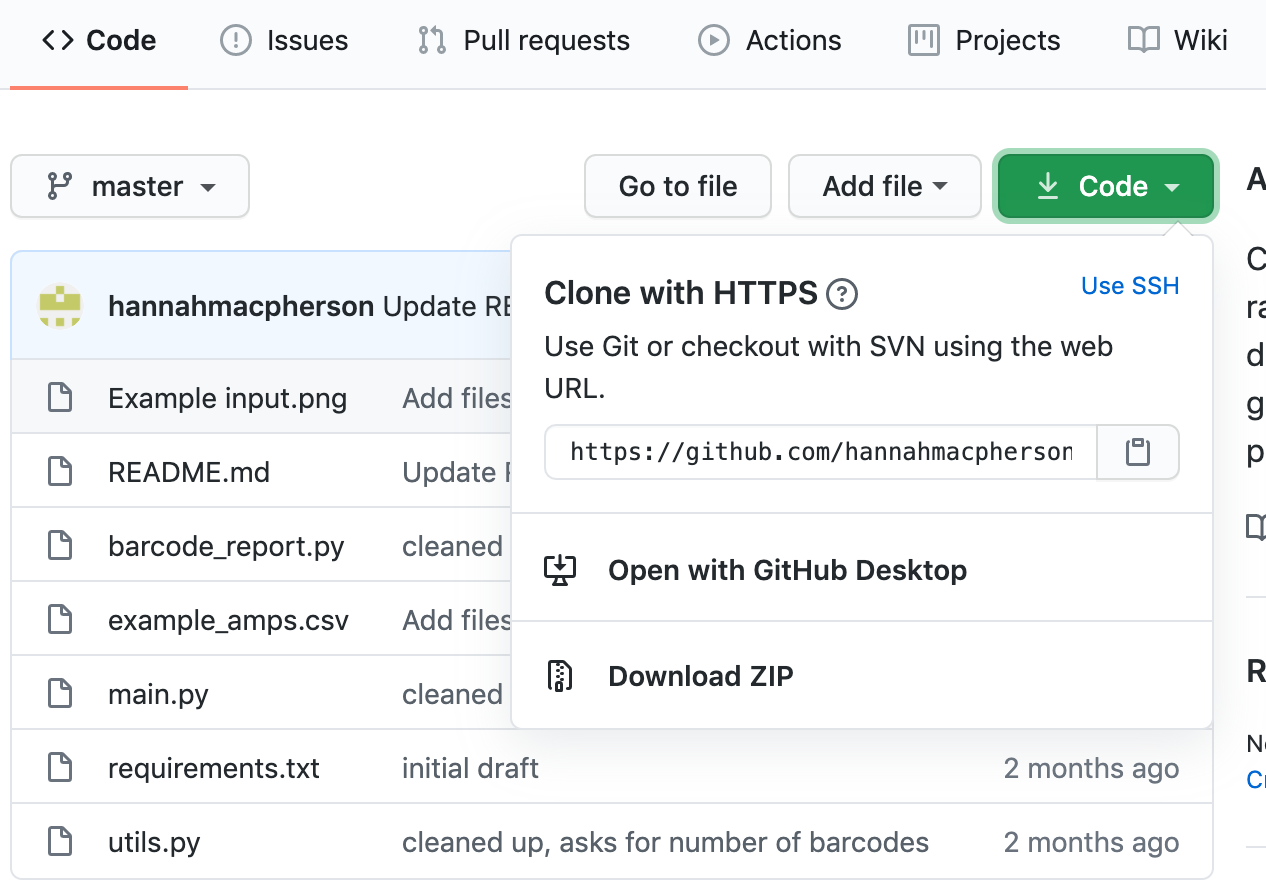
For example:

A screenshot of a cell phone

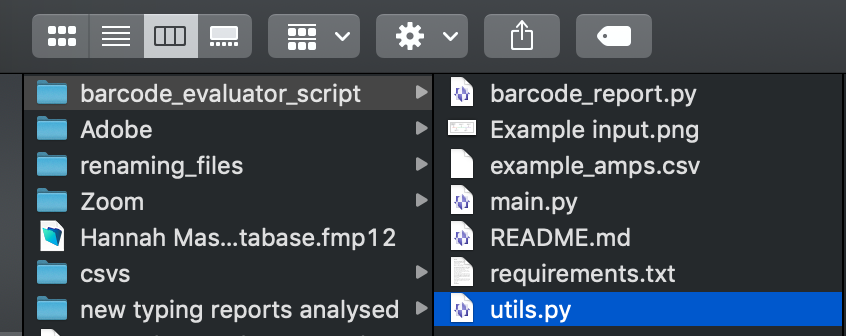
Description automatically generated

**How to set-up the script to run on your machine for the first time:**

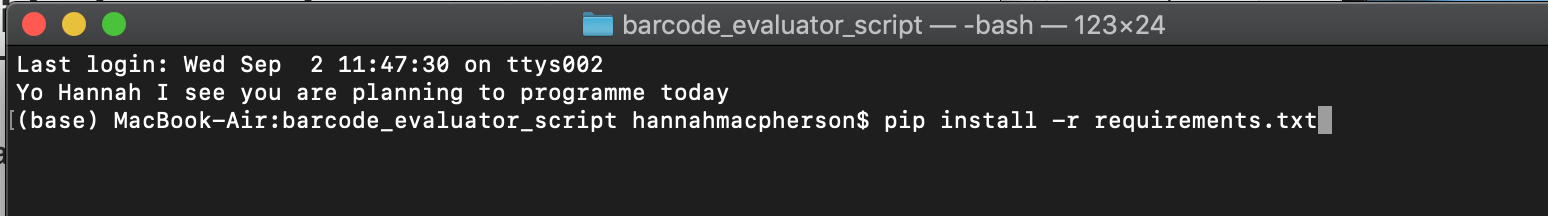
1. Download the folder from GitHub (or onedrive or wherever)



1. Unzip the folder and put it in your Documents (I renamed mine 'barcode\_evaluator\_script', so best to do that too)

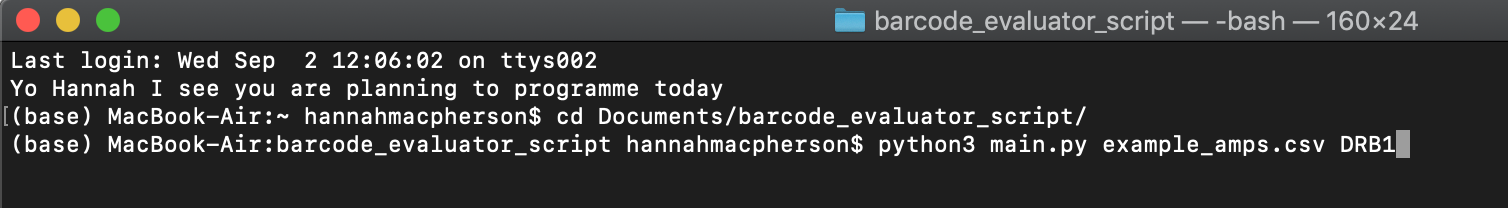


1. Navigate to the folder in Terminal and download the required packages

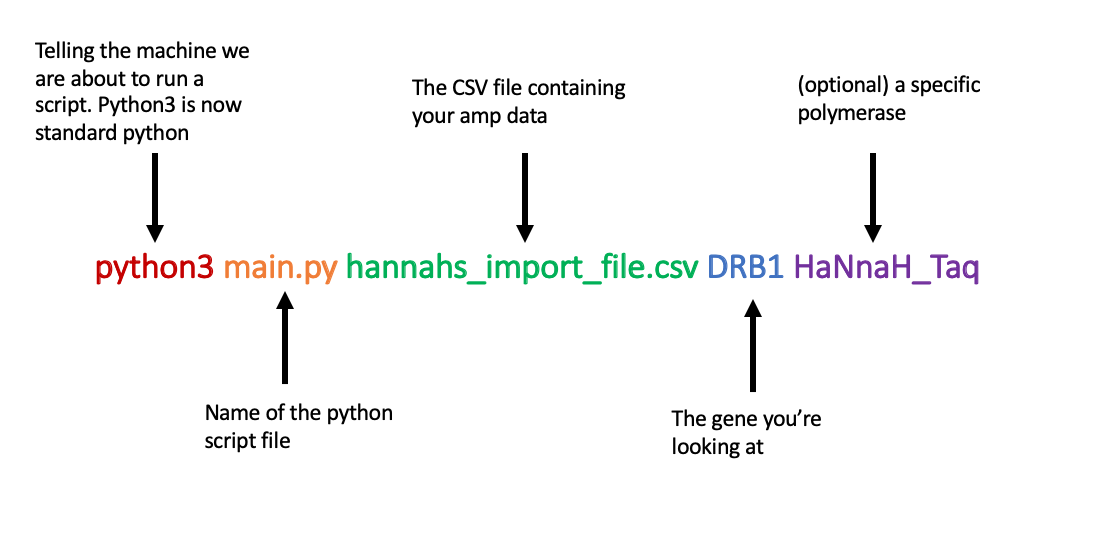


**What to do to run the script each time from now on:**

1. Put your input CSV file in the barcode evaluator folder. Navigate to this folder in terminal and type in what you need from the script (more info below)

[](https://github.com/hannahmacpherson/barcode_evaluator/blob/master/setting%20off%20script.png)

1. The second line should be adapted in the following way:



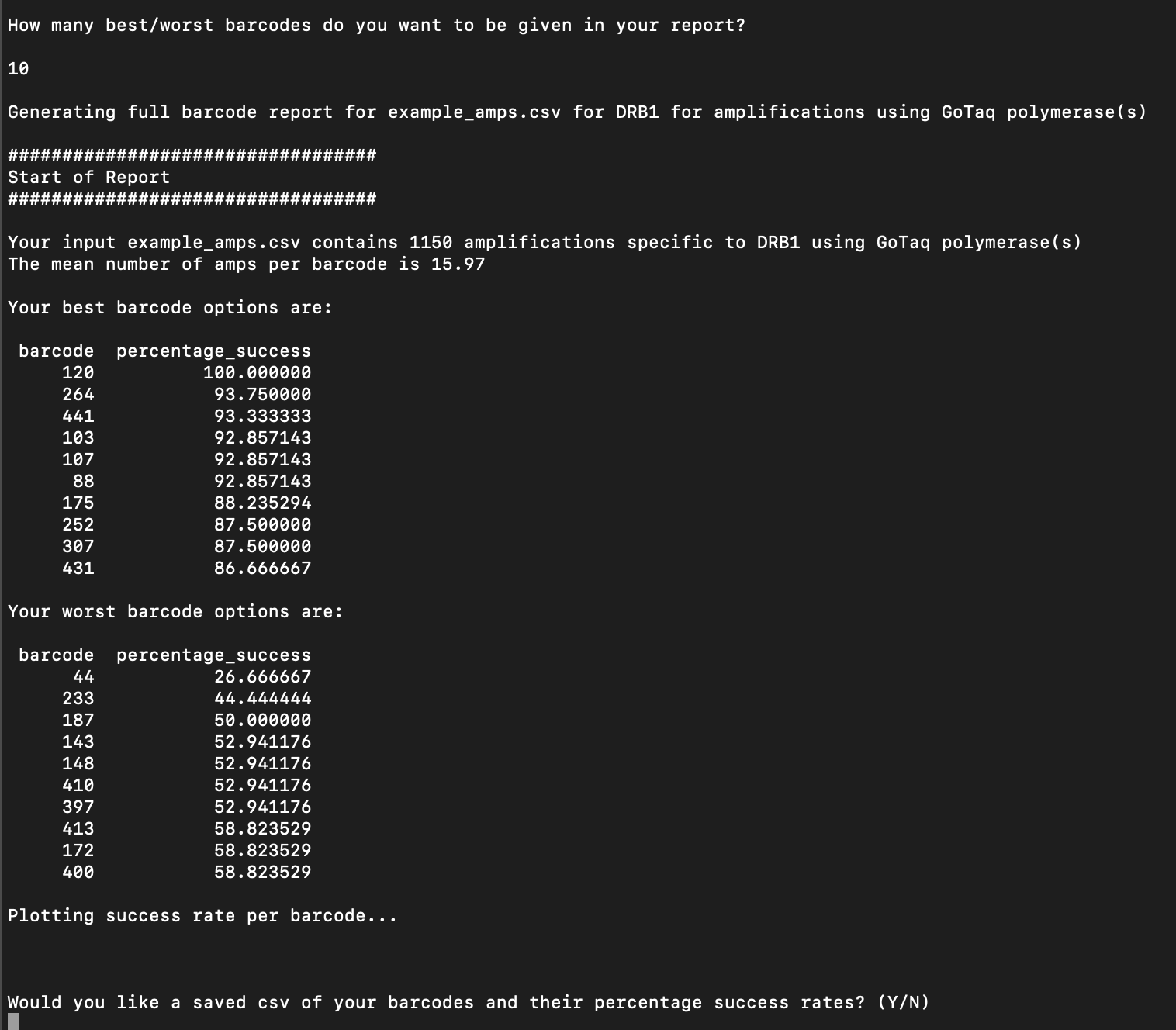
* NB polymerase is optional, if you don't write one then it will default to all data for that gene. Make sure you spell the gene and polymerase as they're written in your input file (case sensitive). You also shouldn't have spaces within the gene or polymerase names\*
* So for example, if I want to find the barcode success rates for DRB1 using a polymerase called HaNnaH\_Taq, I would type:

python3 main.py hannahs\_import\_file.csv DRB1 HaNnaH\_Taq

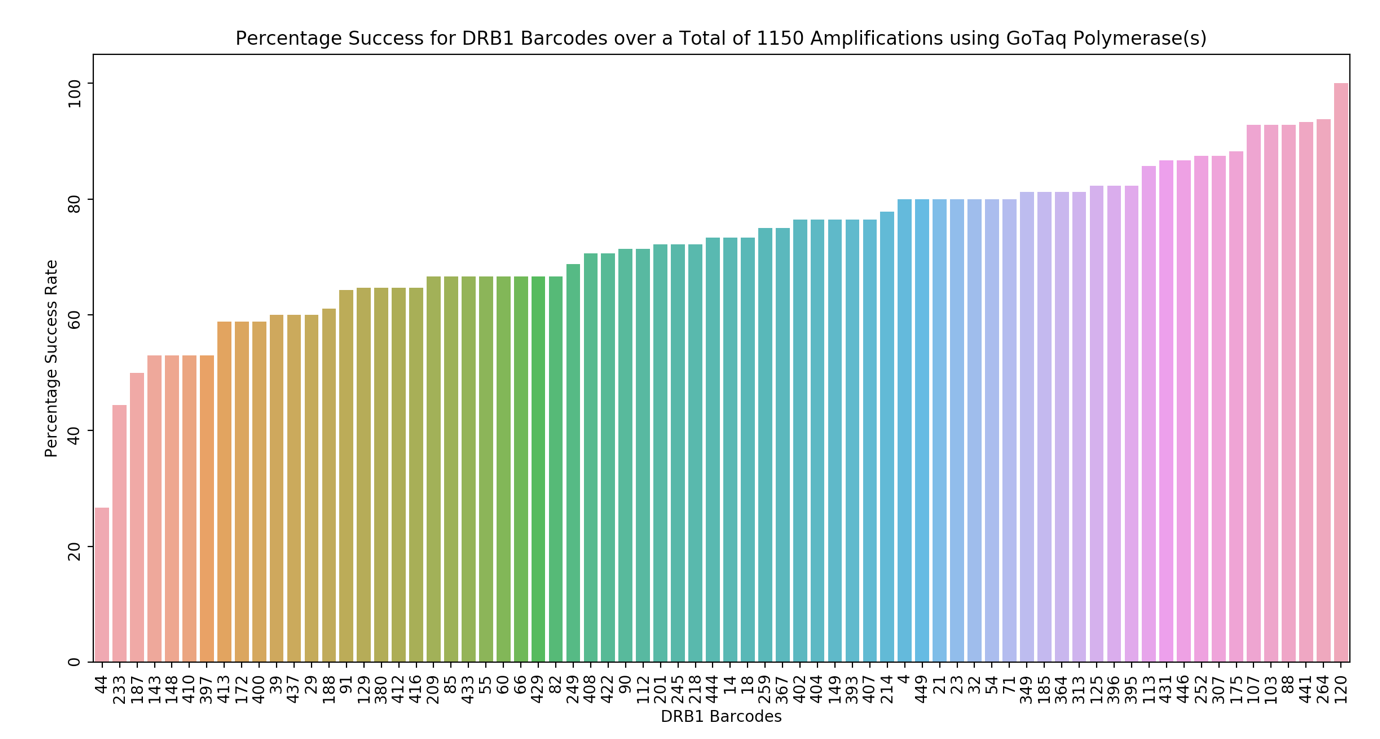
* If I want to find the barcode success rates for DRB1 using all the polymerases I've used in my input file, I would type:

python3 main.py hannahs\_import\_file.csv DRB1

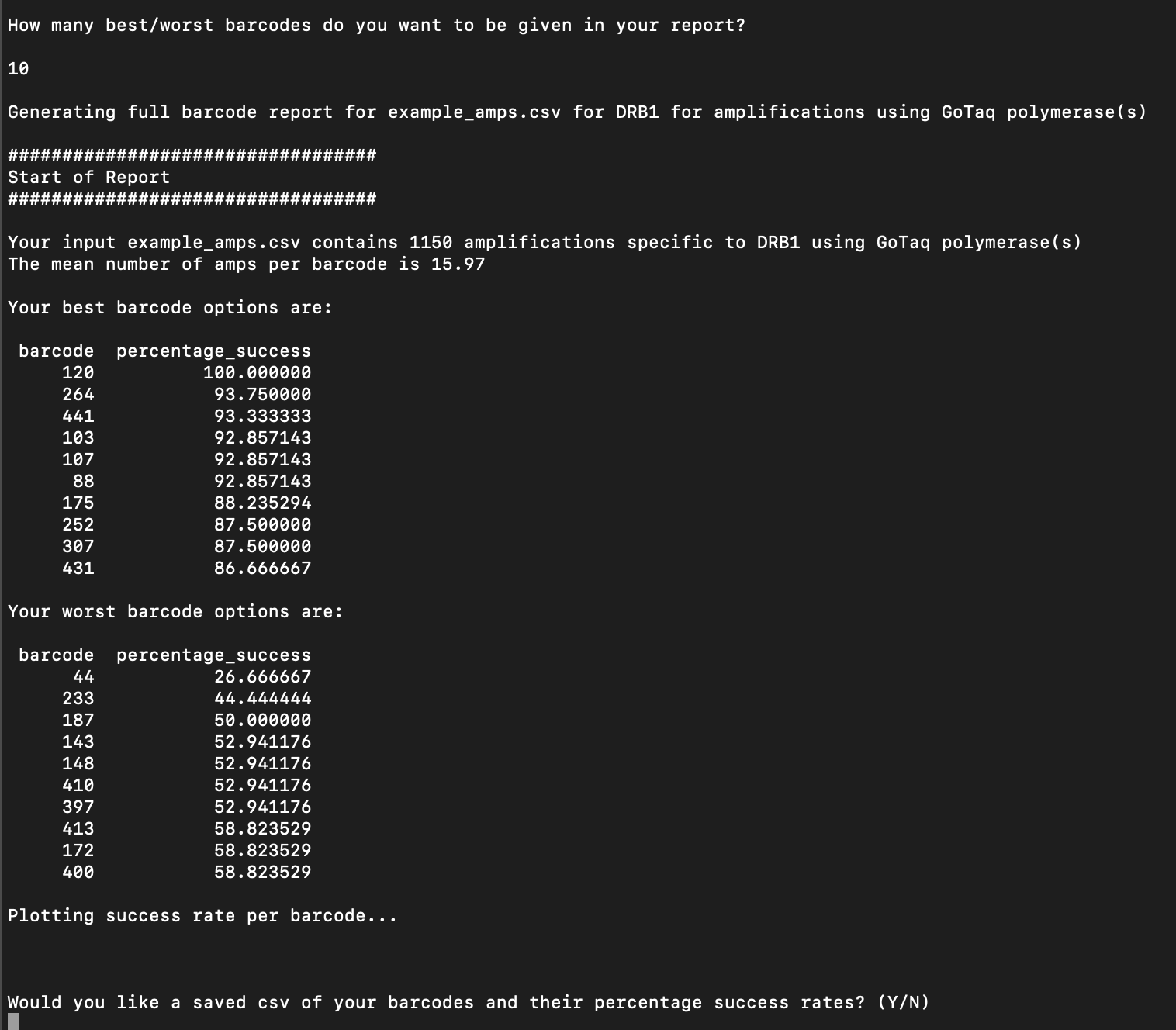
1. You will be asked how many of the best/worst barcodes you want listed.



1. A graph will be plotted (example below). If you want to save it, do that within the pop-up.



1. You'll be asked whether you want to save an additional CSV of every barcode it's used and their % success rate (prompted to type Y or N)



Hannah 04/06/2020