

EDA with Recommendations for CCG Composting

David Snider

8/9/2021

Introduction

The Carolina Community Garden collects weekly surveys on its composting operations. The Garden hopes to use this data to guide composting operations. However, our data collection process is still very much in the beginning phase. The Garden asked me to analyze the survey and the collected data with the goal of improving the survey. I do some basic analysis and make recommendations for how to change the survey, which is in Qualtrics. All changes were implemented except for the last.

Data Cleaning

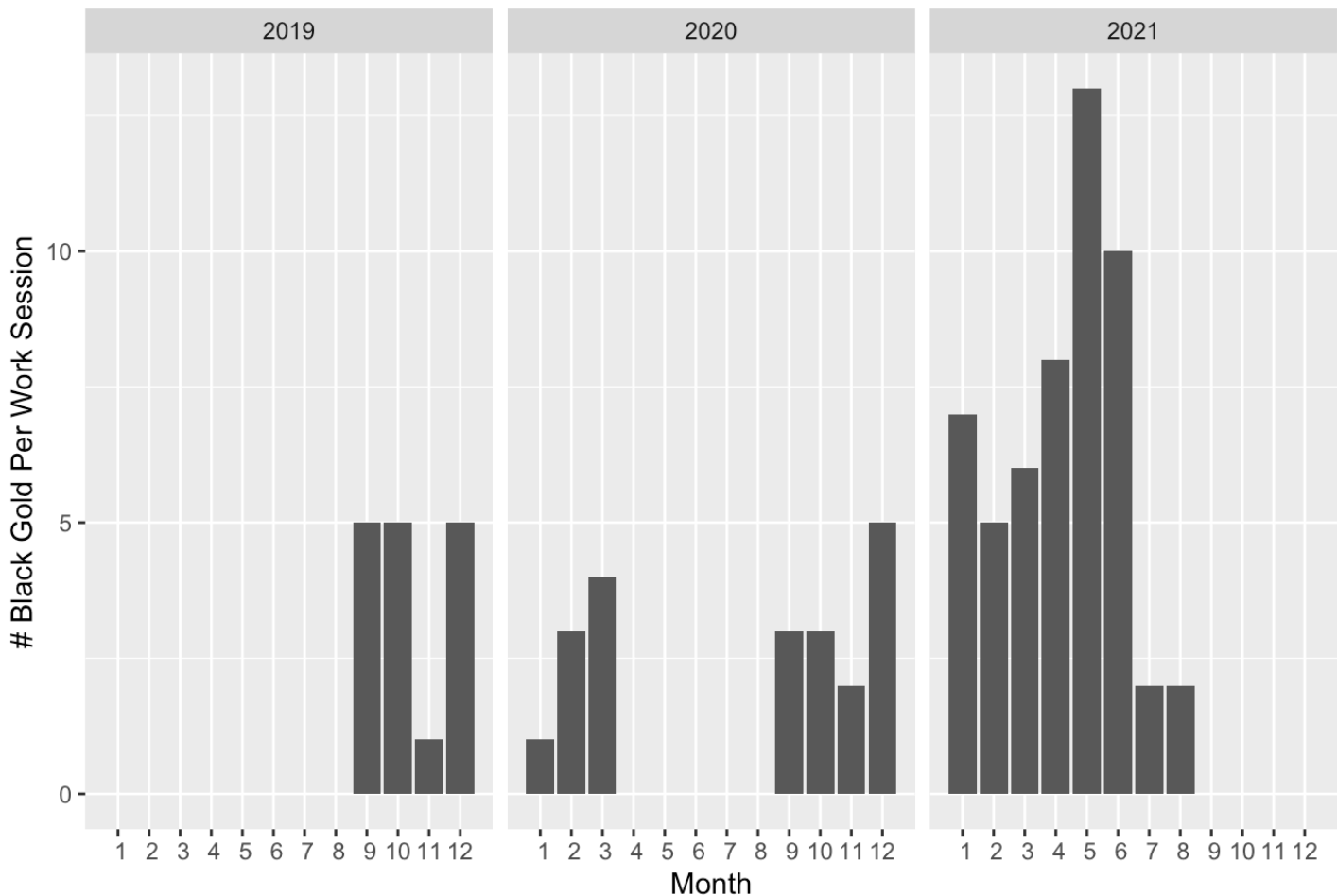
At first the data looked like this, with dimensions of 5 × 43:

StartDate	EndDate	Status	IPAddress
Start Date	End Date	Response Type	IP Address
{“ImportId”:“startDate”,“timeZone”:“America/Denver”}	{“ImportId”:“endDate”,“timeZone”:“America/Denver”}	{“ImportId”:“status”}	{“ImportId”:“ipAddress”}
2019-08-29 09:52:15	2019-08-29 09:53:48	IP Address	107.77.232.197
2019-09-01 13:36:47	2019-09-01 13:39:27	IP Address	174.193.28.33
2019-09-03 13:14:55	2019-09-03 13:15:42	IP Address	99.203.20.248

So we needed to clean it. I removed the first two rows (which were not actual observations), added in missing dates, and tidied the data. Before tidying, each observation was a survey response, corresponding to a particular workday. After tidying, each observation was one workday, which allowed for easier analysis of trends.

For example, it allowed me to make this visualization with ggplot, which could be interpreted as a measure of productivity:

of Black Gold Per Work Session Over Time



Evaluation of Survey Format

The way the data was organized made it rather difficult to analyze the note on each bin. The notes actually describe the activity done on that bin, if you look at the responses that people actually give for that category. Perhaps we should change the name of that category to “activity”, and only allow a few discrete categories. It would make it much easier for the analyst, because otherwise it can be difficult to standardize.

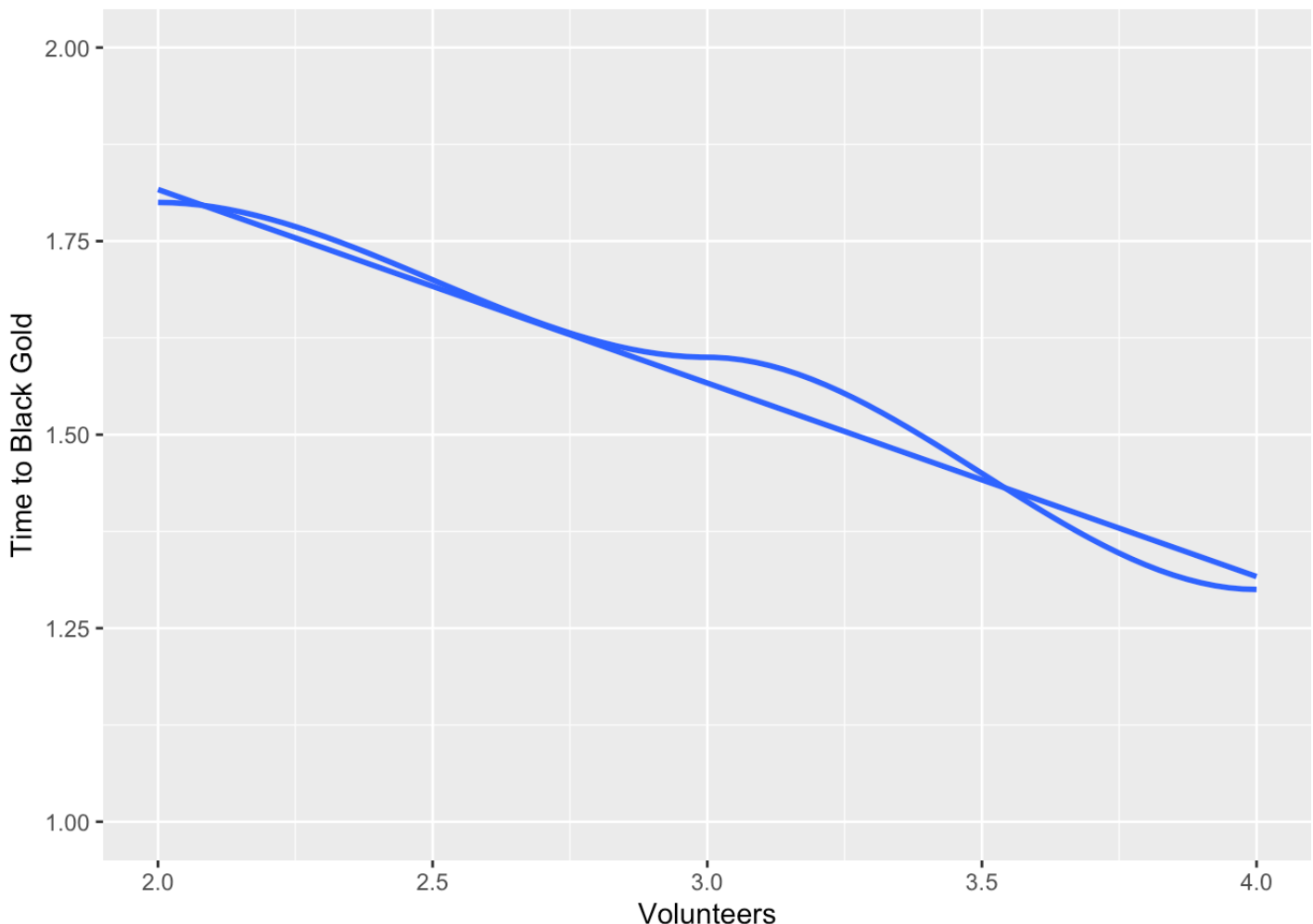
Beyond preliminaries, the visualization above me to question to what extent is increased compost productivity related to warmer weather and to what extent to increased volunteer participation. On the one hand, the numbers seem to be higher during summer months, but the low numbers during July contradict this trend. Simone was gone for the first two weeks of July, and David was gone the first, and stopped coming during the week because of a class conflict. Drained for volunteers, our numbers seem to have dropped.

This discussion is thought-provoking, but it is still speculative. It would be a good idea to quantify this relationship, and we could do so if we collected data on the number of volunteers.

More broadly, it seems that a productive use of the data may be to use information about the weather, and number of compost bins demanded to predict the number of volunteers needed on composting. To do this, we would need to understand both the supply and demand of composting. Supply is a function of our

productivity, which depends on the number of volunteers and the temperature. Demand, however, depends largely on the strategy of the rest of the garden. To measure demand, we could include a section that asks “how much compost was asked for today?” Perhaps the responses could be spaced out in half-bins: 1/2, 1, 3/2, 2, etc. to create basic categories for simple analysis. Moreover, we could encourage the garden to ask for compost liberally—that is, even if we don’t have it. That way we could obtain data to help us understand how we could improve to support the operations of the rest of the garden.

To better understand productivity, it would be good to find a way to measure the lifespan of compost bins. The current method of recording data makes doing so rather difficult. If there were some way to create a virtual “bin” object that could be updated as it is worked upon, then once we have a collection of completed objects, we could observe how their properties, like life-span, relate to variables such as number of volunteers and season. Understanding this relationship could help us understand productivity. The goal could be to get some sort of graph like the one below, that would help us make better predictions. The graph shows that an increase of volunteers is associated with a decrease in the time required to achieve black gold.



There may be a way to achieve this through more sophisticated R programming. Alternatively, Qualtrics may have a way to implement this. I will look into this as I redesign the survey soon.

Recommendations:

- Change the name of the “notes” column to “activity”, and only allow a few discrete values as responses, such as turned, added, combined, etc. (miscellaneous notes can go at the end)
- Include number of composting volunteers in survey
- Measure demand. We already measure supply through black golds. We could measure demand through how much compost is asked for on a given workday. We should encourage Claire to readily ask for compost, even if she thinks we don’t have it, so we can understand how we can improve.
- Identify technique to measure lifespan of composting bins, either through Qualtrics or R Programming