Trent Hogarth

Prof. Wirfs-Brock

CS-215-A

22 October 2023

<div align="center">Project 5: GitHub, Visualization Recreation, Requesting Personal Data</div>
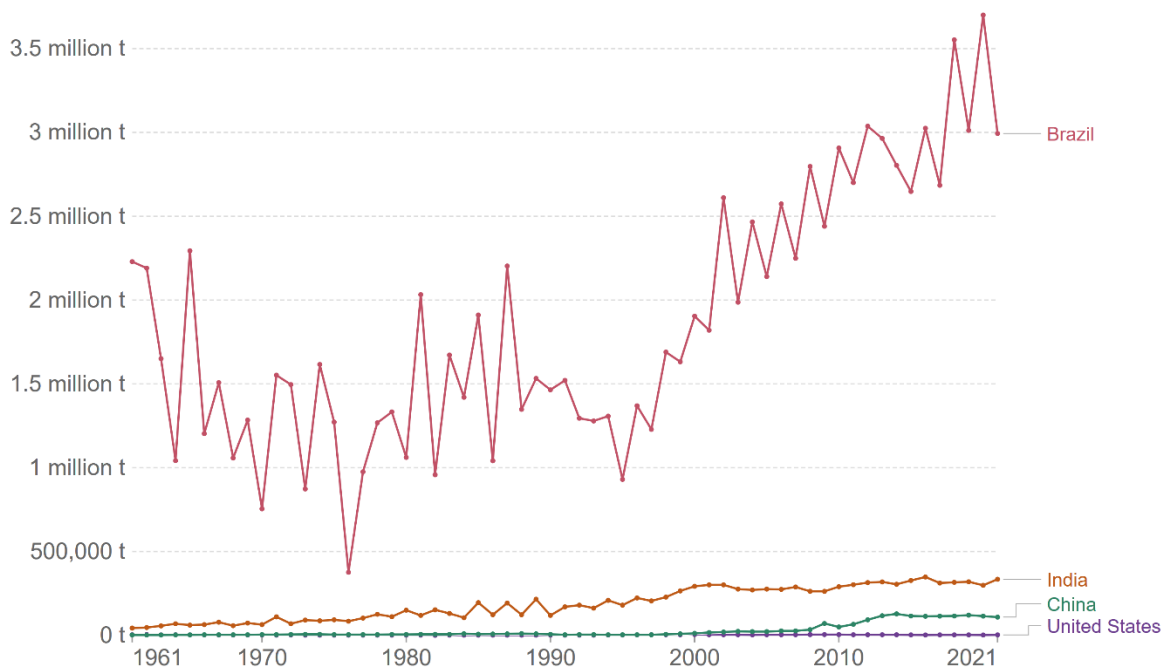
Github: TrentHogarth https://github.com/TrentHogarth

Recreating Data Visualizations:

Original Chart

Coffee bean production, 1961 to 2021. Coffee bean production is measured in tonnes.

https://ourworldindata.org/grapher/coffee-bean-production?tab=chart&country=BRA~CHN~USA~IND

 I drink coffee pretty regularly, so seeing the three most populous nations producing so little compared to Brazil was a bit striking. I also found it strange that Brazil produced so much more of the world's coffee than Colombia, which is more well known to me for its coffee than Brazil. The visualization shows the production of coffee beans for Brazil, India, China, and the US in tons per year over the course of 60 years as a color-coded line graph with each line labeled on the right-hand side, title and subtitle aligned on the left, and the y-axis tick labels in 'million t' at every half-million mark.

Matplotlib

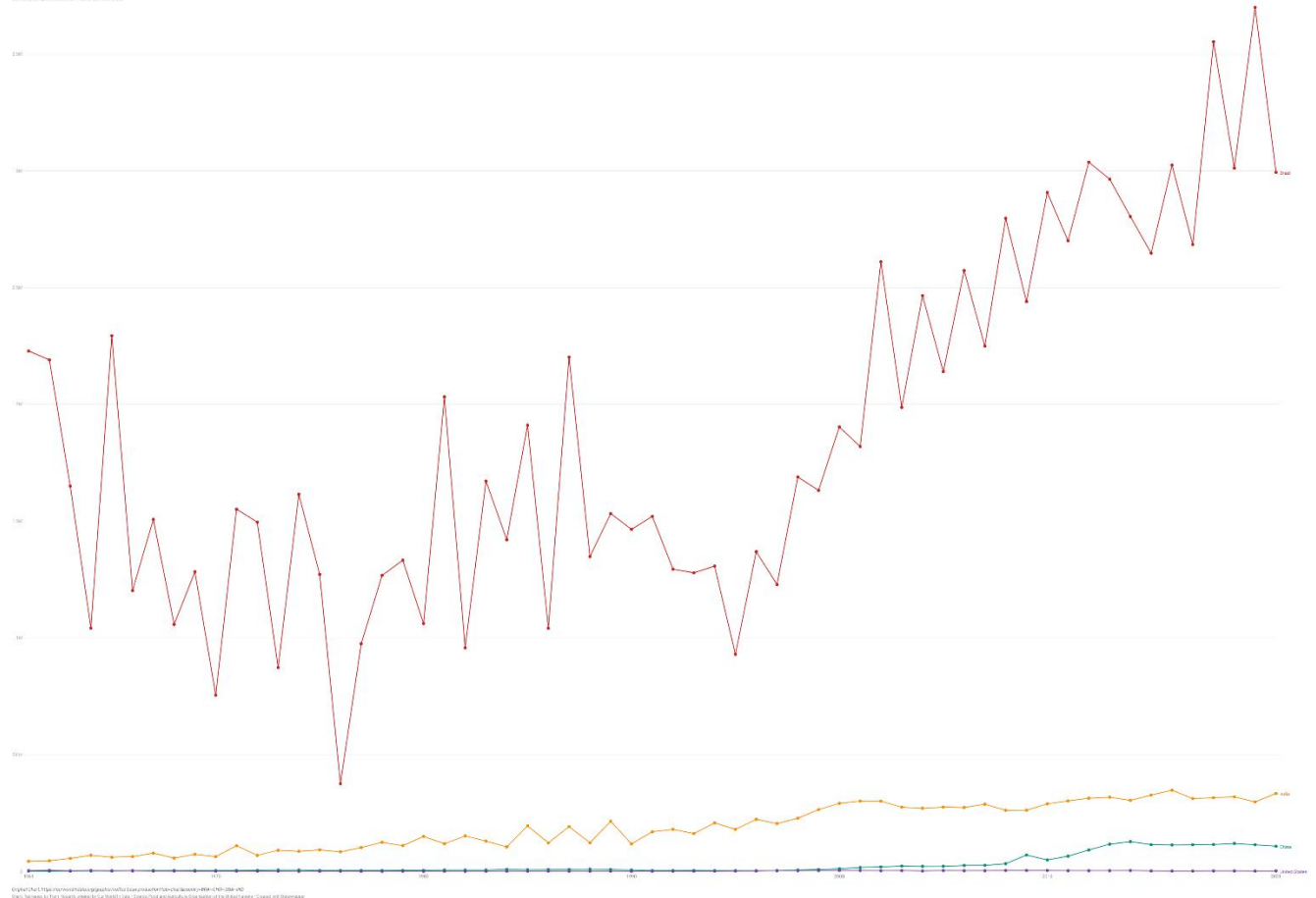Coffee bean production, 1961 to 2021, (Recreation in Matplotlib/pyplot w/ Pandas)
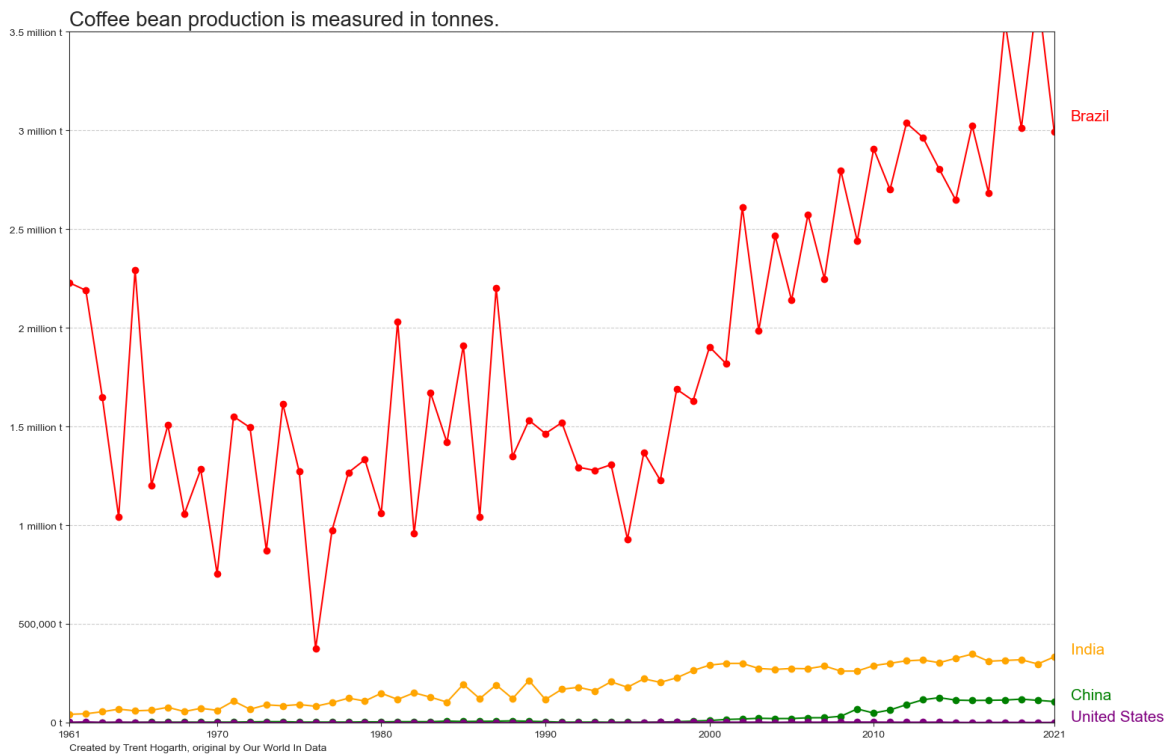
Excel
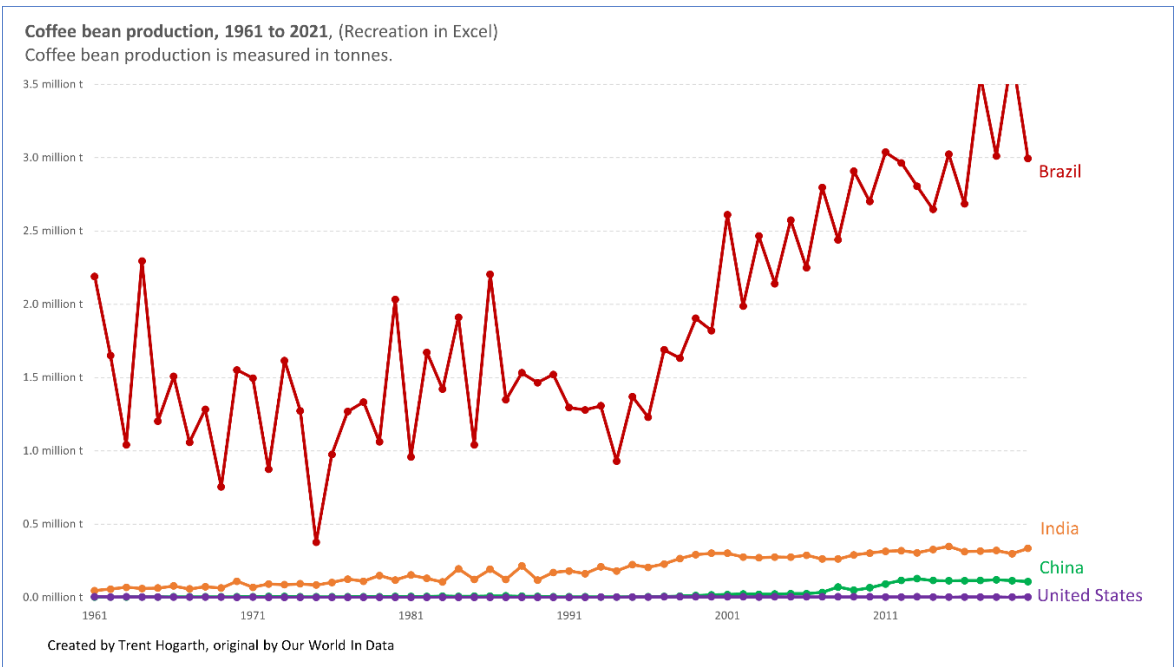
Datawrapper

Coffee bean production, 1961 to 2021, (Recreation in Datawrapper)
Coffee bean production is measured in tonnes.

1. The easiest of these tools by far was Datawrapper, although it was also the only one that required me to modify the data table outside of the tool in order to use the dataset properly. The tool offers only a few options for modifying the graph, and even less for reworking the dataset, but the pairing of features does at least make it the most user-friendly.
2. Matplotlib of course seemed to be the most customizable as far as the number and type of elements that can be added to a graph, but it's also one of the most time consuming to use with having to reorganize the dataset with Pandas before having to configure each part of the plot that you may want to include.
3. Excel also seemed to have a similar amount of utility, especially with modifying the data tables for the graph, which I believe makes it the best out of the three for analyzing the data quickly and in detail. It's also a fair bit faster to add labeling and other elements without having to place them with exact coordinates. It is a bit more difficult to deal with Excel's automatic graph elements, things like configuring what tick labels on an axis are displayed take some work to get right and can still not have the maximum value shown automatically.
4. While Matplotlib might be able to produce the best graphs for communicating data, the time it takes to configure and place labeling and annotations make me believe the few issues Excel has are outweighed by the speed that it offers for making multiple graphs and editing/presenting the dataset in a table.

Requesting Personal Data: Requested personal data from,

Spotify – Data requested, confirmation email claims it will be ready in 30 days at most.

Instagram – Data received in json format.

Garmin – Data requested, website says it will be ready in 48 hours, 30 days at most.