Welcome to Knowledge Check 2!

I hope the videos are going well. Today, we're going to try loading some data and generating a plot using pandas and seaborn. Here are your instructions:

1. Create a new file in your "data_1_checks" repository titled "kc_2". Remember, you can use .py files or .ipynb. The only problem with colab (what this was written in) is that it's a little trickier to commit to GitHub with it (at least as of this writing in April 2022). We really want you to be able to work with git and GitHub seamlessly, so you have to go the slightly longer route now of doing everything locally rather than on this server.
2. Make a folder titled "assets" in your repo. This is where you will put any CSVs or other files you download you want to load in. This is a nice way to organize data - if you drop everything in your main folder (in your case, "data_1_checks") you'll have a bunch of random data floating around intermingled with the files.
3. Load your data into a DataFrame.
4. Plot ANY of your data with seaborn (or matplotlib, or any other plotting tool if you have one you really like).
5. Push your changes to GitHub, and make sure to turn in the GitHub link into Google Classroom.

This will test your ability to find data and load it in. It sounds simple, but working with file paths to get your data into Python can be a little tricky at first. As a hint, Python will always search the current directory you're in. In Jupyter Notebook, you can access the command line with the exclamation point like this:

```
!dir
```

Once you're in your project directory, if you have them named like I do above, your data will be located at:

```
assets/data.csv
```

Note the exact path will depend on your operating system. This is assuming a file named "data.csv", yours will probably be named differently.

CoLab example

Note that CoLab has sample data included that I'm using here. You can also access data that's in your Google Drive with CoLab.

```
import pandas as pd
import seaborn as sns


df = pd.read_csv('sample_data/california_housing_test.csv')
df.head()
```

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	populatio
0	-122.05	37.37	27.0	3885.0	661.0	1537.
1	-118.30	34.26	43.0	1510.0	310.0	809.
2	-117.81	33.78	27.0	3589.0	507.0	1484.
3	-118.36	33.82	28.0	67.0	15.0	49.
4	-119.67	36.33	19.0	1241.0	244.0	850.

This is all we need! Just a basic plot showing two values a

```
sns.scatterplot(x=df['population'],
                y=df['median_house_value'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd874fa6190>
```