Parkinsons Telemonitoring Dataset

This data set was pulled from UCI[1]. The dataset contains biomedical voice measurements from 42 people with early-stage Parkinson's. There are about 200 samples per patient. This dataset contains the following information for each instance:

1. subject# - Integer that uniquely identifies each subject
2. age - Subject age
3. sex - Subject gender '0' - male, '1' - female
4. test_time - Time since recruitment into the trial. The integer part is the number of days since recruitment.
5. motor_UPDRS - Clinician's motor UPDRS score, linearly interpolated
6. total_UPDRS - Clinician's total UPDRS score, linearly interpolated
7. Jitter(%),Jitter(Abs),Jitter:RAP,Jitter:PPQ5,Jitter:DDP - Several measures of variation in fundamental frequency
8. Shimmer, Shimmer(dB), Shimmer:APQ3, Shimmer:APQ5, Shimmer:APQ11, Shimmer:DDA - Each measures of variation in amplitude
9. NHR, HNR - Two measures of ratio of noise to tonal components in the voice
10. RPDE - A nonlinear dynamical complexity measure
11. DFA - Signal fractal scaling exponent
12. PPE - A nonlinear measure of fundamental frequency variation

This data is good for regression and potentially clustering (by patient).

You can use the following code to load the data into google colaboratory:

```
import pandas as pd
url = "https://raw.githubusercontent.com/the-codingschool/TRAIN-datasets/main/parkinsons/parkinsons.csv"
df1 = pd.read_csv(url)
```

[1] A Tsanas, MA Little, PE McSharry, LO Ramig (2009)
'Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests',
IEEE Transactions on Biomedical Engineering (to appear).