

**1. Q: Can I clean my data outside of the notebook?**

A: Ideally all data cleaning will be done within your jupyter notebook. If you're using the survey data, then regex and string manipulation including the .split() method will help you.

**2. Q: What counts as an advanced plot?**

A: Advanced visualizations for the survey data can be anything that hasn't been taught in class. The interesting part of this dataset is that it has multiple dimensions. Think about how you can use color, size, space, and even time to communicate your story from that dataset. Be creative.

Strava Dataset FAQ

**1. Q. What are the units of the data?**

A. The data are in a variety of different units. A previous student noted the following units:

Cadence: rpm

Ground time: milliseconds

Vertical oscillation: centimeters

Distance, Altitude, and Enhanced Altitude: meters

Longitude and Latitude: semicircles (radians)

Air and Form Power: watts

Leg Spring Stiffness: kN/m

Speed: m/s

**2. Q. How can I explore the units more?**

A. Here's some code a student wrote last year using the FitFile package:

```
from fitparse import FitFile
import pandas as pd
import numpy as np

# Get all data messages that are of type record
stryd = FitFile('withstryd.fit').get_messages('record')
my_data = FitFile('my_data.fit').get_messages('record')

def get_data_plus_units(messages):
    data_record = []
    for record in messages:
        # Go through all the data entries in this record
        record_datum = {}
        for record_data in record:
            # Print the records name and value (and units if it has any)
            if record_data.units:
                record_datum[record_data.name] = str(record_data.value) + ' ' + record_data.units
            else:
                record_datum[record_data.name] = str(record_data.value) + ' unknown'
        data_record.append(record_datum)
    return data_record

df = pd.DataFrame(get_data_plus_units(stryd))
df2 = pd.DataFrame(get_data_plus_units(my_data))
df.head(10).transpose()
```

**3. Q. Where can I find the precise meaning of these columns?**

A. The precise meaning of these columns is not defined. This is an "authentically messy and unclear" dataset for you to explore and try and understand, you are welcome to ask me or discuss with peers based on googling around and the details I put in the assignment. I'm not intentionally hiding information I know!

**4. Q. How can I tell the difference between cycling/running activities?**

A. I think some broad understanding of the difference between cycling and running (e.g. speed which can be attained, distances) is probably the best way to start exploring this.

5. **Q. There are fields like cadence/Cadence or altitude/enhanced\_altitude. What's up?**

A. These are \*likely\* from different devices, but that's a best guess. Some of the devices used were [stryd](#), garmin forerunner 245 music, garmin [cadence sensor](#), garmin speed sensor. The last two were used on bike, the first one was used running, and the forerunner was used during both (to receive data).

6. **Q. Where are you able to find the 10 principles to design good figures?**

A. You can watch the design principles at

<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833>

7. **Q. May we extract the information from the .fit files in bonus instead of using the strava.csv file?**

A. Yes, keep in mind that the fit files in the bonus assignment are a superset of the data in strava.csv (they cover a longer time period).

8. **Q. Are we permitted to use any graphing library for Assignment 4?**

A. Yes.

9. **Q. What is an individual exercise versus a portion of an exercise?**

A. This is up to your interpretation! There are at least some activities where I ran then stopped for a while then ran again. Some days have multiple runs in them, for instance. You tell me what is reasonable as far as splitting up this data as you explore it.

10. **Q. Can you explain #4 on the grading. What do you mean by "interesting and defensible analysis that helps you understand what this data means in the context of your activities?"**

A. This is the most subjective portion of the assignment. I want to see you give me a summary of what you discovered that demonstrates your knowledge of the issues discussed in the course. It's vague as there are many ways to provide a reasonable explanation -- I want to see writing at a graduate level which is reasonable for the task/discovery/methods. I know that isn't very clear but my goal isn't to trick or deduct points without reason, I just want to see what insights you have actually hunted down (and your methods).