

Dr. K is very interested in not polluting. He hates creating carbon dioxide. Every trip to work, and every trip home, he watches to see how much fuel he uses.

Unfortunately, he never remembers to write that down.

He suspects that if he studies the GPS data from his travels that he can guess what is the best trip from home to work, or the reverse.

You are provided with raw GPS data. In order to collect this, Dr. K had to create a machine which would record it. He went through great trouble so that you can put on your resume that you worked with raw GPS data. Google has not touched it.

(continued)

Attributes to collect or compute for each GPS file:

Each attribute is worth 5 percentage points unless otherwise listed.

1. What day did the trip occur. Report this as YYYY/MM/DD, For example, 2024/09/23 for Sept 23rd. (5)
2. What time of day did the trip start. Use UTC for this. That's fine. Report this as HH:MM, with HH as a 24-hour clock. (5)
3. Did the trip start near the Professor's house? (5)
4. Did the trip go to RIT? (Yes, or No). (5)
5. Did the trip originate at RIT? (Yes or No). (5)
6. Did the trip go to the Professor's house? (5)
7. Is the trip a full trip? This is only true if the GPS device had a location lock before the trip started, AND after the trip was done. The GPS should be on all the way through the trip, with no sudden jumps from one location to another. You decide how far a sudden jump can be. The device tries to ignore points that occur in a straight line. Report this as Yes, or No, with a comma after the answer. (5)
8. How long did the trip take, not counting the time while the device started or after the device arrived at its destination. (5)
9. How many times did the car come to a stop on this trip? A stop we define as being stationary (or less than 5 mph) for longer than 30 seconds.

You will want to re-define a "stop" to get a better measure. For example, if the car came to a stop for longer than 7 minutes, chances are the car was stopped while the Professor ran an errand. This was probably not a stall in the traffic flow. (5)
10. On a scale of 0 to 100%, what fraction of the time did the car spend going uphill? Uphill is defined as going up by more than a 15 degree angle. You have to do some math to figure this out. Report this with one digit past the decimal point. (5)
11. In terms of minutes and seconds, how long did the car spend climbing hills? Report this as HH:MM::SS. (5)
12. From the bottom of each hill, to the top of each hill, how many meters did the car climb during the trip. If the car is climbing a hill, and then goes flat for a while, and then climbs again, that counts as one hill. You may have to define and describe when a hill is continued, or not. (5)
13. Compute a distance metric, or measure, of badness? What about how much the car accelerated? Can you think of something better to report that the professor might want to minimize when driving – like how much he accelerated when climbing hills? Report this for each trip? (5)

(continued)

Across all GPS files, please estimate and describe:

14. What is the best time of day for the professor to leave his house to go to RIT? Describe what you mean by “best” here and defend it.

For the “best” trip that you found, convert the GPS into a KML stream, render that KML in Google Earth, and do a screen capture. Add this into your Write-up.

15. What is the best time of day for the professor to leave RIT to head for home?

For the “best” trip that you found, convert the GPS into a KML stream, render that KML in Google Earth, and do a screen capture. Add this into your Write-up.

Conclusion: (25)

Write a conclusion. Example questions you might address include:

What did you learn from this exercise?

Did you need to use the Haversine distance?

What kind of data cleaning did you need to do?

Did you need to extract any critical points from the data to log? Is there a “rush hour” or two that the professor needs to avoid?

Can you figure out when the train goes through near RIT, which stops the Professor’s path?

Where there any ethical issues involved in this project?

Can you figure out when the Professor needed to stop for gas?

What else can you derive from this exercise?

In what way was this a good exercise?

Do you think this will help you get a job?

[More Information](#), (Added Nov. 28th) in case you missed the classroom discussion...

Prudential Insurance, and many others, want to offer benefits to people who are safe drivers. Ten years ago, they asked me if I could process GPS files to help them decide when drivers were driving safely, versus aggressively. At the time, I could not. So, I set out to figure out how to collect raw GPS data.

To do that, I built a GPS collection system for my students. You are getting the benefit of it.

So, imagine that Prudential Insurance gives you a set of GPS files. They then ask the question, "What can you tell me about these GPS files?" It is now up to you to determine what you can figure out.

You, as the investigator, will need to define several things. The person or persons evaluating your work will be interested in knowing how you defined them, and what you had to define. What does it mean to "stop going uphill?" What does it mean to "start going uphill?" "How long does the car need to go uphill before you bother paying attention to it?" "If the car is going on a flat road, and goes slightly up and down, will you pay attention to that?" "How will you do noise removal?" "Are there inflection points that matter?" You (and your optional partner) will define all of these things, and write about them in your write-up. You are the investigators, investigate.

In the past, students were asked to figure out how many left and right turns the car took on the trip. That means that the students working on that issue had to define what a left turn, or a right turn, was. And, they had to write about it in their project reports.

It could be worse. You might be asked to figure out which trips it was raining on. No kidding, you can do that. Rain decreases the friction on the road, and the car does not track as well on the road, also I am likely to drive slower in the rain. There are other ways, but don't worry about that this semester.

FEEDBACK FROM PART A:

I have downloaded and looked at all the work done by students in Part A. Here is some thoughts:

- You want to write a program that will take a GPS file as an input argument. Or, perhaps a list of filenames as an input argument. You do not want to hard code the filenames in your program. (Learn to parse command line arguments.)
- There are many packages which handle GPS. Matlab has some. Python has some. Use the packages, or write your own. In your conclusion, discuss what you did.
- GPS files are the input. KML files are the output. There are also many packages for creating KML files: simpleKML, fastKML, pyKML... again, find the packages and figure out how to use them, or you do not have to.

You are creating something that is relatively static. It is okay to create a static

header that starts a KML file, then has the body which contains your data and special markers (if needed), and the trailer for your KML file.

- You can use the Haversine distance, but you don't need to use the Haversine distance because:
 - Each incremental change over the region is relatively small.
 - In our region, Monroe County, longitude and latitude are perpendicular enough.
- For many students, the code handed in for Part A has large and monolithic. You should factor your code into subroutines, and smaller functions.

As a general rule, an individual function should have about one page of comments, and about one page of code – interspersed or separate.

- You are given a problem to solve, you need to decompose that problem into sub-problems, and solve the sub-problems.
- The idea of this is that each GPS file is analyzed, and the resulting analysis put into a CSV file or some kind of database. Why would you want to use a database? Why do we want to use databases?