Segmentation of Trip Data

Dave Hurst

Thursday, December 25, 2014

Preliminary notes for dealing trip segmentation

Now we've got the 99th trip for driver 2591 in memory.

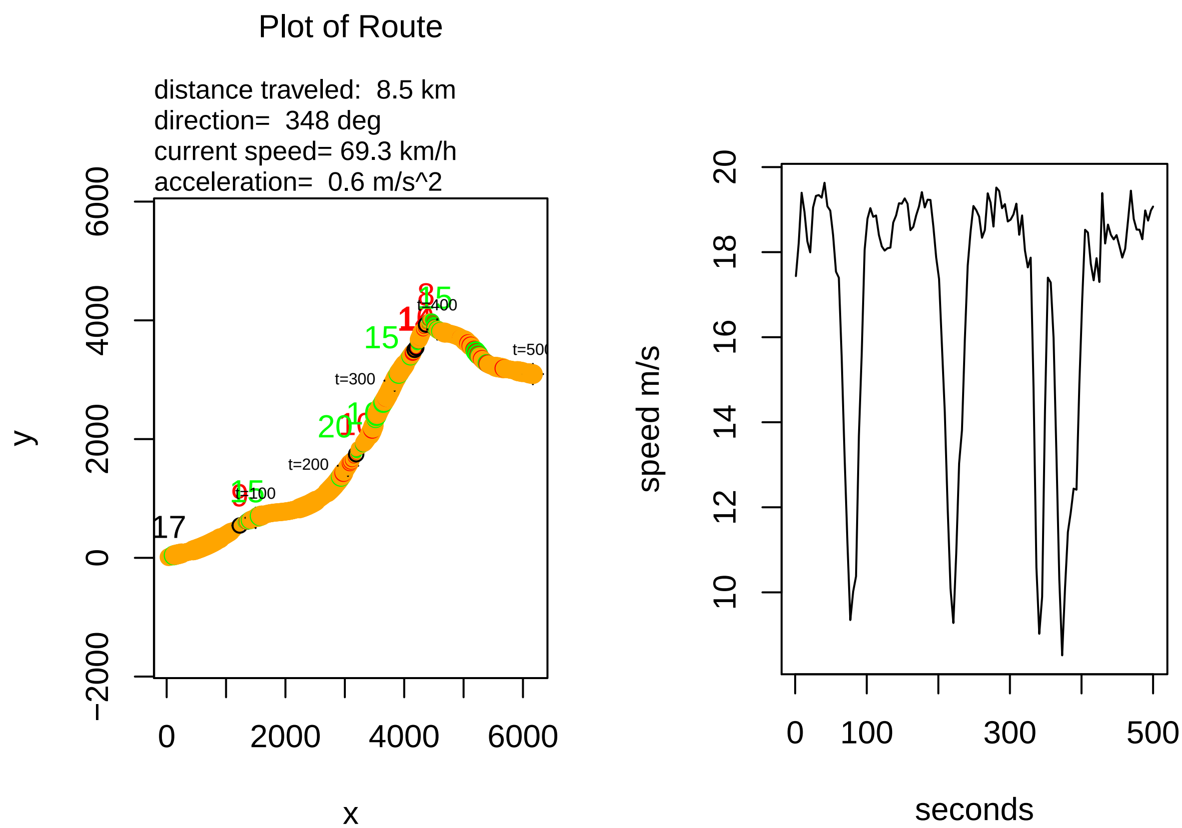
My first attempt will be to identify straight long segments high speed segments (SLH).

- Straight - implies minimal overall change in bearing (what is straight?)
- Long - implies that the constant speed segment is sufficiently (define) long compared to the acceleration/deceleration portion of the segment
- High speed - above some threshold

The goal will be to find the arbitrary thresholds defining these, and write code to automatically discover these sections and the driving characteristics that occur within them.

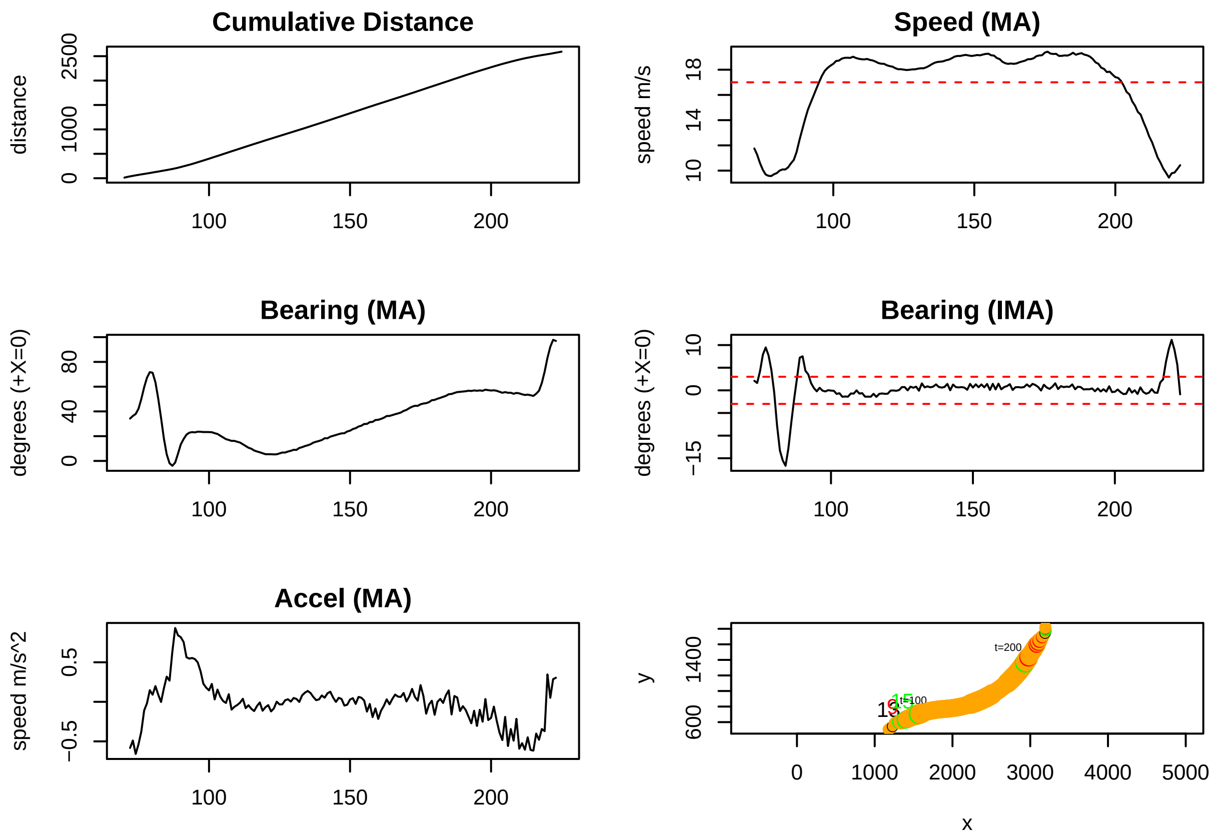
A good candidate to start with is the first portion of the trip we have been working with.

```r
plotTripSegment(trip, 1, 500)
```



I've created a function to plot the characteristics using moving averages.

```r
plotTripSegment6(trip, 70, 225)
```

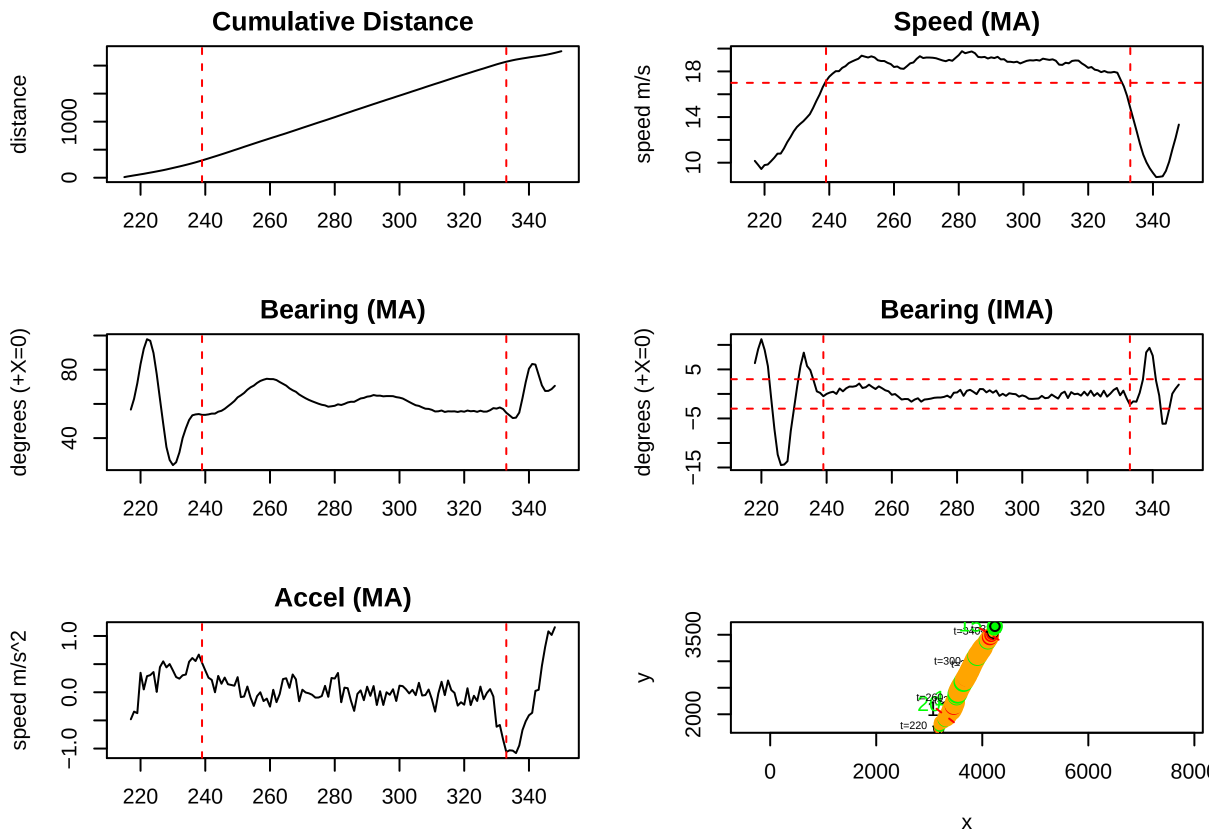


Although this segment is somewhat curved, its straight enough (I think) to not influence behavior. The linear portion of the cumulative distance plot is probably what we want to call the *straight*. The lag shifted bearing plot implies that we should be able to capture *straight* segments by looking for shifts where -3 < B' < 3. Similarly the driver holds relatively constant speed over the straight, and we should be able to capture that as well. Where we have a platuea in speed and a flat line on B', we should be able to identify an SLH. Let's consider the segment *long* if the constant speed section is >75% of the straight segment and involves a >50% increase in speed. And finally, we can define high speed as anything above 17 m/s, which is just shy of 40mph, and seems like a reasonable, if arbitrary cut-off.

On a side note, the acceleration with MA lag paints a much more helpful picture. It looks like the deceleration is about half the acceleration, which wasn't evident from other plots, and I think might be a signature characteristic of drivers.

Given those definitions, let's look at the next segment.

```r
plotTripSegment6(trip, 215, 350, t.mark=20, b.marks=c(239,333))
```

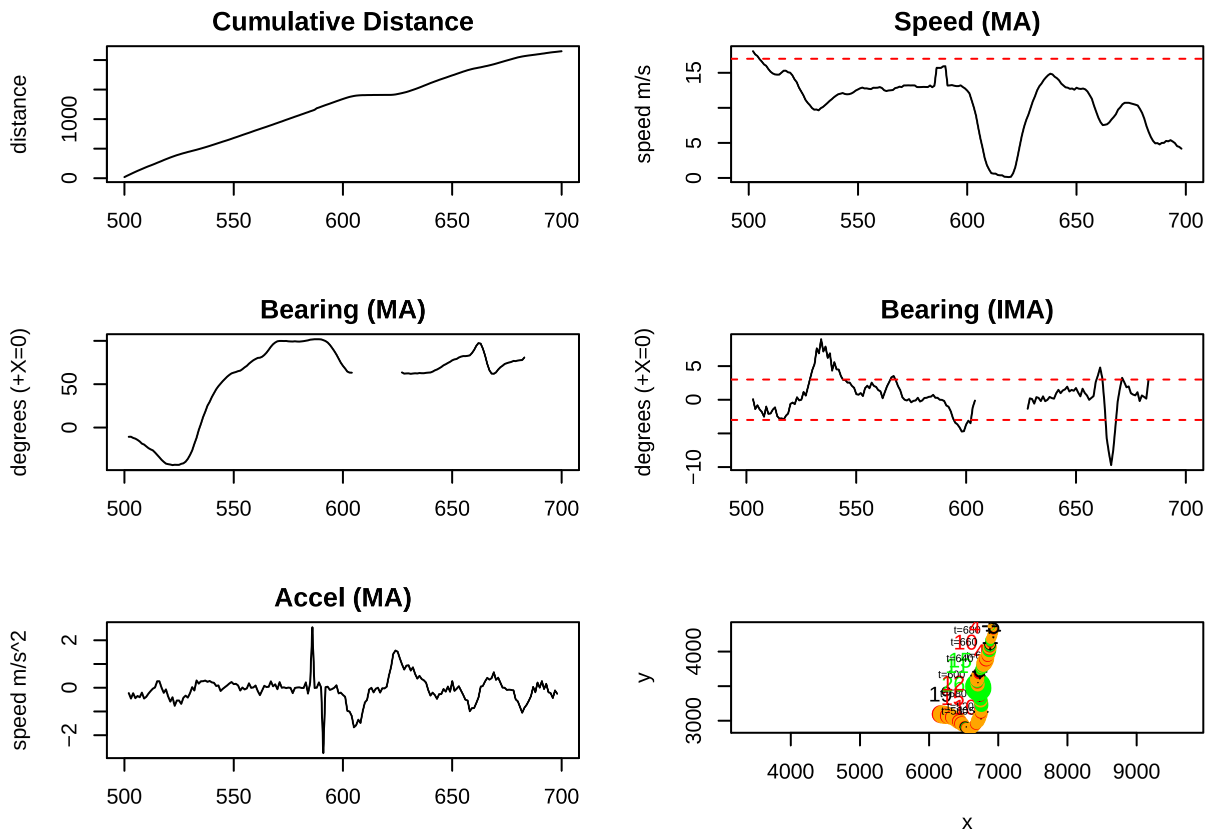


I think the criteria laid out above works well for this one as well. I manually selected the segment boundary (vertical dashed lines) using in the Bearing' plot by looking for the first change in direction that leads to a series within the -3<B'<3 boundary, until the first point that trends out of it. Based on the curvature in the Cum. Distance and the Accel plot, that's a little too agressive. On this plot, putting in a 5 second lag/lead (e.g. t=(244,328)) hits the window pretty well, so I'll build that into the algorithm.

Filtering with Moving Average

I've been worried about how to handle the filtering, but have largely ignored it for now. I expect it will bite us at some point. Since we have a known issue with this trip (between 580< t *issue* <600), I thought I'd try out the new plot6 function. In general it helps, but doesn't resolve the issue completely. In this case, this segment doesn't qualify as an LSH (or any kind of straight segment).

```
plotTripSegment6(trip, 500, 700, t.mark=20)
```


Cumulative Distance
distance
0
1000
Speed (MA)
speed m/s
0
5
15
Bearing (MA)
degrees (+X=0)
0
50
Bearing (IMA)
degrees (+X=0)
−10
0
5
Accel (MA)
speed m/s^2
−2
0
2
500
550
600
650
700
y
3000
4000
x
4000
5000
6000
7000
8000
9000