Filtering Noise in the Data

Dave Hurst

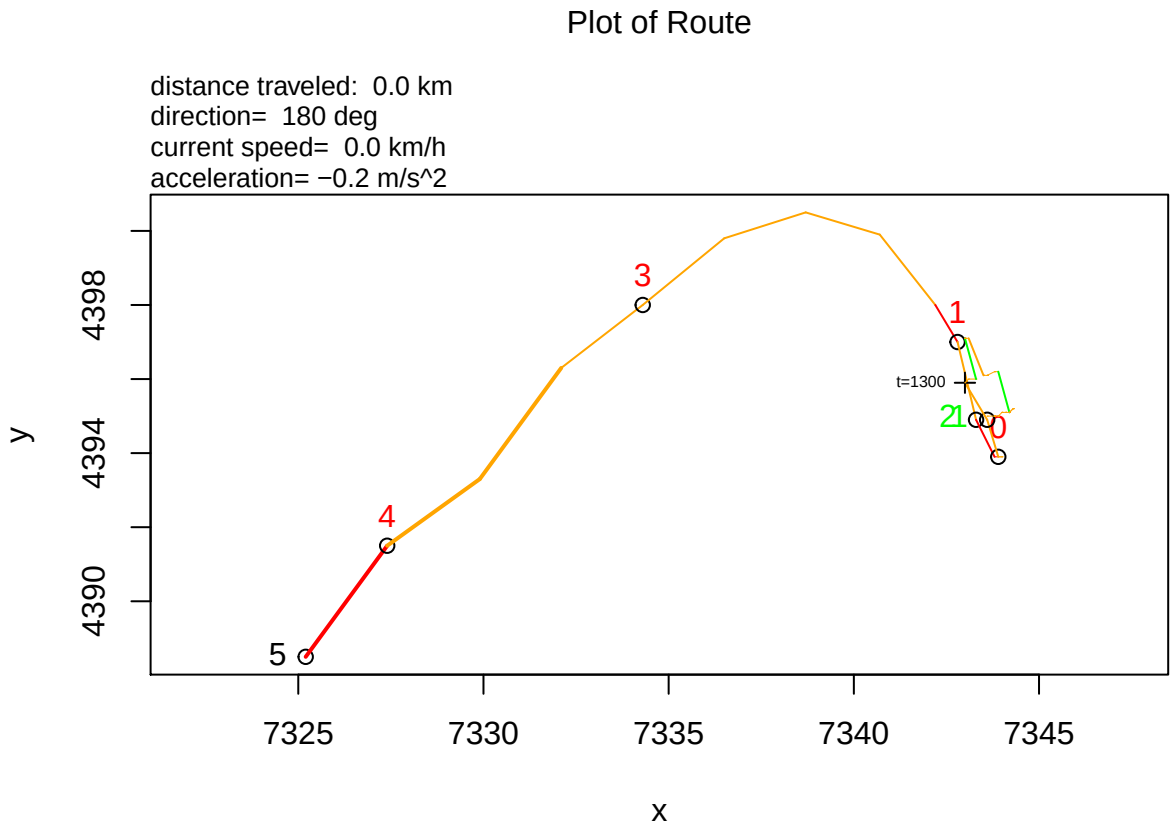
Thursday, December 25, 2014

Preliminary notes for dealing with noise in the data.

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

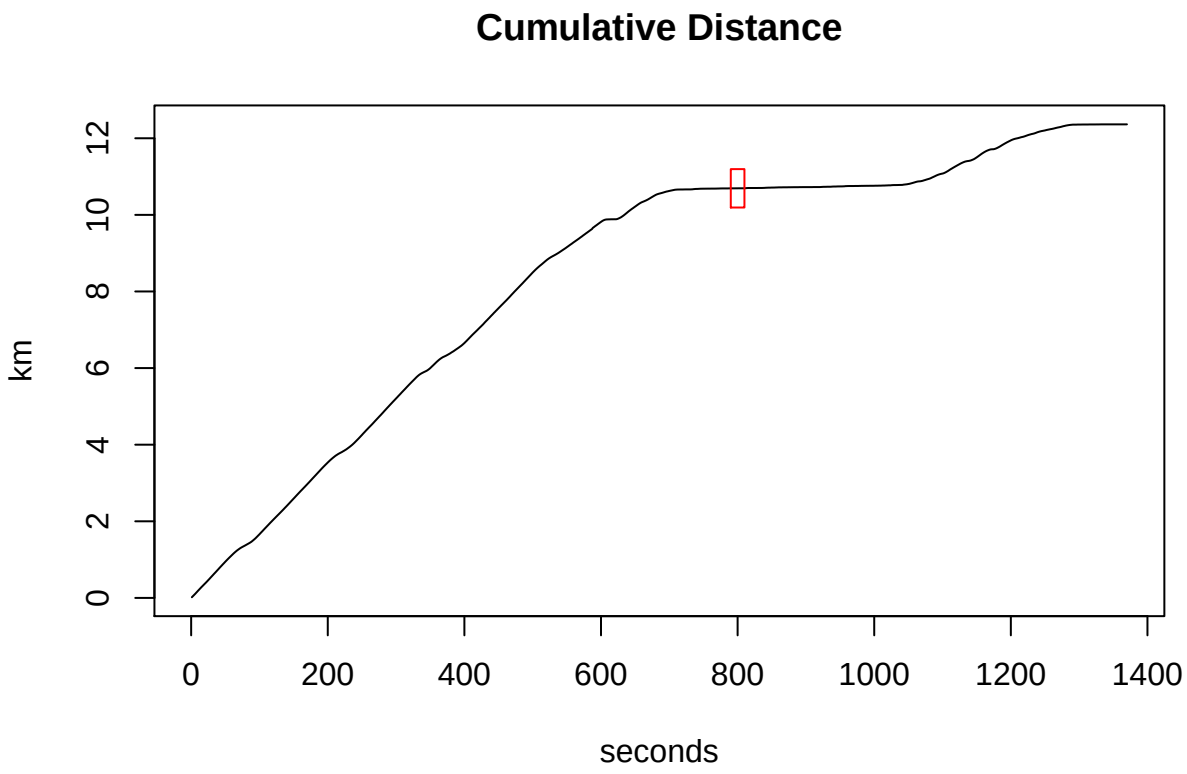
Here's an obvious depiction of the problem. The plot below shows the last 90 seconds of the journey. It's clear that the car is motionless, and yet we get wiggle in the coordinates that show up as 21 m/s velocity.

```
plotTrip(trip, tmin=nrow(trip)-90, v.mark=1)
```



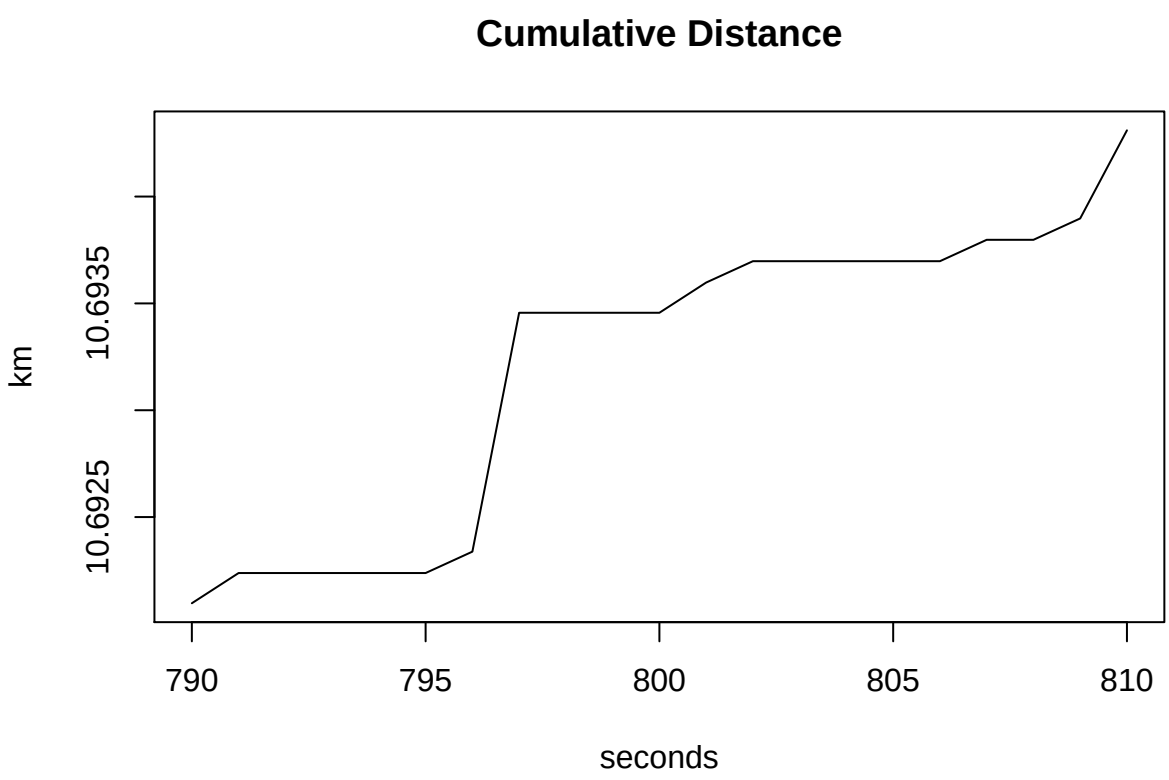
Plot of cumulative distance (note: trip$v can be used as segment distance, since t=1), and pick an arbitrary region to zoom in on (denoted with a red box)

```
plot(cumsum(trip$v)/1000, ylab="km", main= "Cumulative Distance", xlab="seconds", type="l")
t1 <- 790; t2 <- 810
d1 <- cumsum(trip$v)[t1] - 500
d2 <- cumsum(trip$v)[t2] + 500
box.x <- c(t1, t2, t2, t1, t1)
box.y <- c(d1, d1, d2, d2, d1)/1000
lines(box.x, box.y, col="red")
```



Zoom in (on the red box above) to see the noise. This plot is basically saying that car shifted 1 meter in its lane, which is plausible, but could be GPS error. The main take away is that high speed features are going to be more reliable than low speed since the GPS inaccuracy will be a smaller portion of the variance.

```
plot(t1:t2, cumsum(trip$v)[t1:t2]/1000, ylab="km", main= "Cumulative Distance", xlab="seconds",type="l")
```


Cumulative Distance
km
10.6935
10.6925
790
795
800
805
810
seconds