Selected Driver Observations

Dave Hurst

Sunday, December 21, 2014

Playing around with a few driver trip profiles as part of the EDA. *Caveat:* I haven't read any of the forums at this point.

Here's the data read statement. It's a little overkill, but I'm setting things up for multiple trips/drivers

```
data.dir <- "c:/NO_BACKUP/kaggle/telematics/drivers"

drivers.dir <- dir(data.dir)
set.seed(2349)

random.driver <- order(runif(1:length(drivers.dir), ))

N.DRIVERS <- 1
driver.df <- data.frame(id=integer()
                        , speed.avg=numeric(), speed.max=numeric()
                        , break.max=numeric(), accel.max=numeric())
```

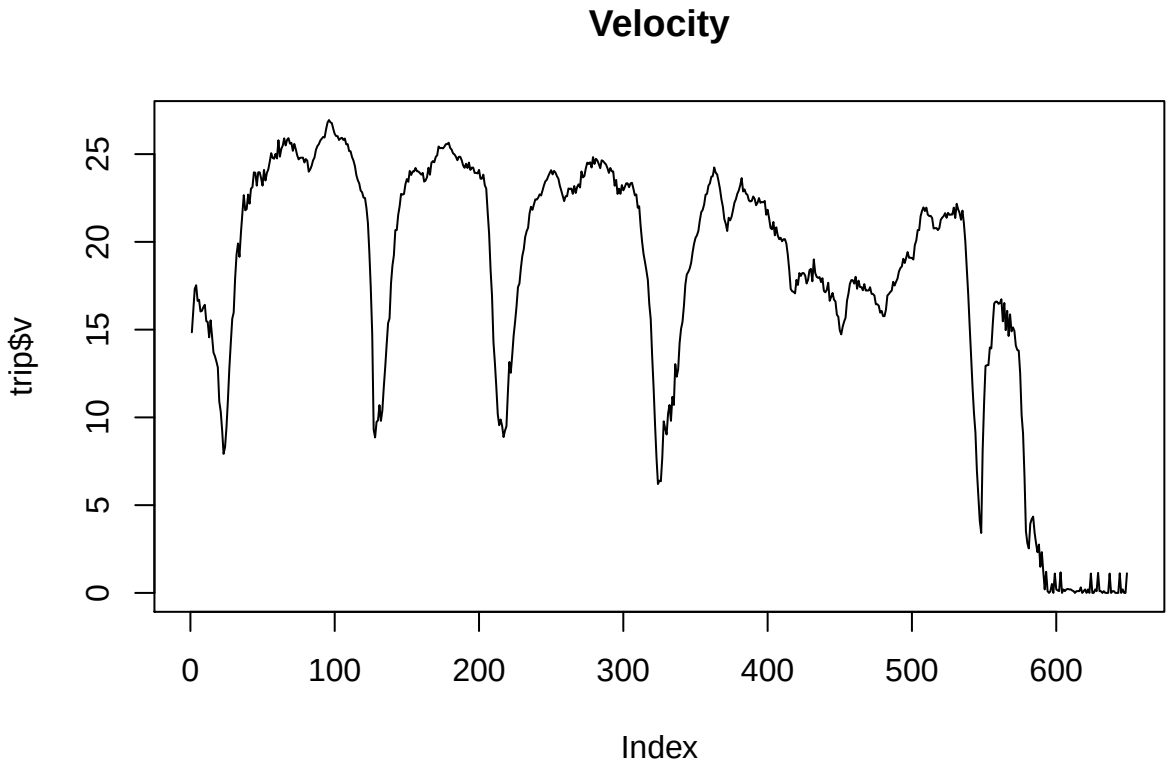
Here's the code for a single trip. I'm calculating *speed* and *acceleration* since these will be used for feature creation

```
i <- 1
j <- 1

    driver.dir <- paste(c(data.dir, drivers.dir[random.driver[1]]), collapse="/")
    driver.files <- dir(driver.dir)

        trip <- read.csv( paste(c(driver.dir, driver.files[j] ), collapse="/"))
        trip.last <- trip[-nrow(trip), ]
        trip <- trip[-1, ]
        trip$x.d <- trip$x - trip.last$x
        trip$y.d <- trip$y - trip.last$y
        trip$v <- sqrt( trip$x.d^2 + trip$y.d^2 )  # distance travelled per second
        trip.last <- trip[-nrow(trip), 3:5 ]
        trip <- trip[-1, ]
        trip$x.d2 <- trip$x.d - trip.last$x.d
        trip$y.d2 <- trip$y.d - trip.last$y.d
        trip$a <- trip$v - trip.last$v

    driver.df[i,"id"] = drivers.dir[random.driver[1]]
    driver.df[i, "speed.avg"] =  mean(trip$v)
    driver.df[i, "speed.max"] =  max(trip$v)
    driver.df[i, "break.max"] =  min(trip$a)
    driver.df[i, "accel.max"] =  max(trip$a)

driver.df[i,]
```

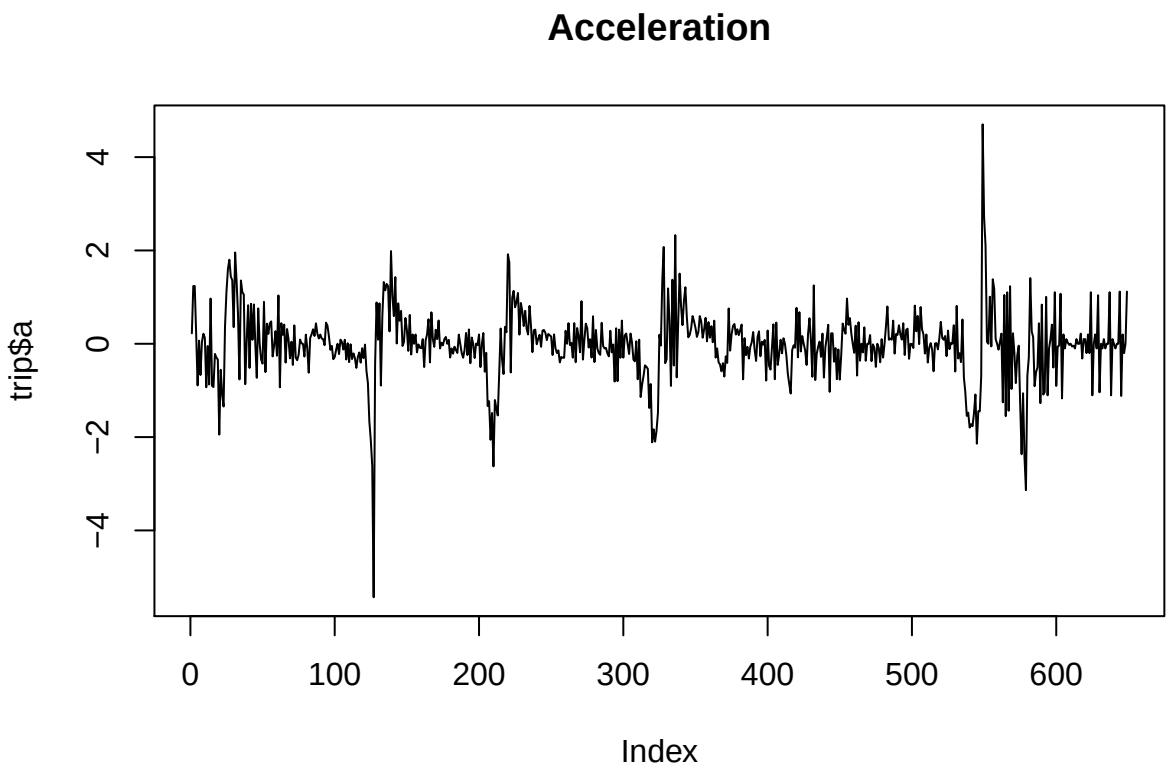
```
##     id speed.avg speed.max break.max accel.max
## 1 2591  17.75485   26.9429 -5.431824  4.702208
```

Here's the speed and acceleration profile for the first trip. *Note:* The R-markdown plots are different than what I'm getting out of R. I see a spike in the Markdown that isn't in R-Studio, and its making the acceleration plot useless. The Velocity is also much better if you plot directly from R – didn't feel like fighting this.

```
plot(trip$v, type="l", main="Velocity")
```




```
plot(trip$a, type="l", main="Acceleration")
```

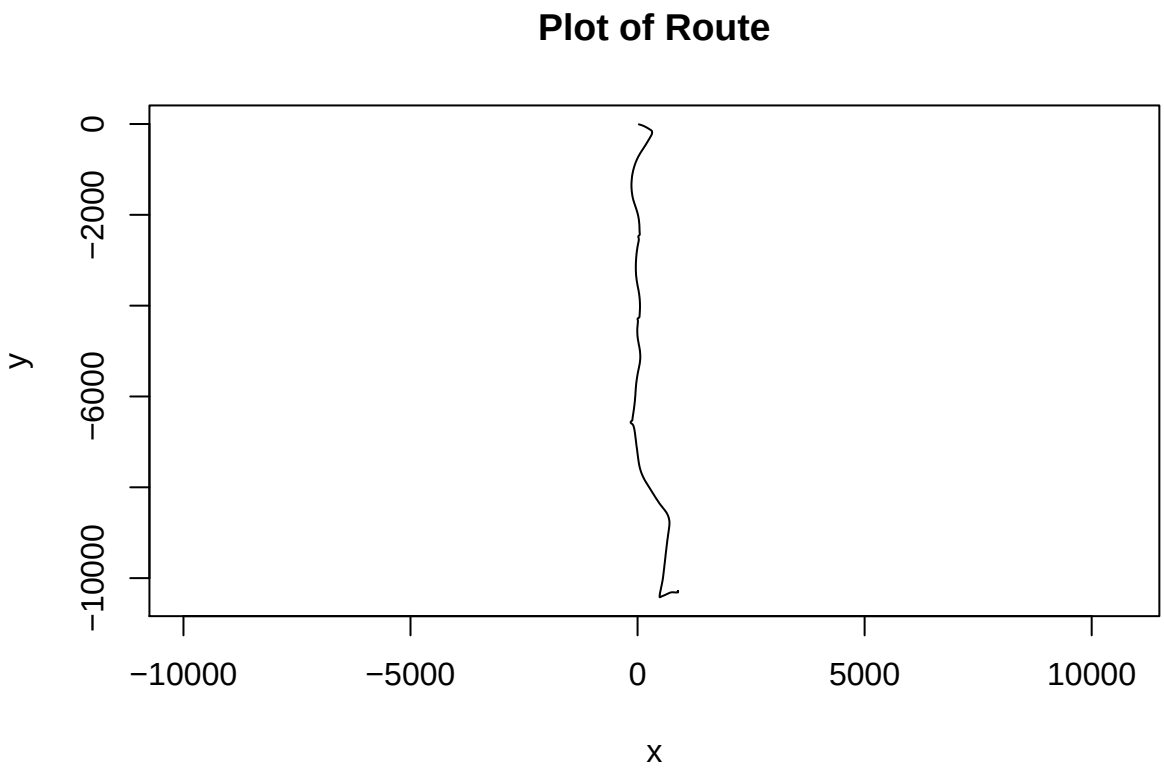


A couple of things to note in this plot: For the first half of the trip, the driver seems to be moving on surface streets at around 25 m/s (~55 mph) with several slow downs, but no stops (timed traffic lights? Country roads with no cops?). the acceleration (slope of the velocity curve) back to the speed limit is consistent, and this probably a characteristic of a particular driver.

Acceleration/Deceleration during "constant" stretches is probably also a characteristic we can capture

Here's the route ... not sure what this tells us.

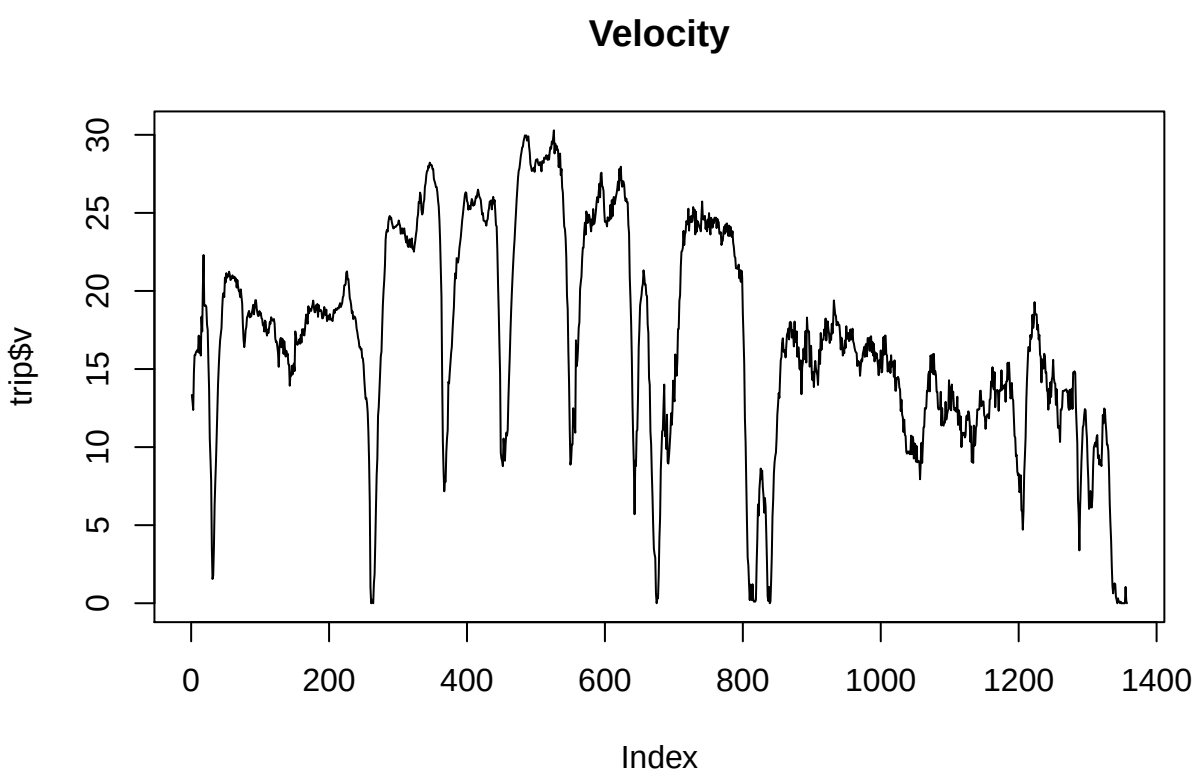
```
plot(trip[, 1:2], type="l", asp=1, main= "Plot of Route")
```

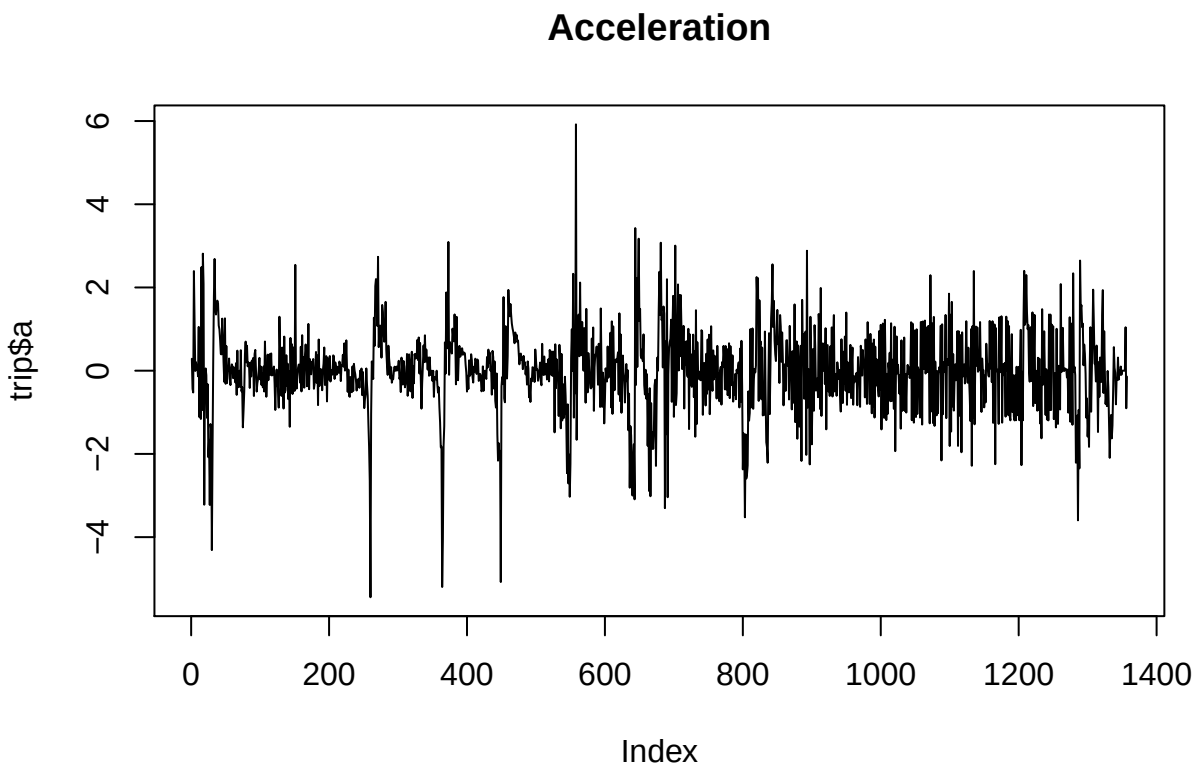


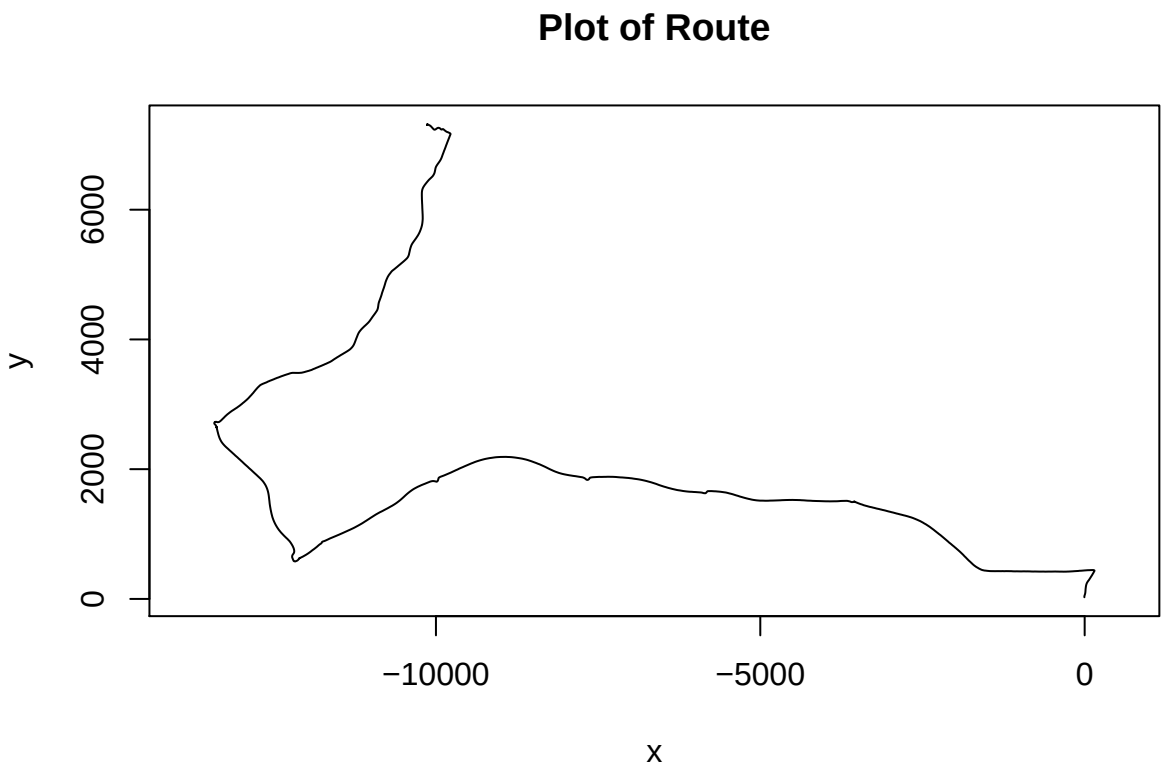
Same plots and calculation for a different trip used for feature creation

```
i <- 1
j <- 103
```

```
##     id speed.avg speed.max break.max accel.max
## 1 2591   16.9457  30.28696 -5.444421  5.920263
```


Velocity
trip$v
0
5
10
15
20
25
30
0
200
400
600
800
1000
1200
1400
Index

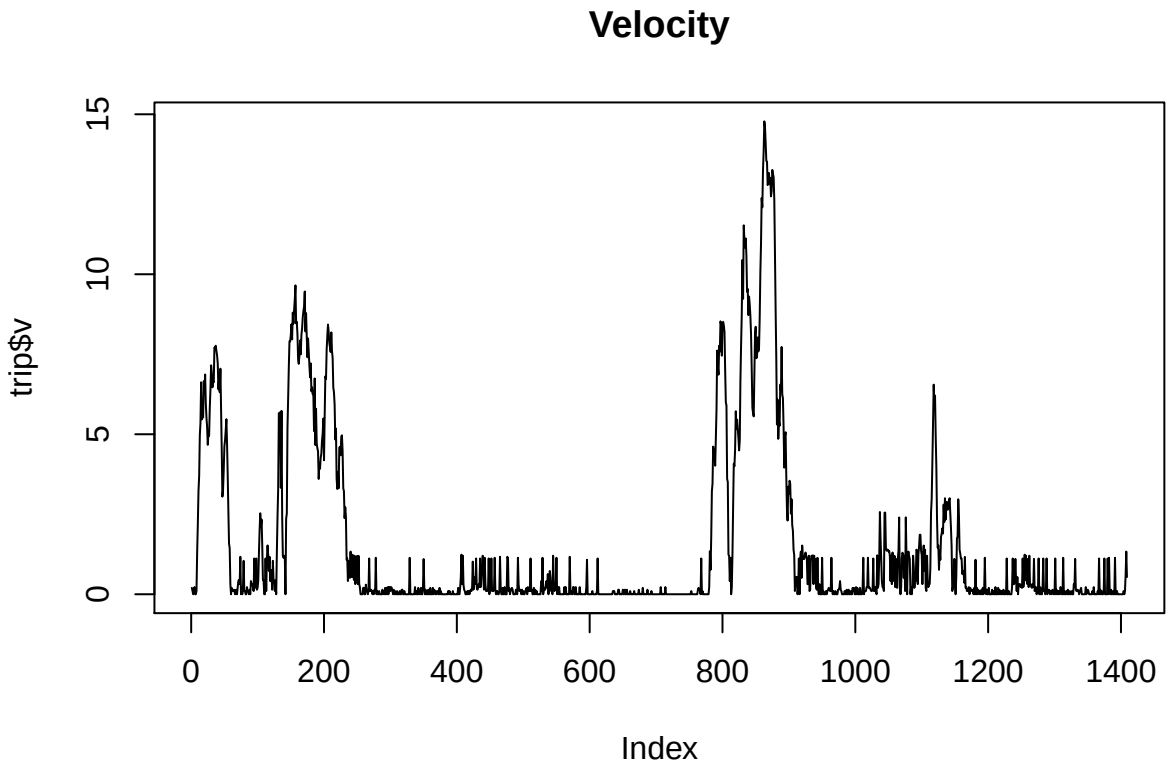

Acceleration
trip$a
6
4
2
0
−2
−4
0
200
400
600
800
1000
1200
1400
Index

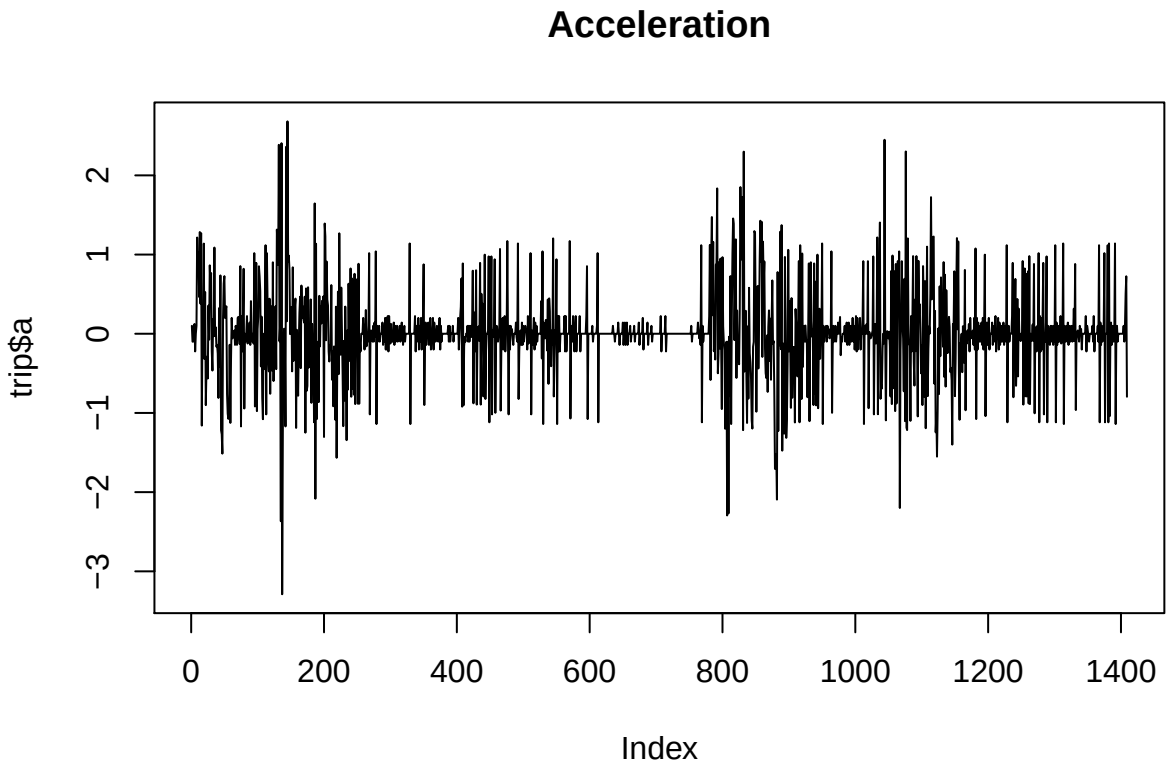


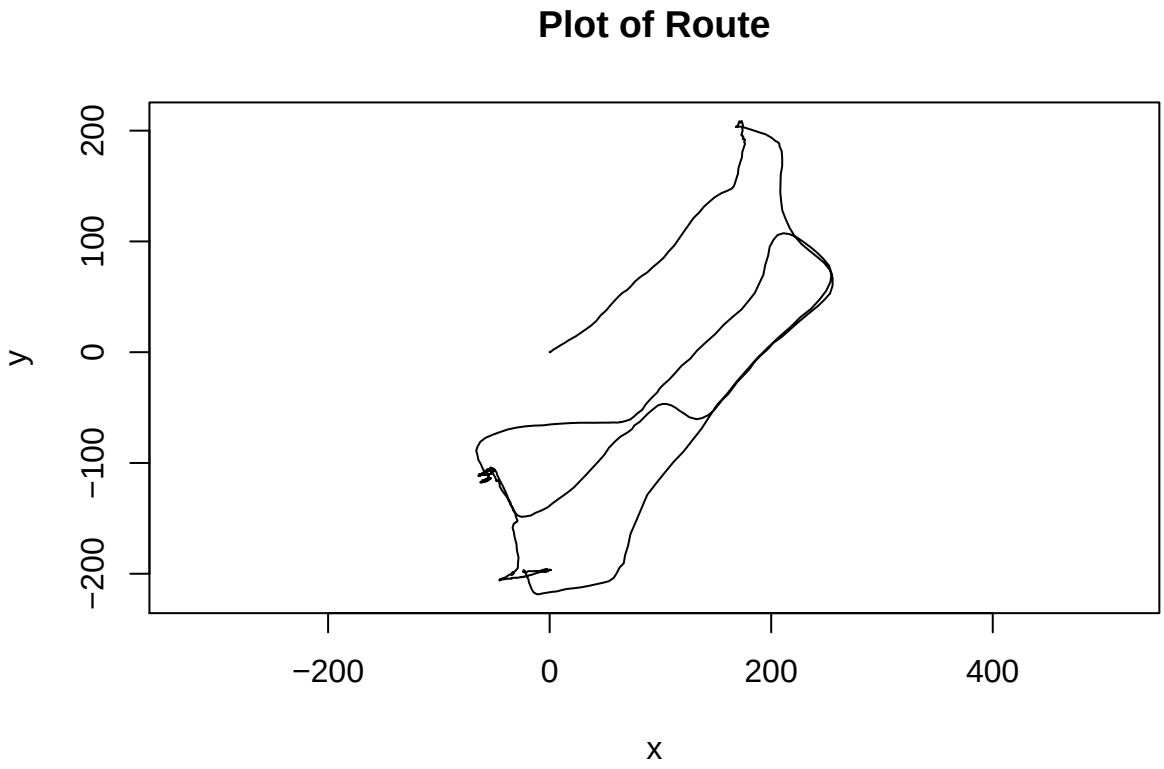
And again for a different driver

```
i <- 62
j <- 72
```

```
##      id speed.avg speed.max break.max accel.max
## 62 2591  1.538971   14.7787 -3.289308  2.682064
```


Velocity
trip$v
0
5
10
15
0
200
400
600
800
1000
1200
1400
Index


Acceleration
trip$a
−3
−2
−1
0
1
2
0
200
400
600
800
1000
1200
1400
Index


Plot of Route
y
200
100
0
−100
−200
−200
0
200
400
x