In-Class Programming Activity, Day 8

Shelby Witherby

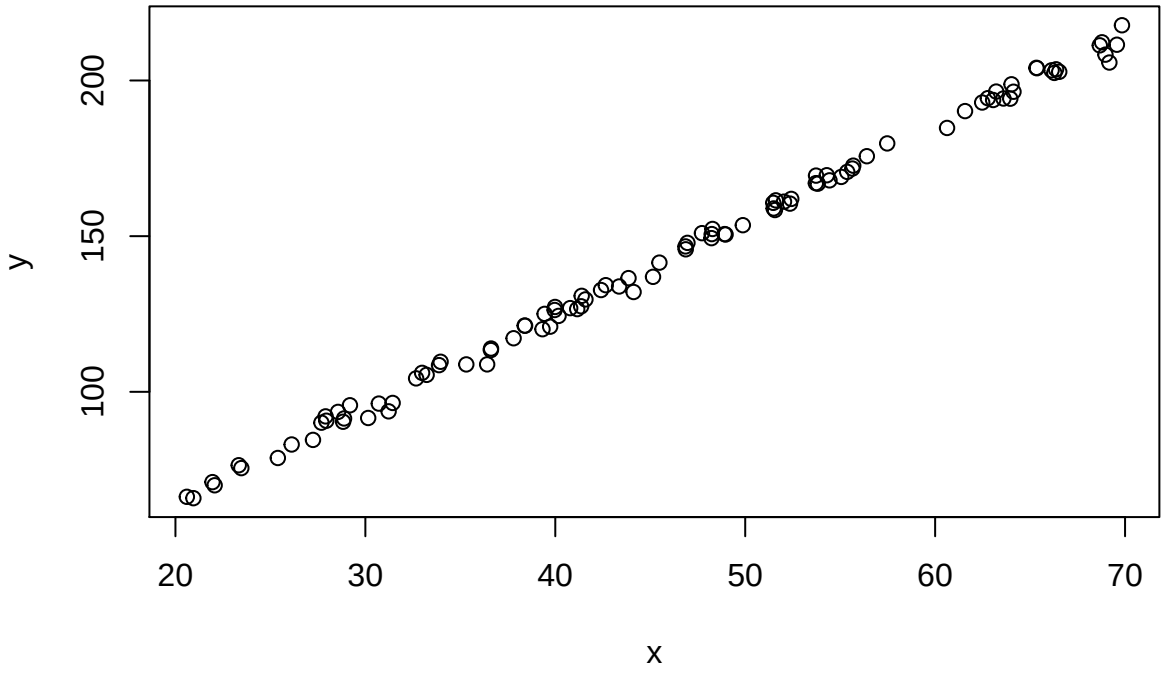
Task 1

1)

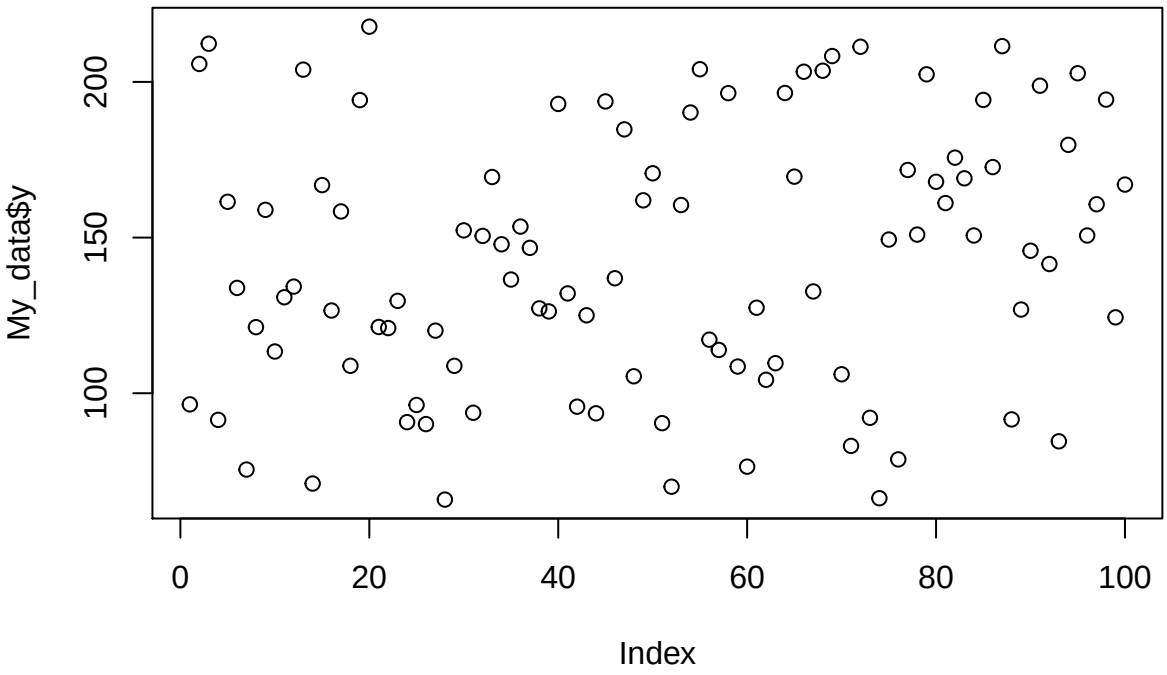
```
library(mosaicData)
library(ggplot2)
My_data <- data.frame(x = runif(100, min = 20, max = 70))
My_data$y <- 5 + 3*My_data$x + 2*rnorm(100)
```

2)

```
plot(My_data)
```




```
plot(My_data$y)
```



3)

```
LL_line <- function(params) {
  m <- params[1]
  b <- params[2]
  sigma <- params[3]
  with(My_data, sum(log(dnorm(y - (m * x + b), sd = sigma))))
}
```

4)

```
testA <- LL_line(c(m = 3, b = 5, sigma = 2))
testB <- LL_line(c(m = 4, b = 1, sigma = 10))
```

5)

```
starting_params <- c(4, 1, 10)
best <- optim(starting_params, LL_line, control = list(fnscale = -1))
```

Task 2 - Taxicab fare structure

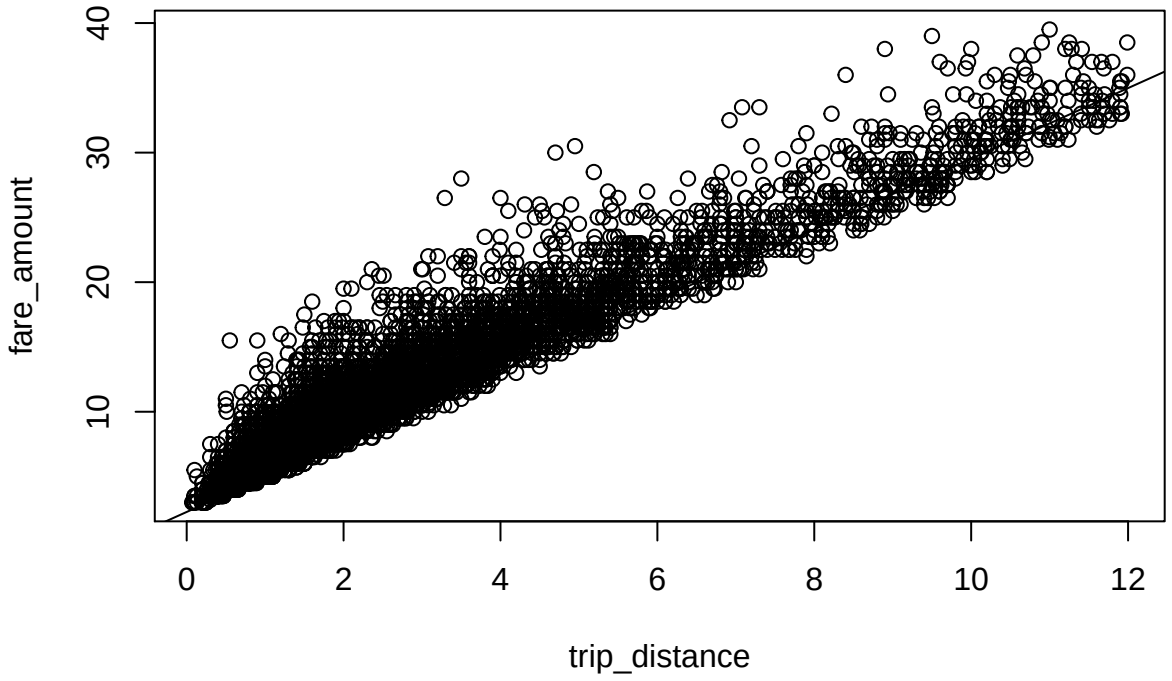
```
load(url("http://tiny.cc/dcf/Taxi_trips.rda"))
attach(Taxi_trips)
#View(Taxi_trips)


taxi_likelihood <- function(params, data = Taxi_trips) {
  base_fare <<- params[1]
  per_mile <<- params[2]
  rate <- 1/params[3]
  extra <- with(data, fare_amount - (base_fare + per_mile*trip_distance))
  sum(log(dexp(extra, rate) + 1e-2))
}
```

```r
best <- optim(c(2.3, 2, 5), taxi_likelihood, control = list(fnscale = -1), data = Taxi_trips)
best
```

```
## $par
## [1] 2.264374 2.726340 1.857986
##
## $value
## [1] -16502.3
##
## $counts
## function gradient
##      423       NA
##
## $convergence
## [1] 0
##
## $message
## NULL
```

```r
plot(fare_amount ~ trip_distance)
abline(base_fare, per_mile)
```



Using base fare as our y-intercept, and per mile as our slope, we see each trip has about an additional 10 dollars due to standing-still time. The component of each trip's fare that's due to standing-still time is thus much more signicant when the trip's distance is small, and is a smaller component of the total trip (proportionally) when it is a long trip.

Test statements

```
scoreActivity::score253(8)
```

```
## --------
## Running test statements for day  08
## Loading required package: scoreActivity
## Loading required package: lazyeval
## passed: object "My_data" exists
## passed: max(My_data$x) <= 70
## passed: object "LL_line" exists
## passed: LL_line(c(m = 3, b = 5, sigma = 2)) > LL_line(c(m = 2, b = 1,
##     sigma = 10))
## passed: object "Taxi_trips" exists
## passed: object "taxi_likelihood" exists
## passed: object "best" exists
## passed: "par" %in% names(best)
```