Extra_credit

Read data

```
df <- read_csv('[sub]Tampa buys 5,000 additional mosquito dunks.csv',show_col_types = FALSE)
```

Prepare data

```
# format to type date
df$as_date <- as.Date(df$Date, format = "%m/%d/%Y")

# fill na = 0
df$'Daily Confirmed Cases'[is.na(df$'Daily Confirmed Cases')] <- 0

# cal prev 1 day
df$Previous_Cases <- dplyr::lag(df$'Daily Confirmed Cases', 1)
df$Previous_Cases[1] <- 0

head(df)
```

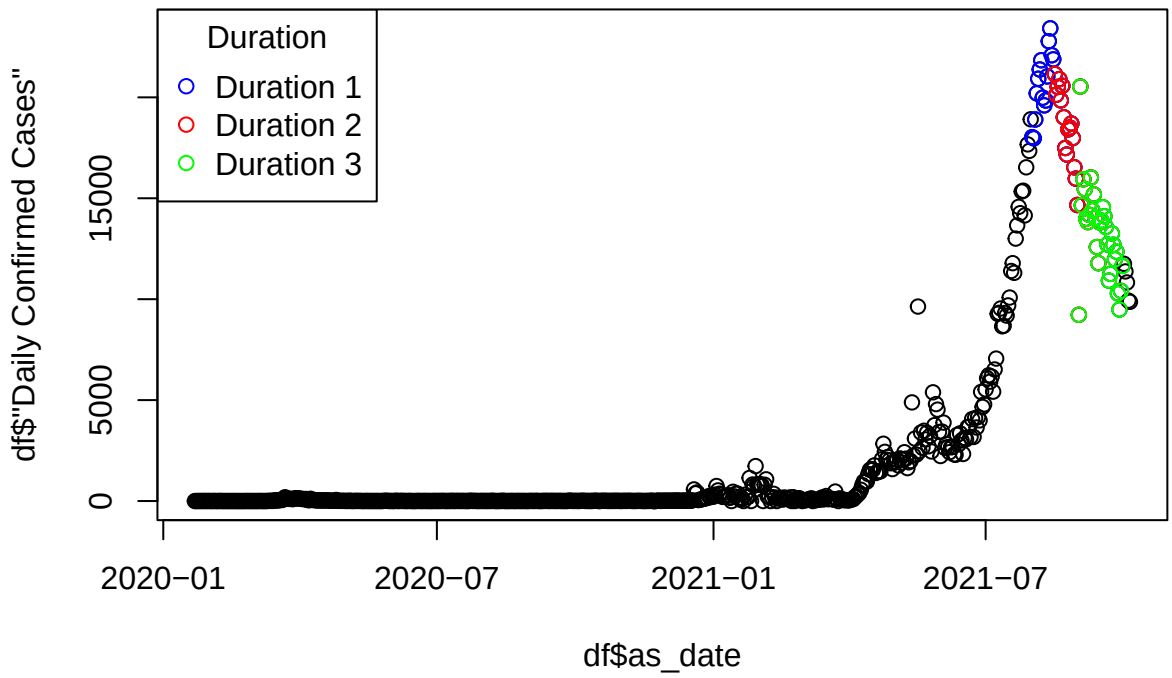
```
## # A tibble: 6 x 5
##   Date   Cumulative Confirmed~1 Daily Confirmed Case~2 as_date    Previous_Cases
##   <chr>                   <dbl>                   <dbl> <date>              <dbl>
## 1 1/22/~                      4                       0 2020-01-22              0
## 2 1/23/~                      4                       0 2020-01-23              0
## 3 1/24/~                      5                       1 2020-01-24              0
## 4 1/25/~                      6                       1 2020-01-25              1
## 5 1/26/~                      8                       2 2020-01-26              1
## 6 1/27/~                      8                       0 2020-01-27              2
## # i abbreviated names: 1: `Cumulative Confirmed Cases`,
## #   2: `Daily Confirmed Cases`
```

```
plot(df$as_date,df$'Daily Confirmed Cases')

# filter data with date duration
du1 <- df[df$as_date >= as.Date("2021-08-01") & df$as_date <= as.Date("2021-08-31"), ]
du2 <- df[df$as_date >= as.Date("2021-08-16") & df$as_date <= as.Date("2021-09-15"), ]
du3 <- df[df$as_date >= as.Date("2021-09-01") & df$as_date <= as.Date("2021-09-30"), ]

points(du1$as_date, du1$'Daily Confirmed Cases', col = "blue", pch = 1)
points(du2$as_date, du2$'Daily Confirmed Cases', col = "red", pch = 1)
points(du3$as_date, du3$'Daily Confirmed Cases', col = "green", pch = 1)

legend("topleft", legend = c("Duration 1", "Duration 2", "Duration 3"), col = c("blue", "red", "green")
```



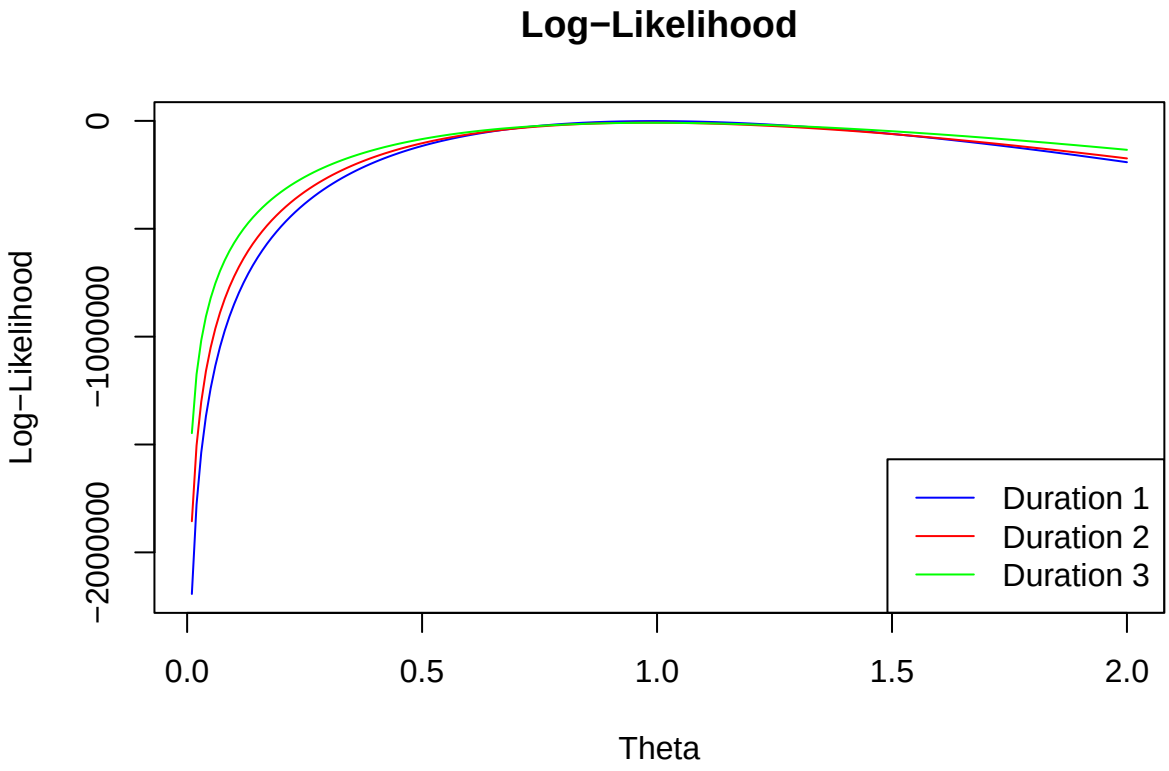
Calculate and Plot Log-Likelihood function

```r
# Cal log-likelihood for the Poisson distribution
log_likelihood <- function(theta, cases, prev_cases) {
  lambda <- theta * prev_cases
  log_lik <- sum(dpois(cases, lambda, log = TRUE))
  return(log_lik)
}

# Calculate log-likelihood for each theta
theta_values <- seq(0.01, 2, by = 0.01)
calculate_log_likelihood <- function(theta, df_subset) {
  cases <- df_subset$'Daily Confirmed Cases'
  prev_cases <- df_subset$Previous_Cases
  log_lik <- log_likelihood(theta, cases, prev_cases)
  return(log_lik)
}
log_likelihood_du1 <- c()
log_likelihood_du2 <- c()
log_likelihood_du3 <- c()

for (theta in theta_values) {
  log_likelihood_du1 <- c(log_likelihood_du1,calculate_log_likelihood(theta, du1))
  log_likelihood_du2 <- c(log_likelihood_du2,calculate_log_likelihood(theta, du2))
  log_likelihood_du3 <- c(log_likelihood_du3,calculate_log_likelihood(theta, du3))
}
```

```r
# Plot log-likelihood function for each duration
plot(theta_values, log_likelihood_du1, type = "l", col = "blue",xlab = "Theta", ylab = "Log-Likelihood"
lines(theta_values, log_likelihood_du2, col = "red")
lines(theta_values, log_likelihood_du3, col = "green")
legend("bottomright", legend = c("Duration 1", "Duration 2", "Duration 3"), col = c("blue", "red", "gre
```



Calculate MLE for each duration

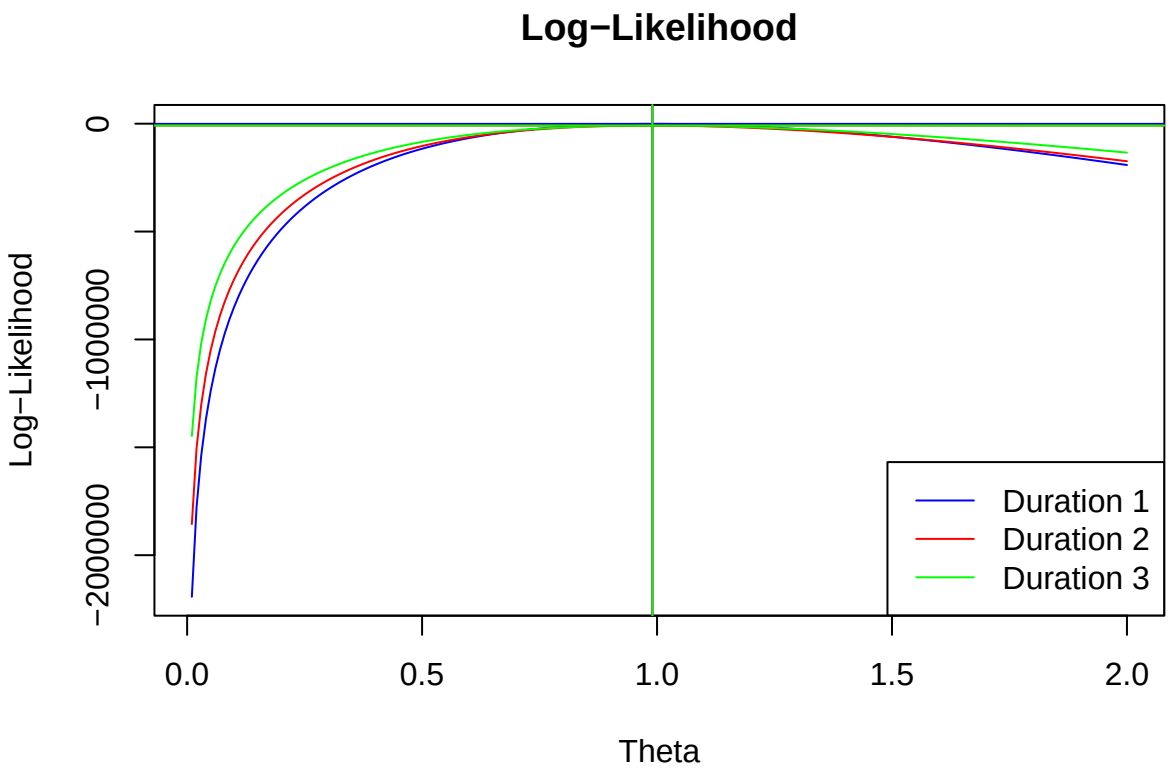
```r
# plot MLE
plot(theta_values, log_likelihood_du1, type = "l", col = "blue",xlab = "Theta", ylab = "Log-Likelihood"
lines(theta_values, log_likelihood_du2, col = "red")
lines(theta_values, log_likelihood_du3, col = "green")
legend("bottomright", legend = c("Duration 1", "Duration 2", "Duration 3"), col = c("blue", "red", "gre

# duration1
abline(h=max(log_likelihood_du1),col="blue")
theta_hat_du1 <- theta_values[which.max(log_likelihood_du1)]
abline(v=theta_hat_du1,col="blue")

# duration2
abline(h=max(log_likelihood_du2),col="red")
theta_hat_du2 <- theta_values[which.max(log_likelihood_du1)]
abline(v=theta_hat_du1,col="red")

# duration3
```

```r
abline(h=max(log_likelihood_du3),col="green")
theta_hat_du3 <- theta_values[which.max(log_likelihood_du3)]
abline(v=theta_hat_du3,col="green")
```




```r
# print MLE
cat('MLE for duration1 :', which.max(log_likelihood_du1),', Optimal-theta:', theta_hat_du1,"\n")
```

```
## MLE for duration1 : 99 , Optimal-theta: 0.99
```

```r
cat('MLE for duration2 :', which.max(log_likelihood_du2),', Optimal-theta:', theta_hat_du2,"\n")
```

```
## MLE for duration2 : 98 , Optimal-theta: 0.99
```

```r
cat('MLE for duration3 :', which.max(log_likelihood_du3),', Optimal-theta:', theta_hat_du3,"\n")
```

```
## MLE for duration3 : 99 , Optimal-theta: 0.99
```

For all durations (duration 1, duration 2, and duration 3) MLE almost same as 99,98,99, and optimal theta values for each duration are all 0.99