Analysis

2024-11-30

```
data <- read.csv("pandemic.csv")
head(data)
```

```
##    peak.inf soc.iso rate.vac quar.dur num.daily
## 1 0.8404217    0.00    0.000        0        10
## 2 0.5455691    0.27    0.000        0        10
## 3 0.7476246    0.40    0.000        0        10
## 4 0.3003751    0.81    0.000        0        10
## 5 0.7808691    0.00    0.003        0        10
## 6 0.6720373    0.27    0.003        0        10
```

```
summary(data)
```

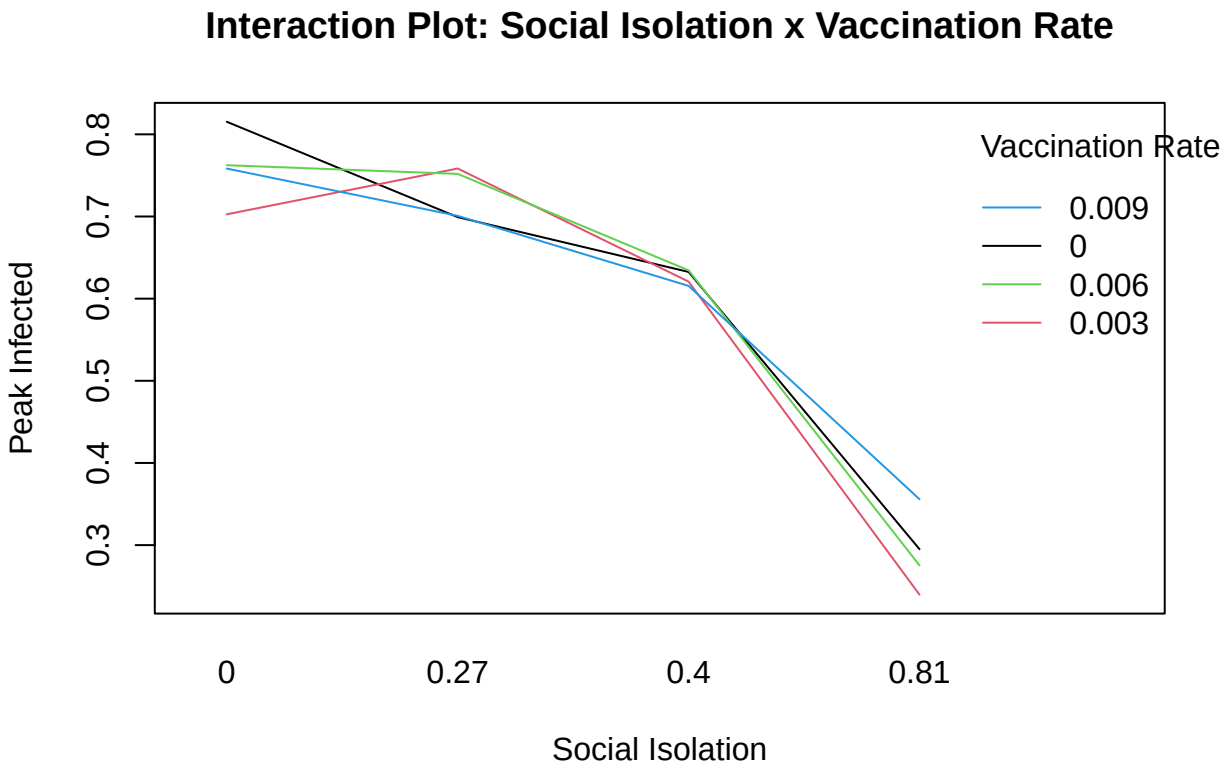
```
##     peak.inf         soc.iso          rate.vac          quar.dur
##  Min.   :0.0100   Min.   :0.0000   Min.   :0.00000   Min.   : 0
##  1st Qu.:0.4646   1st Qu.:0.2025   1st Qu.:0.00225   1st Qu.: 6
##  Median :0.7196   Median :0.3350   Median :0.00450   Median :12
##  Mean   :0.6011   Mean   :0.3700   Mean   :0.00450   Mean   :12
##  3rd Qu.:0.8271   3rd Qu.:0.5025   3rd Qu.:0.00675   3rd Qu.:18
##  Max.   :0.9211   Max.   :0.8100   Max.   :0.00900   Max.   :24
##    num.daily
##  Min.   :10.0
##  1st Qu.:17.5
##  Median :25.0
##  Mean   :25.0
##  3rd Qu.:32.5
##  Max.   :40.0
```

```
# peak.inf is i+q

# TODO need to consider the aliasing structure so that we know what tertiary effects cannot actually be

# TODO I wonder if our levels for vaccination rate are high enough

# TODO Treatment contrasts?

# Anova
anova_model <- aov(peak.inf ~ soc.iso * rate.vac *
                quar.dur * num.daily, data = data)
summary(anova_model)
```

```
##                                  Df Sum Sq Mean Sq F value   Pr(>F)
## soc.iso                           1  8.018   8.018 284.986  < 2e-16 ***
## rate.vac                          1  0.001   0.001   0.033   0.8568
## quar.dur                          1  1.141   1.141  40.553 9.66e-10 ***
## num.daily                         1  4.390   4.390 156.055  < 2e-16 ***
## soc.iso:rate.vac                  1  0.043   0.043   1.543   0.2154
```

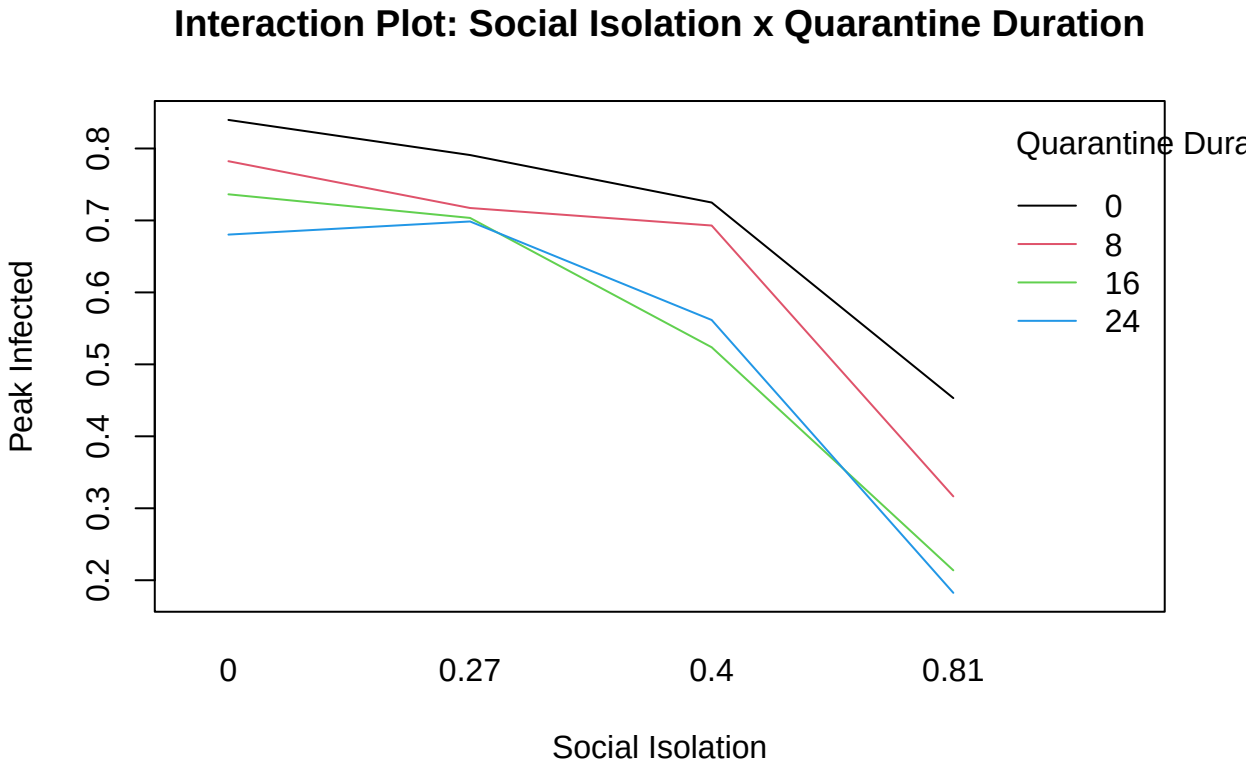
```
## soc.iso:quar.dur                       1  0.093   0.093   3.307   0.0702 .
## rate.vac:quar.dur                      1  0.022   0.022   0.783   0.3772
## soc.iso:num.daily                      1  0.076   0.076   2.700   0.1017
## rate.vac:num.daily                     1  0.160   0.160   5.684   0.0179 *
## quar.dur:num.daily                     1  0.140   0.140   4.959   0.0269 *
## soc.iso:rate.vac:quar.dur              1  0.000   0.000   0.011   0.9183
## soc.iso:rate.vac:num.daily             1  0.022   0.022   0.781   0.3776
## soc.iso:quar.dur:num.daily             1  0.171   0.171   6.064   0.0145 *
## rate.vac:quar.dur:num.daily            1  0.136   0.136   4.836   0.0288 *
## soc.iso:rate.vac:quar.dur:num.daily    1  0.008   0.008   0.283   0.5950
## Residuals                            240  6.752   0.028
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Interaction plots
interaction.plot(
  x.factor = data$soc.iso,
  trace.factor = data$rate.vac,
  response = data$peak.inf,
  col = 1:4,
  lty = 1,
  main = "Interaction Plot: Social Isolation x Vaccination Rate",
  xlab = "Social Isolation",
  ylab = "Peak Infected",
  trace.label = "Vaccination Rate"
)
```

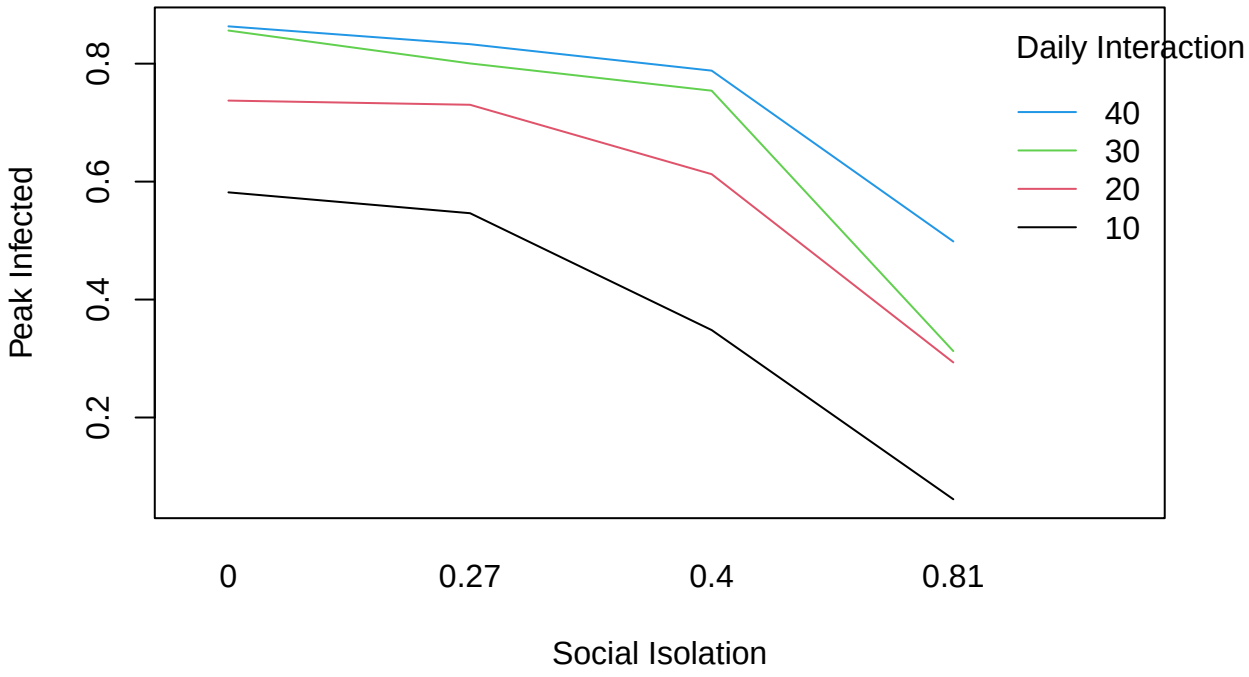



```r
interaction.plot(
  x.factor = data$soc.iso,
  trace.factor = data$quar.dur,
  response = data$peak.inf,
```

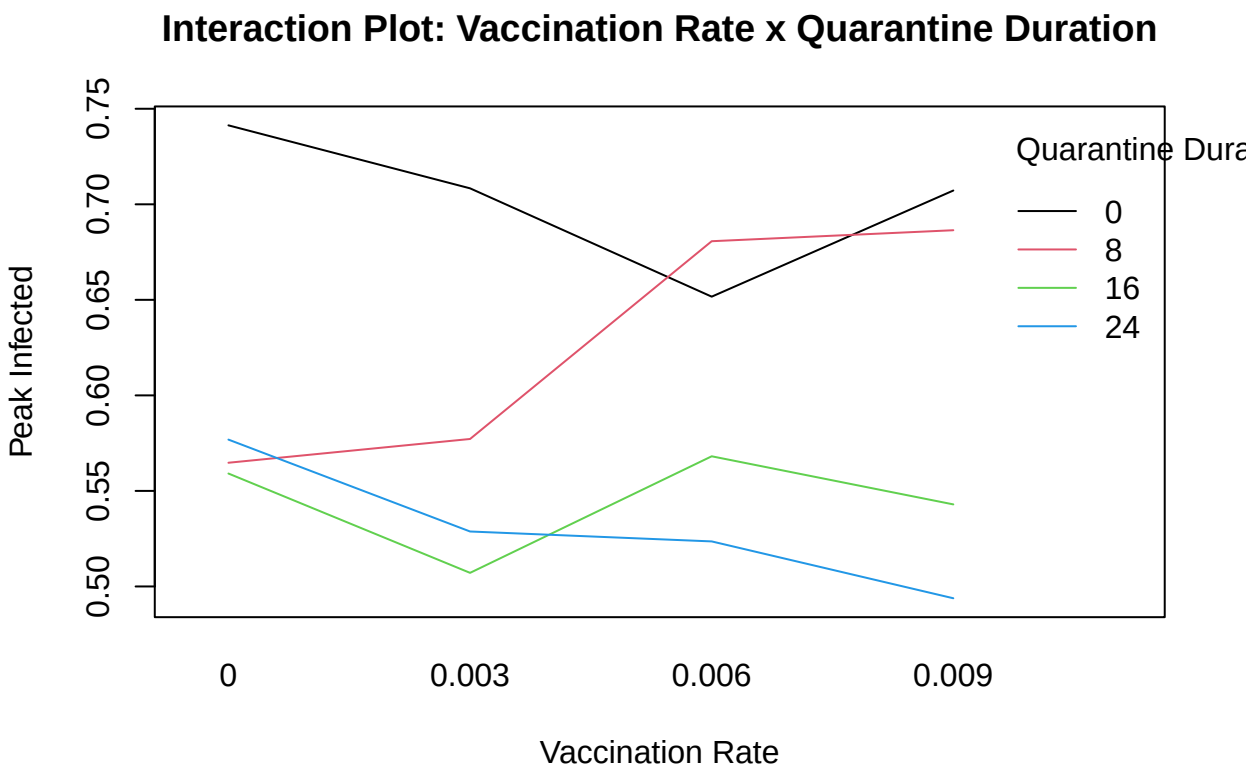
```
  col = 1:4,
  lty = 1,
  main = "Interaction Plot: Social Isolation x Quarantine Duration",
  xlab = "Social Isolation",
  ylab = "Peak Infected",
  trace.label = "Quarantine Duration"
)
```



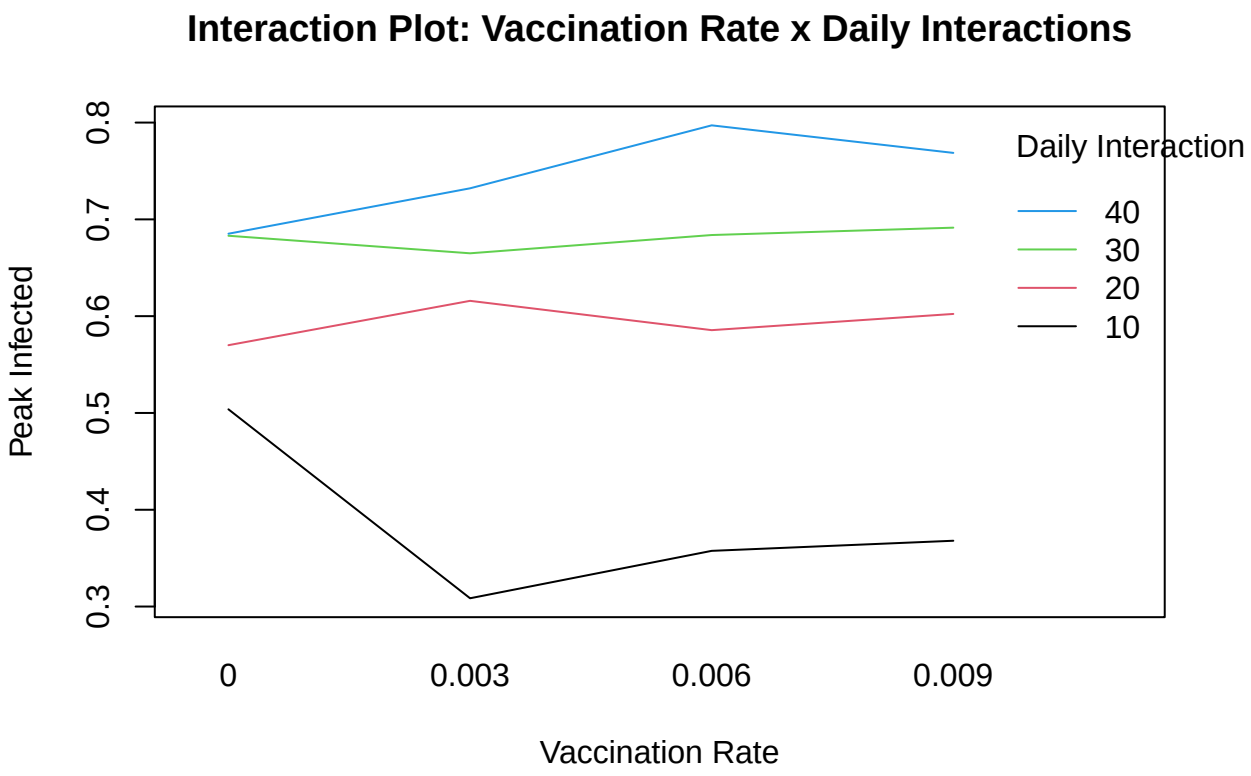

```
interaction.plot(
  x.factor = data$soc.iso,
  trace.factor = data$num.daily,
  response = data$peak.inf,
  col = 1:4,
  lty = 1,
  main = "Interaction Plot: Social Isolation x Daily Interactions",
  xlab = "Social Isolation",
  ylab = "Peak Infected",
  trace.label = "Daily Interactions"
)
```



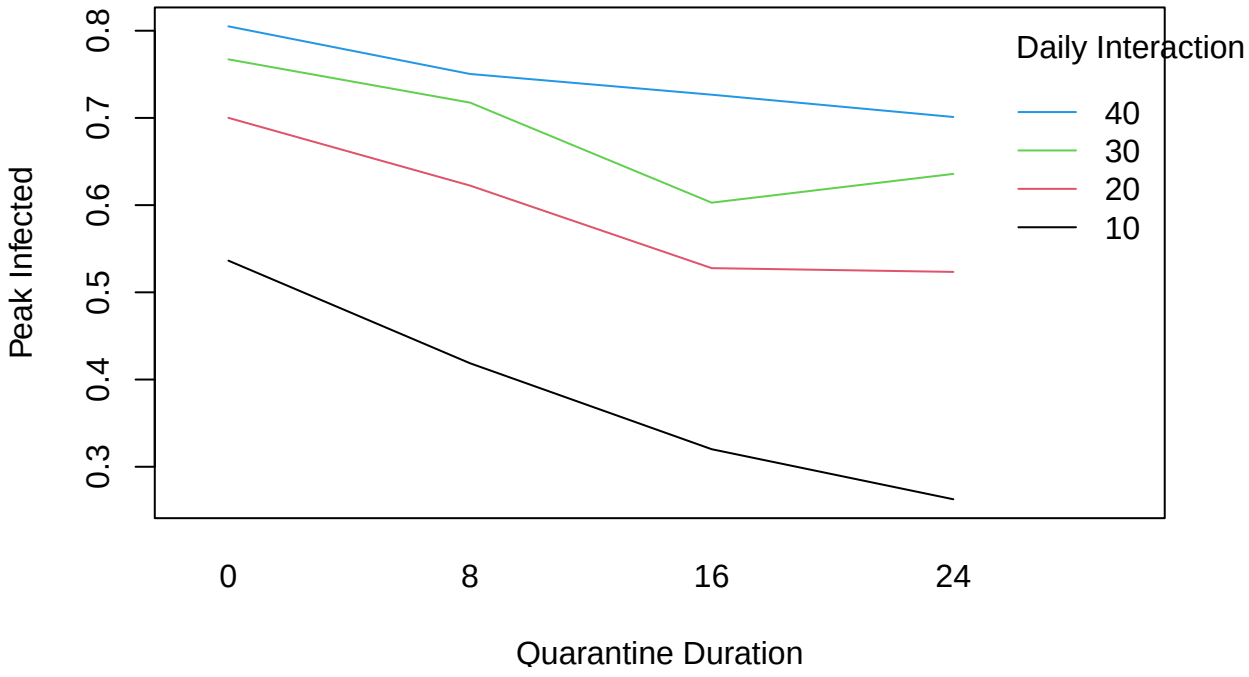

```r
interaction.plot(
  x.factor = data$rate.vac,
  trace.factor = data$quar.dur,
  response = data$peak.inf,
  col = 1:4,
  lty = 1,
  main = "Interaction Plot: Vaccination Rate x Quarantine Duration",
  xlab = "Vaccination Rate",
  ylab = "Peak Infected",
  trace.label = "Quarantine Duration"
)
```




```r
interaction.plot(
  x.factor = data$rate.vac,
  trace.factor = data$num.daily,
  response = data$peak.inf,
  col = 1:4,
  lty = 1,
  main = "Interaction Plot: Vaccination Rate x Daily Interactions",
  xlab = "Vaccination Rate",
  ylab = "Peak Infected",
  trace.label = "Daily Interactions"
)
```




```r
interaction.plot(
  x.factor = data$quar.dur,
  trace.factor = data$num.daily,
  response = data$peak.inf,
  col = 1:4,
  lty = 1,
  main = "Interaction Plot: Quarantine Duration x Daily Interactions",
  xlab = "Quarantine Duration",
  ylab = "Peak Infected",
  trace.label = "Daily Interactions"
)
```

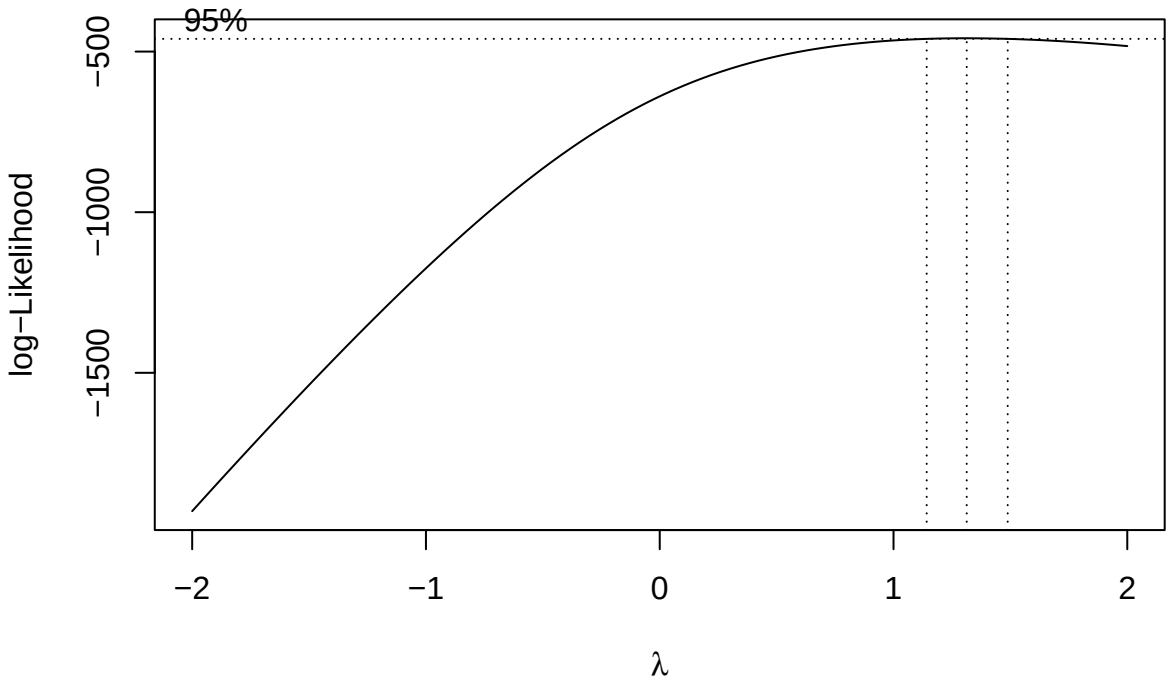



```
# Regression and Box-Cox
lm_model <- lm(peak.inf ~ soc.iso * rate.vac *
                quar.dur * num.daily, data = data)
summary(lm_model)
```

```
##
## Call:
## lm(formula = peak.inf ~ soc.iso * rate.vac * quar.dur * num.daily,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.63218 -0.09738  0.01238  0.10752  0.37489
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          7.255e-01  1.160e-01   6.253 1.83e-09 ***
## soc.iso                             -7.963e-01  2.461e-01  -3.235  0.00139 **
## rate.vac                             1.801e+01  2.067e+01   0.871  0.38459
## quar.dur                            -7.075e-03  7.752e-03  -0.913  0.36233
## num.daily                            6.296e-03  4.237e-03   1.486  0.13854
## soc.iso:rate.vac                    -2.641e+01  4.385e+01  -0.602  0.54748
## soc.iso:quar.dur                     7.836e-03  1.644e-02   0.477  0.63413
## rate.vac:quar.dur                   -2.600e+00  1.381e+00  -1.882  0.06103 .
## soc.iso:num.daily                    8.995e-03  8.986e-03   1.001  0.31785
## rate.vac:num.daily                  -7.241e-01  7.548e-01  -0.959  0.33836
## quar.dur:num.daily                   1.563e-04  2.831e-04   0.552  0.58147
## soc.iso:rate.vac:quar.dur            1.546e+00  2.930e+00   0.528  0.59813
## soc.iso:rate.vac:num.daily           1.529e+00  1.601e+00   0.955  0.34053
## soc.iso:quar.dur:num.daily          -6.273e-04  6.004e-04  -1.045  0.29716
```

```
## rate.vac:quar.dur:num.daily          8.980e-02  5.044e-02   1.781  0.07625 .
## soc.iso:rate.vac:quar.dur:num.daily -5.694e-02  1.070e-01  -0.532  0.59503
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1677 on 240 degrees of freedom
## Multiple R-squared:  0.6811, Adjusted R-squared:  0.6612
## F-statistic: 34.17 on 15 and 240 DF,  p-value: < 2.2e-16
```

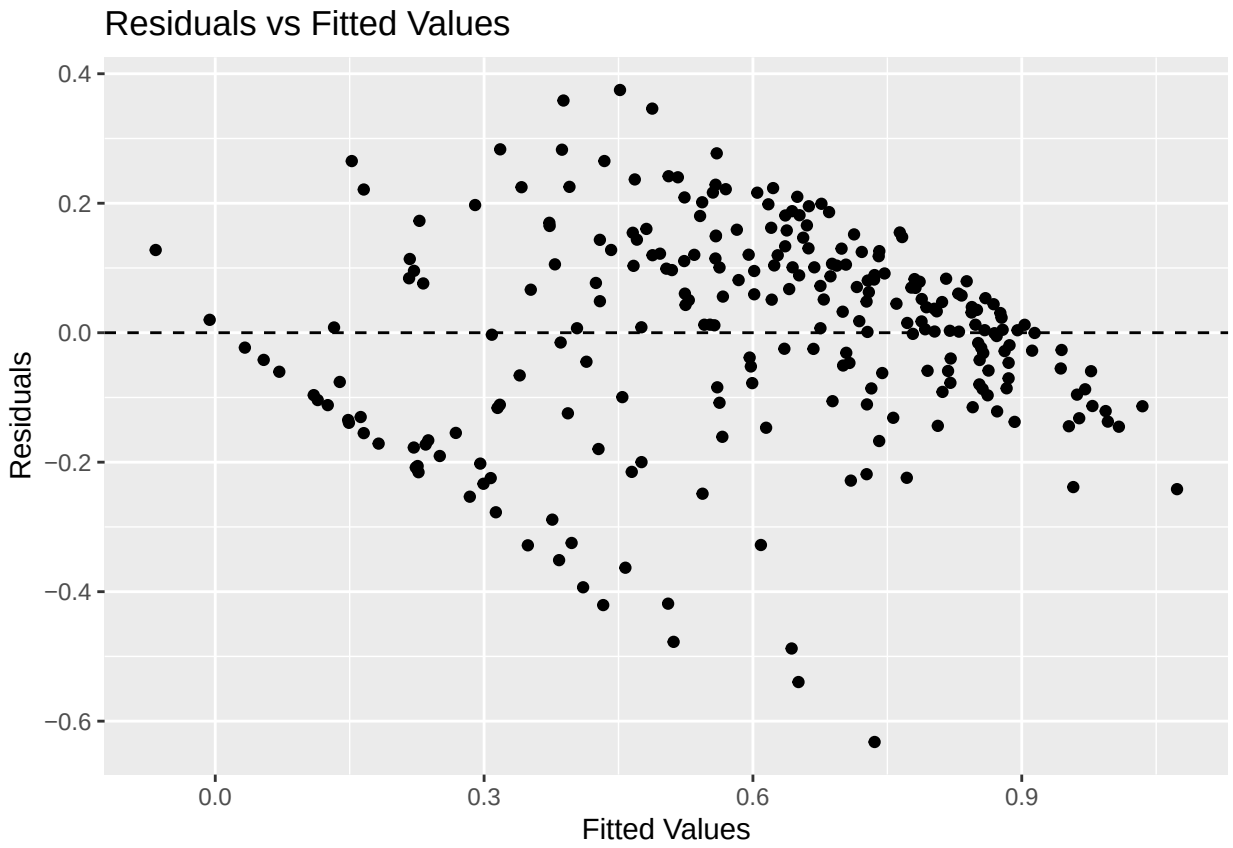
```r
# Box-Cox Transformation
boxcox_result <- boxcox(lm_model, lambda = seq(-2, 2, by = 0.1))
```



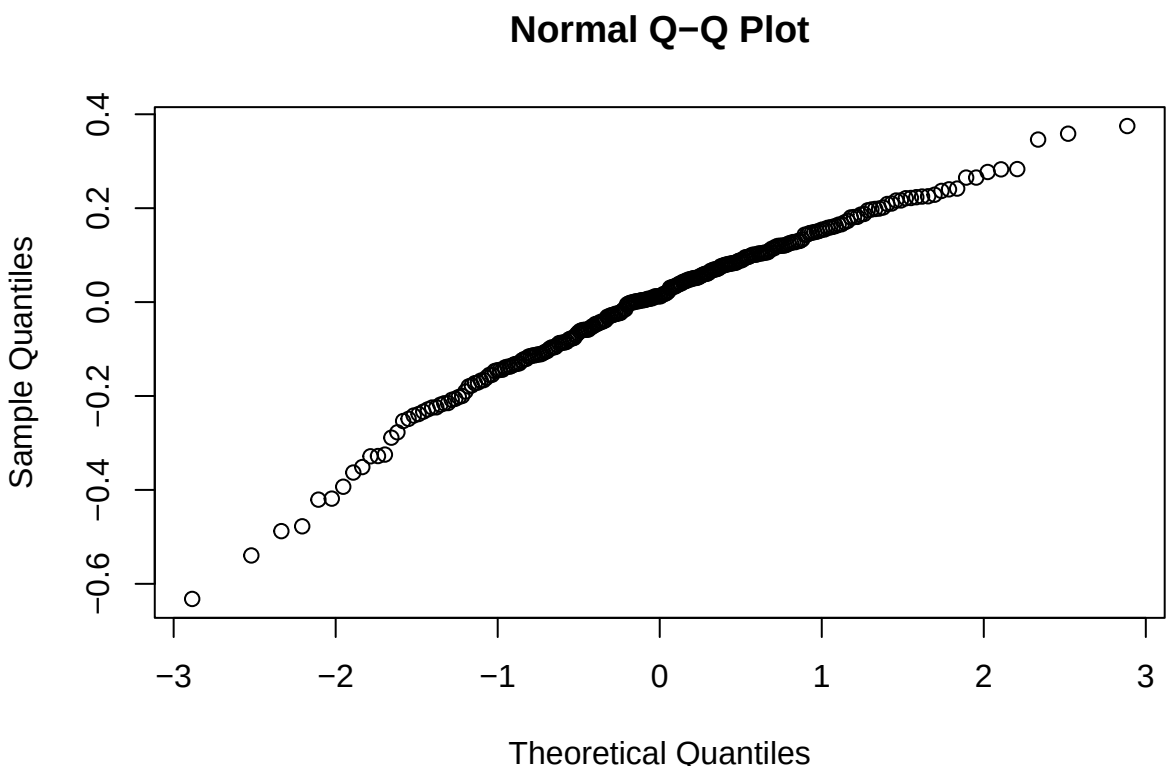

```r
optimal_lambda <- boxcox_result$x[which.max(boxcox_result$y)]
optimal_lambda # no transformation
```

```
## [1] 1.313131
```

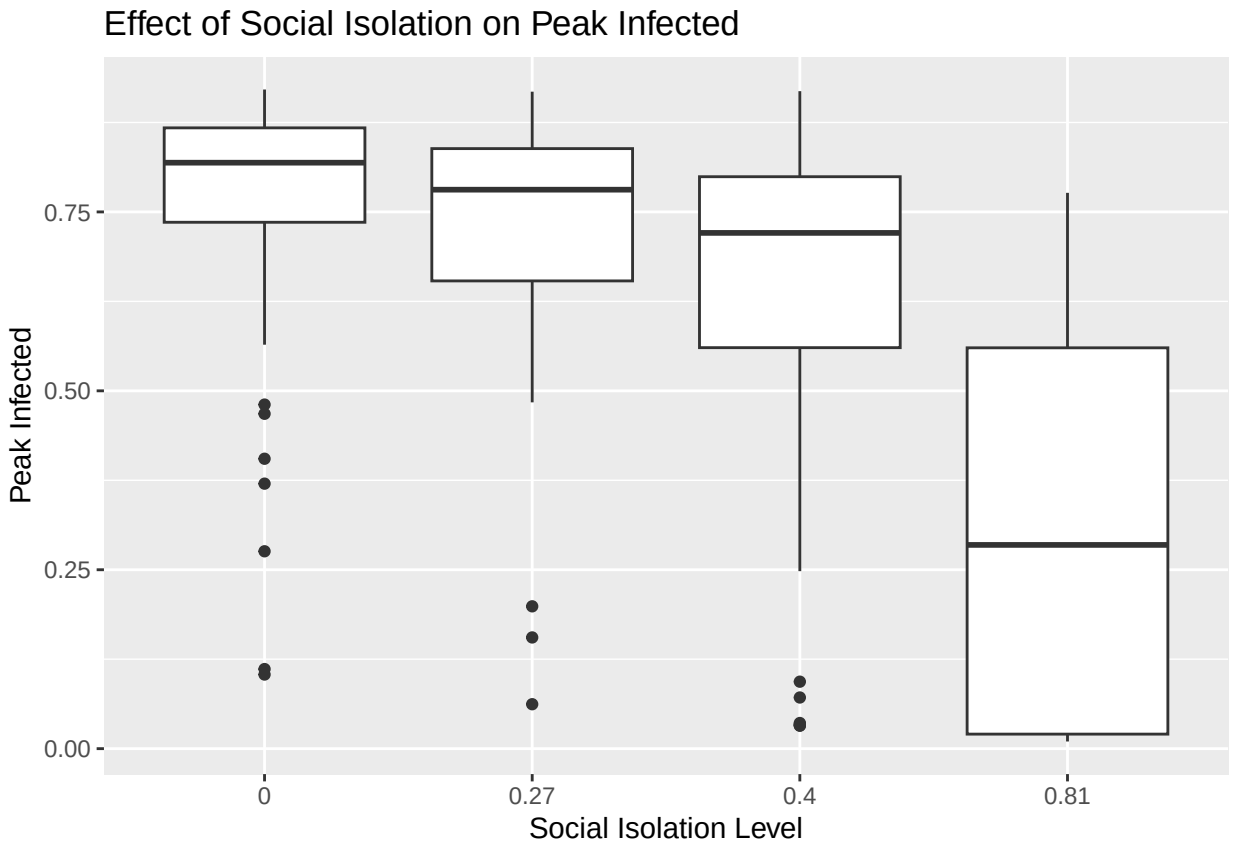
```r
# Residual plots
residuals_plot <- ggplot(data.frame(fitted = lm_model$fitted.values,
                                    residuals = lm_model$residuals),
                         aes(x = fitted, y = residuals)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  labs(title = "Residuals vs Fitted Values", x = "Fitted Values", y = "Residuals")
print(residuals_plot)
```



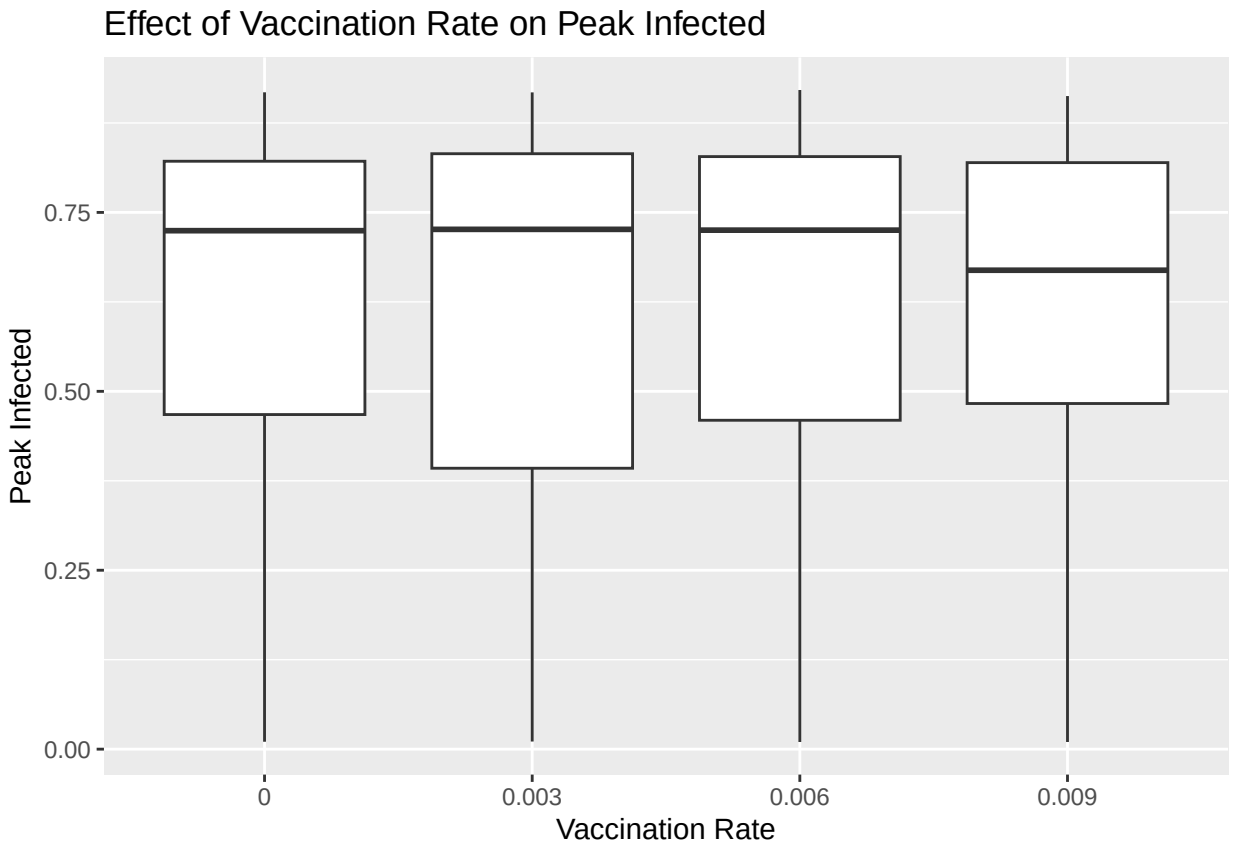

```
qqnorm(resid(lm_model))
```



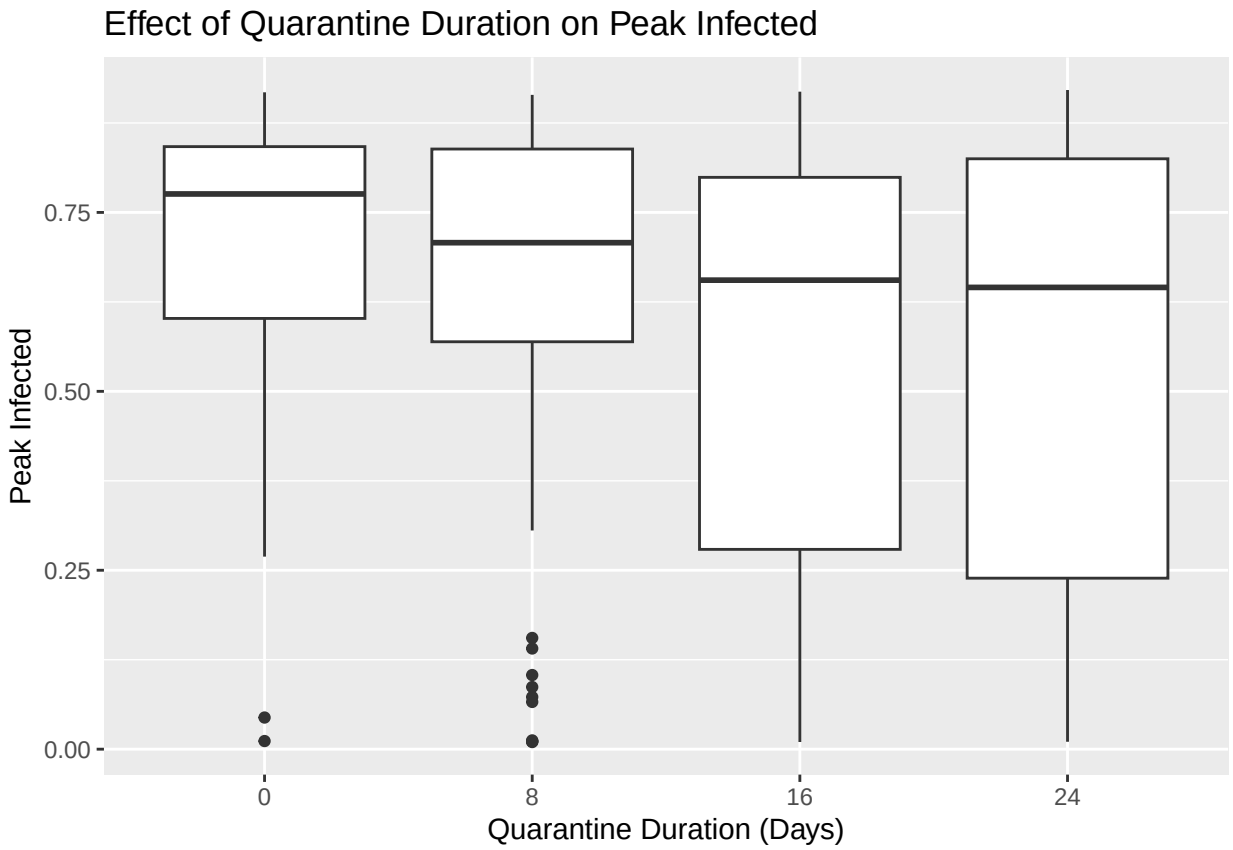
```r
# Bunch of other plots
ggplot(data, aes(x = as.factor(soc.iso), y = peak.inf)) +
  geom_boxplot() +
  labs(title = "Effect of Social Isolation on Peak Infected",
       x = "Social Isolation Level", y = "Peak Infected")
```



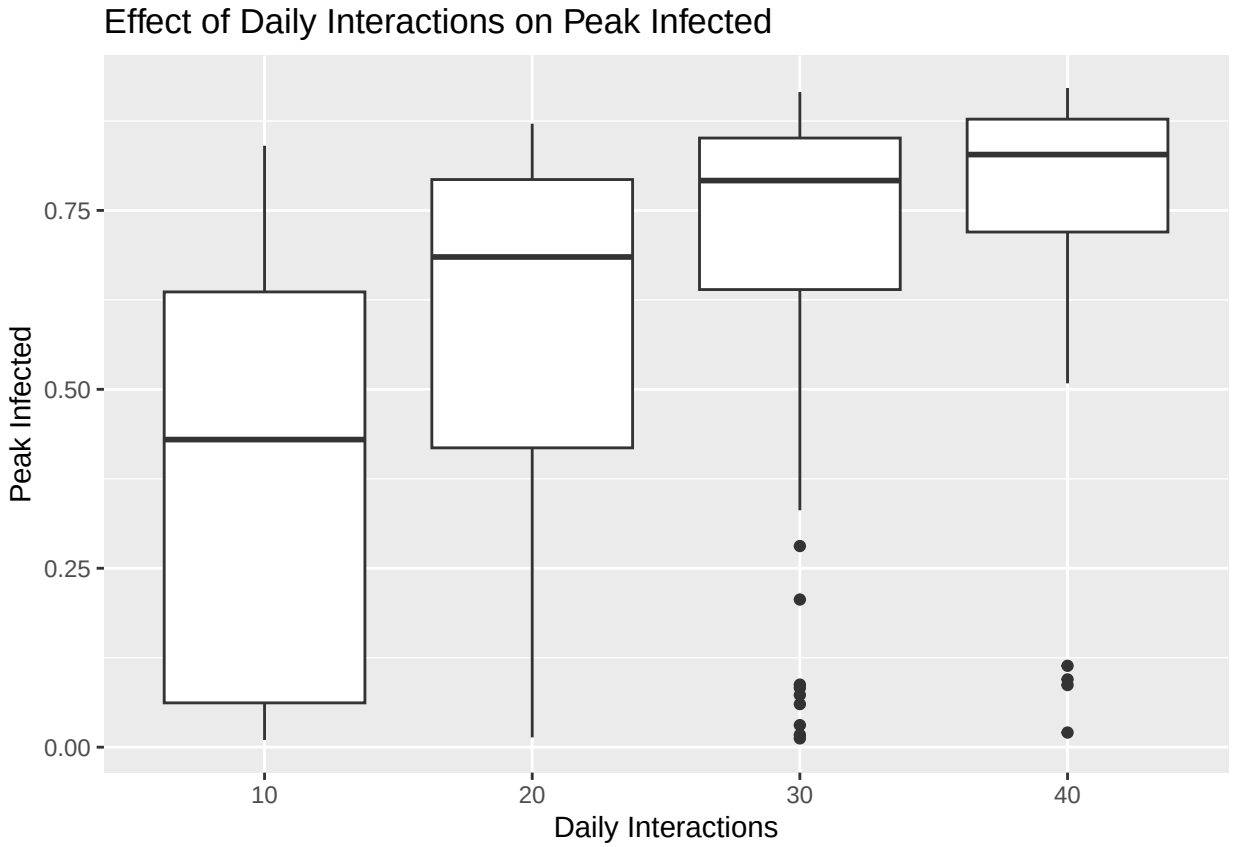

```r
ggplot(data, aes(x = as.factor(rate.vac), y = peak.inf)) +
  geom_boxplot() +
  labs(title = "Effect of Vaccination Rate on Peak Infected",
       x = "Vaccination Rate", y = "Peak Infected")
```



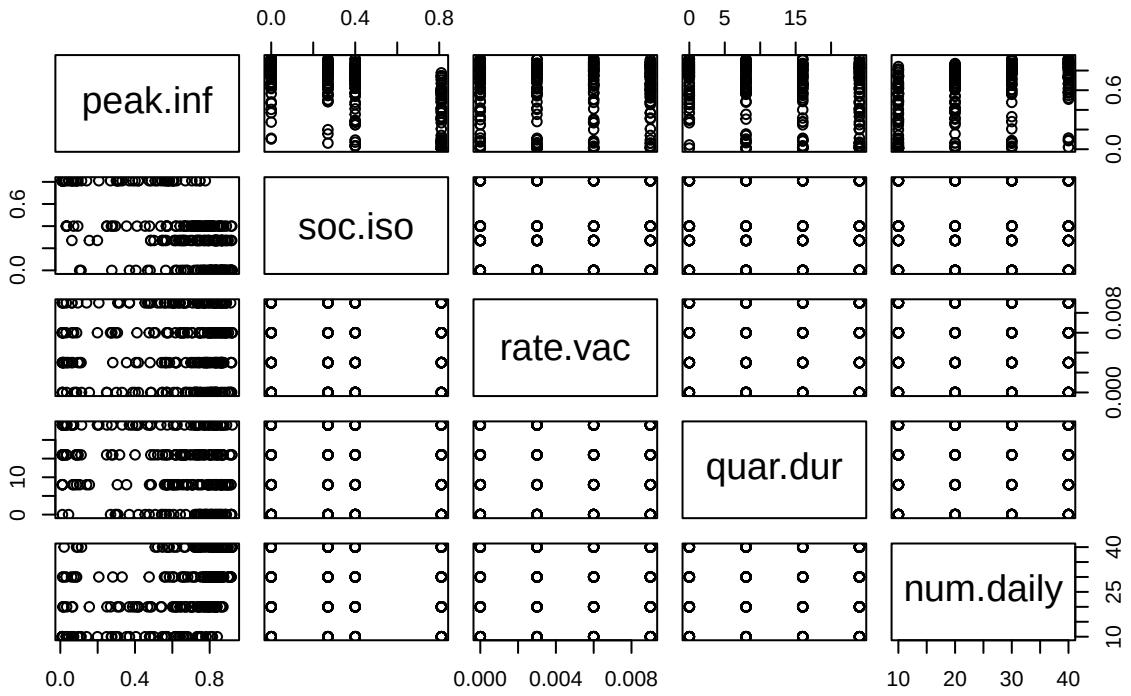

```r
ggplot(data, aes(x = as.factor(quar.dur), y = peak.inf)) +
  geom_boxplot() +
  labs(title = "Effect of Quarantine Duration on Peak Infected",
    x = "Quarantine Duration (Days)", y = "Peak Infected")
```



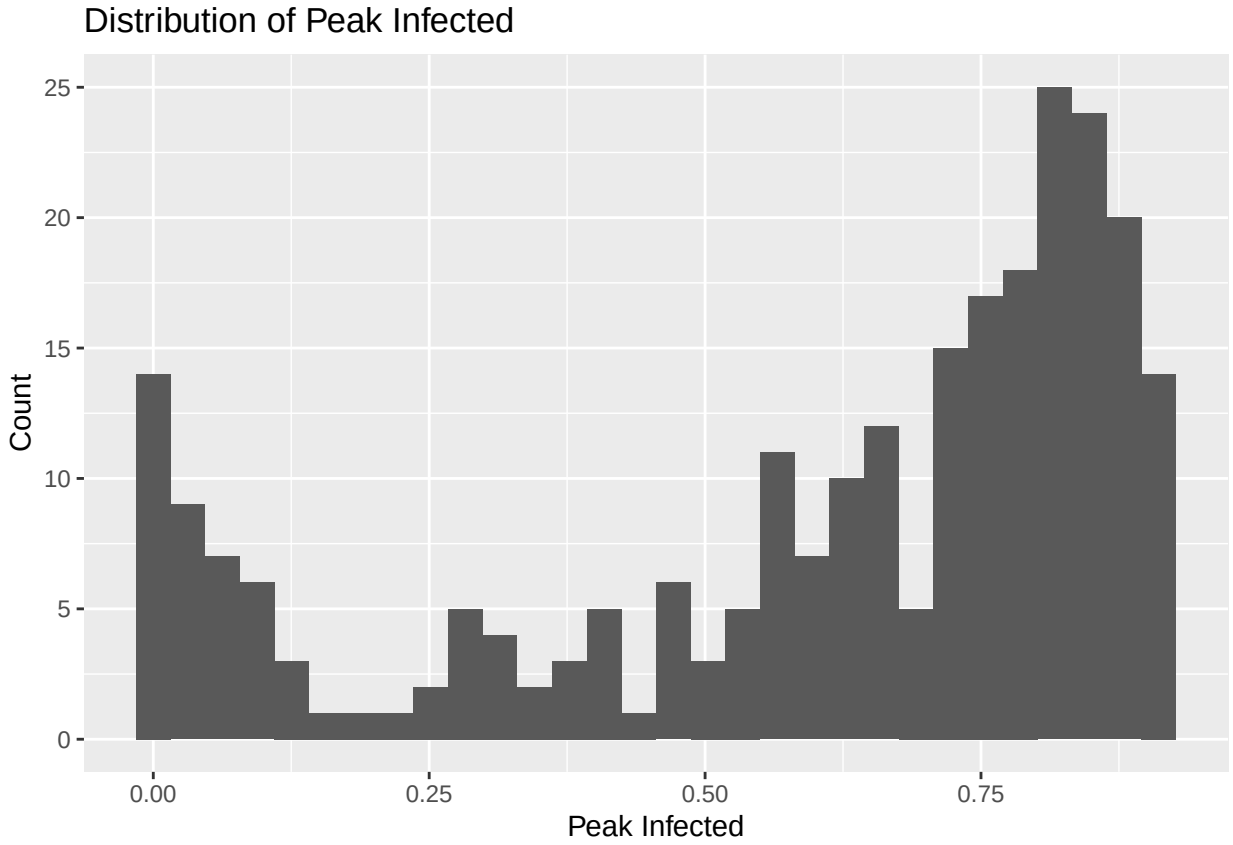

```
ggplot(data, aes(x = as.factor(num.daily), y = peak.inf)) +
  geom_boxplot() +
  labs(title = "Effect of Daily Interactions on Peak Infected",
    x = "Daily Interactions", y = "Peak Infected")
```



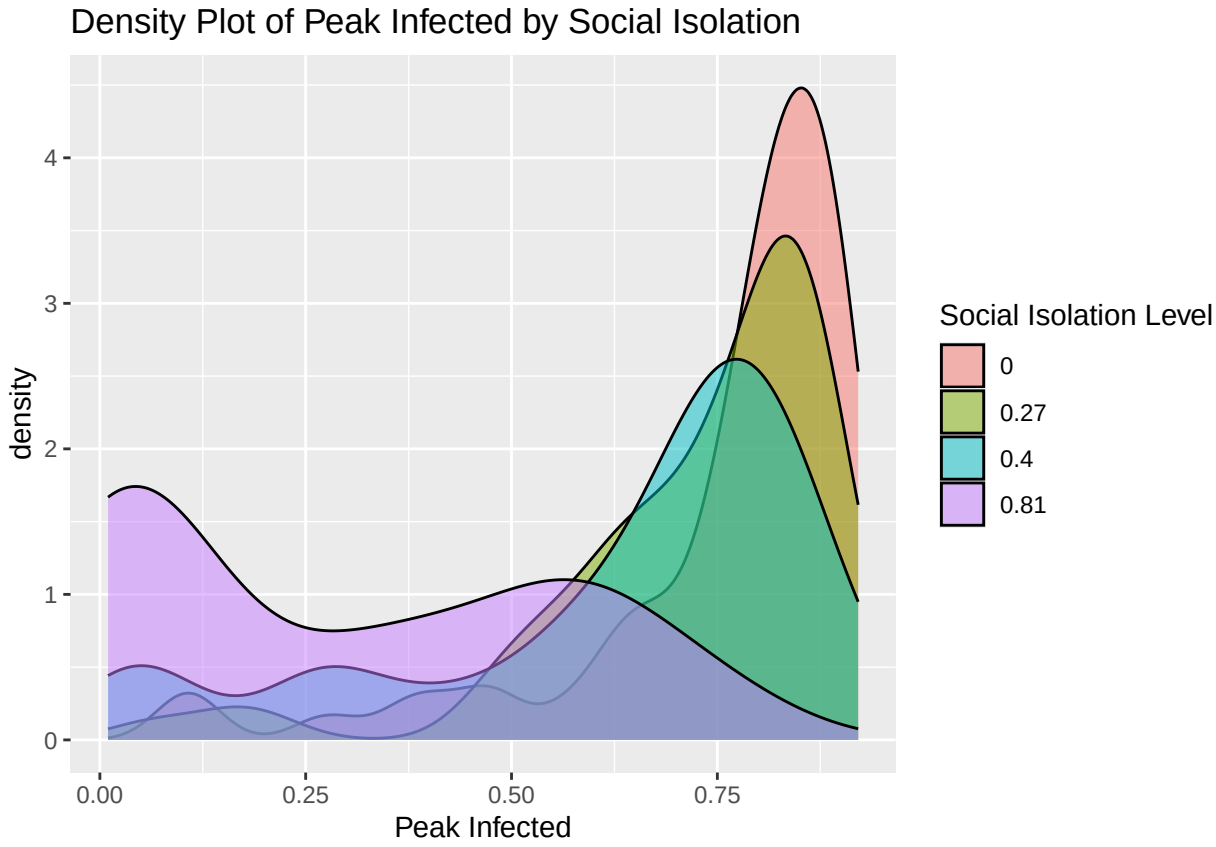

```
pairs(data[, c("peak.inf", "soc.iso", "rate.vac",
               "quar.dur", "num.daily")],
      main = "Scatterplot Matrix") #not useful
```



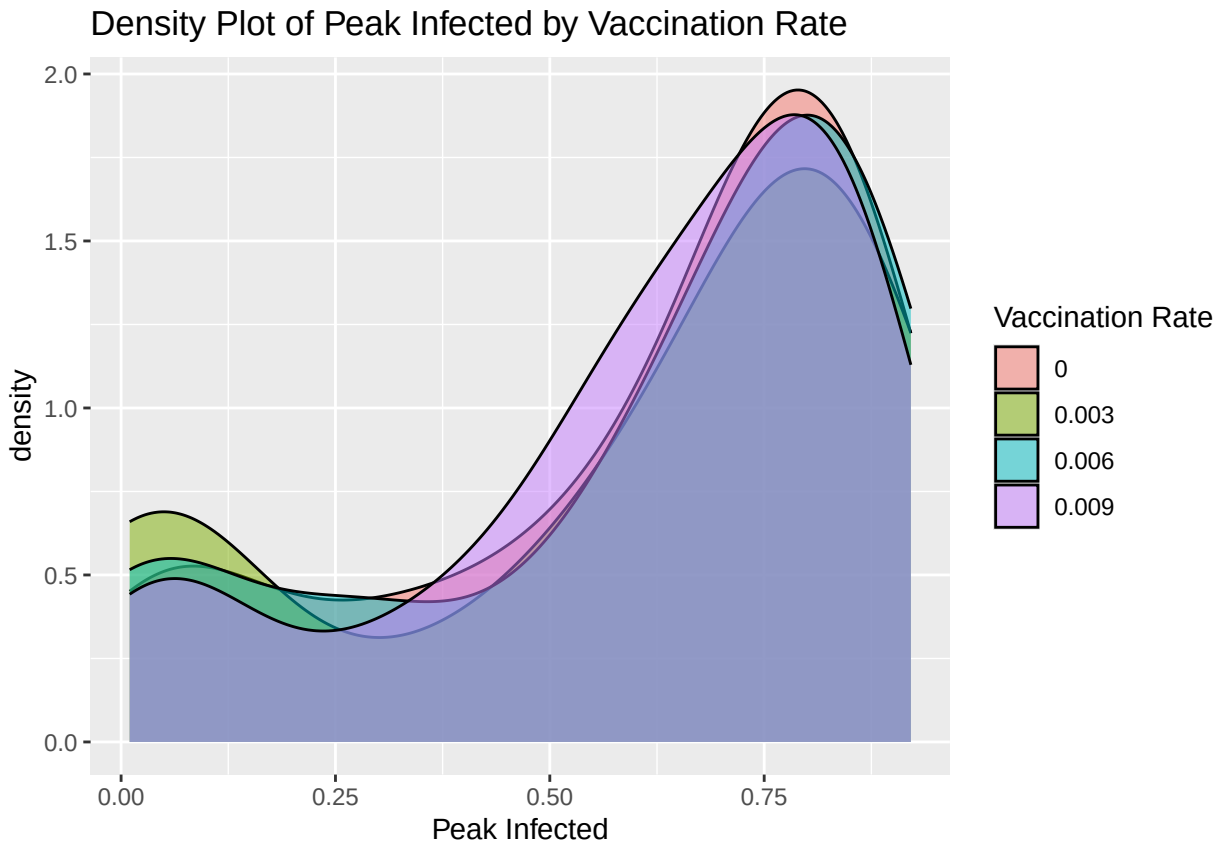

```
ggplot(data, aes(x = peak.inf)) +
  geom_histogram(bins = 30) +
  labs(title = "Distribution of Peak Infected", x = "Peak Infected", y = "Count")
```



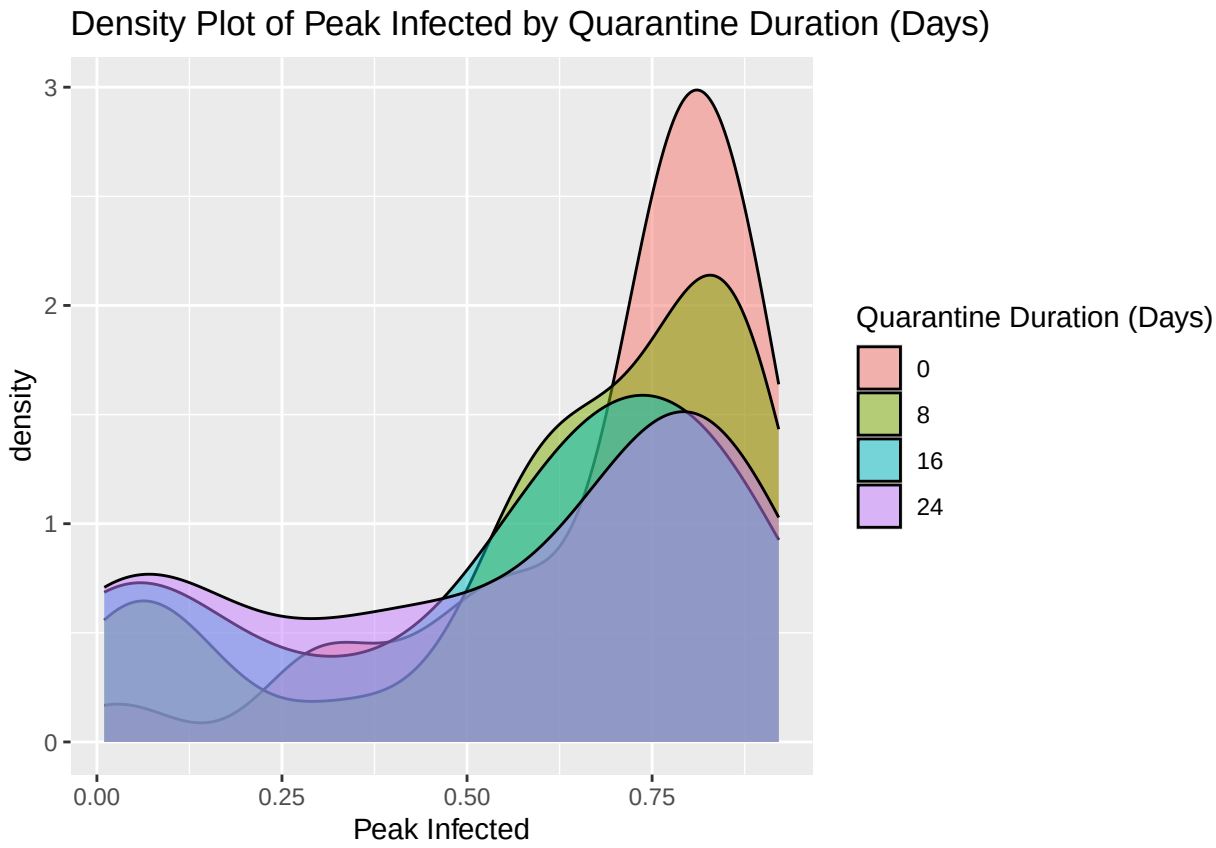
```
ggplot(data, aes(x = peak.inf, fill = as.factor(soc.iso))) +
  geom_density(alpha = 0.5) +
  labs(title = "Density Plot of Peak Infected by Social Isolation",
       x = "Peak Infected", fill = "Social Isolation Level")
```



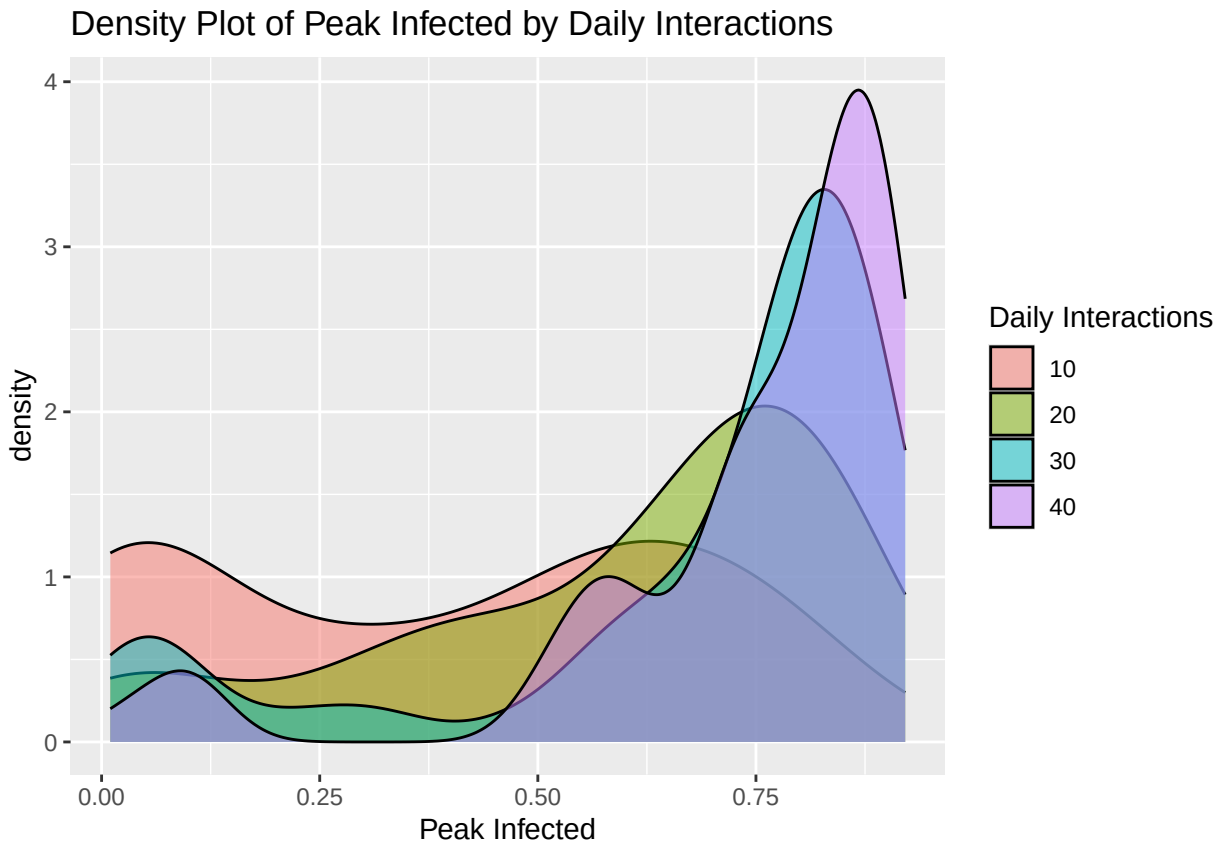

```
ggplot(data, aes(x = peak.inf, fill = as.factor(rate.vac))) +
  geom_density(alpha = 0.5) +
  labs(title = "Density Plot of Peak Infected by Vaccination Rate",
       x = "Peak Infected", fill = "Vaccination Rate")
```




```
ggplot(data, aes(x = peak.inf, fill = as.factor(quar.dur))) +
  geom_density(alpha = 0.5) +
  labs(title = "Density Plot of Peak Infected by Quarantine Duration (Days)",
       x = "Peak Infected", fill = "Quarantine Duration (Days)")
```




```
ggplot(data, aes(x = peak.inf, fill = as.factor(num.daily))) +
  geom_density(alpha = 0.5) +
  labs(title = "Density Plot of Peak Infected by Daily Interactions",
       x = "Peak Infected", fill = "Daily Interactions")
```


Density Plot of Peak Infected by Daily Interactions
density
Peak Infected
0.00
0.25
0.50
0.75
0
1
2
3
4
Daily Interactions
10
20
30
40