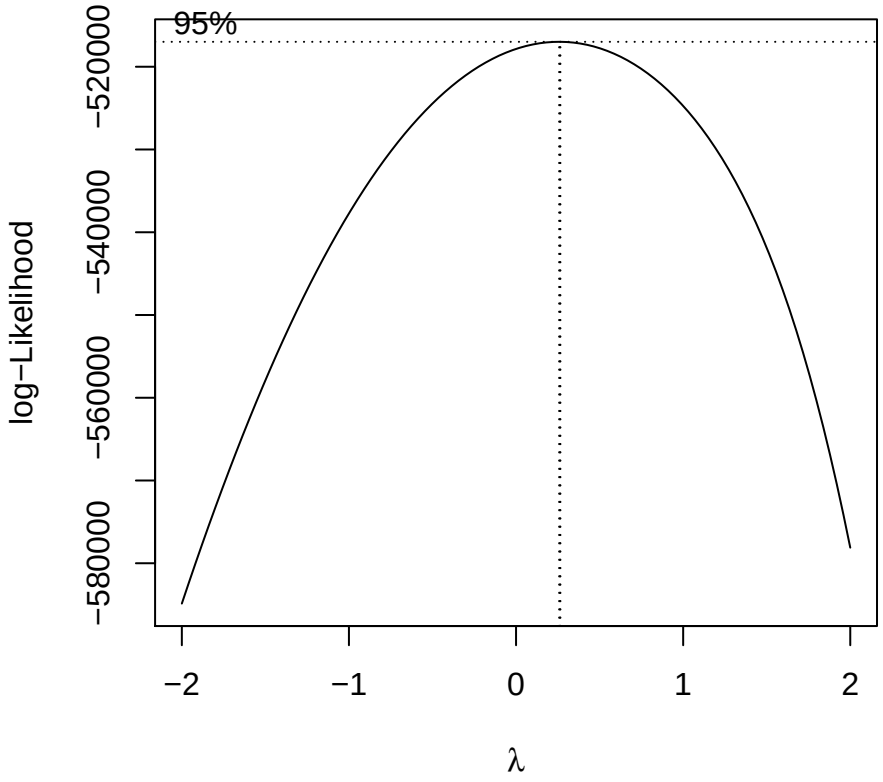
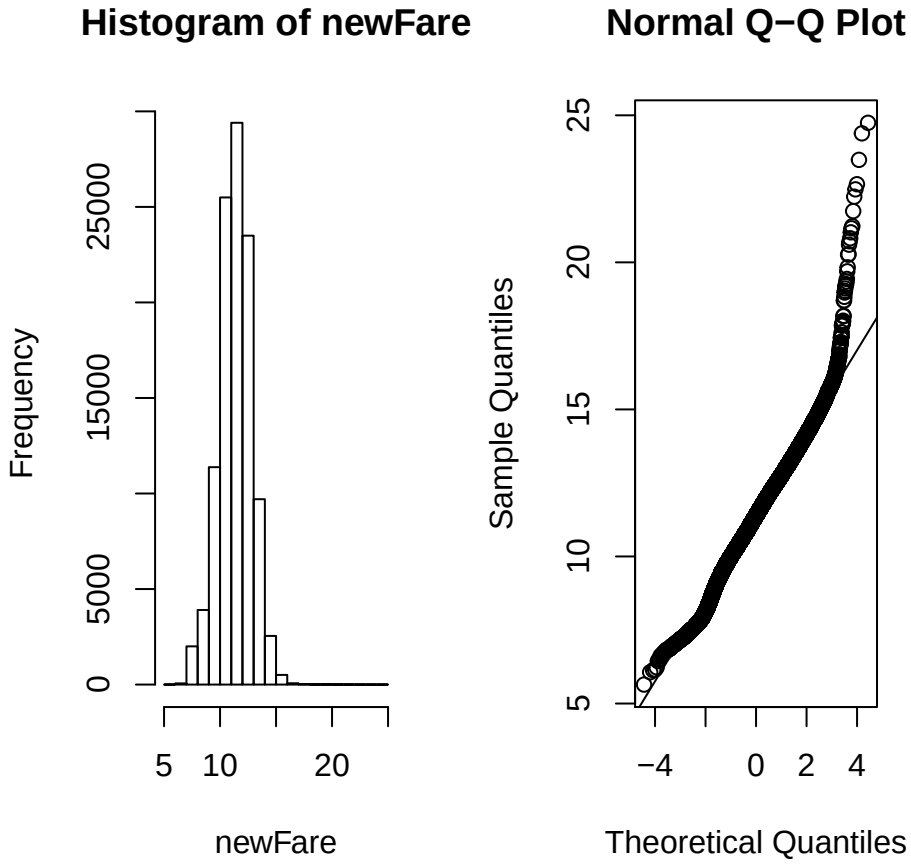
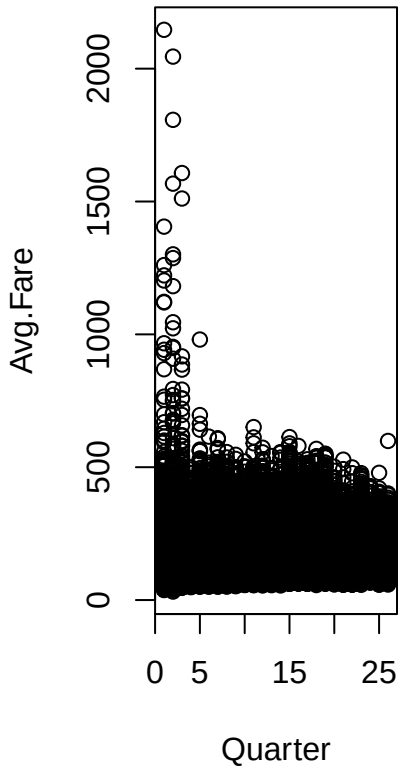
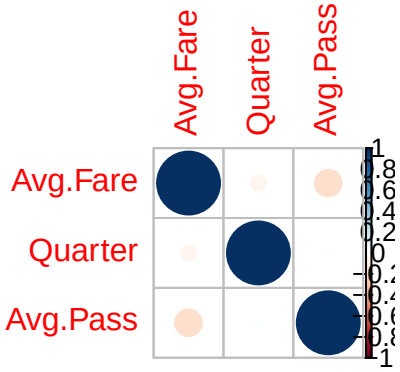



```
# optimal lambda
print(max.bc <- box$x[box$y==max(box$y)])

## [1] 0.2626263

newFare <- (rout$Avg.Fare^max.bc - 1)/(max.bc)
par(mfrow = c(1,2))

hist(newFare)
qqnorm(newFare)
qqline(newFare) #The log-transform is more simple, we will use log
```



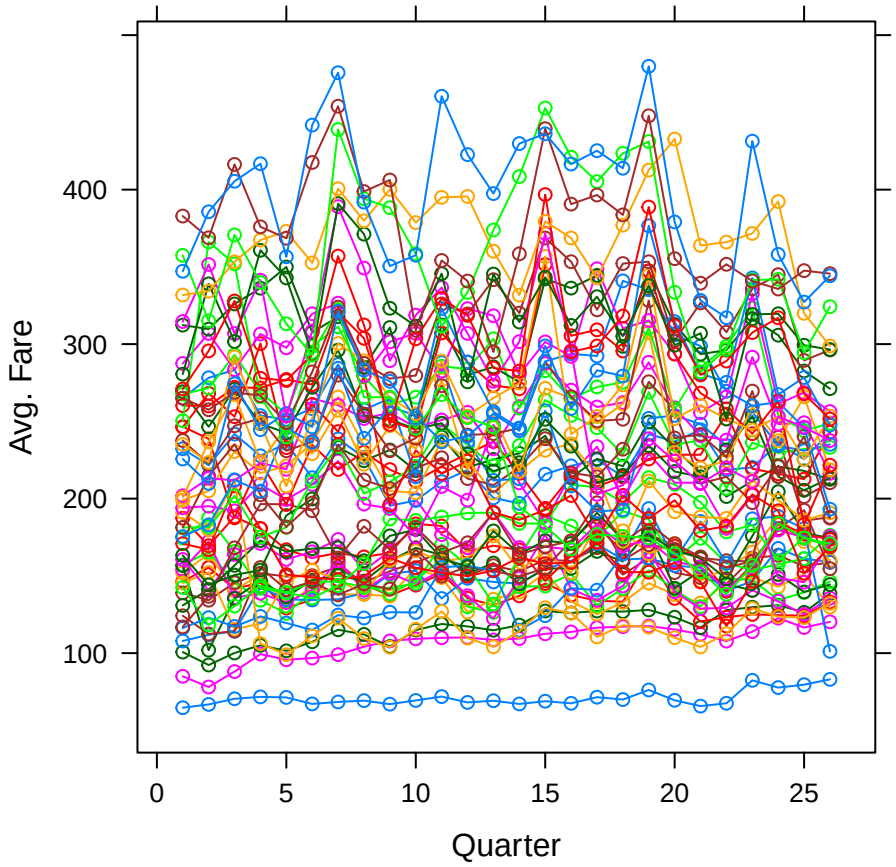

```
newFare <- log(rout$Avg.Fare)

plot(Avg.Fare~Quarter, data = rout)

corrplot(cor(data.frame(Avg.Fare, Quarter, Avg.Pass))) #Covariates seem to be independent
```






```
##############################################################################################################
lme(Avg.Fare ~ 1, random = ~ 1 | Route, data=rout)

## Linear mixed-effects model fit by REML
##   Data: rout
##   Log-restricted-likelihood: -552366.6
##   Fixed: Avg.Fare ~ 1
## (Intercept)
##    202.4471
##
## Random effects:
##  Formula: ~1 | Route
##         (Intercept) Residual
## StdDev:    64.11781 35.94851
##
## Number of Observations: 108602
## Number of Groups: 4177

xyplot(sub2$Avg.Fare~sub2$Quarter, xlab = "Quarter", ylab = "Avg. Fare",
               type="b",
               group=sub2$Route,
```

```
        data=sub2,
        auto.key =F
    ) #Essentially how much each route made over quarters
```

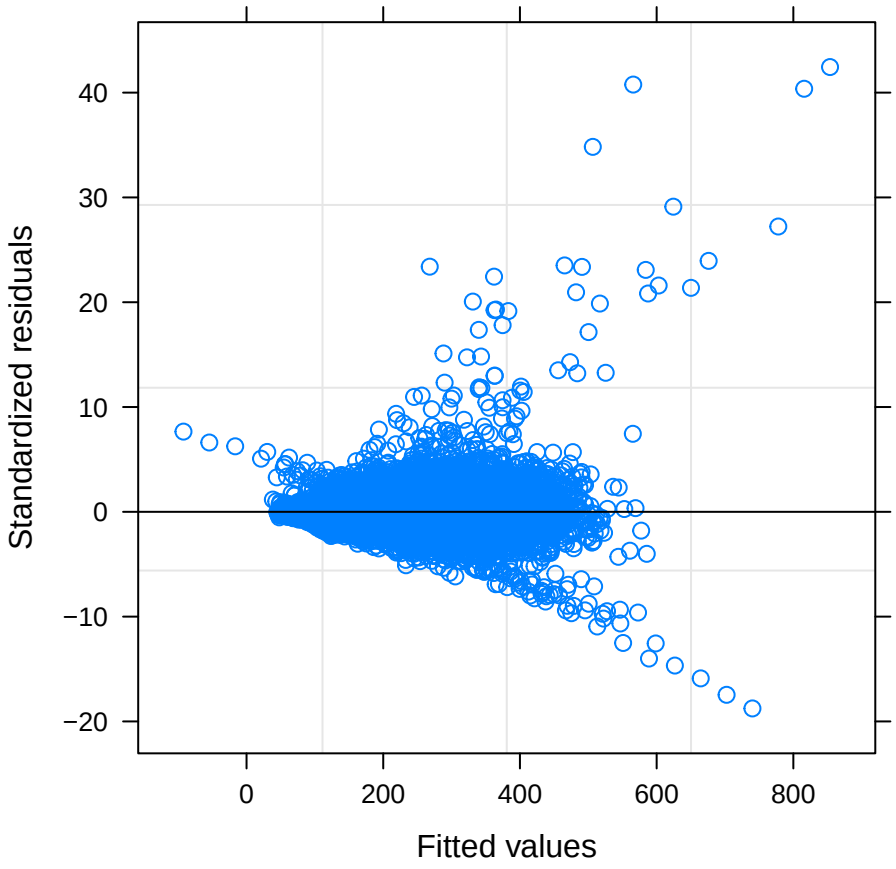



```
#unconditional growth model
fit0 <- lme(Avg.Fare ~ Quarter, random = ~ Quarter | Route, data=rout,
            method = "ML") #random intercept and slope
summary(fit0)
```

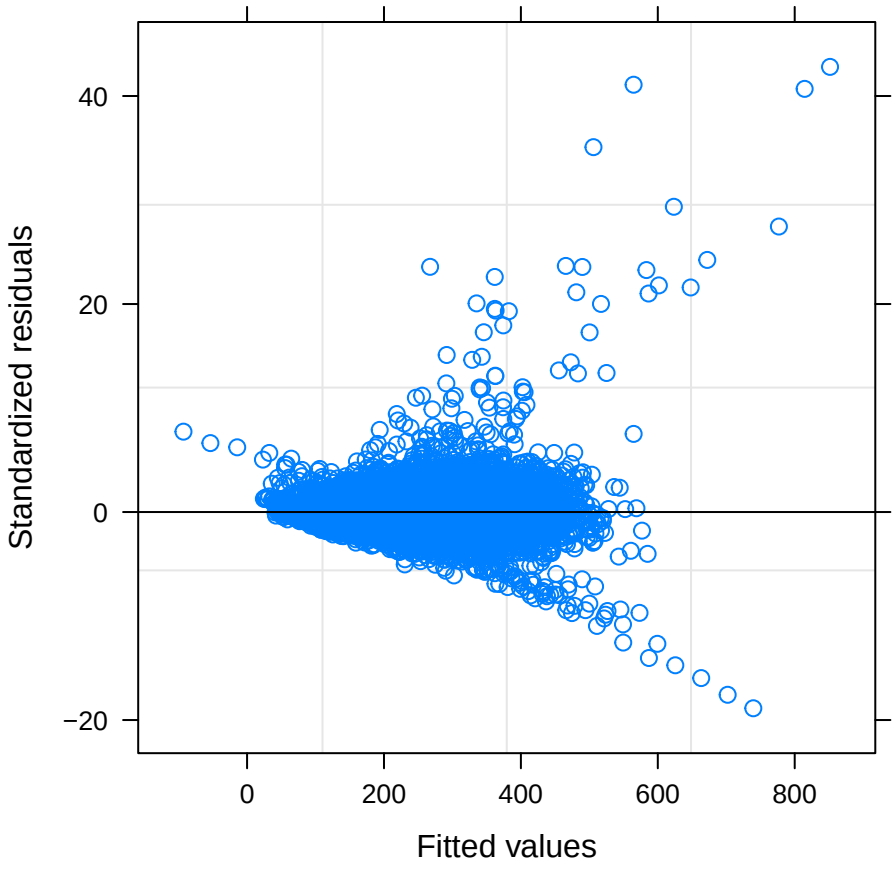
```
## Linear mixed-effects model fit by maximum likelihood
##  Data: rout
##        AIC      BIC    logLik
##    1079268 1079326 -539628.1
##
## Random effects:
##  Formula: ~Quarter | Route
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 82.929646 (Intr)
## Quarter      2.436831 -0.703
```

```
## Residual    30.457995
##
## Fixed effects: Avg.Fare ~ Quarter
##                 Value Std.Error     DF   t-value p-value
## (Intercept) 209.77249 1.2971987 104424 161.71191       0
## Quarter      -0.54262 0.0396676 104424 -13.67915       0
##  Correlation:
##         (Intr)
## Quarter -0.701
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -18.76089612  -0.43303691  -0.02775465   0.37555232  42.43848224
##
## Number of Observations: 108602
## Number of Groups: 4177
```

```
plot(fit0)
```



```
fit00 <- lme(Avg.Fare ~ Quarter + Avg.Pass, random = ~ Quarter | Route, data=rout,
             method = "ML")
plot(fit00)
```




```
anova(fit0, fit00) #Passenger fare is significant
```

```
##       Model df     AIC     BIC    logLik   Test  L.Ratio p-value
## fit0      1  6 1079268 1079326 -539628.1
## fit00     2  7 1077616 1077683 -538800.8 1 vs 2 1654.697  <.0001
```

```
fit00.int <- lme(Avg.Fare ~ Quarter + Avg.Pass + Quarter*Avg.Pass,
                 random = ~ Quarter | Route, data=rout,
             method = "ML")
```

```
summary(fit00.int) #Significant but nearly zero, we exclude for simplicity
```

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: rout
##       AIC     BIC  logLik
##   1077432 1077509 -538708
```

```
##
## Random effects:
##  Formula: ~Quarter | Route
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 82.660605 (Intr)
## Quarter      2.400596 -0.712
## Residual    30.217196
##
## Fixed effects: Avg.Fare ~ Quarter + Avg.Pass + Quarter * Avg.Pass
##                       Value Std.Error     DF   t-value p-value
## (Intercept)      221.79177 1.3340494 104422 166.25453       0
## Quarter           -0.72228 0.0417306 104422 -17.30818       0
## Avg.Pass          -0.05993 0.0016461 104422 -36.40541       0
## Quarter:Avg.Pass   0.00101 0.0000735 104422  13.72380       0
##  Correlation:
##                   (Intr) Quartr Avg.Ps
## Quarter           -0.709
## Avg.Pass          -0.247  0.264
## Quarter:Avg.Pass  0.191 -0.349 -0.777
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -18.91835421  -0.43793150  -0.02725791   0.38153408  42.82078371
##
## Number of Observations: 108602
## Number of Groups: 4177
```

```
fit1 <-lme(Avg.Fare ~ Quarter + Avg.Pass,
           random = ~ 1|Route, method="REML",
           data = rout)

fit1.log <-lme(log(Avg.Fare) ~ Quarter + Avg.Pass,
          random = ~ 1|Route, method="REML",
          data = rout)

summary(fit1); summary(fit1.log)
```

```
## Linear mixed-effects model fit by REML
##  Data: rout
##       AIC     BIC    logLik
##   1100692 1100740 -550340.8
##
## Random effects:
##  Formula: ~1 | Route
##         (Intercept) Residual
## StdDev:    64.25129 35.25245
##
## Fixed effects: Avg.Fare ~ Quarter + Avg.Pass
##                Value Std.Error     DF   t-value p-value
## (Intercept) 220.4409 1.0387180 104423 212.22398       0
## Quarter      -0.5174 0.0142712 104423 -36.25503       0
## Avg.Pass     -0.0530 0.0010192 104423 -51.99880       0
```

```
##  Correlation:
##          (Intr) Quartr
## Quarter  -0.179
## Avg.Pass -0.198 -0.034
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -6.38995569 -0.46994349 -0.01772498  0.42385536 50.00412574
##
## Number of Observations: 108602
## Number of Groups: 4177
## Linear mixed-effects model fit by REML
##  Data: rout
##         AIC       BIC  logLik
##   -92105.81 -92057.83 46057.9
##
## Random effects:
##  Formula: ~1 | Route
##         (Intercept)  Residual
## StdDev:   0.3326033 0.1439553
##
## Fixed effects: log(Avg.Fare) ~ Quarter + Avg.Pass
##                 Value   Std.Error     DF   t-value p-value
## (Intercept)  5.323908 0.005297442 104423 1004.9960       0
## Quarter     -0.000996 0.000058280 104423  -17.0942       0
## Avg.Pass    -0.000313 0.000004358 104423  -71.7455       0
##  Correlation:
##          (Intr) Quartr
## Quarter  -0.142
## Avg.Pass -0.166 -0.036
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -6.86553402 -0.54541787  0.03247829  0.57325132 15.39398078
##
## Number of Observations: 108602
## Number of Groups: 4177
```

```
plot(fit1) #Simple linear model is probably not good enough
```