(2) Comparison of 4 models:
 (i) change-score model, (ii) baseline-as-covariate model, (iii) hybrid model, (iv) a longitudinal model.

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -19.54       3.43  -5.697 8.43e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 282.433)
##
##     Null deviance: 6496  on 23  degrees of freedom
## Residual deviance: 6496  on 23  degrees of freedom
## AIC: 206.53
```

(i) change-score model just accounts for the intercept (mean of the change). Thus, this model not fits the data well and have the highest AIC in all 3 linear models. This model is too simple.

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.15761   16.53937   2.247   0.035 *
## before       0.69807    0.08679   8.044 5.39e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 190.4789)
##
##     Null deviance: 16514.5  on 23  degrees of freedom
## Residual deviance:  4190.5  on 22  degrees of freedom
## AIC: 198.01
```

(ii) baseline-as-covariate model is actually the same with the (iii) hybrid model. The difference of the coefficients of "before" is 1, which is just the move of "1 before" from the left side to the right side of the equation. The residual deviance from these two models are the same, and both less than the change-score model.

```
## (Intercept) 37.15761   16.53937   2.247  0.03503 *
## before      -0.30193    0.08679  -3.479  0.00213 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 190.4789)
##
##     Null deviance: 6496.0  on 23  degrees of freedom
## Residual deviance: 4190.5  on 22  degrees of freedom
## AIC: 198.01
```

This means these 2 models include more variance. Thus they have the same AIC, lower than the first model. Both of models show the significant association between two time points. These models are reasonable and easy to interpret. But they answer slightly different hypotheses.

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: cholesterol ~ time + (1 | subject_id)
##    Data: cholesterol_l
##
## REML criterion at convergence: 421.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.44161 -0.42520 -0.01153  0.41158  1.53751
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  subject_id (Intercept) 767.6     27.71
##  Residual               141.2     11.88
## Number of obs: 48, groups:  subject_id, 24
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  168.250      6.154  27.342
## timebefore    19.542      3.430   5.697
##
```

(iv) the linear mixed model with random intercept. The pre-post dataset has 2 repeat measurements, only enough for a compound symmetry covariance structure, which is the random intercept model. This model includes the "time" (before, after) as the binary fixed effect. This model also shows the significant association between cholesterol levels and time points. This model has a better fit to the data points and using the long form of dataset. The hypotheses is to test the association between cholesterol levels and time,

different from the change & baseline. However, since there are only 2 time points, this model is not necessary and too complex.
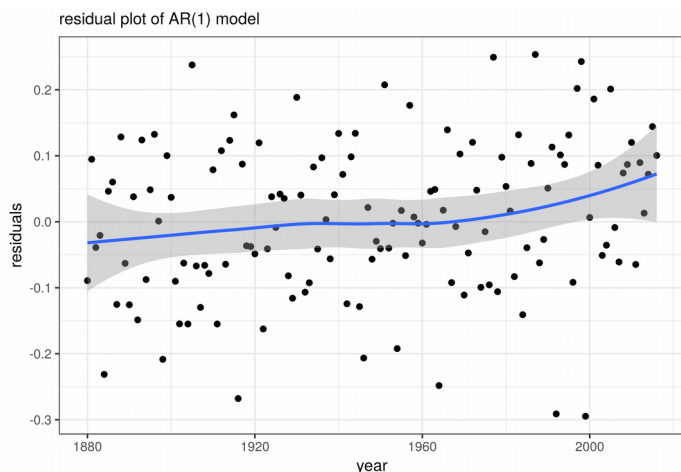
```
paste("The AIC of lmm:", AIC(lmm_prepost))

## [1] "The AIC of lmm: 429.516076427762"
```

(3) Time series data.

residual plot of linear model
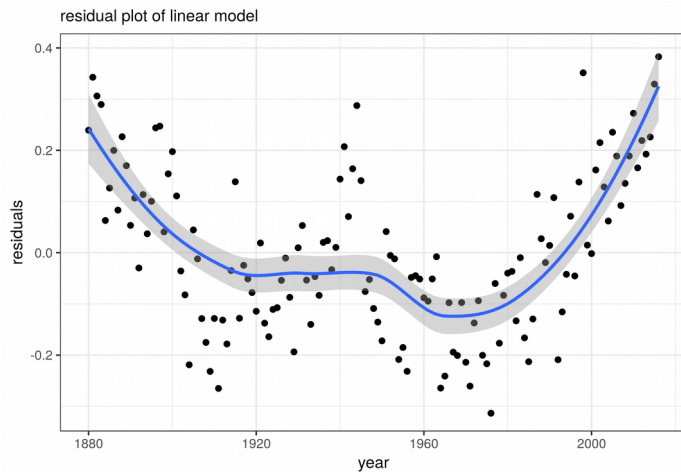
a. the residual plot shows the simple linear regression not works in this case. The residuals are not normally distributed, with a "W" shape pattern. We need try a different model to get an even and balanced residual cloud.

residual plot of AR(1) model

b. and c. The estimated phi in AR(1) process is 0.9432, there is also the correlation between two neighbors. This model fits the data better compared with the simple linear model. The residual is roughly equally distributed. The complexity of the mean by AR(1) model reduces the error of the model.

```
coef(ar_temp)

##      ar1 intercept
## 0.9431780 0.1087432

global_temp$res_ar <- residuals(ar_temp)
```

d. the average increase in temperature per decade is 0.06 ºC.

```
##   (Intercept)         year
## -11.650245508   0.006004107
# average change per decade
paste("average increase in temp. per decade:", round(10 * coef(lm_pred)[2],
    4))

## [1] "average increase in temp. per decade: 0.06"
```