for-loop hints

1. Use `dplyr` to `filter()` the data by the `year` column.

2. Use `lm()` with the formula that starts with `dengue_rate ~ province + ...` and includes all of the the variables. Don't forget to specify your `data`!

3. Use `summary()` on `dengue_fit`. The strongest predictor is the covariate with the largest `z value`.

4. `predict(dengue_fit, test_data, se.fit=T)`

5. Here's the full answer, read through it carefully to see if you understand what is going on:

```r
## create an empty vector to save the results in
prediction_covered <- c()
## the for-loop goes across all values of dengue_preds
for(i in 1:length(dengue_preds)){
    ## rnorm() takes n samples from a normal distribution with a mean and sd
    pred_vector <- rnorm(n = 1000,
                         mean = dengue_preds$fit[i],
                         sd = dengue_preds$se.fit[i])
    ## quantile() finds the probs quantile of the pred_vector
    pred_lb <- quantile(pred_vector, probs = .025) ## 95% lower bound
    pred_ub <- quantile(pred_vector, probs = .975) ## 95% upper bound
    ## if the observed value is between the lower and upper bound then TRUE
    ## otherwise, FALSE
    ## assigned to the i-th value of prediction_covered
    prediction_covered[i] <- test_data$dengue_rate[i] >= pred_lb &
        test_data$dengue_rate[i] <= pred_ub
}
```

Another way of doing this would be to use the `dengue_preds$fit` and `dengue_preds$se.fit` like we do in stats class where the 95% interval is equal to:

$$\widehat{y}\pm$$