This section continues our work with `auto.csv` from the first section, using the matrix operations we learned in the second.

The use of canned routines is not permitted for most of this class; you'll have to write the econometric routines from first principles. First, create matrices of the data, since we will be working mainly with matrix operations. Let **y** be the dependent variable, price, and let **X** be a matrix of the other car characteristics, along with a column of ones prepended. The `cbind()` function binds the columns horizontally and coerces the `matrix` class.

```
y <- matrix(data$price)
X <- cbind(1, data$mpg, data$weight)
head(X)
```

```
     [,1] [,2] [,3]
[1,]    1   22 2930
[2,]    1   17 3350
[3,]    1   22 2640
[4,]    1   20 3250
[5,]    1   15 4080
[6,]    1   18 3670
```

Note that `R` is doing something clever here. We only passed it a single `1` variable, and yet it has created an entire column vector full of ones! This is because `cbind()` automatically ensures that every column is as long as the longest column in the matrix, recycling data as needed.

Just to make sure that our matrices will be conformable when we regress `y` on `X`, check that the number of observations are the same in both variables.

```
dim(X)[1] == nrow(y)
```

```
[1] TRUE
```

Using the matrix operations described in the previous section, we can quickly estimate the ordinary least squared parameter vector.

```
beta <- solve(t(X) %*% X) %*% t(X) %*% y
beta
```

```
            [,1]
[1,] 1946.068668
[2,]  -49.512221
[3,]    1.746559
```

That's it! And although you're not allowed to use it in your problem sets, `lm()` is a nice tool for checking our results.

```
lm(y ~ X - 1)
```

```
Call:
lm(formula = y ~ X - 1)

Coefficients:
      X1          X2          X3
1946.069    -49.512      1.747
```

They match! Thank goodness. If you're interested in knowing what `lm()` does, I highly recommend reading through relevant `R` documentation[1].

Digging deeper into the numbers, consider the projection matrix $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and the residual maker matrix $\mathbf{M} = \mathbf{I}_n - \mathbf{P}$

```
n <- nrow(y)
P <- X %*% solve(t(X) %*% X) %*% t(X)
M <- diag(n) - P
```

`R` is useful for checking the properties of these matrices, including whether $\mathbf{M}$ is symmetric, that is, whether $\mathbf{M} = \mathbf{M}'$. The function `all.equal()` does not test **exact** equality, but instead whether the supplied objects are "close enough" to be considered the same. The problem is the limits of machine precision, and rounding at the tail ends of floating point numbers.

```
all.equal(M, t(M))
```

```
[1] TRUE
```

If we want to test for exact equality, we set the tolerance to zero, and the function will return a message with the mean relative difference between elements — which is clearly very close to zero.

```
all.equal(M, t(M), tol=0)
```

```
[1] "Mean relative difference: 7.266263e-15"
```

The residual maker matrix should also be idempotent, or $\mathbf{M} = \mathbf{M}\mathbf{M}'$.

```
all.equal(M, M %*% t(M))
```

```
[1] TRUE
```

Finally, we can examine the different components of the variation in the dependent variable, as they relate to the OLS estimate. Specifically, we can show that the total sum of square is equal to the sum of the residual and estimated sum of squares:

$$\mathbf{y}'\mathbf{y} = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \mathbf{e}'\mathbf{e} \tag{1}$$

First, define the relevant variables:

---

[1]Remember, you can do this using `?lm`.

```
e <- M %*% y
y.hat <- P %*% y
rss <- t(e) %*% e
ess <- t(y.hat) %*% y.hat
tss <- t(y) %*% y
```

Then check the condition in Eq. (**??**):

```
all.equal(tss, ess + rss)
```

```
[1] TRUE
```

Cool.

Additional puzzles

1. Write a function `wt.coef()` that will return the OLS coefficient on weight from the regression of car price on the covariate matrix described above.

2. Adjust the function to return a list of coefficients from the same linear regression, appropriately named.

3. Find the estimate of the covariance matrix $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ and show that the residuals and covariate matrix are orthogonal.