# Practical Markdown Use

*Chris Mellinger*

*1/25/2018*

## Contents



**See this document here:** https://github.com/Cmell/BBMarkdownTalkSpring2018

# R Markdown as a Text Editor and Processor

R Markdown is an excellent text processing tool. It can *handle* **many** formatting options. It does inline equations, $\pi = 3.14$, with support for many special characters. It will do equations set off from the text too:

$$e = (1 + \frac{1}{n})^n$$

You can include links to great resources, like the official R Markdown cheat sheet, or this fantastic example page with explanations: http://kbroman.org/knitr_knutshell/pages/Rmarkdown.html. Other stuff it can do:

- Bulleted lists, like this one.
- Footnotes[1]
- Images with captions (see the resources above for examples).
- Code expressed inline, `mean(x)`, and this can be evaluated inline as well (more on this soon).
- ~~strikethrough text~~
- Mult-tiered lists
  - list tier 2
    1) Numbered lists
    2) Continued
    3) That will autonumber no matter what you put on the left.
- Tables (there are multiple ways to do this, more on this soon):

| Left Justified | Right Justified | Centered |
|:---|---:|:---:|
| 23 | 45 | 56 |
| 56 | 87 | 42 |

[1] Here is the footnote.

# Combining Analysis and Presentation

R Markdown can run code for you and assimilate the output. This is a great way to organize scripts (which are notoriously annoying to organize), document your code well, and generate a very readable analysis document for sharing with others.

Because inline code can also be evaluated, it is possible to write dynamic documents. An example of a results section that would update dynamically with changes to the analysis is given below.

## Example Analysis

Let's get the packages we will need for this analysis.

```r
library(CMUtils)
library(ResultsHelper)
library(Hmisc)
```

```
## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```r
library(knitr)
library(papaja)
```

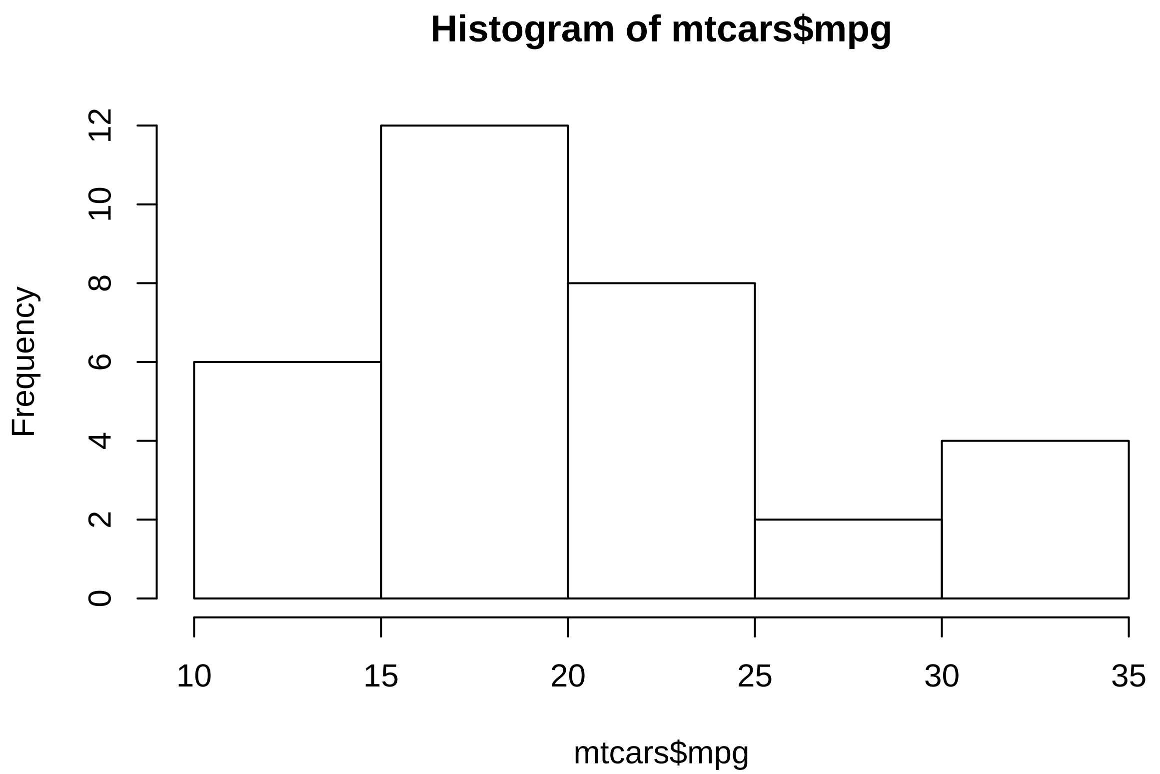
### Descriptives

```r
summary(mtcars)
```

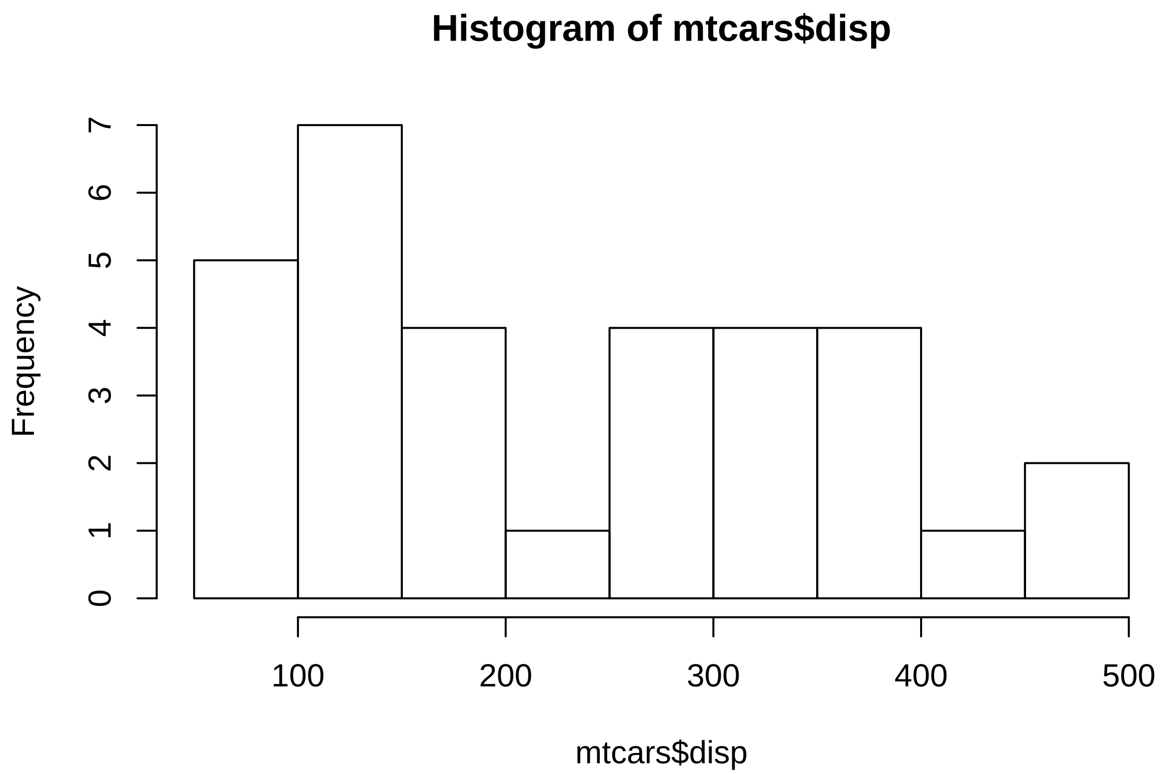
```
##       mpg             cyl             disp             hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##       drat             wt             qsec             vs
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
##        am              gear            carb
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
```

```
##  Median :0.0000   Median :4.000   Median :2.000
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

```r
hist(mtcars$mpg)
```




```r
hist(mtcars$disp)
```




```
mtcars$trans <- ifelse(mtcars$am, 'auto', 'manual')

kable(table(mtcars$trans), col.names = c('Transmission Type', 'Count'))
```

| Transmission Type | Count |
|---|---|
| auto | 13 |
| manual | 19 |

### Engine Size and Efficiency

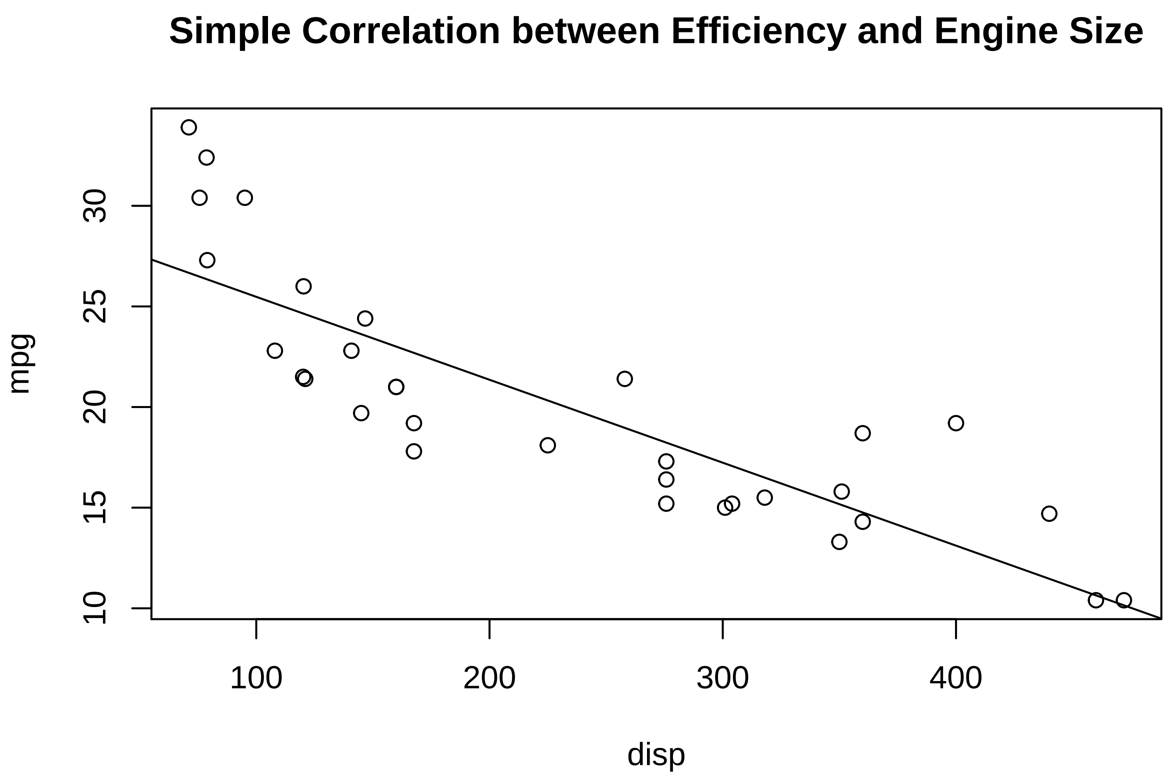
**Question:** What is the relationship between a car's engine size and its efficiency?

First pass with a simple regression:

```
smSum <- lmSummary(simpleMod <- lm(mpg ~ disp, data=mtcars))
```

Looks significant. Let's plot it.

```
with(mtcars, {
  plot(disp, mpg, main="Simple Correlation between Efficiency and Engine Size")
  abline(coef(simpleMod)['(Intercept)'], coef(simpleMod)['disp'])
})
```




What about a different version, controlling for the car's weight?

```
lmSummary(mod1 <- lm(mpg ~ disp + wt, data=mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ disp + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4087 -2.3243 -0.7683  1.7721  6.3484
##
## Coefficients:
##              Estimate Std. Error   t value   f value   R^2 Pr(>|t|)
## (Intercept)  34.96055    2.16454  16.15150 260.87087 0.900 4.91e-16 ***
## disp         -0.01773    0.00919  -1.92861   3.71953 0.114  0.06362 .
## wt           -3.35082    1.16413  -2.87840   8.28518 0.222  0.00743 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.917 on 29 degrees of freedom
## Multiple R-squared:  0.7809, Adjusted R-squared:  0.7658
## F-statistic: 51.69 on 2 and 29 DF,  p-value: 2.744e-10
```

Still marginal, but not as robust as before. It turns out that the car's weight matters too.

#### Transmission and Efficiency

**Question:** Do cars with manual transmissions have better efficiency than cars with automatic transmissions? Does this effect interact with engine size?

The am variable in mtcars is dummy coded 0 for automatic transmissions and 1 for manual transmissions. Let's get some other codes.

```
mtcars <- within(mtcars, {
  amC <- .5 * (am == 1) - .5 * (am == 0)
  amA <- am
  amM <- abs(am - 1)

  # Interactions with displacement
  dispC <- disp - mean(disp)
  amCxDispC <- amC * dispC

  # Simples:
  amAxDispC <- amA * dispC
  amMxDispC <- amM * dispC
})
```

The t-test:

```
lmSummary(modAm <- lm(mpg ~ amC, data=mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ amC, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error  t value  f value   R^2 Pr(>|t|)
## (Intercept)  20.7698     0.8822  23.5429 554.2701 0.949  < 2e-16 ***
## amC           7.2449     1.7644   4.1061  16.8603 0.360 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

The primary interaction model:

```
lmSummary(modAmxDisp <- lm(mpg ~ amC + dispC + amCxDispC, data=mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ amC + dispC + amCxDispC, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6056 -2.1022 -0.8681  2.2894  5.2315
##
## Coefficients:
##               Estimate Std. Error    t value    f value   R^2 Pr(>|t|)
## (Intercept)  19.018804   0.695754  27.335512 747.230193 0.964  < 2e-16 ***
## amC           0.451758   1.391509   0.324653   0.105400 0.004   0.7479
## dispC        -0.043311   0.005729  -7.560373  57.159239 0.671  3.1e-08 ***
```

```
## amCxDispC    -0.031455   0.011457  -2.745378   7.537101 0.212   0.0104 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.907 on 28 degrees of freedom
## Multiple R-squared:  0.7899, Adjusted R-squared:  0.7674
## F-statistic: 35.09 on 3 and 28 DF,  p-value: 1.27e-09
```

In this model, the main effect of displacement indicates that cars with larger engines have lower gas mileage. The interaction suggests that this effect is more negative for manual cars. Let's explore the simples.

```
lmSummary(modAmMxDisp <- lm(mpg ~ amM + dispC + amMxDispC, data=mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ amM + dispC + amMxDispC, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6056 -2.1022 -0.8681  2.2894  5.2315
##
## Coefficients:
##               Estimate Std. Error    t value    f value   R^2 Pr(>|t|)
## (Intercept)  19.244683   1.163583  16.539151 273.543519 0.907 5.54e-16 ***
## amM          -0.451758   1.391509  -0.324653   0.105400 0.004   0.7479
## dispC        -0.059038   0.009623  -6.135338  37.642369 0.573 1.27e-06 ***
## amMxDispC     0.031455   0.011457   2.745378   7.537101 0.212   0.0104 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.907 on 28 degrees of freedom
## Multiple R-squared:  0.7899, Adjusted R-squared:  0.7674
## F-statistic: 35.09 on 3 and 28 DF,  p-value: 1.27e-09
```

```
lmSummary(modAmAxDisp <- lm(mpg ~ amA + dispC + amAxDispC, data=mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ amA + dispC + amAxDispC, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6056 -2.1022 -0.8681  2.2894  5.2315
##
## Coefficients:
##               Estimate Std. Error    t value    f value   R^2 Pr(>|t|)
## (Intercept)  18.792925   0.763132  24.626046 606.442125 0.956  < 2e-16 ***
## amA           0.451758   1.391509   0.324653   0.105400 0.004  0.74786
## dispC        -0.027584   0.006219  -4.435410  19.672863 0.413  0.00013 ***
## amAxDispC    -0.031455   0.011457  -2.745378   7.537101 0.212  0.01044 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.907 on 28 degrees of freedom
## Multiple R-squared:  0.7899, Adjusted R-squared:  0.7674
## F-statistic: 35.09 on 3 and 28 DF,  p-value: 1.27e-09
```

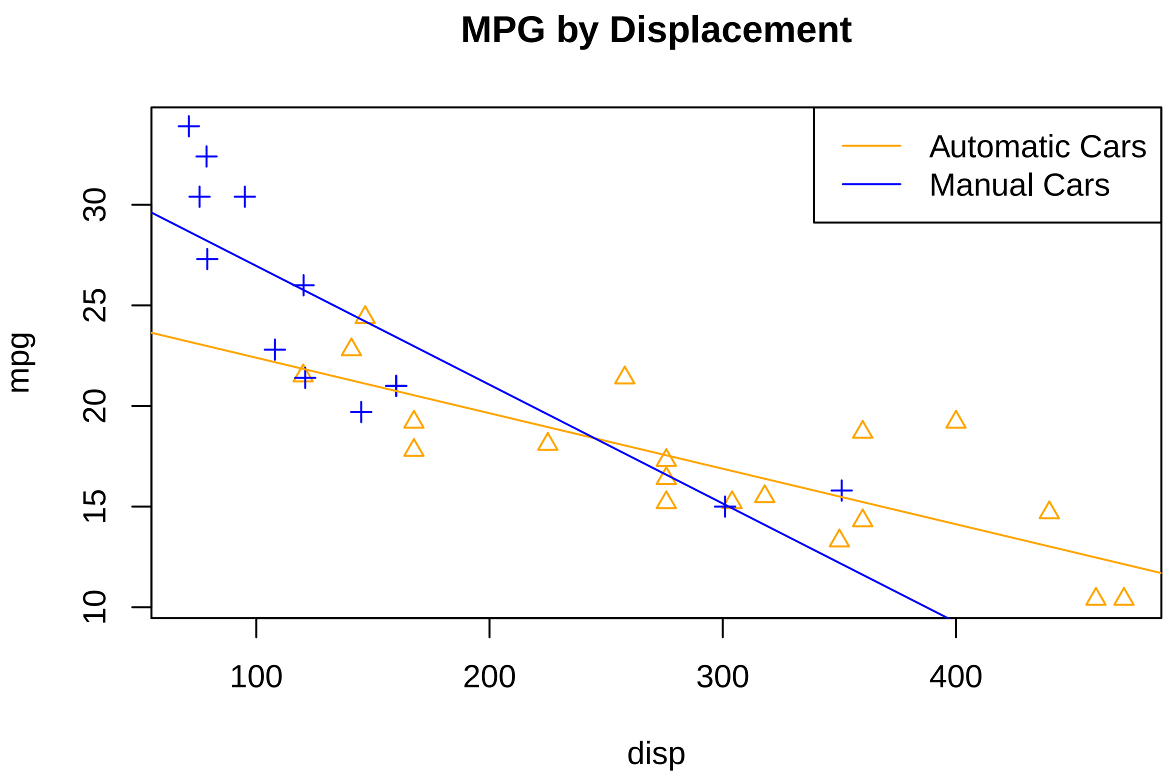
Let's plot the difference for manual and automatic cars.

```r
with(mtcars, {
  # Set up the plot, but specify type is none so that I can add points manually
  plot(disp, mpg, type='n', main="MPG by Displacement")

  # Add points
  points(disp[am==0], mpg[am==0], pch=2, col='orange')
  points(disp[am==1], mpg[am==1], pch=3, col='blue')

  # Add lines
  abline(lm(mpg[!amA] ~ disp[!amA]), col='orange') # Automatic
  abline(lm(mpg[!amM] ~ disp[!amM]), col='blue') # Manual

  # Add a legend
  legend(
    'topright',
    c('Automatic Cars', 'Manual Cars'),
    col = c('orange', 'blue'),
    lty=1
  )
})
```




### Example Results Section

In the first step of the analysis, efficiency (in MPG) was regressed on engine displacement. Engine size sigificantly predicted efficiency, $b = -0.04$, $t(30) = -8.75$.

In a second analysis, engine efficiency was regressed on displacement, but car weight was controlled for. Displacement marginally predicted efficiency, $F(1, 29) = 3.72$, $p = 0.06$, as well as car weight, $b = -0.04$, 95%

CI $[-0.05, -0.03]$, $t(30) = -8.75$, $p < .001$.

### Notes on the Results Section

Three statistical results are presented in the previous section. The first uses base R commands, and I do not recommend doing things this way. To see why, here is the code for that result:

```
In the first step of the analysis, efficiency (in MPG) was regressed on engine
displacement. Engine size sigificantly predicted efficiency,
$b=`r round(coef(simpleMod)['disp'], 2)`$,
$t(`r simpleMod$df.residual`) = `r round(smSum$coefficients['disp', 't value'], 2)`$.
```

It is tremendously flexible, but convoluted to read and difficult to write. In addition, many of the operations, such as rounding, are repeated for every single value.

A much cleaner, easier, and quicker way to acheive a dynamic document is to use functions from packages that do the dirty work. One option is `ResultsHelper` (still under heavy development). Here is the code for the second result:

```
`r fpStr('disp', mod1)`
```

While this is extremely simple to write in code, it does not include all of the desired stats (such as confidence intervals). I hope that soon this will be a flexible package that has the power needed for custom, yet easy to create, dynamic documents.

The third result was created with `papaja`, which is a good package to explore for writing APA style documents. Here is the code:

```
`r apa_print(simpleMod)$full_result$disp`
```

It includes the things we would need for a results section meeting a large range of journal requirements. It also has the advantage of taking a wide range of statistical test objects from R, such as the result from the `anova()` function.

Combining this with the text from the Rmd document, we get:

```
In a second analysis, engine efficiency was regressed on displacement, but car
weight was controlled for. Displacement marginally predicted efficiency,
`r fpStr('disp', mod1)`, as well as car weight, `r apa_print(simpleMod)$full_result$disp`.
```

## Resources

- R Markdown formatting and code use:
  - R Markdown reference
  - R Markdown Cheat Sheet
  - A more thorough introduction to R Markdown
  - An even more thorough introduction
- RStudio's Cheat Sheets:
  - All the cheat sheets
  - Cheat sheet to the RStudio interface
  - The LaTeXcheat sheet
  - Graphing; mostly `ggplot` if you're into that sort of thing
- `Knitr` and chunk options:
  - https://yihui.name/knitr/options/
- Math and other stuff in LaTeX:

  - https://en.wikibooks.org/wiki/LaTeX/Mathematics
  - https://artofproblemsolving.com/wiki/index.php/LaTeX:Symbols
- Some packages you may want to consider:
  - `tidyverse`: A collection of tools to make data operations in R easier and more beautiful to program.
  - `stargazer`: Tools for representing data and results in clean tables.
  - `papaja`: Tools for APA style markdown (showcased above).