STA 444 Practical

Michael Rivera

September 28, 2023

Instructions

You may use the book during the exam and may take until the end of the period.

- **Change the header information within the RMD to contain your own name.**
- **Answer all exercise prompts within the RMD. All code must be shown.**
- **Place answers into the blank R chunks given for each required response.**
- **Compile the RMD into a PDF when finished.**
- **Ensure all code is visible within the PDF.**
- **Submit the PDF through our Canvas portal.**

Exercise 1

Specifically load the packages `ggplot2`, `dplyr`, and `broom`. Be sure the code is displayed. It is okay if warnings or messages are output.

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(broom)
```

Exercise 2

Load the data set `mtcars` from base R. Display the first 6 rows of this data frame.

```
data("mtcars")
head(mtcars, n = 6L)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
```

```
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

Exercise 3

Create a new column within `mtcars` called `wt01` that takes on the value `Light Weight` if less than the mean weight (`wt`) of the data set, or takes on the value `Heavy Weight` if greater than or equal to the mean weight. *This question should use dplyr commands to achieve the result.*

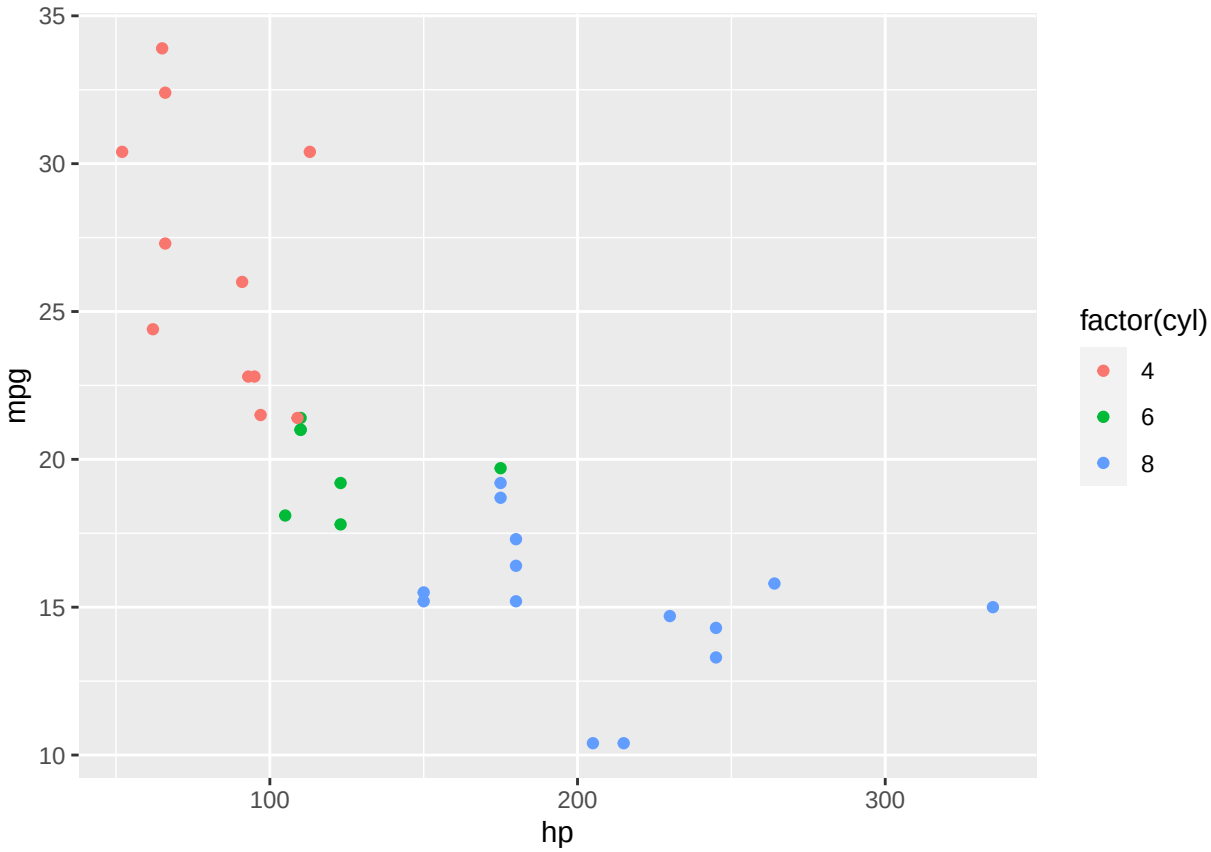
```
mtcars2 <- mtcars %>%
  mutate(
    wt01 = if_else(wt > (mean(c_across('wt'))
  ), 'Heavyweight', 'Lightweight'))

head(mtcars2)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
##                         wt01
## Mazda RX4         Lightweight
## Mazda RX4 Wag     Lightweight
## Datsun 710        Lightweight
## Hornet 4 Drive    Lightweight
## Hornet Sportabout Heavyweight
## Valiant           Heavyweight
```

Exercise 4

Using the `mtcars` data, create a scatter graph with `hp` on the x-axis and `mpg` on the y-axis. Color the graph using the `cyl`. To get unique groups add the command `factor(cyl)` rather than just the variable `cyl` when coloring your graph.

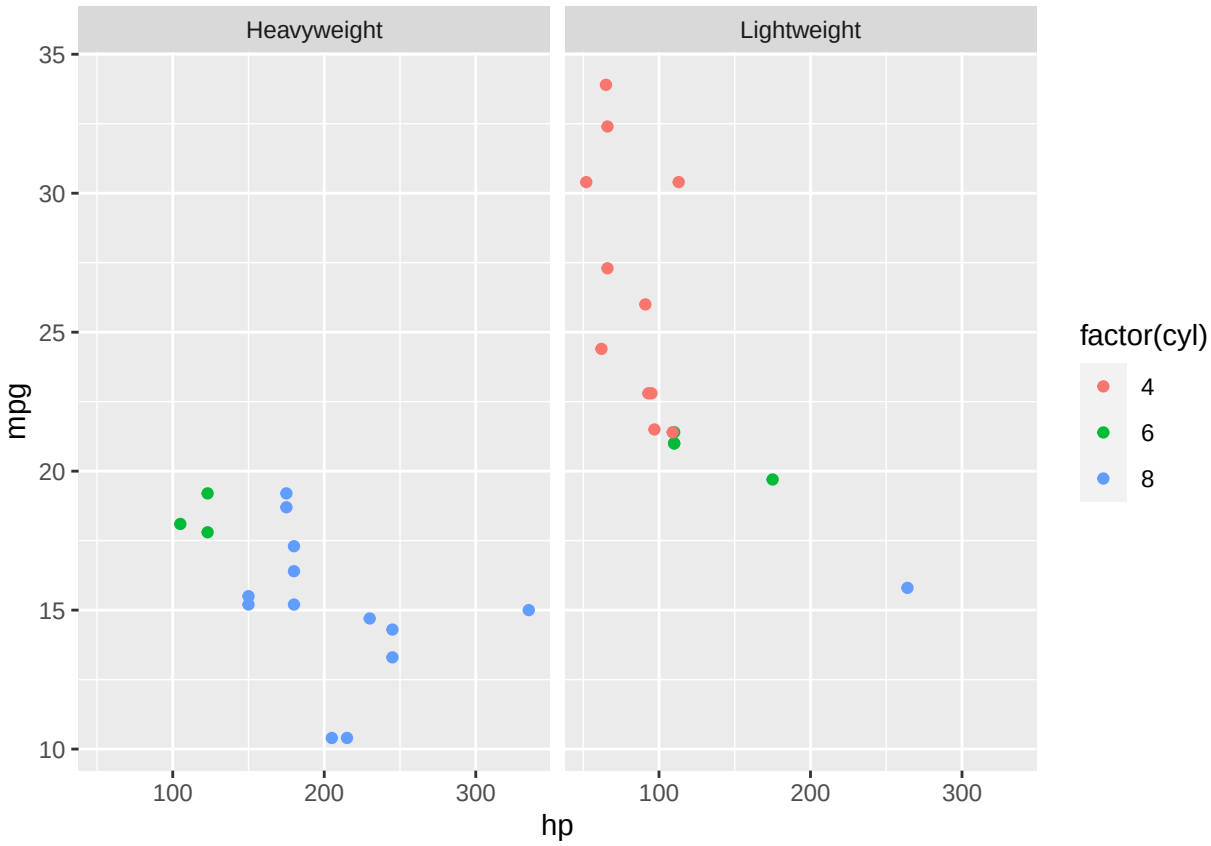
```
ggplot(data=mtcars2, aes(x=hp, y=mpg)) +
  geom_point(aes(color = factor(cyl)))
```



Exercise 5

Split your graph above into a two paneled graph by adding a facet for `wt01`.

```
ggplot(data=mtcars2, aes(x=hp, y=mpg)) +
  geom_point(aes(color = factor(cyl))) +
  facet_grid(cols = vars(wt01))
```



Exercise 6

Using the `mtcars` data, calculate the mean and standard deviation of `mpg` for each `cyl` and `wt01` combination. *Hint: `dplyr` commands.*

```
mtcars3 <- mtcars2 %>%
  group_by(cyl, wt01) %>%
  summarize (mean.mpg = mean(mpg),
             stddev.mpg = sd(mpg)
  )
```

```
## `summarise()` has grouped output by 'cyl'. You can override using the `.groups`
## argument.
```

Exercise 7

Estimate the model below using the `mtcars` data.

```
model.mtcars <- lm(mpg ~ hp*wt01)
```

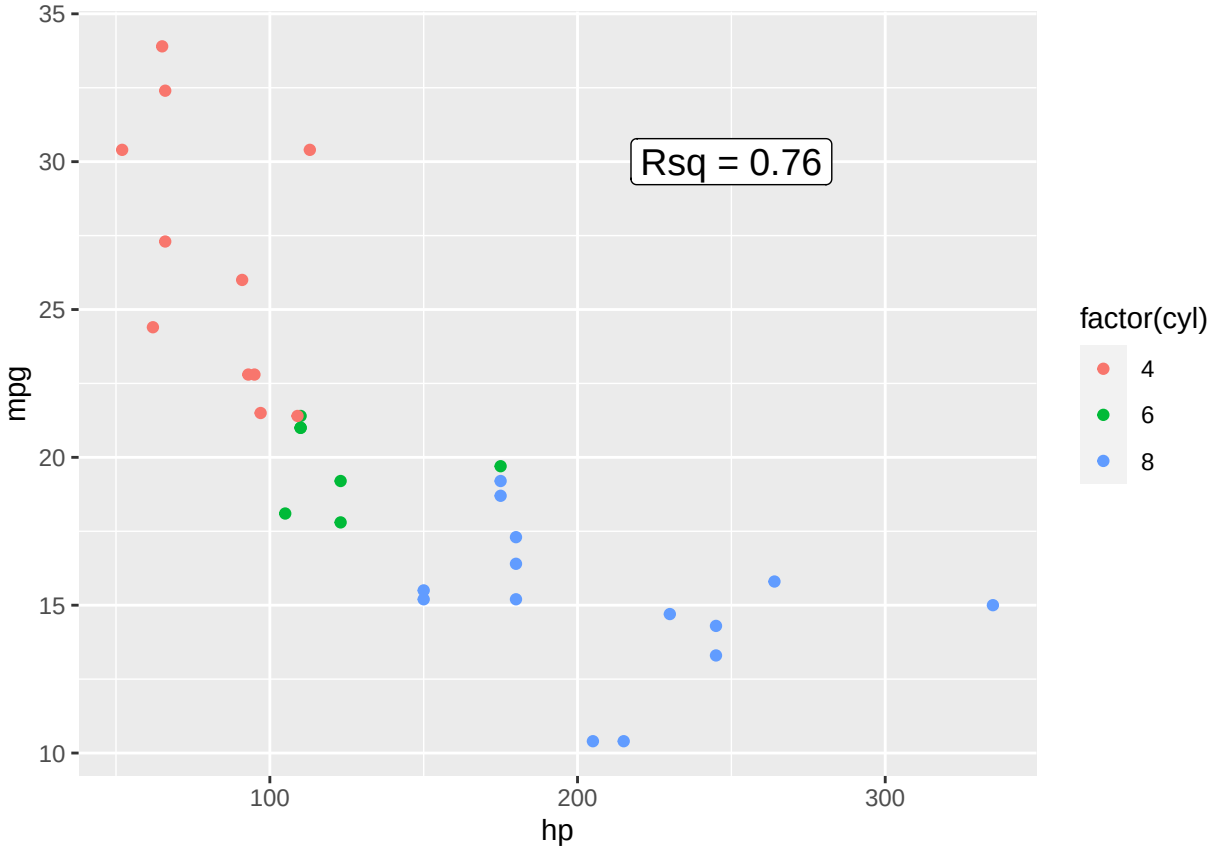
Display the resulting coefficient estimates using the `broom` package. Be sure to include the confidence intervals for each parameter.

```
model.mtcars2 <- lm(mpg ~ hp*wt01, data=mtcars2)
summary(model.mtcars2)
```

```
##
## Call:
## lm(formula = mpg ~ hp * wt01, data = mtcars2)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.8599 -2.5937  0.0661  1.9823  6.6097
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         20.33910    2.76513   7.356 5.21e-08 ***
## hp                  -0.02478    0.01407  -1.761  0.08912 .
## wt01Lightweight     11.47936    3.30937   3.469  0.00171 **
## hp:wt01Lightweight -0.04489    0.02104  -2.133  0.04179 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.134 on 28 degrees of freedom
## Multiple R-squared:  0.7557, Adjusted R-squared:  0.7295
## F-statistic: 28.87 on 3 and 28 DF,  p-value: 1.026e-08
```

```
mtcars4 <- mtcars2 %>%
  dplyr::select(-matches('fit'), -matches('lwr'), -matches('upr')) %>%
  cbind(predict(model.mtcars2, newdata=., interval='confidence'))

Rsq_string <-
  broom::glance(model.mtcars2) %>%
  select(r.squared) %>%
  mutate(r.squared = round(r.squared, digits = 2)) %>%
  mutate(r.squared = paste('Rsq =', r.squared)) %>%
  pull(r.squared)

ggplot(data=mtcars4, aes(x=hp, y=mpg)) +
  geom_point(aes(color = factor(cyl))) +
  annotate('label', label = Rsq_string, x=250, y=30, size = 5)
```



i tried :(# not sure how to make the confidence intervals show up.