Modelling

Stewart Li

9/16/2020

Correlation

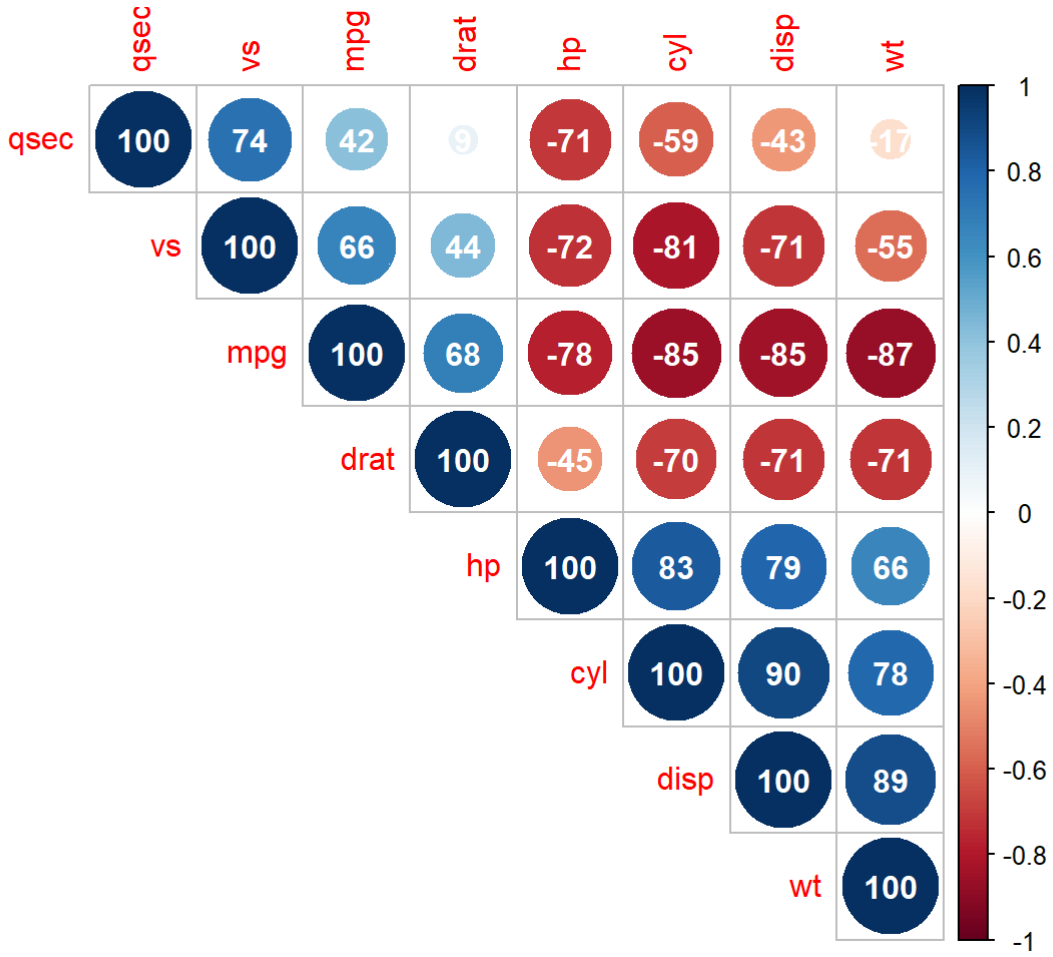
```
df <- mtcars %>%
  as_tibble(rownames = "car") %>%
  mutate(across(c(am, gear, carb), as.factor))
```

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 3.6.1
```

```
## corrplot 0.84 loaded
```

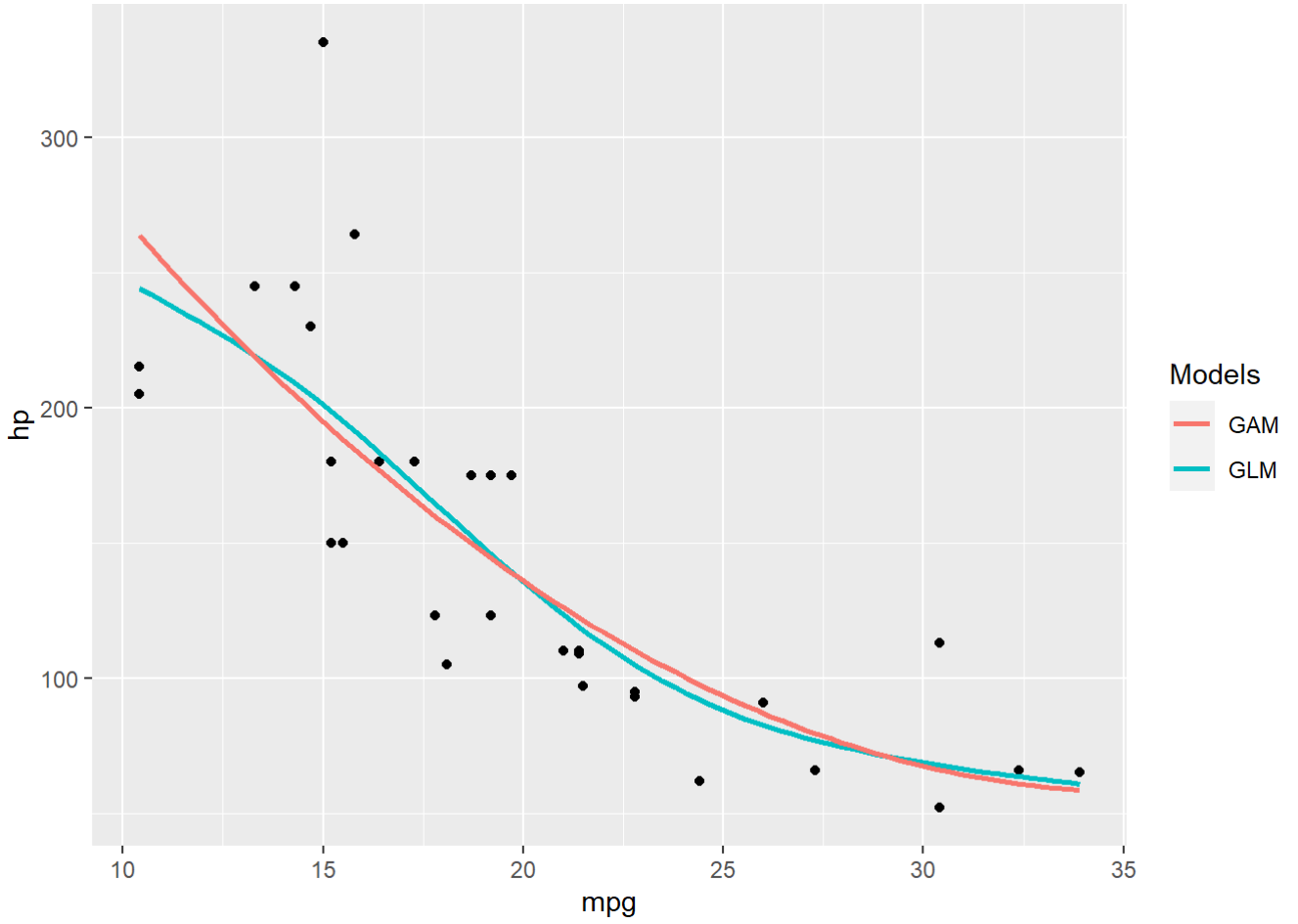
```
corrplot(df %>% select(where(is.numeric)) %>% cor(),
         addCoef.col = "white",
         addCoefasPercent = TRUE,
         type = "upper",
         order = "AOE")
```



Plot

```
df %>%
  ggplot(aes(mpg, hp)) +
  geom_point() +
  geom_smooth(aes(color = 'red'), method = "gam", se = FALSE) +
  geom_smooth(aes(color = 'blue'), method = "glm", formula = y ~ poly(x, 2), se = FALSE)
+
  scale_color_discrete(name = 'Models',
                       labels = c("GAM", "GLM"))
```

```
## `geom_smooth()` using formula 'y ~ s(x, bs = "cs")'
```



Models

```
lm(mpg ~ hp, data = df) %>%
  # str() # list
  names()
```

```
##  [1] "coefficients"  "residuals"     "effects"       "rank"
##  [5] "fitted.values" "assign"        "qr"            "df.residual"
##  [9] "xlevels"       "call"          "terms"         "model"
```

```
df %>%
  group_by(am, gear) %>%
  nest(data = -c(am, gear)) %>%
  mutate(mod = map(data, ~lm(mpg ~ hp + disp, data = .x)),
         mod_tidy = map(mod, broom::tidy)) %>%
  unnest(mod_tidy)
```

```
## # A tibble: 12 x 9
## # Groups:   am, gear [4]
##    am    gear  data        mod    term     estimate std.error statistic  p.value
##    <fct> <fct> <list>      <list> <chr>       <dbl>     <dbl>     <dbl>    <dbl>
##  1 1     4     <tibble [8~ <lm>   (Interc~  42.9      3.63       11.8    7.61e-5
##  2 1     4     <tibble [8~ <lm>   hp        -0.151    0.106      -1.42   2.14e-1
##  3 1     4     <tibble [8~ <lm>   disp      -0.0371   0.0691     -0.537  6.14e-1
##  4 0     3     <tibble [1~ <lm>   (Interc~  26.2      2.39       10.9    1.35e-7
##  5 0     3     <tibble [1~ <lm>   hp        -0.0315   0.0193     -1.63   1.28e-1
##  6 0     3     <tibble [1~ <lm>   disp      -0.0139   0.00969    -1.43   1.78e-1
##  7 0     4     <tibble [4~ <lm>   (Interc~  42.0      8.12        5.17   1.22e-1
##  8 0     4     <tibble [4~ <lm>   hp        -0.0650   0.0323     -2.01   2.93e-1
##  9 0     4     <tibble [4~ <lm>   disp      -0.0925   0.0670     -1.38   3.99e-1
## 10 1     5     <tibble [5~ <lm>   (Interc~  32.9      3.99        8.25   1.44e-2
## 11 1     5     <tibble [5~ <lm>   hp        -0.0364   0.0414     -0.878  4.73e-1
## 12 1     5     <tibble [5~ <lm>   disp      -0.0218   0.0369     -0.590  6.15e-1
```

Model effect

```
lm1 <- lm(mpg ~ hp + disp + am, data = df)
lm2 <- lm(mpg ~ hp + disp + am + gear, data = df)
lm3 <- lm(mpg ~ hp + disp + am + gear + carb, data = df)

# pander::pander(summary(lm1))
# library(stargazer)
```

```
library(ggeffects)
```

```
## Warning: package 'ggeffects' was built under R version 3.6.1
```

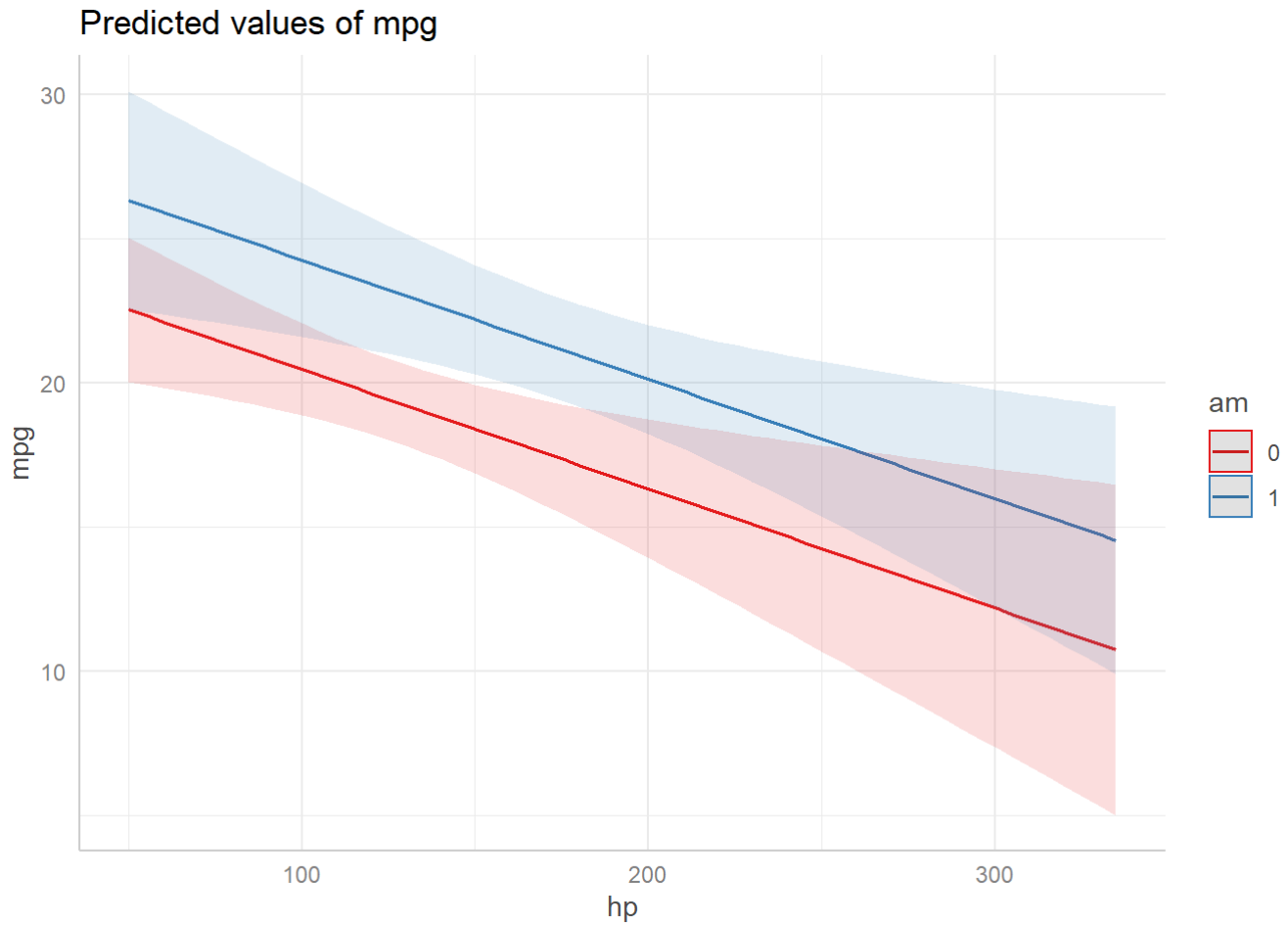
```
lm1_est <- ggpredict(lm1, terms = c("hp", 'am')) # disp hold its mean constant
lm1_est1 <- ggpredict(lm1, terms = c("hp[150, 250]", 'am'))
lm1_est2 <- ggpredict(lm1, terms = c("hp[meansd]", 'disp[minmax]'))

mean(mtcars$disp)
```

```
## [1] 230.7219
```

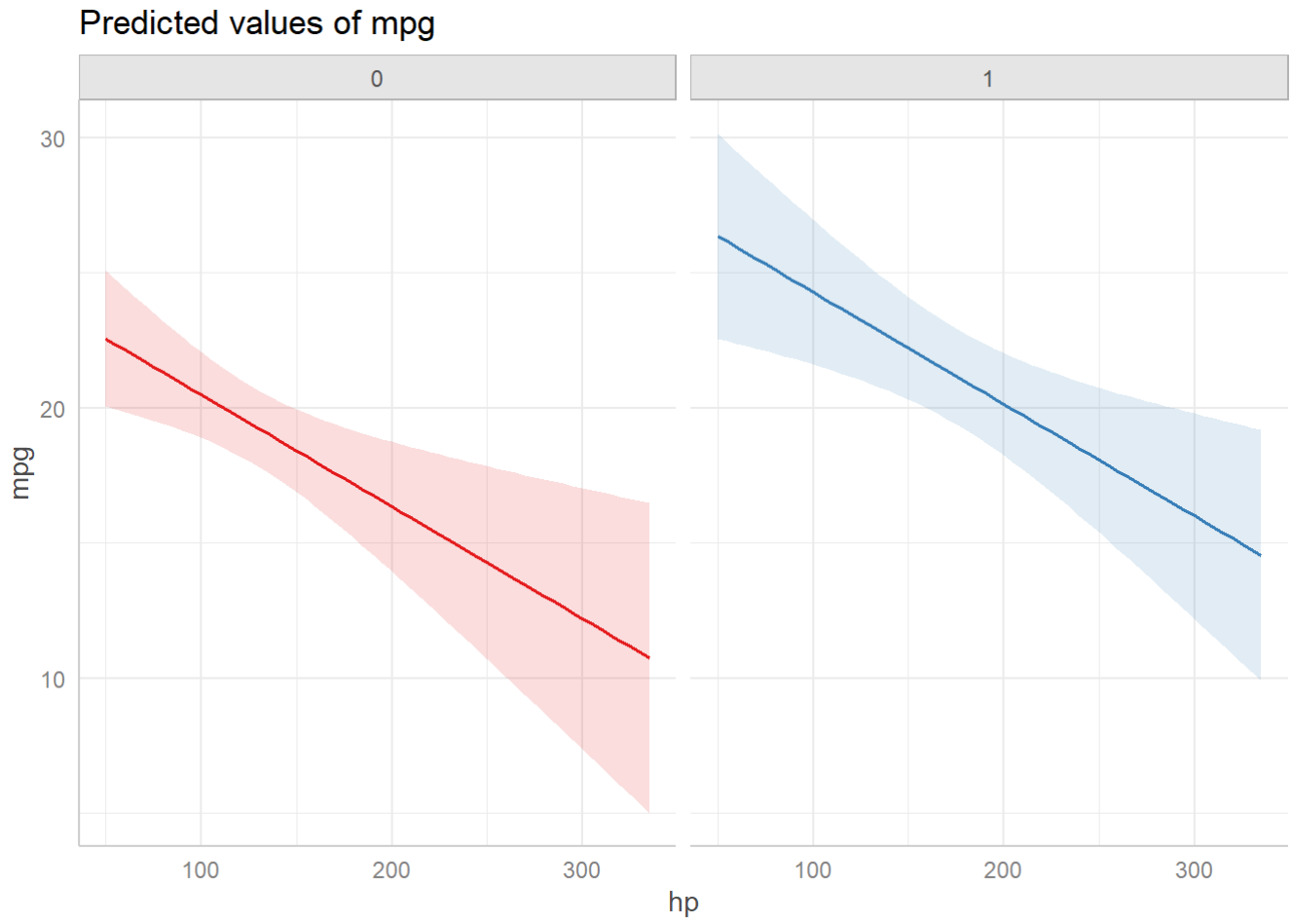
```
mean(mtcars$hp)
```

```
## [1] 146.6875
```

```
plot(lm1_est)
```




```
plot(lm1_est, facets = TRUE)
```



Model report

```
library(gt)
```

```
## Warning: package 'gt' was built under R version 3.6.3
```

```
df %>%
  group_by(gear, carb) %>%
  summarise(across(c(mpg, hp, disp), mean)) %>%
  gt(rowname_col = 'carb') %>%
  tab_header(title = "My report table",
             subtitle = "Grouped by gear and carb") %>%
  tab_spanner(label = "Measurement",
              columns = vars(hp, disp)) %>%
  cols_label(hp = 'HP',
             disp = 'DISP') %>%
  fmt_number(vars(mpg, hp,disp),
             decimals = 0) %>%
  cols_align(align = 'right',
             columns = vars(mpg, hp, disp)) %>%
  row_group_order(groups = c('5', '4', '3'))
```

```
## `summarise()` regrouping output by 'gear' (override with `.groups` argument)
```

My report table			
Grouped by gear and carb			
		Measurement	
	mpg	HP	DISP
5			
2	28	102	108
4	16	264	351
6	20	175	145
8	15	335	301
4			
1	29	72	84
2	25	80	121
4	20	116	164
3			
1	20	104	201
2	17	162	346
3	16	180	276
4	13	228	416

```
library(modelsummary)
```

```
## Warning: package 'modelsummary' was built under R version 3.6.3
```

```
mod_list <- list("Model 1" = lm1, "Model 2" = lm2, "Model 3" = lm3)
# msummary(lm1)
msummary(mod_list, output = "html") # latex
```

	Model 1	Model 2	Model 3

	Model 1	Model 2	Model 3
(Intercept)	27.866	29.125	26.224
	(1.620)	(2.302)	(3.284)
am1	3.796	3.724	2.845
	(1.424)	(1.809)	(1.911)
disp	-0.014	-0.014	-0.013
	(0.009)	(0.011)	(0.014)
hp	-0.041	-0.047	-0.020
	(0.014)	(0.016)	(0.028)
gear4		-1.072	1.694
		(1.996)	(2.863)
gear5		0.698	2.866
		(2.671)	(3.369)
carb2			-1.156
			(1.741)
carb3			-2.718
			(2.454)
carb4			-4.480
			(2.571)
carb6			-6.801
			(3.890)
carb8			-6.234
			(6.128)
R2	0.799	0.806	0.847
R2 Adj.	0.778	0.769	0.774
AIC	163.4	166.3	168.8

	Model 1	Model 2	Model 3
BIC	170.7	176.5	186.3
Log.Lik.	-76.690	-76.127	-72.380
F	37.149	21.626	11.591

```
msummary(mod_list, stars = TRUE) %>%
  kableExtra::row_spec(c(2, 4, 6), background = 'coral')
```

	Model 1	Model 2	Model 3
(Intercept)	27.866***	29.125***	26.224***
	(1.620)	(2.302)	(3.284)
am1	3.796**	3.724**	2.845
	(1.424)	(1.809)	(1.911)
disp	-0.014	-0.014	-0.013
	(0.009)	(0.011)	(0.014)
hp	-0.041***	-0.047***	-0.020
	(0.014)	(0.016)	(0.028)
gear4		-1.072	1.694
		(1.996)	(2.863)
gear5		0.698	2.866
		(2.671)	(3.369)
carb2			-1.156
			(1.741)
carb3			-2.718
			(2.454)
carb4			-4.480*
			(2.571)

* p < 0.1, ** p < 0.05, *** p < 0.01

	Model 1	Model 2	Model 3
carb6			-6.801*
			(3.890)
carb8			-6.234
			(6.128)
R2	0.799	0.806	0.847
R2 Adj.	0.778	0.769	0.774
AIC	163.4	166.3	168.8
BIC	170.7	176.5	186.3
Log.Lik.	-76.690	-76.127	-72.380
F	37.149	21.626	11.591

* p < 0.1, ** p < 0.05, *** p < 0.01

```
library(gtsummary)
```

```
## Warning: package 'gtsummary' was built under R version 3.6.2
```

```
df %>%
  select(-car) %>%
  tbl_summary(by = am) %>%
  as_gt() %>%
  tab_header(title = "My table summary")
```

```
## Warning: The `cells_data()` function is deprecated and will soon be removed
##  * Use the `cells_body()` function instead
```

My table summary

Characteristic[1]	**0**, N = 19	**1**, N = 13
mpg	17.3 (14.9, 19.2)	22.8 (21.0, 30.4)
cyl		
4	3 (16%)	8 (62%)

[1] Statistics presented: median (IQR); n (%)

My table summary		
6	4 (21%)	3 (23%)
8	12 (63%)	2 (15%)
disp	276 (196, 360)	120 (79, 160)
hp	175 (116, 192)	109 (66, 113)
drat	3.15 (3.07, 3.70)	4.08 (3.85, 4.22)
wt	3.52 (3.44, 3.84)	2.32 (1.94, 2.78)
qsec	17.82 (17.18, 19.17)	17.02 (16.46, 18.61)
vs	7 (37%)	7 (54%)
gear		
3	15 (79%)	0 (0%)
4	4 (21%)	8 (62%)
5	0 (0%)	5 (38%)
carb		
1	3 (16%)	4 (31%)
2	6 (32%)	4 (31%)
3	3 (16%)	0 (0%)
4	7 (37%)	3 (23%)
6	0 (0%)	1 (7.7%)
8	0 (0%)	1 (7.7%)

[1] Statistics presented: median (IQR); n (%)