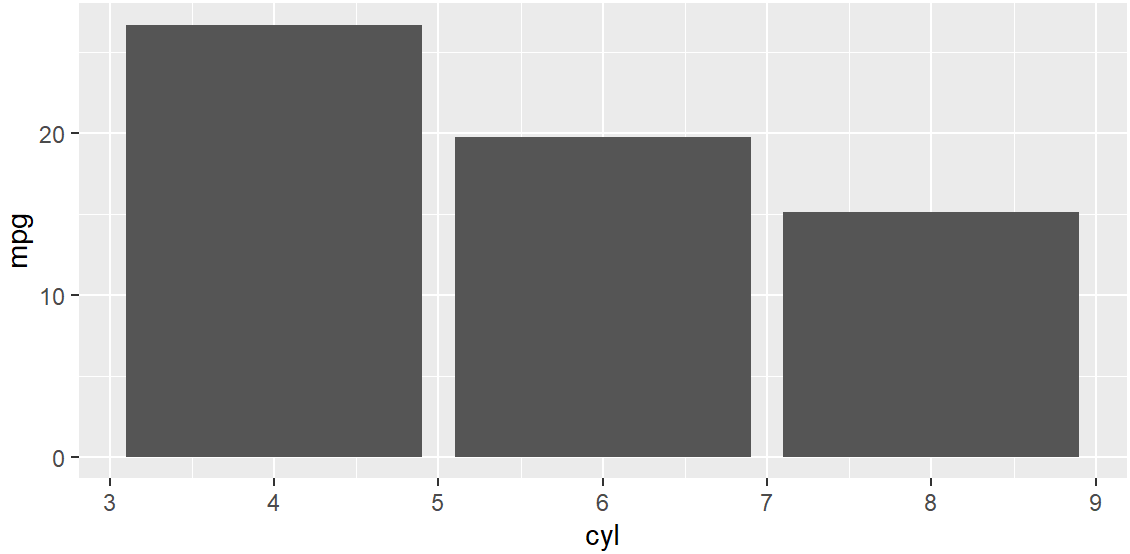
mtcars analysis

Visualize

Aggregate data in Spark, visualize in R.

```
library(ggplot2)
cars %>%
  group_by(cyl) %>% summarise(mpg = mean(mpg)) %>%
  ggplot(aes(cyl, mpg)) + geom_bar(stat="identity")
```



Model

The selected model was a simple linear regression that uses the weight as the predictor of MPG

```
cars %>%
  ml_linear_regression(wt ~ mpg) %>%
  summary()
```

```
## Deviance Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6516 -0.3490 -0.1381  0.3190  1.3684
##
## Coefficients:
## (Intercept)         mpg
##    6.047255   -0.140862
##
## R-Squared: 0.7528
## Root Mean Squared Error: 0.4788
```

rmarkdown::render(“new1.Rmd”)