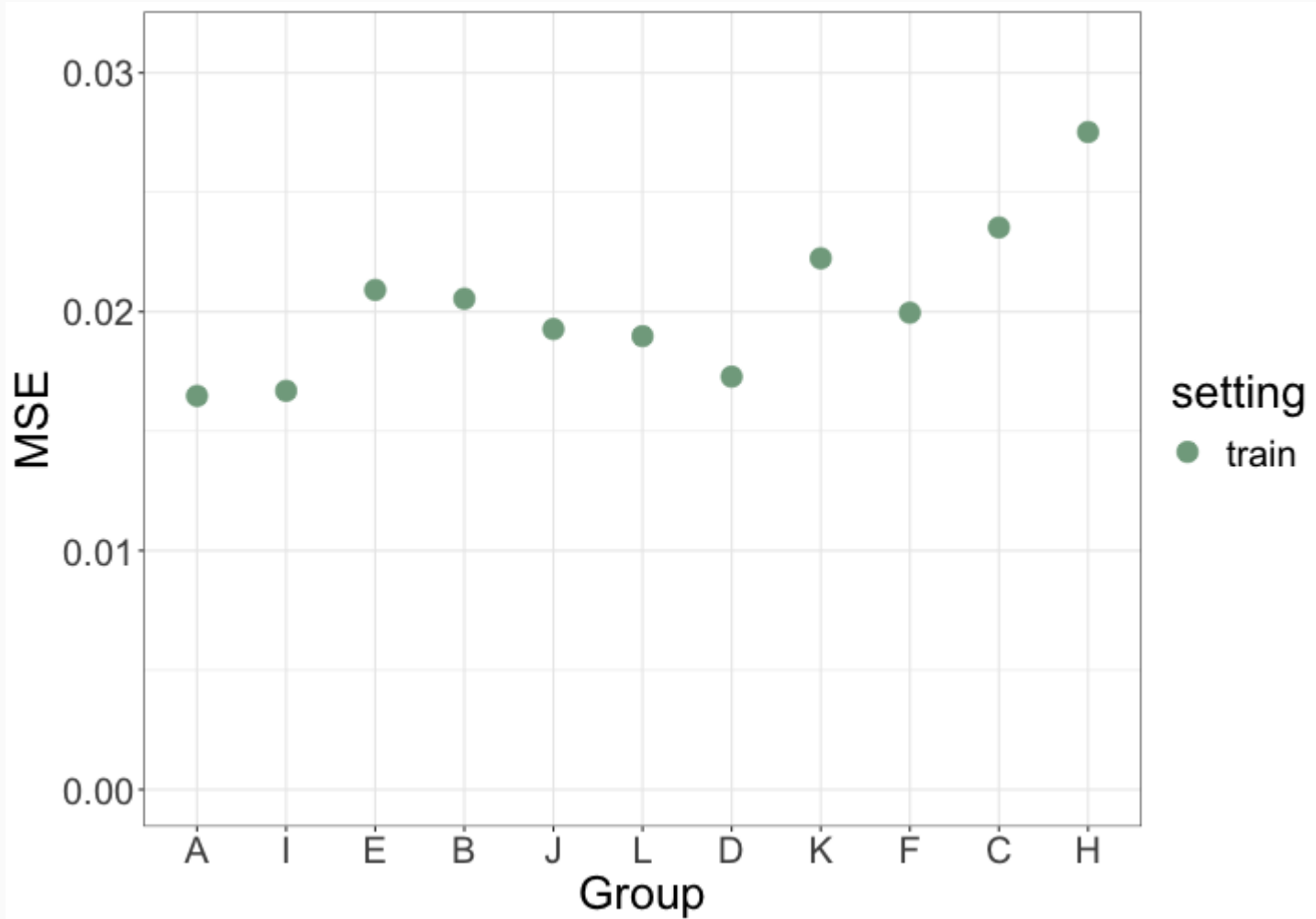# Regression Competition Results
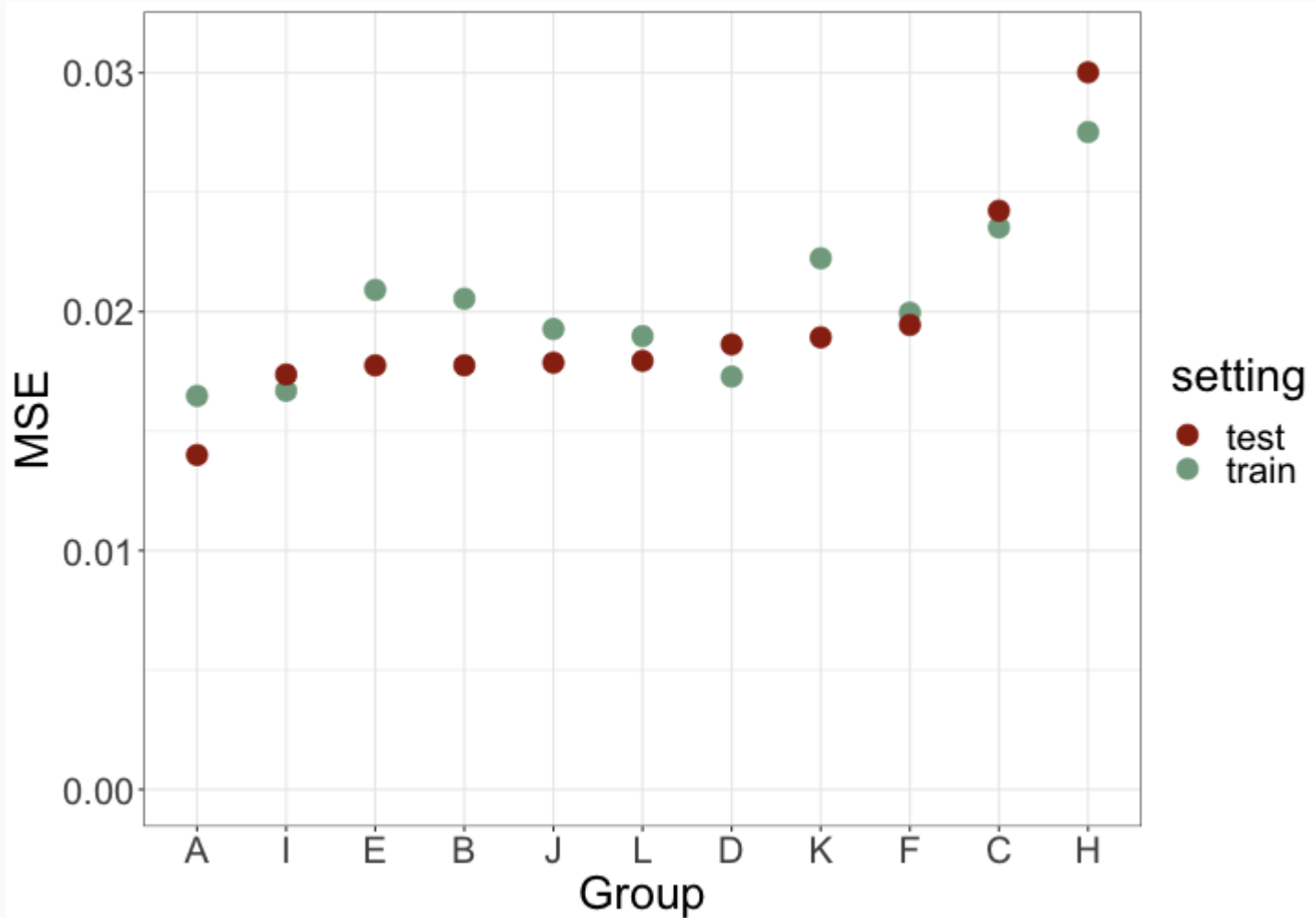
Scoring

- Coding errors were corrected

- For each group, calculate

  - train MSE
  - test MSE on full test data set
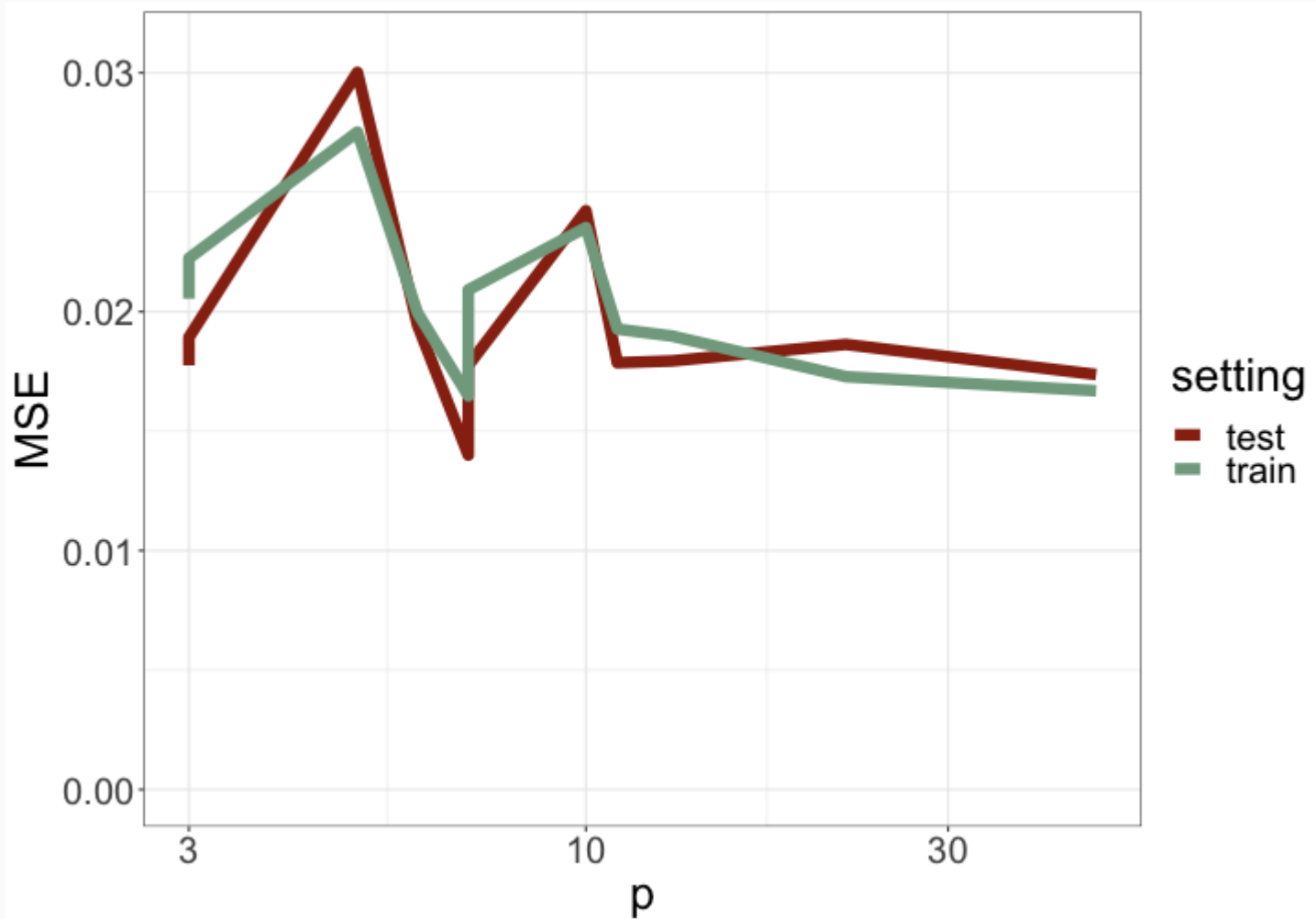
- Remember: sampling variability!

Training MSE

Training MSE + Testing MSE

Bias - Variance Tradeoff

Variable usage

In the top 4 models, the types of variables used were . . .

Twice

high school graduation rate, kids with two parents

Thrice

kids born to unmarried parents

Fource

race

Improving Our Code

1. Pass in what you need

```
myfun <- function(x, y) {
  lm(y ~ x, data = df)
}
```

vs.

```
myfun <- function(x, y, df) {
  lm(y ~ x, data = df)
}
```

2. Take advantage of lm()

```
myfun <- function(df) {
  lm(df$y ~ df$x)
}
```

vs.

```
myfun <- function(df) {
  lm(y ~ x, df)
}
```

3. Use built-in functions where possible

```r
myfun <- function(m1, df) {
  m1$coef[1] + m1$coef[2] * df[ ,1]
}
```

vs.

```r
myfun <- function(m1, df) {
  predict(m1, df)
}
```

4. Keep things tidy (80 character per line)

```r
group_X_fit <- function(training_data) {
  m1 <- lm(ViolentCrimesPerPop~ I(MalePctDivorce^(
              PctUrban + MedRent, d)
  m1
}
```

vs.

```r
group_X_fit <- function(training_data) {
  m1 <- lm(ViolentCrimesPerPop ~
              I(MalePctDivorce^(1/2)) +
              I(PctUrbanr^(1/3)) +
              PctIlleg +
              PctUrban +
              MedRent, d)
  m1
}
```

5. Think towards general usage

```r
myfun <- function(m1, df) {
  mean((df$response - predict(m1))^2)
}
```

vs.

```r
myfun <- function(m1, df) {
  mean((df$response - predict(m1, df))^2)
}
```

Model Selection

Activity

Use the `leaps` package and the `regsubsets()` function to perform model selection on the crimes data set. Start with forward selection to `nvmax = 25`. The general form of the function is:

```
regsubsets(y ~ ., data = train, nvmax = 25, method = "forward")
```

Notes:

- You will need to do some `select()`ing before you can toss the training data set in there.
- Try investigating the output of this function by plotting it using `plot()` and calling on its `attributes()` and `str()`ucture.
- If you complete forward selection, try backwards and compare.
- If you complete both and have a sense of a good model, fit it and compute it's test MSE.

By Tuesday 11:59 pm

Revise `lab-03.Rmd` to include an additional model fitting function called `group_X_automated_fit()` and `group_X_process()`.d