

Git

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.6      v dplyr   1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
## Warning: package 'tibble' was built under R version 4.0.5
```

```
## Warning: package 'tidyr' was built under R version 4.0.5
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
## Warning: package 'forcats' was built under R version 4.0.5
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(broom)
```

```
## Warning: package 'broom' was built under R version 4.0.5
```

```
library(glue)
```

```
## Warning: package 'glue' was built under R version 4.0.5
```

```
library(palmerpenguins)
```

```
## Warning: package 'palmerpenguins' was built under R version 4.0.5
```

```
# fit a linear model
fit <- lm(bill_length_mm ~ body_mass_g, data = penguins)
summary(fit) #summarizes the model
```

```
##
## Call:
## lm(formula = bill_length_mm ~ body_mass_g, data = penguins)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.1251  -3.0434  -0.8089   2.0711  16.1109
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.690e+01  1.269e+00   21.19  <2e-16 ***
## body_mass_g  4.051e-03  2.967e-04   13.65  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.394 on 340 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.3542, Adjusted R-squared:  0.3523
## F-statistic: 186.4 on 1 and 340 DF,  p-value: < 2.2e-16
```

```
glance(fit) # extracts model-level summary data from a fitted object
```

```
## # A tibble: 1 x 12
##   r.squared adj.r.squared sigma statistic p.value    df logLik   AIC   BIC
##   <dbl>      <dbl> <dbl>    <dbl>    <dbl> <dbl> <dbl> <dbl> <dbl>
## 1    0.354        0.352  4.39    186. 3.81e-34     1 -991. 1987. 1999.
## # ... with 3 more variables: deviance <dbl>, df.residual <int>, nobs <int>
```

```
tidy(fit) # extracts info about individual regression coefficients
```

```
## # A tibble: 2 x 5
##   term          estimate std.error statistic p.value
##   <chr>          <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)  26.9      1.27     21.2 3.86e-64
## 2 body_mass_g   0.00405    0.000297  13.7 3.81e-34
```

```

penguin_fit <- penguins %>%
  nest(data = ~species)%>% # nest the data table by species
  mutate(
    fit = map(data, ~lm(bill_length_mm~body_mass_g, data= .x)), # map() a linear model to each nested data table
    glance_output = map(fit, glance) #use map() to apply glance() on each model_fit
  ) %>%
  unnest(cols= glance_output) %>% #unnest output from the column name glance_output
  select(-data, -fit) # select all data except fit and data

# plot the model
# data table manipulation- convert the data into one with 4 columns
label_data <- penguin_fit %>%
  mutate(body_mass_g = 5000, # coordinates of text label
    bill_length_mm = 25, # coordinates of text label
    label=glue("p = {signif(p.value,2)}") #round p-value to 2 decimal places
  ) %>%
  select(species, body_mass_g, bill_length_mm, label)

ggplot(penguins, aes(body_mass_g, bill_length_mm))+
  geom_point(na.rm= TRUE)+
  geom_smooth(method = 'lm', na.rm = TRUE)+
  facet_wrap(~species)+
  geom_text(
    data = label_data,
    aes(label= label)
  )

```

```

## `geom_smooth()` using formula 'y ~ x'

```

