broom practical

Jumping Rivers

Question 1

First, let's load the required packages and data

```
library(broom)
library(tidyverse)
library(jrTidyverse2)
data(beer, package = "jrTidyverse2")
```

This data contains roughly 1500 beers, their alcohol percentage and their colour. Colour is ranked from 1-50, with 50 being the highest.

1) Use `?beer` and `head(beer)` to get a brief overview of the data.

2) We are going to look at how the colour of a beer affects the alcohol percentage using linear regression.

```
fit = lm(ABV ~ Color, data = beer)
```

The above code will run a linear regression model with alcohol percentage, `ABV`, as the response and colour, `Color`, as a variable. Explore the output of `summary(fit)`. Can you grab the p-values?
Hint: use `summary(fit)$coefficients[,4]`

3) That method of grabbing the p-values was tiresome wasn't it? Tidy and store the output of `fit()` such that it is easier to grab the p-values.
Hint: use `tidy()`

4) Using the **ggally** package, produce a coefficient plot.

5) Now we are interested in visualising how well the model has performed. We can do this using the fitted values. Store the data along side the fitted values from the model.
Hint: use `augment()`

6) Amend the code in the notes given to make Figure 1.3 to plot the fitted values against the original data. Alternatively you can use base R to perform this task by using `plot()` and `points()`. Does it look like the model has performed adequately?

7) Grab the adjusted r squared for the model.
Hint: use `glance()`

8) Adjusted R squared is a measure of how well the model is explaining the variation in the data. It is basically a measure of how close all of the original points are to the fitted values. This value can be anything between 0-1. 0 would mean the model explains no variation and therefore is not very good whilst 1 would be the model explains all of the variation and therefore is very good. How good is the model at explaining the variation in the data?