AE 15: Modeling houses in Duke Forest

In this application exercise, we will

- use bootstrapping to quantify the uncertainty around a measure of center – median
- use bootstrapping to quantify the uncertainty around a measure of relationship – slope
- interpret confidence intervals

The dataset are on housing prices in Duke Forest – a dataset you've seen before! It's called `duke_forest` and it's in the **openintro** package. Additionally, we'll use **tidyverse** and **tidymodels** packages.

```
library(tidyverse)
library(tidymodels)
library(openintro)
```

Typical size of a house in Duke Forest

Exercise 1

Visualize the distribution of sizes of houses in Duke Forest. What is the size of a typical house?

```
# add code here
```

Exercise 2

Construct a 95% confidence interval for the typical size of a house in Duke Forest. Interpret the interval in context of the data.

```
# add code here
```

Add interpretation here.

Exercise 3

Without calculating it – would a 90% confidence interval be wider or narrower? Why?

Add response here.

Exercise 4

Construct the 90% confidence interval and interpret it.

```
# add code here
```

Add interpretation here.

Relationship between price and size

The following model predicts price of a house in Duke Forest from its size.

```
df_price_area_fit <- linear_reg() |>
  fit(price ~ area, data = duke_forest)

tidy(df_price_area_fit)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)  116652.   53302.       2.19 3.11e- 2
2 area            159.      18.2      8.78 6.29e-14
```

The slope can be interpreted as:

> For each additional square feet, the model predicts that prices of houses in Duke Forest are higher by $159, on average.

Exercise 5

Quantify the uncertainty around this slope using a 95% bootstrap confidence interval and interpret the interval in context of the data.

```
# add code here
```

Add interpretation here.