AE 4

Brandon Leslie

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.2     v tidyr     1.3.1
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

```
library(openintro)
```

```
Loading required package: airports
Loading required package: cherryblossom
Loading required package: usdata
```

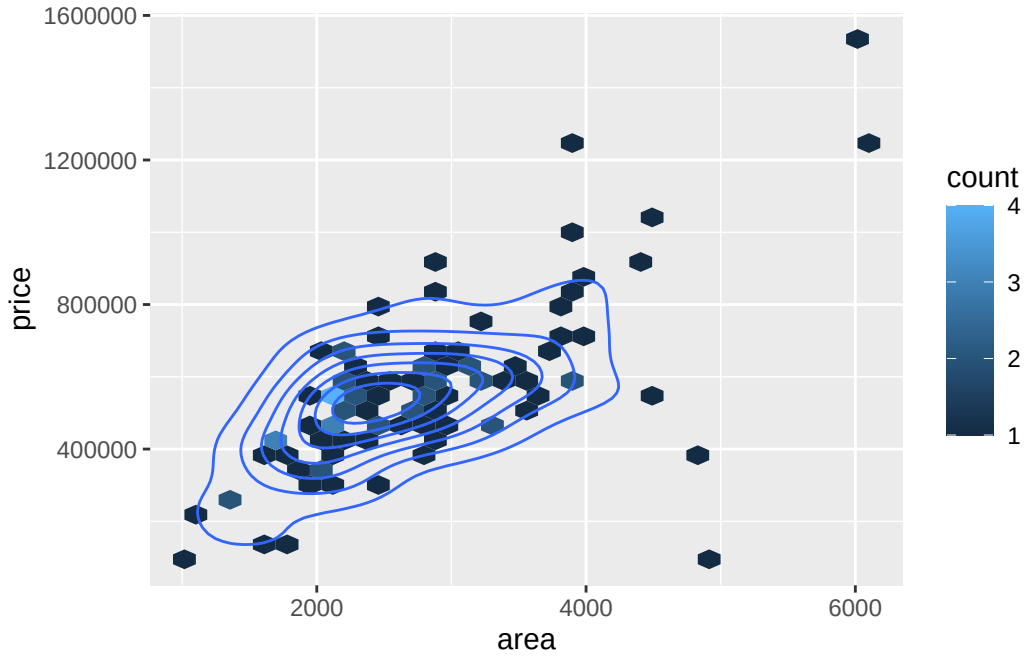
```
library(scales)
```

```

Attaching package: 'scales'

The following object is masked from 'package:purrr':

    discard

The following object is masked from 'package:readr':

    col_factor
```

```
duke_forest <- read_csv("data/duke_forest.csv")
```

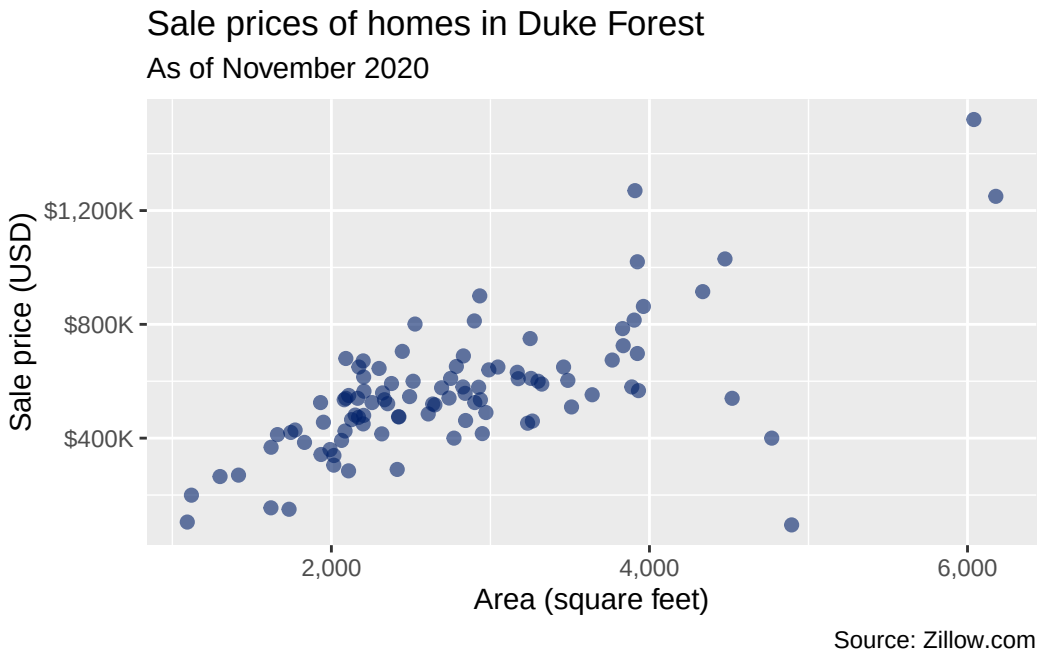
```
Rows: 98 Columns: 13
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (7): address, type, heating, cooling, parking, hoa, url
dbl (6): price, bed, bath, area, year_built, lot

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
ggplot(duke_forest, aes(x = area, y = price)) +
  geom_hex() +
geom_density2d()
```

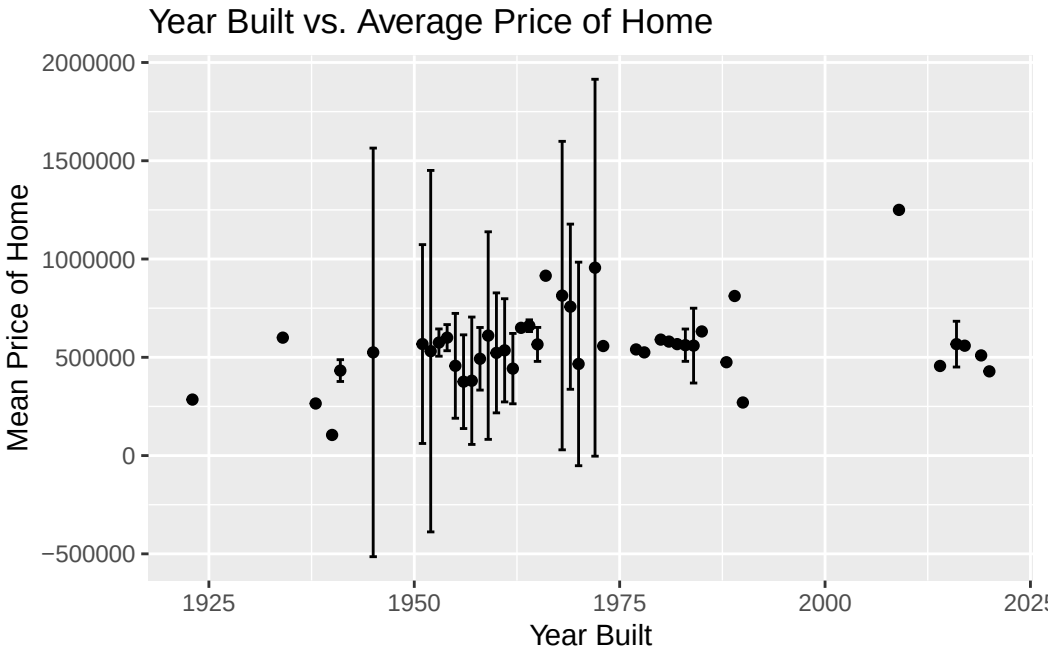



```
ggplot(duke_forest, aes(x = area, y = price)) +
  geom_point(alpha = 0.6, size = 2, color = "#012169") +
  scale_x_continuous(labels = label_number(big.mark = ",")) +
  scale_y_continuous(labels = label_dollar(scale = 1/1000, suffix = "K")) +
  labs(
    x = "Area (square feet)",
    y = "Sale price (USD)",
```

```
    title = "Sale prices of homes in Duke Forest",
    subtitle = "As of November 2020",
    caption = "Source: Zillow.com"
  )
```




```
duke_forest_mean <- duke_forest %>%
  group_by(year_built) %>%
  summarize(avg_price = mean(price), sd_price = sd(price))

duke_forest_mean %>%
  ggplot(aes(x = year_built, y = avg_price)) +
  geom_point() +
  geom_errorbar(aes(ymin = avg_price - 1.96 * sd_price, ymax = avg_price + 1.96 * sd_price)) +
  labs(
    title = "Year Built vs. Average Price of Home",
    x = "Year Built",
    y = "Mean Price of Home"
  )
```


Year Built vs. Average Price of Home
Mean Price of Home
2000000
1500000
1000000
500000
0
−500000
1925
1950
1975
2000
Year Built