Visualizing Texas: `pdf_document`

2019-03-15

Inspired by: https://rmarkdown.rstudio.com/lesson-6.html

Packages

We'll use `ggplot2` for visualization, and some light `dplyr` for data wrangling.

```
library(ggplot2) # plotting
library(dplyr) # wrangling
```

Texas housing data

This data is loaded for you when you install and load the `ggplot2` package.

```
txsamp <- txhousing %>%
  filter(city %in% c("Houston", "Fort Worth", "San Antonio", "Dallas", "Austin"))
glimpse(txsamp)
```

```
## Observations: 935
## Variables: 9
## $ city      <chr> "Austin", "Austin", "Austin", "Austin", "Austin", "A...
## $ year      <int> 2000, 2000, 2000, 2000, 2000, 2000, 2000, 2000, 2000...
## $ month     <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 1, 2, 3, 4, 5...
## $ sales     <dbl> 1025, 1277, 1603, 1556, 1980, 1885, 1818, 1880, 1498...
## $ volume    <dbl> 173053635, 226038438, 298557656, 289197960, 39307377...
## $ median    <dbl> 133700, 134000, 136700, 136900, 144700, 148800, 1493...
## $ listings  <dbl> 3084, 2989, 3042, 3192, 3617, 3799, 3944, 3948, 4058...
## $ inventory <dbl> 2.0, 2.0, 2.0, 2.1, 2.3, 2.4, 2.6, 2.6, 2.6, 2.6, 2....
## $ date      <dbl> 2000.000, 2000.083, 2000.167, 2000.250, 2000.333, 20...
```

Our data is monthly

Here is just a sample of rows from one city to show that we have data for each of the 12 months for each year, except for 2015.

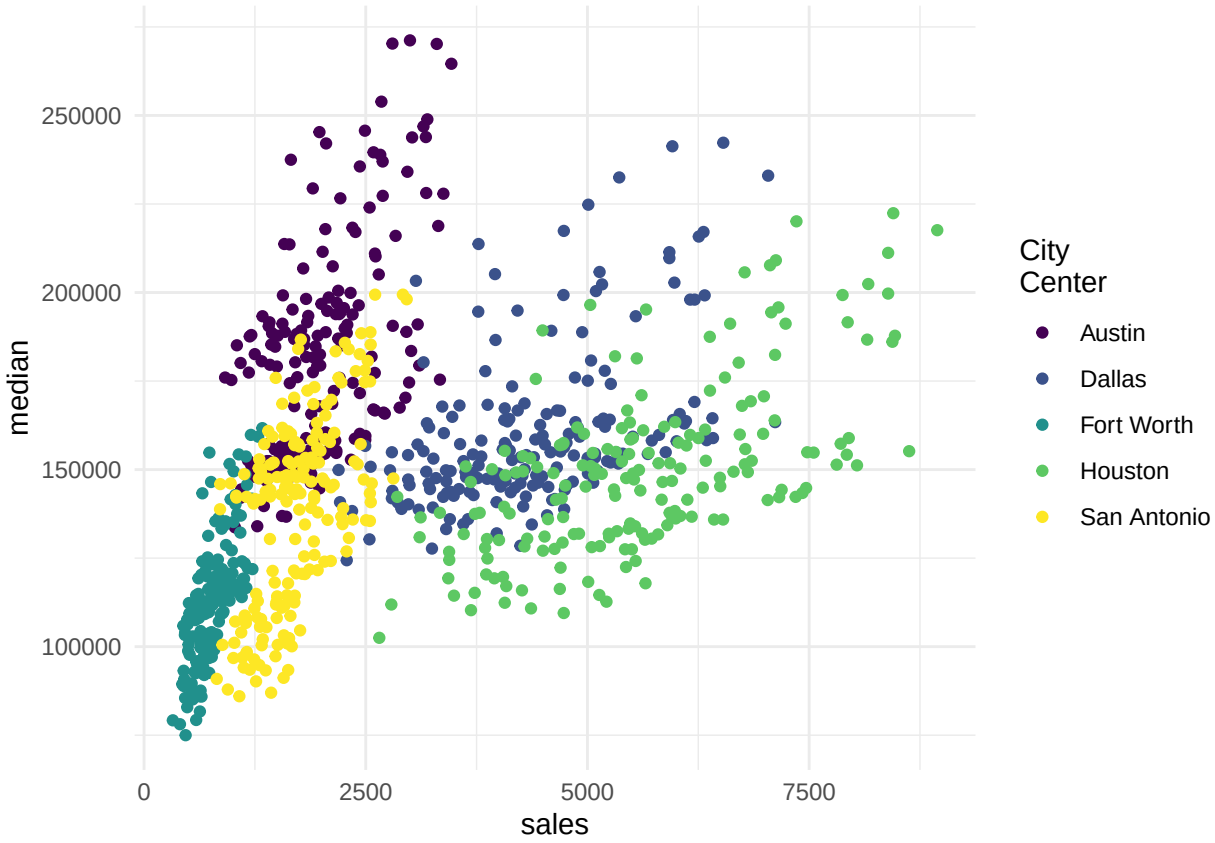
```
txsamp %>%
  filter(city == "Austin") %>%
  count(year)
```

```
## # A tibble: 16 x 2
##     year     n
##    <int> <int>
##  1  2000    12
```

```
##  2  2001    12
##  3  2002    12
##  4  2003    12
##  5  2004    12
##  6  2005    12
##  7  2006    12
##  8  2007    12
##  9  2008    12
## 10  2009    12
## 11  2010    12
## 12  2011    12
## 13  2012    12
## 14  2013    12
## 15  2014    12
## 16  2015     7
```

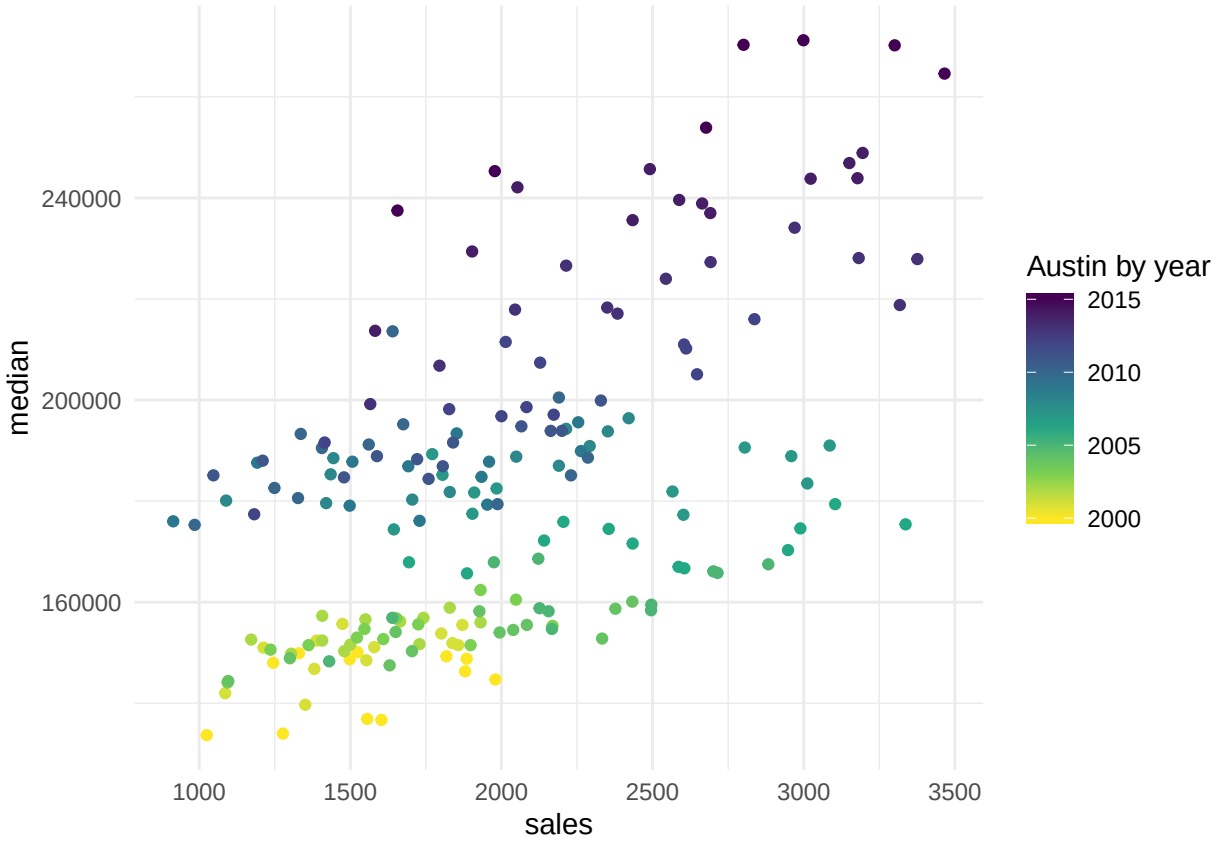
Austin is expensive

```
ggplot(data = txsamp, aes(x = sales, y = median)) +
   geom_point(aes(colour = city)) +
   scale_colour_viridis_d("City\nCenter", option = params$viridis_palette)
```



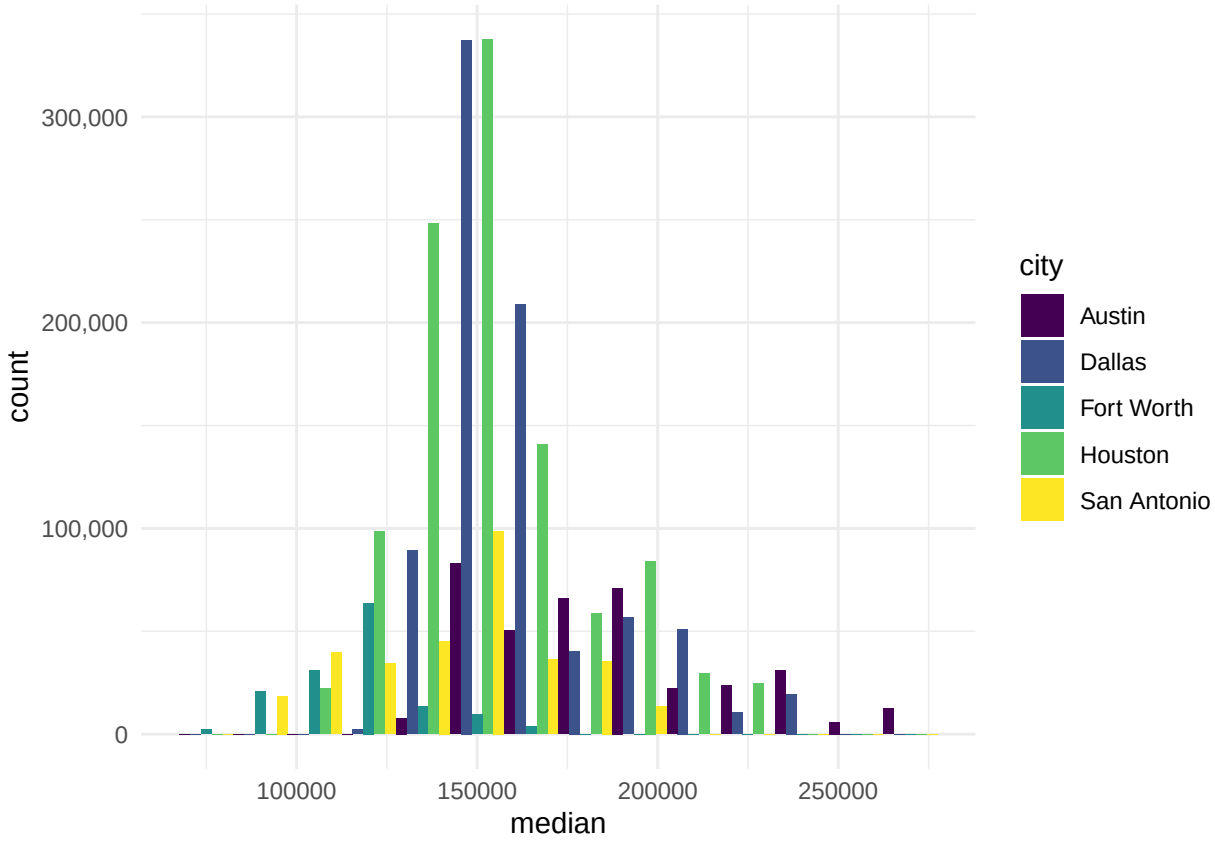
Austin prices on the rise

```
ggplot(data = filter(txsamp, city == "Austin"), aes(x = sales, y = median)) +
   geom_point(aes(colour = year)) +
   scale_colour_viridis_c("Austin by year", option = params$viridis_palette, direction = -1)
```



Fort Worth has more affordable housing

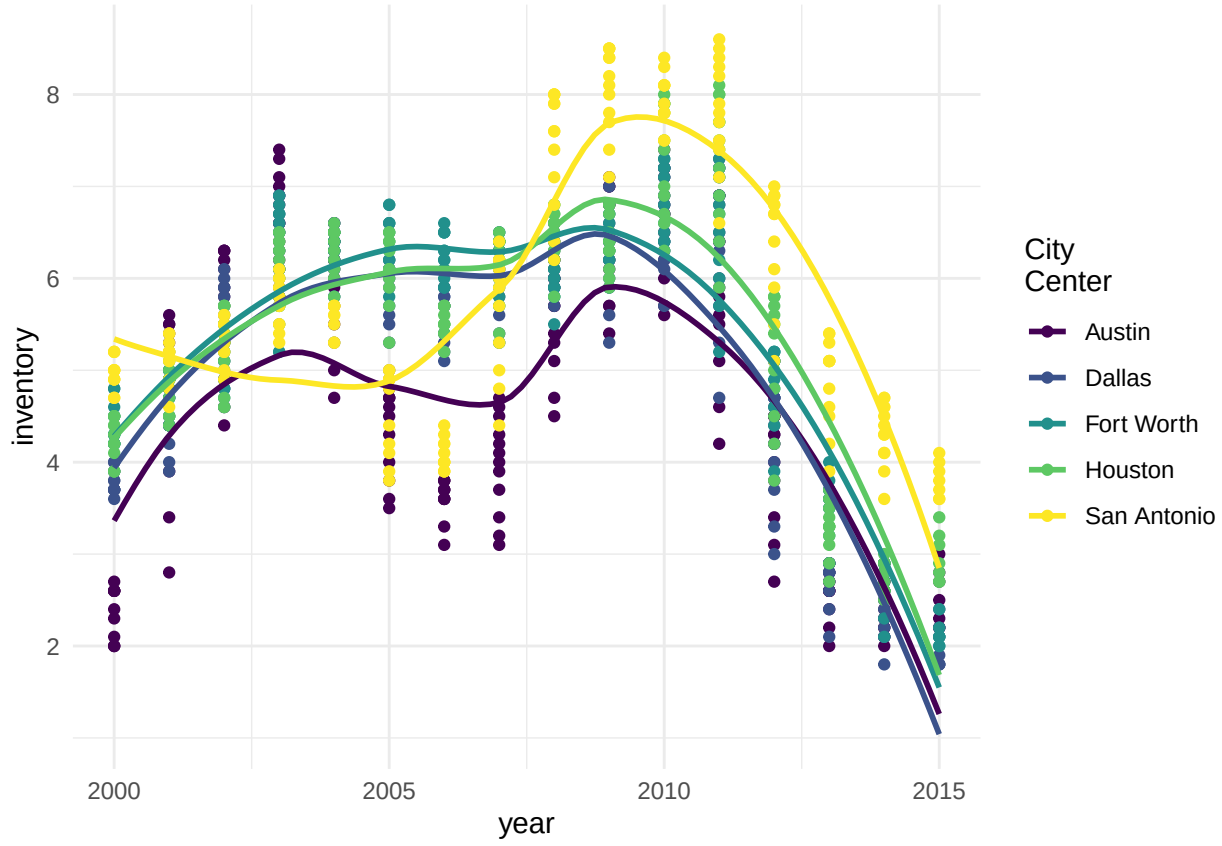
```
library(scales) # to make y-axis in non-scientific notation
ggplot(txsamp, aes(x = median, fill = city)) +
  geom_histogram(aes(weight = sales), position = "dodge", binwidth = 15000) +
  scale_fill_viridis_d(option = params$viridis_palette)+
  scale_y_continuous(labels = comma)
```



The current pace of sales is fast

"Months inventory": amount of time it would take to sell all current listings at current pace of sales.

```
ggplot(data = txsamp, aes(x = year, y = inventory, colour = city)) +
  geom_point() +
  geom_smooth(se = FALSE) +
  scale_colour_viridis_d("City\nCenter", option = params$viridis_palette)
```


inventory
8
6
4
2
2000
2005
2010
2015
year
City
Center
Austin
Dallas
Fort Worth
Houston
San Antonio