Data Science for Everyone – Grammar of Graphics

Dr. Ab Mosca (they/them)

Slides based off slides courtesy of Jordan Crouser (https://jcrouser.github.io/)

Plan for Today

- Connect what we know about visualizations to ggplot in R

ggplot2

- Plot building blocks
  - data
  - aesthetic mappings (how we draw that stuff)
  - geometric objects (the literal stuff we draw)
  - statistical transformations (underlying model)
  - scales (range of values, colors, etc.)
  - faceting (small multiples)

```
ggplot(data, aes()) +
   geom_*
```

data

aesthetic mapping

geometric object

ggplot2

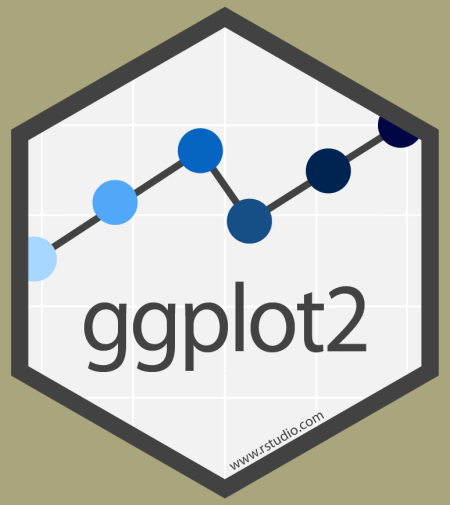
- Plot building blocks
  - data
  - aesthetic mappings (how we draw that stuff)
  - geometric objects (the literal stuff we draw)
  - statistical transformations (underlying model)
  - scales (range of values, colors, etc.)
  - faceting (small multiples)

```
ggplot(data, aes()) +
   geom_*
```

data

aesthetic mapping

geometric object



Data

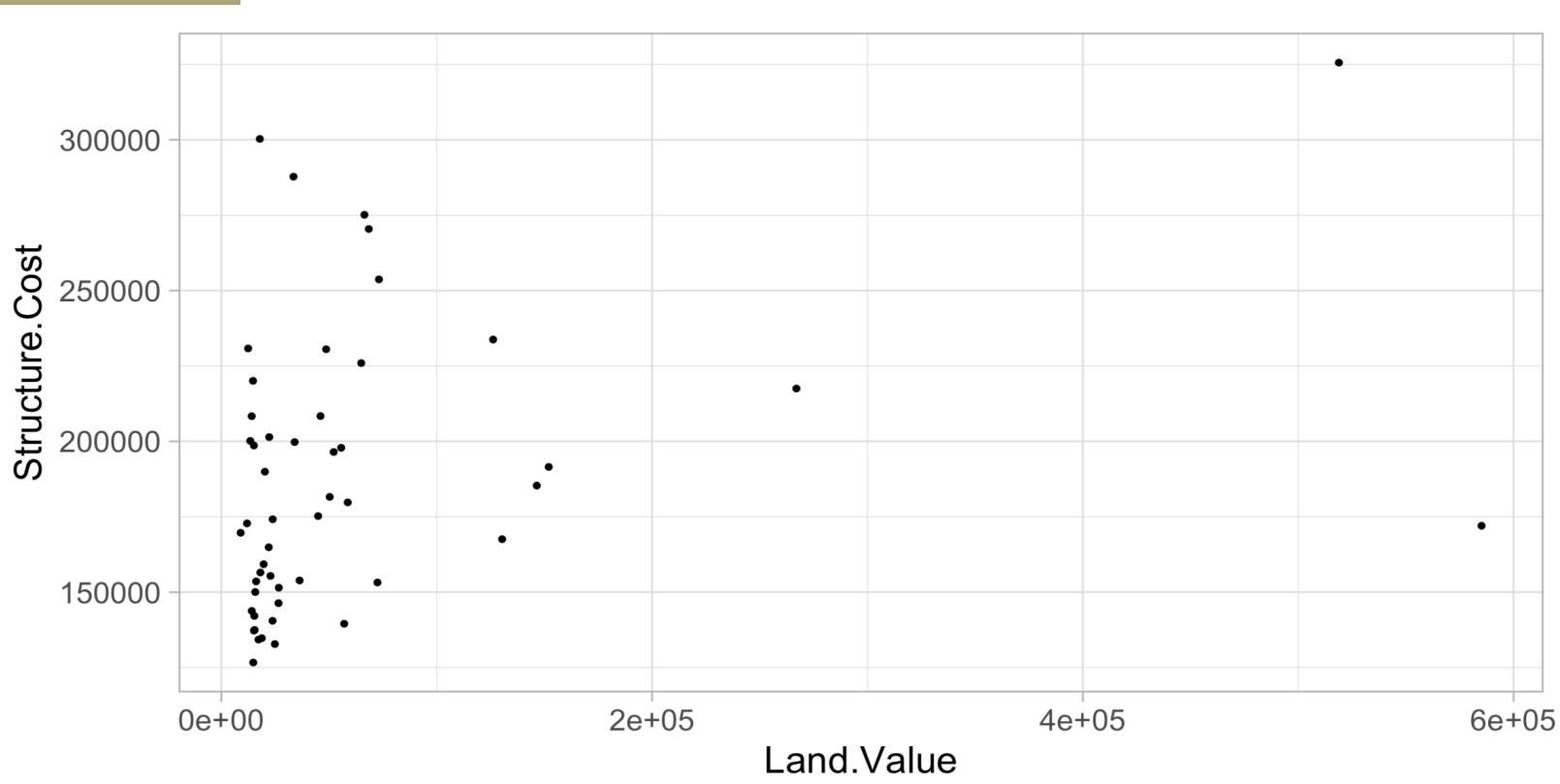
- For today, upload the landdata-states.csv from the course website (under Labs tab) to your R Studio working directory
- Load the data and look at it

```
housing_ds <- read_csv("landdata-states.csv")
glimpse(housing_ds)
```

```
## Rows: 7,803
## Columns: 11
## $ State            <chr> "AK", "AK", "AK", "AK", "AK", "AK", "AK", "AK", "AK",…
## $ region           <chr> "West", "West", "West", "West", "West", "West", "West…
## $ Date             <dbl> 2010.25, 2010.50, 2009.75, 2010.00, 2008.00, 2008.25,…
## $ Home.Value       <int> 224952, 225511, 225820, 224994, 234590, 233714, 23299…
## $ Structure.Cost   <int> 160599, 160252, 163791, 161787, 155400, 157458, 16009…
## $ Land.Value       <int> 64352, 65259, 62029, 63207, 79190, 76256, 72906, 6946…
## $ Land.Share..Pct. <dbl> 28.6, 28.9, 27.5, 28.1, 33.8, 32.6, 31.3, 29.9, 28.7,…
## $ Home.Price.Index <dbl> 1.481, 1.484, 1.486, 1.481, 1.544, 1.538, 1.534, 1.52…
## $ Land.Price.Index <dbl> 1.552, 1.576, 1.494, 1.524, 1.885, 1.817, 1.740, 1.66…
## $ Year             <int> 2010, 2010, 2009, 2009, 2007, 2008, 2008, 2008, 2008,…
## $ Qrtr             <int> 1, 2, 3, 4, 4, 1, 2, 3, 4, 1, 2, 2, 3, 4, 1, 2, 3, 4,…
```

Simple Scatterplot

```
# filter
hp2013Q1 <- housing %>% filter(Date == 2013.25)
# plot
ggplot(hp2013Q1, aes(x = Land.Value, y = Structure.Cost)) +
  geom_point()
```



Geometric Objects (geom) + Aesthetics

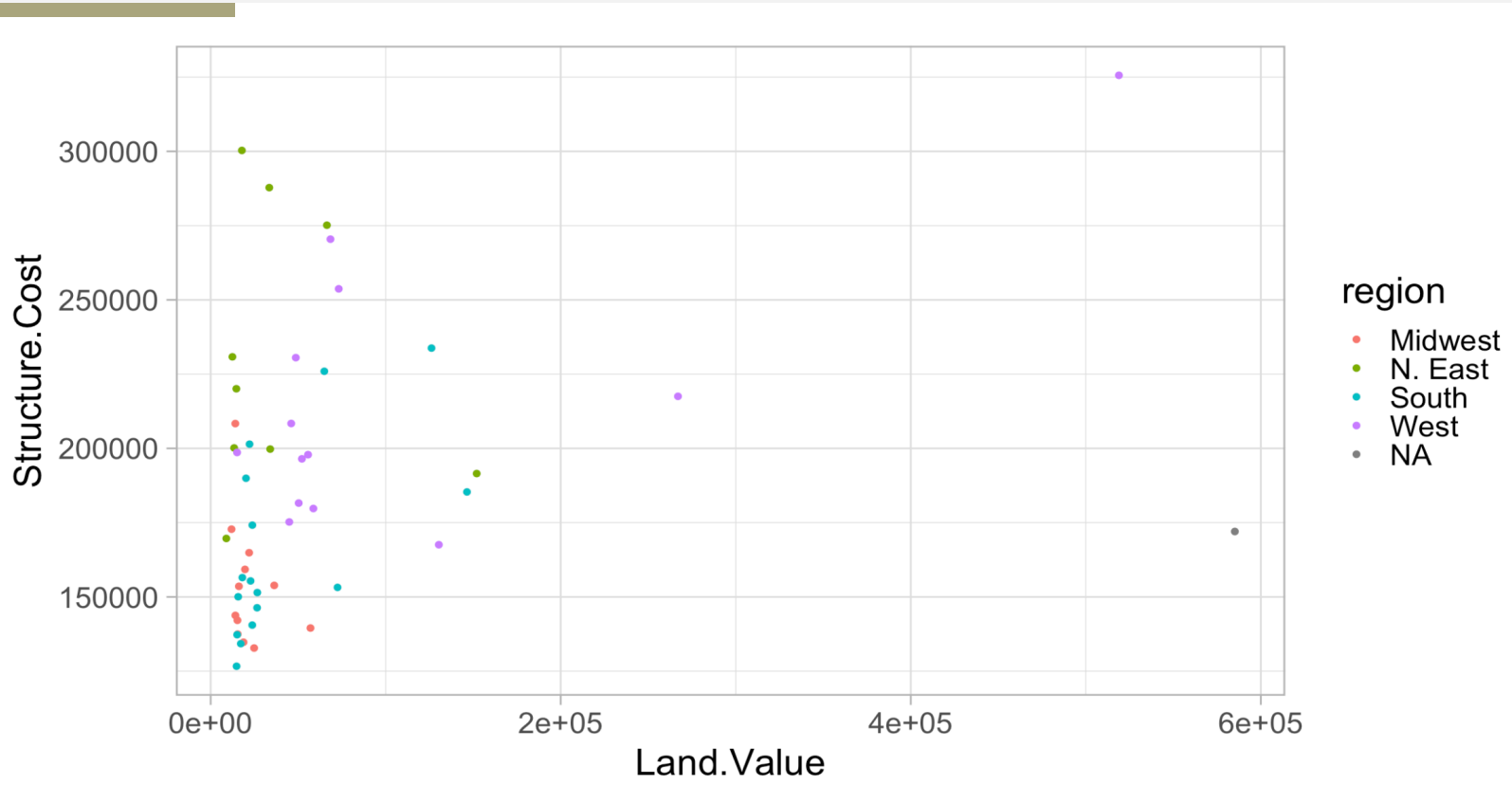
- Apply aesthetics to geometric objects
- Ex.
  - position (i.e., on the x and y axes)
  - color ("outside" color)
  - fill ("inside" color)
  - shape (of points)
  - line type
  - size

Geometric Objects (geom) + Aesthetics

- Apply aesthetics to geometric objects
- Ex.
  - position (i.e., on the x and y axes)
  - color ("outside" color)
  - fill ("inside" color)
  - shape (of points)
  - line type
  - size

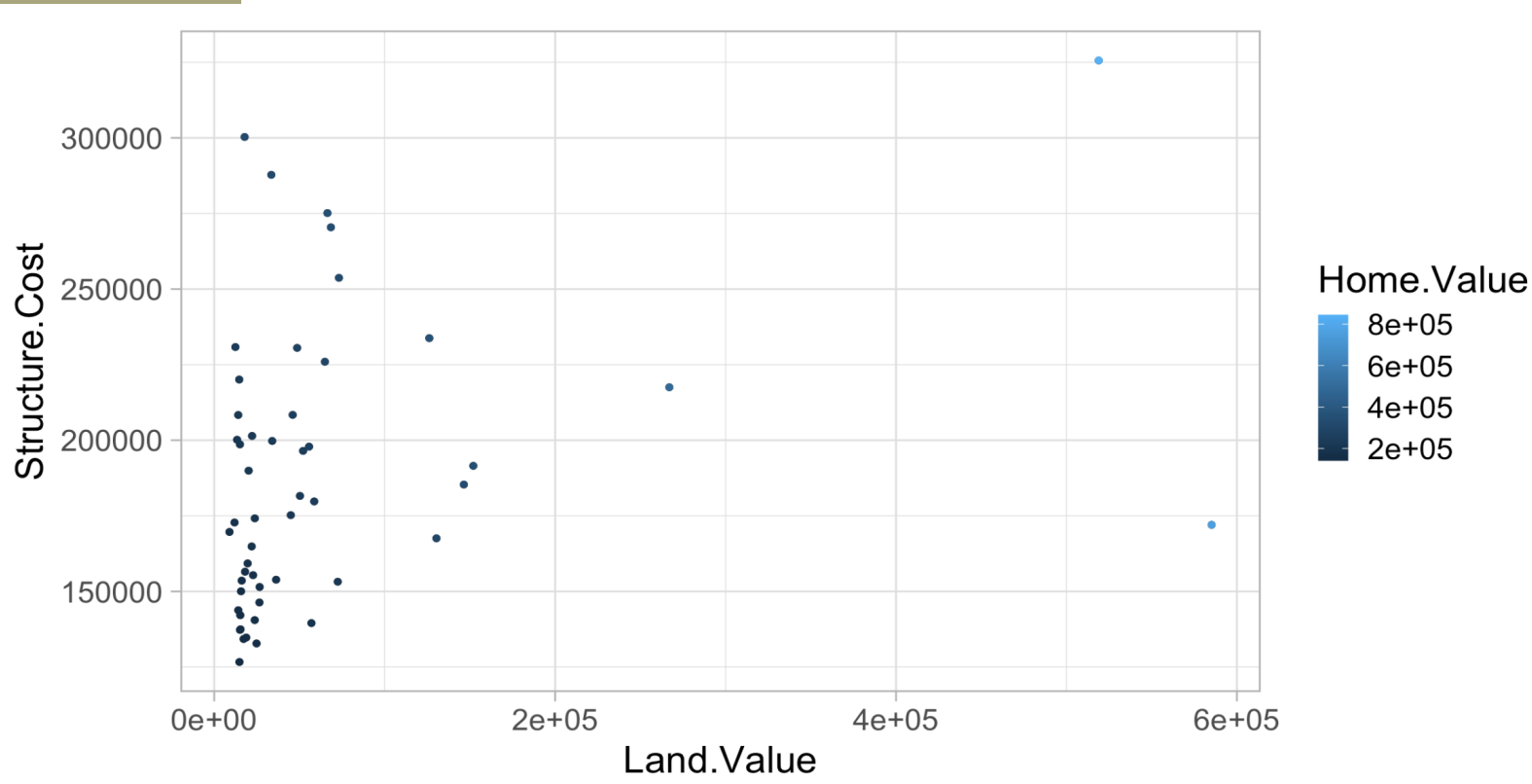
Simple Scatterplot

```
# filter
hp2013Q1 <- housing %>% filter(Date == 2013.25)
# plot
ggplot(hp2013Q1, aes(x = Land.Value, y = Structure.Cost)) +
  geom_point(aes(color = region))
```



Simple Scatterplot

```
# filter
hp2013Q1 <- housing %>% filter(Date == 2013.25)
# plot
ggplot(hp2013Q1, aes(x = Land.Value, y = Structure.Cost)) +
  geom_point(aes(color = Home.Value))
```



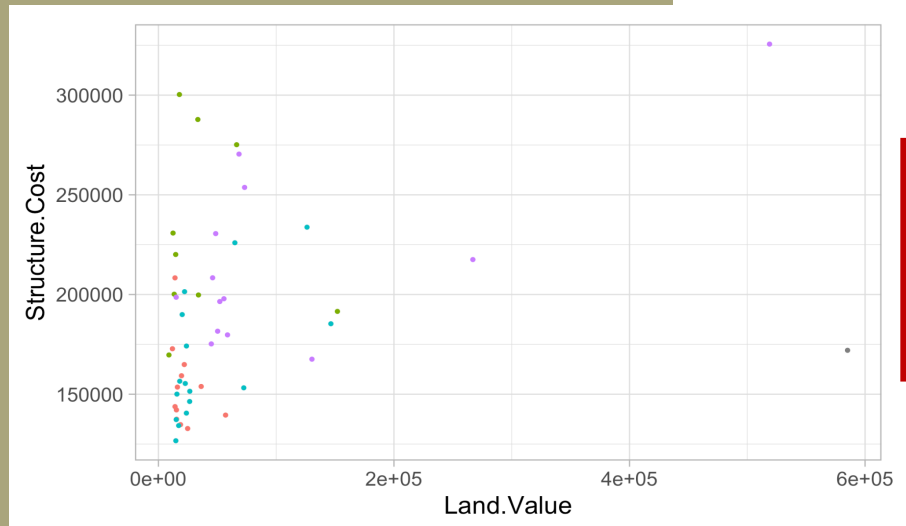
Simple Scatterplot

```
# filter
hp2013Q1 <- housing %>% filter(Date == 2013.25)
# plot
ggplot(hp2013Q1, aes(x = Land.Value, y = Structure.Cost)) +
  geom_point(aes(color = ??))
```

Remember visual channels? Is "color" hue or value?

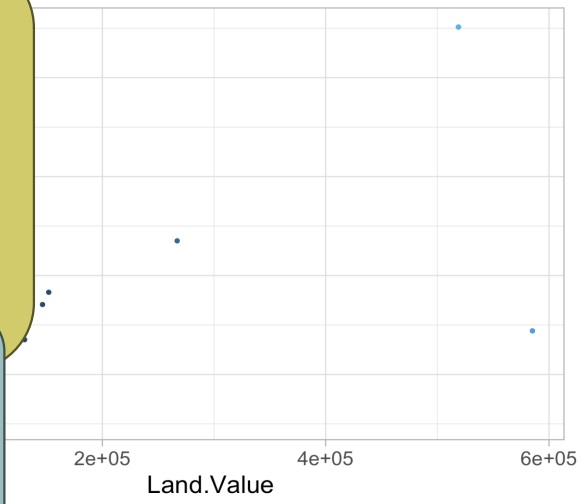
Simple Scatterplot

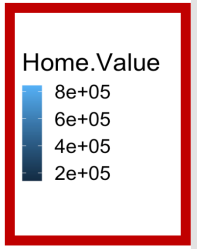
```
# filter
hp2013Q1 <- housing %>% filter(Date == 2013.25)
# plot
ggplot(hp2013Q1, aes(x = Land.Value, y = Structure.Cost)) +
  geom_point(aes(color = ??))
```



Remember visual channels? Is "color" hue or value?

Both! If you do not select a specific color pallet, R automatically picks based on data





region is categorical, so use hue

Home.Value is quantitative, so use value

Note: Assignment vs. Aesthetic Mapping

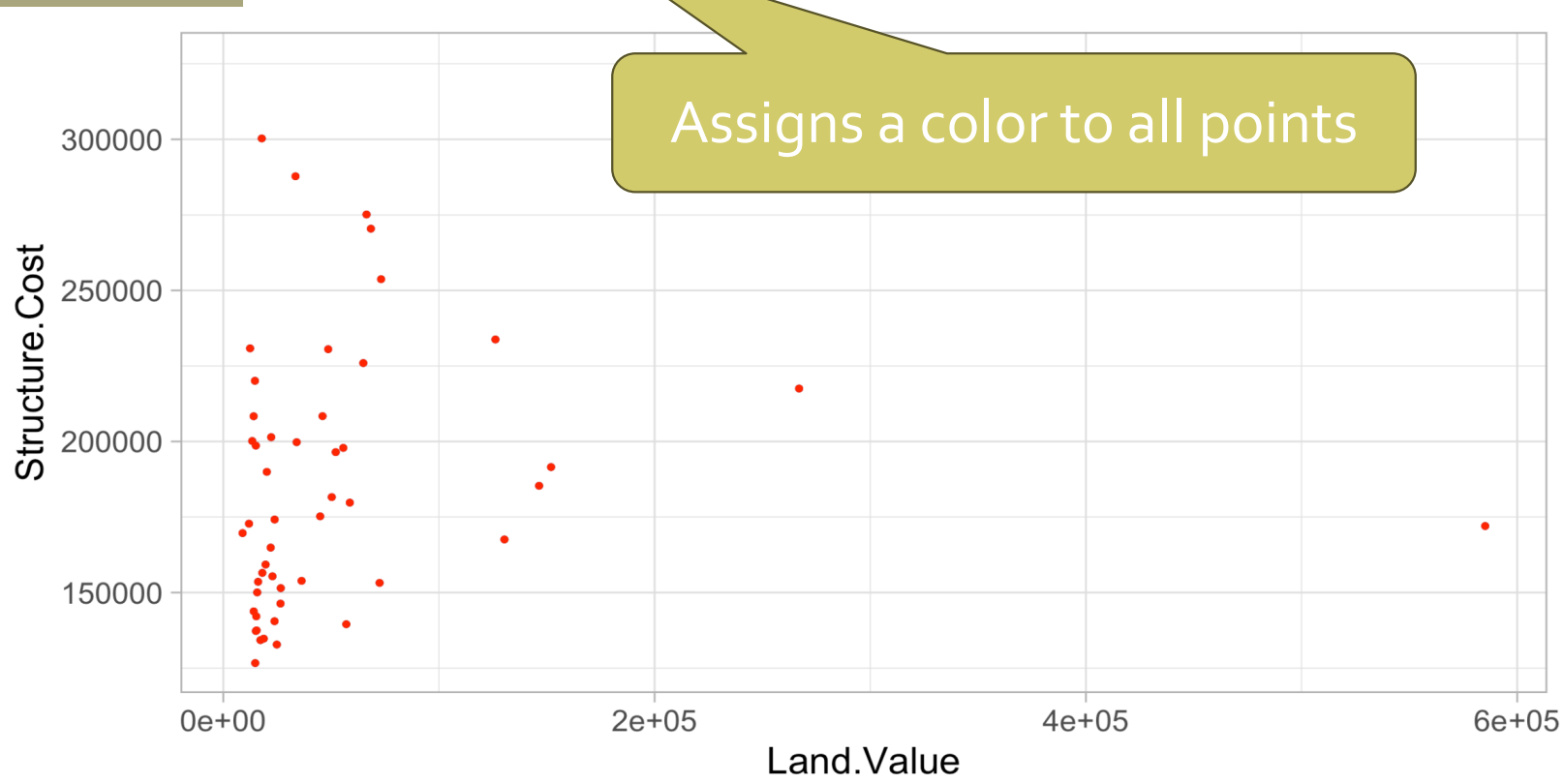
```
# filter
hp2013Q1 <- housing %>% filter(Date == 2013.25)
# plot
ggplot(hp2013Q1, aes(x = Land.Value, y = Structure.Cost)) +
  geom_point(aes(color = ??))
```

Maps color to variable

Note: Assignment vs. Aesthetic Mapping

```
# filter
hp2013Q1 <- housing %>% filter(Date == 2013.25)
# plot
ggplot(hp2013Q1, aes(x = Land.Value, y = Structure.Cost)) +
  geom_point(color = red)
```

Assigns a color to all points



Controlling Aesthetic Mapping

- aes() says to map an aesthetic to a variable, it does not specify *how*

Controlling Aesthetic Mapping

- aes() says to map an aesthetic to a variable, it does not specify *how*
- Ex.

```
# filter
hp2013Q1 <- housing %>% filter(Date == 2013.25)
# plot
ggplot(hp2013Q1, aes(x = Land.Value, y = Structure.Cost)) +
  geom_point(aes(color = Home.Value))
```

Maps color to Home.Value, but doesn't specify *what* color

Controlling Aesthetic Mapping

- aes() says to map an aesthetic to a variable, it does not specify *how*
- This is controlled through scales
- In ggplot scales include:
  - `position`
  - `color and fill`
  - `size`
  - `shape`
  - `linetype`

Controlling Aesthetic Mapping

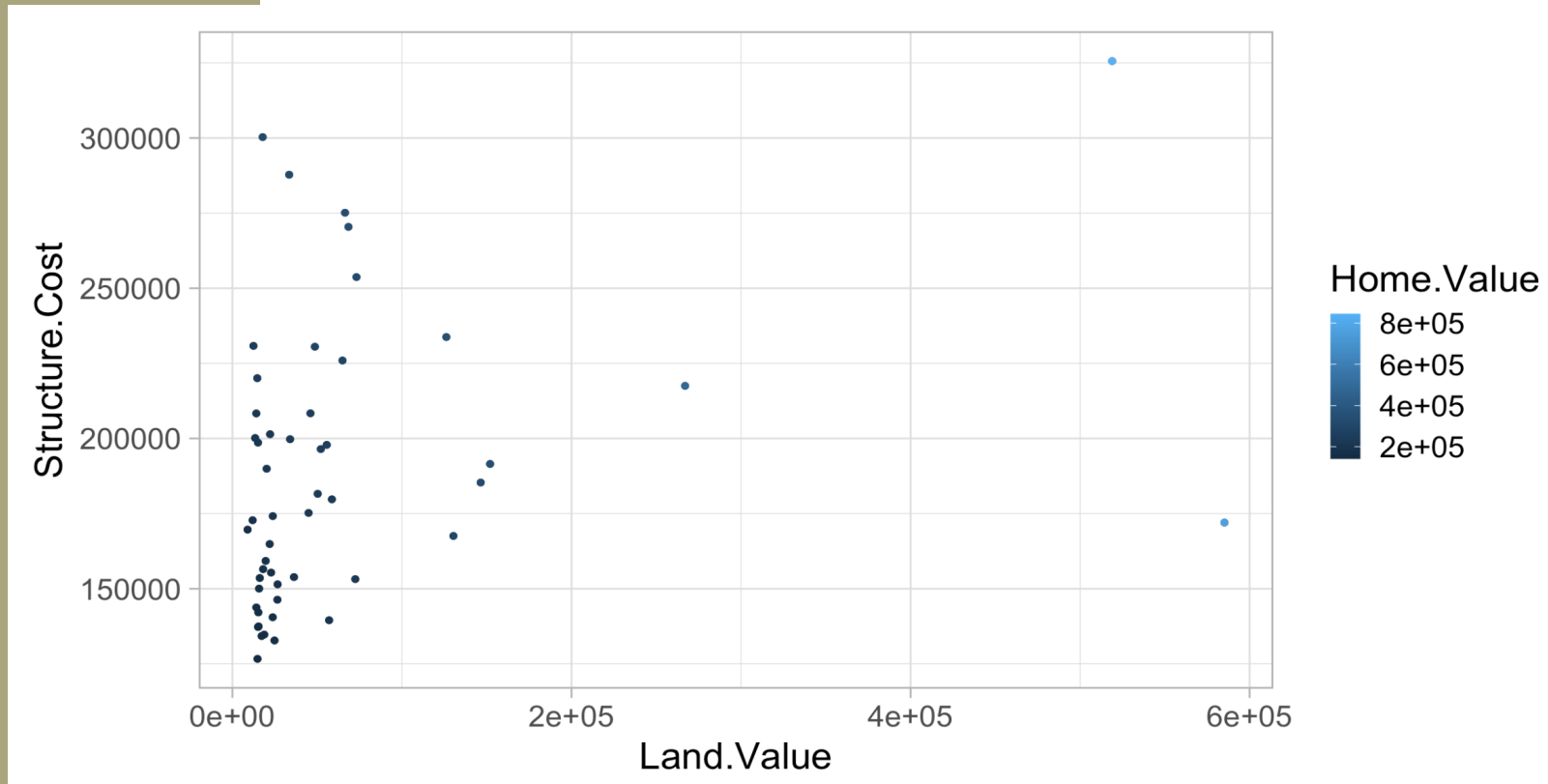
- aes() says to map an aesthetic to a variable, it does not specify *how*
- This is controlled through scales
- In ggplot scales include:
  - `position`
  - `color and fill`
  - `size`
  - `shape`
  - `linetype`

Controlling Aesthetic Mapping

- aes() says to map an aesthetic to a variable, it does not specify *how*
- This is controlled through scales
- The following arguments are common to most scales in ggplot2:
  - **`name`**: the first argument specifies the axis or legend title
  - **`limits`**: the minimum and maximum of the scale
  - **`breaks`**: the points along the scale where labels should appear
  - **`labels`**: the text that appears at each break

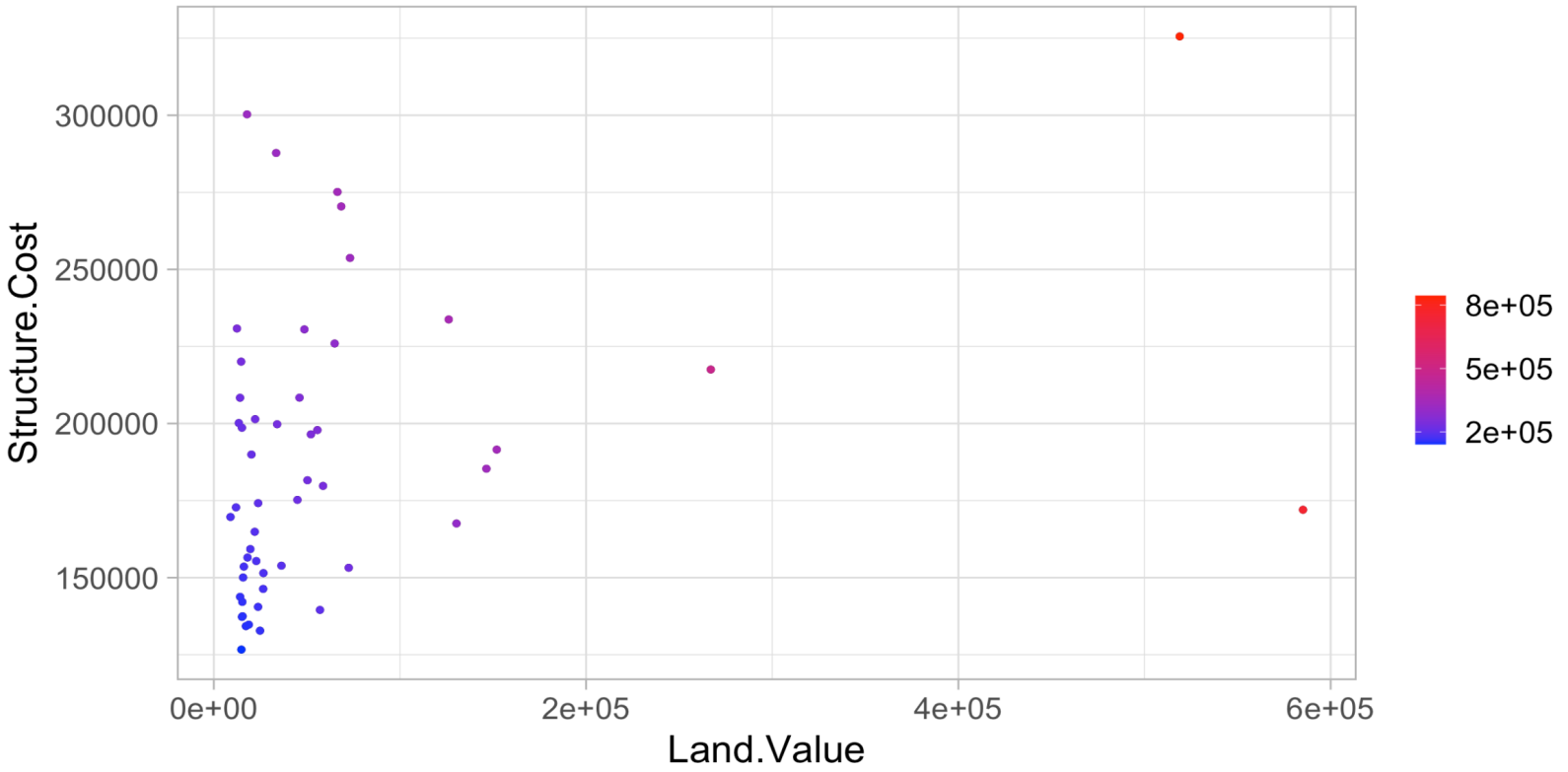
Controlling Aesthetic Mapping

```
p1 <- ggplot(hp2013Q1, aes(x = Land.Value, y = Structure.Cost))+
        geom_point(aes(color = Home.Value))
```



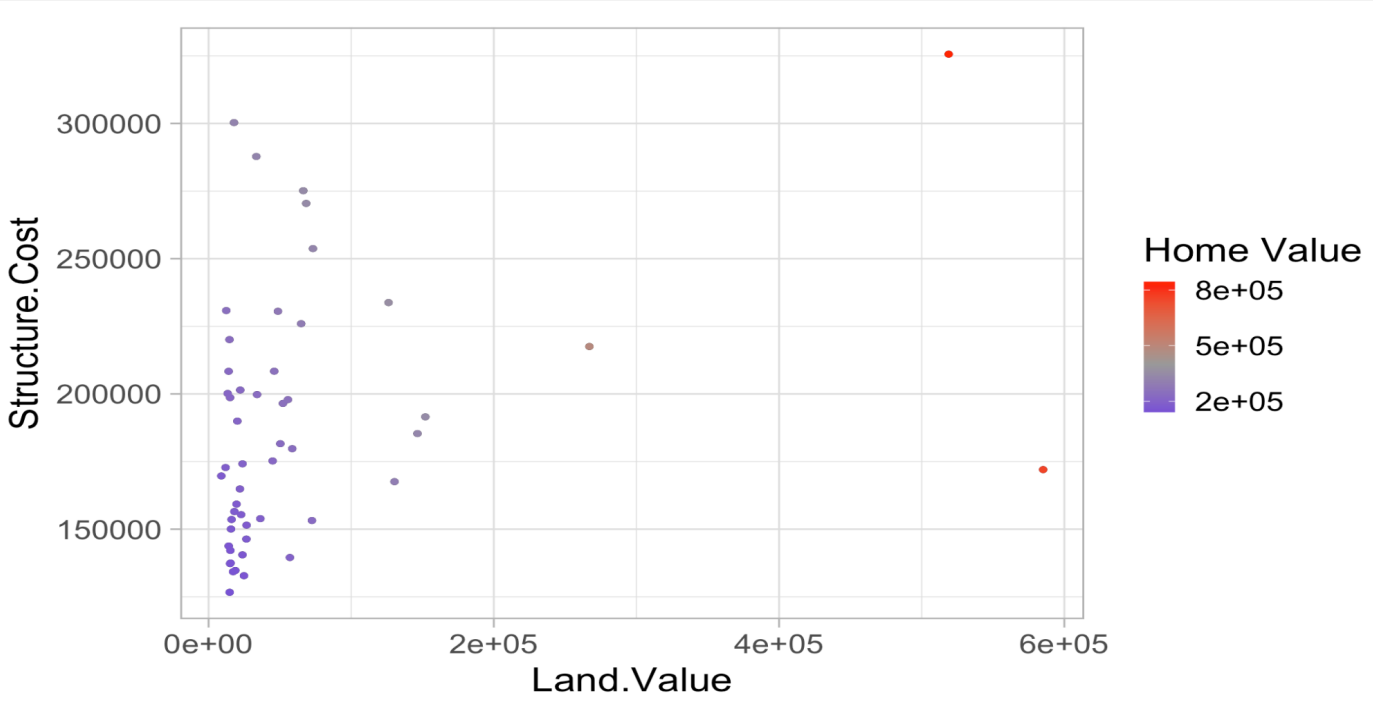
Controlling Aesthetic Mapping

```
# change color scale
p1 +
  scale_color_continuous(name = "",
                         breaks = c(200001, 500001, 800001),
                         labels = c(200000, 500000, 800000),
                         low = "blue", high = "red")
```



Controlling Aesthetic Mapping

```
# change color scale
p1 + scale_color_gradient2(name = ”Home Value",
                           breaks = c(200001, 500001, 800001),
                           labels = c(200000, 500000, 800000),
                           low = "blue", high = "red",
                           mid = "gray60",
                           midpoint = 400000)
```

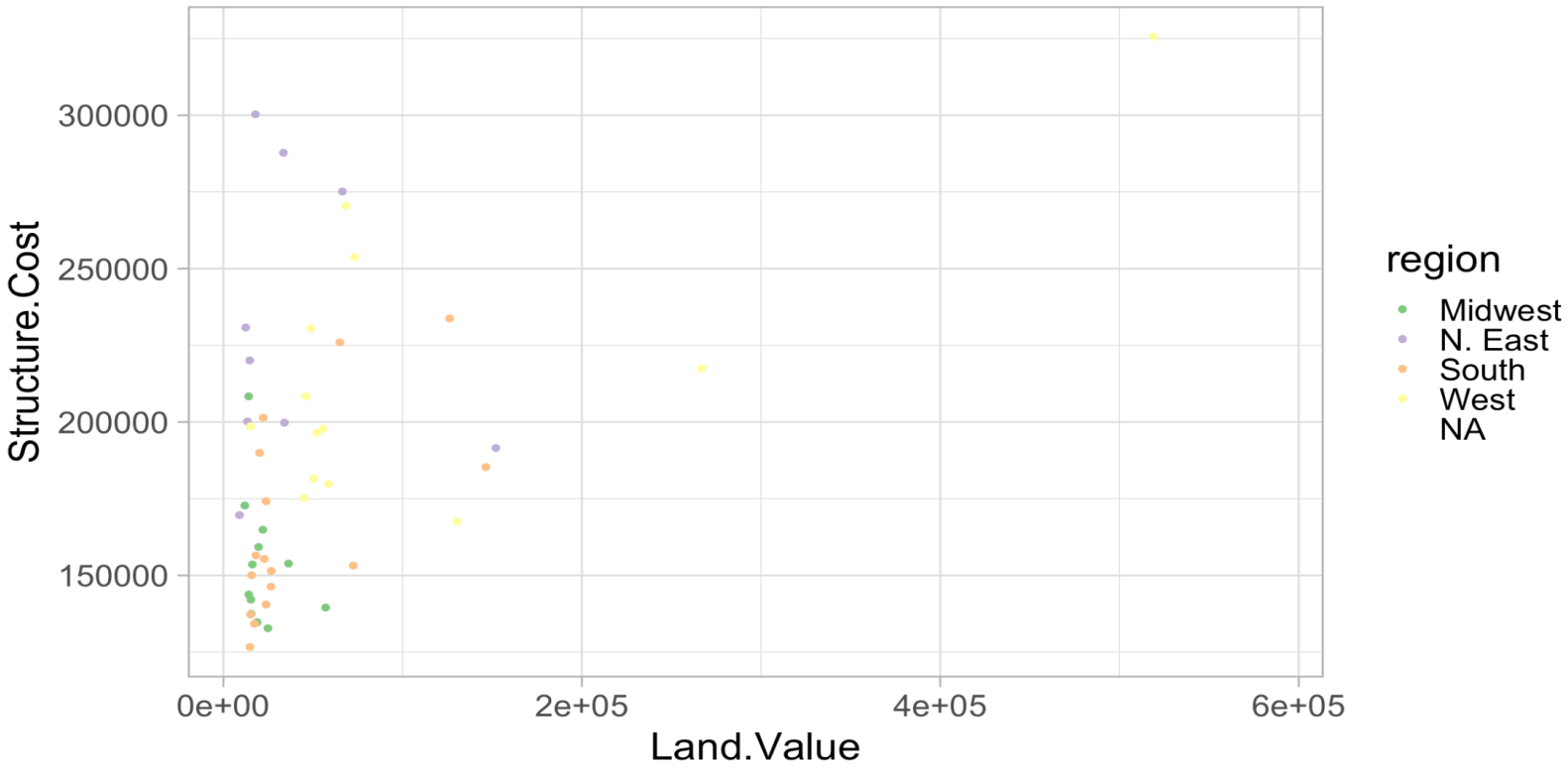


Controlling Aesthetic Mapping – colorbrewer

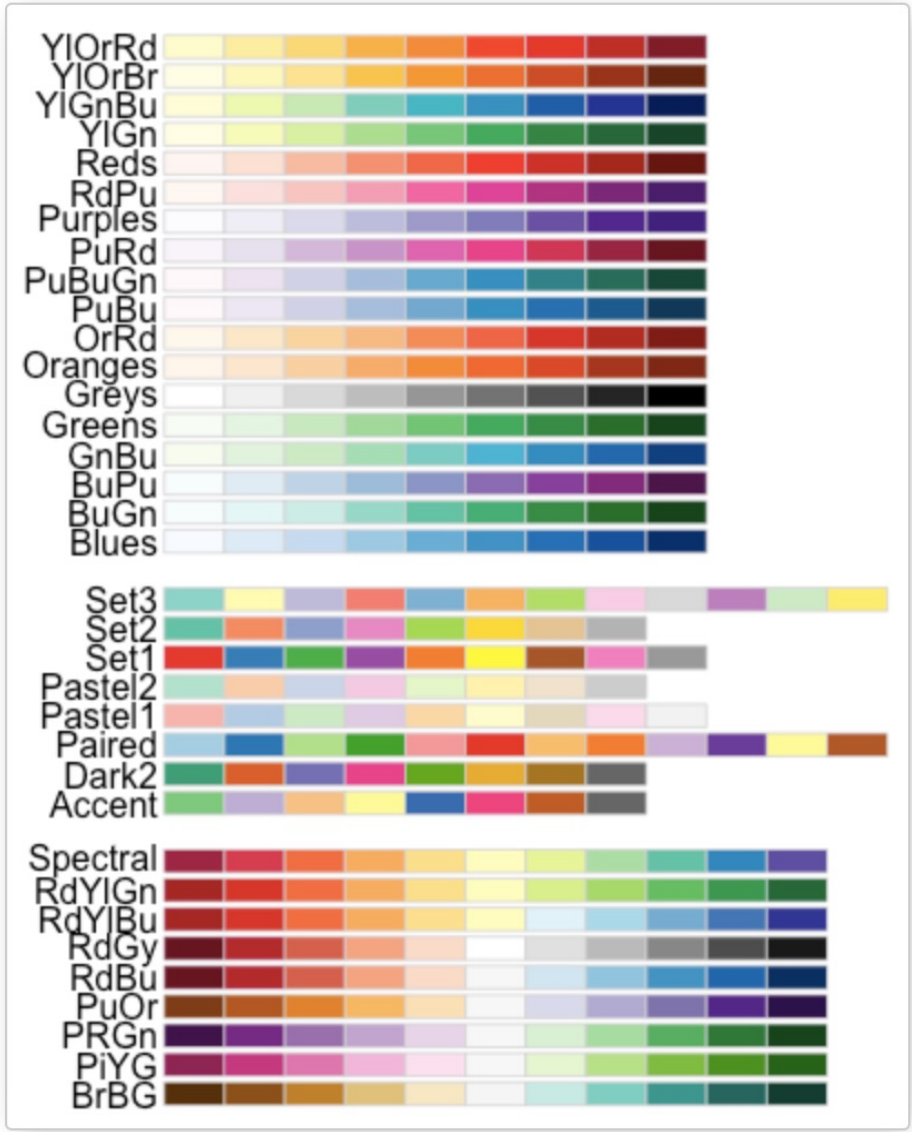
- We can also use built in color maps
- Be sure to choose the correct map for your data type

Controlling Aesthetic Mapping – colorbrewer

```
# use existing color scale
ggplot(hp2013Q1, aes(x = Land.Value, y = Structure.Cost)) +
  geom_point(aes(color = region)) +
  scale_colour_brewer(palette = "Accent")
```



Controlling Aesthetic Mapping – colorbrewer



`ggplot2` tips

- Cheatsheet: https://www.rstudio.com/resources/cheatsheets/

Now, try it out!

Work with 1-2 other people.

Ask a question you can answer with the landdata dataset. Make a graph to answer your question that uses ggplot2, and a colormap.