Introduction 简介

Materials and setup 准备材料和配置

Laptop users: You should have R installed –if not: 笔记本用户-你应该安装 R-如果没有的话：

1. Open a web browser and go to http://cran.r-project.org and download and install it 打开一个网页浏览器然后打开 http://cran.r-project.org 并下载和安装它
2. Also helpful to install RStudio (download from http://rstudio.com) 也可以安装 RStudio（从 http://rstudio.com）
3. In R, type `install.packages("ggplot2")` to install the ggplot2 package. 在 R 里面，输入 `install.packages("ggplot2")` 安装 ggplot2 包。

Everyone: Download workshop materials:

每个人：下载讲习班材料

1. Download materials from http://tutorials.iq.harvard.edu/R/Rgraphics.zip 从 http://tutorials.iq.harvard.edu/R/Rgraphics.zip 下载材料
2. Extract the zip file containing the materials to your desktop 将 zip 文件夹中的材料解压到你的桌面

Workshop Overview

讲习班大纲

Class Structure and Organization:

课程的结构和组织：

- Ask questions at any time. Really!在任何时候问问题，真的！
- Collaboration is encouraged 鼓励大家合作
- This is your class! Special requests are encouraged 这是你们的课堂！鼓励大家有特殊的要求

This is an intermediate R course:

这是一个中级的 R 课程：

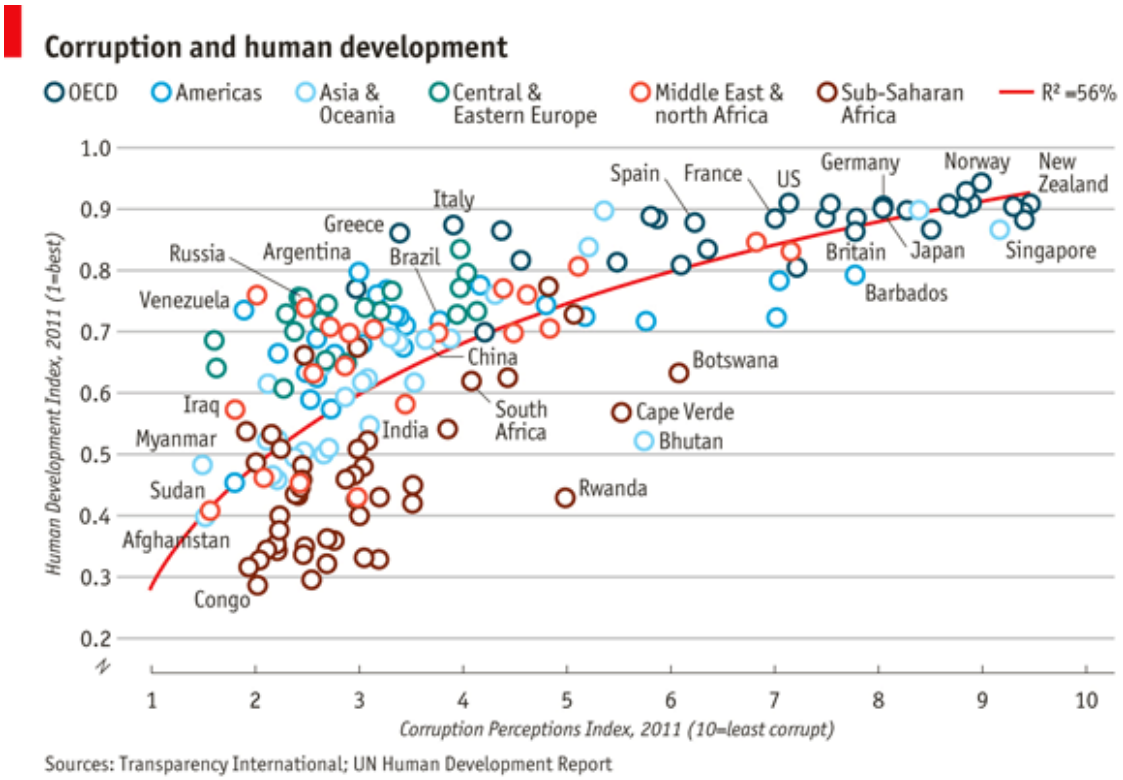
- Assumes working knowledge of R 假设有 R 的工作基础
- Relatively fast-paced 相对快的节奏
- Focus is on `ggplot2` graphics - other packages will not be covered 专注讨论 ggplot2 图像-其他的包讲不会被讲到

Starting At The End

在开始的最后

My goal: by the end of the workshop you will be able to reproduce this graphic from the Economist:

我们的目标：在讲习班的最后，你会知道如何重制这个《经济学人》上的图像：



Why ggplot2?

为什么用 ggplot2？

Advantages of ggplot2

Ggplot2 的优势

- consistent underlying `grammar of graphics` (Wilkinson, 2005)
- 有潜在一致的图像语法 (Wilkinson, 2005)
- plot specification at a high level of abstraction
- 在一个高层次的抽象上画专业图像
- very flexible
- 非常的灵活
- theme system for polishing plot appearance
- 有润色图像的主题系统
- mature and complete graphics system
- 成熟且完整的图像系统
- many users, active mailing list

- 有许多用户、活跃的邮件列表

That said, there are some things you cannot (or should not) do With ggplot2:

然而，这里有很多事情你不能（也不应该）用 ggplot2 来做：

- 3-dimensional graphics (see the rgl package)
- 3 维图像（用 rgl 包）
- Graph-theory type graphs (nodes/edges layout; see the igraph package)
- 图论型图像（点/边 外观：用 igraph 包）
- Interactive graphics (see the ggvis package)
- 交互的图像（用 ggvis 包）

What Is The Grammar Of Graphics?

什么是图像的语法？

The basic idea: independently specify plot building blocks and combine them to create just about any kind of graphical display you want. Building blocks of a graph include:

基础的思想：独立地指明图像的构成区并整合它们来创造几乎任何类型你想要的的图像展示。一个图像的构成区包括：

- Data
- 数据
- aesthetic mapping
- 美学映射
- geometric object
- 几何物体
- statistical transformations
- 统计变换
- scales
- 尺度
- coordinate system
- 坐标系统

- position adjustments
- 位置调整
- faceting
- 分平面

Example Data: Housing prices

例子数据：房子价格

Let’s look at housing prices.

让我们来看看房子价格

```
housing <- read.csv("dataSets/landdata-states.csv")
head(housing[1:5])
```

```
  State region    Date Home.Value Structure.Cost
1    AK   West 2010.25     224952        160599
2    AK   West 2010.50     225511        160252
3    AK   West 2009.75     225820        163791
4    AK   West 2010.00     224994        161787
5    AK   West 2008.00     234590        155400
6    AK   West 2008.25     233714        157458
```

ggplot2 VS Base Graphics

ggplot2 和基础图像的对比

Compared to base graphics, `ggplot2`

对比于基础的图像，`ggplot2`

- is more verbose for simple / canned graphics
- 对于简单/基本的图像来说，是更冗长的
- is less verbose for complex / custom graphics
- 对于复杂/自定义的图像来说，是更简洁的
- does not have methods (data should always be in a `data.frame`)

- 不包含方法（数据永远应该在一个 data.frame 里）
- uses a different system for adding plot elements
- 用一个不同的系统来添加画图元素

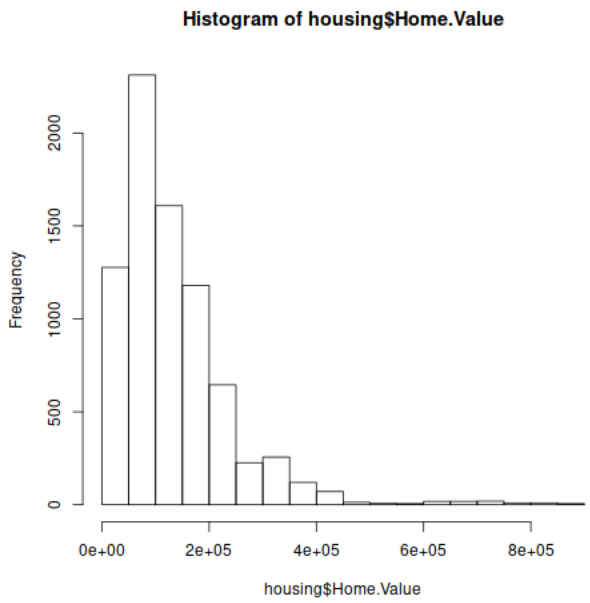
ggplot2 VS Base for simple graphs

ggolot2 与基本画图在简单图上的对比

Base graphics histogram example:

基本图形直方图例子：

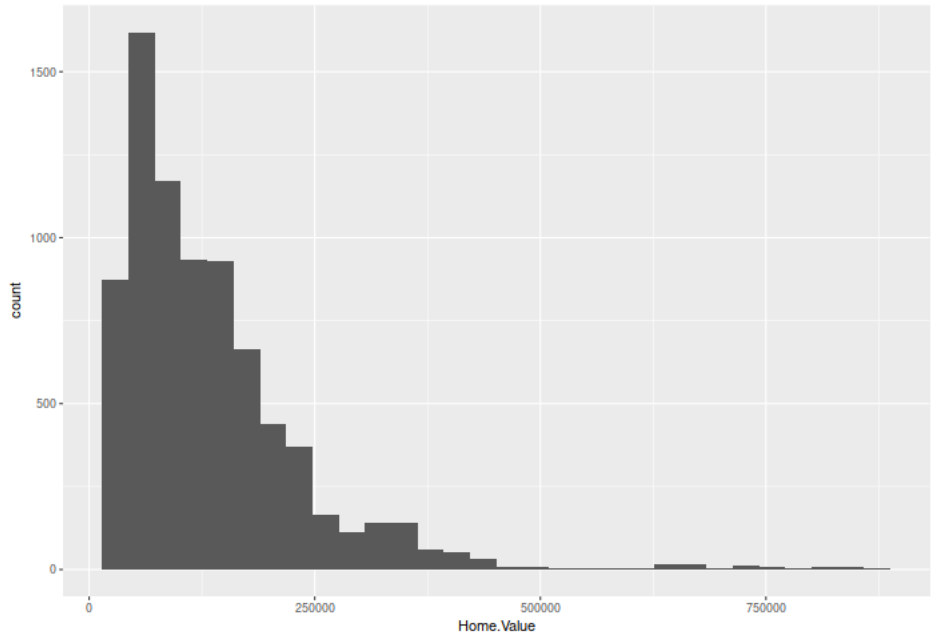
```
hist(housing$Home.Value)
```



ggplot2 histogram example:

ggplot2 直方图例子：

```
library(ggplot2)
ggplot (housing, aes(x = Home.Value)) +
  geom_histogram()
```



Base wins!

基本图像胜利！

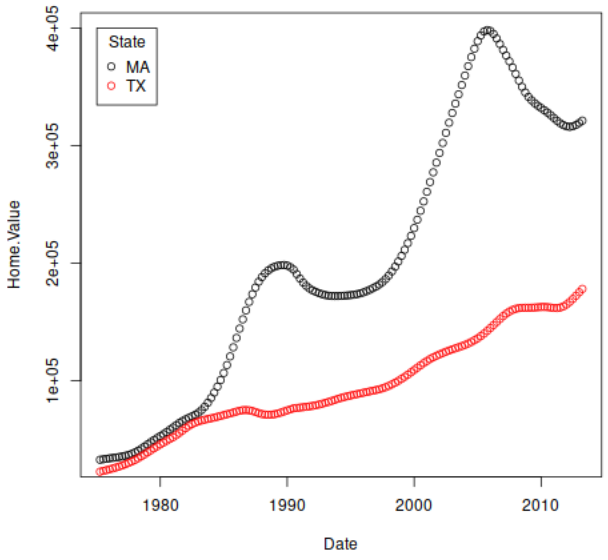
ggplot2 Base graphics VS ggplot for more complex graphs:

ggplot2 基本图像和 ggplot2 在更复杂的图像上的比较：

Base colored scatter plot example:

基本的上色散点图例子：

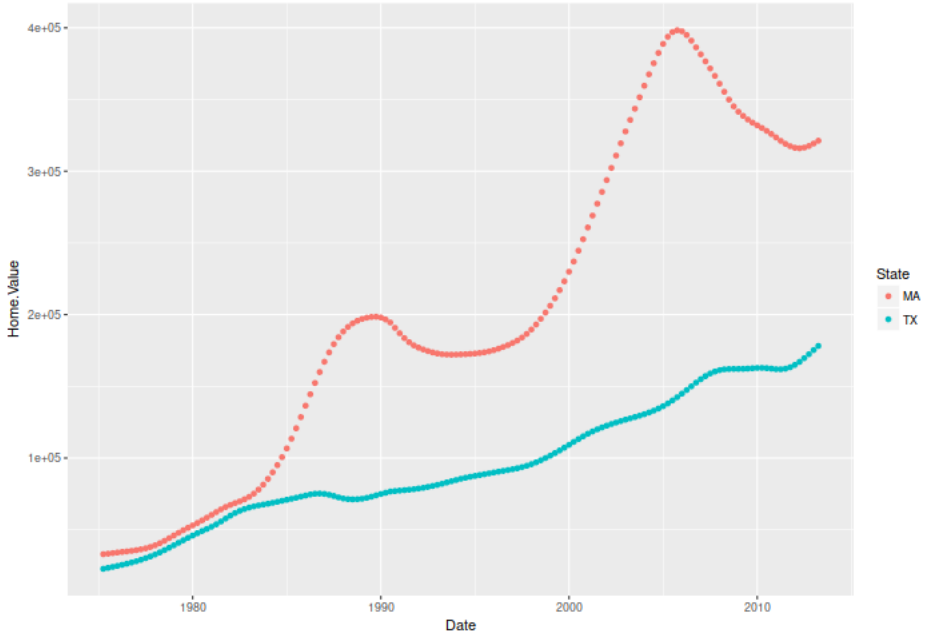
```
plot(Home.Value ~ Date,
     data=subset(housing, State == "MA"))
points(Home.Value ~ Date, col="red",
       data=subset(housing, State == "TX"))
legend(1975, 400000,
       c("MA", "TX"), title="State",
       col=c("black", "red"),
       pch=c(1, 1))
```



ggplot2 colored scatter plot example:

ggplot2 上色散点图例子：

```
ggplot(subset(housing, State %in% c("MA", "TX")),
       aes(x=Date,
           y=Home.Value,
           color=State))+
  geom_point()
```



`ggplot2` wins!

ggplot2 胜利!

Geometric Objects And Aesthetics

几何物体和美学

Aesthetic Mapping

美学映射

In ggplot land *aesthetic* means "something you can see". Examples include:

在 ggplot 中，美学代表着“任何你能看见的东西”。例子包括了：

- position (i.e., on the x and y axes)
- 位置（比如在 x 轴或 y 轴上）
- color ("outside" color)
- 颜色（“外面”的颜色）
- fill ("inside" color)
- 填充（“里面”的颜色）
- shape (of points)
- 形状（点的）
- linetype
- 线型
- size
- 大小

Each type of geom accepts only a subset of all aesthetics - refer to the geom help pages to see what mappings each geom accepts. Aesthetic mappings are set with the `aes()` function.

每个种类的几何只接受所有美学的一个子集-参考几何帮助界面来看每个几何层面什么映射。美学映射是用 aes()函数来定义的。

Geometic Objects (geom)

几何物体（geom）

Geometric objects are the actual marks we put on a plot. Examples include:

几何物体是我们放在一个图形中的标记。例子包括：

- points (`geom_point`, for scatter plots, dot plots, etc)
- 点（几何点，在散点图中、点图中，等等）
- lines (`geom_line`, for time series, trend lines, etc)
- 线（几何线，在时间序列，趋势线中，等等）
- boxplot (`geom_boxplot`, for, well, boxplots!)
- 箱形图（几何箱形图）

A plot must have at least one geom; there is no upper limit. You can add a geom to a plot using the `+` operator

一个图必须有至少一个几何物体：这里没有上限。你可以添加一个几何物体到图中用+运算符

You can get a list of available geometric objects using the code below:

你可以用下面的代码得到一系列的可用几何物体：

```
help.search("geom_", package = "ggplot2")
```

or simply type `geom_<tab>` in any good R IDE (such as Rstudio or ESS) to see a list of functions starting with `geom_`.

或者简单的输入 `geom_<tab>` 在任何好的 R IDE 中（比如 Rstudio 或者 ESS）来查看一系列由 geom 开头的函数。

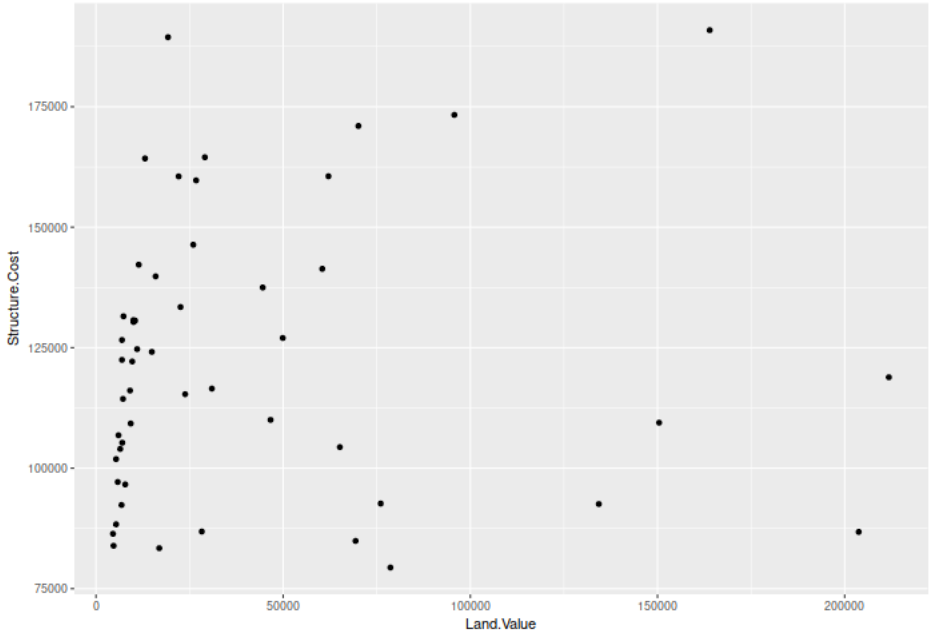
Points (Scatterplot)

点（散点图）

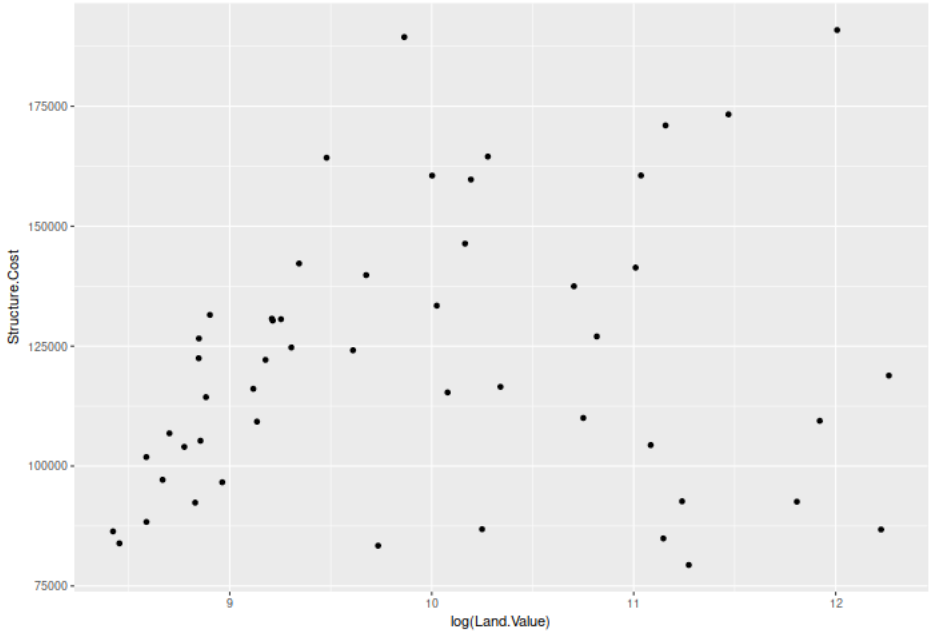
Now that we know about geometric objects and aesthetic mapping, we can make a ggplot. `geom_point` requires mappings for x and y, all others are optional.

既然我们知道了几何物体和美学映射，我们可以画一个 ggplot。geom_point 需要 x 和 y 的映射，所有的其他是可选的。

```
hp2001Q1 <- subset(housing, Date == 2001.25)
ggplot(hp2001Q1,
       aes(y = Structure.Cost, x = Land.Value)) +
  geom_point()
```




```
ggplot(hp2001Q1,
       aes(y = Structure.Cost, x = log(Land.Value))) +
  geom_point()
```



Lines (Prediction Line)

线（预测线）

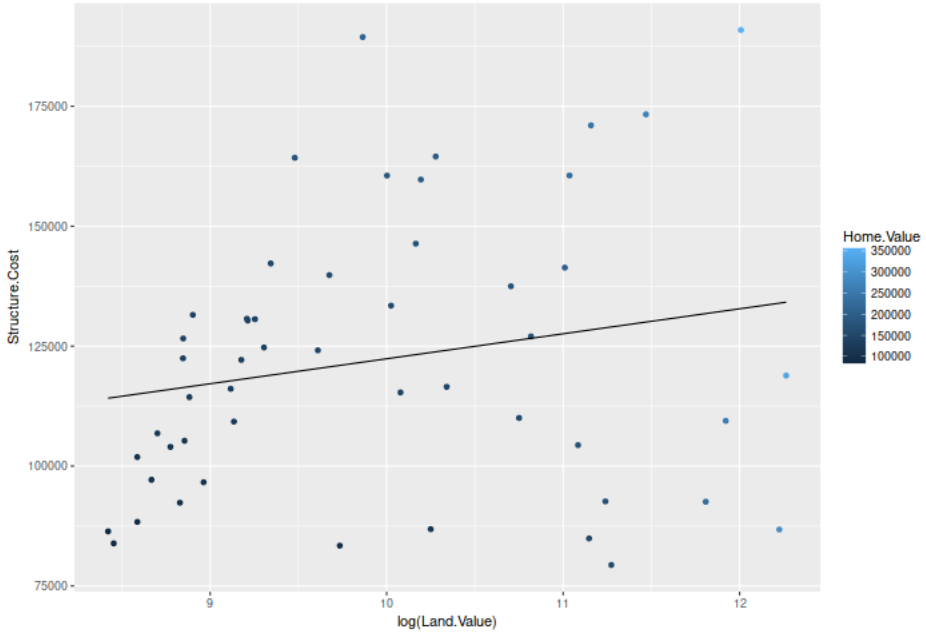
A plot constructed with `ggplot` can have more than one geom. In that case the mappings established in the `ggplot()` call are plot defaults that can be added to or overridden. Our plot could use a regression line:

一个用 ggplot 构建的图可以有不只一种几何物体。在那种映射用 ggplot（）调用建立图像情况下，图像预设值是可以被添加或者重写的。我们的图像可以用一条回归线：

```
hp2001Q1$pred.SC <- predict(lm(Structure.Cost ~ log(Land.Val
ue), data = hp2001Q1))

p1 <- ggplot(hp2001Q1, aes(x = log(Land.Value), y = Structur
e.Cost))

p1 + geom_point(aes(color = Home.Value)) +
  geom_line(aes(y = pred.SC))
```



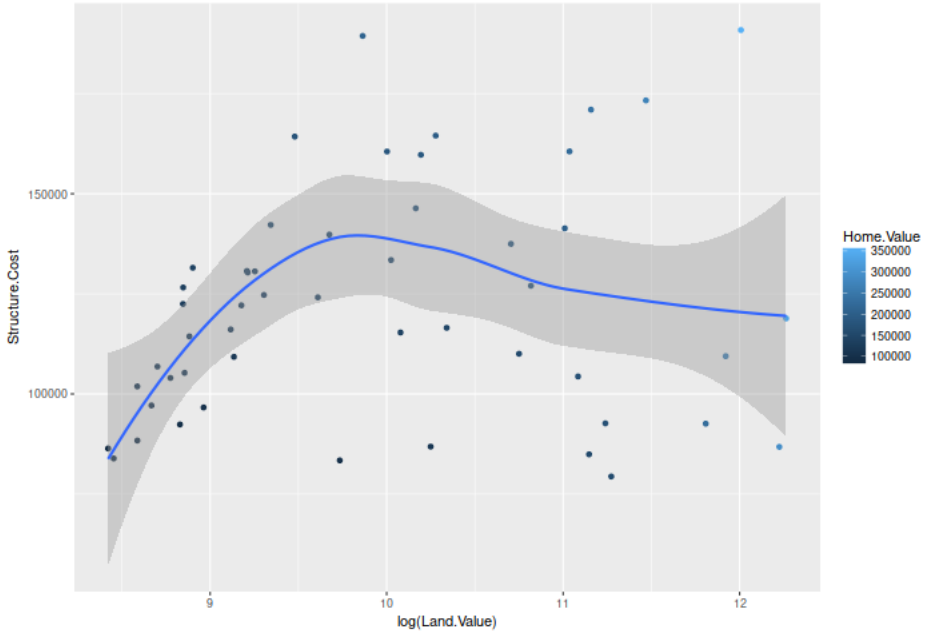
Smoothers

光滑的物体

Not all geometric objects are simple shapes - the smooth geom includes a line and a ribbon.

并非所有的几何物体都是简单的形状-光滑的几何物体包括了一条线和一条带。

```
p1 +
  geom_point(aes(color = Home.Value)) +
  geom_smooth()
```



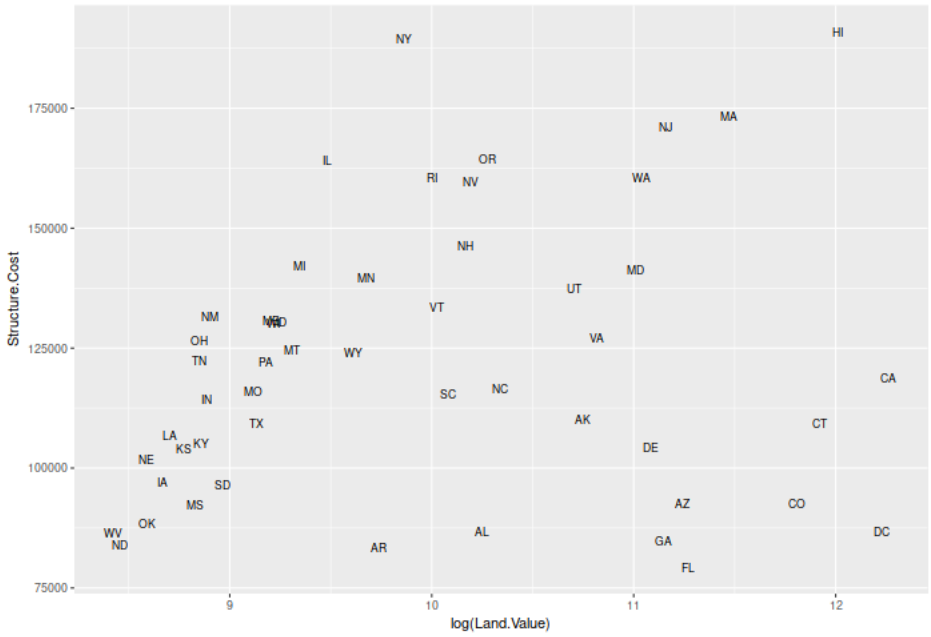
Text (Label Points)

文本（标记点）

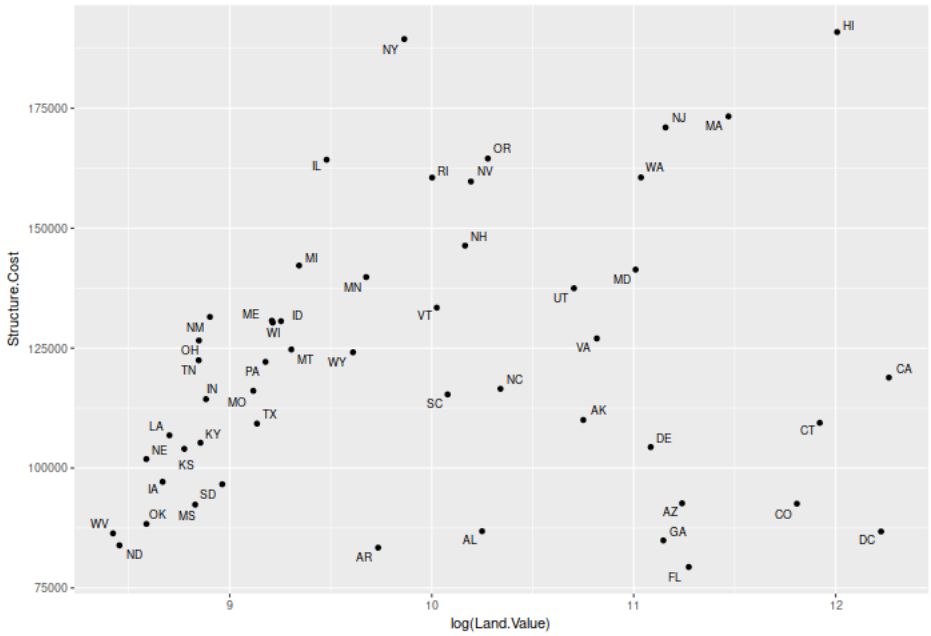
Each `geom` accepts a particular set of mappings - for example `geom_text()` accepts a `labels` mapping.

每一个几何物体可以接受一个特定集合的映射-比如 geom_text()接受一个标记映射。

```
p1 +
  geom_text(aes(label=State), size = 3)
```




```
## install.packages("ggrepel")
library("ggrepel")
p1 +
  geom_point() +
  geom_text_repel(aes(label=State), size = 3)
```



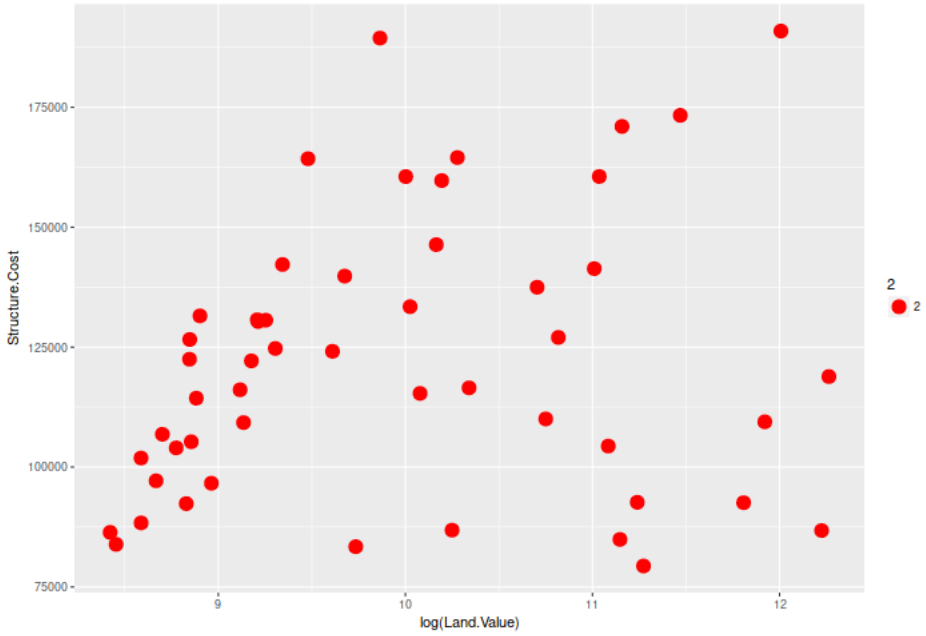
Aesthetic Mapping VS Assignment

美学映射和赋值的对比

Note that variables are mapped to aesthetics with the `aes()` function, while fixed aesthetics are set outside the `aes()` call. This sometimes leads to confusion, as in this example:

注意到变量被用 aes()函数映射到美学特征。然而固定的美学特征是在 aes()调用外被设定的。这有时候会造成困惑。看下这个例子：

```
p1 +
  geom_point(aes(size = 2),# incorrect! 2 is not a variable
             color="red") # this is fine -- all points red
```



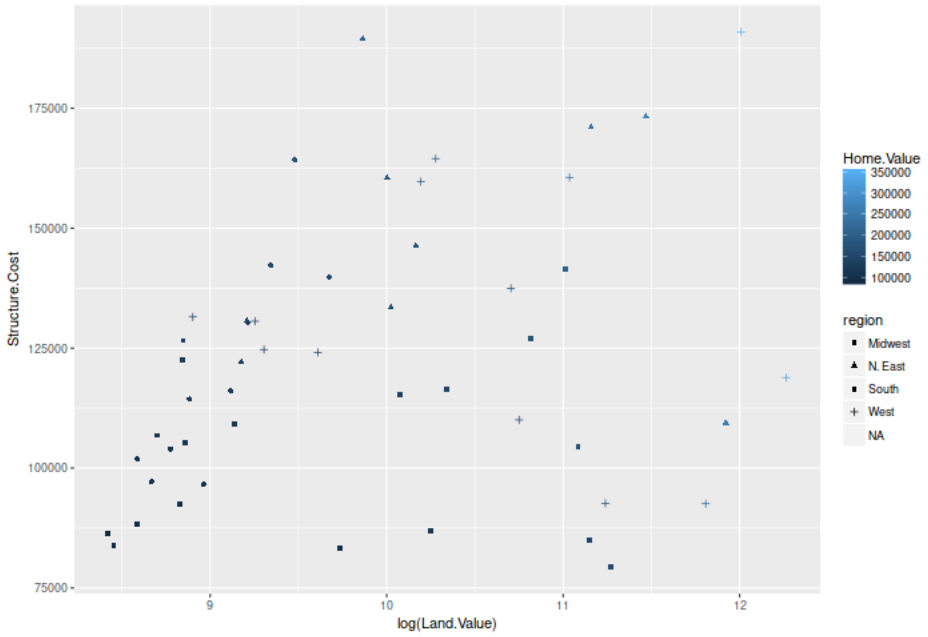
Mapping Variables To Other Aesthetics

映射变量到其他美学特征

Other aesthetics are mapped in the same way as x and y in the previous example.

其他美学特种被用和上面 x 和 y 一样的方法映射。

```
p1 +
  geom_point(aes(color=Home.Value, shape = region))
```



Exercise I

练习 1

The data for the exercises is available in the `dataSets/EconomistData.csv` file. Read it in with

这个练习的数据在 `dataSets/EconomistData.csv` 文件中。打开它通过：

```
dat <- read.csv("dataSets/EconomistData.csv")
head(dat)

ggplot(dat, aes(x = CPI, y = HDI, size = HDI.Rank)) + geom_po
int()
```

```
  X     Country HDI.Rank   HDI CPI           Region
1 1 Afghanistan      172 0.398 1.5      Asia Pacific
2 2     Albania       70 0.739 3.1 East EU Cemt Asia
3 3     Algeria       96 0.698 2.9             MENA
4 4      Angola      148 0.486 2.0              SSA
5 5   Argentina       45 0.797 3.0         Americas
```

```
6 6    Armenia      86 0.716 2.6 East EU Cemt Asia
```

Original sources for these data are http://www.transparency.org/content/download/64476/1031428http://hdrstats.undp.org/en/indicators/display_cf_xls_indicator.cfm?indicator_id=103106&lang=en

数据源是 http://www.transparency.org/content/download/64476/1031428http://hdrstats.undp.org/en/indicators/display_cf_xls_indicator.cfm?indicator_id=103106&lang=en

These data consist of *Human Development Index* and *Corruption Perception Index* scores for several countries.

这些数据包含了一些国家的 *Human Development Index* 和 *Corruption Perception Index* 分数。

1. Create a scatter plot with CPI on the x axis and HDI on the y axis.创造一个散点图，CPI 在 x 轴，HDI 在 y 轴。
2. Color the points blue.把点涂成蓝色。
3. Map the color of the the points to Region. 映射点的颜色到区域。
4. Make the points bigger by setting size to 2 通过把大小设定到 2 将点变大
5. Map the size of the points to HDI.Rank 映射点的大小到 HDI.Rank

Exercise I prototype :prototype:

练习 1 原型：原型：

1. Create a scatter plot with CPI on the x axis and HDI on the y axis. 创造一个散点图，CPI 在 x 轴，HDI 在 y 轴。

```
ggplot(dat, aes(x = CPI, y = HDI)) +
  geom_point()
```

2. Color the points in the previous plot blue.把点涂成蓝色。

```
ggplot(dat, aes(x = CPI, y = HDI)) +
```

```
  geom_point(color = "blue")
```

3. Color the points in the previous plot according to *Region*. 映射点的颜色到区域。

```
ggplot(dat, aes(x = CPI, y = HDI)) +
  geom_point(aes(color = Region))
```

4. Make the points bigger by setting size to 2 通过把大小设定到 2 将点变大

```
ggplot(dat, aes(x = CPI, y = HDI)) +
  geom_point(aes(color = Region), size = 2)
```

5. Map the size of the points to HDI.Rank 映射点的大小到 HDI.Rank

```
ggplot(dat, aes(x = CPI, y = HDI)) +
geom_point(aes(color = Region, size =  HDI.Rank))
```