

## **Part I**

# **Whole Game**

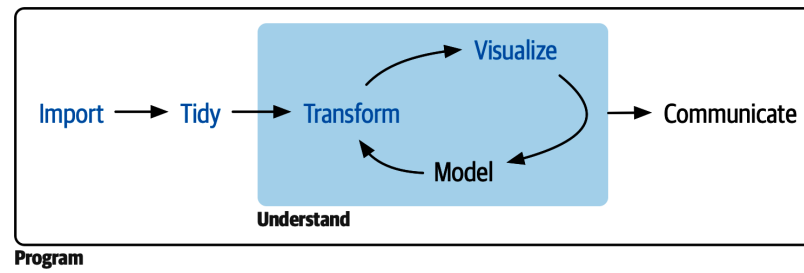


Figure 1: Here we will learn how to import, tidy, transform, and visualize data

Our goal in this part of the book is to give you a rapid overview of the main tools of data science: importing, tidying, transforming, and visualizing data, as shown in [Figure 1](#).

# Chapter 1

## Data Visualization

### 1.1

#### 1.1.1 数值型变量

数值变量分布的另一种可视化方法是密度图。密度图是直方图的平滑版本，也是一种实用的替代方案，特别是对于来自底层平滑分布的连续数据。它显示的细节比直方图少，但可以更轻松地快速收集分布的形状，特别是在众数和偏度方面。

### 1.2 可视化关系

#### 1.2.1 数值变量和分类变量

为了可视化数值变量和分类变量之间的关系，我们可以使用并排箱线图。箱线图是描述分布的位置（百分位数）度量的一种视觉速记形式。它对于识别潜在的异常值也很有用。

请注意我们在这里使用的术语：

- 如果我们希望美学（aesthetic）所代表的视觉属性根据该变量的值而变化，那么我们将变量映射到 `aesthetic`。
- 否则，我们就设置 `aesthetic` 值

#### 1.2.2 两个分类变量

我们可以使用堆积条形图来可视化两个分类变量之间的关系。

### 1.2.3 两个数值变量

可视化两个数值变量之间关系有散点图（使用创建 `geom_point()`）和平滑曲线（使用 `geom_smooth()` 创建）。散点图可能是最常用的用于可视化两个数值变量之间关系的图。

### 1.2.4 三个或更多变量

我们可以通过将更多变量映射到额外的美学效果来将它们合并到图中。然而，在图中添加太多美学映射会使其变得混乱且难以理解。另一种对于分类变量特别有用的方法是将图分割为分面（**facet**），即子图，每个子图显示数据的一个子集。

?? 中，你将了解许多其他几何图形，用于可视化变量的分布及其之间的关系。

## 1.3 保存绘图

绘制完绘图后，您可能希望将其保存为可在其他地方使用的图像，从而将其从 R 中取出。这就是 `ggsave()` 的工作，它将最近创建的绘图保存到磁盘。