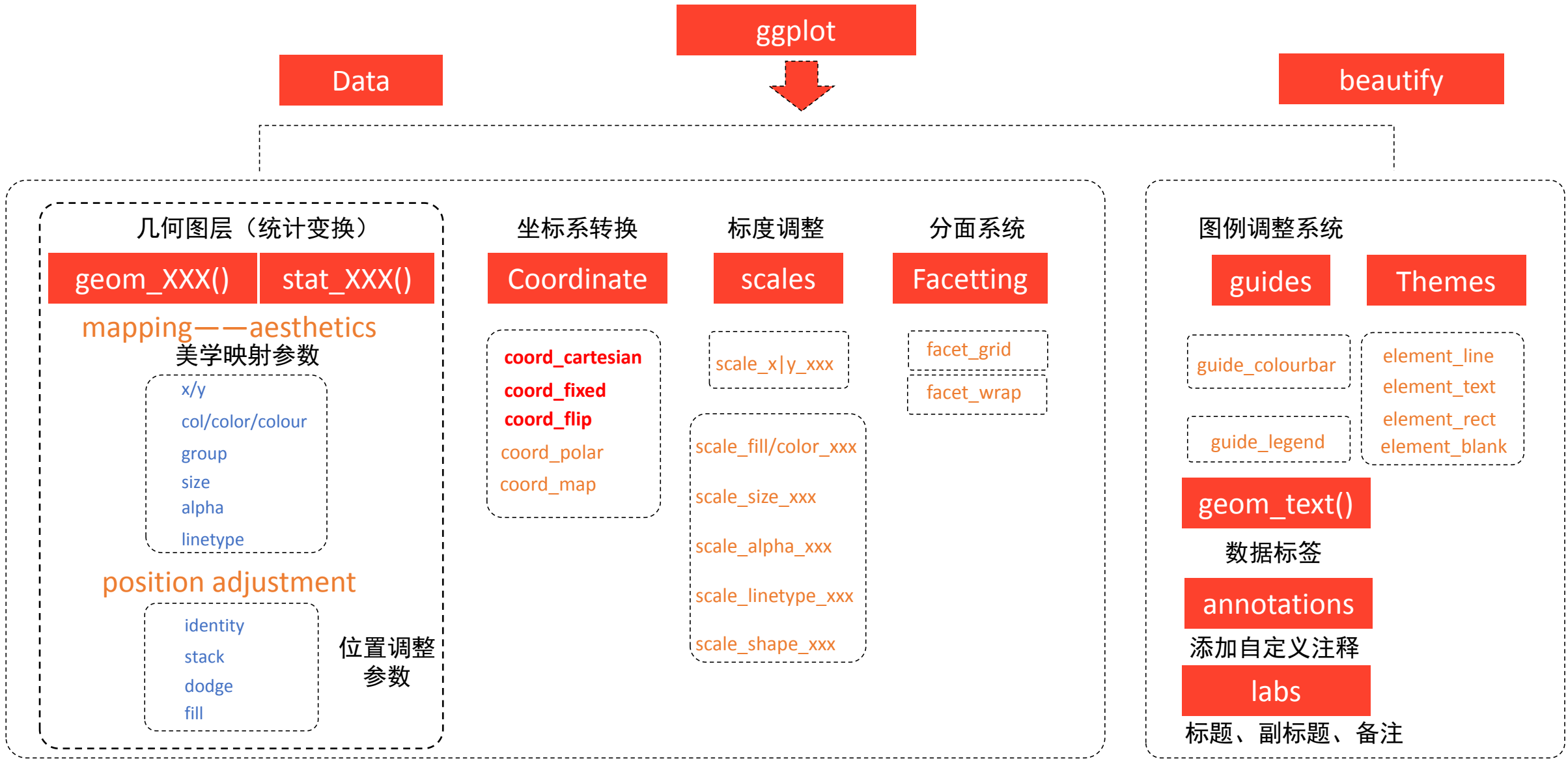


《R语言商务图表与数据可视化》

第六章：ggplot2笛卡尔坐标系应用

ggplot2图层语法框架概览



本章目录



6.1 笛卡尔坐标系主要图表类型

6.2 柱形图（条形图）数据描述



6.5 图形数据标签位置调整规则

6.6 图形颜色映射规则与因子变量的意义



6.3 美学映射参数书写规则

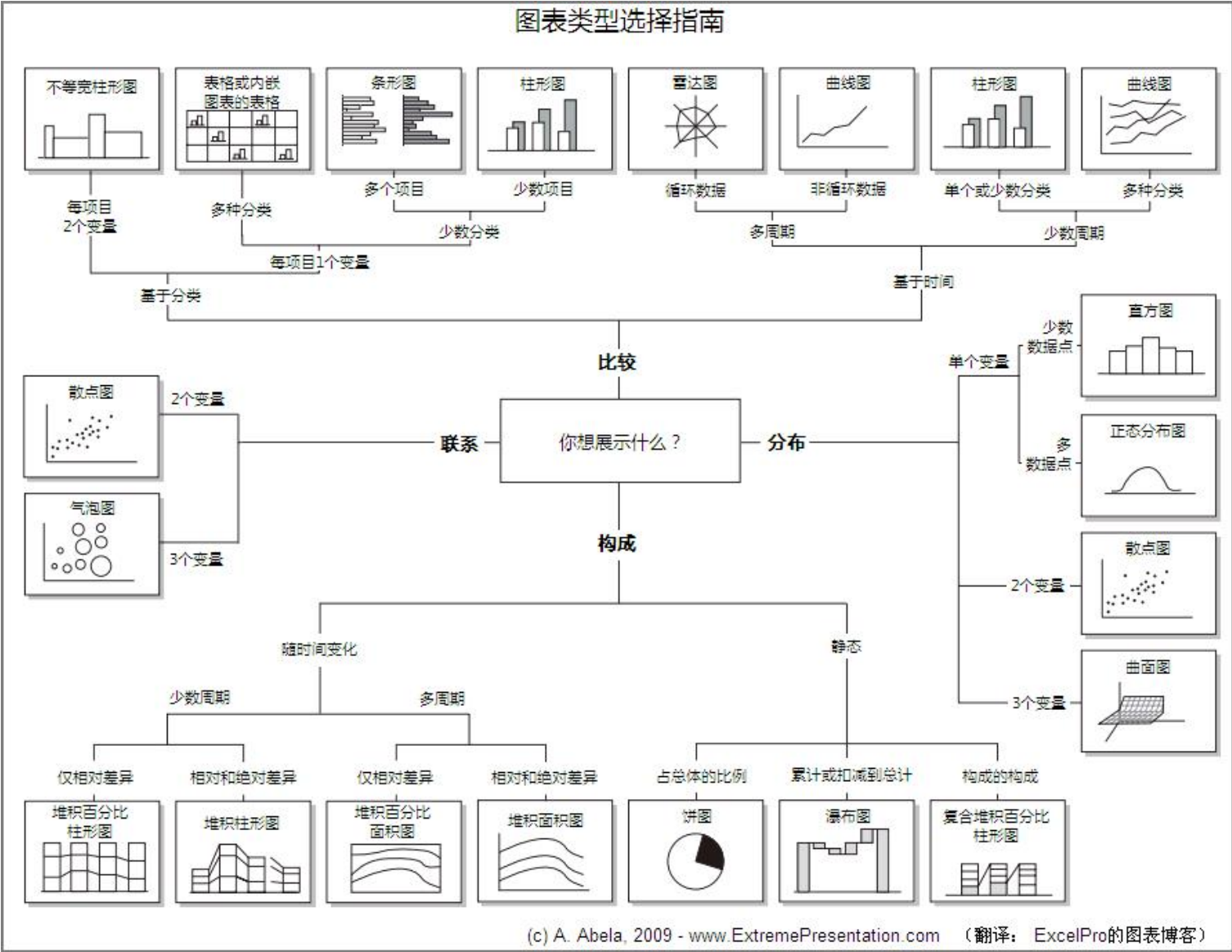
6.4 位置调整参数应用规则



6.7 笛卡尔坐标系下的分面应用

6.8 散点图（气泡图）/折线图（面积图）/箱线图/直方图案例

6.1 笛卡尔坐标系主要图表类



(c) A. Abela, 2009 - www.ExtremePresentation.com (翻译: ExcelPro的图表博客)

比较

- 柱形图
- 雷达图
- 折线图

维度 (类别变量)

分布

- 直方图
- 箱线图

度量 (数值变量)

构成

- 饼图
- 百分比堆积图

维度 (类别变量)

联系 (趋势)

- 散点图
- 气泡图

度量 (数值变量)

6.1 笛卡尔坐标系主要图表类

型表达的数据类型	坐标系划分		
	笛卡尔坐标系	极坐标系	地理信息坐标系
连续型变量表达（趋势、分布）	<ul style="list-style-type: none">• 散点图• 气泡图• 折线图• 面积图• 直方图• 箱线图	<ul style="list-style-type: none">• 日历圆环图	<ul style="list-style-type: none">• 气泡地图• 连续渐变
离散型变量表达（比较、构成）	<ul style="list-style-type: none">• 柱形图（条形图）• 瀑布图• 甘特图• 漏斗图	<ul style="list-style-type: none">• 饼图• 圆环图• 雷达图• 玫瑰图	<ul style="list-style-type: none">• 离散填充

6.2 笛卡尔坐标系主要图表类

图表类型
类型

序列数量

柱形图（条形图）

单序列

多序列（原位）

多序列（簇状柱形图）

多序列（堆积柱形图）

多序列（堆积百分比柱形图）多序列（分面柱形图）

位置参数——position

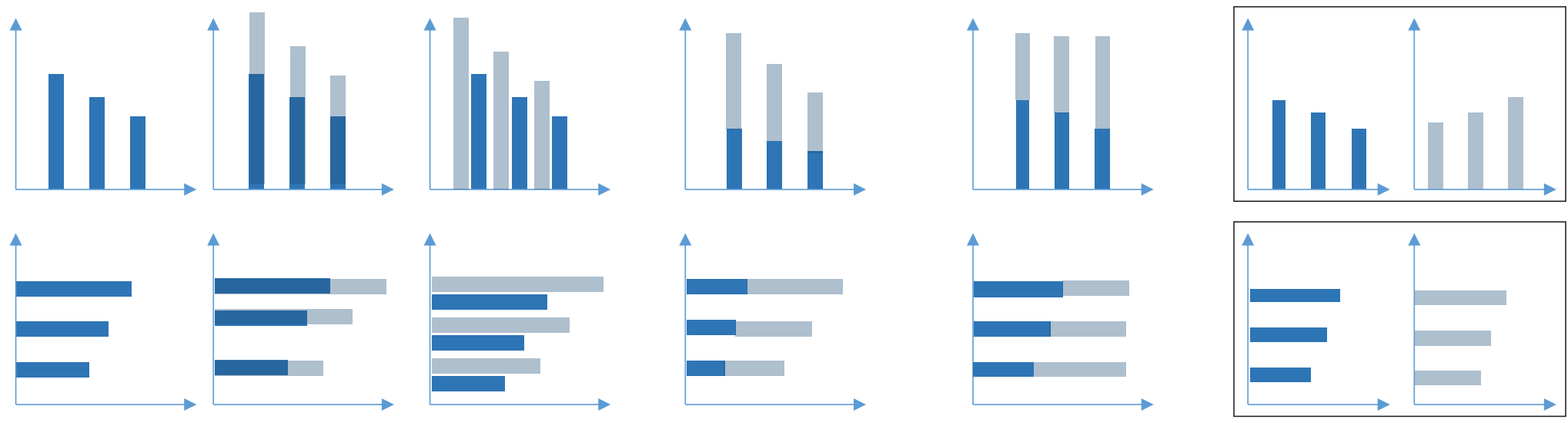
identity

dodge

stack

fill

facet_grid



面临的问题：

序列重叠遮盖

解决措施一：

错开摆放，优化布局

解决措施二：

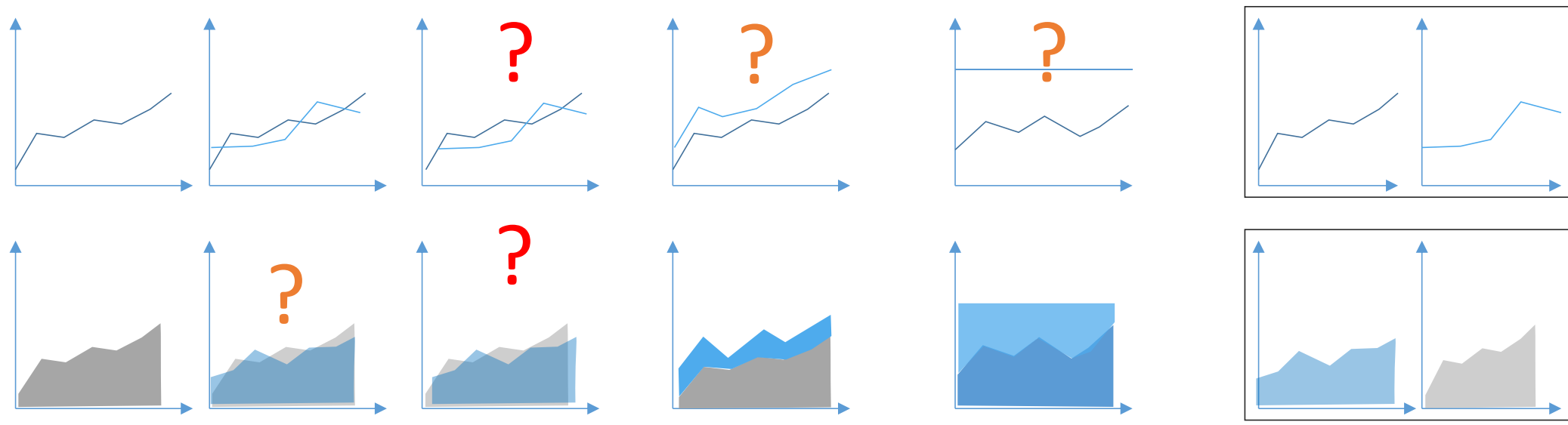
维度组合，子图排列

6.2 笛卡尔坐标系主要图表类

图表类型
折线图

序列数量

折线图	单序列	多序列（原位）	多序列（簇状折线图）	多序列（堆积折线图）	多序列（堆积百分比折线图）	多序列（分面折线图）
位置参数——	position	identity	dodge	stack	fill	facet_grid

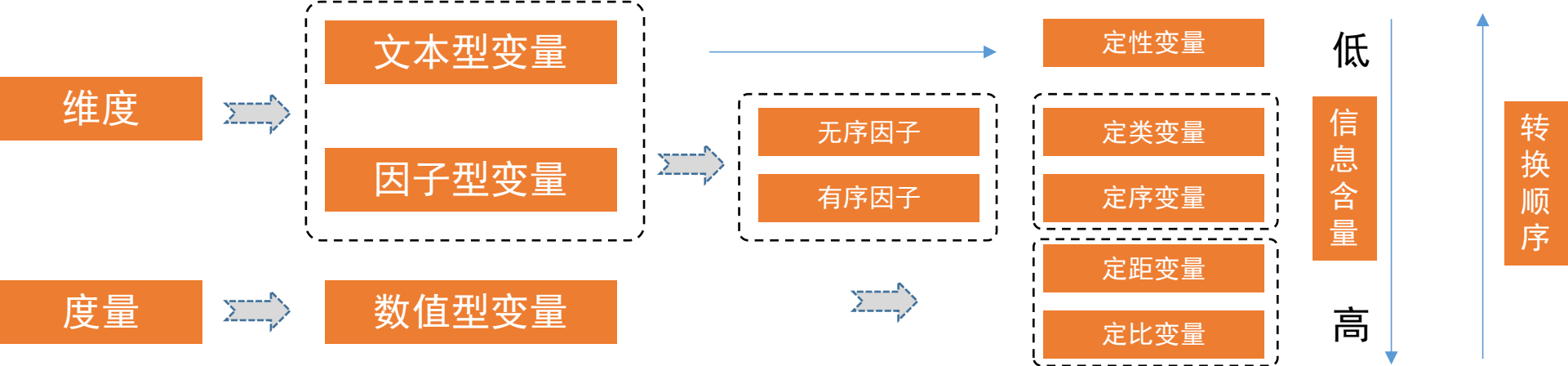


面临的问题：
序列重叠遮盖

解决措施一：
重叠摆放，优化布局

解决措施二：
维度组合，子图排列

6.2 柱形图（条形图）数据描述



6.2 柱形图（条形图）数据描述



↓ ↑
一个维度
一个度量

↓ ↑
二个维度
一个度量

6.3 美学映射参数书写规

关于ggplot与geom对象之间的关系问题：
则

- `ggplot(data=NULL, mapping = aes())`
- `geom_xxx(data=NULL, mapping = aes())`
- ggplot内有data、mapping两个参数，具有全局优先级，可以被之后的所有geom对象所继承（前提是geom内未指定相关参数）。
- geom对象内同样有data和mapping参数，但geom内的data和mapping参数属于局部参数，仅作用于geom对象内部。

所有图层共享数据源和美学映射参数：

所有图层仅共享数据源：

各图层对象均使用独立的数据源与美学映射参数：

- 所有geom对象都使用相同的data和mapping (x,y,size,alpha,linetype,colour,fill),根据参数继承规则，可以data和mapping指定在ggplot函数内，无论之后有多少个图层需要指定data和mapping，你都仅需在ggplot内指定一次即可，后续geom会自动继承。

- 此种情况，根据参数继承规则，将共享的数据源部分data写在ggplot内，将不同图层单独使用的美学映射参数指定在各自的geom内，在遇多图层时，data参数仅需在ggplot内指定一次，之后的geom对象都会自动继承，无需一一指定，但是那些geom内部使用的各自美学映射属性则需按需一一指定。

- 此种情况属于极端情况，仅在涉及到高级制图或者复杂地理信息多图层图表时才会接触，此时因为各图层没有共享任何data和mapping，假设有N个图层需要映射，此时所有的data和mapping参数都需要在各自geom内进行一一制定，因为在geom内指定毫无意义。

6.3 美学映射参数书写规则

则

共性映射与个性映射

- 当我们指定的美学映射参数需要进行个性化映射时（即一一映射），应该写在aes函数内部，即每一个观测值都会按照我们指定的特定变量值进行个性化设定。

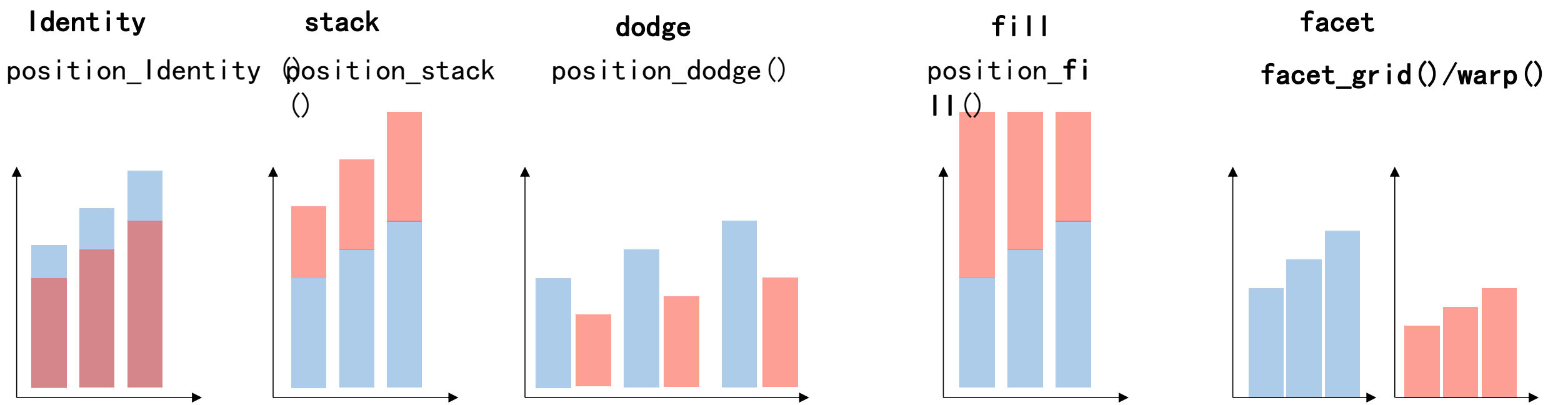
典型情况是需要添加一个维度，将这个维度按照以颜色、大小、线条等方式针对维度向量中每一个记录值进行一一设定

- 当我们需要统一设定某些图表元素对象（共性，统一化）时，此时应该将其参数指定在aes函数外部，即所有观测值都会按照统一属性进行映射，例如size=5, linetype="dash", color= "blue"

典型情况是需要统一所有的点大小、颜色、形状、透明度或者线条颜色、粗细、形状等。这种情况下不会消耗我们数据源中的任何一个维度或者度量指标，仅仅是对已经呈现出来的图形图形元素的外观属性做了统一设定。

6.4 位置调整参数应用规则

- Identity:** 不做任何位置调整，该情况在多分类柱形图中不可行，序列间会遮盖，但是在多序列散点图、折线图即可行，不存在遮盖问题。
- Stack:** 垂直堆叠放置。（堆积柱形图）
- Dodge:** 水平抖动放置（簇状柱形图）
- Fill:** 百分比化（垂直堆叠放置，如百分比堆积面积图，百分比堆积柱形图等）



6.5 图表数据标签位置调整规则

图表数据标签——geom_text()

- geom_text() 函数与其他geom_xxx开头的函数一样都是几何对象，因而其内部参数设置与其他几何对象一致，只是在对应点位置展示数据标签。
- geom_text() 函数可以充分体现美学映射参数的继承关系，一般而言x轴、y轴作为图表几何对象和数据标签几何对象共同的参数，最好写在ggplot() 函数里，而label参数（数据标签独有的参数）写在geom_text() 函数内部。

geom_text()

ggplot2的数据标签系统实质上是一个带有美学映射的几何图层对象，他拥有所有几何对象所拥有的公共属性

geom_text
geom_label

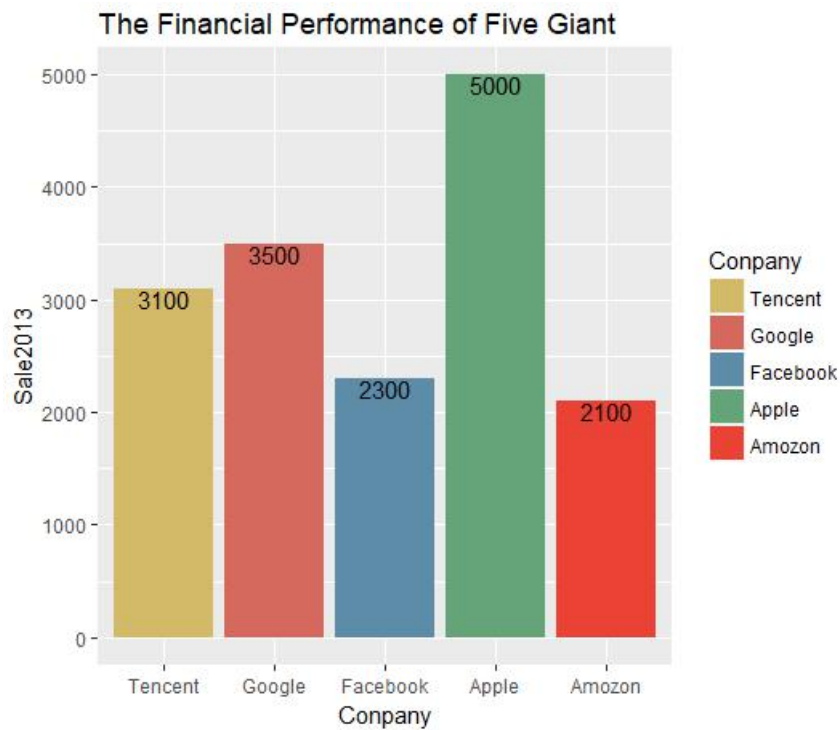
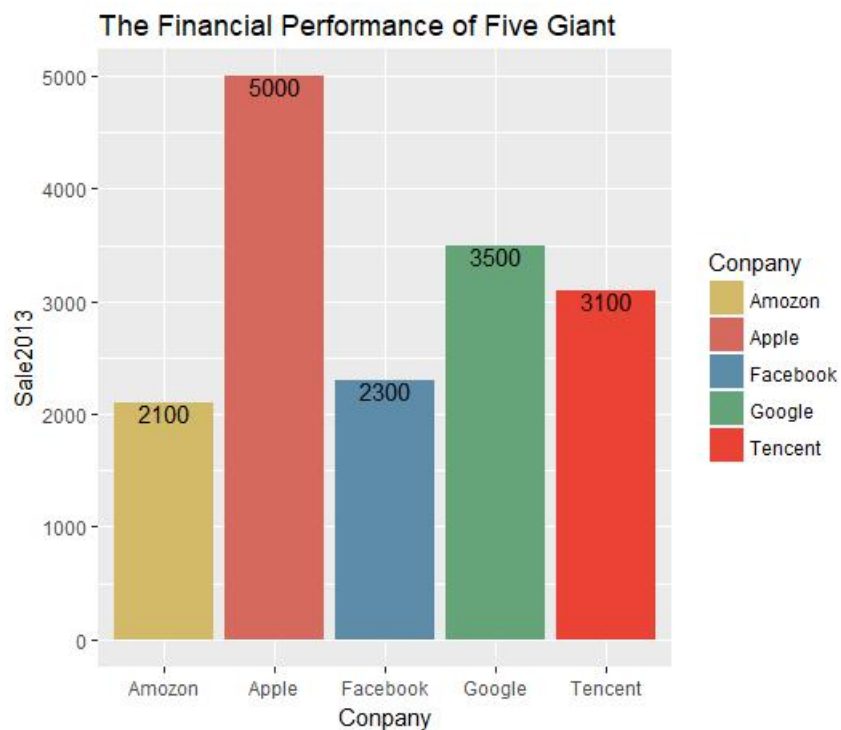
		mapping		
		required aesthetics	general	
geom_text(geom_label	data = NULL, mapping = NULL, stat = "identity", position = "identity"	x y label	alpha angle colour size group	family fontface hjust vjust lineheight

6.6 图形颜色映射规则与因子变量的意义

- 普通类别型变量在美学映射中顺序会被自动还原为英文字母顺序排列（图例中自上而下排列）
- 有序因子变量则按照因子大小顺序自上而下由大到小顺序呈现。

当你对分类变量的映射顺序很在意的时候，不妨将类别变量设定为有序因子变量，

这样可以提前锁定因子顺序，呈现出想要的映射结果，不用在标度调整函数中手动调整映射顺序。



6.7 笛卡尔坐标系下的分面应用

	facet_grid——网格分面	facet_wrap——封装分面
	facet_grid(. ~Year) —— 按行分面	仅限于一维分面
	facet_grid(Year~.) —— 按列分面	
facets——分面规则表达式	varname~. . ~ varname	~variable
Scales——坐标轴是否统一	<ul style="list-style-type: none">• free• fixed• free_x• free_y	<ul style="list-style-type: none">• free• fixed• free_x• free_y
		nrow

6.8 散点图（气泡图）、折线图（面积图）、箱线图、直方图图案

例

散点图（气泡图）

- 分组

折线图（面积图）

- 分面

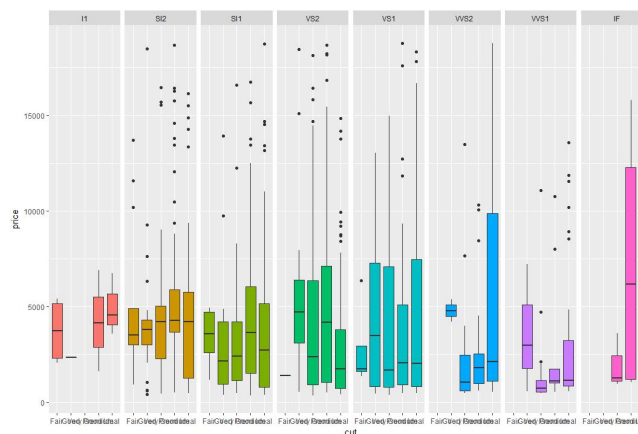
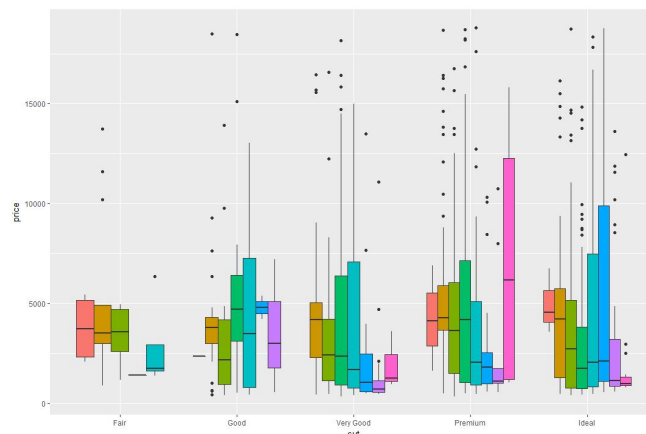
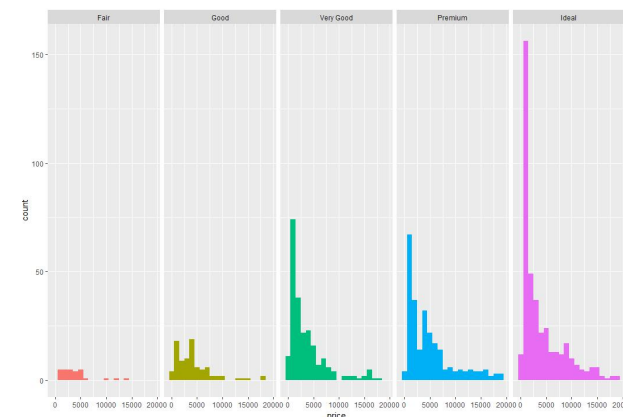
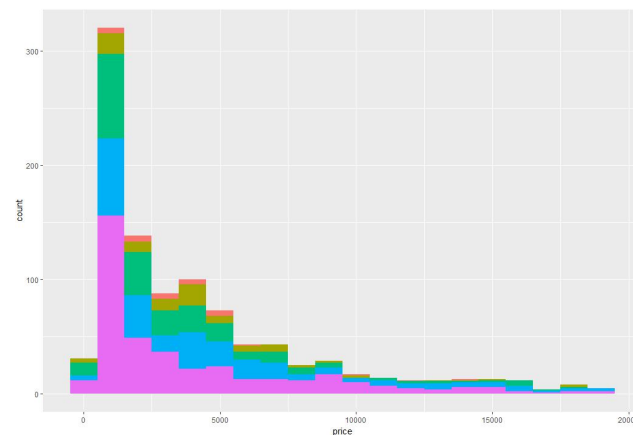
箱线图

- 数据标签

直方图

`geom_text()` 应用（涉及到参数继承规则）

分面表达式应用（纵向、横向、缠绕分面）





谢谢大家！