

[新西兰]哈德利·威克姆 (Hadley Wickham), [美]米内·切廷卡  
亚-郎德尔 (Mine Çetinkaya-Rundel), [美]加勒特·格罗勒芒德  
(Garrett Grolemund) 著  
张敬信 王小宁 黄俊文 译

# R 数据科学 (第 2 版) 链接资源

【致读者】本书包含的数百个链接资源，由 AI 辅助整理生成，以电子版 PDF 形式免费提供。由于网络环境不断变化，个别链接可能访问异常；相关描述文字也系 AI 生成，可能存在不精准之处。本资料仅供参考，敬请知悉。

2026 年 1 月



# 目录

前言 ..... 1

第 1 章 ..... 4

第 2 章 ..... 7

第 3 章 ..... 8

第 4 章 ..... 12

第 5 章 ..... 13

第 6 章 ..... 15

第 7 章 ..... 16

第 8 章 ..... 20

第 9 章 ..... 20

第 10 章 ..... 25

第 11 章 ..... 29

第 12 章 ..... 36

第 13 章 ..... 40

第 14 章 ..... 47

第 15 章 ..... 52

第 16 章 ..... 57

第 17 章 ..... 61

第 18 章 ..... 64

第 19 章 ..... 66

第 20 章 ..... 67

第 21 章 ..... 72

第 22 章 ..... 78

第 23 章 ..... 80

第 24 章 ..... 84

第 25 章 ..... 89

第 26 章 ..... 94

第 27 章 ..... 101

第 28 章 ..... 106

第 29 章 ..... 111



# 前言

1. <https://dplyr.tidyverse.org/reference/mutate.html>: 该资源是 dplyr 包中 mutate() 函数的官方文档页面, 介绍了 mutate() 函数的用法及功能, mutate() 用于在数据框中添加新的变量或修改已有变量。
2. <https://linkedin.com/company/oreilly-media>: 该资源是 O’ Reilly Media 公司在 LinkedIn 上的官方主页, 用于展示公司信息、发布动态以及与专业人士建立联系。
3. <https://oreil.ly/0giAa>: (<https://www.tidymodels.org>) 该资源是一个与 R 语言相关的建模工具包集合网站, tidymodels.org 提供了一整套用于数据建模和机器学习的 R 包, 旨在帮助用户以整洁 (tidy) 且系统化的方式进行建模流程, 适合学习和实践现代统计建模与机器学习方法。
4. <https://oreil.ly/8GLE7>: (<https://r4ds.hadley.nz>) 该资源是 Hadley Wickham 等人编写的《R for Data Science》一书的官方网站, 提供该书的在线版本和相关内容。网站包含书籍的章节内容, 结合了文本与可执行代码, 便于学习和实践数据科学中的 R 语言应用。
5. <https://oreil.ly/8uiH5>: (<https://learning.oreilly.com/library/view/hands-on-programming-with/9781449359089>) 该资源是 O’ Reilly 出版的《Hands-On Programming with R》一书的在线版本或介绍页面, 作者为 Garrett Grolemund。该书适合有一定数理基础和编程经验的读者, 特别适合作为初学者学习 R 语言编程的实践指导教材。
6. <https://oreil.ly/9Op9s>: (<https://learning.oreilly.com/library/view/tidy-modeling-with/9781492096474>) 该资源是由 Max Kuhn 和 Julia Silge 编写的电子书《Tidy Modeling with R》(出版于 O’ Reilly), 主要介绍如何使用 tidymodels 软件包进行数据建模。该书系统讲解了 tidymodels 家族工具的使用方法, 帮助读者掌握与 tidyverse 风格一致的现代 R 语言建模技术, 是学习数据科学中建模部分的重要参考资料。
7. <https://oreil.ly/GG4Et>: (<https://github.com/Rdatatable/data.table>) 该资源是一个名为 data.table 的 R 语言包的官方代码仓库, 提供高效处理大规模数据

（如 10-100 GB 级别）的工具和接口，具有显著的性能优势，适合需要快速数据操作和分析的用户使用。

8. <https://oreil.ly/HNIie>: (<http://ggplot2-book.org>) 该资源是一本专门介绍 ggplot2 数据可视化包的书籍，详细讲解了使用 ggplot2 进行数据可视化的工具和最佳实践，是学习和掌握 ggplot2 的权威参考资料。
9. [https://oreil.ly/Q8z\\_O](https://oreil.ly/Q8z_O): (<https://github.com/hadley/r4ds>) 该资源: <https://github.com/hadley/r4ds> 是《R for Data Science》一书的官方 GitHub 代码仓库，包含该书的源代码和内容，支持读者在线访问、学习和跟进书籍内容的最新更新。
10. [https://oreil.ly/\\_6LNH](https://oreil.ly/_6LNH): (<https://quarto.org>) 该资源是 Quarto 的官方网站，Quarto 是一种现代化的出版工具，支持将文本与可执行代码相结合，用于编写书籍、文章等多种文档，作为 R Markdown 的升级替代，体现了未来文档创作的趋势。
11. [https://oreil.ly/p3\\_RG](https://oreil.ly/p3_RG): (<https://cloud.r-project.org>) 该资源是 R 语言的综合归档网络（Comprehensive R Archive Network, CRAN）的官方镜像站点，用于下载和获取 R 语言的安装包及相关资源。
12. <https://oreil.ly/pRhEK>: (<https://docs.posit.co/ide/user>) 该 URL “<https://docs.posit.co/ide/user>” 指向的是 RStudio（现称 Posit）集成开发环境（IDE）的用户指南，提供关于 IDE 各项功能的详细介绍和使用说明，帮助用户全面了解和高效使用该开发环境。
13. <https://oreil.ly/pxF-k>: (<https://posit.co/download/rstudio-desktop>) 该资源是 RStudio 桌面版的官方下载页面，用户可以在此页面下载适用于本地计算机的 RStudio 集成开发环境（IDE）安装程序。
14. <https://oreil.ly/r-for-data-science-2e>: (<https://learning.oreilly.com/library/view/r-for-data/9781492097396>) 该 URL (<https://learning.oreilly.com/library/view/r-for-data/9781492097396>) 指向的是一本关于 R 语言数据科学的电子书资源，通常包含书籍全文、示例代码、附加资料以及相关学习内容，供读者在线阅读和学习使用。

15. <https://oreilly.com>: 该资源是 O’ Reilly Media 的官方网站, 提供技术和商业领域的培训、知识和洞见。网站包含丰富的图书、文章及在线学习平台, 支持按需访问直播课程、深入学习路径、互动编程环境, 以及来自 O’ Reilly 及其他 200 多家出版社的大量文本和视频内容, 致力于帮助企业和个人提升技能与竞争力。
16. <https://rdrr.io/pkg/nycflights13/man/flights.html>: 该资源是 R 语言包 `nycflights13` 中名为 `flights` 的数据集的官方文档页面, 提供该数据集的详细说明、字段信息及用法示例。
17. <https://rdrr.io/r/base/library.html>: 该资源是关于 R 语言中基础包 (`base package`) 及其函数的文档页面, 特别介绍了如何使用 `library()` 函数加载已安装的 R 包。
18. <https://rdrr.io/r/base/mean.html>: 该资源是 R 语言基础包 (`base package`) 中函数 `mean()` 的文档页面, 介绍了 `mean()` 函数的用法及相关说明。`mean()` 函数用于计算数值向量的算术平均值。
19. <https://rdrr.io/r/base/sum.html>: 该资源是 R 语言基础包中函数 `sum()` 的在线文档页面, 介绍了 `sum()` 函数的用法和功能。
20. [https://tidyverse.tidyverse.org/reference/tidyverse\\_update.html](https://tidyverse.tidyverse.org/reference/tidyverse_update.html): 该资源是 `tidyverse` 包中的一个函数文档页面, 介绍了函数 `tidyverse_update()` 的用法。该函数用于检查已安装的 `tidyverse` 相关包是否有可用的更新, 从而帮助用户保持 `tidyverse` 包的最新版本。
21. <https://twitter.com/oreillymedia>: 该资源是 O’ Reilly Media 在 Twitter 平台上的官方账号页面, 用于发布其相关的技术资讯、图书更新和行业动态。
22. <https://www.oreilly.com/about/contact.html>: 该资源是 O’ Reilly 网站的“联系方式”页面, 通常用于提供用户或访客与 O’ Reilly 公司取得联系的信息, 如客服支持、咨询渠道或反馈入口。
23. <https://www.youtube.com/oreillymedia>: 该资源是 O’ Reilly Media 在 YouTube 上的官方频道, 用户可以通过该频道观看与 O’ Reilly Media 相关的视频内容, 例如技术教程、培训课程、会议讲座等。

## 第 1 章

1. <https://allisonhorst.github.io/palmerpenguins/reference/penguins.html>: 该资源是 R 语言数据包 `palmerpenguins` 中 `penguins` 数据框的官方文档页面，详细介绍了 `penguins` 数据集的结构、变量说明以及数据来源。该数据集包含来自南极帕尔默站的 344 条企鹅观测记录，常用于数据可视化和数据分析教学。通过该页面，用户可以了解数据集中各个变量（如 `bill_depth_mm` 等）的含义和使用方法。
2. [https://forcats.tidyverse.org/reference/fct\\_inorder.html](https://forcats.tidyverse.org/reference/fct_inorder.html): 该资源是 tidyverse 中 `forcats` 包的函数文档，介绍了函数 `fct_inorder()`，该函数用于按因子在数据中出现的顺序来重新排序因子水平，方便对因子变量进行有序处理和分析。
3. <https://ggplot2.tidyverse.org/reference/aes.html>: 该资源是 `ggplot2` 包中用于定义图形映射关系的函数 `aes()` 的官方文档页面。它详细介绍了如何通过 `aes()` 函数将数据集中的变量映射到图形的视觉属性（如坐标轴、颜色、形状等），从而实现数据的美学映射（aesthetics mapping），是使用 `ggplot2` 绘图时指定变量与视觉表现对应关系的关键说明文档。
4. [https://ggplot2.tidyverse.org/reference/facet\\_wrap.html](https://ggplot2.tidyverse.org/reference/facet_wrap.html): 该资源是 `ggplot2` 包中用于绘图的函数 `facet_wrap()` 的官方文档页面。`facet_wrap()` 函数用于根据单个分类变量对图形进行分面（即将数据按该变量的不同类别拆分成多个子图），其第一个参数是一个公式（formula），表示按哪个变量进行分面。文档详细介绍了函数的用法和参数说明。
5. [https://ggplot2.tidyverse.org/reference/geom\\_bar.html](https://ggplot2.tidyverse.org/reference/geom_bar.html): 该资源是 `ggplot2` 包中用于绘制条形图的几何对象函数 `geom_bar()` 的官方文档页面，详细介绍了该函数的用法、参数及示例，帮助用户在数据可视化中创建和定制条形图。
6. [https://ggplot2.tidyverse.org/reference/geom\\_boxplot.html](https://ggplot2.tidyverse.org/reference/geom_boxplot.html): 该资源是 `ggplot2` 包中用于绘制箱线图（boxplot）的函数文档页面，详细介绍了函数 `geom_boxplot()` 的用法，包括如何通过该几何对象展示数据的分布情况和统计特



征，如中位数、四分位数及异常值等。该页面提供了函数的参数说明、示例代码及可视化效果，帮助用户理解并应用箱线图进行数据分析与展示。

7. [https://ggplot2.tidyverse.org/reference/geom\\_density.html](https://ggplot2.tidyverse.org/reference/geom_density.html): 该资源是 ggplot2 包中用于绘制密度图 (density plot) 的函数文档页面。页面详细介绍了函数 `geom_density()` 的用法和原理，说明该函数如何对数值型变量的分布进行平滑估计并绘制密度曲线，作为直方图的平滑替代，帮助用户更直观地观察数据的分布形态、峰值和偏态。
8. [https://ggplot2.tidyverse.org/reference/geom\\_histogram.html](https://ggplot2.tidyverse.org/reference/geom_histogram.html): 该资源是 ggplot2 包中函数 `geom_histogram()` 的官方文档页面，详细介绍了该函数的用法及参数说明，包括 `bins` 参数的作用，即控制直方图的分箱数量。
9. [https://ggplot2.tidyverse.org/reference/geom\\_path.html](https://ggplot2.tidyverse.org/reference/geom_path.html): 该资源是 ggplot2 包中用于绘制路径图 (line charts) 的几何对象 (geom) 函数文档。具体来说，网址 “[https://ggplot2.tidyverse.org/reference/geom\\_path.html](https://ggplot2.tidyverse.org/reference/geom_path.html)” 指向的是 `geom_path()` 函数的官方参考手册，介绍了该函数的用法、参数及示例，帮助用户在数据可视化中通过路径连接数据点来绘制线图。
10. [https://ggplot2.tidyverse.org/reference/geom\\_point.html](https://ggplot2.tidyverse.org/reference/geom_point.html): 该资源是 ggplot2 包中函数 `geom_point()` 的官方文档页面。`geom_point()` 用于向图形中添加点图层，从而创建散点图。该函数是 ggplot2 中众多用于绘制不同几何对象 (geom) 的函数之一，常用于表示数据中的数值点，适合展示两个数值变量之间的关系。文档中详细说明了 `geom_point()` 的功能、参数 (如 `na.rm`)，以及如何在图层中应用美学映射 (aesthetic mappings) 以实现点的颜色、形状等属性的定制。
11. [https://ggplot2.tidyverse.org/reference/geom\\_smooth.html](https://ggplot2.tidyverse.org/reference/geom_smooth.html): 该资源是 ggplot2 包中用于绘制平滑曲线的几何对象函数文档页面。具体来说，`geom_smooth()` 函数用于在散点图等图形上添加拟合的平滑曲线 (如回归线)，帮助可视化两个数值变量之间的趋势关系。文档中详细介绍了如何通过参数 (如 `method = "lm"`) 指定拟合模型类型 (例如线性模型) 来绘制最佳拟合线。
12. <https://ggplot2.tidyverse.org/reference/ggplot.html>: 该 URL (<https://ggplot2.tidyverse.org/reference/ggplot.html>) 指向的是 R 语言数据可视化包

- ggplot2 中核心函数 `ggplot()` 的官方文档页面。该页面详细介绍了 `ggplot()` 函数的用法，包括如何使用该函数开始创建绘图对象，如何通过设置 `data` 参数指定绘图所用的数据集，以及如何通过 `mapping` 参数定义变量与图形美学属性（如 x 轴、y 轴、颜色等）的映射关系。文档还说明了 `ggplot()` 作为绘图的基础画布，配合后续添加的图层函数，实现灵活且强大的数据可视化效果。
13. <https://ggplot2.tidyverse.org/reference/ggsave.html>: 该资源是 ggplot2 包中函数 `ggsave()` 的官方文档页面，介绍了如何使用 `ggsave()` 函数将 R 中最近创建的绘图保存为图像文件（如 PNG、PDF 等），包括设置图像尺寸（宽度和高度）的方法及支持的文件格式等详细说明。
  14. <https://ggplot2.tidyverse.org/reference/labs.html>: 该 URL (<https://ggplot2.tidyverse.org/reference/labs.html>) 指向的是 R 语言数据可视化包 ggplot2 中函数 `labs()` 的官方文档页面。该函数用于为图形添加或修改各种标签，包括图表标题（title）、副标题（subtitle）、坐标轴标签（x、y）以及图例标签（如 color、shape 等），从而提升图表的可读性和美观度。该文档详细说明了 `labs()` 函数的参数用法和示例，帮助用户更方便地自定义图表标签。
  15. <https://ggplot2.tidyverse.org/reference/mpg.html>: 该资源是 ggplot2 包中内置数据集 `mpg` 的官方文档页面，详细介绍了该数据框的结构、变量说明及数据来源。页面包含该数据集的描述、变量的类别（分类变量与数值变量）、观测数量（234 条）、涉及的汽车模型数量（38 款）以及如何在 R 语言环境中调用和使用该数据集的说明。
  16. <https://oreil.ly/01bKz>: (<https://conflicted.r-lib.org>) 该资源是“conflicted”包的官方网站，提供关于 R 语言中“conflicted”包的详细介绍、使用说明及冲突解决相关的信息。
  17. <https://oreil.ly/nwc5>: (<https://allisonhorst.github.io/palmerpenguins>) 该资源是一个名为“palmerpenguins”的 R 语言数据包，包含来自南极巴尔默群岛的企鹅数据，常用于生态学和数据科学的教学与研究。网址 <https://allisonhorst.github.io/palmerpenguins> 提供该数据包的详细介绍、使用说明和示例。

18. <https://pillar.r-lib.org/reference/glimpse.html>: 该资源是 R 语言中 pillar 包的官方文档页面, 介绍了函数 `glimpse()` 的用法。glimpse() 用于以紧凑且易读的格式快速查看数据框的所有变量及其前几条观测值, 方便用户对数据结构进行初步了解。
19. <https://rdr.io/pkg/ggthemes/man/colorblind.html>: 该 URL (<https://rdr.io/pkg/ggthemes/man/colorblind.html>) 指向的是 R 语言中 ggthemes 包的官方文档页面, 具体介绍了函数 `scale_color_colorblind()`。该函数用于在 ggplot2 绘图中应用色盲友好的颜色调色板, 从而提升图表的可访问性和易读性, 特别是对色觉障碍用户更为友好。
20. <https://rdr.io/r/base/library.html>: 该 URL “<https://rdr.io/r/base/library.html>” 指向的是 R 语言中基础包 (base package) 里的 `library()` 函数的文档页面。该页面详细介绍了如何使用 `library()` 函数来加载已安装的 R 包, 是 R 语言中管理和调用包的重要工具。

## 第 2 章

1. <https://rdr.io/r/base/c.html>: 该资源是 R 语言基础包中关于函数 `c()` 的官方文档页面, 介绍了如何使用 `c()` 函数将多个元素组合成一个向量。
2. <https://rdr.io/r/base/seq.html>: 该资源是 R 语言基础包 (base package) 中函数 `seq()` 的官方文档页面, 介绍了 `seq()` 函数的用法及参数说明。`seq()` 函数用于生成规则的数值序列, 页面提供函数的详细功能描述、参数列表以及示例, 帮助用户理解如何在 R 语言中生成连续或间隔自定义的数字序列。

## 第 3 章

1. <https://dplyr.tidyverse.org/reference/arrange.html>: 该资源是 dplyr 包中用于数据框行排序的函数 `arrange()` 的官方文档页面。它详细介绍了 `arrange()` 函数的用途和用法, 说明该函数通过指定一个或多个列 (或复杂表达式) 对数据框的行进行重新排序, 改变行的顺序但不改变行的内容。文档还提到了如何结合 `desc()` 函数实现降序排列, 以及与其他常用行操作函数 (如 `filter()`、`distinct()`) 的区别和配合使用方法。该页面适合需要对数据行进行排序和重新组织的 R 语言数据处理用户参考学习。
2. <https://dplyr.tidyverse.org/reference/context.html>: 该资源是 dplyr 包中用于数据汇总的辅助函数说明页面, 主要介绍函数 `n()` 的用法。`n()` 用于返回当前分组中的行数, 常用于聚合操作中计算样本量, 以确保基于足够数据量进行统计推断。页面详细说明了在调用 `summarize()` 时如何利用 `n()` 来获取每组数据的计数, 是数据变换和汇总时的重要工具函数文档。
3. <https://dplyr.tidyverse.org/reference/count.html>: 该资源是 dplyr 包中函数 `count()` 的官方文档页面, 介绍了 `count()` 函数的用法和功能。`count()` 用于统计数据框中某些变量的出现次数, 并可以通过参数 `sort = TRUE` 将结果按出现次数降序排列, 方便对数据频率进行分析和排序。页面还可能详细解释了 `count()` 函数背后所使用的 dplyr 动词操作。
4. <https://dplyr.tidyverse.org/reference/desc.html>: 该资源是 dplyr 包中函数 `desc()` 的官方文档页面, 介绍了如何在 `arrange()` 函数中使用 `desc()` 对数据框的指定列进行降序排序 (从大到小排列)。
5. <https://dplyr.tidyverse.org/reference/distinct.html>: 该资源是 dplyr 包中用于数据处理的函数文档, 介绍了 `distinct()` 函数的用法。`distinct()` 函数用于从数据集中提取唯一 (不重复) 的行, 主要针对行操作, 也可以指定某些列来获取这些列的唯一组合。该函数会返回每个唯一值的第一条记录, 常用于去重操作。与之相对的是 `count()` 函数, 用于统计各个唯一值的出现次数。此文档详细说明了 `distinct()` 的功能、使用场景及相关示例。

6. <https://dplyr.tidyverse.org/reference/filter.html>: 该资源是 dplyr 包中函数 `filter()` 的官方文档页面。`filter()` 函数用于根据列的条件筛选数据框中的行，保留满足条件的行且不改变行的顺序，也不修改列。该函数是数据变换中处理行的核心操作之一，常用于筛选符合特定条件的子集数据。文档详细介绍了 `filter()` 的用法、参数及示例说明，并指出该函数执行时不会修改原始数据集，而是返回一个新的数据框。
7. [https://dplyr.tidyverse.org/reference/group\\_by.html](https://dplyr.tidyverse.org/reference/group_by.html): 该资源是 dplyr 包中函数 `group_by()` 的官方文档页面。它详细介绍了 `group_by()` 函数的用法和功能，说明该函数用于将数据集按指定变量划分为多个组，从而使后续的数据操作（如汇总、筛选等）能够基于这些分组进行。文档还解释了 `group_by()` 不会改变原始数据，但会给数据框添加分组的属性（class），从而影响后续 dplyr 语句的行为。此外，文档提及了与 `group_by()` 相关的函数如 `ungroup()` 用于取消分组，以及 dplyr 1.1.0 引入的 `.by` 参数，用于在单次操作中实现分组，提供了更灵活的分组机制。总体而言，该页面是学习和理解 dplyr 中分组操作的核心参考资料。
8. <https://dplyr.tidyverse.org/reference/mutate.html>: 该资源是 dplyr 包中函数 `mutate()` 的官方文档页面，介绍了 `mutate()` 函数的用途和用法。`mutate()` 用于在数据框中基于已有列计算并添加新列，且不会改变数据的行数。文档详细说明了如何使用 `mutate()` 创建派生变量，控制新列的位置（如通过 `.before` 和 `.after` 参数），以及通过 `.keep` 参数控制保留哪些列。该函数是 dplyr 中用于列变换的核心工具之一，常用于数据处理和特征工程。
9. <https://dplyr.tidyverse.org/reference/relocate.html>: 该资源是 dplyr 包中用于数据框列操作的函数文档页面，具体介绍了函数 `relocate()` 的用法。`relocate()` 函数用于调整数据框中列的位置，可以将变量列重新排序，比如将相关的变量放在一起，或将重要的变量移动到前面，默认情况下会将指定的变量移动到数据框的最前面。该页面详细说明了 `relocate()` 的功能、用法和示例，帮助用户更灵活地组织和管理数据框的列顺序。
10. <https://dplyr.tidyverse.org/reference/rename.html>: 该资源是 dplyr 包中用于数据框列重命名的函数文档页面。它详细介绍了 `rename()` 函数的用法，说明如

何在保持所有现有列不变的情况下，修改部分列的名称。该页面属于 dplyr 的官方参考手册，提供函数的参数说明、示例及应用场景，帮助用户高效进行数据列的重命名操作。

11. <https://dplyr.tidyverse.org/reference/select.html>: 该资源是 dplyr 包中用于数据框列操作的函数 `select()` 的官方文档页面。它详细介绍了 `select()` 函数的用法，主要功能是选择数据框中的特定列，从而专注于感兴趣的变量。文档中还说明了 `select()` 支持多种基于变量名称的选择操作，包括使用辅助函数和重命名变量的语法。`select()` 是 dplyr 中四个常用的列操作动词之一，其他包括 `mutate()`（新建列）、`rename()`（重名列）和 `relocate()`（调整列位置）。该文档对于理解和掌握如何快速筛选和管理大型数据集中的列非常有帮助。
12. <https://dplyr.tidyverse.org/reference/slice.html>: 该资源是 dplyr 包中与数据切片相关的函数文档，主要介绍了函数如 `slice_min()` 和 `slice_max()` 的用法。这些函数用于根据某列的最小值或最大值对数据框进行行切片，支持处理并保留“并列”值（即多个行具有相同的最小或最大值）。文档还说明了参数如 `n`（指定要选取的行数）和 `with_ties`（是否保留并列值）的作用及使用方法。
13. <https://dplyr.tidyverse.org/reference/summarise.html>: 该资源是 dplyr 包中用于对分组数据进行汇总操作的函数文档页面，详细介绍了 `summarise()`（或 `summarize()`）函数的用法。该函数能够根据分组计算各种汇总统计量，将每个分组的数据缩减为单行，常用于计算均值、计数、最大值等汇总指标。文档中还展示了如何在一次调用中创建多个汇总，以及该函数在数据转换和分组操作中的重要作用。
14. <https://oreil.ly/OjOwY>: ([http://varianceexplained.org/r/empirical\\_bayes\\_baseball](http://varianceexplained.org/r/empirical_bayes_baseball)) 该 URL 指向的资源是 David Robinson 关于“经验贝叶斯方法在棒球数据分析中的应用”的博客文章，详细解释了相关统计问题及其解决方法。
15. <https://oreil.ly/wgS7U>: (<https://www.evanmiller.org/how-not-to-sort-by-average-rating.html>) 该资源是一篇由 Evan Miller 撰写的博客文章，详细解释了“按平均评分排序”这一方法存在的问题，并介绍了如何改进排序策略以获得更合理的结果。

- 
16. <https://oreil.ly/ySpmy>: (<https://www.tidyverse.org/blog/2023/02/dplyr-1-1-0-per-operation-grouping>) 该 URL 指向的是 tidyverse 官网发布的一篇博客文章, 介绍了 dplyr 包 1.1.0 版本的新特性, 特别是关于 “per-operation grouping” (每次操作分组) 功能的内容。文章详细讲解了这一新语法及其使用方法, 展示了该功能如何提升数据变换的灵活性和效率。
  17. <https://pillar.r-lib.org/reference/glimpse.html>: 该 URL (<https://pillar.r-lib.org/reference/glimpse.html>) 指向的是 R 语言中 `glimpse()` 函数的官方文档页面。`glimpse()` 是 `tibble` 包中的一个函数, 用于以紧凑、易读的格式快速查看数据框或 `tibble` 的结构和内容, 显示所有列的名称、类型及部分数据, 方便用户快速了解大型数据集的整体情况。
  18. [https://rdrr.io/pkg/janitor/man/clean\\_names.html](https://rdrr.io/pkg/janitor/man/clean_names.html): 该资源是 R 语言包 `janitor` 中函数 `clean_names()` 的在线文档, 介绍了如何使用该函数对数据框中的列名进行自动化清理和规范化, 以便处理命名不一致的问题。
  19. <https://rdrr.io/pkg/nycflights13/man/flights.html>: 该资源是 R 语言包 `nycflights13` 中 `flights` 数据集的官方文档页面, 详细介绍了 2013 年从纽约市出发的所有航班数据 (共 336,776 条记录), 数据来源于美国交通统计局, 并包含该数据集的结构说明和使用说明。
  20. <https://rdrr.io/r/base/mean.html>: 该资源是 R 语言中基础包 (`base` package) 里函数 `mean()` 的官方文档页面, 介绍了 `mean()` 函数的用法及参数说明。文档中包括如何计算数值的平均值, 以及如何通过设置参数如 `na.rm = TRUE` 来忽略缺失值 (NA) 进行平均值计算。
  21. <https://rdrr.io/r/stats/filter.html>: 该资源是 R 语言官方文档网站上的页面, 具体介绍了基础统计包 (`stats`) 中的函数 `filter()`。页面内容包括该函数的用法说明、参数解释以及示例代码, 帮助用户理解和正确使用 `stats::filter()` 函数。
  22. <https://rdrr.io/r/stats/lag.html>: 该资源是 R 语言中 `stats` 包中函数 `lag()` 的官方文档页面。该页面详细介绍了 `lag()` 函数的用途、用法及参数说明。`lag()` 函数通常用于生成时间序列数据的滞后值, 方便进行时间序列分析。页面还可能包含示例代码和相关函数链接, 帮助用户理解和正确使用该函数。

23. <https://rdr.io/r/Utils/View.html>: 该资源是 R 语言中 `utils` 包的 `View()` 函数的文档页面。`View()` 函数用于在 RStudio 等环境中以表格视图的形式查看数据集，方便用户浏览和检查包含多列的数据。
24. [https://tidyselect.r-lib.org/reference/all\\_of.html](https://tidyselect.r-lib.org/reference/all_of.html): 该资源是 R 语言中 `tidyselect` 包的官方文档页面，具体介绍了函数 `all_of()` 的用法和功能。该页面详细说明了 `all_of()` 如何在选择数据框列时，结合给定的列名向量，确保所选列完全匹配这些名称，是数据选择和处理中的常用工具。
25. [https://tidyselect.r-lib.org/reference/starts\\_with.html](https://tidyselect.r-lib.org/reference/starts_with.html): 该资源是 `tidyselect` 包中函数 `starts_with()` 的官方文档，介绍了如何使用 `starts_with()` 函数根据变量名前缀选择数据框中的变量，属于变量选择工具的参考说明。

## 第 4 章

1. <https://dplyr.tidyverse.org/reference/filter.html>: 该资源是 `dplyr` 包中函数 `filter()` 的官方文档页面，介绍了 `filter()` 函数的用法和参数说明。`filter()` 用于根据指定条件筛选数据框中的行。
2. <https://dplyr.tidyverse.org/reference/mutate.html>: 该资源是 `dplyr` 包中函数 `mutate()` 的官方文档页面，介绍了 `mutate()` 函数的用法。`mutate()` 用于在数据框中新增或修改变量，通常通过对现有变量进行计算生成新变量。文档中包含函数的参数说明、示例代码及使用建议，是学习和参考 `dplyr` 数据变换操作的重要资料。
3. <https://dplyr.tidyverse.org/reference/select.html>: 该资源是 `dplyr` 包中函数 `select()` 的官方文档页面，介绍了 `select()` 函数的用法及参数说明。`select()` 用于从数据框中选择指定的列，常用于数据操作和整理。
4. <https://dplyr.tidyverse.org/reference/summarise.html>: 该资源是 `dplyr` 包中函数 `summarise()`（或 `summarize()`）的官方文档页面，介绍了该函数的用法



和参数说明。`summarise()` 函数用于对数据框进行汇总计算，生成新的汇总变量，常用于数据整理和分析中。

5. [https://oreil.ly/8\\_Z1c](https://oreil.ly/8_Z1c): (<https://styler.r-lib.org>) 该资源是 “styler” 包的官方网站，提供用于 R 语言代码格式化和美化的工具与相关文档，帮助用户自动调整代码风格，使代码更加整洁、一致，提升代码可读性和维护性。
6. <https://oreil.ly/LykON>: (<https://style.tidyverse.org>) 该资源是 “tidyverse 风格指南” (tidyverse style guide)，提供了一套用于 R 语言代码编写的规范与最佳实践，旨在帮助程序员保持代码风格一致、易读易维护，特别适用于使用 tidyverse 生态系统的开发者。

## 第 5 章

1. <https://dplyr.tidyverse.org/reference/distinct.html>: 该资源是 dplyr 包中函数 `distinct()` 的官方文档页面，介绍了 `distinct()` 函数的用法及功能，该函数用于从数据框中提取唯一（不重复）的行或值，帮助用户快速查看和筛选数据中的不同项。
2. <https://dplyr.tidyverse.org/reference/mutate.html>: 该资源是 dplyr 包中 `mutate()` 函数的官方文档页面，介绍了 `mutate()` 函数的用法。`mutate()` 用于对数据框进行变换或新增变量，帮助用户基于已有数据创建新的列，从而便于后续的数据处理和计算。
3. <https://dplyr.tidyverse.org/reference/select.html>: 该资源是 dplyr 包中函数 `select()` 的官方文档页面，介绍了 `select()` 函数的用法及语法，主要用于选择数据框中的指定列。
4. <https://oreil.ly/86uxw>: (<https://www.jstatsoft.org/article/view/v059i10>) 该资源是一篇发表在《Journal of Statistical Software》上的学术论文，题为 “Tidy Data”，介绍了数据整理的理论基础和历史背景。

5. [https://readr.tidyverse.org/reference/parse\\_number.html](https://readr.tidyverse.org/reference/parse_number.html): 该资源是 R 语言中 readr 包的函数文档页面, 介绍了函数 parse\_number()。该函数用于从字符串中提取第一个数字, 忽略字符串中的其他文本, 方便将字符型数据转换为数值型, 从而简化后续的数据处理和计算。
6. <https://tibble.tidyverse.org/reference/tribble.html>: 该资源是关于 R 语言中 tidyverse 包中的函数 tribble() 的官方文档页面。tribble() 函数用于方便地手动创建小型的 tibble 数据框, 特别适用于构造示例数据集, 以便于演示和数据处理操作。
7. <https://tidyr.tidyverse.org/articles/pivot.html>: 该资源是 tidyr 包中的一篇详尽教程 (vignette), 介绍如何使用 pivot\_longer() 和 pivot\_wider() 等函数, 将数据转换为“tidy data” (整洁数据) 的格式。它提供了丰富的示例和应用场景, 帮助用户理解和解决数据整形过程中遇到的各种问题, 是学习 tidyr 数据重塑功能的重要参考文档。
8. [https://tidyr.tidyverse.org/reference/pivot\\_longer.html](https://tidyr.tidyverse.org/reference/pivot_longer.html): 该资源是 tidyr 包中 pivot\_longer() 函数的官方文档页面, 介绍了 pivot\_longer() 函数的用法及示例。pivot\_longer() 用于将数据从宽格式转换为长格式, 即将多个列名中的值“拉长”到行中, 从而实现数据整洁化。文档详细说明了该函数的参数, 如 names\_to、values\_to、values\_drop\_na 等, 展示了如何通过该函数处理不同类型的数据重塑需求, 是 tidyverse 数据整理中常用且重要的工具。
9. [https://tidyr.tidyverse.org/reference/pivot\\_wider.html](https://tidyr.tidyverse.org/reference/pivot_wider.html): 该 URL 指向的是 tidyr 包中函数 pivot\_wider() 的官方文档页面。pivot\_wider() 函数用于将数据从长格式转换为宽格式, 通过增加列数、减少行数, 将分散在多行中的同一观测值合并为单行, 从而使数据集“变宽”。该函数需要指定用于生成新列名的列 (names\_from) 和用于填充新列的值的列 (values\_from), 并且可以通过指定唯一标识每行的列来正确地整理数据。pivot\_wider() 常用于处理政府数据等特定场景, 帮助用户将不规则的“长”数据整理为更易于分析的“宽”数据格式。
10. <https://tidyverse.tidyverse.org>: 该资源是 tidyverse 的官方网站, 提供 R 语言中 tidyverse 包集合的文档、教程和相关信息。

## 第 6 章

1. <https://oreil.ly/L9ip0>: (<https://phdcomics.com/comics/archive.php?comid=1531>) 该 URL 指向的是漫画网站 Piled Higher and Deeper (PhD Comics) 中编号为 1531 的一则漫画条目。根据上下文, 这个漫画条目可能是一幅关于使用 “final” 命名带来 “命运捉弄” 或相关幽默主题的漫画。
2. <https://oreil.ly/coili>: (<https://support.posit.co/hc/en-us/articles/205753617-Code-Diagnostics>) 该资源是关于代码诊断 (Code Diagnostics) 的支持文章, 介绍了 RStudio 诊断工具能够检测和报告的常见代码错误和问题, 帮助用户更好地理解 and 修正代码中的潜在错误。
3. <https://rdr.io/r/base/getwd.html>: 该资源是 R 语言中基础包 (base) 中函数 `getwd()` 的在线文档页面, 介绍了如何使用 `getwd()` 函数获取当前工作目录的相关说明和示例代码。
4. <https://rdr.io/r/utils/install.packages.html>: 该资源是 R 语言中函数 `install.packages()` 的官方文档页面, 详细介绍了如何使用该函数来安装 R 包。文档通常包括函数的用法说明、参数解释以及示例代码, 帮助用户正确安装所需的 R 软件包。
5. <https://twitter.com/rstudiotips>: 该资源是 RStudio Tips 的官方 Twitter 账号, 专门发布与 RStudio 相关的实用技巧和建议, 帮助用户提升 R 语言编程和数据分析的效率。
6. [https://usethis.r-lib.org/reference/use\\_blank\\_slate.html](https://usethis.r-lib.org/reference/use_blank_slate.html): 该资源是 R 包 `usethis` 中的函数 `use_blank_slate()` 的文档页, 介绍如何通过该函数配置 RStudio, 使其在每次启动时不保留之前会话中的工作空间 (即不保存之前运行的代码、创建的对象或读取的数据)。这样可以确保分析过程完全依赖于脚本代码, 避免因依赖旧环境状态而造成的重复劳动或数据丢失风险, 从而促进更规范和可重复的分析工作流程。

## 第 7 章

1. [https://dplyr.tidyverse.org/reference/if\\_else.html](https://dplyr.tidyverse.org/reference/if_else.html): 该资源为 dplyr 包中函数 `if_else()` 的官方文档页面，详细介绍了 `if_else()` 函数的用法。`if_else()` 函数有三个参数，第一个参数是逻辑向量 `test`，当 `test` 为 `TRUE` 时返回第二个参数 `yes` 的值，为 `FALSE` 时返回第三个参数 `no` 的值。该函数常用于根据条件选择性赋值，是 dplyr 中用于条件判断和数据变换的重要工具。页面中还结合示例说明了如何利用 `if_else()` 进行条件替换操作，并关联了逻辑向量的相关内容。
2. <https://oreil.ly/J8GX>: (<http://sfirke.github.io/janitor>) 该资源是 “janitor” 软件包的官方网站，janitor 包提供便捷的数据清理功能，适用于数据管道操作，尤其是在使用管道操作符 (`|>`) 的环境中，帮助用户更高效地进行数据预处理和整理。
3. <https://oreil.ly/4uZOm>: (<https://pos.it/r4ds-03-sales>) 该资源为一个与数据分析或数据科学项目相关的 CSV 文件下载链接，可能包含销售数据集，可用于教学或示例代码运行中的数据输入。具体来说，URL “<https://pos.it/r4ds-03-sales>” 可能指向一个销售数据文件，便于用户下载并存放在项目的 `data` 文件夹中，以配合示例代码使用。
4. <https://oreil.ly/GDubb>: (<https://pos.it/r4ds-students-csv>) 该资源是一个名为 `students.csv` 的 CSV 格式数据文件，通常用于教学或示例项目中，包含学生相关的数据。用户可以通过该 URL 直接下载或在线读取该 CSV 文件以进行数据分析或练习。
5. <https://oreil.ly/RYsgM>: (<https://pos.it/r4ds-02-sales>) 该资源 “<https://pos.it/r4ds-02-sales>” 很可能是与《R for Data Science》相关的第 2 章 (Chapter 2) 或第 2 部分内容中涉及的销售数据 (sales data) 相关的文件或数据集，用于配合书中示例代码的练习和演示，方便读者下载并在本地项目的 `data` 文件夹中使用。
6. <https://oreil.ly/jVd8o>: (<https://pos.it/r4ds-01-sales>) 该资源 “<https://pos.it/r4ds-01-sales>” 可能是一个与销售数据相关的文件或数据集链接，通常用于配

合教程或项目中的示例代码，方便用户下载或在线访问销售相关的 CSV 数据文件。

7. [https://rdrr.io/pkg/janitor/man/clean\\_names.html](https://rdrr.io/pkg/janitor/man/clean_names.html): 该资源是 R 语言包“janitor”中函数 `clean_names()` 的官方文档页面，介绍了该函数的用法及功能。`clean_names()` 函数通过一定的规则和启发式方法，将数据框中的列名统一转换为下划线命名法 (snake\_case)，便于后续的数据处理和分析。
8. <https://rdrr.io/r/base/list.files.html>: 该资源是 R 语言基础包 (base package) 中的函数 `list.files()` 的官方文档页面，介绍如何使用该函数列出当前目录或指定路径下的文件，支持通过匹配文件名模式 (如正则表达式) 来筛选文件，方便批量读取文件操作。
9. <https://rdrr.io/r/base/list.html>: 该资源是 R 语言基础包 (base) 中函数 `list()` 的官方文档页面，介绍了如何使用 `list()` 函数创建列表以及相关用法说明。
10. <https://rdrr.io/r/base/readRDS.html>: 该资源是 R 语言基础包中函数 `readRDS()` 的官方文档页面，介绍了如何使用该函数读取以 R 的自定义二进制格式 (RDS 格式) 保存的单个 R 对象。文档说明了 `readRDS()` 与 `saveRDS()` 配合使用，可以将 R 对象序列化保存到磁盘，并在需要时完整无损地重新加载该对象，保证加载后得到的对象与保存时完全一致。该页面详细描述了函数的用法、参数及示例，帮助用户理解和使用 R 的 RDS 文件读写功能。
11. [https://readr.tidyverse.org/reference/col\\_skip.html](https://readr.tidyverse.org/reference/col_skip.html): 该资源是关于 R 语言包 readr 中函数 `col_skip()` 的官方文档页面。该函数用于在读取数据时跳过某些列，不将这些列包含在结果中，适用于处理大型 CSV 文件时只需使用部分列，从而加快数据读取速度。
12. <https://readr.tidyverse.org/reference/cols.html>: 该资源是 tidyverse 中 readr 包的官方文档页面，主要介绍了 `cols()` 函数及相关函数的用法，帮助用户在读取数据时自定义和控制列的类型和选择，例如通过 `cols()` 指定列类型及默认设置，或使用 `cols_only()` 仅读取指定的列。
13. [https://readr.tidyverse.org/reference/parse\\_atomic.html](https://readr.tidyverse.org/reference/parse_atomic.html): 该资源是 readr 包中关于解析基本原子向量类型的函数文档，详细介绍了如 `col_logical()`、`col_double()`、`col_integer()`、`col_character()` 等函数如何读取和解析逻辑

值、实数、整数和字符串类型的数据列，帮助用户在导入数据时明确指定列的数据类型以优化内存使用和数据处理。

14. [https://readr.tidyverse.org/reference/parse\\_datetime.html](https://readr.tidyverse.org/reference/parse_datetime.html): 该 URL ([https://readr.tidyverse.org/reference/parse\\_datetime.html](https://readr.tidyverse.org/reference/parse_datetime.html)) 指向的是 R 语言 tidyverse 中 readr 包的官方文档页面，具体介绍了 `col_date()` 和 `col_datetime()` 这两个函数的用法。这两个函数用于在读取数据时，将相应的列解析为日期 (date) 和日期时间 (datetime) 数据类型。文档中详细说明了如何使用这两个函数来正确地处理和转换文本数据中的日期和日期时间格式。
15. [https://readr.tidyverse.org/reference/parse\\_factor.html](https://readr.tidyverse.org/reference/parse_factor.html): 该资源为 tidyverse 中 readr 包的函数文档，具体介绍了函数 `col_factor()` 的用法。`col_factor()` 用于在读取数据时将列解析为因子 (factor) 类型。文档详细说明了该函数的参数设置、使用场景及示例，帮助用户正确处理和转换因子数据类型。
16. [https://readr.tidyverse.org/reference/parse\\_number.html](https://readr.tidyverse.org/reference/parse_number.html): 该资源是 readr 包中函数 `col_number()` 的文档页面，介绍了该函数作为一个宽容的数值解析器，能够忽略非数字部分，特别适用于解析货币等带有非数字字符的数值数据。
17. <https://readr.tidyverse.org/reference/problems.html>: 该资源是 tidyverse 中 readr 包的函数 `problems()` 的官方文档页面，用于帮助用户查看和诊断在使用 `read_csv()` 等读取数据函数时遇到的数据读取问题。
18. [https://readr.tidyverse.org/reference/read\\_delim.html](https://readr.tidyverse.org/reference/read_delim.html): 该 URL ([https://readr.tidyverse.org/reference/read\\_delim.html](https://readr.tidyverse.org/reference/read_delim.html)) 指向的是 R 语言中 tidyverse 包里 readr 子包的函数文档页面，具体介绍了用于读取分隔符分隔文件 (如 CSV、TSV 等) 的函数 `read_delim()`。该资源详细说明了 `read_delim()` 及其变体 (如 `read_csv()`、`read_tsv()`) 的用法，包括如何指定文件路径、处理缺失值 (NA)、跳过文件开头的元数据行、处理无列名数据、处理含引号的字符串、合并多个文件数据等常见操作和参数设置。该文档是学习如何高效导入和解析文本格式数据文件的重要参考资料。
19. [https://readr.tidyverse.org/reference/read\\_fwf.html](https://readr.tidyverse.org/reference/read_fwf.html): 该资源是 readr 包中用于读取定宽格式 (fixed-width) 文件的函数 `read_fwf()` 的文档页面。页面介绍了如何使用该函数读取定宽文件，并说明可以通过指定字段的宽度

(`fwf_widths()`) 或字段的位置 (`fwf_positions()`) 来定义字段边界。该文档详细说明了函数的参数及用法, 帮助用户正确解析和导入定宽格式的数据文件。

20. [https://readr.tidyverse.org/reference/read\\_log.html](https://readr.tidyverse.org/reference/read_log.html): 该资源是 `readr` 包中的函数 `read_log()` 的官方文档页面, 提供该函数的使用说明、参数介绍及示例, 帮助用户理解如何使用 `read_log()` 函数读取和解析日志文件。
21. [https://readr.tidyverse.org/reference/read\\_rds.html](https://readr.tidyverse.org/reference/read_rds.html): 该资源是 `readr` 包中函数 `read_rds()` 的官方文档说明页面, 介绍了 `read_rds()` 和 `write_rds()` 函数作为对 R 基础函数 `readRDS()` 和 `saveRDS()` 的统一封装, 用于以 R 的二进制 RDS 格式存储和读取 R 对象, 确保读取时恢复的是与保存时完全相同的 R 对象。页面还包含了函数的示例代码, 展示如何将数据写入和读取 `.rds` 文件。
22. [https://readr.tidyverse.org/reference/read\\_table.html](https://readr.tidyverse.org/reference/read_table.html): 该资源是 `readr` 包中函数 `read_table()` 的官方文档页面, 提供该函数的用法说明、参数介绍及示例, 帮助用户了解如何使用 `read_table()` 从文本文件中读取表格数据。
23. [https://readr.tidyverse.org/reference/write\\_delim.html](https://readr.tidyverse.org/reference/write_delim.html): 该资源是 `readr` 包中用于将数据写入文件的函数文档, 主要介绍了函数 `write_delim()` 及其相关函数 `write_csv()` 和 `write_tsv()`。文档详细说明了这些函数的主要参数, 包括要保存的数据框 (`x`)、文件保存路径 (`file`)、缺失值的写入方式 (`na`) 以及是否追加写入 (`append`) 等, 帮助用户高效地将数据导出为不同分隔符格式的文本文件。
24. <https://tibble.tidyverse.org/reference/tibble.html>: 该资源为 R 语言中 `tidyverse` 生态系统下 `tibble` 包的官方文档页面, 详细介绍了 `tibble()` 函数的用法。该函数用于“手工”创建 `tibble` 对象 (即现代化的数据框), 主要通过按列的方式构建数据表, 便于数据输入和处理。文档中还可能包含函数参数说明、示例代码及相关功能的解释, 帮助用户高效地进行数据导入与整理。
25. <https://tibble.tidyverse.org/reference/tribble.html>: 该资源是 R 语言中 `tidyverse` 包中 `tibble` 子包的一个函数文档, 介绍了函数 `tribble()`。该函数名称来源于“transposed tibble” (转置的 `tibble`), 用于按行 (而非按列) 布局数据, 方便在代码中以易读的形式输入少量数据。其特点是列名以波浪号 ( ) 开头, 数据项用逗号分隔, 适合快速手工录入数据。文档详细说明了该函数的用法和应用场景, 帮助用户更方便地创建和查看数据框。

## 第 8 章

1. <https://oreil.ly/KS82J>: (<https://www.tidyverse.org/blog>) 该资源是 tidyverse 官方博客，内容主要介绍 tidyverse 团队的最新动态、工具更新和相关的 R 语言学习资料，帮助用户及时了解 tidyverse 生态系统的发展与应用。
2. <https://oreil.ly/RxSNB>: (<https://stackoverflow.com>) 该资源是一个知名的程序员问答社区网站 Stack Overflow，用户可以在上面搜索和提问编程相关的问题，获取包括 R 语言在内的多种编程语言的技术帮助和解决方案。
3. <https://oreil.ly/uhknU>: (<https://rweekly.org>) 该资源是一个名为“R Weekly”的社区驱动项目，旨在每周汇总和发布 R 语言社区中最有趣的新闻和动态，帮助用户及时了解 R 语言相关的最新信息和发展。
4. <https://rdr.io/r/base/dput.html>: 该资源是 R 语言中函数 `dput()` 的在线文档页面，介绍如何使用 `dput()` 函数将数据以可重现的 R 代码形式导出，从而方便在其他环境中重建数据对象。页面中还提供了示例说明如何用 `dput()` 生成如 `mtcars` 数据集的重现代码。
5. <https://rdr.io/r/base/library.html>: 该资源是 R 语言中 `base` 包的官方文档页面，详细介绍了 `library()` 函数的用法和功能。`library()` 函数用于加载指定的 R 包（即库），以便在当前会话中使用包内的函数和数据。该页面帮助用户理解如何正确调用和使用 `library()` 来管理 R 包依赖，确保代码可重现。

## 第 9 章

1. <https://ggplot2.tidyverse.org/reference/aes.html>: 该资源是 `ggplot2` 包中关于函数 `aes()` 的官方文档页面，介绍了 `aes()` 函数的用法及其在数据可视化中映射变量到图形属性（如颜色、大小、形状等）的作用。



2. [https://ggplot2.tidyverse.org/reference/coord\\_fixed.html](https://ggplot2.tidyverse.org/reference/coord_fixed.html): 该资源是 ggplot2 包中函数 `coord_fixed()` 的官方文档页面, 介绍了该函数的用法和作用。`coord_fixed()` 用于设置绘图坐标系的纵横比, 使得 x 轴和 y 轴的单位长度相等, 从而保证图形比例的准确性和真实感, 常用于需要保持比例关系的图形展示。
3. [https://ggplot2.tidyverse.org/reference/coord\\_map.html](https://ggplot2.tidyverse.org/reference/coord_map.html): 该 URL ([https://ggplot2.tidyverse.org/reference/coord\\_map.html](https://ggplot2.tidyverse.org/reference/coord_map.html)) 指向的是 ggplot2 包中关于函数 `coord_map()` 的官方文档页面。该资源详细介绍了 `coord_map()` 函数的用法及其作用, 主要用于在 ggplot2 绘图中实现地图投影坐标系的转换, 帮助正确绘制地理空间数据。文档中通常会对比说明 `coord_map()` 与类似函数如 `coord_quickmap()` 之间的区别和适用场景, 便于用户根据需求选择合适的地图坐标系函数。
4. [https://ggplot2.tidyverse.org/reference/coord\\_polar.html](https://ggplot2.tidyverse.org/reference/coord_polar.html): 该资源是 ggplot2 包中函数 `coord_polar()` 的官方文档页面, 介绍了如何使用极坐标系 (polar coordinates) 进行数据可视化。通过该函数, 可以将条形图 (bar chart) 转换为极坐标形式的图表, 如饼图 (pie chart) 和玫瑰图 (Coxcomb chart), 展示数据的不同视角和更丰富的图形表达。
5. [https://ggplot2.tidyverse.org/reference/facet\\_grid.html](https://ggplot2.tidyverse.org/reference/facet_grid.html): 该 URL 指向的是 ggplot2 包中函数 `facet_grid()` 的官方文档页面。`facet_grid()` 函数用于在绘图时, 根据两个变量的组合将图形分面 (即创建一个二维的面板布局), 其第一个参数是一个双侧公式 (`rows cols`), 用于指定行和列的分面变量。与 `facet_wrap()` 函数相比, `facet_grid()` 更适合基于两个变量的组合进行网格状布局, 因此没有 `nrow` 和 `ncol` 参数来控制行列数。该文档详细说明了 `facet_grid()` 的用法、参数及与 `facet_wrap()` 的区别。
6. [https://ggplot2.tidyverse.org/reference/facet\\_wrap.html](https://ggplot2.tidyverse.org/reference/facet_wrap.html): 该 URL ([https://ggplot2.tidyverse.org/reference/facet\\_wrap.html](https://ggplot2.tidyverse.org/reference/facet_wrap.html)) 指向的是 R 语言中 ggplot2 包的函数 `facet_wrap()` 的官方文档页面。该资源详细介绍了 `facet_wrap()` 函数的用法, 该函数用于将一个绘图对象按某个分类变量分割成多个子图 (子面板), 每个子图显示数据的一个子集, 从而实现分面绘图 (faceting)。文档中通常包含函数的参数说明 (如 `nrow`、`ncol`, 用于控制子图

的行数和列数)、示例代码以及如何调整面板布局等内容,帮助用户理解和灵活使用 `facet_wrap()` 进行数据的分面可视化。

7. [https://ggplot2.tidyverse.org/reference/geom\\_abline.html](https://ggplot2.tidyverse.org/reference/geom_abline.html): 该资源是 `ggplot2` 包中函数 `geom_abline()` 的官方文档页面,介绍了该函数的用途和用法。`geom_abline()` 用于在图形中添加一条斜率和截距可控的直线,常用于表示参考线或模型拟合线,帮助用户在数据可视化中更直观地理解数据趋势和关系。
8. [https://ggplot2.tidyverse.org/reference/geom\\_bar.html](https://ggplot2.tidyverse.org/reference/geom_bar.html): 该 URL 指向的是 `ggplot2` 包中函数 `geom_bar()` 的官方文档页面。该页面详细介绍了 `geom_bar()` 的用法及其相关统计转换 (`stat`),默认统计变换为 `stat_count()`,用于绘制条形图 (bar chart),通过对数据分组进行计数来显示类别的频数。此外,页面还说明了 `geom_bar()` 与 `geom_col()` 的区别,解释了如何通过修改 `stat` 参数 (如从默认的 “count” 改为 “identity”) 来映射条形高度到数据的原始数值。页面中还包括对 `geom_bar()` 所计算的变量 (如 `count` 和 `prop`) 的说明,帮助用户理解条形图的统计计算过程。
9. [https://ggplot2.tidyverse.org/reference/geom\\_boxplot.html](https://ggplot2.tidyverse.org/reference/geom_boxplot.html): 该资源是 `ggplot2` 包中函数 `geom_boxplot()` 的官方文档页面,详细介绍了该函数的用法、参数及默认行为,用于在数据可视化中绘制箱线图。
10. [https://ggplot2.tidyverse.org/reference/geom\\_count.html](https://ggplot2.tidyverse.org/reference/geom_count.html): 该资源是 `ggplot2` 包中函数 `geom_count()` 的官方文档页面,介绍了该函数的用法及功能。`geom_count()` 用于在散点图中绘制点的计数,通过点的大小来表示数据中相同坐标点的数量,适合用于显示离散变量的频数分布。
11. [https://ggplot2.tidyverse.org/reference/geom\\_histogram.html](https://ggplot2.tidyverse.org/reference/geom_histogram.html): 该资源是 `ggplot2` 包中关于 `geom_histogram` 函数的官方文档页面,介绍了如何使用 `geom_histogram` 来绘制直方图,是统计变换 `stat_bin` 的实现细节和用法说明。
12. [https://ggplot2.tidyverse.org/reference/geom\\_jitter.html](https://ggplot2.tidyverse.org/reference/geom_jitter.html): 该资源是 `ggplot2` 包中用于数据可视化的函数文档页面,详细介绍了函数 `geom_jitter()` 的用法。`geom_jitter()` 提供了一种在散点图中添加随机抖动 (jitter) 的方法,通过对点的位置进行轻微扰动,避免数据点重叠,使得图形在大尺度上更具信息量

和可读性。文档中说明了如何使用该函数、控制抖动程度的参数，以及与其他相关函数（如 `geom_count()`）的比较。

13. [https://ggplot2.tidyverse.org/reference/geom\\_point.html](https://ggplot2.tidyverse.org/reference/geom_point.html): 该资源是 ggplot2 包中函数 `geom_point()` 的官方文档页面，介绍了如何在绘图中添加散点图层，包括函数的用法、参数说明及示例。
14. [https://ggplot2.tidyverse.org/reference/geom\\_smooth.html](https://ggplot2.tidyverse.org/reference/geom_smooth.html): 该资源是 ggplot2 包中函数 `geom_smooth()` 的官方文档页面，详细介绍了该函数的用法及参数说明。`geom_smooth()` 用于在 ggplot2 图形中添加平滑曲线（如回归线、局部加权回归线等），支持根据分组变量绘制多条平滑线，常用于展示数据趋势和拟合效果。文档中还说明了该函数如何处理美学映射（如 `linetype`）、分组变量的作用，以及参数如 `se`（控制是否显示置信区间）的功能。此外，该页面通常包含与统计变换函数 `stat_smooth()` 相关的信息，帮助用户理解背后的计算机制和可调节选项。
15. <https://ggplot2.tidyverse.org/reference/ggplot.html>: 该资源是 ggplot2 包中函数 `ggplot()` 的官方文档页面，详细介绍了如何使用 `ggplot()` 函数创建绘图对象，包括如何全局设置数据和美学映射（mapping），以及如何通过修改 `ggplot()` 中的参数来控制图层的默认行为。文档还说明了 `ggplot()` 与各类几何对象（geom）函数的关系，指导用户如何在 `ggplot()` 基础上添加不同的几何层，从而构建复杂的图形。
16. [https://ggplot2.tidyverse.org/reference/position\\_dodge.html](https://ggplot2.tidyverse.org/reference/position_dodge.html): 该资源是 ggplot2 包中函数 `position_dodge` 的官方文档页面，提供该函数的用法说明和参数介绍，用于在绘图中调整图形元素的水平避让位置。
17. [https://ggplot2.tidyverse.org/reference/position\\_identity.html](https://ggplot2.tidyverse.org/reference/position_identity.html): 该资源是 ggplot2 包中函数或参数 `position_identity` 的官方文档页面，提供该函数的用法说明和示例，主要用于控制图形元素在绘图中的位置调整方式。
18. [https://ggplot2.tidyverse.org/reference/position\\_jitter.html](https://ggplot2.tidyverse.org/reference/position_jitter.html): 该资源是 ggplot2 包中函数 `position_jitter` 的官方文档页面，提供该函数的用法说明、参数介绍及示例，帮助用户在绘图时对数据点进行抖动处理以避免重叠。

19. [https://ggplot2.tidyverse.org/reference/position\\_stack.html](https://ggplot2.tidyverse.org/reference/position_stack.html): 该资源是 ggplot2 包中关于函数 `position_stack` 的官方文档页面。`position_stack` 用于在图形中对元素进行堆叠位置调整，常用于条形图或面积图，帮助按顺序堆叠不同类别的数据。页面详细说明了该函数的用法、参数和示例，方便用户理解和应用该位置调整方法。
20. [https://ggplot2.tidyverse.org/reference/stat\\_summary.html](https://ggplot2.tidyverse.org/reference/stat_summary.html): 该资源是 ggplot2 包中函数 `stat_summary()` 的官方文档页面。`stat_summary()` 用于对每个唯一的 x 值的 y 值进行统计汇总（如计算均值、中位数等），并将汇总结果用于绘图，从而突出显示统计变换的结果。该页面详细介绍了 `stat_summary()` 的用法、默认的几何对象（geom）、参数设置及示例，帮助用户理解如何在绘图中应用统计汇总函数。
21. <https://oreil.ly/45GHE>: (<https://ggplot2-book.org/maps.html>) 该资源是《ggplot2: Elegant Graphics for Data Analysis》一书中的“Maps”章节，详细介绍了如何使用 ggplot2 包绘制地理地图，包括设置合适的坐标比例（如使用 `coord_quickmap()` 函数）以正确展示空间数据的技术和方法。
22. <https://oreil.ly/8fZzE>: (<https://vita.had.co.nz/papers/layered-grammar.pdf>) 该资源是一篇题为《A Layered Grammar of Graphics》的学术论文，详细介绍了 ggplot2 背后的理论基础和图形语法的分层结构。
23. [https://oreil.ly/ARL\\_4](https://oreil.ly/ARL_4): (<https://exts.ggplot2.tidyverse.org/gallery>) 该资源 (<https://exts.ggplot2.tidyverse.org/gallery>) 是一个展示 ggplot2 扩展包及其各种自定义绘图示例的画廊网站，用户可以在这里浏览和学习使用不同扩展包（如 `ggribes`）实现的多样化图形类型和高级绘图技巧。
24. <https://oreil.ly/NIKZF>: (<https://posit.co/resources/cheatsheets>) 该资源是由 Posit（原 RStudio）提供的各种数据科学相关的速查表（cheatsheets）合集，旨在帮助用户快速查阅和掌握包括 ggplot2 在内的 R 语言包及工具的功能和用法，便于数据可视化和分析工作。
25. <https://oreil.ly/SP6zV>: (<https://ggplot2.tidyverse.org/articles/ggplot2-specs.html>) 该资源是 ggplot2 包中关于“美学映射规范”（aesthetic specifications）的官方说明文档或教程，详细介绍了在绘图中可用的各种美学属

性（如颜色、形状、大小等）的定义和使用方法，帮助用户理解如何在 ggplot2 中映射和设置图形元素的视觉属性。

26. <https://oreil.ly/W6ci8>: (<https://ggplot2.tidyverse.org>) 该资源是 ggplot2 包的官方网站，提供有关 ggplot2 功能的全面介绍、文档和使用指南。
27. <https://oreil.ly/cIFgm>: (<https://ggplot2.tidyverse.org/reference>) 该资源是 ggplot2 包的官方参考页面，提供了该包中所有几何对象 (geoms) 及函数的全面概述，是学习和查阅 ggplot2 中各类图层和功能的权威指南。
28. <https://oreil.ly/pPIuL>: (<https://wilkelab.org/ggribes>) 该资源是 ggribes 包的官方网站，提供关于该 R 语言扩展包的详细文档和说明。ggribes 包用于在 ggplot2 框架下创建 “ridgeline plots” (脊线图)，这类图形有助于展示数值变量在不同类别水平上的密度分布情况。网站中包含该包中函数（如 `geom_density_ridges()`）的用法示例和参数说明，方便用户学习如何实现和定制脊线图。
29. [https://wilkelab.org/ggribes/reference/geom\\_density\\_ridges.html](https://wilkelab.org/ggribes/reference/geom_density_ridges.html): 该资源是 ggribes 包中函数 `geom_density_ridges()` 的官方文档页面，介绍了该函数的用法及参数，主要用于在 ggplot2 中绘制脊线图 (ridgeline plots)，即通过叠加多个密度曲线来展示数值变量在不同类别水平上的分布情况。

## 第 10 章

1. [https://dplyr.tidyverse.org/reference/if\\_else.html](https://dplyr.tidyverse.org/reference/if_else.html): 该资源是 dplyr 包中函数 `if_else()` 的官方文档页面，介绍了如何使用 `if_else()` 函数进行条件判断和替换操作，通常用于根据条件替换数据框中的值，例如将异常值替换为缺失值 (NA)。
2. <https://dplyr.tidyverse.org/reference/mutate.html>: 该资源是 dplyr 包中函数 `mutate()` 的官方文档页面，介绍了 `mutate()` 函数的用法及示例。`mutate()`

用于基于现有变量创建或修改数据框中的列，常用于数据清洗和转换，例如将异常值替换为缺失值（NA）。

3. [https://forcats.tidyverse.org/reference/fct\\_reorder.html](https://forcats.tidyverse.org/reference/fct_reorder.html): 该资源是关于 R 语言中 tidyverse 包中的函数 `fct_reorder()` 的文档页面。该函数用于对无内在顺序的类别型变量（因子）重新排序，以便在数据展示时更具信息性和可读性。页面详细介绍了如何根据某些数值变量（如平均值）对因子的水平进行重新排序，从而改善分类数据的可视化效果。
4. [https://ggplot2.tidyverse.org/reference/aes\\_eval.html](https://ggplot2.tidyverse.org/reference/aes_eval.html): 该资源介绍了 ggplot2 包中的函数 `aes_eval()`，特别是与 `after_stat()` 函数相关的用法。`after_stat()` 用于在绘图映射中引用计算统计量（如密度）之后生成的变量，使用户能够在图形美学映射中使用统计变换的结果。文档详细说明了如何在 `aes()` 映射中评估和使用这些计算后的统计变量。
5. [https://ggplot2.tidyverse.org/reference/coord\\_cartesian.html](https://ggplot2.tidyverse.org/reference/coord_cartesian.html): 该 URL 指向的资源是 ggplot2 包中函数 `coord_cartesian()` 的官方文档页面。该函数用于在绘图时对坐标轴进行缩放（zoom），实现对图形某一区域的放大显示，而不会丢弃坐标轴范围外的数据。与 `xlim()` 和 `ylim()` 函数不同，后者在缩放时会直接丢弃超出范围的数据，`coord_cartesian()` 则保留完整数据，只调整显示范围，适合用来细致观察数据中某些小范围的变化，例如突出显示 y 轴上的小数值或局部特征。
6. [https://ggplot2.tidyverse.org/reference/coord\\_flip.html](https://ggplot2.tidyverse.org/reference/coord_flip.html): 该资源是 ggplot2 包中函数 `coord_flip()` 的官方文档页面，介绍了如何通过添加 `coord_flip()` 这一图层，将垂直箱线图转换为水平箱线图，并讨论了与直接交换 x 和 y 变量的区别。
7. [https://ggplot2.tidyverse.org/reference/cut\\_interval.html](https://ggplot2.tidyverse.org/reference/cut_interval.html): 该资源是 ggplot2 包中关于函数 `cut_interval()` 的官方文档页面。该函数用于将连续变量按照等宽区间（intervals of equal width）进行分割，从而将数据划分为多个区间（bins），常用于数据分箱和分类。文档中详细介绍了该函数的用法、参数说明以及示例，方便用户理解如何基于等宽区间对数据进行切割，从而在数据可视化（如频率多边形、箱线图等）中更好地处理和展示连续变量的分布特征。

8. [https://ggplot2.tidyverse.org/reference/geom\\_bin\\_2d.html](https://ggplot2.tidyverse.org/reference/geom_bin_2d.html): 该资源是 ggplot2 包中函数 `geom_bin2d()` 的官方文档页面, 介绍了如何在二维坐标平面上将数据点按矩形网格进行分箱 (binning), 并通过填充颜色显示每个矩形箱中点的数量。该函数适用于处理大规模数据集的二维分布可视化, 是 `geom_hex()` (使用六边形分箱) 的替代方案之一。
9. [https://ggplot2.tidyverse.org/reference/geom\\_boxplot.html](https://ggplot2.tidyverse.org/reference/geom_boxplot.html): 该资源是 ggplot2 包中函数 `geom_boxplot()` 的官方文档页面, 介绍了该函数用于绘制箱线图的用法和参数说明。
10. [https://ggplot2.tidyverse.org/reference/geom\\_count.html](https://ggplot2.tidyverse.org/reference/geom_count.html): 该资源是 ggplot2 包中的一个函数文档页面, 介绍了 `geom_count()` 函数的用法。  
`geom_count()` 用于在二维图形中可视化两个分类变量之间的共变关系, 通过点的大小表示每种类别组合的观测次数, 从而帮助用户直观了解分类变量的频数分布和关联情况。
11. [https://ggplot2.tidyverse.org/reference/geom\\_density.html](https://ggplot2.tidyverse.org/reference/geom_density.html): 该资源是 ggplot2 包中函数 `geom_density()` 的官方文档页面, 介绍了如何使用该函数在 R 语言中绘制基于核密度估计的平滑密度曲线, 用于可视化数值型变量的分布情况, 尤其适合根据类别变量的不同水平比较数值分布的差异。
12. [https://ggplot2.tidyverse.org/reference/geom\\_hex.html](https://ggplot2.tidyverse.org/reference/geom_hex.html): 该资源是 ggplot2 包中用于二维数据分箱的几何对象函数 `geom_hex()` 的官方文档页面。  
`geom_hex()` 将二维平面划分为六边形的网格单元 (六边形箱), 并通过填充颜色表示落入每个箱内的数据点数量。使用该函数需要先安装 R 语言的 hexbin 包。该方法适合对大规模数据集进行可视化, 帮助缓解透明度叠加带来的显示问题。
13. [https://ggplot2.tidyverse.org/reference/geom\\_histogram.html](https://ggplot2.tidyverse.org/reference/geom_histogram.html): 该 URL ([https://ggplot2.tidyverse.org/reference/geom\\_histogram.html](https://ggplot2.tidyverse.org/reference/geom_histogram.html)) 指向的是 ggplot2 包中 `geom_histogram()` 函数的官方文档页面。`geom_histogram()` 是用于绘制直方图的图层函数, 主要用于对数值型数据进行分箱 (binning) 并展示数据的分布情况。文档中通常包含该函数的用法说明、参数解析、示例代码以及绘图效果, 帮助用户理解如何使用直方图来可视化数值变量的分布特征, 特别是在探索数据集中数值变量和分类变量关系时的应用。

14. [https://ggplot2.tidyverse.org/reference/geom\\_jitter.html](https://ggplot2.tidyverse.org/reference/geom_jitter.html): 该资源是 ggplot2 包中函数 `geom_jitter()` 的官方文档页面, 介绍了该函数的用法。  
`geom_jitter()` 用于在绘制散点图时, 通过向点的位置添加少量随机抖动, 避免数据点重叠 (即过度绘制), 从而更清晰地展示连续变量与分类变量之间的关系。该页面详细说明了函数的参数、示例及应用场景。
15. [https://ggplot2.tidyverse.org/reference/geom\\_point.html](https://ggplot2.tidyverse.org/reference/geom_point.html): 该资源是 ggplot2 包中用于绘制散点图的函数文档页面, 介绍了如何使用 `geom_point()` 来可视化两个数值变量之间的协变关系, 通过散点的分布模式展示变量间的关系, 例如展示钻石的克拉数与价格之间的正相关关系。
16. [https://ggplot2.tidyverse.org/reference/geom\\_tile.html](https://ggplot2.tidyverse.org/reference/geom_tile.html): 该资源是 R 语言中 ggplot2 包的官方文档页面, 详细介绍了函数 `geom_tile()` 的用法。  
`geom_tile()` 用于绘制矩形瓦片图 (tile plot), 常用于展示二维数据的热度图或分类变量之间的关系。文档中包含函数的参数说明、示例代码以及使用场景, 帮助用户通过填充色彩 (fill aesthetic) 直观展示数值大小或类别分布。
17. [https://ggplot2.tidyverse.org/reference/geom\\_violin.html](https://ggplot2.tidyverse.org/reference/geom_violin.html): 该资源是 ggplot2 包中用于绘制小提琴图 (violin plot) 的函数 `geom_violin()` 的官方文档页面。页面详细介绍了如何使用 `geom_violin()` 来可视化数值变量 (如钻石价格) 在不同类别变量 (如分类因子) 水平上的分布情况。文档中包括函数的用法说明、参数设置、示例代码及图形展示, 帮助用户理解并应用小提琴图进行数据分布的比较与分析。
18. <https://ggplot2.tidyverse.org/reference/lims.html>: 该资源是 ggplot2 包中关于函数 `xlim()` 和 `ylim()` 的官方文档页面。文档介绍了这两个函数的用法及其特点, 特别说明它们用于设置坐标轴的显示范围时, 会丢弃超出范围的数据点。  
与 `coord_cartesian()` 函数的 `xlim` 参数相比, `xlim()` 和 `ylim()` 直接筛选数据, 作用机制略有不同, 适用于需要数据裁剪的情形。该页面帮助用户理解如何通过这两个函数控制图形的坐标轴范围及其对数据的影响。
19. <https://rdrr.io/pkg/nycflights13/man/flights.html>: 该 URL 指向的是 R 语言数据包 `nycflights13` 中 `flights` 数据集的文档页面。该页面详细介绍了 `flights` 数据集的结构、变量说明及其含义。具体来说, `flights` 数据集包含 2013 年从纽约



三个机场出发的航班信息，其中变量如 `dep_time` 记录了航班的实际起飞时间，缺失值表示航班被取消。该文档有助于用户理解数据集中各字段的含义及如何处理缺失值。

20. <https://rdr.io/r/base/NA.html>: 该资源是 R 语言基础包中关于函数 `is.na()` 的文档页面。该页面详细介绍了 `is.na()` 函数的用法，该函数用于检测数据中的缺失值 (NA)，返回一个逻辑向量，指示每个元素是否为缺失值。在数据处理中，`is.na()` 常用于判断和处理缺失数据，例如区分缺失和非缺失的观测值，方便后续分析和操作。
21. <https://rdr.io/r/base/mean.html>: 该资源是 R 语言基础包 (base package) 中函数 `mean()` 的官方文档页面，详细介绍了 `mean()` 函数的用法、参数说明 (包括 `na.rm = TRUE` 参数的作用)，以及函数如何计算数值向量的算术平均值。
22. <https://rdr.io/r/base/sum.html>: 该 URL 指向的是 R 语言基础包 (base package) 中函数 `sum()` 的官方文档页面。页面详细介绍了 `sum()` 函数的用法、参数 (包括 `na.rm` 参数)、返回值及示例，帮助用户理解如何对数值向量进行求和操作。

## 第 11 章

1. [https://ggforce.data-imaginist.com/reference/geom\\_mark\\_hull.html](https://ggforce.data-imaginist.com/reference/geom_mark_hull.html): 该资源为 `ggforce` 包中函数 `geom_mark_hull()` 的官方文档页面，介绍了如何使用该函数在 `ggplot2` 图形中通过包围凸包的方式对点集的子集进行标注和高亮，提供了函数的用法说明、参数设置及示例。
2. <https://ggplot2.tidyverse.org/reference/annotate.html>: 该 URL (<https://ggplot2.tidyverse.org/reference/annotate.html>) 指向的是 R 语言数据可视化包 `ggplot2` 中函数 `annotate()` 的官方文档页面。该函数用于向图形中添加注释元素，适合添加一个或少量的标注内容，与通过 `geom` 层突出展示数据子集不同，`annotate()` 更灵活便捷地实现对图形的文字、形状等注释补充。文档中通常包含

函数用法说明、参数详解及示例代码，帮助用户理解并高效使用该函数进行图形注释。

3. [https://ggplot2.tidyverse.org/reference/coord\\_cartesian.html](https://ggplot2.tidyverse.org/reference/coord_cartesian.html): 该资源是 ggplot2 包中函数 `coord_cartesian()` 的官方文档页面。`coord_cartesian()` 用于设置绘图坐标系的显示范围（如 x 轴和 y 轴的限制），通过调整坐标范围实现对图形的“缩放”效果，而不会丢失数据点或改变数据子集。相比直接设置坐标轴刻度的 `limits` 参数，`coord_cartesian()` 更适合用来放大图中某一区域，从而更灵活地观察数据细节。
4. <https://ggplot2.tidyverse.org/reference/element.html>: 该资源是 ggplot2 包中用于自定义图形主题中非数据元素样式的函数参考文档，主要介绍了 `element_*`() 系列函数（如 `element_text()` 和 `element_rect()`）的用法。这些函数用于设置图形中标题、图例边框、字体大小、颜色、字体样式（如加粗）等细节，使用户能够精细控制图形主题中的各个组件的外观。
5. [https://ggplot2.tidyverse.org/reference/geom\\_abline.html](https://ggplot2.tidyverse.org/reference/geom_abline.html): 该资源是 ggplot2 包中关于函数 `geom_abline()` 的官方文档页面，介绍了如何使用 `geom_abline()` 及其相关函数（如 `geom_hline()` 和 `geom_vline()`）来添加参考线（斜线、水平线和垂直线）到图表中，通常通过设置线宽和颜色来使这些参考线既清晰可见又不喧宾夺主，从而辅助数据可视化的解读。
6. [https://ggplot2.tidyverse.org/reference/geom\\_point.html](https://ggplot2.tidyverse.org/reference/geom_point.html): 该 URL 指向的资源是 ggplot2 包中函数 `geom_point()` 的官方文档页面。该函数用于在图形中绘制散点，通常用于展示数据的单个观测点。文档详细介绍了 `geom_point()` 的用法、参数和示例，帮助用户理解如何通过该函数在图表中添加点的几何对象。
7. [https://ggplot2.tidyverse.org/reference/geom\\_segment.html](https://ggplot2.tidyverse.org/reference/geom_segment.html): 该资源是 ggplot2 包中函数 `geom_segment()` 的官方文档页面，介绍了如何使用 `geom_segment()` 函数绘制线段。文档说明该函数可以通过 `arrow` 参数添加箭头，用于在图形中指示和强调特定点的位置。用户需通过美学映射参数 `x` 和 `y` 指定线段的起点坐标，通过 `xend` 和 `yend` 指定终点坐标。
8. [https://ggplot2.tidyverse.org/reference/geom\\_text.html](https://ggplot2.tidyverse.org/reference/geom_text.html): 该 URL 指向的资源是 ggplot2 包中函数 `geom_text()` 的官方文档页面。`geom_text()` 是 ggplot2 中

用于在图形上添加文本标签的几何对象 (geom)，它类似于 `geom_point()`，但额外支持 `label` 美学映射，使用户能够在图中标注单个观测点或观测组的文本信息。该文档详细介绍了 `geom_text()` 的用法、参数及其与分面 (faceting) 等功能的交互方式，帮助用户在数据可视化中实现灵活的文本注释。

9. [https://ggplot2.tidyverse.org/reference/geom\\_tile.html](https://ggplot2.tidyverse.org/reference/geom_tile.html): 该资源是 ggplot2 包中用于绘制矩形图形的函数参考文档页面，详细介绍了如何使用 `geom_tile()` 函数通过设置坐标范围 (如 `xmin`、`xmax`、`ymin`、`ymax`) 来绘制矩形，从而用于图形中的区域标注或热图等效果。
10. <https://ggplot2.tidyverse.org/reference/ggtheme.html>: 该资源是 ggplot2 包中关于主题 (theme) 函数的官方文档页面，详细介绍了 ggplot2 内置的多种图形主题 (如默认的 `theme_gray()`) 及其用法，帮助用户自定义和调整图表的整体视觉风格。
11. [https://ggplot2.tidyverse.org/reference/guide\\_colourbar.html](https://ggplot2.tidyverse.org/reference/guide_colourbar.html): 该资源是 ggplot2 包中关于函数 `guide_colorbar()` 的官方文档页面，介绍了如何使用 `guide_colorbar()` 来控制颜色条 (colorbar) 图例的显示和样式设置，帮助用户自定义图例中颜色渐变条的外观和行为。
12. [https://ggplot2.tidyverse.org/reference/guide\\_legend.html](https://ggplot2.tidyverse.org/reference/guide_legend.html): 该资源是 ggplot2 包中用于图例 (legend) 控制的函数文档页面，具体介绍了函数 `guide_legend()` 的用法。页面内容详细说明了如何通过 `guide_legend()` 来自定义图例的显示方式，例如调整图例的行数 (`nrow` 参数) 和覆盖默认美学属性 (如点的大小)，从而实现更灵活和美观的图例设计。该文档适合需要精细控制 ggplot2 图例样式和布局的用户参考。
13. <https://ggplot2.tidyverse.org/reference/guides.html>: 该资源是 ggplot2 包中用于控制图例显示的函数指南文档。具体介绍了如何使用 `guides()` 函数配合 `guide_legend()` 和 `guide_colorbar()` 来定制单个图例的外观和行为，包括调整图例的行数 (`nrow` 参数) 以及覆盖某些美学属性 (如点的大小)，以便更灵活地展示图中多个图例信息。文档还说明了 `guides()` 函数中参数名称需与对应的美学名称一致，方便用户对图例进行精细化设置。

14. <https://ggplot2.tidyverse.org/reference/labs.html>: 该资源是 ggplot2 包中用于添加和修改图形标签的函数说明文档页面。具体介绍了如何使用 labs() 函数为图形添加标题 (title)、副标题 (subtitle)、坐标轴标签、图例标题以及图形底部的说明文字 (caption), 帮助用户将探索性图形转化为更具说明性的图表。页面还说明了 labs() 函数中各标签参数的用法和作用, 以及与其他函数 (如 guides()) 的关联。
15. [https://ggplot2.tidyverse.org/reference/scale\\_colour\\_discrete.html](https://ggplot2.tidyverse.org/reference/scale_colour_discrete.html): 该资源是 ggplot2 包中关于函数 scale\_color\_discrete() 的官方文档页面, 介绍了如何为离散型变量 (如分类变量) 在图形中选择和设置颜色的比例尺 (scale)。它详细说明了该函数的用法、参数以及示例, 帮助用户自定义和控制离散颜色的映射效果。
16. [https://ggplot2.tidyverse.org/reference/scale\\_continuous.html](https://ggplot2.tidyverse.org/reference/scale_continuous.html): 该资源是 ggplot2 包中关于连续型比例尺函数 scale\_continuous() 的官方文档页面, 详细介绍了如何在图形中对连续变量进行刻度设置和调整, 包括参数说明、用法示例及相关功能。
17. [https://ggplot2.tidyverse.org/reference/scale\\_gradient.html](https://ggplot2.tidyverse.org/reference/scale_gradient.html): 该资源是 ggplot2 包中关于渐变色调色板函数的官方文档, 详细介绍了用于连续颜色映射的函数如 scale\_color\_gradient()、scale\_fill\_gradient() 以及用于区分正负值的发散色标函数 scale\_color\_gradient2(), 帮助用户在数据可视化中通过颜色渐变表现数值的连续变化或区分特定数值范围。
18. [https://ggplot2.tidyverse.org/reference/scale\\_manual.html](https://ggplot2.tidyverse.org/reference/scale_manual.html): 该资源是 ggplot2 包中关于函数 scale\_color\_manual() 的官方文档页面, 详细介绍了如何在绘图时自定义颜色映射, 允许用户预先指定数值与颜色之间的对应关系, 例如为不同类别分配特定的颜色代码。
19. <https://ggplot2.tidyverse.org/reference/theme.html>: 该 URL (<https://ggplot2.tidyverse.org/reference/theme.html>) 指向的是 R 语言数据可视化包 ggplot2 中用于设置和自定义图形主题的函数 theme() 的官方文档页面。该资源详细介绍了如何通过 theme() 函数控制图形中非数据部分的样式和布局, 例如图例的位置 (legend.position)、图例的方向和边框、图标题和图注的位置

(`plot.title.position`、`plot.caption.position`) 等。文档还说明了如何结合 `element_*()` 函数 (如 `element_text()` 和 `element_rect()`) 来进一步定制字体、颜色和边框等细节, 帮助用户灵活调整图形的整体美观和信息表达。

20. [https://oreil.ly/CK\\_sd](https://oreil.ly/CK_sd): (<https://r-graphics.org>) 该资源 “<https://r-graphics.org>” 可能是一个专注于 R 语言绘图与数据可视化的综合性网站, 提供关于 R 图形系统 (如 `ggplot2`) 的教程、示例和相关书籍推荐, 旨在帮助用户深入学习和掌握 R 语言中的图形绘制技巧与最佳实践。
21. <https://oreil.ly/DZtL1>: (<https://ggforce.data-imaginist.com/index.html>) 该资源是 `ggforce` 包的官方文档主页, 提供了关于 `ggforce` 中各种扩展 `ggplot2` 功能的详细介绍和使用说明, 特别是用于数据可视化中标注和高亮数据子集的几何对象 (如 `geom_mark_hull`) 的相关信息。
22. <https://oreil.ly/F1nga>: (<https://jrnold.github.io/ggthemes>) 该资源是由 Jeffrey Arnold 提供的 R 语言扩展包 `ggthemes` 的官方网站, 包含了许多用于 `ggplot2` 的额外图形主题, 方便用户应用多样化的图表风格, 满足特定的企业或期刊风格需求。
23. <https://oreil.ly/IVSL4>: (<https://ggrepel.slowkow.com>) 该资源是由 Kamil Slowikowski 开发的 R 语言包 `ggrepel` 的官方网站或相关介绍页面, 提供用于 `ggplot2` 图形中智能避免标签重叠的扩展功能。
24. <https://oreil.ly/LNHAY>: (<https://colorbrewer2.org>) 该资源是一个名为 ColorBrewer 的在线工具网站, 网址为 <https://colorbrewer2.org>。它提供各种颜色调色板 (ColorBrewer scales), 特别适用于数据可视化中的分类、顺序和分歧数据的配色方案, 帮助用户选择科学合理且易于辨识的颜色组合。该网站的调色板常用于 R 语言的 `RColorBrewer` 包中, 用于增强图表的视觉效果和信息传达。
25. [https://oreil.ly/QIr\\_w](https://oreil.ly/QIr_w): (<http://www.thefunctionalart.com/p/the-truthful-art-book.html>) 该资源是 Albert Cairo 所著的《The Truthful Art》一书的介绍页面, 书中侧重于指导读者如何思考和设计有效的图形视觉表达, 而非具体的制作技术, 适合作为理解数据可视化原则和方法的参考读物。

- 
26. <https://oreil.ly/T4Jxn>: (<https://ggplot2-book.org>) 该资源是《ggplot2: Elegant Graphics for Data Analysis》一书的官方网站，提供了关于 R 语言中 ggplot2 包的全面介绍和详细说明，特别涵盖了图形主题（theming）和数据可视化的高级用法，是学习和深入掌握 ggplot2 绘图系统的重要参考资料。
  27. <https://oreil.ly/i1lyd>: (<https://michelf.ca/projects/sim-daltonism>) 该资源是一个名为 SimDaltonism 的工具网站，提供模拟色盲效果的功能，帮助用户测试和查看图像在不同类型色盲状态下的视觉表现。
  28. <https://oreil.ly/uJRYK>: (<https://clauswilke.com/dataviz>) 该资源是 Claus Wilke 提供的与数据可视化相关的内容，网址为 <https://clauswilke.com/dataviz>。它很可能包含关于数据可视化的教程、示例或书籍《Fundamentals of Data Visualization》的补充材料，旨在帮助读者深入理解和掌握数据可视化技术。
  29. <https://oreil.ly/xWxVV>: (<https://patchwork.data-imaginist.com>) 该资源是 R 语言中用于组合和布局多个 ggplot2 图形的扩展包 patchwork 的官方网站，提供该包的使用指南、示例和相关文档，帮助用户更方便地进行多图排列和展示。
  30. [https://patchwork.data-imaginist.com/reference/guide\\_area.html](https://patchwork.data-imaginist.com/reference/guide_area.html): 该资源是 patchwork 包中关于函数或组件 `guide_area()` 的官方文档或使用指南页面。页面详细介绍了 `guide_area()` 在 patchwork 中的作用，即作为图例（legend）放置区域，允许用户自定义图例的位置和大小，从而统一管理和调整多个子图的图例布局。该页面还可能包含如何通过 `guide_area()` 控制图例高度及其在整体拼接图中的排布方式的说明和示例。
  31. [https://rdr.io/pkg/ggrepel/man/geom\\_text\\_repel.html](https://rdr.io/pkg/ggrepel/man/geom_text_repel.html): 该资源为 R 语言包 ggrepel 中函数 `geom_text_repel()` 的官方文档页面，介绍了该函数的用法和功能。`geom_text_repel()` 用于在绘图时自动调整文本标签的位置，避免标签与点或其他标签重叠，从而提高图表的可读性和美观性。该文档详细说明了如何使用该函数来标注和突出显示图中的特定点。
  32. <https://rdr.io/r/base/cut.html>: 该资源是 R 语言基础包中的函数 `cut()` 的在线文档页面，网址为 <https://rdr.io/r/base/cut.html>。该页面详细介绍了 `cut()` 函

数的用法，该函数用于将连续型变量分割成若干区间，从而将其转换为分类变量，便于分类数据的分析和可视化。

33. <https://rdrr.io/r/base/substitute.html>: 该资源是 R 语言基础包中函数 `substitute()` 的文档说明页面，介绍了如何使用 `substitute()` 函数来替代表达式中的部分内容。文档详细讲解了该函数的用法及参数，尤其是在表达式中使用数学符号和公式时，如何通过 `quote()` 和 `substitute()` 实现动态替换和表达式操作。
34. <https://rdrr.io/r/grDevices/plotmath.html>: 该资源是 R 语言中用于在图形设备中绘制数学表达式的帮助文档，详细介绍了如何使用 `plotmath` 语法在图表中添加数学符号、公式和表达式的各种方法和选项。
35. <https://rdrr.io/r/grid/arrow.html>: 该资源是 R 语言中 `grid` 包的函数 `arrow()` 的文档页面，介绍了 `arrow()` 函数的四个参数及其作用，通常用于在图形中添加箭头效果。页面详细说明了参数的功能并提供示例，帮助用户理解和使用该函数来绘制带箭头的图形元素。
36. <https://rdrr.io/r/utils/install.packages.html>: 该资源是 R 语言官方文档中关于函数 `install.packages()` 的说明页面，介绍了如何在 R 环境中安装软件包。
37. [https://readr.tidyverse.org/reference/parse\\_datetime.html](https://readr.tidyverse.org/reference/parse_datetime.html): 该资源是 tidyverse 中 `readr` 包的函数文档，具体介绍了函数 `parse_datetime()`，用于按照指定格式解析日期时间字符串。
38. [https://scales.r-lib.org/reference/label\\_dollar.html](https://scales.r-lib.org/reference/label_dollar.html): 该资源是 R 语言中 `scales` 包的官方文档页面，详细介绍了函数 `label_dollar()` 的用法。该函数用于格式化数字为货币形式，默认会添加美元符号和千分位逗号分隔符，方便在数据可视化中对货币数值进行美观标注和自定义调整。
39. [https://scales.r-lib.org/reference/label\\_percent.html](https://scales.r-lib.org/reference/label_percent.html): 该资源是 R 语言中 `scales` 包的官方文档页面，介绍了函数 `label_percent()` 的用法。该函数用于将数值格式化为百分比标签，便于在图表或数据展示中以百分数形式显示数据。
40. [https://stringr.tidyverse.org/reference/str\\_wrap.html](https://stringr.tidyverse.org/reference/str_wrap.html): 该资源是 `stringr` 包中函数 `str_wrap()` 的官方文档页面，介绍了如何自动根据指定的每行字符数对长文本进行换行处理。

## 第 12 章

1. <https://dplyr.tidyverse.org/reference/arrange.html>: 该资源是 dplyr 包中函数 `arrange()` 的官方文档页面, 介绍了如何使用 `arrange()` 对数据框进行排序, 包括默认情况下缺失值 (NA) 排序的位置及如何通过结合 `is.na()` 函数来自定义缺失值的排序规则。
2. [https://dplyr.tidyverse.org/reference/case\\_when.html](https://dplyr.tidyverse.org/reference/case_when.html): 该 URL 指向的资源是 dplyr 包中函数 `case_when()` 的官方文档。`case_when()` 函数灵感来源于 SQL 中的 CASE 语句, 提供了一种灵活的方式根据多个条件执行不同的计算。它采用 “condition output” 形式的语法, condition 是逻辑向量, 当条件为 TRUE 时, 返回对应的 output 值。该文档详细介绍了 `case_when()` 的用法、语法特点、条件匹配规则 (未匹配则返回 NA)、类型兼容性要求, 以及如何利用该函数进行条件转换和变量赋值等内容。它是实现复杂条件判断和多分支逻辑的常用工具, 常与 `if_else()` 函数一起使用。
3. <https://dplyr.tidyverse.org/reference/coalesce.html>: 该资源是 dplyr 包中函数 `coalesce()` 的官方文档页面, 介绍了该函数的用法和功能。`coalesce()` 用于按顺序检查多个向量的值, 返回第一个非缺失 (非 NA) 的元素, 常用于数据处理中合并多个向量以填补缺失值。
4. <https://dplyr.tidyverse.org/reference/context.html>: 该 URL (<https://dplyr.tidyverse.org/reference/context.html>) 指向的是 dplyr 包的官方文档页面, 具体介绍了函数 `n()` 的使用上下文 (context)。文档解释了在不同的分组环境下, `n()` 函数如何返回当前分组的元素数量, 例如按天分组时返回每天的延误航班数, 而在整体数据上则返回航班总数。该资源帮助用户理解 `n()` 函数在数据分组操作中的行为和作用。
5. <https://dplyr.tidyverse.org/reference/count.html>: 该资源是 dplyr 包中函数 `count()` 的官方文档页面, 介绍了如何使用 `count()` 函数对数据框中的变量进行计数汇总, 统计各类值出现的频次, 常用于数据分组计数和频率分析。
6. <https://dplyr.tidyverse.org/reference/filter.html>: 该 URL 指向的是 dplyr 包中函数 `filter()` 的官方文档页面。该资源详细介绍了 `filter()` 函数的用法和功能,



`filter()`是用于对数据框 (data frame) 进行行筛选的函数, 可以根据指定的条件保留满足条件的行。文档中还说明了 `filter()` 如何处理缺失值 (NA), 以及示例代码演示了常见的过滤操作。该页面属于 tidyverse 生态系统中 dplyr 包的参考手册, 适合 R 语言用户学习和查阅数据筛选相关内容。

7. [https://dplyr.tidyverse.org/reference/group\\_by.html](https://dplyr.tidyverse.org/reference/group_by.html): 该资源是 dplyr 包中函数 `group_by()` 的官方文档页面, 介绍了如何使用 `group_by()` 函数对数据进行分组操作, 从而便于按组进行后续的数据汇总或变换分析。
8. [https://dplyr.tidyverse.org/reference/if\\_else.html](https://dplyr.tidyverse.org/reference/if_else.html): 该 URL ([https://dplyr.tidyverse.org/reference/if\\_else.html](https://dplyr.tidyverse.org/reference/if_else.html)) 指向的是 R 语言中 dplyr 包的 `if_else()` 函数的官方文档页面。该函数用于基于逻辑条件 (logical vector) 进行向量化的条件判断和赋值操作。具体来说, `if_else()` 接受三个主要参数: 一个逻辑条件向量 (condition), 当条件为 TRUE 时返回的值 (true), 以及当条件为 FALSE 时返回的值 (false)。相比于基础 R 的 `ifelse()` 函数, dplyr 的 `if_else()` 在处理缺失值和类型兼容性方面更为严谨, 能提供更有意义的错误信息, 帮助用户避免类型不匹配的问题。该函数是进行条件转换和数据处理时的常用工具, 常与 `case_when()` 函数一起使用, 支持复杂的多条件判断。
9. <https://dplyr.tidyverse.org/reference/mutate.html>: 该资源是 dplyr 包中函数 `mutate()` 的官方文档页面, 详细介绍了 `mutate()` 函数的用法及其在数据框中创建或修改变量的功能。`mutate()` 常用于基于已有变量计算新变量, 是数据变换和处理的重要工具。
10. <https://dplyr.tidyverse.org/reference/near.html>: 该资源是 dplyr 包中函数 `near()` 的官方文档页面, 介绍了 `near()` 函数的用法和原理。`near()` 用于比较数值时忽略微小的数值误差, 解决直接使用等号 (==) 判断浮点数相等时可能出现的失败问题, 方便判断两个数值是否“足够接近”。
11. <https://rdrr.io/pkg/nycflights13/man/flights.html>: 该资源是 R 语言中“nycflights13”包的官方文档页面, 具体介绍了该包中的“flights”数据集。该数据集包含了 2013 年纽约市三大机场出发的航班详细信息, 常用于数据分析和教学示例。页面详细说明了数据集的结构、变量含义及使用方法, 方便用户理解和应用该数据集进行数据操作与分析。

12. <https://rdr.io/r/base/Extremes.html>: 该资源是 R 语言基础包 (base package) 中关于极值函数 (Extremes functions) 的文档页面, 详细介绍了函数如 `min()` 在不同类型数据 (包括逻辑向量) 上的行为和用法, 以及其与逻辑汇总函数的对应关系。
13. <https://rdr.io/r/base/Logic.html>: 该资源是 R 语言基础包 (base package) 中关于逻辑运算 (Logical operations) 的官方文档页面, 详细介绍了 R 语言中用于处理布尔代数的逻辑向量操作符和函数, 例如 “与” (`&`)、 “或” (`|`)、 “非” (`!`) 以及异或函数 `xor()` 的用法和语义, 帮助用户理解和应用逻辑运算符进行数据筛选和条件判断。
14. <https://rdr.io/r/base/MathFun.html>: 该资源是 R 语言基础包 (base) 中 MathFun 函数集合的文档页面, 详细介绍了数学相关函数的用法和示例, 其中包括对 `abs()` 函数的说明和示例实现。
15. <https://rdr.io/r/base/NA.html>: 该 URL (<https://rdr.io/r/base/NA.html>) 指向的是 R 语言基础包中关于函数 `is.na()` 的官方文档或说明页面。`is.na()` 函数用于检测数据中的缺失值 (NA), 返回一个逻辑向量, 标识哪些元素是缺失值。该函数常用于数据清洗和处理过程中, 用以识别和操作包含缺失值的数据。页面内容通常包括函数的用法、参数说明、返回值以及示例代码, 帮助用户理解如何利用 `is.na()` 来检测和管理缺失数据。
16. <https://rdr.io/r/base/all.html>: 该 URL (<https://rdr.io/r/base/all.html>) 指向的是 R 语言基础包 (base package) 中函数 `all()` 的官方文档页面。该页面详细介绍了 `all()` 函数的用法和功能: 用于对逻辑向量进行汇总判断, 只有当逻辑向量中的所有值均为 TRUE 时, `all()` 函数才返回 TRUE; 如果存在任何 FALSE 或缺失值 (NA), 则返回 FALSE 或 NA (可通过参数 `na.rm=TRUE` 忽略缺失值)。文档中还通常包含函数的语法说明、参数解释、示例代码及应用场景。
17. <https://rdr.io/r/base/any.html>: 该 URL (<https://rdr.io/r/base/any.html>) 指向的是 R 语言基础包 (base) 中函数 `any()` 的文档说明页面。该函数用于对逻辑向量进行总结判断: 如果逻辑向量中有任何一个元素为 TRUE, 则返回 TRUE; 否则返回 FALSE。该函数常用于判断一组条件中是否存在至少一个满足的情况,

支持参数如 `na.rm` 用于忽略缺失值 (NA)。文档中通常包含函数的用法、参数说明及示例，帮助用户理解和应用该函数。

18. <https://rdr.io/r/base/c.html>: 该资源是 R 语言基础包中函数 `c()` 的文档页面。`c()` 函数用于将多个元素合并成一个向量，是创建和操作数据的重要工具。
19. <https://rdr.io/r/base/ifelse.html>: 该 URL (<https://rdr.io/r/base/ifelse.html>) 指向的是 R 语言基础包 (base R) 中函数 `ifelse()` 的文档页面。`ifelse()` 函数用于根据条件对向量的元素进行矢量化的条件判断和赋值，常用于对数据进行分类、筛选或转换。例如，可以用它来根据条件判断将日期向量标记为“工作日”或“周末”，或者计算数值向量的绝对值。该函数是 R 语言中实现条件逻辑的基础工具之一，功能类似于 `dplyr` 包中的 `if_else()`，但后者在处理缺失值和类型不兼容时更为严格和灵活。
20. <https://rdr.io/r/base/length.html>: 该资源是 R 语言基础包中函数 `length()` 的在线文档页面，介绍了 `length()` 函数的用法。`length()` 函数用于返回向量或其他对象的长度，即元素的个数。在上下文中，`length()` 函数与 `sum()` 函数配合使用，用于计算逻辑向量中 `TRUE` 值的比例 (`mean()` 函数即为 `sum()` 除以 `length()`)。
21. <https://rdr.io/r/base/mean.html>: 该资源是 R 语言基础包中函数 `mean()` 的文档页面，介绍了 `mean()` 函数的用法及其在逻辑向量 (logical vector) 上的应用。具体来说，`mean()` 函数用于计算数值向量的平均值，而在逻辑向量中，`TRUE` 被视为 1，`FALSE` 被视为 0，因此 `mean()` 可以用来计算逻辑向量中 `TRUE` 值所占的比例 (即 `TRUE` 的平均值)。该页面还说明了 `mean()` 与 `sum()` 和 `length()` 函数的关系，`mean()` 实际上是 `sum()` 除以 `length()`。文档适用于理解如何对逻辑向量进行数值汇总和统计。
22. <https://rdr.io/r/base/print.html>: 该资源是 R 语言基础包 (base) 中函数 `print()` 的官方文档页面，详细介绍了 `print()` 函数的用法及参数说明，包括如何通过 `digits` 参数控制打印数值的精度。
23. <https://rdr.io/r/base/prod.html>: 该资源是 R 语言中基础包 (base package) 中函数 `prod()` 的官方文档页面。页面详细介绍了 `prod()` 函数的功能、用法和返回值，特别说明了当该函数应用于逻辑向量时的行为及其等价的逻辑汇总函数。用户可以通过该页面了解 `prod()` 函数的定义、参数说明以及示例操作。

24. <https://rdr.io/r/base/sum.html>: 该资源是 R 语言基础包中关于函数 `sum()` 的在线文档页面。它介绍了 `sum()` 函数的用法，特别强调在逻辑向量 (logical vector) 中，TRUE 被视为 1，FALSE 被视为 0，因此 `sum(x)` 可以用来计算逻辑向量中 TRUE 的个数。该函数常用于对逻辑向量进行数值汇总，是统计和数据分析中非常实用的工具。

## 第 13 章

1. <https://dplyr.tidyverse.org/reference/arrange.html>: 该资源是 dplyr 包中函数 `arrange()` 的官方文档页面，介绍了如何对数据框进行排序操作。具体而言，该页面说明了 `arrange()` 函数的用法、参数以及示例，帮助用户根据一个或多个变量对数据进行升序或降序排列，从而实现对数据集的有序组织和分析。
2. [https://dplyr.tidyverse.org/reference/consecutive\\_id.html](https://dplyr.tidyverse.org/reference/consecutive_id.html): 该资源是 dplyr 包中函数 `consecutive_id()` 的官方文档页面，介绍了该函数的用法和功能。`consecutive_id()` 用于在数据处理中按指定变量的值变化创建分组标识，每当参数中的值发生变化时就启动一个新的分组，方便对连续相同值的分组操作。该函数常与 `group_by()` 和 `slice_head()` 等函数联合使用，用于提取每组的首行或进行分组统计。
3. <https://dplyr.tidyverse.org/reference/context.html>: 该资源是 dplyr 包中函数 `n()` 的官方文档页面。页面详细介绍了 `n()` 函数的用法和功能，指出它是一个特殊的汇总函数，不接受任何参数，而是访问当前分组的信息，因此只能在 dplyr 的分组语义（如 `group_by()`、`summarise()` 等函数内部）中使用。文档还可能说明了 `n()` 的变体及其在数据汇总和分组计算中的应用场景。
4. <https://dplyr.tidyverse.org/reference/count.html>: 该 URL (<https://dplyr.tidyverse.org/reference/count.html>) 指向的是 R 语言中 dplyr 包的 `count()` 函数的官方文档页面。`count()` 函数用于快速计算数据框中某个变量或变量组合的频数（计数），支持对缺失值计数以及加权计数（通过 `wt` 参数）。它常用于数据探

索和分析中，通过简洁的语法方便地统计分类变量的出现次数，是数据科学中基础且高效的计数工具。文档中还介绍了 `count()` 函数与其他 `dplyr` 函数（如 `group_by()`、`summarise()`、`arrange()`）的配合使用方法。

5. <https://dplyr.tidyverse.org/reference/cumall.html>: 该资源是 `dplyr` 包中关于函数 `cummean()` 的文档页面，介绍了该函数用于计算数值向量的累积平均值 (cumulative means)，作为对 Base R 中累积和 (`cumsum`)、累积积 (`cumprod`)、累积最小值 (`cummin`) 和累积最大值 (`cummax`) 函数的补充，方便进行累积统计分析。
6. <https://dplyr.tidyverse.org/reference/filter.html>: 该资源是 `dplyr` 包中用于数据框过滤的函数 `filter()` 的官方文档页面。它详细介绍了 `filter()` 函数的用法、参数说明和示例，帮助用户基于条件筛选数据集中的行，是 tidyverse 生态中常用的数据操作函数之一。
7. [https://dplyr.tidyverse.org/reference/group\\_by.html](https://dplyr.tidyverse.org/reference/group_by.html): 该资源是 `dplyr` 包中函数 `group_by()` 的官方文档页面。它介绍了如何使用 `group_by()` 函数对数据进行分组操作，从而方便地基于分组计算汇总统计（如结合 `summarize()`、`count()`、`arrange()` 等函数使用），实现按组执行复杂的数据处理和分析任务。该文档详细说明了 `group_by()` 的用法和应用场景，是进行分组数据操作的重要参考资料。
8. <https://dplyr.tidyverse.org/reference/lead-lag.html>: 该资源是 `dplyr` 包中关于函数 `lead()` 和 `lag()` 的官方文档页面，介绍了这两个函数如何用于引用当前值之前或之后的值，返回与输入向量长度相同的结果，并在开头或结尾用 NA 填充，常用于时间序列或顺序数据的延迟和提前值分析。
9. <https://dplyr.tidyverse.org/reference/mutate.html>: 该资源是 `dplyr` 包中 `mutate()` 函数的官方文档页面，介绍了 `mutate()` 的用法及其在数据变换中的应用。`mutate()` 函数用于对数据框添加新变量或修改已有变量，特别适合进行各种数值和其他类型向量的转换操作。文档还说明了 `mutate()` 与其他函数（如 `count()`、`summarize()`）的配合使用，展示了如何灵活应用 `mutate()` 进行数据处理和分组标准化等任务。
10. [https://dplyr.tidyverse.org/reference/percent\\_rank.html](https://dplyr.tidyverse.org/reference/percent_rank.html): 该资源是 `dplyr` 包中函数 `percent_rank()` 的官方文档页面。该页面详细介绍了

`dplyr::percent_rank()` 函数的用法，该函数用于计算数据中每个值的百分等级（百分比排名），常用于数据排序和排名分析，属于一组排名相关函数的变体之一（如 `row_number()`、`dense_rank()`、`cume_dist()` 等）。文档中提供了函数的参数说明、使用示例及其与其他排名函数的区别，帮助用户根据需求选择合适的排名函数。

11. [https://dplyr.tidyverse.org/reference/row\\_number.html](https://dplyr.tidyverse.org/reference/row_number.html): 该资源是 `dplyr` 包中用于排名的函数文档页面，主要介绍了函数 `row_number()` 的用法。  
`row_number()` 函数用于对数据进行排序编号，可在不传入参数时返回当前行号，常用于数据分组或排序操作中。文档中还提到，`row_number()` 是一系列排名函数（如 `min_rank()`、`dense_rank()`、`percent_rank()` 等）中的一种，用户可根据需求选择合适的函数处理排序和排名，特别是在处理并列名次时有不同的方法。该页面详细说明了这些排名函数的功能、用法及处理并列的规则，适合需要按排名对数据进行操作的 R 语言数据分析用户参考。
12. <https://dplyr.tidyverse.org/reference/slice.html>: 该资源是 `dplyr` 包中用于数据操作的函数文档，具体介绍了函数 `slice_head()` 的用法，`slice_head()` 用于从分组或数据集中提取前几行数据，常用于保留每组的第一条记录。文档详细说明了函数参数、用法示例及应用场景。
13. <https://dplyr.tidyverse.org/reference/summarise.html>: 该资源是 `dplyr` 包中关于函数 `summarise()`（或 `summarize()`）的官方文档页面，详细介绍了该函数的用法和应用场景。`summarise()` 函数主要用于对数据框进行汇总计算，通过对分组后的数据执行各种汇总操作生成简洁的统计结果。文档还说明了 `summarise()` 函数与其他函数（如 `group_by()`、`mutate()`）的配合使用方式，以及如何利用汇总函数进行数据变换和分组标准化。
14. [https://ggplot2.tidyverse.org/reference/cut\\_interval.html](https://ggplot2.tidyverse.org/reference/cut_interval.html): 该资源是 `ggplot2` 包中关于函数 `cut_interval()` 的官方文档页面。该函数及其相关函数 `cut_number()` 和 `cut_width()` 是 `ggplot2` 提供的用于将数据分割为区间的辅助工具，主要用于直方图的计算。这些函数虽然归属于 `ggplot2` 包，稍显特殊，但它们在数据分箱处理中非常实用，并且是在 `tidyverse` 其它组件出现之前就已开发。

- 
15. <https://oreil.ly/XPnjF>: (<https://slider.r-lib.org>) 该资源是 R 语言中的一个名为“slider”的软件包, 提供更复杂的滚动 (rolling) 或滑动 (sliding) 聚合计算功能, 适用于时间序列或序列数据的窗口操作。
  16. <https://oreil.ly/YcbwN>: (<https://rdr.io/r/base/Round.html>) 该资源是 R 语言基础包中函数 `round()` 的官方文档页面, 详细介绍了 `round()` 函数的用法及其参数说明。
  17. <https://oreil.ly/Zse1Q>: ([https://en.wikipedia.org/wiki/Eagle\\_County\\_Regional\\_Airport](https://en.wikipedia.org/wiki/Eagle_County_Regional_Airport)) 该 URL 指向的是“Eagle County Regional Airport” (鹰县区域机场) 的维基百科页面, 介绍该机场的基本信息、地理位置、历史背景及相关航空数据。
  18. <https://oreil.ly/swerv>: (<https://stackoverflow.com/questions/27482712>) 该资源是一个 Stack Overflow 上的问题页面, 讨论如何在数据处理中根据条件创建分组变量, 特别是如何在某些值发生变化时启动新的分组。
  19. <https://rdr.io/r/base/Extremes.html>: 该 URL (<https://rdr.io/r/base/Extremes.html>) 指向的是 R 语言基础包 (base) 中与极值相关函数的文档页面。该页面介绍了用于计算最小值和最大值的函数, 主要包括:
    - `pmin()` 和 `pmax()` 函数: 用于比较两个或多个变量的每一行, 分别返回对应位置的最小值或最大值, 适合逐元素的比较操作。
    - `min()` 和 `max()` 函数: 用于对多个观测值整体求最小值或最大值, 返回单一的极值结果, 区别于 `pmin()/pmax()` 的逐元素操作。页面还提及了这些函数的用法区别及适用场景, 帮助用户正确选择和使用极值计算相关函数。
  20. <https://rdr.io/r/base/Log.html>: 该 URL (<https://rdr.io/r/base/Log.html>) 指向的是 R 语言基础包 (base package) 中与对数相关的函数文档, 主要介绍了 R 中用于计算对数的三个函数:
    - `log()`: 计算自然对数 (以 e 为底)
    - `log2()`: 计算以 2 为底的对数
    - `log10()`: 计算以 10 为底的对数

文档详细说明了这三种对数的用途和特点，推荐使用 `log2()` 或 `log10()`，因为 `log2()` 便于解释倍数关系（差值 1 代表原始值翻倍或减半），`log10()` 便于结果的反向转换（如 3 对应  $10^3=1000$ ）。此外，文档还提到了对数函数的逆运算：`log()` 的逆函数是 `exp()`，而 `log2()` 和 `log10()` 的逆运算分别对应 2 的幂和 10 的幂。

总结来说，该资源是 R 语言中对数函数的官方帮助文档，提供函数说明、用法及数学背景，方便用户理解和应用对数变换。

21. <https://rdr.io/r/base/NA.html>: 该资源是 R 语言基础包（base package）中关于 NA 的文档页面，介绍了 NA 的含义及其在 R 中表示缺失值的用法。页面详细说明了如何识别缺失值（NA），以及如何在数据分析中处理和操作缺失数据。
22. <https://rdr.io/r/base/Round.html>: 该 URL (<https://rdr.io/r/base/Round.html>) 指向的是 R 语言中基础包（base package）里 `round()` 函数的官方文档页面。该页面详细介绍了 `round()` 函数的用法，特别说明了其采用的“四舍六入，五取偶”（也称为“银行家舍入”或“round half to even”）的舍入规则，即当数值正好处于两个整数的中间时，会舍入到最近的偶数，从而避免舍入偏差。此外，该页面还将 `round()` 函数与 `floor()` 函数（向下取整）和 `ceiling()` 函数（向上取整）进行了对比说明。
23. <https://rdr.io/r/base/c.html>: 该资源是 R 语言中 base 包的函数 `c()` 的官方文档页面，介绍了 `c()` 函数的用法及相关说明。`c()` 函数用于组合（合并）多个元素或向量，生成一个新的向量，是 R 语言中最基础且常用的函数之一。
24. <https://rdr.io/r/base/cumsum.html>: 该资源 (<https://rdr.io/r/base/cumsum.html>) 是 R 语言基础包（Base R）中关于函数 `cumsum()` 的文档页面。该函数用于计算一个数值向量的累积和（running or cumulative sums），即返回一个向量，其每个元素是输入向量中对应位置及之前所有元素的和。此外，该页面通常也会介绍与累积计算相关的其他函数，如 `cumprod()`（累积乘积）、`cummin()`（累积最小值）和 `cummax()`（累积最大值）。这些函数用于实现数值的累积统计和转换，是数据聚合和数值处理中的常用工具。
25. <https://rdr.io/r/base/cut.html>: 该资源是 R 语言基础包中的函数文档页面，介绍了函数 `cut()` 的用法。该函数用于将数值型向量分割成若干区间（区间分箱），实现对连续数值进行离散化处理，便于分类或分组分析。



- 
26. <https://rdr.io/r/base/mean.html>: 该 URL (<https://rdr.io/r/base/mean.html>) 指向的是 R 语言基础包 (base) 中函数 `mean()` 的在线文档页面。该资源详细介绍了 `mean()` 函数的用法, 包括如何计算数值向量的算术平均值 (均值), 函数的参数说明及其行为特征。`mean()` 函数用于对一组数据求中心位置的数值总结, 但由于它是所有数值的总和除以数量, 因此对极端值 (异常高或低) 较为敏感。该文档通常包含示例代码和相关函数的链接, 方便用户理解和应用 `mean()` 函数进行数据分析。
  27. <https://rdr.io/r/base/mode.html>: 该资源是 R 语言中基础包 (base) 中函数 `mode()` 的在线文档页面, 介绍了 `mode()` 函数的功能及用法。
  28. <https://rdr.io/r/base/rank.html>: 该资源是 R 语言基础包 (base R) 中函数 `rank()` 的官方文档页面, 介绍了 `rank()` 函数的用法及其参数说明, 特别是如何通过调整 `ties.method` 参数处理并列值, 以及如何使用 `na.last` 参数控制对缺失值 (NA) 的处理。
  29. <https://rdr.io/r/base/sum.html>: 该资源是 R 语言中基础包 (base) 提供的函数 `sum()` 的官方文档页面, 介绍了 `sum()` 函数的用法和功能, 主要用于对数值进行求和计算。
  30. <https://rdr.io/r/stats/IQR.html>: 该 URL (<https://rdr.io/r/stats/IQR.html>) 指向的是 R 语言中 `stats` 包内的函数 `IQR()` 的官方文档页面。该函数用于计算数据的四分位距 (Interquartile Range, IQR), 即数据中位于第 75 百分位数与第 25 百分位数之间的范围, 反映数据的中间 50% 分布的离散程度。具体计算方法是用 `quantile(x, 0.75)` 减去 `quantile(x, 0.25)`。该文档详细介绍了 `IQR()` 函数的用法、参数和示例, 帮助用户理解和应用该函数进行数据离散度的统计分析。
  31. <https://rdr.io/r/stats/median.html>: 该资源链接 (<https://rdr.io/r/stats/median.html>) 指向的是 R 语言中用于计算中位数的函数 `median()` 的官方文档页面。该页面介绍了 `median()` 函数的用途, 即用于找到一组数值数据中位于中间位置的值 (50% 的数据点位于其上方, 50% 位于其下方), 作为衡量数据中心趋势的指标。相比于均值 (`mean`), 中位数对极端值 (异常高或低值) 不敏感, 更

适合描述偏态分布的数据特征。文档还可能包含函数的用法说明、参数介绍以及示例代码。

32. <https://rdrr.io/r/stats/quantile.html>: 该资源是 R 语言中用于计算分位数的函数文档页面，介绍了函数 `quantile()` 的用法。`quantile()` 函数用于求取数据的任意分位点值，是中位数的推广。例如，`quantile(x, 0.25)` 返回数据 `x` 中第 25 百分位的值，`quantile(x, 0.5)` 等同于中位数，`quantile(x, 0.95)` 返回第 95 百分位的值。页面详细说明了该函数的功能、参数及应用场景。
33. <https://rdrr.io/r/stats/sd.html>: 该 URL “<https://rdrr.io/r/stats/sd.html>” 指向的是 R 语言中基础统计包 (`stats`) 中函数 `sd()` 的文档页面。该函数用于计算数据的标准差 (standard deviation)，即衡量数据分散程度的指标。页面详细介绍了 `sd()` 函数的用法、参数说明及示例，是了解和使用标准差计算的官方参考资料。
34. [https://readr.tidyverse.org/reference/parse\\_atomic.html](https://readr.tidyverse.org/reference/parse_atomic.html): 该 URL “[https://readr.tidyverse.org/reference/parse\\_atomic.html](https://readr.tidyverse.org/reference/parse_atomic.html)” 指向的是 R 语言中 `tidyverse` 包中的 `readr` 子包的官方文档页面，具体介绍了 `parse_double()` 函数及相关的基本类型解析函数。该资源详细说明了如何将以字符串形式存储的数字解析为数值型 (`double` 类型)，并提供了函数的用法、参数说明及示例，帮助用户正确处理和转换文本数据中的数字信息。
35. [https://readr.tidyverse.org/reference/parse\\_number.html](https://readr.tidyverse.org/reference/parse_number.html): 该资源是 `readr` 包中函数 `parse_number()` 的官方文档页面，介绍了如何从包含非数字字符的字符串中提取并解析数字。该函数适用于需要忽略字符串中非数字文本（例如货币符号或百分号）时，将字符串转换为数值型数据。
36. <https://tibble.tidyverse.org/reference/tribble.html>: 该资源是 `tidyverse` 包中 `tibble` 子包的函数文档，介绍了函数 `tribble()` 的用法。`tribble()` 用于以易读的行格式快速创建 `tibble`（现代版的数据框），方便用户手工输入小型示例数据。

## 第 14 章

1. <https://dplyr.tidyverse.org/reference/arrange.html>: 该资源是 dplyr 包中函数 `arrange()` 的官方文档页面, 介绍了该函数用于对数据框进行排序的功能, 并说明了其支持通过 `locale` 参数来指定排序时的语言环境, 从而正确处理字符串的排序规则。
2. <https://dplyr.tidyverse.org/reference/coalesce.html>: 该资源是 dplyr 包中函数 `coalesce()` 的官方文档页面, 介绍了如何使用 `coalesce()` 函数来替换数据中的缺失值 (missing values), 提供了该函数的用法说明和示例, 帮助用户根据需求在数据处理流程中灵活处理缺失值。
3. <https://dplyr.tidyverse.org/reference/count.html>: 该资源是 R 语言中 dplyr 包的函数文档页面, 具体介绍了函数 `count()` 的用法。`count()` 函数用于统计数据框中某些变量的频数分布, 方便用户快速计算和汇总数据中各类别的出现次数。
4. <https://dplyr.tidyverse.org/reference/filter.html>: 该资源是 dplyr 包中用于数据操作的函数说明页面, 介绍了 R 语言中函数 `filter()` 的用法, 主要用于根据条件筛选数据框中的行。
5. <https://dplyr.tidyverse.org/reference/mutate.html>: 该 URL (<https://dplyr.tidyverse.org/reference/mutate.html>) 指向的是 R 语言中 tidyverse 套件 dplyr 包的 `mutate()` 函数的官方文档页面。`mutate()` 函数用于对数据框 (data frame) 进行变量的新增或修改, 通常用于基于已有列创建新列。文档介绍了 `mutate()` 的用法及其在数据变换中的应用, 尤其强调它与字符串处理函数 (如 stringr 包中的 `str_c()`、`str_glue()` 和 `str_sub()`) 的结合使用, 帮助用户方便地基于数据中的变量生成新的字符串列。
6. <https://dplyr.tidyverse.org/reference/summarise.html>: 该资源是 dplyr 包中函数 `summarize()` (`summarise()` 的别名) 的官方文档页面。`summarize()` 函数用于对数据框进行汇总操作, 通常结合分组操作使用, 将多个观测值归纳为单个汇总值。文中提到它常与 stringr 包中的字符串汇总函数 `str_flatten()` 配合使用, 以便将字符向量合并成单个字符串, 从而实现对字符串数据的汇总。

7. <https://glue.tidyverse.org/reference/glue.html>: 该资源是 R 语言中 tidyverse 生态系统内的 glue 包的官方文档页面, 介绍了 `glue::glue()` 函数的用法。该函数用于方便地进行字符串插值和拼接, 支持在字符串中嵌入 R 表达式, 实现动态字符串生成。
8. <https://oreil.ly/Fs-YL>: ([https://en.wikipedia.org/wiki/Leaning\\_toothpick\\_syndrome](https://en.wikipedia.org/wiki/Leaning_toothpick_syndrome)) 该 URL 指向的资源是维基百科上关于 “Leaning toothpick syndrome” (倾斜牙签综合症) 的条目, 介绍了一种编程中常见的问题, 即在代码中出现大量连续使用反斜杠 ( ) 进行转义, 导致代码难以阅读和维护的现象。
9. <https://oreil.ly/NHBNe>: (<https://glue.tidyverse.org>) 该资源是 “glue” 包的官方网站, 提供了用于字符串拼接和插值的 R 语言工具, 特别支持在字符串中通过花括号 {} 插入变量或表达式, 从而简化字符串构造代码, 使其更易读和维护。
10. [https://oreil.ly/\\_zF3d](https://oreil.ly/_zF3d): (<https://style.tidyverse.org/syntax.html#character-vectors>) 该资源是 tidyverse 风格指南 (style guide) 中关于字符向量 (character vectors) 语法的说明页面, 详细介绍了如何在 R 语言中创建和使用字符串, 包括单引号和双引号的用法, 以及推荐的编码风格规范。
11. <https://oreil.ly/c1P2g>: ([https://en.wikipedia.org/wiki/List\\_of\\_ISO\\_639-1\\_codes](https://en.wikipedia.org/wiki/List_of_ISO_639-1_codes)) 该 URL ([https://en.wikipedia.org/wiki/List\\_of\\_ISO\\_639-1\\_codes](https://en.wikipedia.org/wiki/List_of_ISO_639-1_codes)) 指向维基百科上关于 ISO 639-1 代码列表的页面。该页面提供了标准化的两字母语言代码, 用于表示各种语言, 这些代码常用于指定语言环境 (locale) 中的语言部分, 帮助软件和系统区分不同的语言及其地区变体。
12. <https://oreil.ly/v8ZQf>: (<http://kunststube.net/encoding>) 该资源是一个关于字符编码 (encodings) 详细讲解的网页, 旨在帮助读者深入理解编码的复杂性和相关知识。
13. [https://rdr.io/pkg/stringi/man/stri\\_locale\\_list.html](https://rdr.io/pkg/stringi/man/stri_locale_list.html): 该资源是 R 语言包 stringi 中的函数文档页面, 介绍了函数 `stri_locale_list()`。该函数用于列出所有可用的语言环境 (locale) 标识符, 这些标识符包括语言代码及可选的区域代码, 用于支持字符串处理中的区域依赖功能。通过该函数, 用户可以查看 stringi

包支持的所有 locale 设置，从而方便地选择合适的语言和区域组合来处理非英文文本和区域差异。

14. <https://rdr.io/r/base/Paren.html>: 该资源是 R 语言基础包中关于圆括号 {} 的文档页面，详细说明了 {} 在 R 代码中的作用和用法。在上下文中，URL 指向的是解释如何在字符串插值函数（如 glue 包的 `str_glue()`）中使用花括号 {}，使得括号内的表达式能够被计算并替换为对应的值，从而实现动态拼接字符串的功能。
15. <https://rdr.io/r/base/Quotes.html>: 该资源是 R 语言基础包中关于字符串中引号和特殊字符转义的帮助文档页面，标题为“Quotes”。页面详细介绍了字符串中如何使用双引号（“）、单引号（’）、反斜杠（\）以及换行符（`\n`）、制表符（`\t`）、Unicode 转义（`\u`、`\U`）等特殊字符的用法和示例。用户可以通过该文档了解并掌握在 R 语言中正确处理和表示各种特殊字符的技巧。
16. <https://rdr.io/r/base/paste.html>: 该 URL（<https://rdr.io/r/base/paste.html>）指向的是 R 语言基础包（base R）中字符串连接相关函数的文档页面，主要介绍了函数 `paste()` 和 `paste0()` 的用法及区别。这些函数用于将多个字符串拼接成一个字符串，`paste()` 允许指定分隔符（默认为空格），而 `paste0()` 则是 `paste(sep = “”)` 的简写，直接连接字符串而不添加分隔符。页面还可能包括函数参数说明、示例及与 tidyverse 中字符串处理函数 `str_c()` 的比较。
17. <https://rdr.io/r/base/rawConversion.html>: 该资源是 R 语言基础包中关于字符编码转换的文档页面，具体介绍了函数 `charToRaw()` 的用法，该函数用于查看字符串的底层字节表示，有助于理解和处理非英文文本的编码问题。
18. <https://rdr.io/r/base/writeLines.html>: 该资源是 R 语言基础包（base R）中函数 `writeLines()` 的在线文档页面，介绍了该函数的用法及相关说明。
19. <https://readr.tidyverse.org/reference/encoding.html>: 该资源是 readr 包中关于编码检测函数 `guess_encoding()` 的官方文档页面，介绍了如何使用该函数帮助用户猜测文本文件的字符编码，以便正确读取数据。文档详细说明了 `guess_encoding()` 的用法、限制及适用场景，是处理数据编码问题时的重要参考。
20. [https://stringr.tidyverse.org/reference/str\\_c.html](https://stringr.tidyverse.org/reference/str_c.html): 该 URL 指向的资源是 R 语言中 stringr 包的函数 `str_c()` 的官方文档页面。`str_c()` 函数用于将任意多个

字符向量按元素合并，返回一个新的字符向量。它类似于基础 R 的 `paste0()` 函数，但更适合 tidyverse 风格的数据操作，支持 tidyverse 的回收规则和缺失值处理。该函数常用于结合固定文本与数据框中的变量字符串，配合 `mutate()` 等数据变换函数使用，方便生成新的字符串列。文档中还介绍了与 `str_c()` 相关的其他字符串处理函数，如 `str_glue()` 和 `str_flatten()`，并比较了它们之间的异同。

21. [https://stringr.tidyverse.org/reference/str\\_equal.html](https://stringr.tidyverse.org/reference/str_equal.html): 该资源是 stringr 包中的函数文档页面，介绍了函数 `str_equal()`。该函数用于比较字符串是否相等，区别于直接使用 `==` 运算符，`str_equal()` 能更准确地判断字符串在视觉上的相同性。
22. [https://stringr.tidyverse.org/reference/str\\_flatten.html](https://stringr.tidyverse.org/reference/str_flatten.html): 该 URL 指向的资源是 stringr 包中函数 `str_flatten()` 的官方文档说明。该函数的主要功能是将一个字符向量的所有元素合并成一个单一的字符串，常用于配合 dplyr 包中的 `summarize()` 函数进行字符串汇总操作。与 `str_c()` 和 `str_glue()` 不同，`str_flatten()` 总是返回长度为 1 的字符串，适合用作聚合函数，方便对多个字符串进行连接并生成一个整体结果。
23. [https://stringr.tidyverse.org/reference/str\\_glue.html](https://stringr.tidyverse.org/reference/str_glue.html): 该 URL ([https://stringr.tidyverse.org/reference/str\\_glue.html](https://stringr.tidyverse.org/reference/str_glue.html)) 指向的是 R 语言中 stringr 包里函数 `str_glue()` 的官方文档页面。`str_glue()` 函数来自 glue 包的集成，主要用于简洁地创建字符串，其特色是支持在字符串中直接通过花括号 `{}` 嵌入并计算 R 表达式，从而方便地将固定文本和变量或表达式结果拼接成字符串。与传统的 `str_c()` 相比，`str_glue()` 减少了大量引号和连接符的使用，使代码更直观易读。该函数常用于结合数据框中的变量生成字符串，适合与 dplyr 的 `mutate()` 函数配合使用。需要注意的是，`str_glue()` 会将缺失值 (NA) 转换成字符串 “NA”。
24. [https://stringr.tidyverse.org/reference/str\\_length.html](https://stringr.tidyverse.org/reference/str_length.html): 该资源是 stringr 包中函数 `str_length()` 的官方文档页面，介绍了 `str_length()` 函数的用法和功能。`str_length()` 用于计算字符串中的字符数（字母数），支持处理包含非英语字符和重音符号的字符串，帮助用户准确获取字符串的长度信息。
25. [https://stringr.tidyverse.org/reference/str\\_sub.html](https://stringr.tidyverse.org/reference/str_sub.html): 该资源是 stringr 包中函数 `str_sub()` 的官方文档页面，介绍了该函数用于从字符串中提取子字符串

的功能。文档说明了当字符串长度不足时，`str_sub()` 不会报错，而是尽可能返回可用的部分；同时展示了如何结合 `dplyr` 包的 `mutate()` 函数使用 `str_sub()` 来获取字符串的首尾字母。此外，还涉及了在处理带重音符号的非英语字符时，`str_sub()` 在字符定位上的注意事项。

26. [https://stringr.tidyverse.org/reference/str\\_view.html](https://stringr.tidyverse.org/reference/str_view.html): 该资源是 `stringr` 包中函数 `str_view()` 的官方文档页面。该函数用于在 R 语言环境中以可视化方式展示字符串内容，帮助用户更直观地识别字符串中的特殊字符和不可见字符（如制表符、空白符等），通过高亮和背景色提示字符串中的空白或转义字符，便于调试和理解字符串数据的实际结构。
27. [https://tidyr.tidyverse.org/reference/pivot\\_longer.html](https://tidyr.tidyverse.org/reference/pivot_longer.html): 该资源是 `tidyr` 包中函数 `pivot_longer()` 的官方文档页面，介绍了 `pivot_longer()` 函数的用法和功能。`pivot_longer()` 用于将数据框从宽格式转换为长格式，通过创建新的行，使数据更适合某些数据处理和分析任务。
28. [https://tidyr.tidyverse.org/reference/pivot\\_wider.html](https://tidyr.tidyverse.org/reference/pivot_wider.html): 该资源是 `tidyr` 包中函数 `pivot_wider()` 的官方文档页面，介绍了如何通过生成新列将输入数据框变“宽”，即将数据从长格式转换为宽格式。
29. [https://tidyr.tidyverse.org/reference/separate\\_longer\\_delim.html](https://tidyr.tidyverse.org/reference/separate_longer_delim.html): 该资源是 `tidyr` 包中函数 `separate_longer_delim()` 的官方文档页面，介绍了如何将字符串根据特定分隔符拆分为多行数据，适用于每行组件数量不固定的情况，是实现字符串按分隔符分割并展开为多行的常用方法。
30. [https://tidyr.tidyverse.org/reference/separate\\_wider\\_delim.html](https://tidyr.tidyverse.org/reference/separate_wider_delim.html): 该资源是 `tidyr` 包中函数 `separate_wider_delim()` 的官方文档页面。该函数用于将字符串按照指定的分隔符拆分成固定数量的列，适用于字符串组件数量固定且需要展开成多列的场景。文档介绍了如何通过传入分隔符和列名来使用该函数，并说明了处理拆分时出现“分段过少”或“分段过多”问题的参数 `too_few` 和 `too_many`，帮助用户诊断和解决宽表拆分中的异常情况。此外，文档还提及了与之类似的函数 `separate_wider_position()` 和 `separate_wider_regex()`。

## 第 15 章

1. <https://dplyr.tidyverse.org/reference/across.html>: 该资源是 dplyr 包中用于变量选择的函数文档页，详细介绍了函数 `across()` 的用法。`across()` 是一个“tidyselect”函数，允许用户在 dplyr 语句中同时对多个变量进行操作或变换，常用于如 `mutate()` 或 `summarise()` 中批量应用函数。页面内容包括参数说明、用法示例及与其他选择辅助函数（如 `matches()`）的配合使用方法。
2. <https://dplyr.tidyverse.org/reference/filter.html>: 该资源是 dplyr 包中用于数据筛选的函数 `filter()` 的官方文档页面，介绍了如何根据逻辑条件从数据框中提取满足条件的行。
3. <https://dplyr.tidyverse.org/reference/mutate.html>: 该资源是 dplyr 包中 `mutate()` 函数的官方文档页面。`mutate()` 用于在数据框中创建或修改变量，常用于数据清洗和变换操作，能够基于已有列生成新列，支持灵活的数据处理流程。
4. <https://dplyr.tidyverse.org/reference/rename.html>: 该资源是 dplyr 包中用于重命名数据框变量的函数说明页面，详细介绍了函数 `rename()` 的用法和示例，帮助用户理解如何在数据操作中重命名列名。
5. <https://dplyr.tidyverse.org/reference/select.html>: 该资源是 dplyr 包中关于函数 `select()` 的官方文档页面，详细介绍了如何使用 `select()` 函数来选择数据框中的变量列，包括支持使用类似 `matches(pattern)` 这样的 tidyselect 选择器。文档中还包含函数的用法说明、参数解释和示例，帮助用户灵活地在数据处理中选择所需变量。
6. <https://dplyr.tidyverse.org/reference/summarise.html>: 该资源是 dplyr 包中函数 `summarise()`（或 `summarize()`）的官方文档页面，介绍了该函数的用法和功能。`summarise()` 用于对数据框进行分组汇总计算，生成新的汇总统计值，如总和、均值等，常与其他函数（如 `str_detect()`、`sum()`、`mean()`）配合使用，用于计算满足特定条件的观测数量或比例。
7. <https://oreil.ly/A9Gbl>: (<https://unicode-org.github.io/icu/userguide/strings/regexp.html>) 该 URL (<https://unicode-org.github.io/icu/userguide/strings/regexp.html>) 指向的是 ICU (International



Components for Unicode) 用户指南中关于正则表达式 (regex) 的文档页面。该资源详细介绍了 ICU 正则表达式的语法和用法, 特别强调了其对 Unicode 字符处理的支持和高级特性, 适用于需要在字符串处理中使用复杂 Unicode 范围和特殊正则表达式功能的开发者。

8. <https://oreil.ly/Db3NF>: (<https://regexcrossword.com/challenges/beginner>) 该资源是一个面向初学者的正则表达式填字游戏 (regex crosswords) 网站, 提供入门级别的正则表达式谜题供用户练习和挑战。
9. <https://oreil.ly/MVwoC>: (<https://www.regular-expressions.info/tutorial.html>) 该资源是一个关于正则表达式的详尽教程网站, 提供正则表达式的基础知识、语法规则及高级特性讲解, 适合用来深入学习正则表达式的原理和用法。
10. <https://oreil.ly/VhVuy>: (<https://www.pcre.org>) 该 URL “<https://www.pcre.org>” 指向的是 PCRE (Perl Compatible Regular Expressions) 引擎的官方网站。PCRE 是一个支持 Perl 风格正则表达式语法的正则表达式库, 广泛用于各种编程语言和工具中以实现强大且灵活的模式匹配功能。
11. <https://oreil.ly/abQNx>: (<https://stringi.gagolewski.com>) 该资源 (<https://stringi.gagolewski.com>) 是 R 语言中用于字符串处理的 stringi 包的官方网站或文档页面。stringi 包基于 ICU (国际化组件) 引擎, 实现了强大且灵活的字符串操作和正则表达式功能, 支持复杂的 Unicode 字符处理, 是 stringr 包的底层依赖。
12. <https://oreil.ly/yGQ5U>: (<https://github.com/laurikari/tre>) 该资源 (<https://github.com/laurikari/tre>) 是 TRE (Tagged Regular Expressions) 正则表达式引擎的开源项目仓库。TRE 是一个支持部分正则表达式扩展功能的轻量级匹配引擎, 通常被用作一些编程语言 (如 R 语言中 base 包) 默认的正则表达式处理底层引擎, 支持基本的正则表达式语法和部分高级特性, 适合嵌入式和性能敏感的场景。
13. <https://rdr.io/r/base/Paren.html>: 该资源是 R 语言基础包中关于括号表达式 (如花括号 {}) 的文档页面, 详细说明了如何在正则表达式 (regex) 中使用花括号来精确指定匹配某个模式的次数。

14. <https://rdr.io/r/base/mean.html>: 该资源是 R 语言基础包 (base package) 中函数 `mean()` 的文档说明页面。页面介绍了 `mean()` 函数的用法和功能, `mean()` 用于计算数值向量的算术平均值 (即均值)。在上下文中, `mean()` 函数被用于结合字符串检测函数 `str_detect()`, 计算匹配某模式的观测值的比例。
15. <https://rdr.io/r/base/sum.html>: 该 URL (<https://rdr.io/r/base/sum.html>) 指向的是 R 语言基础包中函数 `sum()` 的文档页面。该函数用于对数值向量中所有元素求和, 常用于统计计算中。例如, 在字符串检测 (`str_detect`) 与数据汇总 (`summarize`) 结合的场景下, `sum()` 可以统计满足某一条件的观测值数量。页面内容还提到, `sum()` 作为基础数学函数, 在 R 语言中广泛应用, 且在文本搜索时可以通过正则表达式边界符 (如 `b`) 准确匹配该函数名称, 避免匹配到其他类似名称。
16. <https://rdr.io/r/grDevices/colors.html>: 该资源是 R 语言中 `grDevices` 包的 `colors()` 函数的文档页面, 介绍了该函数提供的颜色名称及其变体 (如 “lightgray”、 “darkblue” 等)。页面详细列出了可用颜色名称及其修饰词, 便于用户查询和自动识别颜色修饰符。
17. <https://rdr.io/r/utls/data.html>: 该 URL (<https://rdr.io/r/utls/data.html>) 指向的是 R 语言中 `utls` 包中 `data()` 函数的文档页面。该页面详细介绍了 `data()` 函数的用法, 包括如何列出和加载 R 自带的各种数据集, 是查阅和使用基础 R 数据集的重要参考资料。
18. <https://stringr.tidyverse.org/articles/regular-expressions.html>: 该资源是 `stringr` 包中关于正则表达式的官方教程文档 (vignette), 详细介绍了 `stringr` 支持的完整正则表达式语法, 适合作为学习和参考 `stringr` 正则表达式功能的入门资料。
19. <https://stringr.tidyverse.org/reference/case.html>: 该资源是 `stringr` 包中关于字符串大小写转换函数的官方文档页面, 详细介绍了函数如 `str_to_lower()` 等的用法, 用于将字符串转换为小写、大小写等格式。
20. <https://stringr.tidyverse.org/reference/modifiers.html>: 该资源是 `stringr` 包中用于控制字符串匹配行为的辅助函数和参数的文档, 主要介绍了如何通过函数如 `regex()`、`fixed()` 和 `coll()` 来修改正则表达式的匹配方式 (例如忽略大小

写、指定本地化规则等), 以及这些修饰符 (modifiers) 如何影响字符串匹配和处理。

21. [https://stringr.tidyverse.org/reference/str\\_c.html](https://stringr.tidyverse.org/reference/str_c.html): 该资源是 stringr 包中函数 `str_c()` 的官方文档页面, 介绍了该函数用于将多个字符串或字符串向量连接 (拼接) 成一个字符串的用法及示例。
22. [https://stringr.tidyverse.org/reference/str\\_count.html](https://stringr.tidyverse.org/reference/str_count.html): 该资源是 stringr 包中的函数文档页面, 介绍了函数 `str_count()`。该函数用于统计字符串中匹配特定模式 (通常是正则表达式) 的次数, 返回每个字符串中匹配的数量, 而不是简单的真假值 (如 `str_detect()` 函数)。文档还展示了如何将 `str_count()` 与 dplyr 包中的 `mutate()` 函数结合使用, 例如统计字符串中元音和辅音的个数。
23. [https://stringr.tidyverse.org/reference/str\\_detect.html](https://stringr.tidyverse.org/reference/str_detect.html): 该资源是 stringr 包中函数 `str_detect()` 的官方文档页面。`str_detect()` 函数用于检测字符向量中是否包含符合指定正则表达式模式的字符串元素, 返回一个与输入向量长度相同的逻辑值向量 (TRUE 或 FALSE)。它常用于字符串匹配任务, 并且方便与 dplyr 包的 `filter()`、`summarize()` 等函数配合使用, 以筛选、统计或计算匹配模式的字符串数量或比例。此外, `str_detect()` 与 `str_subset()` 和 `str_which()` 等相关函数一起, 构成了 stringr 包中处理字符串匹配的基础工具。
24. [https://stringr.tidyverse.org/reference/str\\_escape.html](https://stringr.tidyverse.org/reference/str_escape.html): 该资源是 stringr 包的函数文档, 介绍了函数 `str_escape()`。该函数用于对字符串中的特殊字符 (元字符) 进行转义处理, 以确保这些字符串在构建正则表达式模式时能够被字面匹配, 而不会被误解析为正则表达式的控制符号。简单来说, `str_escape()` 帮助用户将任意字符串安全地转换为可直接用于匹配的正则表达式文本。
25. [https://stringr.tidyverse.org/reference/str\\_flatten.html](https://stringr.tidyverse.org/reference/str_flatten.html): 该资源是 stringr 包中函数 `str_flatten()` 的官方文档页面, 介绍了如何将字符串向量合并为单一字符串的用法和示例。
26. [https://stringr.tidyverse.org/reference/str\\_length.html](https://stringr.tidyverse.org/reference/str_length.html): 该资源是 stringr 包中函数 `str_length()` 的官方文档页面, 介绍了如何计算字符串的长度。
27. [https://stringr.tidyverse.org/reference/str\\_match.html](https://stringr.tidyverse.org/reference/str_match.html): 该资源是 R 语言中 stringr 包的函数文档页面, 介绍了函数 `str_match()` 的用法。`str_match()` 用于从

字符串中提取匹配的子串及其分组匹配内容，返回一个矩阵格式的结果。该页面详细说明了函数的参数、返回值及示例代码，帮助用户理解和使用该函数进行正则表达式匹配提取操作。

28. [https://stringr.tidyverse.org/reference/str\\_remove.html](https://stringr.tidyverse.org/reference/str_remove.html): 该资源是 stringr 包中函数 `str_remove()` 的官方文档页面，介绍了 `str_remove()` 及其变体 `str_remove_all()` 用于从字符串中移除匹配的子串，实质上是对 `str_replace(x, pattern, “ ”)` 的简便封装。
29. [https://stringr.tidyverse.org/reference/str\\_replace.html](https://stringr.tidyverse.org/reference/str_replace.html): 该资源是 stringr 包中函数 `str_replace()` 的官方文档页面。它介绍了如何使用 `str_replace()` 函数通过正则表达式匹配并替换字符串中的内容。具体来说，`str_replace()` 用于替换字符串中第一个匹配项，而相应的 `str_replace_all()` 则替换所有匹配项。文档还说明了该函数支持使用反向引用（back references）来灵活修改匹配结果。
30. [https://stringr.tidyverse.org/reference/str\\_subset.html](https://stringr.tidyverse.org/reference/str_subset.html): 该资源是 stringr 包中的函数文档页面，介绍了函数 `str_subset()`。该函数用于从字符向量中筛选出所有匹配指定模式（正则表达式或字符串）的元素，返回仅包含匹配字符串的字符向量。它与函数 `str_detect()` 和 `str_which()` 密切相关，后者分别用于判断匹配结果和返回匹配元素的位置索引。
31. [https://stringr.tidyverse.org/reference/str\\_view.html](https://stringr.tidyverse.org/reference/str_view.html): 该资源是 R 语言中 stringr 包的函数文档页面，介绍了函数 `str_view()` 的用法。`str_view()` 用于可视化字符串向量中与正则表达式匹配的部分，通过将匹配的文本用尖括号 `< >` 包围并高亮显示，帮助用户理解正则表达式模式如何作用于字符串。该函数常用于学习和调试正则表达式匹配效果。
32. [https://stringr.tidyverse.org/reference/str\\_which.html](https://stringr.tidyverse.org/reference/str_which.html): 该资源是 stringr 包中函数 `str_which()` 的官方文档页面。该函数用于在字符串向量中查找匹配特定模式的字符串，并返回这些匹配字符串的位置索引（整数向量）。它与 `str_detect()` 和 `str_subset()` 函数密切相关，后者分别用于检测匹配和返回匹配的字符串本身。

33. <https://stringr.tidyverse.org/reference/stringr-data.html>: 该资源是 `stringr` 包中关于内置数据集 `stringr::words` 的文档页面, 介绍了该数据集包含的常用单词语料库, 用于在字符串处理中构建和测试正则表达式。
34. [https://tidyr.tidyverse.org/reference/separate\\_longer\\_delim.html](https://tidyr.tidyverse.org/reference/separate_longer_delim.html): 该资源是 R 语言包 `tidyr` 中函数 `separate_longer_delim()` 的官方文档页面, 详细介绍了该函数的用法和参数说明。该函数用于根据指定的分隔符 (`delim`) 将字符串拆分成更长格式的数据列, 支持使用固定字符串或正则表达式 (通过 `regex()`) 作为分隔符, 从而灵活处理字符串拆分需求。
35. [https://tidyr.tidyverse.org/reference/separate\\_wider\\_delim.html](https://tidyr.tidyverse.org/reference/separate_wider_delim.html): 该 URL ([https://tidyr.tidyverse.org/reference/separate\\_wider\\_delim.html](https://tidyr.tidyverse.org/reference/separate_wider_delim.html)) 指向的是 R 语言 `tidyr` 包中函数 `separate_wider_delim()` 的官方文档页面。根据上下文推断, `separate_wider_delim()` 函数用于将数据框 (data frame) 中某一列的字符串按照指定的分隔符拆分成多列。它通常与 `separate_wider_regex()` 和 `separate_wider_position()` 函数一起使用, 三者都是 `tidyr` 包中用于“拆分”的函数, 区别在于拆分依据不同: `separate_wider_delim()` 基于分隔符, `separate_wider_regex()` 基于正则表达式, `separate_wider_position()` 基于固定位置。该函数支持通过指定分隔符 (`delim`) 将字符串拆分成多个变量列, 且可以结合正则表达式的灵活匹配来处理复杂的拆分需求。文档还提到, 若拆分失败, 可以使用参数如 `too_short = "debug"` 来调试错误。总体来说, 该页面详细介绍了 `separate_wider_delim()` 的用法、参数及其在数据整理中的应用。

## 第 16 章

1. <https://dplyr.tidyverse.org/reference/count.html>: 该资源是 `dplyr` 包中函数 `count()` 的官方文档页面, 介绍了如何使用 `count()` 函数对数据框 (如 `tibble`) 中的变量进行计数和分组汇总, 特别适用于查看因子 (`factor`) 变量的各个水平及其对应的频数。

2. <https://dplyr.tidyverse.org/reference/mutate.html>: 该资源是 dplyr 包中用于数据变换的函数 `mutate()` 的官方文档页面, 介绍了如何通过 `mutate()` 对数据框进行新变量的添加或已有变量的修改, 以便进行更加复杂的数据转换操作。
3. <https://forcats.tidyverse.org/reference/fct.html>: 该资源是 R 语言中 tidyverse 生态系统下的 forcats 包的官方文档, 具体介绍了 `forcats::fct()` 函数的用法。该函数用于因子 (factor) 变量的处理, 尤其是在因子水平排序时提供比字母排序更安全、合理的排序方式, 通常按因子水平首次出现的顺序进行排序。
4. [https://forcats.tidyverse.org/reference/fct\\_collapse.html](https://forcats.tidyverse.org/reference/fct_collapse.html): 该资源是 tidyverse 包中 forcats 子包的函数文档, 详细介绍了 R 语言中的 `fct_collapse()` 函数。该函数用于将因子 (factor) 变量中的多个水平 (levels) 合并 (折叠) 为较少的新水平, 方便对因子变量进行简化和重编码操作。文档中包含函数用法说明、参数解释及示例。
5. [https://forcats.tidyverse.org/reference/fct\\_inorder.html](https://forcats.tidyverse.org/reference/fct_inorder.html): 该资源是 tidyverse 包中 forcats 子包的函数文档, 具体介绍了函数 `fct_inorder()` 的用法。该函数用于按因子水平在数据中出现的顺序对因子水平进行重新排序, 方便在绘制条形图时保持因子水平的原始顺序。
6. [https://forcats.tidyverse.org/reference/fct\\_lump.html](https://forcats.tidyverse.org/reference/fct_lump.html): 该资源是 forcats 包中关于函数 `fct_lump_x()` 系列的官方文档, 介绍了如何将因子 (分类变量) 中较小的类别合并为 “Other” 类别, 以简化数据的展示和分析。文档详细说明了多个相关函数 (如 `fct_lump_lowfreq()`、`fct_lump_n()`、`fct_lump_min()` 和 `fct_lump_prop()`) 的用法和参数设置, 帮助用户根据不同需求灵活地对因子类别进行合并和分组。
7. [https://forcats.tidyverse.org/reference/fct\\_recode.html](https://forcats.tidyverse.org/reference/fct_recode.html): 该资源是 forcats 包中函数 `fct_recode()` 的官方文档页面。`fct_recode()` 函数用于对因子变量的水平 (levels) 进行重新编码或修改, 可以改变因子水平的名称, 从而简化标签、便于展示和出版。该函数支持对指定的水平进行重命名, 未提及的水平保持不变, 并且在引用不存在的水平时会发出警告。文档详细介绍了该函数的用法、参数和示例, 是处理和转换因子水平的强大工具。

8. [https://forcats.tidyverse.org/reference/fct\\_relevel.html](https://forcats.tidyverse.org/reference/fct_relevel.html): 该资源是关于 R 语言中 tidyverse 包中的函数 `fct_relevel()` 的官方文档说明。`fct_relevel()` 用于调整因子 (factor) 变量的水平顺序, 可以将指定的一个或多个因子水平移动到最前面, 便于对因子水平进行重新排序和管理。
9. [https://forcats.tidyverse.org/reference/fct\\_reorder.html](https://forcats.tidyverse.org/reference/fct_reorder.html): 该资源是 R 语言中 forcats 包的函数文档, 详细介绍了函数 `fct_reorder()`。该函数用于重新排序因子 (factor) 变量的水平 (levels), 以便更好地展示数据, 比如改善绘图的可读性。文档说明了函数的用途、参数以及使用场景, 强调应避免对已有合理顺序的因子使用该函数, 而适用于那些水平顺序任意的因子。
10. [https://forcats.tidyverse.org/reference/fct\\_rev.html](https://forcats.tidyverse.org/reference/fct_rev.html): 该资源是 R 语言中 forcats 包的函数文档页面, 介绍了函数 `fct_rev()` 的用法。`fct_rev()` 用于反转因子 (factor) 变量的水平顺序。结合 `fct_infreq()` 函数使用时, 可以将因子水平按频率升序排列, 从而在条形图中将最大值显示在右侧。
11. [https://forcats.tidyverse.org/reference/gss\\_cat.html](https://forcats.tidyverse.org/reference/gss_cat.html): 该资源是 R 语言包 forcats 中数据集 `gss_cat` 的官方文档页面。`gss_cat` 数据集是从美国“通用社会调查” (General Social Survey, GSS) 中精选的一组示例数据, 包含若干变量, 用于演示和处理因子 (factor) 类型数据时常见的问题和操作。该页面提供了数据集的详细描述、变量说明及使用示例, 方便用户理解和应用该数据集。
12. <https://ggplot2.tidyverse.org/reference/aes.html>: 该资源是 ggplot2 包中函数 `aes()` 的官方文档页面, 详细介绍了 `aes()` 函数的用法和参数, `aes()` 用于在 ggplot2 中定义美学映射 (如坐标轴变量、颜色、形状等), 帮助用户构建可视化图形。
13. <https://oreil.ly/3qBI5>: (<https://gss.norc.org>) 该资源是“General Social Survey” (通用社会调查, 简称 GSS) 的官方网站, 由芝加哥大学下属的独立研究机构 NORC 负责实施和维护。该网站提供了关于 GSS 数据集的信息、访问途径及相关研究资源, 支持用户访问和利用这一长期开展的美国社会调查数据。
14. [https://oreil.ly/J\\_IIg](https://oreil.ly/J_IIg): (<https://forcats.tidyverse.org/reference/index.html>) 该资源是 forcats 包的官方函数参考索引页面, 提供了 forcats 包中所有

函数的详细说明和用法，方便用户查找和使用该包中处理因子变量的各种工具函数。

15. <https://oreil.ly/Z9mkP>: (<https://simplystatistics.org/posts/2015-07-24-stringsasfactors-an-unauthorized-biography>) 该资源是一篇题为“stringsAsFactors: An unauthorized biography”的文章，介绍了 R 语言中字符串自动转换为因子（stringsAsFactors）这一机制的历史背景和演变过程。文章深入探讨了这一默认行为的起因、影响及争议，为理解 R 中分类数据处理提供了背景知识。
16. <https://oreil.ly/phWQo>: (<https://notstatschat.tumblr.com/post/124987394001/stringsasfactors-sigh>) 该资源是一个名为“stringsAsFactors = ”的文章，发布在 notstatschat.tumblr.com 博客上，内容涉及 R 语言中因子（factor）数据类型的历史背景和使用困境，重点讨论了字符串转因子（stringsAsFactors）这一特性及其引发的问题和用户的无奈情绪。文章为理解 R 中分类数据处理提供了背景知识，并为后续的 tidyverse 包如 forcats 的开发提供了参考依据。
17. <https://oreil.ly/zPh8E>: (<https://peerj.com/preprints/3163>) 该资源是一篇由 Amelia McNamara 和 Nicholas Horton 撰写的学术论文，题为“Wrangling categorical data in R”，详细介绍了 R 语言中处理分类数据的方法，涵盖了分类因子（factors）的历史背景、base R 与 tidyverse 方法的对比，并对 categorical data 的整理和管理进行了系统阐述。这篇论文对理解和使用 forcats 包具有重要参考价值。
18. <https://rdr.io/r/base/factor.html>: 该 URL (<https://rdr.io/r/base/factor.html>) 指向的是 R 语言基础包（base package）中关于函数 factor() 和 ordered() 的官方文档页面。该资源详细介绍了因子（factor）这一数据结构的定义和用途，说明了如何使用 factor() 创建因子变量，以及如何通过 ordered() 创建具有顺序关系的有序因子。页面内容涵盖因子的基本操作、因子水平的设置和排序，以及因子在数据分析中的应用，特别是分类变量的处理方法。



19. <https://rdr.io/r/base/levels.html>: 该资源是 R 语言中 base 包的官方文档页面, 具体介绍了函数 `levels()` 的用法及相关信息。该函数用于获取因子 (factor) 对象的有效水平 (levels) 集合。
20. [https://readr.tidyverse.org/reference/parse\\_factor.html](https://readr.tidyverse.org/reference/parse_factor.html): 该资源是 readr 包中用于解析因子 (factor) 数据类型的函数文档页面, 具体介绍了函数 `parse_factor()` 的用法及参数说明, 帮助用户在读取数据时将某列直接转换为因子类型。

## 第 17 章

1. [https://lubridate.tidyverse.org/reference/as\\_date.html](https://lubridate.tidyverse.org/reference/as_date.html): 该 URL ([https://lubridate.tidyverse.org/reference/as\\_date.html](https://lubridate.tidyverse.org/reference/as_date.html)) 指向的是 R 语言中 lubridate 包的函数 `as_date()` 的官方文档页面。该页面介绍了 `as_date()` 函数的用法, 主要用于将数值型的日期时间数据 (通常表示为从 Unix 纪元时间 1970-01-01 起的天数偏移量) 转换为日期对象。文档还提到该函数与 `as_datetime()` 配合使用, 后者用于处理以秒为单位的时间偏移, 方便用户在日期和日期时间之间进行转换。
2. <https://lubridate.tidyverse.org/reference/day.html>: 该资源是 R 语言中 lubridate 包的函数文档页面, 主要介绍日期时间访问器函数中的与“天”相关的函数, 如 `mday()` (获取月份中的某一天)、`yday()` (获取年份中的某一天) 和 `wday()` (获取星期中的某一天)。页面详细说明了如何提取日期的各个组成部分, 并支持通过参数返回星期或月份的名称 (可缩写或全称)。这些函数用于方便地从日期时间对象中提取具体的天信息, 是构建和解析日期时间数据的重要工具。
3. <https://lubridate.tidyverse.org/reference/hour.html>: 该资源是 R 语言中 lubridate 包的函数文档, 具体介绍了函数 `hour()` 的用法。`hour()` 用于从日期时间对象中提取小时部分, 方便用户获取时间的小时信息。该函数属于一组时间组

件访问器函数（如 `year()`、`month()`、`minute()` 等），用于分解和操作日期时间数据。

4. [https://lubridate.tidyverse.org/reference/make\\_datetime.html](https://lubridate.tidyverse.org/reference/make_datetime.html): 该资源是 R 语言中 lubridate 包的函数文档页面，介绍了函数 `make_datetime()` 的用法。该函数用于根据指定的年月日时分秒等组件，创建一个日期时间（date-time）对象。文档还提到，`make_datetime()` 与一组获取日期时间各组成部分（如年、月、日、时、分、秒）的函数相对应，是构造日期时间数据的重要工具。
5. <https://lubridate.tidyverse.org/reference/minute.html>: 该资源是 R 语言包 lubridate 中的函数 `minute()` 的官方文档页面，介绍了 `minute()` 函数的用法及功能。`minute()` 函数用于从日期时间对象中提取分钟部分，是 lubridate 提供的一组访问日期时间各组成部分（如年、月、日、小时、分钟、秒）的函数之一。该函数的功能与 `make_datetime()` 函数相对，后者用于构造日期时间对象。
6. <https://lubridate.tidyverse.org/reference/month.html>: 该资源是 R 语言中 lubridate 包的函数文档，介绍了函数 `month()`。该函数用于从日期时间对象中提取月份部分，支持返回月份的数字形式或通过设置参数 `label = TRUE` 返回月份的缩写名称，且可通过 `abbr = FALSE` 返回完整月份名称。
7. <https://lubridate.tidyverse.org/reference/now.html>: 该 URL (<https://lubridate.tidyverse.org/reference/now.html>) 指向的是 R 语言中 lubridate 包的官方文档页面，具体介绍了函数 `now()` 和 `today()` 的用法。这些函数用于获取当前的日期或日期-时间值，文档中还解释了如 `tzzone`（时区）参数的作用及其重要性。
8. <https://lubridate.tidyverse.org/reference/period.html>: 该资源是 lubridate 包中关于 `period` 函数的官方文档页面，详细介绍了如何使用 `period` 类及相关函数（如 `days()`）来表示和操作时间周期（时间段），以便进行日期时间的计算和转换。
9. [https://lubridate.tidyverse.org/reference/round\\_date.html](https://lubridate.tidyverse.org/reference/round_date.html): 该资源是 R 语言中 lubridate 包的函数文档页面，介绍了用于将日期时间对象舍入到临近时间单位的函数，包括 `floor_date()`（向下舍入）、`round_date()`（四舍五入）和 `ceiling_date()`（向上舍入）。这些函数接受一个日期向量和一个时间单位参数，

方便对日期进行按指定单位的舍入操作，常用于时间序列数据的处理和可视化，例如按周统计航班数量等。

10. <https://lubridate.tidyverse.org/reference/second.html>: 该资源是 R 语言中 lubridate 包的函数文档，介绍了函数 `second()` 的用法。`second()` 用于提取或设置日期时间对象中的“秒”部分，是用于获取时间中秒数的访问器函数之一。它与 `make_datetime()` 函数相对应，后者用于构造完整的日期时间对象。
11. <https://lubridate.tidyverse.org/reference/year.html>: 该资源是 R 语言中 lubridate 包的官方文档页面，介绍了函数 `year()` 的用法。`year()` 函数用于从日期时间对象中提取年份部分，是一类“访问器”函数之一，方便用户获取日期的年、月、日、时、分、秒等组成部分。该页面详细说明了 `year()` 函数的功能、用法及示例，帮助用户进行日期时间数据的处理和分析。
12. <https://lubridate.tidyverse.org/reference/ymd.html>: 该资源是 R 语言中 lubridate 包的官方文档页面，具体介绍了 `ymd()` 函数及其相关函数的用法，这些函数用于解析和创建日期对象。文档还说明了如何通过向函数名后加上下划线和时间单位（如“h”、“m”、“s”）来创建日期时间对象。
13. <https://oreil.ly/19K7t>: (<https://xkcd.com/1179>) 该资源是一个指向 xkcd 漫画编号 1179 的链接，内容与 ISO 8601 日期时间标准相关。该漫画通过幽默的方式解释或讽刺了 ISO 8601 标准在表示日期和时间时的格式规范。
14. <https://oreil.ly/NwvsT>: (<https://www.iana.org/time-zones>) 该资源是 IANA（互联网号码分配局）维护的时区数据库官方网站，提供全球各地时区的详细信息和历史变更记录。该数据库采用城市命名方式来准确反映各地时区及其历史上的调整，以支持计算机系统正确处理不同时区的时间数据。
15. <https://rdr.io/r/base/c.html>: 该资源介绍了 R 语言中基础包（base package）中的函数 `c()`，该函数用于将多个对象合并成一个向量。在上下文中，`c()` 函数被提及用于合并日期时间对象时，可能会丢失时区信息，从而使合并后的日期时间显示为第一个元素的时区。该页面详细说明了 `c()` 函数的用法及行为。
16. <https://rdr.io/r/base/timezones.html>: 该资源是 R 语言基础包中的时间相关函数文档页面，详细介绍了获取当前系统时区（如函数 `Sys.timezone()`）以及列出所有时区名称（如函数 `OlsonNames()`）的方法和使用说明。

17. <https://rdr.io/r/stats/update.html>: 该资源是 R 语言中 stats 包的函数 `update()` 的在线文档页面, 介绍了如何使用 `update()` 函数创建或修改日期时间对象, 并支持一次性设置多个字段的用法说明。
18. [https://readr.tidyverse.org/reference/date\\_names.html](https://readr.tidyverse.org/reference/date_names.html): 该资源是 tidyverse 中 readr 包的函数文档页面, 介绍了与日期名称相关的功能, 具体包括内置的多语言日期名称列表 (通过函数 `date_names_langs()` 获取) 以及自定义日期名称的创建方法 (通过函数 `date_names()` 实现)。用户可以参考该文档来查看支持的语言名称, 或学习如何为非英语环境自定义日期名称, 从而配合 `locale()` 函数正确处理非英文日期格式。
19. <https://readr.tidyverse.org/reference/locale.html>: 该资源是 readr 包中关于函数 `locale()` 的官方文档页面, 介绍了如何设置和使用本地化参数 (locale), 尤其是在处理非英语日期格式 (如使用 `%b` 或 `%B` 格式符时) 时, 如何指定合适的语言环境以正确解析和显示日期字符串。
20. [https://readr.tidyverse.org/reference/parse\\_datetime.html](https://readr.tidyverse.org/reference/parse_datetime.html): 该资源是 readr 包中用于解析日期和日期时间的函数文档, 介绍了如何使用 `col_date()` 和 `col_datetime()` 函数结合自定义的日期时间格式字符串来读取和转换数据中的日期时间字段。文档详细说明了日期时间格式的语法规则 (如 `%Y-%m-%d` 表示 “年-月-日” 格式), 并列出了所有支持的格式选项。

## 第 18 章

1. <https://dplyr.tidyverse.org/reference/coalesce.html>: 该资源是 dplyr 包中函数 `coalesce()` 的官方文档页面, 介绍了如何使用 `coalesce()` 函数来替换缺失值 (missing values), 通常将缺失值替换为某个固定且已知的值 (如 0)。
2. <https://dplyr.tidyverse.org/reference/count.html>: 该资源是 R 语言中 dplyr 包的官方文档页面, 介绍了函数 `count()` 的用法。 `count()` 用于统计数据集中各分

组的观测数量，支持通过参数如 `.drop = FALSE` 来保留所有分组（包括数据中未出现的分组），便于进行分组计数和缺失组的处理。

3. <https://dplyr.tidyverse.org/reference/filter-joins.html>: 该资源是 dplyr 包中关于“filter joins”函数的官方文档页面，详细介绍了如 `anti_join()` 等用于筛选连接的函数。这些函数用于根据匹配关系选择数据框中特定的行，例如 `anti_join()` 用于筛选出左表中没有与右表匹配的行，帮助用户识别数据中的缺失或不匹配部分。
4. [https://dplyr.tidyverse.org/reference/group\\_by.html](https://dplyr.tidyverse.org/reference/group_by.html): 该资源是 dplyr 包中 `group_by()` 函数的官方文档页面，介绍了如何使用 `group_by()` 对数据进行分组操作，并说明了通过参数 `.drop = FALSE` 可以保留所有因子水平（factor levels）。
5. <https://dplyr.tidyverse.org/reference/mutate-joins.html>: 该资源是 dplyr 包中关于“mutate joins”的文档，主要介绍如何使用 dplyr 的连接(join)函数（如 `full_join()`）来合并数据框，从而实现类似 `complete()` 函数的数据补全操作，即通过连接生成包含所有需要行的完整数据集。
6. [https://dplyr.tidyverse.org/reference/na\\_if.html](https://dplyr.tidyverse.org/reference/na_if.html): 该资源是 dplyr 包中函数 `na_if()` 的官方文档页面，介绍了如何将数据中的特定值替换为缺失值（NA），常用于在数据读取后，将某些特殊值（如“99”）转换为缺失值以便后续处理。
7. <https://dplyr.tidyverse.org/reference/select.html>: 该资源是 dplyr 包中函数 `select()` 的官方文档页面，详细介绍了 `select()` 函数的用法和功能。`select()` 用于从数据框中选择指定的列，支持通过列名、位置或辅助函数灵活地筛选和重排列，从而方便地提取和操作数据的子集。
8. <https://rdrr.io/r/base/Extremes.html>: 该 URL “<https://rdrr.io/r/base/Extremes.html>” 指向的是 R 语言基础包（base package）中与极值相关的函数文档页面，主要介绍了函数如 `max()` 和 `min()` 的用法与行为，特别是在处理零长度向量（empty vectors）时的返回值（如返回 `-Inf` 和 `Inf`），以及这些函数在统计汇总中的应用特点。
9. [https://readr.tidyverse.org/reference/read\\_delim.html](https://readr.tidyverse.org/reference/read_delim.html): 该资源是 readr 包中函数 `read_delim()` 的官方文档页面，介绍了如何使用该函数读取分隔符分隔的

数据文件。文档详细说明了函数的参数及用法，特别是如何通过参数（如 `na`）在读取数据时处理缺失值，方便用户高效导入和预处理数据。

10. <https://tidyr.tidyverse.org/reference/complete.html>: 该资源是 `tidyr` 包中 `complete()` 函数的官方文档页面。`complete()` 函数用于通过指定变量的所有组合，生成显式的缺失值（missing values），从而补全数据集中的缺失行。例如，可以用 `complete()` 确保数据中所有年份（year）和季度（qtr）的组合均存在，即使原始数据中缺少某些组合。该函数支持直接使用已有变量名，也支持用户自定义变量取值范围来补全数据。此外，当简单变量组合无法生成完整观测时，也可以通过手动构造完整的数据框，再与原数据集进行合并来实现类似效果。`complete()` 常用于在汇总数据后，将隐含的缺失值转换为显式缺失，方便后续分析和处理。
11. <https://tidyr.tidyverse.org/reference/fill.html>: 该资源是 R 语言 `tidyr` 包中函数 `fill()` 的官方文档页面。它介绍了 `fill()` 函数的用法，该函数用于填补数据框中缺失的值，通常按列进行操作，方便数据清洗和预处理。

## 第 19 章

1. <https://dplyr.tidyverse.org/reference/count.html>: 该资源是 `dplyr` 包中函数 `count()` 的官方文档页面，介绍了如何使用 `count()` 函数对数据框中的主键进行计数，从而检查主键是否唯一，帮助用户验证数据表中每条记录的唯一性。
2. <https://dplyr.tidyverse.org/reference/filter-joins.html>: 该资源是 `dplyr` 包中关于“过滤连接”（filter joins）函数的官方文档，主要介绍了两个函数：`semi_join()` 和 `anti_join()`，它们用于基于另一张表的匹配条件来筛选数据集中的行。文档详细说明了这两个过滤连接的用法和原理，帮助用户理解如何通过过滤连接实现数据的子集选择。
3. [https://dplyr.tidyverse.org/reference/join\\_by.html](https://dplyr.tidyverse.org/reference/join_by.html): 该资源是 `dplyr` 包中用于数据表连接（join）操作的函数说明页面，具体介绍了 `join_by()` 函数。

`join_by()` 提供了一种更清晰且灵活的方式来指定连接条件，支持复杂的键值匹配规则（如复合键和不等式条件），从而避免因默认匹配列含义不同而导致的错误或缺失匹配。

4. <https://dplyr.tidyverse.org/reference/mutate-joins.html>: 该 URL “<https://dplyr.tidyverse.org/reference/mutate-joins.html>” 指向的是 R 语言中 dplyr 包关于 “mutate joins”（变异连接）函数的官方文档页面。该页面详细介绍了四种主要的变异连接函数：`left_join()`、`inner_join()`、`right_join()` 和 `full_join()`。
5. <https://dplyr.tidyverse.org/reference/mutate.html>: 该资源是 dplyr 包中函数 `mutate()` 的官方文档页面，介绍了 `mutate()` 函数的用法。`mutate()` 用于在数据框中添加或修改变量，常用于数据变换和新变量的创建。文档中通常会包含函数的详细说明、参数解释、使用示例及相关注意事项。
6. <https://rdr.io/r/utills/View.html>: 该 URL 指向的是 R 语言中 utills 包中函数 `View()` 的文档页面。`View()` 函数用于在 RStudio 中以表格视图的形式打开数据框或其他对象，方便用户查看和浏览数据内容。

## 第 20 章

1. [https://docs.ropensci.org/writexl/reference/write\\_xlsx.html](https://docs.ropensci.org/writexl/reference/write_xlsx.html): 该资源是 R 语言中 writexl 包的官方文档页面，详细介绍了函数 `write_xlsx()` 的用法。该函数用于将数据写入 Excel 格式的文件，实现将数据框导出为 .xlsx 文件的功能。
2. [https://dplyr.tidyverse.org/reference/bind\\_rows.html](https://dplyr.tidyverse.org/reference/bind_rows.html): 该资源是 dplyr 包中函数 `bind_rows()` 的官方文档页面，介绍了如何使用 `bind_rows()` 函数将多个数据框按行合并的用法和示例。
3. [https://googlesheets4.tidyverse.org/reference/gsheet\\_auth.html](https://googlesheets4.tidyverse.org/reference/gsheet_auth.html): 该资源是 R 语言包 googlesheets4 中函数 `gsheet_auth()` 的官方文档页面，介绍了如何通过该函数进行 Google 账号的身份认证，包括指定特定的 Google 账号、设置认证权限范围等，以便安全地访问和操作 Google Sheets。

4. [https://googlesheets4.tidyverse.org/reference/gsheet\\_create.html](https://googlesheets4.tidyverse.org/reference/gsheet_create.html): 该资源为 R 语言包 `googlesheets4` 中函数 `gsheet_create()` 的官方文档页面, 介绍了如何使用该函数创建新的 Google 电子表格。
5. [https://googlesheets4.tidyverse.org/reference/gsheet\\_examples.html](https://googlesheets4.tidyverse.org/reference/gsheet_examples.html): 该资源是 `googlesheets4` R 包中的一个示例文档页面, 提供了使用函数 `gsheet_example()` 的示例和示范数据, 用于帮助用户快速定位和加载随包提供的示例 Google 表格文件, 以便进行测试和学习如何操作 Google Sheets 数据。
6. [https://googlesheets4.tidyverse.org/reference/range\\_read.html](https://googlesheets4.tidyverse.org/reference/range_read.html): 该 URL ([https://googlesheets4.tidyverse.org/reference/range\\_read.html](https://googlesheets4.tidyverse.org/reference/range_read.html)) 指向的是 R 语言包 `googlesheets4` 中的函数文档, 主要介绍了函数 `read_sheet()` (也称为 `range_read()`)。该函数用于从 Google Sheets 通过 URL 或文件 ID 读取电子表格数据, 并返回一个 `tibble` 格式的数据框。它支持类似于 `readxl` 包中 `read_excel()` 函数的参数设置, 例如自定义列名、指定缺失值字符串、定义列类型, 以及读取电子表格中的指定数据范围。该函数使得用户能够方便地在 R 中直接导入和操作 Google Sheets 数据, 实现与 Excel 数据读取类似的工作流程。
7. [https://googlesheets4.tidyverse.org/reference/sheet\\_properties.html](https://googlesheets4.tidyverse.org/reference/sheet_properties.html): 该资源是关于函数 `sheet_names()` 的官方文档页面, 介绍如何获取 Google 表格中所有工作表名称的相关信息和用法。
8. [https://googlesheets4.tidyverse.org/reference/sheet\\_write.html](https://googlesheets4.tidyverse.org/reference/sheet_write.html): 该资源是 `googlesheets4` R 语言包中的一个函数文档页面, 介绍了函数 `sheet_write()` (或相关函数 `write_sheet()`) 的用法。该函数用于将 R 中的数据框写入到 Google Sheets 表格中, 支持创建新表或向已有表写入数据, 是实现 R 与 Google Sheets 数据交互的重要工具。
9. <https://oreil.ly/03oQy>: (<https://docs.google.com/spreadsheets/d/1yc5gL-a2O0Br8M7B3IsDNX5uR17vBHOyWZq6xSTG2G8>) 该 URL 指向的是一个 Google 电子表格, 内容很可能是一个与文中提到的 Excel 文件 (`survey.xlsx`) 相同或类似的数据集, 方便用户在线查看或下载使用。
10. <https://oreil.ly/CU5XP>: (<https://nacnudus.github.io/tidymodels>) 该资源是一个名为 “`tidymodels`” 的软件包, 主要用于处理 Excel 文件中非传统的表格数据, 比如



通过单元格背景颜色或字体加粗等格式信息来存储的数据，方便用户提取和整理这些复杂的 Excel 数据。

11. <https://oreil.ly/E4dIi>: ([https://docs.google.com/spreadsheets/d/1LgZ0Bkg9d\\_NK8uTdP2uHXm07kAlwx8-Ictf8NocebIE](https://docs.google.com/spreadsheets/d/1LgZ0Bkg9d_NK8uTdP2uHXm07kAlwx8-Ictf8NocebIE)) 该 URL 指向的是一个 Google Sheets 电子表格，内容很可能是一个类似于“roster.xlsx”中的数据集，用于替代或补充文中提到的 Excel 文件。换言之，该文档大概包含一个人员名单或相关数据表，方便用户在线查看或下载使用。
12. <https://oreil.ly/Ejuen>: (<https://doi.org/10.1080/00031305.2017.1375989>) 该资源是一篇由 Karl Broman 和 Kara Woo 撰写的学术论文，题为《Data Organization in Spreadsheets》。论文聚焦于电子表格中数据组织的最佳实践，旨在帮助用户更高效、规范地整理数据，从而在将数据导入 R 语言进行分析和可视化时避免常见问题，提升数据处理的准确性和便捷性。
13. <https://oreil.ly/G28nV>: (<https://googlesheets4.tidyverse.org/articles/auth.html>) 该资源是 googlesheets4 包的认证 (authentication) 使用指南，详细介绍了如何通过浏览器登录谷歌账号以授权访问 Google Sheets，以及如何使用函数 `gs4_auth()` 来指定特定账户、认证范围等高级认证配置。该文档为用户提供了对认证流程的深入说明，帮助用户正确设置和管理访问 Google Sheets 的权限。
14. <https://oreil.ly/Gzphe>: (<https://docs.ropensci.org/writexl>) 该资源是 R 语言包 writexl 的官方文档网站，提供有关 writexl 包及其函数（如 `write_xlsx()`）的详细说明和使用示例，用于将数据写入 Excel 文件。
15. <https://oreil.ly/JtH0t>: (<https://ycphs.github.io/openxlsx>) 该资源 (<https://ycphs.github.io/openxlsx>) 是 R 语言中 openxlsx 包的官方网站或项目主页，提供了关于 openxlsx 包的详细文档和说明。openxlsx 包用于生成和操作 Excel 文件，支持在一个 Excel 文件中写入多个工作表，并且具备丰富的格式化和样式设置功能，适合需要对 Excel 输出进行高级定制的用户使用。
16. <https://oreil.ly/PYENq>: (<https://pos.it/r4ds-survey>) 该资源是一个在线调查问卷，URL “<https://pos.it/r4ds-survey>” 指向一个与数据收集相关的网页，可能用于收集包含宠物数量 (`n_pets`) 等变量的调查数据。

- 
17. <https://oreil.ly/VMlBY>: (<https://developers.google.com/sheets/api>) 该资源是 Google 官方提供的 Sheets API v4 的开发者文档, 介绍了如何使用 Google Sheets API 进行程序化访问和操作 Google 电子表格的数据。
  18. <https://oreil.ly/c7DEP>: ([https://docs.google.com/spreadsheets/d/1V1nPp1tzOuutXFLb3G9Eyx3qxeEhnOXUzL5\\_BcCQ0w](https://docs.google.com/spreadsheets/d/1V1nPp1tzOuutXFLb3G9Eyx3qxeEhnOXUzL5_BcCQ0w)) 该 URL 指向的是一个 Google Sheets 电子表格文件, 通常用于通过函数如 `read_sheet()` 读取数据, 并将其导入为 R 语言中的 tibble 格式数据框。
  19. <https://oreil.ly/clwtE>: (<https://ycphs.github.io/openxlsx/articles/Formatting.html>) 该资源是 openxlsx 包的官方文档或相关文章页面, 详细介绍了如何使用 openxlsx 包在 R 语言中对写入 Excel 文件的数据进行格式化和样式设置, 提供了丰富的功能说明和示例, 帮助用户实现更复杂的 Excel 输出效果。
  20. <https://oreil.ly/jNskS>: (<https://nacnudus.github.io/spreadsheet-munging-strategies>) 该资源 “<https://nacnudus.github.io/spreadsheet-munging-strategies>” 是一个关于处理和转换电子表格数据的策略和方法的网页, 特别是针对那些数据以非传统方式存储 (例如通过单元格背景颜色或文本格式等非表格化信息) 的情况, 提供了实用的技巧和工具参考, 帮助用户更有效地清洗和整理 Excel 中的复杂数据。
  21. <https://oreil.ly/m6q7i>: ([https://docs.google.com/spreadsheets/d/1oCqdXUNO8JR3Pca8fHfiz\\_WXWxMuZAp3YiYFaKze5V0](https://docs.google.com/spreadsheets/d/1oCqdXUNO8JR3Pca8fHfiz_WXWxMuZAp3YiYFaKze5V0)) 该 URL 链接的资源很可能是一个 Google Sheets 电子表格文件, 内容与示例中的 Excel 数据集类似, 供用户查看或下载使用, 方便用户创建或参考相应的销售数据表格。
  22. <https://oreil.ly/qgKTY>: (<https://pos.it/r4ds-penguins>) 该资源是一个包含企鹅数据的 Google 电子表格, 用户可以从其中读取特定工作表 (如 “Torgersen Island”) 以进行数据分析或展示。
  23. <https://oreil.ly/sAjBM>: (<https://pos.it/r4ds-roster>) 该资源是一个 Google Sheet 文档, 标题为 “roster”, 用于存储和展示某个团队或组织的成员名单数据。
  24. [https://rdr.io/pkg/janitor/man/clean\\_names.html](https://rdr.io/pkg/janitor/man/clean_names.html): 该资源是 R 语言包 janitor 中函数 `clean_names()` 的官方文档页面, 介绍了如何使用

`clean_names()` 函数将数据框的列名转换为符合规范的蛇形命名 (`snake_case`), 以便后续数据处理更加方便。

25. [https://readr.tidyverse.org/reference/read\\_delim.html](https://readr.tidyverse.org/reference/read_delim.html): 该 URL ([https://readr.tidyverse.org/reference/read\\_delim.html](https://readr.tidyverse.org/reference/read_delim.html)) 指向的是 R 语言中 `readr` 包的官方文档页面, 具体介绍了 `read_delim()` 函数。该函数用于读取以定界符分隔的文本数据文件 (如 CSV、TSV 等), 其语法与 `read_csv()`、`read_table()` 等类似, 支持灵活地指定文件的列类型和其他读取参数, 方便用户导入结构化的平面文本数据。
26. [https://readr.tidyverse.org/reference/read\\_table.html](https://readr.tidyverse.org/reference/read_table.html): 该资源是 tidyverse 中 `readr` 包的官方文档页面, 介绍了函数 `read_table()`。该函数用于读取以空白字符 (如空格或制表符) 分隔的纯文本表格数据文件, 提供类似于其他读取函数 (如 `read_csv()`) 的语法和用法, 方便用户将表格数据导入 R 进行处理和分析。
27. [https://readxl.tidyverse.org/reference/excel\\_sheets.html](https://readxl.tidyverse.org/reference/excel_sheets.html): 该资源是 R 语言包 `readxl` 中函数 `excel_sheets()` 的官方文档页面, 介绍如何获取 Excel 文件中所有工作表的名称, 以使用户选择并读取感兴趣的工作表数据。
28. [https://readxl.tidyverse.org/reference/read\\_excel.html](https://readxl.tidyverse.org/reference/read_excel.html): 该 URL ([https://readxl.tidyverse.org/reference/read\\_excel.html](https://readxl.tidyverse.org/reference/read_excel.html)) 指向的是 R 语言中 `readxl` 包的官方文档页面, 详细介绍了 `read_excel()` 函数。该函数用于读取 Excel 格式的电子表格文件, 支持 .xls 和 .xlsx 两种格式, 且能够自动识别文件类型。文档说明了 `read_excel()` 的主要参数, 如用于指定读取文件路径的第一个参数、选择读取的工作表 (`sheet` 参数)、以及设置列类型 (`col_types` 参数) 等。函数默认将读取的数据导入为 `tibble` 格式, 方便后续数据处理。此外, 文档还对比了 `read_excel()` 与其他数据读取函数 (如 `readr` 包中的 `read_csv()`) 的相似用法, 强调了该函数在导入 Excel 数据时的灵活性和便利性。
29. [https://readxl.tidyverse.org/reference/readxl\\_example.html](https://readxl.tidyverse.org/reference/readxl_example.html): 该资源是 `readxl` 包中函数 `readxl_example()` 的官方文档页面, 介绍了如何通过该函数获取示例电子表格文件在本地系统中的路径, 以使用户在使用 `read_excel()` 函数时调用该示例文件进行操作和测试。

## 第 21 章

1. <https://dbi.r-dbi.org/reference/dbConnect.html>: 该资源是 R 语言中 DBI 包的函数文档, 详细介绍了函数 `DBI::dbConnect()` 的用法。该函数用于创建与数据库管理系统 (DBMS) 的连接, 文档中说明了第一个参数用于选择具体的数据库类型, 后续参数用于指定连接数据库所需的位置和访问凭证, 帮助用户了解如何通过该函数建立数据库连接。
2. <https://dbi.r-dbi.org/reference/dbGetQuery.html>: 该资源是 R 语言 DBI 包中函数 `dbGetQuery()` 的官方文档页面, 介绍了如何使用该函数通过 SQL 语句在数据库中执行查询, 并获取查询结果返回到 R 环境中。
3. <https://dbi.r-dbi.org/reference/dbReadTable.html>: 该资源是关于 R 语言 DBI 包中函数 `dbReadTable()` 的官方文档页面。该函数用于从数据库中读取指定表的全部内容, 并将其作为一个 `data.frame` 返回, 方便用户在 R 环境中对数据库表数据进行操作和分析。
4. <https://dbi.r-dbi.org/reference/dbWriteTable.html>: 该资源是 R 语言 DBI 包中函数 `dbWriteTable()` 的官方文档页面, 介绍了如何使用该函数将数据框 (`data frame`) 写入数据库中, 创建新的数据库表。文档详细说明了函数的基本用法, 主要参数包括数据库连接对象、目标表名以及要写入的数据框, 适用于加载和存储数据至数据库。
5. <https://dbplyr.tidyverse.org/articles/translation-function.html>: 该资源是 `dbplyr` 包中关于函数翻译机制的详细介绍文档 (vignette)。它说明了 `dbplyr` 如何将常见的字符串处理和日期时间操作的 R 函数自动转换为对应数据库的 SQL 语句, 帮助用户在使用数据库查询时无需手动编写 SQL。文档还提到, 虽然翻译并不完美, 且部分 R 函数尚未支持, 但 `dbplyr` 对大多数常用函数的翻译效果已相当不错。
6. <https://dplyr.tidyverse.org/reference/arrange.html>: 该资源是 `dplyr` 包中函数 `arrange()` 的官方文档页面, 介绍了如何使用 `arrange()` 对数据框中的行进行排序。文档说明了 `arrange()` 与 SQL 中 `ORDER BY` 子句的对应关系, 强调了在数据处理中排序的重要性, 特别是在 SQL 查询中由于表无固有顺序, 使用 `arrange()` 可

以保证结果集的行顺序一致。此外，文档还提到窗口函数中需要单独指定排序，因为主查询的 `ORDER BY` 不会自动应用于窗口函数。

7. [https://dplyr.tidyverse.org/reference/case\\_when.html](https://dplyr.tidyverse.org/reference/case_when.html): 该资源是 dplyr 包中函数 `case_when()` 的官方文档页面，介绍了该函数的用法和示例。  
`case_when()` 是一种类似 SQL 语句中 CASE WHEN 的条件判别函数，用于根据多个条件返回对应的值，常用于数据变换和条件分类。
8. <https://dplyr.tidyverse.org/reference/compute.html>: 该资源是关于 R 语言中 dplyr 包的函数 `collect()` 的官方文档页面。文档详细介绍了 `collect()` 函数的用法及其工作原理：该函数用于将数据库中的查询结果取回到 R 内存中，生成相应的 SQL 语句，调用底层的 `dbGetQuery()` 函数执行查询，然后将结果转换为 tibble（数据框的一种现代形式），方便后续在 R 中进行数据分析和处理。
9. <https://dplyr.tidyverse.org/reference/context.html>: 该资源是 dplyr 包官方文档中关于“context”（上下文）部分的说明页面，主要介绍在 dplyr 中如何处理和翻译函数（如 `n()`）的执行环境和上下文，特别是在将 R 代码转换为 SQL 表达式时函数的行为与实现细节。
10. <https://dplyr.tidyverse.org/reference/desc.html>: 该资源为 dplyr 包中函数 `desc()` 的官方文档页面，介绍了 `desc()` 函数的用法及其功能，特别说明了该函数名称来源于 SQL 中的 DESC 关键字，用于指定排序的降序排列。
11. <https://dplyr.tidyverse.org/reference/distinct.html>: 该 URL (<https://dplyr.tidyverse.org/reference/distinct.html>) 指向的是 R 语言中 dplyr 包中函数 `distinct()` 的官方文档页面。该页面详细介绍了 `distinct()` 函数的用法，该函数用于从数据框或数据库表中提取唯一（不重复）的行，常用于数据去重操作。文档中通常包含函数的参数说明、示例代码以及函数的返回值描述。
12. <https://dplyr.tidyverse.org/reference/explain.html>: 该资源是 dplyr 包中函数 `show_query()` 的官方文档页面，介绍了如何使用该函数查看 dplyr 生成的 SQL 代码，帮助用户理解 dplyr 语句与对应 SQL 语句之间的关系，从而促进学习 SQL。
13. <https://dplyr.tidyverse.org/reference/filter.html>: 该 URL 链接指向的是 R 语言包 dplyr 中函数 `filter()` 的官方文档页面。根据上下文信息，`filter()` 函数

用于对数据进行行筛选，在数据库操作时它会被转换成 SQL 中的 WHERE 子句，用于限定查询条件。此外，当对聚合后生成的变量进行筛选时，`filter()` 会对应生成 SQL 中的 HAVING 子句，因为 WHERE 子句在 SQL 执行顺序中先于 SELECT 和 GROUP BY，而 HAVING 则用于聚合后的条件筛选。该文档详细介绍了 `filter()` 函数的用法及其在数据库查询中如何映射到 SQL 语法的细节。

14. [https://dplyr.tidyverse.org/reference/group\\_by.html](https://dplyr.tidyverse.org/reference/group_by.html): 该资源是 dplyr 包中函数 `group_by()` 的官方文档页面，介绍了该函数的用法及其在数据操作中如何实现类似 SQL 语句中 GROUP BY 子句的功能。
15. [https://dplyr.tidyverse.org/reference/if\\_else.html](https://dplyr.tidyverse.org/reference/if_else.html): 该资源是 dplyr 包中函数 `if_else()` 的官方文档页面，介绍了该函数的用法、参数及示例。`if_else()` 是 dplyr 中用于条件判断和矢量化替代的函数，类似于 SQL 中的 CASE WHEN 语句，常用于根据条件返回不同结果。
16. <https://dplyr.tidyverse.org/reference/lead-lag.html>: 该资源是 dplyr 包中关于窗口函数 `lead()` 和 `lag()` 的官方文档页面，详细介绍了这两个函数的用法和功能。`lead()` 用于获取数据中“下一条”或“后续”的值，而 `lag()` 用于获取“上一条”或“之前”的值，常用于时间序列或顺序数据的比较和计算。
17. <https://dplyr.tidyverse.org/reference/mutate-joins.html>: 该资源是 dplyr 包中关于“连接 (join)”函数的官方文档，详细介绍了如 `inner_join()`、`right_join()` 和 `full_join()` 等函数的用法。这些函数名称与 SQL 中的连接操作紧密对应，文档帮助用户理解如何在 R 语言中使用 dplyr 执行类似 SQL 的数据表连接操作。
18. <https://dplyr.tidyverse.org/reference/mutate.html>: 该资源是 dplyr 包中 `mutate()` 函数的官方文档页面，详细介绍了 `mutate()` 的用法和功能。`mutate()` 用于在数据框中添加新变量或修改已有变量，类似于 SQL 中 SELECT 语句中对列的计算和转换操作。文档中还讲解了 `mutate()` 在与 SQL 语句的对应关系，例如如何生成 SELECT 子句中的表达式，以及在包含汇总函数时如何转换为窗口函数 (window functions) 的细节。该页面适合希望理解 dplyr 中数据变换与 SQL 语法对应关系的用户。

19. <https://dplyr.tidyverse.org/reference/relocate.html>: 该资源是 dplyr 包中函数 `relocate()` 的官方文档页面, 介绍了 `relocate()` 函数的用法。`relocate()` 函数用于重新排列数据框中的列位置, 改变某些列在结果中的显示顺序, 而不修改列的内容或名称。它在 SQL 查询的 SELECT 子句中对应的是调整列的顺序, 帮助用户方便地控制数据列的排列方式。
20. <https://dplyr.tidyverse.org/reference/rename.html>: 该资源是 dplyr 包中关于函数 `rename()` 的官方文档页面。该页面详细介绍了 `rename()` 函数的用法, 功能是为了重命名数据框 (data frame) 中的列名。`rename()` 函数属于 dplyr 包中与 SQL 中 SELECT 子句功能相似的一组函数, 主要用于调整列的名称和位置, 方便数据的处理和分析。
21. <https://dplyr.tidyverse.org/reference/select.html>: 该资源是 dplyr 包中函数 `select()` 的官方文档页面, 详细介绍了 `select()` 函数的用法。`select()` 函数用于从数据框中选择 (或重排) 列, 类似于 SQL 查询中的 SELECT 子句, 能够指定需要保留或重新排列的数据列, 是数据操作中常用的列选择工具。
22. <https://dplyr.tidyverse.org/reference/summarise.html>: 该 URL (<https://dplyr.tidyverse.org/reference/summarise.html>) 所指向的资源是关于 R 语言中 dplyr 包的函数 `summarise()` (或 `summarize()`) 的官方文档。该函数主要用于对数据框进行汇总操作, 通常与 `group_by()` 函数配合使用, 将 R 中的 `summarise()` 操作翻译为 SQL 中的 SELECT 子句, 实现对分组数据的统计汇总, 如计算均值、总和等指标。文档详细说明了 `summarise()` 函数的用法、参数、返回值以及它在数据库查询中的对应关系, 是学习数据汇总和数据库操作时的重要参考资料。
23. <https://dplyr.tidyverse.org/reference/tbl.html>: 该资源是 dplyr 包中函数 `tbl()` 的官方文档页面, 介绍了如何使用 `tbl()` 函数创建一个表示数据库表的对象, 通常用于配合 `dbplyr` 包操作数据库。
24. <https://generics.r-lib.org/reference/setops.html>: 该 URL (<https://generics.r-lib.org/reference/setops.html>) 指向的是 R 语言中 `generics` 包的文档页面, 具体介绍了集合操作相关的函数接口 (如 `intersect()` 等)。这些函数用

于执行集合的交集、并集、差集等操作，是 R 中通用的集合操作泛型函数的参考说明。

25. <https://oreil.ly/-0Usp>: (<https://www.practicalsql.com>) 该资源是一本名为《Practical SQL》的书籍官方网站，作者是 Anthony DeBarros。该网站提供关于如何将数据导入数据库、使用数据库管理系统（DBMS）进行数据操作的实用指导，特别适合数据新闻工作者和数据科学领域的读者。
26. <https://oreil.ly/9Dq5p>: (<https://dtplyr.tidyverse.org>) 该资源是 dtplyr 的官方网站，dtplyr 是 tidyverse 生态系统中的一个 R 语言包，作为 dplyr 的后端，将 dplyr 语法翻译成高效的 data.table 代码，从而提升数据处理的性能。网址“<https://dtplyr.tidyverse.org>”提供了 dtplyr 的文档、使用指南和相关资源。
27. <https://oreil.ly/A8OGW>: (<https://dbplyr.tidyverse.org/reference>) 该资源是 dbplyr 包的官方函数参考手册页面，提供了 dbplyr 中所有函数及其用法的详细说明，帮助用户了解和使用 dbplyr 提供的数据库操作接口。
28. <https://oreil.ly/PTwQz>: (<https://modern-sql.com/concept/three-valued-logic>) 该资源是一篇由 Markus Winand 撰写的文章，介绍了 SQL 中的三值逻辑（Three-Valued Logic），特别讲解了 SQL 中 NULL 值的行为和处理方式。
29. <https://oreil.ly/QfAat>: (<https://sqlfordatascientists.com>) 该资源是《SQL for Data Scientists》一书的官方网站，提供专为数据科学家设计的 SQL 学习内容，涵盖处理现实组织中复杂互联数据的示例和教程。
30. <https://oreil.ly/gmDpk>: (<https://multidplyr.tidyverse.org>) 该资源是“multidplyr”的官方网站，提供了一个 R 语言包，用于将 dplyr 代码并行执行在多核处理器上，从而加速数据处理任务。
31. <https://oreil.ly/k3EaP>: (<https://r-datatable.com>) 该 URL (<https://r-datatable.com>) 对应的资源是关于 R 语言中 data.table 包的官方网站或相关介绍网站。data.table 是 R 语言中一个高效的数据处理包，提供快速的数据操作能力，特别适用于大数据集的处理。该网站可能包含 data.table 的使用教程、文档、示例代码及相关资源，帮助用户理解和掌握 data.table 的功能。
32. <https://oreil.ly/tVS8h>: (<https://cynkra.github.io/dm>) 该资源是一个名为 dm 的软件包（package），由 Tobias Schieferdecker、Kirill Müller 和



Darko Bergant 开发。该包主要用于处理数据库中高度规范化的多表数据，能够自动识别表之间的连接关系（基于数据库管理员提供的约束条件），并帮助用户可视化表之间的关系网络，同时生成连接多表所需的关联操作（joins），极大简化了多表数据整合和分析的过程。

33. <https://oreil.ly/xgmg8>: (<https://github.com/tidyverse/dbplyr/issues>) 该资源是 dbplyr 项目的 GitHub 问题反馈页面，用户可以在此提交和浏览与 dbplyr 相关的问题和建议，帮助开发团队改进数据库翻译和兼容性。具体链接为：<https://github.com/tidyverse/dbplyr/issues>。
34. <https://rdr.io/r/base/cut.html>: 该资源是 R 语言基础包中函数 `cut()` 的官方文档页面，介绍了 `cut()` 函数的用法和功能。`cut()` 函数用于将连续变量按指定的区间划分为若干类别，常用于将数值数据分段分组，以便进行分类处理或统计分析。
35. <https://rdr.io/r/base/library.html>: 该资源是 R 语言中基础包（base package）相关文档的网页，具体介绍了 R 语言中用于加载包的函数 `library()`。页面内容通常包括该函数的用法说明、参数解释及示例，帮助用户理解如何在 R 环境中加载和管理软件包。
36. <https://rdr.io/r/base/mean.html>: 该资源是 R 语言基础包（base）中函数 `mean()` 的官方文档页面，详细介绍了 `mean()` 函数的用法、参数说明以及示例，主要用于计算数值向量的算术平均值。
37. <https://rdr.io/r/stats/median.html>: 该资源是 R 语言中 `median()` 函数的官方文档页面，详细介绍了该函数的用法、参数说明以及计算中位数的相关细节。该函数用于计算数值数据的中位数，是统计分析中常用的汇总函数之一。
38. <https://rdr.io/r/utils/head.html>: 该资源是 R 语言中 `utils` 包的函数文档页面，具体介绍了函数 `head()` 的用法和功能。`head()` 函数用于返回对象（如向量、数据框等）的前几行或前几个元素，默认返回前 6 项，常用于快速查看数据的开头部分。
39. [https://tibble.tidyverse.org/reference/as\\_tibble.html](https://tibble.tidyverse.org/reference/as_tibble.html): 该资源是 R 语言中 tidyverse 包 `tibble` 模块的官方文档页面，介绍了函数 `as_tibble()` 的用法。该函数用于将数据结构（如 `data.frame`）转换为 `tibble`，从而使数据打印输出更加美观和易于阅读。

40. [https://tidyr.tidyverse.org/reference/pivot\\_longer.html](https://tidyr.tidyverse.org/reference/pivot_longer.html): 该资源是 tidyr 包中函数 `pivot_longer()` 的官方文档页面, 介绍了 `pivot_longer()` 函数的用法和功能。`pivot_longer()` 用于将数据从宽格式转换为长格式, 是数据整理和重塑的重要工具。
41. [https://tidyr.tidyverse.org/reference/pivot\\_wider.html](https://tidyr.tidyverse.org/reference/pivot_wider.html): 该资源是 tidyr 包中函数 `pivot_wider()` 的官方文档页面, 介绍了该函数的用法和功能。`pivot_wider()` 用于将数据从“长格式”转换为“宽格式”, 即将某些键值对展开为多列, 便于数据的整形和重构。

## 第 22 章

1. <https://arrow.apache.org/docs/r/reference/acero.html>: 该资源是 Apache Arrow R 语言接口中关于 Acero 组件的官方文档页面, 详细列出了 Acero 当前支持的所有操作和函数, 用于指导用户在 R 语言中使用 Arrow 进行数据处理时, 了解可用的表达式和函数集合。
2. [https://arrow.apache.org/docs/r/reference/open\\_dataset.html](https://arrow.apache.org/docs/r/reference/open_dataset.html): 该资源为 Apache Arrow 官方文档中 R 语言接口部分关于函数 `open_dataset()` 的说明页面。该页面详细介绍了如何使用 `arrow::open_dataset()` 函数高效打开和访问大型数据集 (如 CSV 或 Parquet 文件), 避免一次性将整个数据加载到内存中, 而是通过扫描部分数据获取元数据信息, 并按需读取数据, 从而实现对大规模数据的灵活处理和分析。
3. [https://arrow.apache.org/docs/r/reference/to\\_duckdb.html](https://arrow.apache.org/docs/r/reference/to_duckdb.html): 该资源是 Apache Arrow 官方文档中关于 R 语言接口函数 `to_duckdb()` 的说明页面。该函数用于将 Arrow 数据集无拷贝地转换为 DuckDB 数据库, 实现了在 Arrow 生态系统内不同计算环境之间的高效无缝数据传递。

4. [https://arrow.apache.org/docs/r/reference/write\\_dataset.html](https://arrow.apache.org/docs/r/reference/write_dataset.html): 该资源是 Apache Arrow R 包中函数 `write_dataset()` 的官方文档页面, 介绍了如何使用该函数将数据集写入指定目录, 并支持通过参数设置文件的存储格式和分区方式。
5. <https://dplyr.tidyverse.org/reference/compute.html>: 该资源是 dplyr 包中函数 `collect()` 的官方文档页面, 介绍了如何使用 `collect()` 函数将延迟计算的数据操作 (如在 Apache Arrow 数据上的 dplyr 代码) 执行并将结果以数据框形式返回到 R 环境中。文档详细说明了 `collect()` 在将查询从抽象的 dplyr 操作转换为实际计算并获取结果过程中的作用和用法。
6. [https://dplyr.tidyverse.org/reference/group\\_by.html](https://dplyr.tidyverse.org/reference/group_by.html): 该资源是 dplyr 包中函数 `group_by()` 的官方文档页面, 介绍了如何使用 `group_by()` 函数对数据进行分组操作, 从而定义数据的分区 (partition)。
7. <https://oreil.ly/CIE7D>: (<https://parquet.apache.org>) 该资源是 Apache Parquet 的官方网站, 介绍了 Parquet 格式——一种基于开放标准的列式存储文件格式, 广泛应用于大数据系统中, 具有高效的数据存储和读取能力, 常作为 CSV 等格式的高性能替代方案。
8. <https://oreil.ly/OzpFo>: (<https://parquet.apache.org/docs/file-format/data-pages/encodings>) 该资源介绍了 Parquet 文件格式中用于数据页面的编码方法, 重点讲解了如何通过高效的编码技术减少文件大小, 从而提升数据存储和读写的性能。
9. <https://oreil.ly/TGrH5>: (<https://arrow.apache.org>) 该资源是 Apache Arrow 的官方网站, Apache Arrow 是一个多语言工具箱, 旨在高效分析和传输大型数据集, 支持通过 arrow 软件包使用 dplyr 语法处理超大内存的数据, 具有高速性能。
10. <https://oreil.ly/g60F8>: (<https://arrow.apache.org/docs/r>) 该资源是 Apache Arrow 项目的官方文档, 专门介绍如何在 R 语言中使用 Apache Arrow 包 (arrow package), 包括利用 arrow 提供的 dplyr 后端功能来高效分析和处理大规模数据集。
11. <https://oreil.ly/kACzC>: (<https://hive.apache.org>) 该 URL (<https://hive.apache.org>) 指向的是 Apache Hive 项目的官方网站。Apache Hive 是一个用

于大数据处理的数据仓库软件框架，能够提供数据摘要、查询和分析功能，常用于在分布式存储系统（如 Hadoop）上管理和查询大规模数据集。上下文中提到的“Hive-style partitions”即指 Hive 采用的分区命名规范（key=value），用于高效组织和查询数据文件。

12. <https://oreil.ly/u56DR>: (<https://data.seattle.gov/Community/Checkouts-by-Title/tmmm-ytt6>) 该资源是西雅图公共图书馆的图书借阅数据集，记录了从 2005 年 4 月到 2022 年 10 月期间，每本书每个月的借阅次数，共包含约 4,138 万条记录，数据来源于 Seattle Open Data 平台，供公众查询和分析图书借阅情况。
13. <https://pillar.r-lib.org/reference/glimpse.html>: 该资源是 R 语言中 pillar 包的函数 `glimpse()` 的官方文档页面，介绍了 `glimpse()` 函数的用法及功能。  
`glimpse()` 用于快速查看数据框或类似数据结构的内容和结构，展示行数、列数以及部分数值，方便用户了解数据的基本情况。
14. [https://readr.tidyverse.org/reference/read\\_delim.html](https://readr.tidyverse.org/reference/read_delim.html): 该资源是 R 语言中 tidyverse 包下 readr 库的函数文档页面，详细介绍了函数 `read_delim()` 的用法。  
`read_delim()` 用于读取以自定义分隔符分隔的文本数据文件（如 CSV、TSV 等），支持高效地将大规模数据导入 R 环境，适合处理结构化的文本数据。

## 第 23 章

1. <https://dplyr.tidyverse.org/reference/distinct.html>: 该资源是 dplyr 包中函数 `distinct()` 的官方文档页面，介绍了如何使用 `distinct()` 函数从数据框中提取唯一（不重复）的行。文档详细说明了 `distinct()` 的用法、参数及示例，帮助用户理解并应用该函数来去除重复数据，尤其适用于处理包含重复信息的数据列。
2. <https://oreil.ly/0giAa>: (<https://www.tidymodels.org>) 该资源是 tidymodels 生态系统的官方网站，提供用于构建和管理机器学习模型的 R 语言包集合，支持模型训练、评估和调优等功能。

3. <https://oreil.ly/cdBWZ>: (<https://developers.google.com/maps/documentation/geocoding>) 该 URL (<https://developers.google.com/maps/documentation/geocoding>) 指向的是 Google Maps 的地理编码 (Geocoding) API 官方文档页面。该资源详细介绍了如何使用 Google 的地理编码服务, 通过地址信息获取对应的地理坐标 (经纬度), 以及如何将地理坐标转换回地址。文档中包括 API 的使用方法、请求参数、响应格式、示例代码和最佳实践, 帮助开发者在应用中集成地址与地理位置的转换功能。
4. <https://oreil.ly/ytJvP>: (<https://www.tmw.r.org/base-r.html#combining-base-r-models-and-the-tidyverse>) 该资源是一个关于“结合基础 R 模型与 tidyverse 工具”的教程或文档页面, 内容可能讲解如何在数据建模过程中, 将传统的基础 R 方法与 tidyverse 包 (如 tidyr) 结合使用, 以实现更高效的数据整理和分析工作流。页面地址为: <https://www.tmw.r.org/base-r.html#combining-base-r-models-and-the-tidyverse>
5. <https://rdr.io/pkg/jsonlite/man/fromJSON.html>: 该资源是 R 语言包 jsonlite 中函数 fromJSON() 的官方文档页面, 介绍了该函数的用法及功能。fromJSON() 函数用于将 JSON 数据解析为 R 对象, 且默认执行自动简化 (simplifyVector = TRUE), 能方便地将简单的 JSON 结构转换为向量或数据框, 但在处理复杂嵌套结构时可能不够灵活。
6. [https://rdr.io/pkg/jsonlite/man/read\\_json.html](https://rdr.io/pkg/jsonlite/man/read_json.html): 该 URL 指向的是 R 语言中 jsonlite 包中函数 read\_json() 的官方文档页面。该函数用于从磁盘读取 JSON 格式文件, 并将其转换为 R 中的数据结构。文档详细介绍了 read\_json() 的用法, 适合需要处理 JSON 数据并导入 R 进行分析的用户参考。
7. <https://rdr.io/r/base/AsIs.html>: 该资源是 R 语言基础包 (base package) 中函数或类 AsIs 的文档页面。AsIs 通常用于标记对象, 使其在某些操作中保持“原样”不被自动转换或修改。文中提到的 I() 函数即是 AsIs 类的构造函数, 用于将对象包裹起来以强制 data.frame() 按行处理列表。该页面详细介绍了 AsIs 的用法及相关示例。
8. <https://rdr.io/r/base/Paren.html>: 该 URL (<https://rdr.io/r/base/Paren.html>) 指向的是 R 语言基础包 (base package) 中与括号运算符 () 相关的文档

页面。根据上下文，该链接被用来说明 JSON 中对象用花括号 {} 表示，而 R 语言中 ()（括号）运算符在语法和表达式中的作用。该资源主要介绍了 R 语言中括号的用法及其相关功能，帮助理解类似 JSON 对象的命名列表结构在 R 中的表示和操作。

9. <https://rdr.io/r/base/c.html>: 该资源是 R 语言基础包中函数 `c()` 的文档页面，介绍了该函数的用法和功能。函数 `c()` 用于将多个元素组合成一个扁平的向量，是 R 语言中最常用的基本函数之一。
10. <https://rdr.io/r/base/data.frame.html>: 该 URL (<https://rdr.io/r/base/data.frame.html>) 指向的是 R 语言中基础包 (base package) 里 `data.frame()` 函数的在线文档页面。该页面详细介绍了 `data.frame()` 函数的用法、参数说明及其行为特征，例如如何创建数据框、如何处理列表作为列或行的情况等。
11. <https://rdr.io/r/base/list.html>: 该资源是关于 R 语言中基础函数 `list()` 的文档说明页面，介绍如何使用 `list()` 函数创建可以包含不同类型元素的列表 (list)。
12. <https://rdr.io/r/base/names.html>: 该资源是 R 语言基础包中函数 `names()` 的官方文档页面，介绍了该函数的用法和功能，主要用于获取或设置对象（如列表、数据框等）的名称（列名或元素名）。
13. <https://rdr.io/r/utls/View.html>: 该资源是 R 语言中 `utls` 包下的 `View()` 函数的官方文档页面。`View()` 函数用于在交互式界面中以层级化和可展开的方式查看大型且复杂的列表或数据结构，方便用户浏览和深入检查数据内容，尤其是在 `str()` 函数无法有效展示时。该页面详细介绍了 `View()` 的用法及其交互特性，适合在 RStudio 等环境中使用。
14. <https://rdr.io/r/utls/str.html>: 该资源是 R 语言中 `utls` 包提供的函数 `str()` 的官方文档页面。`str()` 函数用于紧凑地显示 R 对象的内部结构，通过分行展示对象的各个子元素，包括名称（若有）、类型缩写及部分内容，方便用户快速浏览复杂数据结构的层级和组成，特别适合查看列表等复杂对象的结构信息。
15. [https://readr.tidyverse.org/reference/parse\\_atomic.html](https://readr.tidyverse.org/reference/parse_atomic.html): 该资源是 tidyverse 包中 `readr` 子包的函数文档，具体介绍了函数 `parse_double()`。该函

数用于将以字符串形式存储的数字（特别是在 JSON 数据中常见的字符串格式数字）解析并转换为双精度浮点数，从而获得正确的数据类型。简言之，文档说明了如何将字符串格式的数值数据转换成数值型变量。

16. [https://readr.tidyverse.org/reference/parse\\_datetime.html](https://readr.tidyverse.org/reference/parse_datetime.html): 该资源是 tidyverse 中 readr 包的函数文档，具体介绍了函数 `parse_datetime()`。该函数用于将以字符串形式存储的日期时间数据（例如 JSON 中常见的日期时间字符串）解析并转换为正确的日期时间数据结构，方便在 R 语言中进行后续的数据处理和分析。
17. <https://tibble.tidyverse.org/reference/tibble.html>: 该资源是关于 R 语言中 tibble 包的官方文档页面，详细介绍了函数 `tibble()` 的用法。页面说明了 `tibble()` 如何将列表(list)当作向量处理，以及其打印方法如何针对列表进行了优化，方便用户在使用 tibble 时更好地操作和展示列表列。
18. <https://tidyr.tidyverse.org/reference/hoist.html>: 该资源是 tidyr 包中函数 `hoist()` 的官方文档页面，介绍了如何直接提取嵌套数据结构中感兴趣的组件。
19. <https://tidyr.tidyverse.org/reference/unnest.html>: 该资源是 tidyr 包中 `unnest()` 函数的官方文档页面。`unnest()` 函数用于展开嵌套的数据结构，既可以展开行也可以展开列，特别适用于包含二维结构（如数据框）的列表列。该函数在处理 tidymodels 生态系统中的复杂数据时非常有用。
20. [https://tidyr.tidyverse.org/reference/unnest\\_auto.html](https://tidyr.tidyverse.org/reference/unnest_auto.html): 该资源是 R 语言数据整理包 tidyr 中的函数文档，介绍了函数 `unnest_auto()`。该函数能够根据列表列的数据结构，自动选择使用 `unnest_longer()` 还是 `unnest_wider()` 来展开数据。它适合快速探索数据，但不建议长期使用，因为自动选择隐藏了数据结构细节，可能导致代码难以理解。
21. [https://tidyr.tidyverse.org/reference/unnest\\_longer.html](https://tidyr.tidyverse.org/reference/unnest_longer.html): 该 URL ([https://tidyr.tidyverse.org/reference/unnest\\_longer.html](https://tidyr.tidyverse.org/reference/unnest_longer.html)) 指向的是 R 语言中 tidyr 包的函数 `unnest_longer()` 的官方文档页面。该函数用于处理数据框 (tibble) 中包含的列表列，将每个列表元素拆分成多行，从而实现“长格式”的数据展开。其主要特点是保持原有的列数不变，而通过增加行数来“展开”列表数据，适用于每行包含未命名列表元素的情况。文档中还介绍了该函数如何处理

不同类型的列表元素、如何与 `unnest_wider()` 配合使用以处理复杂的嵌套数据结构，以及在实际数据整理（如 JSON 数据转换为规整表格）的应用案例。

22. [https://tidyr.tidyverse.org/reference/unnest\\_wider.html](https://tidyr.tidyverse.org/reference/unnest_wider.html): 该资源是 `tidyr` 包中函数 `unnest_wider()` 的官方文档页面。`unnest_wider()` 用于将数据框中每行包含的命名列表（list-column）“展开”为多列，适用于当每行具有相同数量且命名一致的元素时，将每个元素拆分到独立的列中。它是处理嵌套或层级数据（如 JSON 格式数据）时，将复杂的列表结构转化为矩形数据框的重要工具之一，常与 `unnest_longer()` 配合使用，用于“宽展开”嵌套数据，便于后续的数据分析和操作。

## 第 24 章

1. [http://xml2.r-lib.org/reference/read\\_xml.html](http://xml2.r-lib.org/reference/read_xml.html): 该资源是 R 语言中 `xml2` 包的官方文档页面，详细介绍了函数 `read_xml()`（以及与之相关的 `read_html()`）的用法。该函数用于读取网页或 XML 内容，将其解析为 R 中的 `xml_document` 对象，方便后续使用 `rvest` 包进行网页数据抓取和处理。
2. <https://dplyr.tidyverse.org/reference/rename.html>: 该资源是 `dplyr` 包中用于重命名数据框列的函数说明页面，详细介绍了如何使用 `rename()` 函数对数据框的列名进行修改，以便更方便地进行后续的数据处理和分析。
3. <https://dplyr.tidyverse.org/reference/select.html>: 该 URL 指向的是 `dplyr` 包中函数 `select()` 的官方文档页面。该函数用于从数据框中选择指定的列，支持一次性对列进行重命名和筛选，方便简化数据处理流程。
4. <https://en.wikipedia.org/wiki/Cat>: 该 URL “<https://en.wikipedia.org/wiki/Cat>” 指向维基百科上关于“猫”（Cat）的条目页面，介绍了猫的相关信息，包括其生物学特征、习性、品种、历史以及与人类的关系等内容。



5. <https://en.wikipedia.org/wiki/Dog>: 该 URL (<https://en.wikipedia.org/wiki/Dog>) 指向维基百科中关于“狗” (Dog) 的条目页面, 内容介绍了狗的生物学特征、品种、起源、行为习性及其与人类的关系等相关信息。
6. <https://oreil.ly/0ewJe>: ([https://en.wikipedia.org/wiki/Database\\_right](https://en.wikipedia.org/wiki/Database_right)) 该 URL ([https://en.wikipedia.org/wiki/Database\\_right](https://en.wikipedia.org/wiki/Database_right)) 指向的资源是关于“数据库权利” (Database Right) 的维基百科页面。该页面介绍了数据库权利这一法律概念, 特别是在欧洲法律体系中对数据库内容的特殊保护, 即使数据库中的原始数据本身不受版权保护, 数据库权利仍赋予数据库制作者对其数据库整体结构和内容的专有权利, 从而防止未经授权的复制和提取。这种“特殊权利” (sui generis right) 用于保护数据库制作者的投资和劳动成果, 区别于传统的版权保护。
7. <https://oreil.ly/McJtu>: (<https://flukeout.github.io>) 该资源是一个在线的 CSS 选择器学习和练习工具, 帮助用户通过互动方式掌握 CSS 选择器的用法和原理。
8. <https://oreil.ly/OqUgO>: (<https://www.law.cornell.edu/uscode/text/17/102>) 该 URL “<https://www.law.cornell.edu/uscode/text/17/102>” 指向的是美国法典第 17 编第 102 条, 内容涉及版权法中关于受保护作品范围的具体规定, 明确说明了版权保护适用于“固定在任何有形载体上的原创著作作品”, 并列举了适用的作品类别如文学作品、音乐作品和电影作品等, 同时指出事实和数据不受版权保护。
9. <https://oreil.ly/YoxV7>: (<http://rvest.tidyverse.org>) 该资源是 rvest 包的官方网站, 提供有关 rvest 的详细信息和文档。rvest 是一个用于 R 语言的网页数据抓取工具, 支持模拟浏览器行为, 处理 JavaScript 动态内容, 方便用户从网页中提取数据。
10. <https://oreil.ly/eD6eC>: (<https://kittygiraudel.github.io/selectors-explained>) 该资源 “<https://kittygiraudel.github.io/selectors-explained>” 是一个工具或网站, 能够将复杂的 CSS 选择器翻译成通俗易懂的英文描述, 帮助用户理解 CSS 选择器的含义和结构。

11. <https://oreil.ly/lQNBa>: (<https://xml2.r-lib.org>) 该资源是 R 语言中用于处理 XML 文档的基础包“xml2”的官方网站或文档站点, 提供该包的功能介绍、使用说明及相关技术资料。
12. <https://oreil.ly/lUNa6>: (<https://rvest.tidyverse.org>) 该资源是 rvest 包的官方网站, rvest 是 R 语言中用于网页数据抓取 (web scraping) 的工具包, 提供简便的方法从网页中提取和解析数据, 帮助用户在无 API 支持时获取网页内容。网址 <https://rvest.tidyverse.org> 提供了该包的文档、使用指南和相关示例。
13. <https://oreil.ly/mDAin>: ([https://en.wikipedia.org/wiki/HiQ\\_Labs\\_v.\\_LinkedIn](https://en.wikipedia.org/wiki/HiQ_Labs_v._LinkedIn)) 该资源介绍了美国法院在“HiQ Labs 诉 LinkedIn”案件中关于网站服务条款的法律判决, 特别涉及用户是否因仅浏览公开数据而受限于网站服务条款的约束。案件强调, 单纯将服务条款置于网站底部通常不足以使用户受其约束, 用户需通过明确行为 (如注册账号或勾选同意框) 才能被视为同意条款。该案例在数据访问权限和服务条款法律效力方面具有重要参考意义。
14. <https://oreil.ly/mpfMF>: ([https://developer.mozilla.org/en-US/docs/Web/CSS/CSS\\_Selectors](https://developer.mozilla.org/en-US/docs/Web/CSS/CSS_Selectors)) 该资源是 MDN (Mozilla Developer Network) 提供的关于 CSS 选择器的详细文档, 介绍了各种 CSS 选择器的用法和语法, 帮助开发者理解和使用 CSS 选择器来精准地选中网页元素。
15. <https://oreil.ly/nzJwO>: (<https://gdpr-info.eu>) 该资源 (<https://gdpr-info.eu>) 是一个关于欧盟《通用数据保护条例》(GDPR) 的官方网站或信息平台, 提供该法规的全文内容及相关解读, 旨在帮助用户了解欧盟对个人数据收集、存储和处理的严格法律要求和合规规范。
16. <https://oreil.ly/oFh0->: ([https://en.wikipedia.org/wiki/Fair\\_use](https://en.wikipedia.org/wiki/Fair_use)) 该资源是维基百科上关于“合理使用” (Fair Use) 原则的条目, 介绍了合理使用的定义、适用条件及其在版权法中的作用, 解释了在某些情况下使用受版权保护的内容 (如文本或图像) 可能无需授权而被法律允许, 尤其是在研究或非商业用途时的相关法律考量。
17. <https://oreil.ly/qIgHp>: (<https://developer.mozilla.org/en-US/docs/Web/HTML>) 该资源是 MDN Web Docs (Mozilla 开发者网络) 中关于 HTML 的文

档页面，提供详尽的 HTML 标签和相关技术的解释与使用指南，是学习和查询网页开发中 HTML 规范和用法的重要参考资料。

18. <https://oreil.ly/qNv6l>: (<https://www.youtube.com/watch?v=PetWV5g1Xsc>) 该资源是一个关于如何安装和使用 SelectorGadget 的视频教程，演示了这个 JavaScript 书签工具如何通过用户提供的正负示例自动生成 CSS 选择器，从而简化网页元素的定位和提取过程。
19. <https://oreil.ly/qui0z>: (<https://rvest.tidyverse.org/articles/selectorgadget.html>) 该资源: <https://rvest.tidyverse.org/articles/selectorgadget.html> 是一篇关于 SelectorGadget 工具的介绍文档。SelectorGadget 是一个基于 JavaScript 的书签工具，可以根据用户提供的正面和负面示例自动生成 CSS 选择器，帮助简化网页数据抓取中的元素定位过程。该文档详细说明了 SelectorGadget 的安装和使用方法，适合希望高效提取网页数据的用户参考学习。
20. <https://oreil.ly/rIujg>: (<https://dmi3kno.github.io/polite>) 该资源是由 Dmytro Perepolkin 开发的一个名为“polite”的软件包，用于在网络爬取 (scraping) 过程中自动控制请求频率，避免对服务器造成过大负载。该包会在每次请求之间自动暂停，并缓存已请求的页面，防止重复请求，从而实现更为礼貌和高效的数据抓取。
21. <https://oreil.ly/rzd7z>: (<https://www.wired.com/2016/05/okcupid-study-reveals-perils-big-data-science>) 该 URL 指向的是《Wired》杂志于 2016 年 5 月发表的一篇文章，内容涉及对 OkCupid 数据分析，揭示了大数据科学在实际应用中可能存在的风险和问题。
22. <https://oreil.ly/xaHTf>: (<https://rstudio.github.io/chromote/index.html>) 该资源是 R 语言中名为 chromote 的软件包的在线文档或教程页面，介绍了该包如何通过后台运行 Chrome 浏览器，实现对动态网页（包括 JavaScript 渲染内容）的自动化控制和交互，如模拟用户输入和点击操作，从而辅助网页数据抓取和爬取工作。
23. <https://rvest.tidyverse.org/articles/starwars.html>: 该资源是 rvest 包中的一个示例介绍页面，展示了如何使用 rvest 进行网页数据抓取。页面内容以《星

球大战》(Star Wars) 电影为例, 提供了一个结构简单、HTML 标签层次清晰的网页, 方便用户通过浏览器的“检查元素”功能学习和理解网页元素的层级结构及抓取方法, 适合 rvest 初学者作为入门练习。

24. [https://rvest.tidyverse.org/reference/html\\_attr.html](https://rvest.tidyverse.org/reference/html_attr.html): 该资源是 rvest 包中函数 `html_attr()` 的官方文档页面, 介绍了该函数用于从 HTML 元素的属性中提取数据的用法及相关说明。
25. [https://rvest.tidyverse.org/reference/html\\_element.html](https://rvest.tidyverse.org/reference/html_element.html): 该 URL ([https://rvest.tidyverse.org/reference/html\\_element.html](https://rvest.tidyverse.org/reference/html_element.html)) 指向的是 R 语言中 rvest 包的函数文档, 主要介绍了函数 `html_element()` 的用法和功能。根据上下文信息, 该函数用于从 HTML 文档中根据 CSS 选择器查找元素, 返回的结果数量与输入数量相同, 且在选择器未匹配任何元素时返回缺失值 (而不是零长度向量)。它常与 `html_elements()` 函数配合使用, 后者用于查找所有匹配的元素, 通常用于获取“观察值”, 而 `html_element()` 用于获取“变量”对应的元素, 方便进行网页数据的结构化提取和整理。简单来说, 该资源是关于 rvest 包中 `html_element()` 函数的官方参考手册, 详细说明了该函数的功能、行为差异及应用场景。
26. [https://rvest.tidyverse.org/reference/html\\_table.html](https://rvest.tidyverse.org/reference/html_table.html): 该资源是 rvest 包中函数 `html_table()` 的文档页面, 介绍了如何从网页中提取 HTML 表格数据。该函数能够解析网页中的表格元素, 并将其转换为 tibble 格式的数据框, 方便后续的数据处理与分析。文档中还可能包含函数的用法示例和参数说明。
27. [https://rvest.tidyverse.org/reference/html\\_text.html](https://rvest.tidyverse.org/reference/html_text.html): 该资源是 rvest 包中函数 `html_text()` 和 `html_text2()` 的官方文档页面, 介绍了如何从 HTML 元素中提取纯文本内容。页面特别推荐使用 `html_text2()` 函数, 因为它在将嵌套 HTML 转换为文本方面表现更优。
28. [https://tidyr.tidyverse.org/reference/separate\\_wider\\_delim.html](https://tidyr.tidyverse.org/reference/separate_wider_delim.html): 该资源是 R 语言包 tidyr 中的函数文档, 介绍了函数 `separate_wider_regex()` 的用法。该函数用于根据正则表达式将一系列数据拆分成多列, 更方便地提取和整理表格中的复杂信息, 如将包含多个字段的字符串拆解成独立的变量 (如标题、年份、排名等), 以便后续的数据处理和分析。

---

## 第 25 章

1. <https://dplyr.tidyverse.org/reference/across.html>: 该资源是 dplyr 包中函数 `across()` 的官方文档页面, 介绍了如何使用 `across()` 方便地对数据框的多列同时进行操作或变换, 从而减少代码重复, 提高代码简洁性和可读性。文档详细讲解了该函数的用法及应用场景, 适合在数据处理和变换中批量修改多列时参考。
2. <https://dplyr.tidyverse.org/reference/arrange.html>: 该资源是 dplyr 包中函数 `arrange()` 的官方文档页面, 介绍了如何使用 `arrange()` 函数对数据框中的变量进行排序操作。文档详细说明了函数的用法、参数以及示例, 帮助用户理解并应用该函数对数据进行有序排列。
3. [https://dplyr.tidyverse.org/reference/case\\_when.html](https://dplyr.tidyverse.org/reference/case_when.html): 该资源是 dplyr 包中函数 `case_when()` 的官方文档页面, 介绍了 `case_when()` 函数的用法和示例。`case_when()` 用于根据多个条件对向量元素进行分组分类, 是一种简洁的条件判断工具, 常用于数据处理和清洗。
4. <https://dplyr.tidyverse.org/reference/count.html>: 该资源是 dplyr 包中函数 `count()` 的官方文档页面, 介绍了 `count()` 函数的用法和参数。`count()` 用于对数据框中的某个变量进行计数汇总, 支持数据遮罩 (data masking) 和 tidy evaluation, 方便用户快速计算各类分组频数或比例。文档中详细说明了该函数的参数及其行为, 以及如何与其他 dplyr 函数配合使用。
5. <https://dplyr.tidyverse.org/reference/distinct.html>: 该资源是 dplyr 包中函数 `distinct()` 的官方文档页面, 介绍了 `distinct()` 函数的用法和参数。`distinct()` 用于从数据框中提取唯一 (不重复) 的行, 常用于去重操作。文档中还可能说明了该函数支持的 tidy evaluation (整洁求值) 机制, 包括如何通过变量选择和数据掩码进行操作。
6. <https://dplyr.tidyverse.org/reference/filter.html>: 该资源是 dplyr 包中函数 `filter()` 的官方文档页面, 介绍了 `filter()` 函数的用法。`filter()` 用于根据指定条件筛选数据框中的行, 返回满足条件的子集, 常用于数据处理和清洗环节。文档中详细说明了 `filter()` 的参数、功能及示例, 帮助用户理解如何利用该函数对数据进行行级过滤操作。

7. [https://dplyr.tidyverse.org/reference/group\\_by.html](https://dplyr.tidyverse.org/reference/group_by.html): 该资源是 dplyr 包中函数 `group_by()` 的官方文档页面。它详细介绍了 `group_by()` 函数的用法, 特别说明该函数采用“数据掩码”(data masking) 机制进行数据操作, 而非“整洁选择”(tidy selection)。文档中还可能提供函数的参数说明、示例代码及使用场景, 帮助用户理解如何基于一个或多个变量对数据进行分组操作。
8. <https://dplyr.tidyverse.org/reference/mutate.html>: 该 URL (<https://dplyr.tidyverse.org/reference/mutate.html>) 指向的是 R 语言数据处理包 dplyr 中函数 `mutate()` 的官方文档页面。该资源详细介绍了 `mutate()` 函数的用法, 功能是在数据框中添加新变量或修改已有变量, 返回的数据框行数与输入相同。文档中包含函数的参数说明、使用示例以及相关的函数介绍, 帮助用户理解如何通过 `mutate()` 进行高效的数据变换和列操作。
9. <https://dplyr.tidyverse.org/reference/pick.html>: 该资源是 dplyr 包中用于列选择的函数文档页面。具体来说, `pick()` 函数允许用户在数据掩码 (data masking) 环境中使用 tidy selection 语法, 方便地选择数据框中的列。它常用于解决在某些数据操作函数 (如 `group_by()`) 中不能直接使用 tidy selection 的问题, 同时也可用于生成基于多变量的二维计数表等场景。
10. <https://dplyr.tidyverse.org/reference/relocate.html>: 该资源是 dplyr 包中函数 `relocate()` 的官方文档页面, 介绍了用于重新排列数据框中变量位置的函数, 属于类似于 `select()` 和 `rename()` 的变量选择操作。
11. <https://dplyr.tidyverse.org/reference/rename.html>: 该资源是 dplyr 包中用于重命名数据框列的函数 `rename()` 的官方文档页面, 详细介绍了该函数的用法、参数和示例, 帮助用户理解如何通过 `rename()` 函数对数据集中的变量名进行修改。
12. <https://dplyr.tidyverse.org/reference/select.html>: 该资源是 dplyr 包中关于函数 `select()` 的官方文档页面, 详细介绍了 `select()` 函数的用法和功能, 该函数用于从数据框中选择变量。
13. <https://dplyr.tidyverse.org/reference/slice.html>: 该资源是 dplyr 包中与切片操作相关的函数文档页面, 主要介绍了函数如 `slice_min()` 和 `slice_sample()` 的用法。这些函数用于从数据框中根据某些条件选取最小值的行

或随机抽样的行，文档中详细说明了它们的参数，特别是使用 tidy evaluation（整洁求值）的参数，及其采用的数据掩码（data masking）或整洁选择（tidy selection）机制。

14. <https://dplyr.tidyverse.org/reference/summarise.html>: 该 URL 指向的资源是 dplyr 包中关于函数 summarise()（或 summarize()）的官方文档页面。该文档详细介绍了 summarise() 函数的用法及其在数据汇总中的应用。summarise() 是 dplyr 中用于生成汇总统计的一类函数，通常用于对分组数据进行计算，返回单一汇总值，如均值、计数等。文档还涵盖了与 tidy evaluation 配合使用 summarise() 的技巧、常见参数设置（如 .groups 参数），以及它在数据分析工作流中的重要角色。
15. [https://forcats.tidyverse.org/reference/fct\\_inorder.html](https://forcats.tidyverse.org/reference/fct_inorder.html): 该资源是 tidyverse 中 forcats 包的函数文档页面，具体介绍了函数 fct\_inorder()，该函数用于按照数据中因子水平出现的顺序对因子进行排序，方便在数据可视化（如使用 ggplot2 绘制条形图）时对因子水平进行自定义排序，从而实现例如按频率顺序排列条形的效果。
16. <https://ggplot2.tidyverse.org/reference/aes.html>: 该资源是 ggplot2 包中函数 aes() 的官方文档页面，详细介绍了 aes() 函数的用法和功能。aes() 是一个用于定义图形美学映射（aesthetic mappings）的数据掩码函数，广泛应用于构建各种图形（如直方图）时指定变量与视觉属性（如颜色、大小、坐标等）之间的对应关系。该文档适合希望深入理解和灵活使用 ggplot2 进行数据可视化的用户。
17. <https://oreil.ly/8xygI>: (<https://dplyr.tidyverse.org/articles/programming.html>) 该资源是 dplyr 官方网站上的一篇关于“使用 tidy evaluation 进行编程”的教程文章，详细介绍了如何在 dplyr 中应用编程技巧和数据掩码（data masking）理论，帮助用户更高效地编写和理解基于 tidyverse 的数据操作代码。
18. <https://oreil.ly/MV4kQ>: (<https://gist.github.com/GShotwell/b19ef520b6d56f61a830fabbb3454965b>) 该资源是一个由 GShotwell 在 GitHub Gist 上发布的代码片段，内容可能与 ggplot2（R 语言中的绘图包）相关，展示了如何通过

编程方式增强数据可视化功能，比如利用坐标轴标签显示数据中的最高值，从而实现更具创意和信息丰富的图表展示。

19. <https://oreil.ly/QGH9n>: (<https://tidyr.tidyverse.org/articles/programming.html>) 该资源是 tidyr 包中关于“tidy evaluation”编程方法的官方文章，详细介绍了如何在 tidyverse 生态中使用编程技巧，特别是数据掩码 (data masking) 机制及其应用，帮助用户更高效地利用 tidyr 进行编程。
20. <https://oreil.ly/Vvt6k>: (<https://ggplot2-book.org/programming.html>) 该资源是《ggplot2》一书中的“Programming with ggplot2”章节，主要介绍如何通过编程技巧减少 ggplot2 代码中的重复，提高代码的效率和可维护性。
21. <https://oreil.ly/Ymcmk>: (<https://twitter.com/hadleywickham/status/1571603361350164486>) 该资源是 Hadley Wickham 在 Twitter 上的一条推文，内容可能与 R 语言中的通用函数 (general functions) 或绘图函数 (plotting functions) 相关，提供了这些函数设计理念或示例的原始灵感和动机说明。
22. <https://oreil.ly/eecUd>: (<https://rlang.r-lib.org/reference/topic-data-mask.html>) 该资源是 R 语言官方文档中关于“data mask”（数据掩码）主题的说明页面，介绍了数据掩码的概念及其在 tidy evaluation（整洁求值）编程中的应用，帮助用户理解和使用数据掩码机制以便更有效地进行数据操作和编程。
23. <https://oreil.ly/mXy2q>: (<https://twitter.com/hadleywickham/status/1574373127349575680>) 该 URL “<https://twitter.com/hadleywickham/status/1574373127349575680>” 指向的是 Hadley Wickham 在 Twitter 上的一条推文，这条推文是本 HTML 内容中提及的“original motivating tweets”（最初的灵感推文）之一，内容与 R 语言中通用函数 (general functions) 的设计和使用的模式相关，提供了该领域的原始思想和示例，帮助读者理解和推广相关函数的模式。
24. <https://oreil.ly/rLKSn>: (<https://style.tidyverse.org/functions.html>) 该资源是 tidyverse 的函数风格指南，提供了关于如何编写和命名函数的最佳实践和建议，旨在帮助用户编写符合 tidyverse 规范、风格统一且易于维护的 R 语言代码。



- 
25. <https://rdr.io/r/base/Extremes.html>: 该资源是 R 语言基础包 (base) 中关于极值相关函数 (如 `min()` 和 `max()`) 的文档页面, 详细介绍了这些函数的用法及其在计算数据向量的最小值和最大值时的应用。
  26. <https://rdr.io/r/base/Paren.html>: 该资源是 R 语言基础包中的“括号 (Paren)”帮助文档, 介绍了 R 语言中括号符号 (如大括号 `{}`) 的用法和语法规则, 主要用于代码块的定义和表达式的分组。在上下文中, 该文档被用来说明字符串插值时花括号 `{}` 的作用, 以及函数体必须用大括号 `{}` 包裹并进行缩进的规范。
  27. <https://rdr.io/r/base/mean.html>: 该资源是 R 语言基础包中关于函数 `mean()` 的文档页面, 介绍了 `mean()` 函数的用法和功能。`mean()` 函数用于计算数值向量的算术平均值, 是 R 语言中计算均值的标准函数之一。
  28. <https://rdr.io/r/base/range.html>: 该资源是 R 语言基础包 (base package) 中关于函数 `range()` 的官方文档页面, 详细介绍了 `range()` 函数的用法。`range()` 函数用于同时计算向量的最小值和最大值, 常用于替代分别调用 `min()` 和 `max()` 函数以提高效率。文档中还说明了如何使用该函数的参数来忽略无穷大 (infinite) 等特殊值。
  29. <https://rdr.io/r/stats/Normal.html>: 该 URL “<https://rdr.io/r/stats/Normal.html>” 指向的是 R 语言中 stats 包里关于正态分布 (Normal distribution) 相关函数的文档页面。该页面详细介绍了正态分布的概率密度函数 (如 `dnorm`)、分布函数、分位数函数以及随机数生成函数 (如 `rnorm`) 的用法、参数说明和示例, 方便用户理解和调用这些统计函数进行正态分布相关的计算和模拟。
  30. <https://rdr.io/r/stats/coef.html>: 该资源是 R 语言中 stats 包的函数文档页面, 具体介绍了函数 `coef()`, 该函数用于提取模型的系数 (coefficients)。页面详细说明了函数的用法、参数及返回值, 帮助用户理解如何通过 `coef()` 函数访问统计模型对象中的系数信息。
  31. <https://rlang.r-lib.org/reference/englu.html>: 该资源是 R 语言包 rlang 中函数 `englu()` 的官方文档页面。该页面介绍了 `englu()` 函数的用法, 说明它类似于 stringr 包中的 `str_glu()` 函数, 可以将用大括号 `{}` 包裹的值插入字符串

中，且支持使用双重大括号 `{{ }}` 自动插入变量名，从而方便地解决字符串标签化的问题。

32. [https://stringr.tidyverse.org/reference/str\\_glue.html](https://stringr.tidyverse.org/reference/str_glue.html): 该资源是 `stringr` 包中函数 `str_glue()` 的文档页面，介绍了 `str_glue()` 函数的用法。该函数用于字符串插值，可以将用花括号 `{}` 包裹的表达式的值插入到字符串中，方便动态生成字符串内容。
33. [https://tidyr.tidyverse.org/reference/pivot\\_wider.html](https://tidyr.tidyverse.org/reference/pivot_wider.html): 该资源是 R 语言中 `tidyr` 包的函数文档页面，详细介绍了函数 `pivot_wider()` 的用法。  
`pivot_wider()` 用于将数据从“长格式”转换为“宽格式”，常用于将计数或其他汇总结果重新排列成二维表格。文档中还说明该函数支持 `tidy selection` 语法，方便用户灵活选择用于生成列名的数据变量。

## 第 26 章

1. <https://allisonhorst.github.io/palmerpenguins/reference/penguins.html>: 该资源是“`palmerpenguins`”包中“`penguins`”数据集的官方文档页面，介绍了该数据集的结构、变量说明及使用方法。
2. <https://dbi.r-dbi.org/reference/dbAppendTable.html>: 该资源是 R 语言中 DBI 包的函数文档，介绍了函数 `dbAppendTable()` 的用法。该函数用于将数据追加写入数据库中的指定表。页面详细说明了如何将数据框的数据追加插入已有数据库表中，适合在数据导入或更新时使用。
3. <https://dbi.r-dbi.org/reference/dbCreateTable.html>: 该资源是 R 语言中 DBI 包的函数文档页面，详细介绍了函数 `DBI::dbCreateTable()` 的用法，该函数用于在数据库中创建一个新表。
4. <https://dplyr.tidyverse.org/reference/across.html>: 该 URL 指向的资源是 R 语言中 `dplyr` 包的函数 `across()` 的官方文档。`across()` 函数用于在数据框中对多个列进行批量操作，简化对多列重复应用函数的代码。其主要特点和用途包括：

- 多列选择 (`.cols` 参数)：通过 `tidyselect` 语法选择需要操作的列，比如使用 `starts_with()`、`ends_with()`、`everything()`、`where()` 等函数来灵活指定列范围。
- 函数应用 (`.fns` 参数)：指定对选中列应用的函数，可以是单个函数，也可以是函数列表，且函数以函数名传入（不带括号）。
- 输出列命名 (`.names` 参数)：控制输出列的命名规则，方便与 `mutate()` 等函数配合使用，避免列名冲突或覆盖。
- 配合 `mutate()` 与 `summarize()` 使用：常用于批量计算、转换数据列。
- 变体函数 `if_any()` 和 `if_all()`：用于 `filter()` 中，方便基于多个列的条件筛选，分别表示任一列满足或全部列满足条件。
- 编程友好：支持函数式编程风格，便于编写灵活、简洁的数据操作代码。

总之，该文档详细介绍了 `across()` 函数的语法、参数及使用场景，是学习和掌握 `dplyr` 中高效批量列操作的关键参考。

5. [https://dplyr.tidyverse.org/reference/bind\\_rows.html](https://dplyr.tidyverse.org/reference/bind_rows.html)：该资源是 `dplyr` 包中函数 `bind_rows()` 的官方文档页面，介绍如何使用 `bind_rows()` 函数将多个数据框按行合并。
6. <https://dplyr.tidyverse.org/reference/coalesce.html>：该资源是 `dplyr` 包中函数 `coalesce()` 的官方文档页面，介绍了该函数的用法和功能。`coalesce()` 函数用于对多个向量进行逐元素合并，返回第一个非缺失值（非 NA），常用于替换缺失值，如将 NA 替换为指定的默认值（例如 0）。该文档详细说明了 `coalesce()` 在数据处理中的应用，特别是在与 `mutate()` 和 `across()` 函数结合使用时，如何有效处理缺失值。
7. <https://dplyr.tidyverse.org/reference/filter.html>：该 URL (<https://dplyr.tidyverse.org/reference/filter.html>) 指向的是 R 语言中 `dplyr` 包的函数 `filter()` 的官方文档页面。`filter()` 函数用于根据指定的条件对数据框（data frame）进行行筛选，只保留满足条件的行。文档中介绍了 `filter()` 的用法、参数说明以及与其他函数（如 `across()`、`if_any()` 和 `if_all()`）的结合使用方法，帮助用户高效地进行数据过滤操作。
8. [https://dplyr.tidyverse.org/reference/group\\_by.html](https://dplyr.tidyverse.org/reference/group_by.html)：该资源是 R 语言中 `dplyr` 包的函数说明页面，详细介绍了函数 `group_by()` 的用法。`group_by()` 用于将

数据按一个或多个变量分组，常与 `summarize()` 函数配合使用，从而对每个分组计算汇总统计量。

9. [https://dplyr.tidyverse.org/reference/group\\_nest.html](https://dplyr.tidyverse.org/reference/group_nest.html): 该资源是 dplyr 包中用于数据分组和嵌套的函数 `group_nest()` 的官方文档页面。该函数可以将数据框按照指定的分组变量进行分组，并将每个分组的数据嵌套成一个列表列，方便对每个分组的数据进行独立操作或保存，如将不同组的数据分别写入多个 CSV 文件。
10. <https://dplyr.tidyverse.org/reference/mutate.html>: 该 URL (<https://dplyr.tidyverse.org/reference/mutate.html>) 指向的是 R 语言数据处理包 dplyr 中函数 `mutate()` 的官方文档资源。`mutate()` 函数用于对数据框 (data frame) 进行列的新增或修改操作，常与 `across()` 等函数配合使用，实现对指定列的批量变换和生成新列。文档详细介绍了 `mutate()` 的用法、参数及在数据变换中的应用示例。
11. <https://dplyr.tidyverse.org/reference/relocate.html>: 该资源是 dplyr 包中函数 `relocate()` 的官方文档页面，介绍了如何在数据框中重新排列列的顺序。
12. <https://dplyr.tidyverse.org/reference/select.html>: 该资源是 dplyr 包中函数 `select()` 的官方文档页面，介绍了如何使用 `select()` 函数来选择数据框中的列。文档详细说明了 `select()` 函数的用法，包括支持使用 tidyselect 语法 (如 `starts_with()`、`ends_with()` 等辅助函数) 来根据列名灵活地选择列。该页面是 dplyr 包列选择操作的权威参考资料。
13. <https://dplyr.tidyverse.org/reference/summarise.html>: 该资源是 dplyr 包中函数 `summarize()` 的官方文档页面，详细介绍了 `summarize()` 函数的用法及其功能。`summarize()` 用于计算数据子集的汇总统计量，通常与 `group_by()` 一起使用，实现对分组数据的汇总操作。文档还说明 `summarize()` 会自动保留分组列，并且与 `across()` 函数配合使用时非常方便，适合对多个变量同时进行汇总计算。
14. <https://ggplot2.tidyverse.org/reference/diamonds.html>: 该资源是 R 语言中 ggplot2 包的内置数据集 “diamonds” 的文档页面，提供了关于该数据集的

详细说明和各变量的描述。该数据集包含了大量钻石的属性信息，常用于数据可视化和数据分析示例。

15. [https://ggplot2.tidyverse.org/reference/facet\\_grid.html](https://ggplot2.tidyverse.org/reference/facet_grid.html): 该资源是 ggplot2 包中用于数据分面绘图的函数文档页面，详细介绍了函数 `facet_grid()` 的用法，功能是根据数据的一个或两个变量将图形分面，分别绘制每个子集的图表，从而实现按变量组合进行多面板展示。
16. [https://ggplot2.tidyverse.org/reference/facet\\_wrap.html](https://ggplot2.tidyverse.org/reference/facet_wrap.html): 该资源是 ggplot2 包中函数 `facet_wrap()` 的官方文档页面，介绍了如何使用 `facet_wrap()` 函数根据数据的子集分别绘制多个子图，实现分面绘图功能。
17. <https://ggplot2.tidyverse.org/reference/ggsave.html>: 该资源是 ggplot2 包中用于保存绘图的函数文档页面，介绍了函数 `ggsave()` 的用法和参数，帮助用户将绘制的图形导出为文件。
18. <https://oreil.ly/6vVc4>: ([https://twitter.com/\\_wurli/status/1571836746899283969](https://twitter.com/_wurli/status/1571836746899283969)) 该 URL ([https://twitter.com/\\_wurli/status/1571836746899283969](https://twitter.com/_wurli/status/1571836746899283969)) 所指向的资源很可能是一条 Twitter 推文，内容与 R 语言中 dplyr 包的 `across()` 函数相关。具体来说，这条推文可能由 Jacob Scott 发布，展示了一个利用 `across()` 函数结合 lubridate 包的示例，用于将数据框中的所有日期列拆分成年、月、日三列的代码片段或用法说明。
19. <https://oreil.ly/VmXg4>: (<https://adv-r.hadley.nz/functionals.html>) 该资源是 Hadley Wickham 所著《Advanced R》一书中关于“函数式编程”(Functionals) 章节的在线版本或相关内容，详细介绍了 R 语言中函数式编程的概念和技巧，特别是如何利用函数式编程方法进行高效迭代和数据操作，适合希望深入掌握 R 语言迭代技术的数据科学从业者学习。
20. <https://oreil.ly/f0HWP>: (<http://purrr.tidyverse.org>) 该资源是 purrr 包的官方网站，purrr 是 tidyverse 生态系统中的一个 R 语言包，提供了一组函数式编程工具，用于简化和增强数据处理中的迭代操作，帮助用户更高效地处理和转换数据。
21. <https://oreil.ly/oJsOo>: (<https://docs.ropensci.org/targets>) 该资源是关于 R 语言包 targets 的官方文档，介绍如何使用该工具自动化数据处理流程，使得当

输入数据文件发生变化时，相关的数据清理和分析代码能够自动重新运行，从而实现动态、高效的数据工作流管理。

22. [https://purrr.tidyverse.org/reference/list\\_c.html](https://purrr.tidyverse.org/reference/list_c.html): 该 URL 指向的是 R 语言中 purrr 包的函数 `list_rbind()` 的官方文档页面。根据上下文，这个函数用于将一个包含多个数据框（data frames）的列表合并成一个单一的数据框。它通常与 `list.files()`（列出目录中的文件）和 `map()`（对列表中的每个元素应用函数）等函数配合使用，实现批量读取多个文件并合并数据的自动化处理。`list_rbind()` 还支持通过 `names_to` 参数将列表元素的名称保存为数据框中的新列，方便数据标识。此外，当数据框结构不完全一致时，`list_rbind()` 可能失败，但它能够自动忽略 NULL 值，提高处理异构数据的灵活性。
23. <https://purrr.tidyverse.org/reference/map.html>: 该 URL “<https://purrr.tidyverse.org/reference/map.html>” 指向的是 R 语言中 purrr 包的函数 `map()` 的官方文档。根据上下文信息，`purrr::map()` 是一个用于对向量（或列表）中的每个元素应用同一函数的工具，常用于简化对多个文件或数据元素的批量操作流程。
24. <https://purrr.tidyverse.org/reference/map2.html>: 该资源是 R 语言中 purrr 包的函数文档，介绍了函数 `map2()`。该函数用于对两个输入向量（或列表）的对应元素同时进行迭代操作，适用于需要同时变化两个参数的情况。与单一参数变化的 `map()` 不同，`map2()` 可以并行遍历两个参数，并对其对应元素应用指定函数。文档还提到配套的 `walk2()` 函数，用于执行副作用操作而不关心返回值。
25. [https://purrr.tidyverse.org/reference/map\\_if.html](https://purrr.tidyverse.org/reference/map_if.html): 该 URL ([https://purrr.tidyverse.org/reference/map\\_if.html](https://purrr.tidyverse.org/reference/map_if.html)) 指向的是 R 语言中 purrr 包的函数 `map_if()` 的官方文档。该函数用于对列表中的元素进行条件筛选和修改，即根据元素的值是否满足特定条件，有选择性地对这些元素应用某个函数，实现对列表元素的有条件变换。文档中通常包含函数的用法说明、参数介绍和示例，帮助用户理解如何基于元素值来选择性地修改列表内容。
26. <https://purrr.tidyverse.org/reference/possibly.html>: 该资源是 purrr 包中函数 `possibly()` 的官方文档页面。`possibly()` 是一个函数操作器（function

operator)，它接受一个函数作为输入，返回一个带有修改后行为的新函数。具体来说，`possibly()`使得原本可能报错的函数在遇到错误时不会中断程序，而是返回用户指定的默认值，从而方便安全地处理可能失败的函数调用。

27. <https://rdr.io/r/base/basename.html>: 该资源是 R 语言基础包 (base) 中函数 `basename()` 的官方文档页面，介绍了 `basename()` 函数的用法及功能。该函数用于从完整的文件路径中提取文件名部分。
28. <https://rdr.io/r/base/list.files.html>: 该资源是 R 语言基础包中函数 `list.files()` 的官方文档页面，介绍了该函数的用途和用法。`list.files()` 用于列出指定目录中的所有文件，常用于批量读取和处理多个文件的场景。文档中详细说明了该函数的参数设置及示例，帮助用户高效地获取目录文件列表。
29. <https://rdr.io/r/stats/median.html>: 该资源是 R 语言中 stats 包中的函数说明页面，专门介绍了函数 `median()` 的用法。页面内容包括 `median()` 函数的基本功能，即计算中位数，以及如何处理数据中的缺失值 (missing values)，例如通过传递参数 `na.rm = TRUE` 来忽略缺失值。此外，还涉及如何结合其他函数 (如 `across()`) 一起使用 `median()`，以实现更复杂的数据处理需求，比如同时计算中位数和统计缺失值数量。这个页面适合需要深入理解和灵活使用 `median()` 函数的 R 语言用户参考。
30. <https://rdr.io/r/utils/View.html>: 该资源是 R 语言中 utils 包里的函数 `View()` 的在线文档页面。该页面介绍了 `View()` 函数的用法，该函数用于在 RStudio 的查看器中以表格形式打开和检查数据结构，便于用户对数据进行直观浏览和交互式检查。
31. <https://rdr.io/r/utils/str.html>: 该资源是 R 语言中 utils 包的函数 `str()` 的文档页面。`str()` 函数用于紧凑地显示 R 对象的内部结构，方便用户快速了解数据结构的概要信息。页面详细介绍了 `str()` 的用法、参数及示例，帮助用户理解和使用该函数进行数据结构的简要展示。
32. [https://readr.tidyverse.org/reference/parse\\_number.html](https://readr.tidyverse.org/reference/parse_number.html): 该资源是 tidyverse 包中 readr 子包的函数文档页面，详细介绍了函数 `parse_number()` 的用法，该函数用于从字符串中提取数值部分。

- 
33. [https://rlang.r-lib.org/reference/set\\_names.html](https://rlang.r-lib.org/reference/set_names.html): 该 URL ([https://rlang.r-lib.org/reference/set\\_names.html](https://rlang.r-lib.org/reference/set_names.html)) 指向的是 R 语言中 `rlang` 包的 `set_names()` 函数的官方文档页面。该函数用于为向量或列表设置名称，支持直接赋予名称向量，也可以传入一个函数来动态生成名称。在实际应用中，`set_names()` 常用于给数据路径或文件名向量命名，便于后续的数据操作和管理。
  34. [https://stringr.tidyverse.org/reference/str\\_glue.html](https://stringr.tidyverse.org/reference/str_glue.html): 该资源是 `stringr` 包中函数 `str_glue()` 的官方文档页面，介绍了如何使用 `str_glue()` 函数进行字符串插值和拼接，方便地构建包含变量值的字符串，常用于数据处理和生成动态文本。
  35. [https://tidyr.tidyverse.org/reference/pivot\\_longer.html](https://tidyr.tidyverse.org/reference/pivot_longer.html): 该资源是 `tidyr` 包中用于数据“长格式”转换的函数文档，介绍了 `pivot_longer()` 函数的用法和细节。`pivot_longer()` 用于将数据从宽格式转换成长格式，便于按组进行后续的数据操作和分析。
  36. [https://tidyr.tidyverse.org/reference/pivot\\_wider.html](https://tidyr.tidyverse.org/reference/pivot_wider.html): 该资源是 `tidyr` 包中函数 `pivot_wider()` 的官方文档页面，介绍了如何将数据从“长格式”转换为“宽格式”，即将多行数据扩展成更宽的表格形式，常用于数据整形和重构。
  37. [https://tidyr.tidyverse.org/reference/separate\\_wider\\_delim.html](https://tidyr.tidyverse.org/reference/separate_wider_delim.html): 该资源是 `tidyr` 包中函数 `separate_wider_delim()` 的官方文档页面，介绍了如何将包含多个数据片段的字符串（例如文件路径或文件名）按照指定的分隔符拆分成多个列，从而方便数据整理和分析。
  38. [https://tidyr.tidyverse.org/reference/unnest\\_longer.html](https://tidyr.tidyverse.org/reference/unnest_longer.html): 该资源是 `tidyr` 包中函数 `unnest_longer()` 的官方文档，介绍了如何将列表列中的每个元素拆分成新的行，从而实现数据的“展开”操作，便于长格式数据的整理和分析。
  39. [https://tidyr.tidyverse.org/reference/unnest\\_wider.html](https://tidyr.tidyverse.org/reference/unnest_wider.html): 该资源是 `tidyr` 包中函数 `unnest_wider()` 的官方文档页面，介绍了该函数的用法和功能。`unnest_wider()` 用于将列表列（list column）中的元素展开成新的列，从而将嵌套的数据结构“变宽”，便于进一步的数据处理和分析。
  40. <https://tidyselect.r-lib.org/reference/everything.html>: 该资源是 R 语言包 `tidyselect` 中函数 `everything()` 的官方文档页面。该函数用于选择数据框中除分



组变量外的所有列，常用于配合 `dplyr` 包中的 `across()` 函数进行列选择操作，方便对数据集的所有非分组列进行统一处理。

41. [https://tidyselect.r-lib.org/reference/starts\\_with.html](https://tidyselect.r-lib.org/reference/starts_with.html): 该资源是 `tidyselect` 包中用于列选择的函数文档页面，主要介绍了函数 `starts_with()`，该函数用于根据列名的开头字符匹配并选择数据框中的列，常用于 `tidyverse` 生态中如 `dplyr` 的 `select()` 或 `across()` 函数的 `.cols` 参数中，以便灵活地基于列名模式筛选列。
42. <https://tidyselect.r-lib.org/reference/where.html>: 该资源是 R 语言中 `tidyselect` 包的官方文档页面，介绍了函数 `where()`。`where()` 函数用于基于列的数据类型来选择数据框中的列，通常配合 `dplyr` 包中的 `across()` 一起使用，从而灵活地按列类型进行列选择操作。

## 第 27 章

1. <https://dplyr.tidyverse.org/reference/across.html>: 该资源是 `dplyr` 包中函数 `across()` 的官方文档页面。`across()` 用于在数据框的多个列上同时应用函数，是 `tidyverse` 中实现列级批量操作的核心工具之一。文档中详细说明了该函数的用法、参数及示例，帮助用户高效地进行数据变换和汇总操作。
2. <https://dplyr.tidyverse.org/reference/arrange.html>: 该资源是 `dplyr` 包中函数 `arrange()` 的官方文档页面，介绍了 `arrange()` 函数的用法及其功能，即用于对数据框的行进行排序，类似于基于整数向量（通常由 `order()` 函数生成）进行子集化操作。
3. <https://dplyr.tidyverse.org/reference/filter.html>: 该资源是 `dplyr` 包中用于数据框行筛选的函数 `filter()` 的官方文档页面。页面详细介绍了 `filter()` 函数的用法，说明它如何通过逻辑条件筛选数据框的行，并特别指出与处理缺失值 (NA) 相关的行为。文档还将 `filter()` 与 R 语言中类似功能的函数进行了对比，帮助用户理解其特点和应用场景。

4. <https://dplyr.tidyverse.org/reference/mutate.html>: 该资源是 dplyr 包中函数 `mutate()` 的官方文档页面, 用于介绍如何使用 `mutate()` 函数在数据框中创建或修改列。
5. <https://dplyr.tidyverse.org/reference/pull.html>: 该资源是 dplyr 包中函数 `pull()` 的官方文档页面。 `pull()` 函数用于从数据框中提取指定列 (通过列名或列的位置索引), 并返回该列的向量形式, 相当于访问数据框的 `[[ ]]` 或 `$` 操作符。该函数常用于数据变换管道中, 方便提取单列数据进行后续处理。
6. <https://dplyr.tidyverse.org/reference/relocate.html>: 该资源是 dplyr 包中函数 `relocate()` 的官方文档页面, 介绍了 `relocate()` 函数的用法, 主要用于调整数据框中列的位置, 类似于使用字符向量对子集列进行重新排序或重新定位。
7. <https://dplyr.tidyverse.org/reference/select.html>: 该资源是 dplyr 包中用于数据框操作的函数文档页面, 详细介绍了 R 语言中 `select()` 函数的用法。  
`select()` 函数用于从数据框中选择指定的列, 类似于通过列名字符向量对子集进行选择, 是数据整理和变换过程中常用的列子集操作工具。页面提供函数的参数说明、使用示例及相关函数的对比说明。
8. <https://dplyr.tidyverse.org/reference/summarise.html>: 该资源是 dplyr 包中函数 `summarize()` (或 `summarise()`) 的官方文档页面, 介绍了如何使用该函数对数据进行汇总和摘要操作。页面详细说明了 `summarize()` 的用法、参数和示例, 帮助用户快速掌握基于数据框进行分组统计和汇总的技巧。
9. <https://oreil.ly/VF0sY>: (<https://adv-r.hadley.nz/subsetting.html#subset-multiple>) 该资源是《Advanced R》一书中关于“选择多个元素 (Selecting multiple elements)”的章节, 详细介绍了如何对数据框 (data frame) 进行子集操作, 包括将数据框视为一维对象进行子集化, 以及如何使用矩阵对数据框进行子集选择。
10. <https://oreil.ly/evpcw>: (<https://gist.github.com/hadley/c430501804349d382ce90754936ab8ec>) 该资源是 Hadley Wickham 在 GitHub Gist 上整理的一些使用 R 语言中 base 包函数 (如 `tapply()`) 进行分组汇总的示例和技巧代码, 旨在帮助用户更方便地实现多变量分组和汇总操作。

11. <https://oreil.ly/z6vyT>: (<https://gist.github.com/hadley/1986a273e384fb2d4d752c18ed71bedf>) 该资源是 Hadley Wickham 在 GitHub Gist 上收集的一些使用 R 语言中多种基础函数（如 `transform()`、`with()`、`within()`）创建新列的代码示例。
12. <https://purrr.tidyverse.org/reference/map.html>: 该 URL (<https://purrr.tidyverse.org/reference/map.html>) 指向的是 R 语言中 `purrr` 包的 `map` 系列函数的官方文档页面。该资源详细介绍了 `purrr` 包中 `map` 函数的用法及其功能，`map` 函数是用于对列表或向量中的每个元素应用一个函数，类似于基础 R 中的 `lapply()` 函数，但提供了更多高级特性和更一致的接口。文档中还可能包含 `map` 函数的不同变体（如 `map_vec()`、`walk()` 等）的说明及示例，帮助用户理解和高效使用该系列函数进行函数式编程和迭代操作。
13. <https://rdr.io/r/base/apply.html>: 该资源是 R 语言基函数库中关于函数 `apply()` 的文档页面，详细介绍了 `apply()` 函数的用法及其适用范围。`apply()` 函数主要用于对矩阵和数组的行或列应用指定的函数，文档中还提醒了在对数据框（data frame）使用 `apply(df, 2, ...)` 时的性能和安全性问题，建议使用 `lapply(df, ...)` 替代。该页面适合了解 `apply()` 函数的功能、参数及其在数据处理中的应用场景。
14. <https://rdr.io/r/base/cbind.html>: 该资源是 R 语言基础包中关于函数 `cbind()` 的官方文档页面，介绍了如何将多个向量、矩阵或数据框按列组合成一个新的对象。
15. <https://rdr.io/r/base/do.call.html>: 该资源是 R 语言中 `base` 包里函数 `do.call()` 的官方文档页面，介绍了 `do.call()` 函数的用法及示例。`do.call()` 函数用于将一个函数应用于参数列表（通常是一个列表），在上下文中用来结合多个 `tibble` 数据框。
16. <https://rdr.io/r/base/lapply.html>: 该 URL (<https://rdr.io/r/base/lapply.html>) 指向的是 R 语言中基础包（base R）里的函数 `lapply()` 的官方文档页面。`lapply()` 函数是 R 中用于列表或向量等数据结构的迭代处理的核心函数之一，它接受一个列表和一个函数作为输入，对列表中的每个元素应用该函数，并返回一

个列表。该页面详细介绍了 `lapply()` 的用法、参数和示例，帮助用户理解和掌握其功能。

17. <https://rdr.io/r/base/library.html>: 该资源是 R 语言基础包 (base package) 中函数的在线文档，主要介绍了基础 R 环境下的核心函数，如用于加载包的 `library()` 函数。该页面属于 R 语言官方文档网站 `rdr.io`，提供了基础 R 函数的详细说明和用法示例，是学习和使用基础 R 功能的重要参考资料。
18. <https://rdr.io/r/base/mean.html>: 该资源是 R 语言基础包 (base R) 中函数 `mean()` 的官方文档页面，介绍了 `mean()` 函数的用法及其在计算数值型数据平均值方面的功能。
19. <https://rdr.io/r/base/numeric.html>: 该资源是 R 语言基础包 (base R) 中关于数值类型 (numeric) 相关函数和操作的文档说明页面。它详细介绍了与数值型数据处理相关的函数、用法及示例，帮助用户理解和使用数值类型的判断及操作方法，例如文中提到的 `is.numeric()` 函数即用于检测对象是否为数值型。
20. <https://rdr.io/r/base/order.html>: 该资源是 R 语言基础包 (base package) 中关于函数 `order()` 的官方文档页面，介绍了 `order()` 函数的用法和功能。  
`order()` 函数用于生成排序的索引，常用于对数据框的行进行排序或重新排列。
21. <https://rdr.io/r/base/seq.html>: 该资源是 R 语言基础包中关于函数 `seq_along()` 的文档页面，详细介绍了该函数的用法及示例。`seq_along()` 用于生成一个整数序列，其长度与给定对象（如向量或列表）的元素数量相同，常用于迭代对象的索引。
22. <https://rdr.io/r/base/subset.html>: 该资源是 R 语言基础包 (Base R) 中关于函数 `subset()` 的官方文档页面，详细介绍了 `subset()` 函数的用法。该函数结合了 `filter()` 和 `select()` 的功能，用于从数据框或其他对象中基于条件筛选并选择子集。页面内容包括函数的参数说明、使用示例以及相关细节，便于用户理解和应用该函数进行数据子集操作。
23. <https://rdr.io/r/base/sum.html>: 该资源是 R 语言基础包 (base R) 中函数 `sum()` 的官方文档页面，介绍了 `sum()` 函数的用法及其在数值求和计算中的应用。

- 
24. <https://rdr.io/r/base/tapply.html>: 该 URL (<https://rdr.io/r/base/tapply.html>) 指向的是 R 语言基础包中函数 `tapply()` 的文档页面。该资源详细介绍了 `tapply()` 函数的用途和用法, 说明它是 `apply` 系列函数中的一个重要成员, 用于对数据按照指定的分组变量计算分组汇总统计 (即对数据进行分组后计算单个汇总值)。文档还提到 `tapply()` 返回的是带名称的向量, 使用时可能需要额外处理以便将多个汇总结果和分组变量整合成数据框, 适合有一定 R 语言基础的用户参考和学习该函数的具体用法与应用场景。
  25. <https://rdr.io/r/base/transform.html>: 该资源是 R 语言基础包 (base R) 中关于函数 `transform()` 的官方文档页面, 介绍了如何使用 `transform()` 函数对数据框进行操作, 特别是创建或修改新列。
  26. <https://rdr.io/r/base/vector.html>: 该资源是 R 语言基础包中的函数说明页面, 主要介绍了函数 `vector()` 的用法。`vector()` 函数用于创建指定类型和长度的空向量, 在代码示例中用于预先生成与另一个对象 (如 `paths`) 长度相同的列表结构, 方便后续赋值和处理。
  27. <https://rdr.io/r/base/which.html>: 该资源是 R 语言基础包中函数 `which()` 的官方文档页面, 详细介绍了 `which()` 函数的用法及其行为, 包括如何根据逻辑条件索引元素的位置, 常见应用场景, 以及与其他索引操作的区别和注意事项。
  28. <https://rdr.io/r/base/with.html>: 该资源 (<https://rdr.io/r/base/with.html>) 是 R 语言基础包中函数 `with()` 的文档页面, 介绍了如何在特定数据环境下简化表达式的书写, 常用于在数据框中临时引用变量, 从而方便地进行数据操作和新列的创建。
  29. <https://rdr.io/r/graphics/hist.html>: 该资源是关于 R 语言中绘制直方图 (histograms) 的函数 `hist()` 的文档页面, 介绍了如何使用该函数生成数据的直方图, 通常用于展示数据的分布情况。
  30. <https://rdr.io/r/graphics/plot.default.html>: 该资源是 R 语言中基础绘图函数 `plot()` 的官方文档页面, 介绍了如何使用 `plot()` 函数绘制散点图等基本图形。
  31. [https://tidyselect.r-lib.org/reference/starts\\_with.html](https://tidyselect.r-lib.org/reference/starts_with.html): 该资源是 `tidyselect` 包的官方文档页面, 具体介绍了函数 `starts_with()` 的用法。

`starts_with()` 是一个选择辅助函数，用于在数据选择操作中匹配以特定字符串开头的列名，方便用户在数据处理时进行列的筛选和操作。

## 第 28 章

1. <https://gt.rstudio.com/reference/gt.html>: 该 URL 指向的是 R 语言中 `gt` 包的官方文档页面，具体介绍了 `gt::gt()` 函数的用法和功能。该函数用于创建和显示格式化的表格，便于在 R 中生成美观、易读的数据展示表格。
2. <https://lubridate.tidyverse.org/reference/now.html>: 该资源是 R 语言中 `lubridate` 包的函数文档页面，介绍了函数 `now()` 的用法。`lubridate::now()` 用于获取当前的系统时间，通常返回当前的日期和时间，方便在数据处理和时间分析中使用。
3. <https://oreil.ly/38bld>: (<https://yihui.name/knitr/options>) 该资源是关于 Knitr 中代码块 (chunk) 选项的官方或权威说明页面，详细介绍了用于自定义代码块输出的近 60 个配置选项，帮助用户灵活控制代码执行和结果展示的行为。
4. <https://oreil.ly/Auuh2>: (<https://github.com/hadley/r4ds/tree/main/quarto>) 该资源是哈德利·威克姆 (Hadley Wickham) 维护的《R for Data Science》配套的 Quarto 文件集合，位于 GitHub 仓库 `r4ds` 的 `quarto` 目录下，包含用于教学和学习的 R Quarto 文档示例，如 `diamond-sizes.qmd`，用于展示数据可视化和分析的代码与说明。
5. <https://oreil.ly/BDpHv>: (<https://quarto.org/docs/visual-editor/technical.html#citations-from-zotero>) 该资源是 Quarto 官方文档中关于其可视化编辑器的技术说明页面，具体讲解了如何从 Zotero 个人或群组文献库中导入和使用引用文献的功能。
6. <https://oreil.ly/BpPdW>: (<https://www.crossref.org>) 该资源是 Crossref 官方网站，提供学术文献的数字对象唯一标识符 (DOI) 注册与引用链接服务，支持学术研究中的文献检索与引用管理。

7. <https://oreil.ly/EfKFq>: (<https://www.tidyverse.org/blog/2020/08/taking-control-of-plot-scaling>) 该资源是一篇博客文章，标题为“Taking Control of Plot Scaling”，内容主要介绍了图形大小和缩放的技巧与方法，帮助读者理解如何通过反复试验来调整和优化图表的尺寸和比例，从而提升数据可视化的效果。文章发布在 tidyverse.org 网站上，适合对数据可视化、特别是使用 tidyverse 工具包进行绘图的用户阅读学习。
8. [https://oreil.ly/H\\_Xn-](https://oreil.ly/H_Xn-): (<https://doi.org/10.21105/joss.01686>) 该资源是一篇题为“Welcome to the Tidyverse”的学术论文，发表在《Journal of Open Source Software》(开源软件期刊) 上，DOI 为 10.21105/joss.01686。该论文介绍了 Tidyverse 这个 R 语言中用于数据科学的一整套工具包，旨在帮助用户更高效、整洁地进行数据处理和分析。
9. <https://oreil.ly/Hd2Ey>: (<https://pubmed.ncbi.nlm.nih.gov>) 该资源是美国国家医学图书馆提供的 PubMed 数据库，主要用于检索生命科学和生物医学领域的文献资料，包括医学期刊文章、临床研究报告等。
10. <https://oreil.ly/HvFDz>: (<https://yihui.name/knitr>) 该资源是 knitr 的官方网站或介绍页面，knitr 是一个用于执行嵌入代码块并生成包含代码及其输出的 Markdown 文档的工具，常用于 R 语言的动态文档生成和报告制作。
11. [https://oreil.ly/I9\\_I7](https://oreil.ly/I9_I7): (<https://joss.theoj.org/papers/10.21105/joss.01686>) 该资源是一篇题为《Welcome to the Tidyverse》的学术论文，发表在 Journal of Open Source Software (JOSS) 上，DOI 为 10.21105/joss.01686。论文介绍了 Tidyverse 这一用于数据科学和数据分析的 R 语言软件包集合，旨在帮助用户高效、整洁地处理和可视化数据。
12. <https://oreil.ly/QxUsn>: (<https://pandoc.org>) 该资源是 Pandoc 的官方网站，Pandoc 是一个通用的文档转换工具，能够将 Markdown 等格式的文档转换为多种输出格式（如 HTML、PDF、Word 等），广泛用于生成最终的文档文件。
13. <https://oreil.ly/SF3Pm>: (<https://quartopub.com>) 该资源是 Quarto 的官方网站，提供关于 Quarto 这一支持生成自包含 (self-contained) 文档的工具的相关信息和服务。用户可以通过该网站了解如何使用 Quarto 创建、发布和共享包含所有外部依赖（如图片、样式表、脚本等）的完整 HTML 报告。

14. [https://oreil.ly/\\_6LNH](https://oreil.ly/_6LNH): (<https://quarto.org>) 该资源是 Quarto 的官方网站, 提供 Quarto 工具的官方文档和最新信息, 是学习和了解 Quarto 命令行界面的权威渠道。
15. [https://oreil.ly/\\_I4xb](https://oreil.ly/_I4xb): (<https://rstudio.github.io/renv/index.html>) 该资源是一个关于 R 语言包管理工具 renv 的官方文档或教程页面, 介绍如何使用 renv 来实现项目的包版本管理, 从而保证代码的长期可重现性。页面详细说明了如何将项目依赖的包存储在项目目录中, 便于在未来重新创建相同的运行环境。
16. <https://oreil.ly/bYJez>: (<https://github.com/citation-style-language/styles>) 该资源是 “Citation Style Language (CSL) 风格的官方代码仓库”, 托管在 GitHub 上, 网址为 <https://github.com/citation-style-language/styles>。该仓库收集和维持了广泛使用的参考文献和引文格式样式文件, 供用户下载和应用用于文献管理和排版工具中, 用以统一和规范各种学术文献的引用格式。
17. <https://oreil.ly/bzjrw>: (<https://happygitwithr.com>) 该资源是一本名为《Happy Git with R》的书, 面向 R 语言用户, 提供关于 Git 和 GitHub 的友好入门指导, 帮助数据科学从业者更好地进行版本控制和协作。书籍内容免费在线提供, 网址为 <https://happygitwithr.com>。
18. <https://oreil.ly/nlpLD>: (<https://colinpurrington.com/tips/lab-notebooks>) 该资源是一篇由 Colin Purrington 撰写的关于如何有效使用实验室笔记本 (lab notebooks) 的指南, 提供了实用建议和技巧, 帮助科研人员更好地记录和管理实验数据及过程。
19. <https://oreil.ly/nEiGf>: (<https://en.wikipedia.org/wiki/WYSIWYM>) 该 URL (<https://en.wikipedia.org/wiki/WYSIWYM>) 指向的是关于 “WYSIWYM” (What You See Is What You Mean, 即 “所见即所意”) 的维基百科页面。WYSIWYM 是一种文本编辑理念或界面设计方法, 强调编辑时关注内容的语义结构和含义, 而非最终的视觉排版效果。与传统的 “所见即所得” (WYSIWYG) 编辑器不同, WYSIWYM 界面帮助用户专注于文档的内容和逻辑组织, 从而生成格式规范且结构清晰的文档。该页面详细介绍了 WYSIWYM 的定义、原理及其在文本编辑器中的应用。



- 
20. <https://oreil.ly/qS7tS>: (<https://www.georgegopen.com/the-litigation-articles.html>) 该 URL (<https://www.georgegopen.com/the-litigation-articles.html>) 指向的是 George Gopen 关于写作的系列短文, 这些文章主要面向律师, 但其中的写作技巧和原则对数据科学家等其他专业人士同样适用。文章内容聚焦于如何清晰、有效地传达分析结果, 帮助读者提升句子和段落的结构与表达能力。
  21. <https://oreil.ly/sxxlC>: (<https://quarto.org/docs/visual-editor/technical.html#citations-from-dois>) 该资源是 Quarto 官方文档中关于如何在可视化编辑器中使用 DOI (数字对象唯一标识符) 进行引用的技术说明页面, 介绍了通过 DOI 引用文献的方法和相关技术细节。
  22. <https://oreil.ly/vSwdK>: (<https://datacite.org>) DataCite 是一个致力于为研究数据分配数字对象标识符 (DOI) 并促进数据共享与引用的组织, 提供可靠的数据注册和检索服务。
  23. <https://oreil.ly/yvPFt>: (<https://twitter.com/drob/status/738786604731490304>) 该 URL 指向 Twitter 用户 drob (David Robinson) 发布的一条推文, 内容很可能是关于命名代码模块或数据块的建议, 尤其是以主要创建对象命名代码块, 以便更清晰地理解模块之间的依赖关系。
  24. [https://rdr.io/pkg/knitr/man/clean\\_cache.html](https://rdr.io/pkg/knitr/man/clean_cache.html): 该资源是 R 语言包 knitr 中函数 `clean_cache()` 的官方文档页面, 介绍了如何使用 `clean_cache()` 函数来清理和管理 knitr 生成的缓存文件, 帮助用户在复杂缓存策略下保持缓存的整洁和高效。
  25. <https://rdr.io/pkg/knitr/man/kable.html>: 该资源是 R 语言包 knitr 中函数 `kable()` 的官方文档页面, 提供了该函数的详细说明和用法示例。`knitr::kable()` 用于将数据框或矩阵格式化为美观的表格, 方便在 R Markdown 等报告中展示数据。文档中介绍了如何使用该函数生成基本表格, 并提供了进一步自定义表格样式的选项。
  26. <https://rdr.io/r/base/file.info.html>: 该资源是 R 语言基础包 (base package) 中函数 `file.info()` 的文档页面。该函数用于获取文件或目录的详细信息, 例如文件的大小、权限、创建时间和最后修改时间等。文档页面还介绍了相

关函数如 `file.mtime()`，用于返回文件的最后修改时间，常用于监测文件是否发生变化，从而实现缓存更新或数据同步等功能。

27. <https://rdrr.io/r/base/format.html>: 该资源是 R 语言基础包 (base package) 中函数 `format()` 的官方文档页面，详细介绍了如何使用 `format()` 函数对数字进行格式化处理，包括设置显示的有效数字位数 (digits) 和使用千位分隔符 (big.mark) 以提高数字的可读性。
28. <https://rdrr.io/r/base/getwd.html>: 该资源是 R 语言中基础函数 `getwd()` 的官方文档页面，介绍了该函数的用途和用法。`getwd()` 用于获取当前 R 会话的工作目录路径，帮助用户确认当前的工作目录位置。
29. <https://rdrr.io/r/base/print.html>: 该资源是 R 语言中基础包 (base package) 中函数 `print()` 的官方文档页面，详细介绍了 `print()` 函数的用法、参数和示例，帮助用户理解如何在 R 中显示和检查对象内容。
30. <https://rdrr.io/r/utils/sessionInfo.html>: 该资源是 R 语言中函数 `sessionInfo()` 的文档页面，介绍了该函数用于显示当前 R 会话的详细信息，包括 R 版本、操作系统、已加载的包及其版本等。通过记录这些信息，用户可以追踪代码运行时所依赖的软件环境，有助于实现代码的可重复性和环境管理。
31. <https://rdrr.io/r/utils/str.html>: 该资源是 R 语言中 `utils` 包下的函数 `str()` 的在线文档页面，提供了关于 `str()` 函数的详细说明和用法示例。`str()` 函数用于紧凑地显示 R 对象的内部结构，有助于用户快速了解对象的内容和类型，方便调试和数据分析。
32. <https://tibble.tidyverse.org/reference/tribble.html>: 该资源是 R 语言中 `tibble` 包的函数 `tribble()` 的官方文档页面，介绍了如何使用 `tribble()` 函数以一种清晰、直观的方式在 R 代码中录入小规模的数据片段。

## 第 29 章

1. <https://oreil.ly/iGxF>: (<https://quarto.org/docs/output-formats/all-formats.html>) 该资源是 Quarto 官方文档中关于所有输出格式的详细说明页面, 介绍了 Quarto 支持的各种文档和报告的输出格式及其使用方法。
2. <https://oreil.ly/4Id6V>: (<https://mastering-shiny.org>) 该资源是由 Hadley Wickham 编写的《Mastering Shiny》一书的官方网站, 提供关于 R 语言 Shiny 框架的深入学习材料和教程, 旨在帮助用户掌握构建交互式网页应用的技能。
3. <https://oreil.ly/Jg7T9>: (<https://quarto.org/docs/presentations>) 该资源是 Quarto 官方文档中关于如何创建演示文稿的指南页面, 提供使用 Quarto 工具制作演示文稿的详细说明和示例。
4. <https://oreil.ly/JnOwJ>: (<https://presentationpatterns.com>) 该资源是一本由 Neal Ford、Matthew McCollough 和 Nathaniel Schutta 合著的书籍《Presentation Patterns》, 提供了一系列实用的演示技巧和模式, 帮助读者提升演讲和展示的效果。
5. <https://oreil.ly/LQZud>: (<https://bwlewis.github.io/rthreejs>) 该资源是一个基于 three.js 的交互式三维绘图工具或示例, 展示如何利用 three.js 库进行三维数据的可视化。
6. <https://oreil.ly/P-n37>: (<https://quarto.org/docs/websites>) 该资源是 Quarto 官方网站上的文档页面, 具体介绍了如何使用 Quarto 创建和管理网站的相关内容。
7. <https://oreil.ly/SE3qV>: (<https://rstudio.github.io/dygraphs>) 该资源是一个名为“dygraphs”的工具或库, 用于创建交互式时间序列数据的可视化, 网址为 <https://rstudio.github.io/dygraphs>, 提供相关文档和示例, 方便用户在 R 语言环境中实现动态的时间序列图表。
8. <https://oreil.ly/ST4yc>: (<https://github.com/jtleek/talkguide>) 该资源是“Leek 小组关于如何做学术报告的指南”, 旨在帮助学术演讲者提升演讲技巧和效果。

9. <https://oreil.ly/fiB1h>: (<https://quarto.org/docs/books>) 该资源是 Quarto 官方网站上关于如何使用 Quarto 创建和发布电子书 (books) 的文档说明, 详细介绍了 Quarto 书籍格式的功能、用法和示例。
10. <https://oreil.ly/gQork>: (<https://rich-iannone.github.io/DiagrammeR>) 该资源是 DiagrammeR 的官方网站, 提供用于创建流程图和简单节点链接图等图表的工具和相关文档。
11. <https://oreil.ly/l3tFl>: (<https://rstudio.github.io/DT>) 该资源是 R 语言中用于创建交互式数据表的包 DT 的官方文档或介绍页面, 提供如何使用 DT 包构建动态、可排序、可搜索的数据表的说明和示例。
12. <https://oreil.ly/lXY9u>: (<https://www.coursera.org/learn/public-speaking>) 该资源是一个由 Matt McGarrity 主讲的在线公开演讲课程, 提供于 Coursera 平台, 旨在帮助学习者提升公共演讲技能。
13. <https://oreil.ly/lmdha>: (<https://www.htmlwidgets.org>) 该资源是一个关于 htmlwidgets 的官方网站或资源中心, 提供有关 htmlwidgets 的详细信息以及相关软件包的完整列表, 方便用户学习和使用 htmlwidgets 技术。
14. <https://oreil.ly/mhYNQ>: (<https://quarto.org/docs/output-formats/all-formats.html>) 该资源是 Quarto 官方文档中关于所有支持的输出格式的完整列表页面, 详细介绍了 Quarto 支持的各种输出格式及其对应的配置选项, 帮助用户了解不同格式的通用和特定输出选项, 如目录生成、代码折叠等功能的适用性和使用方法。
15. <https://oreil.ly/ovWgb>: (<https://quarto.org/docs/journals/templates.html>) 该资源是 Quarto 官方提供的期刊文章模板文档, 介绍如何使用 Quarto 的期刊模板来撰写和排版学术期刊文章。
16. <https://oreil.ly/q-E7l>: (<https://quarto.org/docs/reference/formats/ipynb.html>) 该资源是 Quarto 官方文档中关于将 Quarto 文档导出为 Jupyter Notebook 格式 (即 .ipynb 文件) 的参考说明页面。页面介绍了如何使用 `format: ipynb` 选项, 将 Quarto 文档输出为可在 Jupyter 环境中打开和运行的笔记本文件。