

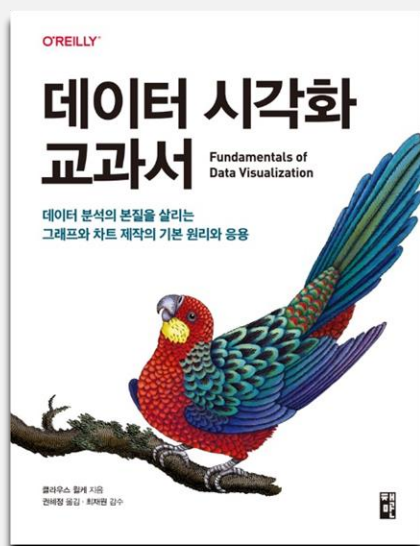
『데이터 시각화 교과서』 원서 깃북 렌더링 방법

= 소스코드 설치와 실행환경 =

고 석 범

신경과 전문의, R 유저 및 저술가, (주)지노바이오 대표

『통계 분석 너머 R의 무궁무진한 활용』(에이콘출판, 2017) 외 다수 저술, 번역



데이터 시각화 교과서

데이터 분석의 본질을 살리는
그래프와 차트 제작의 기본 원리와 응용

클라우드 월케 지음 | 권혜정 옮김 | 최재원 감수 |

376 쪽 | 28,000 원

2020 년 2 월 20 일 출간 |

185*240*16 | ISBN 97989909109

<https://www.onlybook.co.kr/entry/dataviz>

독자 문의 support@onlybook.co.kr



이 책은 저자가 밝힌 대로 특정 언어나 특정 도구를 염두에 두고 쓰이지는 않았다. 하지만 실용적인 측면에서는 예제 코드를 확인하고 실제로 플롯을 만들어 보는 것도 중요하다고 생각해 이 부록을 준비했다.

소스코드와 기본 자료

『데이터 시각화 교과서』 원서인 『Fundamentals of Data Visualization』의 원고는 R 마크다운을 기본으로 하여 깃북으로 렌더링해 주는 북다운(bookdown) 패키지를 사용해 작성되었으며, 전체 프로젝트는 깃허브 사이트(<https://github.com/claustwilke/dataviz>)에 올려져 있다.

실제 마지막으로 렌더링된 책은 <https://serialmentor.com/dataviz/>에서 볼 수 있으며, 북다운으로 만들어진 책들을 호스팅해 누구나 볼 수 있게 하는 북다운 사이트(<https://bookdown.org/>)에서도 확인할 수 있다.

테스트 환경

저자는 R 3.5.0 버전을 사용했다고 했는데, 렌더링은 이 글을 쓰는 2020년 2월 기준 최신 버전인 R 3.6.2을 사용해, RStudio 통합개발환경에서 테스트했다. 이 글을 쓰는 시점에서 이 가이드 문서의 실행환경은 다음과 같으며, 각각의 다운로드 경로도 참조하기 바란다.

- R 버전 3.6.2 (<https://cran.r-project.org/bin/windows/base/>)
- RStudio 버전 1.3.776 (<https://rstudio.com/products/rstudio/download/#download>)
- Rtools 버전 1.2.5033 (<https://cran.r-project.org/bin/windows/Rtools/>)

사전에 검토할 내용

저자가 사용한 R 패키지들이 공식 패키지 저장소인 CRAN(<https://cran.r-project.org>)에 등재되어 있다면 좋았을 테지만, 저자가 개인적으로 만들어 자신의 깃허브에 올린 패키지들을 사용하므로 이 패키지를 독자 여러분이 개인 컴퓨터에 설치하려면 다음과 같은 사전 작업이 필요하다.

R 빌드 툴과 R devtools 패키지 설치

CRAN에 공식 등재되어 있지 않은 R 패키지를 인스톨하기 위해서는, R devtools 패키지 또는 devtools의 서브패키지인 remotes 패키지를 사용할 필요가 있다. 그리고 깃허브에 있는 R 패키지들을 인스톨할 때 코드 빌드(컴파일)가 필요한 경우가 많기 때문에 R 관련 빌드 툴을 설치할 필요가 있다. 만약 컴퓨터에 devtools 패키지가 이미 설치된 경우라면 다음 절로 건너뛰어도 좋다.

R 빌드 툴은 다음과 같은 절차로 설치한다. 설치할 때는 대부분의 기본 옵션을 따라가면 된다.

- (1) 윈도우 사용자인 경우: CRAN에서 Rtools를 설치한다.



CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

About R
[R Homepage](#)
[The R Journal](#)

Software
[R Sources](#)
[R Binaries](#)
[Packages](#)
[Other](#)

Documentation
[Manuals](#)
[FAQs](#)
[Contributed](#)



Subdirectories:

[base](#)

[contrib](#)

[old contrib](#)

[Rtools](#)

Binaries for base distribution. This is what you want to [install R for the first time](#).

Binaries of contributed CRAN packages (for R >= 2.13.x; managed by Uwe Ligges). There is also information on [third party software](#) available for CRAN Windows services and corresponding environment and make variables.

Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).

Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

(2) 맥 사용자인 경우: 엑스코드(Xcode)와 커맨드라인 툴을 사용한다.

R 빌드 툴이 잘 설치되었다면 RStudio 나 베이스 R 에서 devtools 패키지를 설치한다.

```
install.packages("devtools")
```

cowplot 패키지에서 사용되는 글꼴 설치

저자가 만든 cowplot 이라는 패키지에서 Myriad Pro 라고 하는 글꼴이 사용된다. 폰트가 없는 경우에는 웹을 검색하여 저작권을 확인하여 맥 또는 윈도우 컴퓨터에 이 글꼴을 설치하자. 지적 소유권을 보장하지 않는다.

아니면 showtext 같은 R 패키지를 사용하여 cowplot 패키지 함수들을 사용할 때 자신이 원하는 폰트를 사용할 수 있게 해도 좋다.

CRAN 에 등재되지 않은 패키지 설치

저자가 <https://github.com/clauswilke/dataviz> 프로젝트를 저술하면서 사용한 패키지들은 일부 CRAN 에 등재되어 있지 않기 때문에 다음과 같은 코드로 패키지들을 설치해야 한다. 패키지 의존성을 고려해 다음 순서를 그대로 따라서 패키지를 설치하길 권한다. 아울러 이 과정을 진행시킬 때 의존하는 패키지들을 컴파일할지를 물어온다. 그런 경우에는 “All”을 선택하여 설치한다.

윈도우의 경우 ERROR: failed to create lock directory 이라는 오류가 발생하면서 의존하는 패키지들이 설치되지 않는 경우가 발생할 수 있는데 그런 경우에는 다음과 같이 실행한다. 예를 들어 rlang 패키지를 빌드하다가 위와 같은 오류가 발생하면 다음과 같이 실행한다.

```
install.packages("rlang", dependencies = TRUE, INSTALL_opts = '--no-lock')
```

이어서 설치할 패키지들이다.

```
devtools::install_github("clauswilke/colorblindr")
devtools::install_github("clauswilke/dviz.supp")
devtools::install_github("wilkelab/cowplot")
```

```
devtools::install_github("wilkelab/ungeviz")
devtools::install_github("clauswilke/ggtextures")
devtools::install_github("thomasp85/transformr")
devtools::install_github("hrbrmstr/statebins") # 최신 버전
```

statebins 패키지는 CRAN에 정식 등재된 패키지이지만 이것을 사용하면 오류가 생겼다. 따라서 위 주소의 깃허브에 있는 최신 버전을 사용해야 한다.

CRAN에 등재된 패키지 설치

이 책의 실행환경에서는 다음과 같은 패키지들이 사용되므로, 다음 코드를 따라 설치한다. 간혹 코드를 실행하다가 여기에 나열되지 않은 패키지가 필요하다고 오류 메시지를 만날 수도 있다. 그런 경우에는 해당 패키지를 설치하고 나서 다시 실행하길 바란다.

```
install.packages(c(
  "bookdown", "cowplot",
  "tidyverse", "rgeos", "patchwork", "lwgeom",
  "ggthemes", "ggforce", "ggmap", "geofacet",
  "ggstance", "treemapify", "ggrepel",
  "tidybayes", "ggstance", "plot3D", "tinter",
  "ggribes", "emmeans", "mgcv", "sf", "showtext",
  "magick", "nycflights13", "rwordlmap",
  "ganimate"
))
```

저자의 소스 프로젝트를 다운로드 후, 프로젝트 열기

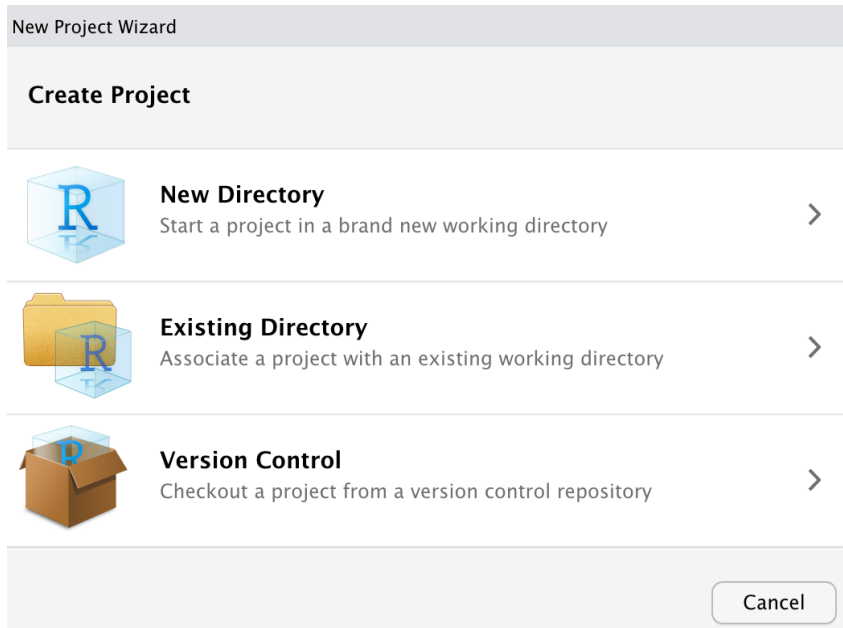
이제 저자의 <https://github.com/clauswilke/dataviz>에 있는 소스 프로젝트를 가지고 온다. 가지고 오는 방법에는 ‘클로닝’과 ‘다운로드’의 2가지 방법이 있다. 되도록 클로닝으로 실행하기를 추천한다.

◆ 클로닝

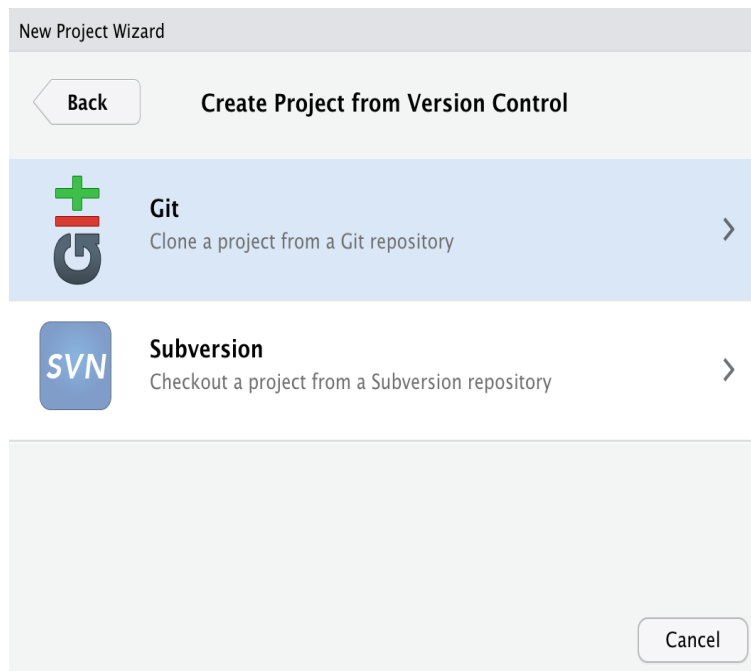
이 방법은 깃(Git)을 사용하는 방법이다. 깃이 설치되어 있고, 깃을 사용할 수 있도록 설정이 완료된 경우라면 이 방법을 사용한다.

- (1) RStudio를 실행해 오른쪽 위 모서리에 있는 프로젝트 버튼을 클릭해 **New Project...**를 선택한다.

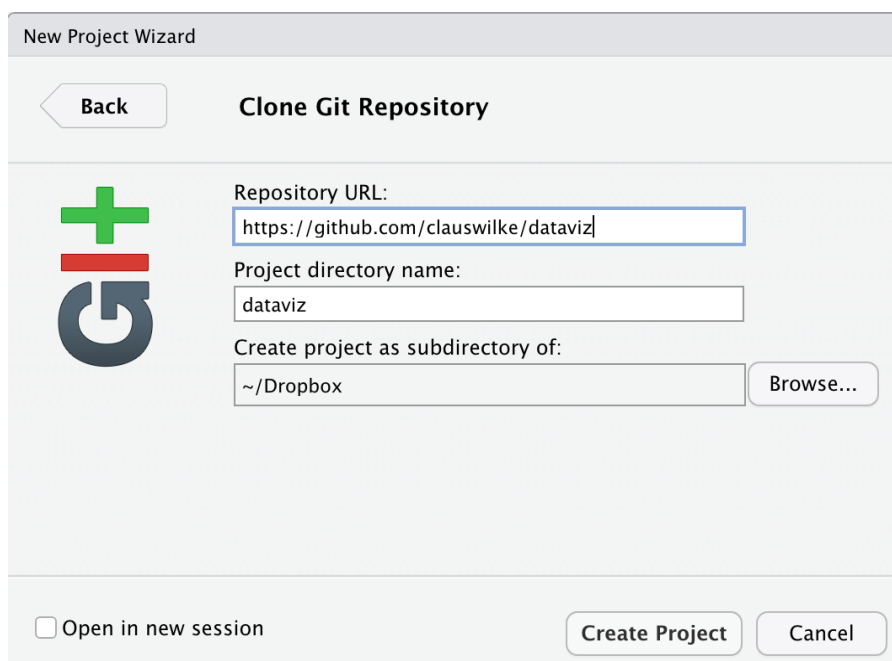
(2) 그다음 창에서 **Version Control** 을 선택한다.



(3) 다음 창에서 **Git** 을 선택한다.



(4) 다음 창에서 “Repository URL”에 저자의 깃허브 사이트 URL 을 입력한다. 가장 밑에 있는 칸은 이 프로젝트로 놓을 폴더를 지정한다. 그런 다음 **Create Project** 를 선택한다. 그리고 나서 프로젝트 클로닝이 끝나면 RStudio 가 다시 프로젝트에 맞춰 실행된다. 정확히 열리면 오른쪽 상단 모서리에 **dataviz** 가 보일 것이다.



◆ 다운로드 방식으로 사용하는 방법

- (1) 저자의 깃허브 사이트 <https://github.com/clauserwilke/dataviz> 로 가서, 오른쪽에 보이는 초록색 **Clone or Download** 를 클릭하고 **Download Zip** 을 클릭한다.
- (2) 다운로드되는 파일은 Zip 으로 압축되어 있다. 이 압축 파일을 원하는 위치에 푼다.
- (3) RStudio 를 실행해 이 프로젝트를 연다. 프로젝트 이름은 dataviz.Rproj 이다. RStudio 의 메뉴 File >> Open Project...를 클릭해 압축 해제한 폴더를 지목하면 열린다. 제대로 열리면 RStudio 오른쪽 상단 모서리에 dataviz-master 라고 프로젝트 이름이 뜰 것이다(클로닝과 조금 다르다).

_book.yml 파일 만들기

저자의 소스 프로젝트는 bookdown 패키지로 만들어져 있는데, 저자는 _bookdown_draft.yml, _bookdown_final.yml, _bookdown_production.yml 파일들을 만들어서 사용하고, 이것을 셸 명령으로 실행한다. 우리는 로컬 컴퓨터에서 렌더링할 것이므로 _bookdown.yml 파일을 만들어 사용한다. 이 파일은 R bookdown 이 책을 렌더링할 때 사용하는 파일이다(자세한 내용은 <https://bookdown.org/yihui/bookdown/output-formats.html> 참조).

- RStudio 에서 **Files** 라는 창에서 _bookdown_production.yml 파일을 클릭하면 이 파일이 편집창에 열린다.
- RStudio 의 메뉴 **File > Save As...**을 선택해 _bookdown.yml 이라는 이름으로 저장한다. 저장 위치는 프로젝트의 루트 디렉터리다.

_output.yml 파일의 내용을 일부 코멘트 처리(옵션)

output.yml 파일을 열어 보면 다음과 같이 되어 있다.

```
bookdown::gitbook:
  css: css/style.css
  split_by: chapter
  config:
    toc:
      collapse: section
      before: |
        <li><a href=".">Data Visualization</a></li>
      after: |
        <li><a href="https://github.com/rstudio/bookdown" target="blank">Published with bookdown</a></li>
bookdown::html_chapters:
  css: css/style.css
```

북다운 패키지가 책을 빌드할 때 여기에 정의된 포맷으로 책을 빌드하는데, 여러 포맷을 정의할 수도 있다. 그런 경우 순차적으로 빌드된다. 아마 독자들은 깃북 형태를 선호할 것이므로, 끝 부분의 포맷을 코멘트 처리한다.

```
# bookdown::html_chapters:
# css: css/style.css
```

그러면 다음과 같이 수정된다.

```
bookdown::gitbook:
  css: css/style.css
  split_by: chapter
  config:
    toc:
      collapse: section
      before: |
        <li><a href=".">Data Visualization</a></li>
      after: |
        <li><a href="https://github.com/rstudio/bookdown" target="blank">Published with
bookdown</a></li>
# bookdown::html_chapters:
# css: css/style.css
```

파일 visualizing_amounts.Rmd 수정하기

visualizing_amounts.Rmd 파일에서 boxoffice-gross 레이블을 가진 청크를 실행할 때 에러가 발생하여 다음과 같이 수정할 것을 권한다. 먼저 파일을 열어 boxoffice-gross 라는 코드 청크를 찾는다.

```
# source: Box Office Mojo
# URL: http://www.boxofficemojo.com/weekend/chart/?view=&yr=2017&wknd=51&p=.htm
# downloaded: 2018-02-11
boxoffice <- data.frame(rank = 1:5,
  title = c("Star Wars: The Last Jedi", "Jumanji: Welcome to the Jungle", "Pitch Perfect 3",
"The Greatest Showman", "Ferdinand"),
  title_short = c("Star Wars", "Jumanji", "Pitch Perfect 3", "Greatest Showman",
```

```

"Ferdinand"),
  amount = c(71565498, 36169328, 19928525, 8805843, 7316746),
  amount_text = c("$71,565,498", "$36,169,328", "$19,928,525", "$8,805,843", "$7,316,746"))
boxoffice_display <- boxoffice %>%
  mutate(A = " ", B = " ", C = " ") %>%
  select(A, rank, title, amount_text, B, C) %>%
  rename(`` = A,
         Rank = rank,
         Title = title,
         `Weekend gross` = amount_text,
         `` = B,
         `` = C)
knitr::kable(
  boxoffice_display,
  caption = 'Highest grossing movies for the weekend of December 22-24, 2017. Data source: Box Office Mojo (http://www.boxofficemojo.com/). Used with permission', booktabs = TRUE,
  row.names = FALSE,
  align = c('c', 'c', 'l', 'r', 'c', 'c')#,
  #format = "html",
  #table.attr = "style = \"width: 75%\""
)

```

여기서 문제가 되는 것은 중간 `boxoffice_display` 객체를 만들 때 사용된 코드다. 아마도 `dplyr` 이전 버전에서 같은 이름의 열을 허용했었던 것 같다(그런 용도로 사용해 보지 않아서 정확히는 모른다.) 이 부분을 다음 부분과 같이 교체한다.

```

library(dplyr)
# source: Box Office Mojo
# URL: http://www.boxofficemojo.com/weekend/chart/?view=&yr=2017&wknd=51&p=.htm
# downloaded: 2018-02-11

boxoffice <- data.frame(rank = 1:5,
  title = c("Star Wars: The Last Jedi", "Jumanji: Welcome to the Jungle", "Pitch Perfect 3",
"The Greatest Showman", "Ferdinand"),
  title_short = c("Star Wars", "Jumanji", "Pitch Perfect 3", "Greatest Showman", "Ferdinand"),
  amount = c(71565498, 36169328, 19928525, 8805843, 7316746),
  amount_text = c("$71,565,498", "$36,169,328", "$19,928,525", "$8,805,843", "$7,316,746"))

boxoffice_display <- boxoffice %>%
  select(rank, title, amount_text) %>%
  rename(
    Rank = rank,
    Title = title,
    `Weekend gross` = amount_text
  )

knitr::kable(
  boxoffice_display,
  caption = 'Highest grossing movies for the weekend of December 22-24, 2017. Data source: Box Office Mojo (http://www.boxofficemojo.com/). Used with permission',
  booktabs = TRUE,
  row.names = FALSE,
  align = c('c', 'c', 'l', 'r', 'c', 'c'),
  format = "html",
  #table.attr = "style = \"width: 75%\""
)

```


챕터별로 코드 실행해 보기

이제 코드를 실행할 준비는 마쳤다. bookdown 패키지를 아는 독자라면 굳이 설명을 하지 않아도 알겠지만, 모른다고 가정하고 설명한다. RStudio 를 사용해 실행한다고 가정한다. 책의 순서는 _bookdown.yml 파일을 열어 보면 알 수 있다. rmd_files 라는 항목을 보자. 일부러 파일과 해당되는 챕터를 명시해 놓았으니 참고하기 바란다.

```
rmd_files: [
  "index.Rmd", # 목차

  "preface.Rmd", # 들어가며
  "introduction.Rmd", # 1 장 시작하며: 이 책의 내용과 구성

  # 1 부 데이터, 시각화로 날개를 달다
  "aesthetic_mapping.Rmd", # 2 장 데이터 시각화: ‘읽는’ 데이터에서 ‘보는’ 데이터로
  "coordinate_systems_axes.Rmd", # 3 장 위치 스케일: 좌표계와 축
  "color_basics.Rmd", # 4 장 색상 스케일
  "directory_of_visualizations.Rmd", # 5 장 다양한 시각화 방식
  "visualizing_amounts.Rmd", # 6 장 수량 데이터의 시각화
  "visualizing_distributions_I.Rmd", # 7 장 데이터 분포의 시각화: 히스토그램과 밀도 도표
  "visualizing_distributions_II.Rmd", # 8 장 데이터 분포의 시각화: 경험적 누적 분포 함수와 QQ 도표
  "boxplots_violins.Rmd", # 9 장 여러 분포 상태의 결합 시각화
  "visualizing_proportions.Rmd", # 10 장 비율 데이터의 시각화
  "nested_proportions.Rmd", # 11 장 비율 데이터를 네포 형태로 시각화
  "visualizing_associations.Rmd", # 10 장 여러 정량 변수의 관계 시각화
  "time_series.Rmd", # 13 장 독립 변수의 시계열 데이터와 함수 시각화
  "visualizing_trends.Rmd", # 14 장 추세의 시각화
  "geospatial_data.Rmd", # 15 장 지리공간 데이터의 시각화
  "visualizing_uncertainty.Rmd", # 16 장 불확실성의 시각화

  # 2 부 그래프 디자인의 기본 원칙

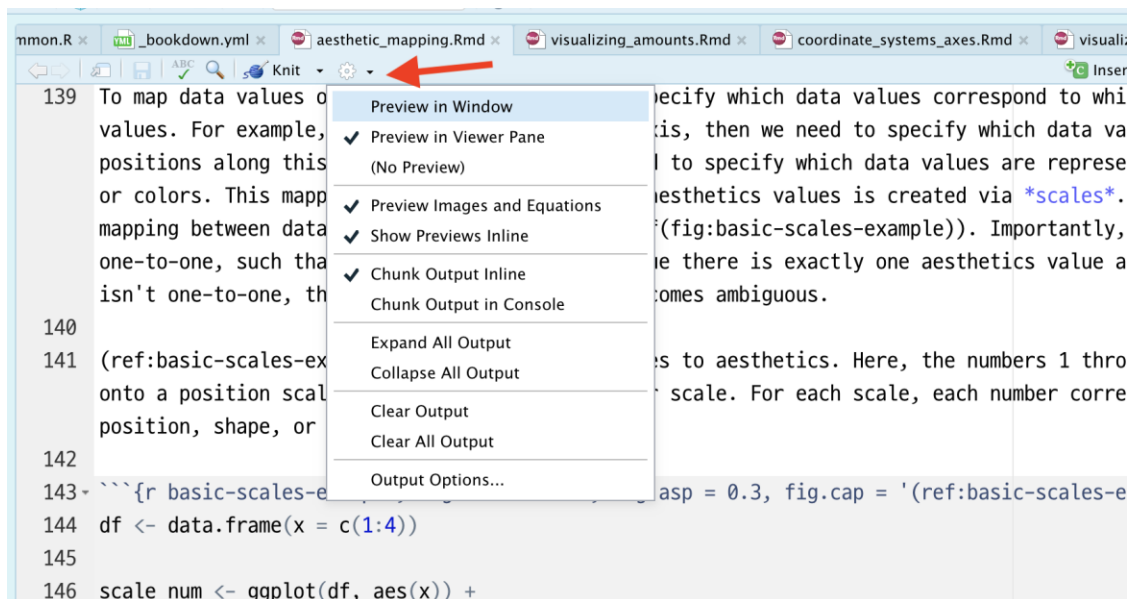
  "proportional_ink.Rmd", # 17 장 잉크 양 비례의 원칙
  "overlapping_points.Rmd", # 18 장 오버플로팅: 겹치는 점 처리 방법
  "pitfalls_of_color_use.Rmd", # 19 장 효과적인 색 사용을 위한 조언
  "redundant_coding.Rmd", # 20 장 불필요한 기호화
  "multi-panel_figures.Rmd", # 21 장 여러 개의 패널로 구성된 도표
  "figure_titles_captions.Rmd", # 22 장 제목, 캡션, 표의 효과적 활용
  "balance_data_context.Rmd", # 23 장 데이터 이해를 돕는 시각화의 주변 요소
  "small_axis_labels.Rmd", # 24 장 축 레이블의 글자는 큼직하게
  "avoid_line_drawings.Rmd", # 25 장 선 그림은 피하자
  "no_3d.Rmd", # 26 장 3 차원 그래픽과 차트는 그만

  # 3 부 시각화 레벨업을 위한 꿀팁
  "image_file_formats.Rmd", # 27 장 가장 흔히 쓰는 이미지 파일 형식
  "choosing_visualization_software.Rmd", # 28 장 올바른 시각화 소프트웨어 선택법
  "telling_a_story.Rmd", # 29 장 스토리텔링과 요점 전달
```

```
"annotated_bibliography.Rmd", # 더 읽을거리
"technical_notes.Rmd", # 소스코드 설치와 실행환경
"references.Rmd" # 참고문헌
]
```

공부하고자 하는 챕터에 해당하는 파일을 연다. 예를 들어 “2 장 데이터 시각화: ‘읽는’ 데이터에서 ‘보는 데이터’로” 장을 공부할 때는 “aesthetic_mapping.Rmd” 파일을 연다. RStudio **Files** 창에서 해당 파일을 클릭해도 되고, 편집창 위에 빈칸에 희미한 글씨로 된 **Go to file/function** 메뉴 칸에 파일 이름을 넣으면 쉽게 열 수 있다.

파일이 열리면 편집창 상단에 **Knit** 라는 버튼이 보인다. 이 버튼을 클릭하면 이 파일에 있는 모든 코드가 실행되고, 텍스트도 프로세싱한 다음에 렌더링된 결과를 볼 수 있다. 먼저 결과를 어디서 볼지를 결정한다. 편집창 위에 기어 모양 오른쪽의 작은 화살표를 클릭한다.



- **Preview in Window** 는 새로운 창에서 결과를 보여준다.
- **Preview in Viewer Pane** 은 RStudio **Viewer** 창에서 결과를 보여준다. 보통 이것을 많이 사용한다.


이제 **Knit** 버튼을 클릭한다. 그럼 우리가 지정한 곳에 문서 렌더링되어 결과를 볼 수 있다.

만약 어떤 패키지가 없어 진행이 안 되는 경우 해당 패키지를 설치하고 다시 **Knit** 버튼을 클릭한다. 만약 출력되어 합쳐진 _main.Rmd 파일이 없어 진행이 안 되는 경우에는 RStudio **Files** 창에서 이 파일을 찾아서 삭제하고 다시 진행시키면 된다.

코드 청크별로 실행해보기

만약 코드 청크별로 실행하려면 다음과 같이 한다. “aesthetic_mapping.Rmd” 파일에서 `basic-scales-example` 이라는 코드 청크를 실행해 보고 싶을 수 있다.

```
142 position, shape, or color and vice versa.
143- ```{r basic-scales-example, fig.width = 5.5, fig.asp = 0.3, fig.cap = '(ref:basic-scales-example)'}
144 df <- data.frame(x = c(1:4))
145
146 scale_num <- ggplot(df, aes(x)) +
147   geom_point(size = 3, color = "#0072B2", y = 1) +
148   scale_y_continuous(limits = c(0.8, 1.2), expand = c(0, 0), breaks = 1, label = "position ") +
149   scale_x_continuous(limits = c(.7, 4.4), breaks = 1:5, labels = c("1", "2", "3", "4", "5"), name = NULL, position
150     = "top") +
151   theme_dviz_grid() +
152   theme(axis.ticks.length = grid::unit(0, "pt"),
153     axis.text = element_text(size = 14),
154     axis.title.y = element_blank(),
155     axis.ticks.y = element_blank())
156
157 scale_color <- ggplot(df, aes(x, color = factor(x), fill = factor(x))) +
158   geom_point(size = 5, shape = 22, y = 1) +
159   scale_y_continuous(limits = c(0.8, 1.2), expand = c(0, 0), breaks = 1, label = "color ") +
160   scale_x_continuous(limits = c(.7, 4.4), breaks = NULL) +
161   scale_color_manual(values = darken(c("#0082A6", "#4EBB9", "#9CDFC2", "#D8F0CD"), .1), guide = "none") +
162   scale_fill_manual(values = c("#0082A6", "#4EBB9", "#9CDFC2", "#D8F0CD"), guide = "none") +
163   theme_dviz_grid() +
164   theme(axis.ticks.length = grid::unit(0, "pt"),
165     axis.text.x = element_blank(),
```

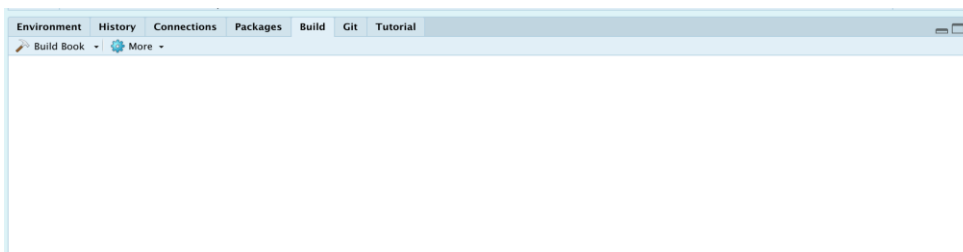


R 마크다운에서 하나의 파일에 들어 있는 코드 청크들은 순서대로 실행되게 되어 있다. 현재 코드 청크가 이전 코드 청크와 완전히 독립된 코드 청크라면, 예를 들어 이전 코드 청크에서 부른 패키지나 객체들을 전혀 사용하지 않는 경우라면, 그림에서 2 번이라고 표시된 버튼을 클릭하여 현재 코드 청크를 실행시킬 수 있다.

그러나 보통은 앞에서 패키지를 로딩해야 할 수도 있고, 이전 청크에서 만든 것을 사용해야 하는 경우가 많다. 그래서 먼저 그림에서 1 번이라고 표시된 버튼을 클릭하여, 이전까지 코드를 모두 실행시키고 나서, 2 번 버튼을 클릭하여 현재 코드 청크의 코드를 실행시킨다.

깃북으로 렌더링하기

어떤 독자는 챕터별로 코드를 실행시키고 나서, 이것을 깃북으로 렌더링하고 싶을 수도 있을 것 같다. 그런 경우에는 RStudio 의 **Build** 라는 창을 찾고, **Build Book** 을 클릭한다.



이런 경우에는 `_main.Rmd` 파일 때문에 진행이 안 된다는 오류가 생기면 이 파일을 찾아서 삭제하고 다시 빌딩하면 된다.