

# Basic elements of file management

ME447/547 Visualizing Data

---

Richard Layton

Spring 2019

Rose-Hulman Institute of Technology

# Effective file management requires two actions














Plan the structure at  
the beginning

- carpentry
- data
- data-raw
- design
- figures
- manage
- practice
- reports
- resources



























Explicitly link files

```
source("carpentry/filename.R")  
read_csv("data/filename.csv")  
source("design/filename.R")  
include_graphics("figures/filename.png")
```














# You recently created the mandatory project-directory structure

 carpentry	dir
 data	dir
 data-raw	dir
 design	dir
 figures	dir
 manage	dir
 practice	dir
 reports	dir
 resources	dir
 .gitignore	file
 .Renviron	file
 README.Rmd	file
 portfolio.Rproj	file













# Open **portfolio.Rproj** to start every work session

 carpentry	
 data	
 data-raw	
 design	
 figures	
 manage	
 practice	
 reports	
 resources	
 .gitignore	
 .Renviron	
 README.Rmd	
 portfolio.Rproj	 Sets the project directory as the working directory














# README introduces your portfolio to the reader

 carpentry	<
 data	<
 data-raw	<
 design	<
 figures	<
 manage	<
 practice	<
 reports	<
 resources	<
 .gitignore	<
 .Renviron	<
 README.Rmd	< Creates the main page of your portfolio website
 portfolio.Rproj	<














## Other top-level files perform administrative duties

 carpentry	<
 data	<
 data-raw	<
 design	<
 figures	<
 manage	<
 practice	<
 reports	<
 resources	<
 .gitignore	< Directs Git to ignore specific files
 .Renvirom	< Stores packages in a library separate from base R
 README.Rmd	<
 portfolio.Rproj	<

## Raw data are never edited manually



























 carpentry	◀
 data	◀
 data-raw	◀ Data in its original form
 design	◀
 figures	◀
 manage	◀
 practice	◀
 reports	◀
 resources	◀
 .gitignore	◀
 .Renviron	◀
 README.Rmd	◀
 portfolio.Rproj	◀

# Data carpentry converts raw data to tidy data














 carpentry	◁ R scripts that create and save tidy data
 data	◁ Tidy data saved here, read by design scripts
 data-raw	◁
 design	◁
 figures	◁
 manage	◁
 practice	◁
 reports	◁
 resources	◁
 .gitignore	◁
 .Renviron	◁
 README.Rmd	◁
 portfolio.Rproj	◁
















# Graph design converts to tidy data to graphs

 carpentry	
 data	
 data-raw	
 design	 R scripts that create and save graphs
 figures	 Graphs saved here, imported by report scripts
 manage	
 practice	
 reports	
 resources	
 .gitignore	
 .Renviron	
 README.Rmd	
 portfolio.Rproj	














# Reports commingle data, scripts, graphs, prose, and references

 carpentry	◀
 data	◀
 data-raw	◀
 design	◀
 figures	◀
 manage	◀
 practice	◀
 reports	◀ One Rmd report per graph
 resources	◀
 .gitignore	◀
 .Renviron	◀
 README.Rmd	◀ Provides links to individual reports
 portfolio.Rproj	◀














## Resource files support the portfolio appearance and format

 carpentry	<
 data	<
 data-raw	<
 design	<
 figures	<
 manage	<
 practice	<
 reports	< Reports explicitly call on resource files
 resources	< Image downloads and bibliography files
 .gitignore	<
 .Renviron	<
 README.Rmd	<
 portfolio.Rproj	<

# Reduce clutter by excusing some resources from version control

 carpentry	<
 data	<
 data-raw	<
 design	<
 figures	<
 manage	< Correspondence and project management
 practice	< Scripts for practicing and learning R
 reports	<
 resources	<
 .gitignore	< Directs Git to ignore specific files
 .Renviron	<
 README.Rmd	<
 portfolio.Rproj	<

# Planning at the start of a project saves time later

 carpentry	◁ R scripts that create and save tidy data
 data	◁ Tidy data saved here, read by design scripts
 data-raw	◁ Data in its original form
 design	◁ R scripts that create and save graphs
 figures	◁ Graphs saved here, imported by report scripts
 manage	◁ Correspondence and project management
 practice	◁ Scripts for practicing and learning R
 reports	◁ Reports explicitly call on resource files
 resources	◁ Image downloads and bibliography files
 .gitignore	◁ Directs Git to ignore specific files
 .Renviron	◁ Stores packages in a library separate from base R
 README.Rmd	◁ Creates the main page of your portfolio website
 portfolio.Rproj	◁ Sets the project directory as the working directory

Files are explicitly linked

---

# Workflow and explicit links using relative file paths

 carpentry directory

```
tidy-data-01.R
```

reads file from raw-data directory

```
read_csv("data-raw/first-data-set-raw.csv")
```

produces tidy data and saves to data directory

```
write_csv("data/first-data-set-tidy.csv")
```

# Workflow and explicit links using relative file paths

📄 reports directory

`d1-report.Rmd`

includes code chunks that run the R script to produce the tidy data

```
source("carpentry/tidy-data-01.R")
```

execute the R script to create the graph

```
source("design/d1-graph.R")
```

and imports the graph

```
include_graphics("figures/d1-graph.png")
```



# Workflow and explicit links using relative file paths

 reports directory

`d1-report.Rmd`

also contains all the prose explaining the data, results, graphs, discussion, and references

# References

Bryan J (2018) Excuse me, do you have a moment to talk about version control? *The American Statistician* 72(1), 20–27 (doi:10.1080/00031305.2017.1399928)

Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L and Teal TK (2017) Good enough practices in scientific computing. *PLoS Computational Biology* 13(6)  
<https://doi.org/10.1371/journal.pcbi.1005510>