

A Minimal Book Example

John Doe

2022-05-09

Contents

1	About	5
1.1	Usage	5
1.2	Render book	5
1.3	Preview book	6
2	Hello bookdown	7
2.1	A section	7
3	Cross-references	9
3.1	Chapters and sub-chapters	9
3.2	Captioned figures and tables	9
4	Methodology	13
4.1	Data	14
5	Footnotes and citations	21
5.1	Footnotes	21
5.2	Citations	21
6	Blocks	23
6.1	Equations	23
6.2	Theorems and proofs	23
6.3	Callout blocks	23

7	Sharing your book	25
7.1	Publishing	25
7.2	404 pages	25
7.3	Metadata for sharing	25

Chapter 1

About

This is a *sample* book written in **Markdown**. You can use anything that Pandoc’s Markdown supports; for example, a math equation $a^2 + b^2 = c^2$.

1.1 Usage

Each **bookdown** chapter is an .Rmd file, and each .Rmd file can contain one (and only one) chapter. A chapter *must* start with a first-level heading: **# A good chapter**, and can contain one (and only one) first-level heading.

Use second-level and higher headings within chapters like: **## A short section** or **### An even shorter section**.

The `index.Rmd` file is required, and is also your first book chapter. It will be the homepage when you render the book.

1.2 Render book

You can render the HTML version of this example book without changing anything:

1. Find the **Build** pane in the RStudio IDE, and
2. Click on **Build Book**, then select your output format, or select “All formats” if you’d like to use multiple formats from the same book source files.

Or build the book from the R console:

```
bookdown::render_book()
```

To render this example to PDF as a `bookdown::pdf_book`, you'll need to install XeLaTeX. You are recommended to install TinyTeX (which includes XeLaTeX): <https://yihui.org/tinytex/>.

1.3 Preview book

As you work, you may start a local server to live preview this HTML book. This preview will update as you edit the book when you save individual .Rmd files. You can start the server in a work session by using the RStudio add-in “Preview book”, or from the R console:

```
bookdown::serve_book()
```

Chapter 2

Hello bookdown

All chapters start with a first-level heading followed by your chapter title, like the line above. There should be only one first-level heading (#) per .Rmd file.

2.1 A section

All chapter sections start with a second-level (##) or higher heading followed by your section title, like the sections above and below here. You can have as many as you want within a chapter.

An unnumbered section

Chapters and sections are numbered by default. To un-number a heading, add a {.unnumbered} or the shorter {-} at the end of the heading, like in this section.

Chapter 3

Cross-references

Cross-references make it easier for your readers to find and link to elements in your book.

3.1 Chapters and sub-chapters

There are two steps to cross-reference any heading:

1. Label the heading: `# Hello world {#nice-label}`.
 - Leave the label off if you like the automated heading generated based on your heading title: for example, `# Hello world = # Hello world {#hello-world}`.
 - To label an un-numbered heading, use: `# Hello world {-#nice-label}` or `{# Hello world .unnumbered}`.
2. Next, reference the labeled heading anywhere in the text using `\@ref(nice-label)`; for example, please see Chapter 3.
 - If you prefer text as the link instead of a numbered reference use: any text you want can go here.

3.2 Captioned figures and tables

Figures and tables *with captions* can also be cross-referenced from elsewhere in your book using `\@ref(fig:chunk-label)` and `\@ref(tab:chunk-label)`, respectively.

See Figure 3.1.

```
par(mar = c(4, 4, .1, .1))  
plot(pressure, type = 'b', pch = 19)
```



Figure 3.1: Here is a nice figure!

Don't miss Table 3.1.

```
knitr::kable(  
  head(pressure, 10), caption = 'Here is a nice table!',  
  booktabs = TRUE  
)
```

Table 3.1: Here is a nice table!

temperature	pressure
0	0.0002
20	0.0012
40	0.0060
60	0.0300
80	0.0900
100	0.2700
120	0.7500
140	1.8500
160	4.2000
180	8.8000

Chapter 4

Methodology

4.0.1 Variables

4.0.1.1 Categories that would be useful (things to predict?)

Owner-occupied? Investor-owned? Vacant? Demolition in the past year (no construction since) Construction in the past year

4.0.1.2 factor analysis Variables

The variables made it into the initial factor analysis were:

- Accessibility
 - Distance to transit (use number of transit stops within 1/2 mile walkshed)
 - Share of old/new homes (use average age of homes within 1/2 mile walkshed)
 - Transit frequency (use transit stops per hour within 1/2 mile walkshed)
- Affordability
 - Average Condition of homes in half-mile walkshed
 - Median rent of block-groups with centroids within 1/2 mile walkshed.
 - Median income of block groups with centroids within 1/2 mile walkshed.
 - Median ownership cost of block groups with centroids within 1/2 mile walkshed.

- Close
 -
- Diverse buildings
 - Entropy of housing types (apartment, townhomes, etc) within 1/2 mile walkshed
- Other
 - Standard deviation of building age within 1/2 mile walkshed

4.1 Data

We obtained data on property addresses, land uses, assessed values (for both land and buildings), and building condition from the property assessment data [Allegheny County Office of Property Assessments, 2022], which includes information on 582,116 properties in Allegheny County.

We also obtained latitude and longitude coordinates for each property from a geocoder file provided by Western Pennsylvania Regional Data Center [2021]. Over 99.5 percent of properties included in the assessment dataset are included in the geocoder file. Properties without geocoded locations are excluded from our analysis, leaving a total of 579,473 properties.

Potential development sites were identified as those

1. classified as “residential” (indicating residential properties with one to four housing units) or “commercial” (which includes mixed-use developments and residential properties with more than four housing units) and
2. with a land use description in one of 59 possible categories. The most common of these are listed Table 4.1¹. One site (3008 Phillip Dr in Clair-

¹The land use descriptions that were classified as potential development sites but are not listed in Table 4.1, which combine to represent less than one percent of all sites are “RIGHTOF WAY - RESIDENTIAL”, “CONDOMINIUM UNIT”, “DWG USED AS OFFICE”, “APART:20-39 UNITS”, “CONDO GARAGE UNITS”, “COMMON AREA”, “CONDO DEVELOPMENTAL LAND”, “CONDEMNED/BOARDED-UP”, “CONDOMINIUM OFFICE BUILDING”, “INDEPENDENT LIVING (SENIORS)”, “DWG USED AS RETAIL”, “OTHER COMMERCIAL”, “MOBILE HOMES/TRAILER PKs”, “RIGHT OF WAY - COMMERCIAL”, “GROUP HOME”, “TOTAL/MAJOR FIRE DAMAGE - COMM”, “OTHER COMMERCIAL HOUSING”, “TOTAL/MAJOR FIRE DAMAGE”, “COMM APRTM CONDOS 5-19 UNITS”, “MUNICIPAL URBAN RENEWAL”, “COMMERCIAL LAND”, “CAMPGROUNDS”, “COMMON AREA OR GREENBELT”, “CHARITABLE EXEMPTION/HOS/HOMES”, “INCOME PRODUCING PARKING LOT”, “DWG APT CONVERSION”, “>10 ACRES VACANT”, “MINOR FIRE DAMAGE”, “COMM APRTM CONDOS 20-39 UNITS”, “COMMERCIAL/UTILITY”, “H.O.A RECREATIONS AREA”, “COMM APRTM CONDOS 40+ UNITS”, “MINOR FIRE DAMAGE - COMM”, “OTHER”, “OTHER RESIDENTIAL STRUCTURE”, “OWNED BY METRO HOUSING AU”, “RESIDENTIAL VACANT LAND”, “HUD PROJ #221”, and “VACANT LAND 0-9 ACRES”

Table 4.1: Most common land uses categorized as potential sites

USEDESC	Number of potential sites	Percent of potential sites	Cumulative
SINGLE FAMILY	371,064	69.8	
VACANT LAND	63,603	12.0	
TWO FAMILY	17,330	3.3	
CONDOMINIUM	16,683	3.1	
TOWNHOUSE	14,953	2.8	
ROWHOUSE	11,129	2.1	
VACANT COMMERCIAL LAND	6,103	1.1	
THREE FAMILY	3,977	0.7	
RES AUX BUILDING (NO HOUSE)	3,635	0.7	
RETL/APT'S OVER	3,366	0.6	
COMM AUX BUILDING	3,040	0.6	
APART: 5-19 UNITS	2,800	0.5	
MOBILE HOME (IN PARK)	2,563	0.5	
FOUR FAMILY	2,064	0.4	
BUILDERS LOT	1,714	0.3	
CONDOMINIUM COMMON PROPERTY	1,307	0.2	
PARKING GARAGE/LOTS	935	0.2	
OFFICE/APARTMENTS OVER	860	0.2	
MOBILE HOME	676	0.1	
APART:40+ UNITS	545	0.1	

ton) is missing a land use description in the assessment data. We checked this address on Zillow to determine that this is a single-family home and classified it as such in our data.

The above criteria yield 531,811 potential sites. Potential building sites were further filtered to exclude those with missing data on the most recent sale (6,574 sites, or about one percent of all sites).² for a total of 526,237 potential sites.

²Four sites had sales prices listed that were unreasonably high. 3039 Liberty Avenue in Pittsburgh is listed as having sold for \$511,945,000 on August 30, 2021. Zillow lists this property as having sold on that date for \$511,945 (https://www.zillow.com/homedetails/3039-W-Liberty-Ave-Pittsburgh-PA-15216/2070262638_zpid/, accessed 5/4/2022), so the value was corrected for what appears to have been a typo. 220 Hyeholde Dr in Coraopolis is listed as having sold for \$28,100,000 in 1967. This may also be a typo, and it also does not seem to be the most recent sale. Zillow lists this home as having sold for \$350,000 in 2004 (https://www.zillow.com/homes/220-hyeholde-dr,-Coraopolis,-PA_rb/11552817_zpid/, accessed 5/4/2022), so the data was corrected to add that as the most recent sale. Two other sites were identified as having unreasonably high sales values: 1339 Arlington Avenue in Pittsburgh is a three-bedroom single-family home that is listed as having sold for \$57,010,813 in 1976 and a 0.06-acre vacant lot with tax ID 0165G00270000000 is listed as having sold for \$24,920,232 in 1936. The sales data for these sites were treated as missing.

The focus of this analysis is on potential development sites rather than on properties. Some properties in the assessor dataset are condominiums where multiple properties share a single parcel of land. We aggregated these to the site level by identifying all properties with an assessed building value greater than zero, a land value of zero, and a land use description that did not indicate the land was vacant. If multiple such properties share an address, we classified all properties at that address as a condominium and aggregated them to the parcel level. This led to a final sample of 518,316 sites.

4.1.1 Aspatial data

Three variables (total assessed fair market value, assessed fair market value of the building, and lot area) were taken directly from the assessment data for use in our analysis. We calculated two additional variables from the assessment data: the inflation-adjusted sales price and the average number of years between sales. The average of:

1. The number of years between the most recent sale and the publication data of the assessment data (May 1, 2022);
2. The number of years between the most recent sale and the second-most recent sale³; and
3. The number of years between the second-most recent sale and the third-most recent sale, if a third-most recent sale is listed.

To aggregate properties identified as condominiums to the site level, we summed the total values for lot area, assessed land value, assessed building value, and inflation-adjusted sale price and averaged values for the average number of years between sales.

```
write_csv(sites,
          here("data",
               "sites.csv"))
```

4.1.2 Accessibility data

Accessibility was calculated from each of the 518,316 sites in our sample to each of several location types described below.

We used land use codes from the county assessor parcel data to identify *destination parcels* that residents might value access to. The most common land use codes of identified destination parcels are listed in 4.2.

³If no date is listed for the second-most recent sale, we used January 1, 1950. The data user guide notes that, although deeds have been recorded in Allegheny County since 1788 (and the

Table 4.2: Land uses identified as potential destinations

USEDESC	Number of identified destinations	Percent of identified destinations
MUNICIPAL GOVERNMENT	10,378	29
CHURCHES, PUBLIC WORSHIP	1,948	5
COMMERCIAL GARAGE	1,735	5
OFFICE - 1-2 STORIES	1,649	4
SMALL DETACHED RET(UNDER 10000)	1,645	4
OFFICE/WAREHOUSE	1,385	3
COUNTY GOVERNMENT	1,287	3
WAREHOUSE	1,255	3
OWNED BY BOARD OF EDUCATION	1,086	3
TOWNSHIP GOVERNMENT	853	2
LIVESTOCK O/T D & P-CAUV	805	2
LIGHT MANUFACTURING	799	2
PUBLIC PARK	710	2
RESTAURANT, CAFET AND/OR BAR	697	2
GENERAL FARM	607	1
OWNED BY COLLEGE/UNIV/ACADEMY	458	1
MEDICAL CLINICS/OFFICES	445	1
RETL/OFF OVER	442	1
OFFICE-ELEVATOR -3 + STORIES	413	1
LODGE HALL/AMUSEMENT PARK	386	1
AUTO SALES & SERVICE	364	1
RETL/STOR OVER	344	0
CEMETERY/MONUMENTS	340	0
STATE GOVERNMENT	331	0
CONVENIENCE STORE/GAS	304	0
BANK	298	0
NEIGH SHOP CENTER	297	0
BARS	240	0
FUNERAL HOMES	194	0
DAYCARE/PRIVATE SCHOOL	191	0
MINI WAREHOUSE	182	0
MEDIUM MANUFACTURING	172	0
NURSING HOME/PRIVATE HOS	171	0
DRIVE IN REST OR FOOD SERVICE	170	0
OFFICE-WALKUP -3 + STORIES	164	0
AUTO SERV STATION	137	0
COMMUNITY SHOPPING CENTER	137	0
OTHER RETAIL STRUCTURES	135	0
WAREHOUSE/MULTI-TENANT	126	0
BOWLING ALLEYS/REC FACILITY	120	0
HOTELS	113	0
DISCOUNT STORE	110	0
CAR WASH	103	0
FAST FOOD/DRIVE THRU WINDOW	85	0
PHARMACY (CHAIN)	85	0
SMALL SHOP	81	0
SUPERMARKETS	75	0
OFFICE/STORAGE OVER	66	0

We identified *job locations* based on data from a Longitudinal Employer-Household Dynamics (LEHD) dataset published by the United States Census Bureau [United States Census Bureau, 2021]. The LEHD dataset provides the total number of jobs in each census block in the United States, based on employment tax records. The location of each job was defined as the centroid of the block in which it was located. We downloaded job location data for Pennsylvania and filtered it to include locations in the Pittsburgh metropolitan area (Allegheny, Armstrong, Beaver, Butler, Fayette, Washington, and Westmoreland counties).

In addition to calculating the accessibility to jobs of all categories, we also calculated accessibility to several subsets of jobs. We disaggregated jobs by earnings, reasoning that the usefulness of a job might vary depending on how well it matches a workers skills or wage expectations. *High-paying job locations* are a subset of job locations where the worker earns more than \$3333 per month. *Low-paying job locations* are those where the worker earns \$1250 per month or less.

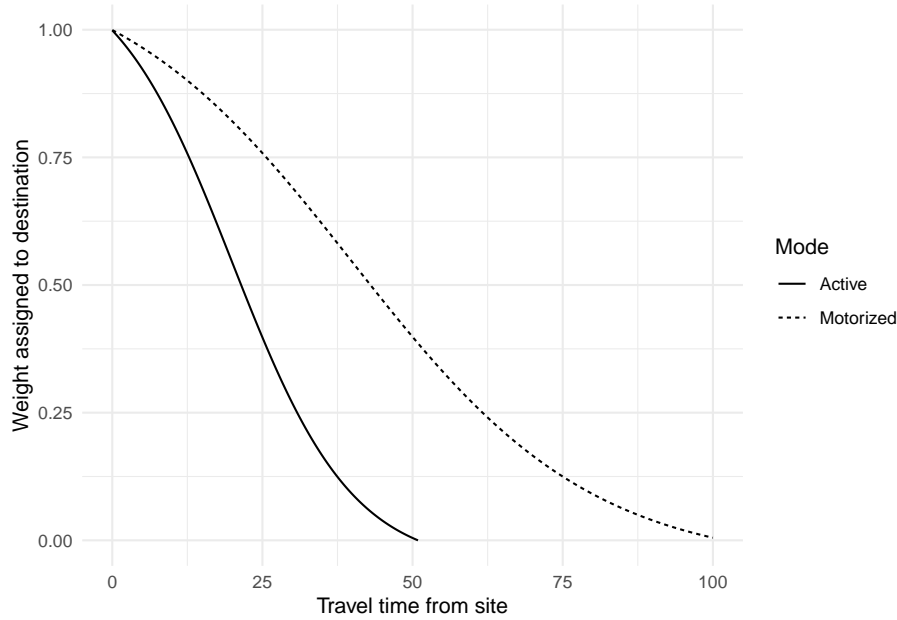
We also disaggregated jobs based on employment industry, based on the North American Industry Classification System (NAICS), reasoning that the presence of jobs particular industries might represent a shopping or recreation destination. *Retail job locations* are a subset of job locations in NAICS sector 44-45 (retail trade); *Entertainment job locations* are those in NAICS sector 71 (arts, entertainment, and recreation); and *Hospitality job locations* are those in NAICS sector 72 (accommodation and food services).

Finally we identified three location types that correspond with common non-work trips: schools, grocery stores, and parks. *Grocery store locations* were identified as vendors participating in the Supplemental Nutrition Program for Women, Infants, and Children (WIC). WIC vendor locations and *school locations* were obtained from the and location data came from the Allegheny County GIS portal [Allegheny County Office of Information Technology, 2018, 2020]. *Park locations* were taken from the Pennsylvania Geospatial Data Clearinghouse [Pennsylvania Department of Conservation and Natural Resources, 2015]. Data were downloaded for Pennsylvania and filtered to Allegheny county.

We used the `r5r` package in the R programming language [Pereira et al., 2021] to calculate accessibility was calculated for each destination type described above for each of four transportation modes (walking, cycling, driving, and transit). The `r5r` package calculates accessibility as the weighted total number of destinations reachable by a given mode, where destinations are weighted according to a decay function, such that destinations that can be reached within less time are assigned greater weight. We used a logistic decay function, as illustrated in ???. For motorized modes, the decay function had a mean (inflection) of 40

earliest sale listed in the data is from November 24, 1806), early sales may not be included in the electronic system and may default to a 1950 sale date [Western Pennsylvania Regional Data Center, 2017]. This effects about one third of parcels in our dataset.

minutes and a standard deviation of 10 minutes. For non-motorized modes, the decay function had a mean of 20 minutes and a standard deviation of 5 minutes.



Calculating accessibility metrics for a combination of four transportation modes and ten destination types yields 40 different accessibility variables.

4.1.3 Density data

For each site, we calculated the density of four different (dis)amenities in the immediate vicinity (defined as the area within a one-kilometer radius): homes, pedestrian paths (including sidewalks and all routes open to pedestrians), roads that are not open to pedestrians (such as freeways), and land uses identified in county assessor data as one of several types that we categorized as disamenities. The land use codes we used to identify disamenities are listed in

- Housing density (use number of homes within 1/2 mile circle)
- Ped network density (use mileage of ped network within 1/2 mile circle)
- Number of disamenities in a 1/2 mile circle
- Mileage of non-ped roads in 1/2 mile circle

Chapter 5

Footnotes and citations

5.1 Footnotes

Footnotes are put inside the square brackets after a caret `^[]`. Like this one ¹.

5.2 Citations

Reference items in your bibliography file(s) using `@key`.

For example, we are using the **bookdown** package [Xie, 2022] (check out the last code chunk in `index.Rmd` to see how this citation key was added) in this sample book, which was built on top of R Markdown and **knitr** [?] (this citation was added manually in an external file `book.bib`). Note that the `.bib` files need to be listed in the `index.Rmd` with the YAML `bibliography` key.

The RStudio Visual Markdown Editor can also make it easier to insert citations: <https://rstudio.github.io/visual-markdown-editing/#/citations>

¹This is a footnote.

Chapter 6

Blocks

6.1 Equations

Here is an equation.

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (6.1)$$

You may refer to using `\@ref{eq:binom}`, like see Equation (6.1).

6.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref{thm:tri}`, for example, check out this smart theorem 6.1.

Theorem 6.1. *For a right triangle, if c denotes the length of the hypotenuse and a and b denote the lengths of the **other** two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here <https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html>.

6.3 Callout blocks

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: <https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html>

Chapter 7

Sharing your book

7.1 Publishing

HTML books can be published online, see: <https://bookdown.org/yihui/bookdown/publishing.html>

7.2 404 pages

By default, users will be directed to a 404 page if they try to access a webpage that cannot be found. If you'd like to customize your 404 page instead of using the default, you may add either a `_404.Rmd` or `_404.md` file to your project root and use code and/or Markdown syntax.

7.3 Metadata for sharing

Bookdown HTML books will provide HTML metadata for social sharing on platforms like Twitter, Facebook, and LinkedIn, using information you provide in the `index.Rmd` YAML. To setup, set the `url` for your book and the path to your `cover-image` file. Your book's `title` and `description` are also used.

This `gitbook` uses the same social sharing data across all chapters in your book—all links shared will look the same.

Specify your book's source repository on GitHub using the `edit` key under the configuration options in the `_output.yml` file, which allows users to suggest an edit by linking to a chapter's source file.

Read more about the features of this output format here:

<https://pkgs.rstudio.com/bookdown/reference/gitbook.html>

Or use:

```
?bookdown::gitbook
```

Bibliography

- Allegheny County Office of Information Technology. Allegheny County WIC Vendor Locations, April 2018. URL https://openac-alcogis.opendata.arcgis.com/datasets/ab9ec54e46d8403db31cff6bdc890aff_0/explore?location=40.458725%2C-79.972398%2C10.20. type: dataset.
- Allegheny County Office of Information Technology. Allegheny County Public Schools / Local Education Agency (LEAs) Locations, December 2020. URL <https://openac-alcogis.opendata.arcgis.com/datasets/AICoGIS::allegheny-county-public-schools-local-education-agency-leas-locations/about>. type: dataset.
- Allegheny County Office of Property Assessments. Allegheny County Property Assessments, May 2022. URL <https://data.wprdc.org/dataset/2b3df818-601e-4f06-b150-643557229491>. type: dataset.
- Pennsylvania Department of Conservation and Natural Resources. Pennsylvania Local Parks Access Points, November 2015. URL <https://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=308>.
- Rafael H. M. Pereira, Marcus Saraiva, Daniel Herszenhut, Carlos Kaue Vieira Braga, and Matthew Wigginton Conway. r5r: Rapid Realistic Routing on Multimodal Transport Networks with R⁵ in R. *Findings*, page 21262, March 2021. doi: 10.32866/001c.21262. URL <https://findingspress.org/article/21262-r5r-rapid-realistic-routing-on-multimodal-transport-networks-with-r-5-in-r>. Publisher: Findings Press.
- United States Census Bureau. LEHD Origin-Destination Employment Statistics (LODES), October 2021. URL <https://lehd.ces.census.gov/data/#lodes>. Type: dataset.
- Western Pennsylvania Regional Data Center. Allegheny County Property Assessment Data User Guide. Technical report, Pittsburgh, PA, April 2017. URL <https://data.wprdc.org/dataset/2b3df818-601e-4f06-b150-643557229491/resource/cc4bafd2-25b6-41d7-83aa-d16bc211b020/download/alleghenycountypropertyassessmentdatauserguide-4.pdf>.

Western Pennsylvania Regional Data Center. Geocoders, February 2021. URL <https://data.wprdc.org/dataset/6bb2a968-761d-48cf-ac5b-c1fc80b4fe6a>.

Yihui Xie. *bookdown: Authoring Books and Technical Documents with R Markdown*, 2022. URL <https://CRAN.R-project.org/package=bookdown>. R package version 0.25.