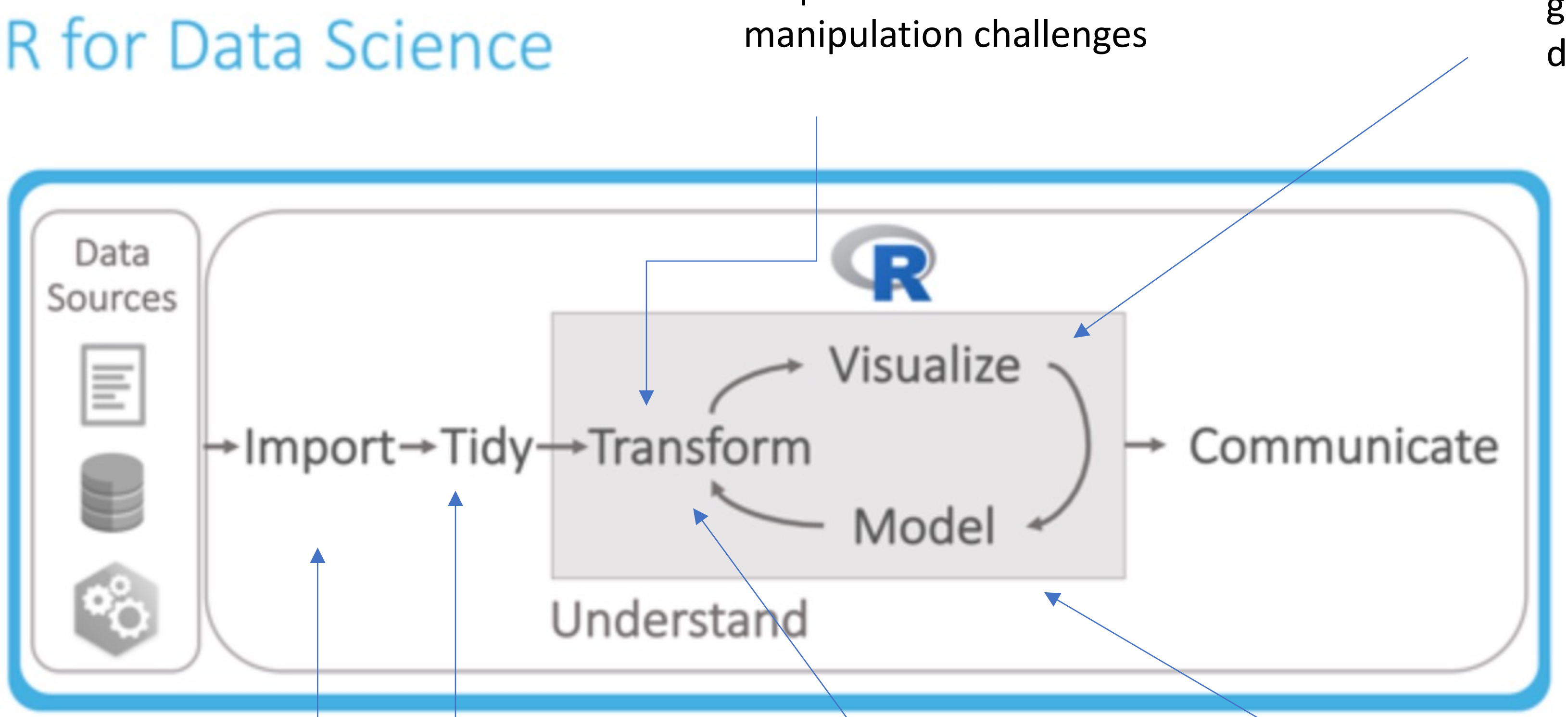


[Dplyr](#)
Helps solve the most common data manipulation challenges

[Ggplot2](#)
Implements the grammar of graphics, a coherent system for describing and building graphs

[R markdown](#)
- Turn analyses into high quality documents, reports, presentations and dashboards

[Knitr](#)
- Enables integration of R code into LaTeX, LyX, HTML, Markdown, AsciiDoc, and reStructuredText documents.

[Readr](#)
Fast and friendly way to read rectangular data

[Tidyr](#)
Helps create tidy data

[Dplyr](#)
”A grammar of data manipulation”

[Broom](#)
- Summarizes statistical model objects in tidy tibbles

Readr

- read_delim() reads files with any delimiter
- read_csv() reads a comma delimited file with period decimal marks
- read_csv2() reads a comma delimited file with comma decimal marks
- read_tsv() reads a tab delimited file
- read_fwf() reads a fixed width file

Helpful links

https://r4ds.had.co.nz/data-import.html

Back to directory

Tidyr

- pivot_longer() “lengthens” data by collapsing several columns into two
- pivot_wider() “widens” data by expanding two columns into several
- drop_na() drop rows containing “NA”
- fill() fill in “NA”
- replace_na() replaces “NA” with a specified value in selected columns
- gather() gathers columns into rows

Helpful links

https://tidyr.tidyverse.org/

https://r4ds.had.co.nz/tidy-data.html

Back to directory

Dyplr

- select() certain columns of data
- filter() your data to select specific rows
- arrange() the rows of your data into an order
- mutate() your data frame to contain new columns
- summarize() chunks of you data in some way
- group_by() splits data into groups
- right_join(), left_join(), semi_join(), anti_join() combines data sets

Helpful links

https://r4ds.had.co.nz/transform.html

https://dplyr.tidyverse.org/

Back to directory

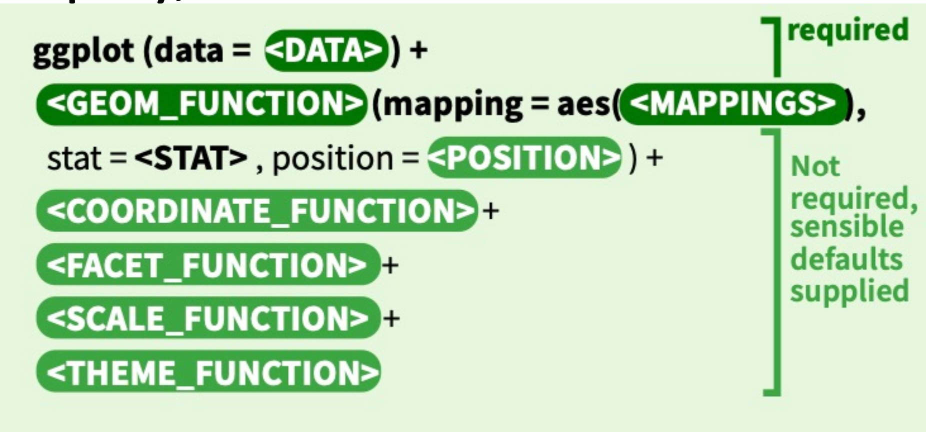
Ggplot2

Back to directory

- <geom_function>   geom represent what you see in the plot: points, lines, polygons, etc.
- <coordinate_function> coord describes how data coordinates are mapped to the plane of the graphic
- <facet_function> facet specifies how to break up and display subsets of data as small multiples
- <scale_function> scales map values in the data space to values in the aesthetic space
- <theme_function> theme controls the finer points of display, like the font size and background color

Helpful links

https://r4ds.had.co.nz/data-visualisation.html



```
ggplot (data = <DATA>) +
<GEOM_FUNCTION>(mapping = aes(<MAPPINGS>),
 stat = <STAT> , position = <POSITION> ) +
<COORDINATE_FUNCTION>+
<FACET_FUNCTION>+
<SCALE_FUNCTION>+
<THEME_FUNCTION>
```

required

Not required, sensible defaults supplied

Broom

- tidy() summarizes information about model components
- glance() reports information about the entire model
- augment() adds information about observations to a dataset

Helpful links

https://broom.tidymodels.org

Back to directory

DS workflow cheat sheet

We created a cheat sheet for data science students like ourselves. We decided on this idea because we felt a simple reference page that explains the workflow of a typical project would be handy. The first page of this cheat sheet shows the flow of a typical data science project. Each step of the workflow includes the libraries used and a brief description of what each library does. To access more details about a specific library (most common/useful functions and resource links), click on a library name. Creating this cheat sheet gave us a better understanding of the workflow for a data science project in r and clarified the purpose for each of the libraries (most of which we have been using in class). If given more time, aesthetic improvements could have been made for our cheat sheet.