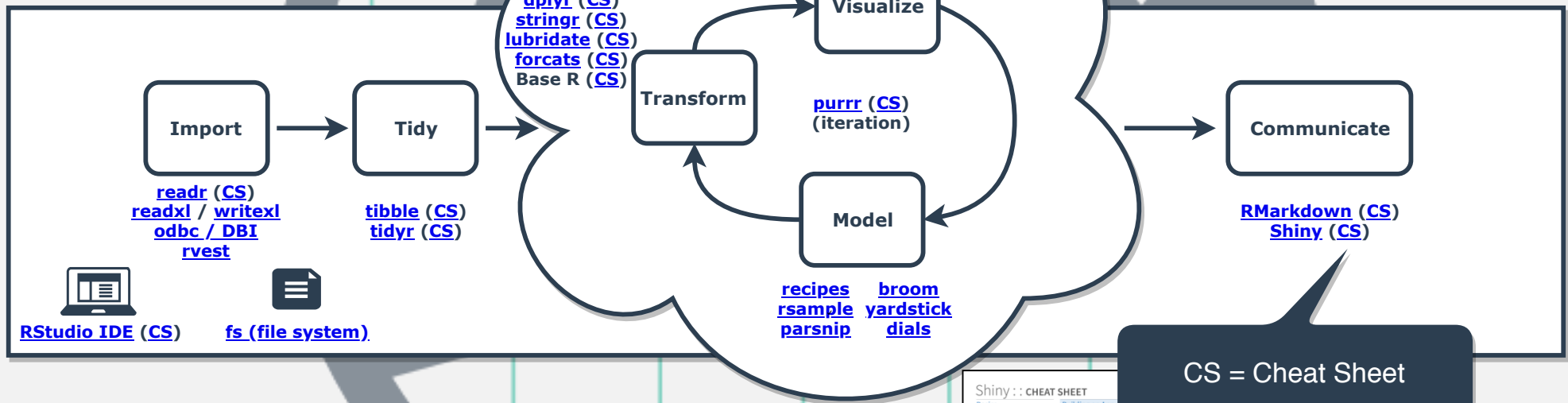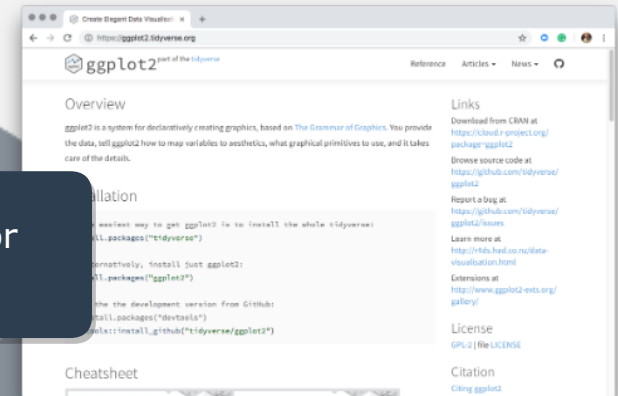# Data Science with R Workflow

The Data Science With R Workflow is available in the book: **R For Data Science**. If you want to learn R and this workflow *for business analysis*, take the **R For Business Analysis (DS4B 101-R) course** through Business Science University.
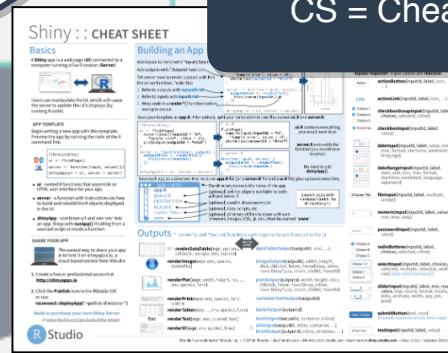
Click the links for Documentation

ggplot2 (CS)

Visualize

dplyr (CS)
stringr (CS)
lubridate (CS)
forcats (CS)
Base R (CS)

Transform

purrr (CS)
(iteration)

Import → **Tidy** →

Communicate

Model

readr (CS)
readxl / writexl
odbc / DBI
rvest

tibble (CS)
tidyr (CS)

RMarkdown (CS)
Shiny (CS)

recipes   **broom**
rsample **yardstick**
parsnip   **dials**

RStudio IDE (CS)   **fs (file system)**

CS = Cheat Sheet

Important Resources

- **R For Data Science Book:** http://r4ds.had.co.nz/
- **Rmarkdown Book:** https://bookdown.org/yihui/rmarkdown/
- **Data Visualization Book**: https://rkabacoff.github.io/datavis/
- **More Cheatsheets:** https://www.rstudio.com/resources/cheatsheets/
- **tidyverse packages:** https://www.tidyverse.org/
- **Connecting to databases:** https://db.rstudio.com/
- **RMarkdown website:** https://rmarkdown.rstudio.com/
- **Shiny web applications website:** http://shiny.rstudio.com/
- **Jenny Bryan's purrr tutorial:** https://jennybryan.org/

"Data Science Courses for Business"

Business Science University
university.business-science.io

version: 1.2

Data Science with R

Special Topics

Text Analysis & NLP

- [Text Mining with R (Book)](#): [tidytext](#)
- NLP:
    - [H2O word2vec](#): Word embeddings
    - [text2vec](#): fast vectorization, topic modeling
    - [udpipe](#): [UDPipe](#) C++ lib in R

Machine Learning

- Multi-Threaded/Scalable/Production ML:
    - [H2O](#) ([CS](#))
    - Extreme Gradient Boosting: [xgboost](#)
    - R + Spark: [sparklyr](#) ([CS](#))
    - Sparkling Water (Spark + H2O): [rsparkling](#)
- ML (Tidy): [parsnip](#)
- ML: [caret](#) ([CS](#))

Deep Learning

- [R Interface to TensorFlow Homepage](#):
    - [Keras](#) ([CS](#))
    - [TF Estimators](#)
    - [TensorFlow (Core)](#)

Time Series Analysis

- Time-aware tibbles: [tibbletime](#) & [tsibble](#)
- Convert between classes: [timetk](#) & [tsbox](#)
- Time Series Index Summary: [timetk](#)
- Generating Future Series: [timetk](#)

Forecasting

- ARIMA, ETS, etc: [forecast](#) & [fable](#)
- Tidy, glance, augment for forecast models: [sweep](#)
- Converting forecast prediction to tibble: [sweep](#)

Anomaly Detection

- Identify anomalies: [anomalize](#)

Network Analysis

- Network Data Transformations (Tidy): [tidygraph](#)
- Network Data Transformations: [igraph](#)

Network Viz

- Static:
    - [ggraph](#) - Graph plotting utilities for ggplot2
- Interactive (JavaScript):
    - [networkD3](#) - D3 Networks in R
    - [plotly](#) ([CS](#)) - plotly.js (network graphs) in R

Speed & Scale

- Fastest Single-Node Speed: [data.table](#) ([CS](#))
- Distributed Cluster (Spark): [sparklyr](#) ([CS](#))
- Parallel Processing: [furrr](#)

Interoperability

- Java: [rJava](#)
- Python: [reticulate](#)
- C++: [Rcpp](#)

Financial Analysis

- Getting financial data: [tidyquant](#) & [quantmod](#)
- Quantitative Analysis: [tidyquant](#) & [xts](#)/[TTR](#)
- Portfolio Analysis: [tidyquant](#) & [PerformanceAnalytics](#)

Financial & Time Viz

- Static:
    - [tidyquant](#) - Financial ggplot2 geoms
- Interactive:
    - [highcharter](#) - highchart.js in R
    - [dygraphs](#) - xts plotting
    - [plotly](#) ([CS](#)) - plotly.js (financial) in R

Geospatial Analysis

- Geocoding (getting lat/long, bboxes, & sf's):
    - [ggmap](#) - Google API (requires key)
    - [osmdata](#) - OpenStreet Overpass API
    - [tmaptools](#) - OpenStreet Nominatum API
- Simple Features (sf objects): [sf](#) ([CS](#)) (tidy)
- Spatial Objects (sp objects): [sp](#) (non-tidy)

Geospatial Viz

- Static:
    - [ggmap](#) - Google API (requires key)
    - [osmplotr](#) - Impressive Maps via OSM
    - [tmap](#) - Thematic Maps
    - [cartography](#) ([CS](#)) - Thematic Maps
- Interactive (JavaScript):
    - [leaflet](#) ([CS](#)) - leaflet.js in R
    - [plotly](#) ([CS](#)) - plotly.js (maps) in R

Miscellaneous Tools

- **Interactive Plotting:** [htmlwidgets for R](#)
- **Building R Packages:** [R packages Book](#)
    - Pkg Development Tools: [devtools](#) ([CS](#))
    - R Templates: [usethis](#)
    - Build Web Doc's: [pkgdown](#)
- **Advanced Concepts ([Advanced R Book](#))**
    - [rlang & Tidy Evaluation](#) ([CS](#))
- **Making Blogs & Books:**
    - [blogdown](#), [bookdown](#)
- **Posting Code (GitHub, Stack Overflow):** [reprex](#)

"Data Science Courses for Business"

Business Science University
[university.business-science.io](#)