

Rotman

INTRO TO R – TIDYMODELS

R Workshop – Part 4 (1)

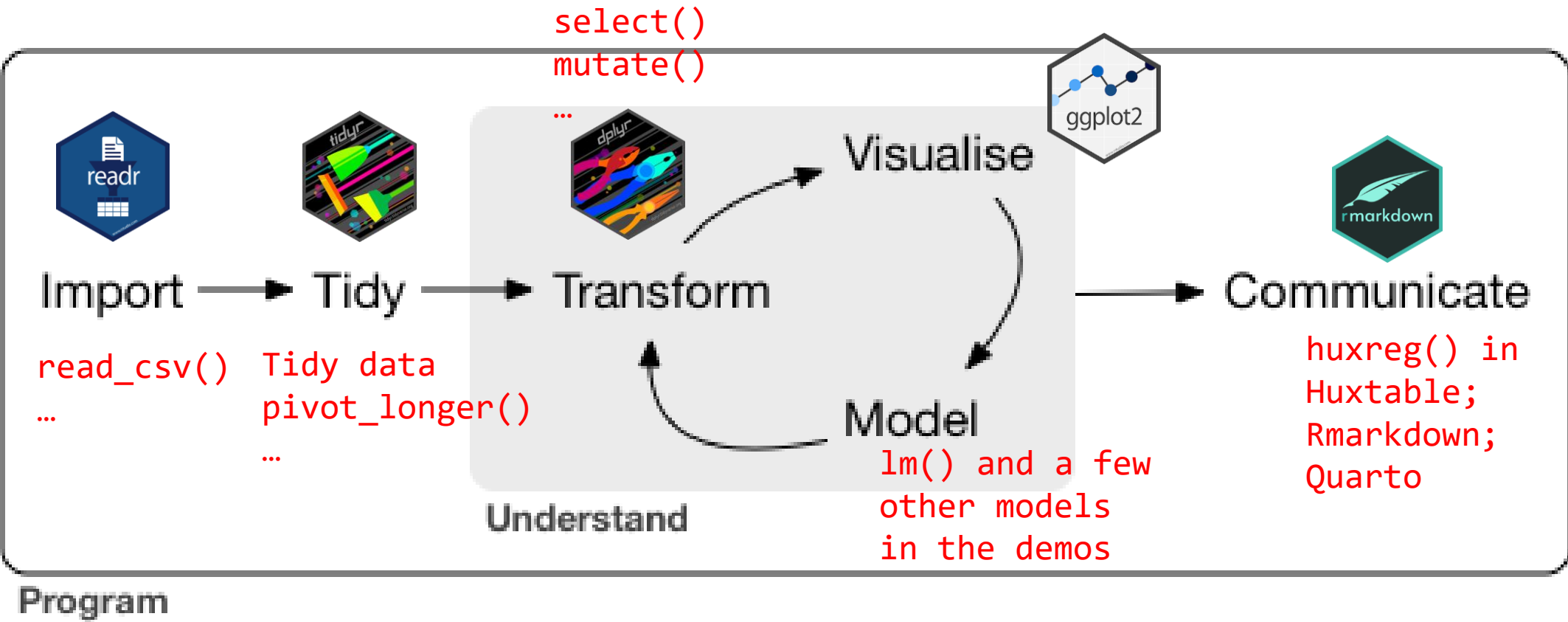
April 4, 2024 Prepared by Jay Cao / TDMDAL

Website: <https://tdmdal.github.io/r-workshop-202324-winter/>



Rotman School of Management
UNIVERSITY OF TORONTO

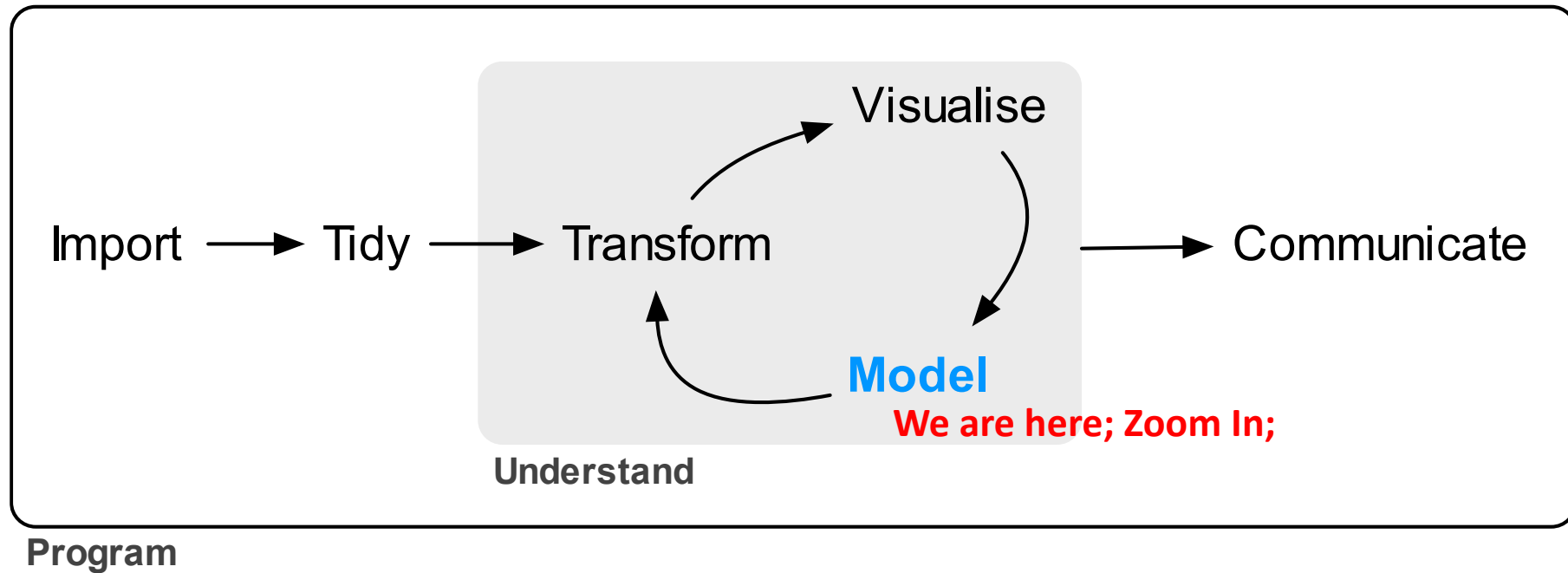
Revisit The Data Analysis Workflow



<https://r4ds.had.co.nz/introduction.html>

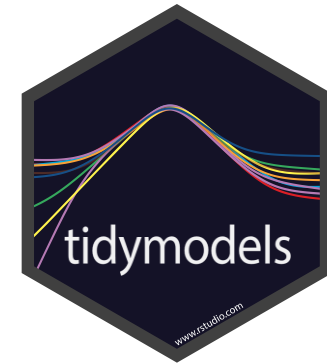
Model Wrangling?

- Recall the data analysis workflow



Tidymodels Framework

- What is it
 - *“a collection of packages for modeling and machine learning using tidyverse principles.”*
 - It is **NOT** a collection of statistical or ML models
- How to think about it
 - tidy data : data :: tidy models : modeling
 - tidyverse : data wrangling :: tidymodels : model wrangling



Model Wrangling Workflow ?



Data resampling



e.g., split data into train, validation, and test set.



Feature engineering



e.g., create dummy variables; normalize predictors.



Manage the steps as a workflow

Provide same interface for models from different packages



Model fitting



Model tuning



Facilitate hyperparameter tuning



Model evaluation



Manage model evaluation metrics using tidy data principles

Notebook Example

- A quick walkthrough of most tidymodels packages
- using [German credit data](#)

Learning Resources

- Package [website](#) has many learning resources
- A good free book: [Tidy Modeling with R](#)
- A Datacamp course: [Modeling with tidymodels in R](#)
 - Not free
 - I will read the free book instead of paying for the course (my opinion only)

