

Let's Git'R Marked Down

- Streamline Model Development Using R Markdown

LIZZY HUANG

DC R CONFERENCE

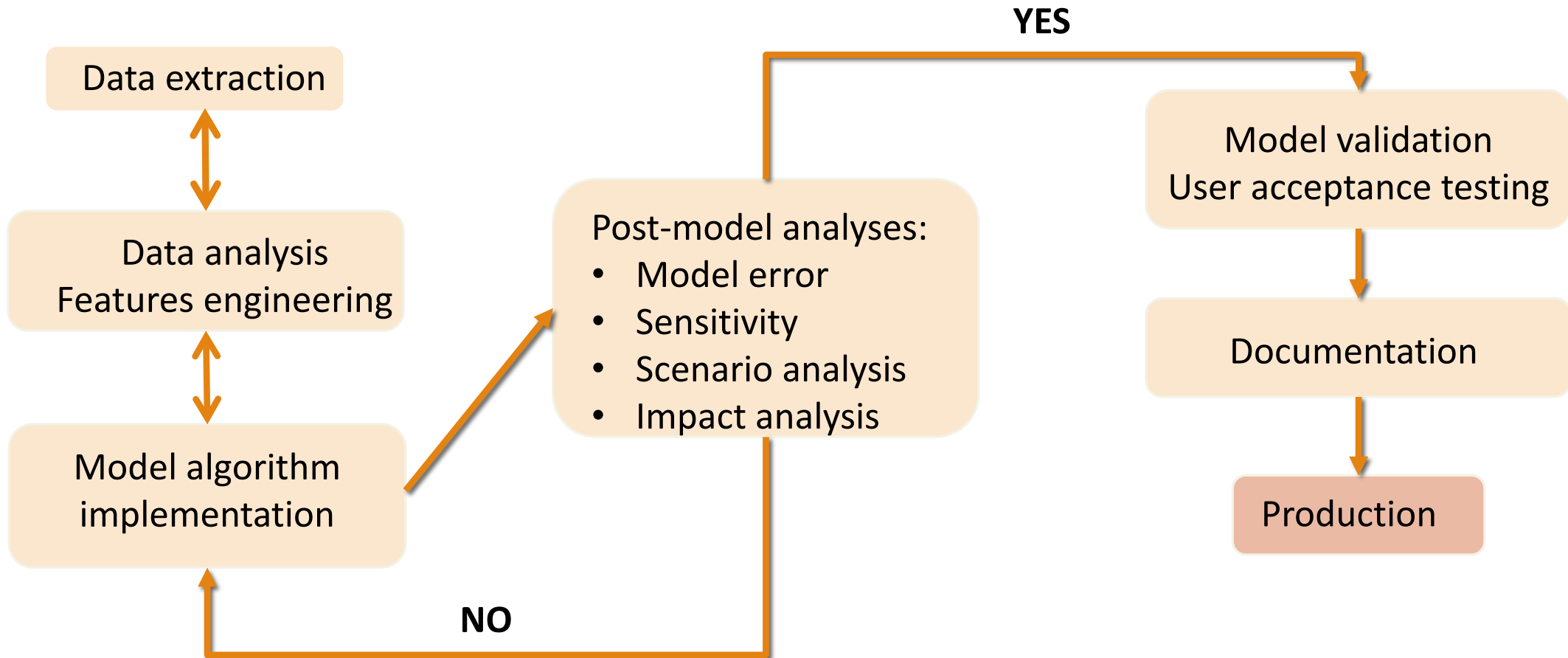
NOVEMBER 8

A solid orange horizontal bar at the bottom of the slide.

Outline

- ❑ General model development workflow
- ❑ Current development process and its challenges
- ❑ Improvement incorporating R Markdown and version control
- ❑ Demo

General Model Development Workflow



Currently Used Tools

Programming tools

- SAS
- SQL
- R
- Python, etc.

Reporting and visualization tools

- Excel
- Word
- Internal GUI applications

Current Process

SAS/SQL:

- Data extraction
- Model estimation
- Model validation
- Visualization of model results

R/Python:

- Data analysis
- Features generation
- Implementation
- Visualization for development purposes

Excel:

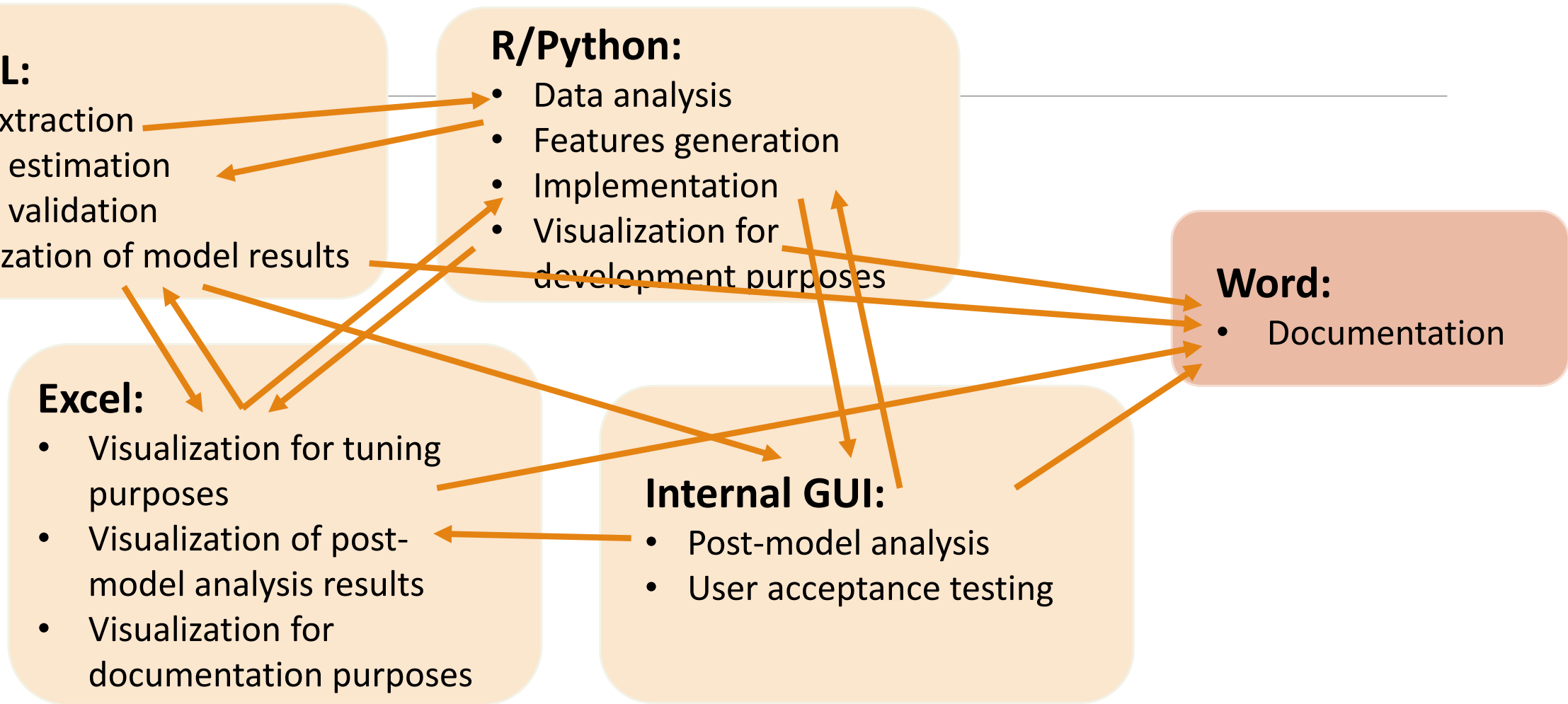
- Visualization for tuning purposes
- Visualization of post-model analysis results
- Visualization for documentation purposes

Internal GUI:

- Post-model analysis
- User acceptance testing

Word:

- Documentation



Challenges

- ❑ Difficult to keep track of code and data changes.
- ❑ Less control of data quality.
- ❑ Unable to standardize estimation algorithms for each product.
- ❑ Impossible to automate process.
- ❑ Inconvenient for collaboration.
- ❑ Hard to reproduce results.
- ❑ Longer time for documentation.

What We Need



Programming platform for
standardized code



Interactive platform for
graphic and text



Centralized platform for
reproducibility



Shared platform with
version control for
collaboration

Ideal Approach

Shared under version control

Interactive and centralized platform

- Connect to database

- Implement code in different languages

- Integrate with graphics and text

- Hook up with internal GUI API

- Documentation

Solution

- ❑ R Markdown for interactive and centralized platform

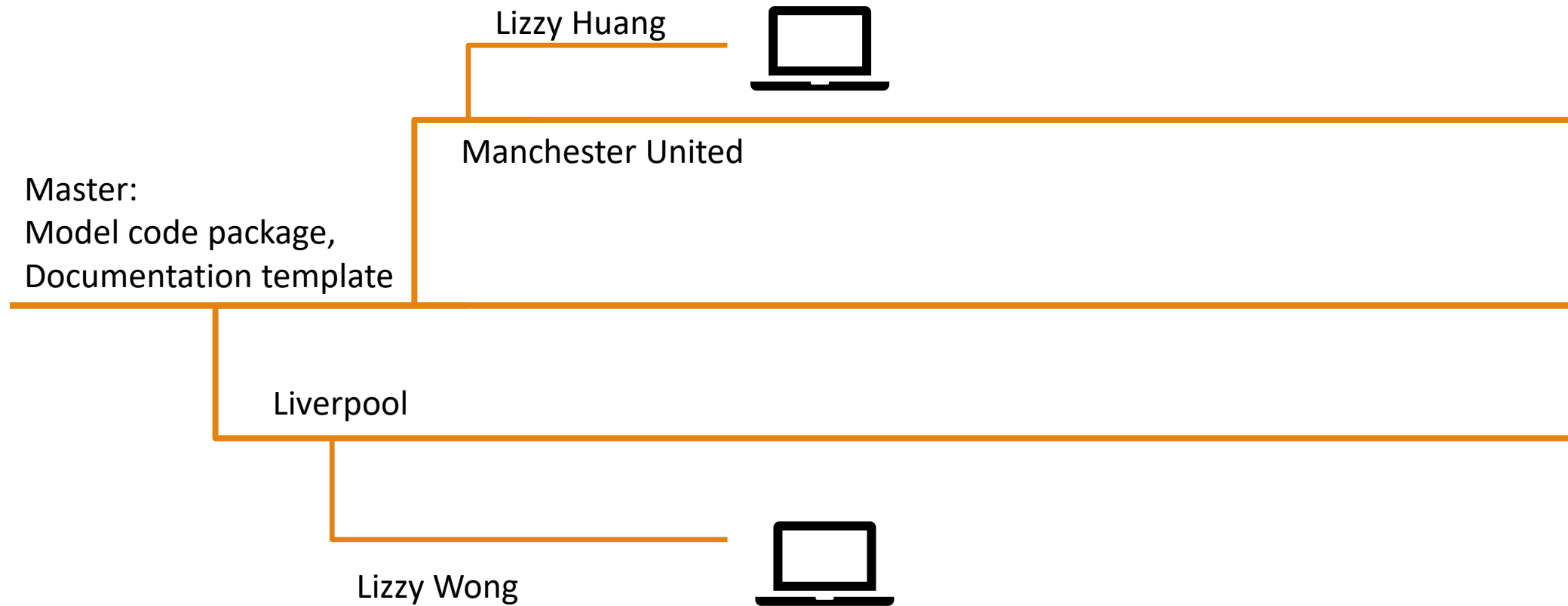
- Generate consistent model documentation template.

- ❑ Git for version control

- Automate entire model development process, post-model analyses, and synchronize documentation.
- Control code version while enable flexibility of team collaboration.

- ❑ Demo

Git



Summary

- ❑ Current model development process uses multiple programming and reporting tools that are not necessarily integrated.
- ❑ Lack of interactive coding platform and version control creates difficulties in tracking code/data changes, standardization, reproducibility, collaboration, and automation.
- ❑ R Markdown together with Git provides the ideal solution involving an interactive and centralized platform in conjunction with version control.