



Using R to Automate the Development of Custom Reports in Word Format

Alexandra Gates, Principal Data Analyst
NORC at the University of Chicago

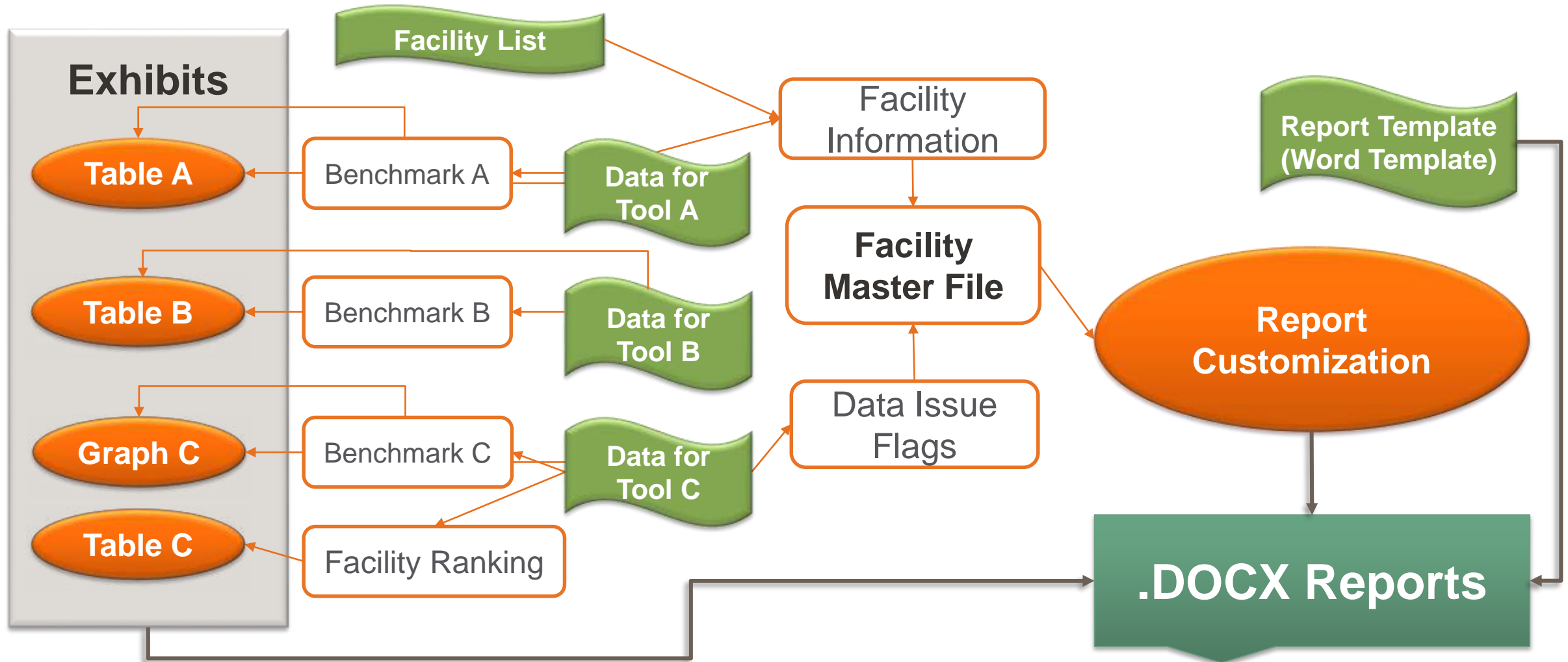
September 23, 2019

- **Background**
- **Preprocessing:** Data → Interim Files
- **Visualization:** Interim Files → Exhibits
- **Customization:** Interim Files → Customization
- **Work with Word:** Template + Exhibits + Customization = Reports!
- **Lessons Learned**

Background

- What are we doing?
 - Primary data collection from multiple data sources for ~500 facilities
 - Reporting data collected per facility vs. benchmark by facility size
- Why are we customizing reports?
 - Data reported & data included varies across facilities
 - Want to ensure continued facility participation
- Why create easy to understand graphics?
 - Want facilities to engage with the data and use it to inform performance
 - Want to reduce back and forth between us and the facilities to explain data included in the report

Overall Process



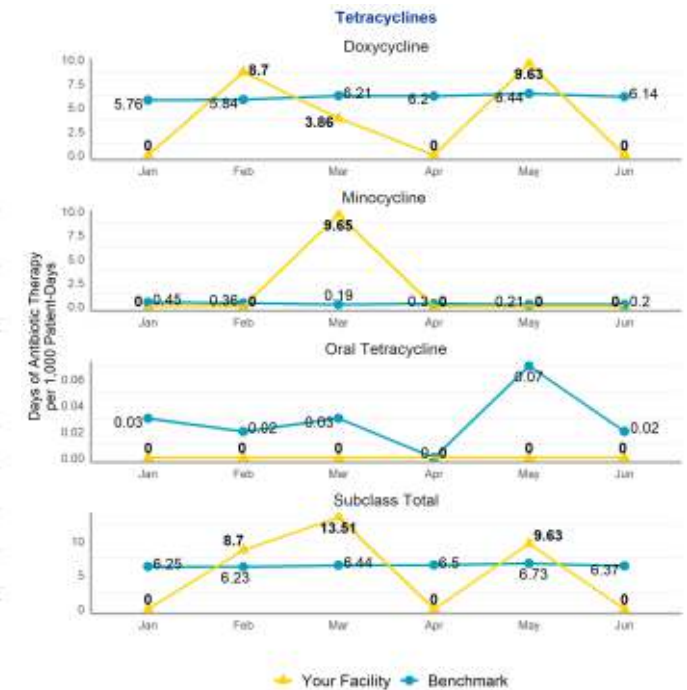
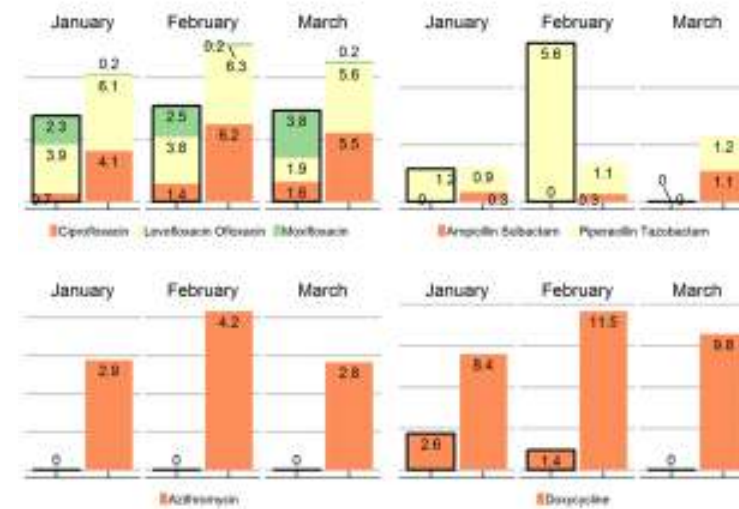
Preprocessing Data

- Analysis is only as good as the data
- Why did we need to do extensive preprocessing?
 - Humans vs. Machines
 - Creation of flags & reporting exclusion reasons
- What libraries did we use?
 - Tidyverse: dplyr, reshape2
 - Gather()
 - Spread()

Visualization

- What packages did we use?
 - ggplot2; ggrepel; gridExtra; ggthemes
- Types of graphs?
 - Line and bar charts
- Ease of consumption
 - Continuity in themes & colors
 - Colors chosen with facilities in mind

Figure 3. Monthly Antibiotic Days of Therapy per 1,000 Patient-Days: Antibiotic Subclasses.



Production

- Starting with Facility Master File
- Main script calls RMD in a loop
 - ④ 1.1 Master Script R File
 - ④ 1.2 Knit Report RMD File
- Main script passes parameters to RMD in each loop
- Parameters: data in Master Facility File → global variables in RMD

Customizations

- Print paragraph (or not)

```
`r if(dot.issue){" Your facility's antibiotic days data may not be presented or not included in the benchmark in this section for several reasons. Below, we summarize relevant data issues of your facility for each month."}`
```

- Print variables (or not)

```
`r dot.reason.1`  
`r sep`  
`r dot.reason.2`  
`r sep`  
`r dot.reason.3`
```



- January: No data was submitted
- January: Data was submitted in an older template version

- Print table (or not)

```
```{r dot_table, echo=FALSE, eval=dot.data}  
kable(dot_data[[2]], booktabs = T, full_width = T)
```
```


Work with WORD (Formatting)

Things you can do:

- Logo as header
- Page numbers in footer
- Font, size, formats for different levels of titles, **bolding**, italic, document margins
- Paragraph spacing

Things that were a little more difficult:

- Table formatting
- Insert blank lines

```
---  
title: "Quarterly Benchmarking Report, 1st Quarter, January to March 2019"  
output:  
  word_document:  
    reference_docx: word-styles-reference.docx
```

Lessons Learned

- Missing versus non-existent data: ggplot knows the difference
- Special characters and extra spaces in names can disrupt report production
- Plan ahead and organize code to have a longer shelf life
- QC is important
- R + word is extremely flexible, after a bit of wrangling
- File sizes can get unruly

Acknowledgement + Contact Information + Thank you!

- Mina Zheng for being my coding partner in crime on this endeavor
- Yue, Tyler, and full NORC team for giving us clean data, providing feedback on templates and graphics, and helping with QC.
- **Email: Gates-Alexandra@norc.org**
- Thank you!