
R Workshop

+ Tips and Tricks

Alea Seifert, Jett Palmer,
Kaatje Matthews, Libby Brill



Agenda



GitHub

Working collaboratively in R



Exploration

How to get started
investigating data



Presentation

What makes an effective
presentation?



General

Overall tips for the DataFest
weekend

Follow Along

- If you have RStudio downloaded, we recommend using the Quarto document sent out in the email
- If not, feel free to use the Posit Cloud link (also in the email):
<https://posit.cloud/content/12064067>
 - Make sure to “Save a Permanent Copy”

GitHub Tips

- When data are too large to push, use identical file paths, include dataset in `.gitignore`
- Store data cleaning steps in an R script and use `source()` to call it
- Avoid merge conflicts by creating separate files for each group member



Exploratory Data Analysis

● Contextualize

Orient yourself to the **project domain**

Learn about **variables** using the codebook, data description, and other documentation

● Identify

Discover **missing data**. Why is it missing and what is the resolution?

Discover **outliers**. Are they real? What do they mean? What is the resolution?

● Observe

Create **simple graphics** to explore relationships and distributions on a high-level. What **interesting patterns** do you find?

Next Steps

- **Narrow the Scope**

It's okay (and even recommended) to not use all the variables in your analysis. If you are overwhelmed, try to **filter or subset**.

- **Feature Engineering**

What variables will you choose to **focus** on? What variables can you **create** (e.g. ratios, aggregate metrics, scores)?

- **External Sources**

Import external data to **fill gaps** in your data or bolster your story with **another angle**.

Presentation

Graphic Tips

- Limit to only 1–2 graphics, so focus on informative graphs
- Highlight what is important
 - Color/bold important features
- Make sure text is readable (axis titles, legends, etc.)

Story Telling

- Be curious during exploratory analysis and “look for a lead”
- Focus on 1–2 interesting ideas and connect them deeply to the domain area
- Consider how your findings might impact stakeholders

Presentation

Curate Content

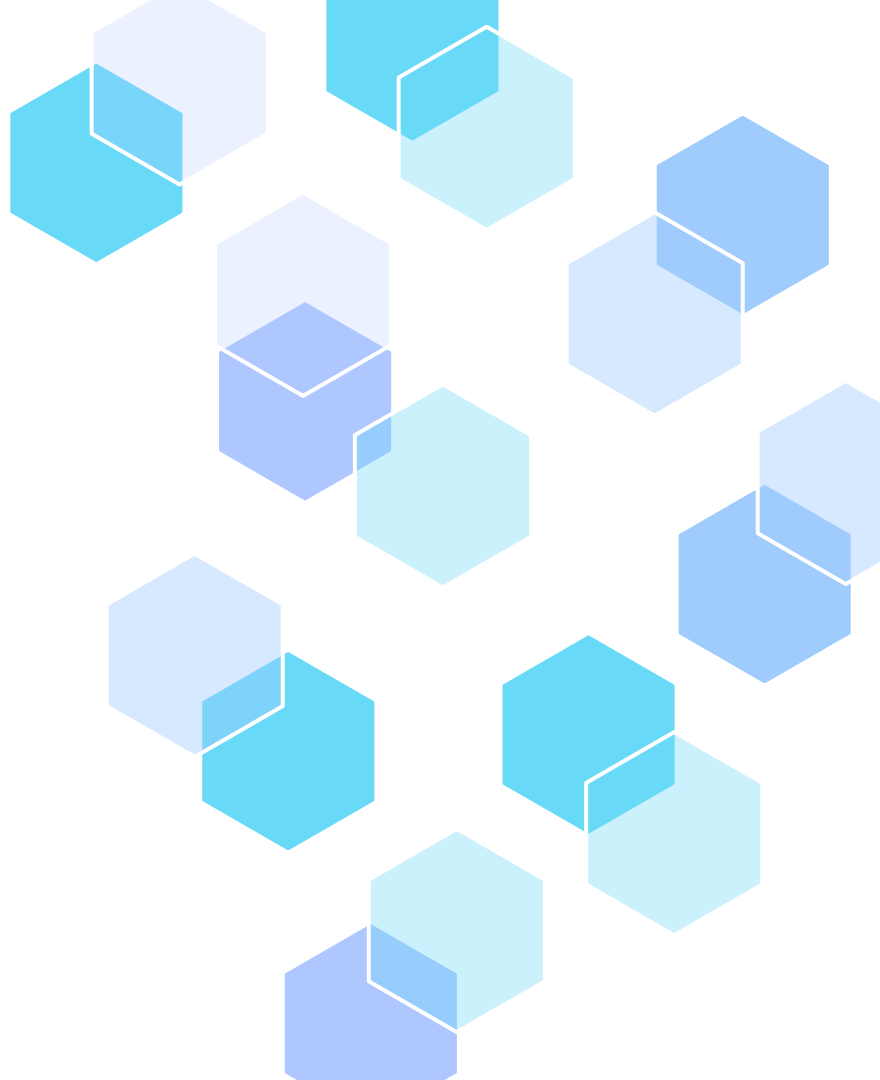
- Less about the journey and more about the key takeaways
- It is okay if you don't discuss everything you have done
 - EDA
 - Assumptions
 - Previous analysis
 - etc.

Practice

- Decide who is presenting each topic
- Use an extemporaneous style
- Ensure everyone understands the key features of the presentation

General Tips

- Being willing to let ideas go
- Don't put all your eggs in one basket
- Working with team
 - Take breaks!
- Use your imagination, be curious
- Take advantage of mentors and professors
- Keep the awards in mind



Thanks!

Questions?

