Presenting Model Results

ID 529 Final Presentation

Shady Abohashem, Ella Douglas-Durham,
Raveena John, Audrey Murchland & Alex Zapf
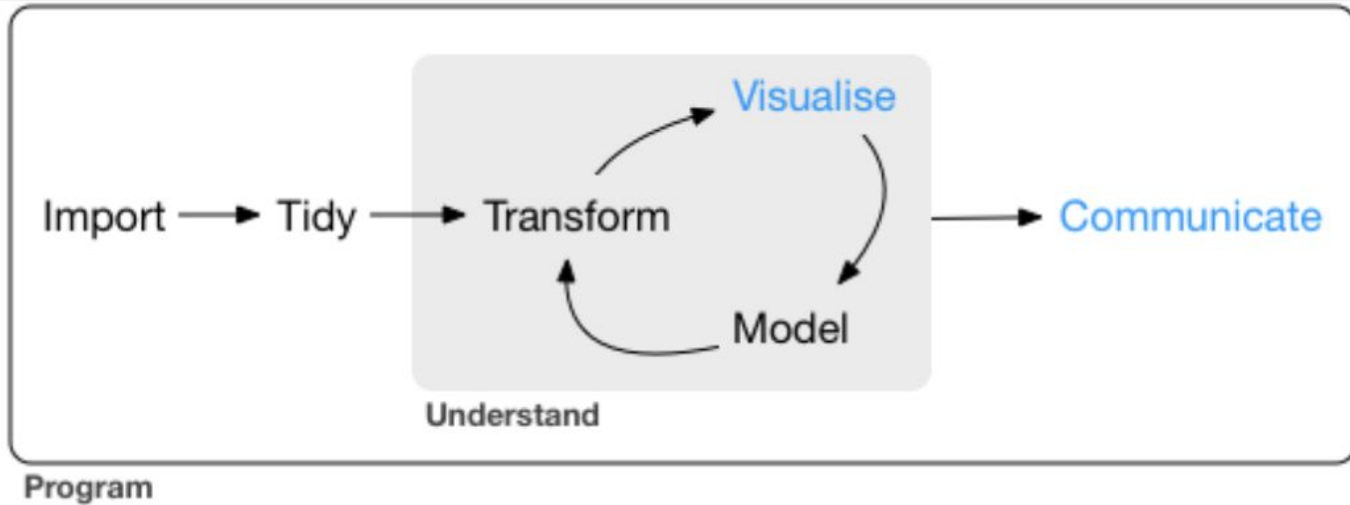
January 20, 2023

Why and to whom do we present model results?

Analyst-to-analyst → decisions return to inform modeling process

Analyst-to-scientist → inform other scientific research hypotheses

Analyst-to-decision makers → inform external policies/programs

Image Source: R for Data Science

Communication made easier: R Markdown

Analyst-to-Analyst

- html_document
  - Can render to documents which will display text, code, and results in various formats: Html documents, pdfs, word documents, opendocument text, rich text formats
- html_notebook
  - notebook in which source code can be *independently* executed
  - focus is primarily on collaboration
  - best to combine with Git and Github

Analyst-to-Scientist/Decision Makers

- Presentations (built-in)
  - HTML: ioslides_present or slidy_presentation
  - PDF: beamer_presentation
  - New slides:
    - first or second level headers (# or ##)
    - Without header (***)
- Dashboards
  - flexdashboard
    - new pages: #
    - new columns: ##
    - new rows: ###
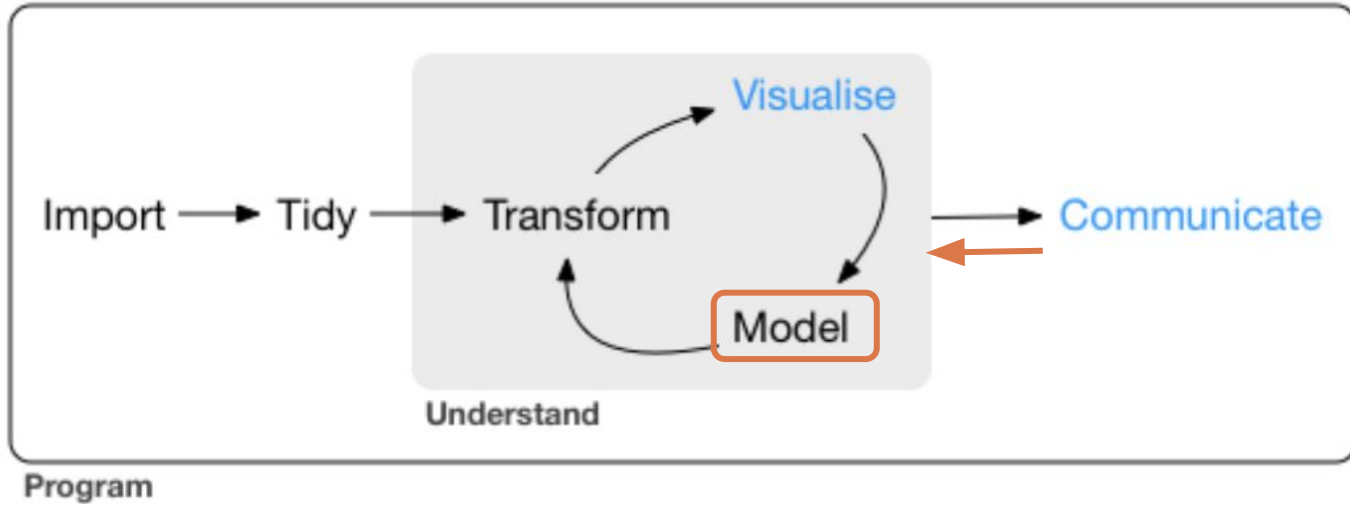
Balance and interaction between research question, data, discipline, and intended audience inform the appropriate presentation format(s) utilized as well as what and how model results are ultimately included.

Recommendations & Best Practices

General Practices.....

- Clearly state the **research question and the methods** employed

- Present results in **a concise and easily interpretable** manner

- Describe any model **limitations** or **assumptions**

- **Interpret the results and discussing their implications**

In the context of the problem being addressed

Zuur, A.F. Methods Ecol Evol 2016.

● Present **statistical model** and Clearly report on **its performance**

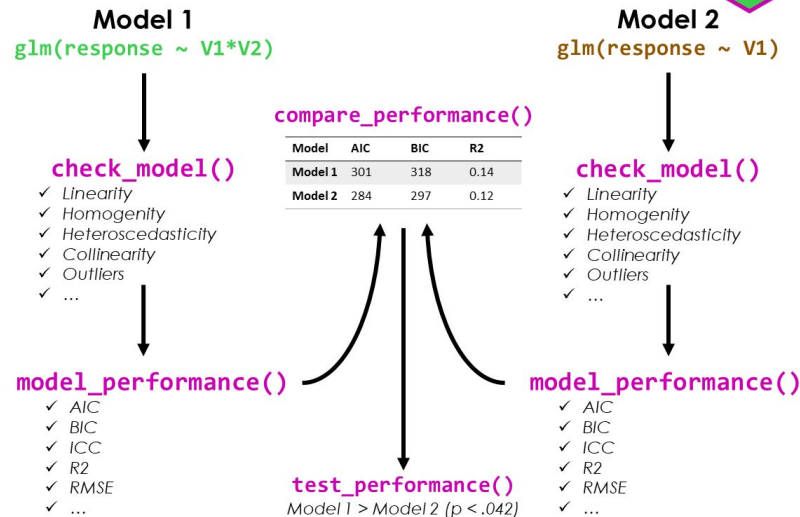**Approach 1**

$$NCalls_{ij} \sim Poisson(\mu_{ij})$$
$$E(NCalls_{ij}) = \mu_{ij}$$
$$\log(\mu_{ij}) = SexParent_{ij} + FoodTreatment_{ij}$$
$$+ArrivalTime_{ij} + SexParent_{ij} \times FoodTreatment_{ij}$$
$$+SexParent_{ij} \times ArrivalTime_{ij} + Nest_i$$
$$Nest_i \sim N(0, \sigma^2)$$

Approach 2

$$RD_{ijk} \sim Beta(\pi_{ijk})$$
$$E(RD_{ijk}) = \pi_{ijk}$$
$$var(RD)_{ijk} = \pi_{ijk} \times (1 - \pi_{ijk})/(1 + \theta)$$
$$logit(\pi_{ijk}) = Time_{ijk} + Relatedness_{ijk} + GroupSize_{ijk}$$
$$+Time_{ijk} \times Relatedness_{ijk}$$
$$+Relatedness_{ijk} \times GroupSize_{ijk} + Groomer_i$$
$$+Hour_j + Receiver_l$$
$$Groomer_i \sim N(0, \sigma^2_{Groomer})$$
$$Hour_j \sim N(0, \sigma^2_{Hour})$$
$$Receiver_l \sim N(0, \sigma^2_{Receiver})$$

Using the *performance* R package

Model 1
`glm(response ~ V1*V2)`

check_model()
✓ Linearity
✓ Homogenity
✓ Heteroscedasticity
✓ Collinearity
✓ Outliers
✓ ...

model_performance()
✓ AIC
✓ BIC
✓ ICC
✓ R2
✓ RMSE
✓ ...

compare_performance()

Model	AIC	BIC	R2
Model 1	301	318	0.14
Model 2	284	297	0.12

test_performance()
Model 1 > Model 2 (p < .042)

Model 2
`glm(response ~ V1)`

check_model()
✓ Linearity
✓ Homogenity
✓ Heteroscedasticity
✓ Collinearity
✓ Outliers
✓ ...

model_performance()
✓ AIC
✓ BIC
✓ ICC
✓ R2
✓ RMSE
✓ ...

Zuur, A.F. Methods Ecol Evol 2016.
Easystats.github.io
Biecek p., exploratory model analysis, 2020.

- **Provide the code used to perform the analysis** *(R. Script or R. markdown)*

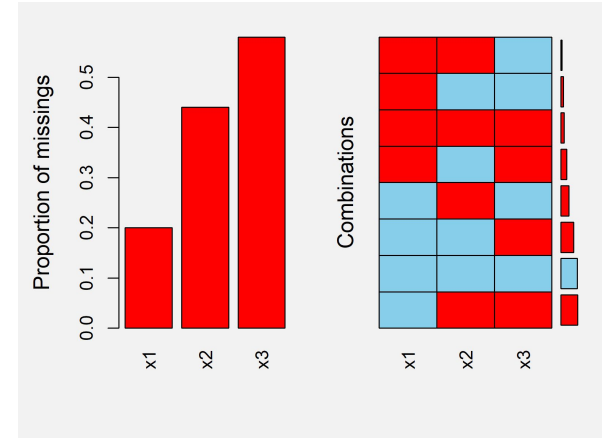- **Data preprocessing, cleaning and handling missing data**

- **Report on missing**
- **Look for trends**
- **Match your imputation**

R **packages**
- **is.na**
- **VIM**
- **MICE**
- **missForest**

Visualizing missing data.

- **Following reporting guidelines for health research studies (e.g. STROBE, PRISMA, CONSORT)**

(EQUATOR website)

Kevin Donovan 2019
Reporting missing values in R.
UC business analytics guide.
Simera et al. EJCI 2009.

9

Recommendations

- Know your **audience**

- Understand your **objective**

- Tailor the **style** to your audience

- Be **concise**, yet sufficient

Adapted from The Art of Data Science, by Roger D. Peng & Elizabeth Matsui

Best Practices for Visualizing Results

1. Present findings in **substantive terms** to be easily interpreted

2. When possible, **show data**

3. Illustrate degree of **confidence** (`geom_pointrange`, `geom_errorbar`)

4. Plot marginal effects and complex survey results with R packages (`margins`, `survey`, `srvyr`)

From Data Visualization: A Practical Introduction, by Kieran Healy

https://socviz.co/modeling.html#get-model-based-graphics-right
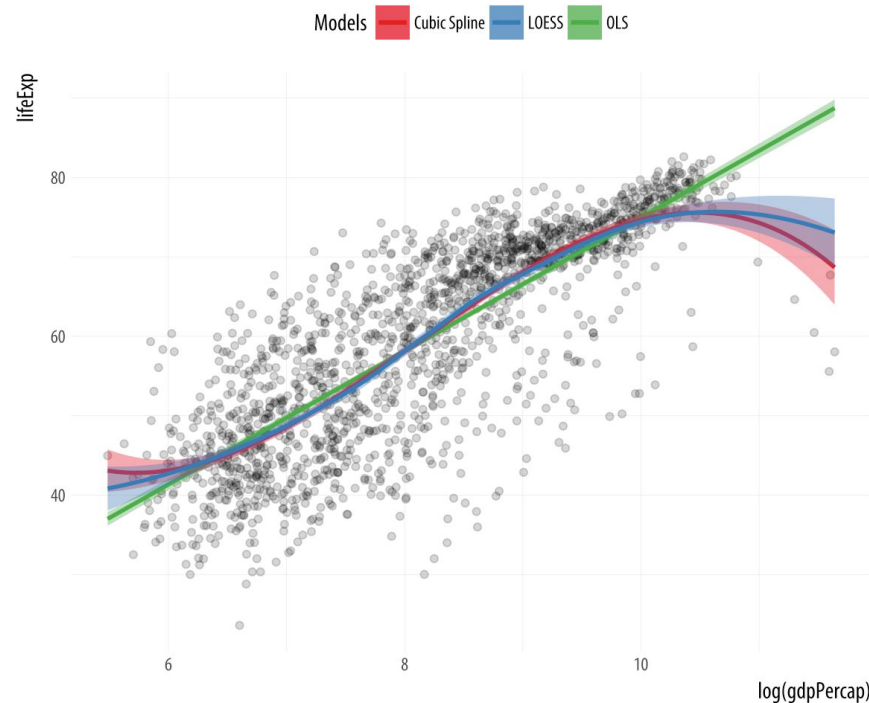
Model components

Key considerations

- Who is the **audience**?
    - determines what to report and in how much detail
- What **type** of model are you working with?
    - Structure, time horizon & cycle length, disease states/pathways, parsimony
- Scarcity/lack of unified **guidelines**
    - Structure, Data, Consistency, Validity, Computer implementation, Transparency, Sponsorship
    - Few guidelines: Health Technology Assessment;
        - TRIPOD statement

Dahabreh et al 2011
Collins et al. 2015
AHRQ 2016
Bennett & Manuel 2012
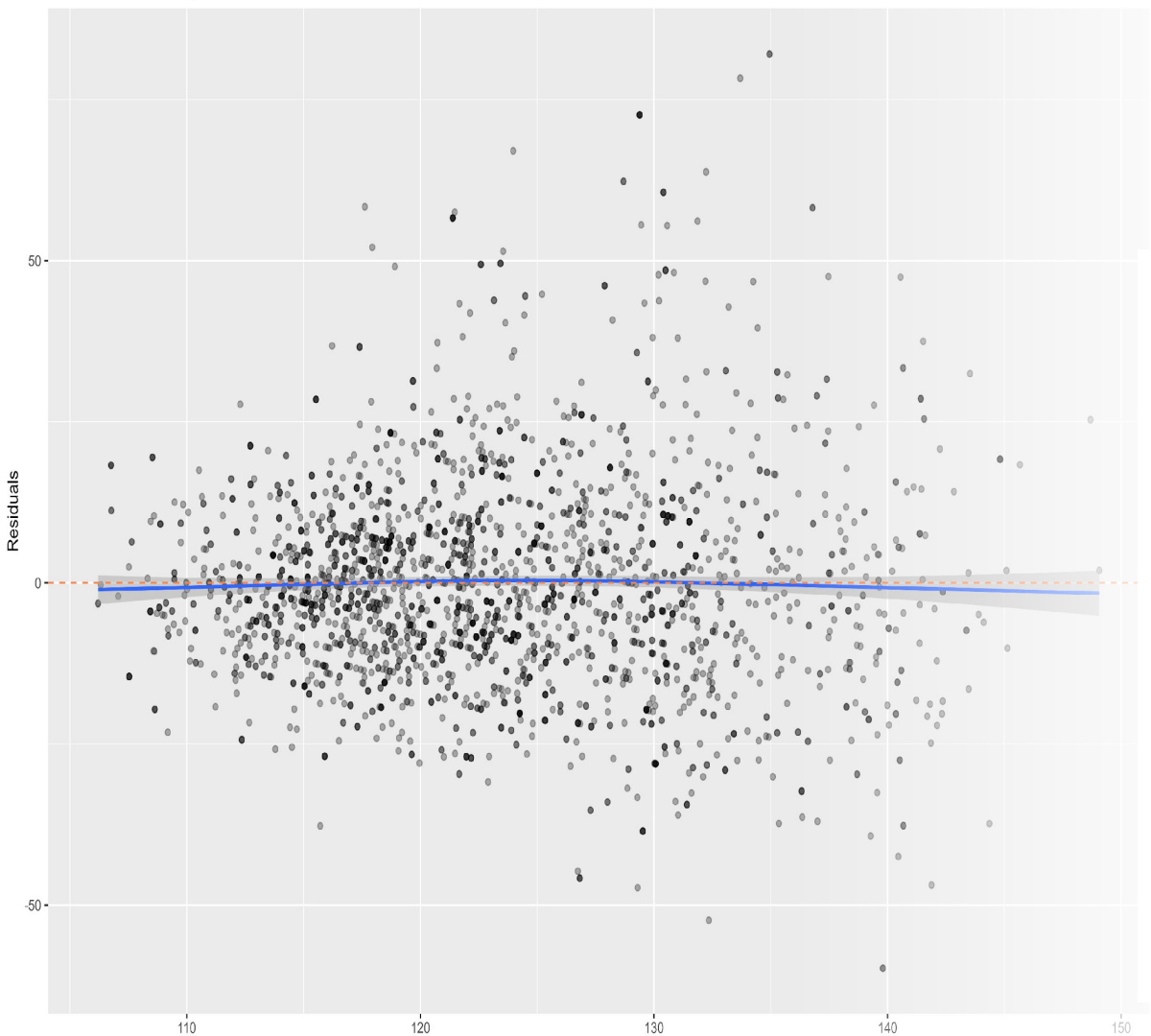
Model diagnostics

- Model **fit statistics**
  - likelihood-based statistics (deviance likelihood, LRT, AIC)
  - `broom::glance(model)`
- **Plotting**
  - fitted vs. residual plots
  - fitting models to observed data
- (Re-)**calibration**
  - `rms::calibrate()`
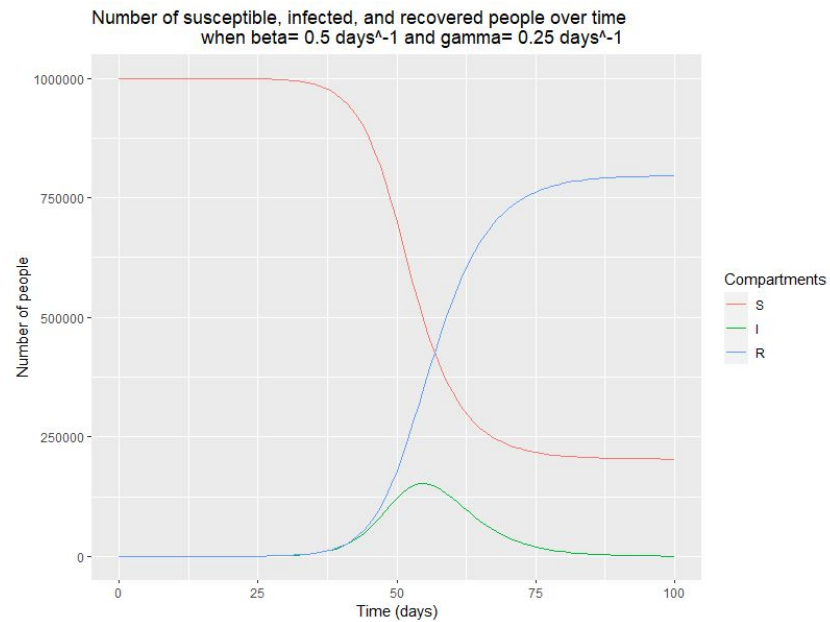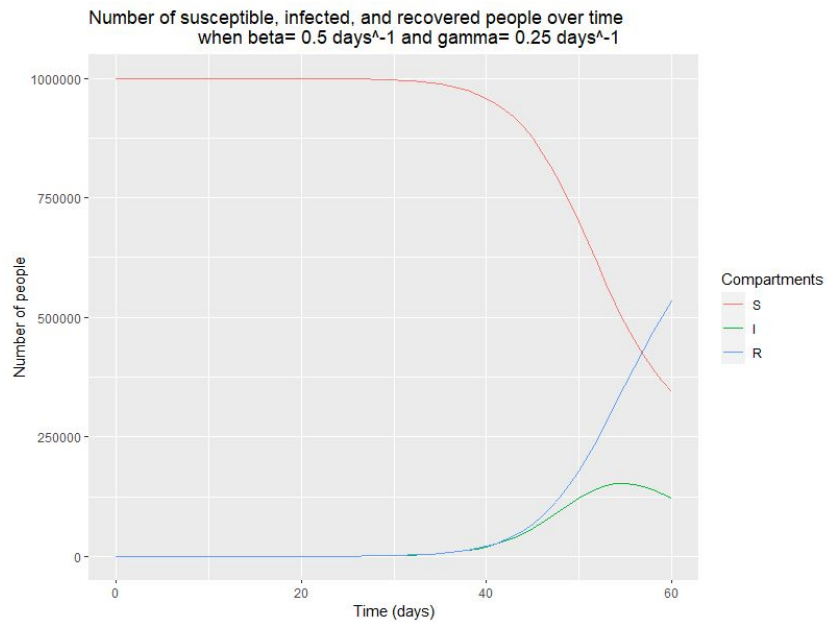 - Tidyverse: `probably` and `tidymodels` packages

14

Scaling

NATURAL VS. LOG (LN) SCALE

TIME SCALES

Number of susceptible, infected, and recovered people over time
when beta= 0.5 days^-1 and gamma= 0.25 days^-1

Number of susceptible, infected, and recovered people over time
when beta= 0.5 days^-1 and gamma= 0.25 days^-1

Resources for Reporting Model Results

gtsummary

- "The {gtsummary} package provides an elegant and flexible way to create publication-ready analytical and summary tables"
- `tbl_summary()` summarizes data frames or tibbles
    - Automatically detects continuous, categorical, and dichotomous variables in your data set, calculates appropriate descriptive statistics, and also includes amount of missingness in each variable
- `tbl_regression()` summarizes regression models
    - Common regression models are automatically identified and the tables are pre-filled with appropriate column headers
- Users can combine tables with `tbl_merge()` or `tbl_stack()`

gtsummary Resources

- ?gtsummary
  - gtsummary github: https://github.com/ddsjoberg/gtsummary
    - Includes a summary of the package
    - Provides examples of how to use the package
    - Links to additional resources
    - FAQ : https://www.danieldsjoberg.com/gtsummary/articles/gallery.html
- *Reproducible Summary Tables with the gtsummary Package*
  - An article in The R Journal by the maintainer and authors of the gtsummary package

Other packages

- finalfit
  - An "all-in-one" function that takes a single dependent variable with a vector of explanatory variable names to produce a final table for publication including summary statistics
- skimr
  - Designed to provide summary statistics about variables in data frames, tibbles, data tables and vectors
- stargazer
  - Produces LaTeX code, HTML code, and ASCII text for well-formatted tables that gold regression analysis results from several models side-by-side and can output summary statistics
- tableone
  - Creates "Table 1", i.e., description of baseline patient characteristics, and supports both continuous and categorical variables, as well as p-values and standardized mean differences

Collected Resources

R Markdown

- R Markdown Online Resources: https://rmarkdown.rstudio.com/
    - Dashboards: https://pkgs.rstudio.com/flexdashboard/

Recommendations & Best Practices

- A protocol for conducting and presenting results of regression-type analyses
- A catalogue of reporting guidelines for health research. EJCI 2009.
- *The Art of Data Science*
- *Data Visualization: A Practical Introduction*

Model Components

- TRIPOD Statement for model reporting
- `rms`::calibrate documentation

Reporting Model Results

- https://github.com/ddsjoberg/gtsummary
- Sjoberg et al. (2021). Reproducible Summary Tables with the gtsummary Package. *The R Journal, 13*(1): 570-580.

Questions?