An Approach to Project Workflow for Professional Biostatistical Services

Paul G. Stevenson and Matthew N. Cooper

Telethon Kids Institute, The University of Western Australia, Perth, Western Australia, Australia

Paul.Stevenson@telethonkids.org.au

BACKGROUND

Large research groups commonly employ a biostatistician to work across their portfolio of research projects, however, this is not feasible for many research active clinicians and "low-profile/establishing" research groups who often struggle to access biostatistical support.

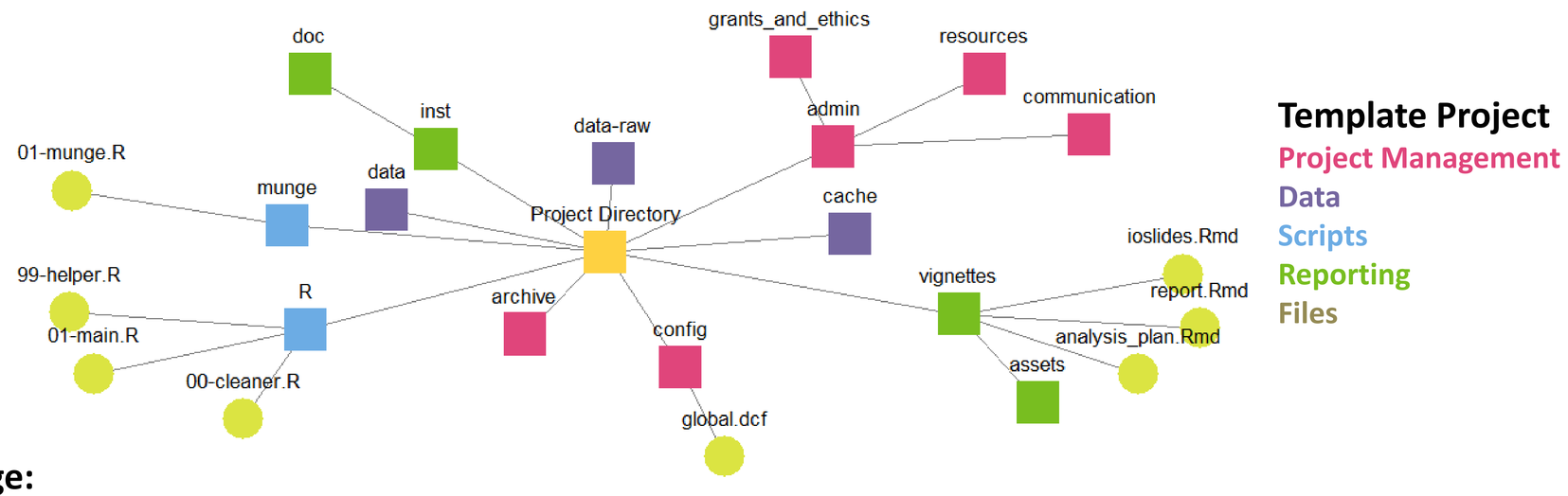
Our group offers "low-barrier" initial consultations for analysis, data management and database development on a "per-project" basis that is attractive to the local health research community as it makes biostatistical expertise affordable and easily accessible.

MOTIVATION

- Standardise our workflow and automate new project initiation using the R programming language.
- Incorporate the principles of "reproducible research" to ensure every step of our analysis is traceable and verifiable.
- Create beautiful HTML documents out-of-the-box that don't require knowledge of HTML or Javascript. (https://github.com/TelethonKids/biometrics).
- Set-up a toolbox to help us and our collaborators to get meaning out of data by focusing on "what does this mean" not "how can we do this."

RESULTS

Our Workflow Relies on a Clean Project Template That Facilitates Reporting, Sharing and Collaboration:



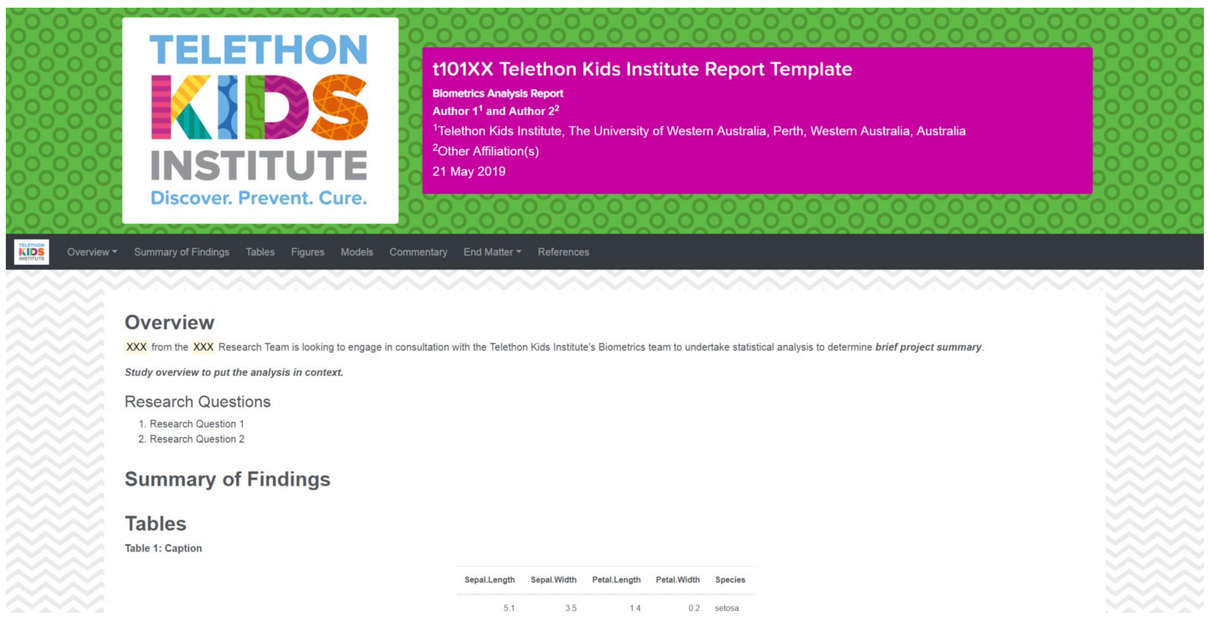
Project as a Package:

Our default projects have been designed so they can be packaged and shared with our collaborators: cleaned data is available with a data dictionary, helper functions are documented, all reports available as vignettes, and an audit history is maintained by using tools like GitLabs. Package versions can be frozen *via.* implementation in a Docker container. (https://github.com/TelethonKids/rstudio).

Here is an "R Toolbox for Biostatistics" So You Can Focus On Your Analysis:

ProjectTemplate	Project initialisation and automation
Captioner	Cross-reference tables/figures/models
KableExtra	Nicely format data frames for reporting
Broom	Extract a model's estimates and statistics
Repmis	Create a bibliography of loaded packages
Devtools::build_vignettes()	Automatically Knit all vignette .Rmd files
jtools::plot_summs()	Visualise a model(s) effect estimates and CIs
Gggally::ggpairs()	Look at your data with a plot pair matrix

Our Default R Markdown HTML Report Template – No HTML needed to create Documents That Are Aligned With Our Institute Style Guide:



Project Creation is Quick and Simple:

```
options(ProjectTemplate.templatedir = "path/to/templates")
create.project(paste("path/to/projects", "00_project", sep = "/"), template = "biometrics_project")
```

CONCLUSION

- This approach allows us to streamline our workflow to expeditiously initiate projects and produce professional looking reports in multiple formats directly from the analysis package without wasting time on the non-analytical aspects of our projects.
- The workflow that we use is identical and successful for both simple and large-scale projects.
- Our researchers have been impressed by the capability of `ggplot2` and `plotly` packages to visualise and better understand their data.

TELETHON KIDS INSTITUTE
Discover. Prevent. Cure.