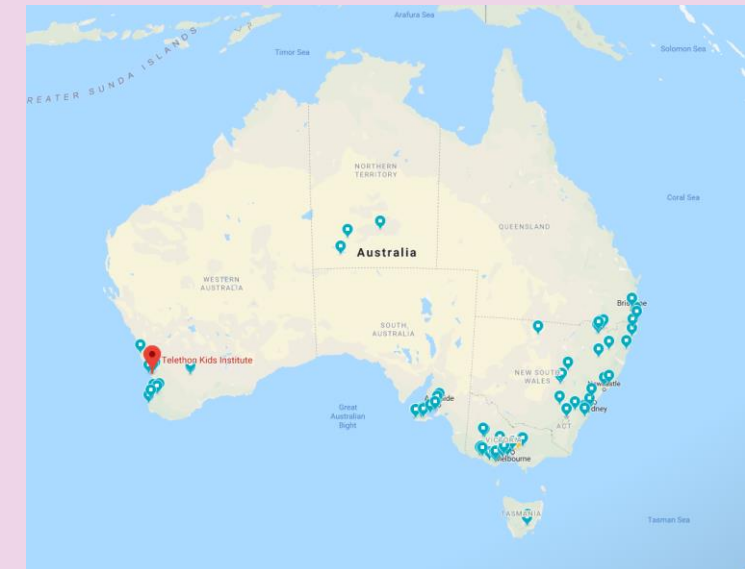# An Approach to Project Workflow for Professional Biostatistical Services

Paul G Stevenson and Matthew N Cooper

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

Background

- Large research groups commonly employ a biostatistician to work across their portfolio of research projects, however not all researches have this capacity

- The Biometrics group at Telethon Kids offers "low-barrier" initial consultations for analysis, data management and database development on a "per-project" basis

- This is attractive to the local health research community as it makes biostatistical expertise affordable and easily accessible.

Background

- Large research groups commonly employ a biostatistician to work across their portfolio of research projects, however not all researches have this capacity

- The Biometrics group at Telethon Kids offers "low-barrier" initial consultations for analysis, data management and database development on a "per-project" basis

- This 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

- 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"

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

- 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"
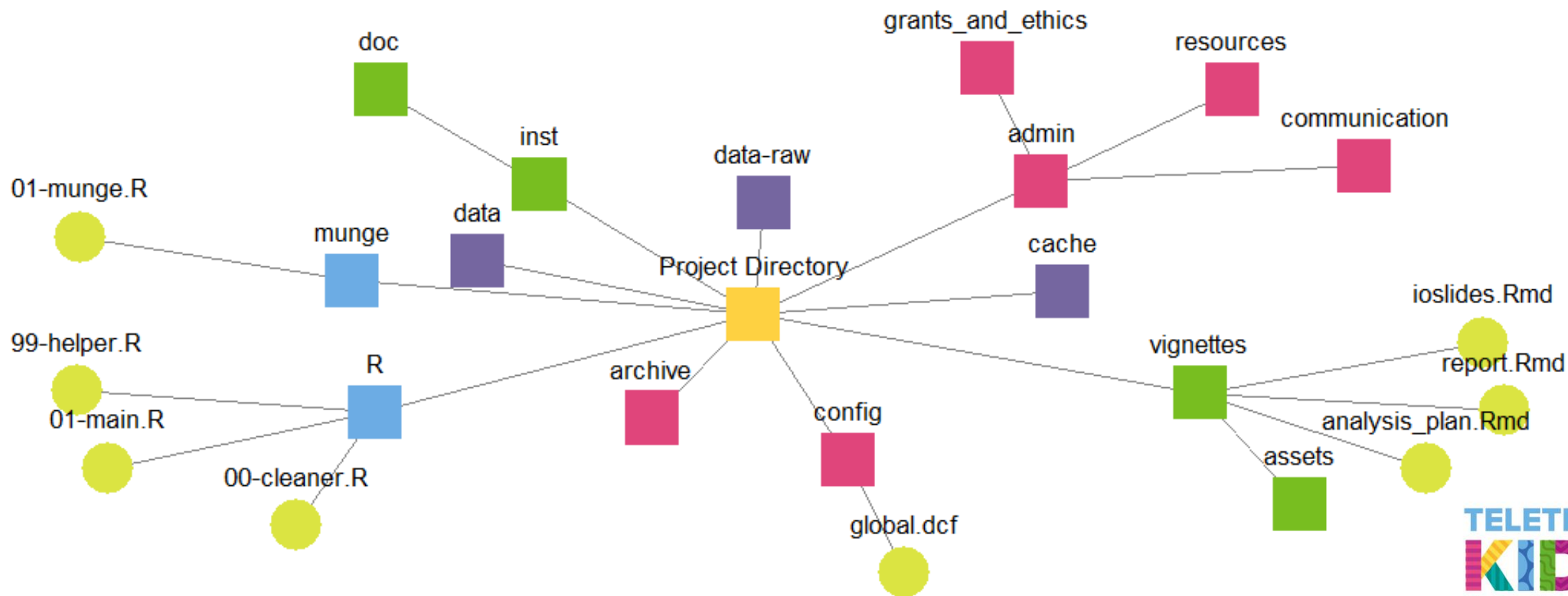
Standardised Workflow

Template Project Directory
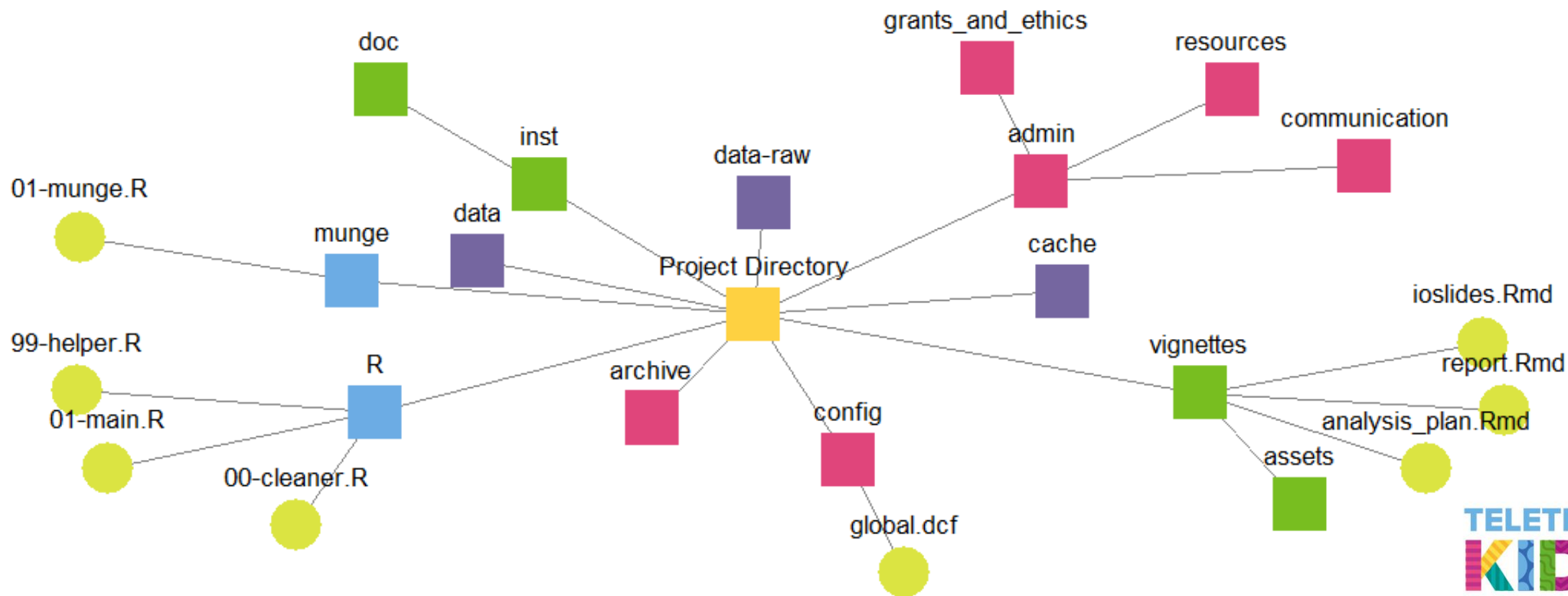
Standardised Workflow

Template Project Directory

Project Automation

- Our default project templates build on the ProjectTemplate package

- Automate:
  - Raw/cached data loading (either, both or none)
  - Execute data wrangling scripts
  - Load project libraries

- Project Creation is Quick and Simple:

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

Reproducible Reporting

- R Markdown allows a seamless transition between text and script in a single document

- Automate tables, figures, models, analysis and commentary in a single document

- When the underlying data is changed, the report can be designed to completely update to reflect the current state of collection

- Beautiful documents can be created with little extra scripting beyond the analysis

- Version control *via.* GitHub/GitLab/BitBucket

Reproducible Reporting

- R Markdown allows a seamless transition between text and script in a single document

- Automate tables, figures, models, analysis and commentary in a single document

- When the underlying data is changed, the report can be designed to completely update to reflect the current state of collection

- Beautiful documents can be created with little extra scripting beyond the analysis

- Version control *via.* GitHub/GitLab/BitBucket

TELETHON
KIDS
INSTITUTE
Discover. Prevent. Cure.

t101XX Telethon Kids Institute Report Template

Biometrics Analysis Report

Author 1[1] and Author 2[2]

[1]Telethon Kids Institute, The University of Western Australia, Perth, Western Australia, Australia

[2]Other Affiliation(s)

21 May 2019

Overview

XXX from the XXX Research Team is looking to engage in consultation with the Telethon Kids Institute's Biometrics team to undertake statistical analysis to determine *brief project summary*.

Study overview to put the analysis in context.

Research Questions

1. Research Question 1
2. Research Question 2

Summary of Findings

Tables

Table 1: Caption

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa

Our Toolbox

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 Toolbox

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

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` graphic tools to help them visualise and better understand their data

Acknowledgements

- UseR!2019 Organising Committee

- Matt Cooper, Bethy McKinnon, Peter Jacoby

- Telethon Kids Institute People and Culture Team

- Out template project is available as a Docker container: https://github.com/TelethonKids/rstudio

- The HTML reporting template is on GitHub: https://github.com/TelethonKids/biometrics