



2019-06-05-Wednesday

What: DSS Session on Building Reproducible Visualizations with *ggplot2*

When: Wednesday, June 5, 11 am – 1 pm Pacific

Where: Institute on Aging and Lifelong Health, [R-Hut](#) room 103a

Remote: [BlueJeans](#) or 1-416-900-2956 (ID 176927559#)

Log: [dss-session-log-2019](#)

Hello everyone,

For the [2019-06-05](#) session of DSS-IALH we would like to shift gears and feature the first **applied session** of the summer. This session will demonstrate a production workflow for generating reproducible data visualisations in *R* using the grammar of *ggplot2* and *dplyr* packages in an *RStudio* project.

Synopsis.

Rarely do applied data science projects produce any given graph only once. The need to generate plots of the same form using different inputs, different options, and in multiple contexts requires the analyst to structure the operations involved in graph production as customizable functions. The technique demonstrated in this session divides the production of data visualizations into three sections (prep - plot - print), each governed by a dedicated custom function. The first function prepares the data for graphing, the second plots the graphic, and the third prints the image to disk. This **applied session** will walk the learner through the stages of developing such a chain of functions and demonstrate the advantages of such an approach in reproducible projects. Data and starter scripts will be provided.

Software in focus

RStudio, R, specifically *ggplot2* and *dplyr* packages.

Key ideas of the session

1. Graph as a set of functions
2. Production sequence = (prep + plot + print) + place
3. Dynamic documents with RStudio: (*.R) + (*.Rmd) = (*.html, *.pdf)

Please let us know if you have any questions about the upcoming session.

Feel free to bring friends.

We look forward to thinking together with you.

Yours,

data science studio

