

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)

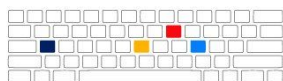
Please let us know if you have any questions about the upcoming session.

Feel free to bring friends.

We look forward to thinking together with you.

Yours,

data science studio



Data pulled from the Canadian Chronic Disease Surveillance System (CCDSS)

Mental Health Services

- URL: <https://health-infobase.canada.ca/ccdss/data-tool/>
- Tab: Trends over time
- Geography: Canada
- Condition: Use of health services for mental illness (annual)
- Measure: Crude prevalence rate
- Sex: Both sexes + Males + Females
- Fiscal Year: 2000:2015
- Local file: ./data-public/raw/scenario-3/PHAC_Infobase_CCDSS_-8586421810172823624.csv

Shared Services Canada [CA] | <https://health-infobase.canada.ca/ccdss/data-tool/?DDL=3&DDL=ASPR&sex=B&yrB=2015&=10&VIEW=0>

Filter items: Showing 1 to 10 of 4,242 entries | Show 10 entries

Download Detailed Table

Detailed Table: Use of health services for mental illness (annual), crude prevalence rate, percent

Geography	Age Group	Sex	Fiscal Year	Rate	95% CI	CV (%)	Number
Canada	1+	Both sexes	2000	14.48	14.46 - 14.49	0.05	4,546,270
Canada	1-19	Both sexes	2000	6.97	6.96 - 6.99	0.14	541,820
Canada	20-34	Both sexes	2000	13.91	13.88 - 13.94	0.11	906,390
Canada	35-49	Both sexes	2000	17.74	17.71 - 17.77	0.08	1,400,180
Canada	50-64	Both sexes	2000	17.96	17.93 - 18.00	0.10	916,880
Canada	65-79	Both sexes	2000	17.34	17.29 - 17.38	0.14	532,070
Canada	80+	Both sexes	2000	23.77	23.67 - 23.86	0.20	248,940
Canada	1+	Females	2000	17.19	17.17 - 17.21	0.06	2,729,050
Canada	1-19	Females	2000	6.28	6.26 - 6.31	0.20	237,980
Canada	20-34	Females	2000	17.88	17.83 - 17.92	0.13	579,310

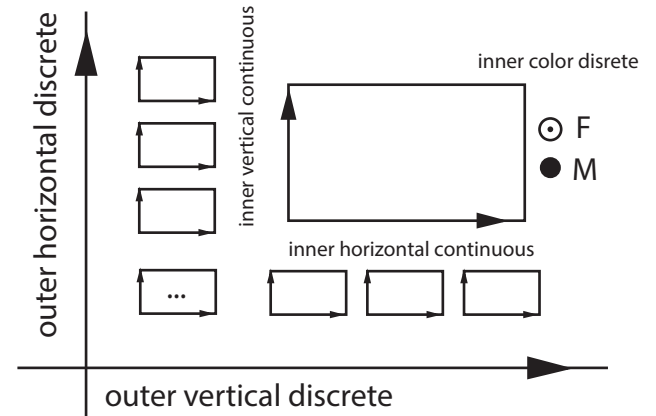
1 2 3 4 5 ... 425 Next

Variables: 11

```

$ area <chr> "Canada", "Canada", "Canada", "Canada", "Canada", ...
$ condition <chr> "Use of health services for mood and anxiety disorder..."
$ age_group <chr> "1+", "1-19", "20-34", "35-49", "50-64", "65-79", ...
$ sex <chr> "Both sexes", "Both sexes", "Both sexes", "Both sexes", ...
$ year <dbl> 2000, 2000, 2000, 2000, 2000, 2000, 2000, 2000, 2000, 2000, ...
$ rate <dbl> 10.71, 3.04, 11.11, 14.29, 14.41, 13.08, 13.35, 13.35, ...
$ rate_cv <dbl> 0.05, 0.21, 0.12, 0.09, 0.12, 0.16, 0.27, 0.07, 0.07, ...
$ rate_95_ci_lower <dbl> 10.70, 3.03, 11.08, 14.26, 14.37, 13.04, 13.28, 13.34, ...
$ rate_95_ci_upper <dbl> 10.72, 3.05, 11.13, 14.31, 14.44, 13.12, 13.42, 13.36, ...
$ number <dbl> 3364580, 236410, 723770, 1127780, 735350, 401440, ...
$ population <dbl> 31412970, 7781830, 6517490, 7893540, 5103900, 3068...
```

Blueprint of Reproducible Graph



I - DATA space

```

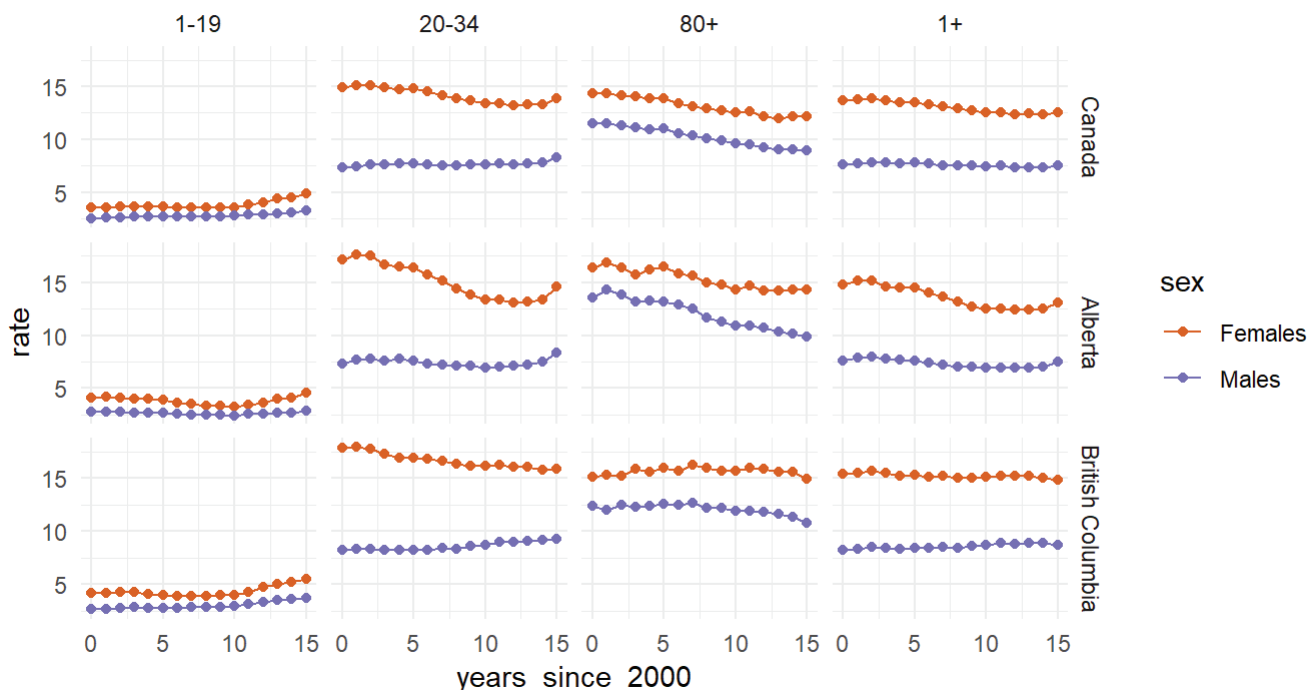
## MEASURE - rate # crude rate (not age-adjusted)
## MEASURE - number # count of cases of the condition
## MEASURE - population # total alive
## TIME - year # fiscal
## DESIGN - area ( 12 provinces + 1 total )
## DESIGN - age_group ( 6 groups + 1 total )
## DESIGN - sex ( 2 groups + 1 total )
## DESIGN - condition ( 1 )
```

II - VISUALIZATION space

```

## INNER - horizontal - TIME - (year)
## INNER - vertical - MEASURE - (rate)
## INNER - color - DESIGN - (sex)
## OUTER - horizontal - DESIGN - (area)
## OUTER - vertical - DESIGN - (age_group)
```

Crude prevalence of MH service utilization



Crude prevalence of MH service utilization in Canada

