



# 2019-06-19-Wednesday

<b>What</b>	Simplified `prep-plot-print` technique: Applications in Mental Health Surveillance
<b>When</b>	Wednesday, June 19, 11 am – 1 pm Pacific
<b>Where</b>	Institute on Aging and Lifelong Health, <a href="#">R-Hut</a> room 103a
<b>Remote Log</b>	<a href="#">BlueJeans</a> or <a href="#">1-416-900-2956</a> (ID 176927559#)
<b>GitHub</b>	<a href="https://github.com/dss-ialh/graph-making-scenarios">https://github.com/dss-ialh/graph-making-scenarios</a> (please clone)
<b>Handout</b>	<a href="#">session handout</a>

For the **2019-06-19** session of the Data Science Studio we would like to continue with the applied series, however the next time (**2019-07-03-Wednesday**) we will return to substantive focus with a dedicated session on Data Privacy and Disclosure Control. More details about this will follow in a separate announcement.

The previous session of DSS focused on building a reproducible graphing system in R that visualized the rate of Mental Health service utilization reported by Canadian Chronic Disease Surveillance System. We launched a reproduction of [github.com/dss-ialh/graph-making-scenario/scenario-3](#) report, which narrated the phases of transforming a concept *ggplot2* graph into a sequence of custom functions that automated graph production. Lessons from that session shaped [another demonstration of this technique](#), delivered during the [VADA 2019 Data Challenge](#) featuring the breast milk data of the [CHILD](#) study. Combining the feedback from both talks, we will further simplify the graphing example and proceed at a slower pace, while offering an opportunity for a self-paced advance through the developmental stages of this minimalistic automated graphing system.

**What should I expect to be doing during the session?** You will be asked to create a *ggplot2* graph using provided data and then create a custom function to govern its serial application.

**What can I do to prepare?** You can **a)** study demonstrations of the `prep-plot-print` technique on [CCDSS](#) and [CHILD](#) data, **b)** study the behavior of the completed system by exploring various combination of the inputs in a [dedicated script](#), and **c)** walk through the [R script underlying the CCDSS report](#), ensuring its reproducibility on your machine.

Please attempt a solution to the following challenges we can discuss during the session:

- How can we extend the existing function to display ALL provinces sorted in the order of population size by default?
- How can we write a for loop to find the optimal values for the `print_plot_1()` function (i.e. width, height, res, etc.) to print a plot with ALL provinces and ALL age groups on a letter size sheet of paper?

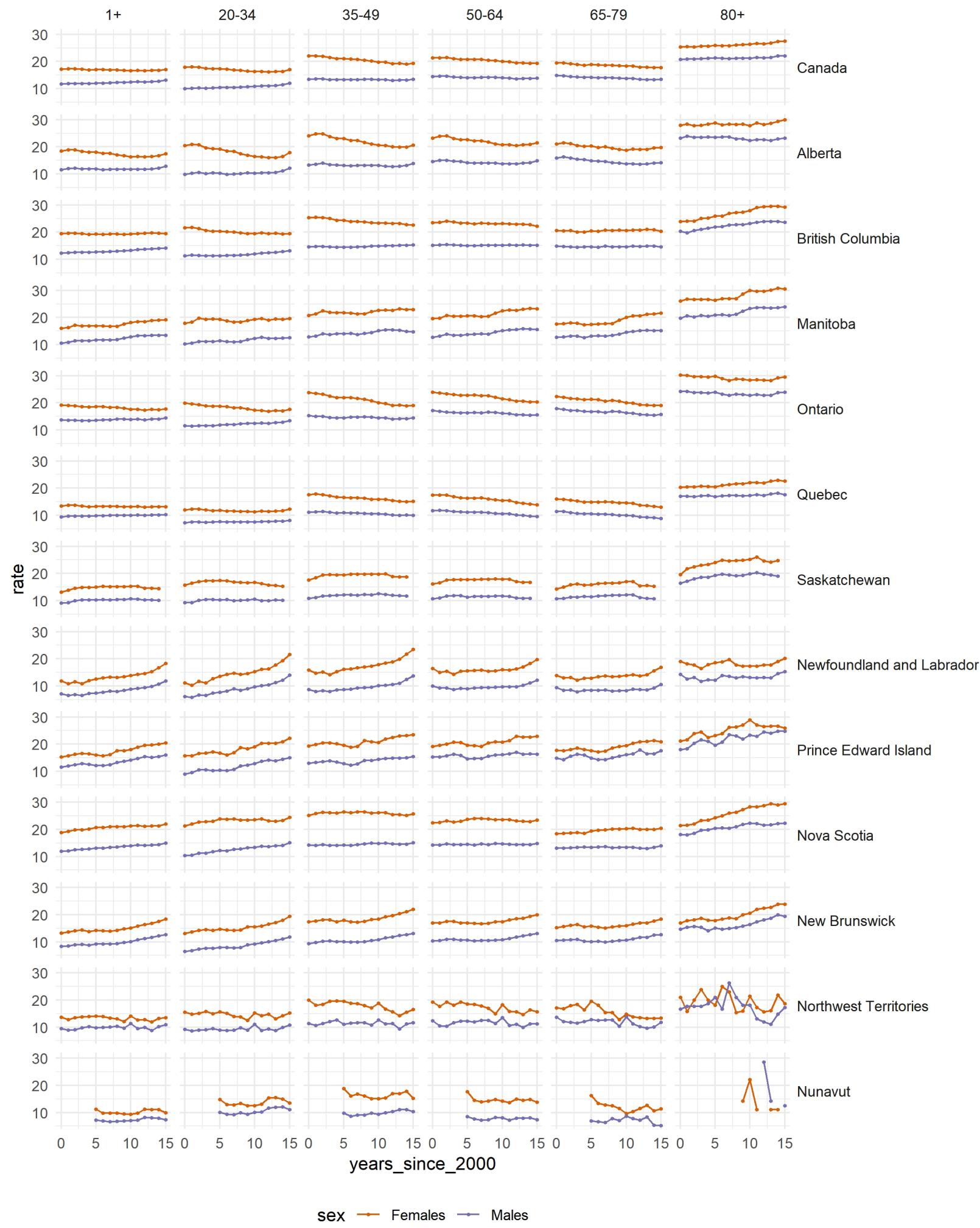
Considering challenges with the teleconferencing during the last meeting we would like to schedule a voluntary **sound check at 10:45**.

Looking forward to thinking with you.

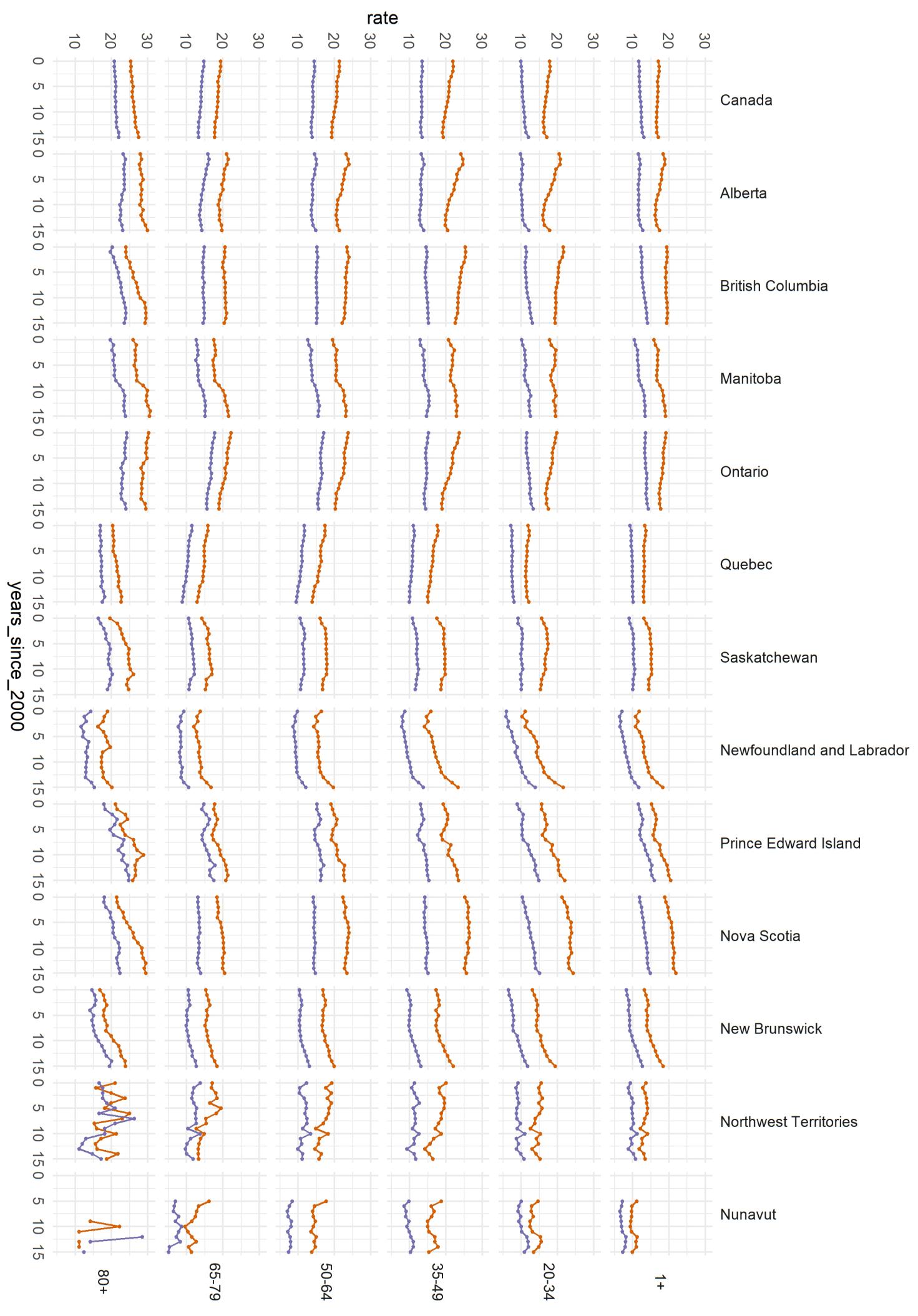
Yours, data science studio



# Crude prevalence of MH service utilization



# Crude prevalence of MH service utilization



# Blueprint of Reproducible Graph

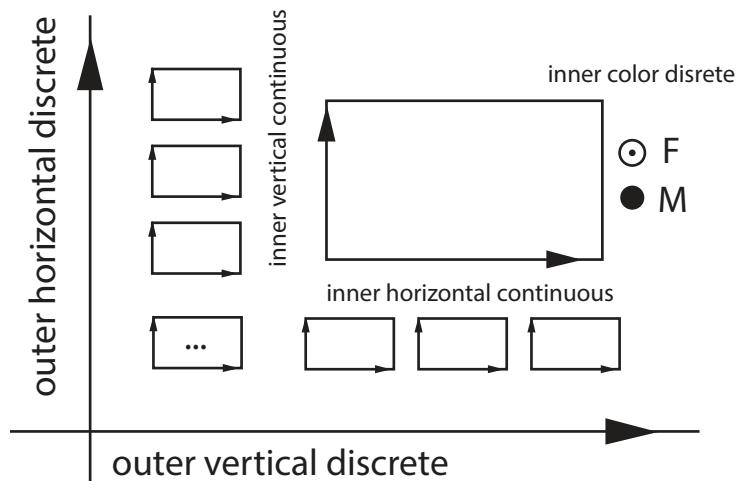
Data pulled from the Canadian Chronic Disease Surveillance System (CCDSS). View data dictionary at <http://infobase.phac-aspc.gc.ca/cubes/ccdss-eng.html>.

## 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

Detailed Table: Use of health services for mental illness (annual), crude prevalence rate, percent						
Geography	Age Group	Sex	Fiscal Year	Rate	95% CI	CV (%)
Canada	1+	Both sexes	2000	14.48	14.46 - 14.49	0.05
Canada	1-19	Both sexes	2000	6.97	6.96 - 6.99	0.14
Canada	20-34	Both sexes	2000	13.91	13.88 - 13.94	0.11
Canada	35-49	Both sexes	2000	17.74	17.71 - 17.77	0.08
Canada	50-64	Both sexes	2000	17.96	17.93 - 18.00	0.10
Canada	65-79	Both sexes	2000	17.34	17.29 - 17.38	0.14
Canada	80+	Both sexes	2000	23.77	23.67 - 23.86	0.20
Canada	1+	Females	2000	17.19	17.17 - 17.21	0.06
Canada	1-19	Females	2000	6.28	6.26 - 6.31	0.20
Canada	20-34	Females	2000	17.88	17.83 - 17.92	0.13

```
Observations: 4,242
Variables: 11
$ area           <chr> "Canada", "Canada", "Canada", "Canada", "Canada", ...
$ condition      <chr> "Use of health services for mood and anxiety disor...
$ age_group      <chr> "1+", "1-19", "20-34", "35-49", "50-64", "65-79", ...
$ sex            <chr> "Both sexes", "Both sexes", "Both sexes", "Both se...
$ year           <dbl> 2000, 2000, 2000, 2000, 2000, 2000, 2000, 20...
$ rate            <dbl> 10.71, 3.04, 11.11, 14.29, 14.41, 13.08, 13.35, 13...
$ rate_cv         <dbl> 0.05, 0.21, 0.12, 0.09, 0.12, 0.16, 0.27, 0.07, 0...
$ rate_95_ci_lower <dbl> 10.70, 3.03, 11.08, 14.26, 14.37, 13.04, 13.28, 13...
$ rate_95_ci_upper <dbl> 10.72, 3.05, 11.13, 14.31, 14.44, 13.12, 13.42, 13...
$ number          <dbl> 3364580, 236410, 72370, 1127780, 735350, 401440, ...
$ population      <dbl> 31412970, 7781830, 6517490, 7893540, 5103900, 3068...
```



# I - DATA space

```

## MEASURE - rate      # crude rate (includes cv, ci_upper, ci_lower)
## 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 gropus + 1 total )
## DESIGN   - sex        ( 2 gropus + 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 - vrtial     - DESIGN  - (age group)

```

