

# Streamflow data analysis with *fasstr* : : CHEAT SHEET



## Getting Started

**fasstr**, Flow Analysis Summary Statistics Tool for R, is a package for tidying, summarizing, performing hydrologic analyses, and visualizing daily streamflow data.

Install **fasstr** from CRAN using :

```
install.packages("fasstr")
```

To use the **station\_number** function argument, a Water Survey of Canada [HYDAT](#) database must be downloaded using:

```
tidyhydat::download_hydat()
```

## Function Usage

**fasstr** functions can be generally categorized into the following groups:

- **Tidying** - preparing data for analyses; `add_*` and `fill_*` functions.
- **Screening** - to look for outliers and missing data; `screen_*` functions.
- **Calculating summary statistics** - long-term, annual, monthly and daily statistics; `calc_*` functions.
- **Visualizing summary statistics** - plotting the various statistics; `plot_*` functions.
- **Computing analyses** - volume frequency analyses and trending; `compute_*` functions.
- **Writing data and plots** - to save your data and plots; `write_*` functions.

## Getting Data

There are two argument options in most functions to choose a data source:

1. **data** Data frame of daily data with dates (YYYY-MM-DD), flow values, and optional groupings. 'data' is the first argument listed to allow for piping (%>%). Arguments for selecting columns in **data** data frame:
  - **dates** Dates column, default 'Date'.
  - **values** Flow values column, default 'Value'.
  - **groups** Groupings columns (optional), default 'STATION\_NUMBER'.
2. **station\_number** Extracts daily data from a HYDAT database using a vector of HYDAT station numbers (ex. '08NM116' or c('08NM116', '08FA002')); downloaded HYDAT required.

Example data with default column names:

STATION_NUMBER	Date	Value
08NM116	1987-04-06	6.230
08NM116	1987-04-07	6.440

## Function Outputs

All outputs from are one, or lists, of the following:

- All **data tables** / data frames produced as **tibbles**.
- All **plots** are produced as lists of **ggplot2** objects.

## Data Tidying

These functions add rows and columns to daily streamflow data frames to prepare for custom analyses.

**fill\_missing\_dates()**

Fills dates with missing data. Other columns filled with NA.

**add\_date\_variables(water\_year\_start = 1)**

Add 'Year', 'Month', 'MonthName', 'WaterYear' 'DayofYear' columns. 'WaterYear' and 'DayofYear' adjust to the selected year start with `water_year_start` argument.

**add\_seasons(seasons\_length)**

Adds column of season identifiers called 'Season' with the length of seasons in months chosen with `seasons_length`.

**add\_rolling\_means(roll\_days, roll\_align)**

Add columns of rolling daily flow means (ex. 7-day means).

**add\_basin\_area(basin\_area)**

Add a basin area column, in square kilometres. See `basin_area` argument on reverse of cheat sheet.

**add\_daily\_volume()**

Add daily volumetric flows, converted from daily mean to cubic metres.

**add\_daily\_yield(basin\_area)**

Add daily yields, converted from daily mean to millimetres based on upstream basin area.

**add\_cumulative\_volume()**

Add daily cumulative volumetric flows on an annual basis, in cubic metres.

**add\_cumulative\_yield()**

Add daily cumulative runoff yield flows on an annual basis, in millimetres based on upstream basin area.

## Data Screening

These functions calculate and plot statistics to screen data for outliers, gaps, and missing dates.

**screen\_flow\_data()**

Calculate annual mean, maximum, minimum, standard deviation, symbol counts, and missing dates.

**plot\_data\_screening()**

Plot annual mean, maximum, minimum, and standard deviation.

**plot\_missing\_dates()**

Plot the number of missing dates for each month and year.

**plot\_flow\_data\_symbols()**

**plot\_flow\_data(plot\_by\_year = FALSE, one\_plot = TRUE)**

Plot the daily mean data set.

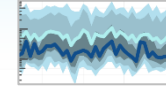
**plot\_annual\_symbols()**

Plot annual symbol counts.

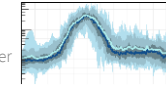
## Basic Summary Statistics

These functions calculate and plot the mean, median, maximum, minimum, and selected percentiles using the 'percentiles' argument. Can select duration of statistics (ex. 7-day) using 'roll\_days' and 'roll\_align' arguments.

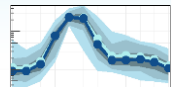
**calc\_annual\_stats()**  
**plot\_annual\_stats()** (lines)  
**plot\_annual\_stats2()** (ribbons)  
Statistics for each year.



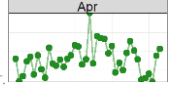
**calc\_daily\_stats()**  
**plot\_daily\_stats(add\_year)**  
Statistics for each day of the year over all years.



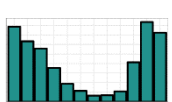
**calc\_longterm\_daily\_stats()**  
**plot\_longterm\_daily\_stats()**  
**calc\_longterm\_monthly\_stats()**  
**plot\_longterm\_monthly\_stats()**  
Statistics for all daily and monthly data for each month over all years.



**calc\_monthly\_stats()**  
**plot\_monthly\_stats()**  
**plot\_monthly\_stats2()**  
Statistics for each month of each year.



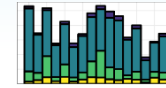
**plot\_monthly\_means()**  
Plot monthly mean statistics over all years. Add long-term mean and % annual means with `percent_MAD` arg.



## Cumulative Statistics

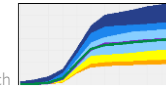
These functions calculate and plot the total flows for years by volume (m³) or by area-based yield (mm) using the 'use\_yield' and 'basin\_area' area arguments.

**calc\_annual\_cumulative\_stats()**  
`include_seasons = TRUE`  
**plot\_annual\_cumulative\_stats()**  
`include_seasons = TRUE`



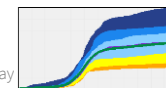
Total annual cumulative discharge for each year (option to include seasonal totals).

**calc\_monthly\_cumulative\_stats()**  
**plot\_monthly\_cumulative\_stats()**  
`add_year`



Cumulative monthly statistics for each month over all years.

**calc\_daily\_cumulative\_stats()**  
**plot\_daily\_cumulative\_stats()**  
`add_year`



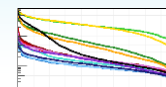
Cumulative daily statistics for each day of year over all years

## Long-term Statistics

These functions calculate and plot various long-term statistics outside of the basic summary statistics.

**plot\_flow\_duration()**

Plot flow durations curves for each month and annually over all years.



**calc\_longterm\_mean(percent\_MAD)**

Calculate the mean discharge over all years with options to include percentages of the long-term mean.

**calc\_longterm\_percentile(percentiles)**

Calculate percentile flow values over all years.

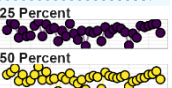
**calc\_flow\_percentile(flow\_value)**

Calculate the percentile rank of a specific flow value from flows over all years.

## Annual Statistics

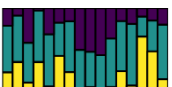
These functions calculate and plot various annual statistics beyond the basic summary statistics.

**calc\_annual\_flow\_timing()**  
`percent_total = c(25,33,3,25,75)`  
**plot\_annual\_flow\_timing()**  
`percent_total = c(25,33,3,25,75)`  
**plot\_annual\_flow\_timing\_year()**



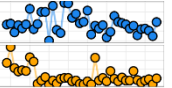
Calculate the day of year when portions of total annual flows have occurred (ex timing of half flows).

**calc\_annual\_normal\_days()**  
`normal_percentiles = c(25,75)`  
**plot\_annual\_normal\_days()**  
`normal_percentiles = c(25,75)`  
**plot\_annual\_normal\_days\_year()**

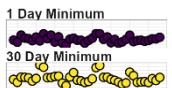


Calculate the number of "normal" days per year and those above or below/below "normal".

**calc\_annual\_extremes()**  
**plot\_annual\_extremes()**  
**plot\_annual\_extremes\_year()**  
Calculate the values and day of occurrence for annual n-day minimum and maximum flow values.



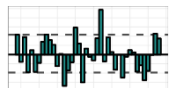
**calc\_annual\_lowflows()**  
`roll_days = c(1,3,7,30)`  
**plot\_annual\_lowflows()**  
`roll_days = c(1,3,7,30)`



Calculate annual minimum flow values and their dates of occurrence.

**plot\_annual\_means()**

Plot annual mean flows with the x-axis centred on the long-term mean



## Arguments and Options

These arguments are used to customize many of the functions. Not all are listed; see function documentation for more specific argument information.

### Date Filtering and Options

**ignore\_missing** Logical value indicating whether dates with missing values should be included in the analysis. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only statistics with no missing dates will return.

**allowed\_missing** Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a statistic 0 (ignore\_missing = FALSE) to 100 percent (ignore\_missing = TRUE). Supersedes ignore\_missing when used, but consistent with usage when not used.

**water\_year\_start** Numeric value indicating the starting month (1 through 12) of years to filter/group data, designated by calendar year in which year ends; default 1.

**start\_years** and **end\_years** Numeric values of the first and last year to consider for analysis. Leave blank to include all years of data provided.

**exclude\_years** Numeric vector of years to exclude from analysis; ex. c(1991:1993, 1995). Leave blank to include all years of data provided.

**complete\_years** Logical value indicating whether to only include years with complete data in analysis. Only in selected analyses; default FALSE.

**months** Numeric vector of months in analysis; default 1:12.

### Data Analysis Options

**roll\_days** Numeric value (or values for some functions) of the number of days to apply a rolling mean; default 1.

**roll\_align** Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations; default 'right'.

**use\_yield** Logical value indicating to use area-based yield, in mm, instead of volumetric for cumulative analysis functions; default FALSE. Requires basin\_area.

**basin\_area** Drainage basin area, in square km, to use when use\_yield = TRUE. Three options: 1) leave blank if column of HYDAT station numbers; 2) single numeric value to apply to all observations; 3) list each basin area for each station c("08NM116" = 795, "08NM242" = 10)

**percentiles** Numeric vector of percentiles to calculate, ex. c(5,25,75,95). Set to NA if none required.

### TibbleOptions

**transpose** Logical value indicating whether to transpose rows and columns of results; default FALSE.

### PlottingOptions

**log\_discharge** Logical value to indicate plotting the discharge axis on a logarithmic scale; default FALSE.

**include\_title** Logical value to indicate adding the group/station number to the plot, if provided.

**add\_year** Numeric value indicating a year of daily flows to add to the daily and long-term statistics plot.

## Writing Functions

These functions help save the outputted objects (tibbles and lists of plots) from the **fasstr** functions.

### write\_flow\_data()

Write a streamflow dataset as a .xlsx, .xls, or .csv file. Can extract and save HYDAT data with this function.

### write\_results(digits = 10)

Write a data frame as a .xlsx, .xls, or .csv file. Can save a data frame and round digits of all numeric columns.

### write\_plots(plots, folder\_name, plot\_filetype, combined\_pdf)

Write plots from a list object into a directory or PDF document. By default will save all plots in a folder. To create a PDF of all plots, set combined\_pdf = TRUE.

### write\_objects(list(list, folder\_name, table\_filetype, plot\_filetype)

Write all tables and plots contained in a list object into a folder. Saves only data frames and ggplot2 objects.

## Annual Trending Analysis

This function computes and plots prewhitened, non-parametric annual trends on streamflow data.

This function calculates prewhitened, non-parametric annual trends using the 'zyp' package. It calculates various annual metrics using the calc\_all\_annual\_stats() function and then calculates and plots the trends. See the zyp package, function documentation, and the trending vignette for more information on the analysis.

### Function

#### compute\_annual\_trends()

Calculate prewhitened nonlinear annual trends on streamflow data.

### Arguments

**zyp\_method** Prewhitening method, either 'zhang' or 'yuepilon'. See [zyp](#) methodology for more information.

**include\_plots** Logical value indicating if annual trending plots should be included. Default TRUE.

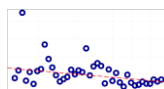
**zyp\_alpha** Numeric value of the significance level (ex. 0.05) of when to plot a trend line. Leave blank for no line.

### Outputs

**\$Annual\_Trends\_Data** A tibble of annual data from the `calc\_all\_annual\_stats()` function used for trending

**\$Annual\_Trends\_Results** A tibble of annual trending results, including significance, confidence intervals, trend values, etc.

**\$Annual\_L\*** A ggplot2 object for each annual statistic trended, with the slope plotted if significance is greater than 'zyp\_alpha' provided.



## Volume Frequency Analyses

These functions compute and plot volume frequency analyses on annual low or high streamflow data.

These functions perform volume frequency analyses by fitting annual minimums or maximums to Log-Pearson Type III or Weibull probability distributions. These functions plot probabilities of data using chosen plotting methods and calculates frequency quantiles (ex. 7Q10) based on fitting data to selected distributions and fitting methods. See function documentation for more information.

### Functions

#### compute\_annual\_frequencies()

Annual frequency analysis from daily streamflow data; calculates minimums or maximums of selected roll\_days.

#### compute\_frequency\_quantile()

Annual frequency analysis from daily streamflow data; calculates minimums or maximums of selected roll\_days and return\_period. Quantile value is returned.

#### compute\_hydat\_peak\_frequencies()

Annual frequency analysis from instantaneous peak data (minimum or maximum) for stations from HYDAT. Data selected using station\_number argument.

#### compute\_frequency\_analysis()

Conduct a frequency analysis with custom data.

### Arguments

**use\_max** Rank data from high to low rather than low to high (for peak analyses); default FALSE.

**use\_log** Log-transform event data before analysis; default FALSE.

**prob\_plot\_positions** Plotting positions used to plot the probabilities; 'weibull' (default), 'hazen', or 'median'.

**prob\_scale\_points** Probabilities to be plotted on the x-axis; default c(.9999, .999, .99, .9, .5, .2, .1, .02, .01, .001, .0001).

**fit\_distr** Distribution used to fit the data; one of Log-Pearson Type III, 'PIII' (default), or Weibull, 'weibull'.

**fit\_dist\_method** Method used to fit the data to the distribution; one of method of moments, 'MOM' (default), or maximum likelihood estimation, 'MLE'.

**fit\_quantiles** Quantiles to be estimated from the fitted distributions (event probabilities); default c(.975, .99, .98, .95, .90, .80, .50, .20, .10, .05, .01).

**plot\_curve** Plot the computed curve on the plot; default TRUE.

### Outputs

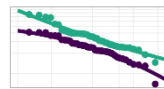
**\$Freq\_Analysis\_Data** Tibble of computed or extracted data used in analysis.

**\$Freq\_Plot\_Data** Tibble of plotting coordinates used in the frequency plot.

**\$Freq\_Plot** ggplot2 object of the frequency plot with return periods and probabilities.

**\$Freq\_Fitting** fitdplus::fitdlist objects of fitted distributions.

**\$Freq\_Fitted\_Quantiles** Tibble of fitted quantiles with probabilities and return periods.



## Computing Full Analyses

These functions calculate a suite of data and plots from many of the **fasstr** functions.

These functions calculate many of the data and plot analyses from the **fasstr** functions, producing tables and plots organized by analysis types. See the function documentation for more information.

### Functions

#### compute\_full\_analysis()

Computes a suite of analyses from **fasstr** functions and produces assorted tables and plots organized in lists grouped by time period and analysis type.

#### write\_full\_analysis()

Writes the compute\_full\_analysis() objects into an Excel workbook and accompanying plot files.

### Arguments

**analyses** Numeric vector of the analyses to include; default is all (1:7). Include those analyses with which statistics are desired: 1: Screening, 2: Long-term, 3: Annual, 4: Monthly, 5: Daily, 6: Trending, 7: Low-flow Frequencies.

### WritingOutputs

**file\_name** Name of Excel workbook, and plots folder if necessary, to save analysis results.

### Outputs

**\$Screening** List of table and plot objects to review and screen data.

**\$Longterm** List of table and plot objects from long-term statistics, including summary statistics and flow duration.

**\$Annual** List of table and plot objects from annual statistics, including summary and cumulative statistics, and other annual metrics.

**\$Monthly** List of table and plot objects from monthly statistics, including summary and cumulative statistics.

**\$Daily** List of table and plot objects from daily statistics, including summary and cumulative statistics.

**\$Trending** List of table and plot objects from an annual trending analysis.

**\$Lowflow\_Frequencies** List of table and plot objects from a low-flow frequency analysis.

### WritingOutputs

Excel '.xlsx' workbook containing tables and plots from selected analyses, and a folder of plots if Daily and/or Trending analyses are computed.

## Further Help

There are several vignettes that provide more information and examples of the many **fasstr** functions:

- [Users Guide](#)
- [Trending Analysis Guide](#)
- [Frequency Analysis Guide](#)
- [Full Analysis Guide](#)
- [fasstr Internal Workflows](#)

See <https://bcgov.github.io/fasstr/> or view them in R using `browseVignettes("fasstr")`

