

Dr. Steffi LaZerte

weathercan

*An R package for accessing
Environment and Climate Change Canada weather data*



@steffilazerte



steffilazerte



steffilazerte.ca

July 2018

A bit of background

Dr. Steffi LaZerte

Biologist (Animal Behaviour)

- BSc University of Toronto
- MSc McGill University
- PhD University of Northern BC



A bit of background

Dr. Steffi LaZerte

Biologist (Animal Behaviour)

- BSc University of Toronto
- MSc McGill University
- PhD University of Northern BC

Working with R since 2007

- Analysis requires a lot of data prep!



A bit of background

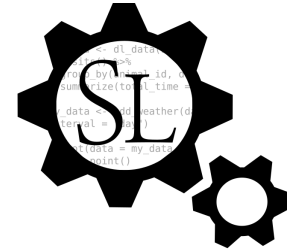
Dr. Steffi LaZerte

Independent biological consultant and R programmer

- Data cleaning, summarizing, and analysis
- Custom R packages and scripts
- R tutoring and workshops

Dr. Steffi LaZerte

Analysis and Data Tools for Science



A bit of background

Dr. Steffi LaZerte

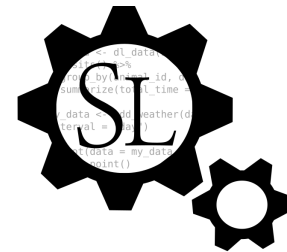
Independent biological consultant and R programmer

- Data cleaning, summarizing, and analysis
- Custom R packages and scripts
- R tutoring and workshops

Solving data problems

Dr. Steffi LaZerte

Analysis and Data Tools for Science



Historical weather data

- Available from Environment and Climate Change Canada (ECCC)
- 1840 to Present
- Hourly, daily, monthly intervals
- > 26,000 stations (past and present)




Historical weather data

- Available from Environment and Climate Change Canada (ECCC)
- 1840 to Present
- Hourly, daily, monthly intervals
- > 26,000 stations (past and present)

Lots of Data!



Accessing data from ECCC website



Government
of Canada

Gouvernement
du Canada

Search Canada.ca

Jobs

Immigration

Travel

Business

Benefits

Health

Taxes

More services

[Home](#) → [Environment and natural resources](#) → [Weather, Climate and Hazard](#) → [Past weather and climate](#)

Historical Data

To determine data availability for a custom location and date, please complete and submit one of the following searches:

Search by Station Name

Search by Province

Search by Proximity

[How to Use - Search by Station Name](#)

Name:

☒ contains ☐ begins with

☒ with data available between:

1840

 to

2017

☐ with data on:

2017

September

25

Display

25

 results per page.

Search

Reset





Government
of Canada

Gouvernement
du Canada

Search Canada.ca



Jobs ▾

Immigration ▾

Travel ▾

Business ▾

Benefits ▾

Health ▾

Taxes ▾

More services ▾

[Home](#) → [Environment and natural resources](#) → [Weather, Climate and Hazard](#) → [Past weather and climate](#) → [Historical Data](#)

Station Results - Historical Data

7 stations found with name containing "Brandon", with data available between 1840 and 2017. Stations are listed in alphabetical order. Confirm the [Data Interval](#) and the date for one of the stations listed and click "GO" to display the historical data.

Station	Prov.	Data Interval	Year	Month	Day	
BRANDON #1 WINTER BAY	MB	<input type="text" value="Daily"/>	<input type="text" value="2002"/>	<input type="text" value="Apr"/>	<input type="text" value="30"/>	<input type="button" value="Go"/>
BRANDON A	MB	<input type="text" value="Hourly"/>	<input type="text" value="2012"/>	<input type="text" value="Dec"/>	<input type="text" value="6"/>	<input type="button" value="Go"/>
BRANDON A	MB	<input type="text" value="Hourly"/>	<input type="text" value="2017"/>	<input type="text" value="Sep"/>	<input type="text" value="25"/>	<input type="button" value="Go"/>
BRANDON CDA	MB	<input type="text" value="Daily"/>	<input type="text" value="2010"/>	<input type="text" value="Mar"/>	<input type="text" value="10"/>	<input type="button" value="Go"/>
BRANDON RCS	MB	<input type="text" value="Hourly"/>	<input type="text" value="2017"/>	<input type="text" value="Sep"/>	<input type="text" value="25"/>	<input type="button" value="Go"/>
BRANDON SOUTH	MB	<input type="text" value="Daily"/>	<input type="text" value="1975"/>	<input type="text" value="Sep"/>	<input type="text" value="30"/>	<input type="button" value="Go"/>





Government
of Canada

Gouvernement
du Canada

Search Canada.ca



Jobs ▾

Immigration ▾

Travel ▾

Business ▾

Benefits ▾

Health ▾

Taxes ▾

More services ▾

[Home](#) → [Environment and natural resources](#) → [Weather, Climate and Hazard](#) → [Past weather and climate](#) → [Historical Data](#)

Hourly Data Report for September 25, 2017

All times are specified in Local Standard Time (LST). Add 1 hour to adjust for Daylight Saving Time where and when it is observed.

BRANDON A
MANITOBA

Latitude:	49°54'36.000" N	Longitude:	99°57'08.000" W	Elevation:	409.30 m
Climate ID:	5010481	WMO ID:	71140	TC ID:	YBR

Related Data

[Almanac Averages & Extremes \(September 25\)](#)

[Daily Data \(September 2017\)](#)

Additional Search Options

[Nearby Stations with Data](#)

[Historical Data Search](#)

Download Data

Hourly Data (September 2017)

☒ CSV ☐ XML

[Download Data](#)

[Get More Data](#)

[← Previous Day](#)

2017

September

25

[Go](#)

Hourly Data Report for September 25, 2017

	Temp °C 	Dew Point Temp °C 	Rel Hum % 	Wind Dir 10's deg	Wind Spd km/h 	Visibility km 	Stn Press kPa 	Hmdx	Wind Chill	Weather
TIME										
00:00 ±	7.6	6.9	95	34	7	24.1	97.12			Cloudy
01:00 ±	8.2	5.7	84	2	13	24.1	97.14			NA





Government
of Canada

Gouvernement
du Canada

Search Canada.ca



Jobs ▾

Immigration ▾

Travel ▾

Business ▾

Benefits ▾

Health ▾

Taxes ▾

More services ▾

[Home](#) → [Environment and natural resources](#) → [Weather, Climate and Hazard](#) → [Past weather and climate](#) → [Historical Data](#)

Hourly Data Report for September 25, 2017

All times are specified in Local Standard Time (LST). Add 1 hour to adjust for Daylight Saving Time where and when it is observed.

BRANDON A
MANITOBA

<u>Latitude:</u>	49°54'36.000" N	<u>Longitude:</u>	99°57'08.000" W	<u>Elevation:</u>	409.30 m
<u>Climate ID:</u>	5010481	<u>WMO ID:</u>	71140	<u>TC ID:</u>	YBR

Related Data

[Almanac Averages & Extremes \(September 25\)](#)

[Daily Data \(September 2017\)](#)

Additional Search Options

[Nearby Stations with Data](#)

[Historical Data Search](#)

Download Data

Hourly Data (September 2017)

☒ CSV ☐ XML

[Download Data](#)

[Get More Data](#)

[← Previous Day](#)

2017

September

25

[Go](#)

Hourly Data Report for September 25, 2017

	<u>Temp</u> °C 	<u>Dew Point Temp</u> °C 	<u>Rel Hum</u> % 	<u>Wind Dir</u> 10's deg	<u>Wind Spd</u> km/h 	<u>Visibility</u> km 	<u>Stn Press</u> kPa 	<u>Hmdx</u>	<u>Wind Chill</u>	<u>Weather</u>
TIME										
00:00 ±	7.6	6.9	95	34	7	24.1	97.12			Cloudy
01:00 ±	8.2	5.7	84	2	13	24.1	97.14			NA



Data good but not ready

"Station Name","BRANDON A"
"Province","MANITOBA"
"Latitude","49.91"
"Longitude","-99.95"
"Elevation","409.30"
"Climate Identifier","5010481"
"WMO Identifier","71140"
"TC Identifier","YBR"
"All times are specified in Local Standard Time (LST). Add 1 hour to adjust for Daylight Saving Time where and when it is observed."

"Legend"
"E","Estimated"
"M","Missing"
"NA","Not Available"
"‡","Partner data that is not subject to review by the National Climate Archives"

"Date/Time","Year","Month","Day","Time","Data Quality","Temp (°C)","Temp Flag","Dew Point Temp (°C)","Dew Point Temp Flag","Rel Hum (%)","Rel Hum Flag","Wind Dir (10s deg)","Wind Dir Flag","Wind Spd (km/h)","Wind Spd Flag","Visibility (km)","Visibility Flag","Stn Press (kPa)","Stn Press Flag","Hmdx","Hmdx Flag","Wind Chill","Wind Chill Flag","Weather"
"2017-09-01 00:00","2017","09","01","00:00","‡","20.8","","17.3","","80","","18","","18","","24.1","","96.21","","26","","","","Cloudy"
"2017-09-01 01:00","2017","09","01","01:00","‡","20.8","","17.2","","80","","17","","27","","24.1","","96.15","","26","","","","NA"
"2017-09-01 02:00","2017","09","01","02:00","‡","20.0","","16.9","","83","","17","","25","","24.1","","96.09","","25","","","","NA"
"2017-09-01 03:00","2017","09","01","03:00","‡","19.4","","16.9","","85","","16","","19","","24.1","","96.07","","","","","Cloudy"
"2017-09-01 04:00","2017","09","01","04:00","‡","19.2","","17.2","","88","","19","","13","","24.1","","96.08","","","","","Thunderstorms"
"2017-09-01 05:00","2017","09","01","05:00","‡","18.8","","17.8","","93","","17","","13","","24.1","","96.05","","","","","NA"
"2017-09-01 06:00","2017","09","01","06:00","‡","18.9","","17.9","","94","","16","","13","","16.1","","96.04","","","","","Mostly Cloudy"
"2017-09-01 07:00","2017","09","01","07:00","‡","18.3","","17.7","","96","","18","","14","","24.1","","96.03","","","","","NA"
"2017-09-01 08:00","2017","09","01","08:00","‡","19.9","","17.8","","88","","20","","18","","24.1","","96.01","","","","","NA"
"2017-09-01 09:00","2017","09","01","09:00","‡","20.6","","18.2","","86","","23","","19","","24.1","","96.02","","27","","","Cloudy"



weathercan: An R package



What's R?



What's R?

An open source, programming language, and software environment



What's R?

An open source, programming language, and software environment

Often used with RStudio IDE 

FileEditCodeViewPlotsSessionBuildDebugProfileToolsHelp

Go to file/function

Addins

example.R

Source on Save

Run

Source

```
1 library(weathercan)
2 library(ggplot2)
3
4 # Get Data
5 w <- weather(50821, start = "2017-09-01")
6
7 # Plot Data
8 ggplot(data = w, aes(x = time, y = temp, colour = station_name)) +
9   theme_bw() +
10  geom_line() +
11  labs(x = "Date", y = "Temperature C", colour = "Station")
12
```

12:1 (Top Level) R Script

EnvironmentHistory

Import Dataset

List

Global Environment

Data

w600 obs. of 35 variables

Console~/

Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> library(weathercan)
> library(ggplot2)
> w <- weather(50821, start = "2017-09-01")
> ggplot(data = w, aes(x = time, y = temp, colour = station_name)) +
+ theme_bw() +
+ geom_line() +
+ labs(x = "Date", y = "Temperature C", colour = "Station")
>

FilesPlotsPackagesHelpViewer

ZoomExportPublish

Temperature C

Date

Station
— BRANDON A

Why use weathercan?



Why use weathercan?

Free

- Free *and* open-source software (FOSS)



Why use weathercan?

Free

- Free *and* open-source software (FOSS)

Fast and Easy

- One line of code to download data from many stations, over many years
- Instantly usable



Why use weathercan?

Free

- Free *and* open-source software (FOSS)

Fast and Easy

- One line of code to download data from many stations, over many years
- Instantly usable

Customizable

- Data is trimmed to start and end times
- You can specify stations, time intervals, timezones, etc.



Why use weathercan?

Reproducible!

- Scripts provide a record of actions
- Make a note of the weathercan version (`packageVersion(weathercan)`)
- Hard to document mouse clicks or website searches



Getting started with **weathercan**

Latest stable version (v0.2.7):

```
install.packages("weathercan")
```

Developmental version (v0.2.7.9000):

```
library(devtools)  
install_github("ropensci/weathercan", build_vignettes = TRUE)
```



Basic usage

Code

```
library(weathercan)  
w <- weather_dl(station_ids = c(50821, 51097), start = "2017-09-01", interval = "hour")
```



Basic usage

Code

```
library(weathercan)
w <- weather_dl(station_ids = c(50821, 51097), start = "2017-09-01", interval = "hour")
```

Output

```
## # A tibble: 15,936 x 29
##   station_name station_id station_operator prov    lat    lon  elev climate_id WMO_id TC_id
## * <chr>          <dbl> <chr>          <fct> <dbl>  <dbl> <dbl> <chr>          <chr>  <chr>
## 1 BRANDON A      50821 NAV Canada    MB    49.9 -100.0  409. 5010481    71140 YBR
## 2 BRANDON A      50821 NAV Canada    MB    49.9 -100.0  409. 5010481    71140 YBR
## 3 BRANDON A      50821 NAV Canada    MB    49.9 -100.0  409. 5010481    71140 YBR
## 4 BRANDON A      50821 NAV Canada    MB    49.9 -100.0  409. 5010481    71140 YBR
## 5 BRANDON A      50821 NAV Canada    MB    49.9 -100.0  409. 5010481    71140 YBR
## # ... with 1.593e+04 more rows, and 19 more variables
```



Basic usage

Code

```
library(weathercan)
```

```
w <- weather_dl(station_ids = c(50821, 51097), start = "2017-09-01", interval = "hour")
```

"day" or "month"

Output

```
## # A tibble: 15,936 x 29
```

```
##   station_name station_id station_operator prov   lat   lon elev climate_id WMO_id TC_id
## * <chr>          <dbl> <chr>              <fct> <dbl> <dbl> <dbl> <chr>      <chr> <chr>
## 1 BRANDON A      50821 NAV Canada        MB    49.9 -100.0 409. 5010481  71140 YBR
## 2 BRANDON A      50821 NAV Canada        MB    49.9 -100.0 409. 5010481  71140 YBR
## 3 BRANDON A      50821 NAV Canada        MB    49.9 -100.0 409. 5010481  71140 YBR
## 4 BRANDON A      50821 NAV Canada        MB    49.9 -100.0 409. 5010481  71140 YBR
## 5 BRANDON A      50821 NAV Canada        MB    49.9 -100.0 409. 5010481  71140 YBR
```

```
## # ... with 1.593e+04 more rows, and 19 more variables
```



Plotting

```
library(ggplot2)
ggplot(data = w, aes(x = time, y = temp, colour = station_name)) +
  theme_bw() +
  geom_line() +
  scale_colour_viridis_d() +
  labs(x = "Date", y = "Temperature C", colour = "Station")
```



And done!

```
library(weathercan)
w <- weather_dl(station_ids = c(50821, 51097), start = "2017-09-01", interval = "hour")

ggplot(data = w, aes(x = time, y = temp, colour = station_name)) +
  theme_bw() +
  geom_line() +
  scale_colour_viridis_d() +
  labs(x = "Date", y = "Temperature C", colour = "Station")
```



Hmmm...

```
library(weathercan)
w <- weather_dl(station_ids = c(50821, 51097), start = "2017-09-01")

ggplot(data = w, aes(x = time, y = temp, colour = station_name)) +
  theme_bw() +
  geom_line() +
  scale_colour_viridis_d() +
  labs(x = "Date", y = "Temperature C", colour = "Station")
```



Hmmm...

```
library(weathercan)
w <- weather_dl(station_ids = c(50821, 51097), start = "2017-09-01")

ggplot(data = w, aes(x = time, y = temp, colour = station_name)) +
  theme_bw() +
  geom_line() +
  scale_colour_viridis_d() +
  labs(x = "Date", y = "Temperature C", colour = "Station")
```

How do we get station ids?



Stations data set

stations



Stations data set

stations

```
## # A tibble: 26,208 x 12
```

##	prov	station_name	station_id	climate_id	WMO_id	TC_id	lat	lon	elev	interval	start	end
##	<fct>	<chr>	<fct>	<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>	<chr>	<int>	<int>
##	1 AB	ABEE AGDM	32232	3010010	71285	XAF	54.3	-113.	664	day	2002	2018
##	2 AB	ABEE AGDM	32232	3010010	71285	XAF	54.3	-113.	664	hour	1990	2018
##	3 AB	ABEE AGDM	32232	3010010	71285	XAF	54.3	-113.	664	month	2002	2007
##	4 AB	ACADIA VALLEY	2030	3020018	<NA>	<NA>	51.2	-110.	732	day	1980	1991
##	5 AB	ACADIA VALLEY	47748	3020035	71048	PAC	51.1	-110.	735	day	2009	2018
##	6 AB	ACADIA VALLEY	2030	3020018	<NA>	<NA>	51.2	-110.	732	hour	NA	NA
##	7 AB	ACADIA VALLEY	47748	3020035	71048	PAC	51.1	-110.	735	hour	2009	2018
##	8 AB	ACADIA VALLEY	2030	3020018	<NA>	<NA>	51.2	-110.	732	month	1980	1991
##	9 AB	ACADIA VALLEY	47748	3020035	71048	PAC	51.1	-110.	735	month	NA	NA
##	10 AB	ACADIA VALLEY C...	2032	3020022	<NA>	<NA>	51.1	-110.	NA	day	1965	1991
##	#	... with 2.62e+04 more rows										



Searching by station name

```
stations_search(name = "Brandon")
```

```
## # A tibble: 15 x 12
```

##	prov	station_name	station_id	climate_id	WMO_id	TC_id	lat	lon	elev	interval	start	end
##	<fct>	<chr>	<fct>	<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>	<chr>	<int>	<int>
##	1 MB	BRANDON #1 WIN...	3474	5010498	<NA>	<NA>	49.8	-100.0	396	day	1987	2002
##	2 MB	BRANDON #1 WIN...	3474	5010498	<NA>	<NA>	49.8	-100.0	396	month	1987	2002
##	3 MB	BRANDON A	3471	5010480	71140	YBR	49.9	-100.0	409.	day	1941	2012
##	4 MB	BRANDON A	3471	5010480	71140	YBR	49.9	-100.0	409.	hour	1958	2012
##	5 MB	BRANDON A	3471	5010480	71140	YBR	49.9	-100.0	409.	month	1941	2012
##	6 MB	BRANDON A	50821	5010481	71140	YBR	49.9	-100.0	409.	day	2012	2018
##	7 MB	BRANDON A	50821	5010481	71140	YBR	49.9	-100.0	409.	hour	2012	2018
##	8 MB	BRANDON CDA	3472	5010485	<NA>	<NA>	49.9	-100.0	363.	day	1890	2010
##	9 MB	BRANDON CDA	3472	5010485	<NA>	<NA>	49.9	-100.0	363.	month	1890	2007
##	10 MB	BRANDON RCS	49909	5010490	71136	PBO	49.9	-100.0	409.	day	2012	2018
##	#	... with 5 more rows										



Searching by station name

```
stations_search(name = "Brandon")
```

```
## # A tibble: 15 x 12
```

##	prov	station_name	station_id	climate_id	WMO_id	TC_id	lat	lon	elev	interval	start	end
##	<fct>	<chr>	<fct>	<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>	<chr>	<int>	<int>
##	1 MB	BRANDON #1 WIN...	3474	5010498	<NA>	<NA>	49.8	-100.0	396	day	1987	2002
##	2 MB	BRANDON #1 WIN...	3474	5010498	<NA>	<NA>	49.8	-100.0	396	month	1987	2002
##	3 MB	BRANDON A	3471	5010480	71140	YBR	49.9	-100.0	409.	day	1941	2012
##	4 MB	BRANDON A	3471	5010480	71140	YBR	49.9	-100.0	409.	hour	1958	2012
##	5 MB	BRANDON A	3471	5010480	71140	YBR	49.9	-100.0	409.	month	1941	2012
##	6 MB	BRANDON A	50821	5010481	71140	YBR	49.9	-100.0	409.	day	2012	2018
##	7 MB	BRANDON A	50821	5010481	71140	YBR	49.9	-100.0	409.	hour	2012	2018
##	8 MB	BRANDON CDA	3472	5010485	<NA>	<NA>	49.9	-100.0	363.	day	1890	2010
##	9 MB	BRANDON CDA	3472	5010485	<NA>	<NA>	49.9	-100.0	363.	month	1890	2007
##	10 MB	BRANDON RCS	49909	5010490	71136	PBO	49.9	-100.0	409.	day	2012	2018

```
## # ... with 5 more rows
```



Searching by station name

```
stations_search(name = "Brandon", interval = "hour")
```



Searching by station name

```
stations_search(name = "Brandon", interval = "hour")
```

```
## # A tibble: 3 x 12
```

##	prov	station_name	station_id	climate_id	WMO_id	TC_id	lat	lon	elev	interval	start	end
##	<fct>	<chr>	<fct>	<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>	<chr>	<int>	<int>
## 1	MB	BRANDON A	3471	5010480	71140	YBR	49.9	-100.0	409.	hour	1958	2012
## 2	MB	BRANDON A	50821	5010481	71140	YBR	49.9	-100.0	409.	hour	2012	2018
## 3	MB	BRANDON RCS	49909	5010490	71136	PBO	49.9	-100.0	409.	hour	2012	2018



Searching by station name (fancy)

```
stations_search(name = "Brandon|Winnipeg", interval = "hour")
```



Searching by station name (fancy)

```
stations_search(name = "Brandon|Winnipeg", interval = "hour")
```

```
## # A tibble: 8 x 12
```

##	prov	station_name	station_id	climate_id	WMO_id	TC_id	lat	lon	elev	interval	start	end
##	<fct>	<chr>	<fct>	<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>	<chr>	<int>	<int>
## 1	MB	BRANDON A	3471	5010480	71140	YBR	49.9	-100.0	409.	hour	1958	2012
## 2	MB	BRANDON A	50821	5010481	71140	YBR	49.9	-100.0	409.	hour	2012	2018
## 3	MB	BRANDON RCS	49909	5010490	71136	PBO	49.9	-100.0	409.	hour	2012	2018
## 4	MB	WINNIPEG A CS	27174	502S001	71849	XWG	49.9	-97.2	239.	hour	2013	2018
## 5	MB	WINNIPEG INTL A	51097	5023227	<NA>	YWG	49.9	-97.2	239.	hour	2013	2018
## 6	MB	WINNIPEG RICHA...	47407	5023226	71852	YWG	49.9	-97.2	239.	hour	2008	2013
## 7	MB	WINNIPEG RICHA...	3698	5023222	71852	YWG	49.9	-97.2	239.	hour	1953	2013
## 8	MB	WINNIPEG THE FO...	28051	5023262	71579	XWN	49.9	-97.1	230	hour	1999	2018



Searching by station name (fancy)

```
stations_search(name = "Brandon|Winnipeg", interval = "hour")
```

```
## # A tibble: 8 x 12
```

##	prov	station_name	station_id	climate_id	WMO_id	TC_id	lat	lon	elev	interval	start	end
##	<fct>	<chr>	<fct>	<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>	<chr>	<int>	<int>
## 1	MB	BRANDON A	3471	5010480	71140	YBR	49.9	-100.0	409.	hour	1958	2012
## 2	MB	BRANDON A	50821	5010481	71140	YBR	49.9	-100.0	409.	hour	2012	2018
## 3	MB	BRANDON RCS	49909	5010490	71136	PBO	49.9	-100.0	409.	hour	2012	2018
## 4	MB	WINNIPEG A CS	27174	502S001	71849	XWG	49.9	-97.2	239.	hour	2013	2018
## 5	MB	WINNIPEG INTL A	51097	5023227	<NA>	YWG	49.9	-97.2	239.	hour	2013	2018
## 6	MB	WINNIPEG RICHA...	47407	5023226	71852	YWG	49.9	-97.2	239.	hour	2008	2013
## 7	MB	WINNIPEG RICHA...	3698	5023222	71852	YWG	49.9	-97.2	239.	hour	1953	2013
## 8	MB	WINNIPEG THE FO...	28051	5023262	71579	XWN	49.9	-97.1	230	hour	1999	2018

Make a note: 50821 (Brandon) and 51097 (Winnipeg)



Alternative: Searching by coordinates

- Search according to location: `c(latitude, longitude)`
- Search within 10km of this location: `dist = 10`

```
stations_search(coords = c(49.84847, -99.95009), dist = 10, interval = "hour")
```



Alternative: Searching by coordinates

- Search according to location: `c(latitude, longitude)`
- Search within 10km of this location: `dist = 10`

```
stations_search(coords = c(49.84847, -99.95009), dist = 10, interval = "hour")
```

```
## # A tibble: 3 x 13
```

##	prov	station_name	station_id	climate_id	WMO_id	TC_id	lat	lon	elev	interval	start	end	distance
##	<fct>	<chr>	<fct>	<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>	<chr>	<int>	<int>	<dbl>
## 1	MB	BRANDON RCS	49909	5010490	71136	PBO	49.9	-100.0	409.	hour	2012	2018	5.73
## 2	MB	BRANDON A	3471	5010480	71140	YBR	49.9	-100.0	409.	hour	1958	2012	6.84
## 3	MB	BRANDON A	50821	5010481	71140	YBR	49.9	-100.0	409.	hour	2012	2018	6.84



Alternative: Searching by coordinates

- Search according to location: `c(latitude, longitude)`
- Search within 10km of this location: `dist = 10`

```
stations_search(coords = c(49.84847, -99.95009), dist = 10, interval = "hour")
```

```
## # A tibble: 3 x 13
```

##	prov	station_name	station_id	climate_id	WMO_id	TC_id	lat	lon	elev	interval	start	end	distance
##	<fct>	<chr>	<fct>	<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>	<chr>	<int>	<int>	<dbl>
## 1	MB	BRANDON RCS	49909	5010490	71136	PBO	49.9	-100.0	409.	hour	2012	2018	5.73
## 2	MB	BRANDON A	3471	5010480	71140	YBR	49.9	-100.0	409.	hour	1958	2012	6.84
## 3	MB	BRANDON A	50821	5010481	71140	YBR	49.9	-100.0	409.	hour	2012	2018	6.84



Alternative: Searching with tidyverse

Note here we're using the stations data frame directly

```
library(tidyverse)
stations %>%
  filter(prov == "MB", start > 2000, interval == "day")
```

```
## # A tibble: 40 x 12
```

##	prov	station_name	station_id	climate_id	WMO_id	TC_id	lat	lon	elev	interval	start	end
##	<fct>	<chr>	<fct>	<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>	<chr>	<int>	<int>
##	1 MB	BRANDON A	50821	5010481	71140	YBR	49.9	-100.0	409.	day	2012	2018
##	2 MB	BRANDON RCS	49909	5010490	71136	PB0	49.9	-100.0	409.	day	2012	2018
##	3 MB	CFB SHILO CCOC	43343	5012673	<NA>	<NA>	49.8	-99.6	372	day	2004	2005
##	4 MB	CHURCHILL	48969	5060595	71618	YYQ	58.7	-94.1	29.3	day	2010	2018
##	5 MB	CHURCHILL AWOS	47408	5060601	<NA>	<NA>	58.7	-94.1	28.7	day	2008	2010
##	6 MB	CHURCHILL CLIM...	44244	5060608	71696	PYQ	58.7	-94.1	28.5	day	2005	2018
##	7 MB	CROSS LAKE JEN...	54658	5060622	71106	PJP	54.5	-98.0	219.	day	2017	2018
##	8 MB	CYPRESS RIVER ...	48128	5010641	71239	POY	49.6	-99.1	374	day	2009	2018
##	9 MB	DAUPHIN	48868	5040689	71036	YDN	51.1	-100.	304.	day	2010	2018



Understanding the data

Flags

For each measurement there is a corresponding 'flag' column

```
w_month <- weather_dl(station_id = 5401, start = "2017-01-01", interval = "month")
names(w_month)  # List all the column names
```

## [1]	"station_name"	"station_id"	"station_operator"
## [4]	"prov"	"lat"	"lon"
## [7]	"elev"	"climate_id"	"WMO_id"
## [10]	"TC_id"	"date"	"year"
## [13]	"month"	"dir_max_gust"	"dir_max_gust_flag"
## [16]	"extr_max_temp"	"extr_max_temp_flag"	"extr_min_temp"
## [19]	"extr_min_temp_flag"	"mean_max_temp"	"mean_max_temp_flag"
## [22]	"mean_min_temp"	"mean_min_temp_flag"	"mean_temp"
## [25]	"mean_temp_flag"	"snow_grnd_last_day"	"snow_grnd_last_day_flag"
## [28]	"spd_max_gust"	"spd_max_gust_flag"	"total_precip"
## [31]	"total_precip_flag"	"total_rain"	"total_rain_flag"
## [34]	"total_snow"	"total_snow_flag"	



Understanding the data

Flags

For each measurement there is a corresponding 'flag' column

```
w_month <- weather_dl(station_id = 5401, start = "2017-01-01", interval = "month")
names(w_month)  # List all the column names
```

## [1]	"station_name"	"station_id"	"station_operator"
## [4]	"prov"	"lat"	"lon"
## [7]	"elev"	"climate_id"	"WMO_id"
## [10]	"TC_id"	"date"	"year"
## [13]	"month"	"dir_max_gust"	"dir_max_gust_flag"
## [16]	"extr_max_temp"	"extr_max_temp_flag"	"extr_min_temp"
## [19]	"extr_min_temp_flag"	"mean_max_temp"	"mean_max_temp_flag"
## [22]	"mean_min_temp"	"mean_min_temp_flag"	"mean_temp"
## [25]	"mean_temp_flag"	"snow_grnd_last_day"	"snow_grnd_last_day_flag"
## [28]	"spd_max_gust"	"spd_max_gust_flag"	"total_precip"
## [31]	"total_precip_flag"	"total_rain"	"total_rain_flag"
## [34]	"total_snow"	"total_snow_flag"	



Understanding the data

Flags

```
select(w_month, date, mean_temp, mean_temp_flag) # Pull out some columns
```

```
## # A tibble: 14 x 3
##   date      mean_temp mean_temp_flag
## * <date>      <dbl> <chr>
## 1 2017-01-01    -4.4 ""
## 2 2017-02-01    -4.3 ""
## 3 2017-03-01    -5.2 ""
## 4 2017-04-01     7.9 ""
## 5 2017-05-01    11.8 E
## 6 2017-06-01    17.5 ""
## 7 2017-07-01    19.3 ""
## 8 2017-08-01    18.4 ""
## 9 2017-09-01    17   ""
## 10 2017-10-01   12.7 ""
## # ... with 4 more rows
```



Understanding the data

Flags

```
vignette("flags", package = "weathercan")
```

code	meaning
E	Estimated
M	Missing
NA	Not Available
‡	Partner data that is not subject to review by the National Climate Archives
A	Accumulated
C	Precipitation occurred, amount uncertain
F	Accumulated and estimated
L	Precipitation may or may not have occurred
N	Temperature missing but known to be > 0



Understanding the data

Units and measurements

```
## # A tibble: 15,936 x 6
##   station_id time                temp temp_dew rel_hum wind_dir
##   *         <dbl> <dtm>                <dbl>   <dbl>   <dbl>   <dbl>
## 1      50821 2017-09-01 00:00:00  20.8    17.3    80     18
## 2      50821 2017-09-01 01:00:00  20.8    17.2    80     17
## 3      50821 2017-09-01 02:00:00  20      16.9    83     17
## 4      50821 2017-09-01 03:00:00  19.4    16.9    85     16
## 5      50821 2017-09-01 04:00:00  19.2    17.2    88     19
## 6      50821 2017-09-01 05:00:00  18.8    17.8    93     17
## 7      50821 2017-09-01 06:00:00  18.9    17.9    94     16
## 8      50821 2017-09-01 07:00:00  18.3    17.7    96     18
## 9      50821 2017-09-01 08:00:00  19.9    17.8    88     20
## 10     50821 2017-09-01 09:00:00  20.6    18.2    86     23
## # ... with 1.593e+04 more rows
```



Understanding the data

Units and measurements

```
vignette("glossary", package = "weathercan")
```

Interval	ECCC Name	Formatted weathercan name	units	Reference
hour	Date/Time	time	ISO date/time	NA
hour	Year	year	year	ECCC glossary page
hour	Month	month	month	ECCC glossary page
hour	Day	day	day	ECCC glossary page
hour	Time	hour	hour	ECCC glossary page
hour	Data Quality	qual	note	ECCC glossary page
hour	Temp (°C)	temp	°C	ECCC glossary page
hour	Temp Flag	temp_flag	note	See the 'flags' vignette
hour	Dew Point Temp (°C)	temp_dew	°C	ECCC glossary page
hour	Dew Point Temp Flag	temp_dew_flag	note	See the 'flags' vignette
hour	Rel Hum (%)	rel_hum	%	ECCC glossary page
hour	Rel Hum Flag	rel_hum_flag	note	See the 'flags' vignette
hour	Wind Dir (10s deg)	wind_dir	10s deg	ECCC glossary page
hour	Wind Dir Flag	wind_dir_flag	note	See the 'flags' vignette



Combining with other data

- Adding weather data to other data sets
- Times don't always line up



Combining with other data

- Adding weather data to other data sets
- Times don't always line up

Sediment data (for example)

```
## # A tibble: 1,392 x 2
##   time          amount
##   <dtm>         <dbl>
## 1 2017-09-01 00:05:34 168.
## 2 2017-09-01 00:35:34 157.
## 3 2017-09-01 01:05:34 176.
## 4 2017-09-01 01:35:34 185.
## 5 2017-09-01 02:05:34 163.
## 6 2017-09-01 02:35:34 169.
## 7 2017-09-01 03:05:34 168.
## # ... with 1,385 more rows
```

Brandon weather data

```
## # A tibble: 7,968 x 3
##   time          temp pressure
##   <dtm>         <dbl>     <dbl>
## 1 2017-09-01 00:00:00 20.8     96.2
## 2 2017-09-01 01:00:00 20.8     96.2
## 3 2017-09-01 02:00:00 20       96.1
## 4 2017-09-01 03:00:00 19.4     96.1
## 5 2017-09-01 04:00:00 19.2     96.1
## 6 2017-09-01 05:00:00 18.8     96.0
## 7 2017-09-01 06:00:00 18.9     96.0
## # ... with 7,961 more rows
```



Interpolating

- Linear interpolation where possible
- Only a single weather station at a time

```
w <- weather_dl(station_ids = 50821, start = "2017-09-01")
```

```
sediment <- weather_interp(data = sediment,  
                           weather = w,  
                           col = c("temp", "pressure"))
```

```
## temp is missing 7 out of 7968 data, interpolation may be less accurate as a result.
```

```
## pressure is missing 7 out of 7968 data, interpolation may be less accurate as a result.
```



Interpolating

- Linear interpolation where possible
- Only a single weather station at a time

```
w <- weather_dl(station_ids = 50821, start = "2017-09-01")
```

```
sediment <- weather_interp(data = sediment,  
                           weather = w,  
                           col = c("temp", "pressure"))
```

```
## temp is missing 7 out of 7968 data, interpolation may be less accurate as a result.
```

```
## pressure is missing 7 out of 7968 data, interpolation may be less accurate as a result.
```

Note messages regarding missing data. These are just a 'heads up', not errors



Interpolating

Sediment & weather data

```
## # A tibble: 1,392 x 4
##   time                amount  temp pressure
##   <dtm>                <dbl> <dbl>    <dbl>
## 1 2017-09-01 00:05:34    168.  20.8     96.2
## 2 2017-09-01 00:35:34    157.  20.8     96.2
## 3 2017-09-01 01:05:34    176.  20.7     96.1
## 4 2017-09-01 01:35:34    185.  20.3     96.1
## 5 2017-09-01 02:05:34    163.  19.9     96.1
## 6 2017-09-01 02:35:34    169.  19.6     96.1
## 7 2017-09-01 03:05:34    168.  19.4     96.1
## # ... with 1,385 more rows
```

Original weather data

```
## # A tibble: 7,968 x 3
##   time                temp pressure
##   * <dtm>                <dbl>    <dbl>
## 1 2017-09-01 00:00:00    20.8     96.2
## 2 2017-09-01 01:00:00    20.8     96.2
## 3 2017-09-01 02:00:00     20     96.1
## 4 2017-09-01 03:00:00    19.4     96.1
## 5 2017-09-01 04:00:00    19.2     96.1
## 6 2017-09-01 05:00:00    18.8     96.0
## 7 2017-09-01 06:00:00    18.9     96.0
## # ... with 7,961 more rows
```



Taking **weathercan** to the next level...



Getting the most out of weathercan

1. Download from multiple stations (spatial)

```
mb <- filter(stations, prov == "MB", start <= "2018", end >= "2018", interval == "hour")
mb
```

```
## # A tibble: 60 x 12
```

##	prov	station_name	station_id	climate_id	WMO_id	TC_id	lat	lon	elev	interval	start	end
##	<fct>	<chr>	<fct>	<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>	<chr>	<int>	<int>
##	1 MB	BERENS RIVER A	53321	5030201	<NA>	YBV	52.4	-97.0	222.	hour	2015	2018
##	2 MB	BERENS RIVER A	53318	5030204	<NA>	YBV	52.4	-97.0	222.	hour	2015	2018
##	3 MB	BERENS RIVER CS	3721	5030203	71158	WCF	52.4	-97.0	222.	hour	1985	2018
##	4 MB	BRANDON A	50821	5010481	71140	YBR	49.9	-100.0	409.	hour	2012	2018
##	5 MB	BRANDON RCS	49909	5010490	71136	PBO	49.9	-100.0	409.	hour	2012	2018
##	6 MB	CARBERRY CS	27741	5010547	71170	WZE	49.9	-99.4	384.	hour	1994	2018
##	7 MB	CARMAN U OF M ...	26857	502I001	71147	WNK	49.5	-98.0	268.	hour	1995	2018
##	8 MB	CHURCHILL	48969	5060595	71618	YYQ	58.7	-94.1	29.3	hour	2010	2018
##	9 MB	CHURCHILL A	50148	5060605	71618	YYQ	58.7	-94.1	29.3	hour	2012	2018
##	10 MB	CHURCHILL CLIM...	44244	5060608	71696	PYQ	58.7	-94.1	28.5	hour	2005	2018

```
## # ... with 50 more rows
```



Getting the most out of weathercan

```
manitoba <- weather_dl(station_ids = mb$station_id, start = "2018-07-01")
```

Big data sets can take time to download: save the output for later

```
write_csv(manitoba, "manitoba_2018-07-01.csv")
```



Getting the most out of weathercan

```
ggplot(manitoba, aes(x = date, y = temp)) +  
  stat_summary(geom = "ribbon", fun.y = median, fun.ymax = max, fun.ymin = min, alpha = 0.5) +  
  stat_summary(geom = "line", fun.y = median) +  
  labs(title = "July 2018 Min/Median/Max Temperatures in Manitoba, Canada",  
        x = "Date", y = "Temperature C")
```



Getting the most out of weathercan



Getting the most out of weathercan

2. Download from multiple years (temporal)

- no 'start' means download until the 'end' date
- no 'end' means download from the 'start' date to the present (or end of the data record)
- no 'start' OR 'end' means download all data

Download ALL data:

```
w <- weather_dl(station_ids = c(3472, 3698), interval = "month")  
write_csv(w, "brandon_historical.csv")
```



Getting the most out of weathercan

```
ggplot(data = w, aes(x = mean_temp, fill = station_name)) +  
  theme_bw() +  
  geom_density(alpha = 0.5) +  
  scale_fill_viridis_d(end = 0.5) +  
  labs(title = "Distribution of Monthly Mean Temperatures",  
        x = "Temperature C", y = "Density", fill = "Station")
```



Recap!



Recap!

1. Load weathercan package

```
library(weathercan)
```



Recap!

1. Load weathercan package

```
library(weathercan)
```

2. Find a station

```
stations_search("Brandon")
```



Recap!

1. Load weathercan package

```
library(weathercan)
```

2. Find a station

```
stations_search("Brandon")
```

3. Download weather

```
w <- weather_dl(station_ids = 50821, start = "2017-09-01")
```



Recap!

1. Load weathercan package

```
library(weathercan)
```

2. Find a station

```
stations_search("Brandon")
```

3. Download weather

```
w <- weather_dl(station_ids = 50821, start = "2017-09-01")
```

4. Add weather data to an existing data set

```
sediment <- weather_interp(data = sediment, weather = w, cols = "temp")
```



Last, but not least

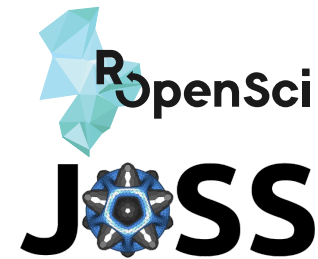
5. Cite weathercan

```
citation("weathercan")
```

```
##  
## To cite 'weathercan' in publications, please use:  
##  
## LaZerte, Stefanie E and Sam Albers (2018). weathercan: Download and format weather data  
## from Environment and Climate Change Canada. The Journal of Open Source Software  
## 3(22):571. doi:10.21105/joss.00571.
```

weathercan is

- Part of the [ropensci project](#)
- Published in [Journal of Open Source Software](#)



Both involve rigorous review of code, best-practices and usability



We invite contributions!

Openly developed on GitHub 

Contribute what you can (**You don't have to be an R programmer!**):

- Ideas / Feature-requests
- Bugs
- Bug-fixes
- Development

: <http://github.com/ropensci/weathercan>



Get help with **weathercan**

Tutorials and Reference: <http://ropensci.github.io/weathercan>

This presentation: <https://ropensci.github.io/Presentations/>

Contact Steffi:



@steffilazerte



steffilazerte



steffilazerte.ca



Get help with **weathercan**

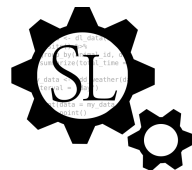
Tutorials and Reference: <http://ropensci.github.io/weathercan>

This presentation: <https://ropensci.github.io/Presentations/>

Contact Steffi:  @steffilazerte  steffilazerte  steffilazerte.ca

Thanks!

Dr. Steffi LaZerte



Analysis and Data Tools for Science

Slides created with the R package [xaringan](#), using [remark.js](#), [knitr](#), and [R Markdown](#)

Compiled on 2018-07-30 with weathercan v0.2.7

