

litteR User Manual

Dennis Walvoort Wageningen Environmental Research, The Netherlands

Willem van LOON Rijkswaterstaat, The Netherlands

September 2021



- 1 Introduction
- 2 Loading the **litteR**-package
- 3 User interface
 - 3.1 Create a new project
 - 3.2 Perform litter analysis
- 4 Input
 - 4.1 Type file
 - 4.2 Data file
 - 4.3 Settings file
 - 4.4 Data Quality Control
- 5 Output
 - 5.1 Report
 - 5.1.1 Settings
 - 5.1.2 Data Quality Control
 - 5.1.3 Outlier analysis
 - 5.1.4 Descriptive statistics
 - 5.1.5 Regional descriptive statistics
 - 5.1.6 Trend analysis
 - 5.1.7 Regional trend analysis
 - 5.2 Statistical summary file
 - 5.3 Log-file
- 6 Troubleshooting
- 7 References
- 8 Appendix

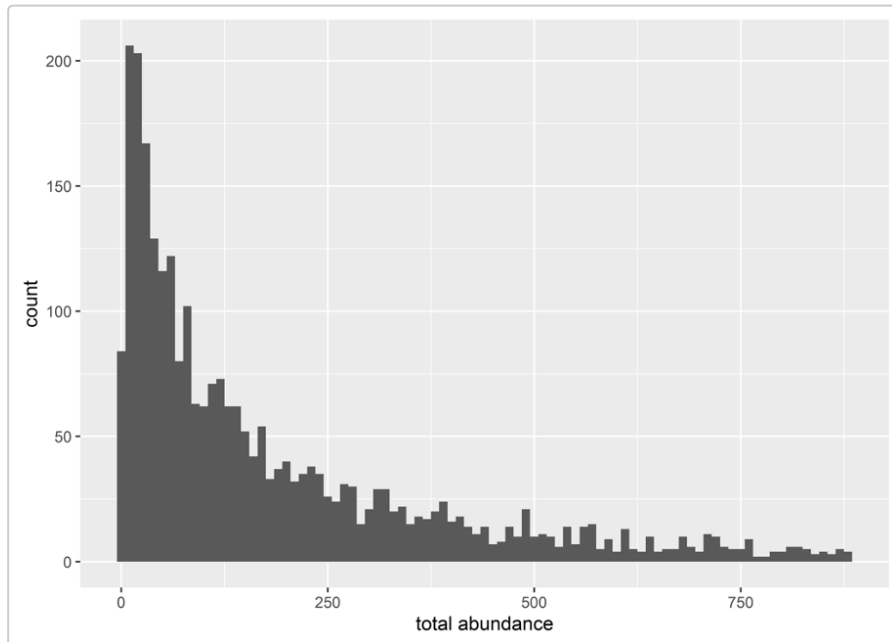
1 Introduction

litteR is a user-friendly tool for analyzing litter data (e.g., beach litter data). The current version (0.9.0) contains routines for:

- data quality control
- outlier analysis
- descriptive statistics, and
- trend analysis

The focus of this version of **litteR** is to provide a user-friendly, flexible, robust, transparent, and relatively simple tool for litter analysis. Although **litteR** is distributed as an [R](#)-package, experience with R is not required. If you need more information on how to install R, RStudio, and **litteR**, please consult our [installation guide](#).

Litter data are [count data](#). As has been illustrated in the histogram below (copied with permission from Hanke *et al.*, 2019), litter data generally have [skewed](#) distributions. All procedures in **litteR** are based on robust statistical methods. They do not require distributional assumptions and are relatively robust for outliers.



This user guide consists of two parts. In the first part, the user interface is described, the second part provides details on the technicalities.

For applications with (a previous version of) **litteR** see Schulz *et al.* (2019). **litteR** is the successor of the Litter Analyst software (Schulz *et al.*, 2017).

2 Loading the litteR-package

Before **litteR** can be used, it should be installed or updated in case you installed **litteR** before. See our [installation guide](#) for details.

You need to *install* **litteR** only once, but you need to *load* this package each time you start RStudio.

The **litterR**-package should be loaded in RStudio before you can use it. This can be done by running the following code in the **R**-console or the [RStudio](#)-console:

```
library(litterR)
```

A startup message appears that gives some essential instructions to start using **litterR**.

3 User interface

3.1 Create a new project

The easiest way to start working with **litterR** is to create an empty project directory. This directory can be filled with example and reference files by running:

```
create_litter_project("d:/work/litter-projects/beach-litter")
```

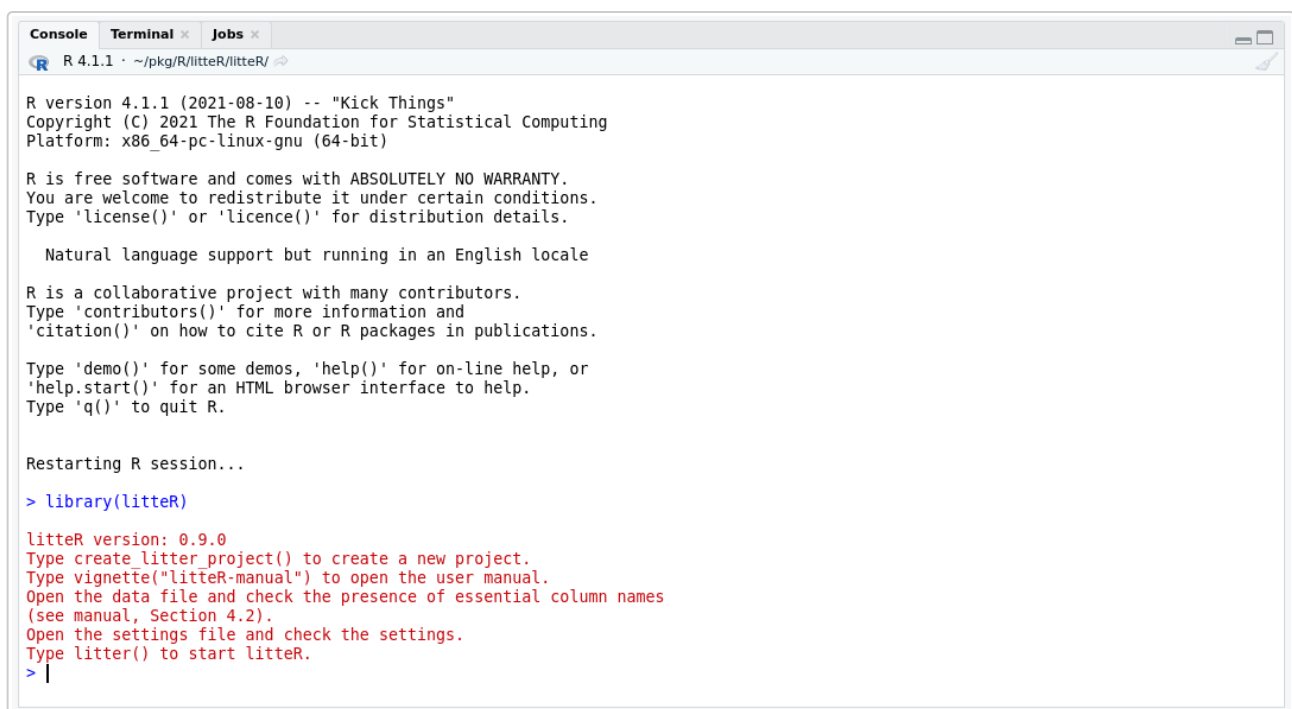
in the [RStudio](#)-console. For more information on how to obtain and use RStudio, consult its [website](#) or read our [installation guide](#).

The argument of function `create_litter_project` (*i.e.*, the quoted part in parentheses) is an existing work directory on your computer. This can be any valid directory name with sufficient user privileges. Note for MS-Windows users: R requires forward slashes!

It is also possible to run `create_litter_project()` without an argument. In that case, a simple graphical user interface pops up for interactive directory selection.

3.2 Perform litter analysis

litterR can be started typing `litter()` in the [RStudio](#) console (see the figure below).



```
Console Terminal x Jobs x
R 4.1.1 · ~/pkg/R/litterR/litterR ↗

R version 4.1.1 (2021-08-10) -- "Kick Things"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

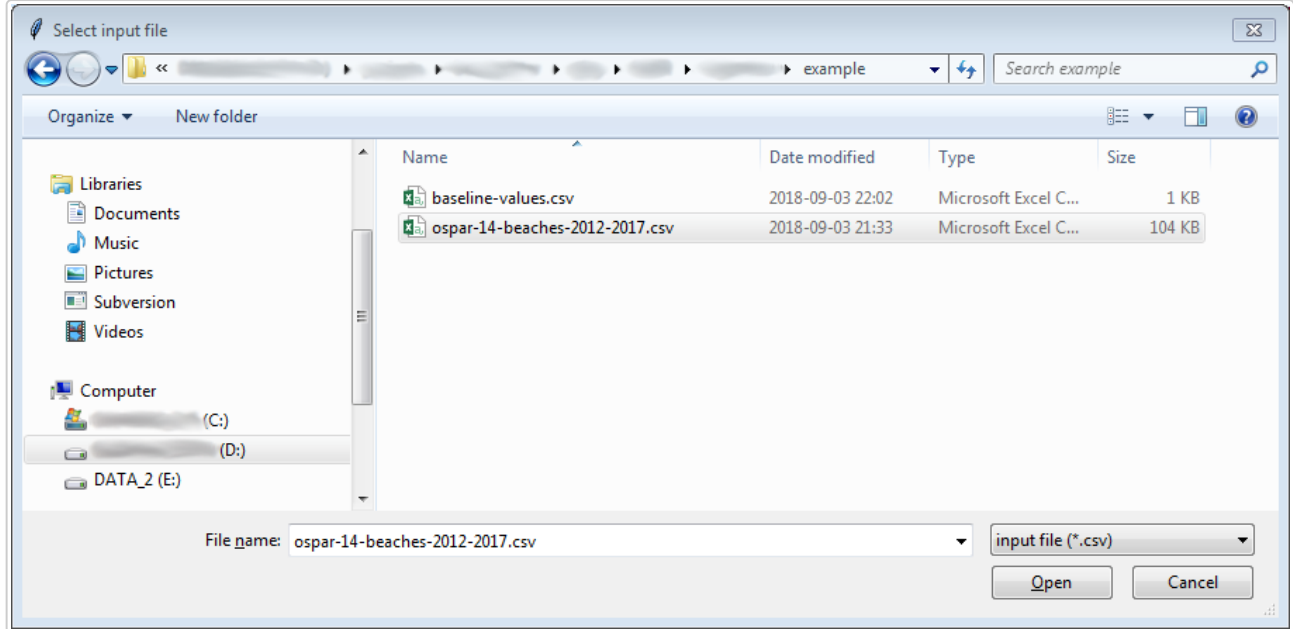
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Restarting R session...

> library(litterR)

litterR version: 0.9.0
Type create_litter_project() to create a new project.
Type vignette("litterR-manual") to open the user manual.
Open the data file and check the presence of essential column names
(see manual, Section 4.2).
Open the settings file and check the settings.
Type litter() to start litterR.
> |
```

After entering `litter()`, a simple graphical user interface pops up for file selection. An example of a file selection dialogue is given below.



4 Input

litteR needs three input files:

1. a type file, which contains all litter types and litter groups that are allowed to use;
2. a data file, which contains litter counts for each monitoring event.
3. a settings file, which contains all settings needed to perform a **litteR** run;

These input files are described below.

4.1 Type file

The type file contains a list of all litter types that are allowed to use in the [data file](#). It also indicates to which litter group each litter type belongs. Two example files, named 'types-ospar.csv' and 'types-ospar-tc-sup-fish-plastic.csv' are automatically generated when using the `create_litter_project`-function, as described [earlier](#) in this tutorial. A type file assigns each litter type (`type_name`) to one or more litter groups. The first 10 rows of 'types-ospar-tc-sup-fish-plastic.csv' are given in the table below.

type_name	included	SUP	FISH	PLASTIC
Plastic: Yokes [1]	x	x		x
Plastic: Bags [2]	x	x		x
Plastic: Small_bags [3]	x	x		x

type_name	included	SUP	FISH	PLASTIC
Plastic: Bag_ends [112]	x	x		x
Plastic: Drinks [4]	x	x		x
Plastic: Cleaner [5]	x	x		x
Plastic: Food [6]	x	x		x
Plastic: Toiletries [7]	x	x		x
Plastic: Oil_small [8]	x			x
Plastic: Oil_large [9]	x			x

The following columns are in this table:

- `type_name`. This column is required and gives all litter types that are allowed in the data file. Litter types given in this column need to be unique;
- `included`: This column indicates whether a type specified in column `type_name` will be used in the analysis or not. Only `type_names` that are included in the analysis will contribute to the total litter count (TC).
- `SUP`, `FISH`, `PLASTIC`, etc.: these columns give the definition of each litter group. In the example above three groups are given: 'single use plastics' (SUP), 'fisheries related litter' (FISH), and 'plastics' (PLASTIC). A cross (x) indicates that a litter type in `type_name` is a member of a litter group or not. A cross (x) means 'a member', an empty cell means 'not a member'.

The user may use one of the provided type files as a template for his own type file. **litteR** will use the type file that has been specified in the [settings-file](#).

litteR performs regional aggregation at the group level. In order to perform regional aggregation at the type level (the columns in the [data file](#)), a group with only one or a few litter types of interest can be constructed in the type file, and then regionally aggregated by running **litteR**.

4.2 Data file

litteR supports a simple and flexible data format. It is similar to the [OSPAR-format](#). The data are stored in so called [wide format](#): each row refers to a single survey, each column to a single litter type or metadata. The table below gives an example of a small part (*i.e.*, the upper left corner) of a data file.

location_code	date	Plastic: Yokes [1]	Plastic: Bags [2]	Plastic...
NL001	2012-01-27	0	3	...
NL001	2012-04-20	0	8	...
NL001	2012-07-22	0	1	...
NL001	2012-10-19	0	2	...
NL001	2013-02-19	0	24	...
:	:	:	:	:

The columns `location_code` and `date` are always required and define unique records (rows) with litter survey data for a specific date and location (e.g., a specific beach, or a location along a river). **litteR** will use these data to estimate statistics (as the median and trend) for each `location_code`.

Column `location_code` may contain location codes (as in the example above), but also full names like 'Bergen', 'Noordwijk', and 'La Grève des Courses'. Full names may be more clear when interpreting the results.

The `date` column gives the monitoring date in [ISO](#) format, i.e., YYYY-mm-dd (for example 2021-09-21, to indicate 21 September 2021). For convenience, the [OSPAR-format](#) (dd/mm/YYYY) is currently also supported (for example 21/09/2021, to indicate 21 September 2021).

Columns `Plastic: Yokes [1]`, `Plastic: Bags [2]`, ... contain the counts for specific litter types. Each litter type (column name) should be listed in the [litter type file](#). Only litter types in the [litter type file](#) are valid column names. All column names that are not valid litter types are considered as optional metadata. These columns are ignored by **litteR** and do not affect the results.

There is one exception: the column `region_code` is optional and should be available when the locations (in column `location_code`) also need to be spatially aggregated. Each `region_code` is related to one or more `location_code`(s) that are part of that region.

In the data file below, one `region_code` (NL) is provided for all locations in `location_code`. Therefore, **litteR** will spatially aggregate the results for all locations (NL001 ... NL004) within the specified region (NL).

<code>region_code</code>	<code>location_code</code>	<code>date</code>	<code>Plastic: Yokes [1]</code>	<code>Plastic: Bags [2]</code>	<code>Plastic...</code>
NL	NL001	2012-01-27	0	3	...
NL	NL001	2012-04-20	0	8	...
NL	NL001	2012-07-22	0	1	...
:	:	:	:	:	:
NL	NL004	2017-04-14	0	0	...
NL	NL004	2017-07-11	1	0	...
NL	NL004	2017-10-18	0	1	...

A data file can be constructed easily from existing litter files. As an example consider the [OSPAR-format](#) below:

<code>Beach ID</code>	<code>Beach name</code>	<code>Country</code>	<code>Survey date</code>	<code>Plastic: Yokes [1]</code>	<code>Plastic: Bags [2]</code>	<code>Plastic...</code>
NL001	Bergen	Netherlands	2012-01-27	0	3	...
NL001	Bergen	Netherlands	2012-04-20	0	8	...
NL001	Bergen	Netherlands	2012-07-22	0	1	...
:	:	:	:	:	:	:

One can simply rename existing columns to the names required by **litteR**. This can be done with a spreadsheet program or a text editor. For instance, renaming `Beach ID`, `Country` and `Survey date` to respectively `location_code`, `region_code`, and `date` gives the following valid **litteR** format:

location_code	Beach name	region_code	date	Plastic: Yokes [1]	Plastic: Bags [2]	Plastic...
NL001	Bergen	Netherlands	2012-01-27	0	3	...
NL001	Bergen	Netherlands	2012-04-20	0	8	...
NL001	Bergen	Netherlands	2012-07-22	0	1	...
:	:	:	:	:	:	:

Column `Beach name` is not recognized by **litteR**, and is therefore ignored.

As an alternative, one may also add new columns with valid **litteR** names to the data file and fill them with the contents of existing columns. See the example below:

region_code	location_code	date	Beach ID	Beach name	Country	Survey date	Plastic...
Netherlands	Bergen	27/01/2012	NL001	Bergen	Netherlands	27/01/2012	...
Netherlands	Bergen	20/04/2012	NL001	Bergen	Netherlands	20/04/2012	...
Netherlands	Bergen	22/07/2012	NL001	Bergen	Netherlands	22/07/2012	...
:	:	:	:	:	:	:	:

This can be done quite easily with a spreadsheet program. The original columns of the [OSPAR-format](#) (`Beach ID`, `Beach name`, `Country`, and `Survey date`) are ignored by **litteR**.

It is advised to use `region_codes` and `location_codes` that are easily recognized by the user. For instance, in the example above, `location_code` 'Bergen' is easier to interpret than `location_code` 'NL001'. Obviously, this choice does not affect the **litteR**-results.

4.3 Settings file

The settings file contains all settings needed to run **litteR**. An example of the contents of a settings file is given in the figure below:

```
# litteR settings file

# Period to analyse (YYYY-mm-dd)
date_min: 2012-01-01
date_max: 2017-12-31

# Percentage of total count to analyse (0 < percentage_total_count <= 100)
percentage_total_count: 80

# Data file.
# Note: the datafile must be in the same path as the settings file
# Note: the file extension should be .csv
file_data: beach-litter-nl-2012-2017.csv
```

```

# Type file. Defines the types and their groups
file_types: types-ospar.csv

# Select trend figures to plot in the report
# Note: this can be zero, one, or more than one location_code, region_code,
# group_code, and/or type_name
location_code: ["NL001", "NL004"]
region_code: ["NL"]
group_code: ["TC", "SUP", "FISH"]
type_name: ["Plastic: Bags [2]"]

# figure quality (high or low)
figure_quality: high

# cutoff value vertical axis with litter counts (percentage)
cutoff_count_axis: 100

```

The settings-file contains the following entries:

- `date_min` and `date_max`, the first and final date of the period to analyze. Dates should be given in [ISO format](#), *i.e.*, YYYY-mm-dd (for example 2021-09-21, to indicate 21 September 2021);
- `percentage_total_count`: the percentage of the total count used to estimate statistics. See the section on [descriptive statistics](#) for more information;
- `file_data`: name of the [data file](#) (including its path, *e.g.*, `c:/my-litter-directory/my-litter-data.csv`);
- `file_types`: name of the [type file](#) (including its path, *e.g.*, `c:/my-litter-directory/types-ospar.csv`);
- `location_code`: name(s) of the location(s) to plot. These should exist in column `location_code` in the [data file](#). As mentioned in the [previous section](#), `location_codes` should be readily interpretable for the user, as these codes are also used in the **litteR**-results (tables and plots);
- `region_code`: name(s) of the region(s) to plot. These should exist in column `region_code` in the [data file](#);
- `group_code`: name(s) of group(s) to plot. Litter groups should be available as column names in the [type file](#);
- `type_name`: name(s) of type(s) to plot; Type names should be available in the [type file](#) and [data file](#);
- `figure_quality`: quality of the plots in the report, either `high` or `low`.
- `cutoff_count_axis`: optional cutoff value as a percentage of the vertical count axis in trend plots. A cutoff value is useful to improve the readability of a plot in case of a few very high litter counts.

4.4 Data Quality Control

All input files are validated by **litteR**. The following validation rules apply:

1. all required columns (see above) should be available;
2. the date format should be valid, *i.e.* preferably YYYY-mm-dd ([ISO](#)). For convenience, the OSPAR-date format (dd/mm/YYYY) is currently also supported;
3. litter type names should be specified in the [type file](#);
4. litter counts in the [data file](#) are [natural numbers](#) (ISO 80000-2). However, in some cases data files contain [real numbers](#) due to preprocessing. Think about normalizing survey lengths to a common length. Although **litteR** prefers natural numbers, real numbers are also allowed for convenience. In case non-natural numbers are found, **litteR** will give a warning but will continue the analysis.;
5. all records should be unique, duplicated records will be removed with a warning;
6. all cells should be filled with the appropriate data type (numbers, text or dates).

7. the data file should be a comma-separated values file (CSV), *i.e.*, a text file where the columns are separated by commas (,) and not by spaces, semicolons (;) or tabs.

5 Output

litteR produces three output files:

1. a report, containing all analysis results
2. a CSV-file, containing all beach litter statistics
3. a log-file, containing all log data.

For convenience, all input and output files are stored as a snapshot in a directory with names like `litteR-results-20210904T221809`, where the final part of the name is a [timestamp](#).

5.1 Report

litteR produces an HTML-report that can best be viewed with modern web browsers like Mozilla FireFox, Google Chrome, or Safari. These browsers are freely available from the internet.

The filename of each report starts with 'litter-results', followed by a timestamp: YYYYmmddTHHMMSS and the extension html. For example: `litteR-results-20210904T221809.html`

This section briefly describes each section in the HTML-report

5.1.1 Settings

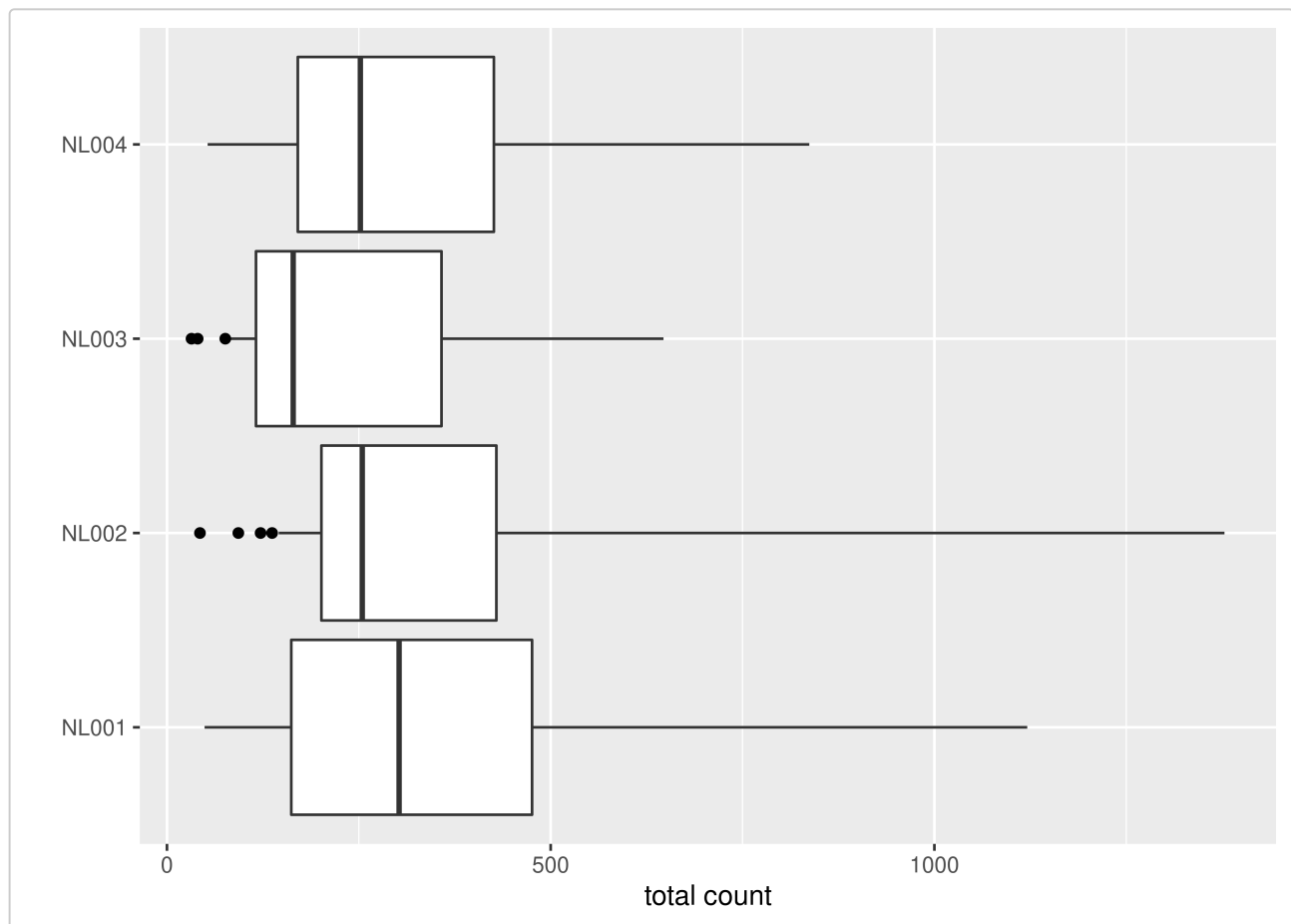
This section gives a summary of the settings in the [settings file](#).

5.1.2 Data Quality Control

In this section (potential) problems in the input files are reported. These problems are also stored in the [log file](#).

5.1.3 Outlier analysis

For each `location_code` in the [data file](#), [adjusted boxplots](#) are given of the total count for the detection of [outliers](#). Outliers are given as dots (if any) in adjusted box-and-whisker plots. [Adjusted boxplots](#) are more suitable for outlier detection in case of skewed distributions than traditional box plots. An example of these box-and-whisker plots are given below.



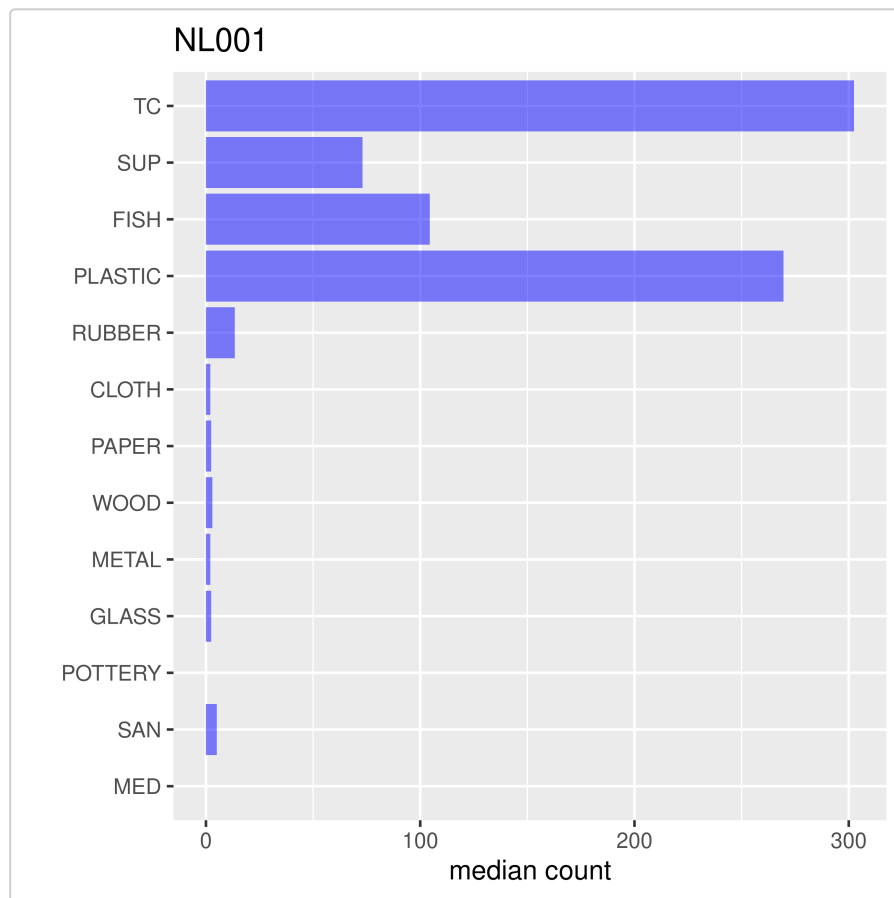
5.1.4 Descriptive statistics

For each `location_code` and group/type name, the following statistics are estimated:

- mean count, *i.e.*, [the arithmetic mean](#) of the counts for each litter type;
- median count, *i.e.*, the [median](#) of the counts for each litter type;
- relative count: the contribution of each litter type to the total count of litter types (%);
- [coefficient of variation](#) (CV): the ratio of the standard deviation to the mean of the counts for each litter type (%);
- ratio of [MAD](#) and [median](#) (RMAD, %);
- number of surveys;
- [Theil-Sen slope](#): a robust non-parametric estimator of slope (litter counts / year);
- [p-value](#): the p-value associated with the one-tailed [Mann-Kendall test](#) to test the null hypothesis of
 - no monotonically *increasing* trend in case the Theil-Sen slope is greater than zero;
 - no monotonically *decreasing* trend in case the Theil-Sen slope is smaller than zero;

These statistics will be estimated for all litter types with the greatest counts making up a percentage of the total count and for all litter groups. This percentage is given as `percentage_total_count` in the [settings file](#).

The descriptive statistics for the litter types and groups are stored in a CSV-file with a name starting with `litter-results` and ending with a [timestamp](#). The statistics for litter groups are also printed as a table and shown as bar plots in the report: one plot for each `location_code` [column of the data file](#). An example is given in the figure below. If you want other groups, or only a subset of groups, you should modify the [type file](#).



5.1.5 Regional descriptive statistics

When the [data file](#) contains column `region_code`, the data for the `location_codes` in that region are spatially aggregated in a stepwise fashion:

1. First a (summary) statistic is estimated for each location (`location_code`) within that region (`region_code`).
2. Next, the same statistic is computed for the results in step 1 to obtain the regional statistic for that region.

Note that these statistics are so called intra-block statistics, *i.e.*, data from individual `location_codes` are not merged.

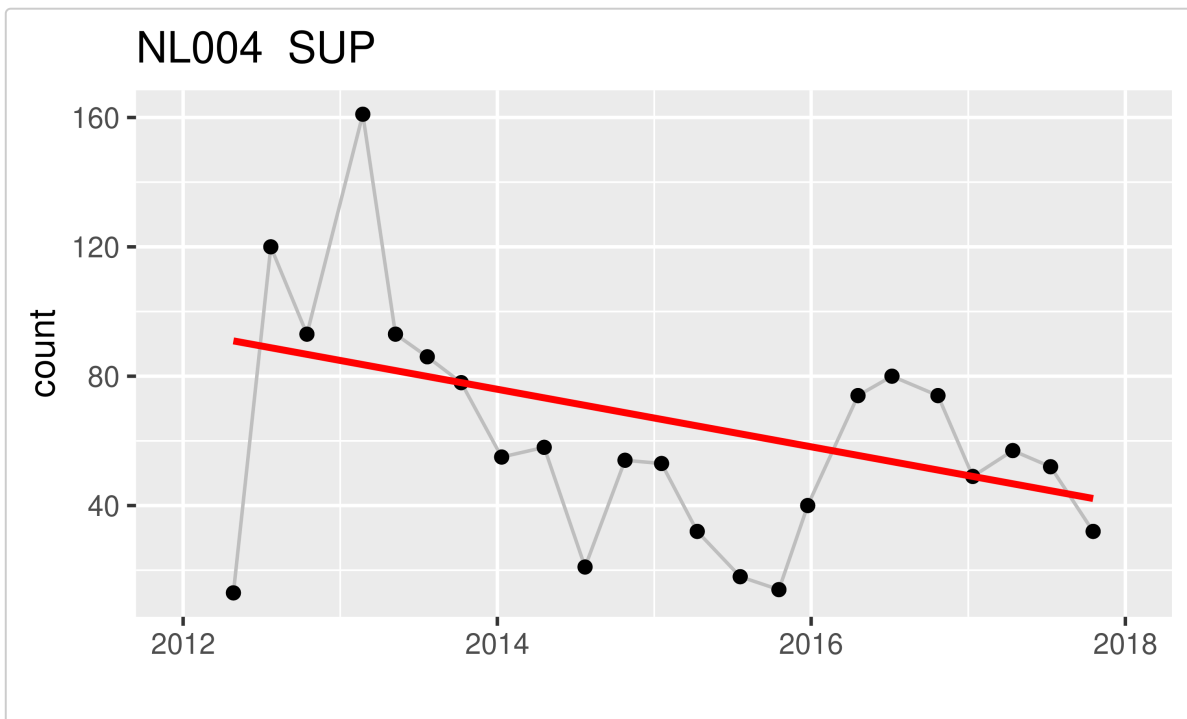
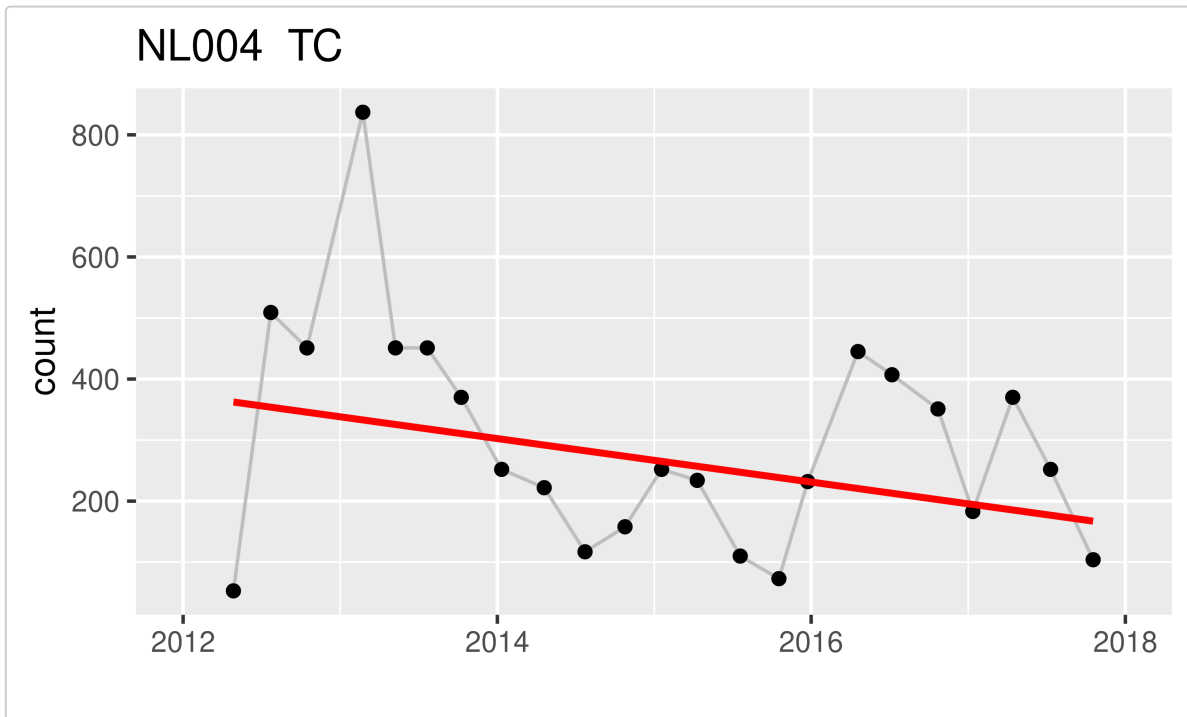
The summary statistics are:

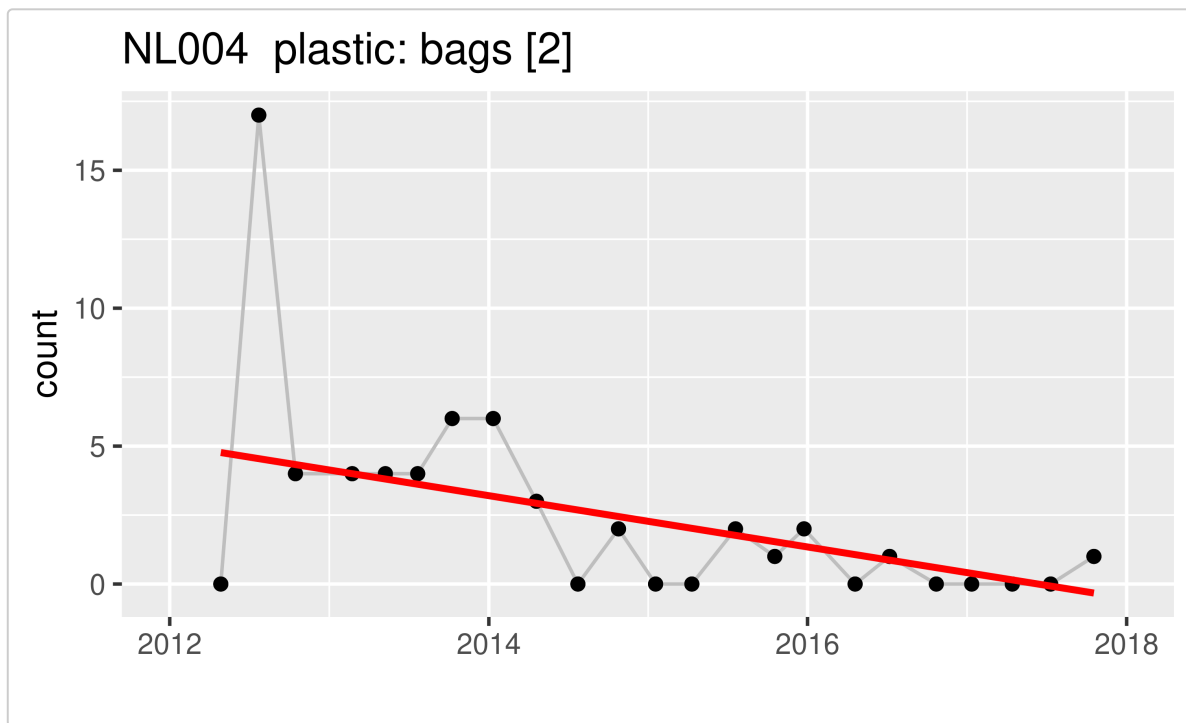
- regional mean: the [mean](#) of the means of the individual locations (`location_code`) within a region (`region_code`) for each litter group;
- regional median: the [median](#) of the medians of the individual locations (`location_code`) within a region (`region_code`) for each litter group;
- regional slope: the median of the Theil-Sen slopes of the individual locations (`location_code`) within a region (`region_code`) for each litter group. Data from different locations have not been mixed in the computation of the Theil-Sen slopes. This method is similar to the one in Gilbert (1987) except that in our procedure all locations within a region contribute equally to the regional trend.
- `p_value`: the p-values for each regional trend (`slope`) are computed by means of the expressions given in [Van Belle & Hughes, 1984](#) (Eqs. 2 and 7) and [Gilbert, 1987](#) (Eqs. 17.1 - 17.5).

5.1.6 Trend analysis

For each `location_code`, and the type names and group codes specified in the [settings file](#), trends are estimated by means of the [Theil-Sen slope estimator](#): a robust non-parametric estimator of slope (counts / year). The significance of the estimated slopes is tested by means of the [Mann-Kendall test](#). The Mann-Kendall test is a non-parametric test and as such does not make distributional assumptions on the data.

The figure below gives examples of trend plots for total count (TC), single use plastics (SUP), and plastic bags at the beach of Terschelling (The Netherlands). In each plot, the black dots are the observations, the thin gray line segments connect the dots and guide the eye, and the red line is the Theil-Sen slope.





5.1.7 Regional trend analysis

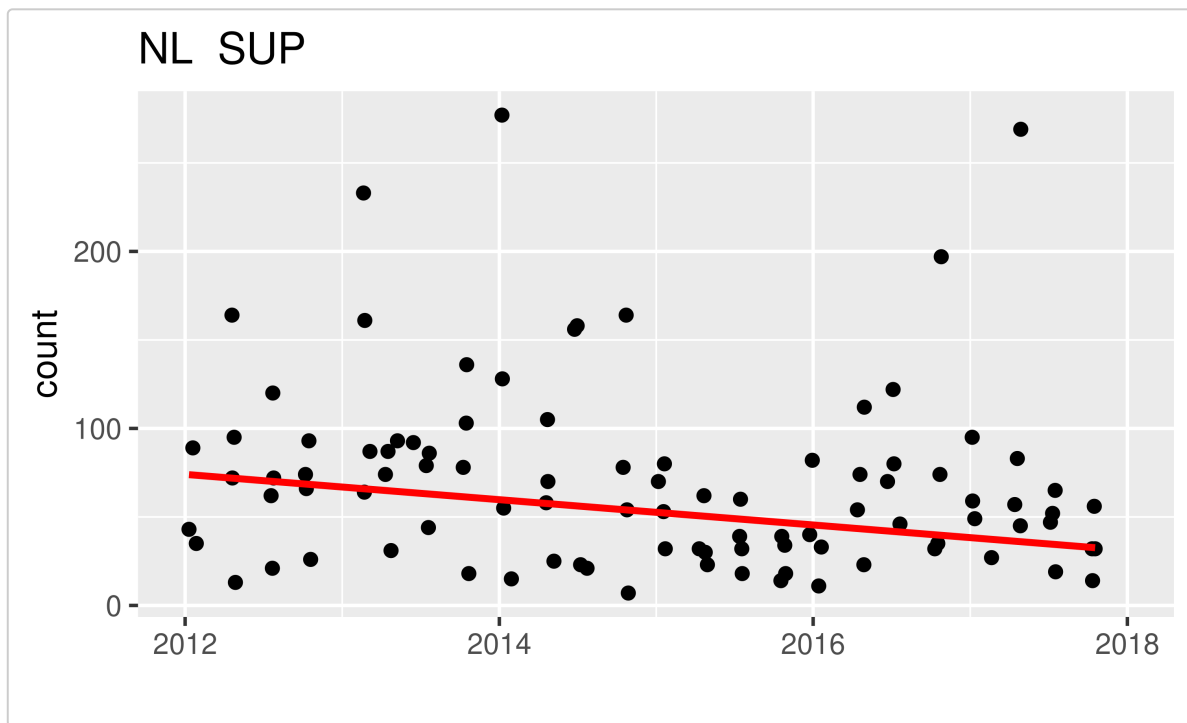
For each `region_code`, and the type names and group codes specified in the [settings file](#), the following statistics have been estimated:

- the number of surveys (N) in the region;
- Theil-Sen slope: the median of all [Theil-Sen slopes](#) within a region;
- [p-value](#): the p-value associated with the one-tailed Regional Kendall test ([Van Belle & Hughes, 1984](#); [Gilbert, 1987](#)) to test the null hypothesis of
 - no monotonically *increasing* trend in case the regional Theil-Sen slope is greater than zero;
 - no monotonically *decreasing* trend in case the regional Theil-Sen slope is smaller than zero;

A p-value less than an *a priori* specified [significance level](#) (e.g., often $\alpha = 0.05$), indicates a significant trend. If the p-value is greater than this significance level, we can't say that there is no trend. We can only conclude that our data do not show evidence for a significant trend (due to lack of data, noise, etc.).

The Regional Kendall test is a non-parametric test and as such does not make distributional assumptions on the data.

An example of a regional trend is given in the figure below:



5.2 Statistical summary file

In addition to a report, a results file (CSV-format) with descriptive statistics and the main trend results for each `location_code` is produced. An example of such a table is given below. See Section [descriptive statistics](#) for more details.

location_code	from	to	type/group_name	%TC	mean	median	cv	rmad	n	slope	p_value
NL001	2012-01-27	2017-10-11	TC	100	376.6	302.5	0.731	0.7793	24	39.98	0.1233
NL001	2012-01-27	2017-10-11	PLASTIC	88.8	341.2	269.5	0.7574	0.8472	24	35.57	0.1134
NL001	2012-01-27	2017-10-11	FISH	41.36	162.4	104.5	0.8622	0.9506	24	13.03	0.1233
NL001	2012-01-27	2017-10-11	plastic: string [32]	27.88	119.8	78	0.9669	1.255	24	17.01	0.0370
NL001	2012-01-27	2017-10-11	SUP	25.38	91.38	73	0.7284	0.7819	24	5.373	0.2230
NL001	2012-01-27	2017-10-11	plastic: plastic_small [117]	9.473	42.46	21.5	1.166	1.103	24	3.121	0.2134
NL001	2012-01-27	2017-10-11	plastic: plastic_large [46]	8.04	24.71	17.5	0.7851	0.5083	24	0.6599	0.4212
NL001	2012-01-27	2017-10-11	plastic: fishing_net_small [115]	7.525	23.25	4	1.544	1.483	24	-2.283	0.0395

location_code	from	to	type/group_name	%TC	mean	median	cv	rmad	n	slope	p_value
NL001	2012-01-27	2017-10-11	plastic: caps [15]	4.978	20.42	16	1.034	0.7413	24	2.312	0.0940
NL001	2012-01-27	2017-10-11	RUBBER	4.75	15	13.5	0.7379	0.6589	24	0.941	0.2061

5.3 Log-file

litter's log-file is very helpful to understand warnings and error messages. The log-file stores the description of all data analysis steps in chronological order. Part of a log-file is given below. The complete log-file is given in the [appendix](#).

```

2021-09-04 22:18:09 [INFO] Starting a new litter session
2021-09-04 22:18:09 [INFO] litter version: 0.9.0
2021-09-04 22:18:09 [INFO] litter release date: 2021-08-20
2021-09-04 22:18:09 [INFO] Reading settings file 'settings.yaml'
2021-09-04 22:18:09 [INFO] Check optional settings...
2021-09-04 22:18:09 [INFO] Check existence of required settings...
2021-09-04 22:18:09 [INFO] All required settings are available
2021-09-04 22:18:09 [INFO] Checking settings 'date_min' and 'date_max'
2021-09-04 22:18:09 [INFO] Settings 'date_min' and 'date_max' are valid
2021-09-04 22:18:09 [INFO] Checking setting 'percentage_total_count'
2021-09-04 22:18:09 [INFO] Setting 'percentage_total_count' is valid
2021-09-04 22:18:09 [INFO] Checking setting 'figure_quality'
2021-09-04 22:18:09 [INFO] Setting 'figure_quality' is valid
2021-09-04 22:18:09 [INFO] Settings file has been read
2021-09-04 22:18:09 [INFO] Constructing filename for report
2021-09-04 22:18:09 [INFO] Filename 'litter-results-20210904T221809.html' created
2021-09-04 22:18:09 [INFO] Construct filename for storing statistics
2021-09-04 22:18:09 [INFO] Filename 'litter-results-20210904T221809.csv' created
2021-09-04 22:18:09 [INFO] Starting litter analysis
2021-09-04 22:18:10 [INFO] Checking parameters in settings file

```

Each line contains a single log-event and always has the following format:

1. timestamp;
2. type of log event: INFO for informative messages, WARN for warnings, ERROR for errors;
3. a log message.

6 Troubleshooting

The runtime error messages and the log file should provide you with clear information about errors in the data file and settings, and about warnings (points of attention). For additional information you can consult the points below.

- Dates should comply with the [ISO date format](#), *i.e.* YYYY-mm-dd (*e.g.*, 2019-09-30) or the OSPAR data format (dd/mm/YYYY). If you prepare your input data with MS-Excel, make sure that the dates in the exported CSV-file also comply with one of these formats. You can easily check this with a text editor (like Notepad on MS-Windows);
- After typing `litter()` in the RStudio-console, a file dialogue should appear. If that is not the case, the file dialogue is probably covered by RStudio (see the task manager or use ALT-TAB on MS-Windows to navigate to the hidden file dialogue);
- **litteR** expects a period (.) as decimal separator (*e.g.*, 8.5) and *not* a comma (*e.g.*, *not* 8,5). In MS-Windows 7, this can be accomplished by means of the 'Region and Language' menu (*e.g.*, by using the English (UK) setting). These settings are particularly important when using MS-Excel for data preparation. In cases **litteR** reports errors, you should check the file format exported by MS-Excel in a text editor (*e.g.*, Notepad);
- When **litteR** complains about an `invalid multibyte string`, there is a character in your input file that is not part of the English alphabet. Substituting this character by a valid character in the range A-Z or a-z usually solves this problem.

7 References

Gilbert, R.O., 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold. 320 pp
<https://www.osti.gov/biblio/7037501-statistical-methods-environmental-pollution-monitoring>

Hanke G., Walvoort D., van Loon W., Addamo A.M., Brosich A., del Mar Chaves Montero M., Molina Jack M.E., Vinci M., Giorgetti A., EU Marine Beach Litter Baselines, EUR 30022 EN, Publications Office of the European Union, Luxemburg, 2019, ISBN 978-92-76-14243-0, <https://doi.org/10.2760/16903>, JRC114129.

Schulz, M., van Loon, W., Fleet, D. M., Baggelaar, P., & van der Meulen, E. (2017). OSPAR standard method and software for statistical analysis of beach litter data. *Marine pollution bulletin*, 122(1-2), 166-175.
<https://doi.org/10.1016/j.marpolbul.2017.06.045>

Schulz, Marcus, Dennis J.J. Walvoort, Jon Barry, David M. Fleet, Willem M.G.M. van Loon, 2019. Baseline and power analyses for the assessment of beach litter reductions in the European OSPAR region. *Environmental Pollution* 248:555-564. <https://doi.org/10.1016/j.envpol.2019.02.030>

Van Belle, G., J.P. Hughes, 1984. Nonparametric Tests for Trend in Water Quality. *Water Resources Research* 20: 127-136. <https://doi.org/10.1029/WR020i001p00127>

8 Appendix

Example of a log-file produced by **litteR**.

```
2021-09-04 22:18:09 [INFO] Starting a new litter session
2021-09-04 22:18:09 [INFO] litter version: 0.9.0
2021-09-04 22:18:09 [INFO] litter release date: 2021-08-20
2021-09-04 22:18:09 [INFO] Reading settings file 'settings.yaml'
2021-09-04 22:18:09 [INFO] Check optional settings...
```



```
2021-09-04 22:18:09 [INFO] Check existence of required settings...
2021-09-04 22:18:09 [INFO] All required settings are available
2021-09-04 22:18:09 [INFO] Checking settings 'date_min' and 'date_max'
2021-09-04 22:18:09 [INFO] Settings 'date_min' and 'date_max' are valid
2021-09-04 22:18:09 [INFO] Checking setting 'percentage_total_count'
2021-09-04 22:18:09 [INFO] Setting 'percentage_total_count' is valid
2021-09-04 22:18:09 [INFO] Checking setting 'figure_quality'
2021-09-04 22:18:09 [INFO] Setting 'figure_quality' is valid
2021-09-04 22:18:09 [INFO] Settings file has been read
2021-09-04 22:18:09 [INFO] Constructing filename for report
2021-09-04 22:18:09 [INFO] Filename 'litter-results-20210904T221809.html' created
2021-09-04 22:18:09 [INFO] Construct filename for storing statistics
2021-09-04 22:18:09 [INFO] Filename 'litter-results-20210904T221809.csv' created
2021-09-04 22:18:09 [INFO] Starting litter analysis
2021-09-04 22:18:10 [INFO] Checking parameters in settings file
2021-09-04 22:18:10 [INFO] Entering data quality control section
2021-09-04 22:18:10 [INFO] Checking existence of 'types-ospar.csv'
2021-09-04 22:18:10 [INFO] 'types-ospar.csv' exists
2021-09-04 22:18:10 [INFO] Validating type file
2021-09-04 22:18:10 [INFO] Checking required columns in type file
2021-09-04 22:18:10 [INFO] Required columns are available
2021-09-04 22:18:10 [INFO] Checking type names for duplicates
2021-09-04 22:18:10 [INFO] No duplicates found
2021-09-04 22:18:10 [INFO] Checking if table cells are either empty or 'x'
2021-09-04 22:18:10 [INFO] All table cells are OK
2021-09-04 22:18:10 [INFO] Checking extension of 'beach-litter-nl-2012-2017.csv' (should be
'csv')
2021-09-04 22:18:10 [INFO] file extension is correct
2021-09-04 22:18:10 [INFO] Checking existence of 'beach-litter-nl-2012-2017.csv'
2021-09-04 22:18:10 [INFO] 'beach-litter-nl-2012-2017.csv' exists
2021-09-04 22:18:10 [INFO] Checking if CSV-file is comma delimited
2021-09-04 22:18:10 [INFO] CSV-file is comma delimited
2021-09-04 22:18:10 [INFO] Reading litter data file
2021-09-04 22:18:10 [INFO] Check if required metadata columns 'location_code' and 'date'
exist.
2021-09-04 22:18:10 [INFO] All required columns are available
2021-09-04 22:18:10 [INFO] Check if optional columns are available.
2021-09-04 22:18:10 [INFO] Optional column(s) found: 'region_code'
2021-09-04 22:18:10 [INFO] Checking date format
2021-09-04 22:18:10 [INFO] All dates are ISO 8601 compliant (YYYY-mm-dd)
2021-09-04 22:18:10 [INFO] Checking consistency of dates
2021-09-04 22:18:10 [INFO] Dates are consistent. All dates are YYYY-mm-dd
2021-09-04 22:18:10 [INFO] Check if all litter types in the type file are present in the data
file
2021-09-04 22:18:10 [INFO] All litter types are present
2021-09-04 22:18:10 [INFO] Select only litter data
2021-09-04 22:18:10 [WARN] The following columns will be excluded from analysis:'region_name',
'country_code', 'country_name', 'location_name', 'survey: old_rope_small [200]', 'survey:
old_rope_large [201]', 'survey: old_plastic_pieces [202]', 'survey: old_gloves [203]',
'survey: old_cartons [204]', 'survey: old_oildrums_new [205]', 'survey: old_oildrums_old
[206]', 'survey: old_human_faeces [207]', 'survey: old_animal_faeces [208]', 'survey:
old_cloth_rope [210]' and 'survey: pellets [998]'
```

2021-09-04 22:18:10 [INFO] Check for empty cells
2021-09-04 22:18:10 [INFO] No empty cells found
2021-09-04 22:18:10 [INFO] Check that litter counts are numbers
2021-09-04 22:18:10 [INFO] Only numbers found
2021-09-04 22:18:10 [INFO] Check that litter counts are nonnegative numbers
2021-09-04 22:18:10 [INFO] No negative numbers found
2021-09-04 22:18:10 [INFO] Check that litter counts are natural numbers
2021-09-04 22:18:10 [INFO] Only natural numbers found
2021-09-04 22:18:10 [INFO] Check if not all litter counts in a single survey (record) are zero
2021-09-04 22:18:10 [INFO] No records found with only zero-counts
2021-09-04 22:18:10 [INFO] Check for duplicated records
2021-09-04 22:18:10 [INFO] No duplicated records found
2021-09-04 22:18:10 [INFO] No records with the same location_code/date found
2021-09-04 22:18:10 [INFO] Computing group totals
2021-09-04 22:18:11 [INFO] Computing relative group totals (relative w.r.t. TC)
2021-09-04 22:18:11 [INFO] Determining top 80% litter...
2021-09-04 22:18:11 [INFO] Entering outlier analysis section
2021-09-04 22:18:12 [INFO] Entering descriptive statistics section
2021-09-04 22:18:12 [INFO] Creating table with litter statistics
2021-09-04 22:18:13 [INFO] Table with litter statistics created
2021-09-04 22:18:17 [INFO] Entering descriptive regional statistics section
2021-09-04 22:18:17 [INFO] Creating table with regional litter statistics
2021-09-04 22:18:21 [INFO] Entering trend analysis section
2021-09-04 22:18:21 [INFO] Creating table with trend statistics
2021-09-04 22:18:21 [INFO] Table with trend statistics created
2021-09-04 22:18:21 [INFO] Creating time-series plots
2021-09-04 22:18:27 [INFO] Entering regional trend analysis section
2021-09-04 22:18:27 [INFO] Creating table with trend statistics
2021-09-04 22:18:27 [INFO] Table with trend statistics created
2021-09-04 22:18:27 [INFO] Creating regional time-series plots
2021-09-04 22:18:30 [INFO] Adding session information.
2021-09-04 22:18:30 [INFO] Report completed
2021-09-04 22:18:30 [INFO] All results have been written to 'litterR-results-20210904T221809'
2021-09-04 22:18:30 [INFO] litterR session terminated