

# Package ‘COVIDreportwriter’

April 27, 2020

**Title** Generation of COVID-19 reports from REDCAP clinical data

**Version** 0.1.0

**Description** Processes raw REDCAP clinical data for the generation of a PDF report containing text and plot summaries of the data.

**Depends** R (>= 3.5.0)

**Imports** binom, boot, broom, filesstrings, fitdistrplus, knitr, ggupset, glue, grid, gridExtra, lubridate, magrittr, rmarkdown, sf, survival, survminer, taRifx, tidyselect, tidyverse, viridis

**Encoding** UTF-8

**LazyData** true

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.1.0

## R topics documented:

adm.outcome.plot . . . . .	2
age.pyramid . . . . .	2
comorbidities.upset . . . . .	3
comorbidity.prevalence.plot . . . . .	3
generate.report . . . . .	4
icu.violin.plot . . . . .	4
import.and.process.data . . . . .	5
modified.km.plot . . . . .	6
onset.adm.plot . . . . .	6
outcomes.by.admission.date . . . . .	7
outcomes.by.country . . . . .	8
recruitment.dat.plot . . . . .	8
sites.by.country . . . . .	9
status.by.time.after.admission . . . . .	9
symptom.heatmap . . . . .	10
symptom.prevalence.plot . . . . .	10
symptoms.upset . . . . .	11
treatment.upset . . . . .	11
treatment.use.plot . . . . .	12
treatment.use.plot.icu . . . . .	12
violin.age.func . . . . .	13
violin.sex.func . . . . .	13
<b>Index</b>	<b>14</b>

---

<code>adm.outcome.plot</code>	<i>Plot distribution of time (in days) from admission to an outcome. Plots a Gamma distribution fit to the lengths of hospital stay (in days) from admission to an outcome - either death or discharge, accounting for unobserved outcomes. See 'Details'.</i>
-------------------------------	--

---

### Description

The estimates of the Gamma distribution were fitted to the observed data were obtained by a maximum likelihood estimation procedure implemented in the `fitdistcens` package in the `fitdistrplus` package. The lengths of stay for patients with unobserved outcomes were treated as interval censored data.

### Usage

```
adm.outcome.plot(data, ...)
```

### Arguments

<code>data</code>	Output of <code>process_data</code> .
-------------------	---------------------------------------

### Value

Plot of the Gamma distribution fit to lengths of hospital stay. The black dashed line indicates the position of the estimated mean of the Gamma distribution. (Note that the expected mean is different from the *observed mean* of lengths of hospital stay, which is estimated using records from patients with observed outcomes only.)

### References

Delignette-Muller, M. L., & Dutang, C. (2015). `fitdistrplus`: An R package for fitting distributions. *Journal of statistical software*, **64**(4), 1-34.

---

<code>age.pyramid</code>	<i>Plot patient demographics by outcome.</i>
--------------------------	--

---

### Description

Plots the age and sex distribution of patients according to clinical outcome.

### Usage

```
age.pyramid(data, ...)
```

### Arguments

<code>data</code>	Output of <code>process_data</code> . This should be a dataframe which includes columns for age group, sex, and outcome of patients. See 'Details'.
-------------------	---

**Value**

Bar plot of the age (in intervals of four years) and sex (male/female) of patients, plotted according to clinical outcome (discharge/death/ongoing care).

The columns of data for age group, sex and outcome should be named "agegp5", "sex" and "outcome" respectively and formatted as follows: the variable "sex" should be numeric with values 1 and 2 for males and females respectively; the variable "agegp5" should be a factor with levels 0-4, 5-9, 10-14, ..., 90+; and the variable "outcome" should be a factor with levels 'discharge', 'censored' and 'death'; in this case, 'censored' patients are those for whom clinical care is ongoing.

---

comorbidities.upset      *Plot prevalence of combinations of comorbidities.*

---

**Description**

Plots the distribution of combinations of the most common comorbidities, amongst all patients for whom these data were recorded.

**Usage**

```
comorbidities.upset(data, max.comorbidities, comorbidities, ...)
```

**Arguments**

data                      Output of [process\\_data](#).

max.comorbidities

The max.comorbidities most frequent comorbidities will be included in the upset plot. Defaults to 4.

**Value**

UpSet plot showing the frequency of combinations of the top max.comorbidities comorbidities. Filled and empty circles below the x-axis of the plot indicate the presence or absence of each comorbidity. The "Any other" category in the upset plot contains all remaining comorbidities which are not included in the top max.comorbidities comorbidities, as well as any other comorbidities recorded as free text by clinical staff.

@examples comorbidities.upset(data = patient.data, max.comorbidities = 4)

---

comorbidity.prevalence.plot      *Plot distribution of comorbidity prevalence.*

---

**Description**

Plots the proportion of patients reporting each comorbidity at admission.

**Usage**

```
comorbidity.prevalence.plot(data, comorbidities, ...)
```



**Value**

Violin plots (with box plots) showing the distribution of the total length of hospital stay for patients who were admitted to ICU/HDU and the distribution of the lengths of stay within ICU/HDU. The coloured areas of the plot indicate the kernel probability density of the observed data and the box plots show the median and interquartile range of the lengths of stay.

---

```
import.and.process.data
```

*Import data from Redcap .csv output for processing*

---

**Description**

Import data from Redcap .csv output for processing

**Usage**

```
import.and.process.data(  
  data.file,  
  data.dict.file,  
  column.table.file = NULL,  
  source.name = NA,  
  message.out.file = NULL,  
  embargo.length = 0,  
  ref.date = today(),  
  verbose = F  
)
```

**Arguments**

data.file	Path of the data file
data.dict.file	Path of the data dictionary file
column.table.file	Path of the column translation file (default required_columns.csv)
source.name	String identifier for the data source (optional)
message.out.file	Path of a file to store messages about fields that were overwritten in the data cleaning process. If not given, this file is not generated
embargo.length	Length of the data embargo in days; patients enrolled less than this period before ref.date will be excluded from reports. Default 0.
ref.date	Date to be taken as the date of the report. Default is today's date.
verbose	Flag for verbose output

---

<code>modified.km.plot</code>	<i>Plot case fatality ratio (CFR) and survival functions for deaths and recovery.</i>
-------------------------------	---

---

### Description

Plots the proportion of deaths and recoveries over time as well as a non-parametric estimate for the CFR using an adapted Kaplan-Meier method. See 'Details'.

### Usage

```
modified.km.plot(data, embargo.limit, ...)
```

### Arguments

`data`                      Output of `process_data`.

### Details

The CFR and survival functions for death and recovery are estimated using a nonparametric Kaplan-Meier-based method proposed by Ghani et al. (2005). This method estimates the CFR with the formula  $\frac{a}{a+b}$ , where  $a$  and  $b$  are the values of the cumulative incidence function for deaths and recoveries respectively, estimated at the last observed time point. See 'References' for details.

### Value

Plot of the survival functions for deaths and recoveries and a line indicating the CFR estimate.

### References

A. C. Ghani, C. A. Donnelly, D. R. Cox, J. T. Griffin, C. Fraser, T. H. Lam, L. M. Ho, W. S. Chan, R. M. Anderson, A. J. Hedley, G. M. Leung (2005). Methods for Estimating the Case Fatality Ratio for a Novel, Emerging Infectious Disease, *American Journal of Epidemiology*, **162**(5), 479 - 486. doi:[10.1093/aje/kwi230](https://doi.org/10.1093/aje/kwi230).

---

<code>onset.adm.plot</code>	<i>Plot distribution of time (in days) from symptom onset to admission. Plots a Gamma distribution fit to durations (in days) from symptom onset to admission. This includes only patients with complete records on the time (in days) between symptom onset and admission.</i>
-----------------------------	---

---

### Description

The estimates of the Gamma distribution were fitted to the observed data were obtained by a maximum likelihood estimation procedure implemented in the `fitdistcens` package in the `fitdistrplus` package.

### Usage

```
onset.adm.plot(data, ...)
```

**Arguments**

data                      Output of [process\\_data](#).

**Value**

Plot of the Gamma distribution fit to lengths of hospital stay. The black dashed line indicates the position of the estimated mean of the Gamma distribution.

**References**

Delignette-Muller, M. L., & Dutang, C. (2015). *fitdistrplus*: An R package for fitting distributions. *Journal of statistical software*, **64**(4), 1-34.

---

outcomes.by.admission.date

*Plot weekly admission counts*

---

**Description**

Plots patient numbers and outcomes by epidemiological week (of 2020) of admission (or, for patients infected in hospital, of symptom onset).

**Usage**

```
outcomes.by.admission.date(data, embargo.limit, ...)
```

**Arguments**

data                      Output of [process\\_data](#). This should be a dataframe which includes columns for the date of admission and outcome for each patient. See 'Details'.

**Value**

Bar plot showing the number of patients per country and by outcome (discharge/ongoing care/death). Bars are annotated with counts.

The variables of data containing the date and outcome information of patients should be named "start.date" and "outcome" respectively. The variable "outcome" should be a factor with levels 'discharge', 'censored' and 'death'; in this case, 'censored' patients are those for whom clinical care is ongoing.

---

`outcomes.by.country`     *Plot the distribution of patients by country and outcome*

---

### Description

Plots the distribution of patients by country and outcome

### Usage

```
outcomes.by.country(data, ...)
```

### Arguments

`data`                      Output of `process_data`. This should be a dataframe which includes columns for the country and outcome associated with each patient. See 'Details'.

### Value

Bar plot showing the number of patients per country and by outcome (discharge/ongoing care/death). Actual counts of the total number of patients for each country are printed on top of each bar.

The variables of data containing the country and outcome information of patients should be named "Country" and "outcome" respectively. The variable "outcome" should be a factor with levels 'discharge', 'censored' and 'death'; in this case, 'censored' patients are those for whom clinical care is ongoing.

---

`recruitment.dat.plot`     *Plot cumulative recruitment of patients.*

---

### Description

Plots the cumulative recruitment of patients, separated by whether follow-up is ongoing or an outcome has been recorded.

### Usage

```
recruitment.dat.plot(data, embargo.limit, ...)
```

### Arguments

`data`                      Output of `process_data`.

`embargo.limit`     The cut-off date for inclusion in the report. Patients recruited after `embargo.limit` are not considered in the analysis. Set `embargo.limit` to the date of the report if all patients are to be considered.

this date have not been included.

### Value

Plot showing the cumulative number of patients in the study. One line plots the cumulative of patients for whom follow-up has been recorded while the other line captures patients for whom follow-up is ongoing. The first dashed black line indicates the `embargo.limit`. The second black line is the cut-off date for the next report, assuming that reports are issued weekly.



---

sites.by.country	<i>Plot the number of sites by country</i>
------------------	--

---

**Description**

Plots the number of sites by country

**Usage**

```
sites.by.country(data, ...)
```

**Arguments**

data	Output of <a href="#">process_data</a> . This should be a dataframe which includes columns for the country and site associated with each patient. See 'Details'. The columns of data containing country and site names should be named "Country" and "site.name" respectively.
------	---

**Value**

Bar plot showing the number of sites per country. Actual counts are printed on top of each bar.

---

status.by.time.after.admission	<i>Plot timelines by patients' status.</i>
--------------------------------	--

---

**Description**

Plots the distribution of patients' status by number of days after admission. Seven statuses are considered: 'Discharged', 'Transferred', 'Unknown', 'Ongoing care', 'Ward', 'ICU' and 'Death'. See 'Details'.

**Usage**

```
status.by.time.after.admission(data, ...)
```

**Arguments**

data	Output of <a href="#">process_data</a> .
------	--

**Details**

Patients with "Unknown" status have left the site at the time of report but have unknown outcomes due to missing data. Patients with "Transferred" status have been transferred to another health facility by the time of the report. Patients still on site at the time of report appear in the "Ongoing care" category for days which are in the future at that time. (For example, a patient admitted 7 days before the date of report and still on site at report would be categorised as "ongoing care" for days 8 and later.) The black line in the plot marks the end of 14 days; due to the cut-off, only a small number of patients appear in the "ongoing care" category left of this line.

**Value**

Plot showing the proportion of patients in each category over time. Each status has been assigned a different colour code to enable easy differentiation.

---

symptom.heatmap	<i>Plot pairwise symptom prevalence.</i>
-----------------	--

---

**Description**

Plots a heatmap for prevalence of pairwise combinations of symptoms. The pairwise prevalence proportions are calculated amongst patients with recorded presence or absence of both symptoms.

**Usage**

```
symptom.heatmap(data, admission.symptoms, ...)
```

**Arguments**

data                      Output of [process\\_data](#).

**Value**

Heatmap showing the proportion of patients for each pairwise combination of symptoms.

---

symptom.prevalence.plot	<i>Plot distribution of symptom prevalence.</i>
-------------------------	---

---

**Description**

Plots the proportion of patients presenting with each symptom at admission.

**Usage**

```
symptom.prevalence.plot(data, admission.symptoms, ...)
```

**Arguments**

data                      Output of [process\\_data](#).

**Details**

Note that the denominators used in the computation of proportions may differ by symptom as symptom information may be incomplete for some patients.

**Value**

Barplot showing the proportion of patients reporting each symptom. Bars are annotated with a fraction representing the number of patients presenting with a symptom over the number of patients for whom presence or absence of that symptom was recorded.

---

symptoms.upset	<i>Plot prevalence of combinations of symptoms.</i>
----------------	---

---

**Description**

Plots the distribution of combinations of the most common symptoms on admission, amongst all patients for whom these data were recorded.

**Usage**

```
symptoms.upset(data, max.symptoms, admission.symptoms, ...)
```

**Arguments**

data	Output of <a href="#">process_data</a> .
max.symptoms	The max.symptoms most frequent symptoms will be included in the plot. Defaults to 4.

**Value**

UpSet plot showing the frequency of combinations of the top max.symptoms symptoms. Filled and empty circles below the x-axis of the plot indicate the presence and absence respectively of each symptom. The “Any other” category in the upset plot contains all remaining comorbidities which are not included in the top max.symptoms symptoms.

**Examples**

```
symptoms.upset(data = patient.data, max.symptoms = 4)
```

---

treatment.upset	<i>Plot frequency of combinations of treatments.</i>
-----------------	--

---

**Description**

Plots the distribution of combinations of the 5 most common treatments administered during hospital stay, across all patients with completed hospital stay and recorded treatment data.

**Usage**

```
treatment.upset(data, ...)
```

**Arguments**

data	Output of <a href="#">process_data</a> .
------	--

**Value**

UpSet plot showing the frequency of combinations of the 5 most common treatments. Filled and empty circles below the x-axis of the plot indicate treatments that were and were not administered respectively.

---

treatment.use.plot	<i>Plot distribution of treatments used.</i>
--------------------	--

---

### Description

Plots the proportion of patients given each treatment during clinical care.

### Usage

```
treatment.use.plot(data, treatments, ...)
```

### Arguments

data	Output of <a href="#">process_data</a> .
------	--

### Details

Note that the denominators used in the computation of proportions may differ by treatment as information on treatment given may be missing/incomplete for some patients.

### Value

Barplot showing the proportion of patients given each treatment. Bars are annotated with a fraction representing the number of patients given a treatment over the number of patients for whom presence or absence of that treatment was recorded.

---

treatment.use.plot.icu	<i>Plot frequency of combinations of Intensive Care Unit (ICU) and High Dependency Unit (HDU) treatments.</i>
------------------------	---

---

### Description

Plots the distribution of combinations of treatments administered during ICU/HDU stay

### Usage

```
treatment.use.plot.icu(data, treatments, ...)
```

### Arguments

data	Output of <a href="#">process_data</a> .
------	--

### Value

UpSet plot showing the frequency of combinations of ICU/HDU treatments. Filled and empty circles below the x-axis of the plot indicate treatments that were and were not administered respectively.

---

violin.age.func	<i>Plot lengths of hospital stay by age group</i>
-----------------	---

---

**Description**

Plots the distribution of lengths of stay by age group. Only cases with reported outcomes (i.e. death/discharge) are considered.

**Usage**

```
violin.age.func(data, embargo.limit, ...)
```

**Arguments**

data                      Output of [process\\_data](#).

**Value**

Violin plots (with box plots) showing the distribution of the total length by age group. Age is plotted in 10-year intervals: 0-9, 10-19, ..., 70+. The coloured areas of the plot indicate the kernel probability density of the observed data and the box plots show the median and interquartile range of the lengths of hospital stay for each age group.

---

violin.sex.func	<i>Plot lengths of hospital stay by sex</i>
-----------------	---

---

**Description**

Plots the distribution of lengths of stay for males and females on the same graph. Only cases with reported outcomes (i.e. death/discharge) are considered.

**Usage**

```
violin.sex.func(data, embargo.limit, ...)
```

**Arguments**

data                      Output of [process\\_data](#).

**Value**

Violin plots (with box plots) showing the distribution of the total length of hospital stay by sex. The coloured areas of the plot indicate the kernel probability density of the observed data and the box plots show the median and interquartile range of the lengths of stay for each sex.

# Index

`adm.outcome.plot`, [2](#)  
`age.pyramid`, [2](#)  
  
`comorbidities.upset`, [3](#)  
`comorbidity.prevalence.plot`, [3](#)  
  
`fitdistcens`, [2](#), [6](#)  
  
`generate.report`, [4](#)  
  
`icu.violin.plot`, [4](#)  
`import.and.process.data`, [5](#)  
  
`modified.km.plot`, [6](#)  
  
`onset.adm.plot`, [6](#)  
`outcomes.by.admission.date`, [7](#)  
`outcomes.by.country`, [8](#)  
  
`process_data`, [2–4](#), [6–13](#)  
  
`recruitment.dat.plot`, [8](#)  
  
`sites.by.country`, [9](#)  
`status.by.time.after.admission`, [9](#)  
`symptom.heatmap`, [10](#)  
`symptom.prevalence.plot`, [10](#)  
`symptoms.upset`, [11](#)  
  
`treatment.upset`, [11](#)  
`treatment.use.plot`, [12](#)  
`treatment.use.plot.icu`, [12](#)  
  
`violin.age.func`, [13](#)  
`violin.sex.func`, [13](#)