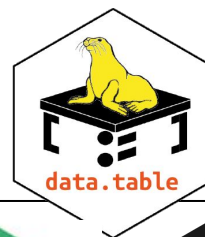
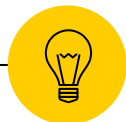


R Package for Data Cleaning

R - Ladies Gaborone

25/01/2022





Overview

Who I am

Messy to tidy data

R Packages for Data Cleaning

Janitor

Naniar

Data wizard

data.table

Demo in RStudio



Simisani Ndaba

Member of R-Ladies Gaborone,

Teaching Assistant since 2016, Department of Computer Science,
University of Botswana www.ub.bw

Msc in Computer Information Systems



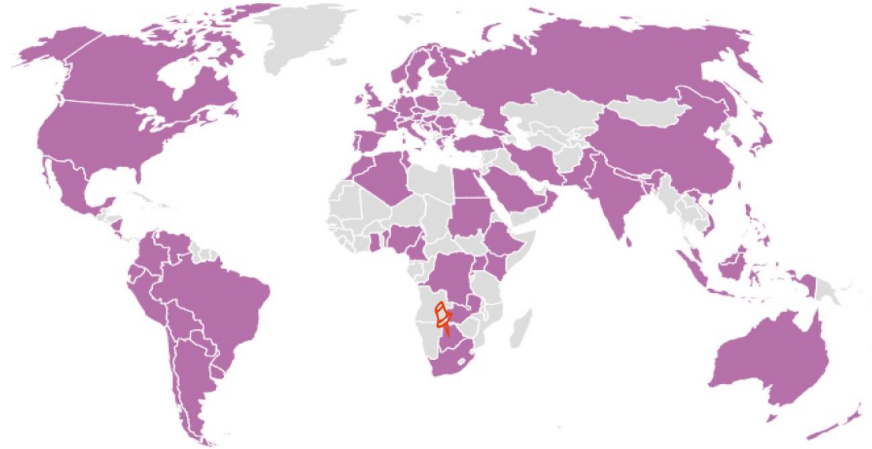
@RLadiesGaborone/@simisani10



simisani.ndaba013@gmail.com



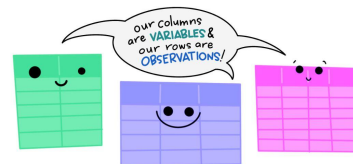
Where I'm from



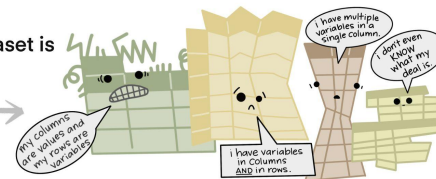


TIDY DATA WORKBENCH

The illustration shows a green, round character with a smiling face and small arms and legs. It is standing next to a wooden table. On the table, there are three data manipulation tools (wrenches) hanging from a pegboard. The tools are labeled 'rename' (red), 'select' (orange), and 'filter' (blue). There are also three data frames (purple, blue, and pink) on the table, each with a smiling face. The background is a light blue grid.

[illegible]

-HADLEY WICKHAM





<https://codata.org/initiatives/data-skills/codata-connect/webinar-series-research-skills-enhancement/webinar-4-importance-of-data-cleaning/>

R Packages for Data Cleaning_20 x Webinar 4: Importance of Data c x meetup_presentations_gaborone x +

codata.org/initiatives/data-skills/codata-connect/webinar-series-research-skills-enhancement/webinar-4-importance-of-data-cleaning/

Apps Invitation: Rladies... Tuesday: You're hos... R Packages for Data...

Reading list

CODATA
COMMITTEE ON DATA
INTERNATIONAL
SCIENCE COUNCIL

About ▾ Membership ▾ Events ▾ Initiatives ▾ Publications ▾ Blog 🔍

Webinar 4: Importance of Data cleaning

The fourth webinar in this series, took place on 5th August 2021.

- The slides are available at the link: [Download Presentation](#)

The recording is available below from Vimeo or in the [CODATA GoToWebinar Channel](#)

Search... Search

Importance of Data cleaning.mp4
from CODATA

CODATA
COMMITTEE ON DATA
INTERNATIONAL
SCIENCE COUNCIL

UNIVERSITY OF BOTSWANA

**Research Skill Enhancement Webinar co-hosted by
RDA CODATA Summer School and CODATA Connect group
presented by
Simisani Ndaka**



R Packages for Data Cleaning

Janitor

Naniar

Datawizard

data.table



Janitor_package

exploring and cleaning data

Functions: tabulations, duplications,

column names and formatting

The adorn_ functions dress up the results of these tabulations calls fast,

basic reporting

<https://cran.r-project.org/web/packages/janitor/vignettes/janitor.html>





Naniar_package

Naniar = NA in r

A suite of tools for generating visualisation of missing values,
imputations and summarising missing data

Data structures used to track missing data





Multiple Imputation using amelia_package

imputation=replacement of missing values

Method:

- Joint Model=Gaussian/Normal Distribution
- Conditional Model=linear regression
- PCA(Principal Component Analysis)

Amelia_package

Joint Modeling

amelia() uses the

- bootstrap, resampling algorithm and
- EM(Expectation Maximum) algorithm, maximum likelihood based model



Data.table_package

Alternative version to data.frame

Used on tabular data

incredibly fast

DT syntax

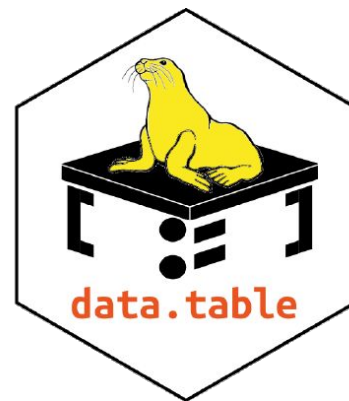
DT[i , j, by]

row, column, group by /SQL

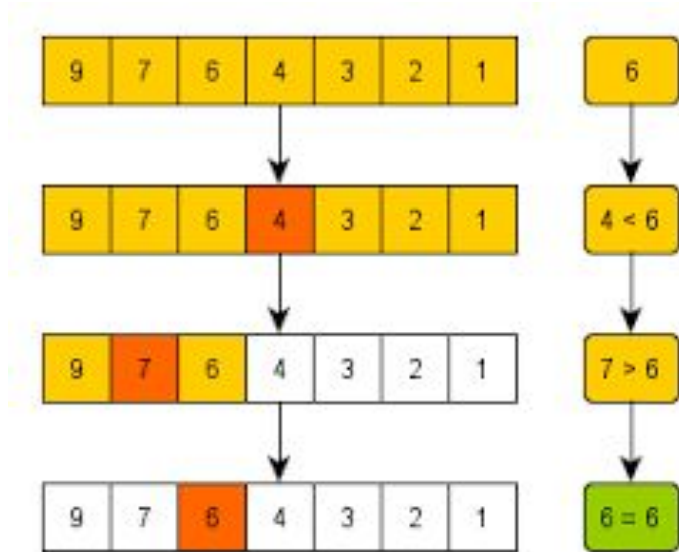
Group using by

Binary search using Keys(reference)

<https://cran.r-project.org/web/packages/data.table/vignettes/datatable-intro.html>



Binary Search....Data.table package cont..





Datawizard_package

A lightweight package to easily manipulate, clean, transform and prepare data for analysis

Data wrangling, transformation, visualisation, text formatting, data properties

<https://easystats.github.io/datawizard/>





Demo in RStudio

```
install.packages(janitor)
```

```
install.packages(naniar)
```

```
install.packages(datawizard)
```

```
install.packages(data.table)
```

```
#extra functionality
```

```
install.packages(amelia)
```

```
install.packages(dplyr)
```

```
install.packages(readr)
```



Thanks!



<https://www.meetup.com/rladies-gaborone/>



<https://youtube.com/RLadiesGaborone>



https://github.com/rladies/meetup_presentations_gaborone

