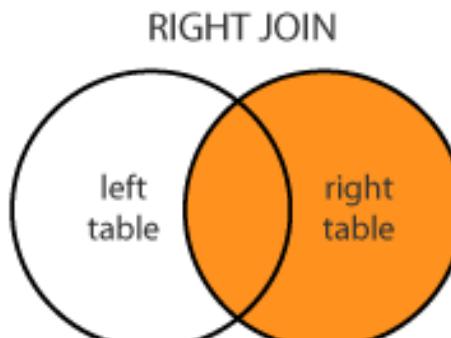
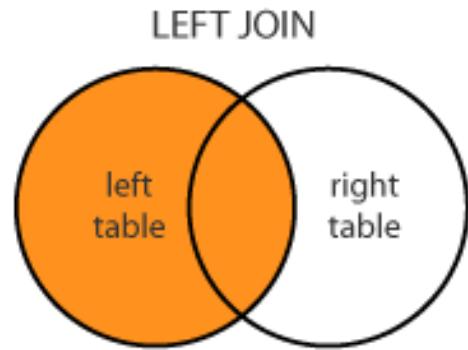
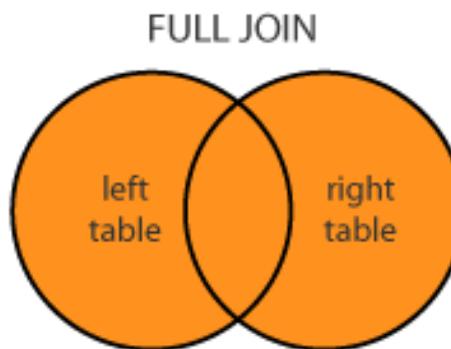
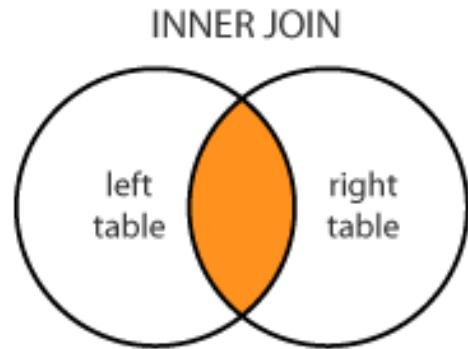


DATAFRAMES PART 2

dr Maria Kubara, WNE UW

JOINING DATSETS



<https://i.stack.imgur.com/iJUML.png>

```
# Inner join  
merge(set1, set2)
```

```
# Outer join  
merge(x = set1, y = set2,  
by = "IdClient", all =  
TRUE)
```

```
# Left outer join  
merge(x = set1, y = set2,  
by = "IdClient", all.x =  
TRUE)
```

```
# Right outer join  
merge(x = set1, y = set2,  
by = "IdKlienta", all.y =  
TRUE)
```

TEXT DATA – USEFUL FUNCTIONS

„Cutting-off” signs at given position:

`substr()`

Merging elements into one text sequence:

`paste()` `paste0()`

Returning the length of elements in subsequent positions of text vector:

`nchar()`

Changing letter capitalization:

`toupper()` `tolower()`

ORDERING DATA

Transpose data/matrix:

`t()`

Sorting a variable by its values:

`sort()`

Ordering dataset by values of given variable:

`order()`

FILTERING DATA

Choosing rows by the condition on column values:

```
set12 [set1$Product ==  
       "Bike", ]
```

Creating a subset using certain feature:

```
subset (set12, Product ==  
        "Bike")
```

CHECKING FOR MISSING DATA

Checking which values
are missing:

`is.na()`

Checking which rows do
not include missing values:

`complete.cases()`

RANDOM NUMBERS

Random number generator:

`runif()`

Specifying the starting point
for random numer generating:

`set.seed()`

USING PACKAGES

Using a package in R:

1. install the package when using it for the first time in the given R version on a given computer

```
install.packages ("packageName")
```

2. load the package every time you'd like to use it - once per working R session

```
library ("packageName")
```

```
library(packageName) #both notations are ok
```

Package documentation website

cluster: "Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.

Methods for Cluster analysis. Much extended the original from Peter Rousseeuw, Anja Struyf and Mia Hubert, bas

Version: 2.1.4 **Pa***version*

Priority: recommended

Depends: R (>= 3.5.0)

Imports: graphics, grDevices, stats, utils

Suggests: MASS, Matrix

Published: 2022-08-22

Author: Martin Maechler [aut, cre], Peter Rousseeuw [aut] (Fortran original), Anja Struyf [au
(volume.ellipsoid({d >= 3}))]

Maintainer: Martin Maechler <[maechler at stat.math.ethz.ch](mailto:maechler@stat.math.ethz.ch)>

License: [GPL-2](#) | [GPL-3](#) [expanded from: [GPL \(> 2\)](#)]

URL: <https://svn.r-project.org/R-packages/trunk/cluster/>

NeedsCompilation: yes

Citation: [cluster citation info](#) Citation

Materials: README NEWS ChangeLog

In views: Cluster Environmetrics, Robust

CRAN checks: cluster results

Documentation:

Reference manual: [cluster.pdf](#)

Downloads:

Package source: [cluster 2.1.4.tar.gz](#) [Source code](#)

Windows binaries: r-devel: cluster 2.1.4 zip r-release: cluster 2.1.4 zip r-oldrel: cluster 2.1.4 zip

macOS binaries: [r-release \(arm64\)](#): `cluster_2.1.4.tgz` [r-oldrel \(arm64\)](#): `cluster_2.1.4.tgz` [r-release \(x86_64\)](#): `cluster_2.1.4.tgz`

Old sources: [cluster archive](#)

[Old sources](#): [cluster archive](#) [Older versions](#)

Reverse dependencies:

Reverse depends: [abodOutlier](#), [aCGH](#), [AssocTests](#), [bios2mds](#), [BoutrosLab.plotting.general](#), [briKmeans](#), [ClassDiscovery](#), [NMF](#), [nomclust](#), [orthpart](#), [Oscone](#), [PAMhm](#), [namr](#), [RnBeads](#), [RPMM](#), [saccadr](#), [STROMA4](#), [Struct](#)

Reverse imports: [ADaCGH2](#), [adiv](#), [ADPclust](#), [adSplit](#), [agricolae](#), [anocva](#), [Anthropometry](#), [aPCoA](#), [aqp](#), [artMS](#), [AclusterExperiment](#), [clusterHD](#), [Clustering](#), [clustrd](#), [CLUSTShiny](#), [cobiclust](#), [coca](#), [CoGAPS](#), [coEvaluateCore](#), [EvoPhylo](#), [factoextra](#), [FactoMineR](#), [FADPclust](#), [FairMclus](#), [fdm2id](#), [flowStats](#), [f](#)

R Updating Loaded Packages



One or more of the packages to be updated are currently loaded. Restarting R prior to install is highly recommended.

RStudio can restart R before installing the requested packages. All work and data will be preserved during restart.

Do you want to restart R prior to install?

Yes

No

Cancel

If you try to install a package that is already loaded to the R session – and it's outdated compared to the newest version available on CRAN:

CANCEL → You will just close this pop-up and stay with the currently loaded package version.

NO → You will reinstall the package to the newest version, then you need to reload it again. R session won't be restarted – there may be some issues with the package running.

YES → You will reinstall the package to the newest version, then you need to reload it again. R session WILL BE RESTARTED. R Studio tries to preserve your work and keep it in the memory but you should be prepared to re-run all your code. Stable solution – updated package will be working as it should.