

Dedicated Aquaphotomics-Software
R-Package „aquap2“
General Introduction and Workshop

Bernhard Pollner & Zoltan Kovacs

Aquaphotomics: Understanding Water in Biology
at the 2nd International Aquaphotomics Symposium
26.-29. November 2016, Kobe, Japan

Outline & Schedule

≤ 40 minutes:

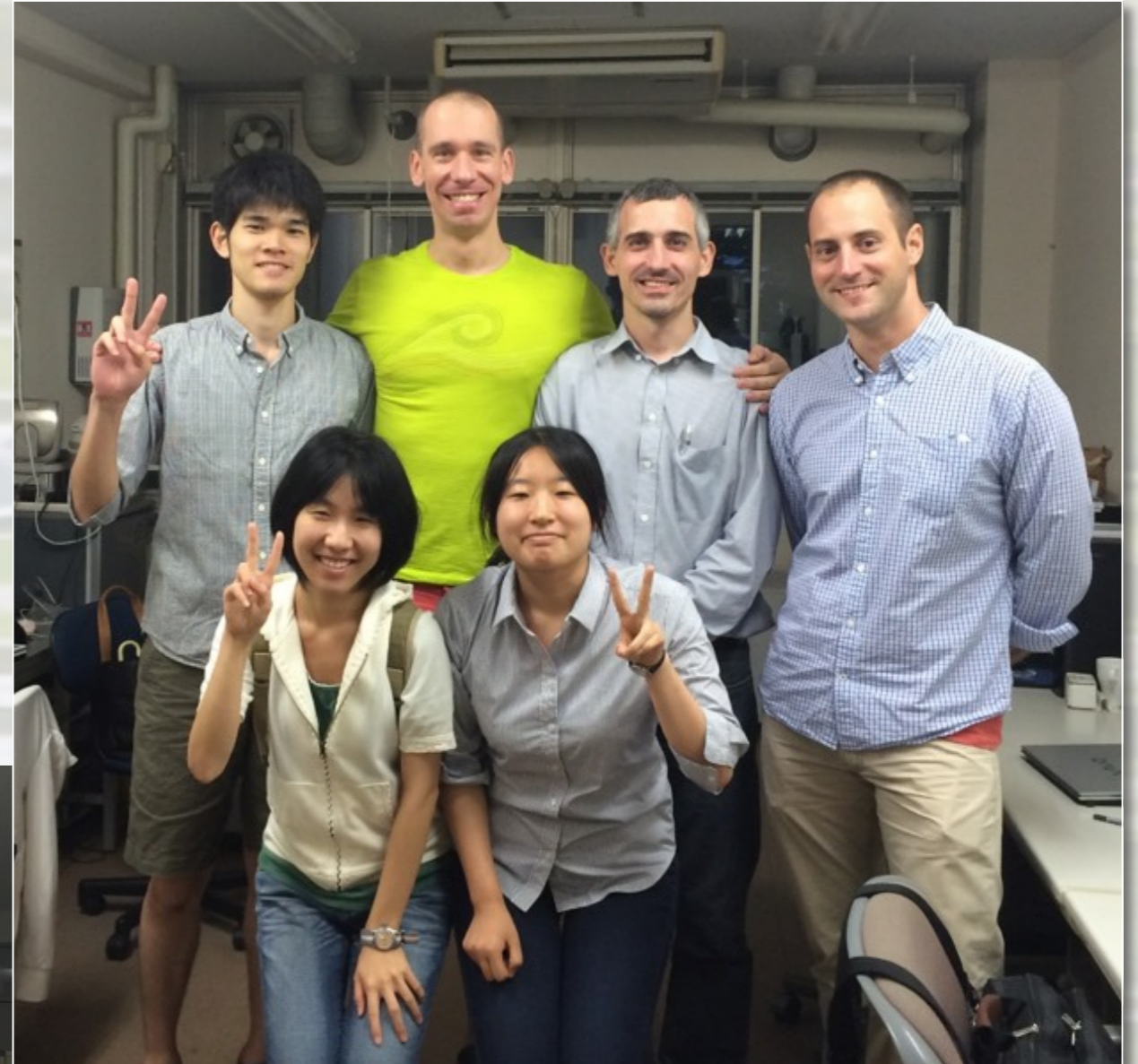
- general introduction - basic concept
- aquap2 helps in:
 - experiment-design,
 - data import and organization
 - data analysis and visualization

~ 2.25 hours:

- practical use & „hands-on“ workshop
- using training datasets
- just bring your laptop!



How it all Started...





The R Project for Statistical Computing

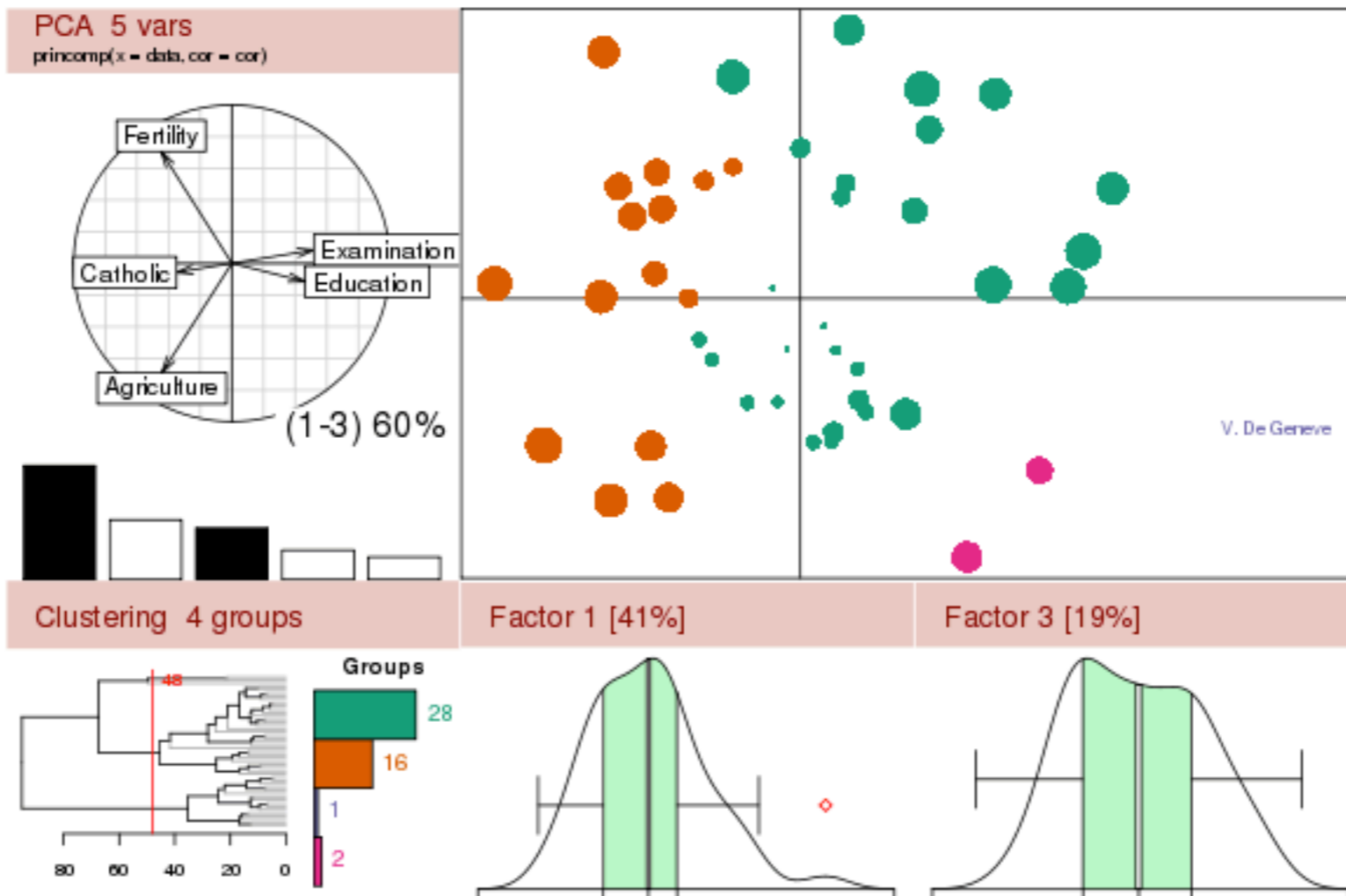
About R
[What is R?](#)
[Contributors](#)
[Screenshots](#)
[What's new?](#)

Download, Packages
[CRAN](#)

R Project
[Foundation](#)
[Members & Donors](#)
[Mailing Lists](#)
[Bug Tracking](#)
[Developer Page](#)
[Conferences](#)
[Search](#)

Documentation
[Manuals](#)
[FAQs](#)
[The R Journal](#)
[Wiki](#)
[Books](#)
[Certification](#)
[Other](#)

Misc
[Bioconductor](#)
[Related Projects](#)
[User Groups](#)
[Links](#)



Getting Started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).
- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News :

- **R version 3.1.1** (Sock it to Me) has been released on 2014-07-10.
- **R version 3.0.3** (Warm Puppy) has been released on 2014-03-06.
- [The R Journal Vol.5/2](#) is available.
- [useR! 2014](#), took place at the University of California, Los Angeles, USA June 30 - July 3, 2014.
- [useR! 2015](#), will take place at the University of Aalborg, Denmark, June 30 - July 3, 2015.



Why R-Project ?

- high level programming language
- open source - freeware !
- cross-platform
- powerful - extendable
- possible: something really new
- widely used in academia

[About R](#)
[What is R?](#)
[Contributors](#)
[Screenshots](#)
[What's new?](#)
[Download Packages](#)
[CRAN](#)

[R Project](#)
[Community](#)
[Members](#) / [Donors](#)
[Mailing Lists](#)
[Bug Tracking](#)
[Development](#)
[Conferences](#)
[Search](#)

[Documentation](#)
[Manual](#)
[FAQ](#)
[The R Journal](#)
[Wiki](#)
[Book](#)
[Certification](#)
[Other](#)

[Use Cases](#)
[Biocollections](#)
[Related Projects](#)
[User Groups](#)
[Links](#)

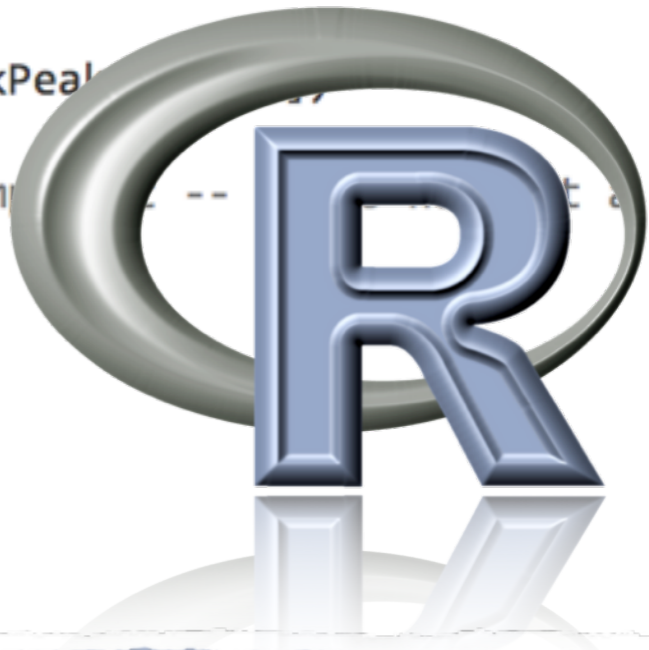
• R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN](#) [mirror](#).

• If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

- R version 3.1.1 (Sock it to Me) has been released on 2014-07-10.
- R version 3.0.3 (Warm Puppy) has been released on 2014-03-06.
- [The R Journal Vol.5/2](#) is available.
- [useR! 2014](#), took place at the University of California, Los Angeles, USA June 30 - July 3, 2014.
- [useR! 2015](#), will take place at the University of Aalborg, Denmark, June 30 - July 3, 2015.

```
the universal input, checking for the class of the input-object
pickPeaks <- function(ObjectToPickPeaks, bandwidth=25, comps=1:4, discrim=FALSE) {      ### universal peak picker
  if (class(ObjectToPickPeaks) == "mvr") {
    allColNames <- colnames(ObjectToPickPeaks$coefficients[, , 1:ObjectToPickPeaks$ncomp])
    lastName <- allColNames[length(allColNames)]
    if (!is.character(lastName)) {          ## problem if there is only one component -- then the name back
      lastName <- "1 comps"
    }
    mat <- ObjectToPickPeaks$coefficients[, , ObjectToPickPeaks$ncomp]
    dfToPickPeaks <- data.frame(X=mat)
    colnames(dfToPickPeaks) <- lastName
  }
  if (class(ObjectToPickPeaks) == "PCA") {
    dfToPickPeaks <- data.frame(ObjectToPickPeaks$loadings[, , comps])
  }
  if (class(ObjectToPickPeaks) == "data.frame") {
    dfToPickPeaks <- ObjectToPickPeaks
  }
  pickResultsList <- pickPeaksInner(dfToPickPeaks, bandwidth, discrim)
}
EOF

needs the pick results object created by pickPeaks as input; contains the vector that was used for picking !
plotPickResults <- function (pickResults, onMain="", onSub="", ncaVariances=NULL, customColor=NULL) {      #####
  if (length(pickResults$rawVector) < 1) {
    stop("An Error at plotPickResults")
  }
  a <- colnames(pickResults$rawVector)
  b <- substr(a, stringFromCharPreview(a), nchar(a)) # to get rid of the "w" in front of the numbers
  wavelengths <- as.numeric(b)          # so we have the wavelength in the column, and the vectors to be picked f
  res <- pickResults$picks$pickResult
  colPos <- stngs$colPosPeaks
  colNeg <- stngs$colNegPeaks
  positionTable <- res[1: (nrow(res)/2) , ]
  heightTable <- res[((nrow(res)/2)+1):nrow(res) , ]
  Yrange <- range(pickResults$rawVector)[2] - range(pickResults$rawVector)[1]
  onSub <- paste(onSub, ", bw=", pickResults$picks$pickWindow, sep="")
}
```



MVA package with additional focus on Aquaphotomics „aquap2“

Package „aquap2“

- still in beta
- yet already powerful and extremely flexible & versatile

Main Advantage

- can dramatically speed up analysis time
- highly repetitive tasks get completely scriptable, i.e. automated

MVA Methods

Already Implemented

- Data pre-treatment
 - smoothing, 2nd derivative, SNV, MSC, EMSC, deTrend, gap derivative, ...
- PCA; SIMCA; PLSR
- Aquagram (classic & extended)

In the Pipeline

- PLS-DA; ANN; SVM; ICA

Key Features I

Experiment Design

- truly randomize samples
- facilitates time resolved experiments

Data Import & Organization

- import of metadata: class- and numerical variables
- import of spectral data
- consistent coloring
- align environmental data (T, rH) to timestamp on spectra
- add custom, device-specific noise to dataset

Data Analysis & Visualization

- **Bootstrapped** Aquagram: custom temperature calibration data
- **User-defined splitting of dataset by wavelength or any class-variable**
- **Highly flexible input and control system**

Key Features II: User-defined splitting of dataset

C_Time	C_Group	w1300	w1300.5	w1301	w...	w1600
T0	Treatment_AAA	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_AAA	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_AAA	0.2362	0.2345	0.4352	0.2362	0.2362
T0	Treatment_BBB	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_BBB	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_BBB	0.2362	0.2345	0.4352	0.2362	0.2362
T0	Treatment_CCC	0.2345	0.4352	0.2362	0.4352	0.2345
T0	Treatment_CCC	0.4352	0.2362	0.2345	0.2362	0.4352
T0	Treatment_CCC	0.2362	0.4352	0.4352	0.2345	0.2362
T1	Treatment_AAA	0.2345	0.2362	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.4352	0.2345	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.2362	0.4352	0.2345	0.2362	0.2345
T1	Treatment_BBB	0.2345	0.2362	0.4352	0.2362	0.4352
T1	Treatment_BBB	0.4352	0.2345	0.2362	0.2345	0.2362
T1	Treatment_BBB	0.2362	0.4352	0.2345	0.4352	0.2362
T1	Treatment_CCC	0.4352	0.2362	0.4352	0.2362	0.2345
T1	Treatment_CCC	0.2362	0.2345	0.4352	0.2362	0.2362
T1	Treatment_CCC	0.2345	0.4352	0.2362	0.2345	0.2345

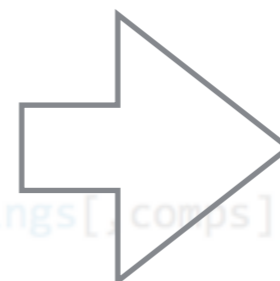
Key Features II: User-defined splitting of dataset

Split by „C_Time“

C_Time	C_Group	w1300	w1300.5	w1301	w...	w1600
T0	Treatment_AAA	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_AAA	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_AAA	0.2362	0.2345	0.4352	0.2362	0.2362
T0	Treatment_BBB	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_BBB	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_BBB	0.2362	0.2345	0.4352	0.2362	0.2362
T0	Treatment_CCC	0.2345	0.4352	0.2362	0.4352	0.2345
T0	Treatment_CCC	0.4352	0.2362	0.2345	0.2362	0.4352
T0	Treatment_CCC	0.2362	0.4352	0.4352	0.2345	0.2362
T1	Treatment_AAA	0.2345	0.2362	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.4352	0.2345	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.2362	0.4352	0.2345	0.2362	0.2345
T1	Treatment_BBB	0.2345	0.2362	0.4352	0.2362	0.4352
T1	Treatment_BBB	0.4352	0.2345	0.2362	0.2345	0.2362
T1	Treatment_BBB	0.2362	0.4352	0.2345	0.4352	0.2362
T1	Treatment_CCC	0.4352	0.2362	0.4352	0.2362	0.2345
T1	Treatment_CCC	0.2362	0.2345	0.4352	0.2362	0.2362
T1	Treatment_CCC	0.2345	0.4352	0.2362	0.2345	0.2345

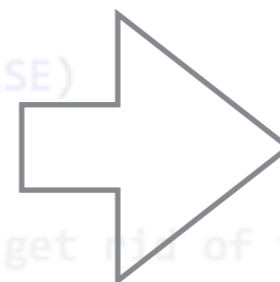
Key Features II: User-defined splitting of dataset

C_Time	C_Group	w1300	w1300.5	w1301	w...	w1600
T0	Treatment_AAA	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_AAA	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_AAA	0.2362	0.2345	0.4352	0.2362	0.2362
T0	Treatment_BBB	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_BBB	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_BBB	0.2362	0.2345	0.4352	0.2362	0.2362
T0	Treatment_CCC	0.2345	0.4352	0.2362	0.4352	0.2345
T0	Treatment_CCC	0.4352	0.2362	0.2345	0.2362	0.4352
T0	Treatment_CCC	0.2362	0.4352	0.4352	0.2345	0.2362



Separate Analysis and Visualization

T1	Treatment_AAA	0.2345	0.2362	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.4352	0.2345	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.2362	0.4352	0.2345	0.2362	0.2345
T1	Treatment_BBB	0.2345	0.2362	0.4352	0.2362	0.4352
T1	Treatment_BBB	0.4352	0.2345	0.2362	0.2345	0.2362
T1	Treatment_BBB	0.2362	0.4352	0.2345	0.4352	0.2362
T1	Treatment_CCC	0.4352	0.2362	0.4352	0.2362	0.2345
T1	Treatment_CCC	0.2362	0.2345	0.4352	0.2362	0.2362
T1	Treatment_CCC	0.2345	0.4352	0.2362	0.2345	0.2345



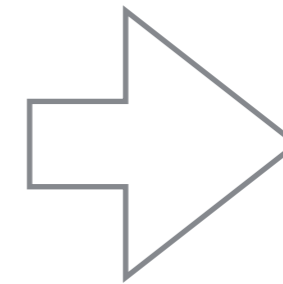
Key Features II: User-defined splitting of dataset

Split by „C_Group“

C_Time	C_Group	w1300	w1300.5	w1301	w...	w1600
T0	Treatment_AAA	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_AAA	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_AAA	0.2362	0.2345	0.4352	0.2362	0.2362
T0	Treatment_BBB	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_BBB	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_BBB	0.2362	0.2345	0.4352	0.2362	0.2362
T0	Treatment_CCC	0.2345	0.4352	0.2362	0.4352	0.2345
T0	Treatment_CCC	0.4352	0.2362	0.2345	0.2362	0.4352
T0	Treatment_CCC	0.2362	0.4352	0.4352	0.2345	0.2362
T1	Treatment_AAA	0.2345	0.2362	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.4352	0.2345	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.2362	0.4352	0.2345	0.2362	0.2345
T1	Treatment_BBB	0.2345	0.2362	0.4352	0.2362	0.4352
T1	Treatment_BBB	0.4352	0.2345	0.2362	0.2345	0.2362
T1	Treatment_BBB	0.2362	0.4352	0.2345	0.4352	0.2362
T1	Treatment_CCC	0.4352	0.2362	0.4352	0.2362	0.2345
T1	Treatment_CCC	0.2362	0.2345	0.4352	0.2362	0.2362
T1	Treatment_CCC	0.2345	0.4352	0.2362	0.2345	0.2345

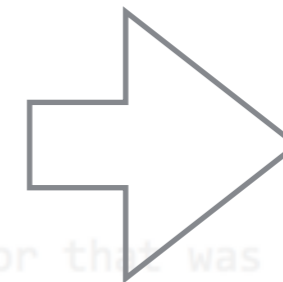
Key Features II: User-defined splitting of dataset

T0	Treatment_AAA	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_AAA	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_AAA	0.2362	0.2345	0.4352	0.2362	0.2362
T1	Treatment_AAA	0.2345	0.2362	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.4352	0.2345	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.2362	0.4352	0.2345	0.2362	0.2345



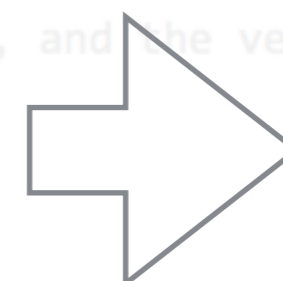
Separate Analysis and Visualization

T0	Treatment_BBB	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_BBB	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_BBB	0.2362	0.2345	0.4352	0.2362	0.2362
T1	Treatment_BBB	0.2345	0.2362	0.4352	0.2362	0.4352
T1	Treatment_BBB	0.4352	0.2345	0.2362	0.2345	0.2362
T1	Treatment_BBB	0.2362	0.4352	0.2345	0.4352	0.2362



Separate Analysis and Visualization

T0	Treatment_CCC	0.2345	0.4352	0.2362	0.4352	0.2345
T0	Treatment_CCC	0.4352	0.2362	0.2345	0.2362	0.4352
T0	Treatment_CCC	0.2362	0.4352	0.4352	0.2345	0.2362
T1	Treatment_CCC	0.4352	0.2362	0.4352	0.2362	0.2345
T1	Treatment_CCC	0.2362	0.2345	0.4352	0.2362	0.2362
T1	Treatment_CCC	0.2345	0.4352	0.2362	0.2345	0.2345



Key Features III: Input & Control

set up once,
modify rarely

Global Settings

experiments obtain parameters
from global settings

modify
experiment-specific

Experiment-wise Settings

Code-based, programmatic
overriding of parameters

modify often during
exploratory analysis

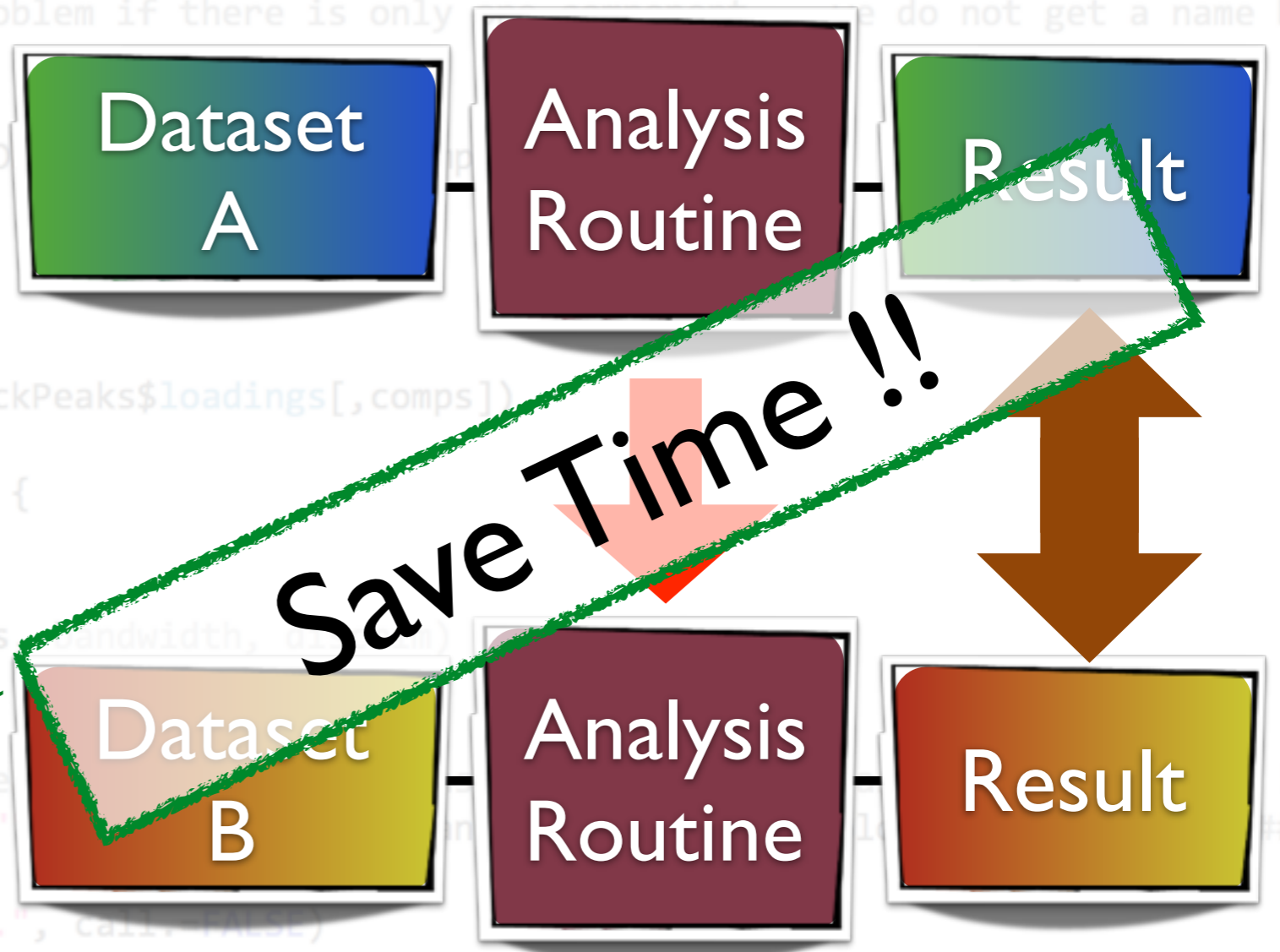
Manual, user-defined Input
when calling Functions

3 levels of control

- flexible, highly customizable
- **global settings**: change input globally in a single, central file („settings.r“)
- **experiment-wise** settings: the specific metadata and analysis procedure file for each experiment.
- **code-based** overriding of parameters: Enables programmatic, i.e. scriptable modification of input-parameters.

Summary - Main Advantages

- **fully scriptable** analysis routines
 - consisting of (user's) custom code, and
 - easy-to-use analysis modules
- **same analysis** of an exact repetition of a previous experiment can so be done **in a few seconds** - just by plugging new data into the old (copied) analysis routine



Think - Copy&Paste - Enjoy

Now: Hands-On aquap2



© StormSignal

© StormSignal

