



ONDERWIJSCONGRES 2017 ONDERWIJS IN DE TOEKOMST

DONDERDAG 19 JANUARI 2017

VERGADERCENTRUM DOMSTAD - UTRECHT

WORKSHOP 1W5

Onderwijs (en Onderzoek) Datawetenschappen in de Cloud?

- **Marc Teunis, Ph.D.;** Hogeschoolhoofddocent Life Sciences, Senior Scientist Lectoraat Innovative Testing, Immunoloog, Klimmer
- **Mathijs van de Venne,** Webdeveloper, Everweb, Klimmer



innovativetesting.nl

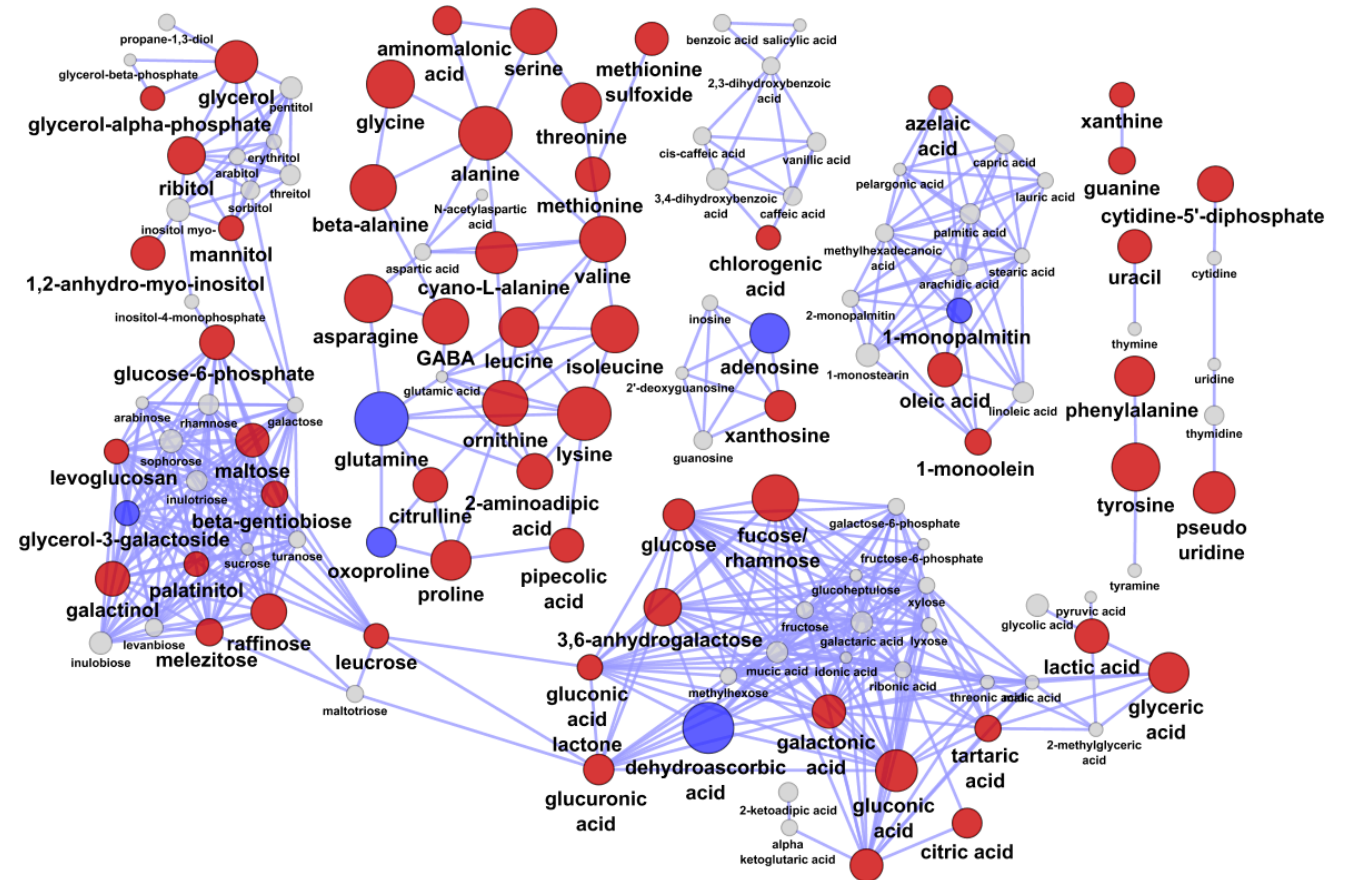


everweb.nl



Doel van de workshop

- 1) Cloud computing voorbeeld
- 2) R in Data Science (Onderwijs)
- 3) Een paar R basics:
 - R graphics
 - Exploratory data analysis
 - Reproducible research



Clone van Github

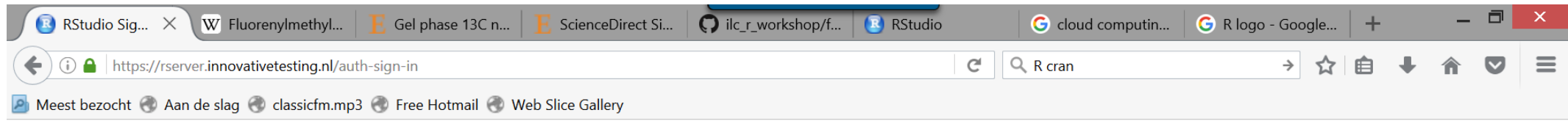
https://github.com/Maddocent/ilc_r_workshop



The screenshot shows the GitHub repository page for "Maddocent / ilc_r_workshop". The browser address bar shows the URL "https://github.com/Maddocent/ilc_r_workshop". The repository page includes a search bar, navigation links for "Pull requests", "Issues", and "Gist", and a header section with "Unwatch", "Star", and "Fork" buttons. Below the header, there are tabs for "Code", "Issues", "Pull requests", "Projects", "Wiki", "Pulse", "Graphs", and "Settings". The main content area shows the repository description "No description or website provided. — Edit" and a summary of repository statistics: 26 commits, 1 branch, 0 releases, and 1 contributor. A table of recent commits is displayed, showing the commit message, the files changed, and the time since the commit.

Commit Message	Files Changed	Time Ago
Maddocent added required packages to circos.R	code, data, documentation, images, packrat, rmarkdown, .Rprofile	2 days ago
added required packages to circos.R	code	2 days ago
gitignore	data	12 days ago
changed some line in Rmd	documentation	3 days ago
changed loaction of files to dedicated folders	images	15 days ago
added bundle unbundle	packrat	12 days ago
added required packages to circos.R	rmarkdown	2 days ago
added required packages to circos.R	.Rprofile	2 days ago

<https://rserver.innovativetesting.nl>



Login gegevens:
papiertje

Sign in to RStudio

Username:
FNT_ond_014

Password:
.....

☐ Stay signed in

Sign In

FNT_ond_014
rprogramm#\$*014

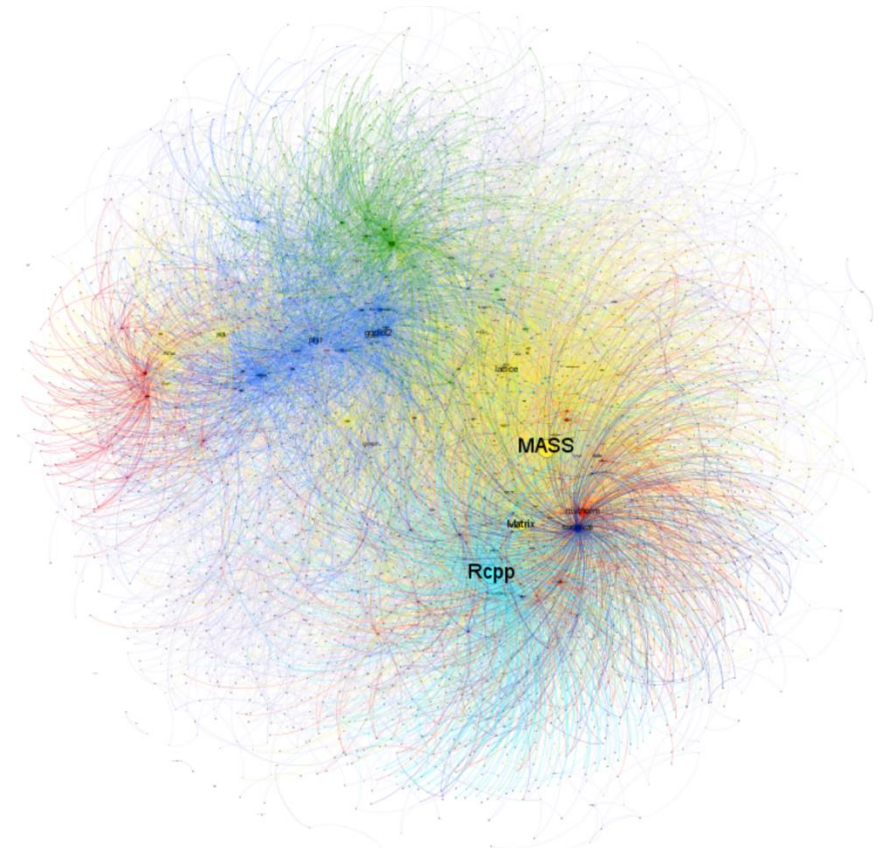


De oplossing?



Waarom R?

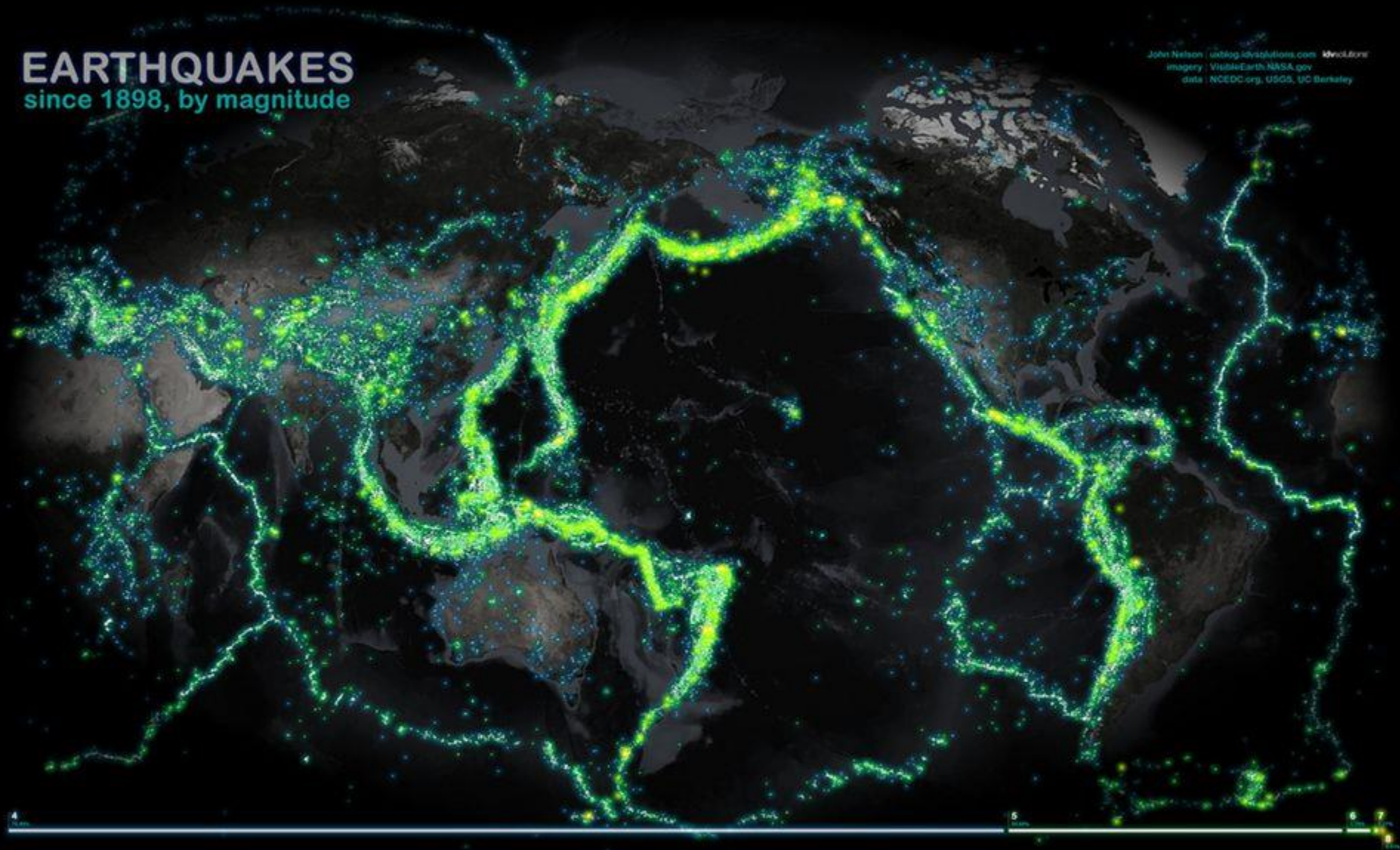
- <https://www.youtube.com/watch?v=ph7Eczs-nTI>
- Reproduceerbaarheid
- Veel applicaties en add-ons
- Gratis documentatie en educatief materiaal (web, books, moocs, courses)
- Gratis!
- Open source!
- Grote R community
- Combinatie met andere talen en 'literate programming'
- Mooie graphics



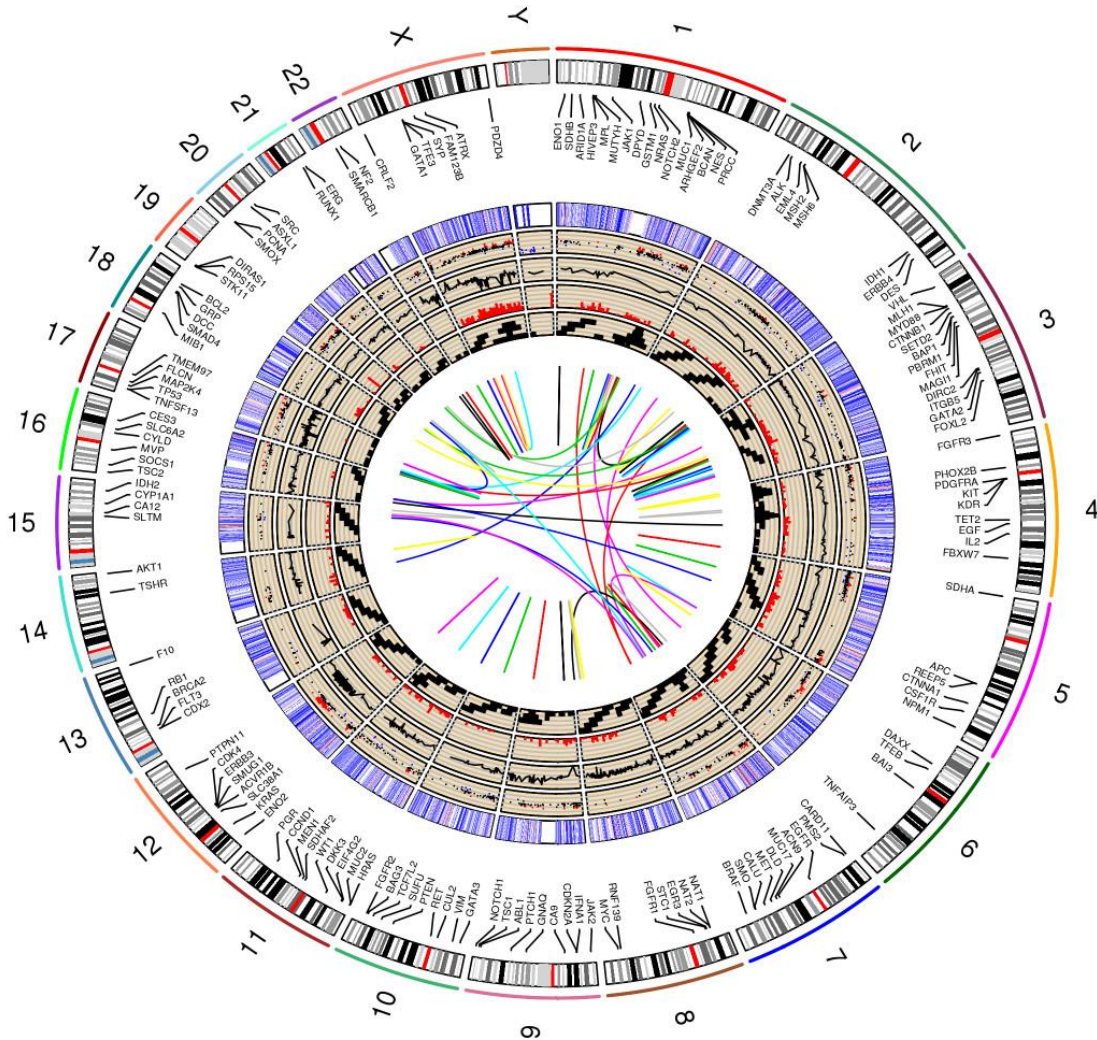
EARTHQUAKES

since 1898, by magnitude

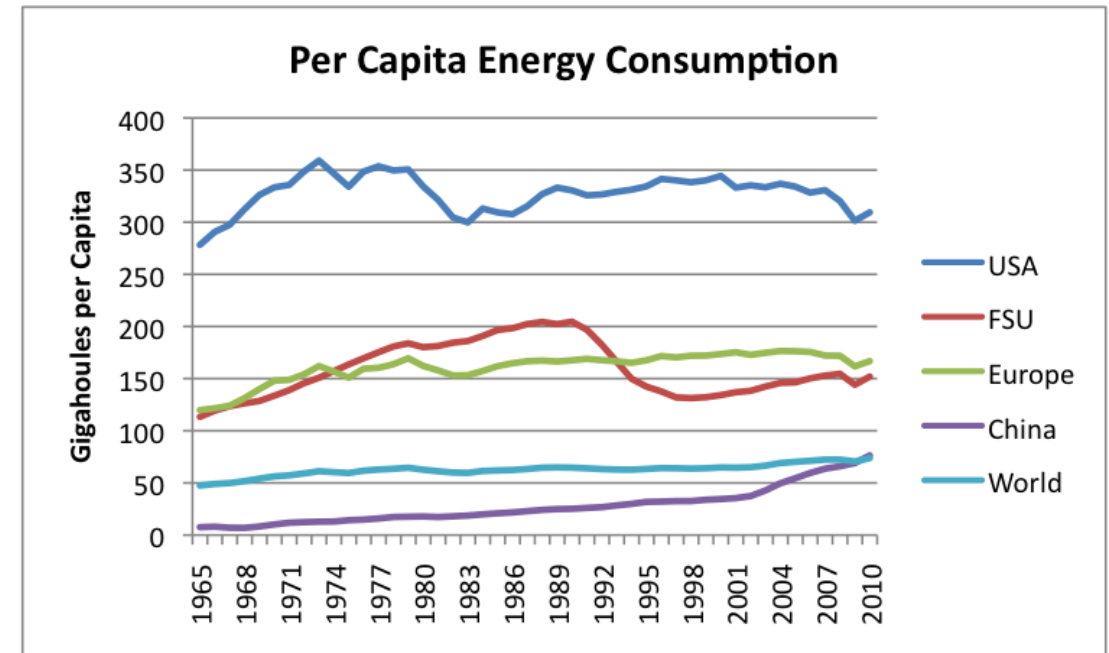
John Nelson | outblog.advantures.com | Mivolution.com
imagery: VisibleEarth.nasa.gov
data: NCEDC.org, USGS, UC Berkeley



RCircos 2D Track Plot with Human Genome



Complex data / Big Data



```
library(circlize)
```

```
# dend: as dendrogram object, usually returned by hclust and as.dendrogram
# maxy: maximum height of the tree
circos.dendrogram = function(dend, maxy) {
  labels = as.character(labels(dend))
  x = seq_along(labels) - 0.5
  names(x) = labels
  is.leaf = function(object) (is.logical(L <- attr(object, "leaf"))) && L
  # recursive function to draw the tree
```

```
draw.d = function(dend, maxy) {
  leaf = attr(dend, "leaf")
  d1 = dend[[1]]
  d2 = dend[[2]]
  height = attr(dend, 'height')
  midpoint = attr(dend, 'midpoint')
  if(is.leaf(d1)) {
    x1 = x[as.character(attr(d1, "label"))]
  } else {
    x1 = attr(d1, "midpoint") + x[as.character(labels(d1))[1]]
  }
  y1 = attr(d1, "height")
  if(is.leaf(d2)) {
    x2 = x[as.character(attr(d2, "label"))]
  } else {
    x2 = attr(d2, "midpoint") + x[as.character(labels(d2))[1]]
  }
  y2 = attr(d2, "height")
  circos.lines(c(x1, x1), maxy - c(y1, height), straight = TRUE)
  circos.lines(c(x1, x2), maxy - c(height, height))
  circos.lines(c(x2, x2), maxy - c(y2, height), straight = TRUE)

  if(!is.leaf(d1)) {
    draw.d(d1, maxy)
  }
  if(!is.leaf(d2)) {
    draw.d(d2, maxy)
  }
}
```

```
draw.d(dend, maxy)
```

```
}
mat = matrix(rnorm(100*10), nrow = 10, ncol = 100)
factors = rep(letters[1:2], 50)
par(mar = c(1, 1, 1, 1))
circos.par(cell.padding = c(0, 0, 0, 0), gap.degree = 5)
circos.initialize(factors, xlim = c(0, 50))
maxy = 0
```

```
f = colorRamp2(breaks = c(-1, 0, 1), colors = c("green", "black", "red"))
```

```
circos.trackPlotRegion(ylim = c(0, 10), bg.border = NA, panel.fun = function(x, y) {
  sector.index = get.cell.meta.data("sector.index")
  m = mat[, factors == sector.index]
```

```
dend.col = as.dendrogram(hclust(dist(t(m))))
```

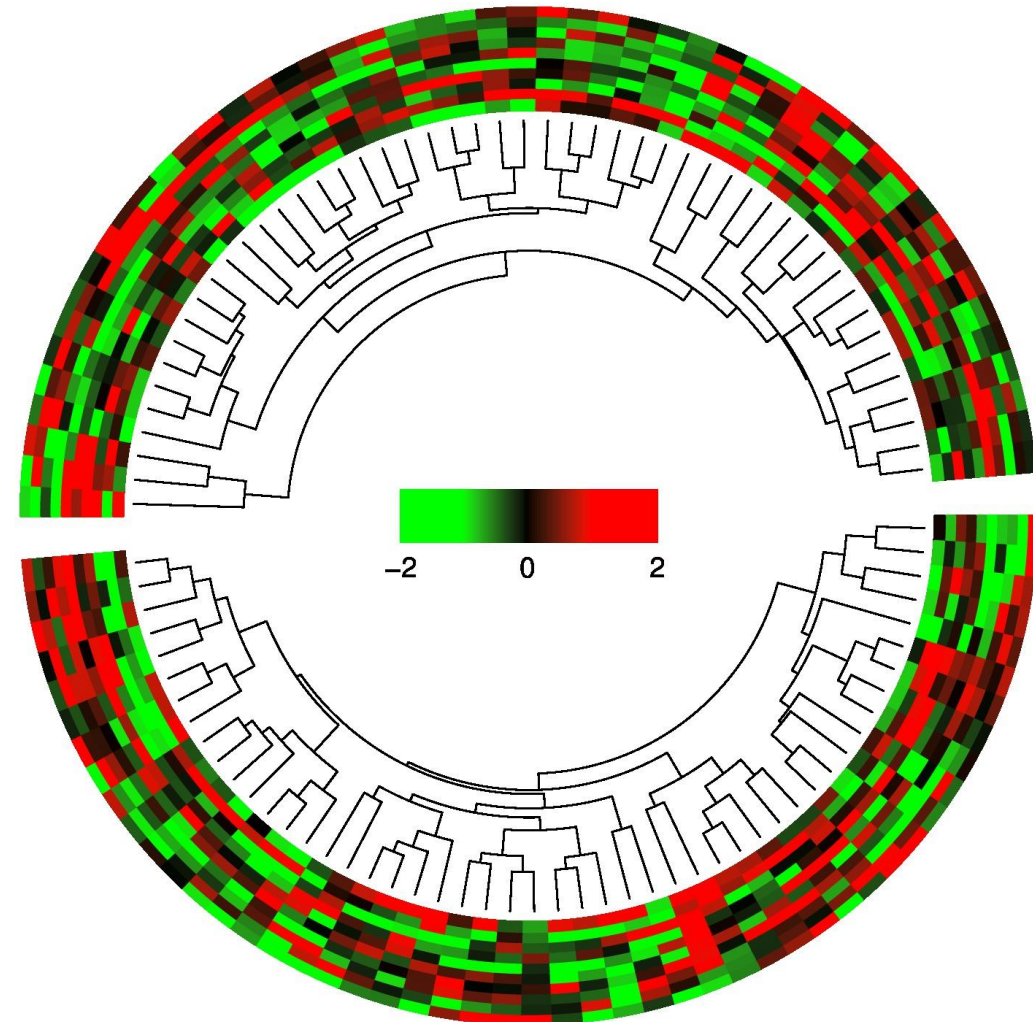
```
maxy = ifelse(maxy > attr(dend.col, "height"), maxy, attr(dend.col, "height"))
assign("maxy", maxy, envir = .GlobalEnv)
```

```
m2 = m[, labels(dend.col)]
```

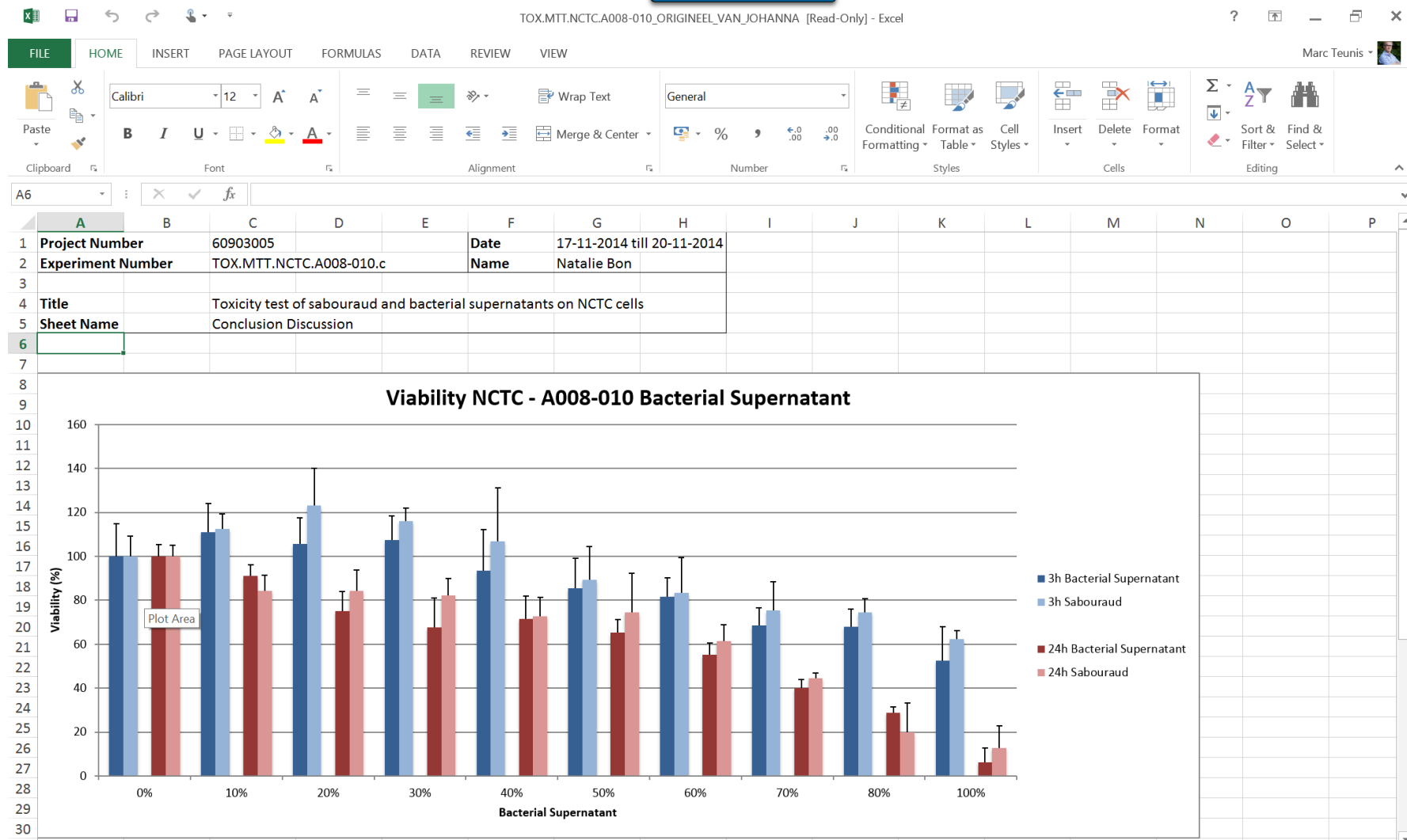
```
nr = nrow(m2)
```

```
nc = ncol(m2)
```

Reproduceren

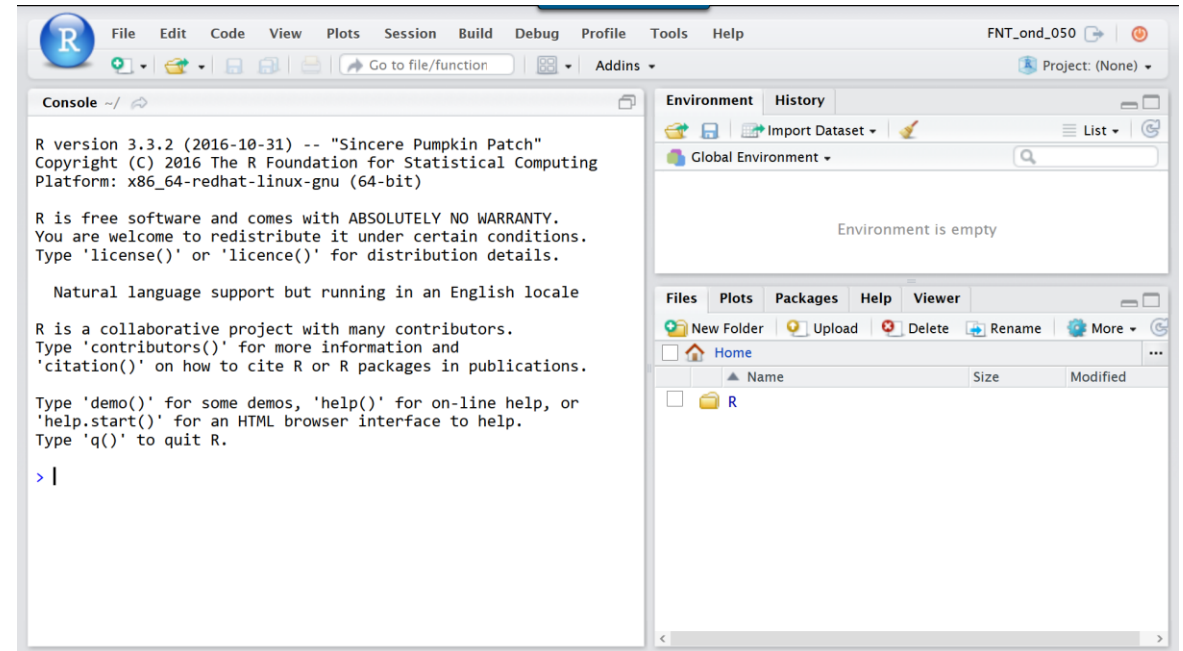


MS Excel Grafieken namaken?

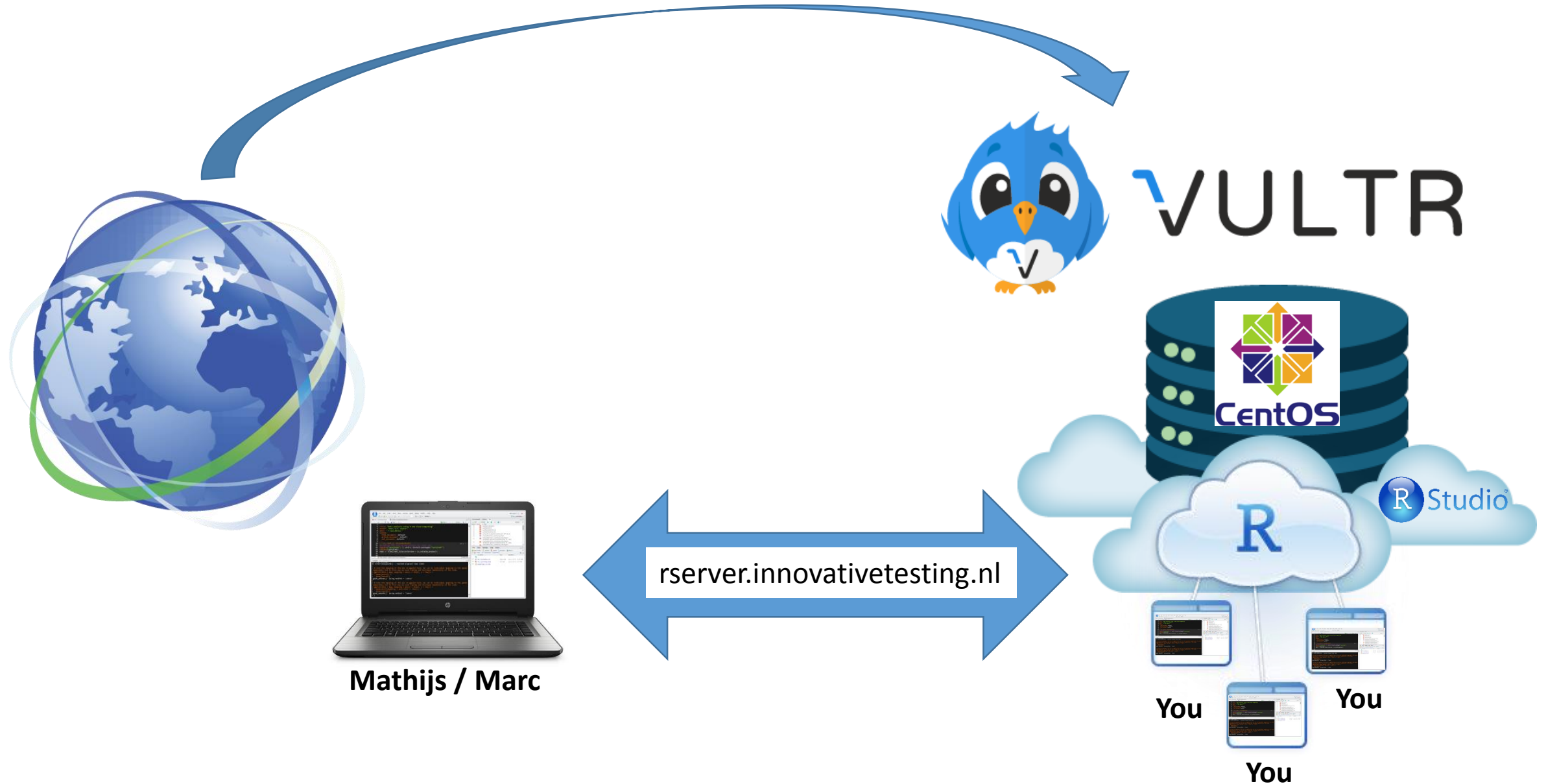


Cursus Tools: Programmeren in R

- Jupyter Notebooks on Amazon's AWS (EC2 - Nano) instances
- Stand alone installaties van R en R's IDE RStudio
- **Github/Bitbucket repositories (`git clone`, `git pull`, `git push`)**
- Vagrant box - Oracle VM
- RStudio AML on Amazon AWS
- RStudio Server corporate
- Docker images
- **VULTR RStudio Cloud server**



RStudio Cloud server



Interactive demo

The screenshot displays the RStudio IDE interface with the following components:

- Main Editor:** Displays a markdown file named `fnt_r_workshop.Rmd`. The content includes a YAML header with fields for `title`, `author`, and `date`, followed by an `output` section for HTML and GitHub documents, and a CSS style block.
- Environment Pane:** Shows the current environment with variables like `b`, `bb`, `cov3R`, `coverplot`, `d`, `dd1`, `dd2`, `dictionary`, `first_time`, `gene`, `ir`, `ir_list`, and `match`, each with its corresponding R class or value.
- Console:** Shows the execution of R code, including the creation of a package maintainer object and a warning message: "Warning message: In citation('sequencing') : no date field in DESCRIPTION file of package 'sequencing'".
- Files Pane:** Displays the file structure of the `fnt_r_workshop` directory, including `fnt_r_workshop.html`, `fnt_r_workshop.md`, `fnt_r_workshop.Rmd`, `fnt_r_workshop_files`, and `formula_example.Rmd`.