

# Visualising big data in R

H. Wickham<sup>a</sup>

RStudio  
h.wickham@gmail.com

**Mots clefs** : Visualisation, big data

R has a notorious reputation for not being able to deal with "big" data (and ggplot2 is a frequent culprit). Fortunately, this isn't an underlying problem with R, and it's something that we can fix with good programming practices and intelligent use of compiled code. In this talk, I'll introduce a new package, bigvis, that aims to make it easier (and faster) to work with very large datasets.

Bigvis makes it possible to visualise 10-100 million observations in just a few seconds. It is built around a pipeline of bin, summarise, smooth and visualise, and makes minimal sacrifices of flexibility to achieve fast performance. As well as discussing the visualisation challenges when you have 10s of millions of observations, I'll also discuss the performance challenges, and how C++ and Rcpp make it pleasurable to integrate compiled code into R.