From Momocs to MomX

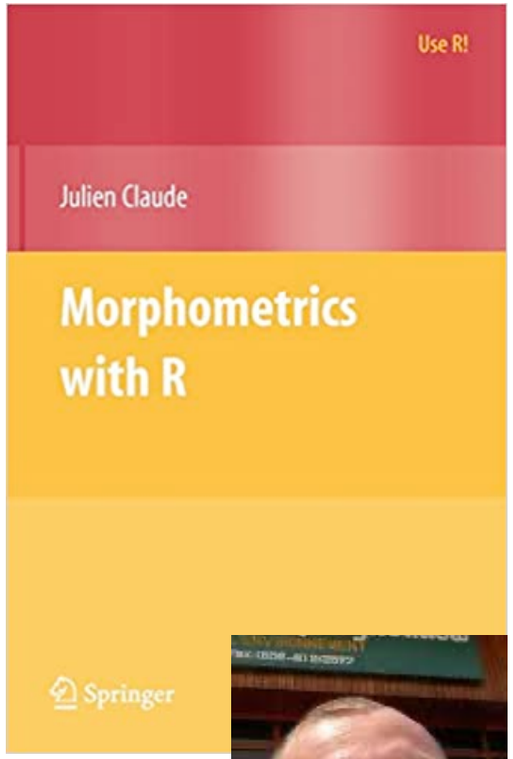
Design principles

Vincent Bonhomme

SMEF 2021 - online ed

History

2008



J Claude

mostly wrapping

mostly wrapping

2014

Journal of Statistical Software
February 2014, Volume 56, Issue 13. *http://www.jstatsoft.org/*

Momocs: Outline Analysis Using R

Vincent Bonhomme
French Institute of Pondicherry

Sandrine Picq
UMR CBAE, Montpellier

Cédric Gaucherel
French Institute of Pondicherry

Julien Claude
University of Montpellier II

+ V Bonhomme

complete refactoring

2021 (?)

+ A Evin

From the very beginning, Momocs aimed to be an unified and complete environment for morphometrics in R

What's fine with Momocs ?

```
library(Momocs)
```

```
bot %>% coo_center %>% efourier %>% PCA %>% LDA(~type) %>% plot_CV
```

Consistant, compact, pipe-friendly grammar

What's fine with Momocs ?

```
library(Momocs)
```

raw coordinates + cofactors/variables

geometrical operations

Principal Component Analysis

Graphical confusion matrix

```
bot %>% coo_center %>% efourier %>% PCA %>% LDA(~type) %>% plot_CV
```

magrittr's pipe

elliptical Fourier transforms

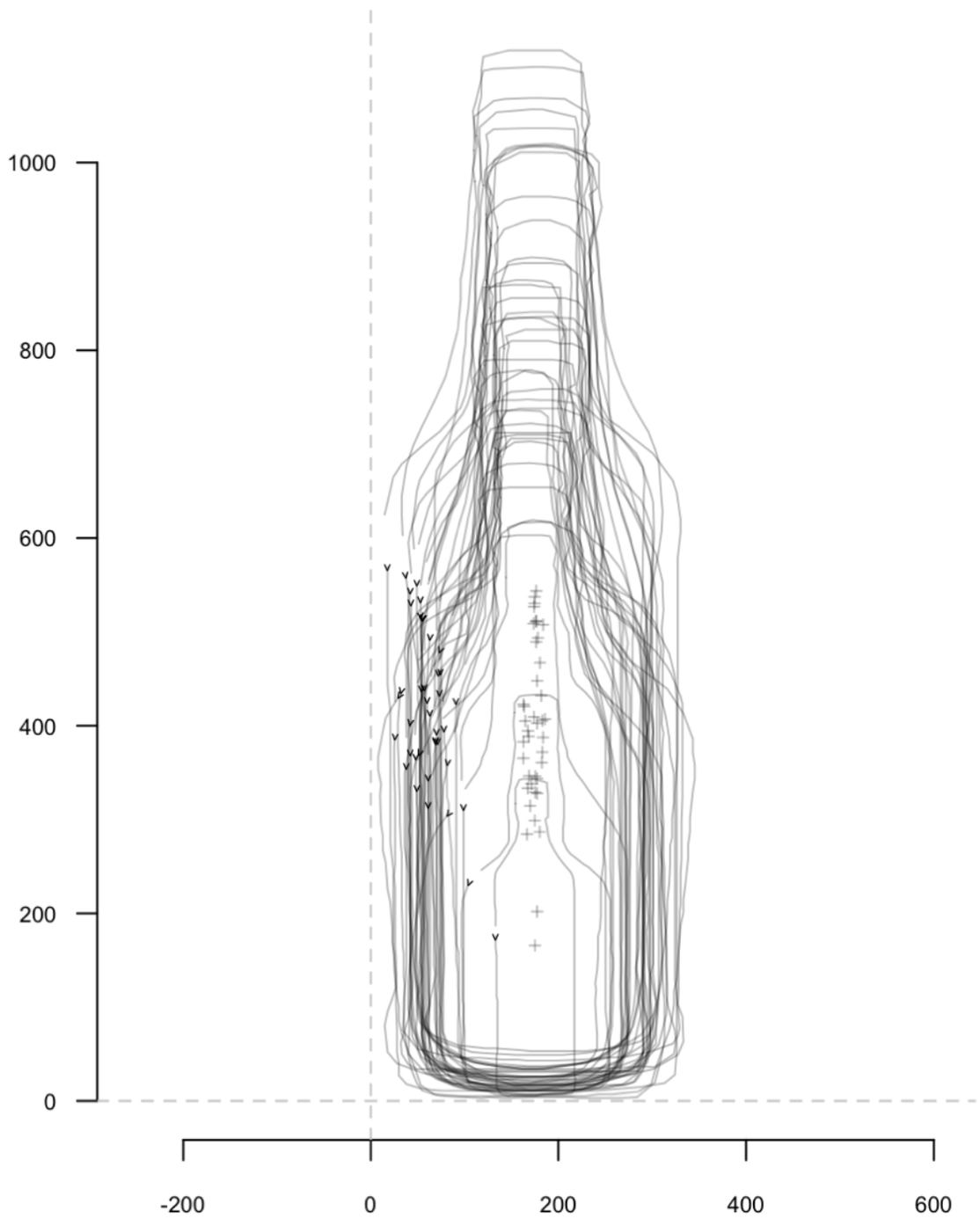
Linear Discriminant Analysis on PC scores

Consistant, compact, pipe-friendly grammar

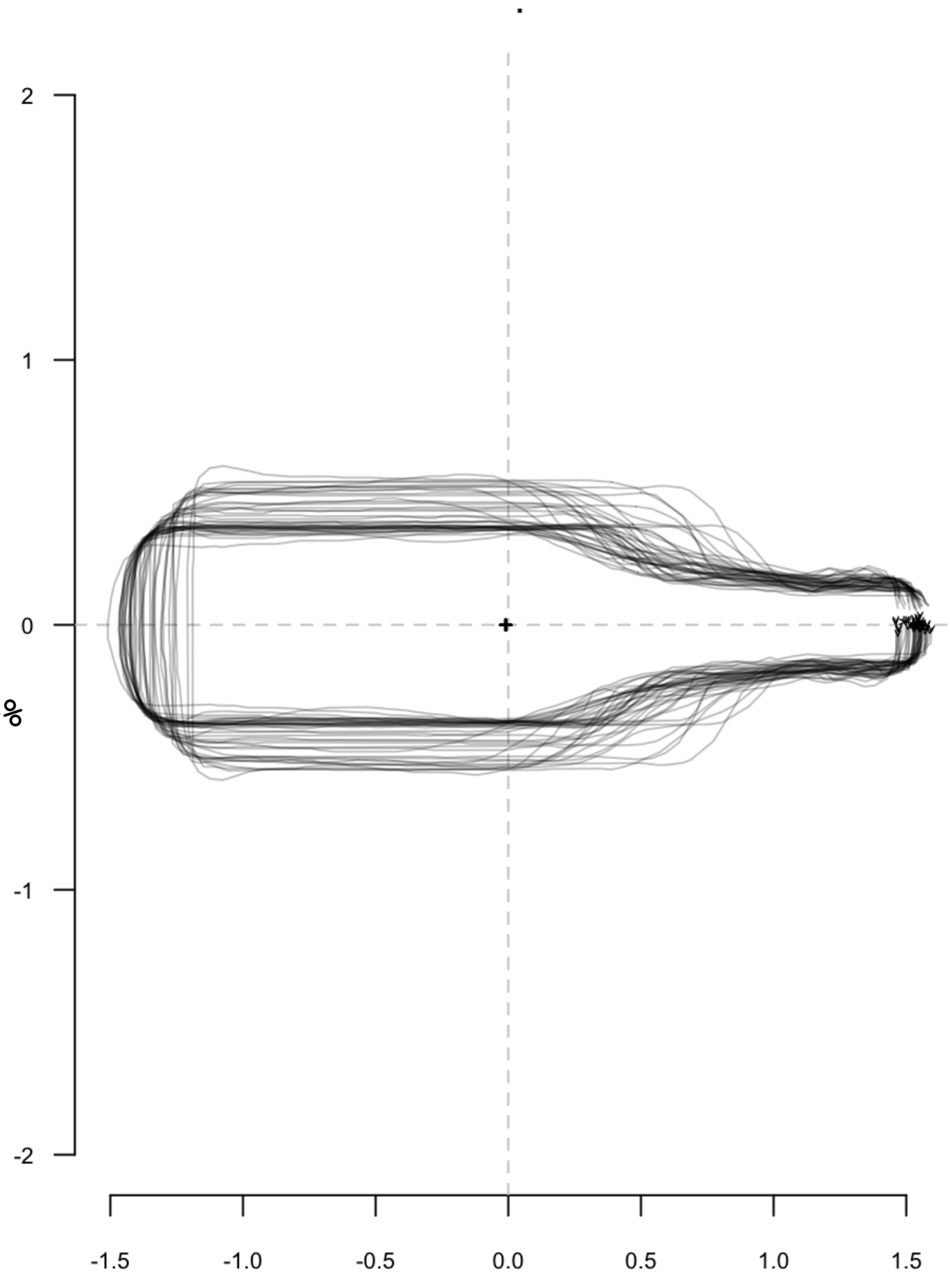
> bot

```
Out (outlines)
  - 40 outlines, 162 +/- 21 coords
  - 2 classifiers (in $fac):
# A tibble: 40 x 2
  type   fake
  <fct>  <fct>
1 whisky a
2 whisky a
3 whisky a
4 whisky a
5 whisky a
6 whisky a
# … with 34 more rows
```

bot %>% stack



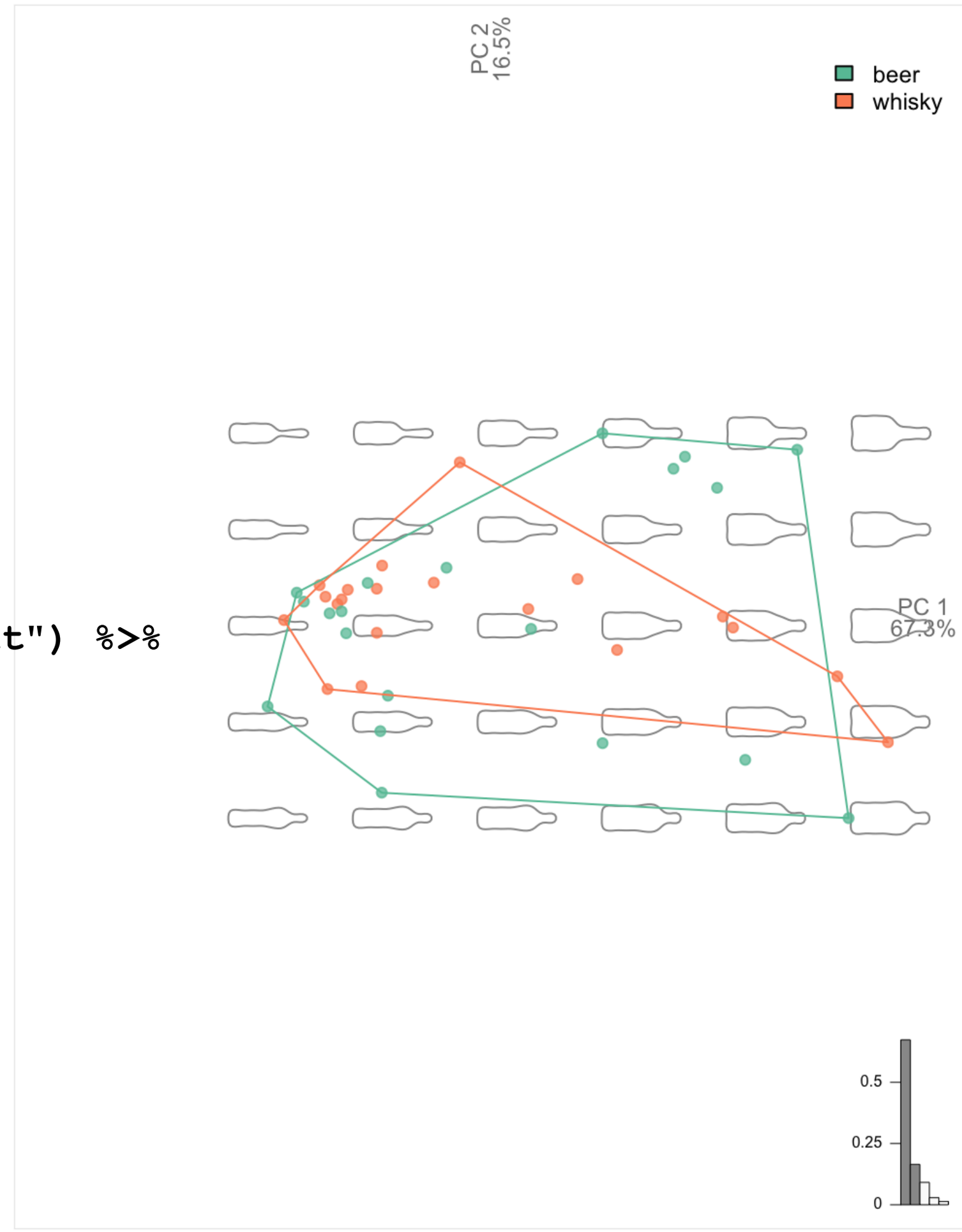
```
bot %>%
  coo_center %>%
  coo_align %>%
  coo_scale %>%
  coo_slidedirection("right") %>%
  stack
```



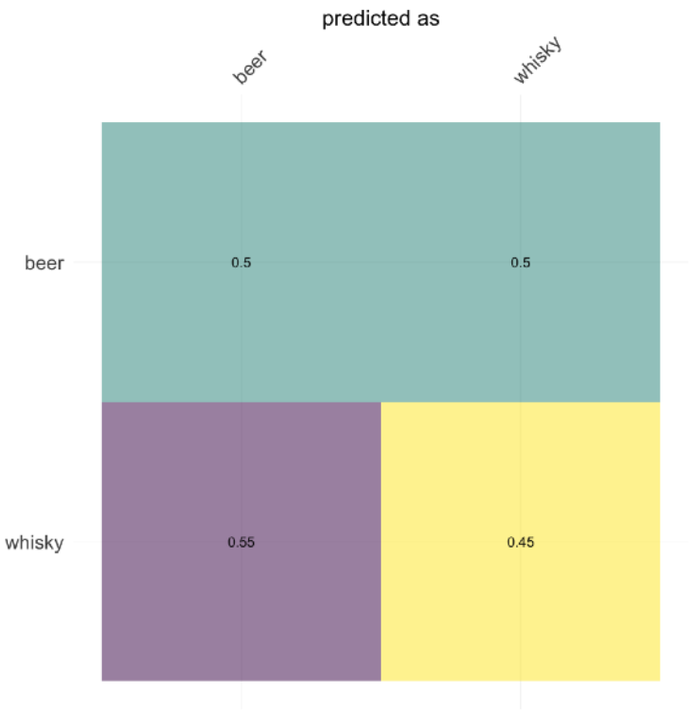
```
bot %>%
  coo_center %>%
  coo_align %>%
  coo_scale %>%
  coo_slidedirection("right") %>%
  efourier(norm=FALSE)

'nb.h' set to 10 (99% harmonic power)
An OutCoe object [ elliptical Fourier analysis ]
--------------------
 - $coe: 40 outlines described, 10 harmonics
# A tibble: 40 x 2
  type   fake
  <fct>  <fct>
1 whisky a
2 whisky a
3 whisky a
4 whisky a
5 whisky a
6 whisky a
# … with 34 more rows
```

```
bot %>%
  coo_center %>%
  coo_align %>%
  coo_scale %>%
  coo_slidedirection("right") %>%
  efourier(norm=FALSE)%>%
  PCA %>%
  plot_PCA(~type)
```



```
bot %>%
  coo_center %>%
  coo_align %>%
  coo_scale %>%
  coo_slidedirection("right") %>%
  efourier(norm=FALSE)%>%
  PCA %>%
  LDA(~type) %>%
  plot_CV()
```




```
 * Cross-validation table ($CV.tab):
        classified
actual   beer whisky
  beer     10     10
  whisky   11      9

 * Class accuracy ($CV.ce):
  beer whisky
  0.50   0.45

 * Leave-one-out cross-validation ($CV.correct): (47.5% - 19/40):
```

What's wrong with Momocs ?

in brief

Not that bad for the regular front-user,
Quite boring for serious coders,
A nightmare for developers (ie myself)

- **Continuously written for 10 years**
  > R (and *my* R) have improved a LOT in the meantime
  > poorly integrate with the tidyverse so far because objects are not native data.frames
  > For complex projects (ie my post-docs) the time *patching* >>> time *exploring :(*
  > *I spend more time handling the email hotline than improving the doc or the code :(*

- **Too big (~20k lines of code + doc)**
  > Hard to extend, to test, to improve
  > API, syntax, …, became quite inconsistent

- **Inherent philosophy was outdated**
  > `data.frame` is the natural object for data science (incl. morphometrics)
  > `tidyverse` is the natural ecosystem for data science (incl. morphometrics)
  > small is beautiful

Hence the motivation for a complete refactoring of Momocs

General architecture

Core

install, load and
update them all

typical path

Acquisition

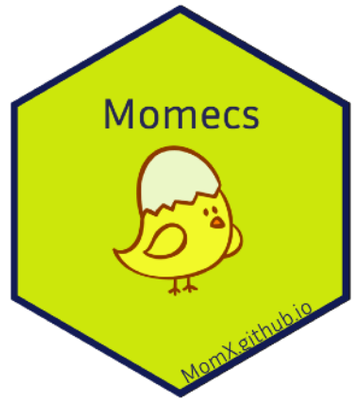
Deep learning - fuelled
geometry extraction

a cross-platform and
open alternative to tspDig

foreign file format to R
(and back)

Morphometrics

geometry toolbox
and morphometrics



interactive
shiny version

Analyses

statistical analyses

+Momoshop, +Mombank, +Momdata

Design principles

1. **Everything is a `tibble` (which is a `data.frame`, and also a `list`)**
   `magrittr, ggplot2, dplyr, tidyr, stringr, purrr, tidyeval`

2. **Every package (and function) does a single task and does it well**
   small is beautiful
   a single bug does not turn down everything

3. **Everything is archived, continuously integrated, tested and documented**
   coding locally -> github -> travis-ci -> code coverage -> pkgdown
   fully automated

momx.github.io

What MomX aims to become

- **Complete**
  > data acquisition and post-processing (**Momacs**)
  > data sharing/import/export, checking, handling (**Momit**)
  > geometric operations on shapes and morphometrics (**Momocs2**)
  > statistics (**Momstats**)
  > interactive version of MomX (**Momecs**)
  > automated reporting (**Momscribe**)
  > share datasets and manage collections (**Mombank**)

- **Convenient and scalable**
  > Arrive and leave whenever you want
  > Interoperable packages (tidyverse ready)
  > Easier to optimise (C++)
  > Packages can be dockerised, deployed
  > Foundations to handle millions of shapes

- **Open-source and gratis**
  > Free plan: written in pure R + GPL3
  > Academic plan: citations + collaborations
  > Training plan: to R and MomX

to fund myself
and MomX development

Towards release

2021

v 0.1 (Momit, Momocs2, Momstats)
release on CRAN
companion paper

State of the union

pkg	website	github	travis	last commit	code coverage
MomX	go	go	build error	last commit april	codecov 24%
Momit	go	go	build passing	last commit may	codecov 14%
Momacs	go	go	build passing	last commit march	-
Momocs2	go	go	build passing	last commit may	codecov 78%
Momstats	go	go	build passing	last commit april	codecov 43%
Momdata	go	go	build passing	last commit april	codecov 100%

2022

fund myself and MomX
Momacs, Momecs, Momdeep
more tutorials
fill the gaps

many details and working things there:

momx.github.io

feel free to join whatever your skills are!

bonhomme.vincent@gmail.com