

tour

Hadley Wickham

Assistant Professor / Dohelman Family Junior Chair
Department of Statistics / Rice University

1. Getting started

2. Saving output

3. Recording paths

4. How it works

5. Learn more

H. Wickham, D. Cook, H. Hofmann, A. Buja.
tourr: An R package for exploring multivariate

```
all.packages("tourr")
for development version
all.packages("tourr",
  repos = "http://r-forge.r-project.org")

every time you load R
library(tourr)

a
ies <- as.numeric(flea$species)
$species <- NULL
(flea)
(flea, col = species)
```

```
ate_xy(flea, grand_tour(d = 2))  
ate_depth(flea, grand_tour(d = 3))  
ate_pcp(flea, grand_tour(d = 4))
```

ess control + break to stop the tour
te: This is research-grade software. Expect
oblems.

t what you learn here will continue to apply
we build better tools for data analysts. T
eory is solid.

Display

Display is terribly crude. Back to the 1980's of computer graphics. But it's a start! And because it's all R code, it's very easy to write and to distribute.

Currently working on high-performance graphics canvas that can quickly draw hundreds of thousands of points.

ick demo

ce("~/Documents/cranvas/demos/tourr-gui.r")

xy(olive)

xy(diamonds)

efully available for everyone to use by the
d of this summer

```
ate_xy(flea, col = species, grand_tour())
```

```
ate_xy(flea, col = species,  
guided_tour(lda_pp(species)))
```

```
ate_xy(flea, col = species, guided_tour(hole
```

in this a few times. What do you observe?
at variables are important?

Saving output


```
ate_xy(flea, grand_tour())

# output to pdf
er(flea, grand_tour(), display_xy(),
df", "flea-tour.pdf")

# control the number of frames
er(flea, grand_tour(), display_xy(),
df", "flea-tour.pdf", frames = 100)

# output to a series of pngs
er(flea, grand_tour(), display_xy(),
ng", "flea-tour-%03d.png")
```

Recording tour paths

```
cording tour paths
resting <- Ida_pp(species)
ies_Ida <- guided_tour(interesting)
path <- save_history(flea, species_Ida)

(path_index(one_path, interesting))
(path_index(interpolate(one_path), interesting))
```

```
ording many tour paths
s <- replicate(100, save_history(
ea, species_Ida, max_bases = 10))
ces <- paths_index(paths, interesting)
: ("Ida-paths.rdata")
t(step, value, data = indices,
oup = try, geom = "line")
t(step, improvement, data = indices,
oup = try, geom = "line")
```

Are there those best indices?

Ly(paths, function(x) x[, , dim(x)[3]],

top = F])

How to make sense of this is an open question!

Some ideas:

Use Frobenius distance + mds / clustering

Interact with plot of indices and projected

How it works

```
ate_xy(flea, col = species, grand_tour())
```

port hand for:

```
ate(  
  flea, # data  
  and_tour(), # tour path  
  splay_xy(col = species) # display method
```

Ordering tour

pieces

```
# The data
flea
str(flea)

# The tour path
grand_tour() in tour-grand.r

# The display method
display_xy() in display-xy.r
```


Read the code!

One of the best ways to improve your R skills. I have spent a lot time rewriting the tour code so it is as clear and readable as possible. (Code is communication too)

Makes precise some of the details that are omitted in the paper. Shows how to deal with problems that arise in practice.

Your turn

Downloading the source package from <http://cran.r-project.org/web/packages/tourr>. Unzip and then read the source code for the following files:

```
tour-local.r
```

```
display-pcp.r
```

```
display-depth.r
```

More code

```
animate() in animate.r  
new_tour() in tour.r  
geodesic_path() in geodesic.r  
geodesic_info() in geodesic.r
```

New methods

Previously, if you wanted to contribute a new display method or new tour path, you'd have to write all the other pieces yourself (probably in C).

With `tourr` you can develop new methods but plug into the existing infrastructure.

As a proof of concept, we've included code for all tours that we know about.

Guided tour

Very easy to write a new index: see [interesting-indices.r](#) for examples

A little harder to write new optimisation functions: see [search-better.r](#) for examples.

Learning more

Key papers

buja-1996.pdf “*Interactive high-dimensional data visualization*”. Motivation, examples and tools.

buja-2004.pdf “*Theory of dynamic projections in high-dimensional data visualization*”. In depth theory

wickham-2010.pdf “*tourr: An R package for exploring multivariate data with projections*”

Other perspectives

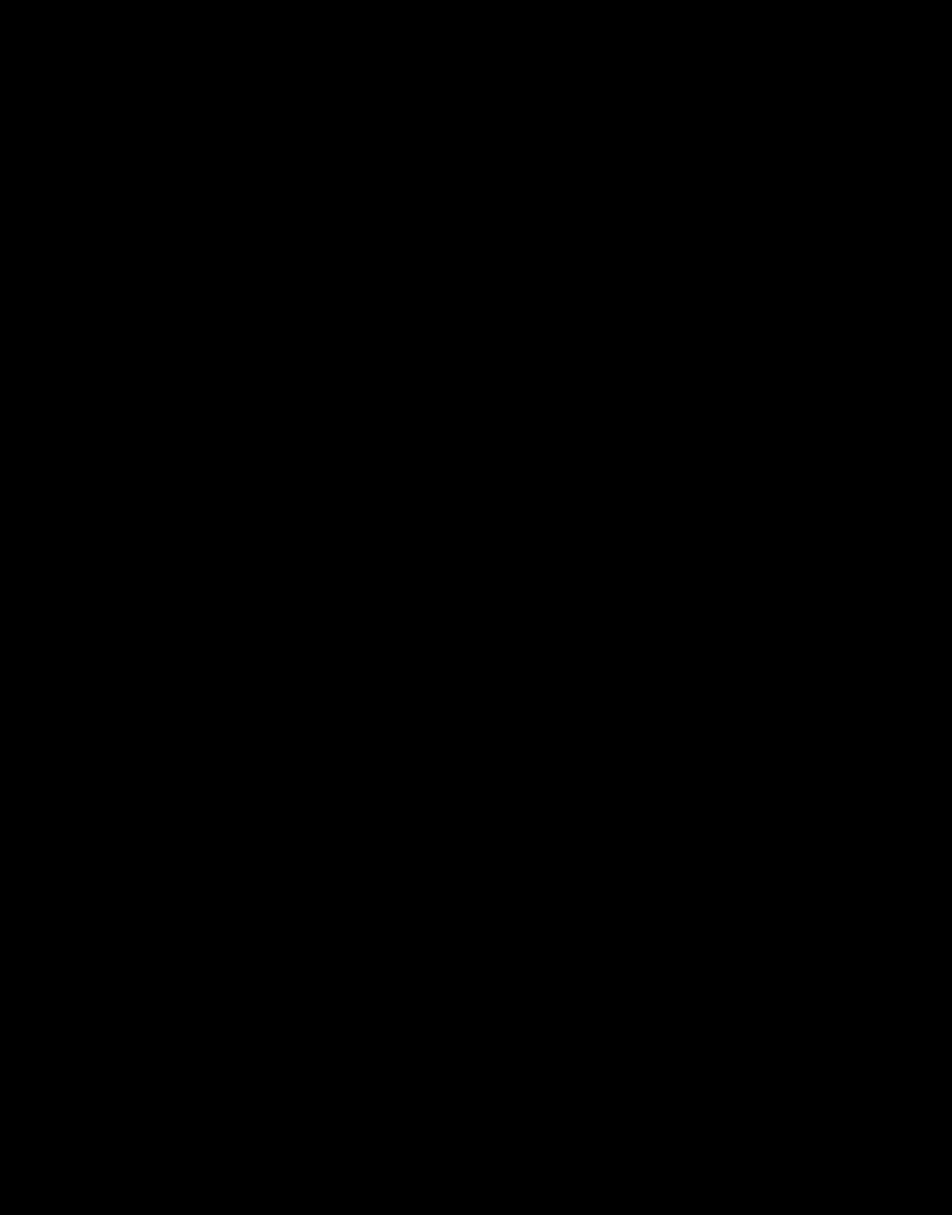
wegman-2002.pdf “*On some mathematics for visualizing high dimensional data*”. Same motivation, but different mathematical approach.

symanzik-2002.pdf “*New applications on the image grand tour*”. Mapping the output of a tour to colour

BOOKS

D. Cook and D. F. Swayne. *Interactive and Dynamic Graphics for Data Analysis: With Examples Using R and GGobi*. Springer, 2007.

Highly recommended if you want to learn more about using interactive and dynamic graphics to better understand high-d data.



This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/3.0/> or send a letter to Creative Commons, 545 Third Street, Suite 300, San Francisco, California, 94105, USA.