The future of interactive graphics in R

Hadley Wickham

Assistant Professor / Dobelman Family Junior Chair

Department of Statistics / Rice University

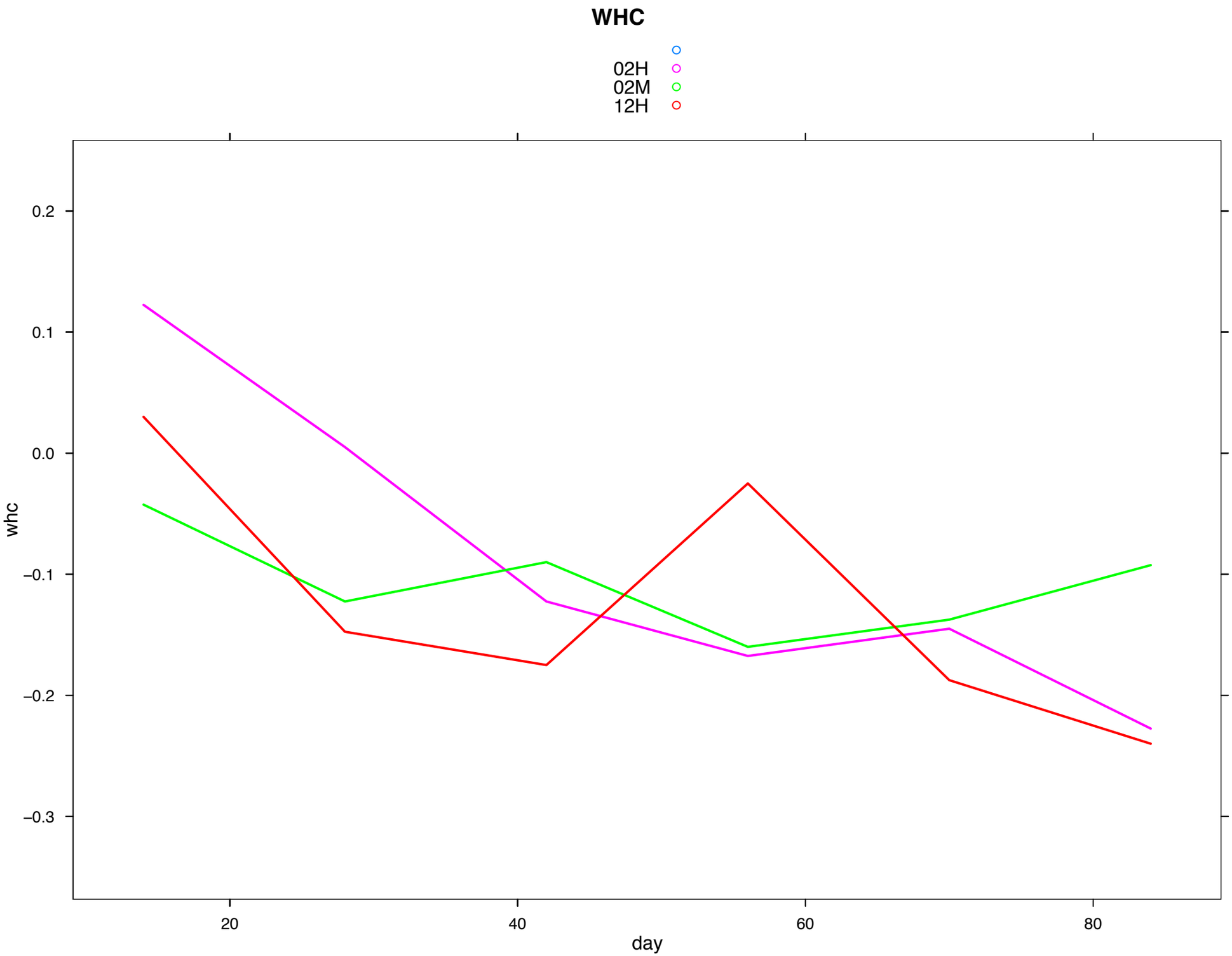
June 2011

CC BY NC

1. **Past**: why ggplot2?
2. **Present**: what's happening in the ggplot2 community right now?
3. **Future**: interactive graphics in R

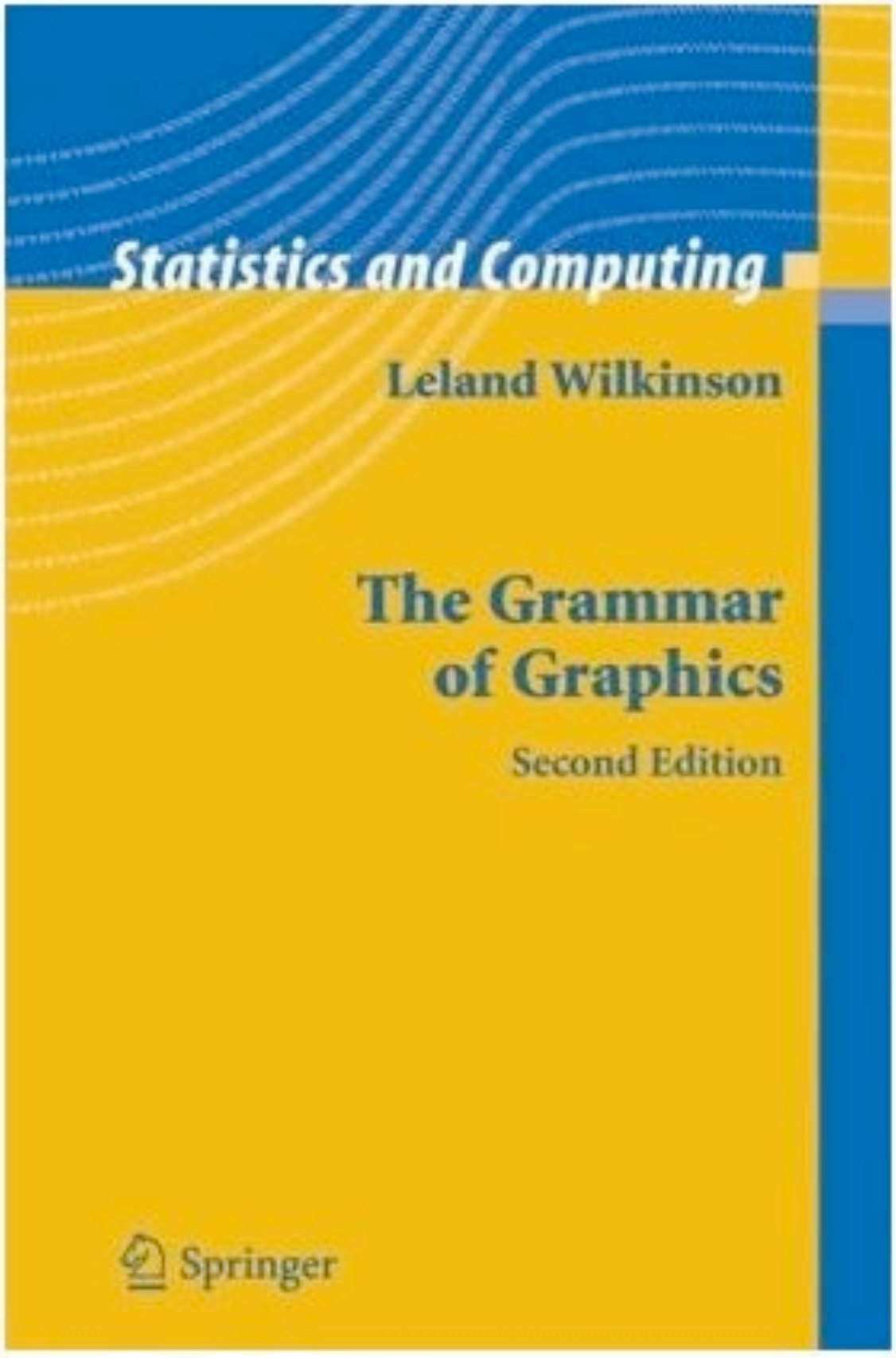
Why ggplot2?

2004



Niggles

- Why didn't `panel.line` automatically order the points? Why couldn't you describe histograms by specifying binwidths?
- Why was `xyplot(panel = panel.bwplot)` the same as `bwplot`?
- Why was it so hard to combine different datasets on the same plot?


Statistics and Computing
Leland Wilkinson
The Grammar of Graphics
Second Edition
Springer

“Nothing is as practical as a good theory”
—*Kurt Lewin*

“[A good model] will bring together in a coherent way things that previously appeared unrelated and which also will provide a basis for dealing systematically with new situations”
—*David Cox*

Milestones

(R 2.2.1)

28 Oct **2005** – ggplot development starts

6 Apr **2006** – first release of ggplot

3 Jul **2006** – first ggplot announcement

10 Jun **2007** – first release of ggplot2

7 Nov **2008** – start of ggplot2 mailing list

7 Aug **2009** – ggplot2 book published

Major versions

0.5: ggplot2, + instead of functional style

0.6: documentation, auto legends

0.7: themes

0.8: facet_wrap, free scales

0.9: namespace, roxygen, S3, diaspora

The "hard" stuff

- Implementing grammar is actually pretty easy
- The hard stuff (as always) is in the details:
  - What is a good default scale for colour? Size? shape?
  - Where should tick marks go?
  - How can the user control the style of the plot?

Milestones

April 2006: how do you put two ggplot figures into the same page?

June 2006: why are the breaks in such bad positions?

July 2006: how do you get rid of the gray background?

Present

Community

Mailing list: 1,500 members, 9,000 messages, 300-500 messages/month

Stackoverflow.com: >450 questions and answers

Code contributions: In last release of ggplot2, 50% of features/fixes contributed by users (mainly by Takahashi Kohske)

Current work

ggplot2 is too complicated - it is hard for me to improve, and hard for new developers to understand.

ggplot2 was my second R package - I have now written around 30. I now know a lot more!

Current work

Break up ggplot2 into simpler pieces: scales, density vis, spatial vis, ...

Aggressively rewrite to make simpler.

Better development practices: S3 instead of proto; roxygen; unit testing; namespaces.

Big data

Working with Revolution Analytics to get ggplot2 working with big data.

Results will be open source.

Focussed on working with RevoScaleR, but will also yield general speed ups

ggplot2 plans

ggplot2 basically complete: still plenty of rough edges, but from a research perspective it's largely done

Aim: Find funding for full-time programmer. Help others become ggplot2 developers

Future

	Orion I (1981)	My laptop (2011)	

	Orion I (1981)	My laptop (2011)	
Cost	$250,000	$2,000	1 / 125

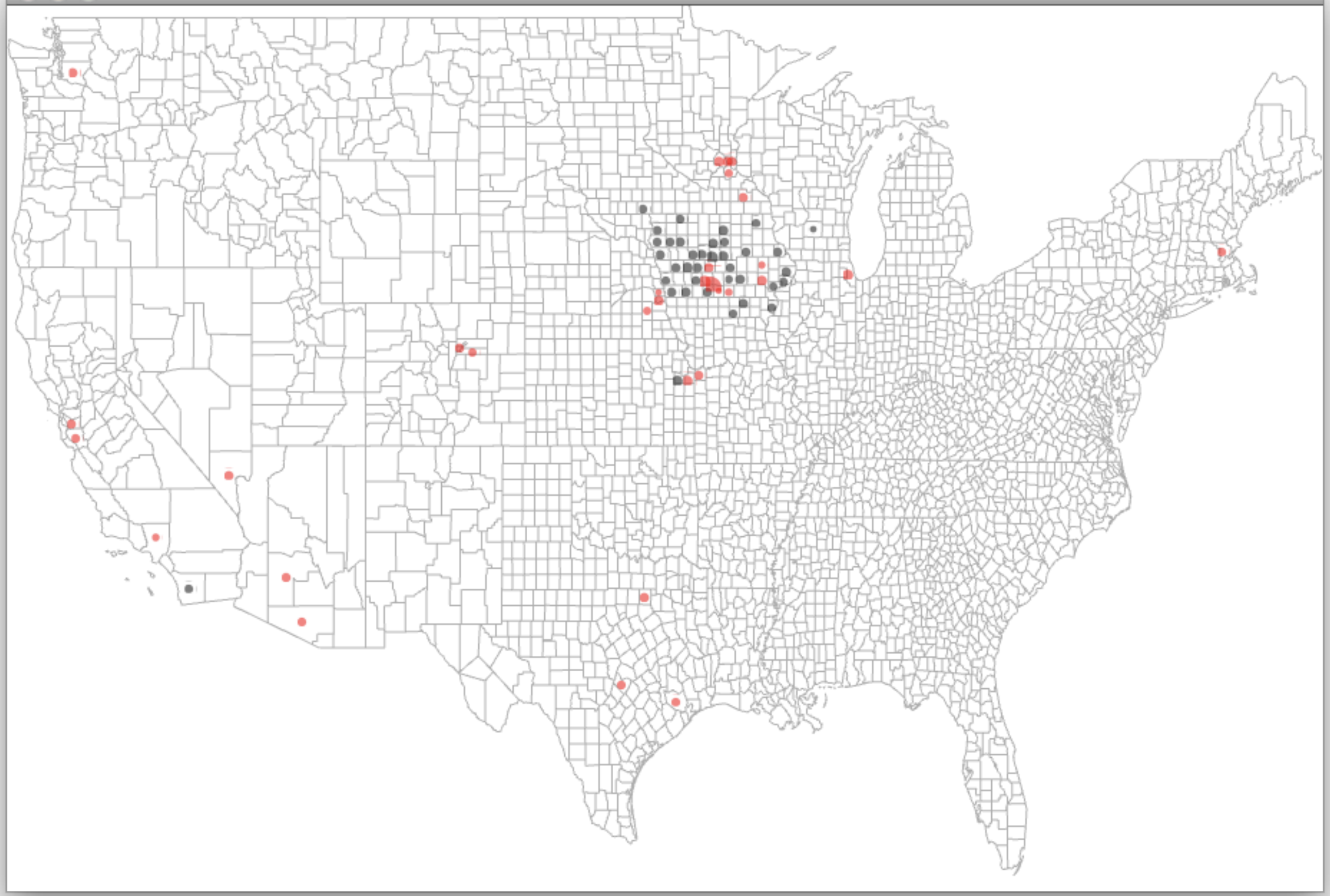
	Orion I (1981)	My laptop (2011)	
Cost	$250,000	$2,000	1 / 125
Speed	8 MHz	2.1 GHz	300 x

	Orion I (1981)	My laptop (2011)	
Cost	$250,000	$2,000	1 / 125
Speed	8 MHz	2.1 GHz	300 x
Memory	128 kb	4 Gb	30,000x

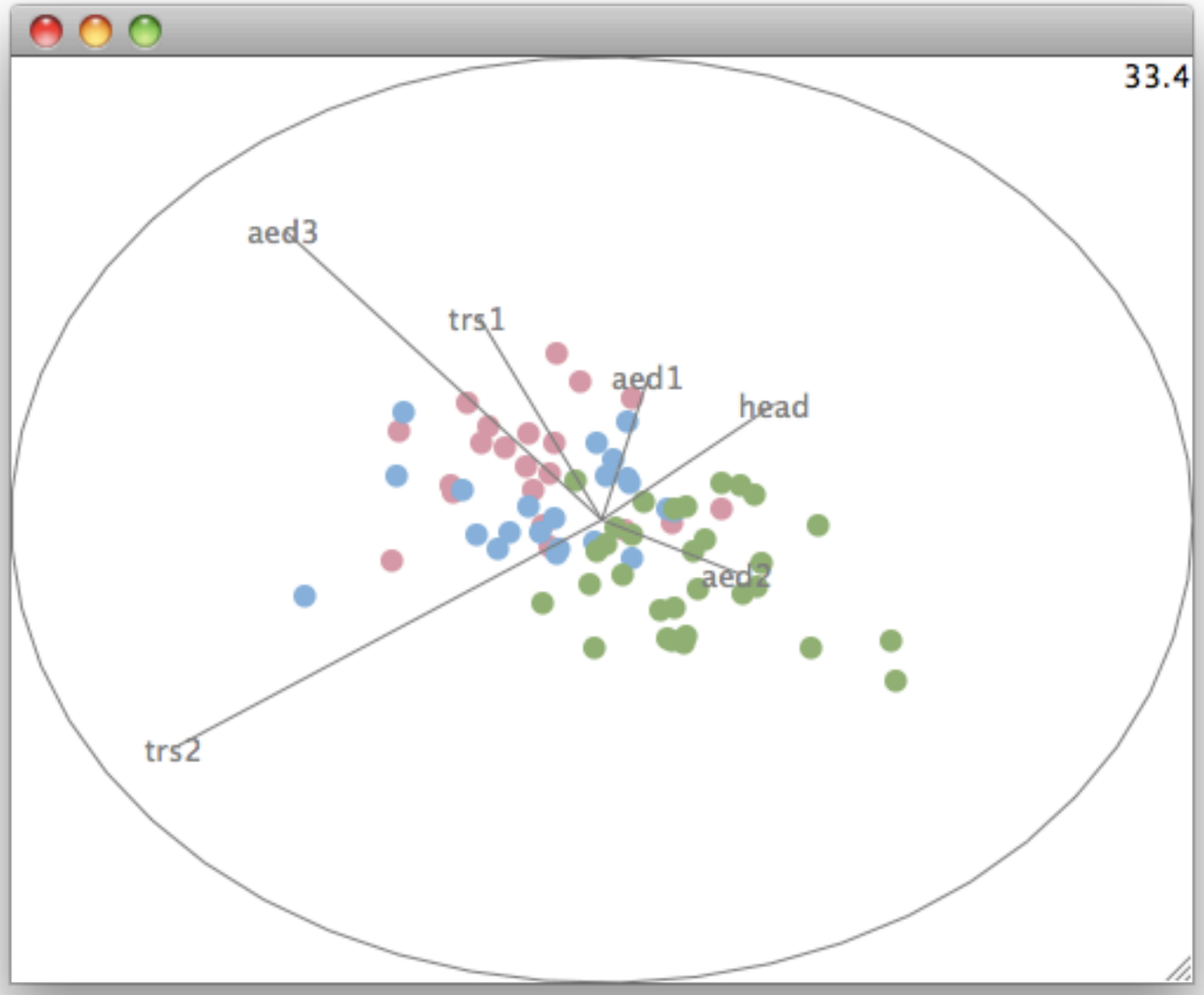
	Orion I (1981)	My laptop (2011)	
Cost	$250,000	$2,000	1 / 125
Speed	8 MHz	2.1 GHz	300 x
Memory	128 kb	4 Gb	30,000x
Screen	1280 x 1024	1440 x 900	1x

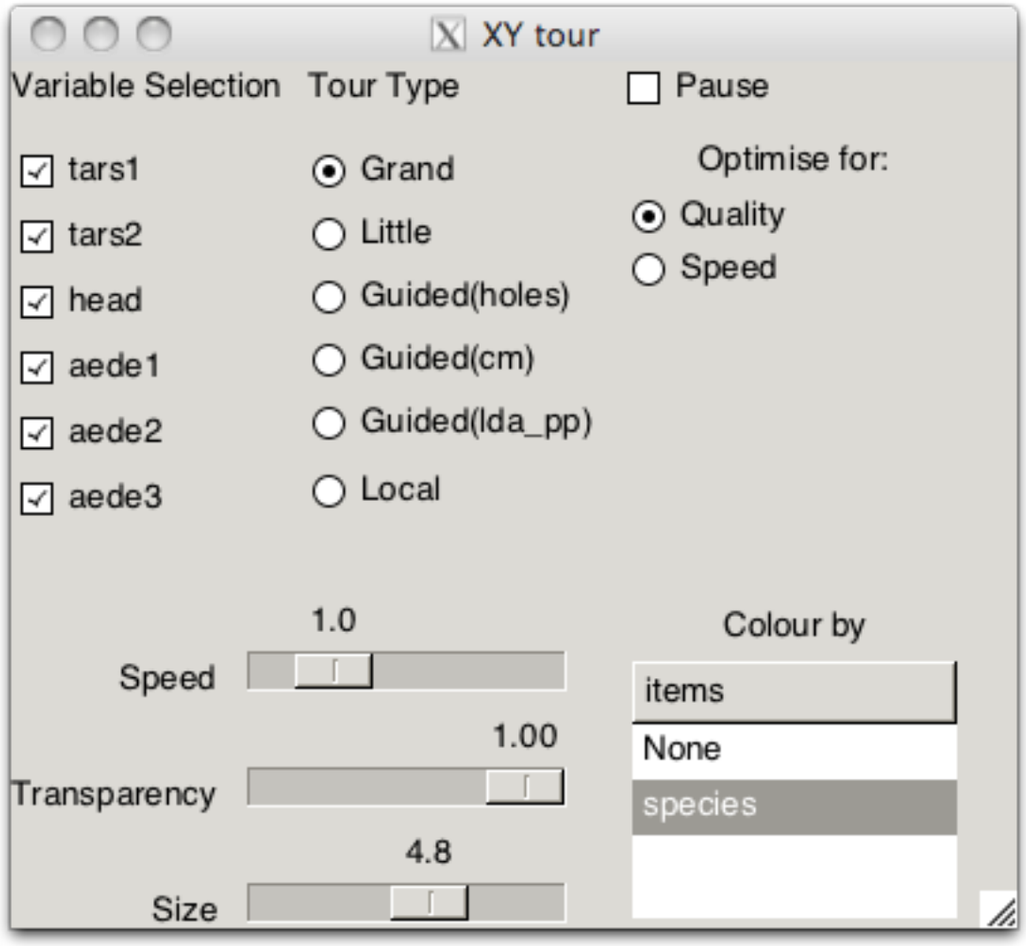
	Orion I (1981)	+ Monitor	
Cost	$250,000	$5,000	1 / 50
Speed	8 MHz	2.1 GHz	300 x
Memory	128 kb	4 Gb	30,000x
Screen	1280 x 1024	2560 x 1600	3 x

Interactivity!

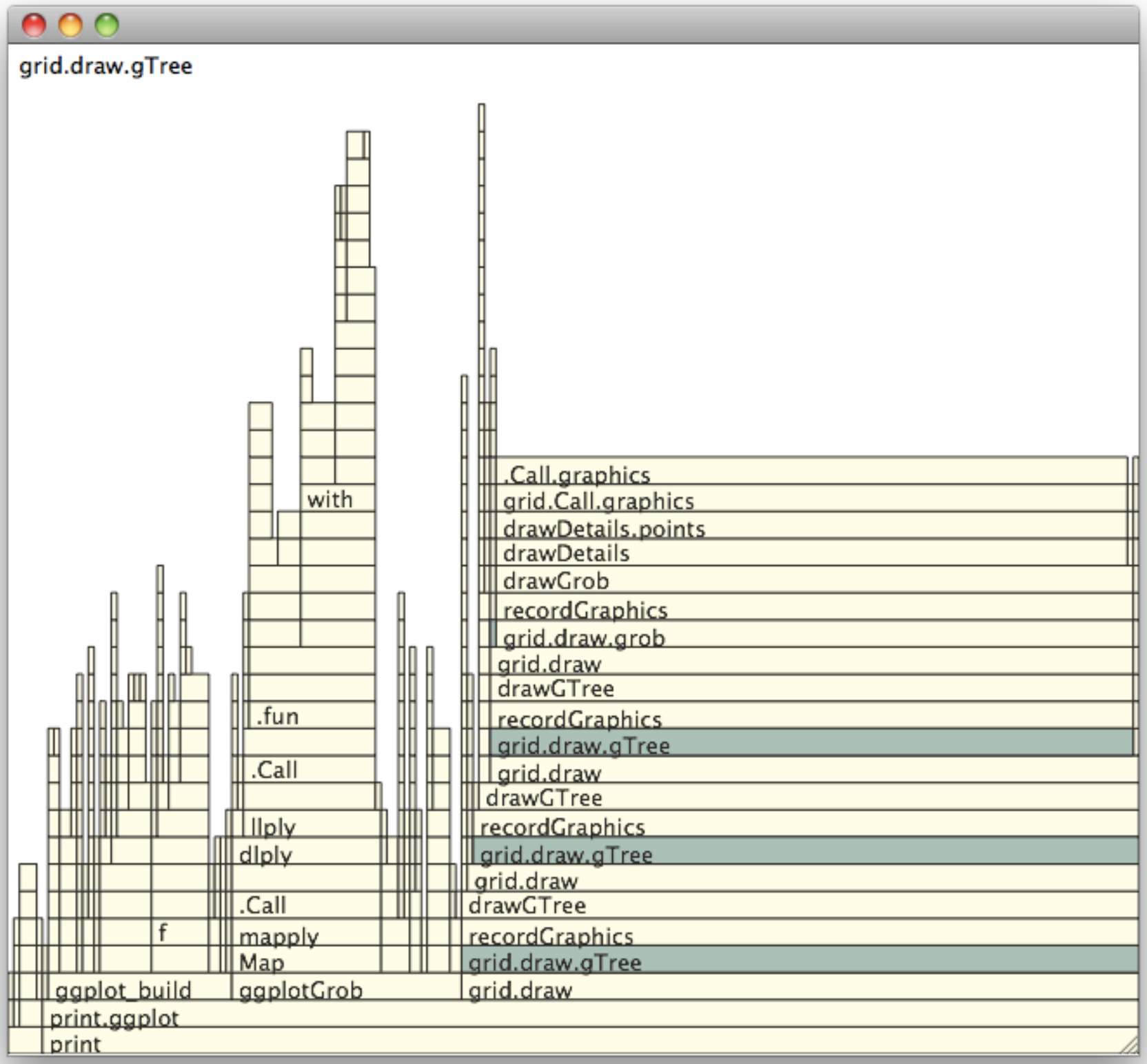


http://www.forbes.com/2010/06/04/migration-moving-wealthy-interactive-counties-map.html

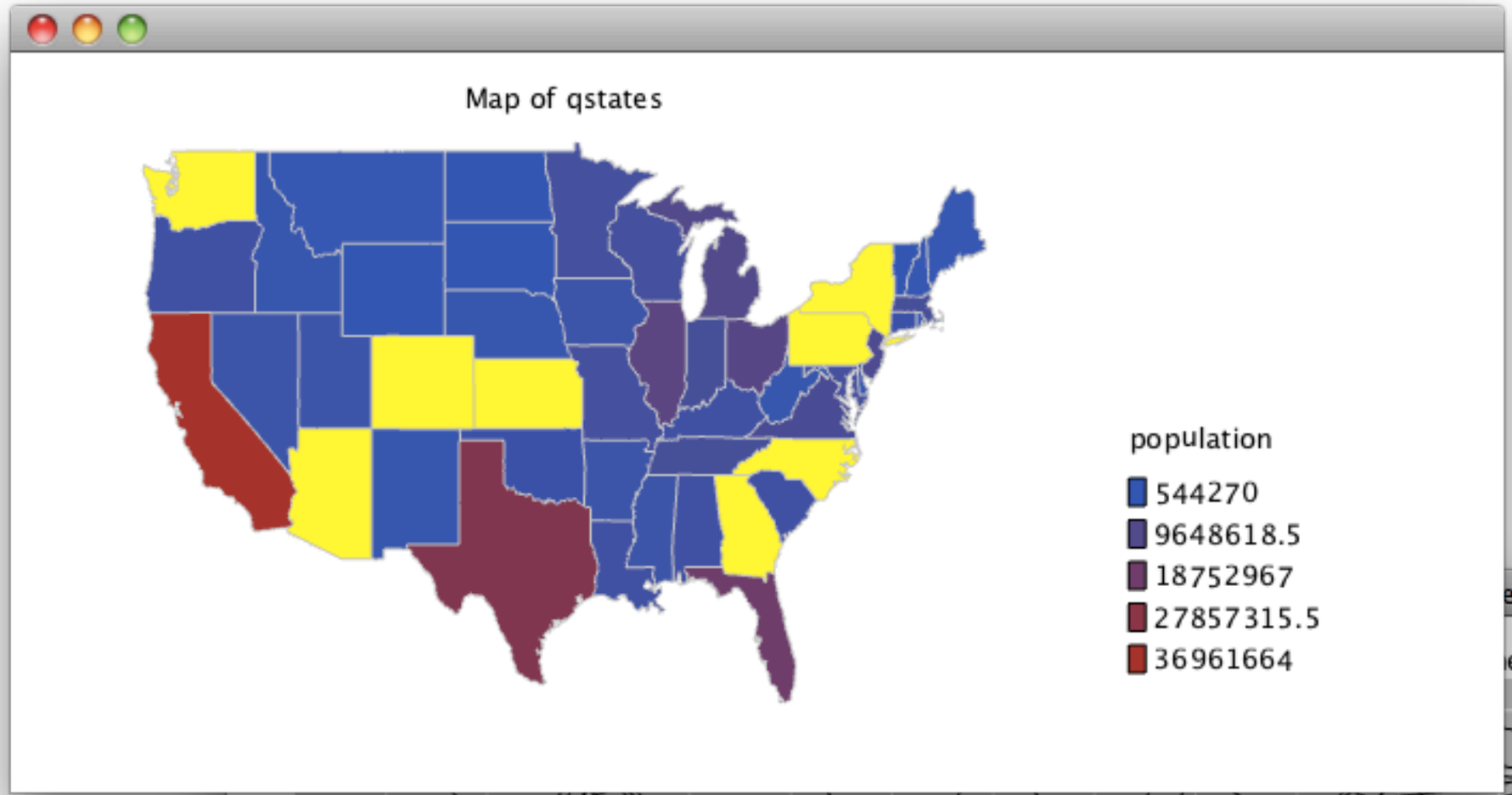


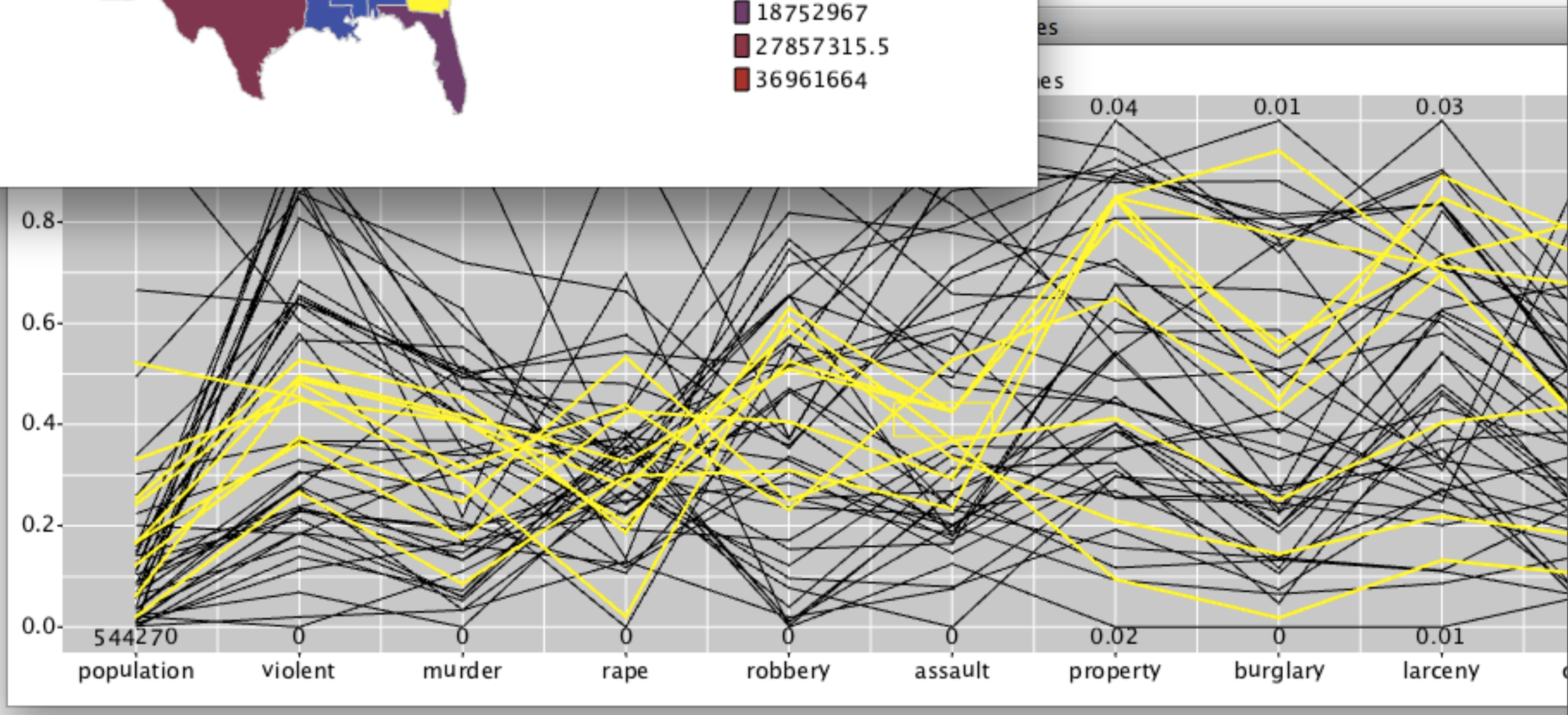


with Bei Huang



with Garrett Grolemund





Yihui Xie, Heike Hofmann, Di Cook, ISU stat graphics team

Components

All examples are pure R code!

qtpaint package provides high performance canvas

tourr package implements grand tour and related algorithms in pure R code

plumbr provides mutable data frames with events

scales package extracts scale related code out of ggplot2

qtbase + qtpaint

Written by Michael Lawrence and Deepayan Sarkar.

qtbase provides complete wrapper to Qt.

qtpaint provides high-performance graphics canvas built on top of Qt and OpenGL.

Together, makes it possible to develop interactive graphics in R.

Getting started

Currently painful - only OS X/linux, and requires a lot of compiling.

Aiming for `install.packages("qtpaint")`, and no external dependencies.

Hopefully by end of summer (on Bioconductor).

Challenges

Have to think carefully about layering together objects that change together.

Need to (re-)learn mutable objects, message passing style. Modularity/ reusability.

Most documentation currently for C++.

OpenGL quirks.

Low-level

High-level

Qtpaint
Processing
Protovis
Quail
Orca
GGobi
Spotfire
Tableau
CrystalVision
Mondrian
MANET

Flexible
Code
Graphics
Research

Targeted
GUI
Data
Data analysis

Low-level

High-level

Qtpaint
Processing
vis
bi
fire
eau
CrystalVision
Mondrian
MANET

But don't want to be limited to just existing statistical approaches!

Flexible
Code
Graphics
Research

Targeted
GUI
Data
Data analysis

gigplot?

interactive

gigplot?

interactive

Check back in 5 years...

Please get in touch if...

- You want to talk about your data analysis process
- You want to learn more about any of the packages I've mentioned
- You have internships or jobs for my students

Saturday, July 23, 2011

This work is licensed under the Creative Commons Attribution-Noncommercial 3.0 United States License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/us/ or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.