R project in Google Summer of Code

Toby Dylan Hocking
toby.hocking@mail.mcgill.ca

12 April 2016

R is a free/open-source software environment for statistics and graphics

- https://www.r-project.org/
- A command interpreter for the R programming language.
- Packages for machine learning, data analysis, visualization.
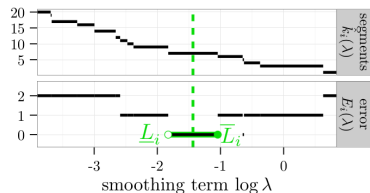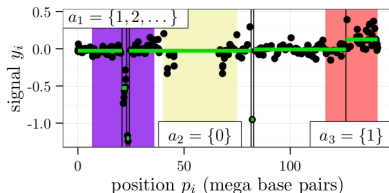- Completely free and open for anyone to use and modify (GNU General Public License version 2).

Figure source: Hocking et al, "Learning sparse penalties for change-point detection using max-margin interval regression," International Conference on Machine Learning 2013.

History of R project and Google Summer of Code

1992 First **R** coded by Robert Gentleman and Ross Ihaka at University of Auckland, New Zealand.

1997 First Comprehensive R Archive Network (CRAN) web site hosting **R packages**.

1998 Google incorporated in Stanford, California.

2005 First **Google Summer of Code** (GSOC).

2008 First time R participates in GSOC (GSOC-R).

2012 My first year as co-administrator of GSOC-R.

Sources: Wikipedia, GSOC web site.

Google Summer of Code (GSOC)

Student gets $5500 for writing open source code for 3 months.

Feb **Admins** for open source organizations e.g. R, Debian, Wikimedia, Mozilla, apply to Google.

Mar **Mentors** suggest projects for each org.
Students submit project proposals to Google.
Google gives funding for n students to an org.

April The top n students get $500 and begin coding.

July Midterm evaluation, pass = $2250.

Aug Final evaluation, pass = $2750.

November Orgs get $500/student mentored.

I have participated as an **admin** and **mentor** for the R project.

Student projects in R Google Summer of Code

year	total students	students I mentored
2008	4	
2009	4	
2010	5	
2011	15	
2012	16	
2013	18	1
2014	17	1
2015	23	3

What makes a good GSOC project?
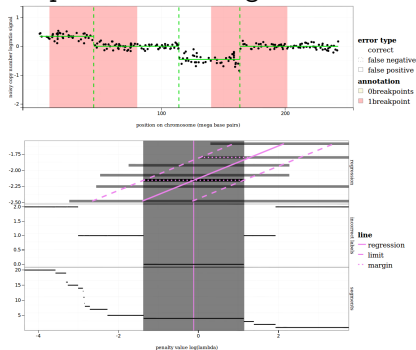
Coding projects should:

- ▶ Result in free/open-source software.
- ▶ Be 3 months of full time work for a student.
- ▶ Include writing documentation and tests.
- ▶ Not include original research.

Examples:

- ▶ Write a new R package.
- ▶ Improve an existing R package.

Example: R package animint, 2013–present

- Useful for interactive data visualization, for example
  `http://bl.ocks.org/tdhock/raw/10f27e4ace80bffa10a0/`

- Students I mentored:

Student	Country	Year	Features coded
Susan VanderPlas	USA	2013	Legends, geoms.
Carson Sievert	USA	2014	Tests, facets, etc.
Tony Tsai	China	2015	Efficient compilation.