

# Leveraging HUBzero to Enable STEM Education for High School Students

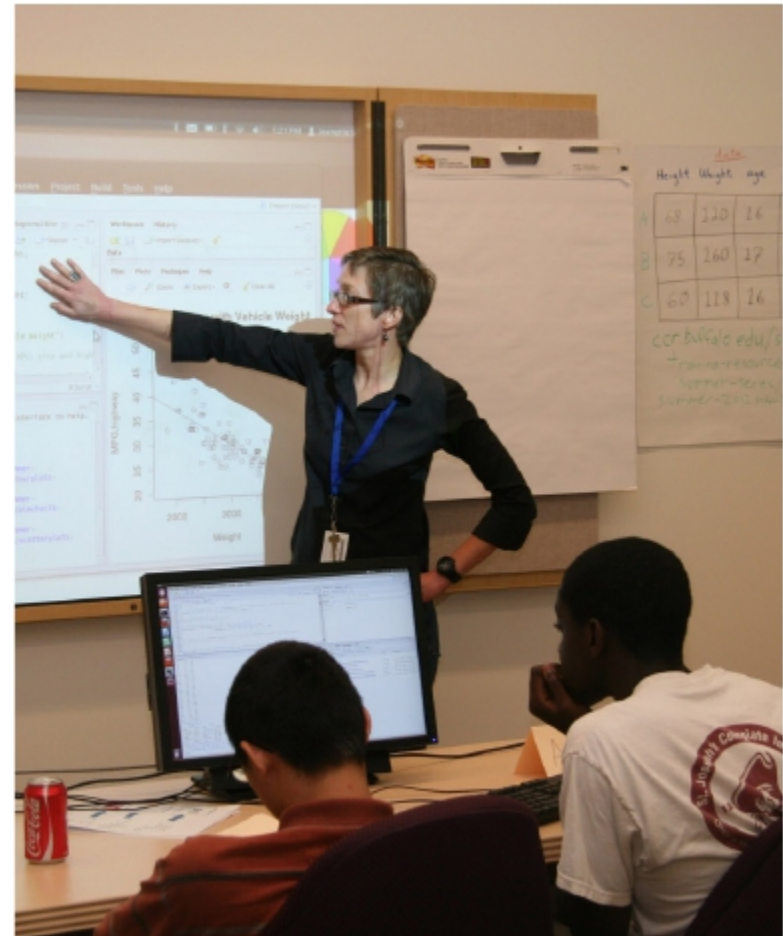
Jeanette Sperhac, Steven M. Gallo, E. Bruce Pitman,  
Thomas R. Furlani, Ryan Mraz, and Sam Steffan

HUBbub, Indianapolis, IN  
5 September 2013



# Outline

1. Introducing the Workshop
2. Choices for 2013
3. Our HUBzero instance
4. Deploying RStudio as a HUB Tool
5. Teaching on the HUB





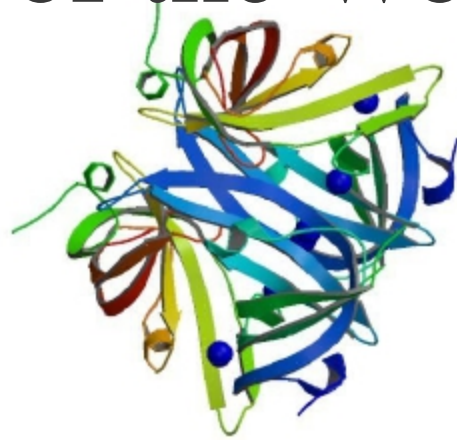
# Eric Pitman Annual Summer Workshop in Computational Science

Center for Computational Research (CCR)

University at Buffalo (UB)

since 1999

# Goals of the Workshop



Protein Structure,  
PDB ID: 3dm3

For two weeks, immerse students in computational science:

- Introduce programming, practice skills
- Tour working laboratories
- Feature specialized guest lectures
- Use computation to investigate a scientific problem

# Participants



Maximum of 12 student participants:

- from local high schools
- entering grades 10 through 12
- with appropriate math and science exposure
- no programming experience is assumed

# Staff



Staff hail from CCR and affiliated institutions:

- Roswell Park Cancer Institute (RPCI)
- Hauptman-Woodward Institute (HWI)
- University at Buffalo (UB)

# Coursework



Problem solving using programming:

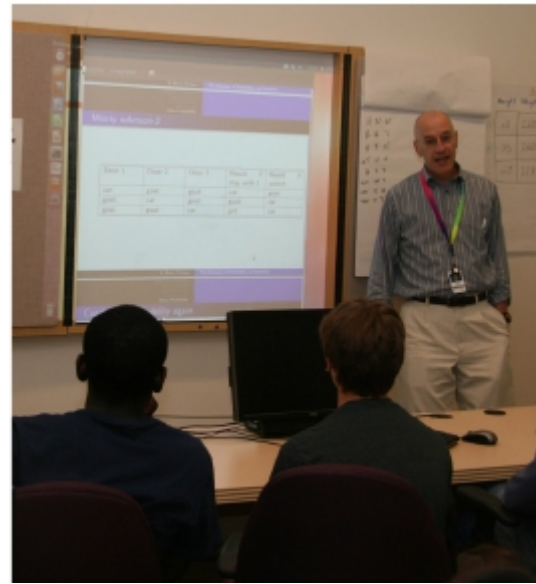
- Variables and data structures
- Writing functions
- Visualizing data

Five half-day units:

- Programming lectures and demonstrations
- Hands-on coding exercises

Students collaborate on shared workstations

# Guest Lectures



Guest lecturers introduce scientific background:

- Protein structure determination
- Gene sequencing and expression
- Basic statistics
- High-performance computing



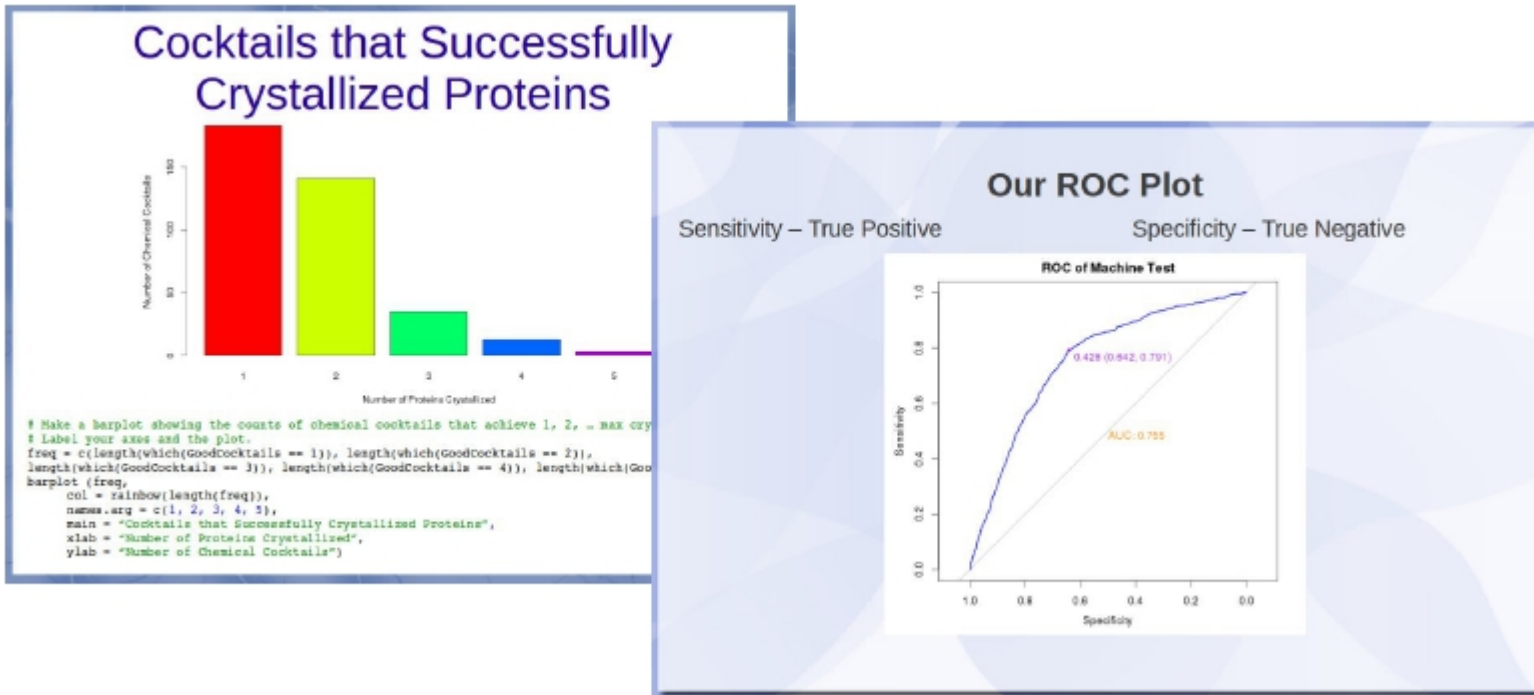
# Lab Tours



Tours provide scientific context:

- CCR machine room
- Protein crystallization and structure labs
- Protein sequencing labs
- Patient radiation treatment facilities

# Final Project



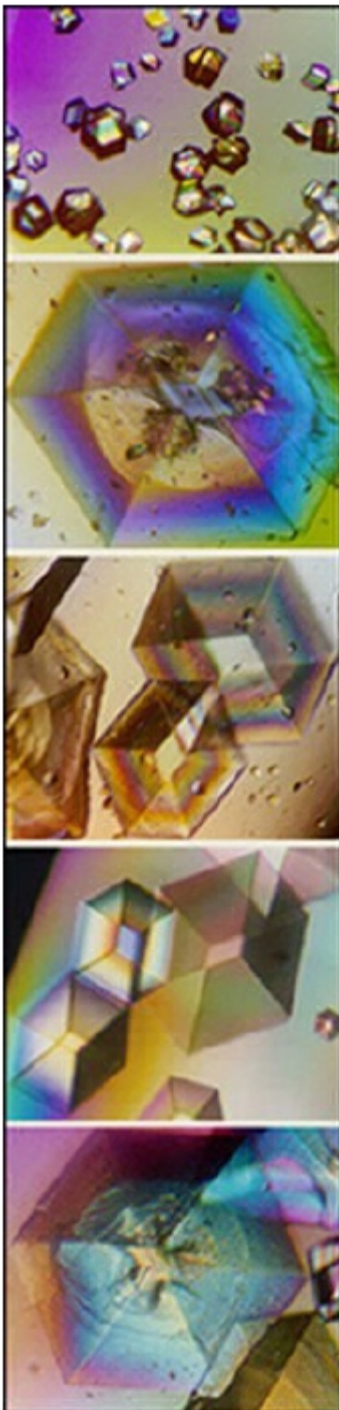
Students exercise their programming skills, working in teams on real scientific data.

The project week culminates in student presentations of their results.

# How do we equip students to investigate a scientific problem?



- Focus on computational *science*
  - not the mechanics of submitting jobs
  - not the Unix command line
  - not the reason the code isn't compiling
- Emphasize getting one line of code right at a time



# 2013: One Solution

## Curriculum

- Data exploration and visualization
- Basic statistics

## Project

- Automated classifiers
- Protein crystallization data

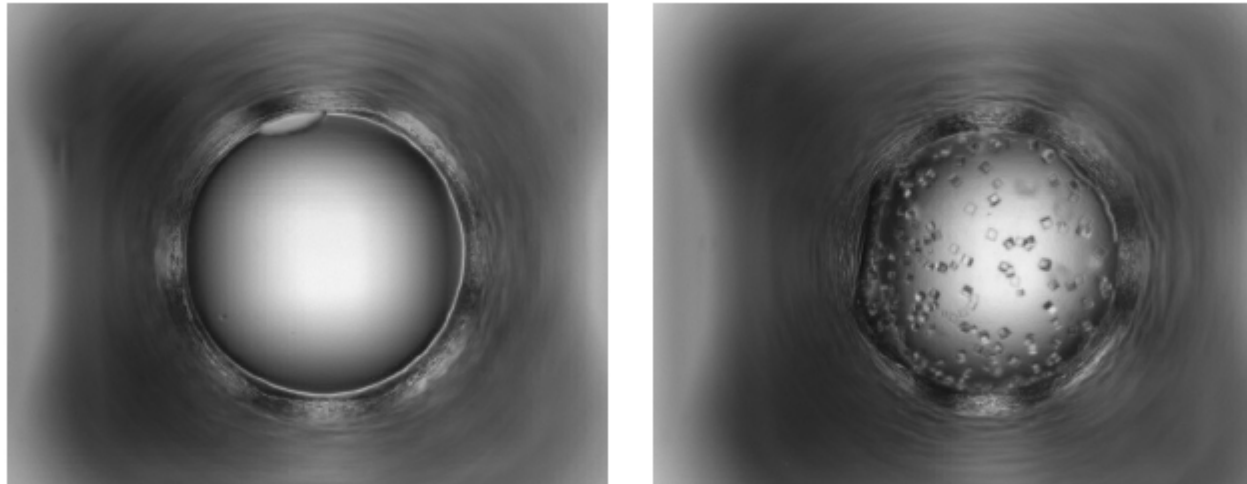
# 2013 Workshop environment



- Each student creates a login on our HUBzero instance
- All coursework is performed on the HUB
- GitHub repository: source code and datasets
- R and RStudio Projects: development
- File transfer: WebDAV

# Workshop Project

40,000 protein crystallization trials:



- Automated classifier assignments of crystal state
- Human expert assignments of actual crystal state
- Protein and chemical cocktail information
- Timeseries data

# Workshop Project

Protein crystallization: all in the neighborhood

- lab work at HWI
- processing on CCR's cluster



# Why we chose HUBzero



- Unified platform for coursework
- Easy on our admins:
  - Obviates software installs on individual student workstations
- Ubiquitous access—anytime, anywhere
- Students retain access after course conclusion





# Why we chose R

R language features:

- Command-line interpreter
- Procedural and object-oriented
- Easily extensible with packages
- Excellent support for *statistics* and *graphing*



# Why we chose RStudio

RStudio Interactive Development Environment (IDE) features:

- Integrates with Git or SVN repositories for source code and dataset distribution
- Supports Projects for ease of development
- Open-source and cross-platform

# RStudio Interface

Editor

The screenshot displays the RStudio interface with four main panes:

- Editor:** Contains R code for loading data and creating a plot. The code is as follows:

```
1 library(ggplot2)
2 source("plots/formatPlot.R")
3
4 view(diamonds)
5 summary(diamonds)
6
7 summary(diamonds$price)
8 aveSize <- round(mean(diamonds$carat), 4)
9 clarity <- levels(diamonds$clarity)
10
11 p <- qplot(carat, price,
12            data=diamonds, color=clarity,
13            xlab="Carat", ylab="Price",
14            main="Diamond Pricing")
15
```
- Console:** Shows the output of the code execution, including summary statistics for the 'diamonds' dataset and the 'price' variable. The output is as follows:

```
      x      y      z
Min.  :0.000 Min.  :0.000 Min.  :0.000
1st Qu.:4.710 1st Qu.:4.720 1st Qu.:2.910
Median :5.700 Median :5.710 Median :3.530
Mean   :5.731 Mean   :5.735 Mean   :3.539
3rd Qu.:6.540 3rd Qu.:6.540 3rd Qu.:4.040
Max.   :10.740 Max.   :58.900 Max.   :31.800
> summary(diamonds$price)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  326   950   2401   3933   5324  18820
> aveSize <- round(mean(diamonds$carat), 4)
> clarity <- levels(diamonds$clarity)
> p <- qplot(carat, price,
+           data=diamonds, color=clarity,
+           xlab="Carat", ylab="Price",
+           main="Diamond Pricing")
>
> format.plot(p, size=24)
>
```
- Workspace and History:** Shows the loaded data object 'diamonds' (53940 obs. of 10 variables) and the function 'format.plot'.
- Plots:** Displays a scatter plot titled 'Diamond Pricing' showing Price (Y-axis, 0 to 15000) versus Carat (X-axis, 0.0 to 3.5). The plot is faceted by Clarity, with a legend on the right showing categories: I1, SI2, SI1, VS2, VS1, VVS2, VVS1, and IF.

Workspace  
and  
History

Console

Files,  
Plots,  
Packages,  
Help

# HUBzero instance hardware



Purchased in January 2012:

- Dell R410 server
- 16GB RAM
  - 4x4GB - 1333MHz dual ranked DIMMs
- Intel Xeon E6520 2.4GHz processor
  - 12MB cache
  - 1066MHz front side bus
- 2x 500GB 7.2K RPM SATA hard drives

# HUBzero instance specifications



- Server runs HUBzero 2012 open-source version
- Debian 6
- Hosted in CCR machine room
- Workspace access restricted through Group
- Tool access restricted to logged-in users

# Student workstation specifications



Commodity workstation, Ubuntu 12.04 LTS:

- Firefox
- LibreOffice (presentation writing)
- Packages to support HUBzero tool sessions and WebDAV:
  - openjdk-6-jdk
  - icedtea-6-plugin
  - davfs2

# Deploying RStudio as a HUBzero Tool

1. Create RStudio tool using the HUBzero tool development workflow.

## Tools: Create New Tool

**About your tool:**

**Tool Name:** *required*

Short name, used for the directory containing this tool. Example: qdot

**Title:** *required*

Full name for this tool. Example: Quantum Dot Lab

**Version:**

Optional version number for this release of the tool. Example: 1.0 or 2.1.5b. Spaces not allowed.

**At a glance:** *required*



# Deploying RStudio as a HUBzero Tool

## 2. Install R inside HUBzero instance's OpenVZ container:

- chroot to OpenVZ container
- edit `sources.list` to include R binaries:\*

```
deb http://cran.mtu.edu/bin/linux/debian squeeze-cran/
```

```
deb-src http://cran.mtu.edu/bin/linux/debian squeeze-cran/
```

- install `r-base` using `apt-get`

\* Refer to “R on Debian” in References slide





# Deploying RStudio as a HUBzero Tool

## 3. Install additional R packages inside OpenVZ container:

- invoke R
- install R packages:
  - > `install.packages("ggplot2")`
- exit container when done



# Deploying RStudio as a HUBzero Tool

## 4. Install RStudio on HUBzero instance:

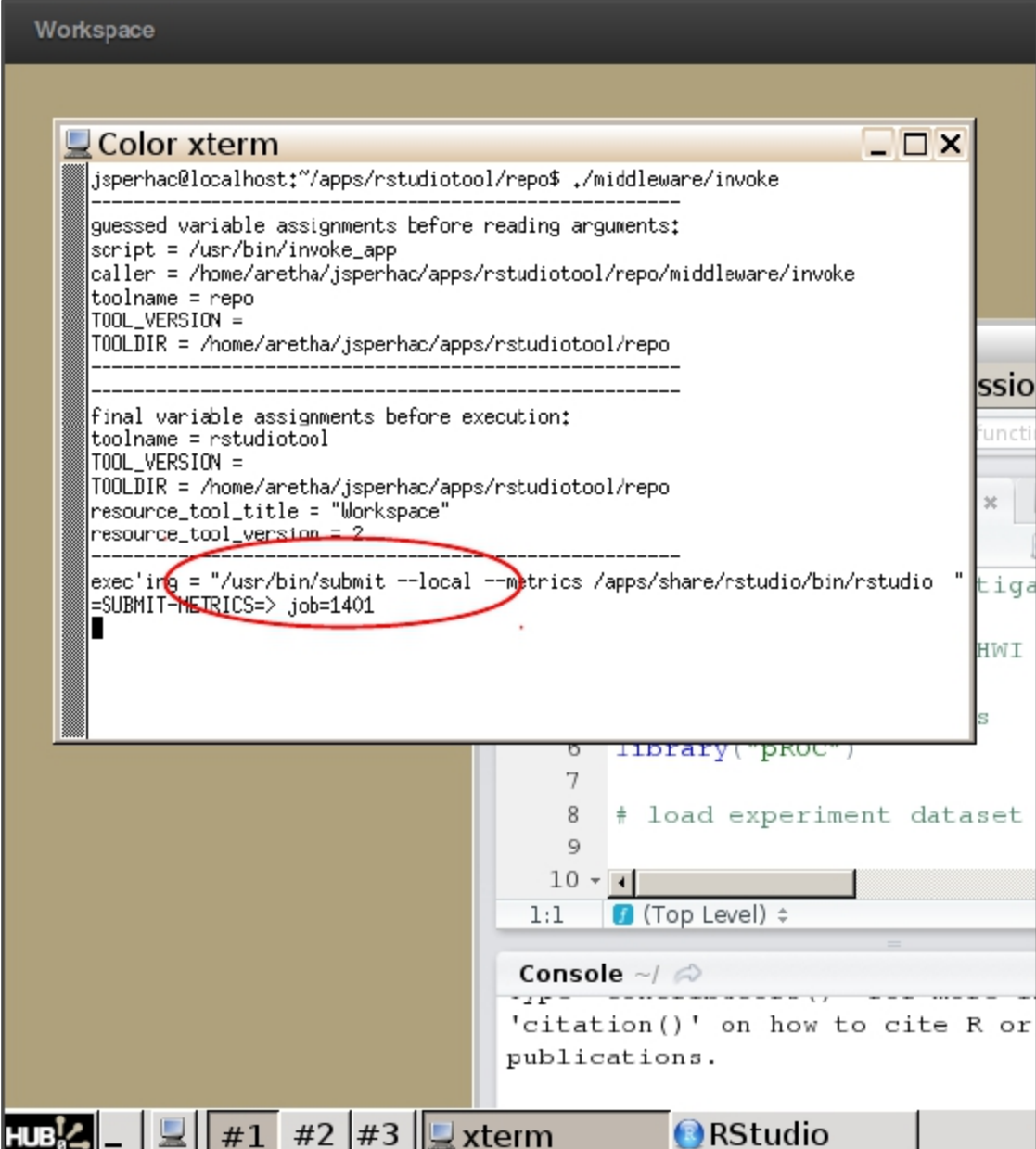
- Unpack RStudio Desktop 0.97.316 tarball in /apps/share/ directory
- Fix qt libraries\*

\* Refer to “Fixing qt libraries during RStudio installation” in References slide



# Deploying RStudio as a HUBzero Tool

6. Verify local call to *submit* for tool invocation



The screenshot shows an RStudio workspace with a terminal window titled "Color xterm". The terminal output displays the execution of a script, showing variable assignments and the final command executed. The command `exec'ing = "/usr/bin/submit --local --metrics /apps/share/rstudio/bin/rstudio"` is circled in red, and its output is `=SUBMIT-METRICS=> job=1401`. Below the terminal, the RStudio console shows the start of an R script with lines for loading a dataset and a citation note.

```
Workspace
```

```
jsperhac@localhost:~/apps/rstudiotool/repo$ ./middleware/invoke
-----
guessed variable assignments before reading arguments:
script = /usr/bin/invoke_app
caller = /home/aretha/jsperhac/apps/rstudiotool/repo/middleware/invoke
toolname = repo
TOOL_VERSION =
TOOLDIR = /home/aretha/jsperhac/apps/rstudiotool/repo
-----
final variable assignments before execution:
toolname = rstudiotool
TOOL_VERSION =
TOOLDIR = /home/aretha/jsperhac/apps/rstudiotool/repo
resource_tool_title = "Workspace"
resource_tool_version = 2
-----
exec'ing = "/usr/bin/submit --local --metrics /apps/share/rstudio/bin/rstudio"
=SUBMIT-METRICS=> job=1401
█
```

```
6 library("PROC")
7
8 # load experiment dataset
9
10
1:1 (Top Level) ↕
```

```
Console ~/ ↻
-----
'citation()' on how to cite R or
publications.
```

HUB #1 #2 #3 xterm RStudio

# Deploying RStudio as a HUBzero Tool

7. Publish RStudio as a tool.

*Done!*

## Tools: Status for rstudiotool - Published

STATUS: Registered Created Uploaded Installed Approved **Published**

This tool is one of 14 tools published on hpc2.

### Tool Information [edit](#)

Title	R Studio Tool (rstudiotool - id #25)
Version	This version 1.1 [ <a href="#">all versions</a> ]
At a glance	RStudio is an IDE for the R language. It provides a multipane window environment for statistical computing and graphics in R.
Description	<a href="#">Preview</a>   <a href="#">Edit description page</a>
VNC geometry	1024x768
Tool execution	open to public
Source code	open source [ <a href="#">change license</a> ]
Project area	open to public

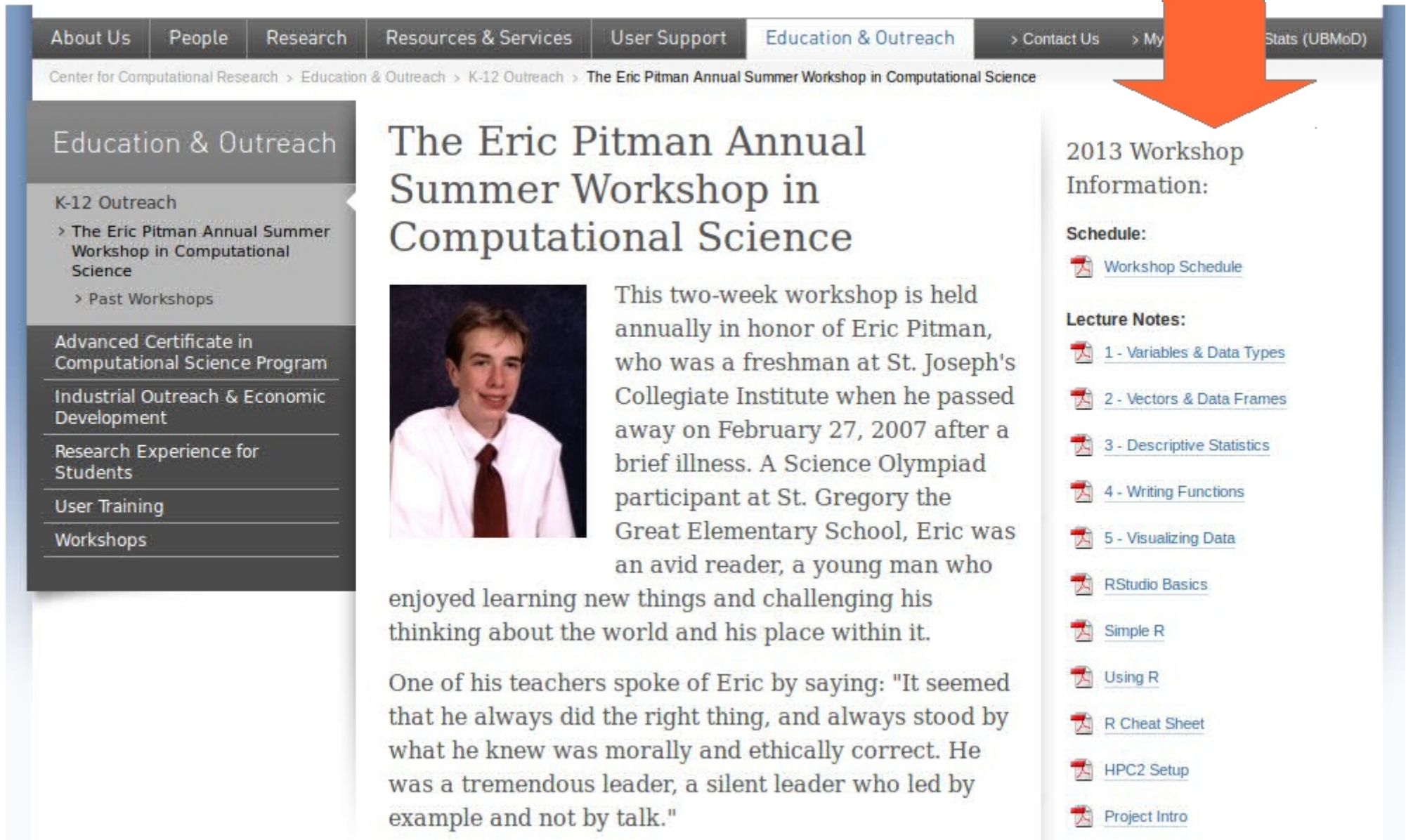
# Teaching on the HUB



Options for distributing handouts and lecture slides:

- Workshop webpage
- Course Resource on HUBzero

# Workshop webpage



The screenshot shows a website for the 'The Eric Pitman Annual Summer Workshop in Computational Science'. The navigation bar includes 'About Us', 'People', 'Research', 'Resources & Services', 'User Support', 'Education & Outreach', 'Contact Us', 'My', and 'Stats (UBMoD)'. The breadcrumb trail is 'Center for Computational Research > Education & Outreach > K-12 Outreach > The Eric Pitman Annual Summer Workshop in Computational Science'. The left sidebar lists 'Education & Outreach' with sub-items: 'K-12 Outreach' (including the current workshop and 'Past Workshops'), 'Advanced Certificate in Computational Science Program', 'Industrial Outreach & Economic Development', 'Research Experience for Students', 'User Training', and 'Workshops'. The main content area features the title 'The Eric Pitman Annual Summer Workshop in Computational Science' and a portrait of a young man. The text describes the workshop as an annual event in honor of Eric Pitman, who passed away in 2007. It mentions his participation in a Science Olympiad and his status as an avid reader. A quote from a teacher describes him as a moral and ethical leader. The right sidebar, titled '2013 Workshop Information', contains a 'Schedule' section with a link to 'Workshop Schedule' and a 'Lecture Notes' section with links to: '1 - Variables & Data Types', '2 - Vectors & Data Frames', '3 - Descriptive Statistics', '4 - Writing Functions', '5 - Visualizing Data', 'RStudio Basics', 'Simple R', 'Using R', 'R Cheat Sheet', 'HPC2 Setup', and 'Project Intro'. A large orange arrow points from the top right towards the '2013 Workshop Information' section.

About Us People Research Resources & Services User Support Education & Outreach > Contact Us > My Stats (UBMoD)

Center for Computational Research > Education & Outreach > K-12 Outreach > The Eric Pitman Annual Summer Workshop in Computational Science

## Education & Outreach

K-12 Outreach

- > The Eric Pitman Annual Summer Workshop in Computational Science
- > Past Workshops

Advanced Certificate in Computational Science Program


Industrial Outreach & Economic Development

Research Experience for Students

User Training

Workshops

## The Eric Pitman Annual Summer Workshop in Computational Science




This two-week workshop is held annually in honor of Eric Pitman, who was a freshman at St. Joseph's Collegiate Institute when he passed away on February 27, 2007 after a brief illness. A Science Olympiad participant at St. Gregory the Great Elementary School, Eric was an avid reader, a young man who enjoyed learning new things and challenging his thinking about the world and his place within it.












One of his teachers spoke of Eric by saying: "It seemed that he always did the right thing, and always stood by what he knew was morally and ethically correct. He was a tremendous leader, a silent leader who led by example and not by talk."

### 2013 Workshop Information:

**Schedule:**

-  [Workshop Schedule](#)

**Lecture Notes:**

-  [1 - Variables & Data Types](#)
-  [2 - Vectors & Data Frames](#)
-  [3 - Descriptive Statistics](#)
-  [4 - Writing Functions](#)
-  [5 - Visualizing Data](#)
-  [RStudio Basics](#)
-  [Simple R](#)
-  [Using R](#)
-  [R Cheat Sheet](#)
-  [HPC2 Setup](#)
-  [Project Intro](#)

# HUBzero Course Resource

## [R Workshop](#)

This two-week summer workshop is geared to high school students with a strong interest in scientific computing. Materials introduce the R language and provide hands-on exercises to practice programming skills. [Learn more >](#)

[View Course Lectures](#)

Lecture	Video	Lecture Notes	Supplemental Material	Suggested Exercises
<a href="#">Lecture 1</a>		<a href="#">Notes</a>		<a href="#">Exercise 1: Variables and Atomic Datatypes</a>
<a href="#">Lecture 2</a>		<a href="#">Notes</a>		<a href="#">Exercise 2: Vectors and Data Frames</a>
Vectors and Data Frames				
<a href="#">Lecture 3</a>		<a href="#">Notes</a>		<a href="#">Exercise 3: Descriptive Statistics</a>
Descriptive Statistics				
<a href="#">Lecture 4</a>		<a href="#">Notes</a>		<a href="#">Exercise 4: Writing Functions</a>
Writing Functions				
<a href="#">Lecture 5</a>		<a href="#">Notes</a>	<a href="#">Plotting Examples</a>	<a href="#">Exercise 5: Visualizing Data</a>
Visualizing Data				
<a href="#">Introductory R</a>			<a href="#">RProg.pdf</a>	
<a href="#">References</a>			<a href="#">introToR.pdf</a>	
Reference material: introductory R textbooks				



# Teaching on the HUB



Options for distributing source code and datasets:

- GitHub, Git, or SVN repositories
- Course Resource on HUBzero
- *data/* and *examples/* subdirectories in HUBzero tool

# Distributing source code and datasets



## RStudio integration with GitHub:

- Support for Projects
  - Each student has own workspace
- Read-only option:
  - No software install or authentication needed
- Caveat: not secured

# GitHub



jsperhac/workshop-dev · GitHub - Mozilla Firefox

hpc2 - resources: Tools: R St... x jsperhac/workshop-dev · Gi... x +

GitHub, Inc. (US) https://github.com/jsperhac/workshop-dev/ ☆ Google

Most Visited Getting Started

Disable Cookies CSS Forms Images Information Miscellaneous Outline Resize Tools

**GitHub** This repository Search or type a command ⚙ Explore Features Enterprise Blog [Sign up](#) [Sign in](#)

PUBLIC jsperhac / **workshop-dev** ☆ Star < 0 🍴 Fork < 0

**Code** Network Pull Requests 0 Issues 0 Graphs

Development space for summer workshop materials. — [Read more](#)

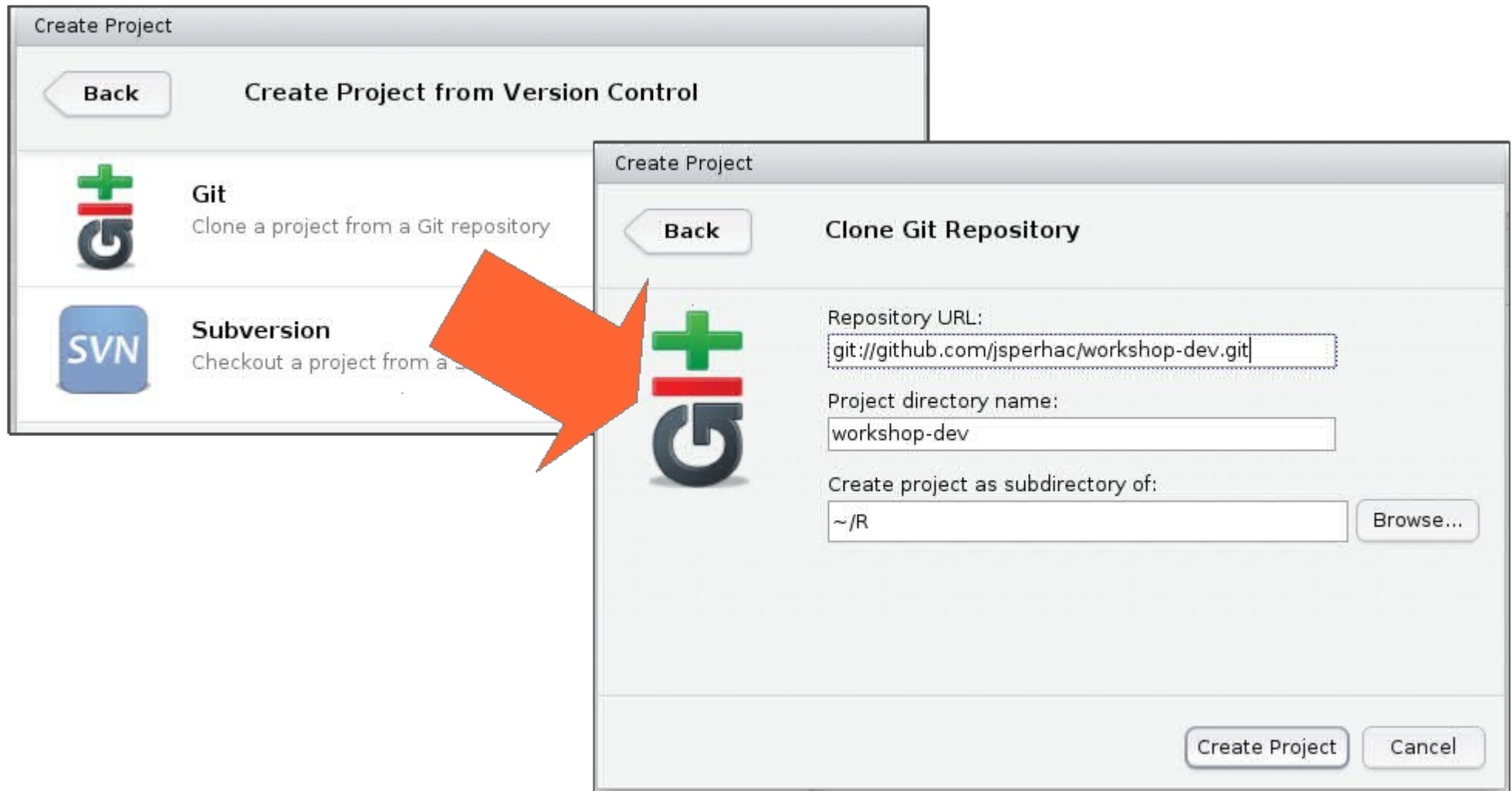
ZIP HTTP SSH Git Read-Only https://github.com/jsperhac/workshop-dev.git Read-Only access

branch: master Files Commits Branches 1 Tags

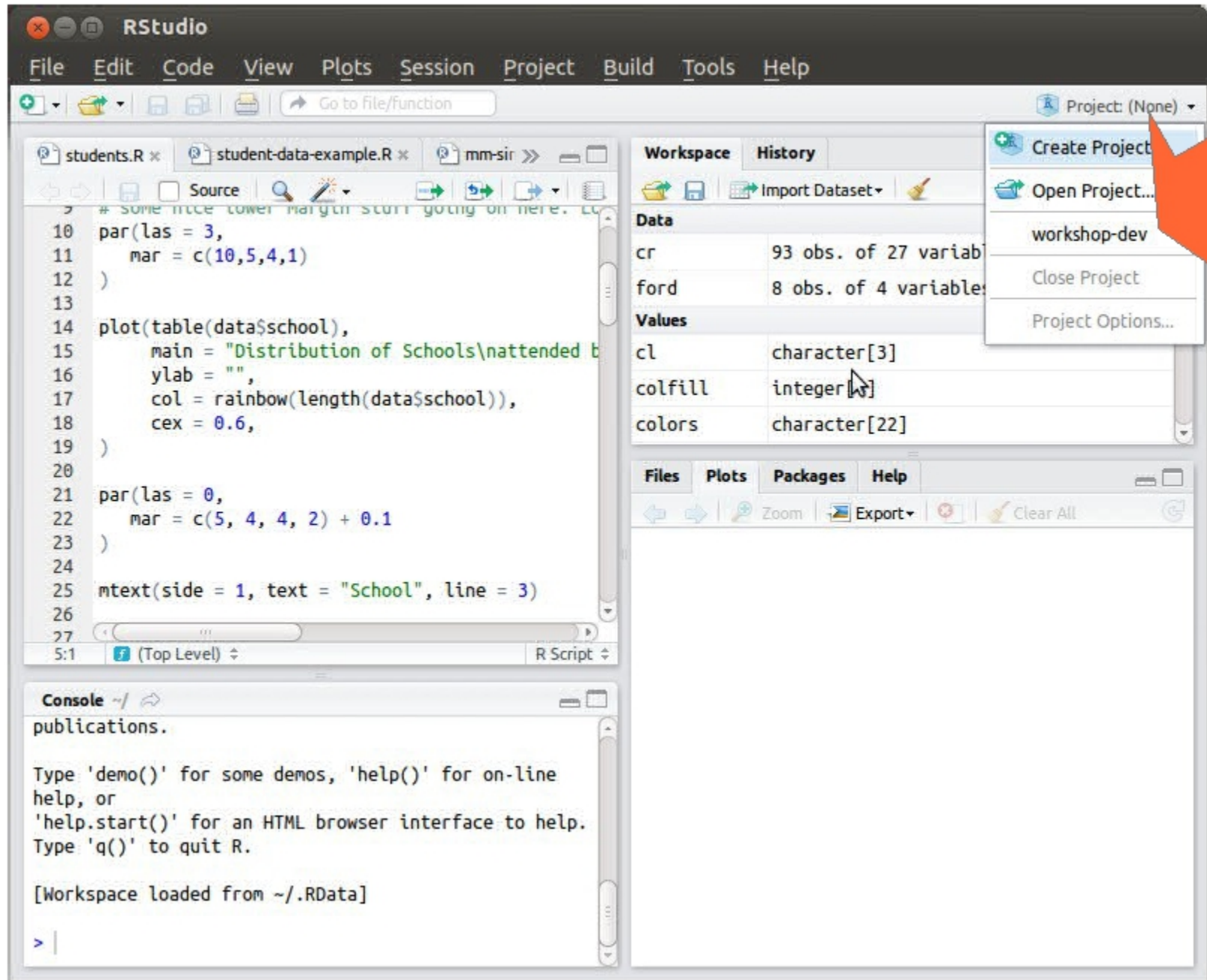
**workshop-dev** / 11 commits

added resource list		
jsperhac	authored 11 days ago	latest commit B232ebd6a1
examples	12 days ago	Added project, resources, exercises files [jsperhac]
exercises	12 days ago	removed answers from two exercise sources [jsperhac]
project	12 days ago	trimmed project data files [jsperhac]
resources	11 days ago	added resource list [jsperhac]

# RStudio data and sources: GitHub integration



# RStudio Project from GitHub



The screenshot shows the RStudio interface with the 'Project' menu open. The menu options are: 'Create Project', 'Open Project...', 'workshop-dev', 'Close Project', and 'Project Options...'. An orange arrow points to the 'Create Project' option. The main editor shows R code for plotting a table of school data. The console shows the R startup message and workspace loading information.

```
File Edit Code View Plots Session Project Build Tools Help
Go to file/function Project: (None)
Workspace History
Data
cr 93 obs. of 27 variables
ford 8 obs. of 4 variables
Values
cl character[3]
colfill integer[1]
colors character[22]
Files Plots Packages Help
Zoom Export Clear All
students.R student-data-example.R mm-sir
# Some nice lower margin stuff going on here. CC
10 par(las = 3,
11     mar = c(10,5,4,1)
12 )
13
14 plot(table(data$school),
15      main = "Distribution of Schools\nattended by",
16      ylab = "",
17      col = rainbow(length(data$school)),
18      cex = 0.6,
19 )
20
21 par(las = 0,
22     mar = c(5, 4, 4, 2) + 0.1
23 )
24
25 mtext(side = 1, text = "School", line = 3)
26
27
5:1 (Top Level) R Script
Console ~/
publications.
Type 'demo()' for some demos, 'help()' for on-line
help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Workspace loaded from ~/.RData]
> |
```

# Distributing source code and datasets



Other options:

- data/ and examples/ subdirectories in HUBzero tool. Caveats:
  - Students must specify paths to resources
  - Students lack own copies in home directory
- Course Resource on HUBzero
  - Caveat: file transfer necessary
  - But: may be secured



# Lessons Learned



Worked well:

- HUBzero instance for coursework
- RStudio deployed as a HUBzero Tool
- GitHub source and data management

Try next year for handout/slide distribution:

- Course Resource on HUBzero



# Additional thoughts



What if HUBzero offered:

- Tighter integration between HUBzero Resources and Workspace?
- Support for submitting student problem sets?
- Support for distributing grades?



# References and How-To

Home page for 2013 CCR Workshop:

- [ccr.buffalo.edu/outreach/k-12-outreach/summer-workshop.html](http://ccr.buffalo.edu/outreach/k-12-outreach/summer-workshop.html)

R on Debian:

- [cran.rstudio.com/bin/linux/debian](http://cran.rstudio.com/bin/linux/debian)

RStudio:

- [rstudio.org](http://rstudio.org)

Fixing qt libraries during RStudio installation:

- [hpc2.org/user-info/kb/development/deploying-qt-application](http://hpc2.org/user-info/kb/development/deploying-qt-application)

Creating Courses in HUBzero, see Section 7.4.2:

- <http://hubzero.org/documentation/1.1.0/managers/components.resources>

GitHub:

- [github.com](http://github.com)

# With Thanks

Sage advice, guidance, and guest lectures:

- E. Bruce Pitman, UB Dean of Arts and Sciences
- Thomas Furlani, CCR Director
- Steve Gallo, CCR

CCR interns:

- Ryan Mraz, Rochester Institute of Technology
- Sam Steffan, University of North Carolina
- Brian Narby, University of Pittsburgh

Photography: Adrian Levesque and Jake Brubaker, CCR



**UB** CENTER FOR COMPUTATIONAL RESEARCH  
University at Buffalo *The State University of New York*