Announcements

Publishing slidify

Hi class,

A quick note about publishing slidify presentations to github. It's appears to be easy with the publish command, which you execute in R in the directory where your slidify Rmd file is.

```
publish(user = "USER", repo = "REPO")
```

(Note the publish command does not appear to be working for some, see the note on how to directly publish to github below). More notes on it can be found here http://slidify.org/publish.html . There's also automatic ways to publish to dropbox and Rpubs (rStudio's web publishing stuff https://rpubs.com/)

You can also publish to github manually, though your github branch must have a .nojekyll fle and be on a branch named gh-pages. There's more on gh-pages here https://pages.github.com/ and below I briefly go through how to do it.

NEW STUFF: Here's how you can directly publish to gh-pages. Specifically, here's what worked for me (you can view the test here http://bcaffo.github.io/testDeck/index.html and view the repo here https://github.com/bcaffo/testDeck)

First I created the slide deck with author('testDeck') in slidify in R. I created the empty repo testDeck on github. I added the slide deck and all of the files at the command line (or you could use the gui git client). The git repo included an empty file named .nojekyll

In the directory I did

```
git init
```

Upcoming Deadlines

Quizzes

Quiz 2
Sun 19 Oct 2014 4:30 PM PDT

Recent Discussions

Uploading application documentation on shinyApps.io
Last post by Michael Stanhope (4 hours ago)

Quiz 2
Last post by Maqdalena Gajdadzieva COMMUNITY TA (8 hours ago)

Slidify, linux and vi
Last post by Rumen Zarev (9 hours ago)

Problem getting Slidify up and running
Last post by Maqdalena Gajdadzieva COMMUNITY TA (12 hours ago)

Issue with Sys.Date()
Last post by Anonymous (16 hours ago)

Browse all discussions »

```
git add *
git commit -a -m "Added all of the files"
git remote add origin *put in your github
origin here*
git push origin master
```

Then I made a branch with

```
git branch gh-pages
```

Then I pushed the branch to gh with

```
git push origin gh-pages
```

Then then files were at http://bcaffo.github.io/testDeck/index.html where testDeck was the name of the repo that I created. You can skip a lot of these steps by just cloning testDeck and modifying it. Remember, though, you have to have a branch called gh-pages and the .nojekyll file for it to show up. Then your page is at http://USERNAME.github.io/REPONAME/HTMLFILENAME

(Note gh-pages is a super easy way to create web sites. Also, having a separate master and gh-pages branches works well since you can work on the site on the master branch and merge the changes into the gh-pages branch only when you're ready for them to be visible.)

I checked the rendering and it worked on chrome, internet explorer and firefox for me.

Wed 15 Oct 2014 9:52 AM PDT

Developing Data Products: Quiz 1 Deadline Extended

It looks like there were extensive Coursera outages yesterday. As a result, we're extending the Quiz 1 deadline by a couple of days. Quiz submissions for full credit are now due by Tuesday at 11:30 PM UTC.

Mon 13 Oct 2014 5:48 AM PDT

Developing Data Products: Week 2

Welcome to Week 2 of *Developing Data Products*. I hope you got Quiz 1 done in time. If you didn't, don't despair; you can still submit it for partial credit until 11:30 PM UTC on Friday. There is a -10% penalty for each day late.

This week, we will cover presenting data analysis, slidify and Rstudio presenter. Make sure to watch the videos and start thinking about your project. Just click on the Course Project link in the left navigation bar to access all of the Course Project details and the submission interface. The project is due by the end of Week 3.

When you feel like you've got a good handle on the Week 2 materials, you can show off your knowledge by taking Quiz 2, which is due BEFORE 11:30 PM on Sunday.

Have a great week.

Brian and the Data Science specialization team

Sun 12 Oct 2014 5:01 PM PDT

Rafa's course CS109 at Harvard

Hey everyone,

A close friend of Brian, Jeff and Roger (former faculty member at Hopkins Biostat, now at Harvard), Rafael Irizarry, along with Verena Kaynig-Fittkau are teaching this Data Science class at Harvard http://cs109.github.io/2014/ and the videos and slides are all online! Check them out.

Brian

Fri 10 Oct 2014 7:25 AM PDT

Data Science Specialization Community Site

Since the beginning of the Data Science Specialization, we've noticed the unbelievable passion students have about our courses and the generosity they show toward each other on the course forums. A couple students have created quality content around the subjects we discuss, and many of these materials are so good we feel that they should be shared with all of our students.

We're excited to announce that we've created a site using GitHub Pages: http://datasciencespecialization.github.io/ to serve as a directory for content that the community has created. If you've created materials relating to any of the courses in the Data Science Specialization, please send us a pull request so we can add a link to your content on our site. You can find out more about contributing here: https://github.com/DataScienceSpecialization/DataSci

We can't wait to see what you've created and where the community can take this site!

- The JHU Data Science Lab Team

Thu 9 Oct 2014 1:30 PM PDT

Difference between a shiny server and shiny app

Hi class,

I want to clarify some questions that have come up on the forums.

Shiny app development can occur on Windows, Linux, Mac OS (and probably even other platforms). When you

do runApp() it will bring up the app in your default browser and your R session will be the server. This is how you should work on your app in development. Make sure that you have a ui.R and server.R functions in the directory where R is working and shiny is loaded (library(shiny)). You can run your app this way whether your connected to the internet or not.

A shiny server is required to deploy your app as a general web page for anyone to reach on the internet. This is a much harder process that involves server administration. In this class we do not cover shiny server administration which gets into administering web services.

Instead, for this class, Rstudio has done the server administration for us. In addition, their service, shinyapps.io offers a nice platform to manage your apps. They do this as a free service and have a premium version where you can build more complex apps with logins and other things. Next week, we'll have details about how to set up your account. In the mean time, only work on shiny locally.

Building and deploying a shiny server yourself is purely extra curricular at this point. Unless you have fairly good linux knowledge, it's not worth your time to bother and should instead focus on the app building part rather than the app serving part.

Brian

Wed 8 Oct 2014 12:05 PM PDT

Statistical Inference: On rCharts

It was pointed out in the forums that rCharts may need to be installed from devtools if you have a recent version of R. Here's the note from Ramnath

https://github.com/ramnathv/rMaps/issues/54

The devtools package on cran is a must to install. Its most useful function is the ability to install packages from github. Note, if you're on windows, I would suggest installing rtools. This allows you to build R packages (and R itself).

http://cran.r-project.org/bin/windows/Rtools/

Brian

Tue 7 Oct 2014 6:22 AM PDT

Introduction and welcome

Hi class,

A couple of first week housekeeping items. First, make sure that you've had R Programming and the Data Scientist's Toolbox. Reproducible Research would be helpful, but is not mandatory. We're running all classes monthly now, so it's worthwhile considering delaying this class until you've had those. At a minimum you must know: very basic git, basic R and very basic knitr.

An important aspect of this class is to peruse the materials in the github repository. All of the most up to date material can be found here

https://github.com/bcaffo/courses/tree/master/09_Developing

You should clone this repository as your first step in this class and make sure to fetch updates periodically. (Please send pull requests too!) It is one of the most essential components of the Specialization that you start to use Git frequently. We're practicing what we preach as well by using the tools in the series to create the series, especially git.

You can clone the whole repo with (http)

git clone https://github.com/bcaffo/courses.git

or (ssh)

git clone git@github.com:bcaffo/courses.git

The lectures are in the index.Rmd lecture files. In this class, we'll cover how to create these sorts of slides.You will see all of the R code to recreate the

lectures. Going through the R code is the best way to familiarize yourself with the lecture materials.

If you'd like to keep up with the instructors I'm @bcaffo on twitter, Roger is @rdpeng and Jeff is @jtleek. The Department of Biostat here is @jhubiostat.

Since today is the first day running all 9 courses simultaneously, we'll be live tweeting at #jhudatascience.

Mon 6 Oct 2014 8:05 AM PDT

Developing Data Products: Week 1

Welcome to Week 1 of *Developing Data Products*. This week we'll cover Shiny, Rcharts and googleVis. Make sure to watch all of the videos. When you feel that you are ready, you can demonstrate your knowledge by taking Quiz 1, which is due by 11:30 PM UTC on Sunday.

Please visit the Data Science Specialization github repository to pull the latest version of the course materials.

Also, the Course Project details are available by clicking the Course Project link in the left navigation bar. Your project is due BEFORE 11:30 PM UTC on the Sunday at the end of Week 3.

I'm glad that you decided to take *Developing Data Products*, part of the Data Science Specialization from Johns Hopkins Biostatistics!

A data product is the production output from a statistical analysis. Data products automate complex analysis tasks or use technology to expand the utility of a data informed model, algorithm or inference. This course covers the basics of creating data products using Shiny, R packages, and interactive graphics. This course focuses on the statistical fundamentals of creating a data product that can be used to tell a story about data to a mass audience.

You will learn how communicate using statistics and statistical products. Emphasis will be paid to communicating uncertainty in statistical results. You will learn how to create simple Shiny web applications and R packages for their data products. In addition, we'll cover reproducible presentations and interactive graphics.

We believe that the key word in Data Science is "science". Our specialization is focused on providing you with three things: (1) an introduction to the key ideas behind working with data in a scientific way that will produce new and reproducible insight, (2) an introduction to the tools that will allow you to execute on a data analytic strategy, from raw data in a database to a completed report with interactive graphics, and (3) on giving you plenty of hands on practice so you can learn the techniques for yourself. This course represents the final cog in a data science application, creating an end-usable data product.

We are excited about the opportunity to attempt to scale Data Science education. We intend for the courses to be self-contained, fast-paced, and interactive. We intend to run them frequently to give people with busy schedules the opportunity to work on material at their own pace.

Finally, we are interested in hearing about your reasons for taking this course and your intentions for the Data Science Specialization. Please take a few moments to complete our Pre-Course Survey: https://www.surveymonkey.com/s/8CW2XZJ?user_id=5e536cd785c4564cb685bc490fcacbf11ff1389f

Have a great week!

Brian and the Data Science Specialization team

Mon 6 Oct 2014 6:01 AM PDT