Gmail **Andrew Ho <kironide@gmail.com>**

# End of Week 1 Announcements and Survey

1 message

**Jonah Sinick** <jsinick@gmail.com> Fri, Feb 19, 2016 at 5:50 PM
To: Ali Bagherpour <ali.bagherp@gmail.com>, Andrew Ho <Kironide@gmail.com>, Chad Groft <clgroft@gmail.com>, David Bolin <david@bolin.at>, Jacob Pekarek <jpekarek@trinity.edu>, Jaiwithani <jaiwithani@gmail.com>, James Cook <cookjw@gmail.com>, Linchuan Zhang <email.linch@gmail.com>, Matthew Gentzel <magw6270@terpmail.umd.edu>, Olivia Schaefer <taygetea@gmail.com>, Sam Eisenstat <sam.eisenst@gmail.com>, Tom Guo <tomguo4@gmail.com>, Trevor Murphy <trevor.m.murphy@gmail.com>

Hi All,

The past week has been a great learning experience for me – I've really enjoyed getting to know you.

I thought a fair amount about the structure of the program ahead of time, but there's no substitute for empirical data, and I've been updating a lot based on my observations and your feedback. I'd like to share what my current thinking and hear your responses.

- **Basic skills:** There's a relatively small amount of core technical material that everyone should learn by the end of the program. I'll make some self-assessments to make it unambiguous what this common core is and for you to check whether you've learned enough.

  Roughly speaking, this set of material can be summarized as:

  – *Advanced R,* Chapters 1–6, 10 and 11
  – Visualizing data using ggplot2, corrplot, mosaicplot, GGally
  – Linear regression and logistic regression.
  – Preventing overfitting (via regularization, cross-validation, and bagging / boosting).
  – General familiarity with methods for handling nonlinearity in data and clustering (no detailed knowledge required).
  – Writing up your work in R Markdown
  – SQL
  – Python at the level of being able of being able to do a simple data science project in it (~1 day long), and write answers to coding questions in Python.

  You're all learning quickly, and I have no doubt that you'll absorb all of these skills before the end of the program.

- **Pair programming** – Some people are benefitting much more than others from working in pairs. There's enough spread on this dimension so that I think it makes sense to make pair programming optional by default.

- **Assignments vs. Personal projects** – Beyond ensuring that you absorb the above basic skills, our goal is to get you to the point where you feel comfortable doing projects of your own choice (whether individually in pairs or in larger groups) as soon as possible.

  One reason for this is that people tend to be most motivated when working on what interests them most. Another is that doing your own projects is the best practice for doing your own work in the future.

  **Please let me know when you're at this point.** We have assignments and projects of our own design intended to ease the transition, but don't want them to be a distraction. Sam and I will be available for extensive consultation on personal projects.

- **Personal project ideas** – this weekend, think about that aspects of what we've been doing so far interest

you the most and what sorts of datasets you'd be interested in exploring. I'll be talking with each of you on Monday to explore further

- **Getting help –** Prompted by some people indicating a preference to not be interrupted, Sam and I have been talking about shifting toward less frequent check-ins. We can individualize this at the level of students.

- **Saturday** – See Robert's email

**Questions for you**

Please take this survey:
https://docs.google.com/forms/d/1AfPWP3_WAumxfz-HVaf1i43laQSsoE6KBb2pHL2yOgc/viewform