STA141 End Notes
Duncan Temple Lang

Practice Data Analysis

Get experience with real data analyses
Start your own projects with available data
  public data sources
  visualize and summarize the data
  make inferences about the population.
redo analyses from blogs, papers, etc. with up-to-date data,
or apply the ideas to a new set of data.
Work in groups - develop teamwork skills

Sustaining what you’ve learned

Continue to practice programming (every day/week) or you will forget it (like any languages/skill).
Problem solve and get comfortable deciding the next steps yourselves based on what makes sense.
Independent thinking

Portfolio

Important to show employers and grad school admission something more than just course work

Need to show you
  can work independently,
  can recognize the important questions and issues
do sensible things without needing constant supervision
  can learn new concepts & technologies.
Get to know faculty beyond class - letters of recommendation.
Concrete Opportunities

I plan to run a data analysis & visualization competition in Winter.

Check the DataScience.ucdavis.edu Web site.

Also, probably some workshops, tutorials, seminars over the next 2 quarters.

More Topics

Many more topics that you would benefit from knowing

Machine learning methods (currently ECS171 or STA208)

Bayesian statistical methods (STA145)

Technology Topics

Many technology topics to learn also.

Read Job Postings (scrape them)

Hacker News, datatau, flowingdata, SimplyStatistics…

Skills from Scraping Job Posts
From 2014
Kaggle

New Related Courses

STA160 - A capstone/project course to gain experience with framing a question, cleaning, exploring, modeling, presenting the data & insights.

STA141B - Data Technologies & Intermediate R
- text mining,
- interactive data visualization (JavaScript, ...)
- advanced R
  - R packages,
  - Object Oriented Programming (OOP),
  - portability,
  - unit testing
- SQL & Databases
- Web Scraping, APIs, ...
- version control

STA141C - Large Data & High Performance Statistical Computing
- shell programming,
- efficient code,
- interfacing to C,
- python for data analysis.
- parallel computing,
- MapReduce & Hadoop/Spark software stack
Office Hours today

Last thing
Office Hours today from 12 - 1.10pm, 1147 MSB.