Assignments STT 5XXX - Spring 2017 - Example

Alan T. Arnholt

Last Updated on: Mar 20, 2019 at 02:11:48 PM

**Jan 17:**

* Go over syllabus.
* Become familiar with the Appstate [RStudio server](https://mathr.math.appstate.edu/). You will use your Appstate user name and password to log in to the server. You must be registered in the class to access the server.
* Sign-up for free accounts on [GitHub](https://github.com) and [Rpubs](https://rpubs.com).
* When you register for a free individual GitHub account, request a [student discount](https://education.github.com) to obtain a few private repositories as well as unlimited public repositories. Please use something similar to **first\_name\_last\_name** as your username when you register with GitHub. For example, my username on GitHub is *alanarnholt*. If you have a popular name such as John Smith, you may need to provide some other distinquishing characteristic in your username. Please use the same **username** for your account on Rpubs.
* Once you have a GitHub account, send an email to [arnholtat@appstate.edu](mailto:arnholtat@appstate.edu) with a Subject line of **STT5812 - GitHub Username**, and tell me in the body of your email your first name, last name, and your GitHub username. I will then manually add you as a team member to the repository in the STAT-ATA-ASU organization that has your name (**Last\_name-First\_name**). This repository will be where you store all of your work for this course. I will also change your repository to a private repository.
* In class lecture: Review of one and two sample inference, permutation tests, and bootstrap procedures — take notes
* Homework:
  + Watch this [video](https://www.youtube.com/watch?v=ZFaWxxzouCY&list=PLjTlxb-wKvXNSDfcKPFH2gzHGyjpeCZmJ&index=3) which provides some background on asking questions.
  + Read this stackoverflow thread details how to create a [minimal R reproducible example](http://stackoverflow.com/questions/5963269/how-to-make-a-great-r-reproducible-example/5963610#5963610).
  + Read [How To Ask Questions The Smart Way](http://www.catb.org/~esr/faqs/smart-questions.html) by Eric Raymond and Rick Moen and heed their advice.
  + Read [Getting used to R, RStudio, and R Markdown](https://ismayc.github.io/rbasics-book/)

1. Most due dates are 5 p.m. Friday for [CrowdGrader](https://www.crowdgrader.org/) assignments.
2. Most [CrowdGrader](https://www.crowdgrader.org/) peer reviews will be due by 11 p.m. on Tuesdays.

**Jan 19:**

* Before class watch the following Videos:
  + [Phillip Guo’s Unix-like command line tutorials (Four tutorials about 10 minutes each)](http://www.pgbovine.net/command-line-tutorial.htm)
  + [Phillip Guo’s Basic Git and Github tutorials (Three tutorials about 10 minutes each)](http://www.pgbovine.net/git-tutorial.htm)
  + [Clone a GitHub Repo with RStudio (Video)](https://www.youtube.com/watch?v=YxZ8J2rqhEM) (5:05)
* In class lecture: Review of one and two sample inference, permutation tests, and bootstrap procedures — take notes
* [Statistics Review](../../STT3851ClassRepo/Rmarkdown/StatReview.html)
* Quiz #1
* Read Chapters 1-2 of [bookdown: Authoring Books and Technical Documents with R Markdown](https://bookdown.org/yihui/bookdown/)
* Link to [Day One Activity and Code](../Rmarkdown/DayOne.html)

**Jan 24:**

* Before class:
  + Read/review [Getting used to R, RStudio, and R Markdown](https://ismayc.github.io/rbasics-book/)
  + Read the [Git and GitHub](http://r-pkgs.had.co.nz/git.html) chapter from Hadley Wickham’s book [R Packages](http://r-pkgs.had.co.nz/)
* Follow the directions from [Happy Git and GitHub for the useR](http://happygitwithr.com/) to [Introduce yourself to Git](http://happygitwithr.com/hello-git.html), [Connect to GitHub](http://happygitwithr.com/push-pull-github.html), and [Cache credentials for HTTPS](http://happygitwithr.com/credential-caching.html).
* Read/review Chapters 1-2 of [bookdown: Authoring Books and Technical Documents with R Markdown](https://bookdown.org/yihui/bookdown/)

**Jan 26:**

* Before class:
  + [Watch Managing Change with RStudio](https://www.rstudio.com/resources/webinars/rstudio-essentials-webinar-series-managing-change-part-1/) (37 minutes)
* In class lecture: Review of one and two sample inference, permutation tests, and bootstrap procedures — take notes
* [Statistics Review](../../STT3851ClassRepo/Rmarkdown/StatReview.html)
* [Schistosomiasis](../../STT3851ClassRepo/Rmarkdown/Schistosomiasis.html)
* Update README file in private repository
* Use an avatar or picture on your GitHub account
* Quiz #2

**Jan 31:**

* Before class:
  + Read chapter 1 [ISLR](http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Sixth%20Printing.pdf)
  + Watch [Opening Remarks and Examples](https://www.youtube.com/watch?v=2wLfFB_6SKI) (18:18)
  + Watch [Supervised and Unsupervised Learning](https://www.youtube.com/watch?v=LvaTokhYnDw) (12:12)
  + Read Section 3.4 of [bookdown: Authoring Books and Technical Documents with R Markdown](https://bookdown.org/yihui/bookdown/)
* Lecture - [Rmarkdown example](../../STT3851ClassRepo/Rmarkdown/FlowControl.html)
* Bootstrap - [Rmarkdown example](../../STT3851ClassRepo/Rmarkdown/TheBootstrap.html)
* [Overplotting Examples and possible Solutions](../../STT3851ClassRepo/Rmarkdown/HousingData.html)
* Clone the repository to your local machine using RStudio by following these instructions:

1. [Fork](https://help.github.com/articles/fork-a-repo/) the repository.
2. Copy the clone URL to the clipboard.
3. Click File > New Project > Version Control > Git
4. Paste the clone URL (https://github.com/YourUserName/STAT-ATA-ASU/STT5812CourseRepo.git) in the Repository URL: box.
5. Use the name (STT5812CourseRepo) in the Project directory name: box.
6. Change if needed the location in the Create project as subdirectory of: box.
7. Click the Create Project box. You should now have a local copy of the forked repository on your local machine. Congratulations!

* Set the upstream remote in your fork to this repository with the command
* git remote add upstream https://github.com/STAT-ATA-ASU/STT5812CourseRepo.git
* Verify with
* git remote -v
* To obtain updates from the upstream repository type
* git pull upstream master
* If the upstream repository is using gh-pages, use gh-pages instead of master to obtain updates.
* git pull upstream gh-pages
* If there are conflicts, you will need to resolve them before proceeding.

**Feb 2:**

* Before class:
  + Read chapter 2 [ISLR](http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Sixth%20Printing.pdf)
  + Watch [Statistical Learning and Regression](https://www.youtube.com/watch?v=WjyuiK5taS8) (11:41)
  + Watch [Curse of Dimensionality and Parametric Models](https://www.youtube.com/watch?v=UvxHOkYQl8g) (11:40)
  + Watch [Assessing Model Accuracy and Bias-Variance Trade-off](https://www.youtube.com/watch?v=VusKAosxxyk) (10:04)
* Lecture - Bias Variance Tradeoff - Take notes
* Quiz #3

**Feb 7:**

* Before class:
  + Review/read chapter 2 [ISLR](http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Sixth%20Printing.pdf)
  + Watch [Classification Problems and K-Nearest Neighbors](https://www.youtube.com/watch?v=vVj2itVNku4) (15:37)
  + Watch [Lab: Introduction to R](https://www.youtube.com/watch?v=jwBgGS_4RQA) (14:12)
* Lecture - Bias Variance Tradeoff - Take notes - [Slides](../../STT3851ClassRepo/Rmarkdown/BiasVarDer.pdf)
* Reproduce [Modified Introduction to R](../../STT3851ClassRepo/Rmarkdown/LabChap2.html). Submit your html file to [CrowdGrader](https://www.crowdgrader.org/crowdgrader/venues/view_venue/2743) no later than 5 pm Feb 10.

**Feb 9:**

* Before class:
  + Review/read chapter 2 [ISLR](http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Sixth%20Printing.pdf)
* Go over [Bias Variance Graphs](../../STT3851ClassRepo/Rmarkdown/BiasVariance.html)
* Go over [Chapter 2 Graphs](../../STT3851ClassRepo/Rmarkdown/Chap2Slides.html)
* Discuss flexible models and when to use them
* Reproduce [Modified Introduction to R](../../STT3851ClassRepo/Rmarkdown/LabChap2.html). Submit your html file to [CrowdGrader](https://www.crowdgrader.org/crowdgrader/venues/view_venue/2743) no later than 5 pm Feb 10.
* Quiz #4

**Feb 14:**

* Before class:
  + Refer to [*bookdown: Authoring Books and Technical Documents with R Markdown*](https://bookdown.org/yihui/bookdown/) as needed
* Reproduce [Writing Assignment](../../STT3851ClassRepo/Rmarkdown/WritingAssignment.html). Submit your html file to [CrowdGrader](https://www.crowdgrader.org/crowdgrader/venues/view_venue/2745) no later than 5 pm Feb 17.

**Feb 16:**

* Work Day
* Reproduce [Writing Assignment](../../STT3851ClassRepo/Rmarkdown/WritingAssignment.html). Submit your html file to [CrowdGrader](https://www.crowdgrader.org/crowdgrader/venues/view_venue/2745) no later than 5 pm Feb 17.
* Quiz #5

**Feb 21:**

* Before class:
  + Read pages 59-82 of [ISLR](http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Sixth%20Printing.pdf)
  + Watch [Simple Linear Regression and Confidence Intervals](https://www.youtube.com/watch?v=PsE9UqoWtS4) (13:01)
  + Watch [Hypothesis Testing](https://www.youtube.com/watch?v=J6AdoiNUyWI) (8:24)
  + Watch [Multiple Linear Regression and Interpreting Regression Coefficients](https://www.youtube.com/watch?v=1hbCJyM9ccs) (15:38)
* Discuss [Chapter 2 and 3 material](../../STT3851ClassRepo/Rmarkdown/Chap2Slides.html)

**Feb 23:**

* Before class:
  + Read pages 82-120 [ISLR](http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Sixth%20Printing.pdf)
  + Watch [Model Selection and Qualitative Predictors](https://www.youtube.com/watch?v=3T6RXmIHbJ4) (14:51)
  + Watch [Interactions and Nonlinearity](https://www.youtube.com/watch?v=IFzVxLv0TKQ) (14:16)
  + Watch [Lab: Linear Regression](https://www.youtube.com/watch?v=5ONFqIk3RFg) (22:10)
* Discuss [Chapter 2 and 3 material](../../STT3851ClassRepo/Rmarkdown/Chap2Slides.html)
* Go over [Bias Variance Graphs Again](../../STT3851ClassRepo/Rmarkdown/BiasVarianceGGplot.html)
* Quiz #6

**Feb 28:**

* Answer questions at the end of [Chapter 2 and 3 material](../Rmarkdown/Chap2Slides.html).
* Reproduce [Lab: Linear Regression](../../STT3851ClassRepo/Rmarkdown/LabChap3.html). Submit your html file to [CrowdGrader](https://www.crowdgrader.org/crowdgrader/venues/view_venue/2746) no later than 5 pm Mar 3.

**Mar 2:**

* Before class:
  + [Start Project 1](http://www.amstat.org/publications/jse/v16n2/datasets.pardoe.html) - Read the hyperlinked paper.
  + The data set can be downloaded from <http://www.amstat.org/publications/jse/datasets/homes76.dat.txt>, while the code book is available from <http://www.amstat.org/publications/jse/datasets/homes76.txt>
* Reproduce [Lab: Linear Regression](../../STT3851ClassRepo/Rmarkdown/LabChap3.html). Submit your html file to [CrowdGrader](https://www.crowdgrader.org/crowdgrader/venues/view_venue/2746) no later than 5 pm Mar 3.
* Quiz #7 (Chapter 3 Assessment)

**Mar 7:**

* [Work on Project 1](../../STT3851ClassRepo/Rmarkdown/DirectionsMHP.html)
* [Code for class notes](../../STT3851ClassRepo/Rmarkdown/BodyTemperatures.html)

**Mar 9:**

* [Work on Project 1](../../STT3851ClassRepo/Rmarkdown/DirectionsMHP.html)
* Quiz #8 (Chapter 3 Assessment)

**~~Mar 13 - Mar 17:~~**  **No Class - Spring Break**

**Mar 21:**

* Commit your final changes for Project 1 to your private repository and email the instructor your SHA prior to 10am March 22, 2017. Use the subject line “Project 1-STT5812-Last\_name First\_name-SHA#” in your email.

**Mar 23:**

* Before class:
  + Read chapter 5 [ISLR](http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Sixth%20Printing.pdf)
  + Watch [Estimating Prediction Error and Validation Set Approach](https://www.youtube.com/watch?v=_2ij6eaaSl0) (14:01)
  + Watch [K-fold Cross-Validation](https://www.youtube.com/watch?v=nZAM5OXrktY) (13:33)
  + Watch [Cross-Validation: The Right and Wrong Ways](https://www.youtube.com/watch?v=S06JpVoNaA0) (10:07)
* Discuss [Cross-Validation Hand Out](../../STT3851ClassRepo/Rmarkdown/Cross-ValidationInClassHO.html)
* Quiz #9

**Mar 28:**

* Before class:
  + Read/re-read chapter 5 [ISLR](http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Sixth%20Printing.pdf)
  + Watch [Lab: Cross-Validation](https://www.youtube.com/watch?v=6dSXlqHAoMk) (11:21)
* Discuss [Cross-Validation Hand Out](../../STT3851ClassRepo/Rmarkdown/Cross-ValidationInClassHO.html)
* [Reproduce *Lab: Cross-Validation*](../../STT3851ClassRepo/Rmarkdown/LabCrossValidation.html)

**Mar 30:**

* Before class:
  + Watch [Linear Model Selection and Best Subset Selection](https://www.youtube.com/watch?v=91si52nk3LA) (13:44)
  + [Forward Stepwise Selection](https://www.youtube.com/watch?v=nLpJd_iKmrE) (12:26)
  + [Backward Stepwise Selection](https://www.youtube.com/watch?v=NJhMSpI2Uj8) (5:26)
* [Finish Cross-Validation Hand Out](../../STT3851ClassRepo/Rmarkdown/Cross-ValidationInClassHO.html)
* Complete [Reproduce *Lab: Cross-Validation*](../../STT3851ClassRepo/Rmarkdown/LabCrossValidation.html). Submit your html file to [CrowdGrader](https://www.crowdgrader.org/crowdgrader/venues/view_venue/2753) no later than 5 pm March 31.

Start Project 2

* [Example Template Directory](https://github.com/STAT-ATA-ASU/STT3851Spring2016/blob/gh-pages/BDbetaTemplate)
* Use the [housedata.csv](https://github.com/STAT-ATA-ASU/STT3851Spring2016/raw/gh-pages/Data/housedata.csv) file - See [King County Housing Data](../../STT3851ClassRepo/Rmarkdown/KingCounty.html)
* Quiz 10

**Apr 4:**

* Before class:
  + Watch [Estimating Test Error Using Mallow’s Cp, AIC, BIC, Adjusted R-squared](https://www.youtube.com/watch?v=LkifE44myLc) (14:06)
  + Watch [Estimating Test Error Using Cross-Validation](https://www.youtube.com/watch?v=3p9JNaJCOb4) (8:43)
* [Notes on Credit Problem](../../STT3851ClassRepo/Rmarkdown/CreditProblemMore.html)

**Apr 6:**

* Before class:
  + Watch [Shrinkage Methods and Ridge Regression](https://www.youtube.com/watch?v=cSKzqb0EKS0) (12:37)
  + Watch [The Lasso](https://www.youtube.com/watch?v=A5I1G1MfUmA) (15:21)
  + Watch [Tuning Parameter Selection for Ridge Regression and Lasso](https://www.youtube.com/watch?v=xMKVUstjXBE) (5:27)
* Quiz #11

**Apr 11:**

* Before class review [Book Down](https://bookdown.org/yihui/bookdown/)
* Before class
  + Watch [Dimension Reduction](https://www.youtube.com/watch?v=QlyROnAjnEk) (4:45)
  + Watch [Principal Components Regression and Partial Least Squares](https://www.youtube.com/watch?v=eYxwWGJcOfw) (15:48)
  + Watch [Lab: Best Subset Selection](https://www.youtube.com/watch?v=3kwdDGnV8MM) (10:36)
  + Watch [Lab: Forward Stepwise Selection and Model Selection Using Validation Set](https://www.youtube.com/watch?v=mv-vdysZIb4) (10:32)
  + Watch [Lab: Model Selection Using Cross-Validation](https://www.youtube.com/watch?v=F8MMHCCoALU) (5:32)
  + Watch [Lab: Ridge Regression and Lasso](https://www.youtube.com/watch?v=1REe3qSotx8) (16:34)

**Apr 13:**

* Work on Project 2
* Work on *Lab: Subset Selection Methods*
* Quiz #12

**~~Apr 17-18:~~ No Class - State Holidays**

**Apr 20:**

* Work on Project 2
* Work on *Lab: Ridge Regression and the Lasso*

**Apr 25:**

* Turn in Project 2 - NLT April 26, 10 am
* Make sure a single .Rmd with the sections outlined in the [Example Paper](../../STT3851ClassRepo/BDbetaTemplate/TemplateBookDown.html) is in your Project2 directory under your private course repository.
* Send an email to the instructor using a subject of “House Data Predictions” with an attached vector (csv file) of your predictions using the data set housedataTEST.csv. Note: your vector should contain 4229 predictions.

**Apr 27:**

* [Poster Template Directory](https://github.com/STAT-ATA-ASU/STT3851Spring2016/tree/gh-pages/Poster)
* Work on Presentation
* Quiz #13

**May 2:**

* Work on Presentation

**May 8:**

* Final - Presentations 9:00-11:30