CWRU DSCI351-451: Week01a-f

Roger H. French

August 28, 2018

Contents

1.1.1.1 Reading, Homeworks, Projects, SemProjects

- Readings:
  - Peng R Programming pages 4 to 33
- Homeworks
  - HW1 given out thursday, due next tuesday
- Data Science Projects:
  -
- 451 SemProjects:
  -
- Friday Comm. Hour
  -

1.1.1.2 Syllabus

1.1.1.3 What we need to do this week

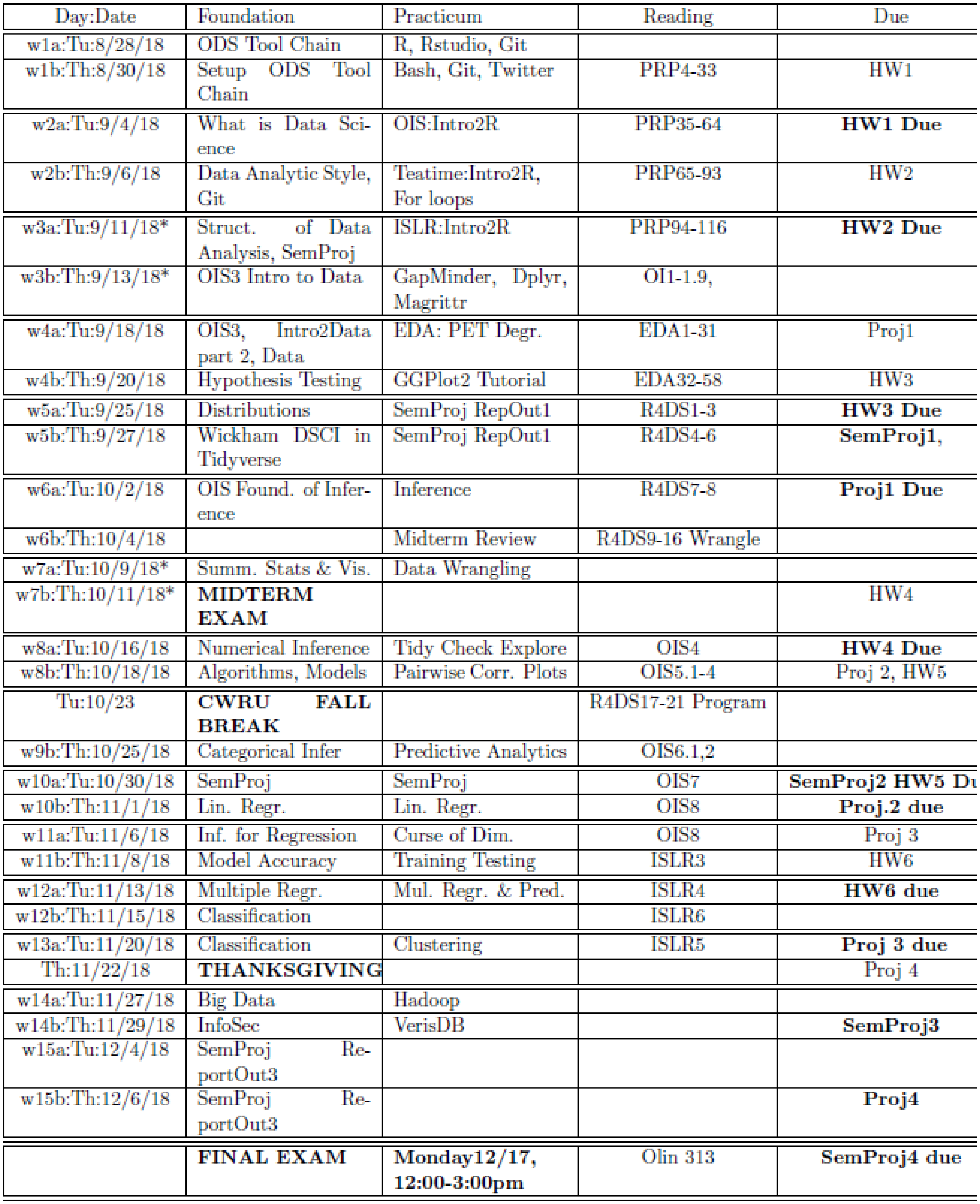
1. Setup VDI

- Rstudio
- Drag icons of R, Rstudio, Git Bash, Spyder, Jupyter Notebook, HipChat to desktop

2. Setup Git

- make H:\Git folder
- git config name and email

3. Setup Bitbucket account
4. Setup Kaggle account
5. Setup Twitter account
6. Setup StackExchange account



Day:Date	Foundation	Practicum	Reading	Due
w1a:Tu:8/28/18	ODS Tool Chain	R, Rstudio, Git		
w1b:Th:8/30/18	Setup ODS Tool Chain	Bash, Git, Twitter	PRP4-33	HW1
w2a:Tu:9/4/18	What is Data Science	OIS:Intro2R	PRP35-64	**HW1 Due**
w2b:Th:9/6/18	Data Analytic Style, Git	Teatime:Intro2R, For loops	PRP65-93	HW2
w3a:Tu:9/11/18*	Struct. of Data Analysis, SemProj	ISLR:Intro2R	PRP94-116	**HW2 Due**
w3b:Th:9/13/18*	OIS3 Intro to Data	GapMinder, Dplyr, Magrittr	OI1-1.9,	
w4a:Tu:9/18/18	OIS3, Intro2Data part 2, Data	EDA: PET Degr.	EDA1-31	Proj1
w4b:Th:9/20/18	Hypothesis Testing	GGPlot2 Tutorial	EDA32-58	HW3
w5a:Tu:9/25/18	Distributions	SemProj RepOut1	R4DS1-3	**HW3 Due**
w5b:Th:9/27/18	Wickham DSCI in Tidyverse	SemProj RepOut1	R4DS4-6	**SemProj1**,
w6a:Tu:10/2/18	OIS Found. of Inference	Inference	R4DS7-8	**Proj1 Due**
w6b:Th:10/4/18		Midterm Review	R4DS9-16 Wrangle	
w7a:Tu:10/9/18*	Summ. Stats & Vis.	Data Wrangling		
w7b:Th:10/11/18*	**MIDTERM EXAM**			HW4
w8a:Tu:10/16/18	Numerical Inference	Tidy Check Explore	OIS4	**HW4 Due**
w8b:Th:10/18/18	Algorithms, Models	Pairwise Corr. Plots	OIS5.1-4	Proj 2, HW5
Tu:10/23	**CWRU FALL BREAK**		R4DS17-21 Program	
w9b:Th:10/25/18	Categorical Infer	Predictive Analytics	OIS6.1,2	
w10a:Tu:10/30/18	SemProj	SemProj	OIS7	**SemProj2 HW5 Du**
w10b:Th:11/1/18	Lin. Regr.	Lin. Regr.	OIS8	**Proj.2 due**
w11a:Tu:11/6/18	Inf. for Regression	Curse of Dim.	OIS8	Proj 3
w11b:Th:11/8/18	Model Accuracy	Training Testing	ISLR3	HW6
w12a:Tu:11/13/18	Multiple Regr.	Mul. Regr. & Pred.	ISLR4	**HW6 due**
w12b:Th:11/15/18	Classification		ISLR6	
w13a:Tu:11/20/18	Classification	Clustering	ISLR5	**Proj 3 due**
Th:11/22/18	**THANKSGIVING**			Proj 4
w14a:Tu:11/27/18	Big Data	Hadoop		
w14b:Th:11/29/18	InfoSec	VerisDB		**SemProj3**
w15a:Tu:12/4/18	SemProj ReportOut3			
w15b:Th:12/6/18	SemProj ReportOut3			**Proj4**
	FINAL EXAM	**Monday12/17, 12:00-3:00pm**	Olin 313	**SemProj4 due**

Figure 1: DSCI351-351M-451 Syllabus

7. Setup Slack account
8. Git Clone

- 18-sdle-teatime
  - for quick introduction to data science techniques and tools
- For Class-Prof Repo
  - Clone your forked Class Repo

1.1.1.4 Your Open Data Science Tool Chain

1.1.1.4.1 Its all about a Data Science Tool Chain

- Use R and build on the communities foundatin
- Use Rstudio as a comfy environment
- Share your Open Data and Open Source Code
- Produce Reproducible Science with Rmarkdown
  - Use Creative Commons Licenses
  - Or other Open Source Licenses
  - Such as the Gnu Public License: GPL
  - Or one of my favorites, the Apache License

Pilot your Data Science studies using available data

- Find available data sets
- Before starting the costly process of making data

Use Git repositories

- For version control
- For Collaboration
- For Open Science sharing

1.1.1.4.2 Online Git Server Communities

- We use BitBucket Account
  - In class, for our class code repositories
  - These are private repositories
- You'll probably also want a GitHub account.
  - Many Rprojects are there, and
  - you can fork their repo's as inspect the code very easily.

1.1.1.4.3 Kaggle Account

- Kaggle started as a data science competition site
- Its recently been bought by Google
  - And give free R and Python Notebooks
  - Including use of free GPUs
- It has a very good Intro to R, Python, Machine Learning etc.
  - First R Tutorial: Getting staRted in R: First Steps
  - 2nd R Tutorial, Level 1, on Modeling
  - 3rd R Tutorial, Level 2, on tidyverse data manipulation

1.1.1.4.4 Twitter used for Data Science

As part of setting up our Data Science Tool Chain

- Signup for a Twitter account
- Using Twitter in university research
- 10 Commandments of Twitter for Academics

Data Science People to follow on Twitter

- @hadleywickham
- @jtleek Jeff Leek JHU
- @rdpeng Roger Peng JHU

- @simplystats
- @Rbloggers
- @JennyBryan
- @hspter Hilary Parker
- @NSSDeviations
- @dataandme
- @rstudio
- @rstudiotips
- @R_Programming
- @CRANberriesFeed
- @timoreilly
- @kaggle
- @SciPyTip
- @PyData
- @debian
- @ubuntu
- @GuardianData
- @UpshotNYT
- @EdwardTufte
- @ProjectJupyter
- @doctorow Cory Doctorow
- @gvanrossum Founder of Python
- @NateSilver538
- @cutting Founder of Hadoop
- @RProgLangRR
- @BitbucketStatus
- @CWRUITS_STATUS

- @cshirky Clay Shirky
-

1.1.1.4.5 Sign up for a Stack Exchange Account

Stack Exchange, Stack Overflow

- are a Q&A community focused on many topics.

Stack Overflow allows you to search by tag

- r and rmarkdown are useful tags for SO

Stack Exchange's Tour of Stack Overflow

Specific Stack Exchange websites

- for SX Data Science

- for SX Statistics on Cross Validated
- for SX Open Data

1.1.1.4.6 Efficiently browse you SX sites

- Google (but more random)
- The Stack Exchange apps
- Using an RSS Feed reader such as Feedly is a good way

1.1.1.4.7 Slack, another component of Agile Sofware Development

- Slack.com
  - We have a CWRU DSCI Slack room
  - There is Slack app for phones
  - And client for computers, its on vdi.
  - Slack client available for windows, mac and Linux
- an online collaboration tool

1.1.1.5 You Online Data Science Portfolio

- Doing open, reproducible data science
- Lets you share a portfolio of codes and projects
- Cite it in your resume
- Build a community of supporters and collaborators

1.1.1.5.1 An Example, Emeline Liu

- emelineliu.com
  - This website, which runs off of Github Pages and Jekyll, is my latest project.
  - Right now, I'm using Poole as a foundation for my website/blog.

1.1.1.6 Links

- http://www.r-project.org
- Rory Winston, for the Learning R Intro
- StackExchange http://stackexchange.com/sites

- Twitter http://twitter.com
- Slack http://slack.com
- CWRU-DSCI Slack
- emelineliu.com
- Github Pages
- Jekyll
- Poole