The basic ideas behind for this new version of the R course are:

1. Coordinate the course with other MSSP courses
2. Organize the course so that it can be taught by other instructors
3. Enrich the course so that it includes more
   1. statistical methodology
   2. variety in data types
4. “Flip” the course so that students come to class after they have prepared for the topics to be discussed. In class, the topics that build on that preparation will be the focus of the class. (This approach is being adopted for all of our classes.) Each class has to be self-contained
5. Complete class objectives and plan (Rproj for each class)
6. Fill in classes that are currently blank

The course has tk specific sections

1. Introduction: Exploratory data analysis, document production, basic R syntax  
   Project: EDA --
2. Cleaning and organizing data  
   Project: Data Cleaning – USDA data?
3. Data manipulation - R programming  
   Project: Text analysis
4. Database access from R - SQL  
   Data access and updating using SQL
5. Working with big data  
   Project: Visualization & EDA with FF
6. Advanced data access & presentations  
   Data access through API
7. Interactive presentations  
   Real-time API  
   Plotly, Shiny

Notes: 24july17

**TO DO ================================================**

1. Complete the 40-class list
   1. The pre-class work (read, do) -- (Add “Bring” that includes material the students should prepare and print for class.)
   2. the in-class work (quiz, objectives, during class)
2. Move some topics into the classes by skipping some RDS chapters or making their content ancillary to other topics
   1. making chapter 17 (class 24) the last RDS chapter that is the class topic by itself OR
   2. making chapter 20 (class 29) the last RDS chapter that is the class topic by itself
3. Topics list – these topics need to be included in the course – currently are not included. The table includes some potentially associated projects

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Topic | Req'd? | Classes | Project |  |
| SQL | Y | 3 | OSHA Data |  |
| Big Data | Y | 3 |  |  |
| API | Y | 3 |  |  |
| Shiny | Y | 3 |  |  |
| Plotly | Y | 1 |  |  |
| Mapping | Y | 1 |  |  |
| R markdown for slides | N | 1 |  |  |
| ggvis | N |  |  |  |
|  |  |  |  |  |

1. Project table – Projects for which we have data, experience, code, etc We don't have enough projects yet

|  |  |  |  |
| --- | --- | --- | --- |
| Project list | Statistical issue/methodology | Associated topic |  |
| NOAA | lm(), t-tests |  |  |
| OSHA |  |  |  |
| Fires |  |  |  |
| Twitter | Text analysis, network analysis |  |  |
| Emoji | Categorical data - logistic? |  |  |
| Securities Analysis |  |  |  |
| Interest rate Analysis |  |  |  |
| Break even analysis | lm() |  |  |
| Branding analysis | Text analysis, network analysis (twitter) |  |  |
| Market segmentation | Clustering ( spatial, trees, models) |  |  |
| Agricultural Data |  | Data Cleaning and organizing |  |
|  |  |  |  |

Random notes

I really like the material in this link, but don’t know what to do with it yet

<http://spatial.ly/2013/12/introduction-spatial-data-ggplot2/>

<https://github.com/marketplace>

<https://guides.github.com/activities/hello-world/>

<https://githubuniverse.com/?utm_source=github&utm_medium=banner&utm_campaign=ww-dotcom-dash-universe-20170717>

<https://guides.github.com/features/pages/>

<http://lib.stat.cmu.edu/datasets/>

<https://fred.stlouisfed.org/categories/121>

<https://bitly.com/blog/bitly-basics-bitlinks/>

<http://gss.norc.org/Get-The-Data>

<https://github.com/caesar0301/awesome-public-datasets>

<https://catalog.data.gov/dataset/u-s-hourly-precipitation-data>

<http://www.nber.org.ezproxy.bu.edu/>

http://www.caida.org/data/overview/

## R materials and online resources

[RDS](http://r4ds.had.co.nz/) ***R for Data Science*,** Garrett Grolemund, Hadley Wickham

[R-Intro](https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf) *An Introduction to R*,   
W. N. Venables, D. M. Smith & the R Core Team

[Adv R](http://adv-r.had.co.nz/Data-structures.html) *Advanced R*, Hadley Wickham   
[Style Guide](http://adv-r.had.co.nz/Style.html)

[R packages](http://r-pkgs.had.co.nz/) ***R packages*,** Hadley Wickham

[Text Mining](http://tidytextmining.com/) *Text mining with R*, Julia Silge and David Robinson

[Pro Git](https://git-scm.com/book/en/v2) *Pro Git*, Scott Chacon and Ben Straub

[Git](https://git-scm.com/) Git Installation

[ggplot2](http://ggplot2.org/) – ggplot2 documentation

[tidyweb](http://tidyverse.org/) an ecosystem of packages

[RStudio](https://www.rstudio.com/) R IDE and information hub --   
RStudio Installation

[TUG](https://www.tug.org/) TeX users group

[Happy Git](http://happygitwithr.com/) Happy Git and GitHub for the useR

Reference Cards: [Ref1](https://cran.r-project.org/doc/contrib/Short-refcard.pdf) [Ref2](https://cran.r-project.org/doc/contrib/Baggott-refcard-v2.pdf) [Ref3](https://www.stats.ox.ac.uk/~snijders/siena/Rrefcard.pdf) [Regex](https://www.rstudio.com/wp-content/uploads/2016/09/RegExCheatsheet.pdf)

## [TeX Live](https://www.tug.org/texlive/) TeX installation

[](https://www.r-project.org/) The R Project for Statistical Computing –   
R Installation

[](https://www.r-bloggers.com/) R news and tutorials

[Plotly](https://plot.ly/) Interactive charts and dashboards

[Cookbook](http://www.cookbook-r.com/) Cookbook for R

[TeXstudio](http://www.texstudio.org/) LaTeX document workbench

## *R for Data Science* Part and chapter numbering:

Part and Chapter numbers differ between the online version and the print version of *R for Data Science*. The content is identical in other respects. This course references the online numbering scheme. The outline below shows part and chapter numbering for the online version.

**R for Data Science (links to online version)**

[Welcome](http://r4ds.had.co.nz/index.html)

* [**1** Introduction](http://r4ds.had.co.nz/introduction.html)
* **I Explore**
* [**2** Introduction](http://r4ds.had.co.nz/explore-intro.html)
* [**3** Data visualisation](http://r4ds.had.co.nz/data-visualisation.html)
* [**4** Workflow: basics](http://r4ds.had.co.nz/workflow-basics.html)
* [**5** Data transformation](http://r4ds.had.co.nz/transform.html)
* [**6** Workflow: scripts](http://r4ds.had.co.nz/workflow-scripts.html)
* [**7** Exploratory Data Analysis](http://r4ds.had.co.nz/exploratory-data-analysis.html)
* [**8** Workflow: projects](http://r4ds.had.co.nz/workflow-projects.html)
* **II Wrangle**
* [**9** Introduction](http://r4ds.had.co.nz/wrangle-intro.html)
* [**10** Tibbles](http://r4ds.had.co.nz/tibbles.html)
* [**11** Data import](http://r4ds.had.co.nz/data-import.html)
* [**12** Tidy data](http://r4ds.had.co.nz/tidy-data.html)
* [**13** Relational data](http://r4ds.had.co.nz/relational-data.html)
* [**14** Strings](http://r4ds.had.co.nz/strings.html)
* [**15** Factors](http://r4ds.had.co.nz/factors.html)
* [**16** Dates and times](http://r4ds.had.co.nz/dates-and-times.html)
* **III Program**
* [**17** Introduction](http://r4ds.had.co.nz/program-intro.html)
* [**18** Pipes](http://r4ds.had.co.nz/pipes.html)
* [**19** Functions](http://r4ds.had.co.nz/functions.html)
* [**20** Vectors](http://r4ds.had.co.nz/vectors.html)
* [**21** Iteration](http://r4ds.had.co.nz/iteration.html)
* **IV Model**
* [**22** Introduction](http://r4ds.had.co.nz/model-intro.html)
* [**23** Model basics](http://r4ds.had.co.nz/model-basics.html)
* [**24** Model building](http://r4ds.had.co.nz/model-building.html)
* [**25** Many models](http://r4ds.had.co.nz/many-models.html)
* **V Communicate**
* [**26** Introduction](http://r4ds.had.co.nz/communicate-intro.html)
* [**27** R Markdown](http://r4ds.had.co.nz/r-markdown.html)
* [**28**Graphics for communication](http://r4ds.had.co.nz/graphics-for-communication.html)
* [**29** R Markdown formats](http://r4ds.had.co.nz/r-markdown-formats.html)
* **30** R Markdown work

| **Class #** | **Pre-Class** | **Class** | **References quizzes, projects, notes** |
| --- | --- | --- | --- |
|  | First class: no pre-class work.  (**REW:** this session needs reformatting and rewriting)  handouts should include an R info sheet that includes R-bloggers, etc – this may be in RDS. Check. **[SL]**  **During class**   * Students install R, Rstudio, git * Simple program example gives students information they need to complete the Pre-Class 2   [R Style Guide](http://adv-r.had.co.nz/Style.html)  [FormatR wiki](https://yihui.name/formatr/) | **Objectives: Intro**   * S knows where to get course information and help.   **Outline**   * Course Objectives * Course Outline (blackboard, daily procedure, assignments, grading) * Install (R, Rstudio, git) – instructions on screen * RStudio tour (windows, projects, packages, r notebook) * Simple program - R notebook **[HW]** * R Markdown preview for next class | initial survey  Notes: |
|  | **Read**  RDS 1 Introduction  RDS26 Introduction, 27 R markdown  R-Intro 1,2,3,4  Do  RDS 1.4 Prerequisites – installations  RDS 27.2.1 Excercises  Problem set **[tk]**  Install LaTex, texStudio  Bring | **Objectives: Markdown**   * S can begin new Project, start an existing project. * S can produce R Markdown docs that combine text, code, & graphics. Know how to use Lorem Ipsum to dummy docs. * S can extract code from R Markdown docs with knitr::purl * S can code tables in knitr::kable, kableExtra * scan, read dat   **[sample doc with tabulated data, graphic, plot, text]**  During class  Make document that includes R-intro coding and homework examples and previews ggplot2() | * quiz * project |
|  | Read  Need reading  Watch  (<https://www.youtube.com/watch?v=Y9XZQO1n_7c>)  Do  script for git, github  cheatsheet  **Bring** | Objectives: git, github  set up remote  push – pull  example program with various reading  need examples to store, revise, recall, etc  During class  Git and github – script files and documents that include material from classes 1 & 2. | **quiz** |
|  | Read  R-Intro 5,6,7  Adv R [Foundations](http://adv-r.had.co.nz/Data-structures.html)  Do  In Adv R [Foundations](http://adv-r.had.co.nz/Data-structures.html) Exercises:  Class 4 Problem Set  **Bring**  **cheatsheet for apply** | Objectives: R Data Structures  S can use R data structures to manipulate  apply  data frames  tibbles vs data frames vs data tables  built-in data – getting doc for  During class  r data structures (vector, lists, matrix, data frame)  apply family  apply family | **quiz** |
|  | Read  RDS Part 1: Explore  [Explore: Introduction](http://r4ds.had.co.nz/explore-intro.html)  [Explore: Data Visualization](http://r4ds.had.co.nz/data-visualisation.html)  cheatsheet  reading for ggmap, ggthemes  Do  http://www.cookbook-r.com/Graphs/Plotting\_distributions\_(ggplot2)/  **Bring** | Objectives: Visualization a  S can  During Class  Visualization : dataset reading and plotting  ggplot  map example  dataset???? | **quiz** |
|  | Read  RDS Part 1: Explore  [Explore: Introduction](http://r4ds.had.co.nz/explore-intro.html)  [Explore: Data Visualization](http://r4ds.had.co.nz/data-visualisation.html)  cheatsheet  Do  assignment  Bring | Objectives: Visualization b  S can  During Class  Visualization  ggplot  ggmap  ggthemes | **quiz** |
|  | Read  RDS Explore – 4 Workflow: Basics  R-Intro 8, 9, 10  Do  **Bring** | Objectives: Coding  During Class  Exercise ggplot  subsetting and sorting  data = USDA??  include apply example with inline function | **quiz** |
|  | Read  RDS 5 Data Transformation with dplyr  cheatsheet  doc  vignettes  Do  dplyr()  **Bring** | Objectives: dplyr a  During Class  data = ag??  Subsetting and Sorting | **quiz** |
|  | Read  RDS 5 Data Transformation with dplyr  Do  dplyr()  **Bring** | Objectives dplyr b  During Class  Subsetting and Sorting | **quiz** |
|  | Read  RDS 6 workflow Scripts  R-Intro 10 Writing your own functions  Do  **Bring** | Objectives: debugging  During Class  Function and Loops | **quiz** |
|  | Read  RDS 7 Exploratory Data Analysis  Do  EDA  **Bring** | Objectives  During Class  Read Data from a variety of sources  ?? is one day enough?? only if we have done most of it in the previous classes | **quiz** |
|  | Read  RDS 10 Tibble  Do  EDA  **Bring** | Objectives  During Class  Assignment  tidyr article start it here – will have to show class how to access the article material on github | project 1 due  project 2  **tidyr article?** |
|  | Read  RDS 11 Data Import  Do  Reading: R-Intro 12 Graphical procedures  Graphics with ggplot2 (<http://www.statmethods.net/advgraphs/ggplot2.html)>  **Bring** | Objectives  During Class  qplot, ggplot | **quiz** |
|  | Read  RDS 12 Tidy data  Reading: Hadley Wickham – Tidy Data  Do  **Bring** | Objectives  During Class | **quiz** |
|  | Read  RDS 13 Relational data  Do  USDA project  **Bring** | Objectives  During Class | **quiz** |
|  | Read  RDS 14 Strings  Do  Reading: Hadley Wickham – Tidy Data  **Bring** | Objectives  During Class |  |
|  | Read  RDS 15 Factors  Do  **Bring** | Objectives  During Class |  |
|  | Read  RDS 16 Dates and times  Lubridate  Do  Reading: Hadley Wickham – Tidy Data  **Bring** | Objectives  During Class | project 2 due |
|  | Read  RDS 17 Introduction  RDS 18 Pipes  Do | Objectives  During Class |  |
|  | Read  RDS 19 functions  Do  **Bring** | Objectives  During Class |  |
|  | Read  RDS 19 Functions  Do  **Bring** | Objectives  During Class |  |
|  | Read  RDS 20 Vectors  Do  **Bring** | Objectives  During Class |  |
|  | Read  RDS 21  Purr & Apply  Do  **Bring** | Objectives  During Class |  |
|  | Read  RDS 21 Purr & Apply  Do  **Bring** | Objectives  During Class |  |
|  | Read  [START SQL here?]  [considering not putting Model part of RDS as specific focus classes – put this information into projects]  RDS 18 Modeling - modelr  Do  **Bring** | Objectives  During Class |  |
|  | Read  RDS 19 Modeling  [considering not putting Model part of RDS as specific focus classes – put this information into projects]  Do  **Bring** | Objectives  During Class |  |
|  | Read  [considering not putting Model part of RDS as specific focus classes – put this information into projects]  RDS 19 Modeling  Do  **Bring** | Objectives  During Class |  |
|  | Read  RDS 20 Modeling  [considering not putting Model part of RDS as specific focus classes – put this information into projects]  Do  **Bring** | Objectives  During Class  Objectives  During Class |  |
|  | Read  RDS 20 Modeling  [considering not putting Model part of RDS as specific focus classes – put this information into projects]  Do  **Bring** | Objectives  During Class |  |
|  | Read  The communication topics MUST be allocated to projects and/or put into the beginning of the semester  RDS Communication – 21 Markdown  RED 22 – ggplot22  Do  **Bring** | Objectives  During Class |  |
|  | Read  The communication topics MUST be allocated to projects and/or put into the beginning of the semester  RDS 23 – r markdown formats  Do  **Bring** | Objectives  During Class |  |
|  | Read  The communication topics MUST be allocated to projects and/or put into the beginning of the semester  RDS 24 r markdown workflow  Do  **Bring** | Objectives  During Class |  |
|  | Read  Do  **Bring** | Objectives  During Class  shiny |  |
|  | Read  Do  **Bring** | Objectives  During Class  Working with big data |  |
|  | Read  Do  **Bring** | Objectives  During Class  plotly |  |
|  | Read  Do  **Bring** | Objectives  During Class  sweave |  |
|  | Read  Do  **Bring**  R Intro 8 Probability distributions | Objectives  During Class  Probability Statistics |  |
|  | Read  Do  **Bring** | Objectives  During Class |  |
|  | Read  Do  **Bring** | Objectives  During Class |  |
|  | Read  Do  **Bring** | Objectives  During Class |  |
| final | Things that must be completed before the end of finals |  |  |