

Topics outline for 679: Data Science.

Week Date	Topics	chapter	Exercises	Project
<b>0</b> <b>1/16</b> <b>MLK no</b> <b>class.</b>				
<b>1</b> <b>1/23</b>	Broad contours data science Importing (readr and fread)	r4ds: 2, 11	fix badRead.csv 11.2.2: 4, 5	<p>What are you interested in? Economics? Sports? Politics? Twitter? Current events? Something more specific? Unemployment? Real estate prices Voting outcomes / ACA signup?</p> <p>Is there a playground that underlies your interest? Once you have thought about this, try to find a a data source.</p> <p>Find a unique data source.</p> <p>Import it into R.</p>
<b>2</b> <b>1/30</b>	dplyr, %>%	r4ds: 5, 12	all exercises in chapter 5 are good!	*Write a one paragraph description of the data and a one paragraph description of your interest. Why are you interested in it? What do you think is fun about it? This step is essential because it will help to guide your future steps with the data. Moreover, this paragraph might serve as a first draft of the first paragraph of your project's introduction. Due 2/6.
<b>3</b> <b>2/6</b>	tidying review: ggplot2	r4ds: 18, 3		<p>Tidy your data. Print out 10 random rows of your tidy data. Make an awesome visualization. Write a caption to describe why it's awesome. Due 2/13.</p> <p>Find more data that “links” to your data! What else to explore?</p>
<b>4</b> <b>2/13</b>	relational data, (playgrounds!)	r4ds: 13		<p>Make your playground by linking the data.</p> <p>What new visualizations can you make? Are they variations on a theme which could be incorporated via “interactions” with knob/dials/buttons/etc?</p>

Week Date	Topics	chapter	Exercises	Project
5 2/20	shiny			Put visualizations into shiny, with some knob/dials/buttons/etc!
6 2/27	<p>Broad contours of “Machine Learning”</p> <ul style="list-style-type: none"> <li>- Prediction vs inference,</li> <li>- unsupervised vs supervised,</li> <li>- assessing model accuracy</li> </ul> <p>Nonlinear regression; splines, lowess, generalized additive models</p> <p>Random Forests and tree based methods</p>	ISLR: 1,2,7,8	p52: 1, 2, 3, 4, 7(a,b,c), 8	
7 3/6	network analysis, PCA, SVD, topic modeling.			
8 3/13	Share shiny app with class. Inspect others app.			
spring break				
9 3/27	Ridge and Lasso.	ISLR 6		<p>Bring your thesis statement to class on 4/3. This statement should be</p> <ol style="list-style-type: none"> <li>(1) argumentative (make a claim, not just broadcast a topic),</li> <li>(2) controversial (debatable, not obvious),</li> <li>(3) supportable (defensible, given the resources from class and the assignment),</li> <li>(4) <b>focused</b> (answers the prompt and fits the page requirements, not too big or too small),</li> <li>(5) significant (why is it worth arguing for?),</li> <li>(6) directed (it is a condensed argument and hints at or almost outlines the direction of the paper)</li> </ol> <p>This is quoted from <a href="#">here</a>. Click for more.</p>

Week Date	Topics	chapter	Exercises	Project
<b>10</b> <b>4/3</b>	Discuss thesis statements.  Resampling techniques for parameter stability (bootstrap) and estimation of prediction error (cross validation)	ISLR: 5		Bring abstracts / executive summaries on 4/10.
<b>11</b> <b>4/10</b>	Presentations			
<b>12</b> <b>4/17</b>	Presentations			
<b>13</b> <b>4/24</b>	TBD			
<b>14</b> <b>5/1</b>	Exam  <b>projects due 5/1.</b>			
<b>15</b> <b>5/8</b>	None. (skipped week 1)			
	Linear regression, Logistic regression	ISLR: 3,4	p120: 3,4,9,15 p168: 4, 6, 7.	