The future of educational content development is plain text
The future of educational content development is plain text.
The future of educational content development is plain text & cloud-based adaptable.
The future of educational content development is plain text & cloud based & distributed.
jtleek.com/talks
The real heros

John Muschelli
@strictlystat

Sean Kross
@seankross
MSD in R
6 Courses
35K+ Enrollments

Executive Data Science
5 Courses
150K+ Enrollments

Genomic Data Science
9 Courses
230k+ Enrollments

Data Science
10 Courses
4.4M+ Enrollments
8K Completers
Course materials for the Data Science Specialization: https://www.coursera.org/specialization/hudatascience/1

Add topics

- 261 commits
- 4 branches
- 0 releases
- 13 contributors

[Files and commit history]

- 01_DataScientistToolbox
- 02_RProgramming
- 03_GettingData
- 04_ExploratoryAnalysis
- 05_ReproducibleResearch
- 06_StatisticalInference
- 07_RegressionModels
- 08_PracticalMachineLearning
- 09_DevelopingDataProducts

Latest commit ce88321 on Mar 16, 2016

- Merge pull request #21 from jaysonvriissimo/patch-4
- Update PDF
- Chicago data
- Fix non-working gplot code
- Edited URL in 05_ReproducibleResearch/organizingADataAnalysis/index.Rmd
- Merge commit '33aa06c9'
- Update index.Rmd
- Update index.Rmd
- Merge remote-tracking branch 'origin/master'
- update
- Initial commit
- update

https://github.com/DataScienceSpecialization/courses
Label issues and pull requests for new contributors

Now, GitHub will help potential first-time contributors discover issues labeled with help wanted or good first issue.

Go to Labels

Clear current search query, filters, and sorts

<table>
<thead>
<tr>
<th>Status</th>
<th>Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open</td>
<td>[link]</td>
</tr>
<tr>
<td>Closed</td>
<td>[link]</td>
</tr>
</tbody>
</table>

- [link] frame = false ===> frame.plot = F
  #168 by mortezasabri was closed on Feb 22

- [link] Line 73 - ln
  #167 by mortezasabri was closed on Feb 21

- [link] Add Data Science Interview Questions & Answers
  #158 by yonidahan was closed on Feb 17, 2017

- [link] update install_github syntax
  #141 by Canuteson was closed on Sep 12, 2018

https://github.com/DataScienceSpecialization/courses
Know about data science

Expensive computer

Access to instruction

Access to connections

Income security

Appropriate programs

Right jobs being posted
“Make data science accessible to the widest audience possible by building a program with widely satisfied prerequisites that you can do entirely through a web browser for little or no cost.”

Source: https://www.youtube.com/watch?v=o72YZcNhUNY
“Make data science accessible to the widest audience possible by building a program with widely satisfied prerequisites that you can do entirely through a web browser for little or no cost”
Your web browser

Through the internet

Software & Computers from Google, Amazon, etc...

Icons made by Smashicons and Freepik from Flaticon and licensed CC-BY 3.0
$450
Course 0: Introduction to CBDS

Course 1: How To Use a Chromebook

Course 2: Google & The Cloud

Course 3: Organizing Data Science Projects

Course 4: Version Control

Course 5: Introduction to R

Course 6: Data Tidying

Course 7: Data Visualization

Course 8: Getting Data

Course 9: Data Analysis

Course 10: Written and Oral Communication in Data Science

Course 11: Getting a Job in Data Science
The Timeline

Feb 2018

- Content development starts

April

- First meeting with Yo

Aug 31: Projected Course Set Completion
The Timeline

Feb 2018

content
development
starts

April

first meeting
with Yo

May

Aug 31:
Projected
Course Set
Completion

May 21st:
Learning
Begins!
Course 1: How To Use a Chromebook
Course 2: Google & The Cloud
Course 3: Organizing Data Science Projects
Course 4: Version Control
Course 5: Introduction to R
Course 6: Data Tidying
Course 7: Data Visualization
Course 8: Getting Data
Course 9: Data Analysis
Course 10: Written and Oral Communication in Data Science
Course 11: Getting a Job in Data Science
<table>
<thead>
<tr>
<th>Content Development</th>
<th>Leek</th>
<th>Johnson</th>
<th>Ellis</th>
<th>Hadavand</th>
<th>Muschelli</th>
<th>Kross</th>
<th>Collado-Torres</th>
<th>Jager</th>
<th>McClymont</th>
<th>Myint</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administration &amp; Tutoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Goals

1. Program must be maintainable
2. Program must be updatable
3. Program must be creatable in distributed way
4. Program must be accessible
5. Program must be free or low cost
Goals

1. Program must be maintainable
2. Program must be updatable
3. Program must be creatable in distributed way
4. Program must be accessible
5. Program must be free or low cost
“The comment is about your presentation, and in particular the way you speak. You nearly always start squeezing your voice as a sentence goes on, many times with ‘vocal fry’ near the end.”
Goals

1. Program must be maintainable
2. **Program must be updatable**
3. Program must be creatable in distributed way
4. Program must be accessible
5. Program must be free or low cost
Repositories

33,499 repository results

- **benjamin-chan/GettingAndCleaningData**
  - Description: Getting and Cleaning Data
  - Updated on Sep 18, 2016

- **Xiaodan/Coursera-Getting-and-Cleaning-Data**
  - Description: Repo for Coursera.com online course: Getting and Cleaning Data
  - Updated on Oct 11, 2015

- **bgentry/coursera-getting-and-cleaning-data-project**
  - Description: course project for Coursera "Getting and Cleaning Data"
  - Updated on May 18, 2016

Languages

- R: 29,958
- HTML: 569
- Rebol: 77
- Python: 15
- Jupyter Notebook: 13
- Java: 7
- Shell: 6
- CSS: 5

Source: https://github.com/search?utf8=%E2%9C%93&q=getting+and+cleaning+data&type=
Number of library calls

Thousands of R scripts after start date

Library call
- library(dplyr)
- library(reshape2)
- library(plyr)
- library(data.table)
- library(tidyr)
- library(stringr)
- library(reshape)
- library(sqlidata)
- library(readr)
- library(downloader)

Source: Wang et al. in prep
Thanks Hadley!

Source: Wang et al. in prep
Goals

1. Program must be maintainable
2. Program must be updatable
3. Program must be creatable in distributed way
4. Program must be accessible
5. Program must be free or low cost
Goals

1. Program must be maintainable
2. Program must be updatable
3. Program must be creatable in distributed way
4. Program must be accessible
5. Program must be free or low cost
Berkeley Will Delete Online Content

Starting March 15, the university will begin removing more than 20,000 video and audio lectures from public view as a result of a Justice Department accessibility order.

By Carl Straumsheim // March 6, 2017

The University of California, Berkeley, will cut off public access to tens of thousands of video lectures and podcasts in response to a U.S. Justice Department order that it make the educational content accessible to people with disabilities.

Today, the content is available to the public on YouTube, iTunes U and the university’s webcast.berkeley site. On March 15, the university will begin removing the more than 20,000 audio and video files from those platforms -- a process that will take three to five months -- and require users sign in with University of California credentials to view or listen to them.

The university will continue to offer massive open online courses on edX and said it plans to create new

Goals

1. Program must be maintainable
2. Program must be updatable
3. Program must be creatable in distributed way
4. Program must be accessible
5. Program must be free or low cost
“But after promising a reordering of higher education, we see the field instead coalescing around a different, much older business model: helping universities outsource their online master's degrees for professionals”
The Elements of Data Analytic Style

A guide for people who want to analyze data.

Jeffrey Leek
Plain text course material
muschellij2 fixed GS links one more time

001_program_philosophy.md update typos ; update all videos to have cc'ing 7 months ago
002_why_automated.md addressing issues on GH 7 months ago
003_data_science_process.md fixed GS links one more time 5 months ago
004_how_to_learn.md fixed GS links one more time 5 months ago
00_welcome.md addressing issues on GH 7 months ago
01_finding_help.md fixed GS links one more time 5 months ago
02_account_setup.md fixed GS links one more time 5 months ago
03_google_account.md fixed GS links one more time 5 months ago
04_other_accounts.md fixed GS links one more time 5 months ago
05_intro_to_project.md fixed GS links one more time 5 months ago
06_googlesheets.md fixed GS links one more time 5 months ago

Source: https://github.com/jhudsl/cds_intro
00_welcome.md
001_program PHILOSOPHY.md
002_why automated.md
003_data science_process.md
004_how to learn.md
01_finding help.md
02_account setup.md
03_google account.md
04_other accounts.md
05_intro to project.md
06_googlesheets.md
07_rstudio cloud.md
08_googledocs.md
09_googleslides.md
10_datacamp.md
### A Data Science Project Example

For this example, we're going to use an example analysis from a data scientist named Hilary Parker. Her work can be found on her blog, and the specific project we'll be working through here is from 2013 and titled "Hilary: the most poisoned baby name in US history". (https://hilaryparker.com/2013/01/30/hilary-the-most-poisoned-baby-name-in-us-history/)
Choosing an appropriate username is important. Some combination of your first and last name is a good idea. For example, if your name were Jane Doe, a username such as "JaneDoe" or "Jane.Doe" would work. If the first username you attempt is taken, you can try another, similar username. In this case, maybe try "J Doe".
Plain text quizzes
Leanpub (mc quizzes)

{quiz, id: quiz_003_data_science_process, random-question-order: true}

### The Data Science Process quiz

{choose-answers: 4}

?1 Which of these is NOT an effective way to communicate the findings of your analysis?

- C) save code locally on your computer
- C) print code out and store in a desk drawer
- o) write a blog post
- o) publish a paper
- o) publish a news article
- o) write a report and share it with your team
- o) write a report for your boss
- o) give a talk at a conference and make materials available online
Within the swirl project in the [Chromebook Data Science Space on RStudio Cloud](https://rstudio.cloud/spaces/3919/join?access_code=RUUQ%2BeEgKea0oMF7EJy4UePldyBBMu7d0amv2KFC) use the `swirl()` function and navigate to the course: `CBDS Introduction to R`. Then, navigate to the lesson `L04 Basic Commands Q01 Swirl`. Complete this swirl module. Once complete, paste the code provided at the end of the swirl module here.
Plain text videos
Our philosophy with building this course and this program is to try to make data science accessible to the widest audience possible.

This course is part of the Chromebook Data Science series of courses all designed with the same philosophy in mind.

These courses are designed to tackle some of the challenges that prevent people from getting into data science in the first place. Some of those challenges are geographic - we’ll talk more about that later. Some are due to the price of education - that is why we are offering these courses as MOOCs. But one of the key barriers is that the type of computer you usually need to do data science is expensive.

Chromebooks, on the other hand, are a very cheap type of computer. Chromebooks aren’t exactly like normal computers and they have a few unique characteristics. They are usually very cheap. They are designed mostly to use the web. You don’t “install” any software on the computer itself. Instead of “apps” and “software” you simply go to websites for your work. A simple way to think about it is that a Chromebook is a computer that only lets you use an internet browser like Chrome. You can’t really do much on the computer itself.

Some people call this way of working - working only through the internet - cloud computing. It’s called cloud computing because the computer you are using most of the time is not the one sitting in front of you. You are using the internet to access tools and computers to do your work. But the physical computers doing the work are stored somewhere else - it could be nearby or on the other side of the globe. That is why people call the computers “in the cloud”.

The goal of Chromebook Data Science is not that you have to use a Chromebook to finish the program, it is just that you could use a Chromebook to finish the whole program. You can finish the entire sequence of courses using any computer with an internet connection and a web browser. We took this approach because we want data science to be accessible to everyone. We have found that in earlier classes we taught online, the cost of computers, difficulties installing software, and lack of computing resources prevented from students from completing our courses. We wanted to strip all those barriers away so that more students would have access to our program.

We also believe that the future of data science is increasingly cloud based. So this educational choice matches a trend we see in the field that we can help you take advantage of. It is less and less likely that you will work only on your laptop as a data scientist. Through the internet you will access data and computing power so that you can magnify the impact of what you are working on. We hope to show you how to use those resources to maximize the value you can bring as a new data scientist.

We do recognize that internet access is also a limiting factor for many people. We have tried to make it so that you don’t have to download data so hopefully the broadband requirements will be minimal. We hope that if internet access is a challenge for you that you can leverage the resources you have - whether they are local libraries, coffee shops, or internet cafes to complete this program. If that isn’t an option for you we’d love to hear from you and see if we can find ways to make data science accessible to everyone, everywhere.

Source: https://github.com/jhudsl/cds_intro
Chromebook Data Science

http://jhudatascience.org/chromebookdatascience/
Create Videos from R Markdown Documents with Ari

The real beauty of R Markdown is that it is becoming the plain text platform from which all kinds of data products can be generated: websites, blogs, books, web applications, slide presentations, and academic publications. Ari is an R package which adds video to this portfolio of data products. The following video was generated by Ari from a single R Markdown document:

Amazon Polly

Turn text into lifelike speech using deep learning

Get started with Amazon Polly

Listen to Amazon Polly >>

Amazon Polly is a service that turns text into lifelike speech, allowing you to create applications that talk, and build entirely new categories of speech-enabled products. Amazon Polly is a Text-to-Speech service that uses advanced deep learning technologies to synthesize speech that sounds like a human voice.

With dozens of lifelike voices across a variety of languages, you can select the ideal voice for your application.

Source: https://aws.amazon.com/polly/
library(ari)

# set up poly credentials

ari_spin(images = list.files("./img"),
          paragraphs = readLines('script.md'),
          output = 'output.mp4',
          voice = 'joanna')
Program Philosophy

How to Use a Chromebook

Introduction to Chromebook Data Science: Program Philosophy

Unlisted

3 views

chromebookdata

Uploaded on May 20, 2018

Category: People & Blogs

Source: https://youtu.be/qNC5OGra4d8
Plain text exercises
swirl teaches you R programming and data science interactively, at your own pace, and right in the R console!
Now that we know there are data from 76 different college majors in this dataset, let's describe the shape of the variable we're interested in.

Load the `ggplot2` package so that we can look at the shape of the variable `percWomen`.

```r
library(ggplot2)
```

Hint: The package has already been installed. Load `ggplot2` from the package library.
Putting it all together
library(didactr)
course_dir="."
check_structure(course_dir)

out <- check_course(course_dir)
created <- create_images(out)

out = check_course(course_dir)
vids_created <- create_videos(out)

out = check_course(course_dir)
vids_uploaded = vids_to_youtube(out, Course=Course)

a = update_youtube_link(out)
<table>
<thead>
<tr>
<th>lesson</th>
<th>md_file</th>
<th>id</th>
<th>gs_name</th>
<th>cnum</th>
<th>course</th>
<th>lesson_name</th>
<th>mod_time_gs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 00_welcome</td>
<td>manuscript/00_welcome.md</td>
<td>18q2gRHXGZxBl7pSwCQh_HTmgo9o99DO39Q72Qk...</td>
<td>00_cdsintro_welcome</td>
<td>00</td>
<td>cdsintro</td>
<td>welcome</td>
<td>2018-05-09 19:26:13</td>
</tr>
<tr>
<td>2 001_program_philosophy</td>
<td>manuscript/001_program_philosophy.md</td>
<td>1s7Sqaq6A4qVa9D9T63B7KjKXMz2c6gG9G4_1ieQl</td>
<td>001_intro_program_philosophy</td>
<td>001</td>
<td>intro_program_philosophy</td>
<td></td>
<td>2018-05-21 12:46:35</td>
</tr>
<tr>
<td>3 002_why_automated</td>
<td>manuscript/002_why_automated.md</td>
<td>1Frdyn8RR8AE8x9cMTzfnFkTPxEhH_5wKAKZQ0zwk8</td>
<td>002_cdsintro_why_automated</td>
<td>002</td>
<td>cdsintro</td>
<td>why_automated</td>
<td>2018-05-14 16:42:54</td>
</tr>
<tr>
<td>4 003_data_science_process</td>
<td>manuscript/003_data_science_process.md</td>
<td>1ST3SyuWjyRyy7myFKwuxESExSThM_wJ7Mx5Ck...</td>
<td>003_cdsintro_data_science_process</td>
<td>003</td>
<td>cdsintro</td>
<td>data_science_process</td>
<td>2018-05-11 17:53:11</td>
</tr>
<tr>
<td>5 01_finding_help</td>
<td>manuscript/01_finding_help.md</td>
<td>1BS05j82c7BxvJZ3F-KrGzQ33vPLWbErCzqyO6g404</td>
<td>01_cdsintro_finding_help</td>
<td>01</td>
<td>cdsintro</td>
<td>finding_help</td>
<td>2018-05-01 16:33:14</td>
</tr>
<tr>
<td>6 02_account_setup</td>
<td>manuscript/02_account_setup.md</td>
<td>1mQMEdr4OpFzurReP9i7e5v8T-kY-DNKhPqVQonZpQ</td>
<td>02_cdsintro_account_setup</td>
<td>02</td>
<td>cdsintro</td>
<td>account_setup</td>
<td>2018-04-25 17:43:05</td>
</tr>
<tr>
<td>7 03_google_account</td>
<td>manuscript/03_google_account.md</td>
<td>1O8twsZq36G6qVCYD_Bw5WQz_y4JcLti1dEgCQ</td>
<td>03_cdsintro_google_account</td>
<td>03</td>
<td>cdsintro</td>
<td>google_account</td>
<td>2018-01-19 14:36:38</td>
</tr>
<tr>
<td>8 04_other_accounts</td>
<td>manuscript/04_other_accounts.md</td>
<td>1Nbqkwy9F4_0uTHWVDoxn7f_sMqJw4LTNcvsWv3A</td>
<td>04_cdsintro_other_accounts</td>
<td>04</td>
<td>cdsintro</td>
<td>other_accounts</td>
<td>2018-01-19 14:36:48</td>
</tr>
<tr>
<td>9 05_intro_to_project</td>
<td>manuscript/05_intro_to_project.md</td>
<td>1auByv5SpghzELH-SMKLwzxWztgXd2dPC4Q53rcT4qLE</td>
<td>05_cdsintro_intro_to_project</td>
<td>05</td>
<td>cdsintro</td>
<td>intro_to_project</td>
<td>2018-05-10 17:25:14</td>
</tr>
<tr>
<td>10 06_googleSheets</td>
<td>manuscript/06_googleSheets.md</td>
<td>1EP7DrM2QjMEIDN3iPWO66oyiMiWlPcR-sropdVx...</td>
<td>06_cdsintro_googleSheets</td>
<td>06</td>
<td>cdsintro</td>
<td>googleSheets</td>
<td>2018-05-04 17:32:04</td>
</tr>
<tr>
<td>11 07_rStudio_cloud</td>
<td>manuscript/07_rStudio_cloud.md</td>
<td>1FAIaQO7qtuANhApu4fCccBoKtF9Nso5QcWULULqS...</td>
<td>07_cdsintro_rStudio_cloud</td>
<td>07</td>
<td>cdsintro</td>
<td>rStudio_cloud</td>
<td>2018-04-26 15:26:55</td>
</tr>
<tr>
<td>12 08_googleDocs</td>
<td>manuscript/08_googleDocs.md</td>
<td>1ArBuF1WPhTca0XCN6B87G1mx8ZCjhxpsqdpDzA</td>
<td>08_cdsintro_googleDocs</td>
<td>08</td>
<td>cdsintro</td>
<td>googleDocs</td>
<td>2018-01-19 14:36:16</td>
</tr>
<tr>
<td>13 09_googleSlides</td>
<td>manuscript/09_googleSlides.md</td>
<td>15OuMmP1xXuvqTMeKiaOOSCqO7T0nWraD67b_pzr...</td>
<td>09_cdsintro_googleSlides</td>
<td>09</td>
<td>cdsintro</td>
<td>googleSlides</td>
<td>2018-01-19 14:36:09</td>
</tr>
<tr>
<td>14 10_datacamp</td>
<td>manuscript/10_dataCamp.md</td>
<td>1Kgpmw0vO_jhHxKf_ULGV4pWjJnuU3uKmd2aqbHUm</td>
<td>10_cdsintro_dataCamp</td>
<td>10</td>
<td>cdsintro</td>
<td>dataCamp</td>
<td>2018-01-19 14:45:24</td>
</tr>
<tr>
<td>file</td>
<td>lesson</td>
<td>url</td>
<td>time_published</td>
<td>kind</td>
<td>etag</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------------------</td>
<td>-------------------------------------------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>-------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>00_welcome.mp4</td>
<td>00_welcome</td>
<td><a href="https://www.youtube.com/watch?v=QzSrCCGS57nE">https://www.youtube.com/watch?v=QzSrCCGS57nE</a></td>
<td>2018-05-21 02:41:32</td>
<td>youtube/video</td>
<td>QzSrCCGS57nE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>001_program_philosophy.mp4</td>
<td>001_program_philosophy</td>
<td><a href="https://www.youtube.com/watch?v=o72YzChNhUNY">https://www.youtube.com/watch?v=o72YzChNhUNY</a></td>
<td>2018-05-21 02:41:54</td>
<td>youtube/video</td>
<td>o72YzChNhUNY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>002_why_automated.mp4</td>
<td>002_why_automated</td>
<td><a href="https://www.youtube.com/watch?v=mn6fyD1t_5g">https://www.youtube.com/watch?v=mn6fyD1t_5g</a></td>
<td>2018-05-21 02:42:00</td>
<td>youtube/video</td>
<td>mn6fyD1t_5g</td>
<td></td>
<td></td>
</tr>
<tr>
<td>003_data_science_process.mp4</td>
<td>003_data_science_process</td>
<td><a href="https://www.youtube.com/watch?v=He8_ShZjm0f">https://www.youtube.com/watch?v=He8_ShZjm0f</a></td>
<td>2018-05-21 02:42:06</td>
<td>youtube/video</td>
<td>He8_ShZjm0f</td>
<td></td>
<td></td>
</tr>
<tr>
<td>01_finding_help.mp4</td>
<td>01_finding_help</td>
<td><a href="https://www.youtube.com/watch?v=aCf1W4mfTg">https://www.youtube.com/watch?v=aCf1W4mfTg</a></td>
<td>2018-05-21 02:42:07</td>
<td>youtube/video</td>
<td>aCf1W4mfTg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>02_account_setup.mp4</td>
<td>02_account_setup</td>
<td><a href="https://www.youtube.com/watch?v=Wy5_0qLEFvS">https://www.youtube.com/watch?v=Wy5_0qLEFvS</a></td>
<td>2018-05-21 02:42:12</td>
<td>youtube/video</td>
<td>Wy5_0qLEFvS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>03_google_account.mp4</td>
<td>03_google_account</td>
<td><a href="https://www.youtube.com/watch?v=8YdM_B_xow">https://www.youtube.com/watch?v=8YdM_B_xow</a></td>
<td>2018-05-21 02:42:17</td>
<td>youtube/video</td>
<td>8YdM_B_xow</td>
<td></td>
<td></td>
</tr>
<tr>
<td>04_other_accounts.mp4</td>
<td>04_other_accounts</td>
<td><a href="https://www.youtube.com/watch?v=0YWi3i3xS64">https://www.youtube.com/watch?v=0YWi3i3xS64</a></td>
<td>2018-05-21 02:42:27</td>
<td>youtube/video</td>
<td>oYWi3i3xS64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>05_intro_to_project.mp4</td>
<td>05_intro_to_project</td>
<td><a href="https://www.youtube.com/watch?v=5MGyVRIxGQ">https://www.youtube.com/watch?v=5MGyVRIxGQ</a></td>
<td>2018-05-21 02:43:12</td>
<td>youtube/video</td>
<td>5MGyVRIxGQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>06_goolesheets.mp4</td>
<td>06_goolesheets</td>
<td><a href="https://www.youtube.com/watch?v=1NgzinoA-nOA">https://www.youtube.com/watch?v=1NgzinoA-nOA</a></td>
<td>2018-05-21 02:43:16</td>
<td>youtube/video</td>
<td>1NgzinoA-nOA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>07_rstudio_cloud.mp4</td>
<td>07_rstudio_cloud</td>
<td><a href="https://www.youtube.com/watch?v=ZEqZk0yxXSM">https://www.youtube.com/watch?v=ZEqZk0yxXSM</a></td>
<td>2018-05-21 02:45:35</td>
<td>youtube/video</td>
<td>ZEqZk0yxXSM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>08_googledocs.mp4</td>
<td>08_googledocs</td>
<td><a href="https://www.youtube.com/watch?v=-4voj_XIsq0">https://www.youtube.com/watch?v=-4voj_XIsq0</a></td>
<td>2018-05-21 02:46:30</td>
<td>youtube/video</td>
<td>-4voj_XIsq0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>09_gooleslides.mp4</td>
<td>09_gooleslides</td>
<td><a href="https://www.youtube.com/watch?v=8QdVuVZVHLY">https://www.youtube.com/watch?v=8QdVuVZVHLY</a></td>
<td>2018-05-21 02:46:34</td>
<td>youtube/video</td>
<td>8QdVuVZVHLY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10_datacamp.mp4</td>
<td>10_datacamp</td>
<td><a href="https://www.youtube.com/watch?v=9g9Nnv8Rlxo">https://www.youtube.com/watch?v=9g9Nnv8Rlxo</a></td>
<td>2018-05-21 02:46:43</td>
<td>youtube/video</td>
<td>9g9Nnv8Rlxo</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Our system
Process

1. Sign up to complete a lesson
2. Pull course repo from Github
3. Create Google Slide Deck (share with team)
4. Write “tutorial” in Markua format
5. Convert to script by removing formatting
6. Create videos with ari
7. Write quizzes
8. Check course with didactr
9. Push for review
10. Get feedback
11. Update and push
The Timeline

Feb 2018: Content development starts

April: First meeting with Yo

May: Learning Begins!

Aug 31: Projected course set completion
The Timeline

- Feb 2018: content development starts
- April: first meeting with Yo
- May: May 21st: Learning Begins!
- June: #1
- July: #2, #3
- Aug: #4, #5
- Sept: Aug 31: Projected Course Set Completion
- Oct: 0-3, 4-5, 6-7, 8-9, 10-11
The Timeline

Feb 2018

content development starts

April

first meeting with Yo

May

May 21st: Learning Begins!

0-3

4-5

6-7

8-9

10-11

June

July

Aug

Sept

#1

#2

#3

#4

#5

Oct

Oct 5th
Learners finish coursework!!!
The future: plain text translations
# Tidy Data

The idea of tidy data was formalized in 2014 in a [paper](http://vita.had.co.nz/papers/tidy-data.pdf) written by a leader in the data science field, [Hadley Wickham](http://hadley.nz/). The principles of tidy data, which are discussed below, provide a standard way of formatting a data set. A tidy dataset follows a number of rules relating to how rows, columns, and spreadsheets are matched up with observations, variables, and types.

# Datos ordenados

La idea de datos ordenados se formalizó en 2014 en un [artículo](http://vita.had.co.nz/papers/tidy-data.pdf) escrito por un líder en el campo de la ciencia de datos, [Hadley Wickham](http://hadley.nz/). Los principios de datos ordenados, que se analizan a continuación, proporcionan una forma estándar de formatear un conjunto de datos. Un conjunto de datos ordenado sigue una serie de reglas relacionadas con la forma en que las filas, las columnas y las hojas de cálculo se combinan con las observaciones, las variables y los tipos.
3. There should be one spreadsheet for each type of data

**Demographic Survey Data**

<table>
<thead>
<tr>
<th>ID</th>
<th>LastName</th>
<th>FirstName</th>
<th>Sex</th>
<th>City</th>
<th>State</th>
<th>Occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Smith</td>
<td>Jane</td>
<td>female</td>
<td>Frederick</td>
<td>MD</td>
<td>Weider</td>
</tr>
<tr>
<td>2</td>
<td>Nayeef</td>
<td>Mohammed</td>
<td>male</td>
<td>Upper Darby</td>
<td>PA</td>
<td>Nurse</td>
</tr>
<tr>
<td>3</td>
<td>Doe</td>
<td>Janice</td>
<td>female</td>
<td>San Diego</td>
<td>CA</td>
<td>Doctor</td>
</tr>
<tr>
<td>5</td>
<td>Jordan</td>
<td>Alex</td>
<td>male</td>
<td>Birmingham</td>
<td>AL</td>
<td>Teacher</td>
</tr>
</tbody>
</table>

**Doctor’s Office Measurements Data**

<table>
<thead>
<tr>
<th>ID</th>
<th>LastName</th>
<th>FirstName</th>
<th>Height_inches</th>
<th>Weight_lbs</th>
<th>Insulin</th>
<th>Glucose</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Smith</td>
<td>Jane</td>
<td>65</td>
<td>180</td>
<td>0.60</td>
<td>163</td>
</tr>
<tr>
<td>2</td>
<td>Nayeef</td>
<td>Mohammed</td>
<td>75</td>
<td>215</td>
<td>1.46</td>
<td>150</td>
</tr>
<tr>
<td>4</td>
<td>Doe</td>
<td>Janice</td>
<td>62</td>
<td>124</td>
<td>0.72</td>
<td>177</td>
</tr>
<tr>
<td>5</td>
<td>Jordan</td>
<td>Alex</td>
<td>77</td>
<td>160</td>
<td>1.23</td>
<td>205</td>
</tr>
</tbody>
</table>

**Datos de la encuesta demostrativa**

<table>
<thead>
<tr>
<th>ID</th>
<th>LastName</th>
<th>FirstName</th>
<th>Sex</th>
<th>City</th>
<th>State</th>
<th>Occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Smith</td>
<td>Jane</td>
<td>female</td>
<td>Frederick</td>
<td>MD</td>
<td>Weider</td>
</tr>
<tr>
<td>2</td>
<td>Nayeef</td>
<td>Mohammed</td>
<td>male</td>
<td>Upper Darby</td>
<td>PA</td>
<td>Nurse</td>
</tr>
<tr>
<td>4</td>
<td>Doe</td>
<td>Janice</td>
<td>female</td>
<td>San Diego</td>
<td>CA</td>
<td>Doctor</td>
</tr>
<tr>
<td>5</td>
<td>Jordan</td>
<td>Alex</td>
<td>male</td>
<td>Birmingham</td>
<td>AL</td>
<td>Teacher</td>
</tr>
</tbody>
</table>

**Dos tipos diferentes de datos**

**Mediciones del consultorio del médico**

<table>
<thead>
<tr>
<th>ID</th>
<th>LastName</th>
<th>FirstName</th>
<th>Height_inches</th>
<th>Weight_lbs</th>
<th>Insulin</th>
<th>Glucose</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Smith</td>
<td>Jane</td>
<td>65</td>
<td>180</td>
<td>0.60</td>
<td>163</td>
</tr>
<tr>
<td>2</td>
<td>Nayeef</td>
<td>Mohammed</td>
<td>75</td>
<td>215</td>
<td>1.46</td>
<td>150</td>
</tr>
<tr>
<td>4</td>
<td>Doe</td>
<td>Janice</td>
<td>62</td>
<td>124</td>
<td>0.72</td>
<td>177</td>
</tr>
<tr>
<td>5</td>
<td>Jordan</td>
<td>Alex</td>
<td>77</td>
<td>160</td>
<td>1.23</td>
<td>205</td>
</tr>
</tbody>
</table>
**Datos de la encuesta demográfica**

<table>
<thead>
<tr>
<th>ID</th>
<th>LastName</th>
<th>FirstName</th>
<th>Sex</th>
<th>City</th>
<th>State</th>
<th>Occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1004</td>
<td>Smith</td>
<td>Jane</td>
<td>female</td>
<td>Frederick</td>
<td>MD</td>
<td>Welder</td>
</tr>
<tr>
<td>4587</td>
<td>Nayef</td>
<td>Mohammed</td>
<td>male</td>
<td>Upper Darby</td>
<td>PA</td>
<td>Nurse</td>
</tr>
<tr>
<td>1727</td>
<td>Doe</td>
<td>Janice</td>
<td>female</td>
<td>San Diego</td>
<td>CA</td>
<td>Doctor</td>
</tr>
<tr>
<td>6879</td>
<td>Jordan</td>
<td>Alex</td>
<td>male</td>
<td>Birmingham</td>
<td>AL</td>
<td>Teacher</td>
</tr>
</tbody>
</table>

**Dos tipos diferentes de datos**

**Mediciones del consultorio del médico.**

<table>
<thead>
<tr>
<th>ID</th>
<th>LastName</th>
<th>FirstName</th>
<th>Height_inches</th>
<th>Weight_lbs</th>
<th>Insulin</th>
<th>Glucose</th>
</tr>
</thead>
<tbody>
<tr>
<td>1004</td>
<td>Smith</td>
<td>Jane</td>
<td>65</td>
<td>180</td>
<td>0.60</td>
<td>163</td>
</tr>
<tr>
<td>4587</td>
<td>Nayef</td>
<td>Mohammed</td>
<td>75</td>
<td>215</td>
<td>1.46</td>
<td>150</td>
</tr>
<tr>
<td>1727</td>
<td>Doe</td>
<td>Janice</td>
<td>62</td>
<td>124</td>
<td>0.72</td>
<td>177</td>
</tr>
<tr>
<td>6879</td>
<td>Jordan</td>
<td>Alex</td>
<td>77</td>
<td>160</td>
<td>1.23</td>
<td>205</td>
</tr>
</tbody>
</table>

Source: https://www.youtube.com/watch?v=sUU4Qha75xc
Creating Courses from Scratch

Source: vignettes/creating_courses_from_scratch.Rmd

Quick rundown

Creating a Course

1. Open RStudio.
2. Install didactr
   ```r
   devtools::install_github("muschellj2/didactr")
   ```
3. Load didactr.
   ```r
   library(didactr)
   ```
4. Create a course. Change `course_name` to your course name, and change `root_path` to where you want the course created on disk:
   ```r
   didactr::create_course(
     course_name = "My Course Name",
     root_path = "path/to/course"
   )
   ```
The future of educational content development is plain text & cloud based & distributed
Thanks

**Students**
Simina Boca, Hilary Parker, Andrew Jaffe, Alyssa Frazee, Nick Carchedi, Leo Collado Torres, Leslie Myint, Prasad Patil, Claire Ruberman, Jack Fu, Sara Wang, Kayode Sosina, Sarah McClymont + many visitors/interns/student collaborators!

**Postdocs**
Abhi Nellore, Kai Kammers, Shannon Ellis, Aboozar Hadavand, Lucy D’Agostino McGowan

**Mentors**
Rafa Irizarry, Giovanni Parmigiani, John Storey, James Powell

**Genomics Collaborators**
Ben Langmead, Andrew Jaffe, Kasper Hansen, Margaret Taub, all of Hopkins Genomics + many more

**JHU DaSL Collaborators**
Roger Peng, Brian Caffo, Stephanie Hicks, John Muschelli, Leah Jager

**JHU DaSL Staff**
Ira Gooding, Jessica Crowell, Sean Kross, Nick Carchedi, Ashley Johnson

**Hopkins Admin**
Karen Bandeen-Roche, Christy Wyskiel, Sukon Kanchanaraksa, Mike Klag, Ellen MacKenzie + many others

**Hebcac/YO**
Ed Sabatino, Liz Torres Brown

**Problem Forward**
Jamie McGovern, Kenny Morales, Tiana Bell