Data Wrangling

Stian Håklev, shaklev@gmail.com

http://reganmian.net

Special Topics in Learning Analytics

March 5, 2015

```
 1  Questions
 2  ----------------
 3  Q-0  <style> .matrix_question td, .matrix_question th { text-align: center; } </style> <script> function eval_matrix(
    question_id, matrix_options) { // Concat the values from the groups var result = ''; for(var opt in matrix_options) {
    result = result + $('input:radio[name='+question_id+'_g'+opt+']:checked').val() + '|'; } $('[name="answer['+question_id+'][
    answer]"]').val(result); } function setup_matrix_question(question_id, matrix_options, matrix_headings) { $('[
    name="answer['+question_id+'][display]"]') .before( '<table class="matrix_question table"> \ <thead> \ <tr
    id="'+question_id+'_head"> \ <th></th> \ </tr> \ </thead> \ <tbody id="'+question_id+'_body"> \ </tbody> \ </table>') .
    hide(); var $qhead = $('#'+question_id+'_head'); for(var opt in matrix_headings) { $qlabel = $('<th />').text(
    matrix_headings[opt]); $qhead.append($qlabel); } var $qbody = $('#'+question_id+'_body'); for(var opt in matrix_options) {
    $qrow = $('<tr />'); $qrow.append($('<td />').css('text-align', 'left').text(matr███████████████████████████████████████
    matrix_headings.length; i++) { $qrow.append( $('<td />') .append($('<input />').a███████ Coursera quiz ████████████████
    '_g' + opt, value: i})) ); } $qbody.append($qrow); } // Hook on form submit $('#q████████████████████████ ix(
    question_id, matrix_options); return true; }); } // ==================================████████████████████████████
    addEventListener('load', function() { setup_matrix_question( 'matrix_q1', [ 'For the amount of time I invested in this
    course, I\'m happy with what I learned', 'The course materials were presented in an engaging manner', 'I would like to take
    a more advanced course in this topic', 'I learned what I was hoping to learn in this course', 'I will use what I learned in
    this course in my job', 'The course increased my interest in Psychology', 'I see the value in Psychology', 'I think
    knowledge of Psychology is important in everyday life' ], [ 'Strongly Disagree', 'Disagree', 'Neutral', 'Agree', 'Strongly
    Agree', ] ); }); </script> If you did complete a fair amount of the course, please answer any of the following questions
    that apply. If a question asks about something you didn't do, just skip it. To begin: Please indicate your level of
    agreement with the following statements:
 4  Q-1 On average, how many hours per week did you dedicate to this course?
 5  Q-2  <script> window.addEventListener('load', function() { setup_matrix_question( 'matrix_q2', [ 'a. Lecture videos', 'b. In
    -video Quizzes', 'c. Exercises', 'd. Assignments', 'e. Completing Peer-Evaluations', 'f. Feedback from Peer Evaluations',
    'g. Community Building and Discussions', ], [ 'Not at all useful', 'Somewhat useful', 'Neutral', 'Useful', 'Very useful', ]
    ); }); </script> Rate the value of each of the following components of the course.
 6  Q-3 How would you rate the pacing of this course?
 7  Q-4 How would you rate the difficulty of this course?
 8  Q-5 How would you rate the length of this course?
 9  Q-6 Which of the following did you use or consult during the course? (Check all that apply)
10  Q-7 In terms of what parts of the course you did (and why) which of the following statements fits you best?
11  Q-8 We're interested in how you space out (or not) the timing of when you watch the lectures. Which of the following
    statements comes closest to describing you?
12  Q-9 We're also interested in how you watched the lectures. Which of the following best describes your typical way of
    watching lectures.
13  Q-10    Of course I felt the need to try out some of our assessment technologies in this class to see how they would work
    in a MOOC context, so I'd love your reaction to that. Let's start with Digital Labcoat. Assuming you tried it, which of the
    following statements best reflects your experience with Digital Labcoat.
14  Q-11    OK, now mTuner. mTuner is interesting because we could have just used the multiple-choice assessment tool within
    Coursera but we really want to use ours … partly because it is our little creation, but mostly to get the reaction from
    learners like you who were in a good position to compare it to other assessment types. So I'd like to ask a few questions
    related to mTuner. Here goes … First of all, clearly one purpose of mTuner assessments is to assess how well you learned
    the course material. With this purpose in mind, which of the following do you most agree with?
15  Q-12    In addition, another goal of mTuner activities is to promote deeper learning during the assessment process. We're
    curious to know if you feel we achieved that goal. So with this in mind, which of the following statements do you most
    agree with?
16  Q-13    OK, now peerScholar. We hope peerScholar helps you learn in many different ways, but I'll focus on two things here;
    communication and thinking. First with respect to communication. In peerScholar you had to write an original composition,
    then you had to read the work of others and comment on it, then you got to see the comments that people gave to your work.
```

```
 23
 24   Student Answers
 25   ---------------
 26
 27   [2054]
 28   Q-0 [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    0.00
 29   Q-0 [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    0.00       4|4|5|4|4|5|4|4|
 30   Q-1 [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    1.00       4 - 6
 31   Q-2 [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    0.00
 32   Q-2 [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    0.00       5|4|2|4|3|3|3|
 33   Q-3 [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    0.00       Slightly too fast
 34   Q-4 [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    1.00       Slightly too difficult
 35   Q-5 [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    1.00       Slightly too long
 36   Q-7 [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    1.00       Initially I only planned to watch the lectures, but at some
      point I decided to also take part in the other assessments and activities.
 37   Q-8 [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    1.00       When the lectures become available I will often watch 2 or 3 at
      a time, then take a break.
 38   Q-9 [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    1.00       I sat in front of the computer and watched the lectures without
      any other distractions.
 39   Q-10    [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    1.00      I enjoyed doing it, but didn't feel I learned much.
 40   Q-11    [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    1.00      mTuner was a little more complicated to understand at
      first, but eventually I did think it was better than the normal Coursera multiple-choice test. That is, I think it provided
      a better assessment of what I learned.
 41   Q-12    [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    1.00      OK, maybe I wouldn't learn more doing a typical multiple-
      choice test, but I don't think the mTuner was really much better than a typical multiple-choice test in terms of promoting
      learning.
 42   Q-13    [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    1.00      Honestly, I don't think it had much effect on my abilities
      to write or understand English. I didn't feel I learned much at all.
 43   Q-14    [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    1.00      Yes I was thinking a lot, but mostly when I wrote my own
      piece, not so much during the assessment of the reflection phases.
 44   Q-15    [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    1.00      I really didn't see much of a difference that mattered.
 45   Q-16    [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    1.00      Good
 46   Q-17    [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    1.00      I feel pretty neutral about this. It did give us a sense of
      who you are, but other than that I didn't really see it as all that relevant or useful. But it wasn't offensive either.
 47   Q-18    [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    1.00      You talking about this did get me to think about it at
      least a little more deeply than I have previously. I didn't think that much about it, but I am a little more aware of some
      of the issues.
 48   Q-19    [01] Sun  7 Jul 2013  2:02 PM UTC (UTC +0000)    0.00
```

```
1  usercat,relevant-to-aca-study-Why-enroll,skills-help-job-Why-enroll,credential-cv-Why-enroll,prestigious-uni-Why-enroll,
   course-fun-Why-enroll,hours-per-week,gender,age-group,highest-level-educ,region,studied-subj-before,work-plan-todo,moocs-
   enrolled-not-completed,moocs-finished-completed,moocs-enrolled,income,english,anything-else
2  Lifelong Learner,Neutral,Strongly Agree,Neutral,Strongly Agree,Neutral,4-6,Male,26-35 years,Less than high school,
   Europe,"Yes, I have done some personal reading or exploring on my own.","I plan to complete all the requirements, including
   watching all videos, completing all quizzes and homework, and earning a statement of accomplishment.",>10,>10,4-6,My income
   is average for my community.,Is my second language.,0
3  Lifelong Learner,Disagree,0,Strongly Agree,Strongly Agree,Strongly Agree,4-6,Male,26-35 years,Completed post-secondary
   education (college or university degree),North America,"Yes, I have taken some courses or have some work experience.","I
   plan to complete all the requirements, including watching all videos, completing all quizzes and homework, and earning a
   statement of accomplishment.",>10,>10,>10,My income is average for my community.,Is my second language.,0
4  Academic / Professor,Disagree,Disagree,Strongly Disagree,Strongly Disagree,Agree,1-3,Male,26-35 years,Completed post-
   graduate education (masters or doctoral degree),Asia,"Yes, I have done some personal reading or exploring on my own.",I
   plan to watch most videos and complete quizzes but not do the homework assignments.,>10,1-3,>10,My income is above average
   for my community.,Is my second language.,0
5  Lifelong Learner,Disagree,Neutral,Disagree,Disagree,Neutral,1-3,Male,26-35 years,Completed some post-secondary education
   but did not earn a degree,Europe,"Yes, I have done some personal reading or exploring on my own.","I plan to complete all
   the requirements, including watching all videos, completing all quizzes and homework, and earning a statement of
   accomplishment.",7-9,7-9,1-3,My income is average for my community.,Is my second language.,0
6  Professional,Strongly Disagree,Disagree,Strongly Disagree,Strongly Agree,Strongly Agree,4-6,Male,46-55 years,Completed post-
   graduate education (masters or doctoral degree),Europe,"No, not at all.",I plan to watch most videos and complete quizzes
   but not do the homework assignments.,7-9,7-9,4-6,My income is average for my community.,is my third or later language.,
   Curiosity
```

usercat	relevant-to-aca-study-W	skills-help-job-Why-en	credential-cv-Why-enr	prestigious-u	course-fun-W	hours-per-we	gender	age-group
Lifelong Learner	Neutral	Strongly Agree	Neutral	Strongly Agre	Neutral	06-Apr	Male	26-35 year
Lifelong Learner	Disagree	0	Strongly Agree	Strongly Agre	Strongly Agre	06-Apr	Male	26-35 year
Academic / Professor	Disagree	Disagree	Strongly Disagree	Strongly Disa	Agree	03-Jan	Male	26-35 year
Lifelong Learner	Disagree	Neutral	Disagree	Disagree	Neutral	03-Jan	Male	26-35 year
Professional	Strongly Disagree	Disagree	Strongly Disagree	Strongly Agre	Strongly Agre	06-Apr	Male	46-55 year
Professional	Strongly Disagree	Agree	Disagree	Neutral	Strongly Agre	06-Apr	Male	46-55 year
Lifelong Learner	Disagree	Disagree	Strongly Disagree	Strongly Disa	Strongly Agre	06-Apr	Male	46-55 year
Researcher	Strongly Disagree	Disagree	Agree	Strongly Disa	Agree	03-Jan	Male	18-25 year
Researcher	Disagree	Disagree	Disagree	Agree	Strongly Agre	03-Jan	Female	26-35 year
Professional	Neutral	Strongly Disagree	Strongly Disagree	Neutral	Strongly Agre	06-Apr	Male	26-35 year
Professional	Strongly Disagree	Neutral	Strongly Disagree	Neutral	Agree	06-Apr	Male	26-35 year
Student	Strongly Agree	Disagree	Neutral	Agree	Strongly Agre	06-Apr	Female	46-55 year
Professional	Strongly Disagree	Strongly Agree	Strongly Disagree	Strongly Disa	Strongly Disa	03-Jan	Male	26-35 year
Professional	Disagree	Agree	Neutral	Strongly Disa	Agree	03-Jan	Male	26-35 year
Lifelong Learner	Strongly Disagree	Neutral	Strongly Disagree	Strongly Disa	Agree	06-Apr	Female	46-55 year
Lifelong Learner	Neutral	Neutral	Strongly Disagree	Neutral	Strongly Agre	09-Jul	Male	26-35 year
None of the above	Strongly Disagree	Agree	Strongly Disagree	Neutral	Strongly Agre	03-Jan	Male	26-35 year
Professional	Strongly Disagree	Neutral	Disagree	Agree	Strongly Agre	06-Apr	Male	26-35 year
Student	Disagree	Strongly Disagree	Strongly Disagree	Strongly Disa	Strongly Agre	03-Jan	Male	18-25 year
Professional	Strongly Disagree	Neutral	Disagree	Neutral	Strongly Agre	06-Apr	Male	18-25 year
Student	Agree	Strongly Disagree	Strongly Disagree	Strongly Disa	Strongly Agre	03-Jan	Female	18-25 year
Lifelong Learner	Agree	Agree	Neutral	Agree	Strongly Agre	06-Apr	Female	18-25 year
Lifelong Learner	Strongly Agree	Strongly Agree	Neutral	Neutral	Strongly Agre	03-Jan	Male	18-25 year
Lifelong Learner	Strongly Disagree	Agree	Neutral	Neutral	Agree	06-Apr	Female	26-35 year
Professional	Strongly Disagree	Neutral	Strongly Disagree	Neutral	Strongly Agre	09-Jul	Male	18-25 year
Lifelong Learner	Neutral	Neutral	Disagree	Neutral	Agree	06-Apr	Male	18-25 year

```
10  Load the files
11 ▾ ```{r}
12  stat <- read.csv("~/Dropbox/Laurie/Enc/precourse-stats.txt.csv")
13  prog2 <- read.csv("~/Dropbox/Laurie/Enc/pre-course-programming2.txt.csv")
14  psych <- read.csv("~/Dropbox/Laurie/Enc/precourse-psych.txt.csv")
15  abed <- read.csv("~/Dropbox/Laurie/Enc/Aboriginal Worldviews//you-and-your-experience.txt.csv")
16  mental <- read.csv("~/Dropbox/Laurie/Enc/Intro to Social Work//you-and-your-experience.txt.csv")
17 ▴ ```
18
19  Add course column, and join
20 ▾ ```{r, results='hide'}
21  stat$course <- "Statistics"
22  prog2$course <- "Programming 2"
23  abed$course <- "Aboriginal Education"
24  psych$course <- "Psychology"
25  mental$course <- "Mental Health"
26  db <- join(stat,prog2, type="full")
27  db <- join(db,abed, type="full")
28  db <- join(db,psych, type="full")
29  db <- join(db,mental, type="full")
30 ▴ ```
```

```r
# Clean up columns

Turn empty strings and 0's into NAs
```{r}
db <- as.data.frame(lapply(db, function(x){
  x <- as.character(x)
  x <- replace(x, x == "0", NA)
  x <- replace(x, x == "", NA)
  x <- as.factor(x)}))
```

Normalize hours per week

```r
```{r, results='hide'}
db$hours <- db$hours.per.week
db$hours.per.week <- NULL

replacelist <- list(
  c(c("10-12", "13-15", ">15"), "10 or more"),
  c(c("0 (I'm not taking the course-just browsing.)", "1"), "1-3"))

db$hours <- multireplace(db$hours, replacelist)

db$hours <- ordered(db$hours, levels=c("1-3",
                                       "4-6",
                                       "7-9",
                                       "10 or more"))
```
```

First a function to order them

```{r}
ordered.likert = function(x) {
  x <- ordered(x, levels=c("Strongly Disagree",
                           "Disagree",
                           "Neutral",
                           "Agree",
                           "Strongly Agree"))}
```

Then defining the columns, and applying the function to them

```{r}
likert.items = c("relevant.to.aca.study.Why.enroll",
                 "skills.help.job.Why.enroll",
                 "credential.cv.Why.enroll",
                 "prestigious.uni.Why.enroll",
                 "course.fun.Why.enroll")

for (column in likert.items) {
  db[[column]] <- ordered.likert(db[[column]])}
```

223 Let's save it to a file, which can be used by other scripts for analysis/visualization
224 ```{r}
225 save(db, file="~/Dropbox/Laurie//Enc/db.Rdata")
226 ```

001_neuralnets
be101x
✓ coursera_001_aboriginaled
coursera_001_mentalhealth
coursera_001_programming1
coursera_001_programming2
coursera_intropsych_001
coursera_introstats_001
coursera_mentalhealth_002
coursera_programming1_002
la101x
mri2_ids
st_test
sta220
test

Structure   Content   Relations   Triggers   Table Info   Query

Search: field   ⌄   =   ⌄ ◦

forum_comments
forum_comments_mri
forum_forums
forum_posts
forum_posts_mri
forum_reporting
forum_reputation_points
forum_reputation_record
forum_subscribe_forums
forum_subscribe_threads
forum_tags
forum_tags_threads
forum_threads
hash_mapping
hash_mapping_mri
hg_assessment_calibrat...
hg_assessment_evaluati...
hg_assessment_metadata
hg_assessment_overall_...

coursera_001_aboriginaled
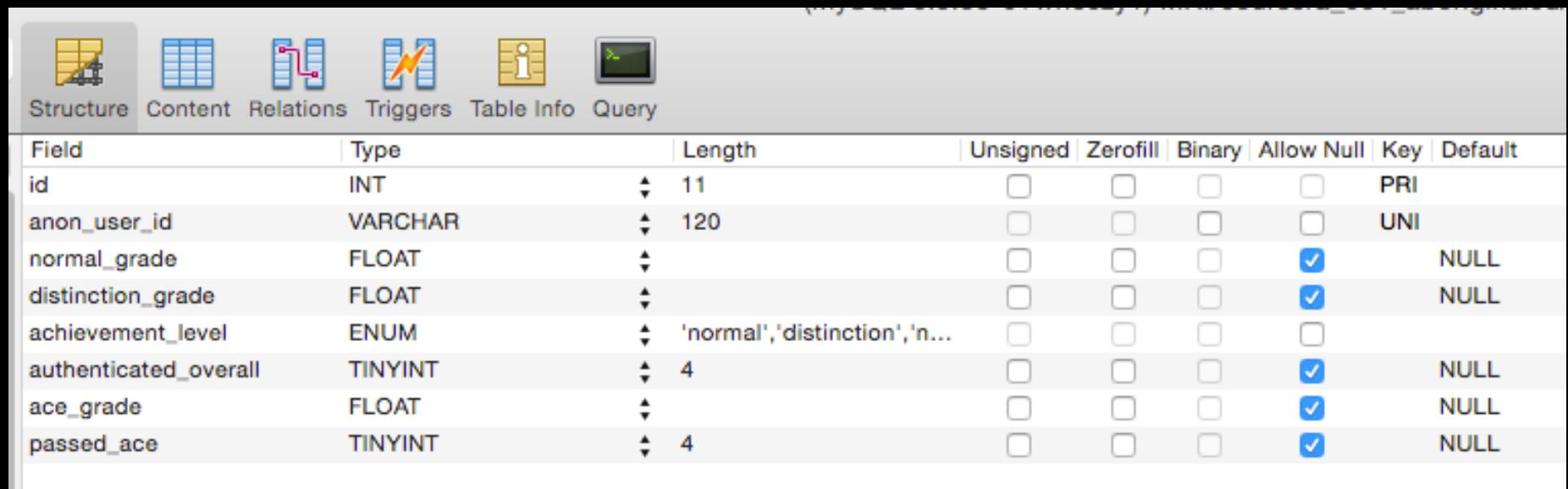
Select Database

Structure | Content | Relations | Triggers | Table Info | Query

Search: | id | = | | |

id	anon_user_id	normal_grade	distinction_grade	achievement_level	authenticated_overall	ace_grade	passed_ace
1	638a63da35c33b68ca63a0399	0	0	none	0	NULL	0
2	b31d0ae770b4906d000b1ffa5	0	0	none	0	NULL	0
3	37fcf2a2bda61eba88a5a81020	0	0	none	0	NULL	0
4	592355ca52d9ae649a768a16a	0	0	none	0	NULL	0
5	df6f970a455a1d54e0f5b91735	0	0	none	0	NULL	0
6	3f92461ba50d9329e09e297f21	0	0	none	0	NULL	0
7	990588572508dd9c4f3f54c59	0	0	none	0	NULL	0
8	694bbc2f3c0831df6a46e5f9f2	0	0	none	0	NULL	0
9	29bd5124fd21cec09c6dca31e7	0	0	none	0	NULL	0
10	b4d4ee2054b4becdeb700900c	0	0	none	0	NULL	0
11	61e6f50d288b3cfa7a1f04a501	0	0	none	0	NULL	0
12	b8e8edce457a8fe7d4f66248c0	0	0	none	0	NULL	0
13	2a4eeb4dbd40a90ddce89263d	0	0	none	0	NULL	0
14	42d98865e4baf1294850348e0	0	0	none	0	NULL	0
15	0f05a917520d716173cf61715	0	0	none	0	NULL	0
16	45e14a60ea939cfb319810b33	0	0	none	0	NULL	0
17	f90619f5a90444f20c2384a09a	0	0	none	0	NULL	0
18	791ac2d799e122ab228408eeb	0	0	none	0	NULL	0
19	a38b660d33f7f239f05458a0c4	0	0	none	0	NULL	0
20	bb73a809b210c78d846322130	98	98	distinction	0	NULL	0
21	04f01578d2a8482c858225163	0	0	none	0	NULL	0
22	38bf92dfd3cb23b12bb762e96	0	0	none	0	NULL	0
23	a514898ed54f118c216c99ec23	0	0	none	0	NULL	0
24	bb153460f174323ecdd212ade	0	0	none	0	NULL	0
25	944777e19df754aea5ace824de	0	0	none	0	NULL	0
26	23c812577d7c0a744ce7301a6	0	0	none	0	NULL	0
27	eb0ac3ded781f90c3121dc483	0	0	none	0	NULL	0
28	f21f33ecc52aa45caf4f3894af4	0	0	none	0	NULL	0
29	26c5ce475028132c3f7aec25a1	0	0	none	0	NULL	0
30	ecdf90f318e6905b3713551e18	0	0	none	0	NULL	0
31	40e942138431694b2b6a02490	0	0	none	0	NULL	0
32	a5aac3b54677c3328dd0ace3b1	0	0	none	0	NULL	0
33	a89c60c86e319995e534eabde	0	0	none	0	NULL	0
34	1e6b360f4602494f358a9bd9b	0	0	none	0	NULL	0
35	368034ec0d6fca3e4df05b9aa0	0	0	none	0	NULL	0
36	ebbe3128c7a7776af10d28e12b	0	0	none	0	NULL	0

TABLES

- access_groups
- activity_log
- announcements
- assignment_metadata
- assignment_part_metad...
- assignment_submission...
- course_grades
- forum_comments
- forum_comments_mri
- forum_forums
- forum_posts
- forum_posts_mri
- forum_reporting
- forum_reputation_points
- forum_reputation_record
- forum_subscribe_forums
- forum_subscribe_threads
- forum_tags
- forum_tags_threads
- forum_threads
- hash_mapping
- hash_mapping_mri
- hg_assessment_calibrat...
- hg_assessment_evaluati...
- hg_assessment_metadata
- hg_assessment_overall_...
- hg_assessment_peer_gr...
- hg_assessment_peer_gr...
- hg_assessment_self_gra...
- hg_assessment_submiss...
- hg_assessment_trainin...
- hg_assessment_training...
- items_sections
- kvs_course.260.assign...
- kvs_course.260.forum_r...
- kvs_course.260.internat...
- kvs_course.260.lecture

TABLE INFORMATION

- created: 2014-04-09
- engine: InnoDB
- rows: 20,996

Field	Type		Length	Unsigned	Zerofill	Binary	Allow Null	Key	Default
id	INT	⬍	11	☐	☐	☐	☐	PRI	
anon_user_id	VARCHAR	⬍	120	☐	☐	☐	☐	UNI	
normal_grade	FLOAT	⬍		☐	☐	☐	☑		NULL
distinction_grade	FLOAT	⬍		☐	☐	☐	☑		NULL
achievement_level	ENUM	⬍	'normal','distinction','n...	☐	☐	☐	☐		
authenticated_overall	TINYINT	⬍	4	☐	☐	☐	☑		NULL
ace_grade	FLOAT	⬍		☐	☐	☐	☑		NULL
passed_ace	TINYINT	⬍	4	☐	☐	☐	☑		NULL

```sql
select anon_user_id from course_grades where achievement_level = 'distinction';
```

TABLES
- access_groups
- activity_log
- announcements
- assignment_metadata
- assignment_part_metad...
- assignment_submission...
- course_grades
- forum_comments
- forum_comments_mri
- forum_forums
- forum_posts
- forum_posts_mri
- forum_reporting
- forum_reputation_points
- forum_reputation_record
- forum_subscribe_forums
- forum_subscribe_threads
- forum_tags
- forum_tags_threads
- forum_threads
- hash_mapping
- hash_mapping_mri
- hg_assessment_calibrat...
- hg_assessment_evaluati...
- hg_assessment_metadata
- hg_assessment_overall_...
- hg_assessment_peer_gr...
- hg_assessment_peer_gr...
- hg_assessment_self_gra...
- hg_assessment_submiss...
- hg_assessment_trainin...
- hg_assessment_training...

Query Favorites ⌄   Query History ⌄

anon_user_id

anon_user_id
bb73a809b210c78d84632213
b97fcfa146ecf57d7ba7b3c044
d448fd3cdc3e8e2a35ad90abe
75fb9879f380756592f8d633e
11cd16e6ee2b1de8f221165d8
7f0b5a5338c22941a94644930
a0c55351e1a0c2e6f1a179ab2
87273aa9cf9b85849d850a582
a3b8294e29a750103dec9414
88ef472c648940b9a197f16ca
81aac4f1f2b26d02706a99726
d23d5532f4bb8b2060d90e01

coursera_001_aboriginaled

Select Database

Struc Switch to the Table Structure tab Table Info Query

Search: id ⟋ = ⟋ ●

TABLES

- access_groups
- activity_log
- announcements
- assignment_metadata
- assignment_part_metad...
- assignment_submission...
- course_grades
- **forum_comments**
- forum_comments_mri
- forum_forums
- forum_posts
- forum_posts_mri
- forum_reporting
- forum_reputation_points
- forum_reputation_record
- forum_subscribe_forums
- forum_subscribe_threads
- forum_tags
- forum_tags_threads
- forum_threads
- hash_mapping
- hash_mapping_mri
- hg_assessment_calibrat...
- hg_assessment_evaluati...
- hg_assessment_metadata
- hg_assessment_overall_...
- hg_assessment_peer_gr...
- hg_assessment_peer_gr...
- hg_assessment_self_gra...

id	thread_id	post_id	forum_user_id	post_time	deleted	is_spam	votes	anonymous
1	13	16	b2fe3153c533760c	1361571936	0	0	4	0
2	13	17	b2fe3153c533760c	1361572380	0	0	10	0
3	13	21	b2fe3153c533760c	1361573409	0	0	0	0
4	13	16	296e97161c5207de	1361574099	0	0	4	0
5	18	37	b4d2b61d902761cl	1361575360	0	0	1	0
6	26	49	91e6aedaef025666	1361576015	0	0	0	0
7	17	39	e38700b2dbb358f8	1361576225	1	0	0	1
8	17	54	91e6aedaef025666	1361576398	0	0	1	0
9	26	73	eef1d3338772253a	1361577087	0	0	2	0
10	28	95	fbf83ee2ab23282c!	1361580230	1	0	0	0
11	13	16	543d2170884b943	1361580293	0	0	10	0
12	36	150	6fd1796d9ac7d056	1361580588	1	0	0	0
13	28	95	63eb3f17a22e4d80	1361580913	0	0	0	0
14	18	62	63eb3f17a22e4d80	1361581007	0	0	0	0
15	13	50	6f3e16917f72ee04:	1361581168	0	0	0	0
16	24	166	d1459ece8d602d48	1361581186	0	0	0	0
17	13	50	f80c83c27fadcc789	1361581821	0	0	0	0
18	11	84	0f59feb06378d53b	1361581834	0	0	0	0
19	13	31	d42231ce5a5ee26c	1361582291	0	0	0	0
20	48	163	d5b963aefd96d7fa	1361582607	0	0	0	0
21	72	174	40e9da9796633be!	1361582715	0	0	2	1
22	25	187	56f11947921f1c38	1361582822	0	0	4	0
23	25	187	d5b963aefd96d7fa	1361583105	0	0	4	0
24	57	145	28d1ab747993fac0	1361583765	0	0	7	0
25	17	64	28d1ab747993fac0	1361584075	0	0	0	0

coursera_001_aboriginaled

Structure | Content | Relations | Triggers | Table Info | Query

Table History | Users | Console

Search: user_id | = | Search | Filter

TABLES

- access_groups
- activity_log
- announcements
- assignment_metadata
- assignment_part_metad...
- assignment_submission_...
- course_grades
- forum_comments
- forum_comments_mri
- forum_forums
- forum_posts
- forum_posts_mri
- forum_reporting
- forum_reputation_points
- forum_reputation_record
- forum_subscribe_forums
- forum_subscribe_threads
- forum_tags
- forum_tags_threads
- forum_threads
- **hash_mapping**
- hash_mapping_mri
- hg_assessment_calibrat...
- hg_assessment_evaluati...
- hg_assessment_metadata
- hg_assessment_overall_...
- hg_assessment_peer_gr...
- hg_assessment_peer_gr...
- hg_assessment_self_gra...
- hg_assessment_submiss...
- hg_assessment_trainin...
- hg_assessment_training_...
- items_sections
- kvs_course.260.assign...
- kvs_course.260.forum_r...

user_id	anon_user_id	forum_user_id	session_user_id	eventing_user_id
326563	0000ea1442390d07c	2e460c588e71283778	07b3e4039c83818bb	f972f62bfc225e4fcfac
780229	0001153485f3e61ad	20d217903b96a5d91	bc74b422299bf56a61	3ed6c54ecfacc49fd0e
571543	0001920e8cc2cd5c4	5c24af67cd40be3d29	dee8060c00110ef073	08888228f98c5c8affc
1690899	000344511929f9408	2991204c6d6bf5babc	db82b229a51a63433	54b8ee94e13fb55ab45
1021214	00049b6fab92327d5	1e2aac08127c73701b	d980d89d218dde412f	231b63764f9d27c11b2
1243882	0009bb3df7bd0ee77	bad64123fc8b0e24ef	d3c3bf59aa9bbc22c5	8294eba2c1519633e3
1134088	000a965ec5309bfc1	5c6b569f440f634fc46	5e550196092e5ac73	ecf9b70cd91026a8a14
1285509	000e32e13b8bc984a	1eee5a30aa7f210d35	430e7463bd1b959ffe	1503c46cb01bd3612c
489253	000f284d8d1140852	a8e040b624b337b36	38887837a7be3e1ed2	e3d923c3a9c5eca4052
2619535	0012347154b8a1a7f	6b8766ec7305b86c0	ad1a6fb88ca2ff7c7c2	856dd0998e2d4cc312
4681718	00144c137d6ee893	bb8f2eb2a60eae9be6	f3d661745eab920b89	070a0c79f80ccf81e8d
1801129	0015fe942e13673fa	cf7f8a806bb4509ae5	1f78faf12d4fbeb4afe0	1210569aa0b06828d
1207767	001bbca357494642	589de67bde47b06d9	b89dbf1813e28f7c67	b35c4be3be2e386844
4762934	00207fe14c7170d9b	c297f13a3776e915fc	5cf7aa71cf8e23dde0e	1926194fab45905375
4254631	00210c1ec52622959	e91e3a6f143fcdd1f2	d24d151a01d68a1489	88658eb238fe010da2
4912799	00230da1041ceba7	fd2a807b84a888c88	3b274fe357f247e7c1	dad5d64376f7bc09702
2234579	002398c9d54e5a5d	2e22a3446e0e1ece21	97b64c65e2c1f9bb8e	affe18b7b2794a6594c
1508329	00271579cf2c092a3	bddc761f4612d047d2	81b70b1032e1e01b89	9401f665c5a7632e45
577687	002adb0740c74e627	60499e60d697bcd247	aca54b6c48ac5cf834b	c5e25f76cd21f1729d8
270409	002b1427e2ab708f	10dd8305bf451de9b	cb17b5c83ff4c9463c	e9a9ef69c75992f53e3
404979	002c55cfe5913ed17	4402227134a8a5f2eb	b561ed5cff4fdb6e04b	2c8946768eb585d37f
1748098	00311ba76a7d43c1	8c03fc4fa7b4ef2864f	5d60573e5afb5d73fd	4cef4f66f47c657de6e
39735	0033b167f45f9312a0	6f97253fd9c9ee9cca0	23d063b6f16962602	d704b237d104d87d4
4175922	003695cc8f9f1153a6	8b42bc3cd5ec739d2	2d255759413821fcf5	00f5ab1f4f0543829a9
2238334	003db3a9532a1b61a	eb80aed6deb1eb0966	26b89e206383684bb	7f75cc0fc26df8e9bb0
477847	003e207127e4ff771	1b983d886288f7448f	cec3d86415bbd5c6ee	13982a81b4b4d9acdb
2457730	003e283a136771ad	c9aceb6b5308792f66	de6bee5fdc9c483896	c485bcf9fda82a83c24
1031809	003e6344b6cc361d9	acbd47e2629ef9928f	47403fe52a55a2a6ae	76830c5664e3a992b3
941108	00401c701a13907d	a69f1299386e44b6ab	68890515ad7f314d6e	82f31be340d63a26d6
676655	0040fadf69fbeed183	0979bf80c8ffd6f0e53	cfe1936ce2f18939d5	bfd5dc5f5e98a60c78b

coursera_001_aboriginaled

Select Database

Structure | Content | Relations | Triggers | Table Info | Query

Q Filter

TABLES

- access_groups
- activity_log
- announcements
- assignment_metadata
- assignment_part_metad...
- assignment_submission...
- course_grades
- forum_comments
- forum_comments_mri
- forum_forums
- forum_posts
- forum_posts_mri
- forum_reporting
- forum_reputation_points
- forum_reputation_record
- forum_subscribe_forums
- forum_subscribe_threads
- forum_tags
- forum_tags_threads
- forum_threads
- hash_mapping
- hash_mapping_mri
- hg_assessment_calibrat...
- hg_assessment_evaluati...
- hg_assessment_metadata
- hg_assessment_overall_...
- hg_assessment_peer_gr...
- hg_assessment_peer_gr...
- hg_assessment_self_gra...
- hg_assessment_submiss...
- hg_assessment_trainin...
- hg_assessment_training...
- items_sections
- kvs_course.260.assign...
- kvs_course.260.forum_r...
- kvs_course.260.internat...
- kvs_course.260.lecture

```sql
1  select
2    C.anon_user_id, count(F.id)
3  from
4    course_grades as C,
5    hash_mapping as H,
6    forum_posts as F
7  where
8    C.achievement_level = 'normal' and
9    C.anon_user_id = H.anon_user_id and
10   H.forum_user_id = F.forum_user_id
11 group by
12   C.anon_user_id;
```

Query Favorites ⌄ | Query History ⌄

anon_user_id	count(F.id)
002adb0740c74e627c6d30e08cf5.	17
00488ce38edcf2e6276c62e7c340(	11
00683a578890073a7baf0dc0914d	17
0088bf394b1ef1a1b6591b531a7cl	7
00b546bb1e08470c20ce19eabbe4	7
00df83da715f894c0f4e3adfb8cf6c	7
0107cc20fcdcb9efe9a9d65cb1fa69	11
016375c8cbd45e1c97b0eee1fd730	15
01a0e132ce9ba2069aa2594ac64a(	23
01c6e685b9bd2fe32447ec821653	4
01f07dd251192050d6bb39ee7bb(	11
02270fe5e3c0bc3e7b72e1a71ca35	4
027150871f73b5d3bc7fe00b12b0	5
027b7b080d8905656e31c5ef616e	6
028a8a83608dda4990747bb729a4	9
0290ed52a0590490de366dbc48ab	5

coursera_001_aboriginaled

Select Database

Structure Content Relations Triggers Table Info Query

Q Filter

TABLES

access_groups
activity_log
announcements
assignment_metadata
assignment_part_metad...
assignment_submission...
course_grades
forum_comments
forum_comments_mri
forum_forums
forum_posts
forum_posts_mri
forum_reporting
forum_reputation_points
forum_reputation_record
forum_subscribe_forums
forum_subscribe_threads
forum_tags
forum_tags_threads
forum_threads
hash_mapping
hash_mapping_mri
hg_assessment_calibrat...

```sql
1  select
2    C.achievement_level, count(distinct F.id)/count(distinct C.id)
3  from
4    course_grades as C,
5    hash_mapping as H,
6    forum_posts as F
7  where
8    C.anon_user_id = H.anon_user_id and
9    H.forum_user_id = F.forum_user_id
10 group by
11   C.achievement_level;
```

⚙ ▾   Query Favorites ⌄   Query History ⌄

achievement_level	count(distinct F.id)/count(distinct C.id)
normal	10.7608
distinction	17.8492
none	4.2557

stats.Rmd ✕ | load-survey.Rmd* ✕ | Load-surveys.Rmd ✕ | Dε »

Knit HTML ▾ | Run | Chunks ▾
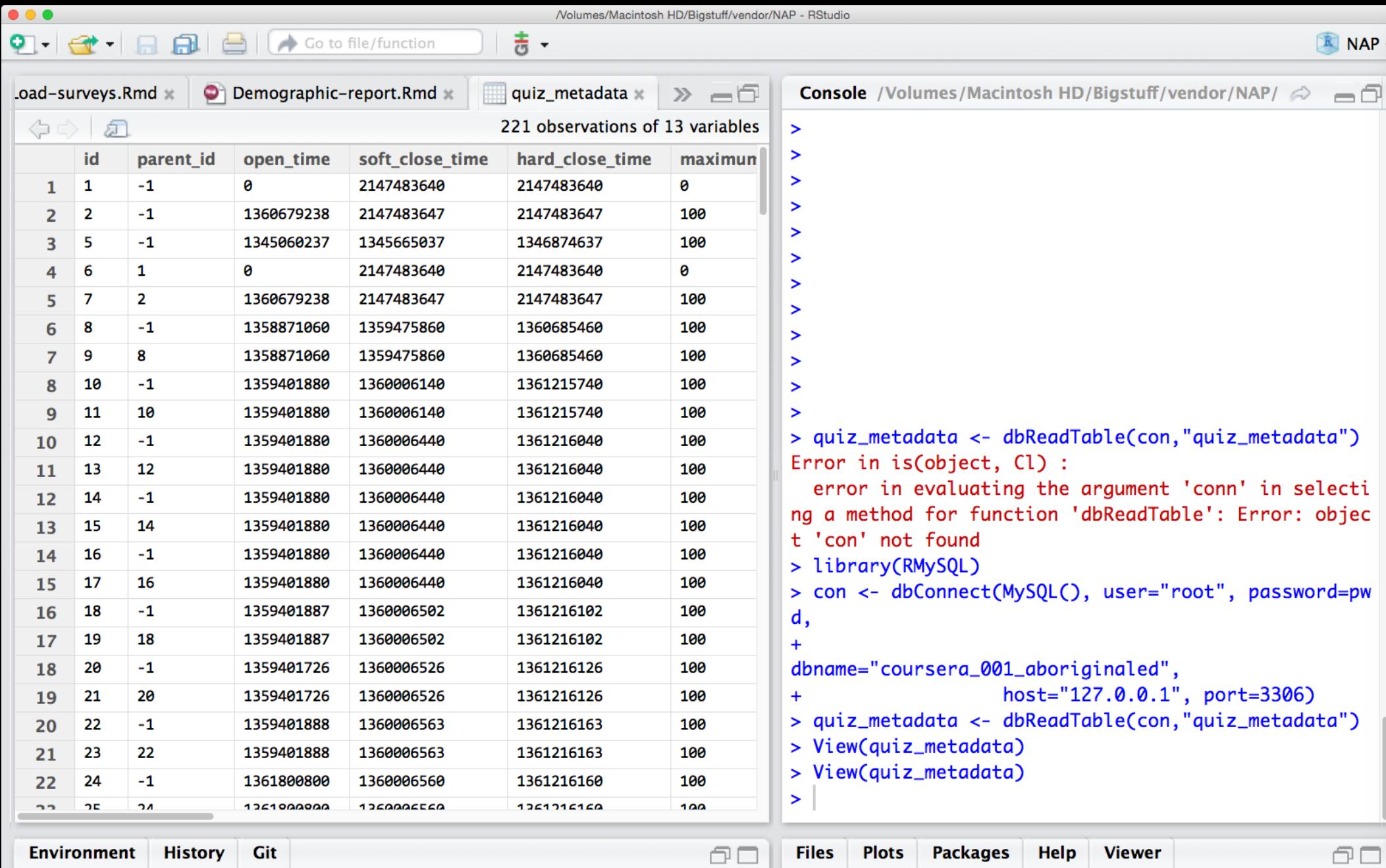
```
1   Connect to database
2 ▾ ```{r}
3   library(RMySQL)
4   con <- dbConnect(MySQL(), user="root", password=pwd,
5                    dbname="coursera_001_aboriginaled",
6                    host="127.0.0.1", port=3306)
7 ▲ ```
8
9
10  Get a full table
11 ▾ ```{r}
12  quiz_metadata <- dbReadTable(con,"quiz_metadata")
13 ▲ ```
```

13:1   C Chunk 2 ⇕   R Markdown ⇕

Console /Volumes/Macintosh HD/Bigstuff/vendor/NAP/

```
>
>
>
>
>
>
>
>
>
>
>
>
>
>
> quiz_metadata <- dbReadTable(con,"quiz_metadata")
Error in is(object, Cl) :
  error in evaluating the argument 'conn' in selecti
ng a method for function 'dbReadTable': Error: objec
t 'con' not found
> library(RMySQL)
> con <- dbConnect(MySQL(), user="root", password=pw
d,
+
dbname="coursera_001_aboriginaled",
+                  host="127.0.0.1", port=3306)
> quiz_metadata <- dbReadTable(con,"quiz_metadata")
>
```

Environment | History | Git

Files | Plots | Packages | Help | Viewer

/Volumes/Macintosh HD/Bigstuff/vendor/NAP - RStudio

NAP

Go to file/function

Load-surveys.Rmd ×   Demographic-report.Rmd ×   quiz_metadata ×   »

221 observations of 13 variables

	id	parent_id	open_time	soft_close_time	hard_close_time	maximun
1	1	-1	0	2147483640	2147483640	0
2	2	-1	1360679238	2147483647	2147483647	100
3	5	-1	1345060237	1345665037	1346874637	100
4	6	1	0	2147483640	2147483640	0
5	7	2	1360679238	2147483647	2147483647	100
6	8	-1	1358871060	1359475860	1360685460	100
7	9	8	1358871060	1359475860	1360685460	100
8	10	-1	1359401880	1360006140	1361215740	100
9	11	10	1359401880	1360006140	1361215740	100
10	12	-1	1359401880	1360006440	1361216040	100
11	13	12	1359401880	1360006440	1361216040	100
12	14	-1	1359401880	1360006440	1361216040	100
13	15	14	1359401880	1360006440	1361216040	100
14	16	-1	1359401880	1360006440	1361216040	100
15	17	16	1359401880	1360006440	1361216040	100
16	18	-1	1359401887	1360006502	1361216102	100
17	19	18	1359401887	1360006502	1361216102	100
18	20	-1	1359401726	1360006526	1361216126	100
19	21	20	1359401726	1360006526	1361216126	100
20	22	-1	1359401888	1360006563	1361216163	100
21	23	22	1359401888	1360006563	1361216163	100
22	24	-1	1361800800	1360006560	1361216160	100
23	25	24	1361800800	1360006560	1361216160	100

Console /Volumes/Macintosh HD/Bigstuff/vendor/NAP/

```
>
>
>
>
>
>
>
>
>
>
>
>
> quiz_metadata <- dbReadTable(con,"quiz_metadata")
Error in is(object, Cl) :
  error in evaluating the argument 'conn' in selecti
ng a method for function 'dbReadTable': Error: objec
t 'con' not found
> library(RMySQL)
> con <- dbConnect(MySQL(), user="root", password=pw
d,
+
dbname="coursera_001_aboriginaled",
+                 host="127.0.0.1", port=3306)
> quiz_metadata <- dbReadTable(con,"quiz_metadata")
> View(quiz_metadata)
> View(quiz_metadata)
>
```

Environment   History   Git

Files   Plots   Packages   Help   Viewer

Knit HTML ▾   Run   Chunks ▾

```
 6                     host="127.0.0.1", port=3306)
 7  ```
 8
 9
10  Get a full table
11  ```{r}
12  quiz_metadata <- dbReadTable(con,"quiz_metadata")
13  ```
14
15  Run a query
16  ```{r}
17  posts <- dbSendQuery(con, "
18  select
19    C.achievement_level, count(distinct F.id)/count(distinct
    C.id)
20  from
21    course_grades as C,
22    hash_mapping as H,
23    forum_posts as F
24  where
25    C.anon_user_id = H.anon_user_id and
26    H.forum_user_id = F.forum_user_id
27  group by
28    C.achievement_level;")
29  ```
```

17:1   Chunk 3 ⇕   R Markdown ⇕

Environment   History   Git

Console /Volumes/Macintosh HD/Bigstuff/vendor/NAP/

```
t 'con' not found
> library(RMySQL)
> con <- dbConnect(MySQL(), user="root", password=pw
d,
+
dbname="coursera_001_aboriginaled",
+                host="127.0.0.1", port=3306)
> quiz_metadata <- dbReadTable(con,"quiz_metadata")
> View(quiz_metadata)
> View(quiz_metadata)
> posts <- dbSendQuery(con, "
+ select
+    C.achievement_level, count(distinct F.id)/count(
distinct C.id)
+ from
+      course_grades as C,
+      hash_mapping as H,
+      forum_posts as F
+ where
+      C.anon_user_id = H.anon_user_id and
+      H.forum_user_id = F.forum_user_id
+ group by
+      C.achievement_level;")
Warning message:
In .local(conn, statement, ...) :
  Decimal MySQL column 1 imported as numeric
>
```

Files   Plots   Packages   Help   Viewer

Editor Panel
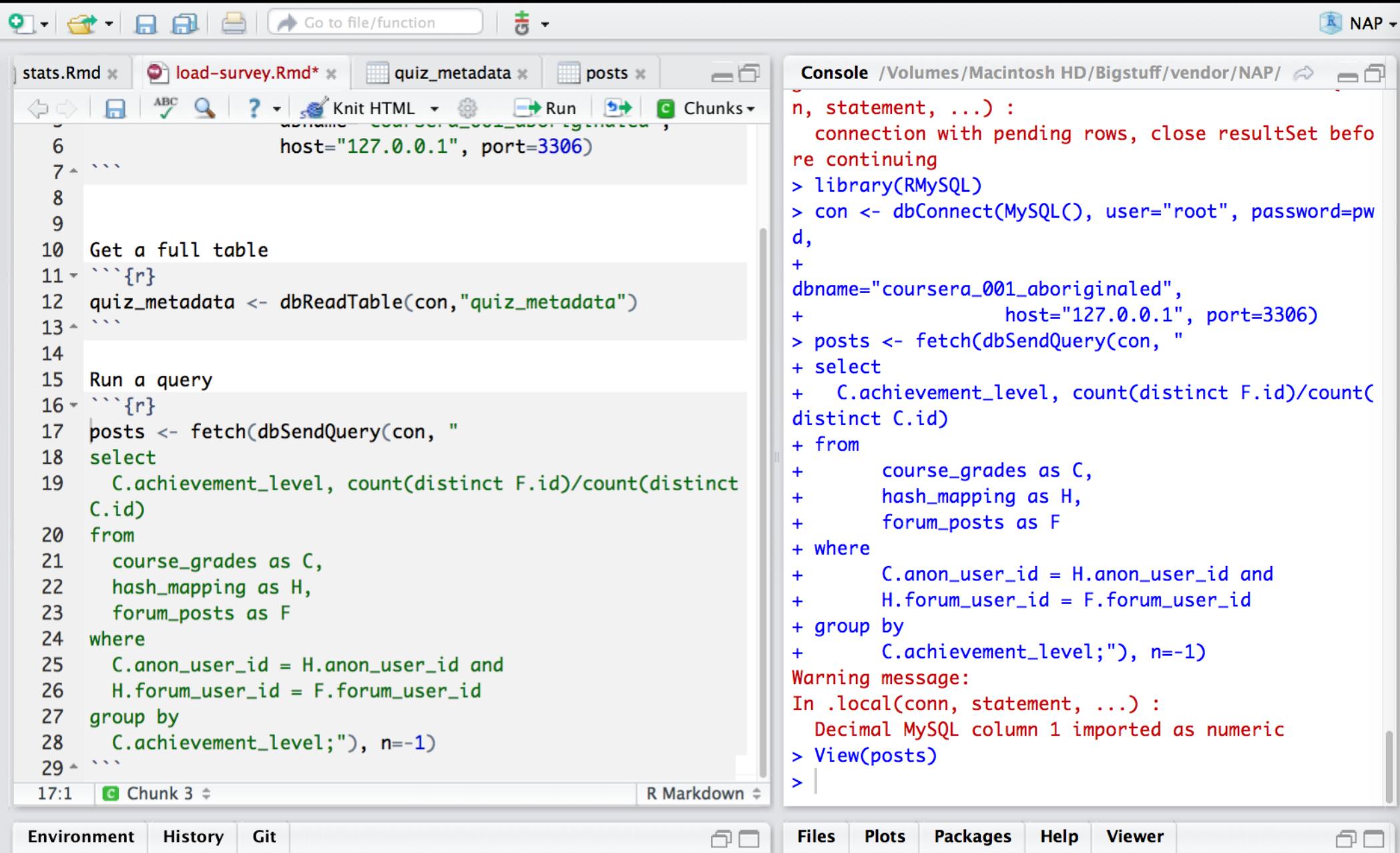
Tabs: md | Demographic-report.Rmd | quiz_metadata | posts | »

3 observations of 2 variables

	achievement_level	count(distinct F.id)/count(distinct C.id)
1	normal	10.7608
2	distinction	17.8492
3	none	4.2557

Environment | History | Git

Console /Volumes/Macintosh HD/Bigstuff/vendor/NAP/

```
n, statement, ...) :
  connection with pending rows, close resultSet befo
re continuing
> library(RMySQL)
> con <- dbConnect(MySQL(), user="root", password=pw
d,
+
dbname="coursera_001_aboriginaled",
+                  host="127.0.0.1", port=3306)
> posts <- fetch(dbSendQuery(con, "
+ select
+    C.achievement_level, count(distinct F.id)/count(
distinct C.id)
+ from
+      course_grades as C,
+      hash_mapping as H,
+      forum_posts as F
+ where
+      C.anon_user_id = H.anon_user_id and
+      H.forum_user_id = F.forum_user_id
+ group by
+      C.achievement_level;"), n=-1)
Warning message:
In .local(conn, statement, ...) :
  Decimal MySQL column 1 imported as numeric
> View(posts)
>
```

Files | Plots | Packages | Help | Viewer

~/Dropbox/Laurie/Enc/Load-surve

Editor pane (load-survey.Rmd):

```
 6                host="127.0.0.1", port=3306)
 7   ```
 8
 9
10   Get a full table
11   ```{r}
12   quiz_metadata <- dbReadTable(con,"quiz_metadata")
13   ```
14
15   Run a query
16   ```{r}
17   posts <- fetch(dbSendQuery(con, "
18   select
19     C.achievement_level, count(distinct F.id)/count(distinct
     C.id)
20   from
21     course_grades as C,
22     hash_mapping as H,
23     forum_posts as F
24   where
25     C.anon_user_id = H.anon_user_id and
26     H.forum_user_id = F.forum_user_id
27   group by
28     C.achievement_level;"), n=-1)
29   ```
```

`17:1   © Chunk 3 ↕`   `R Markdown ↕`

Console /Volumes/Macintosh HD/Bigstuff/vendor/NAP/

```
n, statement, ...) :
  connection with pending rows, close resultSet befo
re continuing
> library(RMySQL)
> con <- dbConnect(MySQL(), user="root", password=pw
d,
+
dbname="coursera_001_aboriginaled",
+               host="127.0.0.1", port=3306)
> posts <- fetch(dbSendQuery(con, "
+ select
+   C.achievement_level, count(distinct F.id)/count(
distinct C.id)
+ from
+       course_grades as C,
+       hash_mapping as H,
+       forum_posts as F
+ where
+       C.anon_user_id = H.anon_user_id and
+       H.forum_user_id = F.forum_user_id
+ group by
+       C.achievement_level;"), n=-1)
Warning message:
In .local(conn, statement, ...) :
  Decimal MySQL column 1 imported as numeric
> View(posts)
>
```

Environment   History   Git

Files   Plots   Packages   Help   Viewer

```r
5        con <- dbConnect(MySQL(), user="root", password=pwd,
                    dbname="coursera_001_aboriginaled",
6                    host="127.0.0.1", port=3306)
7   ```
8
9
10  Get a full table
11  ```{r}
12  quiz_metadata <- dbReadTable(con,"quiz_metadata")
13  ```
14
15  Run a query
16  ```{r}
17  posts_per_user <- fetch(dbSendQuery(con, "
18  select
19    C.achievement_level, count(distinct F.id)
20  from
21    course_grades as C,
22    hash_mapping as H,
23    forum_posts as F
24  where
25    C.anon_user_id = H.anon_user_id and
26    H.forum_user_id = F.forum_user_id
27  group by
28    C.anon_user_id;"), n=-1)
29  ```
```

17:1   © Chunk 3 ⬍   R Markdown ⬍

Environment   History   Git

Console /Volumes/Macintosh HD/Bigstuff/vendor/NAP/

```r
distinct C.id)
+ from
+       course_grades as C,
+       hash_mapping as H,
+       forum_posts as F
+ where
+       C.anon_user_id = H.anon_user_id and
+       H.forum_user_id = F.forum_user_id
+ group by
+       C.achievement_level;"), n=-1)
Warning message:
In .local(conn, statement, ...) :
  Decimal MySQL column 1 imported as numeric
> View(posts)
> posts_per_user <- fetch(dbSendQuery(con, "
+ select
+   C.achievement_level, count(distinct F.id)
+ from
+       course_grades as C,
+       hash_mapping as H,
+       forum_posts as F
+ where
+       C.anon_user_id = H.anon_user_id and
+       H.forum_user_id = F.forum_user_id
+ group by
+       C.anon_user_id;"), n=-1)
>
```

Files   Plots   Packages   Help   Viewer

/Volumes/Macintosh HD/Bigstuff/vendor/NAP - RStudio

~survey.Rmd* ✕ | quiz_metadata ✕ | posts ✕ | posts_per_user ✕ | » ⊟ ☐

4109 observations of 2 variables

	achievement_level	count(distinct F.id)
1	normal	17
2	distinction	4
3	normal	11
4	distinction	86
5	normal	17
6	none	2
7	normal	7
8	normal	7
9	distinction	9
10	none	8
11	distinction	10
12	normal	7
13	normal	11
14	distinction	27

Displayed 1000 rows of 4109 (3109 omitted)

Environment | History | Git

📂 💾 | Import Dataset ▾ | 🧹 Clear | ⟳   ≡ List ▾

Global Environment ▾

Data

● posts   3 obs. of 2 variables
● posts_per_user   4109 obs. of 2 variables
● quiz_metadata   221 obs. of 13 variables
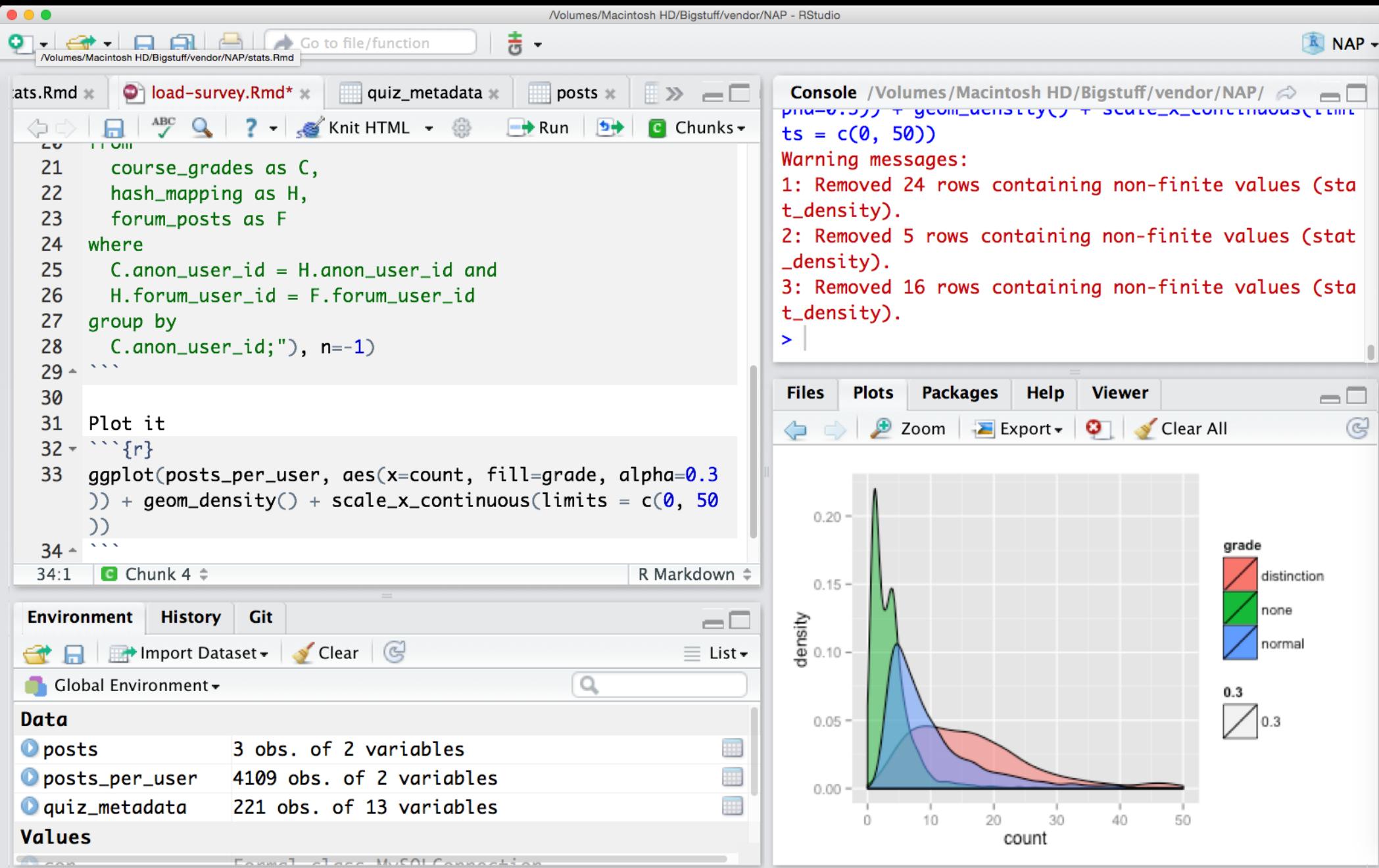
Values
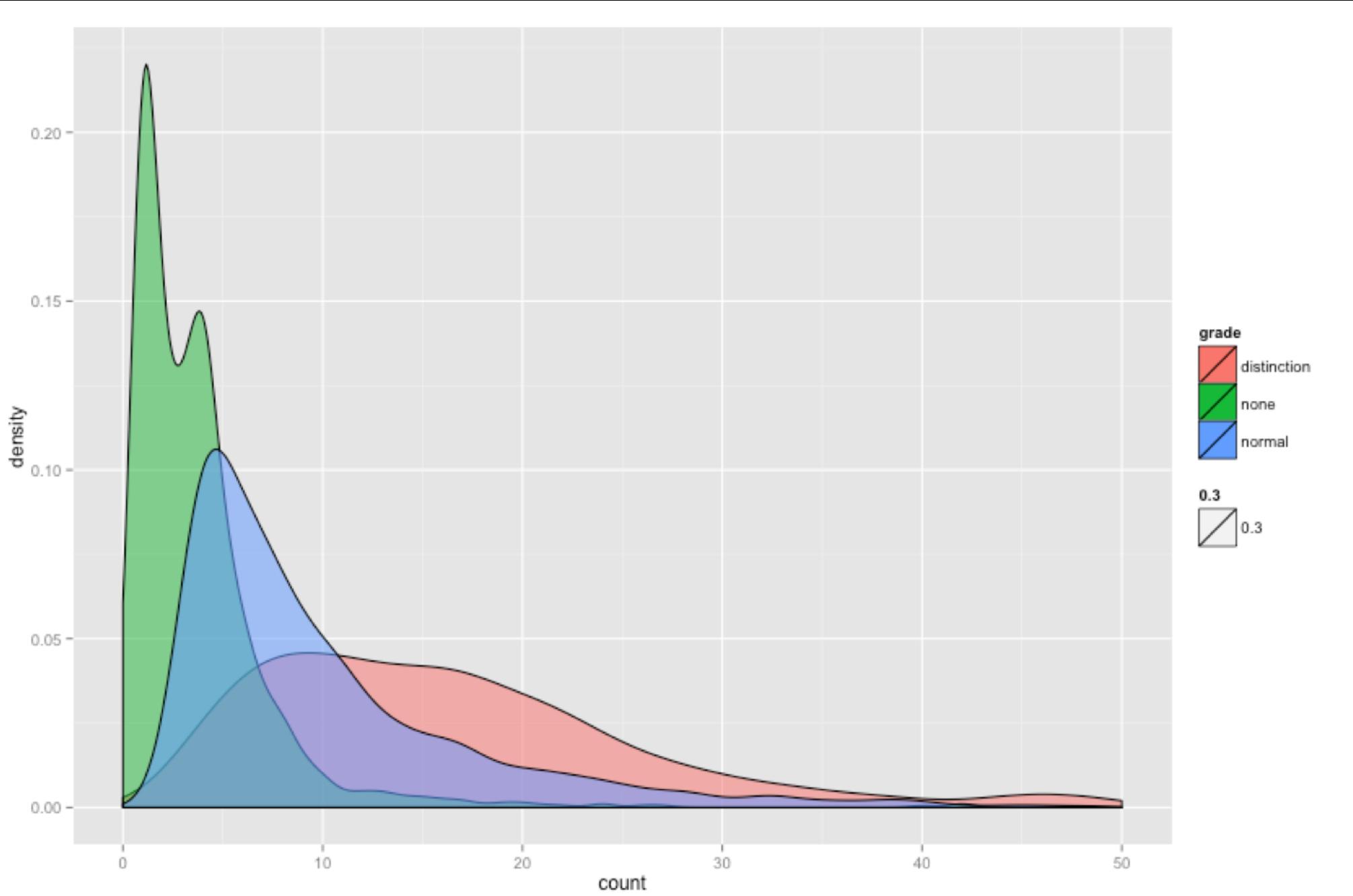
Console /Volumes/Macintosh HD/Bigstuff/vendor/NAP/ ⤤ ⊟ ☐

```
+ from
+     course_grades as C,
+     hash_mapping as H,
+     forum_posts as F
+ where
+     C.anon_user_id = H.anon_user_id and
+     H.forum_user_id = F.forum_user_id
+ group by
+     C.achievement_level;"), n=-1)
Warning message:
In .local(conn, statement, ...) :
  Decimal MySQL column 1 imported as numeric
> View(posts)
> posts_per_user <- fetch(dbSendQuery(con, "
+ select
+   C.achievement_level, count(distinct F.id)
+ from
+     course_grades as C,
+     hash_mapping as H,
+     forum_posts as F
+ where
+     C.anon_user_id = H.anon_user_id and
+     H.forum_user_id = F.forum_user_id
+ group by
+     C.anon_user_id;"), n=-1)
> View(posts_per_user)
>
```

Files | Plots | Packages | Help | Viewer

/Volumes/Macintosh HD/Bigstuff/vendor/NAP/stats.Rmd

Go to file/function

NAP

ats.Rmd × | ⬤ **load-survey.Rmd*** × | **quiz_metadata** × | **posts** × | ≫

Knit HTML ▾ | ⚙ | → Run | ↩ | **C** Chunks ▾

```
20   From
21     course_grades as C,
22     hash_mapping as H,
23     forum_posts as F
24   where
25     C.anon_user_id = H.anon_user_id and
26     H.forum_user_id = F.forum_user_id
27   group by
28     C.anon_user_id;"), n=-1)
29   ```
30
31   Plot it
32   ```{r}
33   ggplot(posts_per_user, aes(x=count, fill=grade, alpha=0.3
       )) + geom_density() + scale_x_continuous(limits = c(0, 50
       ))
34   ```
```

34:1 **C** Chunk 4 ⇕   R Markdown ⇕

Environment | **History** | **Git**

📂 💾 | Import Dataset ▾ | Clear | ⟳   ≡ List ▾

Global Environment ▾   🔍

Data
⬤ posts	3 obs. of 2 variables	▦
⬤ posts_per_user	4109 obs. of 2 variables	▦
⬤ quiz_metadata	221 obs. of 13 variables	▦

Values

Console /Volumes/Macintosh HD/Bigstuff/vendor/NAP/ ↗

```
pha=0.3)) + geom_density() + scale_x_continuous(limi
ts = c(0, 50))
Warning messages:
1: Removed 24 rows containing non-finite values (sta
t_density).
2: Removed 5 rows containing non-finite values (stat
_density).
3: Removed 16 rows containing non-finite values (sta
t_density).
>
```

Files | **Plots** | **Packages** | **Help** | **Viewer**

⬅ ➡ | 🔍 Zoom | 🖼 Export ▾ | ⊗ | 🧹 Clear All   ⟳

localhost:8888/notebooks/demo.ipynb

Apps ★ Bookmarks Food log 2015 - Goo wiki 22 GC Read TStar Spend hdiff #DALMOOC | Data, hdiff HTMLPreview ->Ipad pinb MyA

IP[y]: Notebook demo (autosaved)

File   Edit   View   Insert   Cell   Kernel   Help

Code ▾   Cell Toolbar: None ▾

```
In [201]: useractivity = usrlecture.reset_index().join(grp_u_thread, how='outer', on='username')
```

```
In [205]: useractivity.fillna(0, inplace=True)
          useractivity.head()
```
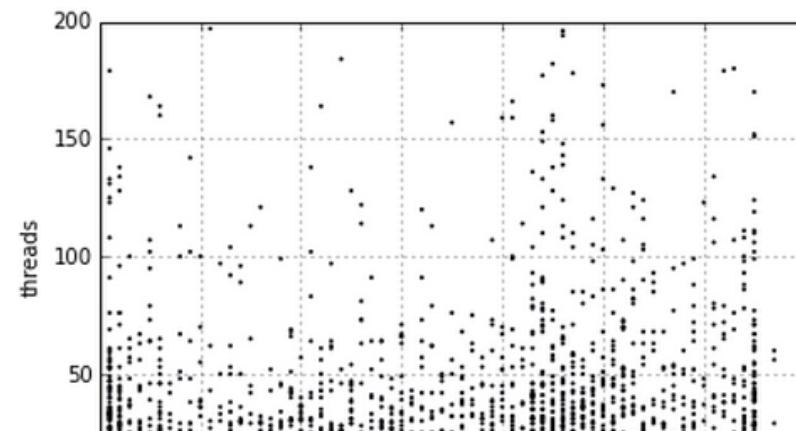
Out[205]:

	username	videos	threads
0	1	10	0
1	2	32	7
2	3	41	1
3	4	14	0
4	5	2	0

5 rows × 3 columns

```
In [219]: useractivity.plot(x='videos', y='threads',kind="scatter", ylim=(0,200), xlim=(0, 70), s=1)
```

Out[219]: <matplotlib.axes.AxesSubplot at 0x247029690>

Clicklog: big data, making it queryable, increasing levels of abstraction

re/view?lecture_id=19","client": "spark","session": "4931011420-1372274741397","language": "en-US,en;q=0.5","from": "https://class.cou
rsera.org/mentalhealth-002/lecture/view?lecture_id=19","user_ip": "67.244.21.168","user_agent": "Mozilla/5.0 (Windows NT 6.2; WOW64; r
v:21.0) Gecko/20100101 Firefox/21.0", "12": ["{\"height\":768,\"width\":1366}"], "13": [0]}
"es-es,es;q=0.8,en-us,en;q=0.5,en;q=0.3","from": "https://class.coursera.org/mentalhealth-002/survey/attempt?quiz_id=146","user_ip": "95.
16.78.198","user_agent": "Mozilla/5.0 (X11; Ubuntu; Linux i686; rv:15.0) Gecko/20100101 Firefox/15.0.1", "12": ["{\"height\":768,\"wid
th\":1366}"], "13": [0]}
"username": "e7e8d3e36244d7f57ffb4951b1bf66a95dda259","timestamp": 1372629807108,"page_url": "https://class.coursera.
org/mentalhealth-002/lecture/view?lecture_id=13","client": "spark","session": "6487801913-1372611846611","language": "en-us","from": "
https://class.coursera.org/mentalhealth-002/lecture/view?lecture_id=13","user_ip": "75.140.96.144","user_agent": "Mozilla/5.0 (iPad; C
PU OS 6_1_3 like Mac OS X)

	action			duration	action_val	lecture_id_val	type_val		quiz_id	session	su
			timestamp								
0	1	N	2013-06-25 19:17:14.594000	16118	survey/attempt	NaN	NaN	N	1	Na	
1	1										

37	71.000	lecture/view	out-of-sequence, seeked, pause
38	31.000	lecture/view	out-of-sequence
39	33.000	lecture/view	out-of-sequence, pause, short-event
40	31.000	lecture/view	seen-before, seeked, short-event
41	33.000	lecture/view	seen-before, short-event
42	31.000	lecture/view	seen-before, seeked, pause
43	33.000	lecture/view	seen-before, pause, short-event
44	35.000	lecture/view	out-of-sequence, pause, short-event
45	33.000	lecture/view	seen-before, seeked, short-event
46	35.000	lecture/view	seen-before, short-event
47	33.000	lecture/view	seen-before, seeked, pause

[1 rows x 4 columns]

		duration	action_val	lecture_id_val	type_val			
timestamp								
2013-06-29 19:51:48.975000		169202	human_grading/	NaN	NaN	N	9	Na
						N	7	Na
						N	10	Na

"7814042312-1372629831884","language": "en-US,en;q=0.8","from": "http
s://www.coursera.org/","user_ip": "24.152.229.234","user_agent": "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like
Gecko) Chrome/27.0.1453.116 Safari/537.36", "12": ["{\"height\":768,\"width\":1366}"], "13": [0]}
{"key": "user.video.lecture.action", "value": "{\"currentTime\":357.7353528,\"playbackRate\":1,\"paused\":true,\"error\":null,\"network
State\":1,\"readyState\":4,\"eventTimestamp\":1372629834044,\"initTimestamp\":1372626984835,\"type\":\"pause\",\"prevTime\":357.508206
9}","username": "0967ed3264df28107090cd8ad80b176e3ba5a5e7","timestamp": 1372629842192,"page_url": "https://class.coursera.org/mentalhe
alth-002/lecture/view?lecture_id=5","client": "spark","session": "1605085030-1372100726269","language": "en-US","from": "https://class