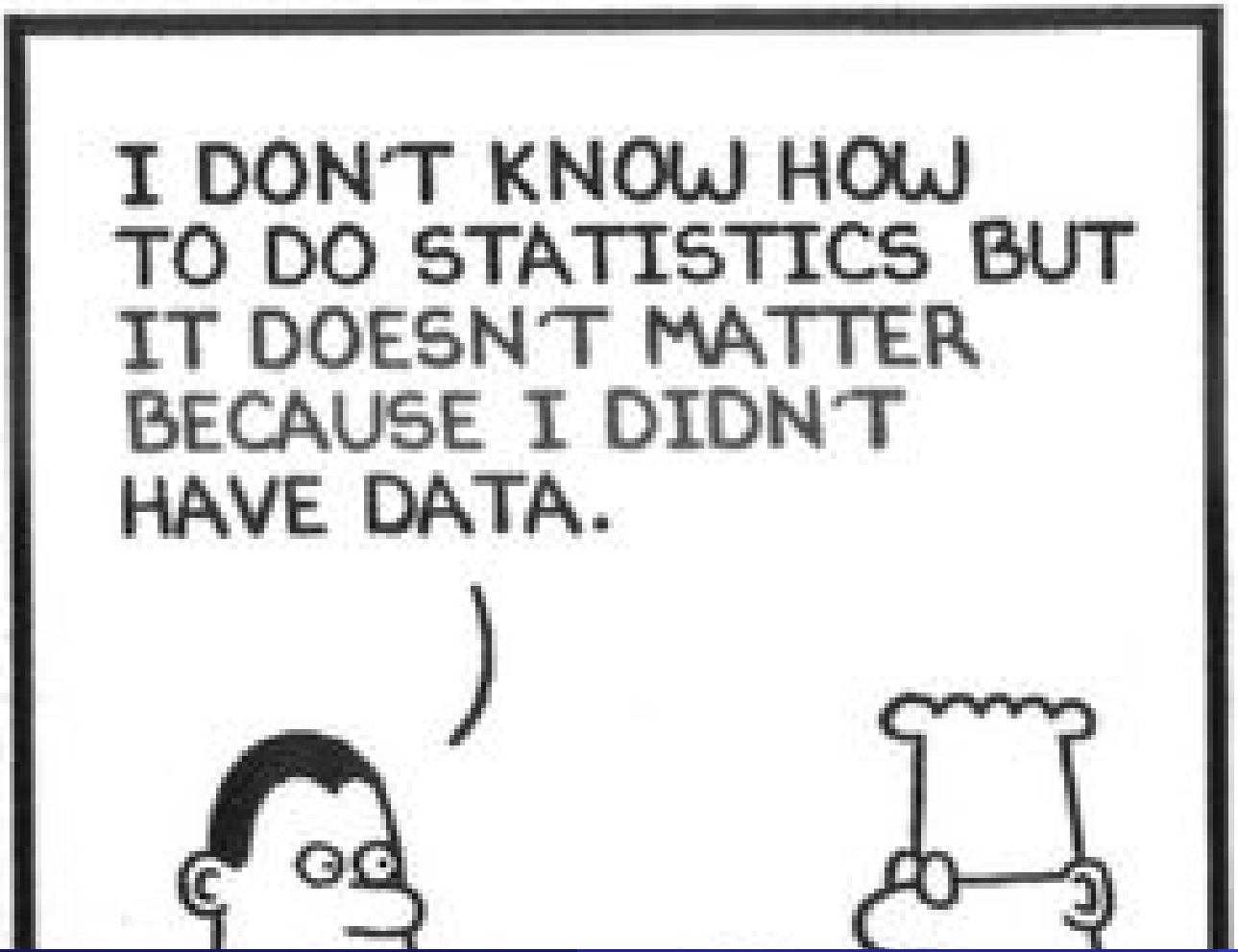
431 Class 02

Thomas E. Love

2017-08-31

Today's Agenda

1. Administration
2. The Class 1 Survey and How To Ask Questions
3. Using R, R Studio and R Markdown



TA Office Hours start Tuesday 2017-09-05

Tuesdays and Thursdays

- 11:30 - 12:30
- 2:30 - 4:30
- 5:45 - 6:45

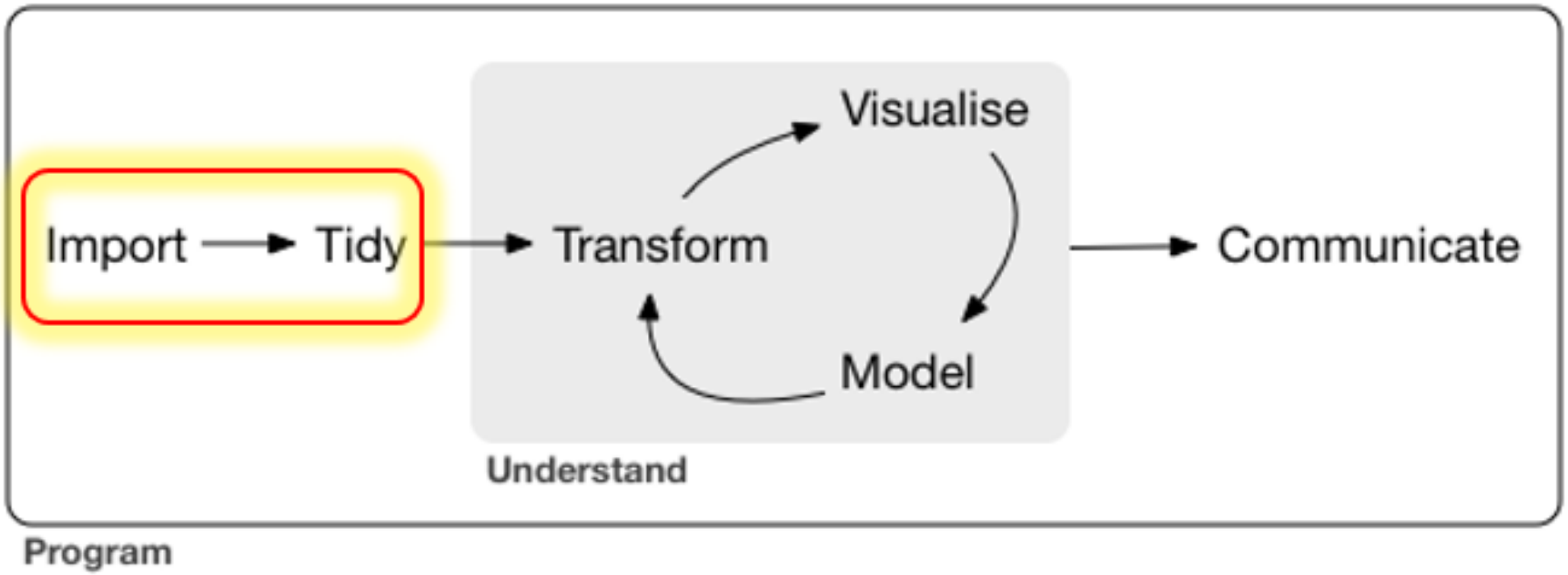
Fridays

- 4:30 - 5:30

TA office hours are held in Wood WG-56 (Computing Lab) or WG-67 (Student Lounge), so be sure to look in both places.

Contact us at 431-help@case.edu

Our web site: https://github.com/thomaselove/431


Import
Tidy
Transform
Visualise
Model
Communicate
Understand
Program

Types of Data (and see the Course Notes, section 4.3)

Data can be **quantitative (numerical)** or **qualitative (categorical)**

- **Quantitative**
  - Variables recorded in numbers that we use as numbers.
  - All quantitative variables must have units of measurement.
  - Can break into *continuous* (may take any value in a range) or *discrete* (limited set of potential values.)
    - Height is certainly continuous as a concept, but how precise is our ruler?
    - Piano vs. Violin
  - (less common) *interval* (equal distances between values, but zero point is arbitrary) as compared to *ratio* variables (a meaningful zero point.)
    - Is *weight* an interval or ratio variable? How about *IQ*?
  - Taking a mean or median is a reasonable idea.

Types of Data (and see Part A Notes, section 2)

Data can be **quantitative (numerical)** or **qualitative (categorical)**

- Qualitative
  - Variables consisting of names of categories.
  - Each possible value is a code for a category (could use numerical or non-numerical codes.)
    - *Binary* categorical variables (two categories, often labeled 1 or 0)
    - *Multi-categorical* variables (usually taken to be 3+ categories)
  - Also, *nominal* (no underlying order) or *ordinal* (categories are ordered.)
    - How is your overall health? (Excellent, Very Good, Good, Fair, Poor)
    - Which candidate would you vote for if the election were held today?
    - Did this patient receive this procedure?

Day 1 Survey Handout

431 First Day Survey (15 Questions)

Please introduce yourself to someone you do not know, ask them these 15 questions, and record their answers on this sheet. At the same time, provide your partner with your answers so they can record your responses on their sheet. Do not place any names on this sheet so that the responses will remain anonymous. Thank you!

1. What is your *sex*? (Male or Female) _____

2. Is English your *most comfortable* language? (Yes or No) _____

3. Fill in the number that best describes your answer to this question:

Has *statistical thinking* been important in your life so far?

Not at all important		Slightly important		Somewhat important		Extremely important
①	②	③	④	⑤	⑥	⑦

4. How old (in years) do you think Professor Love is? _____ years.

5. **Do you smoke?** Fill in the appropriate circle:

No Non-Smoker	I used to. Former Smoker	Yes. Smoker
①	②	③

6. Please indicate which hand you use for each of the following activities by putting a + in the appropriate column, or ++ if you would *never* use the other hand for that activity. If, in any case, you really are indifferent, put + in both columns.

Task	Left	Right
Writing		
Drawing		
Throwing		
Scissors		
Toothbrush		
Knife (without fork)		
Spoon		
Broom (upper hand)		
Striking match (hand that holds the match)		
Opening box (hand that holds the lid)		
Total Count of +s:		

Right − Left = _____ Right + Left = _____ $\frac{Right-Left}{Right+Left}$ = _____

August 30, 2016

431 First Day Survey (15 Questions)

7. **How important do you think statistics will be in your *future career*?**

Not at all important		Slightly important		Somewhat important		Extremely important
①	②	③	④	⑤	⑥	⑦

8. How much did you pay for your most recent haircut? (in $): _____

Please indicate your agreement with the following statements:

	Strongly Disagree				Strongly Agree
9. I prefer to learn from lectures than to learn from activities.	1	2	3	4	5
10. I prefer to work on projects alone than in a team.	1	2	3	4	5

11. What is your height (indicate units of measurement): _____

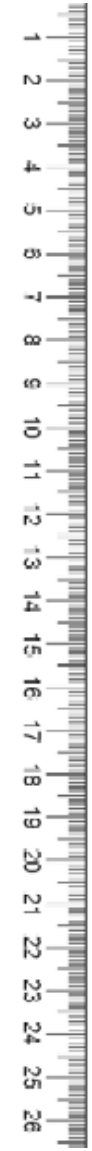
12. Use the ruler provided on the side of this page to measure the span of your right hand (distance from the thumb to the little finger when your fingers are spread apart: _____ cm.

13. What is your favorite color? _____

14. How many hours did you sleep last night? _____ hours.

15. Record your pulse by counting the beats of your heart for 30 seconds, then doubling the result: _____ beats/minute.

August 30, 2016



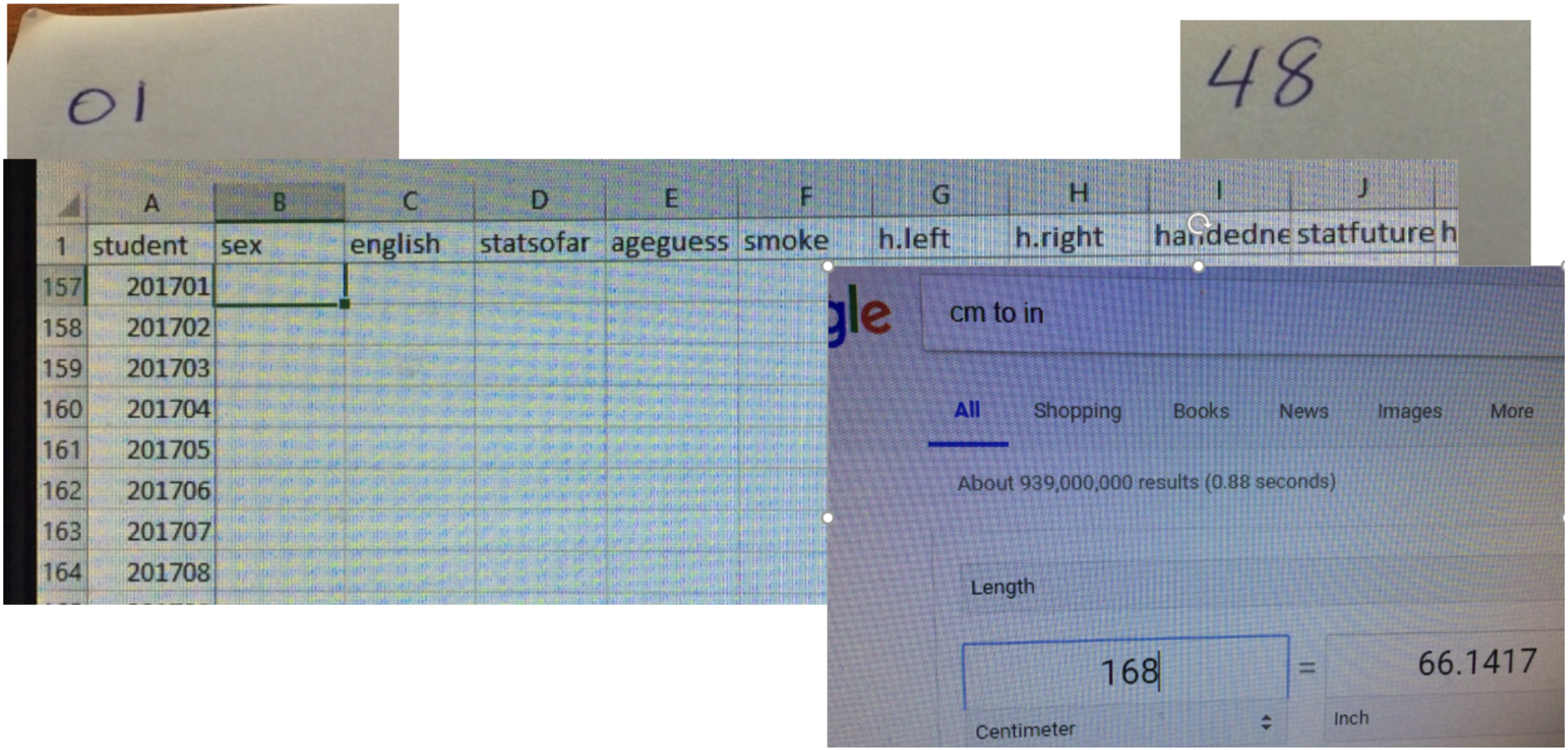
Evaluating some Day 1 Survey variables

1. Do you **smoke**? (1 = Non-Smoker, 2 = Former Smoker, 3 = Smoker)
2. How much did you pay for your most recent **haircut**? (in $)
3. What is your favorite **color**?
4. How many hours did you **sleep** last night?
5. Has statistical thinking been important in your life? (1 = Not at all important to 7 = Extremely important)

Are these quantitative or qualitative?

- If quantitative, are they *discrete* or *continuous*? Do they have a meaningful *zero point*?
- If qualitative, how many categories? *Nominal* or *ordinal*?

Day 1 Survey


01
48
A B C D E F G H I J
1 student sex english statsofar ageguess smoke h.left h.right handedne statfuture h
157 201701
158 201702
159 201703
160 201704
161 201705
162 201706
163 201707
164 201708
cm to in
All Shopping Books News Images More
About 939,000,000 results (0.88 seconds)
Length
168
=
66.1417
Centimeter
Inch

Day 1 Survey

- 48 people completed it Tuesday
- Another 64 did it in 2016, 49 did in 2015, and another 42 in 2014.

Question 1

About how many of those 203 surveys caused *no problems* in recording responses?

Day 1 Survey Handout

431 First Day Survey (15 Questions)

Please introduce yourself to someone you do not know, ask them these 15 questions, and record their answers on this sheet. At the same time, provide your partner with your answers so they can record your responses on their sheet. Do not place any names on this sheet so that the responses will remain anonymous. Thank you!

1. What is your *sex*? (Male or Female) _____

2. Is English your *most comfortable* language? (Yes or No) _____

3. Fill in the number that best describes your answer to this question:

Has *statistical thinking* been important in your life so far?

Not at all important		Slightly important		Somewhat important		Extremely important
①	②	③	④	⑤	⑥	⑦

4. How old (in years) do you think Professor Love is? _____ years.

5. **Do you smoke?** Fill in the appropriate circle:

No Non-Smoker	I used to. Former Smoker	Yes. Smoker
①	②	③

6. Please indicate which hand you use for each of the following activities by putting a + in the appropriate column, or ++ if you would *never* use the other hand for that activity. If, in any case, you really are indifferent, put + in both columns.

Task	Left	Right
Writing		
Drawing		
Throwing		
Scissors		
Toothbrush		
Knife (without fork)		
Spoon		
Broom (upper hand)		
Striking match (hand that holds the match)		
Opening box (hand that holds the lid)		
Total Count of +s:		

Right − Left = _____ Right + Left = _____ $\frac{Right-Left}{Right+Left}$ = _____

August 30, 2016

431 First Day Survey (15 Questions)

7. **How important do you think statistics will be in your *future career*?**

Not at all important		Slightly important		Somewhat important		Extremely important
①	②	③	④	⑤	⑥	⑦

8. How much did you pay for your most recent haircut? (in $): _____

Please indicate your agreement with the following statements:

	Strongly Disagree				Strongly Agree
9. I prefer to learn from lectures than to learn from activities.	1	2	3	4	5
10. I prefer to work on projects alone than in a team.	1	2	3	4	5

11. What is your height (indicate units of measurement): _____

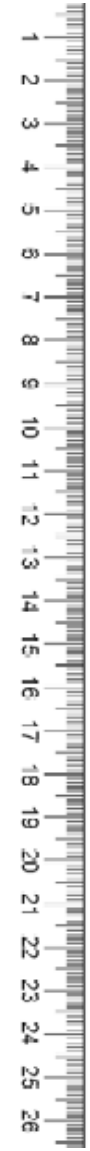
12. Use the ruler provided on the side of this page to measure the span of your right hand (distance from the thumb to the little finger when your fingers are spread apart: _____ cm.

13. What is your favorite color? _____

14. How many hours did you sleep last night? _____ hours.

15. Record your pulse by counting the beats of your heart for 30 seconds, then doubling the result: _____ beats/minute.

August 30, 2016



The 15 Survey Items

#	Topic	#	Topic
Q1	sex	Q9	lectures v activities
Q2	english	Q10	projects alone
Q3	stats so far	Q11	height
Q4	guess TL age	Q12	hand span
Q5	smoke	Q13	color
Q6	handedness	Q14	sleep
Q7	stats future	Q15	pulse rate
Q8	haircut	-	-

Question 1

About how many of those 203 surveys caused *no problems* in recording responses?

- Guesses?

Question 1

About how many of those 203 surveys caused *no problems* in recording responses?

- Guesses?
- 74/203 (38%)

Question 1

About how many of those 203 surveys caused *no problems* in recording responses?

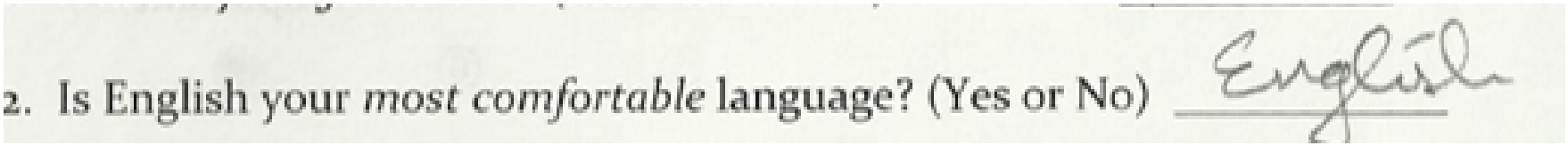
- Guesses?
- 74/203 (38%)
- 20 of the 48 surveys turned in Tuesday had **no** problems (42%)

Guess My Age

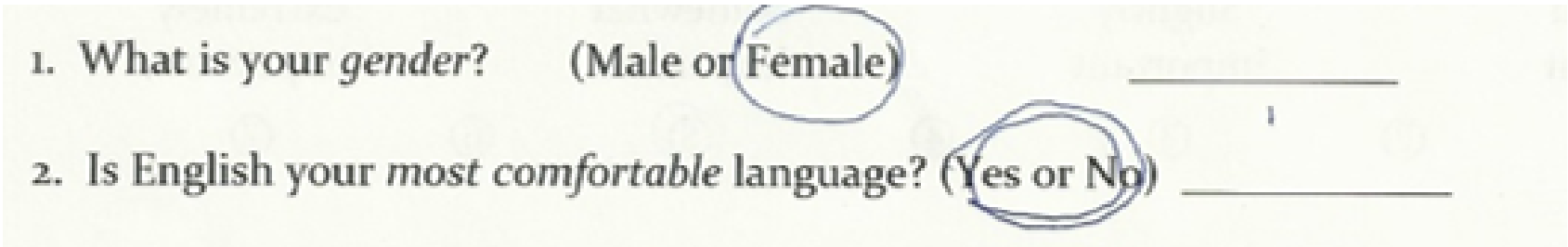
4. How old (in years) do you think Professor Love is? ~~early fifties~~ years

4. How old (in years) do you think Professor Love is? late 50's years.

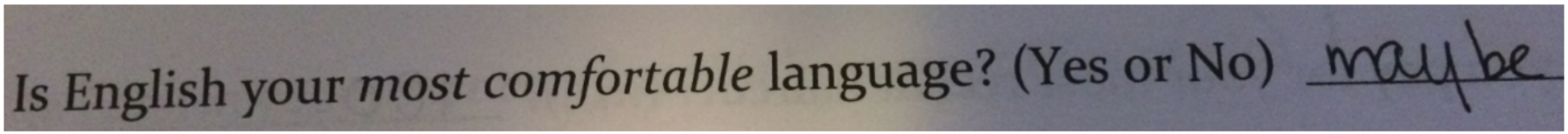
English best language?



TEL Decision: **Yes**

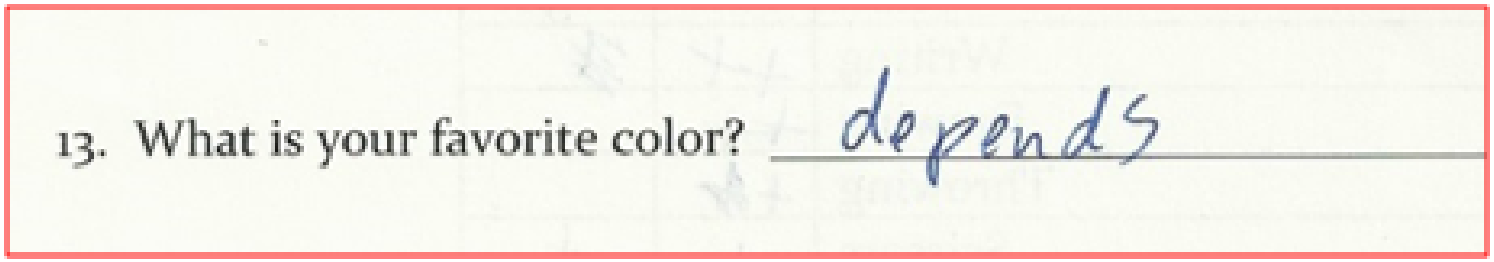


TEL Decision: **NA**

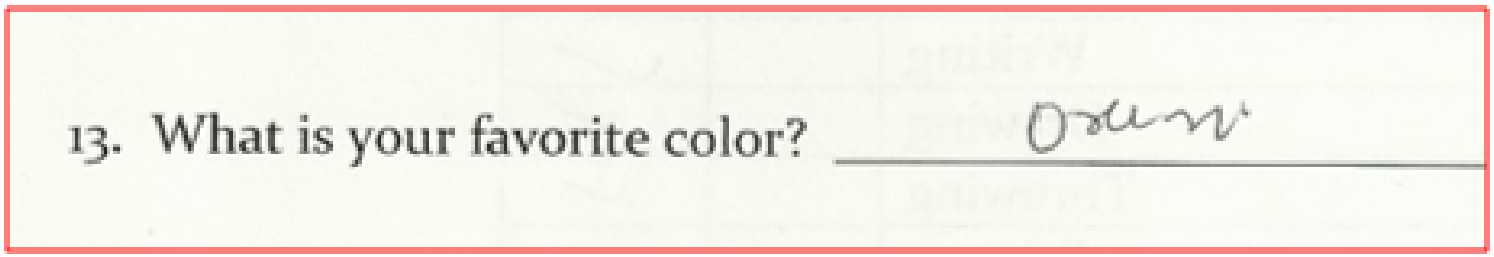


TEL decision: NA

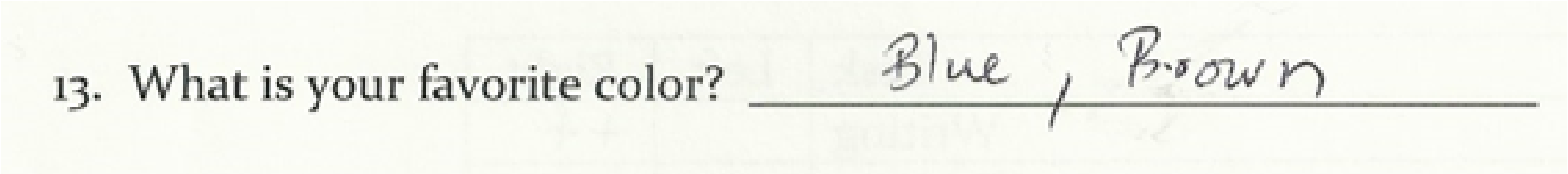
Favorite color



NA



orange



Following the Rules?

15. Record your pulse by counting the beats of your heart for 30 seconds, then doubling the result: 75 beats/minute.

2017 `pulse` responses, sorted ($n = 48$)

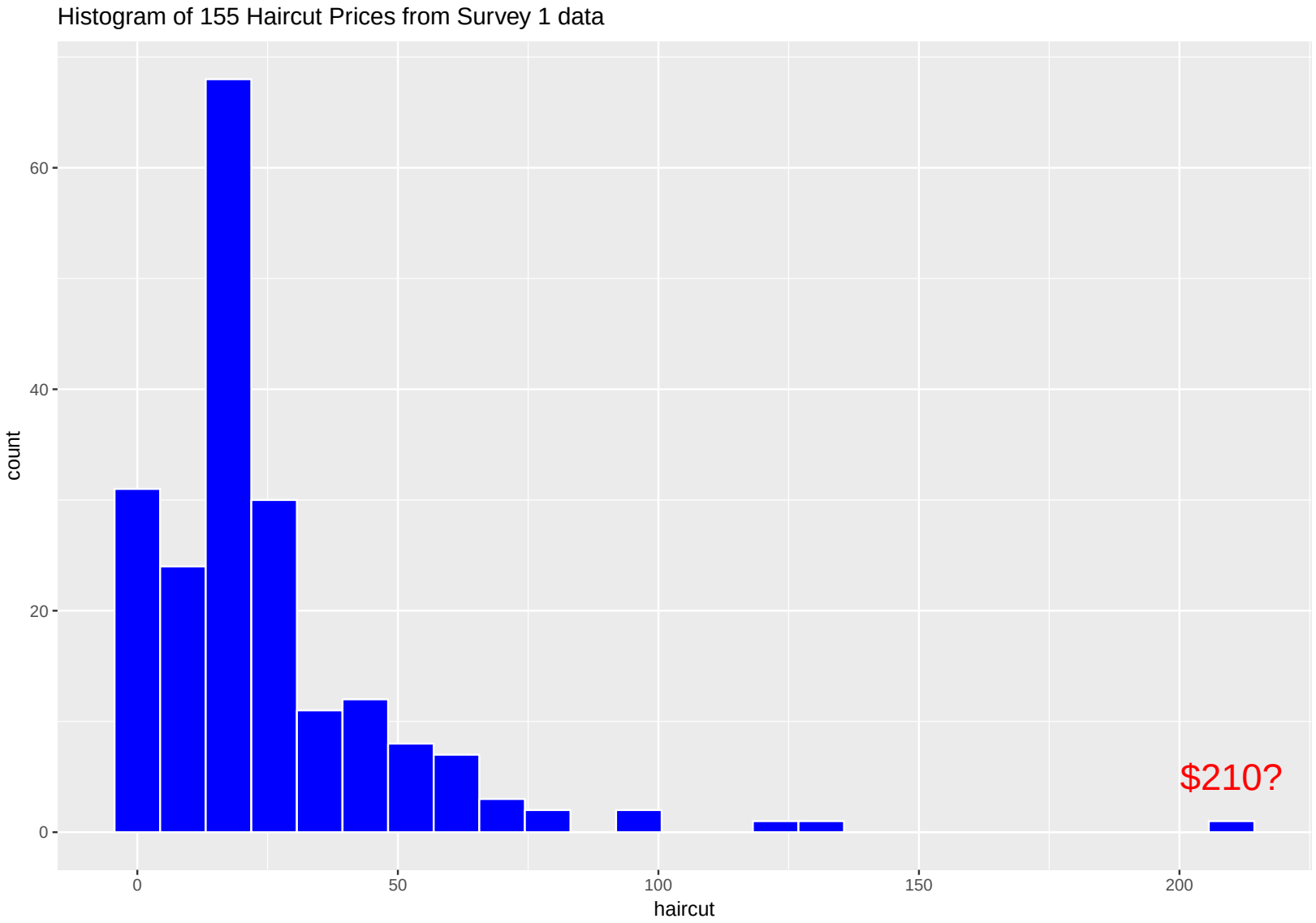
50	54	54	57	58	60
62	64	64	68	68	68
70	70	72	72	74	74
74	75	75	75	76	78
78	80	80	80	80	80
80	82	82	84	84	86
86	86	88	88	91	94
96	98	98	98	100	100

Stem and Leaf: Haircut $ (Thanks, John Tukey)

```
The decimal point is 1 digit(s) to the right of the |

  0 | 00000000000000000000000000003555555677899
  1 | 000022222222444444455555555555555555555677888889
  2 | 00000000000000000000000000012235555555555556688
  3 | 0000000000022222555555
  4 | 000000000558
  5 | 00000555
  6 | 0000000
  7 | 0005
  8 | 0
  9 |
 10 | 00
 11 |
 12 | 0
 13 | 0
 14 |
 15 |
 16 |
 17 |
 18 |
 19 |
 20 |
 21 | 0
```

Haircut Histogram



Hand Span (in cm)

12. Use the ruler provided on the side of this page to measure the span of your right hand (distance from the thumb to the little finger when your fingers are spread apart: 26 cm.

Hand Span Stem-and-Leaf Plot (HT: John Tukey)

```
The decimal point is at the |

 8 | 0
 9 | 55
10 | 00
11 | 0
12 |
13 | 5
14 |
15 |
16 | 00008
17 | 0000055
18 | 000000000000055555555555
19 | 00000000000000022355555555557
20 | 000000000000000000000000023455555555578
21 | 0000000000000000000004555555555555577
22 | 00000000000000000002555568
23 | 000000000055
24 | 000001
25 | 000
```

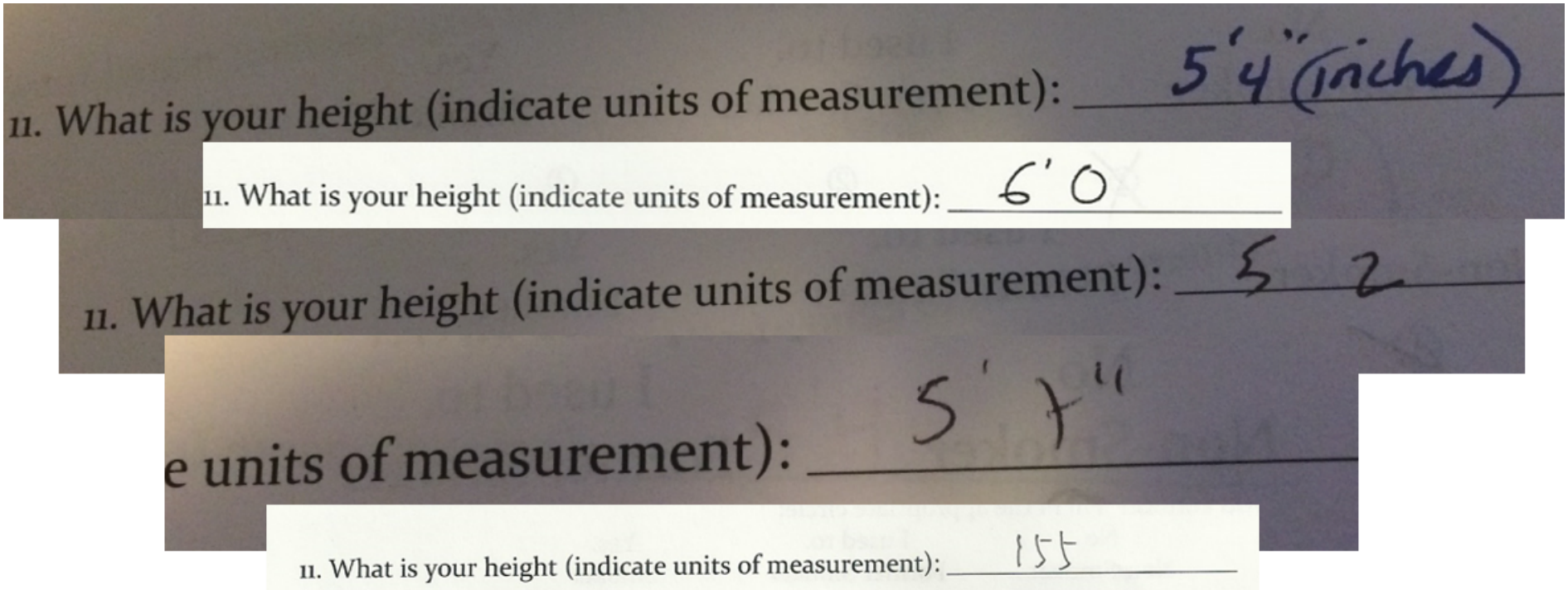
Eight Items had just a few problems (< 15 each)

#	Topic	#	Topic
-	`sex`	-	`lectures v activities`
Q2	*english*	Q10	*projects alone*
-	`stats so far`	-	`height`
Q4	*guess TL age*	Q12	*hand span*
-	`smoke`	Q13	*color*
-	`handedness`	Q14	*sleep*
-	`stats future`	Q15	*pulse rate*
Q8	*haircut*	-	-

Question 2

Of the remaining 7 (sex, stats so far, smoke, handedness, stats future, lectures vs activities, height), 5 had no real problems, and two were messy. Which two?

Height


11. What is your height (indicate units of measurement): 5'4" (inches)
11. What is your height (indicate units of measurement): 6'0
11. What is your height (indicate units of measurement): 5 2
e units of measurement): 5' 7"
11. What is your height (indicate units of measurement): 155

Handedness Scale (2014-15 version)

6. Please indicate which hand you use for each of the following activities by putting a + in the appropriate column, or ++ if you would never use the other hand for that activity. If in any case you really are indifferent, put + in both columns.

Task	Left	Right
Writing		✓
Drawing		✓
Throwing		✓
Scissors		✓
Toothbrush	✓	
Knife (without fork)	✓	
Spoon	✓	✓
Broom (upper hand)		✓
Striking match (hand that holds the match)		✓
Opening box (hand that holds the lid)		✓
Total Count of +s:	3	8

Handedness Scale (2016-17 version)

6. Please indicate which hand you use for each of the following activities by putting a + in the appropriate column, or ++ if you would *never* use the other hand for that activity. If, in any case, you really are indifferent, put + in both columns.

Task	Left	Right
Writing	++	+
Drawing	+ +	+
Throwing	+ +	+
Scissors	+ +	+
Toothbrush	++	+
Knife (without fork)	+ +	+
Spoon	+ +	+
Broom (upper hand)	++	++
Striking match (hand that holds the match)	++	+
Opening box (hand that holds the lid)	+ +	+
Total Count of +s:	20	11

Handedness Scale (2016-17 version)

6. Please indicate which hand you use for each of the following activities by putting a + in the appropriate column, or ++ if you would *never* use the other hand for that activity. If, in any case, you really are indifferent, put + in both columns.

Task	Left	Right
Writing	++	+
Drawing	+	+
Throwing	+	+
Scissors	+	+
Toothbrush	+	+
Knife (without fork)	+	+
Spoon	+	+
Broom (upper hand)	+	+
Striking match (hand that holds the match)	+	+
Opening box (hand that holds the lid)	+	+
Total Count of +s:		10

Right – Left = 10 Right + Left = 10 $\frac{Right-Left}{Right+Left} = \frac{10}{10} = 1$

Garbage in, garbage out . . .

"Data don't make any sense,
we will have to resort to statistics."

Today's Steps

We assume you were able to follow the software installation instructions.

1. Get data from our web site and store it in a new directory on your machine.
2. Open R Studio and start a Project, selecting the new directory for the Project.
3. Open and set up an R Markdown file to do the work.
4. Write code in your R Markdown file to
    - load the data in R
    - load up the `tidyverse` suite of R packages you'll need
    - look at the data
    - visualize the data
    - summarize the data numerically
    - compare the data
    - build a model for the data
5. Compile your R Markdown file to generate an HTML or Word document.

Analyzing the Index Card Guesses of My Age

47 students turned in an index card, meant to contain both a first and a second guess of my age.

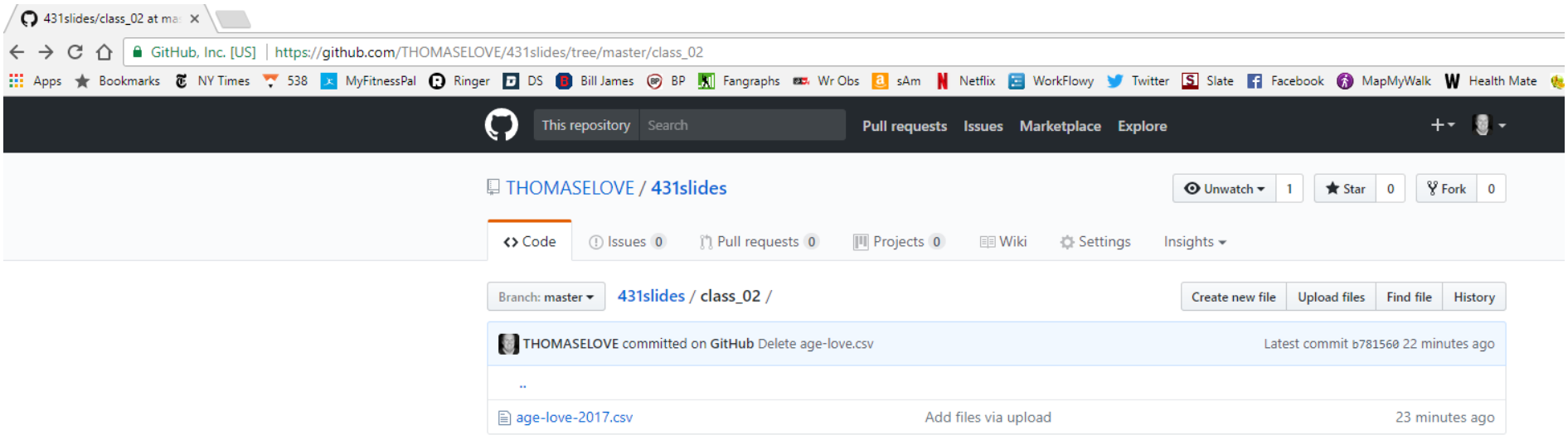
Step 1: Get the Data

- I've stored the data in a .csv file on our web site, for instance, at

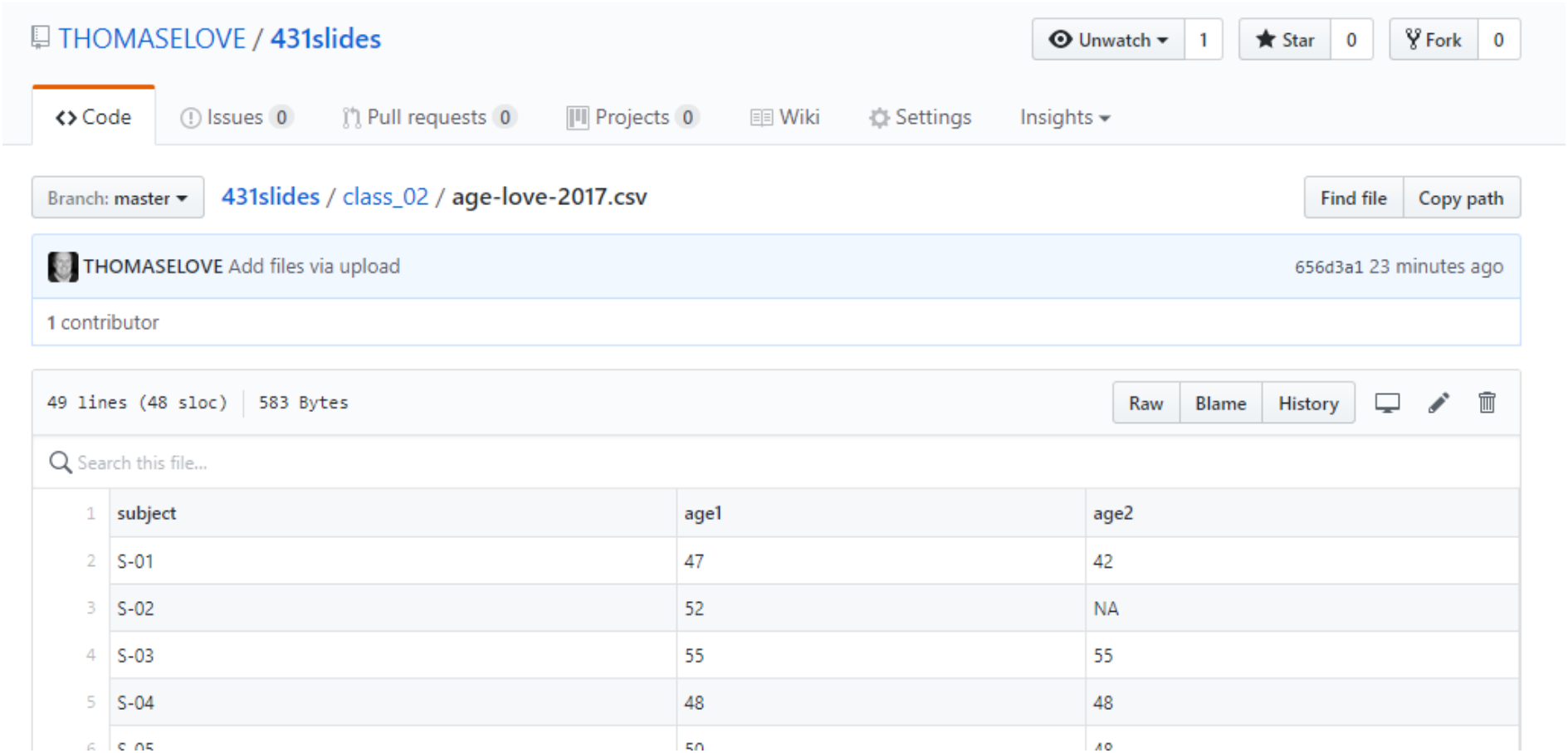
https://github.com/THOMASELOVE/431slides/tree/master/class_02

- We'll grab just that data file, for now, by clicking on it, selecting Raw, and saving the resulting **age-love-2017.csv** file to our computer.
- Specifically, we'll save it to a new directory called **431class2**.

Class 2 Github page - select the age-love-2017.csv file



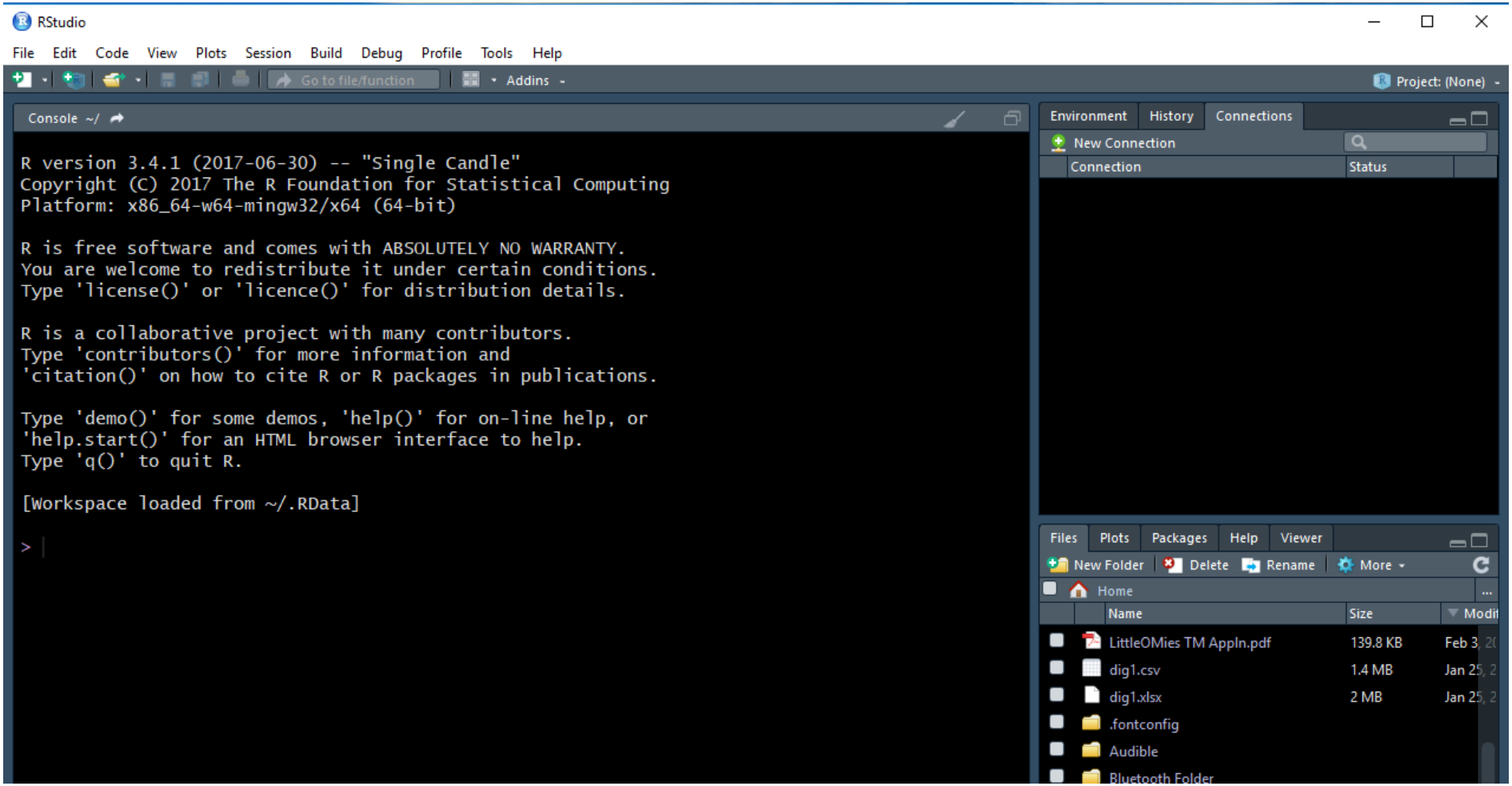
Right-click Raw to download (just) this file, into a 431class2 directory, please.



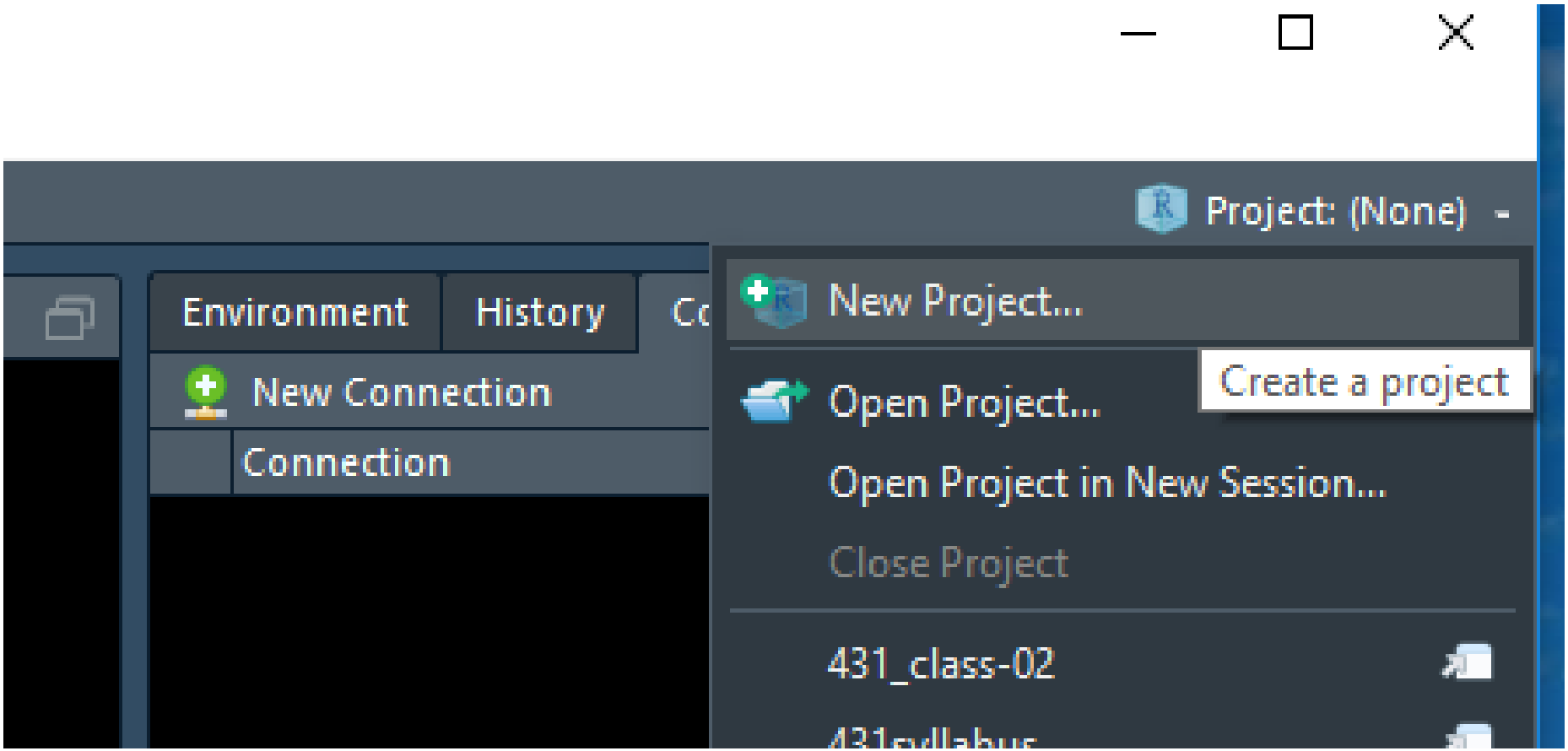
What the file looks like

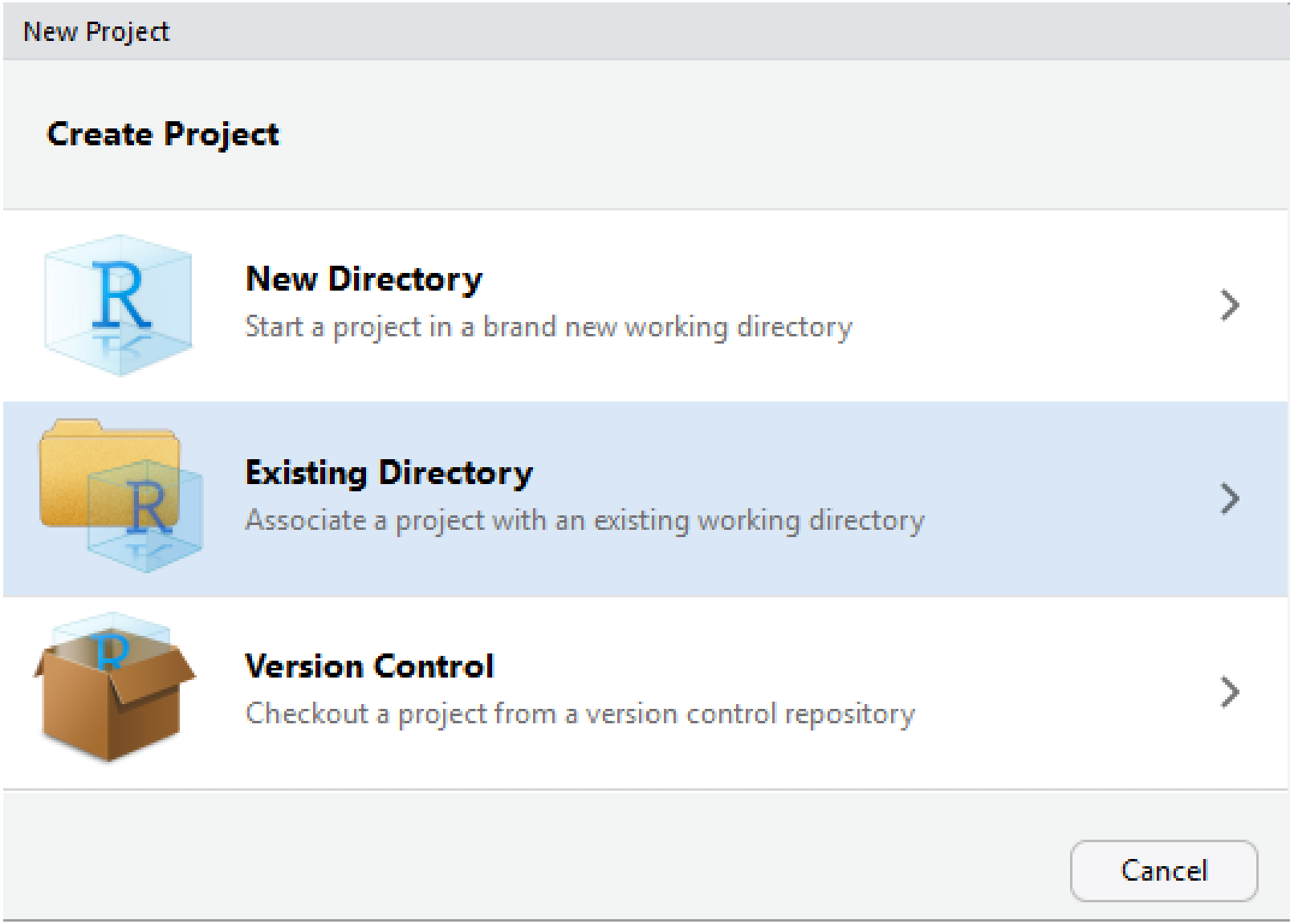
	A	B	C
1	subject	age1	age2
2	S-01	47	42
3	S-02	52	NA
4	S-03	55	55
5	S-04	48	48
6	S-05	50	48
7	S-06	51	50
8	S-07	47	49
9	S-08	45	48
10	S-09	55	55
11	S-10	40	48
12	S-11	50	54

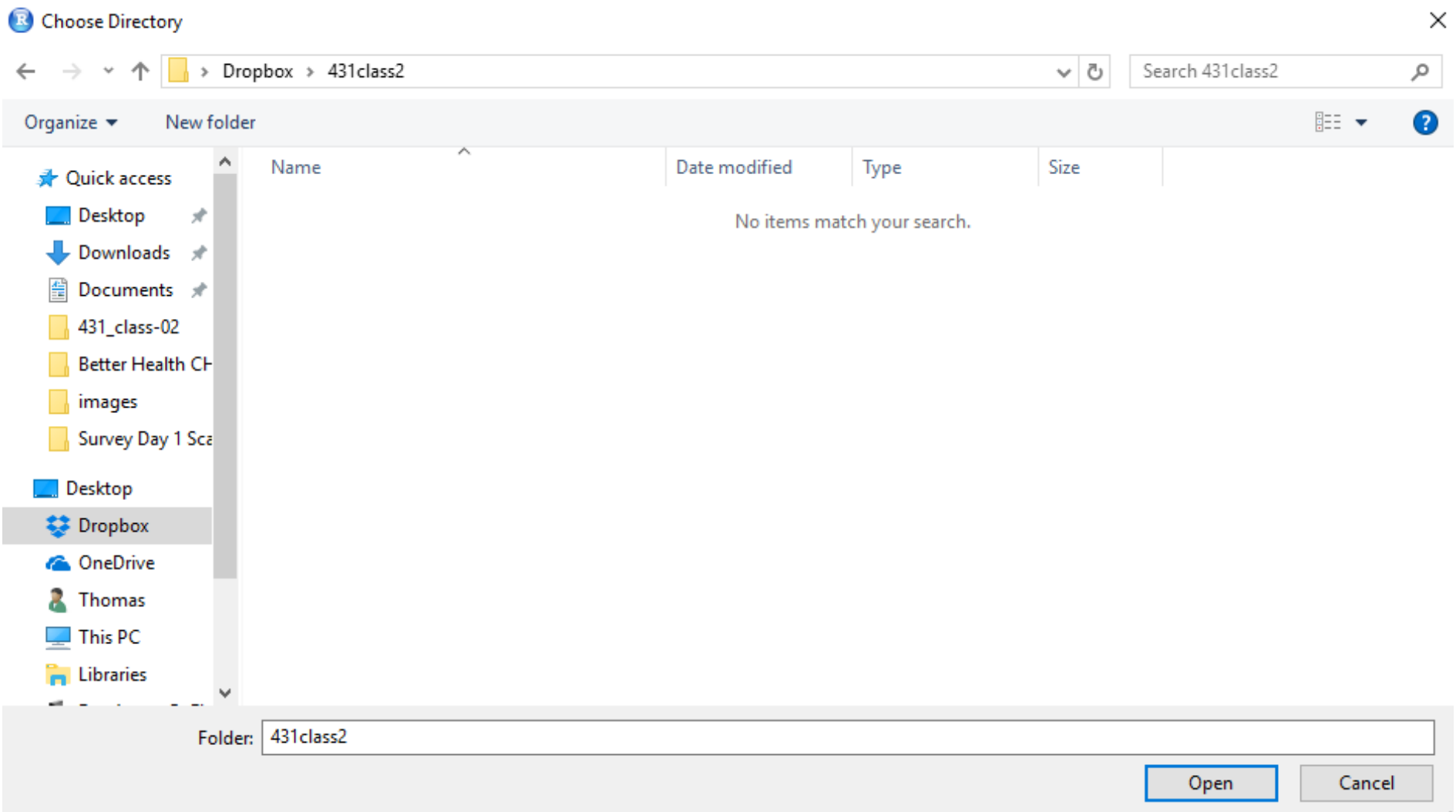
Open R Studio and start a New Project

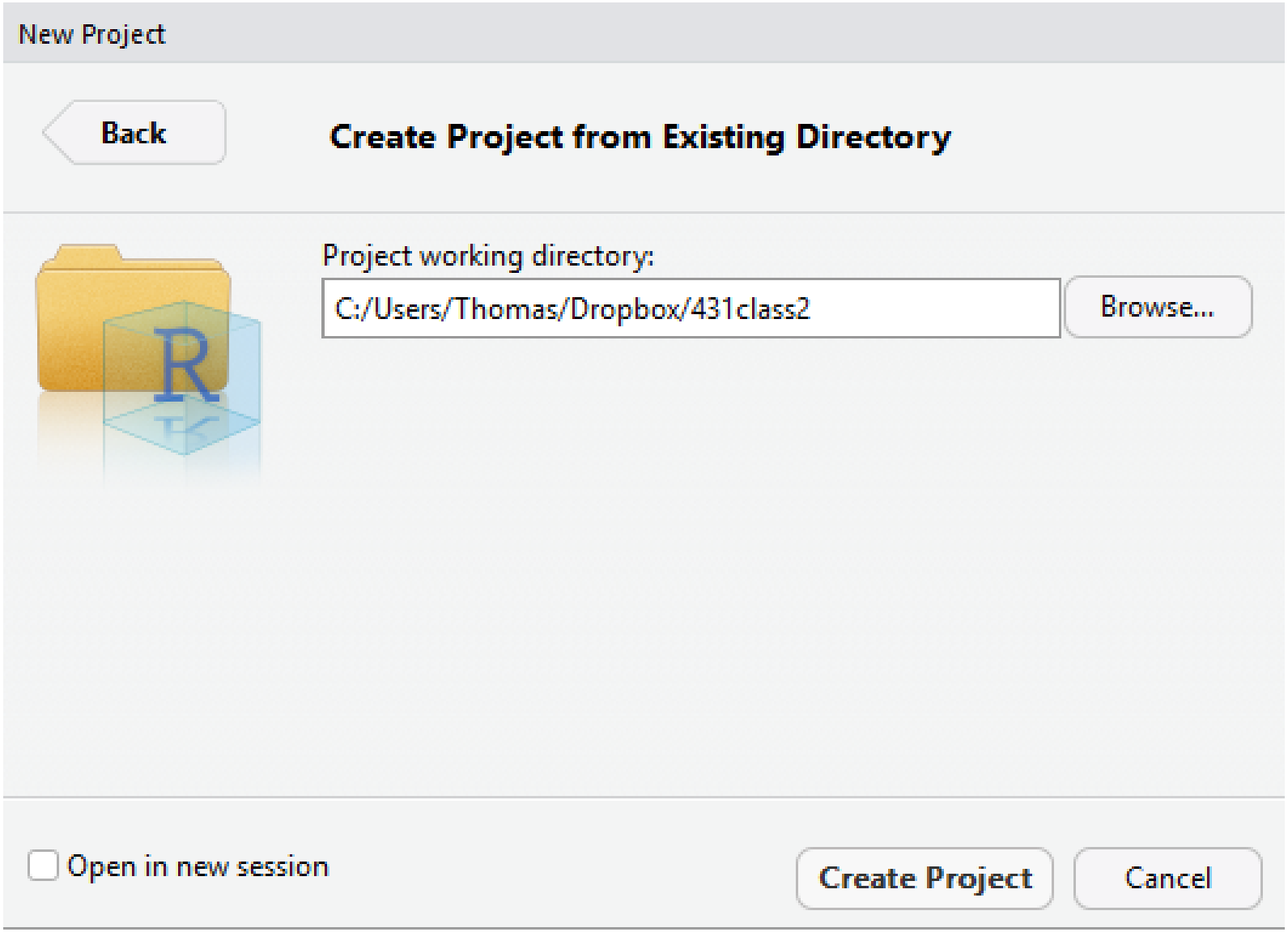


We'll select our existing 431class2 directory for this project.

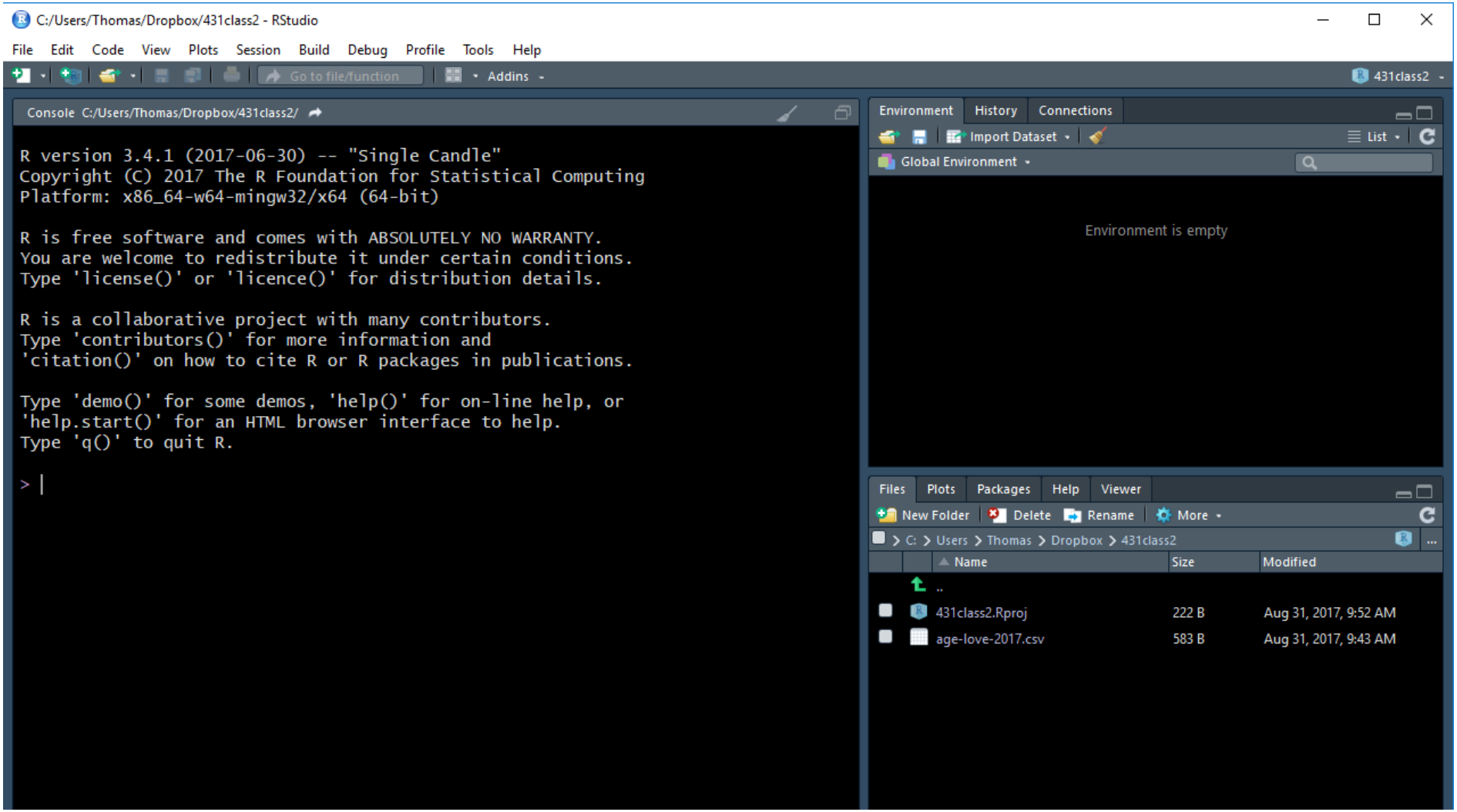



New Project
Create Project
New Directory
Start a project in a brand new working directory
Existing Directory
Associate a project with an existing working directory
Version Control
Checkout a project from a version control repository
Cancel

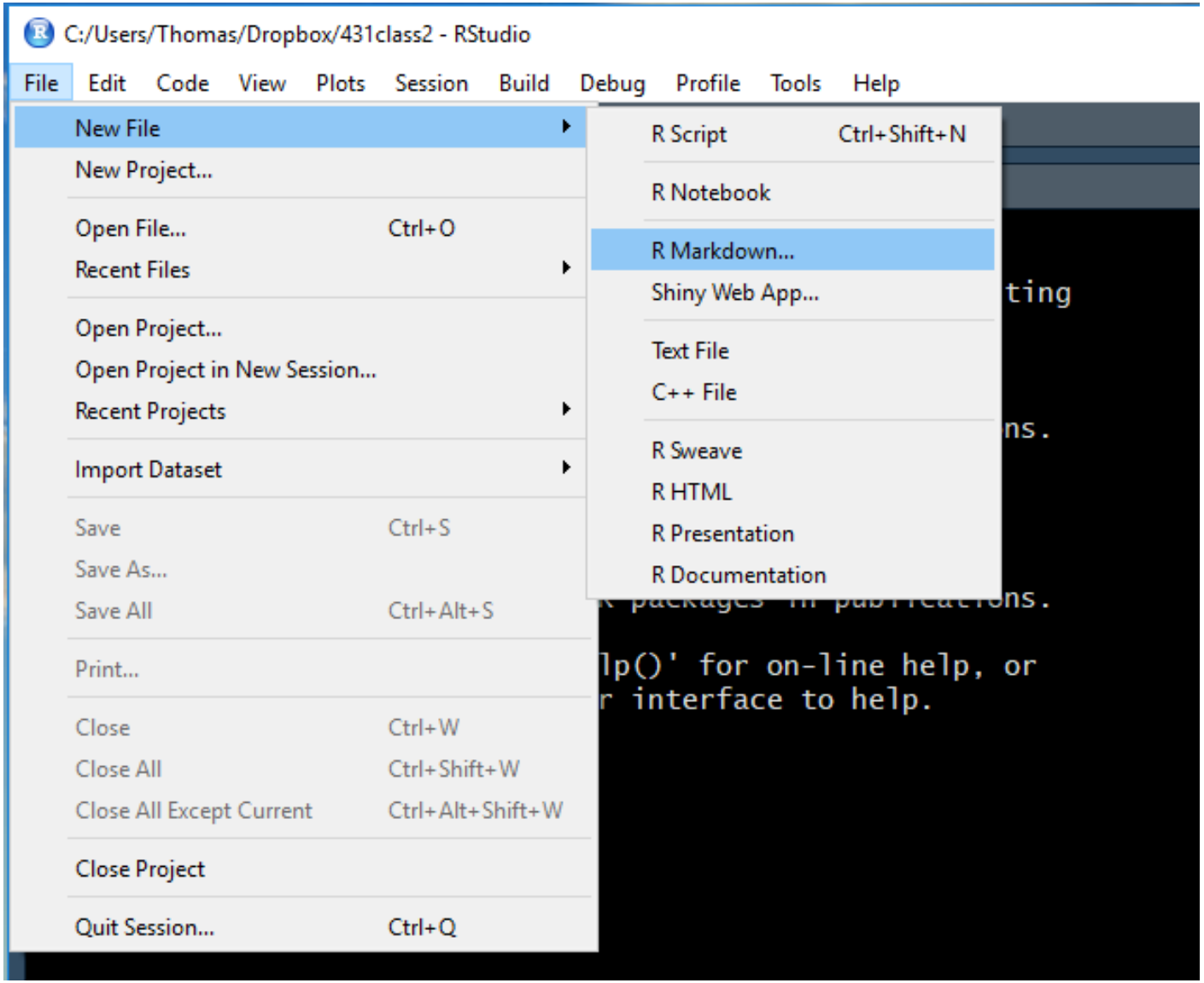

Choose Directory
Dropbox › 431class2
Search 431class2
Organize
New folder
Quick access
Desktop
Downloads
Documents
431_class-02
Better Health CH
images
Survey Day 1 Sca
Desktop
Dropbox
OneDrive
Thomas
This PC
Libraries
Name
Date modified
Type
Size
No items match your search.
Folder:
431class2
Open
Cancel


New Project
Back
Create Project from Existing Directory
Project working directory:
C:/Users/Thomas/Dropbox/431class2
Browse...
Open in new session
Create Project
Cancel

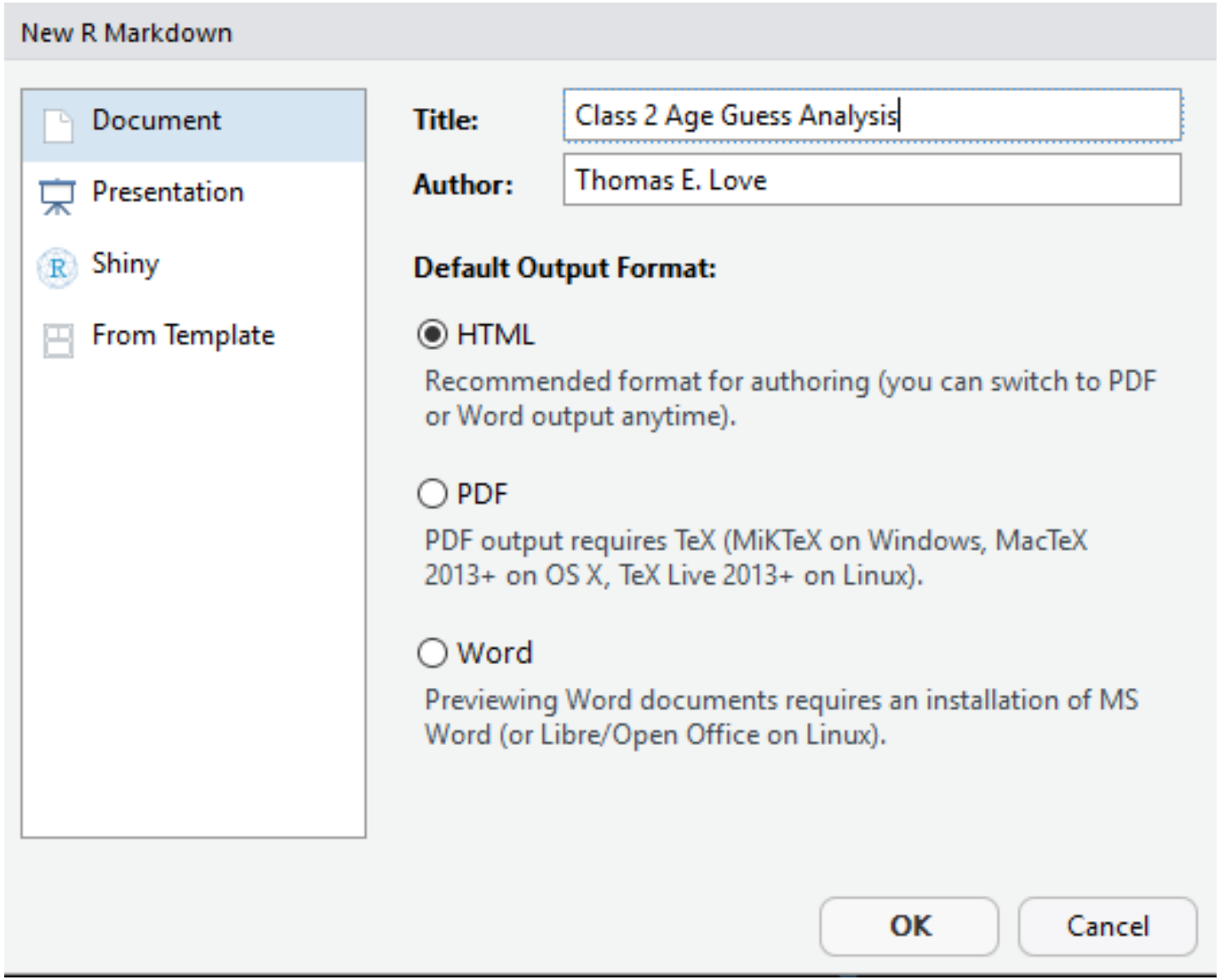
Now we are in our new Project space.



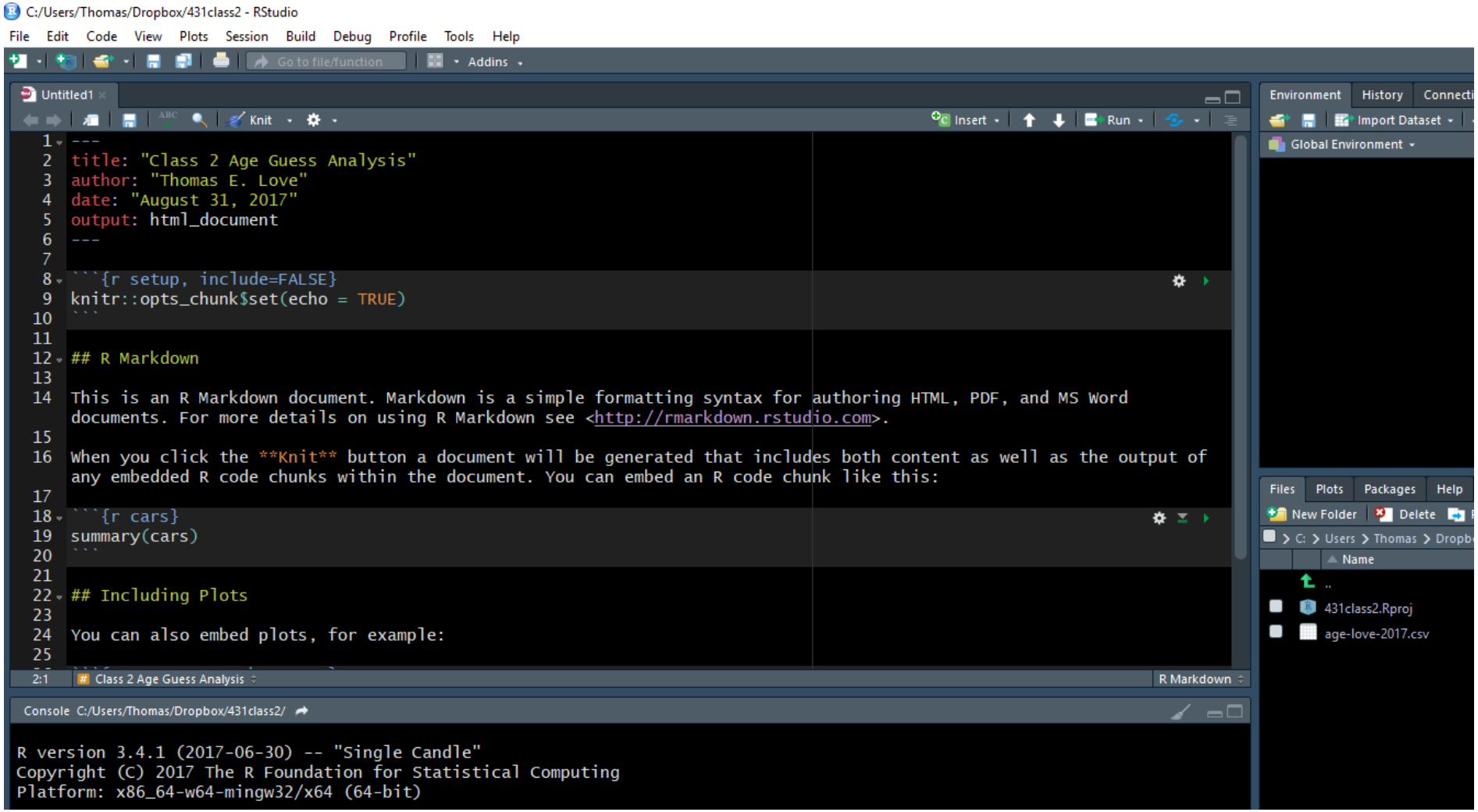
Start a new R Markdown file



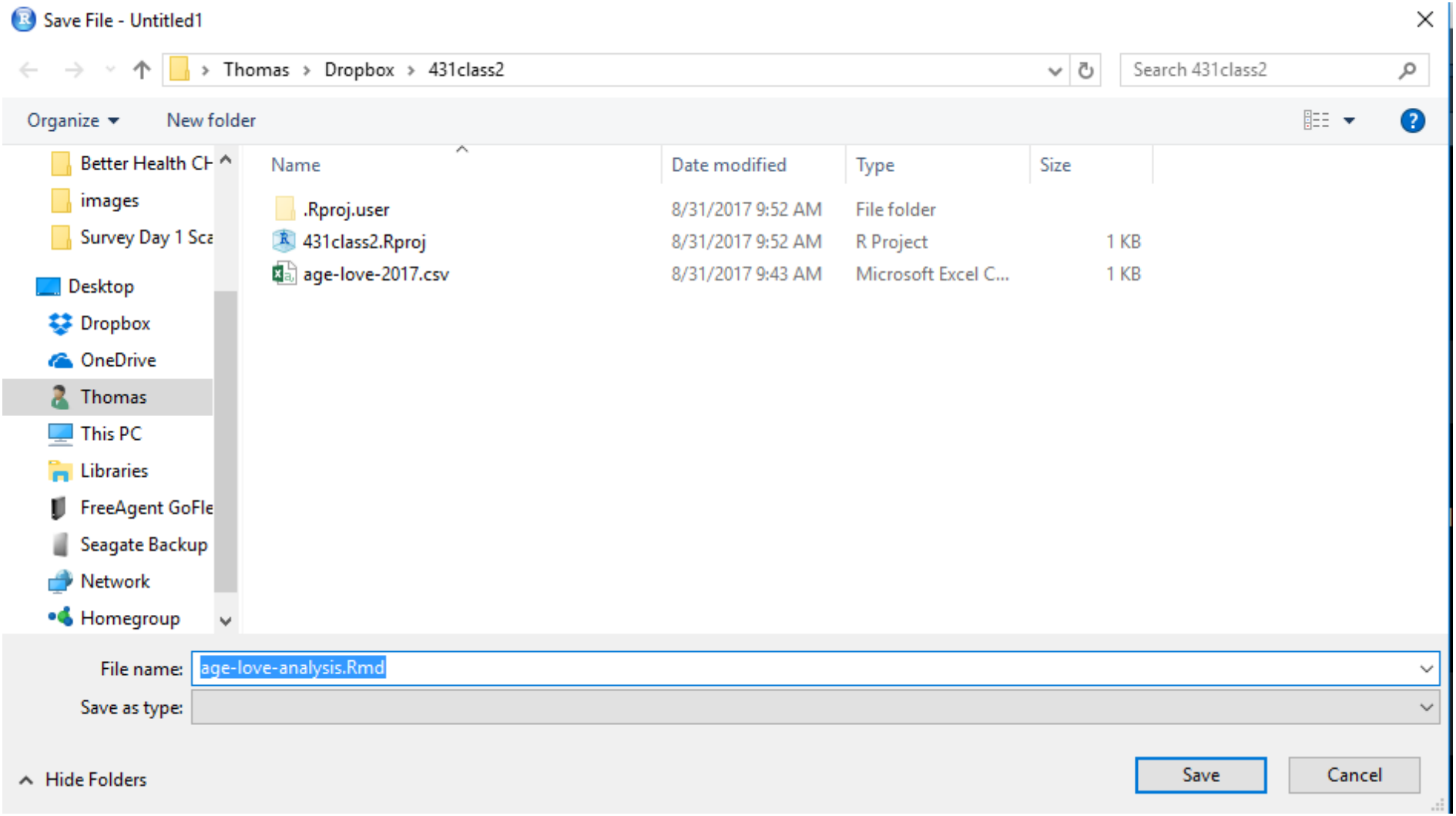
Fill in your name and title for the analysis



A sample Markdown file is generated. Let's knit it into an HTML file.



Save your work as age-love-analysis.Rmd



The result - a web file!

C:/Users/Thomas/Dropbox/431class2/age-love-analysis.html

age-love-analysis.html | Open in Browser | Find | Publish

Class 2 Age Guess Analysis

Thomas E. Love

August 31, 2017

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

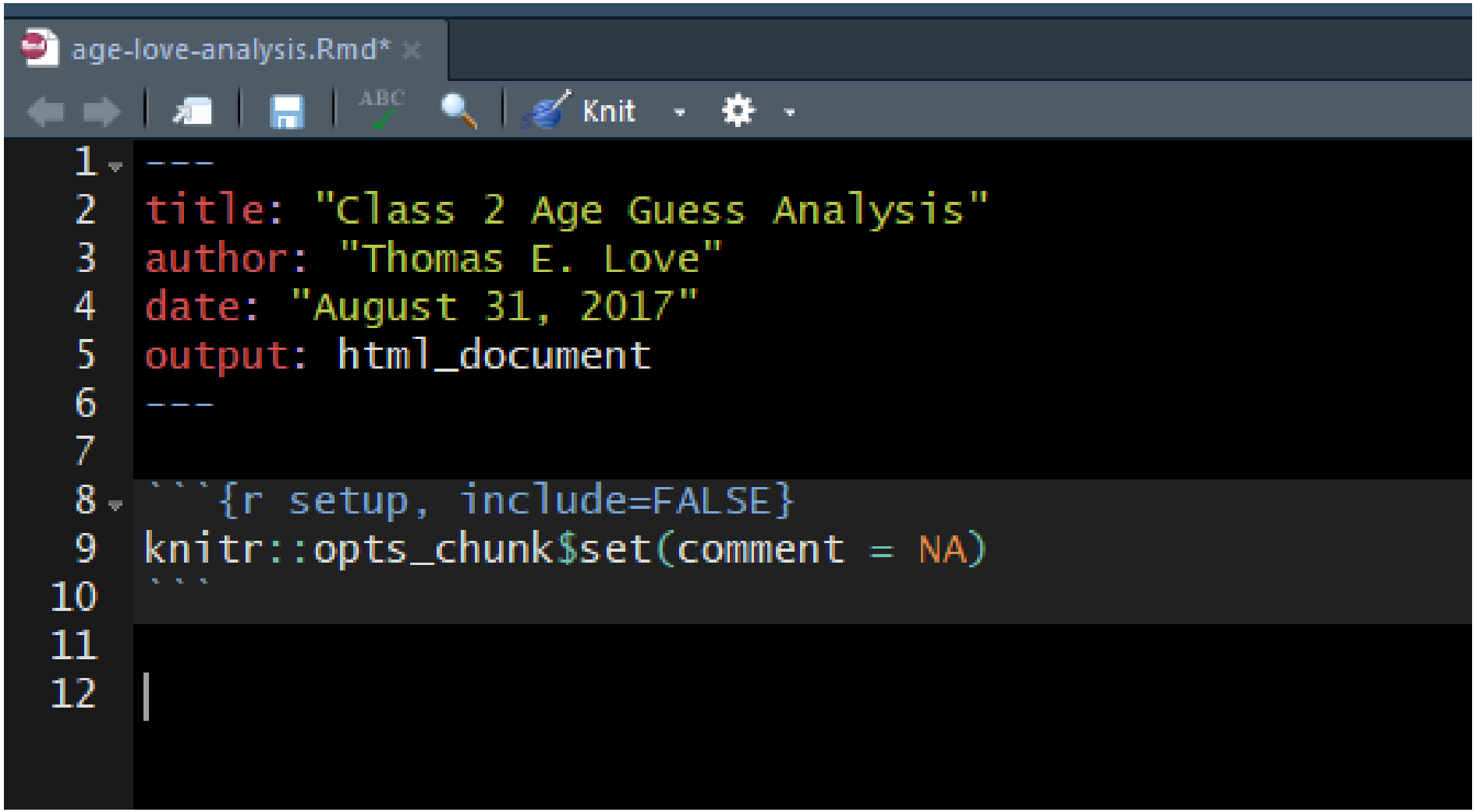
```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

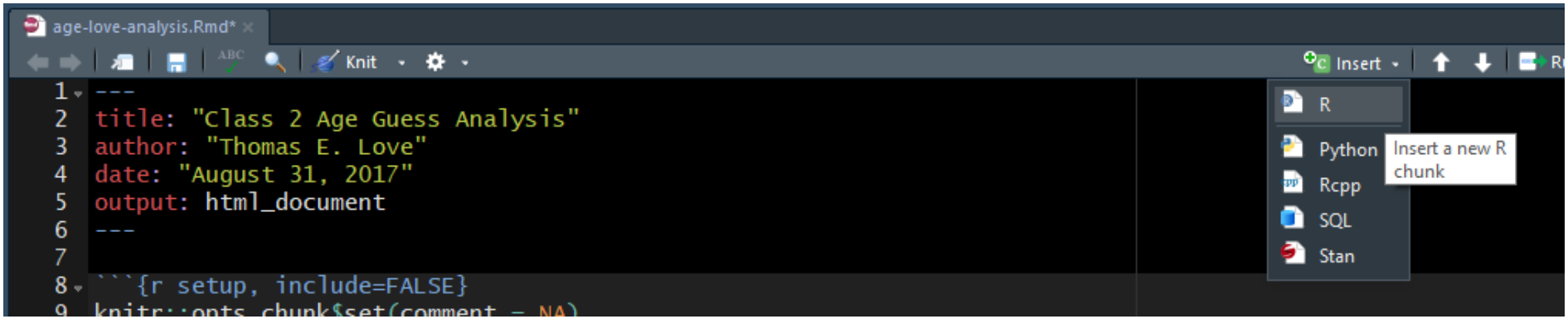
Including Plots

You can also embed plots, for example:

Edit the file to change the setup materials



Insert a new "chunk" of R code

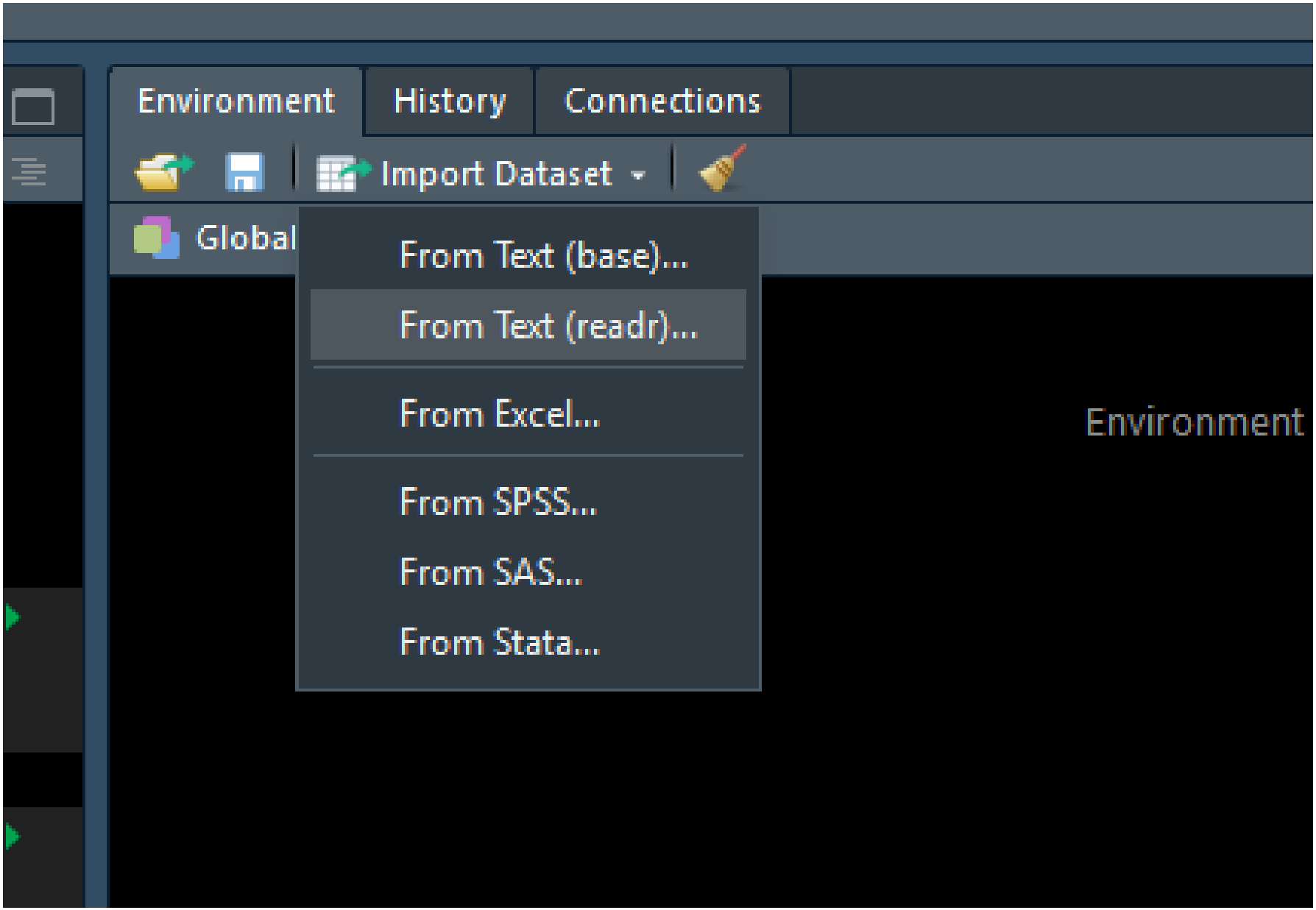


A blank "chunk"

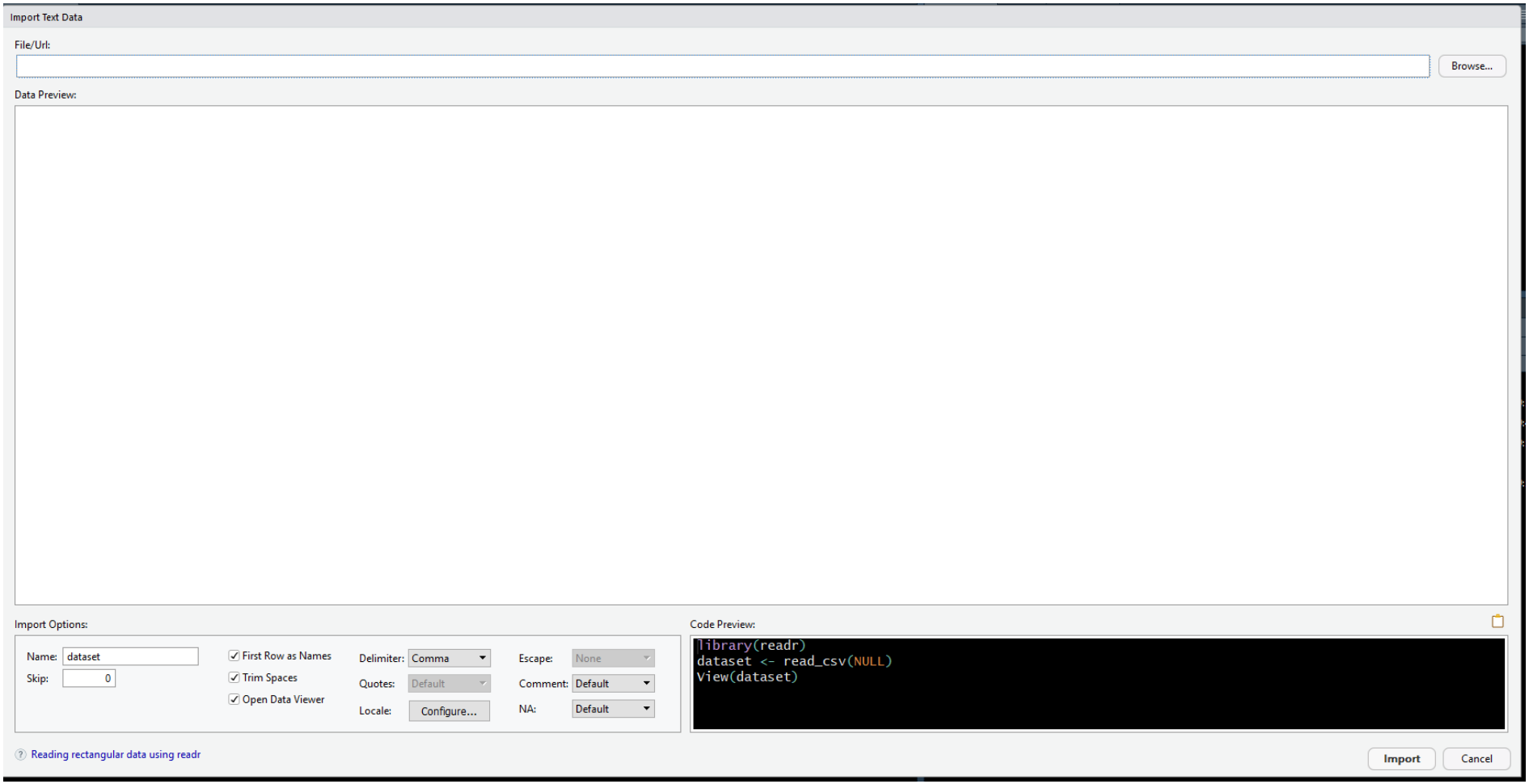
Load up two packages in R (should be installed already)

```
1  ---
2  title: "Class 2 Age Guess Analysis"
3  author: "Thomas E. Love"
4  date: "August 31, 2017"
5  output:
6      html_document:
7          code_folding: show
8  ---
9
10 ```{r setup, include=FALSE}
11 knitr::opts_chunk$set(comment = NA)
12 ```
13
14 ```{r load-packages}
15 library(magrittr); library(tidyverse)
16 ```
```

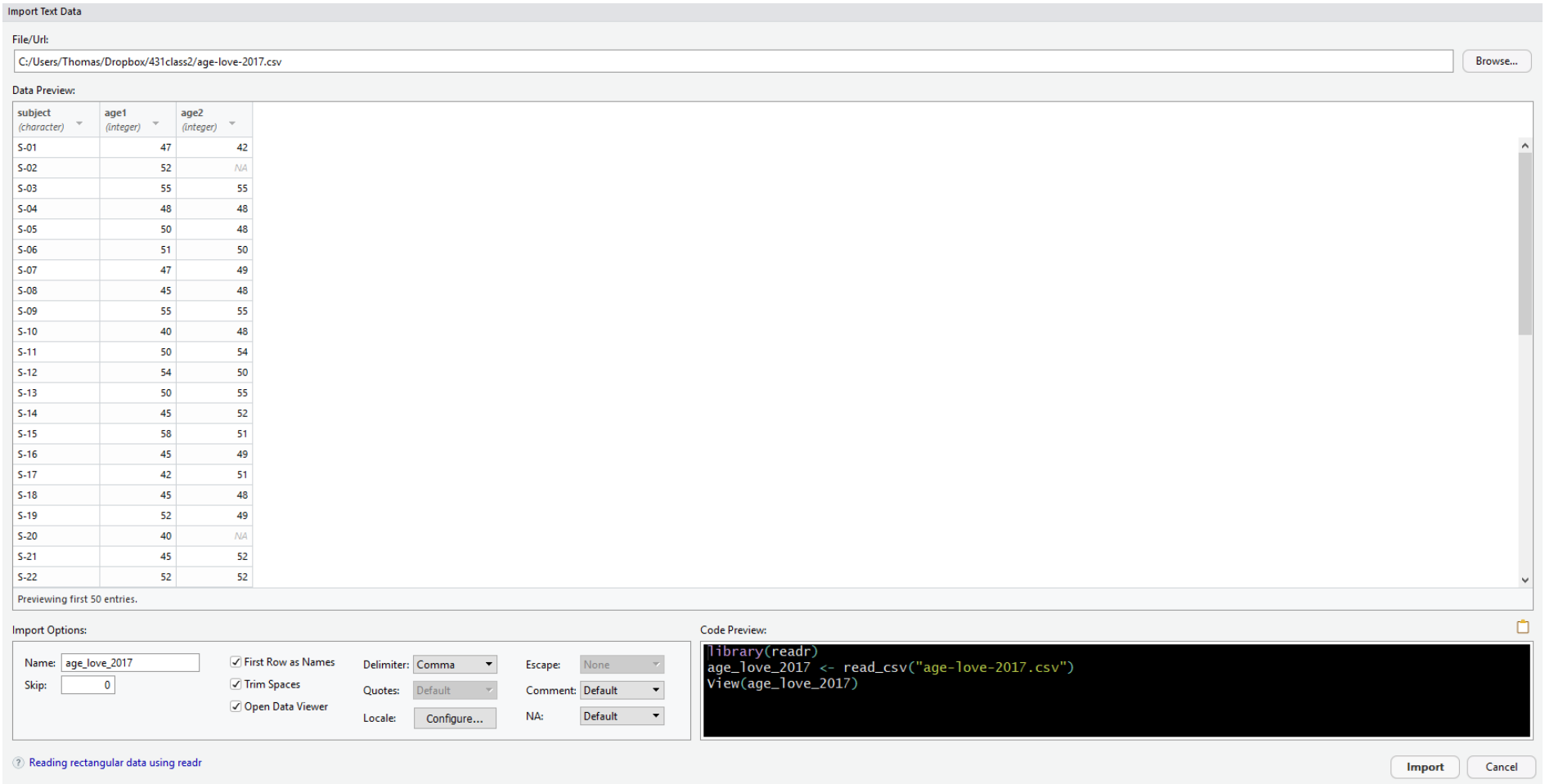
Import the .csv data set



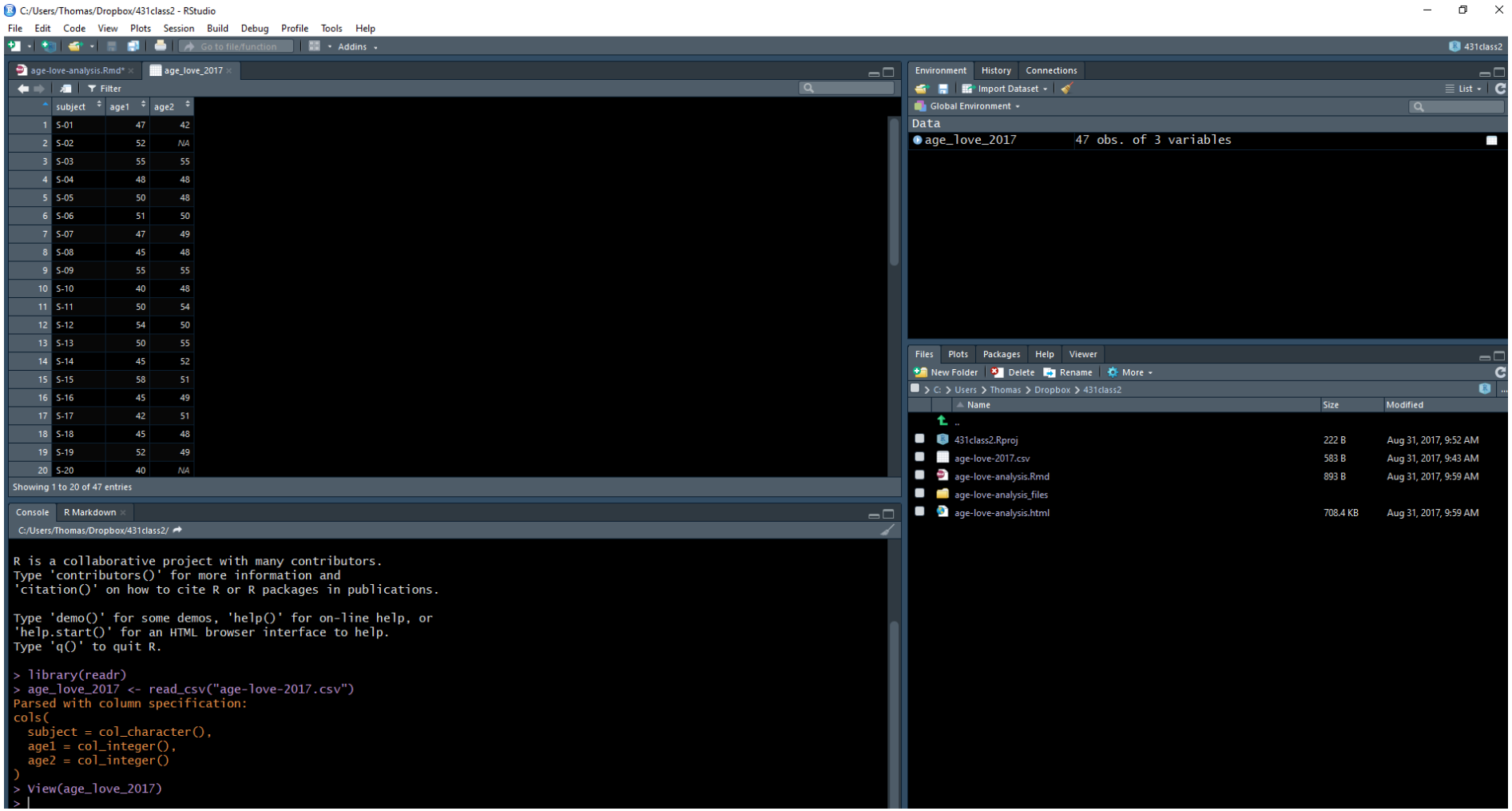
Import window



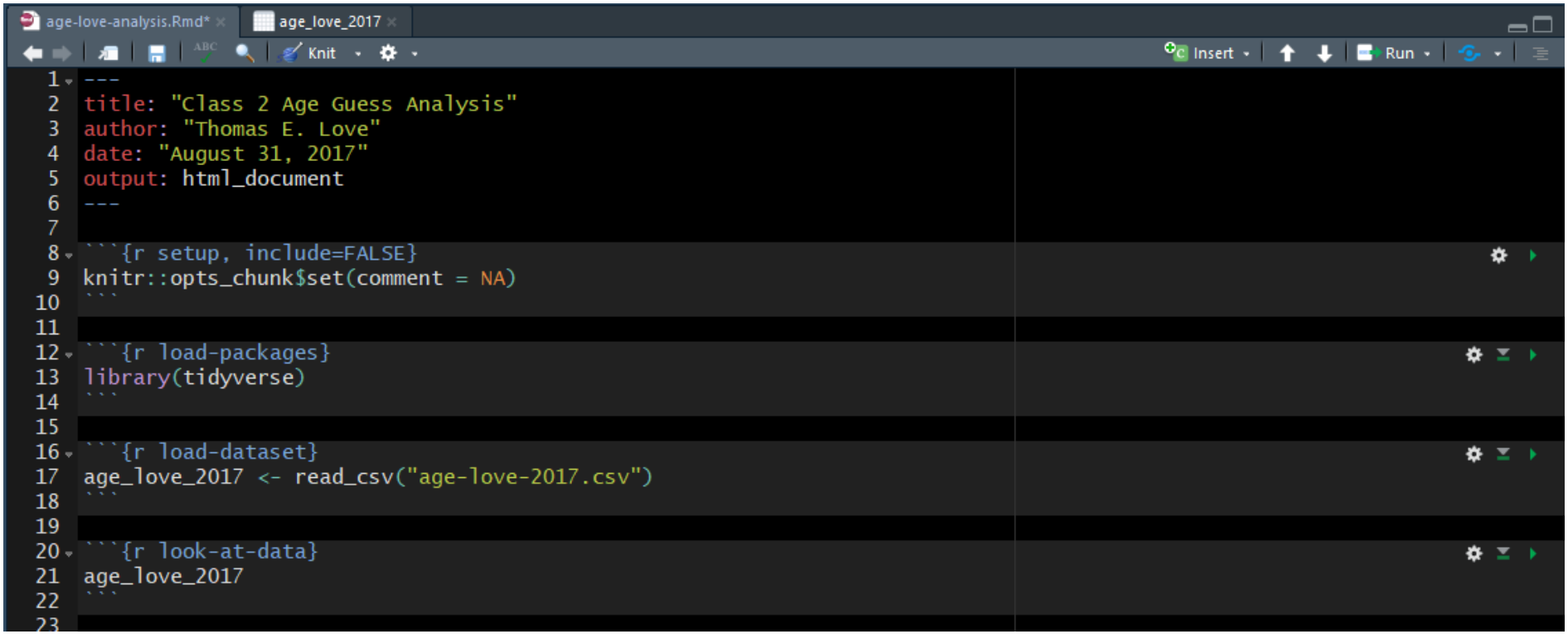
After we make our choices. . .



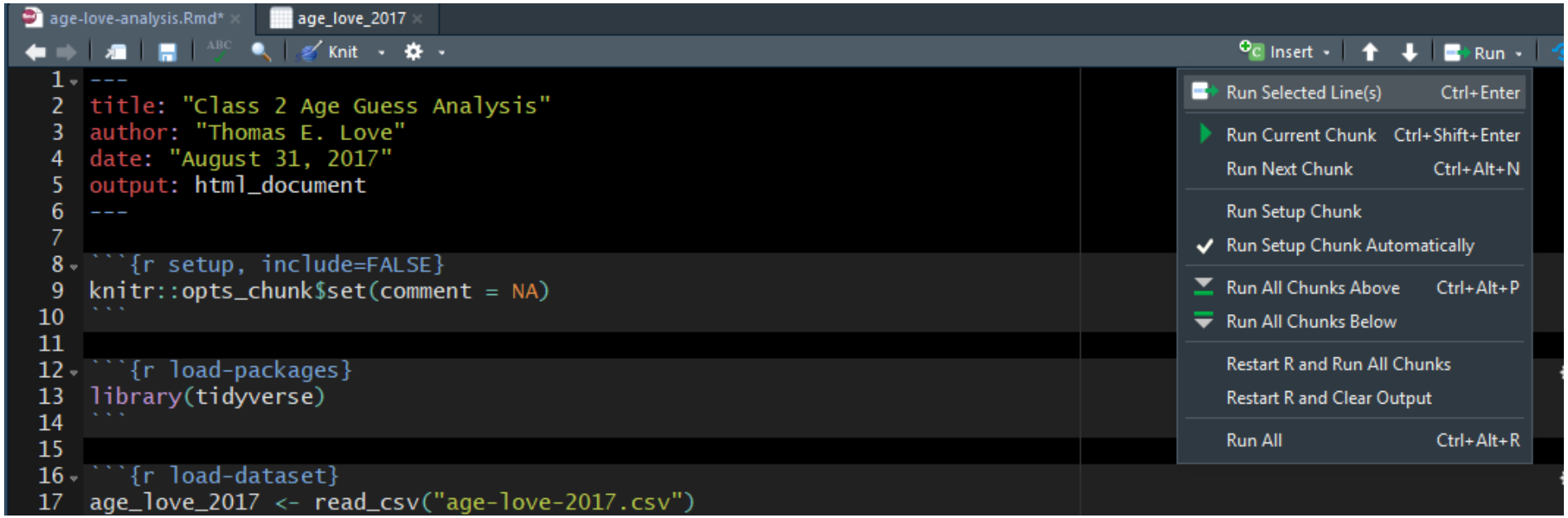
Result (note code in Console)



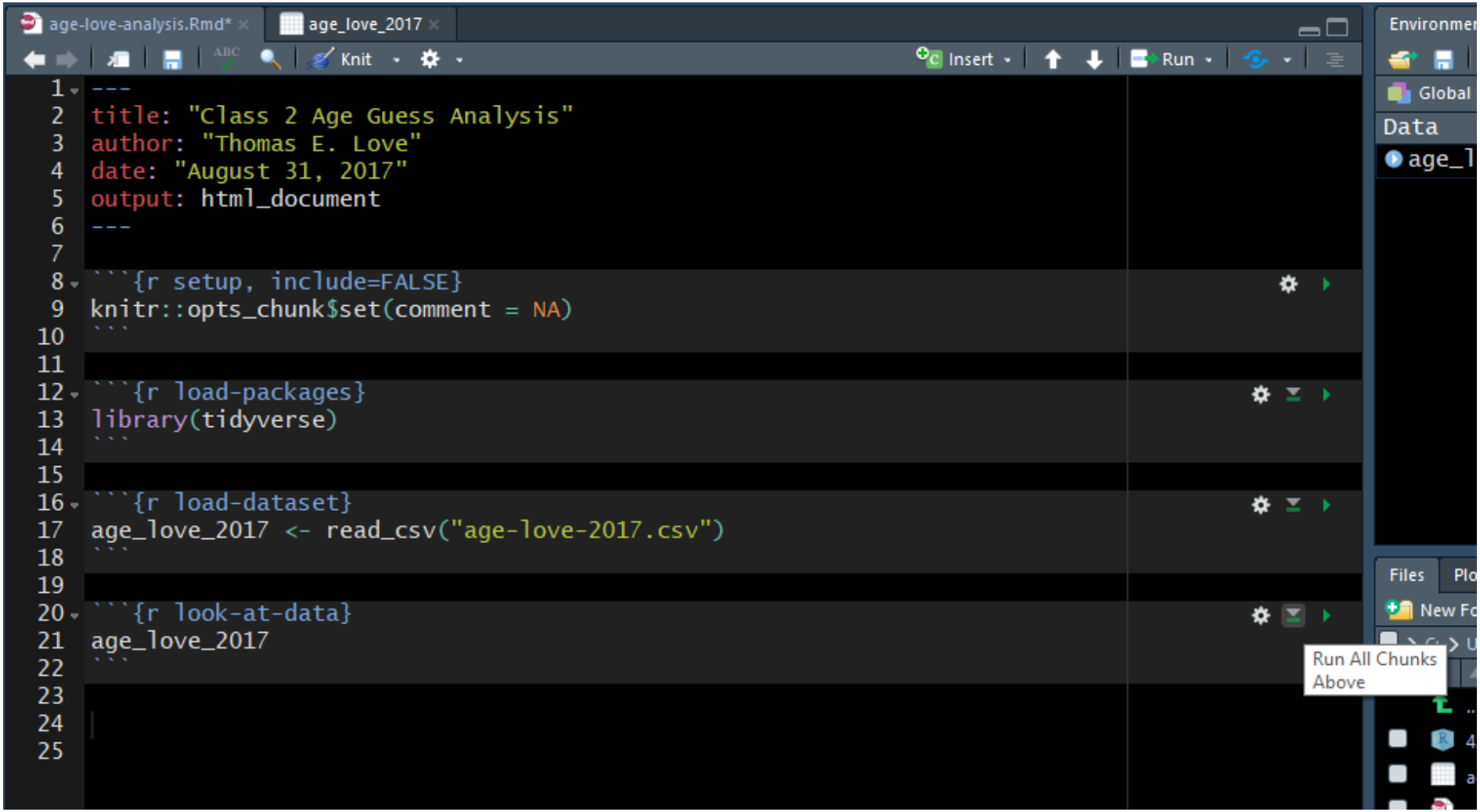
Add data load code to Markdown and also look at the data



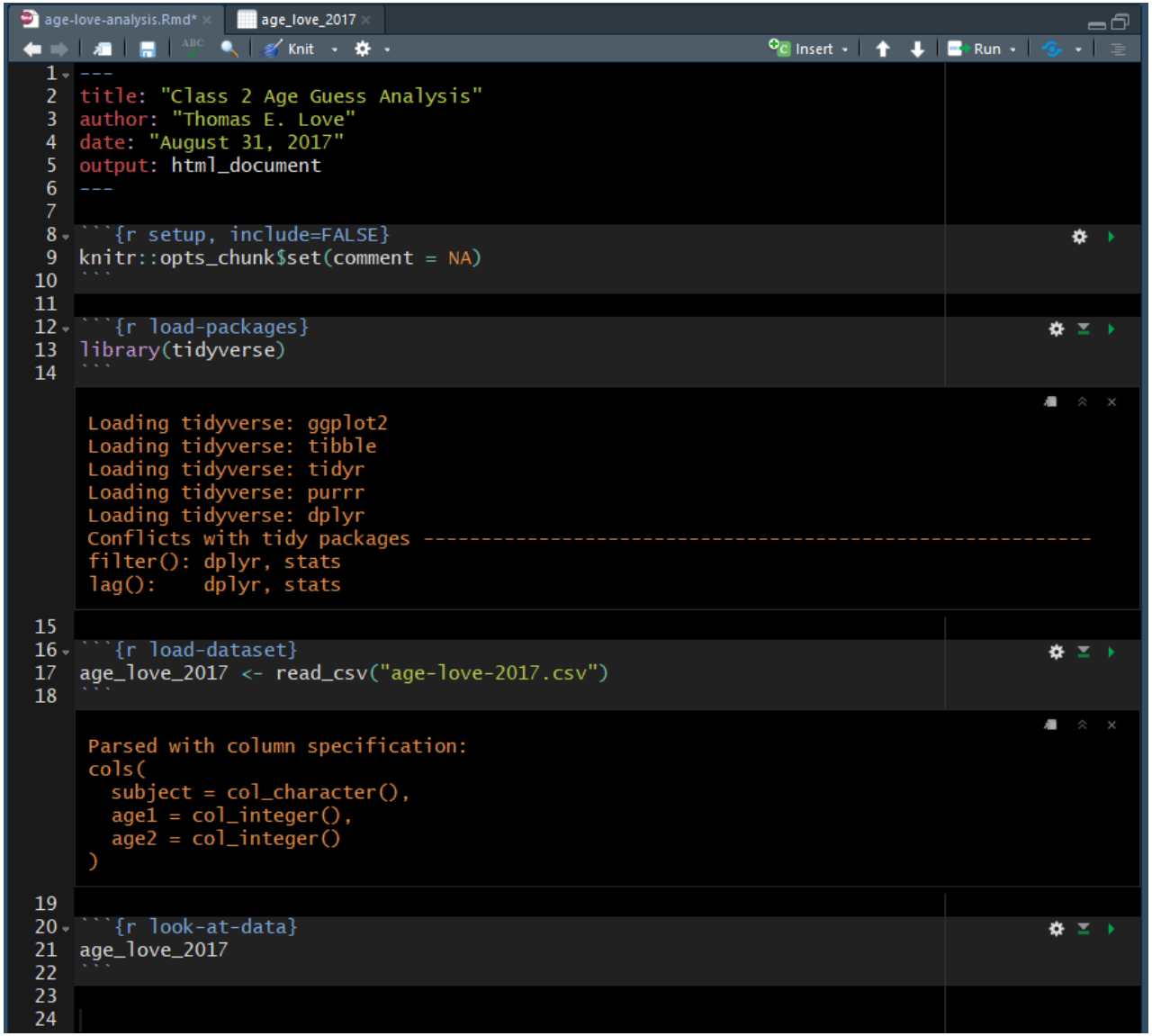
Running the code so we can see results



Or run all of the "chunks" up to a particular point



Running the first three chunks of code



Now, let's look at the data

The age_love_2017 tibble

```{r look-at-data}
age_love_2017
```

subject <chr>	age1 <int>	age2 <int>
S-01	47	42
S-02	52	*NA*
S-03	55	55
S-04	48	48
S-05	50	48
S-06	51	50
S-07	47	49
S-08	45	48
S-09	55	55
S-10	40	48

1-10 of 47 rows

Previous 1 2 3 4 5 Next

Typing in some documentation, mixing text with R code.

````
```{r look-at-data}
age_love_2017
```
````

subject <chr>	age1 <int>	age2 <int>
S-01	47	42
S-02	52	*NA*
S-03	55	55
S-04	48	48
S-05	50	48
S-06	51	50
S-07	47	49
S-08	45	48
S-09	55	55
S-10	40	48

1-10 of 47 rows Previous 1 2 3 4 5 Next

```
Our data set has `r nrow(age_love_2017)` rows, and `r ncol(age_love_2017)` columns.
```

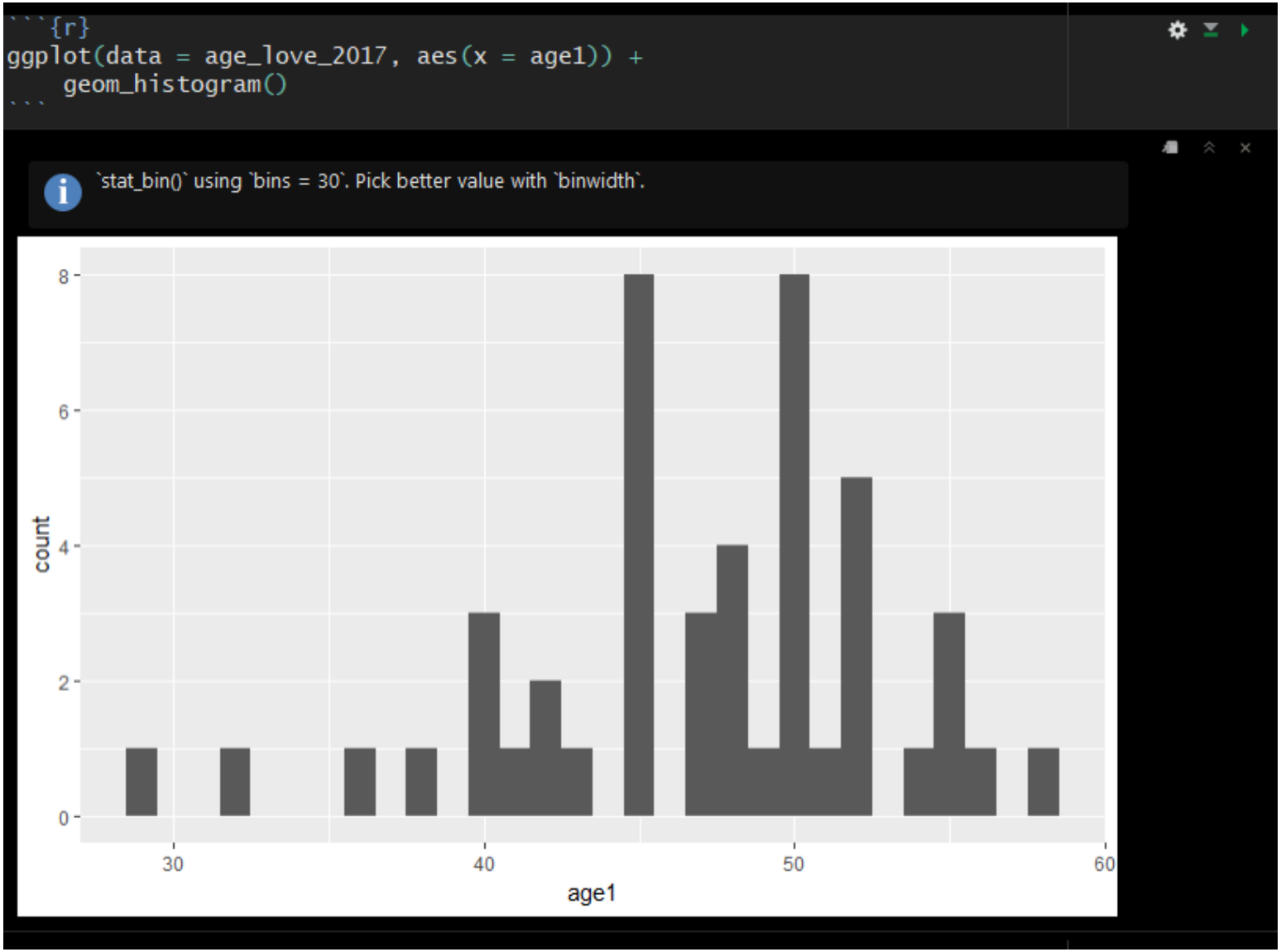
Let's make a histogram

```
# Build a Picture of the First Guesses

```{r}
ggplot(data = age_love_2017, aes(x = age1)) +
    geom_histogram()
```
```

The result. Can we do better?

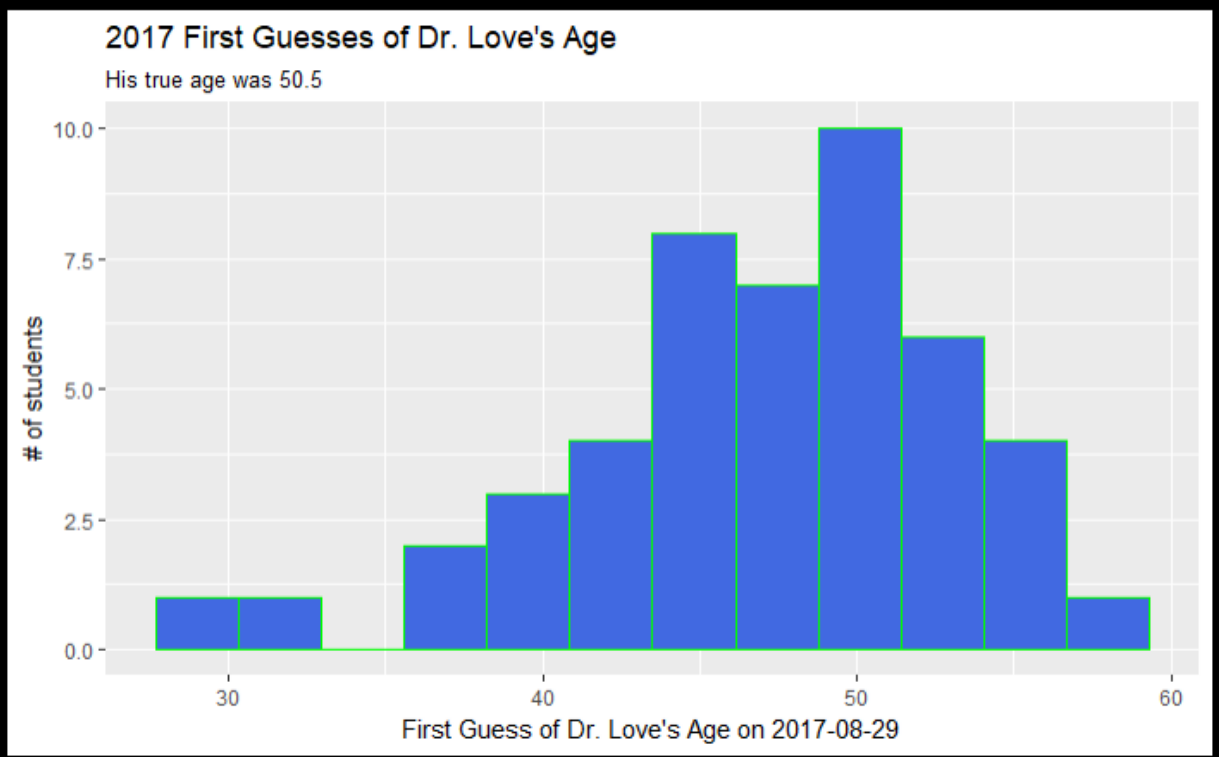
```{r}
ggplot(data = age_love_2017, aes(x = age1)) +
    geom_histogram()
```



A second attempt at the histogram

```
## A nicer picture

```{r second-histogram}
ggplot(data = age_love_2017, aes(x = age1)) +
    geom_histogram(bins = 12, fill = "royalblue", col = "green") +
    labs(x = "First Guess of Dr. Love's Age on 2017-08-29",
         y = "# of students",
         title = "2017 First Guesses of Dr. Love's Age",
         subtitle = "His true age was 50.5")
```
```



A numerical summary of the data

```
44 # A Numerical Summary
45
46 ```{r numerical-summary}
47 summary(age_love_2017)
48 ```
```

```
   subject               age1            age2
 Length:47          Min.   :29.0   Min.   :29.00
 Class :character   1st Qu.:45.0   1st Qu.:48.00
 Mode  :character   Median :48.0   Median :50.00
                    Mean   :47.0   Mean   :48.73
                    3rd Qu.:50.5   3rd Qu.:52.00
                    Max.   :58.0   Max.   :55.00
                                   NA's   :6
```

```
49
50
```

Calculating and Summarizing the Errors

```
49
50 # Better First or Second Guess?
51
52 ```{r calculate-errors}
53 age_love_2017 <- age_love_2017 %>%
54     mutate(error1 = abs(age1 - 50.5),
55            error2 = abs(age2 - 50.5))
56
57 summary(age_love_2017)
58 ```
```

What do these summaries suggest?

```
# Better First or Second Guess?

```{r calculate-errors}
age_love_2017 <- age_love_2017 %>%
    mutate(error1 = abs(age1 - 50.5),
           error2 = abs(age2 - 50.5))

summary(age_love_2017)
```
```

```
   subject              age1            age2           error1          error2
 Length:47          Min.   :29.0   Min.   :29.00   Min.   : 0.500   Min.   : 0.500
 Class :character   1st Qu.:45.0   1st Qu.:48.00   1st Qu.: 1.500   1st Qu.: 0.500
 Mode  :character   Median :48.0   Median :50.00   Median : 4.500   Median : 1.500
                    Mean   :47.0   Mean   :48.73   Mean   : 5.117   Mean   : 3.354
                    3rd Qu.:50.5   3rd Qu.:52.00   3rd Qu.: 6.500   3rd Qu.: 3.500
                    Max.   :58.0   Max.   :55.00   Max.   :21.500   Max.   :21.500
                                   NA's   :6                        NA's   :6
```

Build a scatterplot to compare the errors

```
59
60 # Compare the Guesses
61
62 ```{r guess-1-vs-2}
63 ggplot(data = age_love_2017, aes(x = error1, y = error2)) +
64     geom_point()
65 ```
```

A new scatterplot, with a model for the relationship of age1 to age2

```
66
67 ## Add a Prediction Model?
68
69 ```{r guess1-vs-guess2-with-loess-smooth}
70 ggplot(data = age_love_2017, aes(x = age1, y = age2)) +
71     geom_point(size = 3) +
72     geom_smooth(method = "loess")
73 ```
```

Plot the age1 - age2 differences

```
# Plot the (matched) differences

```{r histogram-of-differences}
ggplot(age_love_2017, aes(x = age1 - age2)) +
    geom_histogram(binwidth = 2,
                   col = "green",
                   fill = "royalblue")
```
```

Numerical summary of the age1 - age2 differences

```
84 # Numerical Summary of the Difference in Ages
85
86 ```{r comparing-guess1-to-guess2}
87 age_love_2017 %$%
88     summary(age1 - age2)
89 ```
90
```

How many people thought I looked younger the second time?

```
92
93 # How many people thought I looked younger in guess 2?
94
95 ```{r count-of-younger-guesses}
96 age_love_2017 %>%
97     count(age1-age2 < 0)
98 ```
```

age1 - age2 < 0 <lgl>	n <int>
FALSE	18
TRUE	23
NA	6

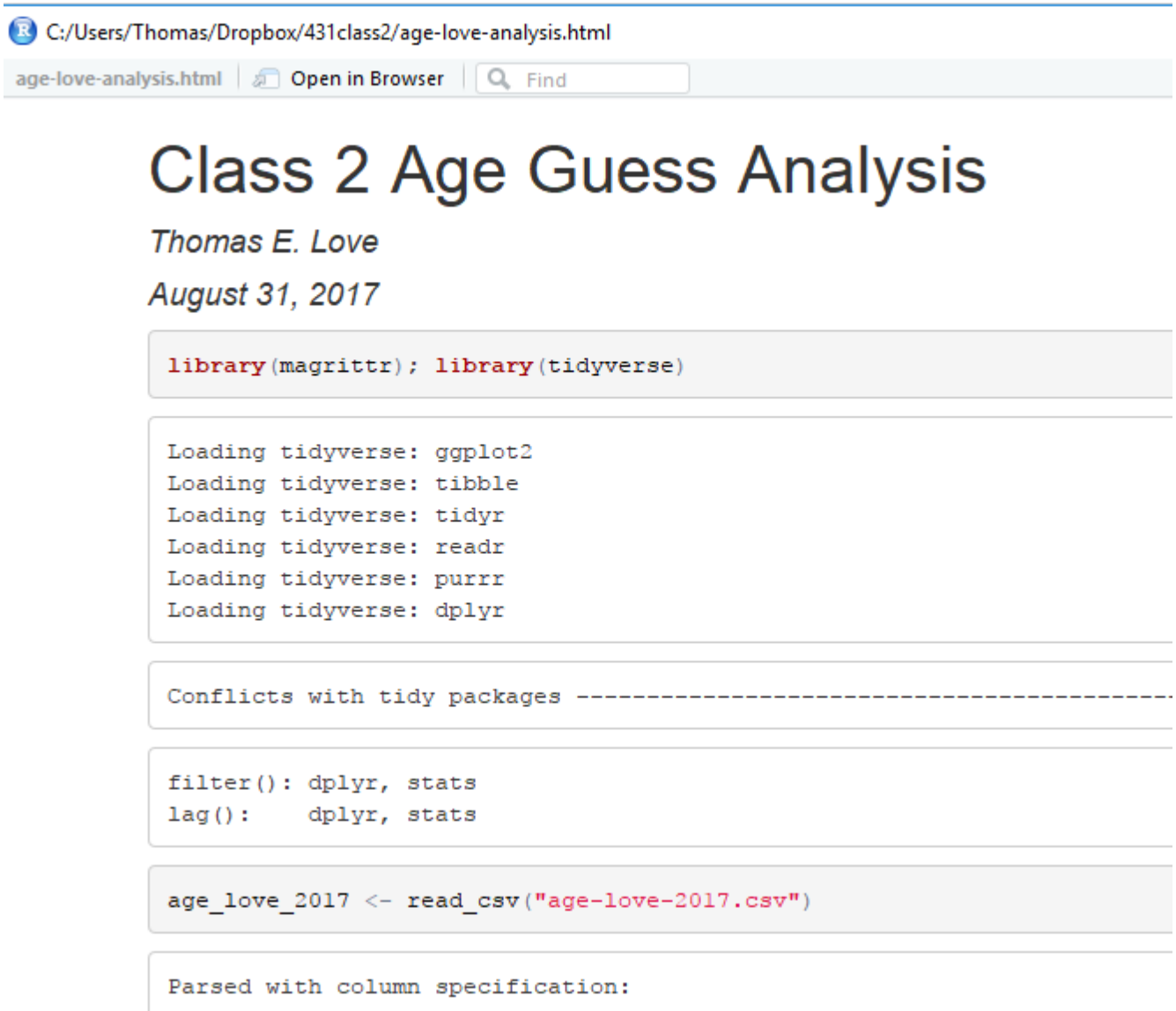
3 rows

```
99
100 I think this is good news! Maybe ...
101
```

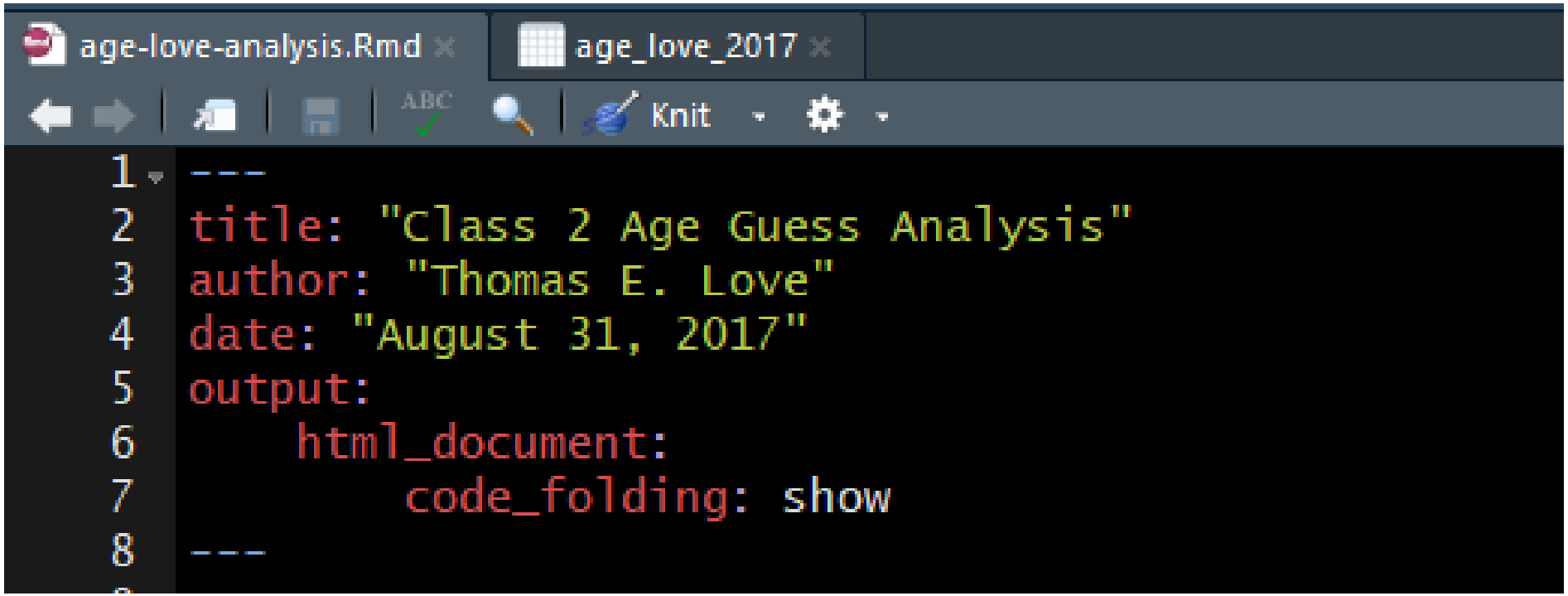
T tests - making a statistical inference

```
97
98  # The Much-Dreaded t test
99
100 ```{r t-test-comparing-guess1-to-guess2}
101 age_love_2017 %$%
102     t.test(age1 - age2)
103 ```
```

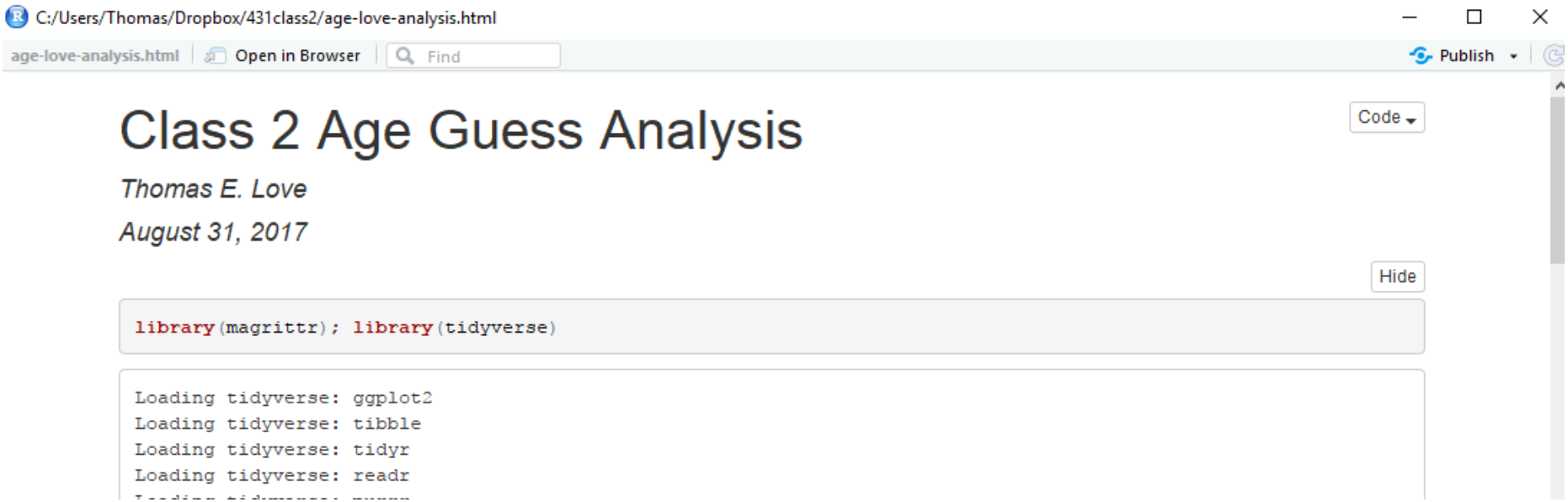
Knit the file into an HTML document



Adjust the YAML to fold the code on demand



New, more final, report



Analyzing the Survey Data - A little challenge

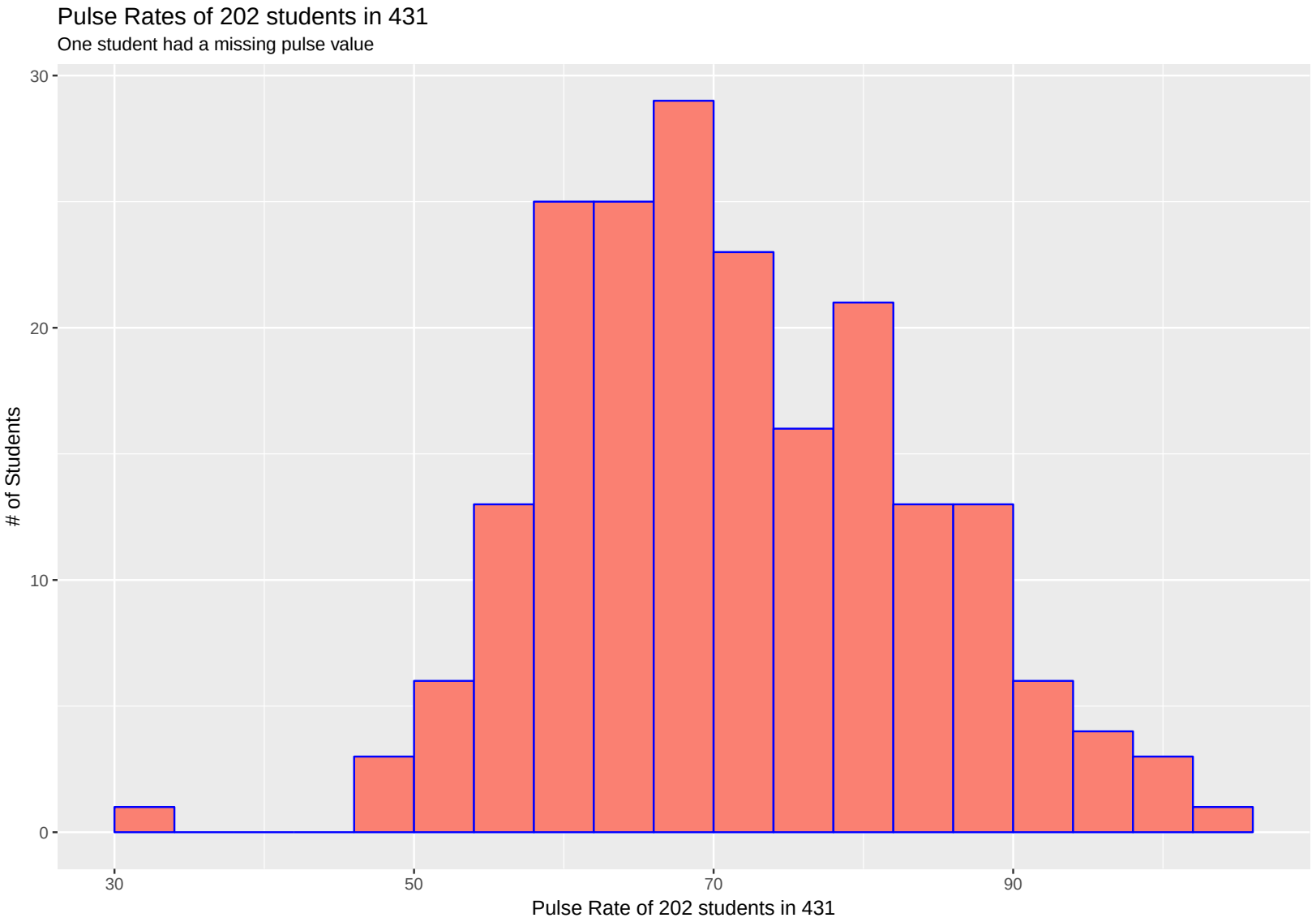
We have data on the site in a file called `surveyday1_2017.csv`. Build a project to study those data.

Put the data in a file called `surv1` in R.

- I'd call my R Markdown file `day1surveyanalysis`

Can you reproduce the following. . .

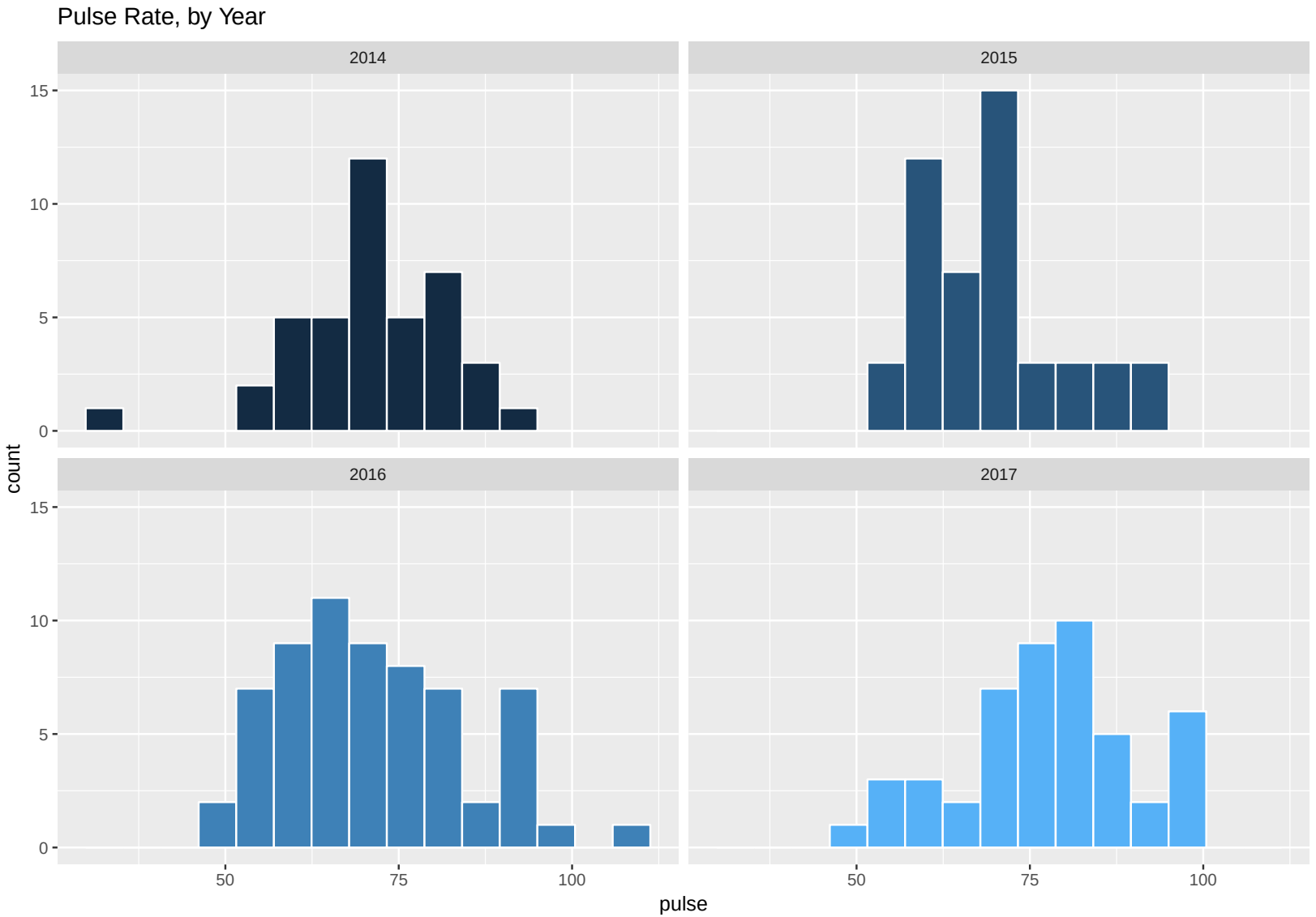
A. That fill color is called *salmon*, I used 20 bins.



Code for Plot A.

```
ggplot(surv1, aes(x = pulse)) +
    geom_histogram(bins = 20,
                   col = "blue", fill = "salmon") +
    labs(x = "Pulse Rate of 202 students in 431",
         y = "# of Students",
         title = "Pulse Rates of 202 students in 431",
         subtitle = "One student had a missing pulse value")
```

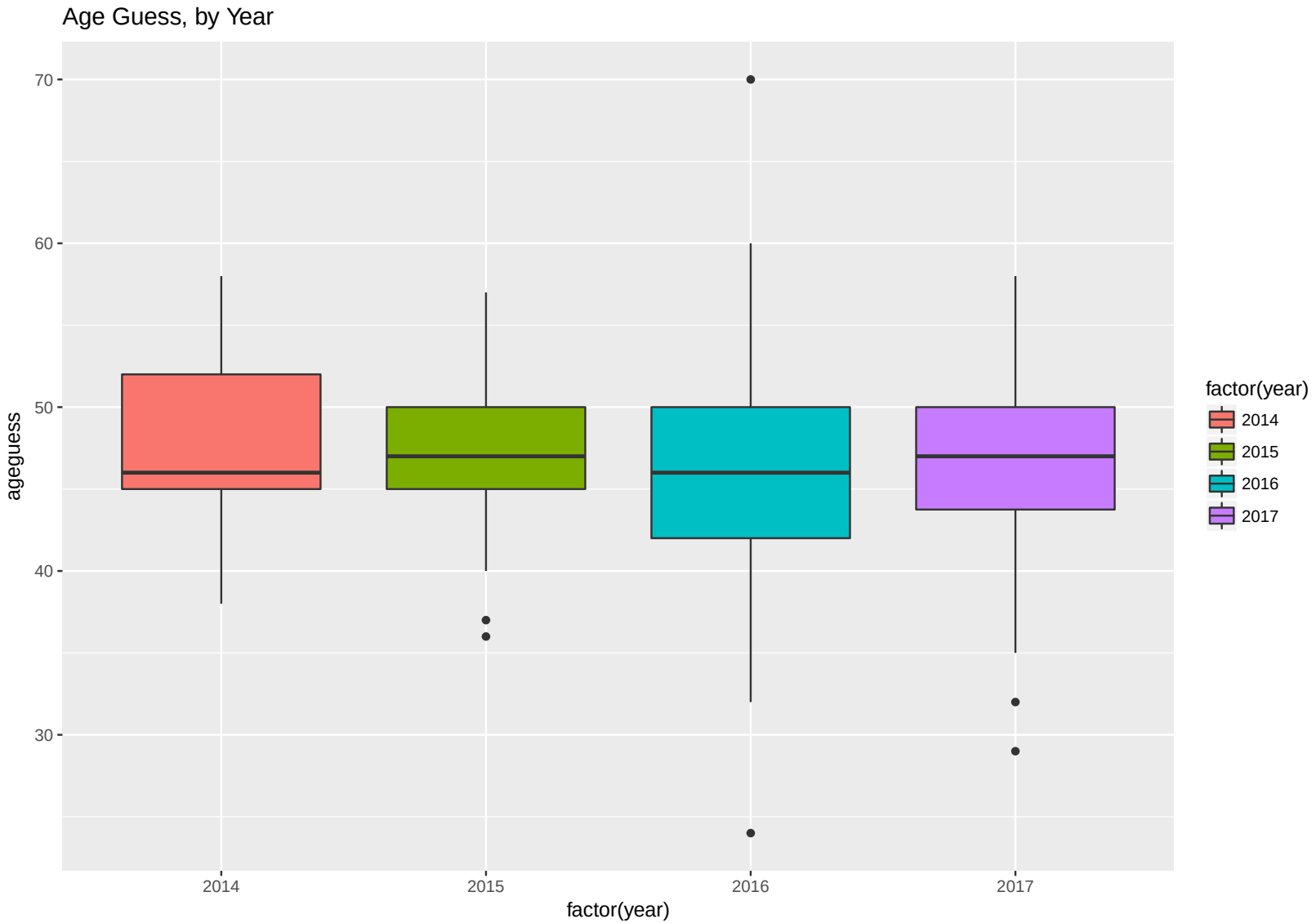
B. Histograms of Pulse Rates, by Year



Code for Plot B.

```
ggplot(surv1, aes(x = pulse, fill = year)) +
    geom_histogram(bins = 15, col = "white") +
    facet_wrap(~ year) +
    guides(fill = FALSE) +
    labs(title = "Pulse Rate, by Year")
```

C.Boxplot of Age Guesses, by Year



Code for Plot C

```
ggplot(surv1, aes(x = factor(year), y = ageguess,
                  fill = factor(year))) +
    geom_boxplot() +
    labs(title = "Age Guess, by Year")
```

Summary Table of Age Guesses, by Year

```
# A tibble: 4 x 5
   year     n     mean       sd median
  <int> <int>    <dbl>    <dbl>  <dbl>
1  2014    42 47.34146 5.213491     46
2  2015    49 47.12245 4.617145     47
3  2016    64 45.96721 6.999922     46
4  2017    48 46.54167 6.146988     47
```

Code for Summary Table

```
surv1 %>%
    group_by(year) %>%
    summarize(n = n(),
              mean = mean(ageguess, na.rm=TRUE),
              sd = sd(ageguess, na.rm=TRUE),
              median = median(ageguess, na.rm=TRUE)
              )
```

What's coming up?

- Running more involved analyses in R and R Studio
- More on exploratory data analysis for distributions and associations
- Discussion of the project requirements is coming next week
- Never too early to get started
    - Read Leek Chapters 5, 9, 10 and 13 (about 30 pages in total) by 2017-09-07 (Class 4)
    - Read Silver Introduction and Chapter 1 (about 50 denser pages) by 2017-09-12 (Class 5)
    - Assignment 1 is due 2017-09-15 at noon