

431 Class 02

thomaseLove.github.io/431

2021-08-26

Instructions for the Quick Survey

Please read these instructions carefully before writing anything down.

- 1 Introduce yourself to someone that you don't know.
- 2 Record the survey answers **for that other person**, while they record your responses.
- 3 Be sure to complete all 15 questions (both sides of the paper).
- 4 When you are finished, thank your partner and raise your hand.
Someone will come to collect your survey.

Regarding Question 4 on the Quick Survey, Professor Love is the large fellow standing in the front of the room.

Today's Agenda

- Data Structures and Variables
- Evaluating some of the Survey variables
- Using R to look at a little data
- Guessing Dr. Love's Age

Class 03 will be a Pre-Recorded Demonstration

We won't meet as a group for Class 03. Instead, I will provide a recording of myself using R and R Studio to:

- create a Project in RStudio
- create an R Markdown document to obtain and describe results
- ingest a data set from the wild into an R data frame
- manage that data frame so that it becomes a tidy tibble
- use that tibble to learn about the data by iterating through
 - visualizing the data
 - transforming the data
 - modeling the data (with a few simple summaries)
- interspersing code and text in my R Markdown document
- “knitting” my R Markdown document into an attractive HTML result

All materials will be posted as soon as possible to the Class 03 page.

Today's Package Loading

```
library(janitor)  
library(googleheets4)  
library(patchwork)  
library(tidyverse)
```

If you actually run this, you will get some messages which we will suppress and ignore today.

Thinking about The Quick Survey

Chatfield's Six Rules for Data Analysis

- 1 Do not attempt to analyze the data until you understand what is being measured and why.
- 2 Find out how the data were collected.
- 3 Look at the structure of the data.
- 4 Carefully examine the data in an exploratory way, before attempting a more sophisticated analysis.
- 5 Use your common sense at all times.
- 6 Report the results in a clear, self-explanatory way.

Chatfield, Chris (1996) *Problem Solving: A Statistician's Guide*, 2nd ed.

Types of Data

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

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?

Our Quick Survey

431 Quick Survey for 2021: Class 02 (15 Questions)

Please introduce yourself to someone near you, 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!

- Do you wear corrective lenses (contacts or glasses)? (Yes or No) _____
- Is English your *most comfortable* language? (Yes or No) _____
- 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

- How tall do you think Dr. Love is? (Please indicate units.) _____
- Do you smoke? Fill in the appropriate circle:

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

6. Please indicate which hand you use for each of the following activities by putting + 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 = _____ $\frac{\text{Right} - \text{Left}}{\text{Right} + \text{Left}} = \text{_____}$

431 Quick Survey for 2021: Class 02 (15 Questions)

- 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

- 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			

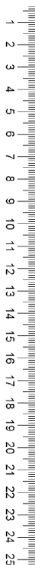
- What is your height (indicate units of measurement): _____

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.

- What is your favorite color? _____

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

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



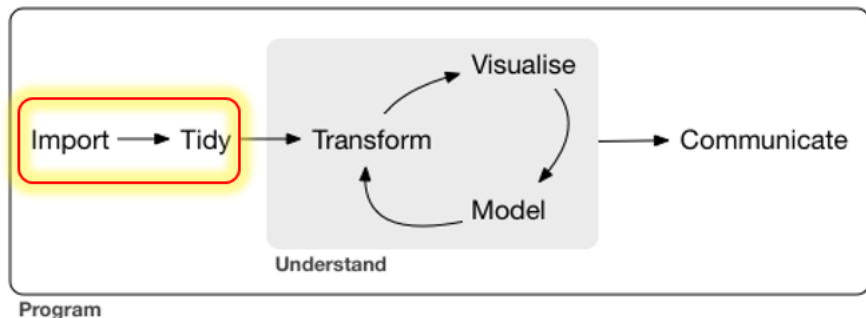
Evaluating Some Quick Survey variables

- 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 Statistical thinking in your future **career**? (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*?

Importing and Tidying Data



Ingesting the Quick Surveys

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	student	sex	glasses	english	statsofar	ageguess	smoke	h.left	h.right	handedness	statfuture	haircut	lecture	alone	height.in	hand.s
2	201901	NA	y	y	6	42	1	1	19							
3	201902	NA	y	y	7	53	1	19	10							
4	201903	NA	y	y	4	45	1	0	10							
5	201904	NA	y	y	7	45	1	16	10							
6	201905	NA	y	y	6	42	1	2	16							
7	201906	NA	y	y	7	50	1	10	0							
8	201907	NA	y	y	5	56	1	1	13							
9	201908	NA	n	n	6	50	1	0	10							
10	201909	NA	n	y	6	52	1	0	17							
11	201910	NA	n	y	4	42	1	18	10							
12	201911	NA	n	n	5	43	1	5	13							
13	201912	NA	y	y	5	52	1	1	13							
14	201913	NA	y	y	7	50	2	1	19							
15	201914	NA	y	y	4	50	1	1	9							

165 cm to inches - Google Search

google.com/search?q=165+cm+to+inches&oq=165+cm+&aqs=chrome.20j69j57j0l4...

Apps ★ Bookmarks NY Times 538 NYT Crossword Slate The Athletic sAm

Google 165 cm to inches

Q All Shopping News Images Videos More Setting

About 20,900,000 results (0.56 seconds)

Length

165 = 64.9606

Centimeter Inch

Formula divide the length value by 2.54

The Quick Survey

315 people before you have taken (essentially) this same survey in the same way.

Fall	2019	2018	2017	2016	2015	2014	Total
<i>n</i>	61	51	48	64	49	42	315

Question

About how many of those 315 surveys caused *no problems* in recording responses?

The 15 Survey Items

#	Topic	#	Topic
Q1	glasses	Q9	lectures_vs_activities
Q2	english	Q10	projects_alone
Q3	stats_so_far	Q11	height
Q4	guess_TL_ht	Q12	hand_span
Q5	smoke	Q13	color
Q6	handedness	Q14	sleep
Q7	stats_future	Q15	pulse_rate
Q8	haircut	-	-

- At one time, I asked about sex rather than glasses.
- In prior years, people guessed my age, rather than height here.
- Sometimes, I've asked for a 30-second pulse check, then doubled.

Response to the Question I asked

About how many of those 315 surveys caused *no problems* in recording responses?

- Guesses?

Response to the Question I asked

About how many of those 315 surveys caused *no problems* in recording responses?

- Guesses?
- 110/315 (35%) caused no problems.

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.

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

50ish years.

What should we do in these cases?

English best language?

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

TEL Decision: Yes

1. What is your *gender*? (Male or Female) _____
2. Is English your *most comfortable* language? (Yes or No) _____

TEL Decision: NA

Is English your *most comfortable* language? (Yes or No) maybe

TEL decision: NA

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
11. What is your height (indicate units of measurement): 5'7"
11. What is your height (indicate units of measurement): 155

Handedness Scale (2016-21 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

Favorite color

13. What is your favorite color? depends

NA

13. What is your favorite color? orange

orange

13. What is your favorite color? Blue, Brown

13. What is your favorite color? N/A

Following the Rules?

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

75 beats/minute.

2019 pulse responses, sorted ($n = 61$, 1 NA)

33	46	48	56	60	60	3		3
62	63	65	65	66	66	4		68
68	68	68	69	70	70	5		6
70	70	70	70	70	70	6		002355668889
71	72	72	74	74	74	7		00000000122444445666888
74	74	75	76	76	76	8		000012445668
78	78	78	80	80	80	9		000046
80	81	82	84	84	85	10		44
86	86	88	90	90	90	11		0
90	94	96	104	104	110			

Stem and Leaf: Pulse Rates, 2014-2019

```
> stem(survey1$pulse)
```

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

```
3 | 03
3 |
4 |
4 | 688
5 | 00022244444
5 | 566666666667888889
6 | 0000000000000000000022222222223344444444444444
6 | 5555556666666666666666666666666666888888888888888888899
7 | 00000000000000000000000000000000001222222222222222244444444444444444
7 | 55555566666666667888888888888888
8 | 00000000000000000000000000000000001222222444444444444
8 | 556666666666666666666666666666666688888888
9 | 00000000000000000000000000000000001222224444
9 | 5668888
10 | 0000444
10 | 6
11 | 0
```

(Thanks, John **Tukey**)

Garbage in, garbage out . . .

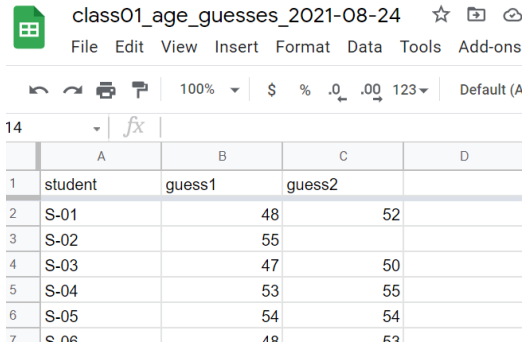


“Data don’t make any sense,
we will have to resort to statistics.”

Guessing My Age (from Class 01)

From our Shared Google Drive

I've placed a Google Sheet called `class01_age_guesses_2021-08-24` on our Shared Google Drive. Remember that you have to log into Google via CWRU to see the Drive I've shared with you.



The screenshot shows a Google Sheet interface. At the top, the title bar reads "class01_age_guesses_2021-08-24" with icons for star, share, and link. Below the title bar is a menu bar with "File", "Edit", "View", "Insert", "Format", "Data", "Tools", and "Add-ons". A toolbar contains various icons for undo, redo, print, and zoom, along with a "100%" zoom level and currency symbols. The spreadsheet grid has a header row with columns A, B, C, and D. Row 1 contains the labels "student", "guess1", and "guess2". Rows 2 through 7 contain student IDs (S-01 to S-06) and their corresponding age guesses (48, 55, 47, 53, 54, 49).

	A	B	C	D
1	student	guess1	guess2	
2	S-01	48	52	
3	S-02	55		
4	S-03	47	50	
5	S-04	53	55	
6	S-05	54	54	
7	S-06	49	53	

Reading from our Shared Google Drive

We'll use the `read_sheet` function from the `googlesheets4` package to read in data from a Google Sheet. My first step is to copy the URL from the Google Sheet into a temporary object I'll call `temp_url`.

```
temp_url <- "https://docs.google.com/spreadsheets/d/1Mgu_Xj0A8"
```

Then I'll ask R to read in the data from the sheet to a new data frame (technically a tibble) called `age_guess`.

```
age_guess <- read_sheet(temp_url)
```

When you do this the first time, R will ask you to verify some things in the browser and allow the browser to pull down the sheet. Let it do so.

What is in the age_guess data frame?

```
age_guess
```

```
# A tibble: 59 x 3
```

	student	guess1	guess2
	<chr>	<dbl>	<dbl>
1	S-01	48	52
2	S-02	55	NA
3	S-03	47	50
4	S-04	53	55
5	S-05	54	54
6	S-06	48	53
7	S-07	53	56
8	S-08	55	NA
9	S-09	52	52
10	S-10	55	54

```
# ... with 49 more rows
```

What do the guess1 values look like?

```
age_guess %>% select(guess1) %>% arrange(guess1)
```

```
# A tibble: 59 x 1
```

```
  guess1
```

```
  <dbl>
```

```
1      30
```

```
2      43
```

```
3      43
```

```
4      47
```

```
5      47
```

```
6      47
```

```
7      48
```

```
8      48
```

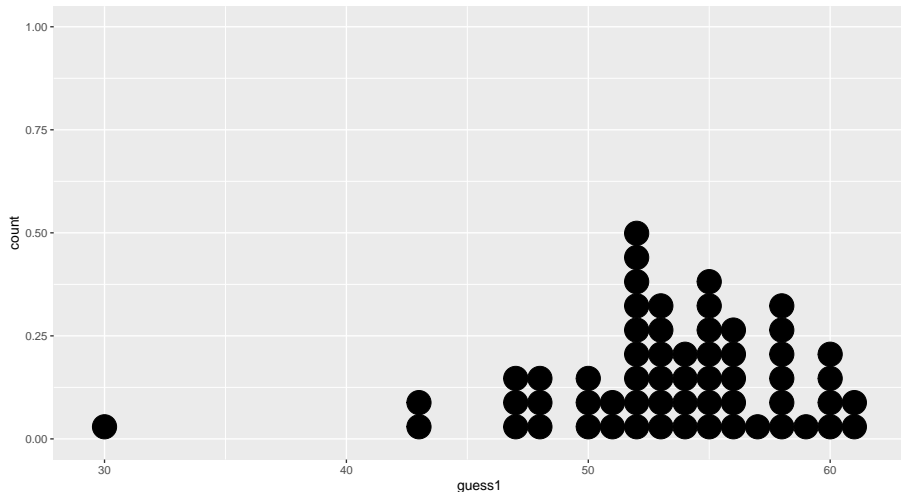
```
9      48
```

```
10     50
```

```
# ... with 49 more rows
```

Plot the guess1 values?

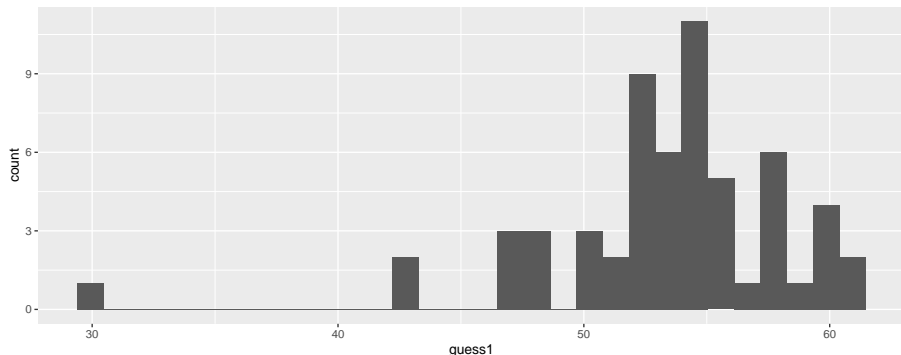
```
ggplot(age_guess, aes(x = guess1)) +  
  geom_dotplot(binwidth = 1)
```



Can we make a histogram?

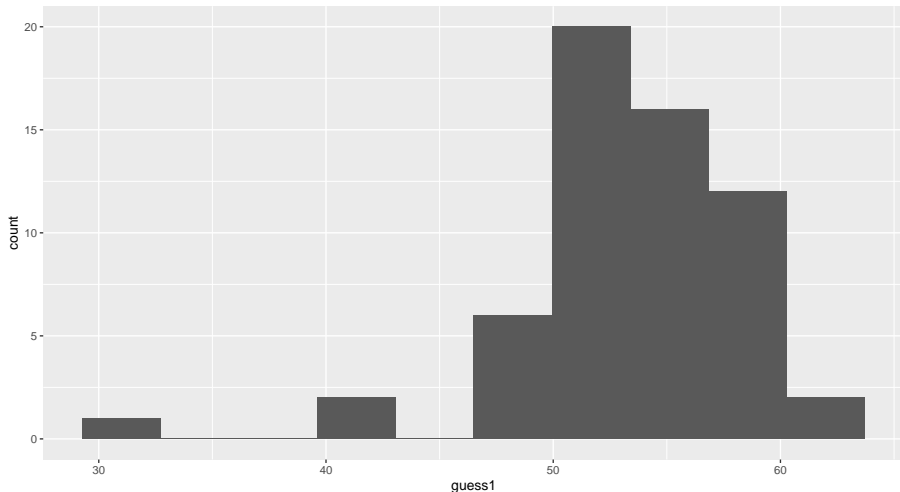
```
ggplot(age_guess, aes(x = guess1)) +  
  geom_histogram()
```

``stat_bin()`` using ``bins = 30``. Pick better value with ``binwidth``.



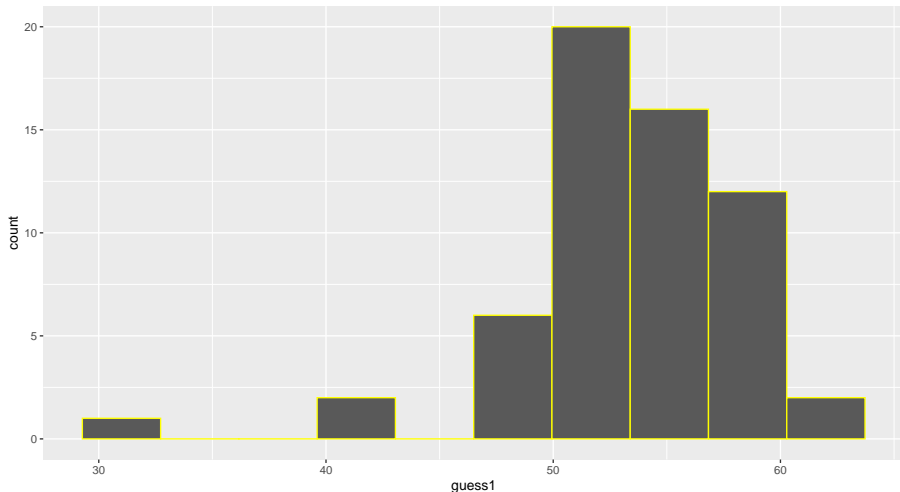
Improving the Histogram, 1

```
ggplot(age_guess, aes(x = guess1)) +  
  geom_histogram(bins = 10)
```



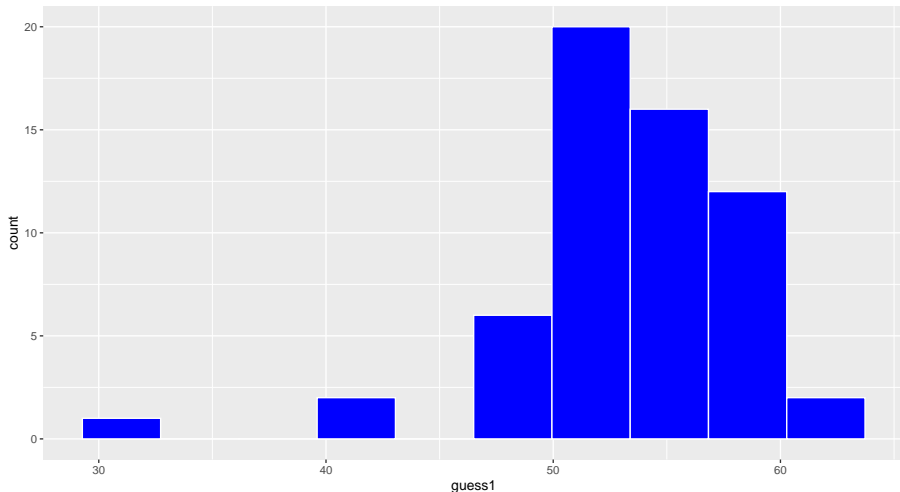
Improving the Histogram, 2

```
ggplot(age_guess, aes(x = guess1)) +  
  geom_histogram(bins = 10, col = "yellow")
```



Improving the Histogram, 3

```
ggplot(age_guess, aes(x = guess1)) +  
  geom_histogram(bins = 10, col = "white", fill = "blue")
```



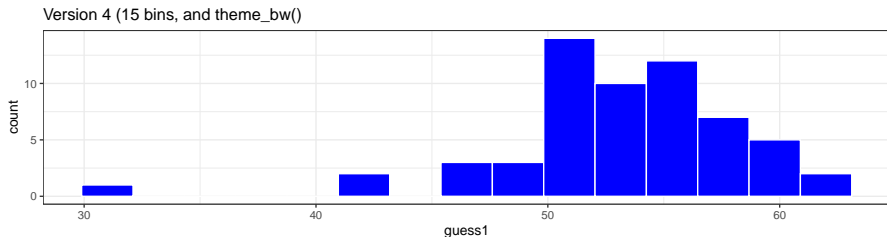
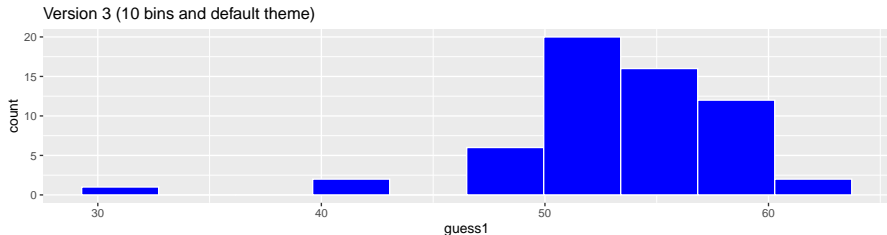
Improving the Histogram, 4 (code only)

```
ggplot(age_guess, aes(x = guess1)) +  
  geom_histogram(bins = 15, col = "white", fill = "blue") +  
  theme_bw()
```

- We've changed the theme to theme_bw
- We've increased the number of bins from 10 to 15.

Results of Version 3 and Version 4 shown on the next slide.

Results for Versions 3 and 4

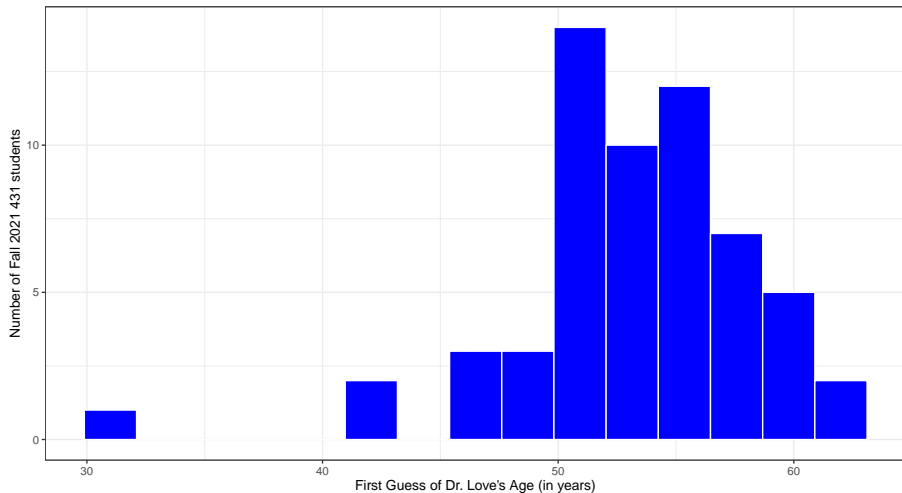


Improving the Histogram, 5 (code only)

```
ggplot(age_guess, aes(x = guess1)) +  
  geom_histogram(bins = 15, col = "white", fill = "blue") +  
  theme_bw() +  
  labs(x = "First Guess of Dr. Love's Age (in years)",  
       y = "Number of Fall 2021 431 students")
```

Here we add axis labels. Result on next slide.

Results of Version 5



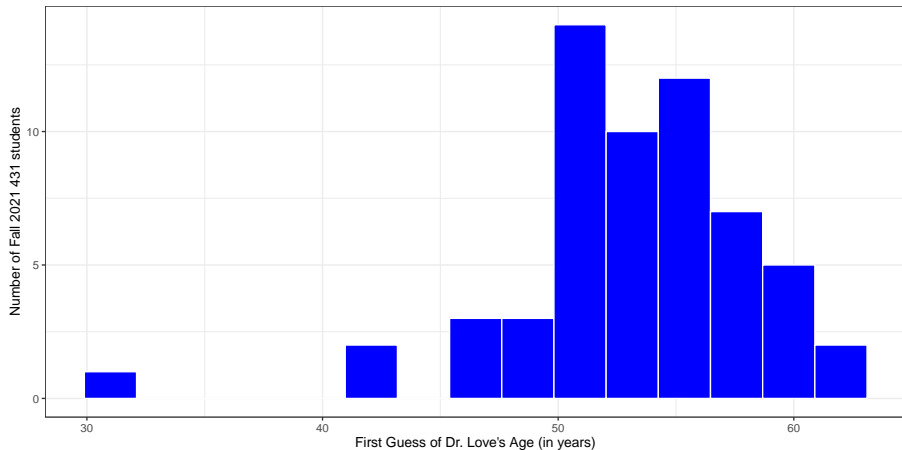
Version 6 adds title and subtitle

```
ggplot(age_guess, aes(x = guess1)) +  
  geom_histogram(bins = 15, col = "white", fill = "blue") +  
  theme_bw() +  
  labs(x = "First Guess of Dr. Love's Age (in years)",  
       y = "Number of Fall 2021 431 students",  
       title = "Most First Guesses were pretty close",  
       subtitle = "Dr. Love's actual age was 54.5")
```


Result of Version 6 code

Most First Guesses were pretty close

Dr. Love's actual age was 54.5



Improving the Histogram, 7

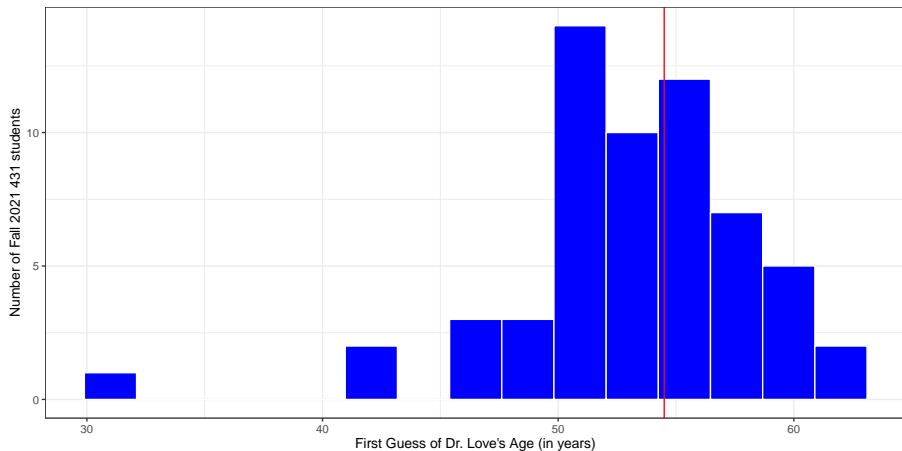
Let's add a vertical line at 54.5 years to show my actual age.

```
ggplot(age_guess, aes(x = guess1)) +  
  geom_histogram(bins = 15, col = "white", fill = "blue") +  
  geom_vline(aes(xintercept = 54.5), col = "red") +  
  theme_bw() +  
  labs(x = "First Guess of Dr. Love's Age (in years)",  
       y = "Number of Fall 2021 431 students",  
       title = "Most First Guesses were pretty close",  
       subtitle = "Dr. Love's actual age was 54.5")
```

Result from Improving the Histogram, 7

Most First Guesses were pretty close

Dr. Love's actual age was 54.5



A Quick Numerical Summary of the Data Frame

```
summary(age_guess)
```

student	guess1	guess2
Length:59	Min. :30.00	Min. :26.00
Class :character	1st Qu.:52.00	1st Qu.:53.00
Mode :character	Median :54.00	Median :55.00
	Mean :53.32	Mean :54.56
	3rd Qu.:56.00	3rd Qu.:57.00
	Max. :61.00	Max. :65.00
		NA's :2

- Was the average guess closer to my actual age (54.5) on the first or second guess?

A Quick Numerical Summary of the Data Frame

```
summary(age_guess)
```

student	guess1	guess2
Length:59	Min. :30.00	Min. :26.00
Class :character	1st Qu.:52.00	1st Qu.:53.00
Mode :character	Median :54.00	Median :55.00
	Mean :53.32	Mean :54.56
	3rd Qu.:56.00	3rd Qu.:57.00
	Max. :61.00	Max. :65.00
		NA's :2

- Was the average guess closer to my actual age (54.5) on the first or second guess?
- What was the range of first guesses? Second guesses?

A Quick Numerical Summary of the Data Frame

```
summary(age_guess)
```

student	guess1	guess2
Length:59	Min. :30.00	Min. :26.00
Class :character	1st Qu.:52.00	1st Qu.:53.00
Mode :character	Median :54.00	Median :55.00
	Mean :53.32	Mean :54.56
	3rd Qu.:56.00	3rd Qu.:57.00
	Max. :61.00	Max. :65.00
		NA's :2

- Was the average guess closer to my actual age (54.5) on the first or second guess?
- What was the range of first guesses? Second guesses?
- What does the NA's : 2 mean in guess2?

A Quick Numerical Summary of the Data Frame

```
summary(age_guess)
```

student	guess1	guess2
Length:59	Min. :30.00	Min. :26.00
Class :character	1st Qu.:52.00	1st Qu.:53.00
Mode :character	Median :54.00	Median :55.00
	Mean :53.32	Mean :54.56
	3rd Qu.:56.00	3rd Qu.:57.00
	Max. :61.00	Max. :65.00
		NA's :2

- Was the average guess closer to my actual age (54.5) on the first or second guess?
- What was the range of first guesses? Second guesses?
- What does the NA's : 2 mean in guess2?
- Why is student not summarized any further?

Some additional summaries

```
mosaic::favstats(~ guess1, data = age_guess)
```

min	Q1	median	Q3	max	mean	sd	n	missing
30	52	54	56	61	53.32203	5.174342	59	0

```
mosaic::favstats(~ guess2, data = age_guess)
```

min	Q1	median	Q3	max	mean	sd	n	missing
26	53	55	57	65	54.5614	5.067465	57	2

How many first guesses were between 53 and 56?

```
age_guess %>% count(guess1 >= 53 & guess1 <= 56)
```

```
# A tibble: 2 x 2
```

	<code>`guess1 >= 53 & guess1 <= 56`</code>	<code>n</code>
	<code><lgl></code>	<code><int></code>
1	FALSE	37
2	TRUE	22

How many second guesses were between 53 and 56?

```
age_guess %>% count(guess2 >= 53 & guess1 <= 56)
```

```
# A tibble: 3 x 2
  `guess2 >= 53 & guess1 <= 56`      n
  <lgl>                        <int>
1 FALSE                        25
2 TRUE                         32
3 NA                           2
```

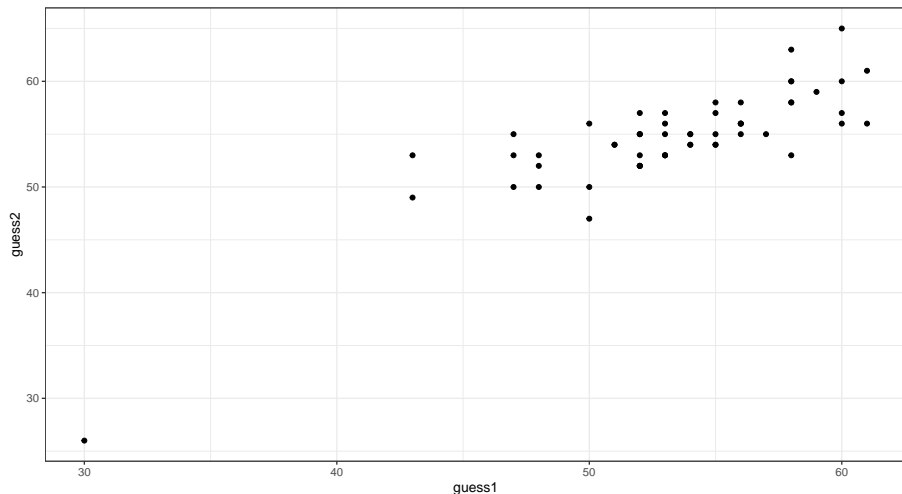
Comparing First Guess to Second Guess

Here's the code. Resulting scatterplot on next slide.

```
ggplot(data = age_guess, aes(x = guess1, y = guess2)) +  
  geom_point() + theme_bw()
```

Comparing First Guess to Second Guess

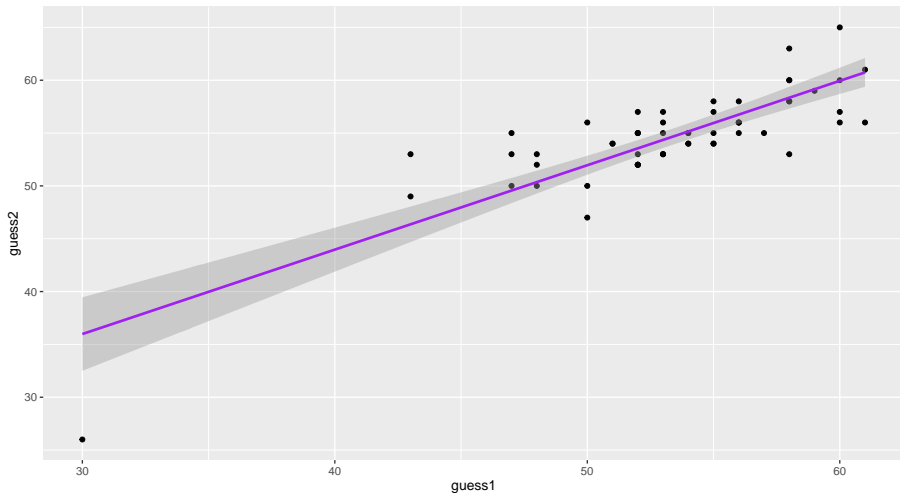
Warning: Removed 2 rows containing missing values
(geom_point).



Filter to complete cases, and add regression line

```
age_guess %>%  
  filter(complete.cases(guess1, guess2)) %>%  
  ggplot(data = ., aes(x = guess1, y = guess2)) +  
  geom_point() +  
  geom_smooth(method = "lm", formula = y ~ x,  
              col = "purple")
```

Resulting Scatterplot



What is that regression line?

```
lm(guess2 ~ guess1, data = age_guess)
```

Call:

```
lm(formula = guess2 ~ guess1, data = age_guess)
```

Coefficients:

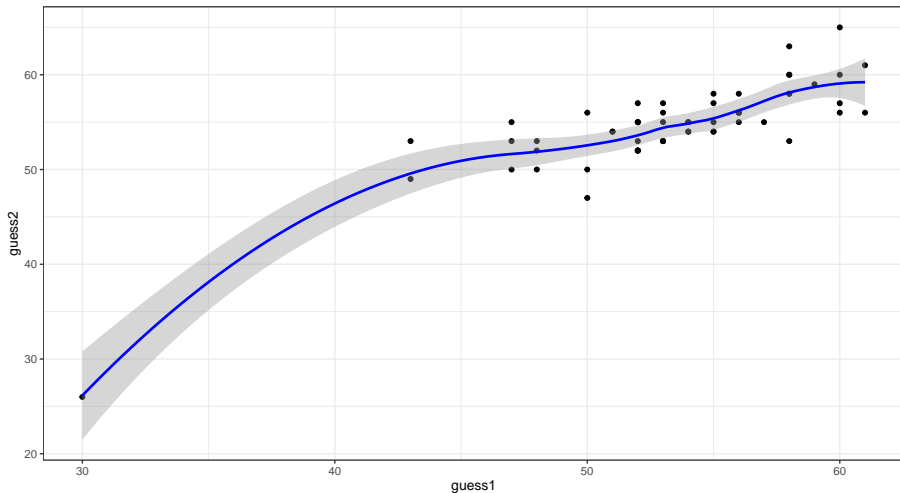
(Intercept)	guess1
12.0219	0.7987

- Note that `lm` filters to complete cases by default.

How about a loess smooth curve, instead?

```
age_guess %>%  
  filter(complete.cases(guess1, guess2)) %>%  
  ggplot(data = ., aes(x = guess1, y = guess2)) +  
  geom_point() +  
  geom_smooth(method = "loess", formula = y ~ x,  
              col = "blue") +  
  theme_bw()
```

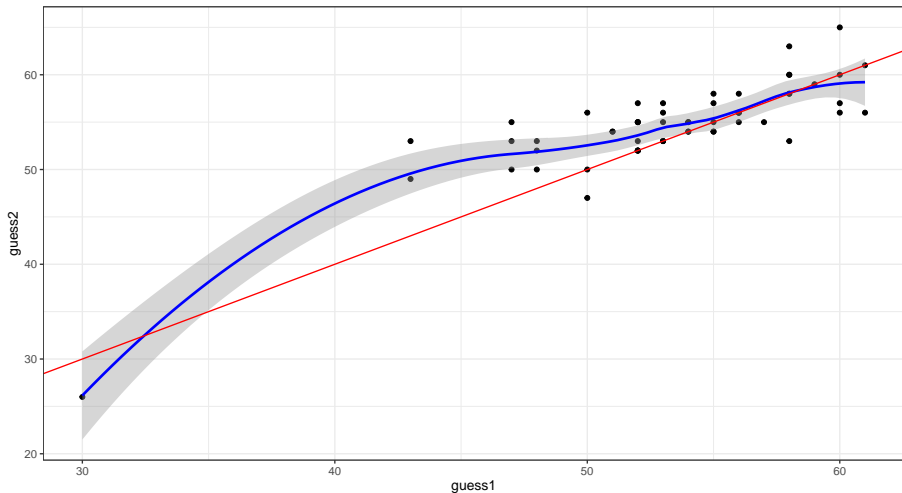

Scatterplot with loess smooth



Add $y = x$ line (no change in guess)?

```
age_guess %>%  
  filter(complete.cases(guess1, guess2)) %>%  
  ggplot(data = ., aes(x = guess1, y = guess2)) +  
  geom_point() +  
  geom_smooth(method = "loess", formula = y ~ x,  
              col = "blue") +  
  geom_abline(intercept = 0, slope = 1, col = "red") +  
  theme_bw()
```

Blue smooth and Red line at $y = x$



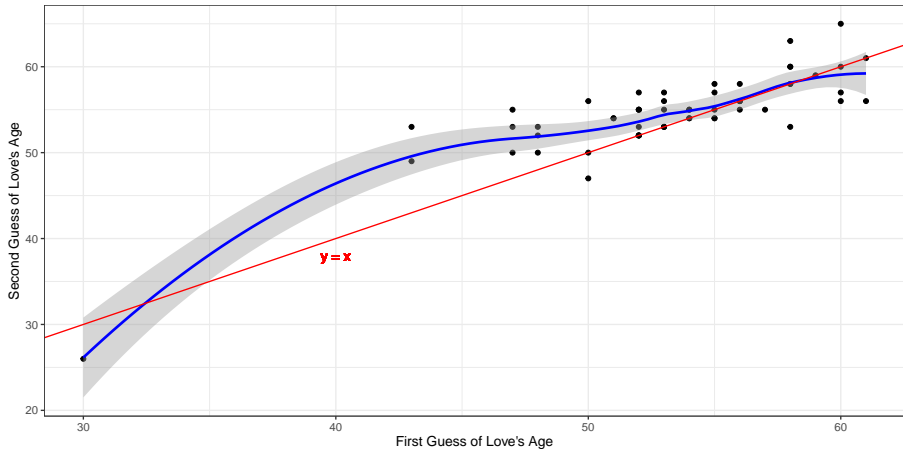
With Better Labels

```
age_guess %>%
  filter(complete.cases(guess1, guess2)) %>%
  ggplot(data = ., aes(x = guess1, y = guess2)) +
  geom_point() +
  geom_smooth(method = "loess", formula = y ~ x,
              col = "blue") +
  geom_abline(intercept = 0, slope = 1, col = "red") +
  geom_text(x = 40, y = 38, label = "y = x", col = "red") +
  labs(x = "First Guess of Love's Age",
       y = "Second Guess of Love's Age",
       title = "Comparing 2021 Age Guesses",
       subtitle = "Love's actual age = 54.5") +
  theme_bw()
```

The Resulting Plot

Comparing 2021 Age Guesses

Love's actual age = 54.5



Decreased / Stayed the Same / Increased

```
age_guess %>% count(sign(guess2 - guess1))
```

```
# A tibble: 4 x 2
```

	`sign(guess2 - guess1)`	n
	<dbl>	<int>
1	-1	10
2	0	19
3	1	28
4	NA	2

How much did people change their guesses?

```
age_guess <- age_guess %>%  
  mutate(change = guess2 - guess1)  
  
summary(age_guess$change)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-5.000	0.000	0.000	1.298	3.000	10.000
NA's					
2					

Table (via `taby1`) of guess changes

```
age_guess %>%  
  tabyl(change) %>%  
  adorn_pct_formatting()
```

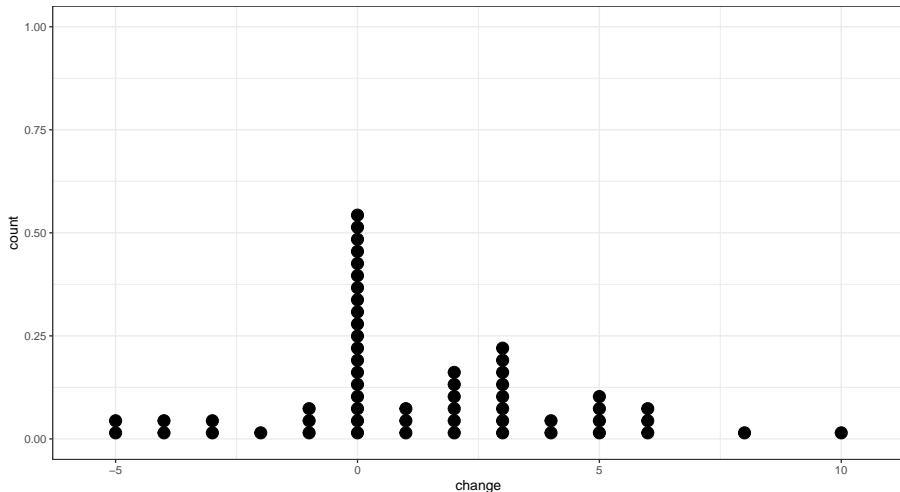
Results on next slide.

Table (via `taby1`) of guess changes

change	n	percent	valid_percent
-5	2	3.4%	3.5%
-4	2	3.4%	3.5%
-3	2	3.4%	3.5%
-2	1	1.7%	1.8%
-1	3	5.1%	5.3%
0	19	32.2%	33.3%
1	3	5.1%	5.3%
2	6	10.2%	10.5%
3	8	13.6%	14.0%
4	2	3.4%	3.5%
5	4	6.8%	7.0%
6	3	5.1%	5.3%
8	1	1.7%	1.8%
10	1	1.7%	1.8%
NA	2	3.4%	–

Dotplot of guess changes

```
ggplot(data = age_guess, aes(x = change)) +  
  geom_dotplot(binwidth = 1, dotsize = 0.25) + theme_bw()
```



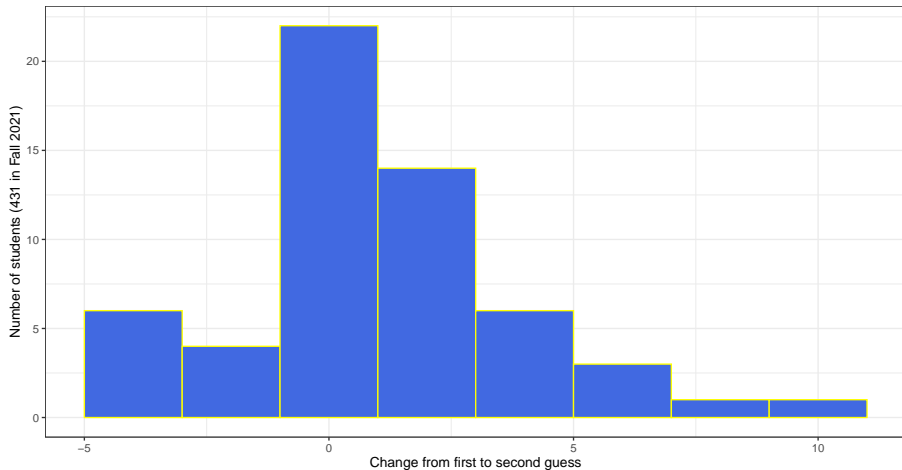
Histogram of guess changes (code)

What will this look like?

```
age_guess %>%  
  mutate(change = guess2 - guess1) %>%  
  filter(complete.cases(change)) %>%  
  ggplot(data = ., aes(x = change)) +  
  geom_histogram(binwidth = 2,  
                 fill = "royalblue", col = "yellow") +  
  theme_bw() +  
  labs(x = "Change from first to second guess",  
       y = "Number of students (431 in Fall 2021)",  
       title = "Most stayed close to their first guess.")
```

Histogram of Changes in Age Guesses

Most stayed close to their first guess.



What Happens Next?

Dr. Love will be taking his younger son to college.

- There will be no class session on Tuesday. Instead, Class 03 will involve watching a recording of Dr. Love walking through a series of analyses in R. That recording will be available soon.
- Class 04 (Thursday) will be held at the usual time and place, and the materials for it will be posted soon.

Lab 01 is due on Monday 2021-09-06 at 9 PM, and so you'll need to get started on that as soon as possible.