

431 Class 02

Thomas E. Love, Ph.D.

2022-09-01

Instructions for the Quick Survey

Please read these instructions **before** writing.

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.)
4. When you are finished, thank your partner and raise your hand. Someone will come to collect your survey.

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

Today's Agenda

- Data Structures and Variables
 - Evaluating some of the Quick Survey variables
- Looking at some of the data collected in Class 01
 - Group Guessing of Ages from 10 Photographs
 - Guessing Dr. Love's Age

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.

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!

1. Do you wear corrective lenses (contacts or glasses)? (Yes or No) _____

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

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

Has <i>statistical thinking</i> 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 tall do you think Dr. Love is? (Please indicate units.) _____

5. 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 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:		

$$\text{Right} - \text{Left} = \underline{\hspace{2cm}} \quad \text{Right} + \text{Left} = \underline{\hspace{2cm}} \quad \frac{\text{Right} - \text{Left}}{\text{Right} + \text{Left}} = \underline{\hspace{2cm}}$$

431 Quick Survey for 2021: Class 02 (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

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
10. I prefer to work on projects alone than in a team.	1	2

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 15 seconds, then quadrupling the result: _____ beats/minute.



Types of Data

The key distinction we'll make is between

- **quantitative** (numerical) and
- **categorical** (qualitative) information.

Information that is quantitative describes a **quantity**.

- All quantitative variables have units of measurement.
- Quantitative variables are recorded in numbers, and we use them as numbers (for instance, taking a mean of the variable makes some sense.)

Continuous vs. Discrete Quantities

Continuous variables (can take any value in a range)
vs. **Discrete** variables (limited set of potential values)

- Is Height a continuous or a discrete variable?
- Height is certainly continuous as a concept, but how precise is our ruler?
- Piano vs. Violin

Quantitative Variable Subtypes

We can also distinguish **interval** (equal distance between values, but zero point is arbitrary) from **ratio** variables (meaningful zero point.)

- Is Weight an interval or ratio variable?
- How about IQ?

Qualitative (Categorical) Data

Qualitative variables consist 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 (three or more categories)
- Can distinguish *nominal* (no underlying order) vs. *ordinal* (categories are ordered.)

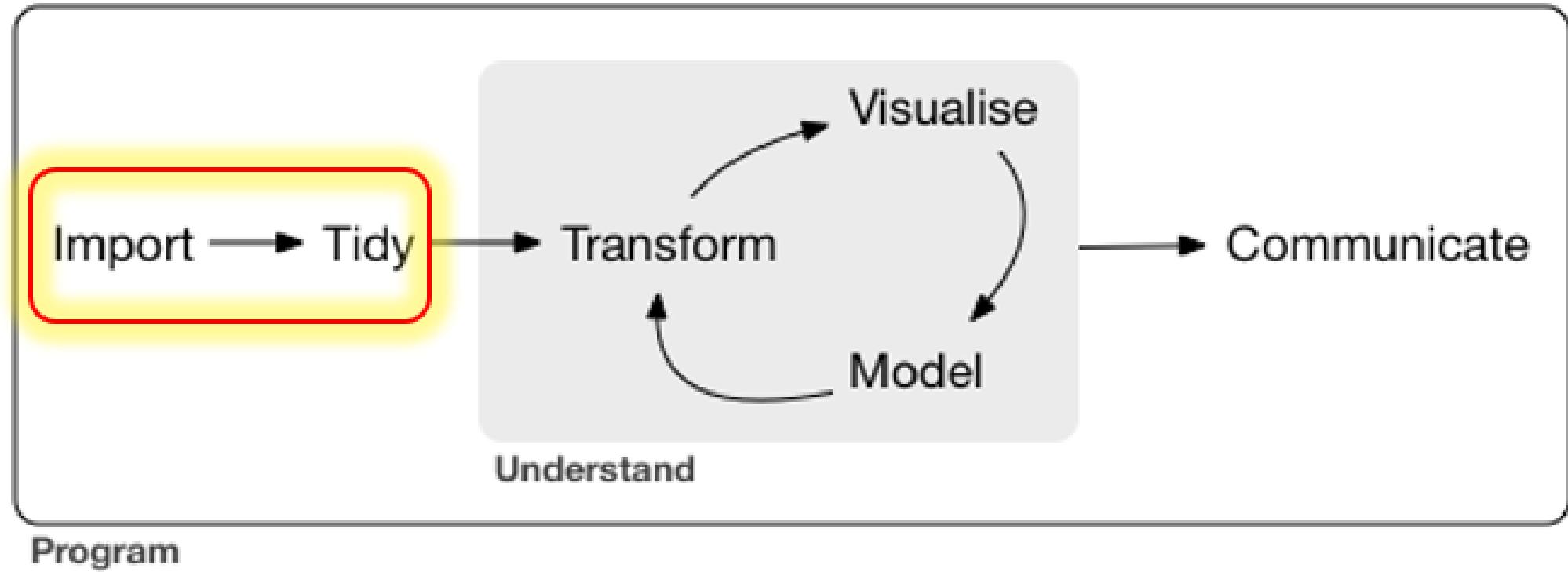
Some Categorical Variables

- 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?
- If you needed to analyze a small data set right away, which of the following software tools would you be comfortable using to accomplish that task?

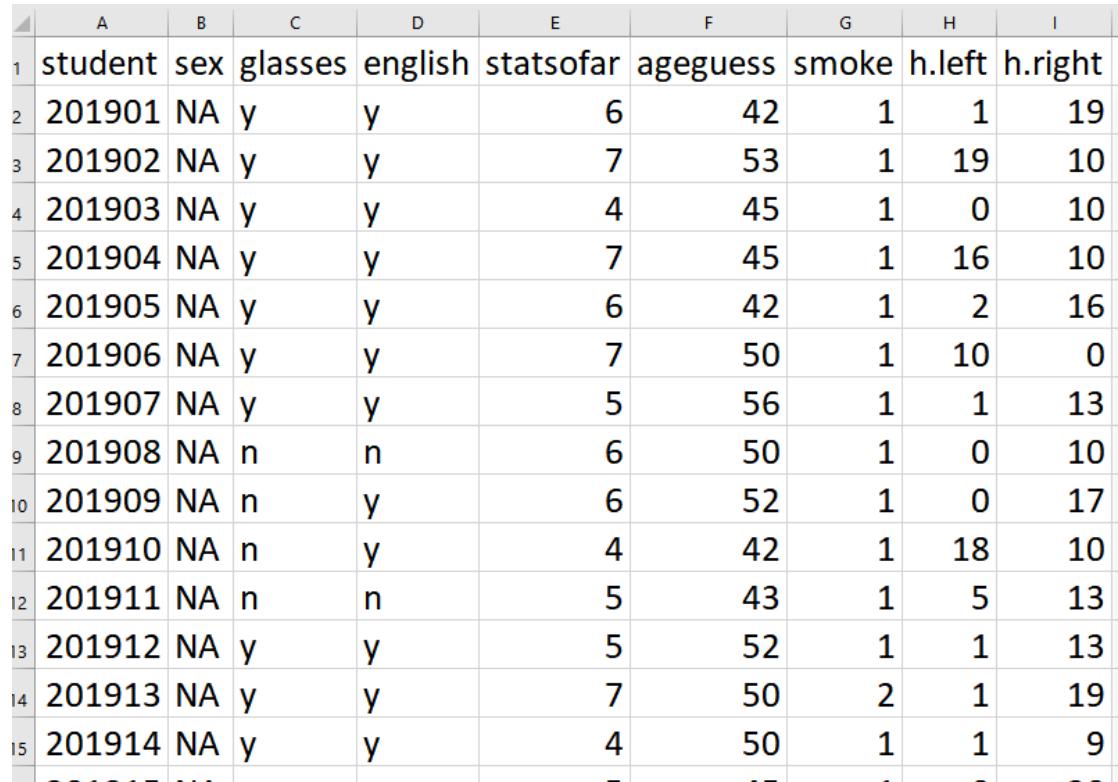
Are these quantitative or categorical?

1. Do you **smoke**? (1 = Non-, 2 = Former, 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)
-
- If quantitative, are they *discrete* or *continuous*? Do they have a meaningful zero point?
 - If categorical, how many categories? *Nominal* or *ordinal*?

Importing and Tidying Data

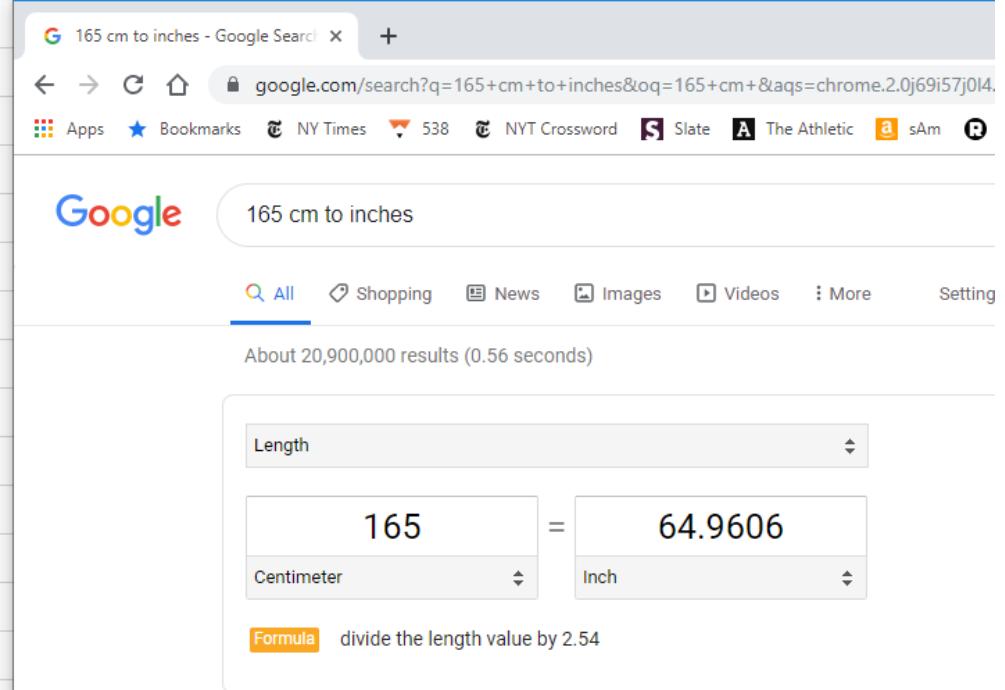


Ingesting the Quick Surveys



A screenshot of a Google Sheets spreadsheet titled "Quick Surveys". The spreadsheet has 15 rows of data, each representing a survey response. The columns are labeled A through P, and the data includes various demographic and behavioral variables.

	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						
	-	--	-	-	-	--						



Google Search Results for "165 cm to inches":

165 cm to inches

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 took (essentially) the 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? • $110/315$ (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? 50-ish 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) Female

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

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? Brown

orange

13. What is your favorite color? Blue, Brown

N/A

13. What is your favorite color?

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 ($n = 61, 1 \text{ NA}$)

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		00235566889
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

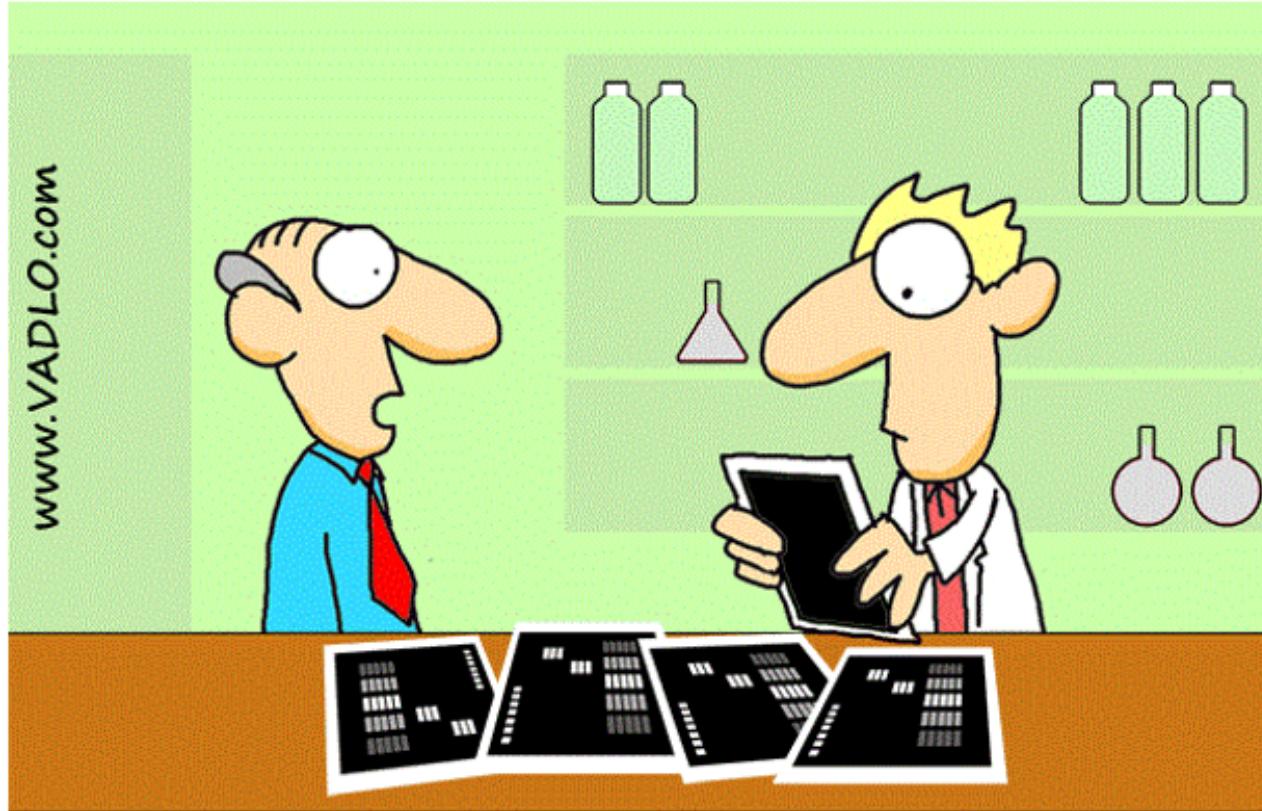
```
> stem(survey1$pulse)
```

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

```
3 | 03
3 |
4 |
4 | 688
5 | 00022244444
5 | 5666666666788889
6 | 000000000000000002222222223344444444444444
6 | 55555666666666666668888888888888888888899
7 | 0000000000000000000000000000122222222222222444444444444444
7 | 55555666666678888888888888
8 | 00000000000000000000000000001222222444444444444
8 | 55666666666888888
9 | 00000000001222224444
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.”

Age Guessing (9 groups, 10 Photos)

Photos 1-5



	#1 Age 21	#2 Age 64	#3 Age 28	#4 Age 14	#5 Age 54
2022	2 2446789	4 88	2 779	1 44455678	5 08
Group	3 25	5 03368	3 012234	2 4	6 02255
Guesses		6 03			7 00
	Mean: 27.4	Mean: 54.3	Mean: 30.6	Mean: 16.3	Mean: 62.4

- Data in [class01_group_guess10ages_2022-08-30](#) Google Sheet on our Shared Drive.

Photos 6-10

#6 Age **74**#7 Age **44**#8 Age **83**#9 Age **24**#10 Age **34**

2022
Group
Guesses

6 | 4458
7 | 03555

4 | **45588**
5 | 003
6 | 8

7 | 2369
8 | 00112

2 | 9
3 | 34555677

2 | 2489
3 | 0115
4 | 2

Mean: 69.9

Mean: 50.1

Mean: 78.2

Mean: 34.6

Mean: 30.2

Comparing the Groups

Group	# Correct	# Within 1	# Within 2	# Too Low	# Too High
Starbucks	2	4	5	2	6
Cyber Squirrels	1	2	4	4	5
Rambunctious Rhinos	0	4	4	4	6
USS Enterprise	0	3	3	4	6
Hot Tukey Sandwich	1	2	2	4	5
Glasses	0	1	2	3	7
Sunday	0	1	1	4	6
The Bootstrappers	0	1	1	5	5
Dasani	0	1	1	4	6

- So ... who wins?
- What other summaries might be helpful?

Summaries of Errors

Group	Mean Error	SD (Errors)	Median Error	(Min, Max) Error
Rambunctious Rhinos	-1.1	7.6	1	(-16, 9)
USS Enterprise	-0.5	6.3	1	(-11, 11)
Cyber Squirrels	0.1	7.6	0.5	(-10, 16)
Sunday	0.5	8.9	2	(-14, 13)
Glasses	1.3	8.0	3	(-16, 11)
The Bootstrappers	1.6	8.3	3	(-11, 11)
Hot Tukey Sandwich	2.3	8.3	2.5	(-12, 14)
Dasani	4.1	10.1	3.5	(-11, 24)
Starbucks	4.4	6.0	2.5	(-2, 16)

- How helpful are these summaries in this setting?

Absolute and Squared Errors

- AE = Absolute Value of Error = $|guess - actual|$, MSE = Mean Squared Error

Group	Mean AE	Range (AE)	Median AE	MSE
USS Enterprise	4.7	(1, 11)	3.5	35.9
Starbucks	5.0	(0, 16)	3.0	52.2
Cyber Squirrels	5.3	(0, 16)	3.5	51.5
Rambunctious Rhinos	5.5	(1, 16)	5.0	52.9
Glasses	6.3	(1, 16)	5.0	58.7
Hot Tukey Sandwich	6.9	(0, 14)	6.0	67.5
Sunday	7.3	(1, 14)	6.5	71.5
The Bootstrappers	7.4	(1, 11)	8.5	65.0
Dasani	8.1	(1, 24)	8.1	108.3

- So ... now who wins?

Guessing My Age from Class 01

The R Packages I'll Load Today

```
1 library(janitor)
2 library(kableExtra)
3 library(googlesheets4)
4 library(patchwork)
5 library(tidyverse)
```

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

From our Shared Google Drive

I've placed `class01_age_guesses_2022-08-30` on our Shared Google Drive.

Reading from Google Drive

We'll use the `read_sheet` function from the `googlesheets4` package to read in data from a Google Sheet.

To do this, we copy the URL from the Google Sheet into a temporary object I'll call `temp_url`.

```
1 gs4_deauth() # lets us work with the file without extra authorization  
2  
3 temp_url <-  
4   "https://docs.google.com/spreadsheets/d/1UMwBxg6CJfReYHNEauJSWmZSKma-DQ1P"
```

- For more on `gs4_deauth()`, visit [this link](#)

Creating the age_guess Tibble

Then I'll read in the sheet's data to a new tibble called `age_guess` using `read_sheet()`. What do we get?

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

# A tibble: 53 × 4
  student   guess1   guess2   actual
  <chr>     <dbl>     <dbl>     <dbl>
1 S-2022-01     57       62     55.5
2 S-2022-02     53       53     55.5
3 S-2022-03     50       50     55.5
4 S-2022-04     48       56     55.5
5 S-2022-05     61       NA     55.5
6 S-2022-06     63       63     55.5
7 S-2022-07     67       58     55.5
8 S-2022-08     50       57     55.5
9 S-2022-09     50       50     55.5
10 S-2022-10    43       56     55.5
# ... with 43 more rows
```

What do the `guess1` values look like?

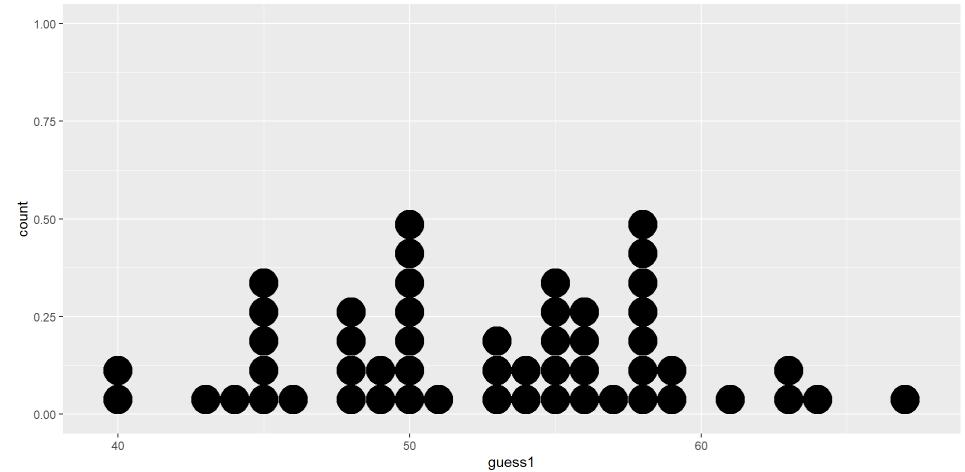
```
1 age_guess |>  
2   select(guess1) |>  
3   arrange(guess1)
```

```
# A tibble: 53 × 1
```

```
  guess1  
  <dbl>  
1     40  
2     40  
3     43  
4     44  
5     45  
6     45  
7     45  
8     45  
9     45  
10    46  
# ... with 43 more rows
```

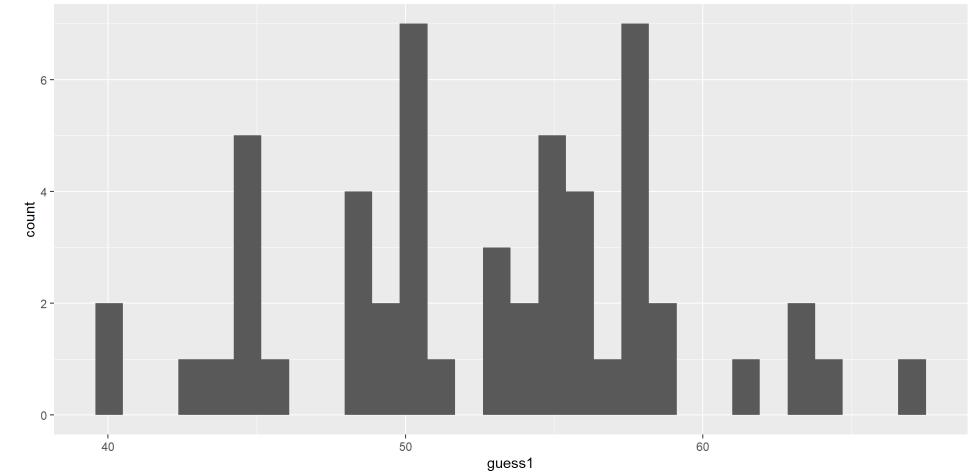
Plot the guess1 values?

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



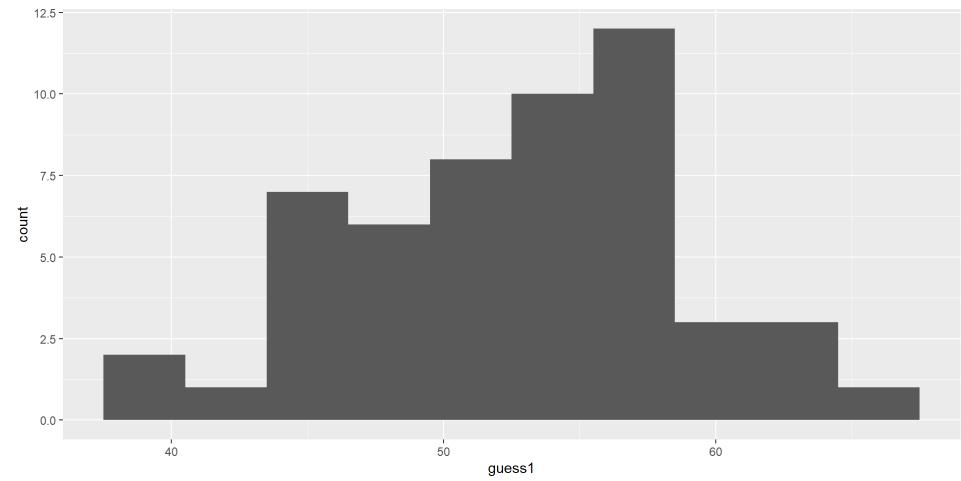
Can we make a histogram?

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



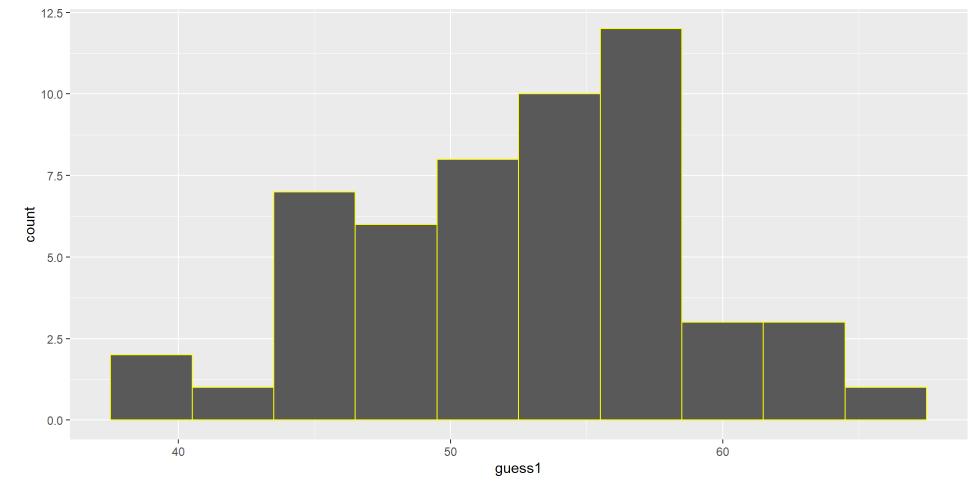
Improving the Histogram, 1

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



Improving the Histogram, 2

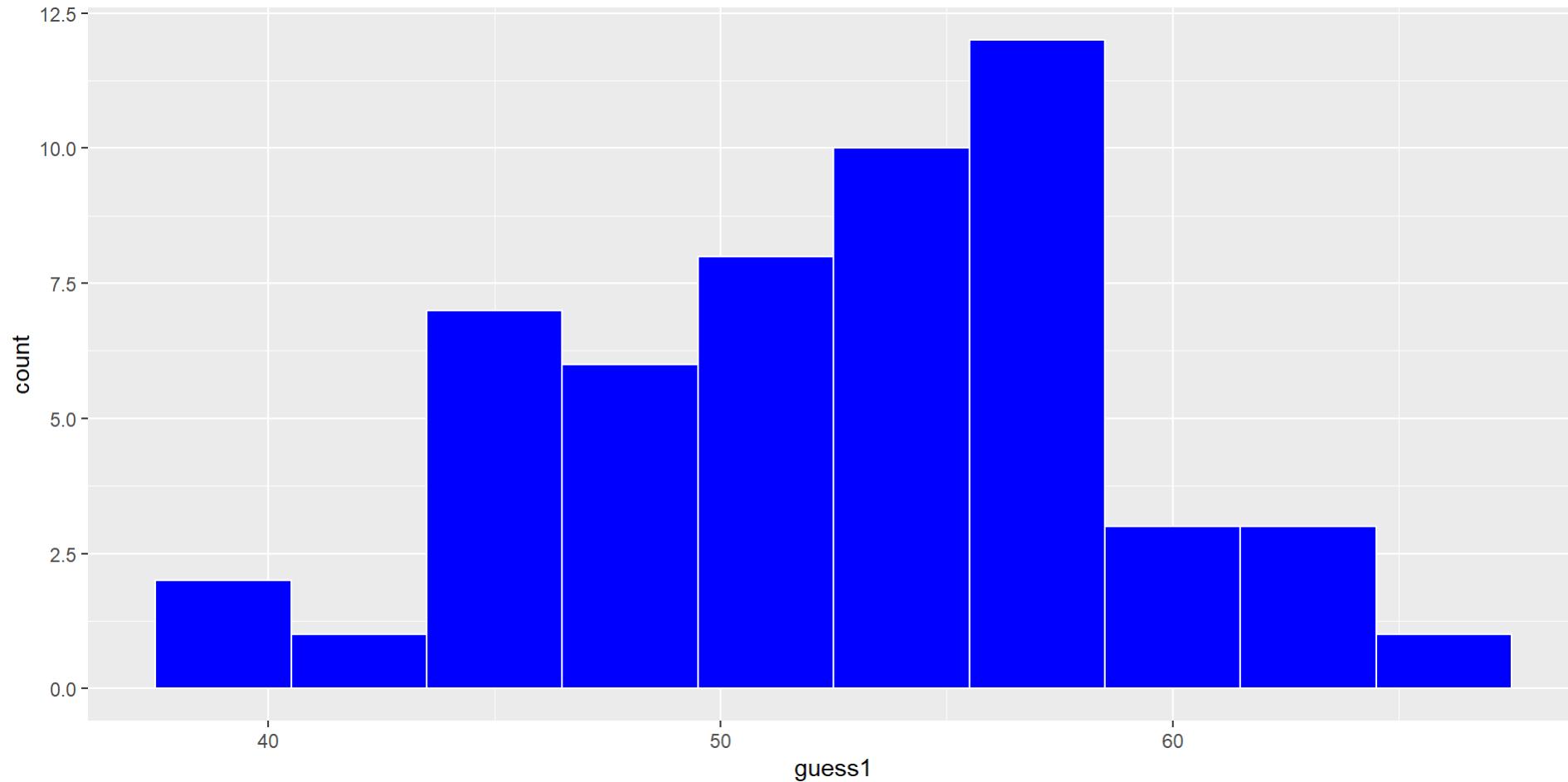
```
1 ggplot(age_guess,  
2         aes(x = guess1)) +  
3     geom_histogram(bins = 10,  
4                      col = "yellow")
```



Improving the Histogram, 3

```
1 ggplot(age_guess,  
2         aes(x = guess1)) +  
3     geom_histogram(bins = 10,  
4                      col = "white",  
5                      fill = "blue")
```

Improving the Histogram, 3

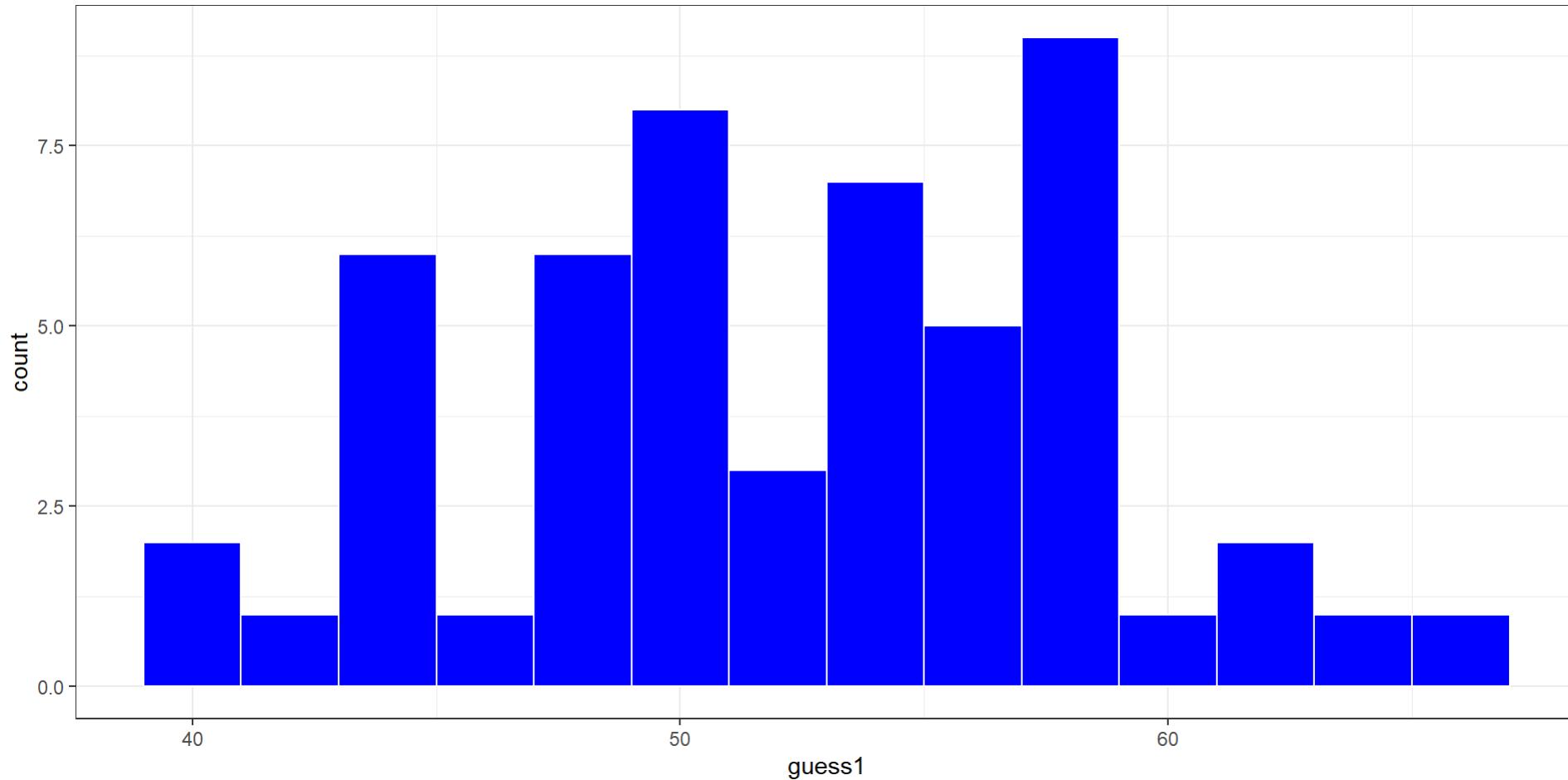


Improving the Histogram, 4

Change theme, specify bin width rather than number of bins

```
1 ggplot(age_guess,  
2         aes(x = guess1)) +  
3     geom_histogram(binwidth = 2,  
4                     col = "white", fill = "blue") +  
5     theme_bw()
```

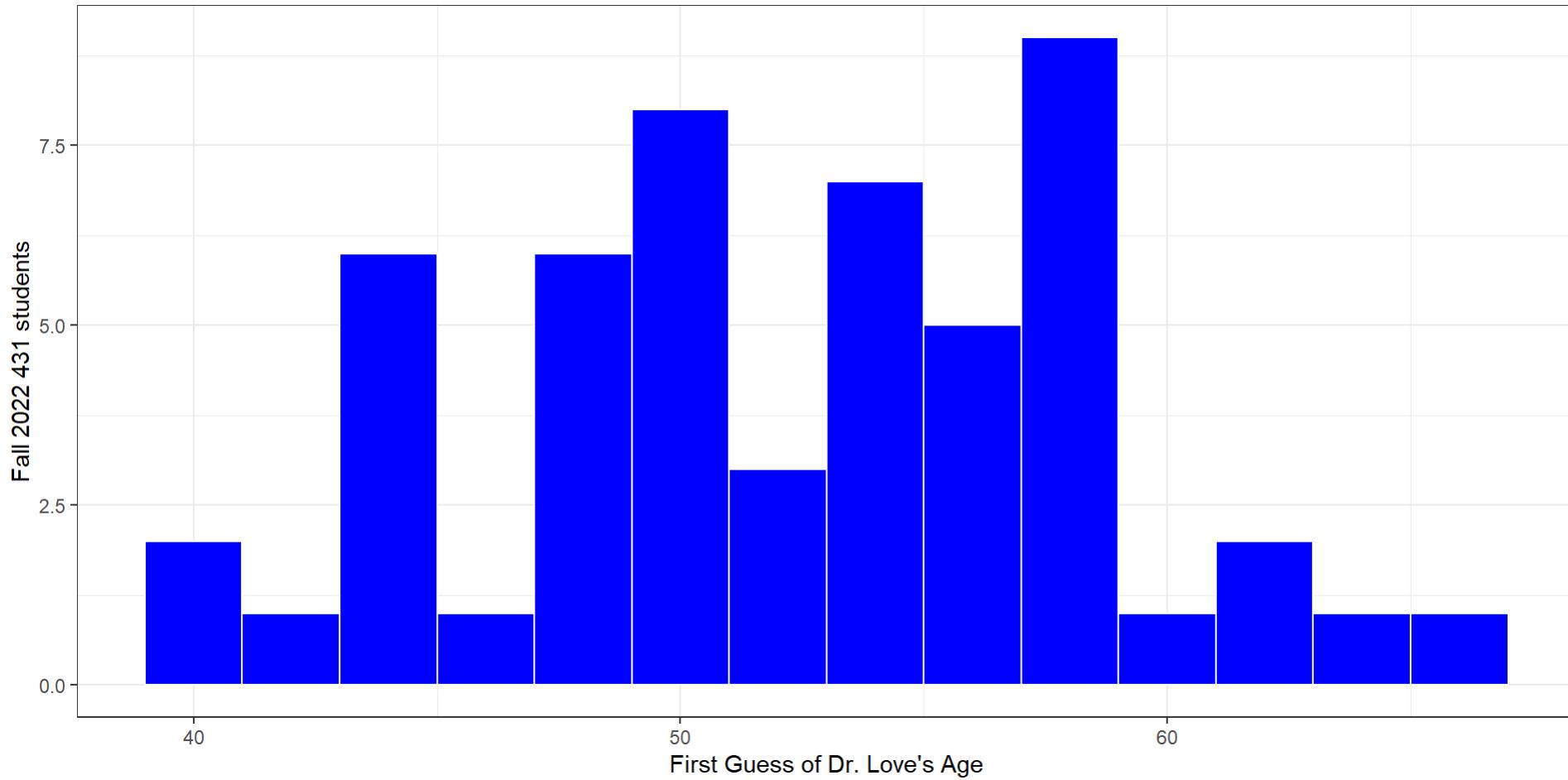
Improving the Histogram, 4



Improving the Histogram, 5

```
1 ggplot(age_guess,
2         aes(x = guess1)) +
3   geom_histogram(binwidth = 2,
4                 col = "white", fill = "blue") +
5   theme_bw() +
6   labs(
7     x = "First Guess of Dr. Love's Age",
8     y = "Fall 2022 431 students")
```

Improving the Histogram, 5



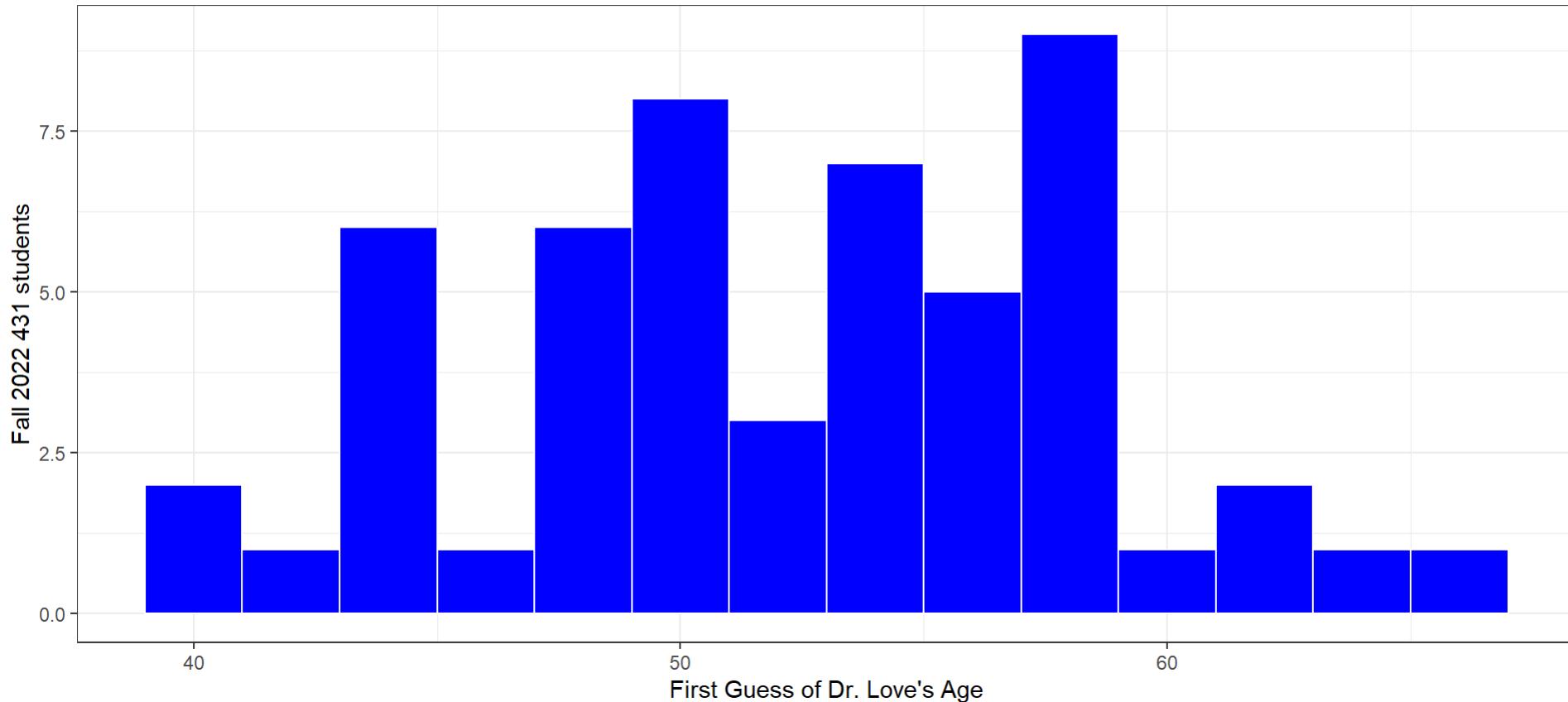
Add title and subtitle (ver. 6)

```
1 ggplot(age_guess,
2         aes(x = guess1)) +
3   geom_histogram(binwidth = 2,
4                 col = "white", fill = "blue") +
5   theme_bw() +
6   labs(
7     x = "First Guess of Dr. Love's Age",
8     y = "Fall 2022 431 students",
9     title = "Pretty wide range of guesses",
10    subtitle = "Dr. Love's Actual Age = 55.5")
```

Add title and subtitle (ver. 6)

Pretty wide range of guesses

Dr. Love's Actual Age = 55.5



Improving the Histogram, 7

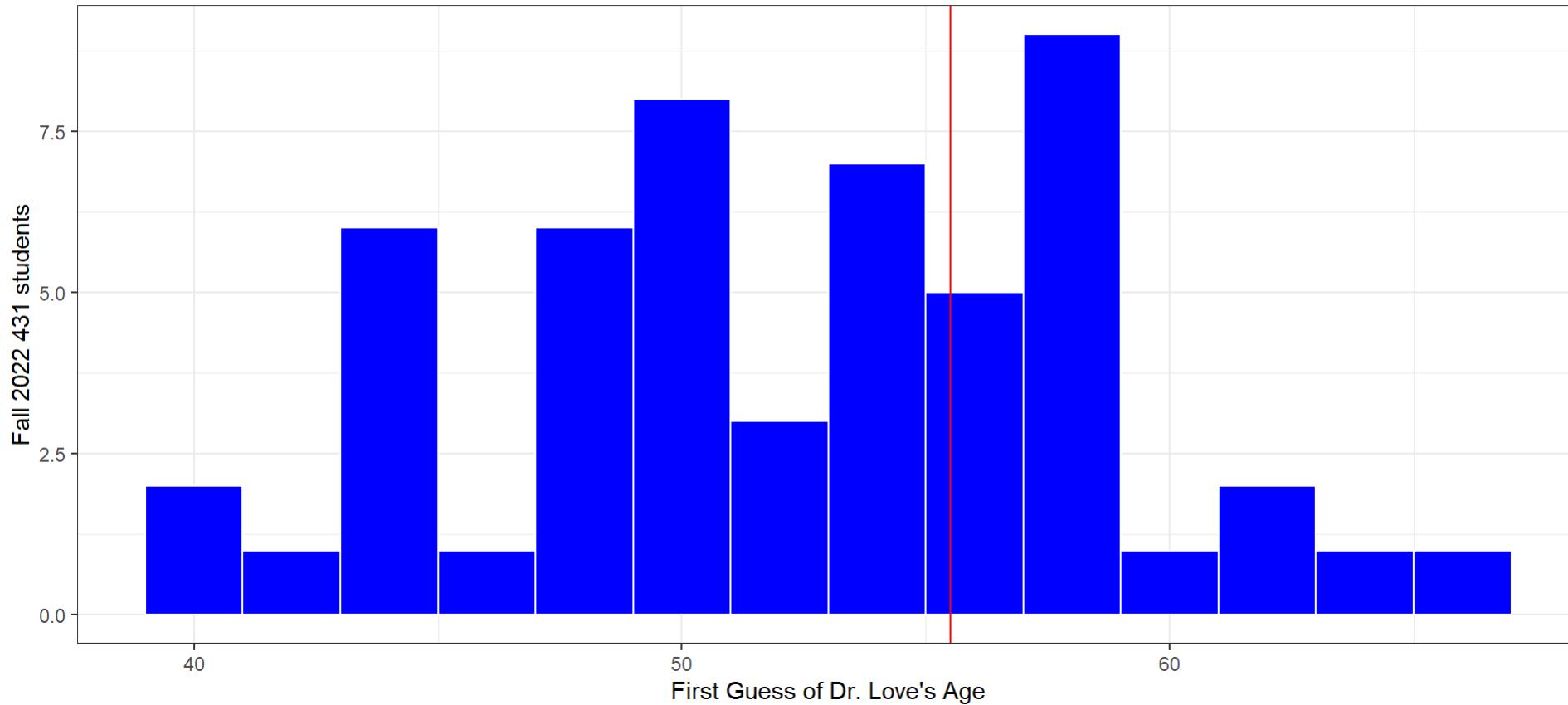
Add a vertical line at 55.5 years to show my actual age.

```
1 ggplot(age_guess,
2         aes(x = guess1)) +
3   geom_histogram(binwidth = 2,
4                 col = "white", fill = "blue") +
5   geom_vline(aes(xintercept = 55.5), col = "red") +
6   theme_bw() +
7   labs(
8     x = "First Guess of Dr. Love's Age",
9     y = "Fall 2022 431 students",
10    title = "Pretty wide range of guesses",
11    subtitle = "Dr. Love's Actual Age = 55.5")
```

Improving the Histogram, 7

Pretty wide range of guesses

Dr. Love's Actual Age = 55.5



Numerical Summary (I was 55.5)

```
1 summary(age_guess)
```

student	guess1	guess2	actual
Length:53	Min. :40.00	Min. :40.00	Min. :55.5
Class :character	1st Qu.:48.00	1st Qu.:50.00	1st Qu.:55.5
Mode :character	Median :53.00	Median :55.00	Median :55.5
	Mean :52.72	Mean :53.88	Mean :55.5
	3rd Qu.:58.00	3rd Qu.:57.00	3rd Qu.:55.5
	Max. :67.00	Max. :63.00	Max. :55.5
	NA's :3		

- Was the average guess closer on guess 1 or 2?
- What was the range of first guesses? Second guesses?
- What does the NA's : 3 mean in `guess2`?
- Why is `student` not summarized any further?

More Numerical Summaries

- Using the `favstats` function from the `mosaic` package

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

min	Q1	median	Q3	max	mean	sd	n	missing
40	48	53	58	67	52.71698	6.178055	53	0

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

min	Q1	median	Q3	max	mean	sd	n	missing
40	50	55	57	63	53.88	5.192774	50	3

- Using the `describe` function from the `psych` package

```
1 age_guess |>
2   select(guess1, guess2) |>
3   psych::describe()
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis
guess1	1	53	52.72	6.18	53	52.65	7.41	40	67	27	0.00	-0.62
se												
0.85												

Guesses between 50 and 60?

- Counting is an under-appreciated activity.

```
1 age_guess |> count(guess1 >= 50 & guess1 <= 60)
```

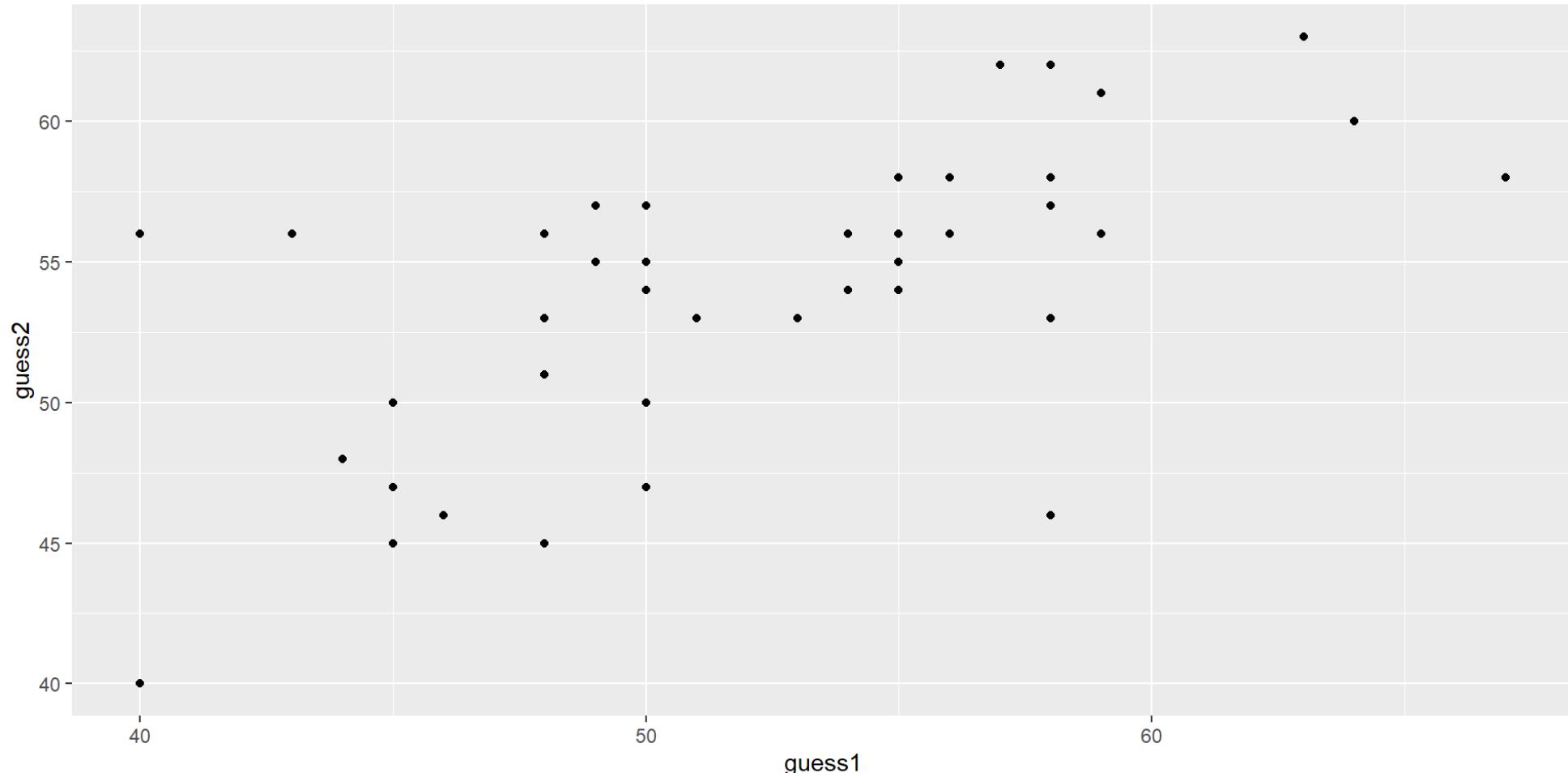
```
# A tibble: 2 × 2
`guess1 >= 50 & guess1 <= 60`      n
<lgl>                                <int>
1 FALSE                               21
2 TRUE                                32
```

```
1 age_guess |> count(guess2 >= 50 & guess1 <= 60)
```

```
# A tibble: 3 × 2
`guess2 >= 50 & guess1 <= 60`      n
<lgl>                                <int>
1 FALSE                               14
2 TRUE                                37
3 NA                                  2
```

Guess 1 - Guess 2 Scatterplot

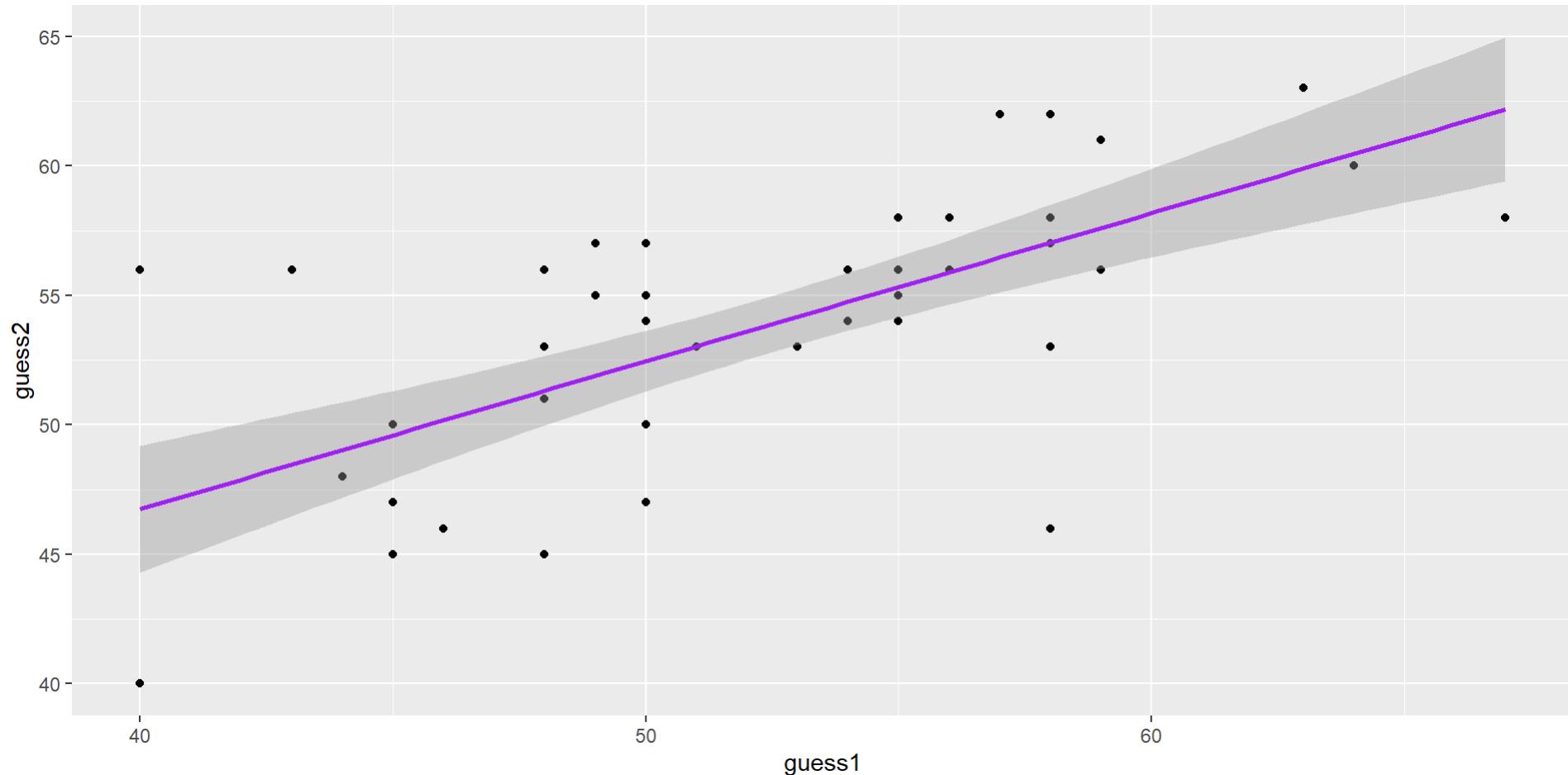
```
1 ggplot(data = age_guess, aes(x = guess1, y = guess2)) +  
2   geom_point()
```



Filter to complete cases, and add regression line

```
1 temp <- age_guess |>  
2   filter(complete.cases(guess1, guess2))  
3  
4 ggplot(data = temp, aes(x = guess1, y = guess2)) +  
5   geom_point() +  
6   geom_smooth(method = "lm", formula = y ~ x, col = "purple")
```

Filter to complete cases, and add regression line



What is that regression line?

Call:

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

Coefficients:

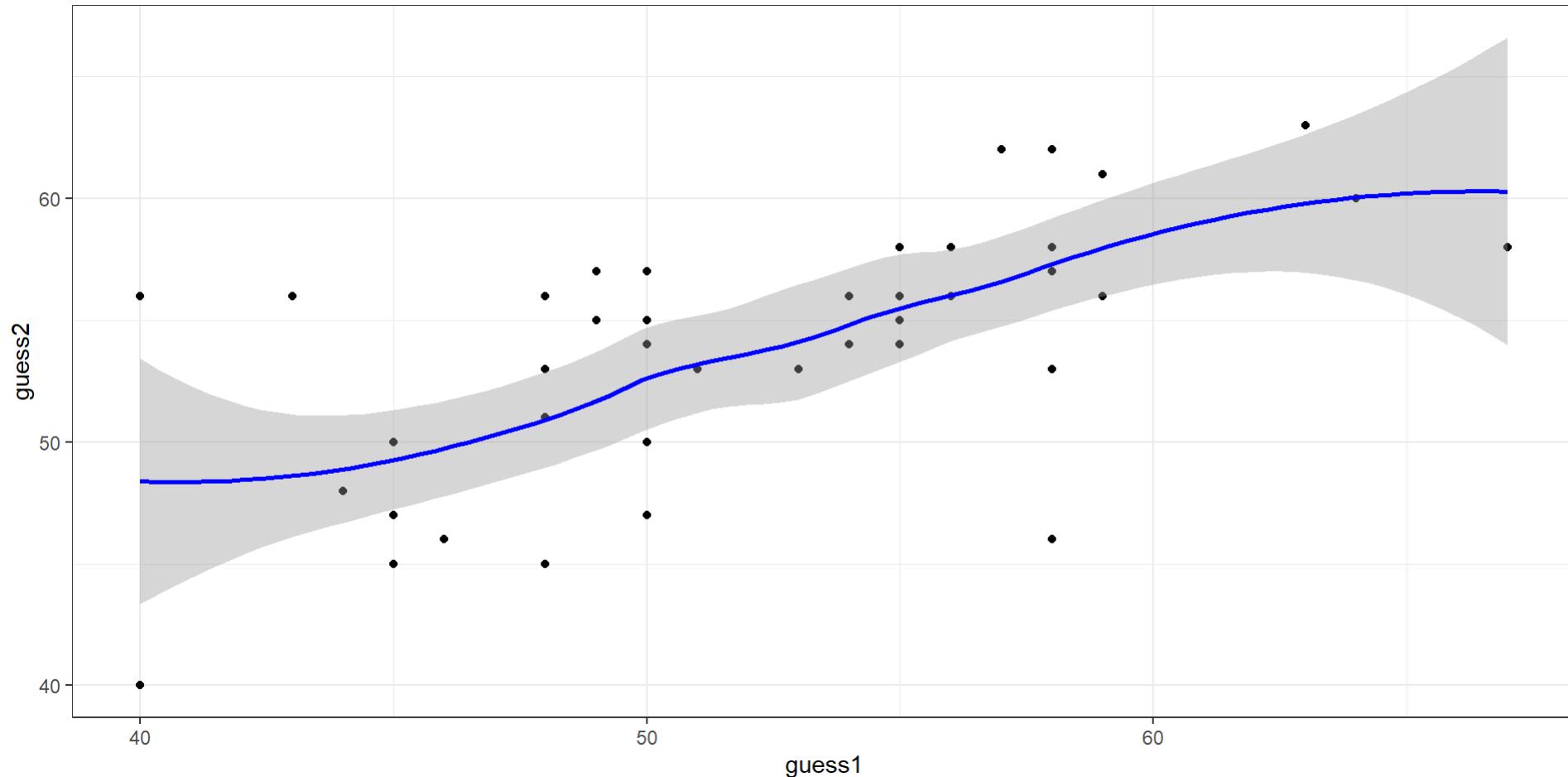
(Intercept)	guess1
23.8708	0.5718

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

How about a loess smooth curve?

```
1 temp <- age_guess |>
2   filter(complete.cases(guess1, guess2))
3
4 ggplot(data = temp, aes(x = guess1, y = guess2)) +
5   geom_point() +
6   geom_smooth(method = "loess", formula = y ~ x, col = "blue") +
7   theme_bw()
```

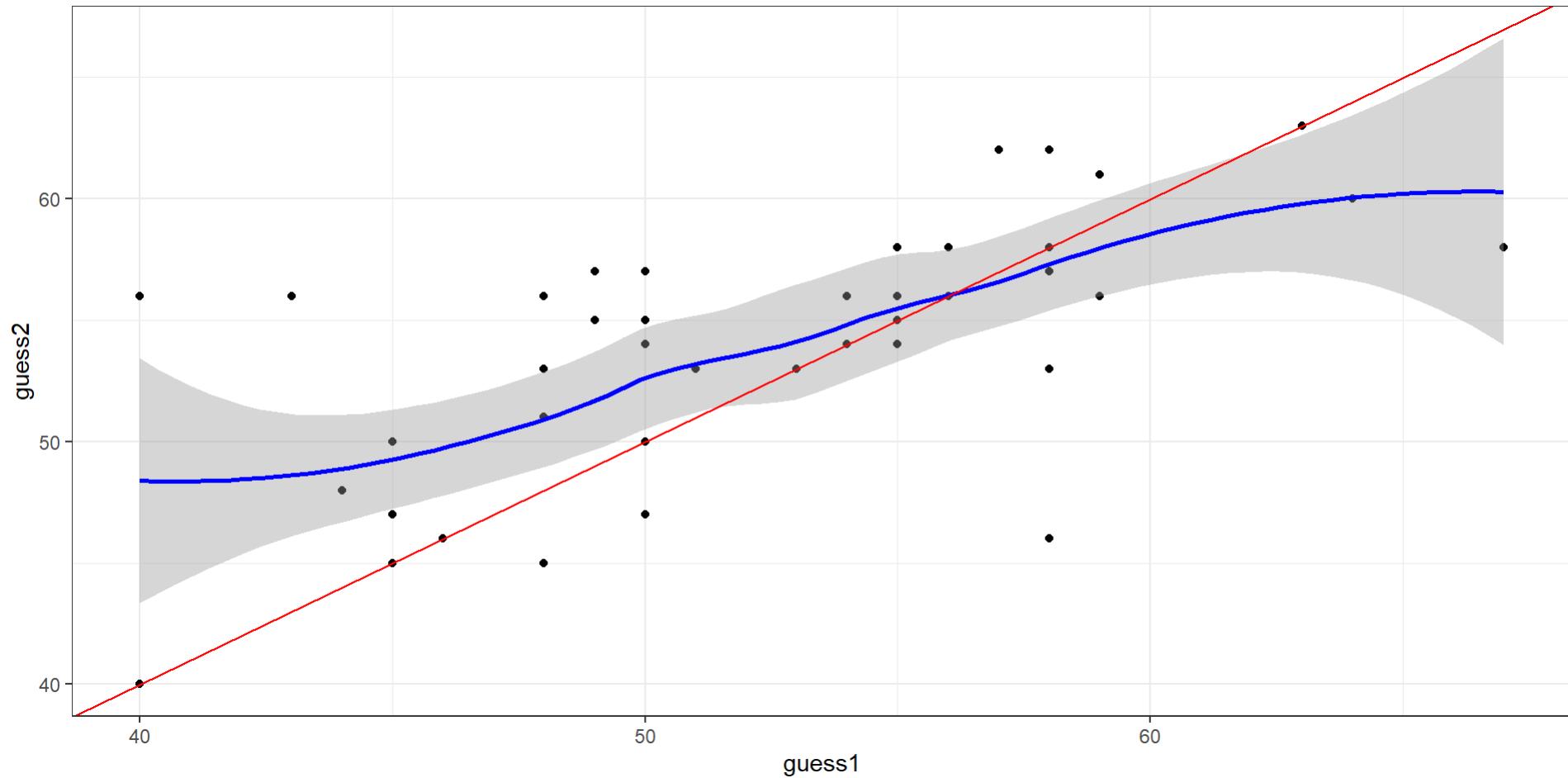
How about a loess smooth curve?



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

```
1 temp <- age_guess |>
2   filter(complete.cases(guess1, guess2))
3
4 ggplot(data = temp, aes(x = guess1, y = guess2)) +
5   geom_point() +
6   geom_smooth(method = "loess", formula = y ~ x, col = "blue") +
7   geom_abline(intercept = 0, slope = 1, col = "red") +
8   theme_bw()
```

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



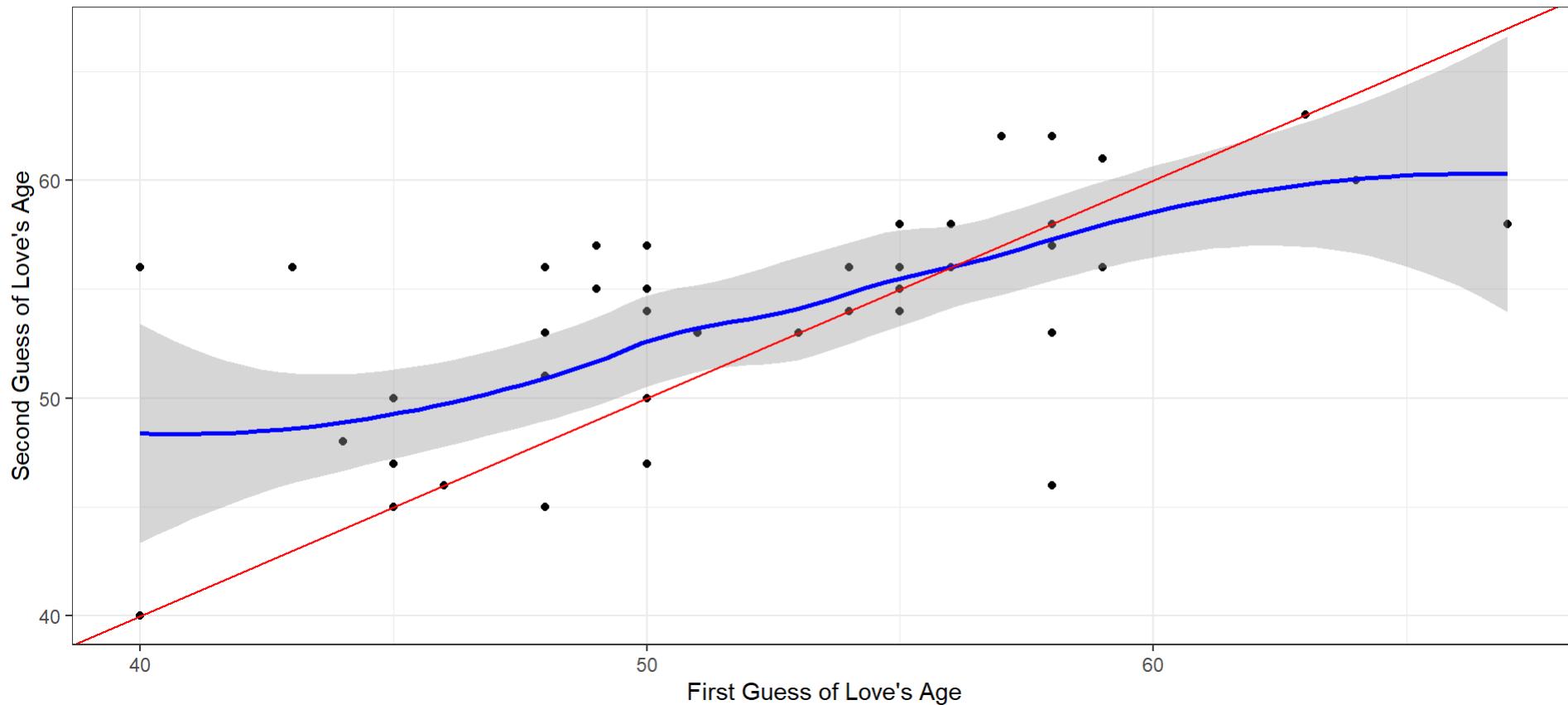
With Better Labels

```
1 ggplot(data = temp, aes(x = guess1, y = guess2)) +
2   geom_point() +
3   geom_smooth(method = "loess", formula = y ~ x, col = "blue") +
4   geom_abline(intercept = 0, slope = 1, col = "red") +
5   geom_text(x = 40, y = 38, label = "y = x", col = "red") +
6   labs(x = "First Guess of Love's Age",
7         y = "Second Guess of Love's Age",
8         title = "Comparing 2022 Age Guesses",
9         subtitle = "Love's actual age = 55.5") +
10  theme_bw()
```

With Better Labels

Comparing 2022 Age Guesses

Love's actual age = 55.5



How did guesses change?

- Did your guesses decrease / stay the same / increase?
- Calculate $\text{guess2} - \text{guess1}$ and examine its sign.

```
1 age_guess |>
2   count(sign(guess2 - guess1))
```

```
# A tibble: 4 × 2
`sign(guess2 - guess1)`      n
                <dbl> <int>
1                  -1    10
2                   0    18
3                   1    22
4                  NA     3
```

How much did guesses change?

Create new variable (change = guess2 - guess1)

```
1 age_guess <- age_guess |>  
2   mutate(change = guess2 - guess1)  
3  
4 age_guess |> select(change) |> summary()
```

```
change  
Min.   :-12.0  
1st Qu.:  0.0  
Median :  0.0  
Mean   :  1.4  
3rd Qu.:  4.0  
Max.   : 16.0  
NA's   :3
```

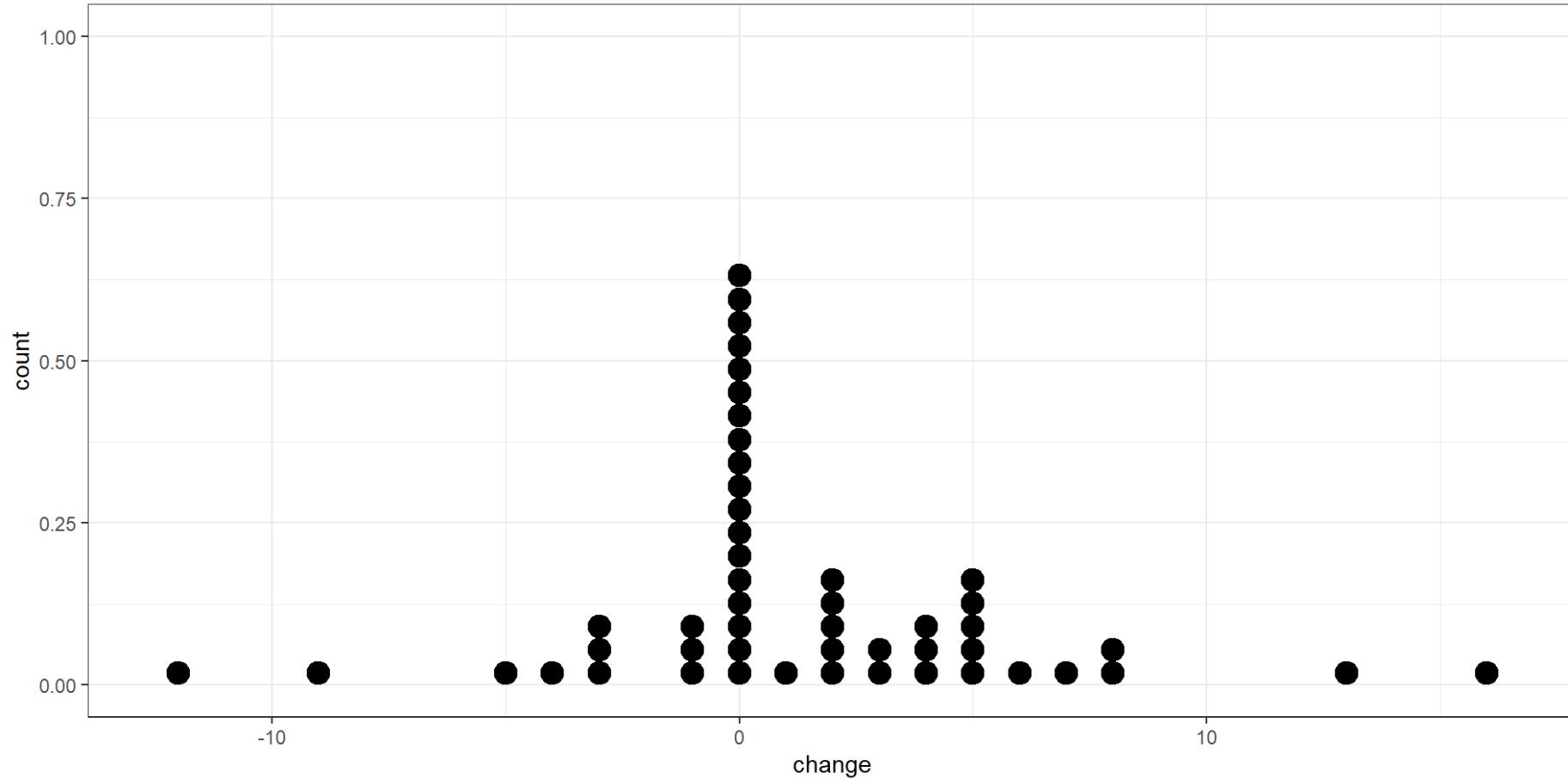
Table (via tabyl) of guess changes

```
1 age_guess |>
2   tabyl(change) |>
3   adorn_pct_formatting()
```

change	n	percent	valid_percent
-12	1	1.9%	2.0%
-9	1	1.9%	2.0%
-5	1	1.9%	2.0%
-4	1	1.9%	2.0%
-3	3	5.7%	6.0%
-1	3	5.7%	6.0%
0	18	34.0%	36.0%
1	1	1.9%	2.0%
2	5	9.4%	10.0%
3	2	3.8%	4.0%
4	3	5.7%	6.0%
5	5	9.4%	10.0%
6	1	1.9%	2.0%
7	1	1.9%	2.0%
∞	∞	∞ ∞	∞ ∞

Dotplot of guess changes

```
1 ggplot(data = age_guess, aes(x = change)) +  
2   geom_dotplot(binwidth = 1, dotsize = 0.5) +  
3   theme_bw()
```



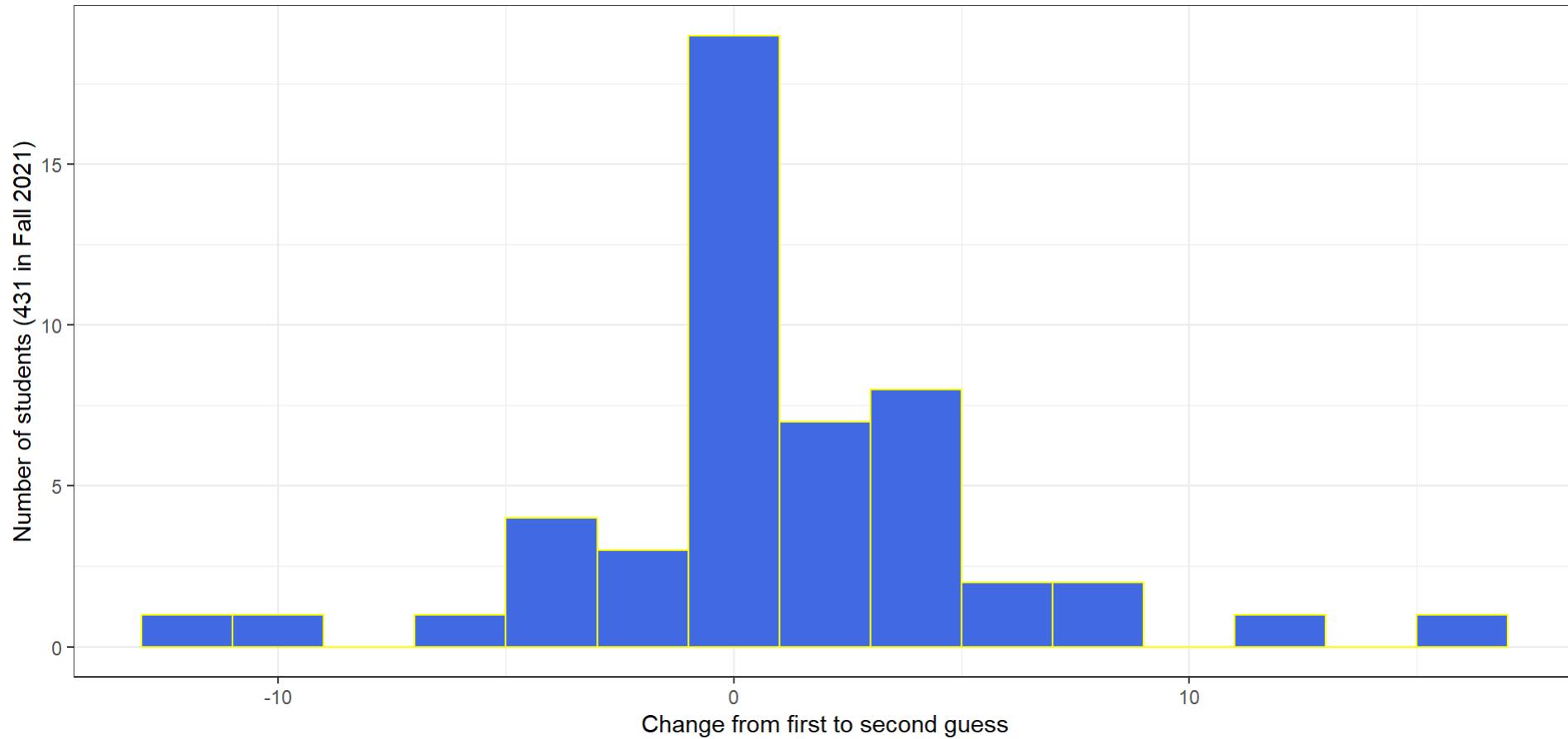
Histogram of Guess Changes

What will this look like?

```
1 ggplot(data = age_guess, aes(x = change)) +  
2   geom_histogram(binwidth = 2, fill = "royalblue", col = "yellow") +  
3   theme_bw() +  
4   labs(x = "Change from first to second guess",  
5         y = "Number of students (431 in Fall 2021)",  
6         title = "Most stayed close to their first guess.")
```

Histogram of Guess Changes

Most stayed close to their first guess.



What Happens Next Time?

To be announced.