Five Old Quiz Questions

Thomas E. Love

2021-09-16

R Packages used to create these questions

```
library(NHANES)
library(patchwork)
library(janitor)
library(magrittr)
library(tidyverse)
```

What is This?

This is a set of five questions from an old quiz. After the questions, I have provided an answer sketch, and an idea of how well people did on these questions in the old quiz. This is just a convenient sample of old questions, and is not meant to be completely representative of what I might ask you this year.

I built this just for the benefit of students worried about Quiz 1 who wanted to understand what it might look like. We're not collecting responses on this set of questions, nor is it required that you even look at this document. If you have questions about this after reading the answer sketch, feel free to ask us.

Were this an actual quiz, you would have four days to do a set of 20-25 questions. You would have some additional instructions, and there would be a separate Google Form answer sheet where you would select or type in your best response for each question.

Occasionally, I ask students to provide a single line of code. This usually comes with a restriction on how many pipes can be included - for these questions, that limit was one pipe, although you may or may not need the pipe in any particular setting.

Moreover, you need not include the library command at any time for any of your code. Assume in all questions that all relevant packages (those listed above) have been loaded in R.

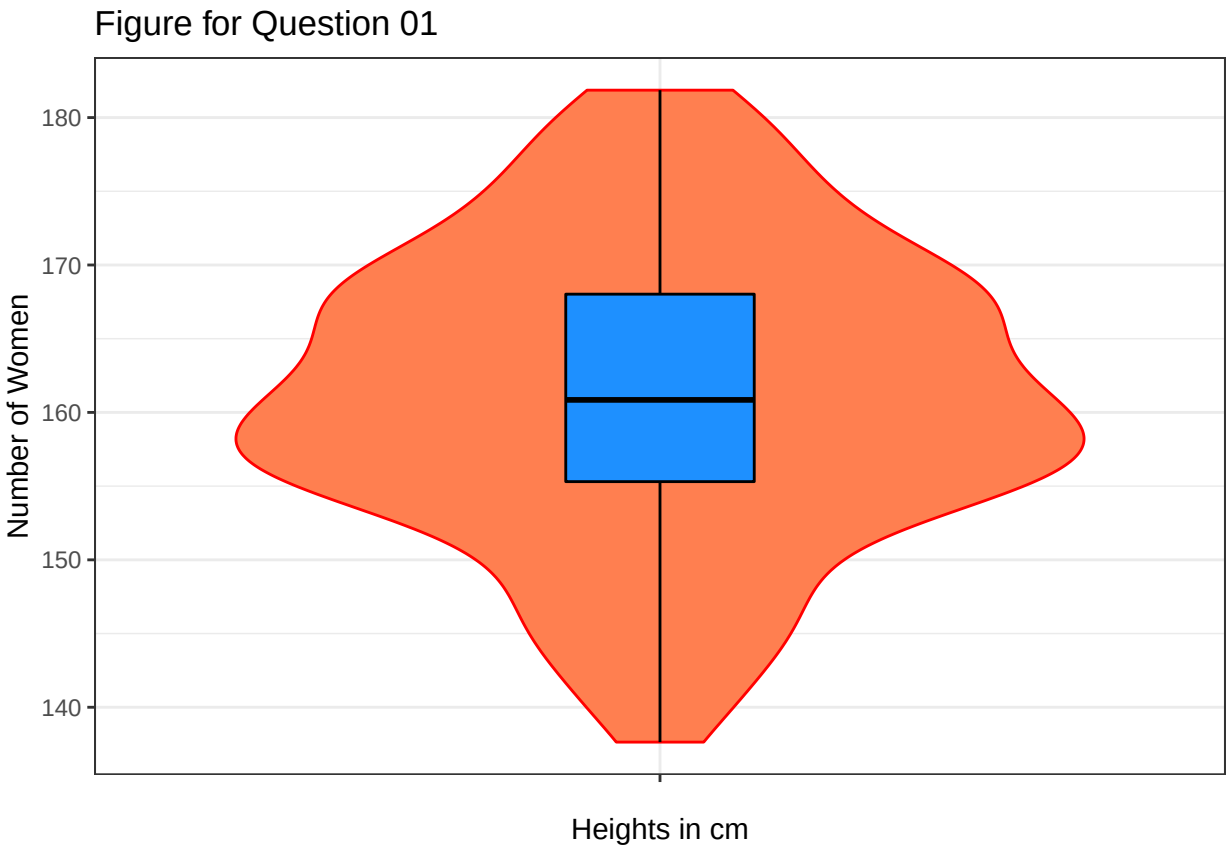
Note that the `mitral.csv` data file (described in Question 2) is available to you on our web site.

1 Question 1

A sample of heights for 148 adult non-Hispanic white women living in the state of Ohio have been stored in the `height` column in a tibble called `dat01`. Which of the following bits of R code was NOT used in generating the Figure for Question 01?

a. `geom_boxplot(fill = "dodgerblue", col = "black", width = 0.2)`
b. `geom_line(fill = "blue", col = "white")`
c. `geom_violin(col = "red", fill = "coral")`
d. `ggplot(data = dat01, aes(x = "", y = height))`
e. `labs(title = "Figure for Question 01")`
f. `labs(x = "Heights in cm", y = "Number of Women")`
g. `theme_bw()`

1.1 Figure for Question 01



Setup for Questions 2-4

Questions 2-4 make use of the `mitral.csv` data that describe 40 patients with either aortic or mitral regurgitation who had heart surgery.

The data have been read into the `mitral` tibble, as shown. The variables are:

- `subj_id` = subject ID
- `ef_pre` = ejection fraction prior to surgery
- `ef_post` = ejection fraction after surgery
- `reg_type` = regurgitation type, either mitral or aortic
- `NYHA` = NYHA class, an ordered four-category variable describing functional limitations
  - NYHA class levels are I, II, III and IV, with I indicating the least and IV indicating the most severe limitations
- `sbp_pre` = systolic blood pressure prior to surgery, in mm Hg.
- `sbp_post` = systolic blood pressure after surgery, in mm Hg.

Note that the `mitral.csv` data file is available to you on our web site.

Tibble (with Code) for Questions 2-4

```
mitral <- read_csv("data/mitral.csv")
```

```
Rows: 40 Columns: 7
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (3): subj_id, reg_type, NYHA
dbl (4): ef_pre, ef_post, sbp_pre, sbp_post

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
mitral
```

```
# A tibble: 40 x 7
   subj_id ef_pre ef_post reg_type NYHA  sbp_pre sbp_post
   <chr>    <dbl>   <dbl> <chr>    <chr>   <dbl>    <dbl>
 1 S-127     0.56    0.34 aortic   III       140      138
 2 S-156     0.62    0.47 aortic   II        110      110
 3 S-174     0.67    0.41 mitral   II        120      120
 4 S-175     0.63    0.48 mitral   III        90      100
 5 S-222     0.74    0.54 mitral   II        130      115
 6 S-263     0.6     0.33 aortic   III       150      135
 7 S-288     0.53    0.47 aortic   I         215      130
 8 S-298     0.69    0.6  mitral   II         95      100
 9 S-300     0.6     0.3  aortic   III       150      130
10 S-341     0.66    0.43 mitral   IV        125      124
# ... with 30 more rows
```

2 Question 2

Write a single line of R code that will specify the coefficients of a linear regression model to predict systolic blood pressure after surgery on the basis of systolic blood pressure prior to surgery, using the `mitral` tibble. Be sure that your code will work, and in particular, that you haven't spelled anything incorrectly.

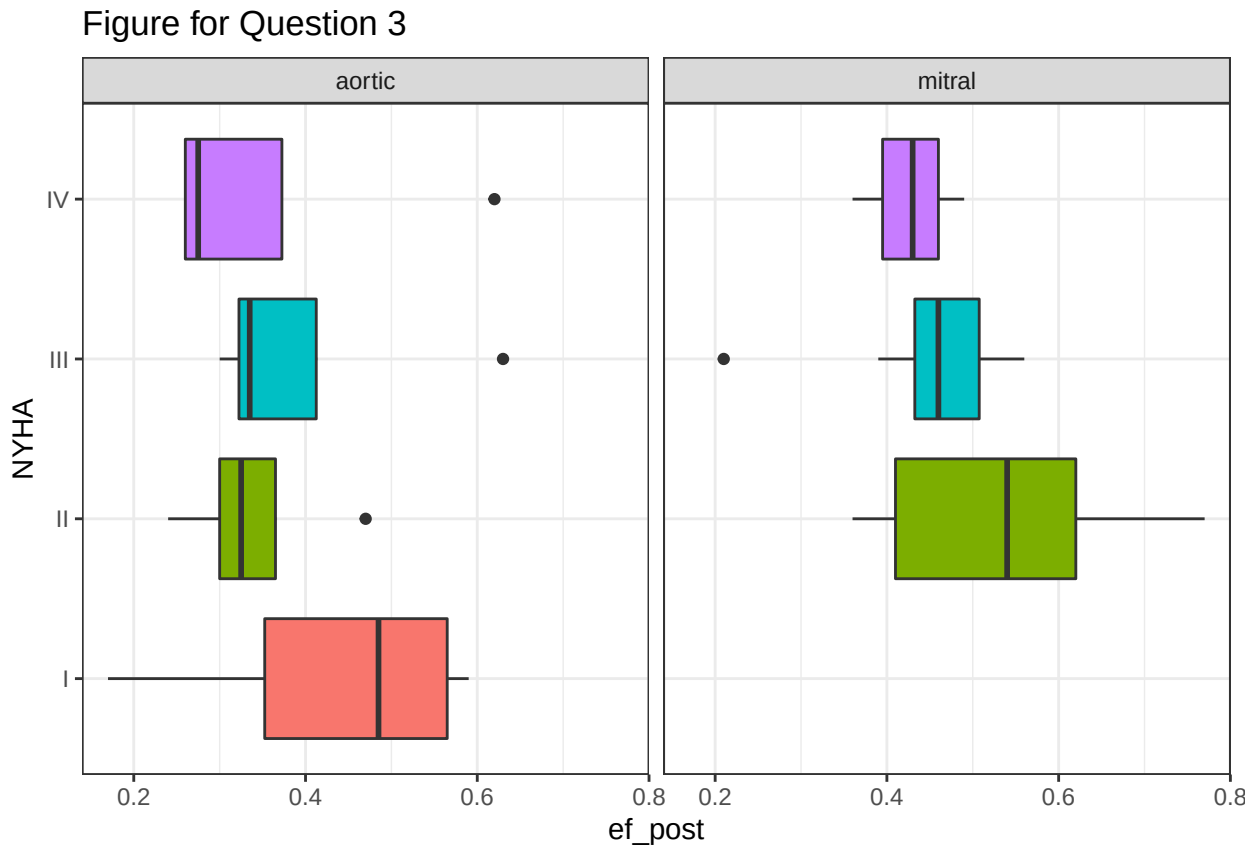
3 Question 3

The Figure for Question 3 shows the ejection fraction after surgery for 40 patients, and the complete code used to develop the Figure, based on the `mitral` tibble, is also provided. Which of the following lines of R code would best inform you as to why there are only seven boxplots in the Figure for Question 3 rather than eight?

a. `drop_na`
b. `facet_grid(~ reg_type, labeller = "label_both")`
c. `summary(mitral)`
d. `mitral %>% group_by(NYHA) %>% summarize(reg_type)`
e. `mitral %>% count(reg_type, NYHA)`
f. None of these would be useful.

3.1 Figure (with Code) for Question 3

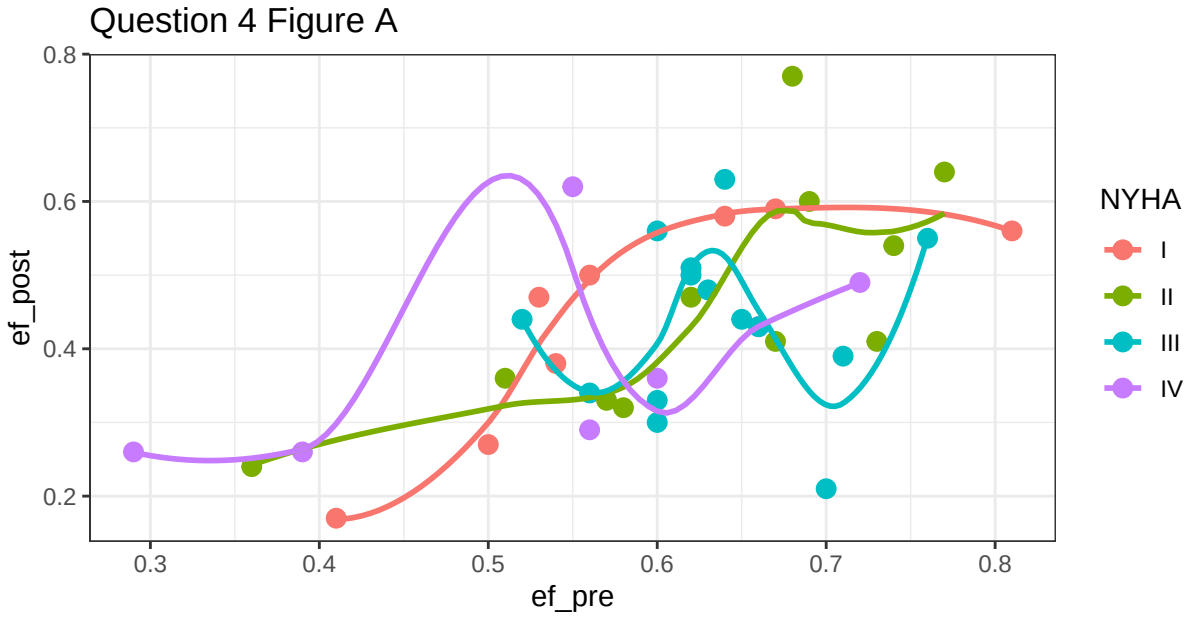
```
ggplot(mitral, aes(x = NYHA, y = ef_post, fill = NYHA)) +
    geom_boxplot() +
    facet_wrap(~ reg_type) +
    coord_flip() +
    guides(fill = "none") +
    labs(title = "Figure for Question 3") +
    theme_bw()
```

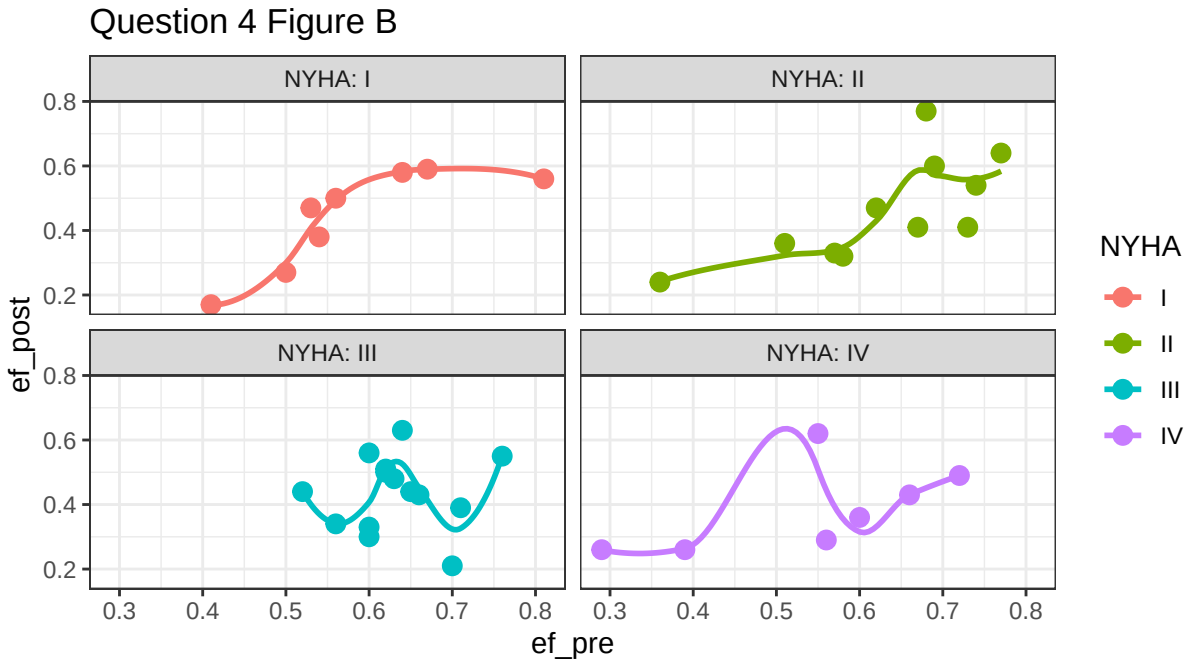


4 Question 4

The Figures for Question 4 (labeled A and B) each show the same data, and the same loess smooths, using two different approaches. To turn Plot A into Plot B, we changed the title and added one line of R code. Please specify the one-line R command used to turn Plot A into Plot B.

4.1 Figures for Question 4





5 Question 5

In the Tibble for Question 5, we show the arm and nose lengths (in cm) and the arm-nose ratio (ANratio), calculated by dividing the arm length by the nose length, for 20 subjects. I have not provided you with an electronic copy of the data set for this question, but you will see a small piece of summary output at the bottom of the tibble.

The Statue of Liberty's nose measures 4 feet, 6 inches, and her arm is 42 feet long. Calculate the Statue's arm/nose ratio, and use it to specify her Z score (# of standard deviations above or below the group mean) as compared to the 20 subjects listed in the Tibble for Question 5. Your response should be the Z score for the Statue of Liberty, with an appropriate sign, rounded to one decimal place.

5.1 Tibble and Output for Question 5

```
dat5
```

```
# A tibble: 20 x 4
   subject   arm  nose ANratio
   <chr>   <dbl> <dbl>   <dbl>
 1 Akari    73.3   4.7    15.6
 2 Beth     60.3   4.5    13.4
 3 Carol    74.7   4.7    15.9
 4 Donna    60.6   4.3    14.1
 5 Early    65.6   4.4    14.9
 6 Feng     63.2   4.3    14.7
 7 Grace    75.7   4.7    16.1
 8 Hanna    67.9   4.3    15.8
 9 Ione     63.4   4.2    15.1
10 Julie    74.4   4.8    15.5
11 Karen    62.6   4.2    14.9
12 Lin      64.1   4.3    14.9
13 Mary     69.7   4.2    16.6
14 Nancy    69.6   4.7    14.8
15 Olive    75.4   4.6    16.4
16 Paris    81.4   5.5    14.8
17 Ruo      70.8   4.4    16.1
18 Sara     82.2   5.2    15.8
19 Tilly    58.5   4.4    13.3
20 Vivi     65.8   4.3    15.3
```

```
dat5 %>% summarize(sum(ANratio))
```

```
# A tibble: 1 x 1
  `sum(ANratio)`
           <dbl>
1            304
```

```
dat5 %$% var(ANratio) %>% round_half_up(., 2)
```

```
[1] 0.81
```

NOTE: This is the final sample question. An answer sketch starts on the next page of this document.

6 Answer Sketch

6.1 Answer 01 is b

The Figure in Question 01 was made by combining the other six bits of code. This is a boxplot and a violin plot, but there's no line chart (as would be implied by `geom_line`) here.

Here's the actual code that was used. . .

```
set.seed(2018004)
temp <- rnorm(148, mean = 161.4, sd = 9.2)

dat01 <- tibble(height = temp)

ggplot(data = dat01, aes(x = "", y = height)) +
    geom_violin(col = "red", fill = "coral") +
    geom_boxplot(fill = "dodgerblue", col = "black", width = 0.2) +
    labs(x = "Heights in cm", y = "Number of Women") +
    labs(title = "Figure for Question 01") +
    theme_bw()
```

6.2 Answer 2 is a line of R code, like `lm(sbp_post ~ sbp_pre, data = mitral)`

Any code that would produce the estimated slope and intercept coefficients for the correct model is OK.

Good options include:

- `lm(sbp_post ~ sbp_pre, data = mitral)`
- `coef(lm(sbp_post ~ sbp_pre, data = mitral))`
- `summary(lm(sbp_post ~ sbp_pre, data = mitral))`
- `broom::tidy(lm(sbp_pre ~ sbp_post, data = mitral))`

6.3 Answer 3 is e

As we can see from the results of applying the code in **e** below, there are no subjects in the NYHA I group who had mitral regurgitation in the data set. That's why no data are plotted in the bottom right of the Figure for Question 3. None of the other codes would provide us with this information, although there are other ways we could have used to figure this out.

```
mitral %>% count(reg_type, NYHA)
```

```
# A tibble: 7 x 3
  reg_type NYHA      n
  <chr>    <chr> <int>
1 aortic   I         8
2 aortic   II        4
3 aortic   III       4
4 aortic   IV        4
5 mitral   II        7
6 mitral   III      10
7 mitral   IV        3
```

6.4 Answer 4 is `facet_wrap(~ NYHA, labeller = "label_both")`

We didn't worry about whether or not you included the + sign after the command.

6.5 Answer 5 is -6.5

To calculate the arm/nose ratio of the Statue of Liberty, we need to get her arm and nose lengths on the same scale. (Note that it doesn't have to be the same scale as was used for the women in the class, mathematically.) So, her arm length is 42 feet, and her nose length is 4.5 feet. Thus, the Statue of Liberty has arm/nose ratio of $42/(4.5) = 9.33$.

As for the 20 women in the tibble, we can obtain their mean and standard deviation:

```
mosaic::favstats(~ ANratio, data = dat5)
```

```
Registered S3 method overwritten by 'mosaic':
  method                           from
  fortify.SpatialPolygonsDataFrame ggplot2
```

```
  min   Q1 median     Q3  max mean        sd  n missing
 13.3 14.8   15.2 15.825 16.6 15.2 0.9002924 20       0
```

The mean of the 20 women is 15.2, since the sum we provided was 304, and since 304/20 is 15.2, and the standard deviation is 0.9 (or, if you like, 0.9003 or 0.9002924, but it won't matter.) Of course, we know the standard deviation was about 0.9 because that's the square root of the variance (0.81) that I provided to you. So you didn't actually need to type in any of the data to get what you needed.

Thus, the Z score for the statue is

$$(9.33 - 15.2)/0.9 = -6.522$$

And so Z = -6.5 for the Statue of Liberty after rounding to one decimal place. Note that if we'd used either 0.9003 for the standard deviation, or even 0.9002924, our Z score would still round to -6.5.

7 Results

When this quiz was given...

- more than 90% of students got Question 1 right
- about 3/4 of available points were awarded for Question 2
- about 7/8 of students got Question 3 right
- just over 1/2 of available points were awarded for Question 4 (lots of people just gave `facet_wrap(~ NYHA)` as a response and so got some partial credit, and others used `nyha` rather than `NYHA` which also lost some credit)
- about 3/5 of students got Question 5 right (some people lost credit for adding extra decimal places)