431 Quiz 1 for Fall 2023

Deadline: Tuesday 2023-10-10 at Noon

Thomas E. Love

2023-10-08

Table of contents

Instructions for Students

There are 26 questions on this Quiz and this PDF is 32 pages long. Be sure you have all 32 pages. It is to your advantage to answer all 26 questions. Your score is based on the number of correct responses, so there's no chance a blank response will be correct, and a guess might be, so you should definitely answer all of the questions.

This is an open book, open notes quiz. You are welcome to consult the materials provided on the course website and that we've been reading in the class, but you are not allowed to discuss the questions on this quiz with anyone other than Professor Love and the teaching assistants. You will be required to complete a short affirmation that you have obeyed these rules as part of submitting the Quiz.

0.1 The Google Form Answer Sheet

All of your answers should be placed in the Google Form Answer Sheet, located at…

- https://bit.ly/431-2023-quiz1-form

All of your answers must be submitted through the Google Form by noon on Tuesday 2023-10-10, without exception. The form will close at 1 PM on that date, and no extensions will be made available, so do not wait until late in the morning on Tuesday to submit your work. We will only accept responses through the Google Form.

The Google Form contains places to provide your responses to each question, and a final affirmation where you'll type in your name to tell us that you followed the rules for the Quiz. You must complete that affirmation before you can submit your responses. When you submit your results (in the same way you submit a Minute Paper) you will receive an email copy of your submission, with a link that will allow you to edit your work.

If you wish to work on some of the quiz and then return later, you can do this by [1] completing the final question (the affirmation) which asks you to type in your full name, and then [2] submitting the quiz. You will then receive a link at your CWRU email which will allow you to return to the quiz as often as you like without losing your progress.

0.2 The Data Sets

I have provided three data sets (called `algae.csv`, `newborn.csv`, and `sleep.csv`) that are mentioned in the Quiz. They may be helpful to you.

In Question 10, I refer to a data set (`nnyfs.Rds`) which you have access to on our 431-data page, and that file will also be useful.

0.3 Getting Help

If you need clarification on a Quiz question, you have exactly one way of getting help:

1. Ask your quiz question via email to **431-help at case dot edu**.

During the Quiz period (5 PM 2023-10-05 through noon 2023-10-10) we will not answer questions about the Quiz through Campuswire or in TA office hours. Instead, we will only answer them through the email address listed above. We promise to respond to all questions received before 9 AM on 2023-10-10 in a timely fashion.

A few cautions:

- Specific questions are more likely to get helpful answers.
- We will not review your code or your English for you.
- We will not tell you if your answer is correct, or if it is complete.
- We will email all students if we find an error in the Quiz that needs fixing.

0.4 When Should I ask for help?

We recommend the following process.

- If you encounter a tough question, skip it, and build up your confidence by tackling other questions.
- When you return to the tough question, spend no more than 10-15 minutes on it. If you still don't have it, take a break (not just to do other questions) but an actual break.
- When you return to the question, it may be much clearer to you. If so, great. If not, spend 5-10 minutes on it, at most, and if you are still stuck, ask us for help.
- This is not to say that you cannot ask us sooner than this, but you should **never, ever** spend more than 20 minutes on any question without asking for help.

0.5 Scoring and Timing

All questions are worth either 3 or 4 points, adding to a total of 100 points. The four questions which are worth 3 points each (specifically, Questions 5, 12, 21 and 25) are marked as such in the Quiz. The questions are not in any particular order, and range in difficulty from "things I expect everyone to get right" to "things that are deliberately tricky".

The Quiz is meant to take 4 hours. I expect most students will take 3-5 hours, and some will take as little as 2 or as many as 8. It is not a good idea to spend a long time on any one question.

0.6 Writing Code into the Google Form

Occasionally, we ask you to provide a single line of code. If not otherwise specified, a single line of code in response can contain **at most** two pipes, although you may or may not need the pipe in any particular setting. Note that I exclusively used the `|>` pipe, and not the `%>%` pipe, in developing this Quiz.

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

R Packages and Love-boost.R script

This doesn't mean you need to use all of these packages, or indeed, that I did in building the Quiz and its answer sketch.

```
library(broom)
library(Epi)
library(ggrepel)
library(glue)
library(gt)
library(gtExtras)
library(Hmisc)
library(janitor)
library(mosaic)
library(naniar)
library(patchwork)
library(simputation)
library(xfun)
library(tidyverse)

source("Love-boost.R")

theme_set(theme_bw())
knitr::opts_chunk$set(comment = NA)
```

1 Question 1.

In the introduction and Chapters 1-4 of *The Art of Statistics*, David Spiegelhalter describes several common features of a strong visualization of data, including some identified by Alberto Cairo. Which of the following features are described as being characteristic of high-quality work in this regard? (CHECK ALL THAT APPLY.)

a. The graph's design is chosen so that relevant patterns become noticeable.
b. The graph contains reliable information.
c. The graph's appearance is presented in an attractive way.
d. The graph is honest, clear, and contains deep insights.
e. The graph should never connect data points gathered at different times.
f. The graph is accompanied by a meaningful title, and clear labels and captions.
g. The graph is of the raw data only.
h. The graph helps to raise more questions and encourage the reader to explore.

2 Question 2.

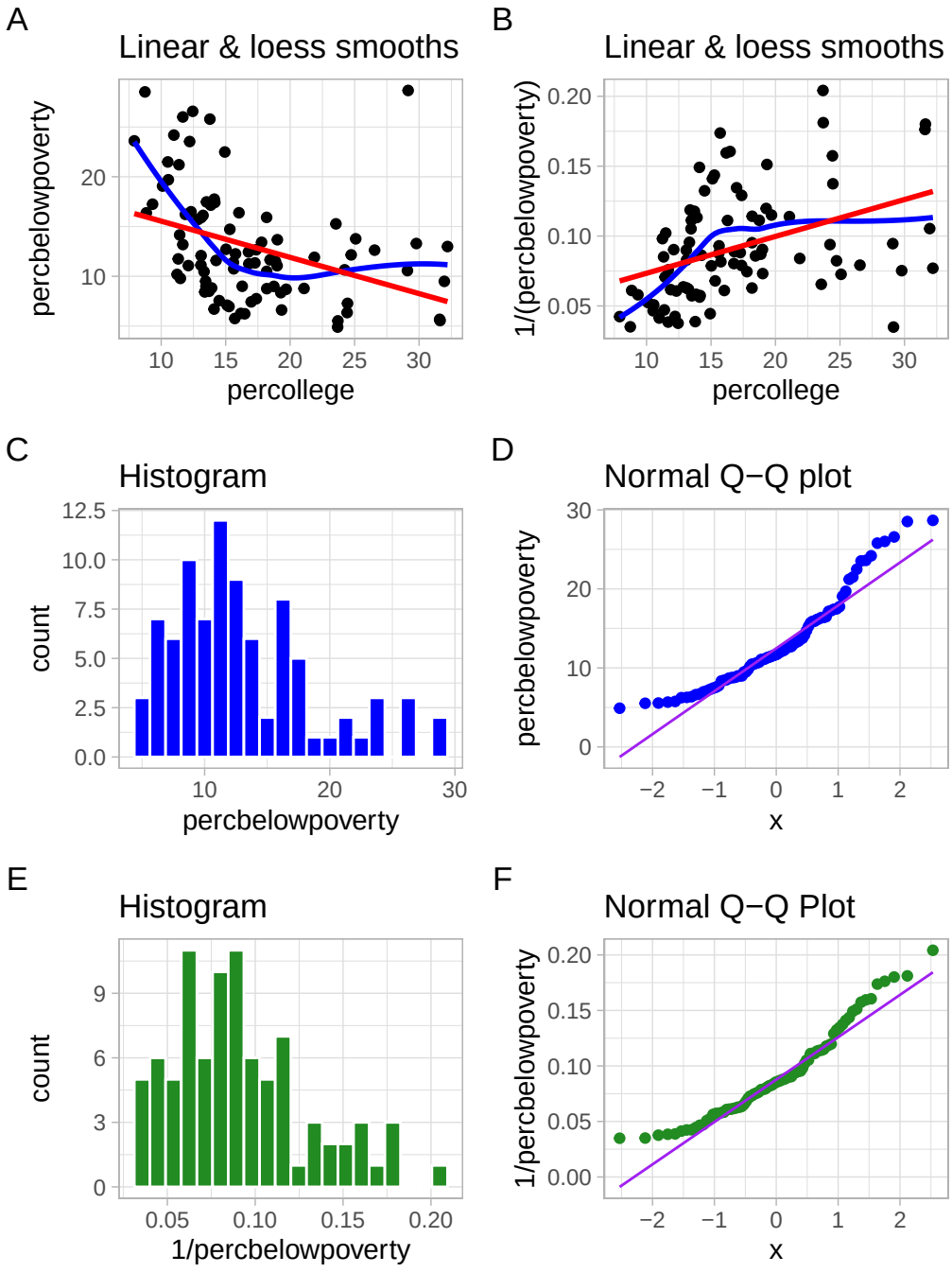
Suppose you are using a subset of the `midwest` data from the `ggplot2` package. You are trying to determine for this subset whether or not a transformation of the outcome (specifically, taking the inverse of the outcome) is necessary to fit a linear regression model to describe the relationship between `percollege` (the predictor, specifically the percent college educated) and `percbelowpoverty` (the outcome, specifically the percent below the poverty level). Which of the Plots shown in the Figure for Question 2 would be of the most help in assessing whether using this transformation would improve the assumption of linearity?

a. Plot A
b. Plot B
c. Plots C and D
d. Plots E and F
e. They would all be equally useful

The Figure for Question 2 is shown on the next page.

Continuation of Question 2

Figure for Question 2



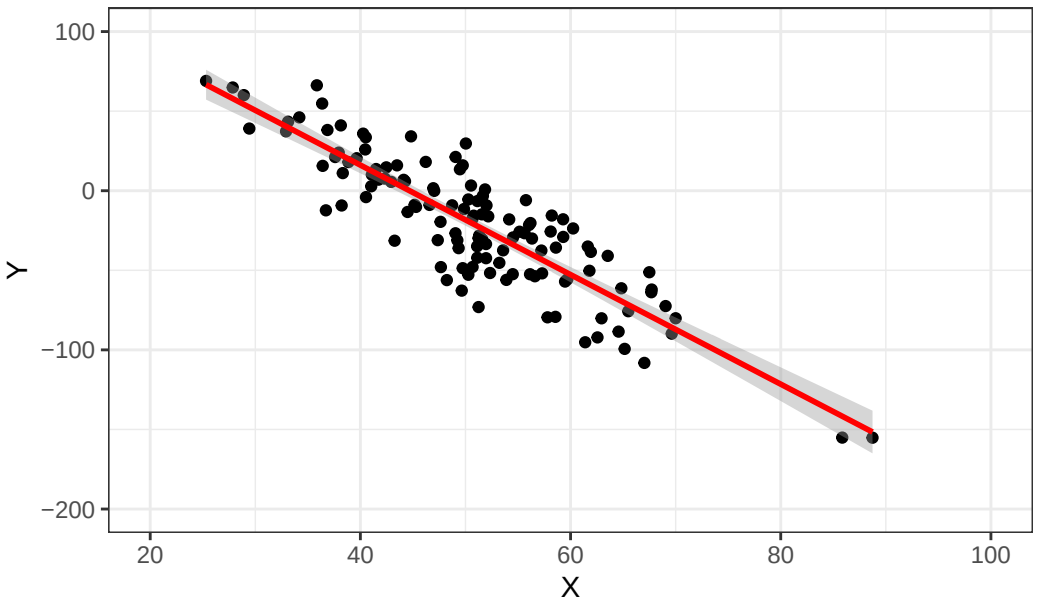
3 Question 3.

Consider the following possible summaries of a linear model fit to predict Y from X, describing the scatterplot shown in the Figure for Question 3. Which of these summaries is correct?

As usual, the R-squared value is defined here as the square of the Pearson correlation coefficient between the outcome and the predictor.

a. Model: y = 3.4 + 154 x, with R-squared = -0.76
b. Model: y = 3.4 - 154 x, with R-squared = -0.26
c. Model: y = -3.4 + 154 x, with R-squared = 0.76
d. Model: y = -3.4 + 154 x, with R-squared = 0.26
e. Model: y = 3.4 + 154 x, with R-squared = 0.76
f. Model: y = 3.4 + 154 x, with R-squared = 0.26
g. Model: y = 154 - 3.4 x, with R-squared = -0.76
h. Model: y = 154 - 3.4 x, with R-squared = -0.26
i. Model: y = 154 + 3.4 x, with R-squared = 0.76
j. Model: y = 154 + 3.4 x, with R-squared = 0.26
k. Model: y = 154 - 3.4 x, with R-squared = 0.76
l. Model: y = 154 - 3.4 x, with R-squared = 0.26

Figure for Question 3



4 Question 4.

The `starwars` data set is part of the `dplyr` package loaded by the tidyverse. You can learn more about it at https://dplyr.tidyverse.org/reference/starwars.html if you like. In that data, we find information on Star Wars characters, specifying 14 different variables, including each character's `homeworld` and `gender`. After loading the packages used in developing this Quiz (specified in the Instructions for Students), try running the following code:

```
starwars |>
  select(name, hair_color, birth_year, homeworld, gender) |>
  gt()|> tab_header("Some Characters in Star Wars")
```

Note that I'm not showing all of the rows in the table here, to save some space.

Some Characters in Star Wars

name	hair_color	birth_year	homeworld	gender
Luke Skywalker	blond	19.0	Tatooine	masculine
C-3PO	NA	112.0	Tatooine	masculine
R2-D2	NA	33.0	Naboo	masculine
Darth Vader	none	41.9	Tatooine	masculine
Leia Organa	brown	19.0	Alderaan	feminine
Owen Lars	brown, grey	52.0	Tatooine	masculine
Beru Whitesun lars	brown	47.0	Tatooine	feminine
R5-D4	NA	NA	Tatooine	masculine
Biggs Darklighter	black	24.0	Tatooine	masculine
Obi-Wan Kenobi	auburn, white	57.0	Stewjon	masculine

Your job is to modify the code I've provided to produce a new table which includes only those characters in the `starwars` data set that have:

1. brown hair (do not include people with both brown and grey hair, for example)
2. a birth year of 15 or higher (years are measured here before the Battle of Yavin)
3. a known, non-missing, homeworld.

Now, review your new table, and answer these questions:

a. How many characters of feminine gender appear in your new table?
b. How many characters of masculine gender appear in your new table?

5 Question 5. (3 points)

In this question, we consider data describing the age at onset (in years) for 17 women with a diagnosis of multiple sclerosis. The oldest age at onset was 44 years. The stem-and-leaf display shows the data for the first 17 subjects.

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

1 | 46788889
2 | 0367
3 | 239
4 | 24
```

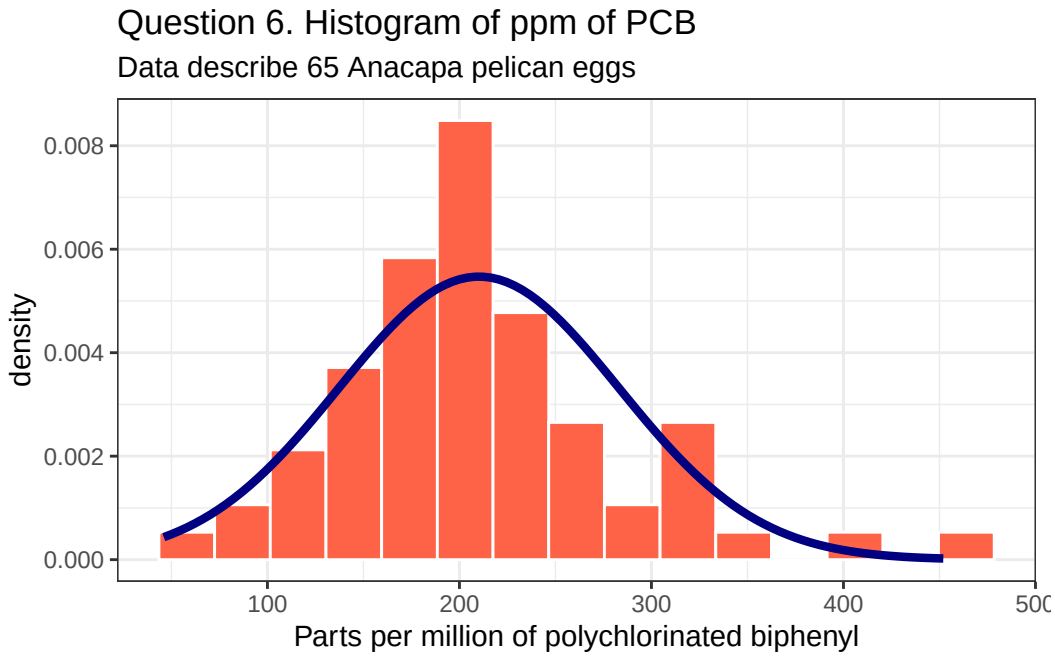
If the next subject added to the data is 28 years of age, which of the following values will decrease, as a result?

```
I. The mean
II. The standard deviation
III. The median
```

a. I only
b. II only
c. III only
d. I and II
e. I and III
f. II and III
g. All three statements
h. None of the three statements

6 Question 6.

The data for this Question represent the concentration in parts per million of PCB (polychlorinated biphenyl, an industrial pollutant) for 65 Anacapa pelican eggs. The tibble containing the data is called `pelican` and the variable of interest is called `ppm`. I have not provided you with these data.



Here are eight lines of code. Note that Dr. Love definitely used lines 1, 2 and 8 in his code. He also used some of the other lines (lines 3-7) but not all of them.

```
1 pelican <- read_csv("data/pelican.csv", show_col_types = FALSE)

2 ggplot(pelican, aes(x = ppm)) +
3   geom_density(col = "navy", lwd = 1.5) +
4   geom_histogram(aes(y = after_stat(density)), bins=15, fill="tomato", col="white") +
5   geom_histogram(bins = 15, fill = "tomato", col = "white") +
6   stat_function(fun = dnorm,
                  args = list(mean = mean(pelican$ppm), sd = sd(pelican$ppm)),
                  col = "navy", lwd = 1.5) +
7   coord_flip() +
8   labs(title = "Question 6. Histogram of ppm of PCB",
         subtitle = "Data describe 65 Anacapa pelican eggs",
         x = "Parts per million of polychlorinated biphenyl")
```

Continuation of Question 6

Please select each of the line numbers that should be REMOVED from the code in order to create the Question 6 plot. (YOU MAY SELECT MORE THAN ONE OPTION.)

a. Line 3
b. Line 4
c. Line 5
d. Line 6
e. Line 7

7 Question 7.

Suppose you are interested in how effectively shell thickness might be used to predict the concentration of environmental pollutants, in a setting like the study developed in Question 6. Which variable should go on the vertical (Y) axis of your scatterplot to display and model this association?

a. the concentration in parts per million of PCB
b. the thickness in micrometers of the egg's shell
c. the egg identification number (1-65)
d. It doesn't matter.
e. It is impossible to tell from the information provided.

8 Question 8.

Sir Austin Bradford Hill's criteria have been a building block for scientific work for over 50 years. In *The Art of Statistics*, David Spiegelhalter outlines recent work by Jeremy Howick and colleagues to separate these criteria into direct, mechanistic and parallel evidence.

For each description below, identify whether what is provided is best categorized as direct, mechanistic or parallel evidence.

Columns:

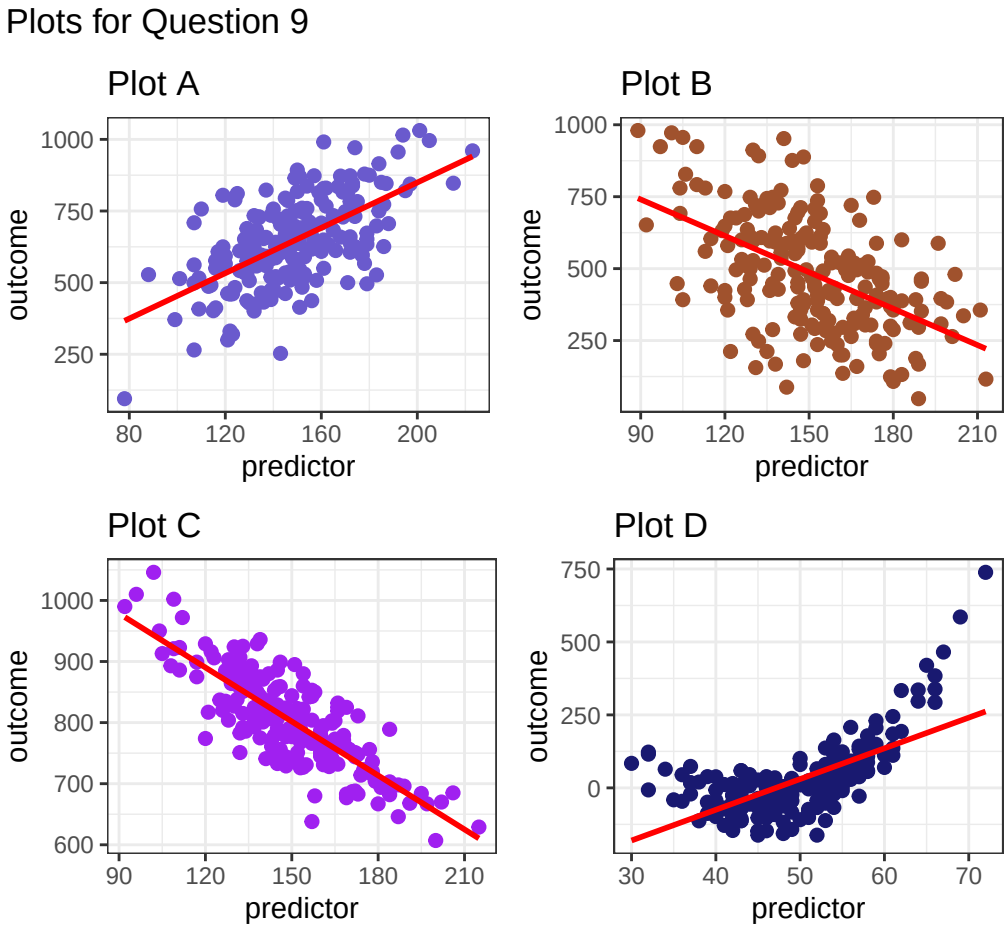
1. direct
2. mechanistic
3. parallel

Rows:

a. The observed effect is consistent with what is already known.
b. The observed effect is consistent with a plausible biological mechanism.
c. The observed effect is seen again in a new, similar study.
d. The observed effect is smaller in subjects who are less exposed.
e. The observed effect is preceded in time by the presumed cause.

9 Question 9.

Consider the four scatterplots provided for Question 9.



Which of the four scatterplots provided for Question 9 is associated with a linear model for `outcome` using `predictor` that has the largest R-square value?

a. Plot A
b. Plot B
c. Plot C
d. Plot D

10 Question 10.

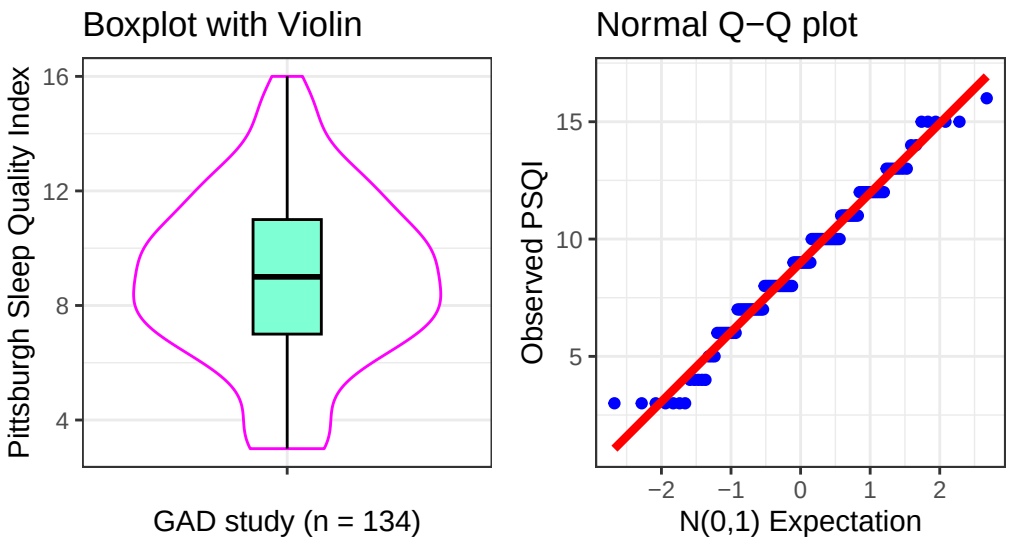
On our 431-data page, in the data-and-code subfolder, you'll find a file called `nnyfs.Rds`. Import that file into R, and then use it to specify answers to the following four questions:

a. How many variables are there in the `nnyfs` tibble?
b. Which variable has the most missingness of the variables in the `nnyfs` tibble?
c. How many missing observations exist in the variable you identified in part b of this question?
d. How many observations in the `nnyfs` tibble have complete data on all listed variables?

11 Question 11.

The Pittsburgh Sleep Quality Index (PSQI) is a self-rated 19-item questionnaire to assess sleep quality and disturbances over the past month. The PSQI yields a global score which ranges from 0 to 21, with higher scores indicating greater overall sleep disturbance. A study of older adults with a diagnosis of generalized anxiety disorder (GAD) yielded a sample of 134 subjects stored in a file provided to you called `sleep.csv`. I used the `sleep` data to produce the Figure for Question 11. Before building this plot, I loaded all necessary packages.

Figure for Question 11



Which of the following bits of R code were **NOT** used in generating the Figure for Question 11? (CHECK ALL THAT APPLY.)

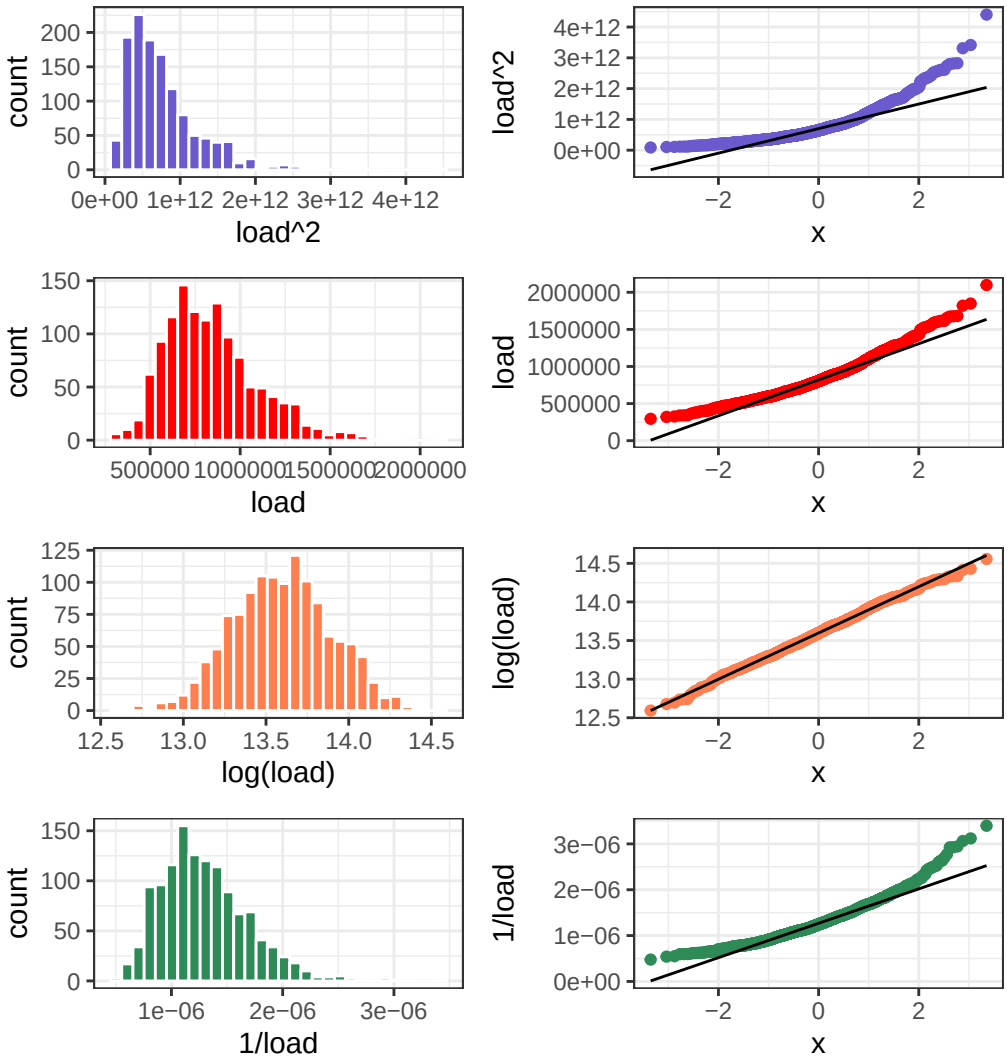
a. `labs(y = "Observed PSQI", x = "N(0,1) Expectation", title = "Normal Q-Q plot")`
b. `geom_boxplot(fill = "aquamarine", col = "black", width = 0.2)`
c. `(p1 / p2) + plot_annotate(title = "Figure for Question 11")`
d. `labs(x = "GAD study (n = 134)", y = "Pittsburgh Sleep Quality Index", title = "Boxplot with Violin")`
e. `p2 <- ggplot(data = sleep, aes(sample = psqi))`
f. `geom_qq(col = "blue") + geom_qq_line(col = "red", lwd = 1.5)`
g. `geom_violin(col = "magenta", fill = "white")`
h. `p1 <- ggplot(data = sleep, aes(x = "", y = psqi))`
i. Check this option only if all the bits of code listed in options a-h were used.

12 Question 12 (3 points)

1,251 subjects were given a hepatitis C RNA quantitative test which measured the amount of Hep C virus present in their blood, in IU/ml, and this is called the viral load. Anything over 800,000 is usually considered high, and anything under that is low. Those with low viral load have a better chance of responding to treatment. Consider the Figure for Question 12.

Figure for Question 12

Exploring Viral Load Power Transformations



Question 12 continues on the next page.

Continuation of Question 12

If our goal is to obtain a transformation of the data which is well fit by a Normal model, which of the following options appears to be our best choice?

a. Taking the square of the viral load.
b. Taking the viral load, untransformed.
c. Taking the natural logarithm of the viral load.
d. Taking the inverse of the viral load.
e. None of these options.

13 Question 13

I have provided you with a data set called `newborn.csv`. After you import that into R as a tibble called `newborn`, the result should contain a variable called `apgar5` that contains scores on the APGAR scale at five minutes for 130 infants, although 4 of the values are listed as NA.

You wish to obtain the standard deviation of the APGAR scores in the `newborn` tibble. If you need to know more about the APGAR score, visit https://goo.gl/9rxkVU. Your task is to mark the box next to **each** of the R commands listed below that produce (perhaps along with other things) the **sample standard deviation** of APGAR scores at five minutes for the 126 infants not marked as NA. (CHECK ALL THAT APPLY.)

a. `favstats(~ apgar5, data = newborn)`
b. `summary(newborn)`
c. `sd(newborn$apgar5)`
d. `newborn |> summarise(sd(apgar5, na.rm = TRUE))`
e. `newborn |> filter(complete.cases(apgar5)) |> summarise(sd = sd(apgar5))`
f. `newborn |> select(complete.cases(apgar5)) |> summarise(sd = sd(apgar5))`
g. None of these will produce the correct value.

14 Question 14.

Suppose you have collected data as part of a cohort study to look at the impact of exposure to an industrial solvent (which is stored in a four-level character variable called `solvent` which can be either none, modest, moderate or profound) on the probability of a bladder cancer diagnosis (stored as a three-level character variable called `diagnosis` which can be either definite, possible, or no.)

You can assume that a tibble containing these variables called `q14` is available to you in R, and that the packages loaded by Dr. Love at the start of this Quiz are also already loaded for you, so you don't need to load them again and there should be no `library()` calls in your response.

Provide a single line of R code (you may use **at most** two pipes) to obtain an appropriate numerical summary of the relationship between the `solvent` and `diagnosis` variables in the `q14` tibble.

15 Question 15.

Suppose now that in continuing your work on the study from Question 14, you now have more granular information on the exposure level to the solvent. Specifically, you now have an `exposure` measure, expressed as the percentage of the Occupational Safety and Health Administration (OSHA) recommended exposure limit, so that 100 = the recommended exposure limit for this solvent, and values above 100 indicate exposures that exceed that limit, while values below 100 indicate exposures that are at least somewhat "safe".

You can assume that a new tibble, called `q15` is available to you in R containing this `exposure` measure as well as the bladder cancer `diagnosis` variable described in Question 14.

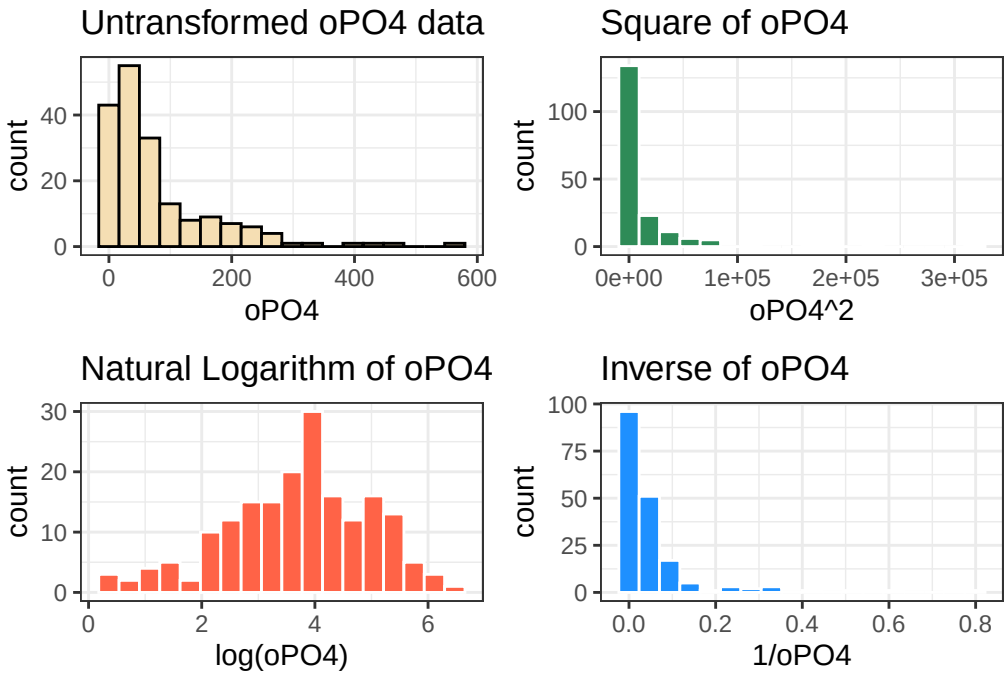
Again, you can also assume that the packages loaded by Dr. Love at the start of this Quiz are also already loaded for you, so you don't need to load them again. You may assume that all of the R packages Dr. Love has asked you to install for this course are installed, as well.

Provide a single line of R code (you may use at most two pipes) to obtain an appropriate numerical summary description of the distribution of `exposure` within each `diagnosis` group in the `q15` tibble.

16 Question 16.

Our next data set is called `algae.csv` and it is available to you with the Quiz materials. This data file describes 200 water samples collected from the same river over a period of 3 months, of which 184 have complete data. Each of those 184 observations contains information on a series of chemical parameters measured in the water samples, one of which is the mean value of orthophosphate, contained in the variable `oPO4`.

Figure for Question 16



Consider the histograms shown in the Figure for Question 16, and suppose your goal is to approximate a Normal distribution with some transformation of the `oPO4` data. Which of the following options describes the most logical transformation to use in trying to accomplish this goal?

a. The square of the oPO4 data
b. The natural logarithm of the oPO4 data
c. The inverse of the oPO4 data
d. The untransformed oPO4 data
e. It is impossible to tell from the information provided.

17 Question 17.

Return to the `algae.csv` file I provided to you, and fit a linear model to predict the natural logarithm of the algae frequency `a1` using the natural logarithm of the `oPO4` (orthophosphate) in the same water sample.

You will have to manage the data to use only those samples with complete data on both variables in your model, and which have values of `a1` that exceed zero.

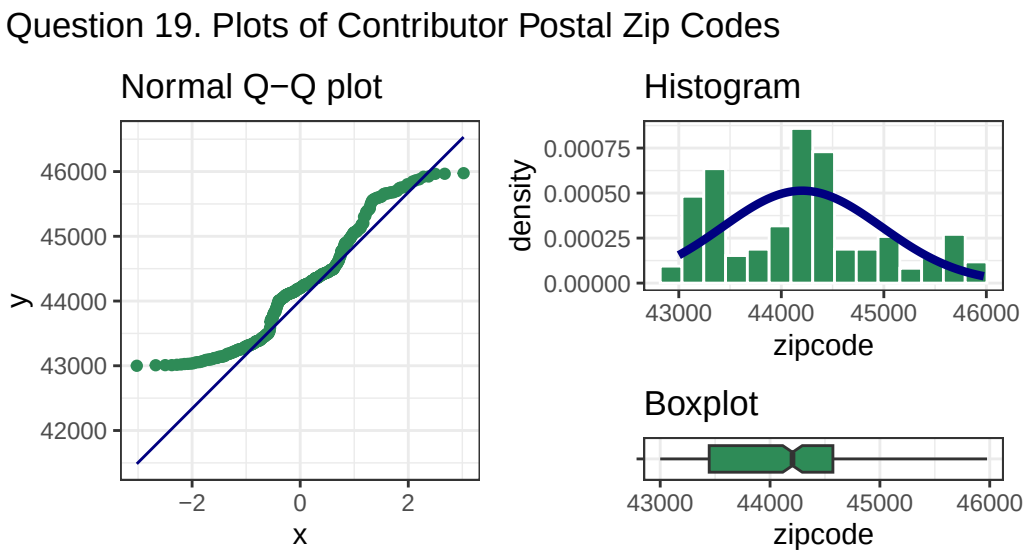
How many water samples are included in your model?

18 Question 18.

Based on your model described in Question 17, what is the predicted value of the actual algae frequency `a1` (be careful: what does your model predict?) for a sample with an `oPO4` value of 64? Round your response to a single decimal place.

19 Question 19.

The plot for Question 19 displays the postal zip codes of a sample of 400 Ohio residents who made a disclosed individual financial contribution to a candidate in the 2020 presidential election.



Which of the following summaries of these data would be most appropriate?

a. The mean.
b. The median.
c. The interquartile range.
d. The mode.
e. It is impossible to tell.

20 Question 20.

The table below shows the most recently measured body-mass index (BMI) of the 15 female patients that are scheduled to be seen this afternoon by a nurse practitioner for primary care of their chronic illness.

Patients scheduled for this afternoon

Patient	BMI (kg/m^2)	Height (m)	Weight (kg)
Allen, L	47.162534	1.65	128.4
Bieber, S	47.122586	1.63	125.2
Carrasco, C	38.220022	1.82	126.6
Civale, A	37.857802	1.71	110.7
Hand, B	34.726353	1.64	93.4
Hill, C	31.000918	1.65	84.4
Karinchak, J	30.884474	1.58	77.1
Maton, P	30.035003	1.63	79.8
McKenzie, T	29.703632	1.73	88.9
Perez, O	28.040197	1.63	74.5
Plesac, Z	27.952452	1.57	68.9
Plutko, A	27.813209	1.68	78.5
Quantrill, C	25.607639	1.44	53.1
Rodriguez, J	25.254996	1.63	67.1
Wittgren, N	20.974482	1.57	51.7
—	—	—	—
Average for these 15 Patients	32.534331	1.64	87.2
Practice Average across all Female Patients	31.169029	1.62	81.8

In one complete English sentence, suggest a worthwhile improvement to this table.

21 Question 21. (3 points)

The initial results of the Systolic Blood Pressure Intervention Trial (SPRINT) received a lot of attention.

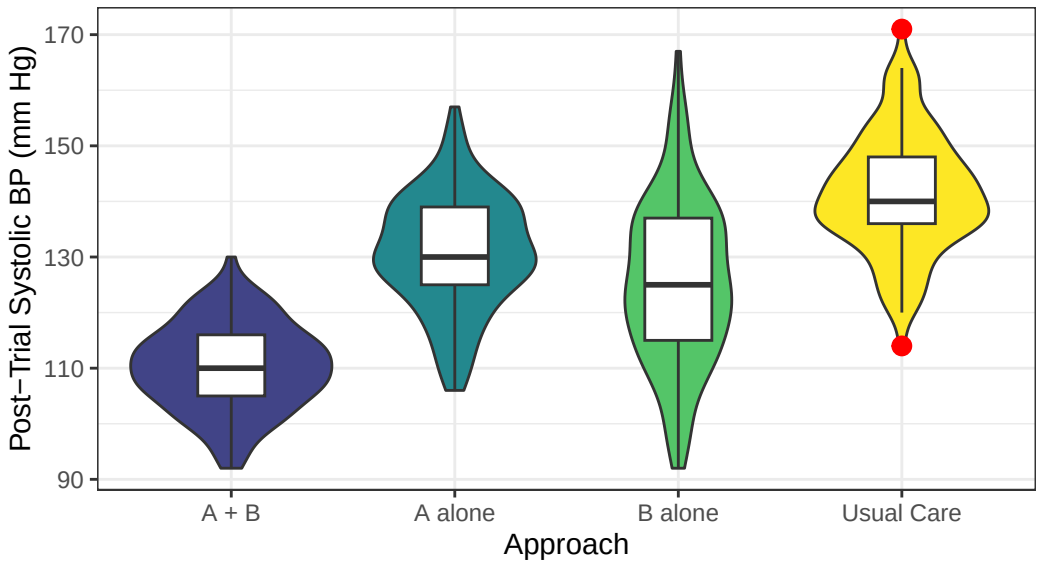
Quoting a press release from the National Institutes of Health:

> SPRINT evaluates the benefits of maintaining a new target for systolic blood pressure, the top number in a blood pressure reading, among a group of patients 50 years and older at increased risk for heart disease or who have kidney disease. A systolic pressure of 120 mm Hg, maintained by this more intensive blood pressure intervention, could ultimately help save lives among adults age 50 and older who have a combination of high blood pressure and at least one additional risk factor for heart disease, the investigators say.

Consider a hypothetical trial, where two different interventions are studied to see whether patients in another population besides that studied in SPRINT may have their blood pressure effectively managed to fall at the target level (120 mm Hg or lower).

500 patients were included in this trial, and were randomly allocated (125 to each intervention) so that we have 125 patients receiving both interventions A and B, 125 receiving A alone, 125 receiving B alone, and 125 receiving usual care (neither A nor B). The post-trial Systolic Blood Pressure results for all 500 patients are shown in the Figure for Question 21.

Figure for Question 21



Question 21 continues on the next page.

Continuation of Question 21

Consider the following statements:

I. The group of patients receiving usual care had the smallest number of patients with SBP at 120 or lower after the trial.

II. The group of patients receiving B alone had the largest spread in their distribution of post-trial systolic blood pressures.

III. The group of patients receiving both A and B had more than 90 patients with post-trial SBP at 120 or lower.

Which of these statements are true?

a. I only
b. II only
c. III only
d. I and II
e. I and III
f. II and III
g. All three statements
h. None of the three statements

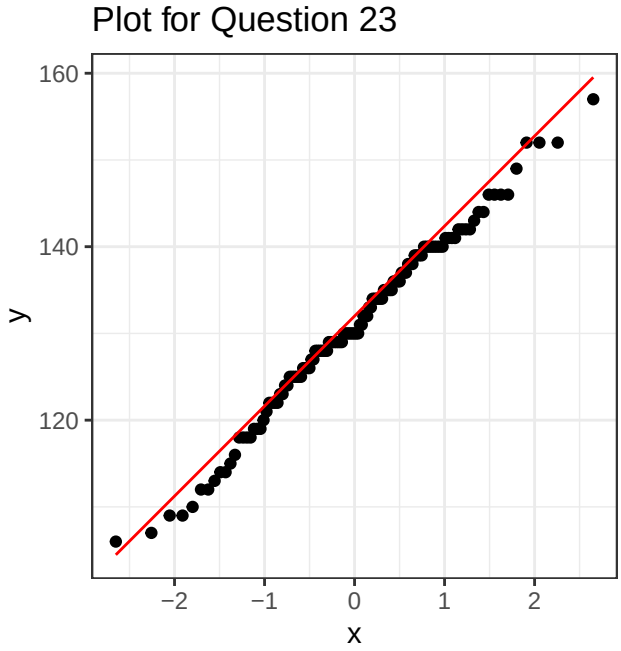
22 Question 22.

Which of the four blood pressure trial groups discussed in Question 21 produced the individual subject with the lowest post-trial systolic blood pressure?

a. The group receiving A alone
b. The group receiving B alone
c. The group receiving usual care
d. The group receiving both A and B
e. It is impossible to tell from the information provided.

23 Question 23.

The normal Q-Q plot shown here is taken from one of the four blood pressure trial groups discussed in Questions 21 and 22. Which one?



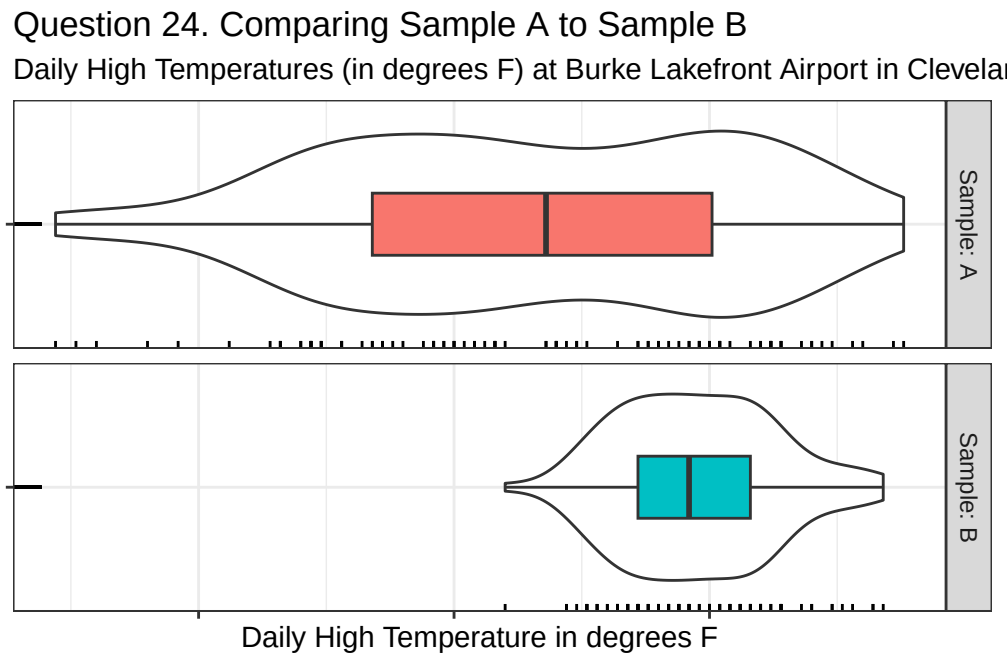
a. The group receiving A alone
b. The group receiving B alone
c. The group receiving usual care
d. The group receiving both A and B

24 Question 24.

The plot for Question 24 shows the high daily temperatures (in degrees Fahrenheit) measured at Burke Lakefront Airport in Cleveland, Ohio in two groups of dates, drawn from the past few years.

- One of the samples was formed from a random selection of 100 dates in the month of September.
- The other sample includes a random selection of 100 dates from the entire year.

Unfortunately, the x-axis (which was the same for each subplot) was left unlabeled, **but the missing x-axis labels are the same** for each of the two samples of data. The plot below provides some evidence regarding the distributions of the two samples.



Which of the following statements are true?

```
I. Sample A describes the data gathered only in September.
II. The interquartile range in Sample A is wider than that of Sample B.
III. Sample A would be less accurately modeled using a Normal distribution than Sample B.
```

a. I only
b. II only
c. III only

d. I and II
e. I and III
f. II and III
g. All three statements
h. None of the three statements

25 Question 25. (3 points)

The process of inductive inference, as described in *The Art of Statistics*, requires us to think hard about how we move from looking at the raw data to making general claims about the target population. Consider the following principles of effective measurement in this context.

```
I. We want to actually measure what we really want to measure
   without introducing systematic bias.

II. We want to sample at random whenever possible from the
    available subjects we are trying to make inferences about.

III. We want to use measures that give us a good chance of getting
     a similar result in a new study using the same measures.
```

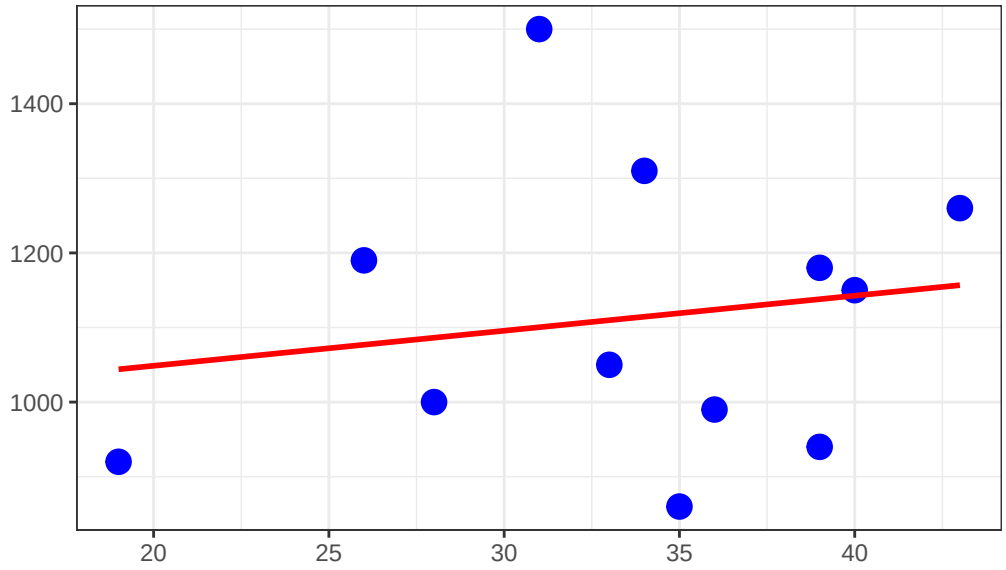
Each of the principles listed above is associated primarily with a particular step in the process of building inductive inference. Identify the step in the process associated with each of the statements above.

a. Moving from the raw data to the sample
b. Moving from the sample to the study population
c. Moving from the study population to the target population

26 Question 26.

Fast food is often high in both fat and sodium. But are the two related? The scatter plot shown in the Figure for Question 26 describes the fat (in g) and sodium (in mg) contents of twelve brands of hamburgers, and includes a linear model fit with geom_smooth, shown in red. In a sentence, what is the MOST IMPORTANT thing that should be done to improve the Figure for Question 26?

Figure for Question 26



This is the end of the Quiz.

Session Information

```
session_info()
```

```
R version 4.3.1 (2023-06-16 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 11 x64 (build 22621)


Locale:
  LC_COLLATE=English_United States.utf8
  LC_CTYPE=English_United States.utf8
  LC_MONETARY=English_United States.utf8
  LC_NUMERIC=C
  LC_TIME=English_United States.utf8

time zone: America/New_York
tzcode source: internal

Package version:
  abind_1.4.5             askpass_1.2.0           backports_1.4.1
  base64enc_0.1-3         bigD_0.2.0              bit_4.0.5
  bit64_4.0.5             bitops_1.0.7            blob_1.2.4
  boot_1.3.28.1           brio_1.1.3              broom_1.0.5
  bslib_0.5.1             cachem_1.0.8            callr_3.7.3
  car_3.1.2               carData_3.0.5           cellranger_1.1.0
  checkmate_2.2.0         class_7.3.22            cli_3.6.1
  clipr_0.8.0             cluster_2.1.4           cmprsk_2.2-11
  codetools_0.2.19        colorspace_2.1-0        commonmark_1.9.0
  compiler_4.3.1          conflicted_1.2.0        cpp11_0.4.6
  crayon_1.5.2            curl_5.1.0              data.table_1.14.8
  DBI_1.1.3               dbplyr_2.3.4            DEoptimR_1.1.2
  desc_1.4.2              diffobj_0.3.5           digest_0.6.33
  doRNG_1.8.6             dplyr_1.1.3             dtplyr_1.3.1
  e1071_1.7.13            ellipsis_0.3.2          Epi_2.47.1
  etm_1.1.1               evaluate_0.22           fansi_1.0.4
  farver_2.1.1            fastmap_1.1.1           fontawesome_0.5.2
  forcats_1.0.0           foreach_1.5.2           foreign_0.8-84
```

Formula_1.2-5	fs_1.6.3	gargle_1.5.2
generics_0.1.3	ggforce_0.4.1	ggformula_0.10.4
ggplot2_3.4.3	ggrepel_0.9.3	ggridges_0.5.4
ggstance_0.3.6	glmnet_4.1.8	glue_1.6.2
googledrive_2.1.1	googlesheets4_1.1.1	gower_1.0.1
graphics_4.3.1	grDevices_4.3.1	grid_4.3.1
gridExtra_2.3	gt_0.9.0	gtable_0.3.4
gtExtras_0.5.0	haven_2.5.3	highr_0.10
Hmisc_5.1-1	hms_1.1.3	htmlTable_2.4.1
htmltools_0.5.6	htmlwidgets_1.6.2	httr_1.4.7
ids_1.0.1	isoband_0.2.7	iterators_1.0.14
itertools_0.1.3	janitor_2.2.0	jquerylib_0.1.4
jsonlite_1.8.7	juicyjuice_0.1.0	knitr_1.44
labeling_0.4.3	labelled_2.12.0	laeken_0.5.2
lattice_0.21-8	lifecycle_1.0.3	lme4_1.1.34
lmtest_0.9.40	lubridate_1.9.3	magrittr_2.0.3
markdown_1.9	MASS_7.3-60	Matrix_1.6-1.1
MatrixModels_0.5.2	memoise_2.0.1	methods_4.3.1
mgcv_1.8-42	mime_0.12	minqa_1.2.6
missForest_1.5	modelr_0.1.11	mosaic_1.8.4.2
mosaicCore_0.9.2.1	mosaicData_0.20.3	munsell_0.5.0
naniar_1.0.0	nlme_3.1-162	nloptr_2.0.3
nnet_7.3-19	norm_1.0.11.1	numDeriv_2016.8-1.1
openssl_2.1.1	paletteer_1.5.0	parallel_4.3.1
patchwork_1.1.3	pbkrtest_0.5.2	pillar_1.9.0
pkgbuild_1.4.2	pkgconfig_2.0.3	pkgload_1.3.3
plyr_1.8.9	polyclip_1.10-6	praise_1.0.0
prettyunits_1.2.0	prismatic_1.1.1	processx_3.8.2
progress_1.2.2	proxy_0.4.27	ps_1.7.5
purrr_1.0.2	quantreg_5.97	R6_2.5.1
ragg_1.2.5	randomForest_4.7.1.1	ranger_0.15.1
rappdirs_0.3.3	RColorBrewer_1.1.3	Rcpp_1.0.11
RcppArmadillo_0.12.6.4.0	RcppEigen_0.3.3.9.3	reactable_0.4.4
reactR_0.4.4	readr_2.1.4	readxl_1.4.3
rematch_2.0.0	rematch2_2.1.2	reprex_2.0.2
rlang_1.1.1	rmarkdown_2.25	rngtools_1.5.2
robustbase_0.99.0	rpart_4.1.19	rprojroot_2.0.3
rstudioapi_0.15.0	rvest_1.0.3	sass_0.4.7
scales_1.2.1	selectr_0.4.2	shape_1.4.6
simputation_0.2.8	snakecase_0.11.1	sp_2.1.0
SparseM_1.81	splines_4.3.1	stats_4.3.1
stringi_1.7.12	stringr_1.5.0	survival_3.5-5
sys_3.4.2	systemfonts_1.0.4	testthat_3.1.10

textshaping_0.3.6 tibble_3.2.1 tidyr_1.3.0
tidyselect_1.2.0 tidyverse_2.0.0 timechange_0.2.0
tinytex_0.47 tools_4.3.1 tweenr_2.0.2
tzdb_0.4.0 UpSetR_1.4.0 utf8_1.2.3
utils_4.3.1 uuid_1.1.1 V8_4.3.3
vcd_1.4.11 vctrs_0.6.3 VIM_6.2.2
viridis_0.6.4 viridisLite_0.4.2 visdat_0.6.0
vroom_1.6.4 waldo_0.5.1 withr_2.5.1
xfun_0.40 xml2_1.3.5 yaml_2.3.7
zoo_1.8-12