

RStudio_091223

2023-09-12

In this worksheet, we will go over ways to quantify behavior and come up with ways to implement these measurements in our studies.

Problem Statement

While a program you have implemented was successful in increasing student focus, you are interested in more quantitative measurements. While listing out the percentage of students who remained focus during a class period is a start, your fellow researchers are interested in how exactly you measure focus. You want to decide which features to look out for and what measurements to make. But which ones?

Preliminaries

We will start by importing any libraries that may be useful.

```
# Import libraries  
library(ggplot2)
```

Quantifying Behavior

Quantifying behavior has many benefits. One primary benefit is that it puts a value to a feature. For example, a statement like “a student raises their hands a few times in class” sounds vague compared to “a student raises their hands once every five questions asked in class.” Having a quantity makes the observation objective rather than subjective, but another key benefit is that it is easier to find patterns around numbers. With numbers, we can create visuals and perform visual analysis on our data.

Question 1

What are some behaviors that you can quantify? Does quantifying these behaviors completely remove subjectivity in an experiment?

(Hint: What counts as a hand raise in the classroom? Does this idea differ from person to person?)

Question 2

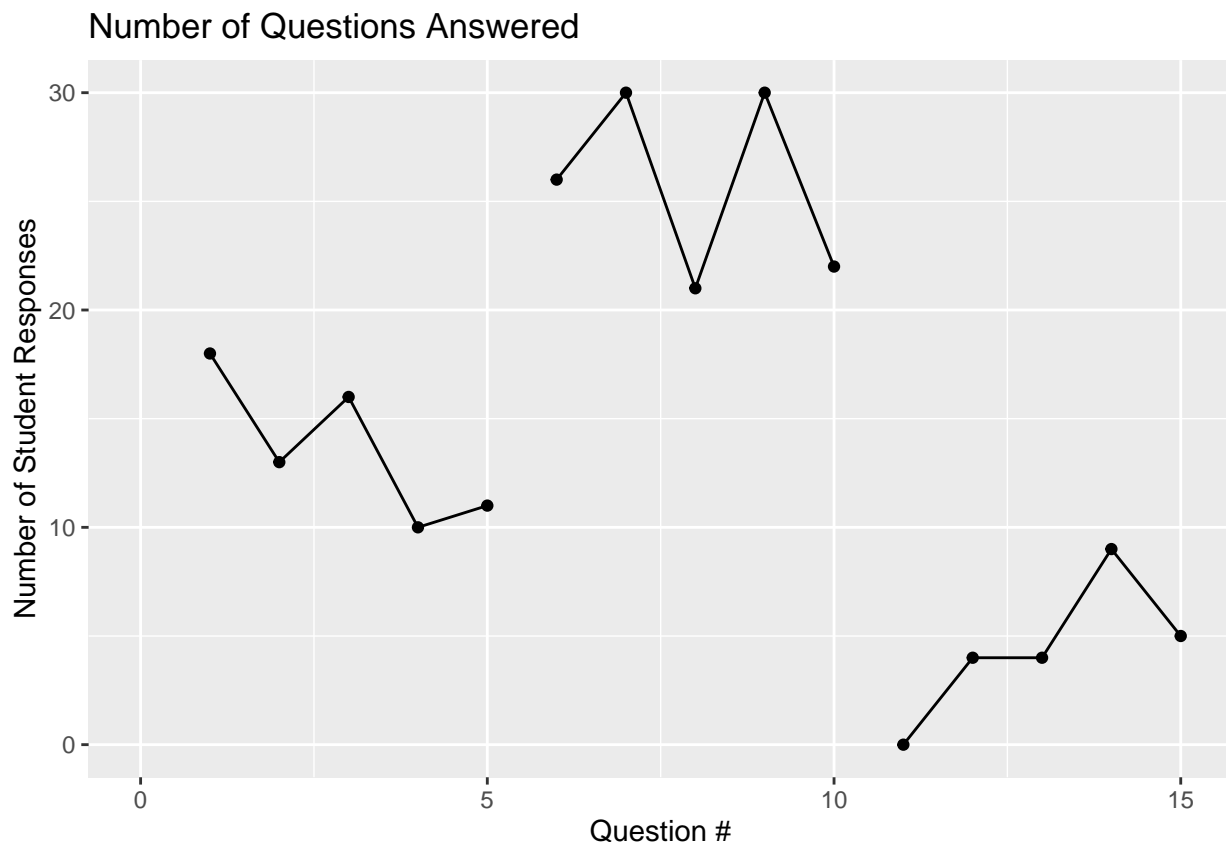
What are some behaviors that you cannot quantify? Should these behaviors not be considered in a study?

Possible Quantities to Consider

Most of the quantities to consider involve two factors: number and time. We like to count the number of events that happen along with recording when they happen. Let's look at a simulated dataset below. The students can respond using hand raises and clickers. There are a total of 15 questions. Students respond with a hand raise for questions 1-5 and 11-15 and with a clicker for questions 6-10.

```
# Generate random data
set.seed(1) # Set seed to make data reproducible
n = 5 # Number of days
A1 = sample(10:20, n, replace = TRUE)
B2 = sample(20:30, n, replace = TRUE)
A3 = sample(0:10, n, replace = TRUE)
data = c(A1, B2, A3)
time = seq(from = 1, to = length(data))
condition = c(rep("A1", n), rep("B2", n), rep("A3", n))
df = data.frame(data, time, condition)

# Make plot
ggplot(data = df, aes(x = time, y = data, group = condition)) +
  geom_line() +
  geom_point() +
  expand_limits(x = 0, y = 0) +
  xlab("Question #") +
  ylab("Number of Student Responses") +
  ggtitle("Number of Questions Answered")
```



Question 3

What is being measured in this study? Could there be a time component to consider in the study?

Question 4

Write some R code that calculates some descriptive statistics for each set of questions.

```
# Data
print(A1)
```

```
## [1] 18 13 16 10 11
```

```
print(B2)
```

```
## [1] 26 30 21 30 22
```

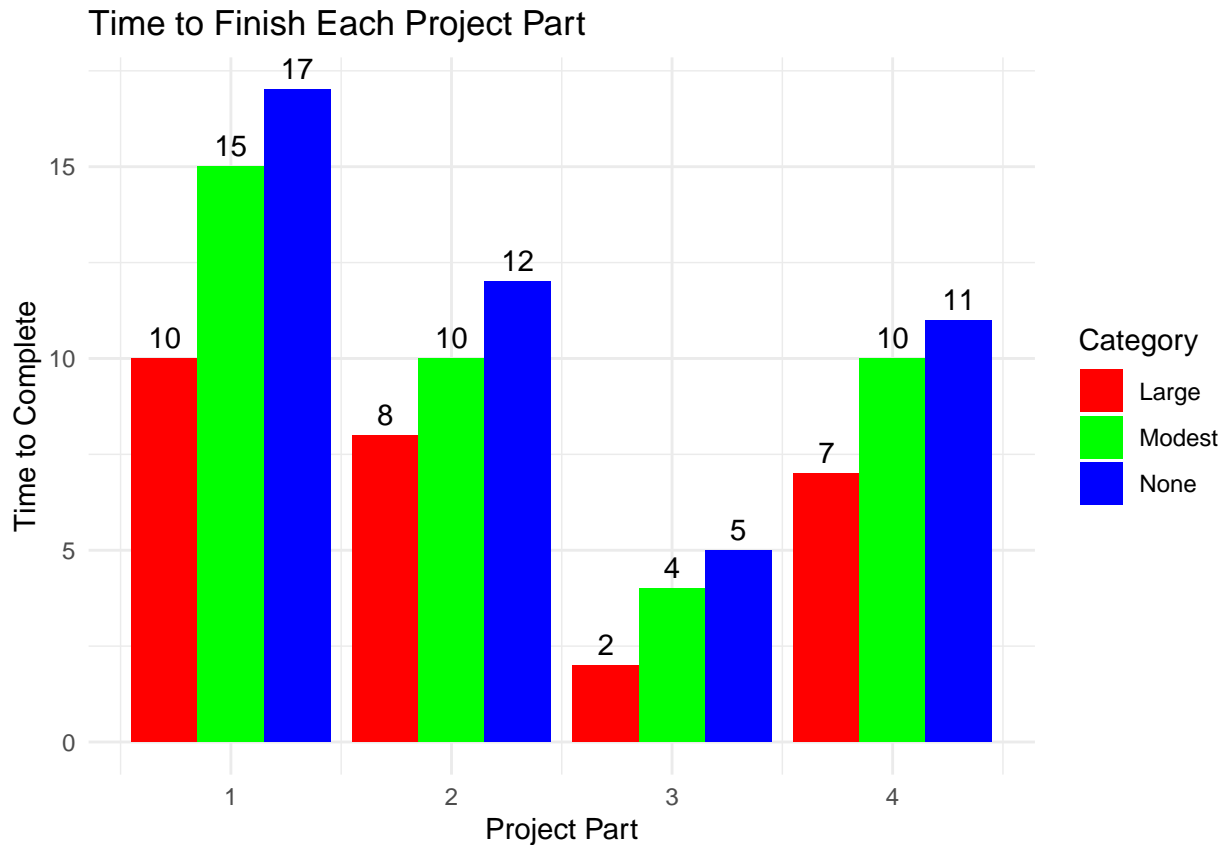
```
print(A3)
```

```
## [1] 0 4 4 9 5
```

Let's look at another simulated dataset below. In this study, 3 groups of student are given a project to complete in class. Each project is divided into 4 parts. At the end of each part, students receive a reward. One group of students is subject to large rewards, one group of students is subject to modest rewards, and one group of students is subject to no rewards.

```
# Create a sample data frame
data <- data.frame(
  Group = rep(1:4, each = 3),
  Category = rep(c("Large", "Modest", "None"), times = 4),
  Value = c(10, 15, 17, 8, 10, 12, 2, 4, 5, 7, 10, 11)
)

# Create the bar plot using ggplot2
ggplot(data, aes(x = Group, y = Value, fill = Category)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  labs(title = "Time to Finish Each Project Part",
       x = "Project Part",
       y = "Time to Complete") +
  scale_fill_manual(values = c("Large" = "red", "Modest" = "green", "None" = "blue")) +
  theme_minimal() +
  geom_text(aes(label = Value),
            position = position_dodge(width = 0.9),
            vjust = -0.5,
            size = 4)
```



Question 5

What conclusions can be drawn from this dataset? How would you determine that your conclusions are significant?

Question 6

The bars in Project Part 1 are notably higher and the bars in Project Part 3 are notably lower. What could be done to have a fair comparison between the different project parts?

The Time Component

Time is a tricky component when measuring behavior. Most commonly, time is used to record when an observation is made. However, it may be interesting to look at the frequency of observations and how long it takes to observe the next observation.

Suppose we have the simulated dataset below in which we relate the difficulty of questions, type of reward, and response latency.

```
# Create two sample datasets
set.seed(1)
x = 1:10
y1 = c()
for (i in 1:length(x)){
  y1 = c(y1, rexp(1, rate = 1/x[i]^2))
}
```

```

}
y2 = c()
for (i in 1:length(x)){
  y2 = c(y2, rexp(1, rate = 0.5/x[i]^2))
}

dataset1 <- data.frame(x, y1)

dataset2 <- data.frame(x, y2)

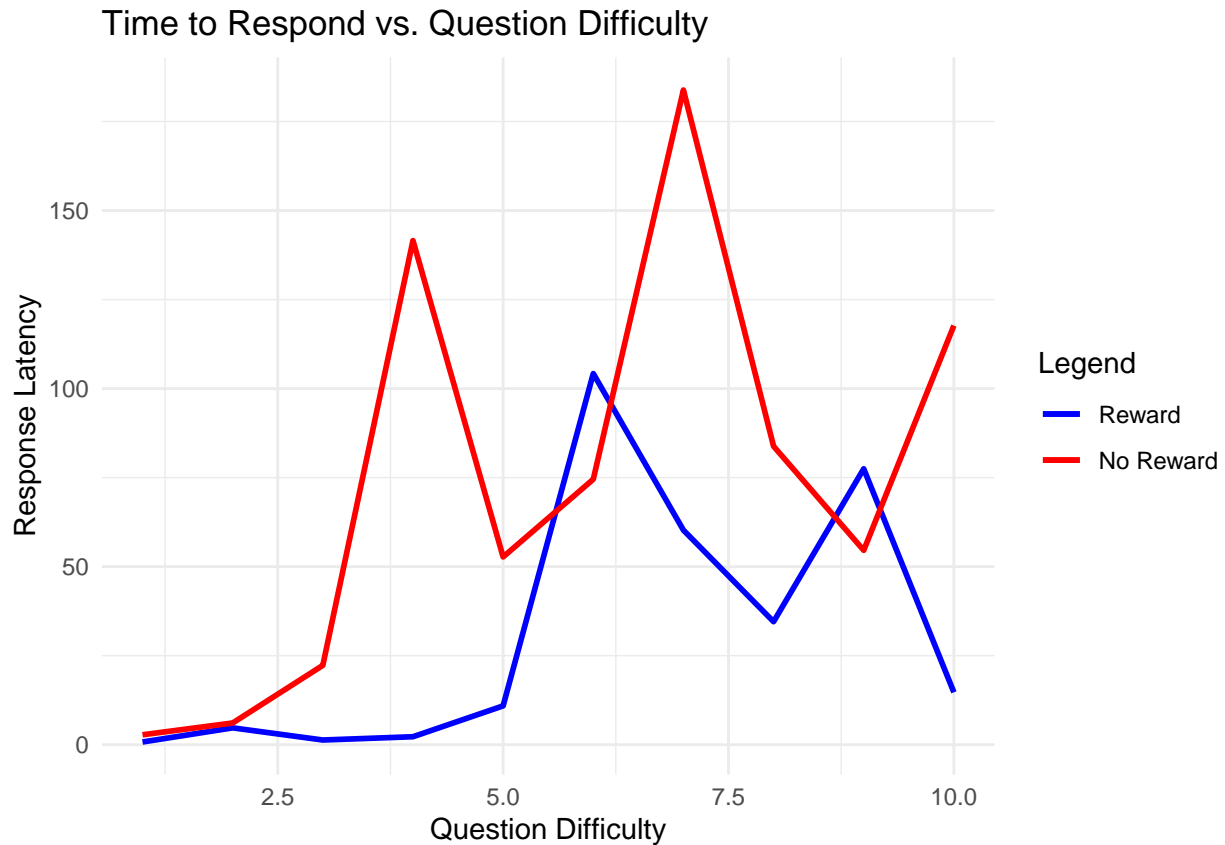
# Create a ggplot with two line plots
ggplot() +
  geom_line(data = dataset1, aes(x = x, y = y1, color = "Line 1"), size = 1) +
  geom_line(data = dataset2, aes(x = x, y = y2, color = "Line 2"), size = 1) +
  labs(
    title = "Time to Respond vs. Question Difficulty",
    x = "Question Difficulty",
    y = "Response Latency",
    color = "Legend"
  ) +
  scale_color_manual(
    values = c("Line 1" = "blue", "Line 2" = "red"),
    labels = c("Reward", "No Reward")
  ) +
  theme_minimal()

```

```

## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.

```



Question 7

Why might response latency be a useful measurement?

Question 8

Based on the figure above, can we definitely attribute differences in response latency to rewards, or is there another variable we should consider?

Question 9

Suppose the study above was conducted in the morning. What kind of results would you expect if the study was repeated in the afternoon?

Question 10

Based on your previous responses, create a list of behaviors that you would measure in your intervention. Consider the time component of these behaviors (e.g. frequency, time of day, etc.).