EDA_consulting

Maggie Sha

4/15/2021

1. Demographic data of respondents (INCLUDE only those who selected Public School)

a. World Language Teacher or Administrator
b. Language(s) Taught
c. Level(s) Taught (elem, middle, high)

2. Response to question: Has your school or district established proficiency targets for each of the levels of world language study?

a. This is asked in the Outcome section for elem, middle, and high
b. Please report on the response by level

3. Internal Consistency of Teacher Practices: start from 1610

a. Check the internal consistency of scores in the Teacher Practices section
b. Cronbach Alpha coefficient

ii. See spreadsheet for sub-scales
iii. Please note certain items need to be reversed when checking for internal consistency (these are noted as "negative" in the spreadsheet)

```
#read file

data_fpath <- file.path(
  "data",
  "World Language Programming Survey in MA - Data to MSSP 6-10-2020.xlsx"
)

first_two_rows <- read_excel(data_fpath, n_max = 1)

qnums <- first_two_rows %>%
  names() %>%
  janitor::make_clean_names()

qnames <- first_two_rows %>%
  flatten_chr()

data <- read_excel(data_fpath, skip = 2, col_names = qnames) %>%
  arrange(`End Date`) %>%
  mutate(response_num = row_number()) %>%
```

```
  rename(district =
           starts_with("Please select your school district"),
         languages_taught =
           matches("Which languages do you currently teach.+ choice$"))
```

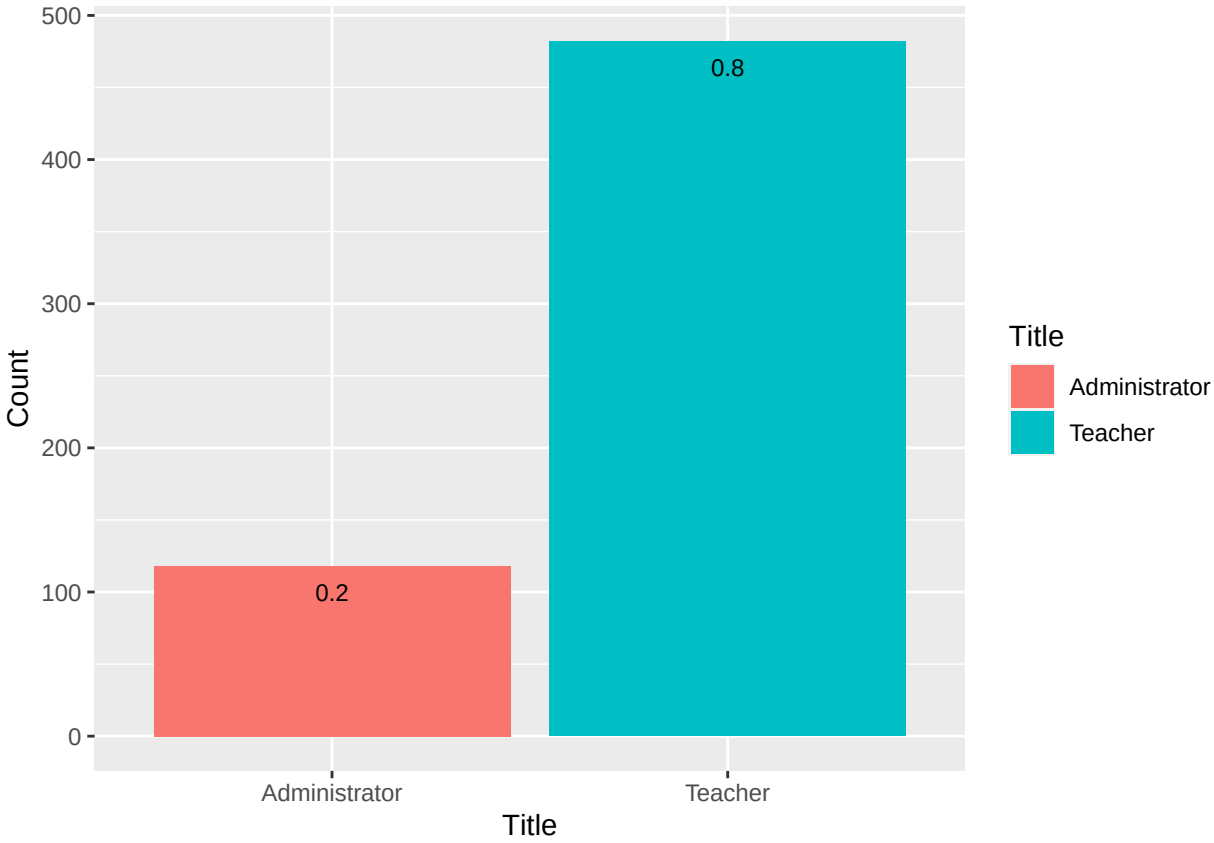
```
## New names:
## * `Name of Elementary School (or Schools if the language program is the same across schools):` -> `Na
## * `For each of the languages taught, indicate the program type from the descriptions above that best
## * `For each of the languages taught, indicate the program type from the descriptions above that best
## * `For each of the languages taught, indicate the program type from the descriptions above that best
## * `For each of the languages taught, indicate the program type from the descriptions above that best
## * ...
```

```
data_clean_names <- data %>%
  janitor::clean_names()

#keep columns that I will use
data <- data_clean_names[,c(9:11, 13)]
```

```
#add clean_title(administrator/teacher) column

# some titles are hard to determine: "specialist", "liaison" "supervisor"
# use answers in language_taught as the second filter
administrator_title <- c("head", "chair", "leader", "chairperson", "director", "supervisor", "administra
data$clean_title <- "Teacher"


# create data frame with teach/administrator title and the corresponding count
for(i in 1:nrow(data)){
  detect <- sum(stri_detect_fixed(tolower(data$part_i_demographics_your_title)[i], administrator_title)
  if(detect >= 1){
    data$clean_title[i] <- "Administrator"
  } else if(grepl("chair/administrator/supervisor", data$languages_taught[i])){
    data$clean_title[i] <- "Administrator"
  }
}

Title.df <- data.frame(table(data$clean_title))
colnames(Title.df) <- c("Title","Count")
```

#1.a

```
#visualize title
ggplot(Title.df, aes(x = Title, y = Count, fill = Title)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = round(Count/sum(Count),digits = 2)), vjust = 2, color= "black", size=3)
```

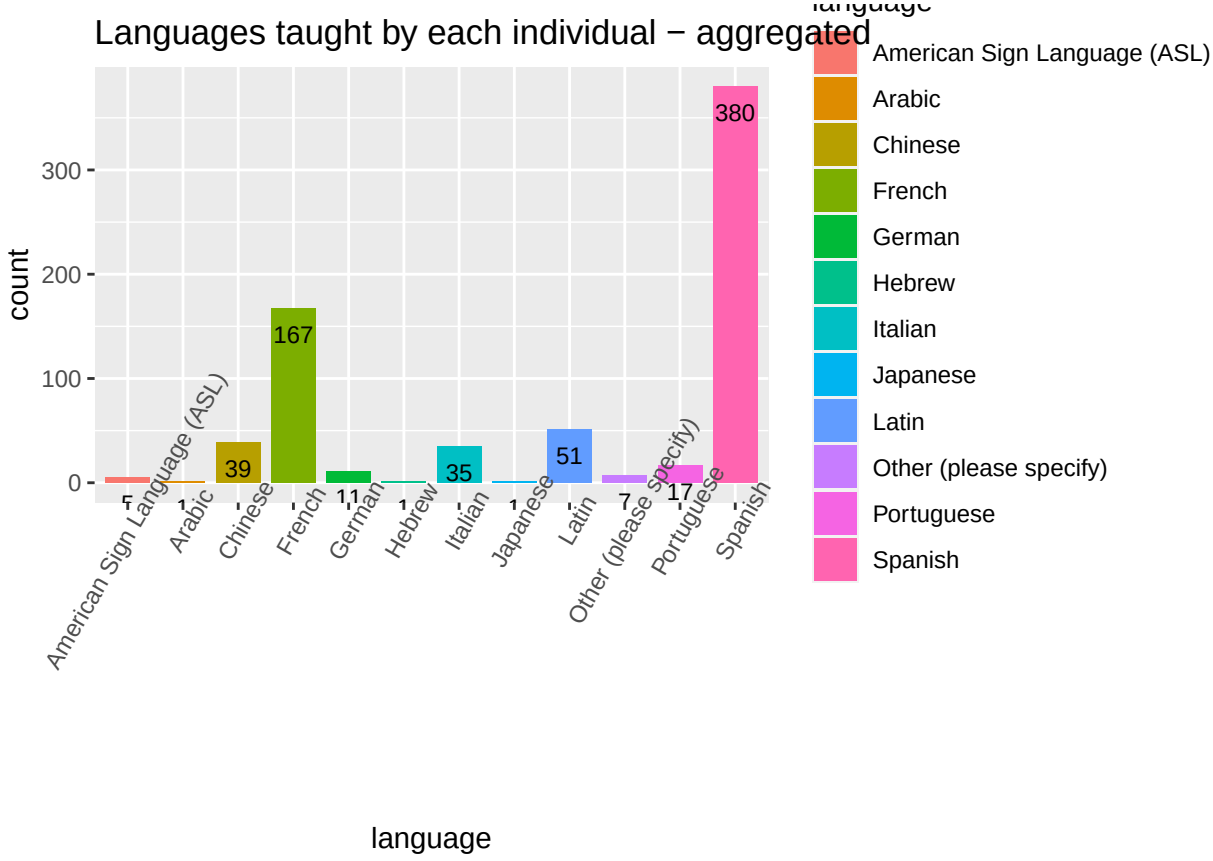



```
# clean columns 'languages_taught' and 'what_level_s_do_you_currently_teach_please_check_all_that_apply
# remove the answer "I am a World Language Program chair/administrator/supervisor"

stat <- "I am a World Language Program chair/administrator/supervisor"

data2 <- data # for language question
data3 <- data # for level question

for(i in 1:nrow(data2)){
  detect2 <- sum(stri_detect_fixed(data2$languages_taught[i], stat))
  if(detect2 >= 1) {
    x <- unlist(strsplit(data2$languages_taught[i], " "))
    data2$languages_taught[i] <- ifelse((c(x[1],x[2]) == c("I", "am")),
                                        str_remove(data2$languages_taught[i], "I am a World Language Progr
                                        str_remove(data2$languages_taught[i], ",I am a World Language Prog
    }
}

for(i in 1:nrow(data3)){
  detect3 <- sum(stri_detect_fixed(data3$what_level_s_do_you_currently_teach_please_check_all_that_apply
  if(detect3 >= 1) {
    y <- unlist(strsplit(data3$what_level_s_do_you_currently_teach_please_check_all_that_apply[i], " "))

    if (c(y[1],y[2]) == c("I", "am")){
      data3$what_level_s_do_you_currently_teach_please_check_all_that_apply[i] <-
        str_remove(data3$what_level_s_do_you_currently_teach_please_check_all_that_apply[i],
```

```
                    "I am a World Language Program chair/administrator/supervisor")
    } else{
      data3$what_level_s_do_you_currently_teach_please_check_all_that_apply[i] <-
        str_remove(data3$what_level_s_do_you_currently_teach_please_check_all_that_apply[i],
                    ",I am a World Language Program chair/administrator/supervisor")
    }
  }
}
```

#1.b

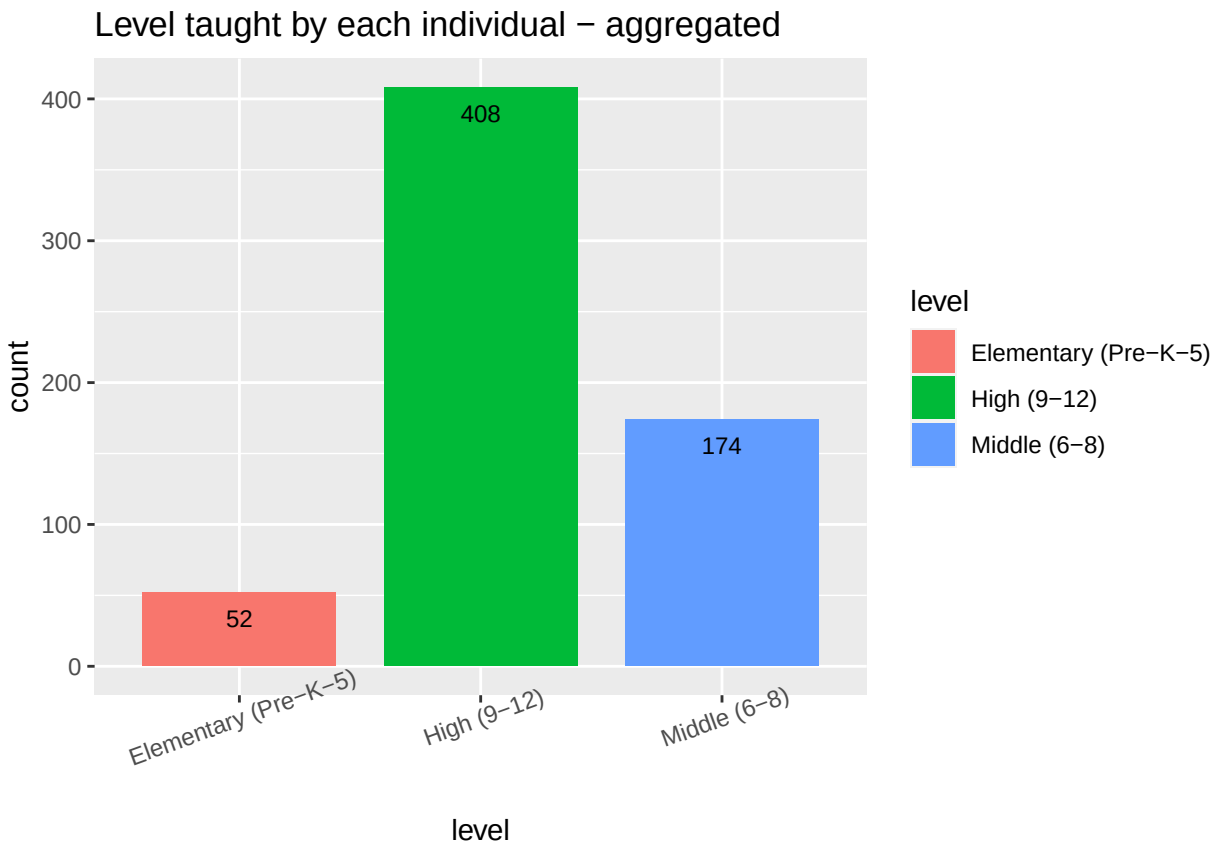
```
# create data frame with each kind of language and the corresponding count
# a person will be counted as 1 in all language levels that this person taught

language.df <- data.frame(language = character(), count = numeric())
data2$languages_taught[data2$languages_taught ==""] <- NA
data2 <- na.omit(data2)

for(i in 1:nrow(data2)){
  split <- unlist(strsplit(data2$languages_taught[i], ","))
  ind <- length(split)
   for (j in 1:ind){
     lan <- split[j]
     if(!split[j] %in% language.df$language){
       # print(i)
       # print(split[j])
       language.df <- rbind.data.frame(language.df, cbind.data.frame(language = lan, count = 1))
     } else{
       ind2 <- which(language.df$language == lan)
       language.df[ind2,]$count <- language.df[ind2,]$count + 1
     }
   }
}

# visualize language_taught distribution
ggplot(language.df, aes(x = language, y = count, fill = language)) +
  geom_bar(stat = "identity", width = 0.8, position = position_dodge(width = 0.9)) +
  geom_text(aes(label = count), vjust = 2, color= "black", size=3) +
  theme(axis.text.x = element_text(angle = 60)) +
  labs(title = "Languages taught by each individual - aggregated")
```



#1.c

```
# create data frame with teach levels (ele,mid,high) and the corresponding counts
# a person will be counted as 1 in all levels that this person taught

level.df <- data.frame(level = character(), count = numeric())
data3$what_level_s_do_you_currently_teach_please_check_all_that_apply[data3$what_level_s_do_you_current
data3 <- na.omit(data3)


for(i in 1:nrow(data3)){
  split <- unlist(strsplit(data3$what_level_s_do_you_currently_teach_please_check_all_that_apply[i], ",
  ind <- length(split)
   for (j in 1:ind){
     lev <- split[j]
     if(!lev %in% level.df$level){
       # print(i)
       # print(split[j])
       level.df <- rbind.data.frame(level.df, cbind.data.frame(level = lev, count = 1))
     } else{
       ind2 <- which(level.df$level == lev)
       level.df[ind2,]$count <- level.df[ind2,]$count + 1
     }
   }
}
```

```r
# only keep level = ele, mid, and high
ele_mid_high <- level.df[c(1,2,4),]

# visualize teach level distrubution
ggplot(ele_mid_high, aes(x = level, y = count, fill = level)) +
  geom_bar(stat = "identity", width = 0.8, position = position_dodge(width = 0.9)) +
  geom_text(aes(label = count), vjust = 2, color= "black", size=3) +
  theme(axis.text.x = element_text(angle = 20)) +
  labs(title = "Level taught by each individual - aggregated")
```



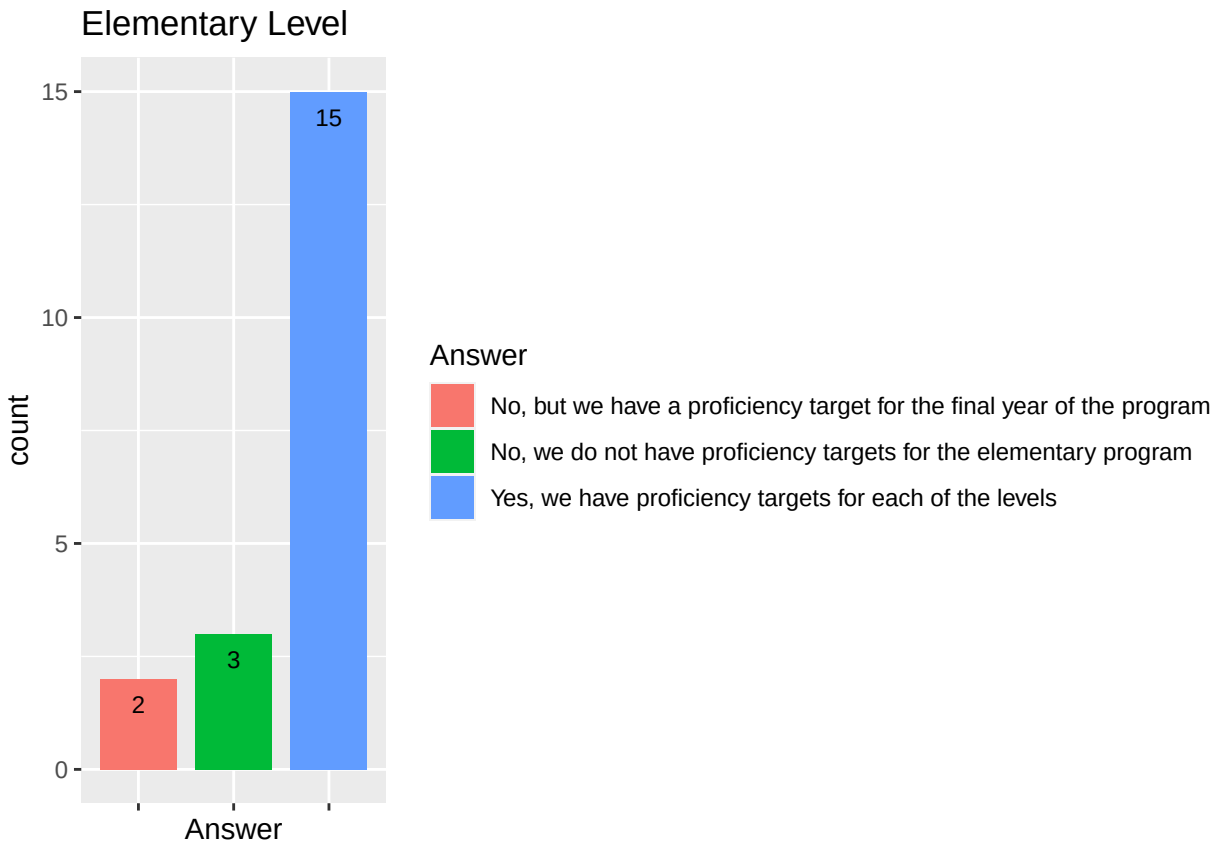
#2

```r
# For each district:
# * If one or more administrators, include the administrator with greater progress
# * If more than one administrators have the same progress, take the first administrator
# * If no admin, take the first teacher to respond to the survey

one_per_district <- data_clean_names %>%
  select(response_num, district, progress, duration_in_seconds, finished,
         languages_taught,has_your_school_or_district_established_proficiency_targets_for_each_of_the_l
         has_your_school_or_district_established_proficiency_targets_for_each_of_the_levels_of_world_lan
         has_your_school_or_district_established_proficiency_targets_for_each_of_the_levels_of_world_lan
  mutate(is_admin = str_detect(languages_taught,
                               "I am a World Language Program chair")) %>%
  group_by(district) %>%
  filter(
```

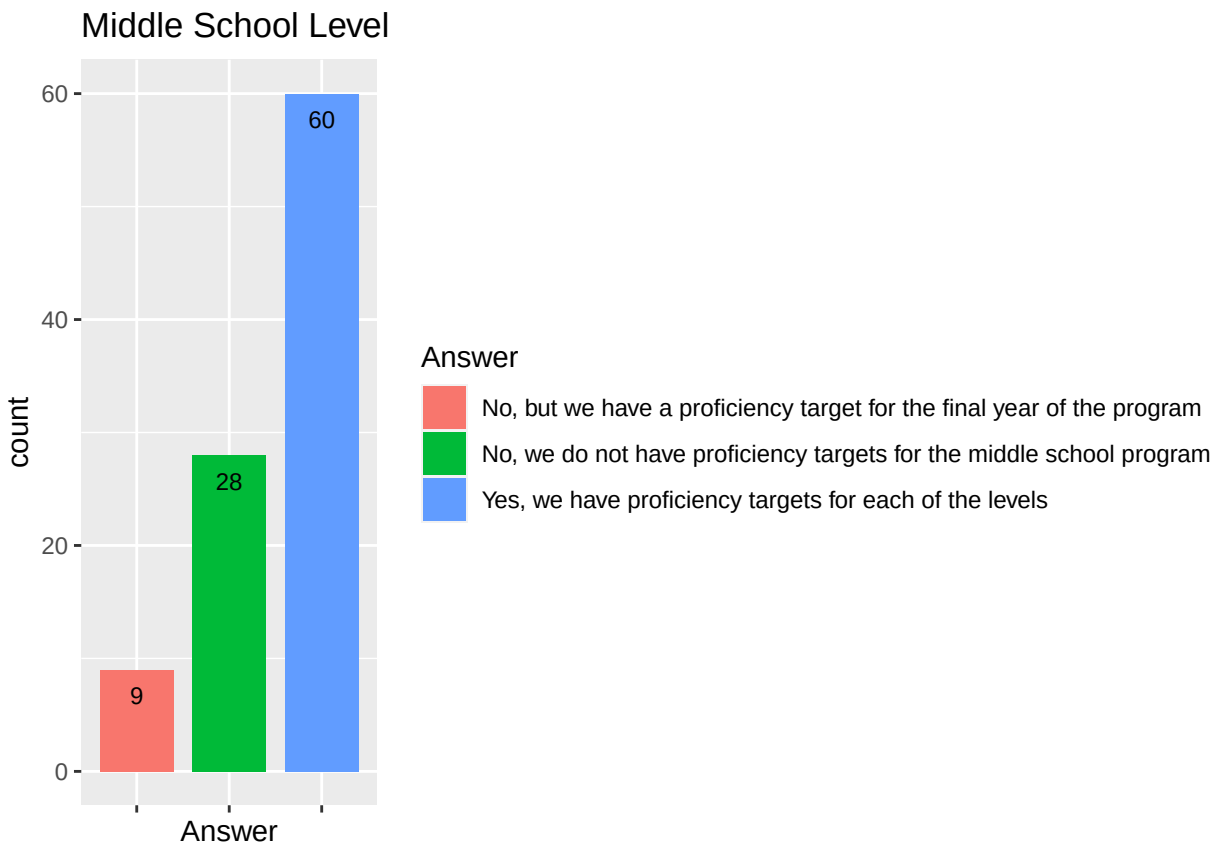
```
    case_when(
      any(is_admin)  ~ is_admin,
      !any(is_admin) ~ response_num == min(response_num)
    )
  ) %>%
  ungroup()

one_per_district <- one_per_district %>% group_by(district) %>%
  filter(progress == max(progress)) %>% filter(response_num == min(response_num))
```

#2.b Response to question: Has your school or district established proficiency targets for each of the levels of world language study?

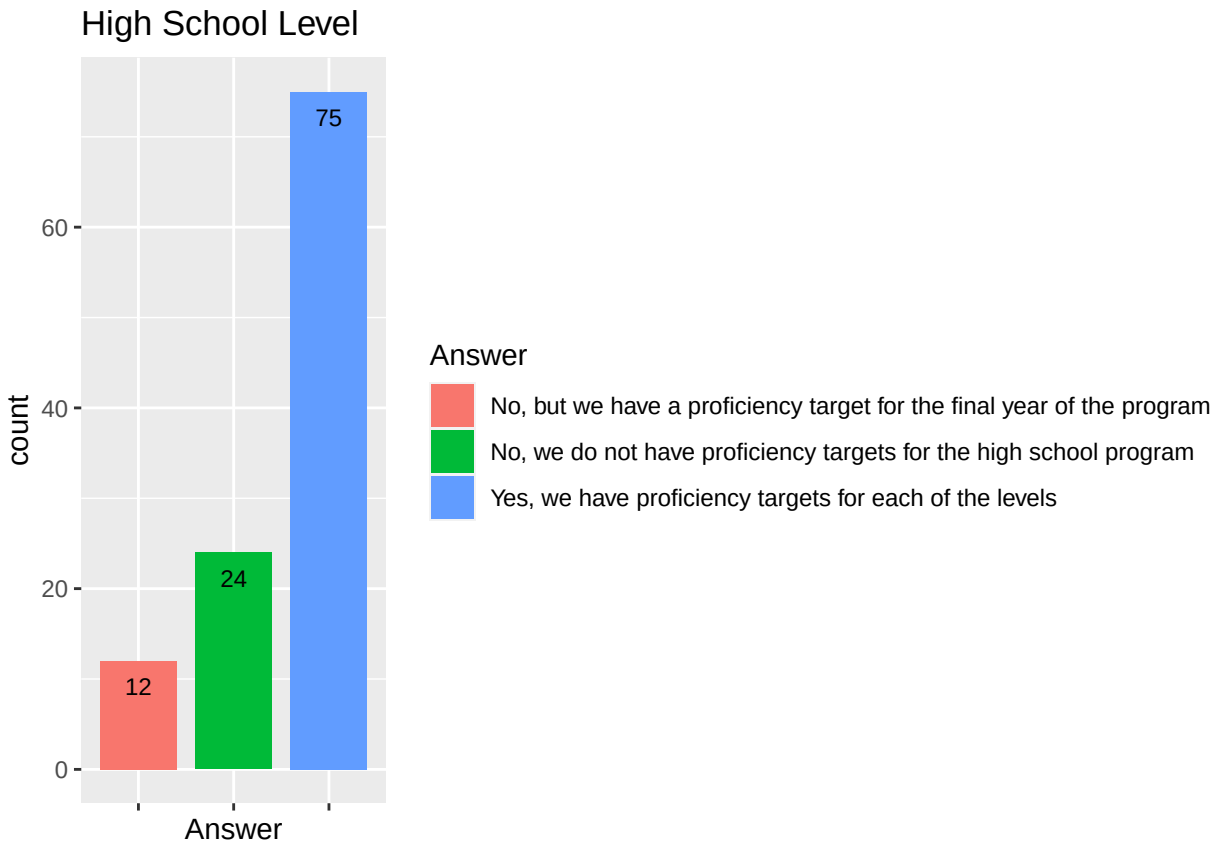
```
# distribution of responses by elementary school, middle school, and high school level
# excluded the answer "Not sure"
ele_targets <- data.frame(table(one_per_district$has_your_school_or_district_established_proficiency_ta
ele_targets <- ele_targets %>% filter(Var1 != "Not sure")
colnames(ele_targets) <- c("Answer","count")

mid_targets <- data.frame(table(one_per_district$has_your_school_or_district_established_proficiency_ta
mid_targets <- mid_targets %>% filter(Var1 != "Not sure")
colnames(mid_targets) <- c("Answer","count")

high_targets <- data.frame(table(one_per_district$has_your_school_or_district_established_proficiency_t
high_targets <- high_targets %>% filter(Var1 != "Not sure")
colnames(high_targets) <- c("Answer","count")


ele <- ggplot(ele_targets, aes(x = Answer, y = count, fill = Answer)) +
  geom_bar(stat = "identity", width = 0.8, position = position_dodge(width = 0.9)) +
  geom_text(aes(label = count), vjust = 2, color= "black", size=3) +
  theme(axis.text.x=element_blank()) +
  labs(title = "Elementary Level")

mid <- ggplot(mid_targets, aes(x = Answer, y = count, fill = Answer)) +
  geom_bar(stat = "identity", width = 0.8, position = position_dodge(width = 0.9)) +
  geom_text(aes(label = count), vjust = 2, color= "black", size=3) +
  theme(axis.text.x=element_blank()) +
  labs(title = "Middle School Level")

high <- ggplot(high_targets, aes(x = Answer, y = count, fill = Answer)) +
  geom_bar(stat = "identity", width = 0.8, position = position_dodge(width = 0.9)) +
  geom_text(aes(label = count), vjust = 2, color= "black", size=3) +
  theme(axis.text.x=element_blank()) +
  labs(title = "High School Level")

ele
```


Elementary Level
15
10
5
0
count
15
3
2
Answer
Answer
No, but we have a proficiency target for the final year of the program
No, we do not have proficiency targets for the elementary program
Yes, we have proficiency targets for each of the levels

```
mid
```


Middle School Level
60
40
20
0
count
60
28
9
Answer
Answer
No, but we have a proficiency target for the final year of the program
No, we do not have proficiency targets for the middle school program
Yes, we have proficiency targets for each of the levels

```
high
```


High School Level
count
60
40
20
0
75
24
12
Answer
Answer
No, but we have a proficiency target for the final year of the program
No, we do not have proficiency targets for the high school program
Yes, we have proficiency targets for each of the levels