

NISS Data Challenge - Education, Employment, and Earnings

Nathan Nguyen *Duke University*

The location of the school where the ninth-graders attended might contribute to their predicted/aspired future occupation. Different locations such as the town, city, rural might sociologically affect one's decision to get a job. To investigate other perspectives in the original data given by the NISS competition organizer, I use the variables S4JobIndustry and X4LOCALE to estimate the percentage of ninth-graders in different high school locations, grouped by 2016 expected occupation.

```
#load data
data <- read.csv("data/HSLSEESAID.csv")
```

Question: How did 2009 ninth-graders attending high school in different regions predict their occupation in 2016?

```
#load packages and themes
library(showtext)
library(tidyverse)
library(viridis)
font_add_google("Lora", "Lora")
showtext_auto()
```

```
order <- c("Don't Know", "Other", "Trades and \n technical", "STEM", "Service",
           "Military and \n protective services", "Healthcare", "Education",
           "Business and \n Management", "Arts and \n entertainment")
```

```
#Processing S4JobIndustry's values so that it saves space in the graph
data <- data %>%
  mutate(S4JobIndustry = case_when(
    S4JobIndustry == "Military and protective services" ~
      "Military and \n protective services",
    S4JobIndustry == "Business and Management" ~
      "Business and \n Management",
    S4JobIndustry == "Arts and entertainment" ~
      "Arts and \n entertainment",
    S4JobIndustry == "Trades and technical" ~
      "Trades and \n technical",
    TRUE ~ S4JobIndustry
  ))
```

```

#process data
processed_data <- data %>%
  mutate(S4JobIndustry = factor(S4JobIndustry, levels = order)) %>%
  group_by(S4JobIndustry, X4LOCALE) %>%
  summarise(count_location = n()) %>%
  group_by(S4JobIndustry) %>%
  mutate(count = sum(count_location)) %>%
  mutate(percent = count_location/count*100) %>%
  ungroup()

#create a new variable for positions of percentages on the graphs
processed_data <- processed_data %>%
  group_by(S4JobIndustry) %>%
  mutate(pos = 100 - cumsum(percent) + (0.5*percent))

#visualizing with a stacked bar chart
g <- ggplot(processed_data, aes(x = factor(S4JobIndustry, levels = order),
                                y = percent,
                                fill = X4LOCALE))+
  geom_bar(stat = "identity", width = .5)+
  coord_flip()+ #flip the chart
  scale_fill_manual(values = c("red", "black", "blue", "grey50"))+ #set color
  theme_linedraw()+
  labs(y = "Percentage",
       x = "Expected Job",
       fill = "Location")+
  theme( #set theme
    text = element_text(family = "Lora", size = 12),
    legend.position = "bottom",
    panel.grid.major.y = element_blank(),
    panel.grid.minor.x = element_blank()
  )+
  scale_y_continuous(labels = function(x) paste0(x, "%"))+
  geom_text(processed_data, mapping = aes(x = S4JobIndustry,
                                          y = pos,
                                          label = paste0(round(percent),"%")),
            size = 4,
            color = "white")
g

```

Based on the visualization, I obtained these key findings:

- Across all the expected occupations, ninth-graders equally attended their last high school in city, rural and suburb area, but the percentage of ninth-graders attended high school in town is remarkably lower (approximately 2 or 2.5 times smaller than the percentage of students in city, rural, and suburb area).

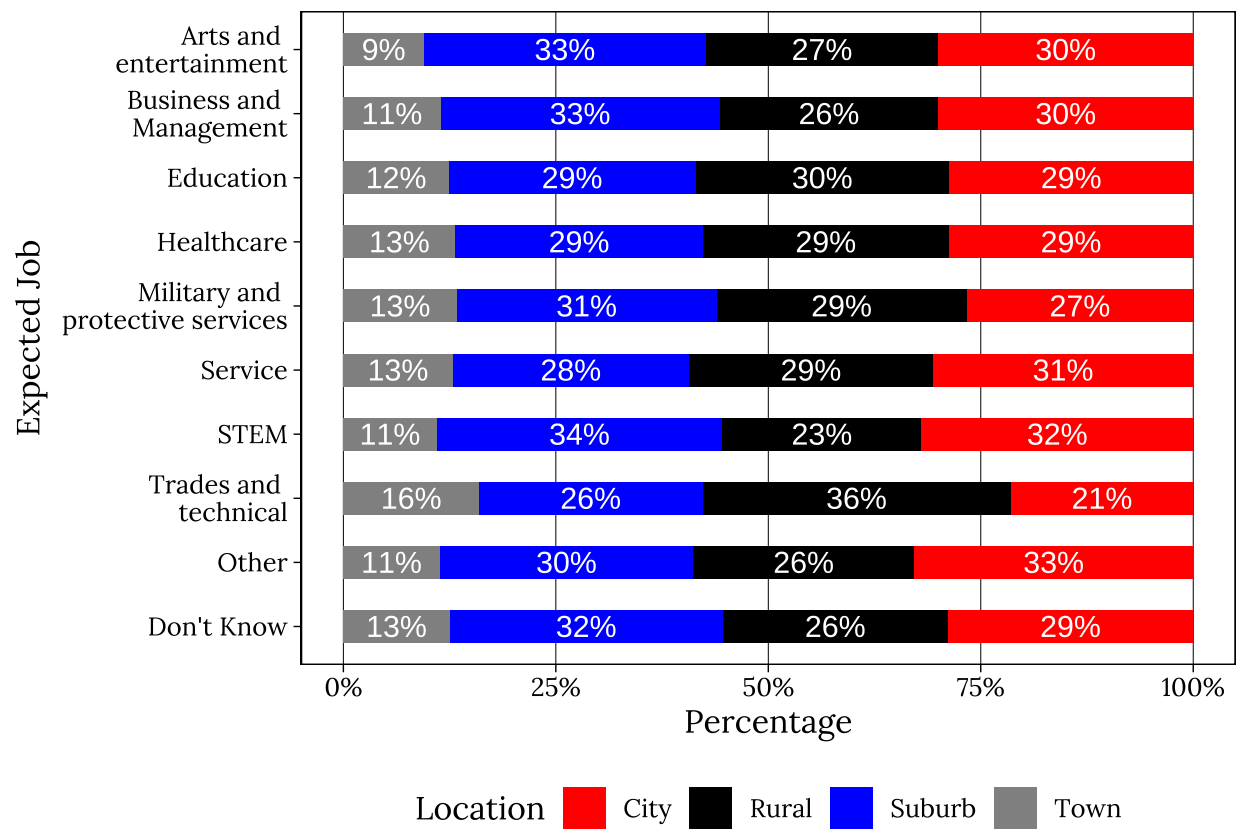


Figure 1: Percentage distribution of 2009 ninth-graders, by geographical region of last attended high school, grouped by expected occupations in 2016

- Ninth-graders expected to have an occupation in STEM field tend to be on the city and suburb area, with respective percentage of 32% and 34% each, whereas the percentage of ninth-graders in rural area and town is remarkably lower compared to the other two (with 23% and 11% each).
- We can see that generally, in each expected occupation, the percentage of ninth-graders attended rural high school is slightly lower than the percentage of ninth-graders attended their last high school in city and suburb. However, we can see that this is not true with the expected jobs in trades and technical, education, and service.
- In service, the percentage of ninth-graders attended rural high school is slightly lower than the percentage of ninth-graders attended their last high school in the suburb (29% compared to 28%), while considering expected job in education, the percentage of ninth-graders attended rural high school is slightly higher than both the percentage of ninth-graders attended their last high school in city and suburb (30% compared to 29% of both city and suburb's percentage).
- The most remarkable difference is in the category trades and technical, where the percentage of ninth-graders attended rural high school is clearly larger than the percentage of ninth-graders attended their last high school in city and suburb, which is 36%, remarkably high compared to other category. We can also see that the percentage of ninth-graders in town is higher when considering the occupation in trades and technical, which is 16%. It is clear that the percentages of ninth-graders expected to have a job in trades and technical whose last attended high school is in town and rural area, are remarkably higher.