NISS Data Challenge - Education, Employment, and Earnings

Nathan Nguyen   *Duke University*

This document provides an introduction to R Markdown, argues for its benefits, and presents a sample manuscript template intended for an academic audience. I include basic syntax to R Markdown and a minimal working example of how the analysis itself can be conducted within R with the knitr package.

Keywords: pandoc, r markdown, knitr

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

Introduction

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

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

```r
#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))
```

```r
ggplot(processed_data, aes(x = factor(S4JobIndustry, levels = order),
                           y = percent,
                           fill = X4LOCALE))+
  geom_bar(stat = "identity", width = .5)+
  coord_flip()+
  scale_fill_manual(values = c("red", "black", "blue", "grey50"))+
  theme_linedraw()+
  labs(y = "Percentage",
       x = "Expected Job",
       fill = "Location")+
  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")
```