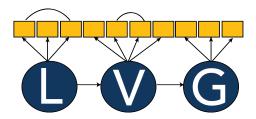# Data Scraping Child Document

## Overview of the Data Scraping Process

Delwin Carter

2025-01-26



**Latent Variable Group**
Department of Education
University of California, Santa Barbara

```r
library(parallel)

# Use the global variable `csv_directory`
csv_files <- list.files(path = csv_directory, pattern = "*.csv", full.names = TRUE)
# Read and process each CSV file
read_csv_file <- function(file) {
  data <- read.csv(file, sep = " ", header = FALSE)
  data$FileName <- gsub("\\.[^.]*$", "", basename(file))
  return(data)
}

# Combine files using parallel processing
num_cores <- detectCores() - 1
cluster_type <- ifelse(.Platform$OS.type == "windows", "PSOCK", "FORK")
cl <- makeCluster(num_cores, type = cluster_type)
clusterExport(cl, c("read_csv_file", "gsub", "basename"))
combined_data <- do.call(rbind, parLapply(cl, csv_files, read_csv_file))
stopCluster(cl)
```