Integrated Postsecondary Education Data System (IPEDS) Data Import

Nathan Alexander, PhD Amleset Abrhale Howard University

Contents


```
knitr::opts_chunk$set(echo = TRUE)

library(tidyverse)
library(ggplot2)
library(dplyr)
library(stringr)
library(weights)
library(here)
here::i_am("ipeds-import.Rmd")
```

We import and prepare for the analysis of the Integrated Postsecondary Education Data System (IPEDS) surveys. We first review the most recent data sets and dictionaries available (2023) here. We also explore some potential options for future analysis.

DATA

We import a series of raw data files.

- **Institutional Characteristics**: Educational offerings, organization, services and athletic associations
- **Graduation Rates**: Graduation rate data, 150% of normal time to complete - cohort year 2017 (4-year) and cohort year 2020 (2-year) institutions
- **Student Financial Aid and Net Price**: Student financial aid and net price: 2022-23

```r
# institutional characteristics
ic <- read.csv("../data/ipeds-2023-ic.csv")

# institutional characteristics
gradrate <- read.csv("../data/ipeds-2023-grad-rate-4year-cohort17-2year-cohort20.csv")

# institutional characteristics
aid <- read.csv("../data/ipeds-2023-stu-fin-aid-net-price.csv")
```

Then we preview the data.

Given the large size of the data frames, we select the first few columns and limiting to the first two rows.

Institutional characteristics

```r
ic %>%
  select(UNITID, PEO1ISTR, PEO2ISTR, PEO3ISTR, PEO4ISTR) %>%
  head(n=2)
```

```
##   UNITID PEO1ISTR PEO2ISTR PEO3ISTR PEO4ISTR
## 1 100654        0        1        0        0
## 2 100663        0        1        1        0
```

Graduation rate

```r
gradrate %>%
  select(UNITID, GRTYPE, CHRTSTAT, SECTION, COHORT) %>%
  head(n=2)
```

```
##   UNITID GRTYPE CHRTSTAT SECTION COHORT
## 1 100654      1       10       1      1
## 2 100654      2       12       1      1
```

Financial aid

```r
aid %>%
  select(UNITID, SCUGRAD, SCUGDGSK, SCUGNDGS, SCUGFFN) %>%
  head(n=2)
```

```
##   UNITID SCUGRAD SCUGDGSK SCUGNDGS SCUGFFN
## 1 100654    5206     5201        5    1547
## 2 100663   13032    12776      256    2172
```

VARIABLES

Amleset will work on the variable labels.

INSTITUTIONS

```
hbcu <- c(
  "220182", # Tennessee State University
  "175772", # Jackson State University
  "234155", # Virginia State University
  "176044", # Mississippi Valley State University
  "198543", # Fayetteville State University
  "198507", # Elizabeth City State University
  "199102", # North Carolina A&T State University
  "131520"  # Howard University
)

hd202.hbcu <- hd2021[which(hd2021$unitid %in%hbcu),]
p <- hd2021.hbcu[, c("instnm", "webaddr", "stabbr", "control")]
names(p) <- c("Institution", "Web Address", "State", "Sector")
p

ivy <- c(
  "186131", # Yale University
  "190150", # Columbia University
  "166027", # Cornell University
  "130794", # Dartmouth College
  "215062", # University of Pennsylvania
  "182670", # Princeton University
  "217156", # Brown University
  "190415"  # Harvard University
)

hd2021.ivy <- hd2021[which(hd2021$unitid %in%ivy),]
q <- hd2021.ivy[, c("instnm", "webaddr", "stabbr", "control")]
names(p) <- c("Institution", "Web Address", "State", "Sector")
q
```