

# cdc\_healthcare\_employment

John Cruz

2023-02-28

## R Markdown

[CDC National Center for Health Statistics](#)

[HCEmpl](#)

each profession and see if employment has been increasing or decreasing and compare this with the change in mean salary. Also, I was thinking about creating a rating or score to see which profession has grown most throughout the last 20 years.

## Required Libraries

```
library(tidyverse)
library(rio)
library(janitor)
```

## Import Data

```
url <- 'https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Publications/Health_US/hus20-21tables/hcempl.xls'
data = import(url)      # use rio to read URL excel file
```

## Drop Non-Data Rows

```
updated_data <-
  data |>
  filter(!row_number() %in% c(1:3, 49:51)) |>      # drop non-data rows
  row_to_names(row_number = 1) |>                  # use janitor to take first row and make them column names
  clean_names()                                     # clean column names
```

## Subset Employment Figures

```
updated_employment <-
  updated_data |>
  select(c(1:8)) |>
```

```
gather('year', 'employment', -c('occupation_title'))

updated_employment$year <-
  updated_employment$year |>
  parse_number()

knitr::kable(head(updated_employment))
```

occupation_title	year	employment
Health care practitioners and technical occupations	2000	NA
Audiologists	2000	11530
Cardiovascular technologists and technicians	2000	40080
Clinical laboratory technologists and technicians	2000	...
Dental hygienists	2000	148460
Diagnostic medical sonographers	2000	31760

## Subset Mean Hourly Wage Figures

```
updated_wages <-
  updated_data |>
  select(c(1, 9:15)) |>
  gather('year', 'mean_hourly_wage', -c('occupation_title'))

updated_wages$year <-
  updated_wages$year |>
  parse_number()

knitr::kable(head(updated_wages))
```

occupation_title	year	mean_hourly_wage
Health care practitioners and technical occupations	2000	NA
Audiologists	2000	22.92
Cardiovascular technologists and technicians	2000	16.809999999999999
Clinical laboratory technologists and technicians	2000	...
Dental hygienists	2000	24.99
Diagnostic medical sonographers	2000	22.03

## Join Employment and Wage Figures

```
employment_wages <-
  inner_join(updated_employment, updated_wages)

## Joining with 'by = join_by(occupation_title, year)'
```

```
knitr::kable(head(employment_wages))
```

occupation_title	year	employment	mean_hourly_wage
Health care practitioners and technical occupations	2000	NA	NA
Audiologists	2000	11530	22.92
Cardiovascular technologists and technicians	2000	40080	16.809999999999999
Clinical laboratory technologists and technicians	2000	...	...
Dental hygienists	2000	148460	24.99
Diagnostic medical sonographers	2000	31760	22.03

## Clean Occupation Names and Figures

```
employment_wages <-  
  employment_wages |>  
  mutate(occupation_title = str_replace(occupation_title, "\\[:digit:]", "")) |> # clean occupation  
  mutate(employment = str_replace(employment, "[^[:alnum:]]+", "")) |> # clean employment  
  mutate(employment = as.integer(employment)) |>  
  mutate(mean_hourly_wage = str_replace(mean_hourly_wage, "[^[:alnum:]]\\.\\.", "")) |> # clean wages  
  mutate(mean_hourly_wage = as.numeric(mean_hourly_wage))
```

## Header