TidyTuesday Challenge

Lisa Zuo

December 9, 2022

Load Data

```
library(tidyverse)
library(tidymodels)
library(tidytuesdayR)
tuesdata <- tt_load('2021-10-05')
```

```
  Downloading file 1 of 1: `nurses.csv`
```

```
nurses <- tuesdata$nurses
```

Data cleaning

```
nurses <- nurses %>%
  rename(state = State,
         year = Year,
         total_employed = `Total Employed RN`,
         employed_se = `Employed Standard Error (%)`,
         avg_hourly = `Hourly Wage Avg`,
         median_hourly = `Hourly Wage Median`,
         avg_annual = `Annual Salary Avg`,
         median_annual = `Annual Salary Median`,
         wage_se = `Wage/Salary standard error (%)`,
         perc_hourly_10 = `Hourly 10th Percentile`,
         perc_hourly_25 = `Hourly 25th Percentile`,
         perc_hourly_75 = `Hourly 75th Percentile`,
```

```
          perc_hourly_90 = `Hourly 90th Percentile`,
          perc_annual_10 = `Annual 10th Percentile`,
          perc_annual_25 = `Annual 25th Percentile`,
          perc_annual_75 = `Annual 75th Percentile`,
          perc_annual_90 = `Annual 90th Percentile`,
          location_quotient = `Location Quotient`,
          total_employed_national_aggregate =
            `Total Employed (National)_Aggregate`,
          total_employed_healthcare_national_aggregate =
            `Total Employed (Healthcare, National)_Aggregate`,
          total_employed_healthcare_state_aggregate =
            `Total Employed (Healthcare, State)_Aggregate`,
          yearly_total_employed_state_aggregate =
            `Yearly Total Employed (State)_Aggregate`)
```

EDA

```
str(nurses)
```

```
spec_tbl_df [1,242 x 22] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ state                                          : chr [1:1242] "Alabama" "Alaska" "Arizona" "
 $ year                                           : num [1:1242] 2020 2020 2020 2020 2020 2020 2
 $ total_employed                                 : num [1:1242] 48850 6240 55520 25300 307060
 $ employed_se                                    : num [1:1242] 2.9 13 3.7 4.2 2 2.8 6.5 11.4 1
 $ avg_hourly                                     : num [1:1242] 29 45.8 38.6 30.6 58 ...
 $ median_hourly                                  : num [1:1242] 28.2 45.2 38 30 56.9 ...
 $ avg_annual                                     : num [1:1242] 60230 95270 80380 63640 120560
 $ median_annual                                  : num [1:1242] 58630 94070 79010 62330 118410
 $ wage_se                                        : num [1:1242] 0.8 1.4 0.9 1.4 1 0.7 1 2.5 1.5
 $ perc_hourly_10                                 : num [1:1242] 20.8 31.5 27.7 21.5 36.6 ...
 $ perc_hourly_25                                 : num [1:1242] 23.7 36.9 32.6 25.7 45.2 ...
 $ perc_hourly_75                                 : num [1:1242] 33.1 53.3 44.7 35.4 71.1 ...
 $ perc_hourly_90                                 : num [1:1242] 38.7 60.7 50.1 39.6 83.3 ...
 $ perc_annual_10                                 : num [1:1242] 43150 65530 57530 44660 76180
 $ perc_annual_25                                 : num [1:1242] 49360 76830 67760 53490 93970
 $ perc_annual_75                                 : num [1:1242] 68960 110890 92920 73630 14783
 $ perc_annual_90                                 : num [1:1242] 80420 126260 104290 82480 1733
 $ location_quotient                              : num [1:1242] 1.2 0.98 0.91 1 0.87 0.95 1.01
 $ total_employed_national_aggregate              : num [1:1242] 1.4e+08 1.4e+08 1.4e+08 1.4e+08
 $ total_employed_healthcare_national_aggregate: num [1:1242] 8632190 8632190 8632190 863219
```

```
$ total_employed_healthcare_state_aggregate   : num [1:1242] 128600 17730 171010 80410 8447
$ yearly_total_employed_state_aggregate       : num [1:1242] 1903210 296300 2835110 1177860
- attr(*, "spec")=
 .. cols(
 ..   State = col_character(),
 ..   Year = col_double(),
 ..   `Total Employed RN` = col_double(),
 ..   `Employed Standard Error (%)` = col_double(),
 ..   `Hourly Wage Avg` = col_double(),
 ..   `Hourly Wage Median` = col_double(),
 ..   `Annual Salary Avg` = col_double(),
 ..   `Annual Salary Median` = col_double(),
 ..   `Wage/Salary standard error (%)` = col_double(),
 ..   `Hourly 10th Percentile` = col_double(),
 ..   `Hourly 25th Percentile` = col_double(),
 ..   `Hourly 75th Percentile` = col_double(),
 ..   `Hourly 90th Percentile` = col_double(),
 ..   `Annual 10th Percentile` = col_double(),
 ..   `Annual 25th Percentile` = col_double(),
 ..   `Annual 75th Percentile` = col_double(),
 ..   `Annual 90th Percentile` = col_double(),
 ..   `Location Quotient` = col_double(),
 ..   `Total Employed (National)_Aggregate` = col_double(),
 ..   `Total Employed (Healthcare, National)_Aggregate` = col_double(),
 ..   `Total Employed (Healthcare, State)_Aggregate` = col_double(),
 ..   `Yearly Total Employed (State)_Aggregate` = col_double()
 .. )
- attr(*, "problems")=<externalptr>
```