Tidy Tuesday: Registered Nurses

Kesava Asam

Report Last Run: 2021-10-08 17:34:55

Contents

1 Packages

```
# load required packages
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(tidytuesdayR)
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(scales)
```

```
##
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':
##
##     discard

## The following object is masked from 'package:readr':
##
##     col_factor
```

```
theme_set(theme_light())
```

2 Data

```
data_tue <- tt_load('2021-10-05')
```

```
## --- Compiling #TidyTuesday Information for 2021-10-05 ----
```

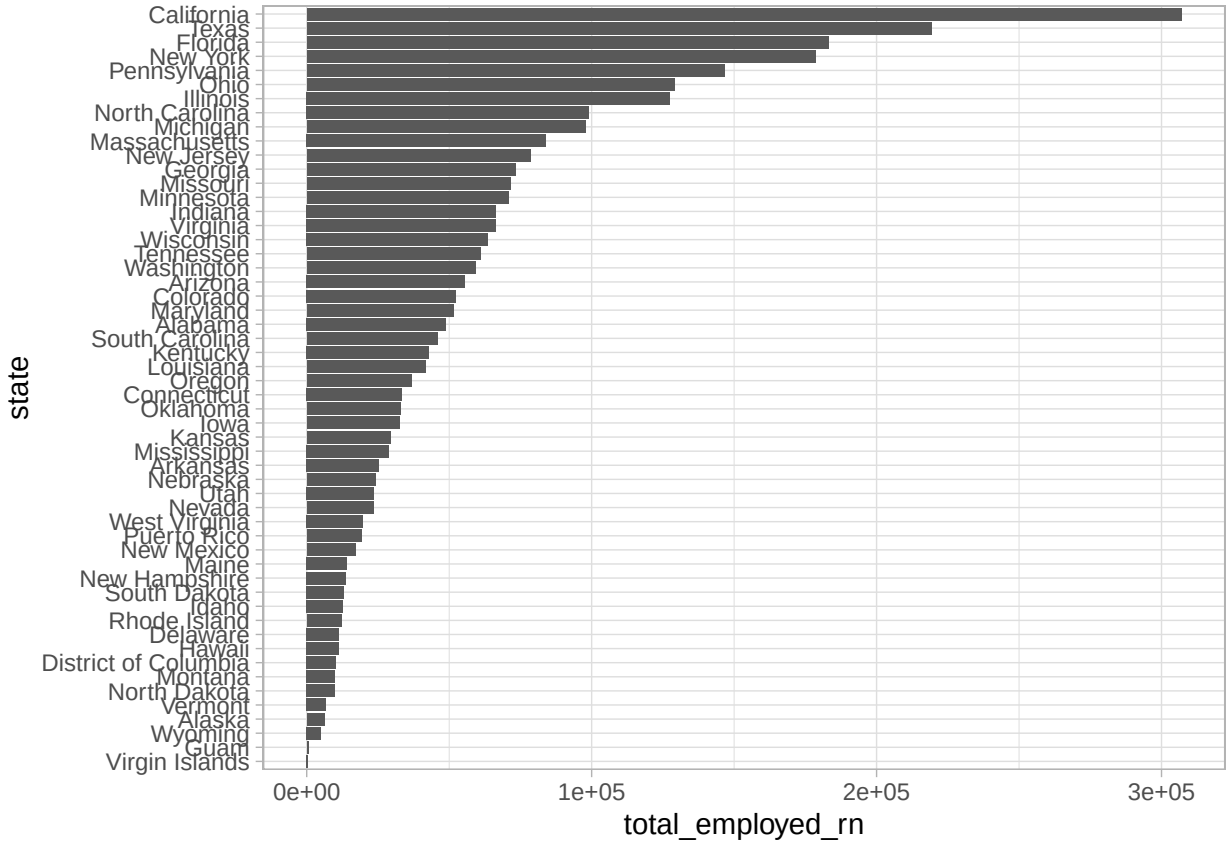
```
## --- There is 1 file available ---

## --- Starting Download ---

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

## --- Download complete ---
```

```
nurses_rn <-
  data_tue$nurses %>%
  clean_names()
```

Registered Nurses - From Data.World

Check the data

```
nurses_rn %>%
  filter(year == 2020) %>%
  mutate(state = fct_reorder(state, total_employed_rn)) %>%
  ggplot(aes(x = total_employed_rn, y = state)) +
  geom_col()
```



Can we do comparison of 2010 to 2020 wage comparison?

```
nurses_rn %>%
  filter(year %in% c(2000, 2010, 2020)) %>%
  select(state, year, total_employed_rn, hourly_wage_median) %>%
  pivot_wider(names_from = year, values_from = c(total_employed_rn, hourly_wage_median)) %>% arrange(des
```

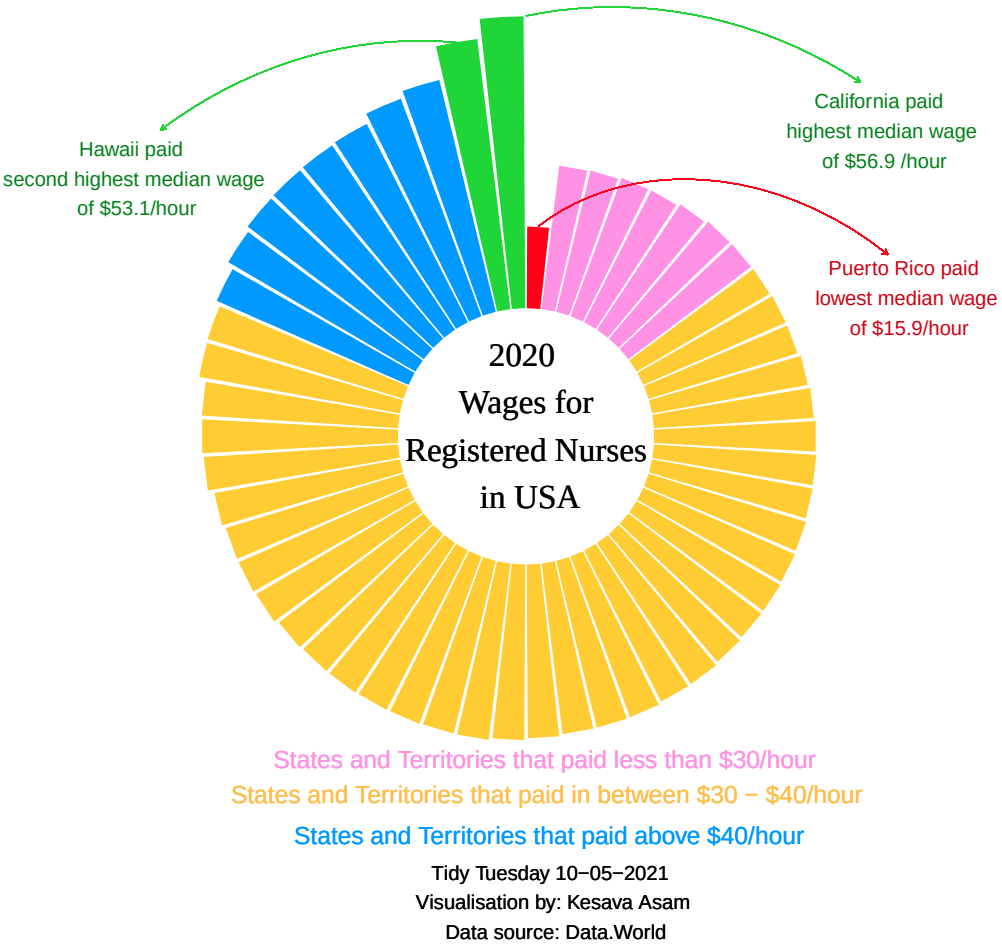
```
## # A tibble: 6 x 2
##   state         hourly_wage_median_2020
##   <chr>                           <dbl>
## 1 California                       56.9
## 2 Hawaii                           53.1
## 3 Oregon                           46.5
## 4 Alaska                           45.2
## 5 Massachusetts                    43.4
## 6 New York                         43.2
```

```
# Make the df needed
df_int <-
  nurses_rn %>%
  filter(year %in% c(2000, 2010, 2020)) %>%
  select(state, year, total_employed_rn, hourly_wage_median) %>%
  # get rns and media wages for 2000, 2010, and 2020 as wide table
  pivot_wider(names_from = year,
              values_from = c(total_employed_rn, hourly_wage_median)) %>%
  # calculate percent change b/w 2010 and 2020 wages
  mutate("decade_diff" =
           (hourly_wage_median_2020 - hourly_wage_median_2010)/hourly_wage_median_2010)
```

```r
# custom colors
col_man <- c("lowest" = "#FF0018",
             "low" = "#FF92E5",
             "highest" = "#1FD537",
             "high" = "#0099FF",
             "in_between" = "#FFCC33")

wages_2020 <-
  df_int %>%
  # bins for specific wages
  mutate(state = fct_reorder(state, hourly_wage_median_2020),
         color_col = factor(case_when(hourly_wage_median_2020 < 16 ~ "lowest",
                                      hourly_wage_median_2020 < 30  ~ "low",
                                      hourly_wage_median_2020 > 50 ~ "highest",
                                      hourly_wage_median_2020 > 40 ~ "high",
                                      TRUE ~ "in_between"))) %>%
  ggplot(aes(x=as.factor(state), y=hourly_wage_median_2020, fill = color_col)) +
  geom_bar(stat="identity") +
  scale_fill_manual(values = col_man, guide = "none") +
  ylim(-25,75) +
  theme_minimal() +
  coord_polar(start = 0) + #dark_mode() +
  theme(
    axis.text = element_blank(),
    axis.title = element_blank(),
    panel.grid = element_blank(),
    plot.margin = unit(rep(-3,4), "cm")) +
  geom_segment(aes(x = 0.5, y = 57, xend = 7, yend = 70),
               arrow = arrow(length = unit(0.15, "cm")),
               size = 0.3, color = "#1FD537") +
  annotate("text", x = 7.9,  y = 66.5,
           label = "California paid \n highest median wage \n of $56.9 /hour",
           colour = "#008018", size =5) +
  geom_segment(aes(x = 1, y = 16, xend = 10, yend = 54),
               arrow = arrow(length = unit(0.15, "cm")),
               size = 0.3, color = "#FF0018")  +
  annotate("text", x = 11,  y = 54,
           label = "Puerto Rico paid \n lowest median wage \n of $15.9/hour",
           colour = "#CC0013",size =5) +
  geom_segment(aes(x = 53, y = 53, xend = 47, yend = 68),
               arrow = arrow(length = unit(0.15, "cm")),
               size = 0.3, color = "#1FD537")  +
  annotate("text", x = 46,  y = 66.5,
           label = "Hawaii paid \n second highest median wage \n of $53.1/hour",
           colour = "#008018", size =5) +
  annotate("text", x = c(27, 27),  y = c(38, 38),
           label = "States and Territories that paid less than $30/hour",
           colour = "#FF92E5",size =6) +
  annotate("text", x = c(27, 27),  y = c(45, 45),
           label = "States and Territories that paid in between $30 - $40/hour",
           colour = "#FFBD4A",size =6) +
  annotate("text", x = c(27, 27),  y = c(53, 53),
           label = "States and Territories that paid above $40/hour",
```

```
          colour = "#0099FF",size =6) +

# annotate("text", x = 1,  y = 70,
#          label = "2020 Wages for Registered Nurses in USA",
#          colour = "black",size =7.5, family = "Times") +

annotate("text", x = c(0.6, 0.6),  y = -23,
         label = "2020 \n Wages for \n Registered Nurses \n in USA",
         colour = "black",size =8.0, family = "Times")
```

```
wages_2020 + annotate("text", x = c(27, 27),  y = c(66, 66),
           label = "Tidy Tuesday 10-05-2021 \n Visualisation by: Kesava Asam \n Data source: Data.World
           colour = "black",size =5)
```



Few more checks

```
df_int %>% arrange(hourly_wage_median_2010) %>% select(state, hourly_wage_median_2010)
```

```
## # A tibble: 54 x 2
##    state          hourly_wage_median_2010
##    <chr>                            <dbl>
##  1 Puerto Rico                       14.3
##  2 Guam                              24.0
##  3 Iowa                              24.1
##  4 South Dakota                      24.3
##  5 West Virginia                     24.7
##  6 Arkansas                          25.3
##  7 North Dakota                      25.3
##  8 Oklahoma                          25.7
##  9 Nebraska                          25.9
## 10 Virgin Islands                    26.0
## # ... with 44 more rows
```

```
df_int %>% arrange(desc(hourly_wage_median_2010)) %>% select(state, hourly_wage_median_2010)
```

```
## # A tibble: 54 x 2
##    state                hourly_wage_median_2010
##    <chr>                                  <dbl>
##  1 California                              41.0
##  2 Hawaii                                  40.4
##  3 Massachusetts                           38.6
##  4 Alaska                                  37.4
##  5 District of Columbia                    36.5
##  6 Oregon                                  36.4
##  7 Maryland                                36.0
##  8 Nevada                                  36.0
##  9 New Jersey                              35.9
## 10 Washington                              35.2
## # ... with 44 more rows
```

```
df_int %>% arrange(hourly_wage_median_2000) %>% select(state, hourly_wage_median_2000)
```

```
## # A tibble: 54 x 2
##    state          hourly_wage_median_2000
##    <chr>                            <dbl>
##  1 Puerto Rico                        9
##  2 Virgin Islands                    17.0
##  3 Iowa                              17.5
##  4 Montana                           17.7
##  5 West Virginia                     17.7
##  6 Kansas                            17.8
##  7 Wyoming                           17.8
##  8 South Dakota                      17.9
##  9 Arkansas                          18.0
## 10 Nebraska                          18.7
## # ... with 44 more rows
```

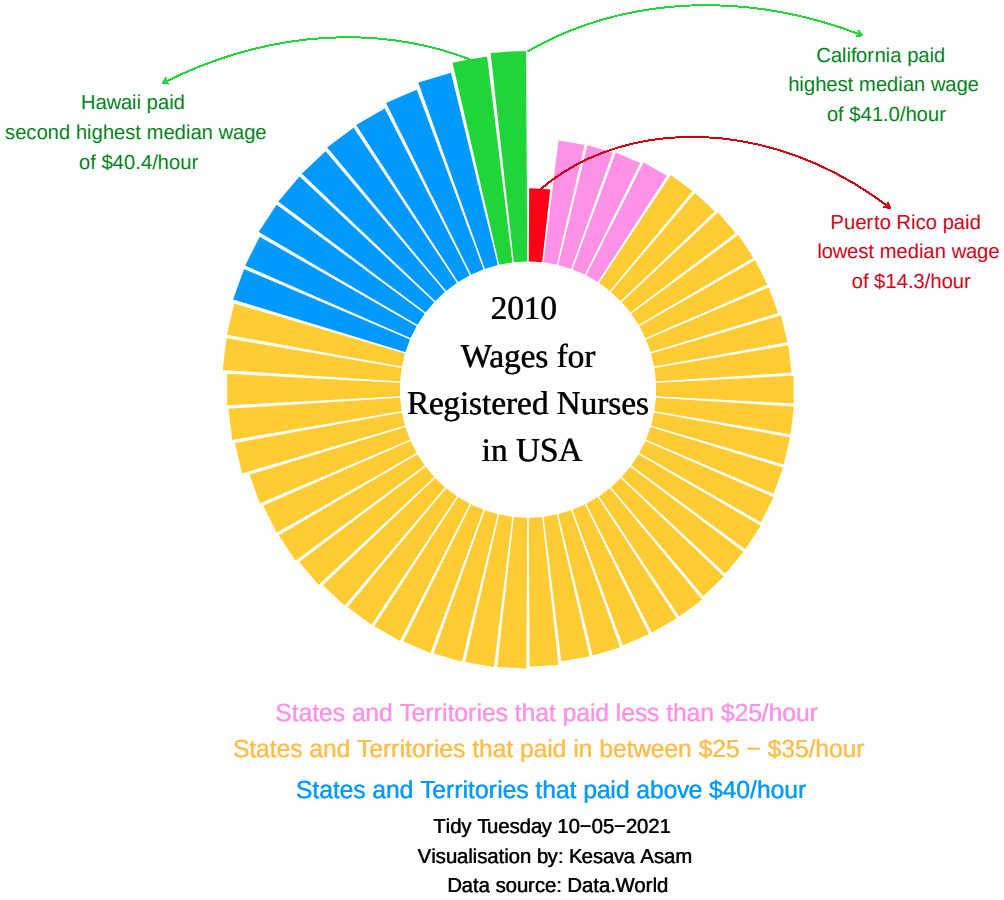
```r
nurses_rn %>% filter(year == 2005) %>%
  arrange(desc(hourly_wage_median)) %>%
  select(state, hourly_wage_median)
```

```
## # A tibble: 54 x 2
##    state         hourly_wage_median
##    <chr>                      <dbl>
##  1 California                  33.2
##  2 Maryland                    32.2
##  3 Hawaii                      31.6
##  4 Massachusetts               31.3
##  5 New Jersey                  30.0
##  6 New York                    29.5
##  7 Connecticut                 29.4
##  8 Washington                  29.2
##  9 Oregon                      29.1
## 10 Nevada                      29.0
## # ... with 44 more rows
```

```r
# Custom colors
col_man <- c("lowest" = "#FF0018",
             "low" = "#FF92E5",
             "highest" = "#1FD537",
             "high" = "#0099FF",
             "in_between" = "#FFCC33")

# Plot for 2010 wages
wages_2010 <-
  df_int %>%
  # bins for specific wages
  mutate(state = fct_reorder(state, hourly_wage_median_2010),
         color_col = factor(case_when(hourly_wage_median_2010 < 15 ~ "lowest",
                                      hourly_wage_median_2010 < 25  ~ "low",
                                      hourly_wage_median_2010 > 40 ~ "highest",
                                      hourly_wage_median_2010 > 35 ~ "high",
                                      TRUE ~ "in_between"))) %>%
  ggplot(aes(x=as.factor(state), y=hourly_wage_median_2010, fill = color_col)) +
  geom_bar(stat="identity") +
  scale_fill_manual(values = col_man, guide = "none") +
  ylim(-25,75) +
  theme_minimal() +
  coord_polar(start = 0) + #dark_mode() +
  theme(
    axis.text = element_blank(),
    axis.title = element_blank(),
    panel.grid = element_blank(),
    plot.margin = unit(rep(-3,4), "cm")) +
  geom_segment(aes(x = 0.5, y = 41, xend = 7, yend = 70),
               arrow = arrow(length = unit(0.15, "cm")),
               size = 0.3, color = "#1FD537") +
  annotate("text", x = 7.9,  y = 66.5,
           label = "California paid \n highest median wage \n of $41.0/hour",
           colour = "#008018", size =5) +
  geom_segment(aes(x = 1, y = 14, xend = 10, yend = 54),
               arrow = arrow(length = unit(0.15, "cm")),
               size = 0.3, color = "#CC0013")  +
  annotate("text", x = 11,  y = 54,
           label = "Puerto Rico paid \n lowest median wage \n of $14.3/hour",
           colour = "#CC0013",size =5) +
  geom_segment(aes(x = 53, y = 40.4, xend = 47, yend = 68),
               arrow = arrow(length = unit(0.15, "cm")),
               size = 0.3, color = "#1FD537")  +
  annotate("text", x = 46,  y = 66.5,
           label = "Hawaii paid \n second highest median wage \n of $40.4/hour",
           colour = "#008018", size =5) +

  annotate("text", x = c(27, 27),  y = c(38, 38),
           label = "States and Territories that paid less than $25/hour",
           colour = "#FF92E5",size =6) +
  annotate("text", x = c(27, 27),  y = c(45, 45),
           label = "States and Territories that paid in between $25 - $35/hour",
           colour = "#FFBD4A",size =6) +
```

```
  annotate("text", x = c(27, 27),  y = c(53, 53),
           label = "States and Territories that paid above $40/hour",
           colour = "#0099FF",size =6) +

  # annotate("text", x = 1,  y = 70,
  #          label = "2020 Wages for Registered Nurses in USA",
  #          colour = "black",size =7.5, family = "Times") +

  annotate("text", x = c(0.6, 0.6),  y = -23,
           label = "2010 \n Wages for \n Registered Nurses \n in USA",
           colour = "black",size =8.0, family = "Times")
```

```
wages_2010 +

annotate("text", x = c(27, 27),  y = c(66, 66),
           label = "Tidy Tuesday 10-05-2021 \n Visualisation by: Kesava Asam \n Data source: Data.World
           colour = "black",size =5)
```

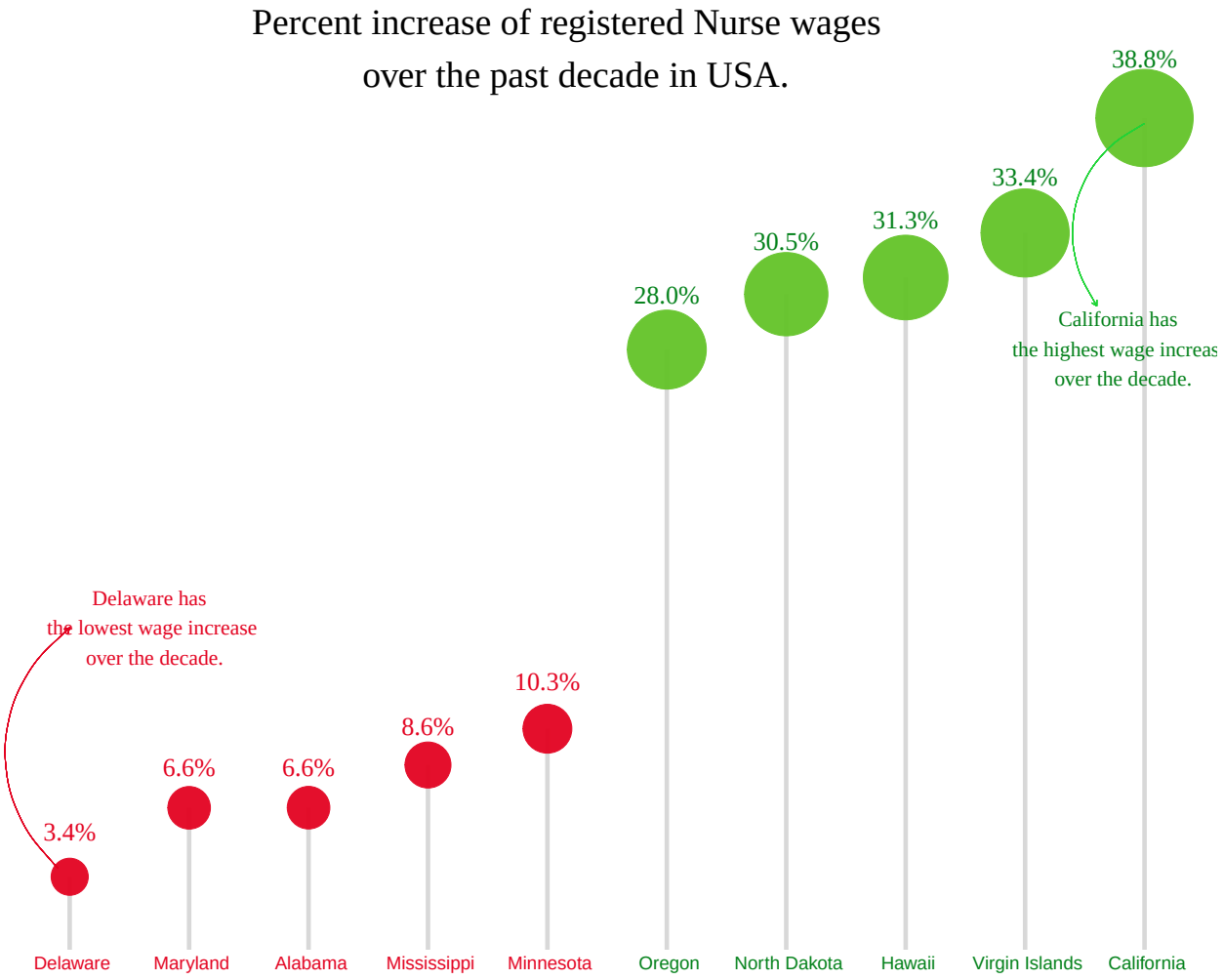


```r
# Display the States that showed highest or lowest wage increase across the decade

states_int <- c("Delaware", "Maryland", "Alabama", "Mississippi", "Minnesota",
                "California", "Virgin Islands", "Hawaii", "North Dakota", "Oregon")


col_pal <- c("low" = "#E30220",
             "high" = "#63C328")

lol_plot <-
  df_int %>%
  arrange(desc(decade_diff)) %>%
  select(state, decade_diff) %>%
  filter(state %in% states_int) %>%
  mutate(color_col = factor(case_when(decade_diff < 0.15 ~ "low",
                                      decade_diff > 0.15  ~ "high")),
         state = fct_reorder(state, decade_diff)) %>%
  ggplot(aes(state, decade_diff)) +
  geom_segment( aes(x=state, xend=state, y=0, yend=decade_diff),
                color="gray", size = 1.5, alpha = 0.6) +
  geom_point(aes(size = decade_diff, color = color_col), alpha = 0.95) +
  scale_radius(range = c(12, 32)) +
  scale_color_manual(values = col_pal, guide = "none") +
  scale_y_continuous(labels = percent_format(), breaks = seq(0,10,0.05)) +
  #coord_flip() +
  labs(y= "",
       x = "") +
  theme_void() +
  theme(
    #axis.text = element_blank(),
    #axis.title = element_blank(),
    panel.grid = element_blank(),
    legend.position = "none")
```

```
# Add annotations to identify the states and the percent increase in wages

lollipop_plot <-

  lol_plot +

  annotate("text", x = 1,  y = 0.055,
           label = "3.4%",
           colour = "#E30220", size =6.5, family = "Times") +
  annotate("text", x = 2,  y = 0.085,
           label = "6.6%",
           colour = "#E30220", size =6.5, family = "Times") +
  annotate("text", x = 3,  y = 0.085,
           label = "6.6%",
           colour = "#E30220", size =6.5, family = "Times") +
  annotate("text", x = 4,  y = 0.104,
           label = "8.6%",
           colour = "#E30220", size =6.5, family = "Times") +
  annotate("text", x = 5,  y = 0.125,
           label = "10.3%",
           colour = "#E30220", size =6.5, family = "Times") +


  annotate("text", x = 6,  y = 0.305,
           label = "28.0%",
           colour = "#008018", size =6.5, family = "Times") +
  annotate("text", x = 7,  y = 0.33,
           label = "30.5%",
           colour = "#008018", size =6.5, family = "Times") +
  annotate("text", x = 8,  y = 0.34,
           label = "31.3%",
           colour = "#008018", size =6.5, family = "Times") +
  annotate("text", x = 9,  y = 0.36,
           label = "33.4%",
           colour = "#008018", size =6.5, family = "Times") +
  annotate("text", x = 10,  y = 0.415,
           label = "38.8%",
           colour = "#008018", size =6.5, family = "Times") +



  annotate("text", x = 1,  y = -0,
           label = "\n Delaware",
           colour = "#E30220", size =4.5, family = "sans") +
  annotate("text", x = 2,  y = -0,
           label = "\n Maryland",
           colour = "#E30220", size =4.5, family = "sans") +
  annotate("text", x = 3,  y = -0,
           label = "\n Alabama",
           colour = "#E30220", size =4.5, family = "sans") +
  annotate("text", x = 4,  y = -0,
           label = "\n Mississippi",
           colour = "#E30220", size =4.5, family = "sans")+
```

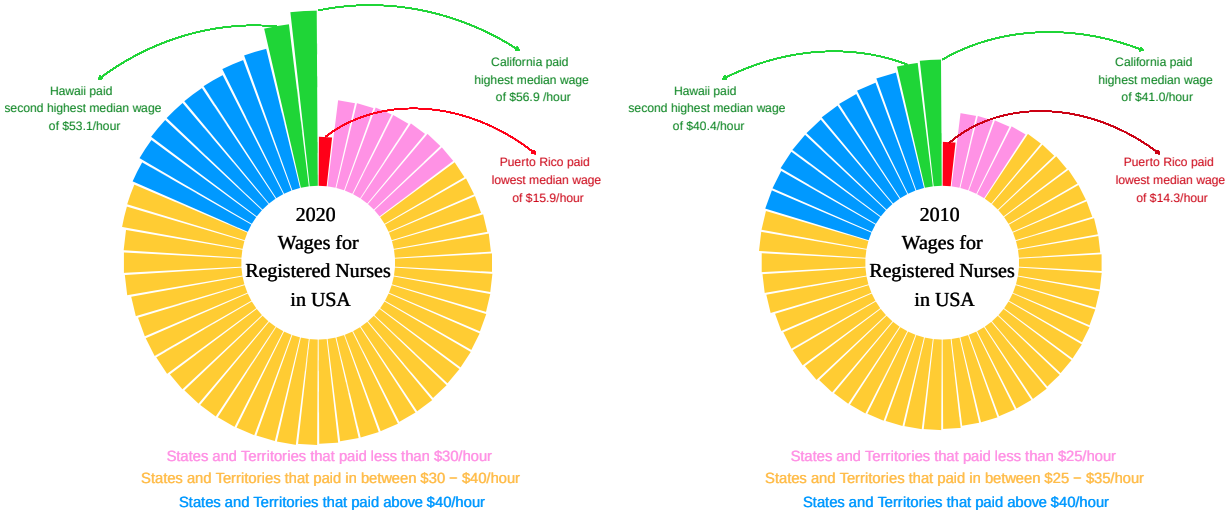
```
annotate("text", x = 5,  y = -0,
         label = "\n Minnesota",
         colour = "#E30220", size =4.5, family = "sans")+


annotate("text", x = 6,  y = -0,
         label = "\n Oregon",
         colour = "#008018", size =4.5, family = "sans")+
annotate("text", x = 7,  y = -0,
         label = "\n North Dakota",
         colour = "#008018", size =4.5, family = "sans")+
annotate("text", x = 8,  y = -0,
         label = "\n Hawaii",
         colour = "#008018", size =4.5, family = "sans")+
annotate("text", x = 9,  y = -0,
         label = "\n Virgin Islands",
         colour = "#008018", size =4.5, family = "sans")+
annotate("text", x = 10,  y = -0,
         label = "\n California",
         colour = "#008018", size =4.5, family = "sans")+


geom_curve(aes(x = 10, y = 0.385, xend = 9.6, yend = 0.30),
             arrow = arrow(length = unit(0.15, "cm")),
             size = 0.2, angle = 90, color = "#1FD537") +
annotate("text", x = 9.8,  y = 0.28,
         label = "California has \n the highest wage increase \n over the decade.",
         colour = "#008018", size =5.2, family = "Times") +


geom_curve(aes(x = 0.9, y = 0.038, xend = 1, yend = 0.15),
             arrow = arrow(length = unit(0.15, "cm")),
             size = 0.2, curvature = -0.5, angle = 90, color = "#E30220") +
annotate("text", x = 1.69,  y = 0.15,
         label = "Delaware has \n the lowest wage increase \n over the decade.",
         colour = "#E30220", size =5.2, family = "Times") +


annotate("text", x = 5.2,  y = 0.42,
         label = "Percent increase of registered Nurse wages \n over the past decade in USA.",
         colour = "black",size =9, family = "Times") +

annotate("text", x = 5.5,  y = -0.01,
         label = "\n\n Tidy Tuesday 10-05-2021. Visualisation by: Kesava Asam. Data source: Data.Worl
         colour = "black",size =5.5, family = "Times")
```

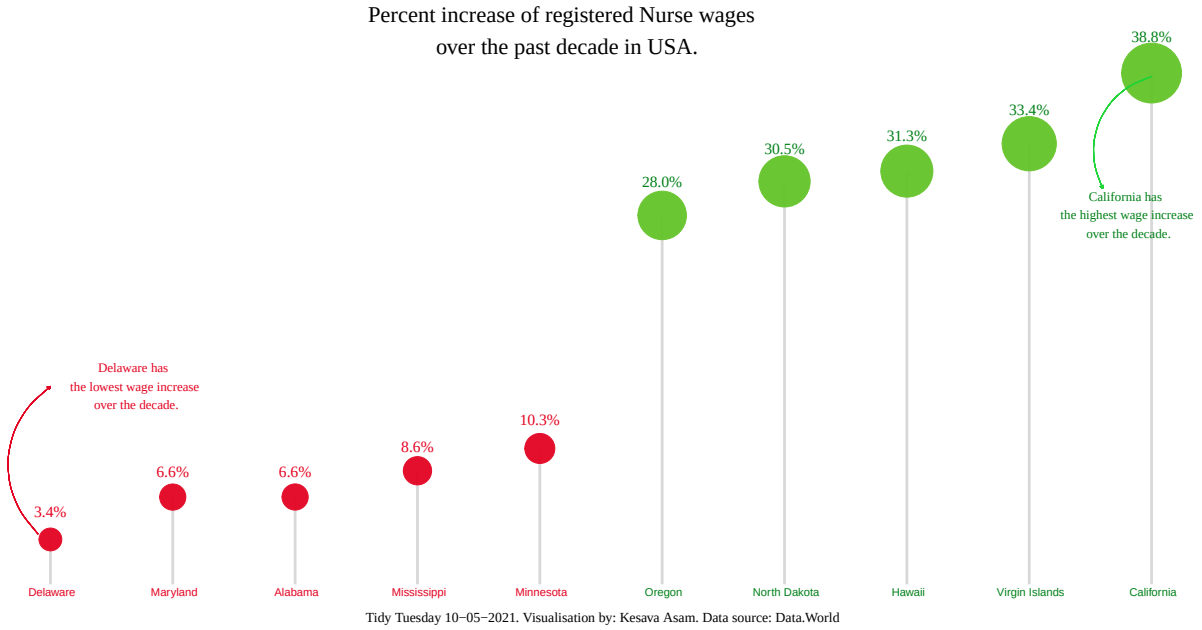
```
lollipop_plot
```



Tidy Tuesday 10−05−2021. Visualisation by: Kesava Asam. Data source: Data.World

```
# Make a collage using cowplot
p1 <- cowplot::plot_grid(wages_2020, wages_2010)

cowplot::plot_grid(p1, lollipop_plot, ncol = 1)
```





```
sessionInfo()
```

```
## R version 4.1.1 (2021-08-10)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] scales_1.1.1       janitor_2.1.0      tidytuesdayR_1.0.1 forcats_0.5.1
##  [5] stringr_1.4.0      dplyr_1.0.7        purrr_0.3.4        readr_2.0.2
##  [9] tidyr_1.1.4        tibble_3.1.5       ggplot2_3.3.5      tidyverse_1.3.1
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.7       lubridate_1.7.10 assertthat_0.2.1 digest_0.6.28
##  [5] utf8_1.2.2       R6_2.5.1         cellranger_1.1.0 backports_1.2.1
##  [9] reprex_2.0.1     evaluate_0.14    highr_0.9        httr_1.4.2
## [13] pillar_1.6.3     rlang_0.4.11     curl_4.3.2       readxl_1.3.1
## [17] rstudioapi_0.13  rmarkdown_2.11   labeling_0.4.2   selectr_0.4-2
## [21] bit_4.0.4        munsell_0.5.0    broom_0.7.9      compiler_4.1.1
## [25] modelr_0.1.8     xfun_0.26        pkgconfig_2.0.3  htmltools_0.5.2
## [29] tidyselect_1.1.1 fansi_0.5.0      crayon_1.4.1     tzdb_0.1.2
## [33] dbplyr_2.1.1     withr_2.4.2      grid_4.1.1       jsonlite_1.7.2
## [37] gtable_0.3.0     lifecycle_1.0.1  DBI_1.1.1        magrittr_2.0.1
## [41] cli_3.0.1        stringi_1.7.4    vroom_1.5.5      farver_2.1.0
## [45] fs_1.5.0         snakecase_0.11.0 xml2_1.3.2       ellipsis_0.3.2
## [49] generics_0.1.0   vctrs_0.3.8      cowplot_1.1.1    tools_4.1.1
## [53] bit64_4.0.5      glue_1.4.2       hms_1.1.1        parallel_4.1.1
## [57] fastmap_1.1.0    yaml_2.2.1       colorspace_2.0-2 rvest_1.0.1
## [61] knitr_1.36       haven_2.4.3      usethis_2.0.1
```