



Accessible data visualizations

Suggested answers

APPLICATION EXERCISE

ANSWERS

MODIFIED

November 21, 2024

```
library(tidyverse)
library(readxl)
library(scales)
library(colorblindr)

theme_set(theme_minimal(base_size = 13))
```

Import nursing data

```
nurses <- read_csv("data/nurses.csv") |> janitor::clean_names()
```

Rows: 1242 Columns: 22

— Column specification —

Delimiter: ","

chr (1): State

dbl (21): Year, Total Employed RN, Employed Standard Error (%), Hourly Wage ...

• Use `spec()` to retrieve the full column specification for this data.

• Specify the column types or set `show_col_types = FALSE` to quiet this message.

```
# subset to three states
nurses_subset <- nurses |>
  filter(state %in% c("California", "New York", "North Carolina"))

# unemployment data
unemp_state <- read_excel(
  path = "data/emp-unemployment.xls",
  sheet = "States",
  skip = 5
) |>
  pivot_longer(
    cols = -c(Fips, Area),
    names_to = "Year",
    values_to = "unemp"
  ) |>
  rename(state = Area, year = Year) |>
```

```
mutate(year = parse_number(year)) |>
filter(state != "United States") |>
# calculate mean unemp rate per state and year
group_by(state, year) |>
summarize(unemp_rate = mean(unemp, na.rm = TRUE))
```

`summarise()` has grouped output by 'state'. You can override using the `.groups` argument.

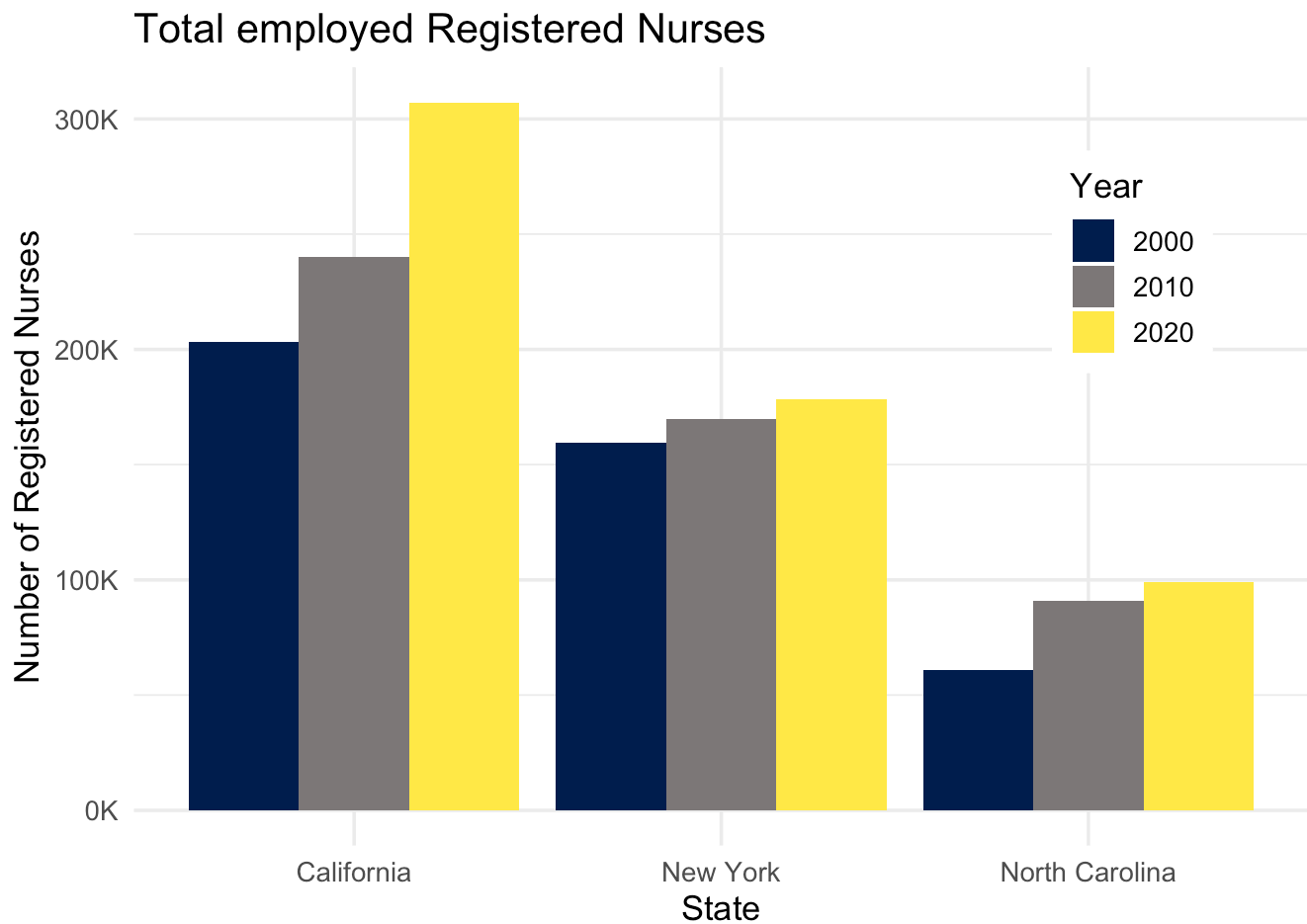
Developing alternative text

Bar chart

Demonstration: The following code chunk demonstrates how to add alternative text to a bar chart. The alternative text is added to the chunk header using the `fig-alt` chunk option. The text is written in Markdown and can be as long as needed. Note that `fig-cap` is not the same as `fig-alt`.

```
```{r}
#| label: nurses-bar
#| fig-cap: "Total employed Registered Nurses"
#| fig-alt: "The figure is a bar chart titled 'Total employed Registered
#| Nurses' that displays the numbers of registered nurses in three states
#| (California, New York, and North Carolina) over a 20 year period, with data
#| recorded in three time points (2000, 2010, and 2020). In each state, the
#| numbers of registered nurses increase over time. The following numbers are
#| all approximate. California started off with 200K registered nurses in 2000,
#| 240K in 2010, and 300K in 2020. New York had 150K in 2000, 160K in 2010, and
#| 170K in 2020. Finally North Carolina had 60K in 2000, 90K in 2010, and 100K
#| in 2020."

nurses_subset |>
 filter(year %in% c(2000, 2010, 2020)) |>
 ggplot(aes(x = state, y = total_employed_rn, fill = factor(year))) +
 geom_col(position = "dodge") +
 scale_fill_viridis_d(option = "E", guide = guide_legend(position = "inside")) +
 scale_y_continuous(labels = label_number(scale = 1 / 1000, suffix = "K")) +
 labs(
 x = "State", y = "Number of Registered Nurses", fill = "Year",
 title = "Total employed Registered Nurses"
) +
 theme(
 legend.background = element_rect(fill = "white", color = "white"),
 legend.position.inside = c(0.85, 0.75)
)
```
```



Total employed Registered Nurses

Line chart

Your turn: Add alternative text to the following line chart.

Tip

Remember the major components of alt text:

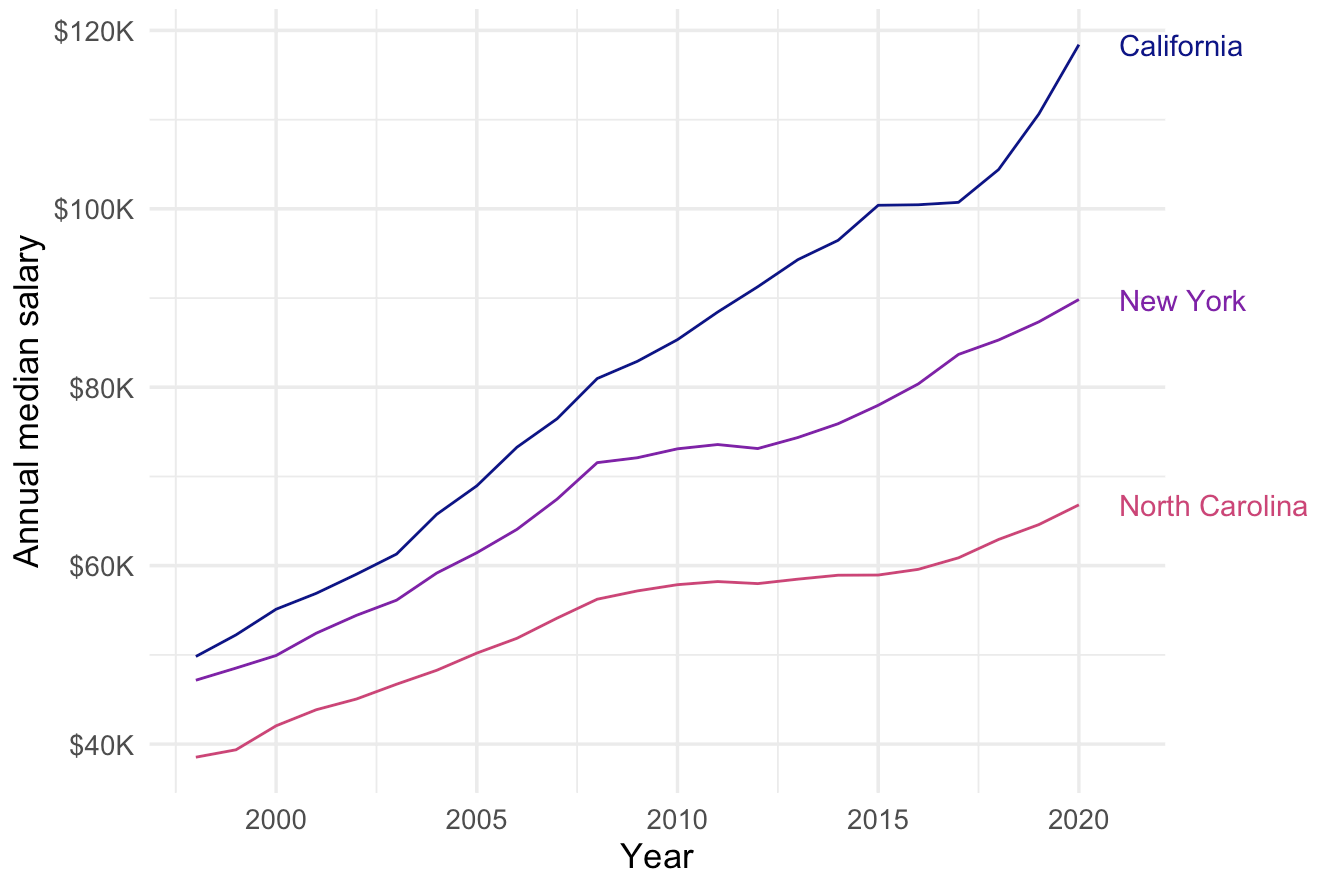
- **CHART TYPE**: It's helpful for people with partial sight to know what chart type it is and gives context for understanding the rest of the visual.
- **TYPE OF DATA**: What data is included in the chart? The x and y axis labels may help you figure this out.
- **REASON FOR INCLUDING CHART**: Think about why you're including this visual. What does it show that's meaningful. There should be a point to every visual and you should tell people what to look for.
- **Link to data source**: Don't include this in your alt text, but it should be included somewhere in the surrounding text.

```
```{r}
#| label: nurses-line
#| fig-alt: 'The figure is titled "Annual median salary of Registered Nurses".
#| There are three lines on the plot: the top labelled California, the middle
#| New York, the bottom North Carolina. The vertical axis is labelled "Annual
```

```
#| median salary", beginning with $40K, up to $120K. The horizontal axis is
#| labelled "Year", beginning with couple years before 2000 up to 2020. The
#| following numbers are all approximate. In the graph, the California line
#| begins around $50K in 1998 and goes up to $120K in 2020. The increase is
#| steady, except for stalling for about couple years between 2015 to 2017.
#| The New York line also starts around $50K, just below where the California
#| line starts, and steadily goes up to $90K. And the North Carolina line starts
#| around $40K and steadily goes up to $70K.'
```

```
nurses_subset |>
 ggplot(aes(x = year, y = annual_salary_median, color = state)) +
 geom_line(show.legend = FALSE) +
 geom_text(
 data = nurses_subset |> filter(year == max(year)),
 aes(label = state), hjust = 0, nudge_x = 1,
 show.legend = FALSE
) +
 scale_color_viridis_d(option = "C", end = 0.5) +
 scale_y_continuous(labels = label_currency(scale = 1 / 1000, suffix = "K")) +
 labs(
 x = "Year", y = "Annual median salary", color = "State",
 title = "Annual median salary of Registered Nurses"
) +
 coord_cartesian(clip = "off") +
 theme(
 plot.margin = margin(0.1, 0.9, 0.1, 0.1, "in")
)
````
```

Annual median salary of Registered Nurses



Scatterplot

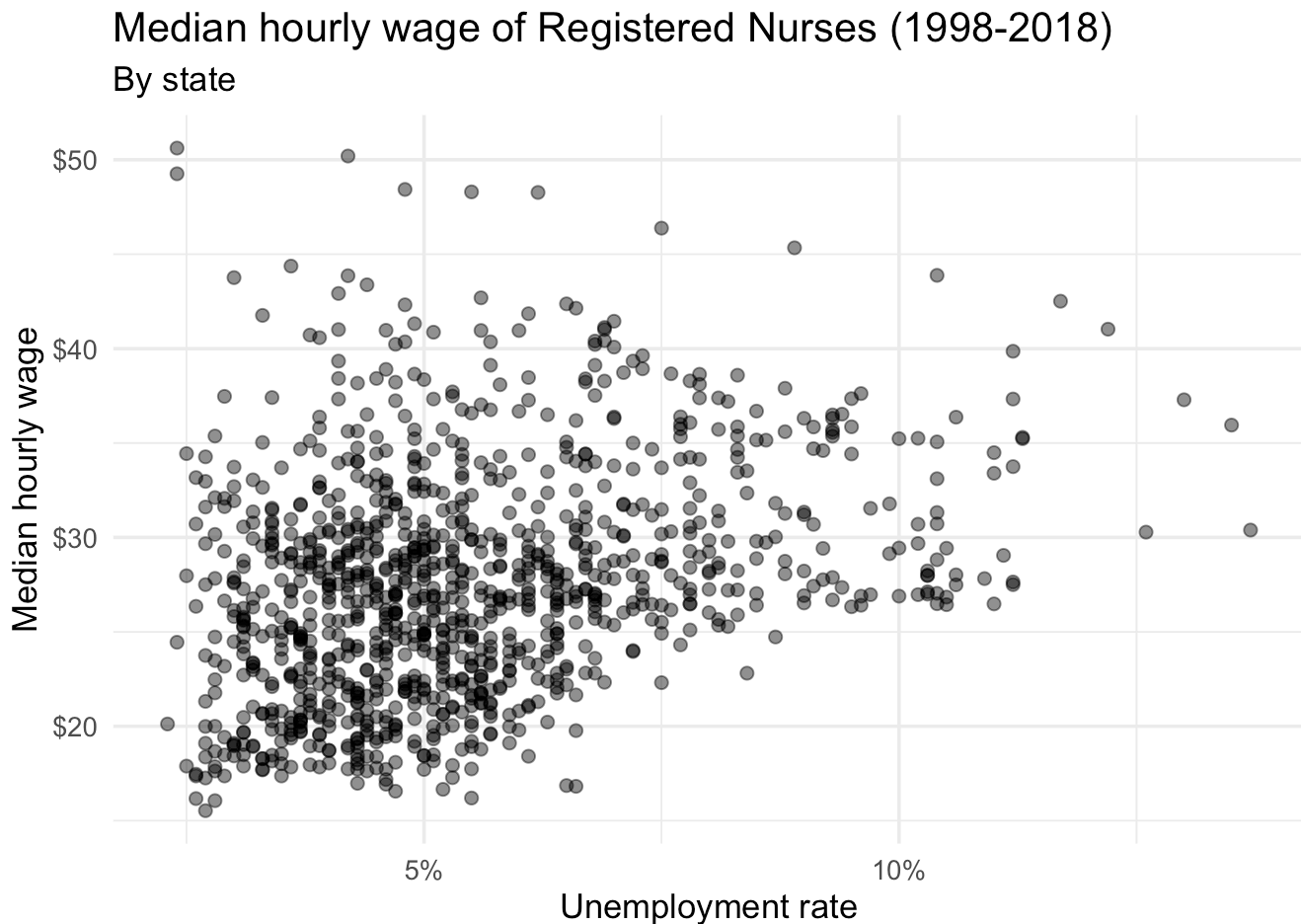
Your turn: Add alternative text to the following scatterplot.

```
```{r}
#| label: nurses-scatter
#| fig-alt: 'The figure is titled "Median hourly wage of Registered Nurses".
#| It is a scatter plot with points for each of the 50 U.S. states from 1998
#| to 2008. The horizontal axis is labeled "Unemployment rate", beginning
#| around 2% up to 14%. The horizontal axis is labelled "Median hourly wage",
#| beginning with amounts under $20 up to approximately $50. The pattern is
#| hard to discern but appears to show a positive correlation between the
#| variables. As unemployment rate increases the median hourly wage also
#| slightly increases. There is more variability in median hourly wage for
#| unemployment rates below 7%.'

nurses |>
 left_join(unemp_state) |>
 drop_na(unemp_rate) |>
 ggplot(aes(x = unemp_rate, y = hourly_wage_median)) +
 geom_point(size = 2, alpha = .5) +
 scale_x_continuous(labels = label_percent(scale = 1)) +
```

```
scale_y_continuous(labels = label_currency()) +
labs(
 x = "Unemployment rate", y = "Median hourly wage",
 title = "Median hourly wage of Registered Nurses (1998-2018)",
 subtitle = "By state"
)
...
```

Joining with ``by = join_by(state, year)``



## Acknowledgments

- Exercise drawn from [STA 313: Advanced Data Visualization](#)

### Session information

