Week9-AE-RMarkdown

2024-03-27

```
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(readxl)
library(scales)
```

```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
#install.packages('janitor')
library(janitor)
```

```
## Warning: package 'janitor' was built under R version 4.3.3
```

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

```
nurses <- read_csv("nurses.csv") |> clean_names()
```

```
## Rows: 1242 Columns: 22
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): State
## dbl (21): Year, Total Employed RN, Employed Standard Error (%), Hourly Wage ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i 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"))
```

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`.

```
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") +
  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 = c(0.85, 0.75)
    )
```

```
## Warning: A numeric `legend.position` argument in `theme()` was deprecated in ggplot2
## 3.5.0.
## i Please use the `legend.position.inside` argument of `theme()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

Task 1. Add alt text to line chart

```
# Your label here
#| label: nurses-salary
# Your caption here
#| fig-cap: "Annual median salary of Registered Nurses"
# Your alt text here
#| fig-alt: "The figure is a line chart titled 'Annual median salary of Registered Nurses' that displays
```

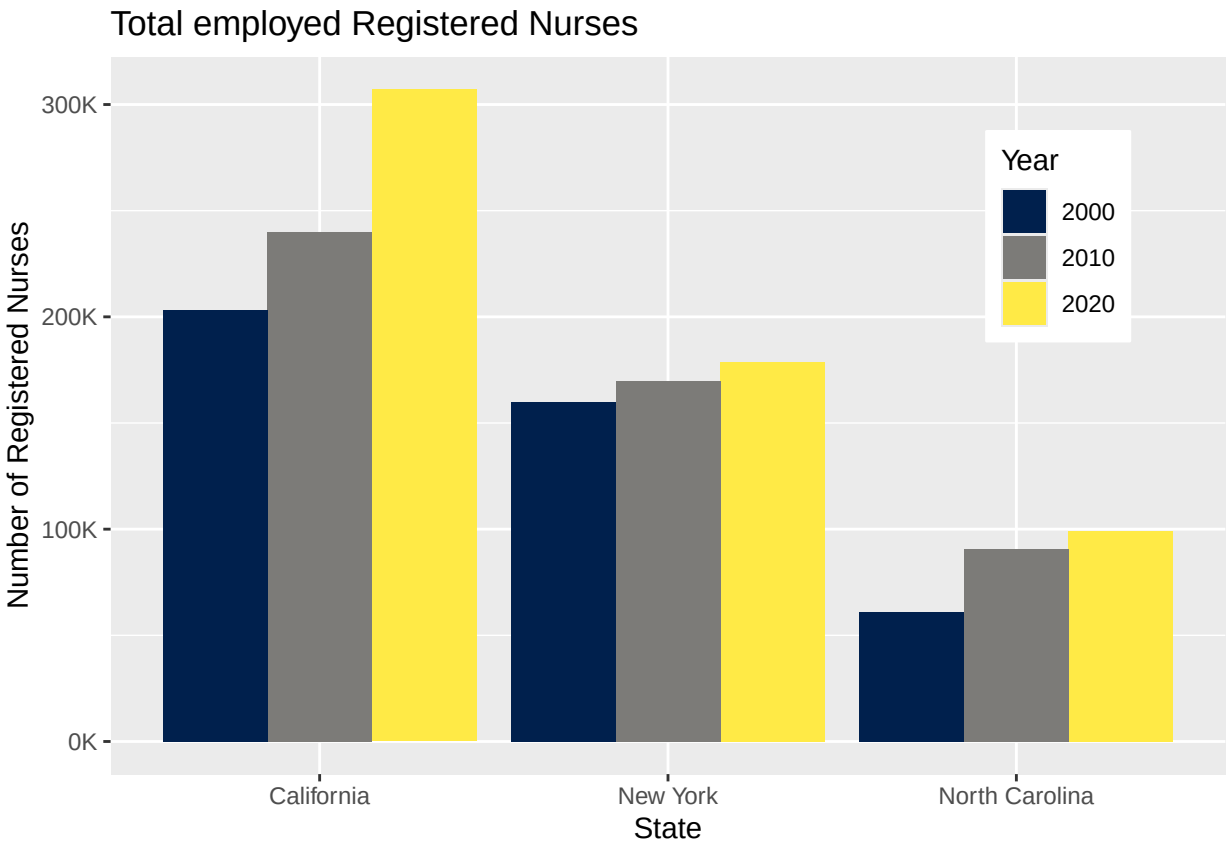
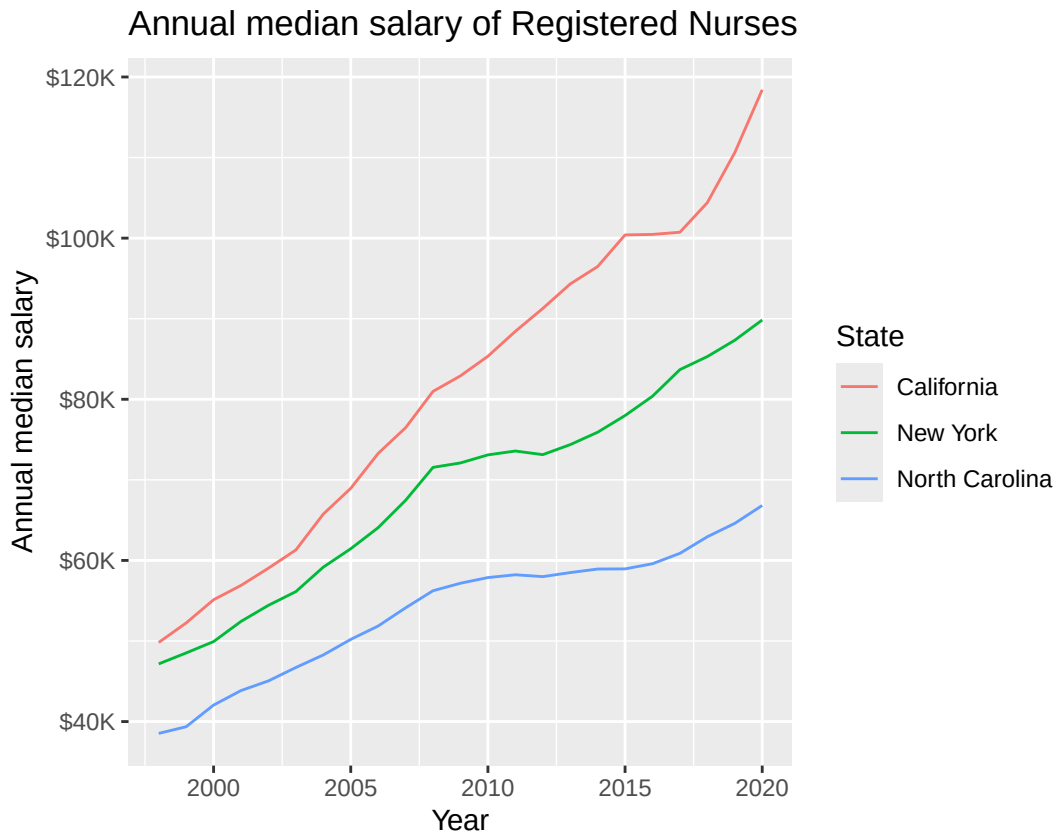


Figure 1: Total employed Registered Nurses

```
#| (California, New York, and North Carolina) from 1998 to 2020. In each state, the
#| median salary of registered nurses increase over time. The following numbers are
#| all approximate. California nurses have the highest median salary, starting from $50K in 1998 to nea

nurses_subset |>
  ggplot(aes(x = year, y = annual_salary_median, color = state)) +
  geom_line() +
  scale_y_continuous(labels = label_dollar(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")
    )
```



Task 2. Direct labelling instead of legends

Create a version of the same line chart but using direct labels instead of legends.

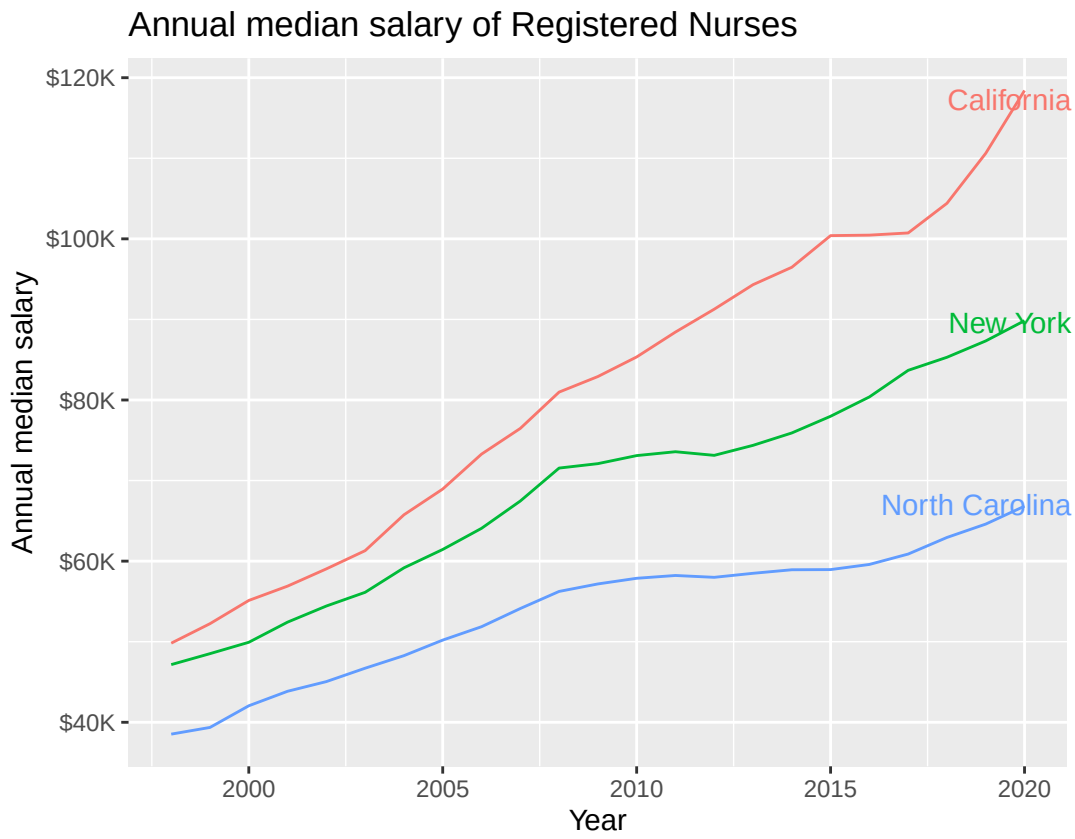
You can achieve this with just `geom_text()` but you can also check out https://r-graph-gallery.com/web-line-chart-with-labels-at-end-of-line.html for a fancier way of achieving this.

```r
library(ggrepel)
```

```
## Warning: package 'ggrepel' was built under R version 4.3.3
```

```r
# YOUR CODE HERE
# Your label here
#| label: nurses-salary
# Your caption here
#| fig-cap: "Annual median salary of Registered Nurses"
# Your alt text here
#| fig-alt: "The figure is a line chart titled 'Annual median salary of Registered Nurses' that displays
#| (California, New York, and North Carolina) from 1998 to 2020. In each state, the
#| median salary of registered nurses increase over time. The following numbers are
#| all approximate. California nurses have the highest median salary, starting from $50K in 1998 to nea

nurses_subset %>%
  mutate(labels = ifelse(year == 2020, state, NA))|>
  ggplot(aes(x = year, y = annual_salary_median, color = state)) +
  geom_line(show.legend = FALSE) +
  geom_text_repel(
    aes(label = labels),
    direction = "y",
    xlim = c(2022, NA),
    segment.size = .7,
    segment.alpha = .5,
    segment.linetype = "dotted",
    box.padding = .8,
    segment.curvature = -0.1,
    segment.ncp = 3,
    segment.angle = 20,
    show.legend = FALSE
  ) +
  scale_y_continuous(labels = label_dollar(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")
    )
```

```
## Warning: Removed 66 rows containing missing values or values outside the scale range
## (`geom_text_repel()`).
```



Task 3. Colorblind-friendly plots

Use `colorblindr` for colorblind-friendly palettes.

```
# remotes::install_github("wilkelab/cowplot")
# install.packages("colorspace", repos = "http://R-Forge.R-project.org")
# remotes::install_github("clauswilke/colorblindr")
library(colorblindr)
```

```
## Loading required package: colorspace
```

Try out colorblind simulations at http://hclwizard.org/cvdemulator/ or |> your plot to `cvd_grid()` to see the plot in various color-vision-deficiency simulations.

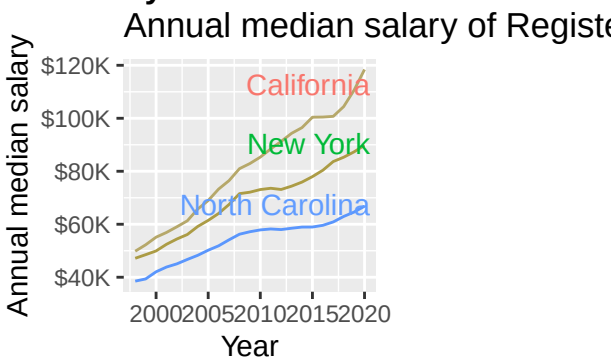
With the line chart from Task 1, create 3 different plots: one with the default color scale, one with the `viridis` color scale, and one with the `OkabeIto` color scale from `colorblindr`. Show the `cvd_grid()` of each plot and describe the simulated effectiveness of the color scales for colorblind viewers.

```
# YOUR CODE HERE for default color scale
# plot <-nurses_subset |>
#   ggplot(aes(x = year, y = annual_salary_median, color = state)) +
#   geom_line() +
#   scale_y_continuous(labels = label_dollar(scale = 1/1000, suffix = "K")) +
#   labs(
```

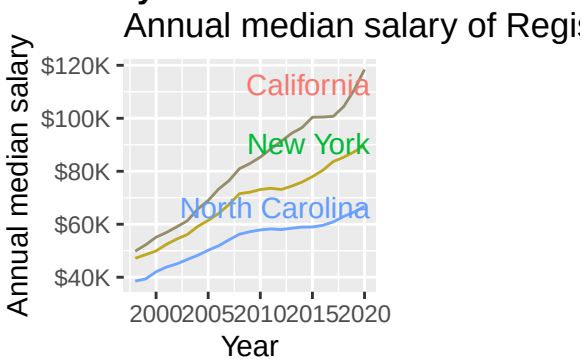
```
#     x = "Year", y = "Annual median salary", color = "State",
#     title = "Annual median salary of Registered Nurses"
#   ) +
#   coord_cartesian(clip = "off") +
#   theme(
#     plot.margin = margin(0.1, 0.9, 0.1, 0.1, "in")
#     )
# cvd_grid(plot)

plot <- nurses_subset %>%
  mutate(labels = ifelse(year == 2020, state, NA))|>
  ggplot(aes(x = year, y = annual_salary_median, color = state)) +
  geom_line(show.legend = FALSE) +
  geom_text_repel(
    aes(label = labels),
    direction = "y",
    xlim = c(2022, NA),
    segment.size = .7,
    segment.alpha = .5,
    segment.linetype = "dotted",
    box.padding = .4,
    segment.curvature = -0.1,
    segment.ncp = 3,
    segment.angle = 20,
    show.legend = FALSE
  ) +
  scale_y_continuous(labels = label_dollar(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"),
    legend.position = "none"
    )
cvd_grid(plot)
```

```
## Warning: Removed 66 rows containing missing values or values outside the scale range
## (`geom_text_repel()`).
## Removed 66 rows containing missing values or values outside the scale range
## (`geom_text_repel()`).
## Removed 66 rows containing missing values or values outside the scale range
## (`geom_text_repel()`).
## Removed 66 rows containing missing values or values outside the scale range
## (`geom_text_repel()`).
```

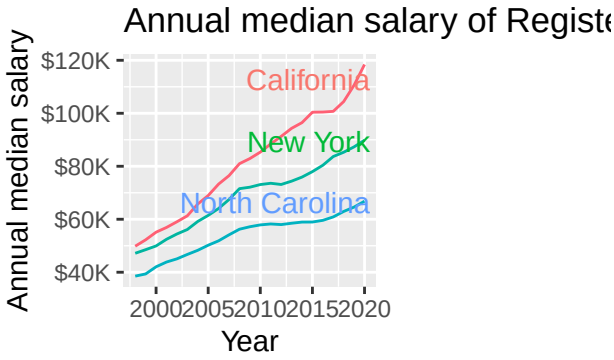
Deutanomaly



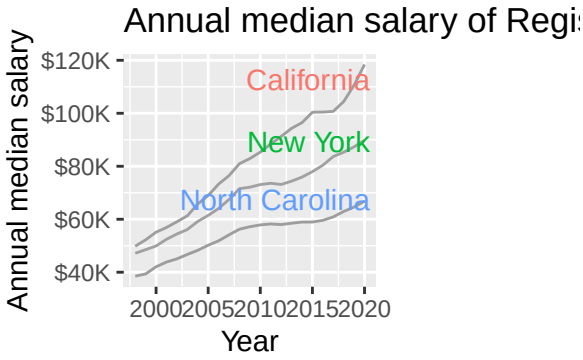
Protanomaly



Tritanomaly



Desaturated



What do you think of the default color scale effectiveness for colorblind viewers?

YOUR ANSWER HERE

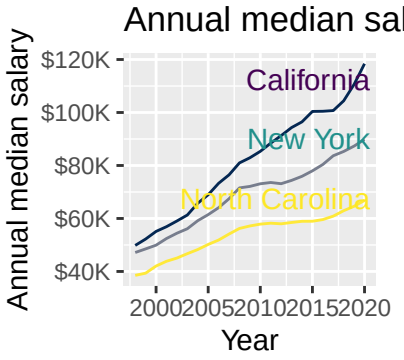
Not very effective. It's hard to differentiate the lines for California and New York, especially in the cases of deutanomaly and desaturated.

```
# YOUR CODE HERE for viridis color scale
plot <- nurses_subset %>%
  mutate(labels = ifelse(year == 2020, state, NA))|>
  ggplot(aes(x = year, y = annual_salary_median, color = state)) +
  geom_line(show.legend = FALSE) +
  geom_text_repel(
    aes(label = labels),
    direction = "y",
    xlim = c(2022, NA),
    segment.size = .7,
    segment.alpha = .5,
    segment.linetype = "dotted",
    box.padding = .4,
    segment.curvature = -0.1,
    segment.ncp = 3,
    segment.angle = 20,
    show.legend = FALSE
  ) +
```

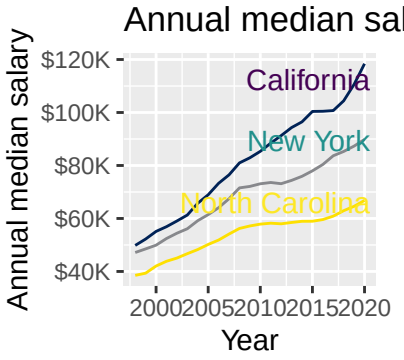
```
  scale_y_continuous(labels = label_dollar(scale = 1/1000, suffix = "K")) +
  labs(
    x = "Year", y = "Annual median salary", color = "State",
    title = "Annual median salary of Registered Nurses"
  ) +
  coord_cartesian(clip = "off") +
  theme(
    plot.margin = margin(0.1, 0.9, 0.1, 0.1, "in"),
    legend.position = "none"
    ) +
  scale_color_viridis_d()
cvd_grid(plot)
```

```
## Warning: Removed 66 rows containing missing values or values outside the scale range
## (`geom_text_repel()`).
## Removed 66 rows containing missing values or values outside the scale range
## (`geom_text_repel()`).
## Removed 66 rows containing missing values or values outside the scale range
## (`geom_text_repel()`).
## Removed 66 rows containing missing values or values outside the scale range
## (`geom_text_repel()`).
```

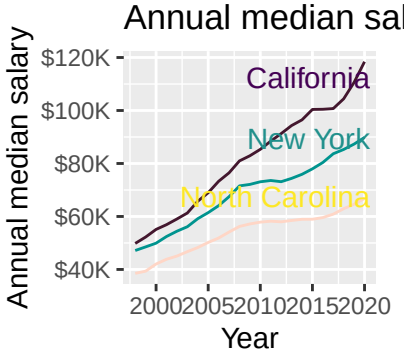
Deutanomaly



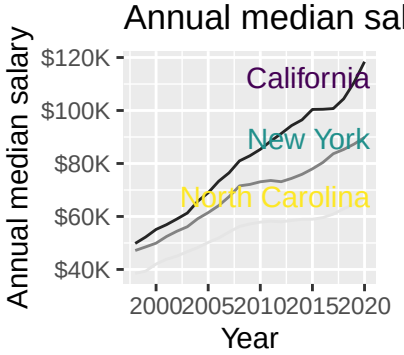
Protanomaly



Tritanomaly



Desaturated



What do you think of the viridis color scale effectiveness for colorblind viewers?

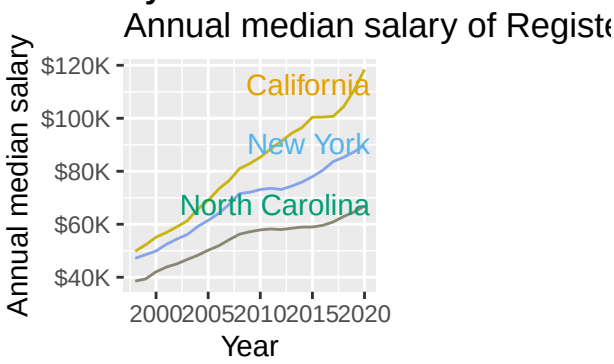
YOUR ANSWER HERE

Still quite hard to differentiate between California's line and New York's line; North Carolina is hard to see overall.

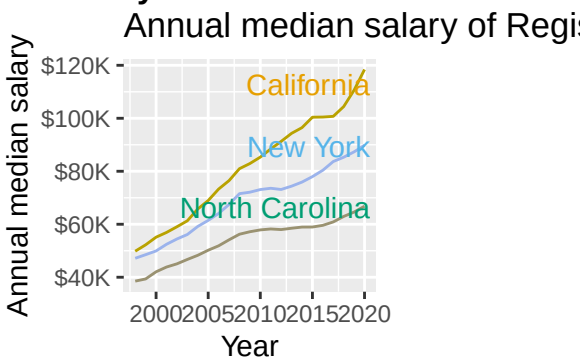
```
# YOUR CODE HERE for OkabeIto color scale
plot <- nurses_subset %>%
  mutate(labels = ifelse(year == 2020, state, NA))|>
  ggplot(aes(x = year, y = annual_salary_median, color = state)) +
  geom_line(show.legend = FALSE) +
  geom_text_repel(
    aes(label = labels),
    direction = "y",
    xlim = c(2022, NA),
    segment.size = .7,
    segment.alpha = .5,
    segment.linetype = "dotted",
    box.padding = .4,
    segment.curvature = -0.1,
    segment.ncp = 3,
    segment.angle = 20,
    show.legend = FALSE
  ) +
  scale_y_continuous(labels = label_dollar(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"),
    legend.position = "none"
    ) +
  scale_color_OkabeIto()
cvd_grid(plot)
```

```
## Warning: Removed 66 rows containing missing values or values outside the scale range
## (`geom_text_repel()`).
## Removed 66 rows containing missing values or values outside the scale range
## (`geom_text_repel()`).
## Removed 66 rows containing missing values or values outside the scale range
## (`geom_text_repel()`).
## Removed 66 rows containing missing values or values outside the scale range
## (`geom_text_repel()`).
```

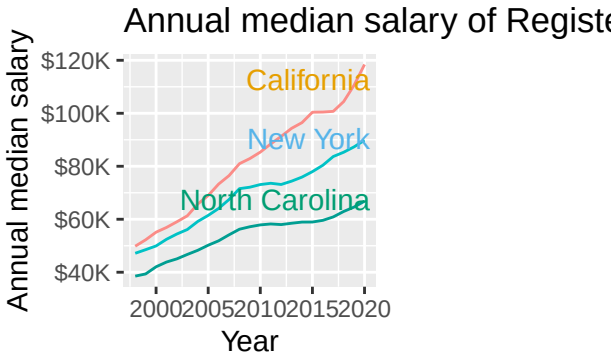
Deutanomaly



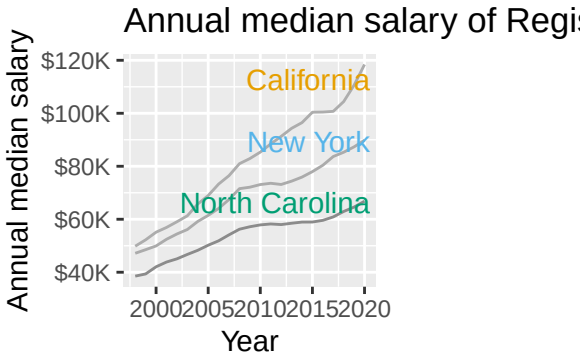
Protanomaly



Tritanomaly



Desaturated



What do you think of the OkabeIto color scale effectiveness for colorblind viewers?

YOUR ANSWER HERE

Easier to differentiate in all cases except perhaps desaturated.