



# AE 06: Take a sad plot, and make it better

## Suggested answers

[APPLICATION EXERCISE](#)[ANSWERS](#)

MODIFIED

February 12, 2025

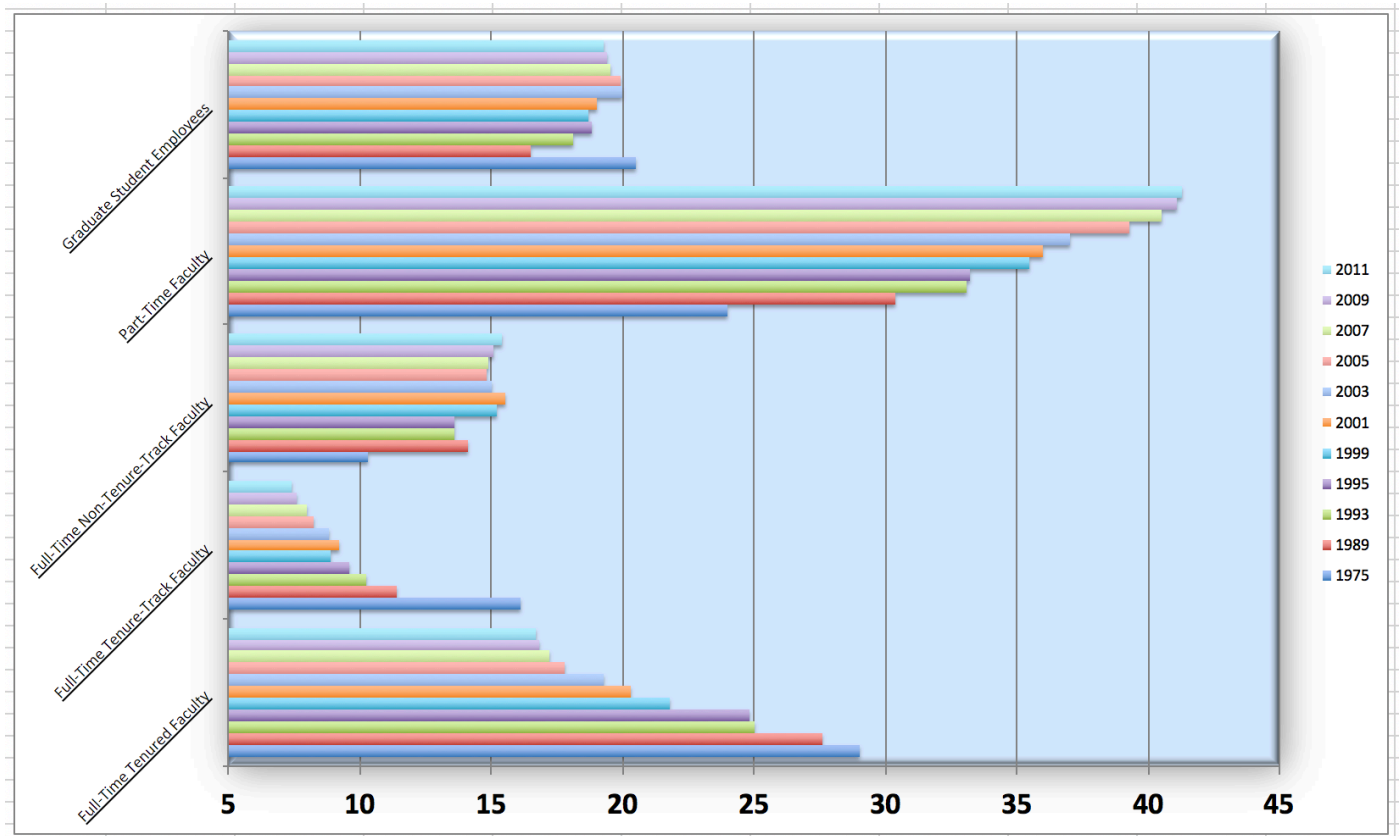
### Important

These are suggested answers. This document should be used as reference only, it's not designed to be an exhaustive key.

```
library(tidyverse)
library(scales)
library(ggthemes)
```

## Take a sad plot, and make it better

The American Association of University Professors (AAUP) is a nonprofit membership association of faculty and other academic professionals. [This report](#) by the AAUP shows trends in instructional staff employees between 1975 and 2011, and contains an image very similar to the one given below.



The data series has been extended through 2021.<sup>1</sup>

Each row in this dataset represents a faculty type, and the columns are the years for which we have data. The values are percentage of hires of that type of faculty for each year.

```
staff <- read_csv("data/instructional-staff.csv")
staff
```

# A tibble: 5 × 12

faculty_type	'1975'	'1989'	'1993'	'1995'	'1999'	'2001'	'2003'	'2005'	'2007'
1 Full-Time Tenu...	29	27.6	25	24.8	21.8	20.3	19.3	17.8	17.2
2 Full-Time Tenu...	16.1	11.4	10.2	9.6	8.9	9.2	8.8	8.2	8
3 Full-Time Non-...	10.3	14.1	13.6	13.6	15.2	15.5	15	14.8	14.9
4 Part-Time Facu...	24	30.4	33.1	33.2	35.5	36	37	39.3	40.5
5 Graduate Stude...	20.5	16.5	18.1	18.8	18.7	19	20	19.9	19.5

# i 2 more variables: '2009' <dbl>, '2011' <dbl>

## Recreate the visualization

In order to recreate this visualization we need to first reshape the data to have one variable for faculty type and one variable for year. In other words, we will convert the data from the **wide format** to **long format**.

**Your turn:** Reshape the data so we have one row per faculty type and year, and the percentage of hires as a single column.

```
staff_long <- staff |>
  pivot_longer(
    cols = -faculty_type,
    names_to = "year",
    values_to = "percentage"
  )
staff_long
```

# A tibble: 55 × 3

	faculty_type	year	percentage
	<chr>	<chr>	<dbl>
1	Full-Time Tenured Faculty	1975	29
2	Full-Time Tenured Faculty	1989	27.6
3	Full-Time Tenured Faculty	1993	25
4	Full-Time Tenured Faculty	1995	24.8
5	Full-Time Tenured Faculty	1999	21.8
6	Full-Time Tenured Faculty	2001	20.3
7	Full-Time Tenured Faculty	2003	19.3
8	Full-Time Tenured Faculty	2005	17.8
9	Full-Time Tenured Faculty	2007	17.2
10	Full-Time Tenured Faculty	2009	16.8

# i 45 more rows

**Your turn:** Attempt to recreate the original bar chart as best as you can. Don't worry about theming or color palettes right now. The most important aspects to incorporate:

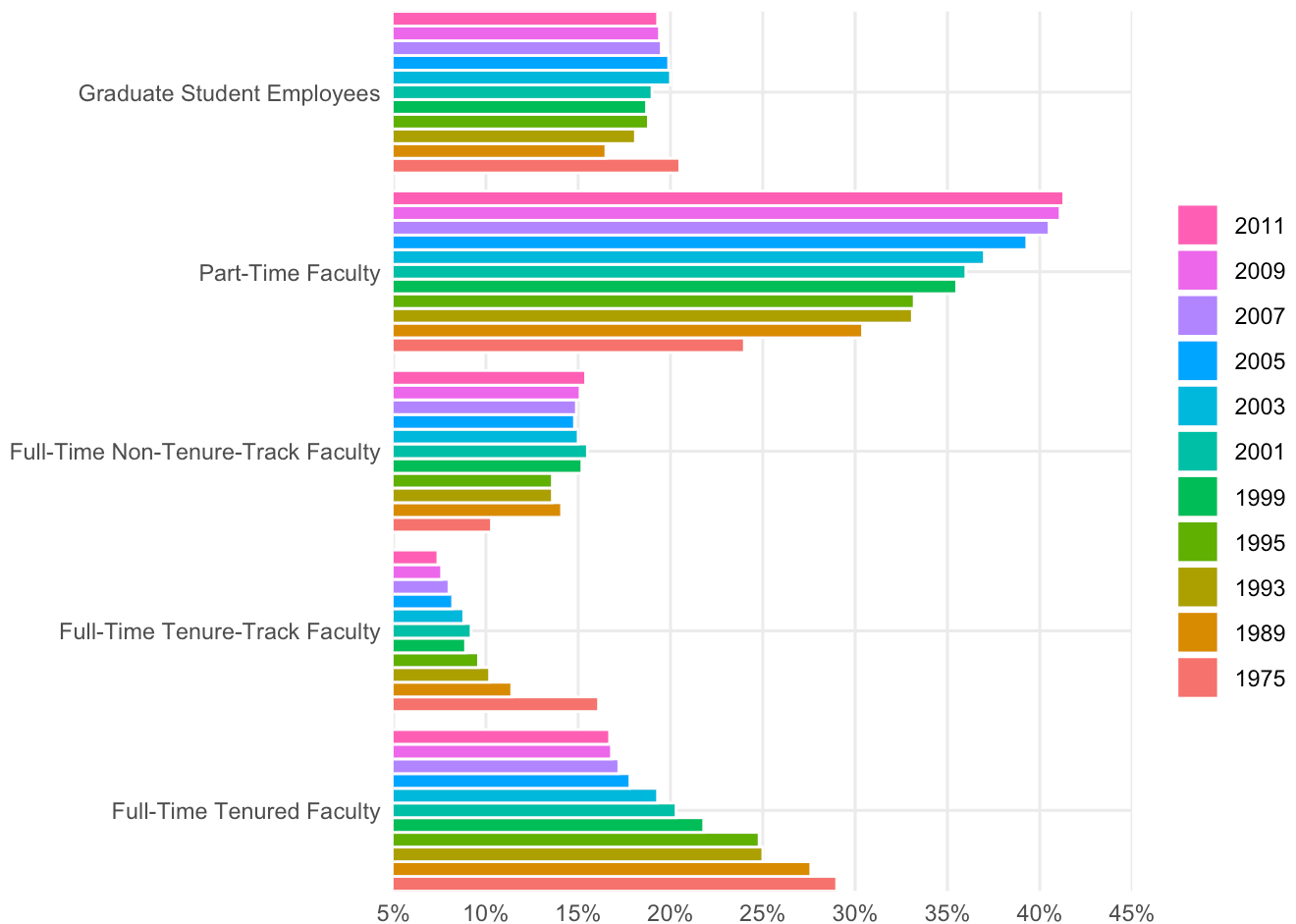
- Faculty type on the *y*-axis with bar segments color-coded based on the year of the survey
- Percentage of instructional staff employees on the *x*-axis
- Begin the *x*-axis at 5%
- Label the *x*-axis at 5% increments
- Match the order of the legend

#### Tip

[{forcats}](#) contains many functions for defining and adjusting the order of levels for **factor** variables. Factors are often used to enforce specific ordering of categorical variables in charts.

```
staff_long |>
  # convert faculty_type to factor to ensure correct order
  mutate(faculty_type = fct_relevel(
    .f = faculty_type,
    "Full-Time Tenured Faculty",
```

```
"Full-Time Tenure-Track Faculty",
"Full-Time Non-Tenure-Track Faculty",
"Part-Time Faculty",
"Graduate Student Employees"
)) |>
ggplot(mapping = aes(x = percentage, y = faculty_type, fill = year)) +
# position dodge to separate the bars
geom_col(position = "dodge", color = "white") +
# generate a sequence of breaks from 5 to 45
scale_x_continuous(
  breaks = seq(from = 5, to = 45, by = 5),
  labels = label_percent(scale = 1)
) +
# reverse the legend values
guides(fill = guide_legend(reverse = TRUE)) +
# no labels on the chart
labs(
  x = NULL,
  y = NULL,
  fill = NULL
) +
# crop the chart to begin at an origin of 5
coord_cartesian(xlim = c(5, 45), expand = FALSE) +
# attempt to match the visual design
theme_minimal() +
theme(
  panel.grid.minor = element_blank()
)
```



## Let's make it better

The original plot is not very informative. It's hard to compare the trends for across each faculty type.

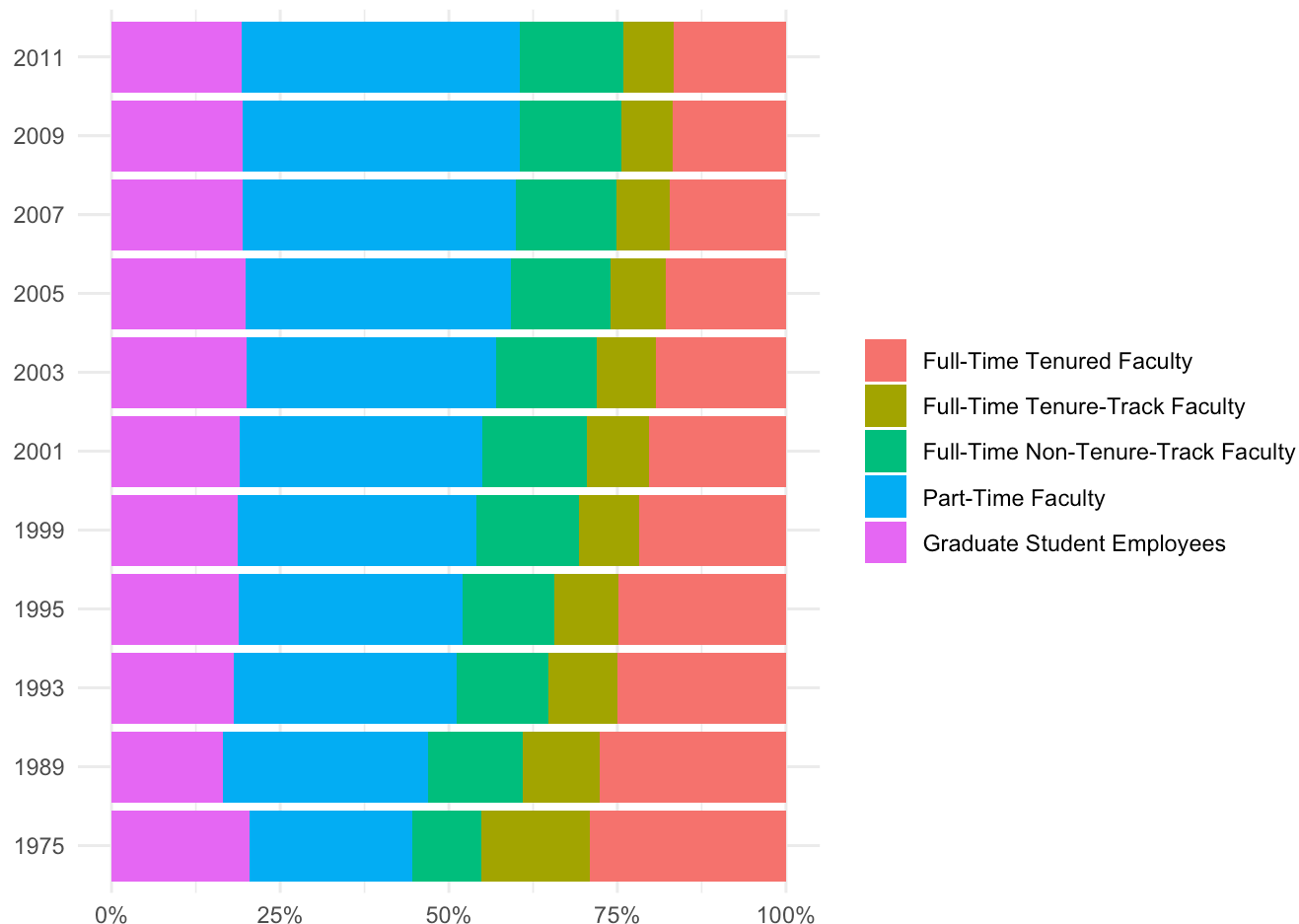
**Your turn:** Improve the chart by using a **relative frequency bar chart** with year on the *y*-axis and faculty type encoded using color.

```
staff_long |>
  mutate(faculty_type = fct_relevel(
    .f = faculty_type, "Full-Time Tenured Faculty",
    "Full-Time Tenure-Track Faculty",
    "Full-Time Non-Tenure-Track Faculty",
    "Part-Time Faculty",
    "Graduate Student Employees"
  )) |>
  ggplot(mapping = aes(x = percentage, y = year, fill = faculty_type)) +
  geom_col(position = "fill") +
  scale_x_continuous(labels = label_percent()) +
  labs(
    x = NULL,
```

```

y = NULL,
fill = NULL
) +
theme_minimal()

```



What are this chart's advantages and disadvantages? [Add response here](#)

This distorts the intervals for the `year` variable. It makes it appear as if the survey was conducted at regular intervals, which is not the case.

## Now we want a line chart

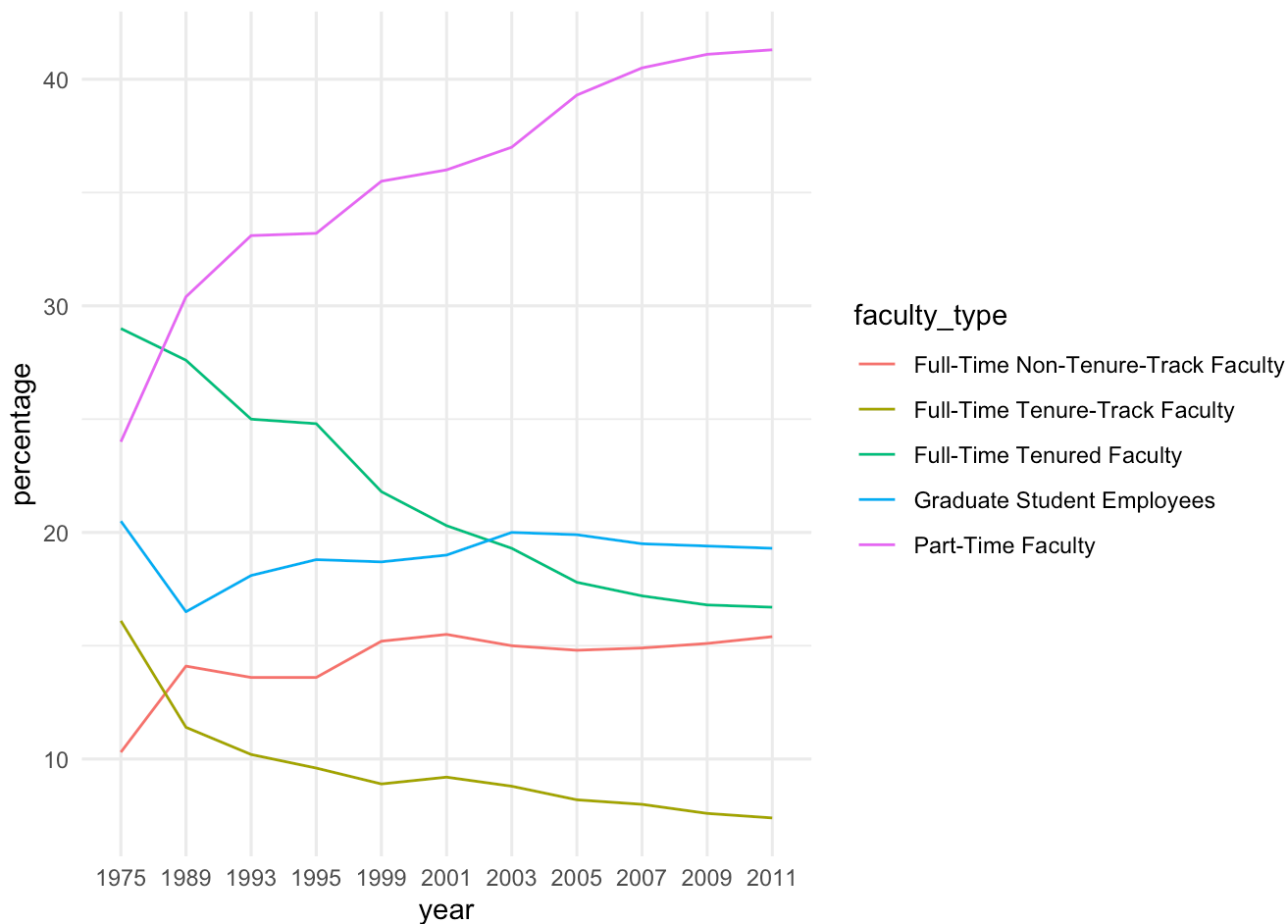
**Your turn:** Let's instead use a **line chart**. Graph the data with year on the  $x$ -axis and percentage of employees on the  $y$ -axis. Distinguish each faculty type using an appropriate aesthetic mapping.

```

staff_long |>
  ggplot(mapping = aes(
    x = year, y = percentage,
    group = faculty_type,
    color = faculty_type
  )) +

```

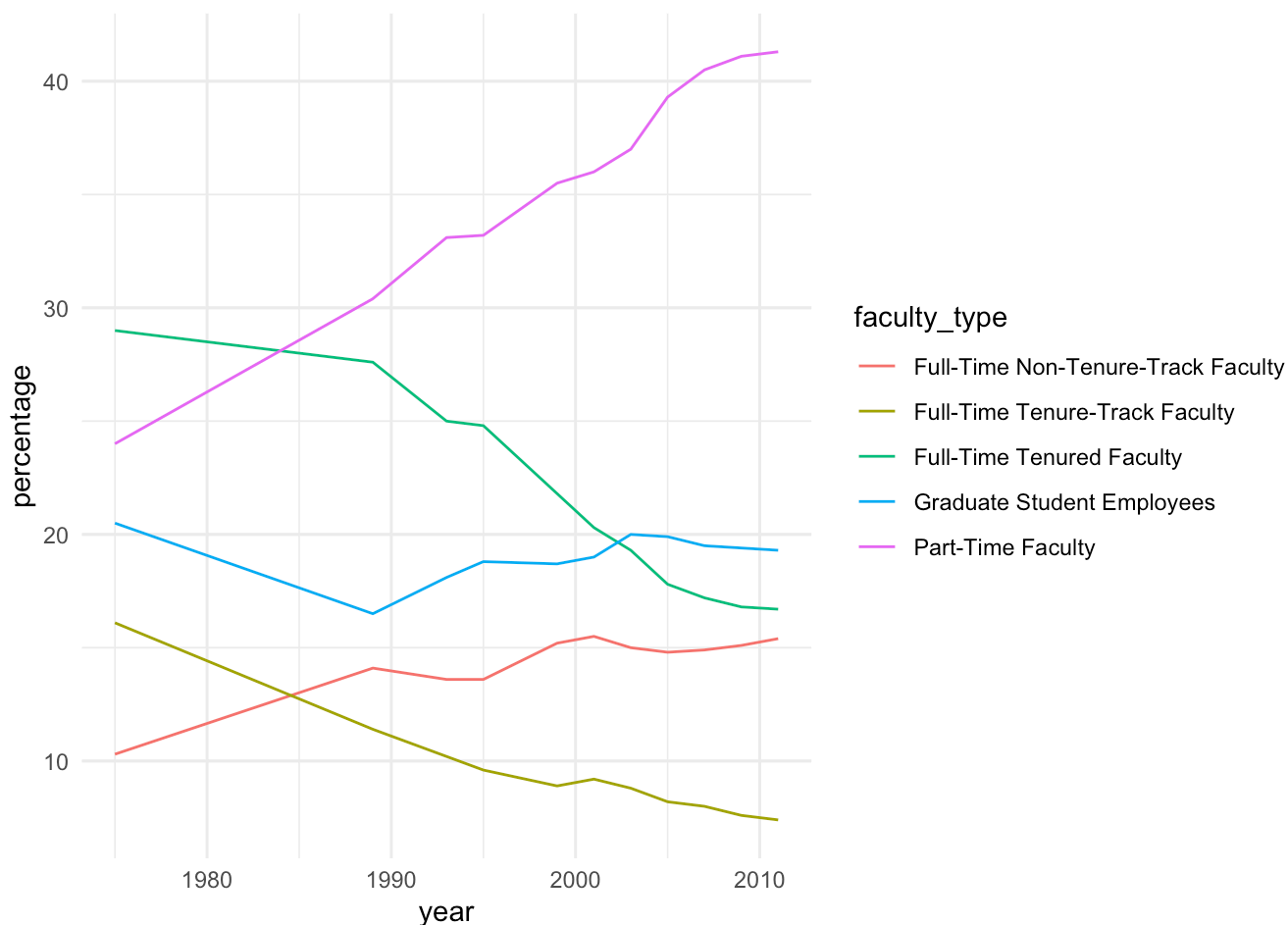
```
geom_line() +
theme_minimal()
```



Ooops, it still is equal intervals because we never ensured `year` was converted to a numeric variable after pivoting it. Let's fix that.

```
staff_long <- staff |>
  pivot_longer(
    cols = -faculty_type,
    names_to = "year",
    values_to = "percentage",
    names_transform = parse_number
  )

staff_long |>
  ggplot(mapping = aes(
    x = year, y = percentage,
    color = faculty_type
  )) +
  geom_line() +
  theme_minimal()
```



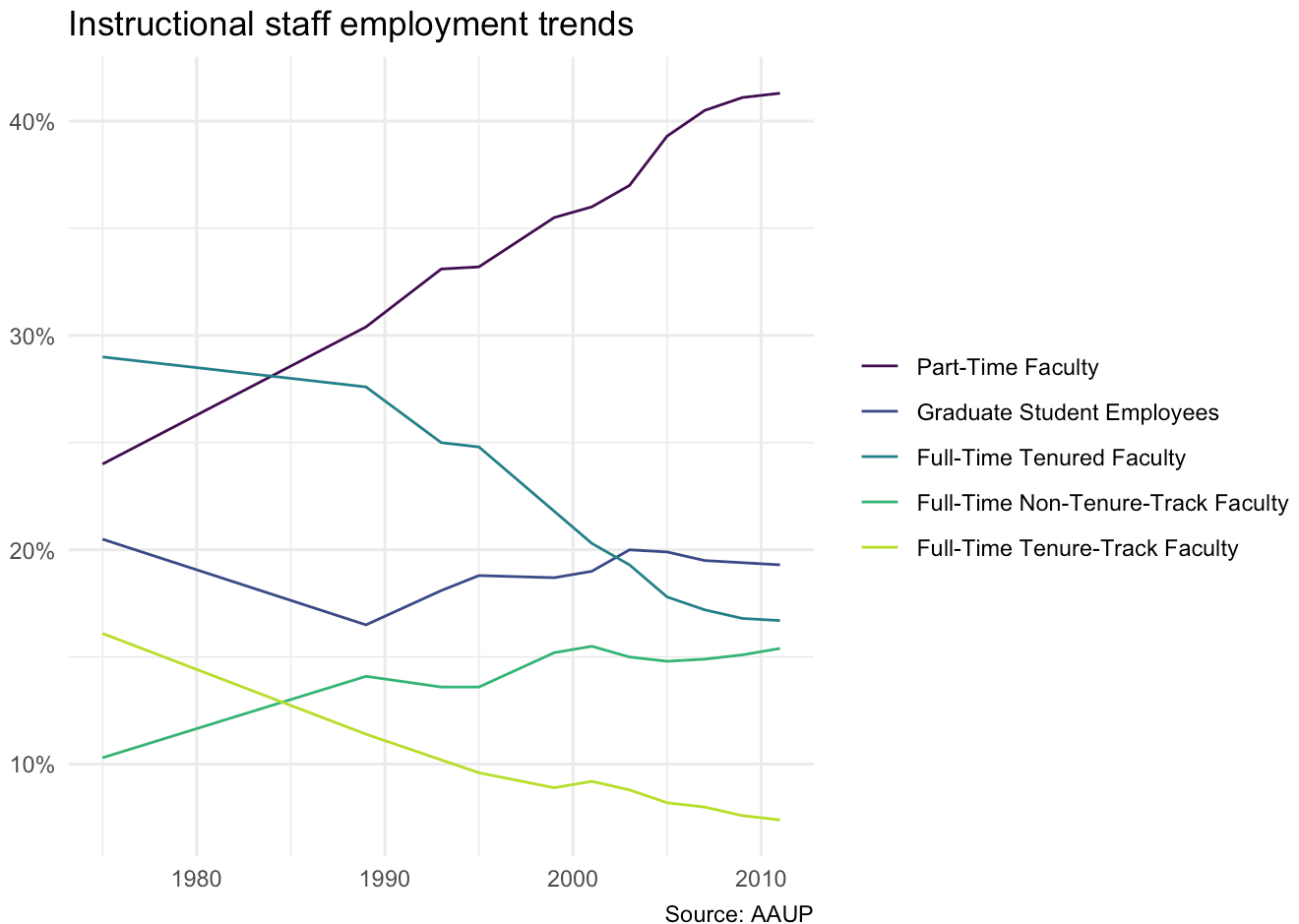
**Your turn:** Now we want to clean it up.

- Add a proper title and labelling to the chart
- Use an optimized color palette<sup>2</sup>
- Order the legend values by the final value of the `percentage` variable

```
staff_long |>
  mutate(
    faculty_type = fct_reorder2(
      .f = faculty_type,
      .x = year,
      .y = percentage
    )
  ) |>
  ggplot(
    mapping = aes(
      x = year, y = percentage,
      color = faculty_type
    )
  ) +
  geom_line() +
  scale_y_continuous(labels = label_percent(scale = 1)) +
```



```
scale_color_viridis_d(end = 0.9) +
labs(
  title = "Instructional staff employment trends",
  x = NULL, y = NULL, color = NULL,
  caption = "Source: AAUP"
) +
theme_minimal()
```



## Goal: even more improvement!

Colleges and universities have come to rely more heavily on non-tenure track faculty members over time, in particular part-time faculty (e.g. contingent faculty, adjuncts). We want to show **academia's reliance on part-time faculty**.

**Your turn:** With your peers, sketch/design a chart that highlights the trend for part-time faculty. What type of geom would you use? What elements would you include? What would you remove?

*Add response here.*

**Your turn:** Create the chart you designed above using {ggplot2}. Post your completed chart to [this discussion thread](#).

**Tip**

When you render the document, your plot images are automatically saved as PNG files in the `ae-05-sad-plot_files/figure-html` directory. You can use these images to post your chart to the discussion thread, or use the `ggsave()` function to directly save your plot as an image file. For example,

```
ggsave(  
  filename = "images/part-time-faculty.png",  
  plot = last_plot(),  
  width = 8, height = 6, bg = "white"  
)
```

saves the last generated plot to a file named `part-time-faculty.png` in the `images` directory. It has a defined height and width (in “inches”) with a white background.

# add code here

## Acknowledgments

- Exercise derived from [Data Science in a Box](#) and licensed under [CC BY-SA 4.0](#).

**Session information**

## Footnotes

1. Data sources: [IPEDS](#) and [Digest of Education Statistics](#). Downloaded February 10, 2025. [↩](#)
2. {viridis} is often a good choice, but you can find others. [↩](#)

Made with  and Quarto.

All content licensed under    [CC BY-NC 4.0](#).